

M1_A4

Sofia Cantu

2024-08-13

```
M <- read.csv("~/Downloads/Codigos R/mc-donalds-menu.csv")

# Se escogieron las variables 'Calorias' y 'Carbohidratos' para analizarlas
calorias <- as.numeric(M$Calories)
carbohidratos <- as.numeric(M$Carbohydrates)

calorias <- na.omit(calorias)
carbohidratos <- na.omit(carbohidratos)
```

Análisis de datos atípicos

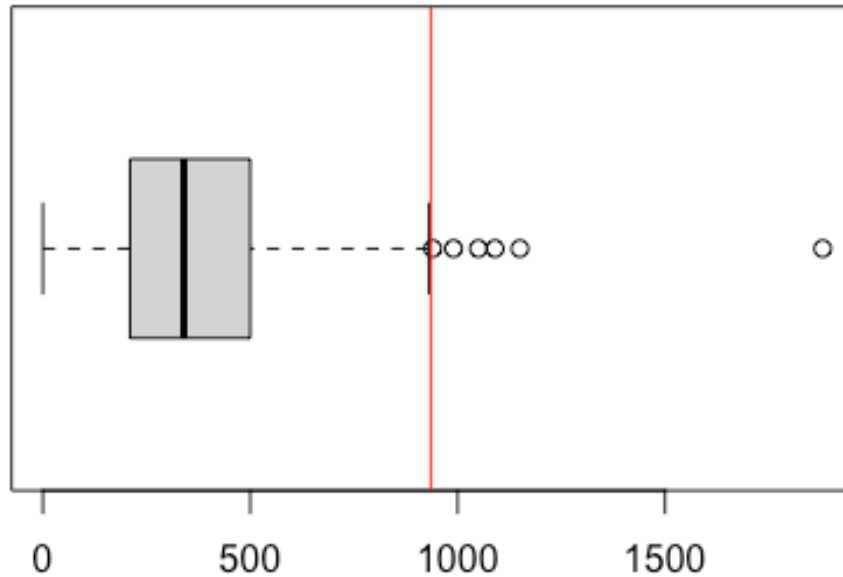
Calorias

```
### Rango Intercuartílico (IQR)
# Calculamos los cuartiles y rango intercuartílico
q1_cal <- quantile(calorias, 0.25)
q3_cal <- quantile(calorias, 0.75)
ri_cal <- q3_cal - q1_cal

# Identificamos datos atípicos (criterio de 1.5 veces el rango intercuartílico)
limite_inferior_cal <- q1_cal - 1.5 * ri_cal
limite_superior_cal <- q3_cal + 1.5 * ri_cal

# Boxplot
boxplot(calorias, horizontal=TRUE, main="Boxplot de Calorías")
abline(v=limite_superior_cal, col="red")
abline(v=limite_inferior_cal, col="blue")
```

Boxplot de Calorías



```
calorias_atipicas_IQR <- calorias[calorias < limite_inferior_cal | calorias >
limite_superior_cal]
print(calorias_atipicas_IQR)

## [1] 1090 1150 990 1050 940 1880

###Desviación Estándar

# Media y desviación estándar. Datos atípicos a más de 3 desviaciones desde
la media
media_cal <- mean(calorias)
sd_cal <- sd(calorias)
limite_inferior_sd_cal <- media_cal - 3 * sd_cal
limite_superior_sd_cal <- media_cal + 3 * sd_cal

calorias_atipicas_sd <- calorias[calorias < limite_inferior_sd_cal | calorias
> limite_superior_sd_cal]
print(calorias_atipicas_sd)

## [1] 1090 1150 1880
```

