

4Es_A01571214_LautaroCoteja

A01571214 - Lautaro Coteja

2024-08-13

R Markdown

4. Explorando Bases

Cargar y Leer Datos

```
file_path = "C:/Users/laute/Downloads/mc-donalds-menu.csv"
```

```
data = read.csv(file_path)
```

```
head(data)
```

```
##      Category                Item  Serving.Size  Calories
## 1 Breakfast              Egg McMuffin 4.8 oz (136 g)      300
## 2 Breakfast      Egg White Delight 4.8 oz (135 g)      250
## 3 Breakfast              Sausage McMuffin 3.9 oz (111 g)      370
## 4 Breakfast      Sausage McMuffin with Egg 5.7 oz (161 g)      450
## 5 Breakfast      Sausage McMuffin with Egg Whites 5.7 oz (161 g)      400
## 6 Breakfast              Steak & Egg McMuffin 6.5 oz (185 g)      430
##  Calories.from.Fat  Total.Fat  Total.Fat....Daily.Value.  Saturated.Fat
## 1                120         13                20             5
## 2                 70          8                12             3
## 3                200         23                35             8
## 4                250         28                43            10
## 5                210         23                35             8
## 6                210         23                36             9
##  Saturated.Fat....Daily.Value.  Trans.Fat  Cholesterol
## 1                 25          0                260
## 2                 15          0                 25
## 3                 42          0                 45
## 4                 52          0                285
## 5                 42          0                 50
## 6                 46          1                300
##  Cholesterol....Daily.Value.  Sodium  Sodium....Daily.Value.  Carbohydrates
## 1                 87        750                31             31
## 2                  8        770                32             30
## 3                 15        780                33             29
## 4                 95        860                36             30
## 5                 16        880                37             30
## 6                100        960                40             31
##  Carbohydrates....Daily.Value.  Dietary.Fiber
```

```

Dietary.Fiber....Daily.Value.
## 1          10          4
17
## 2          10          4
17
## 3          10          4
17
## 4          10          4
17
## 5          10          4
17
## 6          10          4
18
##  Sugars Protein Vitamin.A....Daily.Value. Vitamin.C....Daily.Value.
## 1      3      17          10          0
## 2      3      18           6          0
## 3      2      14           8          0
## 4      2      21          15          0
## 5      2      21           6          0
## 6      3      26          15          2
##  Calcium....Daily.Value. Iron....Daily.Value.
## 1          25          15
## 2          25           8
## 3          25          10
## 4          30          15
## 5          25          10
## 6          30          20

