

Act5

Eliezer Cavazos

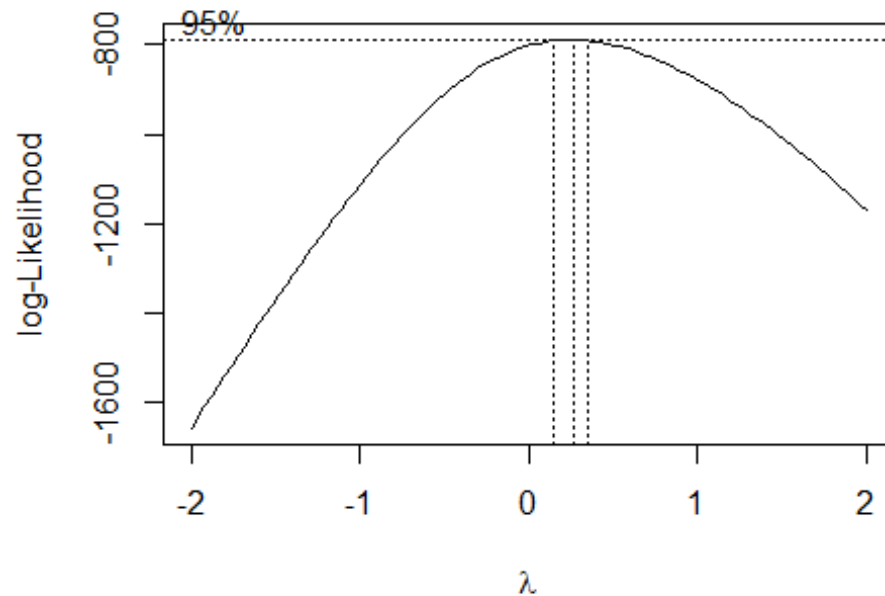
2024-08-14

```
oMcDonalds =  
read.csv("C:\\Users\\eliez\\OneDrive\\Desktop\\Clases\\mc-donalds-menu.csv")  
#Leer la base de datos
```

Box-Cox

```
M1 = oMcDonalds$Sugars
```

```
library(MASS)  
bc<-boxcox((M1+1)~1)# agarra calorias
```



```
lambdaSugar=bc$x[which.max(bc$y)]
```

```
lambdaSugar #Lambda
```

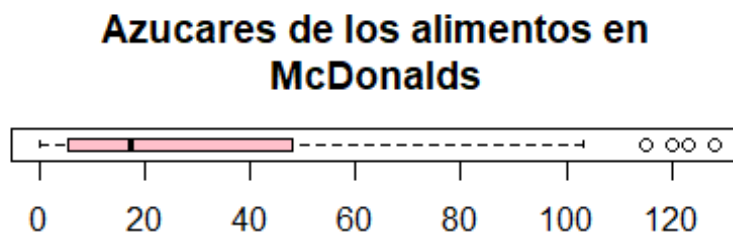
```
## [1] 0.2626263
```

2. Ecuaciones de los modelos

$$\text{Cal1} = \sqrt{x + 1} \quad \text{Cal2} = \frac{(x+1)^{0.2626263} - 1}{0.2626263}$$

3. Analisis Normalidad

```
#M2=subset(M1,M1>0)
par(mfrow=c(2,1))
boxplot(M1, horizontal = TRUE,col="pink", main="Azucares de los alimentos en
McDonalds")
#boxplot(M2, horizontal = TRUE,col="green", main="Calorias de Los alimentos
en
#McDonalds sin ceros")
```



3.1 Comparar las medidas

```
library(nortest)
library(e1071)

cal1=sqrt(M1+1)
cal2=((M1+1)^lambdaSugar-1)/lambdaSugar

D0=ad.test(M1)
D1=ad.test(cal1)
D2=ad.test(cal2)

m0=round(c(as.numeric(summary(M1)),kurtosis(M1),skewness(M1),D0$p.value),3)
```

```
m1=round(c(as.numeric(summary(cal1)),kurtosis(cal1),skewness(cal1),D1$p.value),3)
m2=round(c(as.numeric(summary(cal2)),kurtosis(cal2),skewness(cal2),D2$p.value),3)
```

