Actividad Integradora 1: Precipitaciones máximas mensuales para el diseño de obras hidráulicas

Daniela Jiménez Téllez

2024-10-22

Problema

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

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

Deberás investigar algunas cuestiones como los parámetros de las distribuciones.

Análisis

Importación de librerías

```
library(car)
```

```
## Cargando paquete requerido: carData
```

```
library(MASS)
library(moments)
library(fitdistrplus)
```

```
## Cargando paquete requerido: survival
```

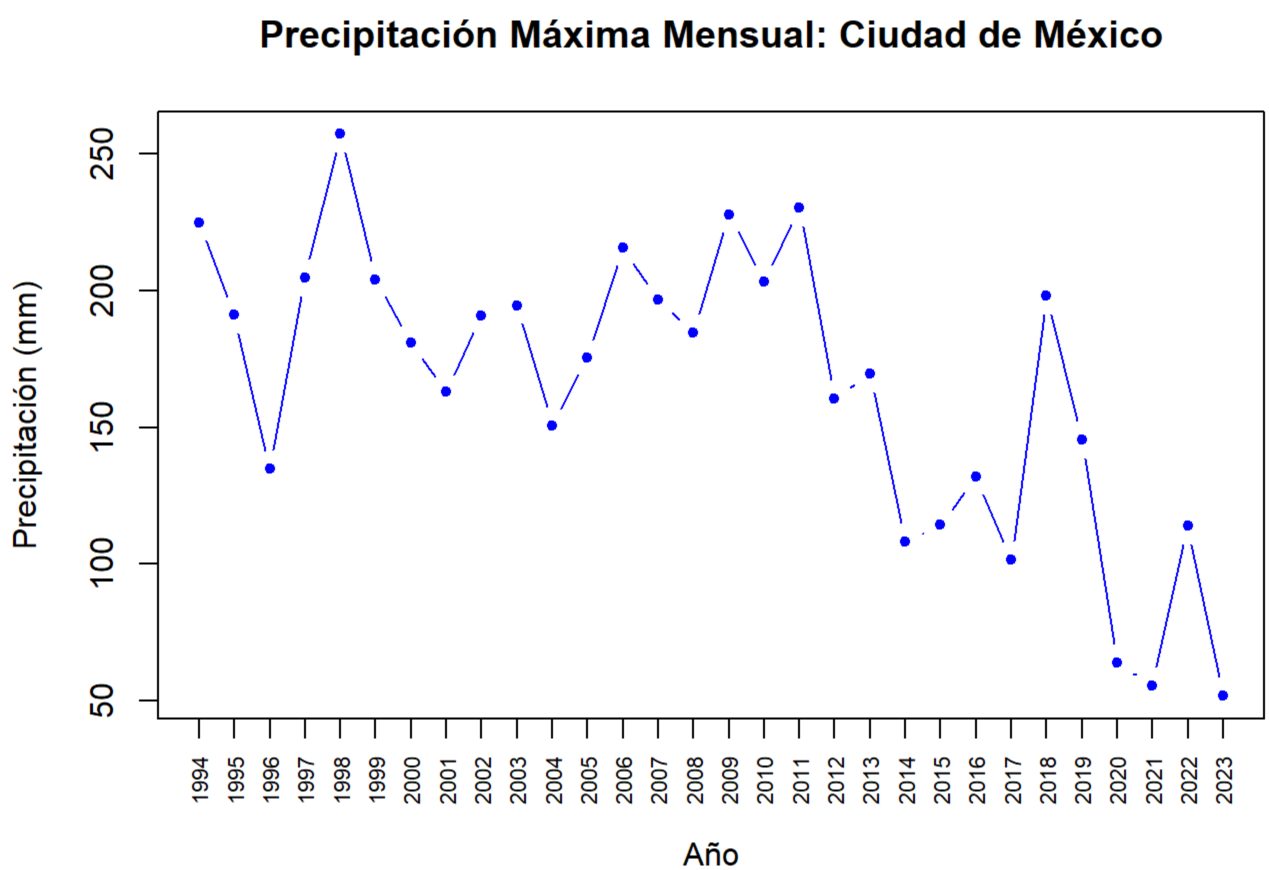
1. Análisis estadístico descriptivo de las precipitaciones históricas máximas mensuales de un estado

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/Links (https://smn.conagua.gob.mx/es/Links) to an external site.). Selecciona un estado que sea diferente a los del resto de tu equipo.

```
data <- read.table("precipitaciones_maximas_mensuales.txt", header = TRUE, sep = "\t", stringsAs
Factors = FALSE)

cdmx <- subset(data, Estado == "Ciudad.de.México")
```

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.

```
years <- unique(cdmx$Anio)

monthly_max <- c()

for (n in 1:length(years)) {monthly_max <- c(monthly_max, max(cdmx$Lluvia[which(cdmx$Anio == yea
rs[n])]))}

names(monthly_max) <- years

plot(monthly_max, type = "b", pch=20, ylab="Precipitación (mm)", main = "Precipitación Máxima Me
nsual: Ciudad de México", xaxt = "n", col = "blue", xlab = "Año")
axis(1, at = 1:length(years), labels=years, cex.axis = 0.7, las = 2)
```



c. Analiza los datos de precipitaciones máximas mensuales del estado seleccionado.

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

- Realiza gráficos que te sirvan para describir la distribución de las lluviar máximas mensuales: histograma y boxplot
- Describe el comportamiento de la distribución: centralización, sesgo, variación, ….

```
# Medidas

cat("Media: ", mean(cdmx$Lluvia, na.rm = TRUE), "\n")
```

```
## Media:  55.20222
```

```
cat("Mediana: ", median(cdmx$Lluvia, na.rm = TRUE), "\n")
```

```
## Mediana:  25.9
```

```
cat("Desviación estándar: ", sd(cdmx$Lluvia, na.rm = TRUE), "\n")
```

```
## Desviación estándar:  61.22487
```

```
cat("Varianza: ", var(cdmx$Lluvia, na.rm = TRUE), "\n")
```

```
## Varianza:  3748.485
```

```
cat("Mínimo: ", min(cdmx$Lluvia, na.rm = TRUE), "\n")
```

```
## Mínimo:  0
```

```
cat("Máximo: ", max(cdmx$Lluvia, na.rm = TRUE), "\n")
```

```
## Máximo:  257.3
```

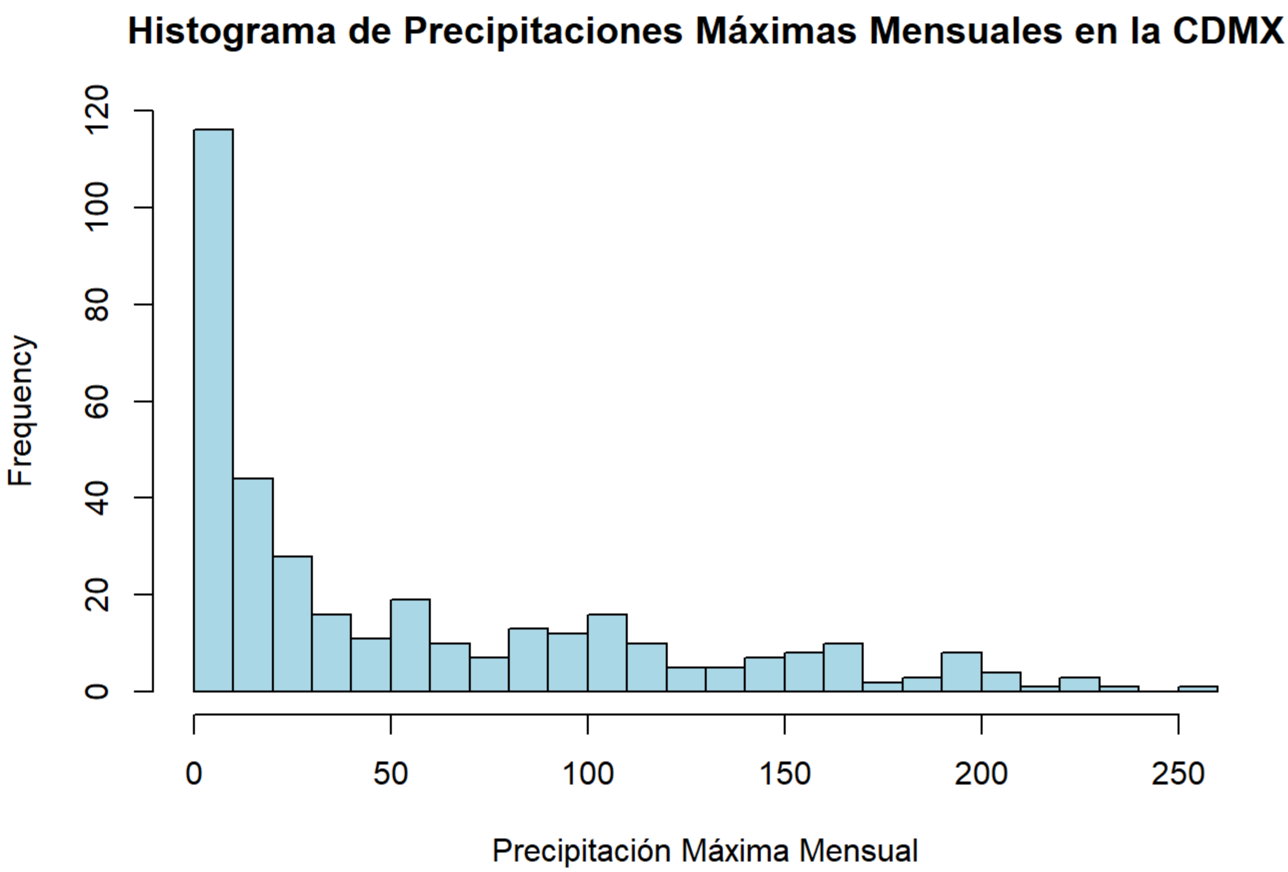
```
cat("Primer cuartil (Q1): ", quantile(cdmx$Lluvia, 0.25, na.rm = TRUE), "\n")
```

```
## Primer cuartil (Q1):  6.175
```

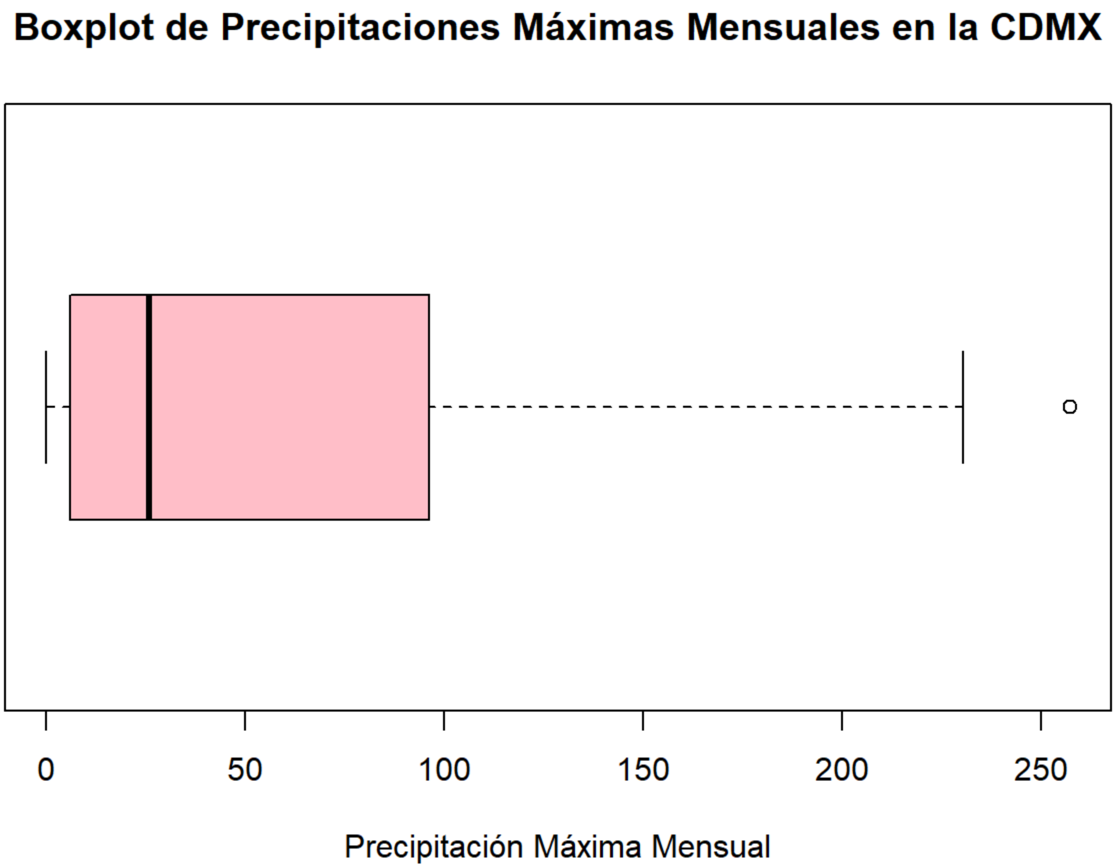
```
cat("Tercer cuartil (Q3): ", quantile(cdmx$Lluvia, 0.75, na.rm = TRUE), "\n")
```

```
## Tercer cuartil (Q3):  96.1
```

```
# Histograma y boxplot

hist(cdmx$Lluvia, breaks = 20, main = "Histograma de Precipitaciones Máximas Mensuales en la CDM
X",
     xlab = "Precipitación Máxima Mensual", col = "lightblue", border = "black")
```




```
boxplot(cdmx$Lluvia, horizontal = TRUE, main = "Boxplot de Precipitaciones Máximas Mensuales en
la CDMX",
        xlab = "Precipitación Máxima Mensual", col = "pink")
```




```
# Comportamiento de la distribución

shapiro_test <- shapiro.test(cdmx$Lluvia)
cat("Prueba de Shapiro-Wilk para normalidad:\n")
```

```
## Prueba de Shapiro-Wilk para normalidad:
```

```
print(shapiro_test)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  cdmx$Lluvia
## W = 0.8314, p-value < 2.2e-16
```

```
asimetria <- skewness(cdmx$Lluvia, na.rm = TRUE)
curtosis <- kurtosis(cdmx$Lluvia, na.rm = TRUE)

cat("Asimetría: ", asimetria, "\n")
```

```
## Asimetría:  1.111968
```

```
cat("Curtosis: ", curtosis, "\n")
```

```
## Curtosis:  3.201182
```

Se observa una distribución asimétrica a la derecha, donde la mayoría de los datos están están en los valores bajos y los demás en los valores más altos.

