Cross-Classified Models, Part 2- Model Fitting

Welcome everybody to cross-classified models, part two model fitting, my name is Professor George Leckie.

So I'm taking over from Bill, who did part one, and in that first lecture he introduced cross-classified models and the practical example, which was the Fife dataset the Scottish data on students in secondary schools and their primary school history that they've been to and how they performed in their end of school exams.

So we're going to start putting some models to these data, but before we do that we need to show you some model equations and so I'm going to introduce some notation and classification diagrams and so on. We will then work our way through a series of cross-classified models very basic just to get a flavour of these models.

We will show you some of the consequences of ignoring cross-classification and I look at ways of describing the kind of complex residual clustering there is in these models these so called Variance Partition Coefficients (VPCs) and Intraclass Correlation Coefficients (ICCs).

We will add explanatory variables to the model and see how we can assess what's going on in different parts of the model and that will lead us in nicely to the follow on video, which will be back to Bill where he'll look at some other applications beyond education of these cross-classified models so that's our battle plan let's go.

So recap of these data. The data are from Fife in Scotland, it was using a very early paper on cross-classified models which actually used maximum likelihood estimation. We are going to be using MCMC Markov chain Monte Carlo methods to fit models. In this lecture as they're much more capable as data get bigger or as models get more complex and so it's a fairly standard approach for fitting these models.

So let me get my pen out what we got here, we got some 3400 students there nested within just 19 secondary schools. So not a huge number of secondary schools. Primary schools are much smaller and these kids had previously attended some 150 primary schools.

The outcome is the end of secondary schooling exam scores. So the the kids are aged 16. It's a total exam score across different subjects and it ranges from 1 to 10, 10 being the highest performance. We've got lots of explanatory variables to play with. We've got things like dummy variable flags for whether mothers and fathers stayed in education beyond compulsory.

Schooling themselves. We've got some earlier test score on verbal reasoning for our students when they enter their secondary schools at age 11. We've got some kind of measure of whether the secondary school which they're attending is their first, second, third, fourth choice okay as well.

We have a measure of the social class of the family and we've got a flag for whether the students are boys or girls, so all of these could explain, of course, why some kids score higher in the end of school exams and others, but they can also explain why some secondary schools appear to be performing a higher than others and also why some primary schools might appear to be higher scoring than others. So we're interested in putting those variables into the model to get more understanding.

So let's look at the very simplest cross-classified model that we could fit the data.It's very simple because it's going to include no explanatory variables initially. So what's the point, the point is we're trying to quantify the degree of secondary school clustering and simultaneously the degree of primary school clustering in our outcome variable, the attainment score for child i.

So the model has only really got an intercept in it to estimate the overall average performance at age 16, but we have these school random effects primary school random effects and student residuals, and so we got this breakdown of their ability and these random effects and residuals we have standard multilevel assumptions that the school effects are normally distributed the primary school effects are normally distributed and the student residuals are normally distributed. And, in each case they're normally distributed around means of zero. These are relative effects and we estimate.

Crucially, the variance of the random effects at each level. Okay, so we have the school variance, the secondary school variance, the primary school variance and the student variance, and so what we're really doing here is we're estimating an overall mean beta zero for your attainment at age 16. But we're doing a decomposition of the variance in attainment scores into separate components of variance associated with different levels of analysis, the secondary school level of analysis.

The primary school level of analysis and what's left over, and what happens within schools, essentially, and we want to know where does the variability lie that's our starting point, always for doing multilevel models.

Now if you're watching this video then you'll know something about two-level models already because that's the kind of prerequisite topic and even three-level models and so you'll be familiar with ijk notation. Now you'll notice here in the equations i'm showing you I'm not showing you the attainment score for child i in primary school j.

In secondary school k because this ijk notation is what we call hierarchical notation. It implies a hierarchy that. Students are nested within primary schools and primary schools are nested within secondary schools. Now we know that's not the case from Bill's, part one video because we saw that each secondary school draws its students from many primary schools. But each primary school simultaneously sends its kids to multiple secondary schools breaking down the naive three-level hierarchy. So that's why we're doing cross-classified models, but because we can't do a hierarchical model we can't use the usual hierarchical notation either and so, rather than ijk notation we're going to use something called classification notation which is what is being used on this slide here and under the equations and that's why things look a little bit different but i'll explain the notation in the next slide. Essentially both primary schools and secondary schools are conceptually at level two in the hierarchy.

