

# Normality Distribution

Leonardo Moreno

2023-04-03

En este proyecto busco analizar la normalidad de los retornos de 5 acciones chilenas tales como Aguas Andinas, Banco Santander Chile, Entel, Falabella y Parque Arauco. Cada una de ellas opera en distintas industrias entre sí. Dentro de las herramientas a utilizar para el análisis de distribución normal existen distintas herramientas, pero para este análisis utilizaremos las siguientes: histogramas, simetría, curtosis y test de Jarque Bera.

Para este proyecto utilizare los paquetes “xts” para trabajar con series de tiempos y “readxl” para importar los precios históricos diarios de estas acciones, los cuales fueron descargados de la página “Investing.com” desde el 03-01-2000 hasta el 31-03-2023. También, utilizaré el paquete “PerformanceAnalytics” para el calculo de los retornos diatios de estos activos. Por último, utilizare el paquete “tseries” para trabajar con el test de Jarque Bera.

A continuación, procederé a importar cada una de las bases de datos de estas acciones, donde se presentan el precio de apertura, el precio más alto y el más precio más bajo, que alcanzó la acción en la jornada respectivamente, el precio de cierre, el precio de cierre ajustado y el volumen transado. Luego de esto convertiré la columna “Date” en un objeto de fecha para que sea reconocido de tal manera para luego convertir cada archivo donde reside la información de cada acción en un archivo “xts”. A modo de ejemplo visualizaré los primeros 3 datos del archivo correspondiente a la acción de Parque Arauco.

```
#Aguas Andinas
AGUASA <- read_excel("AGUASA-A.SN.xlsx")
AGUASA$Date <- as.Date(AGUASA$Date)
AGUASA <- as.xts(AGUASA)

#Bco. Santander Chile
BSAC <- read_excel("BSAC.xlsx")
BSAC$Date <- as.Date(BSAC$Date)
BSAC <- as.xts(BSAC)

#Entel
ENTEL <- read_excel("ENTEL.SN.xlsx")
ENTEL$Date <- as.Date(ENTEL$Date)
ENTEL <- as.xts(ENTEL)

#Falabella
FALABELLA <- read_excel("FALABELLA.SN.xlsx")
FALABELLA$Date <- as.Date(FALABELLA$Date)
FALABELLA <- as.xts(FALABELLA)

#Parque Arauco
PARAUCO <- read_excel("PARAUCO.SN.xlsx")
PARAUCO$Date <- as.Date(PARAUCO$Date)
```

```
PARAUCO <- as.xts(PARAUCO)
head(PARAUCO,3)
```

```
##           Open      High      Low      Close Adj Close Volume
## 2000-01-03 290783081 290783081 285854553 290783081 246489563 895196
## 2000-01-04 281911713 285854553 281911713 281911713 238969482 220755
## 2000-01-05 280926025 280926025 277968903 280926025 238133926 141725
```

Ahora procederé a filtrar con la fecha que realmente trabajaré, para esto utilizaré la función “window” del paquete “xts”. Empezaré con el 01-03-2000 hasta el 31-03-2023.

```
#Filtramos por fecha
#Desde el 30-03-2000 hasta 31-03-2023
AGUASA <- window(AGUASA, start = "2000-03-01", end ="2023-03-31")
BSAC <- window(BSAC, start = "2000-03-01", end ="2023-03-31")
ENTEL <- window(ENTEL, start = "2000-03-01", end ="2023-03-31")
FALABELLA <- window(FALABELLA, start = "2000-03-01", end ="2023-03-31")
PARAUCO <- window(PARAUCO, start = "2000-03-01", end ="2023-03-31")
```

Continuando con el filtro de datos, ahora procederé a dejar solo la columna de interés en cada una de las bases de datos de las acciones, que es “Adj. Close” y a modo de ejemplo visualizaré los 2 primeros Y 2 últimos datos de Falabella.

```
#Seleccionamos la columna que nos interesa de cada base de datos
# En esta caso es Adj Close
AGUASA <- AGUASA[, -c(1,2,3,4,6)]
BSAC <- BSAC[, -c(1,2,3,4,6)]
ENTEL <- ENTEL[, -c(1,2,3,4,6)]
FALABELLA <- FALABELLA[, -c(1,2,3,4,6)]
PARAUCO <- PARAUCO[, -c(1,2,3,4,6)]
head(FALABELLA,2)
```

```
##           Adj Close
## 2000-03-01 535097473
## 2000-03-02 530711426
```

```
tail(FALABELLA,2)
```

```
##           Adj Close
## 2023-03-30 1.838e+09
## 2023-03-31 1.828e+09
```

