

# BACS\_HW\_Week6

106071041

2021/4/5

## Question 1

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

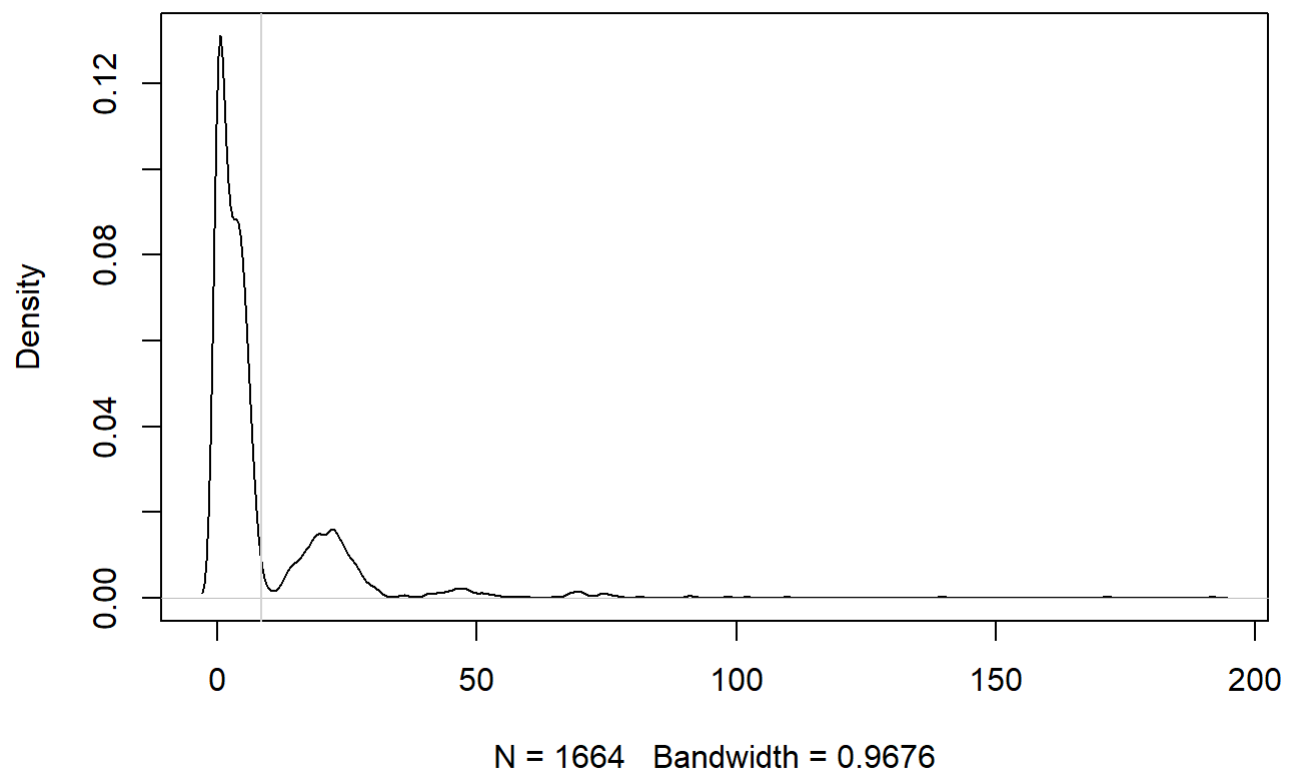
```
Verizon_response <- read.csv("verizon.csv")
```

```
Time <- Verizon_response$Time  
Group <- Verizon_response$Group
```

```
is_ILEC <- Group == "ILEC"  
is_CLEC <- Group == "CLEC"
```

