BACS HW4

Student ID: 110065508

Due: 2022/03/13

Question 1. (a) According to the article, the desired probability is given by the R code

```
> pnorm(-3.7)
```

Which gives 0.0001077997.

(b) $2,200,000 * 0.0001077997 \approx 237$. ■

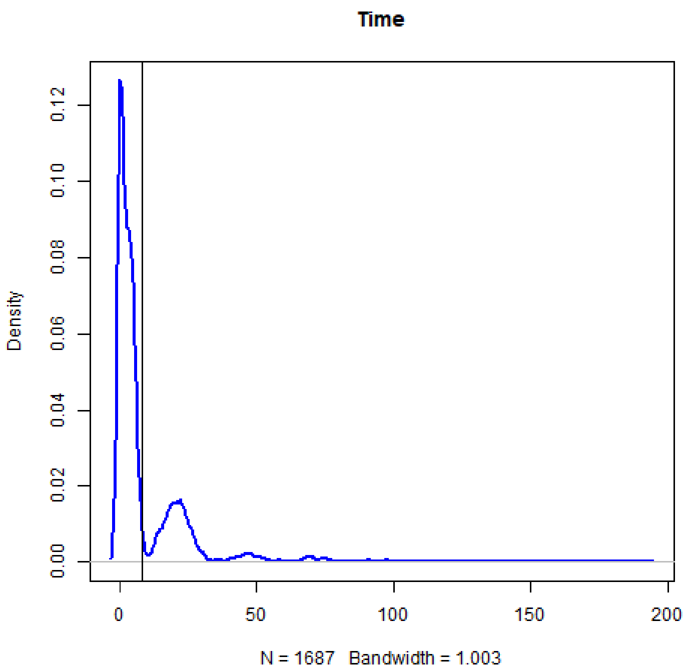
Question 2. Before solving any problems, I read the file `verizon.csv` first and attain some basic statistics.

```
# Question 2
# Read the data and get some basic statistics
library(tidyverse)
ver_time <- read_csv('verizon.csv')$Time # read file
ver_mean <- mean(ver_time) # mean of sample
ver_sd <- sd(ver_time)
ver_size <- length(ver_time)
hyp <- 7.6 # Null hypothesis
```

(a) The Null distribution of t-values:

(i) The following codes done the work:

```
# Question 2 (a)
# (i) Visualize the distribution of Verizon's repair times
png(filename = "2a.png")
plot(density(ver_time), col="blue", lwd=2, main = "Time") # plot pdf
abline(v=mean(ver_time)) # add vertical lines
dev.off()
```



(ii) The null hypothesis H_0 is, "the population mean of repair times μ is equal to 7.6 minutes". On the contrary, the alternative hypothesis H_1 is $\mu \neq 7.6$.

(iii) The sample mean $\bar{X}$ is obtained in the beginning. It is approximately 8.5220. As for the 99% CI, I use the following code:

```
# (iii) Estimate the population mean, and the 99\% CI
CI <- ver_mean + c(-2.58, 2.58)*ver_sd # CI
```

R gives $(0.005868152, 4.599558515)$.

(iv) The following codes done the work:

```r
# (iv) t-statistic and p-value
se <- ver_sd/sqrt(ver_size) # standard error
t <- (ver_mean - hyp) / se # t-statistic
df <- ver_size - 1 # degree of freedom
p <- 1-pt(t,df) # p-value
```

The t-statistic is approximately $t \approx 2.5608$ and the p-value is 0.0053.

(v) If the null hypothesis H_0 were true, we expect t-statistic from the sample near 0. However we have $t = 2.5608$ now. Since the deggre of freedom is large in this case, the Student-t distribution can be approximate by the normal distribution. That is to say, roughly less than 5% of sample t-values can out of $[-2, 2]$. Also, the p-value shows that the probability that we have a more extreme sample mean is about 0.0053.

(vi) The p-value is slightly large than 0.005, so we cannot reject the null hypothesis.

(b) Bootstrapping: Let's setup first. As for (iv), I will attached the required graph after every sub-problems.

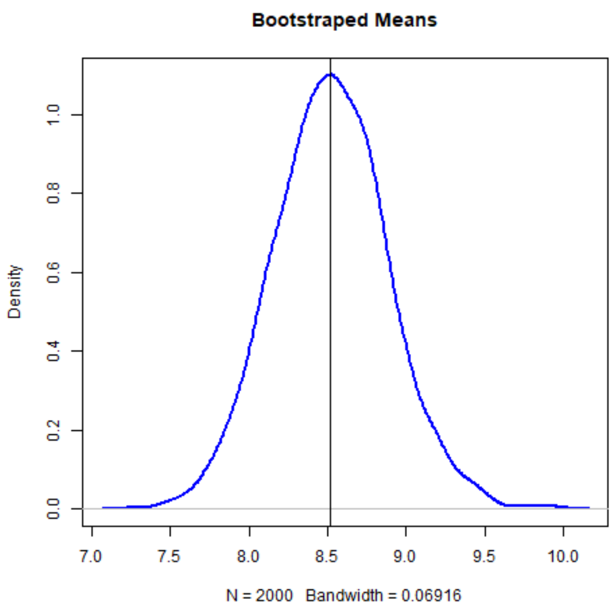
```r
# Question 2 (b)
# bootstrap settings
num_boots <- 2000 # Let's do 2000 times
set.seed(48763)
```

