**#Name – Manish Sanjay Talekar**

**#U number – U54500765**

**#QMB Assignment 2**

**#Preprocessing**

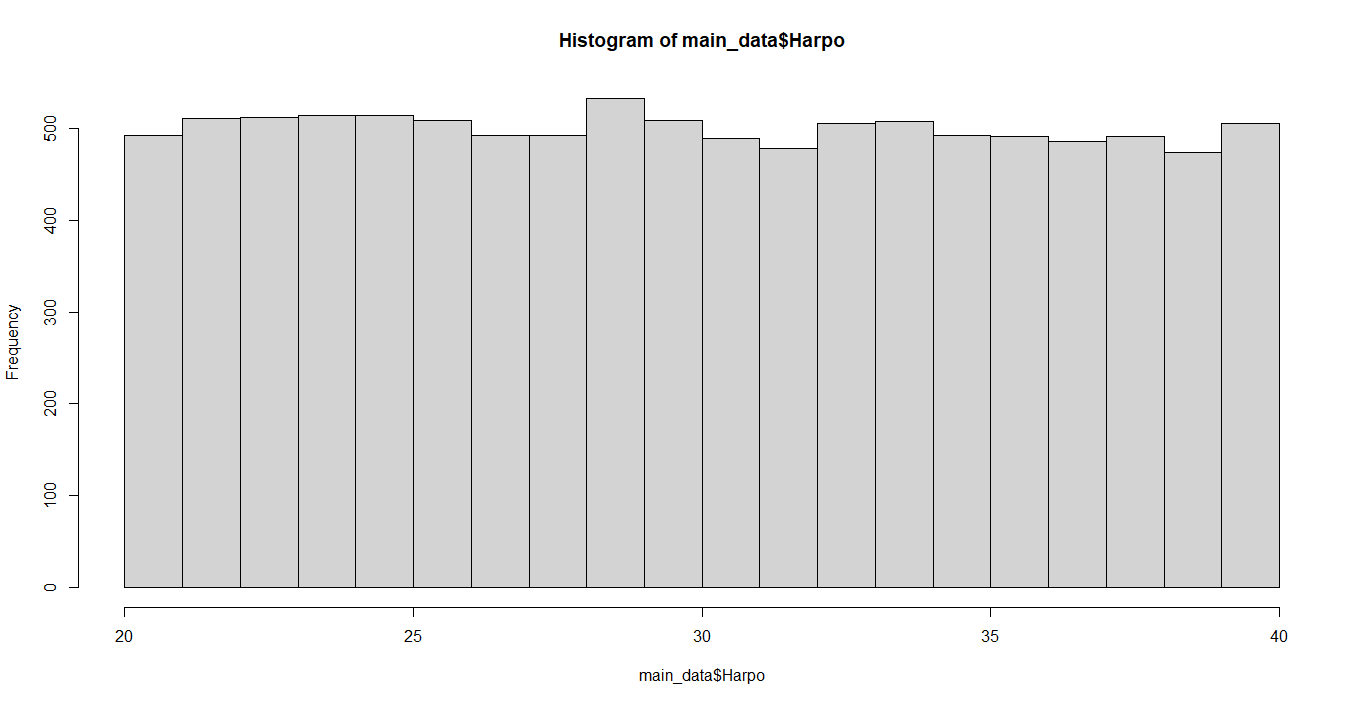
main\_data <- read\_xlsx("6304 Module 2 Assignment Data.xlsx")

**#Analysis**

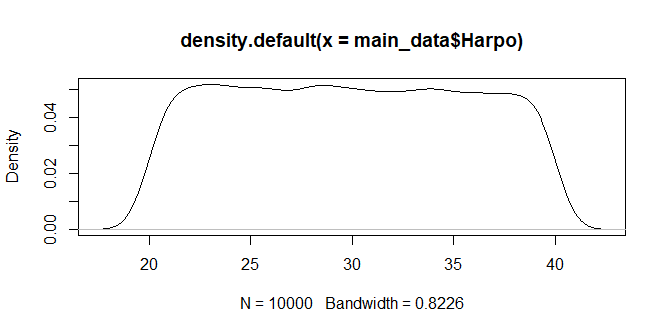
**#question 1**

**#Column 1**

>hist(main\_data$Harpo)