Una vez teniendo la columna de interés procederé a calcular los retornos logarítmicos diarios de cada una de las acciones. Como es de saber, el primer retorno no existe, es decir, el retorno perteneciente al primer día de la base de datos no se puede calcular ya que no existe registro del día anterior, por lo que eliminaré este dato ya que se registra como NA. Visualizaré el resultado obtenido con los 2 primeros y los 2 últimos datos de Entel.

```
#calculamos los retornos logarítmicos diarios de cada una de las acciones
RAGUASA <- Return.calculate(AGUASA, method="log")
colnames(RAGUASA)[1] <- "Return"
RAGUASA <- RAGUASA[-1,]
```

```
RBSAC <- Return.calculate(BSAC, method="log")
colnames(RBSAC)[1] <- "Returnr"
RBSAC <- RBSAC[-1,]
```

```
RENTEL <- Return.calculate(ENTEL, method = "log")
colnames(RENTEL)[1] <- "Return"
RENTEL <- RENTEL[-1,]
head(RENTEL,2)
```

```
##              Return
## 2000-03-02 0.00000000
## 2000-03-03 0.01783502
```

```
tail(RENTEL,2)
```

```
##              Return
## 2023-03-30 0.005888662
## 2023-03-31 -0.026441578
```

```
RFALABELLA <- Return.calculate(FALABELLA, method = "log")
colnames(RFALABELLA)[1] <- "Return"
RFALABELLA <- RFALABELLA[-1,]
```

```
RPARAUCO <- Return.calculate(PARAUCO, method = "log")
colnames(RPARAUCO)[1] <- "Return"
PARAUCO <- RPARAUCO[-1,]
```

Una vez que ya obtuve los retornos diarios, procedere a juntar cada una de las bases de datos, para luego calcular el retorno mensual de cada una de estas 5 acciones.

```
#Retornos mensuales
```

```
Retornos <- merge(RAGUASA, RBSAC, RENTEL, RFALABELLA, RPARAUCO)
colnames(Retornos) <- c("AguasA", "Santander", "Entel", "Falabella", "Parauco")
Retornos <- Retornos[-1,]
Retornos <- to.monthly(Retornos, OHLC = FALSE, indexAt = "lastof")
```

```
## Warning in to.period(x, "months", indexAt = indexAt, name = name, ...): missing
## values removed from data
```

```
head(Retornos,4)
```

