Psychology 613: Multivariate Statistics (Data Analysis III)
Final

The following final exam is due at **5pm** on **Thursday, June 9th**, via Canvas. **All problems must be completed individually from start to finish.**

The document that you turn in should consist of a write-up with ***only the relevant pieces of output*** and a copy of your code (numbered to correspond with the questions). Questions that ask you to interpret results of analyses should be phrased in terms that are substantively meaningful.

This exam will be graded **out of 50 points**. So, **please choose your own combination of problems to add up to 50 possible points**. Extra credit **up to 60 total points** is allowed.

Problems 1 relies on the dataset called **StudentRatings.sav** on Canvas. It contains a number of variables regarding student ratings of instructors /courses (e.g., instructor preparedness, instructor confidence), as well as a number of demographic variables about the instructors (e.g., salary, years of experience) and students (e.g., GPA, student rank). This dataset also has some information about students' political traditionalism and policy preferences. All of the relevant variables have descriptive labels in the SAV file. This dataset was graciously loaned to me by Jim Sidanius and was part of research he conducted at the University of Texas at Austin.

Logistic regression (30 points; 10 points for each part A-C)

1. I have a problem. I want to maximize the number of students who think they will be getting an "A" (because those students give the best ratings), but I don't have enough time to do it. Please help me!

   A) I think helping my students get an "A" (versus any other grade) is a function of three things: focusing my lectures, using clear and relevant examples, and being accessible to students outside of class (all continuous / numeric). Controlling for student GPA (also numeric), compute a logistic regression predicting the likelihood of getting an "A" (vs. "not-an-A") based on those three factors. Do those three teacher factors predict the likelihood of an "A" above and beyond GPA? What about each of them individually?

   B) Based on your findings from Part A, re-run the model with only the one best predictor of "get an A". Generate a plot the probability of getting an A (on the y-axis) as a function of that one best predictor (on the x-axis).

C) Based on your findings from Part A, tell me how much the probability of getting an "A" will increase for a student with an *average* GPA* if I improve by one unit in each of the three domains (give me separate probability changes in each domain, controlling for all other domains). How much would an average-GPA student improve if I put two units of effort into focusing my lectures, but no effort into the other two domains. What is your recommendation? (Assume that I start with an average score on each of the domains.)

*(it may be helpful to center this)

Network Analysis (30 points; 10 points for each part A-C).
**Credit to Cory Costello for this problem*

2. For this question, you'll be working with *real data* available on osf (https://osf.io/zypgf/). These data come from a study on cohort differences in Narcissism by Wetzel et al. (2017). The data I've provided have been cleaned so they contain only responses to the Narcissistic Personality Inventory (NPI): **npi_data_clean.csv**. In addition to the data, the folder I've shared also has a word document with the text of the items: **NPI_w_key.docx** (from Del Paulhus's website); this might be helpful in making sense of what you find.

   A) Produce a correlation network graph of NPI items. Choose a threshold that you can justify and justify it. *Hint*: "sig" is probably easy to justify, but there are other good ones.

   B) Get and interpret Measure(s) of Centrality for the NPI correlation network. What are some highly central items? What is your interpretation of that result?

   C) Assess the extent to which the NPI could be characterized as a smallworld. What do you make of the output you generate?

Multilevel modeling (30 points = 20 points for #3 and 10 points for #4)

3. I would like to understand the factors that contribute to student grades in 613, so I gathered data for the past three years. There have been three different instructors in that time, so my data are structured as students nested within instructors. I don't care about the specific instructors in my sample; *I am interested to generalize to the population of all instructors, not in these three instructors in particular*. I hypothesize that a student's grade in 613 is a function of her/his overall GPA and the number of hours she/he spends on homework per week. Further, across academic quarters, I think that average grades can differ randomly between instructors, that the effect of overall GPA will be moderated by the instructor's years of experience, and that the effect of hours per homework will be moderated by the instructor's average past ratings.

   Please write down the L1 and L2 equations to test my hypothesis and also the single-equation form of the model [don't actually run the model!]. Be sure to indicate what each predictor is, what the meaning of its parameter is (i.e., the "b" or "g" attached to each predictor), and what the variance of each error term is and what that variance represents.

4. In a study on the efficacy of the Drug Abuse Resistance Education (DARE) program, researchers used a MLM to examine students nested within schools that did or did not receive the DARE intervention. Their final model for marijuana use was:

   $\text{Marijuana\_use}_{ij} = b_{0j} + b_{1j}*\text{Pre-DARE\_marijuana\_use} + e_{ij}$,
   where:
   $b_{0j} = 0.033 - 0.044*\text{DARE}_j + u_{0j}$
   $b_{1j} = 0.098 + u_{1j}$

   In the equations, the DV is marijuana use, and the predictors are pre-DARE marijuana use and the DARE intervention (DARE = 0 if the school did not receive the intervention, and DARE = 1 if it did). The only parameters that were significantly different from zero were 0.098, $var(u_{0j})$, and $var(e_{ij})$. Label each of the parameters with the appropriate "g", explain the meaning of each, and interpret the results as you would in a real paper.

5. **Have a great summer!! (0 points)**