So in this classification notation the only index we've gotten a sentence i, which is going to uniquely identify all 3,400 kids in the data okay. And then we have these random effects. The u's in the model, and we have these kind of "2" and "3" superscripts on the random effects and actually you can see them down here as well. The "3" superscripts and the "2" superscripts. Now, this is just to distinguish between different sets of random effects, different higher level classifications.

In our case, the the "3" denotes the secondary school random effect and the "2" denotes the primary school random effect. Now the subscripts on those random effects, particularly if you look at the previous slide is sid and pid. This stands for secondary school ID and primary school ID, and these are what we call classification functions and essentially. In parentheses we have i the index for a student. Now if you you plug in the unique ID for a given student what this function, does is it spits out the secondary school that this kid attends or the primary school that this kid attends if we are using the pid function.

So if, for example, student 1 in our dataset attends secondary school 5 and previously attended primary school 17 then the way these functions would work is if we plugged in a value of 1 for the student identifier the function, the classification function sid of I would spit out 5 telling us that that kid attended secondary school 5. Likewise the pid of student 1 is the primary school student 1 which is primary school 17 and all of this will then kind of feed into the model equation, and so the model equation for student 1 is the attainment score for student 1 is equal to the overall average score across all 3,000 kids.

plus a secondary school random effect and sid of 1 is 5 or so and we now we just have a 5 as the subscript telling us that's the school secondary school random effect for school 5, whereas they attended previously primary school 17, and so this u sub 17 superscript 2 is the primary school effect of primary school 17 and then that's not going to perfectly predict the student's age 16 score. We're going to be left over with the student residual for that particular student and that's the e_i. So that's just how the notation works. It's still a kind of multilevel model with different sets of random effects, but because we don't have a strict hierarchy.

We have to use a different notation and this notation is quite popular in the statistical literature and if you are reading journal articles you'll certainly see it and so that's why we've gone through it in a bit of detail here.

Again, so classification notation is very flexible, and if you had not just a two-way cross-classified model, but a three-way cross-classified model because maybe you're bringing in the neighborhood of residence in there as well, and or maybe even bringing in school districts and regions, you can have all this kind of complexity brought in as so this notation is very scalable which is great.

A limitation is that the notation itself, what we saw on the previous slide with the equations, doesn't really give you a sense of what the data structure looks like. And so we like to compliment the equations on the previous slide with a so-called classification diagram. For this model it is very simple and it's done like this.

The lowest level is student and so the outcome is always measured at the lowest level of the model and students' scores are nested within secondary schools and they are separately nested within primary schools. But the fact that there's no arrow between primary and secondary school means there's no there's no kind of strict nesting of one within the other, they are simply cross-classified.

And so, this this kind of classification diagram communicates the structure we can tag on additional boxes, which are often called nodes and additional arrows to indicate what's going on. You might want to compare that back to the unit diagrams that Bill showed you in Part One so often researchers will include these diagrams in their papers as well.

So let's have a little bit more understanding. Let's consider student 1. Let's pretend that I'm the proud father of student 1 who's an excellent student. And this is the equation written out for my child, child 1, they attended secondary school 5 and primary school 17. And these u's are the effects of those schools. So here is my child and and on the y axis, we have the attainment. And you could see my child is right at the very top there. That child is doing fantastically! So, how does this model work.

Well models always work in the same way in the sense that we start off at zero, an outcome of zero, and then we tag on the estimated intercept, beta_0, and so this thick black line is giving an overall average score in the data across all all 3,400 children, if you like.

Now my child's up here and they go to a particular combination of secondary and primary school, secondary school 5 and primary school 17. And this line up here is, if you like, the predicted mean score for the subset of kids like my own, who first went to primary school 17 then went to secondary school 5.

How do we get to that point. Well let's go back to the overall average line beta_0 and let's just consider kids who attend primary school 17. Okay those kids in general, were scoring above the overall average, for whatever reason, maybe a combination of selection into schools, maybe it's a relatively prosperous neighborhood but also combined with perhaps that primary school is truly effective. So that dotted line there which I've highlighted is showing you the mean performance of kids in primary school 17.

What about from the other perspective, let's just consider all the kids in secondary school 5. Okay, well they are on average performing like this. Okay, which is this much better than the overall average. Okay, so here, you can see the secondary school effect taking us up to the higher of the two dashed lines. But my kid had attended both these schools. First they attended primary school 17 and they got that effect.

