Actividad Integradora 1

Catherine Rojas

2024-10-22

```
# Ocultar todo el código
knitr::opts_chunk$set(echo = TRUE)

# Evitar notación científica
options(scipen=999)

# Cargar la librería printr
library(printr)

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

## Registered S3 method overwritten by 'printr':
##   method                from
##   knit_print.data.frame rmarkdown
```

Precipitaciones máximas mensuales para el diseño de obras hidráulicas

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

Se analizaran los datos históricos (1994-2023) de las precipitaciones máximas mensuales por estado para cumplir el objetivo principal de este estudio que consiste en calcular la precipitación más extrema que se logra con un periodo de retorno seleccionado.

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

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

Estado seleccionado: Chihuahua

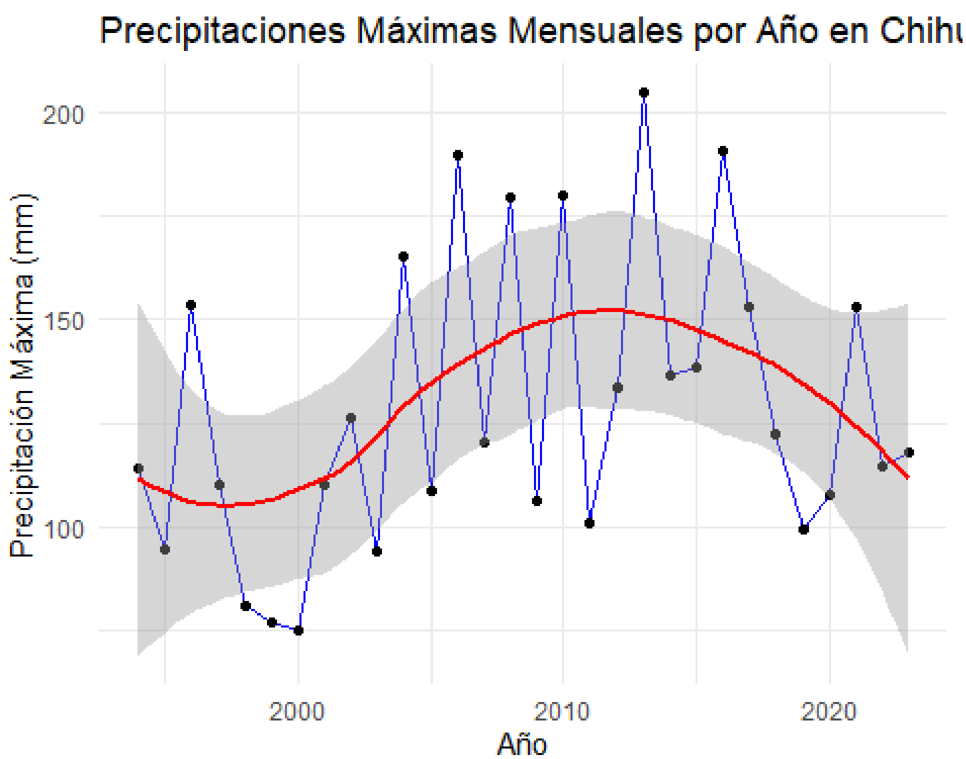
```
# Cargar archivo .txt con read.table
data <- read.table("C:/Users/PC/OneDrive - Instituto Tecnologico y de
Estudios Superiores de Monterrey/Documents/Concentracion
Estadistica/precipitaciones_maximas_mensuales.txt", header = TRUE, sep =
"\t", stringsAsFactors = FALSE)

# Mostrar las primeras filas del archivo
head(data)
```

Anio	Mes	Estado	Lluvia
1994	Ene	Aguascalientes	8.3
1994	Ene	Baja.California	10.3
1994	Ene	Baja.California.Sur	0.0
1994	Ene	Campeche	85.4
1994	Ene	Ciudad.de.México	17.7
1994	Ene	Coahuila	12.8

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.

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.3.3

# Filtrar los datos para el estado de Chihuahua
chihuahua_data <- subset(data, Estado == "Chihuahua")

# Convertir la columna 'Anio' a formato de fecha
chihuahua_data$Anio <- as.numeric(chihuahua_data$Anio)

# Calcular la precipitación mensual máxima por año
max_precipitation <- aggregate(Lluvia ~ Anio, data =  chihuahua_data,
max)

# Graficar las precipitaciones máximas por año
ggplot(max_precipitation, aes(x = Anio, y = Lluvia)) +
  geom_line(color = "blue") +
```

```
  geom_point() +
  geom_smooth(method = "loess", color = "red") +
  labs(title = "Precipitaciones Máximas Mensuales por Año en Chihuahua
(1994-2023)",
       x = "Año", y = "Precipitación Máxima (mm)") +
  theme_minimal()

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



Observaciones

- La gráfica muestra las precipitaciones máximas mensuales registradas en Chihuahua desde 1994 hasta 2023. En azul, vemos los valores de precipitación máxima anual, mientras que la línea roja suavizada con el método LOESS indica la tendencia general a lo largo de los años. La banda sombreada alrededor de la línea roja representa un intervalo de confianza que muestra la variabilidad en los datos.

- La gráfica sugiere variabilidad en las precipitaciones máximas anuales, con varios picos notables, especialmente alrededor de los años 2006, 2013 y 2016, donde las precipitaciones alcanzaron sus valores más altos (aproximadamente entre 190 y 200 mm). Sin embargo, se observa una disminución significativa de estos valores en los años más recientes.

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

C.1) Calcula las medidas de centralización y variación de las precipitaciones máximas mensuales

```
# Máximas precipitaciones mensuales
# Medidas de centralización
media <- mean(chihuahua_data$Lluvia)
mediana <- median(chihuahua_data$Lluvia)

# Medidas de variación
desviacion_estandar <- sd(chihuahua_data$Lluvia)
varianza <- var(chihuahua_data$Lluvia)
rango <- range(chihuahua_data$Lluvia)
cuartiles <- quantile(chihuahua_data$Lluvia)

# Imprimir medidas
cat("Media:", media, "\n")

## Media: 38.62361

cat("Mediana:", mediana, "\n")

## Mediana: 19.9

cat("Desviación estándar:", desviacion_estandar, "\n")

## Desviación estándar: 44.31238

cat("Varianza:", varianza, "\n")

## Varianza: 1963.587

cat("Rango:", rango[1], "-", rango[2], "\n")

## Rango: 0 - 204.9

cat("Cuartiles:\n", "\n")

## Cuartiles:
##

cuartiles

##      0%     25%     50%     75%    100%
##   0.000   4.000  19.900  66.675 204.900
```

Observaciones

Media (38.62 mm): En promedio, la precipitación máxima mensual en Chihuahua es de 38.62 mm. Esto sugiere que, a lo largo del período analizado, las precipitaciones extremas no suelen ser muy altas en la mayoría de los meses.

Mediana (19.9 mm): La mediana es significativamente más baja que la media, lo que sugiere que la distribución de los datos está sesgada hacia la derecha. Esto indica que hay meses con precipitaciones extremas muy altas que aumentan el valor de la media, mientras que la mayoría de los meses tienen precipitaciones máximas más bajas.

-
    - De acuerdo con los resultados de la **Desviación estándar (44.31 mm):**, **Varianza (1963.59 mm^2)** y el **Rango (0 - 204.9 mm):**, podemos decir que existe una gran variabilidad en las precipitaciones máximas mensuales.

```r
# Máximas precipitaciones mensuales por año.
# Medidas de centralización
media <- mean(max_precipitation$Lluvia)
mediana <- median(max_precipitation$Lluvia)

# Medidas de variación
desviacion_estandar <- sd(max_precipitation$Lluvia)
varianza <- var(max_precipitation$Lluvia)
rango <- range(max_precipitation$Lluvia)
cuartiles <- quantile(max_precipitation$Lluvia)

# Imprimir medidas
cat("Media:", media, "\n")

## Media: 128.6633

cat("Mediana:", mediana, "\n")

## Mediana: 119.3

cat("Desviación estándar:", desviacion_estandar, "\n")

## Desviación estándar: 35.48249

cat("Varianza:", varianza, "\n")

## Varianza: 1259.007

cat("Rango:", rango[1], "-", rango[2], "\n")

## Rango: 75.2 - 204.9

cat("Cuartiles:\n", "\n")

## Cuartiles:
##

cuartiles

##      0%     25%     50%     75%    100%
##  75.200 106.525 119.300 153.250 204.900
```

Observaciones

Media (128.66 mm): Este valor representa el promedio de las precipitaciones máximas mensuales durante el período analizado. Es decir, en promedio, la precipitación máxima mensual en Chihuahua fue de 128.66 mm.

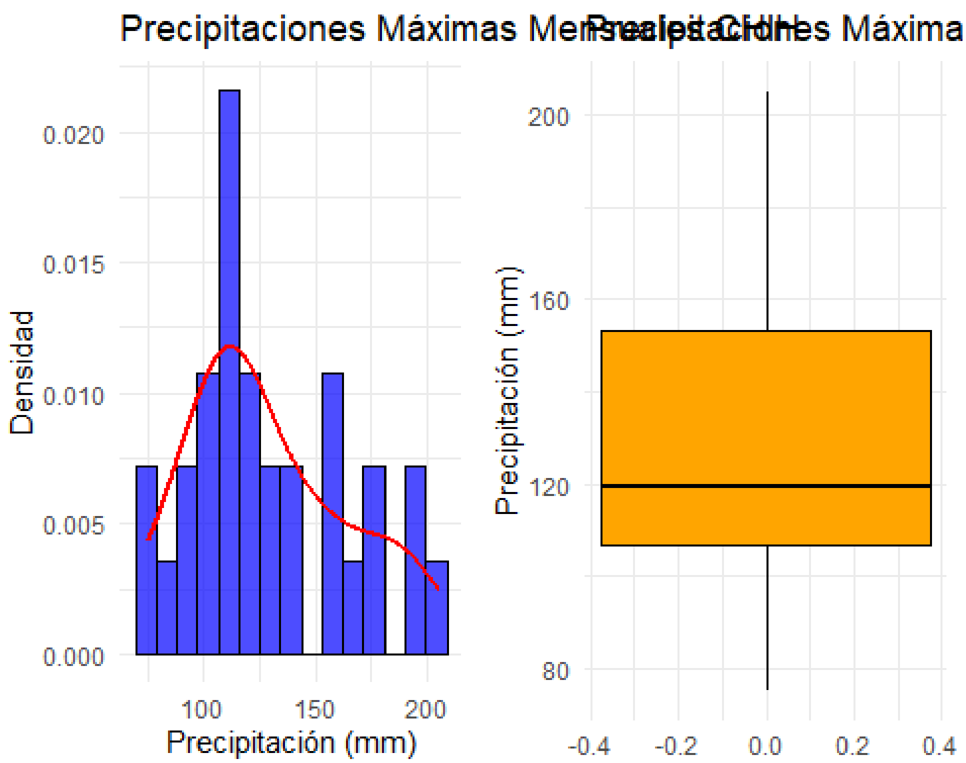
Mediana (119.3 mm): La mediana es el valor que se encuentra en el centro del conjunto de datos. La mitad de los datos están por debajo de 119.3 mm y la otra mitad por encima. Este valor es ligeramente inferior a la media, lo que podría indicar la presencia de algunos valores extremadamente altos (outliers) que incrementan la media.

- De acuerdo con los resultados de la **Desviación estándar (35.48 mm)**, **Varianza (1259.007 mm^2)** y el **Rango (75.2 mm - 204.9 mm)**, podemos decir que existe una considerable variabilidad en las precipitaciones máximas mensuales.

C.2) Realiza gráficos que te sirvan para describir la distribución de las lluviar máximas mensuales: histograma y boxplot

```
library(gridExtra)

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

# Histograma de precipitaciones máximas mensuales
histograma <- ggplot(max_precipitation, aes(x = Lluvia)) +
  geom_histogram(aes(y = ..density..), bins = 15, fill = "blue", color =
"black", alpha = 0.7) +
  geom_density(color = "red", size = 1) +
  labs(title = "Precipitaciones Máximas Mensuales CHIH",
       x = "Precipitación (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.

# Boxplot de precipitaciones máximas mensuales
boxplot <- ggplot(max_precipitation, aes(y = Lluvia)) +
  geom_boxplot(fill = "orange", color = "black") +
  labs(title = "Precipitaciones Máximas Mensuales CHIH",
       y = "Precipitación (mm)") +
  theme_minimal()

# Mostrar ambos gráficos en una sola ventana
grid.arrange(histograma, boxplot, ncol = 2)
```

```
## 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.
```



Interpretación

Histograma: El histograma muestra que las precipitaciones máximas mensuales por año se concentran principalmente entre los 110-120 mm, con una tendencia unimodal. La curva de densidad en rojo refleja esta concentración, aunque con asimetría hacia la derecha, lo que indica la presencia ocasional de precipitaciones extremas (190-200 mm). Sin embargo, la mayoría de las precipitaciones están por debajo de los 160 mm, con los valores más elevados siendo poco frecuentes.

Boxplot: El boxplot muestra la dispersión y los valores atípicos (outliers) de las precipitaciones máximas mensuales. La longitud de la caja indica que hay una variabilidad moderada en los datos, con un rango intercuartílico que va de aproximadamente 106 mm a 153 mm y no hay presencia aparente de valores atípicos significativos más allá de los bigotes.

C.3) Describe el comportamiento de la distribución: centralización, sesgo, variación, ….

Centralización: La distribución está centrada ligeramente por encima de la mediana, posiblemente debido a algunos valores elevados, esto de acuerdo a la mediana y media obtenidos previamente.

Sesgo: El histograma sugiere que la distribución tiene un sesgo positivo o sesgo hacia la derecha. Esto se debe a que la mayor parte de los datos se concentran en el lado izquierdo del gráfico (valores de precipitación más bajos) y se extienden hacia valores más altos de precipitación, con pocos eventos extremos en el rango de 190-200 mm. Esto indica que la mayoría de las precipitaciones máximas son relativamente moderadas, pero existen algunos eventos de lluvia más extrema.

Variación: El rango de los datos, como se observa en el boxplot y en el histograma, va desde aproximadamente 75 mm hasta 205 mm. La desviación estándar de 35.48 mm muestra una variación moderada en los datos. Los valores no están excesivamente dispersos, pero la diferencia entre las precipitaciones más bajas y más altas sigue siendo significativa, especialmente con algunos eventos extremos de lluvia.

Presencia de valores atípicos (outliers): El boxplot no muestra ningún valor atípico evidente, ya que no hay puntos que se extiendan más allá de los "bigotes". Esto significa que, según este gráfico, no se han identificado precipitaciones que sean excepcionalmente extremas en comparación con el resto de los datos. Sin embargo, los valores en el rango de 190-200 mm son más raros, pero no lo suficientemente inusuales como para ser clasificados como outliers.

Forma de la distribución: El histograma muestra una distribución asimétrica o sesgada hacia la derecha, con un solo pico (unimodal) cerca de 110-120 mm, lo que indica que la mayoría de las precipitaciones máximas mensuales están dentro de este rango.

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?

Tendencia general: No parece haber una tendencia clara ascendente o descendente a largo plazo. La gráfica muestra fluctuaciones considerables en las precipitaciones máximas, con años de precipitaciones muy altas seguidos por años con valores mucho menores.

Ciclos o patrones de aumento/disminución: La gráfica sugiere que hay periodos en los que las precipitaciones extremas tienden a aumentar y luego disminuir. Por ejemplo, se observan picos importantes alrededor de 1996, 2006, 2013 y 2016, con años intermedios donde las precipitaciones son menores.

No hay un patrón regular que permita afirmar con certeza que la cantidad de precipitación sube o baja en intervalos constantes de tiempo. Los periodos entre los picos son variables, lo que podría estar relacionado con factores climáticos cambiantes que afectan la precipitación.

Utilidad de las gráficas: Son útiles para analizar tendencias climáticas y evaluar el impacto del cambio climático a nivel regional o global. En un contexto más real, pueden contribuir al diseño de obras hidráulicas considerando picos de precipitación para prevenir inundaciones o daños, además se pueden realizar proyecciones para años con lluvias intensas, ayudando así en la gestión de recursos y planificación urbana.

Conclusión: La gráfica muestra fluctuaciones irregulares en las precipitaciones máximas de Chihuahua, con altos picos de precipitación en años especificos. Sin embargo, no hay una periodicidad clara. Este tipo de análisis es útil para la planificación de infraestructura y gestión de riesgos ante eventos climáticos extremos.

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.

