

Ejercicio4.1

Silvia Pineda

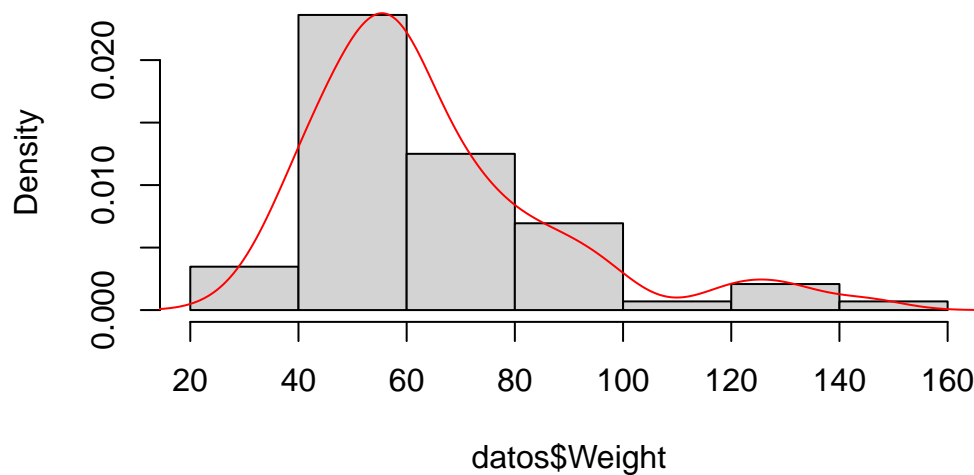
Carga de Datos y Librerías

```
datos<-read.csv("arboles.csv",stringsAsFactors = TRUE)
```

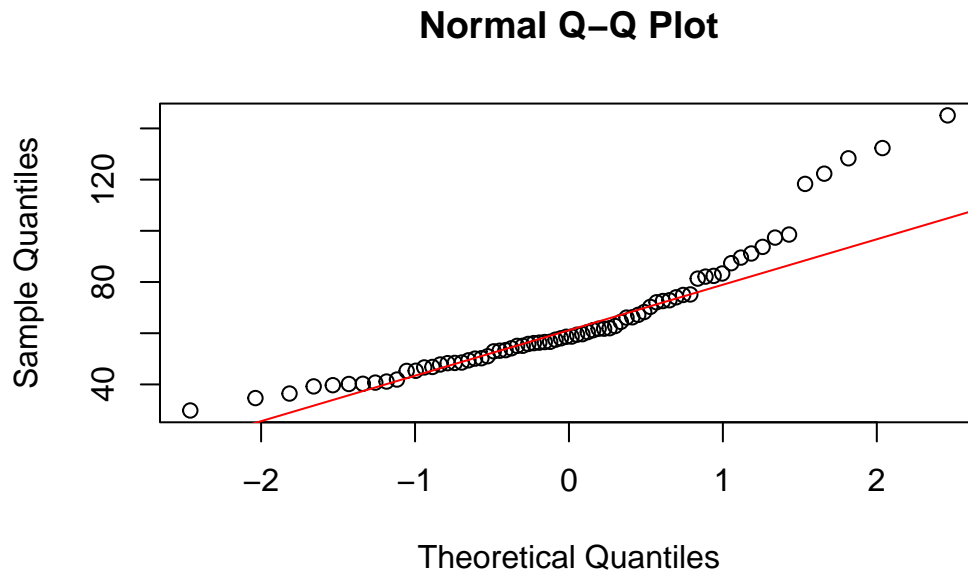
Variable Weight

```
#Realizamos un histograma  
hist(datos$Weight, prob = TRUE)  
lines(density(datos$Weight), col = "red")
```

Histogram of datos\$Weight



```
##Realizamos el qqplot
qqnorm(datos$Weight)
qqline(datos$Weight, col = "red")
```



```
##Hacemos los dos test de normalidad
shapiro.test(datos$Weight)
```

Shapiro-Wilk normality test

```
data:  datos$Weight
W = 0.87575, p-value = 3.714e-06
```

```
ks.test(datos$Weight, "pnorm", mean = mean(datos$Weight), sd = sd(datos$Weight))
```

