Actividad Integradora 1_fer

Fernanda Pérez

2024-10-24

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

ANÁLISIS

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

A) Descarga la base de datos de precipitaciones máximas históricas mensuales de todos los estados de la república. Estado seleccionado: Sinaloa

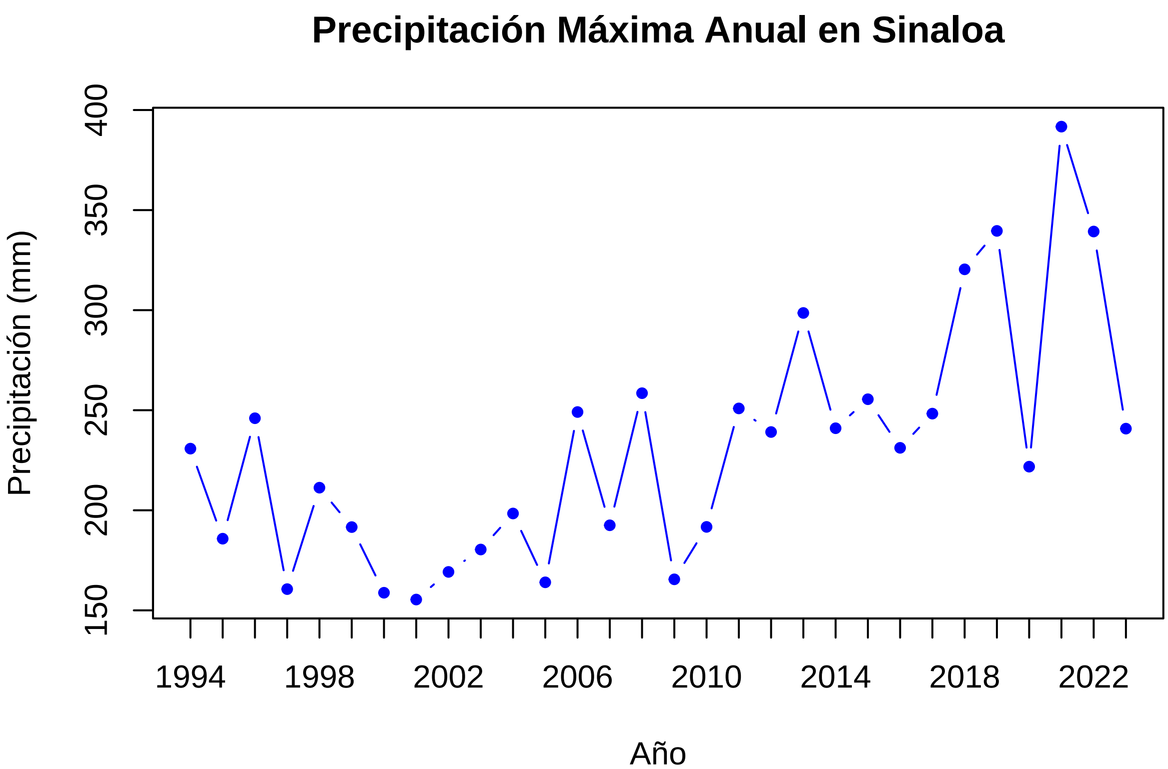
```
rain_data <- read.delim("D:/Downloads/precipitaciones_maximas_mensuales.txt", sep="\t", stringsAsFactor

rain_Sinaloa <- rain_data[rain_data$Estado == "Sinaloa", ]

head(rain_Sinaloa)
```

```
##     Anio Mes  Estado Lluvia
## 25  1994 Ene Sinaloa    0.0
## 58  1994 Feb Sinaloa    0.0
## 91  1994 Mar Sinaloa    4.2
## 124 1994 Abr Sinaloa    1.4
## 157 1994 May Sinaloa    0.0
## 190 1994 Jun Sinaloa   39.7
```

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(rain_Sinaloa$Anio)

monthly_max <- c()
for (n in 1:length(years)) {
  monthly_max <- c(monthly_max, max(rain_Sinaloa$Lluvia[rain_Sinaloa$Anio == years[n]]))
```

```
}

plot(monthly_max, type="b", pch=20, col="blue", xlab="Año", ylab="Precipitación (mm)",
     main="Precipitación Máxima Anual en Sinaloa", xaxt="n")
axis(1, at=1:length(years), labels=years)
```



El grafico nos muestra la precipitación máxima anual en Sinaloa desde 1994 hasta 2023.

Conforme pasan los años, se observa una tendencia general de aumento en las precipitaciones máximas anuales, en especial a partir de 2010. El año que tuvo la precipitación más alta fue alrededor de 2022. *Se observa que los picos más altos de precipitación máxima son alrededor de los años 2017, 2019 y 2022, ya que estos años destacan al tener valores superiores a 350 mm, indicando eventos de precipitación extrema durante esos períodos.* Los periodos más secos fueron entre 1995 y 2000, las precipitaciones máximas tienden a estar por debajo de los 200 mm. *Observamos una mayor amplitud en los valores de precipitación a partir de 2010, o sea que los eventos extremos podrían haberse vuelto más frecuentes.

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

```
media <- mean(rain_Sinaloa$Lluvia)
mediana <- median(rain_Sinaloa$Lluvia)
```

```r
desviacion_estandar <- sd(rain_Sinaloa$Lluvia)
minimo <- min(rain_Sinaloa$Lluvia)
maximo <- max(rain_Sinaloa$Lluvia)

cat("Media:", media, "\n")
```

```
## Media: 59.40694
```

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

```
## Mediana: 14
```

```r
cat("Desviacion Estandar:", desviacion_estandar, "\n")
```

```
## Desviacion Estandar: 81.99136
```

```r
cat("Minimo:", minimo, "\n")
```

```
## Minimo: 0
```

```r
cat("Maximo:", maximo, "\n")
```

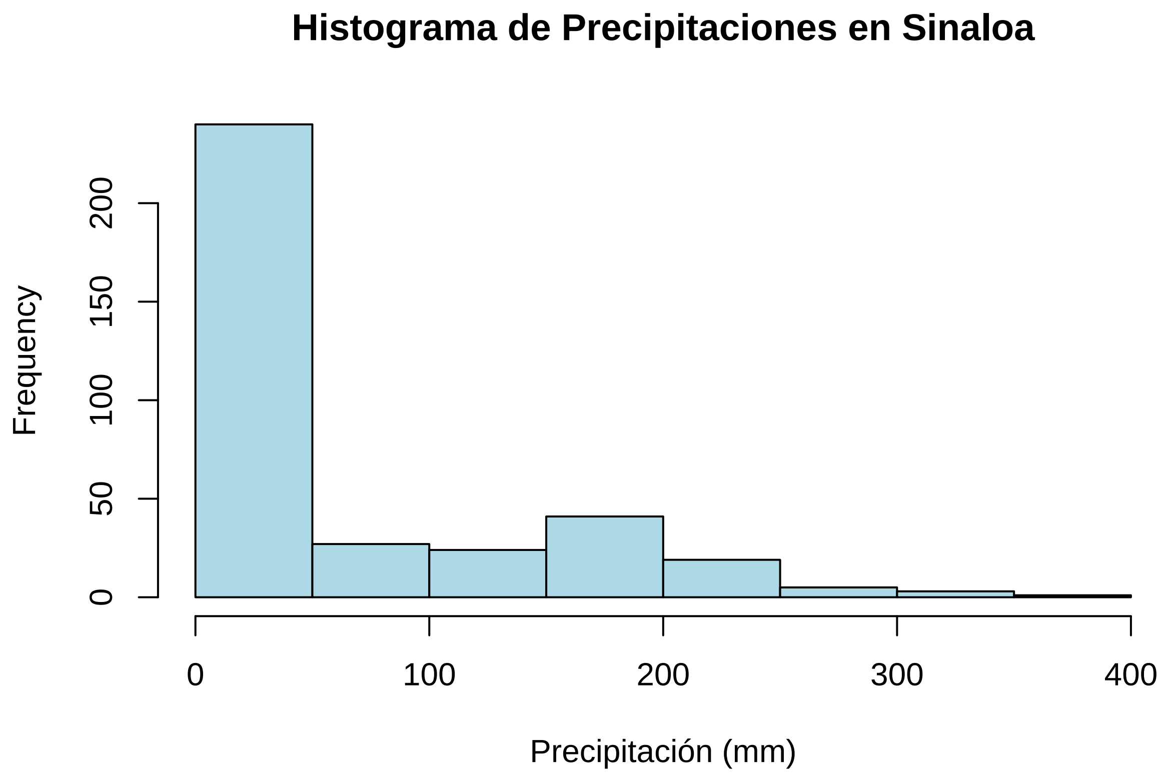
```
## Maximo: 391.7
```

La media nos indica que, en promedio, la precipitación máxima mensual en Sinaloa es de aproximadamente 59.406 mm La desviación estándar es alta en comparación con la media, esto indica una gran variabilidad en las precipitaciones mensuales.Indicandonos que los valores tienden a desviarse considerablemente del promedio, lo cual se alinea con la presencia de eventos de lluvia extremos en ciertos meses. *El valor mínimo de precipitación es 0 mm, lo cual significa que hubo meses en los que no se registró precipitación máxima. O sea que hubo meses sin lluvias significativas, lo cual es congrunte dado el clima seco del estado.* El valor máximo de precipitación mensual es de 391.7 mm, lo que representa que hubo evento extremo de lluvia.

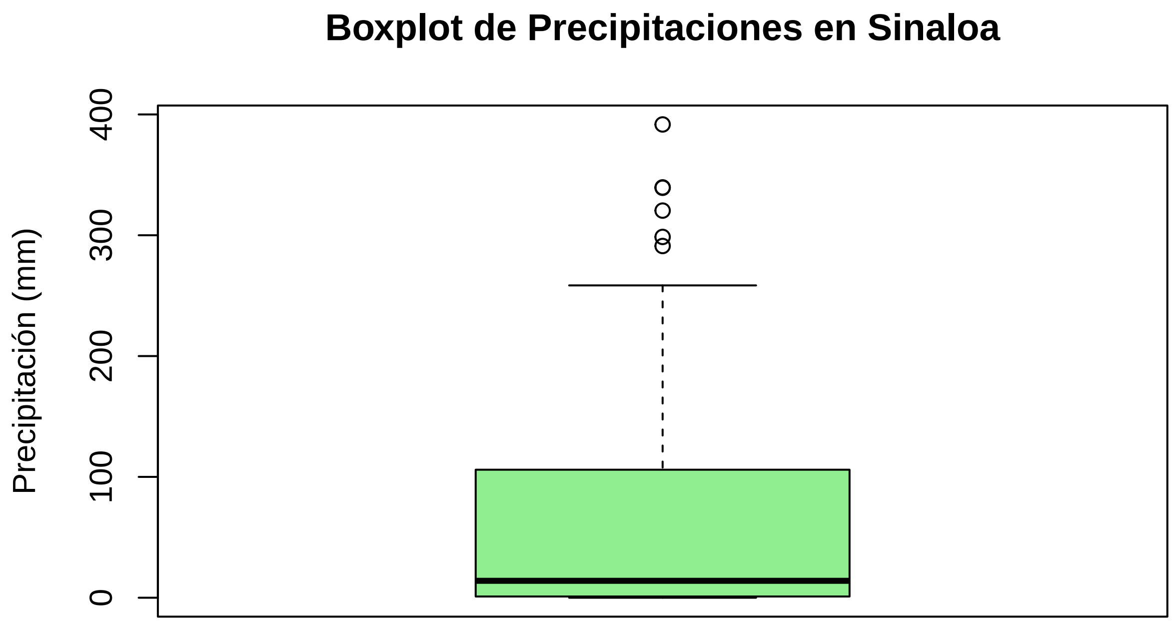
- Distribución sesgada: La diferencia considerable que hay entre la media y la mediana nos alerta de que los datos están sesgados hacia la derecha, o sea que hay varios valores bajos y algunos valores extremadamente altos. -Eventos extremos: La alta desviación estándar y el valor máximo nos indican que Sinaloa cuenta con eventos de lluvia extrema en ciertos meses, generando una gran 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

```r
# Histograma de la distribución de precipitaciones
hist(rain_Sinaloa$Lluvia, main="Histograma de Precipitaciones en Sinaloa",
     xlab="Precipitación (mm)", col="lightblue", border="black")
```




```
# Boxplot de la distribución de precipitaciones
boxplot(rain_Sinaloa$Lluvia, main="Boxplot de Precipitaciones en Sinaloa",
        ylab="Precipitación (mm)", col="lightgreen")
```



-Histograma de precipitaciones: *El histograma muestra una distribución sesgada a la derecha, vemosuna cantidad alta de meses con precipitaciones cercanas a 0 mm. O sea que la mayoría de los meses tienen precipitaciones bajas, lo que es típico en climas con temporadas secas.* Vemos que la barra más alta en el rango de 0 a 50 mm representa más de 200 observaciones, lo que indica que muchos meses tienen precipitaciones muy bajas.

-Boxplot de precipitaciones: *El rango intercuartil es relativamente bajo, lo que nosdice que la mayoría de los valores están concentrados en el rango inferior de precipitaciones, conclusión similar a la que llegamos en el histograma.* En la parte superiror observamos varios outliers (valores atípicos) por encima del límite superior de la caja. Estos outliers representan los eventos de precipitación extrema que hemos identificado, y en algunos meses que exceden los 300 mm. *La mediana al ser mucho más baja que los outliers, nos refleja la naturaleza sesgada de la distribución.* Vemos una distribucón asimetrica ya que la mediana está más cerca de la parte inferior de la caja, lo que indica que la mayoría de los valores se encuentran hacia la parte baja del rango de precipitaciones.

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

D) ¿Qué puedes concluir observando la gráfica de los máximos mensuales anuales para tu Estado?

Al observar la gráfica de los máximos mensuales anuales en el estados seleccionado, Sinaloa, se puede concluir que, aunque hay algunas subidas y bajadas, parece que en general, después de 2010, la cantidad de lluvia en los eventos más extremos está aumentando. Lo cual significa que las lluvias más intensas estan ocurriendo con más frecuencia y/o son más fuertes. Es importante resaltarlo ya que las lluvias extremas pueden causar inundacione severas en el estado, y esto podria relacionarse con los drasticos cambios de clima. Se concluye que la cantidad de lluvia extrema si varía cada año, sin embargo, después de 2010, las lluvias más fuertes son cada vez mayores.

D) ¿Observas alguna tendencia?

Sí, podemos observar una tendencia al aumento de las precipitaciones máximas anuales en los años más recientes, vemos algunos picos significativos alrededor de 2017, 2019 y 2022. Anteror al año de 2010, las precipitaciones máximas fueron más o menos bajas y un poco más constantes, sin embargo en la última década ha habido un aumento en la magnitud de los eventos de lluvia extrema.

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

No se puede establecer un ciclo claro y periódico en el que las precipitaciones suben o bajan, sin embargo si podemos establecer que los eventos de lluvia intensa se han vuelto más frecuentes en los últimos años. Esto podria deberse a factores meteorologicos y/o climaticos que influyen en estos patrones.

D) ¿Para qué nos sirve analizar este tipo de gráficas?

El analizar este tipo de gráficas es sumamente importante para identificar tendencias en eventos climáticos extremos, lo que nos puede ser de gran utilidad para:

Que los ingenieros pueden analisar esta información para poder diseñar sistemas de drenaje, carreteras y edificios que sean capaces de soportar eventos de precipitación extrema, es decir, planificar las infraestructuras adecuadamente. Al anticipar estos eventos extremos se pueden desarrollar estrategias de prevención y mitigación de riesgos relacionados con inundaciones.

Este tipo de análisis es vital para que las autoridades tomen decisiones de prevención y adaptación al cambio climático y la protección contra los fenómenos meteorológicos 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.

