Actividad Integradora 1

Facundo Colasurdo Caldironi

2024-10-22

##ANÁLISIS

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

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.

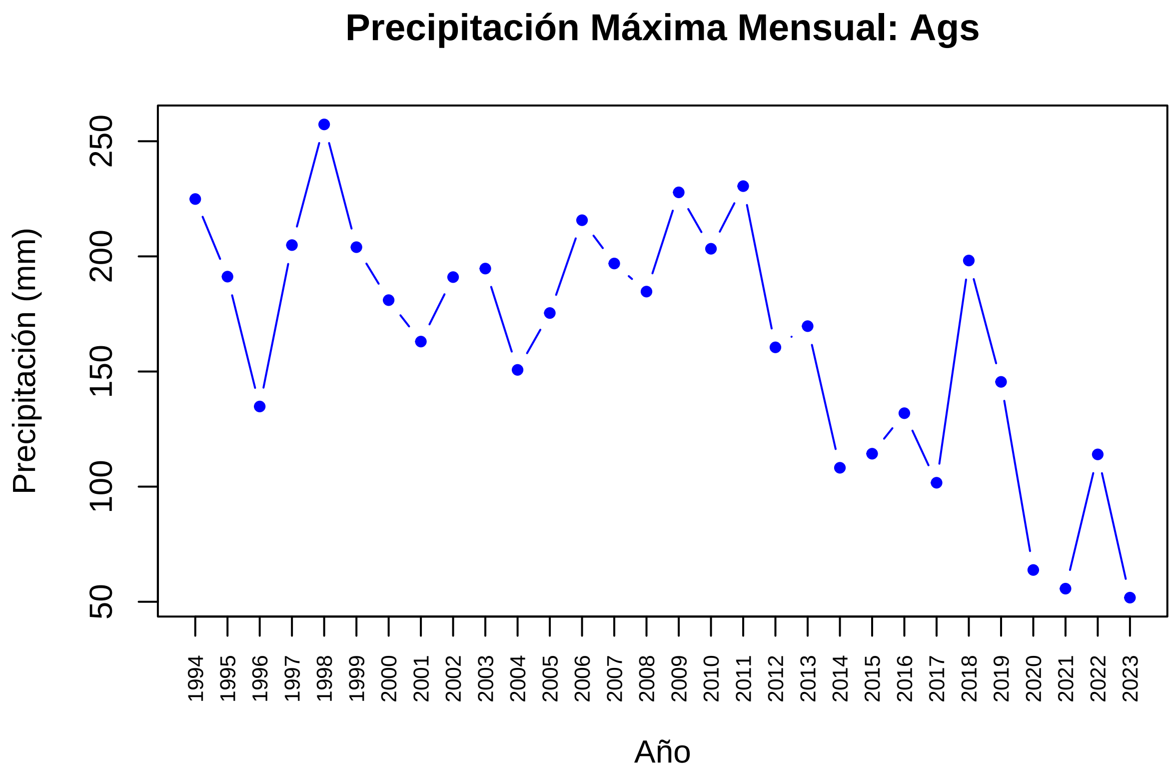
```
rain_data=read.table("file:///Users/facundocolasurdocaldironi/Downloads/precipitaciones_maximas_mensuale
```

```
rain_Ags= rain_data[which(rain_data$Estado =="Ciudad.de.México"), ]
```

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_Ags$Anio)
monthly_max = c()
for (n in 1:length(years)){
  monthly_max = c(monthly_max, max(rain_Ags$Lluvia[which(rain_Ags$Anio == years[n])]))}

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



Analiza los datos de precipitaciones máximas mensuaels del estado seleccionado. Calcula las medidas de centralización y variación de las precipitaciones máximas mensuales

```
mean_max = mean(monthly_max)
median_max = median(monthly_max)
variance_max = var(monthly_max)
sd_max = sd(monthly_max)

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

```
## Media:  164.9033
```

```
cat("Mediana: ", median_max, "\n")
```

```
## Mediana:  178.2
```

```
cat("Varianza:", variance_max, "\n")
```

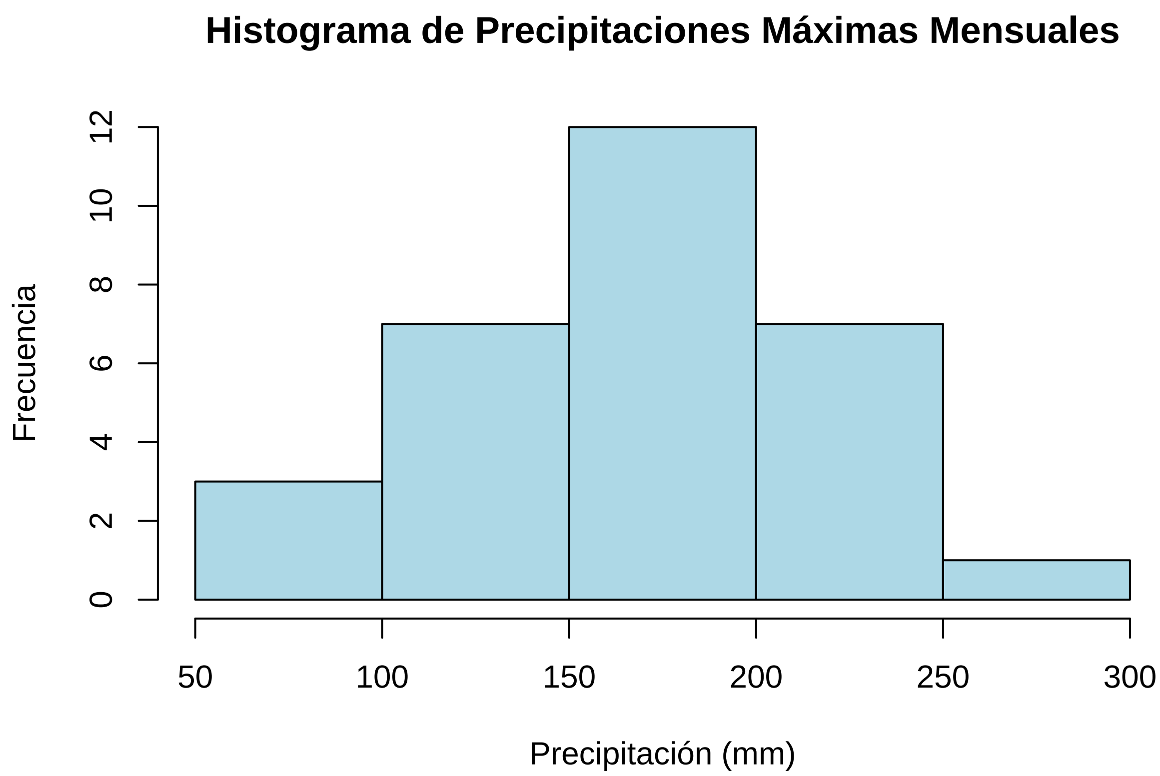
```
## Varianza: 2850.126
```

```
cat("Desviación estándar:", sd_max, "\n")
```

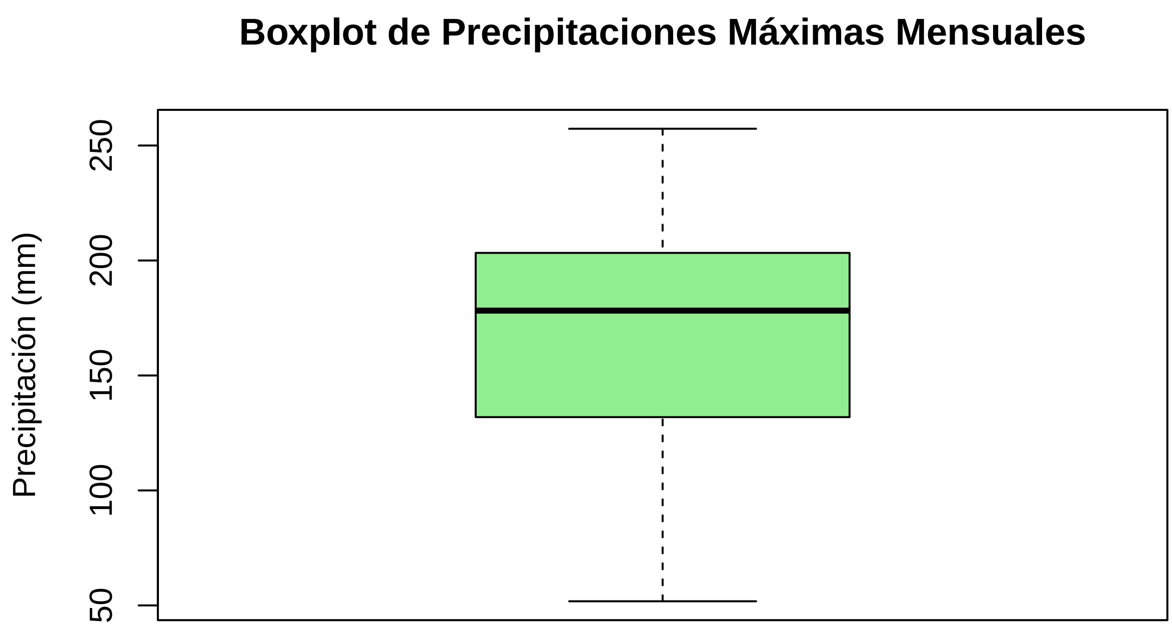
```
## Desviación estándar: 53.38657
```

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

```
hist(monthly_max, main="Histograma de Precipitaciones Máximas Mensuales", xlab="Precipitación (mm)", yla
```




```
boxplot(monthly_max, main="Boxplot de Precipitaciones Máximas Mensuales", ylab="Precipitación (mm)", col
```



Describe el comportamiento de la distribución: centralización, sesgo, variación, …. Al analizar la media y mediana, es posible ver que se encuentra sesgada a la izquierda, por lo que se considera negativa, el boxplot nos demuestra que existen valores atipicos, tanto en los 250 mm a 50 mm.

¿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?

Al analizar la gráfica de máximos mensuales anuales, fue posible ver que la mayor cantidad de precipitación ocurrió en 1998, se puede observar que que tiene una tendencia de lluvia entre 150 a 200 mm, al analizar las fechas, fue posible ver que la cantidad de precipitación bajo en los ultimos años, la razon de por qué nos sirve analizar este tipo de gráficas, ha sido mas que nada brindarnos información sobre las lluvias de una manera más sencilla y entendible.

##2. Análisis de Frecuencias Método Gráfico

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

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

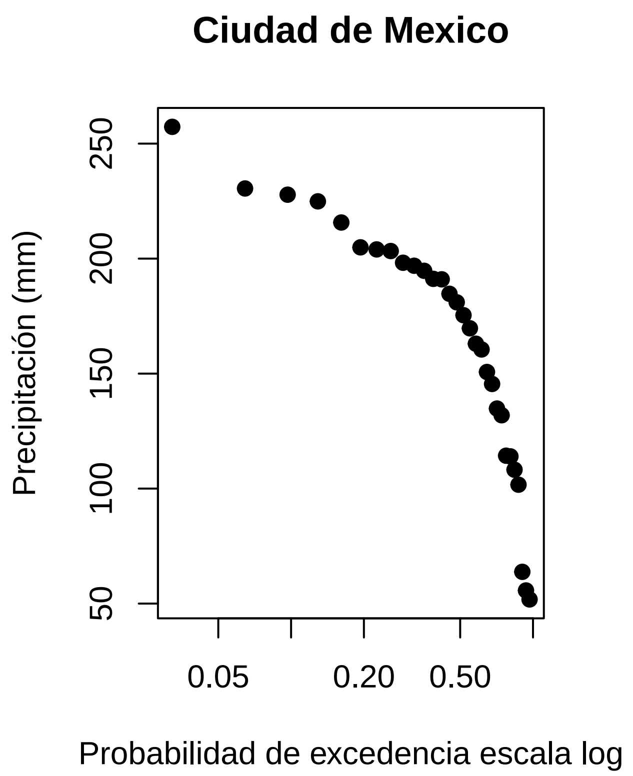
```
rain_analysis = data.frame(max_rain=monthly_max, order_max_rain=sort(monthly_max, decreasing=T)) #creac
rain_analysis$rank_rain = seq(1, nrow(rain_analysis)) #columna del rank (número de orden)
rain_analysis$Pexe = rain_analysis$rank_rain/(nrow(rain_analysis)+1) #Probabilidad de excedencia: rank/
rain_analysis$Pnoexe <- 1-rain_analysis$Pexe #Probabilidad de no excedencia: 1-Pexe, P(Xx)
```

