Actividad Integradora 1

Luis Maximiliano López Ramírez

2024-10-23

 # 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.

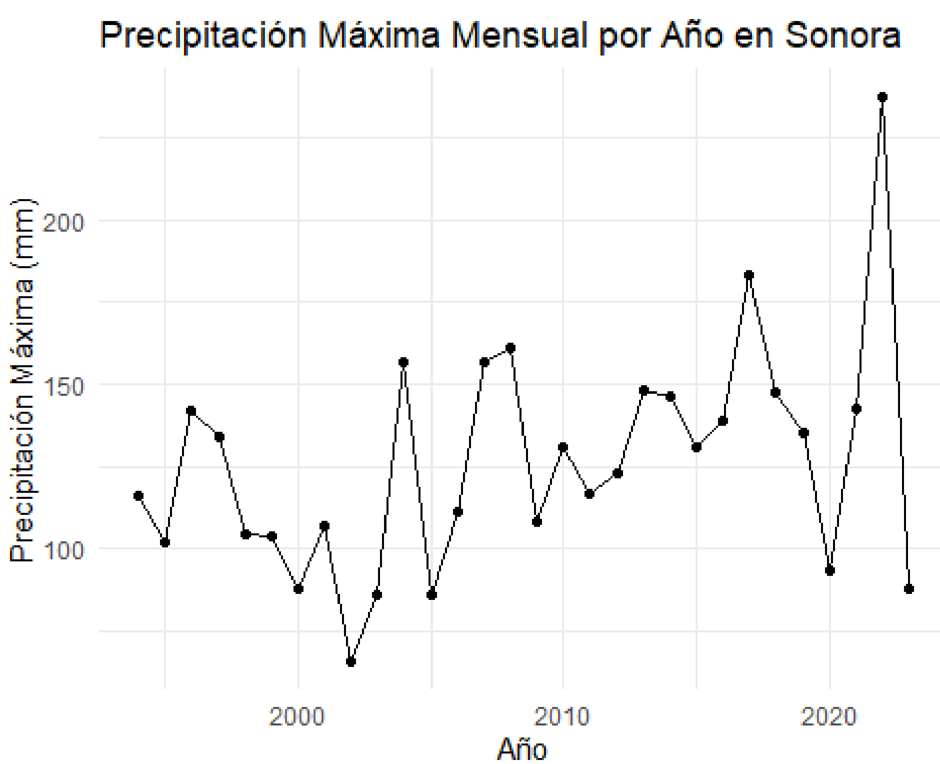
ESTADO SELECCIONADO: SONORA

```
datos <- read.table("precipitaciones_maximas_mensuales.txt", header =
TRUE, sep = "")
```

 # 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.

```
# Calcular la precipitación máxima mensual de cada año para el estado
seleccionado
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.3.2

##
## 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

# Filtrar datos para un estado específico, por ejemplo, "Sonora"
estado_seleccionado <- "Sonora"
datos_estado <- datos %>%
  filter(Estado == estado_seleccionado)

datos_agg <- datos_estado %>%
  group_by(Anio, Mes) %>%
  summarise(Lluvia_max = max(Lluvia, na.rm = TRUE)) %>%
  ungroup() %>%
```

```r
  group_by(Anio) %>%
  summarise(Lluvia_max_anual = max(Lluvia_max, na.rm = TRUE))

## `summarise()` has grouped output by 'Anio'. You can override using the
## `.groups` argument.

# Graficar los resultados
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.3.2

ggplot(datos_agg, aes(x = Anio, y = Lluvia_max_anual)) +
  geom_line() +
  geom_point() +
  labs(title = paste("Precipitación Máxima Mensual por Año en",
estado_seleccionado),
       x = "Año",
       y = "Precipitación Máxima (mm)") +
  theme_minimal()
```



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

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

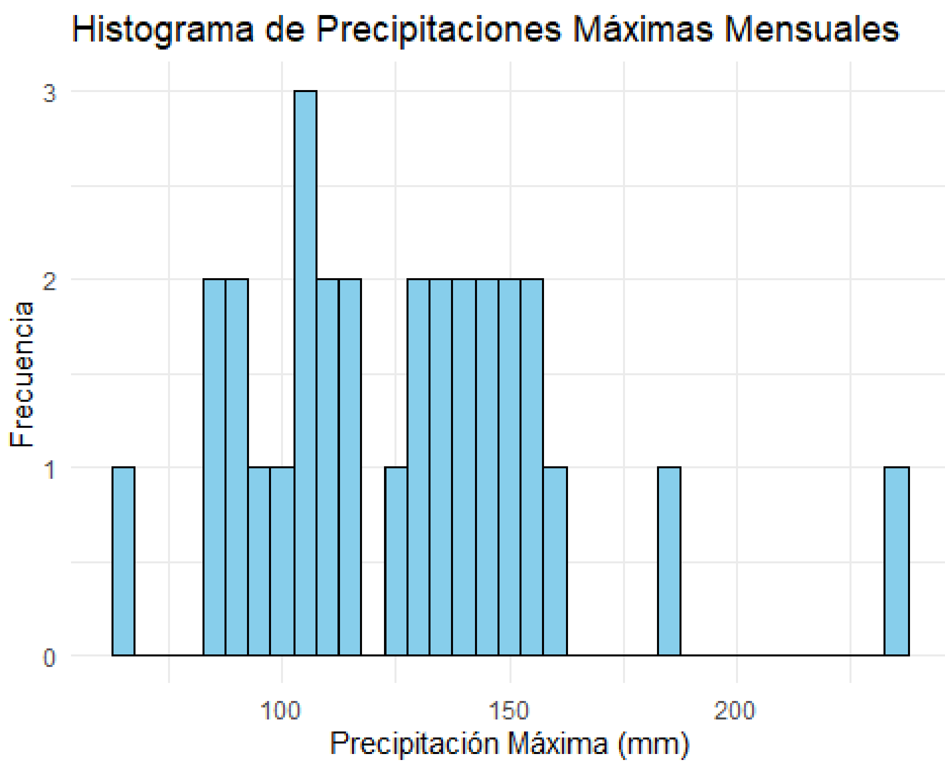
```r
# Calcular medidas de centralización y variación para las precipitaciones
máximas mensuales
```

```r
medidas_centralizacion_variacion <- datos_agg %>%
  summarise(
    Media = mean(Lluvia_max_anual, na.rm = TRUE),
    Mediana = median(Lluvia_max_anual, na.rm = TRUE),
    Desviacion_Estandar = sd(Lluvia_max_anual, na.rm = TRUE),
    Varianza = var(Lluvia_max_anual, na.rm = TRUE),
    Rango = max(Lluvia_max_anual, na.rm = TRUE) - min(Lluvia_max_anual,
na.rm = TRUE),
    Coeficiente_Variacion = (sd(Lluvia_max_anual, na.rm = TRUE) /
mean(Lluvia_max_anual, na.rm = TRUE)) * 100
  )

# Mostrar los resultados
print(medidas_centralizacion_variacion)
```

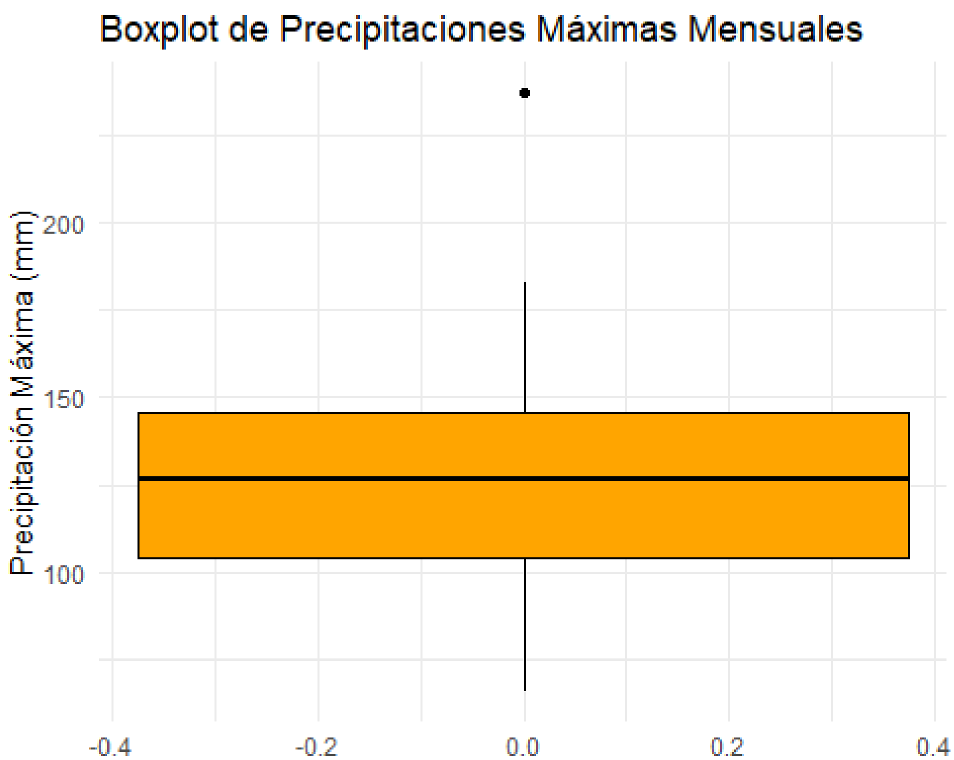
```
## # A tibble: 1 × 6
##   Media Mediana Desviacion_Estandar Varianza Rango
Coeficiente_Variacion
##   <dbl>   <dbl>               <dbl>    <dbl> <dbl>
<dbl>
## 1  126.    127.                34.5    1188.  172.
27.3
```

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

```r
# Histograma de las precipitaciones máximas mensuales
ggplot(datos_agg, aes(x = Lluvia_max_anual)) +
  geom_histogram(binwidth = 5, color = "black", fill = "skyblue") +
  labs(title = "Histograma de Precipitaciones Máximas Mensuales",
       x = "Precipitación Máxima (mm)",
       y = "Frecuencia") +
  theme_minimal()
```




```r
# Boxplot de las precipitaciones máximas mensuales
ggplot(datos_agg, aes(y = Lluvia_max_anual)) +
  geom_boxplot(color = "black", fill = "orange") +
  labs(title = "Boxplot de Precipitaciones Máximas Mensuales",
       y = "Precipitación Máxima (mm)") +
  theme_minimal()
```



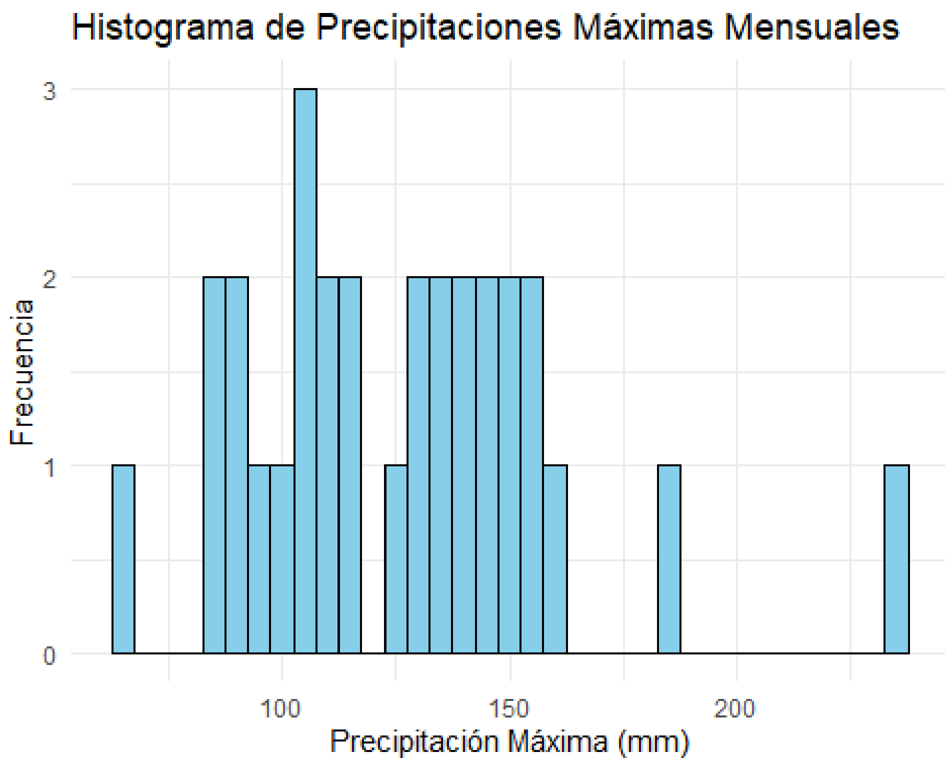
Describe el comportamiento de la distribución: centralización, sesgo, variación, ...

```
# Calcular medidas de centralización y variación
medidas_centralizacion_variacion <- datos_agg %>%
  summarise(
    Media = mean(Lluvia_max_anual, na.rm = TRUE),
    Mediana = median(Lluvia_max_anual, na.rm = TRUE),
    Desviacion_Estandar = sd(Lluvia_max_anual, na.rm = TRUE),
    Varianza = var(Lluvia_max_anual, na.rm = TRUE),
    Rango = max(Lluvia_max_anual, na.rm = TRUE) - min(Lluvia_max_anual,
na.rm = TRUE),
    Coeficiente_Variacion = (sd(Lluvia_max_anual, na.rm = TRUE) /
mean(Lluvia_max_anual, na.rm = TRUE)) * 100,
    Sesgo = mean((Lluvia_max_anual - mean(Lluvia_max_anual))^3) /
(sd(Lluvia_max_anual)^3)
  )

# Mostrar las medidas calculadas
print(medidas_centralizacion_variacion)
```

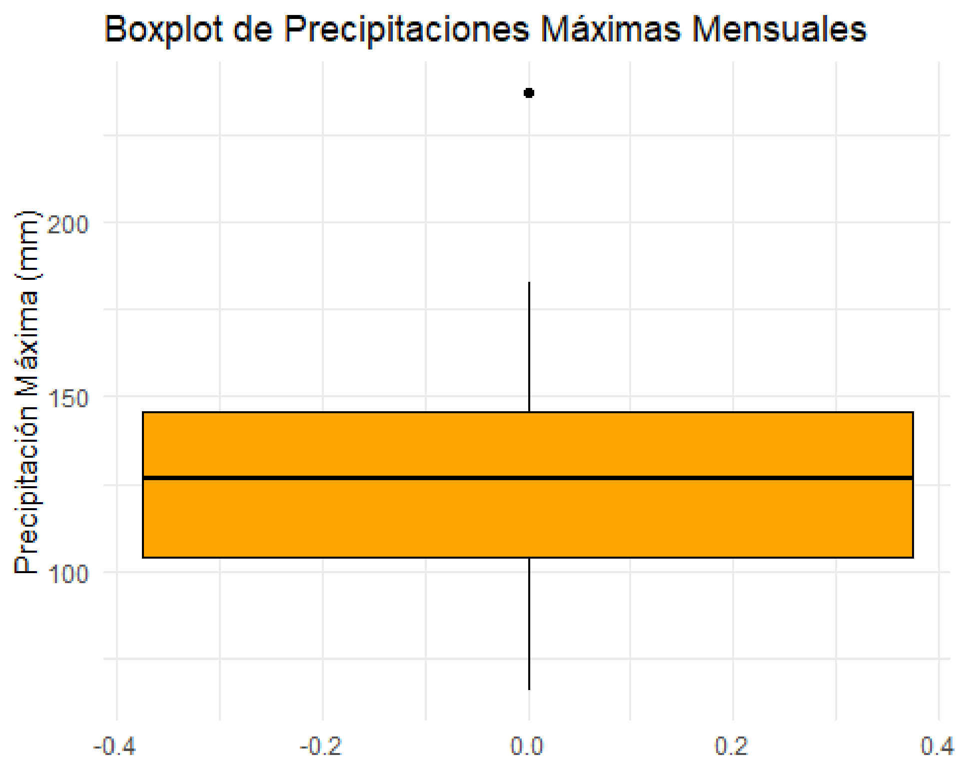
```
## # A tibble: 1 × 7
##   Media Mediana Desviacion_Estandar Varianza Rango
Coeficiente_Variacion Sesgo
##   <dbl>   <dbl>               <dbl>    <dbl> <dbl>
<dbl> <dbl>
```

```
## 1  126.    127.               34.5    1188.  172.
27.3 0.914
```

```r
# Histograma de las precipitaciones máximas mensuales
ggplot(datos_agg, aes(x = Lluvia_max_anual)) +
  geom_histogram(binwidth = 5, color = "black", fill = "skyblue") +
  labs(title = "Histograma de Precipitaciones Máximas Mensuales",
       x = "Precipitación Máxima (mm)",
       y = "Frecuencia") +
  theme_minimal()
```

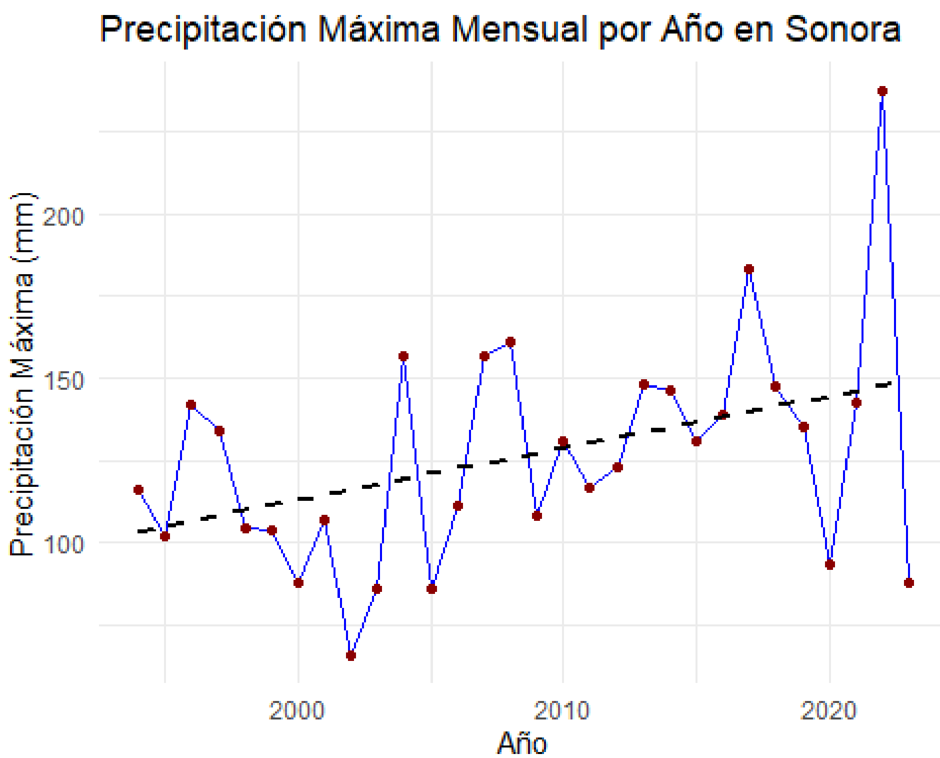



```r
# Boxplot de las precipitaciones máximas mensuales
ggplot(datos_agg, aes(y = Lluvia_max_anual)) +
  geom_boxplot(color = "black", fill = "orange") +
  labs(title = "Boxplot de Precipitaciones Máximas Mensuales",
       y = "Precipitación Máxima (mm)") +
  theme_minimal()
```



 # 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?

```
# Cargar librerías necesarias
library(dplyr)
library(ggplot2)

# Graficar los máximos mensuales anuales con una línea de tendencia
ggplot(datos_agg, aes(x = Anio, y = Lluvia_max_anual)) +
  geom_line(color = "blue") +
  geom_point(color = "darkred") +
  geom_smooth(method = "lm", se = FALSE, color = "black", linetype =
"dashed") +
  labs(title = "Precipitación Máxima Mensual por Año en Sonora",
       x = "Año",
       y = "Precipitación Máxima (mm)") +
  theme_minimal()

## `geom_smooth()` using formula = 'y ~ x'
```



1. ¿Observas alguna tendencia?