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

```
# A) Agregar una columna con los datos de lluvias máximas ordenados de
mayor a menor
max_precipitation$Lluvia_ordenada <- sort(max_precipitation$Lluvia,
decreasing = TRUE)
```

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

- El rango es simplemente el número de orden en la lista de precipitaciones máximas. Este valor es importante para calcular las probabilidades y el periodo de retorno.

```
# B) Agregar una columna con el número de orden (rank o rango)
max_precipitation$Rank <- 1:nrow(max_precipitation)
```

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

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

- La probabilidad de excedencia representa la probabilidad de que ocurra una precipitación mayor o igual a un valor determinado en un año.

```
# C) Calcular la probabilidad de excedencia según Weibull
N <- nrow(max_precipitation)  # Total de datos
max_precipitation$P_exce <- max_precipitation$Rank / (N + 1)
```

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

$$P_{no\ exce} = 1 - P_{exe}$$

- El complemento de la probabilidad de excedencia se interpreta como la probabilidad de que una precipitación sea menor que el valor máximo considerado.

```r
# D) Calcular la probabilidad de no excedencia (complemento de P_exce)
max_precipitation$P_no_exce <- 1 - max_precipitation$P_exce
```

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

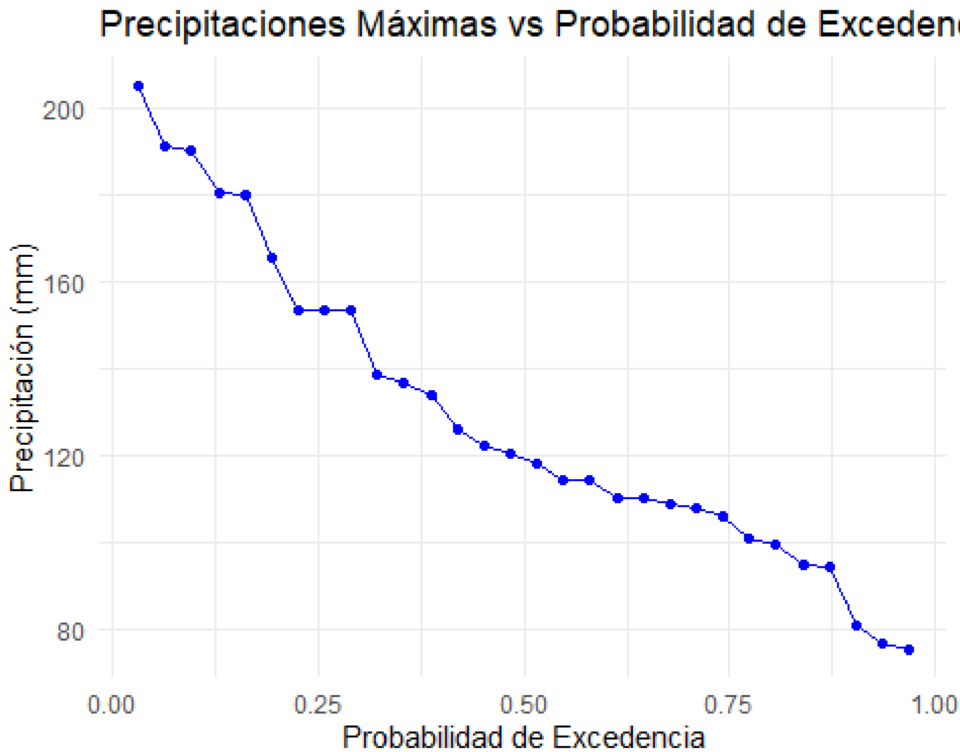
$$P_{ret} = \frac{1}{P_{exe}}$$

- Permite estimar cuántos años en promedio tardará en repetirse un evento de precipitación de la magnitud considerada.

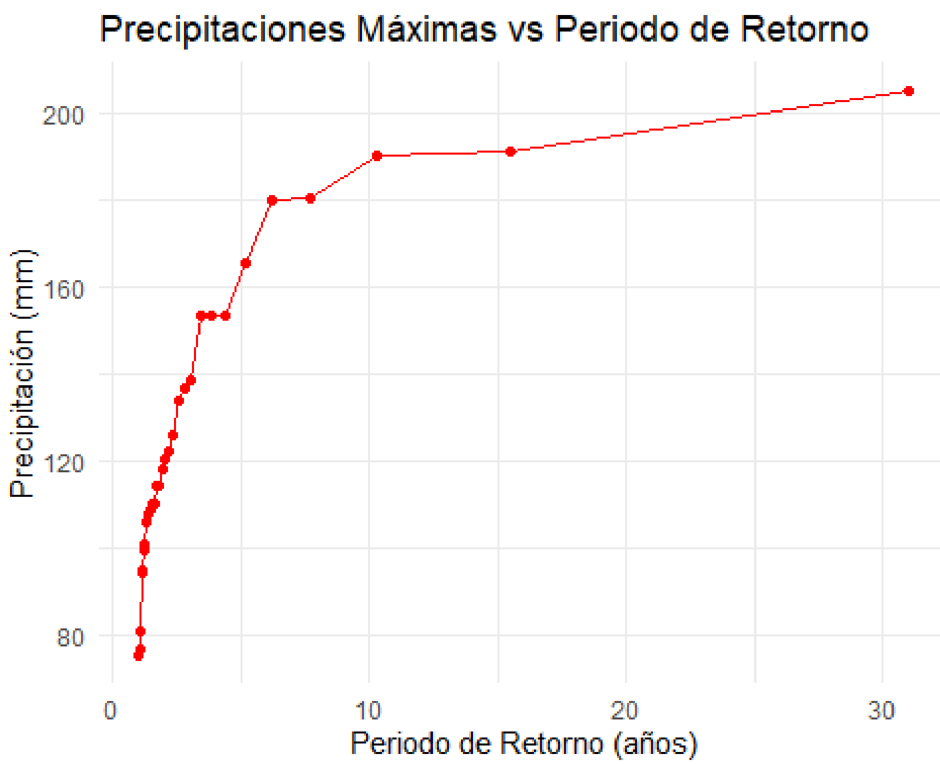
```r
# E) Calcular el periodo de retorno como el inverso de la probabilidad de excedencia
max_precipitation$P_ret <- 1 / max_precipitation$P_exce

# Mostrar las primeras filas del data frame con las nuevas columnas
head(max_precipitation)
```

Anio	Lluvia	Lluvia_ordenada	Rank	P_exce	P_no_exce	P_ret
1994	114.1	204.9	1	0.0322581	0.9677419	31.000000
1995	94.6	190.8	2	0.0645161	0.9354839	15.500000
1996	153.5	190.0	3	0.0967742	0.9032258	10.333333
1997	110.1	180.1	4	0.1290323	0.8709677	7.750000
1998	80.9	179.6	5	0.1612903	0.8387097	6.200000
1999	76.8	165.5	6	0.1935484	0.8064516	5.166667

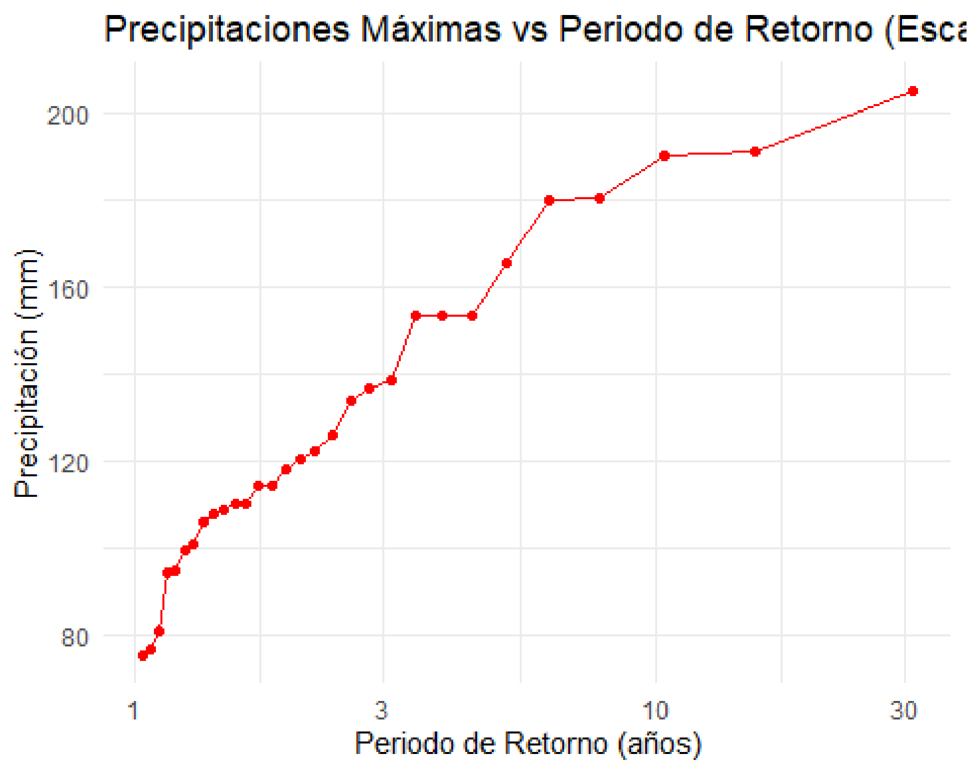
```r
# Gráfico de Precipitaciones vs Probabilidad de Excedencia

grafico_excedencia <- ggplot(max_precipitation, aes(x = P_exce, y =
Lluvia_ordenada)) +
  geom_line(color = "blue") +
  geom_point(color = "blue") +
  labs(title = "Precipitaciones Máximas vs Probabilidad de Excedencia",
       x = "Probabilidad de Excedencia", y = "Precipitación (mm)") +
  theme_minimal()

grafico_excedencia
```




```r
# Gráfico de Precipitaciones vs Periodo de Retorno
grafico_periodo_retorno1 <- ggplot(max_precipitation, aes(x = P_ret, y =
Lluvia_ordenada)) +
  geom_line(color = "red") +
  geom_point(color = "red") +
  labs(title = "Precipitaciones Máximas vs Periodo de Retorno",
       x = "Periodo de Retorno (años)", y = "Precipitación (mm)") +
  theme_minimal()

grafico_periodo_retorno1
```




```
grafico_periodo_retorno <- ggplot(max_precipitation, aes(x = P_ret, y =
Lluvia_ordenada)) +
  geom_line(color = "red") +
  geom_point(color = "red") +
  scale_x_log10() +
  labs(title = "Precipitaciones Máximas vs Periodo de Retorno (Escala
Logarítmica)",
       x = "Periodo de Retorno (años)", y = "Precipitación (mm)") +
  theme_minimal()

grafico_periodo_retorno
```




```
# Mostrar ambos gráficos en una sola ventana
#grid.arrange(grafico_excedencia, grafico_periodo_retorno, ncol = 2)
```

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

Describe las gráficas obtenidas.

- Los gráficos presentados nos permiten interpretar la relación entre las precipitaciones máximas (ordenadas de mayor a menor) y la probabilidad de excedencia, así como entre las precipitaciones y el período de retorno.

Precipitaciones Máximas vs Probabilidad de Excedencia:

- Las precipitaciones más altas (cercanas a 200 mm) tienen una probabilidad de excedencia muy baja (cercana a 0), lo que significa que son eventos muy raros.

- Conforme disminuye la magnitud de las precipitaciones, la probabilidad de que esas precipitaciones sean igualadas o superadas aumenta.

- Las precipitaciones más bajas tienen probabilidades de excedencia cercanas a 1, indicando que estos valores son mucho más comunes.

Precipitaciones Máximas vs Período de Retorno:

- Al aplicar una escala logarítmica, podemos apreciar mejor la distribución de los periodos de retorno para eventos con precipitación más baja, que tienen periodos de retorno más cortos. Los periodos de retorno de precipitaciones más altas son considerablemente más largos, y la escala logarítmica permite visualizar mejor esta relación de crecimiento más gradual. La relación sigue mostrando que las precipitaciones más altas están asociadas con periodos de retorno mayores, lo que significa que son eventos menos frecuentes. Por el contrario, las precipitaciones más bajas tienen un menor periodo de retorno, indicando que ocurren con mayor frecuencia.

¿Qué significa la probabilidad de excedencia?

- La probabilidad de excedencia es la probabilidad de que un evento (en este caso, una precipitación máxima) sea igualada o superada en un año determinado. Si una precipitación tiene una probabilidad de excedencia de 0.01 (1%), significa que hay un 1% de probabilidad de que una precipitación de ese valor o mayor ocurra en un año dado.

¿Qué significa el período de retorno?

- El período de retorno es una medida del tiempo promedio entre eventos de una magnitud específica. Por ejemplo, un período de retorno de 100 años significa que, en promedio, una precipitación máxima de ese nivel ocurre una vez cada 100 años. Sin embargo, esto no significa que el evento ocurra exactamente cada 100 años, sino que es un promedio a largo plazo.

¿Por qué es importante en hidrología?

- En hidrología, tanto la probabilidad de excedencia como el período de retorno son importantes para diseñar infraestructuras hidráulicas (presas, puentes, drenajes, etc.). Es crucial asegurar que las estructuras sean capaces de soportar eventos extremos, como lluvias intensas que podrían causar inundaciones. El diseño basado en estas medidas permite planificar obras que puedan manejar las lluvias más extremas, protegiendo tanto las infraestructuras como las vidas humanas.

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

- Para el diseño de infraestructuras hidráulicas, se busca que la probabilidad de excedencia sea muy baja. En general, se utilizan valores de probabilidad de excedencia asociados con eventos raros y extremos, por ejemplo, una probabilidad de excedencia del 1% a 5%, por ejemplo P_exce = 0.01, lo que corresponde a un período de retorno de 100 años. Esto garantiza que la estructura sea capaz de resistir eventos de precipitación que ocurren con poca frecuencia pero que podrían tener consecuencias graves si no se consideran en el diseño.

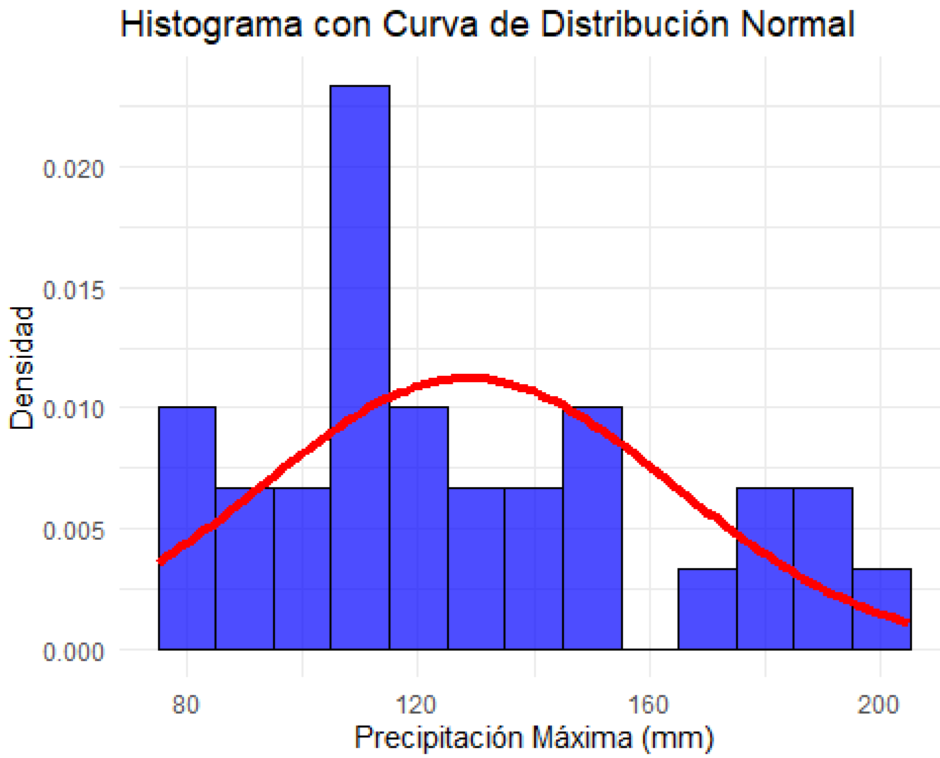
3. Análisis de Frecuencias Método Analítico

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.

A.1) 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?

```r
library(ggplot2)
library(MASS)
library(nortest)
library(stats)

# Ajuste a una Distribución Normal

# Construir el histograma de la función de densidad empírica y sobreponer
la curva de distribución normal

# Calcular la media y desviación estándar
media <- mean(max_precipitation$Lluvia)
desviacion_estandar <- sd(max_precipitation$Lluvia)

# Histograma con curva normal
ggplot(max_precipitation, aes(x = Lluvia)) +
  geom_histogram(aes(y = after_stat(density)), binwidth = 10, fill =
"blue", alpha = 0.7, color = "black") +
  stat_function(fun = dnorm, args = list(mean = media, sd =
desviacion_estandar), color = "red", size = 1.5) +
  labs(title = "Histograma con Curva de Distribución Normal",
       x = "Precipitación Máxima (mm)", y = "Densidad") +
  theme_minimal()
```



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

- De acuerdo con el histograma y la curva de distribución normal sobrepuesta, los datos no parecen ajustarse bien a una distribución normal, pues se observa una forma bastante irregular, con varios picos, lo que sugiere que los datos no siguen una forma simétrica. Además, hay un grupo de valores más altos (alrededor de 200 mm) que están más alejados del centro, lo que contribuye a una asimetría.

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

- La distribución normal tiene dos parámetros:
- Media μ: Representa el valor central o promedio de la distribución. En este caso, es el valor promedio de las precipitaciones máximas mensuales.
- Desviación estándar σ: Mide la dispersión o variabilidad de los datos con respecto a la media. Una desviación estándar más alta implica mayor dispersión en los datos.

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

- Los parámetros de la distribución normal, la media y la desviación estándar, se calculan directamente a partir de los datos utilizando las fórmulas estándar:
- Media: Es simplemente la suma de todos los valores dividida por el número de observaciones.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Donde: - (μ) es la media del conjunto de datos. - $(n)4 es el número total de observaciones. - (x_i) representa cada uno de los valores de los datos. - $(\sum_{i=1}^{n} x_i)$ es la suma de todos los valores (x_i).

- Desviación estándar: Se calcula como la raíz cuadrada de la varianza, que es la suma de los cuadrados de las diferencias entre cada valor y la media, dividida por el número de observaciones menos 1.