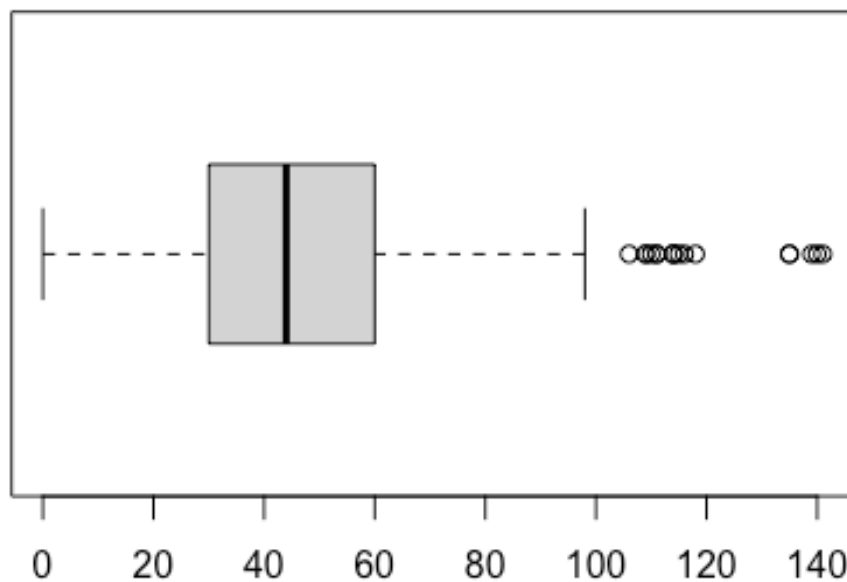
Carbohidratos

```
### Rango Intercuartílico (IQR)
# Calculamos los cuartiles y rango intercuartílico
q1_carb <- quantile(carbohidratos, 0.25)
q3_carb <- quantile(carbohidratos, 0.75)
ri_carb <- q3_carb - q1_carb

# Identificamos datos atípicos (criterio de 1.5 veces el rango
intercuartílico)
limite_inferior_carb <- q1_cal - 1.5 * ri_carb
limite_superior_carb <- q3_cal + 1.5 * ri_carb

# Boxplot
boxplot(carbohidratos, horizontal=TRUE, main="Boxplot de Carbohidratos")
abline(v=limite_superior_cal, col="red")
abline(v=limite_inferior_cal, col="blue")
```

Boxplot de Carbohidratos



```
carbohidratos_atipicos_IQR <- carbohidratos[carbohidratos <
limite_inferior_carb | carbohidratos > limite_superior_carb]
print(carbohidratos_atipicos_IQR)

## [1] 31 30 29 30 30 31 38 43 36 42 34 39 36 42 34 40 41
46
```

```
## [19] 38 48 47 44 48 46 57 55 56 51 56 50 55 111 116 110 115
60
## [37] 61 26 15 66 58 49 47 41 48 46 45 42 32 33 35 51 34
35
## [55] 34 35 44 55 42 58 44 57 43 65 51 43 40 43 41 40 56
42
## [73] 56 42 68 55 61 47 12 18 30 59 118 39 10 22 8 20 42
28
## [91] 37 30 34 27 32 25 30 44 67 15 4 4 30 32 21 22 7
53
## [109] 60 49 39 55 76 28 0 0 0 0 37 53 72 27 0 0 0
0
## [127] 37 54 74 27 12 23 21 34 44 65 0 0 0 0 0 36 45
54
## [145] 27 0 0 0 15 18 24 40 50 62 40 50 62 38 48 60 24
29
## [163] 37 15 19 25 41 51 63 40 51 63 39 49 60 24 30 38 49
60
## [181] 72 49 60 73 45 55 66 45 56 67 50 61 73 50 61 74 23
31
## [199] 47 22 29 43 21 29 43 20 27 41 9 12 18 41 50 70 41
50
## [217] 71 38 46 65 38 47 65 65 80 98 64 79 96 76 91 111 50
62
## [235] 79 47 58 74 50 61 78 86 109 135 90 114 140 91 114 141 109
135
## [253] 96 139 64 80 106 53 114 57
```

###Desviación Estándar

Media y desviación estándar. Datos atípicos a más de 3 desviaciones desde la media

```
media_carb <- mean(carbohidratos)
sd_carb <- sd(carbohidratos)
limite_inferior_sd_carb <- media_carb - 3 * sd_carb
limite_superior_sd_carb <- media_carb + 3 * sd_carb

carbohidratos_atipicos_sd <- carbohidratos[carbohidratos <
limite_inferior_sd_carb | carbohidratos > limite_superior_sd_carb]
print(carbohidratos_atipicos_sd)

## [1] 135 140 141 135 139
```

