Actividad_Integradora1

Rodolfo Jesús Cruz Rebollar

2024-10-28

Precipitaciones máximas mensuales para el diseño de obras hidráulicas

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

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

Análisis

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

Estado seleccionado: Quintana Roo

```r
# Importar base de datos de precipitaciones históricas máximas de los estados

lluvias = read.table("precipitaciones_maximas_mensuales.txt", header = TRUE)

head(lluvias)
```

Anio	Mes	Estado	Lluvia
1994	Ene	Aguascalientes	8.3
1994	Ene	Baja.California	10.3

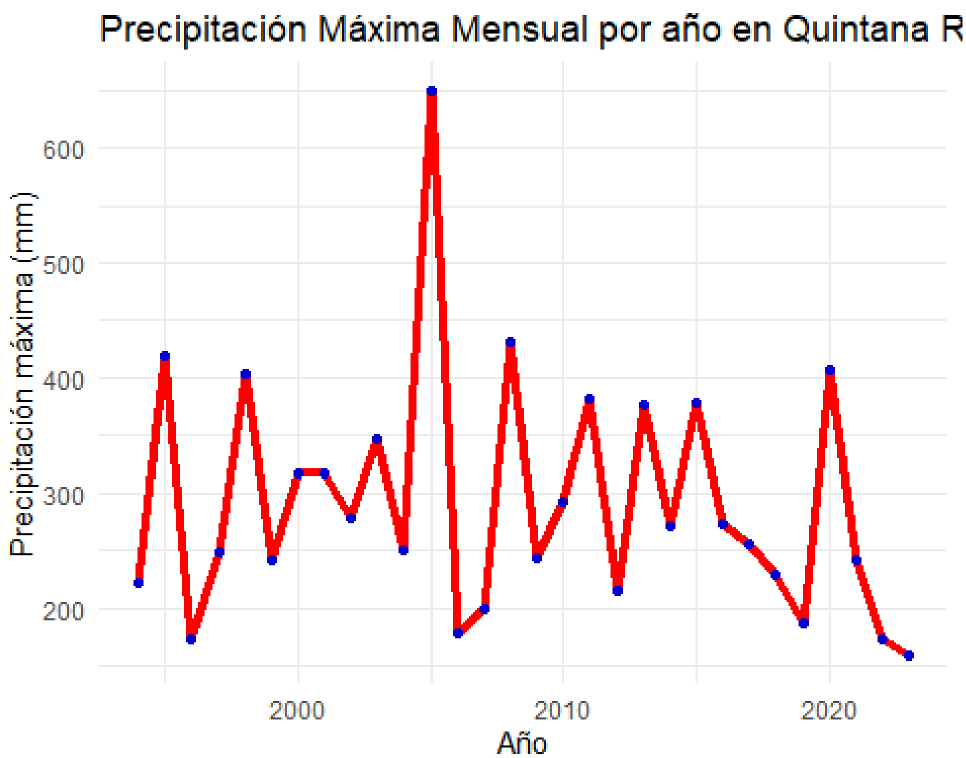
Anio	Mes	Estado	Lluvia
1994	Ene	Baja.California.Sur	0.0
1994	Ene	Campeche	85.4
1994	Ene	Ciudad.de.México	17.7
1994	Ene	Coahuila	12.8

```
# Importar la librería dplyr

library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

# Calcular la precipitación máxima mensual por año para el estado de
Quintana Roo

QRO = lluvias[lluvias$Estado == "Quintana.Roo",]

QRO_max = QRO %>% group_by(Anio) %>% summarise(Precipitacion_max =
max(Lluvia))

QRO_max
```

Anio	Precipitacion_max
1994	222.4
1995	420.0
1996	172.8
1997	249.5
1998	403.6
1999	242.0
2000	317.1
2001	317.2
2002	278.3
2003	346.5
2004	249.9
2005	649.2
2006	178.0

Anio	Precipitacion_max
2007	198.9
2008	431.4
2009	243.3
2010	292.1
2011	382.5
2012	215.8
2013	377.4
2014	271.7
2015	378.8
2016	274.0
2017	255.2
2018	230.2
2019	187.2
2020	406.0
2021	242.3
2022	172.7
2023	159.1

```r
# Graficar la precipitación máxima mensual por año para el estado de
Quintana Roo

library(ggplot2)

ggplot(QRO_max, aes(x = Anio, y = Precipitacion_max)) +
  geom_line(color = "red", linewidth = 1.5) +
  geom_point(color = "blue3") +
  labs(title = "Precipitación Máxima Mensual por año en Quintana Roo",
       x = "Año", y = "Precipitación máxima (mm)") +
  theme_minimal()
```



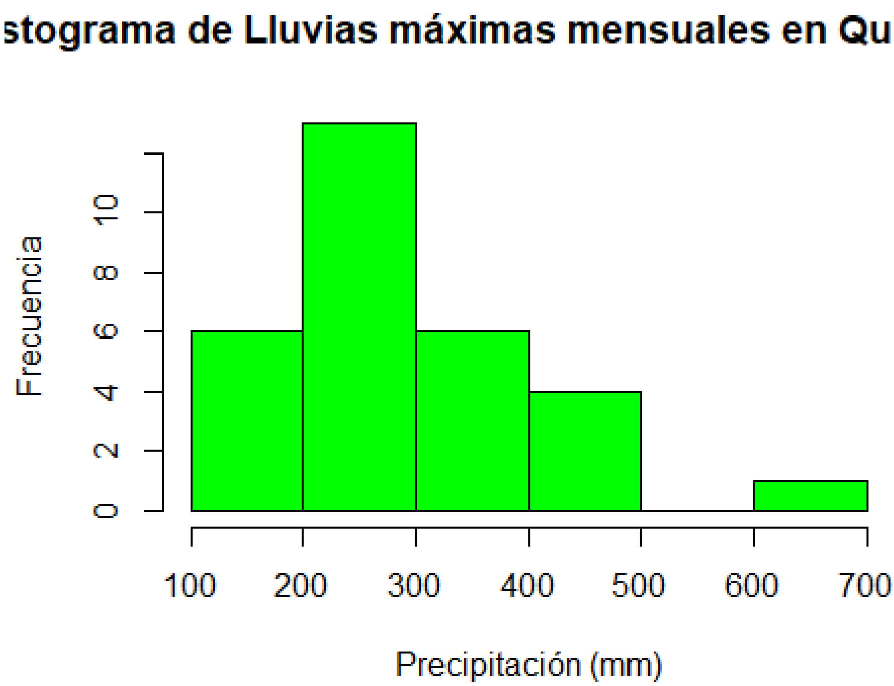
Análisis de precipitaciones máximas mensuales de Quintana Roo

```
# Calcular medidas de centralización y variación de las precipitaciones
máximas mensuales:
# media, mediana, desviación típica, varianza, rango, máximo y mínimo

medidas_CV = QRO_max %>% summarise(promedio =
mean(QRO_max$Precipitacion_max),
                       mediana = median(QRO_max$Precipitacion_max),
                       std_dev = sd(QRO_max$Precipitacion_max),
                       variance = var(QRO_max$Precipitacion_max),
                       range = max(QRO_max$Precipitacion_max) -
min(QRO_max$Precipitacion_max),
                       maximum = max(QRO_max$Precipitacion_max),
                       minimum = min(QRO_max$Precipitacion_max))

# Mostrar las medidas calculadas

medidas_CV
```

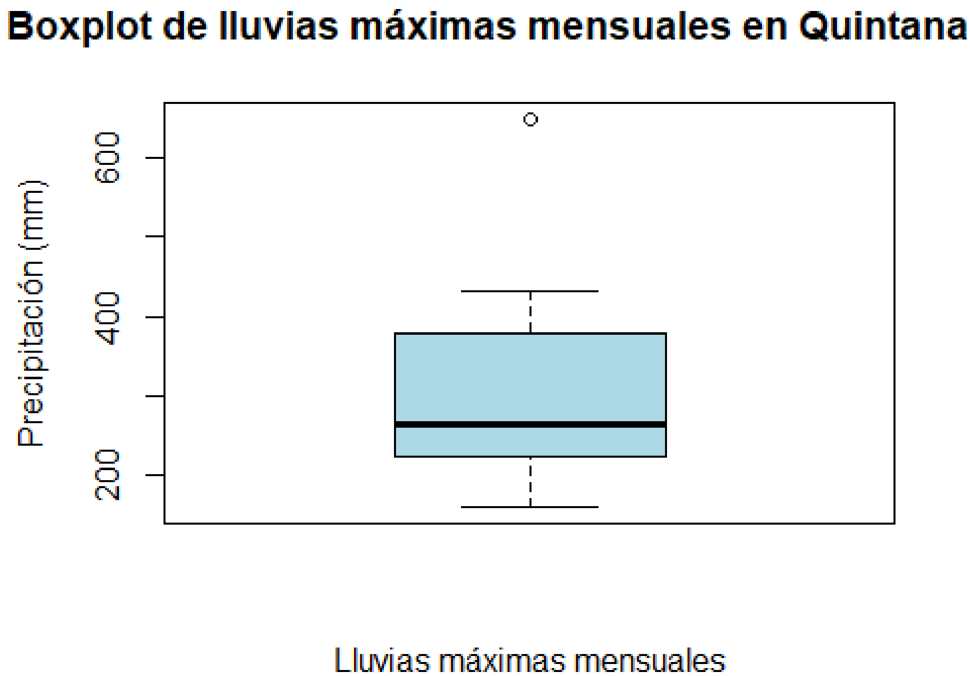
promedio	mediana	std_dev	variance	range	maximum	minimum
292.17	263.45	105.4595	11121.7	490.1	649.2	159.1

```
# Histograma para visualizar la distribución de las lluvias máximas
mensuales

hist(QRO_max$Precipitacion_max, col = "green", xlab = "Precipitación
(mm)",
```

```
    ylab = "Frecuencia", main = "Histograma de Lluvias máximas mensuales en Quintana Roo")
```




```
# Boxplot para visualizar la forma de la distribución de las lluvias máximas mensuales

boxplot(QRO_max$Precipitacion_max, col = "lightblue",
        main = "Boxplot de lluvias máximas mensuales en Quintana Roo",
        xlab = "Lluvias máximas mensuales", ylab = "Precipitación (mm)")
```



