Module Assignment
Module 2
QMB-6304 Foundations of Business Statistics

Write a simple R script to execute the following:

Preprocessing:

Load the data in "6304 Assignment 2 Data.xlsx" into an object. The file includes 10000 observations for each of four variables, which are creatively named for the four Marx Brothers.

```
#Carolina Aldana Yabur
#UID: U25124553

rm(list=ls())
install.packages("rio")
install.packages("moments")
library(rio)
library(moments)
mydata=import("6304 Module 2 Assignment Data.xlsx")
colnames(mydata)=tolower(make.names(colnames(mydata)))
attach(mydata)
```
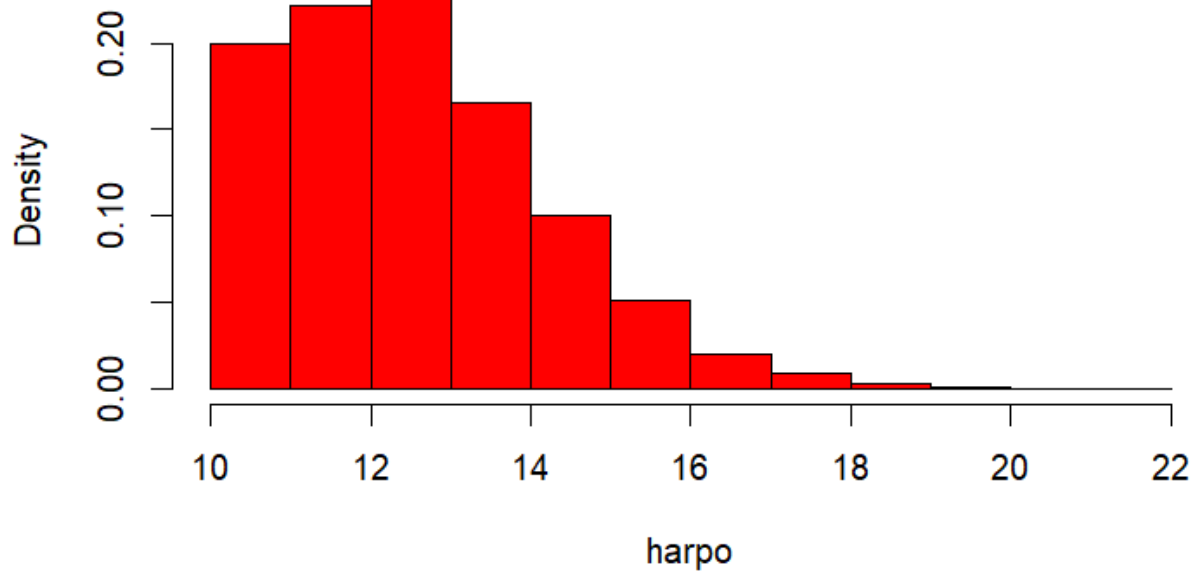
Analysis:

