

# Actividad Integradora 1

Cleber Perez

2024-08-20

## Leer archivo

```
Fd=read.csv("food_data_g.csv") #Leer la base de datos
Fd$variable # para llamar una variable, aunque también la puedes leer con
corchetes cuadrados M[renglón, columna]

## NULL
```

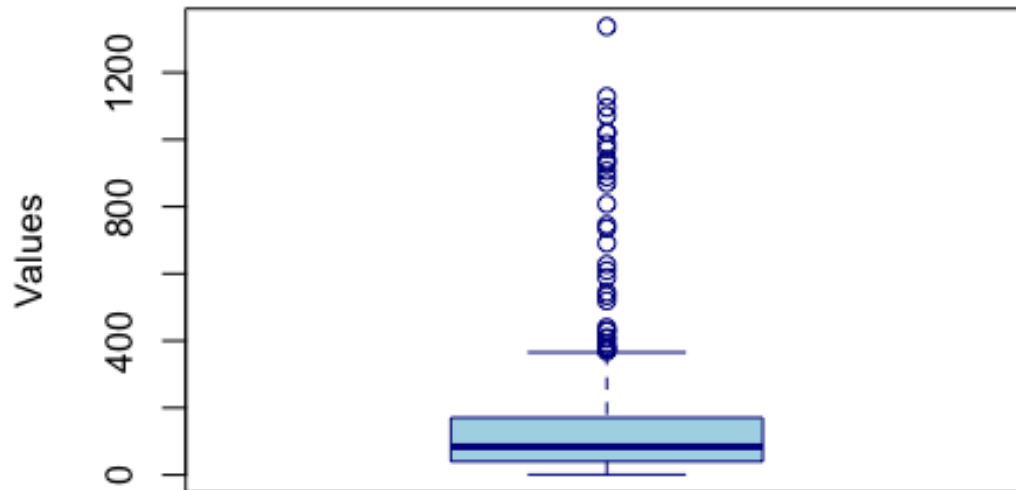
## Punto 1. Análisis descriptivo de la variable

### Análisis descriptivo de la variable (Densidad Nutricional)

#### Graficar el diagrama de caja y bigote

```
boxplot(Fd$Nutrition.Density,
        main = "Densidad Nutricional",
        ylab = "Values",
        col = "lightblue",
        border = "darkblue")
```

## Densidad Nutricional



### Medias que te ayudan a sacar Datos Atipicos

```
media_value <- median(Fd$Nutrition.Density, na.rm = TRUE)
media_value

## [1] 84.1

quartiles <- quantile(Fd$Nutrition.Density, na.rm = TRUE)
quartiles

##      0%      25%      50%      75%     100%
##  1.0250  39.9335  84.1000 170.6915 1337.0000

q1 <- quartiles[1]
q3 <- quartiles[2]
iqr_value <- IQR(Fd$Nutrition.Density, na.rm = TRUE)
iqr_value

## [1] 130.758

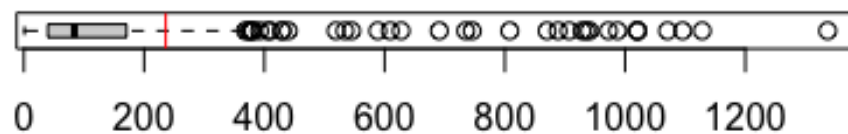
sd_value <- sd(Fd$Nutrition.Density, na.rm = TRUE)
sd_value

## [1] 187.0476
```

### Cota de 1.5 rangos intercuatílicos para Datos Atípicos

```
par(mfrow=c(2,1))
boxplot(Fd$Nutrition.Density, horizontal=TRUE, ylim=c(39.9335, 1337))
abline(v=q3+1.5*iqr_value, col="red")
X1= Fd[Fd$Nutrition.Density<q3+1.5*iqr_value, c("X")]
summary(X1)
```

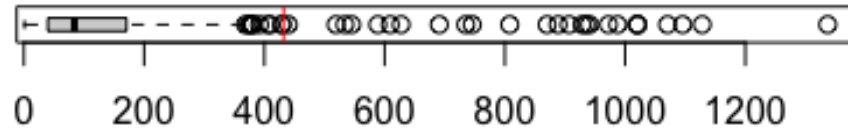
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0   171.0   294.0   292.8   427.0   550.0
```



### Cota de 3 rangos intercuatílicos para Datos Extremos

