

BACS_HW_6

109078505

3/31/2021

Significantly Help: 109078513

Question 1. Verizon claims that mean response time for ILEC and CLEC customers are the same, but the PUC would like to test if CLEC customers were facing greater response times.

```
#Import data
verizon <- read.csv("verizon.csv", header=TRUE)
mean(verizon$Time)
```

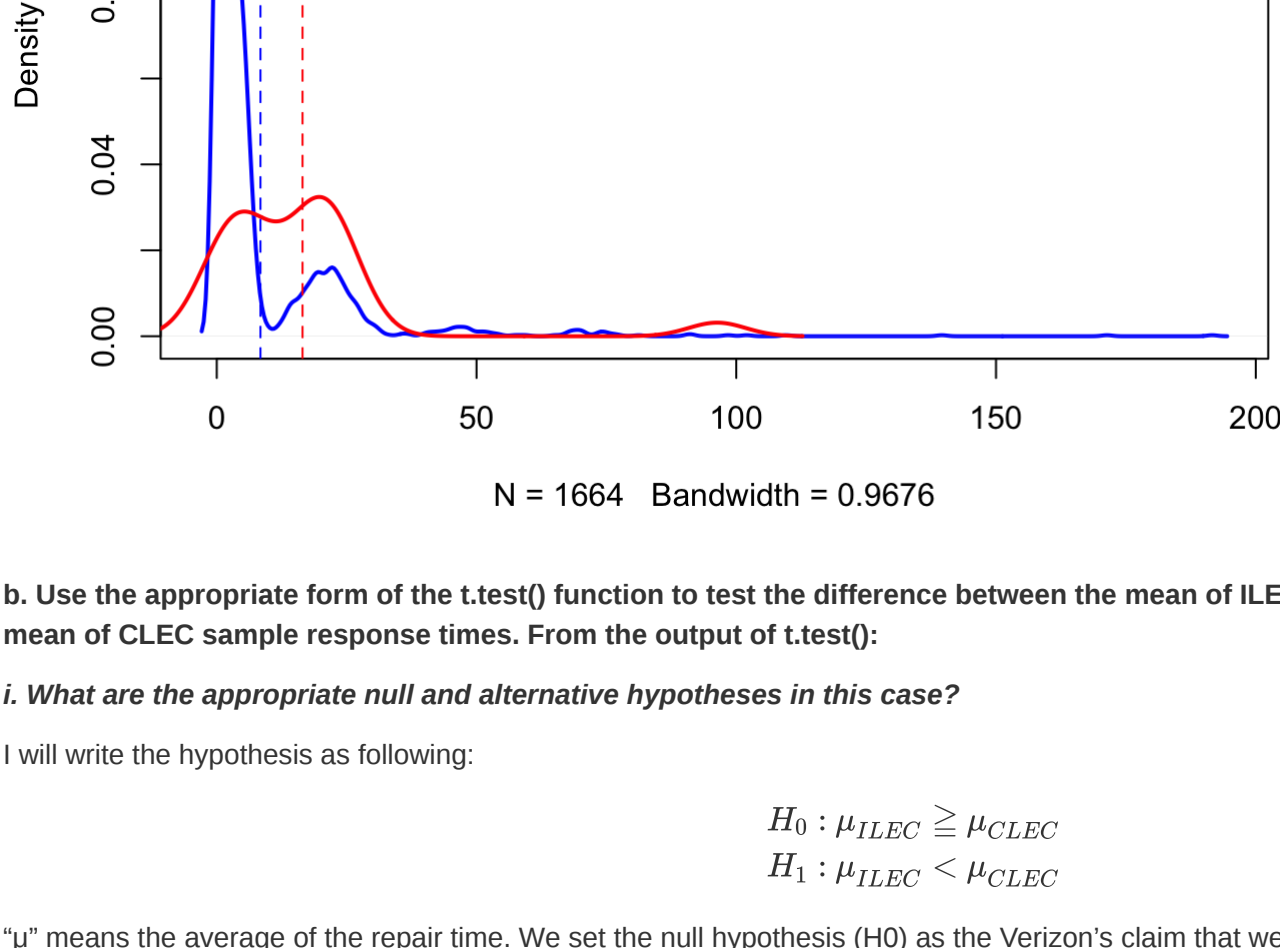
```
## [1] 8.522009
```

```
#Separate ILEC & CLEC
ILEC <- subset(verizon,Group=="ILEC")
ILEC <- ILEC$Time
CLEC <- subset(verizon,Group=="CLEC")
CLEC <- CLEC$Time
```

