Homework 5: The Bootstrap and Jackknife

Instructions: Be sure to electronically submit your answers as an R file. You may work together and discuss the problems with your classmates, but write up your final answers entirely on your own.

1. Write an R function to run a bootstrap resampling for a given data set and statistic. You should use the "matrix trick" demonstrated in class to avoid for loops in this function. Your function should look like this:

   `myboot = function(x, stat, B)`

   The parameters are `x`, a vector (or matrix) of data; `stat`, a function that computes a statistic from a vector of data and an index set; and `B`, the number of bootstrap resamples. Your function should return a list with four elements: `t` the vector of the `B` statistics from resampled data; `t0`, the statistic of the original data; `bias` and `se`, the bootstrap estimates of bias and standard error. (Hint: your function should operate very similarly to the function `boot` in R, and you may want to compare that you get similar results for the same input.)

2. Write an R function to run a jackknife resampling for a given data set and statistic. Your function should look like this:

   `myjackknife = function(x, stat)`

   The parameters are the same as above, and the output should also be a list with the same four elements. Notice in this case `B` is not needed, and the `stat` function does not need to take an index set.

3. This problem uses the data set `ToothGrowth`, which is included with the standard R installation (call `?ToothGrowth` to read about the data). Using your functions from 1 and 2, perform the following on the "vitamin C" subset of data (`ToothGrowth[supp == 'VC']`):

   (a) Run a bootstrap analysis on the correlation between dose and tooth length using 10,000 resamples. Plot a histogram of the bootstrap correlation statistics. Print the bias and standard error. Compute the 95% confidence interval using the bootstrap percentile method.

   (b) Run a jackknife analysis on the correlation between dose and tooth length. Plot a histogram of the leave-one-out statistics. Print the bias and standard error. Do these numbers roughly agree with your bootstrap results in (a)?

   (c) Run a bootstrap analysis on the median of tooth length using 10,000 resamples. Generate the same output as in (a).

   (d) Run a jackknife analysis on the median of tooth length. Plot a histogram of the leave-one-out statistics. Print the bias and standard error. How exactly did this fail?

   (e) Perform a jackknife-after-bootstrap analysis for the correlation example in (a). Print the standard error of the bootstrap estimates of bias and standard error. (Hint: it is ok to copy some code from the example in the book for this.)