Questions and Answers (Q&As)

Mind the Gap Webinar: An Introduction to Cross-classified, Multiple Membership, and Dynamic Group Multilevel Models

Donald Hedeker, Ph.D.

October 20, 2022

Q: How was the 'cons' in the SAS code reflected in the model equations when you talk about the multiple membership modeling using the nurse dataset?

A: The 'cons' variable is a constant of one for each observation (i.e., subject) in the dataset, so it isn't really in the model equation, per se. It identifies the entire sample as a cluster (on the random statement). With that, one then specifies the 25 weights (p1–p25) as independent random effects with the same cluster variance. Thus, this is a bit of a trick to get the software to estimate the parameters of the multiple membership model.

Q: In hierarchical multilevel modeling, we pay special attention to cluster size. I imagine that in cross-classified analysis, we would also be concerned on cluster size. What should we do if clusters have widely variable cluster size?

A: The likelihood equations allow for varying cluster sizes in both the hierarchical and cross-classified models. Nonetheless, there has been research on the effect of varying, and especially small cluster sizes in multilevel models. The issues would seem to be the same for hierarchical and cross-classified models.

Q: Can we analyze cross-classified data with structural equation modeling (SEM)? If we can, what are the pros and cons of SEM versus cross-classified analysis?

A: Sorry, but I don't really know. I am not an expert on SEM software. Based on some googling, it does seem like this might be possible with Mplus, but I can't say for sure.

Q: I am working with a physician mentee who has greater than 1 million measures of pain assessments taken from patients within a hospital. Different providers provide assessments throughout the duration of patient stays in the hospital (i.e., there are repeated measures for each patient). Assessment measures are often binary or ordinal (i.e., not continuous). What type of cross-classified model would you recommend for this type of setup and is it even a computationally feasible problem to contemplate fitting with multilevel models? The physician and his team are currently using the simple cluster-adjusted standard errors approach in Stata.

A: In theory, this might be possible as an acute-effects cross-classified logistic model, either binary or ordinal. Here, it seems like there is a crossing of subjects and providers. So, it would be similar to the example that I provided where subjects and schools were crossed. However, given the data size, it might not be computationally feasible, in which case using the cluster-adjusted approach might be the only way to go.

Q: If the subjects are not measured at the same time points, but measured at different random time points, can we still use the dynamic model?

A: It seems like you could, but with a few caveats. First, if you thought that the cluster effect changed, say linearly across time, then the syntax could be:

RANDOM intercept year/SUB=schoolidccrem TYPE=UN;

This would mimic the random intercept and trend that is specified for the subject, in this example, and allow the cluster effect to have an intercept and linear trend across time.

Another possibility would be to form time bins for the multiple time points (say months 0–6, 7–12, etc.) and then to use the approach that I illustrated.

Q: I'm working on a model of the impact of social determinants of health at the individual and community level on opioid overdose deaths, with clustering around education, health care access, employment, and criminal history/criminality. Would a cross-classified model be applicable?

A: It seems like you have individuals clustered within communities, and then the other factors (education, health care access, employment, and criminal history/criminality) would be subject-level covariates. So, I don't think that a cross-classified model is necessary here, but I might be missing something.

Q: How do you write SAS code for zero-inflated poisson (ZIP) models with repeated observations?

A: As far as I know, SAS does not yet allow for ZIP models in PROC GLIMMIX, which is the PROC for mixed models pertaining to non-normal outcomes. One can potentially do this using PROC NLMIXED, which allows for programming features. This is not necessarily easy, but this site has some material on this topic:

stats.oarc.ucla.edu/sas/faq/how-do-i-run-a-random-effect-zero-inflated-poisson-model-using-nlmixed/

Q: I am investigating the effect of a preschool intervention on several cohorts of third graders. What do you recommend on for controlling for the effect of cohorts? Would this qualify as a cross-classified model where students can be in the same school but in different cohorts and vice versa?

A: It depends on how many cohorts you have. If there were, say 15 or more, you could do this as a cross-classified model with the crossing of cohort and school. However, if there aren't that many cohorts, then simply using cohort as a categorical covariate (e.g., using C-1 dummy codes for the C cohorts) would at least control for overall cohort differences.

Q: How are these analyses applied to binary outcomes?

A: Yes, one can use a logistic mixed model for binary outcomes. So, in SAS, instead of using PROC MIXED, one could use PROC GLIMMIX for this. For example, the PDF below describes how to use PROC GLIMMIX for a multiple membership multilevel model.

www.lexjansen.com/wuss/2017/100_Final_Paper_PDF.pdf

Q: Can sampling weights be applied with cross-classified data?

