

Universidad Complutense de Madrid

FACULTAD DE MATEMÁTICAS

AJUSTE DE DISTRIBUCIONES.

ÍNDICE METEOROLÓGICO DE RIESGO DE
INCENDIO FORESTAL DE AEMET

Memoria prácticas externas en AEMET

Autor:

Jesús Ocaña Cuesta

Tutor AEMET:

Marcelino Núñez Corchero

Mayo 2020

Índice general

0.1. Introducción	3
I FWI	4
1. ENFOQUES	5
2. TEMPORADAS POR SEPARADO	6
2.1. AÑO COMPLETO	6
2.1.1. Análisis descriptivo básico	6
2.1.2. Distribución de frecuencias	9
2.1.3. Histograma	9
2.1.4. Función de distribución	10
2.1.5. Ajustes	10
2.1.6. Bootstrap no paramétrico	19
2.2. TEMPORADA ALTA	19
2.2.1. Análisis descriptivo básico	19
2.2.2. Distribución de frecuencias	20
2.2.3. Histograma	20
2.2.4. Función de distribución	21
2.2.5. Ajustes	21
2.3. TEMPORADA BAJA	24
2.3.1. Análisis descriptivo básico	24
2.3.2. Distribución de frecuencias	24
2.3.3. Histograma	24
2.3.4. Función de distribución	25
2.3.5. Ajustes	25
3. SERIES TEMPORALES	30
3.1. AÑO COMPLETO	32
3.1.1. Ajustes	33
3.2. TEMPORADA ALTA	34
3.2.1. Ajustes	35
3.3. TEMPORADA BAJA	38
3.3.1. Ajustes	39
4. OTRAS LOCALIZACIONES	41
4.1. Santa Elena	41
4.1.1. Año Completo	41

4.1.2.	Temporada alta	42
4.1.3.	Temporada baja	44
4.2.	Alhama de Murcia	45
4.2.1.	Año Completo	45
4.2.2.	Temporada alta	47
4.2.3.	Temporada baja	49
4.3.	Navascués/Nabaskoze	51
4.3.1.	Año Completo	51
4.3.2.	Temporada alta	53
4.3.3.	Temporada baja	54
4.4.	Los Tojos	56
4.4.1.	Año Completo	56
4.4.2.	Temporada alta	58
4.4.3.	Temporada baja	60
4.5.	Vimianzo	62
4.5.1.	Año Completo	62
4.5.2.	Temporada alta	64
4.5.3.	Temporada baja	66
II	RIESGO	68
5.	ENFOQUE	69
6.	ANÁLISIS AÑOS COMPLETOS	70
6.1.	Ajustes	71
6.1.1.	Ajuste a distintas distribuciones	72
7.	ANÁLISIS TEMPORADA ALTA	75
7.1.	Análisis del ajuste de los datos a la distribución multinomial	77
8.	ANÁLISIS TEMPORADA BAJA	78
8.1.	Análisis del ajuste de los datos a la distribución multinomial	80
9.	ANÁLISIS DEL RESTO DE LOCALIZACIONES	81
9.1.	Alhama de Murcia	81
9.1.1.	Años completos	81
9.1.2.	Temporada alta	83
9.1.3.	Temporada baja	85
9.2.	Navascués/Nabaskoze	87
9.2.1.	Años completos	87
9.2.2.	Temporada alta	89
9.2.3.	Temporada baja	91
9.3.	Los Tojos	93
9.3.1.	Años completos	93
9.3.2.	Temporada alta	95
9.3.3.	Temporada baja	97
9.4.	Vimianzo	99

9.4.1. Años completos	99
9.4.2. Temporada alta	101
9.4.3. Temporada baja	103

0.1. Introducción

El presente trabajo pretende conocer la distribución que más se asemeja a los datos que tenemos relativos al Índice de riesgos forestales para la localización de Torrecilla de los Ángeles, situada en la provincia de Cáceres, Extremadura. Para ello se analizarán los datos relativos al periodo 2014-2019.

El análisis también tiene otro propósito, la generación de un código en R que pueda ser usado en el futuro con diferentes localizaciones.

Parte I

FWI

Capítulo 1

ENFOQUES

El primer análisis que se ha hecho ha sido de la variable FWI. Se ha realizado para el conjunto de años por completo. Los datos que manejamos corresponden al periodo 2014-2019.

Los datos que se van a utilizar han sido recogidos diariamente. Esto hace que el factor tiempo tenga importancia ya que los datos pueden estar autocorrelados. Más adelante se corroborará esta situación.

Para trabajar con esto, se pueden tomar diferentes medidas:

1. Hacer submuestras desde los datos. Por ejemplo, escoger 30 días aleatoriamente de cada año estudiado. Después se pueden utilizar técnicas de Bootstrap para obtener resultados. El Bootstrap es una técnica que consiste en realizar muchos remuestreos con reposición de nuestra muestra original y obtener los estadísticos que deseemos. Estos podrán ser después utilizados para conocer las características de nuestra variable y su distribución.
2. Realizar ajustes por temporadas. Realizar análisis para la temporada alta (mayo-octubre) y baja (noviembre-abril) por separado.
3. Eliminar la tendencia de los datos para crear residuos incorrelados (parte no dependiente del pasado) y usarlos como variables aleatorias.

En este trabajo primero se ha intentado obtener la distribución para cada temporada, después se han realizado submuestras y finalmente se ha eliminado la tendencia y obtenido el ajuste.

Capítulo 2

TEMPORADAS POR SEPARADO

2.1. AÑO COMPLETO

2.1.1. Análisis descriptivo básico

Se pretende conocer a qué distribución se asemejan más nuestros datos relativos al FWI. Se trata de una variable continua.

Lo primero que se ha observado son estadísticos básicos que pueden mostrar información básica sobre lo que nos vamos a encontrar.

Media

Un simple pero efectivo estadístico para resumir las características de un conjunto de datos es la media. A veces, es conocido como promedio.

Definición 1 Asumimos que tenemos un conjunto de datos x de N elementos, $x_1, x_2, \dots, x_N$

$$media(x) = \frac{1}{N} \sum_{i=1}^{i=N} x_i \quad (2.1)$$

La media muestra dónde se encuentra el conjunto general de datos, por ello se le conoce como un parámetro de ubicación, permite conocer dónde se encuentran los datos a lo largo de una línea numérica.

La media se puede usar también como una guía que nos diga cómo van a ser los nuevos valores que se puedan introducir, si no se tiene otro tipo de información.

En nuestro caso, la media se sitúa en **16,01**.

Desviación típica

También podemos conocer en qué medida los datos se encuentran cerca de la media. Para ello se usa la desviación típica.

Definición 2 Asumimos que tenemos un conjunto de datos de N elementos, $x_1, x_2, \dots, x_N$. La desviación típica es:

$$std(x_i) = \sqrt{\frac{1}{n} \sum_{i=1}^{i=N} (x_i - media(x))^2} \quad (2.2)$$

Mide el tamaño de la desviación promedio de la media para un conjunto de datos, es decir, qué tan amplia es la difusión de los datos. Es conocida como un parámetro de escala. Cuando la desviación típica es grande, existen muchos elementos con valores más grandes, o más pequeños, que la media. Cuando la desviación típica es pequeña nos encontramos con valores muy cercanos a la media.

Para nuestros datos, la desviación típica es de **17,81**, lo que nos indica que es posible que los datos estén separados. Para saber si realmente los datos se encuentran alejados de la media o no, es necesario basarse también en otros estadísticos, como son el máximo y el mínimo.

Máximo y mínimo

Los valores máximo y mínimo nos dice cuales son los valores de los extremos de nuestro dataset. En nuestro caso el valor mínimo es **0** y el máximo **70,20**.

Teniendo en cuenta el valor de la media (16,01) y de la desviación típica (17,81), podemos decir que existe una longitud muy amplia de datos. Para estos casos la media no tiene mucho sentido. Nuestros datos son en su mayoría bajos, exceptuando algunos valores que llegan a 70, lo que supone que la media no represente bien la estructura de los datos. Es por ello, que se debe tener en cuenta la mediana.

Mediana

Cuando existen valores muy separados, existe mucha variación, es necesario usar otros estadísticos en lugar de la media, que representen mejor la distribución de los datos. Es el caso de la mediana.

Definición 3 La mediana de un dataset se obtiene ordenando los datos y encontrando aquel que se encuentre en la mitad de la lista.

Por ejemplo, la mediana del conjunto de datos 2,5,7 será 5. La mediana del conjunto 3,4,5,6 será 4,5.

En nuestro caso, la mediana es de **7,30**. Si lo comparamos con el valor de la media (16,01) podemos ver que existe mucha diferencia. Esto indica que la mayoría de los datos son valores pequeños.

Cuartiles

Los cuartiles son estadísticos que describen escala, como la desviación típica, pero se ven menos afectados por valores atípicos (valores que se encuentran muy alejados del resto) que la desviación típica. Los cuartiles dividen en cuatro partes iguales un conjunto de datos ordenados.

Definición

4

1. El primer cuartil es el valor tal que el 25 % de los datos es menor o igual a este valor.
2. El segundo cuartil es la mediana. El 50 % de los datos es menor o igual a este valor.
3. El tercer cuartil es el valor tal que el 75 % de los datos es menor o igual a este valor.

En nuestro caso obtenemos como primer cuartil **0,60** y como tercer cuartil **31,50**

Es decir, el 25 % de los datos se encuentran entre 0 y 0,6. Esto significa que muchos datos son tienen valores muy pequeños. La diferencia que hay entre el tercer cuartil (31,50) y el máximo (70) nos indica que hay muy pocos valores que sean elevados.

Resumen

La mayoría de los datos se tienen valores muy pequeños. Los valores grandes no son muchos y el hecho de que haya tanta diferencia entre la media y la mediana nos indica de manera clara que los datos no se distribuirán como una normal, ya que en una distribución normal la media y la mediana coinciden. Esto es debido a que, en una distribución normal, la mitad de los datos son inferiores a la media y la otra mitad superiores. Un ejemplo de distribución normal es la siguiente:

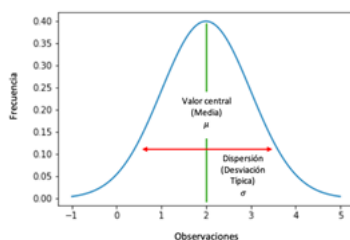


Figura 2.1: Ejemplo de Distribución normal

El resumen, por tanto, de lo visto hasta ahora es el siguiente:

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
0	0,60	7,30	16,01	31,50	70,20	17,81551

2.1.2. Distribución de frecuencias

Con el objetivo de realizar un histograma y conocer cómo es la forma en la que se distribuyen los datos, realizamos primero el cálculo de la distribución de frecuencias. Esto se ha hecho siguiendo la regla de Sturges, que nos identifica el número de clases que se deben tener en cuenta para realizar un histograma. Esta regla nos sugiere el ancho de cada intervalo:

$$k = 1 + \log_2(N) \quad (2.3)$$

Donde N es el tamaño de la muestra a representar y k el número óptimo de intervalos. Siguiendo esta regla, obtenemos 13 intervalos:

Class limits	f	rf	rf(%)	cf	cf(%)
[0,5.454)	994	0.45	45.45	994	45.45
[5.454,10.91)	248	0.11	11.34	1242	56.79
[10.91,16.36)	132	0.06	6.04	1374	62.83
[16.36,21.82)	86	0.04	3.93	1460	66.76
[21.82,27.27)	98	0.04	4.48	1558	71.24
[27.27,32.72)	105	0.05	4.80	1663	76.04
[32.72,38.18)	134	0.06	6.13	1797	82.17
[38.18,43.63)	164	0.07	7.50	1961	89.67
[43.63,49.09)	113	0.05	5.17	2074	94.83
[49.09,54.54)	69	0.03	3.16	2143	97.99
[54.54,59.99)	29	0.01	1.33	2172	99.31
[59.99,65.45)	12	0.01	0.55	2184	99.86
[65.45,70.9)	3	0.00	0.14	2187	100.00

Figura 2.2: Distribución de frecuencias

Se puede observar que el 45 % de los datos se encuentran en el primer intervalo, lo que ya podíamos intuir teniendo en cuenta los valores del primer cuartil.

2.1.3. Histograma

Con el histograma representamos la cantidad de datos que tiene cada intervalo. Podemos ver que la mayoría de los datos están en el primer intervalo (45,45 %).

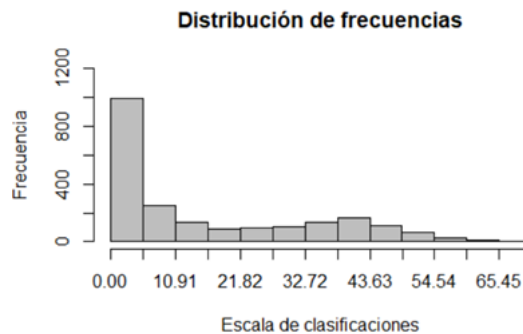


Figura 2.3: Histograma

También observamos que hay otro pico en el intervalo 38,18-43,63. La existencia de dos picos nos indica que se trata de un histograma multimodal, en este caso bimodal. Esto puede suponer que existe la posibilidad de que haya dos distribuciones en el conjunto de datos

debido a características de la variable. En nuestro caso, puede deberse al uso de datos de temporada baja (noviembre-abril) junto con los datos de temporada alta (mayo-octubre).

Una solución sería hacer estudios independientes de ambas temporadas. En este trabajo se ha hecho este estudio para temporada alta y temporada baja, además de para el conjunto de datos.

2.1.4. Función de distribución

Las variables aleatorias pueden ser definidas usando la distribución de probabilidad (probability distribution function). En pocas palabras, no podemos predecir exactamente el valor de una variable aleatoria, pero podemos definir la probabilidad de que la variable aleatoria pueda tomar un cierto rango de valores.

La función de distribución de probabilidad acumulada (CDF) calcula la probabilidad de que una variable aleatoria x sea menor o igual que algún valor especificado de X . Es decir,

$$F(X) = p(x \leq 0) \quad (2.4)$$

Al ser probabilidades, el rango de valores de CDF estará entre 0 y 1.

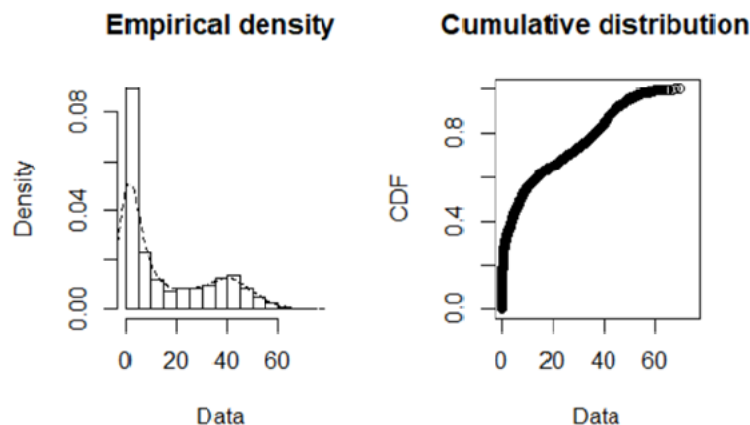


Figura 2.4: Gráfico de la función de distribución

El CDF nos muestra los valores cf% de la tabla de frecuencias (figura 2.2). También podemos ver un pico, en torno a la mitad de la línea, que muestra el pico que vemos en el histograma. Son los valores relativos al intervalo 38,18-43,63.

2.1.5. Ajustes

Nos encontramos ante un conjunto de datos cuya distribución desconocemos. Con los ajustes, se pretenden usar los datos que tenemos para obtener la distribución teórica de la

que provienen. Esto significa que no solo queremos conocer el tipo de distribución que siguen sino también los parámetros (location, shape, scale) que describen la distribución.

Ajustar una distribución significa encontrar la distribución teórica que mejor represente a la población de la variable aleatoria muestreada.

Existen diferentes métodos para realizar los ajustes. En este caso hemos usado el método de máxima verosimilitud (maximum likelihood), definido como:

$$L(\theta) = \prod_{i=1}^n f(x_i | \theta) \quad (2.5)$$

x_i son las n observaciones de la variable X y $f(.|\theta)$ la función de densidad de la distribución paramétrica a la que se pretende ajustar.

En un principio se ha puesto en marcha una función llamada `descdist` que nos muestra los valores de los estadísticos básicos, así como la asimetría (skewness) y la curtosis.

La asimetría nos permite conocer si hay el mismo número de elementos a la izquierda y a la derecha de la media. Existen tres tipos:

1. Asimetría negativa: la cola de la distribución se alarga para valores inferiores a la media.
2. Asimetría positiva. La cola de la distribución se alarga para valores superiores a la media.
3. Simetría: existe el mismo número de elementos a la izquierda y derecha de la media. Media y mediana coinciden. Se trata de distribución normal.

Se usa el coeficiente de Fisher.

$$\gamma_1 = \frac{\mu_3}{\sigma^3} \quad (2.6)$$

Donde μ_3 es el tercer momento en torno a la media y σ es la desviación estándar. Si el valor es mayor a 0 tenemos una asimetría positiva. Si el valor es menor que 0 tenemos asimetría negativa. Un valor de 0 supone simetría.

En nuestro caso obtenemos un valor de **0,84**. Por tanto, asimetría positiva. Esto implica que la mayoría de los valores son inferiores a la media. La mayoría de las observaciones se encuentran en la parte izquierda de la distribución y la cola de la distribución se alarga a la derecha para valores superiores a la media.

La curtosis nos indica el grado de concentración de los datos. Si es grande significa que los datos están muy concentrados y el pico en un histograma será mas puntiagudo. Se calcula:

$$\beta_2 = \frac{\mu_4}{\sigma^4} \quad (2.7)$$

Donde μ_4 es el cuarto momento en torno a la media y σ es la desviación estándar. Existen 3 tipos:

1. Leptocúrtica. Cuando el valor es superior a 3. Mas puntiaguda con colas más gruesas que la normal.

2. Platicúrtica. Cuando el valor es inferior a 3. Menos puntiaguda con colas menos gruesas que la normal.
3. Mesocúrtica. Cuando el valor es igual a 3. Se trata de distribución normal.

El valor obtenido es de **2,33**. Por tanto tenemos una curtosis platicúrtica.

Esta función también nos muestra un gráfico (Cullen and Frey) con el tipo de distribución que podrían seguir los datos a priori teniendo en cuenta los valores de asimetría y curtosis, lo que puede facilitar el trabajo.

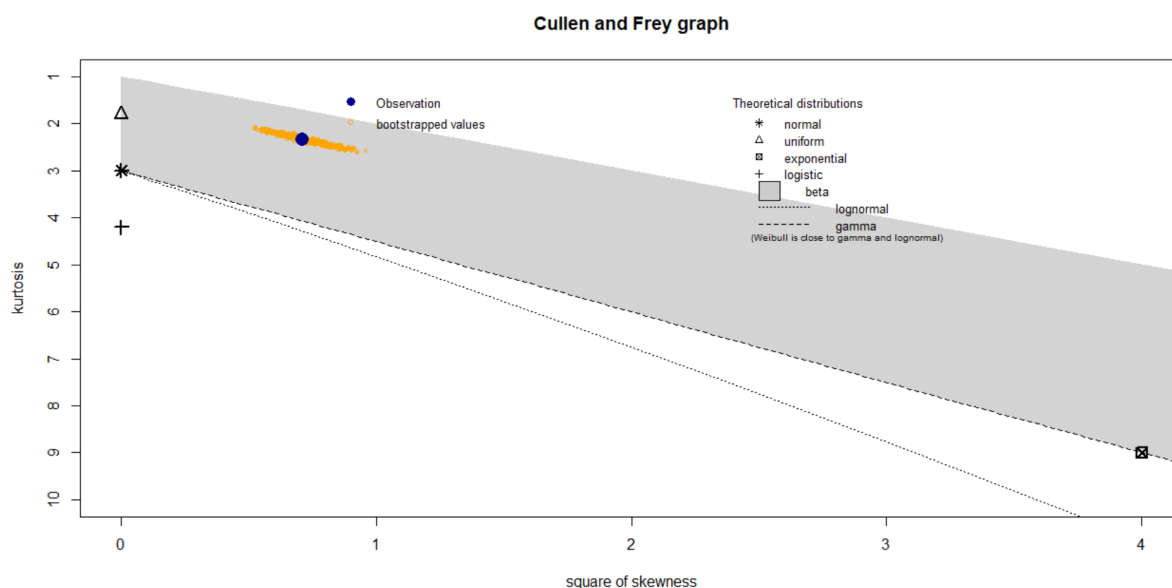


Figura 2.5: Gráfico Cullen and Frey

El punto azul es el valor de la asimetría y la curtosis de nuestros datos. Los puntos naranjas muestran los resultados al hacer Bootstrap no paramétrico (se han realizado 1000 remuestreos de la muestra y calculado sus respectivos valores de asimetría y curtosis).

El resultado que obtenemos no es muy esclarecedor. Nos muestra que nuestra distribución podría seguir una Beta, pero eso no es posible debido a las características de nuestra variable. Una distribución Beta se caracteriza por tener los valores de la función de densidad entre 0 y 1. Al ver el valor mínimo y máximo sabemos nuestros valores se encuentran en el intervalo 0 a 70.20, por lo que no puede seguir en ningún caso una distribución Beta.

Por tanto, se ha decidido realizar ajuste siguiendo diferentes distribuciones conocidas. Una vez hecho el ajuste (como ya hemos dicho se ha utilizado el método de máxima verosimilitud), se evalúa el ajuste para comprobar si es significativo o no. Es decir, si la distribución empírica se parece a la distribución teórica. Para esto, se usan los parámetros que se han estimado con el ajuste en las pruebas de bondad de ajuste.

La prueba de bondad de ajuste usada en la de Kolmogórov-Smirnov. Para la normal también se han utilizado Anderson-Darling y Lilliefors (estas pruebas solo están disponibles para normal)

La prueba de Kolmogórov-Smirnov mide la distancia vertical máxima entre la función de distribución acumulada empírica y una distribución acumulada teórica de referencia. Es decir, se calcula a partir de la diferencia en valor absoluto entre las funciones de distribución acumuladas teórica y observada. Anderson-Darling y Lilliefors son unas modificaciones de la prueba de Kolmogorov-Smirnov donde se les da más peso a las colas de la distribución.

En estas pruebas la hipótesis nula es que la variable se ajusta a una distribución F establecida y la hipótesis alternativa es que no se ajusta. Establecemos como nivel de significación el 5%. Es decir, es la probabilidad de rechazar la hipótesis nula cuando es verdadera (conocido como error de tipo 1). Si el valor que obtenemos es superior a 0,05, aceptamos la hipótesis nula.

También se tienen en cuenta los valores de AIC y BIC, que se obtienen al realizar el ajuste. Estos son unos criterios que nos permiten comparar los distintos ajustes que se hayan realizado.

Al comparar ajustes realizados, nos quedaremos con aquella distribución que tenga los valores de AIC y BIC menores, siempre y cuando hayamos aceptado la prueba de Kolmogorov-Smirnov.

Los tipos de distribuciones teóricas que se han utilizado han sido las siguientes:

1. Normal
2. Logística
3. Exponencial
4. Gamma
5. Pareto

Los resultados se han graficado, obteniendo 4. En todos ellos aparece una línea que representa la forma de la distribución teórica con la que es esta comparando/ajustando. Cuanto más se parezcan los datos a esas líneas, más posibilidad habrá de que la variable siga esa distribución teórica.

1. El primer gráfico muestra el histograma.
2. El gráfico qq-plot usa el cuantil sobre el eje x. compara los cuantiles de la distribución de la variable con la teórica.
3. El gráfico empirical and theoretical cdfs muestra la función de distribución acumulada.
4. El gráfico pp-plot compara la función de distribución acumulada de la variable con la función de distribución acumulada de la distribución teórica.

Distribución Normal El realizar un ajuste para ver si nuestros datos siguen una distribución normal, se ha hecho más por tener el código creado para futuras pruebas que para hacerlo para estos datos. Ya hemos visto, según los valores de media, mediana, asimetría y curtosis, que no siguen una normal. Aun así, se ha calculado el ajuste y realizado la prueba de bondad de ajuste.

```
Fitting of the distribution ' norm ' by maximum likelihood
Parameters :
      estimate Std. Error
mean 16.00796  0.3808680
sd   17.81143  0.2693143
Loglikelihood: -9401.43   AIC: 18806.86   BIC: 18818.24
Correlation matrix:
      mean sd
mean   1  0
sd     0  1
```

Figura 2.6: Resultados ajuste Normal

Los valores estimados de mean y sd son los que se usan en la prueba de bondad de ajuste.

Las pruebas de Kolmogorov-smirnov, Anderson-Darling, Shapirto y Lillie confirman que no se ajusta a una normal.

One-sample Kolmogorov-Smirnov test

```
data: t
D = 0.19189, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Anderson-Darling normality test

```
data: t
A = 149.25, p-value < 2.2e-16
```

Shapiro-wilk normality test

```
data: t
W = 0.82761, p-value < 2.2e-16
```

Lilliefors (Kolmogorov-Smirnov) normality test

```
data: t
D = 0.19186, p-value < 2.2e-16
```

Gráficamente:

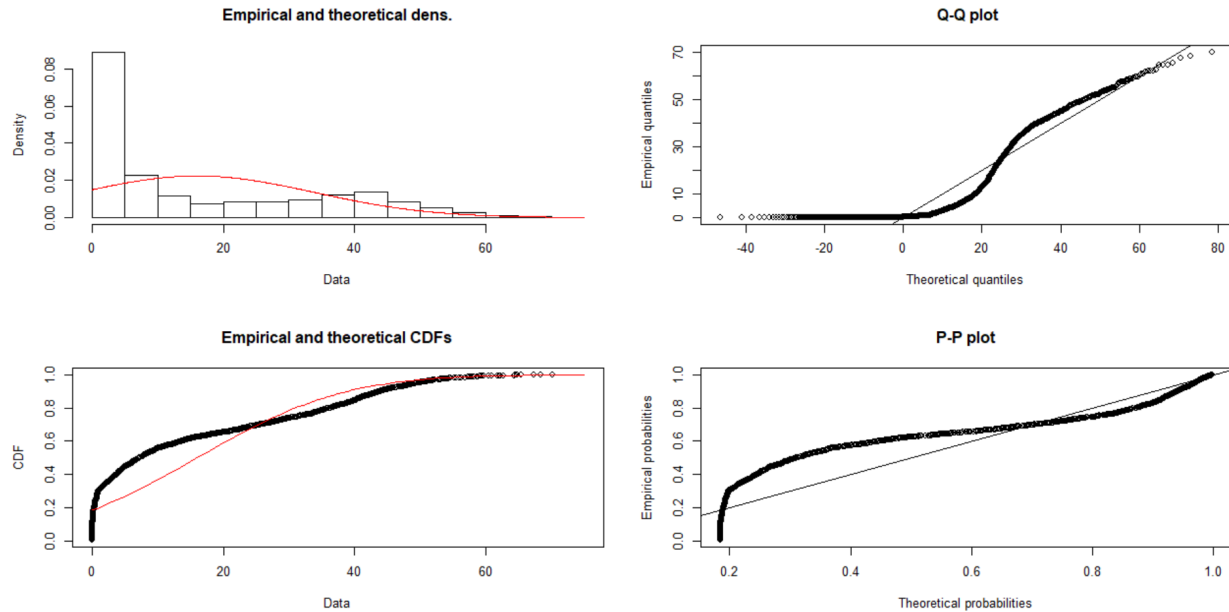


Figura 2.7: Gráfico de distribución, Q-Q plot, Empirical and theoretical CDFs y P-P plot

Como vemos, en todos los gráficos la variable no se parece nada a la línea que representa la distribución teórica.

Distribución Logística Los valores obtenidos son los siguientes:

```
Fitting of the distribution 'logis' by maximum likelihood
Parameters :
      estimate Std. Error
location 13.62878  0.3969665
scale    10.42283  0.1844529
Loglikelihood: -9456.914  AIC: 18917.83  BIC: 18929.21
Correlation matrix:
      location      scale
location 1.0000000  0.1450852
scale    0.1450852  1.0000000
```

Figura 2.8: Resultados ajuste Logística

Al introducirlos en la prueba de bondad de ajuste obtenemos que no se ajusta. Esto es lógico debido a que una distribución logística es muy parecida a una normal con menos curtosis.

One-sample Kolmogorov-Smirnov test

```
data: t
D = 0.21289, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Las representaciones graficas también muestran un resultado similar.

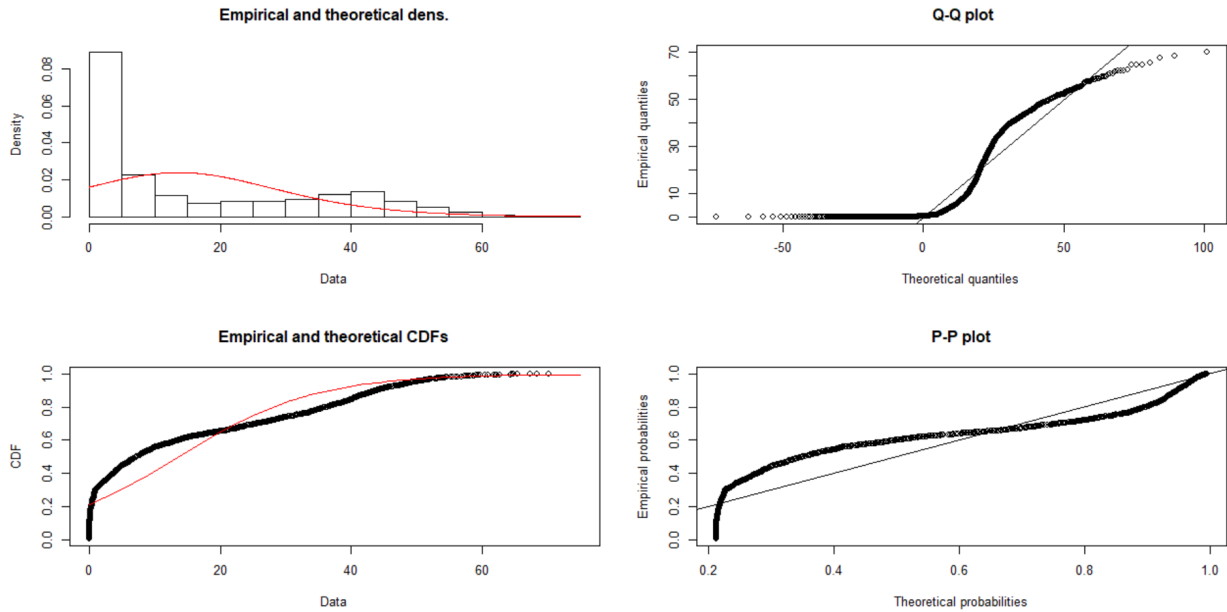


Figura 2.9: Gráfico de distribución, Q-Q plot, Empirical and theoretical CDFs y P-P plot

Distribución Exponencial Este tipo de distribución se asemeja más a la forma de los datos que tenemos, por ello en un principio se podría pensar que la variable se debería ajustar bien. Sin embargo, la prueba de bondad de ajuste nos dice que no es correcto.

```
Fitting of the distribution 'exp' by maximum likelihood
Parameters :
      estimate Std. Error
rate 0.06246894 0.001335452
Loglikelihood: -8251.739 AIC: 16505.48 BIC: 16511.17
```

Figura 2.10: Resultados ajuste Exponencial

La prueba de bondad de ajuste se rechaza.

```
One-sample Kolmogorov-Smirnov test

data: t
D = 0.24488, p-value < 2.2e-16
alternative hypothesis: two-sided
```

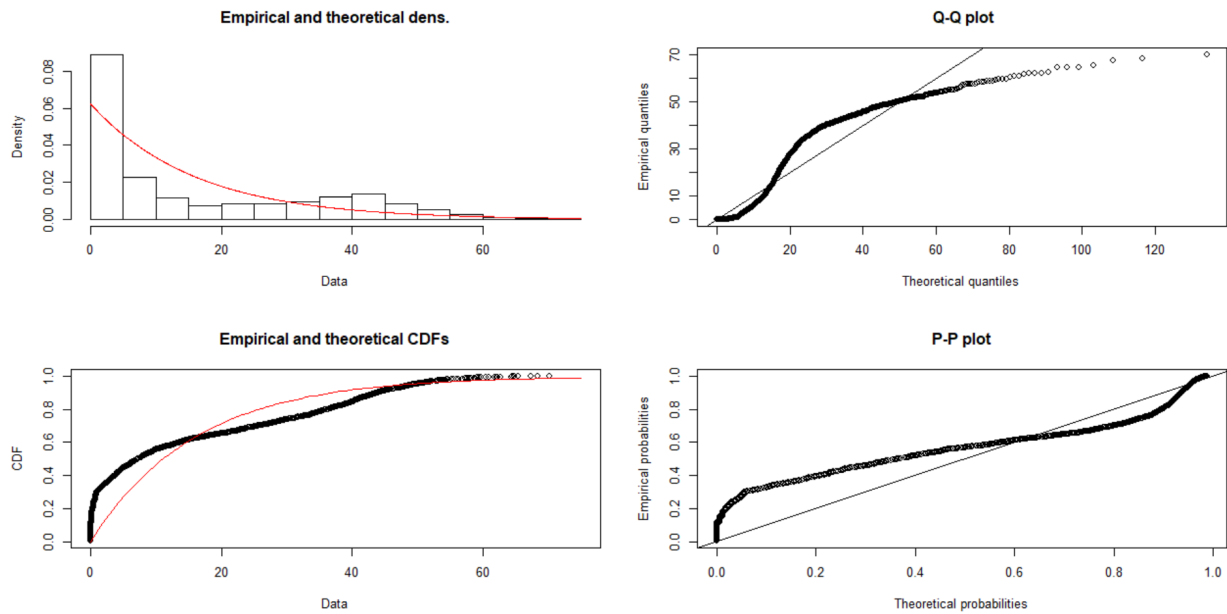


Figura 2.11: Gráfico de distribución, Q-Q plot, Empirical and theoretical CDFs y P-P plot

Distribución Gamma La variable tampoco se ajusta a una distribución gamma.

```
Fitting of the distribution ' gamma ' by matching moments
Parameters :
  estimate
shape 0.80774451
rate 0.05045894
Loglikelihood: Inf AIC: -Inf BIC: -Inf
```

One-sample Kolmogorov-Smirnov test

```
data: t
D = 0.83561, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Figura 2.12: Resultados ajuste Gamma

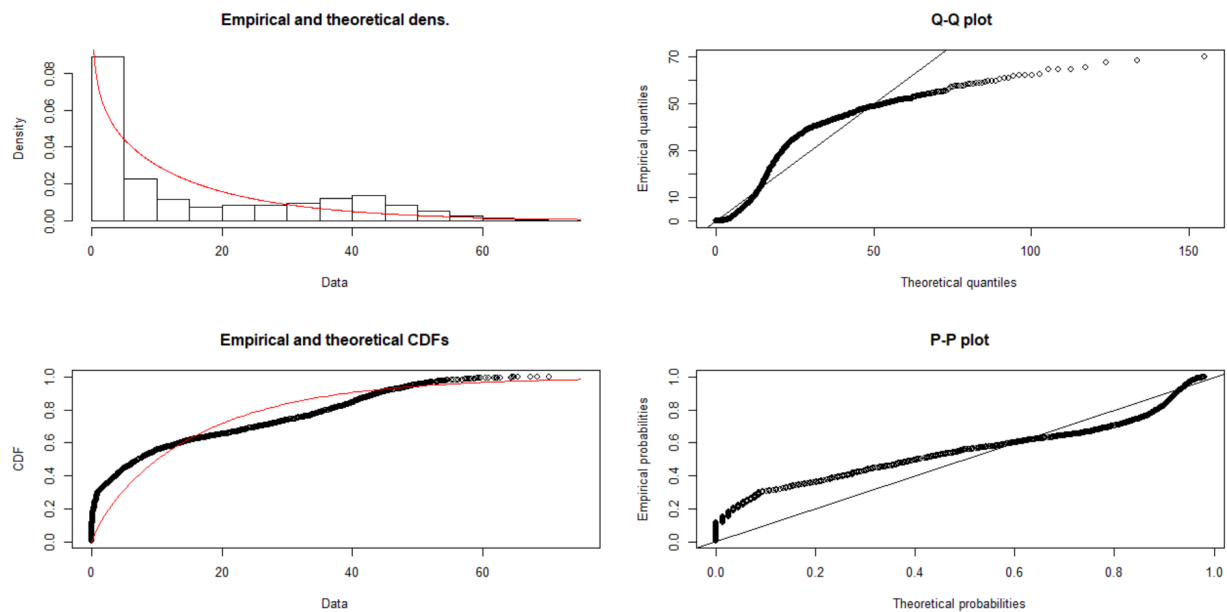


Figura 2.13: Gráfico de distribución, Q-Q plot, Empirical and theoretical CDFs y P-P plot

Distribución de Pareto No se ajusta a una distribución de Pareto.

