Multilevel models for clustered dichotomous outcomes, part 2—2010 MCH EPI pre-conference training workshop

DON HEDECKER: Is that on? Should I move it up? Lower please, right, right. It's like in the class, somebody says, can you speak up because I can't hear you back here; the person in the first row says, can we change spot? Yeah, I've had that happen several times.

All right, so here are the estimates from our logistic multi-level model. And on the left-hand side I've got what's called Fix. That's the model without random effects, that's assuming that these data are not multi-levels. So I put them there for comparison purposes. It's really not a correct model for these data, per se, then I wanted to show you how this full likelihood versus quasi likelihood, what difference it makes. And so what I've got is the two-level model under full likelihood estimation here and then the three-level model under full estimation and then I did the same thing with the quasi-likelihood, the default estimation for Glimmix.

Now if you look at those estimates between the full and the quasi, you'll see they're very, very similar. Standard errors are very similar, so again, for clustered data per se, you don't necessarily need to go the full likelihood route. What you don't get though, you'll notice is on the bottom, I put the value of the deviants minus two times the log likelihood. Okay? Quasi likelihood does not give you a proper likelihood for which you could do likelihood ratios with testing. So you don't get that information out of it. So out of the full likelihood alternatively, if I want to do likelihood ratio test comparing the fixed model to the two-level model, I can compare these two numbers and it's chi-square on one degree of freedom. So you can see that comparing the kits to the two-level, we get a difference there of about, oh what is that 24 or something like that, on one degree of freedom. You don't need a statistician to tell you which way the winds blowing. Right?

Bob Dillon said, "You don't need a weatherman to know which way the wind's blowing." Here you don't need me to tell you that's significant. That's highly significant. Likewise, going from the two-level to the three-level analysis, you could see like, likewise well, what Shuba showed you was continuous. It's not a huge difference really going from the two to the three-level analysis, and per se, that might not be quite at .05, might not be significant. However, I think again, from the design perspective, the fact that it was the schools that were randomized and the intervention was delivered within the classrooms; logically it seems to me that the three-level analysis is the more

correct, more proper analysis. But clearly, the school variation is not that large once you've got the classroom variation accounted for.

Now, this data sets are very useful as I mentioned for teaching purposes for this reason, look at across the board, the classroom curriculum effect. Now what I've indicated here is one asterisk is that marginally significant result, two is significant at .05, and three is highly significant at .01. The classroom curriculum comes out always significant because it's a big effect. Because you know, you could ****, you could do whatever you want, it will always be significant.

However, in the fixed effects analysis, you would find a significant positive effect of TV and a marginally significant negative effect of the interaction. Now, in any of the multi-level analyses, you don't see that. So you're conclusion can really change with or without the multi-level modeling. In the fixed effects approach, you would say, hey there is a significant positive effect of the classroom curriculum, TV, but they interact such that when classroom curriculum is combined with TV it actually brings the things down relative to when they just get the classroom curriculum.

So, the conclusion here does vary if we take this multi-level into account, then we say, hey. TV is doing nothing, there's no interaction, the only thing going on really is the positive effect of classroom curriculum. The kids that got class curriculum were more likely to have a greater probability of answering more of the questions correctly, basically.

Okay, now the other thing that we should point out is for the random effects, here the variances expressed the standard deviation. Again, for the reasons I mentioned is that if you estimate a variance, it's going to be rather smaller than the standard deviation, so computationally speaking, it's easier for the procedure to estimate .5, let's say, than .25. The square of this would be closer to zero.

Another thing that comes into respect with the random effect variances is about inference. It'll print out the estimate; it'll print out a standard error. So typically when you get an estimate in there, what do you do? You take the ratio, do a Wall test. Now I put here Wall test not for standard for standard deviations. Why did I put that comment? Am I lazy? Yes, but that's not necessarily why I put that there. Realize this, the – what's the null hypothesis that we're testing here? That the standard deviation is equal to zero, it's a bounded parameter, Right. Standard deviation can't be negative.

The Wall test has as an assumption that the sampling distribution or that statistic is going to be a standard normal distribution. That's what underlies the Wall test, that's why we an take an estimate, divided by a standard error and compare that to a standard normal based on Wall's work in the '40's, that realizes that if you estimate using maximum likelihood, and you get an estimate and a standard error, you can form that ratio and compare it to a standard normal for inference.

That doesn't work so well when you have a bounded parameter because the sampling distribution for a bounded parameter is not likely to be symmetric. Right? Because if it **** brick wall at zero. So for those reasons, the Wall statistic, the taking of the estimate to the standard error, it doesn't work so well for variances for standard deviations. What works better, and this is not exactly a closed subject matter, there are still people who are working on this topic. You now, what's the proper inference for the random effect variance?

