Latent Class/Profile Analysis Handout

Author: Stephanie Agtarap, PhD updated 8.23.2024

Appropriateness

	Considerations	Notes
What kind of data do I have?	LCA = discrete/categorical cross-sectional data LPA = ordinal/continuous data cross-sectional data Latent growth curve analysis (LGCA) or Latent growth mixture modeling (LGMM) = longitudinal data	
Is my sample size large enough?	N >300 without Monte Carlo simulations	
Does my data meet assumptions of normality?	Local independence of characteristics Normality assumptions	

Generate Hypotheses

	Considerations	Notes
What is my research question?	Do you think outcomes may uniquely differ by a combination of different characteristics?	
How would I define my profiles?	What discrete/continuous indicators independently impact a targeted outcome? How would the patterns of these indicators and/or characteristics inform outcomes above and beyond current single indicators or systems (e.g., diagnostic criteria)?	

Data Cleaning/Setup

	Considerations	Notes
Transformation of scales	Consider for extreme scales differentials (e.g., z-scores)	
Uniform directionality	Ease of interpretation	

Robust categorical data	<10% in one level could be collapsed	
Meeting parametric assumptions	Transformations for normal distribution	
Indicator collinearity	Correlations with indicators	
Missing Data	FIML or other imputation method commonly used	

For Mplus

1. Whatever data you'll be using to identify your subgroups needs to be in worksheet file and formatted according to Mplus rules
    a. No titles in data
    b. Shorter the better → ID, data columns of interest

Estimate Models

2. Mplus script for LPA pretty straightforward:
    a. ANALYSIS
        i. defined as **mixture** model
        ii. Lrtstarts → can be very important if analysis does not converge initially
    b. MODEL
        i. If necessary, can play with model (i.e., mixture model) hold a class variance equal, etc. This is more advanced.
    c. OUTPUT
        i. Tech11 = LRT/Adjusted LRT
        ii. Tech14 = Bootstrapped LRT
        iii. Tech8 = troubleshooting –> can locate areas in covariance matrix with issues, negative values, etc.
    d. PLOT
        i. "type is plot3;"
        ii. "series is"→ this will show each class' average value by each item you entered into the LPA. YOU must define at what x-axis value (e.g., 1.0, 2.0, 3.0, etc.) these will appear.
    e. SAVEDATA
        i. This will save class membership data into a NEW csv file. Original data and ID will also be included.
        ii. Would suggest not doing this until you've decided on the number of classes you choose to retain.

Evaluate Models

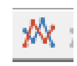
3. **Assessing overall fit of class solutions is first.** In other words: you will manually run models that force the data to fit 1-*k* classes (as desired) and compare their overall fit. Fit is based on numerous factors (See Table below).

Deciding on Number of Classes to Retain

Metric	Description	Standard Threshold	Relative Importance
AIC/BIC/SSBIC	Metric of relative fit of solution with K classes, compared to previous (K-1) model.	Lower = better; Greatest relative change = better;	SSBIC considered most reliable relative index.
Entropy	Indication of “clear delineation of cases”	Closest to 1.00 < .80 considered problematic	Fairly important
Vuong Lo-Mendell Rubin Test / Adjusted Lo-Mendell-Rubin Test	Compares the model with K classes to a model with (K-1) classes.	p < .05 = K classes provides sig. more information than K-1 classes	Adjusted LRT considered reliable sig test index
Boostrapped Lo-Mendell Rubin Test	“ ” (as above) + bootstrapped method applied	“ ” (as above)	Most reliable; better adjusted LRT, but with large *n* usually always significant
Class Sizes (N)	Proportion of your sample that belong in each class	Class sizes < 10% of full sample should be scrutinized as necessary	Practical importance
Class Probability	Average probability of observed sample being in assigned class	Ideally above 90% (need to verify)	Practical importance
Theoretical Sense	Do the classes make sense?	Common sense	Practical importance
Parsimony	Are additional classes really necessary?	Evaluation of all above → if one metric doesn’t hold up, tend to stay at that class solution	Practical importance

4. Analyses will be an **iterative process**. I’d suggest a separate spreadsheet where you can compare the overall fit indices for each class next to each other.

