Introduction to Latent Class Analysis

Hsueh-Sheng Wu

CFDR Workshop Series

November 9, 2020

Outline

- What is Latent Class Analysis (LCA)

- Types of LCA models

- An example of LCA

- Select Stata syntax for LCA

- Stata LCA analysis examples

- Advantages and disadvantages of LCA

- Conclusions

What is Latent Class Analysis (LCA)

- LCA is a multivariate statistical technique estimating the number of unobserved distinctive groups in the population. Once distinctive groups are determined, researchers can examine how likely an individual belongs to each group and what factors differentiate these groups.

- LCA is commonly used when distinctive groups of respondents are in the population, but not variables in the data file readily identify them.

- The distinctive group membership can be inferred from the path coefficients between the latent class variable and the latent class indicators and/or between predictors and the outcome variables.

- LCA is an exploratory data analysis because the exact number of distinctive groups in the population is unknown and needs to be inferred from the data. Therefore, it is important to test a series of LCA model and report the results of the model that best fits the data.

- When a hypothesized LCA model fits the data well, it helps answer the questions such as how many distinctive groups of individuals are in the sample, how likely each respondent belongs to each group, and what variables can differentiate these groups.

- LCA is analyzed within the framework of generalized structural equation model (GSEM). Subsequently, LCA model can also be used for different types of indicators, combined with other kinds of analysis, and tested for different constrains impose on the model parameters.
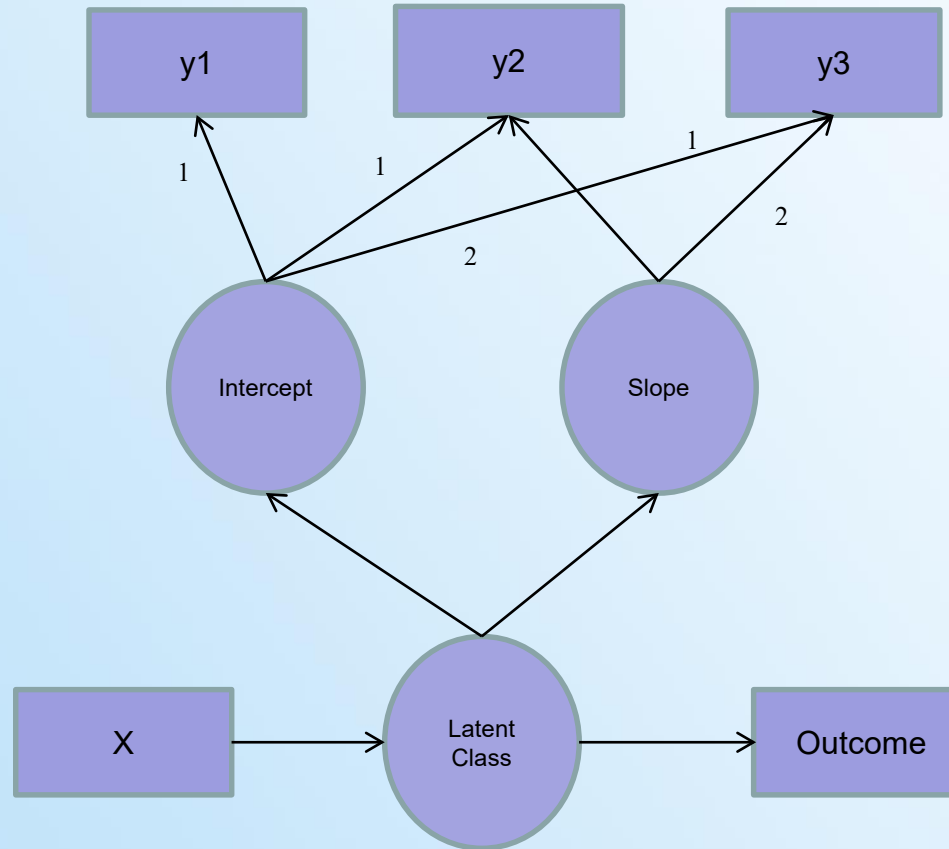
Types of LCA Models
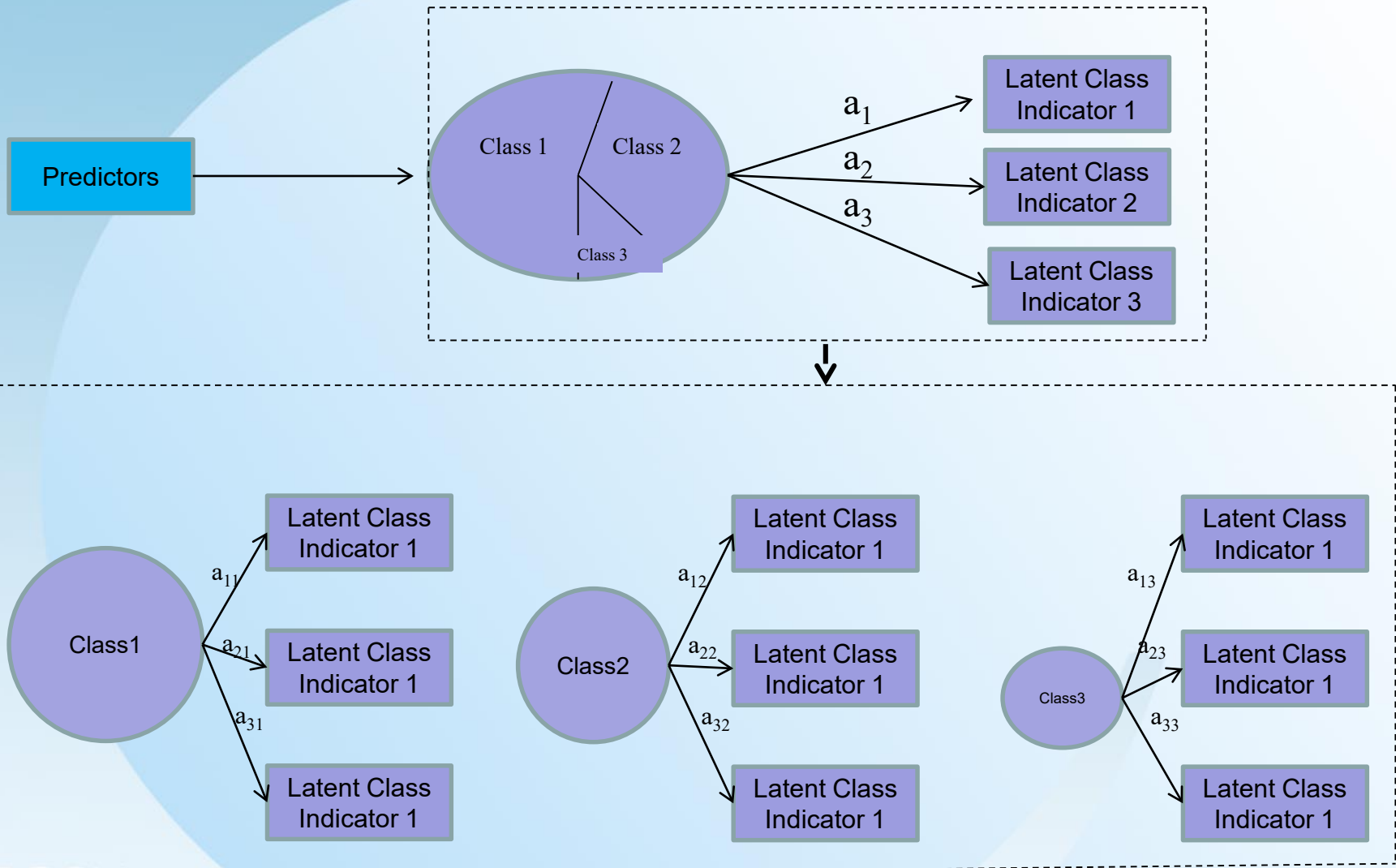
Note. Latent Class Analysis generally refers to models with binary latent class indicators, whereas Latent Profile Analysis refers to models with continuous latent class indicators.