What is better is likelihood ration test. Now, even when you do that, if you take the nominal "P" value, as Shuba mentioned yesterday, an easy adjustment to make that works pretty well is to take that "P" value a divided it by two. Now what's the logic behind that? Why divide by two? The null hypothesis is that the standard deviation is equal to zero or the variance is equal to zero, which ever way you want to state it. What's the alternative hypothesis? That if it – what was it? Not equal zero? That's not precisely correct. It's greater than zero. Right? You really one a one-sided test because it's, again, it's not – the alternative hypothesis – put it up okay. Okay, then I'll be too loud, probably. Okay. I'll talk quietly. I can't do it… I can't do it. That's good.

Yeah, the alternative hypothesis really is that the variance is greater than zero, so the logic behind that divide by two of the "P" values really because we're really interested in the one-sided test theory.

Okay, so that's about the testing of the inference and again, why strictly speaking, I haven't done the Wall Test here even though I do have the estimate and the standard deviation for these variance parameters.

Okay, so Glimmix works quite well for either two or three level. It works really well either full or quasi likelihood. The only disadvantage really with the quasi likelihood solution in this context is that we don't get deviances that can be used for likelihood ratio testing, but the estimates and their standard errors are really pretty good as you can see, you compare them to the full livelihood versions, they're really not off by much at all. Okay?

So from this, let's do a little bit of calculation and then we'll try to run this data set. We can calculate from the two-level model, the ICC and here I've got it for the random classroom analysis. But now again, I've estimated the standard deviation, so I need to square those to get variances. So this is the between classroom variance. And the ratio of the between classroom variance to the between classroom variance plus in classroom variance is about .08. Seven point seven percent of the unexplained variation is at the classroom level. Ninety-three percent of it is at the student level. Right? So classroom is counting for a good deal, but of course, the students account for a lot more. Because what are we modeling? How many answers did the kid get right? You know, classrooms not gonna fix everything. Right?

So the person obviously has the bigger effect, but the classroom effect, that's nothing to sneeze at, you now, 7.7 is pretty large.

Now another way to think of that ICC is, that's like the average correlation between any two individuals in the same classroom in terms of the outcome variable. So that's how correlated the data are within a classroom. That's how similar students are within a classroom. Okay? So the correlation is not zero. It's about .077 for any… it's the average of all pair-wise associations. You know, this kid and this kid. That's what it represents. So you can think of the ICC a couple of different ways. And one way is a proportion of variation that's at the cluster level. Another way it's the average correlation of any two kids in the same cluster.

Okay, so that's for the two-level analysis. Then go to the three-level, we've got three ICC's. The first is gotten by taking the school variance and the numerator and then the denominator, you've got the school variance, the classroom variance, the person variance, or the within classroom variance. That give an ICC here of .03. Three percent of the variation is at the school level.

In that same model, three-level model, I can also calculate the classroom ICC. Well if the kid's in the same classroom, they're also in the same school. That's why in the numerator, we've got the school variance and the classroom variance. The denominator stays the same. It's all sources of variance. And so here we get an ICC again of like what we had in the two-level model. Here it's a .078.

So what this is saying is that kids in the same school are more similar than kids in the same… I'm sorry, kids in the same classroom. So scratch what I… can you erase what I just said there. Please erase it. I don't want the statistical police will arrest me. I've made one too many errors. We're allowed five out of 100, you know, but after that then we're done for.

Okay, so this is the likeness of students in the same classroom. So the classrooms – kids in the same classrooms are more similar than kids in the same school. And that's typically what you see with ICC's. The more defuse, you know, the units, the larger the units… typically the ICC get smaller. You know, kids within the same school are more similar than say the kids in the same neighborhood or something like that. Kids in the same class are more similar than kids in the same school.

And then finally, we can also calculate… though this may not be done as often is how similar are classes within the same school. Now there what you have in a numerator is the school variance and then in the denominator you have the school and classroom variance. It's kind of like, if you took and average for each classroom, basically what's the clustering of the classroom averages within the school. That's kind of what it's like. Say, within the school, how similar are the classroom averages?

And here you can see that that's pretty high, but it's kind of indicating that, you know, kids are quite different within the schools; however, the classroom averages within a given school are pretty similar. That's what that indicated. However, it is less than .5, so what that says is the school contributes slightly less to variation than the class level, class

is contributing more. And again, a way you can think about this is the average classroom posts tobacco health in our schools moderate similar within schools. Okay?

Okay, now again, all of this stuff you can do – you estimated the model, you take the estimates and you can do those simple calculations. Now some of this with SAS you can actually use estimate statements so it can do it for you. But here I've just illustrated how you take those numbers and generate those values.

Now here is something that you might also want to do. Okay, we started this and we had estimates of what's a proportion of kids that had three or more correct in these four cells. Well we began at about 3:00, we had those numbers. So we might want to go back to those numbers. By that, I mean this. You go back to slide number, sorry, 18.