Descripción de la distribución de las precipitaciones máximas mensuales:

En términos generales, tomando en cuenta el boxplot, histograma y las medidas de centralización y variación, es posible observar que la distribución de las precipitaciones máximas mensuales presenta un sesgo positivo (a la derecha), lo cual implica que el promedio de precipitaciones máximas se encuentra situado a la derecha de la mediana de las precipitaciones, además de que también es posible notar que en las precipitaciones se presenta un grado de variación mayormente significativo, dado que en el boxplot se observa que la caja de dicho gráfico es mayormente grande, lo cual indica que en efecto los datos de precipitaciones máximas presentan variaciones mayormente significativas entre ellos, lo cual justifica el elevado nivel de varianza obtenido de dichos datos (11121.7014828). Adicionalmente, también es importante mencionar que el sesgo a la derecha que posee la distribución de las precipitaciones mensuales máximas en el histograma realizado, indica que una gran mayoría de dichas precipitaciones tuvo un volumen comprendido entre los 100 y los 500 mm de lluvia, mientras que al mismo tiempo, solamente una minoría de precipitaciones fueron de un volumen situado entre 600 y 700 mm de lluvia, sin embargo, en el histograma se aprecia que la mayor parte de todas las precipitaciones máximas mensuales registradas en el estado de Quintana Roo, fueron de un volumen entre 200 y 300 mm de lluvia, por lo que el rango entre 200 y 300 mm al ubicarse a la izquierda del gráfico, la mayoría de los datos se desplazan para la región izquierda del histograma quedando solamente una minoría de ellos a la derecha, por lo que se produce un sesgo positivo (a la derecha).

Conclusión del análisis estadístico descriptivo:

Finalmente, en la gráfica de precipitaciones máximas mensuales en Quintana Roo, es posible observar que las precipitaciones máximas mensuales en el estado de Quintana Roo se mantienen en su gran mayoría fluctuando entre los 150 y los 450 mm de lluvia, lo cual implica que los niveles máximos de precipitación se mantienen básicamente estables alrededor del rango entre 150 y 450 mm de lluvia, lo que implica que los datos no presentan ninguna tendencia en particular (creciente o decreciente), además de que también es importante mencionar que se observa que las precipitaciones experimentan una disminución en cuanto a su volumen cada 2 o 3 años, motivo por el cual, en términos generales, analizar éste tipo de gráficas resulta muy útil sobre todo para identificar de forma visual los patrones y/o tendencias que siguen los datos (en este caso de las precipitaciones) y a partir de ellos, poder también identificar si un determinado comportamiento en dichos datos se presenta periódicamente, es decir, que los datos se comporten de una determinada manera cada cierto perido de tiempo, por lo que en este caso específico, es posible concluir en base a la gráfica de precipitaciones máximas mensuales, que en el estado de Quintana Roo, las precipitaciones en su mayoría alcanzan un volumen máximo significativo, además de que cada 2 o 3 años, las precipitaciones presentan una disminución en cuanto a su volumen máximo, pudiendo esto llevar a que ciertos sectores como la agricultura se vean afectados debido a la disminución de agua derivada de las lluvias, por lo que tomando esto en cuenta, cada 2 o 3 años será necesario implementar medidas adicionales para regar los cultivos y de esa manera que éstos no se sequen debido a la disminución medianamente considerable del volumen de las precipitaciones.

2. Análisis de frecuencias método gráfico

```r
# Agregar columna a los datos de precipitación máxima, con los datos de lluvia máxima
# ordenados de frma descendente

QRO_max$lluvias_sorted = sort(QRO_max$Precipitacion_max, decreasing = TRUE)

# Mostrar primeros valores de lluvias máximas ordenados de mayor a menor

head(QRO_max$lluvias_sorted)

## [1] 649.2 431.4 420.0 406.0 403.6 382.5

# Agregar la columna rank a los datos de precipitaciones máximas mensuales

QRO_max$rank_m = 1:nrow(QRO_max)

# Mostrar el rank (número de orden) de los primeros datos de precipitación máxima

head(QRO_max$rank_m)

## [1] 1 2 3 4 5 6
```

```r
# Calcular la probabilidad de excedencia de acuerdo con una distribución
Weibull
# m / (N + 1)

QRO_max$P_exe = QRO_max$rank_m / (nrow(QRO_max) + 1)

# Mostrar el valor de la probabilidad de excedencia para los primeros
datos de
# precipitación

head(QRO_max$P_exe)

## [1] 0.03225806 0.06451613 0.09677419 0.12903226 0.16129032 0.19354839

# Calcular la probabilidad de no excedencia para cada precipitación (1 -
P_exe)

QRO_max$P_no_exe = 1 - QRO_max$P_exe

# Mostrar probabilidad de no excedencia para los primeros datos de
precipitación

head(QRO_max$P_no_exe)

## [1] 0.9677419 0.9354839 0.9032258 0.8709677 0.8387097 0.8064516

# Calcular el periodo de retorno P_ret por cada precipitación

QRO_max$P_ret = 1 / QRO_max$P_exe

# Mostrar el cálculo del periodo de retorno para los primeros datos de
precipitación

head(QRO_max$P_ret)

## [1] 31.000000 15.500000 10.333333  7.750000  6.200000  5.166667
```

Gráfico de precipitaciones máximas vs probabilidad de excedencia

```r
# Graficar las precipitaciones máximas vs su probabilidad de excedencia

plot(QRO_max$lluvias_sorted, QRO_max$P_exe, col = "red",
     xlab = "Precipitación máxima (mm)", ylab = "Probabilidad de
excedencia",
     main = "Precipitación máxima vs probabilidad de excedencia", pch =
19)

lines(QRO_max$lluvias_sorted, QRO_max$P_exe, col = "red", type = "l")
```

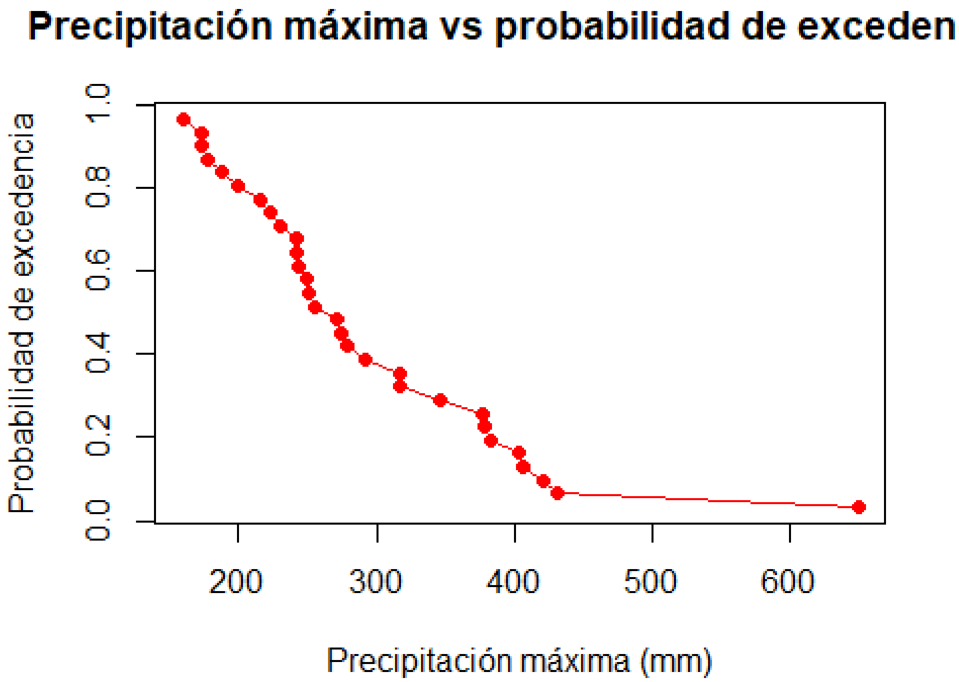
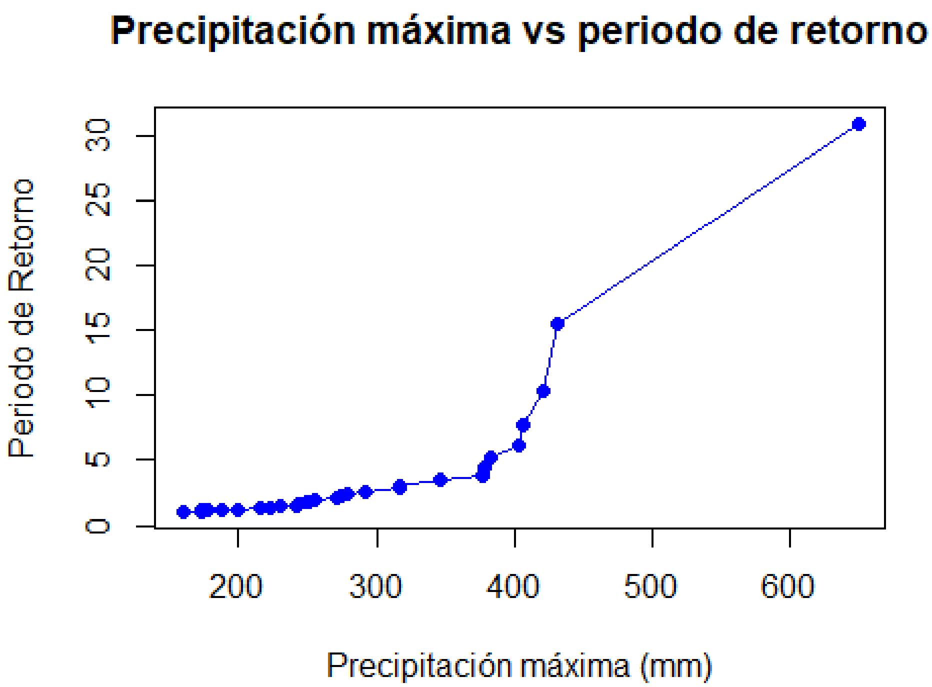