```r
rain_analysis <- data.frame(max_rain = monthly_max,
                            order_max_rain = sort(monthly_max, 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

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

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:

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

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

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

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

```
rain_analysis$Pnoexe <- 1 - rain_analysis$Pexe
```

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

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

```
rain_analysis$Pret <- 1 / rain_analysis$Pexe
```

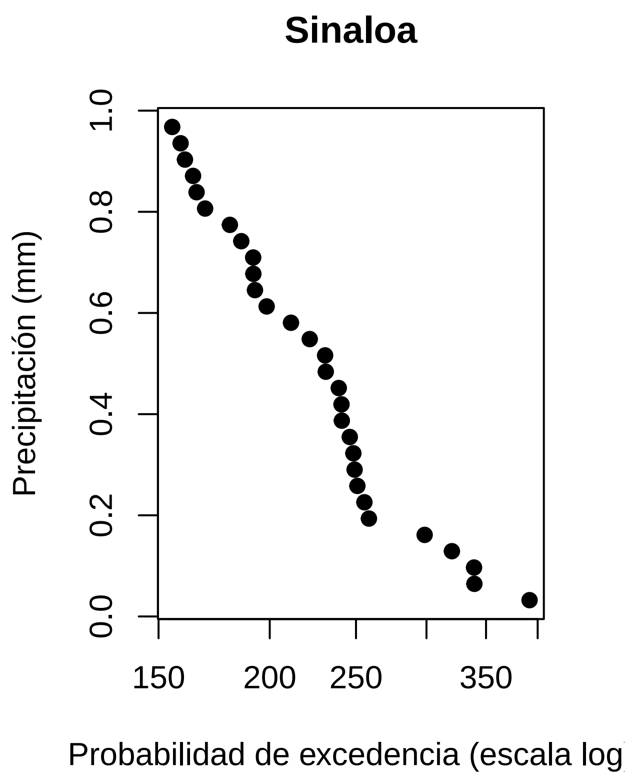
E.1) Se muestran las primeras filas de la tabla resultante:

```
head(rain_analysis)
```

```
##   max_rain order_max_rain rank_rain       Pexe    Pnoexe      Pret
## 1    230.8          391.7         1 0.03225806 0.9677419 31.000000
## 2    185.8          339.6         2 0.06451613 0.9354839 15.500000
## 3    246.0          339.3         3 0.09677419 0.9032258 10.333333
## 4    160.6          320.4         4 0.12903226 0.8709677  7.750000
## 5    211.3          298.6         5 0.16129032 0.8387097  6.200000
## 6    191.6          258.5         6 0.19354839 0.8064516  5.166667
```

F) Describe las gráficas obtenidas. Gráfico de la probabilidad de excedencia:

```r
par(mfrow = c(1, 2))

plot(rain_analysis$order_max_rain, rain_analysis$Pexe, log = "x", pch = 19,
     main = "Sinaloa", xlab = "Probabilidad de excedencia (escala log)",
     ylab = "Precipitación (mm)")
```



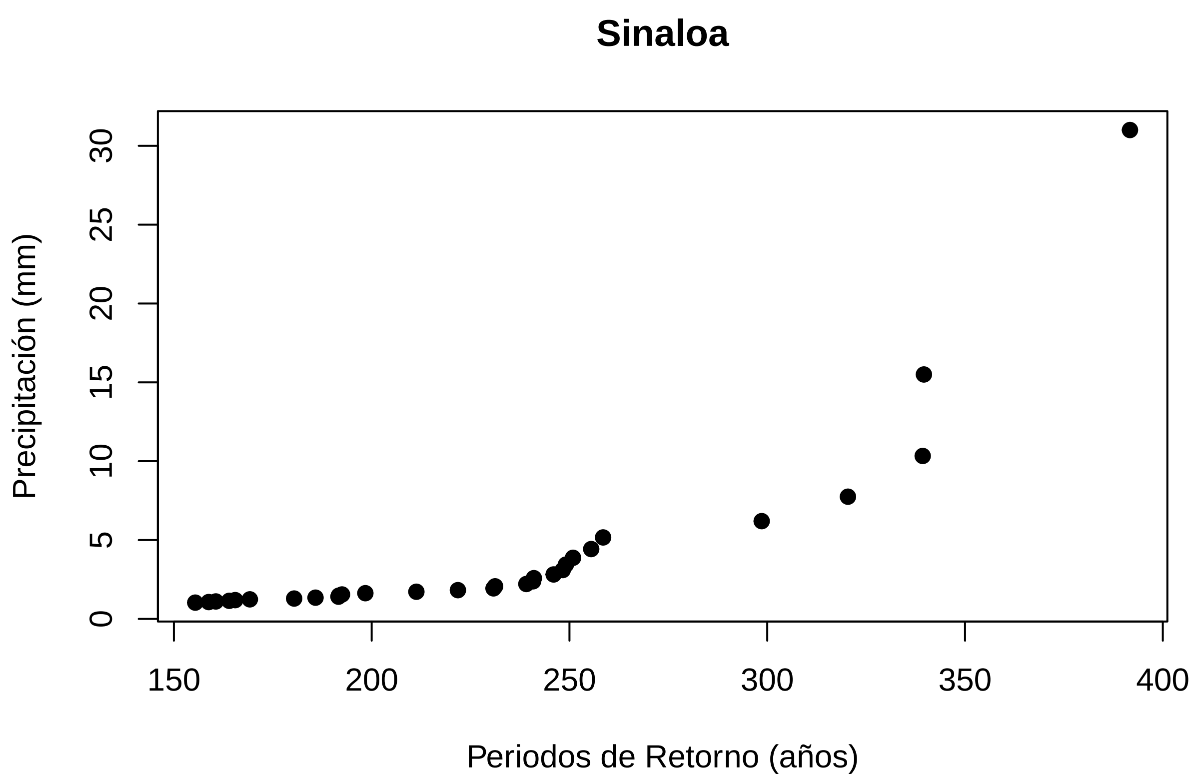
La primera gráfica muestra la probabilidad de excedencia en relación con la precipitación máxima (en mm). En el gráfico de la probabilidad de excedencia para Sinaloa vemos que hay una relación inversa entre la precipitación máxima y la probabilidad de excedencia.Lo cual significa que conforme aumenta la cantidad de precipitación máxima, la probabilidad de que ocurra una precipitación mayor disminuye. O sea que los eventos con precipitación alta son menos frecuentes.

La escala logarítmica en el eje de la probabilidad de excedencia resalta la frecuencia relativa de los eventos. Las precipitaciones más extremas, mayores a los 350mm, tienen una probabilidad de excedencia muy baja y por el contrario los eventos de precipitaciones más bajas son más comunes con una probabilidad de excedencia que se acerca a 1.

Con este gráfico se identifica que tan probable es que un evento de precipitación supere cierto valor. Los eventos con precipitación muy alta tienen una probabilidad muy baja de excedencia, indicando que son eventos raros.

F) Describe las gráficas obtenidas.Gráfico de periodo de retorno:

```
plot(rain_analysis$order_max_rain, rain_analysis$Pret, pch = 19,
     main = "Sinaloa", xlab = "Periodos de Retorno (años)",
     ylab = "Precipitación (mm)")
```



La segunda gráfica muestra la relación entre la precipitación máxima y el periodo de retorno. Vemos que en el grafico de periodo de retorno para Sinaloa hay una relación directa entre la precipitación máxima y el perido de retorno, o sea que mientras mayor sea la precipitación, tambien va a ser mayor el periodo de retorno, o sea que los eventos de precipitación más intensos ocurren con menor frecuencia.

El grafico nos muestra un incremento no lineal conforme aumenta la precipitación, indicandonos que los eventos más extremos tienen una probabilidad muy baja de repetirse en periodos cortos.

Con el grafico de periodo de retorno nos damos una idea del intervalo promedio de tiempo en que podemos esperar eventos de mayor magnitud. Para los fenomenos poco comunes pero de alto impacto hay que estar preparados con la infraestructura adecuada ya que en el grafico se destaca que los eventos extremos a pesar de ser de poco frecuencia si llegan a ocurrir.

F.1) ¿Qué significa la probabilidad de excedencia?

Definición: "Probabilidad de que una magnitud dada de un evento sea igual o excedida." ("Probabilidad de Excedencia," 2024). La probabilidad de excedencia es la probabilidad de que ocurra una precipitación mayor que cierto valor dado en un periodo específico. En este caso en la probabilidad de que en un año dado haya una precipitación que supere un umbal predeterminado. Una probabilidad de excedencia alta significa que son precipitaciones bajas o moderadas, o sea un evento comun; mientras que una probabilidad de excedencia baja indica precipitaciones muy intensas o altas, o sea que el evento es raro o poco comun.

F.2) ¿Qué significa el periodo de retorno?

"El Periodo de Retorno de cualquier evento extremo (lluvias torrenciales, temperaturas extremas, huracanes, etc.), se define como el lapso o número de años que en promedio, se cree que será igualado o excedido, es decir, es la frecuencia con la que se presenta un evento" (Mélice y Reason, 2007). Es decir el periodo de retorno es el intervalo de tiempo promedio entre eventos de una magnitud igual o mayor a un valor determinado. Se usa en hidrología para evaluar la frecuencia de eventos extremos y el riesgo asociado.

F.3) ¿Por qué es importante en hidrología?

En la rama de hidrología tanto la probabilidad de excedencia como el periodo de retorno son cruciales calcularlas y análizarlas ya que son decisivas para el diseño de estructuras y/o infraestructuras que se relacionan con el manejo del agua, como presas, drenajes, alcantarillado, etc.

Al permitir planificar en base en eventos de precipitación extema, las y los ingenieros se pueden asegurar de que la los diseños y las estructuras puedan soportar eventos pocos frecuentes pero desastrosos de alta magnitud.

El tiempo de retorno es trascendental para la gestión de riesgos, este ayuda a los ingenieros a diseñar las infraestructuras de tal manera que puedan soportar estos eventos de manera satisfactoria y sin fallas.

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

En el diseño de infraestructuras críticas usualmente se busca una probabilidad de excedencia baja, porque esto implica un periodo de retorno mayor ademas de un riesgo reducido de que suceda un evento de magnitud superior en un año cualquiera.

El valor deseable de la probabilidad de excedencia va a depender del tipo de obra que se este diseñando, por ejemplo para una obra critica como una presa, se busca que los valores de probabilidad de excedencia sean bajos para poder garantizar que pueda soportar eventos raros y extremos.

---

3. Análisis de Frecuencias Método Analítico

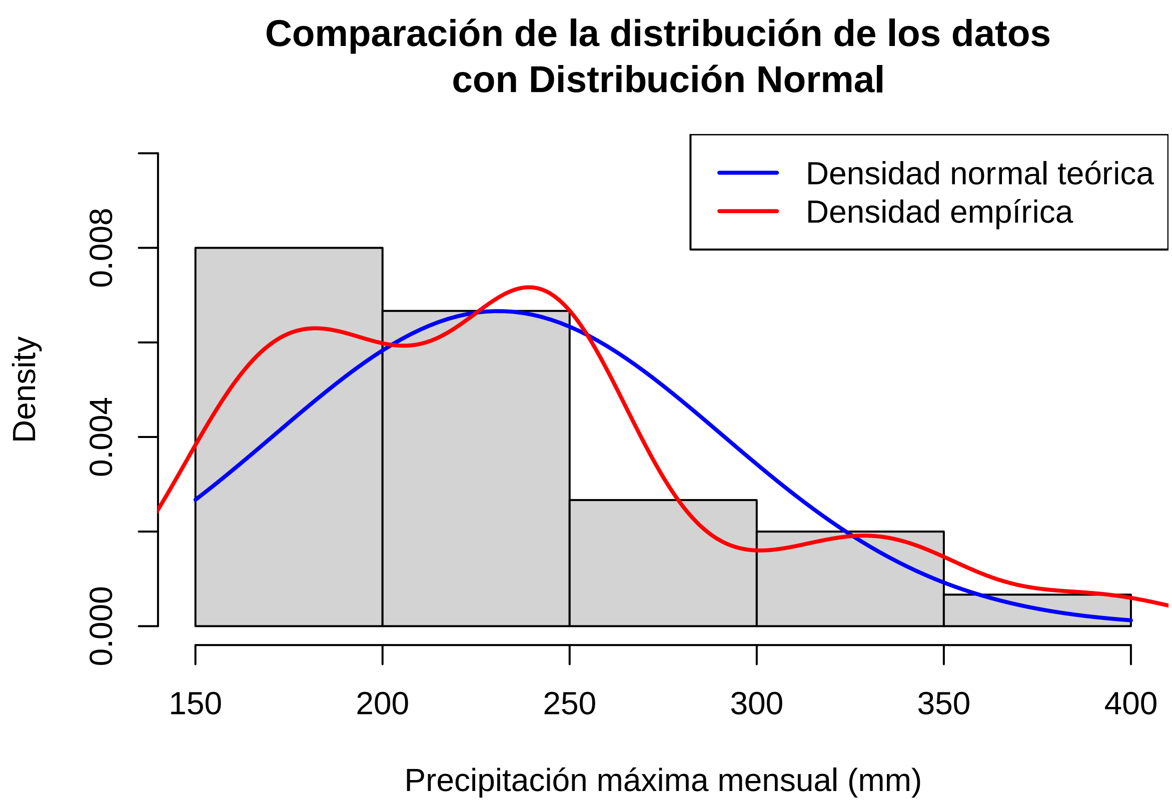
El método analítico consiste en asumir que los datos pueden ser ajustados a través de una función de densidad de probabilidades (FDP) conocida la cual nos servirá para modelar y pronosticar precipitaciones y periodos de retorno. Para ello es necesario probar varias distribuciones y emplear pruebas de 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.

```
hist(monthly_max, xlab="Precipitación máxima mensual (mm)", freq=FALSE, ylim=c(0, 0.01),
     main="Comparación de la distribución de los datos \n con Distribución Normal")
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 empírica"), lwd=2
```



De manera visual, ¿te parece que los datos se ajustan bien a una distribución normal? A primera vista podemos decir que no pareciera que los datos se ajusten perfectamente a una distribución normal. Vemos que la línea azul (la densidad teórica de una distribución normal), no se alinea a la perfección con la línea roja (la densidad empírica basada en los datos observados).

Vemos que en la primera barra del histograma (entre 150 y 200 mm), la densidad empírica supera a la densidad teórica, o sea que hay una mayor concentración de datos en los valores bajos de precipitación en comparación a lo que se espera con lo que predicción de la distribución normal. En la parte central, alrededor de los 225 mm , las curvas se encuentran un poco más alineadas, siendo un rango de valores que pareciera ajustarse un poco mejor a la distribución normal. *Y en la cola derecha, a partir de los 300 mm, la densidad empírica vemos que se separa significativamente de la normal teórica, nos muestra menos eventos extremos de lo que predice la distribución normal, indica que los datos tienen una distribución más sesgada que la normal.

De manera visual podemos decir que los datos no se ajustan bien a una distribución normal, vemos que hay muchas diferencias muy marcadas entre la densidad empírica y la teórica, se notan especialmente en las colas de la distribución. Esto indica que la precipitación máxima mensual tiene un comportamiento mucho más complejo y complicado, el cual no se puede explicar completamente con una distribución normal.

##Explica. ¿Cuántos parámetros tiene la distribución Normal? ¿Cuáles son? La distribución normal cuenta con dos parametros: la media y la desviación estandar.

La media es el valor promedio de la distribución. (El promedio de las precipitaciones máximas mensuales). La desviación estandar mide la dispersión de los datos alrededor de la media. (Refleja cuánta variabilidad hay en los datos de precipitación.)

##¿Por qué los parámetros se calculan de la forma en cómo se hace en el código? Los parámetros de la distribución normal se calculan con la media y la desviación estándar de los datos, dado que la distribución normal se define completamente por estos dos parámetros.

*Para calcular la media se suman todos los valores de precipitación dividida entre el número de observaciones, lo que nos da el centro de la distribución. En el código se calcula con mean(monthly_max)(El promedio de las precipitaciones máximas mensuales).

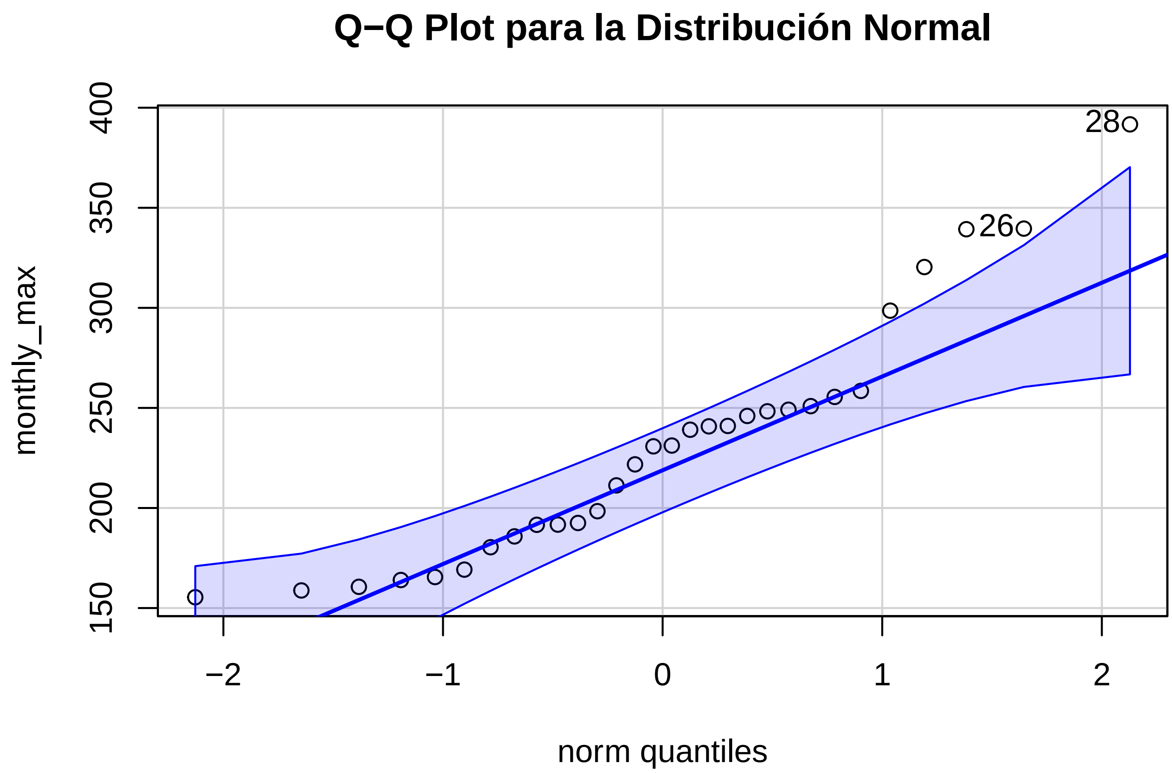
*Para calcular la desviación estándar , es decir, medir la dispersión de los datos alrededor de la media. En el código, se calcula con sd(monthly_max). (Refleja cuánta variabilidad hay en los datos de precipitación.)