```
par(mfrow=c(2,1))
boxplot(Fd$Nutrition.Density, horizontal=TRUE, ylim=c(39.9335, 1337))
abline(v=q3+3*iqr_value, col="red")
X2= Fd[Fd$Nutrition.Density<q3+3*iqr_value, c("X")]
summary(X2)
```

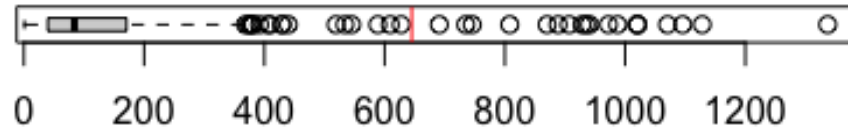
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0   150.0   282.0   281.5   414.0   550.0
```



### Cota de 3 desviaciones estandar alrededor de la media

```
lower_bound_sd <- media_value - 3 * sd_value
upper_bound_sd <- media_value + 3 * sd_value
par(mfrow = c(2, 1))
boxplot(Fd$Nutrition.Density, horizontal = TRUE, ylim = c(39.9335, 1337))
abline(v = lower_bound_sd, col = "blue")
abline(v = upper_bound_sd, col = "red")
X3 <- Fd[Fd$Nutrition.Density < lower_bound_sd | Fd$Nutrition.Density >
upper_bound_sd, c("X")]
summary(X3)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	11.0	25.5	34.0	131.7	75.0	502.0



## Pruebas de Normalidad Univariada

```
library(moments)
jarque.test(Fd$Nutrition.Density)

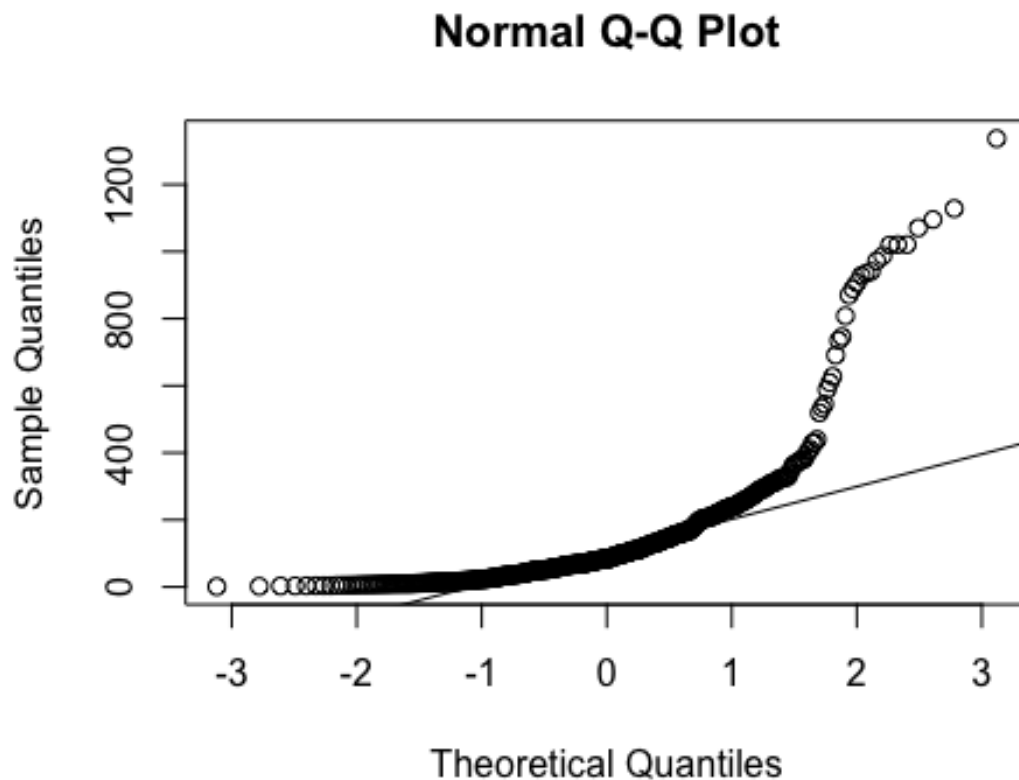
##
##  Jarque-Bera Normality Test
##
## data:  Fd$Nutrition.Density
## JB = 4358.4, p-value < 2.2e-16
## alternative hypothesis: greater
```

```
library(nortest)
ad.test(Fd$Nutrition.Density)