```
Fitting of the distribution 'pareto' by maximum likelihood
Parameters :
      estimate Std. Error
shape  3.777603  0.7526834
scale 48.998792 12.2755567
Loglikelihood: -8370.755  AIC: 16745.51  BIC: 16756.89
Correlation matrix:
      shape  scale
shape 1.0000000 0.9942246
scale 0.9942246 1.0000000
```

One-sample Kolmogorov-Smirnov test

```
data: (t + 1)
D = 0.16518, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Figura 2.14: Resultados ajuste Pareto

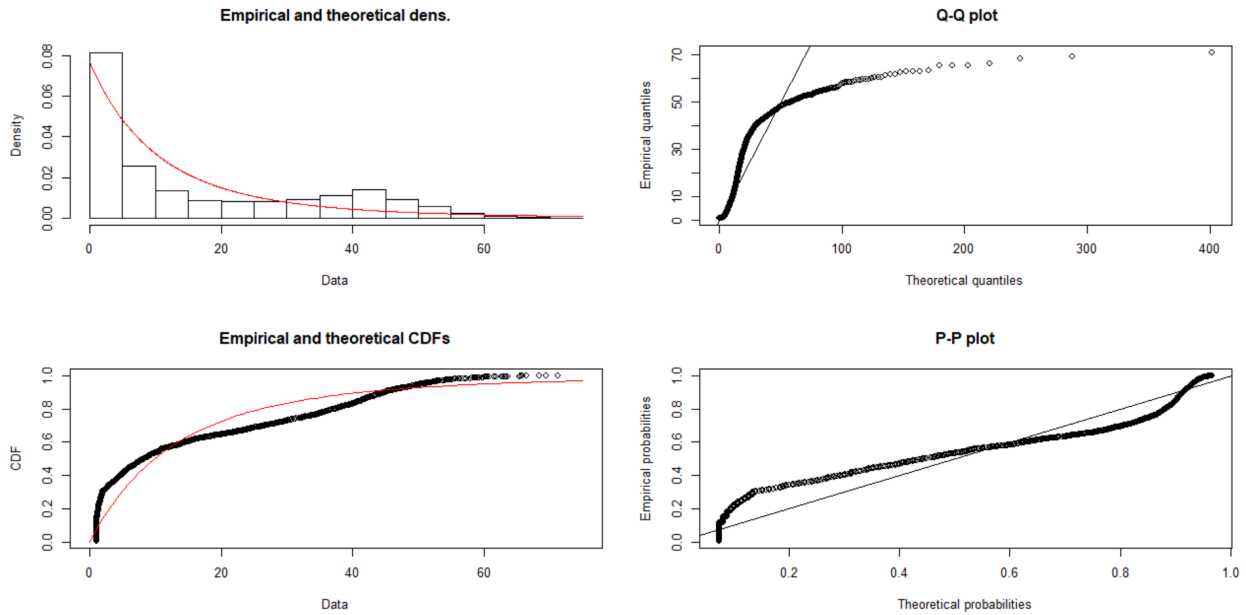


Figura 2.15: Gráfico de distribución, Q-Q plot, Empirical and theoretical CDFs y P-P plot

Tras estos resultados, no se ha encontrado una distribución que se asemeje a la variable FWI para los años completos.

2.1.6. Bootstrap no paramétrico

Aquí, se decidió poner en marcha el Bootstrap no paramétrico que permite obtener parámetros sin tener en cuenta la distribución que sigue.

Se realizaron 2300 interacciones con los que se calculó la media(**16,0161** frente a 16,01 de la muestra), el error estándar(**0,3717** frente a 0,3809 de la muestra) y los intervalos de confianza(**15,27-16,72** frente a 15,26-16,75).

Hecho todo esto, se realizó el mismo análisis solo para los meses de temporada alta.

2.2. TEMPORADA ALTA

2.2.1. Análisis descriptivo básico

En este caso, tenemos 915 observaciones con los siguientes resultados.

Resumen

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
0	19,05	33,90	30,97	42,80	70,20	16,27897

Ahora observamos que la mediana es superior a la media, diferenciándose del análisis anterior donde la mediana era menor.

2.2.2. Distribución de frecuencias

La tabla de frecuencias esta vez tiene 11 intervalos (calculados a través de la regla de Sturges) y la distribución es totalmente diferente del análisis anterior. En esta ocasión, han desaparecido muchos valores bajos, al tratarse de meses mas cálidos. Así, la forma que obtenemos en el histograma es mas parecida a una normal.

Class limits	f	rf	rf(%)	cf	cf(%)
[0,6.446)	102	0.11	11.15	102	11.15
[6.446,12.89)	61	0.07	6.67	163	17.81
[12.89,19.34)	69	0.08	7.54	232	25.36
[19.34,25.78)	83	0.09	9.07	315	34.43
[25.78,32.23)	108	0.12	11.80	423	46.23
[32.23,38.67)	143	0.16	15.63	566	61.86
[38.67,45.12)	177	0.19	19.34	743	81.20
[45.12,51.57)	96	0.10	10.49	839	91.69
[51.57,58.01)	52	0.06	5.68	891	97.38
[58.01,64.46)	18	0.02	1.97	909	99.34
[64.46,70.9)	6	0.01	0.66	915	100.00

Figura 2.16: Distribución de frecuencias

2.2.3. Histograma

Observamos que parece una normal, exceptuando por la parte izquierda, donde todavía existen valores bajos a pesar de estar hablando de la temporada alta.

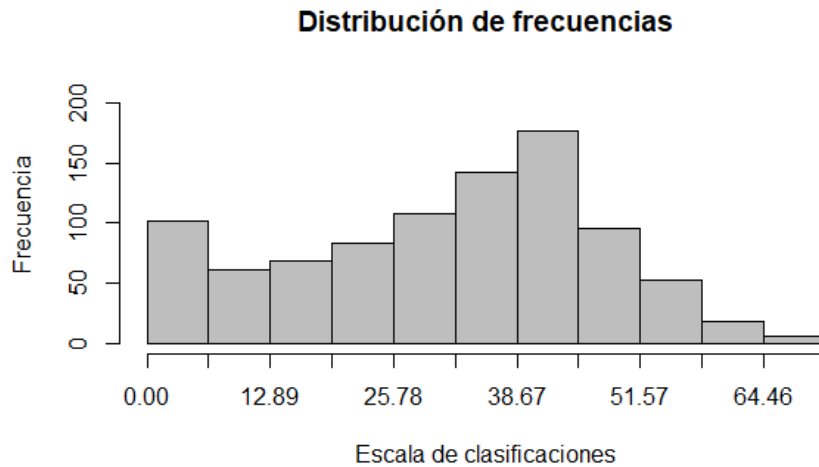


Figura 2.17: Histograma

2.2.4. Función de distribución

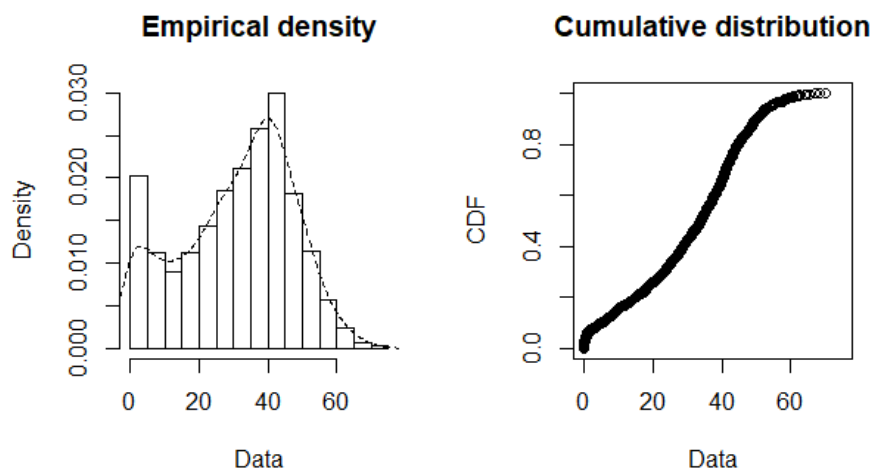


Figura 2.18: Gráfico de la función de distribución

2.2.5. Ajustes

El gráfico de Cullen and Frey coloca la observación cerca de la normal, aunque los valores de asimetría y curtosis no dicen lo mismo, debido seguramente a la cola tan grande que tenemos a la izquierda.

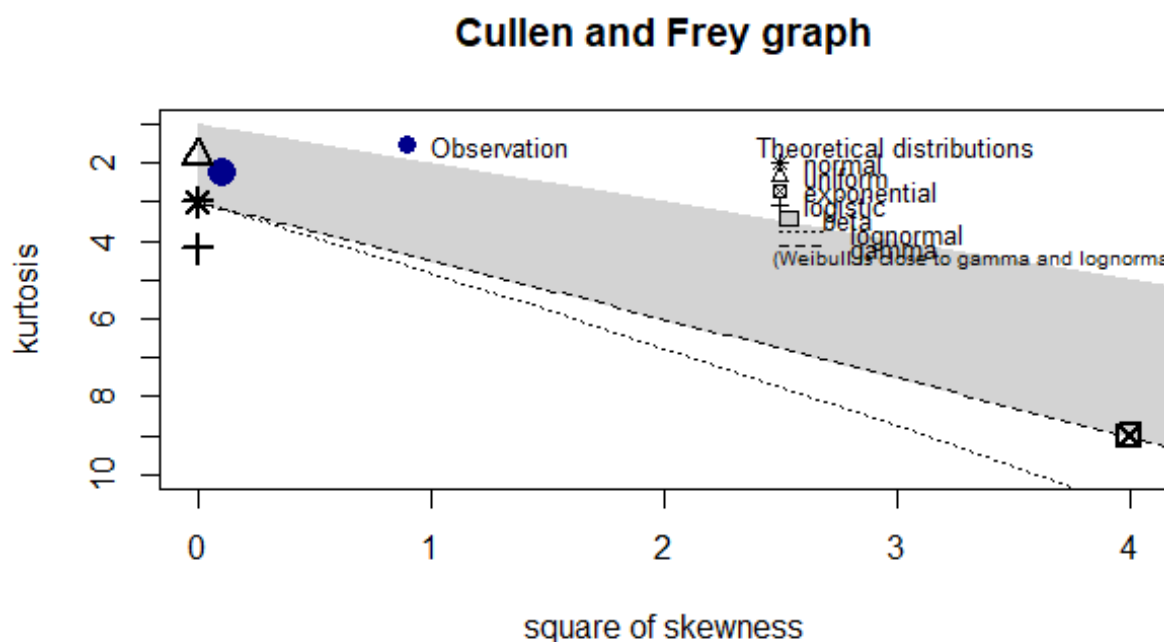


Figura 2.19: Gráfico Cullen and Frey

El valor de la asimetría y al curtosis ha sido de **-0,32** y **2,26** respectivamente.

Distribución Normal El ajuste nos dice que la variable no sigue una distribución normal.

```
Fitting of the distribution ' norm ' by maximum likelihood
Parameters :
  estimate Std. Error
mean 30.96503  0.5378721
sd   16.27008  0.3803330
Loglikelihood: -3850.563  AIC: 7705.127  BIC: 7714.765
Correlation matrix:
      mean sd
mean   1  0
sd     0  1
```

One-sample Kolmogorov-Smirnov test

```
data: tv
D = 0.077754, p-value = 3.135e-05
alternative hypothesis: two-sided
```

Figura 2.20: Resultados ajuste Normal

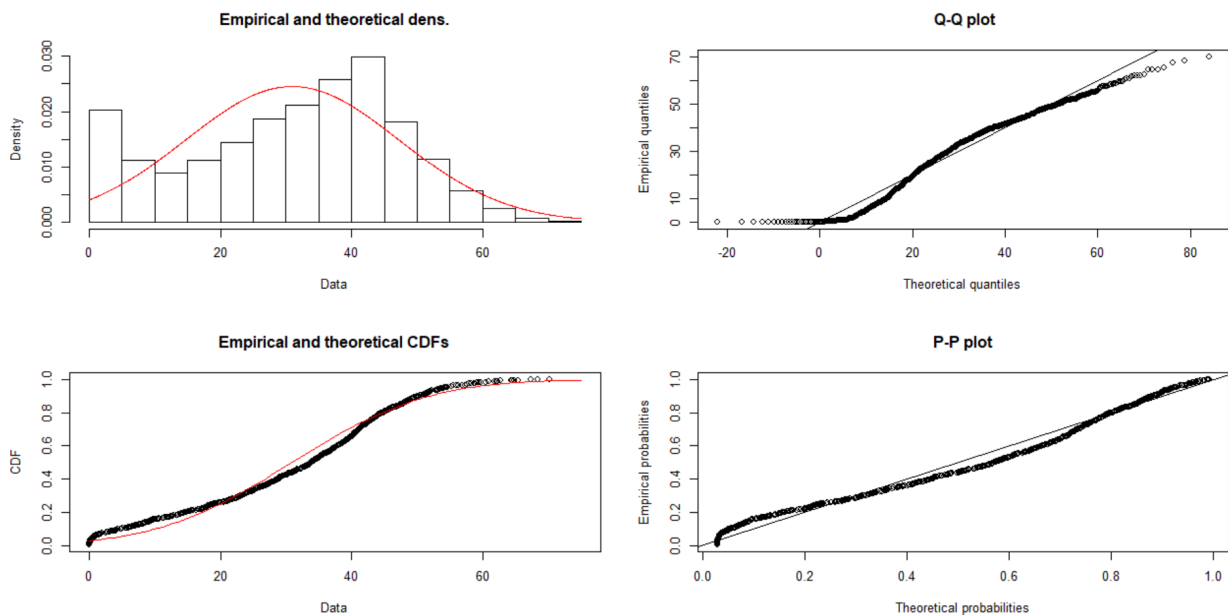


Figura 2.21: Gráfico de distribución, Q-Q plot, Empirical and theoretical CDFs y P-P plot

Distribución Logística Se ha comprobado que la variable no se ajusta a una distribución logística, aunque en este caso el p.valor obtenido no es tan bajo como los resultados anteriores.

```
Fitting of the distribution 'logis' by maximum likelihood
Parameters :
      estimate Std. Error
location 31.873452  0.5573881
scale    9.555285  0.2601539
Loglikelihood: -3875.457   AIC: 7754.915   BIC: 7764.553
Correlation matrix:
      location      scale
location 1.00000000 -0.06040767
scale   -0.06040767 1.00000000
```

One-sample Kolmogorov-Smirnov test

```
data: tv
D = 0.065992, p-value = 0.0006916
alternative hypothesis: two-sided
```

Figura 2.22: Resultados ajuste Logística

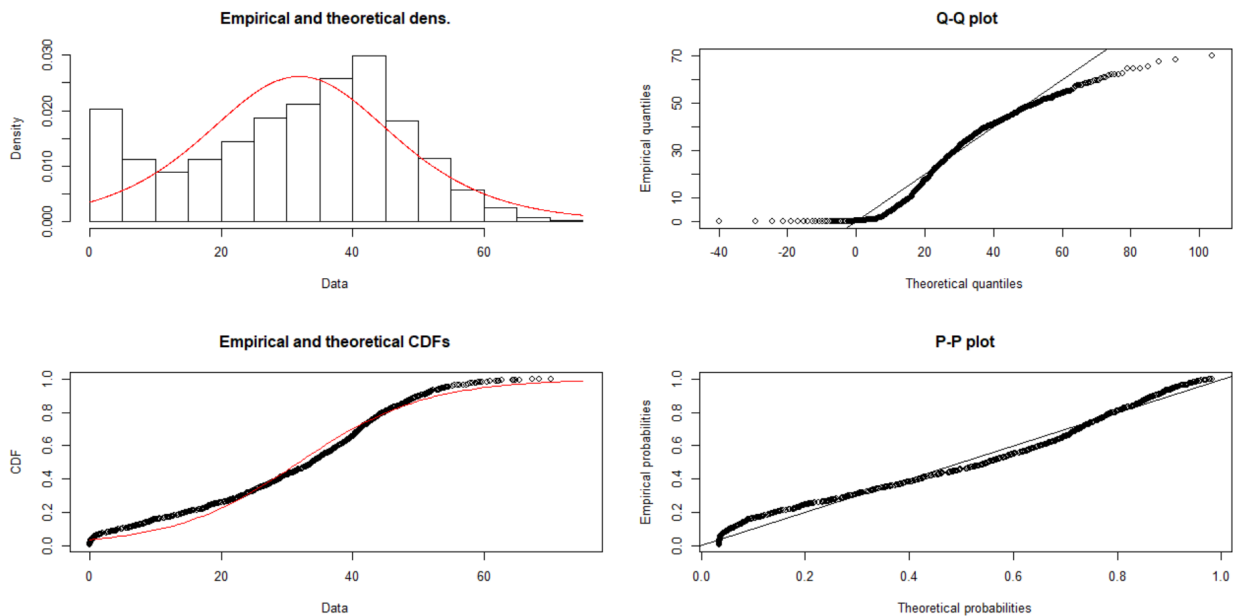


Figura 2.23: Gráfico de distribución, Q-Q plot, Empirical and theoretical CDFs y P-P plot

Para el resto de las distribuciones utilizadas, el resultado no va a ser nunca bueno debido a que tanto la exponencial como la gamma y la Gumbel se caracterizan por tener mas observaciones en los primeros intervalos.

En este caso el mayor numero de observaciones se encuentran en los intervalos medios por lo que en ningún caso se ajustarán a este tipo de distribuciones. No se han puesto en el informe con el fin de no hacerlo demasiado largo.

Tras realizar estos análisis, no se ha encontrado ninguna distribución que se ajuste a los datos ni para los años completos ni para la temporada alta.

2.3. TEMPORADA BAJA

2.3.1. Análisis descriptivo básico

Para la temporada baja (noviembre-abril), nos encontramos con 1272 observaciones.

Resumen

La media se encuentra en **5,25** a diferencia de la temporada alta que estaba en 30,97. Esto, junto con el primer cuartil que está muy cerca del cero, nos dice que las observaciones tienen valores muy bajos.

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
0	0,20	1,50	5,25	6,63	58,80	8,854942

2.3.2. Distribución de frecuencias

La tabla de frecuencias tiene 12 intervalos. Como vemos, la mayoría de los datos (casi el 84 %) se encuentran en los dos primeros intervalos, con valores menores de 9.

Class limits	f	rf	rf(%)	cf	cf(%)
[0,4.949)	879	0.69	69.10	879	69.10
[4.949,9.898)	188	0.15	14.78	1067	83.88
[9.898,14.85)	94	0.07	7.39	1161	91.27
[14.85,19.8)	30	0.02	2.36	1191	93.63
[19.8,24.74)	27	0.02	2.12	1218	95.75
[24.74,29.69)	7	0.01	0.55	1225	96.31
[29.69,34.64)	7	0.01	0.55	1232	96.86
[34.64,39.59)	14	0.01	1.10	1246	97.96
[39.59,44.54)	12	0.01	0.94	1258	98.90
[44.54,49.49)	8	0.01	0.63	1266	99.53
[49.49,54.44)	5	0.00	0.39	1271	99.92
[54.44,59.39)	1	0.00	0.08	1272	100.00

Figura 2.24: Distribución de frecuencias

2.3.3. Histograma

Aparentemente, la variable podría seguir una distribución exponencial, gamma o pareto.

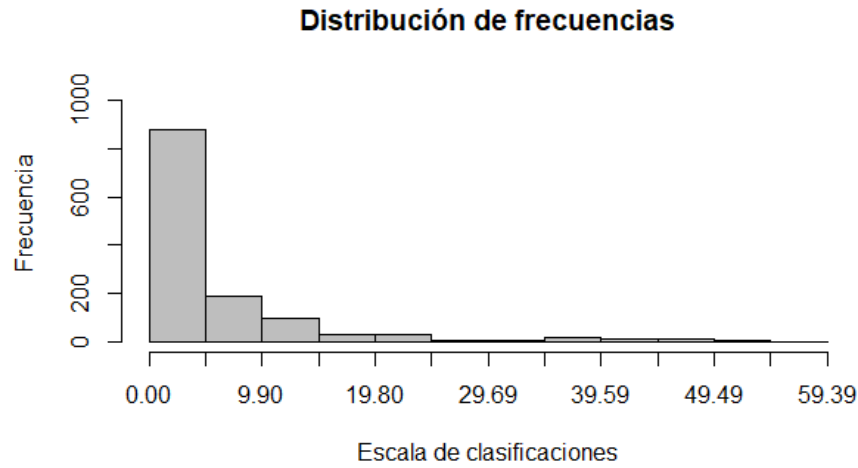


Figura 2.25: Histograma

2.3.4. Función de distribución

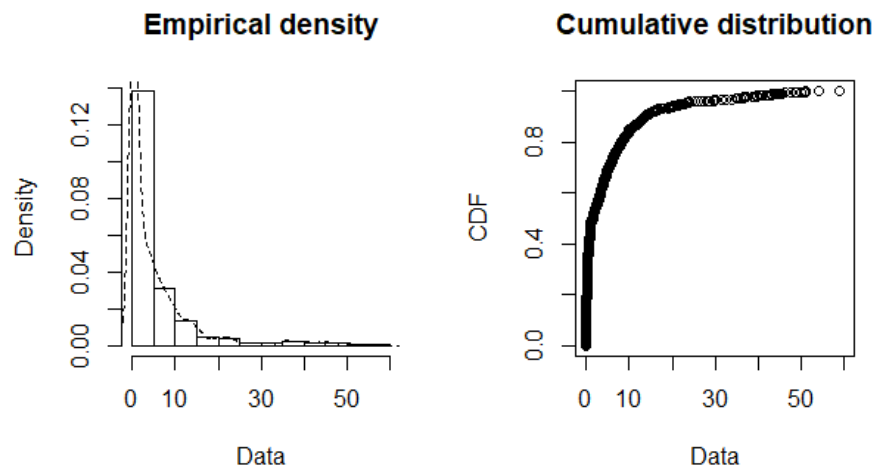


Figura 2.26: Gráfico de la función de distribución

2.3.5. Ajustes

El gráfico de Cullen and Frey no nos esclarece qué tipo de distribución puede seguir la variable. La asimetría es positiva, **2,93**.

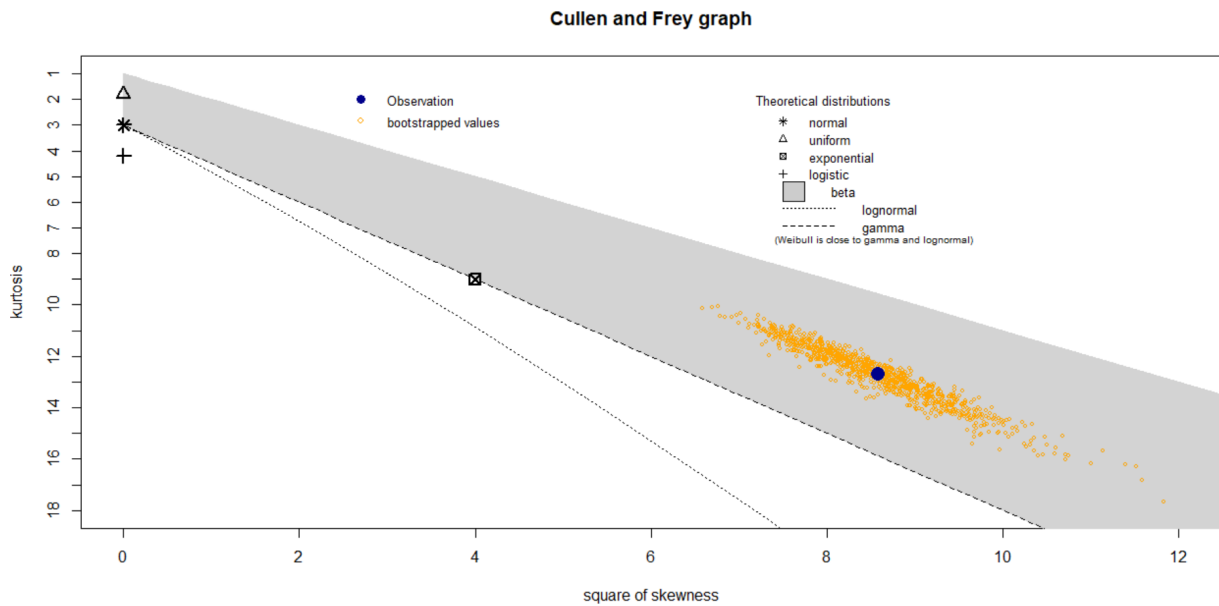


Figura 2.27: Gráfico Cullen and Frey

Debido a su forma, no vamos a probar con normal. Directamente vamos a probar con exponencial, gamma y pareto.

Distribución Exponencial No se ajusta a una distribución exponencial.

```
Fitting of the distribution ' exp ' by maximum likelihood
Parameters :
  estimate Std. Error
rate 0.1905218 0.00534182
Loglikelihood: -3380.961 AIC: 6763.923 BIC: 6769.071
```

One-sample Kolmogorov-Smirnov test

```
data: tv
D = 0.31177, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Figura 2.28: Resultados ajuste Exponencial

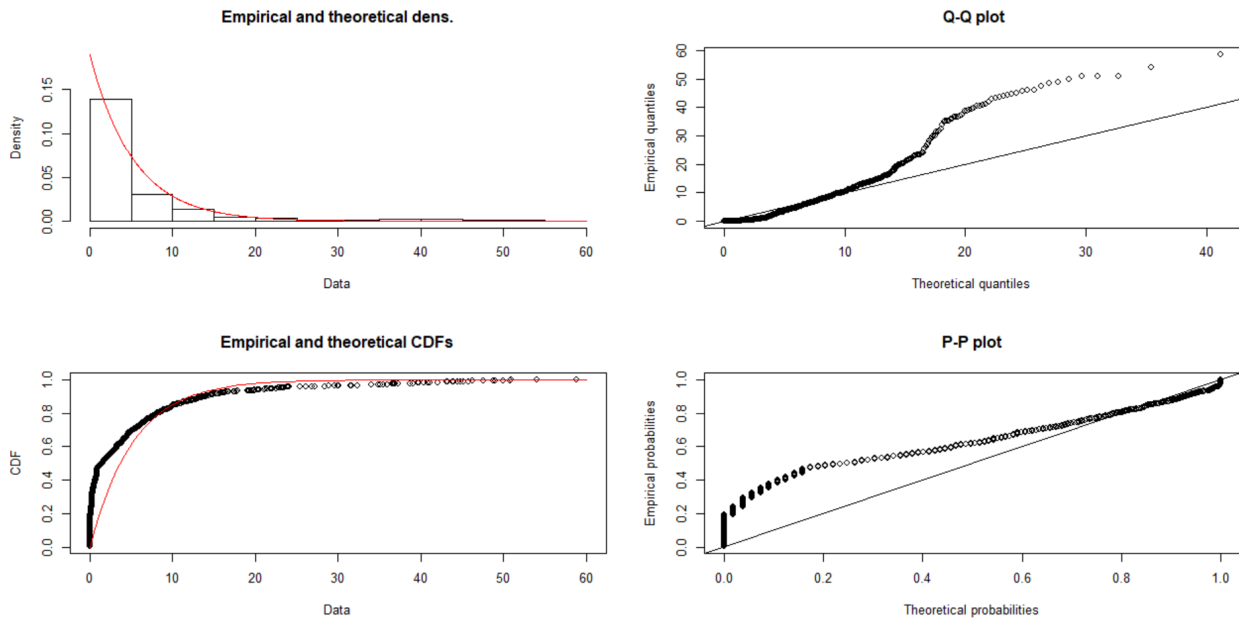


Figura 2.29: Gráfico de distribución, Q-Q plot, Empirical and theoretical CDFs y P-P plot

Distribución Gamma Mediante el test de bondad de ajuste obtenemos que no sigue una distribución gamma. Sin embargo, el problema puede ser parecido al ocurrido con la exponencial. La cola derecha afecta a la distribución. Al observar el grafico de empirical and theoretical CDFs vemos que parece ajustarse correctamente, incluso mejor que la exponencial. Por tanto, aunque no aceptemos que sigue una distribución gamma, podemos decir que es la que gráficamente más se asemeja.

```
Fitting of the distribution ' gamma ' by matching moments
Parameters :
  estimate
shape 0.35162565
rate 0.06699236
Loglikelihood: Inf AIC: -Inf BIC: -Inf
```

One-sample Kolmogorov-Smirnov test

```
data: tv
D = 0.7567, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Figura 2.30: Resultados ajuste Gamma

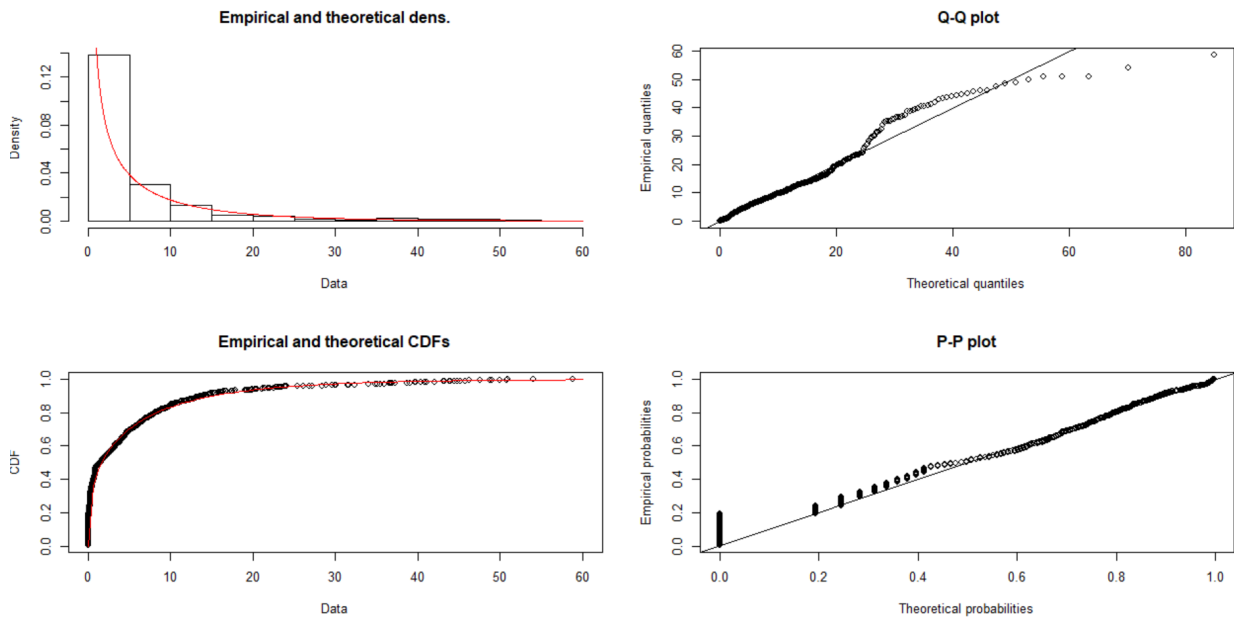


Figura 2.31: Gráfico de distribución, Q-Q plot, Empirical and theoretical CDFs y P-P plot

Distribución de Pareto Al igual que con la distribución Gamma, la distribución de Pareto no es aceptada. Sin embargo, gráficamente observamos que se ajusta relativamente bien.

```
Fitting of the distribution 'pareto' by maximum likelihood
Parameters :
  estimate Std. Error
shape 2.787761  0.291470
scale 11.408984  1.561406
Loglikelihood: -3520.619   AIC: 7045.238   BIC: 7055.534
Correlation matrix:
      shape      scale
shape 1.0000000  0.9633703
scale 0.9633703  1.0000000
```

One-sample Kolmogorov-Smirnov test

```
data: tv
D = 0.27857, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Figura 2.32: Resultados ajuste Pareto

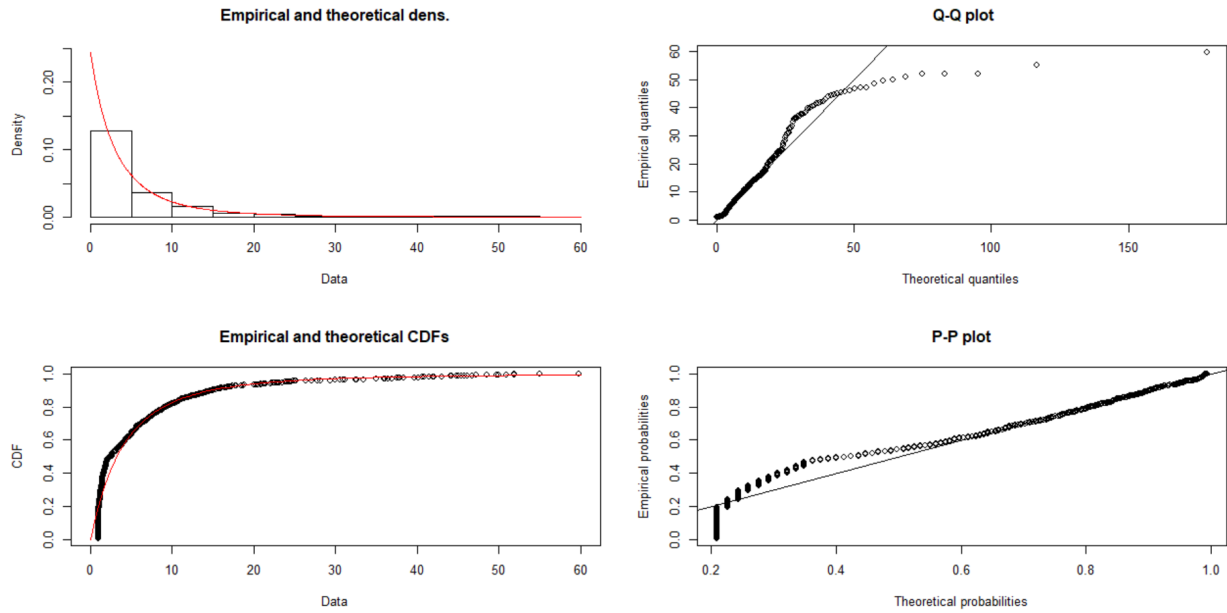


Figura 2.33: Gráfico de distribución, Q-Q plot, Empirical and theoretical CDFs y P-P plot

Por tanto, no se ha conseguido encontrar una distribución que se asemeje a nuestros datos. sin embargo, a pesar de no aceptar los tes de bondad de ajuste, si seguimos el criterio de AIC y BIC y los resultados gráficos, la distribución más parecida es la de Pareto.

Capítulo 3

SERIES TEMPORALES

Aplicar series temporales consiste en tener en cuenta el tiempo a la hora de analizar las variables. El hecho de que los datos se recogen diariamente puede hacer que exista autocorrelación en ellos. Esto es, los datos están relacionados con los datos de días anteriores. Para este caso, tenemos la siguiente serie:

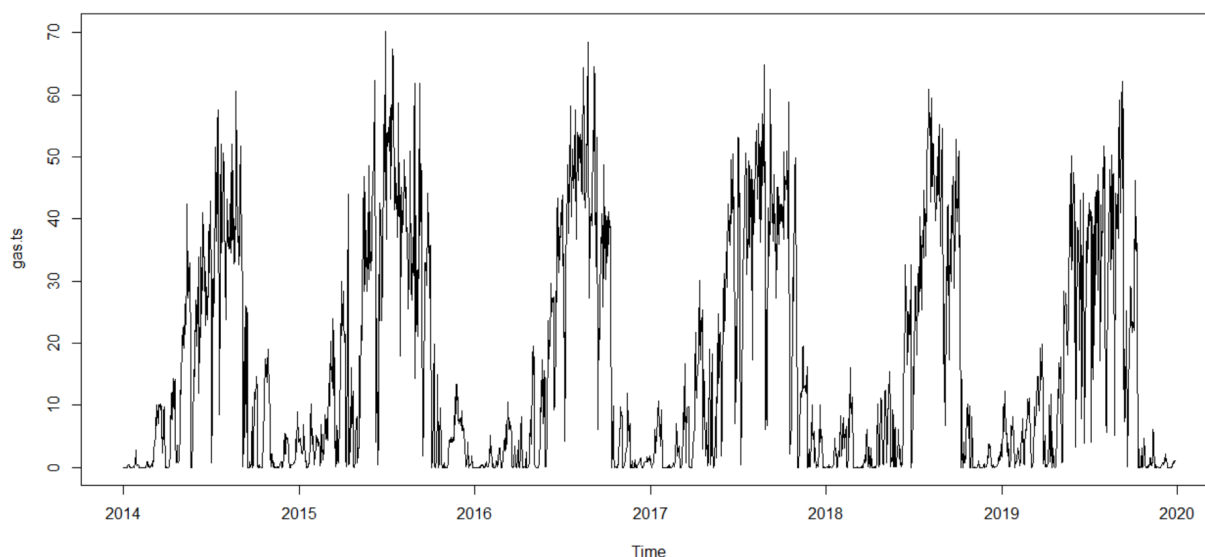


Figura 3.1: Serie

La función de autocorrelación simple y parcial nos muestra si existe autocorrelación en la serie o no. En ella, aparecen los retardos y el nivel de autocorrelación en cada retardo. Las bandas azules muestran los intervalos de confianza al 95 %. Si es superado significa que existe autocorrelación en la serie.

La función de autocorrelación simple mide la dependencia lineal entre un valor y el valor anterior.

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad \hat{\gamma}_k = \frac{1}{n} \sum_{t=k+1}^n \tilde{z}_t \tilde{z}_{t-k} \quad ; \tilde{z}_t = z_t - \bar{z} \quad , k = 1, 2, \dots$$

La función de autocorrelación parcial puede interpretarse como una medida de la dependencia lineal entre un valor y el valor anterior, tras descontar el efecto de los retardos intermedios.

$$\tilde{z}_t = \hat{\phi}_{k1}\tilde{z}_{t-1} + \hat{\phi}_{k2}\tilde{z}_{t-2} + \dots + \hat{\phi}_{kk}\tilde{z}_{t-k} \quad ; k = 1, 2, \dots$$

En nuestro caso obtenemos que existe claramente autocorrelación en la serie. El primer valor que aparece (la primera franja) siempre va a ser alta, por lo que no se debe tener en cuenta.

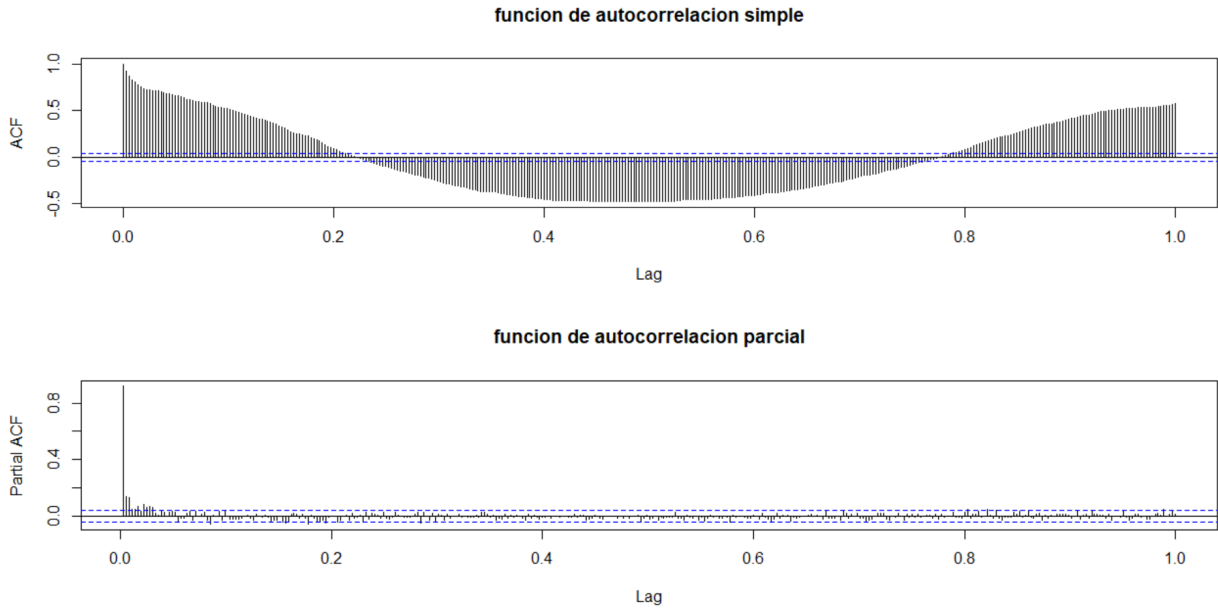


Figura 3.2: Función de autocorrelación simple y parcial

A la hora de estudiar la distribución es bueno quitar la tendencia de la serie con el objetivo de obtener residuos incorrelados.

La tendencia se elimina realizando una diferencia en la serie. Lo que se obtiene es una solución buena. El gráfico de la función de autocorrelación simple muestra que prácticamente todos los retardos están dentro de las bandas azules, lo que hace que podamos tener en cuenta la serie para trabajar. La función de autocorrelación parcial sí sigue mostrando algunos valores que superan las bandas, pero es una situación que se puede aceptar.

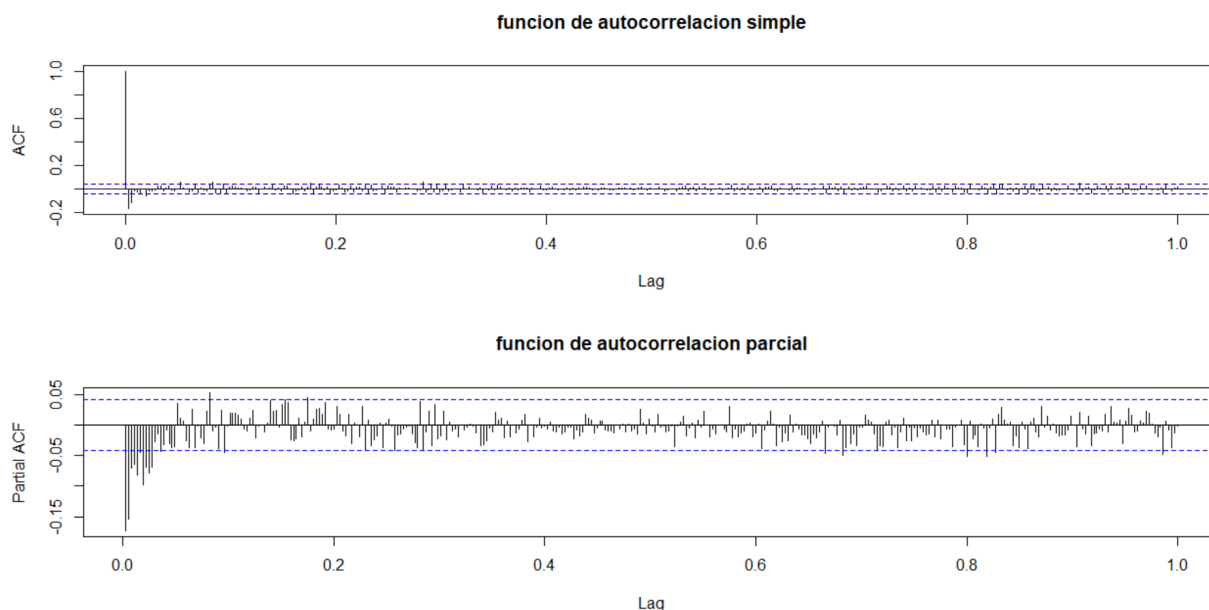


Figura 3.3: Función de autocorrelación simple y parcial tras primera diferencia

Por tanto, a partir de ahora trabajaremos con la serie donde se ha eliminado la tendencia. En este caso, sí se han obtenido resultados positivos en el caso de los datos relativos a la temporada alta.

3.1. AÑO COMPLETO

En este caso, los valores que obtenemos son muy diferentes a los originales, ya que estamos trabajando con los residuos aleatorios (parte no dependiente del pasado)

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
-51,80	-1,5	0,10	0,00046	2,50	29	7,10671

La distribución de frecuencias en este caso vuelve a parecer una normal.

Class limits	f	rf	rf(%)	cf	cf(%)
[-52,-46)	3	0.00	0.14	3	0.14
[-46,-40)	3	0.00	0.14	6	0.27
[-40,-33)	3	0.00	0.14	9	0.41
[-33,-27)	12	0.01	0.55	21	0.96
[-27,-21)	17	0.01	0.78	38	1.74
[-21,-15)	39	0.02	1.78	77	3.52
[-15,-8.4)	85	0.04	3.89	162	7.41
[-8.4,-2.1)	323	0.15	14.78	485	22.19
[-2.1,4.2)	1309	0.60	59.88	1794	82.07
[4.2,10)	285	0.13	13.04	2079	95.11
[10,17)	78	0.04	3.57	2157	98.67
[17,23)	22	0.01	1.01	2179	99.68
[23,29)	7	0.00	0.32	2186	100.00

Figura 3.4: Distribución de frecuencias

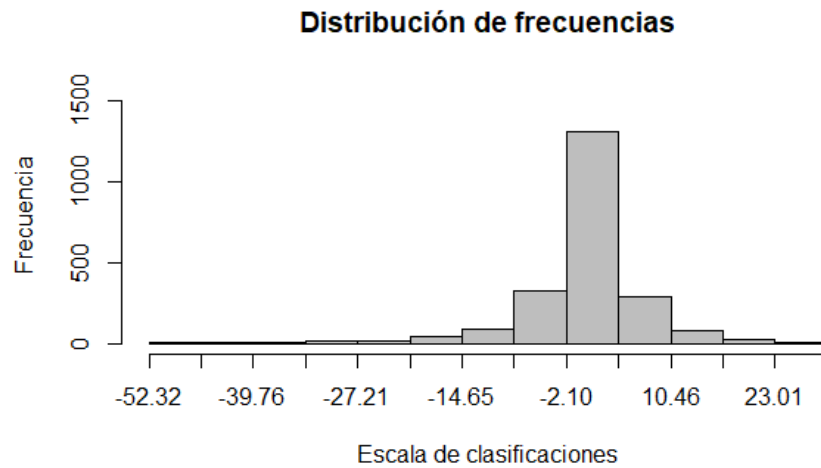


Figura 3.5: Histograma

La mayoría de los datos se encuentran concentrados en el centro, pero hay varios que están en las colas, provocando que estén muy extendidas.

3.1.1. Ajustes

Distribución Normal Tras realizar el ajuste y la prueba de bondad de ajuste, concluimos que no se ajusta a una distribución normal.