Exact one-sample Kolmogorov-Smirnov test

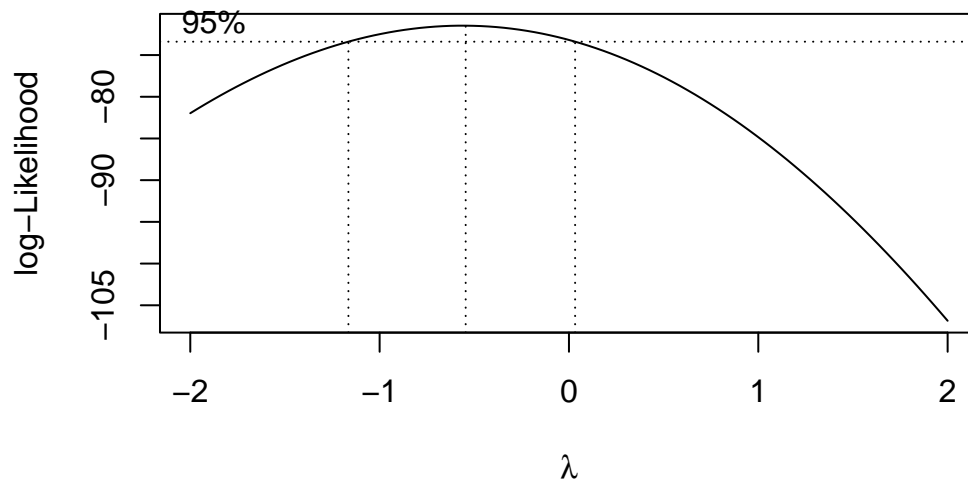
```
data:  datos$Weight
D = 0.15911, p-value = 0.0465
alternative hypothesis: two-sided
```

El histograma muestra asimetría a la derecha y el QQ-plot se separa claramente de la línea teórica.

Ambos tests rechazan la normalidad, indicando que la variable no sigue una distribución normal.

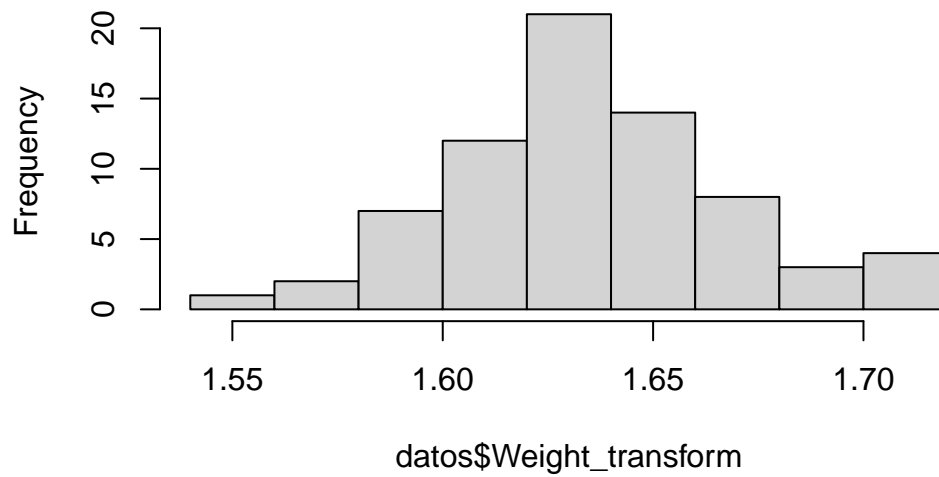
Vamos a ver que pasa transformando la variable:

```
###Vamos a transformar la variable Weight  
library(MASS)  
b<-boxcox(lm(datos$Weight ~ 1))
```



```
lambda <- b$x[which.max(b$y)] # -0.54  
  
##Transformación usando lambda exacto  
datos$Weight_transform <- (datos$Weight ^ lambda - 1) / lambda  
hist(datos$Weight_transform)
```

