PSYC 5143
Spring '24
Problem Set #7

This is due on Wednesday, March 25 at 2pm, via Blackboard, preferably as an R file. #2 is about things we will talk about on Monday after break. Ask for help when needed. **PLEASE!**

1) This problem makes use of the data in ancova.csv. Subjects are each randomly assigned to one of two treatments (*Method*). Before assignment to the groups, each takes a pretest (*Pretest*) that is related to the DV. After experiencing each treatment, the DV (*Posttest*) is measured.

   a. Find the mean of the DV for each of the two groups.

   b. Perform an ANOVA (i.e., ignore the covariate) however you'd like to. Report the *F*-ratio (and its *df*s) that tests for a significant difference between the two treatments. What would you conclude based on the model you've fit?

   c. Perform an ANCOVA using mean-centered pretest scores as the covariate. Report the *F*-ratio (and its *df*s) that tests for a significant difference between the two treatments after controlling for pretest scores.

   d. Report the *y*-intercept and partial slopes for the analysis in part c. State what each tells you.

   e. Write out the regression equation for the full model (i.e., with both the centered covariate and the grouping variable included). Plug the following into the regression equation, calculate and report $\hat{Y}$ for each, and then in the plainest language you can, describe what these predicted values mean:

      i. the mean of the covariate (i.e., 0, because it's been mean-centered), and the lambda value for group 1

      ii. the mean of the covariate and the lambda value for group 2

   f. Check homogeneity of slopes and state whether this assumption is met or not, along with whatever statistics you can muster in support of your statement.

   g. Check whether the covariate is reasonably independent of the group variable in these data.

2) Using the data linked here, please answer the questions below. In this file, there are the following variables:

   - Y = the outcome that's being modeled
   - group = a grouping variable with three levels
   - X1 & X2 = orthogonal contrast codes that represent group membership
   - Z = a good covariate
   - Zbad = a bad covariate

   a. Find group means for both the outcome variable, Z, and Zbad.

   b. Model the outcome using X1, X2, and the Z variable as predictors. **Don't** mean-center the latter.

   c. Model the outcome using X1, X2, and the Zbad variable as predictors. **Don't** mean-center the latter.

   d. Using the parameter estimates from the models in parts b and c, which represent regression equations for making predictions, plug in the lambdas for the groups (i.e., the values of X1 and X2 that correspond to each group) and the overall mean of Z and Zbad, respectively, to generate so-called adjusted means for each group.

   e. Now using the parameter estimates from the model in part c, again plug in the lambdas for each group and – instead of using the overall mean of Zbad – the mean of Zbad for each group to generate predicted scores. (So you'll use a different Zbad value for each group.)

   f. Which is closer to the group means for the outcome: the predicted scores from part d or the predicted scores from part e? Try if you can to state a lesson about using group means vs the overall mean of a covariate to make predictions in a case like this (i.e., where the groups differ substantially with respect to the covariate).