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?

Con base en las gráficas anteriores, se puede observar que las precipitaciones máximas mensuales muestra una distribución asimétrica hacia la derecha, donde se puede ver una concentración de datos en valores bajos y una cola extendida hacia valores más altos. Igualmente, la variabilidad en los datos es alta. Se tiene una desviación estándar de 61.22 y una varianza de 3748.49, lo que nos dice que las lluvias en la CDMX son muy dispersas. Finalmente, en la prueba de normalidad se puede ver que el valor de p es muy bajo, lo que nos dice que los datos no siguen una distribución normal. Esto tiene sentido ya que la curtosis es de 3.20.

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.

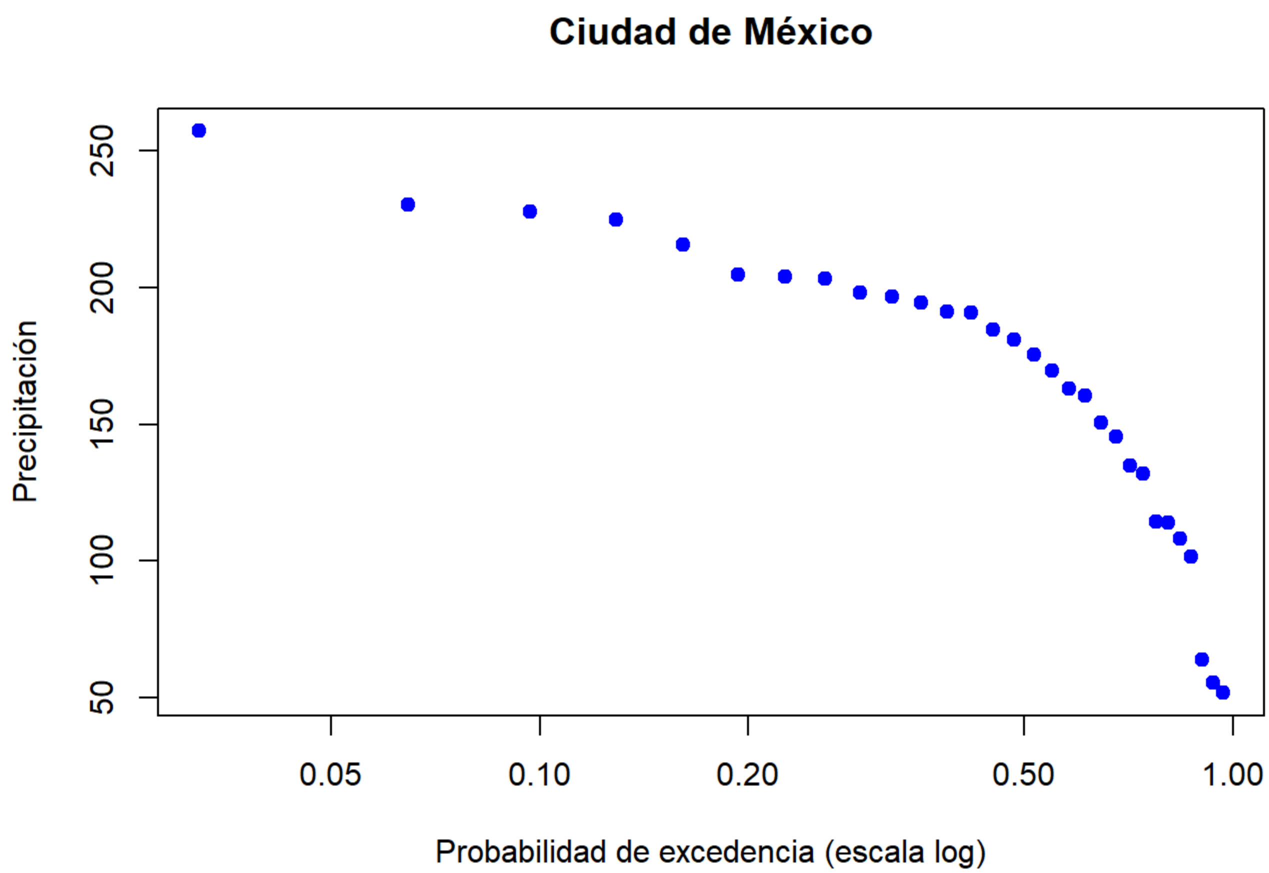
b. Se agrega una columna con el número de orden que tiene asignado cada precipitación máxima. A ese número se le llama “rank” (rango en español) y se simboliza por m

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

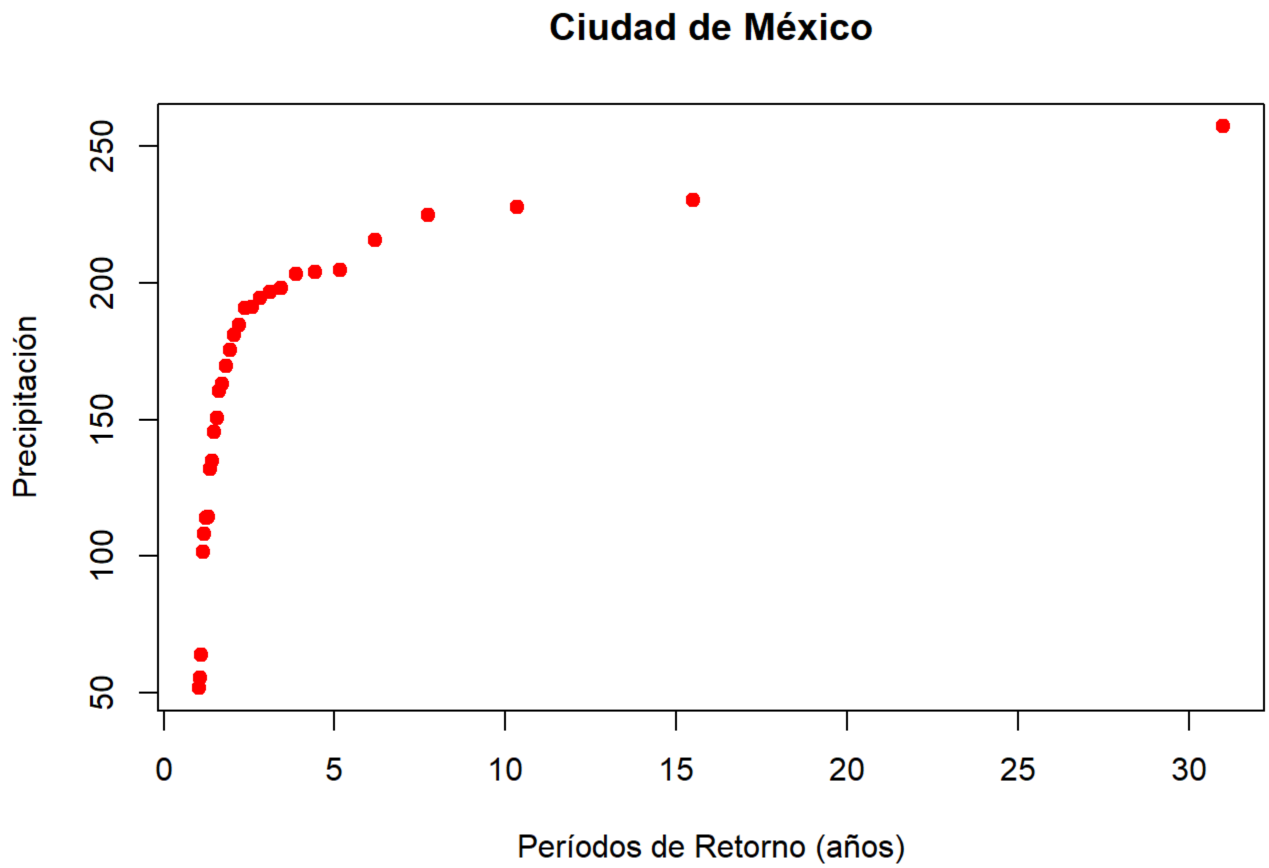
d. Se calcula la probabilidad de no excedencia para cada precipitación (complemento de la probabilidad de excedencia): $P_{no\ exce} = 1 - P_{exe}$

e. Se calcula el periodo de retorno como el inverso de la probabilidad de excedencia: $P_{ret} = \frac{1}{P_{exe}}$

```
rain_analysis <- data.frame(max_rain = monthly_max, order_max_rain = sort(monthly_max, decreasin
g = TRUE))

rain_analysis$rank_rain <- seq(1, nrow(rain_analysis))

rain_analysis$Pexe <- rain_analysis$rank_rain / (nrow(rain_analysis) + 1)
rain_analysis$Pnoexe <- 1 - rain_analysis$Pexe
rain_analysis$Pret <- 1 / rain_analysis$Pexe

plot(y = rain_analysis$order_max_rain, x = rain_analysis$Pexe, log = "x", pch = 19,
  main = "Ciudad de México", xlab = "Probabilidad de excedencia (escala log)", ylab = "Precipita
ción", col = "blue")
```




```
plot(x = rain_analysis$Pret, y = rain_analysis$order_max_rain, pch = 19,
  main = "Ciudad de México", xlab = "Períodos de Retorno (años)", ylab = "Precipitación", col =
"red")
```



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?

En la primera gráfica se puede observar que tan fuertes eran las lluvias contra la probabilidad de excedencia, que es la probabilidad de que una precipitación máxima iguale o supere un valor específico en un año dado. Cuando hay valores altos, hay una baja probabilidad de excedencia, lo que nos dice que eso pasa raramente. Por otro lado, a medida que la probabilidad de excedencia sube, la intensidad de la lluvia disminuye, lo que refleja eventos menos extremos y más frecuentes.

En la segunda gráfica podemos ver la magnitud de las lluvias a medida de que aumenta el periódo de retorno, el cual es el intervalo de tiempo (años) esperado en el que un evento de cierta magnitud ocurre o se excede una vez. Esto quiere decir que a medida que aumenta el tiempo, la intensidad de la precipitación también crece, mostrando eventos menos frecuentes pero más fuertes.

Hablando de hidrología, estos conceptos son importantes para el diseño y las pruebas de obras hidráulicas. Esto es ya que ayudan a prevenir el impacto de eventos extremos, para así poder garantizar la seguridad y resistencia de las obras ante lluvias fuertes que no suelen ocurrir seguido. Al momento de diseñar este tipo de obras, se busca un bajo valor de probabilidad de excedencia, lo cual asegura que la estructura puede soportar eventos raros pero de gran magnitud.

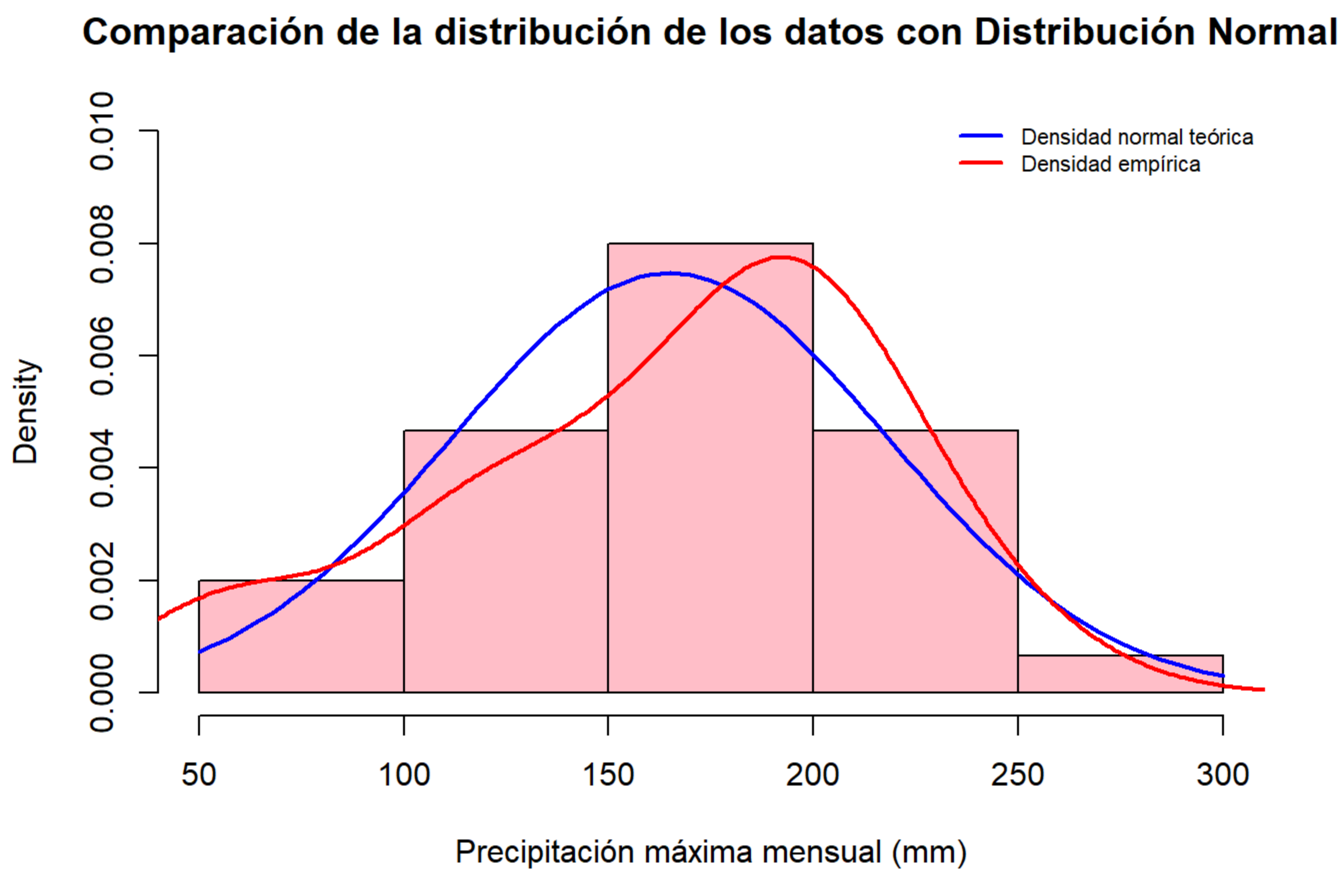
3. Análisis de frecuencias. Método analítico.

