tarea

2024-08-14

Analiza 2 de las siguientes variables en cuanto a sus datos atípicos y normalidad

```
M <- read.csv("downloads/mc-donalds-menu.csv")

Calories = M$Calories
Sugars = M$Sugars
```

Graficar el diagrama de caja y bigote

Calcula el rango intercuartílico y los cuartíles

Identifica la cota de 1.5 rangos intercuartílicos para datos atípicos

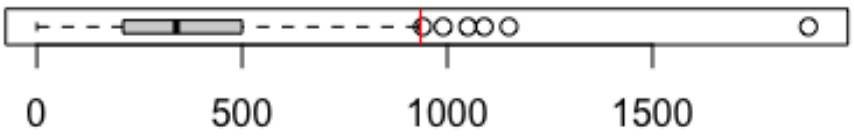
En estos 2 casos si es recomendable quitar los datos atipicos ya que en las calorias si se nota

la diferencia, en los azucares no impacta tanto pero es recomendable. Tambien podemos ver estas diferencias en las comparaciones de las variables con summary y vemos como si afecta en el Max.

```
q1_cal = quantile(Calories, 0.25)
q3_cal = quantile(Calories, 0.75)
ri_cal = IQR(Calories)

par(mfrow=c(2,1))
boxplot(Calories, horizontal=TRUE, ylim=c(0, 1900))
abline(v=q3_cal + 1.5*ri_cal, col="red")

Calories1 <- M[M$Calories < q3_cal + 1.5*ri_cal, "Calories"]

summary(Calories)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0   210.0   340.0   368.3   500.0  1880.0
```

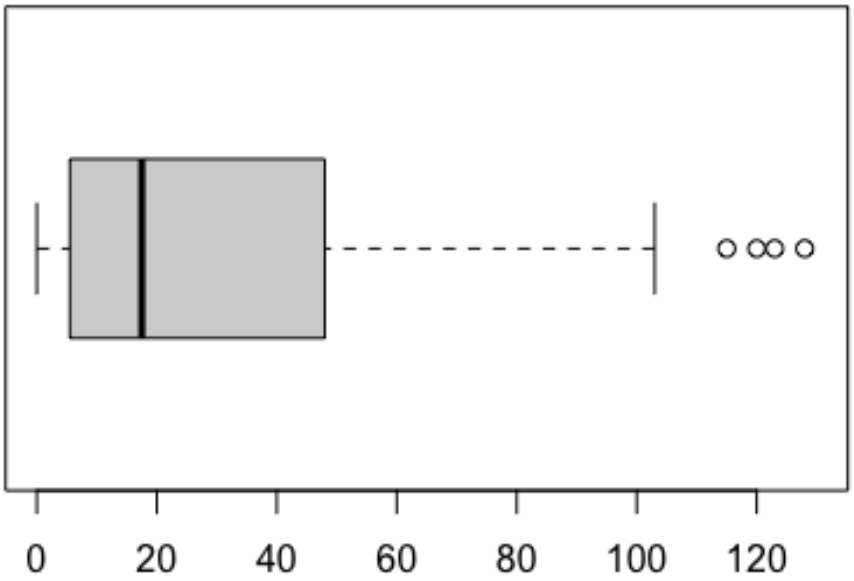
```r
summary(Calories1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0   202.5   335.0   349.0   480.0   930.0
```



Mismo proceso para los Azucares

```r
q1_sug = quantile(Sugars, 0.25)
q3_sug = quantile(Sugars, 0.75)
ri_sug = IQR(Sugars)

boxplot(Sugars, horizontal=TRUE, ylim=c(0, 130))
```