```r
rain_analysis$Pret = 1/rain_analysis$Pexe #Periodo de retorno
head(rain_analysis) #muestra los primeros 10 renglones del data frame
```

```
##      max_rain order_max_rain rank_rain       Pexe    Pnoexe      Pret
## 1994    224.9          257.3         1 0.03225806 0.9677419 31.000000
## 1995    191.2          230.5         2 0.06451613 0.9354839 15.500000
## 1996    134.8          227.8         3 0.09677419 0.9032258 10.333333
## 1997    204.9          224.9         4 0.12903226 0.8709677  7.750000
## 1998    257.3          215.7         5 0.16129032 0.8387097  6.200000
## 1999    204.0          204.9         6 0.19354839 0.8064516  5.166667
```

Se calcula la probabilidad de no excedencia para cada precipitación (complemento de la probabilidad de excedencia): Se calcula el periodo de retorno como el inverso de la probabilidad de excedencia:

```r
par(mfrow=c(1,2))
plot(y=rain_analysis$order_max_rain, x=rain_analysis$Pexe, log="x",pch=19, main="Ciudad de Mexico", xlab
plot(x=rain_analysis$Pret, y=rain_analysis$order_max_rain, pch=19, main="Ciudad de Mexico", xlab="Períod
```



Describe las gráficas obtenidas. ¿Qué significa la probabilidad de excedencia? ¿Qué significa el periodo de retorno? ¿Por qué es importante en hidrología? ¿Qué valores son deseables en la probabilidad de excedencia para una precipitación de diseño de una obra? Puedes consultar el siguiente video de apoyo: https://www.youtube.com/watch?v=WXSIIFcsAFELinks to an external site.

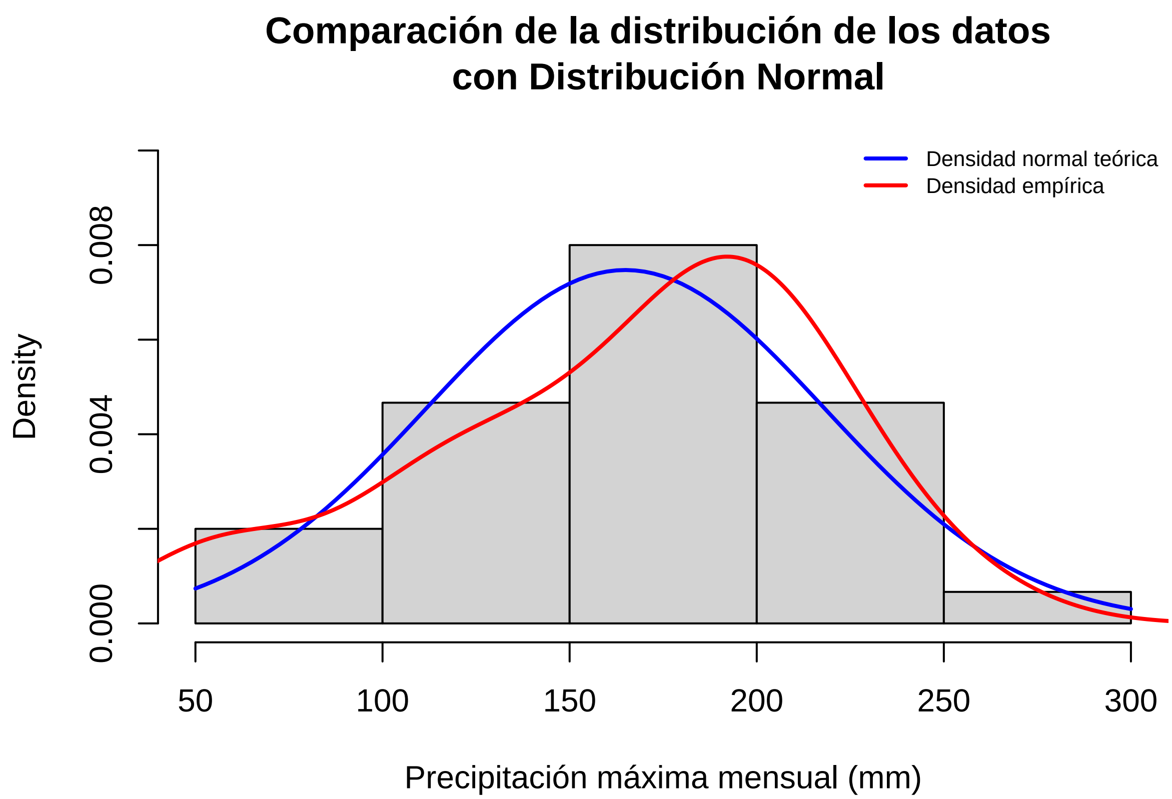
La probabilidad de excedencia nos dice la posibilidad que haya una precipitación igual o mayor a un valor específico durante el año, en donde es fácil ver que es más rara en cantidades grandes, por otra parte, el periodo de retorno representa el tiempo en promedio entre precipitaciones de igual o mayor magnitud, que al igual que en la de excedencia, nos dice que hay un mayor tiempo de espera entre periodos intensos de lluvia. Estos conceptos son clave en hidrología, ya que permiten diseñar infraestructuras capaces de soportar eventos de alta severidad y baja frecuencia, reduciendo los riesgos de inundaciones o fallos estructurales.

##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.

##Ajuste a una Distribución Normal. Hay dos maneras de determinar si un conjunto de datos tiene una distribución normal, una visual y la otra mediante una prueba de bondad de ajuste. Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución normal que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos.

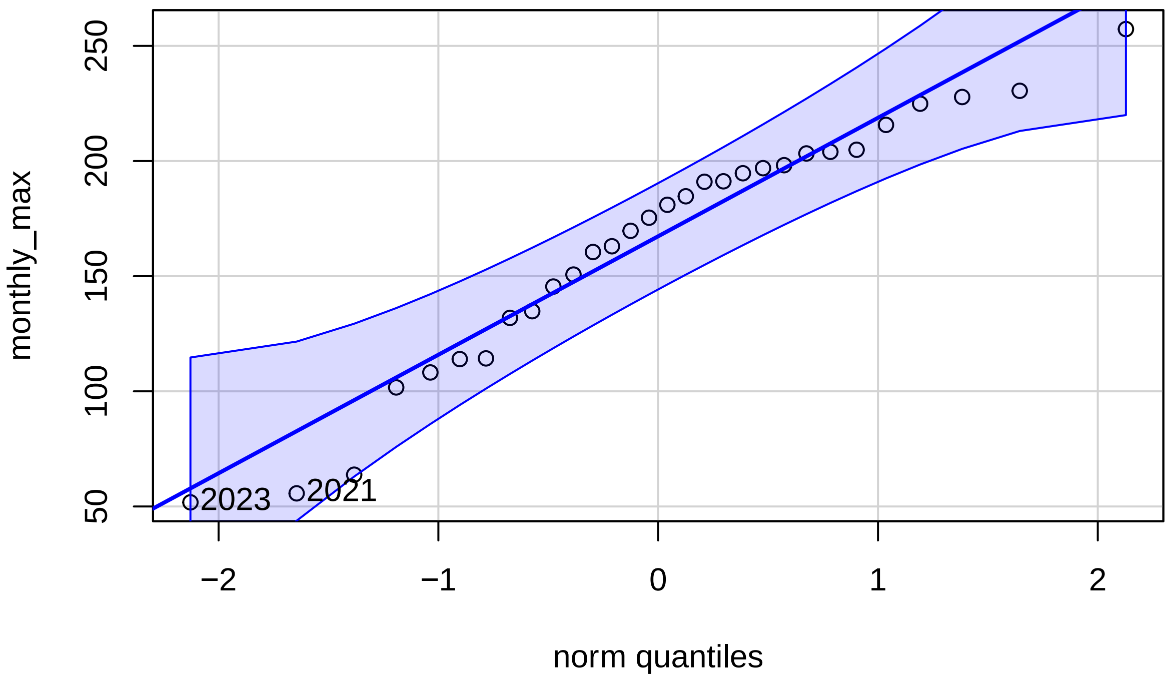
```
#Histograma de densidad empírica superpuesto a la distribución teórica
hist(monthly_max, xlab="Precipitación máxima mensual (mm)", freq=FALSE, ylim=c(0, 0.01), main="Comparaci
curve(dnorm(x, mean=mean(monthly_max), sd=sd(monthly_max)), add=TRUE, col="blue",lwd=2) #Estimación de p
lines(density(monthly_max), col="red", lwd=2)
legend("topright", col=c("blue", "red"), legend =c("Densidad normal teórica","Densidad empírica"),lwd=2
```




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