Pruebas de normalidad univariada

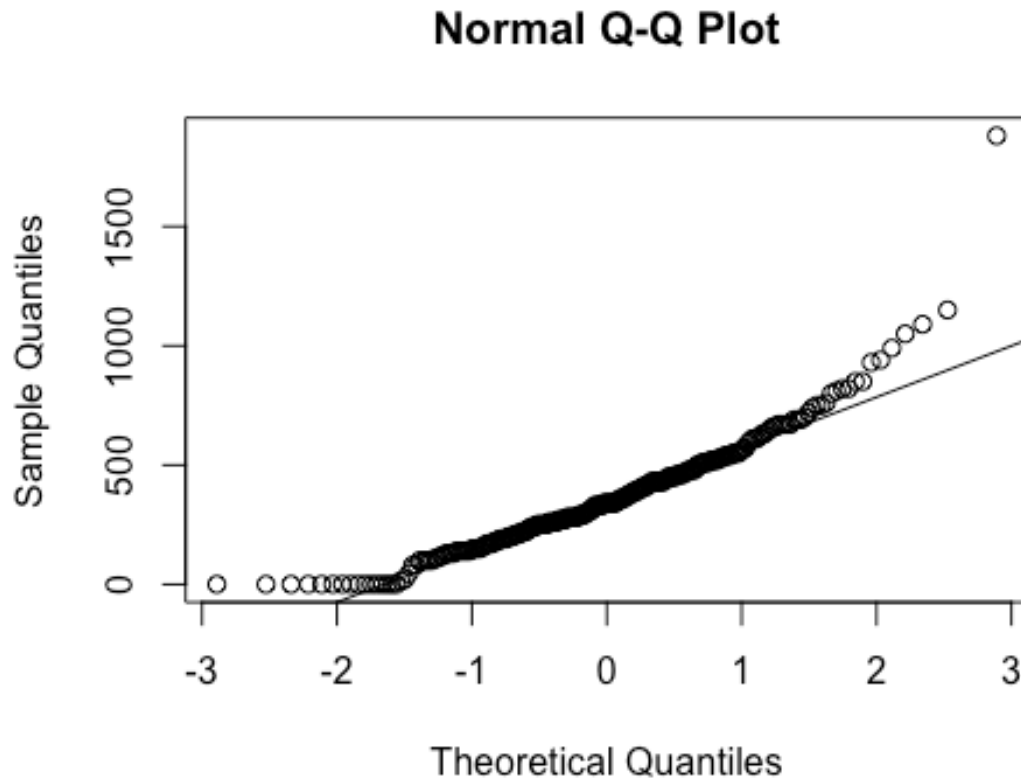
Calorias

Realiza la prueba de Shapiro-Wilk para verificar si los datos provienen de una distribución normal.

```
shapiro.test(calorias)
```

