HW6

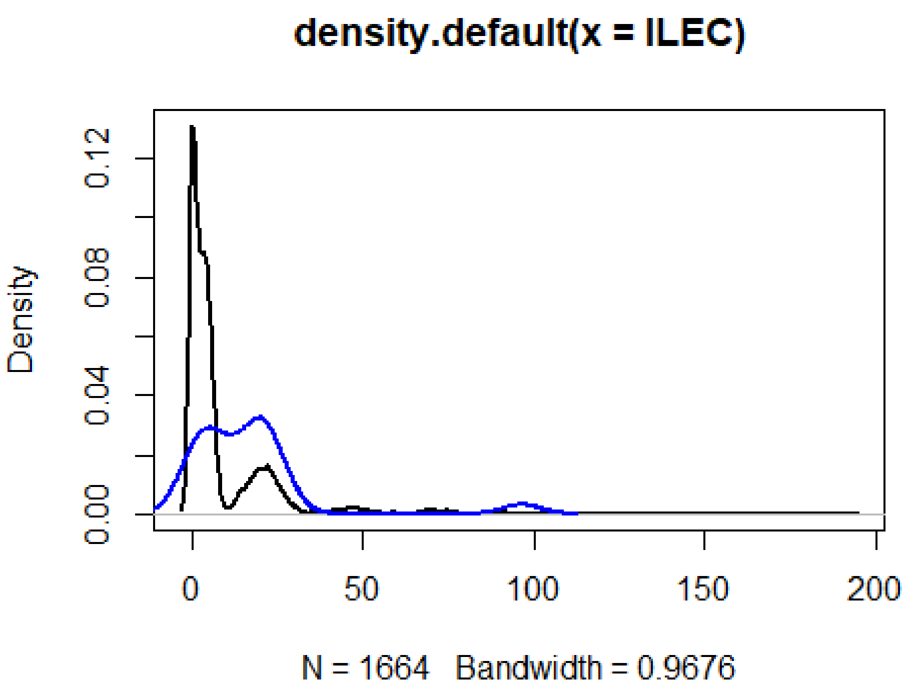
Helped by : 106034054 106070020

Q1

(a) Visualize Verizon's response times for ILEC vs. CLEC customers

```
data <- read.csv("verizon.csv", header = TRUE)
ILEC_time <- subset(data, Group == 'ILEC')
CLEC_time <- subset(data, Group == 'CLEC')
ILEC <- ILEC_time$Time
CLEC <- CLEC_time$Time

plot(density(ILEC), lwd = '2')
lines(density(CLEC),lwd = '2', col = 'blue')
```



(b) test the *difference between the mean of ILEC sample response times versus the mean of CLEC* sample response times.

(i) What are the appropriate null and alternative hypotheses in this case?

H_0:ILEC sample mean >= CLEC sample mean

H_1: ILEC sample mean < CLEC sample mean

(ii) Based on output , would you reject the null hypothesis or not?

```
t.test(ILEC, CLEC, alternative = 'less',
       var.equal = FALSE,conf.level = 0.99)
```

```
##
##  Welch Two Sample t-test
##
## data:  ILEC and CLEC
## t = -1.9834, df = 22.346, p-value = 0.02987
## alternative hypothesis: true difference in means is less th
an 0
## 99 percent confidence interval:
##      -Inf 2.130858
## sample estimates:
## mean of x mean of y
##  8.411611 16.509130
```