Gráfico de precipitaciones máximas vs periodo de retorno

```
# Graficar las precipitaciones máximas vs su periodo de retorno

plot(QRO_max$lluvias_sorted, QRO_max$P_ret, col = "blue", pch = 19,
     xlab = "Precipitación máxima (mm)", ylab = "Periodo de Retorno",
     main = "Precipitación máxima vs periodo de retorno")

lines(QRO_max$lluvias_sorted, QRO_max$P_ret, col = "blue", type = "l")
```



En cuanto a las gráficas obtenidas anteriormente, en la gráfica de precipitación máxima vs probabilidad de excedencia, se observa que la gráfica presenta en general un comportamiento decreciente, dado que se aprecia que conforme incrementa el volumen máximo de las precipitaciones, la probabilidad de excedencia experimenta una disminución gradual, además, también es posible apreciar en dicha gráfica que cuando la precipitación máxima tiene un volumen menor a 200 mm de lluvia, la probabilidad de excedencia es mayormente cercana a 1, mientras que cuando la precipitación posee un volumen superior a los 600 mm de lluvia, la probabilidad de excedencia desciende hasta ser casi nula, por lo cual, de ésta primera gráfica, es posible afirmar que a menor volumen de precipitación, la probabilidad de que dicho volumen de lluvia sea superado posteriormente es mayor, mientras que en caso contrario, si el volumen de precipitación es elevado, entonces la probabilidad de que se presenten posteriormente otras precipitaciones con un volumen aún mayor de lluvia es menor. Por otro lado, en cuanto a la gráfica de precipitaciones máximas vs periodo de retorno, se aprecia que la gráfica presenta en general un comportamiento creciente, dado que es posible observar que conforme aumenta el volumen máximo de las precipitaciones, el periodo de retorno también incrementa de forma gradual, por lo cual, para precipitaciones con un volumen relativamente bajo de lluvia, el periodo de retorno es menor lo cual implica que debe transcurrir un lapso de tiempo corto entre 2 eventos de precipitación de bajo volumen, mientras que por el contrario, el intervalo de tiempo que tiene que transcurrir entre 2 eventos de precipitación de alto volumen es considerablemente más largo que aquel entre 2 precipitaciones de bajo volumen, lo cual significa que en general, las lluvias de bajo volumen tienden a ser mucho más frecuentes que aquellas de alto volumen.

Adicionalmente, en hidrología, la probabilidad de excedencia hace alusión a la probabilidad de que la magnitud o intensidad de un evento hidrológico (precipitaciones, inundaciones, etc) sobrepase un determinado umbral en un periodo de tiempo específico, por ejemplo, si se afirma que una precipitación de 400 mm posee una probabilidad de excedencia del 10% en un lapso de tiempo de 1 año, esto significa que en cualquier año posterior, hay un 10% de probabilidad de que se presenten precipitaciones de un volumen superior a 400 mm. De manera similar, el periodo de retorno en el contexto hidrológico hace referencia al lapso promedio de tiempo que tiene que transcurrir para que se presente un evento como una inundación, o precipitación de una magnitud superior a un cierto umbral, por ejemplo, si una precipitación de 700 mm tiene un periodo de retorno de 2 años, eso significa que existe un 50% de probabilidad de que dicha precipitación tenga lugar en cualquiera de los próximos 2 años.

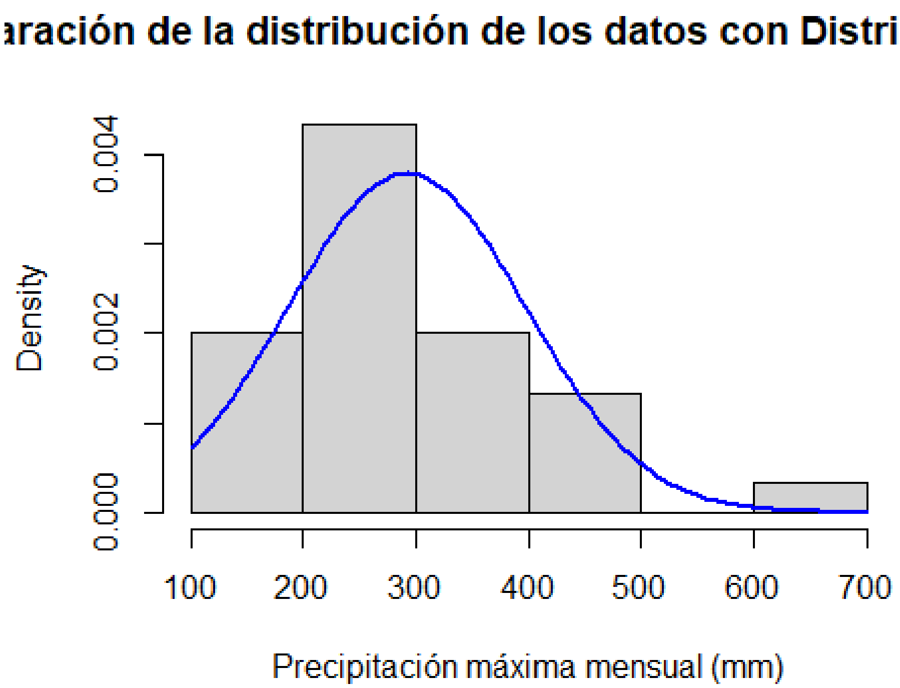
Además de lo anterior, también es importante mencionar que tanto el periodo de retorno como la probabilidad de excedencia son conceptos muy importantes en hidrología, principalmente debido a que se usan para el diseño de infraestructura hidráulica, tales como presas, alcantarillado y puentes que sean capaces de resistir eventos hidrológicos poco recurrentes pero con un alto grado de impacto como el desbordamiento de los ríos e inundaciones severas, además también ayudan en la gestión de los posibles riesgos de dichos eventos extremos, dado que se utilizan también para prever la ocurrencia de dichos eventos de elevada magnitud (por ejemplo inundaciones) y en consecuencia, poder planear con anticipación las medidas de prevención y erradicación que se deben tomar para garantizar la seguridad de la población. Por otra parte, los valores deseables en la probabilidad de excedencia para la precipitación de diseño de una obra dependen tanto de la clase de infraestructura como de la vulnerabilidad y relevancia de la obra ante eventos de extrema magnitud, por ejemplo, para aquellas infraestructuras de elevado riesgo y con una importancia crítica (aeropuertos, plantas de energía, presas, etc), es recomendable una probabilidad de excedencia baja, específicamente valores entre 0.5% y 1%, mientras que al mismo tiempo, para infraestructuras urbanas de menor riesgo (puentes pequeños, sistemas locales de drenaje, etc), se recomienda una probabilidad de excedencia entre el 1% y 5%, y para aquellas infraestructuras que sean temporales o de riesgo leve (por ejemplo senderos rurales), se recomienda una probabilidad de excedencia entre el 5% y el 10%, motivo por el cual, en resumen, para aquella infraestructura que sea crítica para nuestras necesidades como sociedad se recomienda una baja probabilidad de excedencia, mientras que para infraestructura menos vulnerable o crítica, es posible emplear una probabilidad de excedencia mayor, y con ello aceptar un riesgo mayor de fallos en eventos de alta magnitud.

3. Análisis de frecuencias método analítico

A. Ajuste a una distribución normal

Histograma de la función de densidad empírica de los datos:

```r
# Histograma de la distribución empírica de precipitaciones máximas
mensuales

hist(QRO_max$Precipitacion_max, xlab = "Precipitación máxima mensual
(mm)",
     freq = FALSE,
     main = "Comparación de la distribución de los datos con Distribución
Normal")

# Gráfica de la distribución teórica de los datos de precipitaciones
máximas mensuales

curve(dnorm(x, mean = mean(QRO_max$Precipitacion_max),
            sd = sd(QRO_max$Precipitacion_max)), add = TRUE, col =
"blue", lwd = 2)
```

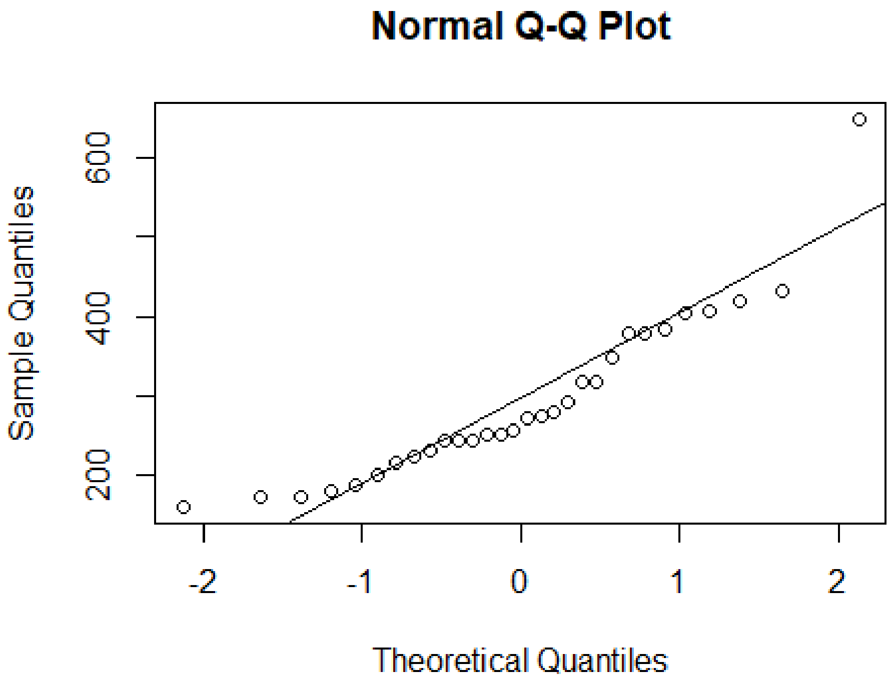


De manera visual, es posible apreciar en el histograma anterior, que los datos de precipitaciones máximas no se ajustan bien a una distribución normal, ya que en el histograma se aprecia que algunas de las frecuencias de las precipitaciones se salen de la curva de color azul que representa la distribución normal teórica de los datos, mientras que de forma similar, también es posible notar que otras frecuencias o barras del histograma, se quedan por debajo de la curva azul, lo cual indica que en términos generales, la dist4ribución verdadera de las precipitaciones máximas presenta ciertas irregularidades en su forma con respecto a una distribución normal, ya sea por exceso (las frecuencias sobrepasan la curva de distribución teórica) o por defecto (las frecuencias se sitúan por debajo de la distribución teórica), además, otro de los motivos por los que las precipitaciones máximas no se ajustan a una distribución normal radica en que la verdadera distribución de los datos presenta un sesgo positivo (a la derecha), puesto que hay una mayor cantidad de datos en la región izquierda del gráfico, provocando que en la región derecha del mismo haya solamente una minoría de ellos, por lo que tomando en cuenta que una distribución normal no tiene sesgo, es posible afirmar que la distribución de los datos de precipitaciones máximas no se ajusta bien a la distribución normal.

