

4. Explorando bases

Eliezer Cavazos

2024-08-13

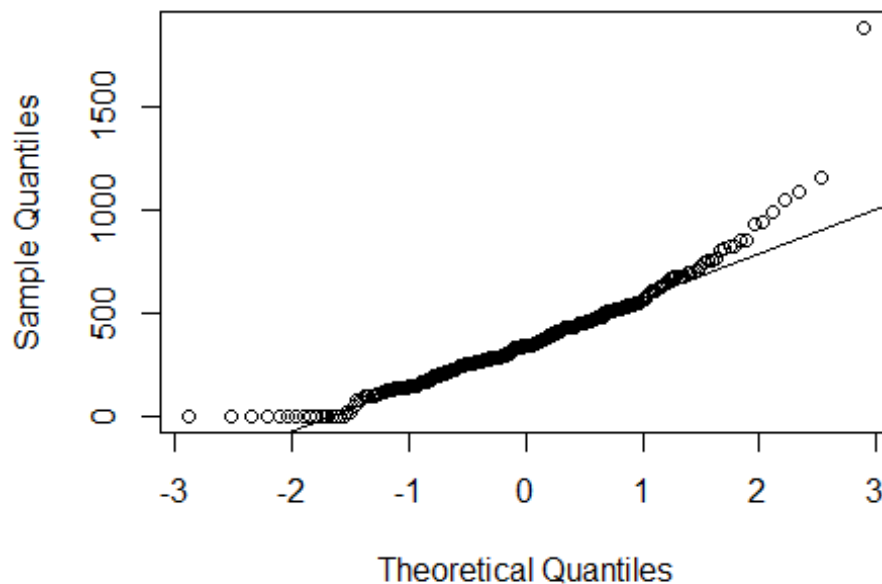
Lectura Archivo

```
oMcDonalds =  
read.csv("C:\\Users\\eliez\\OneDrive\\Desktop\\Clases\\mc-donalds-menu.csv")  
#Leer la base de datos
```

2. Analisis datos atipicos y normalidad

```
#library(nortest)  
shapiro.test(oMcDonalds$Calories)  
  
##  
##  Shapiro-Wilk normality test  
##  
## data:  oMcDonalds$Calories  
## W = 0.91902, p-value = 1.119e-10  
  
qqnorm(oMcDonalds$Calories)  
qqline(oMcDonalds$Calories)
```

Normal Q-Q Plot

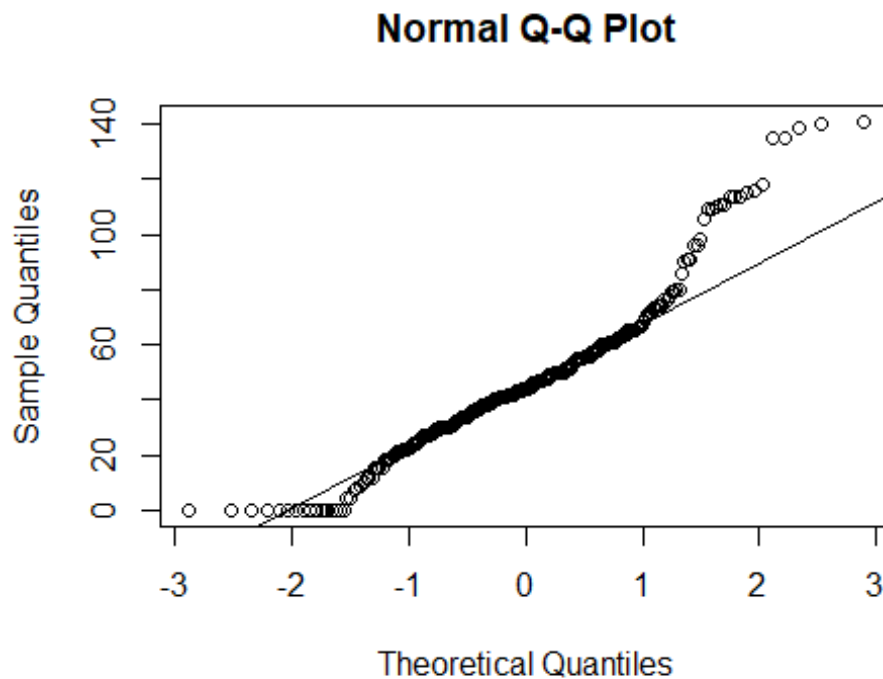


```
shapiro.test(oMcDonalds$Carbohydrates)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: oMcDonalds$Carbohydrates  
## W = 0.93666, p-value = 3.931e-09
```

```
qqnorm(oMcDonalds$Carbohydrates)
```

```
qqline(oMcDonalds$Carbohydrates)
```



