Using Latent Class Analysis as the Basis for Interventions within Higher Education

Cassie West
University of North Texas

What we will cover

- A change in the way we approach data
- Types of analyses
- Highlight one particular method, Latent Class Analysis
- Examples in higher education literature
- Examples in practice

Variable-Centered Approaches

- Assume that the associations among variables are consistent across the population
  - For example: the relationships between common risk factors and retention are the same across all FTIC
- Goal: examine relationships among variables or variance in outcome variables based on predictor variables
- Include: correlation, regression, ANOVA, ANCOVA, t-tests, Chi-square, principle components and factor analysis, structural equation modeling, etc.
- Dominate the statistical landscape in higher education and other social sciences

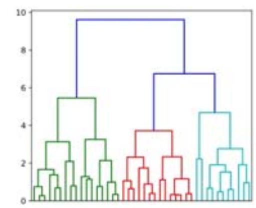
Person-Centered Approaches

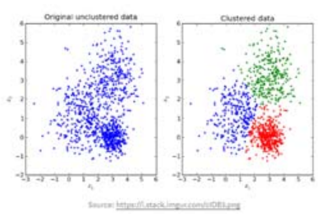
- Assume that the process of how the predictor variables affect the outcome is heterogeneous across population
  - For example: there are subgroups within the FTIC population for which the relationships between the risk factors and retention differ
- Goal: identifying groups of individuals that are qualitatively different from one another but share similar attributes/profiles/experiences
- Can explore individual/group differences in the pattern of responses
  - How the subgroups are structured by demographic variables
- Include: discriminant analysis, logistic regression, cluster analysis, and mixture modeling
  - Known groups verses unknown groups
- Underutilized in Higher Education

Types of Person-Centered Approaches

Cluster Analysis:

- Goal: partition the data into groups based on the observed variables
  - Assume that the subgroups are driven by the observable variables
  - Hierarchical clustering algorithms (nested)
    - Agglomerative or divisive methods
    - Taxonomies
  - Partitional clustering algorithms (un-nested)
    - K-means
    - Minimize squared deviations within a given cluster and maximize distance between the clusters
- Limitations of traditional cluster analysis:
  - Sample dependent and sensitive to the initial cut/decision
  - Cases cannot be reassigned, even if a case is closer to another group at the end of the process
  - Labor intensive to settle on a consistent solution and number of clusters
  - Heuristic in nature and researcher subjectivity may introduce bias





Types of Person-Centered Approaches

Mixture Modeling:

- Under partitional clustering algorithms
- Model based approach to partitioning the data
  - Not simply partitioning the data space
- Goal: determine a model that minimizes homogeneity within subgroups/classes and maximizes heterogeneity between classes
  - Assume the partitions are driven by latent groups
  - Analysis is probabilistic: able to estimate the parameters of the model and prevalence rates of each cluster

Table 1. Taxonomy of models with latent classes

Model Name	*Variable Type*	*Cross-sectional/ Longitudinal*	*Within-Class Variation*
Latent Class Analysis (LCA)	Categorical	Cross-sectional	No
Latent Profile Analysis (LPA)	Continuous	Cross-sectional	No
Latent Transition Analysis (LTA)	Categorical	Longitudinal	No
Mixture Factor Analysis	Categorical Continuous	Cross-sectional	Yes
Mixture Structural Equation Modeling	Categorical Continuous	Cross-sectional	Yes
Growth Mixture Modeling (GMM)	Categorical Continuous	Longitudinal	Yes

- Advantages:
  - Allows for flexibility of group membership
  - Criteria for model selection
  - Model based (generalizable)

Marcoulides, G., & Heck, R. (2013). Mixture models in education. In T. Teo (Ed.), Handbook of quantitative methods for educational research, (pp. 347-366). Rotterdam: SensePublishers.

Latent Class Analysis: Overview

- Cross-sectional categorical data (dichotomous, ordered-category, nominal)
- An extension of K-means clustering
- Widely used in health fields and disease diagnosis (clusters of symptoms and disease subtypes/diagnostic subcategories)
- Goals:
  - Arrive at an array of latent classes that represents the response patterns
    - Provide a prevalence for each class and the error associated with each variable measuring the latent classes
  - Identify items that indicate the classes well and estimate the probabilities
  - Identify any covariates that help explain class membership, and classify individuals correctly within each latent class
- Like a factor analysis- but with people!