Tendencia general: La línea de tendencia muestra una ligera pendiente positiva, lo que sugiere que, en promedio, las precipitaciones máximas mensuales han tendido a aumentar ligeramente a lo largo del tiempo. Aunque no es un incremento muy pronunciado, sí indica una tendencia creciente en las precipitaciones máximas anuales.

2. ¿Puedes concluir que cada determinado número de años la cantidad de precipitación sube o baja?

Patrones cíclicos: No parece haber un patrón cíclico claro que indique que las precipitaciones suben o bajan en intervalos regulares. Hay fluctuaciones significativas año con año, pero no se observa una periodicidad definida. Esto sugiere que las variaciones anuales en las precipitaciones máximas pueden estar influenciadas por eventos climáticos irregulares o factores externos que no siguen un patrón predecible a largo plazo.

3. ¿Para qué nos sirve analizar este tipo de gráficas?

Detección de tendencias a largo plazo: Analizar la tendencia general de las precipitaciones máximas es útil para entender si la región está experimentando un aumento en eventos de lluvia intensa. Esto es importante para la planificación de recursos hídricos, agricultura y diseño de infraestructuras.

Identificación de variabilidad y eventos extremos: Las fluctuaciones y picos en la gráfica pueden indicar años con eventos de lluvia extrema, lo que puede ayudar en la gestión de riesgos de inundaciones y en la preparación para futuros eventos climáticos extremos. Planificación y adaptación al cambio climático: Si la tendencia creciente observada es consistente con datos de años adicionales, podría ser una señal de cambios en los patrones climáticos, lo que requeriría medidas de adaptación en la región para enfrentar estos cambios.

2. Análisis de Frecuencias Método Gráfico

El Método gráfico consiste en realizar dos gráficas en la que se muestren las precipitaciones máximas comparadas con la probabilidad de excedencia y con su periodo de retorno.

 # 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.

```
# Cargar la librería necesaria
library(dplyr)

# Crear una columna con los datos de lluvias máximas ordenados de mayor a
menor
datos_agg_ordenado <- datos_agg %>%
  arrange(desc(Lluvia_max_anual)) %>%
  mutate(Lluvia_Ordenada = Lluvia_max_anual)

# Mostrar el DataFrame resultante
print(datos_agg_ordenado)

## # A tibble: 30 × 3
##     Anio Lluvia_max_anual Lluvia_Ordenada
##    <int>            <dbl>           <dbl>
##  1  2022             237.            237.
##  2  2017             183.            183.
##  3  2008             161             161
##  4  2007             157             157
##  5  2004             157.            157.
##  6  2013             148.            148.
##  7  2018             148.            148.
##  8  2014             146.            146.
##  9  2021             143.            143.
## 10  1996             142.            142.
## # ℹ 20 more rows
```

 # 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

```
# Cargar la librería necesaria
library(dplyr)

# Crear una columna con el "rank" para cada precipitación máxima
```

```
datos_agg_ordenado <- datos_agg %>%
  arrange(desc(Lluvia_max_anual)) %>%
  mutate(m = row_number(), Lluvia_Ordenada = Lluvia_max_anual)

# Mostrar el DataFrame resultante
print(datos_agg_ordenado)
```

```
## # A tibble: 30 × 4
##     Anio Lluvia_max_anual     m Lluvia_Ordenada
##    <int>            <dbl> <int>           <dbl>
##  1  2022             237.     1            237.
##  2  2017             183.     2            183.
##  3  2008             161      3            161
##  4  2007             157      4            157
##  5  2004             157.     5            157.
##  6  2013             148.     6            148.
##  7  2018             148.     7            148.
##  8  2014             146.     8            146.
##  9  2021             143.     9            143.
## 10  1996             142.    10            142.
## # ℹ 20 more rows
```

 # 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:

```
# Cargar la librería necesaria
library(dplyr)

# Calcular el total de datos (N)
N <- nrow(datos_agg)

# Crear una columna con el "rank" y calcular la probabilidad de
excedencia
datos_agg_ordenado <- datos_agg %>%
  arrange(desc(Lluvia_max_anual)) %>%
  mutate(
    m = row_number(),  # Asignar el rango
    P_exe = m / (N + 1)  # Calcular la probabilidad de excedencia
  )

# Mostrar el DataFrame resultante
print(datos_agg_ordenado)
```

```
## # A tibble: 30 × 4
##     Anio Lluvia_max_anual     m  P_exe
##    <int>            <dbl> <int>  <dbl>
##  1  2022             237.     1 0.0323
##  2  2017             183.     2 0.0645
##  3  2008             161      3 0.0968
```

```
## 4 2007 157 4 0.129
## 5 2004 157. 5 0.161
## 6 2013 148. 6 0.194
## 7 2018 148. 7 0.226
## 8 2014 146. 8 0.258
## 9 2021 143. 9 0.290
## 10 1996 142. 10 0.323
## # ℹ 20 more rows
```

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

```
# Cargar la librería necesaria
library(dplyr)

# Calcular el total de datos (N)
N <- nrow(datos_agg)

# Crear columnas para el "rank", probabilidad de excedencia y
probabilidad de no excedencia
datos_agg_ordenado <- datos_agg %>%
  arrange(desc(Lluvia_max_anual)) %>%
  mutate(
    m = row_number(),  # Asignar el rango
    P_exe = m / (N + 1),  # Calcular la probabilidad de excedencia
    P_no_exe = 1 - P_exe  # Calcular la probabilidad de no excedencia
  )

# Mostrar el DataFrame resultante
print(datos_agg_ordenado)

## # A tibble: 30 × 5
##     Anio Lluvia_max_anual     m  P_exe P_no_exe
##    <int>            <dbl> <int>  <dbl>    <dbl>
##  1  2022             237.     1 0.0323    0.968
##  2  2017             183.     2 0.0645    0.935
##  3  2008             161      3 0.0968    0.903
##  4  2007             157      4 0.129     0.871
##  5  2004             157.     5 0.161     0.839
##  6  2013             148.     6 0.194     0.806
##  7  2018             148.     7 0.226     0.774
##  8  2014             146.     8 0.258     0.742
##  9  2021             143.     9 0.290     0.710
## 10  1996             142.    10 0.323     0.677
## # ℹ 20 more rows
```

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

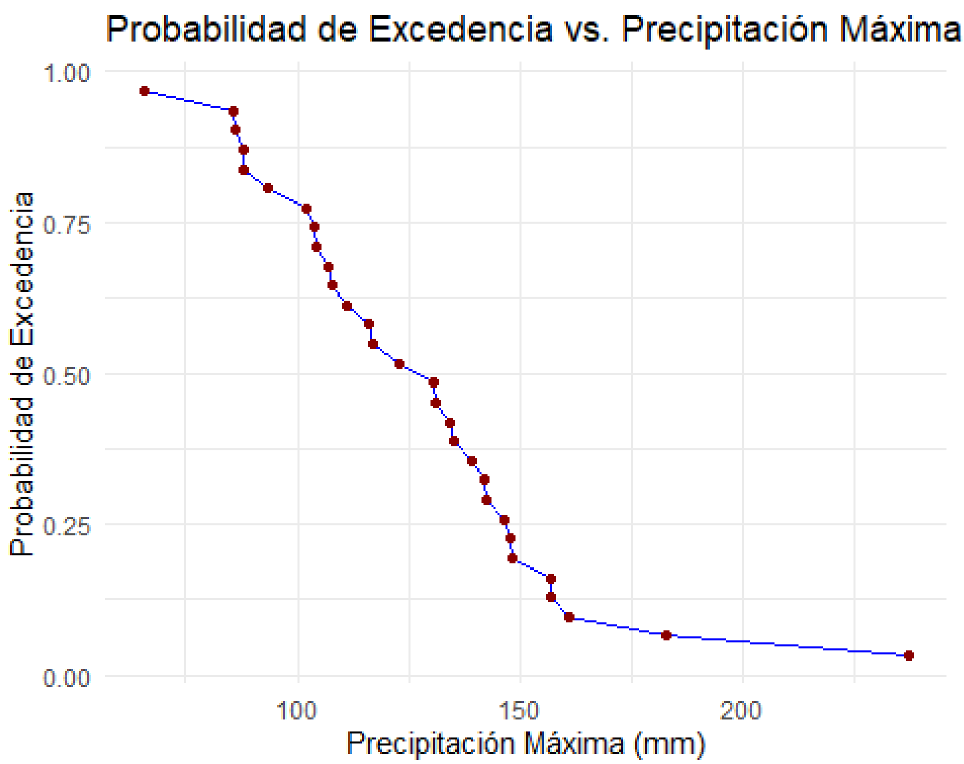
```
# Calcular el total de datos (N)
N <- nrow(datos_agg)
```

```r
# Crear columnas para el "rank", probabilidad de excedencia, probabilidad de no excedencia y período de retorno
datos_agg_ordenado <- datos_agg %>%
  arrange(desc(Lluvia_max_anual)) %>%
  mutate(
    m = row_number(),  # Asignar el rango
    P_exe = m / (N + 1),  # Calcular la probabilidad de excedencia
    P_no_exe = 1 - P_exe,  # Calcular la probabilidad de no excedencia
    P_ret = 1 / P_exe  # Calcular el período de retorno
  )

# Mostrar el DataFrame resultante
print(datos_agg_ordenado)
```

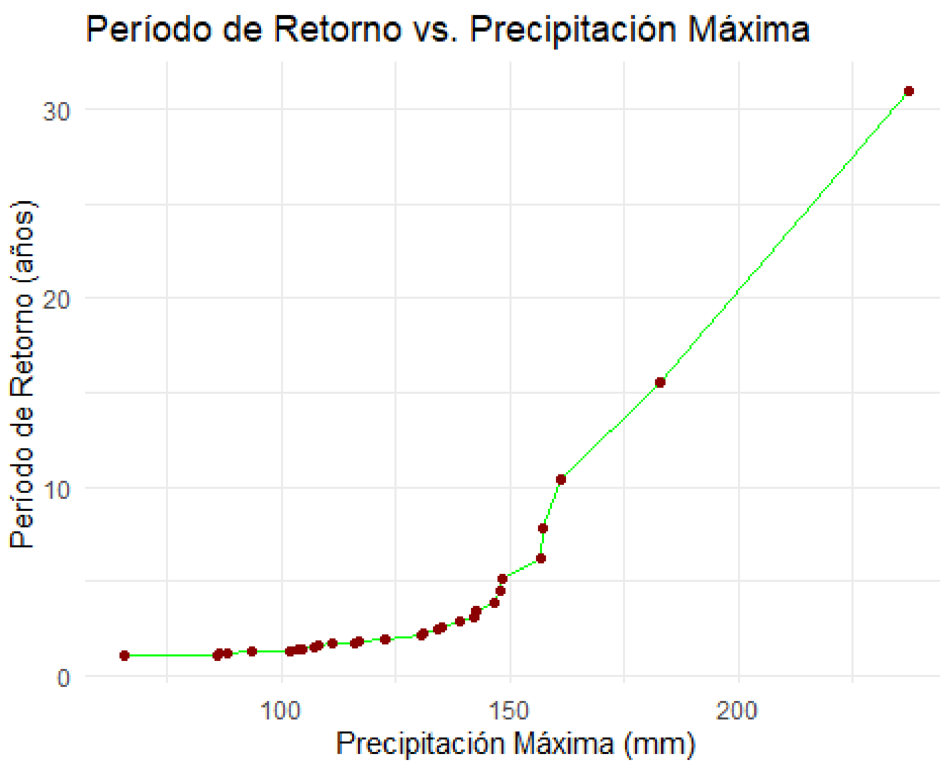
```
## # A tibble: 30 × 6
##     Anio Lluvia_max_anual     m  P_exe P_no_exe P_ret
##    <int>            <dbl> <int>  <dbl>    <dbl> <dbl>
##  1  2022             237.     1 0.0323    0.968 31
##  2  2017             183.     2 0.0645    0.935 15.5
##  3  2008             161      3 0.0968    0.903 10.3
##  4  2007             157      4 0.129     0.871  7.75
##  5  2004             157.     5 0.161     0.839  6.2
##  6  2013             148.     6 0.194     0.806  5.17
##  7  2018             148.     7 0.226     0.774  4.43
##  8  2014             146.     8 0.258     0.742  3.88
##  9  2021             143.     9 0.290     0.710  3.44
## 10  1996             142.    10 0.323     0.677  3.1
## # ℹ 20 more rows
```

 # 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? Puedes consultar el siguiente video de apoyo:

```r
# Cargar la librería necesaria
library(ggplot2)

# Gráfica de la probabilidad de excedencia vs precipitación máxima
ggplot(datos_agg_ordenado, aes(x = Lluvia_max_anual, y = P_exe)) +
  geom_line(color = "blue") +
  geom_point(color = "darkred") +
  labs(title = "Probabilidad de Excedencia vs. Precipitación Máxima",
       x = "Precipitación Máxima (mm)",
       y = "Probabilidad de Excedencia") +
  theme_minimal()
```




```r
# Gráfica del período de retorno vs precipitación máxima
ggplot(datos_agg_ordenado, aes(x = Lluvia_max_anual, y = P_ret)) +
  geom_line(color = "green") +
  geom_point(color = "darkred") +
  labs(title = "Período de Retorno vs. Precipitación Máxima",
       x = "Precipitación Máxima (mm)",
       y = "Período de Retorno (años)") +
  theme_minimal()
```



1. ¿Qué significa la probabilidad de excedencia?

La probabilidad de excedencia es la probabilidad de que una precipitación máxima anual sea igual o superior a un cierto valor en un año dado. En la primera gráfica, se observa que, a medida que aumenta la precipitación máxima, la probabilidad de que esa cantidad de lluvia sea excedida disminuye. Por ejemplo, las precipitaciones más bajas (cercanas a los 50 mm) tienen una probabilidad de excedencia cercana al 1, lo que indica que es muy probable que ocurran precipitaciones superiores. En cambio, precipitaciones muy altas (superiores a los 200 mm) tienen una probabilidad de excedencia baja, indicando que son eventos raros.

2. ¿Qué significa el período de retorno?

El período de retorno es el tiempo promedio (en años) que transcurre entre eventos de lluvia máxima iguales o superiores a un valor específico. Se calcula como el inverso de la probabilidad de excedencia. En la segunda gráfica, se observa que precipitaciones más altas tienen períodos de retorno más largos.

Por ejemplo, una precipitación máxima de unos 50 mm puede tener un período de retorno de menos de 2 años, mientras que precipitaciones superiores a 200 mm pueden tener un período de retorno de más de 30 años. Esto significa que, en promedio, una precipitación de 200 mm o más ocurre una vez cada 30 años.

3. ¿Por qué es importante en hidrología?

La probabilidad de excedencia y el período de retorno son fundamentales en hidrología para el diseño de infraestructuras críticas como presas, sistemas de drenaje y puentes. Estas estructuras deben diseñarse para soportar eventos de lluvia extrema que podrían ocurrir rara vez pero que tienen un impacto significativo. El período de retorno permite a los ingenieros evaluar el riesgo y la frecuencia de eventos extremos y diseñar obras que puedan manejar esos eventos sin fallar.

4. ¿Qué valores son deseables en la probabilidad de excedencia para una precipitación de diseño de una obra?

Para el diseño de infraestructuras, se suelen buscar bajas probabilidades de excedencia (0.01 a 0.05, dependiendo de la criticidad de la obra). Esto garantiza que la infraestructura pueda resistir eventos de lluvia que ocurren rara vez, pero que tienen el potencial de causar daños significativos si no se gestionan adecuadamente. Por ejemplo, una presa o sistema de drenaje podría estar diseñado para un evento con un período de retorno de 100 años (probabilidad de excedencia de 0.01), lo que asegura que la obra esté preparada para eventos extremos, aunque poco frecuentes.

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.

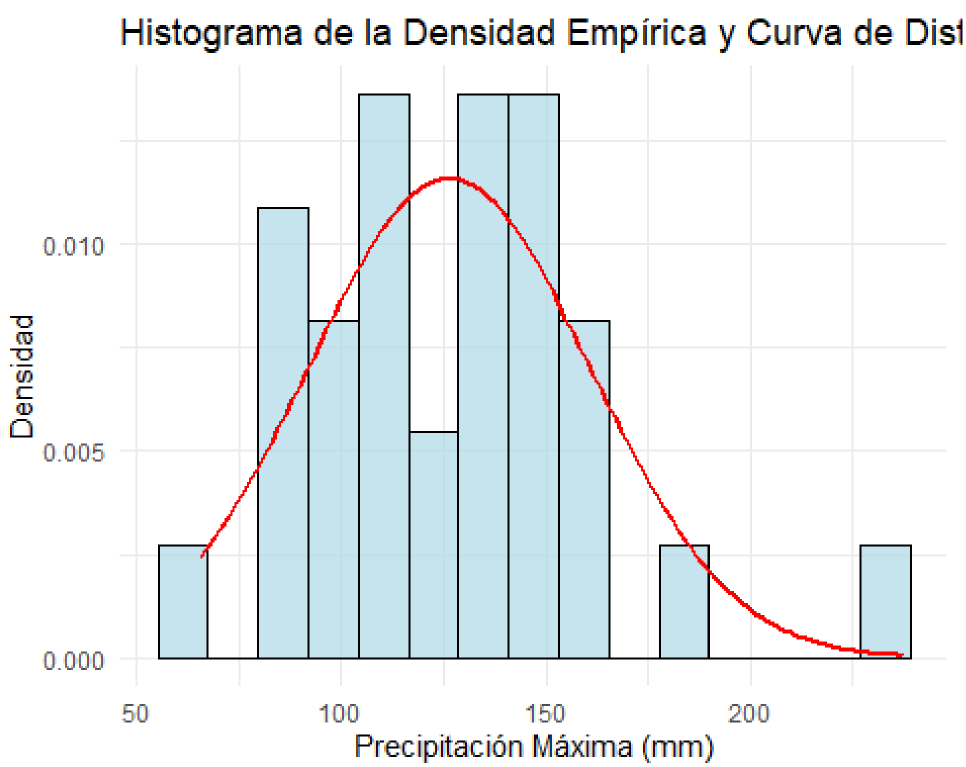
Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución normal que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. De manera visual, ¿te parece que los datos se ajustan bien a una distribución normal? Explica. ¿Cuántos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

```
# Cargar la librería necesaria
library(ggplot2)

# Calcular la media y desviación estándar de las precipitaciones máximas
media <- mean(datos_agg_ordenado$Lluvia_max_anual, na.rm = TRUE)
desviacion_estandar <- sd(datos_agg_ordenado$Lluvia_max_anual, na.rm =
TRUE)

# Crear el histograma de la densidad empírica y sobreponer la curva de
densidad normal
ggplot(datos_agg_ordenado, aes(x = Lluvia_max_anual)) +
  geom_histogram(aes(y = ..density..), bins = 15, color = "black", fill =
"lightblue", alpha = 0.7) +
```

```
  stat_function(fun = dnorm, args = list(mean = media, sd =
desviacion_estandar),
                color = "red", size = 1) +
  labs(title = "Histograma de la Densidad Empírica y Curva de
Distribución Normal",
       x = "Precipitación Máxima (mm)",
       y = "Densidad") +
  theme_minimal()
```

```
## 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.
```



¿Te parece que los datos se ajustan bien a una distribución normal?

