Cross – Classified Models Part 1 – Introduction

Hi Good morning, my name is Professor William Browne and I'm one of the co directors of the Center for multi level modeling that's based at the University of Bristol. We are one of the partner groups with the National Center for Research Methods of the ESRC and as a result, we run training courses, but we are also creating training materials online training materials and online training resources that you could use.

Our speciality is multilevel modelling, which is a form of advanced statistical models. Modelling which is used in the social sciences and, more generally, and along with my colleague Professor George Leckie I'm going to talk to you today about an extension of multi level modelling called cross classified models.

This is the introduction to this topic we're gonna have 3 lecture style videos and then some walk through practical's, so if I carry on through the slides what am I going to talk about in this first lecture well the topic is cross classified models and so I'd like to give you some motivation about what are cross classified models? when would you use them? And what applications might they be useful for? To do that, I want to recap, a little bit on on standard multi-level or nested models so to give you an idea of where we're coming from, a little bit of history, how these models developed.

Before we move on to the example which we will use in the practical materials I just really talk a little bit about the application that practical is based around which is a data set from the field of education that's from a place called Fife in Scotland. That I'll really just introduce that data set and then that will be the end of this introductory lecture and in lecture 2 my colleague, George Leckie will take over, and he will fit models to that Fife dataset. I'll then come back for lecture 3.

Now one feature of NCRM is that we are based in the social sciences, but we reach out to other disciplines and that is certainly true in multi level modelling so I'm going to look at a couple of examples, a couple of extensions, if you like to different application areas and actually expand the modelling so we'll have more levels or more classifications that are crossed in these extended applications so that will be lecture 3 and then, as I say, there will be practical materials, we will start off by running a practical using MLwiN and we may add as time goes on to those practical materials for different software packages.

Okay, so let's start with a bit of background, a bit of revision, probably for some of you that there are certainly materials on the NCRM repository on what we call multi level modelling. Ian Brunton Smith has given a whole series of lectures on that, but I'm just going to recap, a little bit So what are multi level models? well really all of our statistical models that these multi level family of models. What they're designed to do is account for the underlying structure in the dataset. They build on something called linear models and types of linear models are things like regressions and analysis of variance and these models tend to assume independence of the observations once you've put in the predictor variables we assume that the observations are independent.

So originally these multilevel models were developed for nested structures and an example from education would be where we collected data on education we generally wanted to collect say responses, maybe exam scores on pupils and pupils they don't hang around in fields. They go to schools and so pupils are nested within the schools and so two pupils in the same school, we would expect to have the same education experience and, therefore, to be more similar to each other than two randomly chosen pupils.

Okay, so one way of thinking of this is that, therefore, what we need is an extension to our linear modelling which will include some what we call random effects to account for the fact that the pupils are nested within schools.

Because the linear models assumes independence, what would happen if you fitted a linear model to data, where there is this two level structure is you'll be too confident, because you won't have control for the clustering, you would assume everything is fine, everything is independent and your confidence in your results will grow and you'd be overconfident so we can write out - we're going to have a few equations here - we can write out a 2 level model, let me see if I can get my coloured pencil up here so here is the model with a box around it. And, we would normally have some sort of response variable of interest, which we will usually give the letter y, and we will have 2 indices here to indicate that that's an observation And that's y_ij js the response for the ith pupil in this education example, so i is pupils. In the jth school.So j is schools and then we have this is almost like an extension of what we would call the regression models, so we have an intercept beta zero.

Some predictor of interest x_ij so here x_ij, is the predictor variable for the ith pupil in the jth school again and the coefficient beta one that goes with a slope, so this could be, for example an exam scored at 16 (y_ij could be) and x_ij could be some intake variable maybe an exam score at age 11 and then we could see what progress was made for schools.

