**INFORME ESTADÍSTICO:**

**Howatharra**

Irene Pascual Albericio 871627

Sergio Ros Alcázar 874792

Andrea Mateo Sánchez 874649

**ÍNDICE:**

[**1. Introducción a los datos de lluvia. 2**](#_heading=h.yum6hgi39o9t)

[**2. Análisis con todos los datos disponibles. 3**](#_heading=h.9j6tkq5l1evm)

[❖ Medidas de localización, dispersión y forma. 3](#_heading=h.apaeqphtgdnt)

[❖ Histograma. 3](#_heading=h.n8tapplj4huf)

[❖ Diagrama de caja y detección de datos atípicos. 4](#_heading=h.4vk733hu4r5n)

[❖ Estudio de normalidad. 5](#_heading=h.x2ho7iar60ax)

[➔ Test de Shapiro-Wilk. 5](#_heading=h.f7p52xnfg7a)

[➔ Test de Anderson-Darling. 5](#_heading=h.bh1c80y1tb27)

[➔ Test de Kolmogorov-Smirnov. 5](#_heading=h.c92vzwr2fshy)

[❖ Conclusiones. 5](#_heading=h.diu1vy2r59xn)

[**3. Análisis por meses. 6**](#_heading=h.1grgixhf8k2r)

[❖ Medidas de localización, dispersión y forma. 6](#_heading=h.4t30lvglf1zt)

[❖ Histograma. 7](#_heading=h.93udg7i2tyjv)

[❖ Diagrama de caja y detección de datos atípicos. 8](#_heading=h.g6wnqsuub2k7)

[❖ Estudio de normalidad. 8](#_heading=h.vonky58n8mmy)

[➔ Test de Shapiro-Wilk. 8](#_heading=h.8yxvhm77vddw)

[➔ Test de Anderson-Darling. 9](#_heading=h.bw3hk2qdgo6m)

[➔ Test de Kolmogorov-Smirnov. 10](#_heading=h.u6b7htd8zxjf)

[❖ Conclusiones. 10](#_heading=h.joe378m0f85f)

[**4. Análisis por años. 11**](#_heading=h.54obr6wxsse4)

[❖ Medidas de localización, dispersión y forma. 11](#_heading=h.yuf3c2ut3ejn)

[❖ Diagrama de caja y detección de datos atípicos. 12](#_heading=h.p7seb219zt7e)

[❖ Estudio de normalidad. 13](#_heading=h.9aiog8ncvnqg)

[➔ Test de Shapiro-Wilk. 13](#_heading=h.i3tdi4q6spz1)

[➔ Test de Anderson-Darling. 14](#_heading=h.c458tmjbjfqj)

[➔ Test de Kolmogorov-Smirnov. 16](#_heading=h.uz2vte8zw82)

[❖ Conclusiones. 18](#_heading=h.hah1v761djqm)

[**5. Sistemas de componentes independientes. 18**](#_heading=h.yn1d7vfavqb1)

[❖ Media y varianza aproximada. 18](#_heading=h.q964btd0c8gq)

[❖ Cálculo de media y varianza. 19](#_heading=h.wsblanqphs2s)

[❖ Conclusiones. 21](#_heading=h.fa1oe97px098)

# Introducción a los datos de lluvia.

La estación que se nos ha asignado es la estación de Howatharra, la cual está ubicada en la parte más occidental de Australia, concretamente entre las localidades de Geraldton y Binnu.

Los datos que nos han sido proporcionados recogen cifras de precipitaciones de todo tipo desde 1980 hasta 2022, divididas por meses donde indica las lluvias diarias de todo el mes. Los datos se recogen a las 9:00 AM de cada día y acumulan la información de las 24 horas previas.

A lo largo del trabajo, vamos a realizar un análisis estadístico mediante R Commander, el cual incluye resúmenes numéricos, gráficos estadísticos, datos atípicos y estudios de normalidad (test de Shapiro-Wilk, test de Anderson-Darling y test de Kolmogorov-Smirnov).



Fuente: Google Imágenes



Fuente: Google Maps

# Análisis con todos los datos disponibles.

Para comenzar, haremos un estudio de todos los datos disponibles de la estación y comentaremos los resultados obtenidos en el programa “R Commander”.

## Medidas de localización, dispersión y forma.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| media | sd | IQR | skewness | kurtosis | 0% | 25% | 50% | 75% | 100% | n |
| 36.231 | 41.333 | 57.425 | 1.46886 | 2.27071 | 0,0 | 2,5 | 21 | 59.925 | 240.2 | 516 |

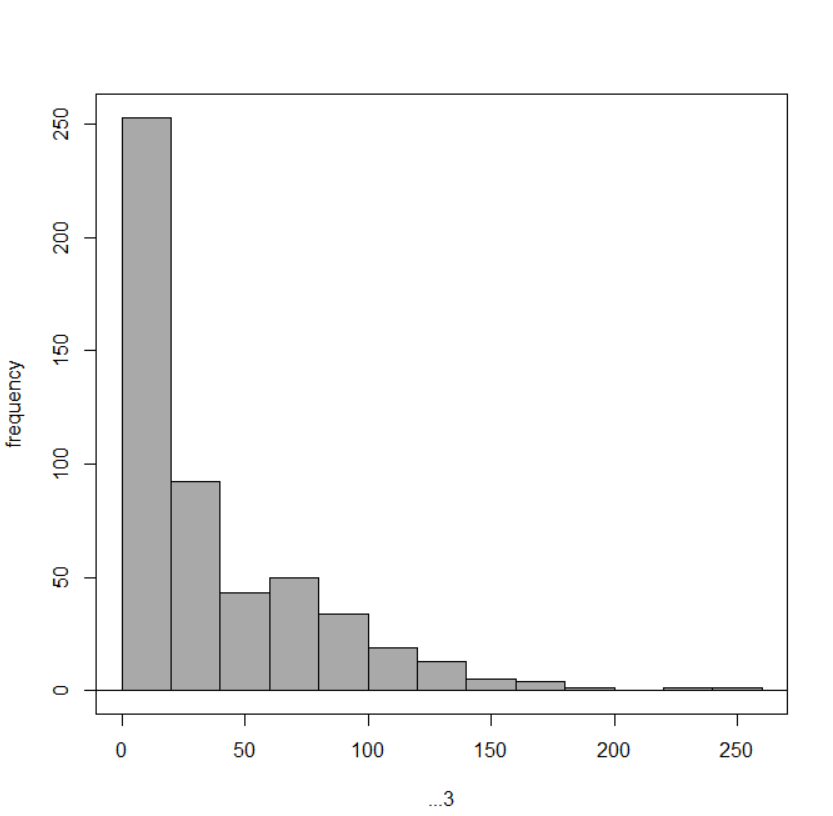
La media general de lluvias en este lugar es 36.231 L/m², lo que nos indica que es un lugar muy seco a comparación de España o Francia, por ejemplo. En cuanto a los cuartiles, sabemos que Q1=2.5, Q2=21, Q3=59.925 y Q4=240.2, por lo que podemos deducir que la mayoría de los datos se encuentran entre Q3 y Q4. La diferencia entre el primer y el tercer cuartil (el rango intercuartílico) es de 57.425.