```
## Loading required package: carData
```

```r
qqPlot(monthly_max)
```

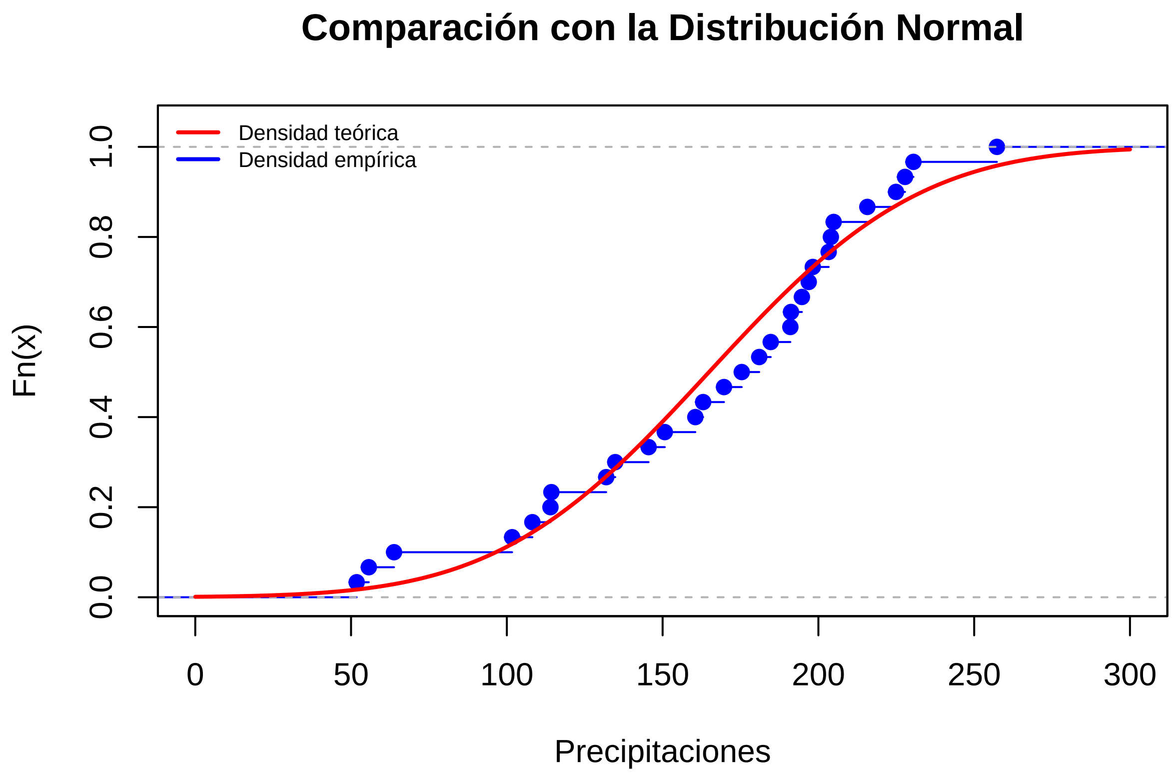



```
## 2023 2021
##   30   28
```

```r
print("QQPLOT")
```

```
## [1] "QQPLOT"
```

```r
#Probabilidad acumulada empírica vs teórica
norm_teorica = pnorm(0:300, mean=mean(monthly_max), sd=sd(monthly_max)) #Estimación de parámetros por m
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),xaxt="n
legend("topleft", col=c("red", "blue"), legend =c("Densidad teórica","Densidad empírica"),lwd=2, bty =
```




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

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

```r
#Prueba de normalilidad de Kolmogorov Smirnov
library(MASS) #Si es la primera vez que la usas, tienes que instalarla: Tools/Install Packages
ks.test(monthly_max, "pnorm", mean=mean(monthly_max), sd=sd(monthly_max)) #usa "plnorm", "pexp", "pgamm
```

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

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?

Los datos no logran ajustarse completamente a la distribución normal, ya que muestra una mayor densidad en los extremos y una menor densidad en el centro en comparación con la curva normal teórica.

Construye la gráfica qqplot. De manera visual, ¿Los datos siguen una distribución normal de acuerdo con la Q-Qplot? Casi todos los daots siguen la distribucion normal, a excepcion de de un par que se pondrian considerar un poco atipicos.

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?

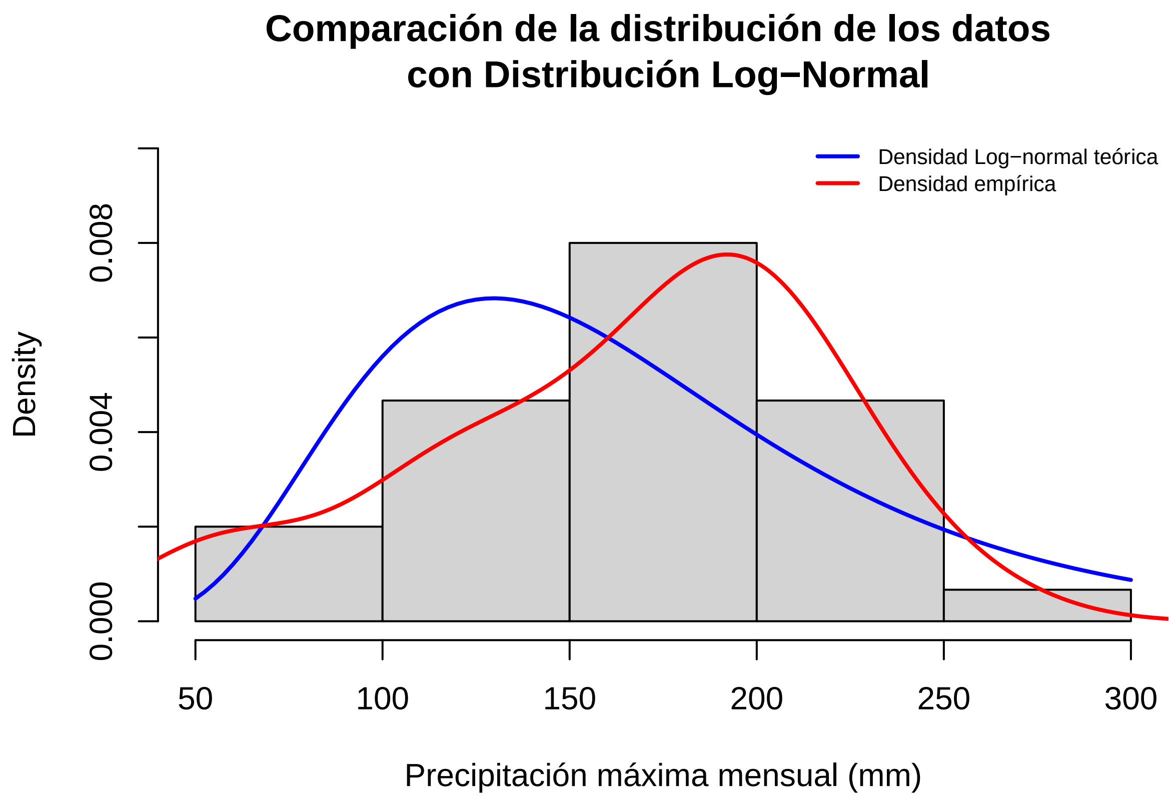
Los datos empíricos son los observados directamente de la muestra, mientras que los datos teóricos provienen de un modelo ajustado a los datos. Las distribuciones de probabilidad acumuladas se parecen bastante, pero hay pequeñas diferencias, especialmente en los extremos, lo que indica que los datos observados no siguen perfectamente la distribución normal ajustada, aunque se aproximan bien en general.

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

Los resultados de las pruebas nos devuelven p-values altos (Shapiro = 0.1552, Ks = 0.7285), lo que implica que no podemos rechazar la hipótesis nula (H0) en ninguno de los casos, gracias a lo anterior, podemos concluir que los datos son consistentes con una distribución normal, todo esto significa que, aunque visualmente pueda haber algunas diferencias en las gráficas, las pruebas estadísticas sugieren que la normalidad no puede descartarse.

##Ajuste a una Distribución Log-Normal. Realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Log-normal para ajustar los datos. Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución Log-normal que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. De manera visual, ¿te parece que los datos se ajustan bien a una distribución Log-normal? Explica. Compara las distribuciones de probabilidad acumuladas (ojiva) empíricos y teóricos de la distribución teórica que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos.

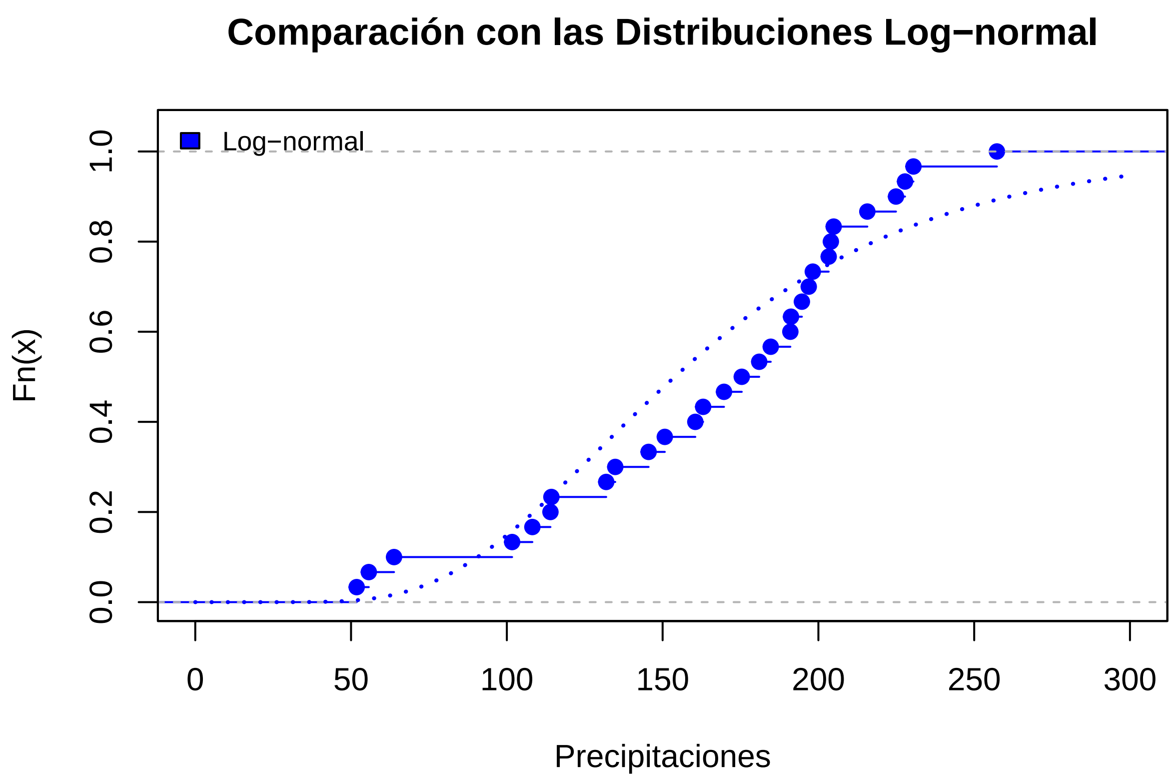
```r
#Histograma de densidad superpuesto a la distribución teórica
hist(monthly_max, xlab="Precipitación máxima mensual (mm)", freq=FALSE, ylim=c(0, 0.01), main="Comparac
curve(dlnorm(x, mean=mean(log(monthly_max)), sd=sd(log(monthly_max))), add=TRUE, col="blue",lwd=2) #Est
lines(density(monthly_max), col="red", lwd=2)
legend("topright", col=c("blue", "red"), legend =c("Densidad Log-normal teórica","Densidad empírica"),lw
```




```r
#Prueba de normalilidad de Kolmogorov Smirnov
library(MASS) #Si es la primera vez que la usas, tienes que instalarla: Tools/Install Packages
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 = 0.1738, p-value = 0.2902
## alternative hypothesis: two-sided
```