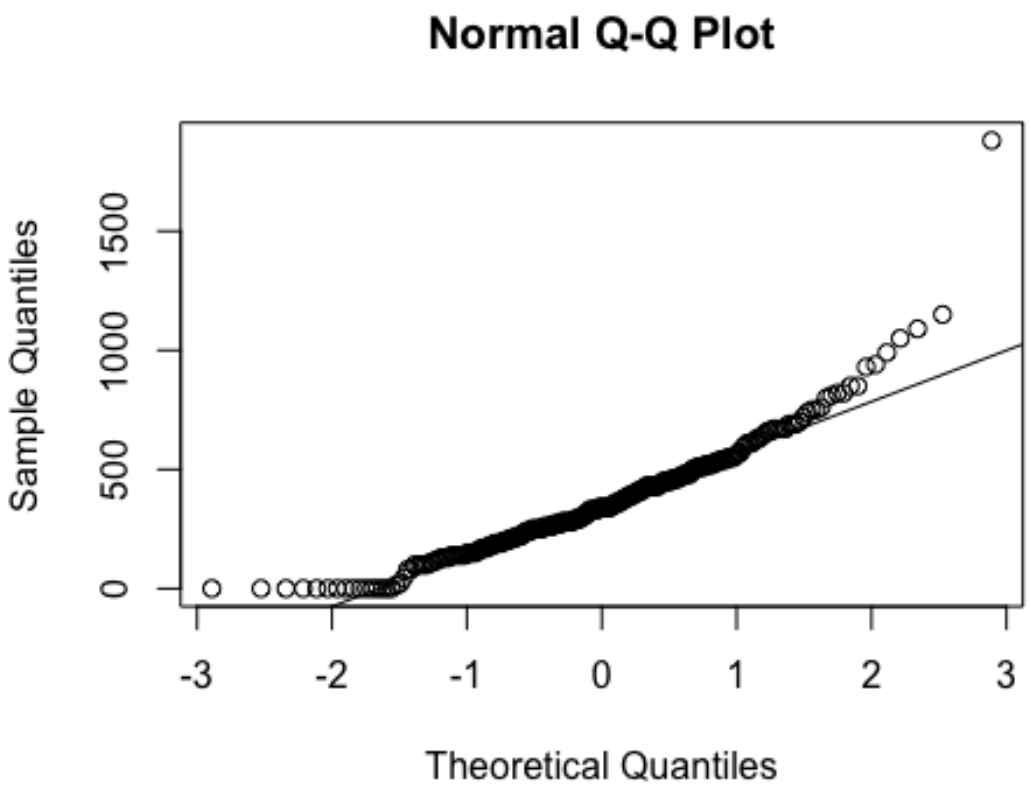



```
Sugars1 = M[M$Sugars < q3_sug + 1.5*ri_sug, "Sugars"]

summary(Sugars)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    5.75   17.50   29.42   48.00  128.00

summary(Sugars1)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    5.00   17.00   27.98   46.25  103.00
```

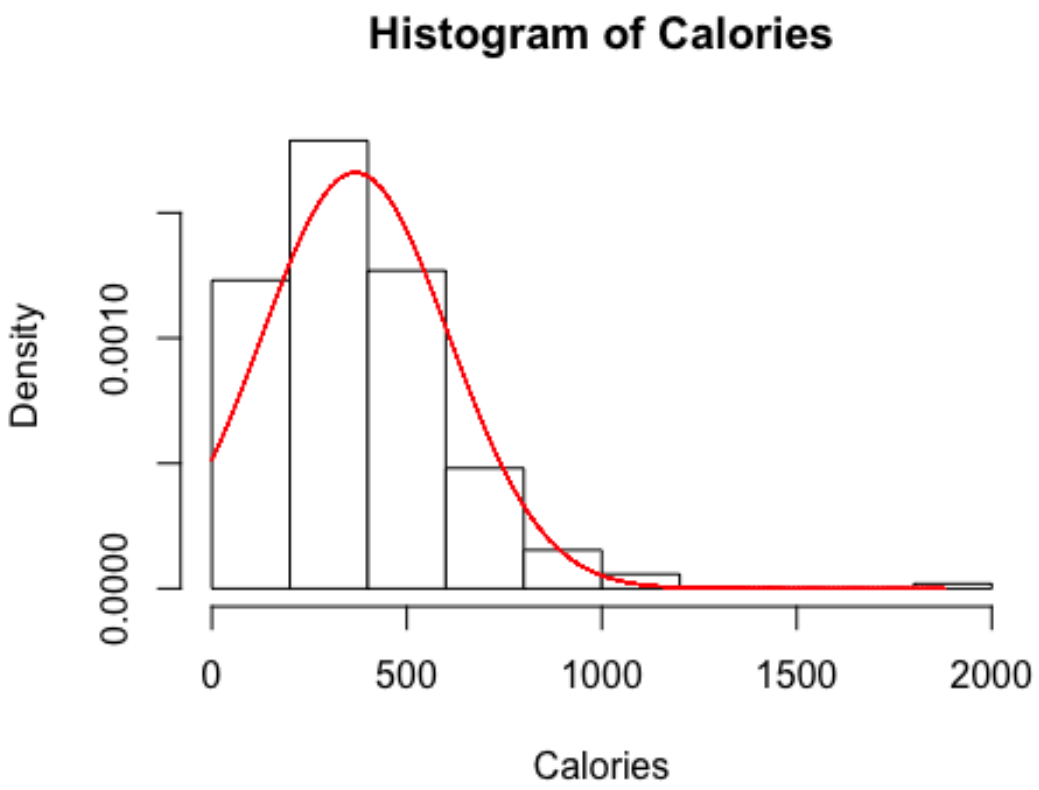
Grafica los datos y su respectivo QQPlot: qqnorm(datos) y qqline(datos) para cada variable