La desviación típica es de 41.333. Es muy elevada, por lo que hay muchos datos que están alejados de la media. Respecto al coeficiente de asimetría, los datos son asimétricos positivos, ya que 1.46886>0 (más adelante lo podremos ver de forma más clara con el diagrama de caja).

La curtosis nos indica el apuntamiento comparado con la distribución normal; pudiendo ser, si es menor que 0, platicúrtica; igual a 0, mesocúrtica y en nuestro caso, mayor que 0, leptocúrtica.

## Histograma.

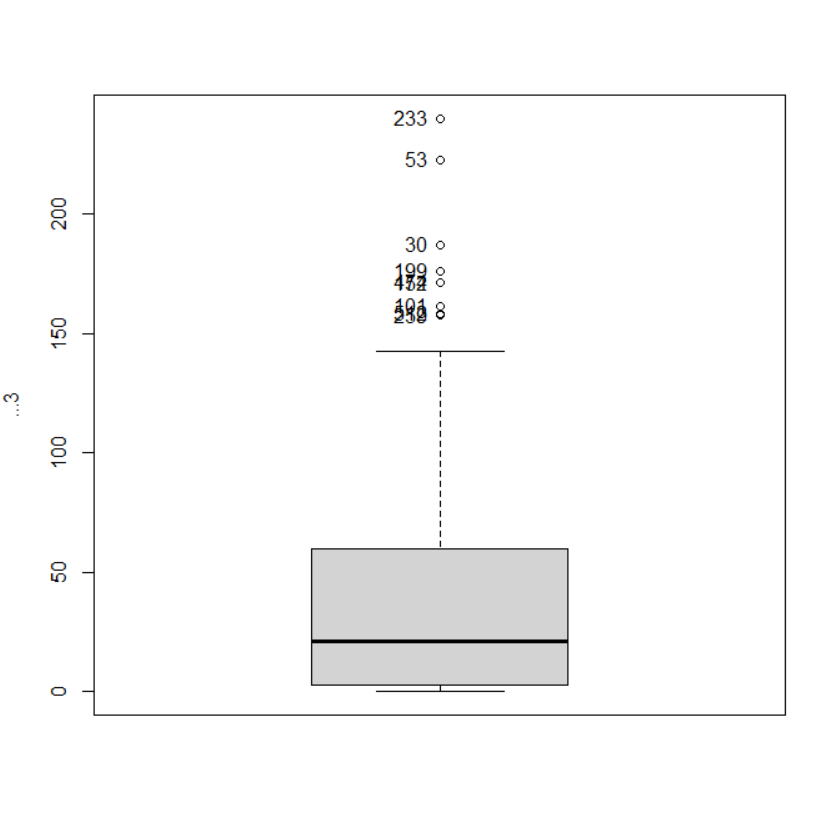
Este histograma indica la frecuencia con la que ha llovido teniendo en cuenta todos los datos desde 1980 hasta 2022. El eje X muestra la cantidad de agua precipitada en litros por metro cuadrado y el eje Y las veces que ha llovido cierta cantidad. Vemos que aproximadamente a partir de 100 L/m² la frecuencia disminuye considerablemente.



## Diagrama de caja y detección de datos atípicos.

Con este diagrama de caja, comprobamos esa asimetría positiva de la que habíamos hablado, y además hay ciertos valores atípicos que distan mucho de la media, y al igual que se puede ver en el histograma, están por encima de los 200 L/m².

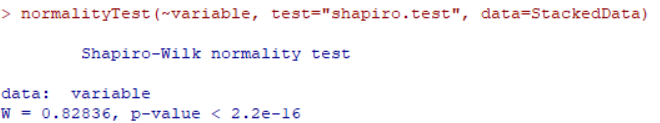
Con esto, es innecesario realizar algún análisis de normalidad puesto que se puede deducir que no sigue una distribución normal. Sin embargo, para comprobarlo, a continuación mostraremos los diversos tests de normalidad.



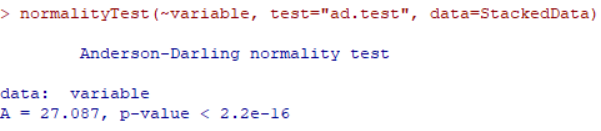
## Estudio de normalidad.

El p-valor en todos los casos es menor que 0.05, y concluimos que efectivamente, como ya habíamos demostrado anteriormente, no sigue la distribución normal.

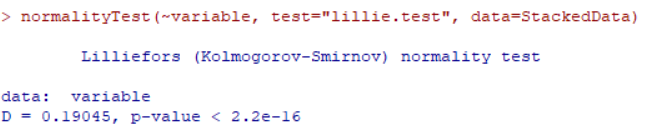
### Test de Shapiro-Wilk.



### Test de Anderson-Darling.



### Test de Kolmogorov-Smirnov.



## Conclusiones.

Podemos observar que existe una relación entre el histograma y las medidas de localización, dispersión y forma. En ambos se puede observar que existe una asimetría positiva. También hemos realizado un estudio de normalidad mediante tres métodos diferentes y en todos hemos llegado a la misma conclusión, los datos en conjunto no siguen ninguna distribución normal, lo cual hemos demostrado además con los datos de los primeros apartados.

# Análisis por meses.

Continuando con el estudio, analizaremos la información agrupada por meses, de enero a diciembre.

## Medidas de localización, dispersión y forma.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Mes | Media | sd | IQR | cv | skewness | kurtosis | 0% | 25% | 50% | 75% | 100% | n |
| Enero | 9.77907 | 19.05070 | 8.00 | 1.94811 | 2.71493 | 7.75907 | 0.0 | 0.00 | 0.0 | 8.00 | 87.0 | 43 |
| Febrero | 10.99069 | 19.64391 | 12.85 | 1.78732 | 2.26897 | 4.89927 | 0.0 | 0.00 | 1.5 | 12.85 | 77.5 | 43 |
| Marzo | 15.81395 | 27.88792 | 19.50 | 1.76350 | 2.61104 | 7.34896 | 0.0 | 0.00 | 3.0 | 19.50 | 128.5 | 43 |
| Abril | 24.69302 | 27.62243 | 32.75 | 1.11863 | 1.52725 | 2.28578 | 0.0 | 4.00 | 14.4 | 36.75 | 115.5 | 43 |
| Mayo | 66.43953 | 51.96404 | 63.35 | 0.78212 | 1.64243 | 3.24407 | 5.0 | 25.30 | 55.5 | 88.65 | 240.2 | 43 |
| Junio | 82.52093 | 40.80385 | 51.85 | 0.49446 | 0.44409 | -0.01488 | 5.0 | 57.00 | 79.5 | 108.85 | 187.0 | 43 |
| Julio | 86.84883 | 34.29209 | 39.80 | 0.39484 | 0.54854 | 0.02063 | 26.5 | 65.00 | 79.6 | 104.80 | 176.0 | 43 |
| Agosto | 71.64186 | 34.49126 | 36.10 | 0.48144 | 0.74039 | 1.08082 | 6.1 | 50.00 | 67.1 | 86.10 | 171.2 | 43 |
| Septiembre | 35.15581 | 22.27999 | 18.15 | 0.63374 | 1.67903 | 3.77637 | 0.0 | 21.10 | 32.0 | 39.25 | 117.0 | 43 |
| Octubre | 15.67907 | 10.71802 | 11.70 | 0.68358 | 1.23768 | 3.24953 | 0.0 | 8.75 | 16.0 | 20.45 | 55.0 | 43 |
| Noviembre | 9.50000 | 11.42126 | 12.35 | 1.20223 | 1.80066 | 3.66814 | 0.0 | 1.15 | 7.0 | 13.50 | 51.5 | 43 |
| Diciembre | 5.49302 | 9.08422 | 7.00 | 1.65376 | 1.92223 | 3.26108 | 0.0 | 0.00 | 0.0 | 7.00 | 37.5 | 43 |

