

Transformacion

Cleber Perez

2024-08-14

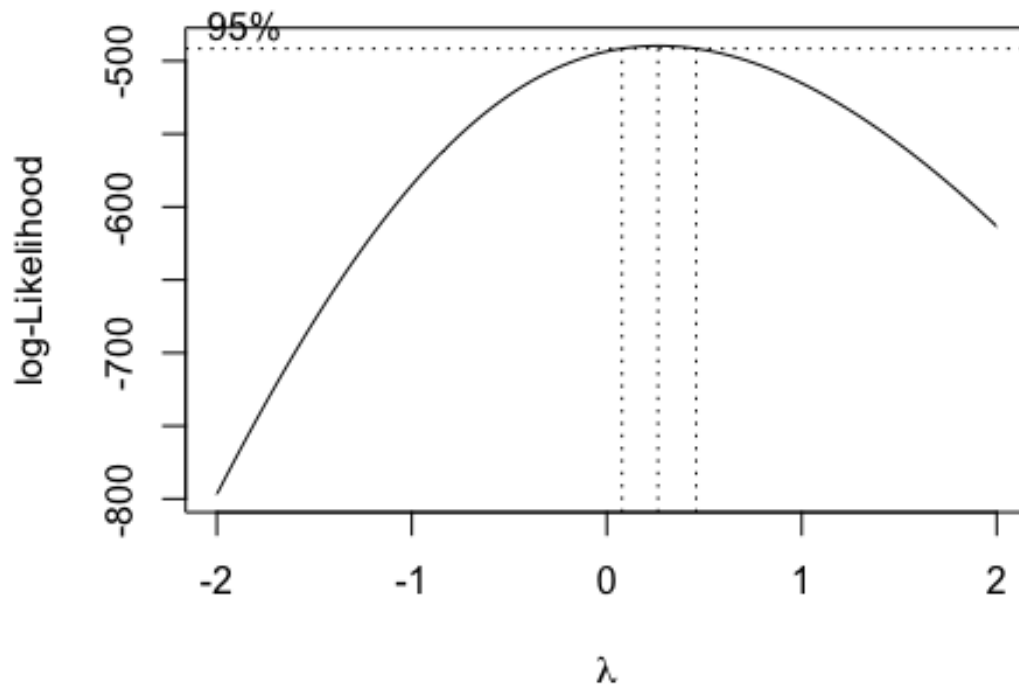
Leer archivo

```
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
```

Transformacion Box-Cox de Carbohydrates

```
library(MASS)
M2 = subset(M, M$Carbohydrates > 0)
min_val <- min(M2$Carbohydrates, na.rm = TRUE)
shift_val <- abs(min_val) + 1
M2$Carbohydrates_shifted <- M2$Carbohydrates + shift_val
boxcox_transform <- boxcox(lm(M2$Carbohydrates_shifted ~ 1, data = M), plotit = TRUE)
```



```
lambda_opt <- boxcox_transform$x[which.max(boxcox_transform$y)]
lambda_opt

## [1] 0.2626263

M2$Carbohydrates_transformed <- (M2$Carbohydrates_shifted^lambda_opt - 1) /
lambda_opt
```

Min, Max, media, mediana, cuartiles, sesgo y curtosis de Carbohydrates

```
min_value <- min(M2$Carbohydrates_transformed, na.rm = TRUE)
min_value

## [1] 2.972952

max_value <- max(M2$Carbohydrates_transformed, na.rm = TRUE)
max_value

## [1] 10.28772

mean_value <- mean(M2$Carbohydrates_transformed, na.rm = TRUE)
mean_value

## [1] 6.896801

median_value <- median(M2$Carbohydrates_transformed, na.rm = TRUE)
median_value

## [1] 6.885678

quartiles <- quantile(M2$Carbohydrates_transformed, probs = c(0.25, 0.75),
na.rm = TRUE)
quartiles

##      25%      75%
## 6.158223 7.634833

q1 <- quartiles[1]
q3 <- quartiles[2]
library(e1071)
skewness_value <- skewness(M2$Carbohydrates_transformed, na.rm = TRUE)
skewness_value

## [1] 0.019585

kurtosis_value <- kurtosis(M2$Carbohydrates_transformed, na.rm = TRUE)
kurtosis_value

## [1] 0.5401768
```