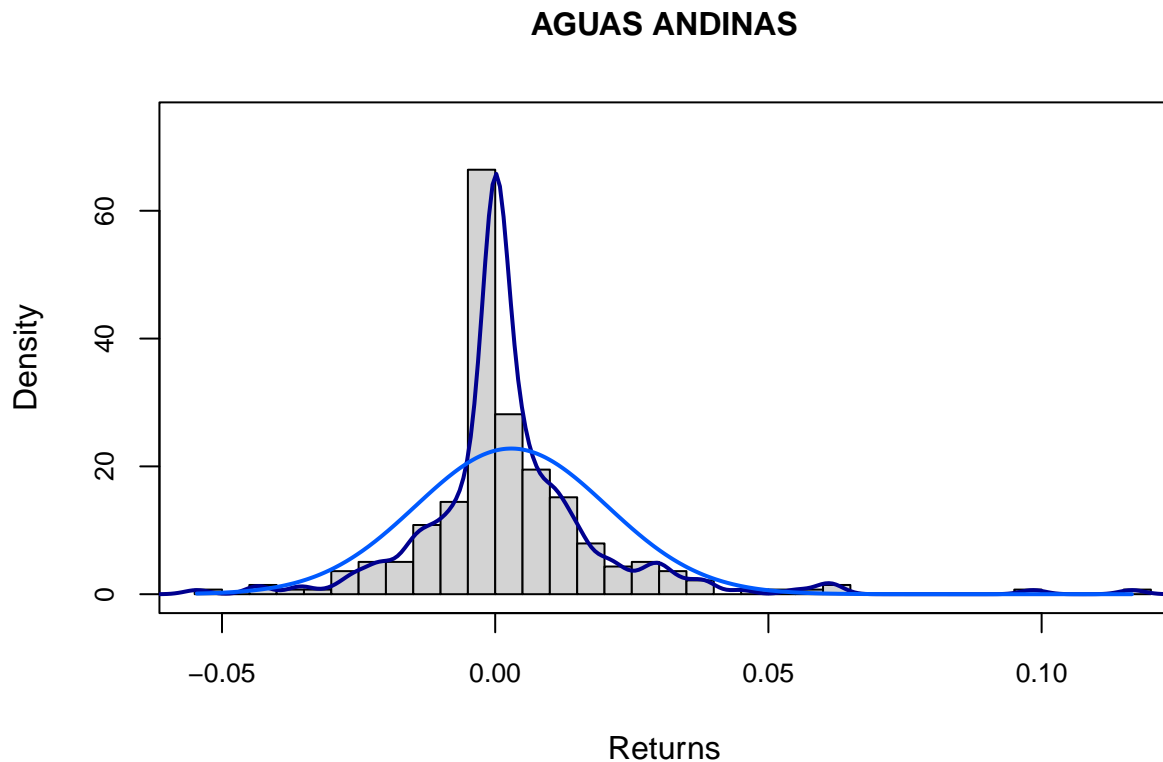
```
##      AguasA    Santander      Entel    Falabella    Parauco
## 2000-03-31    0 0.012739457 -0.022075719 0.026433399 -0.01739190
## 2000-04-30    0 -0.003159567 -0.034662613 0.009132171 -0.01801866
## 2000-05-31    0 0.003478272 -0.010581831 0.000000000 0.03509118
## 2000-06-30    0 -0.003490413 0.004175284 -0.009661784 0.01834917
```

Una vez obtenidos los retornos mensuales, procederemos al análisis de normalidad de estos partiendo por el análisis con histograma, donde la distribución normal estándar estará representada por una curva de color azul claro, y la distribución de los retornos de los activos estará representada por una curva de color azul oscuro.

## #1. HISTOGRAMA

### #1.1 Histograma Aguas A

```
chart.Histogram(Retornos$AguasA, breaks = "FD",  
               main = "AGUAS ANDINAS", methods = c("add.density", "add.normal"), border.col = "black",  
               element.color = "black",  
               note.color = "black")
```