Sabiendo ya el significado de todas las columnas, podemos observar en la media que el mes con más lluvia es julio y el menos lluvioso diciembre. El mes con mayor desviación típica es mayo, y el que tiene la menor es diciembre.

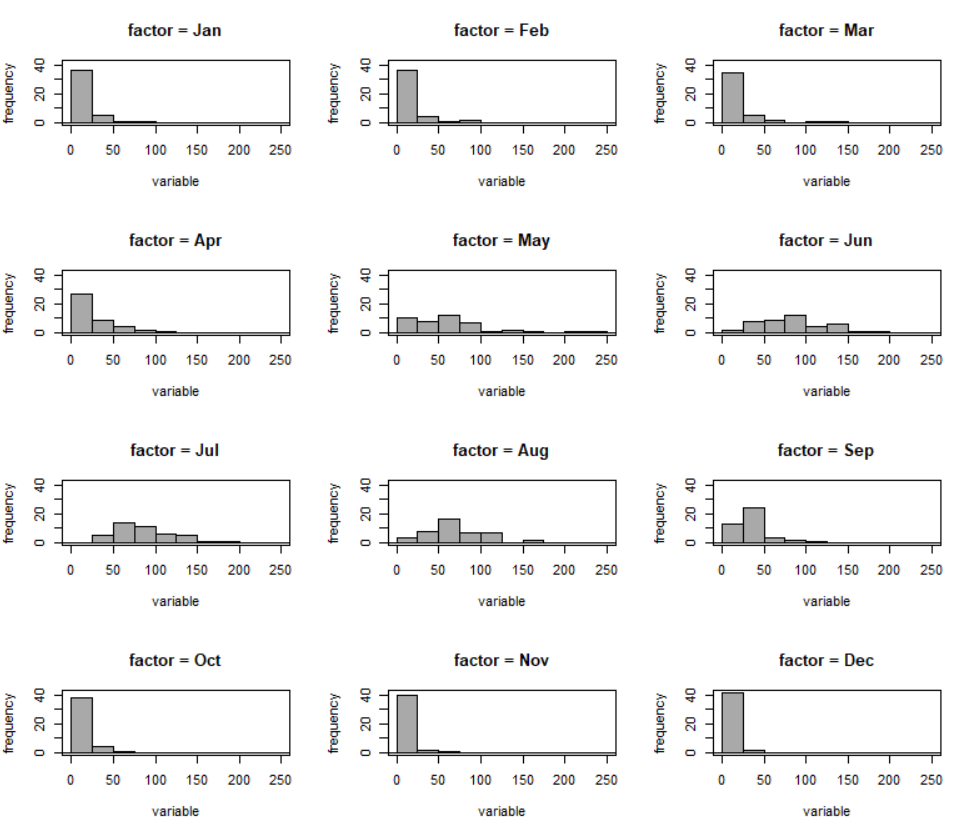
Entre mayo y agosto se encuentran los meses más lluviosos, siendo la media de unos 77 L/m², mientras que en los meses secos, de septiembre a abril, la media es de unos 16 L/m².

Fijándonos en los datos, las lluvias obtenidas son muy bajas lo que nos permite saber que estamos en un clima en el que generalmente llueve poco.

Observamos que el rango intercuartílico varía mucho de un mes a otro, siendo el mayor en mayo (63.35) y el menor en diciembre (7).

En lo referente al coeficiente de asimetría, es mayor que 1 en todos los meses excepto junio, julio y agosto, por lo que los datos tienen una distribución asimétrica positiva, tal y como comprobaremos a continuación. También vamos a realizar los tests de normalidad para determinar en qué meses los datos siguen una distribución normal.

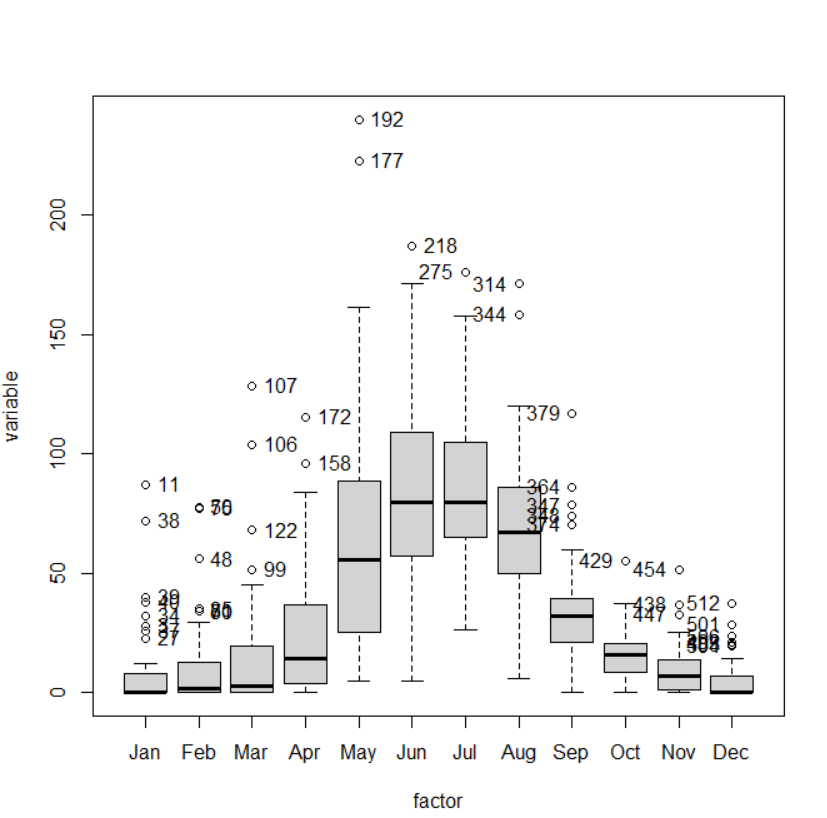
## Histograma.



