

# HW6

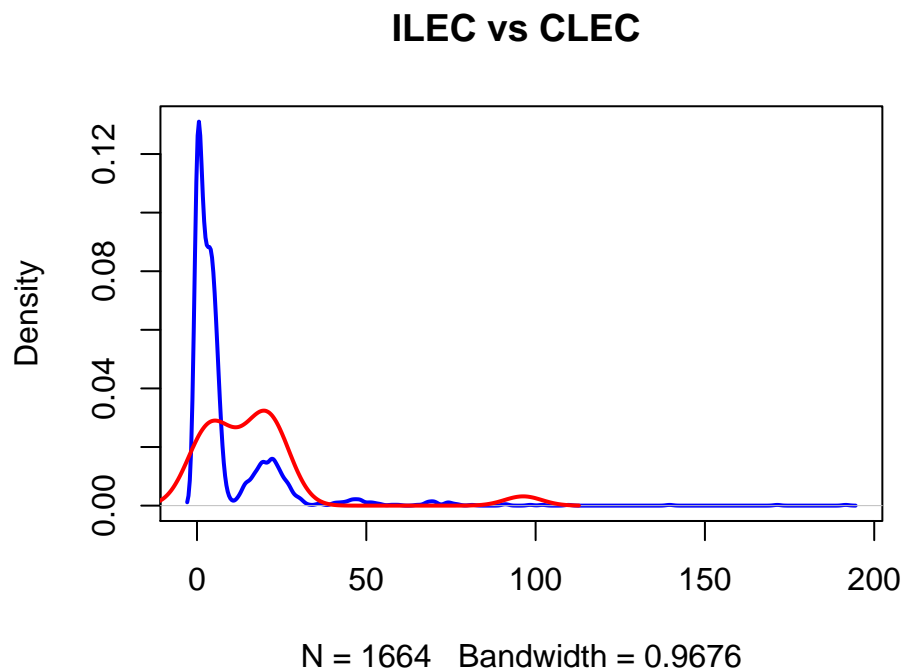
106022113

## Question 1:

```
verizon <- read.csv("verizon.csv")
ILEC <- verizon[verizon$Group == 'ILEC',]
CLEC <- verizon[verizon$Group == 'CLEC',]
ILEC_data <- ILEC$Time
CLEC_data <- CLEC$Time
```

### (a) Visualize the Results

```
plot(density(ILEC_data), lwd = 2, col = 'blue', main = 'ILEC vs CLEC ')
#plot(density(CLEC_data), lwd = 2, col = 'blue' , main = 'CLEC')
lines(density(CLEC_data), lwd = 2, col = 'red')
```



(b) Use `t.test()` to test between means of both response times

```
t.test(ILEC_data,CLEC_data,conf.level = 0.99)

##
## Welch Two Sample t-test
##
## data: ILEC_data and CLEC_data
## t = -1.9834, df = 22.346, p-value = 0.05975
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
## -19.588967 3.393927
## sample estimates:
## mean of x mean of y
## 8.411611 16.509130
```

**i. Appropriate hypothesis:**  $H_0 : \mu_1 = \mu_2$  ,  $H_1 : \mu_1 \neq \mu_2$

Since Verizon claims that the mean reponse time will be the same, the null hypothesis will be for the two means to be the same.

**ii. Results based on the test:** **ANSWER:** Since the 95 percent confidence interval covered 0, we don't have enough evidence to reject the Null hypothesis.

(c) Try the test using bootstrapping.

```
set.seed(38)
bootstrap_null_alt <- function(sample0) {
  resample <- sample(sample0, length(sample0), replace=TRUE)
  resample_se <- sd(resample)/sqrt(length(resample))
  t_stat_null <- (mean(resample)-mean(sample0))/resample_se
  t_stat_alt <- (mean(resample)-mean(boot_CLEC))/resample_se
  return(c(t_stat_alt,t_stat_null))
}
bootstrap <- function(sample0) {
  resample <- sample(sample0, length(sample0), replace=TRUE)
  return(mean(resample))
}
boot_CLEC <- replicate(2000, bootstrap(CLEC_data))
boot_ILEC <- replicate(2000, bootstrap(ILEC_data))
boot_alt_null <- replicate(2000, bootstrap_null_alt(ILEC_data))
```