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

```
#Visualize ILEC & CLEC response times
plot(density(ILEC), col="blue", lwd=2, main = "ILEC & CLEC group comparison chart")
lines(density(CLEC), col="red", lwd=2)
abline(v=mean(ILEC), col="blue", lty=2)
abline(v=mean(CLEC), col="red", lty=2)
```

ILEC & CLEC group comparison chart



b. Use the appropriate form of the t.test() function to test the difference between the mean of ILEC sample response times versus the mean of CLEC sample response times. From the output of t.test():

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

I will write the hypothesis as following:

$$H_0: \mu_{ILEC} \geq \mu_{CLEC}$$
$$H_1: \mu_{ILEC} < \mu_{CLEC}$$

"u" means the average of the repair time. We set the null hypothesis (H0) as the Verizon's claim that we want to reject. We set the alternative hypothesis (H1) as what PUC want to prove.

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

```
t.test(ILEC, CLEC, alternative="less", conf.level = 0.95)
```

```
##
## Welch Two Sample t-test
## data: ILEC and CLEC
## t = -1.0834, df = 22.346, p-value = 0.02987
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -1.091721
## sample estimates:
## mean of x mean of y
## 8.411611 16.509130
```

We will reject null hypothesis, since the p-value is 0.02987, which is smaller than 0.05, under 95% confidence interval.

c. Let's try this using bootstrapping: Estimate bootstrapped null and alternative values of t by using the same t.test() function to compare: bootstrapped samples of ILEC against bootstrapped samples of CLEC (alt t-values); and bootstrapped samples of ILEC against the original ILEC sample (null t-values).

```
#set function to compute null t-values and alternative t-values
bootstrap_null_alt <- function(sample0, sample1) {
  #re-sample ILEC
  resample0 <- sample(sample0, length(sample0), replace=TRUE)
  #re-sample CLEC
  resample1 <- sample(sample1, length(sample1), replace=TRUE)
  #compute t-values
  t_stat_alt <- t.test(resample1, resample0, alternative="greater", var.equal=FALSE, conf.level = 0.99)
  t_stat_null <- t.test(resample0, sample0, var.equal=FALSE, paired=TRUE)
  c(t_stat_alt$statistic, t_stat_null$statistic)
}
```

```
#re-sample for 1000 times
set.seed(42)
boot_t_stats <- replicate(2000, bootstrap_null_alt(ILEC, CLEC))
```

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 (use the same one-vs-two tail logic as 1b).

