Latent Transition Analysis

Dr Oliver Perra
Institute of Child Care Research
Queen’s University Belfast

e-mail: o.perra@qub.ac.uk

Overview of latent class and latent transition models

Latent Class Analysis

• Part of “mixture” models
  — Assumption: unobserved heterogeneity in the population
• Given a set of categorical indicators, individuals can be divided into subgroups (latent classes) based on an unobserved construct (e.g. Disordered v. Non-Disordered)
• Latent classes are mutually exclusive and exhaustive
• Individuals in each class are supposed to behave in the same manner (similar parameter values)
  — Intra-group homogeneity
  — Inter-group heterogeneity
• Latent classes describe the associations among the observed categorical variables

Latent Class Analysis

C

U1 U2 U3 U4
Latent Class Analysis

• Parameters of the model are:
  – Probability of being in each class (membership)
  – Probability of fulfilling each criterion (e.g., endorsing an item) given class membership
    • E.g., Probability of providing correct response to a test given membership in the “Mastery” latent class.
  – Furthermore, the model provides probability of being in each class for each individual (posterior probability)

Categorical indicators: a b c d
Latent class: x
\[ P_{abcdx} = p_x \times p_{a|x} \times p_{b|x} \times p_{c|x} \]
Sum \( p_x = \sum p_{a|x} = \sum p_{b|x} = \sum p_{c|x} = 1 \)

Assumption of conditional independence

• Manifest variables are independent given latent class
  – Put it another way: the observed relationship between manifest variables (answers to questions, success in test items, etc.) is attributable to a common factor

If X is the latent variable with different classes, A and B are categorical outcomes:
\[ P_{abc} = P(a=1|x=1) \times P(b=1|x=1) \times P(x=1) \]
with \( a=1 \rightarrow \text{pass in a} \); \( b=1 \rightarrow \text{pass in b} \); \( x=1 \rightarrow \text{mastery} \)

The probability any mastery respondent passes both tests (P of 111) is equal to the product of their estimated conditional probability of passing test a and estimated probability of passing test b

• Some variables are unlikely to be conditionally independent (e.g., related symptoms).

LC: Model Estimation

• Iterative maximum-likelihood estimation approaches
• Begin with a set of “start values” and proceed with re-estimation iterations until a criterion is met (usually convergence: each iteration in parameter estimation approaches some predesigned small change)
• Expectation-Maximization algorithm: robust with respect to initial start values
• Problems of local optima: convergence to local solutions
Latent Transition Analysis (LTA)

- Longitudinal extension of latent class models

LTA v. Growth models

- In growth models the focus is on average rate of change over time and the growth process is assumed to be continually occurring at the same rate
- In LTA, change can be discontinuous: movement through discrete categories or stages
  - “Qualitative growth”: changes not restricted to quantitative growth
  - Different people may take different paths

Examples of LTA applications - I

- Stages of change for smoking cessation (Martin, Velicer & Fava, 1997)
  - 4 stages:
    - Pre-contemplation
    - Contemplation
    - Action
    - Maintenance
  - Movement was not always linear (forthsliders and backsliders; 2-stage progressions)
  - Probability of forthsliding> backsliding
  - Greater probability to move to adjacent stages than 2-stage progression

Examples of LTA applications - II

- LTA used to evaluate the stability of Typically Developing v. Reading Disability classification across grades 1 to 4 (Compton et al., 2008)
  - Results suggested a fair amount of stability
  - Results also suggested the importance of including a word reading fluency item in the model estimation, particularly after grade 1: inclusion of this indicator reduced “false negatives”
Examples of LTA applications - III

Substance use

A model of substance use onset including both alcohol and tobacco use as possible starting points fit better than a model that included alcohol use as the only starting point. Participants who had tried tobacco but not alcohol in 7th grade seemed to be on an accelerated onset trajectory.

Latent Transition Analysis (LTA)

- Allows specification of number of stages in a model
- Transitions consistent with model, e.g. Cannabis lifetime use → no use (?)
- Estimate prevalence of class membership at first time of measurement
- Incidence of class transitions
- Probability of particular item responses conditional on stage membership

Example of LTA (Nylund, 2007)

- A longitudinal study of over 1,500 middle-school students in US
- Students completed 6-item Peer Victimization Scale in grade 6, 7 and 8 (e.g. being picked on, laughed at, hit and pushed around, etc.)
- Responses to items dichotomised

Note that is not necessary that items have the same number of response categories

<table>
<thead>
<tr>
<th>Item</th>
<th>Grade 6</th>
<th>Grade 7</th>
<th>Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Called bad names</td>
<td>37%</td>
<td>25%</td>
<td>20%</td>
</tr>
<tr>
<td>Talked about</td>
<td>33%</td>
<td>26%</td>
<td>23%</td>
</tr>
<tr>
<td>Picked on</td>
<td>28%</td>
<td>19%</td>
<td>14%</td>
</tr>
<tr>
<td>Hit and pushed</td>
<td>21%</td>
<td>15%</td>
<td>12%</td>
</tr>
<tr>
<td>Things taken/messed up</td>
<td>29%</td>
<td>19%</td>
<td>15%</td>
</tr>
<tr>
<td>Laughed at</td>
<td>30%</td>
<td>20%</td>
<td>18%</td>
</tr>
</tbody>
</table>

Proportion endorsed for 6 binary items by grade
3 classes in Grade 6

<table>
<thead>
<tr>
<th></th>
<th>Victimised (19%)</th>
<th>Sometimes-victimised (29%)</th>
<th>Non-Victimised (52%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Called bad names</td>
<td>.85</td>
<td>.58</td>
<td>.08</td>
</tr>
<tr>
<td>Talked about</td>
<td>.74</td>
<td>.51</td>
<td>.07</td>
</tr>
<tr>
<td>Picked on</td>
<td>.81</td>
<td>.39</td>
<td>.03</td>
</tr>
<tr>
<td>Hit and pushed</td>
<td>.76</td>
<td>.17</td>
<td>.03</td>
</tr>
<tr>
<td>Things taken/messed up</td>
<td>.79</td>
<td>.31</td>
<td>.09</td>
</tr>
<tr>
<td>Laughed at</td>
<td>.86</td>
<td>.36</td>
<td>.06</td>
</tr>
</tbody>
</table>

Conditional item response probability (probability of endorsement) by latent class

3 classes in Grade 7

<table>
<thead>
<tr>
<th></th>
<th>Victimised (13%)</th>
<th>Sometimes-victimised (20%)</th>
<th>Non-Victimised (67%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Called bad names</td>
<td>.76</td>
<td>.59</td>
<td>.05</td>
</tr>
<tr>
<td>Talked about</td>
<td>.69</td>
<td>.53</td>
<td>.09</td>
</tr>
<tr>
<td>Picked on</td>
<td>.82</td>
<td>.26</td>
<td>.03</td>
</tr>
<tr>
<td>Hit and pushed</td>
<td>.68</td>
<td>.12</td>
<td>.05</td>
</tr>
<tr>
<td>Things taken/messed up</td>
<td>.68</td>
<td>.29</td>
<td>.05</td>
</tr>
<tr>
<td>Laughed at</td>
<td>.75</td>
<td>.38</td>
<td>.03</td>
</tr>
</tbody>
</table>

Conditional item response probability (probability of endorsement) by latent class

Transition probabilities grade 6 to 7
(LTA model)

<table>
<thead>
<tr>
<th></th>
<th>6th Grade</th>
<th>7th Grade</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Victimised</td>
<td>Sometimes-victim.</td>
<td>Non-victim.</td>
</tr>
<tr>
<td>Victimised</td>
<td>.42</td>
<td>.41</td>
<td>.17</td>
</tr>
<tr>
<td>Sometimes-victim.</td>
<td>.05</td>
<td>.48</td>
<td>.47</td>
</tr>
<tr>
<td>Non-victim.</td>
<td>.01</td>
<td>.10</td>
<td>.89</td>
</tr>
</tbody>
</table>

N of classes at each occasion

• Many LTA models will consider the same number of classes at each occasion
• However, there may be cases where the number of latent classes may be different across time:
  – e.g.: 2 classes of exposure to violence may be sufficient in early adolescence, but 5 classes may be necessary to describe heterogeneity of violence exposure in late adolescence (more diversity in phenomenon)
• The interpretation of each class is a function of its item response probabilities (see next)
LTA parameters

- Item response probabilities (some refer to these as rho, ρ)
  - Probability of endorsing a category of response at time t (e.g.: 1, 2, ..., t) given latent status membership at time t
  - These allow to interpret latent statuses (e.g. Higher probability of endorsing victimisation items → victimised class)
  - One for each time-status-item combination
    - Constraints can be assumed and tested: E.g. identical across measurement occasions (measurement invariance)?