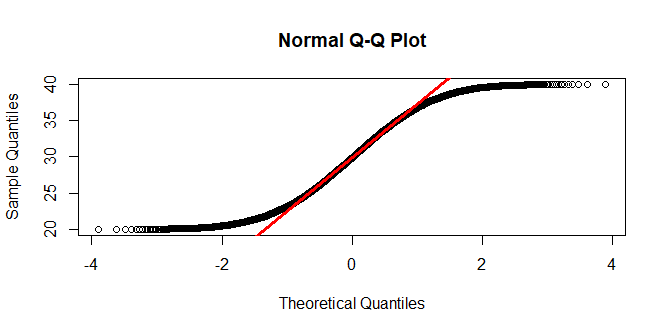


>plot(density(main\_data$Harpo))



>qqnorm(main\_data$Harpo)

>qqline(main\_data$Harpo, col="red", lwd=3)

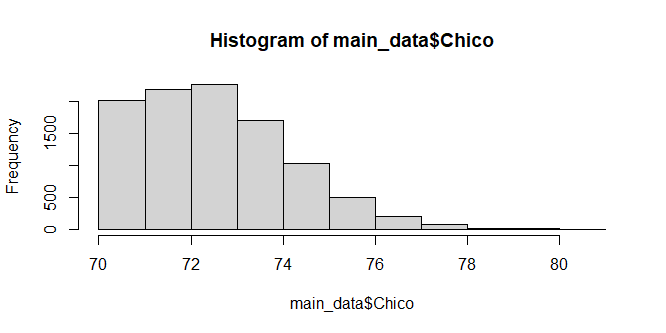


CONCLUSION - After checking histogram and density plot, I would like to conclude that first variable “Harpo” is uniformly distributed. Additionally, qqnorm and qqline shows a lot of data which is beyond the red line this proves that “Harpo” is not normally distributed.

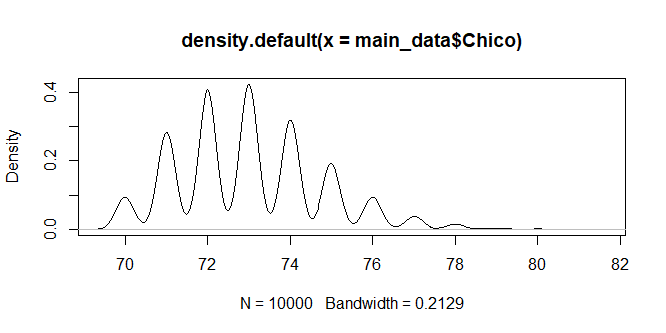
**#question 2** – Data is Uniformly Distributed

**#Column 2**

>hist(main\_data$Chico)

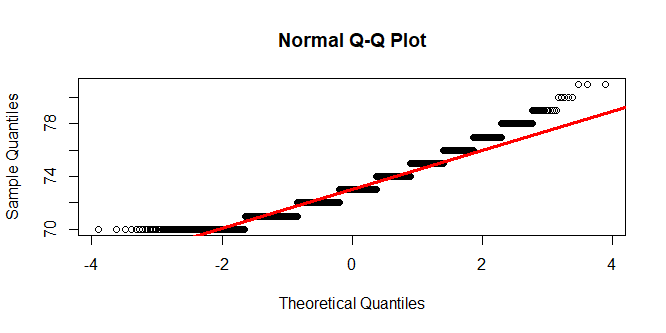


> plot(density(main\_data$Chico))



>qqnorm(main\_data$Chico)

>qqline(main\_data$Chico, col="red", lwd=3)