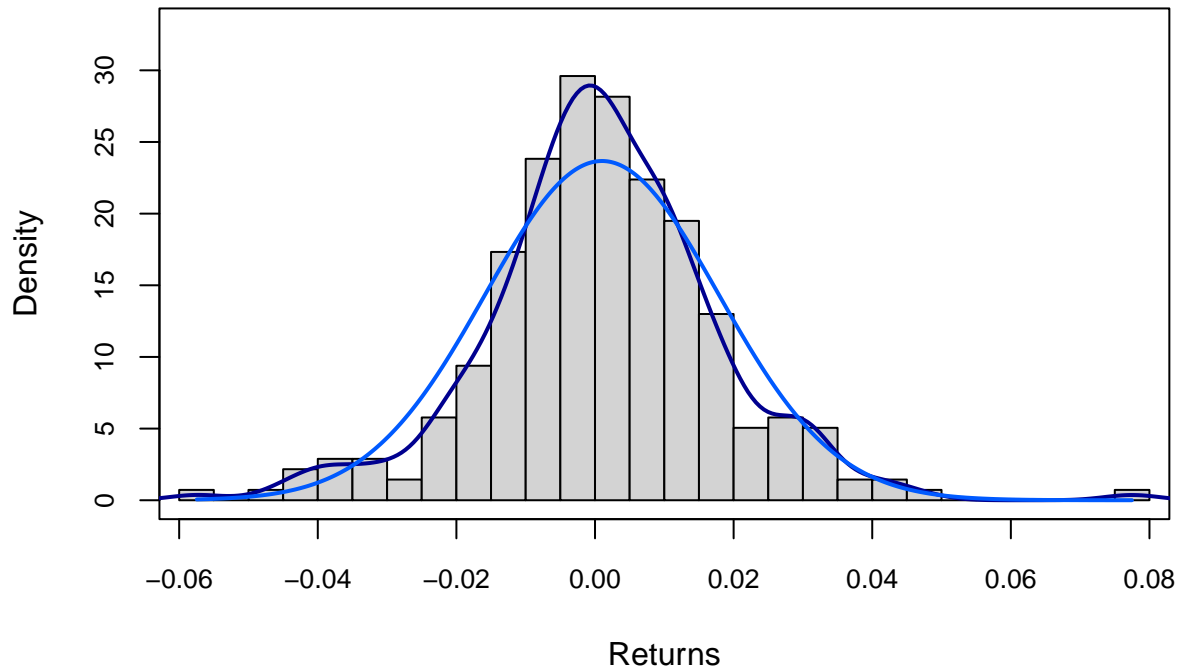


De lo anterior, se puede observar claramente que la distribución de los retornos de Aguas Andinas no es normal, estadísticamente hablando, esto se explica a que los datos se encuentran muy agrupados alrededor de la media, lo que se traduce en una dispersión de los datos muy baja, es decir, posee una desviación estándar muy baja. Además posee pesos en sus colas, evidentemente más de su cola derecha, lo que se traduce en que posee valores extremos positivos.

### #1.2 Histograma BANCO SANTANDER CL

```
chart.Histogram(Retornos$Santander, breaks = "FD",  
               main = "BANCO SANTANDER", methods = c("add.density", "add.normal"), border.col = "black",  
               element.color = "black",  
               note.color = "black")
```