3. Grafico de Densidad

```
oCalories = oMcDonalds$Calories
```

```
y1 = min(oCalories)
```

```
y2 = max(oCalories)
```

```
q1=quantile(oCalories,0.25) #Cuantil 1 de la variable oCalories
```

```
q3 = quantile(oCalories, 0.75)
```

```
#ri= q3-q1 #o
```

```
ri=IQR(oCalories) #Rango intercuantílico de oCalories
```

```
rango = q3+1.5*ri
```

```
par(mfrow=c(2,1)) #Matriz de gráficos de 2x1
```

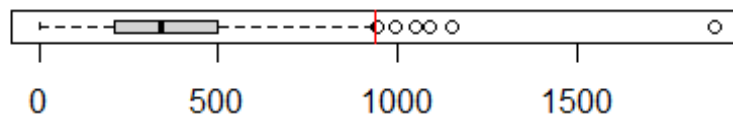
```
boxplot(oCalories, horizontal=TRUE)
abline(v=rango, col="red") #línea vertical en el límite de los datos atípicos o extremos
```

```
oCaloriesClean= oCalories[oCalories<rango] #En la matriz M, quitar datos más allá de 1.5 rangos intercuantílicos arriba de q3 de la variable oCalories
summary(oCaloriesClean)
```

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

```
summary(oCalories)
```

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



```
oCarboHydrates = oMcDonalds$Carbohydrates
```

```
y1 = min(oCarboHydrates)
```

```
y2 = max(oCarboHydrates)
```

```
q1=quantile(oCarboHydrates,0.25) #Cuantil 1 de la variable oCarboHydrates
```

```
q3 = quantile(oCarboHydrates, 0.75)
```

```
#ri= q3-q1 #o
```

```
ri=IQR(oCarboHydrates) #Rango intercuantílico de oCarboHydrates
```

```
rango = q3+1.5*ri
```

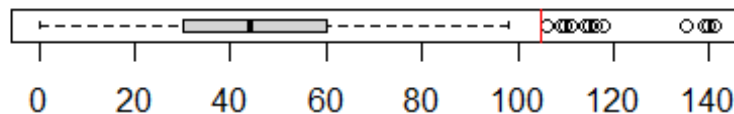
```
par(mfrow=c(2,1)) #Matriz de gráficos de 2x1
boxplot(oCarboHydrates,horizontal=TRUE)
abline(v=rango,col="red") #línea vertical en el límite de los datos atípicos
o extremos
```

```
oCarboHydratesClean= oCarboHydrates[oCarboHydrates<rango] #En la matriz M,
quitar datos más allá de 1.5 rangos intercuantílicos arriba de q3 de la
variable oCarboHydrates
summary(oCarboHydratesClean)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   30.00   43.00   42.28   56.00   98.00
```