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

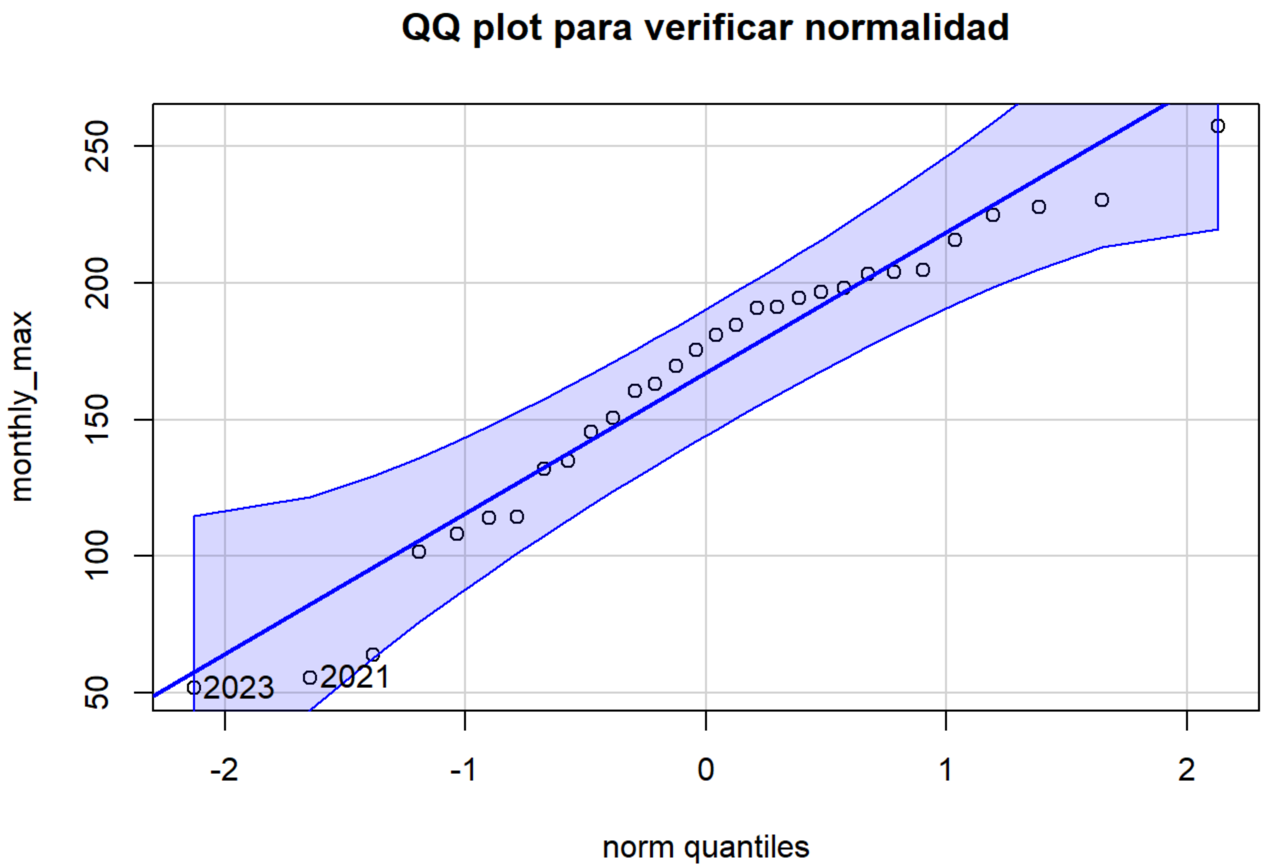
a. **Ajuste a una Distribución Normal.** Hay dos maneras de determinar si un conjunto de datos tiene una distribución normal, una visual y la otra mediante una prueba de bondad de ajuste.

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

```
# Histograma de densidad empírica y distribución normal teórica

hist(monthly_max, xlab="Precipitación máxima mensual (mm)", freq = FALSE, ylim = c(0, 0.01),
     main = "Comparación de la distribución de los datos con Distribución Normal", col = "pink")

curve(dnorm(x, mean = mean(monthly_max), sd=sd(monthly_max)), add=TRUE, col="blue", lwd=2)
lines(density(monthly_max), col="red", lwd=2)

legend("topright", col = c("blue", "red"), legend = c("Densidad normal teórica", "Densidad empír
ica"), lwd = 2, bty = "n", cex = 0.7)
```

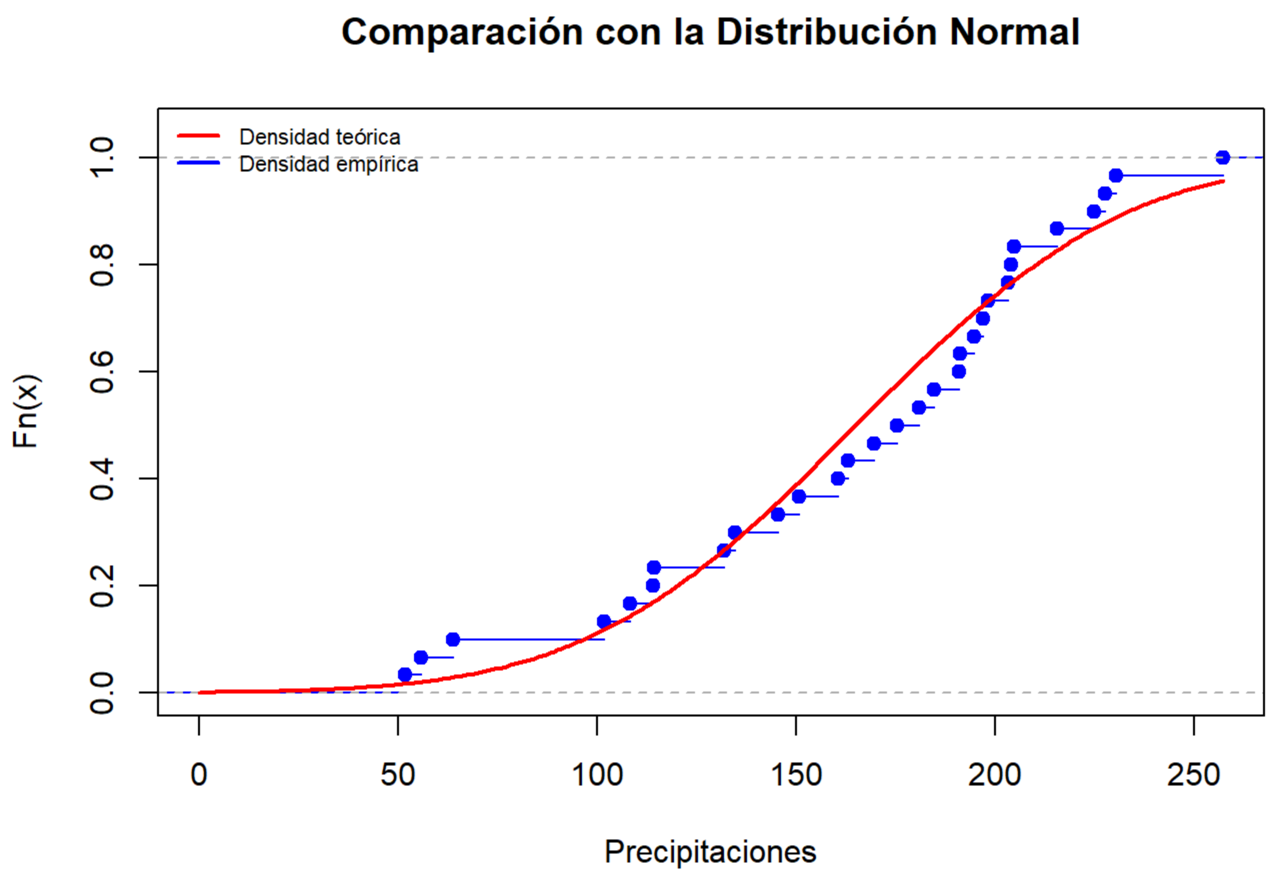



```
# Gráfica QQ plot

qqPlot(monthly_max, main="QQ plot para verificar normalidad")
```




```
## 2023 2021
##   30   28
```

```
# Comparación de probabilidades acumuladas empíricas y teóricas

norm_teorica <- pnorm(0:max(monthly_max), mean=mean(monthly_max), sd=sd(monthly_max))

plot(ecdf(monthly_max), main="Comparación con la Distribución Normal", xlab="Precipitaciones", c
ol="blue", xlim=c(0, max(monthly_max)), ylim=c(0, 1.05))

par(new=TRUE)

plot(0:max(monthly_max), norm_teorica, type="l", main="", xlab="", ylab="", col="red", lwd=2, yl
im=c(0, 1.05), xaxt="n", yaxt="n")

legend("topleft", col=c("red", "blue"), legend = c("Densidad teórica", "Densidad empírica"), lwd
=2, bty="n", cex=0.7)
```




```
# Pruebas de normalidad: Shapiro-Wilks y Kolmogorov-Smirnov

# Prueba de Shapiro-Wilks

shapiro_result <- shapiro.test(monthly_max)
print(shapiro_result)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  monthly_max
## W = 0.94861, p-value = 0.1552
```

```
# Prueba de Kolmogorov-Smirnov

ks_result <- ks.test(monthly_max, "pnorm", mean=mean(monthly_max), sd = sd(monthly_max))

print(ks_result)
```

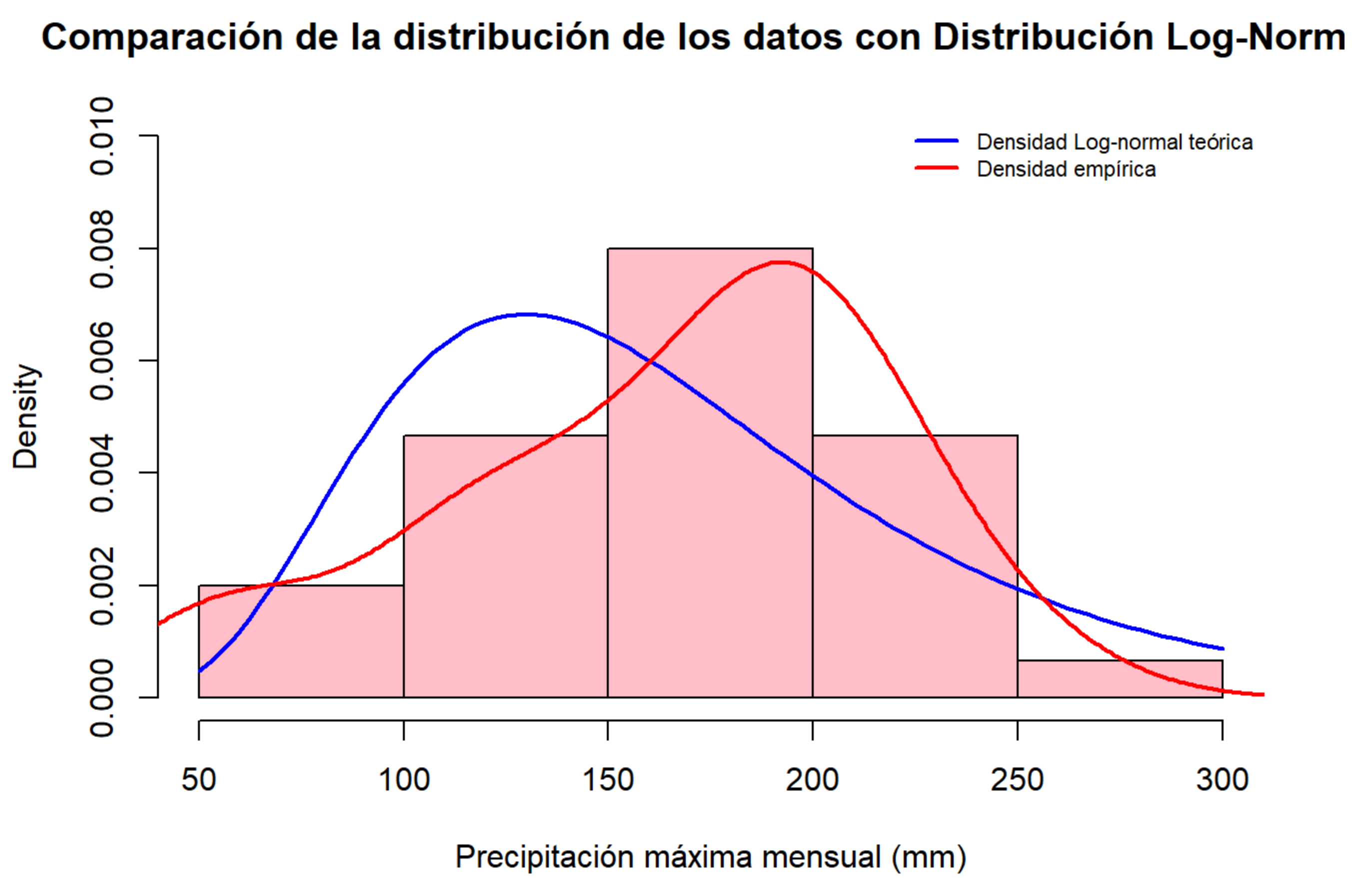
```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  monthly_max
## D = 0.12085, p-value = 0.7285
## alternative hypothesis: two-sided
```

Con base en la siguiente gráfica, se puede observar que los datos no se ajustan perfectamente a una distribución normal. La curva de densidad empírica (línea roja) nos muestra asimetría hacia la derecha, lo que nos dice que esta distribución no parece ser la mejor opción para modelar los datos; sin embargo, la opción azul sí se ajusta. Igualmente, la distribución normal tiene dos parámetros: la media (μ) y la desviación estándar (σ). Estos parámetros determinan la forma de la curva. La media centraliza la distribución en el eje horizontal, y la desviación estándar define qué tan ancha es la curva. En cuanto al código, los parámetros se calcularon usando la media y desviación estándar de los datos de la CDMX. Esto se debe a que la mejor estimación de la media es la media aritmética, y la mejor estimación de la d.s. es la desviación estándar muestral. En cuanto a la gráfica del qqplot, se puede observar cómo se distribuyen los cuantiles de los datos utilizados a los cuantiles de una distribución normal. En este caso, los puntos se desvían de la línea de referencia en las colas, especialmente en los extremos superior e inferior. Esto nos dice que los datos no se ajustan perfectamente una distribución normal, pero no es la peor opción.