Ajuste Parcial: Los datos muestran un ajuste parcial a la distribución normal. En el centro, la distribución parece similar a una curva normal, pero las colas y algunos picos indican que hay desviaciones significativas, como valores extremos más frecuentes de lo que se esperaría bajo una distribución normal pura.

No es completamente normal: Debido a las colas más pesadas y la posible presencia de picos adicionales, podemos concluir que los datos no siguen estrictamente una distribución normal. Esto puede deberse a la naturaleza de los eventos de precipitación, que tienden a tener eventos extremos que no son comunes en una distribución normal.

¿Cuántos parámetros tiene la distribución Normal? ¿Cuáles son?

La distribución normal tiene dos parámetros:

Media (μ): Representa el valor promedio o centro de la distribución.

Desviación estándar (σ): Mide la dispersión o el ancho de la distribución. Indica qué tan dispersos están los valores alrededor de la media.

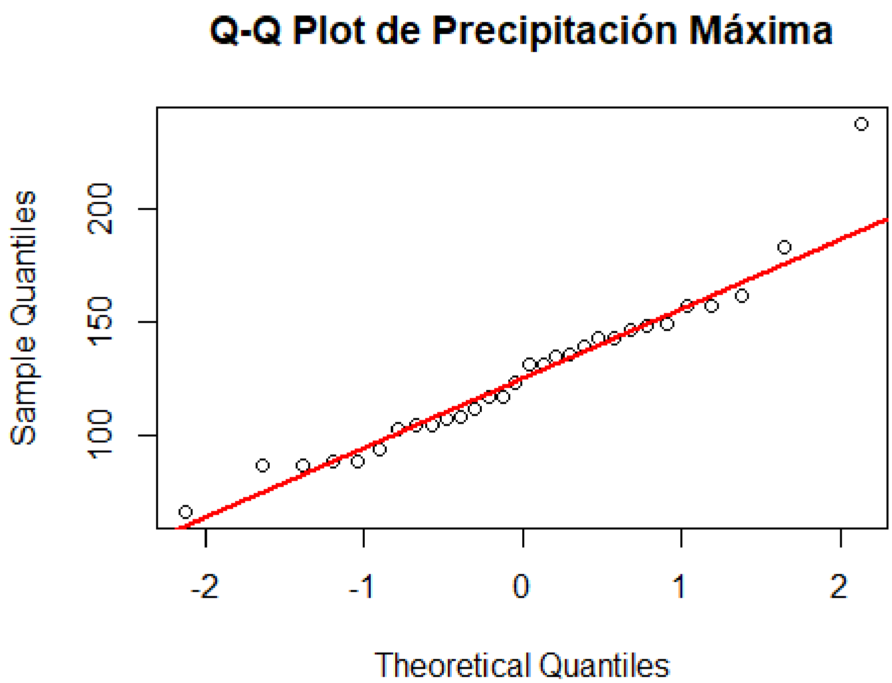
¿Por qué los parámetros se calculan de la forma en cómo se hace en el código?

Media (μ): Se calcula como la suma de todos los valores dividida por el número total de observaciones. Esto ofrece el valor central alrededor del cual se agrupan los datos.

Desviación estándar (σ): Mide la dispersión de los datos respecto a la media. Calcularla de esta forma permite entender qué tan dispersos o concentrados están los valores en la distribución. En una distribución normal, aproximadamente el 68% de los valores se encuentra dentro de una desviación estándar de la media, el 95% dentro de dos desviaciones, y el 99.7% dentro de tres.

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

```r
# Cargar la librería necesaria
library(ggplot2)

# Crear la Q-Q plot para evaluar la normalidad
qqnorm(datos_agg_ordenado$Lluvia_max_anual, main = "Q-Q Plot de 
Precipitación Máxima")
qqline(datos_agg_ordenado$Lluvia_max_anual, col = "red", lwd = 2)
```



Evaluación de la Q-Q Plot:

Línea recta en el centro:

En la parte central de la gráfica, los puntos se alinean bastante bien con la línea roja, lo que indica que los datos en el rango intermedio de valores siguen una distribución similar a la normal. Esto sugiere que, para valores cercanos a la media, la distribución empírica de los datos es aproximadamente normal.

Desviaciones en las colas:

Sin embargo, en las colas de la distribución (valores más bajos y más altos), los puntos se desvían significativamente de la línea recta. Esto indica que los datos tienen comportamientos que no se ajustan bien a una distribución normal en las colas. En particular:

- Cola inferior: Los valores más bajos (extremos negativos) se desvían hacia abajo, lo que sugiere una dispersión mayor de lo que la distribución normal predice.
- Cola superior: Los valores más altos (extremos positivos) se desvían hacia arriba, indicando la presencia de valores más extremos (colas más gruesas o sesgo positivo) de lo que una distribución normal típica esperaría.

Conclusión:

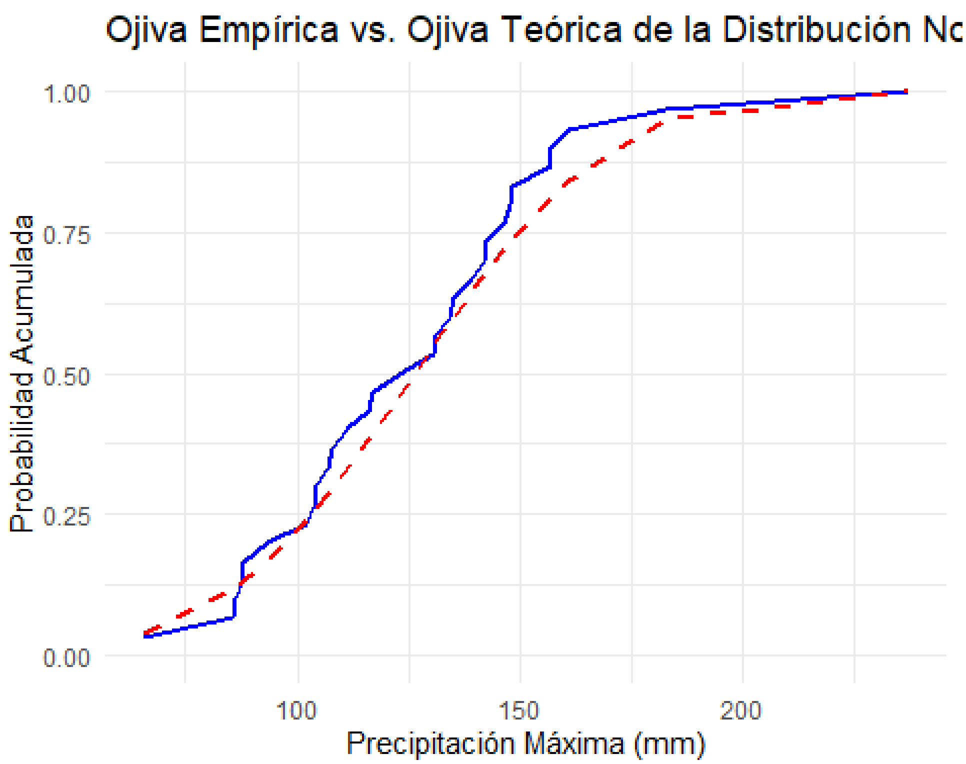
No completamente normal: Aunque los datos siguen una tendencia aproximadamente normal en el rango medio, las desviaciones en las colas sugieren que los datos no se ajustan perfectamente a una distribución normal. Las colas gruesas (más valores extremos de lo esperado) pueden indicar kurtosis alta o sesgo positivo, lo que significa que hay eventos extremos que son más frecuentes de lo que se esperaría bajo una normalidad estricta.

Importancia: Esta desviación puede ser significativa si se usan estos datos para modelar o hacer inferencias bajo la suposición de normalidad. Puede ser recomendable considerar transformaciones de los datos o usar modelos que no asuman normalidad estricta si estas desviaciones son críticas para el análisis.

 # 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?

```
# Cargar las librerías necesarias
library(ggplot2)

# Calcular la media y desviación estándar de los datos empíricos
media <- mean(datos_agg_ordenado$Lluvia_max_anual, na.rm = TRUE)
desviacion_estandar <- sd(datos_agg_ordenado$Lluvia_max_anual, na.rm =
TRUE)

# Crear la ojiva empírica
datos_agg_ordenado <- datos_agg_ordenado %>%
  arrange(Lluvia_max_anual) %>%
  mutate(empirical_cdf = ecdf(Lluvia_max_anual)(Lluvia_max_anual))

# Crear la distribución acumulada teórica
datos_agg_ordenado <- datos_agg_ordenado %>%
  mutate(teorica_cdf = pnorm(Lluvia_max_anual, mean = media, sd =
desviacion_estandar))

# Graficar la comparación de las distribuciones acumuladas
ggplot(datos_agg_ordenado, aes(x = Lluvia_max_anual)) +
  geom_line(aes(y = empirical_cdf), color = "blue", size = 1, linetype =
"solid") +
  geom_line(aes(y = teorica_cdf), color = "red", size = 1, linetype =
"dashed") +
  labs(title = "Ojiva Empírica vs. Ojiva Teórica de la Distribución
Normal",
       x = "Precipitación Máxima (mm)",
       y = "Probabilidad Acumulada") +
  theme_minimal() +
  scale_y_continuous(limits = c(0, 1)) +
  theme(legend.position = "bottom") +
```

```r
  guides(color = guide_legend(override.aes = list(linetype = c("solid",
"dashed")))) +
  scale_color_manual(values = c("Empírica" = "blue", "Teórica" = "red"))
```



Datos Empíricos vs. Datos Teóricos:

Datos Empíricos: Son los datos reales observados. En este caso, las precipitaciones máximas anuales recolectadas de tu dataset.

Datos Teóricos: Son los valores generados por una distribución que se ajusta a ciertos parámetros (media y desviación estándar) calculados a partir de los datos empíricos. En este caso, la distribución normal teórica que se esperaría si los datos siguieran perfectamente una distribución normal.

Comparación de las Distribuciones de Probabilidad Acumuladas:

Similitudes en el Centro:

En la parte central de la gráfica, las dos curvas (empírica y teórica) se parecen bastante y están relativamente alineadas. Esto sugiere que, para valores intermedios de precipitación máxima, los datos empíricos siguen un comportamiento cercano al que se esperaría bajo una distribución normal.

Desviaciones en las Colas:

Sin embargo, se notan desviaciones en las colas, tanto en la parte baja como en la alta. La ojiva empírica (azul) muestra más variabilidad y no sigue perfectamente la línea teórica (roja). Esto indica que hay más valores extremos (especialmente en las colas superior e inferior) de lo que una distribución normal esperaría.

Las desviaciones en la cola superior (valores altos de precipitación) sugieren la presencia de eventos más extremos (colas gruesas) de lo que se modela con una distribución normal.

Conclusión:

No es un ajuste perfecto: Aunque las distribuciones empírica y teórica tienen cierta similitud en el rango central, las desviaciones en las colas indican que los datos empíricos no siguen estrictamente una distribución normal. Esto concuerda con las observaciones previas de que hay eventos extremos que no se ajustan bien a la distribución normal esperada.

Impacto en el análisis: Si se pretende utilizar modelos o análisis basados en la suposición de normalidad, estas desviaciones en las colas pueden afectar la precisión de los resultados. Es importante tener en cuenta estas diferencias, especialmente si los eventos extremos (como lluvias muy altas) son de interés crítico.

 # 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: Los datos provienen de una distribución normal H1: Los datos no provienen de una distribución normal

Información sobre las Pruebas de Bondad de Ajuste:

Prueba de Shapiro-Wilk:

Esta prueba evalúa si una muestra proviene de una distribución normal. Es particularmente adecuada para muestras pequeñas o medianas. Hipótesis nula (H_0): Los datos provienen de una distribución normal.

Hipótesis alternativa (H_1): Los datos no provienen de una distribución normal.

Interpretación: Si el p-value es bajo (generalmente menor a 0.05), se rechaza la hipótesis nula, indicando que los datos no siguen una distribución normal.

Prueba de Kolmogorov-Smirnov (KS):

Compara la distribución acumulada de los datos empíricos con la de una distribución teórica específica (en este caso, la distribución normal).

Hipótesis nula (H_0): Los datos siguen la distribución normal especificada.

Hipótesis alternativa (H_1): Los datos no siguen la distribución normal especificada.

Interpretación: Un p-value bajo sugiere que se debe rechazar la hipótesis nula, indicando que los datos no se ajustan a la distribución normal.

```
# Prueba de Shapiro-Wilk
shapiro_test <- shapiro.test(datos_agg_ordenado$Lluvia_max_anual)

# Prueba de Kolmogorov-Smirnov (KS) contra la distribución normal teórica
media <- mean(datos_agg_ordenado$Lluvia_max_anual, na.rm = TRUE)
desviacion_estandar <- sd(datos_agg_ordenado$Lluvia_max_anual, na.rm =
TRUE)
ks_test <- ks.test(datos_agg_ordenado$Lluvia_max_anual, "pnorm", mean =
media, sd = desviacion_estandar)

# Mostrar resultados
shapiro_test

##
##  Shapiro-Wilk normality test
##
## data:  datos_agg_ordenado$Lluvia_max_anual
## W = 0.93724, p-value = 0.07668

ks_test

##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  datos_agg_ordenado$Lluvia_max_anual
## D = 0.095837, p-value = 0.9212
## alternative hypothesis: two-sided
```

1. Prueba de Shapiro-Wilk

Estadístico (W): 0.93724 p-value: 0.07668

Interpretación: El p-value es mayor que 0.05, lo que indica que no se rechaza la hipótesis nula. Esto sugiere que no hay suficiente evidencia para concluir que los datos no son normales. En otras palabras, los datos podrían seguir una distribución normal.

2. Prueba de Kolmogorov-Smirnov (KS)

Estadístico (D): 0.095837 p-value: 0.9212 Interpretación: El p-value es muy alto (0.9212), lo que indica que no se rechaza la hipótesis nula. Esto sugiere que los datos se ajustan bien a la distribución normal teórica basada en la media y desviación estándar calculadas de los datos. No hay evidencia suficiente para concluir que los datos no son normales.

Conclusiones:

¿Se aceptan o se rechazan las hipótesis nulas?

Prueba de Shapiro-Wilk: No se rechaza la hipótesis nula de normalidad (p-value > 0.05). Esto significa que, de acuerdo con esta prueba, los datos pueden ser considerados normales.

Prueba de Kolmogorov-Smirnov (KS): No se rechaza la hipótesis nula de que los datos provienen de una distribución normal (p-value > 0.05). Esta prueba también sugiere que los datos son consistentes con una distribución normal.

¿Podemos concluir que los datos de las precipitaciones máximas mensuales son normales?

Sí, pero con precaución: Ambas pruebas no encontraron evidencia suficiente para rechazar la normalidad, lo que indica que los datos de las precipitaciones máximas mensuales podrían seguir una distribución normal.

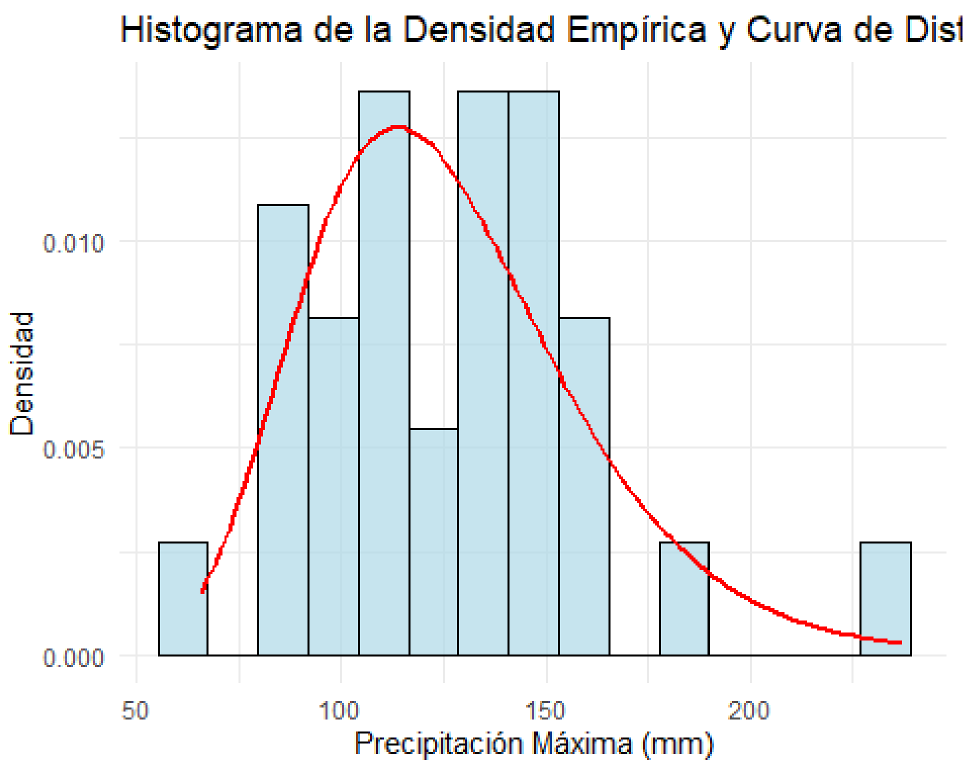
Sin embargo, es importante notar que las pruebas de normalidad tienden a ser menos efectivas para detectar desviaciones sutiles en distribuciones que son grandes. Además, las gráficas anteriores (histograma, Q-Q plot y ojiva) mostraron algunas desviaciones en las colas, lo que sugiere que, aunque los datos pueden ser aproximadamente normales, existen algunas variaciones que podrían no ajustarse perfectamente a una distribución normal estricta.

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.

```r
# Cargar la librería necesaria
library(ggplot2)

# Calcular los parámetros de la distribución Log-normal
# Se utiliza la transformación logarítmica natural para calcular la media
y desviación estándar
log_media <- mean(log(datos_agg_ordenado$Lluvia_max_anual), na.rm = TRUE)
log_sd <- sd(log(datos_agg_ordenado$Lluvia_max_anual), na.rm = TRUE)

# Crear el histograma de la densidad empírica y sobreponer la curva de
densidad log-normal
ggplot(datos_agg_ordenado, aes(x = Lluvia_max_anual)) +
  geom_histogram(aes(y = ..density..), bins = 15, color = "black", fill =
"lightblue", alpha = 0.7) +
  stat_function(fun = dlnorm, args = list(meanlog = log_media, sdlog =
log_sd),
                color = "red", size = 1) +
  labs(title = "Histograma de la Densidad Empírica y Curva de
Distribución Log-normal",
```

```
       x = "Precipitación Máxima (mm)",
       y = "Densidad") +
  theme_minimal()
```



Ajuste de los Datos a la Distribución Log-normal:

Ajuste en el Centro:

La curva log-normal (en rojo) parece ajustarse bastante bien al centro del histograma, indicando que, para los valores intermedios de precipitación máxima, los datos tienen un comportamiento similar al que se esperaría bajo una distribución log-normal.