(i) The bootstrapped 99% CI of the mean is $[7.604745, 9.503061]$.

```r
# (i) Estimate the bootstrapped 99\% CI of the mean
sample_statistic <- function(stat_function, sample0) {
  resample <- sample(sample0, length(sample0), replace=TRUE)
  stat_function(resample)
}

boot_means <- replicate(num_boots, sample_statistic(mean, ver_time))
boot_means_ci_99 <- quantile(boot_means, probs = c(0.005, 0.995)) # 99\%CI

png(filename = "2b_1.png")
plot(density(boot_means), col="blue", lwd=2, main = "Bootstraped Means")
abline(v=mean(boot_means)) # add vertical lines on 99\% CI
dev.off()
```

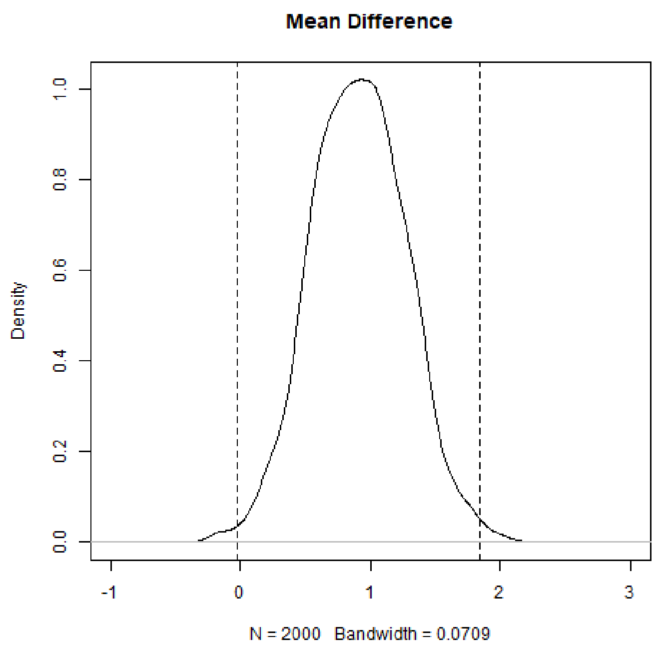


(ii) The 99% CI of the bootstrapped difference is $[-0.02958251, 1.84314182]$.

```
# (ii) Bootstrapped Difference of Means
boot_mean_diffs <- function(sample0, mean_hyp) {
  resample <- sample(sample0, length(sample0), replace=TRUE)
  return( mean(resample) - mean_hyp )
}

mean_diffs <- replicate(num_boots, boot_mean_diffs(ver_time, hyp))
diff_ci_99 <- quantile(mean_diffs, probs=c(0.005, 0.995)) # 99\% CI

png(filename = "2b_2.png")
plot(density(mean_diffs), xlim=c(-1,3), main = "Mean Difference") # plot pdf
abline(v=diff_ci_99, lty="dashed") # add vertical lines on 99\% CI
dev.off()
```

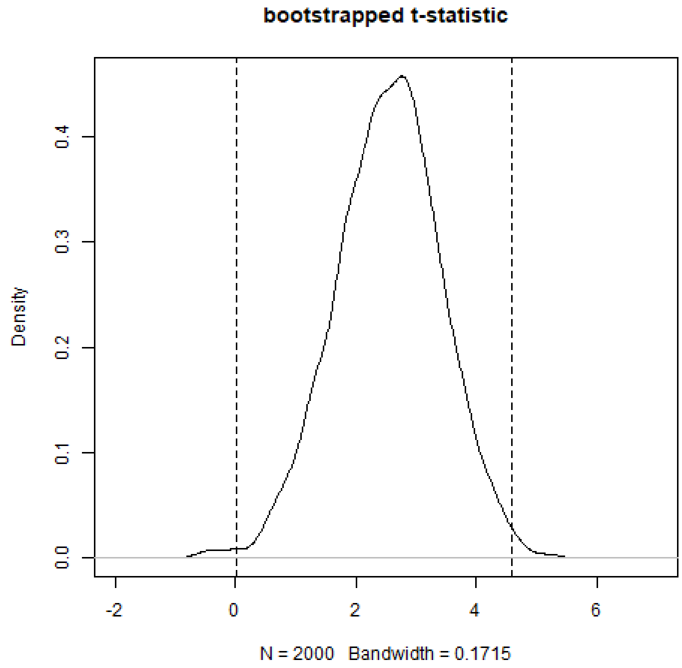


(iii) The 99% CI of the bootstrapped t-statistic is $[0.005868152, 4.599558515]$.

```
# (iii)
boot_t_stat <- function(sample0, mean_hyp) {
  resample <- sample(sample0, length(sample0), replace=TRUE)
  diff <- mean(resample) - mean_hyp
  se <- sd(resample)/sqrt(length(resample))
  return( diff / se )
}

t_boots <- replicate(num_boots, boot_t_stat(ver_time, hyp))
t_ci_99 <- quantile(t_boots, probs=c(0.005, 0.995)) # 99\% CI

png(filename = "2b_3.png")
plot(density(t_boots), xlim=c(-2,7), main="bootstrapped t-statistic") # plot pdf
abline(v=t_ci_99, lty="dashed") # add vertical lines on 99\% CI
dev.off()
```



(c) By (a), the traditional test cannot reject the null hypothesis. The bootstrapped percentile cannot reject the null hypothesis as well. However, 7.6 is really at the edge of its 99% CI. (Things will change if we use 95% CI.)

However, 0 is in the 99% CI of bootstrapped difference of means, so (b)(ii) rejects the null hypothesis, so is bootstrapped t-Interval since $t \approx 2.5608$ is in the 99% CI. ■