Analytical Methods for Business
Pablo Zumba

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 the four variables, which are creatively named for the four Marx Brothers.

```
#Pablo Zumba
rm(list=ls())
library(rio)
marksBrothers = as.data.frame(import("6304 Module 2 Assignment Data.xlsx"))
attach(marksBrothers)
```

Analysis:

1. Use common tools to determine whether any of the four variables are normally distributed. Explain how you arrived at your conclusions.

```
#Part 1. Determining Normality
isNormalDist <- function(data_sample, colorGraph, titleGraph=NULL){
  layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
  hist(data_sample, col=colorGraph, main=titleGraph, probability = TRUE)
  lines(density(data_sample),lwd=3,col="Black")
  boxplot(data_sample,col=colorGraph, main="BoxPlot Distribution")
  qqnorm(data_sample, pch=19, main="QQplot compared to QQline")
  qqline(data_sample, lwd=3, col='red')
  print("Skewness=")
  print(moments::skewness(data_sample))
  print("Kurtosis=")
  print(moments::kurtosis(data_sample))
}
```

```
isNormalDist(Harpo,"Green", "Harpo Distribution")
> isNormalDist(Harpo,"Green", "Harpo Distribution")
[1] "Skewness="
[1] 0.5786532
[1] "Kurtosis="
[1] 3.364361
```

