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

Eliezer Cavazos

2024-10-22

```r
oPrecipitaciones=read.delim("C:\\Users\\eliez\\OneDrive\\Desktop\\Clases\\precipitaciones_maximas_mensuales.txt", sep="\t", stringsAsFactors = F)

oCampeche = oPrecipitaciones[oPrecipitaciones$Estado == "Campeche",]
```

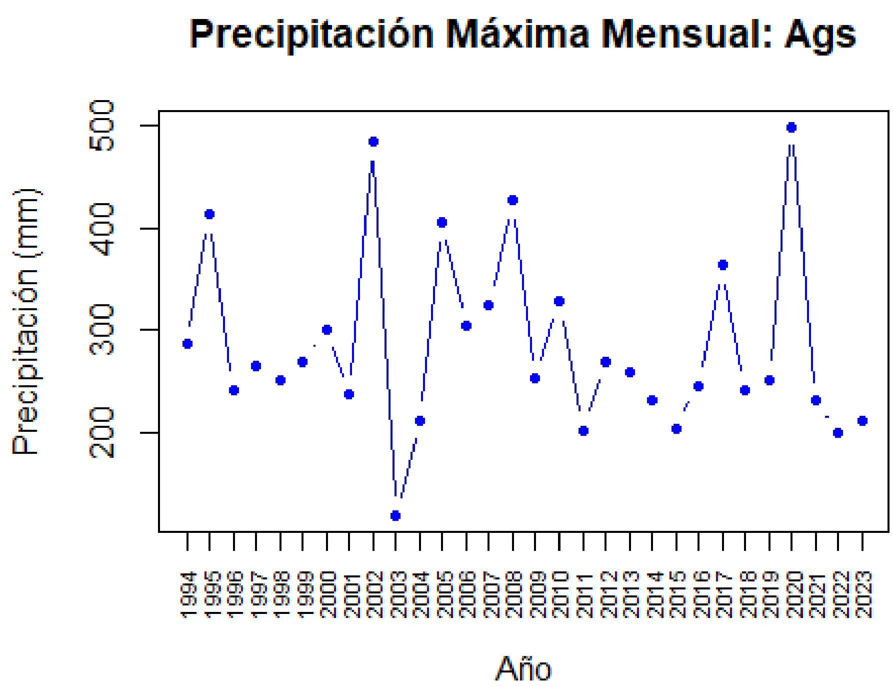
#ANÁLISIS

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

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

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.

```r
years = unique(oCampeche$Anio)
monthly_max = c()
for (n in 1:length(years)){
  monthly_max = c(monthly_max, max(oCampeche$Lluvia[which(oCampeche$Anio == years[n])]))}

names(monthly_max) = years
plot(monthly_max, type="b", pch=20, ylab="Precipitación (mm)", main="Precipitación Máxima Mensual: Ags", xaxt="n", col="blue", xlab="Año")
axis(1, at=1:30,labels=years, cex.axis=0.7, las=2)
```



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

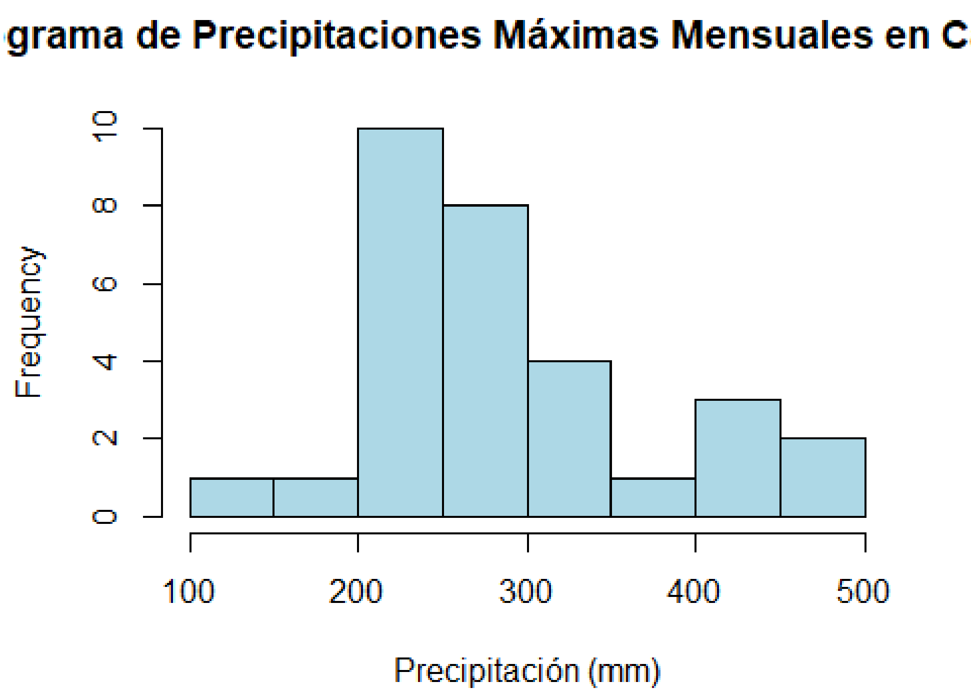
C.1 Calcula las medidas de centralización y variación de las precipitaciones máximas mensuales

```
# Medidas de centralización y variación
fMedia = mean(monthly_max)
fMediana = median(monthly_max)
fSD = sd(monthly_max)
fRango = range(monthly_max)

cat("Media: ", fMedia)

## Media:  284.06

print("\n")

## [1] "\n"

cat("Mediana: ", fMediana)

## Mediana:  255.75

print("\n")

## [1] "\n"

cat("SD: ", fSD)

## SD:  88.14132
```

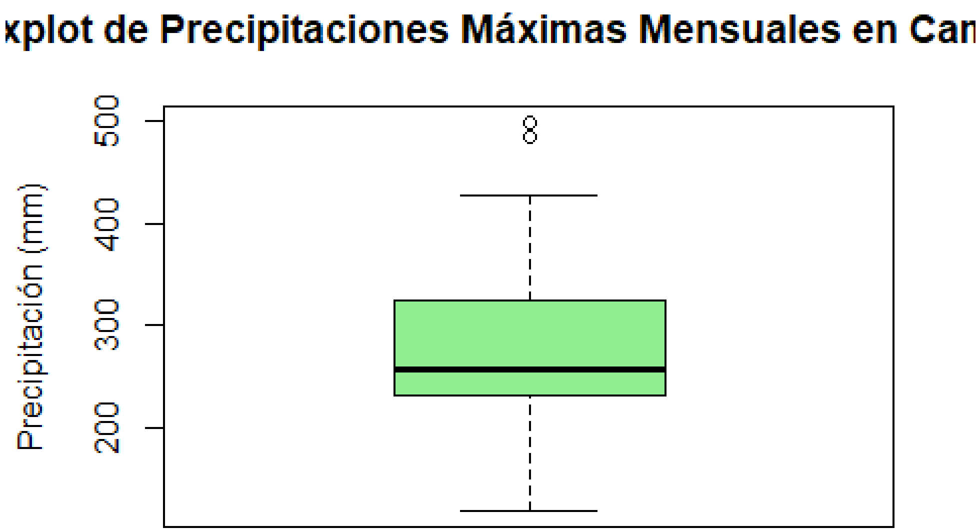
```r
print("\n")

## [1] "\n"

cat("Rango: ", fRango)

## Rango:  117.8 499.1
```

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

```r
# Histograma
hist(monthly_max, breaks=10, col="lightblue", main="Histograma de Precipitaciones Máximas Mensuales en Campeche", xlab="Precipitación (mm)")
```




```r
# Boxplot
boxplot(monthly_max, main="Boxplot de Precipitaciones Máximas Mensuales en Campeche", ylab="Precipitación (mm)", col="lightgreen")
```



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

Centralización: El valor de la media es de 284.06 y la mediana tiene un valor de 255.75 con esto podemos identificar que la distribucion de los datos no tiene un sesgo extremo hacia valores altos o pequeños

Variación: El valor de la desviacion estándar es de 88.14 y el rango esta entre 117.8 y 499.1 esto indica que existe una variabilidad significativa de las precipitaciones máximas mensuales.

Sesgo: Se puede observar por el histograma que los datos casi siguen una tendencia normal con un poco de sesgo a la derecha, esto hace que sea más facil de trabajar los datos sin ocupar de tanta limpieza de datos extremos.

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?

