#amulyageereddy

Commands for 1st and 2nd question:

#Loading the data set.

install.packages(c("rmarkdown", "car", "Hmisc","rio", "moments", "corrplot", "MASS", "ppcor"), dep =TRUE)

install.packages("rio")

install.packages("moments")

library(rio)

library(moments)

amulyadata=import("6304 Module 2 Assignment Data.xlsx",sheet="Sheet 1")

attach(amulyadata)

colnames(amulyadata)=tolower(make.names(colnames(amulyadata)))

##1st question -column 1

set.seed(62067273)

mean(amulyadata$harpo)

sd(amulyadata$harpo)

hist(amulyadata$harpo,col="red",main="Histogram plot for harpo")

plot(density(amulyadata$harpo),lwd=3,main="Harpo Normal Random Deviates")

qqnorm(amulyadata$harpo,pch=16)

qqline(amulyadata$harpo,col="red",lwd=3)

skewness(amulyadata$harpo)

kurtosis(amulyadata$harpo)

#2nd column

mean(amulyadata$chico)

sd(amulyadata$chico)

hist(amulyadata$chico,col="red",main="Histogram plot for chico")

plot(density(amulyadata$chico),lwd=3,main="chico Normal Random Deviates")

qqnorm(amulyadata$chico,pch=16)

qqline(amulyadata$chico,col="red",lwd=3)

skewness(amulyadata$chico)

kurtosis(amulyadata$chico)

#3rd column

mean(amulyadata$zeppo)

sd(amulyadata$zeppo)

hist(amulyadata$zeppo,col="red",main="Histogram plot for zeppo")

plot(density(amulyadata$zepppo),lwd=3,main="Zeppo Normal Random Deviates")

qqnorm(amulyadata$zeppo,pch=19)

qqline(amulyadata$zeppo,col="red",lwd=3)

skewness(amulyadata$zeppo)

kurtosis(amulyadata$zeppo)

#4th column

mean(amulyadata$groucho)

sd(amulyadata$groucho)

hist(amulyadata$groucho,col="red",main="Histogram plot for groucho")

plot(density(amulyadata$groucho),lwd=3,main="Groucho Normal Random Deviates")

qqnorm(amulyadata$groucho,pch=19)

qqline(amulyadata$groucho,col="red",lwd=3)

skewness(amulyadata$groucho)

kurtosis(amulyadata$groucho)

Question 1:

After analyzing the four columns of the data only third column is normally distributed .i.e zeppo. As only for the zeppo column the qqline and qqplot are in interaction with each other and the kurtosis value is 2.99. And the histogram plots and Density plot also clearly indicates the normal curve. Hence only zippo column is normally distributed.

Question 2:

From the above data we can say that only the column 3 .i.e Zeppo follows normal distribution.

1st column:

> set.seed(62067273)

> mean(amulyadata$harpo)

[1] 29.92621

> sd(amulyadata$harpo)

[1] 5.767096

> hist(amulyadata$harpo,col="red",main="Histogram plot for harpo")

> plot(density(amulyadata$harpo),lwd=3,main="Harpo Normal Random Deviates")

> qqnorm(amulyadata$harpo,pch=16)

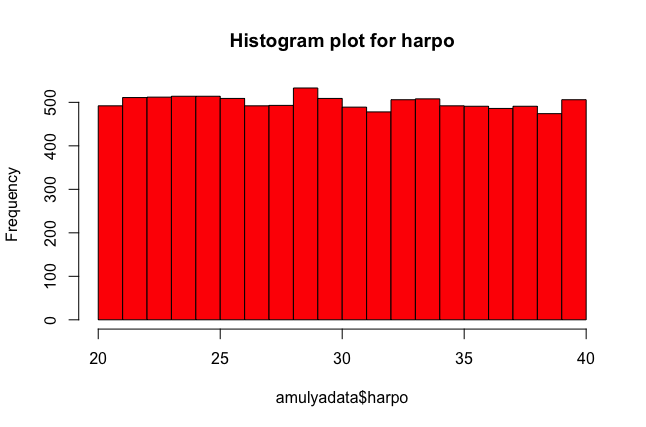
> qqline(amulyadata$harpo,col="red",lwd=3)