As the p-value is bigger than 1%, we should not reject H_0

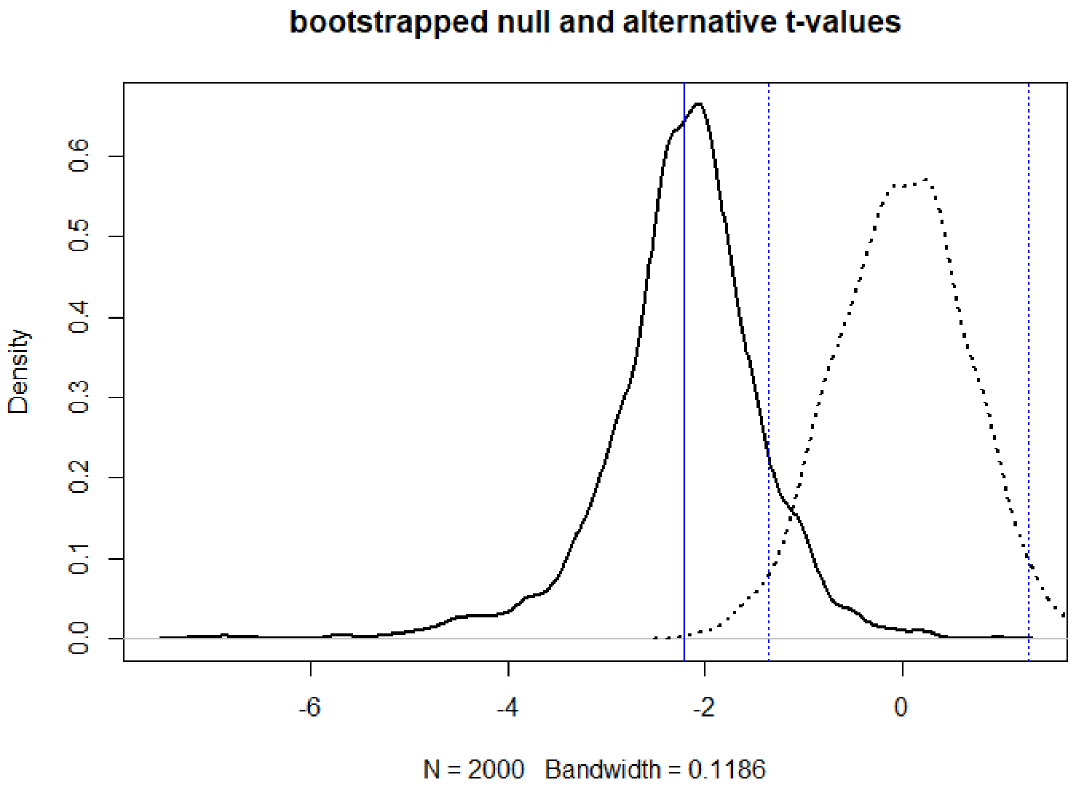
(c) Bootstrapped samples of ILEC against samples of CLEC & samples of ILEC against the original ILEC sample

(i) Plot a distribution of the bootstrapped null t-values and alternative t-values, adding vertical lines to show the 5% rejection zone of the null distribution .

```
boot_t_test <- function(sample0, sample1){
  resample0 <- sample(sample0, length(sample0), replace = 
TRUE)
  resample1 <- sample(sample1, length(sample1), replace = 
TRUE)
  t.test(resample0, resample1, alternative = 'less',
         var.equal = FALSE, conf.level = 0.99)
}
boot_sample0_only <-function(sample0, sample1){
  resample0 <- sample(sample0, length(sample0), replace = 
TRUE)
  t.test(resample0, sample1,
         var.equal = FALSE, conf.level = 0.99)
}
```

```
boot_result_1 <-replicate(2000,boot_t_test(ILEC, CLEC)$statistic)
boot_result_2 <- replicate(2000,boot_sample0_only(ILEC, ILEC)$sta
tistic)

plot(density(boot_result_1),lwd = 2, main = 'bootstrapped null an
d alternative t-values')
abline(v = mean(boot_result_1),col = 'blue')
lines(density(boot_result_2),lwd = 2,lty = 'dotted')
abline(v = quantile(boot_result_2,probs = 0.025),col = 'blue',lty
 = 'dotted')
abline(v = quantile(boot_result_2,probs = 0.975),col = 'blue',lty
 = 'dotted')
```



(ii) should we reject the null hypothesis

Yes, we should reject the null hypothesis because median of alternative is out of the 95% null lines.

Q2 Variance

(a) null and alternative hypotheses in this case?