De las graficas puedo concluir que los datos de precipitaciones maximos mensuales no tienen muchos datos extremos y siguen cierta tendencia que se puede observar a diferencia de que si agarrara otro Estado, puedo ver que cada 5 tiende a subir tantito pero luego empieza a bajar la cantidad de precipitaiones.

2. Análisis de Frecuencias Método Gráfico

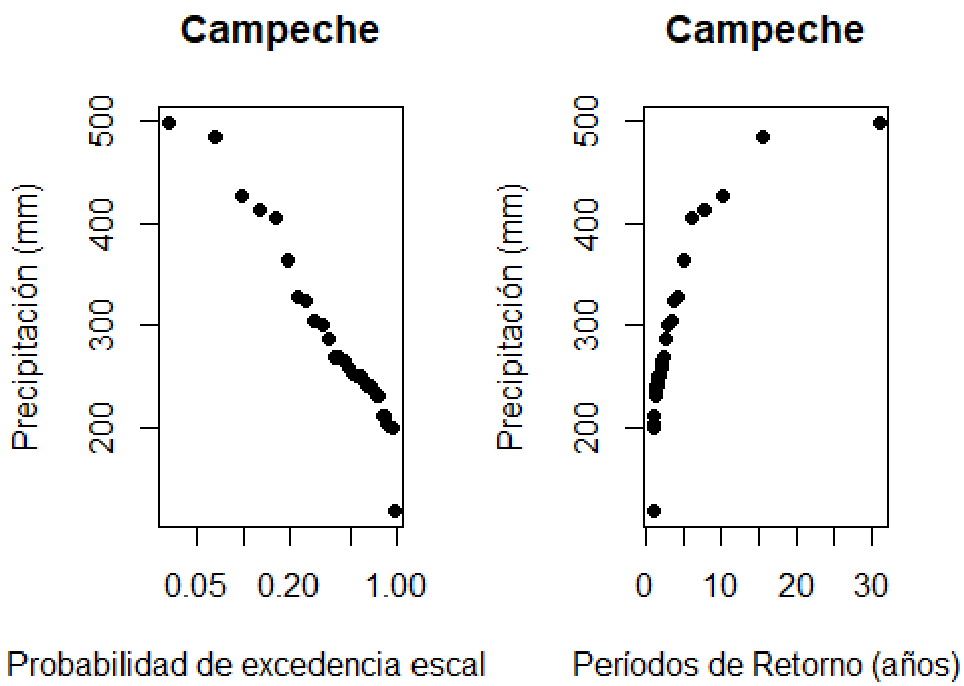
El Método gráfico consiste en realizar dos gráficas en la que se muestren las precipitaciones máximas comparadas con la probabilidad de excedencia y con su periodo de retorno.

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.

```
oAnalisisLluvia = data.frame(max_rain=monthly_max,
order_max_rain=sort(monthly_max, decreasing=T))

oAnalisisLluvia$rank_rain = seq(1, nrow(oAnalisisLluvia)) #columna del rank
(número de orden)
oAnalisisLluvia$Pexe = oAnalisisLluvia$rank_rain/(nrow(oAnalisisLluvia)+1)
#Probabilidad de excedencia: rank/(número de datos +1), P(X≥x)
oAnalisisLluvia$Pnoexe <- 1-oAnalisisLluvia$Pexe #Probabilidad de no
excedencia: 1-Pexe, P(X≤x)
oAnalisisLluvia$Pret = 1/oAnalisisLluvia$Pexe #Periodo de retorno
head(oAnalisisLluvia) #muestra los primeros 10 renglones del data frame

##      max_rain order_max_rain rank_rain       Pexe    Pnoexe      Pret
## 1994    286.2          499.1         1 0.03225806 0.9677419 31.000000
## 1995    413.4          484.3         2 0.06451613 0.9354839 15.500000
## 1996    241.4          426.7         3 0.09677419 0.9032258 10.333333
## 1997    263.9          413.4         4 0.12903226 0.8709677  7.750000
## 1998    250.7          405.5         5 0.16129032 0.8387097  6.200000
## 1999    268.9          363.6         6 0.19354839 0.8064516  5.166667
```

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

```
par(mfrow=c(1,2))
plot(y=oAnalisisLluvia$order_max_rain, x=oAnalisisLluvia$Pexe,
log="x",pch=19, main="Campeche", xlab="Probabilidad de excedencia escala
log", ylab="Precipitación (mm)")
plot(x=oAnalisisLluvia$Pret, y=oAnalisisLluvia$order_max_rain, pch=19,
main="Campeche", xlab="Períodos de Retorno (años)", ylab="Precipitación
(mm)")
```



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:

```
oAnalisisLluvia$Pexe
```

```
##  [1] 0.03225806 0.06451613 0.09677419 0.12903226 0.16129032 0.19354839
##  [7] 0.22580645 0.25806452 0.29032258 0.32258065 0.35483871 0.38709677
## [13] 0.41935484 0.45161290 0.48387097 0.51612903 0.54838710 0.58064516
## [19] 0.61290323 0.64516129 0.67741935 0.70967742 0.74193548 0.77419355
## [25] 0.80645161 0.83870968 0.87096774 0.90322581 0.93548387 0.96774194
```

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

```
oAnalisisLluvia$Pnoexe
```

```
##  [1] 0.96774194 0.93548387 0.90322581 0.87096774 0.83870968 0.80645161
##  [7] 0.77419355 0.74193548 0.70967742 0.67741935 0.64516129 0.61290323
## [13] 0.58064516 0.54838710 0.51612903 0.48387097 0.45161290 0.41935484
## [19] 0.38709677 0.35483871 0.32258065 0.29032258 0.25806452 0.22580645
## [25] 0.19354839 0.16129032 0.12903226 0.09677419 0.06451613 0.03225806
```

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

```
oAnalisisLluvia$Pret
```

```
##  [1] 31.000000 15.500000 10.333333  7.750000  6.200000  5.166667  4.428571
##  [8]  3.875000  3.444444  3.100000  2.818182  2.583333  2.384615  2.214286
## [15]  2.066667  1.937500  1.823529  1.722222  1.631579  1.550000  1.476190
```

```
## [22]  1.409091  1.347826  1.291667  1.240000  1.192308  1.148148  1.107143
## [29]  1.068966  1.033333
```

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?

Viendo las graficas se peude ver que los datos siguen una tendencia exponencial en los periodos de retorno y un decrecimiento en la probabilidad de excedencia escala log.

La probabilidad de excedencia es la probabilidad de que un evento, como una precipitación, supere un cierto valor en un periodo específico (generalmente un año).

El periodo de retorno es el tiempo promedio que se espera entre dos eventos que igualen o superen una cierta magnitud.

La Hidrologia es el estudio de todas las masas de agua de la tierra y en un sentido más estricto a la medida, recopilación y representación de los datos relativos al fondo del océano, las costas, las mareas, las corrientes de agua, cuerpos de agua continentales

Para el cálculo de la precipitación de diseño, se propone trabajar con una probabilidad de excedencia de 0.9 o con una probabilidad ocurrencia de 0.1.

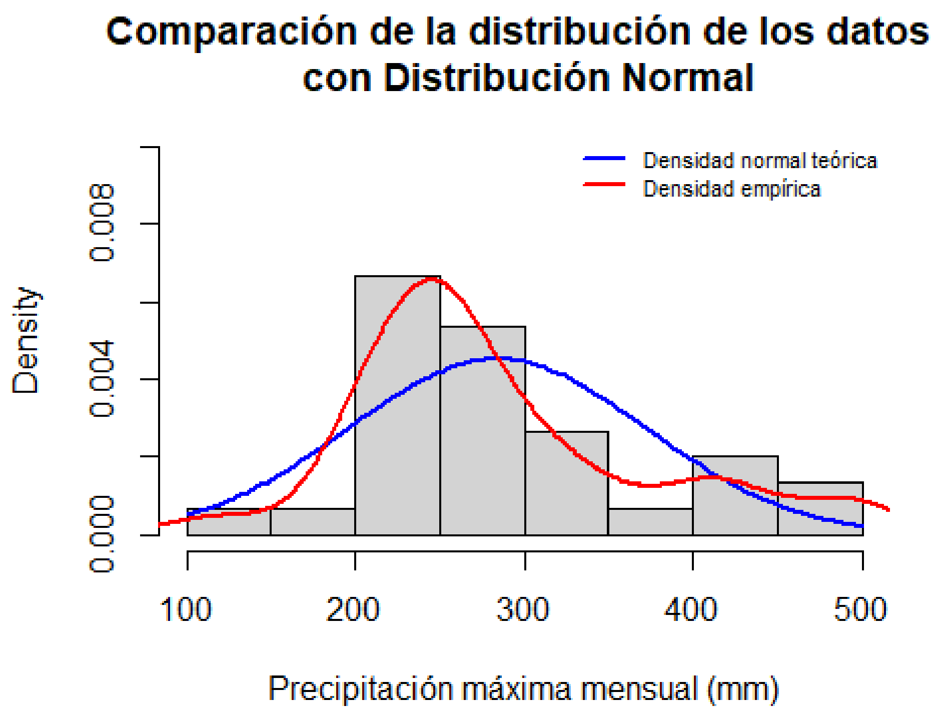
3. Análisis de Frecuencias Método Analítico