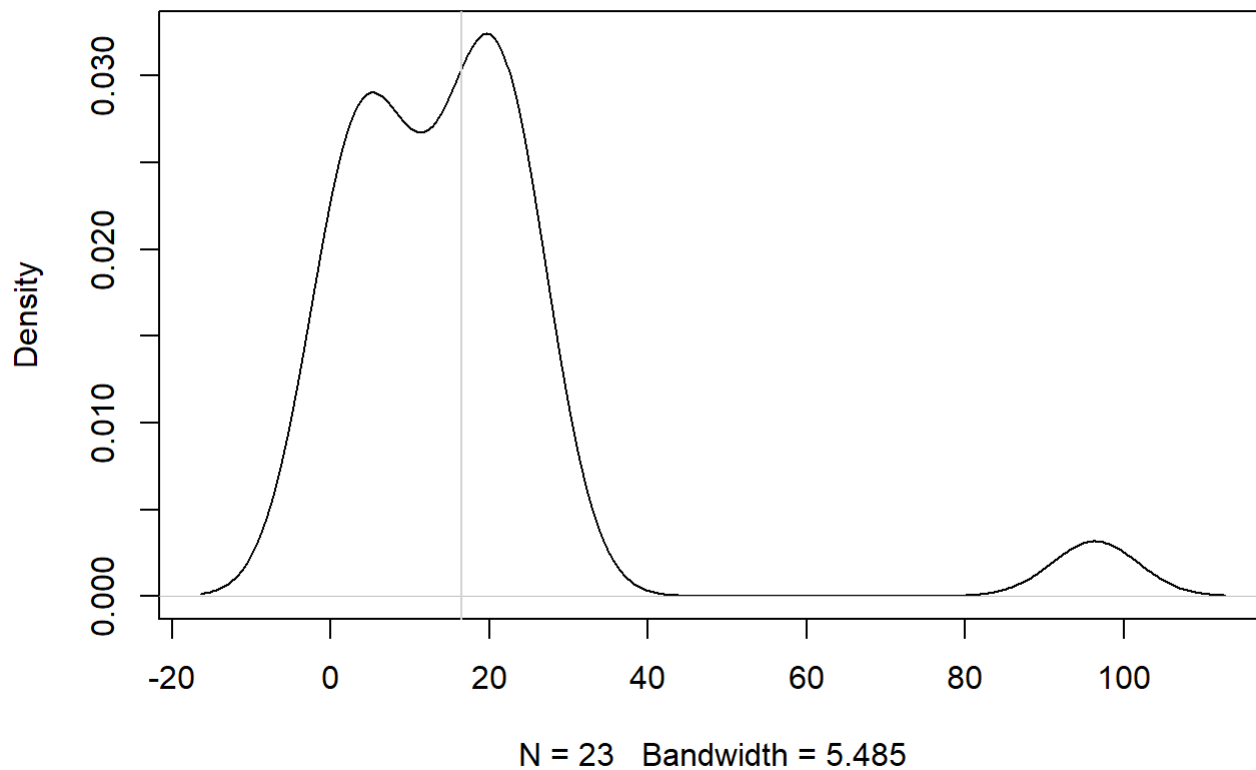
```
response_times_ILEC <- Time[is_ILEC]  
response_times_CLEC <- Time[is_CLEC]
```

```
plot(density(response_times_ILEC))  
abline(v=mean(response_times_ILEC), col = "lightgray")
```

**density.default(x = response\_times\_ILEC)**

```
plot(density(response_times_CLEC))  
abline(v=mean(response_times_CLEC),col = "lightgray")
```

## density.default(x = response\_times\_CLEC)



## b. t.test()

```
t.test(response_times_ILEC, response_times_CLEC, alternative = "l", conf.level = 0.99)
```

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

## i. null and alternative

let  $a$  = the mean of response\_time\_ILEC

$b$  = the mean of response\_time\_CLEC

then  $a-b$  = the difference in means

Null:  $a-b = 0$

Alternative:  $a-b < 0$  (ignore  $a-b > 0$  since it's impossible considering the company's preference)

## ii. reject or not (1% significance level)

not reject null hypothesis ( $-1.98 > -2.32$ )

```
quants_99_1sided <- qt(c(0.01), length(response_times_ILEC)-1)
```

```
quants_99_1sided
```

```
## [1] -2.328592
```

```
t.test(response_times_ILEC, response_times_CLEC, alternative = "l", conf.level = 0.99)
```

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

## c. use t.test() to compare alt t-values and null t-values

### i. plot

```
bootstrap_null_alt <- function(){
  resample_ILEC <- sample(response_times_ILEC, replace = TRUE)
  resample_CLEC <- sample(response_times_CLEC, replace = TRUE)

  t_stat_alt <- t.test(resample_ILEC, resample_CLEC, "l", conf.level = 0.99)
  t_stat_null <- t.test(resample_ILEC, response_times_CLEC, "l", conf.level = 0.99)

  c(t_stat_alt$statistic, t_stat_null$statistic)
}
```

