Assignment 03
Cross-Sectional Multilevel Regression Models

This assignment is intended to give you experience fitting and interpreting multilevel regression models. Use R Markdown to produce an HTML (or PDF) file that includes your responses to each of the questions on this assignment. Please adhere to the following guidelines for further formatting your assignment:

- All graphics should be resized so that they do not take up more room than necessary and should have an appropriate caption. Learn how to do this in a code chunk using knitr syntax.
- Any typed mathematics (equations, matrices, vectors, etc.) should be appropriately typeset within the document using Markdown's display equations. See here for some examples of how mathematics can be typeset in R Markdown.
- All syntax should be hidden (i.e., not displayed) unless specifically asked for. Any messages or warnings produced from loading packages should also be hidden.

For each question, specify the question number using a section header. Submit your HTML (or PDF) file and your RMD file via email that you send to both Jonathan and Andy.

For this assignment, you will use two files *popularLevel1.csv* and *popularLevel2.csv*. The first file contains data on 2000 different students. The variables in this file are:

- `student`: Student ID (within a school)
- `class`: Classroom ID
- `popularity`: Popularity rating of the student (0-10 scale; where 0 = very unpopular and 10 = very popular) based on having all other students in the student's class rate him/her and computing the average of those ratings.
- `teachPop`: Student's popularity as rated by the teacher. Higher values are indicative of higher popularity
- `extra`: Extraversion of the student (measured on a 10-point scale)
- `female`: Sex of the student (Male = 0; Female = 1)

The second file contains data on 100 different classrooms. The variables in this file are:

- `class`: Classroom ID
- `teacherExp`: Experience level of teacher, in years

The data, provided by Hox (2002), were simulated from data collected as part of a sociological study of student popularity. Student popularity, a rating on a scale of 1–10, was derived by a sociometric procedure in which all students in a class rate all the other students. Each students' popularity is the average received popularity rating. Because of the sociometric procedure, group effects as apparent from higher level variance components, are rather strong and should be examined using multilevel regression models.

Hox, J. J. (2002). *Multilevel analysis. Techniques and applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Preparation

To begin the assignment, you will need to merge the `popularLevel2` data into the `popularLevel1` data. This should result in a data frame with 2000 rows and 7 variables.

Baseline Model

To begin the analysis, you will fit the intercept-only model with a random-effect of intercept (henceforth referred to as Model 0). Use the results of fitting this model to answer the questions in this section of the assignment.

1. Using symbolic notation, including variable names where appropriate and all residuals, write the equations for the multilevel model based on Model 0.

2. Compute and interpret the value of the intraclass correlation.

Selecting Level-1 Fixed Effects

In a multilevels analysis, after fitting the baseline varying intercept model, the first thing to do is select the fixed effects that you will include in the model. When doing this it is important to use maximum likelihood (not restricted maximum likelhood) for the estimation. It is also important that when choosing the fixed effects to include, that the models you are comparing have the EXACT same set of random effects.

Refit the baseline model using maximum likelihood estimation (Model 0), and also fit the three other candidate models listed below. Each of the additional model should also include the fixed and random effects for intercept.

- Model 1: Fixed effect of extraversion
- Model 2: Fixed effect of sex
- Model 3: Fixed effects of extraversion and sex

3. Compute evidence based on the AICc to make a decision about which of the four models, fitted with maximum likelihood, should be adopted. Present the evidence along with a clear indication of which model you adopted.

Use the model you adopted from Question 3 to answer the remaining questions in this section.

4. Using symbolic notation, including variable names where appropriate and all residuals, write the equations for the multilevel model based on your adopted model.

5. Based on the `summary()` output for your adopted model, is there evidence to support the inclusion of the fixed effect for extraversion? What about that for sex? Explain.

6. Interpret the fixed effect for extraversion from your adopted model.

7. Interpret the fixed effect for sex from yuor adopted model.

Selecting Random Effects

After selecting the fixed effects that should be included in the model, then we can examine models that also include other potential random effects in addition to the intercept. When doing this it is important to use restricted maximum likelihood (not maximum likelihood) for the estimation. It is also important that when choosing the set of random effects to include, that the models you are comparing have the EXACT same set of fixed effects.

Refit your adopted model using restricted maximum likelihood estimation (Model 0), and also fit the three other candidate models listed below. Each of the additional model should include all of the fixed effects from your adopted model and the random effect for intercept.

- Model 1: Random effect of extraversion
- Model 2: Random effect of sex
- Model 3: Random effects of extraversion and sex

8. Compute evidence based on the AICc to make a decision about which of the four models, fitted with restricted maximum likelihood, should be adopted. Present the evidence along with a clear indication of which model you adopted.

Use the model you adopted from Question 8 to answer the remaining questions in this section.

9. Using symbolic notation, including variable names where appropriate and all residuals, write the equations for the composite model.

10. Compute the proportion/percentage of the variance that is within-class, and that which is between-class from your adopted model.

11. Compute the proportion/percentage of the between-class variance that is attributable to differences in the extraversion effect from your adopted model.

Level-2 Predictors

After selecting the level-1 fixed effects and random effects that should be included in the model, we can examine potential level-2 predictors.

12. The level-2 predictor we will consider in this assignment is teacher experience. The most complex model we can fit will include the predictor of teacher experience in each of the level-2 models. Starting with the model you adopted from the previous section, write the equations for the multilevel model in symbolic notation, including using variable names where appropriate and all residuals, that also includes the predictor of teacher experience in each of the level-2 models.

13. Write the equation for the composite model in symbolic notation, including using variable names where appropriate and all residuals, based on the multilevel model you wrote in Question 12.

14. Examining the models you wrote in Question 12 and 13, which estimation method (ML or REML) should be used to fit this model in order to compare this model to the model you adopted in the previous section? Explain.

15. Fit candidate models that include the level-2 predictor of teacher experience. Fit all the model listed below using the estimation method you identified in Question 14. Compute evidence based on the AICc to make a decision about whether this model, or the one you adopted in the previous section, should be adopted. Present the evidence along with a clear indication of which model you adopted.

   - Model 1: Your adopted model from the previous section (Question 8)
   - Model 2: All effects from Model 1 and the effect of teacher experience in all the level-2 models
   - Model 3: All effects from Model 1 and the effect of teacher experience in the level-2 models for intercept and extraversion
   - Model 4: All effects from Model 1 and the effect of teacher experience in the level-2 models for intercept and sex
   - Model 5: All effects from Model 1 and the effect of teacher experience in the level-2 models for intercept

16. Based on the model you adopt from Question 15, which of the fixed effect estimates should be interpreted? Explain.

Summary Table of the Fitted Models

17. Use the **stargazer** or **texreg** package to create a table presenting the results from (1) the baseline model, (2) the best model from the "Selecting Fixed Effects" section, (3) the best model from the "Selecting Random Effects" section, and (4) the best model from the "Level-2 Predictors" section. Be thoughtful about what is presented; don't just use the defaults. For example, if the sample size is the same for all of the models, put that in the table caption rather than as a row in the table. You may want to find a journal article from your field that includes a table of multilevel model results to inform your decisions about what should be included.