Types of LCA Models
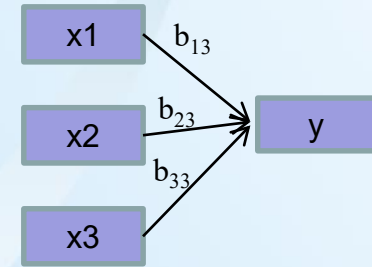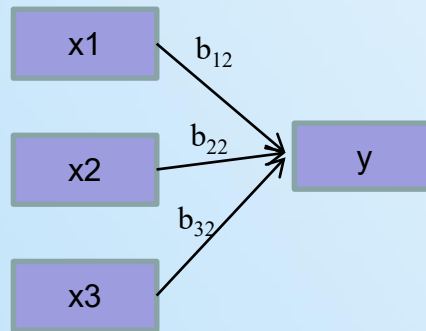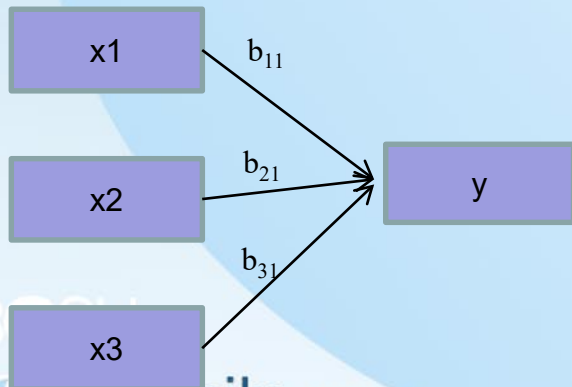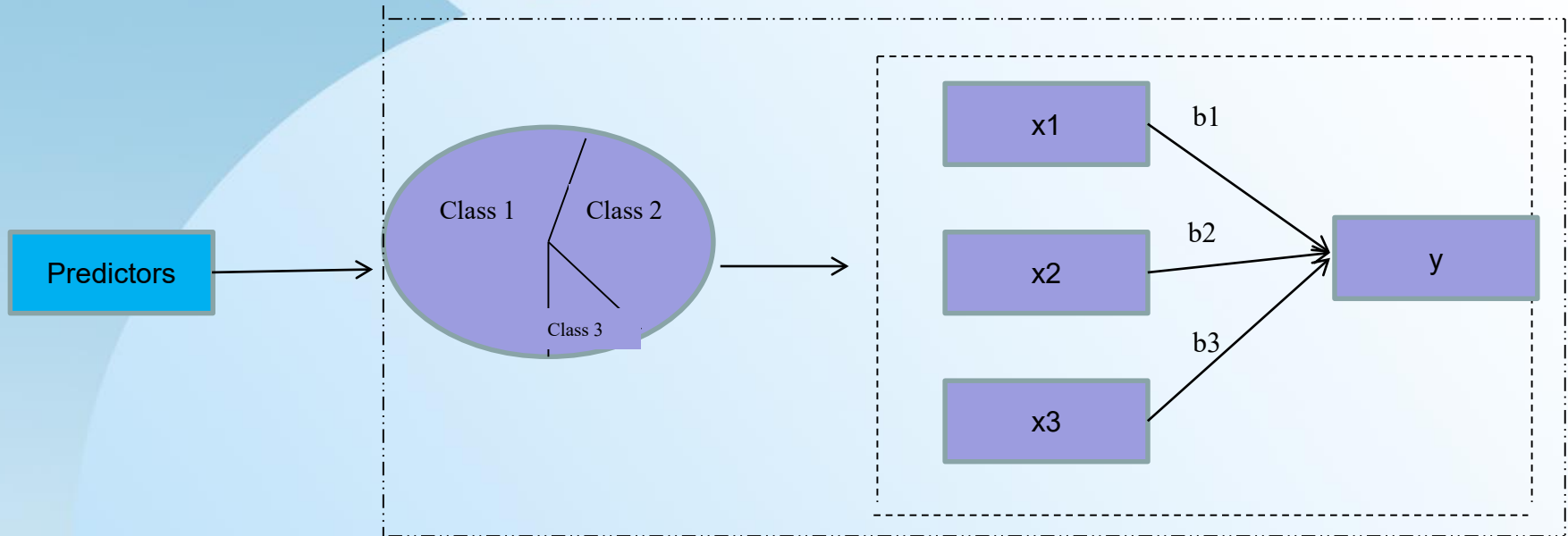
Types of LCA Models