A.2) Construye la gráfica qqplot.

```r
library(car)
```

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

```r
qqPlot(monthly_max, main="Q-Q Plot para la Distribución Normal")
```



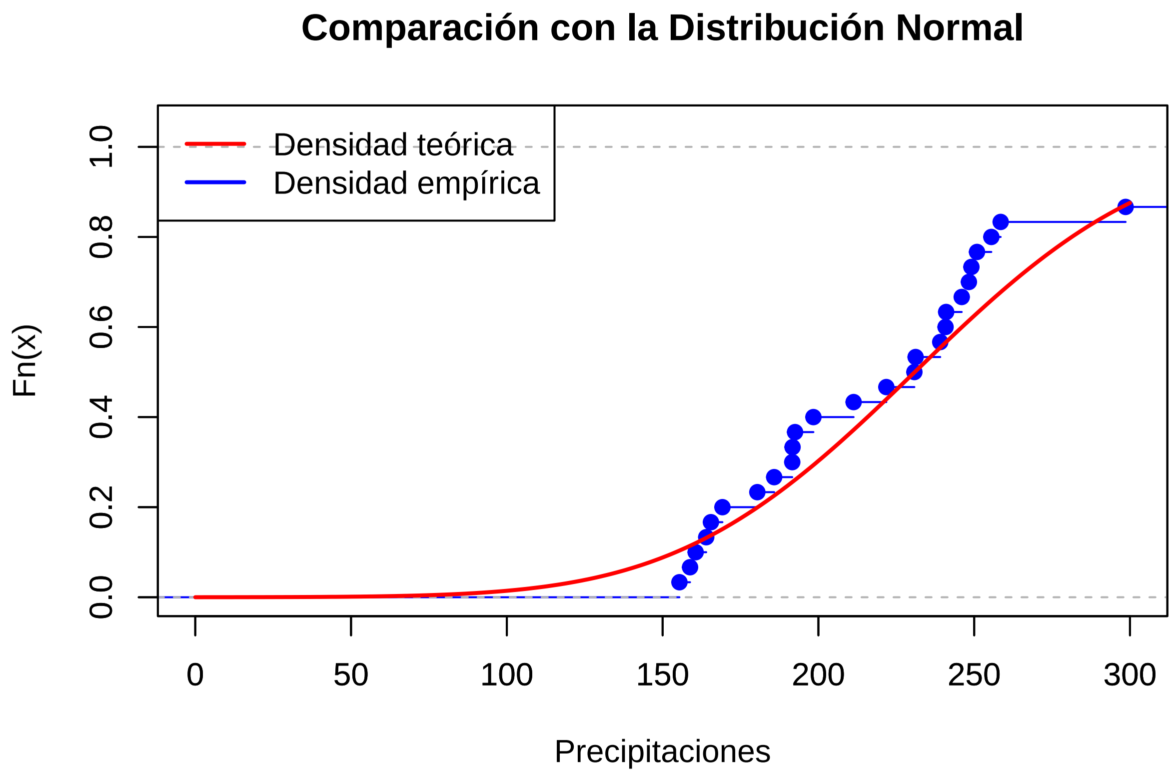

```
## [1] 28 26
```

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

Los puntos que vemos en el Q-Qplot no se ve que sigan completamente una distribución normal, los datos del centro se ven alineados con la línea teórica de la distribución normal, sin embargo en las colas vemos unas desviaciones significativas. En la cola izquierda,valores más bajos de precipitación, los datos estan menos dispersos que los estimados por la distribución normal. Y en la cola derecha, los valores más altos de precipitación vemos que se desvian significativamente hacia arriba.

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.

```
norm_teorica <- pnorm(0:300, mean=mean(monthly_max), sd=sd(monthly_max))
plot(ecdf(monthly_max), main="Comparación con la Distribución Normal", xlab="Precipitaciones", col="blu
par(new=TRUE)
plot(0:300, norm_teorica, type="l", main="", xlab="", ylab="", col="red", lwd=2, ylim=c(0, 1.05))
legend("topleft", col=c("red", "blue"), legend=c("Densidad teórica", "Densidad empírica"), lwd=2)
```



Explica qué son datos empíricos y datos teóricos. ¿Se parecen las distribuciones de probabilidad acumuladas? Los datos empíricos son los datos reales recolectados, en este caso son las precipitaciones máximas mensuales registradas en un periodo específico. Estos datos son los valores reales que ocurrieron en términos de eventos climáticos en la región. En este grafico los puntos azules y las líneas escalonadas representan la función de distribución empírica acumulada (ECDF).

Los datos teóricos son los valores que se esperarían si siguieran un modelo probabilístico en específico, que en este caso seria la distribución normal.

Las distribuciones de probabilidad acumuladas no se parecen completamente. Hay una similitud en la zona de los valores centrales, sin embargo las distribuciones empírica y teórica no coinciden, significando que la distribución normal no es un buen ajuste para describir los datos de precipitación máxima mensual

A.4) Utiliza dos pruebas de bondad de ajuste: Shapiro-Wilks y Kolmogorov-smirnov (KS).

```r
# Prueba de normalidad Shapiro-Wilk
shapiro.test(monthly_max)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  monthly_max
## W = 0.91498, p-value = 0.01991
```

```r
library(MASS)
# Prueba de Kolmogorov-Smirnov (KS) para la normalidad
ks.test(monthly_max, "pnorm", mean=mean(monthly_max), sd=sd(monthly_max))
```

```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  monthly_max
## D = 0.15592, p-value = 0.4167
## alternative hypothesis: two-sided
```

Hipótesis planteadas: H0: Los datos provienen de una distribución normal H1: Los datos no provienen de una distribución normal ## ¿Qué información nos dan las pruebas? Las pruebas de Shapiro-Wilk y Kolmogorov-Smirnov (KS) son pruebas de bondad de ajuste, son utulizadas para confirmar si cierto conjunto de datos sigue una distribución normal o no.

Con la prueba Shapiro-wilk se evalua la normalidad de los datos. Si se tiene un p-valor bajo indica que los datos muy probablemente no siguen una distribución normal.

Y la prueba Kolmogorov-Smirnov (KS) compara la distribución acumulada empírica con una distribución teórica, que en este caso sería la distribución normal. El tener un p-valor bajo señala que hay una diferencia significativa entre las distribuciones empírica y teórica.

¿Cuáles son los valores de los estadísticos?

Para Shapiro-Wilk: Estadistico W:0.91498 y para Kolmogorov-Smirnov: Estadístico D:0.15592. Con estos valores se puede resumir si se ajustan bien o no los datos a la distribución normal, entre más cercano sea el valor a 1, más fuerte va a ser el ajuste a la normalidad.

¿Cuál es el p-value de las pruebas?

Para Shapiro-Wilk: p-valor = 0.01991 y para Kolmogorov-Smirnov (KS): p-valor = 0.4167

¿Se aceptan o se rechazan las hipótesis nulas?

Para Shapiro-Wilk la hipótesis nula: H0: Los datos provienen de una distribución normal, como el p-valor es 0.01991, es menor que el nivel de significancia usual (0.05), por lo que rechazamos la hipotesis nula dando a entender que los datos no siguen una distribución normal de acuerdo a la prueba.

Para Kolmogorov-Smirnov (KS): H1: Los datos no provienen de una distribución normal, como en este caso tenemos un p-valor de 0.4167, es mayor que el nivel de significancia usual (0.05), o sea que no se rechaza la hipótesis nula, es decir no hay la suficiente evidencia para decir que los datos no siguen una distribución normal.

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

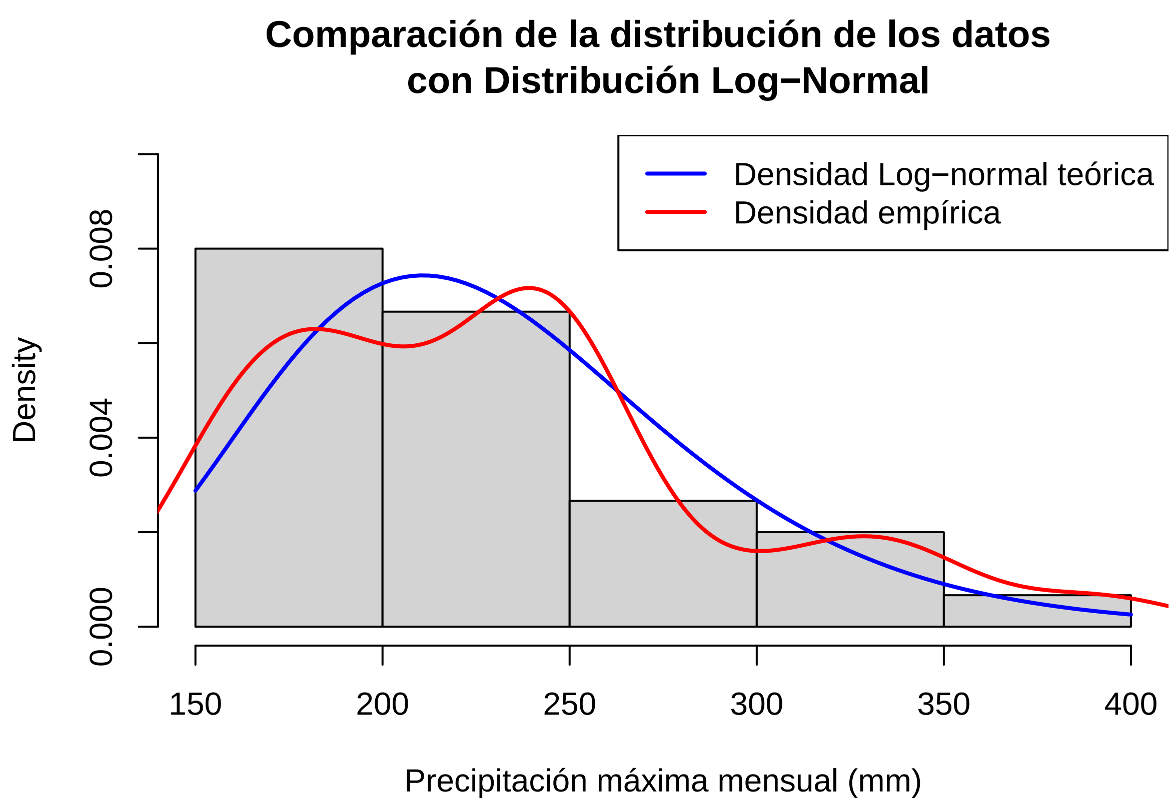
No se puede concluir que los datos sigan una distribución normal, con la prueba de Shapiro-Wilk concluimos que los datos no siguen una distribución normal y por otro lado con la prueba de Kolmogorov-Smirnov no rechaza la normalidad. Una de las dos pruebas rechazó la normalidad (Shapiro-Wilk) y la grafica del Q-Q plot y el histograma nos muestran desviaciones, es decir no tuvimos ajustes buenos, se puede concluir que los datos no siguen una distribución normal, en especial en los extremos(los eventos que se observan en las colas.)

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.

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

curve(dlnorm(x, meanlog=mean(log(monthly_max)), sdlog=sd(log(monthly_max))), add=TRUE, col="blue", lwd=

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

legend("topright", col=c("blue", "red"), legend=c("Densidad Log-normal teórica", "Densidad empírica"), 
```



De manera visual, ¿te parece que los datos se ajustan bien a una distribución Log-normal? Explica. De manera visual el ajuste con la distribución Log-Normal pareciera tener un mejor ajuste que la distribución normal, sin embargo no es perfecta.

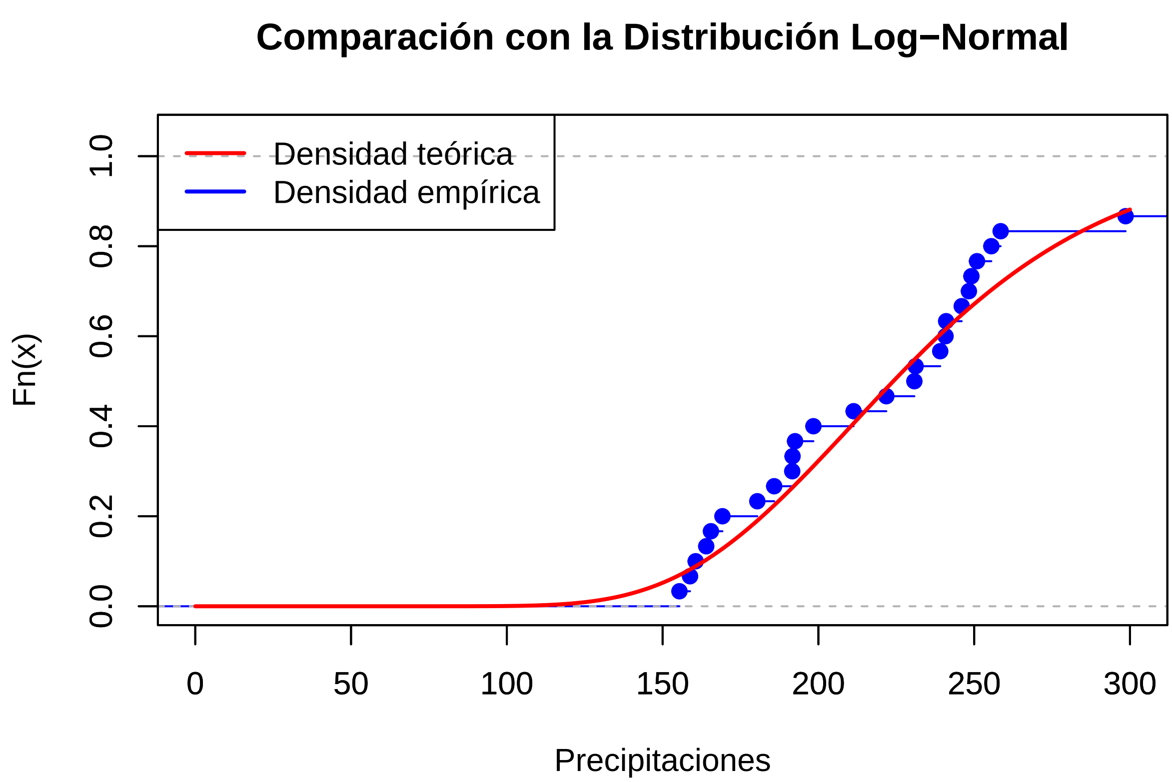
En la parte central entre los 200 y 250 mm, la curva azul (densidad teórica log-normal) sigue más de cerca la curva roja (densidad empírica) en comparación con el ajuste de la nromal.Campturando de manera más precisa los valores centrales de precipitación máxima mensual.

En el rango izquierdo(el más bajo) de 150 a 200, la curva empírica es un poco más alta que la teórica, significando que hay más observaciones en este rango de lo que se predijo con la densidad teorica.

Y en el otro extremo (cola derecha), es decir a partir de los 250 mm , vemos una separación significativa entre las dos curvas, la densidad empírica disminuye más rápidamente que la teórica, indicando que hay menos eventos extremos de lo que predice la distribución log-normal.

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.

```
lognorm_teorica <- plnorm(0:300, meanlog=mean(log(monthly_max)), sdlog=sd(log(monthly_max)))

plot(ecdf(monthly_max), main="Comparación con la Distribución Log-Normal", xlab="Precipitaciones", col=

par(new=TRUE)
plot(0:300, lognorm_teorica, type="l", main="", xlab="", ylab="", col="red", lwd=2, ylim=c(0, 1.05))

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



##Explica qué son datos empíricos y datos teóricos. ¿Se parecen las distribuciones de probabilidad acumuladas? Los datos empíricos son los datos reales recolectados, en este caso son las precipitaciones máximas mensuales registradas en un periodo específico. Estos datos son los valores reales que ocurrieron en términos de eventos climáticos en la región. En este grafico los puntos azules y las líneas escalonadas representan la función de distribución empírica acumulada (ECDF).

Los datos teóricos son los valores que se esperarían si siguieran un modelo probabilístico en específico, que en este caso seria la distribución Log-Normal. La curva roja en esta gráfica representa la función de distribución acumulada teórica (CDF) para una distribución log-normal.

En general las distribuciones acumuladas empírica y teórica se alinean aceptablemente bien en el rango central de los datos, o sea entre los 150 y 250 mm. Sin embargo, la disconformidad en las colas indican que la distribución Log-Normal no captura de manera adecuada la frecuencia de los eventos más extremos o más bajos en los datos.

La distribución Log-Normal tiene un ajuste razonable para la mayoria de los datos, pero no un ajuste perfecto.