##
##  Anderson-Darling normality test
##
## data:  Fd$Nutrition.Density
## A = 52.747, p-value < 2.2e-16
```

## Grafica y respectivo qqplot

```
qqnorm(Fd$Nutrition.Density)
qqline(Fd$Nutrition.Density)
```



### Coeficiente de sesgo y coeficiente de curtosis

```
library(e1071)
```

```
##
```

```
## Attaching package: 'e1071'
```

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

```
##
```

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

```
skewness_value1 <- skewness(Fd$Nutrition.Density, na.rm = TRUE)
```

```
skewness_value1
```

```
## [1] 3.208552
```

```
kurtosis_value1 <- kurtosis(Fd$Nutrition.Density, na.rm = TRUE)
```

```
kurtosis_value1
```

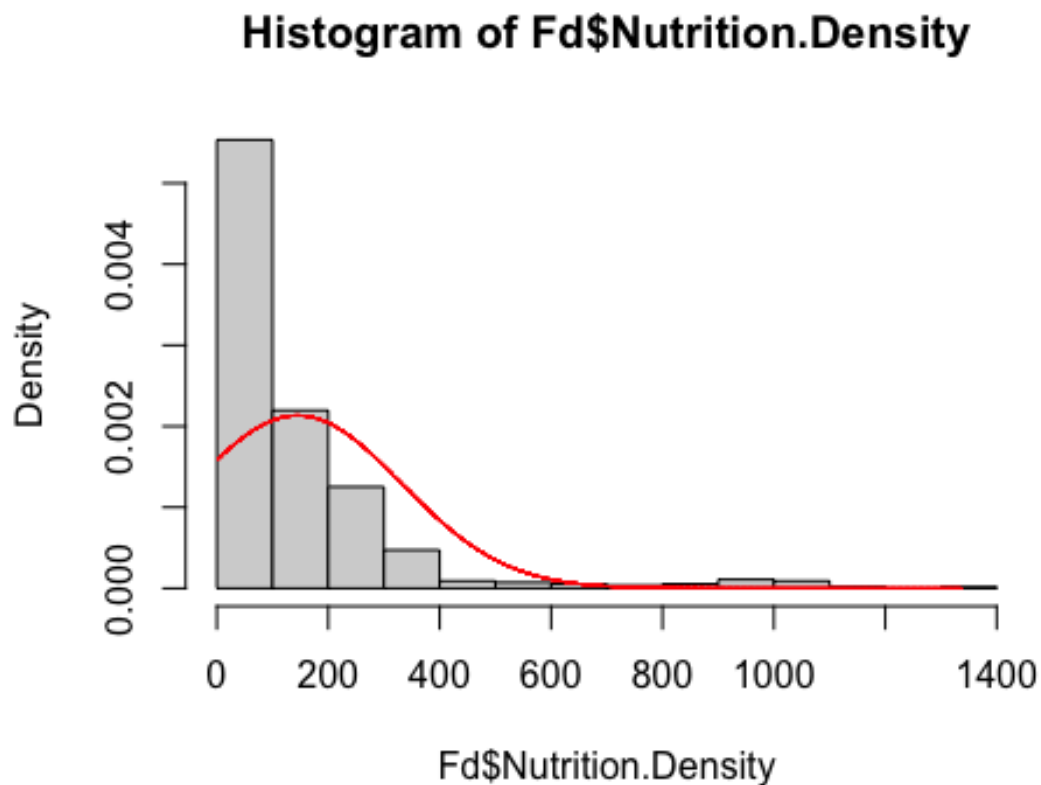
```
## [1] 12.12833
```

### Grafico de densidad empírica y teórica suponiendo normalidad en la variable

```
hist(Fd$Nutrition.Density, freq=FALSE)
```

```
x=seq(min(Fd$Nutrition.Density),max(Fd$Nutrition.Density),0.1)
```

```
y=dnorm(x,mean(Fd$Nutrition.Density),sd(Fd$Nutrition.Density))
lines(x,y,col="red")
```



En cuanto a la grafica podemos suponer que no existe una distribucion normal, dentro del grafico de qqplot podemos observar una clara Asimetria Positiva, tambien dentro del valor de curtosis podemos observar que tiene un valor muy alto que nos indica una distribucion leptocurtica.

## Punto 2. Transformación a normalidad

### Transformacion Yeo-Johnson y Box-Cox

```
library(VGAM)