And down here, this is the random part the model we've got a u_j and an e_ij so we've got what we call a u_j here, this is the school random effect only has a j subscript and e_ij is the residual the pupil random effect if you like. And to make this a multi level model we give distributional assumptions to both u and e, so we say that school random effects have some variance that we're going to estimate and the pupil effects have some variance that we're going to estimate, and if we didn't have this u_j so if this wasn't in the model this bit here then we have a simple regression model so by adding in those u_js we've created this multi level model which allows for or controls for the clustering In the data now, as I say, pupils in schools as a typical example, but I mean i could be cows and j could be herds, I used to work in a vet school and we definitely had data on. Things like milk records - somatic cell counts for cows, so in that scenario you'd have cows nested within herds. But the important thing here is irrespective of the application, the model and the methods that we use to fit it are exactly the same, so let me clear these drawings, I might do some more drawings, as we go and let's move on to the next slide.

Okay, so it might be easy to then, therefore, think about to think about how we might diagrammatically represent what's going on with our data. So in that last example we had pupils within schools, we have what we would call a 2 level model, the pupils were level one the schools were Level two and here we see another two level model, this one from geography you might be collecting information on individual people, but we also might know how they live. And how they live might I mean where they live, might influence their responses, they might be clustering due to geography.

So here we have these 3 people here one, two and three in the first area. These 2 people in the second area and these 2 people in the 3rd area, and we can draw this kind of. These tree like diagrams and we can say, well, can we do more than that can we have more structures, then, than this simple 2 level structure. Now it may turn out that actually we've got more geography than we thought we may know the actual households of these people, these people and so we might introduce a 3 level structure, and here we are introducing a level household in between person and place so these 3 people are actually, the first household and we have a couple of people who are in the same area, area one that belong to a different household, household two. And now we've got the same first Household that we saw before wanted 1,2 here, and that household, and here we have 4 individuals (should be household 2 on the diagram). And then they are in neighbourhood 2 so that's a natural structuring, that's how data might come across, you know we've all recently had to fill in the census, and that census data would have information on each individual household and lots and lots of information, about people in each individual household and, of course, households can naturally just by getting a map and drawing area boundaries can be put into various levels of geography and we would expect actually some dependency and I would expect people from the same household to be dependent, households in the same area to be dependent, you know here they're sharing lots of things they are sharing living space in place, they might be sharing levels of deprivation, you know things like that okay so let's clear the drawings so.

So we really only so far touched on nested models and that's not the topic of this talk what we're interested in are cross classified data structures so here is an interesting data structure. This lines and dots kind of data structure that we have here. So here we have here we have 12 individuals, going from individual one on the left to individual 12 on the right and this is our own level which we've collected we've collected data on individuals, and if we think of that example that we just looked at, maybe we know which neighbourhood those individuals belong to so again we've got these three individuals in neighbourhood one, these 2 in neighbourhood 2 these 3 in neighbourhood 3 these 4 in neighbourhood 4. But maybe actually these individuals are school children. So school children, just like any other individuals live in particular neighbourhoods, but they also go to school, so we are back in that scenario that we looked at in our first motivational example. And maybe these 12 individuals actually go to 3 different schools, school one, school two and school three.

But let's have a look at this which individuals are going to which schools. So for school one we have individuals one, three and five Okay, so you can see here let's see the nesting diagram there it's 3 lines joining school 1 to these 3 individuals and then individuals 2,4,6, 7 and 10 they're all going to school 2 so again we've got this nested diagram of the lines, then individuals eight 9, 11 and 12 they're all going to school 3

So one thing you can look at when you see this diagram is, if you look at the neighbourhood lines, the top half of this diagram that's a nice little nested diagram you know you've got you've captured 3 individuals here and nice straight lines come to neighbourhood one, neighbourhood two etc. when we come to the bottom, what we're seeing is the lines crossing. Okay, so you can see, this line is crossing with this line, and this line and it starts to get a bit messy and that's because what we can see is actually the children living in the same neighbourhoods actually attend different schools and flipping that over schools draw students from several different neighbourhoods. So we can't see like in the last example with households, they were definitely nested within neighbourhoods, schools aren't nested within neighbourhoods, they are crossed with each other and that's why on the right hand side here if we're going to use this level terminology these effectively are both Level two Individuals are definitely nested in schools individuals are definitely nested in neighbourhoods but schools aren't nested within neighbourhoods and neighbourhoods not nested in schools.