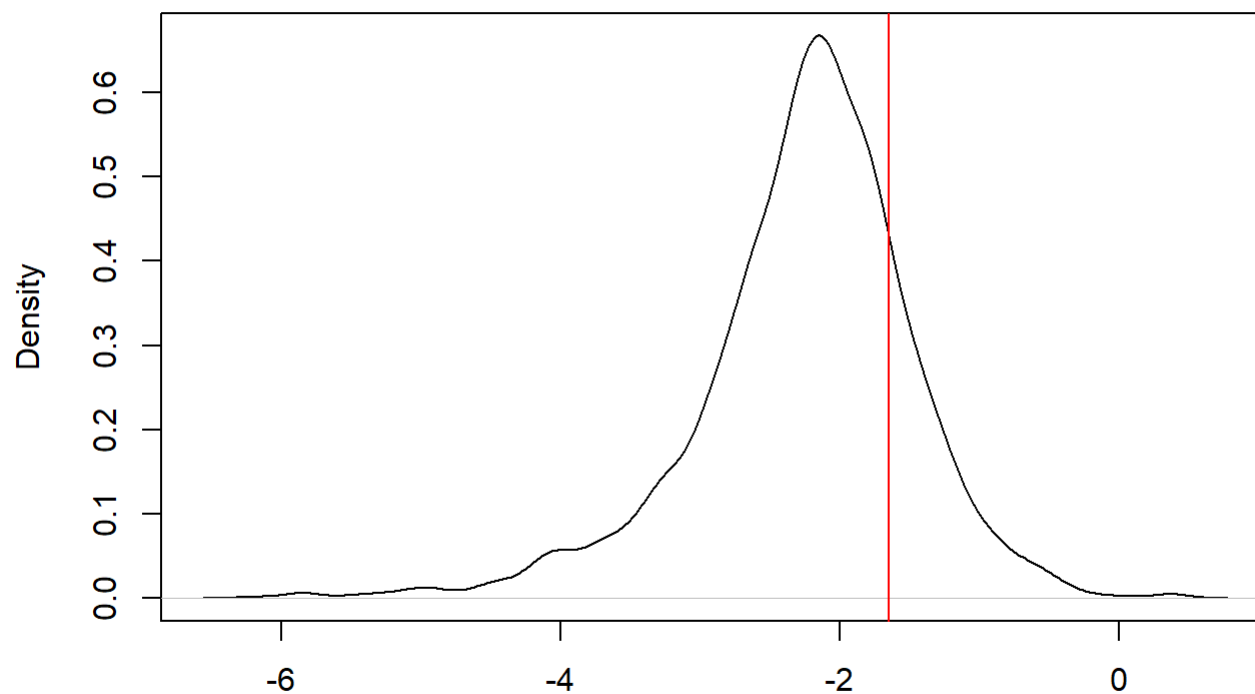
```
set.seed(723482943)
boot_t_alt <- replicate(2000, bootstrap_null_alt()[1])
boot_t_null <- replicate(2000, bootstrap_null_alt()[2])
```

```
quants_95_1sided <- qt(c(0.05), length(response_times_ILEC)-1)
quants_95_1sided
```

```
## [1] -1.64577
```

```
plot(density(boot_t_alt))  
abline(v=quants_95_1sided, col = "red")
```

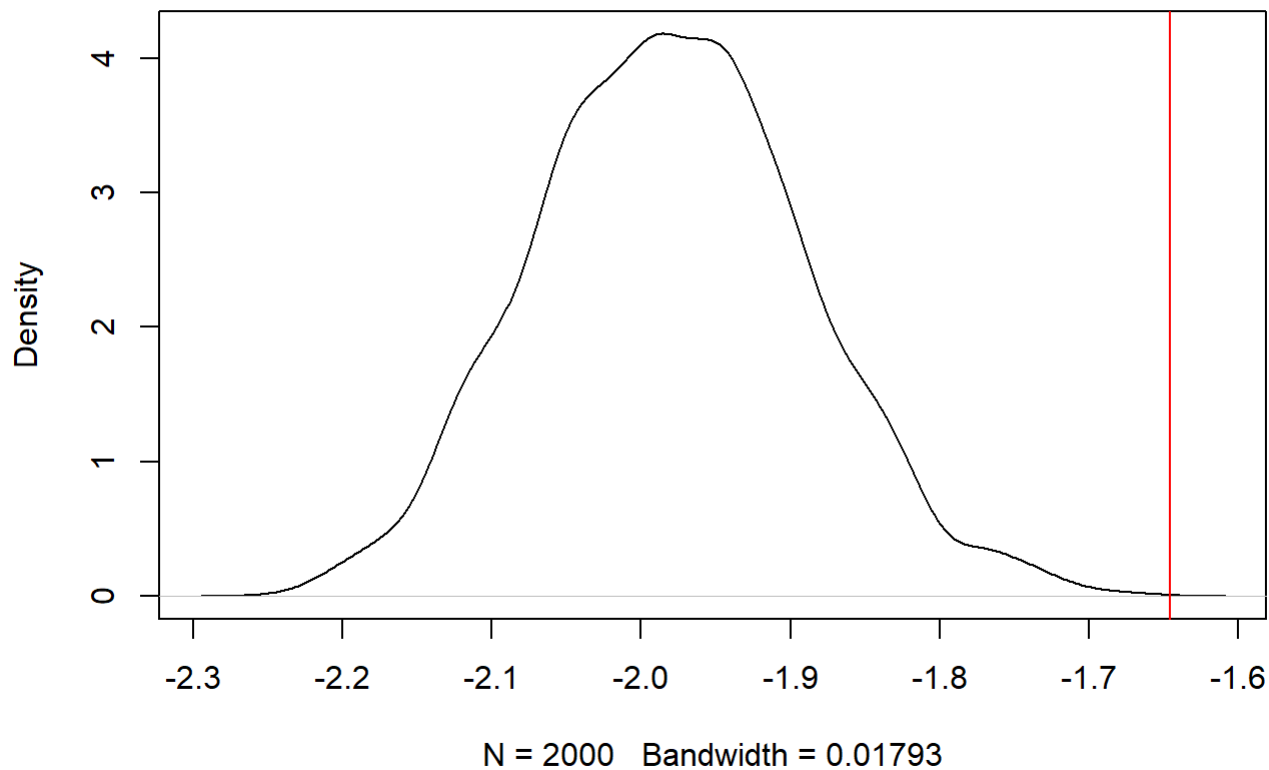
**density.default(x = boot\_t\_alt)**



N = 2000 Bandwidth = 0.125

```
plot(density(boot_t_null))  
abline(v=quants_95_1sided, col = "red")
```

