Final Project

EPsy 8282

Fall 2017

This is an independent project, meaning that you need to complete it individually. You need to do the data analysis and complete the write-up of the results on your own. This being said, you are welcome to discuss the project with others to garner opinions and suggestions for improvement. Think of it like you are the solo analyst and author on a manuscript—while the work is independent, it does not need to be done in isolation. You can also consult with the instructor for additional help and clarification. Additionally, you may use your class notes, books, or any other related material you like (including the internet).

There are two primary parts to this project:

- A proposal, which is due on November 02, 2017.
- The final project report, which is due on Decemeber 13, 2017.

Early submissions on both parts are encouraged.

Goal and Outline

The goal of this project is for you to complete a data analysis from beginning (scientific background and motivation) to end (conclusions, discussion of scientific implications). Your grade will be based on how well you communicate and justify the statistical modeling and analysis principles you use and how well you explain and interpret the results in a way that readers not familiar with longitudinal data models and methods can understand. In particular, a strongly suggested outline of a good report is:

- **Introduction.** What are the objectives of, or motivation for, the work? What substantive back-ground/literature is needed for the audience to understand the work? Why specialized statistical models and methods are required for analysis of longitudinal data (so why methods familiar to the researchers are not appropriate). State clearly the scientific objectives and why they are of interest. Are there specific hypotheses to be addressed?

- **Methods.** What are the variables of interest and how were the data collected? What was the sampled population? What statistical model have you chosen, including its interpretation and how researchers' questions may be cast formally in terms of the model. What technique(s) did you use to analyze the data? Exploratory/descriptive? Modeling? Diagnostics? Summarize the data (this is often most effective when done graphically with few descriptive statistics). Provide a rationale for choosing the model, any assumptions you have made, and why these assumptions are reasonable for the researchers' situation. What method did you use to fit the model?

- **Results.** What are the main results of your analyses? Describe these using the language and in context of the substantive area of interest. How do these results relate to your objectives and research questions?

- **Discussion.** Summarize the main points of your work. Discuss any scientific or statistical problems with the data, the analyses, or the interpretation. Mention any future research directions.

- **References.** In a format suggested by flagship journals in your research domain. References are format-ted differently in *Psychological Methods* than they might be in *Journal of Public Health*.

- **Appendix.** Any syntax that produces results cited in your report should be included in the appendix. This syntax should be commented. Do not include the output; only include syntax here. Also do not include any syntax outside of what you used to produce results referenced in the report. The appendix is for my information only; your report should not refer to the appendix (i.e., the report should not ask readers to go to pages of code or output. Any results that the audience needs to see should be provided in the body of your report.

Audience

When you write-up the project, you need to write for a *non-statistically minded* audience—not for me. You can assume that, although the audience is familiar with basic statistical models and methods (e.g., linear regression), they know virtually nothing about longitudinal data models or methods. Thus, your report should "tell the story" of what you did, why you did it, and what your conclusions are at a level that this audience (or your advisors) can understand.

For example, the non-technical audience to whom you are writing for will not be interested in seeing gory matrix expressions and oddles of equations and formulae, primarily because they will not understand them. Thus, you will need to communicate the above information to them mainly in words, with very few equations and symbols (be sure to define clearly any symbols you do use). They will also likely know little-to-nothing about R procedures you might use or how to interpret their output. Don't include these in the report itself.

It might be useful to think of yourself as a statistical consultant on a longitudinal data analytic research project. You have been called in to conduct an appropriate analysis of the data from a study in order to address relevant research questions. Your job is to conduct the analyses and write the report for the substantive experts—describing what you did, why you did it, and what inferences may be drawn regarding the questions of interest—helping them to understand the analysis enough so that it is useful in interpreting the results, but not using "overly-technical" jargon.

Project Proposal

You will need to submit a project proposal by November 02, 2017. This proposal should be no more than two pages in length. In this proposal, you will include the research question(s) that you propose to answer in your project, a brief substantive rationale for examining those questions (why are they important, etc.), and information about the dataset(s) you plan to utilize to answer those questions. Within this dataset, identify the outcome variable(s) you plan to use in the analysis, along with any predictors and covariates. If you have thoughts or ideas about initial models you may fit to answer the research questions, include those as well.

Details in this proposal may change as you explore the data and carry out your analyses, so you aren't locked-in to the proposal. Rather, it is to give me a more systematic idea of what you are thinking about for the project. It will also give you a basis for writing the *Introduction* and *Methods* sections of the project.

Grading

The project report is worth a total of 40 points (and 40% of your final course grade). Note that the proposal is not graded. The following guidelines will be used to grade your project:

- **Technical Detail (15 points).** Is the analysis problem or research focus of the report formulated clearly using appropriate terminology? Are the appropriate statistical tools used properly? Are the assumptions checked when possible? Are the research or analysis conclusions justified? Are the analysis interpretations and implications described in plain English (no statistical jargon)?

- **Thoroughness (10 points).** Is there enough substantive background provided so that a scientist not experienced in your field can understand the general ideas of your study, and the implications of your study for your population of interest? Was anything obvious overlooked, for example, a too simple procedure performed as a substitute for a more appropriate procedure, an assumption not checked, conclusions not stated for each objective? Is the work of others appropriately cited, and the references for those cited materials included?

- **Clarity (10 points).** Is the report easy to understand? Does it follow a logical structure, both within and across sections? Is it concise and clearly written?

- **Innovation (5 points).** Were the data or was the science particularly challenging or complex? Was any extra information provided which added to the quality of the report (EDA, diagnostics)? Was the report exceptionally well done, smoothly flowing with easily understandable presentation of information? *Note: These are points to be explicitly earned; the default score for this category is 0.*