H_0:ILEC variance = CLEC variance

H_1: ILEC variance < CLEC variance

(b) traditional statistical methods

(i) F-statistic of the ratio of variances

```
data <- read.csv("verizon.csv", header = TRUE)
ILEC_time <- subset(data, Group == 'ILEC')
CLEC_time <- subset(data, Group == 'CLEC')
ILEC <- ILEC_time$Time
CLEC <- CLEC_time$Time
var.test(ILEC, CLEC, alternative = 'less')
```

```
##
##  F test to compare two variances
##
## data:  CLEC and ILEC
## F = 1.7627, num df = 22, denom df = 1663, p-value = 0.01582
## alternative hypothesis: true ratio of variances is greater th
an 1
## 95 percent confidence interval:
##  1.138356      Inf
## sample estimates:
## ratio of variances
##           1.762717
```

(ii) cut-off value of F

```
qf(0.95, df1 = length(CLEC)-1, df2 = length(ILEC)-1)
```

```
## [1] 1.548476
```

(iii) Can we reject the null hypothesis

Reject the null hypothesis. As the F-value in (ii) is bigger than the cut-off value which means in the cut-off zone.

(c) bootstrapping

(i) bootstrapped values of the F-statistic, for both null and alternative hypotheses.

```
f_test_boot <- function(larger_sd_sample, smaller_sd_sample){
  resample_sd1 <- sample(larger_sd_sample, length(larger_sd_sample),
replace = TRUE)
  resample_sd2 <-sample(smaller_sd_sample,length(smaller_sd_sample),
replace = TRUE)
  f_alt <- var(resample_sd1)/var(resample_sd2)
```

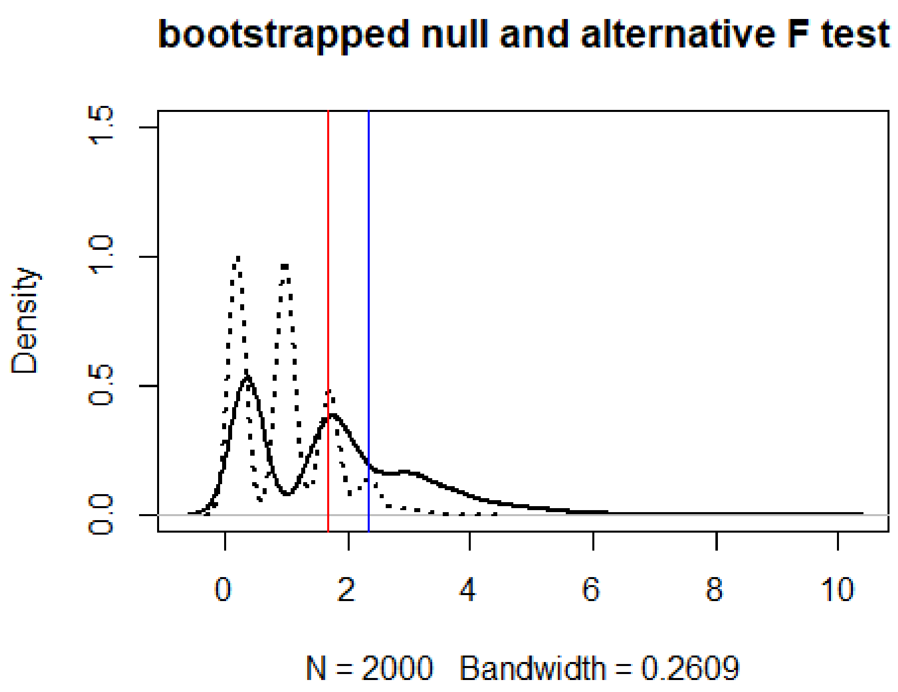
```
    f_null <- var(resample_sd1)/var(larger_sd_sample)
  c(f_alt, f_null)
}
f_boot <- replicate(2000,f_test_boot(CLEC, ILEC))
f_alts <- f_boot[1,]
f_nulls <- f_boot[2,]
```

(ii) 95% cutoff value according to the bootstrapped *null* values of F?

```
boot_var_95 <-quantile(f_nulls, probs = 0.95)
boot_var_95
##      95%
## 2.360329
```

(iii) a visualization of the null and alternative distributions of the bootstrapped F-statistic, with vertical lines at the cutoff value of F nulls.

```
plot(density(f_alts),lwd = 2, main = 'bootstrapped null and
 alternative F test', ylim = c(0,1.5))
lines(density(f_nulls), lwd = 2,lty = 'dotted')
abline(v =boot_var_95,col = 'blue' )
abline(v = median(f_alts), col = 'red')
```



(iv) bootstrap results suggest about the null hypothesis?

