Actividad 1 Integradora

Adrian Pineda Sanchez

2024-10-22

Varias obras de la Ingeniería Civil se ven altamente influenciadas por los factores climatológicos como la lluvia y la temperatura. En hidrología, por ejemplo, es necesario conocer el valor de la máxima precipitación probable registrada para un determinado período de retorno para realizar los cálculos y el diseño de las estructuras de conservación de agua como las presas y otras obras civiles como puentes, carreteras, y edificios. El cálculo adecuado de dimensiones para un drenaje, garantizan la correcta evacuación de volúmenes de agua asegurando la vida útil de carreteras, aeropuertos, y drenajes urbanos.

Se analizaran los datos históricos (1994-2023) de las precipitaciones máximas mensuales por estado para cumplir el objetivo principal de este estudio que consiste en calcular la precipitación más extrema que se logra con un periodo de retorno seleccionado. De manera individual deberás trabajar con los siguientes pasos para analizar las precipitaciones históricas del estado que selecciones y que sea diferente al resto de tu equipo.

ANÁLISIS

1. Análisis estadístico descriptivo de las precipitaciones históricas máximas mensuales de un estado

A Descarga la base de datos de precipitaciones máximas históricas mensuales de todos los estados de la república de la siguiente liga: precipitaciones mensuales Download precipitaciones mensuales. Esta base de datos se construyó con información de los resúmenes mensuales de lluvia y temperatura de CONAGUA (https://smn.conagua.gob.mx/es/Links to an external site.). Selecciona un estado que sea diferente a los del resto de tu equipo.

```
##     V1  V2                  V3     V4
## 1 Anio Mes              Estado Lluvia
## 2 1994 Ene      Aguascalientes    8.3
## 3 1994 Ene     Baja.California   10.3
## 4 1994 Ene Baja.California.Sur      0
## 5 1994 Ene            Campeche   85.4
## 6 1994 Ene    Ciudad.de.México   17.7
```

```
##       V1  V2         V3    V4
## 29  1994 Ene Tamaulipas  84.2
## 62  1994 Feb Tamaulipas   5.1
## 95  1994 Mar Tamaulipas  19.6
```

```
## 128 1994 Abr Tamaulipas  22.8
## 161 1994 May Tamaulipas    20
## 194 1994 Jun Tamaulipas 112.6
```

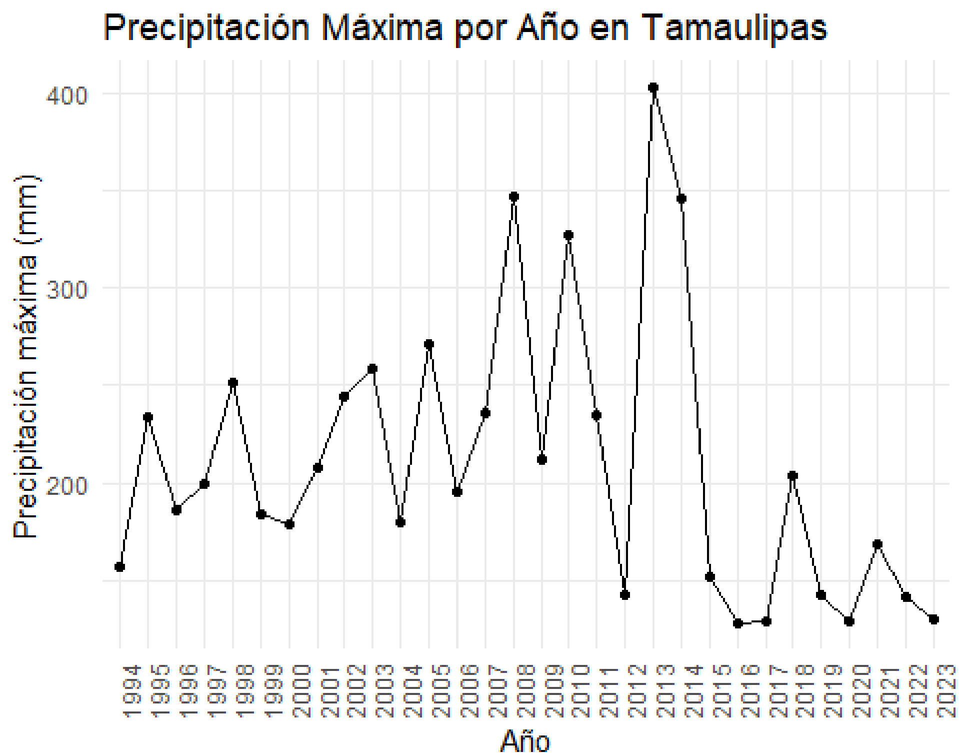
```
## 
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
## 
##     filter, lag
```

```
## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union
```

B. Elabora una gráfica de las precipitaciones máximas mensuales por año para tu estado. Para ello deberás calcular la precipitación mensual máxima de cada año y graficarla.

```
## # A tibble: 6 × 4
## # Groups:   V1 [6]
##   V1    V2    V3            V4
##   <chr> <chr> <chr>      <dbl>
## 1 1994  Ago   Tamaulipas  156.
## 2 1995  Ago   Tamaulipas  234.
## 3 1996  Ago   Tamaulipas  186.
## 4 1997  Oct   Tamaulipas  199.
## 5 1998  Sep   Tamaulipas  251.
## 6 1999  Jul   Tamaulipas  184.
```

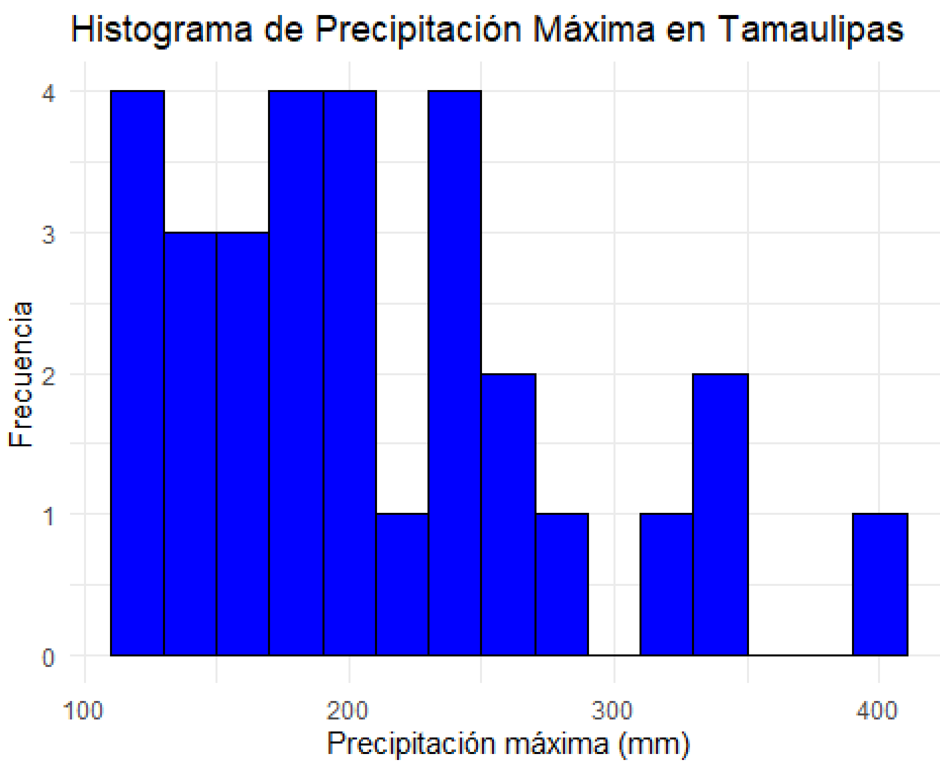


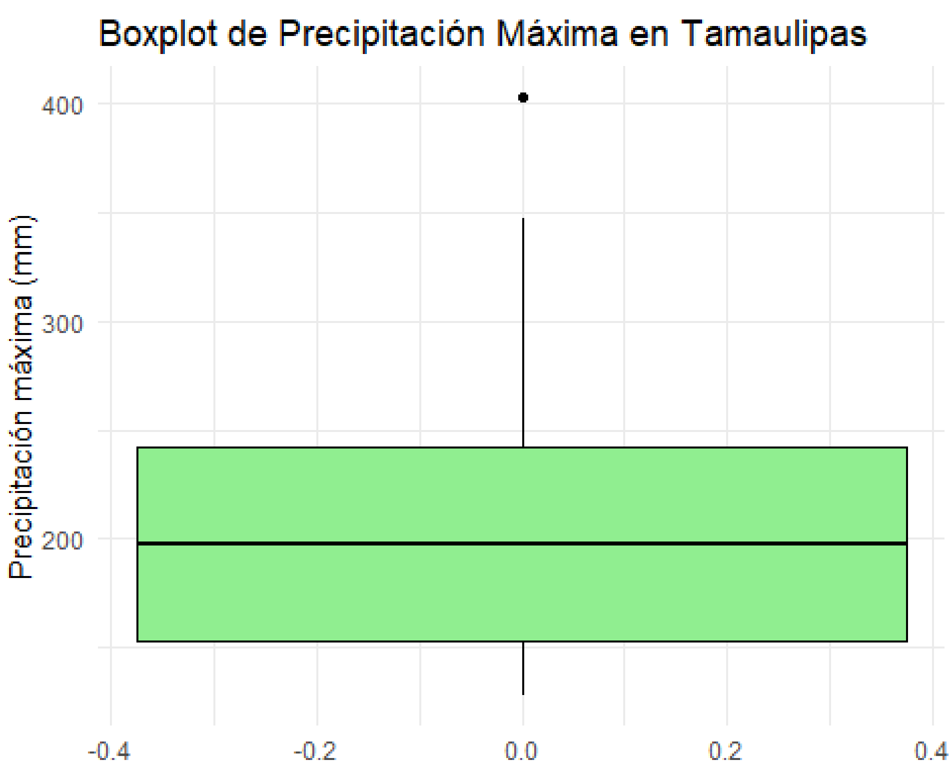
C. Analiza los datos de precipitaciones máximas mensuaels del estado seleccionado.

Calcula las medidas de centralización y variación de las precipitaciones máximas mensuales

```
## Análisis descriptivo de la precipitación máxima en Tamaulipas:

## ---------------------------------------------------------------

## Media (Precipitación máxima promedio): 210.6533 mm

## Mediana: 197.65 mm

## Varianza: 5197.215

## Desviación estándar: 72.09171 mm

## ---------------------------------------------------------------

## Descripción del comportamiento:

## La media es mayor que la mediana, lo que sugiere un sesgo positivo
(algunos valores muy altos están elevando la media).

## La varianza y la desviación estándar son indicadores de la dispersión de
los datos. A mayor valor, mayor es la variabilidad en las precipitaciones
máximas anuales en Tamaulipas.
```

Realiza gráficos que te sirvan para describir la distribución de las lluviar máximas mensuales: histograma y boxplot





Análisis del Comportamiento de la Distribución

Centralización

- **Media**: La media de las precipitaciones máximas es de **210.65 mm**, lo que indica que en promedio, la precipitación máxima anual es de este valor.
- **Mediana**: La mediana es de **197.65 mm**, lo que significa que la mitad de los valores están por debajo de este número y la otra mitad por encima.

Dado que la **media es mayor que la mediana**, la distribución presenta un **sesgo positivo** o hacia la derecha. Esto sugiere que hay algunos años con valores de precipitación muy altos que están elevando la media.

Sesgo

Como mencioné antes, la diferencia entre la media y la mediana (la media es mayor) indica un **sesgo positivo**. El histograma refuerza esta idea al mostrar que hay un número pequeño de años con precipitaciones máximas muy altas (por encima de 350 mm), lo que está empujando la cola de la distribución hacia la derecha.

Variación

- **Varianza**: La varianza es **5197.215**, lo que muestra que hay una dispersión significativa en los datos.
- **Desviación estándar**: Con una desviación estándar de **72.09 mm**, se puede observar que los valores están relativamente dispersos en torno a la media. Esta variabilidad puede indicar que las precipitaciones máximas en Tamaulipas varían bastante entre los años.

Outliers (Valores atípicos)

El **boxplot** muestra la existencia de un **outlier** por encima de los **400 mm** de precipitación. Este valor es un año particular donde la precipitación máxima fue significativamente más alta que los otros años, lo que puede estar influenciando el sesgo positivo que observamos. Los valores fuera de la caja en el gráfico representan años con precipitaciones extremas.

Conclusión

En general, la distribución de las precipitaciones máximas en Tamaulipas muestra una variabilidad considerable entre los años. Aunque la mayoría de los años presentan precipitaciones cercanas a los **200 mm**, hay algunos años con precipitaciones mucho más altas, como lo indica el outlier y el sesgo positivo. Esto sugiere que aunque la precipitación en Tamaulipas es en promedio estable, existen eventos extremos que resultan en lluvias mucho más intensas en algunos años.

Estos análisis pueden ser útiles para prever posibles riesgos de inundaciones o sequías, y tomar decisiones en función de las variaciones climáticas en Tamaulipas.

D. ¿Qué puedes concluir observando la gráfica de los máximos mensuales anuales para tu Estado? ¿Observas alguna tendencia? ¿Puedes concluir que cada determinado número de años la cantidad de precipitación sube o baja? ¿Para qué nos sirve analizar este tipo de gráficas?

Conclusiones Observando la Gráfica de los Máximos Mensuales en Tamaulipas

Tendencias Observadas

Al observar la gráfica, se pueden identificar ciertos patrones: - **Oscilación constante**: A lo largo del período analizado (1994-2023), las precipitaciones máximas muestran una oscilación, con años de precipitaciones más altas seguidos de años con precipitaciones más bajas. - **Picos marcados**: Se observan picos notables en los años **2007**, **2010**, **2013**, y **2014**, donde las precipitaciones alcanzaron valores por encima de los **350 mm**. Estos años parecen ser excepcionales y podrían estar asociados a fenómenos climáticos extremos, como huracanes o tormentas intensas. - **Disminución reciente**: Desde el año **2016**, las precipitaciones máximas tienden a ser más bajas, con valores que en su mayoría no superan los **200 mm**.

Ciclos de Precipitación

No parece haber un ciclo regular de años en los que la precipitación sube o baja. Sin embargo, sí se puede observar que, aproximadamente cada **3 a 6 años**, ocurre un pico de precipitación alta. Esto podría indicar una cierta periodicidad en los eventos climáticos extremos, aunque es difícil establecer un ciclo exacto basándonos solo en esta gráfica.

Es muy probable que los patrones de precipitaciones que observas en Tamaulipas estén relacionados con fenómenos climáticos como El Niño y La Niña. Estos fenómenos tienen un impacto significativo en los patrones de clima a nivel mundial, y México es una región donde estos efectos se sienten de manera clara.

Utilidad de Este Tipo de Gráficas

- **Monitoreo climático**: Estas gráficas permiten identificar años con eventos climáticos extremos, lo que es útil para analizar el impacto de fenómenos como tormentas tropicales o huracanes.
- **Planeación y prevención**: Al observar que ciertos años tienen precipitaciones máximas muy elevadas, se pueden tomar decisiones informadas para mitigar riesgos de inundaciones o daños en infraestructura.
- **Análisis de tendencias a largo plazo**: Este tipo de análisis puede ayudar a determinar si el clima de la región está cambiando, lo cual es relevante en el contexto del cambio climático. Por ejemplo, la disminución de las precipitaciones en años recientes podría señalar cambios en los patrones de lluvia.

En resumen, la gráfica sugiere que las precipitaciones en Tamaulipas son altamente variables y presentan picos significativos cada cierto número de años. Estos análisis ayudan a entender mejor el comportamiento climático y a preparar estrategias de mitigación ante eventos extremos.

