PSY792F SEM

Week 10 – Growth Mixture Modeling

Mark A. Prince, PhD, MS

Example from Prince & Maisto (2013)

- I am going to use data and results from a paper that looked at joinpoint analysis as a post-hoc procedure for GMM
- I will not be going over joinpoint analysis, in detail, but if you are interested in it let me know.
- I will use some of the tables and figures from that paper as well as a modified version of the write-up.
- I'll provide code for a simulated example and the code from that paper.
- In that paper, I used language implying that people are part of a latent class, keep in mind that everyone has a probabilistic membership of being in every class
  - In my more recent papers I have worked harder to remove that type of language

Latent Growth Mixture Modeling

- LGMM or just GMM
- Is an extension of LPA that combines latent growth curve modeling with mixture modeling.
- Looking for latent classes of trajectories across time.
- Used when you have continuous indicators at multiple timepoints
- If you have dichotomous indicators you need Latent Transition Analysis, which I am not going to cover in class.
- All the same rules and assumptions apply for GMM as for LGCM and LPA

Similarities to LGCM and LPA

- LGCM
  - Time scores – can specify linear, quadratic
  - Latent intercepts and slopes
  - Model fit to determine trajectory (growth)

- LPA
  - Seeking to maximize within group homogeneity and between group heterogeneity, but this time it is with regard to intercepts and slopes.
  - Model fit to determine the number of classes
  - Auxiliary statements still work

Limitations to LPA and LGCM

- LPA
  - Used for cross-sectional data
  - Does not provide indices of change, growth
- LGCM
  - Used for longitudinal data
  - Does not identify latent patterns of growth
    - Treats growth as a single homogenous trajectory

