

Ejercicio Práctico Inferencia Estadística

Autor: Ignacio Sánchez Andreu

Índice

1. Definición de variables y justificación del análisis
 1. Definición Variables
 2. Interés de la comparación
 3. Fuente de los datos y obtención
2. Análisis Coca-Cola
 1. Visualización general de los datos
 2. Ajustes a modelos de distribución
 1. Normal
 2. Lognormal
 3. Gamma
 4. T de Student
 5. Modelo más adecuado
3. Análisis IBM
 1. Visualización general de los datos
 2. Ajustes a modelos de distribución
 1. Normal
 2. Lognormal
 3. Gamma
 4. T de Student
 5. Modelo más adecuado
4. Contraste t de Student
 1. Muestras independientes o pareadas
 2. Planteamiento del contraste

1. Definición de variables y justificación del análisis

1.1 Definición Variables

Vamos a comparar los retornos semanales de dos activos financieros cotizados en bolsa: Coca-Cola Company (ticker: KO) y International Business Machines (ticker: IBM)

En el contexto financiero no se estudian directamente las cotizaciones, sino los retornos semanales, que miden la variación relativa del precio de una semana a la siguiente.

Sea X_t el precio de cierre de la acción en el día t . El retorno semanal se define como: $R_t = (X_t - X_{t-5}) / X_{t-5}$ donde el subíndice hace referencia al día al que corresponde la cotización y se considera una semana bursátil de cinco sesiones.

De este modo, para cada empresa se define la variable aleatoria:

R_{KO_t} = Retorno semanal de Coca-Cola en la semana t .

R_{IBM_t} = Retorno semanal de IBM en la semana t.

También añadir que vamos a considerar un nivel de significación $\alpha=0.01$

1.2 Interés de la comparación

El interés de la comparación radica en analizar si los retornos semanales de dos empresas pertenecientes a sectores económicos distintos presentan comportamientos diferentes desde un punto de vista estadístico.

Coca-Cola es una empresa del sector de consumo básico, tradicionalmente considerada más estable, mientras que IBM pertenece al sector tecnológico y de servicios informáticos, normalmente más expuesto a cambios en el ciclo económico. El análisis comparado de sus retornos permite estudiar posibles diferencias en términos de rentabilidad media, variabilidad y forma de la distribución, así como plantear contrastes de hipótesis sobre estas magnitudes.

1.3 Fuente de los datos y obtención

Las muestras se han obtenido a partir de las cotizaciones históricas de Coca-Cola e IBM descargadas de Yahoo Finance (como en el ejercicio 6 de la práctica 1), en el periodo comprendido entre enero de 2023 y julio de 2025.

Los datos corresponden a los precios de cierre, a partir de los cuales se han calculado los retornos semanales siguiendo la definición anterior. La descarga y el tratamiento de los datos se ha realizado utilizando el paquete quantmod de R.

A partir de estas series temporales se obtienen dos muestras de tamaño suficientemente grande de los retornos semanales de ambas compañías, que serán objeto del análisis descriptivo e inferencial posterior.

2. Análisis Coca-Cola

Vamos a comenzar con la obtención de los datos. Primero declaramos que usamos las librerías "quantmod" y "xts" para importarlas al espacio de trabajo.

```
> library(quantmod)
> library(xts)
```

A continuación definimos el intervalo de fechas que vamos a estudiar, obtenemos los símbolos (como se ha mencionado al inicio, Coca-Cola tiene ticker KO), y extraemos los precios de cierre a un vector.

```
> start_date <- as.Date("2023-01-01")
> end_date   <- as.Date("2025-07-01")

> getSymbols("KO", src = "yahoo", from = start_date, to = end_date, auto.assign =
TRUE)

> KO_close <- Cl(KO)    # xts con una columna
> CocaCola <- as.numeric(KO_close) # Dejar los datos en un único vector
```

Como los datos están ordenados de manera creciente en el tiempo, usamos la misma función para obtener los retornos semanales que hay en la práctica 1 pero modificada para que esté bien para datos crecientes (sólo hay que cambiar los índices del vector):

```
> weekly.return <- function(X) {  
  ret <- NULL  
  n <- length(X) - 5  
  for (k in 1:n) {  
    ret <- c(ret, (X[k+5] - X[k]) / X[k])  
  }  
  ret  
}
```

Y así obtenemos los retornos semanales a partir de la cotización:

```
> retornos_K0 <- weekly.return(CocaCola)
```