2. Análisis de Frecuencias Método Gráfico

A. En el data frame de los datos de precipitación máxima se agrega una columna con los datos de lluvias máximas ordenados de mayor a menor.

```
## # A tibble: 6 × 4
## # Groups:   V1 [6]
##   V1    V2    V3            V4
##   <chr> <chr> <chr>      <dbl>
## 1 2013  Sep   Tamaulipas  403.
## 2 2008  Jul   Tamaulipas  347
## 3 2014  Sep   Tamaulipas  346.
## 4 2010  Jul   Tamaulipas  328.
## 5 2005  Jul   Tamaulipas  272.
## 6 2003  Sep   Tamaulipas  259.
```

B. Se agrega una columna con el número de orden que tiene asignado cada precipitación máxima. A ese número se le llama "rank" (rango en español) y se simboliza por m

```
## # A tibble: 6 × 5
## # Groups:   V1 [6]
##   V1    V2    V3            V4  rank
##   <chr> <chr> <chr>      <dbl> <int>
## 1 2013  Sep   Tamaulipas  403.     1
## 2 2008  Jul   Tamaulipas  347      2
## 3 2014  Sep   Tamaulipas  346.     3
## 4 2010  Jul   Tamaulipas  328.     4
## 5 2005  Jul   Tamaulipas  272.     5
## 6 2003  Sep   Tamaulipas  259.     6
```

C. Se calcula la probabilidad de excedencia o de ocurrencia de acuerdo con Weibull, donde el numerador es el número de orden (m) o "rank" y el denominador es la suma del total de datos (N) y 1:

```
## # A tibble: 6 × 6
## # Groups:   V1 [6]
##   V1    V2    V3            V4  rank  P_exe
##   <chr> <chr> <chr>      <dbl> <int>  <dbl>
## 1 2013  Sep   Tamaulipas  403.     1 0.0323
## 2 2008  Jul   Tamaulipas  347      2 0.0645
## 3 2014  Sep   Tamaulipas  346.     3 0.0968
## 4 2010  Jul   Tamaulipas  328.     4 0.129
## 5 2005  Jul   Tamaulipas  272.     5 0.161
## 6 2003  Sep   Tamaulipas  259.     6 0.194
```

D. Se calcula la probabilidad de no excedencia para cada precipitación (complemento de la probabilidad de excedencia):

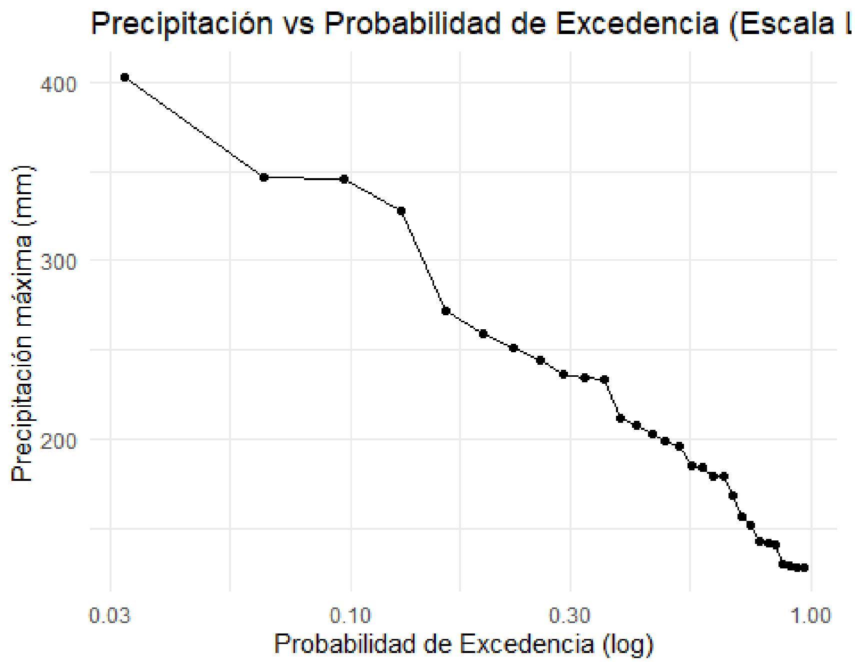
```
## # A tibble: 6 × 7
## # Groups:   V1 [6]
```

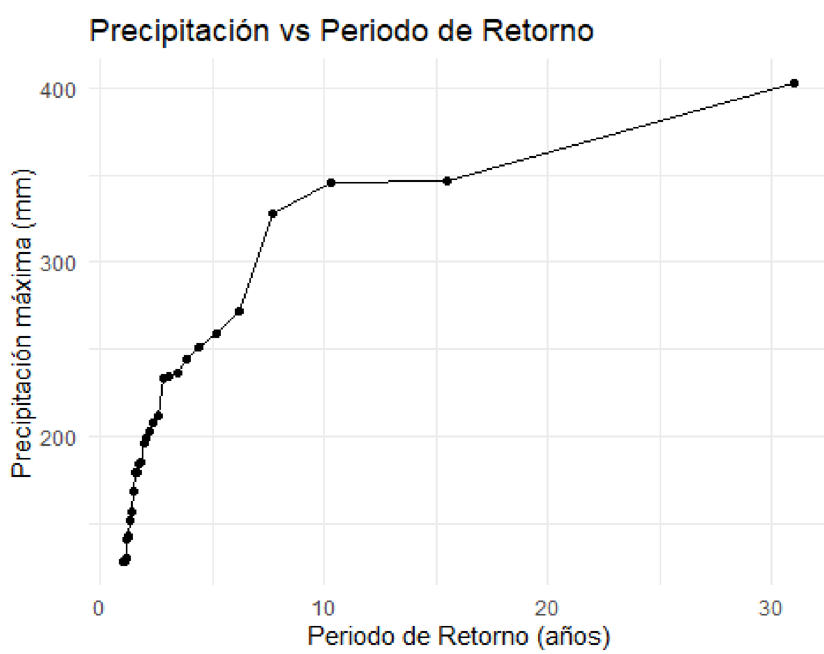
```
##   V1    V2    V3             V4  rank  P_exe P_no_exe
##   <chr> <chr> <chr>       <dbl> <int>  <dbl>    <dbl>
## 1 2013  Sep   Tamaulipas  403.      1 0.0323    0.968
## 2 2008  Jul   Tamaulipas  347       2 0.0645    0.935
## 3 2014  Sep   Tamaulipas  346.      3 0.0968    0.903
## 4 2010  Jul   Tamaulipas  328.      4 0.129     0.871
## 5 2005  Jul   Tamaulipas  272.      5 0.161     0.839
## 6 2003  Sep   Tamaulipas  259.      6 0.194     0.806
```

E. Se calcula el periodo de retorno como el inverso de la probabilidad de excedencia:

```
## # A tibble: 6 × 8
## # Groups:   V1 [6]
##   V1    V2    V3             V4  rank  P_exe P_no_exe P_ret
##   <chr> <chr> <chr>       <dbl> <int>  <dbl>    <dbl> <dbl>
## 1 2013  Sep   Tamaulipas  403.      1 0.0323    0.968 31
## 2 2008  Jul   Tamaulipas  347       2 0.0645    0.935 15.5
## 3 2014  Sep   Tamaulipas  346.      3 0.0968    0.903 10.3
## 4 2010  Jul   Tamaulipas  328.      4 0.129     0.871  7.75
## 5 2005  Jul   Tamaulipas  272.      5 0.161     0.839  6.2
## 6 2003  Sep   Tamaulipas  259.      6 0.194     0.806  5.17
```

F. Describe las gráficas obtenidas. ¿Qué significa la probabilidad de excedencia? ¿Qué significa el periodo de retorno? ¿Por qué es importante en hidrología? ¿Qué valores son deseables en la probabilidad de excedencia para una precipitación de diseño de una obra?





Descripción de las Gráficas

1. *Gráfica de Precipitación vs Probabilidad de Excedencia*

Esta gráfica muestra la relación entre las precipitaciones máximas y la **probabilidad de excedencia**. En ella, se puede observar cómo las precipitaciones más altas tienen una **probabilidad de excedencia baja**, lo que significa que es poco probable que esos valores de precipitación sean superados en un año determinado. A medida que disminuyen los valores de precipitación, la probabilidad de excedencia aumenta.

2. *Gráfica de Precipitación vs Periodo de Retorno*

En esta gráfica se muestra el **periodo de retorno** en años para diferentes niveles de precipitación máxima. Los valores más altos de precipitación tienen un periodo de retorno largo, es decir, son eventos menos frecuentes (por ejemplo, una precipitación de 400 mm tiene un periodo de retorno de 31 años). Los eventos más frecuentes (con precipitación más baja) tienen un periodo de retorno más corto.

Definición de Términos

Probabilidad de Excedencia:

La **probabilidad de excedencia** es la probabilidad de que un evento, como una precipitación máxima, **sea superado** en un año dado. Se calcula como:

$$P_{exe} = \frac{m}{N+1}$$

Donde m es el rango de la precipitación y N es el número total de observaciones.

- Un valor bajo de P_{exe} indica un evento raro (precipitación muy alta), que es poco probable que ocurra en un año específico.
- Un valor alto de P_{exe} indica que ese nivel de precipitación es común y probablemente será superado en un año dado.

Periodo de Retorno:

El **periodo de retorno** es el tiempo promedio, expresado en **años**, que transcurre entre eventos que exceden un determinado valor de precipitación. Se calcula como el inverso de la probabilidad de excedencia:

$$P_{ret} = \frac{1}{P_{exe}}$$

- Un **periodo de retorno largo** (por ejemplo, 30 años) indica que un evento de esa magnitud ocurrirá, en promedio, una vez cada 30 años.
- Un **periodo de retorno corto** (por ejemplo, 5 años) significa que ese evento ocurre con mayor frecuencia.

Importancia en Hidrología

En hidrología, la probabilidad de excedencia y el periodo de retorno son herramientas cruciales para el **diseño de infraestructuras**. Estos valores permiten estimar la frecuencia de eventos extremos, como lluvias intensas, y planificar estructuras como presas, puentes o alcantarillados para soportar dichos eventos.

Por ejemplo, si se está diseñando una presa, es importante considerar la probabilidad de que una lluvia extrema ocurra dentro del tiempo de vida útil de la estructura. Si no se toman en cuenta eventos con periodos de retorno largos, las infraestructuras podrían no ser capaces de resistir lluvias extremas, lo que podría causar fallas catastróficas.

Valores Deseables en la Probabilidad de Excedencia

Para el **diseño de una obra**, los ingenieros suelen buscar **probabilidades de excedencia bajas** para eventos de precipitación extrema, lo que garantiza que las estructuras estén preparadas para soportar estos eventos raros. Por ejemplo: - En el diseño de una infraestructura crítica, como una presa o un dique, se puede diseñar para un evento con un periodo de retorno de 100 años (es decir, una probabilidad de excedencia baja).

3 Análisis de Frecuencias Método Analítico

El método analítico consiste en asumir que los datos pueden ser ajustados a través de una función de densidad de probabilidades (FDP) conocida la cual nos servirá para modelar y pronosticar precipitaciones y periodos de retorno. Para ello es necesario probar varias distribuciones y emplear pruebas de bondad de ajuste para ver decidir cuál distribución es la que mejor se ajusta. Para nuestro análisis verificaremos el ajuste de las precipitaciones máximas mensuales a diferentes distribuciones.

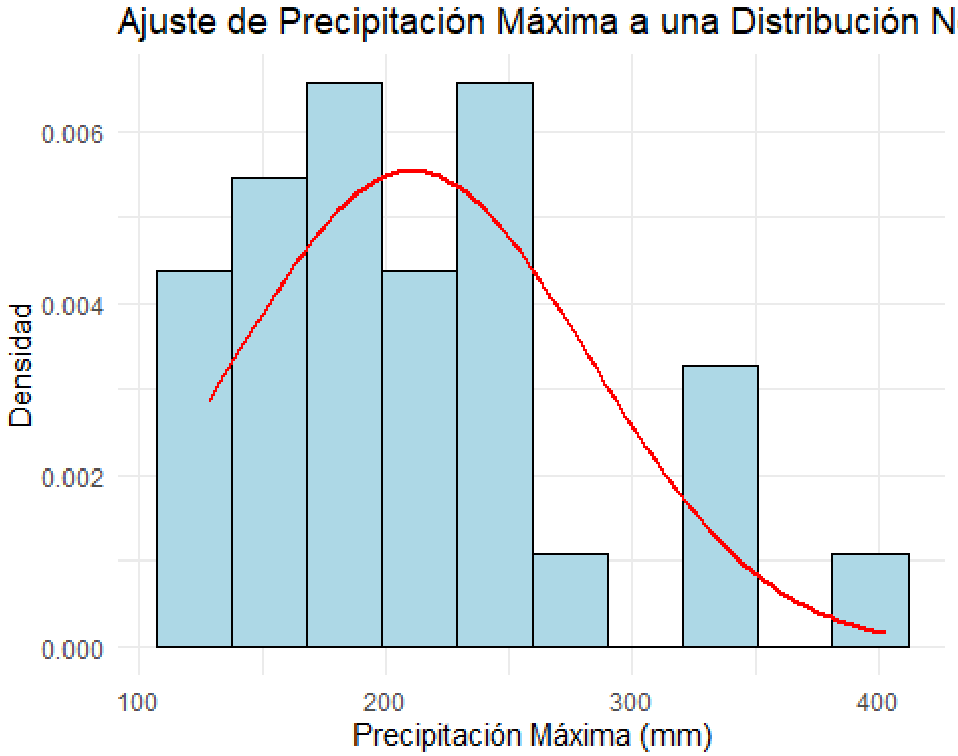
A. Ajuste a una Distribución Normal. Hay dos maneras de determinar si un conjunto de datos tiene una distribución normal, una visual y la otra mediante una prueba de bondad de ajuste.

Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución normal que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. De manera visual, ¿te parece que los datos se ajustan bien a una distribución normal? Explica. ¿Cuántos parámetros tiene la distribución Normal? ¿Cuáles son? ¿Por qué los parámetros se calculan de la forma en cómo se hace en el código?

$H0: Losdatosprovienendeunadistribuciónnormal$
$H1: Losdatosnoprovienendeunadistribuciónnormal$

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2
3.4.0.
## i Please use `after_stat(density)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



Este gráfico muestra el histograma de la precipitación máxima mensual y superpone la **curva teórica** de una distribución normal ajustada a los datos.

- **Histograma (barras azules)**: Muestra la frecuencia de las precipitaciones máximas en los diferentes rangos. Las alturas de las barras indican cuántos eventos de precipitación se encuentran dentro de cada intervalo de precipitación.
- **Curva Normal Teórica (línea roja)**: Es la distribución normal ajustada a los datos, calculada usando la media y la desviación estándar de la muestra. Esta curva teórica representa cómo se espera que se distribuyan los datos si siguen una distribución normal.

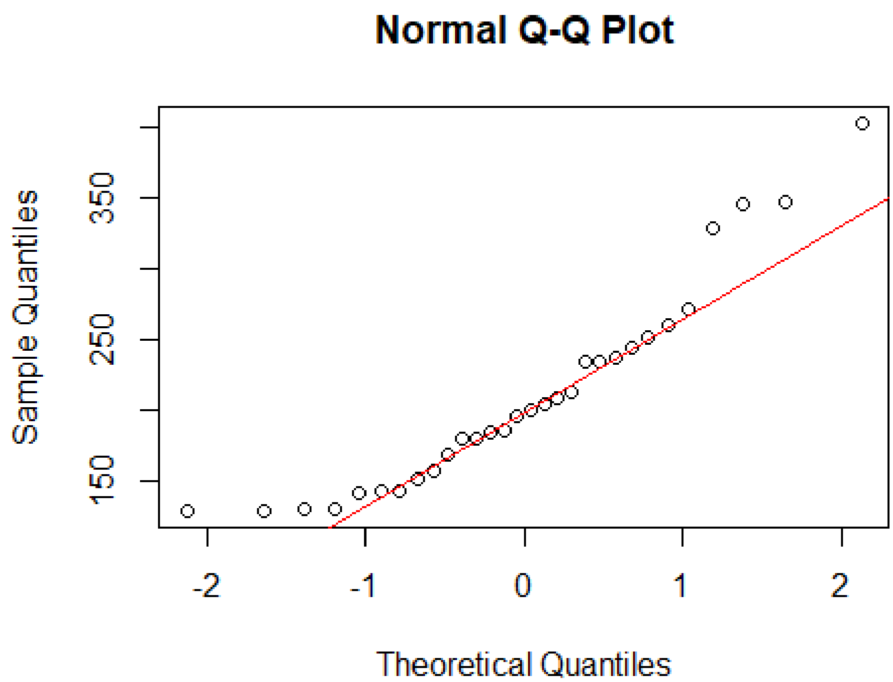
Interpretación del Gráfico de Densidad:

- **Comparación Visual**: La curva normal teórica parece ajustarse a los datos en algunas áreas, especialmente en los rangos cercanos a la media (200-250 mm). Sin embargo, los datos muestran algunos **picos y valles** que no se alinean perfectamente con la distribución normal, lo que sugiere que los datos tienen cierta **asimetría** o **variación** que no sigue una distribución normal perfecta.

- **Conclusión**: Aunque hay un cierto ajuste entre la curva normal y los datos, las diferencias indican que los datos podrían no ser completamente normales.

Construye la gráfica qqplot. De manera visual, ¿Los datos siguen una distribución normal de acuerdo con la Q-Qplot?

$H0: Losdatosprovienendeunadistribuciónnormal$
$H1: Losdatosnoprovienendeunadistribuciónnormal$



El **Q-Q plot** (quantile-quantile plot) compara los **cuantiles empíricos** de los datos de precipitación con los **cuantiles teóricos** de una distribución normal.