Latent Class Analysis: Important Concepts

- Assumption of Local Independence (items are independent)
  - The relationships between the variables are driven by the latent groups
- Contingency table, Sparseness
  - Every possible combination of responses (ex. 9 dichotomous variables: 2^9=512)
- Item response probabilities
- Class homogeneity
- Class separation
- Class membership prevalences
- Identification
  - Sample size and W, N/W >5
  - Different seeds leads to different solutions= poorly identified
  - ML solutions
- Model fit
  - G^2, AIC, BIC/ABIC

Latent Class Analysis: Important Concepts

- Variable considerations:
  - Try to limit to 12 or fewer indicators, even less if more than 2 response categories
  - Need to have a good split between the levels, do not want 90/10
  - Be cautious when changing a continuous variable into a categorical- fundamental loss in variance
    - Collins and Lanza (2010) especially do not recommend that you code a continuous variable into low/medium/high
- Making resources go further
  - Better understanding of our students and their needs can lead to targeted interventions that use limited resources efficiently

Latent Class Analysis in Higher Education

- Boscardin, C. (2012). Profiling students for remediation using latent class analysis. *Advances in Health Sciences Education, 17*, 55-63.
  - Identified an additional 24% of students needing remediation and their specific areas of deficit
- Denson, N., & Ing, M. (2014). Latent class analysis in higher education: An illustrative example of pluralistic orientation. *Research in Higher Education, 55*, 508-526.
  - Looked at skills and disposition toward pluralistic orientation, groups can inform intervention strategies
- Malcom-Piqueux, L. (2015). Application of person-centered approaches to critical quantitative research: Exploring inequities in college financing strategies. *New Directions for Institutional Research, 2014*(163), 59-73.
  - Looked at how students pay for college, created profiles and demonstrated how they were structured by race/ethnicity

Latent Class Analysis in Higher Education

- Fematt, V., Grimm, R. P., Nylund-Gibson, K., Gerber, M., Brenner, M. B., & Solórzano, D. (2019). Identifying transfer student subgroups by academic and social adjustment: A latent class analysis. *Community College Journal of Research and Practice*, DOI: 10.1080/10668926.2019.1657516
  - Looked for meaningful subgroups of transfer students based on their response patterns to measures of academic and social adjustment, first year student success course intervention participation
- Gray, C. (2019). Using profiles of human and social capital to understand adult immigrants' education needs: A latent class approach. *Adult Education Quarterly, 69*, 3-23.
  - Examined the existing human and social capital for adult immigrants, found mismatches between the immigrants' needs and the available adult education resources

Example 1: College Mothers

- Original research examining the relationships between attachment, coping skills, parental and academic self-efficacy, and multiple role stress
- Use the sample to see if there are different subgroups of moms and if they differ on efficacy and stress measures
- How might a university address the needs of different populations of student mothers?

Example 1: College Mothers

- N=343
- Variables:

Variable	Levels		
Employment	Part-Time, 43%	Full-Time, 21%	Not Working, 36%
Traditional Age	Yes, 50%	No, 50%	
Married	Yes, 60%	No, 40%	
1st Generation	Yes, 50%	No, 50%	
Number of Kids*	Single, 68%	Multiple, 32%	
Average Kid Age	Preschool, 57%	School Aged, 43%	
Break between High School and College	Yes, 47%	No, 53%	
Low Income	Yes, 62%	No, 38%	
Course Load*	Part-Time, 20%	Full-Time, 80%	
Classification	Lower, 51%	Upper, 49%	

- Contingency table=1536
  - 343/1536 (what's wrong here?)

Example 1: College Mothers

- 4 Class model fit the best
- Classification and Course Load were not useful in differentiating the groups
- What shall we name them?

Variable	Class 1 (24.19%)	Class 2 (19.59%)	Class 3 (45.84%)	Class 4 (10.37%)
Married	Married	Not married	Not married	-
Break	Took time off	Took time off	No time off	-
Age	Non-traditional age	Non-traditional age	Traditional age	-
Number of Kids	2+ kids	2+ kids	1 kid	1 kid
Average Kid Age	School aged children	School aged children	Preschool aged child	Preschool aged child
Low Income	Not low income	Low Income	(Low income)	Low income
Employment	-	-	(Part time work)	Not working
1st Generation	-	-	-	1st Generation

- No significant difference among the classes in Parental Self-Efficacy or Multiple Role Stress
- Significant differences between Class 1 and Class 3 in Academic Self-Efficacy (Class 3 lower)

Example 2: College Stop-Outs

- Students who reached Senior level at UNT and stopped out
- Advisors wanted to conduct an intervention to re-enroll the students
- Over 3000 students in the last 4 years!
- How do you reach out to that many students?
- Where do you focus resources?