En el eje de abscisas está la cantidad de agua y en el de ordenadas la frecuencia con la que llueve. El eje X representa la cantidad de agua que cae en litros por metro cuadrado y el eje Y representa la frecuencia con la que llueve esa cantidad de litros.

Si nos fijamos en el histograma, nos damos cuenta de que, tal y como habíamos previsto, todos los meses tienen una distribución asimétrica positiva excepto junio, julio y agosto, que siguen una distribución normal. Esto último lo confirmaremos realizando los tests de normalidad.

## Diagrama de caja y detección de datos atípicos.



Ahora pasamos a analizar el diagrama de caja. Podemos ver que los datos de los meses entre septiembre y abril rara vez alcanzan los 50 L/m² y son considerablemente menores que los de mayo a agosto, que llegan hasta los 100 L/m². Observamos además que el mes con mayor lluvia ha sido el mes de junio, ya que la distancia entre el cuartil 3 y la franja entre el rectángulo y la línea final es la mayor.

También notamos la existencia de datos atípicos en todos los meses, entre los que destacan los de junio y julio, que alcanzan los 150 L/m²; y los de mayo, que sobrepasan los 200 L/m².

## Estudio de normalidad.

La columna de la izquierda son los meses y la columna de la derecha representa los p-valores de dichos meses. Los valores cuyo p-valor supera 0.05 (destacados en verde) siguen una distribución normal.

### Test de Shapiro-Wilk.

|  |  |  |
| --- | --- | --- |
| Mes | W | p-valor |
| Enero | 0.0000000007765 | 0.0000000093181 |
| Febrero | 0.0000000052263 | 0.0000000522627 |
| Marzo | 0.0000000034385 | 0.0000000378233 |
| Abril | 0.0000184760040 | 0.0001293300000 |
| Mayo | 0.0000636778498 | 0.0003183900000 |
| Junio | 0.6168413000000 | 0.7940911700000 |
| Julio | 0.3970456000000 | 0.7940911700000 |
| Agosto | 0.1133264000000 | 0.3399792100000 |
| Septiembre | 0.0000493170912 | 0.0002959000000 |
| Octubre | 0.0029885000000 | 0.0119541200000 |
| Noviembre | 0.0000029963580 | 0.0000239708639 |
| Diciembre | 0.0000000185683 | 0.0000001671150 |

Como observamos, los meses de junio, julio y agosto tienen un p-valor mayor de 0.05 por lo tanto siguen una distribución normal.

### Test de Anderson-Darling.

|  |  |  |
| --- | --- | --- |
| Mes | A | p-valor |
| Enero | 2.22e-16 | 8.0478e-16 |
| Febrero | 3.5389e-15 | 3.8928e-14 |
| Marzo | 4.5135e-14 | 4.4151e-13 |
| Abril | 1.9252e-05 | 0.00011551 |
| Mayo | 0.00037162 | 0.00185810 |
| Junio | 0.56275651 | 0.64580088 |
| Julio | 0.32290044 | 0.64580088 |
| Agosto | 0.12361574 | 0.37084722 |
| Septiembre | 1.5218e-05 | 0.00010653 |
| Octubre | 0.04608169 | 0.18432676 |
| Noviembre | 1.2136e-06 | 9.7090e-06 |
| Diciembre | 4.4151e-14 | 4.4151e-13 |

Los resultados de este test confirman lo que hemos obtenido en el anterior, ya que muestra los mismos tres meses. También destaca el mes de octubre, aunque si nos fijamos en el histograma vemos claramente que no sigue una distribución normal.

### Test de Kolmogorov-Smirnov.

|  |  |  |
| --- | --- | --- |
| Mes | D | p-valor |
| Enero | 6.7600e-11 | 0.0000000007436 |
| Febrero | 1.0783e-11 | 0.0000000001294 |
| Marzo | 1.5443e-09 | 0.0000000154428 |
| Abril | 0.00071839 | 0.0043103000000 |
| Mayo | 0.03451771 | 0.1620722000000 |
| Junio | 0.71362352 | 0.7302412000000 |
| Julio | 0.36512058 | 0.7302412000000 |
| Agosto | 0.19011728 | 0.5703518000000 |
| Septiembre | 2.9386e-06 | 0.0000235084804 |
| Octubre | 0.03241445 | 0.1620722000000 |
| Noviembre | 0.00012056 | 0.0008439000000 |
| Diciembre | 1.1509e-08 | 0.0000001035812 |

Este test no es tan preciso como los anteriores, pero vemos que igualmente nos vuelven a salir junio, julio, agosto y octubre como meses que siguen una distribución normal (aunque sabemos que octubre sigue una distribución asimétrica positiva).

## Conclusiones.

En febrero y marzo la desviación típica es mayor que la media por lo que la media no aporta mucha información y no es representativa respecto a la distribución de estos meses. Las medidas de asimetría son positivas en todos los meses, lo que significa que la distribución es asimétrica positiva. Viendo los test que hemos realizado se puede ver que los meses que tienen una distribución gaussiana son aquellos cuyo p-valor es mayor a 0.05.