```
Fitting of the distribution ' norm ' by maximum likelihood
Parameters :
      estimate Std. Error
mean 0.0004574565  0.1519652
sd   7.1050843084  0.1074556
Loglikelihood: -7388.132  AIC: 14780.26  BIC: 14791.64
Correlation matrix:
      mean sd
mean   1  0
sd     0  1
```

One-sample Kolmogorov-Smirnov test

```
data: tt
D = 0.17739, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Figura 3.6: Resultados ajuste Normal

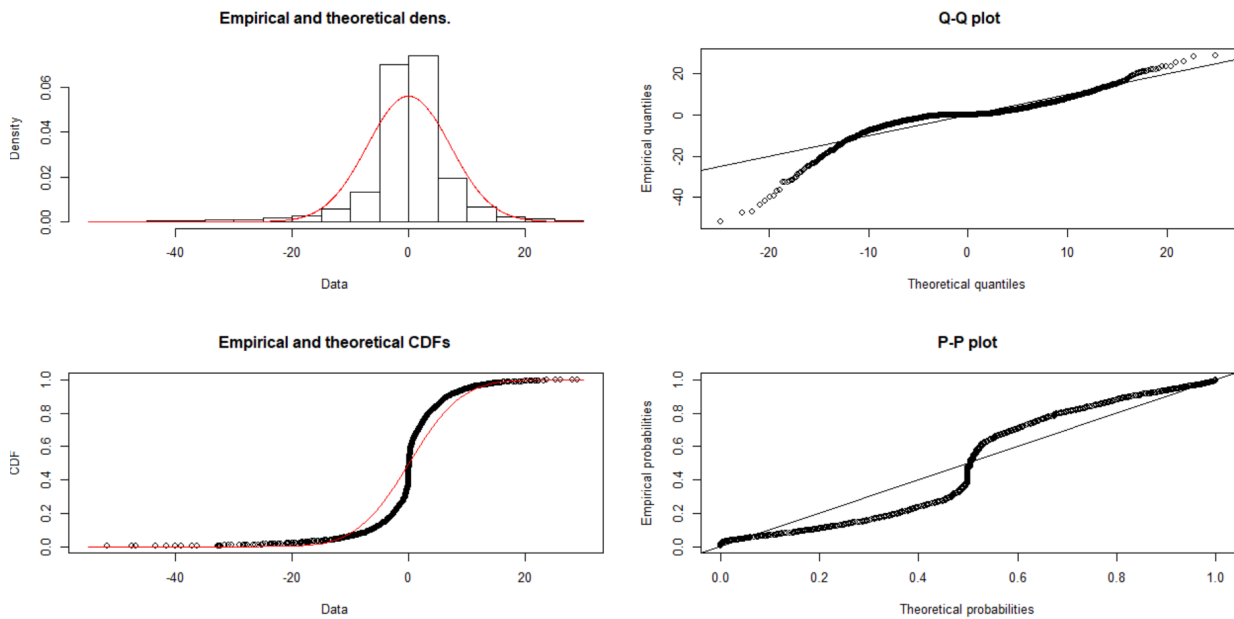


Figura 3.7: Gráfico de distribución, Q-Q plot, Empirical and theoretical CDFs y P-P plot

Distribución Logística No se ajusta a una distribución logística.

```
One-sample Kolmogorov-Smirnov test

data: tt
D = 0.13438, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Figura 3.8: Resultados ajuste Logística

3.2. TEMPORADA ALTA

El estudio de la temporada alta (recordemos, desde mayo a octubre) sí permitió conseguir resultados aceptables.

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
-47,60	-4,6	0,95	0,00044	5,28	29	9,920972

La forma de la distribución parece ser de nuevo normal o logística.

Class limits	f	rf	rf(%)	cf	cf(%)
[-48,-41)	3	0.00	0.33	3	0.33
[-41,-34)	4	0.00	0.44	7	0.77
[-34,-27)	10	0.01	1.09	17	1.86
[-27,-20)	20	0.02	2.19	37	4.05
[-20,-13)	42	0.05	4.60	79	8.64
[-13,-5.9)	115	0.13	12.58	194	21.23
[-5.9,1.2)	273	0.30	29.87	467	51.09
[1.2,8.2)	293	0.32	32.06	760	83.15
[8.2,15)	121	0.13	13.24	881	96.39
[15,22)	25	0.03	2.74	906	99.12
[22,29)	8	0.01	0.88	914	100.00

Figura 3.9: Distribución de frecuencias

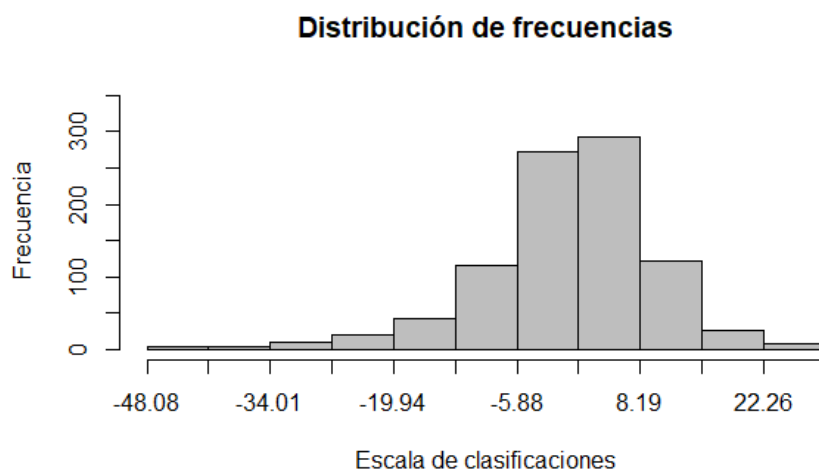


Figura 3.10: Histograma

3.2.1. Ajustes

Distribución Normal No se puede aceptar que los datos sigan una distribución normal.

```
Fitting of the distribution ' norm ' by maximum likelihood
Parameters :
      estimate Std. Error
mean 0.0004376368  0.3279770
sd   9.9155437855  0.2319148
Loglikelihood: -3393.721  AIC:  6791.441  BIC:  6801.077
Correlation matrix:
      mean sd
mean   1  0
sd     0  1
```

One-sample Kolmogorov-Smirnov test

```
data:  tt
D = 0.083651, p-value = 5.57e-06
alternative hypothesis: two-sided
```

Figura 3.11: Resultados ajuste Normal

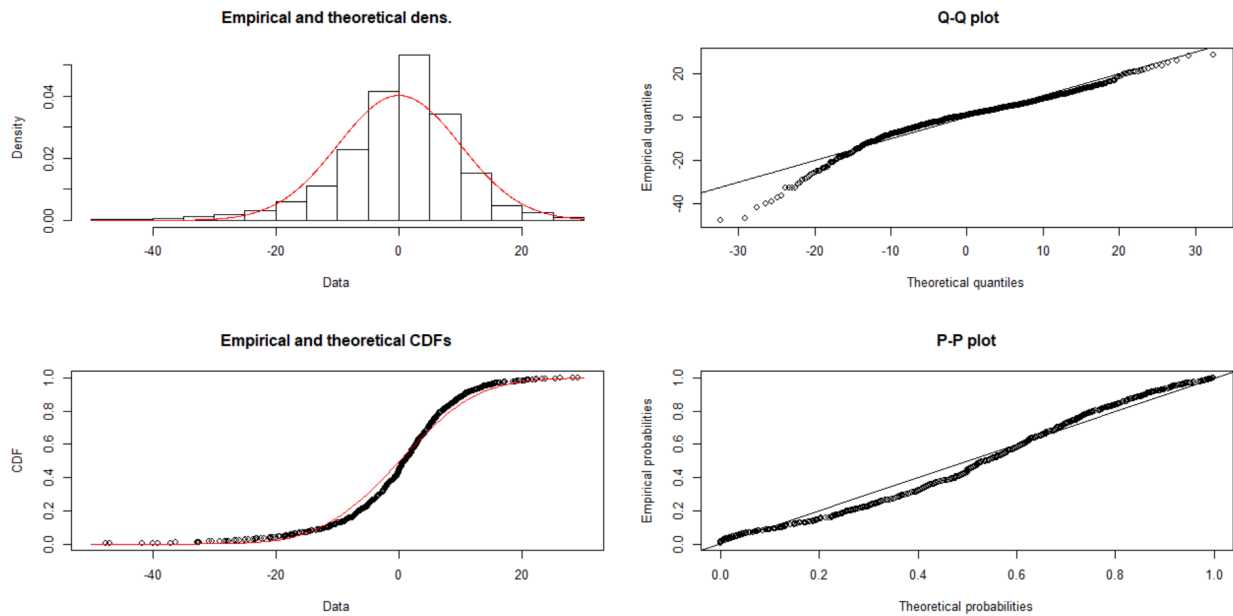


Figura 3.12: Gráfico de distribución, Q-Q plot, Empirical and theoretical CDFs y P-P plot

Distribución Logística SI podemos aceptar (no podemos rechazar) que la serie sin tendencia de los meses de temporada alta sigue una distribución logística.

```
location      scale
0.6343131    5.1900619
(0.2935272) (0.1456533)
```

One-sample Kolmogorov-Smirnov test

```
data: tt
D = 0.043668, p-value = 0.06126
alternative hypothesis: two-sided
```

Figura 3.13: Resultados ajuste Logística

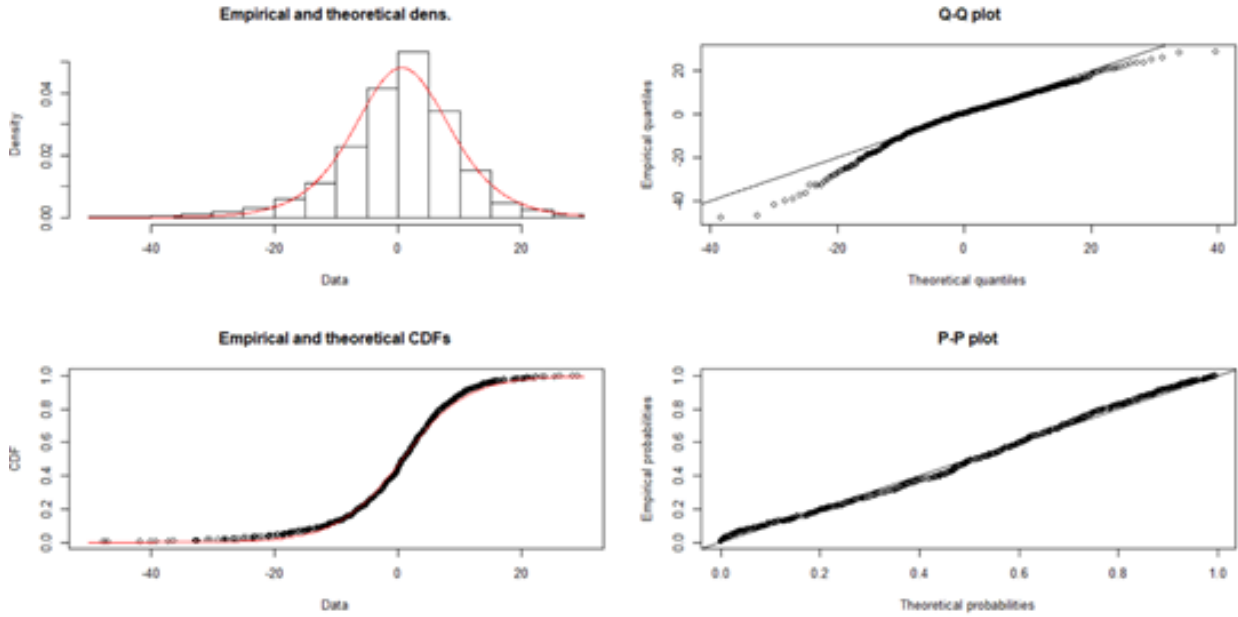


Figura 3.14: Gráfico de distribución, Q-Q plot, Empirical and theoretical CDFs y P-P plot

Conclusión Por tanto, no podemos rechazar la hipótesis de que la serie de temporada alta se asemeje a una distribución logística.

La distribución logística se caracteriza por tener dos parámetros, m (media) y $b > 0$ (desviación típica). Como ya hemos visto, a la media se la conoce por ser una medida de ubicación y a la desviación típica como medida de escala. Estos son los parámetros que se calculan y aparecen en el script como `location` y `scale`.

La función de distribución acumulada es:

$$F(x) = \frac{1}{2} \left[1 + \tanh\left(\frac{x - m}{2b}\right) \right] \quad (3.1)$$

La función de probabilidad de distribución es:

$$f(x) = \frac{1}{4b} \operatorname{sech}^2\left(\frac{x - m}{2b}\right) \quad (3.2)$$

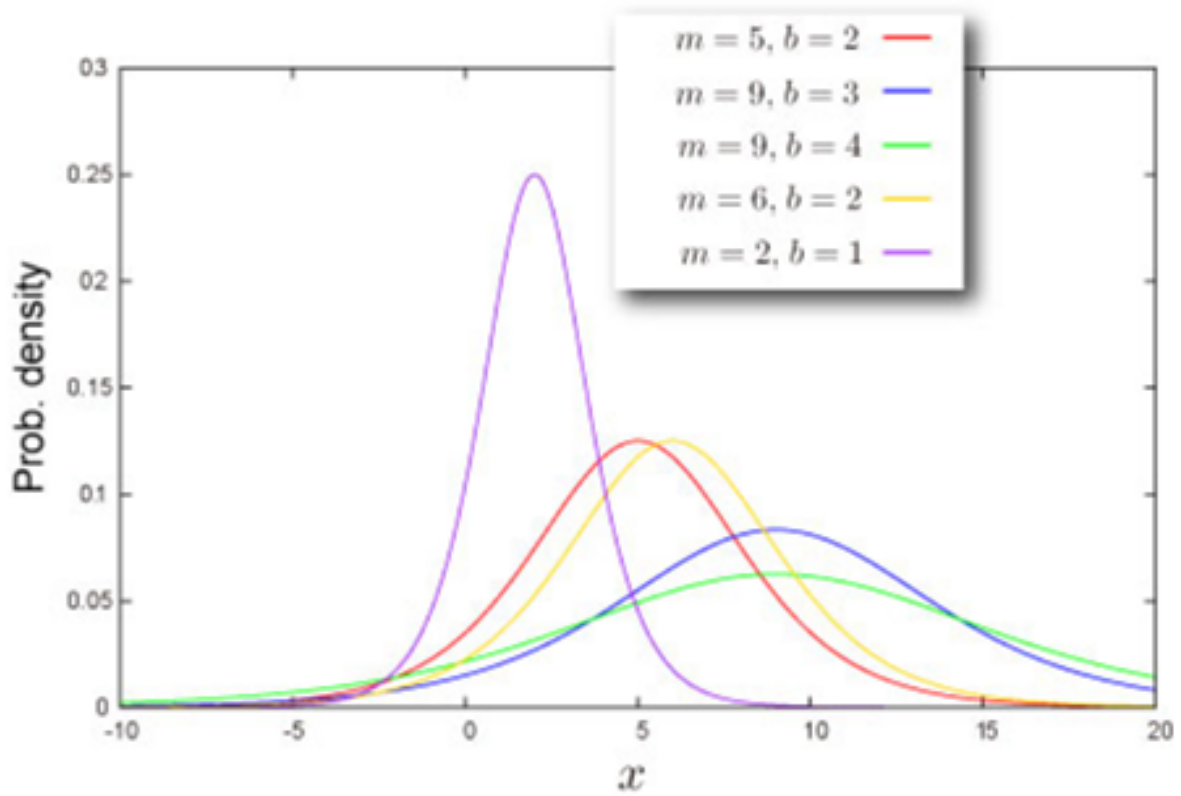


Figura 3.15: Gráfico de CDF Teórico

La función de distribución inversa (función cuantil) es:

$$F^{-1}(P) = 2btanh^{-1}(2P - 1) + m \quad (3.3)$$

La media es m

La desviación estándar es:

$$\sqrt{\frac{\pi^2 b^2}{3}} \quad (3.4)$$

Ejemplos en los que se usa esta distribución son en el estudio de la propagación de epidemias o en el estudio de la difusión y sustitución de energía primaria.

3.3. TEMPORADA BAJA

La temporada baja, compuesta por los meses de noviembre a abril se caracteriza por los siguientes estadísticos:

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
-51,80	-0,50	0,00	0,00079	0,8	41,50	4,573207

La forma de la distribución parece ser de nuevo normal o logística.

Class limits	f	rf	rf(%)	cf	cf(%)
[-52.32,-44.47)	1	0.00	0.08	1	0.08
[-44.47,-36.61)	1	0.00	0.08	2	0.16
[-36.61,-28.76)	3	0.00	0.24	5	0.39
[-28.76,-20.91)	3	0.00	0.24	8	0.63
[-20.91,-13.05)	7	0.01	0.55	15	1.18
[-13.05,-5.201)	48	0.04	3.78	63	4.96
[-5.201,2.651)	1057	0.83	83.16	1120	88.12
[2.651,10.5)	138	0.11	10.86	1258	98.98
[10.5,18.36)	8	0.01	0.63	1266	99.61
[18.36,26.21)	3	0.00	0.24	1269	99.84
[26.21,34.06)	1	0.00	0.08	1270	99.92
[34.06,41.91)	1	0.00	0.08	1271	100.00

Figura 3.16: Distribución de frecuencias

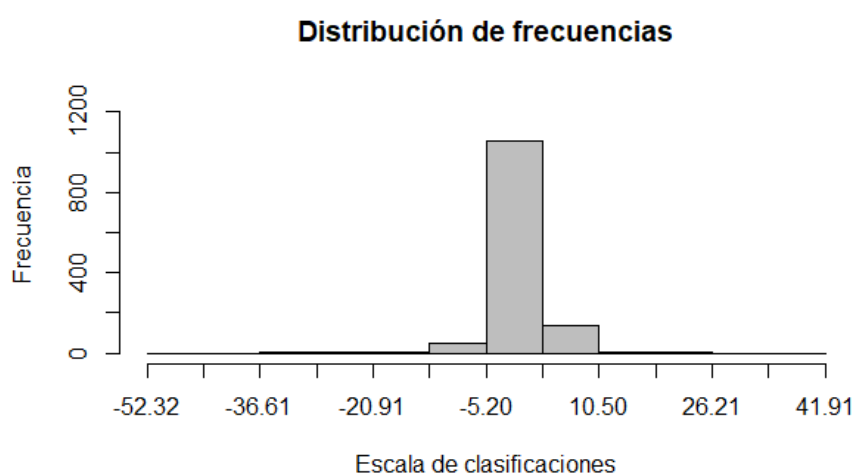


Figura 3.17: Histograma

3.3.1. Ajustes

Distribución Normal No se ajusta a una distribución Normal.

```
Fitting of the distribution ' norm ' by maximum likelihood
Parameters :
      estimate Std. Error
mean 0.0007867821 0.12822631
sd   4.5714073238 0.09066967
Loglikelihood: -3735.164   AIC: 7474.327   BIC: 7484.622
Correlation matrix:
      mean sd
mean   1  0
sd     0  1
```

One-sample Kolmogorov-Smirnov test

```
data: tt
D = 0.23063, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Figura 3.18: Resultados ajuste Normal

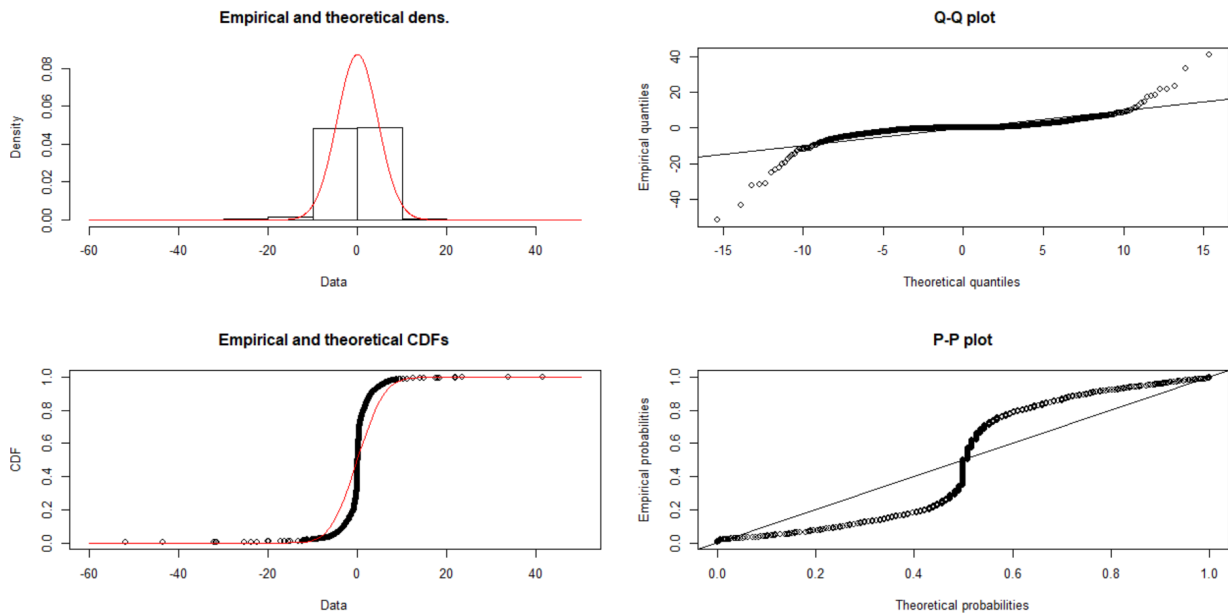


Figura 3.19: Gráfico de distribución, Q-Q plot, Empirical and theoretical CDFs y P-P plot

Distribución Logística No se ajusta a una distribución logística.

```

location      scale
0.13567655   1.65287408
(0.07482107) (0.04156708)

```

One-sample Kolmogorov-Smirnov test

```

data: tt
D = 0.17012, p-value < 2.2e-16
alternative hypothesis: two-sided

```

Figura 3.20: Resultados ajuste Logística

No se ha conseguido encontrar una distribución que se ajuste a nuestros datos.

Capítulo 4

OTRAS LOCALIZACIONES

El análisis de la variable FWI también se ha realizado para otras localizaciones de la Península.

4.1. Santa Elena

4.1.1. Año Completo

Para el caso de Santa Elena (Jaén) se obtienen los siguientes resultados:

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
0	1,65	13,30	21,25	38,55	80	21,1654

Los valores se concentran en los intervalos bajos. Mas del 50 % de los datos son inferiores a 19. También, existen pequeños incrementos en los intervalos medios, sintoma de la necesidad de separar por temporadas.

Class limits	f	rf	rf(%)	cf	cf(%)
[0,6.2)	853	0.39	39.00	853	39.00
[6.2,12)	217	0.10	9.92	1070	48.93
[12,19)	138	0.06	6.31	1208	55.24
[19,25)	130	0.06	5.94	1338	61.18
[25,31)	120	0.05	5.49	1458	66.67
[31,37)	162	0.07	7.41	1620	74.07
[37,44)	140	0.06	6.40	1760	80.48
[44,50)	139	0.06	6.36	1899	86.83
[50,56)	111	0.05	5.08	2010	91.91
[56,62)	81	0.04	3.70	2091	95.61
[62,68)	65	0.03	2.97	2156	98.58
[68,75)	23	0.01	1.05	2179	99.63
[75,81)	8	0.00	0.37	2187	100.00

Figura 4.1: Distribución de frecuencias

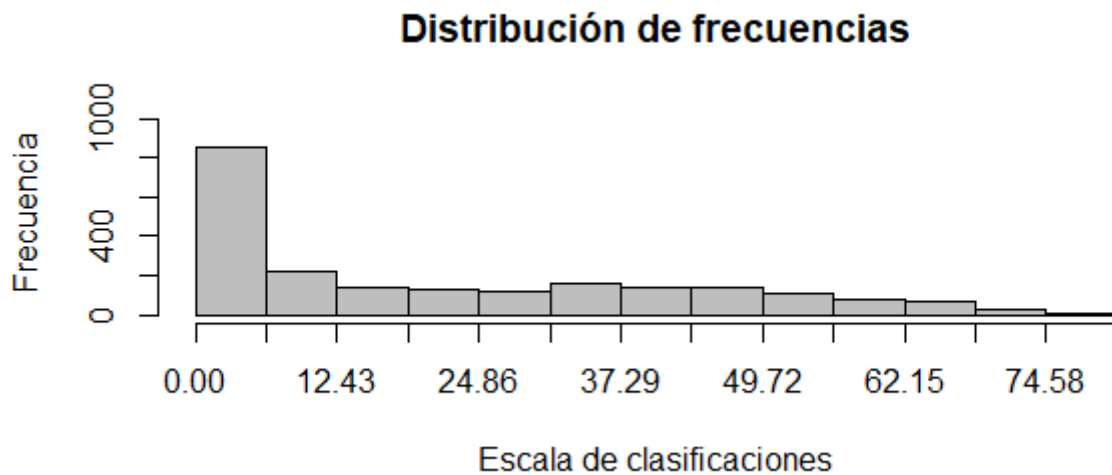


Figura 4.2: Histograma

Al realizar los ajustes correspondientes no se ha encontrado ninguna distribución que se asemeje a nuestros datos.

Distribución	P-value KS
Exponencial	$<2.2e-16$
Gamma	$<2.2e-16$
Pareto	$<2.2e-16$
Gumbel	$<2.2e-16$

4.1.2. Temporada alta

En el caso de la temporada alta, tenemos los siguiente resultados:

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
0	22,12	38,00	35,80	50,77	80	19,94282

Los datos se encuentran su mayoría en los intervalos medios. Sin embargo, se observan muchos valores en el primer intervalo.

Class limits	f	rf	rf(%)	cf	cf(%)
[0,6.7)	151	0.14	13.75	151	13.75
[6.7,13)	58	0.05	5.28	209	19.03
[13,20)	42	0.04	3.83	251	22.86
[20,27)	74	0.07	6.74	325	29.60
[27,34)	117	0.11	10.66	442	40.26
[34,40)	156	0.14	14.21	598	54.46
[40,47)	157	0.14	14.30	755	68.76
[47,54)	132	0.12	12.02	887	80.78
[54,61)	95	0.09	8.65	982	89.44
[61,67)	82	0.07	7.47	1064	96.90
[67,74)	25	0.02	2.28	1089	99.18
[74,81)	9	0.01	0.82	1098	100.00

Figura 4.3: Distribución de frecuencias

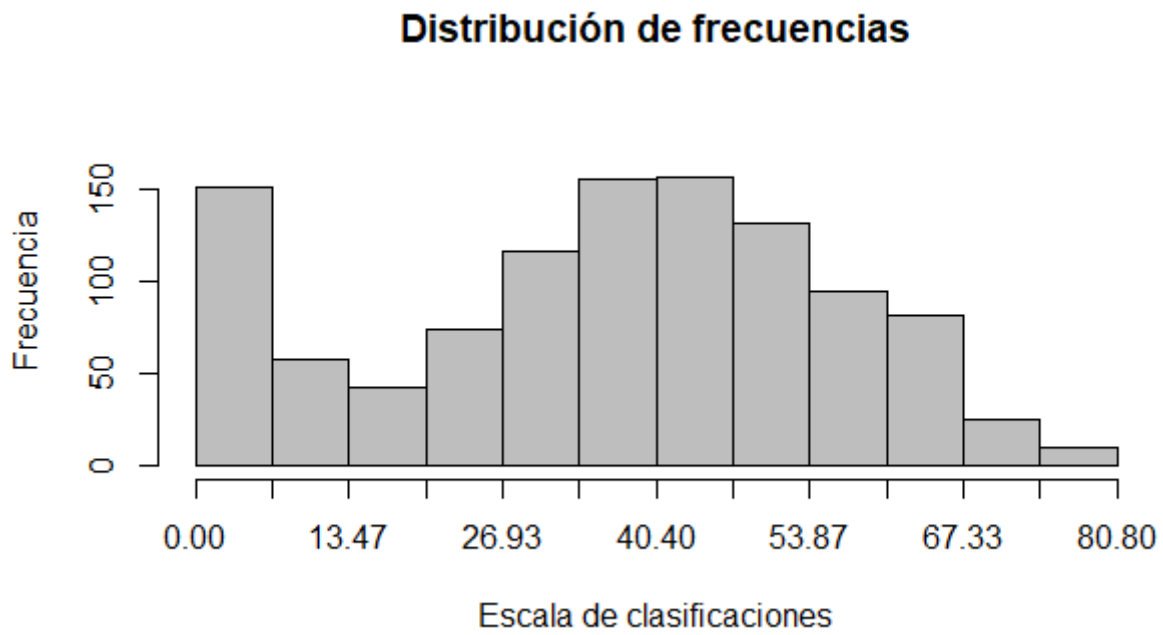


Figura 4.4: Histograma

Con el gráfico de Cullen and Frey observamos que se acerca a una normal. La asimetría es de -0,27 y la curtosis de 2,19.

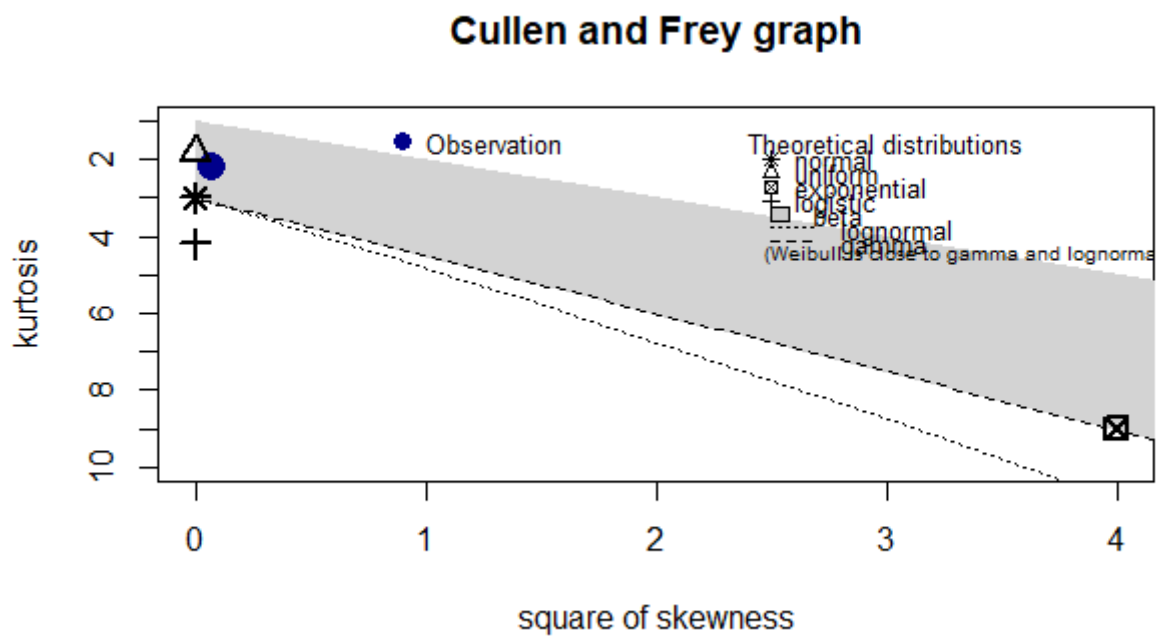


Figura 4.5: Gráfico de Cullen&Frey

Tanto la distribución normal como la logística no son aceptadas.