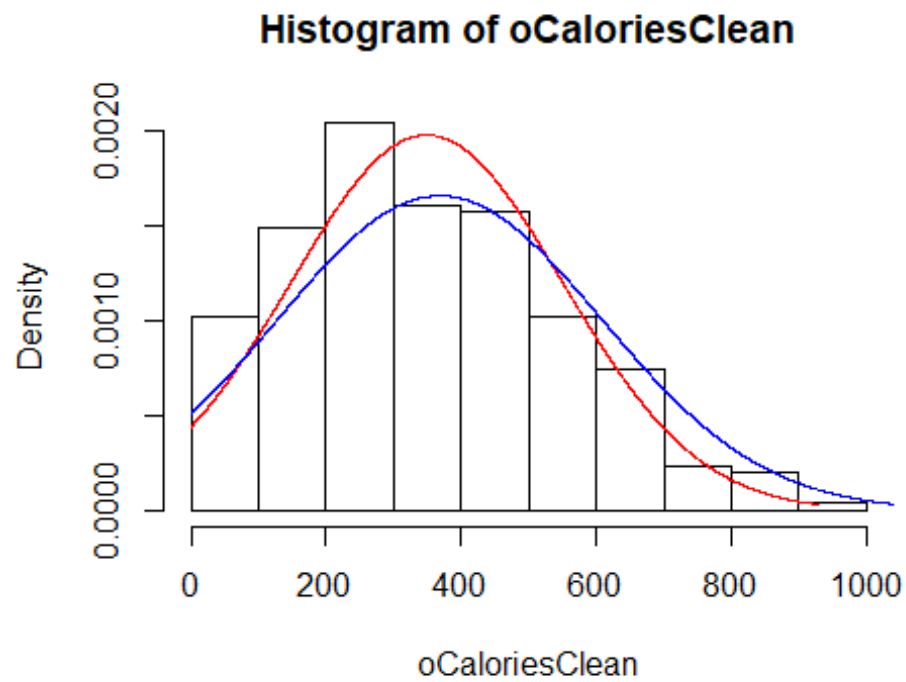
```
summary(oCarboHydrates)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   30.00   44.00   47.35   60.00  141.00
```



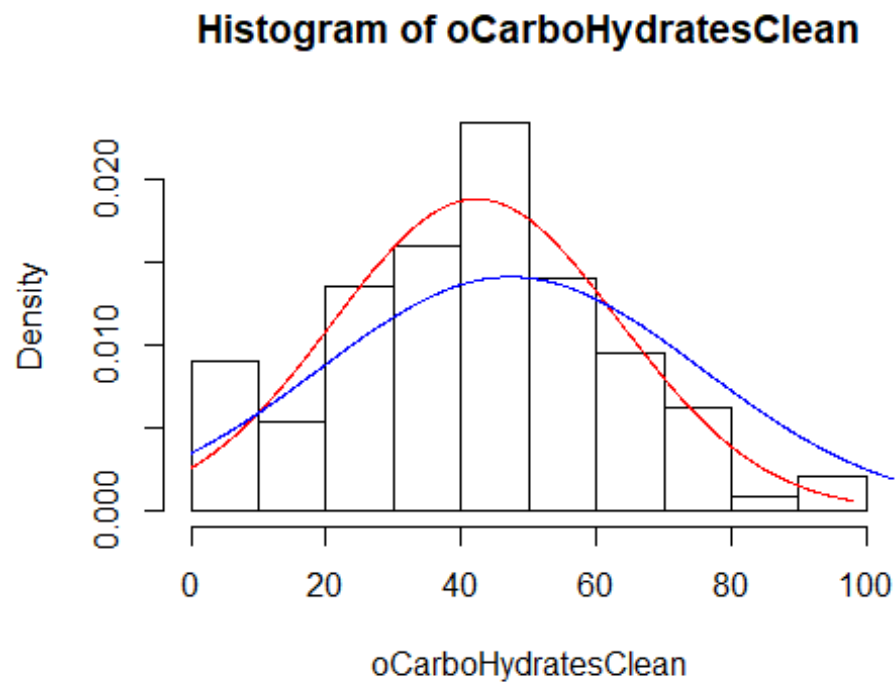
```
hist(oCaloriesClean,prob=TRUE,col=0)
x=seq(min(oCaloriesClean),max(oCaloriesClean),0.1)
y=dnorm(x,mean(oCaloriesClean),sd(oCaloriesClean))
lines(x,y,col="red")

x=seq(min(oCalories),max(oCalories),0.1)
y=dnorm(x,mean(oCalories),sd(oCalories))
lines(x,y,col="Blue")
```



```
hist(oCarboHydratesClean,prob=TRUE,col=0)
x=seq(min(oCarboHydratesClean),max(oCarboHydratesClean),0.1)
y=dnorm(x,mean(oCarboHydratesClean),sd(oCarboHydratesClean))
lines(x,y,col="red")

x=seq(min(oCarboHydrates),max(oCarboHydrates),0.1)
y=dnorm(x,mean(oCarboHydrates),sd(oCarboHydrates))
lines(x,y,col="Blue")
```