# Análisis por años.

## Medidas de localización, dispersión y forma.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Año | Media | sd | IQR | cv | skewness | kurtosis | 0% | 25% | 50% | 75% | 100% | n |
| 1980 | 32.89167 | 36.05505 | 55.050 | 1.09617 | 0.83331 | -0.98536 | 0.0 | 2.800 | 20.70 | 57.850 | 92.9 | 12 |
| 1981 | 38.38333 | 45.79184 | 75.775 | 1.19301 | 0.88078 | -1.14666 | 0.0 | 3.000 | 15.75 | 78.775 | 115.2 | 12 |
| 1982 | 41.69167 | 55.26163 | 66.725 | 1.32548 | 1.86334 | 3.80383 | 0.0 | 3.000 | 21.70 | 69.725 | 187.0 | 12 |
| 1983 | 37.60833 | 47.77766 | 74.875 | 1.27040 | 1.70541 | -0.30137 | 0.0 | 2.100 | 14.20 | 76.975 | 135.3 | 12 |
| 1984 | 52.72500 | 57.98933 | 34.750 | 1.09984 | 2.55033 | 7.80177 | 0.0 | 23.275 | 44.65 | 58.025 | 222.6 | 12 |
| 1985 | 28.85833 | 35.35926 | 43.850 | 1.22527 | 1.25578 | 1.11377 | 0.0 | 0.000 | 16.50 | 43.850 | 110.5 | 12 |
| 1986 | 45.65000 | 49.45296 | 79.925 | 1.08330 | 0.60230 | -1.43033 | 0.0 | 4.850 | 19.50 | 84.775 | 129.0 | 12 |
| 1987 | 34.63333 | 32.26209 | 38.350 | 0.93153 | 1.25301 | 1.97079 | 0.0 | 12.125 | 29.50 | 50.475 | 112.3 | 12 |
| 1988 | 43.08333 | 49.26815 | 53.300 | 1.14355 | 1.55342 | 1.93233 | 3.0 | 11.750 | 19.50 | 65.050 | 161.6 | 12 |
| 1989 | 35.05833 | 44.13684 | 55.950 | 1.25895 | 1.51080 | 1.92841 | 0.0 | 0.000 | 19.70 | 55.950 | 141.5 | 12 |
| 1990 | 38.99167 | 33.65884 | 63.550 | 0.86323 | 0.12903 | -1.69602 | 0.0 | 6.000 | 38.45 | 69.550 | 87.0 | 12 |
| 1991 | 36.20833 | 31.31606 | 24.625 | 0.86488 | 0.99510 | 0.22175 | 0.0 | 18.500 | 29.00 | 43.125 | 95.6 | 12 |
| 1992 | 36.61667 | 51.85635 | 53.625 | 1.41619 | 1.85108 | 3.52363 | 0.0 | 0.000 | 9.818 | 53.625 | 171.2 | 12 |
| 1993 | 37.35000 | 43.79654 | 65.075 | 1.17259 | 0.97954 | -0.29625 | 0.0 | 0.000 | 18.00 | 65.075 | 125.8 | 12 |
| 1994 | 29.15833 | 34.74608 | 67.400 | 1.19163 | 0.74418 | -1.39453 | 0.0 | 0.000 | 16.40 | 67.400 | 86.9 | 12 |
| 1995 | 28.30000 | 40.28654 | 32.550 | 1.14235 | 1.76183 | 2.71300 | 0.0 | 2.250 | 7.55 | 34.800 | 128.8 | 12 |
| 1996 | 48.00000 | 61.30867 | 70.500 | 1.27726 | 1.24900 | 0.28176 | 0.0 | 0.000 | 23.60 | 70.500 | 176.0 | 12 |
| 1997 | 35.97500 | 36.25313 | 68.950 | 1.00773 | 0.45077 | -1.57758 | 0.0 | 1.875 | 28.00 | 70.825 | 92.1 | 12 |
| 1998 | 35.92500 | 42.32580 | 51.000 | 1.17817 | 1.36198 | 1.69423 | 0.0 | 1.500 | 26.10 | 52.500 | 136.7 | 12 |
| 1999 | 61.40000 | 70.87709 | 82.450 | 1.15435 | 1.58422 | 2.80944 | 0.0 | 7.925 | 38.75 | 90.375 | 240.2 | 12 |
| 2000 | 36.95000 | 42.73420 | 66.950 | 1.15654 | 1.09472 | 0.11703 | 0.0 | 2.750 | 19.40 | 69.700 | 128.5 | 12 |
| 2001 | 32.14167 | 44.90921 | 33.450 | 1.39722 | 2.25223 | 5.77750 | 0.0 | 2.625 | 19.70 | 36.075 | 157.5 | 12 |
| 2002 | 28.08333 | 28.01522 | 49.600 | 0.99757 | 0.43751 | -1.68251 | 0.0 | 3.500 | 19.25 | 53.100 | 68.9 | 12 |
| 2003 | 41.62500 | 38.17641 | 65.925 | 0.91715 | 0.44342 | -1.44217 | 0.0 | 6.625 | 33.70 | 72.550 | 97.5 | 12 |
| 2004 | 32.80833 | 36.39139 | 66.050 | 1.10921 | 0.72046 | -1.41305 | 0.0 | 2.375 | 12.80 | 68.425 | 90.0 | 12 |
| 2005 | 35.52500 | 41.22850 | 41.300 | 1.16054 | 1.32886 | 0.73333 | 0.0 | 3.625 | 22.60 | 44.925 | 125.0 | 12 |
| 2006 | 20.75000 | 21.32653 | 19.775 | 1.02778 | 1.42246 | 0.97971 | 3.0 | 5.750 | 11.75 | 25.525 | 66.8 | 12 |
| 2007 | 25.08333 | 28.65293 | 29.750 | 1.14230 | 1.71211 | 3.68011 | 0.0 | 3.750 | 22.25 | 33.500 | 99.8 | 12 |
| 2008 | 40.45833 | 49.42475 | 38.950 | 1.22162 | 1.33612 | 0.29156 | 0.0 | 9.625 | 19.75 | 48.575 | 142.1 | 12 |
| 2009 | 39.61667 | 47.08010 | 69.750 | 1.18839 | 0.80830 | -0.83027 | 0.0 | 0.000 | 15.50 | 69.750 | 126.0 | 12 |
| 2010 | 23.41667 | 28.10671 | 30.125 | 1.20028 | 1.39636 | 1.22965 | 0.0 | 4.500 | 10.60 | 34.625 | 88.0 | 12 |
| 2011 | 47.14167 | 42.59107 | 75.450 | 0.90346 | 0.36091 | -1.88702 | 2.5 | 9.750 | 29.00 | 85.200 | 107.6 | 12 |
| 2012 | 26.52500 | 35.41556 | 24.675 | 1.33517 | 2.37880 | 6.38123 | 0.0 | 5.625 | 15.75 | 30.300 | 127.0 | 12 |
| 2013 | 25.55833 | 24.35609 | 39.200 | 0.95296 | 0.59038 | -0.39072 | 0.0 | 0.000 | 26.20 | 39.200 | 74.0 | 12 |
| 2014 | 34.45833 | 39.81435 | 63.800 | 1.15543 | 0.83606 | -0.38218 | 0.0 | 1.875 | 14.25 | 65.675 | 117.0 | 12 |
| 2015 | 36.79167 | 37.64669 | 59.725 | 1.02323 | 0.87751 | -0.40117 | 0.0 | 9.400 | 19.00 | 69.125 | 113.5 | 12 |
| 2016 | 41.79167 | 42.05487 | 46.375 | 1.00629 | 1.24577 | 0.95225 | 0.0 | 12.375 | 25.75 | 58.750 | 137.0 | 12 |
| 2017 | 35.51667 | 34.40724 | 31.525 | 0.96876 | 1.41882 | 1.73946 | 0.0 | 11.375 | 28.25 | 42.900 | 117.5 | 12 |
| 2018 | 31.70833 | 38.75298 | 45.750 | 1.22217 | 1.37186 | 1.81376 | 0.0 | 0.000 | 22.25 | 45.750 | 125.0 | 12 |
| 2019 | 26.91667 | 48.70404 | 33.500 | 1.80943 | 2.76650 | 8.29163 | 0.0 | 0.000 | 6.75 | 33.500 | 171.5 | 12 |
| 2020 | 31.25000 | 33.77364 | 40.000 | 1.08075 | 1.67113 | 3.94155 | 0.0 | 1.500 | 32.25 | 41.500 | 120.0 | 12 |
| 2021 | 37.25833 | 38.56665 | 41.500 | 1.03511 | 1.76267 | 4.33847 | 0.0 | 7.125 | 38.25 | 48.625 | 140.0 | 12 |
| 2022 | 47.27500 | 49.12822 | 52.675 | 1.03920 | 1.25558 | 1.13304 | 0.0 | 10.825 | 38.25 | 63.500 | 158.5 | 12 |