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

```
ks.test(monthly_max, "plnorm", meanlog=mean(log(monthly_max)), sdlog=sd(log(monthly_max)))
```

```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  monthly_max
## D = 0.1145, p-value = 0.7849
## alternative hypothesis: two-sided
```

B.4) ¿Qué información nos da la prueba KS para una Log-normal?

La prueba de Kolmogorov-Smirnov (KS) es una prueba de bondad de ajuste, se utuliza para confirmar si cierto conjunto de datos provien de una distribución Log-Normal en este caso o no. Compara la distribución empírica de los datos con una distribución teórica (en este caso, la distribución Log-Normal).

¿Cuál es el valor del estadístico de prueba?

El estadistico D de la prueba de Kolmogorov-Smirnov (KS) es 0.1145, valor que representa la mayor diferencia entre la función de distribución acumulada empírica y la teórica. (Entre más pequeño sea este valor más cercanas serán las dos distribuciones.)

¿Cuál es el p-value de la prueba?

El p-value de la prueba es 0.7849, este valor es la probabilida de obtener una diferencia mayor entre las distribuciones empírica y teórica si la hipótesis nula es cierta.

¿Se acepta o se rechaza la hipótesis nula?

Para este caso la hipotesis nula es que los datos provienen de una distribución Log-Normal, como el p-valor es mayor que 0.05 no rechazamos la hipotesis nula, lo cual se traduce a que no tenemos la evidencia suficiente para concluir que los datos no siguen una distribución log-normal.

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

Con la información y análisis realizado podemos concluir que los datos siguen de manera razonable una distribución Log-Normal dado que el p-valor es alto (0.7849) , indicando que no hay una diferencia significativa entre la distribución empírica y la teórica log-normal.

Como no se rechaza la hipótesis nula, nos sugiere que la distribución Log-Normal es adecuada para describir los datos de precipitación máxima mensual y no hay suficiente evidencia para rechazar esta hipótesis.

B.5) ¿Cuántos parámetros tiene la distribución Log-normal? ¿Cuáles son?

La distribución Log-Normal tiene dos parámetros fundamentales: *Media logarítmica(meanlog). Es la media de los logaritmos naturales de los datos observados.* Desviación estándar logarítmica (sdlog): Es la desviación estándar de los logaritmos naturales de los datos observados.

¿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.

La forma en que los parámetros de la distribución Log-Normal se calculan en el código es aplicando el logaritmo a los datos porque la distribución Log-Normal transforma los datos a una distribución normal en su escala logarítmica.

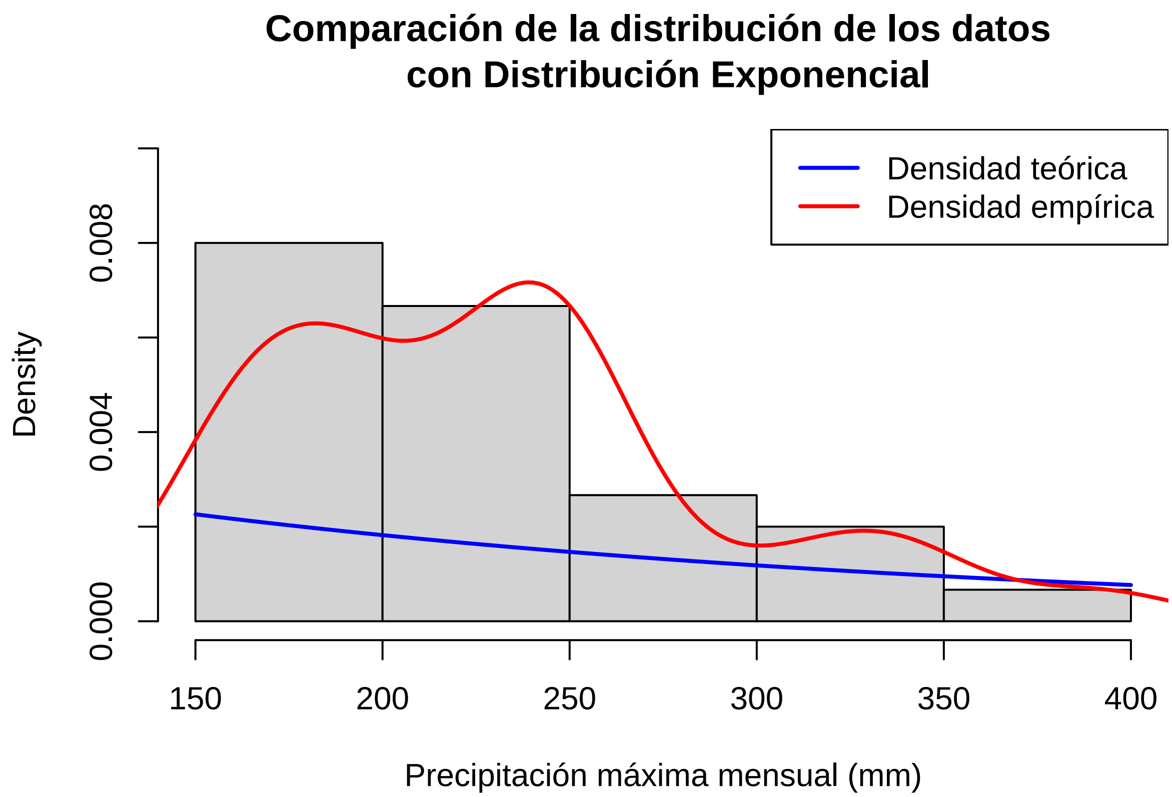
El método de momentos es utilizado para demostrar que los parámetros están bien calculados: *Se relaciona la media y la varianza de los datos originales con los parámetros logarítmicos.* Despejamos la media logaritmica y la desviación estandar logarítmica a partir de las ecuaciones de la media y varianza de la distribución log-normal. *Comparamos los parámetros obtenidos del método de momentos con los del código y confirmamos que coinciden. Esto nos asegura que hemos estimado los parámetros correctamente, usando la media la varianza de los datos transformados mediante logaritmicos.

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.

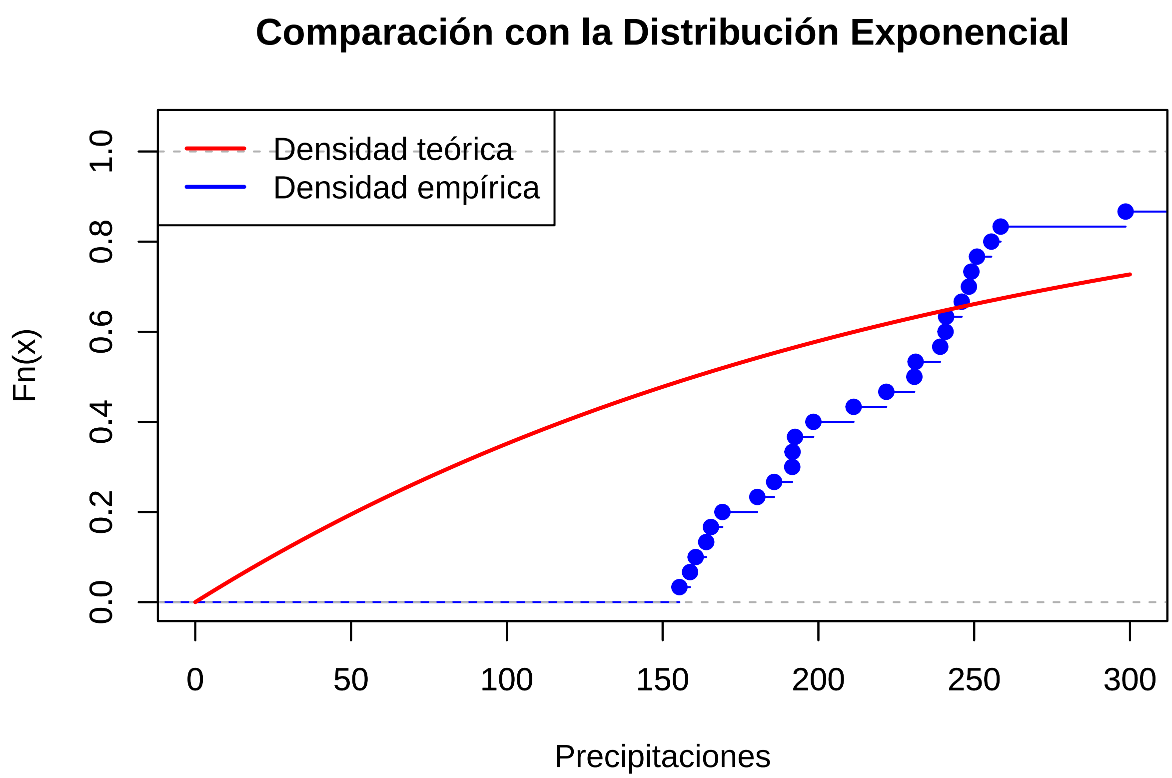
```
hist(monthly_max, xlab="Precipitación máxima mensual (mm)", freq=FALSE, ylim=c(0, 0.01),
     main="Comparación de la distribución de los datos \n con Distribución Exponencial")
curve(dexp(x, 1/mean(monthly_max)), add=TRUE, col="blue", lwd=2)
lines(density(monthly_max), col="red", lwd=2)
legend("topright", col=c("blue", "red"), legend=c("Densidad teórica", "Densidad empírica"), lwd=2)
```



De manera visual, ¿te parece que los datos se ajustan bien a una distribución Exponencial? Explica. De manera visual vemos que los datos no se ajustan bien a una distribución exponencial.En la cola derecha que son los valores de precipitación mayores se ve qeu los incrementos acumulativos no siguen un crecimiento exponenical suave, indicando que los eventos extremos son menos frecuentes de lo que se predice, abiertamente podemos deducir que la distribución exponenicial no pareciera ser un buen ajuste en este caso ya que los datos empíricos no se alinean al patrón teorico de la distribución exponencial.

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.

```
exp_teorica <- pexp(0:300, rate=1/mean(monthly_max))
plot(ecdf(monthly_max), main="Comparación con la Distribución Exponencial", xlab="Precipitaciones", col=
par(new=TRUE)
plot(0:300, exp_teorica, type="l", main="", xlab="", ylab="", col="red", lwd=2, ylim=c(0, 1.05))
legend("topleft", col=c("red", "blue"), legend=c("Densidad teórica", "Densidad empírica"), lwd=2)
```



Explica qué son datos empíricos y datos teóricos. ¿Se parecen las distribuciones de probabilidad acumuladas? Los datos empíricos son los datos reales recolectados, en este caso son las precipitaciones máximas mensuales registradas en un periodo específico. Estos datos son los valores reales que ocurrieron en términos de eventos climáticos en la región. En este grafico los puntos azules y las líneas escalonadas representan la función de distribución empírica acumulada (ECDF).

Los datos teóricos son los valores que se esperarían si siguieran un modelo probabilístico en específico, que en este caso seria la distribución exponencial. La curva roja en esta gráfica representa la función de distribución acumulada teórica (CDF) para una distribución exponencial.

Las distribuciones de probabilidad acumuladas no se parecen ni se alinenan ya que no captura adecuadamente el comportamiento de los datos empíricos, en especial lo vemos en la cola derecha.

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

```r
ks.test(monthly_max, "pexp", rate=1/mean(monthly_max))
```

```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  monthly_max
## D = 0.48979, p-value = 3.686e-07
## alternative hypothesis: two-sided
```

C.4) ¿Qué información nos da la prueba KS para una Exponencial?

La prueba de Kolmogorov-Smirnov (KS) se encarga de comparar la distribución acumulada empírica de los datos con la distribución acumulada teórica exponencial en este caso, y se evalua si los datos observados si se alinean con la distribución propuesta, la exponencial.

¿Cuál es el valor del estadístico de prueba?

El valor estadistico D de la prueba KS es 0.48979, que es la diferencia mayor entre función de distribución acumulada teórica y empírica.

¿Cuál es el p-value de la prueba?

El p-valor de la prueba es 3.686e-07, o sea 0.0000003686, el p-value nos indica la probabilidad de que haya una diferencia mayor entre las distribuciones si la hipótesis nula fuera cierta.

¿Se acepta o se rechaza la hipótesis nula?

Como hipotesis nula tenemos establecido que H0= los datos siguen una distribución exponencial. Y como el p-valor es muchisimo menor que 0.05, podemos rechazar la hipotesis nula.

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

no se puede concluir que que los datos que tenemos siguen una distribució exponencial, se puede argumentar con que el p-valor es muy bajo y con el valor estadistico D, ya que es muy alto indicandonos que hay una diferencia significativa entre los datos empíricos y la distribución exponencial teórica. Con eso se confirma que los datos de las precipitaciones máximas mensuales no siguen una distribución exponencial

C.5) ¿Cuántos parámetros tiene la distribución exponencial? ¿Cuáles son?

La distribución exponencial solamente tiene un parametro, tasa o rate , parametro que define frecuencia (o la rapidez) con la que suceden los eventos.

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

En el código el parametro de rate lo podemos estimar con el inverso de la media de los datos, o sea rate= 1/ media. ya que la media de los datos está inversamente relacionada con el parámetro de rate.

Sigue paso a paso el método de momentos para demostrar que los parámetros están bien calculados.

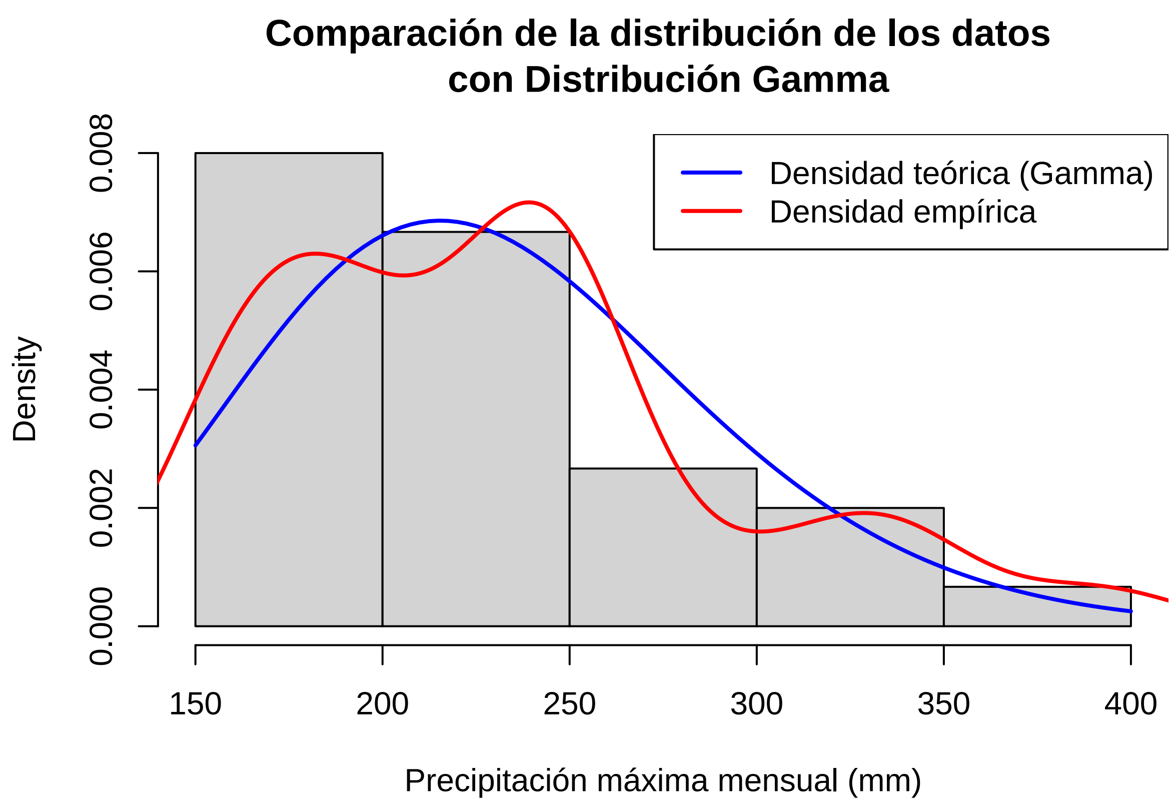
Media teórica de la distribución exponencial: definimos la media teórica con E(x)= 1/lambda. Se despeja lambda de la ecuación : lambda = 1/E(x) *calculamos la media de los datos observados (precipitaciones máximas mensuales) como un estimador de E(x)* la estimación de lambda= 1/media , en el código está estimación se realiza automáticamente al calcular lambda como el inverso de la media de los datos. *Al seguir los pasos en la actividad podemos confirmar que el calculo de lambda como el inverso de la media en tu actividad es una aplicación correcta del método de momentos, confirmando que se calculó bien el parámetro.