We've got the observed proportions. You might want to see how does our model map onto these observed proportions? So that's what I'll illustrate now.

So on slide 30, what I'm doing is those calculations for three different models. First are the fixed effects model, without any random effects. Okay, so here in order to get the estimate of the proportion of probability, I need to use a logistic response function. Remember, you can write the logistic regression model in terms of the probability of the response and it's that function of ****, one over one plus the exponential of minus the logent value. Okay? So I'll call that, I think that's Chi, is that Chi? The Greek letter? I didn't get the classical education when I was growing up. I learned it through statistics. I'm pretty sure that chi. Okay. So, chi, that's a logistic response function.

In the fixed effects model, minus .341 was the estimate of the intercept, namely the log odds for the control group. If I take the logistic response function and apply it to that, I get an estimate of .416. Okay? That's exactly what it was, the requirement. Then for the group that got TV, but not classroom curriculum, we have the intercept plus the TV effect. If some knows the **** applied logistic response function we get our estimate there of the probability of a little bit under a half.

Then for the group that got the, sorry, that got classroom curriculum, but not TV, these two got .632, and then for the group that got classroom curriculum and TV, it's the intercept plus CC plus TV plus the interaction because they're one on everything. So I get those estimates. Okay.

Now let's say I want to do this for the multi-level models. Remember I mentioned that the estimates of the betas are not on the same numerical scale in mixed models as they are on fixed effect models. So we have to take that scaling into account when we are trying to get these marginal estimates. In other words, we have to use the design effect here properly in order to get the estimates of the marginal response probabilities. So that's what I'm doing there is I'm taking the estimate of the logent here for the control group, here for the TV, here for the CC, here for the group that got everything, and then I'm dividing that by the square root of the design effect to marginalize that. To yield the estimates of the response probabilities in these four cells.

Now you see, they're not exactly the same as the fixed effects, but they're not highly off, although I think ****, yeah, yeah, they're not too far off. They ****, but not greatly. But that step of this marginalization needs to be done, again because the betas in a mixed model are not on the same numerical scale as they are on an ordinary fixed effects model. In the three-level model, the marginalization has the same kind of concept, again it's the square root of the design effect, but here we have the school variance, the classroom variance and the "****" variance divided by what would be the variance if we didn't have clustering. I mean, again, that's how the design effect, the way to think about it, what's the variance under cluster sampling, what's the variance under simple random sampling. That ratio is in a sense how much more inflated these estimates are.

You can see you get these results. So again, this can be done after you've estimated the model you can do a little bit of calculation with those estimates to achieve these results. There's no simple option in SAS to actually produce these for you, so it's not like you can say, option whatever, you have to do this by hand, with Excel, or something like that.

Okay, now let's bring in pre-test. So thus far, we have not considered any level one variables, now let's bring in level one covariant pre-test. Okay, so now I've got a logent model for subject J in cluster I at level one it says, basically that the pre-test could include **** logent. And now at level two I'm saying that the person's intercept is… or I'm sorry, the cluster intercept is related to classroom curriculum, TV and the interaction and the random effects here. And notice what I'm saying about the pre-test.

Now if you just ran this with SAS or whatever, you may not appreciate what you're doing, but when you write this equation out in this multi-level manner, it's kind of telling. What I'm saying here, with this statement here, is that the… well, first of all, it's B1I. That's the effect for cluster I of pre-test on the post-test logent. The relationship between pre-test and post-test. And what I'm saying is that relationship is the same for every cluster 'cause there's only one thing here. Right? It doesn't vary by "I". Okay?

So when you just put in a level one covariate, as a covariate, what you're saying is that that relationship of pre-test and post-test, it's the same for every single school, or every single classroom. Now, you might think that that may not be the case. You might think, for example, that the relationship between pre-test and post-test could be different depending on whether they got class curriculum or not.

Let's say you entertain that idea. How would you change the statement to reflect that? You said that the relationship was between pre-test and the logent for post-test vary depending on whether they had classroom curriculum or not. You would have this; you'd have one over here. You'd have like plus Beta5 times CCI. Right? That would allow an interaction. It would say… right. The relationship… think of B1I as the relationship between pre-test and post-test for cluster I. If you said that that was different depending on whether they got CC or not, it's like you put this down here. It's like plus beta5 times CC.

Now let's say getting to what you were saying, let's see, you thought it was different for every single cluster for some clusters that relationship is real strong; real strong and more diffuse. It's like having a random effect over here. All right, you have a model with what's sometimes called a random coefficient because you're allowing something other than the intercept to be random.