```r
x_vals <- 0:300  # Rango para el eje X
log_teorica <- plnorm(x_vals, meanlog=mean(log(monthly_max)), sdlog=sd(log(monthly_max)))
#Probabilidad acumulada empírica vs téorica
plot(ecdf(monthly_max), main="Comparación con las Distribuciones Log-normal", xlab="Precipitaciones", co
par(new=TRUE)
plot(0:300, log_teorica, type="l", main="", xlab="", ylab="", col="blue", lwd=2, ylim=c(0, 1.05), xaxt=
par(new=TRUE)
#Leyenda
legend("topleft", legend=c( "Log-normal"), fill = c( "blue", "darkgreen", "darkorange", "cyan", "magenta
```



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

Los datos tienen similitudes, más no se ajustan completamente, mientras que la empírica tiene un pico cerca de 140 mm, la teórica, tiene el pico en 200mm, la forma es similar, se puede observar que no se ajustan perfectamente.

2.-Explica qué son datos empíricos y datos teóricos. ¿Se parecen las distribuciones de probabilidad acumuladas?

Los datos empíricos son aquellos obtenidos de las observaciones reales, es decir, mediciones reales, tales como las precipitaciones máximas del archivo, mientras que los datos teóricos son aquellos que provienen de un modelo matemático, en donde se sigue una teoría matemática para ajustar los datos observados, aunque se aparecen los datos con una distribución log normal, estas no coinciden exactamente. Las diferencias visuales, especialmente en la forma de la curva en los valores altos y bajos.

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

Podemos concluir que no hay suficiente evidencia para afirmar que los datos no siguen una distribución log-normal, con los valores de D = 0.1738, p-value = 0.2902, con lo que podemos decir que no se rechaza la hipótesis nula, lo no significa que los datos sigan perfectamente una distribución log-normal, sino que la prueba KS no encuentra diferencias significativas entre los datos y la distribución log-normal bajo el umbral de significancia elegido.

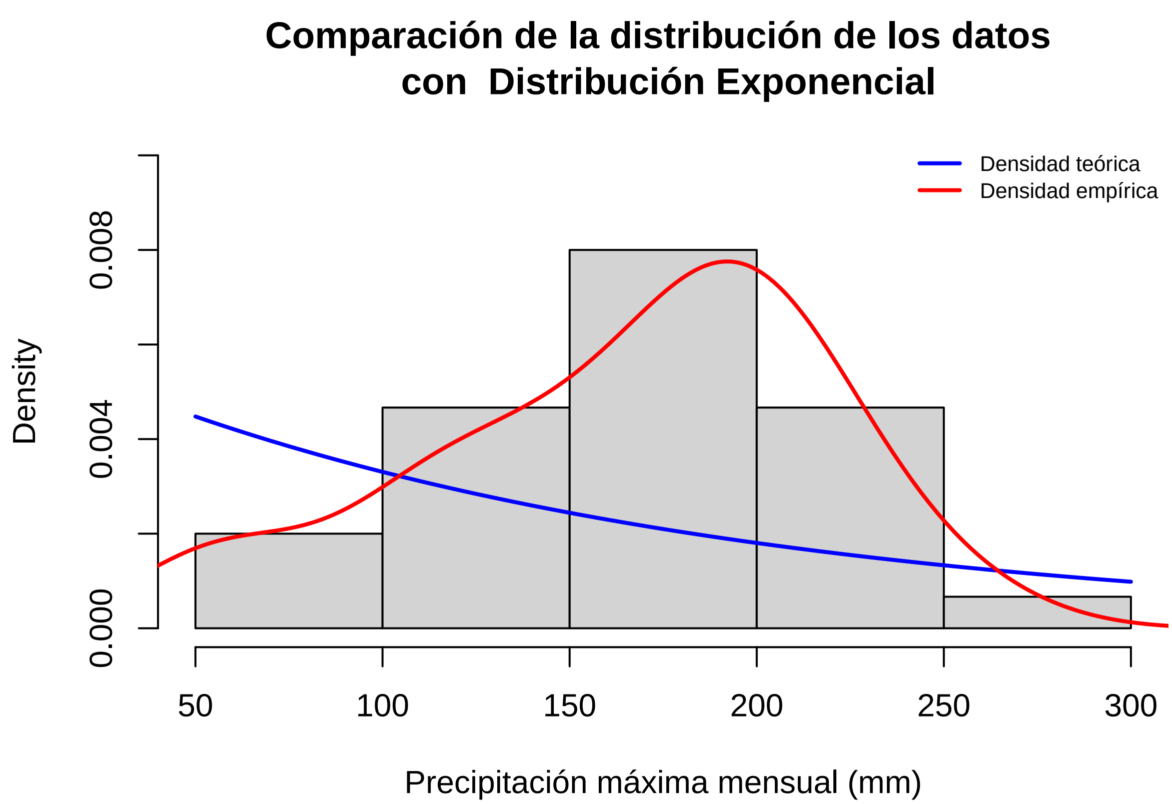
4.-¿Cuántos parámetros tiene la distribución Log-normal? ¿Cuáles son? ¿Por qué los parámetros se calculan de la forma en cómo se hace en el código? Sigue paso a paso el método de momentos para demostrar que los parámetros están bien calculados.

La distribución log-normal tiene dos parámetros Mu (media del logaritmo natural de datos) y sigma (Desviación estándar del logaritmo natural de los datos) , la razón del porqué se calculan de esta manera, utilizando el modelo de máxima verosimilitud, en el cual consiste de igualar los datos muestrales con los teóricos en la distribución.

##Ajuste a una Distribución Exponencial. Realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Exponencial para ajustar los datos. Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución Exponencial que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos.

```
#Histograma de densidad superpuesto a la distribución teórica
hist(monthly_max, xlab="Precipitación máxima mensual (mm)", freq=FALSE, ylim=c(0, 0.01), main="Comparac


curve(dexp(x, 1/mean(monthly_max)), add=TRUE, col="blue",lwd=2) #Estimación de parámetros por método de
lines(density(monthly_max), col="red", lwd=2)
legend("topright", col=c("blue", "red"), legend =c("Densidad teórica","Densidad empírica"),lwd=2, bty =
```



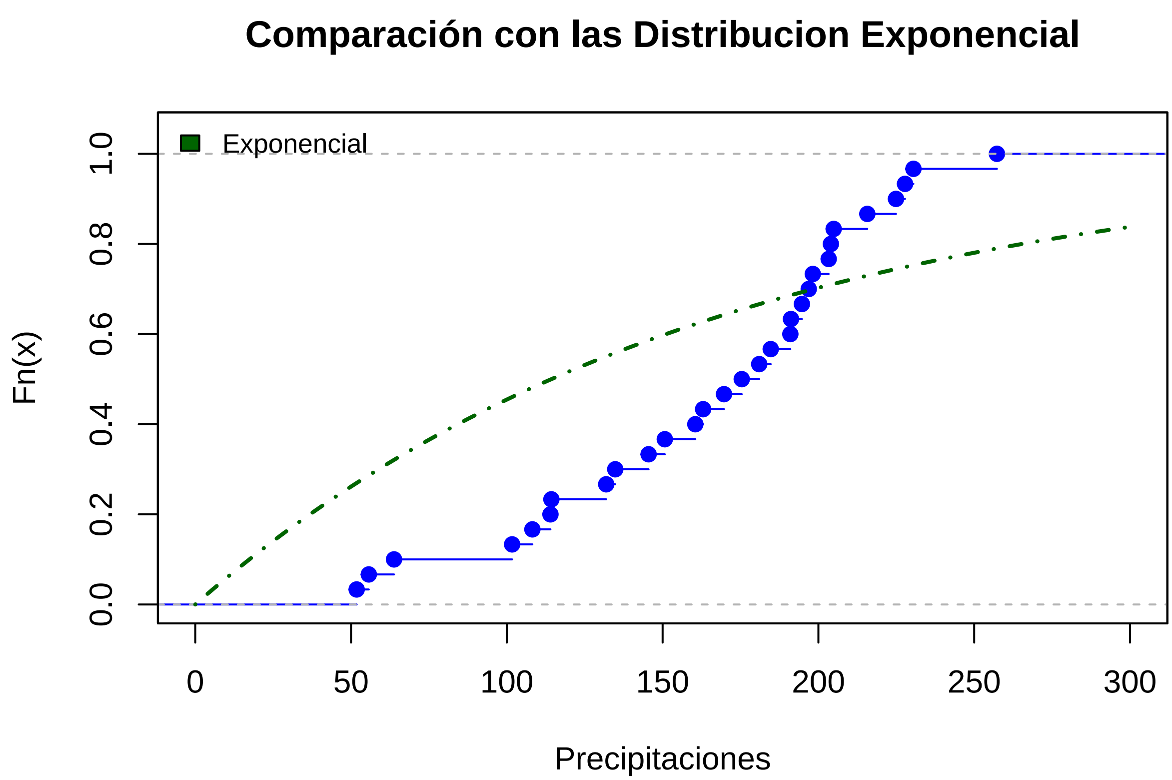

```
#Prueba de normalildad de Kolmogorov Smirnov
ks.test(monthly_max, "pexp", rate=1/mean(monthly_max))
```

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

```
## alternative hypothesis: two-sided
```

```
x_vals <- 0:300  # Rango para el eje X
exp_teorica <- pexp(x_vals, rate=1/mean(monthly_max))