```

Variables seleccionadas

```

calorias = data$Calories
sodio = data$Sodium

```

Normalidad, QQPlot, y Histograma para Calorias

```

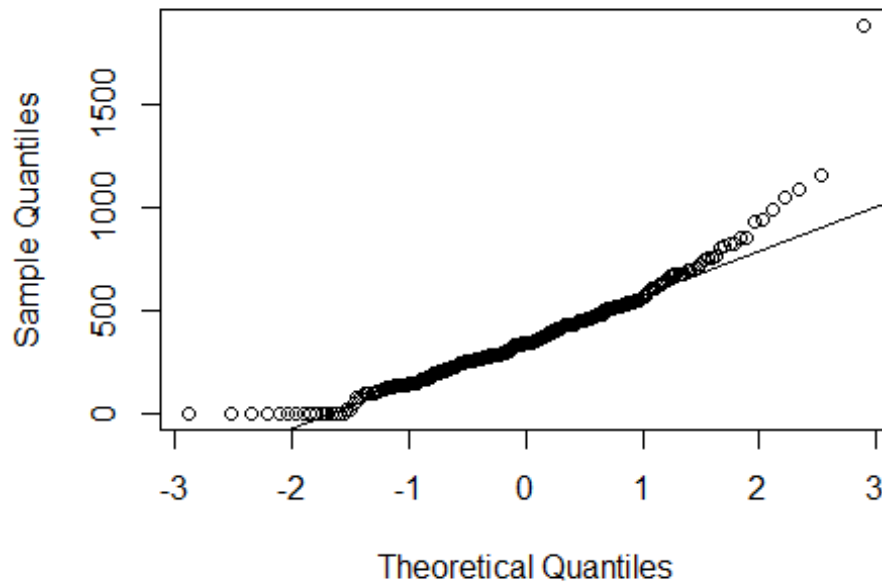
# Prueba de Normalidad para Calorias
shapiro.test(calorias)

##
##  Shapiro-Wilk normality test
##
## data:  calorias
## W = 0.91902, p-value = 1.119e-10

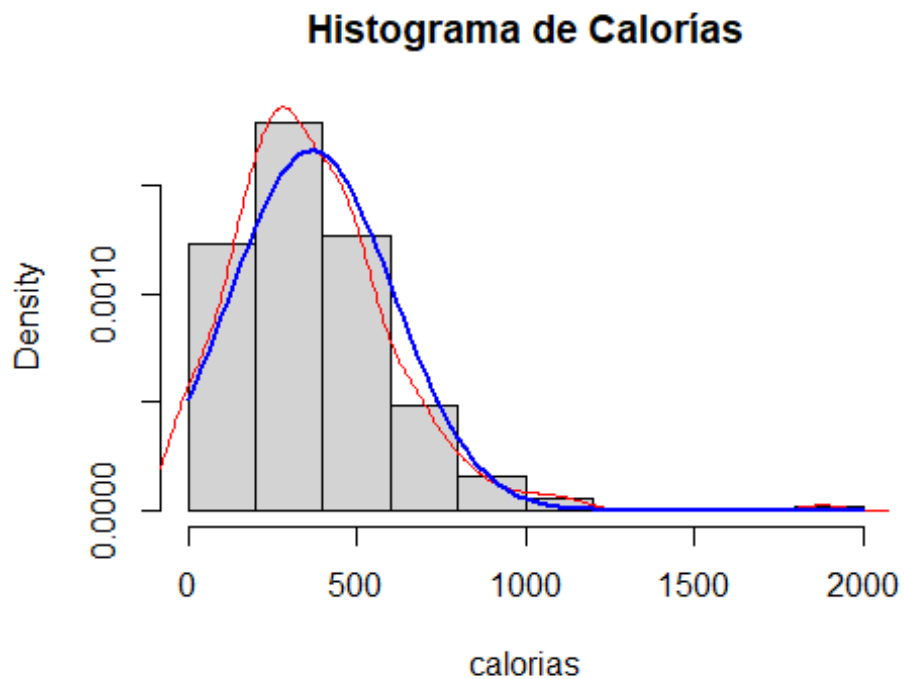
# QQPlot para Calorías
qqnorm(calorias)
qqline(calorias)

```

Normal Q-Q Plot



```
# Histograma para Calorias  
hist(calorias, freq=FALSE, main="Histograma de Calorías")  
lines(density(calorias), col="red")  
curve(dnorm(x, mean=mean(calorias), sd=sd(calorias)), add=TRUE, col="blue",  
lwd=2)
```



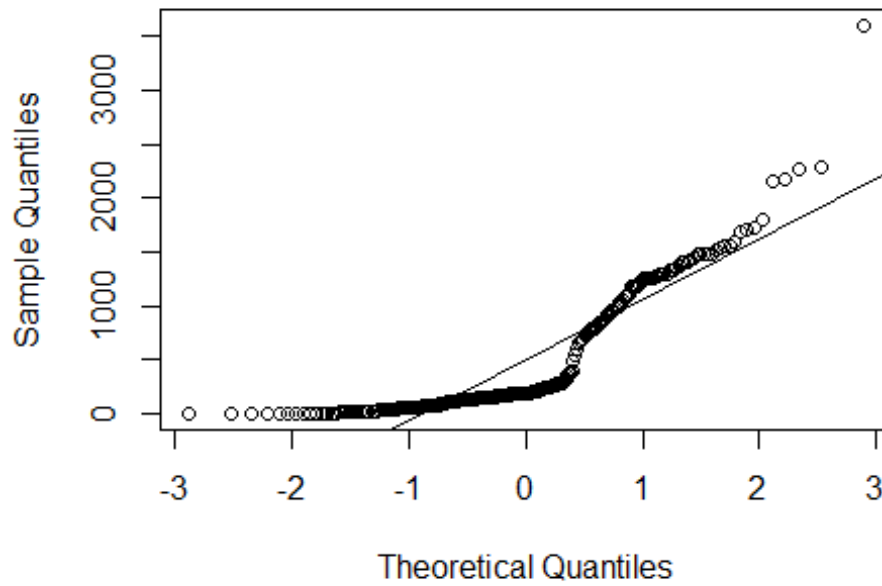
Normalidad, QQPlot, y Histograma para Sodio

```
# Prueba de Normalidad para Sodio  
shapiro.test(sodio)
```