Interpret Optimal Model

5. **Interpretation of given class solution comes *after* you've identified your optimal model.**
    a. Remember *z*-score mean = 0.
        i. Significance tests for latent means can be interpreted as *this class' mean outcome score is significantly different from the mean outcome score of the overall/homogenous sample.*
    b. Generally looking for profiles with *interactions,* i.e., telling a non-linear story across given outcomes.
    c. Interpretation is *relative/specific* to the given sample (e.g., 'higher depression scores' rather than 'high depression'). Be careful not to interpret classes as significantly different from each other.
    d. **In Mplus, you can see visual plots from this button →** .
        i. Greatly assists interpretation. I prefer seeing "*Estimated means and observed individual values*" → "*show all curves in one window*" first, and then look at "*each class per window*".

Examples

Brett, B. L., Kramer, M. D., Whyte, J., McCrea, M. A., Stein, M. B., Giacino, J. T., ... & TRACK-TBI Investigators. (2021). Latent profile analysis of neuropsychiatric symptoms and cognitive function of adults 2 weeks after traumatic brain injury: findings from the TRACK-TBI study. *JAMA Network Open, 4*(3), e213467-e213467.

Faulkner, J. W., Snell, D. L., Theadom, A., de Terte, I., & Low, R. (2023). Patterns of post-traumatic stress symptoms in mild traumatic brain injury and their relationship with outcomes: a latent profile analysis. *Journal of Clinical and Experimental Neuropsychology, 45*(3), 230-241.

Juengst, S. B., Agtarap, S., Venkatesan, U. M., Erler, K. S., Evans, E., Sander, A. M., ... & Whiteneck, G. G. (2024). Developing multidimensional participation profiles after traumatic brain injury: a TBI model systems study. *Disability and Rehabilitation, 46*(11), 2385-2395.

Juengst, S. B., Kumar, R. G., Venkatesan, U. M., O'Neil-Pirozzi, T. M., Evans, E., Sander, A. M., ... & Whiteneck, G. G. (2024). Predictors of multidimensional profiles of participation after traumatic brain injury: a TBI model systems study. *The Journal of Head Trauma Rehabilitation,* 10-1097.

Keenan, H. T., Clark, A. E., Holubkov, R., Cox, C. S., Patel, R. P., Moore, K. R., & Ewing-Cobbs, L. (2020). Latent class analysis to classify injury severity in pediatric traumatic brain injury. *Journal of neurotrauma*, *37*(13), 1512-1520.

Qiu, H., Zador, Z., Lannon, M., Farrokhyar, F., Duda, T., & Sharma, S. (2024). Identification of clinically relevant patient endotypes in traumatic brain injury using latent class analysis. *Scientific Reports, 14*(1), 1294.

Starosta, A. J., Mata-Greve, F., Humbert, A., Zheng, Z., Prado, M. G., Au, M. A., ... & Hoffman, J. M. (2023). Intersectionality of systemic disadvantage on mortality and care following TBI. *The Journal of Head Trauma Rehabilitation, 38*(2), 137-146.

Additional References/Material

Ferguson, S. L., G. Moore, E. W., & Hull, D. M. (2020). Finding latent groups in observed data: A primer on latent profile analysis in Mplus for applied researchers. *International Journal of Behavioral Development*, *44*(5), 458-468.

Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, *4*(4), 440.

Oberski, D. (2016). Mixture models: Latent profile and latent class analysis. *Modern Statistical Methods for HCI*, 275-287.

Sinha, P., Calfee, C. S., & Delucchi, K. L. (2021). Practitioner's guide to latent class analysis: methodological considerations and common pitfalls. *Critical Care Medicine*, *49*(1), e63-e79.

Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. *Journal of Vocational Behavior*, *120*, 103445.

Tein, J. Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(4), 640-657.