1. Use common tools to determine whether any of the four variables are normally distributed. Explain how you arrived at your conclusions.
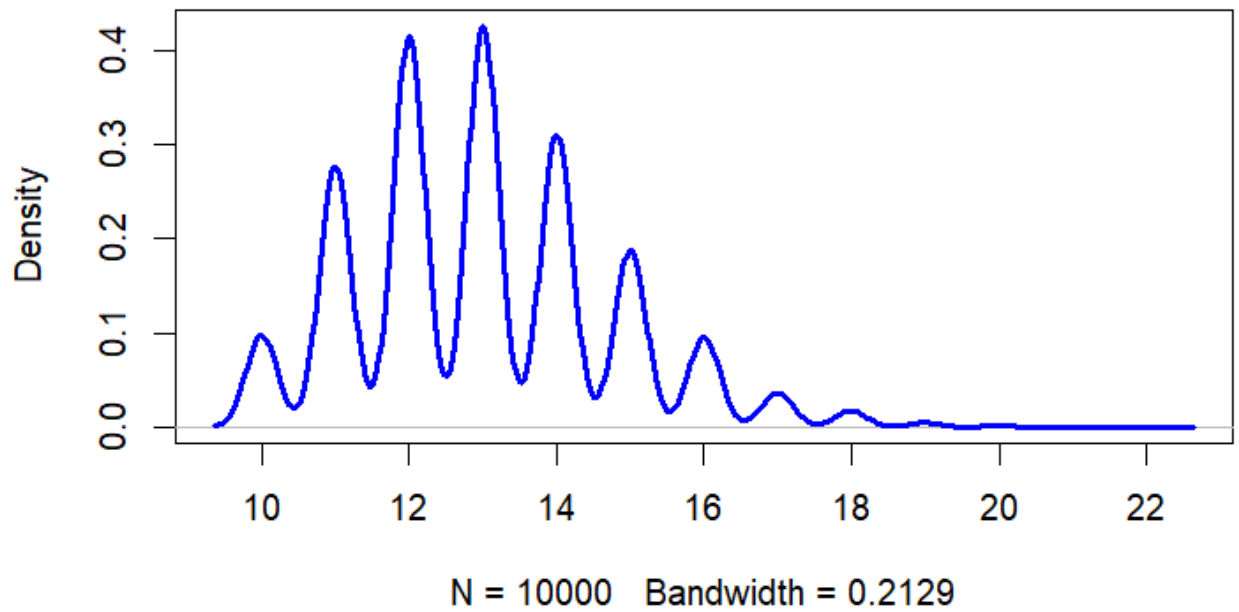
```
#Histogram and Density Plot
hist(harpo, col="red", main= "Harpo Histogram",
probability=TRUE)
plot(density(harpo),lwd=3, main="Harpo Density plot")
lines(density(harpo),lwd=3,col="blue")
```
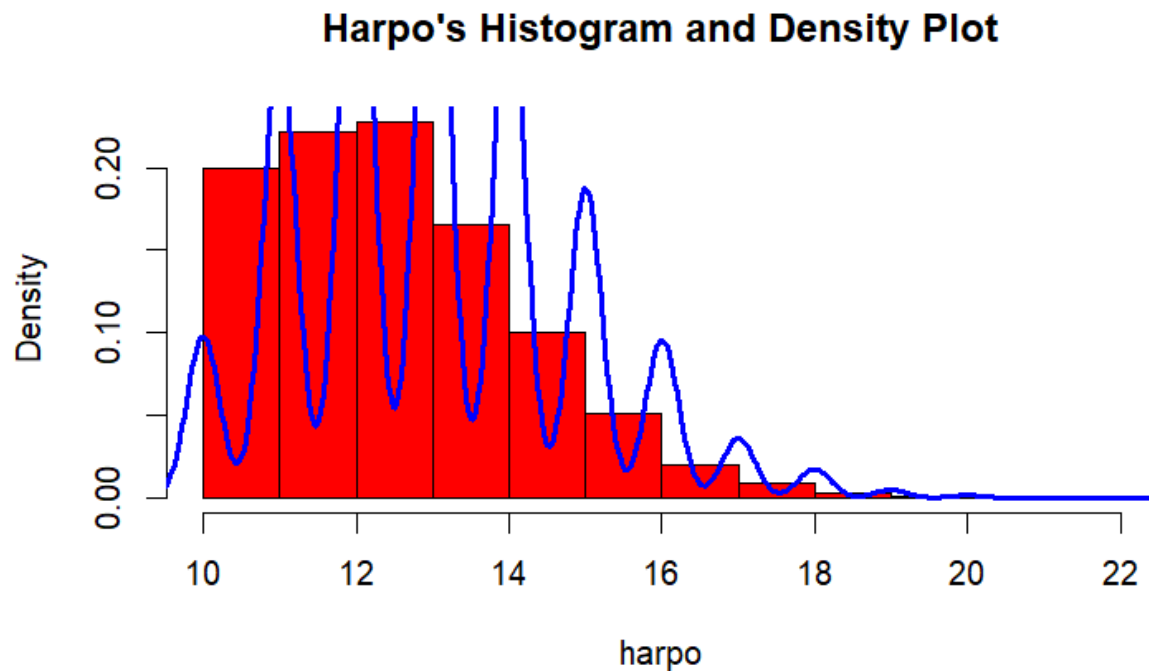
Harpo Histogram

Harpo Density plot

N = 10000   Bandwidth = 0.2129

```
hist(harpo, col="red", main= "Harpo's Histogram and Density
Plot", probability=TRUE)
lines(density(harpo), lwd=3, col="blue")
```