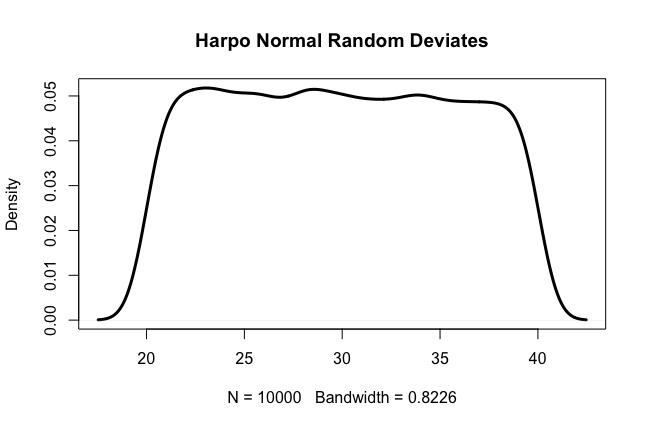
> skewness(amulyadata$harpo)

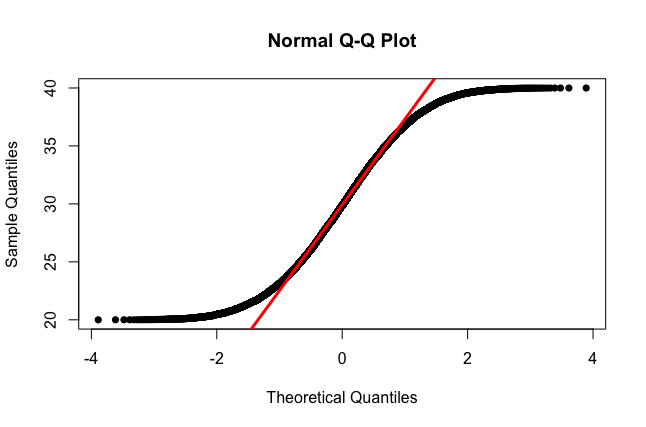
[1] 0.02170411

> kurtosis(amulyadata$harpo)

[1] 1.804474







From the above Q-Q plot of the harpo column of the we can clearly say that the above graph is not a normal distribution and by histogram plot and Density plot we can say that the data follows the **uniformly distributed**. As the kurtosis value also indicates that the data is not normally distributed 1.8.

1st question 2nd column:

> mean(amulyadata$chico)

[1] 72.9924

> sd(amulyadata$chico)

[1] 1.714245

> hist(amulyadata$chico,col="red",main="Histogram plot for chico")

> plot(density(amulyadata$chico),lwd=3,main="chico Normal Random Deviates")

> qqnorm(amulyadata$chico,pch=16)

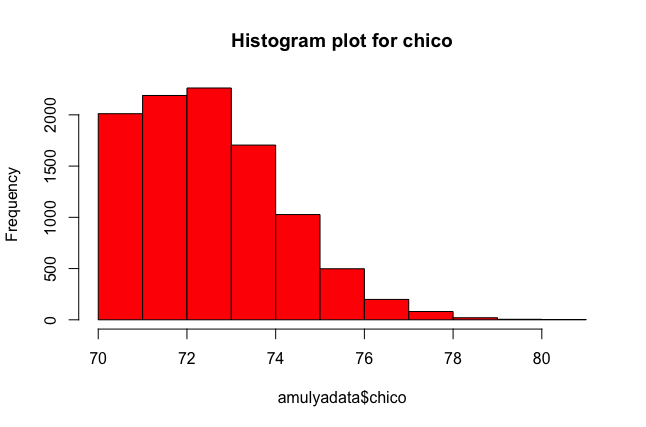
> qqline(amulyadata$chico,col="red",lwd=3)

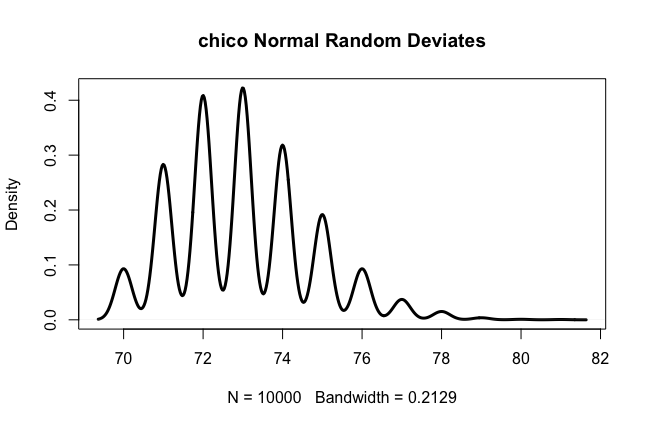
> skewness(amulyadata$chico)

