

▼ Actividad 2: Explorando Bases

Jorge Eduardo de León Reyna - A00829759

```
1 #instalando paquetes
2 install.packages("nortest")
3 install.packages("MVN")

Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

also installing the dependencies 'numDeriv', 'SparseM', 'MatrixModels', 'minqa', 'nloptr', 'RcppEigen', 'Rcpp', 'mnormt', 'g

Warning message in install.packages("MVN"):
"installation of package 'gsl' had non-zero exit status"
Warning message in install.packages("MVN"):
"installation of package 'energy' had non-zero exit status"
Warning message in install.packages("MVN"):
"installation of package 'MVN' had non-zero exit status"
```

▼ 1. Baja el archivo de trabajo: datos de McDonald

```
1 # se importan los datos
2 data = read.csv("/content/mc-donalds-menu-1.csv")
3
4 # se obtienen los valores de las columnas seleccionadas
5 calorias = data[["Calories.from.Fat"]]
6 azucares = data[["Sugars"]]
7
8 # se imprime una muestra de los datos
9 head(data)
10
11 # se obtiene el numero de registros de la tabla
12 nrow(data)
```

	Category	Item	Serving.Size	Calories	Calories.from.Fat	Total.Fat	Total.Fat....Daily.Value.	Saturated.Fat	Saturate
	<chr>	<chr>	<chr>	<int>	<int>	<dbl>	<int>	<dbl>	
1	Breakfast	Egg McMuffin	4.8 oz (136 g)	300	120	13	20	5	
2	Breakfast	Egg White Delight	4.8 oz (135 g)	250	70	8	12	3	
3	Breakfast	Sausage McMuffin	3.9 oz (111 g)	370	200	23	35	8	
4	Breakfast	Sausage McMuffin with Egg	5.7 oz (161 g)	450	250	28	43	10	
5	Breakfast	Sausage McMuffin with Egg Whites	5.7 oz (161 g)	400	210	23	35	8	
6	Breakfast	Steak & Egg McMuffin	6.5 oz (185 g)	430	210	23	36	9	
260									

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

- Calorias
- Carbohidratos
- Proteinas

- Sodio
- Azucares (Sugars)

▼ 3. Analisis de normalidad con pruebas vistas en clase

▼ 3.1 Exploracion de datos atipicos

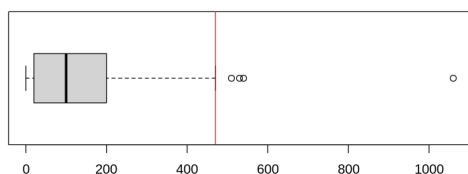
Calorias

```

1 # Calorias
2
3 q1 = quantile(calorias,0.25) # Cuantil 1 de la variable
4 q3 = quantile(calorias, 0.75) # Cuartil 3 de la variable
5 ri = IQR(calorias) # Rango intercuartílico de la variable
6 par(mfrow=c(2,1)) # Matriz de gráficos de 2x1
7
8 boxplot(calorias,horizontal=TRUE) #boxplot
9 abline(v=q3+1.5*ri,col="red") # linea vertical en el límite de los datos atípicos o extremos
10 calorias_limpias = data[data$calorias<q3+1.5*ri,c("calorias")] # En la matriz de datos, quitar datos más allá de 3 rangos inte
11 summary(calorias)

```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	20.0	100.0	127.1	200.0	1060.0



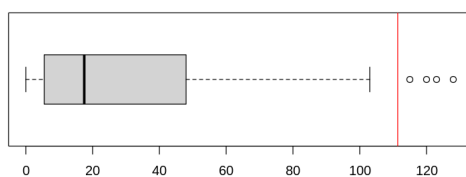
Azucares

```

1 # Azucares
2
3 q1 = quantile(azucares,0.25) # Cuantil 1 de la variable
4 q3 = quantile(azucares, 0.75) # Cuartil 3 de la variable
5 ri = IQR(azucares) # Rango intercuartílico de la variable
6 par(mfrow=c(2,1)) # Matriz de gráficos de 2x1
7
8 boxplot(azucares,horizontal=TRUE) #boxplot
9 abline(v=q3+1.5*ri,col="red") # linea vertical en el límite de los datos atípicos o extremos
10 calorias_limpias = data[data$azucares<q3+1.5*ri,c("azucares")] # En la matriz de datos, quitar datos más allá de 3 rangos inte
11 summary(azucares)
12

```

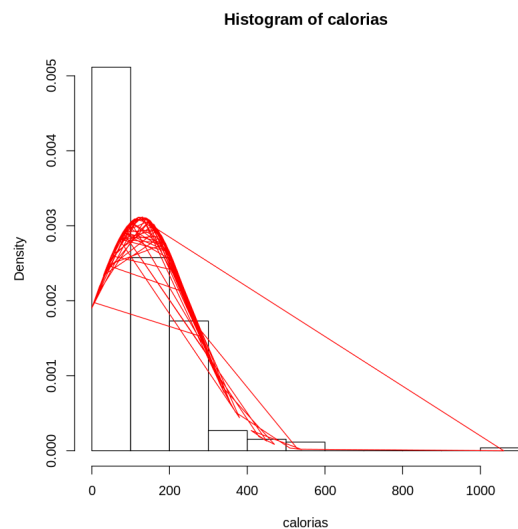
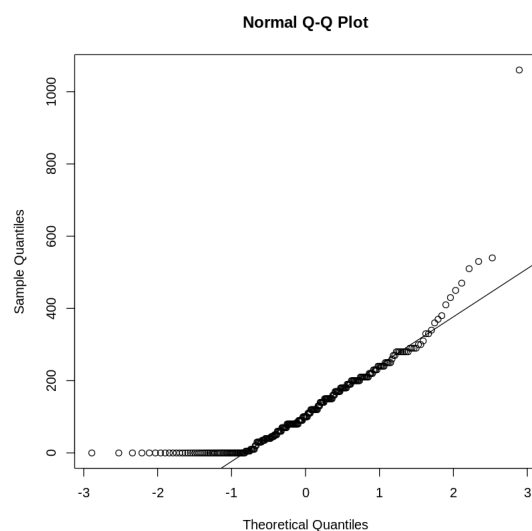
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	5.75	17.50	29.42	48.00	128.00



