Due December 2.

For these questions you should hand in a solution (numbers and/or words and/or graphics) and, separately, the R code used to create it. Don’t just hand in the transcript of an R session.

1. In comparing two groups where the strong null hypothesis happens to be true, the distribution of bootstrap samples will satisfy the strong null hypothesis. Why can’t we test the hypothesis by comparing the observed difference to the bootstrap distribution?

2. For some permutation tests the permutation distribution can be computed analytically. One such test replaces the data by ranks and uses the difference in the mean of the ranks between the two groups as the test statistic, comparing this difference to its distribution under permutation. This is the Wilcoxon rank sum test, implemented by the R function `wilcox.test`. In the randomized trial subset of the PBC data, compare serum bilirubin between treatments using `wilcox.test`, and verify that a permutation test gives the same p-value.

3. Show that if two groups differ only in scale (e.g., N(0,1) and N(0,2)) and have different sample sizes that the Wilcoxon test will give a p-value less than 0.05 more or less than 5% of the time.

Some things to think about:
- When comparing means, the weak null hypothesis is that the means are the same. Is there a natural “weak null hypothesis” for the Wilcoxon test?
- For ordinal categorical data the Wilcoxon test gives the same result no matter what number are assigned to the categories (e.g., 0,1,2,3; 1,10,100,1000). What are the advantages and disadvantages of this invariance?