

# Explorando Bases

Cleber Perez

2024-08-13

## Leer archivo

```
M=read.csv("mc-donalds-menu.csv") #Leer la base de datos
M$variable # para llamar una variable, aunque también la puedes leer con
corchetes cuadrados M[renglón, columna]

## NULL
```

## Datos atípicos y Normalidad (Calorías/Carbohidratos)

### Pruebas de Normalidad univariada de las variables

```
library(nortest)
KS_calories = lillie.test(M$Calories)
KS_calories

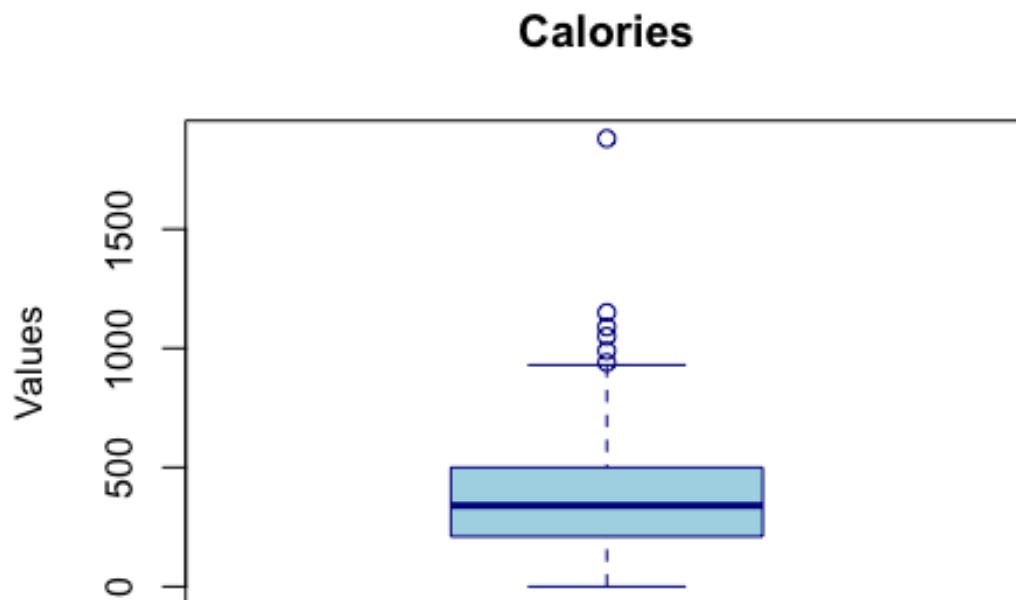
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  M$Calories
## D = 0.073753, p-value = 0.001611

KS_carbs = lillie.test(M$Carbohydrates)
KS_carbs

##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  M$Carbohydrates
## D = 0.098548, p-value = 2.081e-06
```

### Graficas de Caja y bigote de Calorias

```
boxplot(M$Calories,
        main = "Calories",
        ylab = "Values",
        col = "lightblue",
        border = "darkblue")
```



### Rango intercuartílico y los cuartiles de 1.5 rangos de Calorias

```
q1=quantile(M$Calories) #Cuantil 1 de la variable X
q1

##      0%   25%   50%   75%  100%
##       0   210   340   500  1880

ri=IQR(M$Calories)      #Rango intercuartílico de X
ri

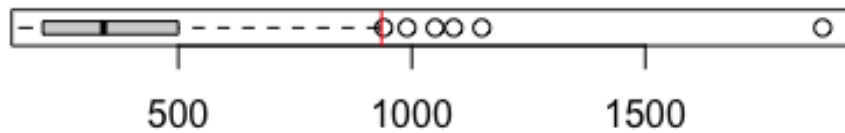
## [1] 290

par(mfrow=c(2,1)) #Matriz de gráficos de 2x1
boxplot(M$Calories,horizontal=TRUE,ylim=c(210,1880))
abline(v=500+1.5*ri,col="red") #línea vertical en el límite de los datos
atípicos o extremos
X1= M[M$Calories<500+1.5*ri,c("X")] #En la matriz M, quitar datos más allá
de 3 rangos intercuartílicos arriba de q3 de la variable X
summary(X1)

## Length Class Mode
##      0  NULL  NULL

summary(M$Calories)
```