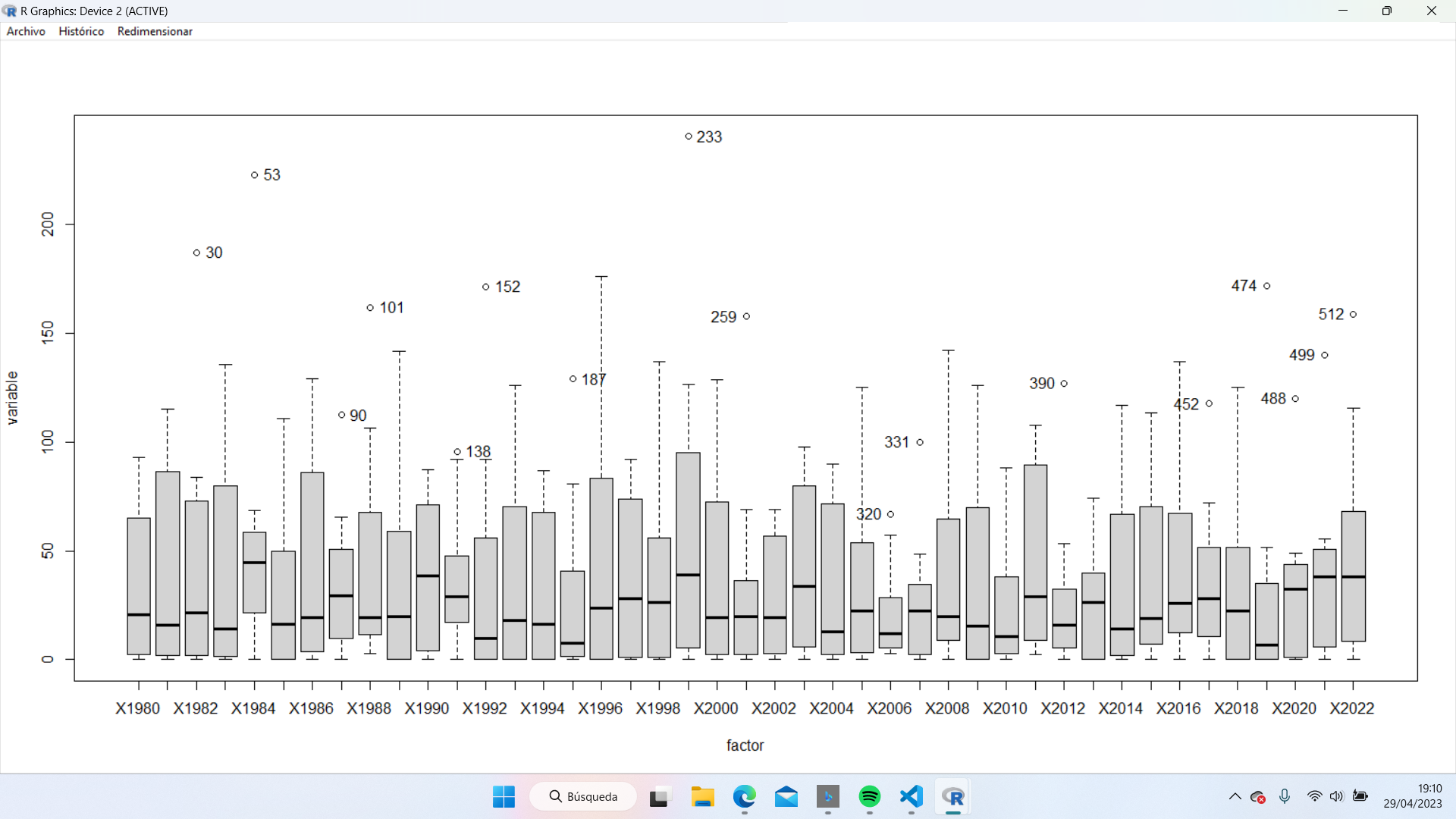
Observando esta tabla y la de los meses, podemos ver que el mes más lluvioso de estos 43 años es julio de 1999.

Gracias a los resultados obtenidos observamos que la media anual está comprendida entre los 25 L/m² y los 45 L/m², con algunas excepciones, como pueden ser 1984, 1986, 1996, 1999, 2011 y 2022 por encima de la media y 1994, 1995, 2002, 2012, 2013 y 2019 por debajo de la media. El año con más lluvias fue 1999 y el año con menos lluvias fue 2006.

La segunda columna representa la desviación típica. Las más notables son en 1996 de 61.31 y en 1999 de 70.88. En estos años los datos recogidos recorren un mayor rango de valores, por lo tanto la media podría ser menos representativa. En cambio, en el año 2006 la desviación típica es bastante menor, de 21.33, por lo que la media puede ser más precisa que en los casos mencionados antes.

Fijándonos en el coeficiente de asimetría en la tercera columna notamos que la mayor parte de los valores son mayores que 0, lo que nos indica que dichos años poseen una asimetría positiva. Los años en los que el coeficiente de asimetría está más próximo a cero, cómo 1990 o 2011, tienden a seguir un modelo gaussiano.

## Diagrama de caja y detección de datos atípicos.



Como podemos observar, el año 1996 es el que más lluvia ha tenido ya que su diagrama de caja es el más grande. Podemos ver que hay datos atípicos en bastantes años, como en 1982, que se alcanzan casi 200 L/m²; 1984, donde se alcanzan los 200 L/m²; o 1999, donde se superan con creces los 200 L/m².

Además, en años como 2006, 2007 o 2012, es evidente que las lluvias han sido considerablemente más escasas en comparación con el resto de años.