So this is this cross classified data structure a messy diagram made even more messy by my pen drawing all over it, which I'll clear now but it's a real example, a real application and we might say, well you know we've seen that from the first few slides that actually data points that are in the same neighbourhood they could be clustered, they could share things in common, they sharing all those resources in this application, etc. Data points in the same school they are basically pupils in the same school they share things in common and have the same teachers, you know the same teaching experience maybe.

So we need to control for those two different sources of clustering and they are not the only examples, the example which we look at actually in the practical is known as an education example it's not quite the same as this it looks at education over time.

So what we have here is students they're actually secondary school students and they do an exam at the end of secondary school, but we have the history of education, so we have which primary school they attend and which secondary school. And what we find is not every people from the same primary school goes to the same secondary school, so there was crossing going on between primary schools and secondary schools. And we can also in a similar context, maybe we took records of children at different time points, maybe we have information of different exams one every year and in this scenario where those exam marks will be nested within students, that they will be nested within schools, but some children will swap schools during their career so again they'll be crossing going on.

We may also have lots of education examples here, we may be having schools being divided on students. This is a bit more like something like 'strictly come dancing' or something where we have raters, raters giving those scores so you're being examined by a rater, it could be a marker actually if you think about it and so different raters mark different papers and different students take different papers so there'd be a crossing going on, so there are lots of examples, and these are just 3, we will see in part three of these lectures that you have examples in different disciplines as well.

So a bit of history. Where did the developments for these models come from ? well many of the nested examples that we've seen so far you know pupils within schools pupils within classrooms they've come from the field of education so as a result lots of researchers statisticians working in education, developed software packages. MLwiN, which is the one which we will use here developed by our Center back in the 1990s HLM was developed in the US and it's still going as well, VARCL was another package in the 90s, which I think now is pretty much, it run its course for a few years and they were all developed with education examples in mind, because of the natural hierarchy students classes schools. Cross-classified models have a different history and actually they appear more, the papers that developed some of the algorithms for these methods appeared more earlier in other fields - things like animal breeding, because if you think about it, when you do an animal breeding experiment you're taking one male and one female animal breeding them and seeing what happens. So you've got natural crosses of the males and the females. And one of the examples that I look at in lecture 3, it's not animal breeding its natural data, but it is, it is animals that are breeding so it'll be birds in this case.

But just a note here so Robin Thompson and colleagues did a lot of pioneering work using restricting maximum likelihood methods and looking at these sorts of cross-classified models. Olay So how do you estimate these models we're going to come on to fitting some cross classified models later and they're more challenging to fit than nested models so we're going to use Markov chain Monte Carlo methods simulation based methods in the lecture and the practicals we see, mainly because they've been implemented in the MlwiN software that we are going to look at, but you can also see that there are packages in R that will do it in Stata and things like winBUGS and they are computationally efficient, they can handle big data sets and they can do complex multiple membership models as well, which we will look at in later lectures so they're useful. Maximum likelihood estimators well.

You will find maximum likelihood estimators for multi level models in most software packages, these days, but usually they've been designed to fit hierarchical models so what they do is they use what we call sparse matrix techniques, so they use the fact that the variance matrix in these models is sparse it's block diagonal in some sense to make it easier to fit the models. One exception being the R package so LME4 is very good at fitting cross classified models, it has no problems with them.

And there was an elegant paper, so the paper by Rasbash and Goldstein 1994 shows clever solutions to recast cross classified models, as some sort of constrained hierarchical model and this was done originally in MLwiN IGLS. it's also available, I believe in the Stata mixed package as well. And these work particularly well in examples like the one we see in the practical where where things are almost nested and there is just you know primary school and secondary school, in our example they're not nested, but there are almost like feeder primary Schools so lots of pupils do indeed go from the same primary school to a particular secondary school, but there is some element of cross classification and it works to a certain degree but it's cumbersome if you've got big data sets or, if you have multiple membership models that will deal with later. So that's a little note on estimation.