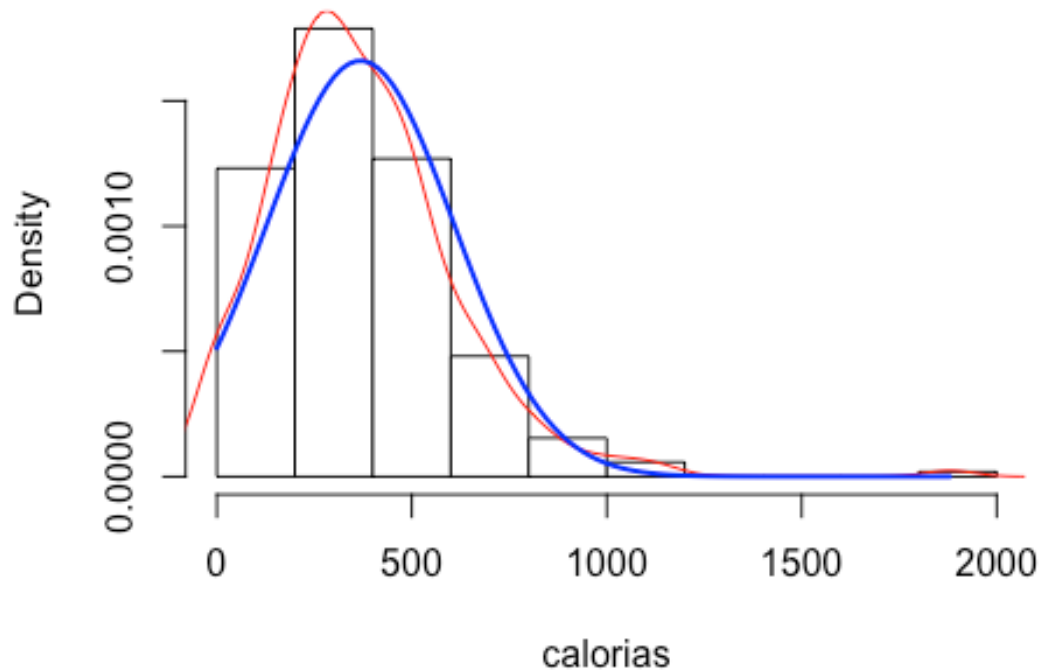
```
##
## Shapiro-Wilk normality test
##
## data:  calorias
## W = 0.91902, p-value = 1.119e-10

qqnorm(calorias)
qqline(calorias)
```



```
hist(calorias, prob=TRUE, col=0)
lines(density(calorias), col="red")
curve(dnorm(x, mean=mean(calorias), sd=sd(calorias)), from=min(calorias),
to=max(calorias), add=TRUE, col="blue", lwd=2)
```

Histogram of calorías



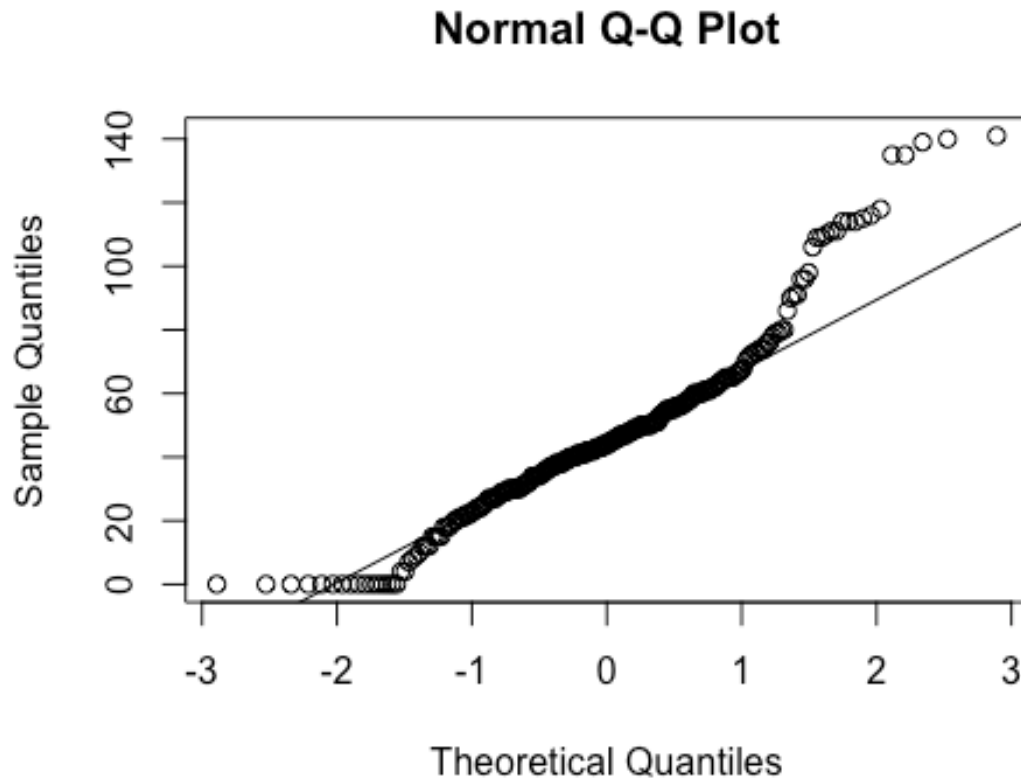
```
skewness(calorías) # Coeficiente de sesgo
## [1] 1.444105
kurtosis(calorías) # Coeficiente de curtosis
## [1] 8.645274
# Comparación de media, mediana y rango medio para Calorías
mean(calorías)
## [1] 368.2692
median(calorías)
## [1] 340
range(calorías)
## [1] 0 1880
```