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

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

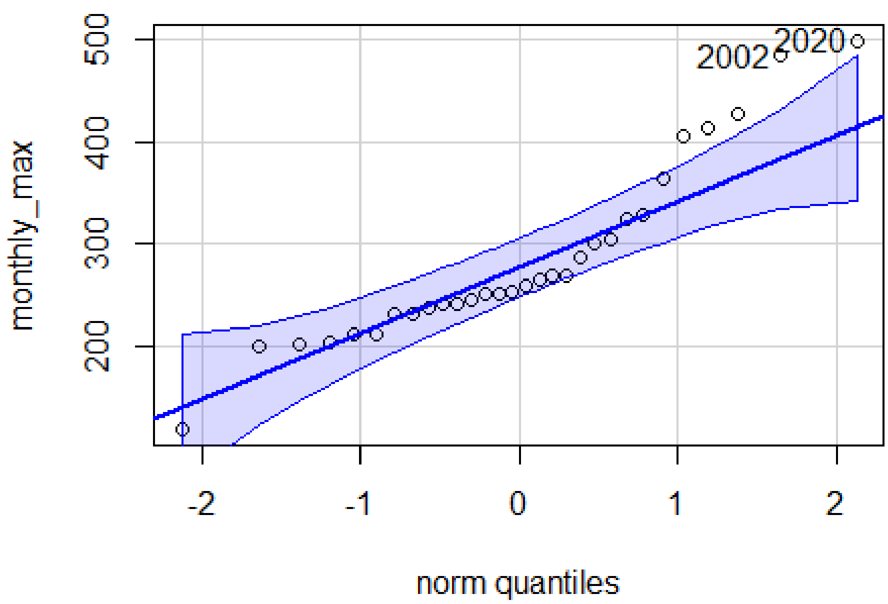
A.1 Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución normal que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. De manera visual, ¿te parece que los datos se ajustan bien a una distribución normal? Explica. ¿Cuántos parámetros tiene la distribución Normal? ¿Cuáles son? ¿Por qué los parámetros se calculan de la forma en cómo se hace en el código?

```r
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) #Estimación de parámetros por método de momentos
lines(density(monthly_max), col="red", lwd=2)
legend("topright", col=c("blue", "red"), legend =c("Densidad normal
teórica","Densidad empírica"),lwd=2, bty = "n", cex=0.7)
```



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

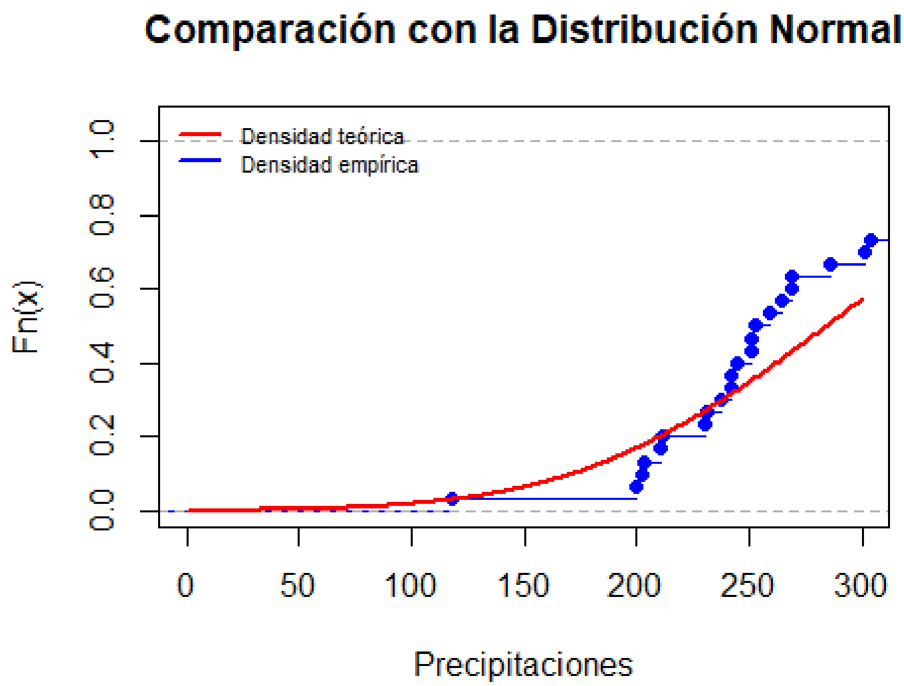
```
library(car) #Si  es la primera vez que la usas tienes que instalarla: Tools/Install Packages

## Loading required package: carData

qqPlot(monthly_max)
```




```
## 2020 2002
##   27    9
```

A.3 Compara las distribuciones de probabilidad acumuladas (ojiva) empíricos y teóricos de la distribución normal que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. Explica qué son datos empíricos y datos teóricos. ¿Se parecen las distribuciones de probabilidad acumuladas?

```
norm_teorica = pnorm(0:300, mean=mean(monthly_max), sd=sd(monthly_max))
#Estimación de parámetros por método de momentos
plot(ecdf(monthly_max), main="Comparación con la Distribución Normal",
xlab="Precipitaciones", col="blue", xlim=c(0,300), ylim=c(0, 1.05))
par(new=TRUE)
plot(0:300, norm_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","Densidad empírica"),lwd=2, bty = "n", cex=0.7)
```



A.4 Utiliza dos pruebas de bondad de ajuste: Shapiro-Wilks y Kolmogorov-smirnov (KS). ¿Qué información nos dan las pruebas? ¿Cuáles son los valores de los estadísticos? ¿Cuál es el p-value de las pruebas? ¿Se aceptan o se rechazan las hipótesis nulas? ¿Podemos concluir que los datos de las precipitaciones máximas mensuales son normales? ¿Por qué?

No te olvides de las hipótesis planteadas: H0: Los datos provienen de una distribución normal H1: Los datos no provienen de una distribución normal

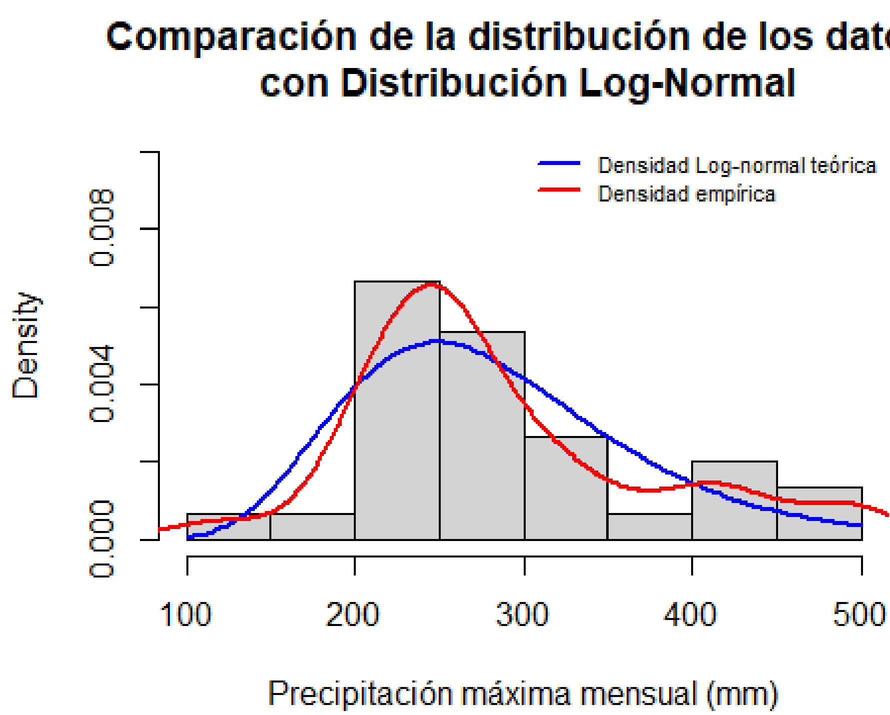
```r
shapiro.test(monthly_max)

##
##  Shapiro-Wilk normality test
##
## data:  monthly_max
## W = 0.90344, p-value = 0.0102

library(MASS)

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

##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  monthly_max
## D = 0.20161, p-value = 0.1517
## alternative hypothesis: two-sided
```

La diferencia del valor P en Shapiro y KS muestran una contradicción que se puede considerar que los datos tienen ligera desviación de normalidad, tambien se puede decir que Shapiro muestra que no es el mejor ajuste mientras que KS dice que la desviación de normalidad no es significativa. Asi que por el momento seria checar otra distribucion para ver si muestra valores más significativos.

