Homework2\_Chapter5

Bethany Smith

January 25, 2018

# Problem 9

## We will now consider the Boston housing data set, from the MASS library

library("MASS")

## Warning: package 'MASS' was built under R version 3.4.3

library("tidyverse")

## -- Attaching packages --------------------------------------- tidyverse 1.2.0 --

## v ggplot2 2.2.1 v purrr 0.2.4  
## v tibble 1.3.4 v dplyr 0.7.4  
## v tidyr 0.7.2 v stringr 1.2.0  
## v readr 1.1.1 v forcats 0.2.0

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()  
## x dplyr::select() masks MASS::select()

attach(Boston)   
library("boot")

## Warning: package 'boot' was built under R version 3.4.3

library("stringr")

## (a) Based on this data set, provide an estimate for the population mean of “medv”. Call this estimate .

mu.hat <- as.table(mean(medv))  
row.names(mu.hat)<- c("Estimate of Population Mean of medv")  
  
mu.hat

## Estimate of Population Mean of medv   
## 22.53281

### *Answer : The estimate of the population mean of medv is 22.53.*

## (b) Provide an estimate of the standard error of . Interpret this result.

se.hat <- as.table((sd(Boston$medv))/(sqrt(length(Boston$medv))))  
row.names(se.hat)<- c("Estimate of Standard Error of population mean of medv")  
se.hat

## Estimate of Standard Error of population mean of medv   
## 0.4088611

### *Answer: The estimate of the standard error for is 0.4088611 rounded to .41.*

## (c) Now estimate the standard error of using the bootstrap. How does this compare to your answer from (b) ?

set.seed(1)  
n = length(Boston$medv)  
B = 1000  
result = rep(NA, B)  
for (i in 1:B) {  
boot.sample = sample(n, replace = TRUE)  
result[i] = mean(sd(Boston$medv[boot.sample])/(sqrt(n)))  
}  
q<-with(Boston, mean(result))  
q <- round(q,2)  
q

## [1] 0.41

### *Answer: The estimate of the standard error of using the bootstrap method is 0.41. This is the same rounded estimate of standard error that we received in answer b. This indicates that the estimate in b can be trusted as a valid estimate of the standard error for .*

## (d) Based on your bootstrap estimate from (c), provide a 95% confidence interval for the mean of “medv”. Compare it to the results obtained using t.test(Boston$medv).

n = length(Boston$medv)  
B = 1000  
result = rep(NA, B)  
for (i in 1:B) {  
boot.sample = sample(n, replace = TRUE)  
result[i] = mean(Boston$medv[boot.sample])  
}  
with(Boston, mean(medv) + c(-1, 1) \* 2 \* sd(result))

## [1] 21.69246 23.37316

t.test(Boston$medv)

##   
## One Sample t-test  
##   
## data: Boston$medv  
## t = 55.111, df = 505, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 21.72953 23.33608  
## sample estimates:  
## mean of x   
## 22.53281

### *Answer: We see from the results above that the 95% confidence interval for the bootsrap estimate of the mean of medv that the lower bound is 21.69 and the upper bound is 23.37. Furthermore when calculating the mean and a 95% confidence interval using a t-test we get a mean of 22.53 (the same mean estimate obtained in question a) and a lower bound CI value of 21.73 and an upper bound CI value of 23.34. Thus, regardless of if we calculate a 95% confidence interval using the bootstrap mean or using a one sample t-test we arrive at nearly the same 95% confidence interval.*

## (e) Based on this data set, provide an estimate, , for the median value of “medv” in the population.

answerE <- as.table(median(Boston$medv))  
row.names(answerE)<- c("Median estimate of medv")  
answerE

## Median estimate of medv   
## 21.2

### *Answer: We can see from the calculation above that the median value of medv for the population, , is 21.2.*

## (f) We now would like to estimate the standard error of . Unfortunately, there is no simple formula for computing the standard error of the median. Instead, estimate the standard error of the median using the bootstrap. Comment on your findings.

n = length(Boston$medv)  
B = 1000  
result = rep(NA, B)  
for (i in 1:B) {  
boot.sample = sample(n, replace = TRUE)  
result[i] = median(Boston$medv[boot.sample])  
}  
G<-with(Boston, result)  
G<-as.table(sd(G))  
row.names(G) <- c("Standard Error of median estimate of medv")  
G

## Standard Error of median estimate of medv   
## 0.3729808

### *Answer: We know from our class time and text book that one of the key advantages of using a bootstrap is that for typical statistics that are difficult to calculate such as the standard error associated with a median estiamte, the boostrap can approximate this hard to calculate value. Thus, from the results above we see that the estimated standard error associated with the median value of is 0.37.*

## (g) Based on this data set, provide an estimate for the tenth percentile of “medv” in Boston suburbs. Call this quantity .

mu.hat.01 <- as.table(quantile(Boston$medv,0.1))  
row.names(mu.hat.01) <- c("Tenth Percentile estimate of medv")  
mu.hat.01

## Tenth Percentile estimate of medv   
## 12.75

### *Answer: The estimated tenth percentile of medv in Boston suburbs, , is 12.75.*

## (h) Use the bootstrap to estimate the standard error of . Comment on your findings.

n = length(Boston$medv)  
B = 1000  
result = rep(NA, B)  
for (i in 1:B) {  
boot.sample = sample(n, replace = TRUE)  
result[i] = quantile(Boston$medv[boot.sample],0.1)  
}  
H<-with(Boston, result)  
H <- as.table(sd(H))  
row.names (H) <- c("Standard Error of 10th quantile estimate")  
H

## Standard Error of 10th quantile estimate   
## 0.5004648

### *Answer: We see from the output above that the standard error for the tenth quantile estimate of medv in Boston suburbs, , is 0.50.*