**Homework Week 7**

**PSC 204B**

**Part I: Regression Inferences**

1. The probability of a child being identified as having an attention-deficit/hyperactivity disorder (ADHD) is approximately 7.5% in the United States. Of those identified with ADHD, there is a 1/2 chance of having the inattention subtype only (of which there is an approximate 68.4% chance of being male), a 1/4 chance of having the hyperactivity/impulsivity subtype (of which there is an approximate 63.2% chance of being male), and a 1/4 chance of having the combined subtype (for which there is an approximate 82.1% chance of being male). Simulate a probability model to determine the number of males out of 10,000 with each subtype of ADHD. Repeat your simulation 1,000 times.
2. Dementia affects approximately 1 in 14 people over 65 years of age. For a normal individual without dementia, cortical thickness is approximately normally distributed with mean equal to 2.5mm and standard deviation equal to .7. For an individual with dementia cortical thickness is also normally distributed, but with an average of only 1.25mm and standard deviation of .4. Simulate a probability model to determine the distribution of cortical thickness in the population of people over 65 years of age, drawing samples of 100 each for each simulation. Repeat your simulation 1,000 times.

BONUS: Use the 'swiss' dataset and estimate the effect of Agriculture on Fertility using a linear model. Then, obtain the confidence intervals with confint(fittedmodel) and bootstrap the CI around the slope and compare the results to the classic CI. After that, bootstrap the confidence interval around the AIC of the model.

**Part II: Multilevel Linear Models**

The High School and Beyond study describes the activities of seniors and sophomores as they progressed through high school, postsecondary education, and into the workplace. The data we are using is from the baseline year of 1980, in which students are nested within schools. The dataset you will use for the following questions is an SPSS dataset, hsb.sav.

You can read this dataset into R by setting your working directory to where the data file is stored and using the following code:

library(foreign)

hsb = read.spss(‘hsb.sav’, to.data.frame = TRUE)

The variables in this dataset include an identification number for school (school), minority status (minority: 0 = not minority, 1 = minority), sex (female: female = 1, male = 0), student-level socioeconomic status (ses), student’s mathematics achievement score (mathach), size of the school (size), private school status (sector: 1 = private, 0 = public), proportion of students in academic track (pracad), disciplinary climate (disclim), minority enrollment (himinty), and school-level mean socioeconomic status (meanses).

1. Select 50 students at random from the larger dataset and plot their math achievement scores against student level socioeconomic status (SES).
2. Compute and interpret the intraclass correlation coefficient (ICC) for student’s math achievement.
3. Write out the multilevel equations for the null model, a random intercept model using SES as a predictor, and a random slopes model using SES as predictor.
4. Fit the null model, random intercept model using SES as a predictor, and random slopes model using SES as a predictor. Use AIC values to determine the best model.
5. Re-run your selected model twice—once using grand-mean centered values of SES, and another using group-mean centered values of SES. Compare the results from these two models.
6. Interpret the parameter estimates from the group-mean centered analysis you conducted in 5.

BONUS: Choose a school-level predictor within the hsb dataset to add to your model. Use this predictor to test an interaction with the student-level predictor of SES. Interpret your results.