HW5_report

107023058

Question 1

Description: Google compute a DOI score for each app to determine whether they are malware. DOI score is binomially distributed.

(a)

Given the critical DOI score(-3.7), find the probability that a randomly chosen app from Google's app store will turn off the Verify security feature.

```
pnorm(-3.7) # those apps whose Z-score are lower than -3.7 have lower retention rate

## [1] 0.0001077997
```

(b)

Assuming there were 2.2 million apps, what number of apps on the Play Store did Google expect would maliciously turn off the Verify feature once installed

```
num_app <- 2.2*10^6
low_retention <- num_app * pnorm(-3.7)
low_retention

## [1] 237.1594
```

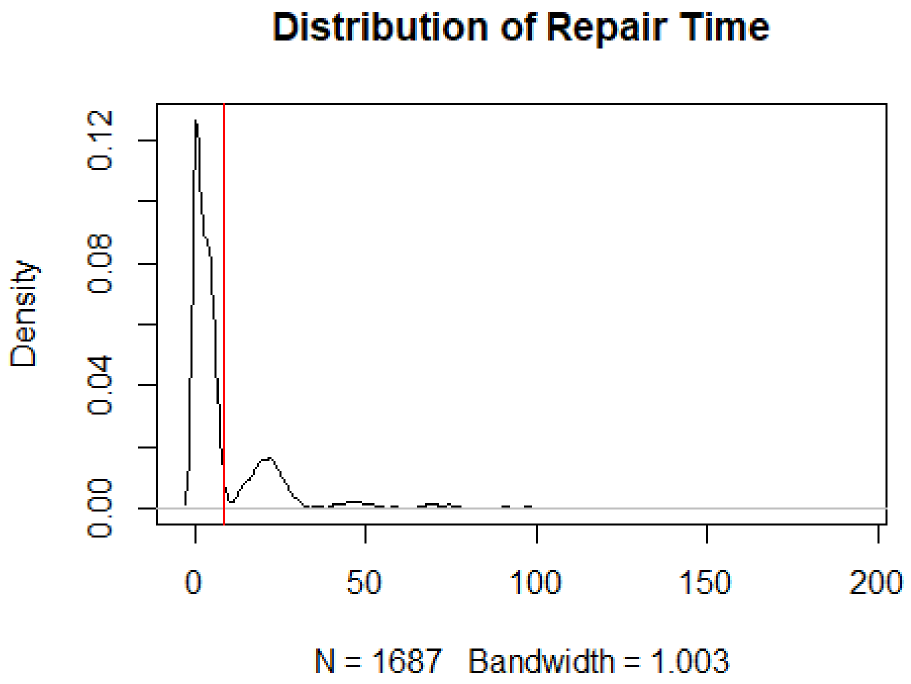
Question 2

Description: Verizon claims that they take 7.6 minutes to repair phone services for its customers on average.

(a) The Null distribution of t-values

(i) Visualize the distribution of Verizon's repair times, marking the mean with a vertical line

```
RepairTime <- read.csv("D:/Retro/NTHU/課程講義/大三/計算統計於商業分析之應用/HW4/verizon.csv",header = TRUE)
repair_time <- RepairTime$Time
plot(density(repair_time),main = 'Distribution of Repair Time')
abline(v = mean(repair_time),col = 'red')
```



(ii) Given what PUC wishes to test, how would you write the hypothesis?

h0: u = 7.6 h1:u!=7.6

(iii) Estimate the population mean, and the 99% confidence interval (CI) of this estimate

```
mean_hypo <- 7.6  # set null mean as 7.6
mean_repair_time <- mean(repair_time)
sd_repair_time <- sd(repair_time)
ci99 <- c(mean_repair_time-2.58*(sd_repair_time/sqrt(length(repair_tim
e))),mean_repair_time+2.58*(sd_repair_time/sqrt(length(repair_time))))
# compute its 99% C.I
ci99

## [1] 7.593073 9.450946
```

(iv) find the t-statistic and p-value of the test

```
se <- sd(repair_time)/sqrt(length(repair_time))

# compute standard error
t_value <- (mean_repair_time-mean_hypo)/se # compute t-value
df<- length(repair_time)-1 # degree of freedom
p_value <- 1- pt(t_value,df) # compute p-value
t_value;p_value

## [1] 2.560762

## [1] 0.005265342

qt(0.995,df) # the critical point of t-test

## [1] 2.578749
```

(v) Briefly describe how these values relate to the Null distribution of t (not graded)

- ANS: In result above, we get the t-value(2.560762) and p-value (0.005265342), and in part (a) we have found the 99%CI of the mean.