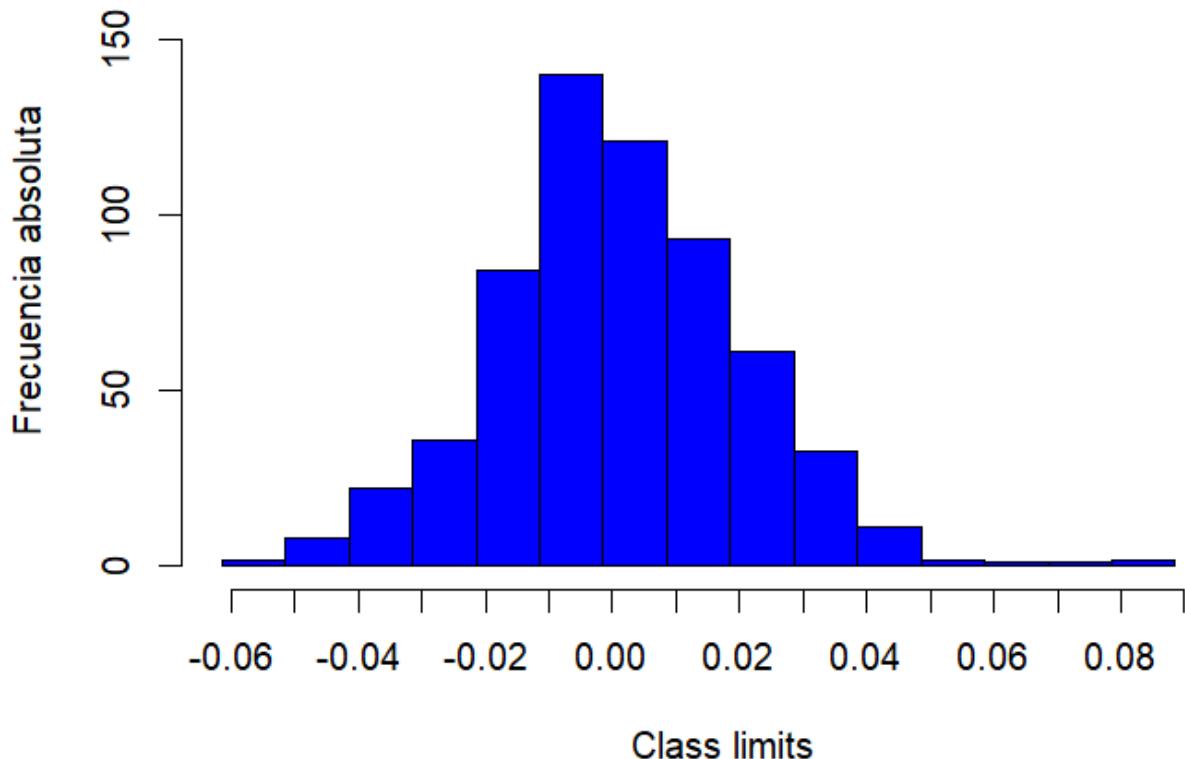
2.1 Visualización general de los datos

Ahora vamos a hacer una primera visualización de los datos usando el paquete fdth.

```
> tabla_intervalos_semanal_CocaCola <- fdth(retornos_K0, start = min(retornos_K0),  
end = max(retornos_K0), h=0.01)  
  
> tabla_intervalos_semanal_CocaCola  
  Class limits   f   rf rf(%)  cf cf(%)  
[-0.0615,-0.0515)  2  0.00  0.32   2  0.32  
[-0.0515,-0.0415)  8  0.01  1.29  10  1.62  
[-0.0415,-0.0315) 22  0.04  3.55  32  5.17  
[-0.0315,-0.0215) 36  0.06  5.82  68 10.99  
[-0.0215,-0.0115) 84  0.14 13.57 152 24.56  
[-0.0115,-0.00146) 140 0.23 22.62 292 47.17  
[-0.00146,0.00854) 121 0.20 19.55 413 66.72  
[0.00854,0.0185)  93 0.15 15.02 506 81.74  
[0.0185,0.0285)  61 0.10  9.85 567 91.60  
[0.0285,0.0385)  33 0.05  5.33 600 96.93  
[0.0385,0.0485)  11 0.02  1.78 611 98.71  
[0.0485,0.0585)  2 0.00  0.32 613 99.03  
[0.0585,0.0685)  1 0.00  0.16 614 99.19  
[0.0685,0.0785)  1 0.00  0.16 615 99.35  
[0.0785,0.0885)  2 0.00  0.32 617 99.68
```

Lo vemos en un histograma para que sea más visual.

Histograma retornos semanales Coca-Cola

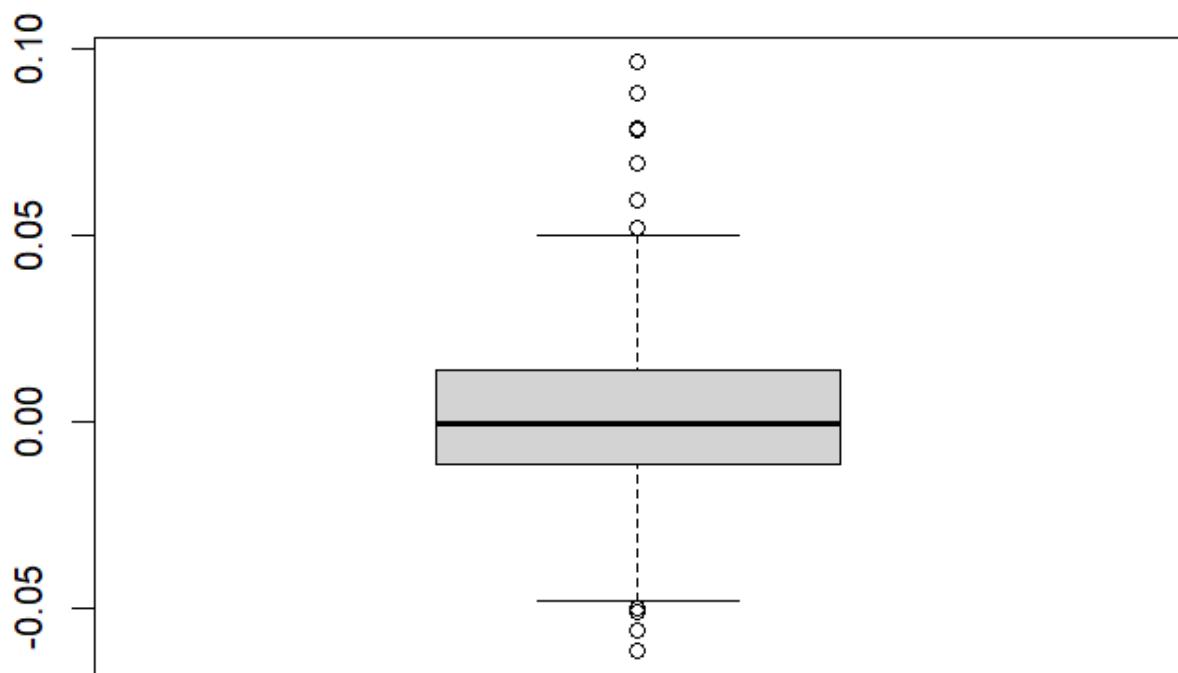


Así a primera vista parece que la mayoría de los datos se concentran alrededor del 0. Usamos ahora un comando para obtener algunas medidas descriptivas significativas como la media, la mediana y algunos cuantiles.

```
> summary(retornos_K0)
  Min.    1st Qu.     Median      Mean      3rd Qu.      Max.
-0.0614585 -0.0110607 -0.0001587  0.0010921  0.0138901  0.0969066
```

A pesar de que la mediana es menor que 0, la media es mayor que 0, por lo que aunque haya más semanas en la que los retornos sean negativas, en el cómputo global parece que hay una subida en la cotización, ya que aunque haya menos semanas en las que el retorno sea positivo, cuando sí lo es el retorno debe ser mayor (por ser la media mayor que 0).