We should not reject the null hypothesis because the median of the alternative is inside the 95% zone.

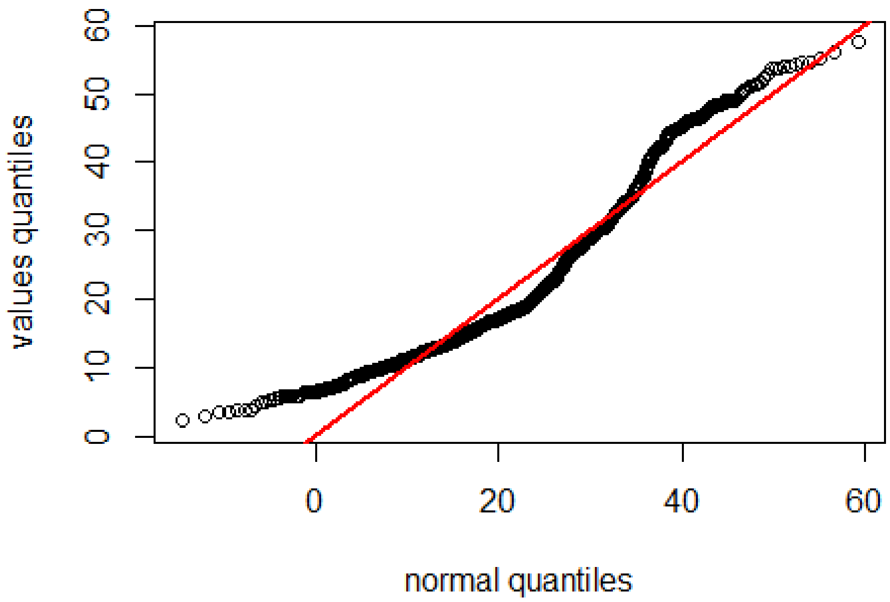
Q3

(a) Let's create a function to see if key statistics/assumptions of *normality* are met in our distributions.

```
norm_qq_plot<- function(values){
  probs1000 <- seq(0, 1, 0.001)
  q_vals <- quantile(values, probs = probs1000)
  q_norm <- qnorm(probs1000,mean = mean(values), sd =
sd(values))
  plot(q_norm, q_vals, xlab ='normal quantiles', ylab =
'values quantiles')
  abline(a=0, b= 1, col = 'red', lwd = 2)
}
```

(b) running it against the values of our `d123` distribution

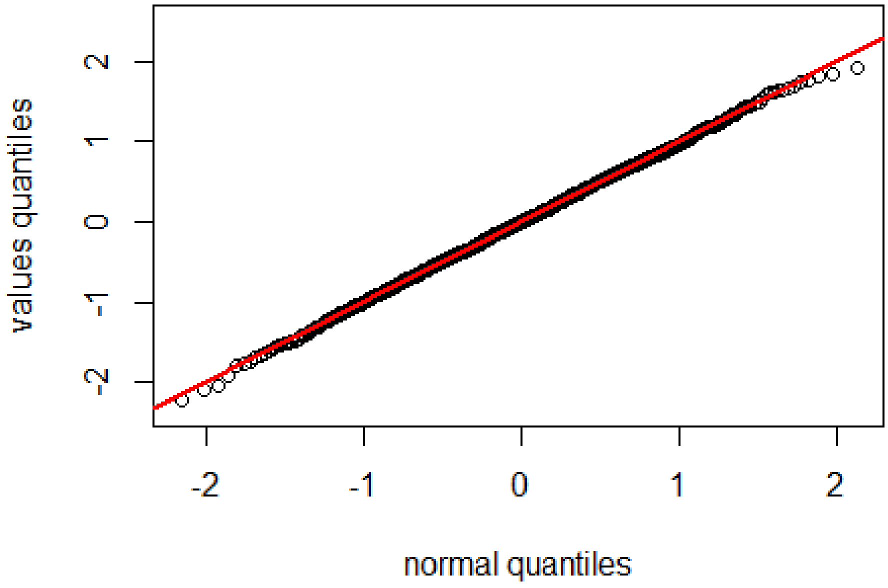
```
norm_qq_plot(d123)
```



The dots doesn't lie on the red line so it is not a normal distribution.