## Loading required package: stats4
## Loading required package: splines

library(nortest)

lp <- seq(0, 1, 0.001) # Valores de Lambda propuestos
```

```

nlp <- length(lp)
n <- length(Fd$Nutrition.Density)
D <- matrix(as.numeric(NA), ncol = 2, nrow = nlp)

for (i in 1:nlp) {
  d <- yeo.johnson(Fd$Nutrition.Density, lambda = lp[i])
  p <- ad.test(d)
  D[i, ] <- c(lp[i], p$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), ]

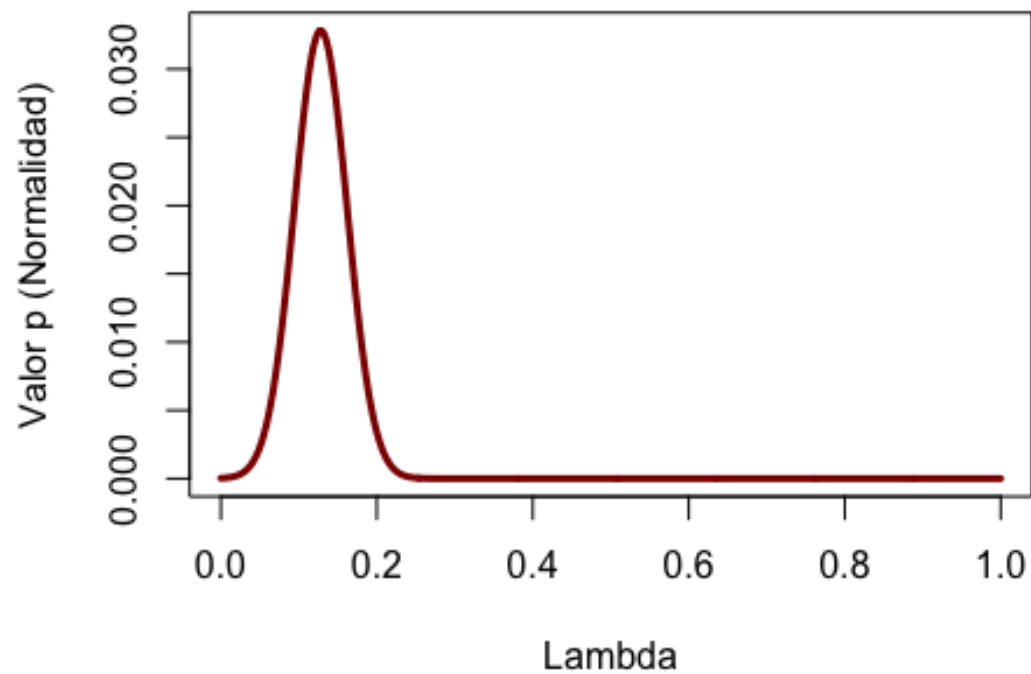
best_lambda <- N$Lambda[which.max(N$Valor_p)]

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)")

```



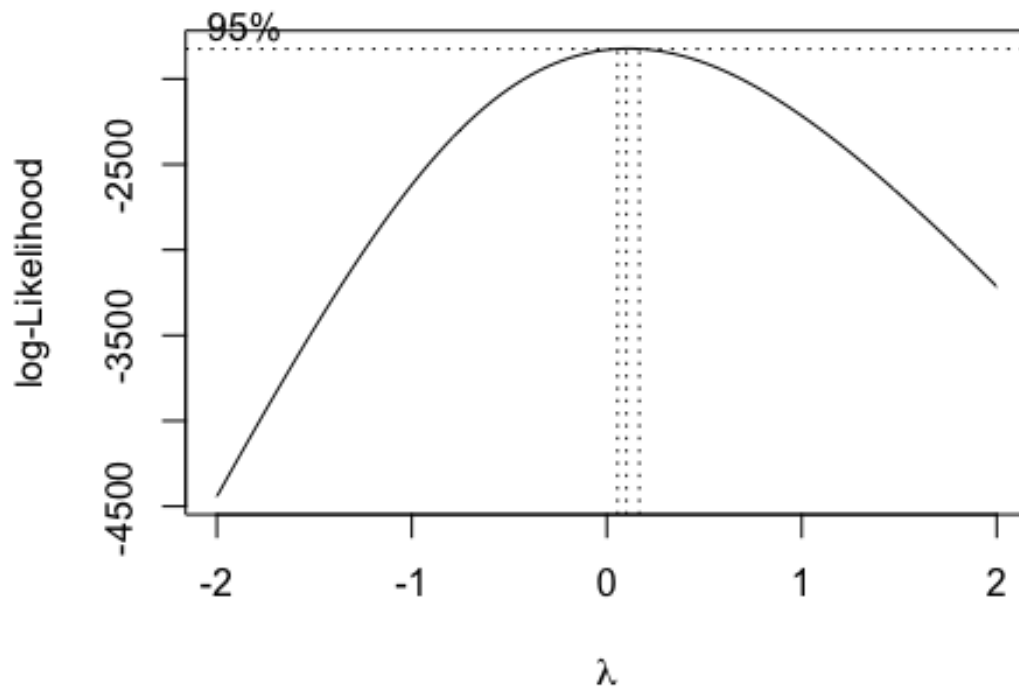