Si graficamos el diagrama de caja y bigotes vemos:



Lo que confirma en parte lo que hemos mencionado antes, las semanas de peor retorno cae menos que las semanas de mayor retorno. En concreto, los datos que están fuera del diagrama de caja y bigotes (los outliers) son:

```
boxplot(retornos_K0)$out
[1] -0.06145852 -0.05073241  0.05195695 -0.04997167 -0.05601151  0.07866604
0.08856147  0.09690656
[9]  0.07879077  0.06971340  0.05967521
```

El peor es -0.06145852, mientras que el mejor 0.09690656, y además aparecen más positivos que negativos entre los outliers, por lo que parece que cuando hay uno tiende más a ser por encima que por debajo del retorno típico.

2.2 Ajustes a modelos de distribución

Tras este pequeño estudio, vamos a ver si los datos se ajustan bien a alguna distribución, en concreto vamos a probar con la normal, lognormal, gamma y la T de Student.

2.2.1 Normal

Vamos a contrastar la hipótesis

H_0 : R_{KO} sigue una distribución normal

frente a

H_1 : R_{KO} no sigue una distribución normal

Resolvemos el contraste con shapiro.test(), obteniendo:

```
> shapiro.test(retornos_KO)

Shapiro-Wilk normality test

data: retornos_KO
W = 0.98308, p-value = 1.378e-06
```

Un pvalor < 0.01, por lo que rechazamos la hipótesis nula, y por tanto asumimos que R_{KO} no sigue una distribución normal.

2.2.2 Lognormal

Planteamos el contraste:

H_0 : R_{KO} sigue una distribución lognormal

frente a

H_1 : R_{KO} no sigue una distribución lognormal

Como hay datos negativos, realizamos una translación de los datos, que no afecta al test para ver si se sigue la distribución lognormal o no. Concretamente los he desplazado todos asegurándome de que son estrictamente positivos y he realizado el contraste:

```
shapiro.test(log(retornos_KO + abs(min(retornos_KO))+1))

Shapiro-Wilk normality test

data: log(retornos_KO + abs(min(retornos_KO)) + 1)
W = 0.98598, p-value = 1.161e-05
```

De nuevo obtenemos pvalor < 0.01, por lo que rechazamos la hipótesis nula, y por tanto R_{KO} no sigue una distribución lognormal.

2.2.3 Gamma

Para hacer el ajuste de la gamma importamos las librerías MASS y fitdistplus, y hacemos la misma translación para no tener datos negativos.

```

> library(MASS)
> fitdistr(retornos_KO + abs(min(retornos_KO))+1, "gamma")
  shape      rate
  2694.2042   2535.6009
( 161.2460) ( 151.7663)

> library(fitdistrplus)
Loading required package: survival

> ajusteCocaCola <- fitdist(retornos_KO + abs(min(retornos_KO))+1, "gamma",
start=list(shape=2694.2042, rate=2535.6009))

> summary(ajusteCocaCola)
Fitting of the distribution ' gamma ' by maximum likelihood
Parameters :
  estimate Std. Error
shape 2709.521   157.1580
rate  2550.013   147.9176
Loglikelihood:  1530.726   AIC: -3057.453   BIC: -3048.597
Correlation matrix:
  shape      rate
shape 1.0000000 0.9999114
rate  0.9999114 1.0000000

```

Si ahora planteamos el contraste:

$H_0: R_{KO}$ sigue una distribución gamma de parámetros shape=2709.521, rate=2550.013

frente a

$H_1: R_{KO}$ no sigue una distribución gamma de parámetros shape=2709.521, rate=2550.013

y lo resolvemos con el test de la chi cuadrado:

```

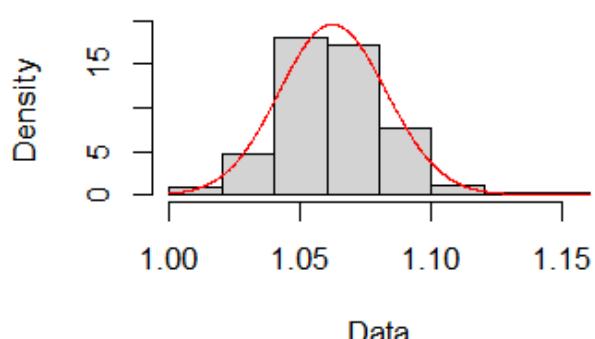
> gofstat(ajusteCocaCola)$chisq
[1] 19.36514
> gofstat(ajusteCocaCola)$chisq$pvalue
[1] 0.4982183

```

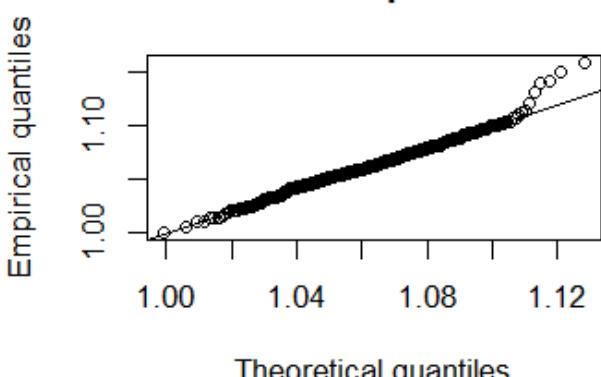
Obtenemos que el estadístico de contraste tiene valor 19.36514 y el pvalor es 0.4982183 > 0.05, por lo que aceptamos la hipótesis nula como verdadera y podemos suponer que R_{KO} sigue una distribución gamma de parámetros shape=2709.521, rate=2550.013.