  For the CI, it's [7.593073, 9.450946],where 7.6 is in the range.

  For the t-value, it's 2.560762. By "qt(0.995,df)", we can know that $T_{0.995,df}$ = 2.578. The t-value is less than 2.578.

  For the p-value, it's 0.005265342, and it's less than α =0.01

(vi) What is your conclusion about the advertising claim from this t-statistic, and why?

- ANS: By the observation in part (v), we can say that under 1% confidence level, we do not reject the null hypothesis.

(b) bootstrapping on the sample data to examine this problem

(i) Bootstrapped Percentile: Estimate the bootstrapped 99% CI of the mean

```
compute_sample_mean <- function(sample0) {
  resample <- sample(sample0, length(sample0), replace=TRUE)
  mean(resample)
}
boostrap_mean<- replicate(2000,compute_sample_mean(repair_time)) # boot
strapping

boostrap_ci99 <- quantile(boostrap_mean,probs = c(0.005,0.995)) # take
0.005% and 0.995% quantile as 99% CI
boostrap_ci99
```

```
##     0.5%    99.5% 
## 7.600487 9.479103
```

(ii) What is the 99% CI of the bootstrapped difference between the population mean and the hypothesized mean?

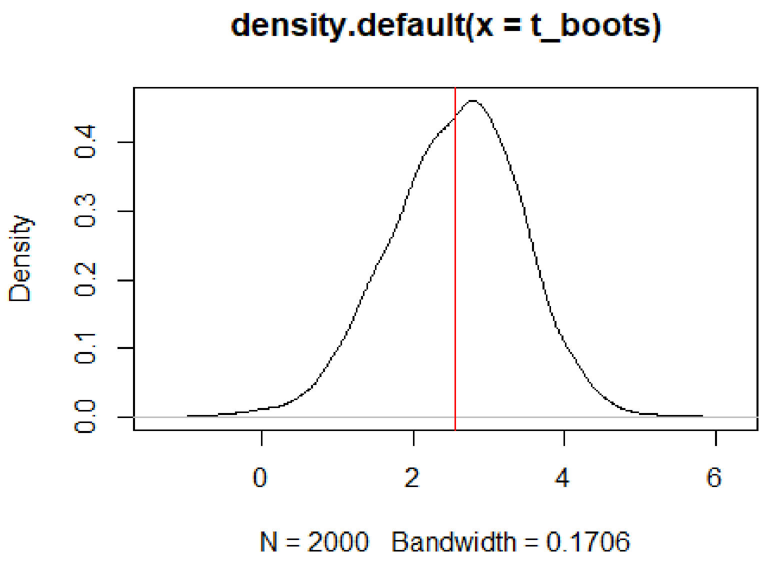
```
boot_mean_diffs <- function(sample0, mean_hyp) {
  resample <- sample(sample0, length(sample0), replace=TRUE)
  return( mean(resample) - mean_hyp )
}
mean_diffs <- replicate(2000,boot_mean_diffs(repair_time,mean_hypo))

# bootstrapping


diff_ci_99 <- quantile(mean_diffs,probs = c(0.005,0.995))

# take 0.005% and 0.995% quantile as 99% CI


diff_ci_99

##        0.5%       99.5% 
## -0.06717288  1.88161938
```

(iii) What is 99% CI of the bootstrapped t-statistic?

