Introduction to the Bootstrap

STA 325: Lab 6, Fall 2018

Today's agenda: Introduction to the Bootstrap

Programming partner's: You should have a programming partner for each lab, and you should switch off who is programming, and use each other for help. We will spend about 30–50 minutes per week on lab exercises and you will be expected to bring you laptops to class to work on these exercises in class. Myself and the TA's will be in class to help you.

In this lab, we'll investigate the bootstrap using linear regression.

Let's first read in the dataset `bootstrap-data.csv`

```
my.data <- read.csv('data/bootstrap-data.csv', header=TRUE)
head(my.data)
```

```
##   temperature pressure
## 1    298.0000 7767.519
## 2    298.0505 7744.627
## 3    298.1010 7762.836
## 4    298.1515 7756.751
## 5    298.2020 7760.371
## 6    298.2525 7753.842
```

1. Perform uni-variate and bi-variate data analysis on the variables in the model. Your response variable is Pressure and your input variable is Temperature. Fit a linear regression model and create three diagnostic plots: an independence plot, a QQ-plot, and a plot of residuals vs. temperature. Additionally produce a table of coefficient estimates with xtable.
2. Write a function that takes as its argument a data frame and returns a resampled 'synthetic' dataset of equal dimension.
3. Write a function that takes as its input a string model specification and a dataframe and produces a fitted model object from lm() with a resampled data frame using the function you have written above.
4. Write a function that takes as its inputs a data frame, a string model and a natural number representing how many bootstrap replications to run and returns the coefficients from each replication using the function you have written above.
5. Write a function that uses all three of the functions you have written above, and takes four arguments: a count of replications, type I error rate, a data frame, and a string model to produce a boot strapped confidence interval for the regression coefficients. Run this function and provide (using xtable) the confidence interval for your regression coefficients. What do you conclude?