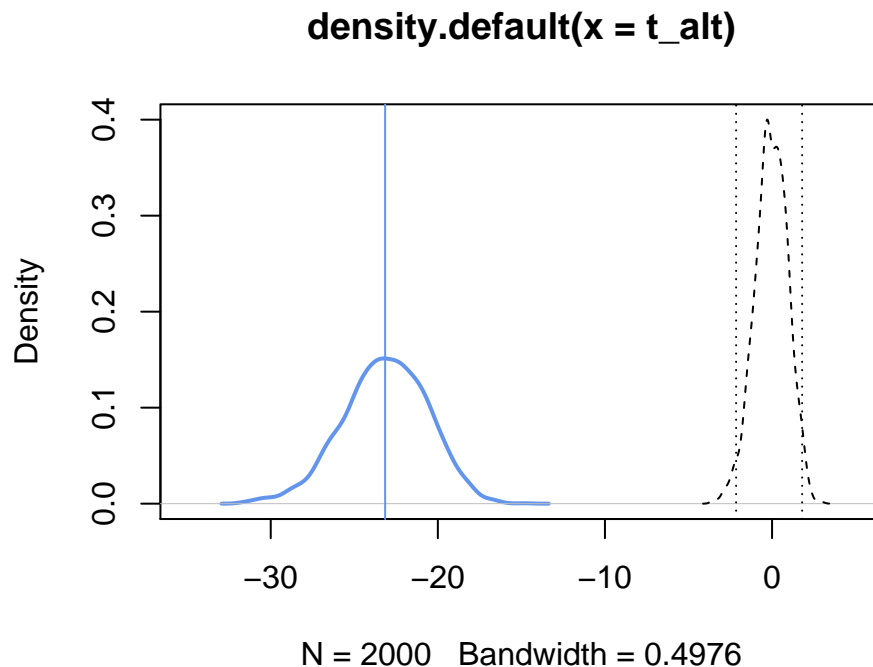
```
t.test(boot_ILEC,boot_CLEC)
```

Test the bootstraps

```
##
## Welch Two Sample t-test
##
## data: boot_ILEC and boot_CLEC
## t = -92.68, df = 2032.5, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -8.427473 -8.078208
## sample estimates:
## mean of x mean of y
## 8.411743 16.664583
```

```
t_alt <- boot_alt_null[1,]
t_null <- boot_alt_null[2,]
plot(density(t_alt),lwd=2,xlim = c(-35,5),ylim=c(0,0.4),col='cornflowerblue')
abline(v=mean(t_alt),col='cornflowerblue')
lines(density(t_null),lty='dashed')
ci_95 <- quantile(t_null, probs = c(0.025,0.975))
abline(v=ci_95,lty='dotted')
```

i. Plot distribution with vertical lines



ii. Should we reject the null hypothesis? **ANSWER:** Since our alternative t values lies outside the 95% CI of null t-distribution, we should reject the null hypothesis

## Question 2: Test variances of ILEC with CLEC

### a. NULL and Alternative Hypothesis

```
var(CLEC_data)-var(ILEC_data)
```

```
## [1] 164.5922
```

**ANSWER:**

$$H_{null} : \text{Var}_{ILEC} = \text{Var}_{CLEC}$$

$$H_{alt} : \text{Var}_{ILEC} > \text{Var}_{CLEC}$$

### b. Tradition statistics:

```
f_value <- var(CLEC_data)/var(ILEC_data)
f_value
```

#### i. F-statistic value:

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

**ANSWER:** Hence we have the f-statistic: 1.762717

```
cutoff <-qf(p =0.95,df1=length(ILEC_data)-1,df2 = length(CLEC_data)-1)
paste("The cutoff Value is: ",cutoff)
```

#### ii. Cutoff Value to reject 95%

```
## [1] "The cutoff Value is: 1.78720338827908"
```

```
var.test(CLEC_data,ILEC_data,alternative = 'greater')
```