Carbohidratos

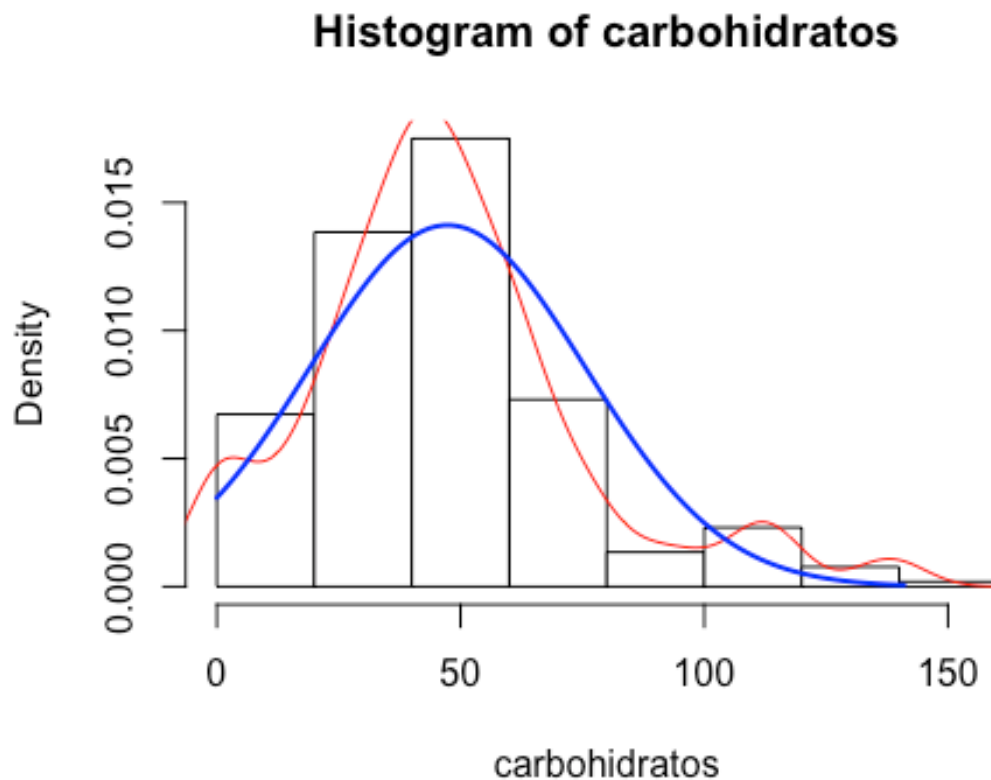
```
shapiro.test(carbohidratos) # Prueba de Shapiro-Wilk
```

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

```
qqnorm(carbohidratos)  
qqline(carbohidratos)
```



```
hist(carbohidratos, prob=TRUE, col=0)  
lines(density(carbohidratos), col="red")  
curve(dnorm(x, mean=mean(carbohidratos), sd=sd(carbohidratos)),  
from=min(carbohidratos), to=max(carbohidratos), add=TRUE, col="blue", lwd=2)
```



```
skewness(carbohidratos) # Coeficiente de sesgo
## [1] 0.9074253

kurtosis(carbohidratos) # Coeficiente de curtosis
## [1] 4.357538

# Comparación de media, mediana y rango medio para Carbohidratos
mean(carbohidratos)
## [1] 47.34615

median(carbohidratos)
## [1] 44

range(carbohidratos)
## [1] 0 141
```

Conclusiones

Para los dos casos, calorías y carbohidratos, mantendría los datos atípicos. Los productos del menu puede variar bastante en terminos de calorías y carbohidratos. Ejemplos son

comparar una ensalada con una hamburguesa, o un refresco azucarado y una taza de café negro. es crucial mostrar esa variedad en nuestro análisis. Los datos atípicos al final del día son productos en el extremo de las categorías, que siguen siendo parte del menú.