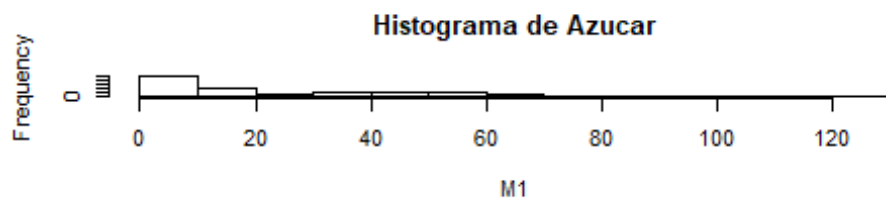
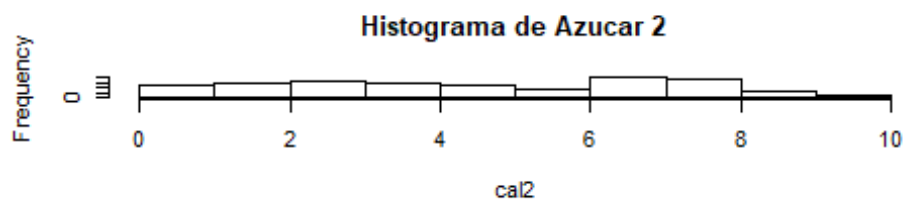
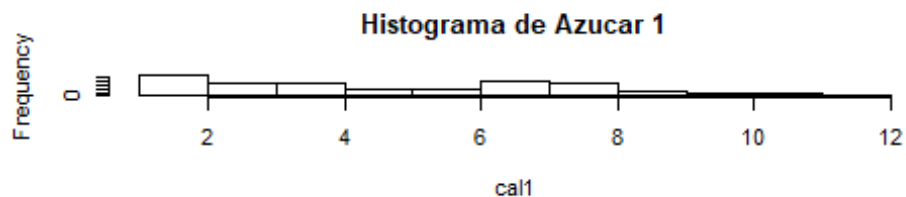
```
m<-as.data.frame(rbind(m0,m1,m2))
row.names(m)=c("Original","Primer modelo","Segundo Modelo")
names(m)=c("Minimo","Q1","Mediana","Media","Q3","Máximo","Curtosis","Sesgo","Valor p")
```

```
m
```

```
##           Minimo    Q1 Mediana  Media      Q3  Máximo Curtosis  Sesgo
## Original          0 5.750  17.500 29.423 48.000 128.000    0.461  1.020
## Primer modelo      1 2.597   4.301  4.825  7.000  11.358   -1.014  0.279
## Segundo Modelo     0 2.477   4.385  4.519  6.774   9.837   -1.113 -0.106
##           Valor p
## Original          0
## Primer modelo      0
## Segundo Modelo     0
```

3.2 Histograma

```
par(mfrow=c(3,1))
hist(cal1,col=0,main="Histograma de Azucar 1")
hist(cal2,col=0,main="Histograma de Azucar 2")
hist(M1,col=0,main="Histograma de Azucar")
```



3.3 Prueba Normalidad Anderson-Darling

```
print("Transformacion 1")

## [1] "Transformacion 1"

ad.test(cal1)

##
## Anderson-Darling normality test
##
## data: cal1
## A = 4.0816, p-value = 3.531e-10

print("Transformacion 2")

## [1] "Transformacion 2"

ad.test(cal2)

##
## Anderson-Darling normality test
##
## data: cal2
## A = 3.3722, p-value = 1.857e-08

print("Original")

## [1] "Original"

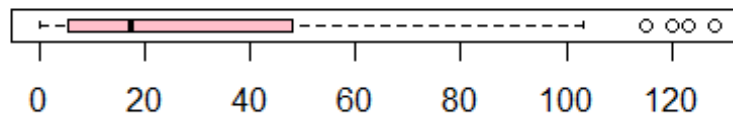
ad.test(M1)

##
## Anderson-Darling normality test
##
## data: M1
## A = 9.9899, p-value < 2.2e-16
```