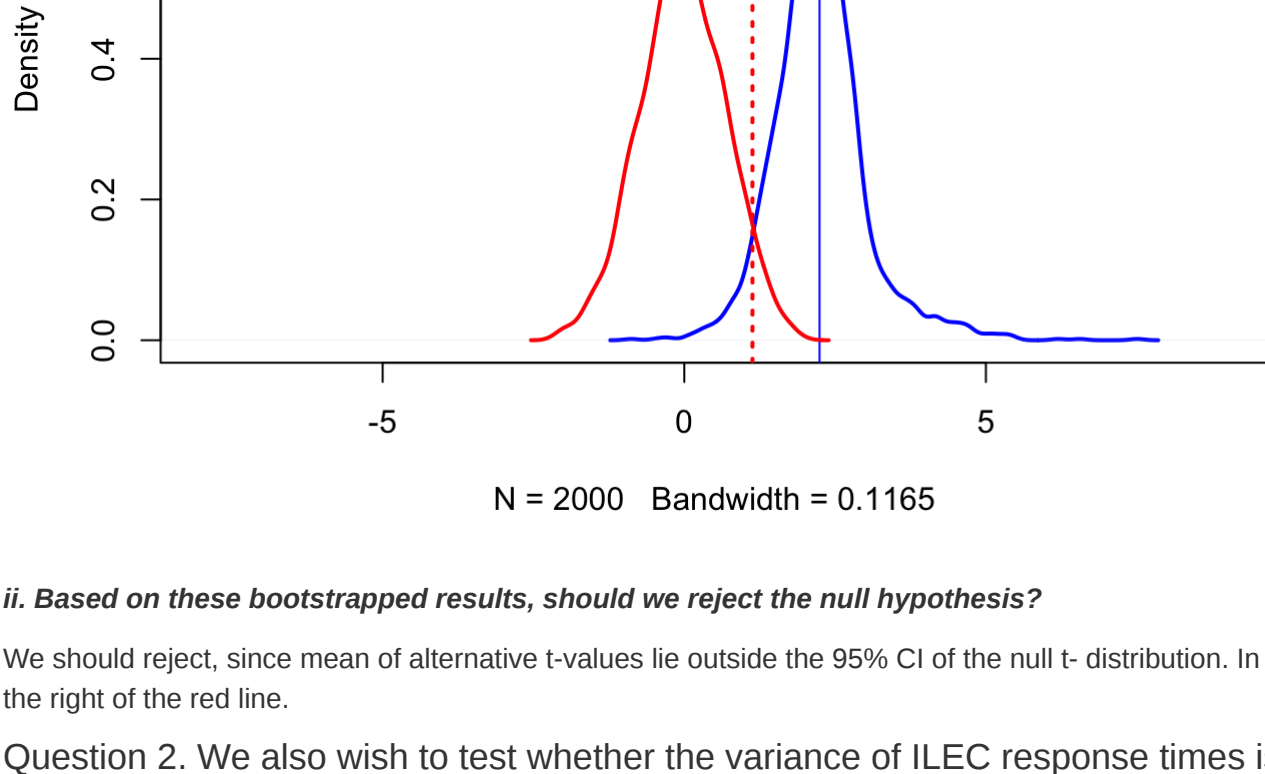
```
#plot null t-values
t_alt <- boot_t_stats[1,]
plot(density(t_alt), xlim=c(-8, 0), ylim=c(0,0.8), col="blue", lwd=2, main = "null t-values and alternative t-values")
abline(v=mean(t_alt), col = "blue")

#plot alternative t-values
t_null <- boot_t_stats[2,]
lines(density(t_null), col="red", lwd=2)

#Left-tail 5% rejection zone
ci_95 <- quantile(t_null, probs=c(0.95))
abline(v=ci_95, lty="dotted", col= "red", lwd=2)

#add legend
legend(y = 0.8, x= -8,
       legend = c("Bootstrapped null t-values", "Bootstrapped alternative t-values"), col = c("red", "blue"),
       bty = "n", lwd = c(2, 2), lty = c(1, 1), cex=0.7, x.intersp=0.6, y.intersp=0.7, text.width=1)
```

null t-values and alternative t-values



ii. Based on these bootstrapped results, should we reject the null hypothesis?

We should reject, since mean of alternative t-values lie outside the 95% CI of the null t- distribution. In other words, most of blue line values are on the right of the red line.

Question 2. We also wish to test whether the variance of ILEC response times is different than the variance of CLEC response times.

a. What is the null and alternative hypotheses in this case?

```
#compute ILEC variance
var(ILEC)
```

```
## [1] 215.7973
```

```
#compute CLEC variance
var(CLEC)
```

```
## [1] 380.3895
```

Since CLEC has bigger variance than ILEC, I will write the hypothesis as following:

$$H_0: S_{ILEC} \leq S_{CLEC}$$
$$H_1: S_{ILEC} > S_{CLEC}$$

b. Let's try traditional statistical methods first:

i. What is the F-statistic of the ratio of variances?

```
#compute F-statistic
var(CLEC)/var(ILEC)
```

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

The F-statistic is 1.762717.

ii. What is the cut-off value of F (reject the 5% most extreme F-values)?