Por otro lado, la distribución normal posee 2 parámetros principales: la media representada por la letra μ y la desviación estándar representada por la letra σ, además, los parámetros de la distribución normal se calculan de la forma indicada en el código, debido a que se necesita un conjunto de puntos en el plano bidimensional para graficar la distribución teórica de las precipitaciones máximas, además de que también dichos puntos se calculan evaluando los datos de precipitaciones en la función de densidad de la distribución normal mediante el comando **dnorm**, además otro motivo por el que la media y desviación estándar de los datos se calculan cómo se señala en el código es porque al ser una distribución teórica, significa que los datos obtenidos para graficarla tienen que ser obtenidos por medio de un determinado modelo matemático/probabilístico, en este caso una distribución normal, para poder establecer el comportamiento esperado de los datos en caso de que sigan esa distribución, no obstante, cabe señalar que el comportamiento ideal de los datos obtenido a partir del modelo de probabilidad normal, no necesariamente corresponde al comportamiento real de los datos, sino que es solamente una expectativa sobre el comportamiento de los mismos basada en este caso, en el modelo de probabilidad de la distribución normal.

Gráfica QQ-plot:

```r
# Graficar el QQ-plot de los datos de precipitaciones máximas en Quintana
Roo

qqnorm(QRO_max$Precipitacion_max)
qqline(QRO_max$Precipitacion_max)
```

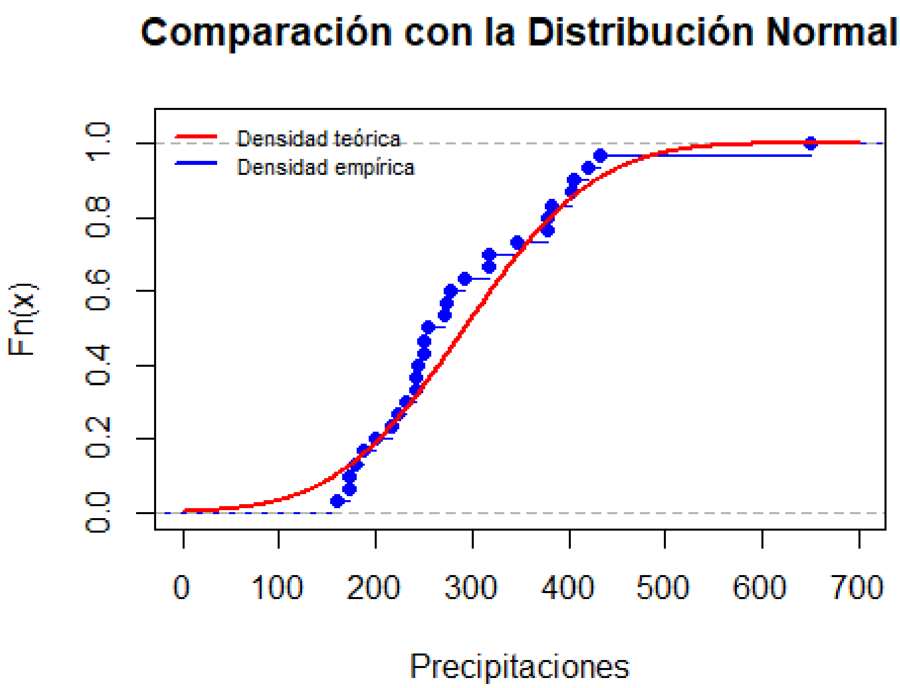


De acuerdo con la gráfica QQ-plot, es posible observar que los datos de precipitaciones máximas no siguen una distribución normal, puesto que algunos de los quantiles graficados (puntos del gráfico) se alejan de la recta principal que representa la distribución normal ideal para los datos, esto debido a que algunos puntos del gráfico se desvían hacia arriba o hacia abajo en los extremos de la recta de la distribución normal ideal, lo cual a su vez es un indicativo de que los datos en cuestión presentan ciertos alejamientos respecto a la distribución normal, además de que con alejamientos nos referimos a que los datos analizados presentan un determinado grado de sesgo en cuanto a su distribución, mismo que en este caso particular, es un sesgo positivo o a la derecha, ya que la mayoría de los datos (puntos del gráfico) se encuentran ubicados en la región izquierda de la recta de la distribución normal, dejando sólo una minoría de datos en la región derecha de la misma, por lo cual, en resumen, dado que la distribución empírica o verdadera de los datos presenta un sesgo además de que también tiene ciertos alejamientos con respecto a la distribución normal, entonces es posible afirmar que las precipitaciones máximas no siguen una distribución normal de acuerdo con la gráfica QQ-plot.

Comparación de distribuciones de probabilidad acumulada empírica y teórica:

```
norm_teorica = pnorm(0:700, mean = mean(QRO_max$Precipitacion_max),
                     sd = sd(QRO_max$Precipitacion_max))

plot(ecdf(QRO_max$Precipitacion_max), main="Comparación con la
Distribución Normal",
     xlab = "Precipitaciones", col = "blue", xlim = c(0, 700), ylim =
c(0, 1.05))
```

```r
par(new = TRUE)

plot(0:700, 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)
```



En la gráfica de la distribución de probabilidad acumulada teórica (en color rojo) y empírica (en color azul), se puede apreciar que la probabilidad acumulada empírica presentan ciertos alejamientos con respecto a la probabilidad acumulada teórica, dado que los puntos en color azul en el gráfico se encuentran en su gran mayoría alejados de la curva roja que representa la distribución de probabilidad acumulada teórica (caso ideal), sin embargo, también cabe destacar que la curva correspondiente a la distribución de probabilidad acumulada teórica (en rojo), es una curva sólida sin ningún tipo de discontinuidad, mientras que la gráfica de la distribución de probabilidad acumulada empírica se presenta como una serie de puntos graficados de abajo hacia arriba siguiendo la trayectoria de la curva de la distribución acumulada teórica, no obstante, la gráfica de la distribución empírica no es una curva continua, sino que tiene una discontinuidad entre cada par de puntos graficados. Adicionalmente, es importante mencionar que por un lado, los datos empíricos son aquellos datos reales recopildos de algún experimento o de alguna otra actividad que implique muestreo dentro de un contexto real, mientras que en cambio, los datos teóricos, son aquellos derivados de algún modelo matemático o de probabilidad, que simbolizan los datos que se obtendrían de dicho modelo si los datos reales siguieran una distribución en particular, además, otro aspecto interesante a mencionar de la gráfica de distribuciones de probabilidades acumuladas radica en que ambas gráficas son mayormente semejantes entre ellas, en el sentido de que ambas poseen una forma bastante similar, ya que al final, ambas tienen la forma de una función particular conocida como la función sigmoide, usada en los modelos de regresión logística, donde Fn(x) corresponde a la probabilidad de ocurrencia de precipitaciones de un cierto volumen, por lo cual en términos generales, de la distribución teórica a la empírica, la probabilidad de ocurrencia correspondiente a cada volumen de precipitación, cambia ligeramente, por lo que a pesar de dichas diferencias entre los datos empíricos y teóricos, los datos de precipitaciones máximas logran seguir en su mayoría la trayectoria de la probabilidad acumulada ideal (teórica) aunque no se ajusten totalmente a ella, por lo que ambas distribuciones de probabilidad acumulada tienen en general una similitud significativa entre ellas.

Hipótesis para pruebas de bondad de ajuste para normalidad:

H_0: Los datos provienen de una distribución normal.

H_1: Los datos no provienen de una distribución normal.

```r
# Realizar prueba de normalidad de Shapiro Wilk para bondad de ajuste

shapiro = shapiro.test(QRO_max$Precipitacion_max)

shapiro

##
##  Shapiro-Wilk normality test
##
## data:  QRO_max$Precipitacion_max
## W = 0.88719, p-value = 0.004141

# Importar librería MASS

library(MASS)

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select

# Realizar prueba de normalidad de Kolmogorov Smirnov (KS) para bondad de
ajuste

ks_normal = ks.test(QRO_max$Precipitacion_max, "pnorm",
```

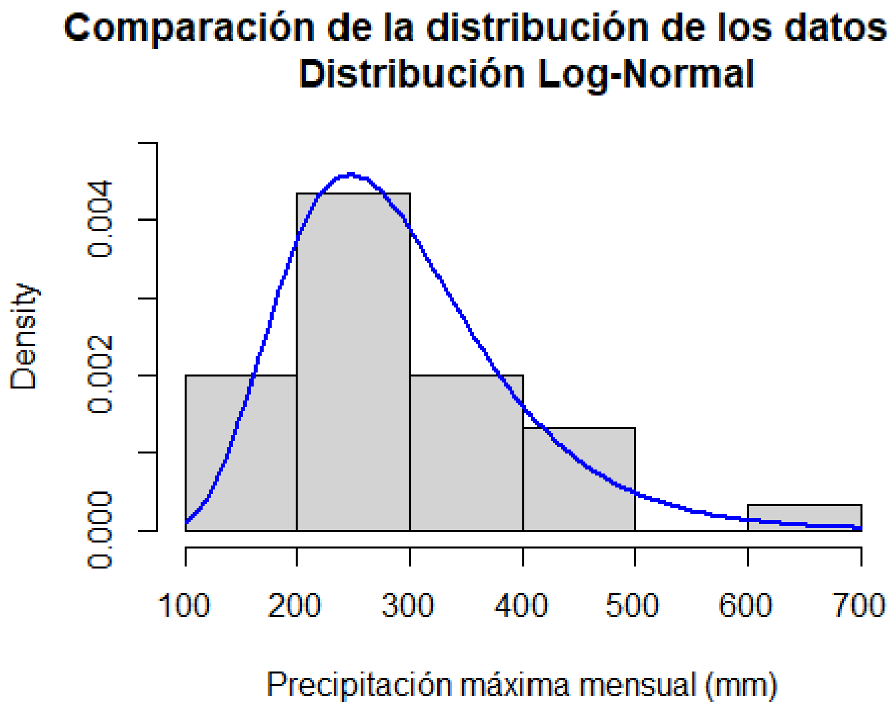
```
                    mean = mean(QRO_max$Precipitacion_max),
                    sd = sd(QRO_max$Precipitacion_max))

ks_normal

##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  QRO_max$Precipitacion_max
## D = 0.15232, p-value = 0.4457
## alternative hypothesis: two-sided
```