Sources: This graph of growth mixture model with covariates is from Jung and Wickrama (2007)

An Example of LCA

- Lin and Wu (2014) used LCA to identify three groups of caregivers that use different strategies to cope with caregiving stress.[5]

- The authors argued that when using the sum of coping strategies to measure coping among caregivers, researchers cannot identify distinctive groups of caregivers using different sets of coping strategies nor examine what factors differentiate these distinct groups of caregivers. They proposed using LCA to overcome this problem.

- They examined how caregivers used 18 coping strategies. Ten of these strategies focused on reducing emotional stress, and 8 focused on solving caregiving-related problems. The examinations involved a series of LCA models with the number of latent classes ranging from 1 to 6.

- AIC and BIC showed the model with three latent classes were superior to other models in terms of parsimoniousness and the goodness of fit.

- Based on this thee-classes LCA model, the authors named these three groups of caregivers as (1) those not using particular coping strategies, (2) those using coping strategies to ameliorate emotional distress only, and (3) those using coping strategies to ameliorate emotional distress and solve problems.

- The authors further examined what factors differentiate these three groups of caregivers.

An Example of LCA

Table 3. Three-Latent-Class Model (N = 1,552)

	Latent Class		
	Unpatterned coping	Emotional coping	Hybrid coping
Probability of membership	.20	.46	.33
Conditional probability of a *Yes* response			
Emotion-focused coping			
Talk with friends or relatives	.22	.91	.92
Pray or meditate	.22	.85	.85
Watch TV	.08	.90	.86
Read	.04	.86	.81
Spend time alone	.14	.72	.75
Spend time on exercise or hobbies	.07	.67	.70
Eat	.05	.50	.61
Take medication to calm	.02	.13	.24
Drink alcohol	.01	.15	.18
Smoke	.03	.16	.14
Problem-focused coping			
Obtain assistive devices	.45	.41	.89
Use personal or nursing care	.21	.08	.63
Make home modifications	.16	.15	.55
Use services to help with housework	.14	.08	.36
Use meal deliver services	.05	.04	.24
Use outside services for transportation	.07	.04	.22
Use respite or caregiver support services	.04	.02	.27
Request information on financial help	.03	.07	.20

Select Stata Syntax for LCA

Specification of different LCA models

- A LCA model with 4 binary latent class indicators and 2 latent classes

    gsem (y1 y2 y3 y4 <-), logit lclass(C 2)

- A LCA model with 4 binary latent class indicators and 3 latent classes

    gsem (y1 y2 y3 y4 <-) , logit lclass(C 3)

- A LCA model with 4 binary latent class indicators, 3 latent classes, and 1 predictor of the latent class variable

    gsem (y1 y2 y3 y4 <-) (C <- x1), logit lclass(C 3)

- A LCA model with 4 binary latent class indicators, 3 latent classes, and 1 predictor of the latent class variable. The whole model varies by the level of a group variable

    gsem (y1 y2 y3 y4 <-) (C <- x1), logit lclass(C 3) group(group)

BGSU

Center for Family and Demographic Research

Select Stata Syntax for LCA

- A LCA model with 4 continuous latent class indicators (i.e., latent profile analysis) and 2 latent classes

    gsem (y1 y2 y3 y4 <-), lclass(C 2)

- A LCA model with 4 continuous latent class indicators, 3 latent classes, and 1 predictor of the latent class variable. The whole model varies by the level of a group variable

    gsem (y1 y2 y3 y4 <-) (C <- x1), lclass(C 3) group(group)

- A LCA model with 4 count latent class indicators and 2 latent classes

    gsem (y1 y2 y3 y4 <-), poisson lclass(C 2)

- A LCA model with different types of latent class indicators and 2 latent classes

    gsem (y1 y2 <-, logit) (y3 <-, poisson) (y4 <-), lclass(C 2)