Histogramas Carbohidratos

```
carb1 = sqrt(M2$Carbohydrates + 1)
library(e1071)
```

```

skewness_value1 <- skewness(carb1, na.rm = TRUE)
skewness_value1

## [1] 0.3480223

kurtosis_value1 <- kurtosis(carb1, na.rm = TRUE)
kurtosis_value1

## [1] 0.590641

library(nortest)
D=ad.test(carb1)
D$p.value

## [1] 0.0001300569

carb2 = ((M2$Carbohydrates + 1)^lambda_opt - 1) / lambda_opt
skewness_value2 <- skewness(carb2, na.rm = TRUE)
skewness_value2

## [1] -0.1533194

kurtosis_value2 <- kurtosis(carb2, na.rm = TRUE)
kurtosis_value2

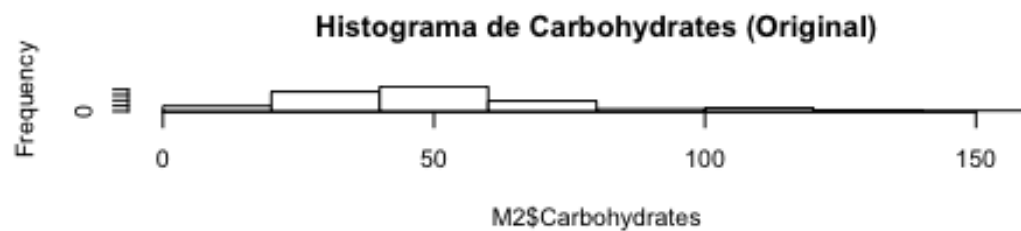
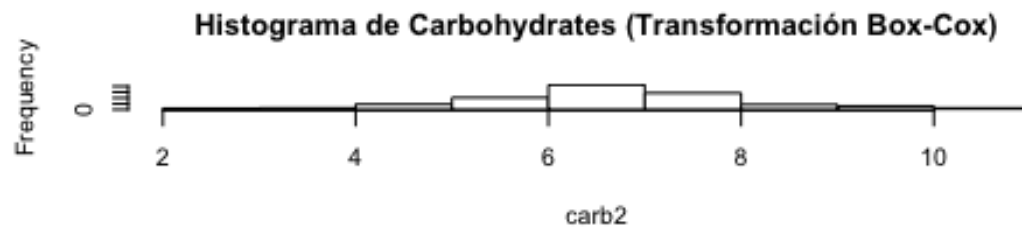
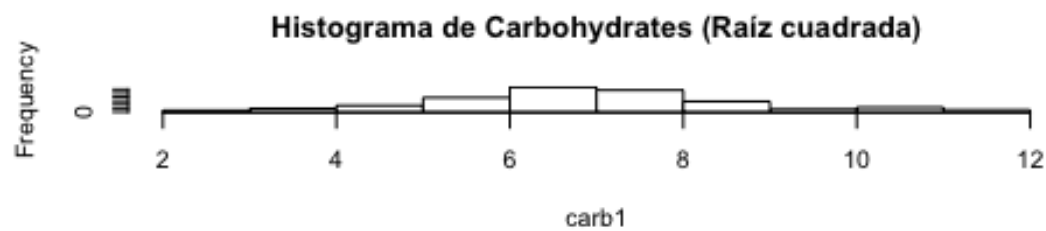
## [1] 0.7661949

D2=ad.test(carb2)
D2$p.value

## [1] 0.001162628

par(mfrow = c(3, 1))
hist(carb1, col = 0, main = "Histograma de Carbohydrates (Raíz cuadrada)")
hist(carb2, col = 0, main = "Histograma de Carbohydrates (Transformación Box-
Cox)")
hist(M2$Carbohydrates, col = 0, main = "Histograma de Carbohydrates
(Original)")