En los resultados anteriores, se aprecia que entre la información que nos proporcionan las pruebas de bondad realizadas anteriormente, se encuentra el valor del estadístico D en el caso de la prueba de Kolmogorov-Smirnov, mismo que es igual a 0.15232, además del valor del estadístico W en el caso de la prueba de Shapiro Wilk, mismo que es igual a 0.88719, además de que en ambas pruebas, también se muestra el valor p del test correspondiente, por lo que en el caso del test de Shapiro Wilk, el p valor es igual a 0.004141, mientras que en el caso del test de Kolmogorov-Smirnov, el p valor es igual a 0.4457, motivo por el cual, es posible afirmar que de acuerdo a la prueba de Shapiro Wilk, dado que el p-value de dicho test es menor a 0.05, se rechaza H_0, por lo que de acuerdo a éste test, los datos no provienen de una distribución normal, mientras que por otro lado, en el caso del test de Kolmogorov-Smirnov, dado que el p-value de dicho test es considerablemente mayor a 0.05, no se rechaza H_0, por lo cual, de acuerdo a éste segundo test realizado, es posible afirmar que los datos provienen de una distribución normal, por lo cual, teniendo en cuenta los resultados de ambas pruebas de bondad de ajuste, es posible concluir que los datos de precipitaciones máximas mensuales no son normales, debido a que a pesar de que la prueba de Kolmogorov-Smirnov arrojó que los datos son normales, también existe evidencia más fuerte de que dichos datos en realidad no son normales, además la evidencia en contra de la normalidad es más fuerte debido a que tanto los gráficos realizados previamente (QQ-plot, gráfico de distribuciones de probabilidad acumulada y gráfico de comparación de la distribución de los datos con la distribución normal) como el test de Shapiro Wilk, apoyan el hecho de que los datos de precipitaciones máximas no siguen una distribución normal, mientras que la evidencia a favor de la normalidad es sólo el test de Kolmogorov-Smirnov, por lo que, en conclusión, los datos de precipitaciones máximas mensuales no son normales.

B. Ajuste a una distribución Log-Normal

```
# Graficar histograma de densidad empírica de los datos con distribución
log-normal

hist(QRO_max$Precipitacion_max, xlab="Precipitación máxima mensual (mm)",
     freq = FALSE, ylim = c(0, 0.005),
     main="Comparación de la distribución de los datos con \n
Distribución Log-Normal")

curve(dlnorm(x, mean=mean(log(QRO_max$Precipitacion_max)),
```

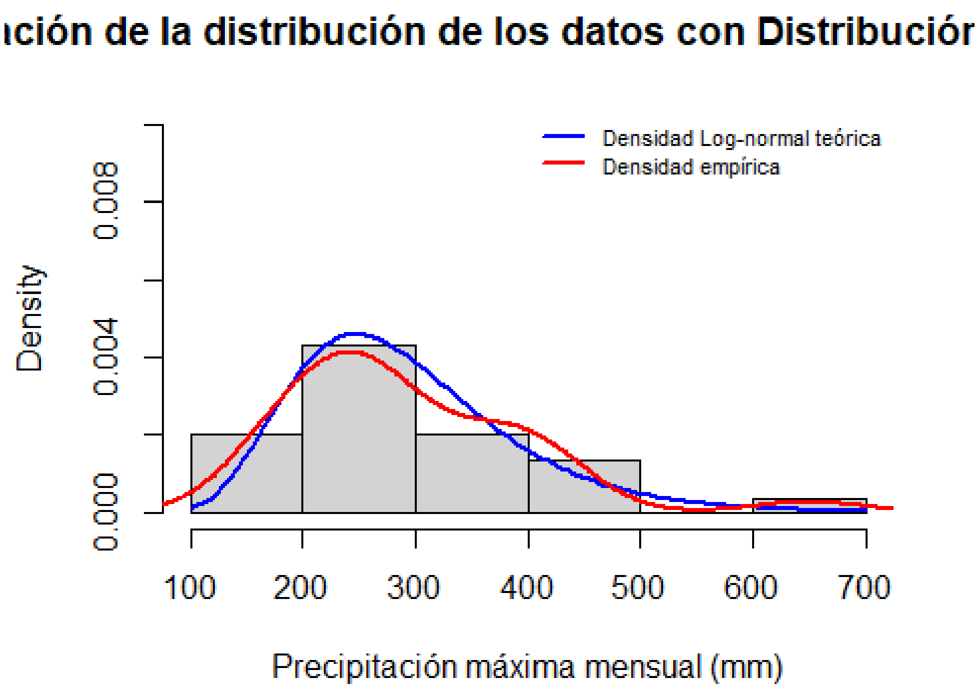
```
              sd=sd(log(QRO_max$Precipitacion_max))), add=TRUE,
col="blue",lwd=2)
```



De forma visual, se puede apreciar que en el histograma anterior, que los datos de precipitaciones máximas mensuales se ajustan de una forma mayormente adecuada a una distribución de tipo log-normal, esto principalmente debido a que en el histograma previo se observa que la distribución de los datos primero experimenta un crecimiento hasta llegar a su máximo pico e inmediatamente después experimenta una disminución gradual bastante evidente hasta llegar a un valor de densidad nulo cuando el máximo volumen de precipitación es de 700 mm, además éste comportamiento coincide con el de una distribución log-normal, ya que la distribución log-normal primero incrementa hasta alcanzar su máximo valor o pico y después desciende de forma gradual hasta llegar al eje horizontal del gráfico.

```
# Graficar la distribución de probabilidad teórica y la empírica

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

curve(dlnorm(x, mean=mean(log(QRO_max$Precipitacion_max)),
              sd=sd(log(QRO_max$Precipitacion_max))), add=TRUE,
col="blue",lwd=2)
```

```
lines(density(QRO_max$Precipitacion_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)
```



En el gráfico anterior se observa que la distribución de probabilidad acumulada empírica (en color rojo) presenta notables acercamientos con respecto a la distribución de probabilidad acumulada teórica, esto debido a que la gráfica correspondiente a la probabilidad acumulada empírica sigue en su gran mayoría la misma trayectoria que la gráfica de la probabilidad acumulada teórica, aunque la gráfica de probabilidad acumulada empírica presenta ciertos alejamientos por encima y por debajo de la distribución teórica, lo cual indica que los datos en cuestión no se terminan de ajustar por completo a una distribución log-normal, no obstante, cabe señalar que ambas curvas graficadas de las 2 distribuciones de probabilidad, también tienen un sesgo a la derecha, además de ser en términos de curtosis mayormente platicúrticas ya que poseen baja altura en el eje vertical del gráfico, motivo por el cual, en términos generales, es posible afirmar que ambas distribuciones de probabilidad (empírica y teórica), tienen una forma bastante similar entre ellas, lo cual es indicativo de que la distribución log-normal se ajusta mejor a los datos que la distribución normal analizada anteriormente.

Nota: los datos teóricos son aquellos que se obtienen dentro de un contexto real, por ejemplo las mediciones hechas de un cierto fenómeno en un experimento, mientras que los datos teóricos, son aquellos que se obtienen en base a un modelo de naturaleza matemática, específicamente de probabilidad, que representan el comportamiento que se esperaría que tuvieran los datos analizados en caso de que éstos mismos siguieran una distribución en específico.

Hipótesis para la prueba de bondad de ajuste para la log-normal:

H_0: Los datos provienen de una distribución Log-Normal.

H_1: Los datos no provienen de una distribución Log-Normal.

```
# Realizar la prueba de bondad de ajuste KS para determinar si los datos
se ajustan a una
# distribución log-normal

ks_lognormal = ks.test(QRO_max$Precipitacion_max, "plnorm",
        mean=mean(log(QRO_max$Precipitacion_max)),
        sd=sd(log(QRO_max$Precipitacion_max)))

ks_lognormal
```

```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  QRO_max$Precipitacion_max
## D = 0.09462, p-value = 0.9279
## alternative hypothesis: two-sided
```

Además, en términos generales, la prueba KS para una distribución Log-Normal nos arroja cierta información, entre la que se encuentra el valor del estadístico D, mismo que es de 0.09462, además del p-value que es igual a 0.9279, por lo cual, en base a dicha información, es posible apreciar que dado el hecho de que el p-value de la prueba resulta ser bastante superior a 0.05, entonces eso implica que no se rechaza H_0, motivo por el cual, es posible afirmar a partir de esto que los datos de precipitaciones máximas mensuales provienen de una distribución log-normal, esto último debido a que la distribución empírica de los datos presentan ciertos desviaciones o alejamientos de la distribución teórica aunque son mínimos, lo cual ocasiona que al momento de aplicar la prueba KS a los datos, la distribución de los datos analizados ya se identifique mayormente como una log-normal, ya que los mínimos alejamientos observados no cambian radicalmente la forma de la distribución real de los datos, misma que ya coincide aproximadamente con la forma de la distribución log-normal ideal, por lo cual, en resumen, los datos de precipitaciones máximas mensuales sí siguen una distribución log-normal.