LTA Parameters (ctd.)

- Latent class prevalence at time t: probability of being in latent class a at time t
- Some (e.g. Collins) refer to these parameters as delta δ (with a subscript for class and time, e.g. δ \text{v6}
  - E.g. In Nylund’s study, prevalence of “victimised” class in grade 6 was 19%, thus δ \text{v6} = .19

LTA Parameters (ctd.)

- Transition probabilities: Probability of class b membership at time 2 given membership to class a at time 1
  - E.g. Probability of being in “victimised” class in grade 7 given membership to “non-victimised” in grade 6 (= .01)
  - Usually referred to as tau τ and underscript indicating class membership at time t given membership at time 1, e.g.: 
    - τ_{b|a}
    - τ_{1|3}

LTA Parameters (ctd.)

- τ parameters arranged in a transition probability matrix like this:

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Time 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Class 1</td>
<td>Class 2</td>
</tr>
<tr>
<td>Class 1</td>
<td>τ_{11}</td>
<td>τ_{12}</td>
</tr>
<tr>
<td>Class 2</td>
<td>τ_{21}</td>
<td>τ_{22}</td>
</tr>
<tr>
<td>Class 3</td>
<td>τ_{31}</td>
<td>τ_{32}</td>
</tr>
</tbody>
</table>
LTA Parameters (ctd.)

• Restrictions and constraints can also be imposed on transition parameters:

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Time 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Victimised</td>
<td>Sometimes victim.</td>
</tr>
<tr>
<td>(\tau_{11})</td>
<td>(\tau_{21})</td>
</tr>
<tr>
<td>Sometimes-victim.</td>
<td>(\tau_{12})</td>
</tr>
<tr>
<td>Non-victimised</td>
<td>(\tau_{13})</td>
</tr>
</tbody>
</table>

• E.g. \(\tau_{13} = 0\) → fixing probability of transitioning from non-victimised to victimised to 0

• Absorbing class: one that has a zero probability of exiting: \(\tau_{11} = 1\) → 100% probability of being victimised at time 2 if victimised at time 1

LTA Parameters (ctd.)

• Other restrictions and constraints can be imposed on transition parameters:

  – Transition probabilities to be the same across time points:
    
    E.g. : The probability of transitioning from victimised to non-victimised between grades 6 and 7 the same as between grades 7 and 8
    
    \(\tau_{n7|v6} = \tau_{n8|v7}\)
    
    Change process assumed stationary: individuals are transitioning between classes with the same probabilities across time points

Summary so far

• Latent Class Analysis: fundamentally a measurement model

• Latent Transition Analysis: measurement and structural model. Describes qualitative change across measurements points (2 or more)

• LTA parameters:
  
  – Conditional item response probabilities \(p\) (measurement model)
  
  – Prevalence of latent statuses at each time point \(\delta\)
  
  – Transition probabilities between two time points \(\tau\)

LTA Steps

• Step 1: Investigate measurement model alternatives for each time point (separately for each time point)

• Step 2: Test for measurement invariance across time

• Step 3: Explore specification of the latent transition model without covariates
  
  – Investigate transition probability specifications

• Step 4: Include covariates in LTA model

• Step 5: Include distal outcomes
### Step 1: Investigate measurement model alternatives

- Decision does not involve only statistical indicators of fit to data, but also interpretability of results and aims of the study.
- "The choice of factor analysis or LCA is a matter of which model is most useful in practice. It cannot be determined statistically, because data that have been generated by an m-dimensional factor analysis model can be fit perfectly by a latent class model with m+1 classes" (Muthén & Muthén, 2000).
- If the aim is *diagnosis* or *categorisation*, then use LCA (avoids the use of arbitrary cut-points or ad-hoc rules).

### Step 1: investigate measurement model

<table>
<thead>
<tr>
<th>Sub-step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>If LCA → determine number of classes at each time point</td>
</tr>
<tr>
<td>1.2</td>
<td>Test restrictions on item response parameters</td>
</tr>
<tr>
<td>1.3</td>
<td>Validate results including covariates</td>
</tr>
</tbody>
</table>

### Determining n of classes

- The standard procedure is to test a series of LC models: from 2-class to n-class.
- No accepted single indicator to decide on the appropriate number of classes:
  - Although log-likelihood value is provided in estimation, this cannot be used to compare models with different n classes (e.g. 2- vs. 3-class) via Likelihood Ratio Test (LRT).
Determining n of classes (ctd.)

- Consider $\chi^2$ and likelihood ratio chi-square test $G^2$
- Use information criteria (the lower the value the better the fit)
  - AIC penalises by number of parameters $\Rightarrow$ preference for “simpler” models
  - BIC penalises by number of parameters and sample size
  - Mplus provides the sample-size adjusted BIC

LC statistics and information criteria

- $\chi^2 = \text{Sum } [ (\text{observed f. } - \text{expected f. })^2 / \text{exp. f. } ]$
- $G^2 = 2 \text{ sum } [\text{obs. f. } \times \ln(\text{obs. f. } / \text{exp. f. }) ]$
- AIC = $G^2 - 2 \text{ df}$
- BIC = $G^2 - \text{df } \times [\ln(N)]$
- Sample-size adjusted BIC: $N^* = (N + 2) / 24$

Practical

- Introduction to Mplus language
- Estimation of LC model using Mplus
- Imposing constraints on measurement parameters using Mplus

Intro to LCA in Mplus

- Mplus uses:
  - input files to instruct how to read separate data file, to specify type of analysis and model and to request information in output file and other functions (additional files, plots, etc.).
  - Results are reported in the output file
  - It can also provide (under request in input) files that can be used to create graphs
  - It can provide (under request) files with model parameters
**Intro to Mplus (ctd.)**

**ANALYSIS:**

- **TYPE:** MIXTURE;
- **STARTS:** 100 10;
- **STITERATIONS:** 20;

The other essential bit to conduct LCA:

- **TYPE:** MIXTURE in the ANALYSIS command invokes a mixture model algorithm (necessary for “mixture” models such as LCA, LTA, LGCA, GMM, etc).
- The default estimator for this type of analysis is Maximum Likelihood with robust standard errors (MLR in Mplus). [This can be changed with command ESTIMATOR = …].
- By default, ML optimization in two stages: initial one with 10 random sets of starting values; 2 optimisations with highest likelihoods used as starting values in the final stage. This is what would happen if you do not provide the STARTS command in ANALYSIS.
- In the example above, 100 random sets are used, with 10 values with highest likelihood used in the final stage. Increase n starts is often necessary for the model to converge.
- The max number of iterations allowed in initial stage is 10 by default, but can be increased (in the example STITERATIONS = 20) for more thorough investigation of multiple solutions.