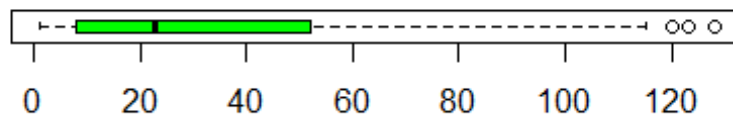
4 Anomalias

```
M2=subset(M1,M1>0)
par(mfrow=c(2,1))
boxplot(M1, horizontal = TRUE,col="pink", main="Azucares de los alimentos en McDonalds")
boxplot(M2, horizontal = TRUE,col="green", main="Azucares de los alimentos en McDonalds sin ceros")
```

Azucares de los alimentos en McDonalds

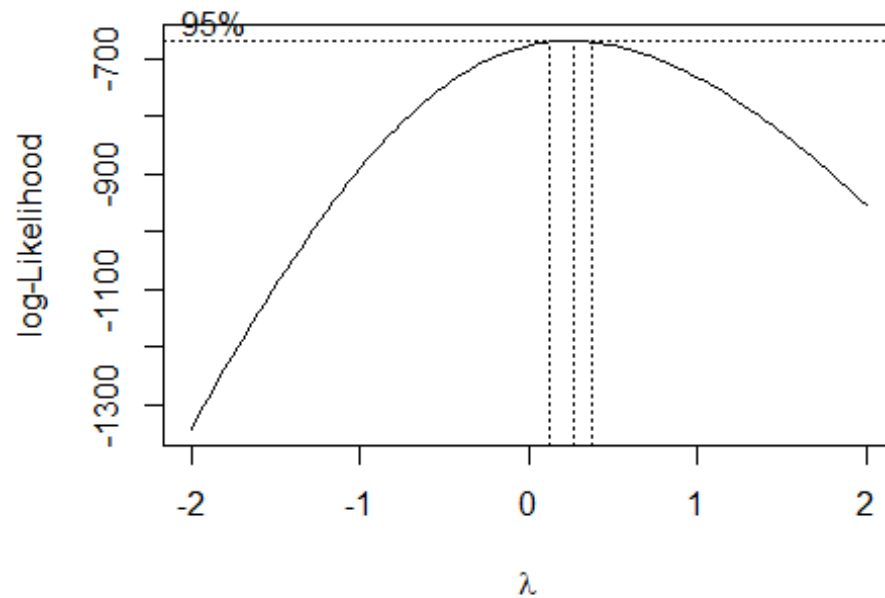


Azucares de los alimentos en McDonalds sin ceros



5. Encuentra Valor Lambda

```
bc<-boxcox(M2~1)# agarra calorías
```



```
lambdaSugar=bc$x[which.max(bc$y)]
```

```
lambdaSugar #Lambda
```

```
## [1] 0.2626263
```