Harpo's Histogram and Density Plot

The histogram and density plot of the variable Harpo suggest that the distribution is not normal, since both histogram and density plot show that the bulk of data is on the left and the longer tail for both is on the right side, appearing right skewed.

```
#QQ Plot
qqnorm(harpo, pch=19, main="Harpo QQ Plot")
qqline(harpo, col="red", lwd=3)
```
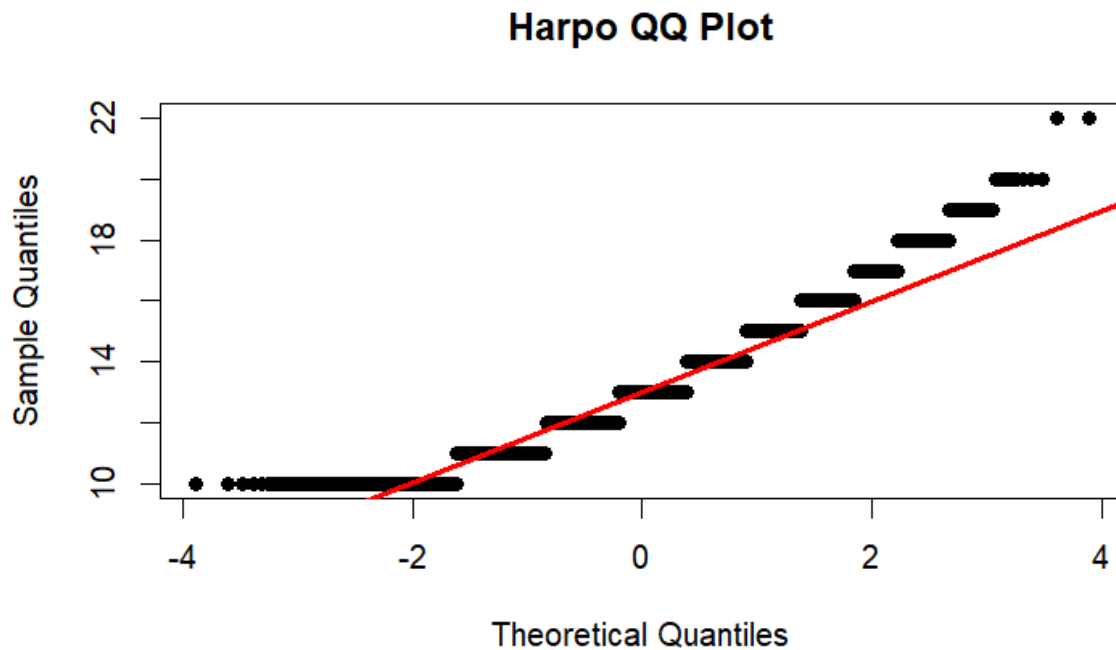
Harpo QQ Plot

```
> #Objective Measures
> skewness(harpo)
[1] 0.5786532
> kurtosis(harpo)
[1] 3.364361
```

The skewness is greater than 0 which suggests that the data points are concentrated on the left side. On the other hand, the kurtosis is greater than 3, which suggests that the distribution is leptokurtic (heavy tails and sharp peak) and not normal.

```
#Zeppo
#Histogram and Density Plot
hist(zeppo, col="red", main= "Zeppo Histogram",
probability=TRUE)
plot(density(zeppo),lwd=3, main="Zeppo Density plot")
lines(density(zeppo),lwd=3,col="blue")
```