**MODEL:**

- **%OVERALL%**
  - This is the part of the model common for all classes [x#1].
  - %#1% [a$1-d$1](1-5);
  - %#2% [a$1-d$1](6-10);

The other important part is the MODEL command. It is not necessary to specify a model if you are conducting a simple LCA, with no covariates and no restrictions on parameters (omit the MODEL command completely in this case).
- **%overall%** describes the part of the model that is common to ALL latent classes (e.g. latent class affiliation is regressed on covariate x).
- **%#1%** is used to specify the part of the model that differs for class 1,
- **%#2%** specifies the part of the model specific to class 2
... And so on (if more than 2 classes)
Intro to Mplus (ctd).

• Mplus thinks of categorical variables (binary or with more categories) as continuous latent variables that are “cut” into different categories.
• The points in which to “cut” the underlying latent variable are called thresholds.
• If we take a binary variable:
  - Categories of response are “No” (category 1) and “Yes” (category 2)
  - Indicators have one threshold each \([a_1 b_1 c_1 d_1]\);
  - The threshold represents the point in which the underlying distribution is cut to create the two response categories
  - We want to fit a two-class model: \(x\) (latent class) \(\rightarrow x\#1\) (latent class 1) \(\rightarrow x\#2\) (latent class 2)
    - In the same manner as for observed categorical variables, we need to estimate a threshold for \(x\) \(\rightarrow [x\#1]\) that cuts the distribution into two categories


Intro to Mplus (ctd.)

• We are considering a model with 4 binary indicators:
  - a b c d
• Categories of response are “No” (category 1) and “Yes” (category 2)
• Indicators have one threshold each \([a_1 b_1 c_1 d_1]\);
  - the threshold represents the point in which the underlying distribution is cut to create the two response categories
• We want to fit a two-class model: \(x\) (latent class) \(\rightarrow x\#1\) (latent class 1) \(\rightarrow x\#2\) (latent class 2)
  - In the same manner as for observed categorical variables, we need to estimate a threshold for \(x\) \(\rightarrow [x\#1]\) that cuts the distribution into two categories

Intro to Mplus (ctd.)

• Number of thresholds = n of categories -1 (a binary variable needs only one cut to create two categories).
• Thresholds are indicated by the name of the variable followed by \$ and the progressive number: all within square brackets.
• A variable a with 3 categories (e.g. not yet, sometimes, often) would have 2 thresholds:
  - \([a_1 ; a_2]\)
• The asterisk * is used to free a parameter. If followed by a number, it assigns a starting value to the thresholds;
• @ is used to fix the value of a thresholds to some pre-defined value (e.g. -15)
Thresholds are in a logit scale:
The LCA model with p observed binary items u, has a categorical latent variable C with K classes \(C = k; k = 1, 2, ..., K\). The marginal item probability for item \(u_j = 1\) \((j = 1, 2, ..., p)\) is given by:
\[
P(u_j = 1) = \sum P(C = k) * P(u_j = 1 | C = k)
\]
where the conditional item probability in a given class is defined by:
\[
P(u_j = 1 | C = k) = \frac{1}{1 + \exp(-v_{jk})}
\]
where the \(v_{jk}\) is the logit for each of the \(u_j\)s for each of the latent classes, \(k\).

For example, if we want to constrain \(P(a=1|c=1) = .05\), we fix logit threshold \(v(k)\) to \(-2.95\);
A threshold = 0 will make \(P(a=1|c=1) = .50\) ...and so on

Intro to Mplus (ctd.)

The parentheses after the indicators’ thresholds assign a name (if a letter is used) or post a constraint (if a number used) to each of these parameters.
If we wanted the thresholds of \(a, b, c\) and \(d\) to be the same for \(x1\) and \(x2\), we would have written:
\[
\begin{align*}
&\text{MODEL:} \\
&\text{%OVERALL%} \\
&\text{this is the part of the model common for all} \\
&\text{classes} \\
&[x#1]; \\
&[x#1] (1); \\
&[x#2] (2); \\
&\text{MODEL CONSTRAINT:} \\
&p2 = p1;
\end{align*}
\]

Constraints on measurement model: Parallel indicators
In this example, the thresholds for the latent class estimators \((a:d)\) are equal to each other within each class, but not equal across classes \(\Rightarrow\) given membership in class 1, the probability of endorsing indicator \(a\) is the same as the probability of endorsing indicator \(b\), and so on.
Referred as parallel indicators: have identical error rates with respect to each of the latent classes (if we consider one type of response within class as an error).
Constraints on measurement model (ctd.)

I added a statement to fix the thresholds of d in class 1 to (the logit value of) -15 (@ fixes the value of parameters). This means that individuals in class 1 have probability=1 of endorsing the item. By placing the threshold at the lower limit of the underlying distribution, all scores will be above the “cut”, hence in category 2.

Intro to Mplus (ctd.)

Command OUTPUT allows you to choose options regarding information in the output. TECH1 for example will report arrays containing parameter specifications and starting values for all free parameters in the model (useful to check what the model is actually doing). TECH10 reports univariate, bivariate and response pattern model fit information for the categorical dependent variables in the model.

The PLOT command creates graph files that can be useful for inspecting results. TYPE = PLOT3 provides plots with histograms, scatterplots, sample proportions and estimated probabilities (e.g. item response conditional probabilities).

What the output looks like:

A successfully converged model will have the best log likelihood values repeated at least twice. If the best (highest closest to 0) value is not replicated in at least two final stage solutions, it is possible a local solution has been reached (the solution is not trustworthy).
Success

Loglikelihood values at local maxima, seeds, and initial stage start numbers:

-10148.718  987174         1689
-10148.718  777300         2522
-10148.718  406118         3827
-10148.718  51296          3485
-10148.718  997836         1208
-10148.718  119680         4434
-10148.718  338892         1432
-10148.718  765744         4617
-10148.718  636396         168
-10148.718  189568         3651
-10148.718  469158         1145
-10148.718  90078          4008
-10148.718  373592         4396
-10148.718  73484          4058
-10148.718  154192         3972
-10148.718  203018         3813
-10148.718  785278         1603
-10148.718  235356         2878
-10148.718  681680         3557
-10148.718  92764          2064

A solution (-10155.482) is replicated 2 times, but is not the best solution. The best log-likelihood solution must be replicated for a trust-worthy solution.

What if log likelihood not replicated?

If already increased STARTS (e.g. = 100 10) and STITERATIONS (e.g. =20) then:

- Increase the initial stage random sets of starting values further to 500 (e.g. STARTS = 500 10) or more.
- Take the seed value of the best loglikelihood values, then use the OPTSEED option in the ANALYSIS command indicating these seeds:
  E.g. ANALYSIS: TYPE=mixture; OPTSEED=370560;
  If estimates are replicated using different seeds of best log-likelihoods, we can trust we did not find local solutions.

Note: problems in converging indicate the model is not well defined for the data: e.g. too many classes extracted.

Not there yet

Loglikelihood values at local maxima, seeds, and initial stage start numbers:

-10153.627  23688          4596
-10153.678  150818         1050
-10154.388  584226         4481
-10155.122  735928         916
-10155.373  309852         2802
-10155.437  925994         1386
-10155.482  370560         3292
-10155.482  662718         460
-10155.630  320864         2078
-10155.833  873488         2965
-10156.017  212934         568
-10156.231  98352          3636
-10156.339  12814          4104
-10156.497  557806         4321
-10156.644  134830         780
-10156.741  80226          3041
-10156.793  276392         2927
-10156.819  304762         4712
-10156.950  468300         4176
-10157.011  83306          2432

What does the output look like?

TESTS OF MODEL FIT

Loglikelihood

H0 Value -2663.146
H0 Scaling Correction Factor 1.020
for MLR

Information Criteria

Number of Free Parameters 9
Akaike (AIC) 5344.293
Bayesian (BIC) 5388.462
Sample-Size Adjusted BIC 5359.878
[n^2 / (n - 2) / 24]

What does the output look like?

