**conductma@bsc.ccop**

[**concomchair@bsc.coop**](mailto:concomchair@bsc.coop)

**cc victor**

**criteria to be considered clear and present danger**

**Statistics 135 Lab #1**

**Part 1: Simple Random Sampling/Bootstrap, due Tuesday Feb 13, 11:55pm**

This assignment will use the infants data in the file KaiserBabies.rda, which is one of the files in bcourses. An example of how to load it into R is in the lecture code from January 19.

1. Take a simple random sample of 10 observations (births) using the two lines of code below. The function set.seed makes it so that everyone will be using the same sample.

set.seed(7)

mysample=sample(na.omit(infants$wt),10)

* 1. Use the sample average to estimate the average weight of the mothers, calculate the estimated standard error of these estimates and form a 95% confidence interval for the average of the population (assuming normality works).
  2. Repeat 1000 times (without using the set.seed function) to get 1000 different confidence intervals. How many of them do you expect to cover the true average? How many do? Note that in practice you would be unable to do this since you only get one sample.
  3. Calculate the SD of the sample averages. Is it close to the estimated standard error from a)? Make a histogram of the sample averages to see if it seems plausible that the probability histogram for the sample average follows the normal curve pretty closely. Make a quantile-quantile plot to further investigate. Does it seem like the confidence interval is valid?

1. Start with your original sample and use it to construct a bootstrap population. You can use any code from the lecture code from Feb 2, but you’ll need to understand how it works. Don’t worry about the fact that since the population size is not an integer multiple of the sample size, the bootstrap population won’t be exactly the same size as the original population. This part of the lab WOULD be possible to do in practice.
   1. Using that bootstrap population, get 1000 simple random samples of size 10. For each, get the sample average and make a histogram of these sample averages. Put a vertical line through the average of the bootstrap population. Calculate the SD of the sample averages. Is it close to the estimated SE from 1a) above?
   2. Construct a 95% bootstrap confidence interval by taking the 2.5 percentile and the 97.5 percentile of the bootstrap sample averages. How does it compare to the confidence interval you got in 1a)?
2. Repeat all parts of 1) and 2) with a sample size of 100 instead of 10.