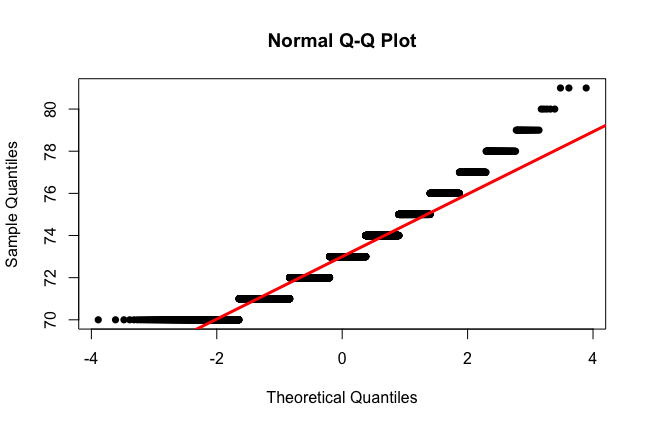
[1] 0.5306944

> kurtosis(amulyadata$chico)

[1] 3.217501







From the above data and the Q-Q plot the qqcurve and qqline of the column chico are not in interaction this indicates that it does not follow normal distribution. And the histogram plot and Denisty plot indicates that it follows **weibull distribution**.

1st Question –3rd coloumn

> mean(amulyadata$zeppo)

[1] 10.04221

> sd(amulyadata$zeppo)

[1] 15.03914

> hist(amulyadata$zeppo,col="red",main="Histogram plot for zeppo")

> plot(density(amulyadata$zeppo),lwd=3,main="Zeppo Normal Random Deviates")

> qqnorm(amulyadata$zeppo,pch=19)

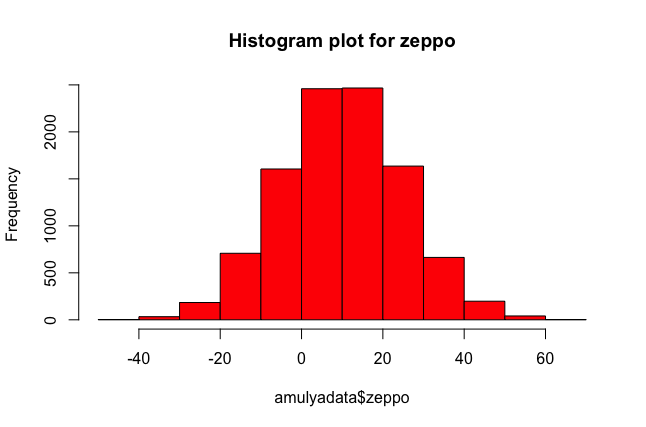
> qqline(amulyadata$zeppo,col="red",lwd=3)

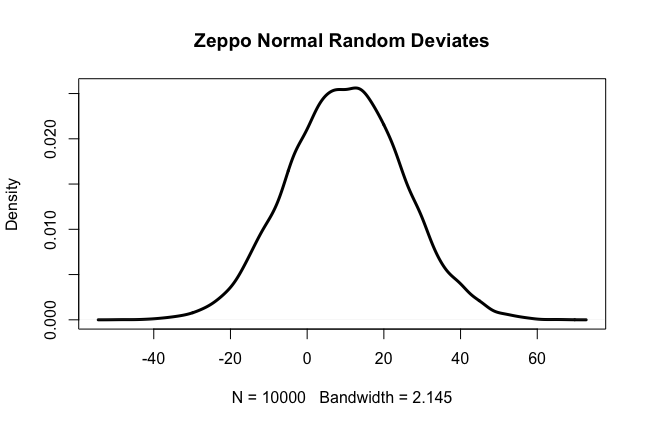
> skewness(amulyadata$zeppo)