#Probabilidad acumulada empírica vs téorica
plot(ecdf(monthly_max), main="Comparación con las Distribucion Exponencial", xlab="Precipitaciones", co
par(new=TRUE)
plot(0:300, exp_teorica, type="l", main="", xlab="", ylab="", col="darkgreen", lwd=2, ylim=c(0, 1.05), 
par(new=TRUE)
legend("topleft", legend=c("Exponencial"), fill = c("darkgreen"), cex = 0.8, bty="n")
```



1.-De manera visual, ¿te parece que los datos se ajustan bien a una distribución exponencial? Explica. Se puede observar que no se ajustan bien, en donde la empírica tiene un pico, mientras que la exponencial sólo desciende desde el inicio, es evidente el desajuste.

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?

Los datos empíricos son aquellos obtenidos de las observaciones reales, es decir, mediciones reales, tales como las precipitaciones máximas del archivo, mientras que los datos teóricos son aquellos que provienen de un modelo matemático, al comparar las distribuciones dee probabilidad acumuladad, see puede ver que no coinciden del todo. La distribución acumulada empírica probablemente tendría un aumento más abrupto que la exponencial, que sería más gradual.

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

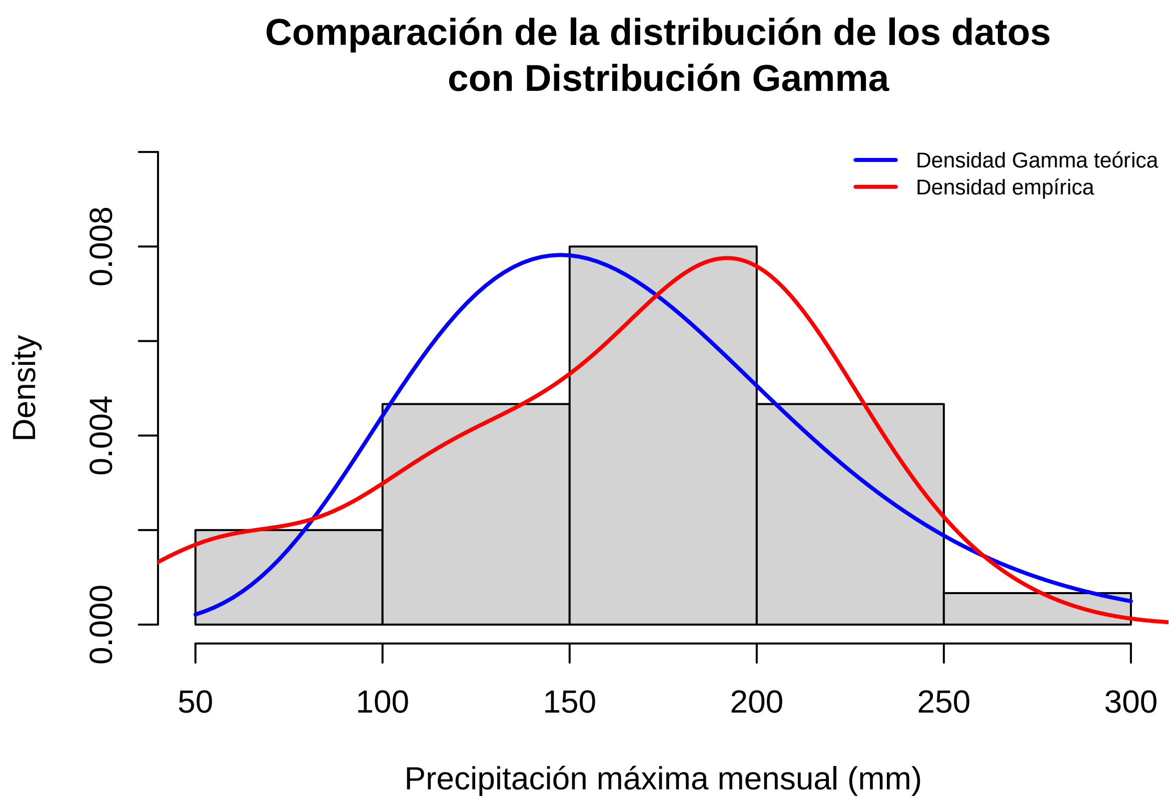
La prueba de KS nos ayuda a determinar si los datos siguen una distribución exponencial, comparando la distribución empírica con la teórica. El resultado del estadístico D es 0.36029 y el p-value correspondiente es 0.0005399. Con estos resultados, se rechaza la hipótesis nula, que sostiene que los datos siguen una distribución exponencial, ya que tanto el valor de D como el p-value indican una diferencia significativa entre las distribuciones.

4.-¿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.

Los dos parámetros de gama son Alfa (Forma de distribución) y beta (Rango de valores), estos parámetros se calculan utilizando el método de momentos o máxima verosimilitud, en donde se igualan los valores muestrales con los teóricos.

##Ajuste a una Distribución Gamma. Realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Gamma para ajustar los datos. Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución Gamma que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos.

```
#Histograma de densidad superpuesto a la distribución teórica
hist(monthly_max, xlab="Precipitación máxima mensual (mm)", freq=FALSE, ylim=c(0, 0.01), main="Comparaci
curve(dgamma(x, mean(monthly_max)^2/var(monthly_max), mean(monthly_max)/var(monthly_max)), add=TRUE, col
lines(density(monthly_max), col="red", lwd=2)
legend("topright", col=c("blue", "red"), legend =c("Densidad Gamma teórica","Densidad empírica"),lwd=2,
```



```r
#Prueba de normalildad de Kolmogorov Smirnov
ks.test(monthly_max, "pgamma", shape = mean(monthly_max)^2 / var(monthly_max), scale = var(monthly_max)
```

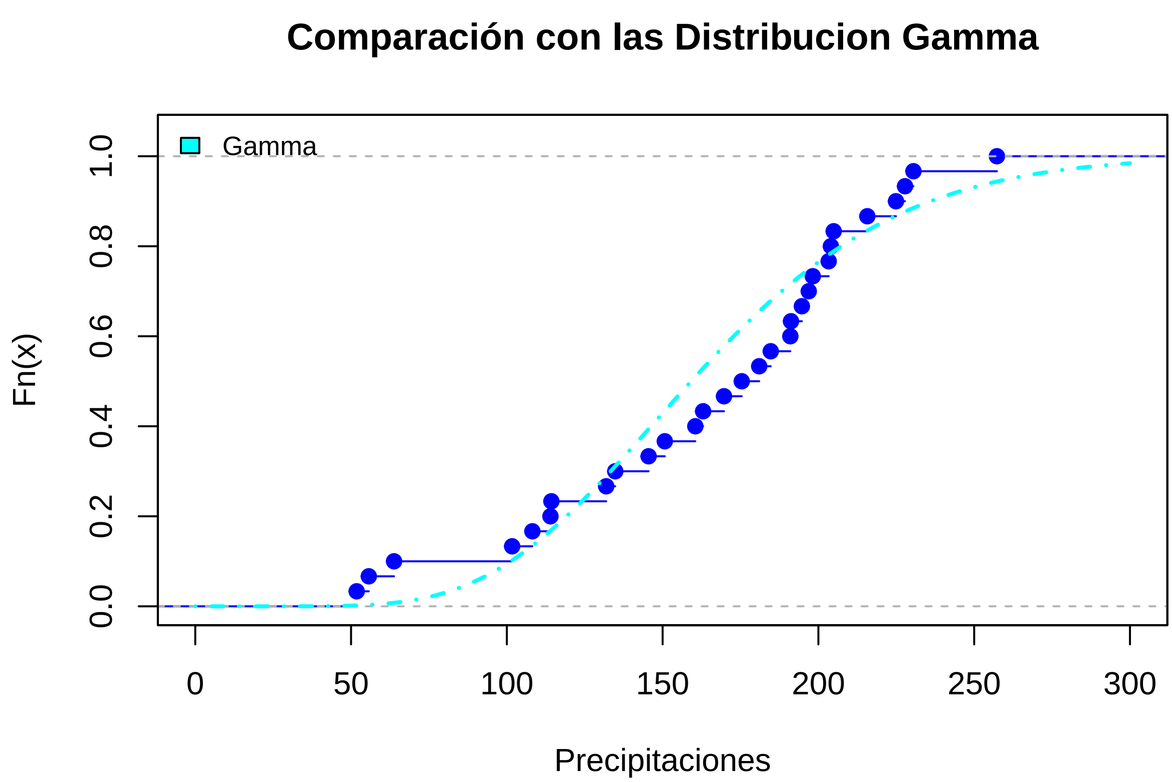
```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  monthly_max
## D = 0.15526, p-value = 0.4219
## alternative hypothesis: two-sided
```

```r
x_vals <- 0:300  # Rango para el eje X
gamma_teorica <- pgamma(x_vals, shape=mean(monthly_max)^2/var(monthly_max), rate=mean(monthly_max)/var(m

#Probabilidad acumulada empírica vs téorica
plot(ecdf(monthly_max), main="Comparación con las Distribucion Gamma", xlab="Precipitaciones", col="blue
par(new=TRUE)

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

legend("topleft", legend=c("Gamma"), fill = c("cyan", "magenta"), cex = 0.8, bty="n")
```



De manera visual, ¿te parece que los datos se ajustan bien a una distribución Gamma? Explica. Visualmente, ambos datos tienen similitudes, mas no son iguales, esto debido a que la empirica tiene un pico alrededor de 200 mm, mientras que la gama teorica tiene siu pico en 150 mm.

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?

Los datos empíricos son aquellos obtenidos de las observaciones reales, tales como las precipitaciones máximas del archivo, mientras que los datos teóricos son aquellos que provienen de un modelo matemático, en donde se sigue una teoría matemática para ajustar los datos observados, aunque se aparecen los datos con una distribución Gamma, estas no coinciden exactamente, especificamente en los picos de cada uno.

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

La prueba de KS nos ayuda a determinar si los datos siguen una distribución Gamma, comparando la distribución empírica con la teórica. El resultado del estadístico D es 0.15526, y el p-value correspondiente es 0.4219. Con estos resultados, se puede decir que no se rechaza la hipótesis nula, que establece que los datos siguen una distribución Gamma, gracias al valor de P, se puede decir que no hay evidencia suficiente para concluir que existe una diferencia significativa entre las distribuciones empírica y teórica.

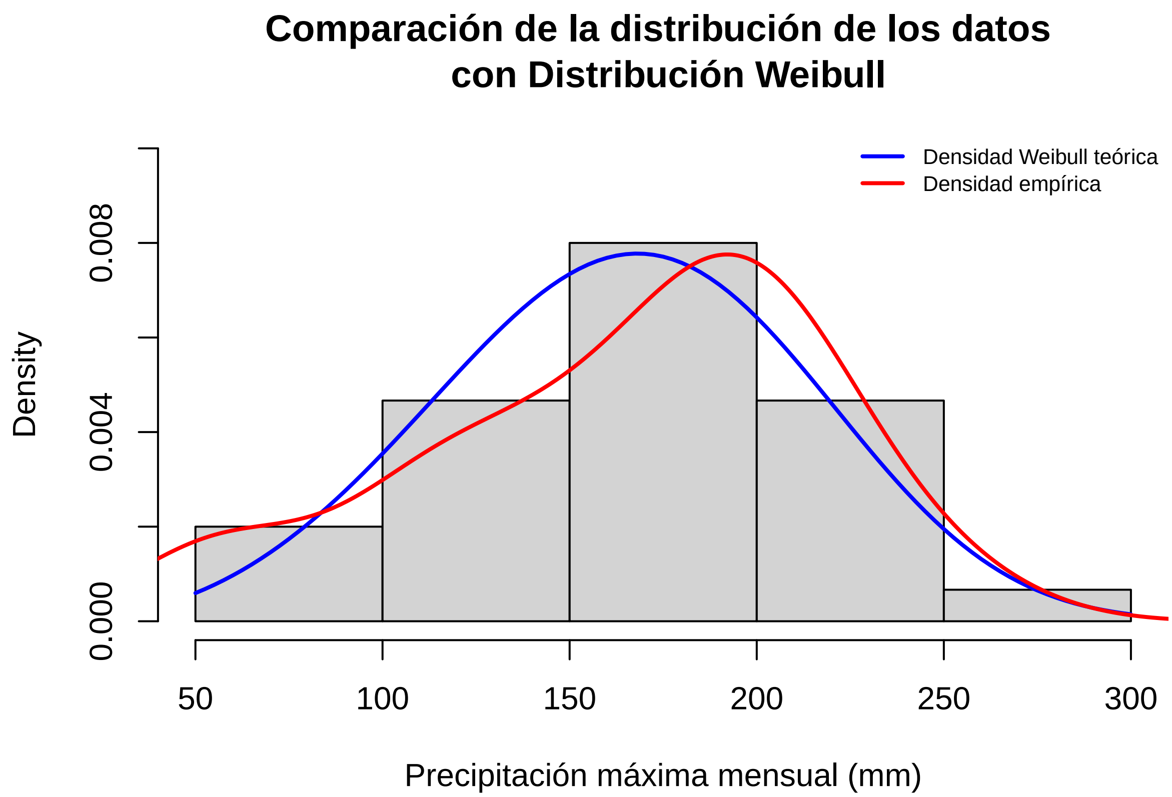
¿Cuántos parámetros tiene la distribución Gamma? ¿Cuáles son? ¿Por qué los parámetros se calculan de la forma en cómo se hace en el código? Sigue paso a paso el método de momentos para demostrar que los parámetros están bien calculados.

Tiene dos parametros, forma y escala,se calculan utilizando el método de momentos, que ajusta la distribución teórica a los momentos (media y varianza) de los dato, la forma se coonsigue media al cuadrado sobre varianza, mientras que escala se obtiene de varianza sobre medoia

##Ajuste a una Distribución Weibull. Realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Weibull para ajustar los datos. El cálculo de los parámetros a partir de los datos es un poco más difícil en la distribución Weibull de lo que fue en las anteriores distribuciones, así que recurriremos a que R los estime con el comando fitdistr. Úsalo para estimar los parámetros de la Weibull. Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución Weibull que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos.