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.

```
media_observada <- mean(monthly_max)
varianza_observada <- var(monthly_max)
k <- (media_observada^2) / varianza_observada
theta <- varianza_observada / media_observada
hist(monthly_max, xlab="Precipitación máxima mensual (mm)", freq=FALSE,
     main="Comparación de la distribución de los datos \n con Distribución Gamma")
curve(dgamma(x, shape=k, scale=theta), add=TRUE, col="blue", lwd=2)
lines(density(monthly_max), col="red", lwd=2)
legend("topright", col=c("blue", "red"), legend=c("Densidad teórica (Gamma)", "Densidad empírica"), lwd=
```

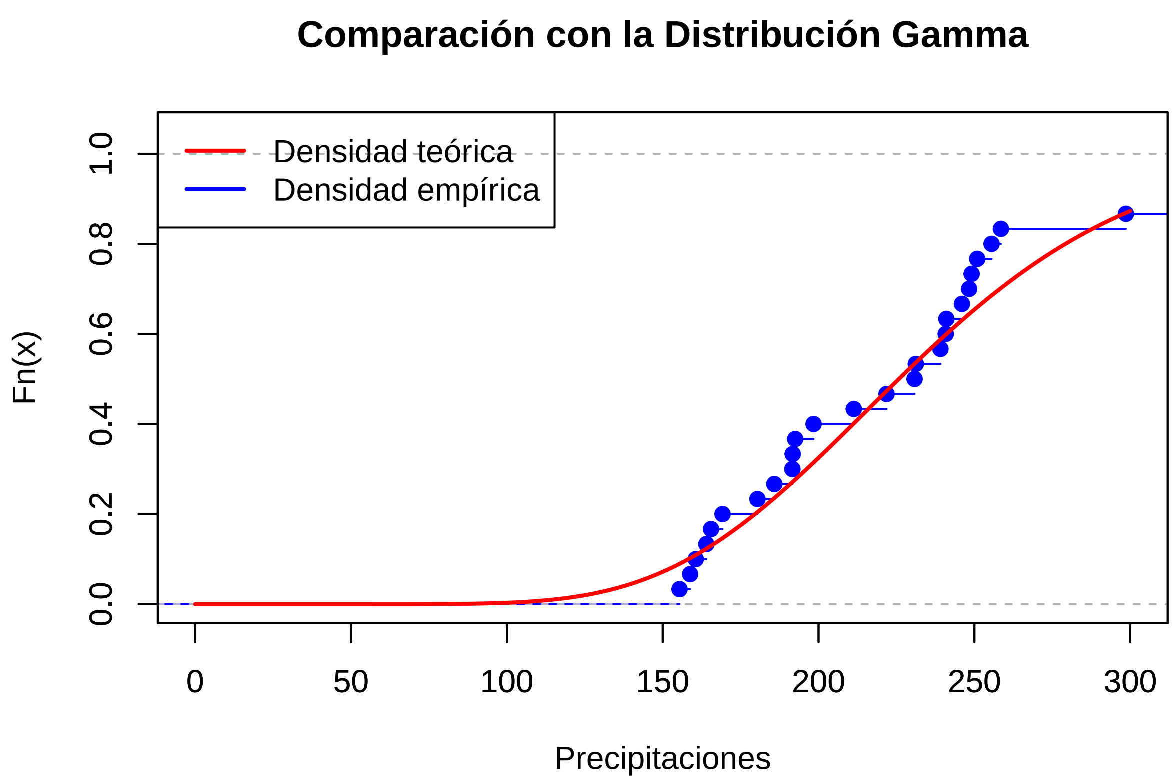


De manera visual, ¿te parece que los datos se ajustan bien a una distribución Gamma? Explica. De manera visual se puede decir que los datos se ajustan razonablemente bien a una distribución gamma, no perfectamente. En un rango de 150mm a 300mm ambas densidades son parecidas. En los valores bajos, o sea la cola izquierda la densidad empírica es un poco más alta que la teórica y en la cola derecha, los valores altos,apartir de los 300mm la curva teórica disminuye suavemente y la densidad empírica tiene una caída

más repentina.La distribución es un modelo de captura razonablemente bien la mayor parte de los datos de precipitación máxima mensual.

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.

```
gamma_teorica <- pgamma(0:300, shape=k, scale=theta)
plot(ecdf(monthly_max), main="Comparación con la Distribución Gamma", xlab="Precipitaciones", col="blue
par(new=TRUE)
plot(0:300, gamma_teorica, type="l", main="", xlab="", ylab="", col="red", lwd=2, ylim=c(0, 1.05))
legend("topleft", col=c("red", "blue"), legend=c("Densidad teórica", "Densidad empírica"), lwd=2)
```



Explica qué son datos empíricos y datos teóricos. ¿Se parecen las distribuciones de probabilidad acumuladas? Los datos empíricos son los datos reales recolectados, en este caso son las precipitaciones máximas mensuales registradas en un periodo específico. Estos datos son los valores reales que ocurrieron en términos de eventos climáticos en la región. En este grafico los puntos azules y las líneas escalonadas representan la función de distribución empírica acumulada (ECDF).

Los datos teóricos son los valores que se esperarían si siguieran un modelo probabilístico en específico, que en este caso seria la distribución gamma. La curva roja en esta gráfica representa la función de distribución acumulada teórica (CDF) para una distribución gamma.

Tienen similitudes ya que las distribuciones acumuladas empírica y teórica se alinean bastante bien. En la mayoria de los rangos la curva teórica y la curva empírica se mantienen cerca, lo cual es bueno.

Como diferencias podemos destacar que en las colas hay diferencias, o sea que la distribución gamma no capta a la perfección los eventos de muy baja precipitación o los extremos.

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

```
ks_test_result <- ks.test(monthly_max, "pgamma", shape=k, scale=theta)
ks_test_result
```

```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  monthly_max
## D = 0.13171, p-value = 0.6282
## alternative hypothesis: two-sided
```

D.4) ¿Qué información nos da la prueba KS para una distribución Gamma?

La prueba de Kolmogorov-Smirnov (KS) se encarga de comparar la distribución acumulada empírica de los datos con la distribución acumulada teórica gamma en este caso, y se evalua si los datos observados si se alinean con la distribución propuesta, la exponencial.

D.5) ¿Cuál es el valor del estadístico de prueba?

El valor estadistico D de la prueba KS es 0.13171, que es la diferencia mayor entre función de distribución acumulada teórica y empírica.

¿Cuál es el p-value de la prueba?

El p-valor de la prueba es 0.6282, el p-value nos indica la probabilidad de que haya una diferencia mayor entre las distribuciones si la hipótesis nula fuera cierta.

¿Se acepta o se rechaza la hipótesis nula?

Como hipotesis nula tenemos establecido que H0= los datos siguen una distribución gamma. Y como el p-valor es mayor que 0.05 , no rechazamos la hipotesis nula.

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

Podemos decir que si, los datos se ajustan a la distribución gamma de manera razonable, ya que el p-valor al ser alto nos dice que no existe una diferencia significativa entre los datos empíricos y la distribución teórica gamma, justificando que la Distribución Gamma es un buen ajuste para describir las precipitaciones máximas mensuales.

D.6) ¿Cuántos parámetros tiene la distribución Gamma? ¿Cuáles son?

La distrubución gamma cuenta con dos parametros: *Forma(k), que define la forma de la distribución y afecta la simetria.* Escala(theta), que define la escala (tamaño) medio de los datos en relación con la forma.

##¿Por qué los parámetros se calculan de la forma en cómo se hace en el código? En el codigo los parámetros se calculan usando el método de momentos porque este método relaciona los momentos (media y varianza) de los datos con los parámetros teóricos de la distribución. * Parámetro de forma : k <- (media_observada^2) / varianza_observada *Parámetro de escala : theta <- varianza_observada / media_observada

##Sigue paso a paso el método de momentos para demostrar que los parámetros están bien calculados. *Media y varianza teóricas de la Distribución Gamma: -Media E(x) = ktheta -varianza Var(x) = k(theta)^2

Se despejan los parámetros usando la media y la varianza Se calcula K y tehta con los datos observados *Y al realizar el procedimiento necesario se demuestra que nuestros parámetros k y theta se calcularon de manera correcta usando el método de momentos, basandonos en los momentos (media y varianza) observados en los datos.

E) Ajuste a una Distribución Weibull.

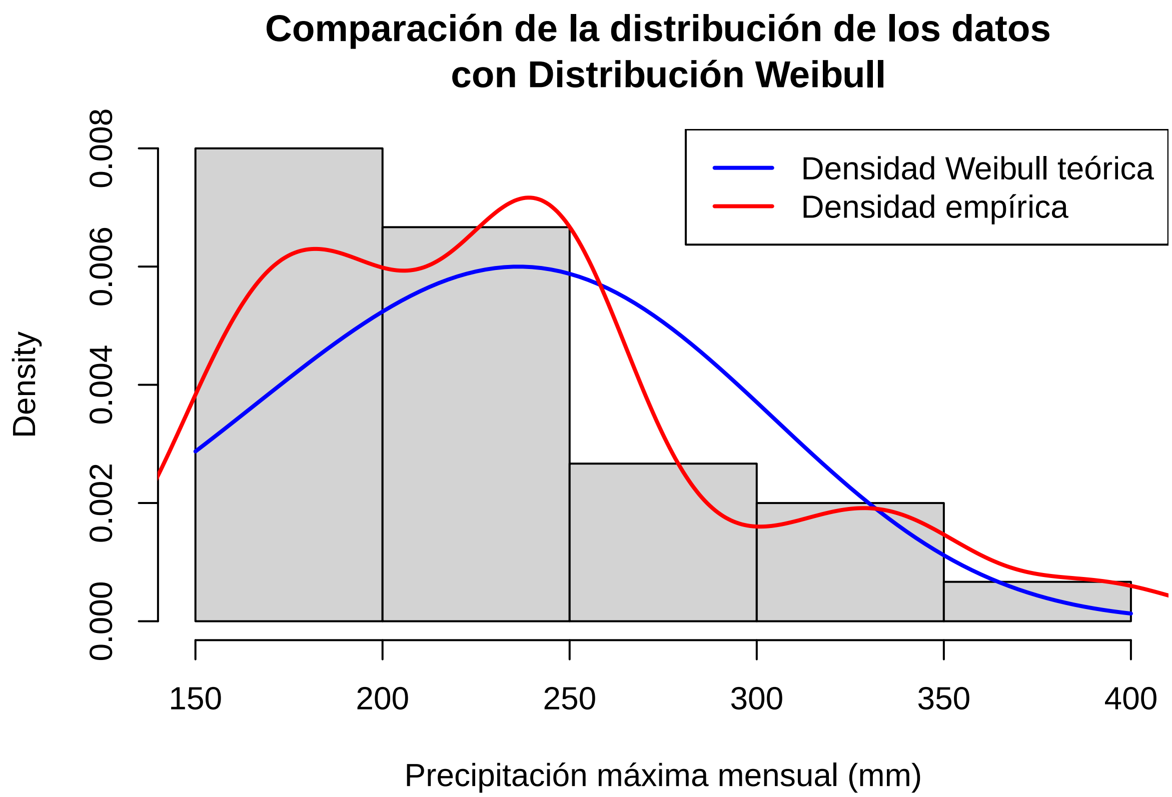
Realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Weibull para ajustar los datos.

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.

```
library(MASS)
weibull_fit <- fitdistr(monthly_max, "weibull", lower=c(0,0))
```

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.

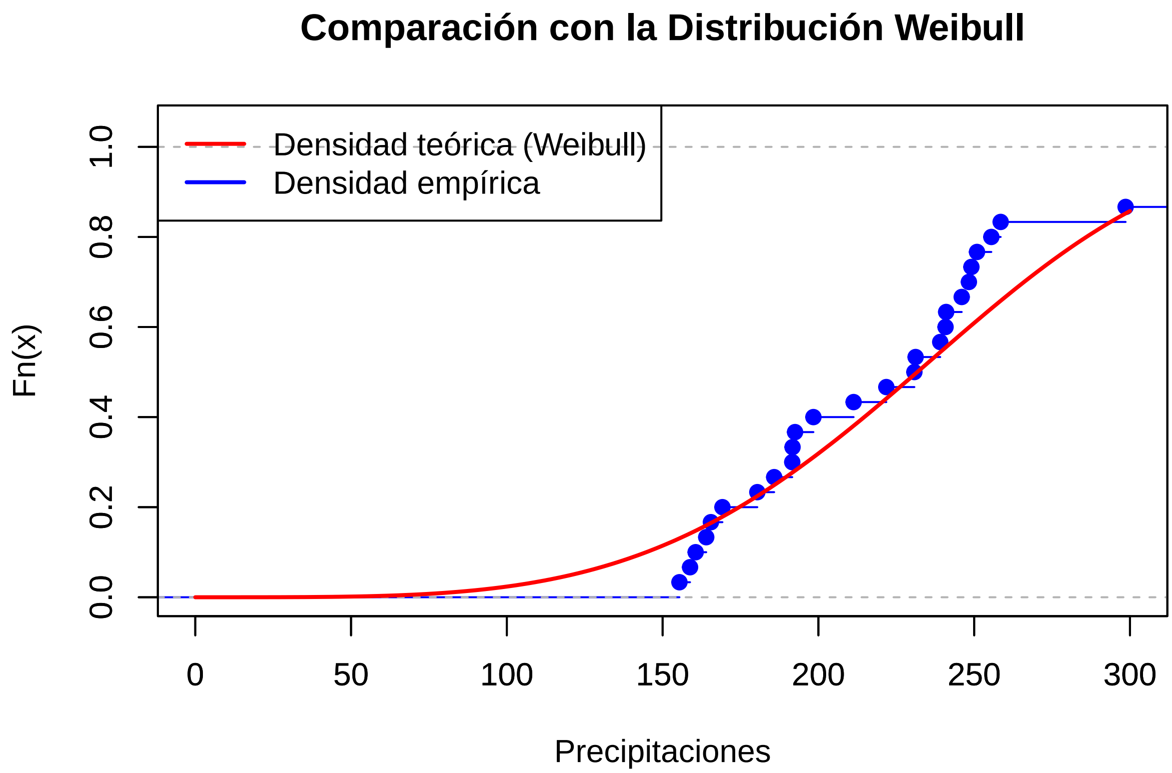
```
hist(monthly_max, xlab="Precipitación máxima mensual (mm)", freq=FALSE,
     main="Comparación de la distribución de los datos \n con Distribución Weibull")
curve(dweibull(x, shape=weibull_fit$estimate[1], scale=weibull_fit$estimate[2]),
      add=TRUE, col="blue", lwd=2)
lines(density(monthly_max), col="red", lwd=2)
legend("topright", col=c("blue", "red"), legend=c("Densidad Weibull teórica", "Densidad empírica"), lwd=
```



De manera visual, ¿te parece que los datos se ajustan bien a una distribución Weibull? Explica. De manera visual podemos concluir que los datos no se ajsutan bien a una distribución Weibull, en el rango de entre 150mm y 250mm las densidades no coinciden y la densidad empirica muestra muchas fluctuaciones, y en la cola derecha, apartir de los 300mm las curvas se alinean un poco. Pero en general pareciera que los datos no se ajustan a la distribución weibull.

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.

```
weibull_teorica <- pweibull(0:300, shape=weibull_fit$estimate[1], scale=weibull_fit$estimate[2])
plot(ecdf(monthly_max), main="Comparación con la Distribución Weibull", xlab="Precipitaciones", col="blu
par(new=TRUE)
plot(0:300, weibull_teorica, type="l", main="", xlab="", ylab="", col="red", lwd=2, ylim=c(0, 1.05))
legend("topleft", col=c("red", "blue"), legend=c("Densidad teórica (Weibull)", "Densidad empírica"), lw
```



Explica qué son datos empíricos y datos teóricos. ¿Se parecen las distribuciones de probabilidad acumuladas? Los datos empíricos son los datos reales recolectados, en este caso son las precipitaciones máximas mensuales registradas en un periodo específico. Estos datos son los valores reales que ocurrieron en términos de eventos climáticos en la región. En este grafico los puntos azules y las líneas escalonadas representan la función de distribución empírica acumulada (ECDF).

Los datos teóricos son los valores que se esperarían si siguieran un modelo probabilístico en específico, que en este caso seria la distribución Weibull. La curva roja en esta gráfica representa la función de distribución acumulada teórica (CDF) para una distribución Weibull.

En general, la curva teórica (roja) y la curva empírica (azul) se alinean razonablemente bien, lo que sugiere que la distribución Weibull captura razonablemente bien la tendencia acumulativa de los datos, en especial en el rango entre 150 y 250mm.

Sin embargo en los valores muy bajos y extremos, o sea las colas de la distribución, vemos algunas diferencias entre las curvas.

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

```
ks_test_result <- ks.test(monthly_max, "pweibull", shape=weibull_fit$estimate[1], scale=weibull_fit$esti
ks_test_result
```

```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  monthly_max
```

```
## D = 0.17511, p-value = 0.2821
## alternative hypothesis: two-sided
```

E.5) ¿Qué información nos da la prueba KS para una Weibull?

La prueba de Kolmogorov-Smirnov (KS) se encarga de comparar la distribución acumulada empírica de los datos con la distribución acumulada teórica weibull en este caso, y se evalua si los datos observados si se alinean con la distribución propuesta, la weibull.

¿Cuál es el valor del estadístico de prueba?

El valor estadistico D de la prueba KS es 0.17511, que es la diferencia mayor entre función de distribución acumulada teórica y empírica.