Harpo Distribution

BoxPlot Distribution

QQplot compared to QQline
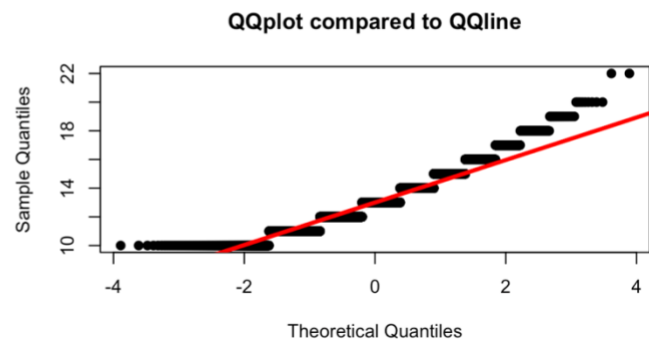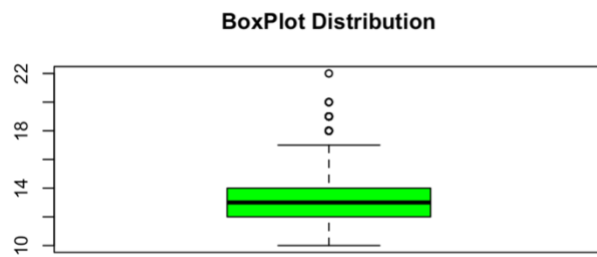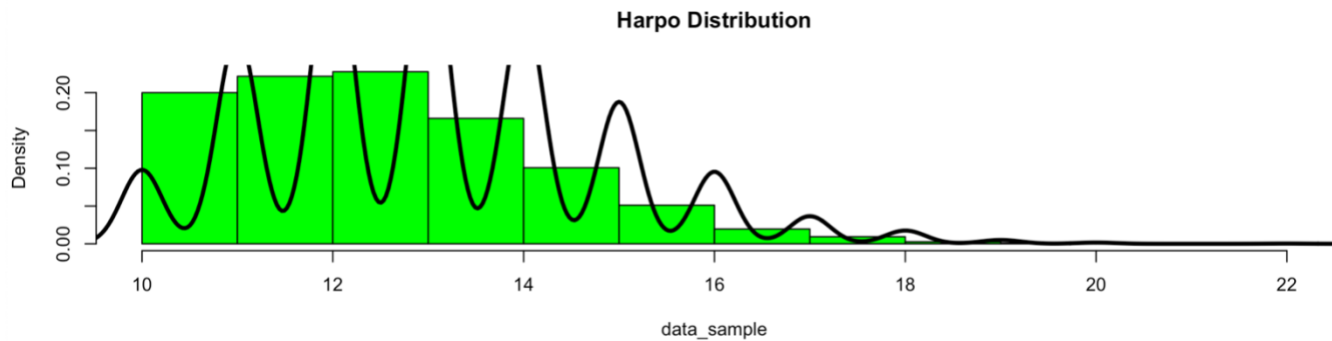
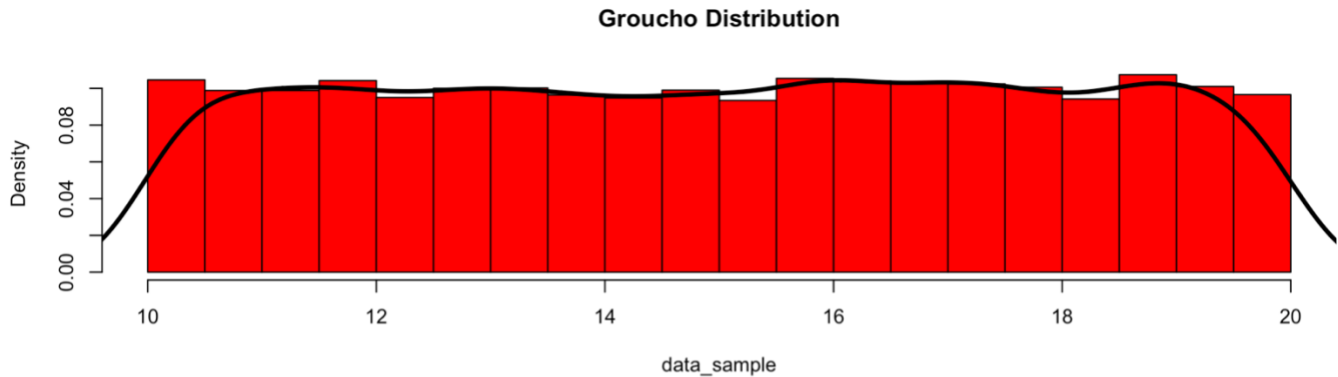"Harpo" data is NOT normally distributed but tends to normality. Based on the Histogram, the "Harpo" data seems to be right skewed which can be corroborated by looking at the skewness value of 0.579. The boxplot tells us there are 4 outliers and a possible tendency to normality due to the position of the median and the quantiles of the data compared to the theoretical quantiles show they follow the QQline but not perfectly.

```
isNormalDist(Zeppo,"Blue", "Zeppo Distribution")
> isNormalDist(Zeppo,"Blue", "Zeppo Distribution")
[1] "Skewness="
[1] -0.0245893
[1] "Kurtosis="
[1] 3.019482
```