Si además hacemos un plot(ajusteCocaCola) vemos:

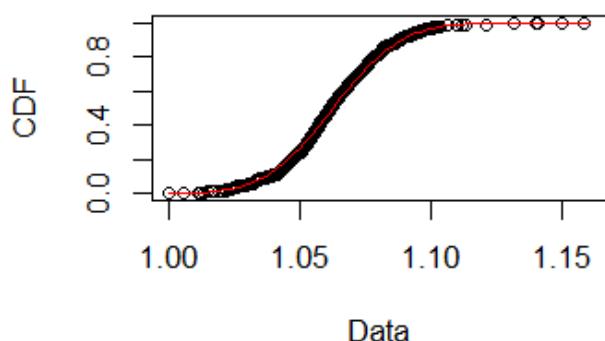
Empirical and theoretical dens.



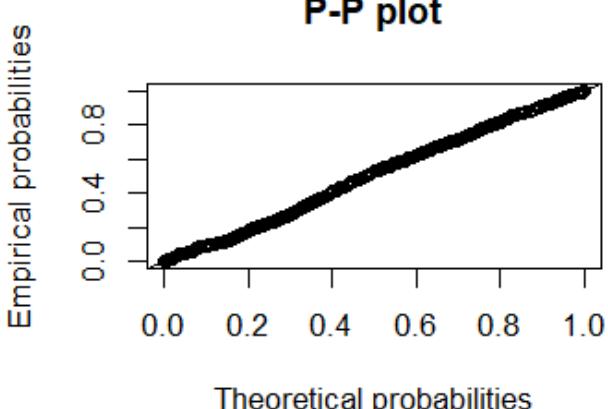
Q-Q plot



Empirical and theoretical CDFs



P-P plot



Las gráficas se adaptan bastante bien, y el gráfico Q-Q es bastante bueno a excepción de la cola superior, pero los datos parecen aceptar la hipótesis nula.

2.2.4 T de Student

Para obtener los estimadores de la T de Student hacemos, al igual que en la práctica 1,

```
> fitdistr(retornos_K0, "t")
      m           s           df
0.0007332938  0.0177170252  8.2228490811
(0.0007871535) (0.0008224514) (2.5050826904)
There were 12 warnings (use warnings() to see them)

> 0.0177170252*sqrt((8.2228490811)/(8.2228490811-2))
[1] 0.02036607
```

Ahora importamos la librería fGarch y hacemos el ajuste con los valores que hemos obtenido como estimadores.

```

> library(fGarch)
> ajusteTstudentCocaCola <- fitdist(retornos_KO, "std", start =
list(mean=0.0007332938, sd=0.02036607, nu=8.2228490811))
> summary(ajusteTstudentCocaCola)
Fitting of the distribution ' std ' by maximum likelihood
Parameters :
      estimate Std. Error
mean 0.0007252018 0.0007858782
sd   0.0204253595 0.0007568063
nu   7.7909553307 2.1794096715
Loglikelihood: 1540.199   AIC: -3074.399   BIC: -3061.114
Correlation matrix:
      mean        sd        nu
mean 1.00000000 -0.01524006 0.05622912
sd   -0.01524006 1.00000000 -0.44551004
nu    0.05622912 -0.44551004 1.00000000

```

Finalmente, realizamos el test de la chi cuadrado para resolver el contraste

$H_0: R_{KO}$ sigue una distribución T de Student de parámetros $\text{mean}=0.0007252018$, $\text{sd}=0.0204253595$, $\text{nu}=7.7909553307$

frente a

$H_1: R_{KO}$ no sigue una distribución T de Student de parámetros $\text{mean}=0.0007252018$, $\text{sd}=0.0204253595$, $\text{nu}=7.7909553307$

```

> gofstat(ajusteTstudentCocaCola)$chisq
[1] 12.94949
> gofstat(ajusteTstudentCocaCola)$chisq$pvalue
[1] 0.8411468

```

Obtenemos que el valor del estadístico para este contraste es $W=12.94949$ y el $p\text{valor}=0.8411468>0.05$, por lo que aceptamos la hipótesis nula, es decir, podemos asumir que los retornos semanales de Coca-Cola siguen una distribución T de Student con los parámetros mencionados.

2.2.5 Modelo más adecuado

Tenemos que tanto la distribución Gamma como la T de Student se ajustan bien a las observaciones, por lo que para decantarnos por una u otra vamos a hacer uso del criterio de información de Akaike. Aquel modelo con el valor más pequeño para AIC es el que mejor se ajusta, y en este caso hemos obtenido unos valores (se puede ver en la salida de los comandos) de:

- AIC= -3057.453 para la distribución Gamma.
- AIC= -3074.399 para la distribución T de Student.

Como la distribución T de Student tiene AIC menor (y también menor BIC si nos fijamos en la salida), concluimos que es el modelo de distribución más adecuado para describir las observaciones.

3. Análisis IBM

Como ya hemos declarado antes todas las librerías necesarias en el mismo script no hace falta que las volvamos a importarlas, por lo que en este análisis obviaremos ir mencionando que librerías son necesarias para cada cosa.

El intervalo de fechas que vamos a estudiar para IBM es el mismo que antes, podemos usar la misma variable en la que hemos guardado la fecha antes, y obtenemos los símbolos (IBM tiene ticker IBM), y extraemos los precios de cierre a un vector.

```
> getSymbols("IBM", src = "yahoo", from = start_date, to = end_date, auto.assign = TRUE)
[1] "IBM"
> # Extraemos el precio de cierre
> IBM_close <- Cl(IBM) # xts con una columna
> IBM <- as.numeric(IBM_close)
> # Retornos semanales
> retornos_IBM <- weekly.return(IBM)
```

La función `weekly.return` es la misma que para Coca-Cola. Así hemos obtenido los retornos semanales a partir de la cotización de cierre.

