STAT/BIOSTAT 571: Homework 2

To be handed in on Weds January 20th, in class. Please see 'Chapter 0' of the slides for a summary of how to answer questions appropriately, and the guidelines from 570. Where solutions require use of `R`, summarize your findings in a written answer, and append your **annotated** code, to show what you did. For each question, write up your solution on your own, using **full sentences**.

1. **[EEs with vector outcomes – optional]**

   (a) The class site contains a 'wide' format version of the education data, where each row contains the outcomes from each class. Using it, implement your own version of the bootstrap inference seen in class (e.g. slide 2.12) and check that your answers agree with that analysis, up to Monte Carlo error.

   (b) For this analysis, resampling clusters, approximately how much Monte Carlo error is there in the estimation of β using $B = 1,000$ resamples? How about $B = 10,000$?

2. **[EEs with vector outcomes]** On the class site you will find data on 30 patients with leprosy. The data are from a clinical trial, with outcomes (counts of leprosy bacilli) measured at two times. No-one is treated prior to the first measurement, but subjects receive one of three different treatments prior to the second measurement. Interest lies in the following four parameters;

   - θ_0 At baseline, the population average counts of leprosy bacilli
   - θ_1 Among those on placebo, the ratio of average count at the end of the study period to the average count at baseline
   - θ_2 At the end of the study period, the ratio of the average count for those on treatment A to the average count for those on placebo
   - θ_3 At the end of the study period, the ratio of the average count for those on treatment B to the average count for those on placebo

   (a) Give a set of estimating equations which provides consistent estimates of all four parameters, or some transformation of them; mild regularity conditions may be assumed implicitly.

   (b) Solve these estimating equations for the data, and briefly describe — in full sentences — how you computed this solution. (Hint: a simpler answer in a) makes this part easier)

   (c) Use the nonparametric bootstrap to calculate approximate 95% intervals for each of the four univariate parameters of interest.

   (d) Calculate sandwich estimates of the standard error of your four univariate estimates, and use them to construct approximate 95% intervals for each of the four univariate parameters of interest. Compare your answers with those in c).