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

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

The cut-off value is 1.548476.

iii. Can we reject the null hypothesis?

Yes, since our F-value (1.762717) is beyond 95% Critical F-value (1.548476).

c. Let's try bootstrapping this time:

```
#set function to compute null F-values and alternative F-values.
set.seed(43)
sd_providers_test <- function(larger_sd_sample, smaller_sd_sample) {
  #re-sample
  resample_larger_sd <- sample(larger_sd_sample, length(larger_sd_sample), replace=TRUE)
  resample_smaller_sd <- sample(smaller_sd_sample, length(smaller_sd_sample), replace=TRUE)
  #compute f-value
  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)
}
```

```
f_stats <- replicate(10000, sd_providers_test(CLEC, ILEC))

# Create bootstrapped values of the F-statistic, for both null and alternative hypotheses.
```

```
#F-statistic for null
f_nulls <- f_stats[2,]
#F-statistic for alternative
f_alts <- f_stats[1,]
```

ii. What is the 95% cutoff value according to the bootstrapped null values of F?

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

```
##      95%
## 2.331735
```

The 95% cutoff value is 2.331735.

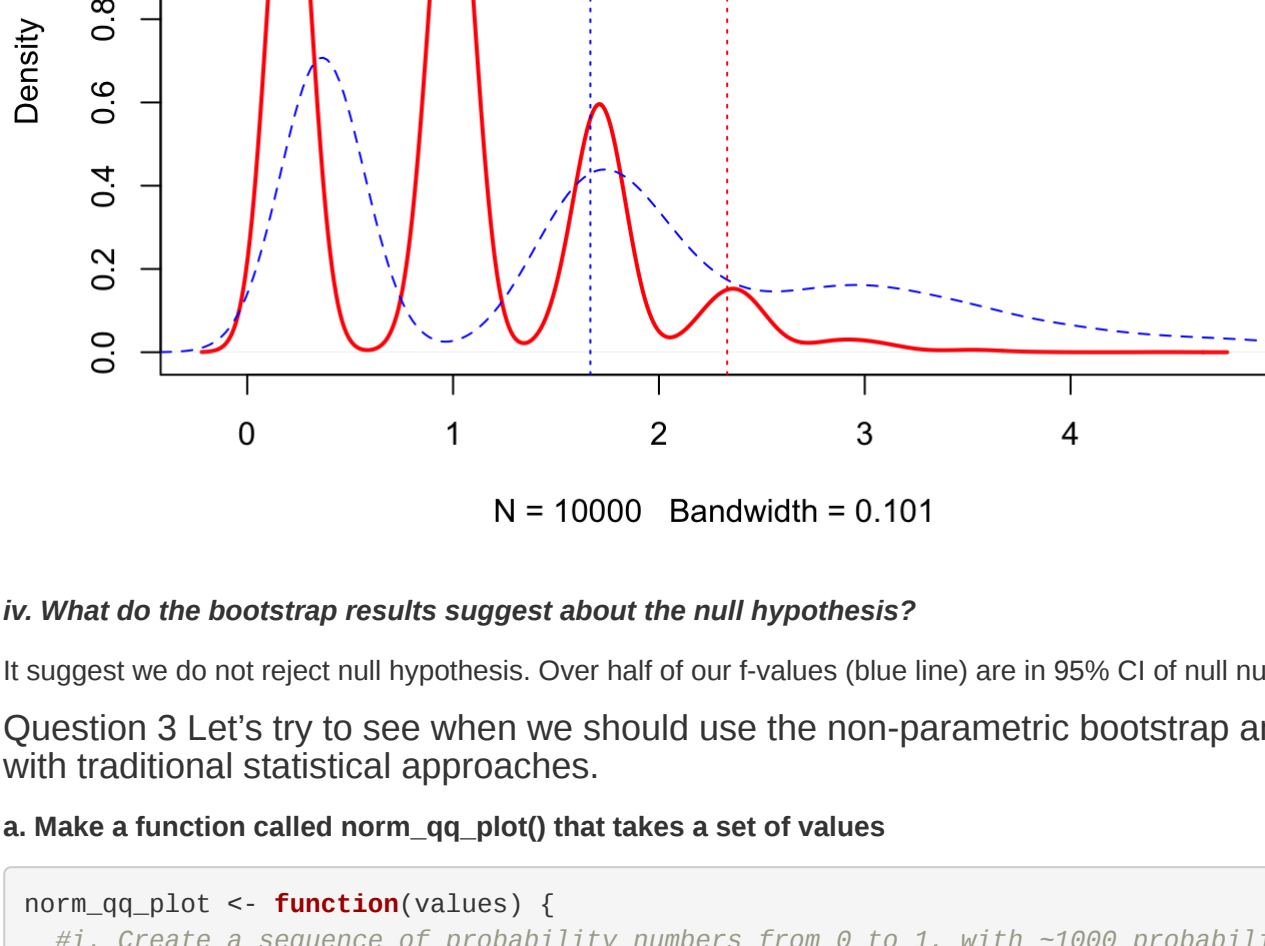
iii. Plot a visualization of the null and alternative distributions of the bootstrapped F-statistic, with vertical lines at the cutoff value of F nulls.

```
#plot null F-values
plot(density(f_nulls), lwd=2, col="red")

#plot alternative t-values
lines(density(f_alts), lty="dashed", col="blue" )

#95% cutoff value line
ci_95 <- quantile(f_nulls, probs=c(0.95))
abline(v=ci_95, lty="dotted", col= "red")
abline(v=median(f_alts), lty="dotted", col= "blue")
```