Let's talk about the example that George will take you through most of the model fitting of, and this is an example from Fife, which is an area of Scotland in the North of the UK or course if you're from elsewhere and it's been used in many places. Rasbash and Goldstein back in 1994 used it originally when they were showing off their extensions of the IGLS algorithm I just talked about for cross classified models.

So it's quite it's a reasonable size data set we've got 3435 data points and these are students and, as I said earlier, this is exam scores at the end of secondary education so in the UK, what we have is we have primary and secondary education and then tertiary education and university. So primary education is from the ages of about four to 11, secondary education from 11 to 16 and then there's this this area between 16 and 18 which which is sometimes called six form. It fits in tertiary education in some settings is the pre university years, then university from 18 onwards.

So what we have is, we have a score at age 16, and although we have a lot of pupils they come from pretty much all of the secondary schools in Fife. In this data set and 19 secondaries that's a small number of secondary schools 148 primary schools that's typical in the UK education system where primary schools are usually closer to where people live and maybe have one or two classes and may only have 30 students in a year in a primary school. Secondary schools, then are the bigger schools and they take students from a wider area so you have often primary schools that are feeders to secondary schools so that's the nature of the data set.

The response is actually an attainment score, total attainment score and it ranges from one to 10. Okay, so we're going to do some, we're going to assume these scores are normally distributed or at least the residuals in the models, we fit are normally distributed, but it's worth noting that, of course, this is, this is actually an integer score so everyone gets a score from one to 10 on their attainment age 16. And there were lots of student level predictors - things that might influence the score so for each of these pupils we know how well their mother and their father did in education, we do have a score an earlier score, the verbal reasoning score, which I think was taken at age 11.

We have this variable called choice of school so at the time, this is quite old data, as you can imagine, given it was analyzed in 1994. At the time, in Scotland you were allowed to fill in a form which said which school you'd want to go to, which secondary, and then depending on spaces you'd be allocated to a secondary so this choice of school will be one if you got your first choice 2 etc, we have some measure of social class and we have the gender of the child and these can then be used to explain what we see in terms of variation in attainment.

So let's look at what the data looks like so what we've done here, this is just actually a randomly sampled eight students from 2 of the secondary schools, schools one and two. So what we see is I've sorted this data by secondary school just this small sub sample so we've got these first four pupils in secondary school one this next 4 in secondary school 2 and primary school varies as we go down so these first 2 are in primary 1 this next 2 in primary 5. These two primary 12 and these last 2 are in primary 14 so actually these data points form a nested data structure with primary nested in secondary, but what you'll find if I put the whole dataset there is that your see some of these number ones here appearing lower down for different secondaries so it isn't really nested data but to a large degree, it is close to nested data so some of these primaries will feed the secondaries.

We've got the pupil number here this is just the numbers that we've given to each pupil, we've got their scores this first pupil scored nine, the next one six. Their verbal reasoning score for age 11, how well their mother did in education, so this is just a binary whether they can pass a particular level so one would be higher education and then the choice of secondary school so seven of these eight children got their first choice this child here got their second choice of secondary schools - they didn't really want to go to that secondary school but they were allocated to it.

So a little bit of a point here. In a cross classified data structure, what we really need is we need unique identifiers for classifications okay. So when it says primary 1 here, that is the first primary school. Now, when you have nested repeated levels of nested data you might have the first primary school in secondary school 1 and the first primary school in secondary school 2 and they might be different schools. Here because we're treating cross classified models, these have to be unique. So that gives us some motivation for the data we've got a few predictors here, and you can see the format we require for the data set.