Select Stata Syntax for LCA

Post-Estimation commands used after the LCA model is fitted

- Show the goodness of fit statistics
    estat lcgof

- Store the goodness of fit statistics
    estimates store model_name

- Show the stored goodness of fit statistics
    estimates stats model_name

- Show the predicted probability of latent classes
    estat lcprob

- Showed the predicted means of observed variable
    estat lcmean

- generate the predicted probabilities for each latent class
    predict classpost*, classposteriorpr

BGSU

Center for Family and Demographic Research

Stata Syntax for LCA

- Use the line show the predicted probability of the latent classes
    margins, predict(classpr class(1)) predict(classpr class(2)) ///
    predict(classpr class(3))

    marginsplot, xtitle ("") ytitle ("")   ///
    xlabel (1 "Class 1" 2 "Class 2" 3 "Class 3") ///
    title("Predicted Latent Class Probabilities with 95% CI")

- Use the Bar chart to show the predicted probability of the latent classes
    margins, predict(classpr class(1)) predict(classpr class(2)) ///
    predict(classpr class(3))

    marginsplot, recast(bar) xtitle ("") ytitle ("") ///
    xlabel (1 "Class 1" 2 "Class 2" 3 "Class 3")   ///
    title("Predicted Latent Class Probabilities with 95% CI")

Stata Syntax for LCA

- Show the predicted probability of the outcome variable for class 1

```
margins, predict(outcome(y1) class(1)) ///
predict(outcome(y2) class(1)) ///
predict(outcome(y3) class(1))

marginsplot, recast(bar) xtitle ("") ytitle ("") ///
xlabel (1 "outcome 1" 2 "outcome 2" 3 "outcome 3") ///
title("Predicted Probability of Outcomes For Class 1 with 95% CI")
```

LCA Examples

- See the Stata log file

Advantages of LCA

- LCA is a statistical method examining the heterogeneity of the population by identifying distinctive, but unmeasured, classes of subjects in the population

- LCA is a type of generalized linear model, so it can analyze both categorical and/or continuous observed variables

- LCA can be combined with other techniques such as measurement models and structural models to better understand whether and how population heterogeneity is associated with results of these models

Disadvantages of LCA

- LCA is an exploratory data analysis technique because researchers do not exactly know how many distinctive groups of respondents are in the population. Therefore, it is important to test LCA models with different numbers of latent classes and choose the one that fits the data best.

- LCA is a classification technique that groups subjects in a particular data set. If the data set does not contain a sufficient number of a certain group of respondents, LCA will not be able to identify this group.

- The LCA result is determined by what items are used and what models are specified. If the items are not good indicators of the class membership or the LCA model is mis-specified for the relations among the items, the LCA result will be inaccurate.

- The identification of the latent classes are not theory-based. Thus, researchers need to carefully examine these latent classes and consider whether these latent classes are theoretically sensible.

Conclusions

- LCA is a statistical technique for identifying unmeasured class membership in the population with the use of categorical and/or continuous variable.

- LCA is a type of exploratory data analysis and the accuracy of its results is conditioned on the sample, variables, the model specification used. Researchers need to ascertain the LCA model is not mis-specified. When the sample size is sufficiently large, researchers can randomly divide the sample into two sub-samples. One sub-sample is used to for identify the LCA model that fits the data, and the other is for cross-validating the identified LCA model.

- LCA helps identify distinctive groups of respondents in the population. However, it is researchers' responsibility to use existing literature to evaluate whether the groupings is theoretically sensible,

- In practice, researches often find their hypothesized LCA models do not converge. Subsequently, they may need to try different sets of starting values and see if the models will converge (see pages 131-133, 495,517-518 of Stata Structural Equation Modeling Reference Manual Release 16).

- LCA can be combined with other statistics techniques. However, some of complex LCA models can only be estimated in Mpus, rather than in Stata. Students and researchers can download the demonstration version of Mplus (https://www.statmodel.com/demo.shtml) and test whether these complex models can be fitted.

Center for Family and Demographic Research