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

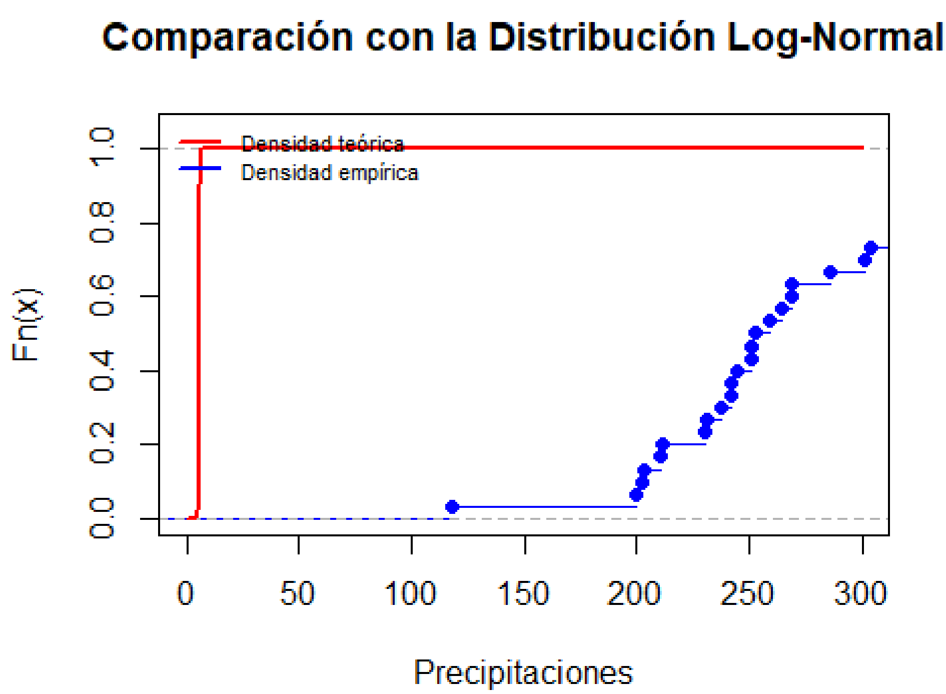
B.1 Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución Log-normal que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. De manera visual, ¿te parece que los datos se ajustan bien a una distribución Log-normal? Explica.

```r
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, mean=mean(log(monthly_max)), sd=sd(log(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 Log-normal
teórica","Densidad empírica"),lwd=2, bty = "n", cex=0.7)
```



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

```
log_teorica = pnorm(0:300, mean=mean(log(monthly_max)),
sd=sd(log(monthly_max))) #Estimación de parámetros por método de momentos
plot(ecdf(monthly_max), main="Comparación con la Distribución Log-Normal",
xlab="Precipitaciones", col="blue", xlim=c(0,300), ylim=c(0, 1.05))
par(new=TRUE)
plot(0:300, log_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","Densidad empírica"),lwd=2, bty = "n", cex=0.7)
```



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

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

##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  monthly_max
## D = 0.99919, p-value = 1.221e-15
## alternative hypothesis: two-sided
```

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

La distribución Log-Normal puede no ser apropiada para modelar estos datos, y se sugiere explorar otras distribuciones ya que el valor P es muy bajo, con esto se rechaza la hipotesis nula y podemos concluir que es mejor usar otra distribución

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

Tiene 2 parametros que son μ (media de la variable logarítmica) y σ (desviación estándar de la variable logarítmica).

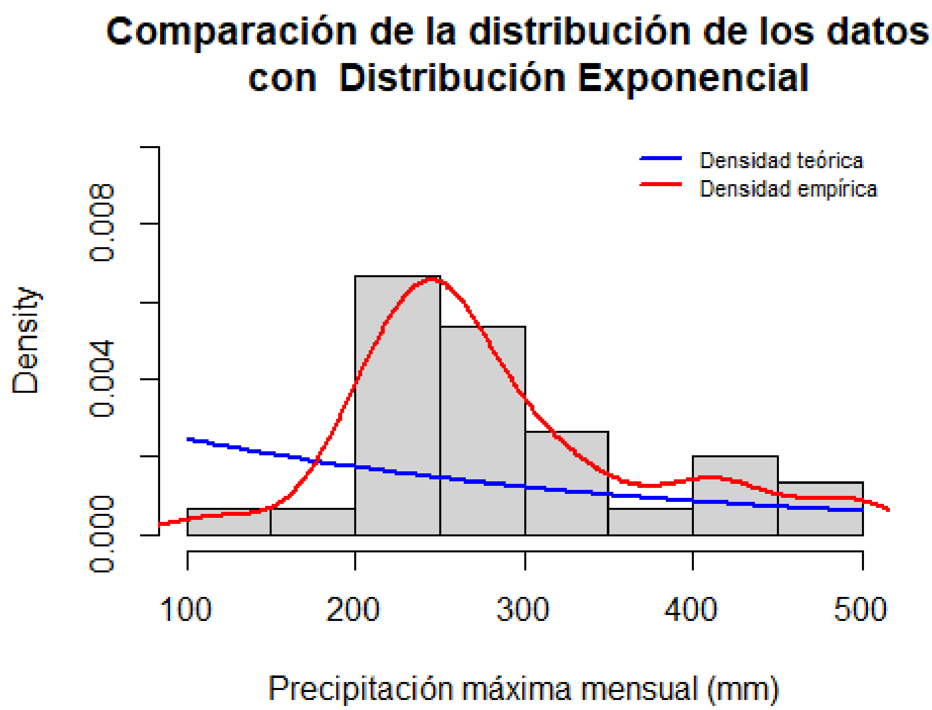
Se calculan en el método de momentos. Para la distribución log-normal.

1. Primero se calcula la media y la varianza de los datos originales.
2. Luego, con estas estimaciones, se puede calcular la media y la varianza teórica
3. Se comprara la media y varianza teórica con la media y varianza de los datos originales

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

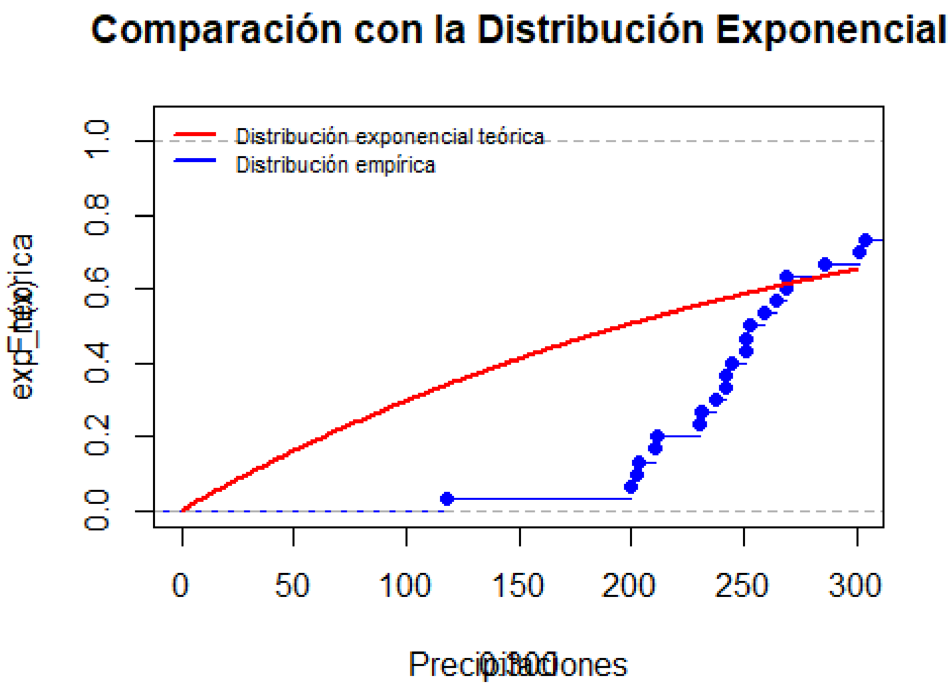
C.1 Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución Exponencial que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. De manera visual, ¿te parece que los datos se ajustan bien a una distribución Exponencial? Explica.