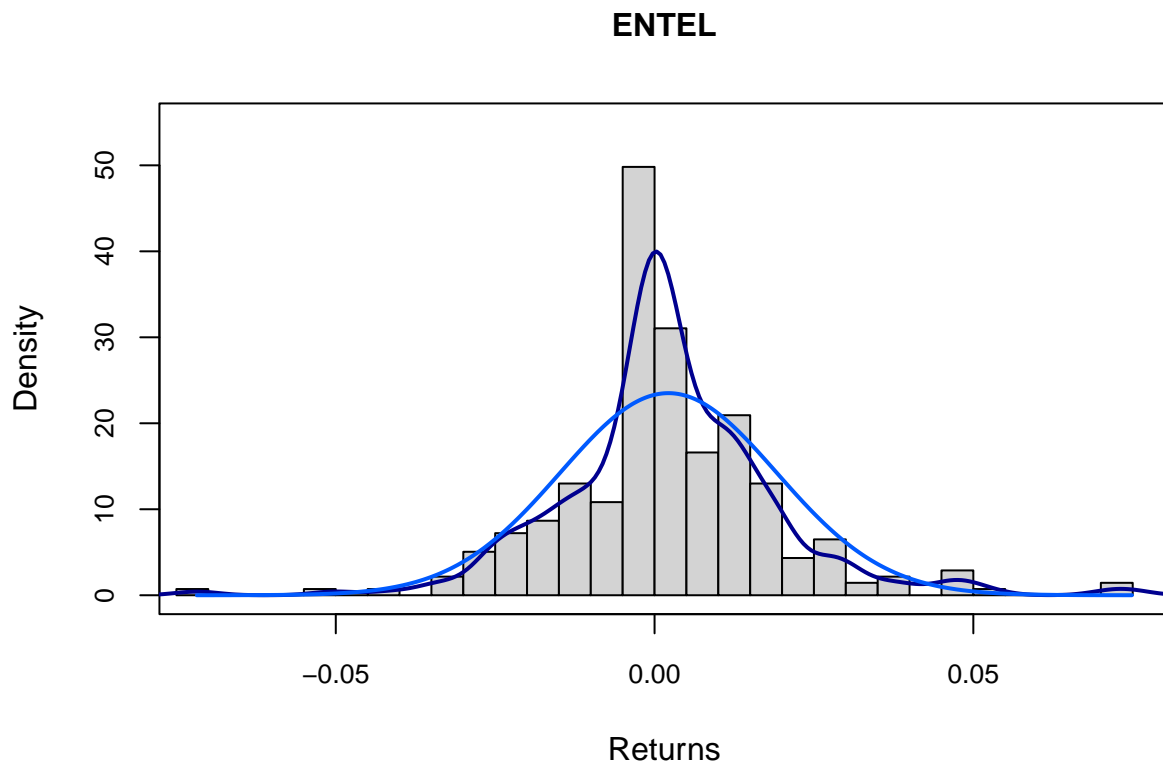
## BANCO SANTANDER



Al contrario de Aguas Andinas, Banco Santander Chile presenta una distribución de sus retornos que se asimila bastante a una normal. No posee una gran elevación, por lo que sus datos no se concentran en gran porcentaje alrededor de la media. También se puede ver hay gran simetría tanto en el lado izquierdo como el lado derecho pero, si posee peso en sus colas.

### #1.3 Histograma ENTEL

```
chart.Histogram(Retornos$Entel, breaks = "FD",  
  main = "ENTEL", methods = c("add.density", "add.normal"), border.col = "black",  
  element.color = "black",  
  note.color = "black")
```

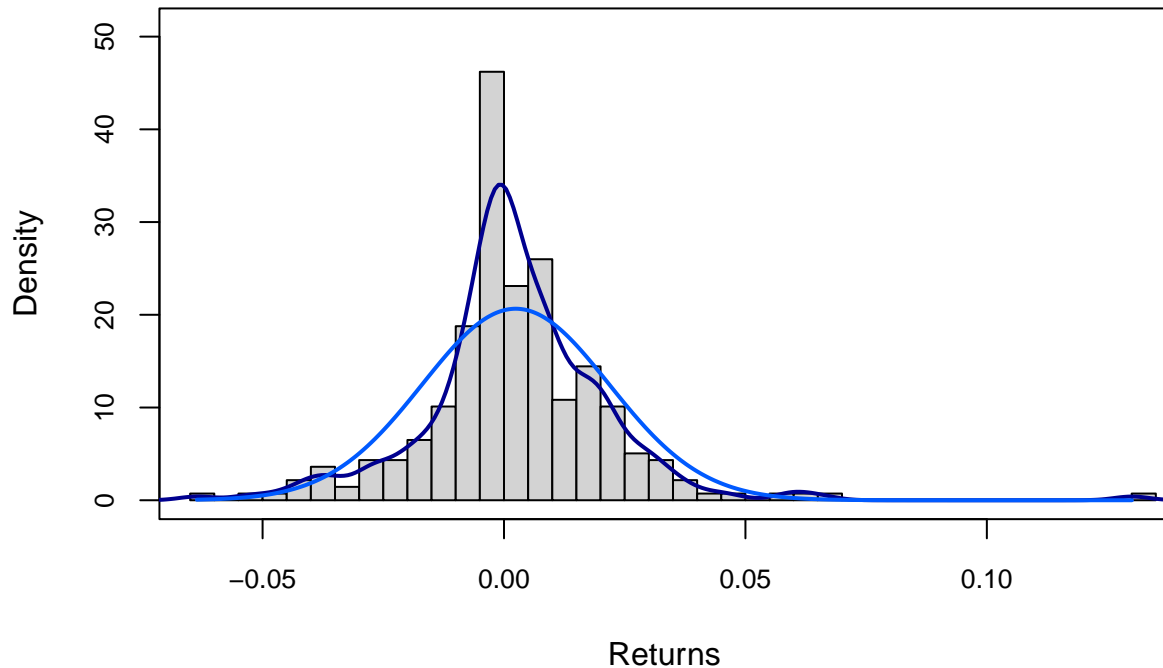


Entel por su parte, no representa una distribución normal estandar ya que presenta una leve elevación de su cumbre, es decir, los datos se acumulan alrededor de su media. Además posee peso en sus colas, lo que se traduce en valores extremos y también tiene colas largas.