Then they attended secondary school 5 and got that effect and we add those two effects together, that is how we get up to this higher solid line up here, which is giving you the predictive performance of kids in that particular combination of primary, secondary.

So these are modeled as additive effects, you can see that here. We are simply adding the primary school and secondary school effects together. There's no interaction. That is an extension that one can do with cross-classified models, but we don't discuss it here.

But, having got to the subset of kids and the combination of primary 17 and secondary 5 we see we have 1,2,3,4,5 kids orbiting around that mean performance and some with, below with negative student residuals, some above with positive student residuals, and I'm that delighted parent whose kid is somehow magically doing the best amongst all five of those kids there. And you can see the residual e_i.

So hopefully, you can see mechanically how our models working how its decomposing the variability in attainment into kind of mean performance associated with primary school attended, secondary school attended, and then residual variability attributable to a given student.

Okay let's think a little bit more about the data structure and some implications for results, we already know some of the things that we're going to see when we come to fit the models just by having an understanding of the dimensions of the data and, in particular, we have only 19 secondary schools.

So my between secondary school variance is not going to be estimated very precisely because it's kind of the variance of 19 means. Well 19 is a small dataset right. So I'm going to have a big old standard error on my between second school variance.

But on any given secondary school we have lots of students and, indeed, on average, we have 181 students in each secondary school, which means I can get a relatively precise estimate of the specific effect of, for example, secondary school 5, my kids secondary school.

So we get relatively precise estimates of those guys. But at the primary level this this kind of story flips around because here we now have lots of schools, some 150 and so my between primary school variance Sigma_u2^2, that will be estimated relatively precisely with relatively small standard error because we've got a big sample of 148.

But, in contrast, for any given primary school, let's say primary school 17, the one that my child went to. On average we've only got 23 students per primary school, which is a much smaller sample size and so we're going to get bigger confidence intervals less precise estimate of individual school effects at the primary level just because there is data. So it's quite good to kind of have a sense of your data structure, because it's kind of telling you what you're going to see and giving you understanding once you do see the results shortly.

So let's fit our first models so, as I indicated already we're fitting all models by Markov Chain Monte Carlo methods but we are doing so using kind of sensible starting values from the MLwiN software. Our priors that we're using are minimally informative and so the results are effectively the kind of same results that you really get from a maximum likelihood estimation.

But MCMC is much more scalable to bigger datasets and more complex models, so this is quite good to use. So the model which I've shown you so far is actually model 2, that cross-classified model with no explanatory variables, but with the secondary school. Random effects and variance and the primary school random effects and variance. So we estimate the variance of these effects directly. And that is what is being summarized here.

Now we actually have an even simpler model, model 1 where we do not include tThe secondary school variance or the primary school variance, because in this simple model we have deliberately excluded the secondary school random effects entirely and the primary school random effects entirely, why?

Well, the contrast of model 1 to model 2 provides us with a test of whether we have any clustering in general, because model 1 ignores the clustering, model 2 takes it into account, so difference in the results between the two models is giving evidence that there is clustering. And by difference, I mean the fit statistic so we're funding model by MCMC, and so we have the Bayesian Deviance Information Criteria (DIC). It's a bit like the AIC or BIC if you fit models by maximum likelihood and so just as for those statistics we like smaller values.

And so, as you fit more complex models, the DIC drops, and you want it to drop by at least 2 to 5 points, really, but any drop of that magnitude bigger is suggesting that the more complex model is statistically preferred.

So we've got a whopping drop of some 383 points when we account for secondary school clustering or effects and primary school clustering or effects, and that is telling us that we have meaningful variation at those higher levels and, indeed, as we see shortly if we add up these variance components, we will get an estimate of the total variance which will actually be pretty similar to the 9.363 variance we had in the first model.

And what we'll see is that those two higher levels, secondary schools and primary schools are accounting for about 16% of the overall variance in the attainment and so in a cross sectional setting such as this that's pretty meaningful and in line with the literature on school effects, and so we have clustering, it's substantial and it's statistically significant and we should proceed with a cross-classified multilevel model.