```
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) #Estimación
de parámetros por método de momentos
lines(density(monthly_max), col="red", lwd=2)
legend("topright", col=c("blue", "red"), legend =c("Densidad
teórica","Densidad empírica"),lwd=2, bty = "n", cex=0.7)
```



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

```
exp_teorica = pexp(0:300, rate=1/mean(monthly_max))
plot(ecdf(monthly_max), main="Comparación con la Distribución Exponencial",
xlab="Precipitaciones", col="blue", xlim=c(0,300), ylim=c(0, 1.05))
par(new=TRUE)
plot(0:300, exp_teorica, type="l", col="red", lwd=2, ylim=c(0, 1.05),
xaxt="n", yaxt="n")
legend("topleft", col=c("red", "blue"), legend =c("Distribución exponencial
teórica", "Distribución empírica"), lwd=2, bty = "n", cex=0.7)
```



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

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

##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  monthly_max
## D = 0.47088, p-value = 1.265e-06
## alternative hypothesis: two-sided
```

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

La prueba de KS rechaza la hipotesis nula ya que el valor de P es muy pequeño y se puede decir que la distribución Exponencial no es adecuada para los datos, ya que los datos no siguen una distribución exponencial.

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

La distribución Exponencial tiene un parametro que es λ (tasa de ocurrencia).

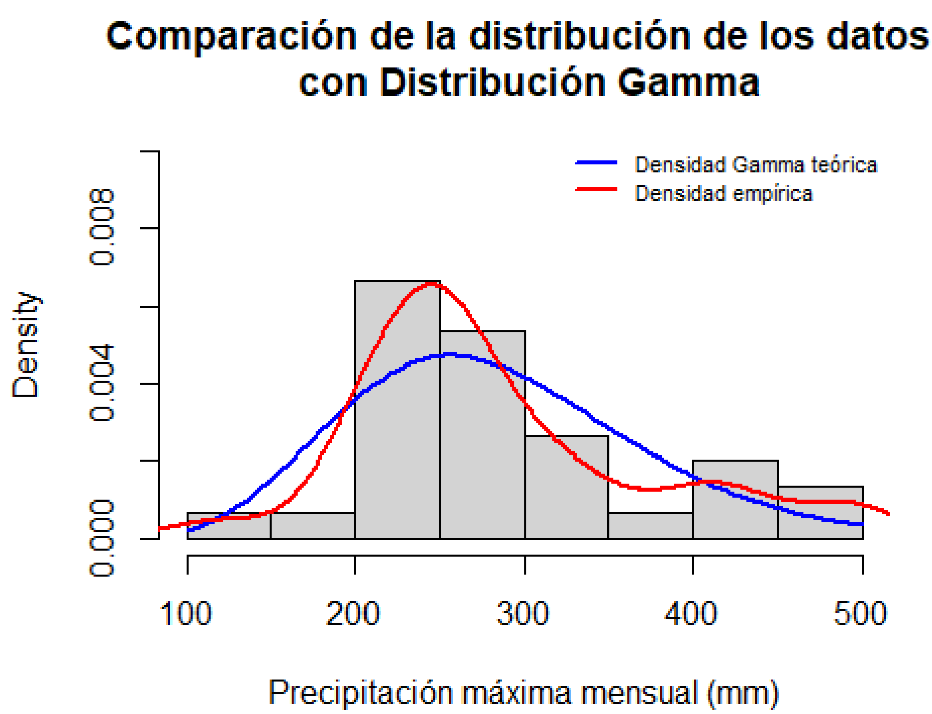
Para calcular λ se calcula como el inverso de la media de los datos ” 1/media “, entonces si se conoce la media muestral de los datos, se puede calcular λ.

1. Primero se calcula la media de los datos
2. Ya con la media calculada apartir de los datos se debe checar que la media teórica de la distribución exponencial son iguales para demostrar que los parámetros están bien calculados.

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

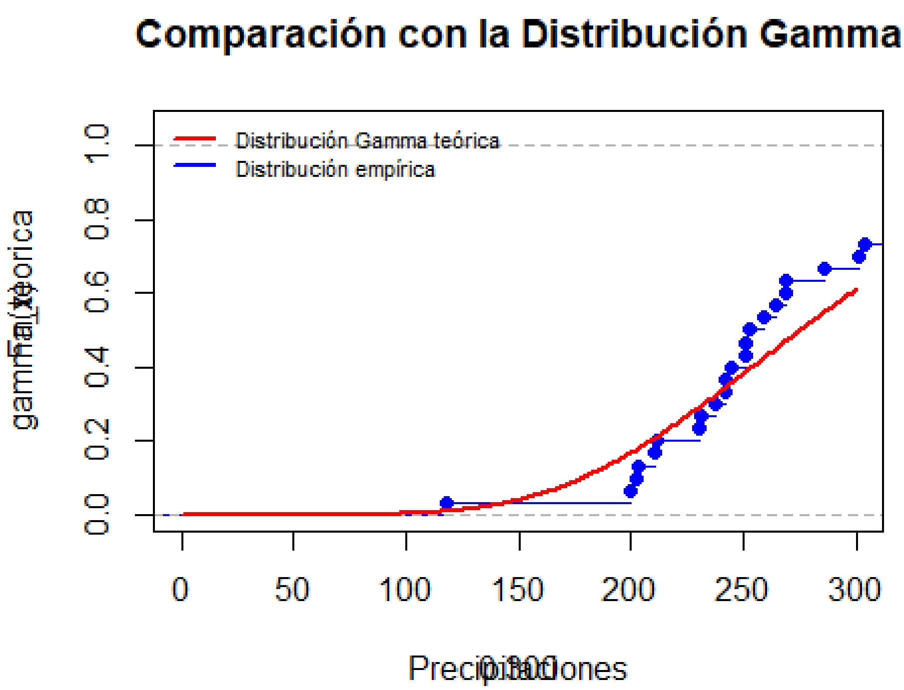
D.1 Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución Gamma que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. De manera visual, ¿te parece que los datos se ajustan bien a una distribución Gamma? Explica.

```
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 Gamma")
curve(dgamma(x, mean(monthly_max)^2/var(monthly_max),
mean(monthly_max)/var(monthly_max)), add=TRUE, col="blue",lwd=2) #Estimación
de parámetros por método de momentos
lines(density(monthly_max), col="red", lwd=2)
legend("topright", col=c("blue", "red"), legend =c("Densidad Gamma
teórica","Densidad empírica"),lwd=2, bty = "n", cex=0.7)
```



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

```r
alpha <- mean(monthly_max)^2 / var(monthly_max)
beta <- var(monthly_max) / mean(monthly_max)

gamma_teorica <- pgamma(0:300, shape=alpha, scale=beta)

plot(ecdf(monthly_max), main="Comparación con la Distribución Gamma",
xlab="Precipitaciones", col="blue", xlim=c(0,300), ylim=c(0, 1.05))

par(new=TRUE)
plot(0:300, gamma_teorica, type="l", col="red", lwd=2, ylim=c(0, 1.05),
xaxt="n", yaxt="n")

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



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

```r
ks.test(monthly_max, "pgamma", shape=alpha, scale=beta)

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

```
## D = 0.16165, p-value = 0.373
## alternative hypothesis: two-sided
```

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

La prueba KS muestra que no se puede rechazar aun la hipótesis nula, ya que el valor P es mayor a 0.05, esto significa que los datos si se pueden ajustar a una distribución Gamma.

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

La distribución Gamma tiene dos parametros k (también llamado shape) y θ (también llamado scale).

k se puede calcular: (media)^2 / varianza.

θ se puede calcular: varianza / media.

1. Primero se calcula la media y la varianza de los datos
2. Con estos datos se calcula los parametros k y θ
3. Despues se verifica asegurandose que:

- k * θ = media
- k * (θ)^2 = varianza

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.