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



### Rango intercuartilico y los cuartiles de 3 rangos de Calorias

```
q1=quantile(M$Calories) #Cuantil 1 de la variable X
q1
```

##	0%	25%	50%	75%	100%
##	0	210	340	500	1880

```
ri=IQR(M$Calories) #Rango intercuartilico de X
ri
```

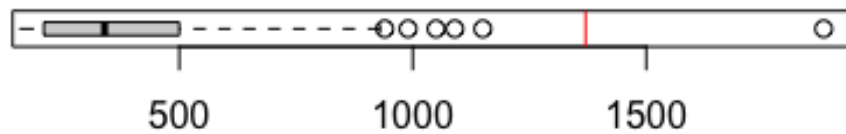
```
## [1] 290
```

```
par(mfrow=c(2,1)) #Matriz de gráficos de 2x1
boxplot(M$Calories,horizontal=TRUE,ylim=c(210,1880))
abline(v=500+3*ri,col="red") #linea vertical en el límite de los datos
atípicos o extremos
X1= M[M$Calories<500+3*ri,c("X")] #En la matriz M, quitar datos más allá de
3 rangos intercuartílicos arriba de q3 de la variable X
summary(X1)
```

```
## Length Class Mode
##      0  NULL  NULL
```

```
summary(M$Calories)
```

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



### Pruebas de normalidad en Calorias

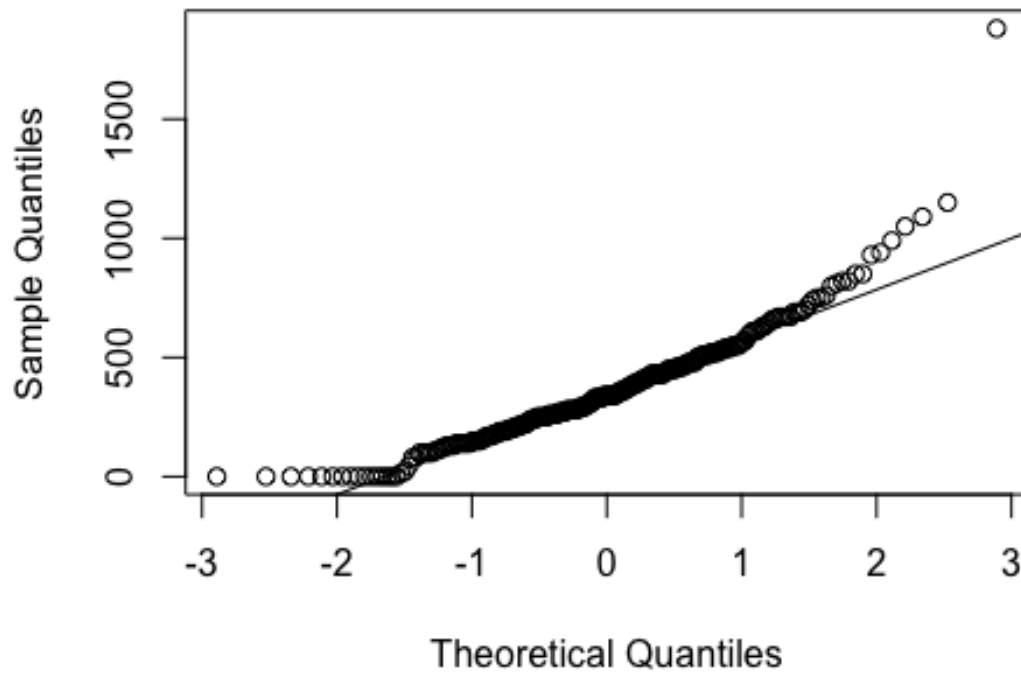
```
library(nortest)
n = lillie.test(M$Calories)
n

##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  M$Calories
## D = 0.073753, p-value = 0.001611
```