#### #1.4 Histograma FALABELLA

```
chart.Histogram(Retornos$Falabella, breaks = "FD",
  main = "FALABELLA", methods = c("add.density", "add.normal"), border.col = "black",
  element.color = "black",
  note.color = "black")
```

## FALABELLA

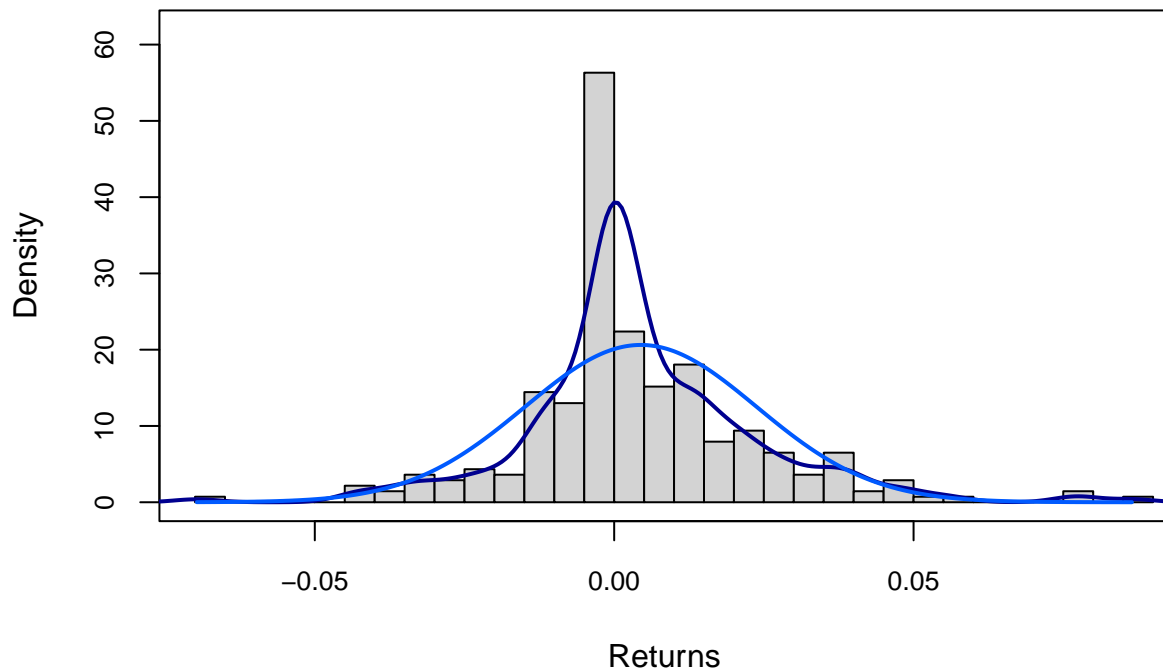


Con respecto a Falabella lo que podemos decir es que analizando su histograma no representa lo que es una distribución normal estándar en sus retornos debido a que hay concentración de datos alrededor de su media, por eso su elevación. Además, no posee gran asimetría en sus datos. Por último, también posee peso de sus colas.

### #1.5 Histograma PARQUE ARAUCO

```
chart.Histogram(Retornos$Parauco, breaks = "FD",  
  main = "PARQUE ARAUCO", methods = c("add.density", "add.normal"), border.col = "black",  
  element.color = "black",  
  note.color = "black")
```

## PARQUE ARAUCO



Por último, analizando el histograma de los retornos de Parque Arauco podemos concluir que tampoco posee una distribución normal en sus retornos ya que posee una concentración de datos alrededor de su media, posee una clara asimetría y peso en sus colas.

A continuación, utilizaré las otras herramientas para evaluar la normalidad de los retornos de nuestras acciones, estas serán la asimetría, la curtosis y el Test de Jarque Bera. La asimetría mide la simetría, una distribución normal es simétrica, es decir, este valor es 0. La curtosis es una medida de colas pesadas, es decir, la tendencia de la distribución a generar valores extremos o atípicos, en una distribución normal, la curtosis es igual a 3.

```
#Asimetria Y kURTOSIS
#AGUASA
SKAGUASA <- skewness(Retornos$AguasA)
SKAGUASA
```

```
## [1] 1.847527
```

```
KAGUASA <- kurtosis(Retornos$AguasA)
KAGUASA
```

```
## [1] 10.16648
```