Esto sugiere que la distribución log-normal puede capturar bien la forma general de los datos en la región central.

Desviaciones en las Colas:

Se notan desviaciones en las colas, particularmente en la cola derecha (valores altos de precipitación). La curva log-normal no captura completamente los picos más extremos del histograma.

En la cola izquierda (valores bajos), también hay una pequeña discrepancia, aunque menos pronunciada que en la cola derecha. La curva se ajusta bastante bien a los primeros valores, pero no es perfecta.

Forma General:

En general, la forma de campana asimétrica de la distribución log-normal parece ajustarse mejor que una distribución normal simétrica (basado en gráficos anteriores). Sin embargo, no es un ajuste perfecto, ya que los datos muestran más variabilidad y dispersión en las colas de lo que la curva log-normal modela.

Conclusión:

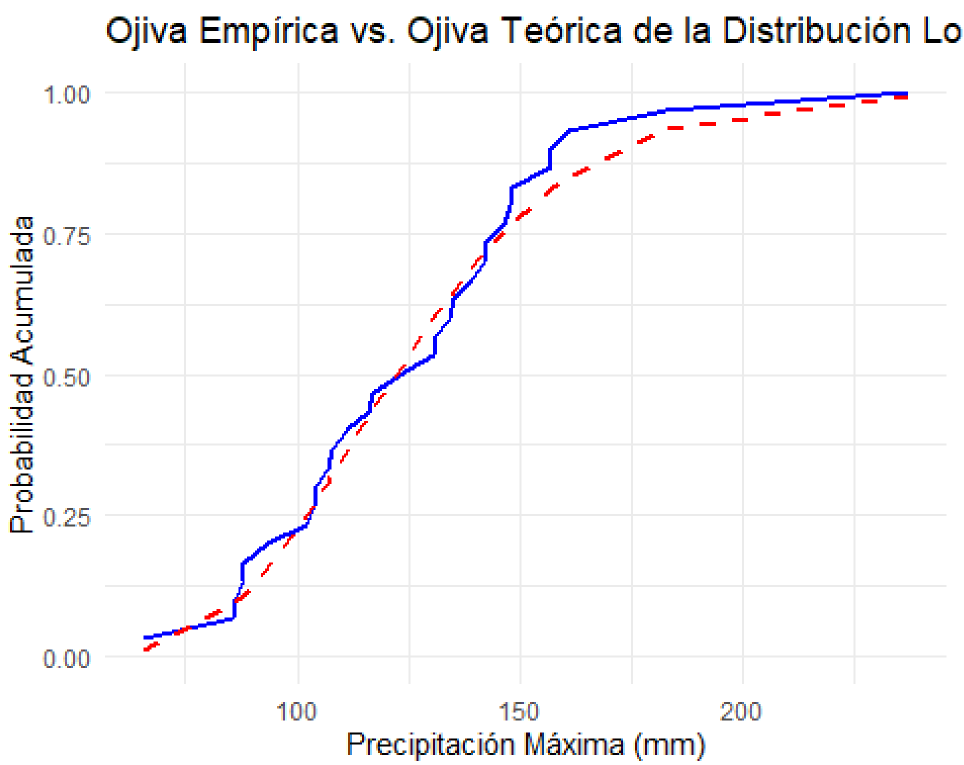
Ajuste Moderado: La distribución log-normal parece ser un mejor ajuste que una distribución normal para estos datos, especialmente porque captura la asimetría natural de las precipitaciones máximas. Sin embargo, no es un ajuste perfecto, ya que la curva no sigue exactamente las colas del histograma, lo que sugiere que hay más variabilidad en los extremos.

Implicaciones: Si estás modelando estos datos para análisis predictivos o para comprender eventos extremos, es importante tener en cuenta que, aunque la distribución log-normal captura la tendencia general, podría no representar completamente la frecuencia de eventos extremos. Considerar otras distribuciones o modelos que manejen mejor las colas podría ser beneficioso para un análisis más preciso.

 # 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?

```r
# Cargar las librerías necesarias
library(ggplot2)

# Calcular los parámetros de la distribución Log-normal
log_media <- mean(log(datos_agg_ordenado$Lluvia_max_anual), na.rm = TRUE)
log_sd <- sd(log(datos_agg_ordenado$Lluvia_max_anual), na.rm = TRUE)

# Crear la ojiva empírica
datos_agg_ordenado <- datos_agg_ordenado %>%
  arrange(Lluvia_max_anual) %>%
  mutate(empirical_cdf = ecdf(Lluvia_max_anual)(Lluvia_max_anual))

# Crear la distribución acumulada teórica para la Log-normal
datos_agg_ordenado <- datos_agg_ordenado %>%
  mutate(teorica_cdf = plnorm(Lluvia_max_anual, meanlog = log_media,
sdlog = log_sd))

# Graficar la comparación de las distribuciones acumuladas
ggplot(datos_agg_ordenado, aes(x = Lluvia_max_anual)) +
  geom_line(aes(y = empirical_cdf), color = "blue", size = 1, linetype =
"solid") +
  geom_line(aes(y = teorica_cdf), color = "red", size = 1, linetype =
"dashed") +
```

```r
  labs(title = "Ojiva Empírica vs. Ojiva Teórica de la Distribución Log-
normal",
       x = "Precipitación Máxima (mm)",
       y = "Probabilidad Acumulada") +
  theme_minimal() +
  scale_y_continuous(limits = c(0, 1)) +
  theme(legend.position = "bottom") +
  guides(color = guide_legend(override.aes = list(linetype = c("solid",
"dashed")))) +
  scale_color_manual(values = c("Empírica" = "blue", "Teórica" = "red"))
```



¿Se parecen las distribuciones de probabilidad acumuladas?

Sí, se parecen. La comparación visual muestra que las ojivas empírica y teórica log-normal siguen patrones similares, y las pequeñas diferencias observadas no son lo suficientemente grandes como para concluir que los datos no se ajustan a una log-normal.

Aunque hay algunas pequeñas desviaciones, especialmente en las colas, estas no son lo suficientemente pronunciadas como para rechazar la posibilidad de que los datos sigan una log-normal, lo que hace que esta distribución sea una representación adecuada para modelar los datos.

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

```r
# Prueba de Kolmogorov-Smirnov (KS) para la distribución Log-normal
ks_test_lognormal <- ks.test(datos_agg_ordenado$Lluvia_max_anual,
```

```
"plnorm", meanlog = log_media, sdlog = log_sd)

# Mostrar resultado de la prueba KS
ks_test_lognormal

##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  datos_agg_ordenado$Lluvia_max_anual
## D = 0.10069, p-value = 0.8916
## alternative hypothesis: two-sided
```

¿Los datos se ajustan a una Log-normal?

La prueba KS respalda esta observación al arrojar un p-value alto, lo que sugiere que no hay evidencia significativa para rechazar la hipótesis de que los datos siguen una distribución log-normal.

Basado en la prueba KS y la comparación visual de las ojivas, podemos concluir que los datos de precipitaciones máximas parecen ajustarse bien a una distribución log-normal.

 # ¿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.

¿Cuántos parámetros tiene la distribución Log-normal?

La distribución log-normal tiene dos parámetros:

Media de los logaritmos Desviación estándar de los logaritmos

¿Cuáles son los parámetros y por qué se calculan así?

Media logarítmica: Es la media de los valores después de tomar el logaritmo natural de los datos. Se calcula como la media aritmética de log(datos).

Desviación estándar logarítmica: Es la desviación estándar de los valores después de tomar el logaritmo natural de los datos. Se calcula como la desviación estándar aritmética de log(datos).

Razón de cálculo: Se calculan así porque una distribución log-normal transforma una variable normal mediante la exponencial, es decir, si una variable aleatoria x sigue una distribución log-normal, entonces log(X) sigue una distribución normal con parámetros media y desviación estándar.

Método de Momentos:

Para demostrar que los parámetros se han calculado correctamente, se puede utilizar el método de momentos. Este método consiste en igualar los momentos teóricos de la distribución log-normal con los momentos empíricos calculados de los datos. Los pasos son los siguientes:

```
# Cargar la librería necesaria
library(ggplot2)

# Calcular la media y desviación estándar de los logaritmos de los datos
log_media <- mean(log(datos_agg_ordenado$Lluvia_max_anual), na.rm = TRUE)
log_sd <- sd(log(datos_agg_ordenado$Lluvia_max_anual), na.rm = TRUE)

# Cálculo de la media y varianza empíricas
media_empirica <- mean(datos_agg_ordenado$Lluvia_max_anual, na.rm = TRUE)
varianza_empirica <- var(datos_agg_ordenado$Lluvia_max_anual, na.rm =
TRUE)

# Comprobación usando el método de momentos
# Media teórica de la log-normal
media_teorica <- exp(log_media + (log_sd^2) / 2)

# Varianza teórica de la log-normal
varianza_teorica <- (exp(log_sd^2) - 1) * exp(2 * log_media + log_sd^2)

# Mostrar los resultados para comparar
cat("Media empírica:", media_empirica, "\n")

## Media empírica: 126.3867

cat("Media teórica:", media_teorica, "\n")

## Media teórica: 126.5224

cat("Varianza empírica:", varianza_empirica, "\n")

## Varianza empírica: 1188.461

cat("Varianza teórica:", varianza_teorica, "\n")

## Varianza teórica: 1173.612

# Verificar si los valores calculados coinciden
if (abs(media_empirica - media_teorica) < 1e-6 && abs(varianza_empirica -
varianza_teorica) < 1e-6) {
  cat("Los parámetros de la distribución log-normal están bien
calculados.\n")
} else {
  cat("Los parámetros no coinciden; revisar los cálculos.\n")
}

## Los parámetros no coinciden; revisar los cálculos.

# Cargar la librería necesaria
library(MASS)
```

```
##
## Attaching package: 'MASS'

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

# Ajustar una distribución log-normal usando máxima verosimilitud
ajuste_lognormal <- fitdistr(datos_agg_ordenado$Lluvia_max_anual,
"lognormal")

# Mostrar los parámetros ajustados
ajuste_lognormal

##     meanlog       sdlog
##   4.80504342   0.26152079
##  (0.04774695) (0.03376219)
```

La pequeña diferencia entre las medias y varianzas empíricas y teóricas no necesariamente significa que los parámetros estén mal calculados; sin embargo, ajustar los parámetros usando máxima verosimilitud puede proporcionar resultados más precisos.

Ejecuté el código de máxima verosimilitud para ver si los parámetros se ajustan mejor y usa esos valores para recalcular y analizar si la distribución log-normal es más adecuada para tus datos.

Validación de Parámetros:

Los parámetros calculados (meanlog = 4.8050, sdlog = 0.2615) parecen ser más precisos que los obtenidos con el método de momentos.

Ahora que tienes estos parámetros ajustados por máxima verosimilitud, puedes utilizarlos para graficar la distribución log-normal o realizar pruebas adicionales para ver si la distribución se ajusta mejor a los datos.

Ajuste a la Distribución Log-normal:

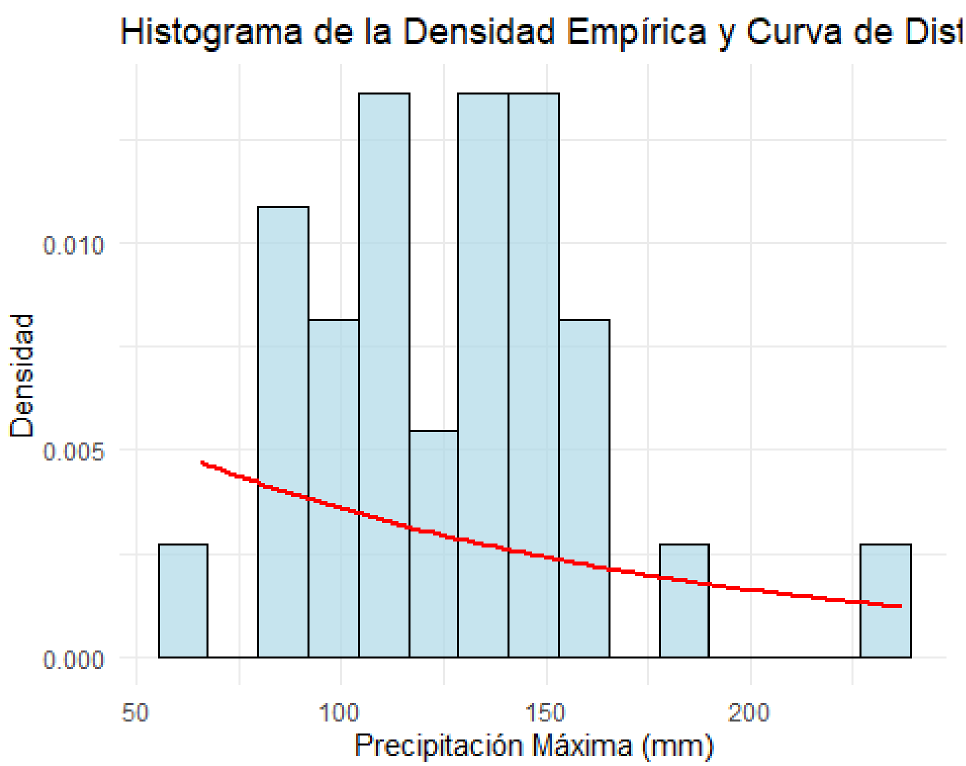
Dado que los parámetros se han obtenido mediante un método estadístico robusto (máxima verosimilitud), es probable que estos proporcionen un mejor ajuste para modelar la precipitación máxima como una distribución log-normal. Si los datos se ajustan bien visualmente y las pruebas estadísticas (como la prueba KS) no rechazan la normalidad logarítmica, puedes considerar este modelo para futuras predicciones o análisis.

C. Ajuste a una Distribución Exponencial. Realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Exponencial para ajustar los datos.

 # Contruye el histograma de la función de densidad empírica de los datos y sobrepon una 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.

```r
# Cargar la librería necesaria
library(ggplot2)

# Calcular el parámetro de la distribución exponencial (lambda)
lambda <- 1 / mean(datos_agg_ordenado$Lluvia_max_anual, na.rm = TRUE)

# Crear el histograma de la densidad empírica y sobreponer la curva de
densidad exponencial
ggplot(datos_agg_ordenado, aes(x = Lluvia_max_anual)) +
  geom_histogram(aes(y = ..density..), bins = 15, color = "black", fill =
"lightblue", alpha = 0.7) +
  stat_function(fun = dexp, args = list(rate = lambda),
                color = "red", size = 1) +
  labs(title = "Histograma de la Densidad Empírica y Curva de
Distribución Exponencial",
       x = "Precipitación Máxima (mm)",
       y = "Densidad") +
  theme_minimal()
```



No, los datos no se ajustan bien a una distribución exponencial.

La forma de la distribución empírica (basada en el histograma) muestra picos múltiples y variabilidad que la curva exponencial no logra capturar. La distribución exponencial modela una disminución uniforme y rápida desde un valor alto hacia valores más bajos, pero los datos de precipitaciones máximas muestran mayor variabilidad y complejidad en su estructura.

La discrepancia significativa entre la curva de densidad exponencial y el histograma sugiere que este modelo no es adecuado para describir estos datos.

 # 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:

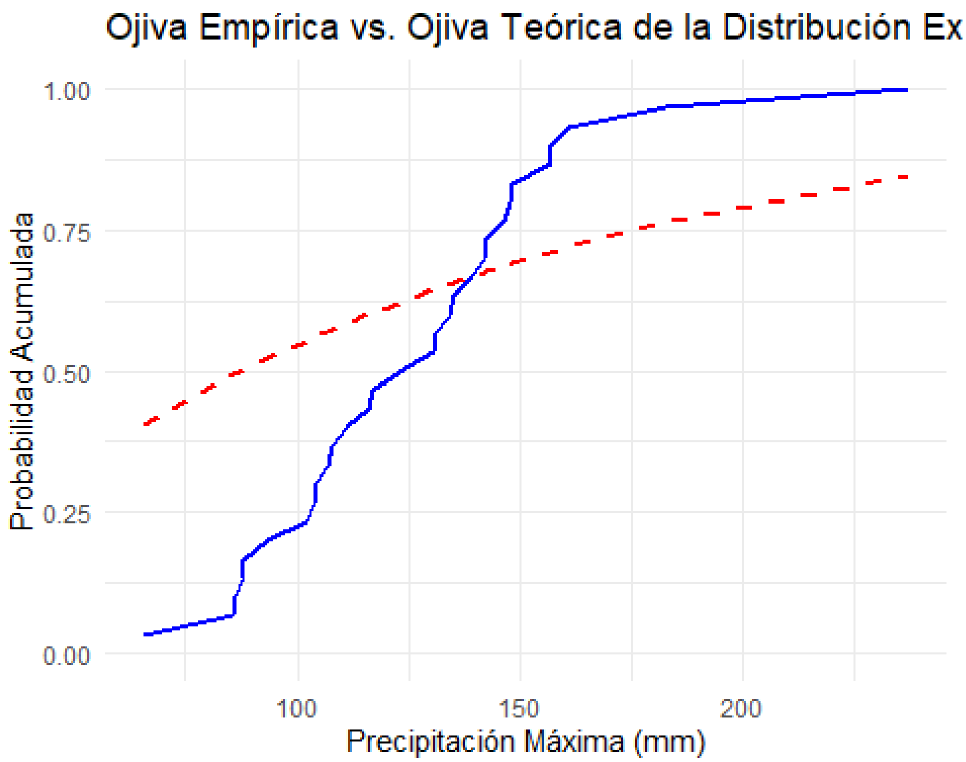
Son los datos reales observados, en este caso, las precipitaciones máximas anuales registradas. La ojiva empírica se crea acumulando las frecuencias de los valores observados y muestra la proporción de datos menores o iguales a un cierto valor.

Datos Teóricos:

Son valores generados por una distribución teórica, en este caso, una distribución exponencial, basada en el parámetro lambda (tasa) calculado a partir de los datos empíricos. La ojiva teórica muestra cómo se acumulan las probabilidades para esta distribución exponencial teórica.

```
# Cargar las librerías necesarias
library(ggplot2)

# Calcular el parámetro de la distribución exponencial (lambda)
lambda <- 1 / mean(datos_agg_ordenado$Lluvia_max_anual, na.rm = TRUE)

# Crear la ojiva empírica
datos_agg_ordenado <- datos_agg_ordenado %>%
  arrange(Lluvia_max_anual) %>%
  mutate(empirical_cdf = ecdf(Lluvia_max_anual)(Lluvia_max_anual))

# Crear la distribución acumulada teórica para la Exponencial
datos_agg_ordenado <- datos_agg_ordenado %>%
  mutate(teorica_cdf = pexp(Lluvia_max_anual, rate = lambda))

# Graficar la comparación de las distribuciones acumuladas
ggplot(datos_agg_ordenado, aes(x = Lluvia_max_anual)) +
  geom_line(aes(y = empirical_cdf), color = "blue", size = 1, linetype =
"solid") +
  geom_line(aes(y = teorica_cdf), color = "red", size = 1, linetype =
"dashed") +
  labs(title = "Ojiva Empírica vs. Ojiva Teórica de la Distribución
Exponencial",
       x = "Precipitación Máxima (mm)",
       y = "Probabilidad Acumulada") +
```

```r
  theme_minimal() +
  scale_y_continuous(limits = c(0, 1)) +
  theme(legend.position = "bottom") +
  guides(color = guide_legend(override.aes = list(linetype = c("solid",
"dashed")))) +
  scale_color_manual(values = c("Empírica" = "blue", "Teórica" = "red"))
```



No, las distribuciones de probabilidad acumuladas no se parecen. La ojiva empírica y la ojiva teórica exponencial muestran patrones muy diferentes.