4. Analizar

Normalidad

```
library(nortest)

print("Calories")

## [1] "Calories"

ad.test(oCalories)

##
## Anderson-Darling normality test
##
## data: oCalories
## A = 2.5088, p-value = 2.369e-06

ad.test(oCaloriesClean)

##
## Anderson-Darling normality test
##
## data: oCaloriesClean
## A = 0.89786, p-value = 0.02166

print("CarboHydrates")

## [1] "CarboHydrates"

ad.test(oCarboHydrates)
```

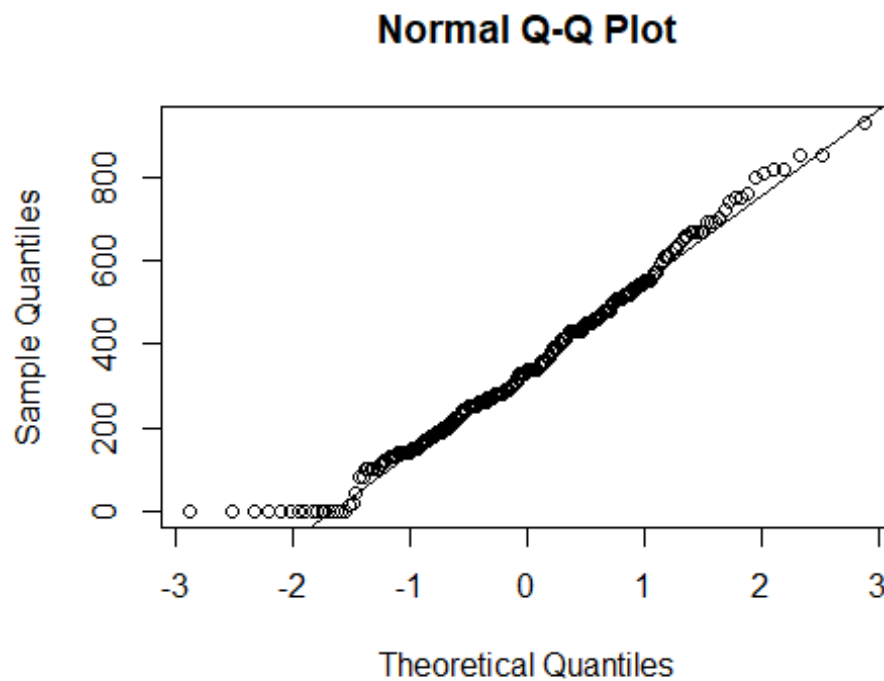
```
##  
## Anderson-Darling normality test  
##  
## data: oCarboHydrates  
## A = 4.1402, p-value = 2.547e-10
```

```
ad.test(oCarboHydratesClean)
```

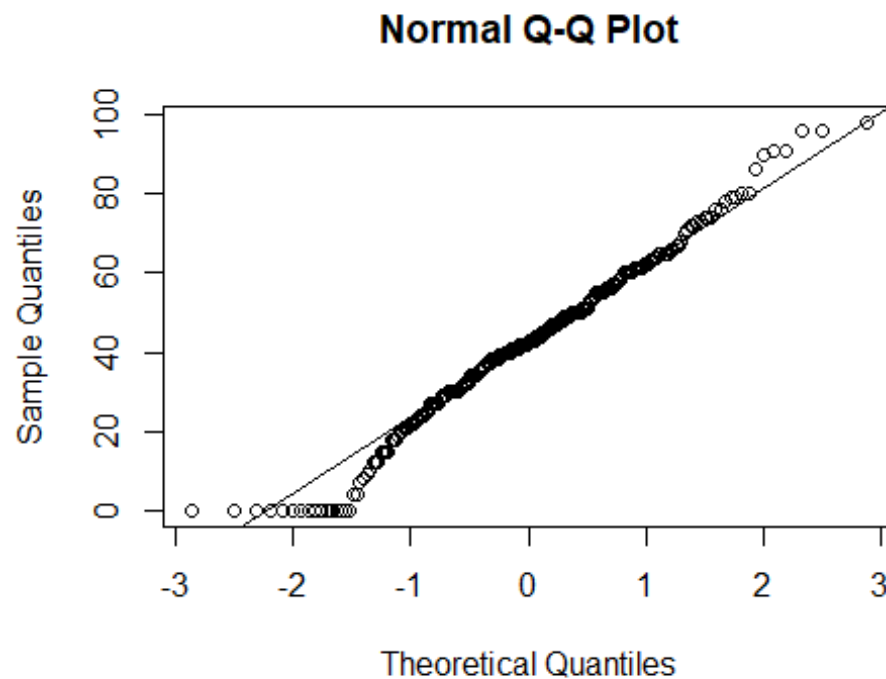
```
##  
## Anderson-Darling normality test  
##  
## data: oCarboHydratesClean  
## A = 0.74917, p-value = 0.05048
```