5. Yeo Johnson

```
library(VGAM)
```

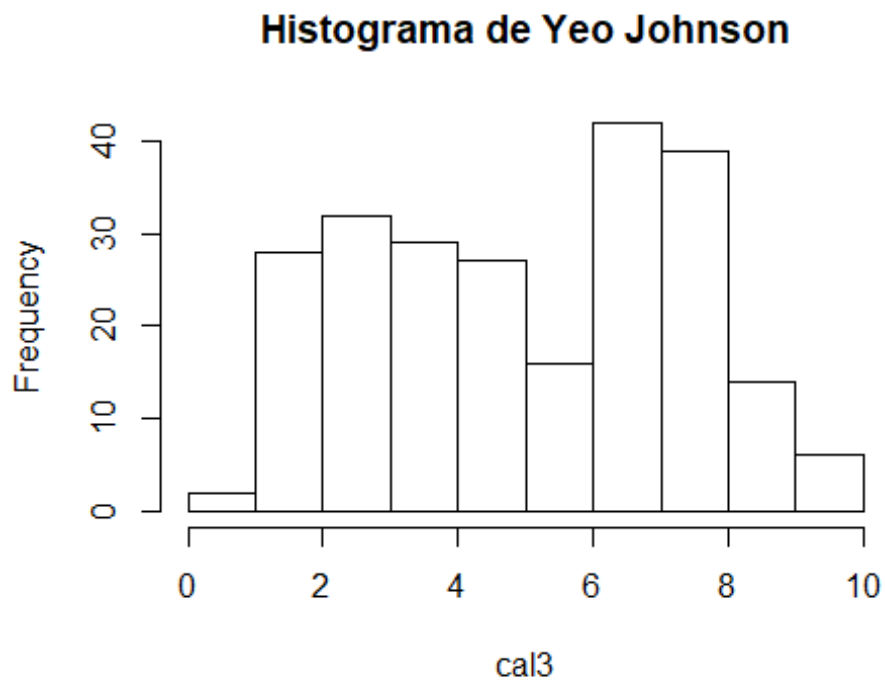
```
## Loading required package: stats4
```

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

```
cal1=sqrt(M2+1)
```

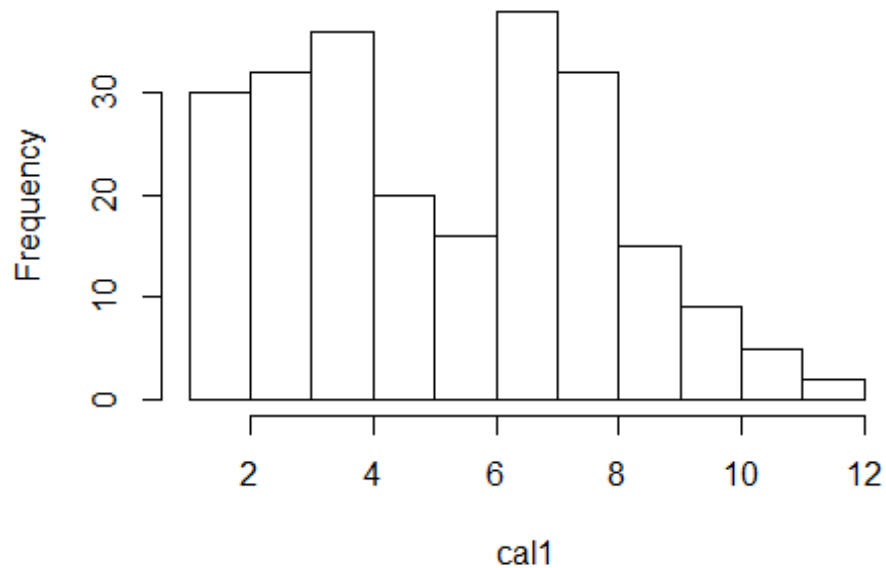
```
cal3<- yeo.johnson(M2, lambda = lambdaSugar)
```

```
hist(cal3,col=0,main="Histograma de Yeo Johnson")
```



```
hist(cal1,col=0,main="Histograma Original")
```