**density.default(x = boot\_t\_null)**



ii. should we reject?

looks like we should reject

## Question 2

### a. null & alternative

let

c = the variance of response\_time\_ILEC

d = the variance of response\_time\_CLEC

then c-d = the difference in variances

Null: c-d = 0

Alternative: c-d is not equal to 0

### b. F-statistic

#### i. F-statistic

```
var_ILEC <- var(response_times_ILEC)
var_CLEC <- var(response_times_CLEC)
smaller_var <- ifelse(min(var_ILEC, var_CLEC) == var_CLEC, "var_CLEC", "var_ILEC")
smaller_var
```

```
## [1] "var_ILEC"
```

```
F_statistic <- var_CLEC/var_ILEC
F_statistic
```

```
## [1] 1.762717
```

```
var.test(response_times_CLEC, response_times_ILEC, ratio = 1, "g")
```

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

## ii. cut-off value of F-statistic

```
qf(p=0.95, df1 = length(response_times_CLEC), df2 = length(response_times_ILEC))
```

```
## [1] 1.53575
```

## iii. can we reject?

Yes, reject it!!!

## c. bootstrapping

### i. bootstrapped F-statistics for both null and alternative hypothesis

```
set.seed(23423424)
sd_providers_test <- function(larger_sd_sample, smaller_sd_sample){
  resample_larger_sd <- sample(larger_sd_sample, replace = TRUE)
  resample_smaller_sd <- sample(smaller_sd_sample, replace = TRUE)
  f_alt <- var(resample_larger_sd) / var(resample_smaller_sd)
  f_null <- var(resample_larger_sd) / var(larger_sd_sample)
  c(f_alt, f_null)
}
```

```
boot_f_stats <- replicate(10000, sd_providers_test(response_times_CLEC, response_times_ILEC))
# bootstrapped null hypothesis
f_alts <- boot_f_stats[1,]
# bootstrapped alternative hypothesis
f_nulls <- boot_f_stats[2,]
```

## ii. 95% cutoff value of null F-statistics

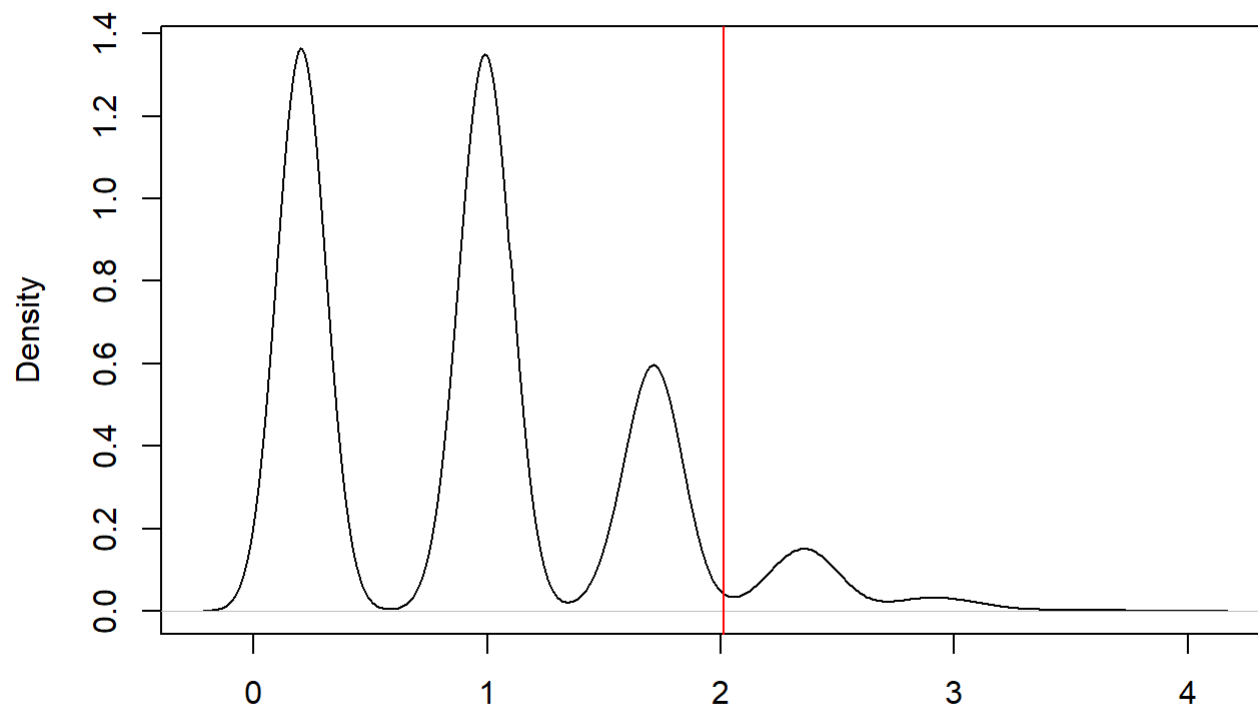
```
cutoff_Fstatistics <- qf(p=0.95, df1 = length(response_times_CLEC), df2 = length(response_times_
CLEC))
cutoff_Fstatistics
```

```
## [1] 2.014425
```