Distribución	P-value KS
Normal	5.777e-05
Logística	1.13e-05

4.1.3. Temporada baja

Los resultados son los siguientes:

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
0	0,50	3,60	7,39	12	44,80	8,754802

El 52,26 % de los datos se concentran en el primer intervalo.

Class limits	f	rf	rf(%)	cf	cf(%)
[0,4.113)	475	0.52	52.26	475	52.26
[4.113,8.227)	117	0.13	12.87	592	65.13
[8.227,12.34)	97	0.11	10.67	689	75.80
[12.34,16.45)	72	0.08	7.92	761	83.72
[16.45,20.57)	55	0.06	6.05	816	89.77
[20.57,24.68)	41	0.05	4.51	857	94.28
[24.68,28.79)	22	0.02	2.42	879	96.70
[28.79,32.91)	18	0.02	1.98	897	98.68
[32.91,37.02)	5	0.01	0.55	902	99.23
[37.02,41.13)	5	0.01	0.55	907	99.78
[41.13,45.25)	2	0.00	0.22	909	100.00

Figura 4.6: Distribución de frecuencias

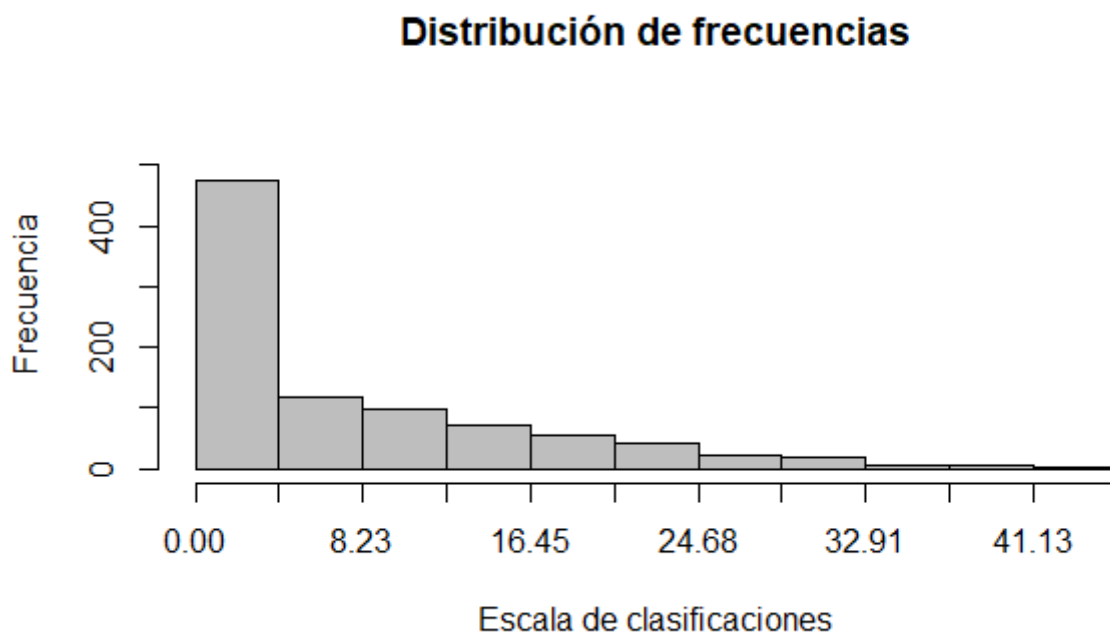


Figura 4.7: Histograma

El valor de asimetría es de 1,38.

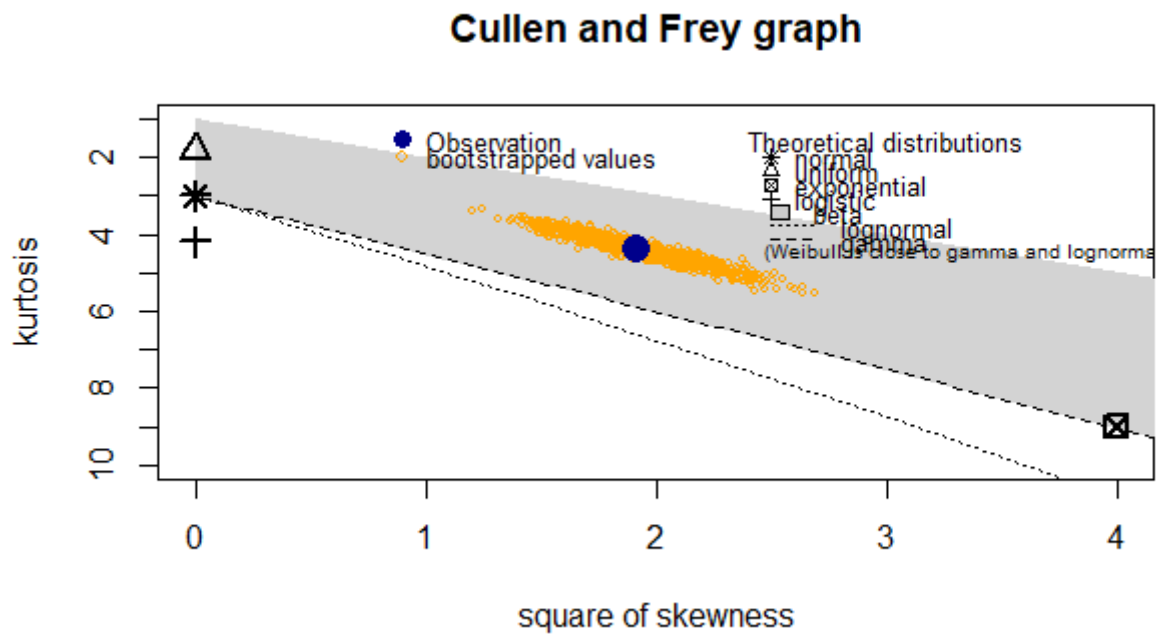


Figura 4.8: Gráfico de Cullen&Frey

Finalmente, obtenemos que ninguna distribución de las utilizadas se ajusta.

Distribución	P-value KS
Exponencial	<2.2e-16
Gamma	<2.2e-16
Pareto	<2.2e-16
Gumbel	9.28273e-236

4.2. Alhama de Murcia

4.2.1. Año Completo

Para el caso de Alhama de Murcia (Murcia) se obtienen los siguientes resultados:

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
0	8,20	17,60	19,12	29,20	76	13,23197

Los valores parecen concentrarse en bajos. Más del 50 % de los datos son inferiores a 18, aunque la distribución tiene una forma escalonada.

Class limits	f	rf	rf(%)	cf	cf(%)
[0,5.9)	436	0.20	19.94	436	19.94
[5.9,12)	321	0.15	14.68	757	34.61
[12,18)	357	0.16	16.32	1114	50.94
[18,24)	279	0.13	12.76	1393	63.69
[24,30)	261	0.12	11.93	1654	75.63
[30,35)	246	0.11	11.25	1900	86.88
[35,41)	160	0.07	7.32	2060	94.19
[41,47)	87	0.04	3.98	2147	98.17
[47,53)	27	0.01	1.23	2174	99.41
[53,59)	10	0.00	0.46	2184	99.86
[59,65)	2	0.00	0.09	2186	99.95
[65,71)	0	0.00	0.00	2186	99.95
[71,77)	1	0.00	0.05	2187	100.00

Figura 4.9: Distribución de frecuencias

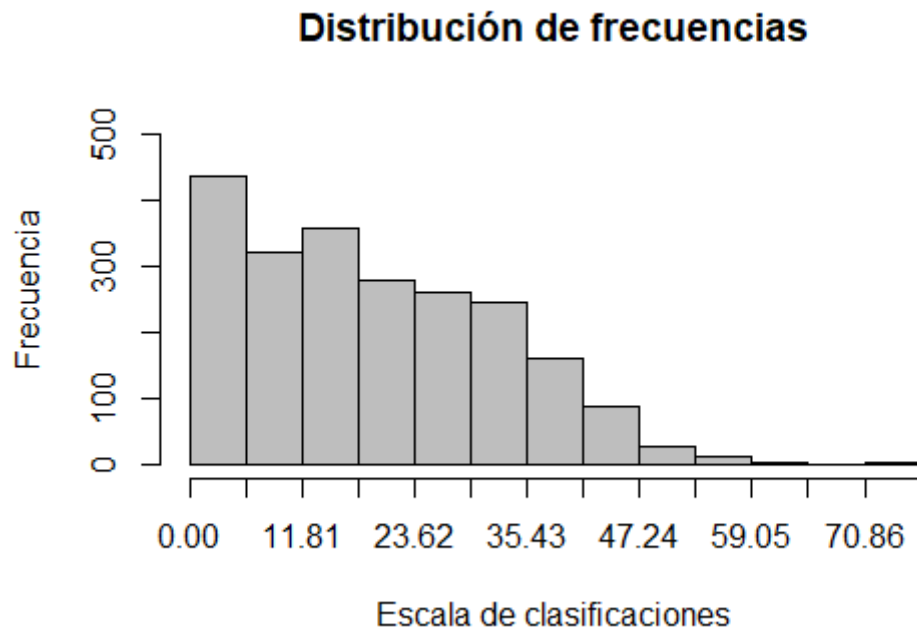


Figura 4.10: Histograma

La asimetría es de 0,45.

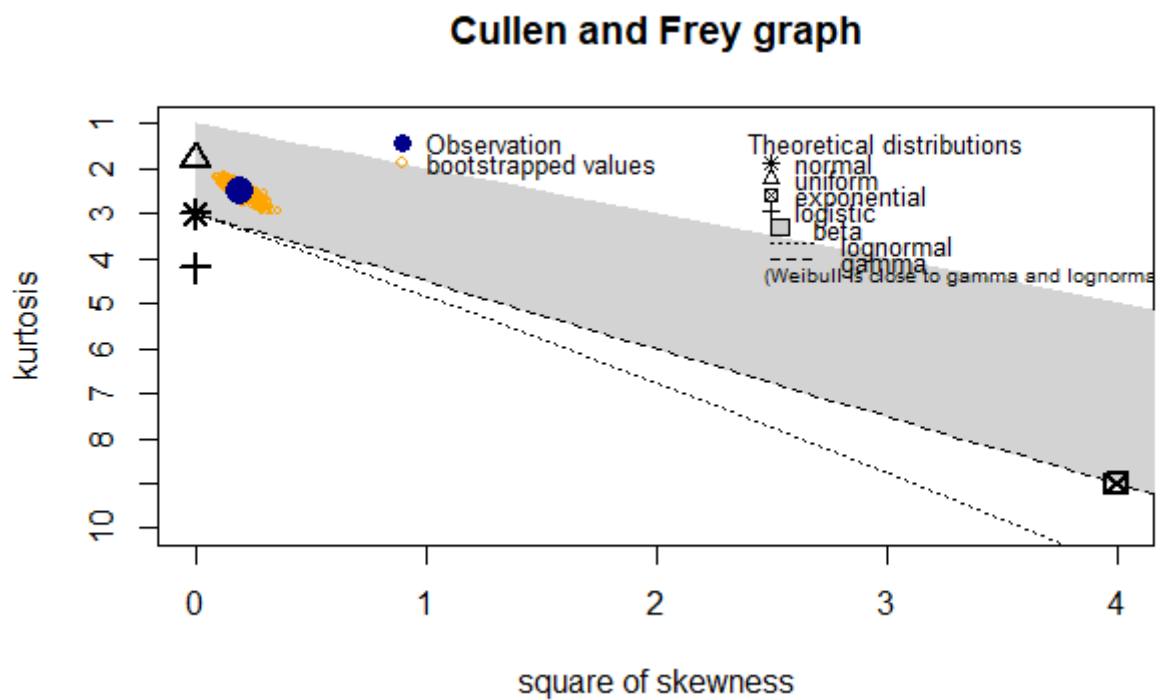


Figura 4.11: Gráfico de Cullen&Frey

No se acepta ninguna distribución.

Distribución	P-value KS
Normal	7.076e-11
Logística	1.987e-14
Exponencial	<2.2e-16
Gamma	<2.2e-16
Pareto	<2.2e-16
Gumbel	2.664705e-68

4.2.2. Temporada alta

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
0	15,32	26,75	25,12	34,90	76	13,25927

Los datos se encuentran su mayoría en los intervalos medios. Sin embargo, se observan muchos valores en el primer intervalo.

Class limits	f	rf	rf(%)	cf	cf(%)
[0,5.399)	85	0.09	8.84	85	8.84
[5.399,10.8)	61	0.06	6.35	146	15.19
[10.8,16.2)	78	0.08	8.12	224	23.31
[16.2,21.6)	86	0.09	8.95	310	32.26
[21.6,26.99)	134	0.14	13.94	444	46.20
[26.99,32.39)	173	0.18	18.00	617	64.20
[32.39,37.79)	156	0.16	16.23	773	80.44
[37.79,43.19)	112	0.12	11.65	885	92.09
[43.19,48.59)	50	0.05	5.20	935	97.29
[48.59,53.99)	18	0.02	1.87	953	99.17
[53.99,59.39)	8	0.01	0.83	961	100.00

Figura 4.12: Distribución de frecuencias

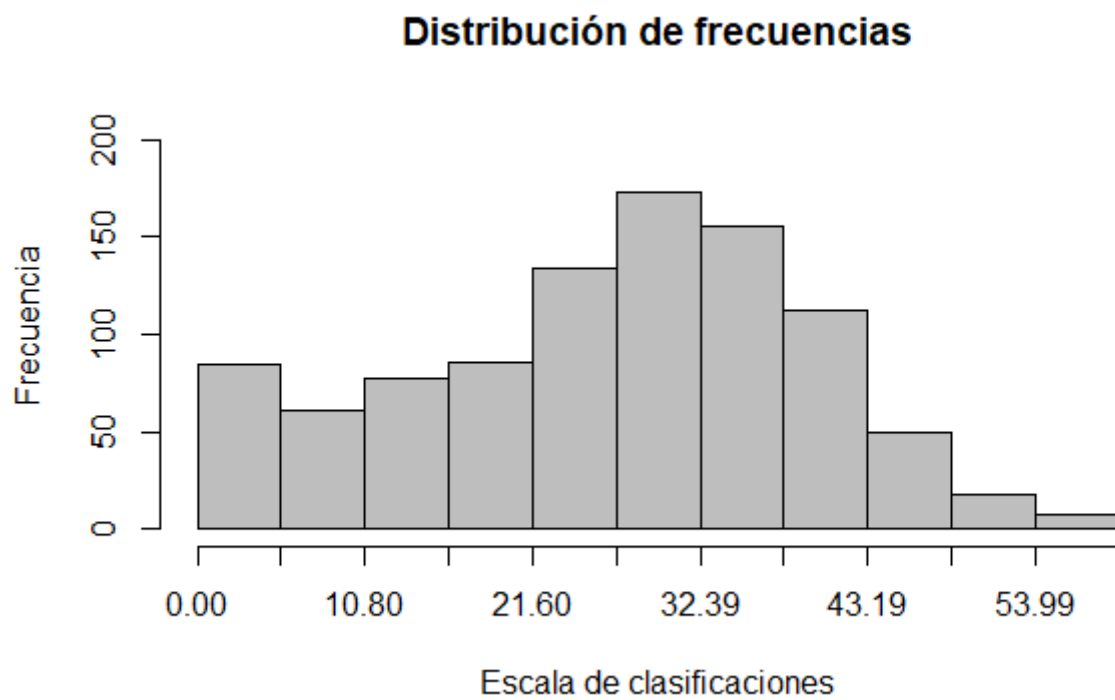


Figura 4.13: Histograma

La asimetría es de -0,12 y la curtosis es 2,4.

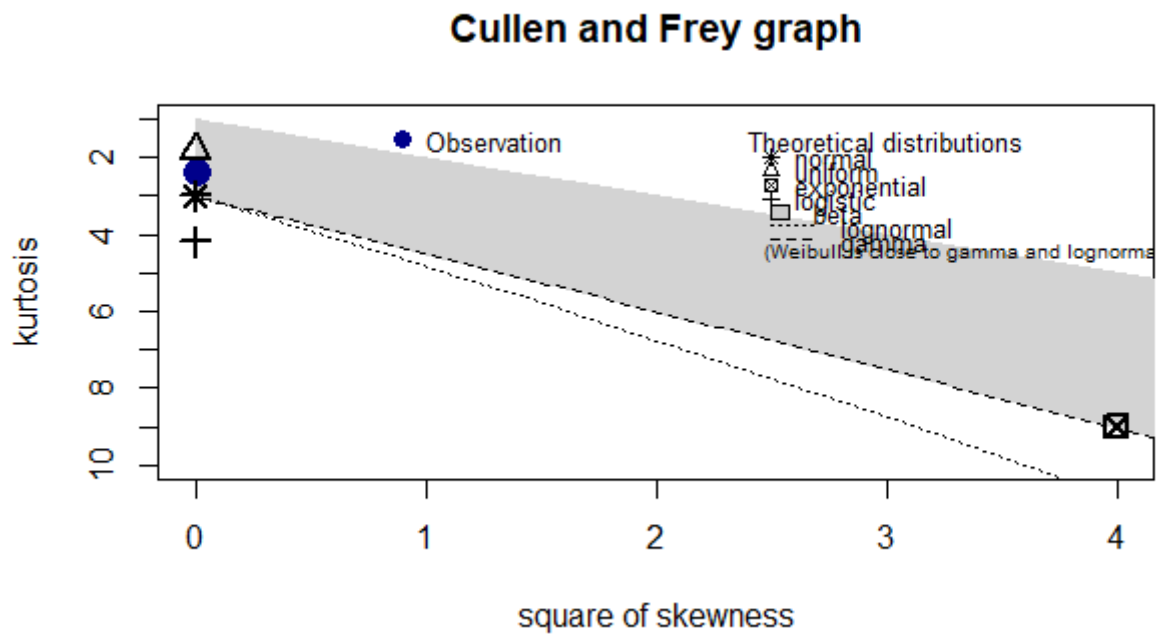


Figura 4.14: Gráfico de Cullen&Frey

Sin llegar a aceptarse el test de bondad de ajuste para normal y para logística, ésta última se acerca más a la distribución de nuestros datos.

Distribución	P-value KS
Normal	0.0009425
Logística	0.003479

4.2.3. Temporada baja

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
0	5,00	11,80	13,04	19,20	63,10	10,06331

Los datos se asemejan a una exponencial.

Class limits	f	rf	rf(%)	cf	cf(%)
[0,5.311)	296	0.26	26.45	296	26.45
[5.311,10.62)	212	0.19	18.95	508	45.40
[10.62,15.93)	231	0.21	20.64	739	66.04
[15.93,21.24)	162	0.14	14.48	901	80.52
[21.24,26.55)	113	0.10	10.10	1014	90.62
[26.55,31.87)	56	0.05	5.00	1070	95.62
[31.87,37.18)	26	0.02	2.32	1096	97.94
[37.18,42.49)	11	0.01	0.98	1107	98.93
[42.49,47.8)	4	0.00	0.36	1111	99.29
[47.8,53.11)	5	0.00	0.45	1116	99.73
[53.11,58.42)	1	0.00	0.09	1117	99.82
[58.42,63.73)	2	0.00	0.18	1119	100.00

Figura 4.15: Distribución de frecuencias

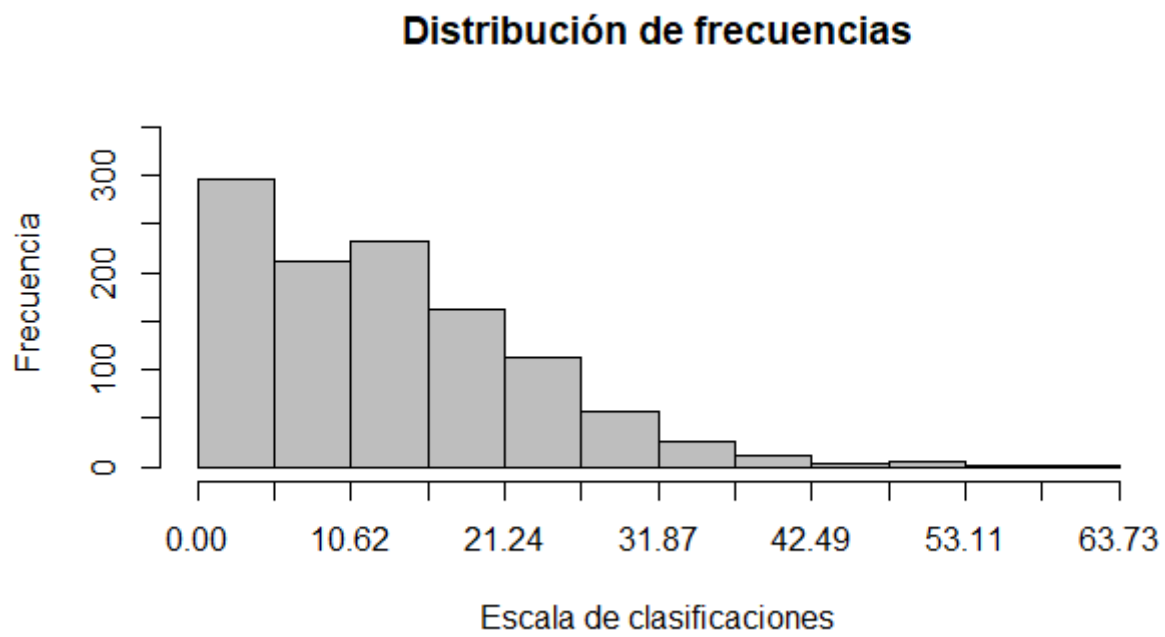


Figura 4.16: Histograma

La asimetría es de 0,98. El gráfico de Cullen&Frey nos muestra que podría ser una gamma.

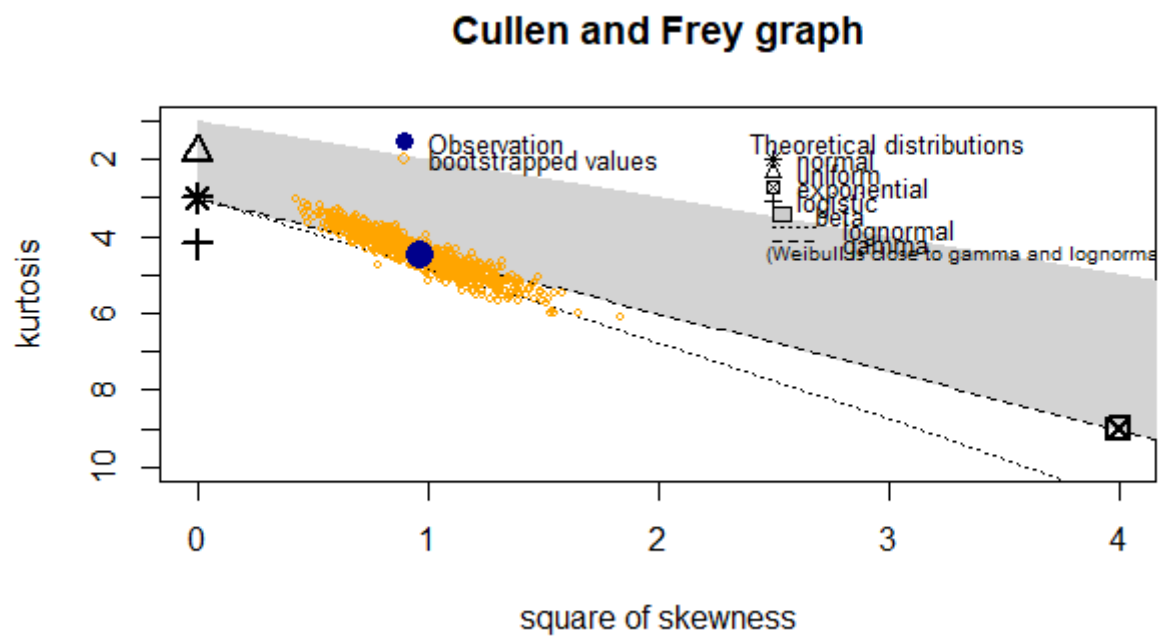


Figura 4.17: Gráfico de Cullen&Frey

Los test no lo confirman.

Distribución	P-value KS
Logística	1.661e-10
Lognormal	2.068e-11
Exponencial	4.156e-13
Gamma	<2.2e-16
Pareto	<2.2e-16
Gumbel	1.077916e-17

4.3. Navascués/Nabaskoze

4.3.1. Año Completo

Para el caso de Navascués (Navarra) se obtienen los siguientes resultados:

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
0	0,20	2,20	6,47	10,10	53,40	8,9448

La mayoría de datos se encuentran en el primer intervalo.

Class limits	f	rf	rf(%)	cf	cf(%)
[0,4.149)	1291	0.59	59.03	1291	59.03
[4.149,8.298)	253	0.12	11.57	1544	70.60
[8.298,12.45)	201	0.09	9.19	1745	79.79
[12.45,16.6)	154	0.07	7.04	1899	86.83
[16.6,20.74)	97	0.04	4.44	1996	91.27
[20.74,24.89)	80	0.04	3.66	2076	94.92
[24.89,29.04)	33	0.02	1.51	2109	96.43
[29.04,33.19)	31	0.01	1.42	2140	97.85
[33.19,37.34)	22	0.01	1.01	2162	98.86
[37.34,41.49)	12	0.01	0.55	2174	99.41
[41.49,45.64)	10	0.00	0.46	2184	99.86
[45.64,49.79)	2	0.00	0.09	2186	99.95
[49.79,53.93)	1	0.00	0.05	2187	100.00

Figura 4.18: Distribución de frecuencias

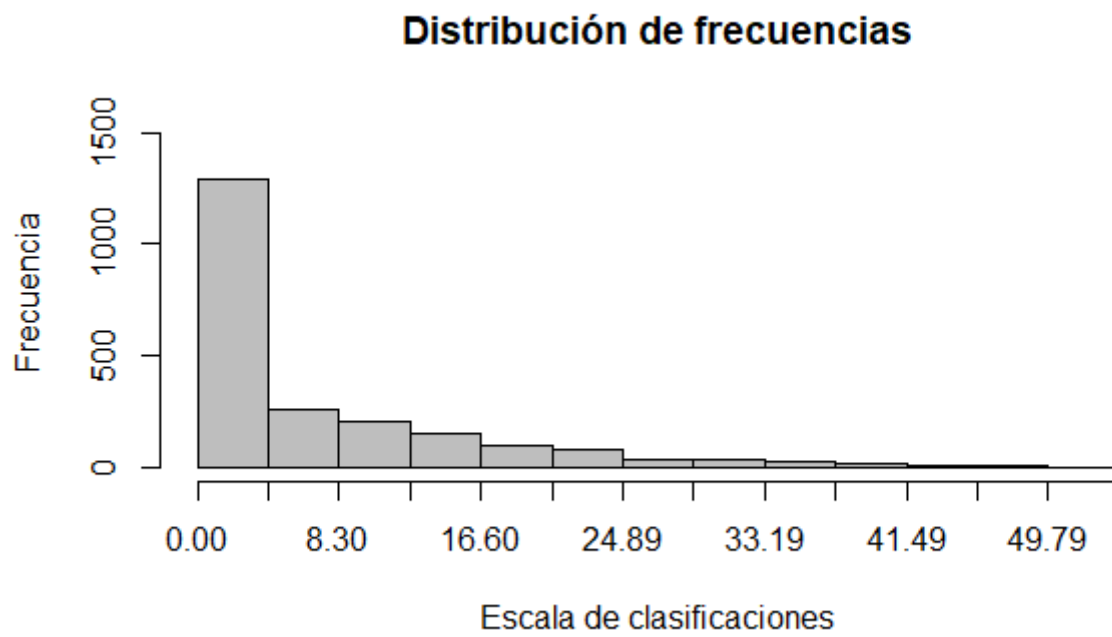


Figura 4.19: Histograma

El gráfico de Cullen&Frey nos muestra que podría ser una exponencial.

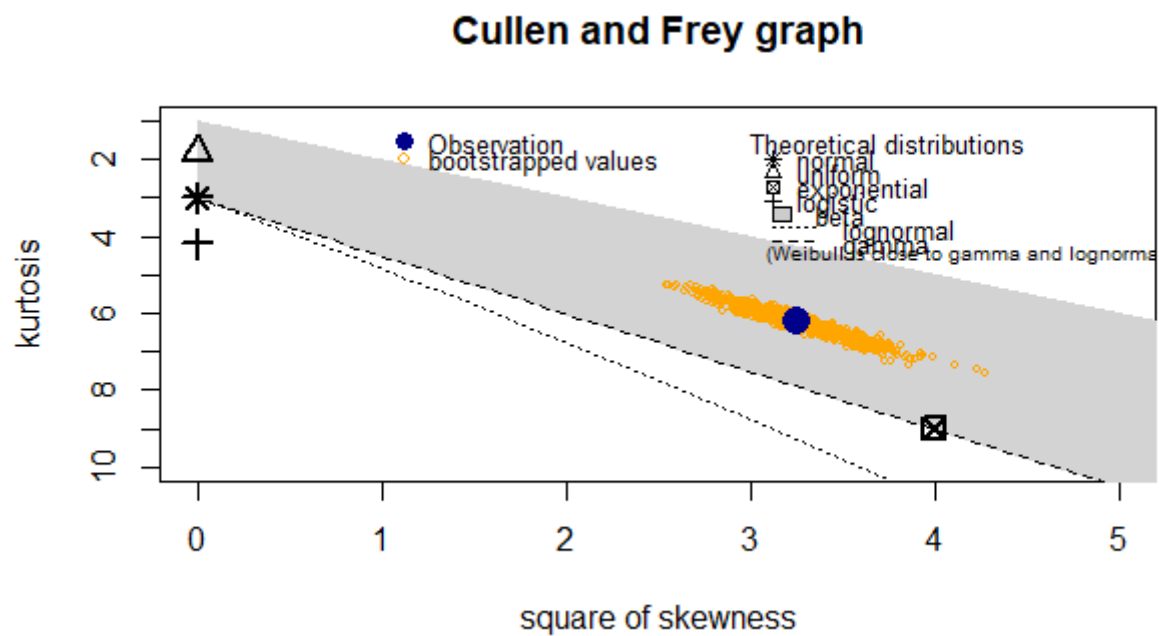


Figura 4.20: Gráfico de Cullen&Frey

No se acepta que la variable siga ninguna de las distribuciones utilizadas.

Distribución	P-value KS
Exponencial	<2.2e-16
Gamma	<2.2e-16
Pareto	<2.2e-16
Gumbel	0

4.3.2. Temporada alta

En el caso de la temporada alta tenemos:

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
0	1,80	8,30	10,74	16,50	53,40	10,40892

La distribución de frecuencias es la siguiente.

Class limits	f	rf	rf(%)	cf	cf(%)
[0,4.494)	416	0.38	37.89	416	37.89
[4.494,8.989)	158	0.14	14.39	574	52.28
[8.989,13.48)	145	0.13	13.21	719	65.48
[13.48,17.98)	146	0.13	13.30	865	78.78
[17.98,22.47)	78	0.07	7.10	943	85.88
[22.47,26.97)	56	0.05	5.10	999	90.98
[26.97,31.46)	42	0.04	3.83	1041	94.81
[31.46,35.96)	25	0.02	2.28	1066	97.09
[35.96,40.45)	18	0.02	1.64	1084	98.72
[40.45,44.94)	8	0.01	0.73	1092	99.45
[44.94,49.44)	5	0.00	0.46	1097	99.91
[49.44,53.93)	1	0.00	0.09	1098	100.00

Figura 4.21: Distribución de frecuencias

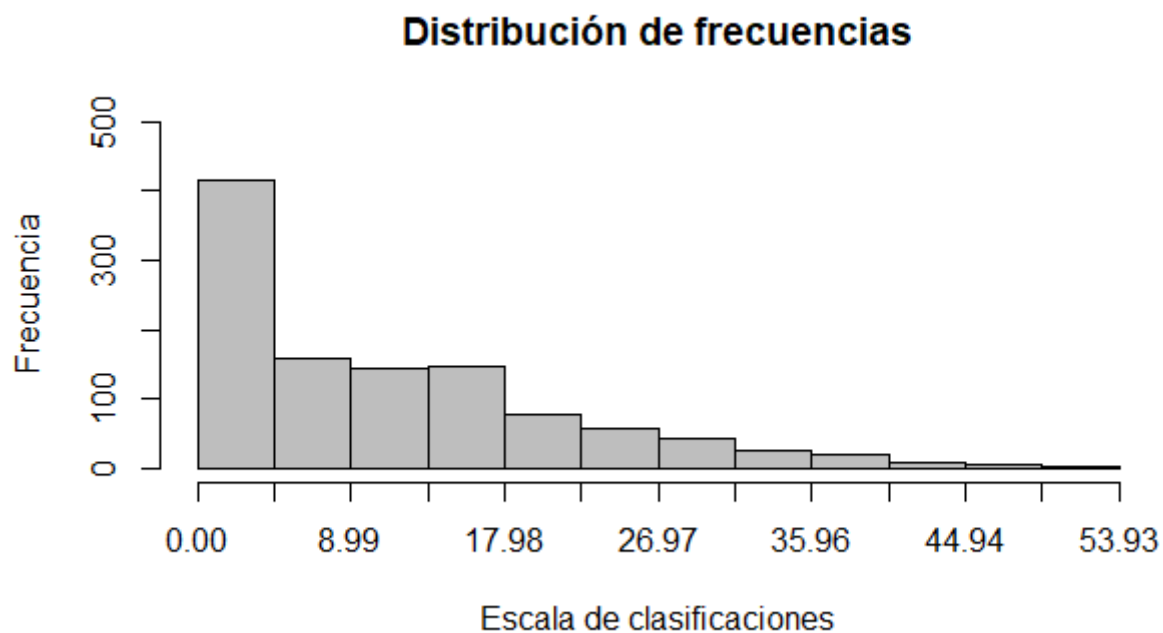


Figura 4.22: Histograma

El gráfico de Cullen&Frey nos muestra que podría ser una exponencial.

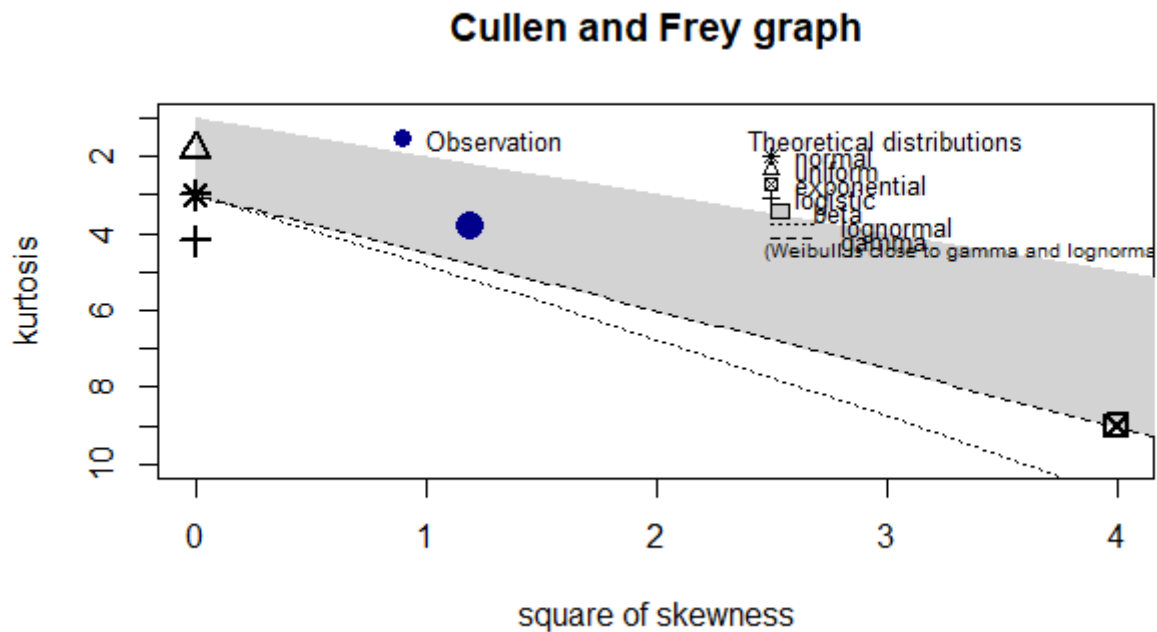


Figura 4.23: Gráfico de Cullen&Frey

La variable no se ajusta a ninguna distribución.

Distribución	P-value KS
Exponencial	<2.2e-16
Gamma	<2.2e-16
Pareto	<2.2e-16
Gumbel	3.265834e-128

4.3.3. Temporada baja

En el caso de la temporada baja tenemos:

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
0	0	0,40	2,17	2,7	27,30	3,82037

La mayoría de los datos están en el primer intervalo.