#### iii. Can we reject the null hypothesis?

```
##
## F test to compare two variances
##
## data: CLEC_data and ILEC_data
## 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
```

**ANSWER:** We don't have enough evidence to reject the null hypothesis since the p-value is greater than 0.05.

### c. Bootstrapping

```
set.seed(43)
var_verizon_test <- function(larger_var_sample, smaller_var_sample){
  resample_larger_var <- sample(larger_var_sample, length(larger_var_sample), replace = T)
  resample_smaller_var <- sample(smaller_var_sample, length(smaller_var_sample), replace = T)
  f_alt <- var(resample_larger_var) / var(resample_smaller_var)
  f_null <- var(resample_larger_var) / var(larger_var_sample)
  return(c(f_alt, f_null))
}
```

```
f_stats <- replicate(2000, var_verizon_test(CLEC_data, ILEC_data))
f_alts <- f_stats[1,]
f_nulls <- f_stats[2,]
```

### i. Bootstrap f-statistics of null and alternative hypothesis

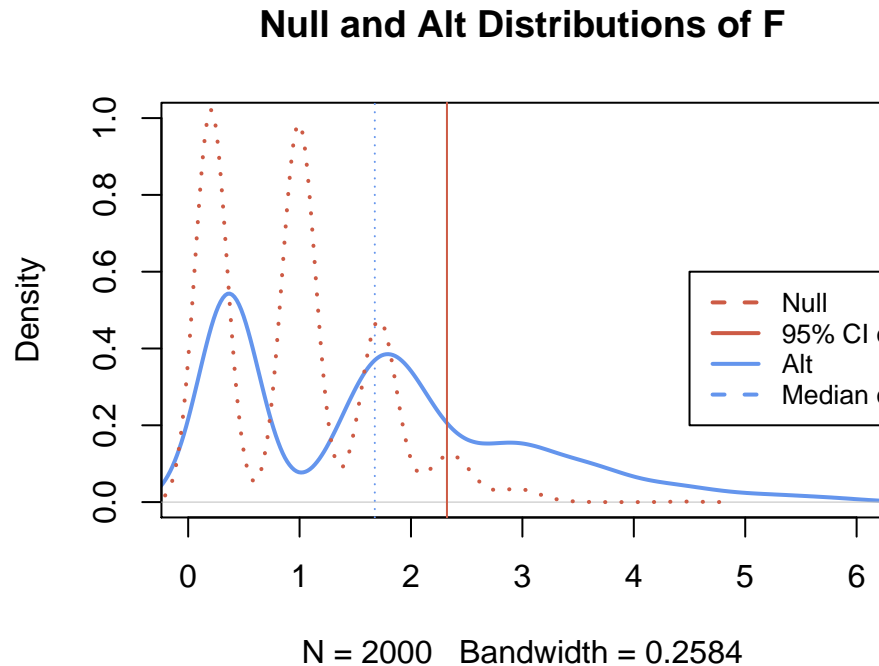
```
quantile(f_nulls, probs = 0.95)
```

### ii. 95% cutoff value according to the bootstrapped null values of F

```
##      95%
## 2.322896
```

```
plot(density(f_alts), col="cornflowerblue", ylim = c(0,1), xlim=c(0,6), main="Null and Alt Distributions of F",
lines(density(f_nulls), col="coral3", lwd=2, lty="dotted")
abline(v=quantile(f_nulls, probs=0.95), col="coral3")
abline(v=median(f_alts), lty="dotted", col="cornflowerblue")
legend(4.5, 0.6, c("Null", "95% CI of Null", "Alt", "Median of Alt"),
      lty=c(2,1,1,2),
      col = c("coral3", "coral3", "cornflowerblue", "cornflowerblue"),
      lwd = c(2,2,2,2),
      cex = 0.8
    )
```

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



iv. What do the bootstrap results suggest about the null hypothesis? **ANSWER:** It suggests that we do not reject the null hypothesis. If I plotted out the median of alternative hypothesis, it lies in side the 95% CI of the null distribution.