Histogram of datos\$Weight_transform



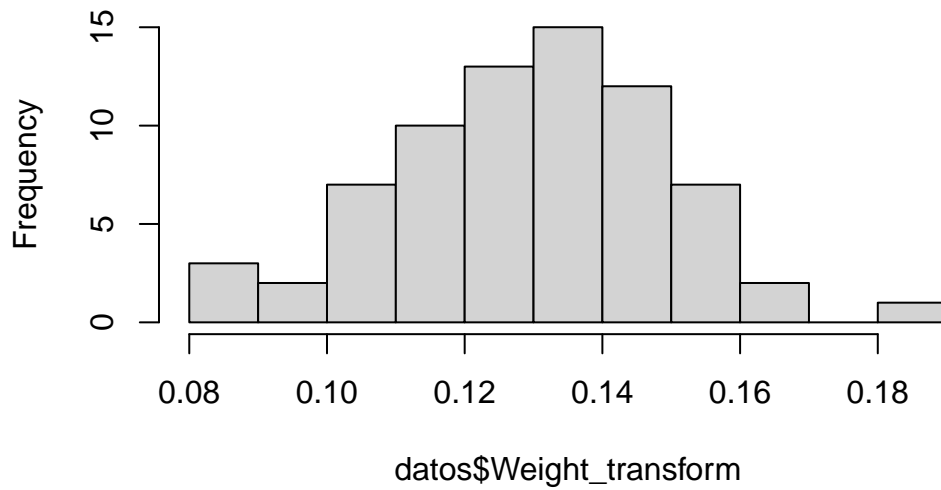
```
ks.test(datos$Weight_transform, "pnorm", mean = mean(datos$Weight_transform), sd = sd(datos$W
```

Exact one-sample Kolmogorov-Smirnov test

```
data: datos$Weight_transform
D = 0.062243, p-value = 0.9266
alternative hypothesis: two-sided
```

```
##Transformación usando 1/sqrt(y)
datos$Weight_transform <- 1/sqrt(datos$Weight)
hist(datos$Weight_transform)
```

Histogram of datos\$Weight_transform



```
ks.test(datos$Weight_transform, "pnorm", mean = mean(datos$Weight_transform), sd = sd(datos$W
```

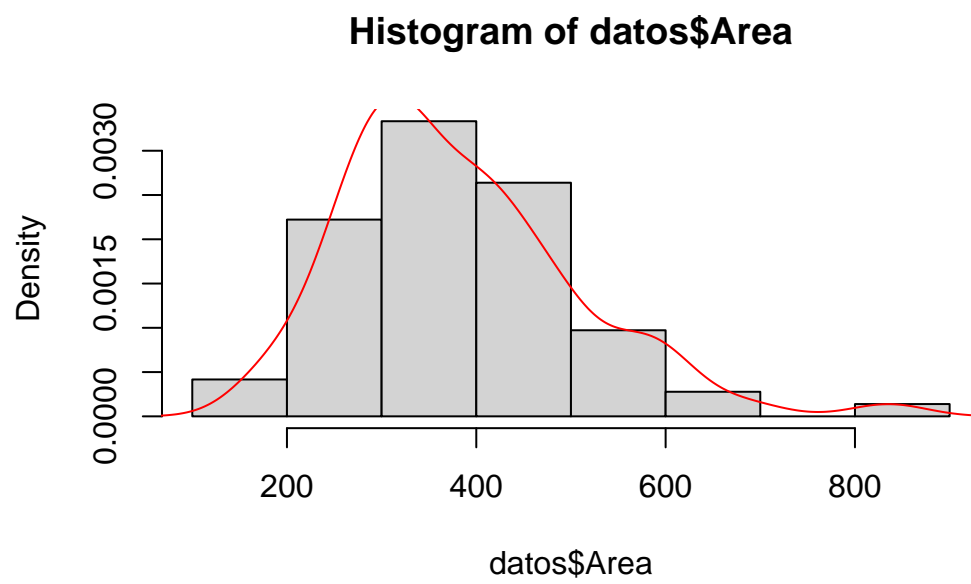
Exact one-sample Kolmogorov-Smirnov test

```
data: datos$Weight_transform
D = 0.065121, p-value = 0.9006
alternative hypothesis: two-sided
```

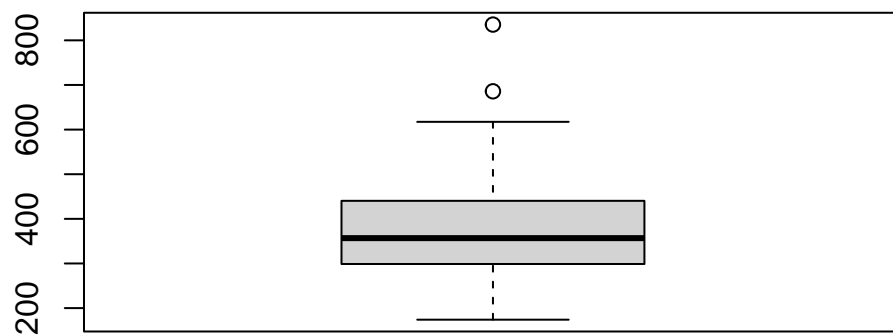
Tras aplicar la transformación Box-Cox, la distribución se vuelve mucho más simétrica y ajustada a la normal.