3.1 Visualización general de los datos

Ahora vamos a hacer una primera visualización de los datos usando el paquete `fdth`.

```
> tabla_intervalos_semanal_IBM <- fdt(retornos_IBM, start = min(retornos_IBM), end = max(retornos_IBM), h=0.01)
> tabla_intervalos_semanal_IBM
   Class limits  f  rf rf(%)  cf cf(%)
[-0.1197,-0.1097)  2  0.00  0.32    2  0.32
[-0.1097,-0.09966) 1  0.00  0.16    3  0.48
[-0.09966,-0.08966) 2  0.00  0.32    5  0.81
[-0.08966,-0.07966) 2  0.00  0.32    7  1.13
[-0.07966,-0.06966) 3  0.00  0.48   10  1.62
[-0.06966,-0.05966) 4  0.01  0.65   14  2.26
[-0.05966,-0.04966) 4  0.01  0.65   18  2.91
[-0.04966,-0.03966) 19 0.03  3.07   37  5.98
[-0.03966,-0.02966) 29 0.05  4.68   66 10.66
[-0.02966,-0.01966) 49 0.08  7.92  115 18.58
[-0.01966,-0.009656) 73 0.12 11.79  188 30.37
[-0.009656,0.0003437) 72 0.12 11.63  260 42.00
[0.0003437,0.01034) 81 0.13 13.09  341 55.09
[0.01034,0.02034) 86 0.14 13.89  427 68.98
[0.02034,0.03034) 61 0.10  9.85  488 78.84
```

```

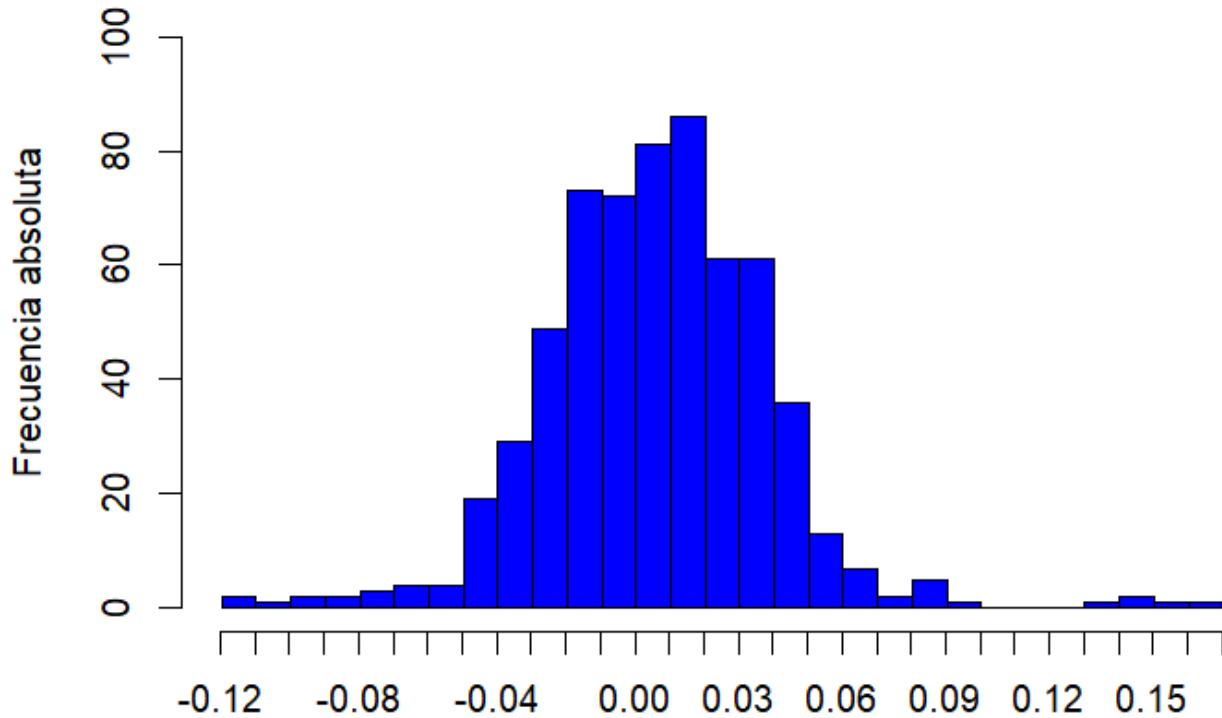
[0.03034,0.04034) 61 0.10 9.85 549 88.69
[0.04034,0.05034) 36 0.06 5.82 585 94.51
[0.05034,0.06034) 13 0.02 2.10 598 96.61
[0.06034,0.07034) 7 0.01 1.13 605 97.74
[0.07034,0.08034) 2 0.00 0.32 607 98.06
[0.08034,0.09034) 5 0.01 0.81 612 98.87
[0.09034,0.1003) 1 0.00 0.16 613 99.03
[0.1003,0.1103) 0 0.00 0.00 613 99.03
[0.1103,0.1203) 0 0.00 0.00 613 99.03
[0.1203,0.1303) 0 0.00 0.00 613 99.03
[0.1303,0.1403) 1 0.00 0.16 614 99.19
[0.1403,0.1503) 2 0.00 0.32 616 99.52
[0.1503,0.1603) 1 0.00 0.16 617 99.68
[0.1603,0.1703) 1 0.00 0.16 618 99.84

```

Hay bastantes más intervalos que los que salían con Coca-Cola y parece haber una mayor variabilidad, los datos parecen más dispersos.

Lo vemos en un histograma para que sea más visual.