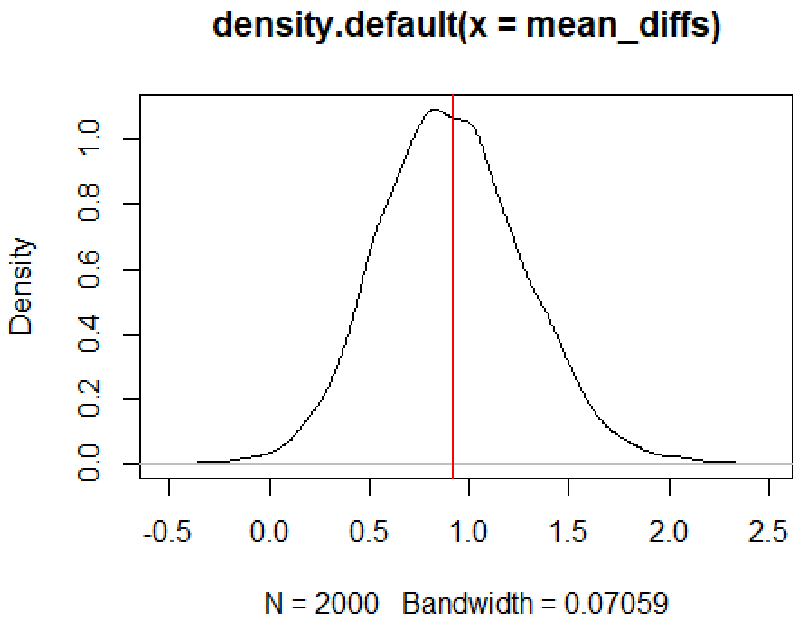
```
boot_t_stat <- function(sample0, mean_hyp) {
  resample <- sample(sample0, length(sample0), replace=TRUE)
  diff <- mean(resample) - mean_hyp
  se <- sd(resample)/sqrt(length(resample))
  return( diff / se )
}
t_boots <- replicate(2000, boot_t_stat(repair_time, mean_hypo)) # boots
trapping
t_stat_ci99 <- quantile(t_boots,probs= c(0.005,0.995)) # take 0.005% an
d 0.995% quantile as 99% CI
t_stat_ci99

##       0.5%      99.5% 
## 0.04584611 4.61176939
```

(iv) Plot separate distributions of all three bootstraps above

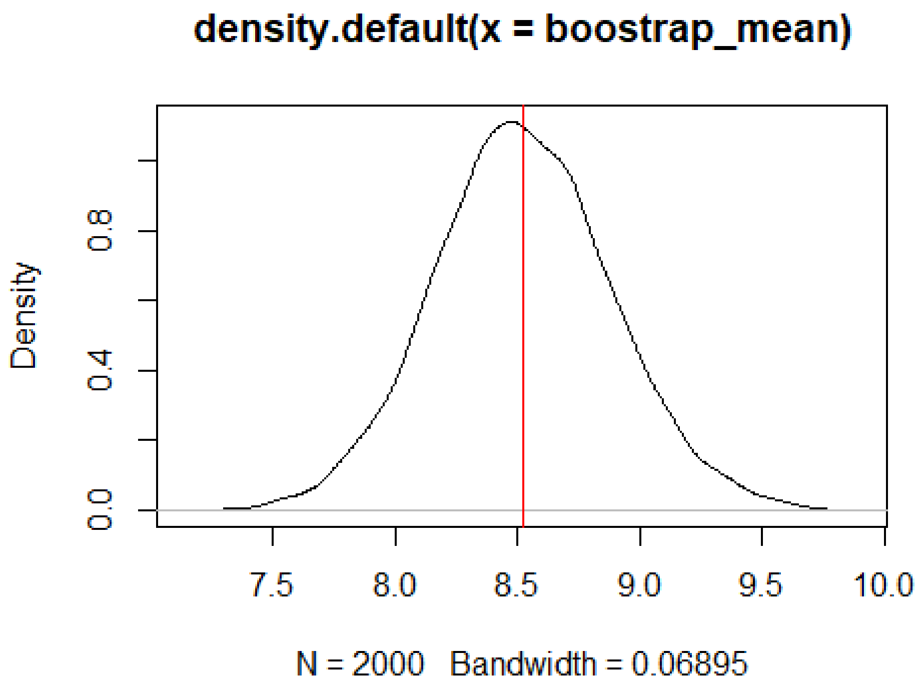
```
plot(density(t_boots))
abline(v = mean(t_boots),col = 'red')
```



```
plot(density(mean_diffs))
abline(v = mean(mean_diffs),col = 'red')
```



```
plot(density(boostrap_mean))
abline(v = mean(boostrap_mean),col = 'red')
```



(c)

Do the four methods (traditional test, bootstrapped percentile, bootstrapped difference of means, bootstrapped t-Interval) agree with each other on the test?

ANS: Yes, the four method give the same judge.

First, in part (i) , we can find that the bootstrapped 99% CI contains 7.6 in its range. Second, in part (ii), 0 is included in its CI, which means bootstrapped mean – hypothesis mean is 0.

Third, the 99% CI of bootstrapped t-static is [0.04584611 ,4.61176939], the t-static we found in part(a) is also included.

As the result, we can make a conclusion that the four method give the same judege.