```r
#Cálculo de los parámetros
weibull_fit <- fitdistr(monthly_max, "weibull", lower=c(0,0)) #fitdistr da los parámetros calculados a p

#Histograma de densidad superpuesto a la distribución teórica
hist(monthly_max, xlab="Precipitación máxima mensual (mm)", freq=FALSE, ylim=c(0, 0.01), main="Comparaci
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
```



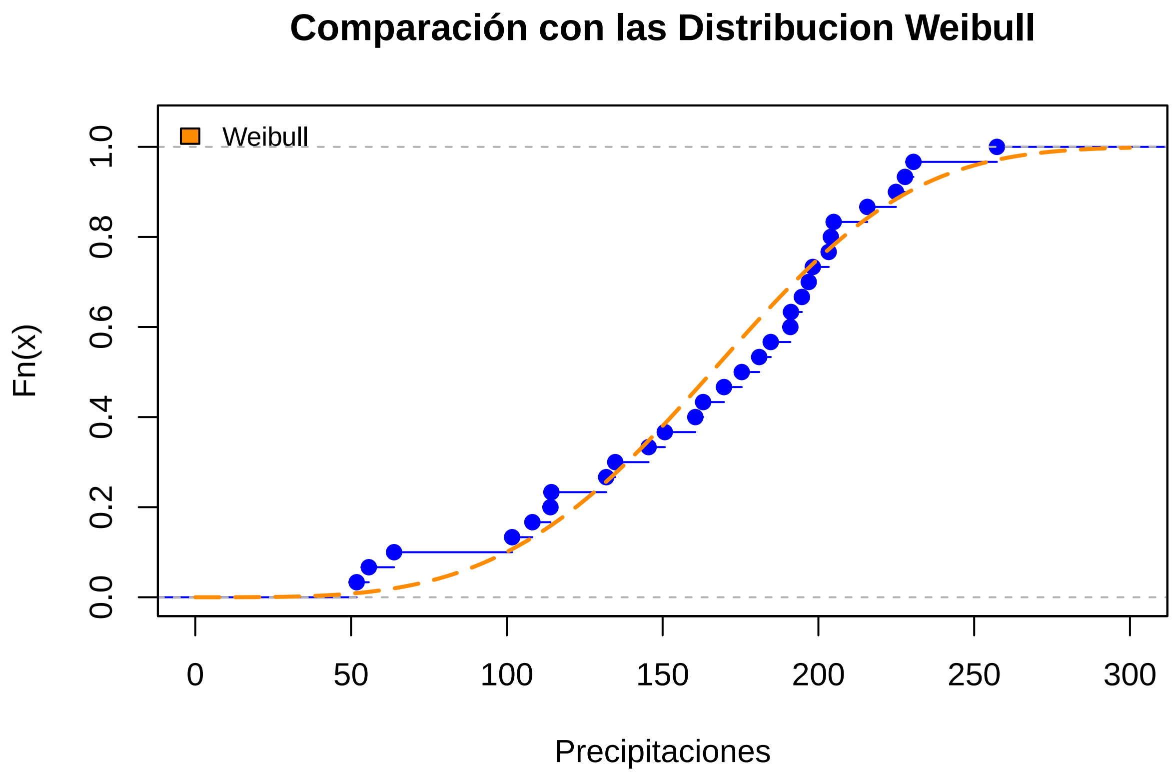

```
#Prueba de normalildad de Kolmogorov Smirnov
ks.test(monthly_max, "pweibull", shape = weibull_fit$estimate[1], scale = weibull_fit$estimate[2])
```

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

```
x_vals <- 0:300  # Rango para el eje X
weibull_teorica <- pweibull(x_vals, shape=weibull_fit$estimate[1], scale=weibull_fit$estimate[2])

#Probabilidad acumulada empírica vs téorica
plot(ecdf(monthly_max), main="Comparación con las Distribucion Weibull", xlab="Precipitaciones", col="bl
par(new=TRUE)

plot(0:300, weibull_teorica, type="l", main="", xlab="", ylab="", col="darkorange", lwd=2, ylim=c(0, 1.(
par(new=TRUE)
legend("topleft", legend=c("Weibull"), fill = c("darkorange"), cex = 0.8, bty="n")
```



De manera visual, ¿te parece que los datos se ajustan bien a una distribución Weibull? Explica. La distribucion wibul y la empirica no se sobreponen, mas se puede decir que sus picos estan muy cerca uno del otro.

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?

Los datos empíricos son aquellos obtenidos de las observaciones reales, tales como las precipitaciones máximas del archivo, mientras que los datos teóricos son aquellos que provienen de un modelo matemático, en donde se sigue una teoría matemática para ajustar los datos observados, aunque se aparecen los datos con una distribución Weibull, aunque sus picos se encuentren muy cerca, aun asi no se consideran que se sobreponen uno sobre el otro.

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

Ks nos ayuda a determinar si los datos se ajustan a una distribución Weibull, en este caso, el valor del estadístico de prueba D es 0.12421 y el valor de P es 0.6977. Dado que el p es considerablemente mayor que 0.05, se dice que no se rechaza la hipótesis nula, lo que indica que no hay suficiente evidencia para concluir que los datos difieren de una distribución Weibull. Por lo tanto, ppodemos decir que los datos podrían seguir una distribución Weibull, ya que no hay diferencias significativas entre la distribución teórica y la empírica según la prueba KS.

¿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 parámetros: el parámetro de forma y escala . La estimación de los parámetros en la distribución Weibull es más compleja que en otras distribuciones, como la Gamma o la Exponencial, debido a que los momentos (media y varianza) no tienen una relación directa y simple con los parámetros. Además, la función de verosimilitud para la Weibull no tiene una solución analítica directa, para estimar los parámetros. Esto contrasta con la distribución Gamma, cuyos parámetros pueden ser estimados directamente a partir de la media y varianza de los datos usando el método de momentos.

##Ajuste a una Distribución Gumbel. Realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Gumbel para ajustar los datos. Para probar si los datos de precipitación máxima se ajustan a una distribución Gumbel, se necesita definir las funciones de densidad de acuerdo con la función Gumbel. Creálas con las fórmulas de la Distribución Gumbel. Para estimar los parámetros y hacer el ajuste de la Distribución Gumbel con la biblioteca "fitdistrplus". Haz las gráficas de histograma de densidad empírica y teórica, la probabilidad de acumulada empírica y teórica y el QQplot. Haz la prueba KS para determinar si los datos se ajustan a una Gumbel

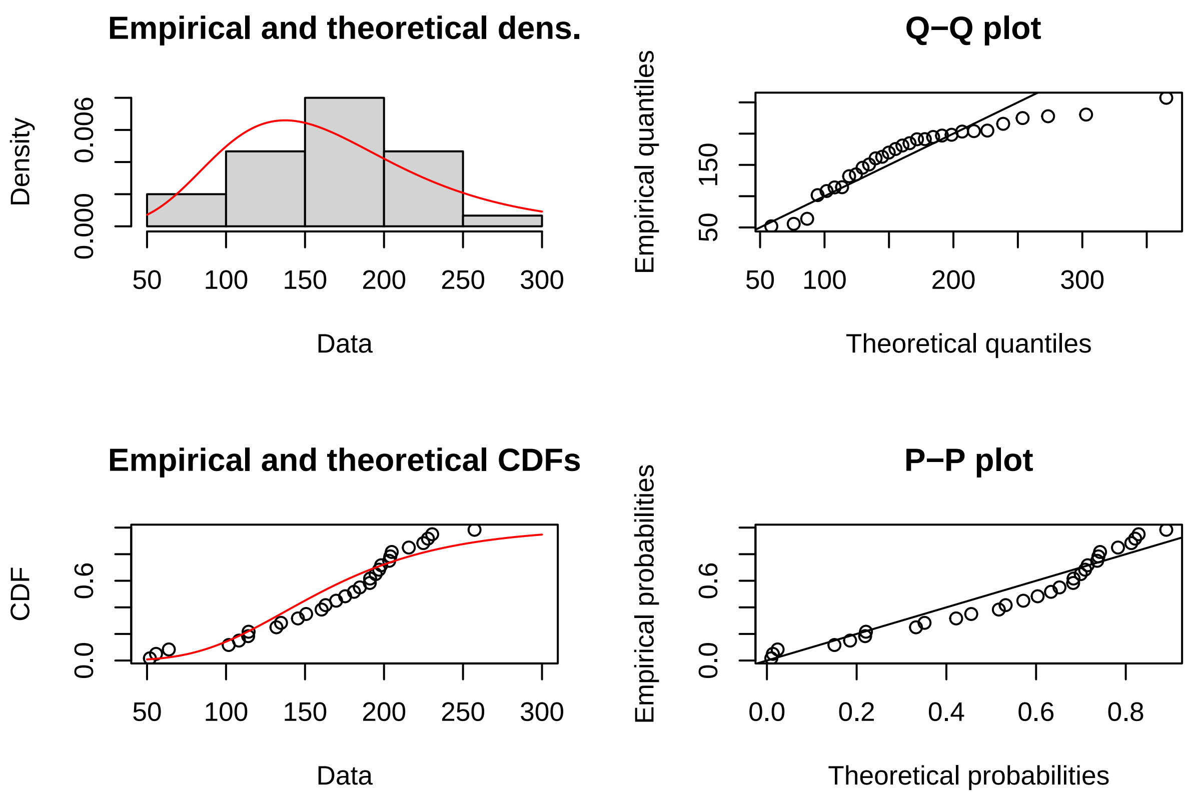
```r
#Para crear la función de cálculo de probabilidades con la Gumbel (no está en las bibliotecas ni en la
dgumbel = function(x, a, b) 1/b*exp((a-x)/b)*exp(-exp((a-x)/b))
pgumbel = function(q, a, b) exp(-exp((a-q)/b))
qgumbel = function(p, a, b) a-b*log(-log(p))

#Estimación de los parámetros de la Distribución Gumbel
library(fitdistrplus) #si es la primera vez que lo usas, necesitas instalarlo: Tools/Install Packages,
```