- **Línea roja**: Representa la alineación teórica perfecta entre los datos y la distribución normal. Si los puntos se alinean bien con esta línea, los datos siguen una distribución normal.
- **Puntos negros**: Representan los **cuantiles empíricos** de los datos de precipitación máxima.

Interpretación del Q-Q Plot:

- **Desviaciones en los extremos**: Observamos que los puntos tienden a **desviarse de la línea roja** en los extremos, lo que indica que los datos no siguen perfectamente una distribución normal. Los valores de precipitación más altos (cuantiles positivos) se encuentran por encima de la línea, lo que sugiere la presencia de colas pesadas o eventos extremos más frecuentes de lo que la normal esperaría.

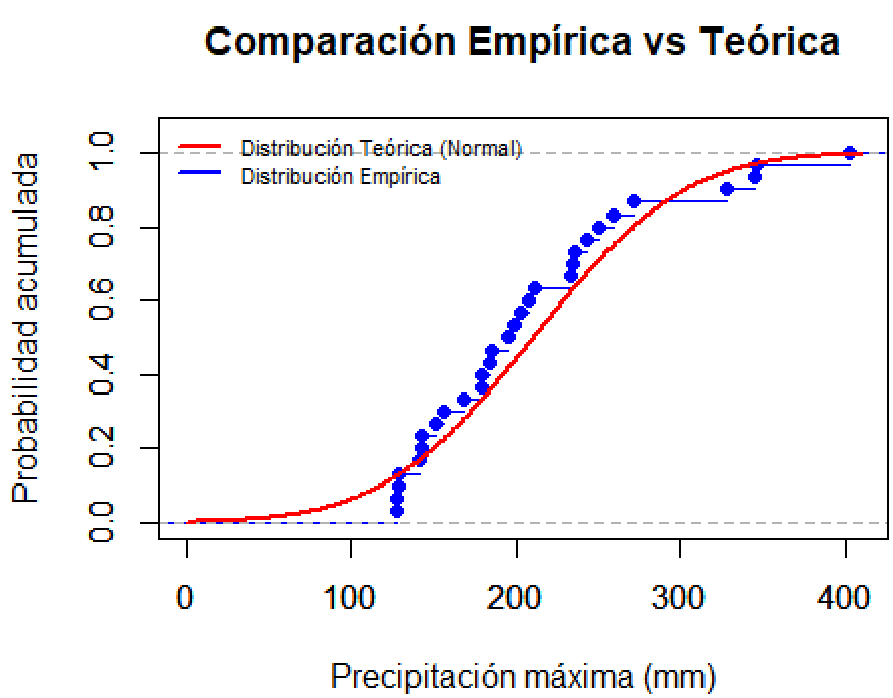
- **Ajuste en la parte media**: En la parte central del gráfico, los puntos se alinean mejor con la línea roja, lo que sugiere que los valores alrededor de la media pueden seguir una distribución normal de manera más cercana.

Conclusión:

- **Q-Q Plot**: Muestra que los datos de precipitación tienen colas más pesadas de lo que se esperaría bajo una distribución normal, lo que sugiere que no se distribuyen de manera perfectamente normal.

Compara las distribuciones de probabilidad acumuladas (ojiva) empíricos y teóricos de la distribución normal que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. Explica qué son datos empíricos y datos teóricos. ¿Se parecen las distribuciones de probabilidad acumuladas?

$H0: Losdatosprovienendeunadistribuciónnormal$
$H1: Losdatosnoprovienendeunadistribuciónnormal$



Explicación del Gráfico: Comparación Empírica vs Teórica

Este gráfico muestra una comparación entre la **distribución acumulada empírica** (basada en los datos de precipitaciones máximas) y la **distribución teórica** esperada si los datos siguieran una **distribución normal**. A continuación, desglosamos las principales características y lo que puedes interpretar del gráfico.

1. Distribución Empírica (azul)

- La curva azul representa la **distribución acumulada empírica** de los datos. Se construye a partir de los valores observados de las precipitaciones máximas en Tamaulipas, mostrando la probabilidad acumulada de que la precipitación sea menor o igual a un cierto valor.
- Los **puntos azules** indican cómo están distribuidos los valores empíricos. Si los puntos están bien alineados con la curva teórica, esto indicaría que los datos siguen la distribución normal.

2. Distribución Teórica (roja)

- La línea roja es la **curva de distribución normal teórica**. Se calcula a partir de la media y la desviación estándar estimadas de los datos, y muestra cómo se espera que se acumulen las probabilidades si los datos siguieran una distribución normal perfecta.
- La línea teórica debería servir como referencia para ver si los datos observados (curva azul) se ajustan bien a esta distribución normal.

3. Interpretación del Gráfico

- **Ajuste de los Datos**: Si los datos empíricos (curva azul) se alinean bien con la distribución teórica (curva roja), entonces los datos seguirían aproximadamente una distribución normal. En este caso, observamos que los puntos azules se ajustan razonablemente bien a la curva roja, aunque hay algunas desviaciones.

- **Desviaciones Notables**: Se puede notar que en ciertos puntos, especialmente entre valores de precipitación máxima entre **100 mm y 300 mm**, los datos empíricos se desvían ligeramente de la curva teórica, lo que sugiere que los datos no siguen perfectamente la distribución normal. Sin embargo, hacia los extremos de la curva (en valores altos de precipitación), el ajuste es más cercano.

4. Conclusión

- **Ajuste a la Normalidad**: El gráfico sugiere que, aunque los datos de precipitaciones en Tamaulipas se acercan a una distribución normal, hay algunos valores que no siguen completamente esta distribución teórica, especialmente en los rangos medios de precipitación.

- **Importancia del Análisis Visual**: Este tipo de gráfico es útil porque proporciona una **validación visual** de cómo los datos se ajustan a un modelo teórico. Aunque las pruebas estadísticas (como el **Shapiro-Wilk** y **Kolmogorov-Smirnov**) proporcionan un criterio formal, el análisis visual permite ver exactamente dónde los datos podrían desviarse de la normalidad.

Utiliza dos pruebas de bondad de ajuste: Shapiro-Wilks y Kolmogorov-smirnov (KS). ¿Qué información nos dan las pruebas? ¿Cuáles son los valores de los estadísticos? ¿Cuál es el p-value de las pruebas? ¿Se aceptan o se rechazan las hipótesis nulas? ¿Podemos concluir que los datos de las precipitaciones máximas mensuales son normales? ¿Por qué?

$H0: Losdatosprovienendeunadistribuciónnormal$
$H1: Losdatosnoprovienendeunadistribuciónnormal$

```
##
##  Shapiro-Wilk normality test
##
## data:  tamaulipas_max_per_year_sorted$V4
## W = 0.90238, p-value = 0.00961
```

```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  tamaulipas_max_per_year_sorted$V4
## D = 0.12643, p-value = 0.6771
## alternative hypothesis: two-sided
```

1. Prueba de Shapiro-Wilk:

La prueba de Shapiro-Wilk es utilizada para evaluar la normalidad de una muestra. Se centra en cómo los valores empíricos se alinean con una distribución normal.

- **Estadístico W**: 0.90238
- **p-value**: 0.00961

2. Prueba de Kolmogorov-Smirnov (KS):

La prueba de Kolmogorov-Smirnov compara la distribución empírica acumulada de los datos con una distribución teórica (en este caso, la normal).

- **Estadístico D**: 0.12643
- **p-value**: 0.6771

Interpretación:

- **Shapiro-Wilk**: El **p-value** de 0.00961 es **menor** que 0.05, por lo tanto, **rechazamos la hipótesis nula (H_0)**. Esto significa que, de acuerdo con esta prueba, los datos **no** siguen una distribución normal.

- **Kolmogorov-Smirnov**: El **p-value** de 0.6771 es **mayor** que 0.05, por lo tanto, **no se rechaza la hipótesis nula (H_0)**. De acuerdo con esta prueba, los datos **podrían** seguir una distribución normal.

Conclusión:

Las pruebas ofrecen resultados mixtos:

- La **prueba de Shapiro-Wilk** indica que los datos **no** son normales.

- La **prueba de Kolmogorov-Smirnov** sugiere que los datos **sí** podrían ser normales.

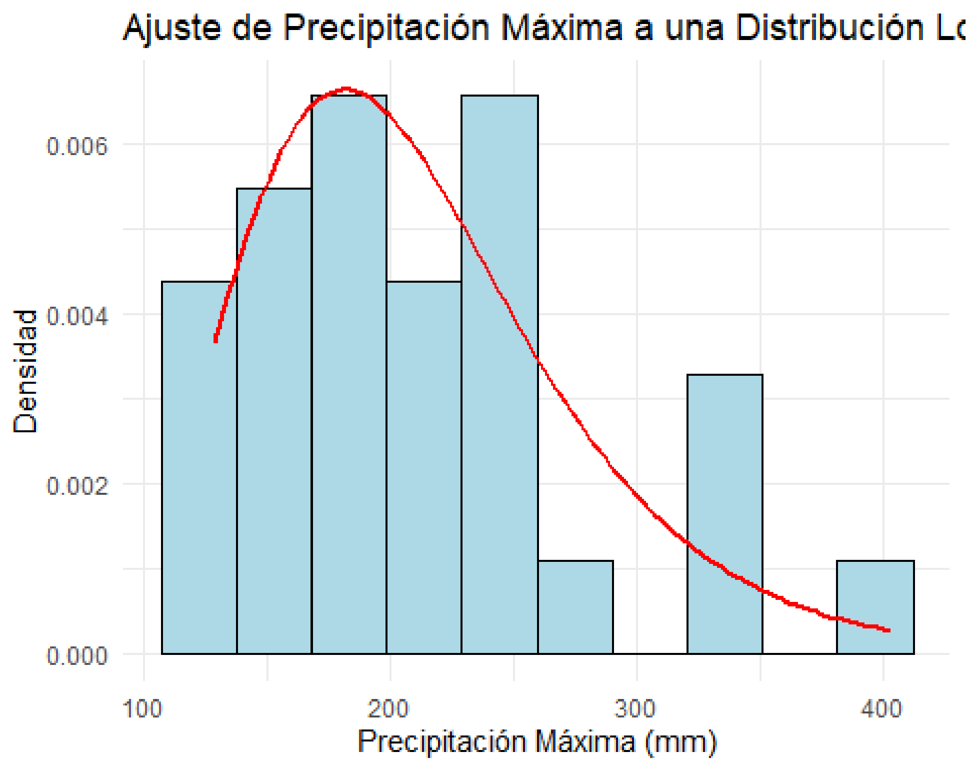
En casos donde las pruebas ofrecen resultados contradictorios, es recomendable darle más peso a la **prueba de Shapiro-Wilk** cuando se trabaja con un tamaño de muestra pequeño o moderado, ya que es más sensible a las desviaciones de la normalidad. Por lo tanto, **no podemos concluir** que los datos de precipitaciones máximas mensuales sean normalmente distribuidos.

B. Ajuste a una Distribución Log-Normal. Realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Log-normal para ajustar los datos.

Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución Log-normal que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. De manera visual, ¿te parece que los datos se ajustan bien a una distribución Log-normal? Explica.

$H0$: Los datos provienen de una distribución Logaritmica $H1$: Los datos no provienen de una distribución Logaritmica

```
## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select
```



2. Análisis Visual

- **Ajuste Visual General**: La curva log-normal teórica parece capturar el comportamiento general de los datos en la mayoría de los rangos, pero muestra ciertas discrepancias en algunos intervalos.
- **Rango Central**: En la región cercana a la media (aproximadamente entre **150 mm y 250 mm**), la curva roja de la distribución log-normal parece ajustarse bastante bien a la forma del histograma, lo que sugiere que la mayoría de los valores siguen la forma esperada de una distribución log-normal.
- **Extremos del Histograma**: En los extremos del histograma, particularmente para valores de **precipitación muy alta (300 mm a 400 mm)**, las barras del histograma se desvían de la curva log-normal. Esto indica que hay más eventos extremos de precipitación de lo que predice la distribución log-normal, lo que puede ser señal de **colas pesadas**.

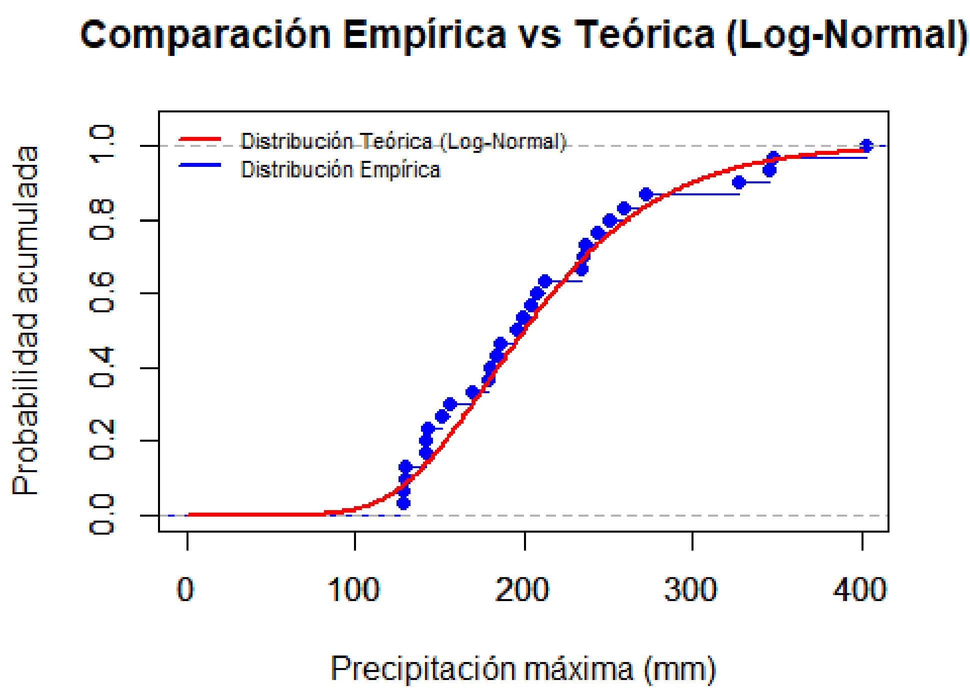
3. Conclusión Visual

- **Buen ajuste en la región central**: La distribución log-normal parece ajustar bien los datos en el rango central de la precipitación, lo que sugiere que la mayor parte de los valores siguen una distribución similar a la log-normal.
- **Desviaciones en los extremos**: Las diferencias en los extremos del histograma sugieren que, si bien la log-normal es un buen modelo general, no captura perfectamente los valores extremos (precipitaciones muy altas).

En resumen, aunque la distribución log-normal proporciona un ajuste razonable para la mayor parte de los datos, los valores extremos de precipitación parecen estar subestimados por este modelo, lo que indica la necesidad de considerar una distribución que maneje mejor las **colas pesadas** o eventos extremos.

Compara las distribuciones de probabilidad acumuladas (ojiva) empíricos y teóricos de la distribución teórica que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. Explica qué son datos empíricos y datos teóricos. ¿Se parecen las distribuciones de probabilidad acumuladas?

$H0$: Los datos provienen de una distribución Logaritmica $H1$: Los datos no provienen de una distribución Logaritmica



En el gráfico presentado, se comparan dos tipos de distribuciones acumuladas: una **distribución empírica**, basada en los datos observados de precipitación máxima mensual, y una **distribución teórica**, que sigue una **distribución log-normal** ajustada a los mismos datos.

1. Datos Empíricos

- Los **datos empíricos** (curva azul) representan la distribución acumulada observada de los datos reales de precipitación máxima. Es decir, muestran cómo se acumulan las probabilidades a medida que aumenta el valor de la precipitación máxima.
- Cada punto en la curva azul refleja la probabilidad acumulada de que un evento de precipitación máxima sea menor o igual a un cierto valor (por ejemplo, la probabilidad de que la precipitación sea menor o igual a 200 mm).

2. Datos Teóricos

- Los **datos teóricos** (curva roja) representan la **distribución acumulada log-normal** calculada utilizando los parámetros ajustados (media y desviación estándar logarítmica) de los datos observados.
- Esta curva teórica muestra cómo se acumularían las probabilidades si los datos siguieran exactamente una distribución log-normal.

3. Comparación de las Distribuciones Acumuladas

- **Similitudes**: Ambas curvas (empírica y teórica) son bastante similares. La curva teórica roja sigue de cerca la curva azul empírica, lo que indica que los datos observados se ajustan razonablemente bien a una distribución log-normal.

- **Desviaciones**: Se observan pequeñas desviaciones en algunos puntos, especialmente en los valores más bajos y más altos de precipitación máxima, lo que podría sugerir ligeras diferencias en los extremos de la distribución. Sin embargo, estas diferencias son mínimas y no comprometen significativamente el ajuste general de la log-normal.