Chi-Square Test of Model Fit for the Binary and Ordered Categorical (Ordinal) Outcomes

Pearson Chi-Square
Value  3.309
Degrees of Freedom 6
P-Value 0.7428

Likelihood Ratio Chi-Square
Value  3.496
Degrees of Freedom 6
P-Value 0.7445
What does the output look like?

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THE ESTIMATED MODEL
Latent Classes
1  524.25270  0.52425
2  475.74730  0.47575

ENTROPY serves as a measure of the precision of individual classification. It ranges from 0 (everybody has an equal posterior probability of membership in all classes) to 1 (each individual has posterior probability 1 of membership in a single class and probability 0 of membership in the remaining classes). High entropy indicates clear class separation.

Determined the number of classes
• Compare statistics and information criteria (BIC, AIC, sample-size adjusted BIC) → the lower, the better fit
• Likelihood Ratio Test (LRT) not applicable: but Mplus provides a Bootstrap LRT (OUTPUT: TECH14).
  – if CLASSES=x(3), the test provides p value of 3-class vs. 2-class fit. A significant value (p < .05) would indicate a significant improvement in fit with the inclusion of a third class.
• Mplus provides another similar test (Vu-Luong-Mendell-Rubin → TECH11)
• Consider Entropy (if the aim is finding homogenous clusters)
• Inspect bivariate and response patterns standardised residuals (TECH10): the model with more significant residuals (>|1.96|) has lower fit
• Interpretability of results

What does the output look like?

MODEL RESULTS

Two-Tailed
Estimate    S.E.  Est./S.E.    P-Value

Latent Class 1
Thresholds
A$1               -0.948      0.187     -5.056      0.000
B$1               -0.764      0.169     -4.529      0.000
C$1               -1.103      0.185     -5.957      0.000
D$1               -0.895      0.184     -4.860      0.000

Latent Class 2
Thresholds
A$1                1.272      0.250      5.093      0.000
B$1                0.953      0.174      5.492      0.000
C$1                0.901      0.205      4.397      0.000
D$1                1.023      0.191      5.372      0.000

Categorical Latent Variables

W991               0.097      0.241      0.402      0.688

The conditional item response probabilities help attach meaning to each class (similarly to factor loadings in factor analysis).
In this case, class 1 includes individuals that have higher probability of endorsing the items (e.g. if items are symptoms, this class would be the “disorder” class).
Class 2 includes individuals with lower probabilities of endorsing items.
In this case, the profiles do not cross, but is possible to have classes where, for example, individuals in one class have higher probability of endorsing items a and b and individuals in another class endorse items c and d.
Step 1: Validate results of LCA

- Test associations between latent classes (cross-sectional) and covariates: do they make sense?
  - E.g. Does the “victimised” latent class at each age point relate to known risk factors of this process (e.g. School safety)?
- It is also possible to investigate differential item functioning:
  - Two individuals in the same latent class have different item endorsement probabilities.

*Note that the introduction of covariates (and distal outcomes) may change the model parameters, including class profiles and their respective size (more on this later)*
Validate results of LCA (ctd.)

- In Mplus, the regression of one variable on another one is expressed by "ON" in the MODEL command.
- To regress latent variable class on covariate gender (coded male) →
  class ON male;
  Regression of class on male: the dependent (class) is regressed on the covariate (male)

Summary Step 1

- Assuming classification is the aim, determine the number of classes at each time point (consider information criteria, model residuals, interpretability of results, etc.)
- It is possible to test constraints on measurement model
- Test associations with covariates and DIF

Step 2: Investigate measurement invariance

- Assume we have settled for a measurement model at each time point (LCA), identified the number of classes and decided on other parameters constraints (e.g. parallel indicators)
- If the same number and type of classes across time, we can explore measurement invariance:
  - Equality of parameters of the measurement models, the conditional item response probabilities
- Measurement invariance assures that latent statuses can be interpreted in the same way across time
Types of measurement invariance

- Full invariance: conditional item probabilities are invariant across measurement occasions
  - Same number and type of classes occur at each time point
- Full measurement non-invariance: no constraints on measurement parameters across time
  - Even if the same n of classes, their profile and their meaning may be different
- Partial measurement invariance: equality of constraints for some measurement parameters across time

Assumptions tested before imposing relationships between latent variables

Measurement invariance

- Reduces the number of parameters estimated (as well as computation)
- Makes interpretation of parameters straightforward
- However, it may not be plausible, depending on the nature of latent classes, indicators, period spanned by measurement points

Mplus: cross-sectional LCA

- We assume 4 indicators (a b c d) measured at time 1 (a1 b1...d1) and at time 2 (a2...d2)
- We estimate two latent categorical variables, with two classes each (latent variables are x at time 1 and y at time 2).
- How can you make sure indicators a1 to d1 are regressed on x and a2 to d2 are regressed on y using Mplus?

Mplus: cross-sectional LCA (cd.)

VARIABLE:
  NAMES ARE a1 b1 c1 d1
  a2 b2 c2 d2
  cova;
  usevar are a1-d1 a2-d2;
  categorical = a1-d1 a2-d2 ;
  classes = x (2) y(2) ;

ANALYSIS:
  TYPE = MIXTURE;
  STARTS = 100 10;
  STITERATIONS = 20;
Mplus: cross-sectional LCA (cd.)

MODEL:

%OVERALL%

MODEL x:

% x#1%

[a1$1-d1$1*-1]

% x#2%

[a1$1-d1$1*1]

No specification regarding the relationship between x and y as yet

MODEL y:

% y#1%

[a2$1-d2$1*-1]

% y#2%

[a2$1-d2$1*1]

Thresholds (therefore: response probabilities) are estimated for a1 b1 c1 and d1 within classes of latent variable x and its categories is preceded by MODEL x:

Thresholds (therefore response probabilities) are estimated for a2 b2 c2 and d2 within classes of latent variable y.

No constraints on thresholds (freely estimated): conditional item response probabilities freely estimated (non-invariance)

Mplus: cross-sectional LCA (ctd.)

Full measurement invariance

MODEL:

%OVERALL%

MODEL x:

% x#1%

[a1$1-d1$1*-1] (1-4);

% x#2%

[a1$1-d1$1*1] (5-8);

Thresholds (therefore: response probabilities) are constrained to be the same for a1 in x1 and a2 in y1, or else: P(a1=1 | x=1) = P (a2 = 1 | y =1). The same is true for indicator b in class 1 of x and y, and so on.

Similar constraints are imposed for class 2 of x and y (5-8)

In this way, we specify a full-measurement invariance model

Mplus output: measurement non-invariance

Latent Class Pattern 1 1

A1

Category 1 0.279 0.038 7.401 0.000

Category 2 0.721 0.038 19.097 0.000

Prob of endorsing (category 2) item a1 if x=1 is 0.729; prob of endorsing item a2 if y=1 is 0.746

B1

Category 1 0.318 0.037 8.697 0.000

Category 2 0.682 0.037 18.662 0.000

Mplus output: measurement invariance

Latent Class Pattern 1 1

A1

Category 1 0.785 0.030 28.290 0.000

Category 2 0.215 0.030 7.401 0.000

Prob of endorsing (category 2) item a1 if x=1 is 0.785 and is the same probability of endorsing item a2 if y=1 (equality constraint imposed)

B1

Category 1 0.722 0.035 20.709 0.000

Category 2 0.278 0.035 7.986 0.000

Latent Class Pattern 2 2

A1

Category 1 0.771 0.026 29.694 0.000

Category 2 0.229 0.026 8.842 0.000

Prob of endorsing (category 2) item a1 if x=1 is 0.771 and is the same probability of endorsing item a2 if y=1 (equality constraint imposed)

