Final Data Analysis Project
Analysis of Longitudinal Data of Your Own Choosing

Abstract See syllabus for due date. This is Tuesday week 7 for 2022. See below for guidelines for the abstract.

Meeting You must meet with Weiss about the final project before week 9 or during week 9. Note: I am not available Thursdays, nor May 12/13 pf 2022.

Final Project is due finals week, on Monday, by 11:59pm. Please submit through CCLE.

ABSTRACT: Should be typed 1 paragraph. Describe the data, your tentative model(s), and purpose of the analysis. Please supply the following information about your project:

1. Data set name.
2. Data set source.
3. Outcome(s) name(s).
4. Predictor(s) name(s).
5. Definition of time, spacing of observations, number of repeated measures, amount of missingness, any related topics.
6. Sample size information (people, observations).
7. Anticipated/possible issues/difficulties with the analysis.

MEETING: Expect the meeting to be about 15 minutes. In addition to the information discussed for the abstract, bring *profile plots* and *empirical summary plots*, and *output from one reasonable model fit* to the data.

DATA SET: You will need to find a data set to analyze. Identify an appropriate data set and purpose for analyzing it early in the quarter. Places to search for data sets are a) collecting your own data; b) Your current or previous work; c) Colleague of your employer; d) Occasionally your advisor may have a suggestion; e) Doctoral thesis or master's paper data set; f) Consulting client. Getting a 'real' data set (for example from NCHS) as opposed to a book/web data set is strongly encouraged. If you get a set from the web or from a book you will need to do additional research to create a purpose for the analysis. You may NOT take a data set from the notes, the book, my web pages or from a previous stat or biostat course (for example not from Biostat 200ABC), nor the UCI Machine Learning Repository. Do not use Kaggle data.

If you would like to create a data set, it is possible using census data on states/counties/cities for example, but be careful on choices of covariates, as time varying covariates typically require bivariate longitudinal analysis. The US federal government (NCHS, CDC, NCES) offer up interesting and important data sets for analysis (LSOA, NHANES, NHIS, NNHS, ECLS-B, ECLS-K) many of which are longitudinal.

General requirements for your report

- Please keep your report to 3 typed double spaced pages, pdf.
- Double spacing and good margins are important to leave space for comments.
- Figures and Tables
  - Put all figures and tables in an appendix at the end of your report.
  - The appendix does not count towards the three pages.
  - Label all figures and tables with an appropriate and complete caption.
  - Number figures and tables separately and sequentially. The first figure (table) mentioned in the text is figure (table) 1, second mentioned is figure (table) 2, and so on.
  - Do not put additional text in the appendix.
  - You must refer (however briefly) to *all* figures and tables in the text.
  - Table captions: Define all columns/rows/entries carefully and accurately.
  - Figure captions: Define the data plotted, all symbols/lines/colors in the caption.
- Supply sufficient output to justify all claims.
- In your main text, explain what conclusions you take from each table and figure.