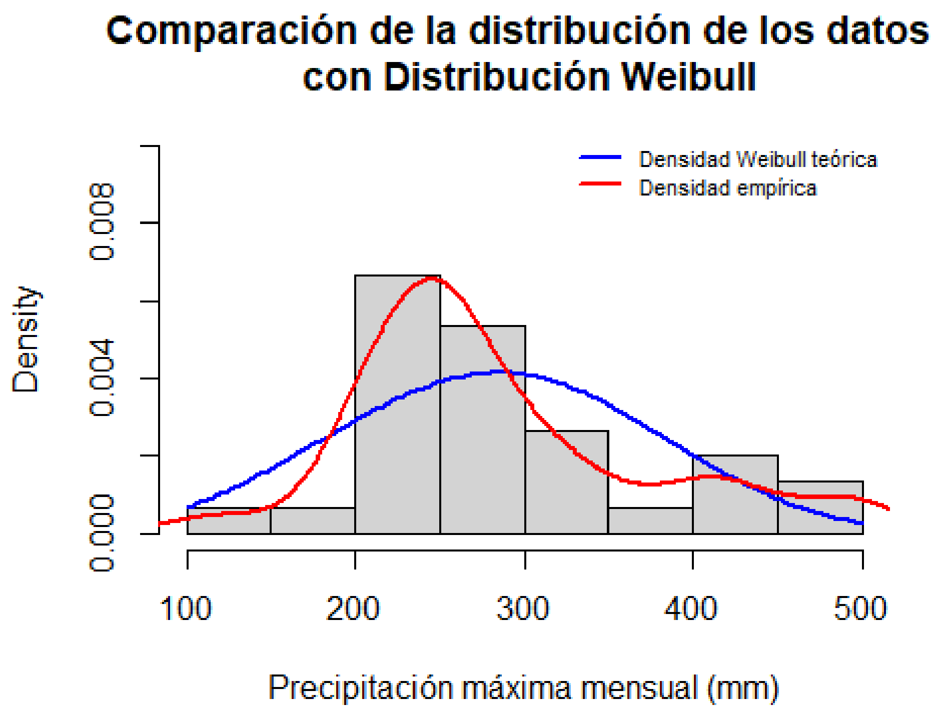
E.1 El cálculo de los parámetros a partir de los datos es un poco más difícil en la distribución Weibull de lo que fue en las anteriores distribuciones, así que recurriremos a que R los estime con el comando fitdistr. Úsalo para estimar los parámetros de la Weibull.

```
library(MASS)
weibull_fit <- fitdistr(monthly_max, "weibull", lower=c(0,0))
shape_weibull <- weibull_fit$estimate["shape"]
scale_weibull <- weibull_fit$estimate["scale"]
```

E.2 Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución Weibull que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. De manera visual, ¿te parece que los datos se ajustan bien a una distribución Weibull? Explica.

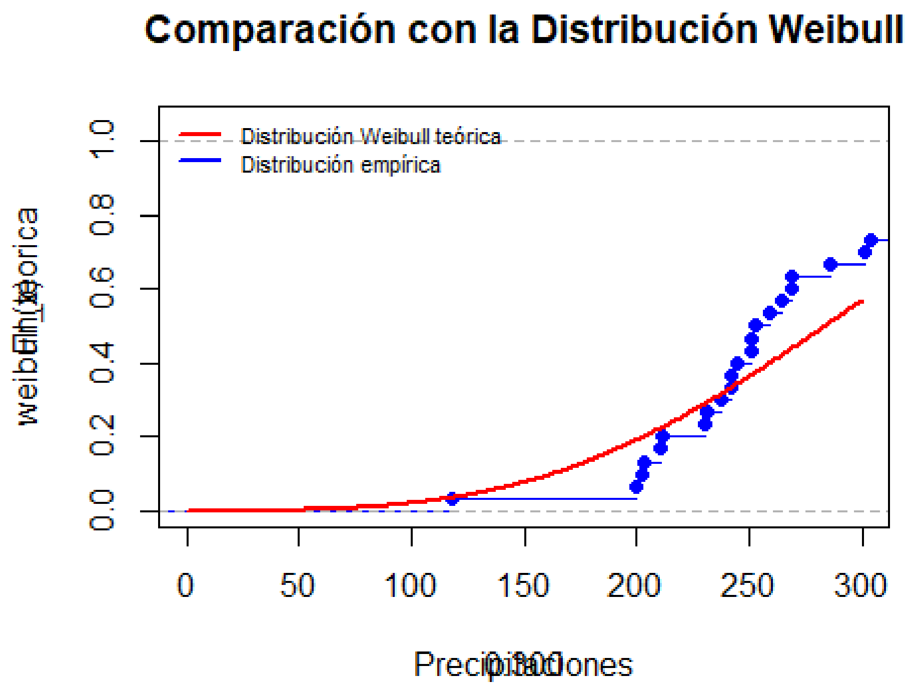
```
hist(monthly_max, xlab="Precipitación máxima mensual (mm)", freq=FALSE,
ylim=c(0, 0.01), main="Comparación de la distribución de los datos \n con
Distribución Weibull")
curve(dweibull(x, weibull_fit$estimate[1], weibull_fit$estimate[2]),
add=TRUE, col="blue",lwd=2)
lines(density(monthly_max), col="red", lwd=2)
```

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



E.3 Compara las distribuciones de probabilidad acumuladas (ojiva) empíricos y teóricos de la distribución teórica que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. Explica qué son datos empíricos y datos teóricos. ¿Se parecen las distribuciones de probabilidad acumuladas?

```
weibull_teorica <- pweibull(0:300, shape=shape_weibull, scale=scale_weibull)
plot(ecdf(monthly_max), main="Comparación con la Distribución Weibull",
xlab="Precipitaciones", col="blue", xlim=c(0,300), ylim=c(0, 1.05))
par(new=TRUE)
plot(0:300, weibull_teorica, type="l", col="red", lwd=2, ylim=c(0, 1.05),
xaxt="n", yaxt="n")
legend("topleft", col=c("red", "blue"), legend=c("Distribución Weibull
teórica", "Distribución empírica"), lwd=2, bty="n", cex=0.7)
```



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

```
ks.test(monthly_max, "pweibull", shape=shape_weibull, scale=scale_weibull)

##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  monthly_max
## D = 0.19292, p-value = 0.1879
## alternative hypothesis: two-sided
```

E.5 ¿Qué información nos da la prueba KS para una Weibull? ¿Cuál es el valor del estadístico de prueba? ¿Cuál es el p-value de la prueba? ¿Se acepta o se rechaza la hipótesis nula? ¿Podemos concluir que los datos de las precipitaciones máximas mensuales siguen una distribución Weibull? ¿Por qué?

La prueba de KS nos dice que la hipotesis nula no se puede rechazar, porque el valor P es mayor a 0.05 y se sugiere que los datos podrian seguir una distribución Weibull.

E.6 ¿Cuántos parámetros tiene la distribución Weibull? ¿Cuáles son? Explica por qué es más compleja la estimación de los parámetros a partir de los datos en esta distribución a diferencia de las distribuciones anteriores.

La distribución Weibull tiene dos parametros k (también llamado shape) y θ (también llamado scale).

A diferencia de la Distribución Gamma los parametros tienen un valor diferente.

θ: Este parámetro afecta la "dispersión" de la distribución. En términos generales, un valor mayor de θ desplaza la distribución hacia la derecha, lo que significa que las observaciones son en promedio más grandes.

k: Este parámetro forma determina la forma de la distribución. Dependiendo de su valor - Si $k<1$, la distribución tiene una forma decreciente, lo que indica que la tasa de fallos disminuye con el tiempo. - Si $k=1$, la distribución se convierte en una distribución exponencial, lo que indica que la tasa de fallos es constante. - Si $k>1$, la distribución tiene una forma creciente, lo que significa que la tasa de fallos aumenta con el tiempo.

Algunas de las razones por la que la estimacion de los datos en esta distribución son por: - No linealidad: A diferencia de las distribuciones normales, donde los momentos (media y varianza) pueden ser usados directamente para estimar parámetros, la relación entre k, θ, la media y la varianza en Weibull no es directa. -Dependencia de k: El comportamiento del parámetro k cambia drásticamente el modelo de la distribución, lo que significa que pequeñas variaciones en los datos pueden resultar en cambios significativos en las estimaciones de θ. - Sensibilidad a los datos extremos: La forma de la distribución Weibull puede ser muy sensible a los valores extremos (outliers), lo que puede complicar la estimación de parámetros.

