STATISTICAL HORIZONS

Latent Class Analysis

Stephanie Lanza, Ph.D. and Bethany Bray, Ph.D.

Upcoming Seminar:
December 10-12, 2020, Remote Seminar

LATENT CLASS ANALYSIS

Statistical Horizons

STEPHANIE T. LANZA, PH.D.

- C. Eugene Bennett Chair in Prevention Research
- Director, Edna Bennett Pierce Prevention Research Center
- Professor, Biobehavioral Health
- The Pennsylvania State University

BETHANY C. BRAY, PH.D.

- Associate Director for Scientific Outreach, Center for Dissemination and Implementation Science
- Associate Professor, Medicine
- The University of Illinois at Chicago

LATENT CLASS ANALYSIS

(LCA)

LATENT CLASSES OF ADOLESCENT DRINKING BEHAVIOR

DRINKING IN 12TH GRADE

- Data from 2004 cohort of Monitoring the Future public release
- *n* = 2490 high school seniors who answered at least one question about alcohol use (48% boys, 52% girls)
- Goals of the study:
  - Alcohol use behavior among U.S. 12th graders
  - Gender differences in measurement and behavior
  - Predict behavior from skipping school and grades

DRINKING IN 12TH GRADE

Seven indicators of drinking behavior

Item	Proportion 'Yes'
Lifetime alcohol use	82%
Past-year alcohol use	73%
Past-month alcohol use	50%
Lifetime drunkenness	57%
Past-year drunkenness	49%
Past-month drunkenness	29%
5+ drinks in past 2 weeks	26%

WE WILL USE LCA TO…

- Identify and describe underlying classes of drinking behavior in U.S. 12th grade students

What would you name these 5 classes?

THE 5-CLASS MODEL

	Probability of 'Yes' response				
Item	Class 1 (18%)	Class 2 (22%)	Class 3 (9%)	Class 4 (17%)	Class 5 (34%)
Lifetime alcohol use	.00	1.00	1.00	1.00	1.00
Past-year alcohol	.00	.61	1.00	1.00	1.00
Past-month alcohol	.00	.00	1.00	.39	1.00
Lifetime drunk	.00	.24	.29	1.00	1.00
Past-year drunk	.00	.00	.00	1.00	1.00
Past-month drunk	.00	.00	.00	.00	.92
5+ drinks past 2 wk	.00	.00	.16	.00	.73

What would you name these 5 classes?

THE 5-CLASS MODEL

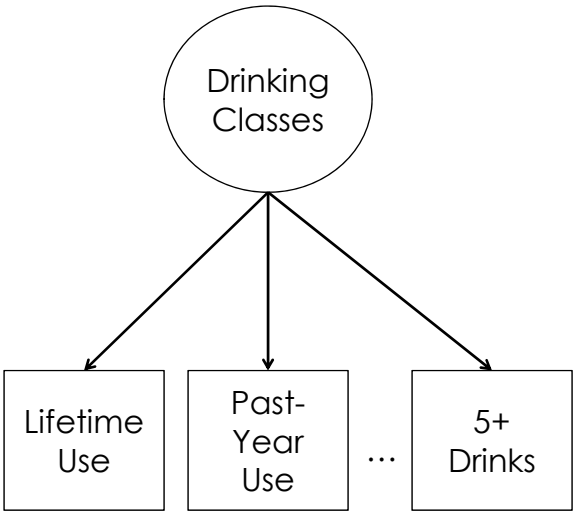
	Probability of 'Yes' response				
Item	Class 1 (18%)	Class 2 (22%)	Class 3 (9%)	Class 4 (17%)	Class 5 (34%)
Lifetime alcohol use		√	√	√	√
Past-year alcohol		√	√	√	√
Past-month alcohol			√		√
Lifetime drunk				√	√
Past-year drunk				√	√
Past-month drunk					√
5+ drinks past 2 wk					√

What would you name these 5 classes?

THE 5-CLASS MODEL

	Probability of 'Yes' response				
Item	Non-Drinkers	Experi-menters	Light Drinkers	Past Partiers	Heavy Drinkers
Lifetime alcohol use		√	√	√	√
Past-year alcohol		√	√	√	√
Past-month alcohol			√		√
Lifetime drunk				√	√
Past-year drunk				√	√
Past-month drunk					√
5+ drinks past 2 wk					√

GRAPHICAL REPRESENTATION



RESOURCES

- Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. New York, NY: Wiley.

