# Latent Variable Models for Social Research – video 2

## Transcript

Full resource, see: <https://www.ncrm.ac.uk/resources/online/all/?id=20835>

Chris Playford: Hello. In this video I want to show how you might estimate and interpret the output for a latent class model using Stata. I will first describe the rationale for the model and the variables included.

In this example I’m working with a latent class model. These models are suitable when seeking to estimate a latent variable model where the manifest or observed indicator variables are categorical and when the analyst chooses to treat the latent variable as categorical also. In other latent variable models you might instead with to treat the latent variable as continuous, and there are other latent variable models which do this. In this model there are five manifest or observed variables relating to outcomes in school subjects. These are based on five subject groupings, English, mathematics, science, humanities and other subjects. Each variable is coded as a binary variable, 1 or 0, where 1 indicates a pass grades A\* to C, and 0 otherwise, based on General Certificate of Secondary Education qualifications taken by school pupils in 1992.

The model estimated using the Stata gsem command is based on a logistic regression model framework. The first part of the code specifies the five manifest variables, the binary GCSE subject outcomes. The lclass option then specifies how many classes are to be estimated. This is up to the analyst. In this example there are four latent classes to be estimated. As the logit(?) scale can be a little difficult to interpret, Stata provides further commands to re-express the model parameters as probabilities. To look at the latent class probabilities or prior probabilities, we run the estat lcprob command. We can see that four latent classes have been estimated. The margin column reports the probability that a randomly selected individual belongs to the latent class. We can see the probabilities of class membership are 0.23 or as a percentage 23% for class 1, 0.22 or 22% for class 2, 0.01 or 10% for class 3, 0.46 or 46% for class 4.

Now let’s look at the conditional probabilities for the manifest or observed variables. We do this using the estat lcmean command. These are the probabilities of gaining an A\* to C pass in each of the GCSE subjects given membership of a particular latent class. I have re-expressed these probabilities as percentages. Class 1 is characterised by the lowest levels of overall attainment, i.e. the probability of gaining an A\* to C in GCSE English is 16%, 5% in GCSE maths, 5% in GCSE science and so on. Class 2 have higher levels of attainment in GCSE English, probability of an A\* to C is 81%, or GCSE other subjects 70%, but much lower probabilities in gaining an A\* to C pass in GCSE maths 31% or GCSE science 14%. Class 3 demonstrates an inverse pattern to the conditional probability reported by members of class 2. Members of class 3 have higher probabilities of gaining an A\* to C in GCSE maths 55%, or GCSE science 76%, but lower probabilities of gaining an A\* to C in GCSE English 48%, or GCSE other subjects 61%. Class 4 has the highest levels of attainment, with the probability of gaining A\* to C passes in each of the five GCSE subject groups being greater than 90%. If you want to calculate overall goodness of fit, you can also use the estat lcgof command.

Thank you for watching this video. I hope you found this helpful when interpreting latent class model output using Stata.

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