Variable Area

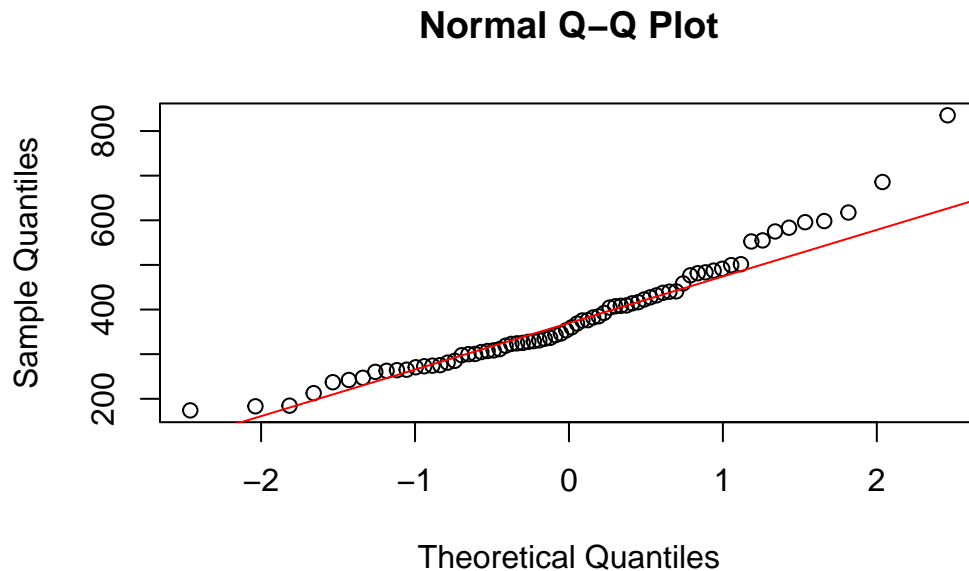
```
#Realizamos un histograma
hist(datos$Area, prob = TRUE)
lines(density(datos$Area), col = "red")
```



```
boxplot(datos$Area)
```



```
##Realizamos el qqplot
qqnorm(datos$Area)
qqline(datos$Area, col = "red")
```



```
##Hacemos los dos test de normalidad
shapiro.test(datos$Area)
```

Shapiro-Wilk normality test

```
data:  datos$Area
W = 0.94301, p-value = 0.002689
```

```
ks.test(datos$Area, "pnorm", mean = mean(datos$Area), sd = sd(datos$Area))
```

Exact one-sample Kolmogorov-Smirnov test

```
data:  datos$Area
D = 0.10072, p-value = 0.4301
alternative hypothesis: two-sided
```

El histograma muestra asimetría a la derecha y el QQ-plot se desvía de la línea teórica, lo que indica falta de normalidad.

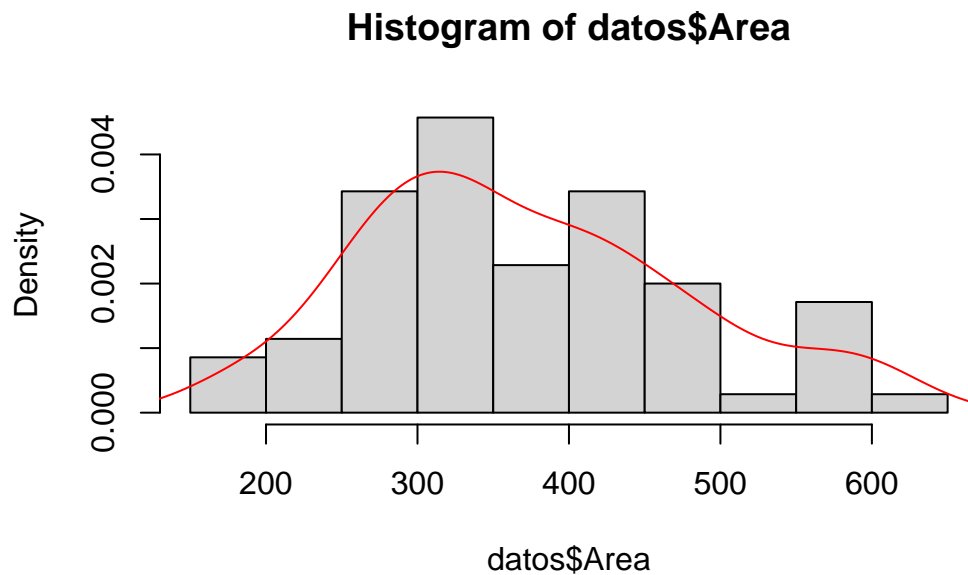
Shapiro rechaza normalidad, pero KS no lo hace; esto es habitual cuando hay outliers, porque Shapiro es más sensible.

