Psychology 613: Multivariate Statistics (Data Analysis III)
Problem Set 5

The following problem set is due at **5pm** on **Thursday, June 2nd**, on Canvas.

For these questions we'll be using data from the Adolescents Training and Learning to Avoid Steroids (ATLAS) project. Collection of these data was supported by National Institute on Drug Abuse grant R01- DA07356. Dr. Linn Goldberg and Dr. David MacKinnon have given us permission to use this data in homework exercises for this class (note: we do NOT have permission to use this data for any other purpose, e.g., publication).

The ATLAS program is a team-based intervention designed to decrease steroid use among high school football players. Thirty-one teams from different high schools in the northwest U.S. were randomly assigned to intervention or control conditions. The seven-week intervention consisted of both classroom and weight training sessions. Although the ultimate object of the ATLAS program was to reduce steroid use, a more immediate aim was to increase the adoption of healthy alternatives to steroids such as strength training. Therefore, levels of strength training self-efficacy were measured prior to and immediately following the intervention period.

The **ATLAS.sav** file is an SPSS file which contains the following variables:
SCHOOL – school ID
GRADE – student grade (9-12)
ID – individual student ID
INTERV – indicates whether a school was assigned to control (0) or intervention (1) conditions
STRTRN – individual report of strength training self-efficacy at pretest
CSTRTRN – individual report of strength training self-efficacy at immediate posttest
USE – dichotomous measure of individual steroid use at pretest (1=yes, 0=no)
CCHTL – individual perception of coach tolerance of steroid use at pretest
PROAS – individual report of number of reasons for using steroids at pretest
ESTEEM – individual report of self-esteem at pretest
CESTEEM – individual report of self-esteem at posttest

In general, these variables are scored such that higher numbers indicate more of the named construct. There is some missing data, indicated by the '.' symbol. Use all available data in the analysis for each question. There are a total of 1226 observations in the file, sorted by ID.

The document that you turn in should consist of a write-up, a copy of your code (numbered to correspond with the questions), and a copy of your output (also numbered). Questions that ask you to interpret results of analyses should be phrased in terms that are substantively meaningful. For example, instead of saying something like "the predictor was significant," you should phrase interpretations in terms of the measured constructs, e.g., "homework was significantly positively related to strength training self-efficacy."

Multilevel modeling (***Problems 3, 4, 5, and 6 required; all others optional***)

1. Use R to estimate a disaggregated model predicting post-test strength training self-efficacy from the pretest measure of this variable. What do the results of this analysis suggest? What is the problem with conducting the analysis in this manner?

2. Estimate an aggregate model predicting mean posttest strength training self-efficacy from mean pretest strength training self-efficacy. What do the results of this analysis suggest? Is this analysis consistent with the disaggregated analysis?

3. Conduct an OLS regression within each school (i.e., the first step of intercepts- and slopes-as-outcomes). Predict posttest strength training self-efficacy from pretest strength training. Plot the lines on a single set of axes (using R, EXCEL, or by hand). Does it look like the ANCOVA assumption (homogeneity of regression) would be met?

4. Finish the intercepts- and slopes-as-outcomes analysis: use the group-level intervention variable to (separately) predict intercepts and slopes. Summarize the results of the analysis.

5. Compute the ICC of posttest strength training self-efficacy. In your estimation, is an ICC of this size likely to be problematic? Is most of the variance between or within groups?

6. Write down the first- and second-level equations for a random coefficient model in which posttest strength training self-efficacy is predicted by pretest strength training self-efficacy, intervention, and their interaction. What is the meaning of each parameter? (By "parameter" here, I mean to include both fixed and random effects.)

7. Examine the model from Problem 6. Suppose the slope of pretest strength training self-efficacy was found not to vary significantly across schools (i.e., that the slope was not really a random coefficient). How would this change the model? How would it change your interpretation of the data?

8. Run the model from Problem 6 (with a random intercept and random pretest slope) in R using lmer(). What are the parameters? How do the results compare to the results from the intercepts- and slopes-as-outcomes model?