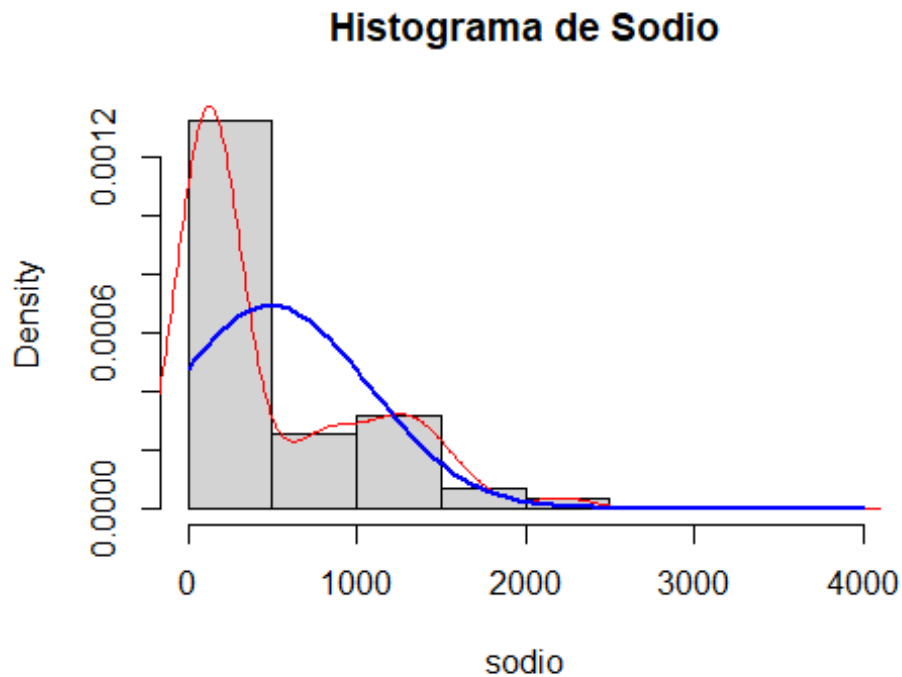
```
##  
##  Shapiro-Wilk normality test  
##  
## data:  sodio  
## W = 0.78191, p-value < 2.2e-16
```

```
# QQPlot para Sodio  
qqnorm(sodio)  
qqline(sodio)
```

Normal Q-Q Plot



```
# Histograma para Sodio
hist(sodio, freq=FALSE, main="Histograma de Sodio")
lines(density(sodio), col="red")
curve(dnorm(x, mean=mean(sodio), sd=sd(sodio)), add=TRUE, col="blue", lwd=2)
```



Coeficiente de sesgo y curtosis para Calorias y Sodio

```
# Libreria para calculo de sesgo y curtosis  
library(moments)
```

```
# Calorías  
skewness(calorias)
```

```
## [1] 1.444105
```

```
kurtosis(calorias)
```

```
## [1] 8.645274
```

```
# Sodio  
skewness(sodio)
```

```
## [1] 1.535166
```

```
kurtosis(sodio)
```

```
## [1] 5.796412
```