Vamos a ver que pasa si eliminamos los outliers

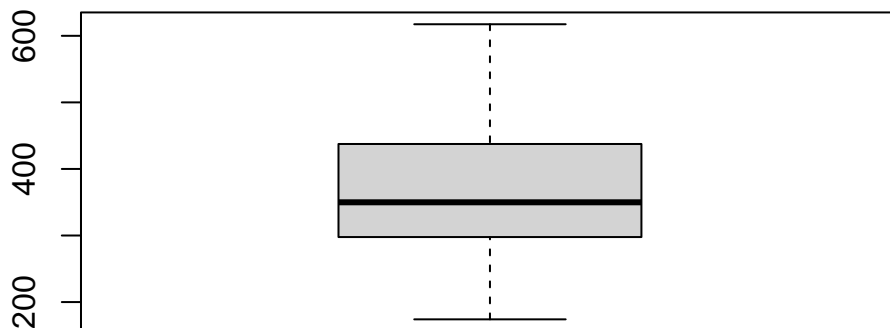
```
###Eliminar los outliers
###Los valores atípicos son:

outlier_values <- boxplot.stats(datos$Area)$out # outlier values.
out_ind <- which(datos$Area %in% c(outlier_values))
datos$Area[out_ind]<-NA

#Realizamos un histograma
hist(datos$Area, prob = TRUE)
lines(density(na.omit(datos$Area)), col = "red")
```

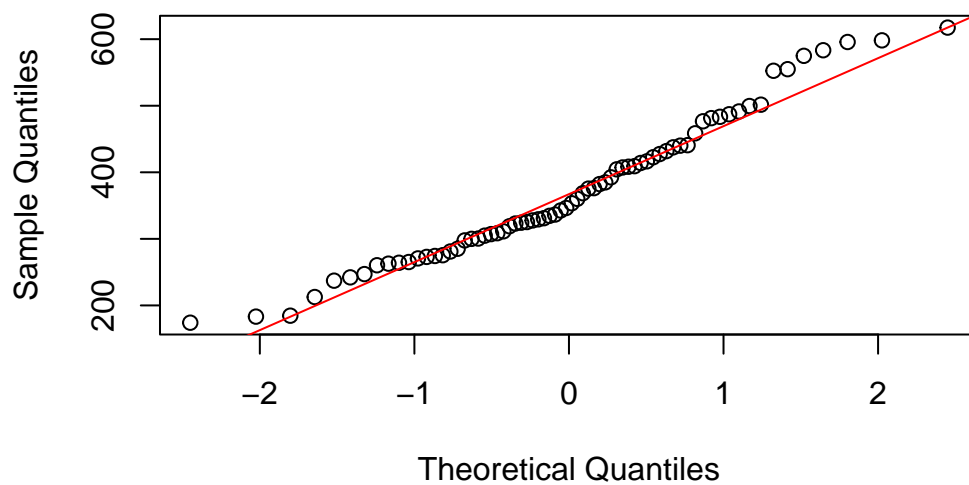


```
boxplot(datos$Area)
```

```
##Realizamos el qqplot  
qqnorm(datos$Area)  
qqline(datos$Area, col = "red")
```

Normal Q-Q Plot



```
##Hacemos los dos test de normalidad  
shapiro.test(datos$Area)
```

Shapiro-Wilk normality test

```
data:  datos$Area  
W = 0.96805, p-value = 0.07072
```

```
ks.test(datos$Area, "pnorm", mean = mean(datos$Area), sd = sd(datos$Area))
```