Realiza el histograma y su distribución teórica de probabilidad

```
qqnorm(Calories)
qqline(Calories)
```

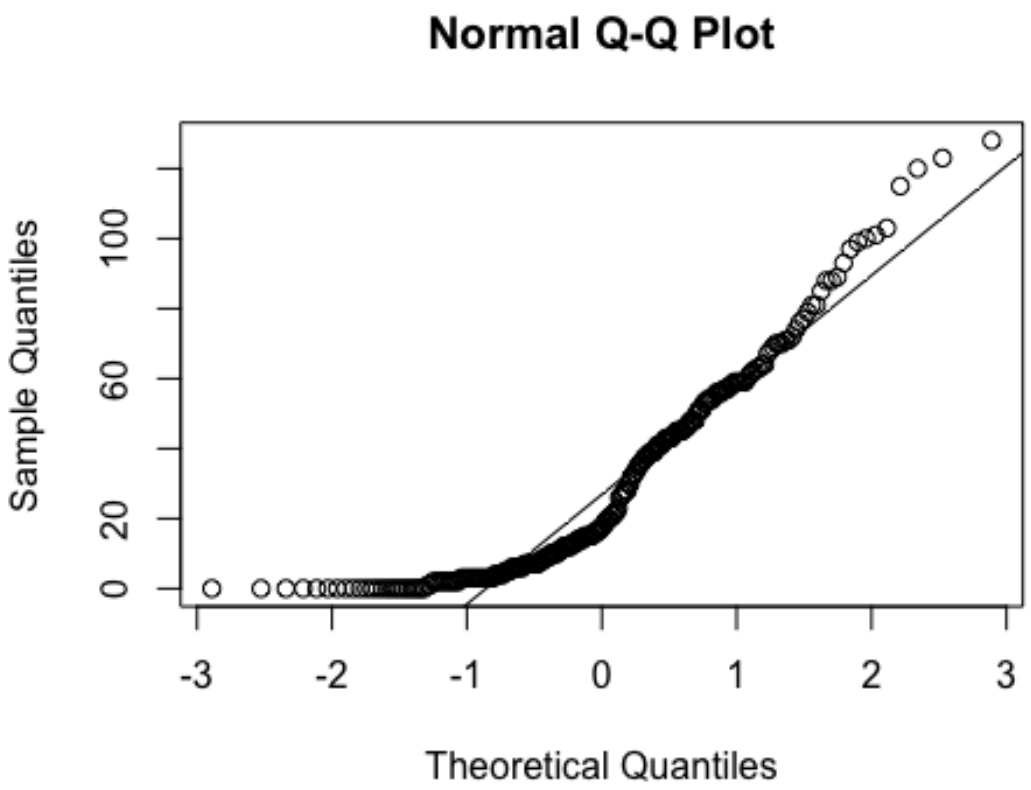



```r
hist(Calories, freq=FALSE, col=0)
x_cal = seq(min(Calories), max(Calories), 0.1)
y_cal = dnorm(x_cal, mean=mean(Calories), sd=sd(Calories))
lines(x_cal, y_cal, col="red")
```