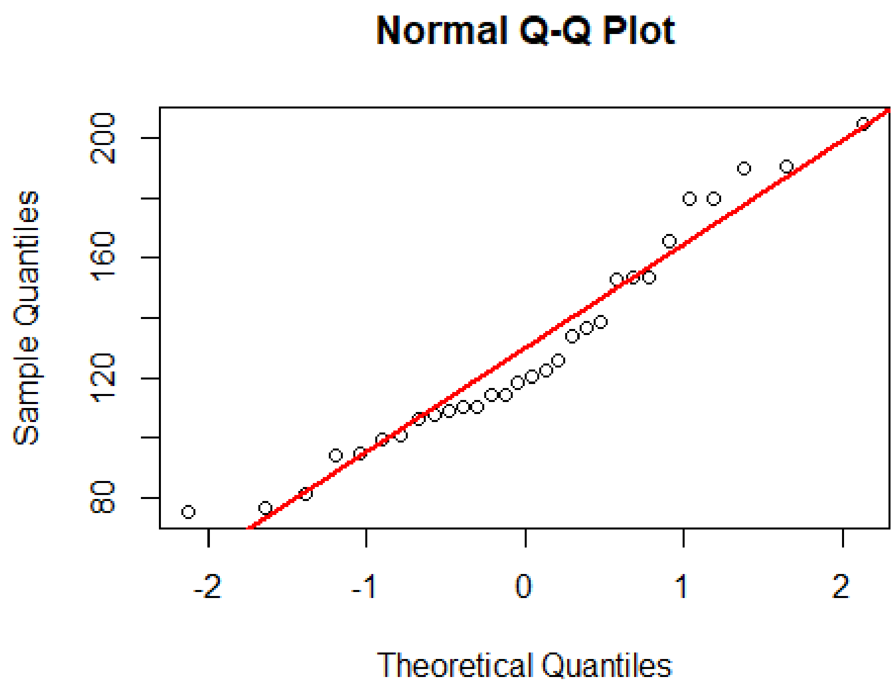
$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2}$$

Donde: - (σ) es la desviación estándar. - (n) es el número total de observaciones. - (x_i) es cada valor de los datos. - (μ) es la media del conjunto de datos. - $(\sum_{i=1}^{n}(x_i - \mu)^2)$ es la suma de los cuadrados de las diferencias entre cada valor (x_i) y la media.

- Estos parámetros son utilizados para generar la curva normal teórica sobre el histograma de los datos empíricos. Al comparar ambas representaciones, podemos ver visualmente si los datos siguen una distribución normal o si se desvían de ella, como es el caso en esta gráfica.

A.2) Construye la gráfica qqplot. De manera visual, ¿Los datos siguen una distribución normal de acuerdo con la Q-Qplot?

```
# Construir el Q-Q plot para verificar si los datos siguen una
distribución normal
qqnorm(max_precipitation$Lluvia)
qqline(max_precipitation$Lluvia, col = "red", lwd = 2)
```



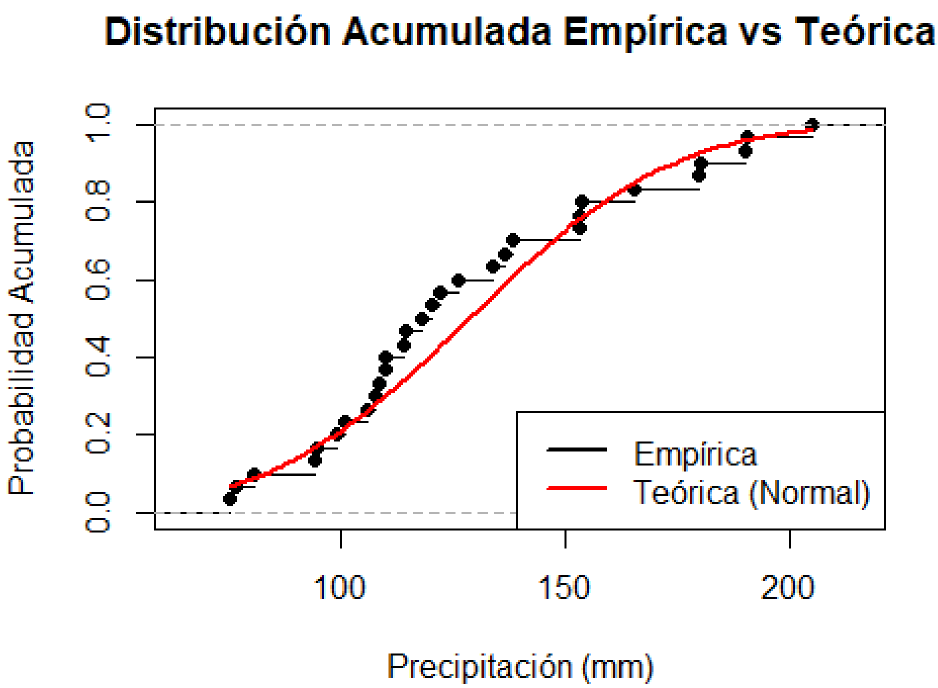
¿Los datos siguen una distribución normal de acuerdo con la Q-Qplot?

- De acuerdo con el gráfico Q-Q plot obtenido, los datos no siguen una distribución normal. Podemos observar que los puntos se desvían significativamente de la línea roja en las colas de la distribución, particularmente en la parte superior derecha y en la parte inferior izquierda. Esto indica que los valores extremos (tanto altos como bajos) no se ajustan a lo esperado en una distribución normal.

A.3) 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?

```
# Comparar las distribuciones acumuladas empíricas y teóricas (ojiva)
# CDF empírica
empirical_cdf <- ecdf(max_precipitation$Lluvia)

# Crear un rango de valores para la CDF teórica
x_vals <- seq(min(max_precipitation$Lluvia),
max(max_precipitation$Lluvia), length.out = 100)

# CDF teórica basada en los parámetros de los datos
teoretical_cdf <- pnorm(x_vals, mean = media, sd = desviacion_estandar)

# Gráfico de CDF empírica y teórica
plot(empirical_cdf, main = "Distribución Acumulada Empírica vs Teórica",
```

```r
xlab = "Precipitación (mm)", ylab = "Probabilidad Acumulada")
lines(x_vals, teoretical_cdf, col = "red", lwd = 2)
legend("bottomright", legend = c("Empírica", "Teórica (Normal)"), col =
c("black", "red"), lwd = 2)
```



Explica qué son datos empíricos y datos teóricos.

Datos empíricos: Son los datos que se observan directamente de la realidad, en este caso, las precipitaciones máximas mensuales en Chihuahua. Estos datos se basan en mediciones y observaciones del mundo real.

Datos teóricos: Son los datos que se generan a partir de un modelo matemático o estadístico que describe cómo se espera que los datos se comporten bajo ciertos supuestos. En este caso, la distribución teórica es la distribución normal, la cual se ajusta a los datos utilizando los parámetros (media y desviación estándar) calculados a partir de los mismos datos empíricos.

¿Se parecen las distribuciones de probabilidad acumuladas?

- Al observar la gráfica, se puede notar que las distribuciones empírica y teórica no coinciden completamente. Aunque coinciden razonablemente en el rango medio, las diferencias en los extremos sugieren que la distribución normal no captura adecuadamente el comportamiento extremo de las precipitaciones. En los valores de precipitación más bajos, la distribución empírica está por debajo de la teórica, mientras que en los valores más altos, la distribución empírica crece más rápido, lo que indica menos eventos de precipitaciones bajas y más eventos extremos de precipitación alta que los previstos por la distribución normal.

A.4) 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?

No te olvides de las hipótesis planteadas:

$$H_0\text{:Los datos provienen de una distribución normal}$$

$$H_1\text{:Los datos no provienen de una distribución normal}$$

```
# Prueba de bondad de ajuste: Shapiro-Wilk
shapiro_test <- shapiro.test(max_precipitation$Lluvia)
shapiro_test

##
##  Shapiro-Wilk normality test
##
## data:  max_precipitation$Lluvia
## W = 0.94298, p-value = 0.1094

# Prueba de bondad de ajuste: Kolmogorov-Smirnov
ks_normal <- ks.test(max_precipitation$Lluvia, "pnorm", mean = media, sd
= desviacion_estandar)
ks_normal

##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  max_precipitation$Lluvia
## D = 0.13672, p-value = 0.5818
## alternative hypothesis: two-sided

# Mostrar los resultados de las pruebas de bondad de ajuste
cat("Resultados de las pruebas de bondad de ajuste:\n")

## Resultados de las pruebas de bondad de ajuste:

cat("Shapiro-Wilk Test: p-value =", shapiro_test$p.value, "\n")

## Shapiro-Wilk Test: p-value = 0.1094479

cat("Kolmogorov-Smirnov Test: p-value =", ks_normal$p.value, "\n")

## Kolmogorov-Smirnov Test: p-value = 0.5818443

# Conclusión según los p-values y las hipótesis:
if (shapiro_test$p.value < 0.05) {
  cat("La prueba de Shapiro-Wilk rechaza la hipótesis nula (los datos no
provienen de una distribución normal).\n")
```

```r
} else {
  cat("La prueba de Shapiro-Wilk no rechaza la hipótesis nula (los datos pueden provenir de una distribución normal).\n")
}
```

```
## La prueba de Shapiro-Wilk no rechaza la hipótesis nula (los datos
pueden provenir de una distribución normal).
```

```r
if (ks_normal$p.value < 0.05) {
  cat("La prueba de Kolmogorov-Smirnov rechaza la hipótesis nula (los datos no provienen de una distribución normal).\n")
} else {
  cat("La prueba de Kolmogorov-Smirnov no rechaza la hipótesis nula (los datos pueden provenir de una distribución normal).\n")
}
```

```
## La prueba de Kolmogorov-Smirnov no rechaza la hipótesis nula (los
datos pueden provenir de una distribución normal).
```

¿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?

Prueba Shapiro-Wilk:

- Esta prueba se utiliza para verificar si los datos siguen una distribución normal.

Estadístico W: El valor obtenido es W = 0.94298.

p-value: El valor p = 0.1094, lo que significa que no podemos rechazar que los datos siguen una distribución normal, aunque esto no significa que lo sigan con certeza.

Prueba Kolmogorov-Smirnov (KS):

- Esta prueba compara la función de distribución acumulada (CDF) de la muestra con una distribución teórica (en este caso, una normal).

Estadístico D: El valor obtenido es D = 0.13672.

p-value: El valor p es p = 0.5818. Dado que el p-value es mucho mayor que el nivel de significancia común (0.05), no rechazamos la hipótesis nula. Esto sugiere que los datos no son significativamente diferentes de una distribución normal.

- En ambas pruebas (Shapiro-Wilk y Kolmogorov-Smirnov), el p-value es mayor que 0.05, lo que significa que no se rechaza la hipótesis nula H_0 en ninguno de los casos. Por lo tanto, no podemos concluir que los datos se desvíen de una distribución normal.

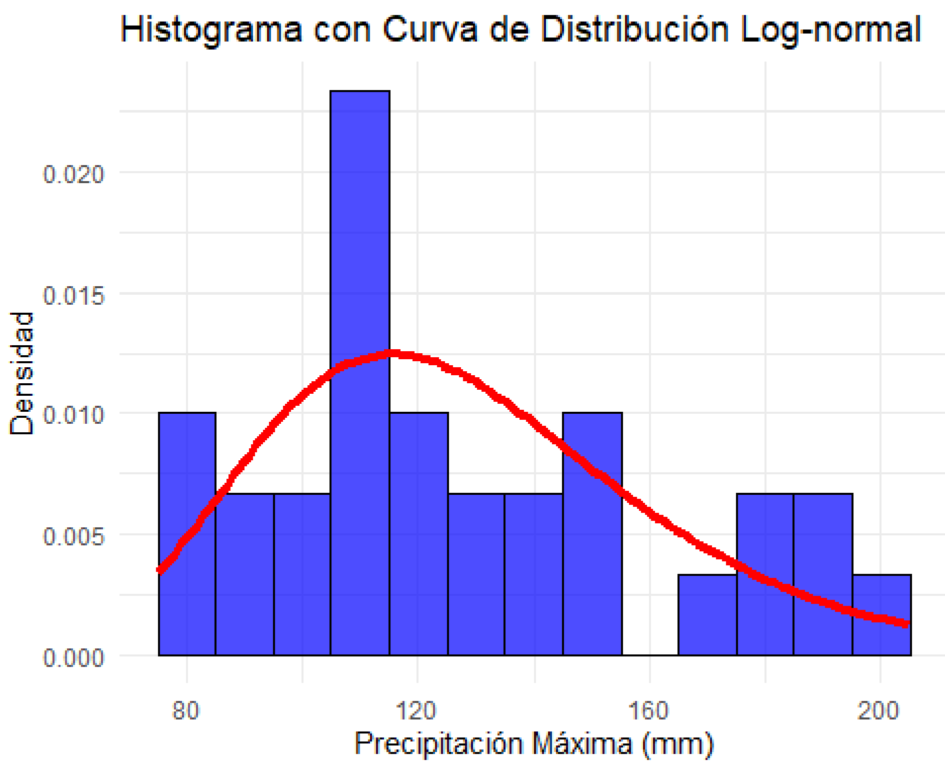
¿Podemos concluir que los datos de las precipitaciones máximas mensuales son normales? ¿Por qué?

- Aunque las pruebas no permiten rechazar la normalidad, basándonos en los gráficos previos (histograma, Q-Q plot y CDF), se puede sospechar que la normalidad no es un ajuste perfecto, especialmente en los extremos. Sin embargo, desde el punto de vista estadístico, los datos no muestran una desviación significativa de la normalidad según estas pruebas.

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.

B.1) 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
library(fitdistrplus)

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

## Loading required package: survival

# Filtrar los datos para asegurarse de que solo queden valores positivos
chihuahua_data_filtered <- max_precipitation[max_precipitation$Lluvia >
0, ]

# Ajuste a la distribución Log-normal
# Ajustar la distribución Log-normal usando el método de máxima
verosimilitud (MLE)
lognormal_fit <- fitdistr(chihuahua_data_filtered$Lluvia, "lognormal")

# Extraer parámetros estimados
mu_hat <- lognormal_fit$estimate[1]
sigma_hat <- lognormal_fit$estimate[2]

# Histograma con curva log-normal sobrepuesta
ggplot(chihuahua_data_filtered, aes(x = Lluvia)) +
  geom_histogram(aes(y = after_stat(density)), binwidth = 10, fill =
"blue", alpha = 0.7, color = "black") +
  stat_function(fun = dlnorm, args = list(meanlog = mu_hat, sdlog =
sigma_hat), color = "red", size = 1.5) +
  labs(title = "Histograma con Curva de Distribución Log-normal",
       x = "Precipitación Máxima (mm)", y = "Densidad") +
  theme_minimal()
```



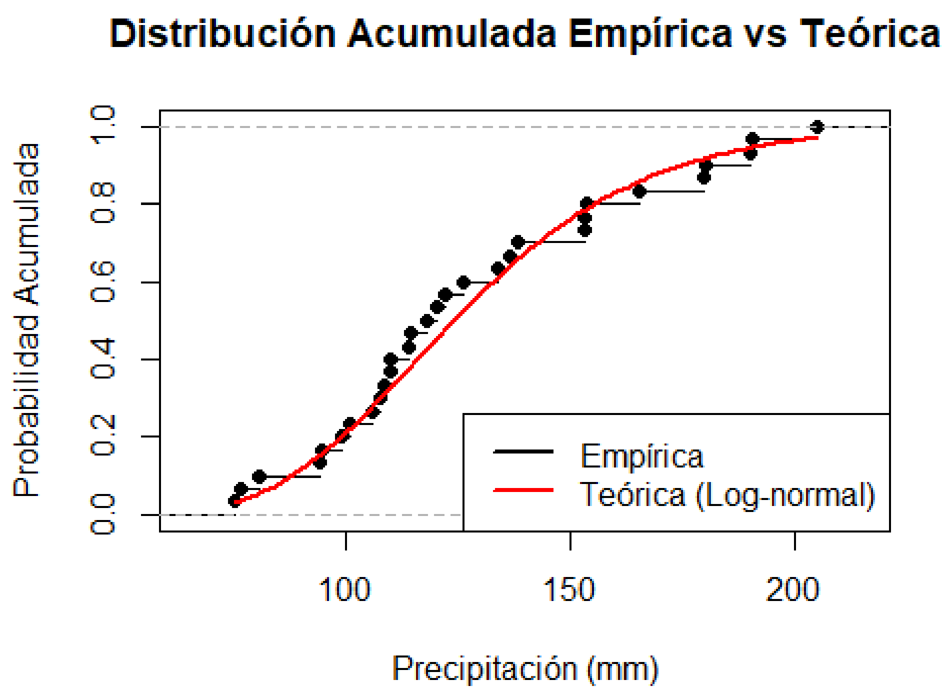
De manera visual, ¿te parece que los datos se ajustan bien a una distribución Log-normal? Explica.

- De manera visual, el ajuste a una distribución log-normal parece más adecuado que el ajuste a una distribución normal, especialmente en la cola derecha, donde se encuentran los valores más altos. Sin embargo, el ajuste aún no es perfecto, particularmente en los picos del histograma, lo que indica que aunque la distribución log-normal puede ser un buen modelo para estos datos, podría no capturar todos los detalles con exactitud.

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

```r
# Comparación de CDF empírica vs teórica (ojiva)

# CDF empírica
empirical_cdf <- ecdf(max_precipitation$Lluvia)

# Crear un rango de valores para la CDF teórica
x_vals <- seq(min(max_precipitation$Lluvia),
max(max_precipitation$Lluvia), length.out = 100)

# CDF teórica log-normal basada en los parámetros ajustados
teoretical_cdf <- plnorm(x_vals, meanlog = mu_hat, sdlog = sigma_hat)
```