## iii. visualizaiton of bootstrapped null and alternative F-statistics and draw the cutoff value vertical line

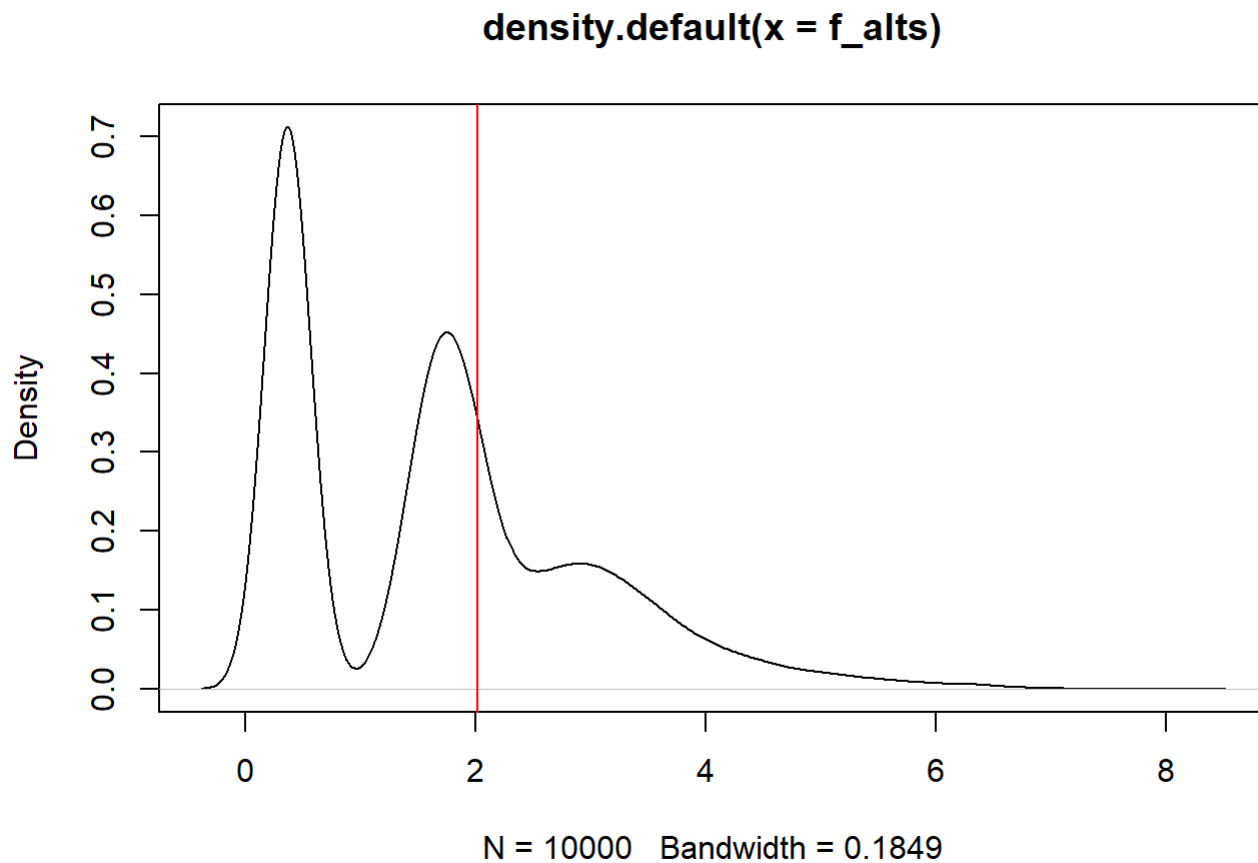
```
# bootstrapped null F-statistics
plot(density(f_nulls),)
abline(v=cutoff_Fstatistics, col = "red")
```



**density.default(x = f\_nulls)**

N = 10000 Bandwidth = 0.09987

```
# bootstrapped alternative F-statistics  
plot(density(f_alts))  
abline(v=cutoff_Fstatistics, col = "red")
```



iv. what do the bootstrap results suggest about the *null* hyp?