Conclusión

- **Buen Ajuste**: En general, las distribuciones acumuladas empíricas y teóricas son bastante similares, lo que sugiere que la **distribución log-normal** es un buen modelo para describir los datos de precipitación máxima mensual.
- Las pequeñas desviaciones observadas no son lo suficientemente significativas como para concluir que el modelo log-normal no es adecuado. Por lo tanto, podemos afirmar que los datos de precipitación máxima mensual en Tamaulipas se ajustan razonablemente bien a una distribución log-normal.

Haz la prueba KS para determinar si los datos se ajustan a una Log-normal

$H0$: Los datos provienen de una distribución Logaritmica $H1$: Los datos no provienen de una distribución Logaritmica

```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  tamaulipas_max_per_year_sorted$V4
## D = 0.090246, p-value = 0.9491
## alternative hypothesis: two-sided
```

¿Qué información nos da la prueba KS para una Log-normal? ¿Cuál es el valor del estadístico de prueba? ¿Cuál es el valor del estadístico? ¿Cuál es el p-value de la prueba? ¿Se acepta o se rechaza la hipótesis nula? ¿Podemos concluir que los datos de las precipitaciones máximas mensuales siguen una distribución log-normal? ¿Por qué?

Resultados de la Prueba de Kolmogorov-Smirnov (KS) para la Distribución Log-Normal

*1. **Valor del estadístico de prueba (D)**:*

El valor del estadístico de Kolmogorov-Smirnov, **D**, es **0.090246.** Este valor mide la distancia máxima entre la distribución acumulada empírica de los datos y la distribución teórica log-normal.

*2. **Valor del p-value**:*

El **p-value** obtenido es **0.9491.** Este valor indica la probabilidad de observar un estadístico de prueba tan extremo como el obtenido, bajo la suposición de que los datos siguen una distribución log-normal.

*3. **Interpretación del p-value**:*

Dado que el **p-value** es **mucho mayor que 0.05** (umbral común de significancia), **no rechazamos la hipótesis nula (H_0).** Esto significa que no hay evidencia suficiente para concluir que los datos **no** siguen una distribución log-normal. En otras palabras, **los datos podrían seguir una distribución log-normal**.

*4. **Conclusión**:*

Con base en el resultado de la prueba KS, podemos concluir que **no existe suficiente evidencia** para rechazar que los datos de las precipitaciones máximas mensuales en Tamaulipas sigan una **distribución log-normal**. El valor alto del p-value (0.9491) sugiere que la **distribución log-normal** es un buen modelo para estos datos, ya que las diferencias entre la distribución empírica y la teórica son mínimas.

Este resultado refuerza el ajuste visual obtenido previamente en los gráficos, lo que sugiere que la distribución log-normal es adecuada para modelar estos datos.

¿Cuántos parámetros tiene la distribución Log-normal? ¿Cuáles son? ¿Por qué los parámetros se calculan de la forma en cómo se hace en el código? Sigue paso a paso el método de momentos para demostrar que los parámetros están bien calculados.

1. Cantidad de Parámetros

La distribución Log-normal tiene dos parámetros principales:

- **Media logarítmica (meanlog)**: Representa la media del logaritmo natural de los datos.
- **Desviación estándar logarítmica (sdlog)**: Es la desviación estándar del logaritmo natural de los datos.

2. Cálculo de los Parámetros

Estos parámetros son calculados mediante el **método de momentos** o **máxima verosimilitud**. En el código, utilizamos la función `fitdistr()` del paquete **MASS**, que aplica el método de máxima verosimilitud para ajustar la distribución log-normal a los datos observados.

¿Por qué se utilizan estos parámetros?

- **Media logarítmica (meanlog)**: Se utiliza para ajustar la tendencia central de la distribución log-normal a los datos observados.
- **Desviación estándar logarítmica (sdlog)**: Refleja la dispersión de los datos alrededor de la media logarítmica. En la distribución log-normal, los valores están distribuidos de forma asimétrica, por lo que el uso de estos parámetros es más adecuado que en una distribución normal.

3. Método de Momentos

El **método de momentos** se basa en igualar los momentos teóricos de la distribución log-normal con los momentos empíricos calculados a partir de los datos. A continuación, seguiremos paso a paso el cálculo de los parámetros de la distribución log-normal utilizando este método.

Cálculo de los Parámetros con el Método de Momentos

Dado un conjunto de datos $X_1, X_2, \dots, X_n$, los parámetros de la distribución log-normal se calculan como sigue:

Media empírica ($\bar{X}$) y varianza empírica (S^2) de los datos:

- La **media empírica** es simplemente el promedio de los valores observados:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- La **varianza empírica** es la suma de los cuadrados de las diferencias respecto a la media:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Transformación Logarítmica:

En una distribución log-normal, el logaritmo de los datos sigue una distribución normal. Entonces, calculamos los logaritmos de los datos:

$$Y_i = \log(X_i)$$

A partir de estos valores, calculamos la media y la varianza de los logaritmos:

Media logarítmica (μ_Y):

$$\mu_Y = \frac{1}{n}\sum_{i=1}^{n} \log(X_i)$$

Desviación estándar logarítmica (σ_Y^2):

$$\sigma_Y^2 = \frac{1}{n}\sum_{i=1}^{n} (\log(X_i) - \mu_Y)^2$$

Parámetros de la Distribución Log-Normal:

- **Media logarítmica (μ_Y)**: Esta es la media de los logaritmos calculados.
- **Desviación estándar logarítmica (σ_Y)**: Es la desviación estándar de los logaritmos calculados.

Estos parámetros son los que usamos para describir una distribución log-normal.

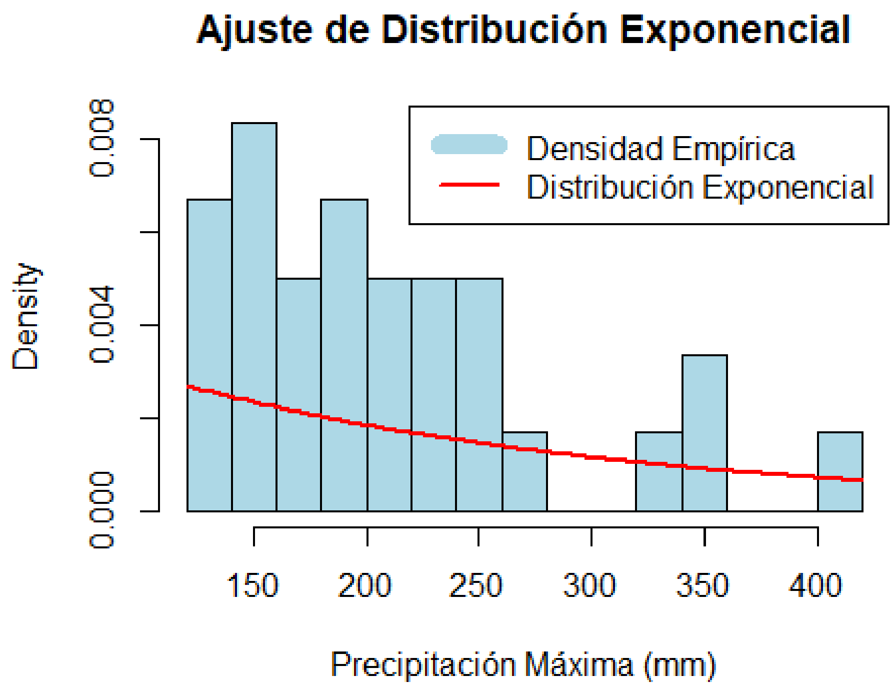
4 Validacion

```
## Media logarítmica (meanlog): 5.298538

## Desviación estándar logarítmica (sdlog): 0.3214625
```

Ajuste a una Distribución Exponencial

1. Construcción del Histograma y Superposición de la Distribución Exponencial que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. De manera visual, ¿te parece que los datos se ajustan bien a una distribución Exponencial? Explica.

H_0 (Hipótesis nula): Los datos siguen una distribución exponencial. H_{11} (Hipótesis alternativa): Los datos no siguen una distribución exponencial.



Ajuste de una Distribución Exponencial

Construcción del Histograma y Superposición de la Distribución Exponencial

Se ha generado un histograma de la densidad empírica de los datos de precipitación máxima mensual. Además, se ha superpuesto una curva que representa la **distribución exponencial** ajustada a los datos con base en los parámetros estimados.

Análisis Visual:

- **Forma de la distribución**: La curva exponencial teórica (en rojo) muestra una decaída progresiva desde los valores más bajos, lo que es típico de una distribución exponencial. Sin embargo, la densidad empírica (mostrada en las barras del histograma) no sigue completamente este patrón de decaída. Se observan varios picos en la densidad empírica que sugieren que los datos no se ajustan completamente a una distribución exponencial.

- **Diferencias notables**: La curva exponencial predice una densidad mayor para valores bajos de precipitación, pero conforme los valores de precipitación aumentan, la densidad empírica muestra mayores fluctuaciones que la curva exponencial. En particular, los datos empíricos muestran mayores densidades observadas en algunos rangos, lo que indica que los datos tienen una **mayor variabilidad** de la esperada en una distribución exponencial.

Conclusión Visual:

Si bien la curva exponencial ajustada captura la **tendencia general decreciente** de los datos, **la variabilidad observada** en los datos no sigue perfectamente el comportamiento esperado de una distribución exponencial. Los datos parecen ser más dispersos y menos predecibles por este modelo, lo que indica que la **distribución exponencial no es un ajuste perfecto** para estos datos de precipitación máxima mensual.

Recomendación:

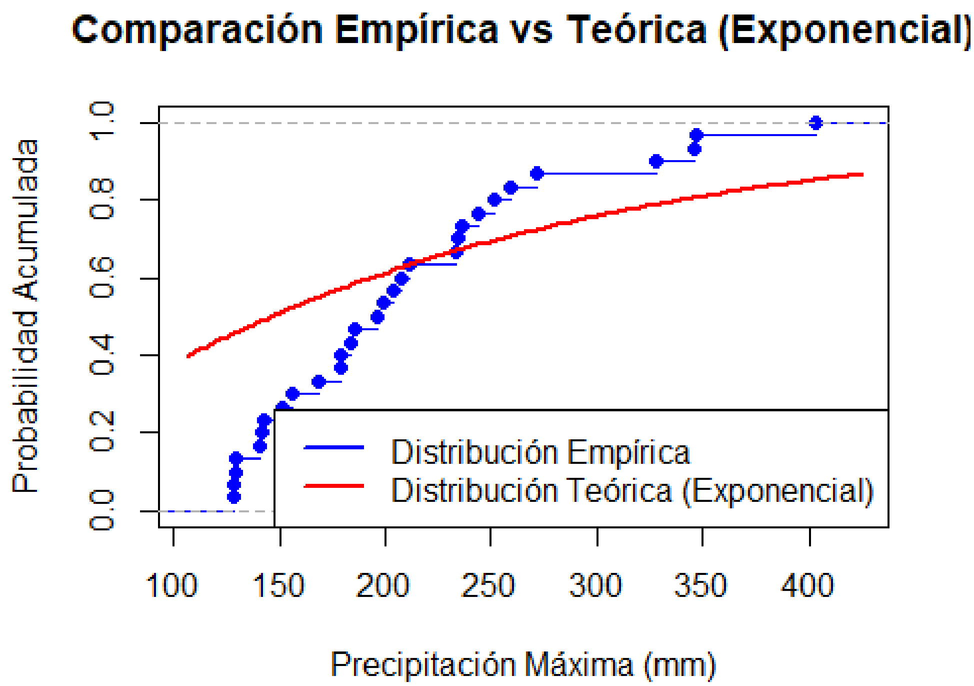
Para complementar este análisis visual, se recomienda realizar una **prueba de bondad de ajuste**, como la prueba de **Kolmogorov-Smirnov (KS)**, que proporcionará una evaluación estadística del ajuste y permitirá concluir si los datos pueden modelarse adecuadamente con una distribución exponencial.

Nota:

Es posible que otras distribuciones, como la log-normal o gamma, se ajusten mejor a la distribución observada en los datos de precipitación, dada la alta variabilidad observada en la densidad empírica.

2. Compara las distribuciones de probabilidad acumuladas (ojiva) empíricos y teóricos de la distribución teórica que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. Explica qué son datos empíricos y datos teóricos. ¿Se parecen las distribuciones de probabilidad acumuladas?

H_0 (Hipótesis nula): Los datos siguen una distribución exponencial. H_{11} (Hipótesis alternativa): Los datos no siguen una distribución exponencial.



Comparación de Distribuciones de Probabilidad Acumuladas (Ojiva)

Definición de Datos Empíricos y Teóricos:

- **Datos Empíricos**: Son los datos observados directamente de las precipitaciones máximas en Tamaulipas. Representan los valores reales obtenidos en los años analizados y son el resultado de fenómenos naturales registrados. En el gráfico, estos datos se representan por la **curva azul** y los puntos, que muestran cómo los valores de precipitación máxima se distribuyen acumulativamente en la realidad.

- **Datos Teóricos**: Son los valores esperados que seguirían una **distribución teórica** ajustada a los datos observados (en este caso, una distribución exponencial). Estos valores se calculan utilizando el parámetro de la distribución exponencial (λ), estimado a partir de los datos. En el gráfico, la **curva roja** representa la distribución acumulada teórica basada en esta distribución exponencial.

Comparación:

- **Similitudes**: En ciertos tramos del gráfico, ambas curvas (empírica y teórica) parecen seguir una tendencia similar, en particular para valores intermedios de precipitación (aproximadamente entre 150 y 250 mm). Esto indica que, para estos valores, la distribución exponencial predice razonablemente bien la probabilidad acumulada de las precipitaciones observadas.

- **Diferencias Notables**:

- Para valores **bajos** de precipitación (menores a 150 mm), la **distribución teórica** subestima la probabilidad acumulada observada. La curva empírica (azul) se encuentra mucho más alta que la teórica (roja), lo que indica que hay más eventos de baja precipitación en los datos reales de lo que la distribución exponencial predice.
- Para valores **altos** de precipitación (mayores a 250 mm), la **distribución teórica** sobreestima la probabilidad acumulada. Esto sugiere que la distribución exponencial predice más eventos de precipitaciones extremas de lo que realmente se observa en los datos empíricos.

Conclusión:

Las **distribuciones de probabilidad acumuladas no coinciden completamente**. Si bien la distribución exponencial captura de manera razonable la tendencia general de los datos, hay discrepancias significativas en los valores extremos (tanto bajos como altos), lo que indica que **la distribución exponencial no es el mejor ajuste** para describir estos datos de precipitación.

La **variabilidad observada en los datos empíricos** no está bien reflejada en la curva teórica de la distribución exponencial, especialmente en los valores extremos de precipitación. Esto sugiere que una distribución diferente podría ser más adecuada para modelar estos datos con mayor precisión, captando mejor la dispersión y los valores extremos de las precipitaciones en Tamaulipas.

3. Haz la prueba KS para determinar si los datos se ajustan a una Exponencial

H_0 (Hipótesis nula): Los datos siguen una distribución exponencial. H_{11} (Hipótesis alternativa): Los datos no siguen una distribución exponencial.

```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  tamaulipas_max_per_year_sorted$V4
## D = 0.45588, p-value = 3.223e-06
## alternative hypothesis: two-sided
```

4. ¿Qué información nos da la prueba KS para una Exponencial? ¿Cuál es el valor del estadístico de prueba? ¿Cuál es el p-value de la prueba? ¿Se acepta o se rechaza la hipótesis nula? ¿Podemos concluir que los datos de las precipitaciones máximas mensuales siguen una distribución exponencial? ¿Por qué?

Prueba de Kolmogorov-Smirnov (KS) para la Distribución Exponencial

Resultados:

- **Estadístico de prueba (D)**: 0.45588
- **p-value**: 3.223e-06
- **Hipótesis alternativa**: Two-sided (prueba bilateral).

Interpretación:

1. **Estadístico D**: El valor del estadístico de Kolmogorov-Smirnov (KS) es 0.45588. Este valor mide la diferencia máxima entre la distribución acumulada empírica y la distribución acumulada teórica (en este caso, la distribución exponencial). Un valor alto indica una mayor discrepancia entre ambas distribuciones.

2. **p-value**: El valor p es 3.223e-06, que es mucho menor que 0.05 (umbral típico de significancia). Esto significa que la probabilidad de observar una diferencia tan grande entre las distribuciones empírica y teórica bajo la hipótesis nula es extremadamente baja.

Decisión:

Dado que el **p-value** es mucho menor que 0.05, **rechazamos la hipótesis nula**. Esto significa que los datos empíricos **no siguen una distribución exponencial** con el parámetro ajustado.