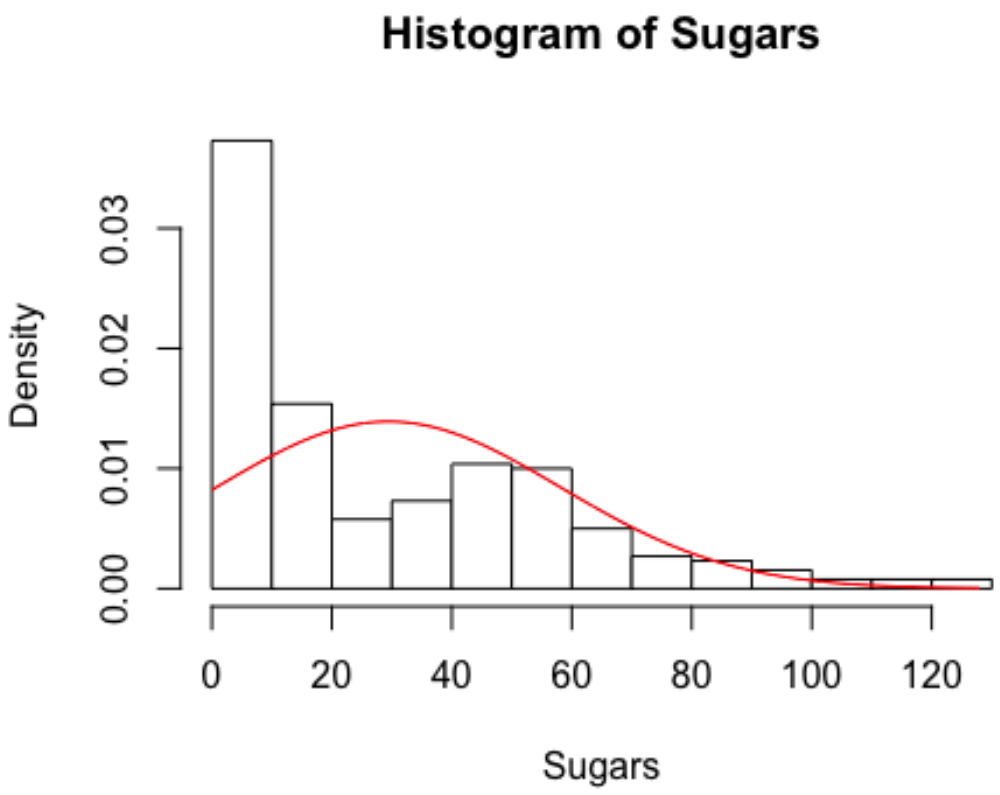


Mismo proceso para Azucares

```
qqnorm(Sugars)
qqline(Sugars)
```




```r
hist(Sugars, freq=FALSE, col=0)
x_sug = seq(min(Sugars), max(Sugars), 0.1)
y_sug = dnorm(x_sug, mean=mean(Sugars), sd=sd(Sugars))
lines(x_sug, y_sug, col="red")
```



#

Calcula el coeficiente de sesgo y el coeficiente de curtosis de cada variable.

```
library(moments)

cal_skewness = skewness(Calories)
cal_kurtosis = kurtosis(Calories)

sug_skewness = skewness(Sugars)
sug_kurtosis = kurtosis(Sugars)

cat("Sesgo de Calories:", cal_skewness, "\n")

## Sesgo de Calories: 1.444105

cat("Curtosis de Calories:", cal_kurtosis, "\n")

## Curtosis de Calories: 8.645274

cat("Sesgo de Sugars:", sug_skewness, "\n")

## Sesgo de Sugars: 1.025977

cat("Curtosis de Sugars:", sug_kurtosis, "\n")

## Curtosis de Sugars: 3.487744
```

```r
library(moments)
ad.test(Calories)
ad.test(Sugars)
```

```
        Anderson-Darling normality test

data:  Calories
A = 2.5088, p-value = 2.369e-06


        Anderson-Darling normality test

data:  Sugars
A = 9.9899, p-value < 2.2e-16
```