```r
# Gráfico de CDF empírica y teórica
plot(empirical_cdf, main = "Distribución Acumulada Empírica vs Teórica",
xlab = "Precipitación (mm)", ylab = "Probabilidad Acumulada")
lines(x_vals, teoretical_cdf, col = "red", lwd = 2)
legend("bottomright", legend = c("Empírica", "Teórica (Log-normal)"), col
= c("black", "red"), lwd = 2)
```



Explica qué son datos empíricos y datos teóricos.

Datos empíricos: Son los datos observados directamente de la realidad, en este caso, las precipitaciones máximas mensuales medidas en Chihuahua. La curva negra en la gráfica representa la distribución acumulada empírica, que muestra cómo se distribuyen los valores observados en el conjunto de datos.

Datos teóricos: Son los datos generados a partir de un modelo teórico, en este caso, la distribución log-normal ajustada con los parámetros estimados de los datos empíricos. La curva roja representa la distribución acumulada teórica log-normal, que muestra cómo se distribuirían los datos si siguieran una distribución log-normal perfecta basada en esos parámetros.

¿Se parecen las distribuciones de probabilidad acumuladas?

- Al observar el gráfico, las distribuciones empírica (línea negra) y teórica log-normal (línea roja) coinciden bien en el rango medio de los datos, lo que sugiere que la log-normal captura adecuadamente ese intervalo. En los extremos inferiores hay pequeñas diferencias, mientras que en los superiores la log-normal modela mejor los datos que la distribución normal. Aunque las distribuciones acumuladas se parecen bastante, especialmente en el rango medio y alto, persisten pequeñas diferencias en los extremos, lo que indica que la log-normal es un mejor ajuste, pero no perfecto.

B.3) Haz la prueba KS para determinar si los datos se ajustan a una Log-normal

```
# Prueba de bondad de ajuste: Kolmogorov-Smirnov para Log-normal
ks_log <- ks.test(max_precipitation$Lluvia, "plnorm", meanlog = mu_hat,
sdlog = sigma_hat)
ks_log

##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  max_precipitation$Lluvia
## D = 0.087439, p-value = 0.9605
## alternative hypothesis: two-sided

# Mostrar los resultados de la prueba de KS
cat("Resultados de la prueba de Kolmogorov-Smirnov (KS):\n")

## Resultados de la prueba de Kolmogorov-Smirnov (KS):

cat("Estadístico de prueba:", ks_log$statistic, "\n")

## Estadístico de prueba: 0.08743888

cat("p-value:", ks_log$p.value, "\n")

## p-value: 0.9605171
```

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

```
# Interpretación de la prueba KS
if (ks_log$p.value < 0.05) {
  cat("La prueba de Kolmogorov-Smirnov rechaza la hipótesis nula (los
datos no siguen una distribución log-normal).\n")
} else {
  cat("La prueba de Kolmogorov-Smirnov no rechaza la hipótesis nula (los
datos pueden seguir una distribución log-normal).\n")
}

## La prueba de Kolmogorov-Smirnov no rechaza la hipótesis nula (los
datos pueden seguir una distribución log-normal).
```

¿Qué información nos da la prueba KS para una Log-normal? ¿Cuál es el valor del estadístico de prueba? ¿Cuál es el valor del estadístico? ¿Cuál es el p-value de la prueba? ¿Se acepta o se rechaza la hipótesis nula?

- La prueba de Kolmogorov-Smirnov (KS) compara la distribución acumulada empírica de los datos con la distribución acumulada teórica de una distribución log-normal ajustada a los datos. El objetivo de esta prueba es evaluar si los datos observados pueden haber sido generados por una distribución log-normal.

Estadístico de prueba: El valor del estadístico de prueba **D** es 0.0874. Este valor mide la distancia máxima entre las distribuciones empírica y teórica. Un valor más cercano a 0 indica un mejor ajuste.

p-value: El p-value obtenido es 0.9605 Este valor nos indica la probabilidad de obtener una diferencia tan grande como la observada (o mayor) entre las distribuciones empírica y teórica, asumiendo que la hipótesis nula es verdadera.

- H_0: Los datos provienen de una distribución log-normal.
- H_1: Los datos no provienen de una distribución log-normal.
- En este caso, dado que el p-value es 0.9605, muy superior a 0.05, no se rechaza la hipótesis nula. Esto significa que no se puede descartar que los datos sigan una distribución log-normal..

¿Podemos concluir que los datos de las precipitaciones máximas mensuales siguen una distribución log-normal? ¿Por qué?

- Aunque no podemos concluir con total certeza que los datos siguen una distribución log-normal (debido a las diferencias observadas en los extremos), el ajuste es bastante aceptable según las pruebas estadísticas (como la prueba KS) y los análisis gráficos. Por lo tanto, podemos decir que la distribución log-normal es una buena aproximación para modelar los datos de precipitaciones máximas mensuales, pero podría no capturar todos los detalles de la distribución, especialmente en los valores extremos.

B.5)¿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.

```
# Parámetros de la distribución Log-normal y explicación
cat("Los parámetros de la distribución Log-normal son:\n")

## Los parámetros de la distribución Log-normal son:

cat("Mu (media del logaritmo de los datos):", mu_hat, "\n")

## Mu (media del logaritmo de los datos): 4.821242

cat("Sigma (desviación estándar del logaritmo de los datos):", sigma_hat,
"\n")

## Sigma (desviación estándar del logaritmo de los datos): 0.2677591
```

¿Cuántos parámetros tiene la distribución Log-normal? ¿Cuáles son?

- La distribución log-normal tiene dos parámetros:
- μ: Es la media del logaritmo natural de la variable aleatoria que sigue una distribución normal.
- σ: Es la desviación estándar del logaritmo natural de la variable aleatoria.

¿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.

- Los parámetros de la distribución log-normal se calculan utilizando el método de máxima verosimilitud (MLE) en el código. Este método estima los parámetros que maximizan la probabilidad de observar los datos dados, bajo el supuesto de que los datos siguen una distribución log-normal.

Paso a paso con el Método de Momentos

- El método de momentos se basa en igualar los momentos teóricos de la distribución log-normal con los momentos empíricos obtenidos de los datos. Para la distribución log-normal, los momentos en términos de la media μ y la varianza σ^2 son los siguientes:

1. **Media empírica de los datos** $\overline{X}$:

- La media $E(X)$ de una variable aleatoria X que sigue una distribución log-normal está dada por:

$$E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

- Para la media empírica, simplemente se calcula:

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

* donde X_i son los datos de las precipitaciones máximas mensuales.

2. **Varianza empírica de los datos** $Var(X)$:

- La varianza de una variable aleatoria X que sigue una distribución log-normal está dada por:

$$\mathrm{Var}(X) = [\exp(\sigma^2) - 1]\exp(2\mu + \sigma^2)$$

- La varianza empírica es:

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

3. **Resolviendo para** μ y σ^2:

- Tomando los logaritmos naturales de ambos momentos teóricos, se pueden despejar μ y σ^2 en términos de los momentos empíricos.

- Primero, despejamos σ^2 utilizando la relación entre la media y la varianza empírica:

$$\sigma^2 = \ln\left(\frac{S^2}{\overline{X}^2} + 1\right)$$

- Luego, despejamos μ:

$$\mu = \ln(\overline{X}) - \frac{\sigma^2}{2}$$

```
# 1. Cálculo de la media y varianza empírica de los datos
media_empirica <- mean(chihuahua_data_filtered$Lluvia)
varianza_empirica <- var(chihuahua_data_filtered$Lluvia)

# 2. Aplicar el método de momentos para calcular los parámetros de la
Log-normal
# Cálculo de sigma^2 y mu mediante el método de momentos
sigma2_momentos <- log(varianza_empirica / (media_empirica^2) + 1)
mu_momentos <- log(media_empirica) - sigma2_momentos / 2

# Mostrar los resultados del método de momentos
cat("Método de momentos:\n")

## Método de momentos:

cat("Mu =", mu_momentos, "\n")

## Mu = 4.820549

cat("Sigma^2 =", sigma2_momentos, "\n")

## Sigma^2 = 0.07330009

# 3. Comparación con el método de máxima verosimilitud (MLE)
# Ajustar la distribución log-normal usando máxima verosimilitud (MLE)
lognormal_fit <- fitdistr(chihuahua_data_filtered$Lluvia, "lognormal")

# Extraer parámetros ajustados por MLE
mu_MLE <- lognormal_fit$estimate[1]
sigma2_MLE <- lognormal_fit$estimate[2]^2

# Mostrar los resultados de MLE
cat("Método de máxima verosimilitud (MLE):\n")

## Método de máxima verosimilitud (MLE):
```

```r
cat("Mu =", mu_MLE, "\n")

## Mu = 4.821242

cat("Sigma^2 =", sigma2_MLE, "\n")

## Sigma^2 = 0.07169496

# Comparar los resultados
cat("Comparación:\n")

## Comparación:

cat("Mu (Momentos) =", mu_momentos, "vs Mu (MLE) =", mu_MLE, "\n")

## Mu (Momentos) = 4.820549 vs Mu (MLE) = 4.821242

cat("Sigma^2 (Momentos) =", sigma2_momentos, "vs Sigma^2 (MLE) =",
sigma2_MLE, "\n")

## Sigma^2 (Momentos) = 0.07330009 vs Sigma^2 (MLE) = 0.07169496
```

Conclusión

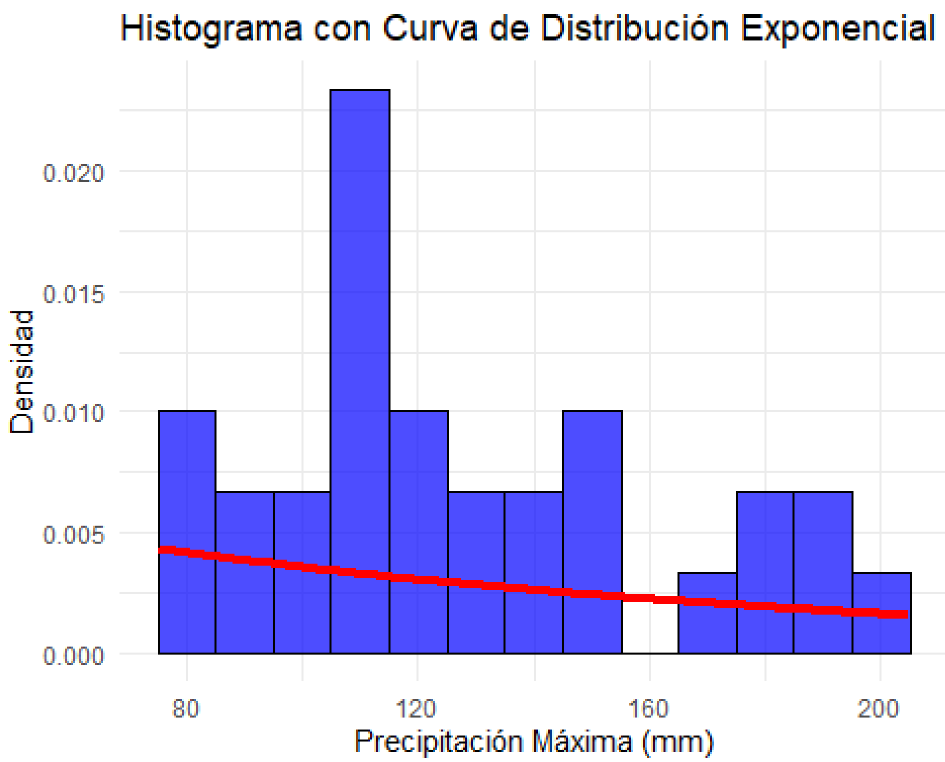
- Ambos métodos ofrecen estimaciones similares, pero el método de máxima verosimilitud (MLE) es más preciso al maximizar la probabilidad de los datos observados bajo la distribución ajustada. Aunque la diferencia en Mu es insignificante, MLE proporciona una mejor estimación de Sigma^2, lo que lo convierte en la mejor opción en este caso, aunque el método de momentos también es una buena aproximación.

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.

C.1) 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
# Ajuste a la distribución Exponencial

# Ajustar la distribución Exponencial usando el método de máxima
verosimilitud (MLE)
exponential_fit <- fitdistr(chihuahua_data_filtered$Lluvia,
"exponential")

# Extraer el parámetro lambda (inverso de la media)
lambda_hat <- 1 / mean(chihuahua_data_filtered$Lluvia)

# Histograma con curva de distribución Exponencial sobrepuesta
ggplot(chihuahua_data_filtered, aes(x = Lluvia)) +
```

```
  geom_histogram(aes(y = after_stat(density)), binwidth = 10, fill =
"blue", alpha = 0.7, color = "black") +
  stat_function(fun = dexp, args = list(rate = lambda_hat), color =
"red", size = 1.5) +
  labs(title = "Histograma con Curva de Distribución Exponencial",
       x = "Precipitación Máxima (mm)", y = "Densidad") +
  theme_minimal()
```



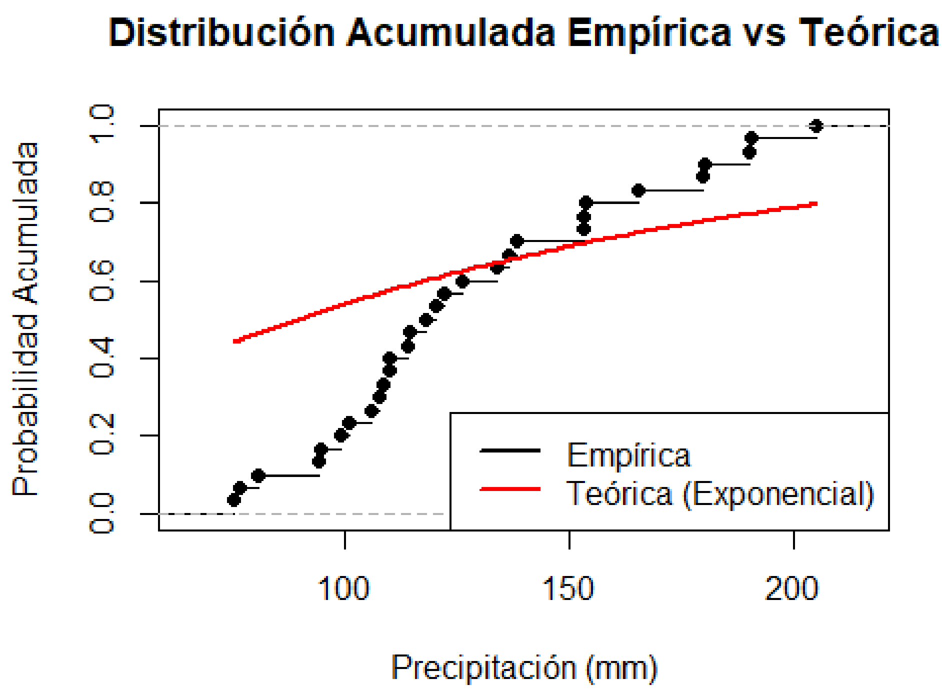
De manera visual, ¿te parece que los datos se ajustan bien a una distribución Exponencial? Explica.

- Los datos no se ajustan bien a una distribución exponencial. El histograma muestra varios picos en diferentes intervalos que no siguen la tendencia decreciente suave de la distribución exponencial. Además, la distribución exponencial no captura adecuadamente los valores más altos de precipitación, lo que indica un mal ajuste en los extremos. En resumen, la distribución exponencial no es un buen modelo para estos datos.

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

```
# Comparación de CDF empírica vs teórica (ojiva)
# CDF empírica
empirical_cdf <- ecdf(chihuahua_data_filtered$Lluvia)
```

```r
# Crear un rango de valores para la CDF teórica
x_vals <- seq(min(chihuahua_data_filtered$Lluvia),
max(chihuahua_data_filtered$Lluvia), length.out = 100)

# CDF teórica exponencial basada en el parámetro ajustado
teoretical_cdf <- pexp(x_vals, rate = lambda_hat)

# Gráfico de CDF empírica y teórica
plot(empirical_cdf, main = "Distribución Acumulada Empírica vs Teórica",
xlab = "Precipitación (mm)", ylab = "Probabilidad Acumulada")
lines(x_vals, teoretical_cdf, col = "red", lwd = 2)
legend("bottomright", legend = c("Empírica", "Teórica (Exponencial)"),
col = c("black", "red"), lwd = 2)
```



Explica qué son datos empíricos y datos teóricos.

Datos empíricos: Los datos empíricos son los datos reales observados de las precipitaciones máximas. La distribución acumulada empírica muestra la proporción de datos por debajo de un valor específico de precipitación.

Datos teóricos: Los datos teóricos provienen de una distribución exponencial ajustada a los datos reales. Es decir, utilizamos los parámetros estimados (en este caso, la tasa lambda) para generar la distribución acumulada que teóricamente deberían seguir los datos si estos provinieran de una distribución exponencial.

¿Se parecen las distribuciones de probabilidad acumuladas?

- Las distribuciones de probabilidad acumuladas no se parecen. La distribución empírica se desvía significativamente de la teórica exponencial. En el rango medio, la distribución exponencial subestima las precipitaciones, y en los extremos superiores (más de 150 mm), no captura adecuadamente los eventos extremos.

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

```
# Prueba de bondad de ajuste: Kolmogorov-Smirnov para Exponencial
ks_exp <- ks.test(chihuahua_data_filtered$Lluvia, "pexp", rate =
lambda_hat)
ks_exp

##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  chihuahua_data_filtered$Lluvia
## D = 0.4426, p-value = 0.000007148
## alternative hypothesis: two-sided

# Mostrar los resultados de la prueba de KS
cat("Resultados de la prueba de Kolmogorov-Smirnov (KS):\n")

## Resultados de la prueba de Kolmogorov-Smirnov (KS):

cat("Estadístico de prueba:", ks_exp$statistic, "\n")

## Estadístico de prueba: 0.4425994

cat("p-value:", ks_exp$p.value, "\n")

## p-value: 0.000007148288
```

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

```
# Interpretación de la prueba KS
if (ks_exp$p.value < 0.05) {
  cat("La prueba de Kolmogorov-Smirnov rechaza la hipótesis nula (los
datos no siguen una distribución exponencial).\n")
} else {
  cat("La prueba de Kolmogorov-Smirnov no rechaza la hipótesis nula (los
datos pueden seguir una distribución exponencial).\n")
}

## La prueba de Kolmogorov-Smirnov rechaza la hipótesis nula (los datos
no siguen una distribución exponencial).
```

¿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?