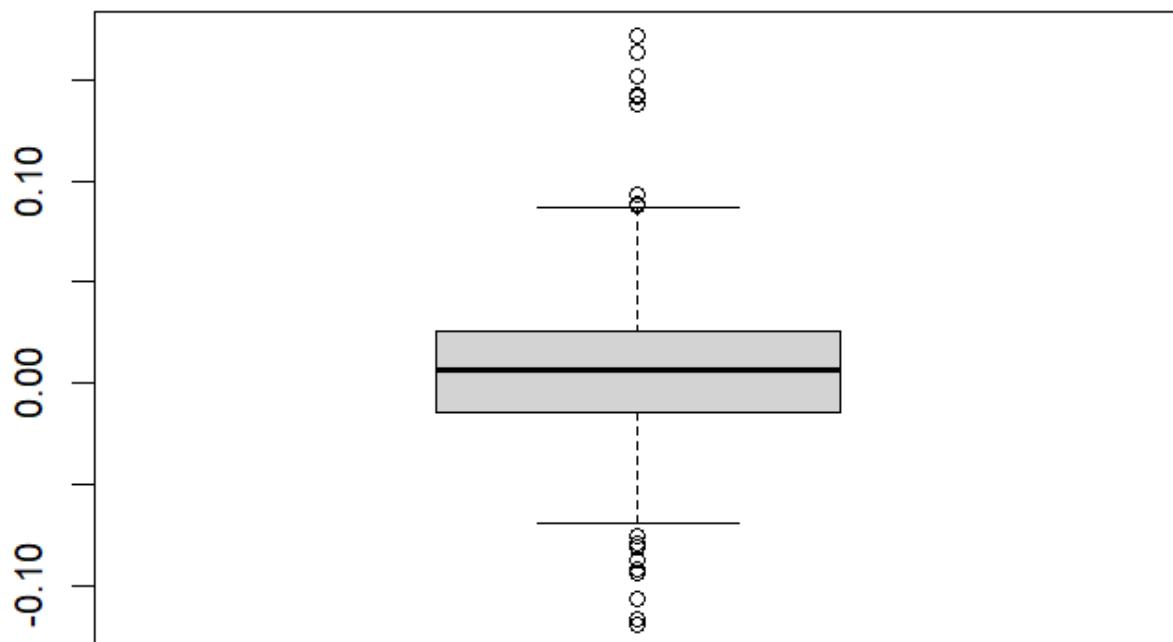
Histograma retornos semanales IBM



```
> summary(retornos_IBM)
   Min.    1st Qu.     Median      Mean   3rd Qu.      Max.
-0.119656 -0.014327  0.006843  0.006364  0.026118  0.171940
```

Tanto la media como la mediana son mayores que 0, lo que nos indica que, en general, el retorno semanal ha sido positivo, y por tanto el precio de cotización ha crecido en este periodo.

Si graficamos el diagrama de caja y bigotes vemos:



Las semanas de peor retorno cae menos que las semanas de mayor retorno. En concreto, los datos que están fuera del diagrama de caja y bigotes (los outliers) son:

```
boxplot(retornos_IBM)$out
[1]  0.14139293  0.09295546 -0.07957923 -0.07954921 -0.08776555 -0.10684418
-0.07549526 -0.08129448
[9] -0.09395052 -0.11965631  0.08771157  0.08837100  0.14258537  0.13745549
0.16329811  0.17194003
[17]  0.15164231 -0.09201321 -0.11708076  0.08899244
```

El peor es -0.11965631, mientras que el mejor 0.17194003, y además aparecen más positivos que negativos entre los outliers, por lo que parece que cuando hay uno tiende más a ser por encima que por debajo del retorno típico. También observar que el bigote del tercer cuantil es, en valor absoluto, ligeramente mayor que el del primer cuantil (que es negativo).

3.2 Ajustes a modelos de distribución

Tras este pequeño estudio, vamos a ver si los datos se ajustan bien a alguna distribución, en concreto vamos a probar con la normal, lognormal, gamma y la T de Student.

3.2.1 Normal

Vamos a contrastar la hipótesis

$H_0: R_{IBM}$ sigue una distribución normal

frente a

$H_1: R_{IBM}$ no sigue una distribución normal

Resolvemos el contraste con shapiro.test(), obteniendo:

```
> shapiro.test(retornos_IBM)

Shapiro-Wilk normality test

data: retornos_IBM
W = 0.95452, p-value = 6.826e-13
```

Un pvalor < 0.01, por lo que rechazamos la hipótesis nula, y por tanto R_{IBM} no sigue una distribución normal.

3.2.2 Lognormal

Planteamos el contraste:

$H_0: R_{IBM}$ sigue una distribución lognormal

frente a

$H_1: R_{IBM}$ no sigue una distribución lognormal

Como hay datos negativos, realizamos una translación de los datos, que no afecta al test para ver si se sigue la distribución lognormal o no. Concretamente los he desplazado todos asegurándome de que son estrictamente positivos y he realizado el contraste:

```
> shapiro.test(log(retornos_IBM + abs(min(retornos_IBM))+1))

Shapiro-Wilk normality test
```

```

data: log(retornos_IBM + abs(min(retornos_IBM))) + 1)
W = 0.95938, p-value = 4.779e-12

```

De nuevo obtenemos pvalor < 0.01, por lo que rechazamos la hipótesis nula, y por tanto R_{IBM} no sigue una distribución lognormal.