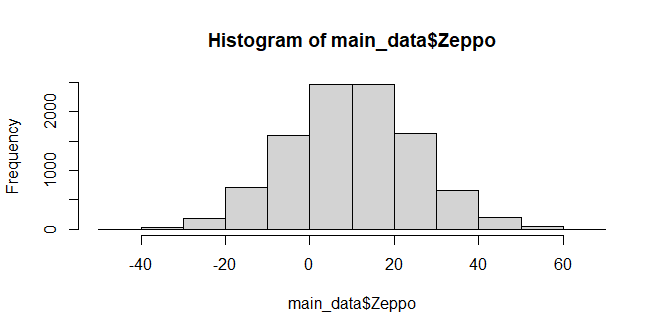


CONCLUSION – Histogram shows the data to be right skewed. Apart from that QQ plot shows that a lot of data is out of the QQ line. Which proves that the second variable “Chico” is not normally distributed.

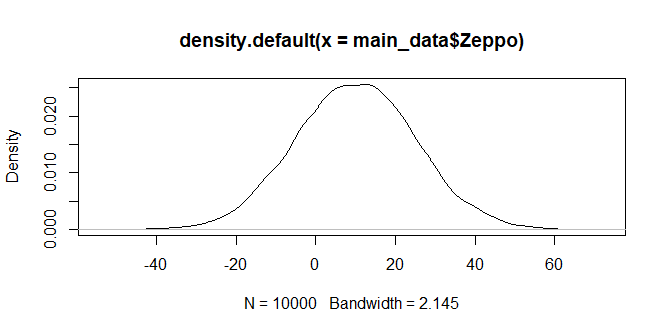
**#question 2** – The distribution seems to be in the wave form.

**#Column 3**

> hist(main\_data$Zeppo)

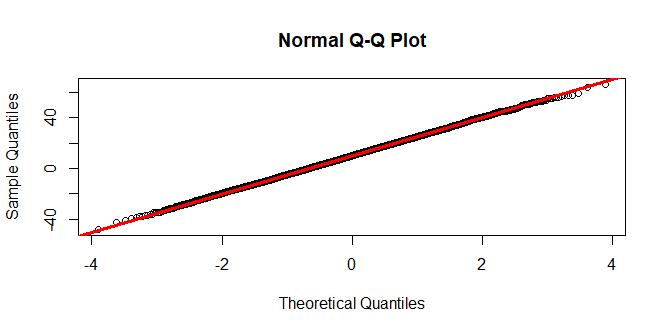


> plot(density(main\_data$Zeppo))



> qqnorm(main\_data$Zeppo)

> qqline(main\_data$Zeppo, col="red", lwd=3)

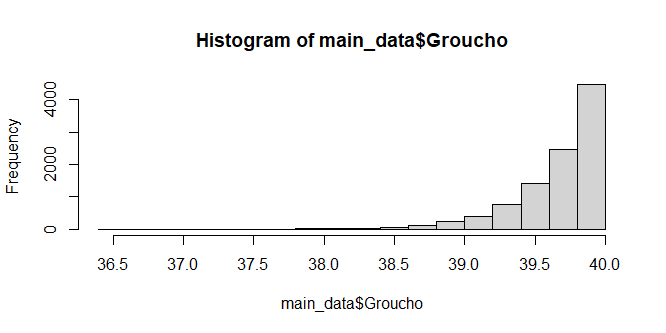


CONCLUSION – Histogram and density plot shows the data to be normally distributed. And QQ plot confirms that third variable “Zeppo” is normally distributed as all the data falls on the QQ line.