density.default(x = f_nulls)



iv. What do the bootstrap results suggest about the null hypothesis?

Question 3 Let's try to see when we should use the (blue line) are in 95% CI of null null hypothesis (red line).

I suggest we do not reject null hypothesis. Over half of our t-values are in 95% CI of null null hypothesis (red line).

It suggests that we should use the values (blue line) are in 95% CI of null null hypothesis (red line) and when we might be better off with traditional statistical approaches.

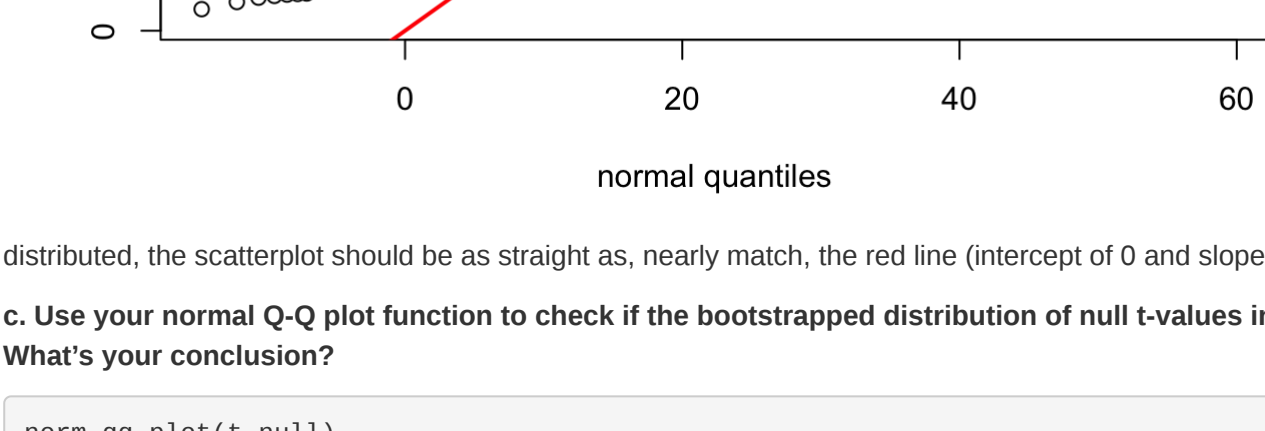
a. Make a function called norm_qq_plot() that takes a set of values

```
norm_qq_plot <- function(values) {
  #1. 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
  q_vals <- quantile(values, probs=probs1000)
  #iii. Calculate ~1000 quantiles of a normal distribution as our values # 某百分位的z-score是幾
  q_norm <- qnorm(probs1000, mean = mean(values), sd = sd(values))
  #iv. Create a scatterplot (quantiles of 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
  abline(a=0, b=1, col="red", lwd=2)
}
```

