R Homework Number 9

Fisher Ankney

November 13th, 2018

Statistics 5193

*Note: this document was created using R Markdown.

Question 1a

```r
library(readxl)
StudentData <- read_excel("~/Documents/data_science/r_stat_5193/data/StudentData.xlsx")
```

```r
boot.mean <- function(my.data, trim.percent = 0.05) {
  dat.sample <- sample(my.data, replace = T)
  mean(dat.sample, trim = trim.percent)
}
```

Question 1b

```r
# the use of attach is not recommended due to the convoluted nature of the
# function, it reduces reproducibility, therefore I will reference the
# dataframe in the suggested code

set.seed(1)
boot.mean(StudentData$HSClass)
```

```
## [1] 309
```

```r
boot.mean(StudentData$HSClass, trim.percent = .10)
```

```
## [1] 344
```

Question 2a

```r
my.boot.sample <- vector('double', length = 10000)

set.seed(1)
for (i in 1:10000) {
  my.boot.sample[i] <- boot.mean(StudentData$HSClass)
}

mean(my.boot.sample)
```
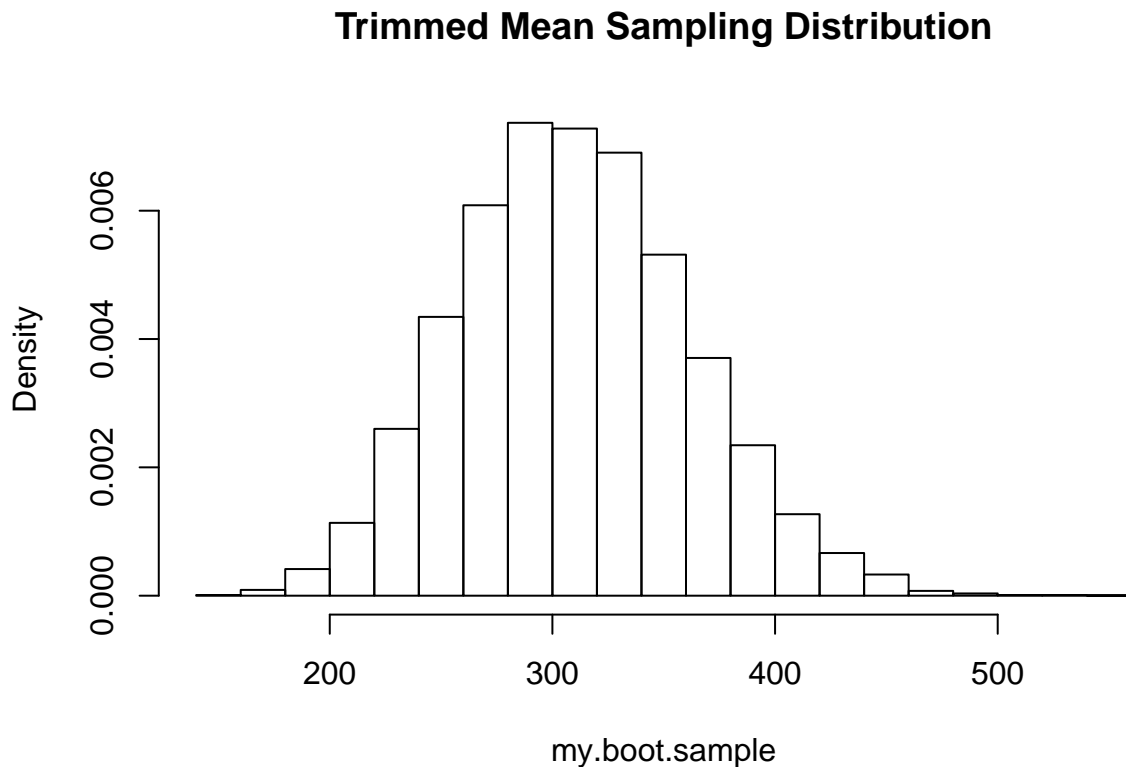
```
## [1] 310.1323
```

```r
sd(my.boot.sample)
```

```
## [1] 51.98851
```
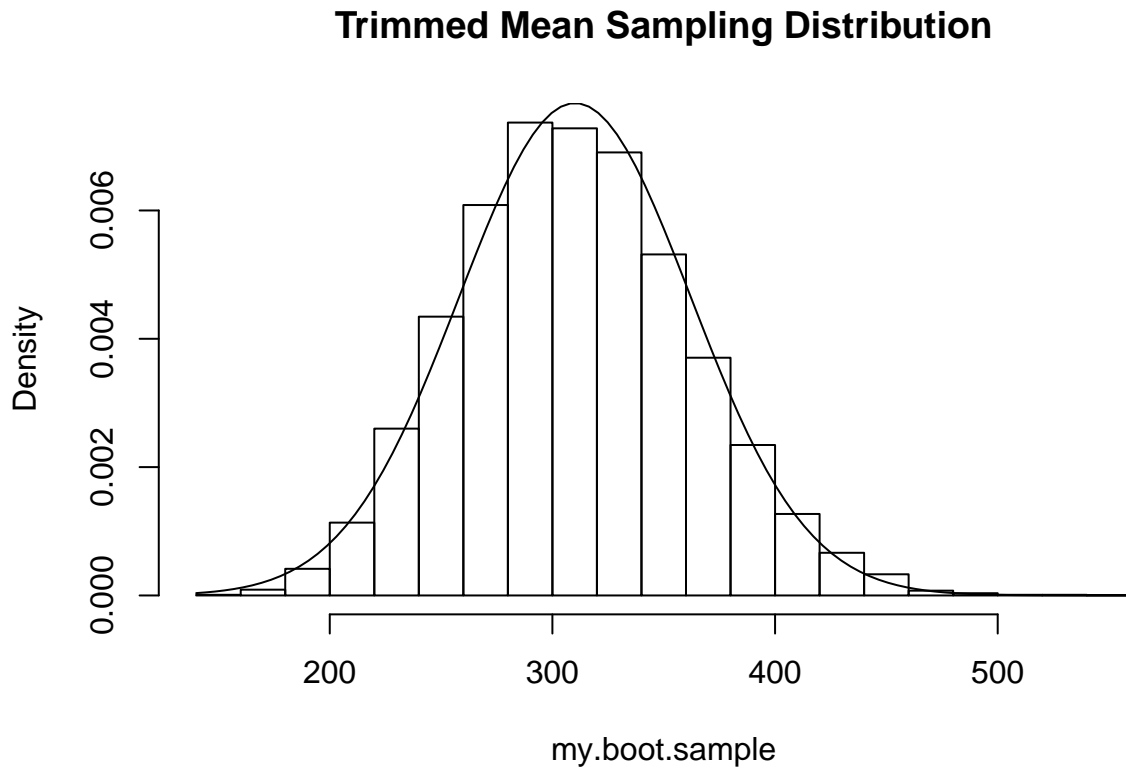
Question 2b

```r
hist(my.boot.sample,
     probability = T,
     main = "Trimmed Mean Sampling Distribution")
```


2

Question 2c

```r
hist(my.boot.sample,
     probability = T,
     main = "Trimmed Mean Sampling Distribution")

curve(dnorm(x, mean = mean(my.boot.sample), sd = sd(my.boot.sample)),
      add = T)
```

Trimmed Mean Sampling Distribution

Question 2d

```r
quantile(my.boot.sample, c(0.025, 0.975))
```

```
##     2.5%    97.5%
## 214.9068 417.5470
```