Zeppo Histogram

Zeppo Density plot

N = 10000   Bandwidth = 1.426
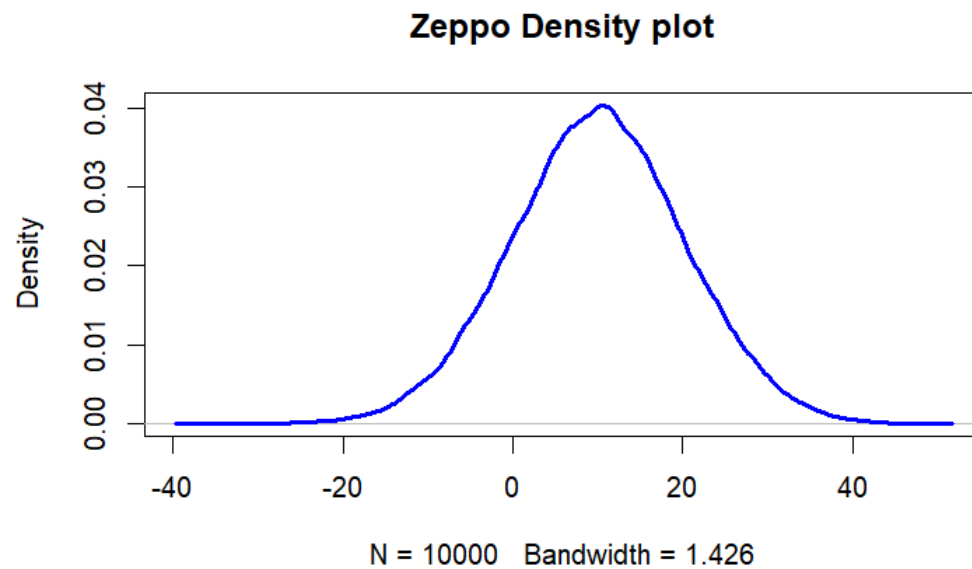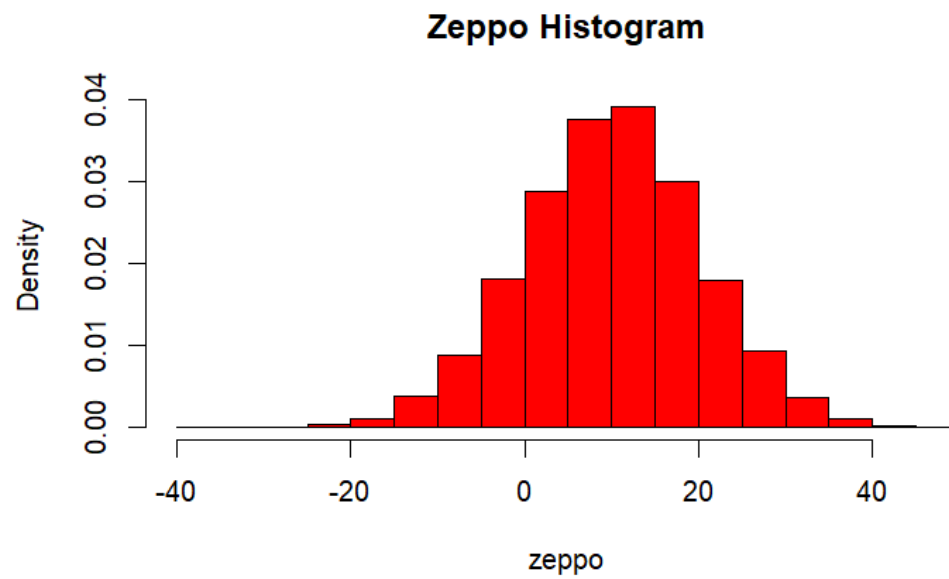
```
hist(zeppo, col="red", main= "Zeppo's Histogram and Density
Plot", probability=TRUE)
lines(density(zeppo), lwd=3, col="blue")
```