▼ 3. 2 Gráfico de densidad de probabilidad y comparacion con normalidad hipótetica

Calorias

```
1 # Calorias
2
3 qqnorm(calorias) # grafico qqplot
4 qqline(calorias)
5 hist(calorias,prob=TRUE,col=0) # histograma
6 x=seq(min(calorias),max(calorias),0.1)
7 y=dnorm(calorias,mean(calorias),sd(calorias))
8 lines(calorias,y,col="red")
```

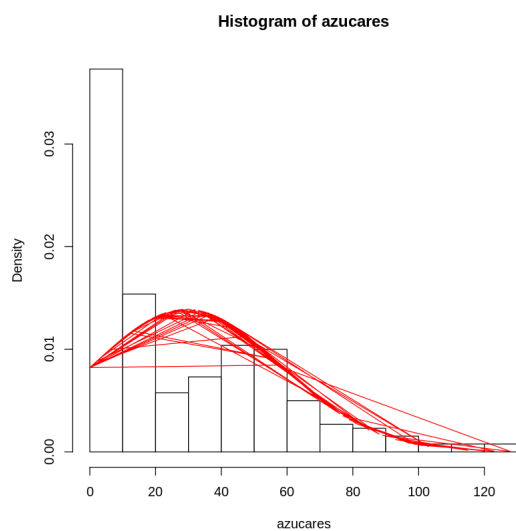
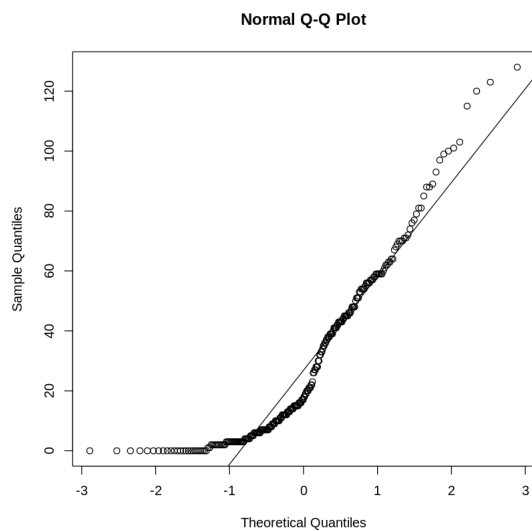


Azucares

```

1 # Azucares
2
3 qqnorm(azucares) # grafico qqplot
4 qqline(azucares)
5 hist(azucares,prob=TRUE,col=0) # histograma
6 x=seq(min(azucares),max(azucares),0.1)
7 y=dnorm(azucares,mean(azucares),sd(azucares))
8 lines(azucares,y,col="red")

```



▼ 3.3 Analisis de Curtosis y Sesgo

Calorias

```

1 #calorias
2
3 library(moments)
4 sesgo = skewness(calorias)
5 curtosis = kurtosis(calorias)
6
7 print(paste("Sesgo = ", sesgo))
8 print(paste("Curtosis = ", curtosis))

[1] "Sesgo = 2.13311030970577"
[1] "Curtosis = 13.3797589671064"

```

Azucares

```
1 #azucares
2
3 library(moments)
4 sesgo = skewness(azucares)
5 curtosis= kurtosis(azucares)
6
7 print(paste("Sesgo = ", sesgo))
8 print(paste("Curtosis = ", curtosis))

[1] "Sesgo = 1.02597720707632"
[1] "Curtosis = 3.48774420501022"
```

▼ 4. Analisis final e interpretación

Gracias a los diversos tests de normalidad aplicados a las dos variables seleccionadas (azúcares y calorías), se ha llegado a la conclusión de que en ambos casos **los datos no siguen una distribución normal**.

En un primer análisis, al representar gráficamente los datos mediante un diagrama de cajas (boxplot), se pudo observar que ambas variables presentan asimetría hacia la izquierda (sesgo hacia la derecha). Además, se identificó la presencia de un número significativo de valores atípicos, lo que sugiere una posible falta de normalidad.

El histograma de densidad de probabilidad confirmó visualmente la ausencia de normalidad en ambas variables. Se aprecia claramente que la media se encuentra inclinada hacia el lado izquierdo del gráfico. Este mismo patrón se reflejó en el gráfico Q-Q, donde los datos tienden a agruparse en un solo lado de la línea.

Finalmente, al calcular la curtosis y el sesgo para ambas variables, se obtuvo una confirmación cuantitativa de la falta de distribución normal. El sesgo positivo, evidenciado en la acumulación de datos hacia la izquierda, y el valor alejado de cero en la curtosis, reflejado en colas más gruesas en el gráfico correspondiente, respaldan la conclusión de la no normalidad de las variables, especialmente en el caso de las calorías.

1

✓ 0 s se ejecutó 09:12

● ×