B1

Category 1 0.755 0.023 33.225 0.000

Category 2 0.245 0.023 10.789 0.000

Latent Class Pattern 1 1

A1

Category 1 0.785 0.030 28.290 0.000

Category 2 0.215 0.030 7.401 0.000

Prob of endorsing (category 2) item a1 if x=1 is 0.785 and is the same probability of endorsing item a2 if y=1 (equality constraint imposed)
Test for measurement invariance

Run LRT test:

- Non-invariance:
  TESTS OF MODEL FIT
  Loglikelihood
  H0 Value: -5295.298
  H0 Scaling Correction Factor: 1.011
  Information Criteria
  Number of Free Parameters: 18

- Invariance:
  TESTS OF MODEL FIT
  Loglikelihood
  H0 Value: -5300.601
  H0 Scaling Correction Factor: 1.012
  Information Criteria
  Number of Free Parameters: 10

\[
\text{Log-likelihood null model (model with equality constraints)}
\]
\[
\text{Log-likelihood of unconstrained model}
\]

\[
\text{LR} = -2 \cdot (\text{L0} - \text{L1}) / \text{cd}
\]
\[
cd = \left( \frac{\text{p0} \cdot \text{c0} - \text{p1} \cdot \text{c1}}{\text{p0} - \text{p1}} \right)
\]
\[
c0 = \text{scaling factor null model}
\]
\[
c1 = \text{scaling factor alternative model}
\]
\[
p0 = \text{parameters in null model}
\]
\[
p1 = \text{parameters in alternative model}
\]

\[
\text{Cd} = \left( \frac{10 \cdot 1.012 - 18 \cdot 1.011}{10 - 18} \right) = 1.0097
\]

\[
\text{LR} = -2 \cdot \left( \frac{-5300.601 - (-5295.298)}{1.0097} \right) = 10.50
\]

Df = p1 - p0 = 18 - 10 = 8

Chi square (10.50, 8) = .23

The LRT indicates no significant worsening of fit if equality constraints imposed: assume measurement invariance

Partial Measurement Invariance

- Many different options, e.g.:
  - Time-specific structure of one class: in the example class 1 of x and y (time 1 and 2) is freely estimated across time, while equality constraints are imposed on class of x and y (this class is invariant)

MODEL x:
- \( \%a1\%1\% \)
- \( \%a1\%1\%2\%1\% \)
- \( \%a1\%2\%1\% \)

MODEL y:
- \( \%a1\%1\% \)
- \( \%a1\%2\%1\% \)
- \( \%a1\%1\%2\%1\% \)
- \( \%a1\%2\%1\% \)
Partial Measurement Invariance

- Many different options, e.g.:
  - **Differential item functioning with respect to time**: one item (or more) within a class is non-invariant across time (a in class 1 of x and y), while the rest of the parameters are held invariant.

MODEL x:

\[
\begin{align*}
\alpha_{x1} & = \delta1; \\
\alpha_2 & = \delta2; \\
\end{align*}
\]

MODEL y:

\[
\begin{align*}
\alpha_{y1} & = \delta1; \\
\alpha_2 & = \delta2; \\
\end{align*}
\]

Explore transitions based on cross-sectional results

- Before imposing relationships between latent variables, it may be useful to inspect transitions between latent classes estimated cross-sectionally to get some preliminary idea of the type of movement in the sample across time.
- Use the modal class assignment (each individual assigned to the class with highest posterior probability).
- In Mplus: include “IDVAR=idnumber” in VARIABLE This tells Mplus to include an ID variable (idnumber) in the data file.
- Command SAVEDATA writes a file:
  ```
  SAVEDATA:
  SAVE = cprob;
  !cprob includes the modal class assignment and the probability of being in each class for each individual in the sample
  ```

Summary Step 2

- Measurement invariance needs to be investigated before imposing a relationship between latent statuses at each time point.
- Full measurement invariance facilitates estimation and interpretation, but may sometimes not be a plausible assumption.
- If full measurement invariance not tenable, test partial measurement invariance (e.g. a time invariant “normative” class of non-victimised adolescents or non-violent children).

Step 3: Explore specification of the latent transition model without covariates
Step 3: Explore specification of the latent transition model without covariates

- LTA is an autoregressive model: one stage directly related to previously measured stage.
- First order effects ($x \rightarrow y$); Second order effects ($x \rightarrow z$)

Step 3: Explore LTA solution

- 3.1 Impose constraints on transition probabilities
- 3.2 First and second order effects
- 3.3 Stationary transitions (if 3 time measurements and no covariates)
- 3.4 Latent higher-order covariates (Mover-Stayer model)
- 3.5 Model fit

Step 3

- We have settled on class specifications and measurement characteristics of classes across time
- We can now impose auto-regressive relationships between latent variables across time
- In Mplus:
  \[
  \text{CLASSES=} \ x(2) \ y(2);
  \]
  \[
  \text{MODEL:}
  \]
  \[
  \%\text{overall}\%
  \]
  \[
  y \ ON \ x;
  \]
  \[
  \chi{\text{#1}} \%
  \]
  \[
  \text{MODEL x:}
  \]
  \[
  \%\text{overall}\%
  \]
  \[
  y \ ON \ x;
  \]
  \[
  \chi{\text{#1}} \%
  \]
  \[
  \text{This can also be written as:}
  \]
  \[
  \%\text{overall}\%
  \]
  \[
  [x11 \ldots d11] \ (1-4); \ 
  \]
  \[
  \chi{\text{#1}} \%
  \]

Step 3.1: Restricting transition probabilities

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class1</td>
<td>$\tau_{11}$</td>
<td>$\tau_{21}$</td>
</tr>
<tr>
<td>Class2</td>
<td>$\tau_{12}$</td>
<td>$\tau_{22}$</td>
</tr>
<tr>
<td>Class3</td>
<td>$\tau_{13}$</td>
<td>$\tau_{23}$</td>
</tr>
</tbody>
</table>

- Some tau parameters can be fixed
- This can help express a model of development (e.g. No backsliding)
Step 3.1: Restricting transition probabilities (ctd)

- A model of No backsliding among ordered classes:
  - If the classes represented degrees of ability (from 1=less able to 3=more able), the probability of transitioning from a more advanced level to a less advanced one is fixed to 0.

- In this example, we assume there are no transitions from a class at one extreme to a class at the other end (only transitions between adjacent stages allowed).

How to calculate transition probabilities

- The transition probabilities from x to y are given by unordered logistic regression expressions:
  - \( P(y|x=1) = \frac{\exp(a_1 + b_1)}{\text{sum}} \)
  - \( P(y|x=2) = \frac{\exp(a_1 + b_2)}{\text{sum}} \)
  - \( P(y|x=3) = \frac{\exp(a_1 + b_3)}{\text{sum}} \)

  - sum, etc. represent the sum of the exponentiations across the classes of y in rows x (= 1, 2, 3).

  - The values in column y=3 are all 0 (a3=0; b31 =0 ; ...; b33=0) because the last class is the reference class.

How to calculate transition probabilities (ctd.)

- The parameters in the table are in Mplus:
  - a1: [y#1];
  - a2: [y#2];
  - b11: y#1 ON x#1;
  - b12: y#1 ON x#2;
  - b21: y#2 ON x#1;
  - b22: y#2 ON x#2;
How to calculate transition probabilities (ctd.)