Now, if you're just trying to account for the multi-level structure, typically you don't necessarily need to go that route, but if you have scientific interest in that, is it the case that the pre-test and the post-test relationship varies across the different clusters, that's where you go that route and possibly have a model with a second random effect that would introduce a random effect for the pre-test. And what that would be saying is that the relationship between pre-test and post-test varies from cluster to cluster. For some clusters it's strong for others it's week, something like that.

Okay, again. I think those ideas you wouldn't necessarily come upon you know, when you just run the software, but when you run it like this, you just kind of, you know, you kind of say, well look what I'm doing for the intercept, why don't I do that for the pre-test effect. What would it mean if I did that? Okay, so we won't go that route today. Again, tomorrow we will when we do longitudinal, but that's basically what it would represent.

Okay, so in terms of ****, beta in audit is now the pre-test adjusted logent for the control group. Beta1 is the effective pre-test on post-test. Beta2 is the pretest adjusted logent different between CCS, with CC, that was the classroom curriculum effect. Beta3 is a TV effect pre-test adjusted. Beta4 is the pre-test adjusted interaction and we've got a random cluster deviation in this two-level model the way I've written it here.

Here's the three-level version of this model with pre-test in there. Okay, admittedly, when you get to more than two levels, there's a lot of bookkeeping to be done and a lot of times I make a mistake when I write it out this way. You know, you've just got to think carefully what's, level one students. In level two classrooms, in level 3 schools. In the covariant part, you only put the variables that are measured at that level. Pre-test is measured at the student level. It changes from student to student. So it goes in over here. CC, TV, this, it's the same for every kid in the same school. So it doesn't vary by kid… student. It doesn't vary by classroom. It varies by schools. Okay? Because if you look at a given school, say that's got four classrooms, they all have the same level of TV, let's say. They don't vary, so it's not a level two variable. It's level three. Okay?

This is a full-blown level three analysis that has basically as covariates; it's got only these three, these four covariates. It's got pre-test, these three, and then it's got two random effects. Okay? This is the way it's written in the multi-level, or HLM kind of structure.

Question: John, I have a question about the data management. **** where we have to structure this **** to give me an alpha section. There's no way of referencing data set that have just, you know…

Answer: Yeah, yeah. HLM does it that way actually I think. I haven't run HLM in a long time, but it used to be that you could access, like in this case, it's different data sets. You know, one that had the level one **** information, another one that had level two. Because you're right, if you have a level 3 variable, there's a lot of redundancy there. Yeah. But in FAS, it's stata and those packages… see the thing is, the way they've written it, they don't distinguish between levels… distinguishing between level one, level two, level three, I've done it here, but the software doesn't care about that, really. So you know, those are just different variables. And it doesn't care that it's redundant for this you now, set of observations, per se. It helps us to understand the model, to interpret it, but in terms of the algorithm, it's really not essential. But HLM, I think, still you can do it that way where you can enter them as three data sets. Of course, you're going to have ID's and all of them so that everything gets glued back together again. Okay.

Okay. So here are there results. Again, I've shown the fixed, the **** and the quasi likelihood for this three-level analysis that now has pre-test in and what you'll see here is, again, the interesting story that pre-test and classroom curriculum are significant regardless, so there it, again, it doesn't matter what you do really, you get the right result… and when I say "right", you're getting the same conclusion, let's say.

For TV and the interaction again, by the model that doesn't take into account any kind of clustering, the fixed effects model, you would see a significant TV effect, a significant interaction, none of those… those two effects are not significant in any of the multi-level models. And really there are only two multi-level models here, but have estimated them these two different ways. Again, the quasi-likelihood the easier thing does quite fine here in relation to the full likelihood procedure, so there's no necessity here to do that. The gain is that you do get the **** values that you can use for likelihood testing.

Now, notice this, here, when we go from the fixed to the two-level model, we can see a difference of about 16 **** freedom, so that's significant. Going from the two to the three is really pretty minor. That difference is you know, about one point, right? So again, one would argue… I would make the argument that this is a three-level design, so we should have the school variance in there, but you can appreciate that the model solution is not very different between here and here.

If you can do the three-level, might as well because logically it makes sense that you know, because schools were randomized to the different conditions, but in terms of the fit of the model, it's really not much of a difference between the model that accounts for the classrooms only and the model that comes from the classrooms and schools.

Okay. And then again, we can calculate the two-level ICC's here we get six percent now at the classroom level, it's diminished somewhat because we put in this **** of a pre-test, so again, still a fair amount of variation, still at the classroom level, six percent. And then we can do the three-level calculations for the three-level model, so schools – okay, we saw that the school variances you know, was not really doing a whole lot, it was what now at about two percent, classroom six percent, and then the likeness of classes in the same school about .29. Again, what that reflects is the fact that the classrooms are having

more of an effect than the schools per se. So again, schools is not so important here especially where we've accounted for the pre-test, but again, for logical design reasons, it would be worthwhile to keep it in the model.