Adicionalmente, la distribución log-normal posee 2 parámetros principales: μ que hace referencia a la media en escala logarítmica y σ que representa la desviación estándar también en escala logarítmica, además a continuación se presenta el método de momentos para verificar que los parámetros de la distribución log-normal están correctamente calculados:

Considerando que los momentos teóricos de una distribución log-normal son:

$E[X] = e^{\left(\mu+\frac{\sigma^2}{2}\right)}$ (media teórica)

$\mathrm{Var}(X) = \left[e^{\sigma^2} - 1\right] \cdot e^{\left(2\mu+\sigma^2\right)}$ (varianza teórica)

Además, los momentos muestrales son:

media muestral: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

varianza muestral: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Ahora se procede a establecer la ecuación para la media:

$$\bar{x} = e^{\left(\mu+\frac{\sigma^2}{2}\right)}$$

$$\mu = \ln(\bar{x}) - \frac{\sigma^2}{2}$$

Establecer la ecuación para la varianza:

$$s^2 = \left[e^{\sigma^2} - 1\right] \cdot e^{\left(2\mu+\sigma^2\right)}$$

$$\frac{s^2}{\bar{x}^2} = e^{\sigma^2} - 1$$

$$e^{\sigma^2} = 1 + \frac{s^2}{\bar{x}^2}$$

$$\sigma^2 = \ln\left(1 + \frac{s^2}{\bar{x}^2}\right)$$

Finalmente, establecer los valores estimados para la desviación estándar y la media de la distribución log-normal:

$$\sigma = \sqrt{(}\ln\left(1 + \frac{s^2}{\bar{x}^2}\right))$$

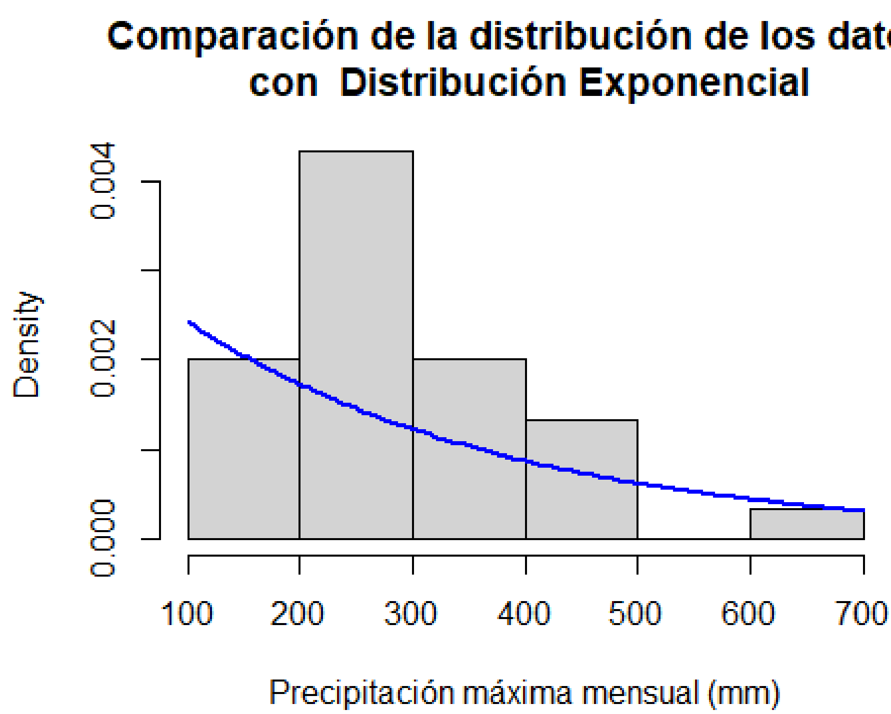
$$\mu = \ln(\bar{x}) - \frac{\sigma^2}{2}$$

```r
# Calcular el valor del parámetro sigma para que los datos se ajusten a
una log-normal

lognormal_sigma = sqrt(log(1 + (var(QRO_max$Precipitacion_max)) /
                             ((mean(QRO_max$Precipitacion_max)) ^ 2)))

# Calcular el valor del parámetro mu para que los datos sigan una log-
normal
```

```
lognormal_mu = log(mean(QRO_max$Precipitacion_max)) - (((lognormal_sigma)
^ 2) / 2)

# Mostrar valores de los parámetros mu y sigma para ajustar los datos a
la log-normal

cat("sigma para ajustar datos a una log-normal: ", lognormal_sigma, "\n")

## sigma para ajustar datos a una log-normal:  0.349959

cat("mu para ajustar datos a una log-normal: ", lognormal_mu)

## mu para ajustar datos a una log-normal:  5.6161
```

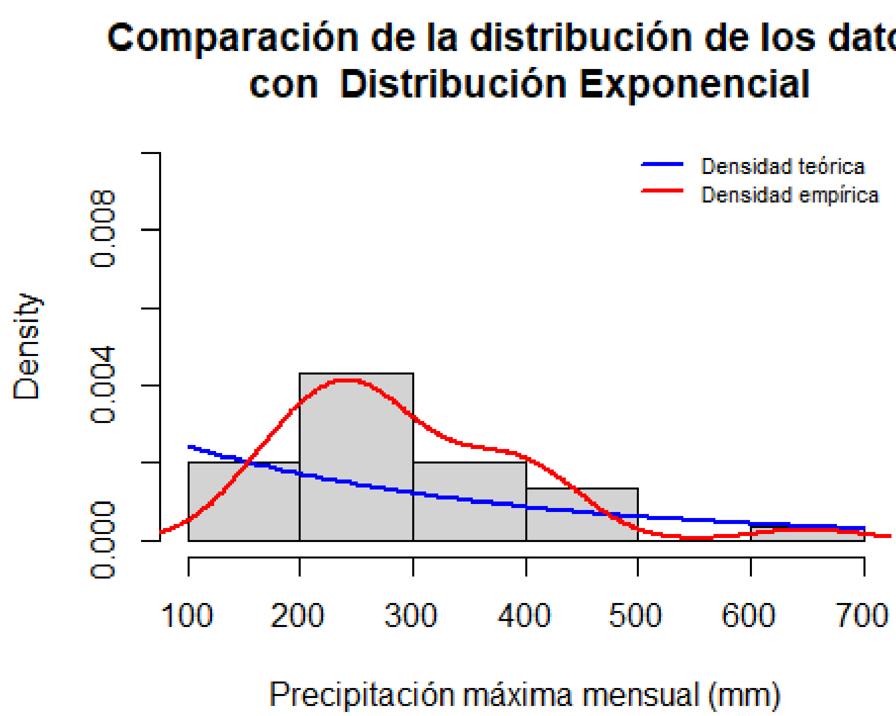
C. Ajuste a una distribución exponencial

```
# Histograma de densidad empírica de los datos cpn la distribución
teórica superpuesta

hist(QRO_max$Precipitacion_max, xlab="Precipitación máxima mensual (mm)",
     freq=FALSE,
     main="Comparación de la distribución de los datos \n con
Distribución Exponencial")

curve(dexp(x, 1/mean(QRO_max$Precipitacion_max)), add=TRUE,
col="blue",lwd=2)
```



De forma visual, se aprecia en el histograma anterior que los datos no se ajustan adecuadamente a una distribución exponencial, debido a que se aprecia que la distribución de los datos presenta al principio un aumento hasta llegar a su máximo valor o pico para después descender de manera gradual hasta alcanzar un valor de densidad casi nulo, lo cual no corresponde con el comportamiento de una distribución exponencial, ya que dicha distribución se caracteriza por descender continuamente durante a lo largo de todo el gráfico y en ningún momento presenta incrementos, motivo por el cual, los datos de precipitaciones máximas mensuales no parecen ajustarse adecuadamente a una distribución exponencial.

```
# Graficar distribuciones de probabilidad acumuladas: empírica y teórica

hist(QRO_max$Precipitacion_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(QRO_max$Precipitacion_max)), add=TRUE,
col="blue",lwd=2)

lines(density(QRO_max$Precipitacion_max), col="red", lwd=2)

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



En el gráfico anterior de las distribuciones de probabilidad empírica y teórica, se observa que la distribución de probabilidad empírica (en color rojo) presenta alejamientos signficativos respecto a la distribución de probabilidad teórica (en color azul), esto principalmente debido a que la distribución empírica supera en tamaño a la distribución exponencial teórica, además de que también la forma de ambas distribuciones probabilísticas es significativamente distinta, en el sentido de que el comportamiento de ambas difiere significativamente, motivo por el cual, es posible afirmar que ambas distribuciones de probabilidad tanto empírica como teórica, no son similares entre sí.

Hipótesis para la prueba KS para distribución exponencial:

H_0: los datos provienen de una distribución exponencial.

H_1: los datos no provienen de una distribución exponencial.

```
# Realizar prueba de bondad de ajuste KS para determinar si los datos se
ajustan a una
# distribución exponencial

ks_exp = ks.test(QRO_max$Precipitacion_max, "pexp")

