

# Optional Homework: Bootstrap Practice Problems

*Fall 2016, Math 107, Prof. Adam Loy*

*Due Wednesday, October 26 at the beginning of class*

## Instructions

Since I cannot guarantee that I will be able to return an assignment before the Midterm II (which is on 10/28), I am making this an optional homework assignment. I strongly encourage everyone to work on these problems, since the bootstrap will be on the exam, but if you do not submit this assignment it will not adversely impact your grade.

If you choose to submit the assignment, it will replace your lowest homework grade, regardless of whether your grade on this assignment is higher. When calculating your final homework grade, I will still drop the lowest two grades (in terms of percentages).

Additionally, if you choose to submit this assignment, you should complete it in this R markdown file. When you are done, please knit the document to Word and print a hard copy for submission. Make sure to change the **author**: so that you get credit for your work!

This assignment requires the use of the **CarletonStats** package, which I load below.

```
library(CarletonStats)
```

## Problem 1

The General Social Survey<sup>1</sup> (GSS) is a sociological survey used to collect data on demographic characteristics and attitudes of residents of the United States. In 2010, the survey collected responses from 1,154 US residents. The survey is conducted face-to-face with an in-person interview of a randomly selected sample of adults. One of the questions on the survey is “After an average work day, about how many hours do you have to relax or pursue activities that you enjoy?” A 95% confidence interval from the 2010 GSS survey is 3.53 to 3.83 hours.

### Part a

Provide an interpretation of this interval within the context of the problem.

*Write your answer here, replacing this text.*

### Part b

In this context, what does it mean to be 95% confident?

*Write your answer here, replacing this text.*

### Part c

Suppose the researchers think a 90% confidence level would be more appropriate for this interval. Will this new interval be narrower or wider than the 95% confidence interval?

*Write your answer here, replacing this text.*

---

<sup>1</sup>National Opinion Research Center, General Social Survey, 2010.

### Part d

If a new survey asking the same questions was to be done with 500 Americans, would the standard error of the estimate most likely be larger, smaller, or about the same?

*Write your answer here, replacing this text.*

## Problem 2

A poll conducted in 2013 found that 52% (383) of 736 randomly selected U.S. adult Twitter users get at least some news on Twitter.<sup>2</sup> The data are available in the data folder of this assignment and can be loaded using the following code:

```
twitternews <- read.csv("data/twitternews.csv")
```

### Part a

Describe how to construct a bootstrap distribution. You may find it easiest to describe this using some tactile simulation method.

*Write your answer here, replacing this text.*

### Part b

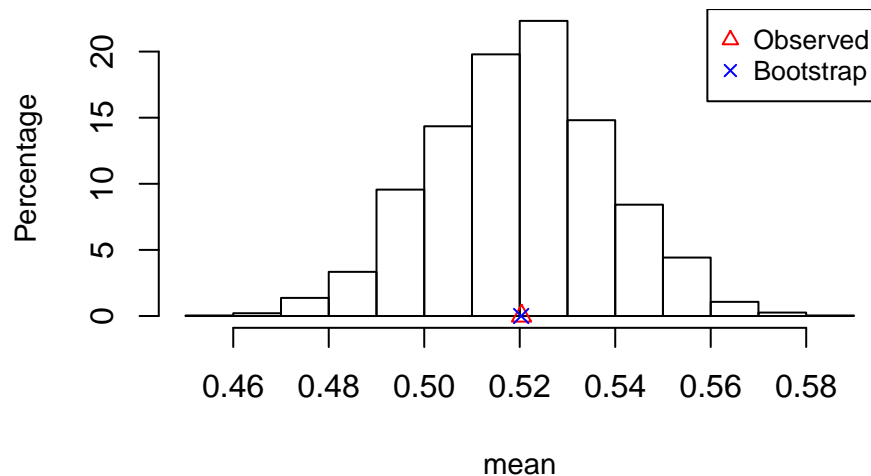
Run the below code to create a bootstrap distribution of the sample proportion consisting of 10,000 simulations.

```
boot(as.numeric(twitternews$news.on.twitter == "yes"), B = 10000)

##
## ** Bootstrap interval for mean **
##
## Observed as.numeric(twitternews$news.on.twitter == "yes") : 0.52038
## Mean of bootstrap distribution: 0.5203
## Standard error of bootstrap distribution: 0.01845
##
## Bootstrap percentile interval
##      2.5%      97.5%
## 0.4836957 0.5557065
##
##      *-----*
```

---

<sup>2</sup>Twitter News Consumers: Young, Mobile and Educated. November 4, 2013. Pew Research.



### Part c

Report the bootstrap standard error of the sample proportion.

*Write your answer here, replacing this text.*

### Part d

Using the plug-in method, calculate a 95% bootstrap confidence interval for the proportion of U.S. adult Twitter users get at least some news on Twitter.

*Write your answer here, replacing this text.*

### Part e

Provide an interpretation of your interval from the previous part within the context of the problem.

*Write your answer here, replacing this text.*

## Problem 3

The data file `CommuteStLouis.csv` contains a random sample of commute times (in minutes) for 500 workers in St. Louis, Missouri. The data are available in the data folder of this assignment and can be loaded using the following code:

```
commute <- read.csv("data/CommuteStLouis.csv")
```

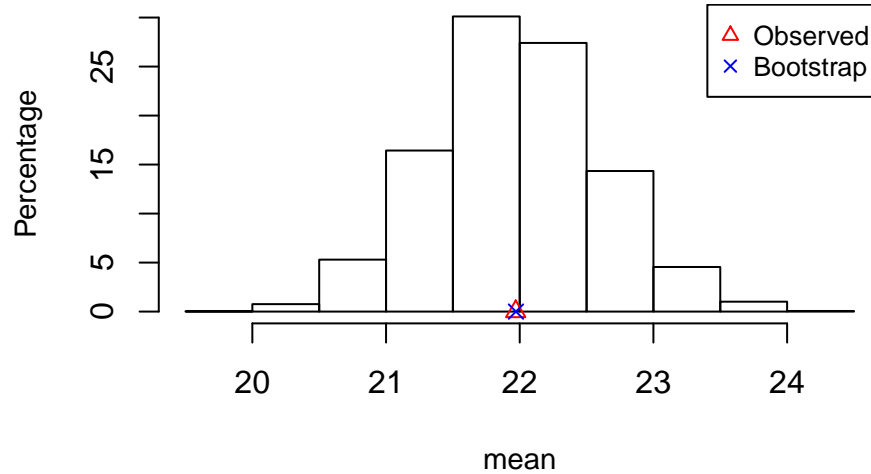
### Part a

Run the below code to create a bootstrap distribution of the sample mean consisting of 10,000 simulations.

```
boot(commute$Time, B = 10000)
```

```
##
## ** Bootstrap interval for mean **
##
## Observed commute$Time : 21.97
## Mean of bootstrap distribution: 21.97166
## Standard error of bootstrap distribution: 0.63279
```

```
##
## Bootstrap percentile interval
## 2.5% 97.5%
## 20.758 23.272
##
## *-----*
```



### Part b

Report the bootstrap standard error of the sample mean.

*Write your answer here, replacing this text.*

### Part c

Using the plug-in method, calculate a 95% bootstrap confidence interval for the average commute time in St. Louis.

*Write your answer here, replacing this text.*

### Part d

Provide an interpretation of your interval from the previous part within the context of the problem.

*Write your answer here, replacing this text.*