4.2 Grafica QQPLOT

```
qqnorm(oCaloriesClean)  
qqline(oCaloriesClean)
```



```
qqnorm(oCarboHydratesClean)  
qqline(oCarboHydratesClean)
```



4.3 Analisis Sesgo y Curtosis

```
library(e1071)
print("Calories")

## [1] "Calories"

skewness(oCaloriesClean)

## [1] 0.3469956

kurtosis(oCaloriesClean)

## [1] -0.3045226

print("Carbohydrates")

## [1] "Carbohydrates"

skewness(oCarboHydratesClean)

## [1] -0.02844112

kurtosis(oCarboHydratesClean)

## [1] -0.09271988
```

4.4 Comparar media, mediana y rango medio

```
print("Calories")
```



```
## [1] "Calories"

summary(oCaloriesClean)

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

summary(oCalories)

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

print("Carbohydrates")

## [1] "Carbohydrates"

summary(oCarboHydratesClean)

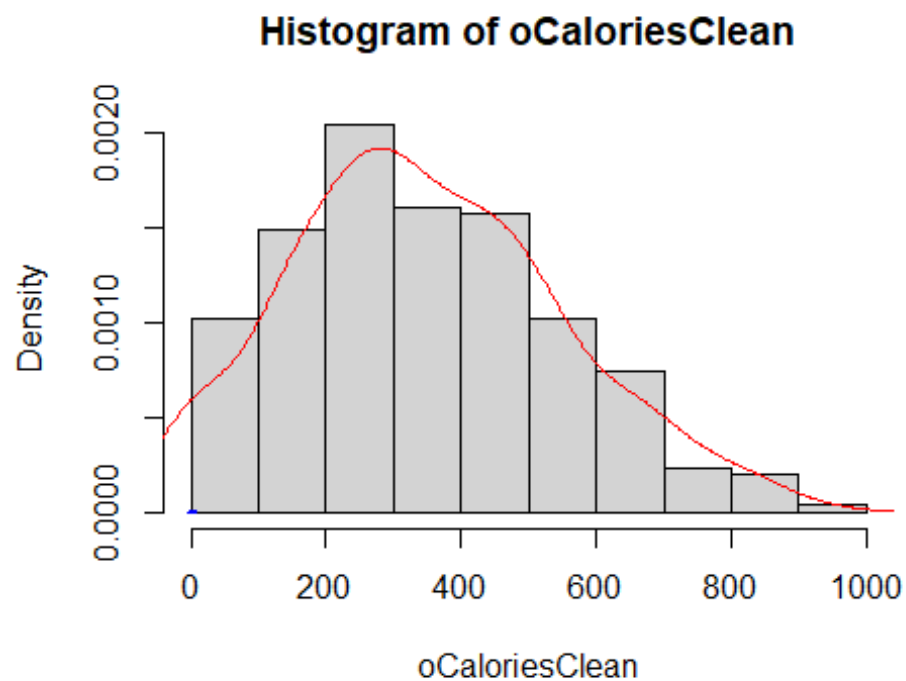
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   30.00   43.00   42.28   56.00   98.00

summary(oCarboHydrates)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   30.00   44.00   47.35   60.00  141.00
```

4.5 Histograma

```
hist(oCaloriesClean,freq=FALSE)
lines(density(oCaloriesClean),col="red")
curve(dnorm(x,mean=mean(oCaloriesClean,sd=sd(oCaloriesClean))), from=-6,
to=6, add=TRUE, col="blue",lwd=2)
```



```
hist(oCarboHydratesClean,freq=FALSE)
lines(density(oCarboHydratesClean),col="red")
curve(dnorm(x,mean=mean(oCarboHydratesClean,sd=sd(oCarboHydratesClean)),
from=-6, to=6, add=TRUE, col="blue",lwd=2)
```

Histogram of oCarboHydratesClean