Zeppo's Histogram and Density Plot
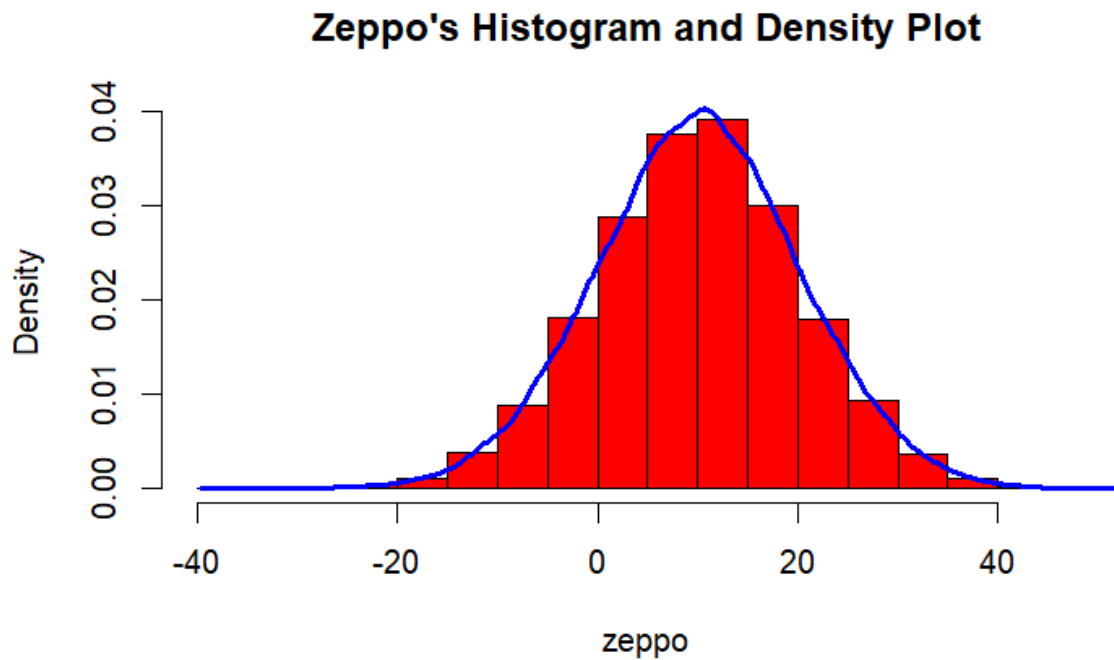
The histogram and density plot of the variable Zeppo suggest that the distribution is normal, since the bulk of data is concentrated around the mean of the variable (10.06). Both of them look like a smooth bell curve.

```
#QQ Plot
qqnorm(zeppo, pch=19, main="Zeppo QQ Plot")
qqline(zeppo, col="red", lwd=3)
```
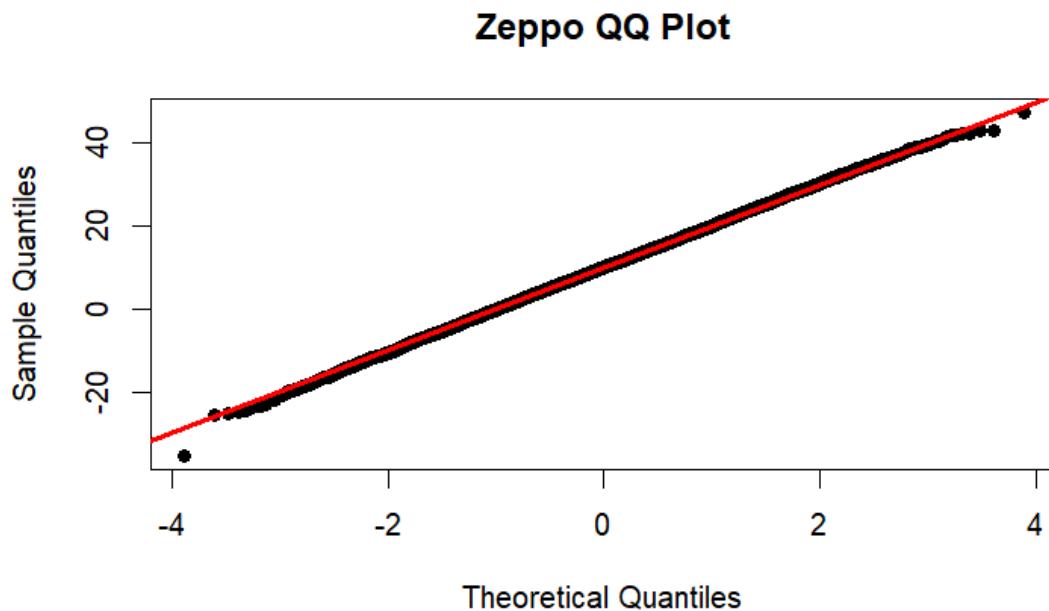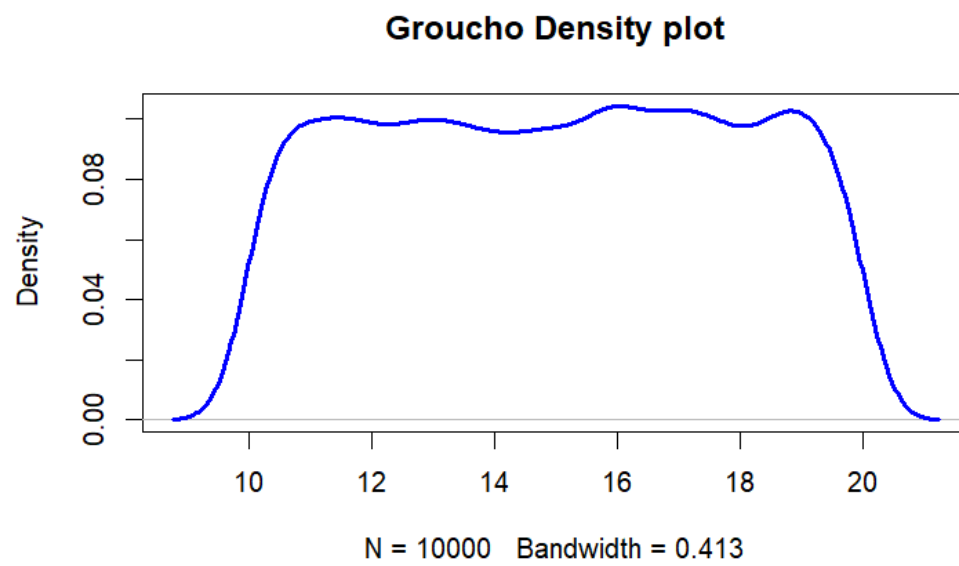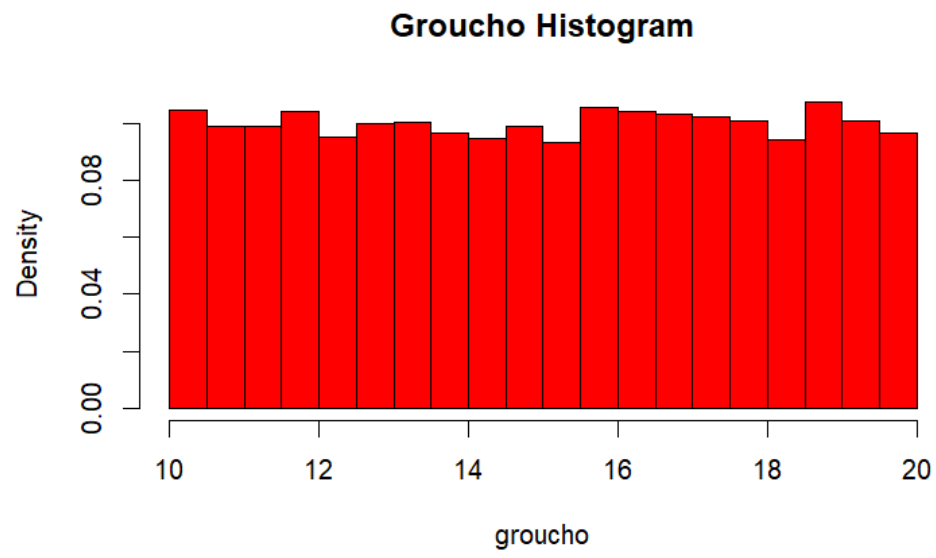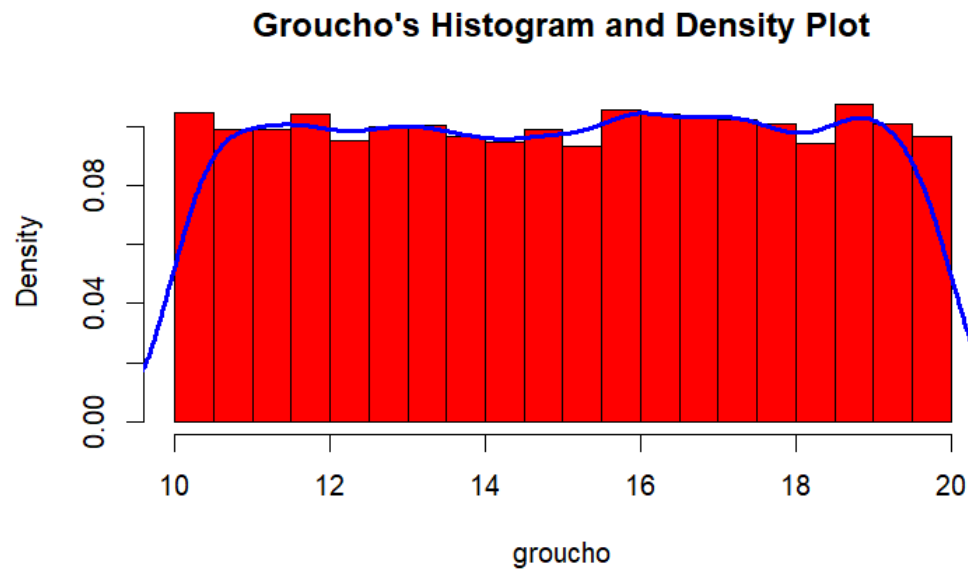
Zeppo QQ Plot

```
> #Objective Measures
> skewness(zeppo)
[1] -0.0245893
> kurtosis(zeppo)
[1] 3.019482
```