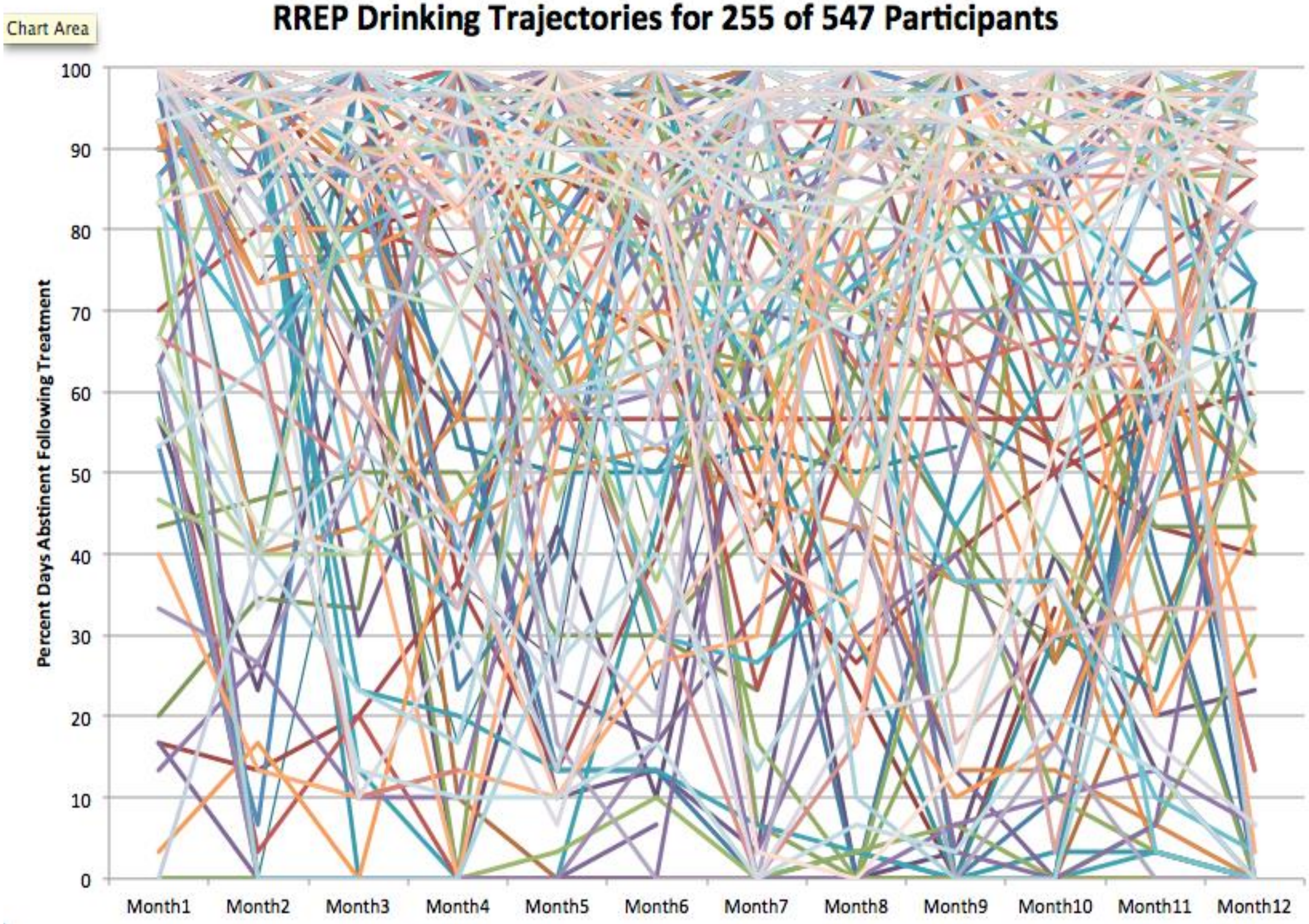