ILEC and CLEC have the same variances

## Question 3

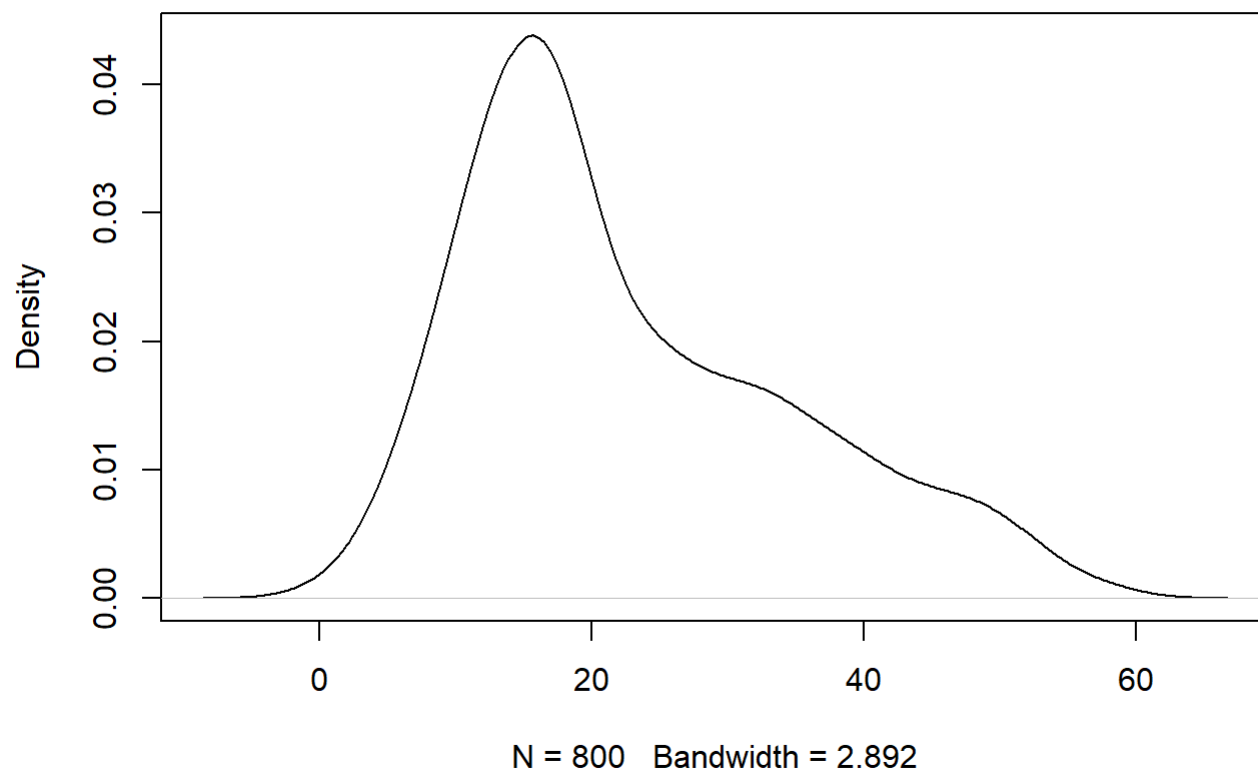
a. Create a function to see if key statistics/assumptions of normality are met in our distributions

```
norm_qq_plot <- function(values){  
  # i. Create a sequence of probability numbers from 0 to 1, with ~1000 probabilities in between  
  probs1000 <- seq(0, 1, 0.001)  
  # ii. Calculate ~1000 quantiles of our values (you can use probs=probs1000)  
  q_vals <- quantile(values, probs = probs1000)  
  # iii. Calculate ~1000 quantiles of a perfectly normal distribution with the same mean and standard deviation as our values  
  q_norm <- qnorm(probs1000, mean = mean(values), sd = sd(values))  
  # iv. Create a scatterplot comparing the quantiles of a normal distribution versus quantiles of values  
  plot(q_norm, q_vals, xlab="normal quantiles", ylab="values quantiles")  
  # v. Finally, draw a red line with intercept of 0 and slope of 1, comparing these two sets of quantiles  
  abline(0, 1, col="red", lwd=2)  
}
```