```
nutritionden <- yeo.johnson(Fd$Nutrition.Density, lambda = best_lambda)

print(G)

##      Lambda      Valor_p
## 129  0.128 0.03282063

library(MASS)
boxcox_transform <- boxcox(lm(Fd$Nutrition.Density ~ 1, data = Fd), plotit =
TRUE)
```



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

## [1] 0.1010101

Fd$Nutrition.Density_transformed <- (Fd$Nutrition.Density^lambda_optbc - 1) /
lambda_optbc
```

### Min, Max, media, mediana, cuartiles, sesgo y curtosis de nueva variable

```
min_value <- min(nutritionden, na.rm = TRUE)
min_value

## [1] 0.7384119

max_value <- max(nutritionden, na.rm = TRUE)
max_value

## [1] 11.82004

mean_value <- mean(nutritionden, na.rm = TRUE)
mean_value

## [1] 6.039103
```

```

median_value <- median(nutritionden, na.rm = TRUE)
median_value

## [1] 5.985655

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

##      25%      75%
## 4.751730 7.282633

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

## [1] 0.131189

kurtosis_value <- kurtosis(nutritionden, na.rm = TRUE)
kurtosis_value

## [1] 0.04284343

```

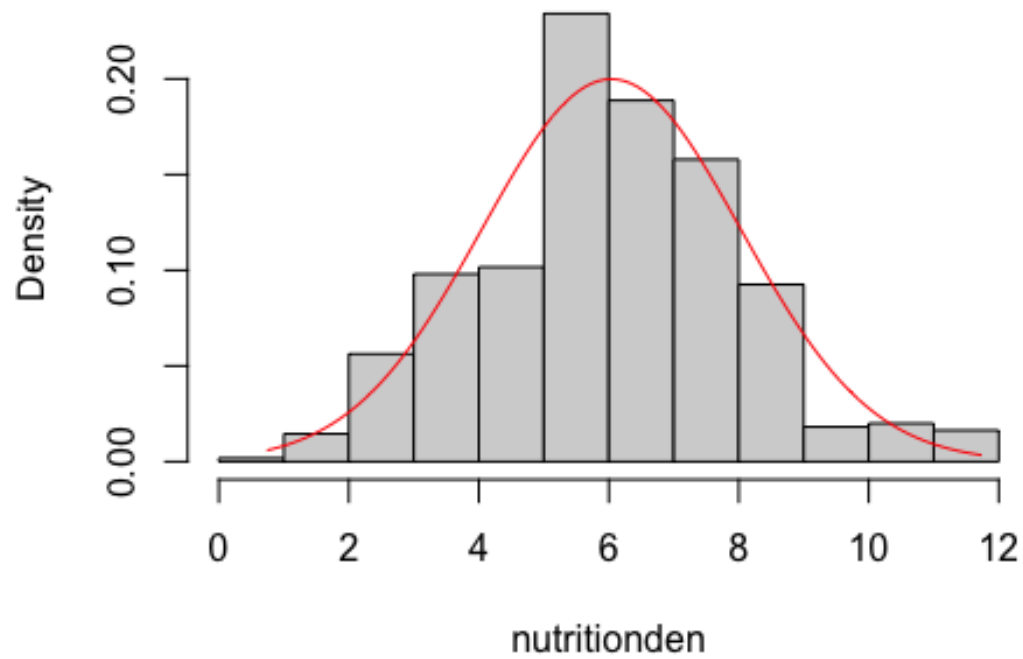
### Grafico de densidad empírica y teórica de nueva variable

```

hist(nutritionden, freq=FALSE)
x=seq(min(nutritionden),max(nutritionden),0.1)
y=dnorm(x,mean(nutritionden),sd(nutritionden))
lines(x,y,col="red")