¿Cuál es el p-value de la prueba?

El p-valor de la prueba es 0.2821, el p-value nos indica la probabilidad de que haya una diferencia mayor entre las distribuciones si la hipótesis nula fuera cierta.

¿Se acepta o se rechaza la hipótesis nula?

Como hipotesis nula tenemos establecido que H0= los datos siguen una distribución weibull. Y como el p-valor es mayor que 0.05 , no rechazamos la hipotesis nula.

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

De acuerdo a las pruebas de bondad podemos decir que si ya que el p-valor alto nos sugiere que no hay una diferencia significativa entre los datos empíricos y la distribución teórica Weibull, sin embargo en la primera grafica no vemos que haya un buen ajuste.

E.6) ¿Cuántos parámetros tiene la distribución Weibull? ¿Cuáles son?

La distribución con la que estamos trabajando, la weibull tiene dos parametros: *forma(shape), que controla la forma de la distribución.* escala(scale), que afectala escala de los datos

##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. En diferencia con las distribuciones anteriores, el ajuste de los parámetros de weibull es más complejo, por lo que se opta por usar el método de máxima verosimilitud usando fitdistr con el fin de obtener una estimación mucho mejro y más precisa, para análisar de manera visual y estadisticamente el ajuste de la distribución Weibull a los datos de precipitación máxima mensual.

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.

```
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))
```

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.

```
if (!requireNamespace("fitdistrplus", quietly = TRUE)) install.packages("fitdistrplus")
library(fitdistrplus)
```

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

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

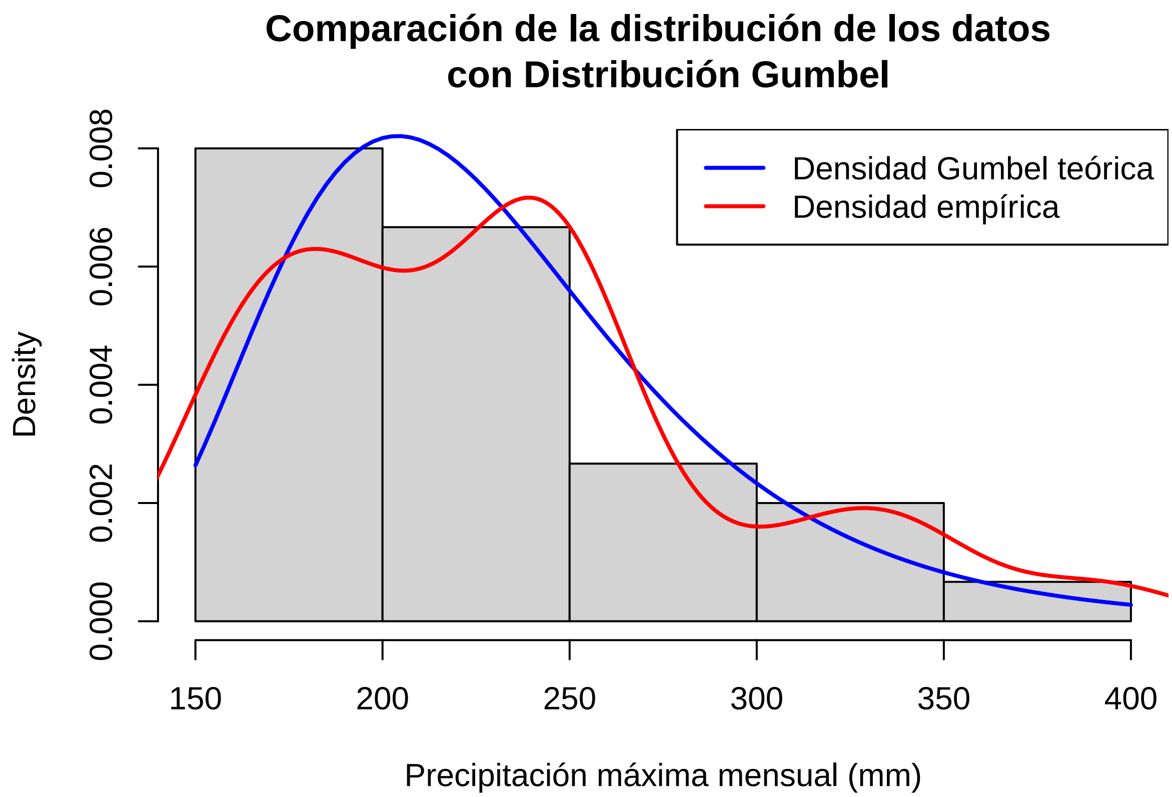
```
## Fitting of the distribution ' gumbel ' by maximum likelihood
## Parameters:
##    estimate Std. Error
## a 204.07290   8.604255
## b  44.81446   6.609607
```

En esta tabla vemos las estimaciones de los parametros de la distribución Gumbel:

*El parámetro de ubicación (a): 204.07290 con un error estándar de 8.604255

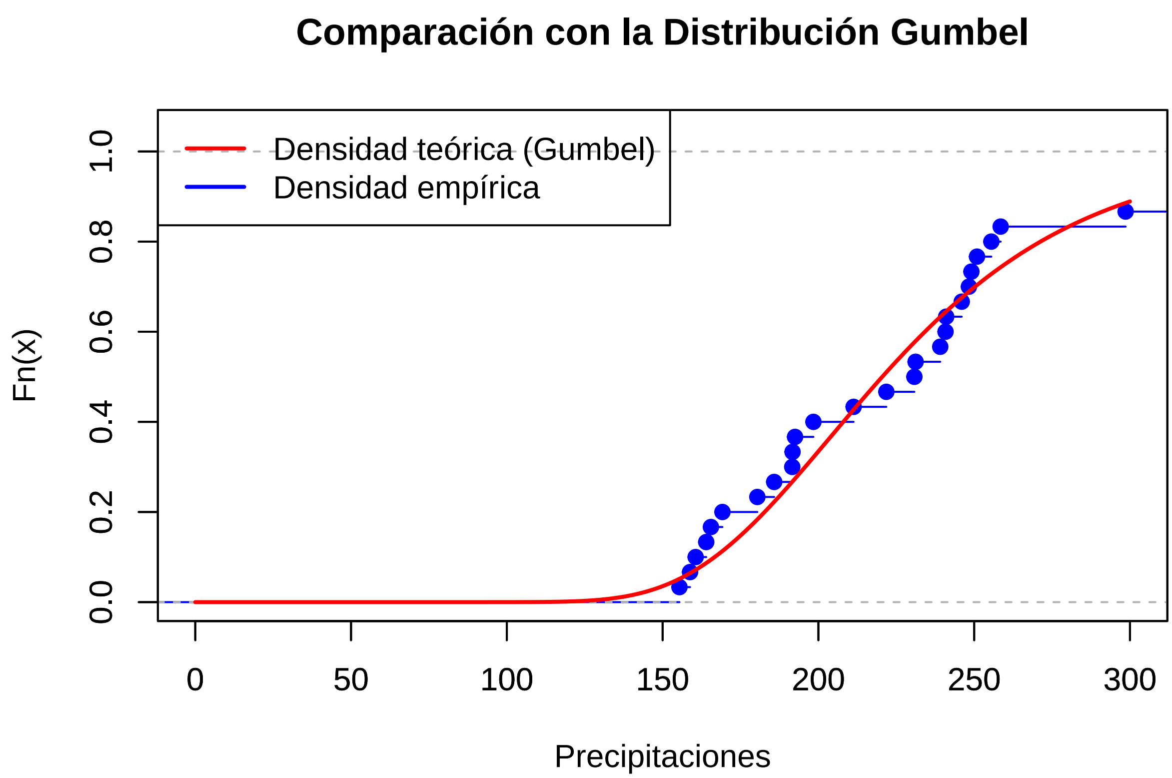
*El parámetro de escala (b): 44.81446 con un error de estimación de 6.609607

```
hist(monthly_max, xlab="Precipitación máxima mensual (mm)", freq=FALSE,
     main="Comparación de la distribución de los datos \n con Distribución Gumbel")
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 empírica"), lwd=
```



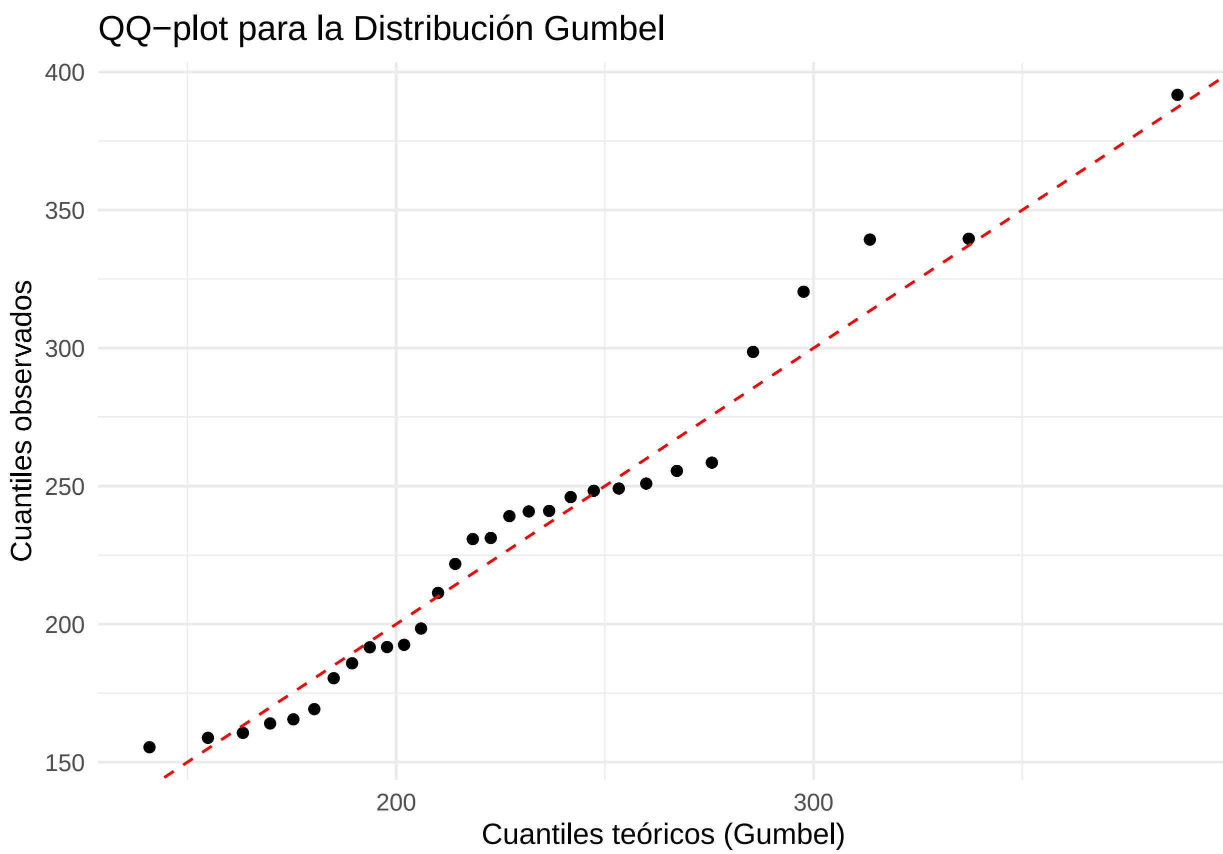
Como análisis visual vemos que en el rango de 150 a 300mm la curva teórica se alinea relativamente bien con la densidad empírica. Y en la cola derecha, o sea los eventos extremos, apartir de los 300mm los datos difieren un poco.

```
gumbel_acum_teorica <- pgumbel(0:300, gumbel_fit$estimate[1], gumbel_fit$estimate[2])
plot(ecdf(monthly_max), main="Comparación con la Distribución Gumbel", xlab="Precipitaciones", col="blu
par(new=TRUE)
plot(0:300, gumbel_acum_teorica, type="l", main="", xlab="", ylab="", col="red", lwd=2, ylim=c(0, 1.05)
legend("topleft", col=c("red", "blue"), legend=c("Densidad teórica (Gumbel)", "Densidad empírica"), lwd=
```



Vemos que el ajuste en la función de distribución acumulada (CDF) es bueno, es decir vemos que la curva teórica de la distribución Gumbel se ajusta bastante bien a la curva empírica en la mayoria de los rangos. Sin embargo en las colas es donde vemos unas ligeras desviaciones.

```
if (!requireNamespace("fitdistrplus", quietly = TRUE)) install.packages("fitdistrplus")
library(fitdistrplus)
gumbel_fit <- fitdist(monthly_max, "gumbel", start = list(a = 1, b = 1))
library(ggplot2)

gumbel_quantiles <- function(p, a, b) a - b * log(-log(p))

observed_quantiles <- sort(monthly_max)
theoretical_quantiles <- gumbel_quantiles(ppoints(length(observed_quantiles)),
                                          gumbel_fit$estimate[1],
                                          gumbel_fit$estimate[2])
ggplot(data.frame(theoretical = theoretical_quantiles, observed = observed_quantiles),
       aes(sample = observed)) +
  stat_qq(distribution = function(p) gumbel_quantiles(p, gumbel_fit$estimate[1], gumbel_fit$estimate[2]
  geom_abline(intercept = 0, slope = 1, color = "red", linetype = "dashed") +
  labs(title = "QQ-plot para la Distribución Gumbel",
       x = "Cuantiles teóricos (Gumbel)",
       y = "Cuantiles observados") +
  theme_minimal()
```



Vemos que la gran mayoria de los puntos parecen alinearse a la linea de referencia, indicando que los datos se ajustan bien a los cuantiles teóricos de la distribución gumberl, sin embargo en ambos extremos los puntos se desvian. Con el qq-plor vemos que la distribución gumbel nos proporciona un ajuste razonable con algunas desviaciones en los extremos.

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

```r
gumbel_excede <- 1 - pgumbel(monthly_max, gumbel_fit$estimate[1], gumbel_fit$estimate[2])
ks_test_gumbel <- ks.test(monthly_max, "pgumbel", a=gumbel_fit$estimate[1], b=gumbel_fit$estimate[2])
ks_test_gumbel
```

```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  monthly_max
## D = 0.10983, p-value = 0.8241
## alternative hypothesis: two-sided
```

F.4) ¿Qué información nos da la prueba KS para una Gumbel?

La prueba de Kolmogorov-Smirnov (KS) se encarga de comparar la distribución acumulada empírica de los datos con la distribución acumulada teórica Gumbel en este caso, y se evalua si los datos observados si se alinean con la distribución propuesta, la Gumbel.

¿Cuál es el valor del estadístico de prueba?

El valor estadistico D de la prueba KS es 0.10983, que es la diferencia mayor entre función de distribución acumulada teórica y empírica.

¿Cuál es el p-value de la prueba?

El p-valor de la prueba es 0.8241, el p-value nos indica la probabilidad de que haya una diferencia mayor entre las distribuciones si la hipótesis nula fuera cierta.

¿Se acepta o se rechaza la hipótesis nula?

Como hipotesis nula tenemos establecido que H0= los datos siguen una distribución Gumbel. Y como el p-valor es mayor que 0.05 , no rechazamos la hipotesis nula.

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

Podemos concluir que los datos siguen razonablemente una distribución gumbel ya que el p-valor alto indica que no hay una diferencia significativa entre la distribución empírica y la teórica gumbel, con lo que se justifica que la distribución gumbel es un buen ajuste.

F.5) ¿Cuántos parámetros tiene la distribución Gumbel? ¿Cuáles son?

La distribución de Gumbel tiene dos parámetros: *ubicación (a), que define el valor central de la distribución* escala (b), que define la dispersión de los datos alrededor del valor de ubicación.

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?

```r
if (!requireNamespace("fitdistrplus", quietly = TRUE)) install.packages("fitdistrplus")
library(fitdistrplus)

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

mu <- mean(monthly_max)
sigma <- sd(monthly_max)
gamma <- 0.5772

moment_b <- sigma * sqrt(6) / pi
moment_a <- mu - moment_b * gamma

cat("Parámetros estimados con fitdistrplus (máxima verosimilitud):\n")
```

```
## Parámetros estimados con fitdistrplus (máxima verosimilitud):
```

```
cat("a =", fit_a, "\n")
```

```
## a = 204.0729
```

```
cat("b =", fit_b, "\n\n")
```

```
## b = 44.81446
```

```
cat("Parámetros estimados con el método de momentos:\n")
```

```
## Parámetros estimados con el método de momentos:
```

```
cat("a =", moment_a, "\n")
```

```
## a = 203.9791
```

```
cat("b =", moment_b, "\n")
```

```
## b = 46.68678
```

No se obtienen exactamente los mismos valores por los 2 metodos, los resultados son ligeramnete distintos.

¿por qué crees que se dé esta diferencia?

fitdistrplus utiliza la máxima verosimilitud, que se encarga de encontrar los valores de los parámetros que maximizan la probabilidad de los datos observados. El método de momentos se basa en igualar la media y la varianza teóricas de la distribución con las de los datos observados.

