Homework 8

Santhosh Rajendran

Problem 1

```
setwd("/Users/santhoshrajendran/Documents/3rd Year 1st Sem HW Files/STAT 3080/")
deltaDelays <-
  read.csv("Delta delays.csv", header=TRUE)
B<-10000
samp_mean<-IQR(deltaDelays$Arrival.Delay)
boot_samp <- replicate(B, sample(deltaDelays$Arrival.Delay, replace=T))
boot_means <- apply(boot_samp,2,IQR)
means_df <- data.frame(boot_means)
boot_err <- boot_means - samp_mean
boot_err_sort <- sort(boot_err)
p5 <- B*0.05
p95 <- B*0.95
samp_mean - boot_err_sort[c(p95,p5)]
```

```
## [1] -3.5 20.5
```

Problem 2

a

The values in the previous problem indicate that there is evidence that the 90th percentile of arrival delays is a late arrival. The 90th percentile value is close to 6 minutes past the recorded arrival.

b

I don't think my answer is 100% indisputable but I am more inclined to trust it since I bootstrapped it with 10,000 repetitions. The lower bound of the IQR is also negative.

Problem 3

```
setwd("/Users/santhoshrajendran/Documents/3rd Year 1st Sem HW Files/STAT 3080/")
american=read.csv('American delays.csv',header=TRUE)
diff <- quantile(deltaDelays$Arrival.Delay, .9) -
  quantile(american$Arrival.Delay, .9)
random.test <- function(x){
  random_comb <- sample(c(deltaDelays$Arrival.Delay,american$Arrival.Delay))
  percentile1 <- quantile(random_comb[1:x], .9)
  percentile2 <- quantile(random_comb[(x+1):(length(deltaDelays$Arrival.Delay)+
                                               length(american$Arrival.Delay))]
                          , .9)
  percentile1 - percentile2
}
bootdiff <- replicate(B, random.test(length(deltaDelays$Arrival.Delay)))
sum(bootdiff <= diff |
      bootdiff >= 2*mean(bootdiff) - diff) / B
```

```
## [1] 0.8357
```

Using a p-value of 0.05, the test fails to reject the null hypothesis, indicating the two 90th percentiles are statistically indifferent.