Histograma Original



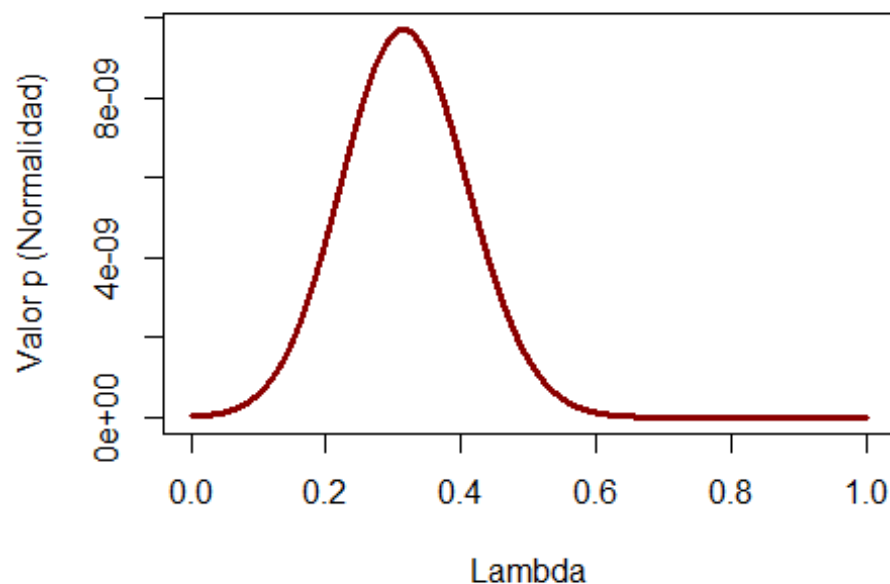
```
library(VGAM)
lp <- seq(0,1,0.001) # Valores de Lambda propuestos
nlp <- length(lp)
n=length(M2)
D <- matrix(as.numeric(NA),ncol=2,nrow=nlp)
d <- NA
for (i in 1:nlp){
  d= yeo.johnson(M2, lambda = lp[i])
  p=ad.test(d)
  D[i,]=c(lp[i],p$p.value)}

```

```
N=as.data.frame(D)
```

```
plot(N$V1,N$V2, type="l",
col="darkred", lwd=3,
xlab="Lambda",
ylab="Valor p (Normalidad)")

```



Lambda que maximiza el valor de P

```
G=data.frame(subset(N,N$V2==max(N$V2)))
```

```
G
```

```
##          V1          V2
## 315 0.314 9.726043e-09
```

6. Escribe la ecuación del modelo encontrado.

$$\text{Cal2} = \frac{(x+1)^{.314} - 1}{0.314}$$

7.1 Comparar Medidas

```
lambdaSugar = 0.314
```

```
cal1=sqrt(M2+1)
cal3=((M2+1)^lambdaSugar-1)/lambdaSugar
```

```
D0=ad.test(M2)
D1=ad.test(cal1)
D3=ad.test(cal3)
```

```
m0=round(c(as.numeric(summary(M2)),kurtosis(M2),skewness(M2),D0$p.value),3)
m1=round(c(as.numeric(summary(cal1)),kurtosis(cal1),skewness(cal1),D1$p.value),3)
m3=round(c(as.numeric(summary(cal3)),kurtosis(cal3),skewness(cal3),D1$p.value
```



```
),3)
```

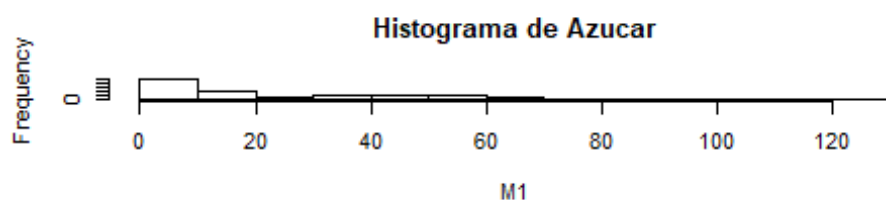
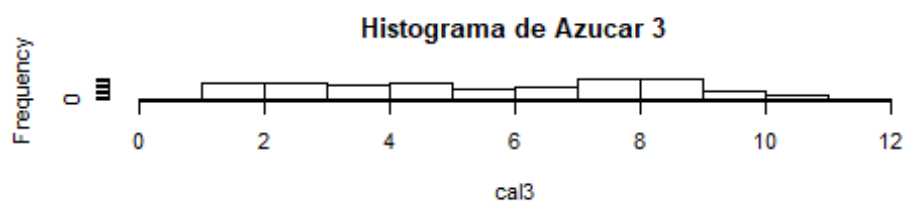
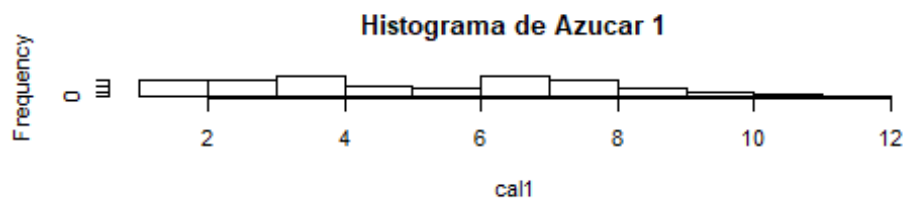
```
m<-as.data.frame(rbind(m0,m1,m2))
row.names(m)=c("Original","Primer modelo","Nuevo modelo")
names(m)=c("Minimo","Q1","Mediana","Media","Q3","Máximo","Curtosis","Sesgo","
Valor p")
```