Exact one-sample Kolmogorov-Smirnov test

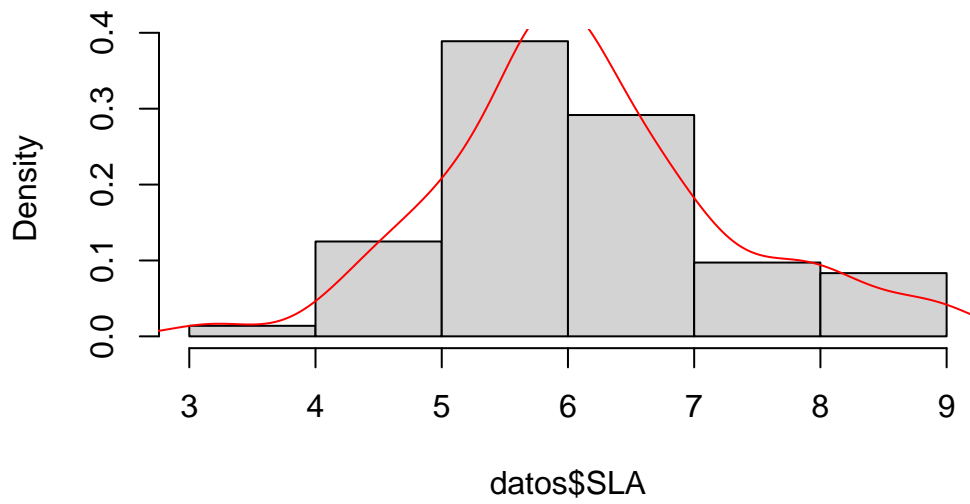
```
data:  datos$Area  
D = NA, p-value = NA  
alternative hypothesis: two-sided
```

Al eliminar los outliers, la distribución se hace más simétrica y se ajusta mejor a la normalidad. Además Shapiro ya no rechaza la normalidad.

Variable SLA

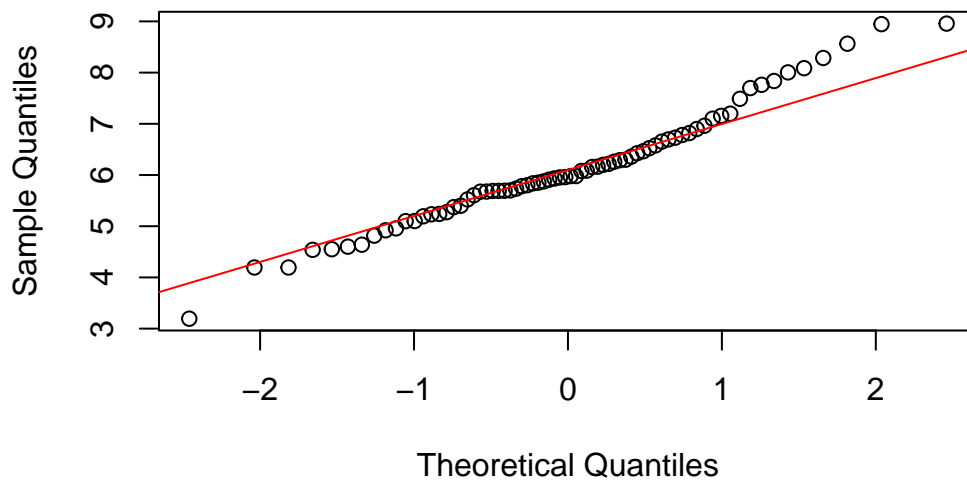
```
#Realizamos un histograma  
hist(datos$SLA, prob = TRUE)  
lines(density(datos$SLA), col = "red")
```

Histogram of datos\$SLA



```
##Realizamos el qqplot  
qqnorm(datos$SLA)  
qqline(datos$SLA, col = "red")
```

Normal Q-Q Plot



```
##Hacemos los dos test de normalidad  
shapiro.test(datos$SLA)
```

Shapiro-Wilk normality test

```
data:  datos$SLA  
W = 0.97559, p-value = 0.1744
```

```
ks.test(datos$SLA, "pnorm", mean = mean(datos$SLA), sd = sd(datos$SLA))
```

Exact one-sample Kolmogorov-Smirnov test

```
data:  datos$SLA  
D = 0.091851, p-value = 0.5471  
alternative hypothesis: two-sided
```