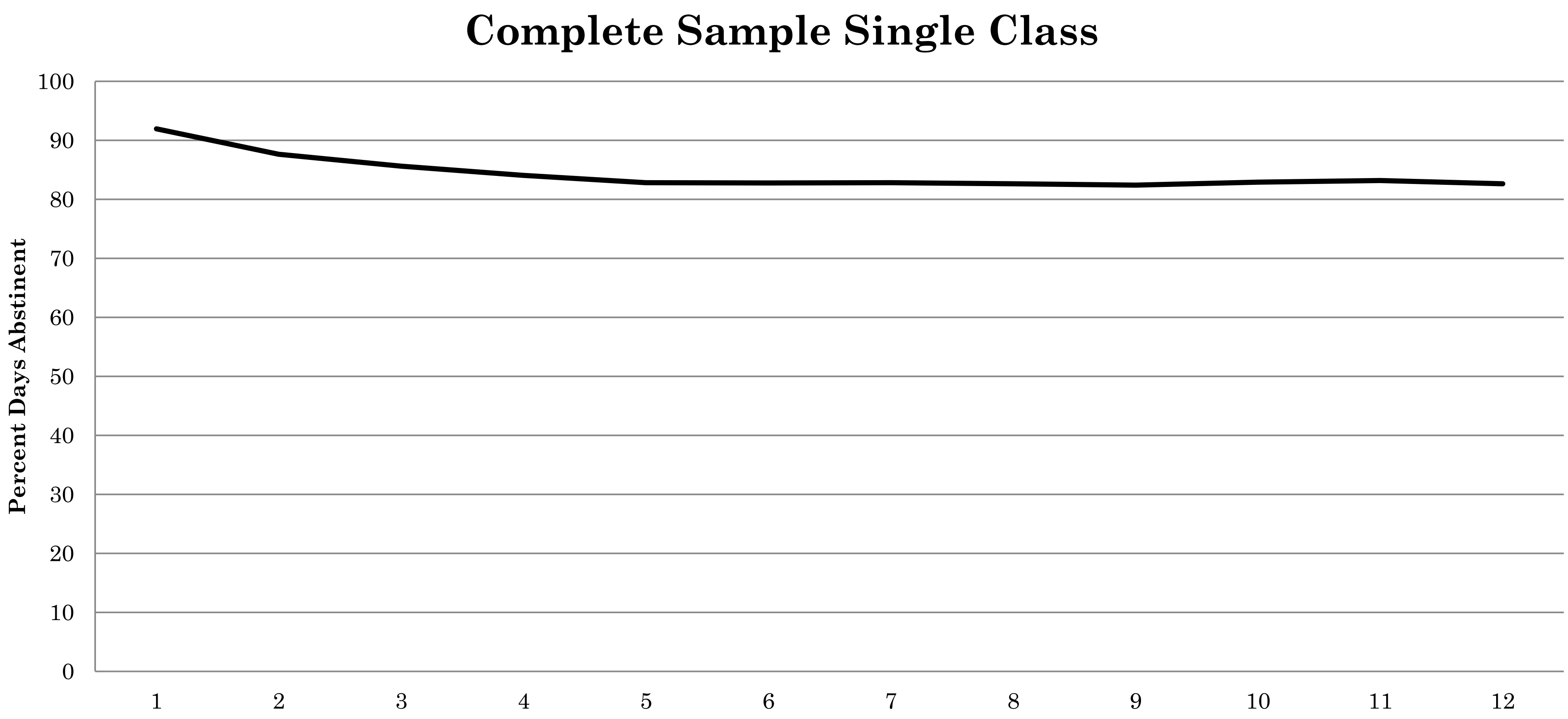


