# Homework 8

Will Scheib

## Problem 1

```
set.seed(06072000)

delta_delays <- read.csv("data/Delta delays.csv")

samp <- delta_delays %>% pull(Arrival.Delay)
samp_IQR <- samp %>% IQR()
reps <- 10000

## Draw the bootstrap samples and determine their IQRs
boot_iqrs <- replicate(reps, IQR(sample(samp, replace=T)))

## Determine the 90% bootstrap confidence interval of the population IQR
boot_err_sort <- (boot_iqrs - samp_IQR) %>% sort()
p0.5 <- reps*0.05
p99.5 <- reps*0.95
boot_ci <- samp_IQR - boot_err_sort[c(p99.5,p0.5)]
boot_ci

## [1] -3.0 20.5
```

## Problem 2

### Part a

```
## Determine the test statistic
samp_90 <- quantile(samp, .90)

boot_samp <- replicate(reps, sample(samp, replace=T))
boot_90s <- apply(boot_samp,2,quantile, .90)

## Determine the sampling distribution of the test statistic under the null hypothesis
pct90_0 <- 3
boot_90s_null <- boot_90s - mean(boot_90s) + pct90_0

## Determine the p-value
sum(boot_90s_null > samp_90)/reps
```

```
## [1] 0.0664
```

There is not evidence that the 90th percentile of arrival delay lengths on these holiday travel days is a late arrival because the p-value of 0.0664 is greater than the assumed alpha of 0.05.

### Part b

No, because differing sets of samples might generate different results, particularly because our calculated p-value was so close to 0.05.

## Problem 3

```
# Calculate second sample 95th percentile
american_delays <- read.csv("data/American delays.csv")

samp2 <- american_delays %>% pull(Arrival.Delay)
samp_90_2 <- samp2 %>% quantile(0.90)

# Determine the test statistic
samp_diff <- samp_90 - samp_90_2

# Determine the sampling distribution of the test statistic under the null hypothesis
rand.test <- function(samp, samp2){
  x <- length(samp)
  rand_comb <- c(samp, samp2) %>% sample()
  b90_1 <- rand_comb[1:x] %>% quantile(0.90)
  b90_2 <- rand_comb[(x+1):(length(samp)+length(samp2))] %>% quantile(0.90)
  b90_1 - b90_2
}

boot_diffs_null <- replicate(reps, rand.test(samp, samp2))

# Determine the p-value
sum(boot_diffs_null <= samp_diff |
      boot_diffs_null >= 2*mean(boot_diffs_null) - samp_diff)/reps
```

```
## [1] 0.8471
```

There is not evidence that the 90th percentile of arrival delay lengths is different for the two airlines because the p-value of 0.8448 is greater than the assumed alpha of 0.05.