- La prueba de Kolmogorov-Smirnov (KS) nos proporciona una comparación entre la distribución empírica de los datos observados y una distribución teórica (en este caso, la distribución exponencial).

Valor del estadístico de prueba: El valor del estadístico de prueba es D = 0.4426 Este valor mide la máxima diferencia entre las distribuciones acumuladas empírica y teórica (en este caso, la exponencial). Mientras más grande sea este valor, más diferente será la distribución teórica de la empírica.

P-value: El valor p es 0.000007148 (muy pequeño). El valor p indica la probabilidad de observar una diferencia tan grande (o mayor) entre las distribuciones si la hipótesis nula fuera verdadera.

H_0: Los datos provienen de una distribución exponencial.

H_1: Los datos no provienen de una distribución exponencial.

- Dado que el valor p es extremadamente pequeño ($p < 0.05$), rechazamos la hipótesis nula H_0. Esto significa que, con un nivel de significancia común (como 0.05), existe suficiente evidencia para concluir que los datos no siguen una distribución exponencial.

¿Podemos concluir que los datos de las precipitaciones máximas mensuales siguen una distribución exponencial? ¿Por qué?

- Los resultados de la prueba KS sugieren que los datos de las precipitaciones máximas mensuales no se ajustan a una distribución exponencial.

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

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

- La distribución Exponencial tiene un solo parámetro. El parámetro de tasa, que se suele denotar como λ. Este parámetro λ está relacionado con la media de la distribución exponencial de la siguiente manera:

$$\lambda = \frac{1}{\mu}$$

donde:

λ es el parámetro de tasa.

μ es la media de los datos.

¿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.

El cálculo del parámetro λ de la distribución Exponencial sigue principios fundamentales tanto en el método de máxima verosimilitud (MLE) como en el método de momentos.

1. **Distribución Exponencial y su parámetro λ**:

- La distribución Exponencial se caracteriza por un solo parámetro λ , conocido como la tasa de ocurrencia. La función de densidad de probabilidad (PDF) es:

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

Donde λ es el parámetro que define la tasa de ocurrencia de los eventos. La media de una distribución Exponencial es el inverso de λ:

$$\mu = \frac{1}{\lambda}$$

2. **Método de Máxima Verosimilitud (MLE)**:
   El MLE busca el valor del parámetro que maximiza la probabilidad de observar los datos. Para la distribución Exponencial, el estimador MLE del parámetro λ es el inverso de la media de los datos observados:

$$\hat{\lambda} = \frac{1}{\overline{x}}$$

- donde $\overline{x}$ es la media de los datos observados.
- En el código, se calcula λ como el inverso de la media de los datos, ya que:
- Para una muestra $(x_1, x_2, \dots, x_n)$, el estimador de máxima verosimilitud de λ se calcula como:

$$\hat{\lambda} = \frac{1}{\overline{x}}$$

- donde $(\overline{x})$ es la media de los datos observados.
- Por lo tanto, el código está aplicando la fórmula del MLE de manera correcta, calculando λ a partir de la media de los datos filtrados de las precipitaciones.
- La razón de que λ se calcule como el inverso de la media de los datos observados es que, en la distribución Exponencial, la media (esperanza matemática) de los datos es directamente proporcional al parámetro de tasa λ:

La esperanza matemática o media de la distribución Exponencial es:

$$E[X] = \frac{1}{\lambda}$$

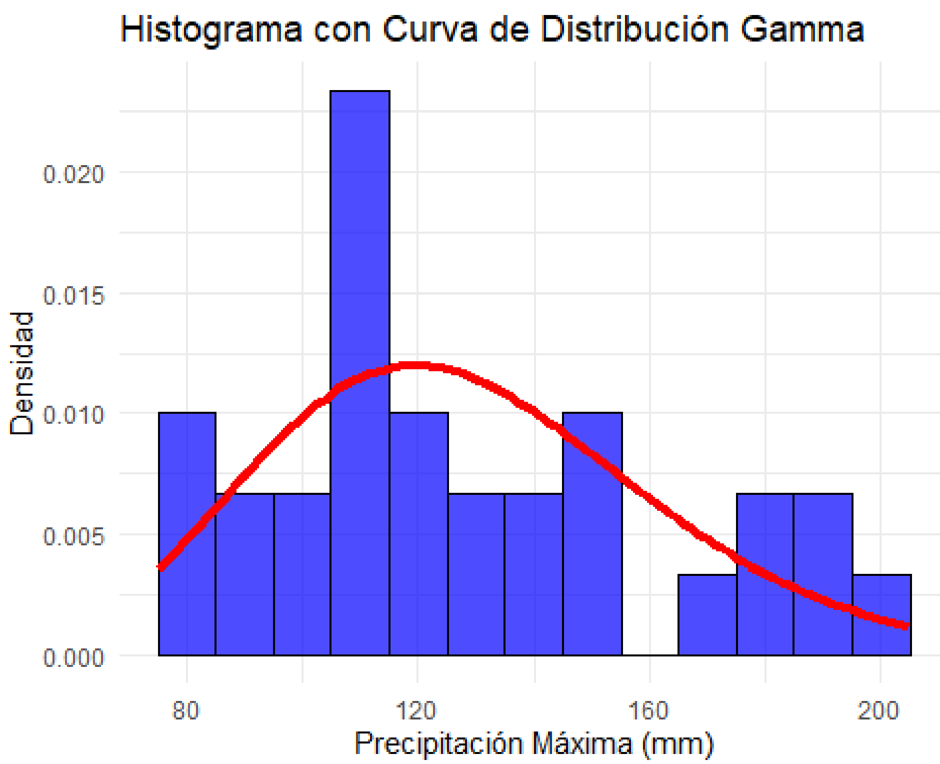
- Por lo tanto, para obtener el parámetro λ, simplemente tomamos el inverso de la media observada de los datos, ya que eso maximiza la probabilidad de los datos según la función de densidad de probabilidad de la distribución Exponencial.

```
# Paso 1: Calcular la media de los datos (Momento Muestral)
mean_precipitation <- mean(chihuahua_data_filtered$Lluvia)
cat("Momento Muestral (Media de los datos):", mean_precipitation, "\n")

## Momento Muestral (Media de los datos): 128.6633

# Paso 2: Estimar el parámetro lambda usando el método de momentos
lambda_hat_momentos <- 1 / mean_precipitation
cat("Parámetro lambda estimado por el método de momentos:",
lambda_hat_momentos, "\n")

## Parámetro lambda estimado por el método de momentos: 0.007772222

# Paso 3: Comparar con el estimador de máxima verosimilitud (MLE)
lambda_hat_mle <- 1 / mean(chihuahua_data_filtered$Lluvia)
cat("Parámetro lambda estimado por MLE:", lambda_hat_mle, "\n")

## Parámetro lambda estimado por MLE: 0.007772222
```

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.

D.1) 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.

```
# Ajuste a la distribución Gamma usando el método de máxima verosimilitud
(MLE)
gamma_fit <- fitdistr(chihuahua_data_filtered$Lluvia, "gamma")

# Extraer los parámetros estimados: shape (forma) y rate (tasa)
shape_hat <- gamma_fit$estimate[1]
rate_hat <- gamma_fit$estimate[2]

# Histograma con curva de distribución Gamma sobrepuesta
ggplot(chihuahua_data_filtered, aes(x = Lluvia)) +
  geom_histogram(aes(y = after_stat(density)), binwidth = 10, fill =
"blue", alpha = 0.7, color = "black") +
  stat_function(fun = dgamma, args = list(shape = shape_hat, rate =
rate_hat), color = "red", size = 1.5) +
  labs(title = "Histograma con Curva de Distribución Gamma",
       x = "Precipitación Máxima (mm)", y = "Densidad") +
  theme_minimal()
```



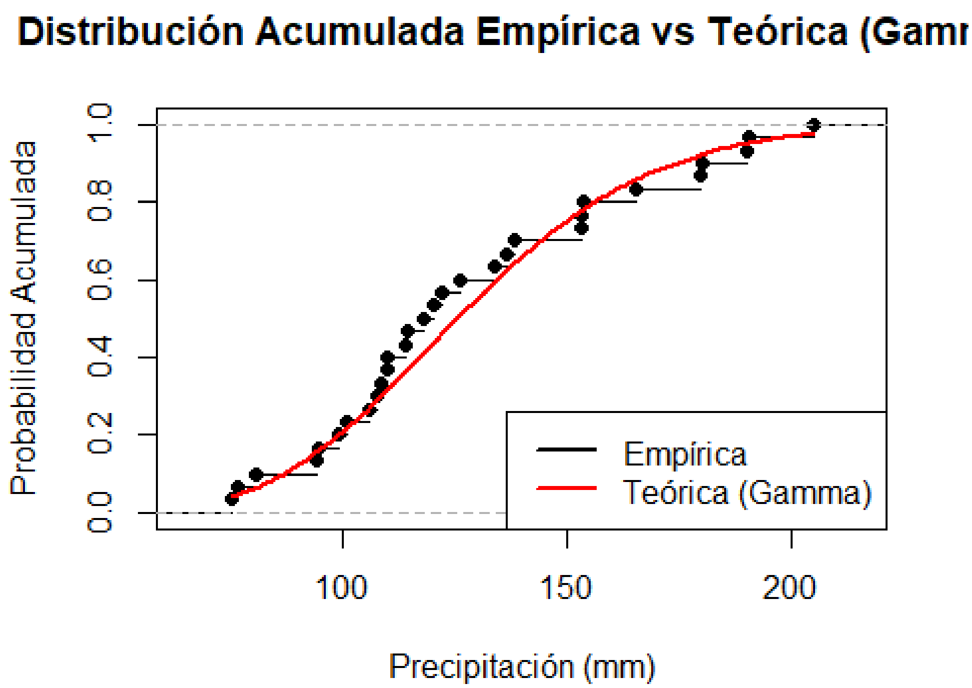
De manera visual, ¿te parece que los datos se ajustan bien a una distribución Gamma? Explica.

- Visualmente, los datos no se ajustan completamente a una distribución Gamma. Aunque en el rango medio y en los valores bajos de precipitaciones la curva sigue razonablemente el patrón del histograma, no captura bien los picos irregulares ni los eventos extremos en los valores altos de precipitación. En resumen, la distribución Gamma no es el mejor modelo para estos datos.

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

```
# Comparación de CDF empírica vs teórica (ojiva)
# CDF empírica
empirical_cdf <- ecdf(chihuahua_data_filtered$Lluvia)

# Crear un rango de valores para la CDF teórica
x_vals <- seq(min(chihuahua_data_filtered$Lluvia),
max(chihuahua_data_filtered$Lluvia), length.out = 100)

# CDF teórica basada en los parámetros ajustados
theoretical_cdf <- pgamma(x_vals, shape = shape_hat, rate = rate_hat)

# Gráfico de CDF empírica y teórica
plot(empirical_cdf, main = "Distribución Acumulada Empírica vs Teórica
```

```
(Gamma)", xlab = "Precipitación (mm)", ylab = "Probabilidad Acumulada")
lines(x_vals, theoretical_cdf, col = "red", lwd = 2)
legend("bottomright", legend = c("Empírica", "Teórica (Gamma)"), col =
c("black", "red"), lwd = 2)
```



Explica qué son datos empíricos y datos teóricos.

Datos empíricos: Los datos empíricos representan las observaciones reales de las precipitaciones máximas mensuales en Chihuahua. En el gráfico, estos datos se muestran en la línea negra, que representa la probabilidad acumulada basada en las observaciones.

Datos teóricos: Los datos teóricos se basan en la distribución Gamma ajustada a los datos empíricos utilizando los parámetros shape y rate estimados por máxima verosimilitud (MLE). En el gráfico, la línea roja representa la distribución teórica de Gamma.

¿Se parecen las distribuciones de probabilidad acumuladas?

Las distribuciones de probabilidad acumuladas se parecen en el rango medio, pero existen diferencias en los extremos, especialmente en las precipitaciones más bajas y más altas. Aunque la distribución Gamma es un mejor ajuste que otras, no modela perfectamente todo el rango de los datos.

D.3) Haz la prueba KS para determinar si los datos se ajustan a una Gamma

```
# Prueba de bondad de ajuste: Kolmogorov-Smirnov para Gamma
ks_gamma <- ks.test(chihuahua_data_filtered$Lluvia, "pgamma", shape =
```

```r
shape_hat, rate = rate_hat)
ks_gamma

##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  chihuahua_data_filtered$Lluvia
## D = 0.10502, p-value = 0.8613
## alternative hypothesis: two-sided

# Mostrar los resultados de la prueba de KS
cat("Resultados de la prueba de Kolmogorov-Smirnov (KS) para la
distribución Gamma:\n")

## Resultados de la prueba de Kolmogorov-Smirnov (KS) para la
distribución Gamma:

cat("Estadístico de prueba:", ks_gamma$statistic, "\n")

## Estadístico de prueba: 0.1050164

cat("p-value:", ks_gamma$p.value, "\n")

## p-value: 0.8613266
```

D.4) ¿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?

```r
# Interpretación de la prueba KS
if (ks_gamma$p.value < 0.05) {
  cat("La prueba de Kolmogorov-Smirnov rechaza la hipótesis nula (los
datos no siguen una distribución Gamma).\n")
} else {
  cat("La prueba de Kolmogorov-Smirnov no rechaza la hipótesis nula (los
datos pueden seguir una distribución Gamma).\n")
}

## La prueba de Kolmogorov-Smirnov no rechaza la hipótesis nula (los
datos pueden seguir una distribución Gamma).
```

¿Qué información nos da la prueba KS para una distribución Gamma? ¿Cuál es el valor del estadístico de prueba? ¿Cuál es el valor del estadístico de prueba? ¿Cuál es el p-value de la prueba? ¿Se acepta o se rechaza la hipótesis nula?

- La prueba de Kolmogorov-Smirnov (KS) se utiliza para comparar la distribución acumulada empírica de los datos con la distribución acumulada teórica de una distribución específica (en este caso, la distribución Gamma). La prueba KS verifica si los datos observados siguen una distribución Gamma ajustada.

Valor del estadístico de prueba: El estadístico de prueba D para esta prueba es 0.10502. Este valor indica la distancia máxima entre la distribución empírica acumulada y la distribución teórica Gamma ajustada. Cuanto menor sea este valor, más cercano es el ajuste de la distribución teórica a los datos empíricos.

p-value de la prueba: El p-value de la prueba es 0.8613, lo que representa la probabilidad de obtener una distancia entre las distribuciones empírica y teórica al menos tan grande como la observada si la hipótesis nula fuera cierta.

- H_0: Los datos observados provienen de una distribución Gamma.
- H_1: Los datos observados no provienen de una distribución Gamma.
- Dado que el p-value es alto, no hay suficiente evidencia para rechazar la hipótesis nula. Esto significa que no podemos descartar que los datos de precipitaciones máximas sigan una distribución Gamma, ya que no se observa una diferencia significativa entre las distribuciones empírica y teórica.

¿Podemos concluir que los datos de las precipitaciones máximas mensuales siguen una distribución Gamma? ¿Por qué?

- los resultados sugieren que los datos de las precipitaciones máximas mensuales podrían seguir una distribución Gamma, ya que no hay evidencia suficiente para rechazar esta hipótesis según la prueba KS. Aunque la prueba KS indica que la distribución Gamma es un ajuste plausible, el análisis visual todavía muestra algunas diferencias en los extremos, por lo que la conclusión no es definitiva, pero el ajuste es razonable.

D.5) ¿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.

```r
# Método de Momentos para estimar los parámetros de la distribución Gamma
# Calcular la media y varianza muestral
mean_precipitation <- mean(chihuahua_data_filtered$Lluvia)
var_precipitation <- var(chihuahua_data_filtered$Lluvia)

# Estimación de los parámetros de la Gamma mediante el método de momentos
shape_moment <- mean_precipitation^2 / var_precipitation
rate_moment <- mean_precipitation / var_precipitation

cat("Parámetro shape estimado por el método de momentos:", shape_moment,
"\n")

## Parámetro shape estimado por el método de momentos: 13.14866

cat("Parámetro rate estimado por el método de momentos:", rate_moment,
"\n")

## Parámetro rate estimado por el método de momentos: 0.1021943
```

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

- La distribución Gamma tiene dos parámetros:

Shape (Forma), denotado como k o **shape**:

- Controla la forma de la distribución. Un valor bajo de k resulta en una distribución más sesgada, mientras que un valor mayor de k hace que la distribución se parezca más a una normal.

Rate (Tasa), denotado como λ o **rate**:

- Controla la escala de la distribución. El parámetro de rate está relacionado con la escala θ como $\lambda = \frac{1}{\theta}$.

*La función de densidad de probabilidad de la distribución Gamma es:

$$f(x;k,\lambda) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)}, \quad x \geq 0$$

- Donde:

k es el parámetro de forma.

λ es el parámetro de tasa.

$(\Gamma(k))$ es la función Gamma, que generaliza el factorial para números reales.

¿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.

Los parámetros se calculan de dos maneras comunes: máxima verosimilitud (MLE) y método de momentos.

Máxima Verosimilitud (MLE):

- El método de máxima verosimilitud (MLE) busca los valores de los parámetros que maximizan la probabilidad de observar los datos dados. En el código se utiliza la función fitdistr() para ajustar una distribución Gamma a los datos. Esto devuelve los estimadores de máxima verosimilitud para los parámetros de forma k y tasa λ.

- MLE es uno de los métodos más precisos para estimar los parámetros porque maximiza la probabilidad de que los datos observados provengan de la distribución Gamma.

Método de Momentos:

- El método de momentos estima los parámetros de la distribución ajustando los momentos de la muestra (media y varianza observadas) a los momentos teóricos de la distribución.

- Para la distribución Gamma:

- La media de la distribución Gamma es:

$$\mu = \frac{k}{\lambda}$$

- La varianza es:

$$\sigma^2 = \frac{k}{\lambda^2}$$

- El método de momentos ajusta estas ecuaciones a los valores de la media y varianza de los datos observados para obtener los parámetros estimados.