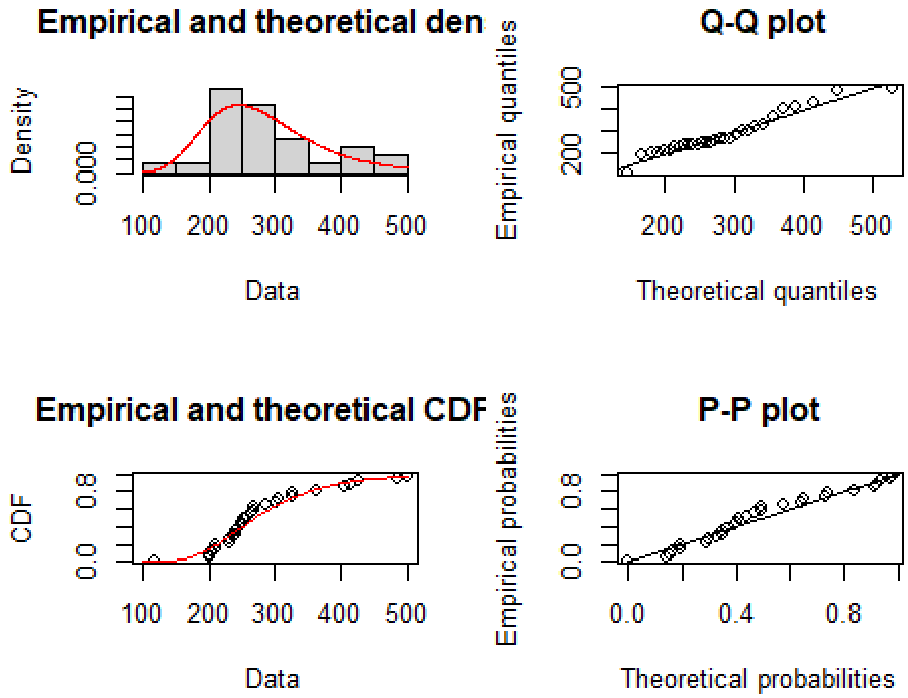
F. Ajuste a una Distribución Gumbel. Realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Gumbel para ajustar los datos.

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.

```
library(fitdistrplus) #si es la primera vez que lo usas, necesitas instalarlo: Tools/Install Packages, además, te pedirá instalar la biblioteca survival

## Loading required package: survival

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




```
gumbel_exe = 1-pgumbel(oAnalisisLluvia$order_max_rain,
gumbel_fit$estimate[1], gumbel_fit$estimate[2]) #Calculo de la probabilidad
de excendencia teórica
```

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

```
ks.test(oAnalisisLluvia$Pexe, gumbel_exe)

##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  oAnalisisLluvia$Pexe and gumbel_exe
## D = 0.13333, p-value = 0.9578
## alternative hypothesis: two-sided
```

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

De todas las distribuciones creo que este ha sido el mejor ya que el valor D es muy bajo esto dice que hay una discrepancia pequeñas entra la distribucion empirica, eso dice que las distribuciones son muy similares. Ademas el valor P es mayor por mucho a 0.05 y no se puede rechazar la hipótesis nula.

F.5 ¿Cuántos parámetros tiene la distribución Gumbel? ¿Cuáles son? Estima los parámetros de la Gumbel partiendo de la media y la desviación estándar de los datos y con la fórmula de la media y la desviación estándar de la Gumbel. Compara los valores obtenidos con los estimados con el comando "fitdistrplus" ¿se obtienen los mismos valores por ambos métodos? ¿por qué crees que se dé esta diferencia? ¿por qué crees que se dé esta diferencia?

La distribución Gumbel tiene dos parámetros μ (ubicación) y β (escala)

μ: Este parámetro representa la ubicación de la distribución. Es el valor alrededor del cual la distribución se centra.

β: Este parámetro de escala controla la dispersión de la distribución. Un valor mayor de β indica una mayor dispersión de los datos alrededor de μ.

Como estimar los parametros apartir de Media y Desviación Estándar

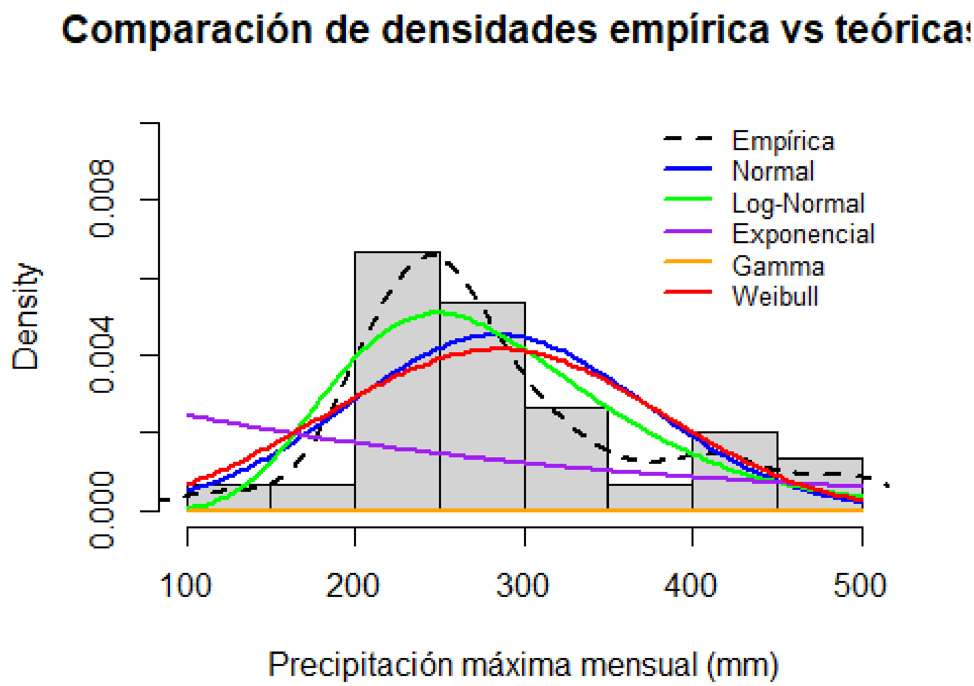
$_G = + 0.5772$ $\sigma_G = \frac{\beta\pi}{\sqrt{6}}$

La diferencia se puede dar por el método de estimación o la suposición de la normalidad o tambien por la variabilidad de los datos.

G. Compara los ajustes de las distribuciones que analizaste

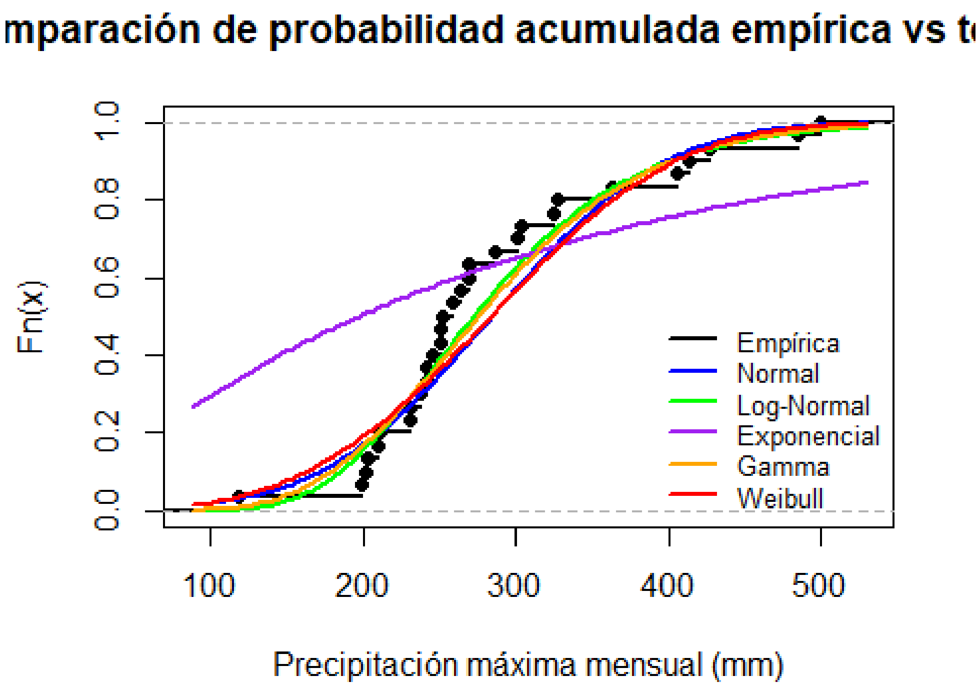
G.1 Haz un gráfico comparativo de los histogramas de densidad empírica vs densidad teórica de todas las distribuciones que analizaste (todas las distribuciones en un solo gráfico.