### Graficas de densidad de pobabilidad de Calorias

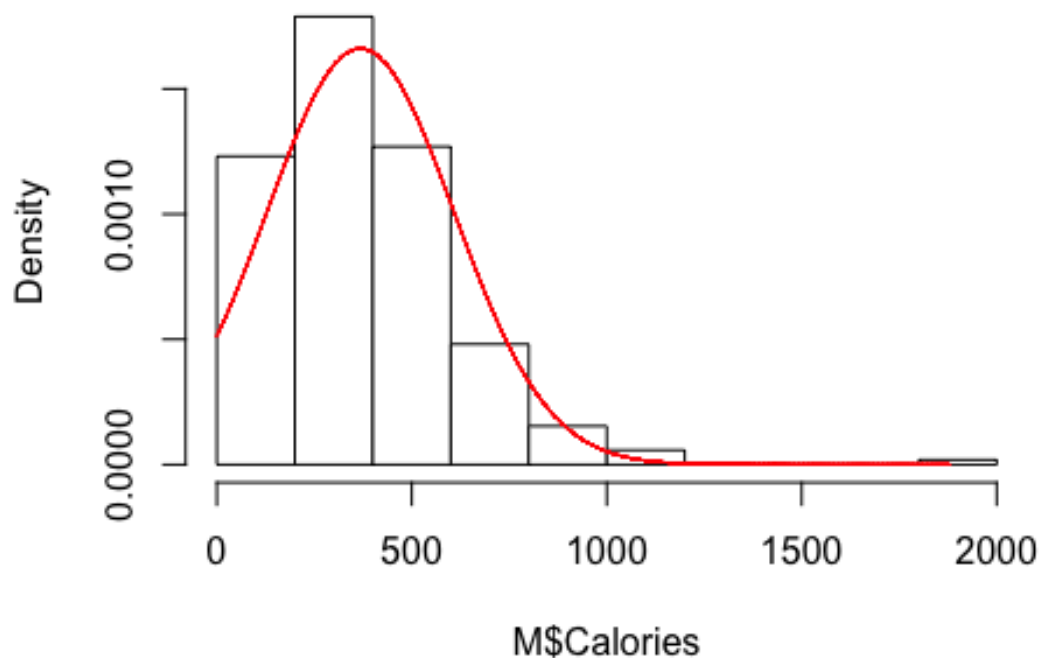
```
qqnorm(M$Calories)
qqline(M$Calories)
```