Shape (Forma) se calcula como el cuadrado de la media dividido por la varianza porque la forma de la distribución Gamma depende tanto de la dispersión de los datos como de su promedio.

**Rate (Tasa)* se calcula como la media dividida por la varianza porque la tasa controla cómo se distribuyen los datos alrededor de la media.

- El método de momentos ofrece una interpretación más simple de los parámetros basada en los datos observados, mientras que el MLE maximiza la probabilidad de los datos dados.

Pasos del Método de Momentos para la Distribución Gamma

1. La media teórica de la distribución Gamma es:

$$\mu = \frac{k}{\lambda}$$

- donde k es el parámetro de forma y λ es el parámetro de tasa.

2. La varianza teórica de la distribución Gamma es:

$$\sigma^2 = \frac{k}{\lambda^2}$$

3. Igualamos la media muestral y la varianza muestral a las fórmulas de la media y varianza teóricas para resolver los parámetros k (forma) y λ (tasa).

4. A partir de las ecuaciones anteriores, resolvemos los parámetros:

$$k = \frac{\mu^2}{\sigma^2}$$

$$\lambda = \frac{\mu}{\sigma^2}$$

```
# Paso 1: Calcular la media y la varianza de los datos (momentos muestrales)
mean_precipitation <- mean(chihuahua_data_filtered$Lluvia)
var_precipitation <- var(chihuahua_data_filtered$Lluvia)

cat("Momento Muestral - Media de los datos:", mean_precipitation, "\n")

## Momento Muestral - Media de los datos: 128.6633

cat("Momento Muestral - Varianza de los datos:", var_precipitation, "\n")

## Momento Muestral - Varianza de los datos: 1259.007

# Paso 2: Calcular los parámetros de la distribución Gamma usando el método de momentos
shape_moment <- mean_precipitation^2 / var_precipitation
rate_moment <- mean_precipitation / var_precipitation

cat("Parámetro 'shape' estimado por el método de momentos:", shape_moment, "\n")

## Parámetro 'shape' estimado por el método de momentos: 13.14866

cat("Parámetro 'rate' estimado por el método de momentos:", rate_moment, "\n")

## Parámetro 'rate' estimado por el método de momentos: 0.1021943
```

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.

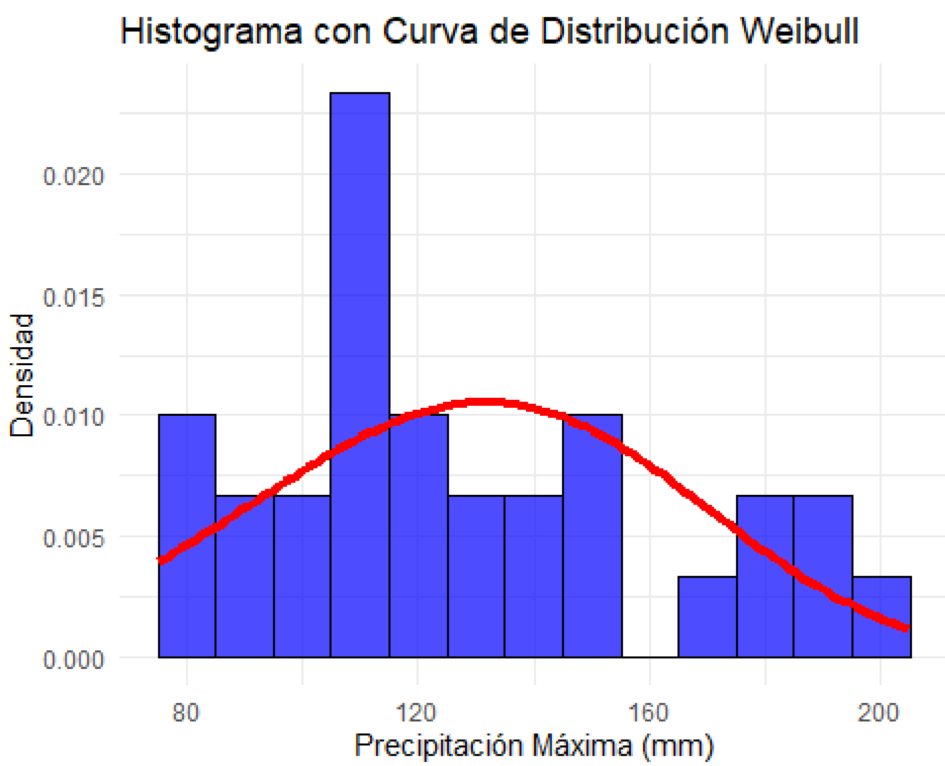
E.1) El cálculo de los parámetros a partir de los datos es un poco más difícil en la distribución Weibull de lo que fue en las anteriores distribuciones, así que recurriremos a que R los estime con el comando fitdistr. Úsalo para estimar los parámetros de la Weibull.

```
# Ajustar la distribución Weibull usando el comando fitdistr
weibull_fit <- fitdistr(chihuahua_data_filtered$Lluvia, "weibull")

# Extraer los parámetros estimados: shape (forma) y scale (escala)
shape_hat <- weibull_fit$estimate[1]
scale_hat <- weibull_fit$estimate[2]

cat("Parámetro shape (forma):", shape_hat, "\n")

## Parámetro shape (forma): 3.930887

cat("Parámetro scale (escala):", scale_hat, "\n")
```

```
## Parámetro scale (escala): 142.0152
```

E.2) 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.

```
# Histograma con curva de distribución Weibull sobrepuesta
ggplot(chihuahua_data_filtered, aes(x = Lluvia)) +
  geom_histogram(aes(y = after_stat(density)), binwidth = 10, fill =
"blue", alpha = 0.7, color = "black") +
  stat_function(fun = dweibull, args = list(shape = shape_hat, scale =
scale_hat), color = "red", size = 1.5) +
  labs(title = "Histograma con Curva de Distribución Weibull",
       x = "Precipitación Máxima (mm)", y = "Densidad") +
  theme_minimal()
```

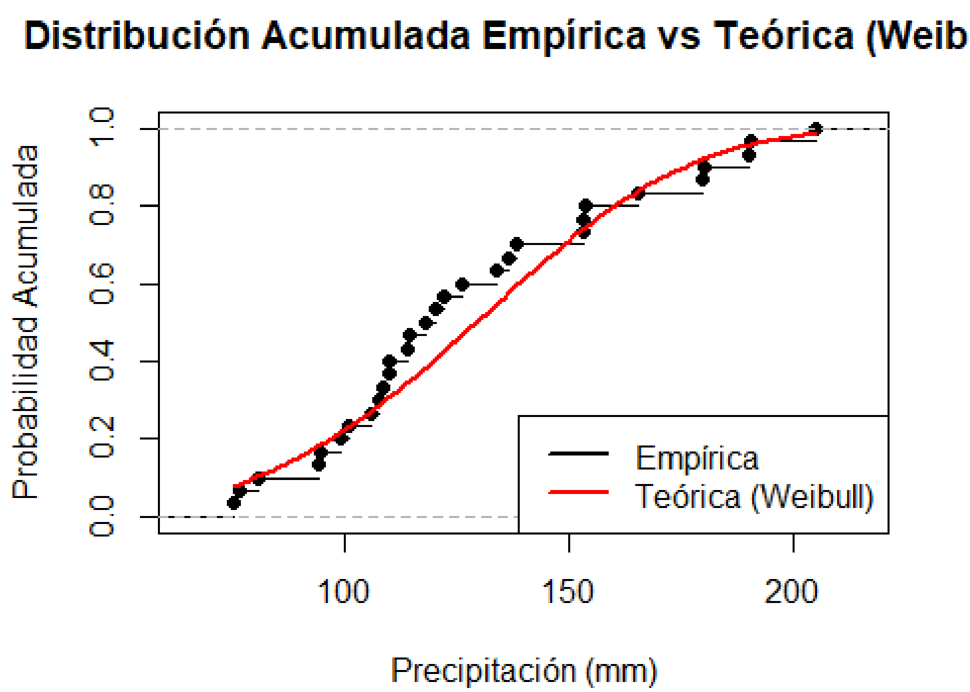


De manera visual, ¿te parece que los datos se ajustan bien a una distribución Weibull? Explica.

- Visualmente, la distribución Weibull no se ajusta bien a los datos. Aunque ofrece un ajuste razonable en el rango medio, subestima las precipitaciones bajas y no captura adecuadamente los eventos extremos en los valores altos.

E.3) 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?

```
# Comparación de CDF empírica vs teórica (ojiva)
# CDF empírica
empirical_cdf <- ecdf(chihuahua_data_filtered$Lluvia)

# Crear un rango de valores para la CDF teórica
x_vals <- seq(min(chihuahua_data_filtered$Lluvia),
max(chihuahua_data_filtered$Lluvia), length.out = 100)

# CDF teórica basada en los parámetros ajustados
theoretical_cdf <- pweibull(x_vals, shape = shape_hat, scale = scale_hat)

# Gráfico de CDF empírica y teórica
plot(empirical_cdf, main = "Distribución Acumulada Empírica vs Teórica
(Weibull)", xlab = "Precipitación (mm)", ylab = "Probabilidad Acumulada")
lines(x_vals, theoretical_cdf, col = "red", lwd = 2)
legend("bottomright", legend = c("Empírica", "Teórica (Weibull)"), col =
c("black", "red"), lwd = 2)
```



Explica qué son datos empíricos y datos teóricos.

Datos empíricos: Son los datos observados directamente de las precipitaciones máximas registradas. Estos reflejan las observaciones reales o medidas experimentales.

Datos teóricos: Representan el comportamiento que se esperaría si los datos siguieran perfectamente la distribución teórica Weibull con los parámetros calculados (forma y escala).

¿Se parecen las distribuciones de probabilidad acumuladas?

- Las distribuciones acumuladas se parecen en el rango medio, pero presentan diferencias en los extremos. En los valores bajos, la distribución Weibull subestima las precipitaciones, mientras que en los valores altos sigue más de cerca los datos empíricos. En general, la distribución Weibull ofrece un ajuste razonable, pero no perfecto.

E.4) Haz la prueba KS para determinar si los datos se ajustan a una Weibull

```r
# Prueba de bondad de ajuste: Kolmogorov-Smirnov para Weibull
ks_weibull <- ks.test(chihuahua_data_filtered$Lluvia, "pweibull", shape =
shape_hat, scale = scale_hat)
ks_weibull

##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  chihuahua_data_filtered$Lluvia
## D = 0.1393, p-value = 0.5583
## alternative hypothesis: two-sided

# Mostrar los resultados de la prueba de KS
cat("Resultados de la prueba de Kolmogorov-Smirnov (KS) para la
distribución Weibull:\n")

## Resultados de la prueba de Kolmogorov-Smirnov (KS) para la
distribución Weibull:

cat("Estadístico de prueba:", ks_weibull$statistic, "\n")

## Estadístico de prueba: 0.1393037

cat("p-value:", ks_weibull$p.value, "\n")

## p-value: 0.5583486
```

E.5) ¿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?

```r
# Interpretación de la prueba KS
if (ks_weibull$p.value < 0.05) {
  cat("La prueba de Kolmogorov-Smirnov rechaza la hipótesis nula (los
```

```
datos no siguen una distribución Weibull).\n")
} else {
  cat("La prueba de Kolmogorov-Smirnov no rechaza la hipótesis nula (los
datos pueden seguir una distribución Weibull).\n")
}

## La prueba de Kolmogorov-Smirnov no rechaza la hipótesis nula (los
datos pueden seguir una distribución Weibull).
```

¿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?

- La prueba de Kolmogorov-Smirnov (KS) para la distribución Weibull nos proporciona información sobre la bondad de ajuste entre los datos empíricos y la distribución Weibull teórica ajustada.

Valor del estadístico de prueba (D): 0.1393. Este valor mide la máxima diferencia absoluta entre las distribuciones empírica y teórica. Cuanto más bajo sea este valor, mejor es el ajuste.

p-value: 0.5583, que es mayor que el nivel de significancia típico de 0.05.

H_0: Los datos siguen una distribución Weibull.

H_1: Los datos no siguen una distribución Weibull.

- Dado que el p-value es alto, no hay suficiente evidencia para rechazar la hipótesis nula. Esto significa que no podemos descartar que los datos de precipitaciones sigan una distribución Weibull.

¿Podemos concluir que los datos de las precipitaciones máximas mensuales siguen una distribuciónWeibull? ¿Por qué?

- Los resultados de la prueba KS indican que los datos de las precipitaciones máximas mensuales podrían seguir una distribución Weibull, ya que no hay suficiente evidencia para rechazar esta hipótesis. Aunque la prueba sugiere que Weibull es un ajuste razonable, todavía se observan algunas diferencias en el ajuste visual, por lo que el ajuste no es completamente perfecto.

E.6) ¿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? ¿Cuáles son?

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

1. **Parámetro de forma k**: Este parámetro define la forma de la distribución. Dependiendo de su valor, la distribución Weibull puede tomar diferentes formas:

- Si ($k < 1$), la distribución tiene una alta probabilidad de valores pequeños, y su curva cae empinadamente al principio.
- Si ($k = 1$), la distribución Weibull es equivalente a una distribución exponencial.
- Si ($k > 1$), la distribución tiene una curva más suavizada, lo que refleja una menor probabilidad de valores bajos y un aumento gradual antes de disminuir.

2. **Parámetro de escala** λ: Este parámetro define la “escala” o el “rango” de la distribución, controlando el estiramiento o compresión de la curva. Si λ es grande, la distribución estará más extendida (más “ancha”), y si λ es pequeña, estará más concentrada en valores pequeños.

Explica por qué es más compleja la estimación de los parámetros a partir de los datos en esta distribución a diferencia de las distribuciones anteriores.

La estimación de los parámetros de la distribución Weibull es más compleja que en otras distribuciones como la normal o exponencial, y esto se debe a varios factores:

1. **No linealidad**: A diferencia de la distribución normal, la distribución Weibull involucra una relación no lineal entre los parámetros y la densidad de probabilidad. El parámetro de forma k afecta de manera no lineal tanto la varianza como la media, lo que complica su estimación.
2. **Interdependencia de los parámetros**: Los parámetros de forma k y de escala λ están interrelacionados. Un cambio en uno de los parámetros afecta la probabilidad global de los valores, lo que hace que la estimación simultánea de ambos parámetros sea difícil.
3. **Métodos de estimación**: Para distribuciones como la normal o exponencial, la media y la varianza se pueden estimar directamente a partir de los datos. Sin embargo, para la distribución Weibull, los parámetros no tienen fórmulas simples y directas basadas en los momentos (media y varianza), lo que requiere métodos numéricos como la **Máxima Verosimilitud (MLE)** o el **método de mínimos cuadrados**.
4. **Sensibilidad de k**: El parámetro de forma k afecta considerablemente la forma de la curva y, por lo tanto, pequeños errores en la estimación de k pueden resultar en grandes desviaciones en el ajuste de la distribución.

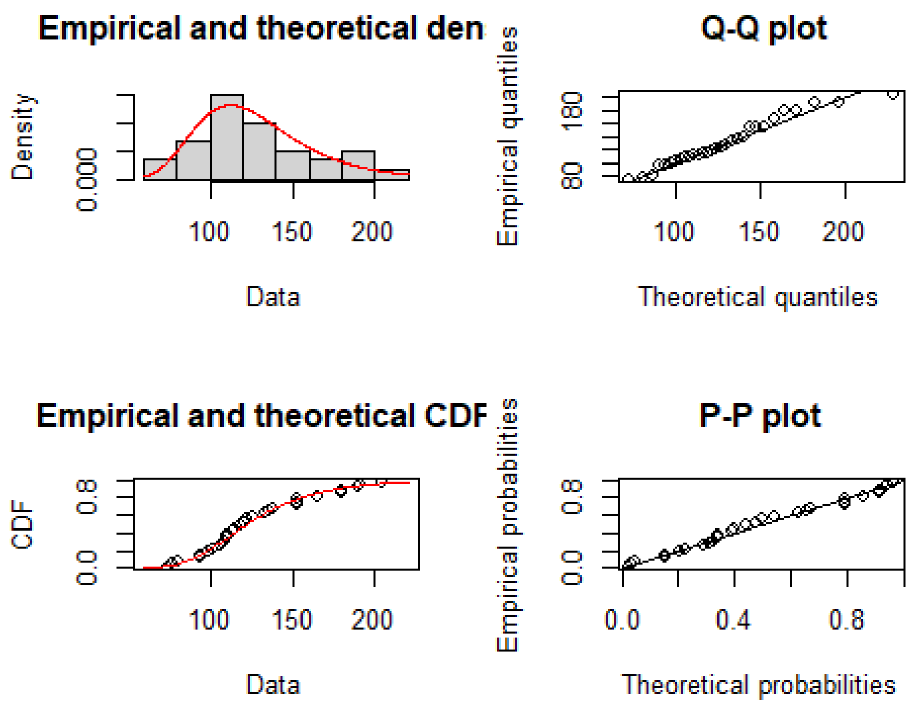
F) Ajuste a una Distribución Gumbel. Realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Gumbel para ajustar los datos.