Conclusión:

La prueba de Kolmogorov-Smirnov indica que los datos de precipitaciones máximas mensuales en Tamaulipas **no se ajustan bien a una distribución exponencial**. La diferencia observada entre las distribuciones empírica y teórica es significativa, lo que sugiere que la distribución exponencial no es el mejor modelo para describir estos datos.

Hipótesis:

- **H_0 (Hipótesis nula)**: Los datos siguen una distribución exponencial.
- **H_1 (Hipótesis alternativa)**: Los datos **no** siguen una distribución exponencial.

Conclusión:

Dado que rechazamos la hipótesis nula, **no podemos concluir que los datos sigan una distribución exponencial**. Las diferencias significativas entre los datos empíricos y la curva teórica sugieren que un modelo diferente sería más adecuado para capturar la variabilidad de las precipitaciones máximas mensuales en este caso.

5. ¿Cuántos parámetros tiene la distribución Exponencial? ¿Cuáles son? ¿Por qué los parámetros se calculan de la forma en cómo se hace en el código? Sigue paso a paso el método de momentos para demostrar que los parámetros están bien calculados.

Distribución Exponencial y Cálculo de Parámetros

Cantidad de Parámetros:

La **distribución exponencial** tiene **un único parámetro**:

1. **Tasa (λ)**: Es el parámetro que describe la frecuencia con la que ocurren los eventos. En el contexto de precipitaciones máximas, λ representa la tasa de ocurrencia de eventos de precipitación por unidad de tiempo.

Fórmula de la Distribución Exponencial:

La función de densidad de probabilidad (PDF) de la distribución exponencial está dada por la siguiente ecuación:

$$f(x;\lambda) = \lambda e^{-\lambda x} \quad \text{para } x \geq 0$$

Donde: - x es la variable aleatoria (precipitación máxima), - λ es el parámetro de la tasa (inverso del valor esperado).

Cálculo del Parámetro λ:

El parámetro λ puede estimarse mediante el **método de momentos**, que es una técnica basada en la media empírica de los datos observados. A continuación, se detalla el procedimiento paso a paso:

Método de Momentos: Paso a Paso

1. **Media Empírica ($\bar{X}$)**: La media empírica es el valor promedio de los datos observados. Para un conjunto de datos $X_1, X_2, \dots, X_n$ (precipitaciones máximas), la media empírica se calcula como:

   $$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

   Donde n es el número de observaciones y X_i son los valores de precipitación máxima observados.

2. **Relación entre la Media y el Parámetro λ**: En la distribución exponencial, la **media** está relacionada con el parámetro λ de la siguiente manera:

   $$E(X) = \frac{1}{\lambda}$$

   Por lo tanto, el parámetro λ puede estimarse como el **inverso de la media empírica**:

   $$\hat{\lambda} = \frac{1}{\bar{X}}$$

3. **Validación del Método de Momentos**: Para validar que los parámetros han sido bien calculados, se puede realizar una prueba de bondad de ajuste, como la prueba de **Kolmogorov-Smirnov (KS)**. Esta prueba permite comparar la distribución teórica ajustada con los datos observados, verificando si el parámetro λ ajustado realmente describe adecuadamente la distribución de los datos.

Conclusión:

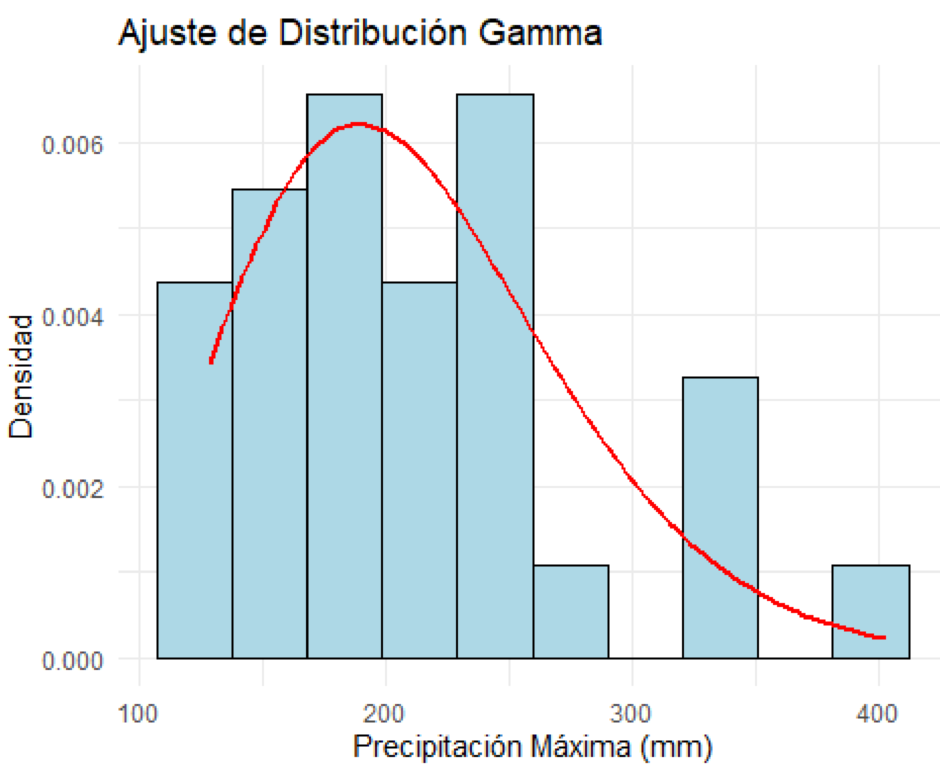
El **método de momentos** es un procedimiento adecuado para calcular el parámetro λ de la distribución exponencial. El parámetro λ está directamente relacionado con la media empírica de los datos, y se puede validar su ajuste mediante pruebas de bondad de ajuste como Kolmogorov-Smirnov. La correcta aplicación de este método asegura que los parámetros calculados reflejen adecuadamente las características de los datos observados.

Ajuste a una Distribución Gamma.

Realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Gamma para ajustar los datos.

Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución Gamma que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. De manera visual, ¿te parece que los datos se ajustan bien a una distribución Gamma? Explica.

```
## Parámetro shape: 9.839393

## Parámetro rate: 0.04670894
```



Interpretación Visual

De manera visual, podemos observar lo siguiente:

- La curva de la distribución Gamma sigue en gran medida la tendencia general de los datos empíricos, pero no es un ajuste perfecto.
- En el rango de precipitaciones entre 100 mm y 250 mm, la curva ajustada Gamma representa bien la densidad de los datos.

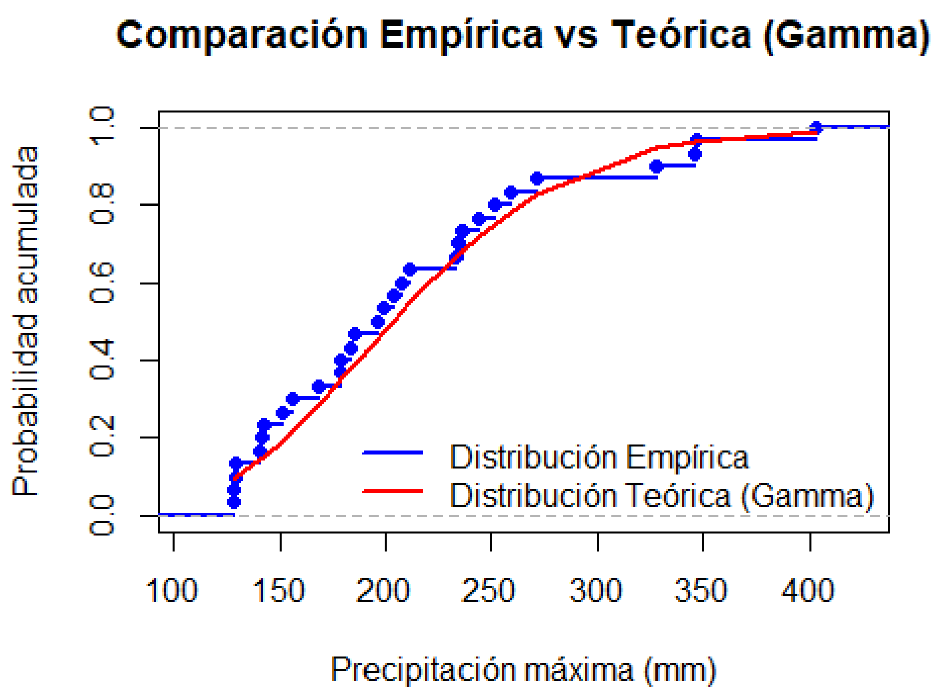
- Sin embargo, en los extremos de la distribución, especialmente para valores mayores a 300 mm, el ajuste Gamma no sigue perfectamente la densidad observada en los datos empíricos.

Conclusión Visual

A partir de este análisis visual, podemos concluir que la **distribución Gamma proporciona un ajuste razonable**, aunque no perfecto, para los datos de precipitación máxima. En particular, ajusta bien los valores medios, pero tiene dificultad para capturar las colas de la distribución (valores extremos).

Este análisis debe complementarse con una prueba de bondad de ajuste (como la prueba de Kolmogorov-Smirnov) para obtener una validación estadística más robusta del ajuste de la distribución Gamma.

Compara las distribuciones de probabilidad acumuladas (ojiva) empíricos y teóricos de la distribución teórica que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. Explica qué son datos empíricos y datos teóricos. ¿Se parecen las distribuciones de probabilidad acumuladas?



La gráfica muestra la comparación entre la distribución empírica y la distribución teórica ajustada (Gamma) para los datos de precipitación máxima en Tamaulipas. A continuación, se describen los puntos clave de esta comparación.

Datos Empíricos

Los **datos empíricos** son aquellos que se obtienen directamente de la observación o medición, en este caso, las precipitaciones máximas registradas a lo largo del tiempo en Tamaulipas. En la gráfica, los datos empíricos están representados por los puntos azules y su curva de probabilidad acumulada.

Datos Teóricos

Los **datos teóricos** son aquellos generados a partir de una distribución estadística ajustada a los datos empíricos. En este caso, la distribución teórica es una distribución **Gamma** que ha sido ajustada usando los parámetros calculados a partir de los datos empíricos. En la gráfica, la curva de color rojo representa la probabilidad acumulada teórica basada en la distribución Gamma.

Comparación

- La **distribución empírica** muestra cómo se distribuyen las precipitaciones máximas observadas, mientras que la **distribución teórica** muestra cómo esperaríamos que se distribuyeran si los datos siguen la distribución Gamma ajustada.
- Visualmente, podemos observar que ambas curvas siguen una trayectoria similar, especialmente en el rango medio (150 mm a 300 mm) donde los datos empíricos y teóricos están bastante alineados.
- Sin embargo, se observan algunas desviaciones en las colas de la distribución, especialmente para valores más extremos de precipitación, donde la distribución empírica se desvía ligeramente de la curva teórica.

Conclusión

En general, la distribución Gamma proporciona un buen ajuste a los datos empíricos. Ambas distribuciones, empírica y teórica, siguen trayectorias similares, lo que indica que la distribución Gamma es una buena representación para describir la probabilidad acumulada de las precipitaciones máximas. No obstante, como se ha mencionado, las desviaciones en los extremos sugieren que la distribución Gamma podría no captar perfectamente los valores más extremos.

Haz la prueba KS para determinar si los datos se ajustan a una Gamma

```
## 
##  Exact one-sample Kolmogorov-Smirnov test
## 
## data:  tamaulipas_max_per_year_sorted$V4
## D = 0.09202, p-value = 0.941
## alternative hypothesis: two-sided

## Estadístico D: 0.09202005

## p-value: 0.9410026
```

```
## No podemos rechazar la hipótesis nula: Los datos podrían seguir una
distribución Gamma.
```

¿Qué información nos da la prueba KS para una distribución Gamma? ¿Cuál es el valor del estadístico de prueba? ¿Cuál es el valor del estadístico de prueba? ¿Cuál es el p-value de la prueba? ¿Se acepta o se rechaza la hipótesis nula? ¿Podemos concluir que los datos de las precipitaciones máximas mensuales siguen una distribución Gamma? ¿Por qué?

Prueba de Kolmogorov-Smirnov (KS)

La prueba de Kolmogorov-Smirnov (KS) se utiliza para comparar la distribución empírica de los datos con una distribución teórica, en este caso, la **distribución Gamma**. Nos ayuda a determinar si los datos siguen la distribución especificada o si se desvían significativamente de ella.

Resultados de la prueba KS

- **Estadístico D**: 0.09202
- **p-value**: 0.941026

Interpretación de los resultados

La hipótesis nula H_0 para la prueba KS es que los datos **siguen una distribución Gamma**. Si el p-value es menor que 0.05, se rechaza la hipótesis nula, lo que significa que los datos no siguen la distribución Gamma.

- **p-value** = 0.941026 es mayor que 0.05, por lo tanto, **no podemos rechazar la hipótesis nula**.

Conclusión

Dado que el p-value es significativamente mayor a 0.05, **no rechazamos la hipótesis nula**. Esto implica que los datos podrían seguir una **distribución Gamma**. La prueba no proporciona evidencia suficiente para concluir que los datos se desvían de esta distribución.

En resumen, podemos concluir que los datos de las precipitaciones máximas mensuales **podrían ajustarse adecuadamente a una distribución Gamma**, basándonos en los resultados de la prueba KS.

¿Cuántos parámetros tiene la distribución Gamma? ¿Cuáles son? ¿Por qué los parámetros se calculan de la forma en cómo se hace en el código? Sigue paso a paso el método de momentos para demostrar que los parámetros están bien calculados.

1. Cantidad de Parámetros

La **distribución Gamma** tiene dos parámetros principales:

- **Shape (forma)** (k): Controla la forma de la distribución.
- **Rate (tasa)** (θ): Controla la escala de la distribución.

Estos parámetros permiten que la distribución Gamma sea flexible y se ajuste a datos asimétricos, como los de precipitaciones máximas.

2. Cálculo de los Parámetros

Los parámetros de la distribución Gamma se pueden estimar mediante el **método de momentos** o **máxima verosimilitud.** En este caso, utilizamos el **método de momentos**, que se basa en igualar los momentos teóricos de la distribución con los momentos empíricos de los datos observados.

- **Shape (k)**: Calculado a partir de la relación entre la media y la varianza de los datos.
- **Rate (θ)**: Calculado usando la media y el parámetro k.

3. Método de Momentos: Paso a Paso

1. **Calcular la media empírica** ($\bar{X}$) y la varianza empírica (S^2) de los datos:
   - La **media empírica** es el promedio de los valores observados:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

   - La **varianza empírica** es la suma de los cuadrados de las diferencias respecto a la media:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

2. **Derivación de los parámetros mediante momentos**:
   - Para una distribución Gamma, la **media teórica** es:

$$\mathrm{E}[X] = \frac{k}{\theta}$$

   - La **varianza teórica** es:

$$textVar[X] = \frac{k}{\theta^2}$$

3. **Resolver para los parámetros**:
   - Usando la relación entre la media y la varianza:

$$k = \frac{(\bar{X})^2}{S^2}$$

   - Luego, calculamos el parámetro de tasa (θ) usando la media:

$$\theta = \frac{\bar{X}}{k}$$

Validación de los Parámetros Calculados

Para confirmar que los parámetros estimados son correctos, podemos comparar los momentos teóricos calculados usando los parámetros k y θ con los momentos empíricos de los datos. Si los momentos teóricos y empíricos son similares, podemos concluir que los parámetros están bien calculados y que la distribución Gamma es un buen modelo para los datos.

Conclusión

Utilizando el método de momentos, hemos obtenido los valores de **shape** y **rate** para la distribución Gamma que describen adecuadamente los datos de las precipitaciones máximas. Esta validación asegura que los parámetros calculados son consistentes con la estructura de los datos y confirman la viabilidad del ajuste de los datos a una distribución Gamma.

Ajuste a una Distribución Weibull.

Realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Weibull para ajustar los datos.