### Question 3. Non-parametric Bootstrap

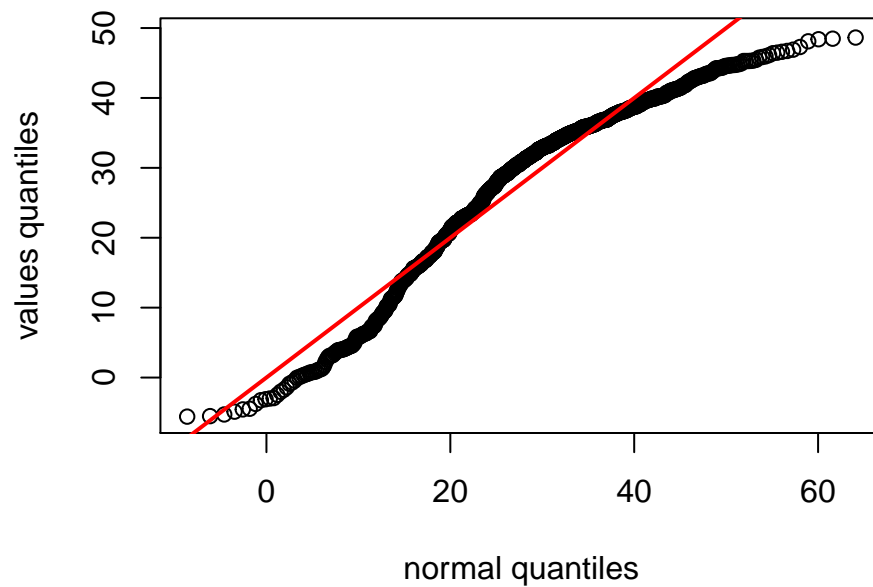
a. Create a function to see if key statistics/assumptions of normality are met in our distributions.

```
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. Confirm function works by running it against d123 distribution

```
d1 <- rnorm(n=100, mean=5, sd=5)
d2 <- rnorm(n=200, mean=20, sd=5)
```

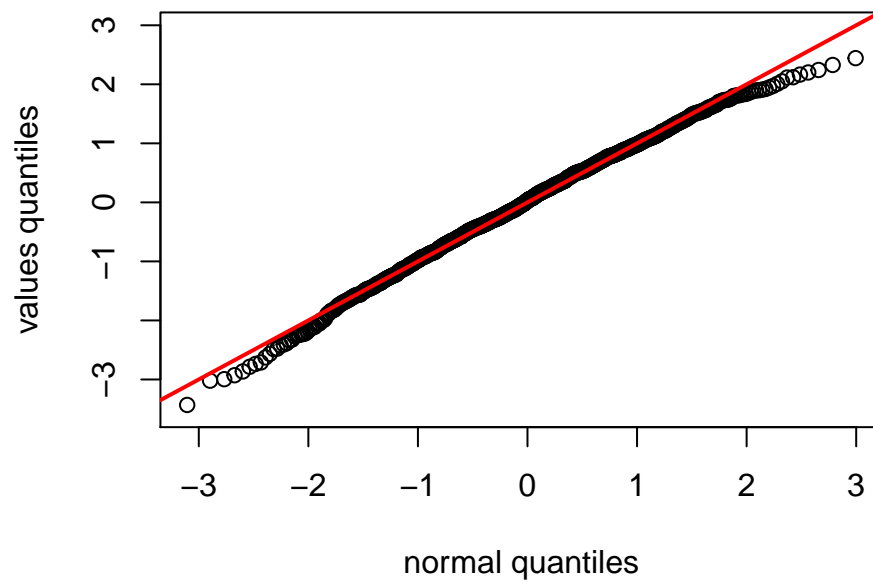
```
d3 <- rnorm(n=500, mean=35, sd=5)
d123 <- c(d1, d2, d3)
norm_qq_plot(d123)
```



**ANSWER:** The points in the Q-Q plot form a relatively straight line since the quantiles of the d123 nearly match what the quantiles of the d123 would theoretically be if the dataset was normally distributed. Hence, we can see that the black dots form closely to the red line, indicating it is near being normal.

c. Check bootstrap of null t-values in question(1c) with QQ plot