Chart Area
RREP Drinking Trajectories for 255 of 547 Participants
Percent Days Abstinent Following Treatment
100
90
80
70
60
50
40
30
20
10
0
Month1
Month2
Month3
Month4
Month5
Month6
Month7
Month8
Month9
Month10
Month11
Month12

LGCM

- Aggregate everyone to the mean
- Treats the sample as Homogeneous



Growth Mixture Models (GMM)

- Identifies subgroups of individuals who have similar trajectories to one another within a group, but have different trajectories from individuals in the other groups.
  - Maximizes within group homogeneity while also maximizing between group heterogeneity

- NOTE: Individuals are not assigned to groups but each person has a probabilistic membership in all groups
  - Model fit takes into consideration how well the class structure fits the data

Model Fit

- Same as for LPA
- I still prefer LMR test, but another option is the Bootstrapped Likelihood Ratio Test, which is a similar test comparing a model with k classes to a model with k-1 classes
- Also, sample size adjusted BIC is an alternative to BIC, which I tend to use if I have under ~500 observations

Selecting # of Groups

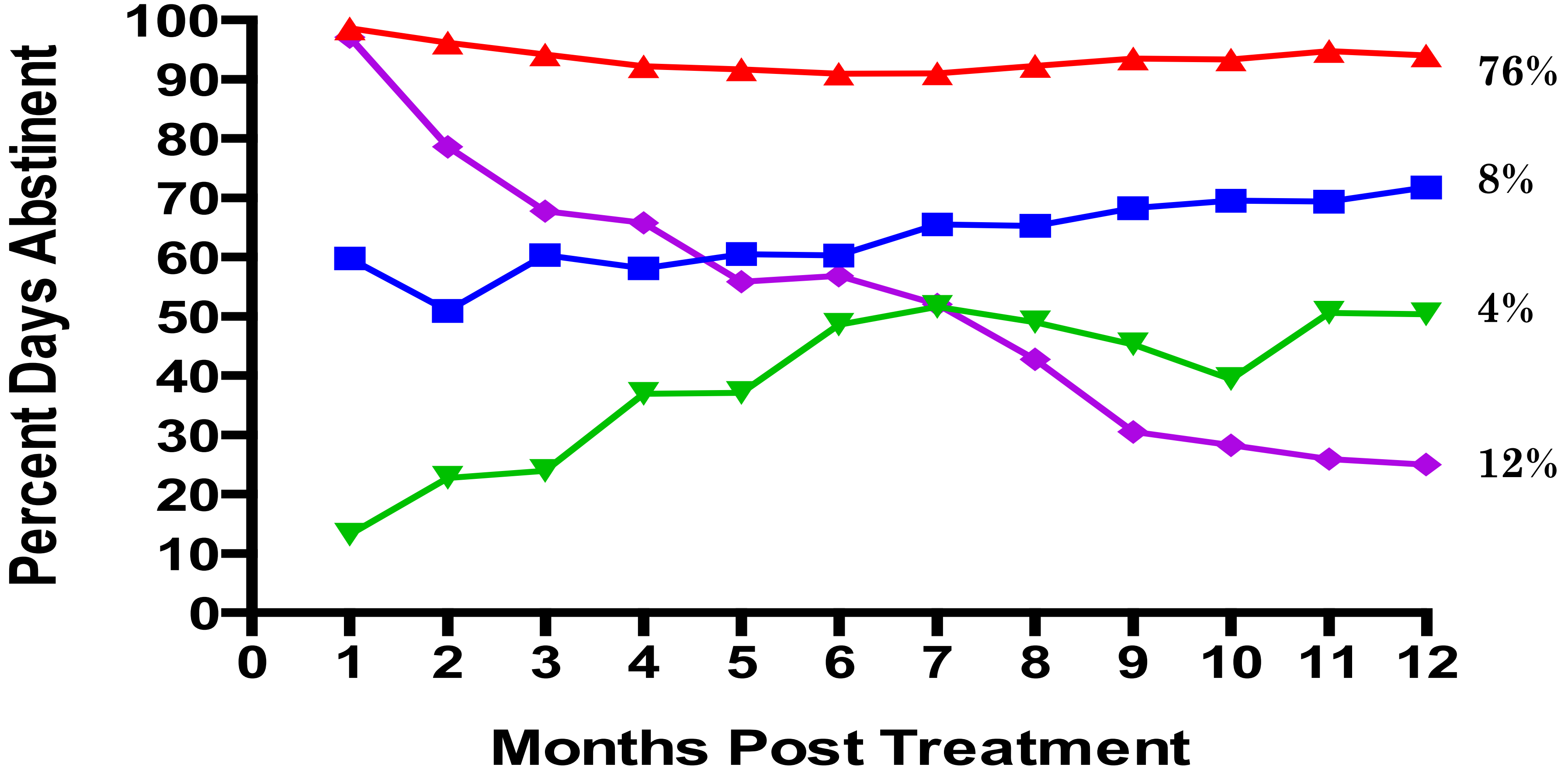
PDA					
	1	2	3	**4**	5
saBIC	54660	54160	53772	**53588**	53584
Entropy	1	0.99	0.995	**0.97**	0.97
BLRT Improvement	N/A	<.01	<.01	**<.01**	<.01
# People/Class					
1	549	495	483	**44**	3
2		54	44	**416**	22
3			22	**22**	44
4				**67**	416
5					64
4 class model	SDFD	SOD	GDFD	IFD	

Classification Quality

Average Latent Class Probabilities for Most Likely Latent Class Membership

	1	2	3	4
1	**0.981**	0.019	0.000	0.000
2	0.000	**0.990**	0.000	0.010
3	0.001	0.000	**0.999**	0.000
4	0.000	0.073	0.000	**0.927**

RREP PDA 4-Class Model



■ SDFD = slightly decreasing frequency drinkers

▲ SOD = stable occasional drinkers

▼ GDFD = greatly decreasing frequency drinkers

◆ IFD = increasing frequency drinkers

Conceptual Model

In Mplus the categorical latent variable derived by GMMs are treated as second order latent variables

C – the latent class variable is defined by the intercept, the slope, and any other mixture indicators (U)