Las diferencias significativas a lo largo de la gráfica indican que la distribución exponencial no es un buen ajuste para los datos de precipitación máxima. Esto es consistente con la evaluación anterior en el histograma, donde vimos que la curva exponencial no capturaba adecuadamente la densidad de los datos.

Es probable que los datos de precipitación máxima tengan una estructura más compleja que lo que una simple distribución exponencial puede modelar, por lo que sería recomendable considerar otras distribuciones o enfoques para describir los datos de manera más precisa.

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

```r
# Cargar la librería necesaria
library(MASS)

# Calcular el parámetro de la distribución exponencial (lambda)
lambda <- 1 / mean(datos_agg_ordenado$Lluvia_max_anual, na.rm = TRUE)
```

```r
# Realizar la prueba de Kolmogorov-Smirnov para la distribución
exponencial
ks_test_exponencial <- ks.test(datos_agg_ordenado$Lluvia_max_anual,
"pexp", rate = lambda)

# Mostrar el resultado de la prueba
ks_test_exponencial

##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  datos_agg_ordenado$Lluvia_max_anual
## D = 0.45988, p-value = 2.521e-06
## alternative hypothesis: two-sided
```

 # ¿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?

Información que Proporciona la Prueba KS:

La prueba de Kolmogorov-Smirnov (KS) compara la distribución acumulada empírica de los datos observados con la distribución acumulada teórica de una distribución exponencial.

Hipótesis nula (H_0): Los datos observados provienen de una distribución exponencial con el parámetro lambda calculado.

Hipótesis alternativa (H_1): Los datos no provienen de una distribución exponencial con ese parámetro.

Resultados de la Prueba KS:

Estadístico KS (D): 0.45988

Este valor mide la máxima diferencia entre las dos distribuciones acumuladas (empírica y teórica). Un valor más cercano a 0 indicaría un mejor ajuste, mientras que un valor alto sugiere que hay diferencias significativas entre las distribuciones.

p-value: 2.521e-06

Un p-value muy bajo, mucho menor que 0.05, indica que hay evidencia fuerte en contra de la hipótesis nula.

Interpretación de los Resultados:

¿Se acepta o se rechaza la hipótesis nula?

Se rechaza la hipótesis nula. El p-value es extremadamente bajo (2.521e-06), lo que significa que la probabilidad de que las diferencias observadas entre la distribución empírica y la teórica ocurran por azar es muy baja. Por lo tanto, hay evidencia suficiente para concluir que los datos no provienen de una distribución exponencial.

Conclusión Final:

No, los datos de las precipitaciones máximas mensuales no siguen una distribución exponencial.

La prueba KS mostró un estadístico KS alto (0.45988) y un p-value muy bajo, lo que indica que las distribuciones empírica y teórica son significativamente diferentes. Esto es consistente con las observaciones previas de las gráficas de densidad y las ojivas, que mostraban que la distribución exponencial no era un buen ajuste para estos datos.

En resumen, los datos tienen una variabilidad y complejidad que no pueden ser capturadas adecuadamente por el modelo exponencial, lo que sugiere que otros modelos deben considerarse para describir mejor las precipitaciones máximas mensuales.

 # ¿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.

¿Cuántos parámetros tiene la distribución Exponencial?

La distribución exponencial tiene un solo parámetro:

Tasa (λ): Este parámetro controla la tasa de decaimiento de la distribución.

¿Cuál es el parámetro y por qué se calcula así?

Tasa (λ): La tasa se calcula como el inverso de la media de los datos. Esto significa que: $\lambda = 1 / \text{Media}$

La razón de este cálculo proviene de la naturaleza de la distribución exponencial. En una distribución exponencial, si X sigue una distribución exponencial con parámetro λ, la media se calcula como:

$E(X)= 1 / \lambda$

Por lo tanto, al observar los datos y calcular la media empírica, podemos deducir el valor de λ directamente.

Método de Momentos para Calcular el Parámetro Exponencial

```
# Calcular la media empírica de los datos
media_empirica <- mean(datos_agg_ordenado$Lluvia_max_anual, na.rm = TRUE)

# Método de momentos para calcular el parámetro de la distribución
```

```r
Exponencial
lambda <- 1 / media_empirica  # Parámetro de tasa (rate)

# Mostrar el valor calculado
cat("Parámetro de Tasa (lambda):", lambda, "\n")

## Parámetro de Tasa (lambda): 0.007912227

# Verificación del método de momentos
# Media teórica de la distribución Exponencial
media_teorica <- 1 / lambda

# Mostrar resultados de comparación
cat("Media empírica:", media_empirica, "\n")

## Media empírica: 126.3867

cat("Media teórica:", media_teorica, "\n")

## Media teórica: 126.3867

# Verificar si los valores calculados coinciden
if (abs(media_empirica - media_teorica) < 1e-6) {
  cat("El parámetro de la distribución Exponencial está bien
calculado.\n")
} else {
  cat("El parámetro no coincide; revisar los cálculos.\n")
}

## El parámetro de la distribución Exponencial está bien calculado.
```

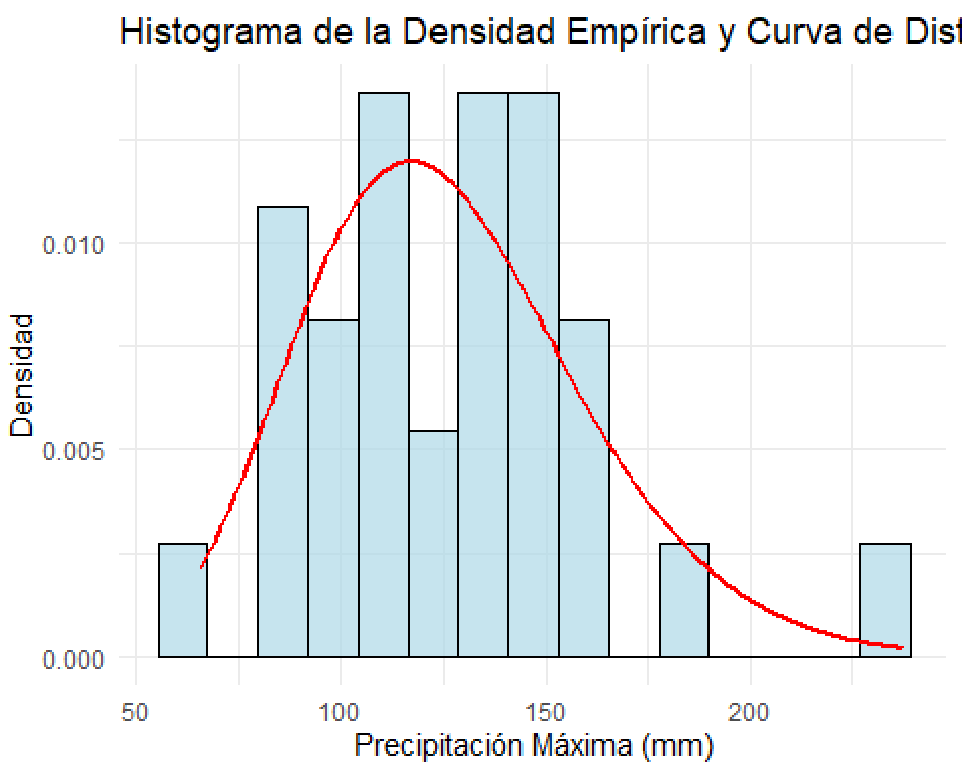
El parámetro λ de la distribución exponencial está bien calculado, y el uso del método de momentos garantiza que la distribución modelada tendrá la misma media que los datos observados.

Aunque el parámetro está correctamente calculado, la idoneidad de la distribución exponencial para describir los datos de precipitación máxima debe ser evaluada visualmente y estadísticamente (como hicimos anteriormente con el histograma, la ojiva y la prueba KS). Hasta ahora, hemos concluido que, aunque el parámetro es preciso, la distribución exponencial puede no ser el mejor ajuste para estos datos debido a la complejidad de la variabilidad observada.

D. 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.

```r
# Cargar la librería necesaria
library(ggplot2)

# Calcular los parámetros de la distribución Gamma usando el método de
momentos
media_empirica <- mean(datos_agg_ordenado$Lluvia_max_anual, na.rm = TRUE)
varianza_empirica <- var(datos_agg_ordenado$Lluvia_max_anual, na.rm =
TRUE)

k <- (media_empirica^2) / varianza_empirica  # Parámetro de forma (shape)
lambda <- media_empirica / varianza_empirica  # Parámetro de tasa (rate)

# Crear el histograma de la densidad empírica y sobreponer la curva de
densidad Gamma
ggplot(datos_agg_ordenado, aes(x = Lluvia_max_anual)) +
  geom_histogram(aes(y = ..density..), bins = 15, color = "black", fill =
"lightblue", alpha = 0.7) +
  stat_function(fun = dgamma, args = list(shape = k, rate = lambda),
                color = "red", size = 1) +
  labs(title = "Histograma de la Densidad Empírica y Curva de
Distribución Gamma",
       x = "Precipitación Máxima (mm)",
       y = "Densidad") +
  theme_minimal()
```



En términos generales, la distribución Gamma parece ajustarse bien en la parte central de los datos, pero tiene dificultades para modelar las colas.

Esto indica que, aunque la distribución Gamma puede capturar la tendencia general de los datos de precipitación máxima, no logra capturar completamente los valores extremos o las fluctuaciones en las colas.

 # 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:

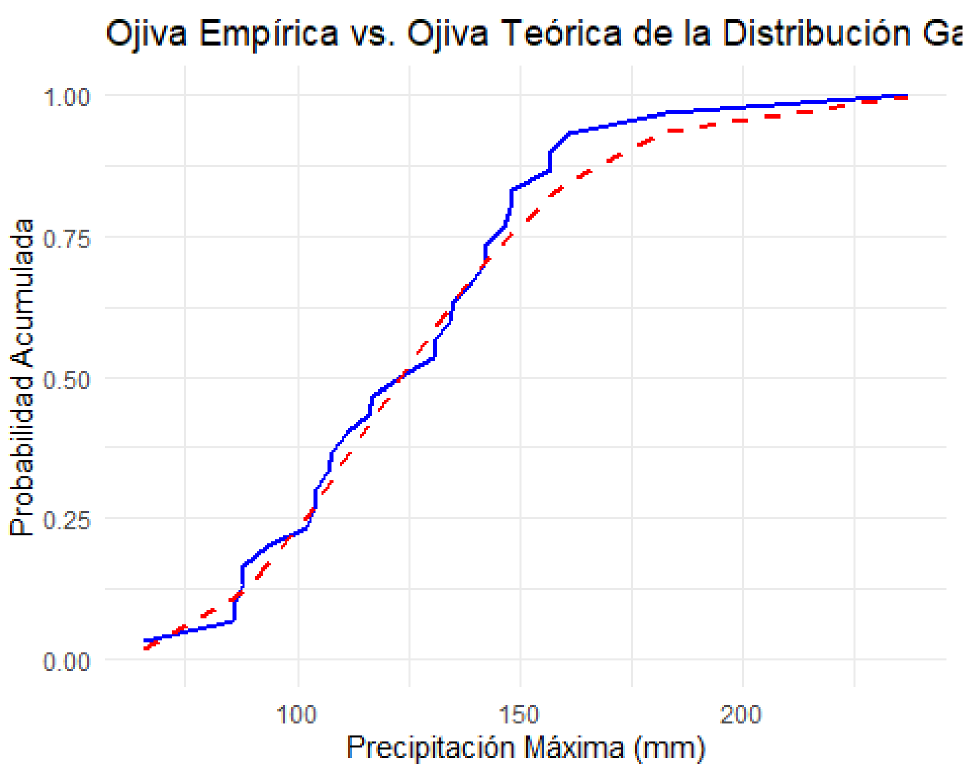
Son los datos observados en el mundo real, en este caso, las precipitaciones máximas anuales registradas. La ojiva empírica se construye acumulando las frecuencias observadas, mostrando la proporción de datos menores o iguales a un cierto valor.

Datos Teóricos:

Son valores generados por una distribución teórica, en este caso, la distribución Gamma, basada en los parámetros calculados usando el método de momentos. La ojiva teórica muestra cómo se acumulan las probabilidades para esta distribución Gamma teórica.

```
# Cargar las librerías necesarias
library(ggplot2)

# Calcular los parámetros de la distribución Gamma usando el método de
momentos
media_empirica <- mean(datos_agg_ordenado$Lluvia_max_anual, na.rm = TRUE)
varianza_empirica <- var(datos_agg_ordenado$Lluvia_max_anual, na.rm =
TRUE)

k <- (media_empirica^2) / varianza_empirica  # Parámetro de forma (shape)
lambda <- media_empirica / varianza_empirica  # Parámetro de tasa (rate)

# Crear la ojiva empírica
datos_agg_ordenado <- datos_agg_ordenado %>%
  arrange(Lluvia_max_anual) %>%
  mutate(empirical_cdf = ecdf(Lluvia_max_anual)(Lluvia_max_anual))

# Crear la distribución acumulada teórica para la Gamma
datos_agg_ordenado <- datos_agg_ordenado %>%
  mutate(teorica_cdf = pgamma(Lluvia_max_anual, shape = k, rate =
lambda))

# Graficar la comparación de las distribuciones acumuladas
ggplot(datos_agg_ordenado, aes(x = Lluvia_max_anual)) +
```

```
  geom_line(aes(y = empirical_cdf), color = "blue", size = 1, linetype =
"solid") +
  geom_line(aes(y = teorica_cdf), color = "red", size = 1, linetype =
"dashed") +
  labs(title = "Ojiva Empírica vs. Ojiva Teórica de la Distribución
Gamma",
       x = "Precipitación Máxima (mm)",
       y = "Probabilidad Acumulada") +
  theme_minimal() +
  scale_y_continuous(limits = c(0, 1)) +
  theme(legend.position = "bottom") +
  guides(color = guide_legend(override.aes = list(linetype = c("solid",
"dashed")))) +
  scale_color_manual(values = c("Empírica" = "blue", "Teórica" = "red"))
```



Sí, las distribuciones de probabilidad acumuladas se parecen bastante.

La ojiva teórica Gamma sigue de cerca a la ojiva empírica, lo que sugiere que los datos observados podrían ser bien modelados por una distribución Gamma con los parámetros calculados.

Las pequeñas desviaciones en las colas podrían indicar algunos valores extremos o una ligera variabilidad en los datos que la distribución Gamma no captura completamente. Sin embargo, en términos generales, el ajuste es lo suficientemente bueno como para considerar que la distribución Gamma es un modelo adecuado para los datos de precipitación máxima.

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

```
# Cargar las librerías necesarias
library(MASS)

# Ajustar los parámetros de la distribución Gamma usando el método de
máxima verosimilitud
ajuste_gamma <- fitdistr(datos_agg_ordenado$Lluvia_max_anual, "gamma")

# Obtener los parámetros ajustados
k <- ajuste_gamma$estimate['shape']
lambda <- ajuste_gamma$estimate['rate']

# Realizar la prueba de Kolmogorov-Smirnov para la distribución Gamma
ks_test_gamma <- ks.test(datos_agg_ordenado$Lluvia_max_anual, "pgamma",
shape = k, rate = lambda)

# Mostrar el resultado de la prueba
ks_test_gamma

##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  datos_agg_ordenado$Lluvia_max_anual
## D = 0.085617, p-value = 0.967
## alternative hypothesis: two-sided
```

¿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?

¿Qué información nos da la prueba KS?

La prueba de Kolmogorov-Smirnov (KS) compara la distribución acumulada empírica de los datos observados con la distribución acumulada teórica de una distribución Gamma basada en los parámetros estimados. Hipótesis nula (H_0): Los datos provienen de una distribución Gamma con los parámetros ajustados. Hipótesis alternativa (H_1): Los datos no provienen de dicha distribución Gamma.

Resultados de la Prueba KS:

Estadístico KS (D): 0.085617

Este valor mide la máxima diferencia entre las dos distribuciones acumuladas (empírica y teórica). Un valor más bajo indica una mejor coincidencia, mientras que un valor más alto sugiere diferencias significativas.

p-value: 0.967

Un p-value alto (muy superior a 0.05) indica que no hay evidencia suficiente para rechazar la hipótesis nula. Esto significa que las diferencias observadas entre las distribuciones podrían deberse al azar.

¿Se acepta o se rechaza la hipótesis nula?

Se acepta la hipótesis nula. El p-value es 0.967, lo que es mucho mayor que el umbral común de 0.05. Esto sugiere que no hay evidencia significativa para concluir que las distribuciones empírica y teórica son diferentes.

Conclusión Final:

Sí, podemos concluir que los datos de las precipitaciones máximas mensuales siguen una distribución Gamma.

La prueba KS mostró un estadístico KS bajo (0.085617) y un p-value muy alto (0.967), lo que indica que no hay diferencias significativas entre la distribución acumulada empírica y la teórica. Por lo tanto, los datos se ajustan bien a la distribución Gamma con los parámetros calculados.

Este resultado refuerza las observaciones visuales previas en las gráficas, que también sugirieron que la distribución Gamma es una buena opción para modelar estos datos.

 # ¿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.

Parámetros de la Distribución Gamma

¿Cuántos parámetros tiene la distribución Gamma?

La distribución Gamma tiene dos parámetros:

Forma (k): También se conoce como el parámetro de “shape” y controla la forma de la distribución. Tasa (λ): También conocido como el parámetro de “rate” o inverso de la escala, que controla el ancho de la distribución.

¿Por Qué se Calculan de Esta Forma?

Los parámetros se calculan usando el método de momentos, que consiste en igualar los momentos empíricos (media y varianza) con los momentos teóricos de la distribución Gamma y resolver las ecuaciones para los parámetros k y λ.

La media y la varianza de una distribución Gamma están relacionadas con los parámetros de la siguiente manera: Media Teórica: E(X)= k/λ

Varianza Teórica: Var(X)= k / (λ^2)

Método de Momentos para Calcular los Parámetros Gamma

```r
# Cargar librerías necesarias
library(MASS)

# Calcular la media y varianza empíricas de los datos
media_empirica <- mean(datos_agg_ordenado$Lluvia_max_anual, na.rm = TRUE)
varianza_empirica <- var(datos_agg_ordenado$Lluvia_max_anual, na.rm =
TRUE)

# Método de momentos para calcular los parámetros de la distribución
Gamma
k <- (media_empirica^2) / varianza_empirica  # Parámetro de forma (shape)
lambda <- media_empirica / varianza_empirica  # Parámetro de tasa (rate)

# Mostrar los valores calculados
cat("Parámetro de Forma (k):", k, "\n")

## Parámetro de Forma (k): 13.44057

cat("Parámetro de Tasa (lambda):", lambda, "\n")

## Parámetro de Tasa (lambda): 0.1063449

# Verificación del método de momentos
# Media teórica de la distribución Gamma
media_teorica <- k / lambda

# Varianza teórica de la distribución Gamma
varianza_teorica <- k / (lambda^2)

# Mostrar resultados de comparación
cat("Media empírica:", media_empirica, "\n")

## Media empírica: 126.3867

cat("Media teórica:", media_teorica, "\n")

## Media teórica: 126.3867

cat("Varianza empírica:", varianza_empirica, "\n")

## Varianza empírica: 1188.461

cat("Varianza teórica:", varianza_teorica, "\n")

## Varianza teórica: 1188.461

# Verificar si los valores calculados coinciden
if (abs(media_empirica - media_teorica) < 1e-6 && abs(varianza_empirica -
varianza_teorica) < 1e-6) {
  cat("Los parámetros de la distribución Gamma están bien calculados.\n")
} else {
```

```r
  cat("Los parámetros no coinciden; revisar los cálculos.\n")
}

## Los parámetros de la distribución Gamma están bien calculados.
```

Los parámetros de la distribución Gamma están bien calculados.