- Lanza, S. T., Bray, B. C., & Collins, L. M. (2013). An introduction to latent class and latent transition analysis. In J. A. Schinka, W. F. Velicer, & I. B. Weiner (Eds.), *Handbook of Psychology* (2nd ed., Vol. 2, pp. 691-716). Hoboken, NJ: Wiley.

SOME TECHNICAL DETAILS: LCA

LATENT CLASS NOTATION

- ***Y*** represents the vector of all possible response patterns
  - ***y*** represents a particular response pattern
    - Example: ***y*** = (Y, Y, N, N, N, N, N)
- ***X*** represents the vector of all covariates of interest
  - ***x*** represents a particular covariate

LATENT CLASS NOTATION

- The latent class model can be expressed as

$$P(\mathbf{Y}_i = \mathbf{y}_i \mid \mathbf{X}_i = \mathbf{x}_i) = \sum_{c=1}^{C} \pi_c(\mathbf{x}_i) \prod_{m=1}^{M} \prod_{r=1}^{R_m} \rho_{mr_m|c}^{I(y_{im}=r_m)}$$

where

$$\pi_c(\mathbf{x}_i) = P(C_i = c \mid \mathbf{X}_i = \mathbf{x}_i) = \frac{\exp[\beta_{0c} + \beta_{1c} x_{i1} + \cdots + \beta_{Kc} x_{iK}]}{1 + \sum_{c=1}^{C-1} \exp[\beta_{0c} + \beta_{1c} x_{i1} + \cdots + \beta_{Kc} x_{iK}]}$$

LATENT CLASS NOTATION

…with (c = 1,2,…,K) latent classes and (m = 1,2,…,M) indicators, each with (r_m = 1,2,…,R_m) response options.

γ_c = probability of membership in latent class c (latent class membership probabilities)

$\rho^{I(y_m=r_m)}_{mr_m|c}$ = probability of response r_m to indicator m, conditional on membership in latent class c (item-response probabilities)

ITEM-RESPONSE PROBABILITIES

- ρ parameters express the relation between…
  - The discrete latent variable in an LCA and
  - The observed indicator variables
- Similar conceptually to factor loadings
  - Basis for interpretation of latent classes
- Are probabilities (between 0 and 1)

ITEM-RESPONSE PROBABILITIES

- p parameters analogous to factor loadings; both…
  - Express relation between manifest and latent variables
  - Form basis for interpreting latent structure

- But…
  - Factor loadings are β-weights
  - p parameters are <u>probabilities</u>

INTERPRETATION?

Probability of Correctly Performing Task	Latent Class 1	Latent Class 2
Task 1	Low	High
Task 2	Low	High
Task 3	Low	High
Task 4	Low	High
Task 5	Low	High

INCLUDING GROUPING VARIABLES

MULTIPLE-GROUPS LCA

- Two reasons to include a grouping variable:
- To explore measurement invariance
  - e.g., "Do the items map onto the latent construct in the same way for males and females?"
- To divide sample into groups for comparison purposes
  - e.g., "How does the probability of membership in the HEAVY DRINKERS latent class differ in the experimental and control groups?"

PREDICTING LATENT CLASS MEMBERSHIP

OUR DRINKING EXAMPLE

- Remember…
- Data from 2004 cohort of Monitoring the Future public release
- *n* = 2490 high school seniors who answered at least one question about alcohol use (48% boys, 52% girls)
- Goals of the study:
  - Alcohol use behavior among U.S. 12th graders
  - Gender differences in measurement and behavior
  - Predict behavior from skipping school and grades

OUR DRINKING EXAMPLE

	Probability of 'Yes' response				
Item	Non (18%)	Exper (22%)	Light (9%)	Past (17%)	Heavy (34%)
Lifetime alcohol use		√	√	√	√
Past-year alcohol		√	√	√	√
Past-month alcohol			√		√
Lifetime drunk				√	√
Past-year drunk				√	√
Past-month drunk					√
5+ drinks past 2 wk					√

WE WILL USE LCA TO…

- Identify and describe underlying classes of drinking behavior in U.S. 12th grade students
- Include a grouping variable (i.e., sex)
  - Test for measurement invariance across males and females
  - Examine sex differences in prevalence of behavior types
- Explore whether grades and skipping school predict drinking class membership

Exercise 1

Using the data provided (`exercise-1.sas`), fit a 4-class latent class model for marijuana use and attitudes using 7 indicators of the latent class variable. Use 4893 as the random seed. Interpret all parameters in the model.