F.1) 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 las funciones para la distribución Gumbel
dgumbel <- function(x, a, b) {
  return(1 / b * exp((a - x) / b) * exp(-exp((a - x) / b)))
}

pgumbel <- function(q, a, b) {
  return(exp(-exp((a - q) / b)))
}

qgumbel <- function(p, a, b) {
  return(a - b * log(-log(p)))
}
```

F.2) 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
# Estimar los parámetros de la distribución Gumbel usando fitdist
(fitdistrplus)

# Ajustar la distribución Gumbel a los datos utilizando fitdist y
funciones definidas por el usuario
gumbel_fit <- fitdist(chihuahua_data_filtered$Lluvia, "gumbel", start =
list(a = 1, b = 1))

# Extraer los parámetros estimados: a (localización) y b (escala)
a_hat <- gumbel_fit$estimate[1]
b_hat <- gumbel_fit$estimate[2]

cat("Parámetro a (localización):", a_hat, "\n")

## Parámetro a (localización): 112.268

cat("Parámetro b (escala):", b_hat, "\n")

## Parámetro b (escala): 28.35365

# Generar los gráficos asociados (histograma y densidad)
plot(gumbel_fit)
```



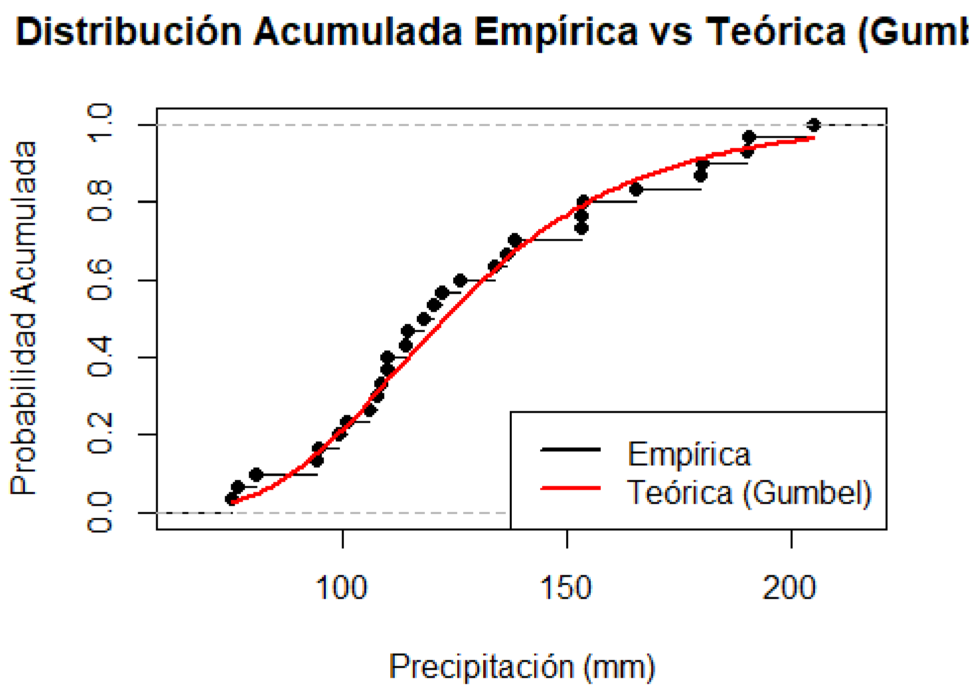
Interpretación

Visualmente, se muestra que el ajuste teórico es razonablemente bueno en general, aunque con algunas desviaciones en los valores extremos.

- Histograma: La densidad empírica de los datos, y la curva roja corresponde a la densidad teórica ajustada. Se observa un ajuste razonable, aunque no perfecto, especialmente en los extremos.
- Q-Q plot y P-P plot: Ambos gráficos muestran pequeñas discrepancias en los valores extremos, lo que confirma que la distribución Gumbel no captura por comportamiento de los datos.

```
# Comparación de CDF empírica vs teórica (ojiva)

# CDF empírica
empirical_cdf <- ecdf(chihuahua_data_filtered$Lluvia)

# Crear un rango de valores para la CDF teórica
x_vals <- seq(min(chihuahua_data_filtered$Lluvia),
max(chihuahua_data_filtered$Lluvia), length.out = 100)

# CDF teórica basada en los parámetros ajustados
theoretical_cdf <- pgumbel(x_vals, a_hat, b_hat)

# Gráfico de CDF empírica y teórica
plot(empirical_cdf, main = "Distribución Acumulada Empírica vs Teórica
```

```
(Gumbel)", xlab = "Precipitación (mm)", ylab = "Probabilidad Acumulada")
lines(x_vals, theoretical_cdf, col = "red", lwd = 2)
legend("bottomright", legend = c("Empírica", "Teórica (Gumbel)"), col =
c("black", "red"), lwd = 2)
```



Interpretación

- La distribución Gumbel ofrece un ajuste aceptable para el rango medio de los datos, pero no captura perfectamente los extremos, lo que indica que no es completamente adecuada para modelar todos los datos.

F.3) Haz la prueba KS para determinar si los datos se ajustan a una Gumbel

```
# Prueba de bondad de ajuste: Kolmogorov-Smirnov para Gumbel

# Calculamos la probabilidad de excedencia teórica
gumbel_exe <- 1 - pgumbel(chihuahua_data_filtered$Lluvia, a_hat, b_hat)

# Prueba KS comparando las probabilidades de excedencia empíricas vs
teóricas
ks_gumbel <- ks.test(chihuahua_data_filtered$Lluvia, "pgumbel", a_hat,
b_hat)
cat("Resultados de la prueba de Kolmogorov-Smirnov:\n")

## Resultados de la prueba de Kolmogorov-Smirnov:

cat("Estadístico de prueba:", ks_gumbel$statistic, "\n")

## Estadístico de prueba: 0.08906633
```

```
cat("p-value:", ks_gumbel$p.value, "\n")

## p-value: 0.9541215

# Interpretar el resultado de la prueba KS
if (ks_gumbel$p.value < 0.05) {
  cat("Rechazamos la hipótesis nula: Los datos no siguen una distribución Gumbel.\n")
} else {
  cat("No se rechaza la hipótesis nula: Los datos podrían seguir una distribución Gumbel.\n")
}

## No se rechaza la hipótesis nula: Los datos podrían seguir una distribución Gumbel.
```

F.4) ¿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?

- La prueba de Kolmogorov-Smirnov (KS) para la distribución Gumbel nos proporciona información clave sobre la similitud entre los datos observados y una distribución teórica (en este caso, la distribución Gumbel).

- Estadístico de prueba (D): 0.0891, lo que indica una pequeña diferencia entre la distribución acumulada empírica y la teórica Gumbel. Un valor bajo de D sugiere que las distribuciones se ajustan bien.

- p-value: El p-value es 0.9541, lo que es significativamente mayor que el nivel de significancia típico de 0.05.

- No se rechaza la hipótesis nula, lo que indica que los datos podrían seguir una distribución Gumbel.

- Con base en la prueba KS y el alto p-value, podemos concluir que las probabilidades de excedencia de las precipitaciones máximas mensuales podrían seguir una distribución Gumbel, ya que no se rechaza la hipótesis nula. La distribución Gumbel parece ser un buen ajuste para estos datos.

F.5) ¿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?

La distribución Gumbel tiene **dos parámetros**:

1. **Parámetro de localización** α o μ: Este parámetro define la ubicación del máximo de la distribución.

2. **Parámetro de escala** b o β: Este parámetro determina la dispersión o el “ancho” de la distribución.

Función de densidad de probabilidad de la distribución Gumbel:

$$f(x; a, b) = \frac{1}{b} \exp\left(-\frac{x-a}{b} - \exp\left(-\frac{x-a}{b}\right)\right)$$

Donde: - a es el parámetro de localización. - b es el parámetro de escala.

Fórmulas para la media y la desviación estándar de la distribución Gumbel:

- La media μ y la desviación estándar σ de la distribución Gumbel están relacionadas con los parámetros de localización a y escala b de la siguiente manera:

Media μ de la distribución Gumbel:

$$\mu = a + b \cdot \gamma$$

- donde ($\gamma \approx 0.5772$) es la constante de Euler-Mascheroni.

Desviación estándar σ de la distribución Gumbel:

$$\sigma = \frac{\pi}{\sqrt{6}} \cdot b$$

Cálculo de los parámetros a y b a partir de la media y desviación estándar:

- Sabemos que:
- La media de la distribución Gumbel es μ.
- La desviación estándar es σ.
- Podemos despejar los parámetros a y b utilizando estas fórmulas:

Despejar el parámetro de escala b:

$$b = \frac{\sigma \cdot \sqrt{6}}{\pi}$$

Despejar el parámetro de localización a:

$$a = \mu - b \cdot \gamma$$

```r
# Calcular la media y la desviación estándar de los datos
mean_precipitation <- mean(chihuahua_data_filtered$Lluvia)
sd_precipitation <- sd(chihuahua_data_filtered$Lluvia)

# Constante de Euler-Mascheroni
gamma <- 0.5772
```

```r
# Estimar el parámetro b (escala) a partir de la desviación estándar
b_est <- sd_precipitation / (pi / sqrt(6))

# Estimar el parámetro a (localización) a partir de la media
a_est <- mean_precipitation - b_est * gamma

cat("Parámetro a estimado a partir de la media y desviación estándar:",
a_est, "\n")

## Parámetro a estimado a partir de la media y desviación estándar:
112.6948

cat("Parámetro b estimado a partir de la media y desviación estándar:",
b_est, "\n")

## Parámetro b estimado a partir de la media y desviación estándar:
27.66559

# Comparar con los valores obtenidos con fitdistrplus
cat("Parámetro a estimado con fitdistrplus:", a_hat, "\n")

## Parámetro a estimado con fitdistrplus: 112.268

cat("Parámetro b estimado con fitdistrplus:", b_hat, "\n")

## Parámetro b estimado con fitdistrplus: 28.35365

# Comparar las diferencias
cat("Diferencia en el parámetro a:", abs(a_hat - a_est), "\n")

## Diferencia en el parámetro a: 0.4267547

cat("Diferencia en el parámetro b:", abs(b_hat - b_est), "\n")

## Diferencia en el parámetro b: 0.6880676
```

Método basado en la media y desviación estándar:

- Este método utiliza las fórmulas de la media y desviación estándar para despejar los parámetros de localización y escala de la distribución Gumbel. Es un método aproximado basado en los momentos de la muestra (media y varianza).

Ajuste por máxima verosimilitud (MLE) con fitdistrplus:

- El ajuste con MLE intenta maximizar la probabilidad de observar los datos dados los parámetros de la distribución. Este método ajusta los parámetros de manera que se maximice la probabilidad de los datos observados, lo que generalmente produce estimaciones más precisas.

Elección del método:

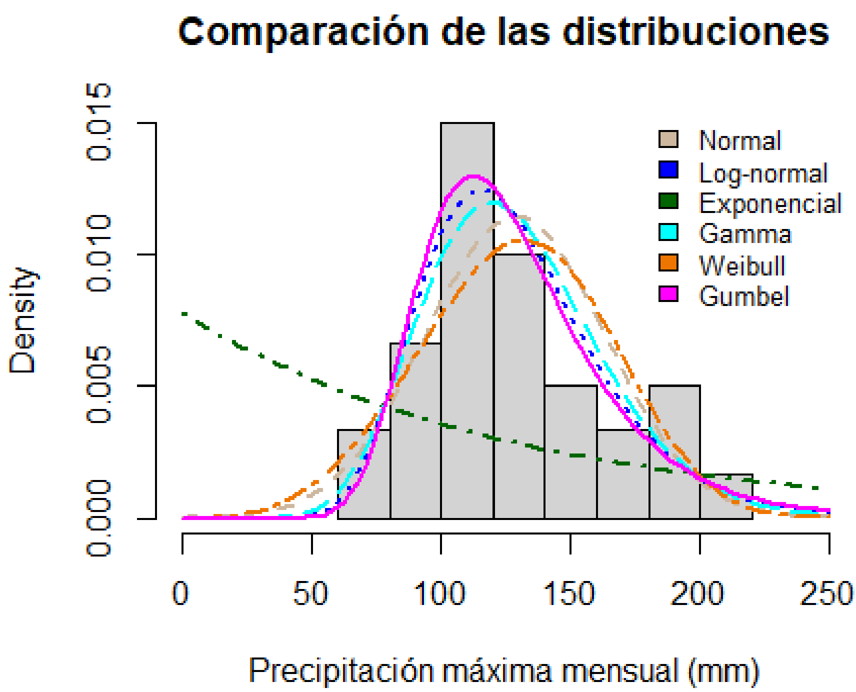
- En general, el ajuste por MLE tiende a ser más preciso cuando el objetivo es obtener un buen ajuste de la distribución a los datos. El método de momentos es útil como una estimación rápida o aproximada de los parámetros.

G) Compara los ajustes de las distribuciones que analizaste

G.1) 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.

```
# Ajustar varias distribuciones
# Normal
normal_fit <- fitdist(chihuahua_data_filtered$Lluvia, "norm")

# Log-Normal
lognormal_fit <- fitdist(chihuahua_data_filtered$Lluvia, "lnorm")

# Exponencial
exp_fit <- fitdist(chihuahua_data_filtered$Lluvia, "exp")

# Gamma
gamma_fit <- fitdist(chihuahua_data_filtered$Lluvia, "gamma")

# Weibull
weibull_fit <- fitdist(chihuahua_data_filtered$Lluvia, "weibull")

# Gumbel
gumbel_fit <- fitdist(chihuahua_data_filtered$Lluvia, "gumbel", start =
list(a=1, b=1))

# Comparación de histogramas de densidad empírica vs teórica
# Ajustar los márgenes
par(mar = c(5, 5, 4, 2))

# Ajustar el tamaño del gráfico y los límites del eje x
hist(chihuahua_data_filtered$Lluvia, xlab = "Precipitación máxima mensual
(mm)", freq = FALSE, ylim = c(0, 0.015), xlim = c(0, 250), main =
"Comparación de las distribuciones", col = "lightgray")
curve(dnorm(x, mean = normal_fit$estimate[1], sd =
normal_fit$estimate[2]), add = TRUE, col = "bisque3", lwd = 2, lty = 2)
curve(dlnorm(x, mean = lognormal_fit$estimate[1], sd =
lognormal_fit$estimate[2]), add = TRUE, col = "blue", lwd = 2, lty = 3)
curve(dexp(x, rate = exp_fit$estimate[1]), add = TRUE, col = "darkgreen",
lwd = 2, lty = 4)
curve(dgamma(x, shape = gamma_fit$estimate[1], rate =
gamma_fit$estimate[2]), add = TRUE, col = "cyan", lwd = 2, lty = 5)
curve(dweibull(x, shape = weibull_fit$estimate[1], scale =
weibull_fit$estimate[2]), add = TRUE, col = "darkorange2", lwd = 2, lty =
6)
curve(dgumbel(x, gumbel_fit$estimate[1], gumbel_fit$estimate[2]), add =
```

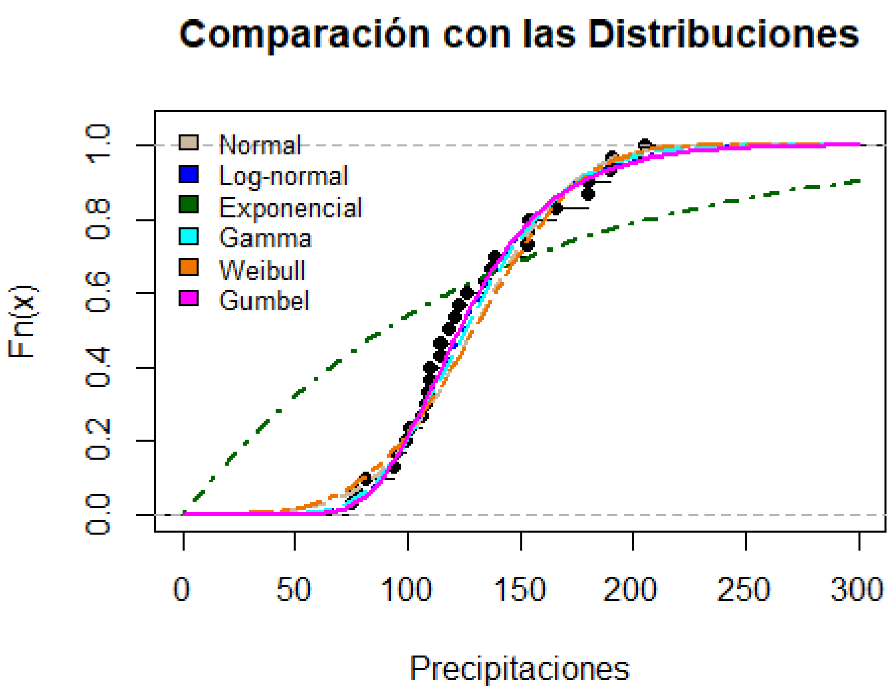
```
TRUE, col = "magenta", lwd = 2, lty = 7)

# Leyenda
legend("topright", legend = c("Normal", "Log-normal", "Exponencial",
"Gamma", "Weibull", "Gumbel"),
       fill = c("bisque3", "blue", "darkgreen", "cyan", "darkorange2",
"magenta"),
       cex = 0.8, bty = "n")
```



G.2)Haz un gráfico comparativo de las probabilidades acumuladas empírica vs teóricas de todas las distribuciones que analizaste (todas las distribuciones en un solo gráfico.

```
# Comparación de las probabilidades acumuladas empíricas vs teóricas
plot(ecdf(chihuahua_data_filtered$Lluvia), main = "Comparación con las
Distribuciones", xlab = "Precipitaciones", col = "black", xlim = c(0,
300), ylim = c(0, 1.05))
curve(pnorm(x, mean = normal_fit$estimate[1], sd =
normal_fit$estimate[2]), add = TRUE, col = "bisque3", lwd = 2, lty = 2)
curve(plnorm(x, mean = lognormal_fit$estimate[1], sd =
lognormal_fit$estimate[2]), add = TRUE, col = "blue", lwd = 2, lty = 3)
curve(pexp(x, rate = exp_fit$estimate[1]), add = TRUE, col = "darkgreen",
lwd = 2, lty = 4)
curve(pgamma(x, shape = gamma_fit$estimate[1], rate =
gamma_fit$estimate[2]), add = TRUE, col = "cyan", lwd = 2, lty = 5)
curve(pweibull(x, shape = weibull_fit$estimate[1], scale =
weibull_fit$estimate[2]), add = TRUE, col = "darkorange2", lwd = 2, lty =
6)
```

```
curve(pgumbel(x, gumbel_fit$estimate[1], gumbel_fit$estimate[2]), add =
TRUE, col = "magenta", lwd = 2, lty = 7)

# Leyenda para el gráfico de probabilidades acumuladas
legend("topleft", legend = c("Normal", "Log-normal", "Exponencial",
"Gamma", "Weibull", "Gumbel"),
       fill = c("bisque3", "blue", "darkgreen", "cyan", "darkorange2",
"magenta"),
       cex = 0.8, bty = "n")
```




```
# Crear un dataframe para almacenar los resultados
resultados_ks <- data.frame(
  Distribución = c("Normal", "Log-normal", "Exponencial", "Gamma",
"Weibull", "Gumbel"),
  KS_Statistic = c(ks_normal$statistic, ks_log$statistic,
ks_exp$statistic, ks_gamma$statistic, ks_weibull$statistic,
ks_gumbel$statistic),
  p_value = c(ks_normal$p.value, ks_log$p.value, ks_exp$p.value,
ks_gamma$p.value, ks_weibull$p.value, ks_gumbel$p.value)
)

# Mostrar la tabla con los resultados
print(resultados_ks)

##   Distribución KS_Statistic        p_value
## 1       Normal   0.13672355 0.581844272797
## 2   Log-normal   0.08743888 0.960517144639
```

```
## 3  Exponencial   0.44259941 0.000007148288
## 4        Gamma   0.10501644 0.861326558653
## 5      Weibull   0.13930375 0.558348575215
## 6       Gumbel   0.08906633 0.954121487414
```

G.3)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.