```
# Histograma empírico con densidades teóricas superpuestas
hist(monthly_max, probability=TRUE, main="Comparación de densidades empírica
vs teóricas", xlab="Precipitación máxima mensual (mm)", ylim=c(0, 0.01))

# Densidad empírica
lines(density(monthly_max), col="black", lwd=2, lty=2)

# Densidad teórica para cada distribución ajustada
curve(dnorm(x, mean=mean(monthly_max), sd=sd(monthly_max)), col="blue",
lwd=2, add=TRUE)          # Normal
curve(dlnorm(x, meanlog=mean(log(monthly_max)), sdlog=sd(log(monthly_max))),
col="green", lwd=2, add=TRUE) # Log-normal
curve(dexp(x, rate=1/mean(monthly_max)), col="purple", lwd=2, add=TRUE)
# Exponencial
curve(dgamma(x, mean(monthly_max)^2/var(monthly_max)), col="orange", lwd=2,
add=TRUE) # Gamma
curve(dweibull(x, shape=weibull_fit$estimate["shape"],
scale=weibull_fit$estimate["scale"]), col="red", lwd=2, add=TRUE) # Weibull

# Leyenda para identificar cada distribución
legend("topright", legend=c("Empírica", "Normal", "Log-Normal",
"Exponencial", "Gamma", "Weibull"),
       col=c("black", "blue", "green", "purple", "orange", "red"), lwd=2,
lty=c(2, 1, 1, 1, 1, 1), bty="n", cex=0.8)
```



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

```
# Ojiva empírica
plot(ecdf(monthly_max), main="Comparación de probabilidad acumulada empírica
vs teóricas", xlab="Precipitación máxima mensual (mm)", col="black", lwd=2,
ylim=c(0, 1))

# Ojivas teóricas de cada distribución ajustada
curve(pnorm(x, mean=mean(monthly_max), sd=sd(monthly_max)), col="blue",
lwd=2, add=TRUE)           # Normal
curve(plnorm(x, meanlog=mean(log(monthly_max)), sdlog=sd(log(monthly_max))),
col="green", lwd=2, add=TRUE) # Log-normal
curve(pexp(x, rate=1/mean(monthly_max)), col="purple", lwd=2, add=TRUE)
# Exponencial
curve(pgamma(x,shape=alpha, scale=beta), col="orange", lwd=2, add=TRUE) #
Gamma
curve(pweibull(x, shape=weibull_fit$estimate["shape"],
scale=weibull_fit$estimate["scale"]), col="red", lwd=2, add=TRUE) # Weibull

# Leyenda
legend("bottomright", legend=c("Empírica", "Normal", "Log-Normal",
"Exponencial", "Gamma", "Weibull"),
       col=c("black", "blue", "green", "purple", "orange", "red"), lwd=2,
bty="n", cex=0.8)
```



G.3 Define cuál es la mejor distribución que se ajusta a tus datos. Argumenta interpretando la comparación entre los gráficos y analizando las pruebas de ajuste de curva.

La mejor distribucion que puedo identificar por la prueba de KS fue la distribucion de Gumbel ya que es el que tiene el valor D se acerca más a 0 significa que la distribución teórica se ajusta bien a los datos y ademas el que tiene el valor P más grande que esto indica que los datos no se alejan significativamente de la distribución teórica.

#DISEÑO DE OBRAS HIDRÁULICAS

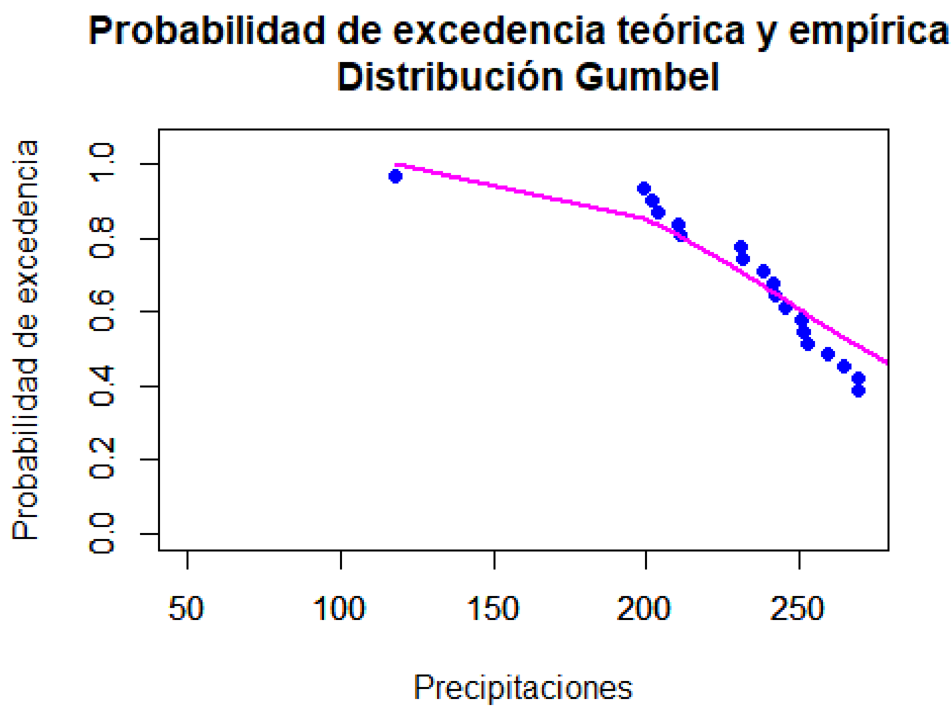
4. Precipitación de diseño de obras hidráulicas

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

```
plot(oAnalisisLluvia$order_max_rain, oAnalisisLluvia$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) #gráfico de probabilidad empírica, calculado en el punto 2
par(new=TRUE)
plot(oAnalisisLluvia$order_max_rain, gumbel_exe, type="l", main="", xlab="",
```

```
ylab="", col="magenta", lwd=2, xlim=c(50,270), ylim=c(0, 1.05)) #gráfico de
probabilidad teórica
```



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:

```
# Probabilidad de excedencia para 100 años que fue lo que se encontro en el
pdf
Pexe_100 = 1 / 100
Pnoexe_100 = 1 - 100

# Valor de precipitación máxima mensual para 50 años de periodo de retorno
precipitacion_max_50 = qgumbel(p = Pnoexe_100, a = gumbel_fit$estimate[1], b
= gumbel_fit$estimate[2])

## Warning in log(p): NaNs produced

precipitacion_max_50

##   a
## NaN
```

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

```
Pexe = 1/200
Pnoexe = 1 - Pexe
qgumbel(p = Pnoexe, a=gumbel_fit$estimate[1], b=gumbel_fit$estimate[2])

##        a
## 612.1259
```

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

El valor del caudal máximo representa la precipitación máxima mensual esperada en Campeche, correspondiente a un periodo de retorno de 200 años. Es decir, se espera que una precipitación tan alta o mayor ocurra, en promedio, una vez cada 200 años, lo cual significa una probabilidad de ocurrencia baja pero no imposible.

Al aumentar el periodo de retorno, se obtendría un valor de precipitación máxima aún mayor, reflejando eventos de precipitación más extremos, aunque con menor probabilidad de ocurrencia anual.

Los valores de precipitación máxima para el mismo periodo de retorno pueden diferir significativamente entre estados debido a variaciones climáticas y geográficas, por lo que el uso de datos locales es esencial para obtener estimaciones realistas y precisas.

Las infraestructuras deben diseñarse tomando en cuenta periodos de retorno sugeridos, ya que estos reflejan la probabilidad de ocurrencia de eventos extremos. Al utilizar periodos de retorno largos, se garantiza mayor seguridad para la infraestructura ante eventos de alta precipitación.

Conocer la distribución que mejor se ajusta a los datos permite modelar con precisión el comportamiento de las precipitaciones y calcular los valores extremos de manera confiable, lo cual es crucial para la planificación y el diseño de obras resistentes.