So, here we go, next slide here so model 2 is just the model we had already and we might at this point, go well okay so we have established that there is this kind of clustering in general attributable to secondary schools and primary schools. But what we should really check is whether maybe just one or other of these sources of variation is dominating. So can we get away with simpler model which only accounts for let's say the secondary school clustering and ignores the primary school clustering. Because if we could be a two-level model, and we could get away with models that we've already learned, we wouldn't have to learn these fancy new cross-classified models. So we should always justify to our audience that the full complexity of the model we are presenting is required.

So here I compare model 2, the full cross-classified model to model 3, a simpler model, which ignores the primary school effects, its sets them all to zero, and therefore the primary school variance to zero, effectively.

So what happens, though, when we do that is that the DIC, which the smaller the better, now get bigger OK, and now increases by 260 points so that's telling us that model 3.

Is a statistically worse fit of the data than model 2. We should stick with our cross-classified model, not go to the simpler two-level model. And what is really telling us is that. You know, over and above acknowledging the different secondary schools that students go to, there is still meaningful variation between primary schools.

So students who attend the same primary school are significantly more alike, more similar, more highly correlated attainment scores, than the students from different primary schools, and that is not simply because they went on to experience a common secondary school that's the standard story there.

So that's telling us that we can't shift to a simpler students within secondary schools model. Well, how about this, can we instead shift to a simpler students within primary schools model. Okay, so model 4 is going to this time knock out the secondary school effects to see whether we can get away with that. So it's a bit like setting all the secondary school effects, all the u's to zero, therefore the secondary school variance to zero, what happens?

Well, once again, the DIC of the cross-classified model then goes up by more than that two to five points. It actually goes up by some 30 points so statistically model 4 is a worse model than model 2. OK, so the more complex model is statistically preferred and so we've really justified now that we need a cross-classified approach, we have two sources of important clustering going on. Clustering in students' age 16 scores due to common secondary schools attended and clustering is students' age 16 scores because of this history of earlier primary schools attended. So, students from the same secondary are significantly more like the students from different secondaries, and that is not simply because of shared primary school experiences. There's something over and above that associated with secondary schools is the story. O

kay next up we've got Variance Partition Coefficients (VPCs) and Intraclass Correlation Coefficients (ICCs). So these are these great statistics in multilevel modeling which you get to to quantify the degree of residual clustering you have.

On the previous slide we have established, we have statistically significant clustering, but what we haven't done is additionally communicate, whether that is also a practical significance. OK, so the clustering gotta be not just statistically significant because in a large dataset everything is statistically significant, but the clustering has to also be meaningful, meaningful correlation between kids within the same context.

These are very easy to calculate so the secondary school VPC is the proportion of overall variance in attainment, the sum of the three variances which is attributable to secondary schools, so we divide the secondary school variance by the sum of all three. So we get out of proportion, which is 0.043 and we say that 4.3% of the variation in age 16 scores lies between secondary schools.

Now expressed as an intraclass correlation, what we're saying is that the correlation between two kids who share the same secondary school but attended different primary schools is 0.043, so a small positive correlation there some kind of similarity due perhaps to true genuine differences in quality of education between secondary schools, but also clearly due to selection effects into secondary schools.

At the primary level, we just do the same thing, but this time we divide the primary school variance by the summation of all three variances. And that gives us a larger 0.119. So some 11.9% of the variation in age 16 scores lies between primary schools. There is s a stronger clustering attributed to primary schools than there is secondary schools. And so the correlation between two kids in the same primary school but who went on to attend two different secondary schools is 0.119.

Now our highest VPC and ICC relates to when we're saying well how much of the variation is attributable to schools in general, both primary schools and secondary schools combined, if you like? So we put both these variance components on the numerator divide by the summation of all three on the denominator. That gives us 0.162. So 16.2% of the variation in test scores at age 16 is attributable to the two sources of clustering combined or considered simultaneously.

As a correlation, we are saying the correlation between two kids who attend, not only the same primary school, but then went on to the same secondary school is 0.162. So in the literature, these would be considered of practical importance and they're inline with other studies and so not only to we have statistically significant clustering but it's also meaningful in the literature, these magnitudes.

Right, this brings us on to our next and model, where we are going to introduce covariates. And all the covariates that I'm going to introduce are actually characteristics of the students, not characteristic of the secondary schools or primary schools, although these can be introduced and would often be introduced in the next model.