m

```
##           Minimo      Q1 Mediana  Media      Q3  Máximo Curtosis  Sesgo
## Original      1.000 8.000  23.000 32.553 52.000 128.000   0.356  0.947
## Primer modelo  1.414 3.000   4.899  5.232  7.280  11.358  -1.001  0.277
## Nuevo modelo   0.000 2.477   4.385  4.519  6.774   9.837  -1.113 -0.106
##           Valor p
## Original           0
## Primer modelo      0
## Nuevo modelo       0
```

7.2 Histograma Modelos Obtenidos

```
par(mfrow=c(3,1))
hist(cal1,col=0,main="Histograma de Azucar 1")
hist(cal3,col=0,main="Histograma de Azucar 3")
hist(M1,col=0,main="Histograma de Azucar")
```



7.3 Prueba de Normalidad de Anderson-Darling

```
print("Transformacion 1")
```

```
## [1] "Transformacion 1"

ad.test(cal1)

##
## Anderson-Darling normality test
##
## data: cal1
## A = 3.826, p-value = 1.459e-09

print("Transformacion 3")

## [1] "Transformacion 3"

ad.test(cal3)

##
## Anderson-Darling normality test
##
## data: cal3
## A = 3.4867, p-value = 9.726e-09

print("Original")

## [1] "Original"

ad.test(M2)

##
## Anderson-Darling normality test
##
## data: M2
## A = 7.4424, p-value < 2.2e-16
```

8. Definir mejor transoformacion

Pude identificar que la mejor transformación para el caso de los datos que hay de McDonald's es la transformación de Box-Cox, ya que con base en las necesidades y que los datos de la columna de Azúcares no contienen valores negativos hace que no se necesite una transformación tan Robusta como la de Yeo Johnson y quede mejor usar Box-Cox, aunque en este caso en los dos casos salio un valor P muy pequeño despues de las transformaciones eso significa que tal vez se necesite una diferente transformación para este caso

9. Ventajas y Desventajas de los modelos de Box Cox y de Yeo Johnson.

Ventajas Box-Cox: - Cuando no se tiene un valor de lambda - Útil cuando la variabilidad de Y cambia como una función de X. Desventajas Box-Cox: - Solo acepta numeros positivos

Ventajas Yeo Johnson: - Mas robusto que Box-Cox - Acepta valores positivos, negativos y Cero Desventajas Yeo Johnson: - Es una extension de Box-Cox

10.1 3 diferencias entre lo que es la transformación y el escalamiento de los datos

1. El escalamiento de datos se encarga de que cada variable de nuestro set tenga el mismo rango de valores
2. La transformación cambia la distribución de los datos
3. La transformación es utilizada para corregir sesgos o estabilizar la varianza

10.2 Indica cuando es necesario utilizar cada uno

La transformación se usa cuando queremos cambiar la distribución de los datos por ejemplo de una no normalizada a una distribución normalizada, mientras que el escalamiento de los datos se encarga de mantener el mismo rango de valores en nuestro set de datos