```
norm_qq_plot(t_null)
```

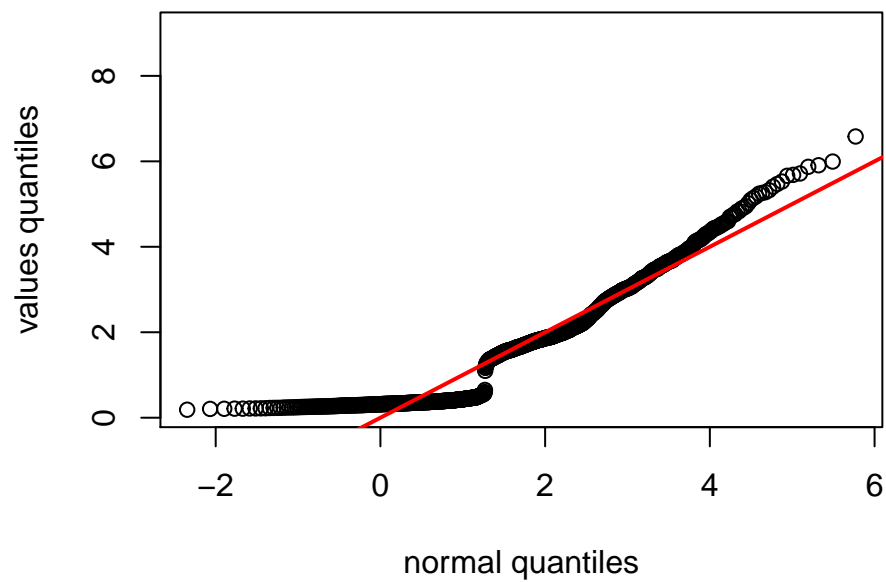


**ANSWER:** Since the bootstrapped distribution of null t-values is close to the diagonal line, we can assume that it is normally distributed. Hence, we don't need to use bootstrapping for the hypothesis test.

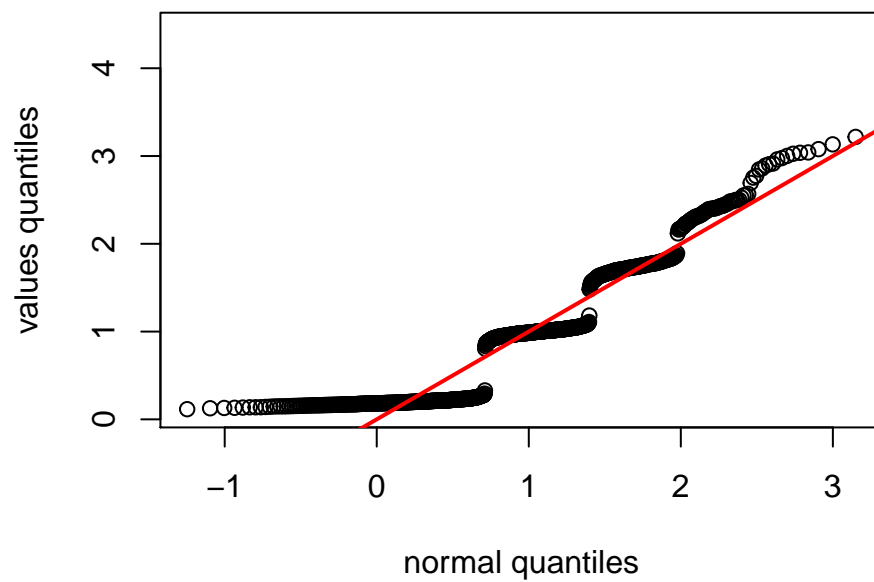
d. Check the two samples in question 2 with QQ plot

```
norm_qq_plot(f_alts)
```





```
norm_qq_plot(f_nulls)
```



**ANSWER:** The alternative and null values of the f-test are not close to the diagonal line, instead they are

heavily concentrated in some areas, we can interpret it as heavily tailed. Hence, it is not viable for us to use the f-test in question 2.