But in our data all we've got here are these student characteristics, to play with. We have a measure of their attainment at age 11 when they entered their secondary schools, verbal reasoning test, we have a measure of their social class, high values indicate kids from more affluent backgrounds, we have dummy variable indicators of whether the father and mother stayed in school themselves beyond, beyond age 16 essentially, and we have this measure of whether the secondary school attended was the first, second, third or fourth choice for the child as well. Putting those explanatory variables into the model, we're going to explain the variation in the outcome, and so we expect the residual variances to go down at all levels. Why all levels? Because kids aren't randomly assigned to schools. There's going to be kind of social class clustering in different schools.

So if we adjust for social class we're going to explain part of the variability between secondary schools. And so, if we do so, then the magnitude of the secondary school effects will, in general, reduce towards zero. Likewise the primary school variance will go down and their effects will get smaller in absolute magnitude.

And even within secondary schools and primary schools, some kids perform better than others and that's related to these characteristics and so residual variance will go down as well. So I probably expect to see positive coefficient on the verbal reasoning. So kids that are doing better at age 11 carry on doing better at age 16.

Positive relationship between social class and attainment I expect to see: kids from more affluent backgrounds go on to score better at age 16, all else equal. Likewise kids whose fathers stay in education themselves, or mother stayed in education themselves will likely score higher at age 16. The only negative coefficient is likely to be on choice.

If you don't get it, the school you're attending is not your first choice, but your second or third or fourth, it is less and less the school you wanted to go to, perhaps that might be actually negatively associated with how you go on to score.

Here we go. Here are the results, and so model 5 is the new model. It's a cross-classified model extended to include this set of covariates which have just gone through and all the coefficients are positive and statistically significant apart from the hypothesized choice variable which was negative and significant.

I can judge significance in the normal way. Divide point estimates by standard errors. Is the ratio effectively 2 or higher? And of course you can translate these into p-values or calculate confidence intervals in the normal way as well, but we can see very easily they're all significant.

Now from a multilevel perspective, as well as being interested in the regression coefficients, we're very interested in the variance components and the proportion of variance that we explain at different levels of analysis, essentially R-squareds at the secondary school level, the primary school level, and at the student level and so we can calculate things like the proportion reduction in a variance component and we see the secondary school variance drops massively, and indeed it goes down some 97%.

So, actually, that these five predictors explain away nearly all the variability between secondary schools it's quite an interesting finding.We also have very powerful explanatory power at the primary school level. That variance also dropped massively as well, drops by some 82%. And then, even within primary secondary school combinations, residual variance which is pretty large accounting for 84% of the variance in total, well that effectively halves as well on the inclusion of covariates.

So we got big explanatory power associated with these covariates, but especially at the secondary school level and at the primary school level.

So those wide educational disparities that we saw between both types of schools are kind of pretty much wiped out when we take into account these background characteristics of kids suggesting rather strangely perhaps, that the schools don't appear to have particularly differential effects from one another, once you've ruled out the kind of selection into schools, which is clearly a big feature of the Fife data.

It is very sensible to always do residual diagnostics after any statistical model, including multilevel models. Now in our cross-classified models we're particularly interested in higher level effects, and so this is a so called quantile-quantile (QQ) plot of the secondary school predicted effects.

And so we've got these predicted effects essentially on the y-axis have been standardized against so called normal scores on the x-axis and the reason we do this plot is that we have made a normality assumption about the secondary school effects and so you want these effects to lie on a kind of 45-degree line and what we see is that they're pretty good in the big scheme of things. There are only 19 of them.

We do the same at the primary school level. These again, there is some kind of curvature rather than a straight line relationship, but in the big scheme of things most researchers will be happy that that looks approximately normally distributed.

Now if you've not seen quantile-quantile plots before, just do histograms and then check the histogram of predicted school effects looks like a bell curve. Here superimpose a normal curve okay. That's fine to do that. So the plots on the left are kind of or residual diagnostics, checking model assumptions. Normality assumptions are pretty good.

Plots on the right are actually more for substantive interpretation. So having adjusted kind of quite richly for the different types of kids that go to different schools, including a measure of prior attainment, these residualised secondary school effects and residualised primary school effects lie much closer to being perhaps measures of the actual quality of education imparted by secondary schools and primary schools. And, so that's what these caterpillar plots are kind of showing. On the y-axis we've got the predicted secondary school effects here. And the predicted primary school effects down here. And, in each case we ranked schools in rank order from the least effective secondary school in this case to the most. And the least effective primary school to the most effective primary school, but we got big old, wide 95% confidence intervals around these effects. And they'll be wider for the primary schools, but note these are on different scales okay these plots and they're not directly comparable. But they'll be wider for the primary schools, because in general there's fewer kids per primary school, then there are per secondary school.