<table>
<thead>
<tr>
<th>x</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>b1 + b1</td>
<td>b2 + b3</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>b1</td>
<td>b2 + b3</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>b1</td>
<td>b2</td>
<td>0</td>
</tr>
</tbody>
</table>

Example:

\[ a_1 \rightarrow [y_1] = -1.8; \]
\[ a_2 \rightarrow [y_2] = 0.3; \]
\[ b_{11} \rightarrow y_1 \text{ ON } x_1 = 2.6; \]
\[ b_{12} \rightarrow y_1 \text{ ON } x_2 = 2.1; \]
\[ b_{21} \rightarrow y_2 \text{ ON } x_1 = -1.3; \]
\[ b_{22} \rightarrow y_2 \text{ ON } x_2 = 0.5; \]

\[ P(y=1|x=1) = \frac{\text{exp}(a_1+b_{11})}{\text{exp}(a_1+b_{11}) + \text{exp}(a_2+b_{21}) + \text{exp}(0)} \]
\[ P(y=1|x=1) = \frac{\text{exp}(-1.8+2.6)}{\text{exp}(-1.8+2.6)+\text{exp}(0.3+(-1.3))+1} \]
\[ P(y=2|x=1) = \frac{\text{exp}(a_2+b_{21})}{\text{exp}(a_1+b_{11}) + \text{exp}(a_2+b_{21}) + \text{exp}(0)} \]
\[ P(y=2|x=1) = \frac{\text{exp}(0.3+(-1.3))}{\text{exp}(-1.8+2.6)+\text{exp}(0.3+(-1.3))+1} \]
\[ P(y=3|x=1) = \frac{\text{exp}(a_3+b_{31})}{\text{exp}(a_1+b_{11}) + \text{exp}(a_2+b_{21}) + \text{exp}(0)} \]
\[ P(y=3|x=1) = \frac{\text{exp}(0)}{\text{exp}(-1.8+2.6)+\text{exp}(0.3+(-1.3))+1} \]

If we want to fix \( P(y=3|x=1) \) to 0 we refer to the formula for its probability:

\[ P(y=3|x=1) = \frac{\text{exp}(a_3+b_{31})}{\text{exp}(a_1+b_{11}) + \text{exp}(a_2+b_{21}) + \text{exp}(0)} \]

In this case, to ensure the result is 0, we make the value of \( \text{exp}(a_1) \) very small by assigning to \( a_1 \) a large negative number (for example, -15).

Since parameter \( a_1 \) in Mplus is indicated by the logit intercept \([y_1] \):

\[ \text{MODEL} \]
\[ \text{[y1@-15];} \]

Thus, we should obtain that, whatever the other parameters, \( P(y=1|x=3) \approx 0 \)

\[ \tau_{1|3} = 0 \]

Second order effects

- First order effects \((x \rightarrow y ; y \rightarrow z)\): if no second order effects, non-adjacent latent variables are indirectly related
- Second order effects \((x \rightarrow z)\): lasting direct effects that being in category of \( x \) has on later class membership
Second order effects (ctd.)

VARIABLES:
- Classes = x(2) y(2) z(2);

MODEL:
%overall%
y ON x;  
z ON y;  
z ON x;  

Can also be written:
y ON x;  
z ON y;

MODEL:
%overall%
y ON x;

z ON y;

z ON x;

2nd order x

Can also be written:
y on x;

z ON x y;

Can also be written:
y#1 ON x#1;  
y#1 ON y#1;  

Can also be written:

y#1 ON x#1;  
y#1 ON y#1;

Second order effects (ctd.)

• Inspection of transition probabilities matrices estimated under different assumption (first- vs. second-order effects) help highlight impact of previous classification

<table>
<thead>
<tr>
<th></th>
<th>Grade 6</th>
<th>Grade 7</th>
<th>Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Victimised</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Victim</td>
<td>02</td>
<td>10</td>
<td>08</td>
</tr>
<tr>
<td>Victim</td>
<td>02</td>
<td>10</td>
<td>08</td>
</tr>
<tr>
<td>some victim</td>
<td>04</td>
<td>04</td>
<td>02</td>
</tr>
<tr>
<td>Non victim</td>
<td>00</td>
<td>06</td>
<td>05</td>
</tr>
</tbody>
</table>

Stationary transitions

• Assume transitions across time points (> 2) are stationary: same probabilities to transition from a stage to another between time 1- time 2 and between time 2-time 1, and so on...
• However, if covariates are included, stationarity is no longer meaningful (it would bias estimation of covariates' coefficients)

\[
\begin{align*}
x & \rightarrow y \rightarrow z \\
x & \rightarrow y & y \rightarrow z \\
& \rightarrow \text{transitions} \\
& \text{Non-Victim, Victim} \\
& \text{Stationary transitions (ctd.)} \\
\end{align*}
\]
Stationary transitions (ctd.): Output

**LATENT TRANSITION PROBABILITIES BASED ON THE ESTIMATED MODEL**

X Classes (Rows) by Y Classes (Columns)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.339</td>
<td>0.661</td>
</tr>
<tr>
<td>2</td>
<td>0.863</td>
<td>0.137</td>
</tr>
</tbody>
</table>

Y Classes (Rows) by Z Classes (Columns)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.339</td>
<td>0.661</td>
</tr>
<tr>
<td>2</td>
<td>0.863</td>
<td>0.137</td>
</tr>
</tbody>
</table>

Same transition probabilities: change happens at the same rate across time points

**Higher-order latent variables**

- It is possible to estimate a further latent variable to investigate unobserved heterogeneity in developmental process.
- For example: a latent class of "movers" (individuals that transition between stages across measurement occasions) and one of "stayers" (individuals that remain in the same class across measurement occasions).
- E.g. If x and y are classes of depression, mover/stayer model help identify individuals chronically depressed

**Movers/stayers model**

- Allows more accurate estimation of transition probabilities if, indeed, there are individuals with zero probability of transitioning.
- Pre-requisite: same number of classes with same meaning (measurement invariance).

Movers: freely estimate the probability of transitioning across time points

Stayers: fix the probability of transitioning across time points to 0

**How to calculate transition probabilities with covariates**

<table>
<thead>
<tr>
<th>STAYERS</th>
<th>Time 1</th>
<th>Time 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Class2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Class3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Time 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>y2</td>
</tr>
<tr>
<td>x1</td>
<td>a1∗b11+g1∗(ms)</td>
</tr>
<tr>
<td>x2</td>
<td>a1+g1*ms</td>
</tr>
</tbody>
</table>

- The Mover/Stayer latent variable (ms) is a (latent) covariate of the two latent variables x (time 1) and y (time 2).
- The latent variable ms has two categories.
  - One category of ms (the last one) is the reference category.
  - The coefficient g describes the change in log odds for one category of ms as compared to the reference category.

If ms = 1

\[ P(y=1|x=1) = \frac{\text{EXP}(a1 + b11+g1*ms)}{\text{EXP}(a1+b11+g1*ms) + \text{EXP}(0)} \]

If ms = 2 (ref. Cat.)

\[ P(y=1|x=1) = \frac{\text{EXP}(a1 + b11) }{\text{EXP}(a1+b11) + \text{EXP}(0)} \]
How to calculate transition probabilities with covariates (ctd.)

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Time 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>y1</td>
</tr>
<tr>
<td>y2</td>
<td>y2</td>
</tr>
<tr>
<td>x1</td>
<td>a1+b1+g1(ms)</td>
</tr>
<tr>
<td>x2</td>
<td>a1+g1(ms)</td>
</tr>
</tbody>
</table>

If ms = 1
\[ P(y=1|x=1) = \frac{\exp(a1 + b1 + g1(ms))}{\exp(a1 + b1 + g1(ms)) + \exp(0)} \]

If ms = 2 (ref. Cat.)
\[ P(y=1|x=1) = \frac{\exp(a1 + g1(ms))}{\exp(a1 + g1(ms)) + \exp(0)} \]

Assume ms=1 is the mover class and ms=2 the stayer.