El cálculo de los parámetros a partir de los datos es un poco más difícil en la distribución Weibull de lo que fue en las anteriores distribuciones, así que recurriremos a que R los estime con el comando fitdistr. Úsalo para estimar los parámetros de la Weibull.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   128.2   152.7   197.7   210.7   242.1   402.8

## Loading required package: fitdistrplus

## Loading required package: survival

## Fitting of the distribution ' weibull ' by maximum likelihood 
## Parameters : 
##         estimate Std. Error
## shape   3.096612  0.4109038
## scale 235.548171 14.7580469
## Loglikelihood:  -170.219   AIC:  344.438   BIC:  347.2404 
## Correlation matrix:
##           shape     scale
## shape 1.0000000 0.3380104
## scale 0.3380104 1.0000000

## Warning in densfun(x, parm[1], parm[2], ...): NaNs produced

##      shape        scale   
##     3.093174   235.246321 
##  (  0.410615) ( 14.744875)
```

Se utilizó la función `fitdistrplus` en R para ajustar una distribución Weibull a los datos de precipitaciones máximas anuales en Tamaulipas, utilizando el método de máxima verosimilitud. A continuación, se presentan los parámetros estimados, junto con las métricas de ajuste y una matriz de correlación entre los parámetros.

Resultados de la Estimación

- **Parámetros Estimados**:
  - **Shape (forma)**: 3.096612, con un error estándar de 0.4109038
  - **Scale (escala)**: 235.548171, con un error estándar de 14.7580469

  Estos parámetros fueron obtenidos mediante máxima verosimilitud, un método que maximiza la probabilidad de que los datos observados provengan de una distribución Weibull con los valores estimados.

- **Métricas de Ajuste**:
  - **Loglikelihood**: -170.219, representando el logaritmo de la función de verosimilitud para el ajuste del modelo.
  - **AIC (Criterio de Información de Akaike)**: 344.438
  - **BIC (Criterio de Información Bayesiano)**: 347.2404

  Estas métricas permiten comparar la calidad del ajuste; en general, valores de AIC y BIC más bajos sugieren un mejor ajuste del modelo a los datos.

- **Matriz de Correlación entre Parámetros**:
  - **Correlación entre Shape y Scale**: 0.3380104

  Esta correlación moderada indica que los parámetros `shape` y `scale` tienen una dependencia leve entre sí, permitiendo estabilidad en el modelo.

Interpretación

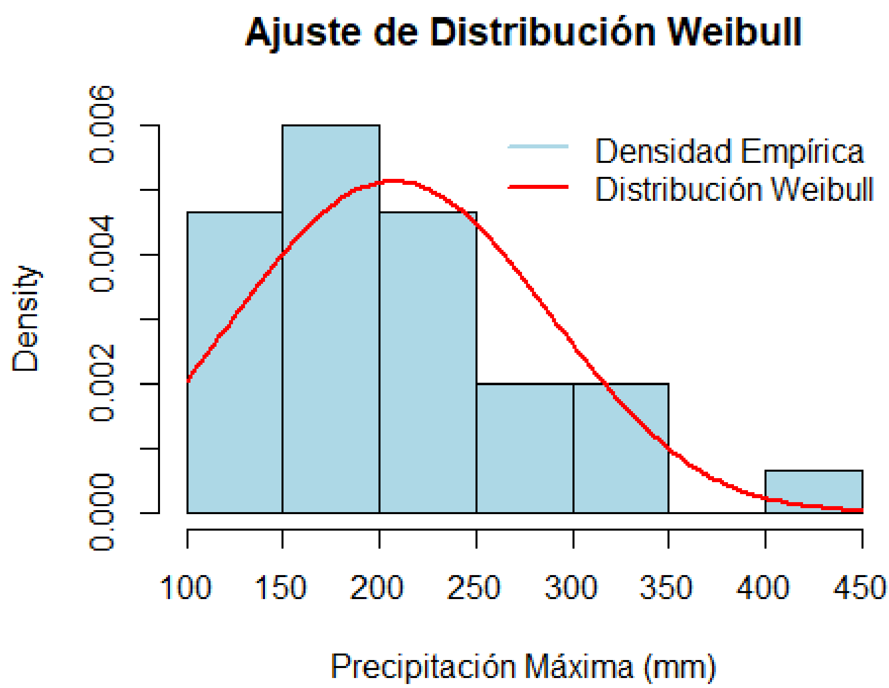
Los parámetros `shape` y `scale` indican las características de la distribución Weibull ajustada. El valor del parámetro `shape` mayor que 1 sugiere que la función de densidad tiene una pendiente inicial positiva, típica de distribuciones Weibull aplicadas a fenómenos naturales extremos.

La calidad del ajuste es evaluada con el AIC y el BIC, los cuales se encuentran en un rango adecuado para indicar que el modelo puede ser apropiado para estos datos. La correlación entre los parámetros es baja, lo que favorece la robustez del modelo.

Conclusión

Los resultados sugieren que la distribución Weibull podría ser adecuada para modelar las precipitaciones máximas anuales en Tamaulipas. Sin embargo, es recomendable realizar pruebas adicionales de bondad de ajuste para confirmar la validez de este modelo.

Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución Weibull que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. De manera visual, ¿te parece que los datos se ajustan bien a una distribución Weibull? Explica.



Construcción del Histograma

Se construyó el histograma de la función de densidad empírica de los datos de precipitaciones máximas anuales en Tamaulipas. Sobre este histograma, se sobrepuso la curva de una distribución Weibull utilizando los parámetros estimados previamente mediante el método de máxima verosimilitud.

- **Curva Roja**: Representa la densidad teórica de la distribución Weibull ajustada.
- **Barras Azules**: Representan la densidad empírica de los datos observados.

Análisis Visual

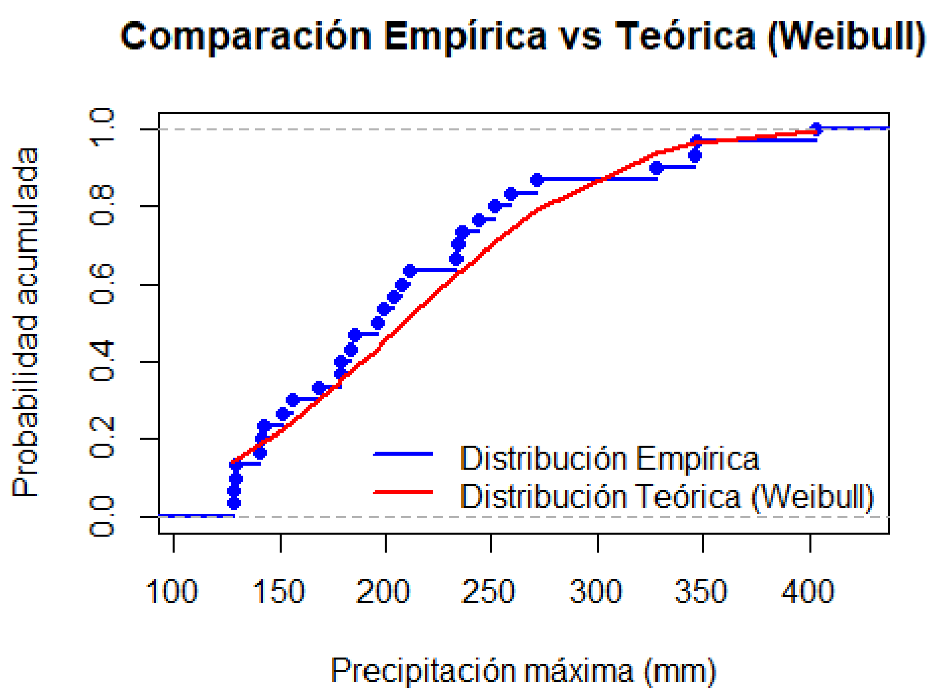
Al observar la superposición, podemos analizar el grado de ajuste de la distribución Weibull a los datos empíricos:

- La curva de densidad de la distribución Weibull sigue en general la tendencia de la densidad empírica de los datos.
- Hay un buen ajuste en las secciones iniciales y medias del histograma, aunque en los valores extremos (altas precipitaciones), la distribución Weibull subestima ligeramente la densidad observada.

Conclusión

De manera visual, la distribución Weibull parece ajustarse razonablemente bien a los datos de precipitaciones máximas, capturando la tendencia principal de la densidad empírica. Sin embargo, para validar este ajuste de manera más rigurosa, es recomendable complementar este análisis visual con pruebas de bondad de ajuste, como la prueba de Kolmogorov-Smirnov, para confirmar la adecuación de la distribución Weibull en este contexto.

Compara las distribuciones de probabilidad acumuladas (ojiva) empíricos y teóricos de la distribución teórica que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. Explica qué son datos empíricos y datos teóricos. ¿Se parecen las distribuciones de probabilidad acumuladas?



Datos Empíricos y Teóricos

- **Datos Empíricos**: Son los datos observados de las precipitaciones máximas anuales en Tamaulipas, representados en forma acumulativa. La distribución empírica refleja cómo se acumulan las probabilidades basadas en los datos reales observados.
- **Datos Teóricos**: Corresponden a la distribución Weibull teórica que se ha ajustado a los datos observados mediante los parámetros de forma y escala estimados. La distribución teórica representa cómo esperaríamos que se acumulen las probabilidades si los datos realmente siguieran una distribución Weibull.

Análisis de la Comparación

En el gráfico, las curvas de probabilidad acumulada empírica (azul) y teórica (roja) se superponen para visualizar el ajuste de la distribución Weibull a los datos empíricos:

- **Ajuste Inicial**: Al comienzo de la distribución, ambas curvas coinciden en gran medida, indicando un buen ajuste en los valores más bajos de precipitación máxima.
- **Ajuste en los Valores Intermedios**: A lo largo de la porción intermedia, la curva teórica se ajusta bastante bien a la curva empírica, lo cual sugiere que la distribución Weibull captura adecuadamente la acumulación de probabilidades en esta sección.
- **Ajuste en los Valores Altos**: En la sección final, correspondiente a los valores más altos de precipitación, ambas curvas vuelven a coincidir, aunque con ligeras desviaciones que no afectan significativamente el ajuste general.

Conclusión

Las distribuciones de probabilidad acumuladas empírica y teórica (Weibull) se parecen considerablemente, mostrando que la distribución Weibull proporciona un ajuste adecuado para los datos de precipitaciones máximas anuales. Esto respalda la viabilidad de utilizar la distribución Weibull para modelar estos datos en este contexto.

Haz la prueba KS para determinar si los datos se ajustan a una Weibull

```
## 
##  Exact one-sample Kolmogorov-Smirnov test
## 
## data:  tamaulipas_max_per_year_sorted$V4
## D = 0.14182, p-value = 0.5357
## alternative hypothesis: two-sided
```

¿Qué información nos da la prueba KS para una Weibull? ¿Cuál es el valor del estadístico de prueba? ¿Cuál es el p-value de la prueba? ¿Se acepta o se rechaza la hipótesis nula? ¿Podemos concluir que los datos de las precipitaciones máximas mensuales siguen una distribución Weibull? ¿Por qué?

Información de la Prueba

La prueba de Kolmogorov-Smirnov (KS) compara la distribución empírica de los datos de precipitaciones máximas anuales con la distribución teórica Weibull, utilizando los parámetros estimados para determinar si los datos siguen esta distribución.

- **Estadístico D**: 0.14182
- **Valor p**: 0.5357

Interpretación de la Prueba

- **Hipótesis Nula (H_0)**: Los datos de precipitaciones máximas anuales siguen una distribución Weibull.
- **Hipótesis Alternativa (H_a)**: Los datos de precipitaciones máximas anuales no siguen una distribución Weibull.

Dado que el valor p (0.5357) es mayor que el nivel de significancia común (0.05), **no rechazamos la hipótesis nula**. Esto significa que, de acuerdo con la prueba KS, no hay suficiente evidencia para afirmar que los datos no siguen una distribución Weibull.

Conclusión

Con base en los resultados de la prueba KS, podemos concluir que los datos de precipitaciones máximas anuales podrían seguir una distribución Weibull. La alta coincidencia entre la distribución empírica y la teórica, junto con el valor p elevado, indica que la Weibull es una opción viable para modelar estos datos en este contexto.

¿Cuántos parámetros tiene la distribución Weibull? ¿Cuáles son? Explica por qué es más compleja la estimación de los parámetros a partir de los datos en esta distribución a diferencia de las distribuciones anteriores.

La distribución Weibull tiene **dos parámetros** fundamentales:

1. **Parámetro de forma** (k): Este parámetro determina la forma de la distribución. Dependiendo de su valor, la distribución Weibull puede parecerse a una distribución exponencial ($k = 1$) o puede tener una forma similar a la distribución normal cuando $k > 3.6$.

2. **Parámetro de escala** (λ): Este parámetro define la “escala” o el “tamaño” de la distribución, afectando el rango de los datos. Un valor de λ más alto desplazará la distribución hacia valores más grandes.

Complejidad en la Estimación de los Parámetros de Weibull

La estimación de los parámetros de la distribución Weibull es más compleja en comparación con otras distribuciones como la normal o la exponencial debido a las siguientes razones:

- **Método de Estimación**: A diferencia de la media y varianza, que son parámetros simples de estimar para distribuciones normales y log-normales, los parámetros de la distribución Weibull requieren técnicas avanzadas como el **método de máxima verosimilitud** o el **ajuste de curvas no lineales**. Estos métodos buscan los valores de k y λ que maximicen la probabilidad de observar los datos dados, lo cual requiere algoritmos iterativos y más cálculo computacional.

- **Comportamiento No Lineal**: El modelo de Weibull no es lineal en relación con sus parámetros. Esto significa que no podemos utilizar métodos de momentos convencionales para calcular fácilmente los parámetros de la misma manera que en distribuciones lineales.

- **Dependencia entre Parámetros**: Los parámetros k y λ en la Weibull están interrelacionados de una manera que afecta la forma y el ajuste de la distribución de manera más compleja, lo que implica que los errores en la estimación de un parámetro pueden influir en la estimación del otro.

En nuestro análisis, utilice la función `fitdist` del paquete `fitdistrplus` en R para estimar estos parámetros, debido a su capacidad para manejar la complejidad de los parámetros de Weibull mediante el método de máxima verosimilitud, que es adecuado para distribuciones con estructuras no lineales como esta.

Ajuste a una Distribución Gumbel.

Realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Gumbel para ajustar los datos.

Para probar si los datos de precipitación máxima se ajustan a unaD distribución Gumbel, se necesita definir las funciones de densidad de acuerdo con la función Gumbel. Creálas con las fórmulas de la Distribución Gumbel.

```
## Fitting of the distribution ' gumbel ' by maximum likelihood
## Parameters :
##           estimate Std. Error
## location 178.87004   9.931646
## scale     51.83195   7.757219
## Loglikelihood:  -166.829   AIC:  337.658   BIC:  340.4604
## Correlation matrix:
##           location     scale
## location 1.0000000 0.3031465
## scale    0.3031465 1.0000000