Como podemos ver de los resultados anteriores, la asimetría nos entrega un valor cercano a 0, por lo que se podría decir que los datos no son particularmente asimétricos, sin embargo, la curtosis es considerablemente mayor que 3 por lo que se concluye que tiene peso en sus colas, es decir, estamos en presencia de valores extremos, tal como se dijo con el análisis de histograma.



```
#BSCA
SKBSCA <- skewness(Retornos$Santander)
SKBSCA
```

```
## [1] 0.0977506
```

```
KBSCA <- kurtosis(Retornos$Santander)
KBSCA
```

```
## [1] 1.845219
```

Banco Santander por su parte, posee un valor en su asimetría que es mucho más cercano a 0 de lo que es Aguas Andinas. Esto nos dice que sus datos no están concentrados alrededor de su media, esto puede significar que hay una alta dispersión de los datos. En cambio, el valor de la curtosis es bastante más alto de 3, por lo que también posee valores atípicos.

```
#ENTEL
SKENTEL <- skewness(Retornos$Entel)
SKENTEL
```

```
## [1] 0.3482389
```

```
KENTEL <- kurtosis(Retornos$Entel)
KENTEL
```

```
## [1] 3.315688
```

Entel presenta una asimetría de 1.3 aproximadamente, lo que nos indica que es una distribución asimétrica y además, posee una alta curtosis, por lo que posee valores extremos. Lo anterior coincide con el análisis realizado con su histograma.

```
#FALABELLA
SKFALABELLA <- skewness(Retornos$Falabella)
SKFALABELLA
```

```
## [1] 0.9270385
```

```
KFALABELLA <- kurtosis(Retornos$Falabella)
KFALABELLA
```

```
## [1] 7.375727
```

Por su parte, Falabella, presenta una asimetría de -0.03 aproximadamente, que es un valor muy cercano a 0 por lo que podemos decir que sus datos no son asimétricos. En cambio, a pesar de que su curtosis es la menor de los 5 activos, está muy por encima del valor de una normal, por lo que presenta peso en sus colas.

```
#PARAUCO
SKPARAUCO <- skewness(Retornos$Parauco)
SKPARAUCO
```

```
## [1] 0.5723459
```

```
KPARAUCO <- kurtosis(Retornos$Parauco)
KPARAUCO
```

```
## [1] 2.787589
```

Por último Parque Arauco, presenta una asimetría negativa de -0,4, que es un valor cercano a 0, por lo que se puede decir que por lo general sus datos no presentan asimetría. Por otro lado, su curtosis presenta un gran valor de 14,33 aproximadamente, lo que es un valor muy por encima de 3. Por lo que quiere decir que tiene peso en sus colas.

A continuación analizaremos la última herramienta para evaluar la normalidad de los retornos de nuestros 5 activos. Esta herramienta es el Test de Jarque Bera. Este test se basa en un estadístico de prueba que compara simultáneamente la asimetría y curtosis de los datos con sus valores para una distribución normal, entonces, la prueba puede detectar desviaciones de la normalidad causadas tanto por una o por otra, y ambas. El estadístico de prueba T se compara con una distribución chi-cuadrado con 2 grados de libertad.

```
#Test de Jarque Bera
```

```
apply(Retornos, 2, jarque.bera.test)
```

```
## $AguasA
##
##   Jarque Bera Test
##
## data:  newX[, i]
## X-squared = 1350.5, df = 2, p-value < 2.2e-16
##
##
## $Santander
##
##   Jarque Bera Test
##
## data:  newX[, i]
## X-squared = 39.739, df = 2, p-value = 2.349e-09
##
##
## $Entel
##
##   Jarque Bera Test
##
## data:  newX[, i]
## X-squared = 132.49, df = 2, p-value < 2.2e-16
##
##
## $Falabella
##
##   Jarque Bera Test
##
## data:  newX[, i]
## X-squared = 667.56, df = 2, p-value < 2.2e-16
##
##
```

```
## $Parauco
##
##  Jarque Bera Test
##
## data:  newX[, i]
## X-squared = 104.81, df = 2, p-value < 2.2e-16
```

Como se puede ver del resultado anterior, el valor p-value es menor a 0.05 por lo que