Homework 8

Noah McIntire

Problem 1

```
library(ggplot2)
set.seed(07312001)
delays<-read.csv("/Users/noahmcintire/Downloads/Delta delays.csv")
data_samp<-delays$Arrival.Delay
delays_IQR<- IQR(data_samp)
B<-10000
## Draw the bootstrap samples
boot_samp <- replicate(B, sample(data_samp, replace=T))
#boot_samp[,1:5]

## Determine the sample IQR from each bootstrap sample
boot_IQR <- apply(boot_samp,2,IQR)


boot_err <- boot_IQR - delays_IQR
boot_err_sort <- sort(boot_err)
p5 <- B*0.05 #location of a
p95 <- B*0.95 #location of b
boot_ci <- delays_IQR - boot_err_sort[c(p95,p5)]
boot_ci

## [1] -3 21

boot_err_sort[c(p95,p5)]

## [1] 17.5 -6.5
```

Problem 2

a)

```
q90 <- quantile(data_samp, 0.9)
boot_samp2 <- replicate(B, sample(delays$Arrival.Delay, replace=T))
boot_q2 <- apply(boot_samp2,2,quantile, 0.9)
null = 0
boot_quant_null <- boot_q2 - mean(boot_q2) + null
pval = sum(boot_quant_null >= q90) / B
pval
```

```
## [1] 0.0693
```

Because our p value is 0.0679, there is not sufficient evidence and we fail to reject the null hypothesis, meaning there that the 90th quantile of delays is not a late arrival at Dulles on the days after thanksgiving in 2019.

b)

The conclusion in part a is disputable. Because the p value is so close to 0.05, there is a possibility that with a larger sample size (or even potentially with a different set of samples obtained while bootstrapping) that a p value under 0.05 is obtained, meaning that there would be evidence that for the 90th quantile being a late arrival, which would dispute my conclusion in part a.

Problem 3

```
Adelays<-read.csv("/Users/noahmcintire/Downloads/American delays.csv")
## the below was done by following the idea of a randomized test in ref 1
a_samp <- sample(Adelays$Arrival.Delay)
a90 <-quantile(a_samp, 0.9)
diff90 <- q90 - a90
aDelay<- Adelays$Arrival.Delay
All_delay<-append(a_samp, data_samp)
#length(All_delay)
randomTest <- function(x){
  samp <- sample(All_delay, replace = T)
  q1 <- quantile(samp[1:22], 0.9)
  q2 <- quantile(samp[23:45], 0.9)
  q1-q2
}
All_boot_diffs <- replicate(B, randomTest(1))
pval = sum(All_boot_diffs <= diff90 | All_boot_diffs >= 2*mean(All_boot_diffs)- diff90)
pval
```

```
## [1] 0.7606
```

Because our p value is not under the significance level of 0.05, we are unable to reject the null hypothesis. Therefore, there is not statistical evidence that the 90th percentile of the two data sets are different.

References

1. https://measuringu.com/randomization-test/