ks_exp
```

```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  QRO_max$Precipitacion_max
## D = 1, p-value = 0.000000000000001221
## alternative hypothesis: two-sided
```

En términos generales, la prueba KS para una distribución exponencial da como información principal el valor del estadístico de prueba D que en este caso es igual a 1, además del p valor de la prueba, mismo que es igual a 1.221e-15, motivo por el cual, dado que el p valor del test es mucho menor que 0.05, entonces se rechaza H_0, lo cual implica que de acuerdo a la prueba de bondad de ajuste KS, los datos de precipitaciones máximas mensuales no siguen una distribución exponencial, esto debido principalmente a que la forma de la distribución de los datos (distribución empírica) no se ajusta bien a la distribución exponencial ideal que se esperaría que siguieran los datos, lo cual es respaldado por el hecho de que en la prueba KS se rechazó que los datos sigan una distribución exponencial, por lo que en base a esta evidencia, se afirma finalmente que los datos de precipitaciones máximas mensuales no siguen una distribución exponencial.

Además de lo anterior, en el caso de la distribución exponencial, ésta misma solamente tiene 1 parámetro, que corresponde a la tasa promedio de ocurrencia de un evento por cada unidad de tiempo transcurrida que a su vez se representa como λ, por lo que para demostrar que dicho parámetro está ocrrectamente calculado, se siguen los pasos a continuación:

La media teórica para una variable que sigue una distribución exponencial es:

$$E(X) = \frac{1}{\lambda}$$

Además, la media empírica para esa misma variable está representada por:

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (esto simboliza la estimación empírica de la media teórica).

Ahora se procede a igualar la media teórica con la media empírica y resolver dicha ecuación resultante para el parámetro λ:

$$\frac{1}{\lambda} = \bar{x}$$

$$\bar{x}\lambda = 1$$

$$\lambda = \frac{1}{\bar{x}}$$

Entonces, la estimación por el método de momentos para el parámetro λ de la distribución exponencial es:

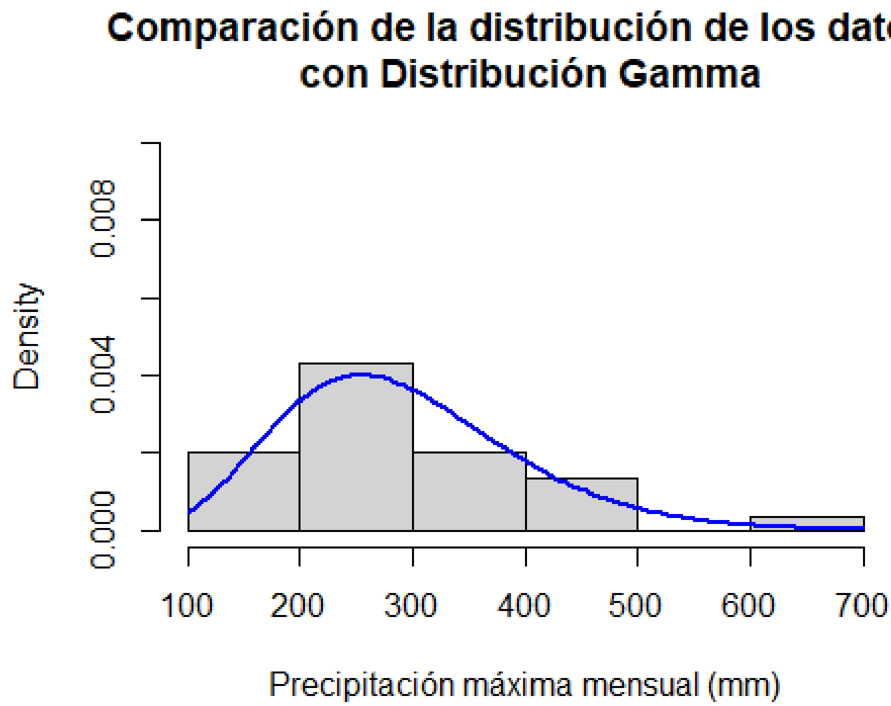
$$\lambda = \frac{1}{\bar{x}}$$

```
# Calcular el valor del parámetro lambda para ajustar los datos a una
exponencial

exp_lambda = 1 / mean(QRO_max$Precipitacion_max)

# Mostrar el valor de lambda para ajustar los datos a una distribución
exponencial

cat("lambda para ajustar datos a una exponencial: ", exp_lambda)

## lambda para ajustar datos a una exponencial:  0.003422665
```

D. Ajuste a una distribución Gamma

```
# Histograma de la función de densidad empírica de los datos con
distribución Gamma

hist(QRO_max$Precipitacion_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(QRO_max$Precipitacion_max)^2/var(QRO_max$Precipitacion_max),
```

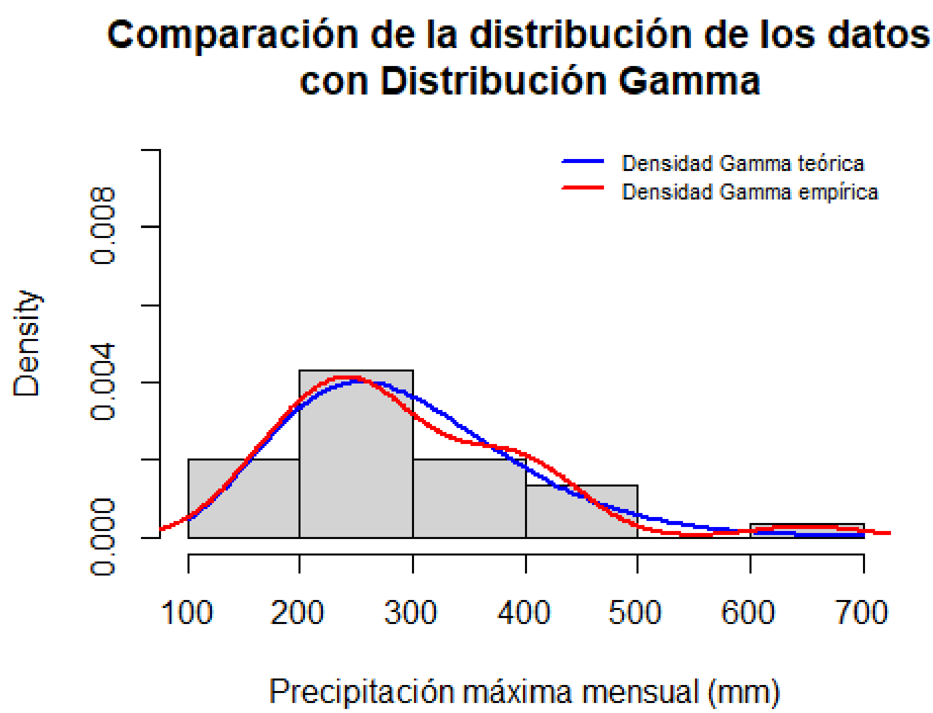
```
mean(QRO_max$Precipitacion_max)/var(QRO_max$Precipitacion_max)),
add=TRUE, col="blue",lwd=2)
```



De acuerdo al histograma anterior, es posible apreciar que a simple vista, los datos de precipitaciones máximas mensuales se ajustan mayormente bien a una distribución Gamma, debido a que la forma de la distribución real o empírica de los datos se parece bastante a la forma característica de una distribución Gamma, esto principalmente debido a que la distribución verdadera de los datos se encuentra sesgada a la derecha, lo cual es una característica propia de la distribución Gamma, además de que también se observa en el histograma que a pesar de que las distribución empírica de los datos presenta ciertos alejamientos con respecto a la distribución Gamma ideal, aún así dicha distribución empírica de los datos se acerca mucho a una distribución Gamma, por lo que en base a dicha evidencia gráfica, es posible afirmar que de manera visual, los datos sí parecen ajustarse bien a una distribución Gamma.

```
hist(QRO_max$Precipitacion_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")

# Estimación de parámetros por método de momentos

curve(dgamma(x,
mean(QRO_max$Precipitacion_max)^2/var(QRO_max$Precipitacion_max),
mean(QRO_max$Precipitacion_max)/var(QRO_max$Precipitacion_max)),
```

```
add=TRUE,
      col="blue",lwd=2)

lines(density(QRO_max$Precipitacion_max), col="red", lwd=2)

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



En el gráfico previo de distribuciones de probabilidad teóricas y empíricas se observa que ambas distribuciones de probabilidad se parecen mucho entre sí, dado que ambas siguen la misma trayectoria, además de que tienen cualidades similares, tales como sesgo a la derecha, un nivel de curtosis mayormente bajo (platicúrticas), además de que la distribución empírica de los datos (curva de color rojo) se encuentra con la distribución teórica (se superponen entre sí) en varias ocasiones a lo largo de todo el histograma, además de que los alejamientos entre ambas distribuciones son mayormente pequeños, por lo cual, éstas 2 distribuciones tienen una forma bastante similar, lo cual significa que ambas distribuciones presentan variaciones menores en cuanto a su forma y por lo tanto, también en cuanto a su comportamiento, por lo tanto, en resumen, se puede afirmar que ambas distribuciones de porbabilidad acumulada (empírica y teórica) se parecen mucho entre ellas.

Hipótesis para la prueba KS para verificar ajuste a una distribución Gamma:

H_0: los datos provienen de una distribución Gamma.

H_1: los datos no provienen de una distribución Gamma.

```
# Ajustar datos a una distribución Gamma

gamma_fit = fitdistr(QRO_max$Precipitacion_max, "gamma")

# Realizar una prueba KS para verificar si los datos se ajustan a una
distribución Gamma

ks_gamma = ks.test(QRO_max$Precipitacion_max, "pgamma",
        shape = gamma_fit$estimate["shape"],
        rate = gamma_fit$estimate["rate"])

ks_gamma

##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  QRO_max$Precipitacion_max
## D = 0.11397, p-value = 0.7895
## alternative hypothesis: two-sided
```

De forma general, la prueba KS para una distribución Gamma nos da como información el valor del estadístico de prueba D que en este caso es igual a 0.11397, además del p valor que en este caso particular es igual a 0.7895, motivo por el cual, dado que dicho p valor resulta ser mucho mayor a 0.05, entonces no se rechaza la hipótesis nula H_0, por lo cual se puede afirmar que los datos de precipitaciones máximas mensuales sí siguen una distribución Gamma, esto debido a que además de que la prueba KS lo confirma, también es posible observar en el gráfico de distribución empírica y teórica que dichas distribuciones tienen muchas similitudes entre ellas, hasta el punto en que ambas se encuentran bastante cerca entre sí, aunque con ligeros alejamientos, no obstante, dado que las 2 distribuciones tienen prácticamente datos altamente semejantes entre sí, entonces se concluye de esta parte del análisis que los datos de precipitaciones máximas mensuales siguen una distribución Gamma.

Además de lo anterior, la distribución Gamma posee 2 parámetros principales: forma representada por k y parámetro de escala, representado a su vez por θ, por lo cual, a continuación se demostrará mediante el método de momentos que éstos 2 parámetros están correctamente calculados:

¿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.