The skewness is very close to 0 and suggests that is approximately symmetric, the kurtosis is also very close to 3 which suggests the same thing.

```
#Groucho
#Histogram and Density Plot
hist(groucho, col="red", main= "Groucho Histogram",
probability=TRUE)
plot(density(groucho),lwd=3, main="Groucho Density plot")
lines(density(groucho),lwd=3,col="blue")
```
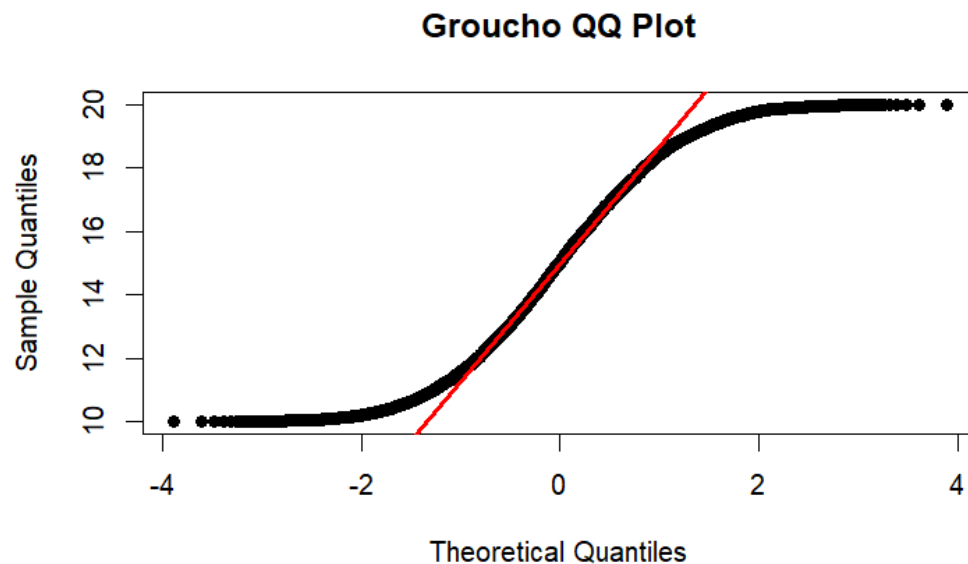
Groucho Histogram