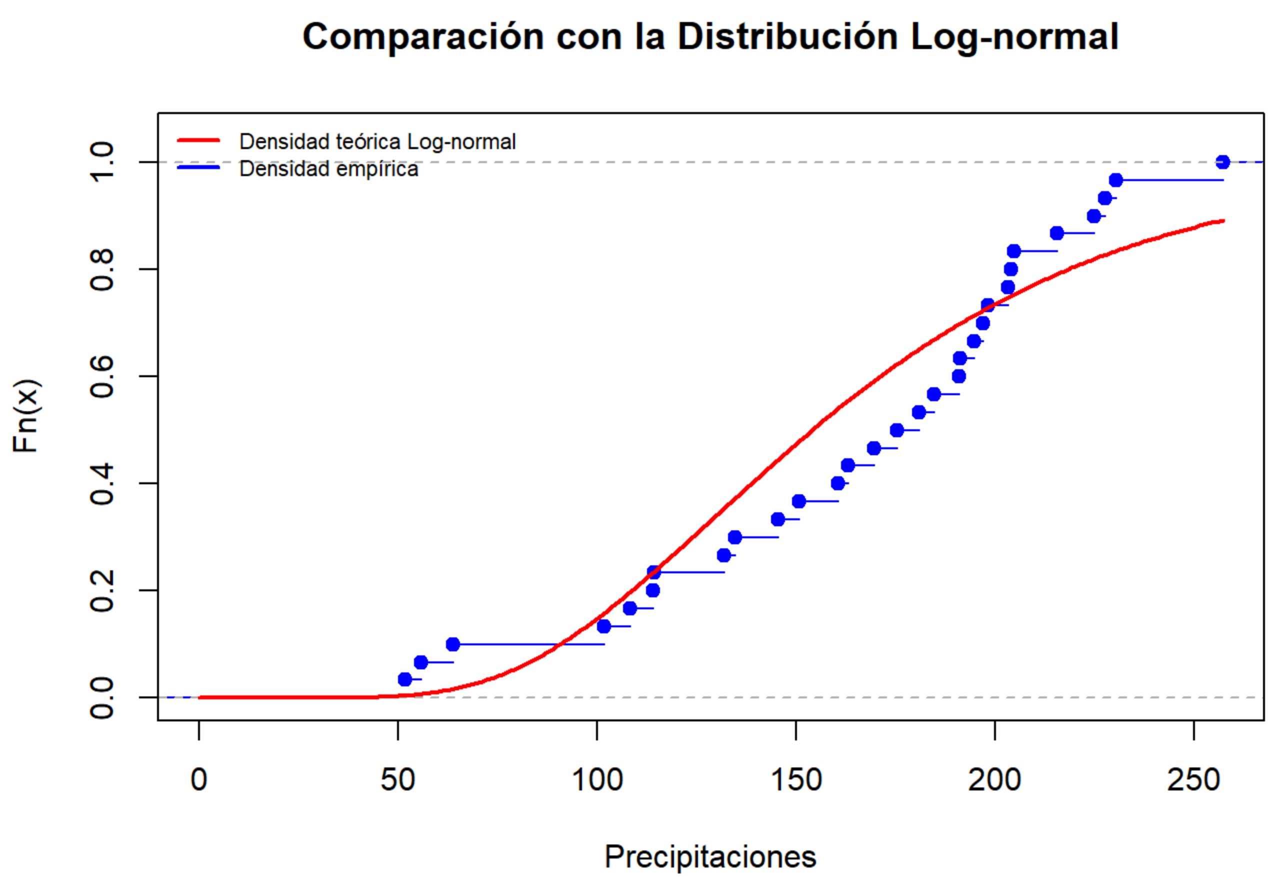
b. **Ajuste a una Distribución Log-Normal.** Realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Log-normal para ajustar los datos.

- Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución Log-normal que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. De manera visual, ¿te parece que los datos se ajustan bien a una distribución Log-normal? Explica.
- 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?
- Haz la prueba KS para determinar si los datos se ajustan a una 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? ¿Podemos concluir que los datos de las precipitaciones máximas mensuales siguen una distribución log-normal? ¿Por qué?
- ¿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.

```
# Histograma de densidad empírica superpuesto a la densidad Log-normal teórica

hist(monthly_max, xlab = "Precipitación máxima mensual (mm)", freq = FALSE, ylim = c(0, 0.01),
     main="Comparación de la distribución de los datos con Distribución Log-Normal", col = "pin
k")

curve(dlnorm(x, mean = mean(log(monthly_max)), sd = sd(log(monthly_max))), add = TRUE, col = "bl
ue", lwd = 2)
lines(density(monthly_max), col = "red", lwd = 2)

legend("topright", col = c("blue", "red"), legend = c("Densidad Log-normal teórica", "Densidad e
mpírica"), lwd = 2, bty = "n", cex = 0.7)
```



```
# Comparación de probabilidades acumuladas empíricas y teóricas

lognorm_teorica <- plnorm(0:max(monthly_max), mean=mean(log(monthly_max)), sd=sd(log(monthly_ma
x)))

plot(ecdf(monthly_max), main="Comparación con la Distribución Log-normal", xlab="Precipitacione
s", col="blue", xlim=c(0, max(monthly_max)), ylim=c(0, 1.05))

par(new=TRUE)

plot(0:max(monthly_max), lognorm_teorica, type="l", main="", xlab="", ylab="", col="red", lwd=2,
ylim=c(0, 1.05), xaxt="n", yaxt="n")

legend("topleft", col=c("red", "blue"), legend=c("Densidad teórica Log-normal", "Densidad empíri
ca"), lwd=2, bty="n", cex=0.7)
```



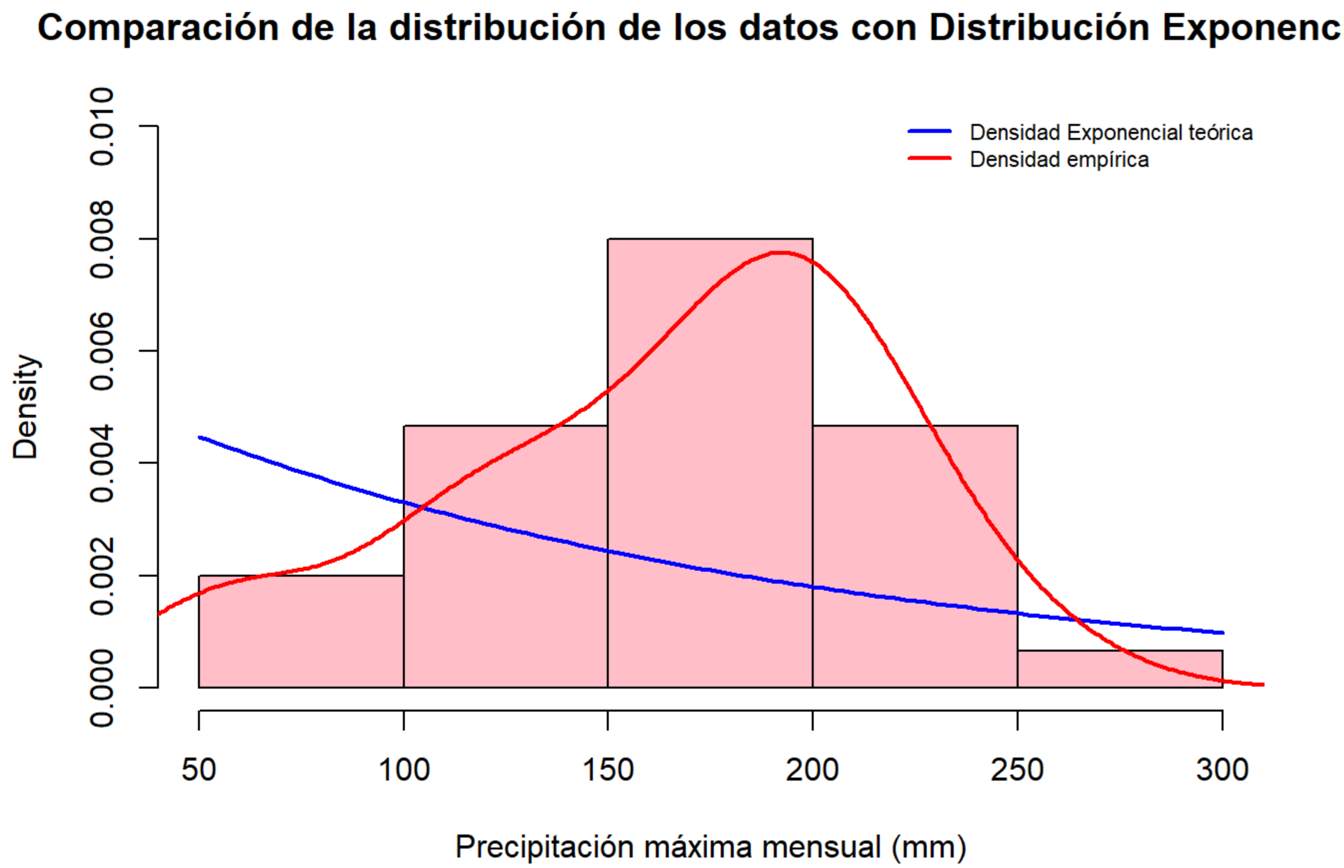

```
# Prueba de Kolmogorov-Smirnov (KS) para la Log-normal

ks_lognorm_result <- ks.test(monthly_max, "plnorm", mean=mean(log(monthly_max)), sd=sd(log(month
ly_max)))
print(ks_lognorm_result)
```

```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  monthly_max
## D = 0.1738, p-value = 0.2902
## alternative hypothesis: two-sided
```

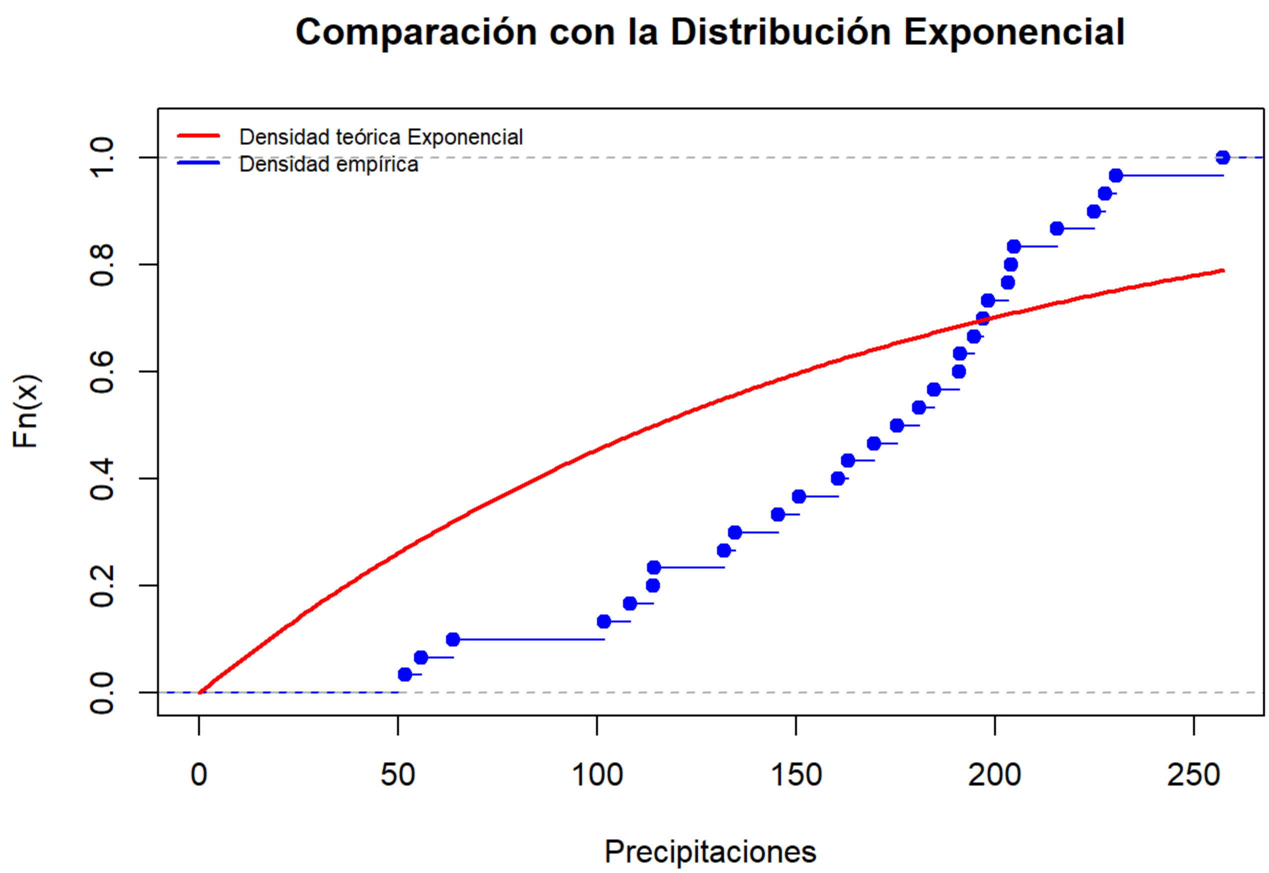
c. **Ajuste a una Distribución Exponencial.** Realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Exponencial para ajustar los datos.

- Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución Exponencial que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. De manera visual, ¿te parece que los datos se ajustan bien a una distribución Exponencial? Explica.
- 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?
- Haz la prueba KS para determinar si los datos se ajustan a una 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? ¿Podemos concluir que los datos de las precipitaciones máximas mensuales siguen una distribución exponencial? ¿Por qué?
- ¿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.

```
# Histograma de densidad empírica superpuesto a la densidad Exponencial teórica

hist(monthly_max, xlab = "Precipitación máxima mensual (mm)", freq = FALSE, ylim = c(0, 0.01), m
ain="Comparación de la distribución de los datos con Distribución Exponencial", col = "pink")

curve(dexp(x, rate = 1/mean(monthly_max)), add = TRUE, col = "blue", lwd = 2)

lines(density(monthly_max), col = "red", lwd = 2)

legend("topright", col = c("blue", "red"), legend = c("Densidad Exponencial teórica", "Densidad
empírica"), lwd = 2, bty = "n", cex = 0.7)
```




```
# Comparación de probabilidades acumuladas empíricas y teóricas

exp_teorica <- pexp(0:max(monthly_max), rate = 1/mean(monthly_max))

plot(ecdf(monthly_max), main="Comparación con la Distribución Exponencial", xlab="Precipitacione
s", col="blue", xlim=c(0, max(monthly_max)), ylim = c(0, 1.05))

par(new = TRUE)

plot(0:max(monthly_max), exp_teorica, type = "l", main = "", xlab = "", ylab = "", col = "red",
lwd = 2, ylim = c(0, 1.05), xaxt = "n", yaxt = "n")

legend("topleft", col=c("red", "blue"), legend=c("Densidad teórica Exponencial", "Densidad empír
ica"), lwd=2, bty="n", cex=0.7)
```




```
# Prueba de Kolmogorov-Smirnov (KS) para la Exponencial

ks_exp_result <- ks.test(monthly_max, "pexp", rate=1/mean(monthly_max))
print(ks_exp_result)
```

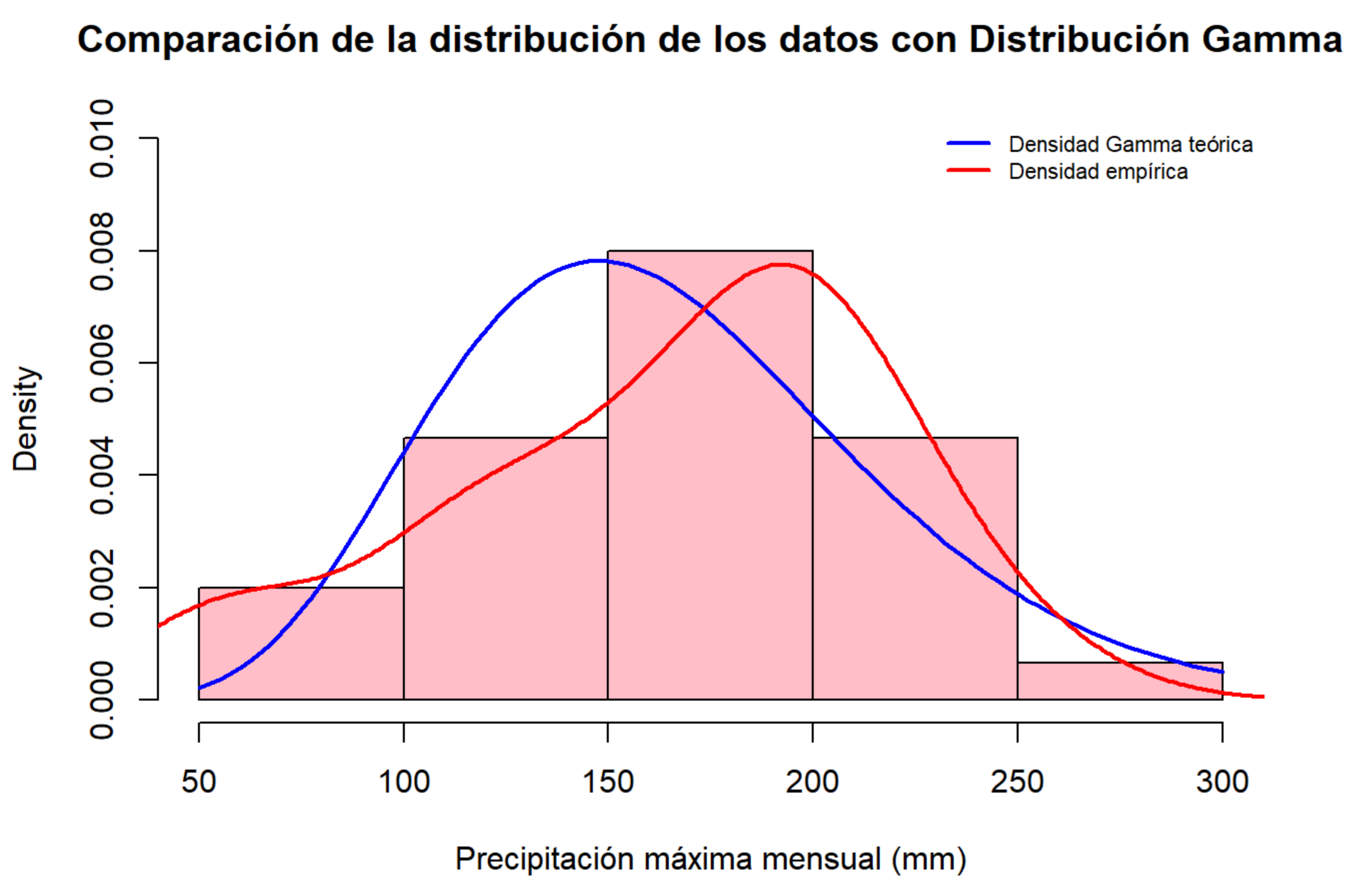
```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  monthly_max
## D = 0.36029, p-value = 0.0005399
## alternative hypothesis: two-sided
```

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

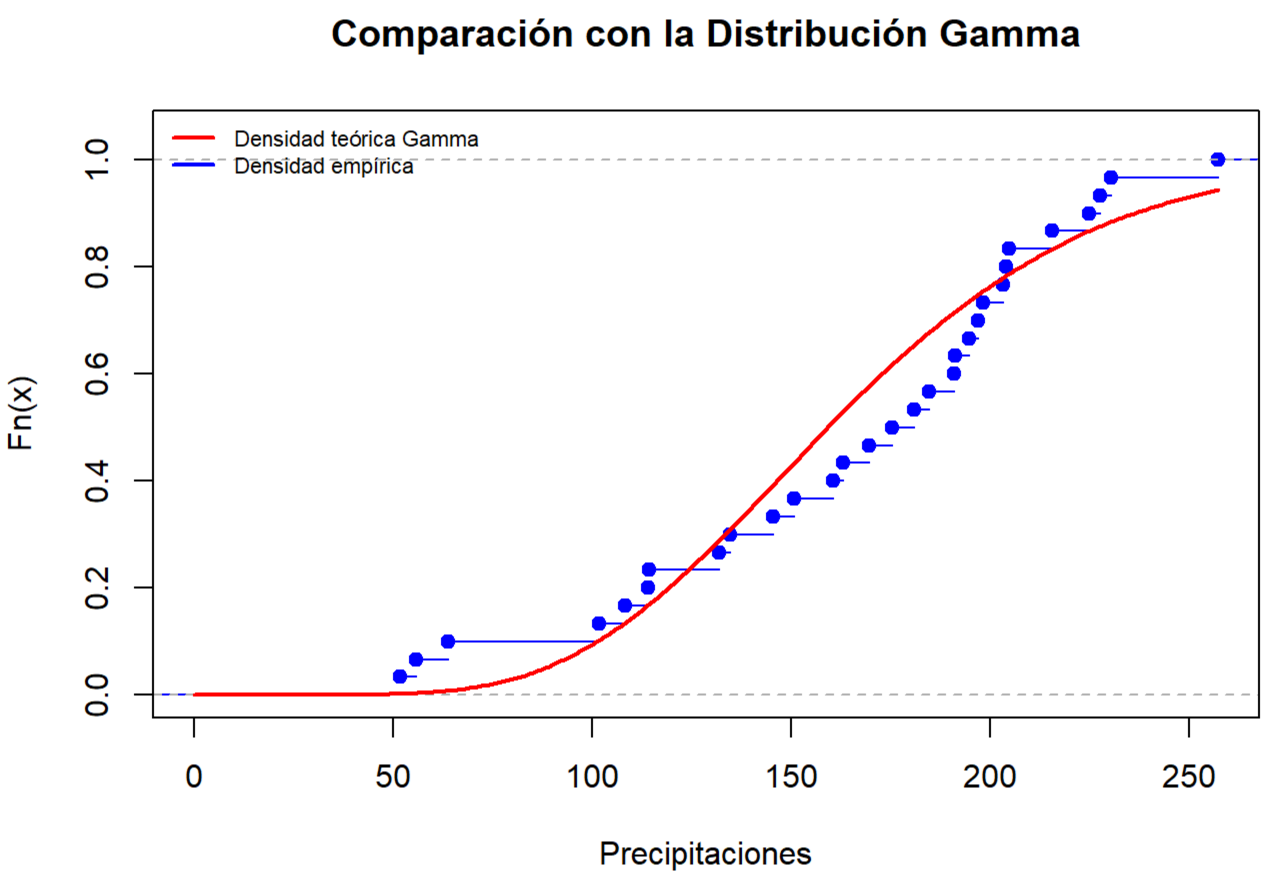
- Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución Gamma que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. De manera visual, ¿te parece que los datos se ajustan bien a una distribución Gamma? Explica.
- Compara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

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

```
# Histograma de densidad empírica superpuesto a la densidad Gamma teórica

shape_param <- mean(monthly_max)^2 / var(monthly_max)
rate_param <- mean(monthly_max) / var(monthly_max)

hist(monthly_max, xlab="Precipitación máxima mensual (mm)", freq=FALSE, ylim=c(0, 0.01), main="C
omparación de la distribución de los datos con Distribución Gamma", col = "pink")

curve(dgamma(x, shape=shape_param, rate=rate_param), add=TRUE, col="blue", lwd=2)

lines(density(monthly_max), col="red", lwd=2)

legend("topright", col=c("blue", "red"), legend=c("Densidad Gamma teórica", "Densidad empíric
a"), lwd=2, bty="n", cex=0.7)
```



```
# Comparación de probabilidades acumuladas empíricas y teóricas

gamma_teorica <- pgamma(0:max(monthly_max), shape=shape_param, rate=rate_param)

plot(ecdf(monthly_max), main="Comparación con la Distribución Gamma", xlab="Precipitaciones", co
l="blue", xlim=c(0, max(monthly_max)), ylim=c(0, 1.05))

par(new=TRUE)

plot(0:max(monthly_max), gamma_teorica, type="l", main="", xlab="", ylab="", col="red", lwd=2, y
lim=c(0, 1.05), xaxt="n", yaxt="n")

legend("topleft", col=c("red", "blue"), legend=c("Densidad teórica Gamma", "Densidad empírica"),
lwd=2, bty="n", cex=0.7)
```



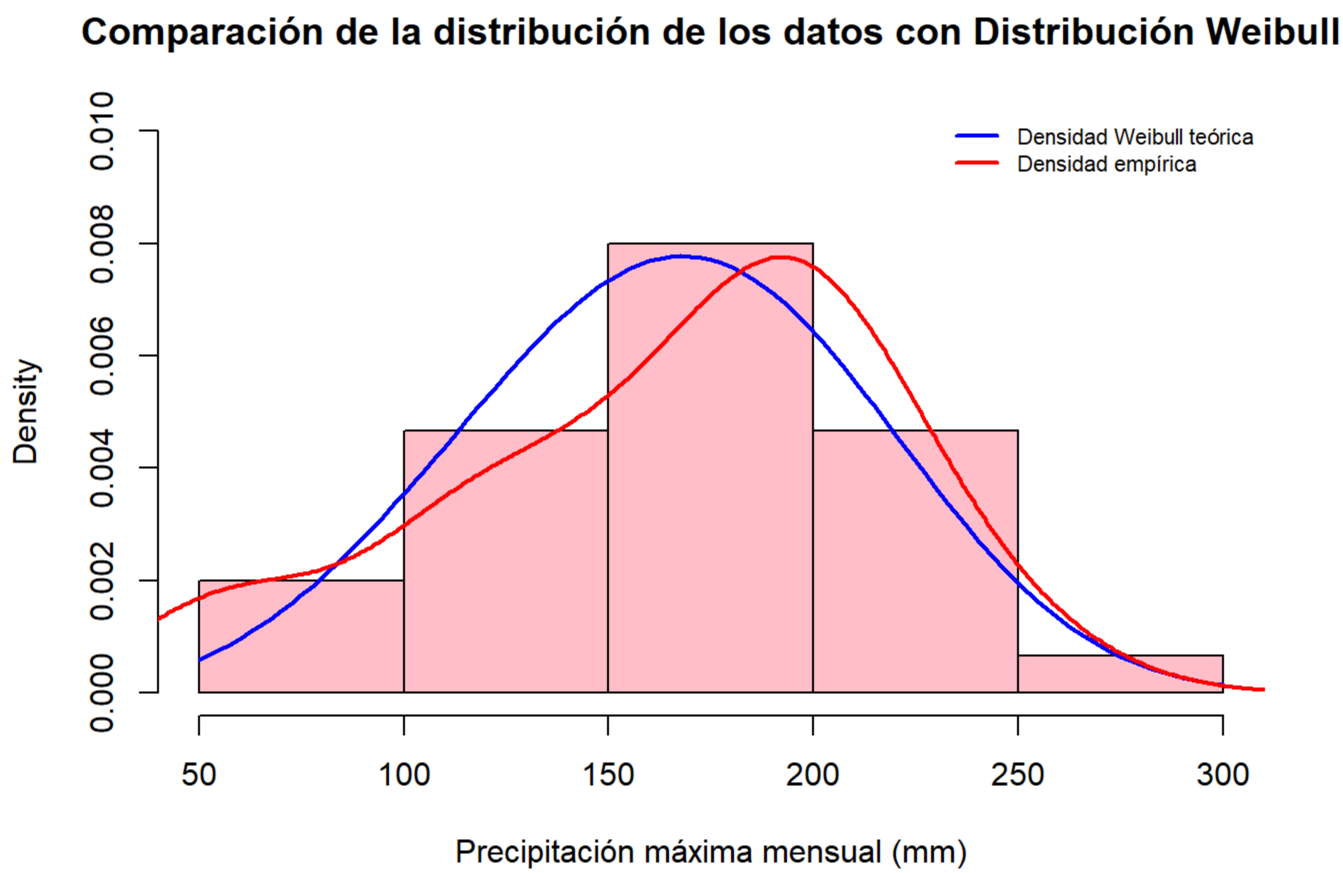

```
# Prueba de Kolmogorov-Smirnov (KS) para la Gamma

ks_gamma_result <- ks.test(monthly_max, "pgamma", shape=shape_param, rate=rate_param)
print(ks_gamma_result)
```

```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  monthly_max
## D = 0.15526, p-value = 0.4219
## alternative hypothesis: two-sided
```

