CS5d:
Latent Class and Latent Transition Analysis:
LCA and LTA: A Guide to Practice

Date: Sat, Oct 8, 2022
Time: 8:45 AM - 10:15 AM CDT

Zhe Wang, Ph.D
Assistant Professor, Data Analytics
Denison University

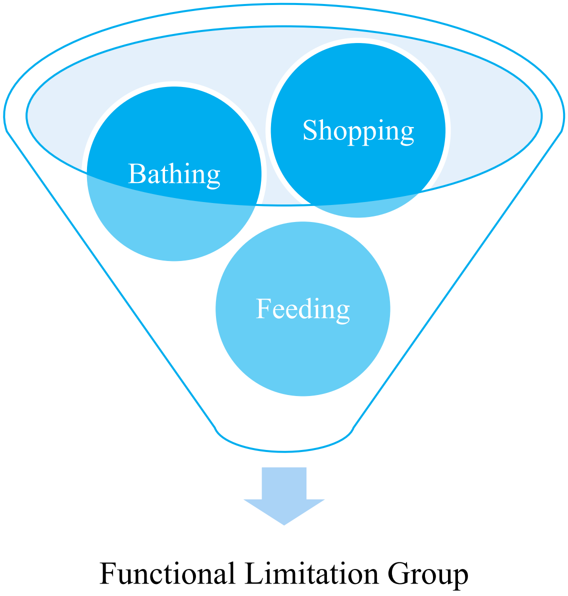
DENISON

Outline

- LCA and LTA Model Specification
- LCA and LTA Output Interpretation
- Model Selection
- Assumption and Limitation
- Summary

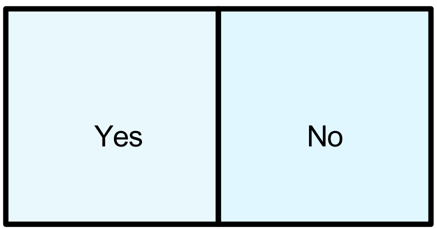
Observed Variable and Latent Variable

- Observed Variables:
  - Can be measured
  - Categorical variables
- Latent Variables:
  - Can not be measured directly
  - Categorical variables

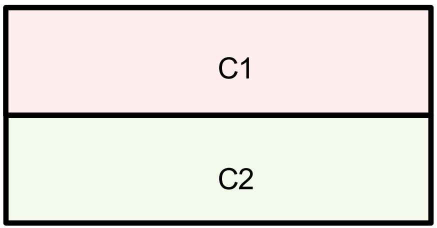


LCA Model Specification

- Observed Categorical Variables:
  - Y: Have difficulty dressing?
  - 1 = Yes, 2 = No
  - with marginal probability p_1, p_2



- Pre-selected 2 Latent Class
  - C: latent functional limitation group 1 or 2
  - with latent prevalence γ_1, γ_2



LCA Model Specification

- Observed Categorical Variables:
  - Have difficulty dressing?
  - 1 = Yes, 2 = No
  - with marginal probability p_1, p_2
- Pre-selected 2 Latent Class
  - C: latent functional limitation group 1 or 2
  - with latent prevalence γ_1, γ_2

Yes & C1	No & C1
Yes & C2	No & C2

$$P(Yes) = P(Yes\ \&\ C1) + P(Yes\ \&\ C2)$$

$$P(Y=1) = P(Y=1|C=1)P(C=1) + P(Y=1|C=2)P(C=2)$$

$$p_1 = \rho_{1|1}\gamma_1 + \rho_{1|2}\gamma_2$$

$$p_1 = \sum \gamma_c \rho_{1|c}$$

LCA General Parameters Definition

- Observed categorical variable: $Y_1, \dots, Y_M$;
- Each $Y_m (m = 1,2, \dots, M)$ has $r_m(\mathbf{1}, \mathbf{2}, \dots, \boldsymbol{R_m})$ different levels;
- Unobserved latent variable has $C (c = 1,2, \dots, C)$ different groups;

- Latent class prevalence $\boldsymbol{\gamma_c}$
- Item-response probability $\boldsymbol{\rho_{r_m|c}}$
- The marginal probability of observing $y_1, \dots, y_M$ is

$$P(Y_i = y_i) = \sum_{c=1}^{C} \boldsymbol{\gamma_c} \prod_{m=1}^{M} \prod_{r_m=1}^{R_m} \boldsymbol{\rho_{r_m|c}^{I(y_m = r_m)}}$$

Output Interpretation: Prevalence

Latent Group 1	Latent Group 2	….	….
0.80	0.20	…	….

80% of the patients were in Latent Group 1.

20% of the patients were in Latent Group 2.

Output Interpretation: Item Response Probabilities

Indicator		Latent Group 1	Latent Group 2	…
Do you have difficulty in dressing?	Yes	…	**0.85**	…
	No	…	0.15	…
Var 2	1	…	…	…
	2	…	…	…
	3	…	…	
…	…	…	…	…

Among patients who were in Latent Group 2,
85% of them had difficulty dressing.