## Normal Q-Q Plot



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

## Histogram of M\$Calories



### Coeficiente de sesgo y curtosis de Calorias

```
mean_calories = mean(M$Calories, na.rm = TRUE)
mean_calories

## [1] 368.2692

median_calories = median(M$Calories, na.rm = TRUE)
median_calories

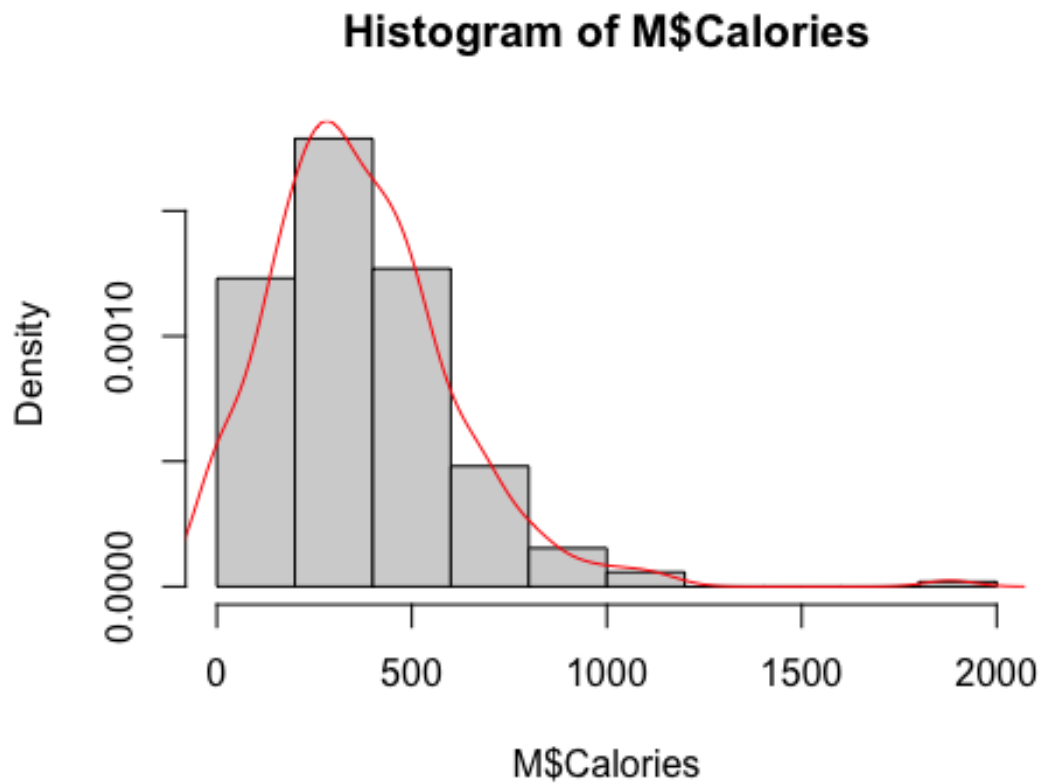
## [1] 340

range_calories = range(M$Calories, na.rm = TRUE)
rango_m_calories = diff(range_calories)
rango_m_calories

## [1] 1880
```

### Histograma de Calorias

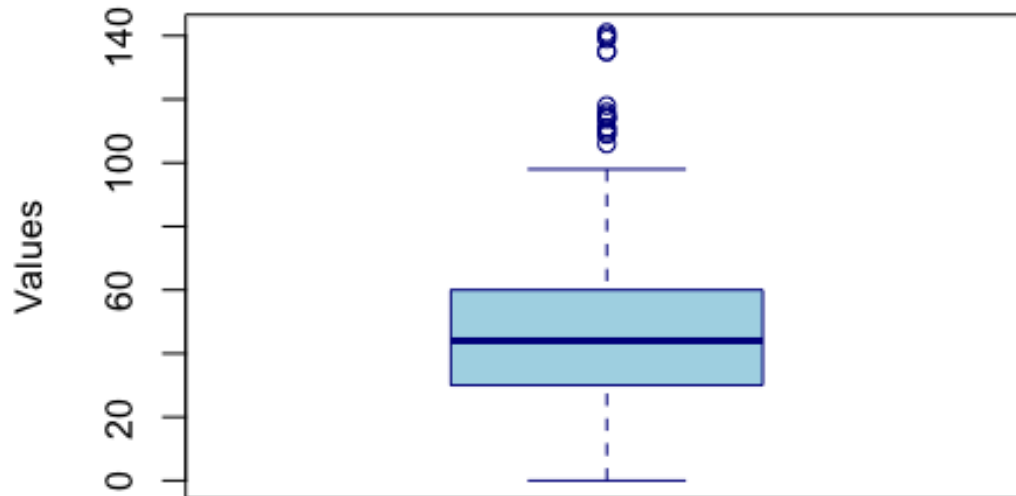
```
hist(M$Calories, freq=FALSE)
lines(density(M$Calories), col="red")
```



#### Graficas de Caja y bigote de Carbohidratos

```
boxplot(M$Carbohydrates,  
        main = "Carbohydrates",  
        ylab = "Values",  
        col = "lightblue",  
        border = "darkblue")
```

