# Mixture and Group-Based Trajectory Models – part 2.

## Transcript

Video: https://youtu.be/ocGXO7ST3TM

 First, I will summarise the key concepts I introduced in my first presentation. Growth mixture models and latent class growth analysis are person-centred approaches that identify groups with distinctive developmental trajectories. These models are based on probability, so they provide robust and transparent methods for classification of individuals into different groups. And the difference between the two methods is that growth mixture models allow infra-group variation in the growth parameters, whereas latent class growth analysis doesn’t allow any variation in the growth parameters of the individuals within one class. And I also emphasized the importance of exerting some judgement in selecting the models that can represent variation in trajectories of variables we observe across individuals.

 In this second presentation, I will give a more formal introduction and outline of the trajectory-based group models, and I will make the case for looking at categorical outcomes for uncategorical variables, introducing some of the key concepts that allow us to work with these type of variables. I will then talk about how we can interpret the parameters and how we can plot trajectories, both for growth mixture models and latent class growth analysis.

 So I will focus on categorical variables because when we are studying issues such as drug use or mental health, often we deal with variables that are ordered categorical. In the example I provided here, the question about low mood has category of responses that are not evenly spaced, and are ordered categorical. This means that we cannot apply models that have been developed for continuous outcomes, and models that often have prescriptive assumptions about the distribution of residuals.

 So I will take as an example, a variable like antisocial behaviour, that I labelled A, where A can take five ordered values from zero to four, but these values are not evenly spaced. In a similar case, we can use a logging link to model the association between the probability of the variable taking a category of response, for example, two or higher, versus the probability of the variable assuming a lower category in the response. This ratio between the two probabilities represents the odds of variable A being in category two or higher, versus the probability of a lower category. And by calculating the logarithm of these odds, we give to this outcome a symmetric distribution which allows us to model the outcome as a linear function of covariant X.

 In a category called outcomes, as a function of covariates, it’s facilitated further by a latent response formulation. In this case, the category called outcome is considered a rough categorisation of an underlying latent response, as you can see in this graph I produced here. This continuous underlying response that you see in blue is divided into different categories by thresholds, which are (inaudible 00:03:54) that divide the latent and the line response into the different categories you see represented in green in the graph. The frequency of the categories depends on where we cut the latent response, and we called this latent response Y\*. And depending on the value of the underlying Y\* value of an individual, the individual will have a corresponding categorical score. So for example, if Y\* is less than the first threshold, the value of the observed categorical variable for that individual will be zero. So in this way, we can express the latent response Y\* as a function of different covariates, in a formula that is similar to a standard regression model that we are more familiar with.

 So the equation where the latent response Y\* is a function of covariates in an equation that is very similar to a standard regression can also be expanded to and be applied in order to define a growth model for category covariables. So here I’m considering latent response Y\* at time T for individual I as a linear function of the individual’s intercept, which I called B0 here. And as a function of the individual’s slope, B1, here, that represents the amount of change in the outcome by unit of time. We also have an epsilon term here that represents variability in the individual’s score at time T. But this epsilon TI term has a fixed mean and a variance, so I will not consider it. So individual I has their own intercept B0 and slope B1, as you can see here. But how are these terms defined? The individual intercept as the sum of the sample’s average intercept G0, plus the individuals’ variation above or below this average, U0. Similarly, the individual slope is the sum of the average, a sample average slope plus the variation of individual I above or below this average slope, so here represented by U1.

[00:06:42]

 So if we put all this together, we have this equation here, where the variation around the sample average intercept and the sample average slope are supposed to be normally distributed with mean zero and variance (inaudible 00:06:42), which is the covariance matrix of the two individual use.

 So the two terms that I have highlighted here in blue represent the sample average intercept plus the individual’s I variation above or below the sample average, whereas the other two terms here highlighted in blue represent the sample average slope and the individual’s variation above or below the slope. When we are defining a growth model for a categorical variable, the scale of the underlying latent variable Y\* is arbitrary, so to obtain a point where we can anchor the distribution, we fix the intercept to zero. So the question to estimate Y\* at time T for the individual is even simpler because the average intercept is fixed to zero. And it’s important here to mention a key assumption of this model and that’s the assumption of proportional odds. This says that the facts of covariance on the odds of an individual being one category is the same for all the categories, which allows us to have a single coefficient to represent the association between covariant and ordinal outcome.

 Another assumption when we apply this model to growth modelling assumes that the thresholds that cut the underlying linked variable are constant across time. So the cut points that define the categories of response in the categorical variable we observe are the same across time or across age. So in this graph, I have plotted the thresholds of a fictional variable A – antisocial behaviour, and these thresholds are the same across time, as you can see. So if someone scores zero, for example, here, in the latent variable Y\*, they will be in category two of the categorical ordinal variable we observe.

 So we centre time at time point one, so that the different time points are zero, one and so on. And if we estimated that the sample average slope G1 is equal to point 18, using the equation here, we can plot the sample average trajectory by multiplying the slope coefficient by the centre time variable. So that the sample average trajectory will be zero at time zero, at point 15 at time one, point 30 at time two and so on. So using the formula here, the equation here, we can calculate the average trajectory of individuals in this sample once we know this parameter G1, which is the average slope.

 So the sample average trajectory will look something like this, you see it here represented in black. We can see that on average, individuals move from category two of the variable to category three at the last point, the last measurement occasion of the study.