3.2.3 Gamma

Hacemos la misma translación para no tener datos negativos y las mismas operaciones que en el caso de Coca-Cola.

```

> fitdistr(retornos_IBM + abs(min(retornos_IBM))+1, "gamma")
  shape      rate
1145.09160   1016.93655
( 64.23383) ( 57.05772)

> ajusteIBM <- fitdist(retornos_IBM + abs(min(retornos_IBM))+1, "gamma",
start=list(shape=1145.09160, rate=1016.93655))

> summary(ajusteIBM)
Fitting of the distribution ' gamma ' by maximum likelihood
Parameters :
  estimate Std. Error
shape 1153.742   64.31517
rate 1024.624   57.13044
Loglikelihood: 1230.633   AIC: -2457.267   BIC: -2448.411
Correlation matrix:
  shape      rate
shape 1.0000000 0.9997748
rate  0.9997748 1.0000000

```

Si ahora planteamos el contraste:

$H_0: R_{IBM}$ sigue una distribución gamma de parámetros $shape=1153.742$, $rate=1024.624$

frente a

$H_1: R_{IBM}$ no sigue una distribución gamma de parámetros $shape=1153.742$, $rate=1024.624$

y lo resolvemos con el test de la chi cuadrado:

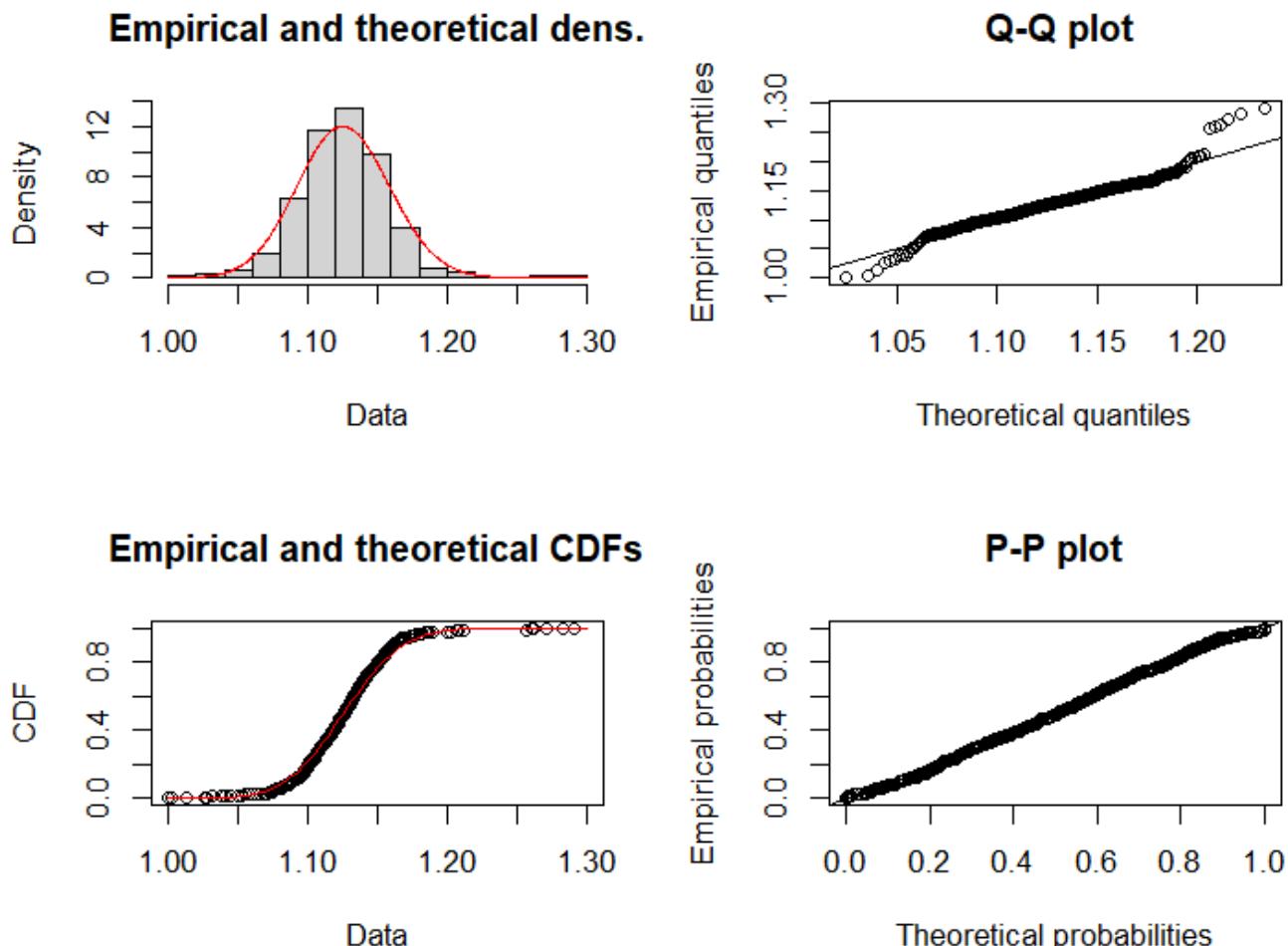
```

> gofstat(ajusteIBM)$chisq
[1] 33.81887
> gofstat(ajusteIBM)$chisq$pvalue
[1] 0.0273764

```

Obtenemos que el estadístico de contraste tiene valor 33.81887 y el pvalor es 0.0273764 > 0.01, por lo que aceptamos la hipótesis nula como verdadera y podemos suponer que R_{IBM} sigue una distribución gamma de parámetros shape=1153.742, rate=1024.624.

Si además hacemos un plot(ajustelBM) vemos:



Las gráficas se adaptan bastante bien, aunque el gráfico Q-Q falla bastante en las colas de la distribución, de ahí que si hubiéramos considerado un nivel de significación de 0.05 habríamos tenido que rechazar la hipótesis nula.