Class limits	f	rf	rf(%)	cf	cf(%)
[0,2.298)	793	0.73	72.82	793	72.82
[2.298,4.596)	103	0.09	9.46	896	82.28
[4.596,6.893)	69	0.06	6.34	965	88.61
[6.893,9.191)	49	0.04	4.50	1014	93.11
[9.191,11.49)	38	0.03	3.49	1052	96.60
[11.49,13.79)	13	0.01	1.19	1065	97.80
[13.79,16.08)	9	0.01	0.83	1074	98.62
[16.08,18.38)	9	0.01	0.83	1083	99.45
[18.38,20.68)	0	0.00	0.00	1083	99.45
[20.68,22.98)	2	0.00	0.18	1085	99.63
[22.98,25.28)	3	0.00	0.28	1088	99.91
[25.28,27.57)	1	0.00	0.09	1089	100.00

Figura 4.24: Distribución de frecuencias

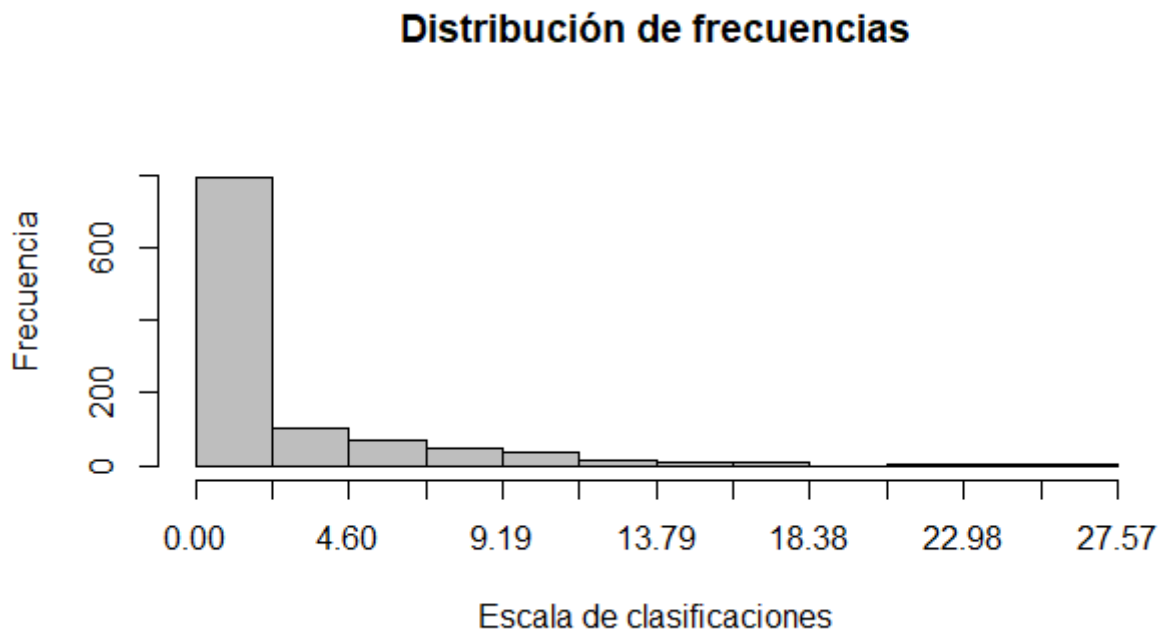


Figura 4.25: Histograma

El gráfico de Cullen&Frey nos muestra que podría ser una exponencial.

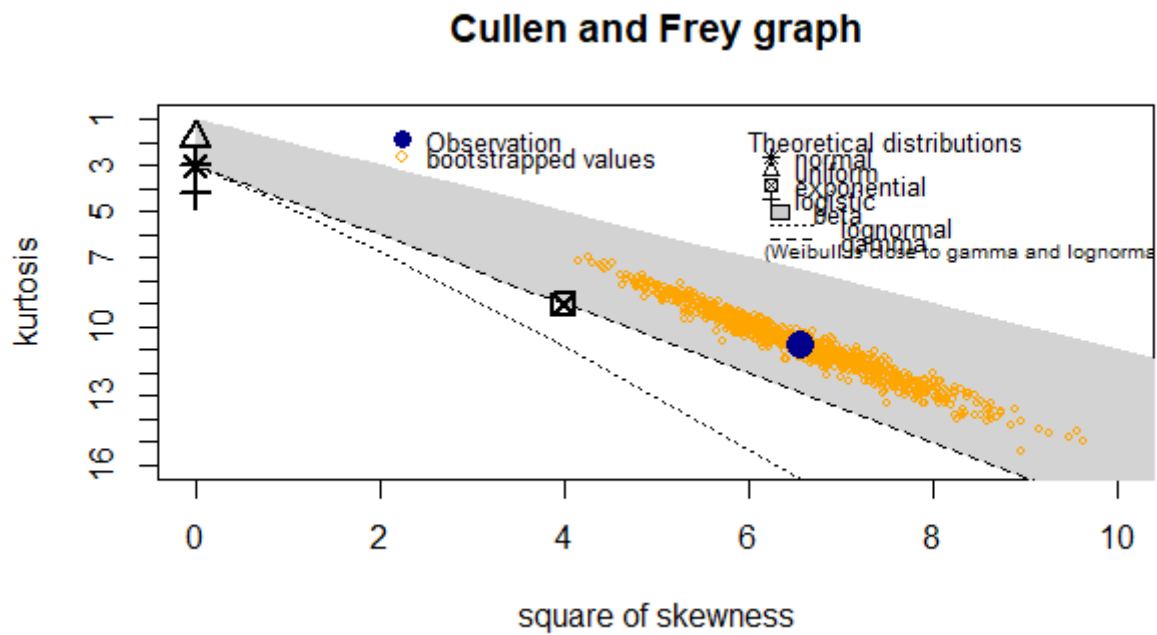


Figura 4.26: Gráfico de Cullen&Frey

La variable no se ajusta a ninguna distribución.

Distribución	P-value KS
Exponencial	<2.2e-16
Gamma	<2.2e-16
Pareto	<2.2e-16
Gumbel	5.469846e-235

4.4. Los Tojos

4.4.1. Año Completo

Para el caso de Los Tojos (Cantabria) se obtienen los siguientes resultados:

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
0	0	2,20	6,47	10,10	53,40	8,9448

En este caso, se trata de una zona lluviosa donde el número medio de días con precipitación superior o igual a 1,0 mm es de 115.

Class limits	f	rf	rf(%)	cf	cf(%)
[0,3.294)	1505	0.69	68.82	1505	68.82
[3.294,6.588)	208	0.10	9.51	1713	78.33
[6.588,9.882)	140	0.06	6.40	1853	84.73
[9.882,13.18)	126	0.06	5.76	1979	90.49
[13.18,16.47)	74	0.03	3.38	2053	93.87
[16.47,19.76)	56	0.03	2.56	2109	96.43
[19.76,23.06)	36	0.02	1.65	2145	98.08
[23.06,26.35)	17	0.01	0.78	2162	98.86
[26.35,29.65)	5	0.00	0.23	2167	99.09
[29.65,32.94)	10	0.00	0.46	2177	99.54
[32.94,36.24)	4	0.00	0.18	2181	99.73
[36.24,39.53)	4	0.00	0.18	2185	99.91
[39.53,42.82)	2	0.00	0.09	2187	100.00

Figura 4.27: Distribución de frecuencias

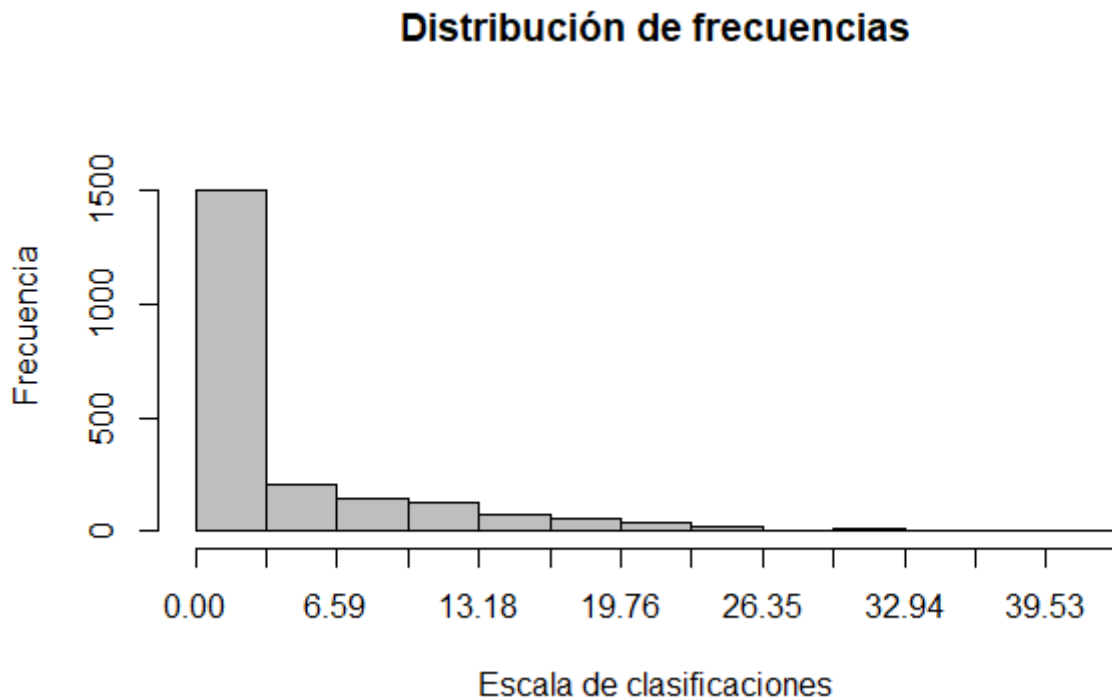


Figura 4.28: Histograma

El gráfico de Cullen&Frey nos muestra que podría ser una exponencial.

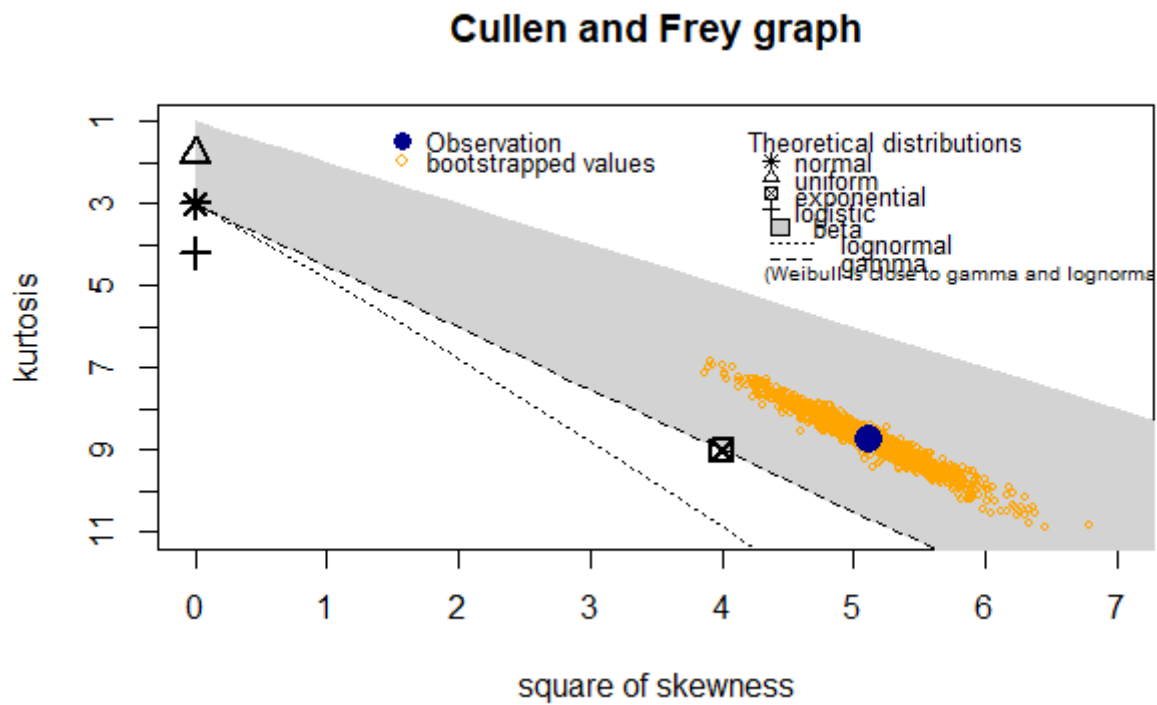


Figura 4.29: Gráfico de Cullen&Frey

Sin embargo, no se ajusta a ninguna distribución utilizada.

Distribución P-value KS

Exponencial	<2.2e-16
Gamma	<2.2e-16
Pareto	<2.2e-16
Gumbel	0

4.4.2. Temporada alta

En el caso de la temporada alta tenemos:

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
0	0,30	2,40	5,85	9,78	41,50	7,494281

La mayoría de datos se encuentran en el primer intervalo, como podemos ver con la distribución de frecuencias.

Class limits	f	rf	rf(%)	cf	cf(%)
[0,3.493)	607	0.55	55.28	607	55.28
[3.493,6.986)	135	0.12	12.30	742	67.58
[6.986,10.48)	102	0.09	9.29	844	76.87
[10.48,13.97)	90	0.08	8.20	934	85.06
[13.97,17.46)	69	0.06	6.28	1003	91.35
[17.46,20.96)	43	0.04	3.92	1046	95.26
[20.96,24.45)	21	0.02	1.91	1067	97.18
[24.45,27.94)	11	0.01	1.00	1078	98.18
[27.94,31.44)	7	0.01	0.64	1085	98.82
[31.44,34.93)	8	0.01	0.73	1093	99.54
[34.93,38.42)	4	0.00	0.36	1097	99.91
[38.42,41.91)	1	0.00	0.09	1098	100.00

Figura 4.30: Distribución de frecuencias

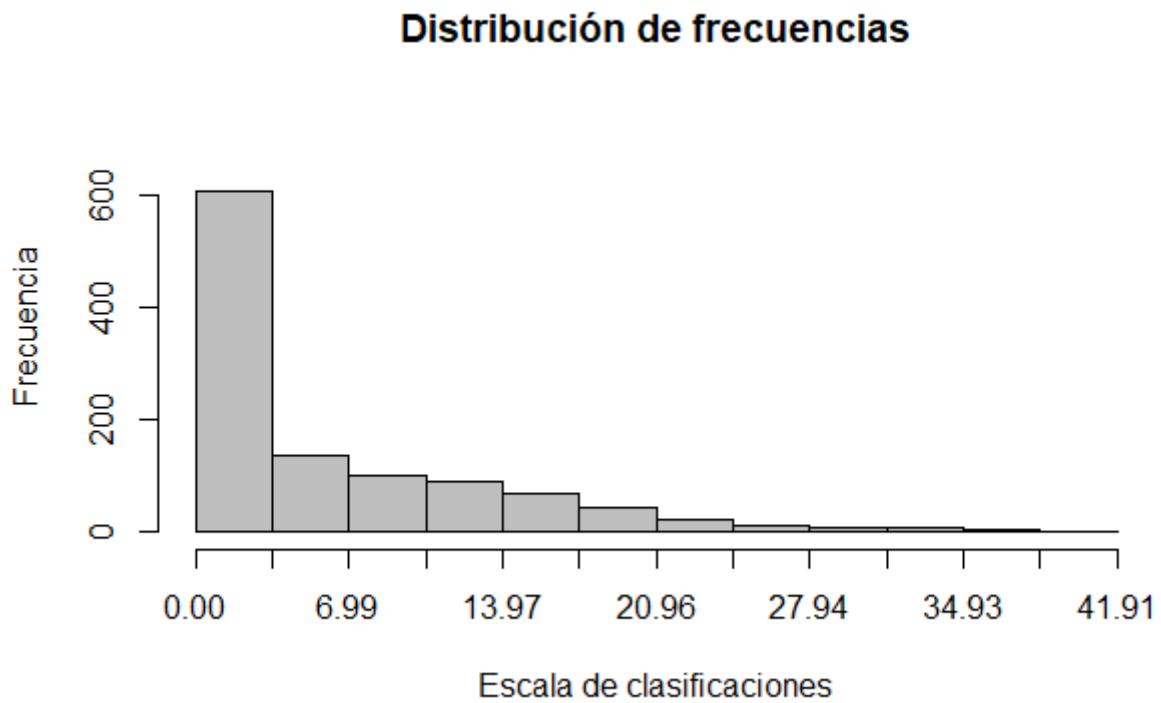


Figura 4.31: Histograma

El gráfico de Cullen&Frey nos muestra que podría ser una exponencial. La asimetría es de 1,58.

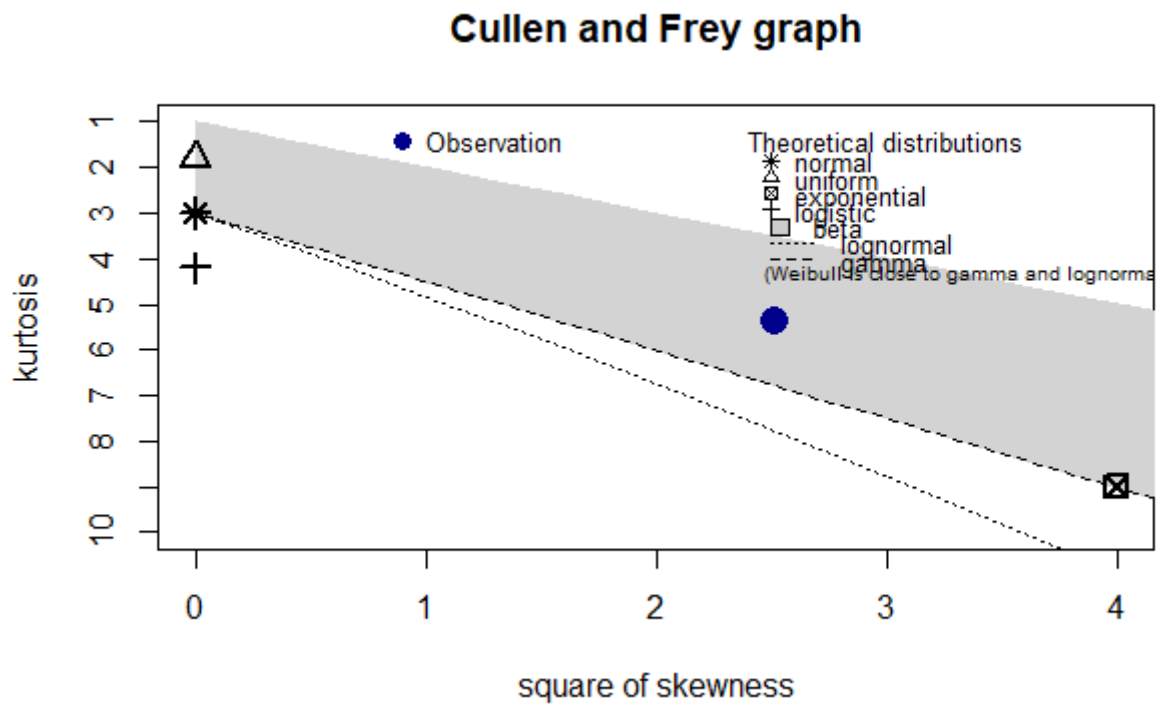


Figura 4.32: Gráfico de Cullen&Frey

La variable no se ajusta a ninguna distribución utilizada.

Distribución	P-value KS
Exponencial	<2.2e-16
Gamma	<2.2e-16
Pareto	<2.2e-16
Gumbel	2.401163e-283

4.4.3. Temporada baja

En el caso de la temporada baja tenemos:

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
0	0,00	0,20	1,81	1,00	42,40	3,960267

Los datos se concentran en el primer intervalo, por lo que la búsqueda de una distribución no tiene sentido. Simplemente se trata de una temporada con muchos días lluviosos donde la posibilidad de lluvia es nula o casi nula.

Class limits	f	rf	rf(%)	cf	cf(%)
[0,3.569)	911	0.84	83.65	911	83.65
[3.569,7.137)	84	0.08	7.71	995	91.37
[7.137,10.71)	43	0.04	3.95	1038	95.32
[10.71,14.27)	29	0.03	2.66	1067	97.98
[14.27,17.84)	8	0.01	0.73	1075	98.71
[17.84,21.41)	10	0.01	0.92	1085	99.63
[21.41,24.98)	1	0.00	0.09	1086	99.72
[24.98,28.55)	0	0.00	0.00	1086	99.72
[28.55,32.12)	2	0.00	0.18	1088	99.91
[32.12,35.69)	0	0.00	0.00	1088	99.91
[35.69,39.26)	0	0.00	0.00	1088	99.91
[39.26,42.82)	1	0.00	0.09	1089	100.00

Figura 4.33: Distribución de frecuencias

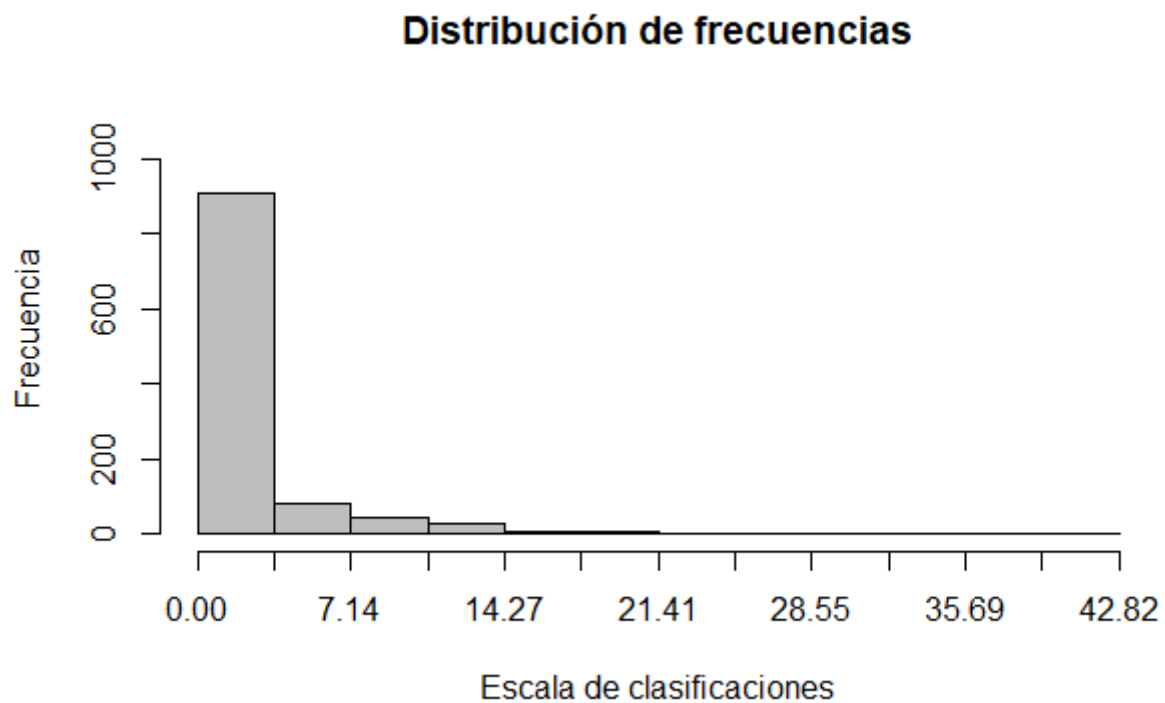


Figura 4.34: Histograma

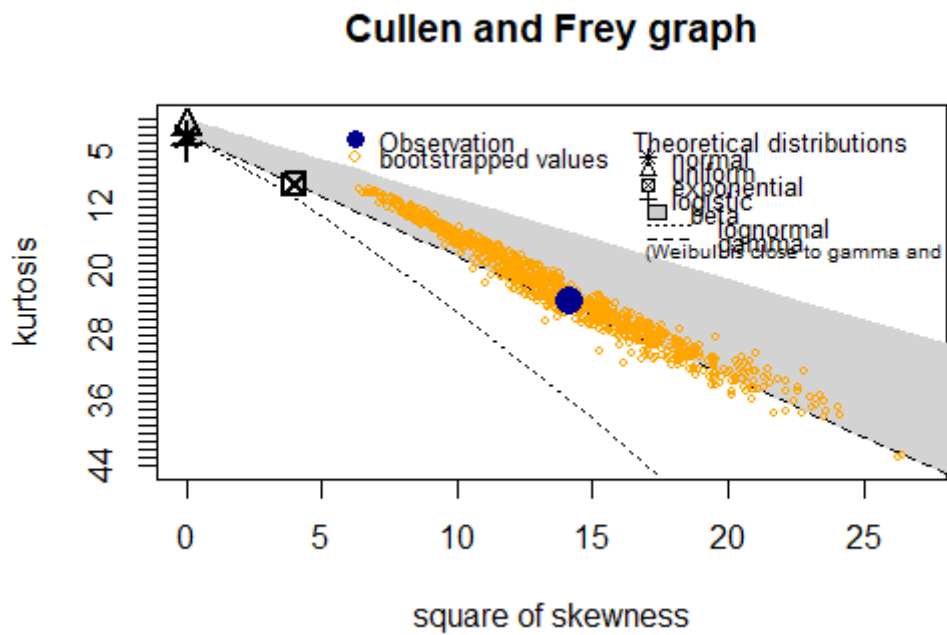


Figura 4.35: Gráfico de Cullen&Frey

La variable no se ajusta a ninguna distribución utilizada.

Distribución	P-value KS
Exponencial	<2.2e-16
Gamma	<2.2e-16
Pareto	<2.2e-16
Gumbel	0

4.5. Vimianzo

4.5.1. Año Completo

Para el caso de Vimianzo (A Coruña) se obtienen los siguientes resultados:

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
0	0	0,50	3,68	5,20	40,30	6,113191

En este caso, también se trata de una zona lluviosa donde el número medio de días con precipitación superior o igual a 1,0 mm es de más de 150.

Class limits	f	rf	rf(%)	cf	cf(%)
[0,3.131)	1491	0.68	68.18	1491	68.18
[3.131,6.262)	215	0.10	9.83	1706	78.01
[6.262,9.393)	163	0.07	7.45	1869	85.46
[9.393,12.52)	110	0.05	5.03	1979	90.49
[12.52,15.65)	72	0.03	3.29	2051	93.78
[15.65,18.79)	51	0.02	2.33	2102	96.11
[18.79,21.92)	32	0.01	1.46	2134	97.58
[21.92,25.05)	15	0.01	0.69	2149	98.26
[25.05,28.18)	20	0.01	0.91	2169	99.18
[28.18,31.31)	10	0.00	0.46	2179	99.63
[31.31,34.44)	5	0.00	0.23	2184	99.86
[34.44,37.57)	2	0.00	0.09	2186	99.95
[37.57,40.7)	1	0.00	0.05	2187	100.00

Figura 4.36: Distribución de frecuencias

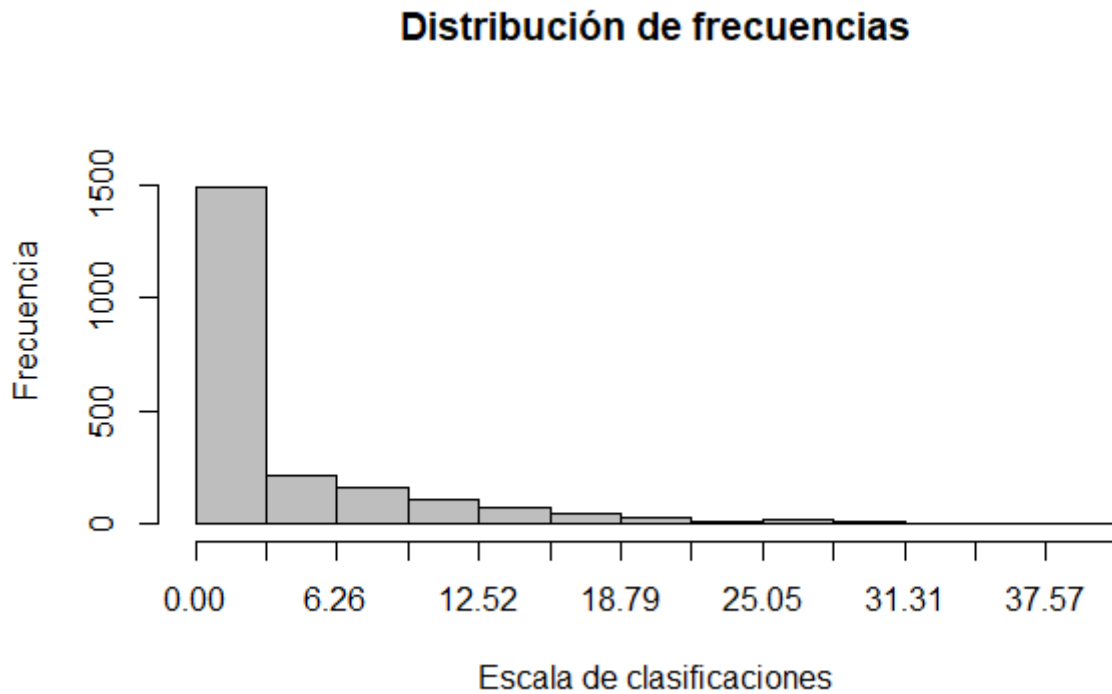


Figura 4.37: Histograma

El gráfico de Cullen&Frey nos muestra que se acerca a una exponencial. La asimetría es de 2,25

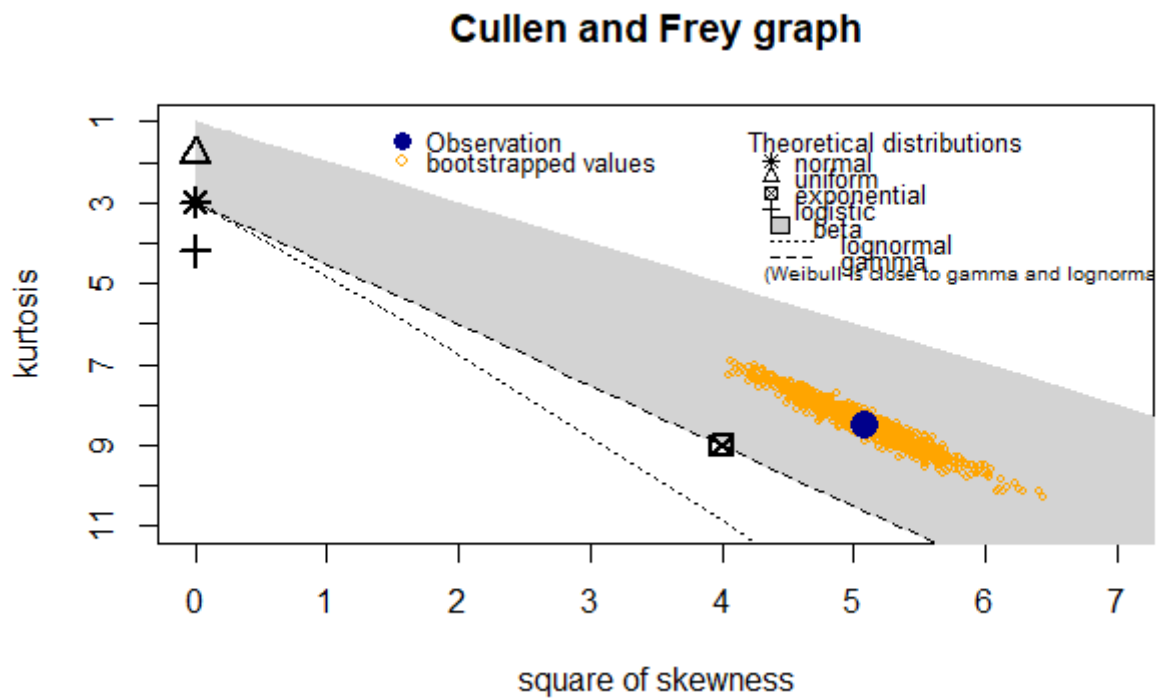


Figura 4.38: Gráfico de Cullen&Frey

Sin embargo, no se ajusta a ninguna distribución utilizada.

Distribución	P-value KS
Exponencial	<2.2e-16
Gamma	<2.2e-16
Pareto	<2.2e-16
Gumbel	0

4.5.2. Temporada alta

En el caso de la temporada alta tenemos:

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
0	0,30	2,75	5,64	8,80	40,30	7,034541

Los valores se siguen concentrando en el primer intervalo, pero en menor medida si lo comparamos con el conjunto del año.

Class limits	f	rf	rf(%)	cf	cf(%)
[0,3.392)	580	0.53	52.82	580	52.82
[3.392,6.784)	173	0.16	15.76	753	68.58
[6.784,10.18)	116	0.11	10.56	869	79.14
[10.18,13.57)	85	0.08	7.74	954	86.89
[13.57,16.96)	53	0.05	4.83	1007	91.71
[16.96,20.35)	35	0.03	3.19	1042	94.90
[20.35,23.74)	18	0.02	1.64	1060	96.54
[23.74,27.14)	19	0.02	1.73	1079	98.27
[27.14,30.53)	12	0.01	1.09	1091	99.36
[30.53,33.92)	4	0.00	0.36	1095	99.73
[33.92,37.31)	2	0.00	0.18	1097	99.91
[37.31,40.7)	1	0.00	0.09	1098	100.00

Figura 4.39: Distribución de frecuencias

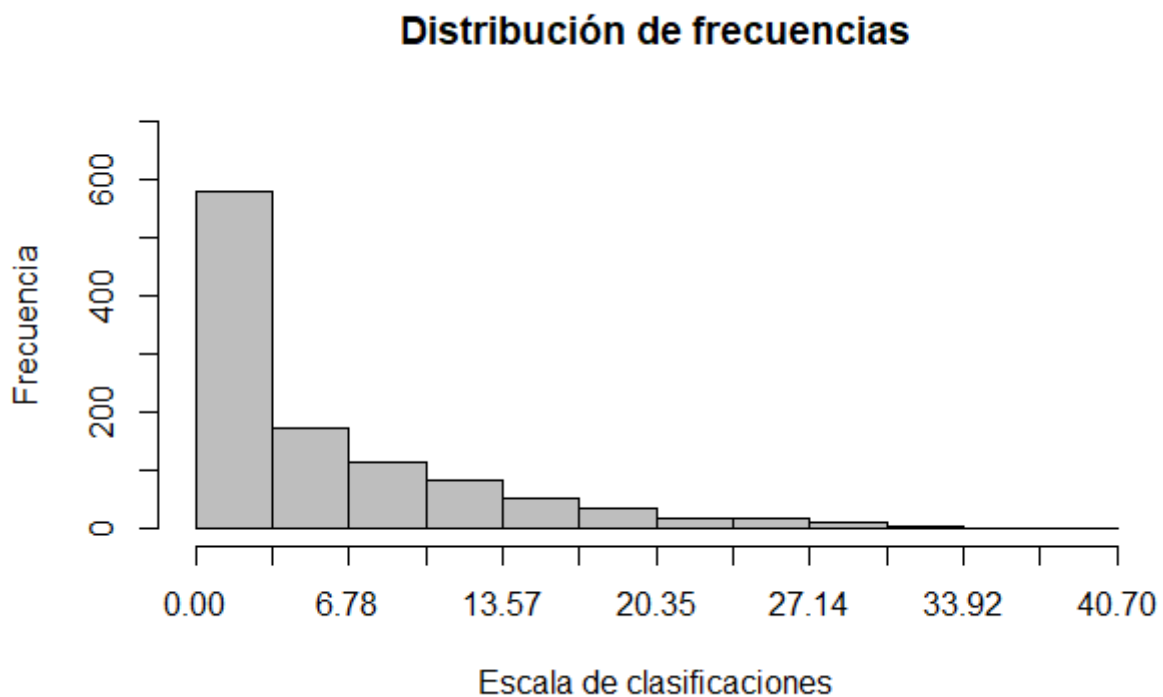


Figura 4.40: Histograma

El gráfico de Cullen&Frey nos muestra que podría ser una exponencial. .

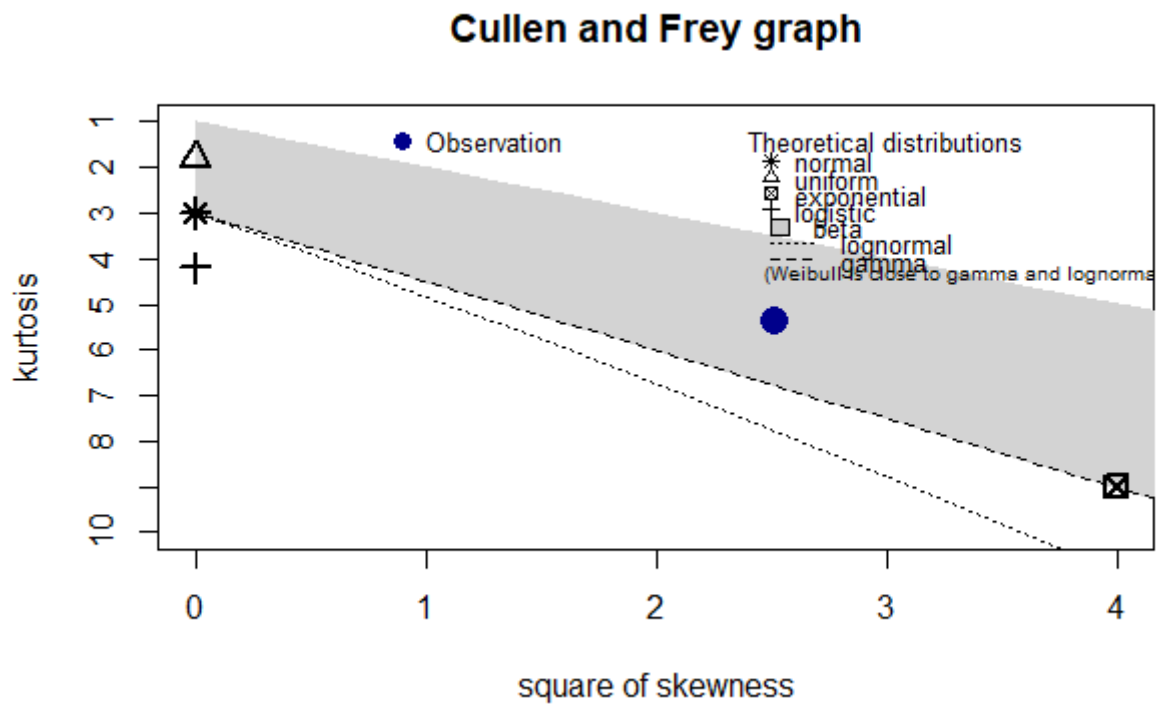


Figura 4.41: Gráfico de Cullen&Frey

La variable no se ajusta a ninguna distribución utilizada.

Distribución	P-value KS
Exponencial	<2.2e-16
Gamma	<2.2e-16
Pareto	<2.2e-16
Gumbel	4.852484e-155

4.5.3. Temporada baja

En el caso de la temporada baja tenemos:

Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo	Desviación típica
0	0,00	0,00	1,74	0,80	37,10	4,149569

Al tratarse de una zona lluviosa, la mayoría de los datos se encuentran en el primer intervalo.