A: In theory, this would seem possible, but I don't know for sure if there is software for this. Stata does allow this for ordinary multilevel models, but I am not sure about these cross-classified models.

Q: How do we choose weights for the multiple membership random effect model?

A: That really depends on the particular application. It might be the amount of time that the person spends with the different clusters, or it could be something else. If there are multiple ways of defining the weights, then a sensitivity analysis would seem to be useful. In this case, you run the analyses using the different ways of defining the weights, and compare results to see if your conclusions are similar/different. This will give a sense about the robustness of the results to the varying ways of defining the weights.

Q: In the nurse example, the weights are negatively correlated. Can types other than TOEPLITZ(1) work?

A: Not sure that I understand about the weights being negatively correlated. The use of TOEPLITZ(1) just ensures that there will be a single cluster variance estimated attributable to the clustering of patients within nurses. As mentioned in the first question, this is something of a trick to get the software to estimate the parameters of the multiple membership multilevel model.

Q: Are there methods for calculating sample sizes in these settings?

A: This would probably best be done via simulation, since the particulars of the data would be pretty unique for each case. For example, there might be a great deal of crossing or not quite so much. So, it seems like it would be a bit difficult to come up with simple sample size equations in this case. The use of simulation for cross-classified models is described in this PowerPoint link:

www.bristol.ac.uk/cmm/media/software/mlpowsim/ssxc.ppt

Q: Could you comment on applications of these methods to spatial models?

A: Yes, it would seem that there would be cases where these models apply. For example, suppose subjects are clustered by census tracts, but subjects move around and therefore can spend a different amount of time in different census tracts. If this information is available, then one might try a multiple membership model using the amount of time a person spends in the different census tracts as the weights.

Q: Do these models apply for other kinds of outcomes (e.g., binary, ordinal, counts)?

A: Yes, but other statistical software programs would be necessary. For example, in SAS one could use PROC GLIMMIX for other outcome types.

Q: Is it possible to model data where some subjects are fully nested (e.g., students in schools) and others are cross classified (e.g., students within a crossing of groups and schools)?

A: Yes, such data are called partially cross-classified multilevel data, and a modeling approach is described in the following article, which includes sample code in its appendix:

Luo W, Cappaert KJ, Ning L. Modelling partially cross-classified multilevel data. British Journal of Mathematical and Statistical Psychology. 2015;68(2):342–362. doi: 10.1111/bmsp.12050. PMID:25773173.

Based on simulations, they write: "Modelling partially cross-classified data with the fully nested model and the fully crossed model results in biased parameter estimates."

Q: What is the impact of inappropriate modeling of cross-classified data?

A: This topic was investigated in the following paper:

Meyers JL and Beretvas SN. The impact of inappropriate modeling of cross-classified data structures. Multivariate Behavioral Research*. 2006;41(4):473–497. doi: 10.1207/s15327906mbr4104_3. PMID:26794915.*

These authors provided an illustrative example and performed a simulation study. From the simulation study, they concluded: "Results indicated that when the structure is ignored, fixed-effect estimates were unaffected, but standard error estimates associated with the variables modeled incorrectly were biased. Estimates of the variance components also displayed bias, which was related to several study factors."

Q: What is the impact of ignoring multiple membership data structure when it exists?

A: This was investigated in:

Chung H and Beretvas SN. The impact of ignoring multiple membership data structures in multilevel models. British Journal of Mathematical and Statistical Psychology. 2012;65(2):185–200. doi: 10.1111/j.2044-8317.2011.02023.x. PMID:21732931.

In this paper, they used simulation with students (level-1) clustered within schools (level-2), with the condition that some students were members of multiple schools (either 10% or 20% of the students). Their results indicated that standard multilevel modeling of such data (using the last school that a student attended, rather than all schools and a multiple membership model) resulted in biased estimates of the cluster-level (level-2) covariates, as well as biased estimates of both level-1 and level-2 variance parameters.

Q: What is the impact of ignoring interaction between the crossed factors when it is present?

A: This was examined in:

Shi Y, Leite W, Algina J. The impact of omitting the interaction between crossed factors in cross-classified random effects modelling. British Journal of Mathematical and Statistical Psychology. 2010;63(Pt 1):1–15. doi: 10.1348/000711008X398968. PMID:19243680.

They report: "Results from a Monte Carlo simulation study indicate that, for fixed effects, both coefficients estimates and accompanied standard error estimates are not biased. For random effects,

results are affected at level 2 but not at level 1 by the presence of an interaction variance and/or a correlation between the residual of level two factors."