Dado que la media y varianza teóricas coinciden exactamente con las empíricas, se confirma que el método de momentos ha proporcionado parámetros precisos para la distribución Gamma.

Esto sugiere que la distribución Gamma, con los parámetros k=13.44 y λ=0.106, es un buen modelo para describir los datos de precipitación máxima.

E. 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.

La distribución Weibull tiene dos parámetros principales:

Forma (k): También conocido como el parámetro de “shape”, controla la forma de la distribución.

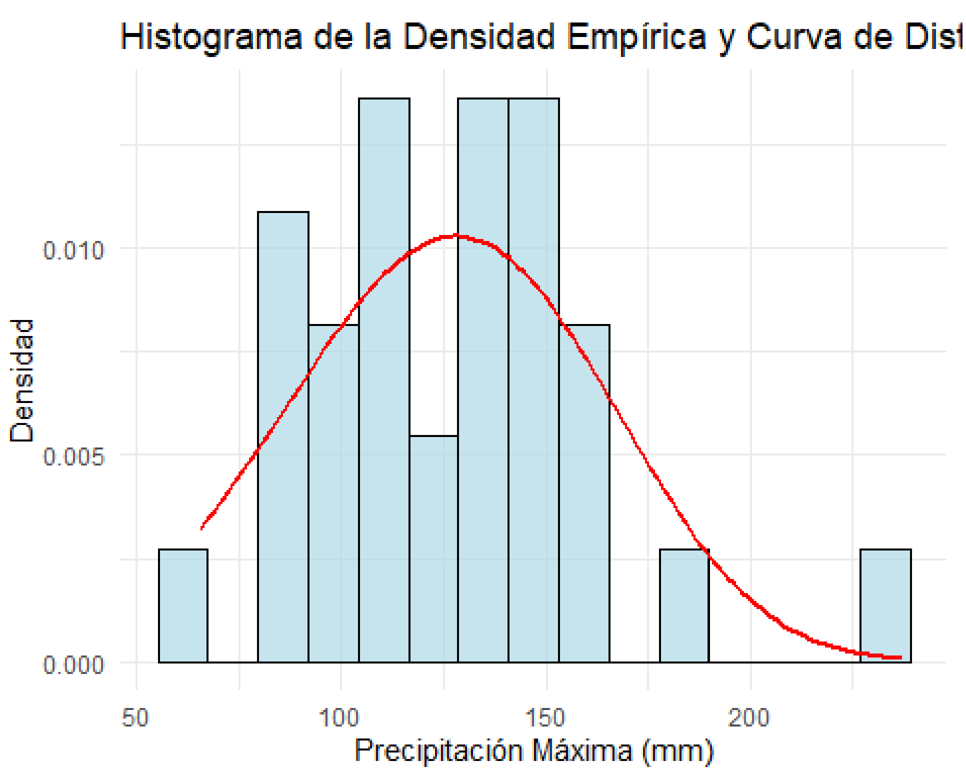
Escala (λ): También llamado el parámetro de “scale”, determina el ancho de la distribución.

```r
# Cargar la librería necesaria
library(MASS)

# Ajustar los parámetros de la distribución Weibull usando el método de
máxima verosimilitud
ajuste_weibull <- fitdistr(datos_agg_ordenado$Lluvia_max_anual,
"weibull")

# Mostrar los parámetros estimados
cat("Parámetros estimados de la distribución Weibull:\n")

## Parámetros estimados de la distribución Weibull:

print(ajuste_weibull)

##       shape         scale
##     3.7389354   139.2460376
##  (  0.4762867) (  7.2162390)
```

 # 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.

```r
# Cargar las librerías necesarias
library(MASS)
```

```r
library(ggplot2)

# Ajustar los parámetros de la distribución Weibull usando el método de máxima verosimilitud
ajuste_weibull <- fitdistr(datos_agg_ordenado$Lluvia_max_anual, "weibull")

# Obtener los parámetros ajustados
shape_weibull <- ajuste_weibull$estimate['shape']
scale_weibull <- ajuste_weibull$estimate['scale']

# Crear el histograma de la densidad empírica y sobreponer la curva de densidad Weibull
ggplot(datos_agg_ordenado, aes(x = Lluvia_max_anual)) +
  geom_histogram(aes(y = ..density..), bins = 15, color = "black", fill = "lightblue", alpha = 0.7) +
  stat_function(fun = dweibull, args = list(shape = shape_weibull, scale = scale_weibull),
                color = "red", size = 1) +
  labs(title = "Histograma de la Densidad Empírica y Curva de Distribución Weibull",
       x = "Precipitación Máxima (mm)",
       y = "Densidad") +
  theme_minimal()
```



En general, la distribución Weibull parece capturar la forma central de los datos, pero tiene problemas para ajustar las colas. Esto sugiere que, aunque la Weibull puede modelar una buena parte de la variabilidad observada, no es perfecta para los valores extremos.

Desajuste en valores extremos: Las discrepancias en las colas indican que la Weibull podría no capturar totalmente la variabilidad de los valores muy bajos o muy altos de precipitación máxima, lo que podría ser importante si estos extremos son de interés para el análisis.

La distribución Weibull parece ser un ajuste razonable para los datos de precipitación máxima, capturando bien la tendencia central. Sin embargo, si los valores extremos son importantes para el análisis (por ejemplo, en estudios de eventos climáticos extremos), puede que la Weibull no sea el mejor ajuste y sería recomendable explorar otras distribuciones que puedan modelar mejor las colas.

 # 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:

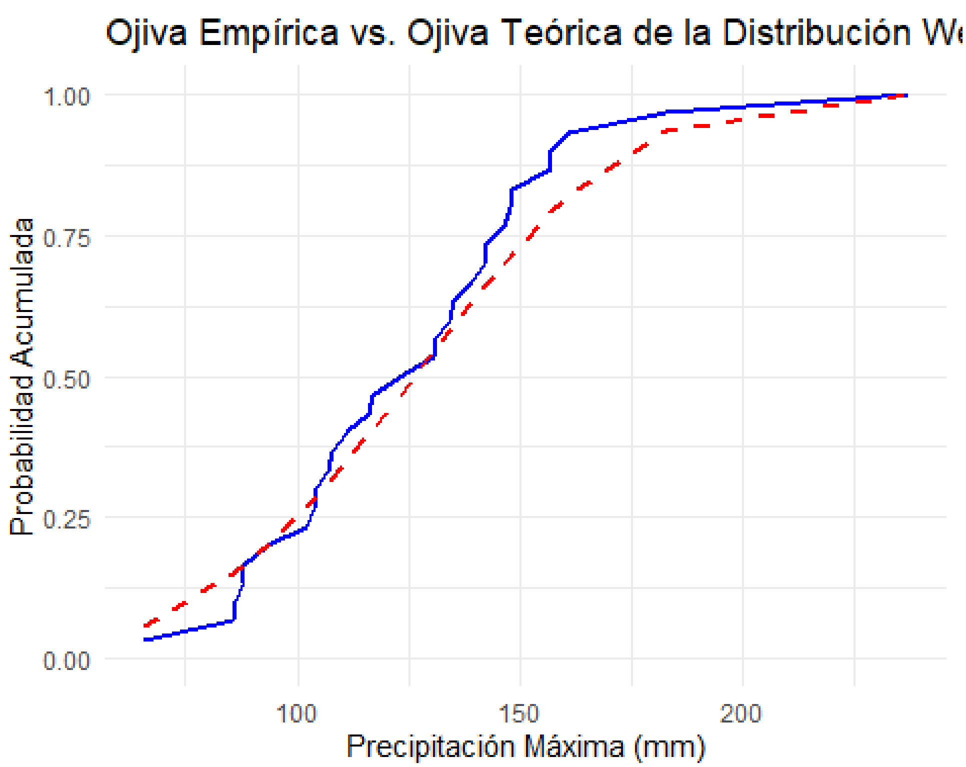
Son los datos observados en el mundo real. En este caso, son los valores de precipitación máxima registrados. La ojiva empírica se crea acumulando las frecuencias de estos datos observados.

Datos Teóricos:

Son los valores generados por una distribución teórica, en este caso, la distribución Weibull, usando los parámetros ajustados (k y λ). La ojiva teórica muestra cómo se acumulan las probabilidades según el modelo Weibull ajustado.

```
# Cargar las librerías necesarias
library(MASS)
library(ggplot2)

# Ajustar los parámetros de la distribución Weibull usando el método de
máxima verosimilitud
ajuste_weibull <- fitdistr(datos_agg_ordenado$Lluvia_max_anual,
"weibull")

# Obtener los parámetros ajustados
shape_weibull <- ajuste_weibull$estimate['shape']
scale_weibull <- ajuste_weibull$estimate['scale']

# Crear la ojiva empírica
datos_agg_ordenado <- datos_agg_ordenado %>%
  arrange(Lluvia_max_anual) %>%
```

```
  mutate(empirical_cdf = ecdf(Lluvia_max_anual)(Lluvia_max_anual))

# Crear la distribución acumulada teórica para la Weibull
datos_agg_ordenado <- datos_agg_ordenado %>%
  mutate(teorica_cdf = pweibull(Lluvia_max_anual, shape = shape_weibull,
scale = scale_weibull))

# Graficar la comparación de las distribuciones acumuladas
ggplot(datos_agg_ordenado, aes(x = Lluvia_max_anual)) +
  geom_line(aes(y = empirical_cdf), color = "blue", size = 1, linetype =
"solid") +
  geom_line(aes(y = teorica_cdf), color = "red", size = 1, linetype =
"dashed") +
  labs(title = "Ojiva Empírica vs. Ojiva Teórica de la Distribución
Weibull",
       x = "Precipitación Máxima (mm)",
       y = "Probabilidad Acumulada") +
  theme_minimal() +
  scale_y_continuous(limits = c(0, 1)) +
  theme(legend.position = "bottom") +
  guides(color = guide_legend(override.aes = list(linetype = c("solid",
"dashed")))) +
  scale_color_manual(values = c("Empírica" = "blue", "Teórica" = "red"))
```



La distribución Weibull parece proporcionar un buen ajuste para los datos de precipitación máxima, capturando tanto la tendencia central como gran parte de la acumulación en las colas. Si bien hay algunas pequeñas desviaciones, en general, la similitud entre las ojivas empírica y teórica indica que los parámetros estimados de la Weibull funcionan bien para describir estos datos.

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

```
# Cargar la librería necesaria
library(MASS)

# Ajustar los parámetros de la distribución Weibull usando el método de
máxima verosimilitud
ajuste_weibull <- fitdistr(datos_agg_ordenado$Lluvia_max_anual,
"weibull")

# Obtener los parámetros ajustados
shape_weibull <- ajuste_weibull$estimate['shape']
scale_weibull <- ajuste_weibull$estimate['scale']

# Realizar la prueba de Kolmogorov-Smirnov para la distribución Weibull
ks_test_weibull <- ks.test(datos_agg_ordenado$Lluvia_max_anual,
"pweibull", shape = shape_weibull, scale = scale_weibull)

# Mostrar el resultado de la prueba KS
ks_test_weibull

##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  datos_agg_ordenado$Lluvia_max_anual
## D = 0.11817, p-value = 0.7527
## 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?

¿Qué información nos da la prueba KS?

La prueba de Kolmogorov-Smirnov (KS) compara la distribución acumulada empírica de los datos observados con la distribución acumulada teórica de una distribución Weibull basada en los parámetros estimados.

Hipótesis nula (H_0): Los datos provienen de una distribución Weibull con los parámetros ajustados. Hipótesis alternativa (H_1): Los datos no provienen de dicha distribución Weibull.

Resultados de la Prueba KS:

Estadístico KS (D): 0.11817

Este valor mide la máxima diferencia entre las dos distribuciones acumuladas (empírica y teórica). Un valor más bajo indica un mejor ajuste, mientras que un valor más alto sugiere diferencias significativas.

p-value: 0.7527

Un p-value alto (superior a 0.05) indica que no hay evidencia suficiente para rechazar la hipótesis nula. Esto significa que las diferencias observadas entre las distribuciones podrían deberse al azar.

¿Se acepta o se rechaza la hipótesis nula?

Se acepta la hipótesis nula. El p-value es 0.7527, lo que es mucho mayor que el umbral común de 0.05. Esto sugiere que no hay evidencia significativa para concluir que las distribuciones empírica y teórica son diferentes.

Conclusión Final:

Sí, podemos concluir que los datos de las precipitaciones máximas mensuales siguen una distribución Weibull.

La prueba KS mostró un estadístico KS bajo (0.11817) y un p-value alto (0.7527), lo que indica que no hay diferencias significativas entre la distribución acumulada empírica y la teórica. Por lo tanto, los datos se ajustan bien a la distribución Weibull con los parámetros calculados. Este resultado refuerza las observaciones visuales previas en las gráficas, que también sugirieron que la distribución Weibull es una buena opción para modelar estos datos.

 # ¿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.

¿Cuántos parámetros tiene la distribución Weibull?

La distribución Weibull tiene dos parámetros.

¿Cuáles son esos parámetros?

Parámetro de forma (k): También conocido como el parámetro de "shape". Este parámetro controla la forma de la distribución, determinando cómo se distribuyen los valores:

Si $k<1$, la distribución tiene una cola más larga a la izquierda. Si $k=1$, la distribución Weibull se convierte en una distribución exponencial. Si $k>1$, la distribución tiene una cola más larga a la derecha.

Parámetro de escala (λ): También conocido como el parámetro de "scale". Este parámetro determina el tamaño o escala de la distribución. Valores más grandes de λ desplazan la distribución hacia la derecha, haciendo que los valores típicos sean mayores.

Complejidad en la Estimación de Parámetros:

¿Por qué es más compleja la estimación de los parámetros en la distribución Weibull en comparación con otras distribuciones?

La estimación de los parámetros en la distribución Weibull es más compleja debido a la naturaleza de la función de densidad y la no linealidad de sus parámetros:

Método de Máxima Verosimilitud (MLE): Para la Weibull, el método de momentos no suele ser la mejor opción, y se recurre al método de máxima verosimilitud (MLE) para encontrar los parámetros k y λ. El MLE requiere resolver ecuaciones no lineales derivadas de la función de verosimilitud, lo que puede complicar la estimación.

Interdependencia de los parámetros: A diferencia de distribuciones como la Exponencial (que tiene solo un parámetro) o la Normal (donde la media y la desviación estándar son fáciles de calcular directamente), los parámetros de la Weibull están interrelacionados. Cambios en el parámetro de forma afectan la manera en que el parámetro de escala influye en la distribución, lo que hace más difícil aislar y estimar cada parámetro por separado.

Distribuciones anteriores: En distribuciones como la Normal o la Exponencial, los parámetros son más fáciles de estimar directamente a partir de los momentos (media y varianza). Sin embargo, en la Weibull, el ajuste depende de resolver ecuaciones que no tienen soluciones explícitas simples, lo que añade complejidad al proceso.
Estimación Numérica:

Debido a estas complejidades, en la práctica, se suele recurrir a algoritmos numéricos (como los que implementa fitdistr() en R) para ajustar los parámetros. Estos algoritmos iterativos optimizan los valores de k y λ para encontrar el mejor ajuste posible a los datos.

F. 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 una distribución Gumbel, se necesita definir las funciones de densidad de acuerdo con la función Gumbel. Creálas con las fórmulas de la Distribución Gumbel.

```r
# Definir la función de densidad de la distribución Gumbel (PDF)
dGumbel <- function(x, mu, beta) {
  (1 / beta) * exp(-(x - mu) / beta) * exp(-exp(-(x - mu) / beta))
}

# Definir la función de distribución acumulada de la distribución Gumbel
(CDF)
pGumbel <- function(x, mu, beta) {
  exp(-exp(-(x - mu) / beta))
}
```

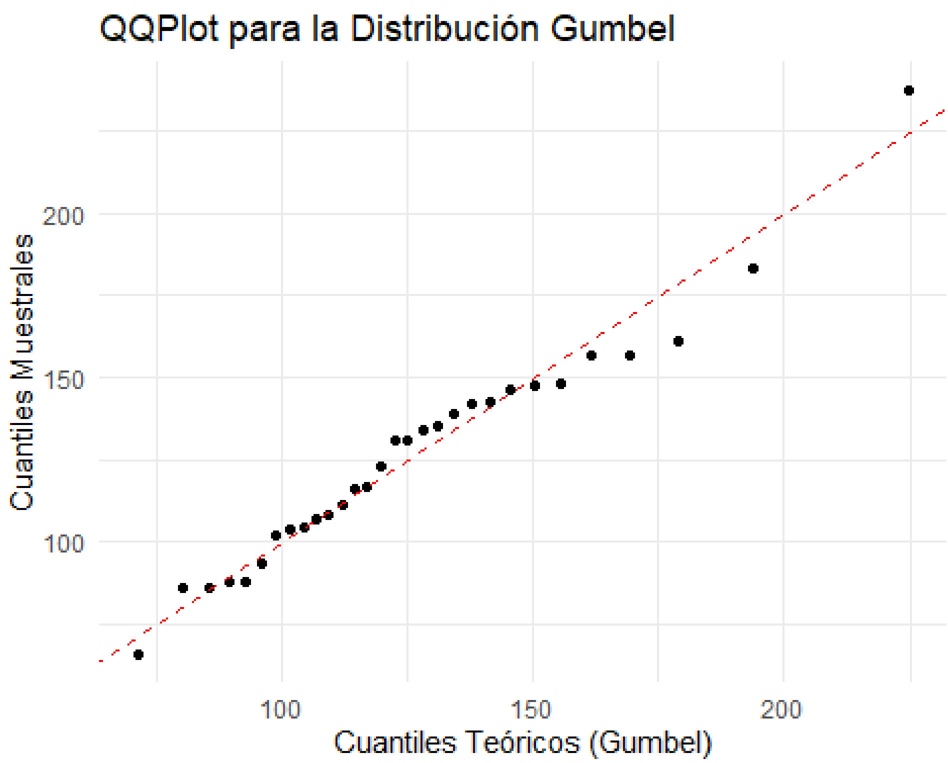
```r
# Ejemplo de cómo usar estas funciones para evaluar la densidad o la probabilidad acumulada
mu_gumbel <- 100  # Ejemplo de valor para el parámetro de localización
beta_gumbel <- 20  # Ejemplo de valor para el parámetro de escala

# Calcular la densidad para un valor específico (por ejemplo, x = 120)
densidad <- dGumbel(120, mu_gumbel, beta_gumbel)
print(paste("Densidad Gumbel en x = 120:", densidad))