Fixing \( x \rightarrow \{y|=1\} = -15 \) ensures that in category 2 of ms (reference category)
\[ P(y=2|x=1) = 0 \quad (0 \text{ prob. of moving from 1 to 2}) \]

If ms = 2 \( \rightarrow g1(ms)=0 \)
\[ P(y=2|x=1) = \exp(a1 + g1(ms)) \]

This had been fixed
\[ P(y=1|x=2)=0 \text{ in ms#2} \]

When a higher-order latent class is
introduced

This specifies measurement invariance: thresholds of x#2 the same as y#2
It is possible to specify different measurement constraints in ms1 and
ms2, or for combinations of ms, x, y

Mover / Stayer model in Mplus

**VARIABLE:**
- CLASSES = ms(2) x(2) y(2)

**MODEL:**
- OVERALL:
  - Regresses x and y ON ms
  - (mover/stayer)
  - Fixes prob of y=1|x=2 in ms2 to 0
- freed in ms1 (i.e. not equal across ms classes)

- MS1:
  - y#1 ON x#1
  - y#1 ON x#1@30
  - Fixes P(y=1|x=1)=1 in ms2

- MS2:
  - y#1 ON x#1
  - This had been fixed
  - P(y=1|x=2)=0 in ms#2

**output**

Category Latent Variables

<table>
<thead>
<tr>
<th>X#1</th>
<th>MS#1</th>
<th>Y#1</th>
<th>MS#1</th>
</tr>
</thead>
<tbody>
<tr>
<td>X#1</td>
<td>3.012</td>
<td>2.104</td>
<td>-1.432</td>
</tr>
<tr>
<td>MS#1</td>
<td>14.045</td>
<td>0.000</td>
<td>999.000</td>
</tr>
<tr>
<td>Y#1</td>
<td>-15.000</td>
<td>0.000</td>
<td>999.000</td>
</tr>
</tbody>
</table>

Means

<table>
<thead>
<tr>
<th>Subject</th>
<th>MS#1</th>
<th>Y#1</th>
</tr>
</thead>
<tbody>
<tr>
<td>X#1</td>
<td>2.284</td>
<td>2.056</td>
</tr>
<tr>
<td>Y#1</td>
<td>-15.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

This had been fixed
\[ P(y=1|x=2)=0 \text{ in ms#2} \]

This had been freed in
ms#1 (i.e. not equal
across ms classes)

This had been fixed
\[ P(y=1|x=1)=1 \text{ in ms#2} \]
Mover / Stayer model in Mplus: output (ctd.)

**Final Class Counts and Proportions for the Latent Classes Based on Estimated Posterior Probabilities**

<table>
<thead>
<tr>
<th>Latent Class</th>
<th>Pattern</th>
<th>Count</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1 1</td>
<td>ms=1 x=1 y=1</td>
<td>5.64118</td>
<td>0.00564</td>
</tr>
<tr>
<td>1 1 2</td>
<td>ms=1 x=1 y=1</td>
<td>236.08630</td>
<td>0.23609</td>
</tr>
<tr>
<td>1 2 1</td>
<td>ms=1 x=1 y=2</td>
<td>208.96989</td>
<td>0.20897</td>
</tr>
<tr>
<td>1 2 2</td>
<td>ms=1 x=1 y=2</td>
<td>244.08220</td>
<td>0.24408</td>
</tr>
<tr>
<td>2 1 1</td>
<td>ms=2 x=2 y=2</td>
<td>279.44841</td>
<td>0.27945</td>
</tr>
<tr>
<td>2 1 2</td>
<td>ms=2 x=2 y=2</td>
<td>0.00009</td>
<td>0.00000</td>
</tr>
<tr>
<td>2 2 1</td>
<td>ms=2 x=2 y=2</td>
<td>0.00001</td>
<td>0.00000</td>
</tr>
<tr>
<td>2 2 2</td>
<td>ms=2 x=2 y=2</td>
<td>25.77192</td>
<td>0.02577</td>
</tr>
</tbody>
</table>

**Movers (ms=1)**

**Stayers (ms=2)**

---

**Model fit of LTA models**

- The Chi-Square statistics (Pearson or Likelihood-ratio based) not recommended (distribution not well approximated when large number of sparse cells)
- Nested models (e.g. Stationary transitions vs. non-stationary) → compare with LRT (remember correction by scaling factor if MLR estimator)
- Consider residuals (less significant residuals → better fit)

*Important to build model step by step*

---

**Summary Step 3**

- Impose autoregressive relationships (current status predicted by previous status)
- Consider and test constraints on transition probabilities
- If more than 2 time points, it is possible to consider stationary transitions (but not meaningful if covariates are included) and second-order effects
- It is possible to include higher-order latent covariates (e.g. Movers / Stayers model)

**Step 4: Include covariates in the LTA model**
Step 4: Include covariates in the LTA model

- Categorical, nominal and continuous covariates can be included as predictors of class membership and transition probabilities
- Covariates can be time-varying or time-invariant
- They can have time-varying or time-invariant effects (independently of their being time-varying or not)

Step 4: Categorical covariates

- If covariates are categorical (e.g. Gender) : multiple-groups LTA
  - It is possible to explore measurement invariance across groups: e.g. Do items map onto the latent variables in the same way for males and females?
  - Explore differences in latent class membership at start point
    - E.g. Does probability of being victimised in Grade 6 differ between males and females?
  - Explore differences in transition probabilities
    - E.g. Does probability of transitioning from victimised to non-victimised differ between males and females?

Investigating measurement invariance

- This model can be extended to LTA models
Explore differences in latent class membership at start point

- E.g. Does probability of being victimised in Grade 6 differ between males and females?
  - VARIABLES: usevar are male a1 b1 c1 d1 a2 b2 c2 d2;
  - classes = x(2) y(2);
  - Model:
    - x on male;
    - y on x;
    - Model x:
      - %x#1%

In this case, class membership at time 2 is predicted by class membership at time 1 (x) and NOT by gender: transition probabilities from x to y are the same for males and females.

Logistic regression of x on male

Explore differences in transition probabilities

- E.g. Does probability of transitioning from victimised to non-victimised differ between males and females?
  - VARIABLES: usevar are male a1 b1 c1 d1 a2 b2 c2 d2;
  - classes = x(2) y(2);
  - Model:
    - x on male;
    - y on x male;
    - Model x:
      - %x#1%

In this case, class membership at time 2 is also predicted by gender while controlling for previous latent status: transition probabilities from x to y differ between males and females.

Logistic regression of x on male

Transition probabilities with categorical covariates

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Time 2</th>
<th>y1</th>
<th>y2</th>
<th>y3</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>a1+b11+g1(male)</td>
<td>a2+b21+g2(male)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>x2</td>
<td>a1+b12+g1(male)</td>
<td>a2+b22+g2(male)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>x3</td>
<td>a1+g1(male)</td>
<td>a2+g2(male)</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

$g_1$ and $g_2$ are the logistic coefficients: change in the log odds of being in class $y_1$ or $y_2$ compared to class $y_3$ (reference) for males (male=1) as compared to females (male=0).

In Mplus these parameters are:

- $g_1$ and $g_2$ terms are equal to 0 (reference class).
- $P(y=1|x=1)$ for females (male=0) is calculated adding $g_1$ and $g_2$ parameters.
- $P(y=1|x=1)$ for males (male=1) is:

\[
P(y=1|x=1) = \frac{\exp(a1+b11+g1)}{\exp(a1+b11+g1) + \exp(a2+b21+g2) + \exp(0)}
\]

Transition probabilities for females (male=0) can be calculated considering that the $g_1$ and $g_2$ terms are equal to 0 (reference class).

\[
P(y=1|x=1) = \frac{\exp(a1+b11)}{\exp(a1+b11) + \exp(a2+b21) + \exp(0)}
\]
**Transition probabilities with categorical covariates**

Table 3.7: Estimated transition probabilities by gender (males on the left, females on the right) based on model with only gender as a covariate

<table>
<thead>
<tr>
<th></th>
<th>Males Gender 7</th>
<th>Females Gender 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>VT</td>
<td>0.42 0.44 0.16</td>
<td>VT 0.42 0.26 0.19</td>
</tr>
<tr>
<td>NT</td>
<td>0.03 0.21 0.44</td>
<td>NT 0.03 0.44 0.53</td>
</tr>
<tr>
<td>VT</td>
<td>0.01 0.11 0.88</td>
<td>VT 0.01 0.69 0.31</td>
</tr>
</tbody>
</table>

One can thus obtain transition matrices for males and females.

