ED 201F – Applied Mixtures

Lab Activity: Investigating Classification in LCA

Learning goal:

The goal of this lab activity is to provide an opportunity for you to "get your hands into" the posterior probabilities and get practice calculating the Average Posterior Probabilities (AvePP) for each class. Specifically, we are going to recreate this table from the Mplus output. One way to think of the AvePP values are as the class-specific classification errors in the model.

Average Latent Class Probabilities for Most Likely Latent Class Membership (Row)by Latent Class (Column)

	1	2	3	4
1	0.868	0.024	0.000	0.109
2	0.008	0.861	0.131	0.000
3	0.000	0.126	0.846	0.028
4	0.038	0.004	0.016	0.941

In this lab we use the PYDI dataset as an example. The LCA model has already been estimated with the response patterns for the 4-class solution included in the Excel spreadsheet.

How do we incorporate classification error in LCA?

Since this is "pattern" data, each line represents a specific response pattern. Thus we need to weight each pattern by the frequency to get the total error across all observations:

Weighted probability
(number of people in that pattern) * CPROB

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1		Freq	PYDI1	PYDI2	PYDI3	PYDI4	PYDI5	PYDI6	PYDI7	CPROB1	CPROB2	CPROB3	CPROB4	C					
2		889	1	1	1	1	1	1	1	0.000	0.935	0.064	0.000	2		B2*J2			
3		271	0	1	1	1	1	1	1	0.000	0.644	0.355	0.001	2					
4		23	1	1	1	1	0	1	1	0.389	0.601	0.000	0.010	2					
5		7	1	1	1	1	1	0	1	0.000	0.710	0.274	0.016	2					
6	TOTAL>>>																		
7																			

total frequency count, which is the same as number of people who were modally assigned to this class.

total weighted probability (also thought of as error)

Steps for estimating the Average Posterior Probabilities (AvePP):

1. We will begin by computing the AvePP values for observations in modal class 1 (Most Likely Class Membership):
    a. Go to the second spreadsheet in the file named **C = 1**

2. Compute the weighted error for each response pattern by class:
    a. In the first input column and row (shaded yellow) in the cell "**P2**" multiply the frequency for that specific response pattern by the posterior probability for class 1 (**CPROB1**)
    b. The weighted error will be repeated for each response pattern (down the column)
    c. Repeat the weighted error calculation for each of the remaining classes **(CPROB2-4)**

3. Compute the total frequency count by summing the frequency column: Insert the total frequency count in the empty row beneath the table

4. Take the sum of each of the weighted error columns computed in step 2 to compute the **total weighted error**
    a. Divide the **total weighted error** by the **total frequencies count**

5. These four values are the first row of the AvePP table from the Mplus output.

6. Repeat steps 1 – 4 to compute the values in the remaining 3 rows of the AvePP table (i.e., for observations where most likely class membership are classes 2 – 4. Check to make sure the values correspond with the Mplus output.

Follow up questions:

1. Look at those who were assigned to class 1:
    - Looking at the tab C=1, what column do you think would have the largest weighted posterior probability (cprob)? Why?
    - What is the response pattern with the largest CPROB1?
    - What is the response pattern with the smallest CPROB1?
    - How about for class 2?
2. How should we interpret the diagonal values of the AVEpp table?
        - How should we interpret the off-diagonals?