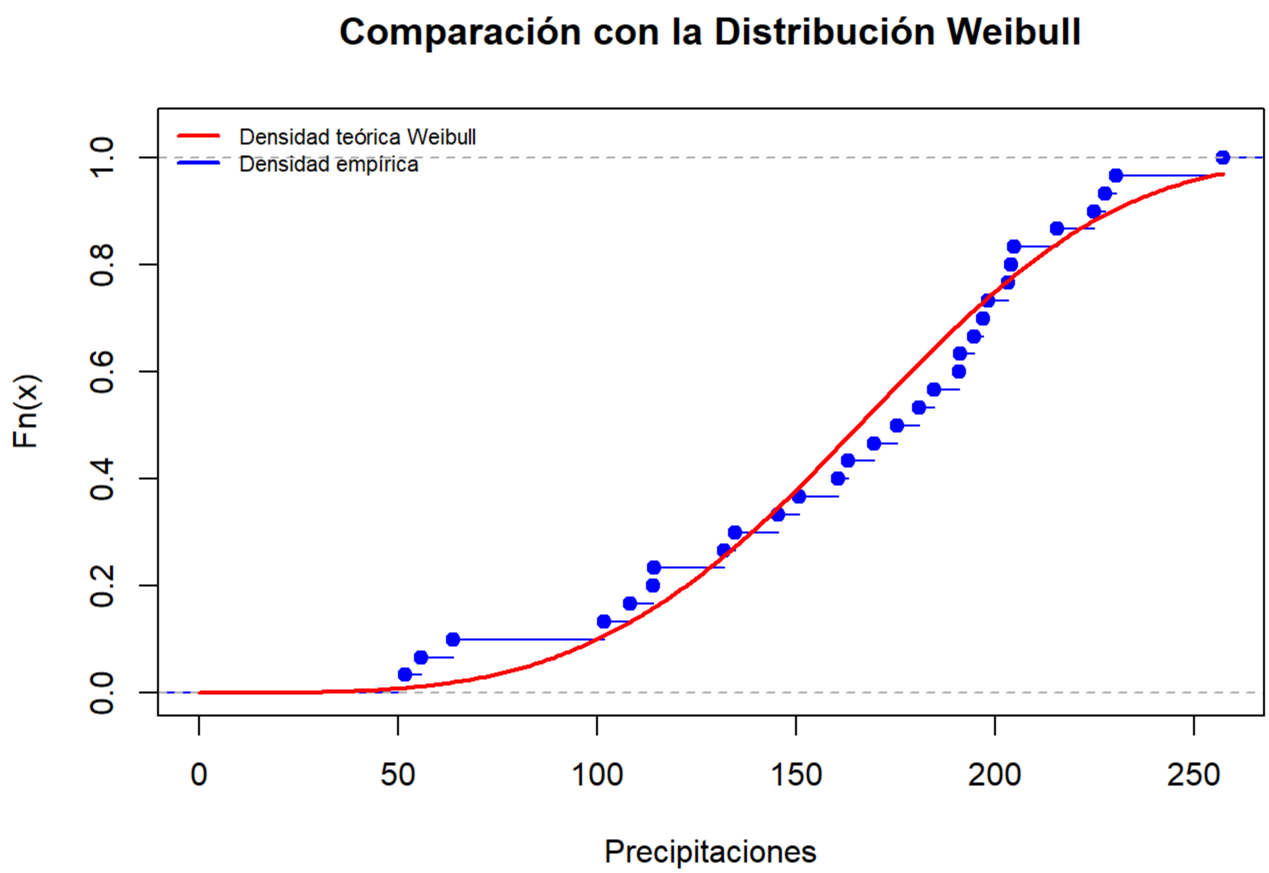
e. **Ajuste a una Distribución Weibull.** Realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Weibull para ajustar los datos.

- 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.
- 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.
- 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?
- Haz la prueba KS para determinar si los datos se ajustan a una 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? ¿Podemos concluir que los datos de las precipitaciones máximas mensuales siguen una distribución Weibull? ¿Por qué?
- ¿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.

```r
# Estimación de los parámetros de la distribución Weibull usando fitdistr

weibull_fit <- fitdistr(monthly_max, "weibull", lower=c(0, 0))
shape_param <- weibull_fit$estimate[1]
scale_param <- weibull_fit$estimate[2]

# Histograma de densidad empírica superpuesto a la densidad Weibull teórica

hist(monthly_max, xlab="Precipitación máxima mensual (mm)", freq=FALSE, ylim=c(0, 0.01),
     main="Comparación de la distribución de los datos con Distribución Weibull", col = "pink")

curve(dweibull(x, shape=shape_param, scale=scale_param), add=TRUE, col="blue", lwd=2)

lines(density(monthly_max), col="red", lwd=2)

legend("topright", col=c("blue", "red"), legend=c("Densidad Weibull teórica", "Densidad empíric
a"), lwd=2, bty="n", cex=0.7)
```




```
# Comparación de probabilidades acumuladas empíricas y teóricas

weibull_teorica <- pweibull(0:max(monthly_max), shape=shape_param, scale=scale_param)

plot(ecdf(monthly_max), main="Comparación con la Distribución Weibull", xlab="Precipitaciones",
col="blue", xlim=c(0, max(monthly_max)), ylim=c(0, 1.05))

par(new=TRUE)

plot(0:max(monthly_max), weibull_teorica, type="l", main="", xlab="", ylab="", col="red", lwd=2,
ylim=c(0, 1.05), xaxt="n", yaxt="n")

legend("topleft", col=c("red", "blue"), legend=c("Densidad teórica Weibull", "Densidad empíric
a"), lwd=2, bty="n", cex=0.7)
```




```
# Prueba de Kolmogorov-Smirnov (KS) para la Weibull

ks_weibull_result <- ks.test(monthly_max, "pweibull", shape=shape_param, scale=scale_param)
print(ks_weibull_result)
```

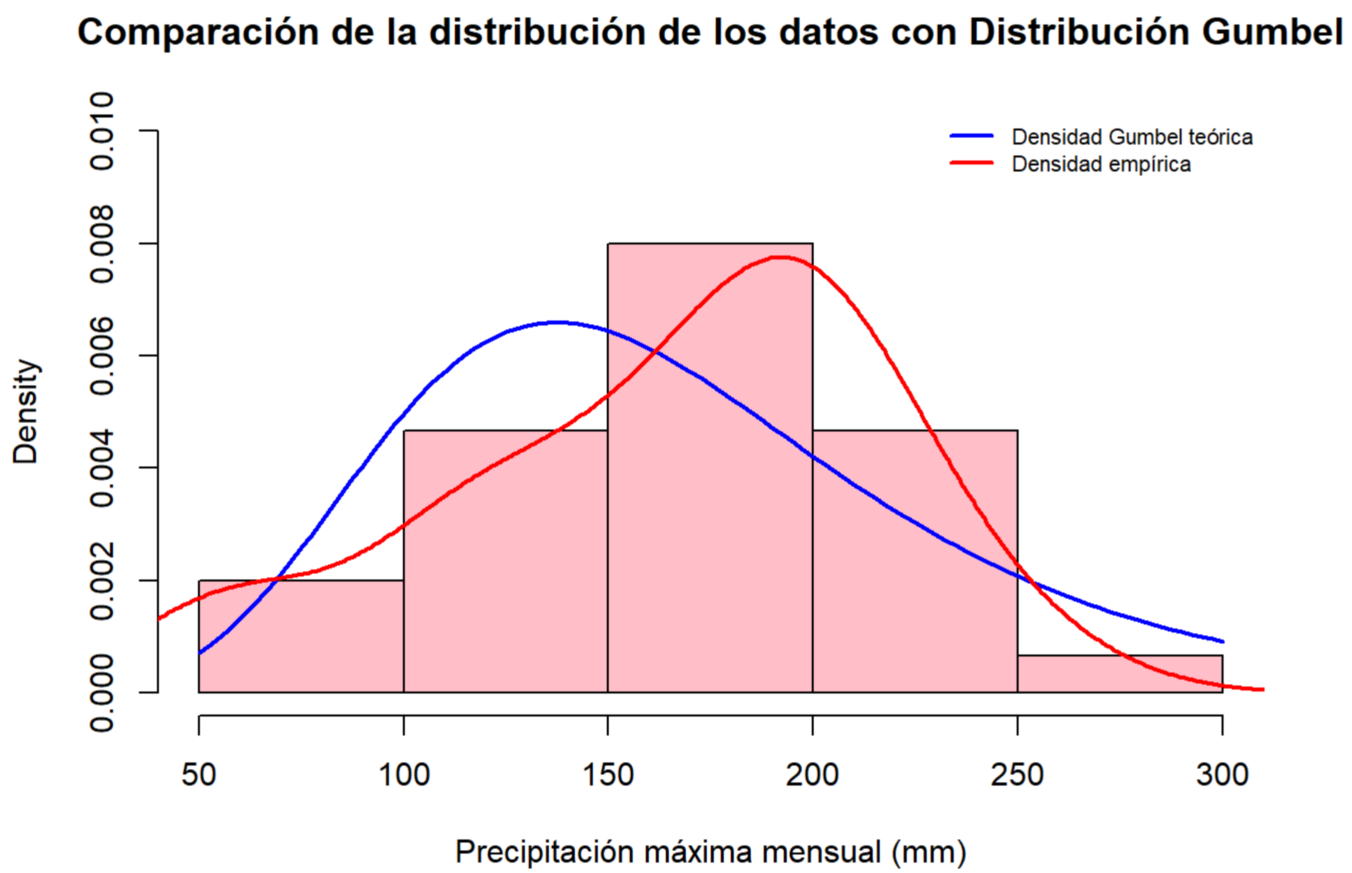
```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  monthly_max
## D = 0.12421, p-value = 0.6977
## alternative hypothesis: two-sided
```

Viendo lo anterior, los datos parecen ajustarse bien a una distribución Weibull, como se observa en las gráficas de densidad y probabilidad acumulada, donde las curvas teóricas siguen de cerca las curvas empíricas. La prueba de Kolmogorov-Smirnov también muestra buenos resultados, con un estadístico D = 0.12421 y un p-value de 0.6977, lo cual es un buen nivel de significancia. Esto indica que no hay evidencia significativa para rechazar la hipótesis nula, lo que sugiere que la distribución Weibull es adecuada para modelar estos datos. La distribución Weibull tiene dos parámetros: el de forma y el de escala, los cuales se estiman mediante métodos como el de máxima verosimilitud, lo que la hace más compleja de ajustar en comparación con distribuciones como la normal, cuyos parámetros se calculan directamente a partir de la media y la desviación estándar.

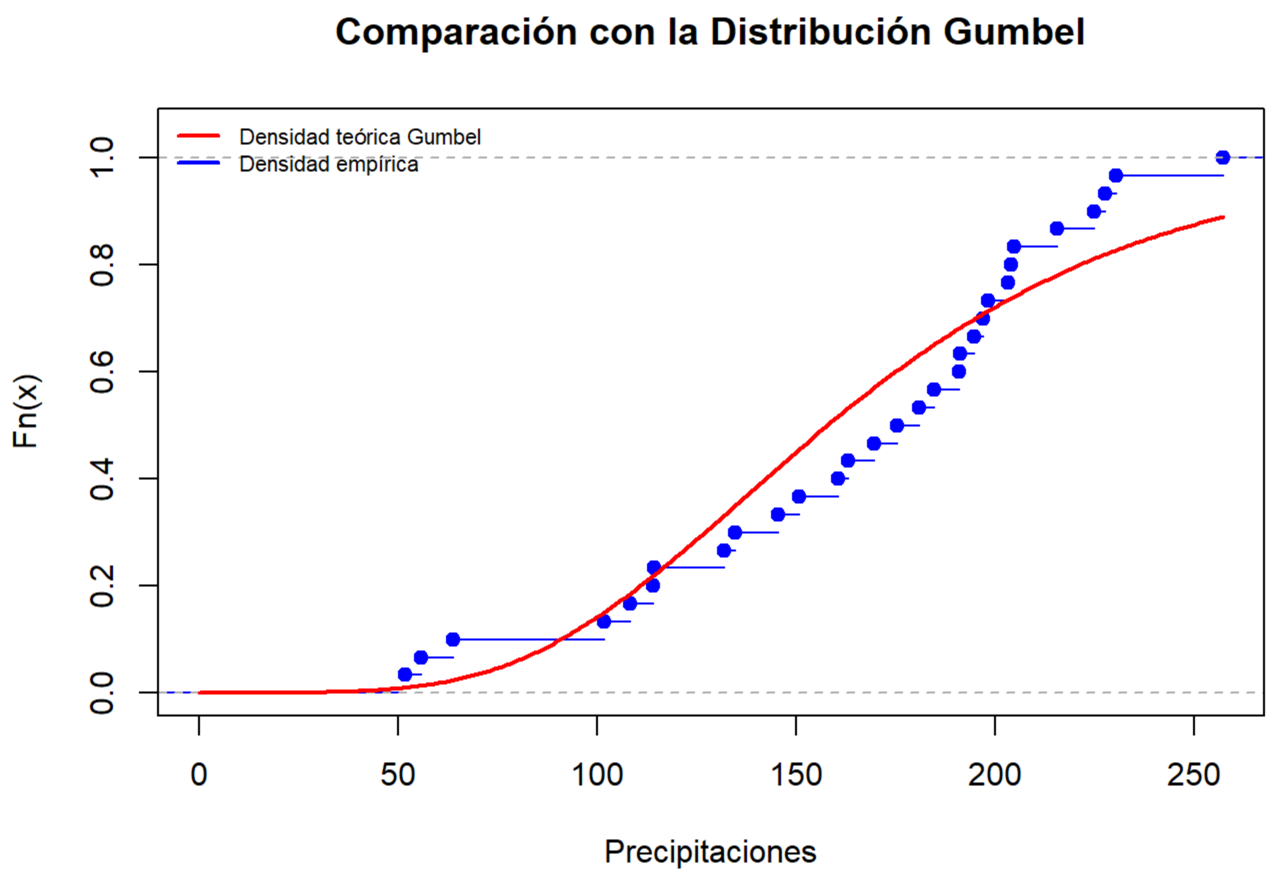
f. **Ajuste a una Distribución Gumbel.** Realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Gumbel para ajustar los datos.

