Homework 9
36-325/725
due Friday Nov 9

(1) Get the data on eruption times and waiting times between eruptions of the old faithful geyser.

```
data(faithful)
names(faithful)
help(faithful)
x <- faithful$waiting
y <- faithful$eruptions
plot(table(x),xlab="waiting times")
plot(table(y),xlab="duration")
```

Make sure you understand what these commands do. What does the table command do? Why am I using it?

Estimate the mean waiting time and give a standard error for the estimate. Also, give a 90 per cent confidence interval for the mean waiting time. (Note: 90 per cent, not 95 per cent!) Now estimate the median waiting time. Give a standard error and 90 per cent confidence interval. Use the bootstrap to get the standard error. For the confidence interval, try two approaches: the Normal interval and the percentile interval.

(2) 100 people are given a standard antibiotic to treat an infection and another 100 are given a new antibiotic. In the first group, 90 people recover; in the second group, 85 people recover. Let p_1 be the probability of recovery under the standard treatment and let p_2 be the probability of recovery under the new treatment. We are interested in estimating $\theta = p_1 - p_2$. Provide an estimate, standard error, an 80 per cent confidence interval and a 95 per cent confidence interval for θ. (Hint: look at the mouse data example.)

(3) In 1975, an experiment was conducted to see if cloud seeding produced rainfall. 26 clouds were seeded with silver nitrate and 26 were not. The decision to seed or not was made at random. You can get the data from

http://lib.stat.cmu.edu/DASL/Stories/CloudSeeding.html

Let θ be the difference in the median precipitation from the two groups. Estimate θ. Estimate the standard error of the estimate and produce a 95 per cent confidence interval using the bootstrap.

(4) Let $X_1, \ldots, X_n$ be distinct observations (no ties). Let $X_1^*, \ldots, X_n^*$ denote a bootstrap sample and let $\overline{X}_n^* = n^{-1} \sum_{i=1}^n X_i^*$. Find: $\mathrm{E}(\overline{X}_n^* | X_1, \ldots, X_n)$, $\mathrm{Var}(\overline{X}_n^* | X_1, \ldots, X_n)$, $\mathrm{E}(\overline{X}_n^*)$ and $\mathrm{var}(\overline{X}_n^*)$.

(5) In this question, we will investigate how accurate the bootstrap estimate of the standard error is when the sample size is small. Let $X_1, \ldots, X_n \sim N(\theta, 1)$. The plug-in estimate of θ is $\hat{\theta}_n = n^{-1} \sum_{i=1}^n X_i$. Find an exact expression for the standard error of $\hat{\theta}_n$. Call this se. Now take $\theta = 0$, $n = 5$, and conduct a simulation. Each time you simulate a data set, estimate the standard error using the bootstrap. To keep this manageable, use $B = 200$ in the bootstrap and use 100 simulations. You will thus get 100 bootstrap standard errors $\hat{se}_1, \ldots, \hat{se}_{100}$. To assess the accuracy of the bootstrap, compute the mean of $|\hat{se}_1 - se|/se, \ldots, |\hat{se}_{100} - se|/se$. What is your conclusion about the accuracy of the bootstrap standard error? Repeat for $n = 50$.