b. Confirm that your function works

```
set.seed(978234)
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)
norm_qq_plot(d123)
```

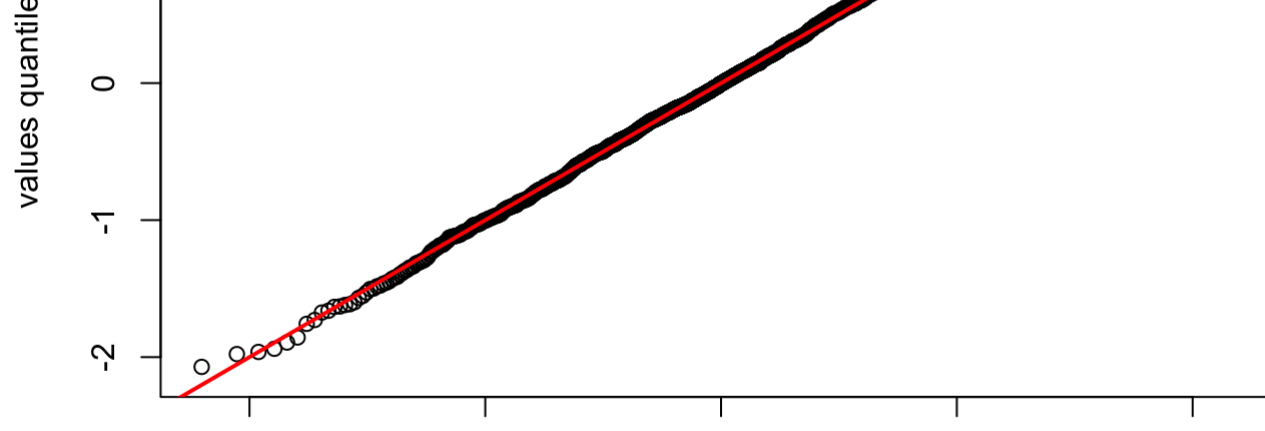
If a distribution is perfectly normally



distributed, the scatterplot should be as straight as, nearly match, the red line (intercept of 0 and slope of 1).

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(t_null)
```



From the Q-Q plot, the bootstrapped distribution of null t-values in question 1c was normally distributed.

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)
```

```
norm_qq_plot(CLEC)
```

```
norm_qq_plot(ILEC)
```

```
norm_qq_plot(CLEC)
```

```
norm_qq_plot(ILEC)
```

```
norm_qq_plot(CLEC)
```

```
norm_qq_plot(ILEC)
```

```
norm_qq_plot(CLEC)
```

```
norm_qq_plot(ILEC)
```

```
norm_qq_plot(CLEC)
```

```
norm_qq_plot(ILEC)
```

```
norm_qq_plot(CLEC)
```

```
norm_qq_plot(ILEC)
```

```
norm_qq_plot(CLEC)
```

```
norm_qq_plot(ILEC)
```

```
norm_qq_plot(CLEC)
```

```
norm_qq_plot(ILEC)
```

```
norm_qq_plot(CLEC)
```