Pruebas de bondad de ajuste:

- Según la tabla de resultados de la prueba de Kolmogorov-Smirnov (KS), la distribución que muestra el valor más bajo de KS_Statistic es la Log-normal (0.08743888), con un p-value alto (0.960517144639), lo que significa que no se rechaza la hipótesis nula de que los datos siguen una distribución Log-normal. Además, la distribución Gumbel también tiene un p-value alto y un KS_Statistic bajo (0.08906633), siendo la segunda mejor opción.

Gráficas de ajuste:

- Al observar las gráficas, tanto la Log-normal como la Gumbel muestran un buen ajuste en el rango medio y alto de las precipitaciones, especialmente en la distribución acumulada empírica frente a la teórica.

- La distribución Log-normal es la que mejor se ajusta a los datos de precipitaciones máximas mensuales, tanto por los resultados de la prueba KS como por las gráficas de ajuste.

4) DISEÑO DE OBRAS HIDRÁULICAS. 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.

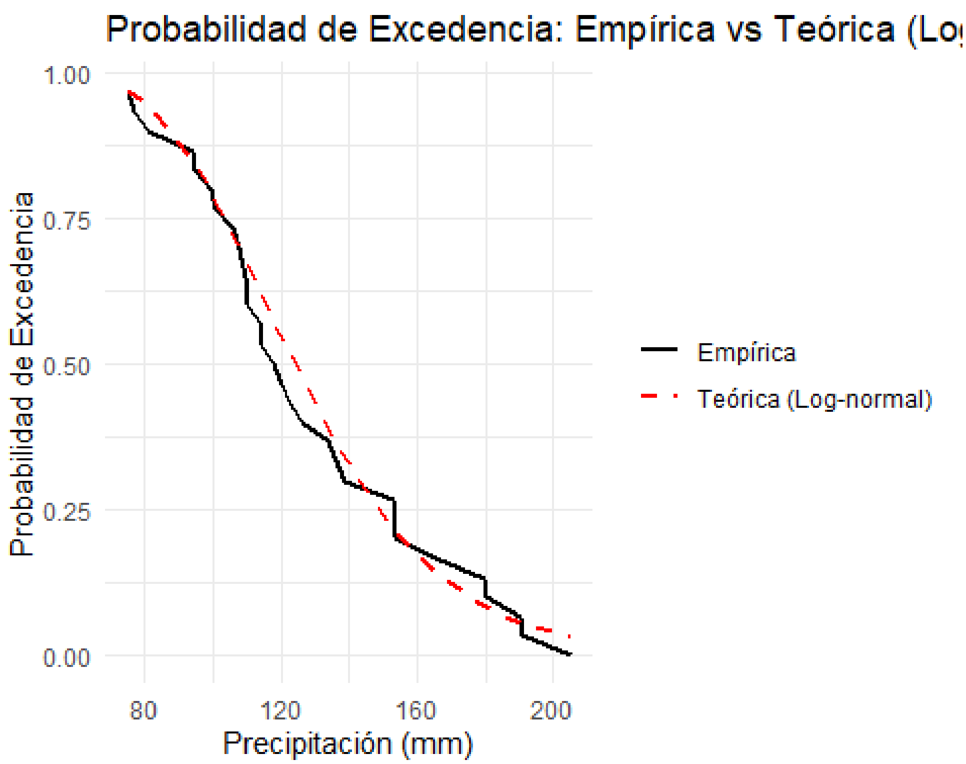
Para el diseño de una presa derivadora en una zona de riego mediana, el periodo de retorno recomendado es de 100 a 500 años. Este rango está diseñado para garantizar la seguridad y eficacia de la presa en la gestión de los caudales de agua para riego en áreas de tamaño medio (1,000 a 10,000 Ha).

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

- Para realizar el gráfico comparativo de la probabilidad de excedencia teórica vs empírica y evaluar la selección de la distribución elegida, se debe calcular la probabilidad de excedencia empírica a partir de los datos observados, y compararla con la probabilidad de excedencia teórica de la distribución elegida anteriormente (Log-normal).

```
# Definir la función de probabilidad de excedencia empírica
empirical_exceedance <- 1 -
```

```r
ecdf(chihuahua_data_filtered$Lluvia)(chihuahua_data_filtered$Lluvia)

# Calcular la probabilidad de excedencia teórica usando la distribución
Log-normal ajustada
theoretical_exceedance_lognormal <- 1 -
plnorm(chihuahua_data_filtered$Lluvia, meanlog = mu_hat, sdlog =
sigma_hat)

# Crear un dataframe para graficar
exceedance_data <- data.frame(
  Precipitation = chihuahua_data_filtered$Lluvia,
  Empirical_Exceedance = empirical_exceedance,
  Theoretical_Exceedance = theoretical_exceedance_lognormal
)

# Graficar las probabilidades de excedencia empírica vs teórica
(distribución Log-normal)
ggplot(exceedance_data, aes(x = Precipitation)) +
  geom_line(aes(y = Empirical_Exceedance, color = "Empírica"), size = 1)
+    # Probabilidad de excedencia empírica
  geom_line(aes(y = Theoretical_Exceedance, color = "Teórica (Log-
normal)"), size = 1, linetype = "dashed") +  # Probabilidad de excedencia
teórica
  labs(title = "Probabilidad de Excedencia: Empírica vs Teórica (Log-
normal)",
       x = "Precipitación (mm)",
       y = "Probabilidad de Excedencia") +
  scale_color_manual(values = c("Empírica" = "black", "Teórica (Log-
normal)" = "red")) +
  theme_minimal() +
  theme(legend.title = element_blank())
```



¿Qué te indica ese gráfico?

- El gráfico muestra la comparación entre la probabilidad de excedencia empírica (curva negra) y la probabilidad de excedencia teórica basada en la distribución Log-normal (curva roja discontinua).
- En el rango medio de los datos (entre aproximadamente 80 mm y 160 mm), las curvas están bastante alineadas. Esto indica que la distribución Log-normal se ajusta bien a los datos empíricos en esta región. En los valores extremos (tanto para precipitaciones bajas como para precipitaciones altas), la curva teórica de la distribución Log-normal tiende a desviarse ligeramente de los valores empíricos. Esto sugiere que la Log-normal no captura perfectamente los eventos extremos, pero las diferencias son relativamente pequeñas.

Interpreta y argumenta la certeza de la selección de la distribución elegida.

- La distribución Log-normal parece ser una buena elección para modelar las precipitaciones máximas mensuales, ya que el gráfico muestra que las probabilidades de excedencia empírica y teórica están alineadas de manera consistente.
- Las pruebas de bondad de ajuste, incluyendo el p-valor de la prueba KS, respaldan esta selección. Por lo tanto, esta la distribución es adecuada para describir las probabilidades de excedencia de las precipitaciones máximas mensuales en este caso.

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

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

- Para encontrar la probabilidad de excedencia utilizando el límite inferior del periodo de retorno sugerido (en este caso, 100 años), usamos la fórmula:

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

Despejando para la probabilidad de excedencia (P_{exe}):

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

Podemos calcular la probabilidad de excedencia para un periodo de retorno de 100 años.

```
# Definir el periodo de retorno sugerido (límite inferior)
periodo_retorno_inferior <- 100  # 100 años

# Calcular la probabilidad de excedencia (P_exe)
probabilidad_excedencia <- 1 / periodo_retorno_inferior

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

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

- Hay una probabilidad del 1% de que el evento de diseño (precipitación) sea excedido en cualquier año dado. Este valor de probabilidad de excedencia se puede utilizar para evaluar la seguridad y la eficacia de la obra hidráulica, en este caso, una presa derivadora.

C) Conociendo la probabilidad de excedencia, calcula su complemento (1 - Pexe) y utiliza esta probabilidad para encontrar el valor de la precipitación máxima mensual que tendrá ese periodo de retorno.

```
# Calcular el complemento de la probabilidad de excedencia
probabilidad_no_excedencia <- 1 - probabilidad_excedencia

# Utilizar la distribución Log-normal ajustada para encontrar el valor de
precipitación máxima mensual
# Usar la función cuantil (qlnorm) de la distribución Log-normal ajustada
precipitacion_maxima <- qlnorm(probabilidad_no_excedencia, meanlog =
lognormal_fit$estimate[1], sdlog = lognormal_fit$estimate[2])
```

```r
# Mostrar el resultado
cat("La precipitación máxima mensual para un periodo de retorno de",
periodo_retorno_inferior, "años es:", precipitacion_maxima, "mm\n")
```

```
## La precipitación máxima mensual para un periodo de retorno de 100 años
es: 231.399 mm
```

- Usamos la CDF inversa (cuantil) de la distribución Log-normal ajustada para encontrar el valor de la precipitación correspondiente a la probabilidad de no excedencia. La función qgamma() calcula este valor utilizando los parámetros shape y rate ajustados previamente con fitdist()

D) ¿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.

¿Qué significa este valor?

- El valor de la precipitación máxima mensual para un periodo de retorno, como el que se calculó para 100 años, nos indica la cantidad de precipitación que se espera que ocurra en promedio una vez cada 100 años. Es un valor importante en el diseño de obras hidráulicas, como presas, canales y sistemas de drenaje, ya que estas infraestructuras deben ser capaces de manejar eventos extremos de precipitación para evitar daños por inundaciones.

¿Qué pasa si incrementamos el periodo de retorno?

- Si incrementamos el periodo de retorno, es decir, consideramos un evento que ocurre menos frecuentemente, el valor de la precipitación máxima mensual aumentará, ya que los eventos más raros tienden a ser más extremos. Esto es crucial para asegurar que las obras estén preparadas para eventos de mayor magnitud, aunque menos probables.

¿El caudal máximo para este periodo de retorno será el mismo si utilizamos datos históricos de otro estado?

- Si utilizamos datos históricos de otro estado, el caudal máximo para el mismo periodo de retorno podría ser diferente. Las condiciones climáticas, geográficas y ambientales varían entre regiones, lo que afecta los patrones de precipitación y, por ende, la magnitud de los eventos extremos. Esto subraya la importancia de utilizar datos locales para el diseño de infraestructuras.

¿Por qué crees que las obras hidráulicas deben diseñarse en base a periodos de retorno sugeridos?

- Diseñar en base a periodos de retorno sugeridos es vital para garantizar la seguridad y funcionalidad de las obras hidráulicas a lo largo del tiempo. Si se diseña solo para eventos comunes, el sistema podría fallar en eventos extremos, causando grandes daños. Por otro lado, sobredimensionar las obras para eventos extremadamente raros podría ser innecesariamente costoso.

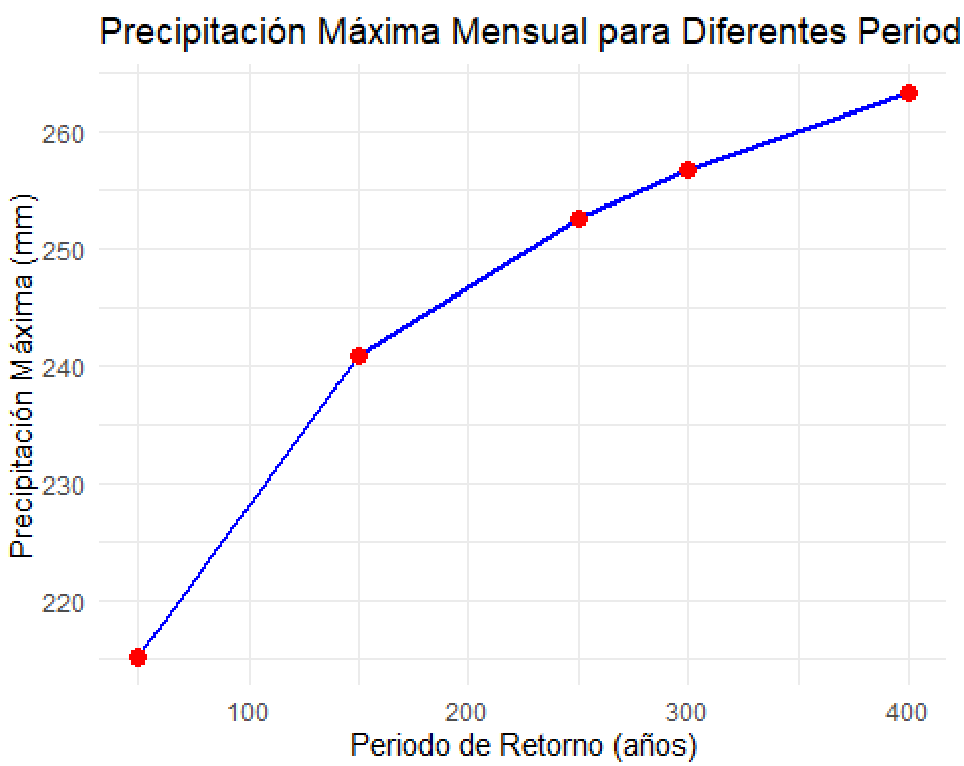
¿Por qué es importante conocer la distribución de probabilidad a la que mejor se aproximan los datos históricos? Explora otros periodos de retorno diferentes a los que se proporcionan en los periodos sugeridos para contestar esta pregunta.

```r
# Definir una lista de periodos de retorno para explorar
periodos_retorno <- c(50, 150, 250, 300, 400)  # Puedes agregar más valores aquí

# Crear un vector vacío para almacenar los resultados de las precipitaciones máximas
precipitaciones_maximas <- c()

# Calcular la precipitación máxima para cada periodo de retorno
for (periodo in periodos_retorno) {

  # Calcular la probabilidad de excedencia (P_exe)
  probabilidad_excedencia <- 1 / periodo

  # Calcular el complemento de la probabilidad de excedencia (P_no_exce)
  probabilidad_no_excedencia <- 1 - probabilidad_excedencia

  # Utilizar la distribución Log-normal ajustada para encontrar la precipitación máxima mensual
  precipitacion_maxima <- qlnorm(probabilidad_no_excedencia, meanlog = lognormal_fit$estimate[1], sdlog = lognormal_fit$estimate[2])

  # Guardar el resultado en el vector
  precipitaciones_maximas <- c(precipitaciones_maximas, precipitacion_maxima)
}

# Crear un dataframe con los periodos de retorno y sus precipitaciones máximas
resultados <- data.frame(Periodo_de_Retorno = periodos_retorno, Precipitacion_Maxima = precipitaciones_maximas)

# Mostrar los resultados
resultados
```

Periodo_de_Retorno	Precipitacion_Maxima
50	215.1106

Periodo_de_Retorno	Precipitacion_Maxima
150	240.7783
250	252.4866
300	256.6432
400	263.1835

```r
# Visualizar los resultados con un gráfico
library(ggplot2)
ggplot(resultados, aes(x = Periodo_de_Retorno, y = Precipitacion_Maxima))
+
  geom_line(color = "blue", size = 1) +
  geom_point(color = "red", size = 3) +
  labs(title = "Precipitación Máxima Mensual para Diferentes Periodos de
Retorno",
       x = "Periodo de Retorno (años)",
       y = "Precipitación Máxima (mm)") +
  theme_minimal()
```



- Conocer la distribución de probabilidad a la que mejor se ajustan los datos históricos es esencial para realizar predicciones precisas sobre la frecuencia y magnitud de eventos extremos. Explorando otros periodos de retorno, como los que se calcularon en la tabla, podemos observar cómo la precipitación máxima aumenta conforme el periodo de retorno se incrementa.

Conclusión final El uso de métodos estadísticos avanzados para el ajuste de distribuciones y la comparación de probabilidades de excedencia resultó en una selección confiable de la distribución Log-normal para modelar las precipitaciones máximas en la región de estudio. Los cálculos realizados para distintos periodos de retorno proporcionan información esencial para la planificación y diseño de infraestructuras hidráulicas robustas y seguras, asegurando que la presa derivadora será capaz de gestionar adecuadamente las precipitaciones extremas, minimizando el riesgo de fallos estructurales y garantizando un suministro de agua seguro para la zona de riego.

El análisis realizado establece una sólida base para el diseño de obras hidráulicas en la región, lo que demuestra el valor de integrar técnicas estadísticas rigurosas en el proceso de toma de decisiones en ingeniería civil.