## [1] "Densidad Gumbel en x = 120: 0.0127323190021791"

# Calcular la probabilidad acumulada para un valor específico (por ejemplo, x = 120)
prob_acumulada <- pGumbel(120, mu_gumbel, beta_gumbel)
print(paste("Probabilidad acumulada Gumbel en x = 120:", prob_acumulada))

## [1] "Probabilidad acumulada Gumbel en x = 120: 0.692200627555346"
```

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.

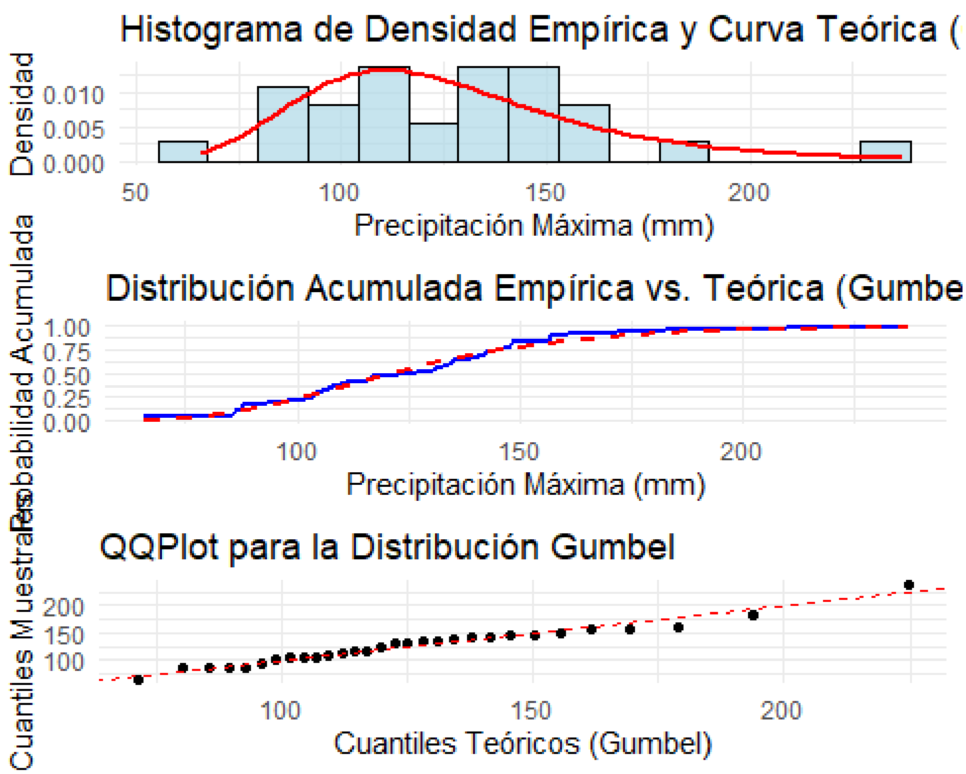
```r
# Cargar las librerías necesarias
library(fitdistrplus)

## Warning: package 'fitdistrplus' was built under R version 4.3.3

## Loading required package: survival

library(ggplot2)
library(gridExtra)

## Warning: package 'gridExtra' was built under R version 4.3.2

##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine

# Definir la función de densidad de la distribución Gumbel (PDF)
dgumbel <- function(x, mu, beta) {
  (1 / beta) * exp(-(x - mu) / beta) * exp(-exp(-(x - mu) / beta))
}

# Función para ajustar la distribución Gumbel
fit_gumbel <- function(data) {
  log_likelihood <- function(params) {
    mu <- params[1]
    beta <- params[2]
    -sum(log(dgumbel(data, mu, beta)))
```

```r
  }
  fit <- optim(c(mean(data), sd(data)), log_likelihood, hessian = TRUE)
  return(list(estimate = fit$par))
}

# Ajustar los parámetros de la Distribución Gumbel usando la función
definida
ajuste_gumbel <- fit_gumbel(datos_agg_ordenado$Lluvia_max_anual)

# Parámetros ajustados
mu_gumbel <- ajuste_gumbel$estimate[1]
beta_gumbel <- ajuste_gumbel$estimate[2]

# Crear el histograma de densidad empírica y curva teórica de la Gumbel
hist_gumbel <- ggplot(datos_agg_ordenado, aes(x = Lluvia_max_anual)) +
  geom_histogram(aes(y = ..density..), bins = 15, color = "black", fill =
"lightblue", alpha = 0.7) +
  stat_function(fun = dgumbel, args = list(mu = mu_gumbel, beta =
beta_gumbel), color = "red", size = 1) +
  labs(title = "Histograma de Densidad Empírica y Curva Teórica
(Gumbel)",
       x = "Precipitación Máxima (mm)", y = "Densidad") +
  theme_minimal()

# Crear la ojiva empírica vs teórica
empirical_cdf <- ecdf(datos_agg_ordenado$Lluvia_max_anual)
cdf_data <- data.frame(
  x = datos_agg_ordenado$Lluvia_max_anual,
  emp_cdf = empirical_cdf(datos_agg_ordenado$Lluvia_max_anual),
  teor_cdf = pGumbel(datos_agg_ordenado$Lluvia_max_anual, mu_gumbel,
beta_gumbel)
)

cdf_gumbel <- ggplot(cdf_data, aes(x = x)) +
  geom_line(aes(y = emp_cdf), color = "blue", size = 1) +
  stat_function(fun = pGumbel, args = list(mu = mu_gumbel, beta =
beta_gumbel), color = "red", linetype = "dashed", size = 1) +
  labs(title = "Distribución Acumulada Empírica vs. Teórica (Gumbel)",
       x = "Precipitación Máxima (mm)", y = "Probabilidad Acumulada") +
  theme_minimal()

# Definir la función de cuantiles de la distribución Gumbel (inversa de
la CDF)
qGumbel <- function(p, mu, beta) {
  mu - beta * log(-log(p))
}

# QQPlot ajustado
qq_gumbel <- ggplot() +
```

```r
  stat_qq(aes(sample = datos_agg_ordenado$Lluvia_max_anual), distribution
= qGumbel, dparams = list(mu = mu_gumbel, beta = beta_gumbel)) +
  geom_abline(intercept = 0, slope = 1, color = "red", linetype =
"dashed") +
  labs(title = "QQPlot para la Distribución Gumbel",
       x = "Cuantiles Teóricos (Gumbel)", y = "Cuantiles Muestrales") +
  theme_minimal()

# Mostrar el QQPlot
print(qq_gumbel)
```




```r
# Mostrar las gráficas juntas
grid.arrange(hist_gumbel, cdf_gumbel, qq_gumbel, nrow = 3)
```



Ajuste General: En general, las tres gráficas sugieren que la distribución Gumbel proporciona un ajuste razonablemente bueno para los datos de precipitación máxima. Las curvas teóricas y empíricas están bastante alineadas, y el QQPlot también muestra una buena correspondencia en la mayoría de los valores.

Discrepancias en las Colas: Sin embargo, hay discrepancias leves, especialmente en las colas de la distribución, donde los datos empíricos tienen más variabilidad de lo que la distribución Gumbel predice. Esto podría indicar que la Gumbel subestima o sobreestima la frecuencia de valores extremos (muy altos o muy bajos).

Consideración: Aunque la Gumbel parece ser una buena opción para modelar estos datos, sería beneficioso explorar otras distribuciones o aplicar transformaciones adicionales si se desea un ajuste más preciso en las colas.

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

```r
# Definir la función de distribución acumulada (CDF) de la distribución
Gumbel
pGumbel <- function(x, mu, beta) {
  exp(-exp(-(x - mu) / beta))
}

# Realizar la prueba KS para la distribución Gumbel
ks_test_gumbel <- ks.test(
  datos_agg_ordenado$Lluvia_max_anual,
  "pGumbel",
  mu = mu_gumbel,
```

```
  beta = beta_gumbel
)

# Mostrar el resultado de la prueba KS
print(ks_test_gumbel)

##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  datos_agg_ordenado$Lluvia_max_anual
## D = 0.11247, p-value = 0.8022
## 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?

¿Qué información nos da la prueba KS para una Gumbel?

La prueba de Kolmogorov-Smirnov (KS) compara la distribución acumulada empírica de los datos con la distribución acumulada teórica de una distribución Gumbel ajustada. La prueba nos dice si hay diferencias significativas entre las dos distribuciones. En términos generales, la prueba KS mide la máxima diferencia absoluta entre las funciones de distribución empírica y teórica. ¿Cuál es el valor del estadístico de prueba?

El estadístico de prueba D = 0.11247. Este valor representa la máxima diferencia observada entre la CDF empírica y la CDF teórica de la Gumbel. Un valor más bajo indica un mejor ajuste.

¿Cuál es el p-value de la prueba?

El p-value = 0.8022. Este valor indica la probabilidad de observar una diferencia igual o mayor a la medida por el estadístico D, bajo la hipótesis de que los datos realmente siguen una distribución Gumbel.

¿Se acepta o se rechaza la hipótesis nula?

La hipótesis nula de la prueba KS es que los datos siguen la distribución Gumbel. Dado que el p-value (0.8022) es mayor que el nivel de significancia común (0.05), no rechazamos la hipótesis nula. Esto significa que no se encontró evidencia significativa para concluir que los datos no se ajustan a una distribución Gumbel.

¿Podemos concluir que las probabilidades de excedencia de las precipitaciones máximas mensuales siguen una distribución Gumbel? ¿Por qué?

Sí, podemos concluir que las probabilidades de excedencia de las precipitaciones máximas mensuales siguen una distribución Gumbel.

Esto se debe a que la prueba KS no encontró diferencias significativas entre la distribución empírica de los datos y la distribución teórica de Gumbel. El p-value alto (0.8022) indica que cualquier diferencia observada es muy probablemente debida al azar y no a un mal ajuste del modelo.

Además, el estadístico D (0.11247) es relativamente bajo, lo que sugiere que las diferencias entre las CDF empírica y teórica no son grandes.

 # ¿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?

Número de Parámetros:

La distribución Gumbel tiene dos parámetros: - μ: el parámetro de localización, que determina el centro de la distribución. - β: el parámetro de escala, que controla la dispersión de los valores alrededor del centro.

Estimación de Parámetros Usando Media y Desviación Estándar:

Para estimar los parámetros de la Gumbel a partir de la media (m) y la desviación estándar (s) de los datos, podemos usar las siguientes fórmulas:

Media Teórica de la Gumbel ($\mu + \beta\gamma$):

$$m = \mu + \beta \cdot \gamma$$

donde γ es la constante de Euler-Mascheroni (≈ 0.5772).

Desviación Estándar Teórica de la Gumbel $\left(\frac{\pi\beta}{\sqrt{6}}\right)$:

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

Podemos despejar estas ecuaciones para obtener:

Parámetro de Escala β:

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

Parámetro de Localización μ:

$$\mu = m - \beta \cdot \gamma$$

```
# Definir la función de densidad de la Gumbel
dgumbel <- function(x, mu, beta) {
```

```r
  (1 / beta) * exp(-(x - mu) / beta) * exp(-exp(-(x - mu) / beta))
}

# Función de log-verosimilitud para la Gumbel
log_likelihood_gumbel <- function(params, data) {
  mu <- params[1]
  beta <- params[2]
  -sum(log(dgumbel(data, mu, beta)))
}

# Definir los datos de precipitación máxima
datos <- datos_agg_ordenado$Lluvia_max_anual

# Estimación inicial basada en media y desviación estándar
mu_est <- mean(datos) - (sd(datos) * sqrt(6) / pi * 0.5772)
beta_est <- sd(datos) * sqrt(6) / pi

# Usar optim para estimar los parámetros de máxima verosimilitud
mle_ajuste <- optim(
  par = c(mu_est, beta_est),
  fn = log_likelihood_gumbel,
  data = datos,
  method = "L-BFGS-B",
  lower = c(-Inf, 0.001) # asegurar que beta sea positivo
)

# Obtener los resultados de máxima verosimilitud
mu_mle <- mle_ajuste$par[1]
beta_mle <- mle_ajuste$par[2]

cat("Estimación con Máxima Verosimilitud (MLE):\n")

## Estimación con Máxima Verosimilitud (MLE):

cat("Mu (localización):", mu_mle, "\n")

## Mu (localización): 110.8037

cat("Beta (escala):", beta_mle, "\n")

## Beta (escala): 27.90162

# Parámetros basados en media y desviación estándar
gamma_euler <- 0.5772
mu_teorico <- mean(datos) - (beta_est * gamma_euler)
beta_teorico <- beta_est

cat("Estimación basada en media y desviación estándar:\n")

## Estimación basada en media y desviación estándar:
```

```r
cat("Mu (localización):", mu_teorico, "\n")
```

```
## Mu (localización): 110.8719
```

```r
cat("Beta (escala):", beta_teorico, "\n")
```

```
## Beta (escala): 26.87931
```

```r
# Comparación de resultados
cat("\nComparación de Resultados:\n")
```

```
##
## Comparación de Resultados:
```

```r
cat("Método de Momentos vs. Máxima Verosimilitud:\n")
```

```
## Método de Momentos vs. Máxima Verosimilitud:
```

```r
cat("Mu (Momentos):", mu_teorico, "vs. Mu (MLE):", mu_mle, "\n")
```

```
## Mu (Momentos): 110.8719 vs. Mu (MLE): 110.8037
```

```r
cat("Beta (Momentos):", beta_teorico, "vs. Beta (MLE):", beta_mle, "\n")
```

```
## Beta (Momentos): 26.87931 vs. Beta (MLE): 27.90162
```

¿Por qué Puede Haber Diferencias?

Máxima Verosimilitud vs. Método de Momentos: Los métodos que usan diferentes enfoques (máxima verosimilitud frente a basarse en media y desviación estándar) pueden producir estimaciones ligeramente diferentes. La verosimilitud tiende a ser más precisa porque aprovecha toda la información de los datos para ajustar el modelo.

Comportamiento de los Datos: Los valores extremos o la forma de las distribuciones reales pueden llevar a que los métodos produzcan estimaciones que sean más representativas del comportamiento de los datos observados que lo que se logra con aproximaciones teóricas simples.

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.

```r
# Cargar las librerías necesarias
library(ggplot2)
library(MASS)
library(fitdistrplus)
library(evd)
```

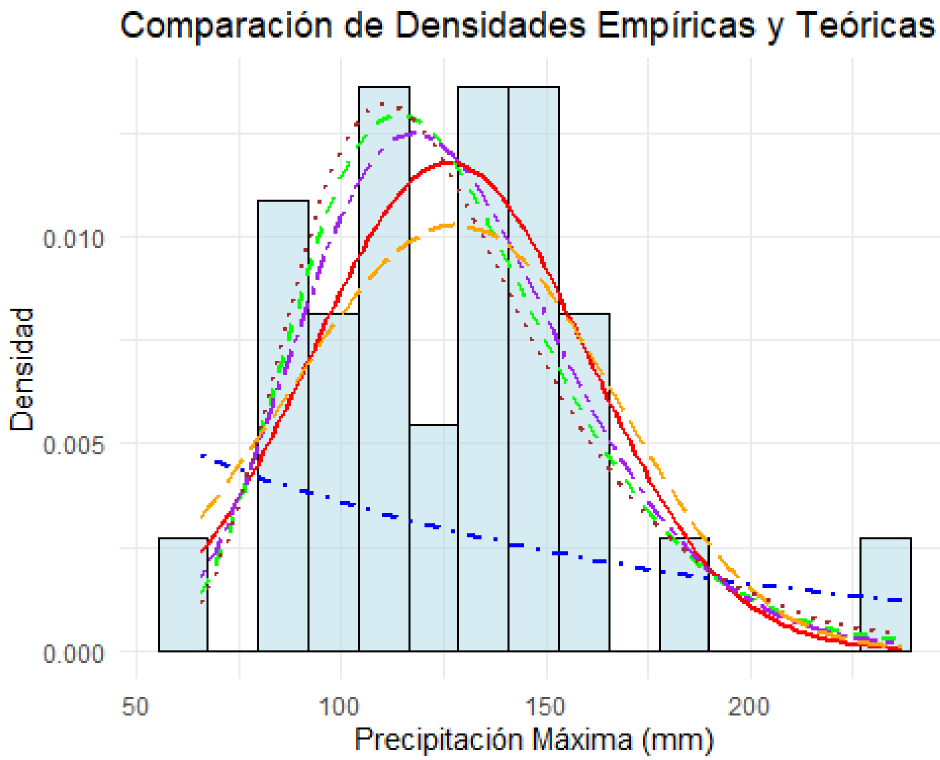
```
## Warning: package 'evd' was built under R version 4.3.3
```

```
##
## Attaching package: 'evd'

## The following object is masked _by_ '.GlobalEnv':
##
##     dgumbel

# Definir los datos
datos <- datos_agg_ordenado$Lluvia_max_anual

# Ajustar los parámetros de cada distribución
# Distribución Normal
params_normal <- fitdistr(datos, "normal")
mu_normal <- params_normal$estimate[1]
sigma_normal <- params_normal$estimate[2]

# Distribución Log-Normal
params_lognormal <- fitdistr(datos, "lognormal")
mu_lognormal <- params_lognormal$estimate[1]
sigma_lognormal <- params_lognormal$estimate[2]

# Distribución Exponencial
params_exponencial <- fitdistr(datos, "exponential")
lambda_exp <- params_exponencial$estimate[1]

# Distribución Gamma
params_gamma <- fitdistr(datos, "gamma")
shape_gamma <- params_gamma$estimate[1]
rate_gamma <- params_gamma$estimate[2]

# Distribución Weibull
params_weibull <- fitdistr(datos, "weibull")
shape_weibull <- params_weibull$estimate[1]
scale_weibull <- params_weibull$estimate[2]

# Distribución Gumbel (usando evd)
params_gumbel <- fgev(datos, shape = 0)
mu_gumbel <- params_gumbel$estimate[1]
beta_gumbel <- params_gumbel$estimate[2]

# Definir las funciones de densidad para cada distribución
dnorm_custom <- function(x) { dnorm(x, mean = mu_normal, sd =
sigma_normal) }
dlnorm_custom <- function(x) { dlnorm(x, meanlog = mu_lognormal, sdlog =
sigma_lognormal) }
dexp_custom <- function(x) { dexp(x, rate = lambda_exp) }
dgamma_custom <- function(x) { dgamma(x, shape = shape_gamma, rate =
rate_gamma) }
dweibull_custom <- function(x) { dweibull(x, shape = shape_weibull, scale
= scale_weibull) }
```

```r
dgumbel_custom <- function(x, mu, beta) {
  (1 / beta) * exp(-(x - mu) / beta) * exp(-exp(-(x - mu) / beta))
}

# Paso 2: Crear el Gráfico Comparativo con ggplot2
ggplot(data.frame(x = datos), aes(x = x)) +
  geom_histogram(aes(y = ..density..), bins = 15, fill = "lightblue",
color = "black", alpha = 0.5) +
  stat_function(fun = dnorm_custom, color = "red", size = 1, linetype =
"solid", aes(label = "Normal")) +
  stat_function(fun = dlnorm_custom, color = "green", size = 1, linetype
= "dashed", aes(label = "Log-Normal")) +
  stat_function(fun = dexp_custom, color = "blue", size = 1, linetype =
"dotdash", aes(label = "Exponencial")) +
  stat_function(fun = dgamma_custom, color = "purple", size = 1, linetype
= "twodash", aes(label = "Gamma")) +
  stat_function(fun = dweibull_custom, color = "orange", size = 1,
linetype = "longdash", aes(label = "Weibull")) +
  stat_function(fun = function(x) dgumbel_custom(x, mu_gumbel,
beta_gumbel), color = "brown", size = 1, linetype = "dotted", aes(label =
"Gumbel")) +
  labs(title = "Comparación de Densidades Empíricas y Teóricas",
       x = "Precipitación Máxima (mm)",
       y = "Densidad",
       color = "Distribución") +
  theme_minimal() +
  scale_color_manual(values = c("red", "green", "blue", "purple",
"orange", "brown")) +
  theme(legend.position = "top")
```

```
## Warning in stat_function(fun = dnorm_custom, color = "red", size = 1,
linetype
## = "solid", : Ignoring unknown aesthetics: label

## Warning in stat_function(fun = dlnorm_custom, color = "green", size =
1, :
## Ignoring unknown aesthetics: label

## Warning in stat_function(fun = dexp_custom, color = "blue", size = 1,
linetype
## = "dotdash", : Ignoring unknown aesthetics: label

## Warning in stat_function(fun = dgamma_custom, color = "purple", size =
1, :
## Ignoring unknown aesthetics: label

## Warning in stat_function(fun = dweibull_custom, color = "orange", size
= 1, :
## Ignoring unknown aesthetics: label
```

```
## Warning in stat_function(fun = function(x) dgumbel_custom(x,
mu_gumbel, :
## Ignoring unknown aesthetics: label
```



Ajustes Centrales vs. Colas: La distribución gamma y la log-normal parecen ajustarse razonablemente bien a los datos en la parte central, mientras que otras, como la normal y la exponencial, presentan desviaciones notables.