Class limits	f	rf	rf(%)	cf	cf(%)
[0,3.392)	1051	0.83	82.63	1051	82.63
[3.392,6.784)	84	0.07	6.60	1135	89.23
[6.784,10.18)	55	0.04	4.32	1190	93.55
[10.18,13.57)	40	0.03	3.14	1230	96.70
[13.57,16.96)	13	0.01	1.02	1243	97.72
[16.96,20.35)	12	0.01	0.94	1255	98.66
[20.35,23.74)	6	0.00	0.47	1261	99.14
[23.74,27.14)	2	0.00	0.16	1263	99.29
[27.14,30.53)	4	0.00	0.31	1267	99.61
[30.53,33.92)	1	0.00	0.08	1268	99.69
[33.92,37.31)	3	0.00	0.24	1271	99.92
[37.31,40.7)	1	0.00	0.08	1272	100.00

Figura 4.42: Distribución de frecuencias

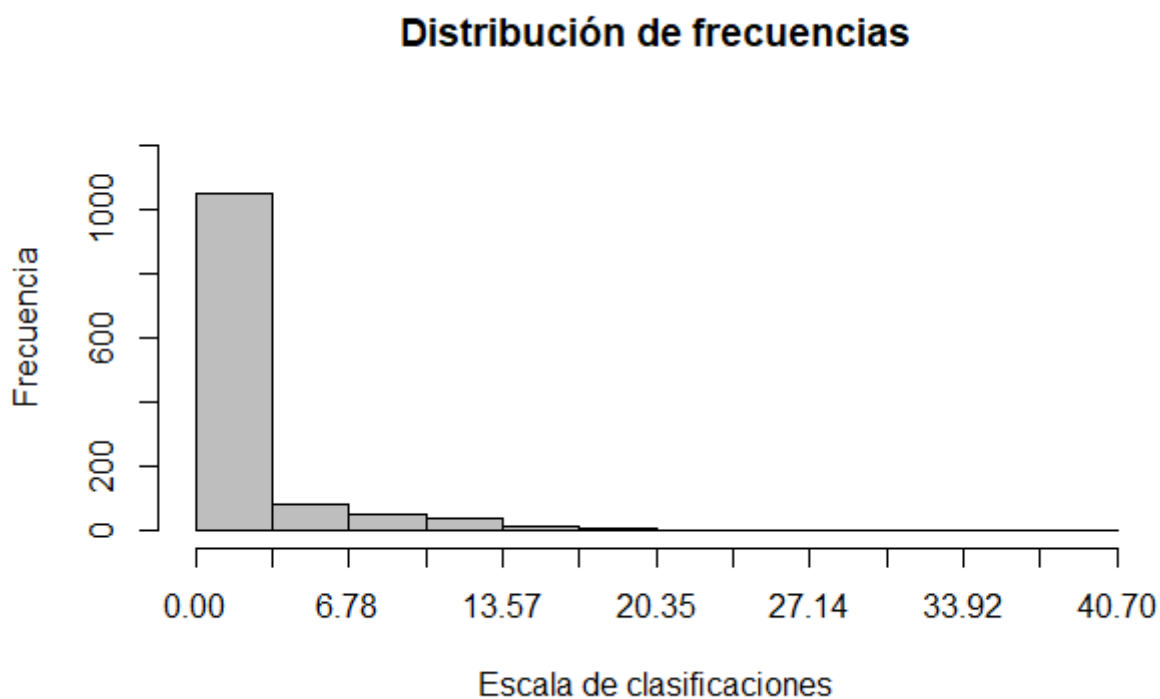


Figura 4.43: Histograma

La variable no se ajusta a ninguna distribución utilizada.

Distribución P-value KS

Exponencial	<2.2e-16
Gamma	<2.2e-16
Pareto	<2.2e-16
Gumbel	0

Parte II

RIESGO

Capítulo 5

ENFOQUE

La segunda parte del trabajo consistió en estudiar la variable RIESGO, también conocida como IMRI. El análisis de esta variable ha sido más difícil pues las referencias encontradas no han sido suficientes.

A la hora de trabajar con variables categóricas se han utilizado tablas de frecuencias y se ha intentado ajustar a las distribuciones discretas más conocidas, según la información encontrada.

El trabajo se ha realizado utilizando los datos de Torrecilla de los Ángeles para el conjunto de años, para la temporada alta y para la temporada baja. Después, se realizó el análisis para el resto de localizaciones sugeridas.

Capítulo 6

ANÁLISIS AÑOS COMPLETOS

Como ya vimos al estudiar la variable FWI, la mayoría de los datos son valores bajos, esto supone que a la hora de analizar la variable RIESGO ocurra lo mismo.

	1	2	3	4	5
Frecuencia	1116	545	351	114	61
Proporción	0,51	0,25	0,16	0,05	0,03

Gráficamente se observa una forma escalonada

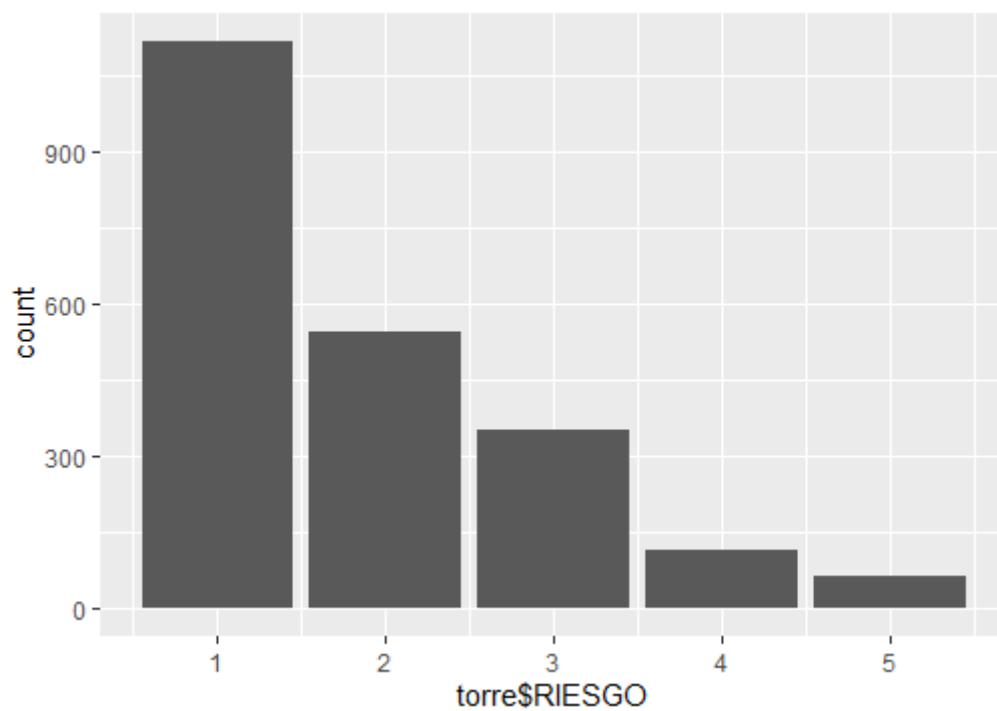


Figura 6.1: Histograma

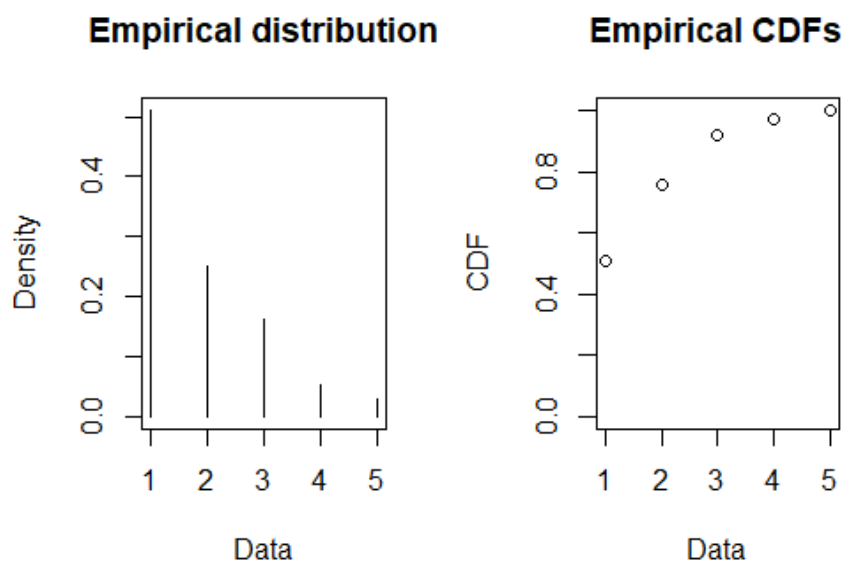


Figura 6.2: Gráfico de la función de distribución

Existe la posibilidad de hacer el gráfico de Cullen&Frey. En este caso no obtenemos una información clara sobre la distribución que podría tener la variable.

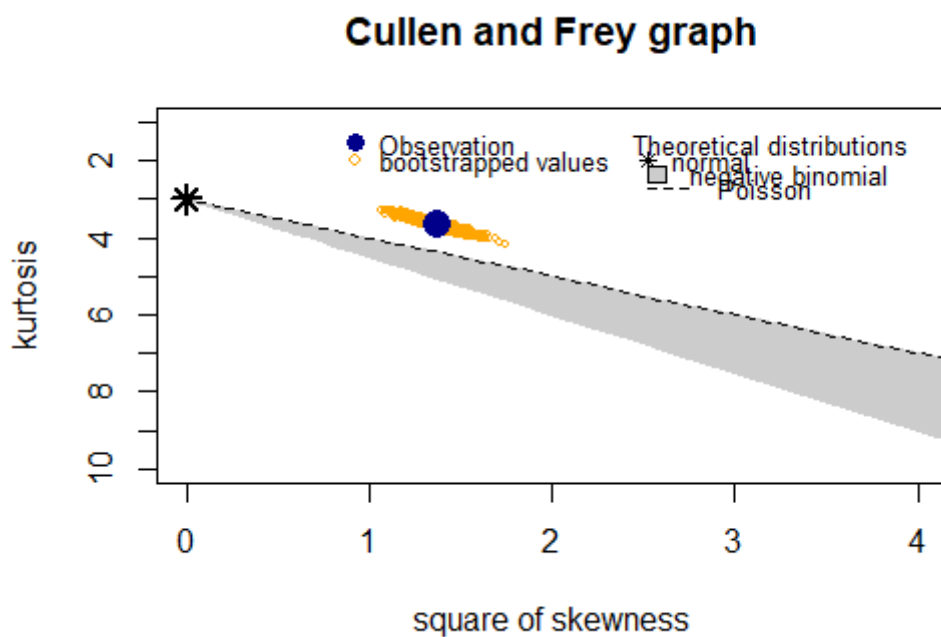


Figura 6.3: Gráfico de Cullen&Frey

6.1. Ajustes

El análisis de ajuste se ha hecho desde dos perspectivas:

1. Primero, se han realizado ajustes a diferentes distribuciones.
2. Segundo, se ha centrado el análisis en la distribución multinomial. Aquí, al tratarse de varias posibles categorías (5) se obtienen las probabilidades de que pertenezca a cada categoría. Para obtener esas probabilidades se han utilizado los datos del periodo 2014-2017. Luego, los datos del periodo 2018-2019 se han usado para saber si realmente los datos siguen las proporciones obtenidas anteriormente.

6.1.1. Ajuste a distintas distribuciones

Se ha intentado ajustar a las distribuciones más conocidas sin obtener resultados positivos.

Aquí, para conocer la bondad del ajuste no se utiliza la prueba de Kolgomorov-Smirnov, sino la prueba de bondad de ajuste de la chi-cuadrado. La hipótesis nula es la misma que para Kolgomorov-Smirnov, los datos se ajustan a una distribución dada.

Análisis del ajuste de los datos a la distribución de Poisson Para hacer un ajuste a una distribución Poisson es necesario conocer el valor de la estimación de lambda, que en este caso es de **1.838134**. El cálculo es sencillo, se trata de la media. Una vez hecho el ajuste a una Poisson, se ha obtenido lo siguiente.

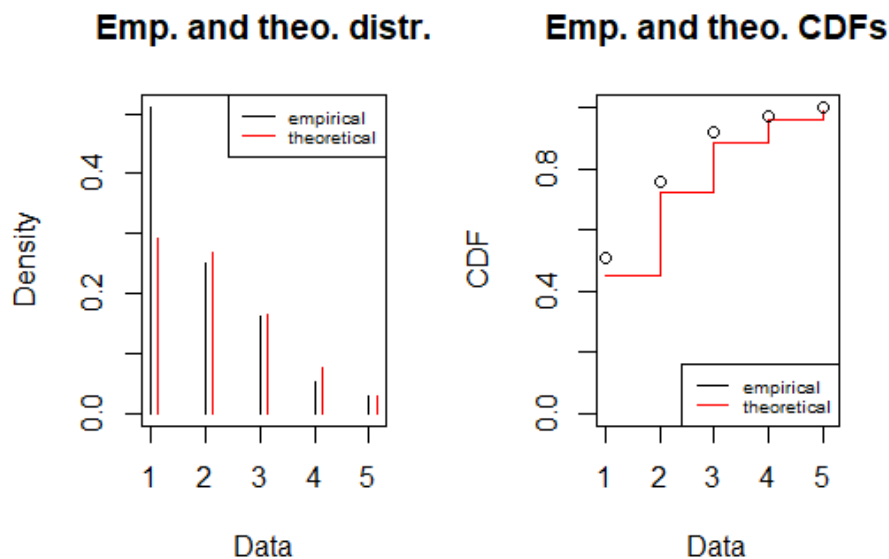


Figura 6.4: Gráfico de la función de distribución teórica y empírica

El p-valor asociado a la prueba de bondad de ajuste de chi-cuadrado ha sido de **2.1773e-09**.

Por tanto, rechazamos H_0 , no se ajusta a una distribución de Poisson.

Análisis del ajuste de los datos a la distribución Geométrica Esta distribución está relacionada con la realización de pruebas hasta la consecución del éxito. Por tanto, no es una distribución que se pueda usar para nuestros datos. Aun así, se ha incluido en el trabajo.

Los resultados confirman lo dicho.

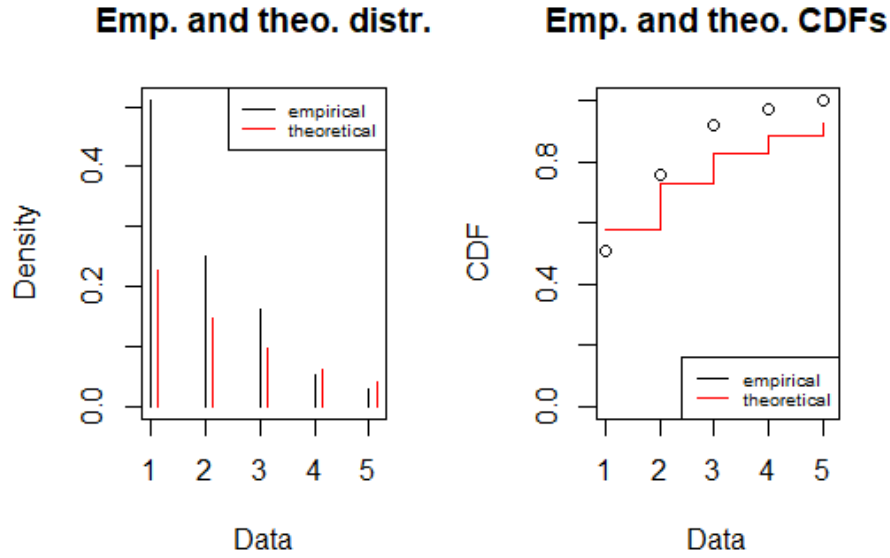


Figura 6.5: Gráfico de la función de distribución teórica y empírica

El p-valor asociado a la prueba de chi-cuadrado es de **5.031805e-89**. No se ajusta a una distribución geométrica.

Análisis del ajuste de los datos a la distribución multinomial En este tipo de variables, se va a tener en cuenta las frecuencias para comparar una distribución teórica con la distribución de la variable de estudio. Se simulará una distribución teórica, se calculará la probabilidad de que el valor este en cada categoría y se comparará con la variable RIESGO.

La distribución multinomial se define como:

Supongamos que se tienen c posibles categorías. Dados n sucesos, se puede definir una variable aleatoria X_i (para $i = 1, \dots, c$) que indica el número de veces que aparece la categoría i .

Se denota la probabilidad de obtener cada categoría i como $(\pi_1, \pi_2, \dots, \pi_c)$ donde

$$\sum_i \pi_i = 1 \quad (6.1)$$

Para n observaciones independientes, la probabilidad de que n_1 observaciones caigan en la categoría 1, n_2 caigan en la categoría 2, ..., n_c caigan en la categoría c , (donde $\sum_i n_i = n$) es igual a

$$P(n_1, n_2, \dots, n_c) = \left(\frac{n!}{n_1! n_2! \dots n_c!} \right) \pi_1^{n_1} \pi_2^{n_2} \dots \pi_c^{n_c} \quad (6.2)$$

Para la inferencia, los parametros a estimar son $(\pi_1, \dots, \pi_c)$.

Los estimadores por máxima verosimilitud son:

$$\hat{\pi}_j = \frac{n_j}{n} \quad (6.3)$$

Para $j = 1, \dots, c$

Es decir, los estimadores son simplemente las proporciones muestrales de cada categoría.

Para el caso, se necesitan unas proporciones para saber si la muestra sigue las mismas proporciones que debería. Para ello, se han calculado mediante los datos de 2014-2017. Los datos de 2018-2019 serán los utilizados para saber si se ajustan a los datos de 2014-2017.

Una vez obtenidas ambas proporciones, se realiza la prueba de bondad de ajuste de chi cuadrado.

Por tanto, con esto se puede saber si los datos siguen una distribución multinomial o no. H_0 sería que los datos siguen una distribución multinomial.

Para nuestro caso, el p-valor obtenido es **0,001517**. No se puede decir que la muestra de los años 2018-2019 sigue las estadísticas de los años anteriores. Es decir, que no siguen las probabilidades que tenían los datos de los años anteriores a un 5 % de significación.

Capítulo 7

ANÁLISIS TEMPORADA ALTA

A continuación se va a analizar la temporada alta.

	1	2	3	4	5
Frecuencia	278	295	339	114	61
Proporción	0,26	0,27	0,31	0,10	0,06

El histograma muestra que la concentración de datos está principalmente en las tres primeras categorías.

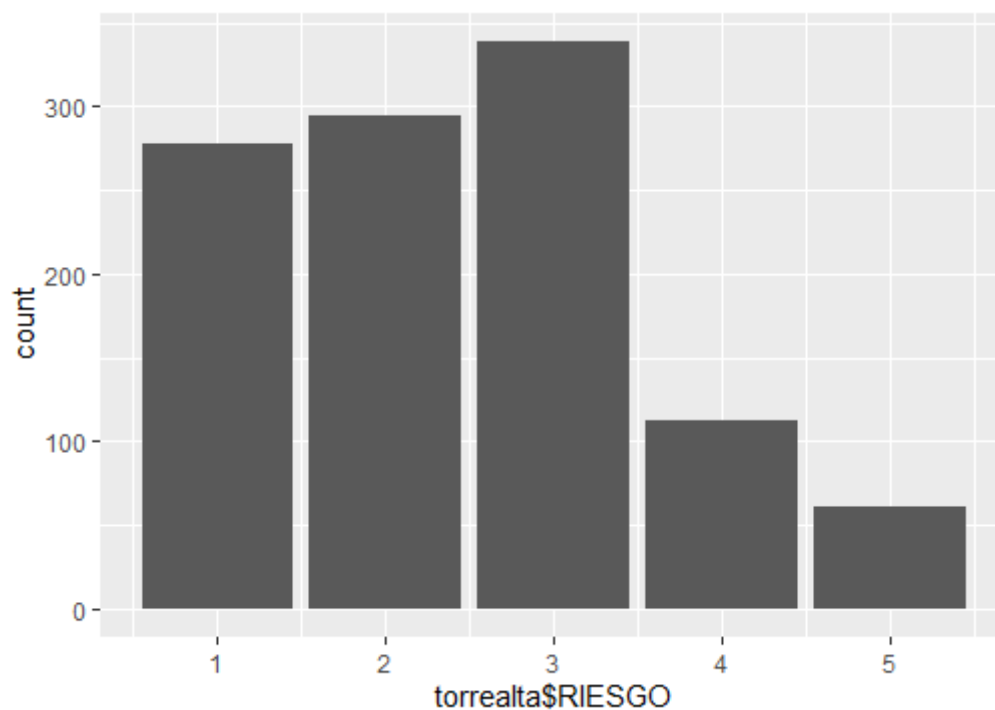


Figura 7.1: Histograma

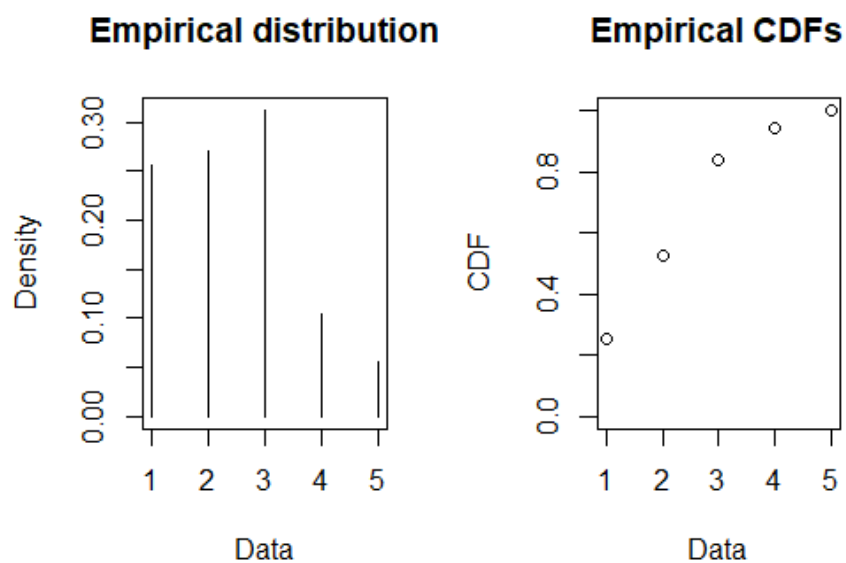


Figura 7.2: Gráfico de la función de distribución

El gráfico de Cullen&Frey nos muestra que podría seguir una distribución normal. Los valores de asimetría y curtosis son 0,42 y 2,47 respectivamente.

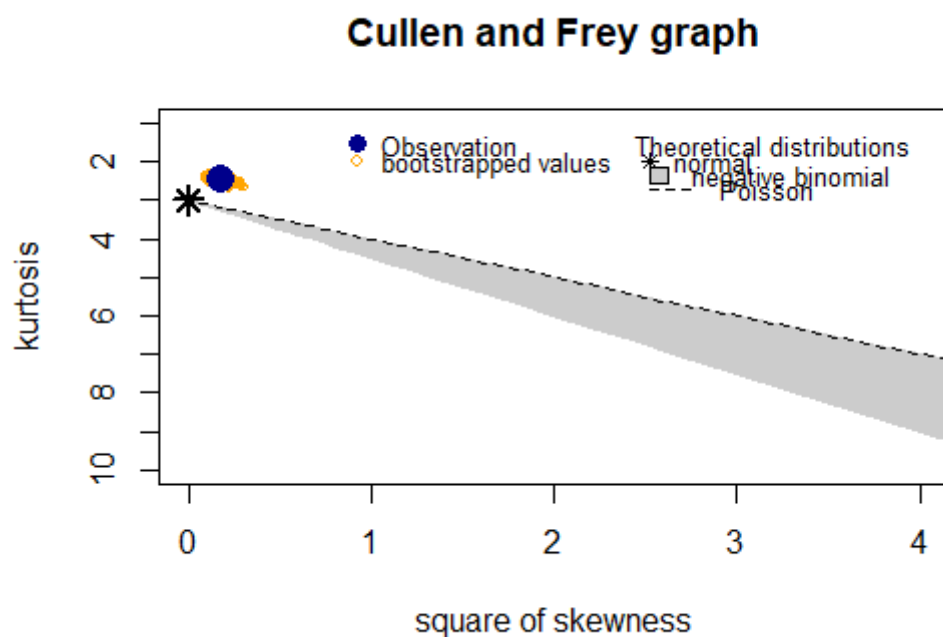


Figura 7.3: Gráfico de Cullen&Frey

Los valores de asimetría y curtosis son **0,42** y **2,47** respectivamente.

Tras realizar el ajuste obtenemos que no se distribuye como ninguna de las distribuciones utilizadas.

Distribución	P-value KS
Normal	3.190605e-71
Logística	2.371363e-22
Geométrica	6.700921e-183

7.1. Análisis del ajuste de los datos a la distribución multinomial

2014-2017	1	2	3	4	5
Frecuencia	173	184	233	85	49
Proporción	0,24	0,25	0,32	0,12	0,07
2018-2019	1	2	3	4	5
Frecuencia	105	111	106	28	12
Proporción	0,29	0,31	0,29	0,08	0,03

El p-valor obtenido en la prueba de chi cuadrado es de **0.0004662**, no aceptamos H0.

Capítulo 8

ANÁLISIS TEMPORADA BAJA

Para la temporada baja tenemos:

	1	2	3	4	5
Frecuencia	834	250	12	1	0
Proporción	0,76	0,23	0,01	0,0009	0

En este caso no hay valores en la categoría 5.

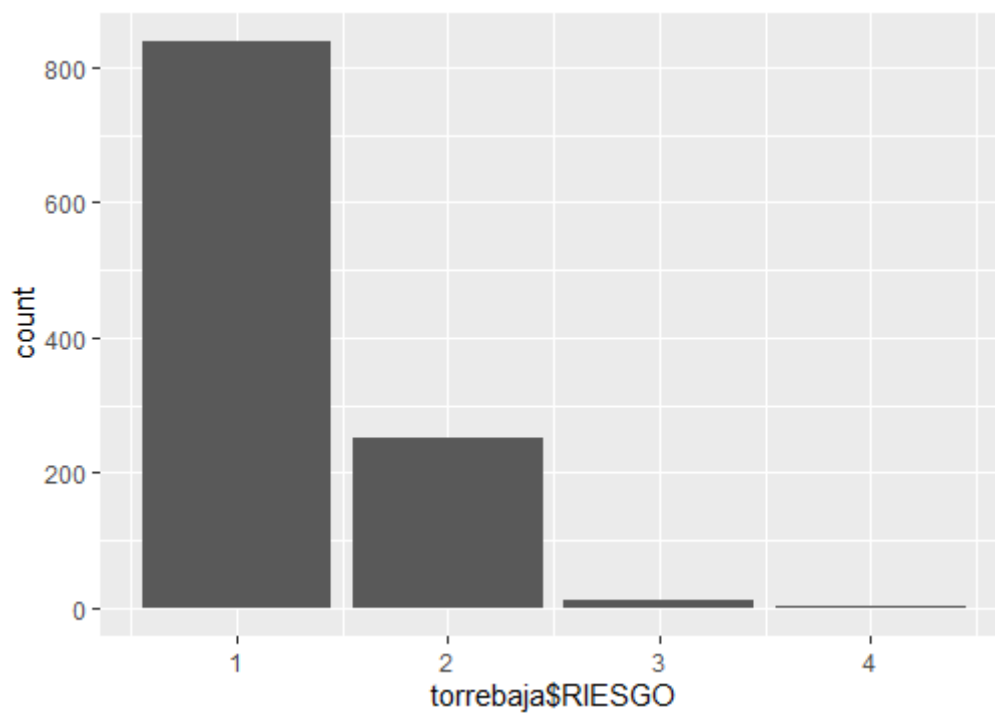


Figura 8.1: Histograma

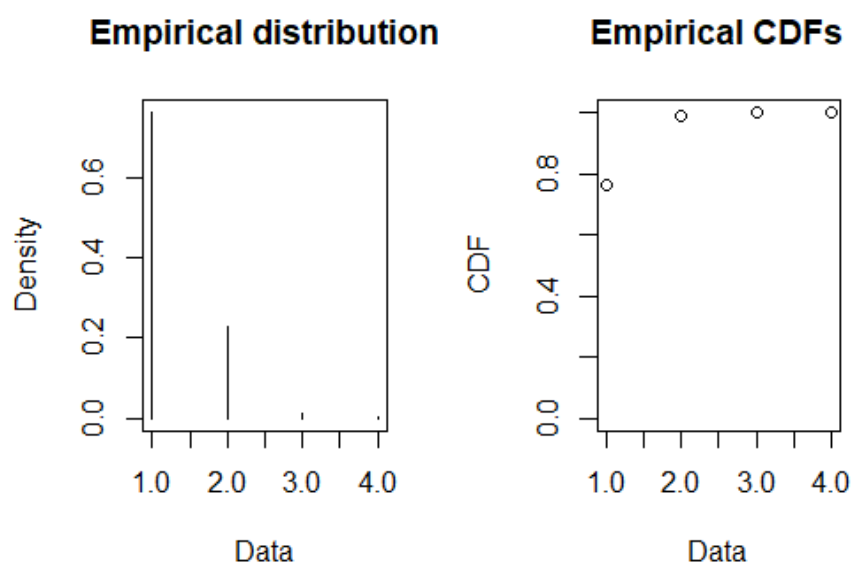


Figura 8.2: Gráfico de la función de distribución

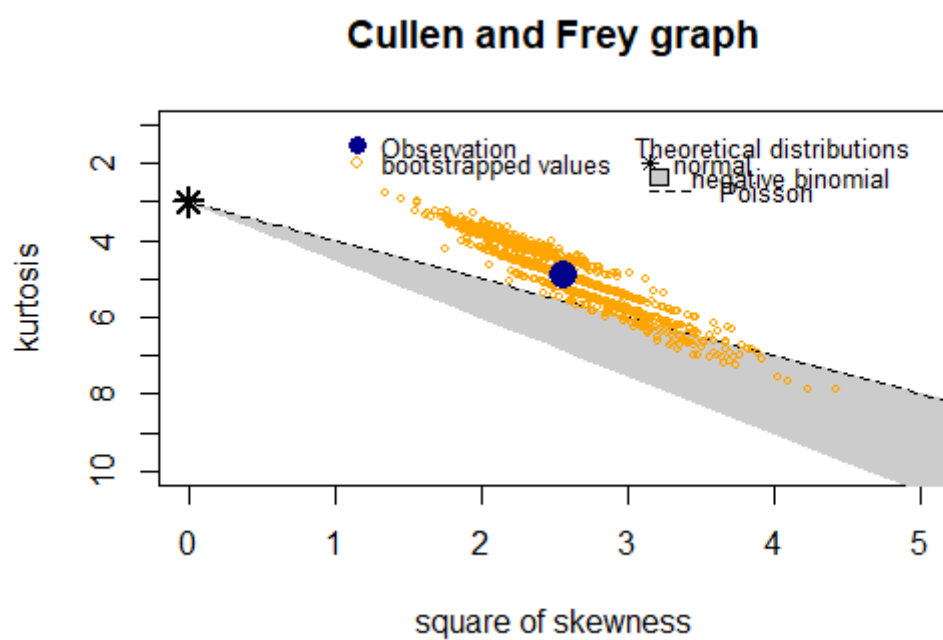


Figura 8.3: Gráfico de Cullen&Frey

La asimetría es de **1.41**.

No se ha encontrado una distribución que ajuste bien.

Distribución	P-value KS
Poisson	2.716122e-20
Geométrica	1.093268e-193

8.1. Análisis del ajuste de los datos a la distribución multinomial

La tabla de proporciones y frecuencias para cada serie de años es la siguiente:

2014-2017	1	2	3	4	5
Frecuencia	494	194	47	0	0
Proporción	0,67	0,26	0,06	0	0
2018-2019	1	2	3	4	5
Frecuencia	259	83	22	2	0
Proporción	0,71	0,23	0,06	0,005	0

El p-valor obtenido es de **2.2e-16**. No aceptamos H0.

Capítulo 9

ANÁLISIS DEL RESTO DE LOCALIZACIONES

9.1. Alhama de Murcia

9.1.1. Años completos

	1	2	3	4	5
Frecuencia	982	591	368	146	100
Proporción	0,45	0,27	0,17	0,07	0,05

La situación en Alhama de Murcia es parecida a Santa Elena.

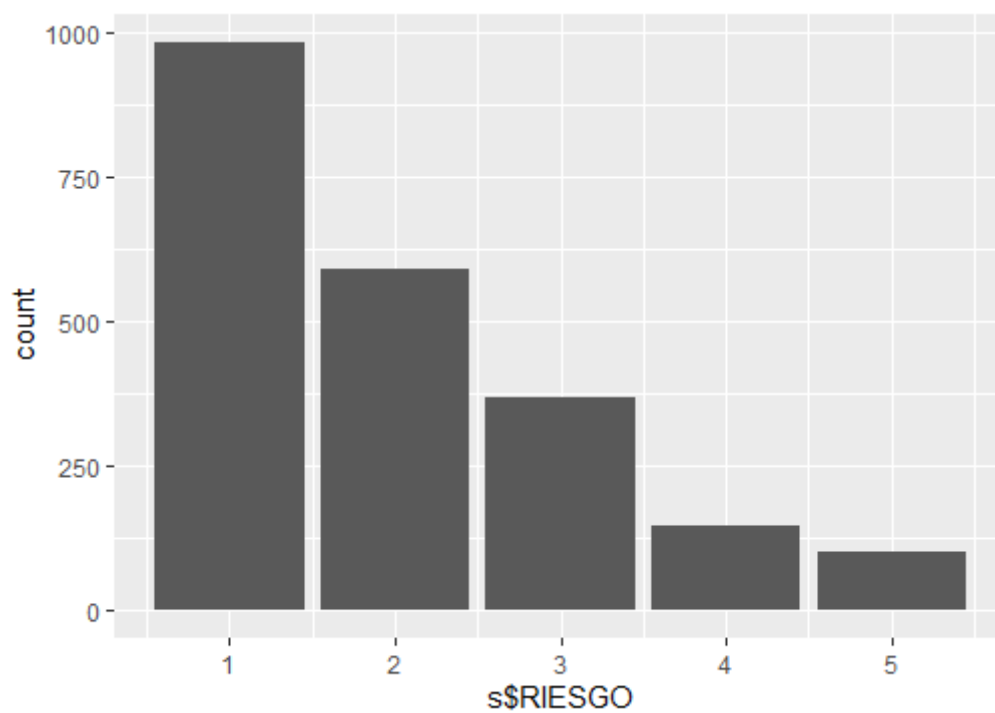


Figura 9.1: Histograma

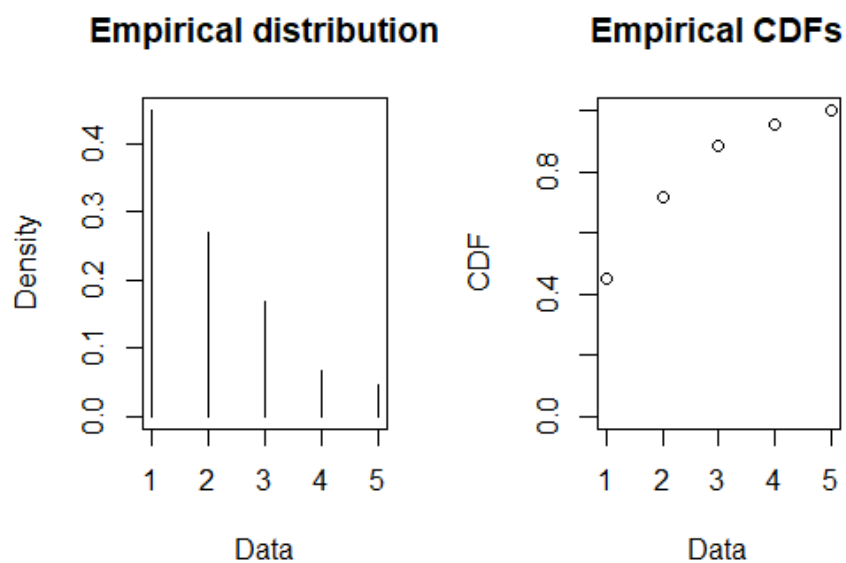


Figura 9.2: Gráfico de la función de distribución

La asimetría es de **1,04**.

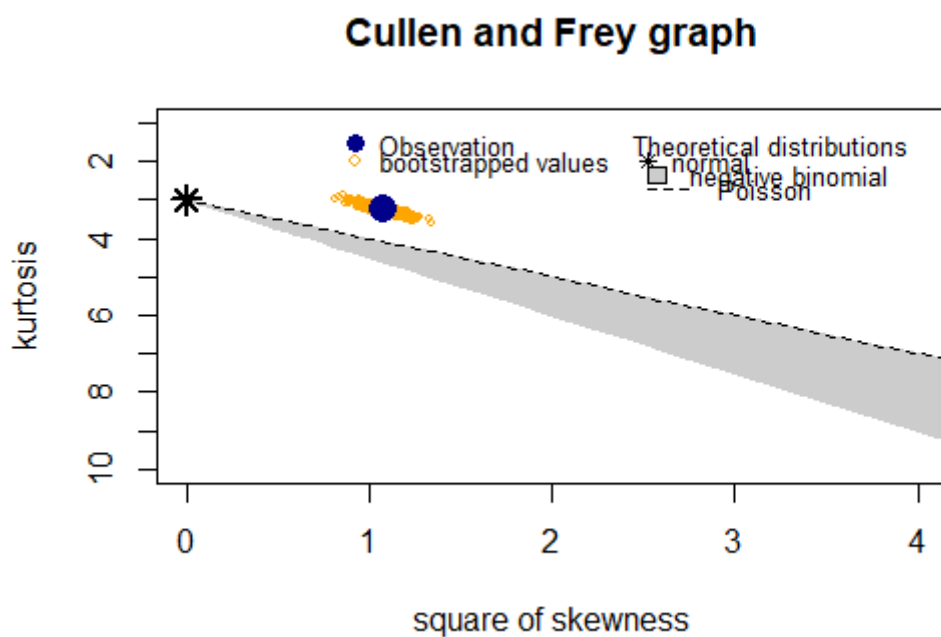


Figura 9.3: Gráfico de Cullen&Frey

El gráfico de Cullen&Frey no aclara la situación.
Las distribuciones probadas no se han aceptado.