```
## Loading required package: survival
```

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

#Estimación de los parámetros de la Distribución Gumbel y gráficos asociados
library(fitdistrplus) #si es la primera vez que lo usas, necesitas instalarlo: Tools/Install Packages,
gumbel_fit = fitdist(monthly_max, "gumbel", start = list(a=1, b=1))
plot(gumbel_fit)
```

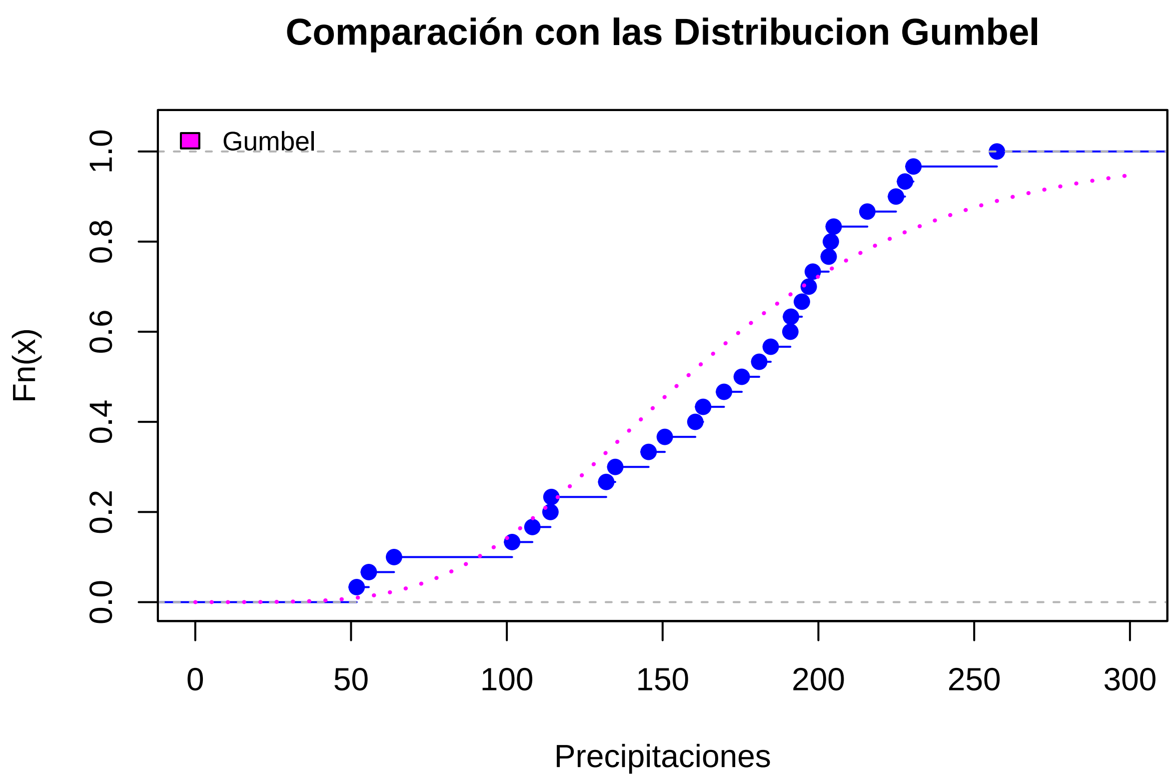



```
#Prueba de Kolmogorov-Smirnov
gumbel_exe = 1-pgumbel(rain_analysis$order_max_rain, gumbel_fit$estimate[1], gumbel_fit$estimate[2]) #C
ks.test(rain_analysis$Pexe, gumbel_exe)
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  rain_analysis$Pexe and gumbel_exe
## D = 0.16667, p-value = 0.808
## alternative hypothesis: two-sided
```

```
x_vals <- 0:300  # Rango para el eje X

plot(ecdf(monthly_max), main="Comparación con las Distribucion Gumbel", xlab="Precipitaciones", col="blu
par(new=TRUE)
gumbel_teorica = pgumbel(0:300, gumbel_fit$estimate[1], gumbel_fit$estimate[2])
plot(0:300, gumbel_teorica, type="l", main="", xlab="", ylab="", col="magenta", lwd=2, ylim=c(0, 1.05),
#Leyenda
legend("topleft", legend=c("Gumbel"), fill = c( "magenta"), cex = 0.8, bty="n")
```



¿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?

KS para una distribución Gumbel nos da información sobre si las probabilidades de excedencia de las precipitaciones máximas mensuales siguen esta distribución. En este caso, el valor del estadístico de prueba es 0.16667 y el valor de p es 0.808, gracias al valor p, no se rechaza la hipótesis nula. Esto significa que no hay evidencia suficiente para decir que las probabilidades de excedencia no siguen una distribución Gumbel. Por lo tanto, podemos concluir que los datos podrían ajustarse a una distribución Gumbel.

¿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. Los dos principales parametros son mu y beta, estos dos parámetros permiten ajustar la forma de la distribución Gumbel y son esenciales para modelar fenómenos de valores extremo

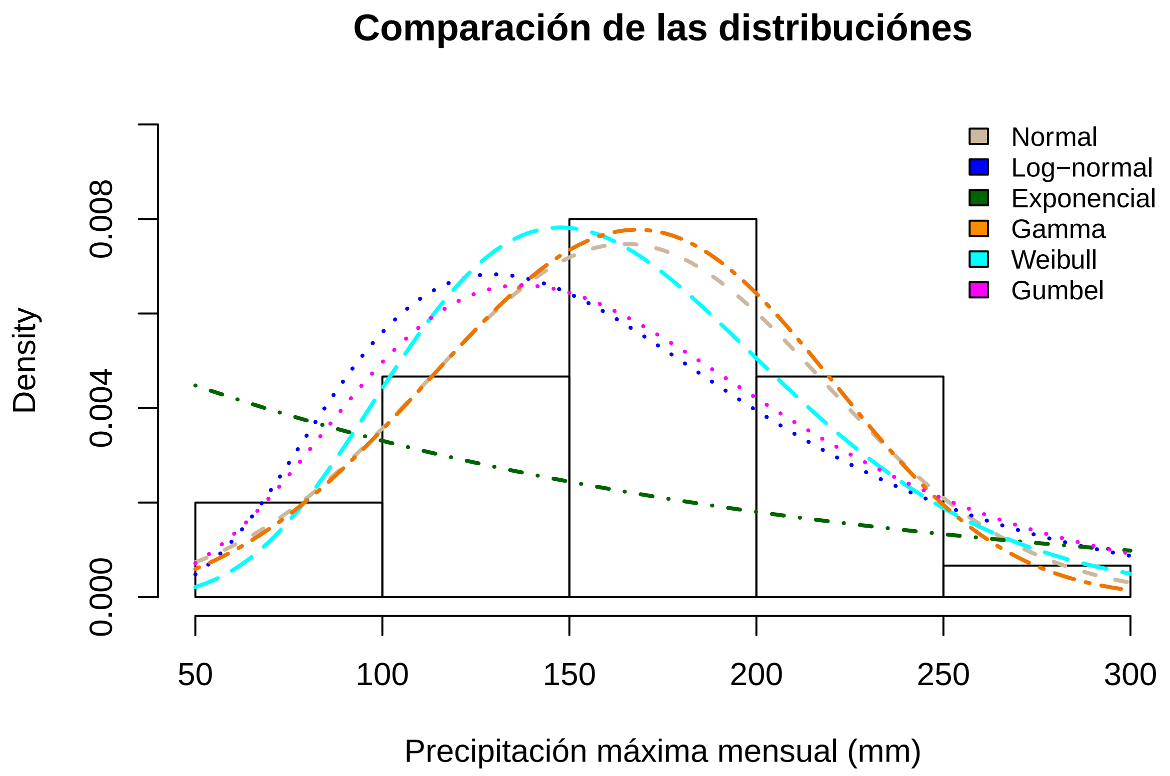
Compara los valores obtenidos con los estimados con el comando "fitdistrplus" ¿se obtienen los mismos valores por ambos métodos? ¿por qué crees que se dé esta diferencia? ¿por qué crees que se dé esta diferencia? Realemnte no hay mucha diferencia, probablemente a la manera en que analisa los datos

##Compara los ajustes de las distribuciones que analizaste Haz un gráfico comparativo de los histogramas de densidad empírica vs densidad teórica de todas las distribuciones que analizaste (todas las distribuciones en un solo gráfico. Haz un gráfico comparativo de las probabilidades acumuladas empírica vs teóricas de todas las distribuciones que analizaste (todas las distribuciones en un solo gráfico.

```
#Histograma de función de densidad empírica vs teórica
hist(monthly_max, xlab="Precipitación máxima mensual (mm)", freq=FALSE, ylim=c(0, 0.01), main="Comparaci
curve(dnorm(x, mean=mean(monthly_max), sd=sd(monthly_max)), add=TRUE, col="bisque3",lwd=2, lty=2)
```

```r
curve(dlnorm(x, mean=mean(log(monthly_max)), sd=sd(log(monthly_max))), add=TRUE, col="blue",lwd=2, lty=3
curve(dexp(x, 1/mean(monthly_max)), add=TRUE, col="darkgreen",lwd=2,lty=4)
curve(dgamma(x, mean(monthly_max)^2/var(monthly_max), mean(monthly_max)/var(monthly_max)), add=TRUE, col
curve(dweibull(x, weibull_fit$estimate[1], weibull_fit$estimate[2]), add=TRUE, col="darkorange2",lwd=2,
curve(dgumbel(x, gumbel_fit$estimate[1], gumbel_fit$estimate[2]), add=TRUE, col="magenta",lwd=2, lty=15

#Leyenda
legend("topright", legend=c("Normal", "Log-normal", "Exponencial", "Gamma", "Weibull", "Gumbel"), fill =
```