## Estudio de normalidad.

### Test de Shapiro-Wilk.

|  |  |  |
| --- | --- | --- |
| Años | W | p-valor |
| 1980 | 0.82142 | 0.01659 |
| 1981 | 0.77284 | 0.004678 |
| 1982 | 0.76754 | 0.004099 |
| 1983 | 0.78452 | 0.00628 |
| 1984 | 0.70082 | 0.0008581 |
| 1985 | 0.82374 | 0.01767 |
| 1986 | 0.82352 | 0.01756 |
| 1987 | 0.89854 | 0.1519 |
| 1988 | 0.79628 | 0.008502 |
| 1989 | 0.80605 | 0.01099 |
| 1990 | 0.89169 | 0.1239 |
| 1991 | 0.88242 | 0.09415 |
| 1992 | 0.75354 | 0.00291 |
| 1993 | 0.83876 | 0.02674 |
| 1994 | 0.77564 | 0.005017 |
| 1995 | 0.74961 | 0.002647 |
| 1996 | 0.79306 | 0.00782 |
| 1997 | 0.85185 | 0.03869 |
| 1998 | 0.8412 | 0.02864 |
| 1999 | 0.83321 | 0.02292 |
| 2000 | 0.82982 | 0.02087 |
| 2001 | 0.71615 | 0.001211 |
| 2002 | 0.83913 | 0.02702 |
| 2003 | 0.87215 | 0.06958 |
| 2004 | 0.79975 | 0.009307 |
| 2005 | 0.814 | 0.01358 |
| 2006 | 0.79431 | 0.008078 |
| 2007 | 0.81823 | 0.01521 |
| 2008 | 0.7608 | 0.003473 |
| 2009 | 0.80766 | 0.01147 |
| 2010 | 0.81874 | 0.01542 |
| 2011 | 0.83177 | 0.02203 |
| 2012 | 0.72023 | 0.001329 |
| 2013 | 0.90489 | 0.1834 |
| 2014 | 0.82225 | 0.01696 |
| 2015 | 0.86738 | 0.06054 |
| 2016 | 0.86557 | 0.05743 |
| 2017 | 0.84519 | 0.03204 |
| 2018 | 0.8214 | 0.01658 |
| 2019 | 0.6057 | 0.0001197 |
| 2020 | 0.81241 | 0.01301 |
| 2021 | 0.81327 | 0.01332 |
| 2022 | 0.87201 | 0.06931 |

En este test, los años que están destacados en verde son aquellos cuyo p-valor es mayor que 0.05. Como podemos ver, solamente 8 datos pasan el test.

### Test de Anderson-Darling.

|  |  |  |
| --- | --- | --- |
| Años | A | p-valor |
| 1980 | 0.86123 | 0.01848 |
| 1981 | 1.2081 | 0.002219 |
| 1982 | 1.0237 | 0.006842 |
| 1983 | 1.161 | 0.002959 |
| 1984 | 1.3087 | 0.001201 |
| 1985 | 0.81479 | 0.02455 |
| 1986 | 0.91806 | 0.01305 |
| 1987 | 0.39919 | 0.3074 |
| 1988 | 0.99645 | 0.008083 |
| 1989 | 0.90401 | 0.01422 |
| 1990 | 0.45695 | 0.2176 |
| 1991 | 0.56831 | 0.1111 |
| 1992 | 1.133 | 0.00351 |
| 1993 | 0.77915 | 0.03054 |
| 1994 | 1.1976 | 0.002367 |
| 1995 | 1.2212 | 0.002049 |
| 1996 | 1.0426 | 0.006095 |
| 1997 | 0.68567 | 0.05413 |
| 1998 | 0.68017 | 0.05598 |
| 1999 | 0.68078 | 0.05577 |
| 2000 | 0.85739 | 0.01892 |
| 2001 | 1.2476 | 0.001743 |
| 2002 | 0.75091 | 0.0363 |
| 2003 | 0.54703 | 0.1242 |
| 2004 | 1.0214 | 0.006939 |
| 2005 | 0.93044 | 0.0121 |
| 2006 | 1.0182 | 0.007076 |
| 2007 | 0.72601 | 0.04228 |
| 2008 | 1.3148 | 0.001157 |
| 2009 | 0.97726 | 0.009089 |
| 2010 | 0.87759 | 0.01672 |
| 2011 | 0.84217 | 0.02076 |
| 2012 | 1.2062 | 0.002245 |
| 2013 | 0.42182 | 0.2688 |
| 2014 | 0.92033 | 0.01287 |
| 2015 | 0.67418 | 0.05807 |
| 2016 | 0.65566 | 0.06505 |
| 2017 | 0.76433 | 0.03344 |
| 2018 | 0.78912 | 0.02873 |
| 2019 | 1.84 | 0.00004732 |
| 2020 | 0.7505 | 0.03639 |
| 2021 | 0.7438 | 0.03791 |
| 2022 | 0.5672 | 0.1119 |

En este test, se observa que los años que han pasado el test anterior también pasan este (subrayados en verde) y seguirán una distribución normal, aunque aparecen datos nuevos que superan este test y no el anterior (1997,1998 y 1999).

### Test de Kolmogorov-Smirnov.

|  |  |  |
| --- | --- | --- |
| Años | D | p-valor |
| 1980 | 0.21624 | 0.1257 |
| 1981 | 0.31293 | 0.001938 |
| 1982 | 0.26591 | 0.01894 |
| 1983 | 0.3151 | 0.001725 |
| 1984 | 0.31079 | 0.002171 |
| 1985 | 0.21892 | 0.1156 |
| 1986 | 0.25986 | 0.02455 |
| 1987 | 0.14152 | 0.7282 |
| 1988 | 0.25487 | 0.03025 |
| 1989 | 0.27432 | 0.01303 |
| 1990 | 0.15475 | 0.5933 |
| 1991 | 0.22083 | 0.1089 |
| 1992 | 0.27945 | 0.01029 |
| 1993 | 0.22629 | 0.09053 |
| 1994 | 0.29261 | 0.005488 |
| 1995 | 0.26739 | 0.01775 |
| 1996 | 0.26457 | 0.02007 |
| 1997 | 0.2125 | 0.1409 |
| 1998 | 0.198 | 0.2166 |
| 1999 | 0.20198 | 0.1931 |
| 2000 | 0.25144 | 0.03482 |
| 2001 | 0.28849 | 0.006708 |
| 2002 | 0.22168 | 0.1061 |
| 2003 | 0.16496 | 0.4893 |
| 2004 | 0.2807 | 0.009713 |
| 2005 | 0.2533 | 0.03227 |
| 2006 | 0.24254 | 0.04958 |
| 2007 | 0.19067 | 0.2655 |
| 2008 | 0.31237 | 0.001997 |
| 2009 | 0.26027 | 0.02413 |
| 2010 | 0.25401 | 0.03135 |
| 2011 | 0.26043 | 0.02397 |
| 2012 | 0.22889 | 0.08253 |
| 2013 | 0.18633 | 0.298 |
| 2014 | 0.28159 | 0.009317 |
| 2015 | 0.2459 | 0.04347 |
| 2016 | 0.22969 | 0.08018 |
| 2017 | 0.26526 | 0.01947 |
| 2018 | 0.20662 | 0.1684 |
| 2019 | 0.29087 | 0.005978 |
| 2020 | 0.21626 | 0.1256 |
| 2021 | 0.23478 | 0.06655 |
| 2022 | 0.18513 | 0.3073 |

A diferencia de los dos primeros tests, vemos que en este caso, los datos que podían seguir una distribución normal, aumentan en cantidad. Por tanto, para definir los años que posiblemente sigan una distribución normal, nos quedaremos únicamente con aquellos que cumplen la condición de tener un p-valor mayor de 0.05 en los tres tests, que son los siguientes: 1987, 1990, 1991, 2003, 2013, 2016 y 2022.