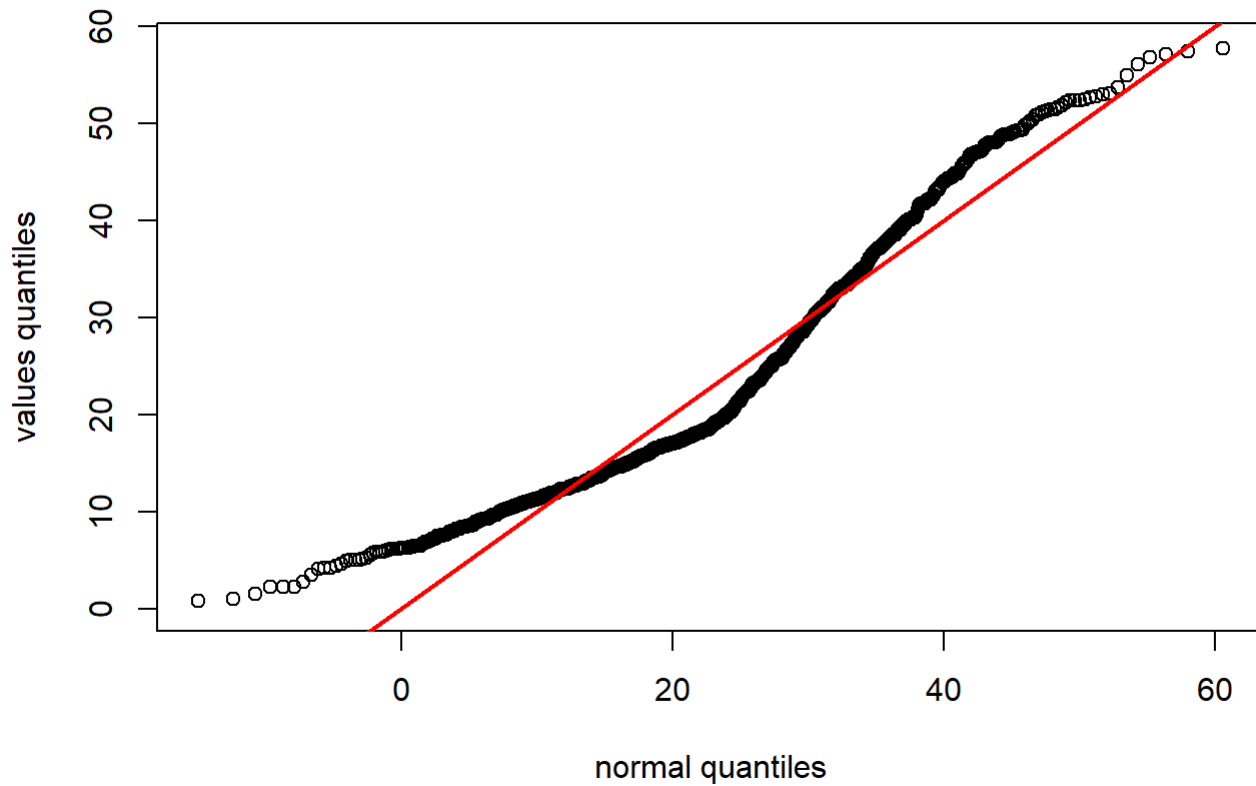
## b. Test the function by d123

a Q-Q plot is trying to answer the question: "How similar are the quantiles in my dataset compared to what the quantiles of my dataset would be if my dataset followed a theoretical probability distribution?"

```
set.seed(32423)  
d1 <- rnorm(n=500, mean=15, sd=5)  
d2 <- rnorm(n=200, mean=30, sd=5)  
d3 <- rnorm(n=100, mean=45, sd=5)  
d123 <- c(d1, d2, d3)  
  
plot(density(d123))
```