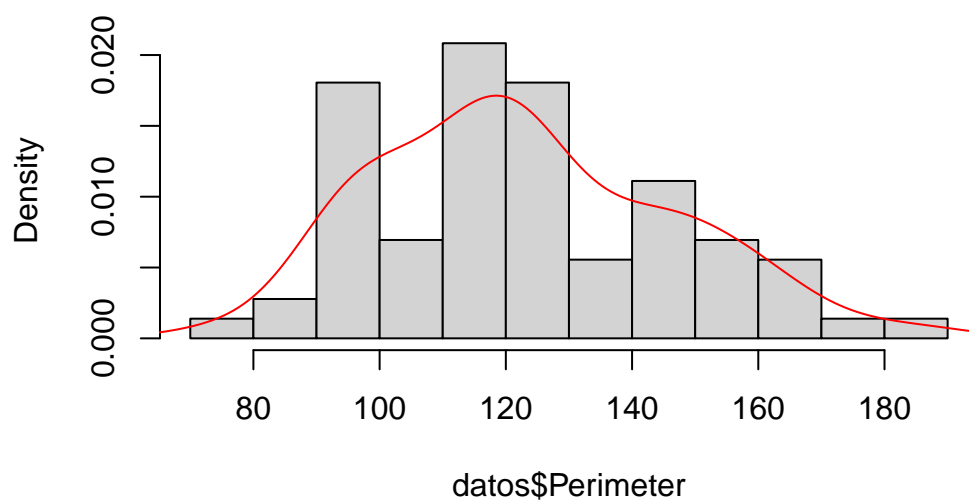
El histograma es aproximadamente simétrico y el QQ-plot sigue bien la línea teórica. Tanto Shapiro como KS no rechazan la normalidad.

Esta variable presenta un comportamiento claramente normal.

Variable Perimeter

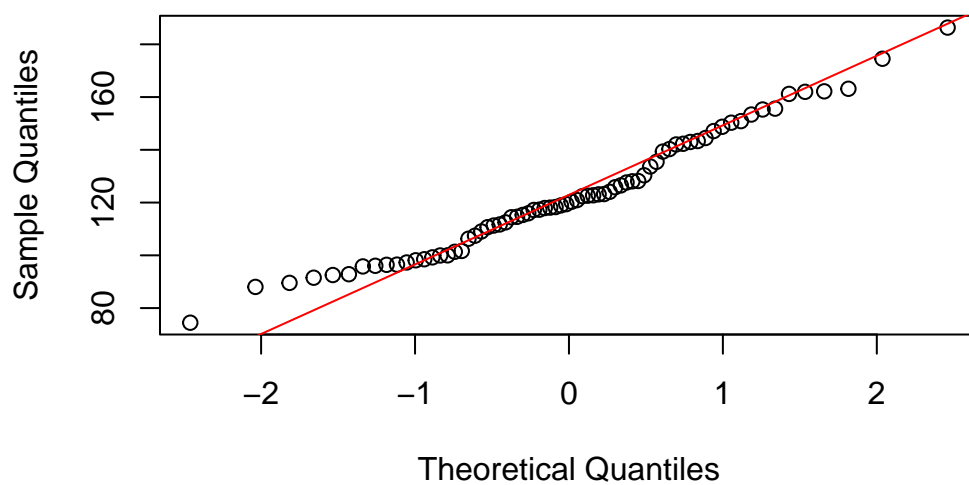
```
#Realizamos un histograma  
hist(datos$Perimeter, prob = TRUE)  
lines(density(datos$Perimeter), col = "red")
```

Histogram of datos\$Perimeter



```
##Realizamos el qqplot  
qqnorm(datos$Perimeter)  
qqline(datos$Perimeter, col = "red")
```

Normal Q-Q Plot



```
##Hacemos los dos test de normalidad  
shapiro.test(datos$Perimeter)
```

Shapiro-Wilk normality test

```
data:  datos$Perimeter  
W = 0.97434, p-value = 0.148
```

```
ks.test(datos$Perimeter, "pnorm", mean = mean(datos$Perimeter), sd = sd(datos$Perimeter))
```

Exact one-sample Kolmogorov-Smirnov test

```
data:  datos$Perimeter  
D = 0.091077, p-value = 0.5579  
alternative hypothesis: two-sided
```