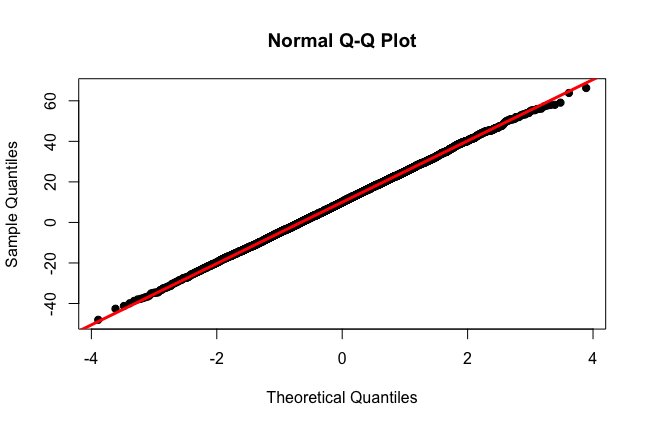
[1] 0.01012317

> kurtosis(amulyadata$zeppo)

[1] 2.954868







From the data we can say that only column 3 zeppo follows the normal distribution and the kurtosis value for zeppo is 2.99. This clearly indicates that the zeppo column follows the normal distribution. And the histogram plots and Density plot also clearly indicates the normal curve.In Q-Q plot the qqline and qqnorm are almost equal.

Question 1 column 4:

> mean(amulyadata$groucho)

[1] 39.66701

> sd(amulyadata$groucho)

[1] 0.3294301

> hist(amulyadata$groucho,col="red",main="Histogram plot for groucho")

> plot(density(amulyadata$groucho),lwd=3,main="Groucho Normal Random Deviates")

> qqnorm(amulyadata$groucho,pch=19)

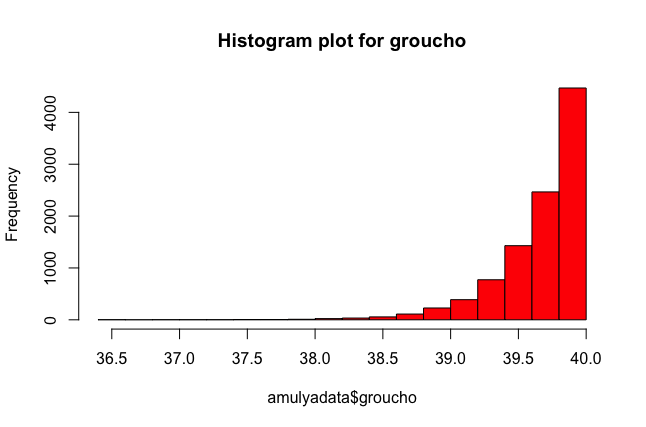
> qqline(amulyadata$groucho,col="red",lwd=3)

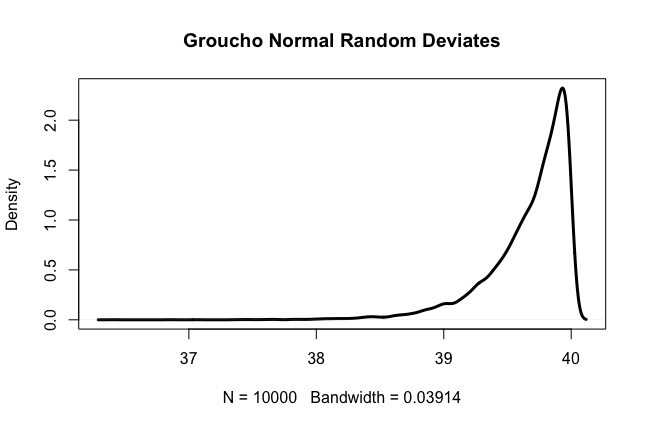
> skewness(amulyadata$groucho)

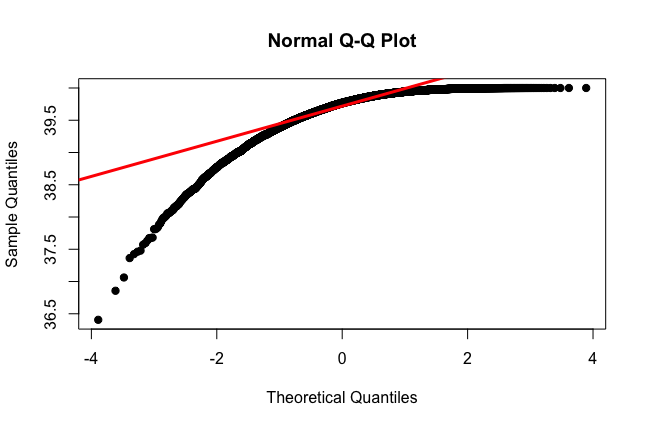
[1] -2.034274

> kurtosis(amulyadata$groucho)

[1] 9.600076







From the above data we can say that the distribution is not normal as there are several points which do not lie on qqline intersection is not perfect. And from the histogram graph and density plot clearly indicates that it follows **Exponential distribution**.