And then finally, again, to do that fitting of the marginal observed proportions again, I have just illustrated these calculations here for the… for arms. Now here what I've done for the treatment effect… I'm sorry, for the pre-test effect is, here's the coefficient for the pre-test effect. And what I've done is to use the mean of pre-test score for the given arm, you know to calculate this estimated probability I the control arm, in the TV, in the CC and then in the TV plus CC. Okay, so point four here is, again, the estimated coefficient for the pre-test effect. Of course, pre-test is different for these four different arms and that's what these means reflect. It's pretty similar but that's what those reflect. And I did the same thing here for the random classroom in the three-level models. Again, when you do these latter two, you need to do these marginalization factors that I mentioned to get the proper estimates. Okay.

Okay, well let me just summarize a little bit and then I'll actually run these things and you can run it with me if you like till the end today. So, I would say, mixed logistic regression modeling, direct extension of ordinary logistic regression, I would hope you would see it's useful **** for multi-level data. Software is available in SAS and other programs. Extended models are available for ordinal nominal count outcomes. As we'll see tomorrow, similar models can be used for longitudinal, albeit, more issues. Okay? Why is that? Again, the data within clusters are usually correlated to some degree. The data within subjects are usually correlated to a great degree. You know, when you think about it yourself you get the same answer every day, right. But if you ask kids in a classroom, you'll get some similarity, but not as similar as asking the same person that we… longitudinally.

So we usually need more random effects there. We do a little bit more modeling and then also, what we'll also get into tomorrow with longitudinal, residual and missing data and attrition because they raise their head in a way that we don't see so much in clustered. And so similar models, but more extensions that we want to think about.

Okay, well, let's try running some of these things that we went over now. We'll take – and so we'll get to SAS. So previously we had this TDFFP mix example and we did all this running with Procmix, so let me get instead… and let me clear out the output and the log file, and then I'll bring in the newer Syntax file. Okay. And we'll take a look at the dataset that we're actually going to be using.

Okay, so does everybody have this folder with these SAS examples and things? The one we want to use now is called TVSFP Glimmix **** Mix. It has these examples that we've looked at this afternoon using Glimmix, and it also has a Syntax using Anomix as well. And "B" is for binary. So that's the logic of that naming convention. So let's bring that in. Let's try again, open, okay, now I've got it.

Now, if you open it, you'll have this Syntax as well as the previous Syntax file that you used earlier today. I'm gonna close this one out because I don't want to deal with this. I'm going to be focusing on this, but you could have it both around, it doesn't really matter. I guess I'll just leave it, it doesn't matter.

So this syntax again, was actually in our handouts, we went over this a little bit. What it's going to read is this dataset TVSFPOrs.dat, and it's reading it from in this case, a directory called Data. So, let me copy that dataset into that location. Or, I imagine, oh it hasn't been copied then?

Question: Changing, everybody is changing to a ****.

Okay. Isn't it, is it in C drive? Training. Okay. I see. So here it's in training and then it's – oh, it's in here okay. Okay, so all we got to do is change the name. Okay. So we'll go here and do that. Yeah, yeah, yeah, let me try that… that seems like that would be pretty slick. Oh, from mixed, oh that's good – good idea. There we go. Okay, so we'll go there. Right. Here's the correct folder location. Okay. Training, San Antonio. I'll tell you, it's a lot nicer being in San Antonio than Chicago today. It's really cold this morning. You know, I came here, it's like, people with short sleeves, it'd December. What is this? Is this a meeting of the Polar Bear Society? No, it's warm. Is it really. Oh good. Well when you want cold, come up and see us. But you know what I found, I was reading in a – I took American Airlines and they had a magazine and there was this whole article about how the state of North Dakota was actually one of the few states that, in these bad economic times is actually flourishing. They're – they've got like 3% unemployment in North Dakota. But you've got to live in Fargo and you've got to deal with, I mean, compared to Chicago, that's much, much colder. Well it seems in terms of economy, they're doing great. Okay here we go.

Okay, so what you want to do is to have as you in file, this location. It's in the folder Training, and then within that, San Antonio10.dat.SAS.Examples and it's got those little underscores there, and it'll read this dataset. And what I'm running here respectively is the ordinary logistic regression and then Glimmix quasi likelihood, then Glimmix 3 level quasi likelihood, then after that we get the full likelihood versions in Glimmix and then the two-level model in Anomix. And if you're really curious, you can see that the estimates that I get out of this and this are exactly the same. So that's what helped me switch to Glimmix because I used to say, "Ah Glimmix is sort of proximate, I don't like it. I like to do the full likelihood." But I actually thought I could do it in Glimmix, and it gives me exactly the same results as the full likelihood, Anomix.