**Multigroup LTA in Mplus**

- Another possibility is to use the KNOWNCLASS option:

  - VARIABLES: usevar are a1 b1 c1 d1 a2 b2 c2 d2; **NOTE: no male classes = cmale(2) y(2);**
  - KNOWNCLASS = cmale (male = 0 male = 1);
  - MODEL:
    - x on cmale;
    - y on x cmale;
    - Model x:
      - %xa#1%
    - ...

  Defines a new class for which class membership is known (observed).

- The known class is used as a predictor of latent class membership at time 1 and time 2.

- In this case, different thresholds (item response prob.) are estimated for females and males, but these are invariant at time 1 and 2 within groups.

  - Model cmale:x:
    - %cmale#1.x#1% [a1S1-d1S1] [1-4];
    - %cmale#1.x#2% [a1S1-d1S1] [5-8];
    - %cmale#2.x#1% [a1S1-d1S1] [9-12];
    - %cmale#2.x#2% [a1S1-d1S1] [12-16];

  - Model cmale:y:
    - %cmale#1.y#1% [a2S1-d2S1] [1-4];
    - %cmale#1.y#2% [a2S1-d2S1] [5-8];
    - %cmale#2.y#1% [a2S1-d2S1] [9-12];
    - %cmale#2.y#2% [a2S1-d2S1] [12-16];
Estimation with covariates

- The inclusion of covariates changes estimation of LTA parameters, including class profiles, class size and transition probabilities (see formulae for calculating transition probabilities with covariates).
  - This is also the reason why stationary transition probabilities are not meaningful when covariates are included in the model. Imposing these constraints would bias estimation of covariates coefficients.
- If adding covariates changes the class structure substantially, this might point to the need to allow for measurement non-invariance (more investigation needed).

Relating LCA results


If one does not want to include covariates while estimating latent classes, there are different approaches where the latent class membership is regressed on covariates. E.g.:
- Consider the most likely class (modal class assignment based on posterior probabilities)
  - In this case, class membership is used as an observed variable ignoring the fact that individuals have different probabilities of being in one class.
- Weight regression by each individual's posterior probability of being in a given class.
- Clark & Muthén: including covariates while forming latent classes still performed the best.

Summary Step 4

- Covariates can be time-varying or time invariant.
- Interval or categorical covariates can be used to predict class affiliation at first measurement point and changes in transition probabilities.
- Categorical covariates: multiple groups LTA.
- Covariates may substantially change LTA parameters, including measurement parameters. This may warrant further investigation (e.g. DIF).
Step 5: Include distal outcomes

- Variables measured after the period considered by the model can be included as long-term outcomes related to the change process.
- Distal outcomes can be included in different ways. E.g.:
  - Can be related to a higher-order latent variable such as Mover-Stayer classification
  - Can be related to the latent status at the last time point of measurement

Distal interval outcomes in Mplus

- The interval variable is testscor

VARIABLES are male a1 b1 c1 d1 a2 b2 c2 d2 testscor;
usevar are a1-d2 testscor;
categorical are a1-d2;
classes are x(2) y(2);

MODEL:
%overall%
y ON x;

MODEL x:
%#1%
[a1-d1] (1-4);
%#2%
[a2-d2] (5-8);

MODEL y:
%#1%
[a2-d2] (1-4);
%#2%
[a2-d2] (1-4);

MODEL TEST:
p1 = p2;

Estimates means of testscor in y1 and y2: in MODEL command an interval variable name between brackets indicates the variable mean

testscor is in the USEVAR but not CATEGORICAL statement (therefore: interval variable)

This provides Wald test for H0 : p1 = p2

Step 5: Include distal outcomes (ctd)

- Distal outcomes of different type (e.g. categorical or interval variables) can be included in LTA.
- In the case of interval variables, the variable means can be estimated for each class of the latent variable; these means can be compared to investigate significant differences.
- In the case of binary variables, proportions are estimated for each class of the latent variable.
A binary distal outcome (or a categorical one) can be included in the same way that other categorical indicators are regressed on the latent variable.

According to Muthén, statistically the distal outcome is another latent class indicator (although one thinks of it in substantively different terms).

The inclusion of a distal covariate may change some LTA parameters:
- If this is the case, this warrants further investigation.

The outcome (binary) variable is `testbin`.

Variables: names are male a1 b1 c1 d1 a2 b2 c2 d2 testbin;

Usevar are a1-d2 testbin;

Categorical are a1-d2 testbin;

Classes are x(2) y(2);

Model:

%overall%
y ON x;

MODEL x:

%x#1%
[a1$1-d1$1] (1-4);

%x#2%
[a1$1-d1$1] (5-8);

MODEL y:

%y#1%
[a2$1-d2$1] (1-4);

[\text{testbin$1$}];

%y#2%
[a2$1-d2$1] (1-4);

[\text{testbin$1$}];

Estimates thresholds of testbin in y1 and y2 (hence, proportions):

Proportion of “pass” scores (cat.2) in testbin is 60% in y1 and 43% in y2.

OUTPUT:

RESULTS IN PROBABILITY SCALE

Latent Class Pattern 1 1

A1
Category 1 0.974 0.005 199.237 0.000
Category 2 0.026 0.005 5.409 0.000

[\text{testbin}]

Category 1 0.395 0.012 33.182 0.000
Category 2 0.605 0.012 50.832 0.000

Latent Class Pattern 1 2

A1
Category 1 0.974 0.005 199.237 0.000
Category 2 0.026 0.005 5.409 0.000

[\text{testbin}]

Category 1 0.568 0.023 24.839 0.000
Category 2 0.432 0.023 18.866 0.000

We specified estimation of testbin only in latent variable y, so will consider the different y classes.
• OUTPUT:

LATENT CLASS ODDS RATIO RESULTS

Latent Class Pattern 1 1 Compared to Latent Class Pattern 1 2

<table>
<thead>
<tr>
<th>TESTBIN Category</th>
<th>Odds Ratio</th>
<th>SE</th>
<th>Est/SE</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category &gt; 1</td>
<td>2.017</td>
<td>0.222</td>
<td>9.081</td>
<td>0.000</td>
</tr>
</tbody>
</table>

This means that compared to individuals in class 2 of y, individuals in class 1 of y are 2.017 times more likely to have a TESTBIN score greater than category 1 than they are to have a score in category 1. Since there are only 2 categories and category 2 is the “pass” score: compared to individuals in y2 individuals in y1 are 2 times more likely to have a pass score than they are to have a fail score.

• If you want to treat the distal binary outcome as a different variable (not a latent class indicator) some options available:
  - create a binary latent variable measured by the binary indicator (your outcome) without error, then regress this variable on the latent class of interest (the predictor)
  - create a binary latent variable measured by the binary indicator with error (a LC measurement model of your outcome), then regress this on the predictor latent class

• These approaches are not encouraged
Further application

- Associative Latent Transition Analysis (ALTA):
  – Multiprocess model → examine change over time in two or more discrete developmental processes

Further applications
References and resources

- A great resource to learn about stats in general:
  http://www.ats.ucla.edu/stat/
  Including examples from LCA textbooks:
  http://www.ats.ucla.edu/stat/mplus/examples/
- Mplus web page (visit the “Mplus Web Notes” and the “Short Course Videos and Handouts” pages for tutorials and examples)
  http://www.statmodel.com/
References and resources

- Nylund’s dissertation on LTA (includes input files of some of the models tested):
- Bray’s dissertation on “advanced latent class modeling techniques” (also includes Mplus input files):
  http://www.statmodel.com/download/Bray%20Dissertation%20%282007%29

References and resources


References and resources