But, in terms of physical separation, because the primary school variance is much bigger than the secondary school variance and we've got many more primary schools as well, we've got some statistical separation, so for primary schools at least, there's a little kind of subset of primary schools which appear to be significantly less effective than the average given by the zero line and another subset up here, that would be judged to be significantly more effective provision of education. But, the majority can't be statistically separated out from the overall average or one another.

Secondary schools, we've only got 19. There variance is smaller and it turns out, really, none of these schools can be statistically separated from one another, or from the overall average. And, remember once we adjusted for all the covariates, that secondary school variance was pretty tiny and that's what being depicted there. One thing which you might notice is that the school down here is perhaps bit outlying, and indeed it's down here as well, it's school 19.

Okay, so these residual diagnostics do hint that one school appears to be kind of substantially different perhaps from the rest. So, there's very little variability between secondary schools and a fair bit of that is attributable perhaps just one school, school 19 differing from the rest.

So that brings us all nicely to our very last model where we enter in a dummy variable for being in school 19 because it is an outlier and so rather than just dropping the school we're actually going to build in a departure from normality, if you like, into the model.

Simultaneously, and we could do this in two steps if we want, but simultaneously we've actually dropped the secondary school random effects. If you left them in, they would get even smaller and effectively zero so that's why we're taking them out.

So we're kind of interested in whether this model is preferred to the previous one. This one's perhaps got a better characterization of secondary schools as really just been one cluster of schools, 1 to 18 performing relatively modestly, and then one other school, school 19, which is performing significantly worse from that kind of cluster.

And, so that's what we're doing here rather than allowing every school to have its unique effect. So let's just see what we get back from this final model. And, it should be noted, before I do that, but because I don't include the secondary school random affects for the other schools, this model just drops to a two-level model. Students at level one nested within primary schools at level two. And so we could fit this as a regular two-level model but let's see whether it's preferred.

So here we go last set of results. Model 5 is the previous model which included the secondary school variance. And, you can see a very small variance, very imprecisely estimated, hinting that we could probably get away with removing it. We did remove it in model 6 but instead we put in a dummy variable for that outlying school, which seemed to be far less effective than the rest and indeed in contrast to schools 1 to 18, school 19 is scoring some points 0.6 points less on average, kids in that school, and that's pretty sizable you know it's kind of pushing on to be three times the effect size of kids whose parents stayed in school, okay, or mothers there as well. So there's a pretty sizable effect when compared to some of the other covariates. Crucially, we should check that model 6 is statistically preferred to model 5, as well as being a simpler, more parsimonious model.

And it's pretty similar really because the DIC drops, okay, suggesting model 6 is preferred. But it only drops by some kind of four points which is pretty borderline statistically. Model 5 and model 6 are pretty comparable, okay.

And so, in reality, you could go with either of them really whatever kind of pleased you most in terms of the story you're trying to tell they are statistically comparable fits. So picking the one which makes more substantive sense would be how you would proceed.

So just to end up,in this lecture, part two, we've actually fitted cross-classified models to the Fife data, and to do so, we first had to introduce model equations and that required some new notation and diagrams because we couldn't use regular hierarchical notation because we don't have a hierarchy.

But we could use concepts such as measures of residual clustering, the ICC and VPC that apply in simpler models, and we've adapted those to the cross-classified settingto be able to detect which source of variance is more important with primary schools vs. secondaries. Of course we can add in predictive variables, just like we normally do to explain the variability at different levels of analysis.

And, interesting in this example just those five student characteristics pretty much wiped out all of the secondary school variability and a lot of the primary school variability and actually that kind of led us ultimately to reject the cross-classified model once the covariates were included and you could actually revert to a simpler model at that point. So that's often the case, actually, when you add explanatory variables in, the residual part of the model is going to change and you could lead to a different structure than what you thought initially.

So that's me done, I hope you found that interesting and enlightening andwhat follows after this video is the third and final video of this cross-classified series, and that will be Bill presenting some further applications of cross-classified models to some veterinary science examples and he'll introduce some exciting new model extensions to the basic cross-classified models introduced here. So thank you very much for listening and good luck with your learning.