So let's run these things now. And that's -- that's all I have here, so we'll just run them and see how long they take. Okay, so now it's running Logistic, and it's dong with that one not it's running Glimmix, so there's going to be four runs. Two three-levels. Okay, so it's gotten through a few of those. The one that's going to take the longest to be honest is NLMIXED. And it's kind of weird because NLMIXED is doing exactly the same thing as one of those Glimmix runs, but the amount of computation times if you

look at the log file is quite a bit different actually and I'm not sure why, but that's the one that takes the longest. Okay, now it's all done.

So you can see we will need the 21 points on three of those and then the other one was – those other two were three-level – I'm sorry, 21 point on two, because two of them were quasi likelihood and then as we did two three-level analyses and it all ran pretty fast. Right? So that wasn't too bad really. And it's quite a – it's not like it's a toy dataset either, it's 1,600 observations. So it's. I mean you probably have datasets that are larger than this, but it's not – it's not that small per se.

Okay, so let's go to the top here. This is first the logistic model results. Number of observations, 1,600, that's good news. It tells you now this nice message because I put descended. "Probability model is T… you now, is equal to 1". Okay, so we're modeling the probability of these students getting things correct, or three or more correct. Okay.

Now this is what I always like to see, "convergence criterion satisfied." You want to see that. If you don't see that, not so good. But you want to look for that. The model fits statistics. What it's doing is comparing via AIC, SC and the deviants, modeled with or without the covariance. Okay, so here for example, do you want to do a likelihood ratio test comparing the model with or without the covariance. You can look at this deviance value, this deviance value, the difference there is chi squared on three degrees of freedom. In fact, it actually does carries that out for you there… here with the likelihood ratio chi squared value here of 50.0177. So, that's indicating that's highly significant, but realize this inference is carried out under the assumption that these 1,600 subjects are independent of each other. So it's treating this not as clustered, not as multi-level, not yet. Because we're just doing ordinary logistic regression here.

So the inference here is not strictly speaking, correct, because this is a multi-level dataset, but we're not treating it as multi-level yet. That's why you see these significant results for TV and the close to significant result for the interaction. Okay. Again, if you want to use NLMIXED though, these are pretty good things as starting values for those coefficients.

Okay, it also in logistic takes the estimates and exponentiates them to give you the odd ratio and then 95 **** presents competence limits for those.

Okay, then we get to Glimmix. Now the first version that I'm doing is with the quasi likelihood two-level model. It gives a lot of information, binary, logent, there's all kinds of options you could potentially change. I used the more or less defaults on many of them. We used descending so it's indicating we're modeling the probability of a one response.

Subjects. This is actually the number of classrooms. Okay? It's the terminology is really more appropriate for longitudinal where you have repeated observations within subjects. So this is not subjects per se, but really clusters, or you can think of it here as level two

units. There's 135 classrooms. What the maximum observation per cluster? 28. So there's a classroom there with 28 in them.

It gives some additional information about how it maximizes the likelihood. Now again, this is the most interesting part of the output, if you're having trouble sleeping tonight, look at this over and over again. Boom! You'll be sleeping.

Iteration history. What is iteration? That's **** the likely that the solution is sort of kin to getting to the top of a multi-dimensional hill. Why? Because you are looking at a function, you're seeing where… the solution is where the derivatives don't change at all. So you're trying to get to the top of the hill, that's what you're getting here is how much change was there in terms of what the gradient… the first derivatives. And when that gets rather small, it then converges. Every – none of the parameters are changing by much, you're at the solution point. Okay? So, you can specify different convergence criteria if you like, but this is you know, you can see it's like to a negative eight, so it's very precise by default.

Then here are the results. Now again, this first run is with the pseudo-likelihood. So it does identify it, minus two times the residual logs, pseudo likelihood. I really don't know what purpose you can use that for, because again, you can't use that necessarily for likelihood ratio testing, but it does print that out. It also prints out here the estimate of the classroom standard deviation. Okay, because again, I've asked for this **** version. So that's the standard deviation. So again, in **** you'd want to square that, .25 or thereabouts.

Here are the solutions for the fixed effects. Okay. It expresses them as estimates, standard errors, degrees of freedom, the "T" value and then the "P" value. So again, note that the latter two are highly non-significant, whereas in the model where I ignored the clustering, you know, one was slightly below .05 the other one was slightly above, here they're like .25, .26. Okay, so this is the two-level analysis taking into account the clustering of kids in the classrooms. It also gives "F" test results. Realize this, if you have a "T" statistic, you square it; you get an "F" test with one numerator degrees of freedom. So all of these "F" values here are just the squaring of these "T" values right here and the "P" values you see I this column are identical to what you see for the regressers here. Okay, so it's just a different way of reporting the test result.