Zeppo Distribution

BoxPlot Distribution

QQplot compared to QQline
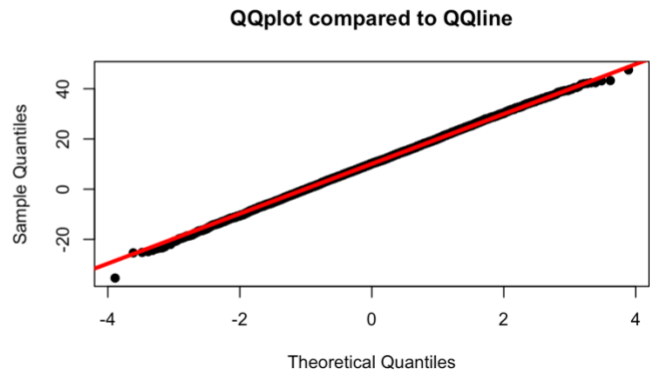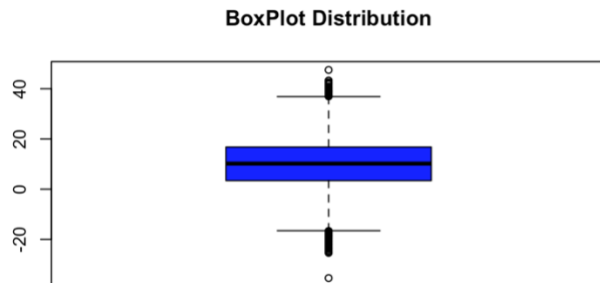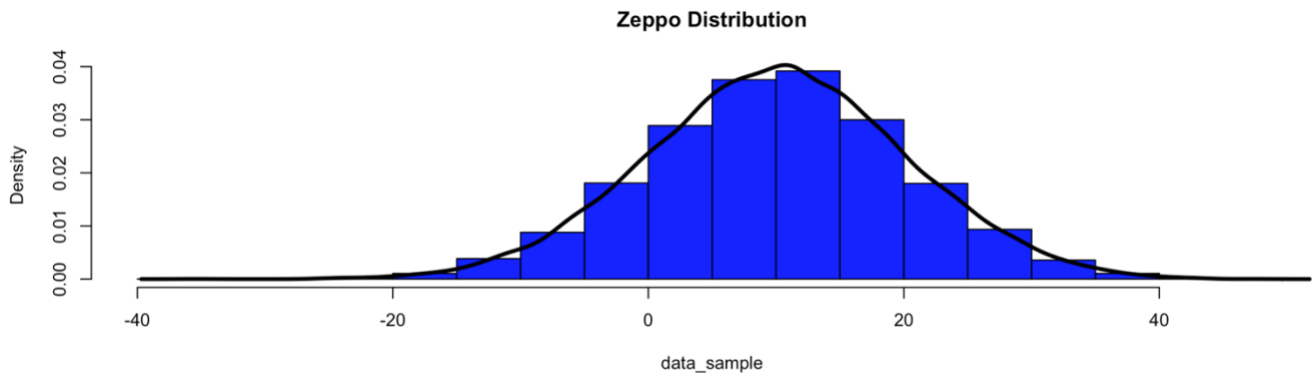
Based on the Histogram, the "Zeppo" data seems slightly skewed to the left (Skewness=-0.023). The boxplot tells us there is 1 outlier on each side and the median is located at the center of the plot, making the plot symmetric and evenly spread. The quantiles of the data and the theoretical quantiles follow the QQline almost perfectly. For those reasons, the "Harpo" data is normally distributed.

```
isNormalDist(Groucho,"Red", "Groucho Distribution")
> isNormalDist(Groucho,"Red", "Groucho Distribution")
[1] "Skewness="
[1] -0.01466599
[1] "Kurtosis="
[1] 1.795703
```


Groucho Distribution

BoxPlot Distribution

QQplot compared to QQline
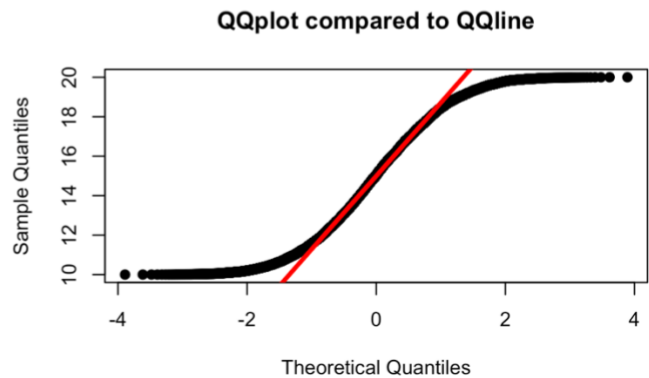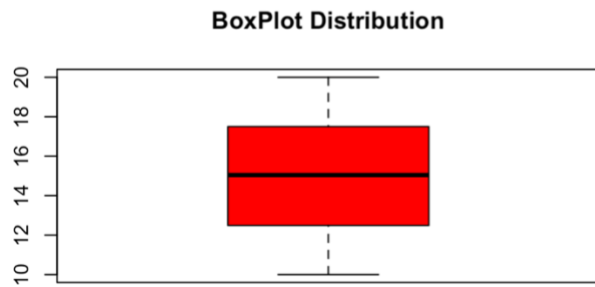
"Groucho" data is normally distributed. Based on the Histogram, the "Groucho" data seems to be evenly flat spread corroborated with a value of Skewness=-0.0147 and Kurtosis=1.79 (far from 3 and not a bell shape). The boxplot is also showing signs of normality due to its shape and the symmetric position of the median line with no outliers. Finally, the quantiles of the data compared to the theoretical quantiles show they follow just an interval of the QQline due to the distributions having no tails.

```
isNormalDist(Chico,"Purple", "Chico Distribution")
> isNormalDist(Chico,"Purple", "Chico Distribution")
[1] "Skewness="
[1] -2.031371
[1] "Kurtosis="
[1] 9.268749
```