The variables in `exercise-1.sas` are shown on the next page.

Optional (advanced SAS programming):

Plot the item-response probabilities using the SAS macro `LCAgraphicsV1.sas`. The macro has a user's guide that describes its use.

Hints:

(1) Save the macro file to your hard drive.

(2) Specify that path in an `%include` statement prior to running LCA.

(3) Execute the macro using the following syntax after running LCA:
`%ItemResponsePlot(ParamDataset=filename);`

LIFETIME	Frequency of marijuana use over participant's lifetime 1 = use 2 = no use	POL_BLF1	Political beliefs (conservative) 0 = not conservative 1 = conservative ((0,0) is liberal)
PREV_YR	Frequency of marijuana use over the previous year 1 = use 2 = no use	POL_BLF2	Political beliefs (moderate) 0 = not moderate 1 = moderate ((0,0) is liberal)
PREV_MO	Frequency of marijuana use over the previous month 1 = use 2 = no use	RLG_IMP1	Importance of religious beliefs (not important) 0 = important to some degree 1 = not important ((0,0) is very important)
NEXT_MO	How likely it is that the participant will use marijuana in the next year 1 = will use 2 = will not use	RLG_IMP2	Importance of religious beliefs (important) 0 = not important or very important 1 = important ((0,0) is very important)
APRV_TRY	Does participant disapprove of people trying marijuana one or twice 1 = do not disapprove 2 = disapprove	SKP_CLS	Number of skipped classes (0-25)
APRV_OCC	Does participant disapprove of people smoking marijuana occasionally 1 = do not disapprove 2 = disapprove	GRADE	Grades (on average, percent) (60-100)
APRV_REG	Does participant disapprove of people smoking marijuana regularly 1 = do not disapprove 2 = disapprove	GOOUT	Number of evenings out per week on average (0-7)
SEX	Gender 0 = male 1 = female	YEAR	Survey year 1 = 1999 2 = 2000 3 = 2001
RACE	Race/Ethnicity 0 = white 1 = non-white		

Data Summary, Model Information, and Fit Statistics (EM Algorithm)

```
Number of subjects in dataset:        2587
Number of subjects in analysis:       2587

Number of measurement items:             7
Response categories per item:            2 2 2 2 2 2 2
Number of groups in the data:            1
Number of latent classes:                4
Rho starting values were randomly generated (seed = 4893).

No parameter restrictions were specified (freely estimated).

The model converged in 108 iterations.

Maximum number of iterations: 5000
Convergence method: maximum absolute deviation (MAD)
Convergence criterion:  0.000001000

============================================
Fit statistics:
============================================

Log-likelihood:     -6512.18
G-squared:            253.06
AIC:                  315.06
BIC:                  496.66
CAIC:                 527.66
Adjusted BIC:         398.17
Entropy:                0.93
Degrees of freedom:       96

Test for MCAR
      Log-likelihood:     -6385.65
      G-squared:            222.25
      Degrees of freedom:      514
```

Parameter Estimates

(Standard errors could not be computed; please see the log file for details.)
Gamma estimates (class membership probabilities):

Class:	1	2	3	4
	0.1423	0.5197	0.2447	0.0932

Rho estimates (item response probabilities):

Response category 1:

Class:	1	2	3	4
LIFETIME :	1.0000	0.1199	1.0000	0.3097
PREV_YR :	0.9016	0.0000	1.0000	0.0000
PREV_MO :	0.2611	0.0000	0.7339	0.0000
NEXT_MO :	0.3058	0.0159	0.8749	0.2010
APRV_TRY :	0.6672	0.1658	1.0000	0.9807
APRV_OCC :	0.1978	0.0054	0.9975	0.9850
APRV_REG :	0.0206	0.0019	0.6077	0.4254

Response category 2:

Class:	1	2	3	4
LIFETIME :	0.0000	0.8801	0.0000	0.6903
PREV_YR :	0.0984	1.0000	0.0000	1.0000
PREV_MO :	0.7389	1.0000	0.2661	1.0000
NEXT_MO :	0.6942	0.9841	0.1251	0.7990
APRV_TRY :	0.3328	0.8342	0.0000	0.0193
APRV_OCC :	0.8022	0.9946	0.0025	0.0150
APRV_REG :	0.9794	0.9981	0.3923	0.5746