(c) Use your normal Q-Q plot function to check if the bootstrapped distribution of null t-values in question 1c was normally distributed. What's your conclusion?

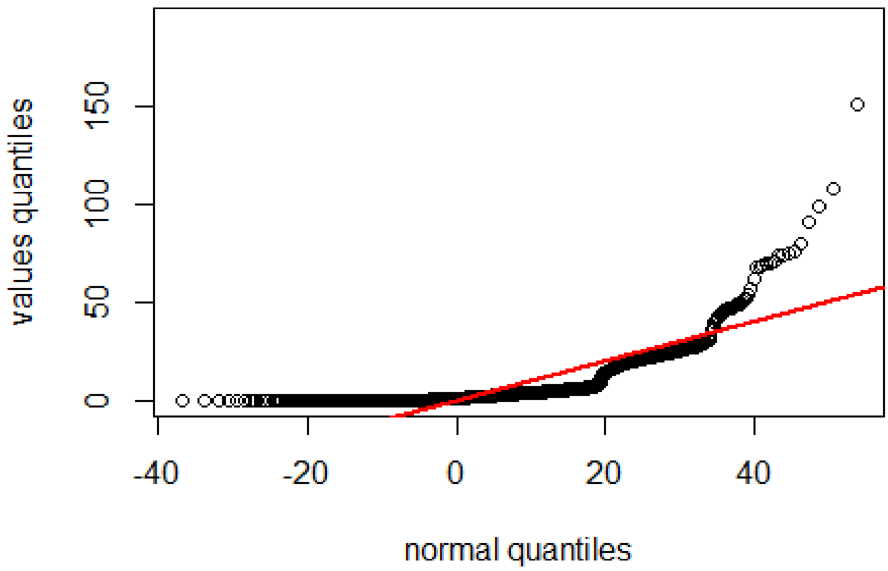
```
norm_qq_plot(boot_result_2)
```



The dots lie on the red line so it is not a normal distribution.

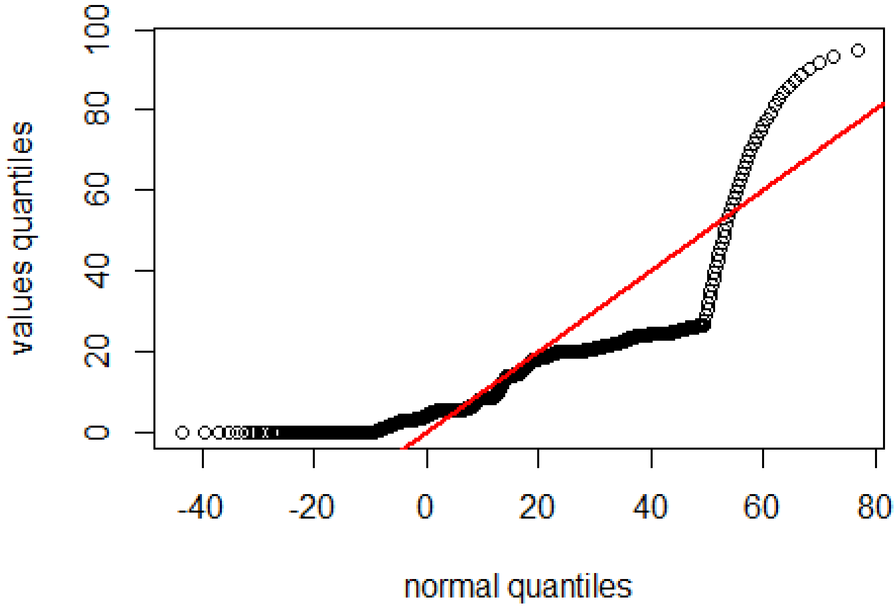
(d) Use your normal Q-Q plot function to check if the two samples we compared in question 2 could have been normally distributed. What's your conclusion?

```
norm_qq_plot(ILEC)
```



The dots doesn't lie on the red line so it is not a normal distribution.

```
norm_qq_plot(CLEC)
```



The dots doesn't lie on the red line so it is not a normal distribution.