Chico Distribution

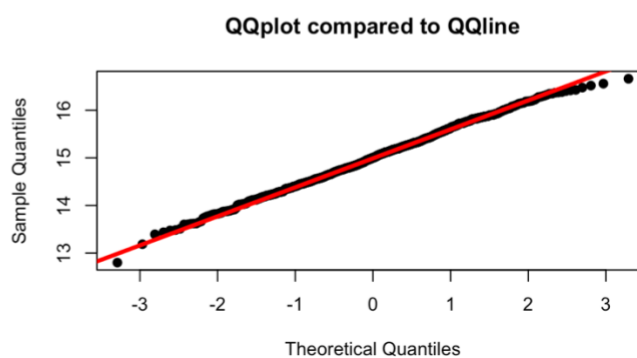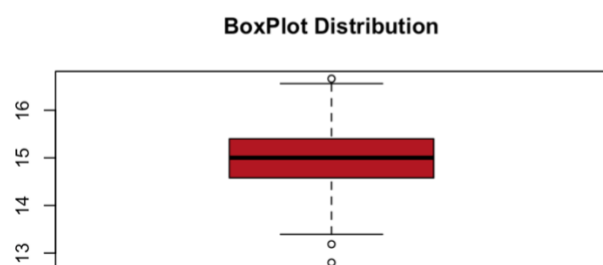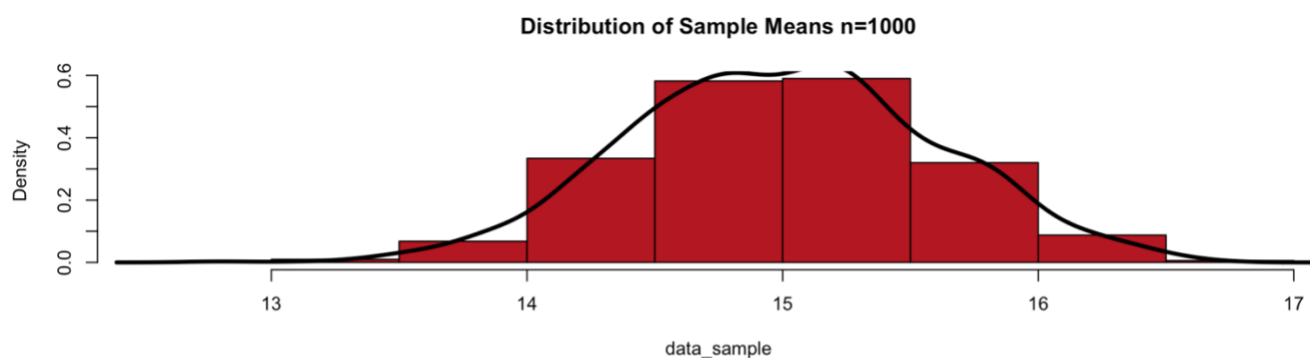**BoxPlot Distribution**