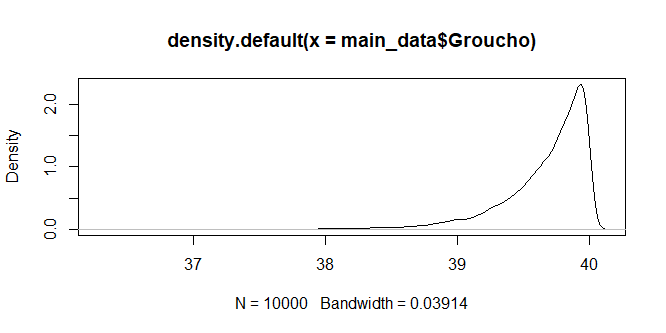
**#question 2** – Data is normally distributed

**#Column 4**

> hist(main\_data$Groucho)

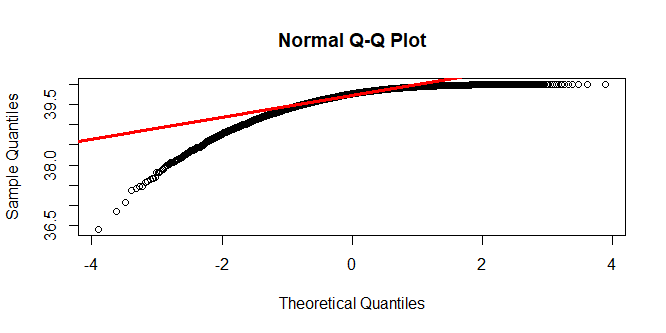


> plot(density(main\_data$Groucho))



> qqnorm(main\_data$Groucho)

> qqline(main\_data$Groucho, col="red", lwd=3)



CONCLUSION - Histogram and density plot shows the data to be left skewed. Additionally, a lot of data is out of bound from the QQ line as shown in the QQ plot. This proves that the fourth variable is not normally distributed.

#**question 2** – This distribution seems to be inverse Weibull distribution.

**#question 3** –

> df\_ques3 <- data.frame()

> for (i in 1:1000){

+ samp <- sample\_n(main\_data, 50)

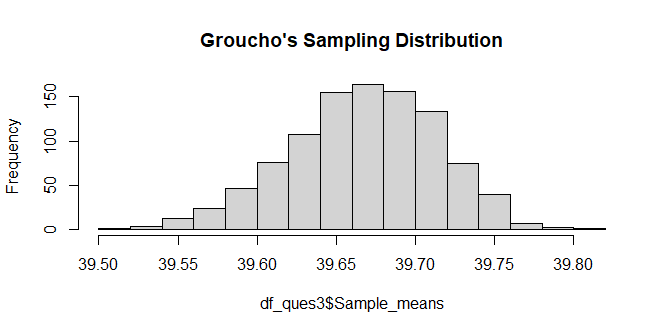
+ df\_ques3[i,1] <- i

+ df\_ques3[i,2] <- mean(samp$Groucho)

+ }

> colnames(df\_ques3) <- c('Groucho\_Samples', "Sample\_means")

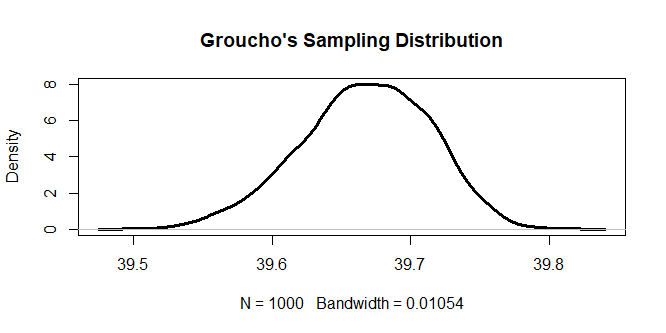
> hist(df\_ques3$Sample\_means, main="Groucho's Sampling Distribution")



> plot(density(df\_ques3$Sample\_means),

+ main="Groucho's Sampling Distribution",

+ lwd=3)



> skewness(df\_ques3$Sample\_means)

[1] -0.2642173

> kurtosis(df\_ques3$Sample\_means)

[1] 2.861898

CONCLUSION – Histogram and Density plots shows the sampling distribution of “Groucho” to be normally distributed. This is further proved by analytical tools Skewness (does not significant value to be considered) and Kurtosis (whose value is close to 3). Therefore, the Central Limit Theorem has been proved.