groucho

Groucho Density plot

N = 10000   Bandwidth = 0.413

Groucho's Histogram and Density Plot

```
#QQ Plot
qqnorm(groucho, pch=19, main="Groucho QQ Plot")
qqline(groucho, col="red", lwd=3)
```

Groucho QQ Plot

© 2020 Ronald K. Satterfield

```
> #Objective Measures
> skewness(groucho)
[1] -0.01466599
> kurtosis(groucho)
[1] 1.795703
```

```
#Chico
#Histogram and Density Plot
hist(chico, col="red", main= "Chico Histogram",
probability=TRUE)
plot(density(chico),lwd=3, main="Chico Density plot")
lines(density(chico),lwd=3,col="blue")
```
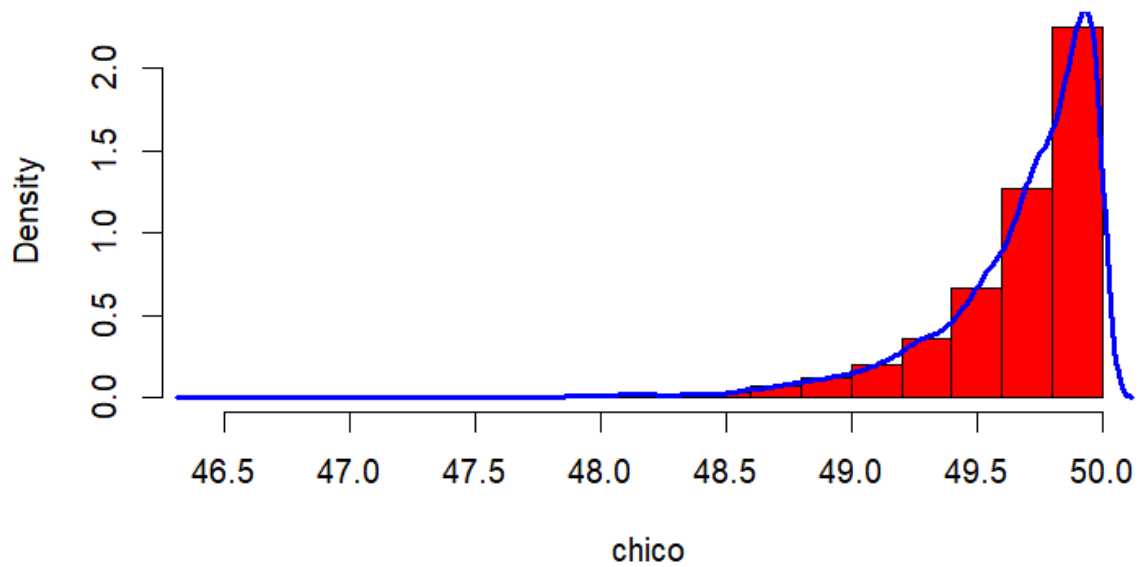