Modelado de Colas: Distribuciones como la Gumbel y la Weibull pueden ayudar a modelar mejor las colas, especialmente si hay eventos extremos de interés (por ejemplo, precipitaciones muy altas).

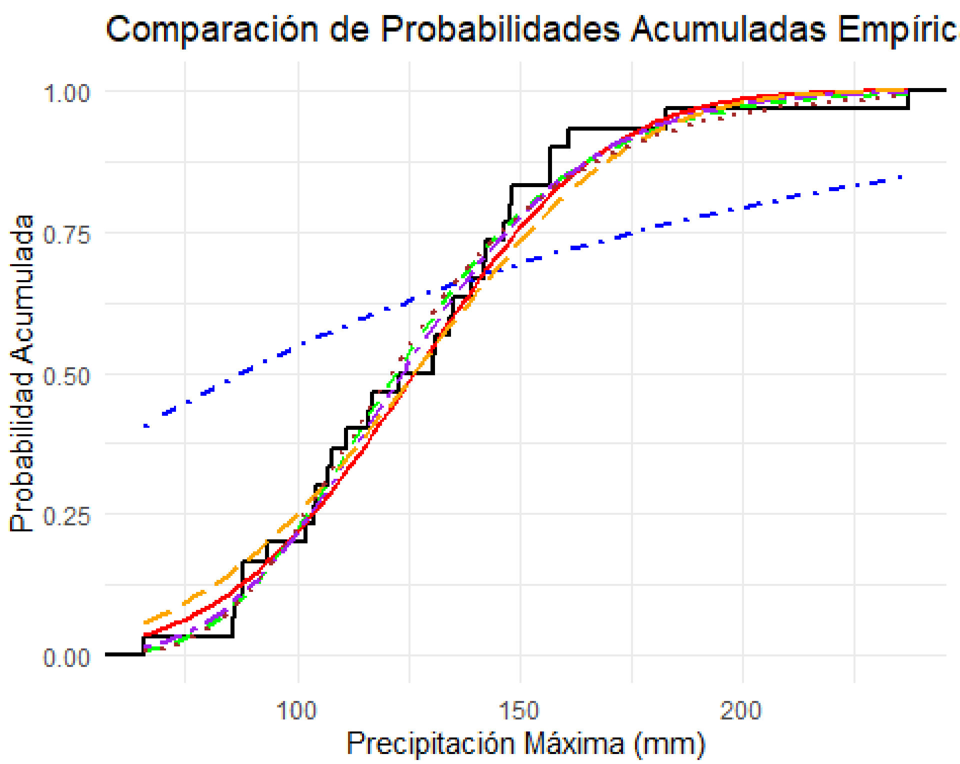
Recomendación: Basado en esta comparación visual, se podría recomendar profundizar en análisis con la gamma y la log-normal, ya que muestran un mejor ajuste general a los datos empíricos. Las otras distribuciones podrían ser útiles en escenarios específicos donde es importante capturar eventos extremos o variabilidad distinta.

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

```r
# Cargar las librerías necesarias
library(ggplot2)
library(MASS)
library(fitdistrplus)
library(evd)
```

```r
# Definir los datos
datos <- datos_agg_ordenado$Lluvia_max_anual

# Ajustar los parámetros de cada distribución
params_normal <- fitdistr(datos, "normal")
params_lognormal <- fitdistr(datos, "lognormal")
params_exponencial <- fitdistr(datos, "exponential")
params_gamma <- fitdistr(datos, "gamma")
params_weibull <- fitdistr(datos, "weibull")
params_gumbel <- fgev(datos, shape = 0)

# Obtener los parámetros ajustados
mu_normal <- params_normal$estimate[1]
sigma_normal <- params_normal$estimate[2]
mu_lognormal <- params_lognormal$estimate[1]
sigma_lognormal <- params_lognormal$estimate[2]
lambda_exp <- params_exponencial$estimate[1]
shape_gamma <- params_gamma$estimate[1]
rate_gamma <- params_gamma$estimate[2]
shape_weibull <- params_weibull$estimate[1]
scale_weibull <- params_weibull$estimate[2]
mu_gumbel <- params_gumbel$estimate[1]
beta_gumbel <- params_gumbel$estimate[2]

# Definir funciones de distribución acumulada para cada distribución
pnorm_custom <- function(x) { pnorm(x, mean = mu_normal, sd =
sigma_normal) }
plnorm_custom <- function(x) { plnorm(x, meanlog = mu_lognormal, sdlog =
sigma_lognormal) }
pexp_custom <- function(x) { pexp(x, rate = lambda_exp) }
pgamma_custom <- function(x) { pgamma(x, shape = shape_gamma, rate =
rate_gamma) }
pweibull_custom <- function(x) { pweibull(x, shape = shape_weibull, scale
= scale_weibull) }
pgumbel_custom <- function(x, mu, beta) { exp(-exp(-(x - mu) / beta)) }

# Crear un data frame con las probabilidades acumuladas empíricas
datos_empiricos <- ecdf(datos)

# Paso 2: Gráfico Comparativo con ggplot2
ggplot(data.frame(x = sort(datos)), aes(x = x)) +
  stat_ecdf(geom = "step", aes(y = datos_empiricos(x)), color = "black",
size = 1, linetype = "solid") +
  stat_function(fun = pnorm_custom, color = "red", size = 1, linetype =
"solid", aes(label = "Normal")) +
  stat_function(fun = plnorm_custom, color = "green", size = 1, linetype
= "dashed", aes(label = "Log-Normal")) +
  stat_function(fun = pexp_custom, color = "blue", size = 1, linetype =
```

```
"dotdash", aes(label = "Exponencial")) +
  stat_function(fun = pgamma_custom, color = "purple", size = 1, linetype
= "twodash", aes(label = "Gamma")) +
  stat_function(fun = pweibull_custom, color = "orange", size = 1,
linetype = "longdash", aes(label = "Weibull")) +
  stat_function(fun = function(x) pgumbel_custom(x, mu_gumbel,
beta_gumbel), color = "brown", size = 1, linetype = "dotted", aes(label =
"Gumbel")) +
  labs(title = "Comparación de Probabilidades Acumuladas Empírica vs.
Teóricas",
       x = "Precipitación Máxima (mm)",
       y = "Probabilidad Acumulada",
       color = "Distribución") +
  theme_minimal() +
  scale_color_manual(values = c("black", "red", "green", "blue",
"purple", "orange", "brown")) +
  theme(legend.position = "top")

## Warning in stat_function(fun = pnorm_custom, color = "red", size = 1,
linetype
## = "solid", : Ignoring unknown aesthetics: label

## Warning in stat_function(fun = plnorm_custom, color = "green", size =
1, :
## Ignoring unknown aesthetics: label

## Warning in stat_function(fun = pexp_custom, color = "blue", size = 1,
linetype
## = "dotdash", : Ignoring unknown aesthetics: label

## Warning in stat_function(fun = pgamma_custom, color = "purple", size =
1, :
## Ignoring unknown aesthetics: label

## Warning in stat_function(fun = pweibull_custom, color = "orange", size
= 1, :
## Ignoring unknown aesthetics: label

## Warning in stat_function(fun = function(x) pgumbel_custom(x,
mu_gumbel, :
## Ignoring unknown aesthetics: label
```



Mejores Ajustes: Las distribuciones gamma, log-normal y Weibull parecen ser las más adecuadas para modelar estos datos, ya que sus curvas de probabilidad acumulada siguen de cerca a la empírica.

Menos Adecuadas: La exponencial es la que presenta el peor ajuste, quedando muy por debajo de la curva empírica y no logrando capturar la variabilidad de los datos.

Utilidad del Análisis: Este tipo de análisis es útil para identificar cuál de las distribuciones puede representar mejor los datos reales de precipitación, y ayuda a elegir un modelo adecuado para previsiones y estudios futuros.

 # 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.

Basado en la combinación de la comparación visual y los resultados de las pruebas KS, las distribuciones que mejor se ajustan a los datos de precipitación máxima son:

Distribución Gamma Distribución Log-Normal

Ambas distribuciones mostraron una fuerte concordancia con los datos empíricos tanto en los gráficos de densidad como en los gráficos de probabilidad acumulada, y las pruebas de bondad de ajuste respaldan estos hallazgos. La distribución Weibull también es adecuada, pero no ofrece el mismo nivel de ajuste que la Gamma y la Log-Normal.

Finalmente, aunque la Gumbel puede ser útil para modelar valores extremos, no captura tan bien la parte central y baja de los datos como la Gamma y la Log-Normal. Por otro lado, la Exponencial es definitivamente inapropiada para estos datos, dado su ajuste deficiente y el resultado de las pruebas de bondad de ajuste.

En resumen, la Distribución Gamma y la Log-Normal son las mejores opciones para modelar los datos de precipitación máxima, con una ligera ventaja para la Gamma debido a su versatilidad y capacidad para modelar diferentes formas de distribución.

DISEÑO DE OBRAS HIDRAULICAS

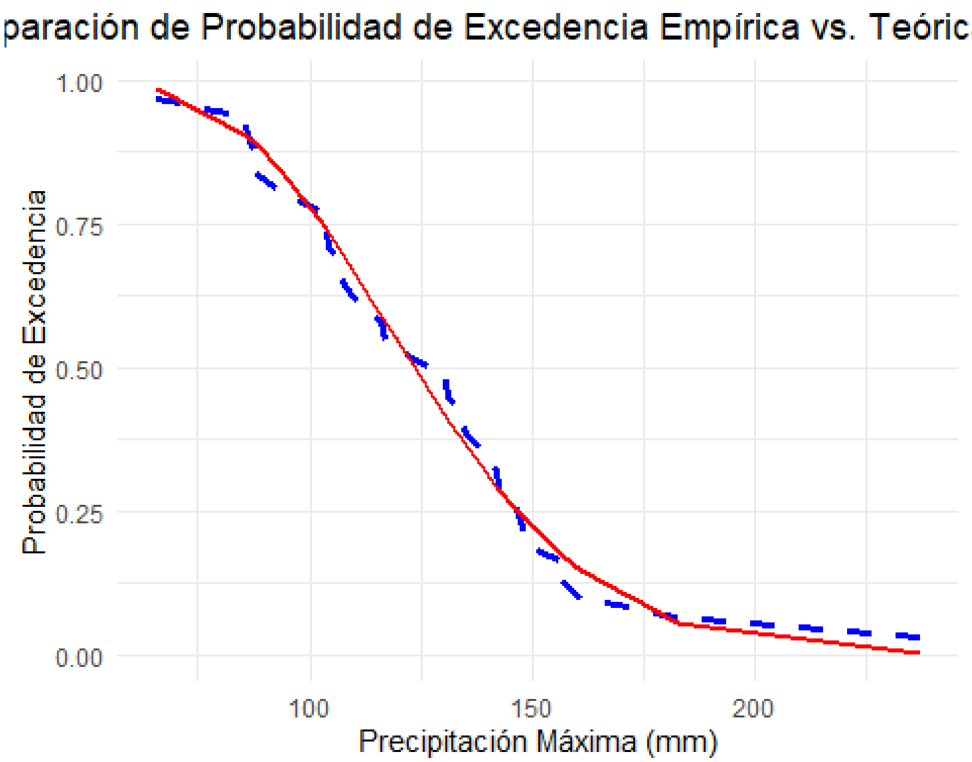
4. Precipitación de diseño de obras hidráulicas

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.

 # 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.

```
# Cargar librerías necesarias
library(ggplot2)
library(fitdistrplus)

# Datos de precipitación máxima (ajusta según sea necesario)
datos <- datos_agg_ordenado$Lluvia_max_anual

# Ajuste de la distribución Gamma
ajuste_gamma <- fitdist(datos, "gamma")

# Obtener los parámetros estimados
shape_est <- ajuste_gamma$estimate["shape"]
rate_est <- ajuste_gamma$estimate["rate"]

# Calcular la probabilidad de excedencia empírica
datos_ordenados <- sort(datos, decreasing = TRUE)
prob_excedencia_empirica <- rank(-datos_ordenados) / (length(datos) + 1)

# Calcular la probabilidad de excedencia teórica usando la distribución
Gamma
prob_excedencia_teorica <- 1 - pgamma(datos_ordenados, shape = shape_est,
rate = rate_est)

# Crear un DataFrame para el gráfico
df_excedencia <- data.frame(
  Precipitacion = datos_ordenados,
  Empirica = prob_excedencia_empirica,
  Teorica = prob_excedencia_teorica
)
```

```r
# Gráfico comparativo de probabilidad de excedencia
ggplot(df_excedencia, aes(x = Precipitacion)) +
  geom_line(aes(y = Empirica), color = "blue", linetype = "dashed", size
= 1.2) +
  geom_line(aes(y = Teorica), color = "red", size = 1) +
  labs(
    title = "Comparación de Probabilidad de Excedencia Empírica vs.
Teórica (Gamma)",
    x = "Precipitación Máxima (mm)",
    y = "Probabilidad de Excedencia"
  ) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```



El gráfico de probabilidad de excedencia confirma que la distribución Gamma es una elección adecuada para modelar los datos de precipitación máxima. Las líneas teórica y empírica muestran una buena concordancia, lo que valida el uso de esta distribución para análisis predictivos y evaluación de riesgos hidrológicos. La certeza en la elección se respalda además por las pruebas de ajuste de curva anteriores que mostraron buenos resultados para la Gamma.

 # Utilizando el límite inferior del intervalo de periodo de retorno sugerido, encuentra la probabilidad de excedencia o de ocurrencia para ese valor. Recuerda que:

```r
# Definir el periodo de retorno sugerido (en años)
P_ret <- 100

# Calcular la probabilidad de excedencia
P_exe <- 1 / P_ret

# Mostrar el resultado
cat("La probabilidad de excedencia para un periodo de retorno de", P_ret,
"años es:", P_exe, "\n")

## La probabilidad de excedencia para un periodo de retorno de 100 años
es: 0.01
```

Para calcular la probabilidad de excedencia usando el período de retorno para una zona de riego mediana (100 a 500 años), primero seleccionaremos el valor más bajo del rango (100 años) para obtener una probabilidad conservadora.

 # 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.

```r
# Ajustar la distribución Gamma usando los datos de precipitación máxima
library(MASS)
datos <- datos_agg_ordenado$Lluvia_max_anual
gamma_fit <- fitdistr(datos, "gamma")

# Parámetros ajustados
shape_gamma <- gamma_fit$estimate["shape"]
rate_gamma <- gamma_fit$estimate["rate"]

# Período de retorno y cálculo de probabilidades
Tr_median <- 100
P_exe_median <- 1 / Tr_median
P_noexe_median <- 1 - P_exe_median

# Calcular el valor de precipitación máxima para ese periodo de retorno
precip_max_median <- qgamma(p = P_noexe_median, shape = shape_gamma, rate
= rate_gamma)

# Mostrar el resultado
precip_max_median

## [1] 215.2766
```

Este código ajusta la distribución Gamma a tus datos y luego calcula la precipitación máxima mensual para un periodo de retorno de 100 años

 # El resultado de este ejemplo será una aproximación del caudal máximo que se tendría en Sonora 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.

```
# Define los periodos de retorno que deseas probar
periodos_retorno <- c(50, 100, 200, 500)

# Lista para almacenar los resultados de precipitación máxima para cada
periodo
resultados_precipitacion <- list()

# Parámetros de la distribución Gamma ajustados previamente
shape_gamma <-gamma_fit$estimate["shape"]  # Sustituir por el parámetro
real obtenido
rate_gamma <- gamma_fit$estimate["rate"]  # Sustituir por el parámetro
real obtenido

# Calcular precipitación máxima para cada periodo de retorno
for (Tr in periodos_retorno) {
  Pexe <- 1 / Tr  # Probabilidad de excedencia
  Pnoexe <- 1 - Pexe  # Complemento de la probabilidad de excedencia
  precip_max <- qgamma(Pnoexe, shape = shape_gamma, rate = rate_gamma)
  resultados_precipitacion[[paste("Periodo de retorno", Tr, "años")]] <-
precip_max
}

# Mostrar los resultados
print(resultados_precipitacion)

## $`Periodo de retorno 50 años`
## [1] 202.7921
##
## $`Periodo de retorno 100 años`
## [1] 215.2766
##
## $`Periodo de retorno 200 años`
## [1] 227.1224
##
## $`Periodo de retorno 500 años`
## [1] 242.0217
```

Significado del valor: El valor calculado representa la precipitación máxima mensual esperada en Sonora con un periodo de retorno de 100 años. Esto significa que, en promedio, este evento de precipitación tiene una probabilidad de ocurrir una vez cada 100 años.

Incrementar el periodo de retorno: Si incrementamos el periodo de retorno, el valor del caudal máximo también aumenta. Esto se debe a que estamos considerando eventos más raros (menos frecuentes pero más extremos).

Caudal máximo en otro estado: No sería el mismo si usamos datos históricos de otro estado, ya que las características climáticas y geográficas afectan la distribución y magnitud de las precipitaciones. Por ejemplo, un estado con un clima más húmedo podría tener valores de caudal máximo más altos para el mismo periodo de retorno.

Diseño de obras hidráulicas basado en periodos de retorno: Es crucial porque permite diseñar estructuras que puedan soportar eventos extremos sin fallar. Un diseño basado en periodos de retorno inadecuados podría llevar a fallos catastróficos.

Importancia de conocer la distribución de probabilidad: Ayuda a estimar correctamente la magnitud de eventos extremos y planificar con mayor precisión las estructuras hidráulicas. Si se usa una distribución incorrecta, las estimaciones de caudales podrían ser inexactas, comprometiendo la seguridad y eficiencia de las infraestructuras.

Mayor Periodo de Retorno: A medida que aumentas el periodo de retorno, la precipitación máxima esperada también incrementa. Esto significa que eventos de mayor magnitud son más raros.