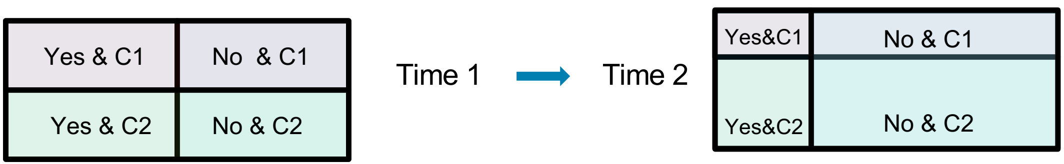
From LCA to LTA

- Latent Transition Analysis (LTA) is a longitudinal extension of LCA
- In LCA, class membership is static
- In LTA, class membership is dynamic
- Development can be represented as movement through discrete categories or stages
- Provides a way to estimate and test models of stage-sequential development

LTA Model Specification

- Observed Categorical Variables:
  - Y: Have difficulty dressing?
  - 1 = Yes, 2 = No
  - with marginal probability at **2** time points (p_{1_1}, p_{2_1}), (p_{1_2}, p_{2_2})

- Pre-selected 2 Latent Class
  - C: latent functional limitation group 1 or 2
  - with latent prevalence at the initial time points $(\gamma_{1_1}, \gamma_{2_1})$



LTA General Parameters Definition

- Observed categorical variable: $Y_1, \ldots, Y_M$;
- Each $Y_m (m = 1,2, \ldots, M)$ has $r_m (\mathbf{1, 2, \ldots, R_m})$ different levels;
- Unobserved latent variable has $C (c = 1,2, \ldots, C)$ different groups;
- T$(t = 1,2, \ldots, T)$ different time points.

- $\boldsymbol{\gamma_{c_1}}$ = Latent class prevalence at Time 1
- $\boldsymbol{\rho_{r_m|c_t}}$ = Item-response probabilities at Time t
- $\boldsymbol{\tau_{c_t|c_{t-1}}}$ = Transition probabilities of latent class c_t at time t, conditional on membership in latent class c_{t-1} at time $t-1$, e.g. $\boldsymbol{\tau_{1_2|2_1}}$

$$P(Y_i = y_i) = \sum_{c_1=1}^{C} \ldots \sum_{c_T=1}^{C} \boldsymbol{\gamma_{c_1}} \boldsymbol{\tau_{c_2|c_1}} \ldots \boldsymbol{\tau_{c_t|c_{t-1}}} \prod_{m=1}^{M} \prod_{r_m=1}^{R_m} \prod_{t=1}^{T} \boldsymbol{\rho_{r_m|c_t}^{I(y_m=r_m)}}$$

Output Interpretation: Prevalence

	Latent Group 1	Latent Group 2	….
Time 1	**0.80**	**0.20**	…
Time 2	0.70	0.30	…
…	…	…	…

At Time 1, **80%** of the patients were in Latent Group 1, and **20%** of the patients were in Latent group 2.

The numbers at time 2 are 70% and 30%.

Output Interpretation: Item Response Probabilities

Indicator (Time 1)		Latent Group 1	Latent Group 2	…
Do you have difficulty in dressing?	Yes	…	**0.85**	…
	No	…	**0.15**	…
Var 2	1	…		…
	2	…		…
…	…	…	…	…

Among patients who were in Latent Group 2,
85% of them had difficulty dressing,
and **15%** of didn't have this issue.

Output Interpretation: Transition Probabilities

Time 1 \ Time 2	Latent Group 1	Latent Group 2	…
Latent Group 1	**0.80**	**0.20**	…
Latent Group 2	…	**…**	…
…	…	…	…

For patients who were in Latent Group 1 at Time 1,

80% of them were in Latent Group 1 at Time 2;

20% of them transferred to Latent Group 2 at Time 2.

Model Selection Criteria and Model Diagnostics

Fit Statistics	Thresholds	Recommendation
AIC	The lower the better	For analyses where n < 300, we advise using and reporting the AIC and BIC.
CAIC		
BIC		For all analyses, we recommend using and reporting BIC and SABIC.
SABIC		
VLMR-LRT	<0.05	Use to test if a model with k classes is better than model with k–1 class.
BLRT		
Log-Likelihood	The higher the better	Log-likelihood will be maximized using EM algorithm. It cannot be used to compare models.
Entropy	>0.8	We advise reporting entropy for model diagnostics but not relying on the value to determine a final class solution.

* Aflaki, Kayvan, Simone Vigod, and Joel G. Ray. "Part II: A Step-by-Step Guide to Latent Class Analysis." *Journal of Clinical Epidemiology* (2022).

