Wolfe\_Mod1\_HW

Tyler Wolfe

8/30/2020

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

summary(cars)

## speed dist   
## Min. : 4.0 Min. : 2.00   
## 1st Qu.:12.0 1st Qu.: 26.00   
## Median :15.0 Median : 36.00   
## Mean :15.4 Mean : 42.98   
## 3rd Qu.:19.0 3rd Qu.: 56.00   
## Max. :25.0 Max. :120.00

## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

#Problem 1

library(ISLR) dim(College) help(College)

set.seed(2020)  
collegeApps <- College$Apps mean(collegeApps)

set.seed(2020) boots <- NULL for (i in 1:10000) { meanCollegeApps <- mean(sample(collegeApps,1000,replace = TRUE)) boots<- c(boots,meanCollegeApps) } hist(boots, main = “Sample Distribution for Mean number of College Apps”)

quantile(boots,c(.1,.9))

mean(boots)

is.matrix(College) is.data.frame(College)

## College is a dataframe, which reads like a ;list but prints like a matrix. I will be using a list (option3)

mylist<-boots

sapply(mylist, mean)

quantile(mylist,c(.1,.9))

##Problem 2

set.seed(0830) boots <- NULL for (i in 1:10000) { meanCollegeApps <- mean(sample(collegeApps,1000,replace = TRUE)) boots<- c(boots,meanCollegeApps) } mylist<-boots

# CI = 80% with new seed

quantile(mylist,c(.1,.9))

#third seed (3190)

set.seed(3190) boots <- NULL for (i in 1:10000) { meanCollegeApps <- mean(sample(collegeApps,1000,replace = TRUE)) boots<-c(boots,meanCollegeApps) } mylist<-boots

#CI = 80% with third seed

quantile(mylist,c(.1,.9))

#Re-examing CI with “5000” sample for all three seeds (2020,0830,3190)

set.seed(2020) boots <- NULL for (i in 1:10000) { meanCollegeApps <- mean(sample(collegeApps,5000,replace = TRUE)) boots<- c(boots,meanCollegeApps) }

mylist2020<-boots

quantile(mylist2020,c(.1,.9))

set.seed(0830) boots <- NULL for (i in 1:10000) { meanCollegeApps <- mean(sample(collegeApps,5000,replace = TRUE)) boots<- c(boots,meanCollegeApps) }

mylist0830<-boots

quantile(mylist0830,c(.1,.9))

set.seed(3190) boots <- NULL for (i in 1:10000) { meanCollegeApps <- mean(sample(collegeApps,5000,replace = TRUE)) boots<- c(boots,meanCollegeApps) }

mylist3190<-boots

quantile(mylist3190,c(.1,.9))

#Problem 3

newSamp <- sample(CollegeApps, PubPriv=College$Private) numPubPriv

tapply(numPubPrivPubPriv, mean)

newMeans <- tapply(numPubPrivPubPriv, mean) newPubMeans <- newMeans[1] newPrivMeans <-newMeans[2] newPubMeans - newPrivMeans

set.seed(42) randomResults <- NULL for(i in 1:10000){ newSample <- sample (CollegeApps, PubPriv=newSample) newMeans <-tapply(numPubPrivPubPriv, mean) randomResults <- c(randomResults, newMeans[1]-newMeans[2]) } hist(randomResults, xlim=c(-4000,4000)) abline(v=newPubMeans - newPrivMeans, col=“red”, lwd=2)

greaterThanOrig <- sum(randomResults < newPubMeans - newPrivMeans)

sum(greaterThanOrig)/10000