## Carbohydrates



### Rango intercuartilico y los cuartiles de 1.5 rangos de Carbohydrates

```
q1=quantile(M$Carbohydrates) #Cuantil 1 de la variable X
q1

##      0%    25%    50%    75%   100%
##       0     30     44     60    141

ri=IQR(M$Carbohydrates) #Rango intercuartilico de X
ri

## [1] 30

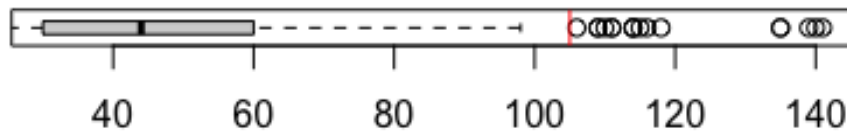
par(mfrow=c(2,1)) #Matriz de gráficos de 2x1
boxplot(M$Carbohydrates, horizontal=TRUE, ylim=c(30,141))
abline(v=60+1.5*ri, col="red") #linea vertical en el límite de los datos
atípicos o extremos
X1= M[M$Carbohydrates<60+1.5*ri, c("X")] #En la matriz M, quitar datos más
allá de 3 rangos intercuartílicos arriba de q3 de la variable X
summary(X1)

## Length Class Mode
##      0  NULL  NULL

summary(M$Carbohydrates)
```



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



### Rango intercuartilico y los cuartiles de 3 rangos de Carbohydrates

```
q1=quantile(M$Carbohydrates) #Cuantil 1 de la variable X
q1

##      0%   25%   50%   75%  100%
##       0    30    44    60   141

ri=IQR(M$Carbohydrates)      #Rango intercuartilico de X
ri

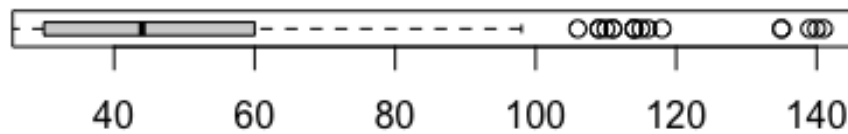
## [1] 30

par(mfrow=c(2,1)) #Matriz de gráficos de 2x1
boxplot(M$Carbohydrates,horizontal=TRUE,ylim=c(30,141))
abline(v=60+3*ri,col="red") #linea vertical en el límite de los datos
atípicos o extremos
X1= M[M$Carbohydrates<60+3*ri,c("X")] #En la matriz M, quitar datos más allá
de 3 rangos intercuartilicos arriba de q3 de la variable X
summary(X1)
```

```
## Length Class Mode
##      0  NULL  NULL
```

```
summary(M$Carbohydrates)
```

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



### Pruebas de normalidad en Carbohydrates

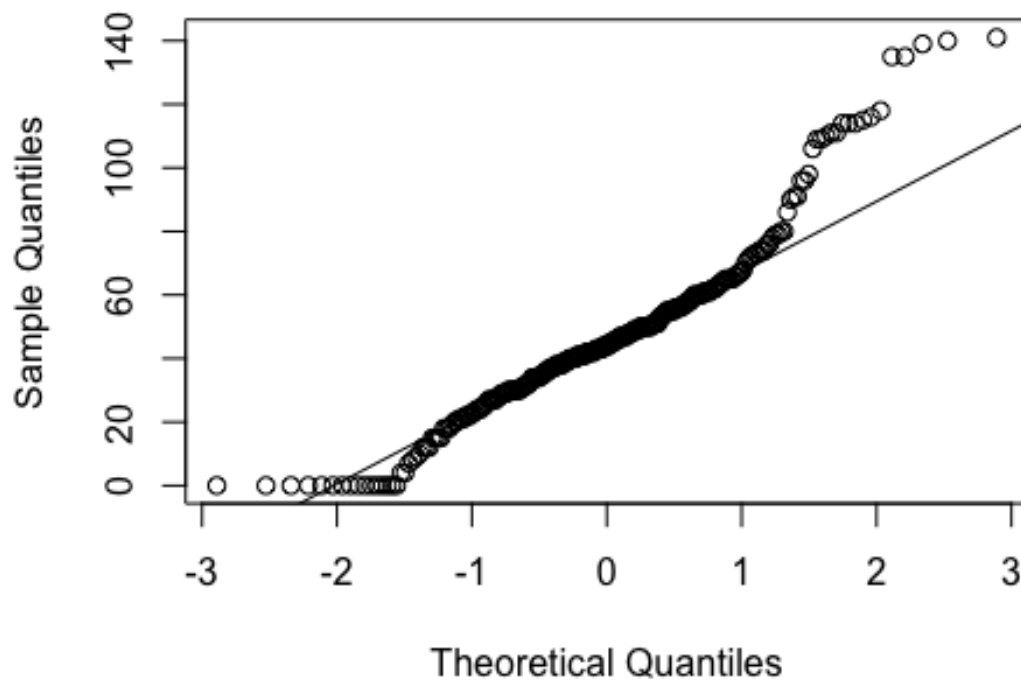
```
library(nortest)
n = lillie.test(M$Carbohydrates)
n

##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  M$Carbohydrates
## D = 0.098548, p-value = 2.081e-06
```