 Using the same equation, we can also estimate each individual’s variation around the sample average intercept and slope, and if the variation around the intercept for individual I is point 0.50, for example, here, and the variation around the slope is point 14, we can then use these parameters to estimate the individual I’s values of Y\* at each time point as I show here. So using the equation and using these parameters here, I can calculate that the value of Y\* for individual I at time point zero will be point 0.50 and so on. And here in the graph, you see that I’ve plotted the individual I trajectory in red. And we can see that individual I moves from being in category two of the ordinal variable observed to being in category four by the end of the study.

 The same model can be represented in this way where we have the latent variables that represent the intercept and the slope and the change, and that explains the change in the observed categorical variable. So the intercept, which is zero as represented here, and the sample average slope G1 is conceptualised as a latent variable that influenced the observed categorical outcomes. The variance U0 and U1 round this sample’s average parameters represent individual variation above or below this average. And as described earlier, there are supposed to be distributed normally with variance representing the equal variance matrix.

[00:12:45]

 If we wanted to define a trajectory that is not linear but can accelerate or decelerate over time, we can…we should consider further slow here, G2, which is multiplied by the square of time to define patterns of acceleration and deceleration. So a similar model could create a trajectory as you can see represented here. It is also possible to model individual variation around this quadratic term or this quadratic slope so that each individual will have different quadratic trajectories. In the exercises I provided with this material, there is an example of quadratic growth that you can use.

 So we can extent the growth model I presented to model different trajectory groups, for example, around the growth mixture model. In this case, I also assume that there is a latent class variable, a further latent variable that is categorical. And latent class affiliation affects the average intercept and slope of the individuals within one class, within each class. Furthermore, within each class, individuals vary around these class-specific trajectories, which means that within each class, individuals still show their own trajectories, and to represent this model, we index the parameters of the equation I presented before, and you see here, by latent class indicator, okay.

 When we are modelling growth mixture analysis, we assume that the thresholds, the category underlined latent response Y\* are invariant across time but also invariant across classes, so each class will have the same thresholds, and therefore, again, we can represent for each threshold in a graph like this, where the thresholds remain constant across time and will be the same for each latent class that we estimate. Because the scale of the underlying Y\* variable is arbitrary, we fix the average intercept of one of the classes to be zero, leaving the intraclass average intercept of other classes to be estimated freely. In N+, for example, the intercept of the last latent class estimated is fixed to, say, zero by default. Once we estimated the intercept and the slope parameters for different classes, we can calculate the Y\* scores for the average sample in class one at time zero, one and so on, as well as the values of Y\* for those that are in latent class two, and in here, you can see that even the parameters, although they’re not specific to each class, I can calculate the Y\* values for an average individual in latent class one, and even the parameters for latent class two, I can calculate the average values of Y\* for the individual in latent class two.

 And once I plot these trajectories, they will look something like this, where we can see that whereas in the average individual in latent class one, in latent class two start at similar levels, those in latent class one end up to be more likely to be in category three of the variable A, whereas those in latent class two follow a different trajectory. However, there is individual variation within classes when we apply growth mixture models, so we can estimate individual variation and calculate the scores for an individual in latent class two, for example, and plot them as I did here. So once we know the parameters that pertain to the individual, we can calculate the expected trajectory of an individual in latent class two, and as you can see, even within a latent class here, individuals have their own expected trajectories that follow a general class-specific pattern, but with some variation.

 And it’s important also to remember that these are expected trajectories for average individuals in one class and for the individual within one class, where the actual trajectory of the individual will be different to what we expect due to measurement error. Measurement error of variants, it is important to consider the size of residuals when we compare different models.

 Latent class growth analysis model is less complex because we assume no individual variation around the intraclass average growth parameters, so individuals within a class are supposed to follow the same trajectory, so those are the parameters that we had in the growth mixture models are fixed to zero.

 When we apply latent class growth analysis, we also are working with the same assumptions as growth mixture models, so we assume that the thresholds that cut beyond the line response Y\* are the same across time and across latent classes. And we can, by estimating the parameters that are class specific, so the intercept and the slope of each latent class, then we can plot the expected trajectory of individuals within each class. But because we are not assuming various variation across individuals within each class, then we are…each individual is supposed to follow the same trajectory of that class, and the only differences are due to measurement error, so there is no individual-specific trajectory in the latent class growth analysis analysing growth mixture models.

[00:19:27]

 So in this presentation, I’ve demonstrated how we can apply growth models to categorical outcomes, talking about the logged link function and latent response formulation, and I’ve also expanded this to show how growth parameters are then indexed by latent classes in growth mixture models and latent class growth analysis. I’ve emphasized again, that the key difference between the two models is that growth mixture models allow variation around the average intraclass trajectories, whereas latent class growth analysis assumes that only individuals within that class will follow the same trajectory, save for measurement error. So thank you very much for your attention, and in the exercises I provided with this resource, there are also examples of different growth mixtures models and latent class growth analysis applied to real data. So if you’re interested, look at that. Thank you very much, bye.

National Centre for Research Methods (NCRM)
Social Sciences
Murray Building (Bldg 58)
University of Southampton
Southampton SO17 1BJ
United Kingdom

**Web** www.ncrm.ac.uk
**Email** info@ncrm.ac.uk
**Tel** +44 23 8059 4539
**Twitter** @NCRMUK