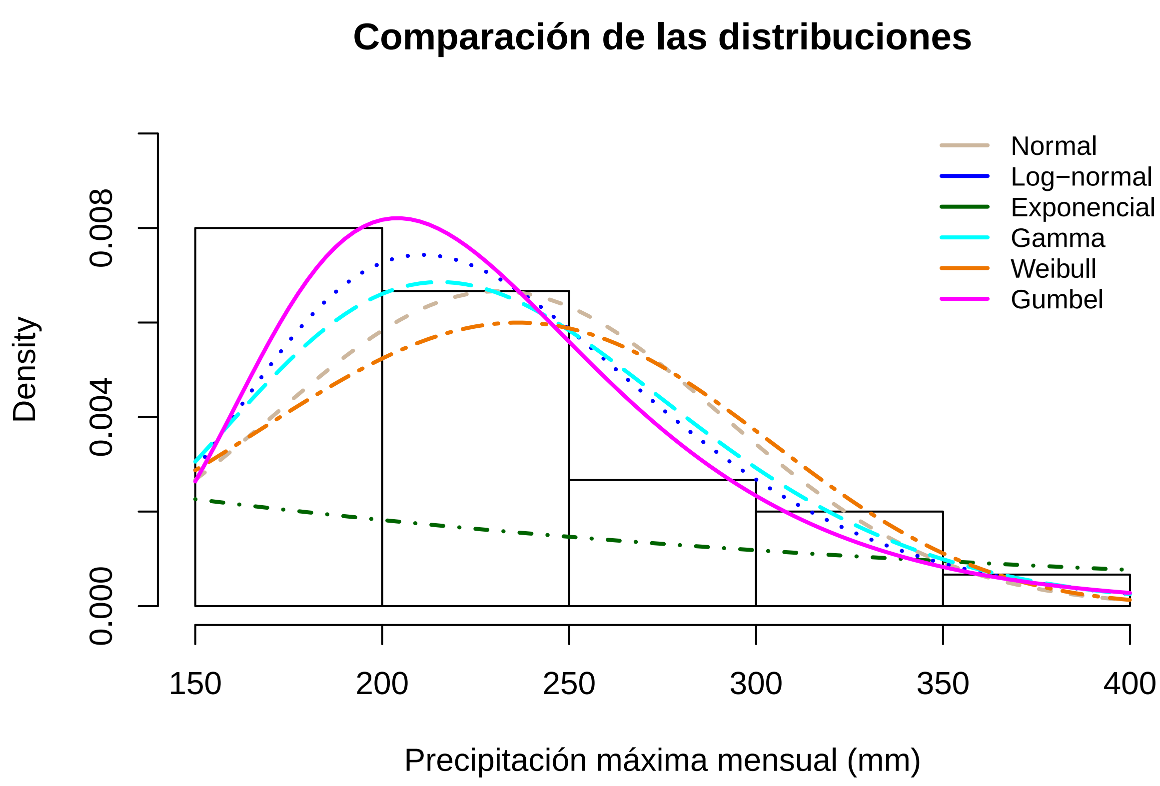
En general las diferencias se deben a que tienen diferente objetivos de ajuste, o sea la maxima verosimilutud trata de capturar la forma completa de la distribución y por otro lado el método de momwntos solamente se ajusta a la media y a la variazna.

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.

```
hist(monthly_max, xlab="Precipitación máxima mensual (mm)", freq=FALSE,
     ylim=c(0, 0.01), main="Comparación de las distribuciones", col=0)

curve(dnorm(x, mean=mean(monthly_max), sd=sd(monthly_max)), add=TRUE, col="bisque3", lwd=2, lty=2)
curve(dlnorm(x, meanlog=mean(log(monthly_max)), sdlog=sd(log(monthly_max))), add=TRUE, col="blue", lwd=
curve(dexp(x, 1/mean(monthly_max)), add=TRUE, col="darkgreen", lwd=2, lty=4)
```

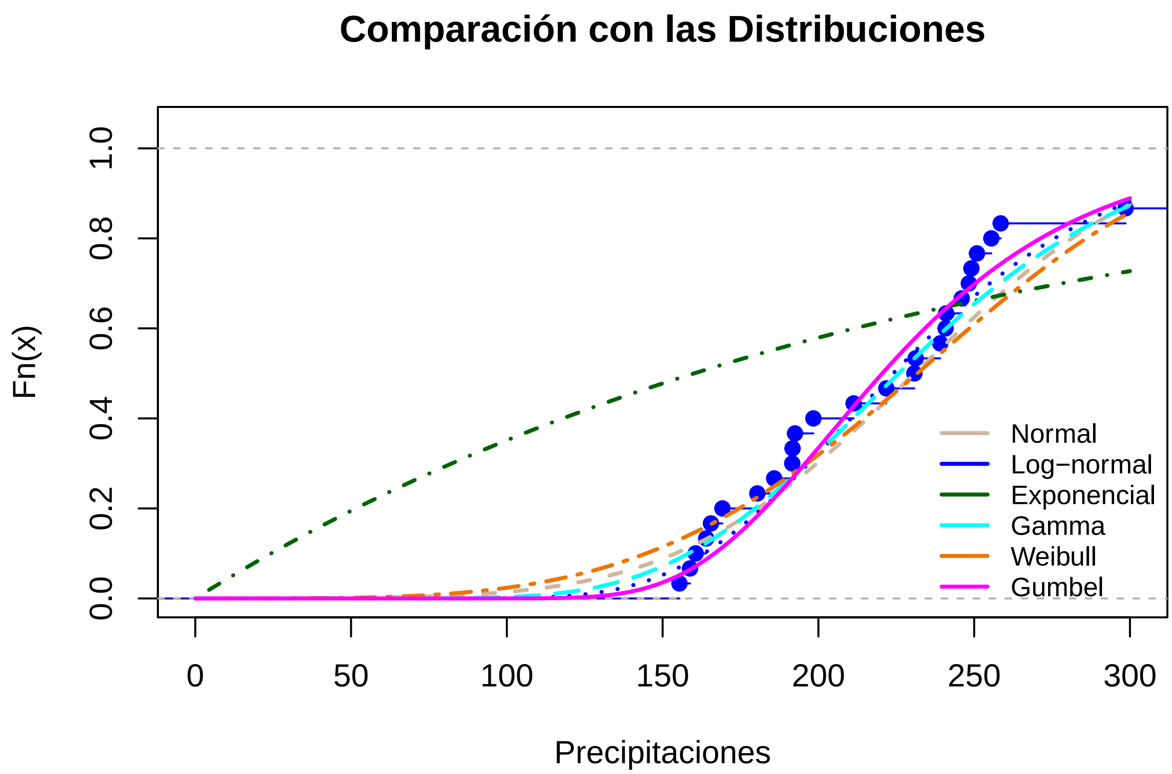
```r
curve(dgamma(x, shape=(mean(monthly_max)^2/var(monthly_max)), rate=mean(monthly_max)/var(monthly_max)),
      add=TRUE, col="cyan", lwd=2, lty=5)
curve(dweibull(x, shape=weibull_fit$estimate[1], scale=weibull_fit$estimate[2]),
      add=TRUE, col="darkorange2", lwd=2, lty=6)
curve(dgumbel(x, gumbel_fit$estimate[1], gumbel_fit$estimate[2]),
      add=TRUE, col="magenta", lwd=2, lty=7)
legend("topright", legend=c("Normal", "Log-normal", "Exponencial", "Gamma", "Weibull", "Gumbel"),
       col=c("bisque3", "blue", "darkgreen", "cyan", "darkorange2", "magenta"),
       lwd=2, bty="n", cex=0.8)
```



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

```r
plot(ecdf(monthly_max), main="Comparación con las Distribuciones", xlab="Precipitaciones", col="blue",
     xlim=c(0,300), ylim=c(0, 1.05))

lines(seq(0, 300, length=100), pnorm(seq(0, 300, length=100), mean=mean(monthly_max), sd=sd(monthly_max
lines(seq(0, 300, length=100), plnorm(seq(0, 300, length=100), meanlog=mean(log(monthly_max)), sdlog=sd
lines(seq(0, 300, length=100), pexp(seq(0, 300, length=100), rate=1/mean(monthly_max)), col="darkgreen"
lines(seq(0, 300, length=100), pgamma(seq(0, 300, length=100), shape=(mean(monthly_max)^2/var(monthly_ma
lines(seq(0, 300, length=100), pweibull(seq(0, 300, length=100), shape=weibull_fit$estimate[1], scale=we
lines(seq(0, 300, length=100), pgumbel(seq(0, 300, length=100), gumbel_fit$estimate[1], gumbel_fit$esti
```

```r
legend("bottomright", legend=c("Normal", "Log-normal", "Exponencial", "Gamma", "Weibull", "Gumbel"),
       col=c("bisque3", "blue", "darkgreen", "cyan", "darkorange2", "magenta"), lwd=2, bty="n", cex=0.8
```



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.

Para lograr definir cual es la mejor distribución que se ajusta a los datos se analizarán el histograma de densidad como la probabilidad acumulada de cada distribución y tambien las pruebas de bondad de ajuste.

Análisis del histograma de densidad:

*Las distribucines: gumbel y weibull se acercan mucho a la forma de la densidad empírica, en especial en el rango de los datos del centro.

*Las distribuciones log-normla y gamma se acercan a la densidad empírica, sin embargo es más obvio que presentan algunas desviaciones en las colas.

*La distribución exponencial es la que menos se alinea y la que menos se ajusta debido a su forma única de decaimiento, y por eso vemos una diferencia tan marcada en comparación de las otras distribuciones.

Análisis de la probabilidad acumulada:

*Nuevamente las distribuciones gumbell y weibull estan muy cercanas a alinearsea con la distribución empírica en muchos rangos de valores, lo cual es bueno porque nos indica un buen ajuste en la probabilidad acumulada.

*Las distribuciones log-normal y la gamma nos presentan un buen ajuste, es decir aceptable, pero no tan bueno como el de las 2 distribuciones pasadas.

*Finalmente la distribución exponencial vemos que no se ajusta de buena manera , lo que sugiere que no captura de manera correcta los datos.

Pruebas de bondad de ajuste

```r
resultados_ks <- data.frame(
  Distribucion = c("Normal", "Log-normal", "Exponencial", "Gamma", "Weibull", "Gumbel"),
  D = c(0.1559, 0.1145, 0.4897, 0.1317, 0.1751, 0.1098),
  p_value = c(0.4167, 0.7849, 0.00000003686, 0.6282, 0.2821, 0.8241)
)
resultados_ks
```

```
##   Distribucion      D   p_value
## 1       Normal 0.1559 4.167e-01
## 2   Log-normal 0.1145 7.849e-01
## 3  Exponencial 0.4897 3.686e-08
## 4        Gamma 0.1317 6.282e-01
## 5      Weibull 0.1751 2.821e-01
## 6       Gumbel 0.1098 8.241e-01
```

Para determinar la mejor distribución bajo el criterio de la prueba Kolmogorov-Smirnov (KS). SE tienen en cuenta dos puntos clave:

*Menor valor de D, se busca tener el valor más chico de D, ya que cuanto menor sea el valor de D, mejor será el ajuste de la distribución teórica a los datos observados.

*Mayor p-valor, se busca tener un p-valor alto, ya que eso indica que no hay suficiente evidencia para rechazar la hipótesis de que los datos siguen esa distribución (con la que se este trabajando), lo que sugiere un buen ajuste.

Interpretación:

La distribución exponencial tiene el mayor valor D (0.4897) y un p-valor muy bajo (0.00000003686), dando a entender que la distribución exponencial tiene el peor ajuste, se rechaza la hipótesis de que los datos siguen esta distribución, por lo que queda automaticamente descartada esta distribución.

Las distribcuiones normal, log-normal, gamma y weibull tienen resultados aceptables en el D valor sin embargo ninguno es tan satisfactorio ni tan bajo como el de la distribución gumbel, y en cuanto al p-valor tampoco son perfectos ni tan buenos como el p-avlor de la distribución gumbel, asi que por descarte las eliminamos a todas las anteriores y nos quedamos con la distribución gumbel.

La distribución gumbel tiene el valor D más bajo de todas las distribuciones (0.1098) y tiene el p-valor más alto de todas las distribuciones (0.8241), indicandonos que esta es la distribución con mejor ajuste , el valor bajo de D y el p-valor alto indican que no hay suficiente evidencia para rechazar la hipótesis de que los datos siguen esta distribución, por lo que es la distribución que mejor satisface y la que mejor ajuste tiene, por lo que se puede decir que la distribución gumbel es la mejor opción en este caso.

Tomando en cuenta los tres análisis realizados: densidad, probabilidad acumulada y las pruebas de bondad de ajuste, se puede afirmar que la distribución gumbel pareciera ser la mejor opción para modelar los datos de precipitación máxima mensual. En ambos graficos realiza un ajsute cercano en la densidad y probabilidad acumulada y en cuento a la prueba de Kolmogorov-Smirnov (KS) se acepta con un p-valor alto. La distribución weibull tambien se toma como un buen ajuste, sin embargo la distribución gumbel es mejor y representa mejor la naturaleza asimétrica.

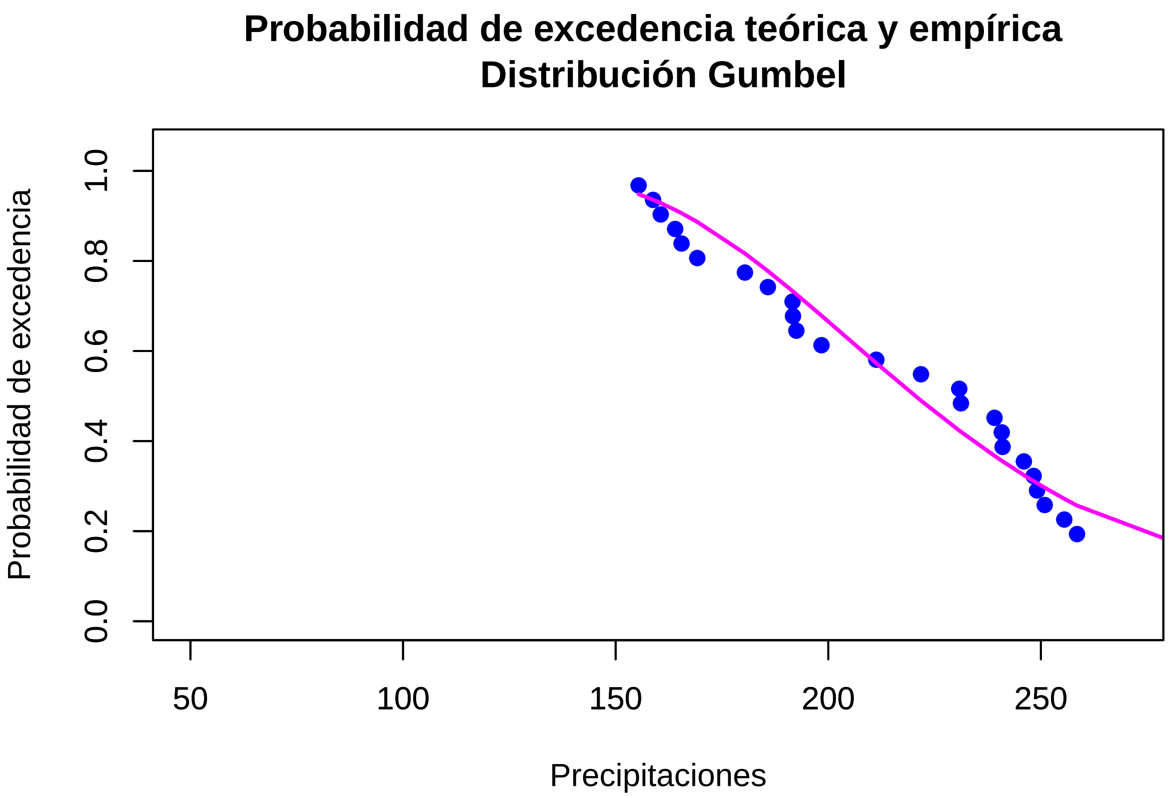
DISEÑO DE OBRAS HIDRÁULICAS

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.Links to an external site.

A) Haz el gráfico comparativo de la probabilidad de excedencia teórica vs empírica.