Covariates can be added to predict C, i, or s

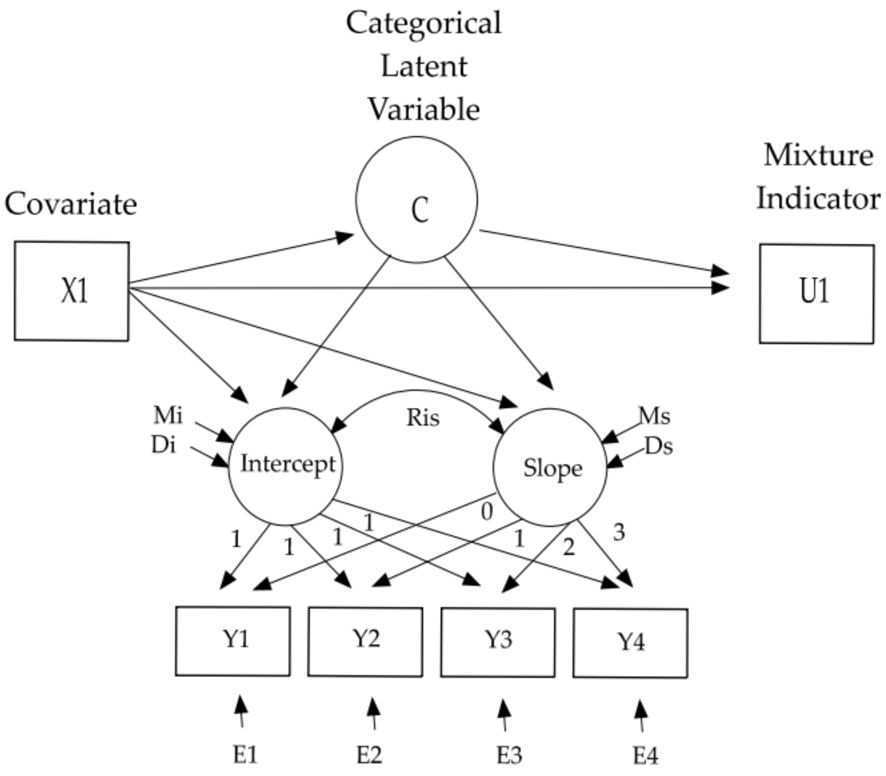


FIG. 1. Representation of the growth mixture model.

Describing the Trajectories

- LGMM models allow for the estimation of trends for each subgroup
  - These trends often do not fully capture the abrupt changes commonly seen in alcohol use across time
    - See Prince & Maisto 2013 for an example of Joinpoint Analysis which can be used as a post-hoc procedure

How to write the code (simulated data)

```
DATA:	FILE IS ex8.1.dat;

VARIABLE:	NAMES ARE y1-y4 x;

	CLASSES = c (2);

ANALYSIS:
	TYPE = MIXTURE;
	STARTS = 40 8; !This is a small number I often use 500 25;
MODEL:
	%OVERALL%
	i s | y1@0 y2@1 y3@2 y4@3;
	i s ON x; !x is a predictor of the intercept and linear slope
	c ON x; !x is also a predictor of the latent class variable
OUTPUT:
TECH1 TECH8; !I like tech7 tech11 and tech14
```

Real Data

- From Prince & Maisto 2013
- See examples “6.11.10 Drinking GMM PDA 1 class.inp” through “6.11.10 Drinking GMM PDA 5 class.inp”

How to write up the results

- Same as before
  - Any information about your variables
  - Model building strategy and model selection
    - Model fit indices
    - Growth patterns you compared
    - Number of classes you compared
  - Table comparing overall model fit
  - Figure of final model
  - Describe the final model
  - Any auxiliary variables

Model Fitting Issues

- What to try when models don't converge
  - Typically due to EM algorithm reaching a local maxima
  - Increase starting values
  - Set starting values based on previous models
    - E.g., if your 3 class model converged but your 4 class won't you can start with the starting values from the 3-class model
  - Fix variances of intercepts and slopes to 0
    - Assumes residuals are constant across classes
    - i@0 s@0;