Assignment 4

EPsy 8282

Fall 2017

Please submit your responses to each question in a printed document. Also, 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** and **labels**.
- Any typed mathematics (equations, matrices, vectors, etc.) should be appropriately typeset within the document using Equation Editor, Markdown, or LaTeX.

This assignment is worth 18 points. Each question is worth 1 point unless otherwise noted.

The National Longitudinal Survey of Youth (NLSY) is a large scale study that began in 1979 by the U.S. Department of Labor to examine the transition of young people into the labor force. It was carefully designed to broadly represent young people in the country. The original cohort consisted of approximately 12,500 subjects who were in the age range of 14–21 in 1979.

For this assignment, you will use the file *reading.csv*. These data are a subset of NLSY data that includes longitudinal data on students' reading achievement The assessment used was the Peabody Individual Achievement Test (PIAT) Reading Recognition subtest, which measures word recognition and pronunciation ability. Scores range from 0 to 84, with higher scores indicating higher levels of reading achievement. The data consists of six variables, which are:

- `id`: Student ID number
- `male`: Dummy coded sex variable (0 = female, 1 = male)
- `read_1986`: Reading score in 1986
- `read_1988`: Reading score in 1988
- `read_1990`: Reading score in 1990
- `read_1992`: Reading score in 1992

Preparation

Before you begin the analysis, you will need to tidy the data into the long format and also create a new variable indicating the year centered on 1986. In other words, the year value for 1986 will be 0, that for 1988 will be 2, etc. Use this variable as the time predictor in the analyses for this assignment.

Description

1. Create a plot of the (1) individual profiles and (2) observed mean change profile over time. Facet this plot on sex.
2. What does the plot suggest about the structural form for the fixed-effects structure of the longitudinal growth pattern? Does the longitudinal change in reading achievement seem linear? Nonlinear? Explain.
3. What does the plot suggest about including potential random effects? Based on your response to Question 2, which random effects might be included? Explain.
4. What does the plot suggest about whether sex should be included in the model as a covariate? Explain.

Linear-Mixed Effects Models

Fit the following models using unrestricted maximum likelihood estimation and assuming an independent error structure:

- **Model 1:** $\text{read}_{ij} = \beta_0 + [b_{0i} + \epsilon_{ij}]$
- **Model 2:** $\text{read}_{ij} = \beta_0 + \beta_1(\text{year}_{ij}) + [b_{0i} + \epsilon_{ij}]$
- **Model 3:** $\text{read}_{ij} = \beta_0 + \beta_1(\text{year}_{ij}) + \beta_2(\text{year}^2_{ij}) + [b_{0i} + \epsilon_{ij}]$
- **Model 4:** $\text{read}_{ij} = \beta_0 + \beta_1(\text{year}_{ij}) + \beta_2(\text{year}^2_{ij}) + \left[b_{0i} + b_{1i}(\text{year}_{ij}) + \epsilon_{ij}\right]$
- **Model 5:** $\text{read}_{ij} = \beta_0 + \beta_1(\text{year}_{ij}) + \beta_2(\text{year}^2_{ij}) + \left[b_{0i} + b_{1i}(\text{year}_{ij}) + b_{2i}(\text{year}^2_{ij}) + \epsilon_{ij}\right]$
- **Model 6:** $\text{read}_{ij} = \beta_0 + \beta_1(\text{year}_{ij}) + \beta_2(\text{year}^2_{ij}) + \beta_3(\text{male}_{ij}) + \left[b_{0i} + b_{1i}(\text{year}_{ij}) + b_{2i}(\text{year}^2_{ij}) + \epsilon_{ij}\right]$

5. Create a table of model results. In this table, report the model parameter estimates, the number of total parameters estimated in the model, and the AICc value for the model. If a parameter is not estimated in a model, leave that associated cell blank in the table. Use the table on the next page as a guide for creating your table. **(3pts)**
6. Based on the AICc value, and using Models 1, 2, and 3, which structure will you adopt for the fixed-effects? Explain.
7. Based on the AICc value, and using Models 3, 4, and 5, which random-effects will you adopt? Explain.
8. Based on the AICc values, is there an effect of sex? Explain.
9. Interpret the following estimated fixed-effects for Model 6 using the context of the data (i.e., relate the numerical values of the estimated model parameters back to the substantive context for this data set). (Note: Interpret them, even if they are not statistically significant.)
   - Intercept
   - Quadratic effect of year
   - Effect of sex
10. Create a plot of the predicted change curve for both males and females based on the parameters estimated in Model 6.
11. Explain how the predicted change curve is consistent with your interpretation of the fixed-effects in Question 9.

Table 1
Results of Fitting a Taxonomy of Linear Mixed-Effects Models Predicting Reading Achievement

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Fixed-effects						
Year						
Year^2						
Male						
Intercept						
Random-effects						
$\hat{\sigma}^2_{\text{Intercept}}$						
$\hat{\sigma}^2_{\text{Year}}$						
$\hat{\sigma}^2_{\text{Year}^2}$						
$\hat{\sigma}^2_{\epsilon}$						
Model-level summaries						
Number of parameters						
AICc						

Error Structures

For all remaining questions on this assignment, use the following model:

$$\text{read}_{ij} = \beta_0 + \beta_1(\text{year}_{ij}) + \left[b_{0i} + b_{1i}(\text{year}_{ij}) + \epsilon_{ij}\right]$$

12. Fit the model above using unrestricted maximum likelihood and an indpendent error structure. Create and include residual plots for the level-1 and level-2 residuals (random-effects). Use those plots to comment on whether or not the model assumptions are tenable.
13. Create and report the correlation matrix of the residuals for the repeated measures. Use those results to comment on whether it is plausible to assume an independent error structure.

Re-fit the same model using unrestricted maximum likelihood and each of the following error structures.

- Compound symmetry with a fixed correlation of 0.3
- First order autoregressive—AR(1)
- Unstructured (no constraint)

14. Report the estimated correlation matrix based on the error structure for each of the four models fitted: independent, compound symmetry, AR(1), and unstructured. **(2pts)**
15. Based on the AICc value, which error structure will you adopt? Explain.