

Name _____

Psychology 5068
Hierarchical Linear Models
Homework 1
Due February 5, 2018

For this assignment, you will use the High School and Beyond data (HSB.csv) to perform some basic R data manipulation, analyses, and displays. These represent a sampling of the kinds of things we will need to do with our data in this class. This is meant more as a warm-up exercise than a comprehensive exposure to R. Some of these things are easy to accomplish; others are a bit more challenging. For the latter, Google is your friend. I've yet to come across an R-related question that didn't yield fairly quickly to a Google search.

We introduced the HSB data in class but focused on just three variables (math achievement, SES, and sector). Here is a bit more information about the file contents. There 7185 cases (students) in the data set. They come from 160 different schools. The following four variables were measured at the student level (Level 1):

- *minority*, a dummy variable for student ethnicity (1 = minority, 0 = other)
- *female*, a dummy variable for student sex (1 = female, 0 = male)
- *ses*, a composite scale constructed from variables measuring parental education, occupation, and income
- *mathach*, a measure of mathematics achievement

The following six variables were measured at the school level (Level 2):

- *size* (school enrollment)
- *sector*, a dummy variable for type of school (1 = Catholic, 0 = public)
- *pracad*, the proportion of students in the academic track
- *disclim*, a scale measuring disciplinary climate
- *himnty*, a dummy variable indicating minority enrollment (1 = more than 40% minority enrollment, 0 = less than 40%)
- *meanses*, the mean of the SES values for the students in a particular school who are included in the level-1 data

1. Read in the HSB.csv file and save it in a dataframe called HSB_Data, excluding the variables, *pracad* and *disclim*. **Verify you have done this correctly by printing the first several lines of the dataframe.**
2. Produce basic descriptive information for just these two variables: *mathach* and *ses*, **using a single command.**
3. What is the overall correlation between *mathach* and *ses*?
4. Produce a cross-classification table for *female* and *minority*. Make sure that the rows and columns of the table have appropriate labels (not just numbers).
5. Are the two variables in Question 4 independent of each other?

6. Produce a histogram for *ses*. Include a blue vertical line indicating the mean, a red vertical line indicating the median, and the normal density curve (in green). Make sure the axes are appropriately labeled. Do the data seem to be normally distributed?
7. Produce a Q-Q plot for *ses*. Does this change your opinion regarding normality?
8. Produce a scatterplot of *meanse*s (y axis) versus *size* (x axis). Make sure the axes are appropriately labeled. Add the best-fitting linear regression line as well as a loess (nonlinear) fit line. Color the plot symbols so that public schools are red and Catholic schools are blue. What does this figure tell you?
9. Plot the best-fitting line relating *mathach* to *ses* and include the 99% confidence interval around the line.
10. Produce a two-panel plot. In the upper panel, show the boxplots of *mathach* separately for each public school. In the lower panel, show the boxplots of *mathach* separately for each Catholic school.