- Para probar si los datos de precipitación máxima se ajustan a una distribución Gumbel, se necesita definir las funciones de densidad de acuerdo con la función Gumbel. Creálas con las fórmulas de la Distribución Gumbel.
- 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.
- Haz la prueba KS para determinar si los datos se ajustan a una Gumbel
- ¿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?
- ¿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?

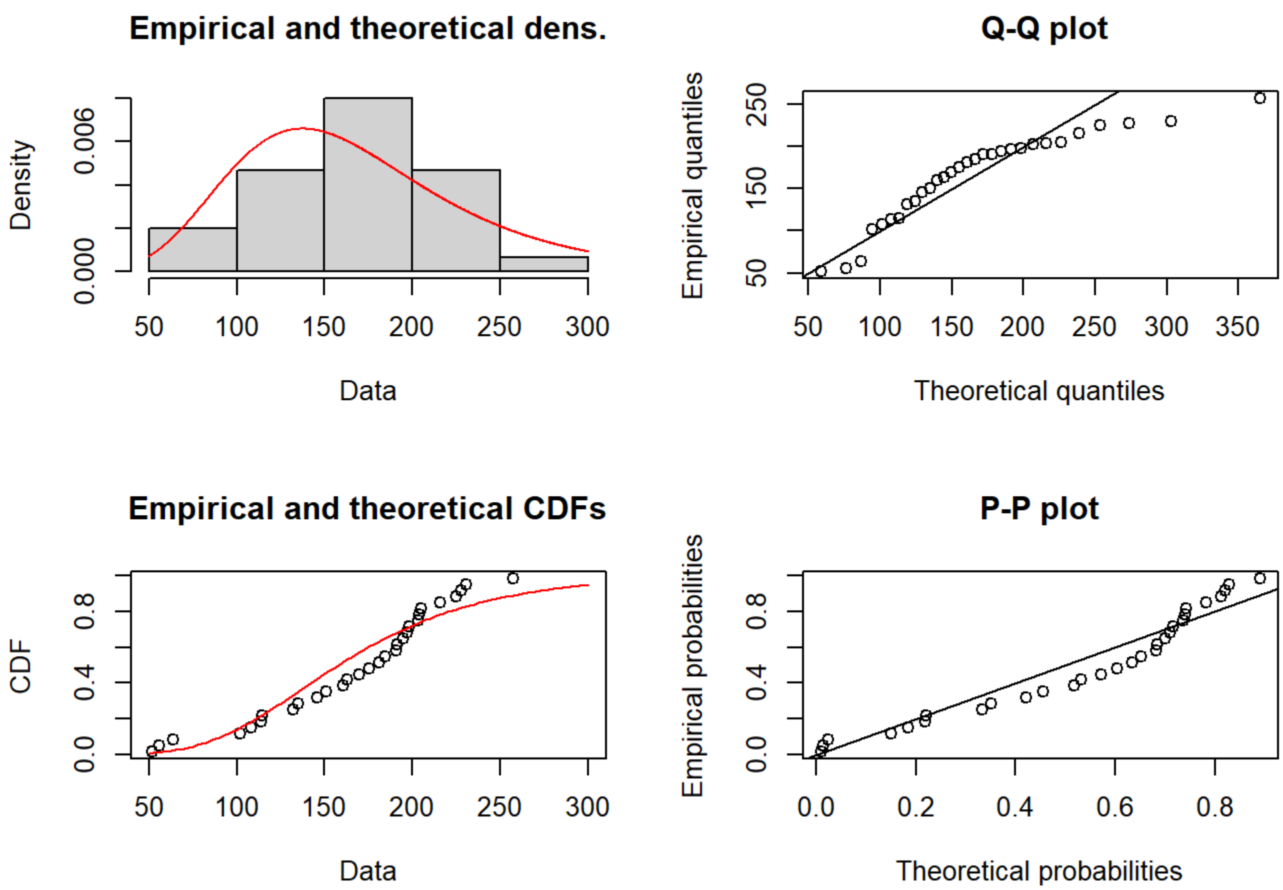
```
# Definición de las funciones de densidad y distribución acumulada de Gumbel

dgumbel <- function(x, a, b) (1/b) * exp((a - x)/b) * exp(-exp((a - x)/b))
pgumbel <- function(q, a, b) exp(-exp((a - q)/b))
qgumbel <- function(p, a, b) a - b * log(-log(p))

# Estimación de los parámetros de la distribución Gumbel

gumbel_fit <- fitdist(monthly_max, "gumbel", start = list(a = 1, b = 1))

# Histograma de densidad empírica y teórica (Gumbel)

hist(monthly_max, xlab = "Precipitación máxima mensual (mm)", freq = FALSE, ylim = c(0, 0.01),
     main = "Comparación de la distribución de los datos con Distribución Gumbel", col = "pink")

curve(dgumbel(x, gumbel_fit$estimate[1], gumbel_fit$estimate[2]), add = TRUE, col = "blue", lwd
= 2)

lines(density(monthly_max), col="red", lwd = 2)

legend("topright", col = c("blue", "red"), legend = c("Densidad Gumbel teórica", "Densidad empír
ica"), lwd = 2, bty = "n", cex = 0.7)
```




```
# Comparación de las probabilidades acumuladas empíricas y teóricas

gumbel_teorica <- pgumbel(0:max(monthly_max), a = gumbel_fit$estimate[1], b = gumbel_fit$estimat
e[2])

plot(ecdf(monthly_max), main = "Comparación con la Distribución Gumbel", xlab = "Precipitacione
s", col="blue", xlim = c(0, max(monthly_max)), ylim = c(0, 1.05))

par(new=TRUE)

plot(0:max(monthly_max), gumbel_teorica, type = "l", main = "", xlab = "", ylab = "", col = "re
d", lwd = 2, ylim = c(0, 1.05), xaxt = "n", yaxt = "n")

legend("topleft", col = c("red", "blue"), legend = c("Densidad teórica Gumbel", "Densidad empíri
ca"), lwd = 2, bty = "n", cex = 0.7)
```




```
# QQ plot

plot(gumbel_fit)
```




```
# Prueba de Kolmogorov-Smirnov (KS) para la Gumbel

gumbel_exceedance <- 1 - pgumbel(monthly_max, gumbel_fit$estimate[1], gumbel_fit$estimate[2])

ks_result <- ks.test(monthly_max, "pgumbel", gumbel_fit$estimate[1], gumbel_fit$estimate[2])

print(ks_result)
```

```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  monthly_max
## D = 0.15035, p-value = 0.4619
## alternative hypothesis: two-sided
```

```
# Estimación de parámetros

mean_data <- mean(monthly_max)
sd_data <- sd(monthly_max)
a_approx <- mean_data - 0.5772 * (sd_data * sqrt(6) / pi)
b_approx <- (sd_data * sqrt(6)) / pi
cat("Parámetros aproximados: a =", a_approx, ", b =", b_approx, "\n")
```

```
## Parámetros aproximados: a = 140.8772 , b = 41.62534
```

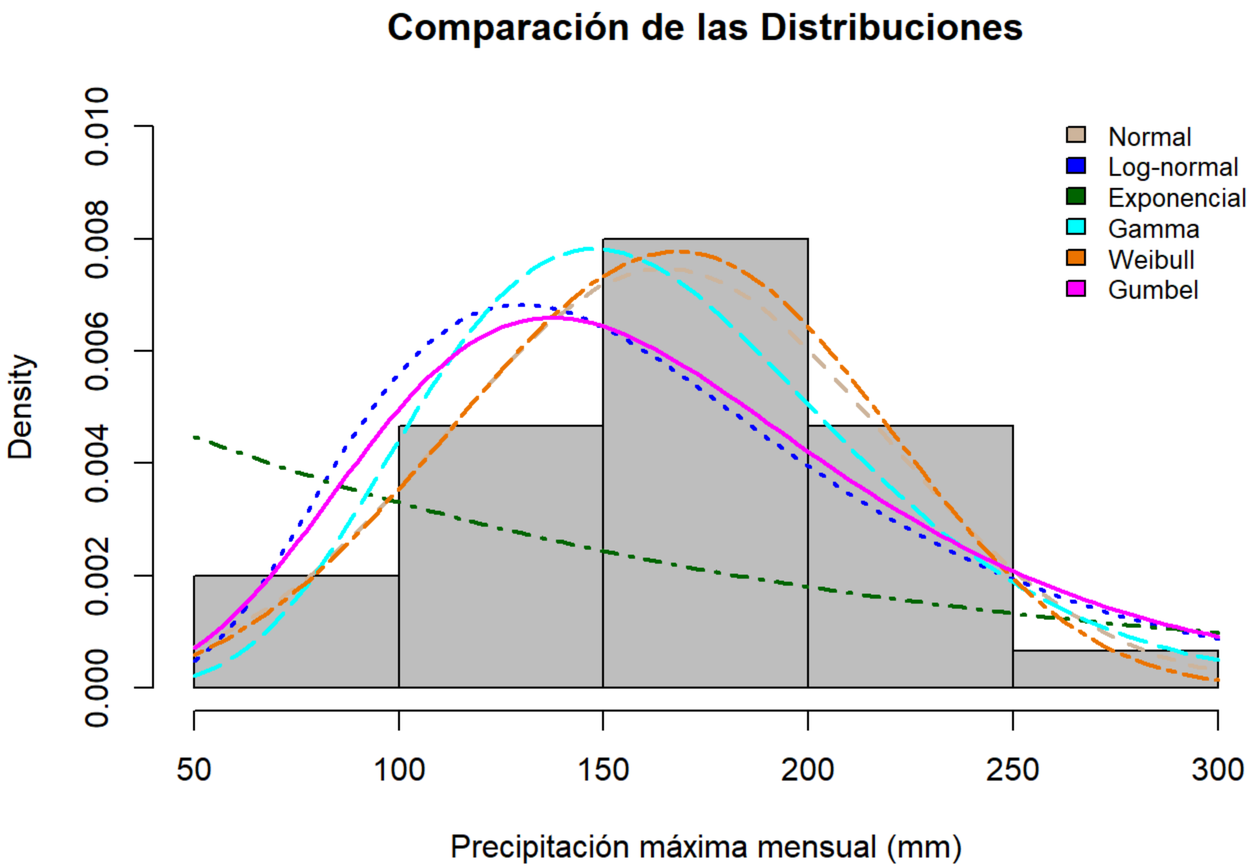
```
cat("Parámetros estimados con fitdistrplus: a =", gumbel_fit$estimate[1], ", b =", gumbel_fit$estimate[2], "\n")
```

```
## Parámetros estimados con fitdistrplus: a = 137.3022 , b = 55.76204
```

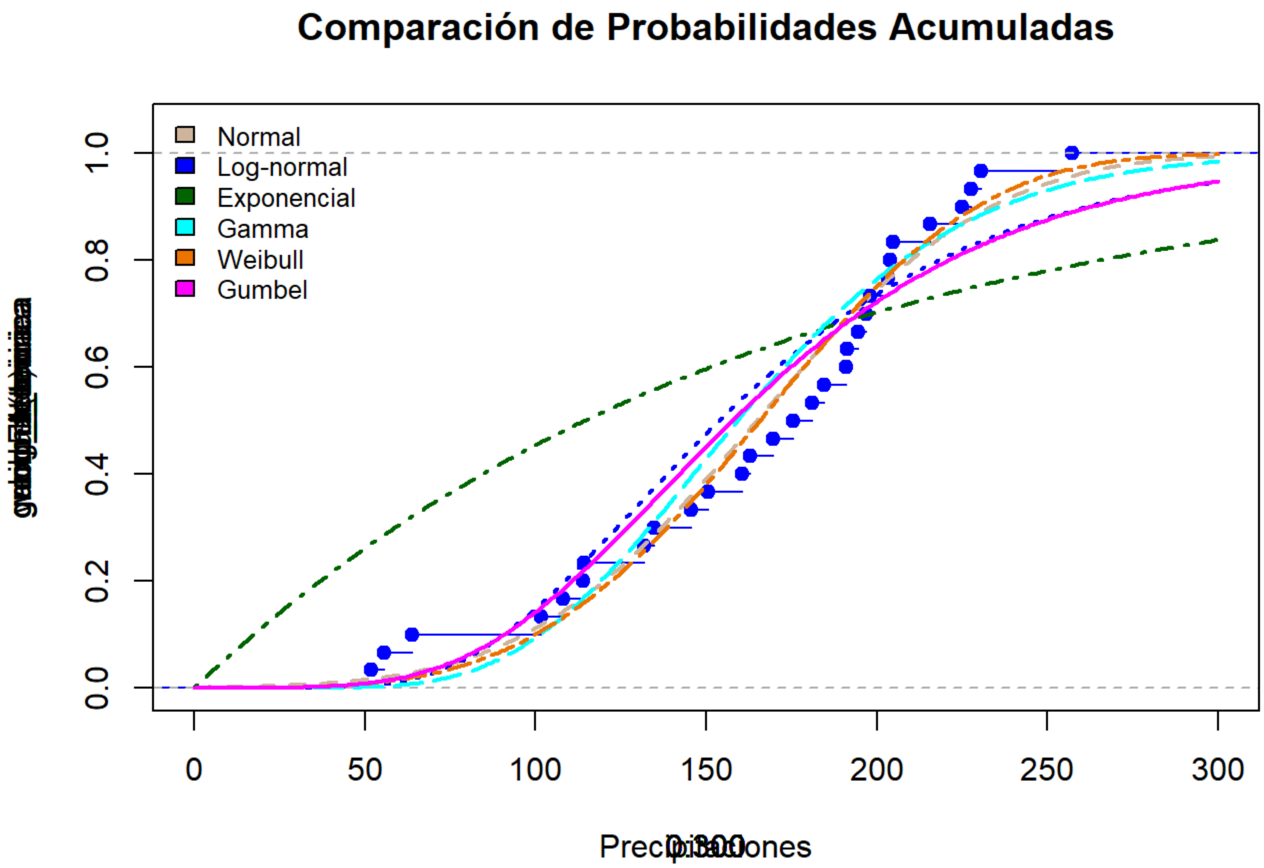
g. **Compara los ajustes de las distribuciones que analizaste**