**density.default(x = d123)**

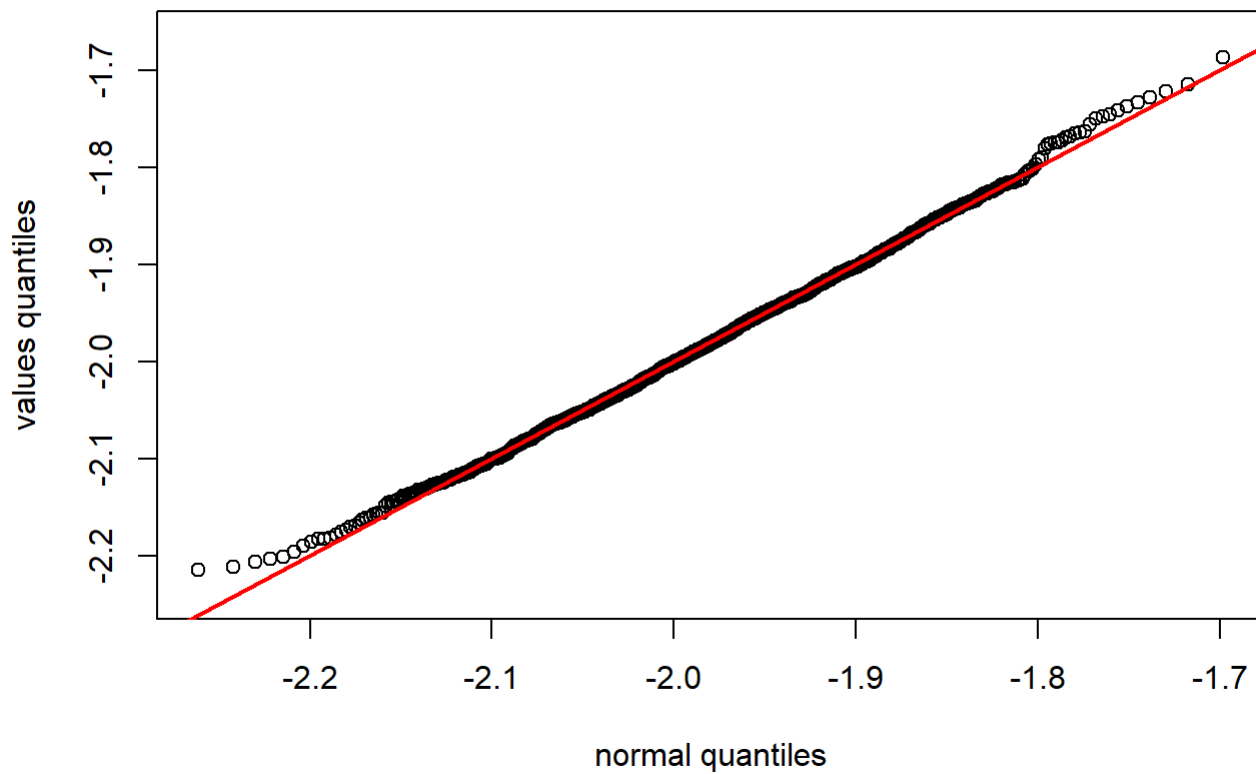
```
norm_qq_plot(d123)
```



c. Check the bootstrapped distribution of null t-values was normally distributed

Yes, it was normally distributed

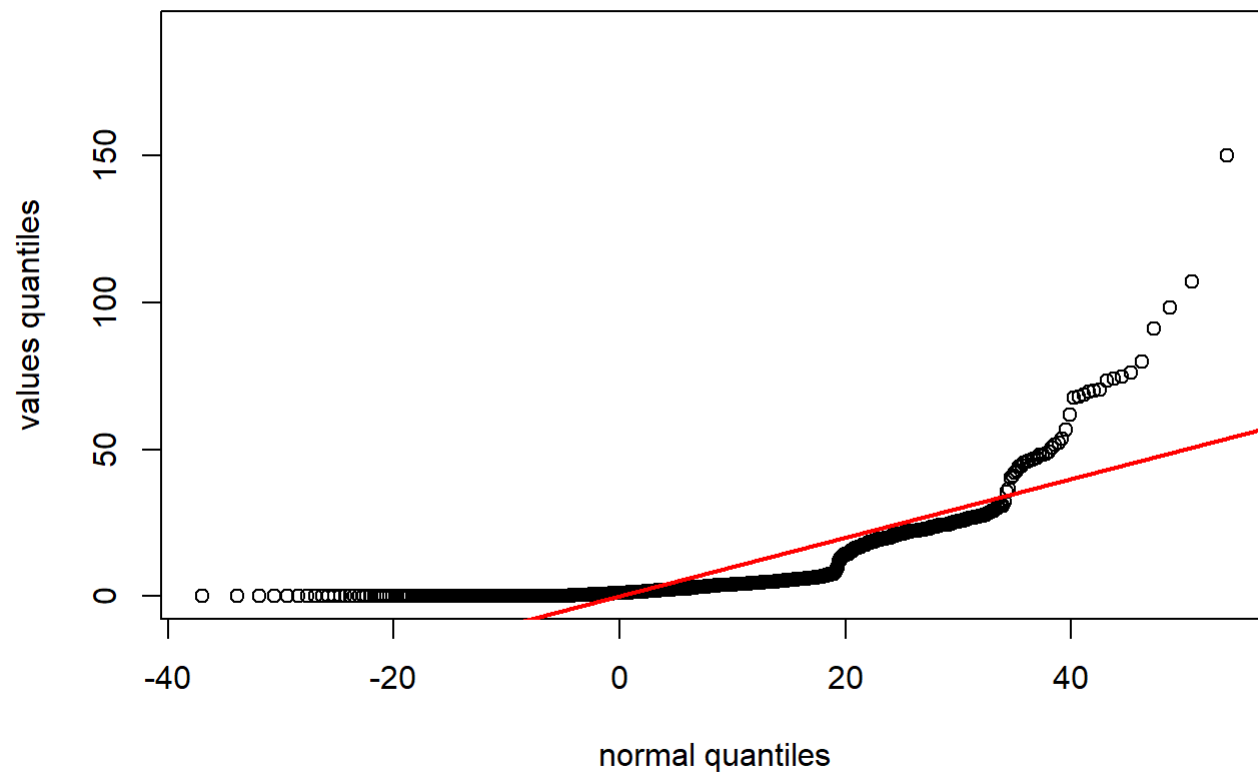
```
norm_qq_plot(boot_t_null)
```



d. Check the two samples in Q2 were normally distributed

It seems that they are not.

```
norm_qq_plot(response_times_ILEC)
```



```
norm_qq_plot(response_times_CLEC)
```