Distribución	P-value KS
Poisson	3.279285e-13
Geométrica	1.036725e-120

Análisis del ajuste de los datos a la distribución multinomial La tabla con las frecuencias y proporciones de cada serie de años es la siguiente:

2014-2017	1	2	3	4	5
Frecuencia	603	412	264	110	70
Proporción	0,41	0,28	0,18	0,08	0,05

2018-2019	1	2	3	4	5
Frecuencia	379	179	104	36	30
Proporción	0,52	0,25	0,14	0,05	0,04

El p-valor obtenido es de **2.028e-07**. No aceptamos H0.

9.1.2. Temporada alta

	1	2	3	4	5
Frecuencia	319	274	273	135	85
Proporción	0,29	0,25	0,25	0,12	0,08

Los datos se distribuyen de forma homogénea entre las primera categorías.

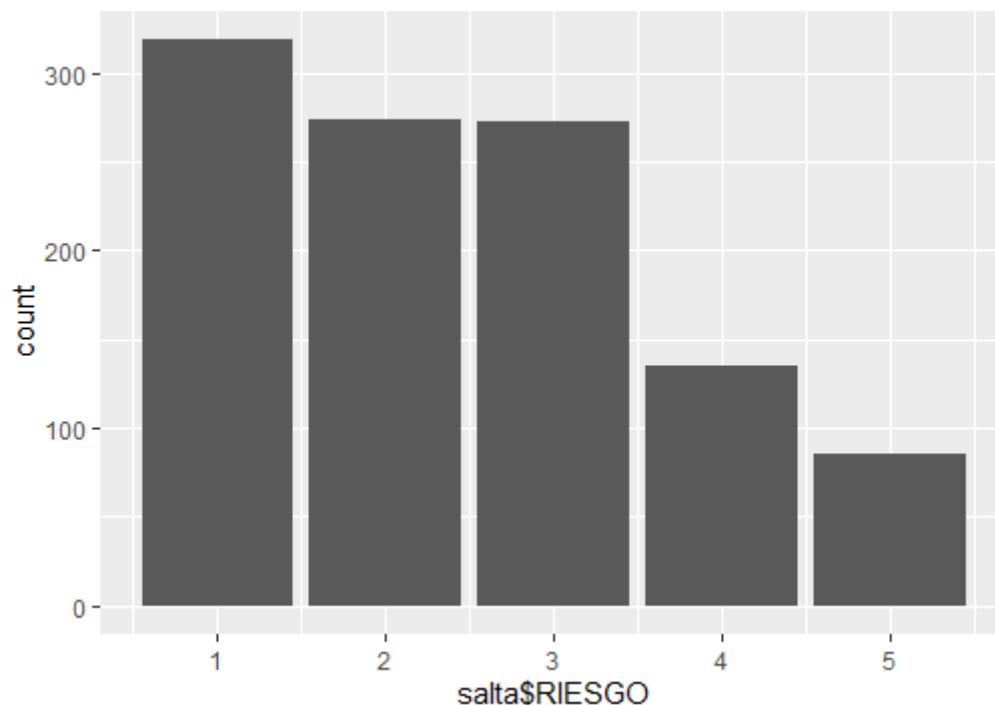


Figura 9.4: Histograma

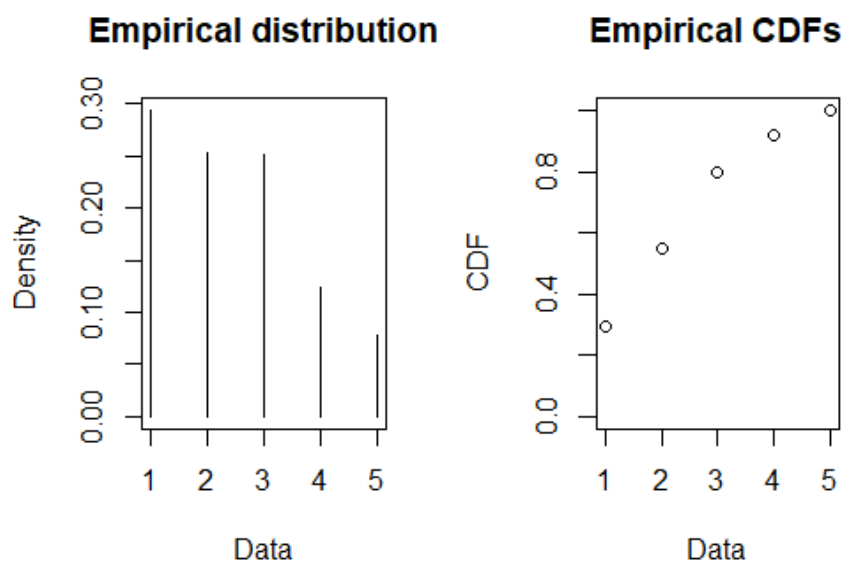


Figura 9.5: Gráfico de la función de distribución

La asimetría es de **0,48**.

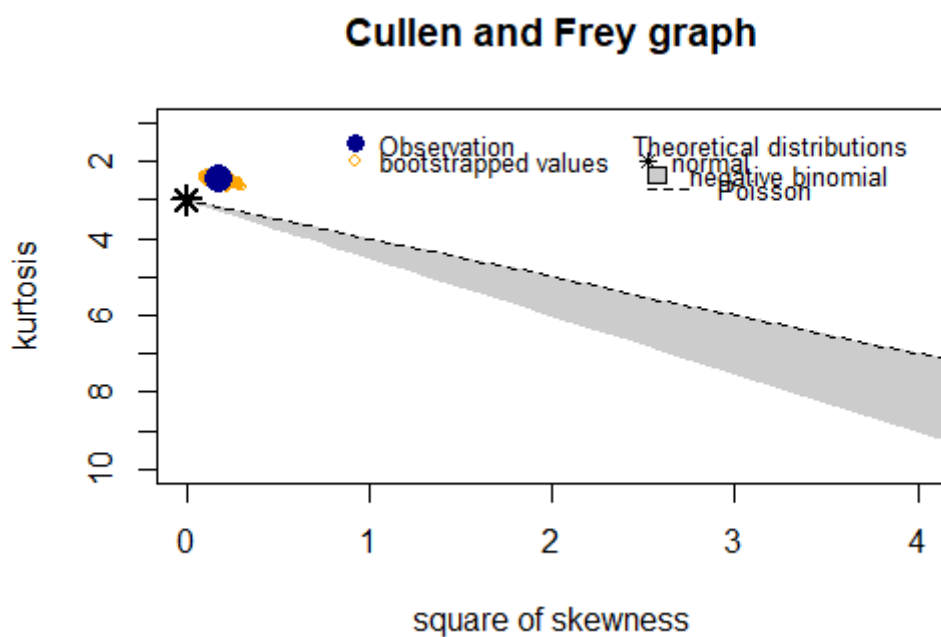


Figura 9.6: Gráfico de Cullen&Frey

Las distribuciones utilizadas no se ajustan.

Distribución	P-value KS
Normal	1.627927e-72
Poisson	4.269437e-11
Geométrica	7.238039e-127

Análisis del ajuste de los datos a la distribución multinomial La tabla con las frecuencias y proporciones de cada serie de años es la siguiente:

2014-2017	1	2	3	4	5
Frecuencia	173	192	202	100	57
Proporción	0,25	0,27	0,28	0,14	0,08

2018-2019	1	2	3	4	5
Frecuencia	146	82	71	35	28
Proporción	0,40	0,23	0,20	0,10	0,08

El p-valor obtenido es de **1.662e-11**. No aceptamos H0.

9.1.3. Temporada baja

	1	2	3	4	5
Frecuencia	663	317	95	11	15
Proporción	0,60	0,29	0,09	0,01	0,01

La distribución de los datos es la siguiente:

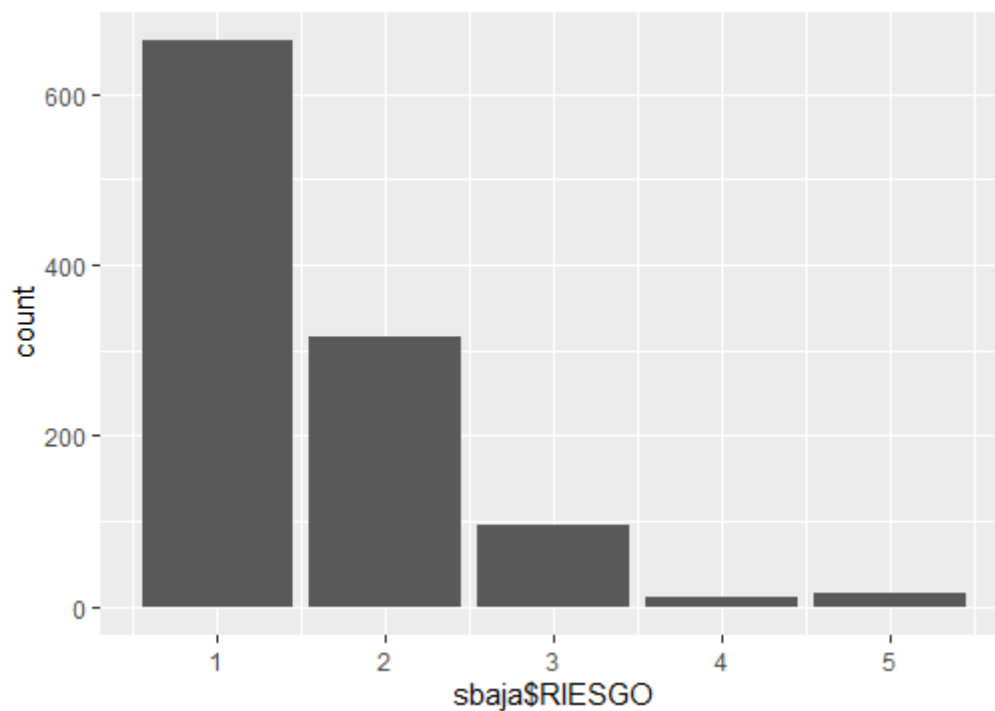


Figura 9.7: Histograma

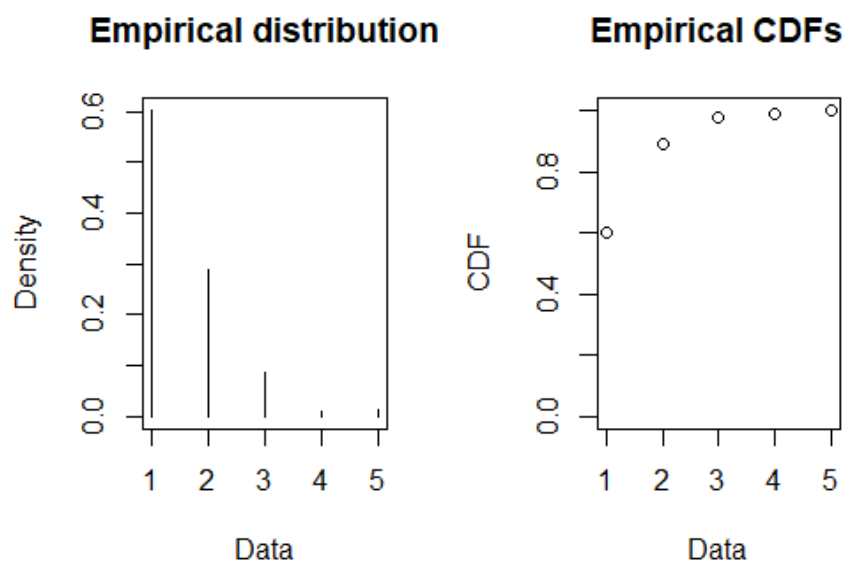


Figura 9.8: Gráfico de la función de distribución

La asimetría es de **1,75**.

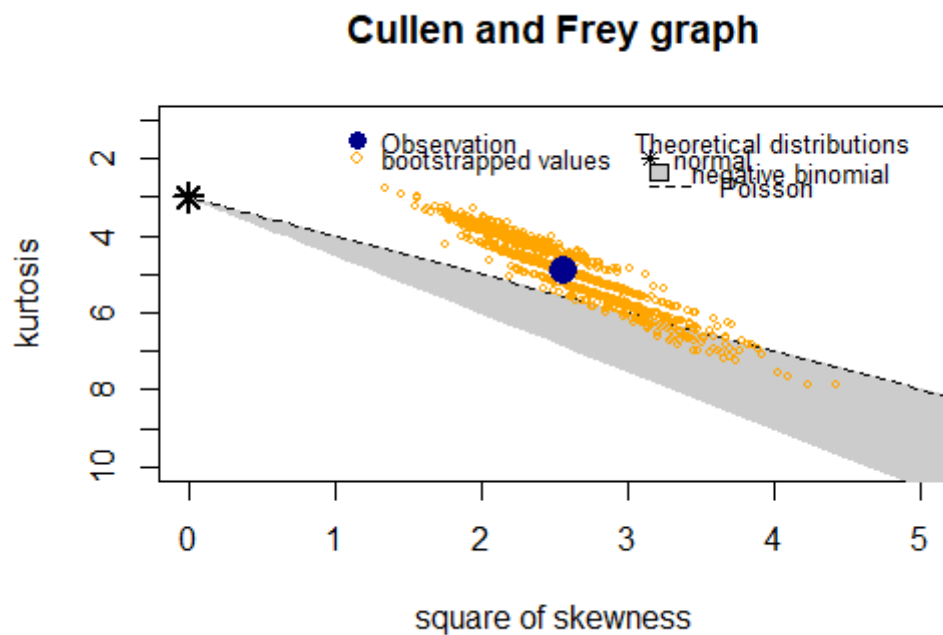


Figura 9.9: Gráfico de Cullen&Frey

No se ajusta ninguna distribución utilizada.

Distribución	P-value KS
Poisson	1.248985e-14
Geométrica	7.238039e-127

Análisis del ajuste de los datos a la distribución multinomial La tabla con las frecuencias y proporciones de cada serie de años es la siguiente:

2014-2017	1	2	3	4	5
Frecuencia	430	220	62	10	13
Proporción	0,59	0,30	0,08	0,01	0,01

2018-2019	1	2	3	4	5
Frecuencia	233	97	33	1	2
Proporción	0,63	0,27	0,09	0,003	0,005

El p-valor obtenido es de **0.04933**. Está muy cerca de aceptarse, pero hay que tener en cuenta que hay categorías que tienen muy pocos valores (menores a 5 implica que la bondad de ajuste no sea correcta), por lo que aceptar H_0 sería un error.

9.2. Navascués/Nabaskoze

9.2.1. Años completos

La tabla de frecuencias y proporciones para los años completos es la siguiente:

	1	2	3	4	5
Frecuencia	1051	586	356	124	70
Proporción	0,48	0,27	0,16	0,06	0,03

La distribución de los datos es la siguiente:

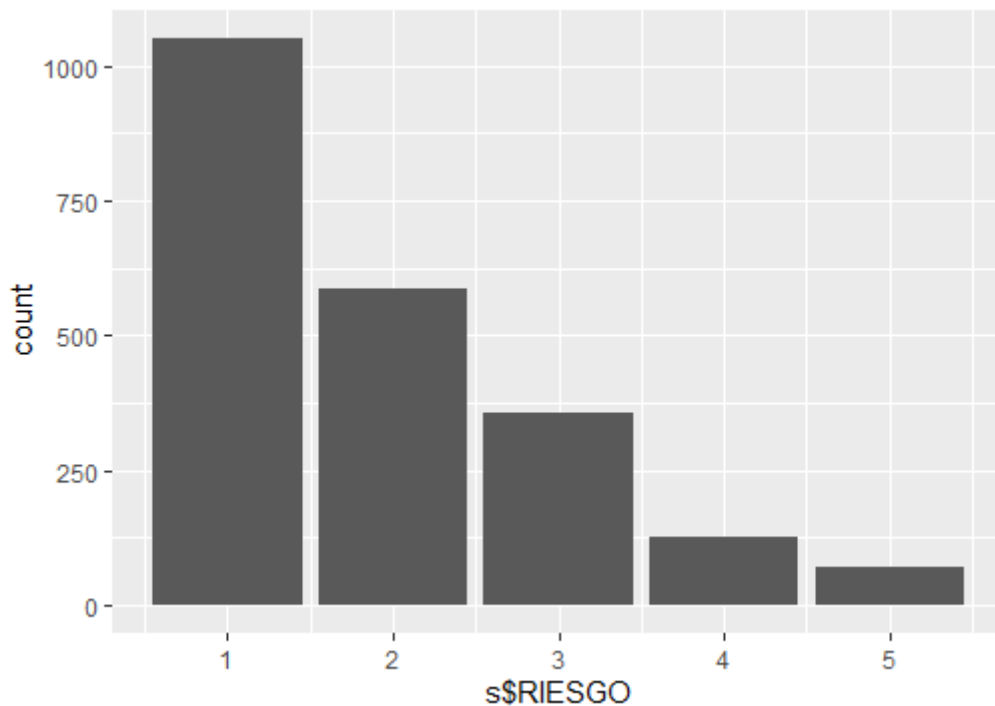


Figura 9.10: Histograma

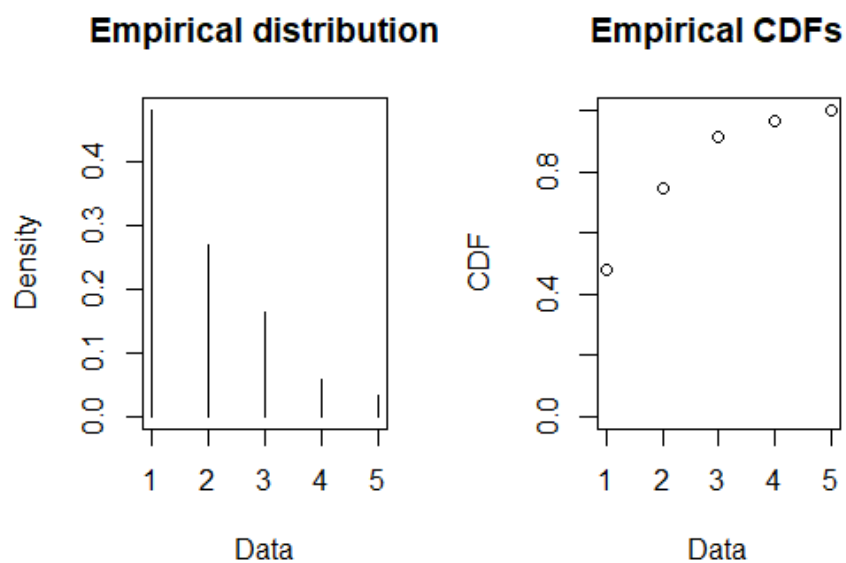


Figura 9.11: Gráfico de la función de distribución

La asimetría es de **1,12**.

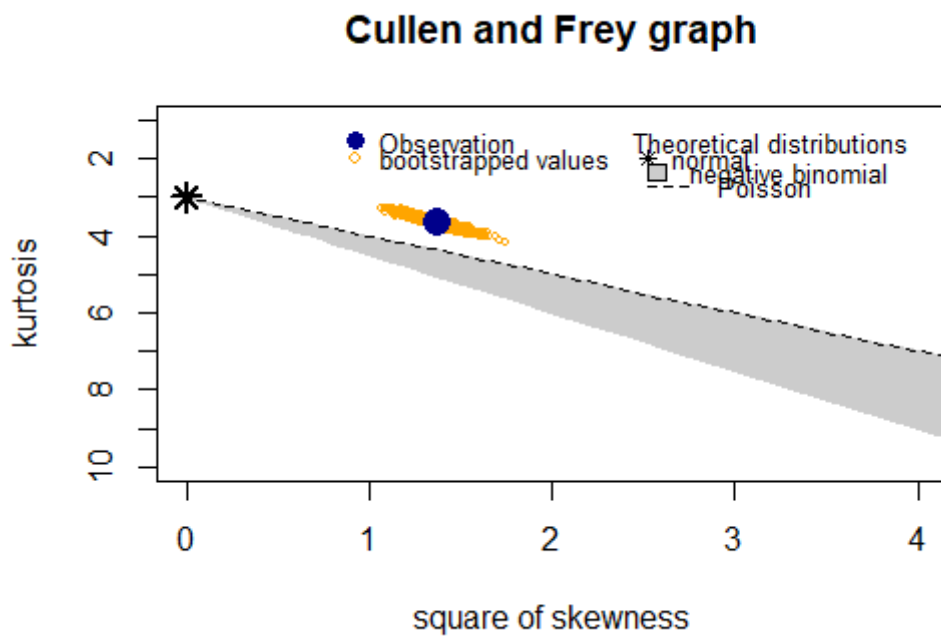


Figura 9.12: Gráfico de Cullen&Frey

No se ajusta ninguna distribución utilizada.

Distribución	P-value KS
Poisson	4.052515e-07
Geométrica	7.840118e-105

Análisis del ajuste de los datos a la distribución multinomial La tabla con las frecuencias y proporciones de cada serie de años es la siguiente:

2014-2017	1	2	3	4	5
Frecuencia	667	425	234	81	52
Proporción	0,46	0,29	0,16	0,06	0,04

2018-2019	1	2	3	4	5
Frecuencia	384	161	122	43	18
Proporción	0,53	0,22	0,17	0,06	0,02

El p-valor obtenido es **0.0001265**. No aceptamos H_0 .

9.2.2. Temporada alta

	1	2	3	4	5
Frecuencia	330	318	251	117	70
Proporción	0,30	0,29	0,23	0,11	0,06

Los datos se distribuyen de forma escalonada.

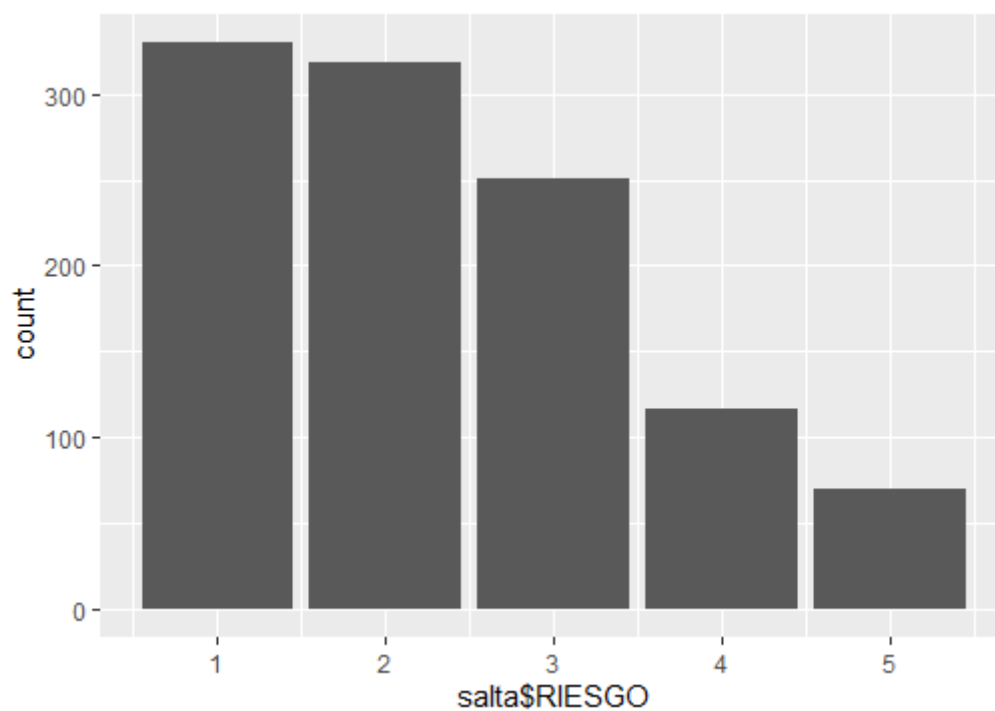


Figura 9.13: Histograma

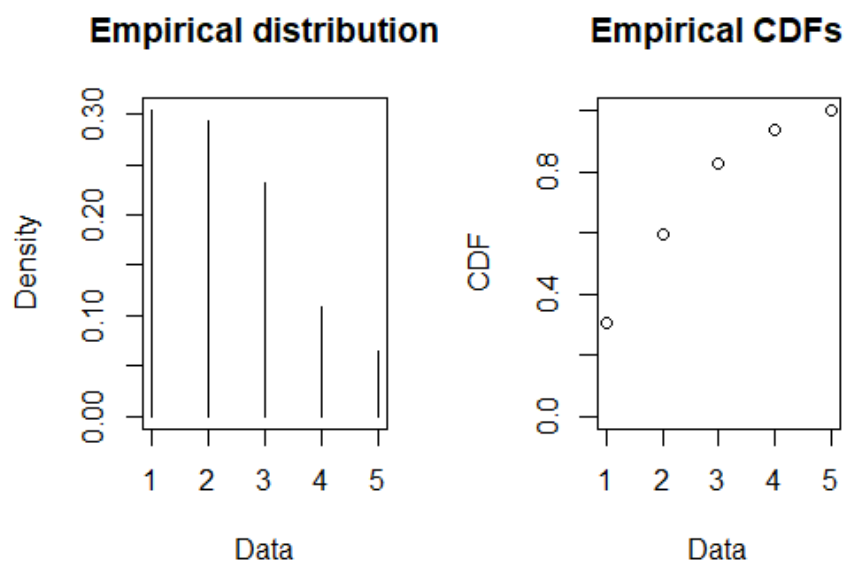


Figura 9.14: Gráfico de la función de distribución

La asimetría es de **0,61**.

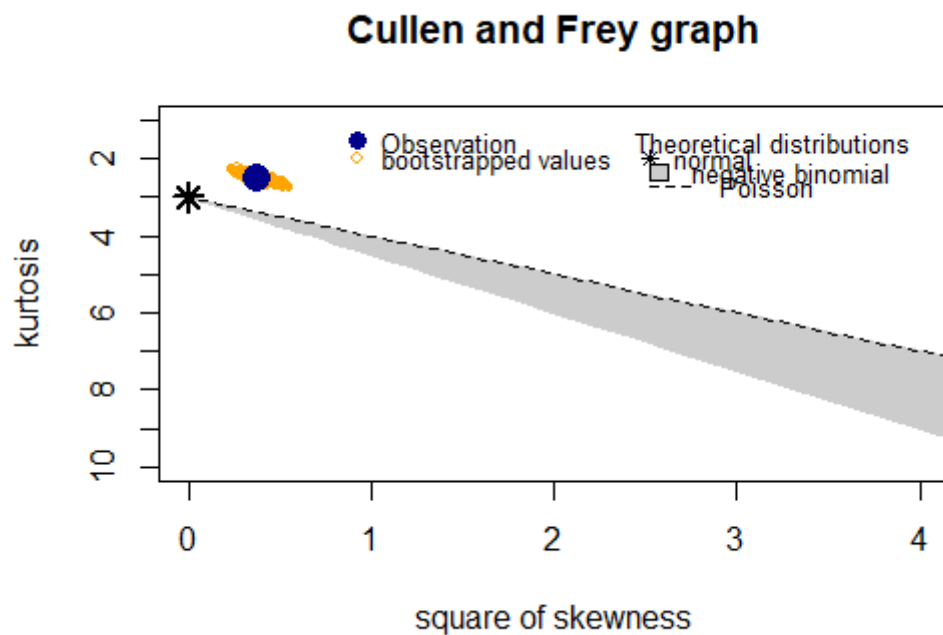


Figura 9.15: Gráfico de Cullen&Frey

No se ajusta ninguna distribución utilizada.

Distribución	P-value KS
Poisson	3.072559e-09
Geométrica	1.605736e-122

Análisis del ajuste de los datos a la distribución multinomial La tabla con las frecuencias y proporciones de cada serie de años es la siguiente:

2014-2017	1	2	3	4	5
Frecuencia	207	225	166	74	52
Proporción	0,29	0,31	0,23	0,10	0,07

2018-2019	1	2	3	4	5
Frecuencia	123	93	85	43	18
Proporción	0,34	0,26	0,23	0,12	0,05

El p-valor obtenido es **0.0323**. Está más cerca de 0,05, pero no podemos aceptar H0.

9.2.3. Temporada baja

	1	2	3	4	5
Frecuencia	721	268	105	7	0
Proporción	0,65	0,24	0,10	0,06	0

No existen valores en la categoría 5.

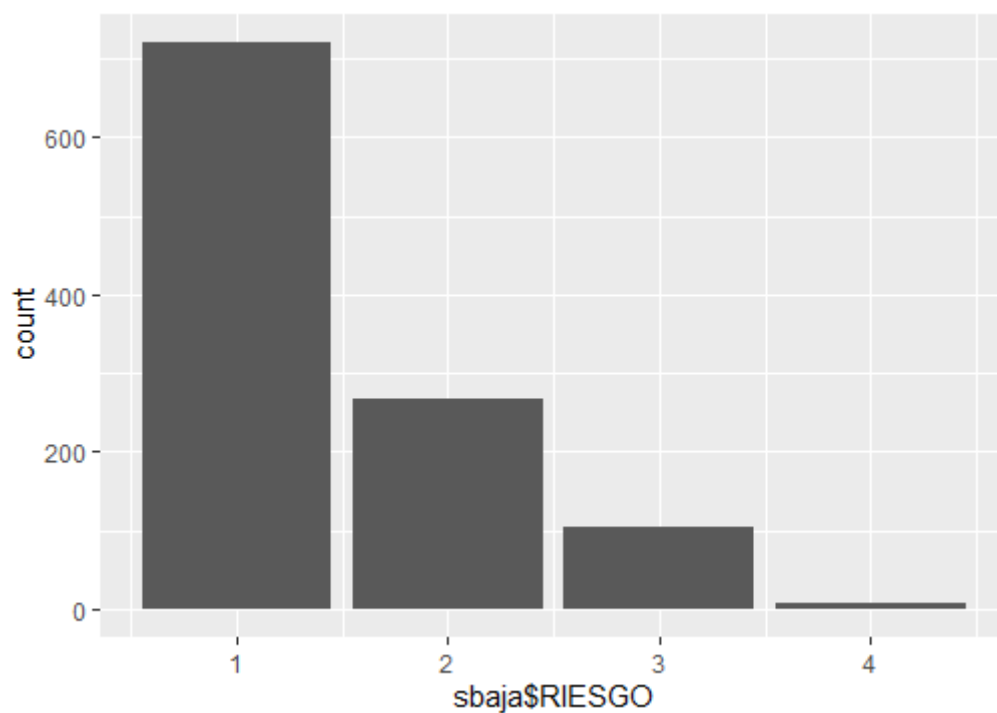


Figura 9.16: Histograma

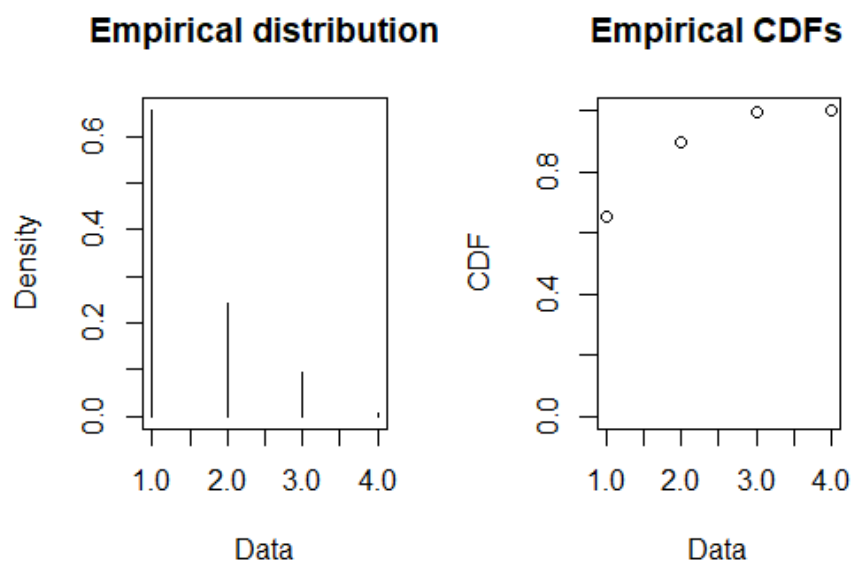


Figura 9.17: Gráfico de la función de distribución

La asimetría es de **1,32**.

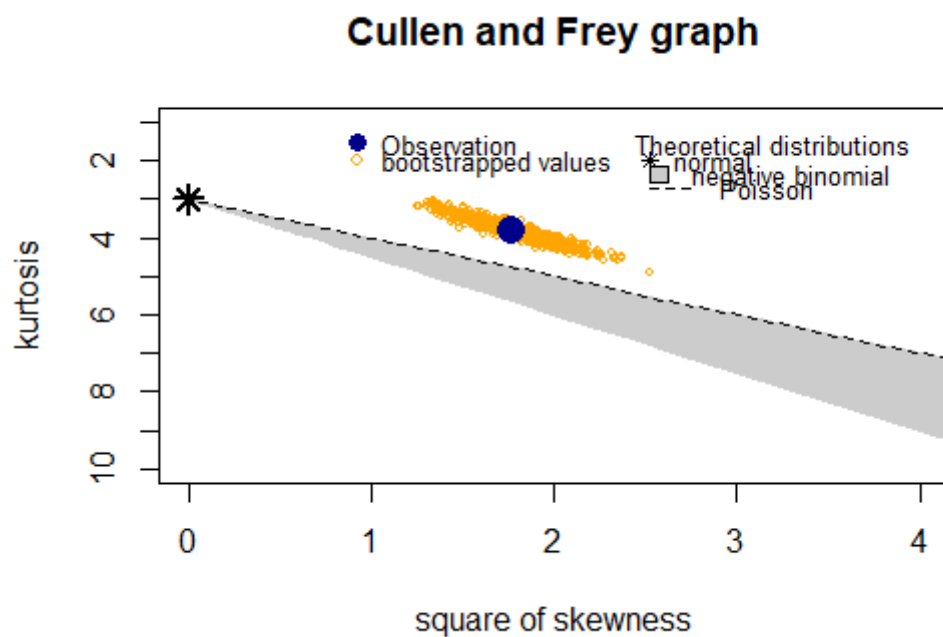


Figura 9.18: Gráfico de Cullen&Frey

No se ajusta ninguna distribución utilizada.

Distribución	P-value KS
Poisson	4.152194e-16
Geométrica	1.605736e-122

Análisis del ajuste de los datos a la distribución multinomial La tabla con las frecuencias y proporciones de cada serie de años es la siguiente:

2014-2017	1	2	3	4	5
Frecuencia	460	200	68	7	0
Proporción	0,63	0,27	0,09	0,01	0

2018-2019	1	2	3	4	5
Frecuencia	261	68	37	0	0
Proporción	0,71	0,19	0,10	0	0

El p-valor obtenido es **0.0003903**. No podemos aceptar H0.

9.3. Los Tojos

9.3.1. Años completos

La tabla de frecuencias y proporciones para los años completos es la siguiente:

La tabla de frecuencias y proporciones para los años completos es la siguiente:

	1	2	3	4	5
Frecuencia	929	567	380	183	128
Proporción	0,42	0,26	0,17	0,08	0,06

La distribución de los datos es la siguiente:

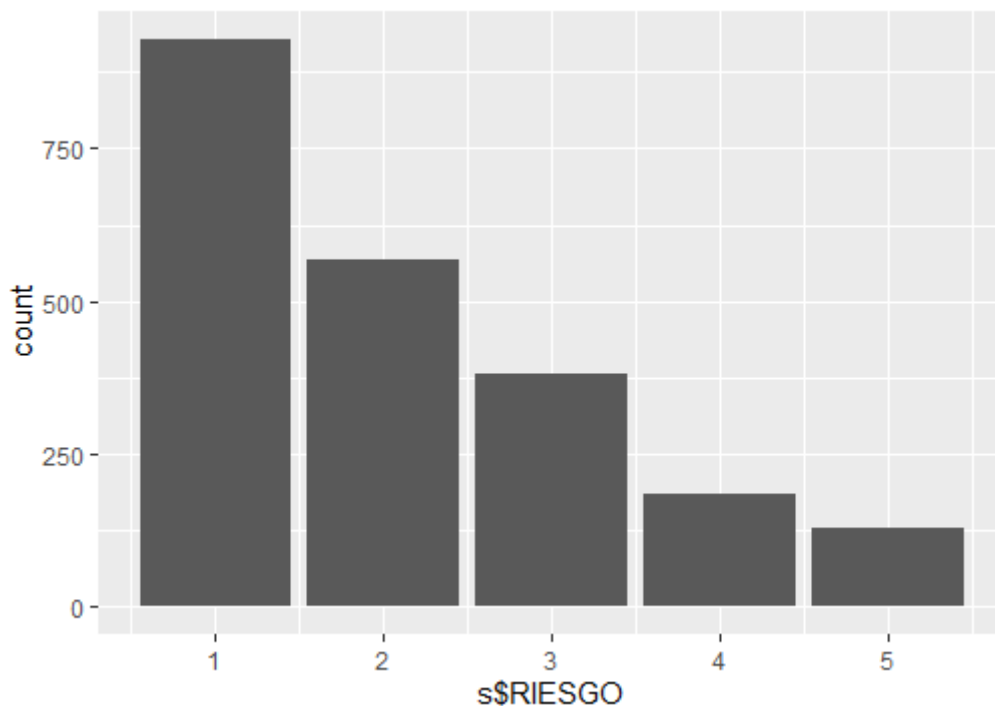


Figura 9.19: Histograma

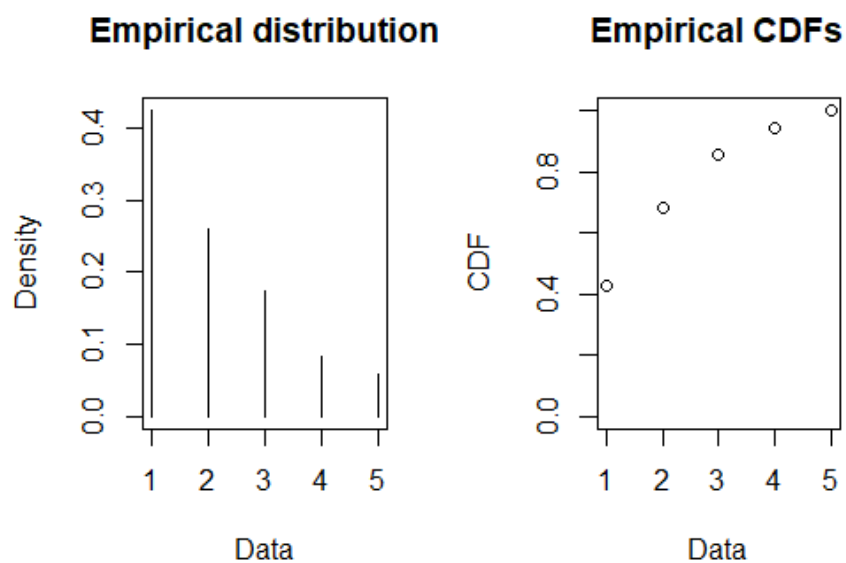


Figura 9.20: Gráfico de la función de distribución

La asimetría es de **0,91**.