- Haz un gráfico comparativo de los histogramas de densidad empírica vs densidad teórica de todas las distribuciones que analizaste (todas las distribuciones en un solo gráfico.
- Haz un gráfico comparativo de las probabilidades acumuladas empírica vs teóricas de todas las distribuciones que analizaste (todas las distribuciones en un solo gráfico.
- Define cuál es la mejor distribución que se ajusta a tus datos. Argumenta interpretando la comparación entre los gráficos y analizando las pruebas de ajuste de curva.

```
# Normal

mean_norm <- mean(monthly_max)
sd_norm <- sd(monthly_max)

# Log-normal

mean_lognorm <- mean(log(monthly_max))
sd_lognorm <- sd(log(monthly_max))

# Exponencial

rate_exp <- 1 / mean(monthly_max)

# Gamma

shape_gamma <- mean(monthly_max)^2 / var(monthly_max)
rate_gamma <- mean(monthly_max) / var(monthly_max)

# Weibull

weibull_fit <- fitdistr(monthly_max, "weibull", lower=c(0, 0))
shape_weibull <- weibull_fit$estimate[1]
scale_weibull <- weibull_fit$estimate[2]

# Gumbel

gumbel_fit <- fitdist(monthly_max, "gumbel", start=list(a=1, b=1))
a_gumbel <- gumbel_fit$estimate[1]
b_gumbel <- gumbel_fit$estimate[2]

# Gráfico comparativo de histogramas de densidad empírica vs teórica

hist(monthly_max, xlab="Precipitación máxima mensual (mm)", freq=FALSE, ylim=c(0, 0.01), main="C
omparación de las Distribuciones", col = "gray")
curve(dnorm(x, mean=mean_norm, sd=sd_norm), add=TRUE, col="bisque3", lwd=2, lty=2)
curve(dlnorm(x, mean=mean_lognorm, sd=sd_lognorm), add=TRUE, col="blue", lwd=2, lty=3)
curve(dexp(x, rate=rate_exp), add=TRUE, col="darkgreen", lwd=2, lty=4)
curve(dgamma(x, shape=shape_gamma, rate=rate_gamma), add=TRUE, col="cyan", lwd=2, lty=5)
curve(dweibull(x, shape=shape_weibull, scale=scale_weibull), add=TRUE, col="darkorange2", lwd=2,
lty=6)
curve(dgumbel(x, a_gumbel, b_gumbel), add=TRUE, col="magenta", lwd=2, lty=7)
legend("topright", legend=c("Normal", "Log-normal", "Exponencial", "Gamma", "Weibull", "Gumbe
l"),
       fill=c("bisque3", "blue", "darkgreen", "cyan", "darkorange2", "magenta"), cex=0.8, bty
="n")
```


Comparación de las Distribuciones
Density
0.000
0.002
0.004
0.006
0.008
0.010
50
100
150
200
250
300
Precipitación máxima mensual (mm)
Normal
Log-normal
Exponencial
Gamma
Weibull
Gumbel

```
# Gráfico comparativo de probabilidades acumuladas empírica vs teóricas

plot(ecdf(monthly_max), main="Comparación de Probabilidades Acumuladas", xlab="Precipitaciones",
col="blue", xlim=c(0, 300), ylim=c(0, 1.05))
norm_teorica <- pnorm(0:300, mean = mean_norm, sd=sd_norm)
log_teorica <- plnorm(0:300, mean = mean_lognorm, sd=sd_lognorm)
exp_teorica <- pexp(0:300, rate = rate_exp)
gamma_teorica <- pgamma(0:300, shape=shape_gamma, rate=rate_gamma)
weibull_teorica <- pweibull(0:300, shape=shape_weibull, scale=scale_weibull)
gumbel_teorica <- pgumbel(0:300, a=a_gumbel, b=b_gumbel)

par(new=TRUE); plot(0:300, norm_teorica, type = "l", col = "bisque3", lwd = 2, ylim = c(0, 1.0
5), xaxt = "n", yaxt = "n", lty = 2)
par(new=TRUE); plot(0:300, log_teorica, type = "l", col = "blue", lwd = 2, ylim = c(0, 1.05), xa
xt = "n", yaxt = "n", lty = 3)
par(new=TRUE); plot(0:300, exp_teorica, type="l", col="darkgreen", lwd=2, ylim=c(0, 1.05), xaxt
="n", yaxt="n", lty=4)
par(new=TRUE); plot(0:300, gamma_teorica, type="l", col="cyan", lwd=2, ylim=c(0, 1.05), xaxt
="n", yaxt="n", lty=5)
par(new=TRUE); plot(0:300, weibull_teorica, type="l", col="darkorange2", lwd = 2, ylim = c(0, 1.
05), xaxt="n", yaxt="n", lty=6)
par(new=TRUE); plot(0:300, gumbel_teorica, type="l", col="magenta", lwd=2, ylim=c(0, 1.05), xaxt
="n", yaxt="n", lty=7)
legend("topleft", legend = c("Normal", "Log-normal", "Exponencial", "Gamma", "Weibull", "Gumbe
l"),
       fill=c("bisque3", "blue", "darkgreen", "cyan", "darkorange2", "magenta"), cex=0.8, bty
="n")
```



Con base en la gráfica anterior se puede observar que de las 6 distribuciones usadas, solo dos se ajustan bien a los datos, las cuales son la Normal y la Weibull. En este caso, la Distribución Weibull es la que se utilizará para el diseño de las obras hidráulicas porque muestra un mejor ajuste. En el caso de las otras distribuciones se tiene que la peor fue la Exponencial, y las demás simplemente no se ajustaron bien a los datos de la Ciudad de México.

Diseño de obras hidráulicas

4. Precipitación de diseño de obras hidráulicas

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

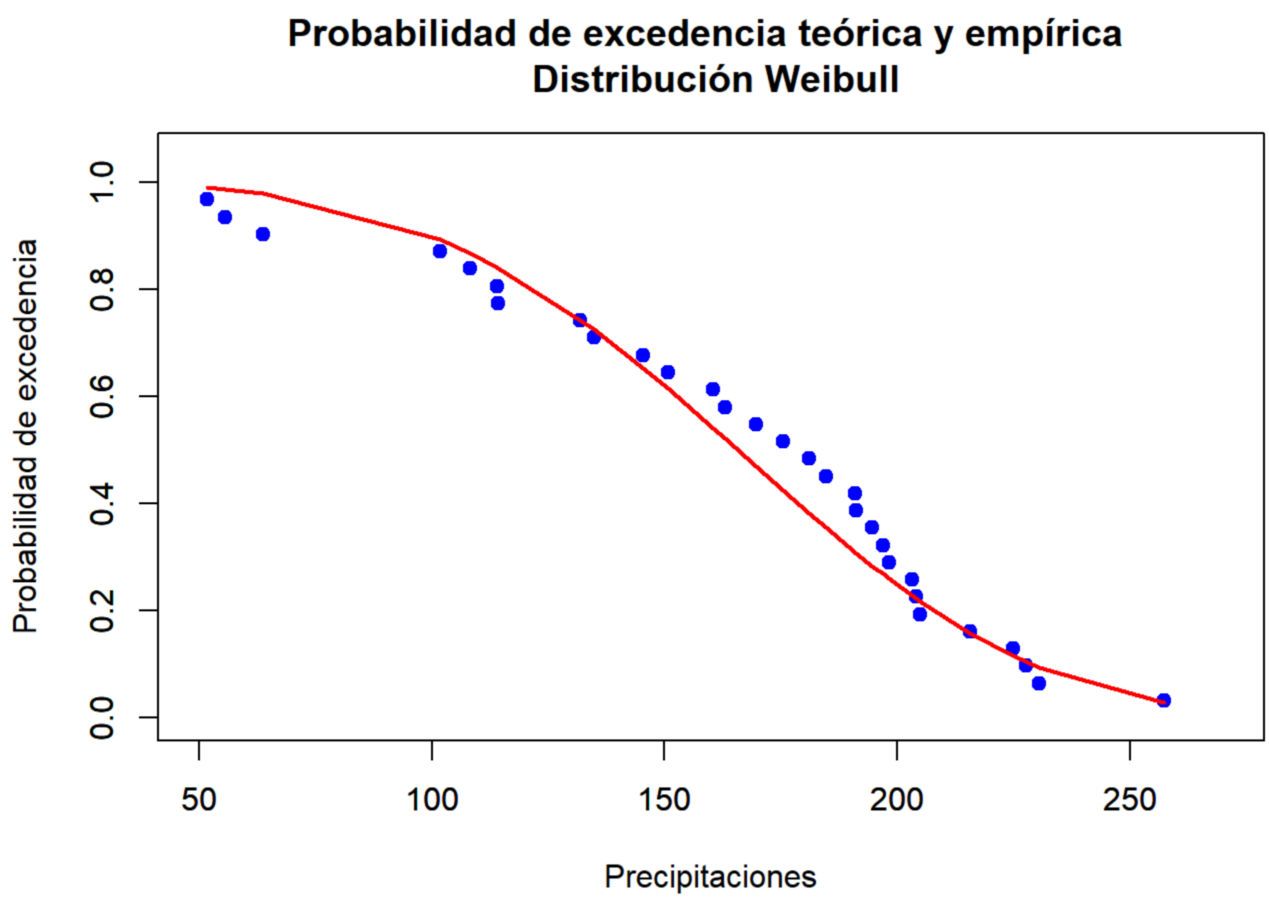
a. Haz el gráfico comparativo de la probabilidad de excedencia teórica vs empírica. ¿Qué te indica ese gráfico? interpreta y argumenta la certeza de la selección de la distribución elegida.

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

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

d. El resultado de este ejemplo será una aproximación del caudal máximo que se tendría en Aguascalientes con un periodo de retorno de 200 años. ¿Qué significa este valor? ¿Qué pasa si incrementamos el periodo de retorno? ¿El caudal máximo para este periodo de retorno será el mismo si utilizamos datos históricos de otro estado? ¿Por qué crees que las obras hidráulicas deben diseñarse en base a periodos de retorno sugeridos? ¿Por qué es importante conocer la distribución de probabilidad a la que mejor se aproximan los datos históricos? Explora otros periodos de retorno diferentes a los que se proporcionan en los periodos sugeridos para contestar esta pregunta.

```
# Gráfico comparativo

weibull_exe <- 1 - pweibull(rain_analysis$order_max_rain, shape = shape_weibull, scale = scale_w
eibull)

plot(rain_analysis$order_max_rain, rain_analysis$Pexe, main = "Probabilidad de excedencia teóric
a y empírica \n Distribución Weibull", xlab = "Precipitaciones", ylab = "Probabilidad de exceden
cia", col = "blue", xlim = c(50,270), ylim = c(0, 1.05), pch = 19)

par(new=TRUE)

plot(rain_analysis$order_max_rain, weibull_exe, type = "l", main = "", xlab = "", ylab = "", col
= "red", lwd = 2, xlim = c(50,270), ylim = c(0, 1.05))
```



```
# Cálculo de la probabilidad de excedencia para el periodo de retorno

periodo_retorno <- 100
Pexe <- 1 / periodo_retorno

cat("Probabilidad de excedencia para un periodo de retorno de", periodo_retorno, "años:", Pexe,
"\n")
```

```
## Probabilidad de excedencia para un periodo de retorno de 100 años: 0.01
```

```
# Cálculo de la precipitación máxima mensual para el periodo de retorno

Pnoexe <- 1 - Pexe

# Cálculo del cuantil para la distribución Weibull

precipitacion_diseño <- qweibull(Pnoexe, shape = shape_weibull, scale = scale_weibull)

cat("Precipitación de diseño para un periodo de retorno de", periodo_retorno, "años:", precipita
cion_diseño, "mm\n")
```

```
## Precipitación de diseño para un periodo de retorno de 100 años: 275.9524 mm
```

El valor de 275.95 mm representa la precipitación máxima mensual que, en promedio, se espera que ocurra una vez cada 100 años en la CDMX. Este valor es esencial para el diseño de infraestructura hidráulica, ya que permite anticipar el impacto de eventos de precipitación extrema. Si incrementamos el periodo de retorno, la precipitación de diseño aumentará, reflejando eventos más raros y de mayor magnitud. Esto es importante porque estructuras como presas deben ser capaces de resistir estos eventos extremos para minimizar riesgos de fallas o inundaciones. El caudal máximo calculado para un periodo de retorno específico no sería el mismo si se utilizaran datos históricos de otro estado, ya que las características climatológicas y patrones de precipitación varían entre regiones. Esto subraya la necesidad de basar el diseño en datos locales y representativos. Diseñar obras hidráulicas en base a periodos de retorno recomendados asegura que las estructuras estén preparadas para eventos extremos dentro de un rango de frecuencia aceptable, equilibrando costos y seguridad. Conocer la distribución de probabilidad que mejor ajusta los datos históricos es crucial, ya que permite realizar estimaciones más precisas de eventos raros. Experimentar con diferentes periodos de retorno permite evaluar la resiliencia de la infraestructura ante eventos de distintas magnitudes, lo cual es clave para una gestión de riesgos efectiva.