Example 2: College Stop-Outs

- N=2977
- Variables:

Variable	Levels	
Coursework Level	Senior (51.39%)	Lower (48.61%)
Academic Decline	Yes (49.01%)	No (50.99%)
Excess Hours*	Yes (27.11%)	No (72.89%)
Non-Traditional Age	Yes (50.66%)	No (49.34%)
1st Generation	Yes (43.27%)	No (56.73%)
Admit Type*	Native (23.14%)	Transfer (76.86%)
Academic Standing*	Good (67.95%)	Poor (32.05%)
Course Load*	Full-time (39.84%)	Part-time (60.16%)
Pell Grant Recipient	Yes (34.46%)	No (65.54%)

- Contingency table=512
  - 2977/512 (much better!)

Example 2: College Stop-Outs

- Although a 5 class model fit better, the solutions were unstable
  - Identification issues
- A 4 class model was chosen
- Pell Grant and 1st Generation were not useful in differentiating the groups
- Admit Type was removed and examined as a grouping variable

Variable	Class 1 (17.43%)	Class 2 (10.51%)	Class 3 (33.45%)	Class 4 (38.62%)
Coursework Level	-	Senior Level	-	-
Academic Decline	-	Yes	(No)	-
Excess Hours	-	(No)	(No)	No
Non-Traditional Age	Yes	-	Yes	No
Academic Standing	Poor	-	Good	Good
Course Load	Part-time	Full-time	Part-time	-
Native	3.68%	7.34%	15.17%	73.81%
Transfer	20.47%	8.77%	38.33%	32.47%

Additional Resources/References

- Curran-Bauer Analytics, *"Introduction to latent class/profile analysis,"* https://www.youtube.com/watch?v=NIZFm_EI8OM
- Collins, L., & Lanza, S. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. Wiley.
- DiStefano, C. (2012). Cluster analysis and latent class clustering techniques. In B. Laursen, T. Little, & N. Card (Eds.), *Handbook of developmental research methods* (pp. 645-666). New York, NY: The Guilford Press.
- DiStefano, C., & Mindrila, D. (2013). Cluster analysis. In T. Teo (Ed.), *Handbook of quantitative methods for educational research*, (pp. 103-122). Rotterdam: SensePublishers.
- Lanza, S., & Rhoades, B. (2013) Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. *Prevention Science, 14*, 157-68.
- Lanza, S., Bray, B., & Collins, L. (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., pp. 691-716). Hoboken, NJ: Wiley.

If you have any difficulty locating these resources, please let me know!

I have them all in PDF.

Additional Resources

- Website for SAS PROC LCA/LTA plugin and additional resources:

https://www.methodology.psu.edu/downloads/proclcalta/

- Users Guide: https://www.methodology.psu.edu/files/2019/03/proc_lca_lta_1-3-2-1_users_guide-2ggq4d3.pdf

SAS Coding

- When coding your categorical variables start the levels at 1, do not use 0
  - Example: 1=Native, 2=Transfer
- Basic coding structure:

***4 classes from dataset named LCA;**

PROC LCA DATA=LCA OUTPARAM=param1 OUTPOST=post1 OUTEST=est1

OUTSTDERR=stderr1; *these statements will output your parameters and predicted classed;

NCLASS **4**; *indicates the number of classes;

ID EMPLID; *keeps your ID variable in the output data, you can add multiple variables here;

ITEMS SR_LEVEL DECLINE EXCESS NON_TRAD FULLTIME GOOD_STAND; *list of your items to be included in the model;

CATEGORIES **2 2 2 2 2 2** ; *list of the number of levels for each variable in the analysis, must have same order as previous line;

SEED **4875**; *random number;

NSTARTS **100**; *this statement will iterate the model 100 times, will tell you how many times you land on the same solution- important for identification;

RHO PRIOR=**1**; *this statement will ensure that you get standard errors estimated;

run;

SAS Code with grouping variable

<u>*4 classes from dataset named LCA, with a grouping variable;</u>

```
PROC LCA DATA=LCA OUTPARAM=param1 OUTPOST=post1 OUTEST=est1
OUTSTDERR=stderr1; *these statements will output your parameters and predicted classed;
 NCLASS 4; *indicates the number of classes;
 ID EMPLID; *keeps your ID variable in the output data, you can add multiple variables here;
ITEMS SR_LEVEL DECLINE EXCESS NON_TRAD FULLTIME GOOD_STAND; *list of your items to be included in the model;
CATEGORIES 2 2 2 2 2 2 ; *list of the number of levels for each variable in the analysis, must have same order as previous line;
GROUPS ADMIT; *grouping variable;
GROUPNAMES NATIVE TRANSFER ; *grouping variable levels named, make sure they match your numeric coding (1=Native, 2=Transfer);
MEASUREMENT GROUPS; *indicates measurement invariance;
SEED 4875; *random seed, will need to match previous runs if trying to ensure you replicate the same groups;
RHO PRIOR=1; *this statement will ensure that you get standard errors estimated;
RUN;
```