### Graficas de densidad de pobabilidad de Carbohydrates

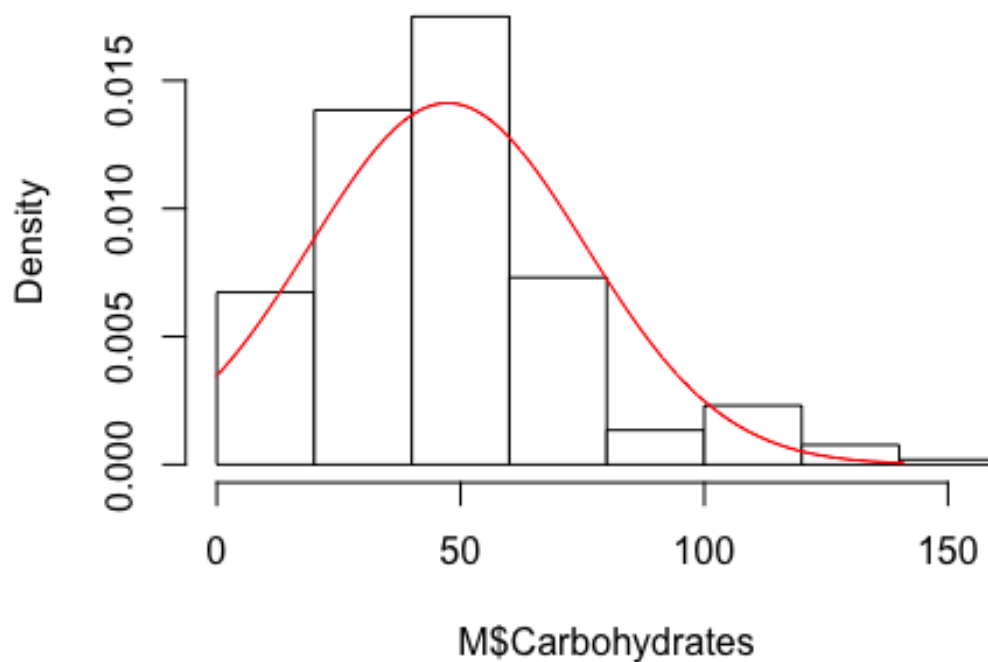
```
qqnorm(M$Carbohydrates)
qqline(M$Carbohydrates)
```

## Normal Q-Q Plot



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

## Histogram of M\$Carbohydrates



### Coeficiente de sesgo y curtosis de Carbohydrates

```
mean_carbohydrates = mean(M$Carbohydrates, na.rm = TRUE)
mean_carbohydrates

## [1] 47.34615

median_carbohydrates = median(M$Carbohydrates, na.rm = TRUE)
median_carbohydrates

## [1] 44

range_carbohydrates = range(M$Carbohydrates, na.rm = TRUE)
rango_m_carbohydrates = diff(range_carbohydrates)
rango_m_carbohydrates

## [1] 141
```

### Histograma de Carbohydrates

```
hist(M$Carbohydrates, freq=FALSE)
lines(density(M$Carbohydrates), col="red")
```

**Histogram of M\$Carbohydrates**