Question 3a

```r
boot.ci <- function(my.data, trim.percent = 0.05, plot.it = T) {

# recreate boot.mean function
  boot.mean <- function(my.data, trim.percent = trim.percent) {
    dat.sample <- sample(my.data, replace = T)
    mean(dat.sample, trim = trim.percent)
  }

# create empty vector to store results
  my.boot.sample <- vector('double', length = 10000)

# loop
  for (i in 1:10000) {
    my.boot.sample[i] <- boot.mean(my.data, trim.percent = trim.percent)
  }

# save mean and sd for plotting
data.mean <- mean(my.boot.sample)
data.sd <- sd(my.boot.sample)

# return confidence interval
  print(quantile(my.boot.sample, c(0.025, 0.975)))

# conditional plot.it
  if (plot.it == T) {
    hist(my.boot.sample,
         probability = T,
         main = "Trimmed Mean Sampling Distribution")

    curve(dnorm(x, mean = data.mean, sd = data.sd), add = T)

  }
}
```

Question 3b

```r
# again, I'd rather specify the dataset instead of use attach()

set.seed(1)
boot.ci(StudentData$HSClass)
```
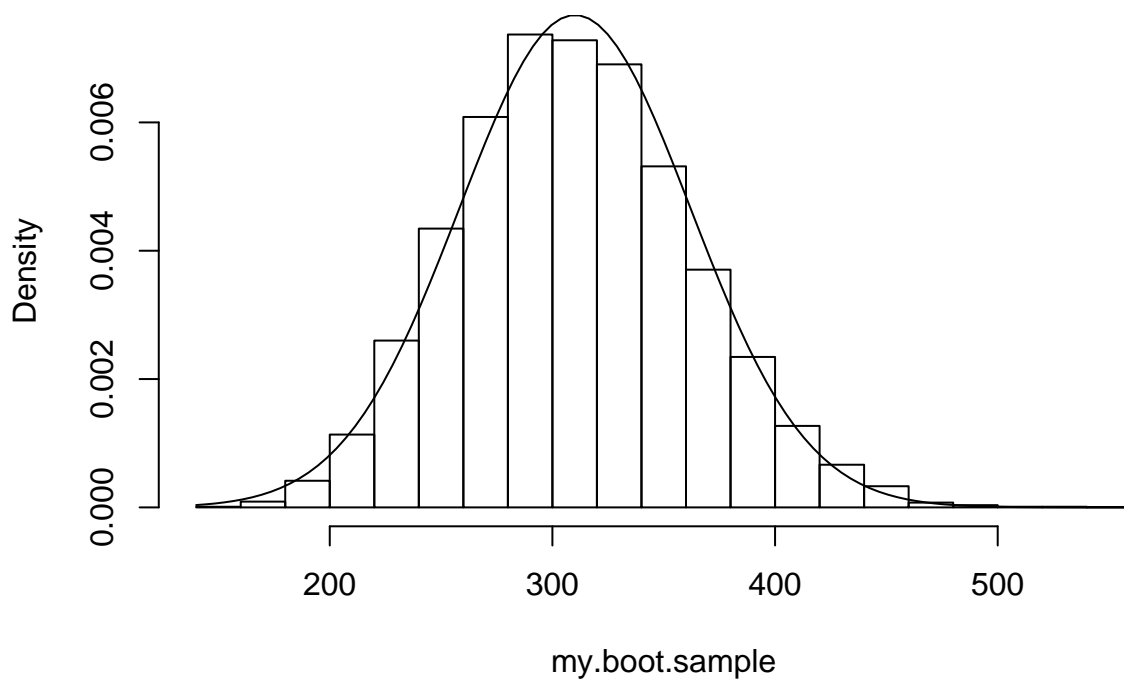
```
##     2.5%    97.5%
## 214.9068 417.5470
```

Trimmed Mean Sampling Distribution

my.boot.sample

Question 3c

```
# again, I'd rather specify the dataset instead of use attach()

set.seed(1)
boot.ci(StudentData$HSClass, trim.percent=.25)
```

```
##      2.5%     97.5%
## 152.2092 382.6329
```

```
axis(side=1, at=c(214.9068,417.5470 ),
     labels=F, col='red')
```


Trimmed Mean Sampling Distribution