Okay, then after that, we get to three-level quasi likelihood. Okay, see where that other information we get here. Okay, so we got a binary, let's see, that's obs per subject that must be 137, that's per school, there's 28 schools. I guess it doesn't give us information on the 135 classrooms here in terms of the dimensions. Okay.

Again, the iteration history, that's so interesting. And then the results of this, we get the estimates for the standard deviation for classrooms nested within schools and for schools that are standard errors and then the solution of the fixed effects. Now, sometimes you see this and you say, shall I worry about this? Degrees of freedom, zero. I'm not precisely sure how they're figuring out the degrees of freedom, but realize this, have you

ever studied a “T” distribution compared to a standard normal? Well, haven’t done that? One of my favorite things to do.

However, those “T”, “P” values, are compared in the standard normal. Guess what? They’re not very different if the degrees of freedom are like 50 or higher. Okay, so it’s reporting here a “T” value, which is just the estimate divided by the standard error and to get those “P” values, it’s comparing them to a… “T” distribution in this case, these many degrees of freedom. Do you think these “P” values are going to be any different if they got it on the “Z” as opposed to “T”? No. The degrees of freedom are huge.

So I’m not sure why you get zero here necessarily, but if you really want a “P” value for this “T” static? Just look at a standard normal. I’m not sure what SAS has done there. I wouldn’t worry about that per se, that zero, but for whatever reason, you see that on the output and you’re first reaction might be, oh, something’s wrong here. Something might be wrong with how fast it did that, I don’t know, but it’s not something that I’d worry about. Okay?

Okay, so then this is then the three-level analysis. Yes.

Question: So where does the degree of freedom of 1,465?

Answer: Yeah, where’s that coming from?

Question: Where are they from?

Answer: Right, that’s a good question? Well, it’s kind of like this. Like in a Nova models, they are very precise equations for figuring out degrees of freedom, right? **** regression models, very precise formulas for figuring out degrees of freedom. In mixed models, there’s different ways to get at degrees of freedom for this reason. Let’s say this was a – you know, there’s 1,600 observations here, there are 1,600 observations that are correlated to some degree and that correlation plays a role in how it calculates the degrees of freedom. Think about this, if you have 1,600 observations and they’re totally independent of each other, you’ve got more information than if you have 1,600 observations that are totally dependent. Right? So that association plays a role in the calculation of the degrees of freedom column. And there’s different ways of doing that. So here, I’m using the default and to be honest, I can’t describe it in great accurate detail, but that’s what it involves. It involves essentially how much data do you have, how correlated are those data, and how many parameters are you estimating. Those are the elements that go into this.

Again, if you have relatively large datasets, you don’t’ have to worry about this really because again, if this is, you can compare this to a standard normal and you get the same result. You are essentially doing an **** test once you get past 50 anyway. It’s not like it matters greatly. But that’s sort of the elements that go into that.

Okay, so those are the pseudo likelihood versions and in the time we have left, we'll go into the full likelihood versions and then to finish things off, you always want to finish on a high point. Right? You know, we're going to go with the NL mixed results. That'll really drive you out of this room real quick. You won't be flicking your lighters for an encore at that point. I don't think so.

Okay, so here's then, Glimmix, where I used full Likelihood. See here's where the degrees of freedom method containment. There's a couple of different options for how it actually calculates those things. And I'm just using the default here because again, my sample size is rather large. Where you have to worry about that is, let's say you've got, I don't know, 20 clusters and you've got three observations per cluster, maybe something like that. You're more in a smaller sample situation. There, how you calculate the degrees of freedom might play a role. Here where you've got 1,600 within 135, 28 schools, it really doesn't matter because you've got so much data. But if you do have small data, in general, the one to use is one called the Kenward Rogers adjustment for degrees of freedom. SAS has that. If you're in small sample situation, that's probably the way to go. Okay.

Okay, this is again, the two-level analysis because I have 135 classrooms, there again, it's giving a lot of information, some of which is helpful, some of which is not. I'm using 21 quadrentur points here. All right. Here's a result. Now here in terms of **** statistics, that is the value of the deviance that I can use for likelihood ration testing. So that 2138, I can use that for example, compared to my deviance value from my prologistix to do a statistical test of whether or not the cluster variance is significant or not. Because that's the new parameter in this model that wasn't in that model.

Okay, and then here, you know, you see everything else pretty much as stated. Again, the classroom standard deviation. Again, it's not really… the point estimate ****here is pretty similar to what we got in our quasi likelihood estimation. So the quasi likelihood looks pretty well in this scenario. We get our fixed effects. Again, very similar results to what we saw previously. So no change in the story really there.

We go to the three- level analysis, a lot of information here. You shouldn't skim over this so much. I mean, when you first run a dataset, you want to take a look at these to make sure they kind of make sense, you know, that you've got this many observations so that it's not throwing out observations for some crazy reason or not. I mean, a lot of times when people bring me output that's the first thing I take a look at and I ask questions. Do you have, you know 28 schools or whatever. Does this make sense because you don't want to jump to the output for garbage results? So here I'm glossing over that because I have looked over this many times, but it is not something that should be always glossed over like I am doing right now.