**Question 3 :**

> myydata = data.frame()

> for(i in 1:1000)

+ {

+ myydata[i,1] = mean(sample(amulyadata$groucho,50,replace=FALSE))

+ }

> colnames(myydata) = c("Data")

> hist(myydata$Data ,col="red",main="Histogram plot for my sample data")

> plot(density(myydata$Data),lwd=3,main="plot of sampling distribution")

> qqnorm(myydata$Data)

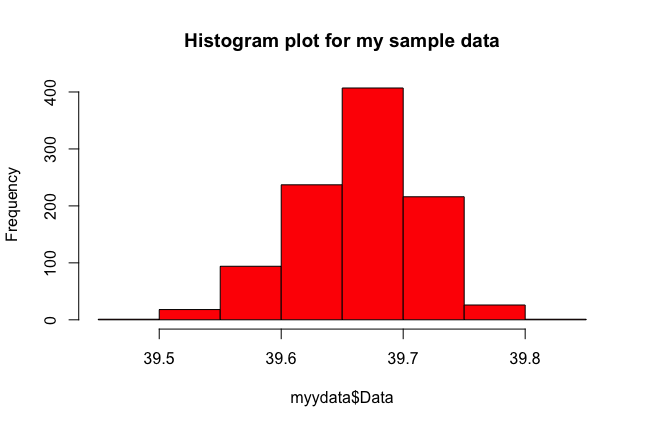
> qqline(myydata$Data,col="red",lwd=3)

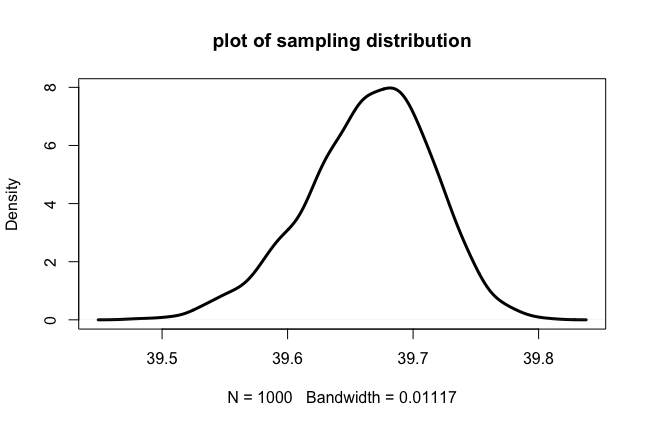
> skewness(myydata$Data)

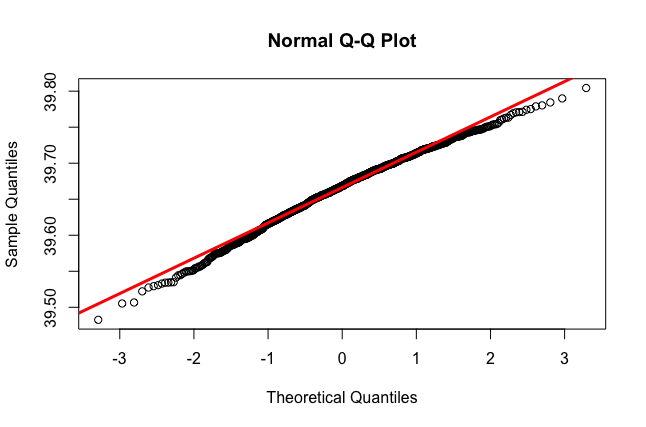
[1] -0.3788458

> kurtosis(myydata$Data)

[1] 3.058426







From the plots of sampling distribution for the Groucho column with the data set of 50 from 1000 sample data. We can clearly observe that the data follows normal distribution. In the Q-Q plot of the sampling distribution the qqline and qqcurve are aligned in the same path and the kurtosis value is 3.05. This clearly indicates that data follows normal distribution, Histogram and desity plot also clearly indicating the normal curve in the graphs.

From the Central limit theorem as the size of data increases the graph becomes normalized, As the sample data is greater that 30 it clearly indicates that the data is normally disturbed. Hence form the Q-Qplot , kurtosis value and the central limit theorem the sampling distribution follows normal distribution.