```



Prueba Normalidad Jarque Bera

```
library(moments)
```

```
##
```

```
## Attaching package: 'moments'
```

```
## The following objects are masked from 'package:e1071':
```

```
##
```

```
## kurtosis, moment, skewness
```

```
jarque.test(M2$Carbohydrates)
```

```
##
```

```
## Jarque-Bera Normality Test
```

```
##
```

```
## data: M2$Carbohydrates
```

```
## JB = 93.745, p-value < 2.2e-16
```

```
## alternative hypothesis: greater
```

```
jarque.test(carb1)
```

```
##
```

```
## Jarque-Bera Normality Test
```

```
##
```

```
## data: carb1
## JB = 8.8979, p-value = 0.01169
## alternative hypothesis: greater
```

```
jarque.test(carb2)
```

```
##
## Jarque-Bera Normality Test
##
## data: carb2
## JB = 7.4299, p-value = 0.02436
## alternative hypothesis: greater
```

Transformacion Yeo Johnson de Carbohydrates

```
library(VGAM)
```

```
## Loading required package: stats4
```

```
## Loading required package: splines
```

```
library(nortest) # Para usar ad.test()
```

```
lp <- seq(0, 1, 0.001) # Valores de Lambda propuestos
nlp <- length(lp)
n <- length(M2$Carbohydrates)
D <- matrix(as.numeric(NA), ncol = 2, nrow = nlp)
```

```
for (i in 1:nlp) {
  d <- yeo.johnson(M2$Carbohydrates, lambda = lp[i])
  p <- ad.test(d)
  D[i, ] <- c(lp[i], p$p.value)
}
```

```
N <- as.data.frame(D)
colnames(N) <- c("Lambda", "Valor_p")
```

```
# Filtrar valores no finitos
```

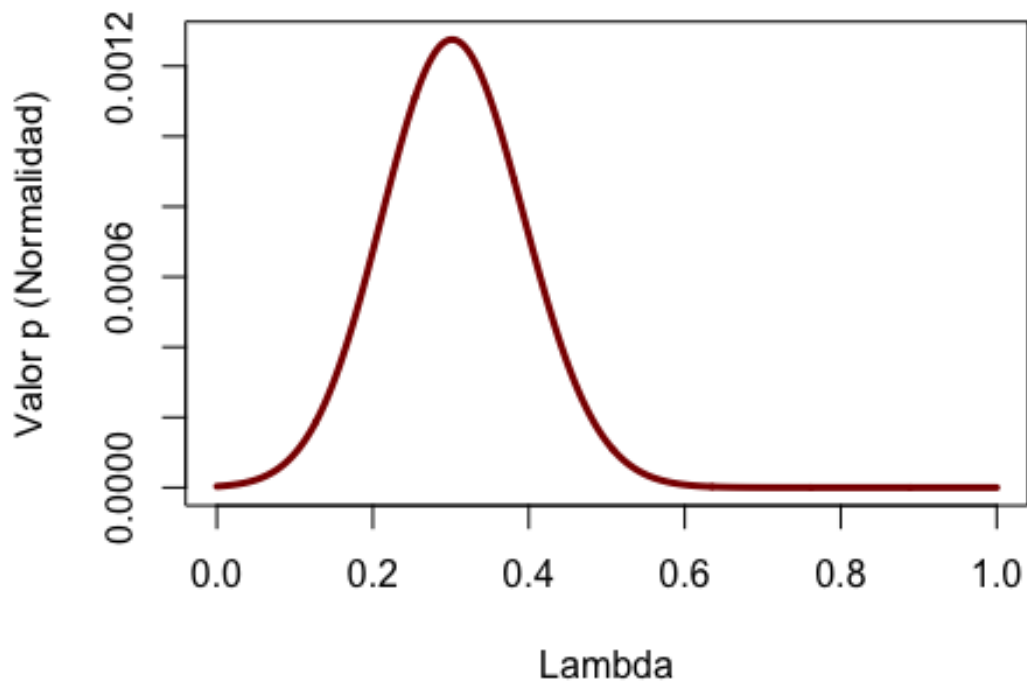
```
N <- N[is.finite(N$Lambda) & is.finite(N$Valor_p), ]
```

```
best_lambda <- N$Lambda[which.max(N$Valor_p)]
```

```
G <- data.frame(subset(N, N$Valor_p == max(N$Valor_p)))
```

```
# Graficar
```

```
plot(N$Lambda, N$Valor_p, type = "l",
     col = "darkred", lwd = 3,
     xlab = "Lambda",
     ylab = "Valor p (Normalidad)")
```



```
carb3 <- yeo.johnson(M2$Carbohydrates, lambda = best_lambda)
```

```
print(G)
```

```
##      Lambda      Valor_p
## 303  0.302 0.001275547
```