## Conclusiones.

Como punto final cabe recalcar que los años con una desviación estándar mayor en la tabla del principio del apartado no se podrían tomar muy en cuenta ya que esto nos dice que muchos de sus datos están bastante alejados de la media.

# Sistemas de componentes independientes.

Vamos a realizar un análisis del siguiente sistema. Esto lo haremos mediante la técnica de simulación estadística para obtener la media y varianza aproximadas de la variable T. Para ello tomaremos valores que van de 10.000 a 500.000.

## Media y varianza aproximada.

El programa utilizado para calcular los datos de la tabla es R Commander y el código es el siguiente:

rm(list=ls()) n<-10000 X123<-pmin(X1,X2,X3)

X1<-rexp(n,1/(4\*1)) X456<-pmin(X4,X5,X6)

X2<-rexp(n,1/(4\*2)) X123456<-pmax(X123,X456)

X3<-rexp(n,1/(4\*3)) T<-pmin(X123456,X7)

X4<-rexp(n,1/(4\*4)) muT<-mean(T)

X5<-rexp(n,1/(4\*5)) varT<-var(T)

X6<-rexp(n,1/(4\*6)) sdT<-sd(T)

X7<-rexp(n,1/(4\*7)) muT;varT;sdT

|  |  |  |  |
| --- | --- | --- | --- |
| n | Aproximación de E[T] | Aproximación de Var[T] | Aproximación de |
| 10.000 | 5.6893 | 25.4658 | 5.0463 |
| 50.000 | 5.7416 | 26.0852 | 5.1073 |
| 100.000 | 5.7434 | 26.1854 | 5.1171 |
| 500.000 | 5.7411 | 25.8430 | 5.0836 |

## Cálculo de media y varianza.

Para calcular la media y la varianza exacta tendremos que hacer uso de un programa de cálculo simbólico, en este caso hemos utilizado Maple, ya que debido a ciertos problemas Sage no nos daba algunos resultados. El código utilizado es el siguiente:

CdfExp:=(x,lambda)->1-exp(-lambda\*x)



Cdf123:=(x,lambda1,lambda2,lambda3)->1-(1-CdfExp(x,lambda1))\*(1-CdfExp(x,lambda2))\*(1-CdfExp(x,lambda3))



Cdf456:=(x,lambda4,lambda5,lambda6)->1-(1-CdfExp(x,lambda4))\*(1-CdfExp(x,lambda5))\*(1-CdfExp(x,lambda6))



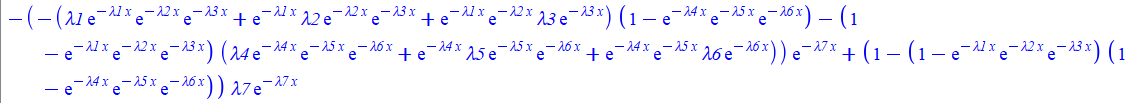
Cdf123456:=(x,lambda1,lambda2,lambda3,lambda4,lambda5,lambda6)->Cdf123(x,lambda1,lambda2,lambda3)\*Cdf456(x,lambda4,lambda5,lambda6)



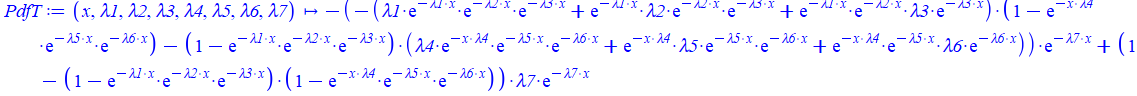
CdfT:=(x,lambda1,lambda2,lambda3,lambda4,lambda5,lambda6,lambda7)->1-(1-Cdf123456(x,lambda1,lambda2,lambda3,lambda4,lambda5,lambda6))\*(1-CdfExp(x,lambda7))



diff(CdfT(x,lambda1,lambda2,lambda3,lambda4,lambda5,lambda6,lambda7),x$1)



PdfT:= (x,lambda1,lambda2,lambda3,lambda4,lambda5,lambda6,lambda7)->-(-(lambda1\*exp (-lambda1\*x)\*exp(-lambda2\*x)\*exp(-lambda3\*x)+ lambda2\*exp(-lambda1\*x)\*exp(-lambda2\*x)\*exp(-lambda3\*x)+ lambda3\*exp(-lambda1\*x)\*exp(-lambda2\*x)\*exp(-lambda3\*x))\*(1-exp(-lambda4\*x)\*exp(-lambda5\*x)\*exp(-lambda6\*x))-(1-exp(-lambda1\*x)\*exp(-lambda2\*x)\*exp(-lambda3\*x))\*(lambda4\*exp (-lambda4\*x)\*exp(-lambda5\*x)\*exp(-lambda6\*x)+ lambda5\*exp(-lambda4\*x)\*exp(-lambda5\*x)\*exp(-lambda6\*x)+ lambda6\*exp(-lambda4\*x)\*exp(-lambda5\*x)\*exp(-lambda6\*x)))\*exp(-lambda7\*x)+(1-(1-exp(-lambda1\*x)\*exp(-lambda2\*x)\*exp(-lambda3\*x))\*(1-exp(-lambda4\*x)\*exp(-lambda5\*x)\*exp(-lambda6\*x)))\*lambda7\*exp(-lambda7\*x)



ETk:=(k)->int(x^k\*PdfT(x,1/(4\*1),1/(4\*2),1/(4\*3),1/(4\*4),1/(4\*5),1/(4\*6),1/(4\*7)),x=0..infinty)



ETk(0);



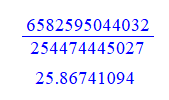
ETk(1);evalf(%);



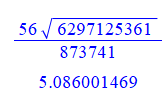
ETk(2);evalf(%);



ETk(2)-(ETk(1))^2;evalf(%);



Sqrt(ETk(2)-(ETk(1))^2);evalf(%);



La media tiene un valor exacto de 5.747854341, la varianza tiene un valor de 25.86741094 y la desviación estándar es 5.086001469.

## Conclusiones.

Como podemos observar tras realizar la simulación estadística y la resolución teórica de nuestro circuito, los valores obtenidos en la simulación convergen al mismo valor, que coincide con el obtenido mediante la resolución teórica.

Al realizar simulaciones con diferentes tamaños de muestras, acabamos dándonos cuenta de que cuanto mayor es la muestra analizada más exacto es el valor obtenido, el cual converge al valor obtenido al realizar este problema de manera teórica en Maple.

En el primer experimento, con la muestra de 10.000, el valor de la esperanza es de 5.6893, el cual dista aproximadamente 0.058 del valor real. Sin embargo, en el último experimento, con la muestra de 500.000, el valor de la esperanza es 5.7411, que tan solo dista 0.0067 del valor real de la esperanza. Lo mismo ocurre si nos fijamos en la varianza o en la desviación típica.

Es por esto que podemos afirmar que la técnica de simulación estadística, con una población elevada, es un buen método para el cálculo aproximado del valor de la esperanza y de la varianza.