U.S. Young Adult Arrest Rate Trajectory

U.S. Average County-level Young Adult Arrest Rate Trajectory

All model results

Chart, line chart

Description automatically generated

best model, k=6, x = 3

Chart, line chart

Description automatically generated

**Individual trajectories by group**

A picture containing text, grass

Description automatically generated

A picture containing diagram

Description automatically generated

Predicted Values for Trajectory group vs weighted means vs non-weighted means  
It ends up that often the weighted means are very nearly equivalent to the unweighted means (just aggregating means based on the classified group). In this example the predicted values are a colored line, the weighted means are a colored line with superimposed points, and the non-weighted means are just a black line. You can see the non-weighted means are almost exactly the same as the weighted ones. For group 3 you typically need to go to the hundredths to see a difference.

Chart, line chart

Description automatically generated

#plot of maximum posterior probabilities

Two types of absolute fit measures I’ve seen advocated in the past are the average maximum posterior probability per group and the odds of correct classification. The occ function calculates these numbers given two vectors (one of the max probabilities and the other of the group classifications). We can get this info from our long data by just selecting a subset from one time period. Here the output at the console shows that we have quite large average posterior probabilities as well as high odds of correct classification. (Also updated to included the observed classified proportions and the predicted proportions based on the posterior probabilities. Again, these all show very good model fit.) Update: Jeff Ward sent me a note saying I should be using the predicted proportion in each group for the occ calculation, not the assigned proportion based on the max. post. prob. So I have updated to include the occ\_pp column for this, but left the old occ column in as a paper trail of my mistake.

Graphical user interface, application, table

Description automatically generated

> occ(longD)

group AvePP Freq OCC ClassProp PredProp occ\_pp

1 1 0.9998505 239 81189.9811 0.07609042 0.07622467 81035.2081

2 2 0.9995052 488 10982.6300 0.15536453 0.15563904 10959.6966

3 3 0.9996334 521 13713.2468 0.16587074 0.16645738 13655.3077

4 4 0.9986013 621 2897.2372 0.19770774 0.19795498 2892.7268

5 5 0.9990743 621 4379.5840 0.19770774 0.19824776 4364.7142

6 6 0.9907537 651 409.8429 0.20725883 0.20547617 414.3282

**#scatterplot matrix**

Remember that these latent class models are fuzzy classifiers. That is each point has a probability of belonging to each group. A scatterplot matrix of the individual probabilities will show how well the groups are separated. Perfect separation into groups will result in points hugging along the border of the graph, and points in the middle suggest ambiguity in the class assignment. You can see here that each group closer in number has more probability swapping between them.

Calendar

Description automatically generated