Media, Mediana y Rango Medio de Sodio y Calorias

```
# Calorías  
mean_calorias = mean(calorias)
```

```

median_calorias = median(calorias)
range_calorias = mean(range(calorias))

# Sodio
mean_sodio = mean(sodio)
median_sodio = median(sodio)
range_sodio = mean(range(sodio))

# Comparación
cat("Calorías - Media:", mean_calorias, "Mediana:", median_calorias, "Rango Medio:", range_calorias, "\n")

## Calorías - Media: 368.2692 Mediana: 340 Rango Medio: 940

cat("Sodio - Media:", mean_sodio, "Mediana:", median_sodio, "Rango Medio:", range_sodio, "\n")

## Sodio - Media: 495.75 Mediana: 190 Rango Medio: 1800

```

Datos Atipicos Calorias

```

# Cuartiles y rango intercuartílico
q1_calorias = quantile(calorias, 0.25)
q3_calorias = quantile(calorias, 0.75)
ri_calorias = IQR(calorias)

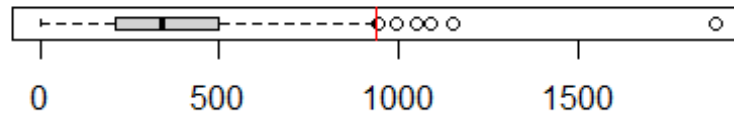
# Gráfico de caja y eliminación de atípicos
par(mfrow=c(2,1))
boxplot(calorias, horizontal=TRUE, main="Boxplot de Calorías")
abline(v=q3_calorias + 1.5 * ri_calorias, col="red")

# Eliminar atípicos
calorias_sin_outliers <- calorias[calorias < (q3_calorias + 1.5 *
ri_calorias)]
summary(calorias_sin_outliers)

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

```

Boxplot de Calorías



Datos Atipicos Sodio

```
# Cuartiles y rango intercuartílico
```

```
q1_sodio <- quantile(sodio, 0.25)
```

```
q3_sodio <- quantile(sodio, 0.75)
```

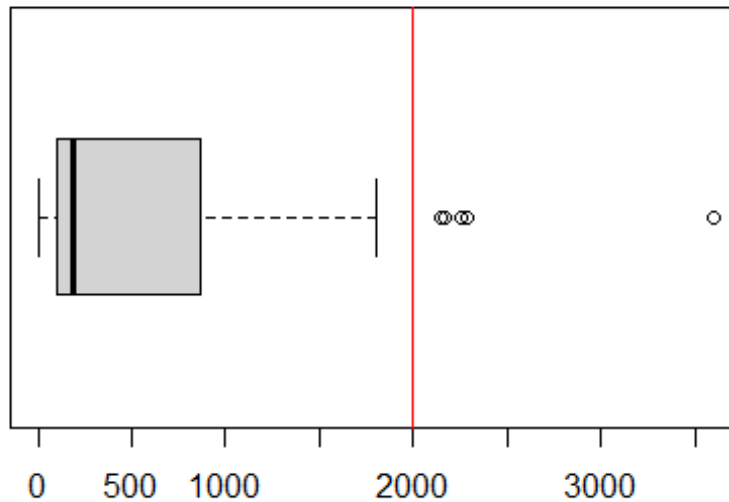
```
ri_sodio <- IQR(sodio)
```

```
# Gráfico de caja y eliminación de atípicos
```

```
boxplot(sodio, horizontal=TRUE, main="Boxplot de Sodio")
```

```
abline(v=q3_sodio + 1.5 * ri_sodio, col="red")
```


Boxplot de Sodio



```
# Eliminar atípicos
sodio_sin_outliers <- sodio[sodio < (q3_sodio + 1.5 * ri_sodio)]
summary(sodio_sin_outliers)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.0	95.0	190.0	456.6	830.0	1800.0

Conclusion de Resultados

En cuanto a normalidad, la prueba de Shapiro-Wilk indica si las variables tienen una distribución normal, el QPlot asiste en la visualización de si los datos siguen la línea de normalidad y los histogramas comparan la distribución observada con una normal / teórica. El sesgo indica si la distribución está sesgada positivamente (a la derecha) o negativamente (hacia la izquierda), la curtosis es como la agudeza de la distribución en comparación a una distribución normal. La mediana, media y rango, nos ayudan a entender la simetría de distribución. Y por último, los datos atípicos, ayudan a eliminar datos atípicos que puedan afectar al momento de analizar cualquier dato.