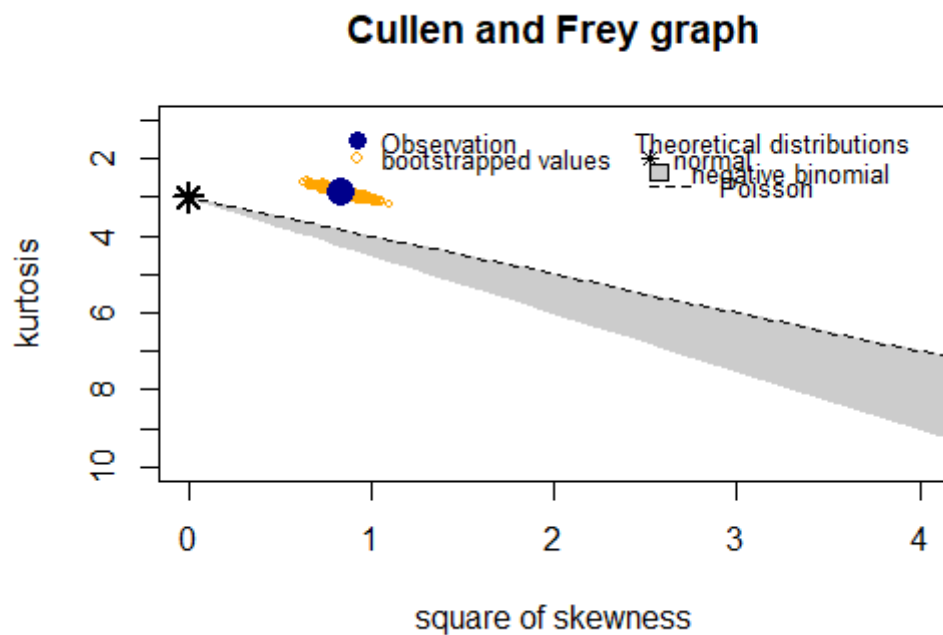


Figura 9.21: Gráfico de Cullen&Frey

No se ajusta ninguna distribución utilizada.

Distribución	P-value KS
Poisson	3.060549e-16
Geométrica	1.113985e-124

Análisis del ajuste de los datos a la distribución multinomial La tabla con las frecuencias y proporciones de cada serie de años es la siguiente:

2014-2017	1	2	3	4	5
Frecuencia	619	355	279	129	77
Proporción	0,42	0,24	0,19	0,09	0,05

2018-2019	1	2	3	4	5
Frecuencia	310	212	101	54	51
Proporción	0,43	0,29	0,14	0,07	0,07

El p-valor obtenido es **0.0001185**. No podemos aceptar H0.

9.3.2. Temporada alta

	1	2	3	4	5
Frecuencia	319	287	230	140	110
Proporción	0,29	0,26	0,21	0,13	0,10

Los datos se distribuyen de forma escalonada.

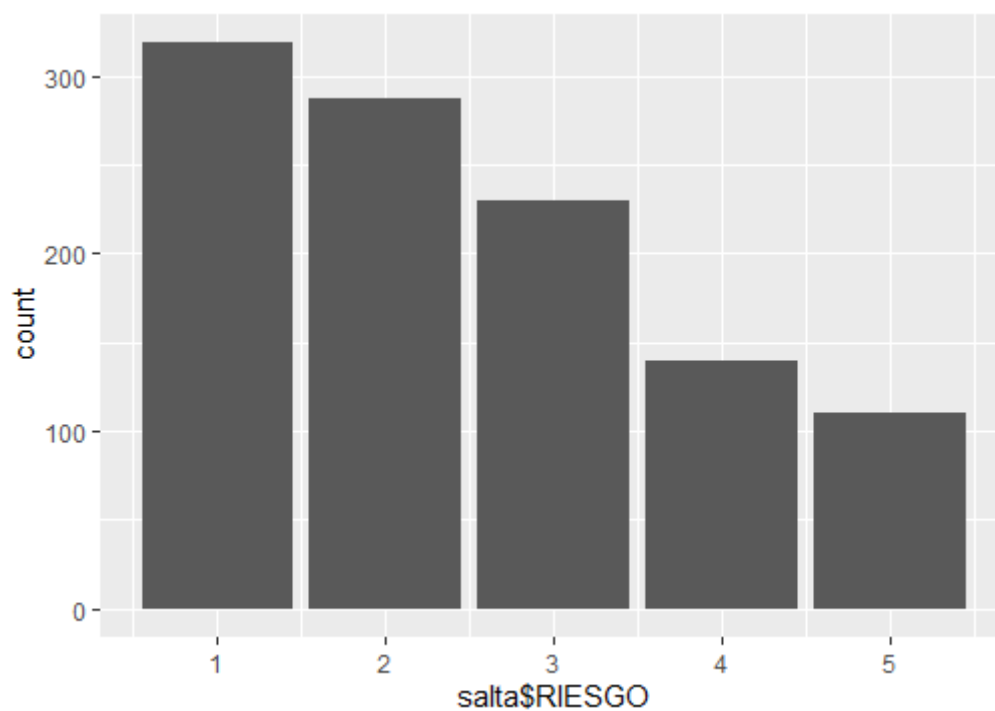


Figura 9.22: Histograma

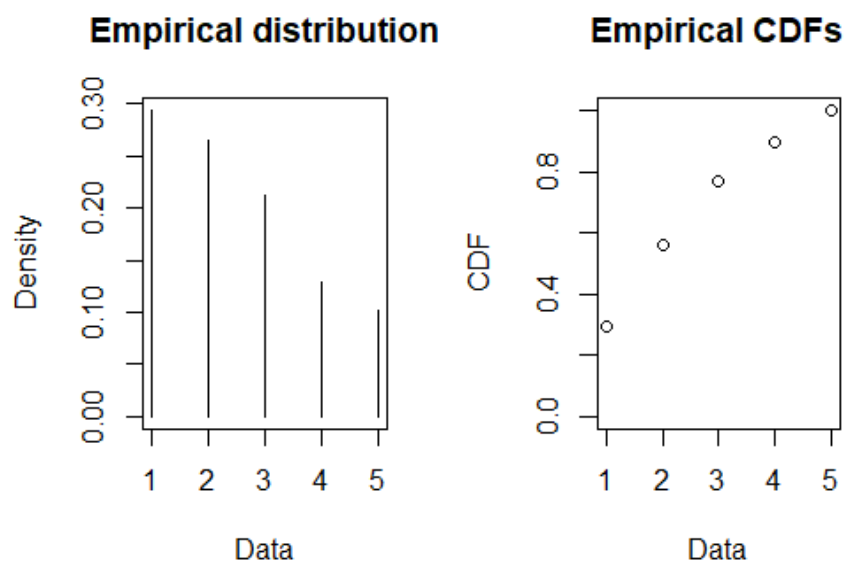


Figura 9.23: Gráfico de la función de distribución

La asimetría es de **0,51**.

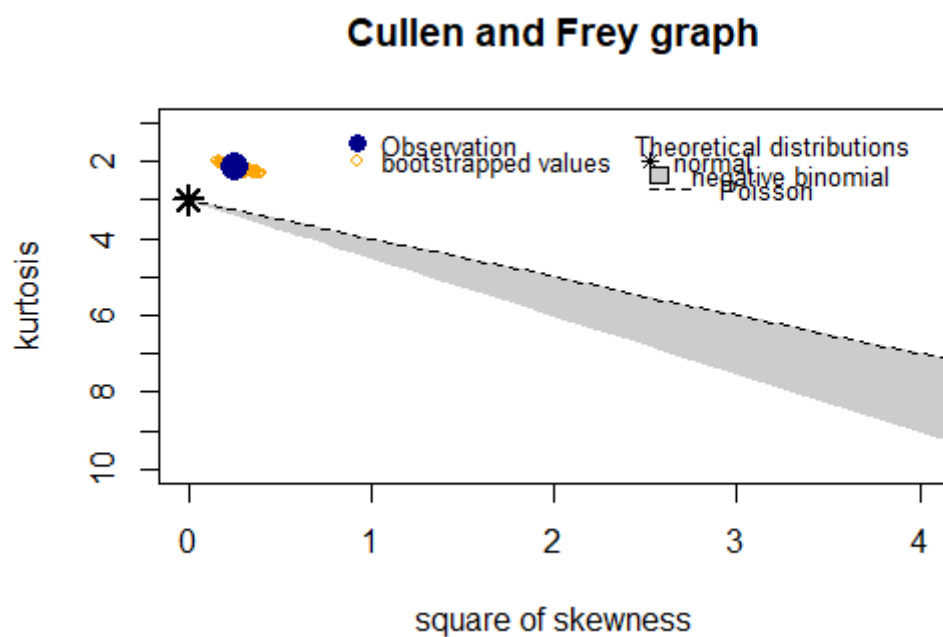


Figura 9.24: Gráfico de Cullen&Frey

No se ajusta ninguna distribución utilizada.

Distribución	P-value KS
Normal	8.253065e-73
Poisson	1.758241e-13
Geométrica	4.659165e-117

Análisis del ajuste de los datos a la distribución multinomial La tabla con las frecuencias y proporciones de cada serie de años es la siguiente:

2014-2017	1	2	3	4	5
Frecuencia	195	190	174	102	63
Proporción	0,27	0,26	0,24	0,14	0,09

2018-2019	1	2	3	4	5
Frecuencia	124	97	56	38	47
Proporción	0,34	0,27	0,15	0,10	0,13

El p-valor obtenido es **7.015e-06**. No podemos aceptar H0.

9.3.3. Temporada baja

	1	2	3	4	5
Frecuencia	610	280	150	43	18
Proporción	0,55	0,25	0,14	0,04	0,02

Más del 50 % de los datos están en la primera categoría.

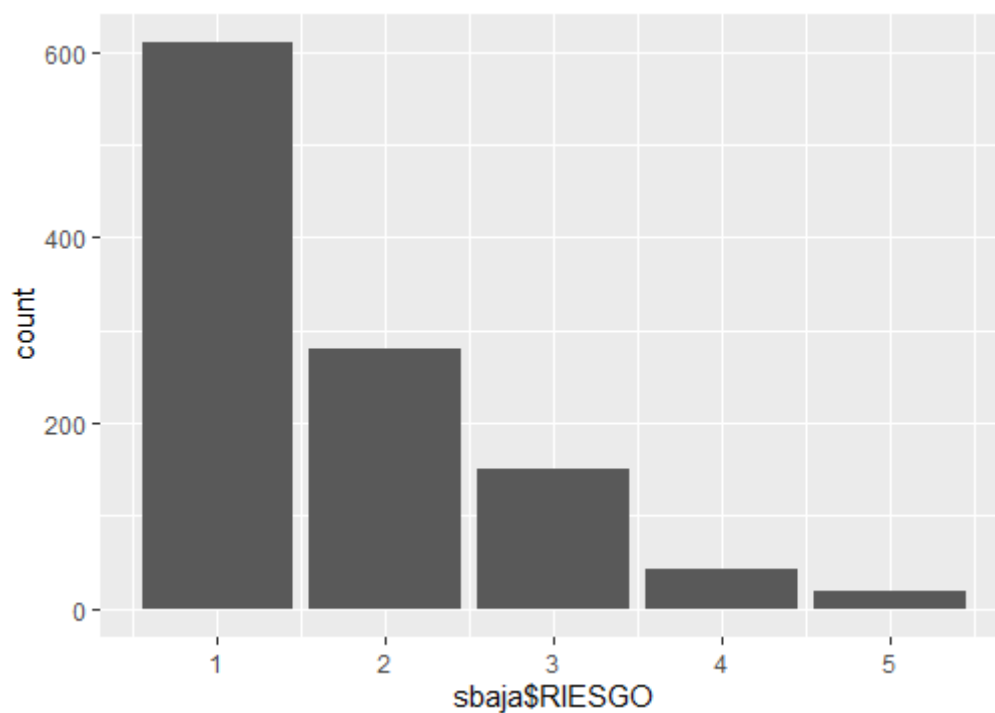


Figura 9.25: Histograma

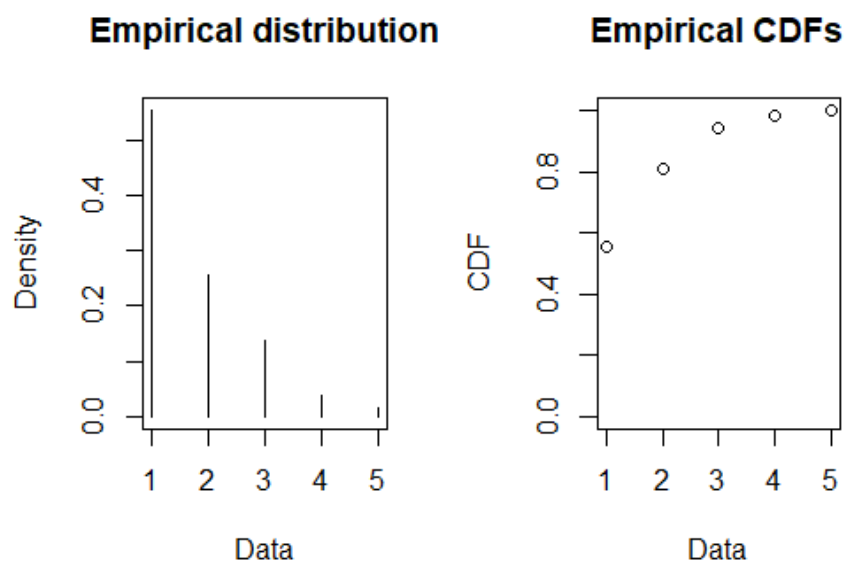


Figura 9.26: Gráfico de la función de distribución

La asimetría es de **1,33**.

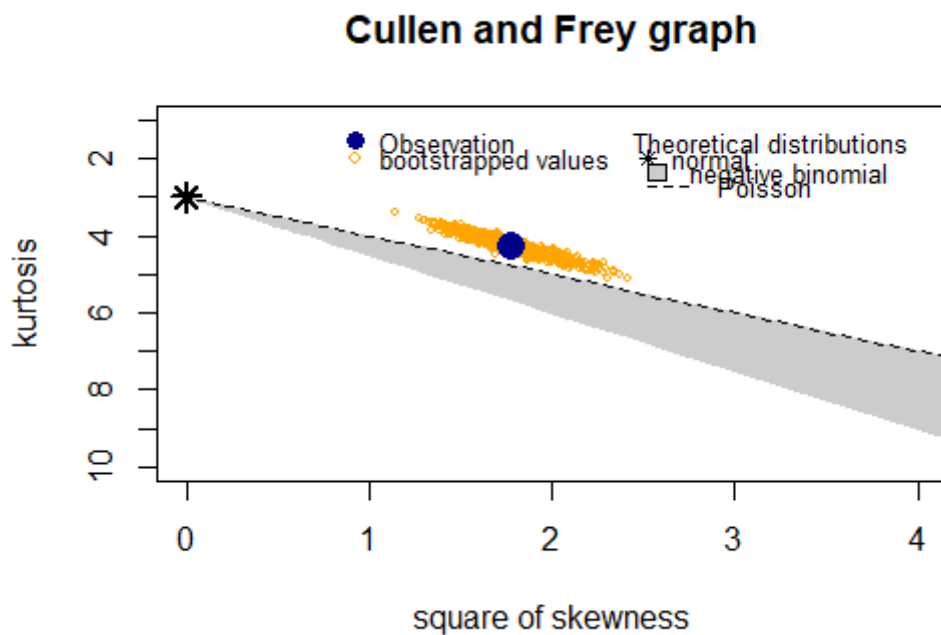


Figura 9.27: Gráfico de Cullen&Frey

No se ajusta ninguna distribución utilizada.

Distribución	P-value KS
Poisson	2.211995e-06
Geométrica	4.659165e-117

Análisis del ajuste de los datos a la distribución multinomial La tabla con las frecuencias y proporciones de cada serie de años es la siguiente:

2014-2017	1	2	3	4	5
Frecuencia	424	165	105	27	14
Proporción	0,58	0,22	0,14	0,04	0,02

2018-2019	1	2	3	4	5
Frecuencia	186	115	45	16	4
Proporción	0,51	0,37	0,12	0,04	0,01

El p-valor obtenido es **0.0008287**. No podemos aceptar H0.

9.4. Vimianzo

9.4.1. Años completos

La tabla de frecuencias y proporciones para los años completos es la siguiente:

La tabla de frecuencias y proporciones para los años completos es la siguiente:

	1	2	3	4	5
Frecuencia	1005	576	339	161	106
Proporción	0,46	0,26	0,16	0,07	0,05

La distribución de los datos es la siguiente:

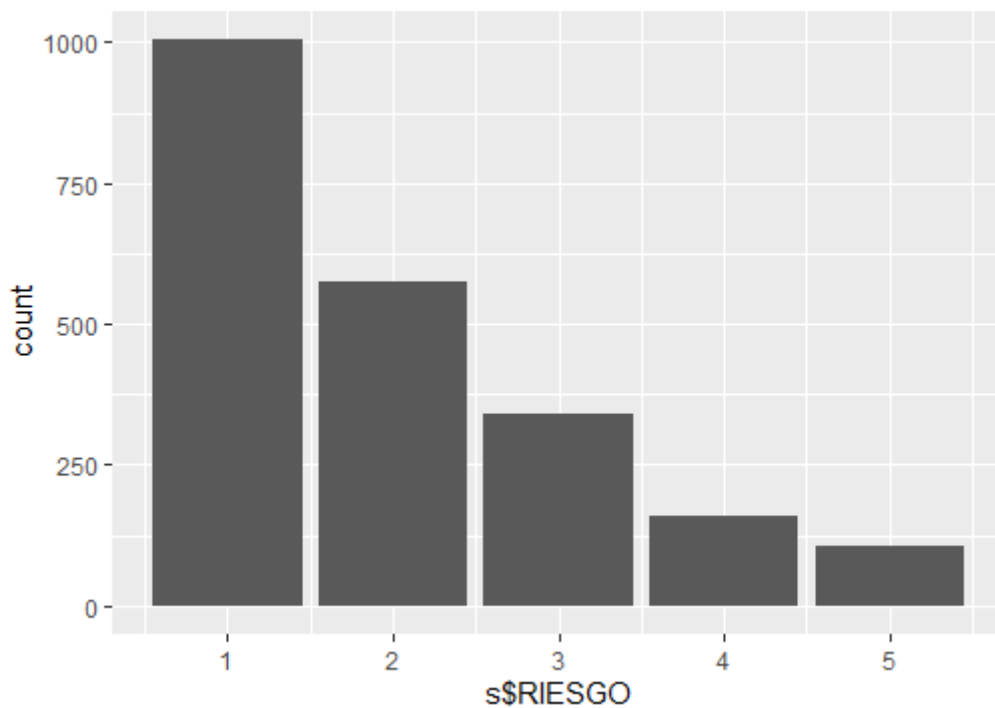


Figura 9.28: Histograma

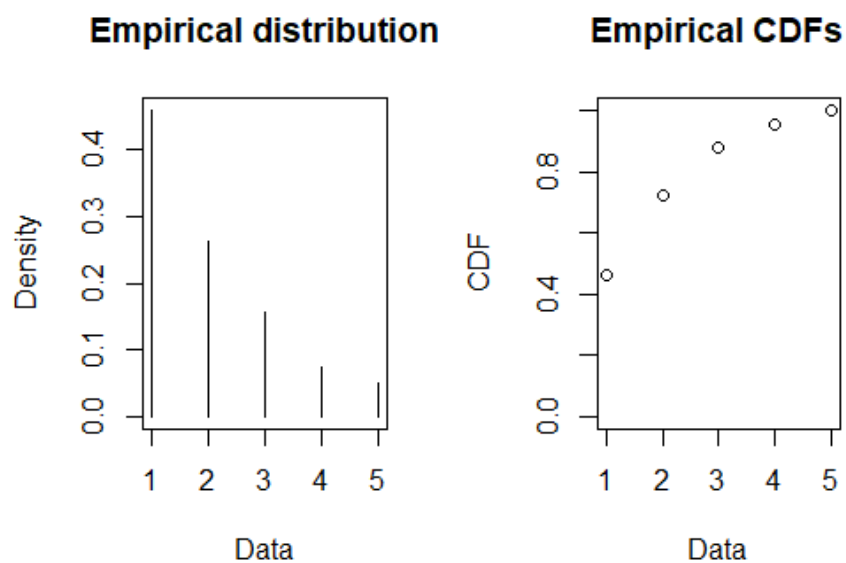


Figura 9.29: Gráfico de la función de distribución

La asimetría es de **1,05**.

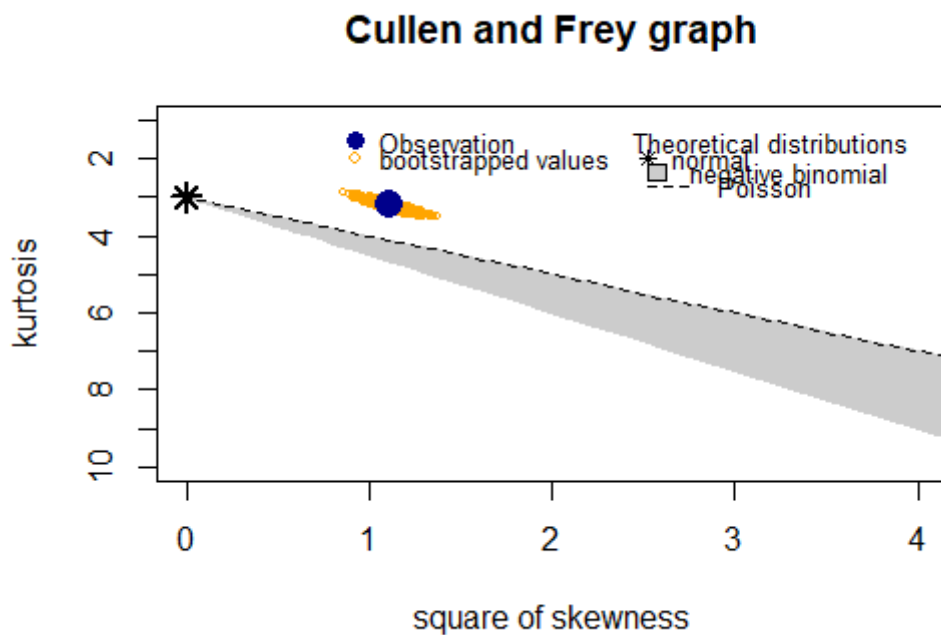


Figura 9.30: Gráfico de Cullen&Frey

No se ajusta ninguna distribución utilizada.

Distribución	P-value KS
Poisson	5.650479e-15
Geométrica	2.630314e-106

Análisis del ajuste de los datos a la distribución multinomial La tabla con las frecuencias y proporciones de cada serie de años es la siguiente:

2014-2017	1	2	3	4	5
Frecuencia	666	392	217	108	76
Proporción	0,46	0,27	0,15	0,07	0,05

2018-2019	1	2	3	4	5
Frecuencia	339	184	122	53	30
Proporción	0,46	0,25	0,17	0,07	0,04

El p-valor obtenido es **es 0.3757**. **No podemos rechazar H0.**

9.4.2. Temporada alta

	1	2	3	4	5
Frecuencia	329	308	240	125	84
Proporción	0,30	0,28	0,22	0,12	0,08

Los datos se distribuyen de forma escalonada.

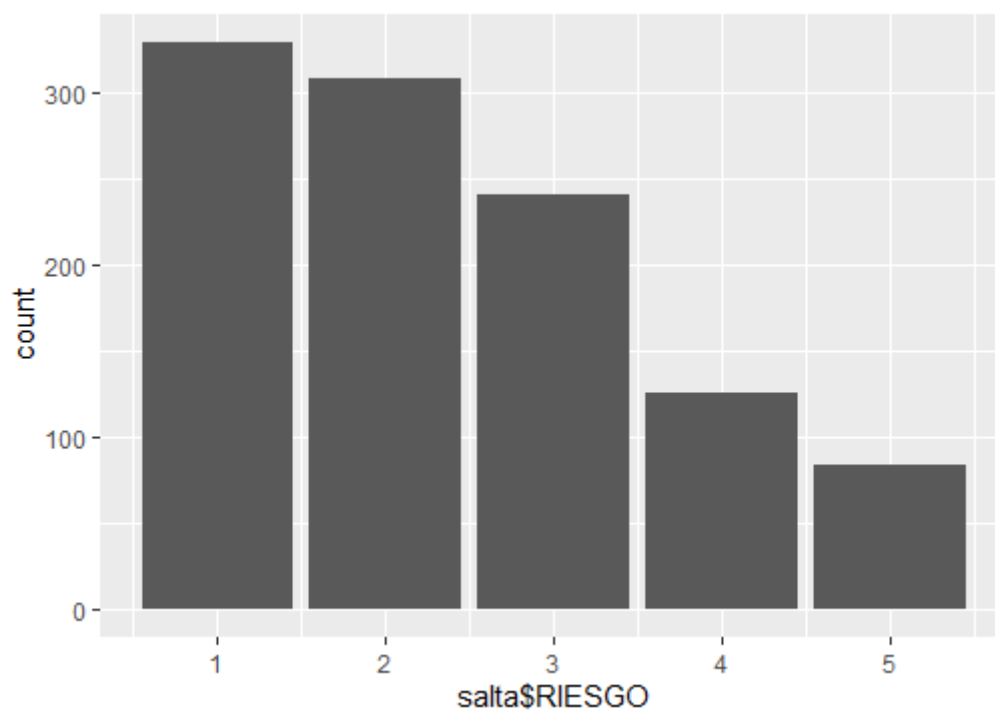


Figura 9.31: Histograma

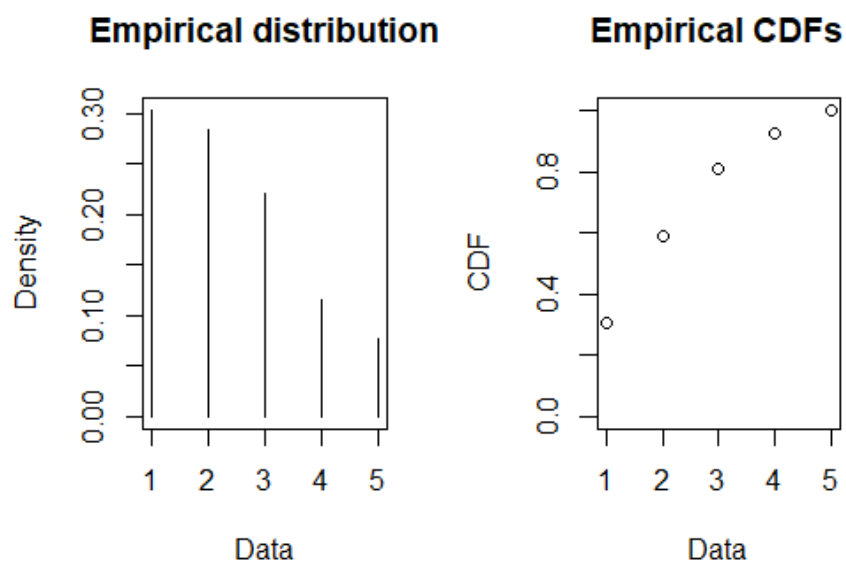


Figura 9.32: Gráfico de la función de distribución

La asimetría es de **0,59**.

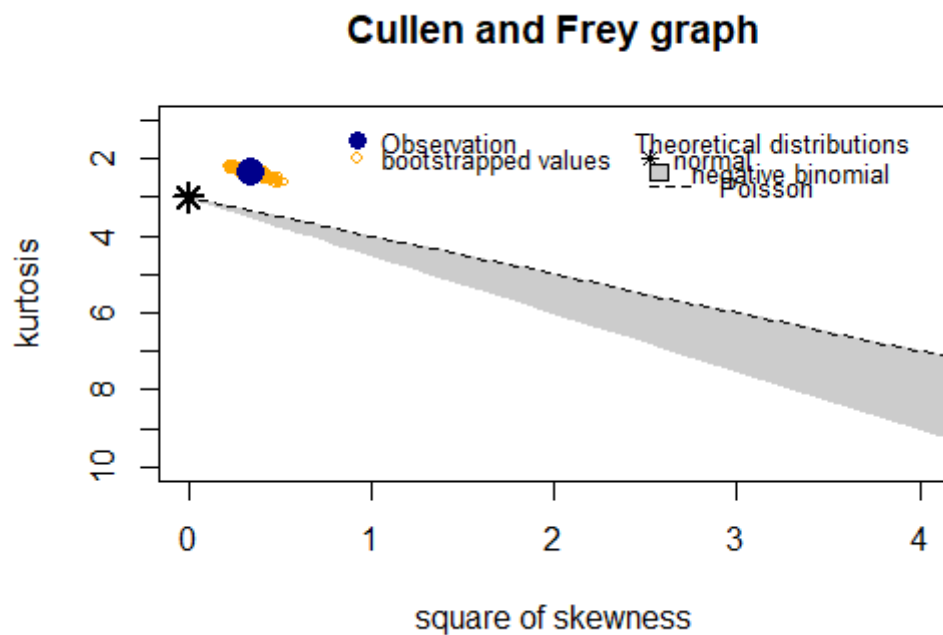


Figura 9.33: Gráfico de Cullen&Frey

No se ajusta ninguna distribución utilizada.

Distribución	P-value KS
Poisson	1.186933e-09
Geométrica	1.052006e-116

Análisis del ajuste de los datos a la distribución multinomial La tabla con las frecuencias y proporciones de cada serie de años es la siguiente:

2014-2017	1	2	3	4	5
Frecuencia	213	212	155	85	59
Proporción	0,29	0,29	0,21	0,12	0,08

2018-2019	1	2	3	4	5
Frecuencia	116	96	85	40	25
Proporción	0,32	0,27	0,23	0,11	0,07

El p-valor obtenido es **0.501**. **No podemos rechazar H0.**

9.4.3. Temporada baja

	1	2	3	4	5
Frecuencia	676	268	99	36	22
Proporción	0,61	0,24	0,09	0,03	0,02

Más del 60 % de los datos están en la primera categoría.

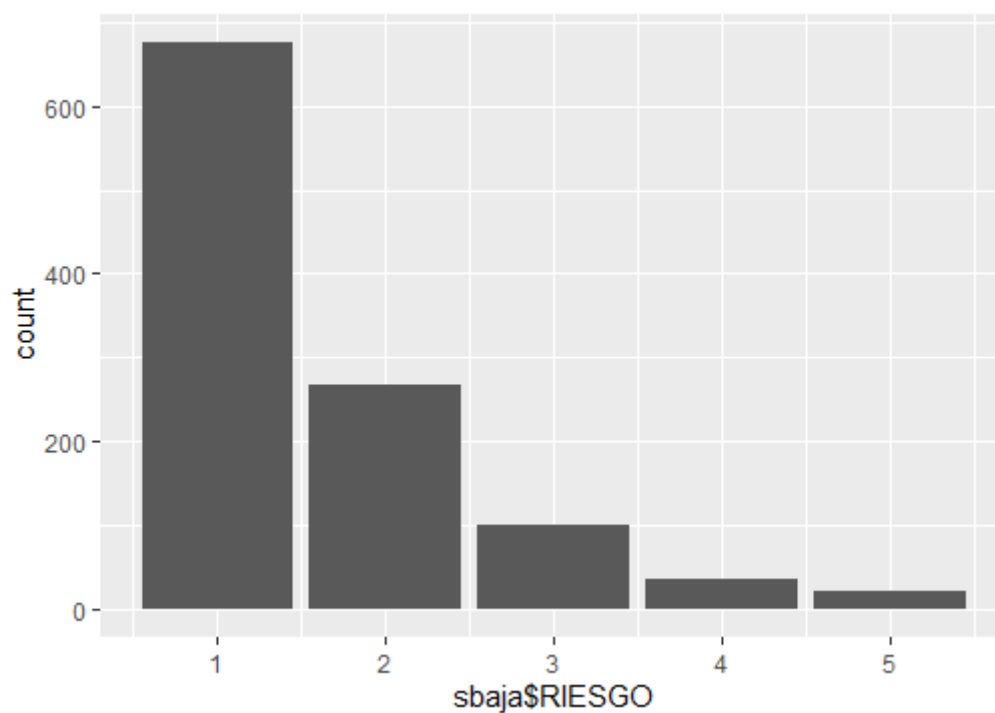


Figura 9.34: Histograma

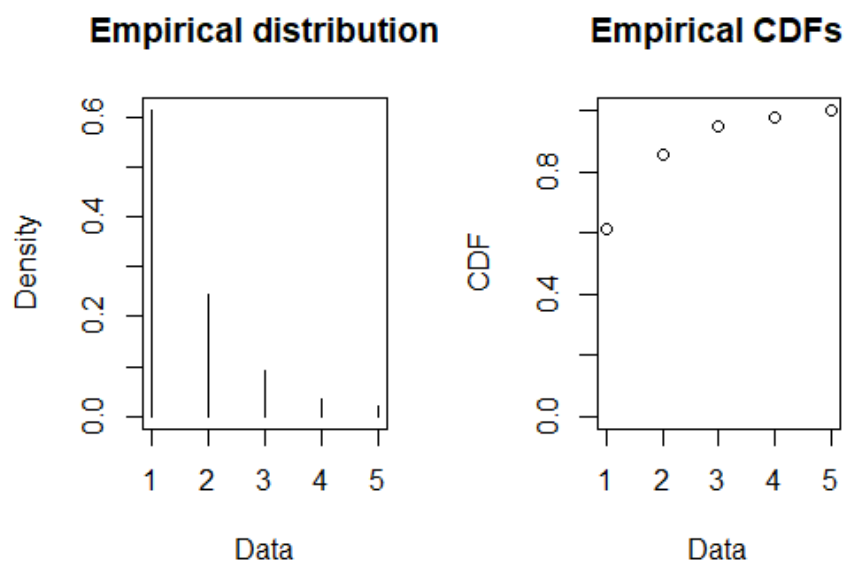


Figura 9.35: Gráfico de la función de distribución

La asimetría es de **1,72**.

En el gráfico de Cullen&Frey se muestra la observación muy cerca de una Poisson.

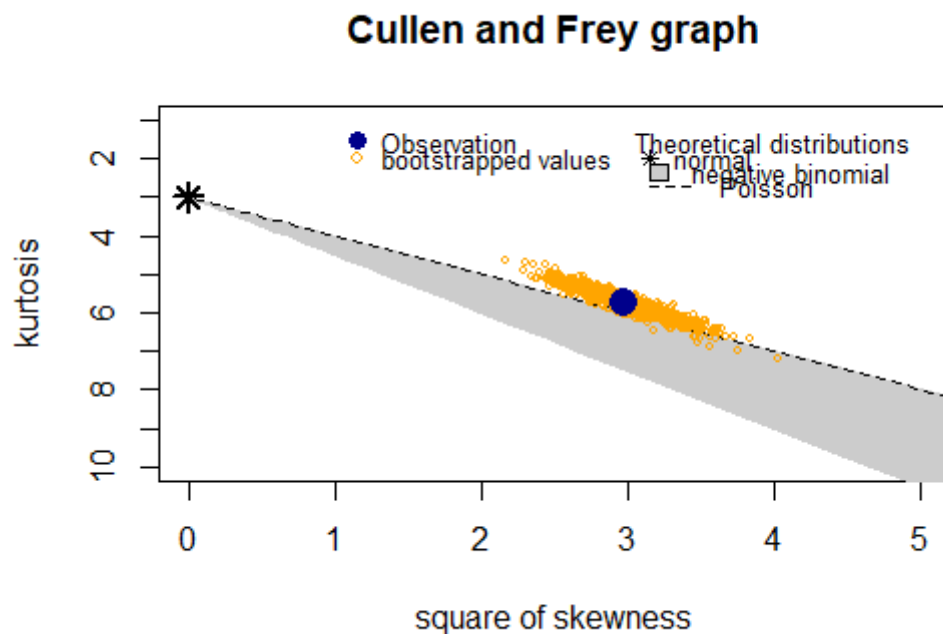


Figura 9.36: Gráfico de Cullen&Frey

El ajuste muestra que no se acepta que sea una distribución de Poisson.

Distribución	P-value KS
Poisson	1.26124e-10
Geométrica	1.052006e-116

Análisis del ajuste de los datos a la distribución multinomial La tabla con las frecuencias y proporciones de cada serie de años es la siguiente:

2014-2017	1	2	3	4	5
Frecuencia	453	180	62	23	17
Proporción	0,62	0,24	0,08	0,03	0,02

2018-2019	1	2	3	4	5
Frecuencia	223	88	37	13	5
Proporción	0,61	0,24	0,10	0,04	0,01

El p-valor obtenido es **0.5743**. No podemos rechazar H0.