Assignment 07

Multilevel Regression Models II

This assignment is intended to give you more experience working with multilevel regression models. *Do not include any R syntax or output unless it is specifically required in the question.* Please submit your responses to each of the questions below. Please submit your responses to each of the questions below in a printed document. All graphics should be resized so that they do not take up more room than necessary and should have an appropriate caption. All tables should also have an appropriate caption.

This assignment is worth 20 points. Each question is worth 1 point unless otherwise noted.

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.

The data come from 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. Also, you need to fit the following four models. For each of these models, account for the correlated nature of the observations within classroom.

Model A

$$\text{Popularity}_{ij} = \beta_0^* + \epsilon_{ij}$$
$$\beta_0^* = \beta_{00} + \eta_{0j}$$

Model B

$$\text{Popularity}_{ij} = \beta_0^* + \beta_1^*(\text{Extraversion}) + \epsilon_{ij}$$
$$\beta_0^* = \beta_{00} + \eta_{0j}$$
$$\beta_1^* = \beta_{10} + \eta_{1j}$$

Model C

$$\text{Popularity}_{ij} = \beta_0^* + \beta_1^*(\text{Extraversion}) + \beta_2^*(\text{Female}) + \epsilon_{ij}$$
$$\beta_0^* = \beta_{00} + \eta_{0j}$$
$$\beta_1^* = \beta_{10} + \eta_{1j}$$
$$\beta_2^* = \beta_{20}$$

Model D

$$\text{Popularity}_{ij} = \beta_0^* + \beta_1^*(\text{Extraversion}) + \beta_2^*(\text{Female}) + \epsilon_{ij}$$
$$\beta_0^* = \beta_{00} + \beta_{01}(\text{Teacher Experience}) + \eta_{0j}$$
$$\beta_1^* = \beta_{10} + \beta_{11}(\text{Teacher Experience}) + \eta_{1j}$$
$$\beta_2^* = \beta_{20}$$

Summarizing the Models

1. Add all four models to an appropriate summary regression table. This table should include rows for each of the fixed effects, and all variance components. At the model-level, the log-likelihood and AICc values for each model should also be included in this table. **(3pts)**

Pseudo R^2 Values

2. Compute the proportional reduction in the Level-1 (residual) variance component for Model B, Model C, and Model D. Add these to your summary regression table.

3. Compute the proportional reduction in the Level-2 intercept variance component for Model B, Model C, and Model D. Add these to your summary regression table.

4. Compute the proportional reduction in the Level-2 slope variance component (for extraversion) for Model C and Model D. Add these to your summary regression table.

5. Compute the squared correlation between the observed and fitted popularity values for all four models. Add these to your summary regression table.

6. Explain why it is not unexpected that the proportional reduction in the Level-1 (residual) variance for Model C is greater than zero.

7. Explain why it is not unexpected that the proportional reduction in the Level-1 (residual) variance for Model C and Model D are similar.

Model Selection and Interpretation

8. Based on the AICc values, which of the four models should be adopted? Explain.

Use the model you adopt to answer all remaining questions on the assignment.

9. Interpret the proportional reduction in the Level-1 variance component for your adopted model.

10. Interpret the proportional reduction in the Level-2 intercept variance component for your adopted model.

11. Interpret the proportional reduction in the Level-2 slope variance component for your adopted model.

12. Interpret the squared correlation between the observed and fitted popularity values for your adopted model.

Examination of Model Assumptions

13. Create density plots of the level-1 and level-2 residuals from your adopted model to help you evaluate the assumptions of normality for the mixed-effects model.

14. In 1–2 sentences comment on whether you believe these assumptions have been satisfied.

15. Create scatterplots of the level-1 and level-2 residuals versus the fitted values from your adopted model to help you evaluate the assumptions of linearity and homogeneity of variance for the mixed-effects model.

16. In 1–2 sentences comment on whether you believe these assumptions have been satisfied.

Plotting the Model

17. Create a plot that displays the fixed and random effects for the model. Show the effect of popularity for a range of extraversion controlling for teacher experience. Show the sex predictor by facetting. Give you plot an appropriate caption. **(2pts)**