MATH 185 – Homework 2
Due Monday, 04/18/2016, by 11:59 PM

Send your code to `math185ucsd@gmail.com`. *Follow the following format exactly. For Homework 1, in subject line write "MATH 185 (HW 1)" and nothing else in the body. There should only be one file attached, named* `hw1-lastname-firstname.R`. *Make sure your code is clean, commented and running. Keep your code simple, using packages only if really necessary. If your code does not run, include an explanation of what is going on.*

Problem 1. (A test for symmetry) Consider the setting where we have a numerical sample $X_1, \ldots, X_n$ that is assumed iid from a given distribution. We want to test the distribution is symmetric about 0. The following is a well-known test based on random sign flipping. For a sign vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n) \in \{-1, +1\}^n$, compute the sample mean of $\varepsilon_1 X_1, \ldots, \varepsilon_n X_n$, denoted $Y_\varepsilon = \frac{1}{n}\sum_i \varepsilon_i X_i$. Let $Y_\star$ denote the original sample mean. The one-sided version of the test returns the p-value

$$P = 2^{-n}\#\{\varepsilon : Y_\varepsilon \geq Y_\star\}. \tag{1}$$

A. Make sense of this in your mind. To probe your understanding, explain why the factor 2^{-n} is there.

B. In practice there are too many sign vectors to compute the p-value exactly and one has to resort to Monte Carlo sampling. Write a function `flipSignTest1(x, B=999)` which implements this. `x` is the vector of observations and `B` is the number of MC iterations.

C. How about a two-sided test? To answer this question, write a function `flipSignTest2(x, B=999)` which implements the two-sided version. [This should be a simple modification of your previous function.]

Problem 2. (Father and son's heights) Consider the `father.son` dataset in the `UsingR` package. [This package accompanies the book *Using R for Introductory Statistics*.] Suppose we want to compare the father and son's heights. Particularly, say we want to test whether a son tends to be taller than his father.

A. Produce a plot that would help answer this question (at least in an informal way).

B. Denote the data $(X_i, Y_i), i = 1, \ldots, n$, where X is the father's height and Y the son's height. Let $Z_i = Y_i - X_i$. There are several ways to formalize the question into a test of hypotheses. One of them is to test for symmetry of the distribution of Z about 0. Use your function `flipSignTest1` to perform such a test. Offer some brief comments. [This test is in fact a permutation test. Can you see why?]

Problem 3. (Bootstrap confidence interval) Write code to compute a *one-sided* 90% bootstrap pivotal confidence interval of the mean difference in height between son and father. How would that help answer the question in Problem 2? What population are we drawing an inference about?