Model Assumptions

Local independence assumption:

- latent class variable accounts for all relations between the observed variables

Conditional independence assumption:

- observed variables are independent conditional on the latent variable

Missingness assumption:

- the model assume data are missing at random

Markov assumption(LTA):

- the change over time only depends on one measurement time before

Limitation 1:

- **Proper class assignment is not guaranteed**

C	LL	AIC	CAIC	BIC	SABIC	BLRT p	VLMR-LRT p
1	-11238.18	9428.707	9572.890	9548.890	9472.660	-	-
2	-9909.352	6821.055	7115.427	7066.427	6910.791	<0.001	<0.001
3	-9663.649	6379.651	6824.213	6750.213	6515.171	<0.001	**0.0342**
4	-9480.225	6062.801	**6657.554**	6558.554	6244.106	<0.001	0.3282
5	-9384.330	5921.011	6665.953	**6541.953**	6148.099	<0.001	0.1519
6	-9303.139	5808.629	6703.762	6554.762	6081.501	<0.001	0.7559
7	-9234.914	**5722.180**	6767.503	6593.503	**6040.836**		

Simulation studies suggest the first time the p-value of LMR test is non-significant might be good indication to stop increasing the number of class. *

Nylund, K.L., Asparouhov, T.I.H.O.M.I.R. and Muthén, B.O., 2007. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, *14*, pp.535-69.

Limitation 2:

- **Naming fallacy for latent classes**

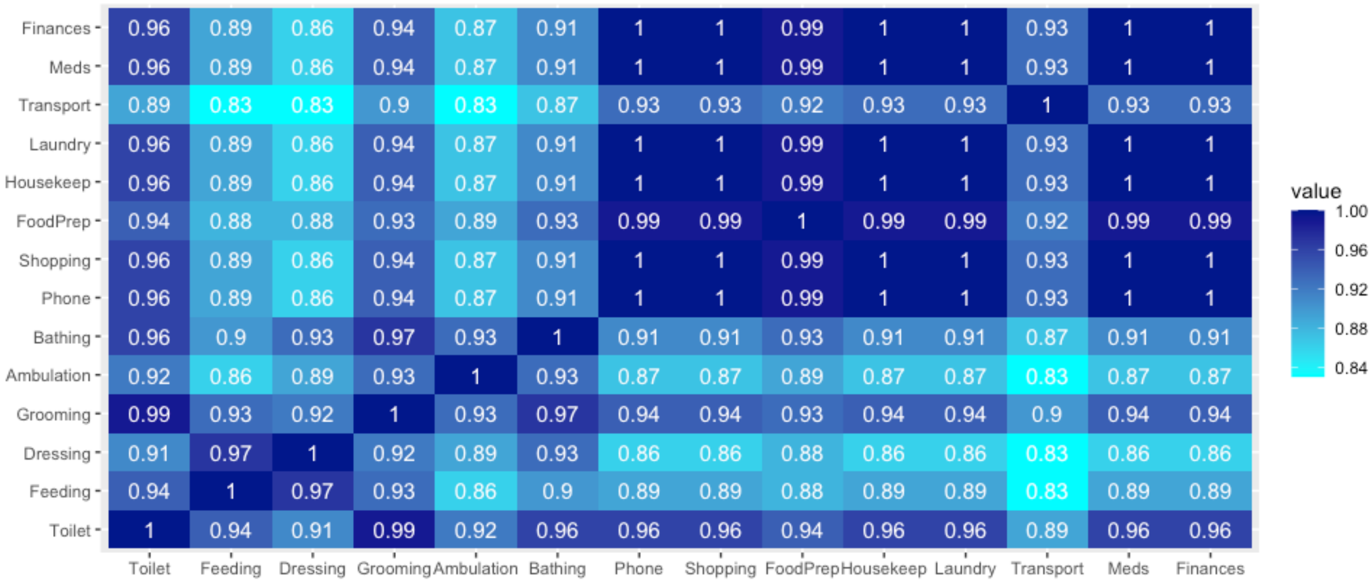
Have deficits in..	Class 1	Class 2
Eating	**0.710**	0.178
Walking	**0.895**	0.040
Managing Money	0.113	**0.911**
Taking Med.	0.131	**0.860**

Have deficits in..	Class 1	Class 2
Eating	**0.710**	**0.835**
Walking	0.411	0.040
Managing Money	**0.956**	**0.911**
Taking Med.	0.231	**0.860**

Table: Item response probabilities for patients having deficits in the listed activities.

Limitation 3:

- Can not deal with multicollinearity and complex patterns



LCA Summary

- Step 0: Study descriptive statistics and test on assumptions
- Step 1: Starting with a one-class model, and add one class at a time,
- Step 2: Compare model fit statistics to identify the best model
- Step 3: Explore specification of the LCA without covariates
- Step 4: Extend the model complexity, e.g. add covariates or distal outcomes
- Step 5: Report the results

LTA Summary

- Step 0: Study descriptive statistics and test on assumptions
- Step 1: Conduct LCA at each time point. Compare model fit statistics to identify the best model at each time point
- Step 2: Test measurement invariance if the same number of classes emerge in Step 1
- Step 3: Explore specification of the latent transition model without covariates
- Step 4: Extend the model complexity, e.g. add covariates or distal outcomes
- Step 5: Report the results

The End

Please email: wangz@denison.edu

- To discuss further about these studies or your ideas