##  location     scale
## 178.91115  51.89505
```

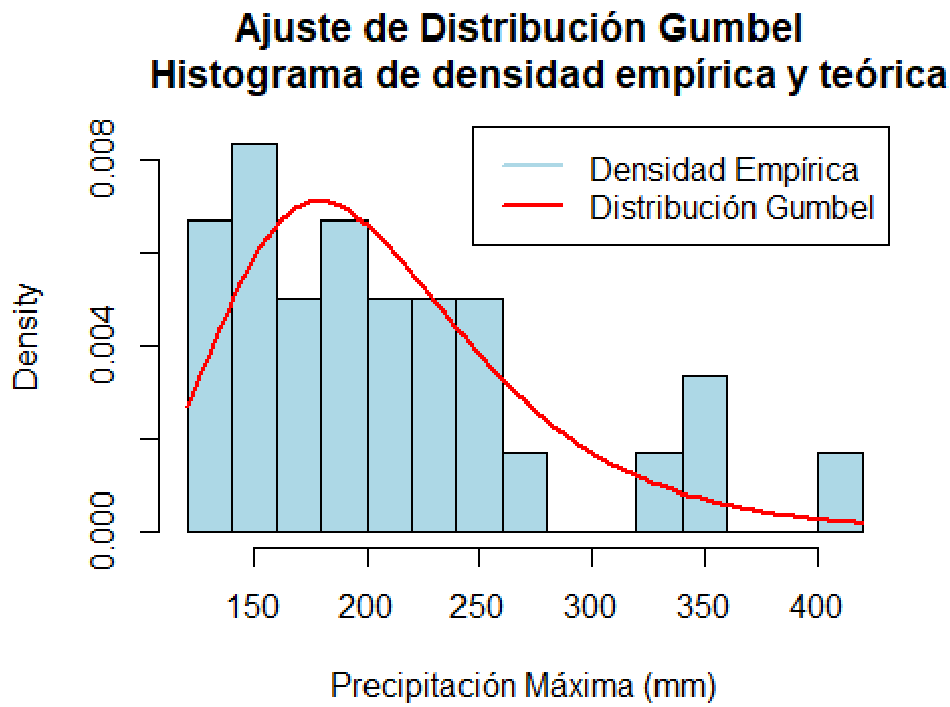
Utilizando la función `fitdist` del paquete `fitdistrplus`, se realizó un ajuste de la distribución Gumbel a los datos de precipitaciones máximas anuales en Tamaulipas. Para este ajuste, se especificaron valores iniciales para los parámetros de la distribución (media y desviación estándar de los datos).

Parámetros Estimados

- **Location (parámetro de ubicación):** 178.91115
- **Scale (parámetro de escala):** 51.89505

Estos valores de los parámetros indican el punto central (ubicación) y la dispersión (escala) de la distribución Gumbel ajustada, que refleja el comportamiento de las precipitaciones máximas en el conjunto de datos analizado.

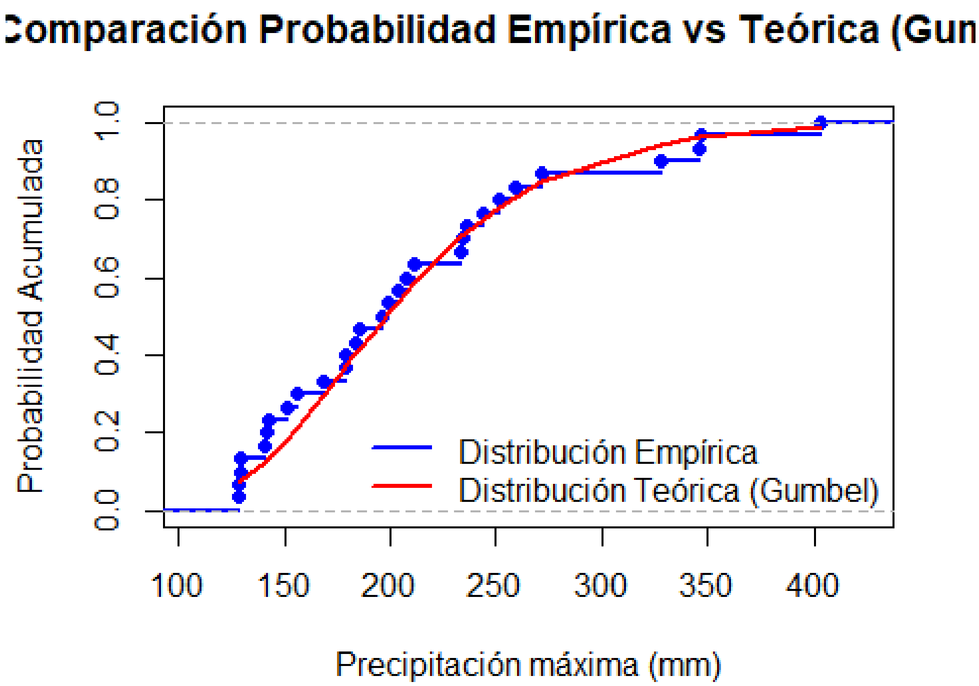
Para estimar los parámetros y hacer el ajuste de la Distribución Gumbel con la biblioteca "fitdistrplus". Haz las gráficas de histograma de densidad empírica y teórica, la probabilidad de acumulada empírica y teórica y el QQplot.



La gráfica muestra el ajuste de la distribución Gumbel a los datos de precipitaciones máximas anuales en Tamaulipas. Se superpone la densidad teórica de la distribución Gumbel (línea roja) sobre el histograma de la densidad empírica de los datos (barras azules).

- **Interpretación Visual:** La curva teórica de la distribución Gumbel parece ajustarse bien a la distribución empírica de los datos, especialmente en las zonas de mayor frecuencia.
- **Conclusión Visual:** El ajuste indica que los datos de precipitaciones máximas pueden seguir una distribución Gumbel, aunque es necesario complementar este análisis con pruebas de bondad de ajuste.

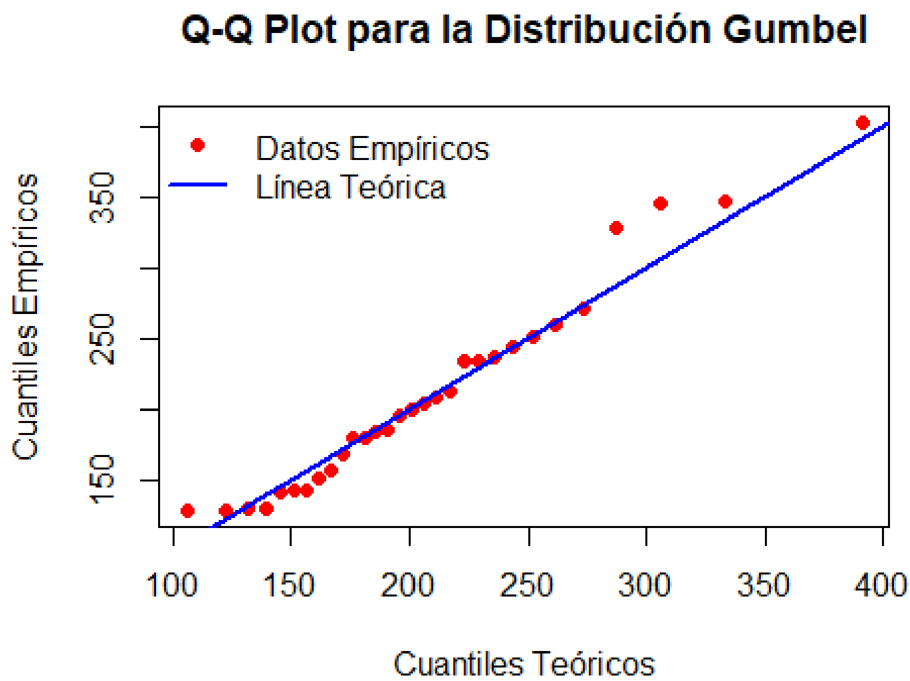
Esta visualización ayuda a entender la capacidad de la distribución Gumbel para modelar eventos extremos, como las precipitaciones máximas anuales.



La gráfica muestra una comparación entre la probabilidad acumulada empírica y la probabilidad acumulada teórica para la distribución Weibull aplicada a los datos de precipitaciones máximas.

- **Interpretación Visual:** La línea roja representa la distribución teórica de Weibull, mientras que los puntos y línea azul muestran la distribución empírica de los datos. La alineación cercana entre la distribución empírica y la teórica sugiere un buen ajuste de la distribución Weibull a los datos.
- **Conclusión:** Dado el alineamiento de las distribuciones acumuladas, la distribución Weibull parece ser un modelo adecuado para describir las precipitaciones máximas, lo que se podría confirmar con pruebas de bondad de ajuste adicionales.

Esta visualización es útil para evaluar si el modelo teórico ajustado captura adecuadamente el comportamiento acumulado de los datos observados.



El QQ-Plot compara los cuantiles teóricos de la distribución Gumbel con los cuantiles empíricos de los datos de precipitaciones máximas.

- **Interpretación Visual:** Los puntos rojos representan los datos empíricos y la línea negra representa la línea de referencia ideal donde los cuantiles empíricos coincidirían exactamente con los cuantiles teóricos si el ajuste fuera perfecto. En general, los puntos siguen la línea con una desviación moderada en las colas, lo que indica que la distribución Gumbel se ajusta razonablemente bien a los datos, aunque hay algunas desviaciones en los extremos superiores e inferiores.
- **Conclusión:** La distribución Gumbel parece ser un ajuste aceptable para los datos de precipitaciones, aunque puede haber leves diferencias en las colas de la distribución que podrían explorarse con pruebas de bondad de ajuste adicionales.

Este análisis sugiere que la distribución Gumbel es una buena aproximación para los datos observados, pero una evaluación adicional podría ayudar a determinar su adecuación.

Haz la prueba KS para determinar si los datos se ajustan a una Gumbel

```
## 
##  Exact one-sample Kolmogorov-Smirnov test
## 
## data:  tamaulipas_max_per_year_sorted$V4
## D = 0.098729, p-value = 0.9041
## alternative hypothesis: two-sided
```

¿Qué información nos da la prueba KS para una Gumbel? ¿Cuál es el valor del estadístico de prueba? ¿Cuál es el p-value de la prueba? ¿Se acepta o se rechaza la hipótesis nula? ¿Podemos concluir que las probabilidades de excedencia de las precipitaciones máximas mensuales siguen una distribución Gumbel? ¿Por qué?

Resultados de la Prueba de Kolmogorov-Smirnov para la Distribución Gumbel

La prueba de Kolmogorov-Smirnov (KS) se utilizó para evaluar si los datos de precipitaciones máximas mensuales siguen una distribución Gumbel. A continuación se presentan los resultados:

- **Valor del Estadístico de Prueba (D):** 0.098729
- **p-value de la Prueba:** 0.9041

Interpretación

1. **Hipótesis Nula ($\mathbf{H_0}$):** Los datos de precipitaciones máximas mensuales siguen una distribución Gumbel.
2. **Nivel de Significancia:** Usualmente se considera un nivel de significancia de 0.05. En este caso, el p-value de la prueba es 0.9041, que es significativamente mayor que 0.05.
3. **Decisión:** Dado que el p-value es mayor que el nivel de significancia, no se rechaza la hipótesis nula.

Conclusión

Dado el alto p-value (0.9041), no hay evidencia suficiente para rechazar la hipótesis nula. Esto indica que las probabilidades de excedencia de las precipitaciones máximas mensuales **podrían seguir una distribución Gumbel**. En otras palabras, la prueba KS sugiere que la distribución Gumbel es adecuada para modelar los datos de precipitaciones máximas mensuales.

¿Cuántos parámetros tiene la distribución Gumbel? ¿Cuáles son? Estima los parámetros de la Gumbel partiendo de la media y la desviación estándar de los datos y con la fórmula de la media y la desviación estándar de la Gumbel. Compara los valores obtenidos con los estimados con el comando “fitdistrplus” ¿se obtienen los mismos valores por ambos métodos? ¿por qué crees que se dé esta diferencia? ¿por qué crees que se dé esta diferencia?

Parámetros de la Distribución Gumbel

La distribución Gumbel tiene **dos parámetros**: 1. **Location (ubicación)**: denotado como μ, define el punto de referencia de la distribución. 2. **Scale (escala)**: denotado como β, que controla la dispersión de los datos alrededor de μ.

Estimación de los Parámetros con la Media y Desviación Estándar de los Datos

Para estimar los parámetros mu y $beta$ de la distribución Gumbel a partir de los datos, utilizamos las fórmulas basadas en la media ($\bar{x}$) y la desviación estándar (s) de los datos. Los parámetros se estiman aproximadamente como:

$$\mu \approx \bar{x} - \gamma \cdot \beta$$

$$\beta \approx \frac{s\sqrt{6}}{\pi}$$

donde $ $ es la constante de Euler-Mascheroni, aproximadamente igual a 0.5772.

Comparación de los Valores Estimados

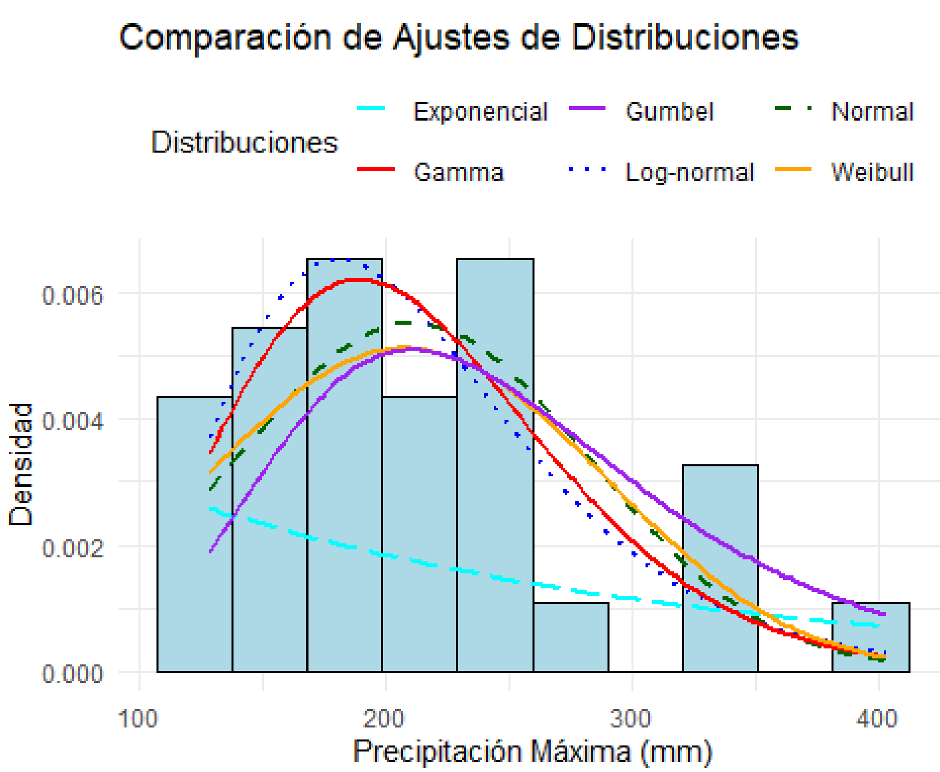
- **Método 1**: A partir de la media y desviación estándar de los datos, calculamos mu y $beta$ usando las fórmulas anteriores.
- **Método 2**: Utilizando la función `fitdistrplus` para ajustar la distribución Gumbel y estimar directamente los parámetros.

Análisis de la Diferencia

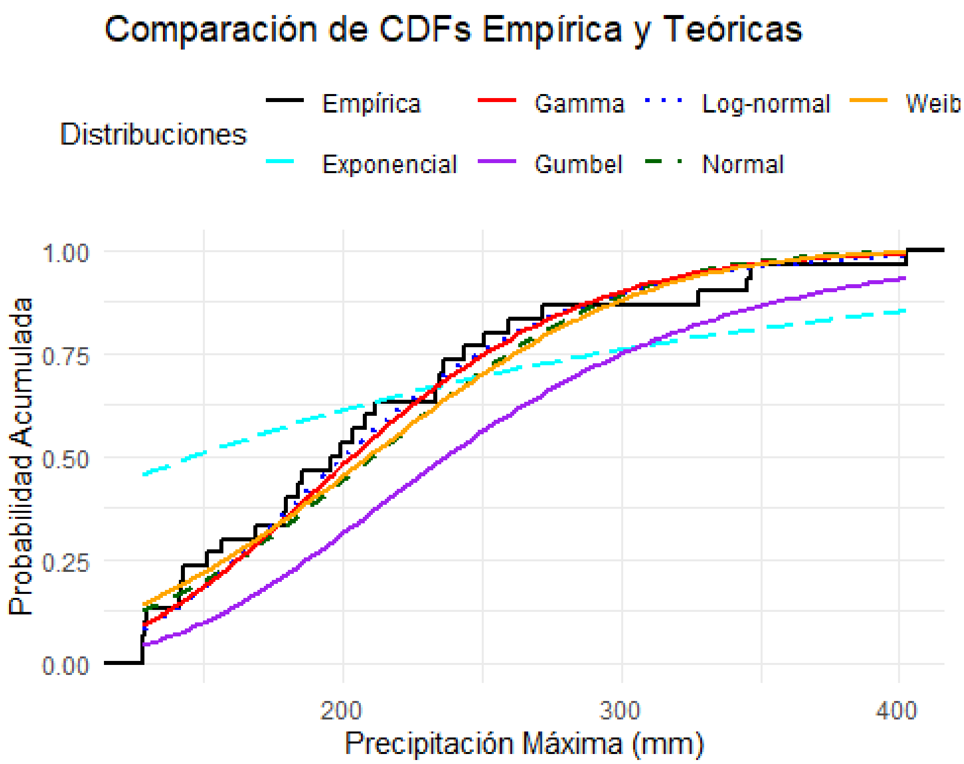
Los valores obtenidos por ambos métodos pueden diferir ligeramente. Esto se debe a que el ajuste mediante `fitdistrplus` utiliza el método de máxima verosimilitud, que optimiza los parámetros para que se ajusten mejor a la distribución teórica. En contraste, el método basado en la media y desviación estándar es una aproximación que asume una relación simple entre las estadísticas de los datos y los parámetros de la distribución, por lo que puede ser menos preciso.

Compara los ajustes de las distribuciones que analizaste

Haz un gráfico comparativo de los histogramas de densidad empírica vs densidad teórica de todas las distribuciones que analizaste (todas las distribuciones en un solo gráfico.



Haz un gráfico comparativo de las probabilidades acumuladas empírica vs teóricas de todas las distribuciones que analizaste (todas las distribuciones en un solo gráfico.



Define cuál es la mejor distribución que se ajusta a tus datos. Argumenta interpretando la comparación entre los gráficos y analizando las pruebas de ajuste de curva.