Let's just do some preliminary looking at the data set so what I've done here is I've worked out the average score in attainment for each pupil - sorry - for each higher level so on the left hand side here we have the secondary schools and on the right hand side, the primary schools so we're got our 19 secondary schools and we can see that the average attainment if I draw these lines. Just about lies between about seven and below there about four so there is a little bit of spread. This school here seems to be a bit of an outlier it is a bit low, this school here seems to be the best of the bunch when we compare that with the primary school, bearing in mind that the secondary school there are lots of pupils per secondary school.

In primary school they were probably less, there is much more variability. So we've got one here where everyone got 10 and down here we've got a few which are down to 2 so we see that variability going on, so that's quite interesting so when we actually fit models, we might expect. The variance of the primary school to be bigger than the variance for secondary school and we might think that's the more important predictor - where you went to primary school might be a more important, predictor than where you went to secondary school in defining how well you do.

And one more thing to look at with the data let's do a quick cross classification of a cross tabulation sorry of primary school by secondary school, so what i've done here at the top, here we can see 1,2,3, 4 5, 10 primary schools 10 that I have chosen and four of the secondary schools now this is part of a much bigger matrix, much bigger table. I just put this bit to give you some idea of what's going on.

So we can see for these four schools, at least primary school 1 - 45 pupils and they all go to secondary school 9. Primary School 2 well all seven of them are going to school 7. primary school 4 here we have a bit of a difference so most of the pupils go to school 9, but one of them goes to school six and we can see similar things if we come across. These are all got one secondary one primary Okay, then we've got one pupil who is going to number eight, when the rest of his peers are going to school six and here's one with more split okay 15 go to 6, 1 to 7 and 4 go to 8.

So, what you have got to realize is some of this will be geography so some primary schools will be close to the boundary between two secondary schools and some of it will be choice, so we can see that secondaries are drawing their students from many different primaries and primaries are sending their children to different secondary schools, we have examples of that though many of them are what I would call feeder schools where the bulk of those children go to one secondary so we're not far away from nested models in actual fact we call this data sparse, because you can see that lots of the cells here are empty. Now this is very different from some cross classified examples that you might see where that's not the case.

So how cross-classified in practice, are these data? Well, so if we looked at them from a secondary school perspective than one secondary school, we could find has only seven primary schools that are giving it all of its pupils where as the one which has the most variety has 32 of the primary schools for it's pupils. But never 148, 32 of them is the most being sent into one secondary so of the 148 what we find is that again, they don't send them up to all 19 different secondary schools, they actually send to between one and six.

And in actual fact 57, so over a third of the primary schools, are simply a subset of pupil that go to the secondary, so all those pupils go to the same secondary. Another 50 of them, so we are now well over half are split between 2, 26 between 3 and then we have got the tail there are 15 primaries there that go to four or five or six. So we do have this feeder school idea so really what we're looking at this system really is a bit of a feeder schooling system so there's this natural idea of primary sending on the secondaries and so all the same factors of geography and underlying deprivation, etc, will persist in both primary and secondary rather than, say, a school choice system where kids are moving lots and lots and in actual fact if I was to move 8.4% of students right take them pick them up and put them into the main secondary school for their primary school I could create a strict hierarchy so it's cross classified data but it's not hugely cross-classified

Okay, so that's introduced the data set, and I think that's a natural point to stop, and what I will do is is I will hand over to George Leckie in the second example second lecture but now that you've got an idea of this Fife in Scotland data set, you can see how we might fit some models to maybe explain the variation we see in education, so what I’ve covered in this first lecture is I’ve done some recap. So i've recapped the ideas of multilevel models in general and then shown just by showing you these line and dot diagrams shown that not all data structures have this hierarchical or nested data structure i've given you a little bit of historical background, a little bit about estimation of these models and then we've begun to introduce an example from educational research and this example has a cross classified structure.

As I say, George will now take over and in the second lecture we're going to fit some models to this structure and then we're finished with that example and I’ll return the third lecture analogy do some extensions and additional applications, so I hope that was a straightforward introduction, thank you very much for listening, I look forward to speaking to the third lecture.