```

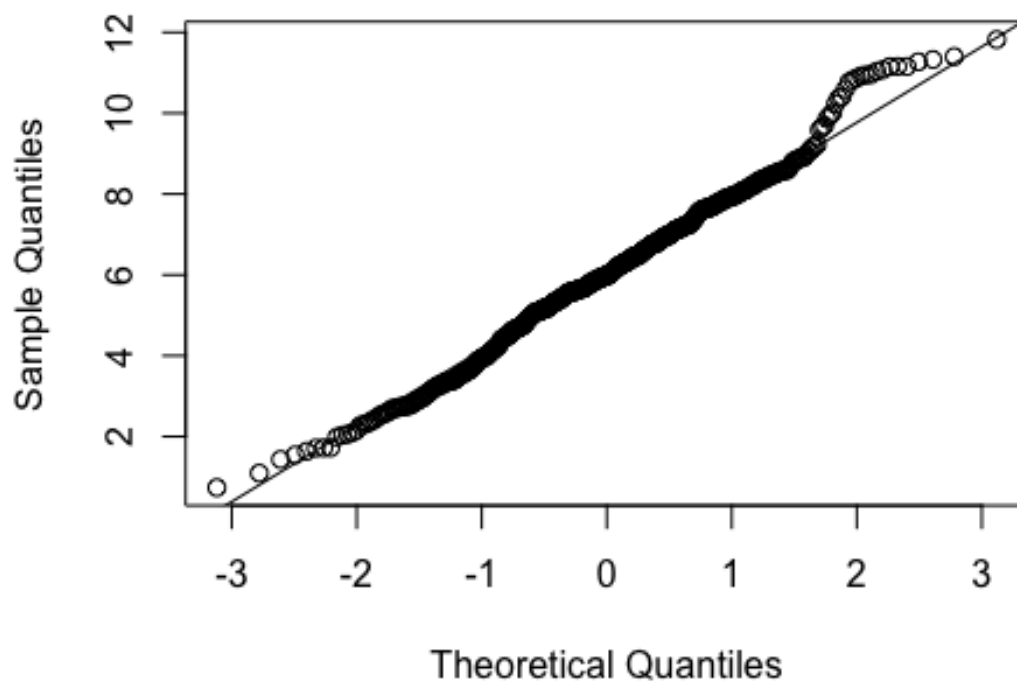
## Histogram of nutritionden



### Grafica y respectivo qqplot de la nueva variable

```
qqnorm(nutritionden)  
qqline(nutritionden)
```

## Normal Q-Q Plot



### Pruebas de Normalidad Univariada de nueva variable

```
library(moments)
jarque.test(nutritionden)

##
##  Jarque-Bera Normality Test
##
## data:  nutritionden
## JB = 1.6559, p-value = 0.4369
## alternative hypothesis: greater

library(nortest)
ad.test(nutritionden)

##
##  Anderson-Darling normality test
##
## data:  nutritionden
## A = 0.82618, p-value = 0.03282
```

## Comentarios de Resultados

Como podemos observar utilice la transformacion de Yeo-Johnson para que pudiera encontrar una mejor  $\lambda$  a nuestros datos, aunque la transformacion de BoxCox encontro una  $\lambda$  muy cerca, despues de estos cambios podemos observar un cambio muy notorio dentro de las graficas de antes y despues de las transformaciones, en el histograma de la variable transformada se puede observar un mayor equilibrio en los datos, encontrando una campana mas normalizada en vez de una Leptocurtica extrema que obtuvimos de la variable normal, no solo en esto, podemos observar que dentro de curtosis obtuvimos un dato no tan elevado como con la primera variable (15.18 vs 3.05), que nos dice que hay una distribucion mucho mas normalizada que antes, aunque dentro de las pruebas de normalidad la de Jarque-Bera nos indica que los datos podrian estar normalizados y la de Anderson-Darling nos indica que no son normales, esto podria suceder a las metricas que usan respectivas pruebas. Como conclusion es que no se pudo lograr una normalizacion total ya que se sigue teniendo una curva algo leptocurtica dentro del qqplot de la nueva variable pero se puede observar una mejoria notoria dentro de las variables.