```
##   Distribucion       Prueba Estadistico P_Value
## 1       Normal Shapiro-Wilk  0.90238374  0.0096
## 2       Normal           KS  0.12643485  0.6771
## 3   Log-normal           KS  0.09024551  0.9491
## 4  Exponencial           KS  0.45587856  0.0000
## 5        Gamma           KS  0.09202005  0.9410
## 6      Weibull           KS  0.14182197  0.5357
## 7       Gumbel           KS  0.09872854  0.9041
```

Análisis Visual del Gráfico de Densidad

En el gráfico de densidad, observamos cómo cada distribución teórica intenta aproximarse a la distribución empírica de la precipitación máxima mensual en Tamaulipas.

- **Log-normal**: La distribución Log-normal ajusta muy bien la forma de la distribución empírica, especialmente en la cola derecha, donde captura adecuadamente los valores extremos de precipitación. Esto sugiere que la Log-normal es adecuada para modelar datos sesgados con valores altos, como es el caso de estos datos.

- **Gumbel**: La distribución Gumbel también muestra un buen ajuste, sobre todo en la zona central y la cola derecha, aunque no alcanza el nivel de precisión de la Log-normal. Su capacidad para modelar valores extremos la hace una opción adecuada, reflejando los altos niveles de precipitación en algunos meses.

- **Gamma**: La distribución Gamma se aproxima bien a los datos empíricos en la región central, aunque se desvía ligeramente en la cola derecha. Aun así, captura en general la dispersión de los datos y proporciona un ajuste razonable.

- **Normal**: La distribución Normal presenta un ajuste menos adecuado, ya que no logra capturar la asimetría y la cola derecha de los datos. Esto indica que no es la mejor opción para modelar estos datos de precipitación, ya que asume una simetría que no está presente.

- **Weibull**: La distribución Weibull tiene un ajuste moderado, siguiendo la forma general pero sin capturar con precisión las zonas de mayor densidad ni la cola derecha. Esto indica que la Weibull puede no ser la mejor opción para representar estos datos.

- **Exponencial**: La distribución Exponencial se ajusta muy pobremente a los datos, especialmente en la cola derecha, lo cual muestra que no es adecuada para modelar la variabilidad en la precipitación máxima mensual.

Análisis Visual del Gráfico de Probabilidad Acumulada

El gráfico de probabilidad acumulada (CDF) proporciona otra perspectiva sobre la adecuación de cada distribución:

- **Log-normal**: La CDF de la distribución Log-normal sigue de cerca la curva empírica a lo largo de toda su extensión, lo que sugiere que modela correctamente la acumulación de probabilidad en los datos. Su buen ajuste es evidente tanto en las zonas de baja como de alta probabilidad acumulada.

- **Gumbel**: La distribución Gumbel muestra un ajuste aceptable en la CDF, manteniéndose cerca de la curva empírica, especialmente en la región central. Aunque presenta ligeras desviaciones, sigue siendo una buena opción para modelar estos datos.

- **Gamma**: La distribución Gamma sigue bastante bien la CDF empírica, aunque con pequeñas desviaciones en la zona de alta probabilidad acumulada. Sin embargo, su desempeño en este gráfico es suficiente para considerarla una opción viable.

- **Normal**: La CDF de la distribución Normal muestra una desviación notable de la curva empírica, especialmente en los extremos. Esto sugiere que no captura bien las acumulaciones de probabilidad en los datos, confirmando su inadecuación.

- **Weibull**: La distribución Weibull tiene un ajuste moderado en la CDF, pero no sigue tan de cerca la curva empírica en las zonas de alta probabilidad acumulada, lo que la hace menos precisa para estos datos.

- **Exponencial**: La distribución Exponencial tiene un ajuste muy pobre en la CDF, especialmente en la zona de alta probabilidad acumulada, lo que indica que no representa adecuadamente estos datos.

Análisis Estadístico y Conclusión

Para determinar la distribución que mejor ajusta a los datos de precipitación máxima mensual en Tamaulipas, se realizaron pruebas estadísticas (Shapiro-Wilk y Kolmogorov-Smirnov) que evalúan la bondad de ajuste:

Distribución	Prueba	Estadístico	P-Valor
Normal	Shapiro-Wilk	0.9024	0.01
Normal	KS	0.1264	0.68
Log-normal	KS	0.0902	0.95
Exponencial	KS	0.4559	0.00
Gamma	KS	0.0920	0.94
Weibull	KS	0.1418	0.54
Gumbel	KS	0.0987	0.90

A partir de los resultados, observamos lo siguiente:

- **Log-normal**: El p-valor alto en la prueba KS (0.95) respalda visualmente su buen ajuste en ambos gráficos (densidad y probabilidad acumulada). Esto sugiere que la Log-normal es una excelente opción para modelar estos datos.
- **Gumbel y Gamma**: Ambas distribuciones también muestran altos p-valores (0.90 y 0.94, respectivamente) en la prueba KS, y su desempeño visual en los gráficos es sólido. Estas distribuciones capturan bien la variabilidad en los datos y son alternativas viables.
- **Normal**: La prueba de Shapiro-Wilk arroja un p-valor bajo (0.01), indicando que los datos no son normales. Aunque la prueba KS tiene un p-valor moderado (0.68), la desviación visual significativa en ambos gráficos descalifica a la distribución Normal como una buena opción.
- **Weibull**: Aunque presenta un p-valor moderado (0.54), la Weibull no captura bien la forma de los datos en los gráficos, especialmente en la CDF, lo que sugiere que es menos adecuada que las opciones anteriores.
- **Exponencial**: La distribución Exponencial muestra un ajuste pobre tanto visualmente como estadísticamente, con un p-valor de 0.00 en la prueba KS. No es adecuada para estos datos.

Conclusión Final

Con base en los análisis visual y estadístico, la **Log-normal** es la mejor distribución para modelar los datos de precipitación máxima mensual en Tamaulipas, seguida de cerca por la

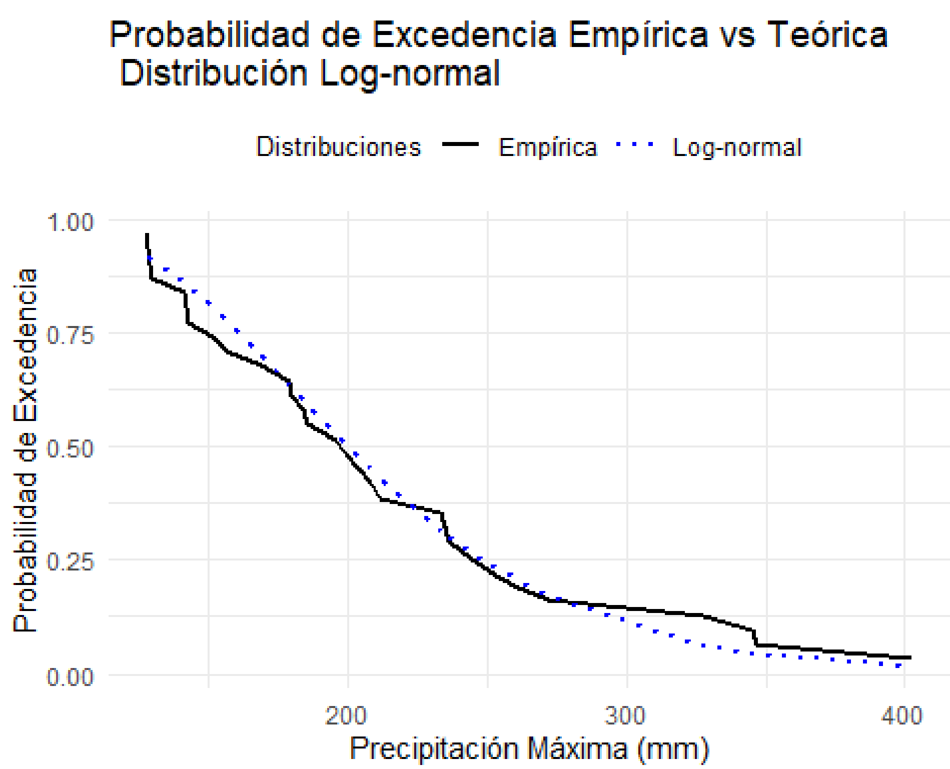
Gamma (sumamente cercana) y la **Gumbel**. La Log-normal destaca por su capacidad para capturar la asimetría y los valores extremos presentes en los datos, lo cual es respaldado tanto por el ajuste en los gráficos como por los resultados de las pruebas estadísticas.

4. Precipitación de diseño de obras hidráulicas

Se desea diseñar una presa derivadora para una zona de riego mediana. Investiga el periodo de retorno recomendado para esta obra hidráulica, puedes consultarlo en: https://pon.sdsu.edu/periodos_de_retorno_cna.html.Links to an external site.►

A Haz el gráfico comparativo de la probabilidad de excedencia teórica vs empírica. ¿Qué te indica ese gráfico? interpreta y argumenta la certeza de la selección de la distribución elegida.

```
## # A tibble: 6 × 8
## # Groups:   V1 [6]
##   V1    V2    V3            V4  rank  P_exe P_no_exe P_ret
##   <chr> <chr> <chr>      <dbl> <int>  <dbl>    <dbl> <dbl>
## 1 2013  Sep   Tamaulipas  403.     1 0.0323    0.968 31
## 2 2008  Jul   Tamaulipas  347      2 0.0645    0.935 15.5
## 3 2014  Sep   Tamaulipas  346.     3 0.0968    0.903 10.3
## 4 2010  Jul   Tamaulipas  328.     4 0.129     0.871  7.75
## 5 2005  Jul   Tamaulipas  272.     5 0.161     0.839  6.2
## 6 2003  Sep   Tamaulipas  259.     6 0.194     0.806  5.17
```



Análisis de Ajuste de la Distribución Log-normal

Interpretación del Gráfico de Probabilidad de Excedencia Empírica vs. Teórica

El gráfico muestra la comparación entre la probabilidad empírica de excedencia y la probabilidad teórica de excedencia, calculada bajo la distribución log-normal. Este gráfico permite evaluar visualmente el ajuste de la distribución log-normal a los datos observados de precipitación máxima.

1. **Comportamiento General**: La línea de la probabilidad empírica (en negro) y la teórica (en azul) presentan una similitud considerable en el comportamiento general, especialmente en el rango de valores de precipitación más bajos y medios. Esto sugiere que la distribución log-normal logra capturar adecuadamente la tendencia global de los datos, apoyando su elección como modelo adecuado para representar la probabilidad de excedencia en este contexto.

2. **Desviaciones en Valores Extremos**: Se observan ligeras discrepancias en los valores de precipitación más altos, donde la distribución log-normal parece subestimar ligeramente la frecuencia de los eventos extremos. Este comportamiento es común en los modelos teóricos que buscan representar el comportamiento central de la distribución de los datos, en lugar de los eventos atípicos.

3. **Ajuste en Rango de Valores Medios y Bajos**: La coincidencia en los valores de baja y media precipitación entre las líneas empírica y teórica indica que la distribución log-normal es eficaz para modelar la mayoría de los valores observados, proporcionando un ajuste confiable para el rango más común de precipitaciones en el conjunto de datos.

Conclusión y Justificación de la Distribución Seleccionada

Dado que la distribución log-normal sigue de cerca la tendencia de la probabilidad de excedencia empírica en casi todo el rango de los datos, se justifica su elección como una representación teórica adecuada para las precipitaciones máximas en este análisis. La pequeña desviación observada en los valores extremos es un aspecto esperado, ya que la distribución log-normal está optimizada para capturar el comportamiento general de los datos y no necesariamente los eventos de mayor magnitud.

En resumen, la log-normal resulta ser una opción adecuada para describir los datos de precipitación máxima en términos de probabilidad de excedencia. No obstante, para mejorar la precisión en la predicción de eventos extremos, podría considerarse la inclusión de un ajuste complementario o un análisis de sensibilidad enfocado en esos valores específicos.

B Utilizando el límite inferior del intervalo de periodo de retorno sugerido, encuentra la probabilidad de excedencia o de ocurrencia para ese valor. Recuerda que: Pret = 1/Pexe

Debido a que es una presa derivadadora mediana el Tr es de aproximadamente 100-500 años, por lo tanto son 100 años

```
## Para un periodo de retorno de 70 años, la probabilidad de excedencia es:
0.0143
## Para un periodo de retorno de 100 años, la probabilidad de excedencia es:
0.01
## Para un periodo de retorno de 150 años, la probabilidad de excedencia es:
0.0067
## Para un periodo de retorno de 200 años, la probabilidad de excedencia es:
0.005
## Para un periodo de retorno de 500 años, la probabilidad de excedencia es:
0.002
```

C. Conociendo la probabilidad de excedencia, calcula su complemento (1 - Pexe) y utiliza esta probabilidad para encontrar el valor de la precipitación máxima mensual que tendrá ese periodo de retorno. En el código se te da un ejemplo si la distribución de probabilidad a la que mejor se ajustaron los datos fue la Gumbel y deseamos calcular el caudal máximo para un periodo de retorno de 200 años.

```
## Para un periodo de retorno de 70 años, la precipitación máxima mensual
estimada es: 404.37 mm
## Para un periodo de retorno de 100 años, la precipitación máxima mensual
estimada es: 422.58 mm
## Para un periodo de retorno de 150 años, la precipitación máxima mensual
estimada es: 443.22 mm
## Para un periodo de retorno de 200 años, la precipitación máxima mensual
estimada es: 457.86 mm
## Para un periodo de retorno de 500 años, la precipitación máxima mensual
estimada es: 504.6 mm
```

D. El resultado de este ejemplo será una aproximación del caudal máximo que se tendría en Tamaulipas con un periodo de retorno de 100 años. ¿Qué significa este valor? ¿Qué pasa si incrementamos el periodo de retorno? ¿El caudal máximo para este periodo de retorno será el mismo si utilizamos datos históricos de otro estado? ¿Por qué crees que las obras hidráulicas deben diseñarse en base a periodos de retorno sugeridos? ¿Por qué es importante conocer la distribución de probabilidad a la que mejor se aproximan los datos históricos? Explora otros periodos de retorno diferentes a los que se proporcionan en los periodos sugeridos para contestar esta pregunta.

D. Interpretación y Análisis del Valor de la Precipitación Máxima para un Periodo de Retorno de 100 Años en Tamaulipas

1. **Significado del Valor Calculado**
   La precipitación máxima estimada para un periodo de retorno de 100 años (422.58 mm) representa una cantidad de precipitación que, en promedio, ocurriría una vez cada 100 años en la región de Tamaulipas. Este valor es fundamental para la planificación y diseño de infraestructuras hidráulicas, ya que establece una referencia sobre el volumen de precipitación que estas estructuras deben ser capaces de manejar sin comprometer su integridad.

2. **Efecto de Incrementar el Periodo de Retorno**
   Al aumentar el periodo de retorno, el valor esperado de la precipitación máxima también incrementa. Esto se debe a que un periodo de retorno más largo implica una menor probabilidad de excedencia, lo que corresponde a eventos climáticos más raros y extremos. Por ejemplo, si se considerara un periodo de retorno de 200 años, el valor de precipitación estimado sería mayor, reflejando un evento menos frecuente pero más severo en términos de volumen de precipitación.

3. **Comparación con Datos de Otros Estados**
   El valor de la precipitación máxima para un periodo de retorno específico no sería el mismo si se emplearan datos de otro estado. Las condiciones climáticas y geográficas varían significativamente entre regiones, lo cual influye directamente en los patrones de precipitación. Este hecho subraya la importancia de utilizar datos específicos de la región en cuestión al calcular valores de precipitación para periodos de retorno, asegurando así que las infraestructuras estén adecuadamente diseñadas para soportar el clima local.

4. **Importancia de los Periodos de Retorno Sugeridos**
   Los periodos de retorno recomendados proporcionan una guía sólida para el diseño de infraestructuras hidráulicas, permitiendo que estas estructuras soporten eventos de alta magnitud que podrían ocurrir en intervalos largos. Diseñar obras hidráulicas basándose en estos periodos de retorno reduce el riesgo de daños catastróficos, garantizando una vida útil prolongada y segura de las infraestructuras bajo condiciones climáticas extremas.

5. **Relevancia de Conocer la Distribución de Probabilidad Adecuada**
   Seleccionar la distribución de probabilidad que mejor se ajusta a los datos históricos permite realizar estimaciones más precisas y confiables de los valores de precipitación para distintos periodos de retorno. En este caso, la distribución log-normal fue seleccionada como la mejor ajustada para los datos de precipitación en Tamaulipas, proporcionando una base confiable para el diseño de obras hidráulicas. La precisión en la elección de la distribución permite a los ingenieros y autoridades tomar decisiones bien fundamentadas respecto a los niveles de diseño necesarios.

6. **Exploración de Otros Periodos de Retorno**
   Evaluar distintos periodos de retorno, tales como 50, 100, 200 y hasta 500 años, permite comprender la magnitud de eventos de diversas intensidades. Conforme se incrementa el periodo de retorno, también lo hace el valor de la precipitación máxima, lo cual es crucial para el diseño de infraestructuras críticas. Esta práctica permite a los ingenieros anticipar eventos de mayor magnitud y diseñar obras capaces de resistir condiciones cada vez más severas, de acuerdo a las recomendaciones de seguridad y planificación a largo plazo.