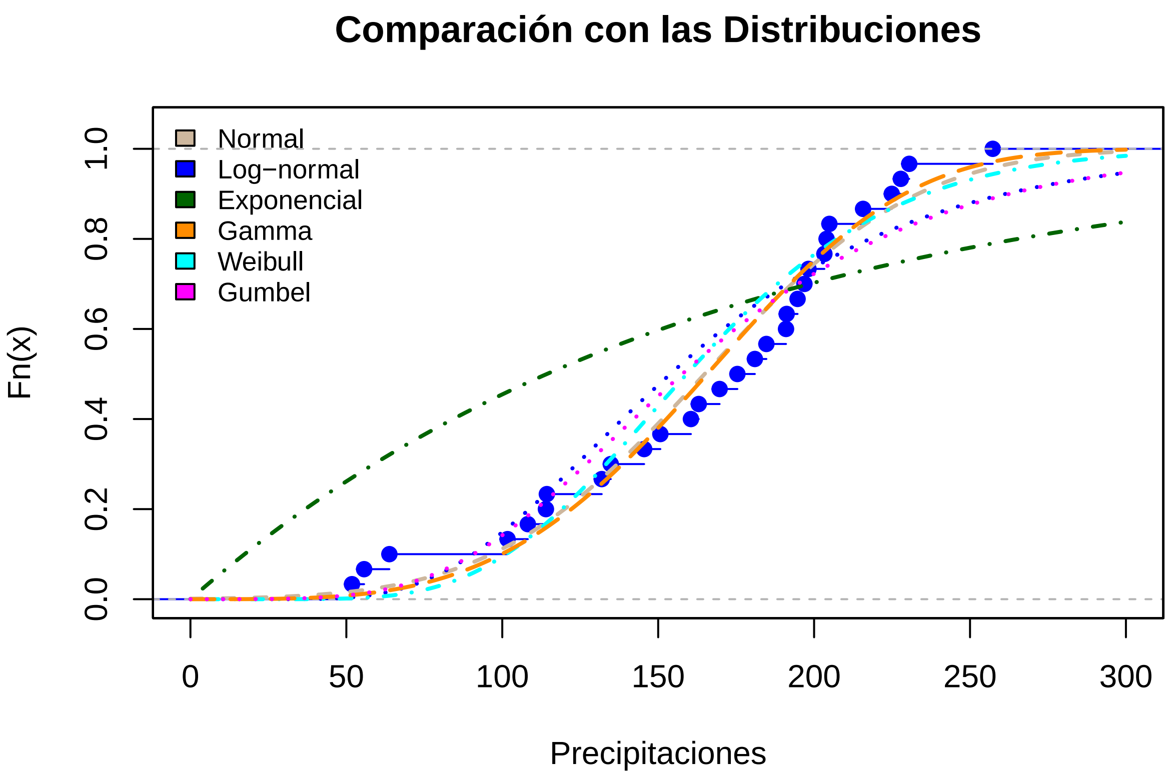



```r
x_vals <- 0:300  # Rango para el eje X

norm_teorica <- pnorm(x_vals, mean=mean(monthly_max), sd=sd(monthly_max))
log_teorica <- plnorm(x_vals, meanlog=mean(log(monthly_max)), sdlog=sd(log(monthly_max)))
exp_teorica <- pexp(x_vals, rate=1/mean(monthly_max))
gamma_teorica <- pgamma(x_vals, shape=mean(monthly_max)^2/var(monthly_max), rate=mean(monthly_max)/var(m
weibull_teorica <- pweibull(x_vals, shape=weibull_fit$estimate[1], scale=weibull_fit$estimate[2])

#Probabilidad acumulada empírica vs téorica
plot(ecdf(monthly_max), main="Comparación con las Distribuciones", xlab="Precipitaciones", col="blue", 
par(new=TRUE)
plot(0:300, norm_teorica, type="l", main="", xlab="", ylab="", col="bisque3", lwd=2, ylim=c(0, 1.05), xa
par(new=TRUE)
plot(0:300, log_teorica, type="l", main="", xlab="", ylab="", col="blue", lwd=2, ylim=c(0, 1.05), xaxt=
par(new=TRUE)
plot(0:300, exp_teorica, type="l", main="", xlab="", ylab="", col="darkgreen", lwd=2, ylim=c(0, 1.05), 
```

```
par(new=TRUE)
plot(0:300, gamma_teorica, type="l", main="", xlab="", ylab="", col="cyan", lwd=2, ylim=c(0, 1.05), xaxt
par(new=TRUE)
plot(0:300, weibull_teorica, type="l", main="", xlab="", ylab="", col="darkorange", lwd=2, ylim=c(0, 1.0
par(new=TRUE)
gumbel_teorica = pgumbel(0:300, gumbel_fit$estimate[1], gumbel_fit$estimate[2])
plot(0:300, gumbel_teorica, type="l", main="", xlab="", ylab="", col="magenta", lwd=2, ylim=c(0, 1.05),
#Leyenda
legend("topleft", legend=c("Normal", "Log-normal", "Exponencial", "Gamma", "Weibull", "Gumbel"), fill =
```



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.

Al analiizar las dos graficas, fue posible observar que la mayoria de las distribuciones aparte de la exponencial se acercaban de manera adecuada a los datos, mas al analizar los valores d y P de cada distribucion, fue posible darse cuenta que la distribucion Weibull se considera como la mejor opcion, ya que esta proporciona tanto visualmente un ajuste adecuado, como estadisticamente, en donde supera a la gamma, la cual es la segunda mejor opcion

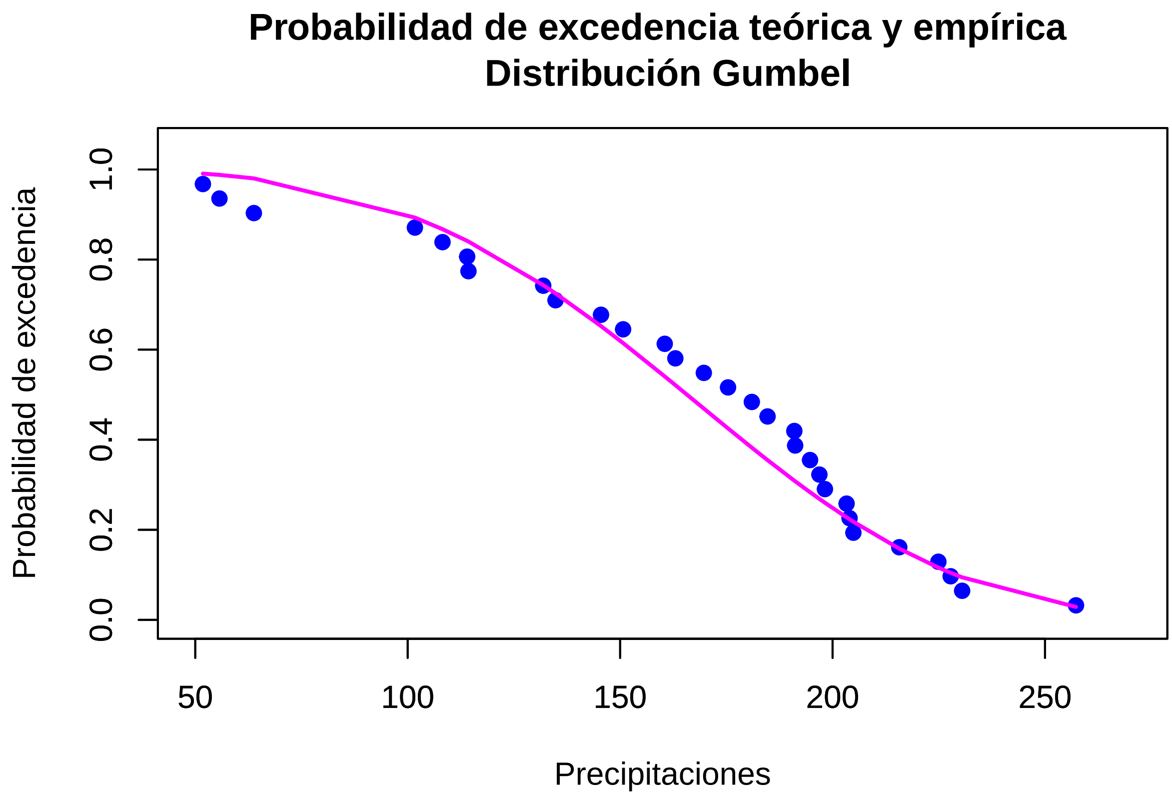
##DISEÑO DE OBRAS HIDRÁULICAS ##4. Precipitación de diseño de obras hidráulicas

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.

```
#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 empír
par(new=TRUE)
```

```r
weibull_exe <- 1 - pweibull(rain_analysis$order_max_rain, shape = weibull_fit$estimate[1], scale = weibu

plot(rain_analysis$order_max_rain, weibull_exe, type="l", main="", xlab="", ylab="", col="magenta", lwd=
```



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.

Utilizando la distribucion weibull, es posible ver que se ajusta de buena medida con los puntos azules, aunque existe una pequeña desviacion en el medio de la misma, siendoo eestas las precipitaciones alrededor de 150 a 200 mm, por lo que se puede decir que aunque captura de buena manera los demas datos, si los datos de las precipitaciones entre esos valores son de interes, esta distribucion podria no ser la mas adecuada.

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

```
##    shape
## 972.4546
```

```r
Pexe = 1/100
Pnoexe = 1 - Pexe
qgumbel(p = Pnoexe, a=weibull_fit$estimate[1], b=weibull_fit$estimate[2])
```

```
## shape
## 845.2
```

```r
Pexe = 1/50
Pnoexe = 1 - Pexe
qgumbel(p = Pnoexe, a=weibull_fit$estimate[1], b=weibull_fit$estimate[2])
```

```
##    shape
## 717.4793
```

```r
Pexe = 1/10
Pnoexe = 1 - Pexe
qgumbel(p = Pnoexe, a=weibull_fit$estimate[1], b=weibull_fit$estimate[2])
```

```
##    shape
## 415.3644
```

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

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.

El resultado de este ejemplo será una aproximación del caudal máximo que se tendría en Ciudad de mexico 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 obtenido de 972,45 nos representa el caudal maximo que se puede generar en ciudad de mexico para un periodo de 200 años, si incrementamos el periodo, es sensible comprender que este tambien aumente, al mismo tiempo que este no sera el mismo si se utilizan los datos de otro estado, debido a las precipitacionees unicas de cada ubicacion. La razon de por que es tan iiimpoortante que se tengan que diseñar en los periodos sugeridos es muy sencilla, estas impactan las estimaciones y sus frecuencias, lo que puede resultar en tomas de riesgos inecesarios de inundaciones por la lluvia,, al explorar otros periodos, nos permite observar como estos cambian dependiendo de la cantidad de años, mientras mayor numero de años, mayor cantidad de lluvia.

##ENTREGA ##5. Reporte

El reporte deberá incluir una introducción, metodología, conclusiones, interpretaciones, respuestas a las preguntas, y descripción de referencias consultadas (por lo menos 3). En el reporte podrás incluir gráficas con el objetivo de describir los resultados encontrados.

Con el trabajo realizado en los puntos 1 a 4 elabora un reporte de tu trabajo con las siguientes secciones: Introducción. Haz una introducción de a lo más media página sobre el tema. Debe tener citas de referencias confiables sobre el tema. Metodología. Resumen el análisis realizado en los puntos anteriores. Responde las preguntas e interpreta tus resultados Para argumentar tu discusión e interpretaciones, relaciona los gráficos y las medidas. No interpretes cada gráfico, cada pregunta y cada tabla de medidas por separado: Intégralas. Incluye gráficas que te ayuden a describir los resultados encontrados. Reporta las medidas (valor p y estadísticos, por ejemplo) en forma de tabla Esta parte debe tener 5 páginas (máximo 6) a espacio sencillo. Discusión y conclusiones. Incluye: Reporta una conclusión final de tu análisis tomando en cuenta que el objetivo de este trabajo: calcular la precipitación más extrema que se logra con un periodo de retorno seleccionado basándote en la bibliografía. Esta parte debe tener 1 página (máximo página y media) Referencias bibliográficas en formato APA (al menos 3) Anexo: Adicionalmente deberás entregar un documento de Rmarkdown en formato "pdf" y "Rmd" en donde incluyas código, figuras y comentarios. Para elaborar este reporte considera: El reporte no debe contener código de R. En el primer chunck de R de tu documento, cambia echo = TRUE por echo = FALSE para que al imprimir no se incluya el código. Tiene que tener una longitud máxima de 8 páginas a espacio sencillo.