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

par(new=TRUE)
gumbel_exe <- 1 - pgumbel(rain_analysis$order_max_rain,
                          gumbel_fit$estimate[1],
                          gumbel_fit$estimate[2])
plot(rain_analysis$order_max_rain, gumbel_exe, type="l", col="magenta",
     lwd=2, xlim=c(50, 270), ylim=c(0, 1.05), xlab="", ylab="")
```



¿Qué te indica ese gráfico? interpreta y argumenta la certeza de la selección de la distribución elegida.

En este gráfico vemos la comparación entre la probabilidad de excedencia empírica (puntos en azul) y la probabilidad de excedencia teórica de la distribución gumbel (línea magenta). Vemos que la distribución gumbel logra ajuste muy bueno a los datos empíricos en la gran mayoria de los rangos de precipitación, con ligeras desviaciones en algunos rangos.

Se observa que la cercania que hay entre los puntos azules y la línea magenta es indicio de que la distribución Gumbel es una buena aproximación para modelar la probabilidad de excedencia de las precipitaciones para este ejercicio.

Dado que la distribución si sigue el comportamiento de los datos, podemos afirmar que la distribución Gumbel fue una buena elección, ya que se ajusta bien en especial en el rango de valores medios donde la probabilidad de excedencia es más representativa para periodos de retorno típicos en el diseño de obras hidráulicas.

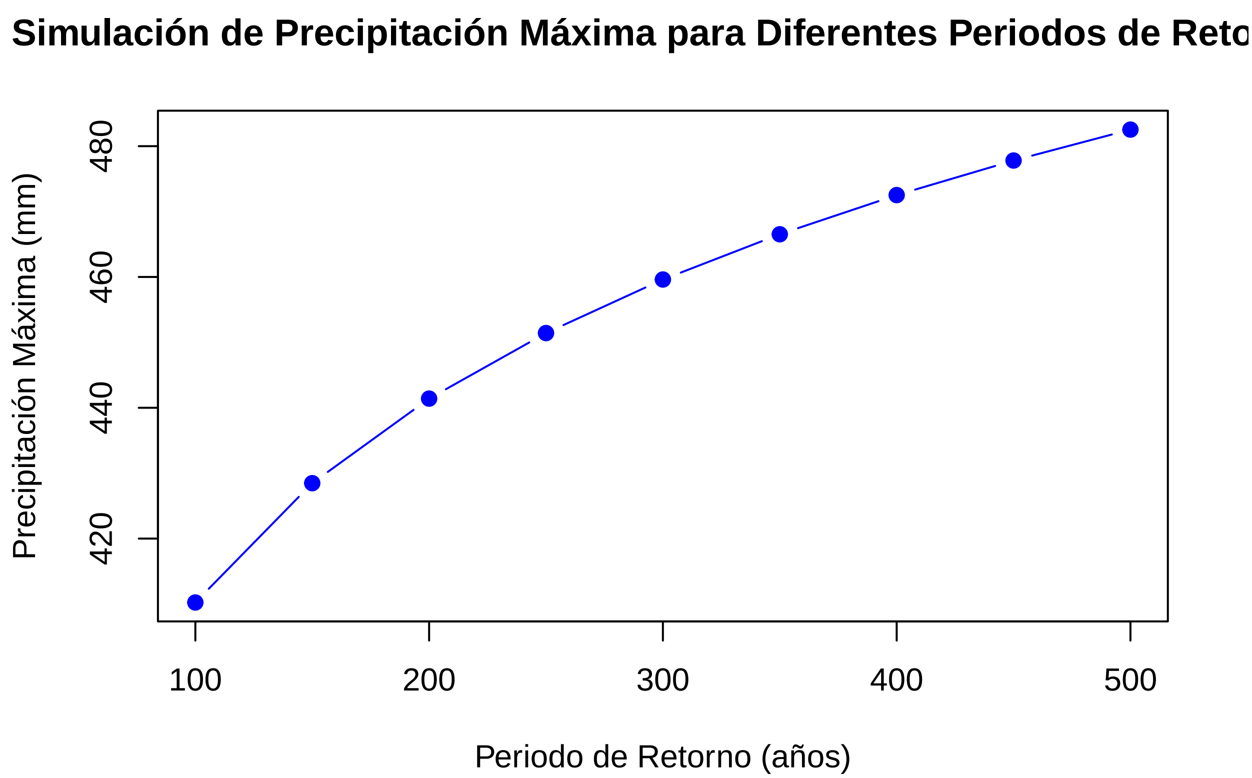
A continuación se calcula la precipitación máxima para varios períodos de retorno (de 100 a 500 años, con incrementos de 50 años) usando la distribución de Gumbel, primero se calcula la probabilidad de excedencia para cada período y luego se obtiene la precipitación máxima correspondiente usando los parámetros ajustados de Gumbel.

```
periodos_retorno <- seq(100, 500, by=50)

probabilidad_excedencia <- 1 / periodos_retorno
precipitacion_maxima <- qgumbel(1 - probabilidad_excedencia,
                                gumbel_fit$estimate[1],
                                gumbel_fit$estimate[2])

resultados <- data.frame(Periodo_Retorno = periodos_retorno,
                         Probabilidad_Excedencia = probabilidad_excedencia,
```

```r
                                    Precipitacion_Maxima = precipitacion_maxima)

print(resultados)
```

```
##   Periodo_Retorno Probabilidad_Excedencia Precipitacion_Maxima
## 1             100             0.010000000             410.2261
## 2             150             0.006666667             428.4720
## 3             200             0.005000000             441.4019
## 4             250             0.004000000             451.4244
## 5             300             0.003333333             459.6100
## 6             350             0.002857143             466.5289
## 7             400             0.002500000             472.5211
## 8             450             0.002222222             477.8057
## 9             500             0.002000000             482.5324
```

```r
plot(periodos_retorno, precipitacion_maxima, type="b", pch=19, col="blue",
     xlab="Periodo de Retorno (años)", ylab="Precipitación Máxima (mm)",
     main="Simulación de Precipitación Máxima para Diferentes Periodos de Retorno")
```



Análisando los resultados vemos que la precipitación máxima aumenta con el período de retorno. Conforme el período de retorno incrementa, la probabilidad de excedencia disminuye, indicando que precipitaciones extremas son menos probables en ciertos años.

Para lograr determinar cuál es el mejor período de retorno, se toman en cuenta diversos factores en función del análisis realizado y el diseño de la obra hidráulica.

*Si se quiere que la presa derivadora resista eventos extremos inusuales, es recomendable un perído de retorno mayor dado que esto reduciría la probabilidad de excedencia y por lo tanto reduciría el riesgo de falla debido a los eventos de precipitación.

*El aumentar el perído de retorno también va a aumentar los requisitos de diseño, lo cual por ende va a aumentar significativamente los costos de contrucción y mantenimeinto, si se cuenta con un presupuesto limitado, se puede optar por un período de retorno intermedio ya que se tienen un balance de seguridad y costo.

*De acuerdi a la tabla de perídos de retorno, el rango que se sugiere para presas de riesgo medianas, como en este caso, es de 100 a 500 años, por lo cual se optó por un periodo de retorno en este rango intermedio de entre 200 a 300 años, ya que es la práctica común en zonas de riesgo medio.

La eleciión que se hizo de un período de retorno de 300 años representa un equilibrio adecuado entre seguridad y practicidad, ya que representa un nivel de protección suficientemente bueno para la obra, considerando que eventos más extremos serían raros pero posibles. Este período esta dentro de los rangos de recomendación para estructuras hidráulicas de tipo mediano , por lo que hace equilibrio con los costos y la seguridad sin llegar a caer en exagerar costos para eventos extremadamente improbables.

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:

```r
Tr <- 300
Pexe <- 1 / Tr
cat("Probabilidad de excedencia:", Pexe)
```

```
## Probabilidad de excedencia: 0.003333333
```

Al ajustar el periodo de retorno a 300 años, la probabilidad de excedencia va a disminuir a 0.33% lo que implica una menor probabilidad de que la precipitación máxima esperada sea superada en un año dado. Esta cifra considera tambien un evento de menor frecuencia y mayor magnitud, el utilizar el período de retorno de 300 años nos asegura una protección adicional contra eventos extremos de precipitación, lo cual reduce el riesgo de que el sistema se vea superado.

C) Conociendo la probabilidad de excedencia, calcula su complemento (1 - Pexe) y utiliza esta probabilidad para encontrar el valor de la precipitación máxima mensual que tendrá ese periodo de retorno. En el código se te da un ejemplo si la distribución de probabilidad a la que mejor se ajustaron los datos fue la Gumbel y deseamos calcular el caudal máximo para un periodo de retorno de 300 años.

```r
Pnoexe <- 1 - Pexe

precipitacion_maxima <- qgumbel(Pnoexe, gumbel_fit$estimate[1], gumbel_fit$estimate[2])
cat("precipitacion maxima", precipitacion_maxima)
```

```
## precipitacion maxima 459.61
```

Con el período de retorno de 300 años, se obtiene una precipitación máxima estimada de aproximadamente 459.61 mm. Este valor refleja una mayor intensidad de precipitación que se espera que ocurra en intervalos de 300 años, lo cual proporciona un diseño más seguro y conservador para la obra hidráulica.

D) El resultado de este ejemplo será una aproximación del caudal máximo que se tendría en Sinalos con un periodo de retorno de 300 años. ¿Qué significa este valor?

El valor que se obtuvo 459.61 mm representa una estimación de la precipitación máxima que podría esperarse en Sinaloa en un periodo de retorno de 300 años, lo cual implica que un evento de precipitación de esta magnitud tiene una probabilidad de ocurrencia del 0.33% en cualquier año (1/300). Este es un parámetro fundamental en el diseño de obras hidraulicas ya que le permite a las y los ingenieros dimensionar la infraestructura para soportar eventos extremos y poder reducir y evitar el riesgo de fallos estructurales.

¿Qué pasa si incrementamos el periodo de retorno?

Si se incrementa el periodo de retorno , el valor de la precipitación máxima estimada va a a umentar, dado que un mayor periodo de retorno implica diseñar la infraestructura para resistir eventos más extremos que ocurren con menor frecuencia.El incremento del periodo de retorno implicaria un amento de costos de construcción dado que se diseña más robustamente para soportar mayores volúmenes de agua, es por eso que se busca el equilibrio entre seguridad y costos.

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

No, el caudal máximo estimado no será el mismo si se utilizan datos de otro estado, dado que cada estado tienen sus propias características climáticas y patrones de precipitación, por lo que los valores de precipitación máxima para un mismo periodo de retorno pueden variar.

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

Por seguridad, el diseñar obras hidráulicas en base de periodos de retorno sugeridos va a garantizar que la infraestructura pueda manejar eventos extremos de precipitación que son probables, aunque poco frecuentes, en cierta región. Esto ayuda con el balance de seguridad y costo de la obra, proporciona un buen y suficiente nivel de seguridad sin llegar a gastos inncesarios, si las obras hidráulicas se diseñan en base a periodos de retorno sugeridos los ingenieros se aseguran de que la infraestructura tenga la resilencia suficienteante los eventos climáticos más extremos del estado.

¿Por qué es importante conocer la distribución de probabilidad a la que mejor se aproximan los datos históricos?

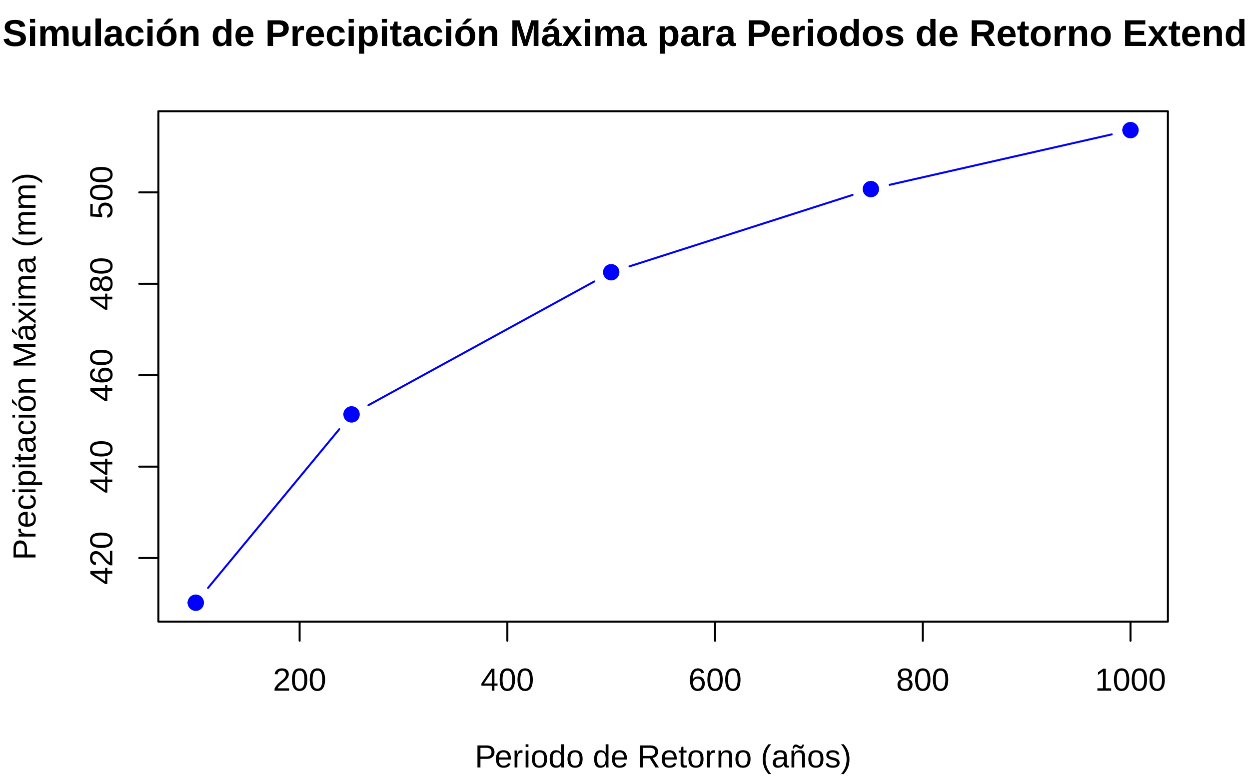
El conocer la distribución de probabilidad que mejor se ajusta a los datos históricos va a permitir hacer estimaciones más precisas y más confiables, lo cual es vital en el diseño de obras hidráulicas dado que al predecir la frecuencia y magnitud de eventos como inundaciones o tormentas severas, la distribución correcta garantiza la seguridad de la sociedad ya que se aseguran de que los valores de diseño no subestimen ni sobrestimen el riesgo real.

Explora otros periodos de retorno diferentes a los que se proporcionan en los periodos sugeridos para contestar esta pregunta.

```r
periodos_retorno_extendidos <- c(100, 250, 500, 750, 1000)
probabilidad_excedencia_extendida <- 1 / periodos_retorno_extendidos
precipitacion_maxima_extendida <- qgumbel(1 - probabilidad_excedencia_extendida,
                                          gumbel_fit$estimate[1],
                                          gumbel_fit$estimate[2])

resultados_extendidos <- data.frame(Periodo_Retorno = periodos_retorno_extendidos,
                                    Probabilidad_Excedencia = probabilidad_excedencia_extendida,
                                    Precipitacion_Maxima = precipitacion_maxima_extendida)
print(resultados_extendidos)
```

```
##   Periodo_Retorno Probabilidad_Excedencia Precipitacion_Maxima
## 1             100             0.010000000             410.2261
## 2             250             0.004000000             451.4244
## 3             500             0.002000000             482.5324
## 4             750             0.001333333             500.7180
## 5            1000             0.001000000             513.6178
```

```r
plot(periodos_retorno_extendidos, precipitacion_maxima_extendida, type="b", pch=19, col="blue",
     xlab="Periodo de Retorno (años)", ylab="Precipitación Máxima (mm)",
     main="Simulación de Precipitación Máxima para Periodos de Retorno Extendidos")
```



Se realizó la simulación para explorar otros periodos de retorno: 100, 250, 500, 750 y 1000 años ; se calculó la probabilidad de excedencia para cada uno de estos periodos, se calculó la precipitación máxima correspondiente usando la distribución Gumbel ajustada y se graficó la precipitación máxima en función del periodo de retorno extendido. Con el fin de análisar como aumentan los valores de precipitación máxima conforme incrementa el periodo de retorno, y justificar el uso de periodos de retorno razonables en el diseño.

Se elegió el período de retorno de 200 años dado que se priorizó el balance entre la seguridad hidráulica y el costo. En la simulación se observa que conforme incrementa el período de retorno los valores de la columna de precipitación máxima tambien aumenta, lo cual implica mayores costos y mayores dimensiones en las obras de infraestructura, elegir un perído de retorno de 300 años es adecuado para una zona de riego mediana y se ajusta a las recomendaciones de diseño hidráulico, lo cual entra en el margen de seguridad adicional en comparación de un período de 200 años, asegurando la seguridad pero sin incrementar los costos inecesariamente.

Referencias

- Probabilidad de excedencia. (2024). Retrieved October 24, 2024, from Probabilidad de excedencia website: https://www.osman.es/diccionario/definicion.php?id=13700#:~:text=Definici%C3%B3n%3A%20Probabilidad%2

- Lozano, G., Virginia, V. T., Romero Rodríguez, Moisés, José, C., Manuel, & Espinoza, S. (2024). Periodos de retorno de lluvias torrenciales para el estado de Tamaulipas, México. Investigaciones Geográficas, (76), 20–33. Retrieved from https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0188-46112011000300003#:~:text=El%20Periodo%20de%20Retorno%20de,M%C3%A9lice%20y%20Reason%2C%2020

*recomendaciones de periodos de retorno para la estimacion del gasto maximo de diseno en las obras hidraulicas, Comision Nacional del Agua, hidrologia de avenidas, Mexico. (2024). Retrieved October 28, 2024, from Sdsu.edu website: https://pon.sdsu.edu/periodos_de_retorno_cna.html