3.2.4 T de Student

Igual que en el caso de Coca-Cola, obtenemos los estimadores y hacemos el ajuste para la T de Student para obtener los estimadores de máxima verosimilitud.

```
> fitdistr(retornos_IBM, "t")
      m           s          df
 0.005994205  0.025530359  4.985766437
(0.001189135) (0.001139583) (0.879378648)
There were 14 warnings (use warnings() to see them)

> 0.025530359*sqrt((4.985766437)/(4.985766437-2))
[1] 0.03299096
```

```

> ajusteTstudentIBM <- fitdist(retornos_IBM, "std", start = list(mean=0.005994205,
sd=0.03299096, nu=4.985766437))
> summary(ajusteTstudentIBM)
Fitting of the distribution ' std ' by maximum likelihood
Parameters :
      estimate Std. Error
mean 0.006004694 0.001188622
sd   0.032983014 0.001572991
nu   4.977038382 0.876137986
Loglikelihood: 1262.969 AIC: -2519.937 BIC: -2506.653
Correlation matrix:
      mean        sd        nu
mean 1.0000000 0.00440025 -0.01013375
sd   0.00440025 1.0000000 -0.66671149
nu   -0.01013375 -0.66671149 1.00000000

```

Finalmente, realizamos el test de la chi cuadrado para resolver el contraste:

$H_0: R_{IBM}$ sigue una distribución T de Student de parámetros $\text{mean}=0.006004694$, $\text{sd}=0.032983014$, $\text{nu}=4.977038382$

frente a

$H_1: R_{IBM}$ no sigue una distribución T de Student de parámetros $\text{mean}=0.006004694$, $\text{sd}=0.032983014$, $\text{nu}=4.977038382$

```

> gofstat(ajusteTstudentIBM)$chisq
[1] 23.63372
> gofstat(ajusteTstudentIBM)$chisq$pvalue
[1] 0.2105797

```

Obtenemos que el valor del estadístico para este contraste es $W=23.63372$ y el $p\text{valor}=0.2105797 > 0.05$, por lo que aceptamos la hipótesis nula, es decir, podemos asumir que los retornos semanales de Coca-Cola siguen una distribución T de Student con los parámetros mencionados.

3.2.5 Modelo más adecuado

Tenemos que tanto la distribución Gamma como la T de Student se ajustan bien a las observaciones, por lo que para decantarnos por una u otra vamos a hacer uso del criterio de información de Akaike. Si hubiéramos optado por un nivel de significancia de 0.05 sólo habríamos aceptado la T de Student, pero con 0.01 hemos aceptado ambas distribuciones, por lo que en este caso aquel modelo con el valor más pequeño para AIC es el que mejor se ajusta, y hemos obtenido unos valores (se puede ver en la salida de los comandos) de:

- AIC= -2457.267 para la distribución Gamma.
- AIC= -2519.937 para la distribución T de Student.

Como la distribución T de Student tiene AIC menor (y también menor BIC si nos fijamos en la salida), concluimos que es el modelo de distribución más adecuado para describir las observaciones.

4. Contraste t de Student

Una vez obtenidas las dos muestras correspondientes a los **retornos semanales de Coca-Cola e IBM**, se plantea el contraste de la t de Student con el objetivo de comparar sus **medias poblacionales** y analizar si existen diferencias significativas en la rentabilidad media semanal de ambas empresas.

4.1 Muestras independientes o pareadas

Para decidir el tipo de contraste adecuado, es necesario determinar si las muestras son **independientes** o **pareadas**.

En este caso, aunque los retornos de ambas empresas se observan en el mismo periodo temporal, **no existe una correspondencia natural uno a uno** entre las observaciones que justifique un emparejamiento. Cada retorno semanal procede de la evolución propia de la cotización de una empresa distinta y no de la medición repetida de una misma variable bajo dos condiciones. Además, son empresas que se centran en sectores bastante distintos, operando Coca-Cola en el sector de bebidas y refrescos mientras que IBM se caracteriza por sus servicios de Software, Hardware y consultoría tecnológica.

Por tanto, los retornos semanales de Coca-Cola e IBM se consideran realizaciones de **dos variables aleatorias distintas** y las muestras se tratarán como **muestras independientes**.

4.2 Planteamiento del contraste

La normalidad de las distribuciones ya la hemos estudiado anteriormente. Como vimos, ni los retornos semanales de Coca-Cola ni los de IBM siguen una distribución normal ni una lognormal, sin embargo, como el tamaño de las muestras es mayor a 30, podemos hacer el contraste de la t de Student (aunque no nos garantice nada de las medias).

Bajo estas condiciones, se utiliza primero el **contraste de la F de Snedecor** para ver si las varianzas son iguales o distintas. Por lo tanto, primero hay que contrastar las hipótesis:

$$H_0: \sigma_{KO}^2 = \sigma_{IBM}^2$$

frente a

$$H_1: \sigma_{KO}^2 \neq \sigma_{IBM}^2$$

donde σ_{KO}^2 y σ_{IBM}^2 representan las varianzas de los retornos semanales de Coca-Cola e IBM, respectivamente.

Resolvemos el contraste con el comando:

```
> var.test(retornos_KO, retornos_IBM, paired=FALSE)
```

```
F test to compare two variances
```

```

data: retornos_KO and retornos_IBM
F = 0.37846, num df = 618, denom df = 618, p-value < 2.2e-16
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
0.3232007 0.4431595
sample estimates:
ratio of variances
0.3784567

```

Obtenemos un p-valor < 0.01, por lo que rechazamos la hipótesis nula.

Ahora planteamos, para ver si tienen misma media, el contraste de la t de Student para las hipótesis:

$$H_0: \mu_{KO} = \mu_{IBM}$$

frente a

$$H_1: \mu_{KO} \neq \mu_{IBM}$$

y lo resolvemos mediante el comando:

```

> t.test(retornos_KO, retornos_IBM, alternative = "two.sided", var.equal = FALSE)

Welch Two Sample t-test

data: retornos_KO and retornos_IBM
t = -3.3576, df = 1027.2, p-value = 0.0008151
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.008353657 -0.002191006
sample estimates:
mean of x   mean of y
0.001092057 0.006364389

```

Obtenemos un p-valor < 0.01, por lo que rechazamos la hipótesis nula, los retornos semanales de IBM y Coca-Cola no tienen la misma media.