QQplot compared to QQline
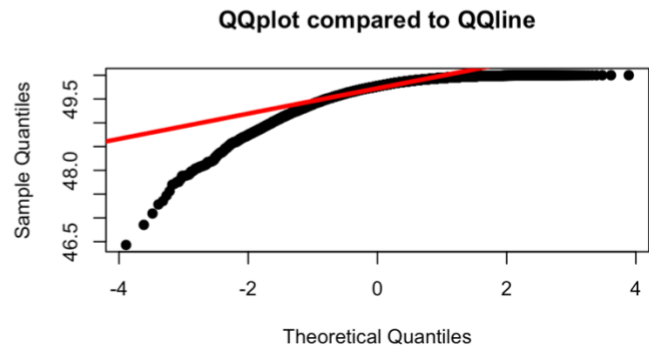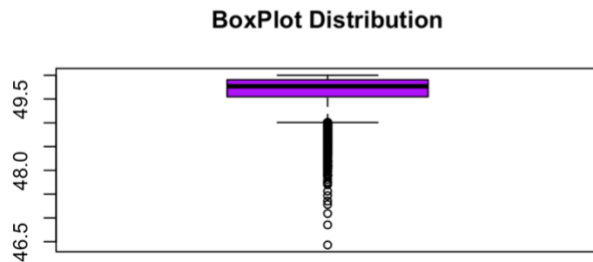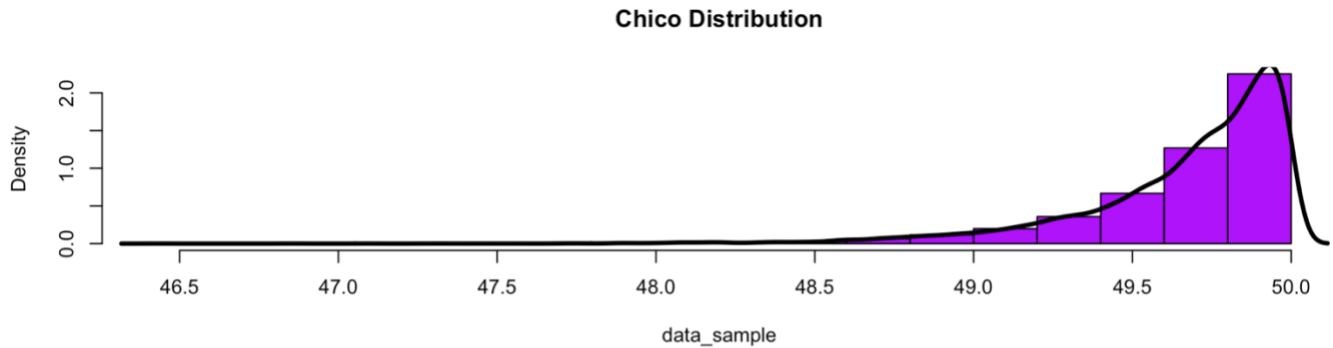
Due to the histogram being left-skewed (Skewness=-2.03), the value of kurtosis not close to 3, the asymmetrical shape of the boxplot with many outliers, and the QQplot not following the theoretical QQline, we can say that the "Chico" data is not normally distributed.

2. Focus on the Groucho variable. Build a sampling distribution of the population mean by:
   a. Taking 1000 samples of n=25.
   b. Calculating the mean for each of the 1000 samples and storing them.
   c. Using the tools you know to show the 1000 sample means are normally distributed.

```
#Part 2. Groucho variable and sampling distributions.
set.seed(54252888)
sample_means = data.frame()
for(i in 1:1000){
  sample_means[i,1] = mean(sample(Groucho,25))
}
attach(sample_means)
isNormalDist(V1, "Brown", "Distribution of Sample Means n=1000")
> isNormalDist(V1, "Brown", "Distribution of Sample Means n=1000")
[1] "Skewness="
```

```
[1] -0.02201137
[1] "Kurtosis="
[1] 2.833495
```

Distribution of Sample Means n=1000

BoxPlot Distribution

QQplot compared to QQline

The histogram shows a bell-shaped distribution with a value of skewness close to 0 and kurtosis close to 3. The box plot distribution is symmetric about the median with only 3 outliers. The quantiles of the data and the theoretical quantiles follow the QQline almost perfectly. For these reasons, we can say the sample means in the "Groucho" data is normally distributed.