Min, Max, media, mediana, cuartiles, sesgo y curtosis de Carbohydrates

```
M2$Carbohydrates_transformed2 <- (M2$Carbohydrates_shifted^best_lambda - 1) /
best_lambda
```

```
min_value <- min(M2$Carbohydrates_transformed2, na.rm = TRUE)
min_value
```

```
## [1] 3.118199
```

```
max_value <- max(M2$Carbohydrates_transformed2, na.rm = TRUE)
max_value
```

```
## [1] 11.60389
```

```
mean_value <- mean(M2$Carbohydrates_transformed2, na.rm = TRUE)
mean_value
```

```
## [1] 7.572487
```

```

median_value <- median(M2$Carbohydrates_transformed2, na.rm = TRUE)
median_value

## [1] 7.544977

quartiles <- quantile(M2$Carbohydrates_transformed2, probs = c(0.25, 0.75),
na.rm = TRUE)
quartiles

##      25%      75%
## 6.700136 8.424072

q1 <- quartiles[1]
q3 <- quartiles[2]
library(e1071)
skewness_value <- skewness(M2$Carbohydrates_transformed2, na.rm = TRUE)
skewness_value

## [1] 0.09262205

kurtosis_value <- kurtosis(M2$Carbohydrates_transformed2, na.rm = TRUE)
kurtosis_value

## [1] 3.545537

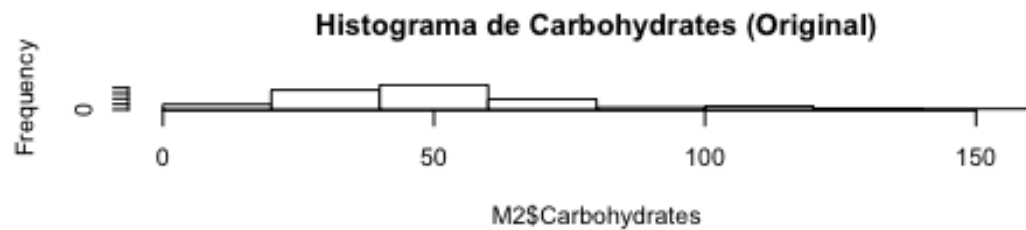
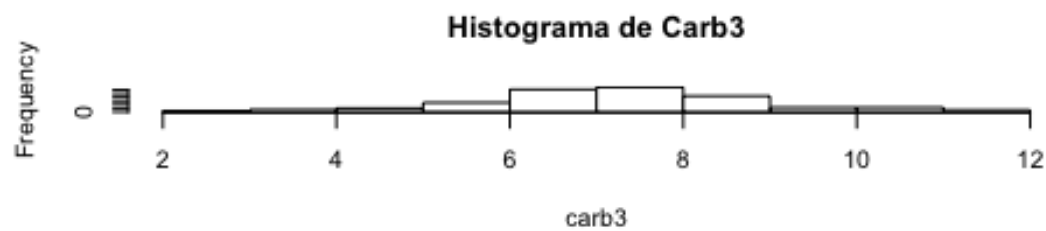
```

Histogramas Carbohidratos

```

par(mfrow = c(3, 1))
hist(carb3, col = 0, main = "Histograma de Carb3")
hist(M2$Carbohydrates, col = 0, main = "Histograma de Carbohydrates
(Original)")

```



Prueba Normalidad Jarque Bera

```
library(moments)
jarque.test(M2$Carbohydrates_transformed2)
```

```
##
##  Jarque-Bera Normality Test
##
## data:  M2$Carbohydrates_transformed2
## JB = 3.3746, p-value = 0.185
## alternative hypothesis: greater
```

```
jarque.test(M2$Carbohydrates)
```

```
##
##  Jarque-Bera Normality Test
##
## data:  M2$Carbohydrates
## JB = 93.745, p-value < 2.2e-16
## alternative hypothesis: greater
```