Okay, so here we get our deviance for this model, and again, we could do a likelihood ratio test to compare to the previous model. Here we get estimates of the classroom standard and school standard deviations, fixed effects. Okay, again, the solution hasn't really changed here. CC is highly significant; again, the latter two are not. Again, when

we're working with the three levels, we still have this degrees of freedom thing with the intercept. Who cares? Who tests the intercepts, right? I mean… what does the intercept represent here?

Okay, that's a good test question on an exam. What does the intercept represent? Right. Because it's not somebody ever pays attention to. What it represents here is, that's the logent for the control group. Okay.

And so by testing whether it's zero or not, you're testing whether the… the logent equals zero, the odds equals 1, the probability equals a half. That's what we're testing with this parameter. You know, and it is significant in this model. It's above… so we're saying that basically, for the control group, the probability of three or more correct is lower than 50%, right. It's statistically… it's lower. The point estimate is negative. Negative logent, odds less than one, probability less than half, and it's significant.

So for the control group, right? And what did they have, right? They have an observed proportion of about 40%. So it's basically the same, that 40% is statistically different than 50%. That's what the intercept represents in this model. So it actually does have some interpretational quality here.

Okay, so that's – those are the results. Then finally, we have just a few more minutes left, okay. It's not time for the big hit. Okay. NLMIXED. Right. NLMIXED. Here it tells the binary that you're… the **** variable. Here's the thing I love the best about NLMIXED, distribution for random effects. You have to state normal. That gives you the impression, when I first saw that, that oh, it can fit non-normal. No it can't. You have to write normal, and it prints out normal, but that's the only thing it'll do. It just prints normal.

Adaptive quadrant … a thousand quadrant, that's what NLMIXED does by fault. That why, prior to version 9, I always used NLMIXED to estimate these models. I liked full likelihood estimation, again, now Glimmix can do that, so there's no difference in term of that. All right. It does the 21 points here, here there are 1,600 observations. I'm doing the nesting here within classrooms because NLMIXED cannot do three level and they also state that is one of its problems or limitations I should say. It's not really a problem, it's a limitation.

It gives you that fascinating iteration history we see at the end it converged. Here's the deviance. Again, we can use this for likelihood ratio testing. And then it gives various penalized versions of the deviants. Where to use those? If you want to compare models that are not nested to each other, for example, in model selection. We often sometimes use AIC when we want to compare models that are not nested to each other and NLMIXED tells us that smaller values are better. Okay, so that's what that is all about.

And then finally here, we get our parameter estimates. Now again in NLMIXED you have to actually name those things and I have chosen those names, B0 for the intercept, BCC for the classroom curriculum, BTV, I've named those things. When I first was

running it I was naming things B0, B1, B2, B3, and then I'd have to look, oh what is B1? It's like… then I remembered, look I can name it anything I want. I don't have to call B1, B2, B3… I got a little smarter. Well, let's put the name of the variable and the name of the coefficient. You know, it's ****.

Also, here I have done an estimate command in the NLMIXED Syntax to give me an estimate of the ICC and so here I get that value of point, about eight percent of the variance at the classroom level. And here by doing so you get a statistical test of the ICC. You can see it's highly significant. And it actually gives you that information.

The other columns that I wouldn't pay too much attention to, well it gives you a confidence interval, I guess, the low and the upper, that might be of interest here. Ninety-five percent confidence interval for the parameters. The gradient is basically at the solution, first derivative; you should be close to zero for those things.

And that's about it. So that's how to run mixed models, multi-leveled models for clustered data. That's where we went through sort of ad nauseum today, and at this point you might be a little nauseas, but tomorrow we're going to hit longitudinal data and there's gonna be more challenges there, but more interesting things as well. And really, though these things are separated, you could really also have, obviously, a combination. You could have repeated measures on a subject… subjected nested within a cluster. Right? You could easily have that. And think, in a way, with that is like is a three-level analysis. Here I've got students within classrooms in schools. But I could also have repeated observations within subjects within clinics. I could have four-level. With this dataset even.

I mentioned there were two time points, pre-test, post-test. I lied, there were four time points. Those pre-test, post-test, and then two follow-ups. If I want to model that, I could have… right. The repeated measures within subjects, within classrooms, within schools.

We won't get to the four level… sorry. But we'll meet again tomorrow and what time do we start in the morning? Eight o'clock. Oh, okay.

Eight o'clock, sounds great. And thanks for you're attention today and sorry about the logistics, but I think Subash again, did a wonderful job and I'd like to thank him once again for covering for me.