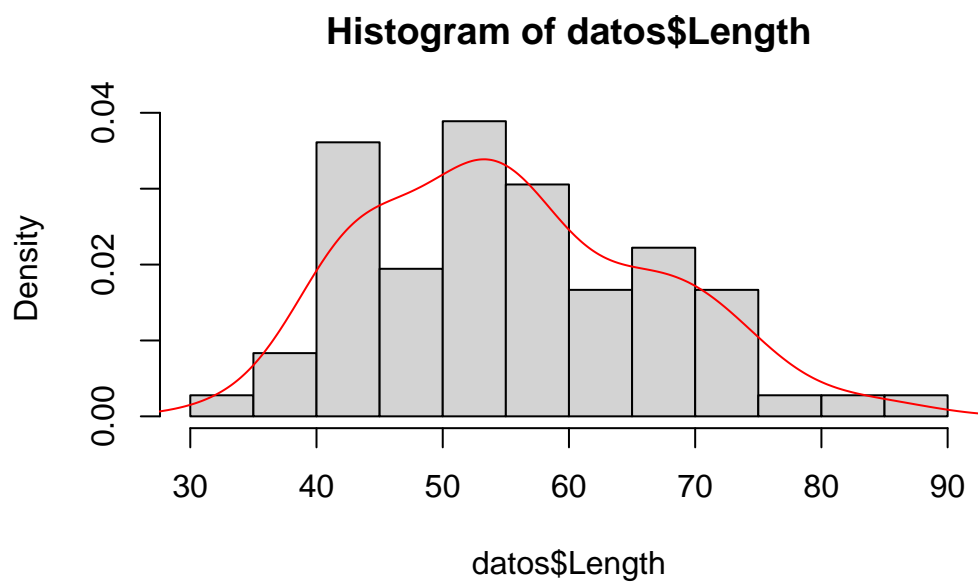
El histograma es relativamente simétrico y la densidad tiene forma de campana.

El QQ-plot ajusta bien salvo ligeras desviaciones en las colas.

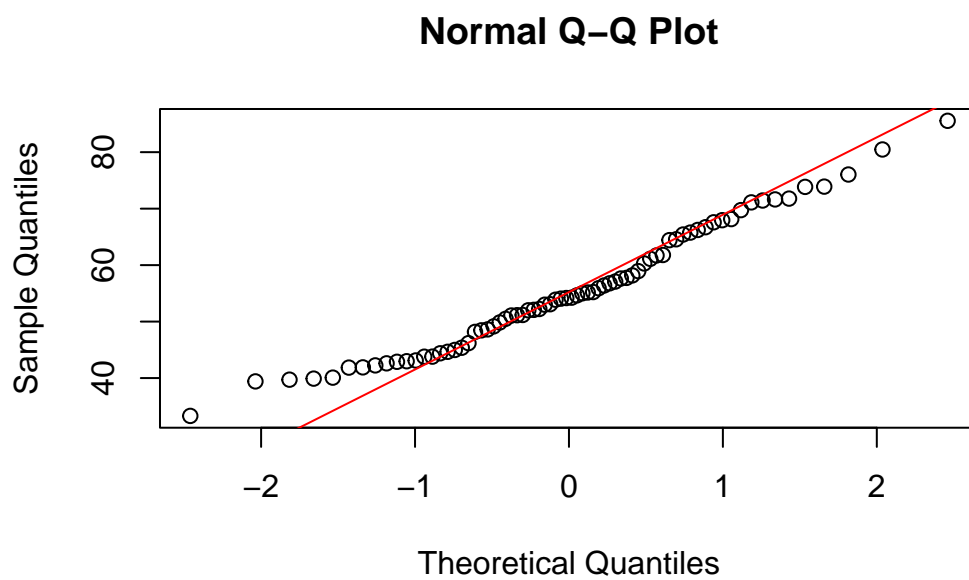
Ambos tests no rechazan la normalidad, por lo que puede considerarse normal.

Variable Length

```
#Realizamos un histograma  
hist(datos$Length, prob = TRUE)  
lines(density(datos$Length), col = "red")
```



```
##Realizamos el qqplot  
qqnorm(datos$Length)  
qqline(datos$Length, col = "red")
```



```
##Hacemos los dos test de normalidad  
shapiro.test(datos$Length)
```

Shapiro-Wilk normality test

```
data:  datos$Length  
W = 0.97244, p-value = 0.1149
```

```
ks.test(datos$Length, "pnorm", mean = mean(datos$Length), sd = sd(datos$Length))
```

Exact one-sample Kolmogorov-Smirnov test

```
data:  datos$Length  
D = 0.078993, p-value = 0.7299  
alternative hypothesis: two-sided
```

La distribución es bastante simétrica y con ligera cola a la derecha.

El QQ-plot se mantiene cerca de la línea, con desviaciones moderadas al inicio y al final.

Shapiro y KS no rechazan la normalidad, por lo que la variable puede considerarse normal.