Chico Histogram

Chico Density plot

N = 10000   Bandwidth = 0.03822

```
hist(chico, col="red", main= "Chico's Histogram and Density
Plot", probability=TRUE)
lines(density(chico), lwd=3, col="blue")
```
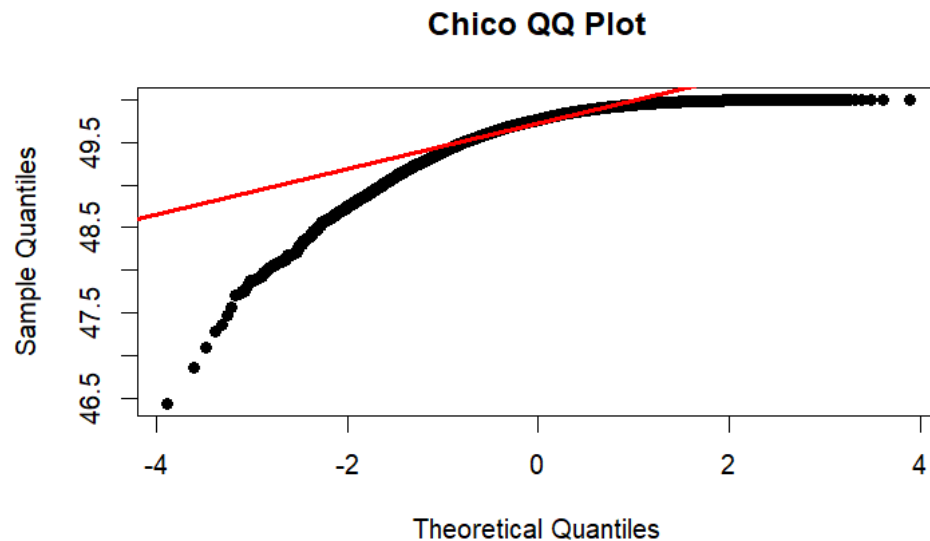
Chico's Histogram and Density Plot

```
#QQ Plot
qqnorm(chico, pch=19, main="Chico QQ Plot")
qqline(chico, col="red", lwd=3)
```

Chico QQ Plot

```
> #Objective Measures
> skewness(chico)
[1] -2.031371
> kurtosis(chico)
[1] 9.268749
```

2. Focus on the Chico variable. Build a sampling distribution of the population mean by calculating the mean for each of 1000 samples of n=40. Plot the sampling distribution with an appropriate graph title. Verify whether these 1000 means are in fact normally distributed and justify your conclusion with appropriate analytics and/or graphical tools.

```
set.seed(1000)
n=40
brothers_sample=rep()
for (x in 1:1000){
  brothers_sample[x]=mean(sample(chico,40))
  hist(brothers_sample, col="red",
       main= "Sample Means of Variable Data",
       xlab="Sample Means")
  plot(density(brothers_sample), lwd=3,
       main= "Density Plot of Variable Samples")
  hist(brothers_sample, col="red",
       main="Density - Histogram of Sample Data",
       xlab= "Sample Means", probability= TRUE)
  lines(density(brothers_sample), lwd=3, col="blue")
  qqnoem(brothers_sample, col="red", lwd=3,
         main= "QQ Plot of Variable Samples")
  qqline(brothers_sample, col="red", lwd=3)
  skeweness(brothers_sample)
  kurtosis(brothers_sample)
}
```

Your deliverable will be a single MS-Word file showing 1) the R script which executes the above instructions and 2) the results of those instructions. The first line of your script file should be a "#" comment line showing your name as it appears in Canvas. Results should be presented in the order in which they are listed here. Deliverable due time will be announced in class and on Canvas. **This is an individual assignment to be completed before you leave the classroom. No collaboration of any sort is allowed on this assignment.**