Por un lado, los parámetros de la distribución Gamma se relacionan con sus momentos teóricos de la siguiente manera:

Media teórica: $\mu = k\theta$

Varianza (σ^2): $\sigma^2 = k\theta^2$

Además, los momentos muestrales de dicha distribución Gamma son:

Media muestral: $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Varianza muestral: $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$

Ahora se procede a igualar la media teórica con la media muestral, además de la varianza teórica con su correspondiente variante muestral:

Relación de medias:

$$\hat{\mu} = k\theta$$

Relación de varianzas:

$$\hat{\sigma}^2 = k\theta^2$$

Con las 2 ecuaciones formadas anteriormente, se forma un sistema de 2 ecuaciones y 2 incógnitas que es necesario resolver mediante los métodos ya conocidos:

- Despejar θ de la primera ecuación y sustituir el resultado en la ecuación 2 que involucra a $\hat{\sigma}^2$:

$$\theta = \frac{\hat{\mu}}{k}$$

$$\hat{\sigma}^2 = k\left(\frac{\hat{\mu}}{k}\right)^2$$

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

Por último, despejar para el parámetro k (parámetro de forma de la distribución Gamma):

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

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

$$k = \left(\frac{\hat{\mu}}{\hat{\sigma}}\right)^2$$

Finalmente, sustituir el resultado de la estimación para k en la primera ecuación para obtener el resultado de la estimación para el parámetro θ:

$$\theta = \frac{\hat{\mu}}{k} = \frac{\hat{\mu}}{\left(\frac{\hat{\mu}}{\hat{\sigma}}\right)^2}$$

$$\theta = \hat{\mu}\left(\frac{\hat{\sigma}^2}{\hat{\mu}^2}\right)$$

$$\theta = \frac{\hat{\sigma}^2}{\hat{\mu}}$$

De esta forma, se obtienen las siguientes estimaciones para los parámetros k y θ de la distribución Gamma:

$$k = \left(\frac{\hat{\mu}}{\hat{\sigma}}\right)^2$$

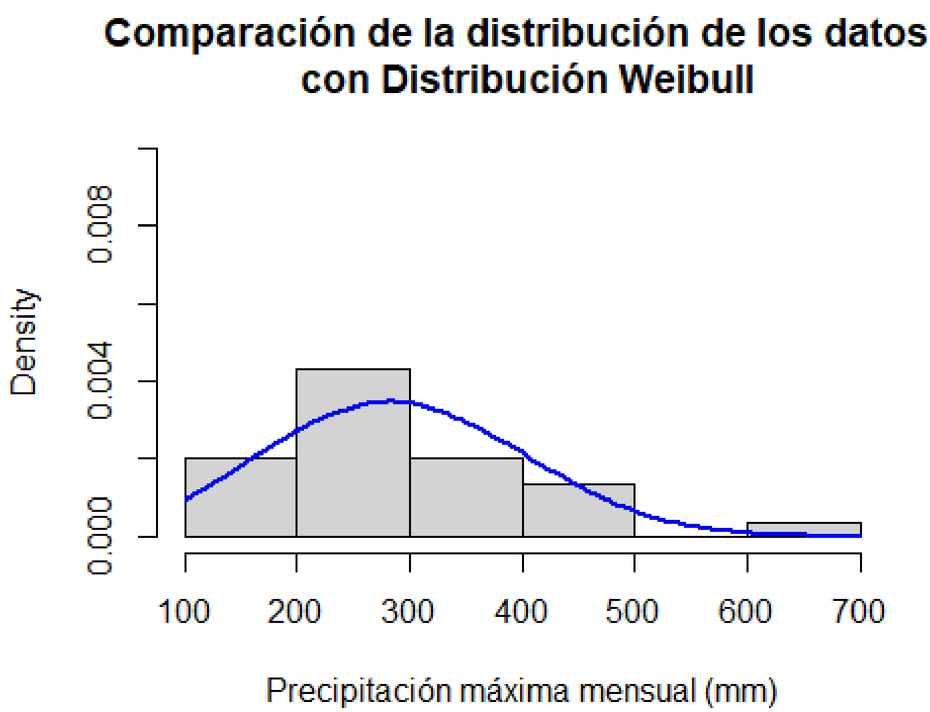
$$\theta = \frac{\hat{\sigma}^2}{\hat{\mu}}$$

```
# Calcular valor del parámetro de forma k para ajustar los datos a una
Gamma

gamma_k = (mean(QRO_max$Precipitacion_max) /
sd(QRO_max$Precipitacion_max)) ^ 2

# Calcular valor del parámetro del parámetro de escala theta para ajustar
datos a una
# Gamma

gamma_theta = var(QRO_max$Precipitacion_max) /
mean(QRO_max$Precipitacion_max)

# Mostrar los valores de k y theta para ajustar los datos a una
distribución Gamma

cat("k para ajustar datos a una gamma: ", gamma_k, "\n")

## k para ajustar datos a una gamma:  7.675382

cat("theta para ajustar datos a una gamma: ", gamma_theta)

## theta para ajustar datos a una gamma:  38.06586
```

E. Ajuste a una distribución Weibull

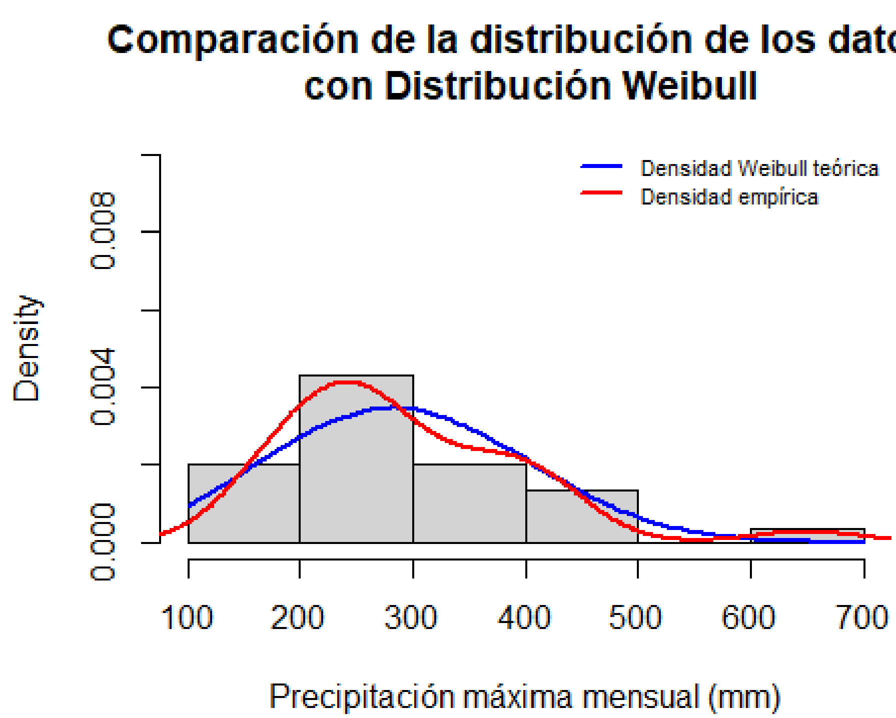
```
# Estimar los parámetros de la distribución Weibull a la que trataremos
de ajustar los
# datos

weibull_fit = fitdistr(QRO_max$Precipitacion_max, "weibull",
lower=c(0,0))

# Mostrar los parámetros calculados para la distribución Weibull

print("Valores de forma y escala para ajustar datos a una Weibull:")
```

```
## [1] "Valores de forma y escala para ajustar datos a una Weibull:"

weibull_fit$estimate

##      shape      scale
##   2.876854 327.214095

# Graficar histograma de la función de densidad empírica de los datos con
distribución
# Weibull

hist(QRO_max$Precipitacion_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)
```



En términos generales, a simple vista parece que los datos en custión no se ajustan de forma totalmente decuada a una distribución Weibull, esto debido a que en el histograma anterior se aprecia que la distribución verdadera o empírica de los datos presenta ciertos alejamientos con respecto a la distribución Weibull teórica que se esperaría que tuvieran los datos, esto debido a que en ciertas regiones del histograma, la distribución empírica de los datos sobrepasa el máximo valor alcanzado por la distribución Weibull ideal, mientras que en otras regiones del histograma, sucede lo contrario, es decir que la distribución real de los datos se ubica por debajo de los valores teóricos correspondientes a dichos valores empíricos, lo cual provoca que debido a los tipos de alejamientos mencionados, los datos analizados no se ajusten adecuadamente a una distribución Weibull, por lo cual, en pocas palabras, los datos de precipitaciones máximas mensuales no parecen ajustarse bien a una distribución Weibull.

```
# Histograma de distribuciones de probabilidad teóricas y empíricas con
distribución
# Weibull

hist(QRO_max$Precipitacion_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(QRO_max$Precipitacion_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)
```



En términos generales, en el gráfico previo es posible observar que la distribución empírica de los datos (en color rojo) posee una forma mayormente irregular en comparación con la distribución teórica cuya forma no presenta irregularidades, esto debido a que la distribución empírica posee datos entre ciertos rangos de valores que son más frecuentes que otros, por lo cual se aprecian ciertas irregularidades en la forma de dicha distribución empírica, además de que también es importante mencionar que ésta distribución (empírica) también tiene algunas características en común con la distribución teórica, entre las que se encuentran el hecho de que ambas distribuciones poseen un sesgo a la derecha, dado que la mayor parte de los datos se encuentran desplazados hacia la región izquierda del gráfico, dejando solamente una minoría de los datos en la región derecha del mismo, además de que ambas distribuciones también son parecidas en el sentido de que inicialmente ambas presentan un incremento hasta alcanzar su respectivo valor máximo, para luego descender gradualmente hasta alcanzar un valor de densidad aproximadamente nulo, motivo por el cual, es posible afirmar que ambas distribuciones de probabilidad acumulada (teórica y empírica) sí se parecen, aunque también tienen ciertas diferencias evidentes en el gráfico previo, lo que provoca que la distribución real de los datos no se ajuste completamente bien a una distribución Weibull.

Hipótesis para la prueba de bondad de ajuste (distribución Weibull):

H_0: los datos provienen de una distribución Weibull.

H_1: los datos no provienen de una distribución Weibull.

```
# Realizar una prueba KS de bondad de ajuste para comprobar si los datos
se ajustan a una
# distribución Weibull, en base a los datos ajustados a la distribución
Weibull

ks_weibull = ks.test(QRO_max$Precipitacion_max, "pweibull",
                     shape = weibull_fit$estimate["shape"],
                     scale = weibull_fit$estimate["scale"])

ks_weibull

##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  QRO_max$Precipitacion_max
## D = 0.13386, p-value = 0.6082
## alternative hypothesis: two-sided
```

Adicionalmente, la prueba KS para una distribución Weibull nos arroja como información el valor del estadístico de prueba D, que en este caso dicho estadístico D tiene un valor de 0.13386, además de que también otra información que nos brinda la prueba KS para una distribución Weibull es el p valor de la prueba que en este caso es igual a 0.6082, por lo cual, dado que el p valor es mucho mayor a 0.05, entonces no se rechaza H_0, lo cual implica que los datos analizados provienen de una distribución

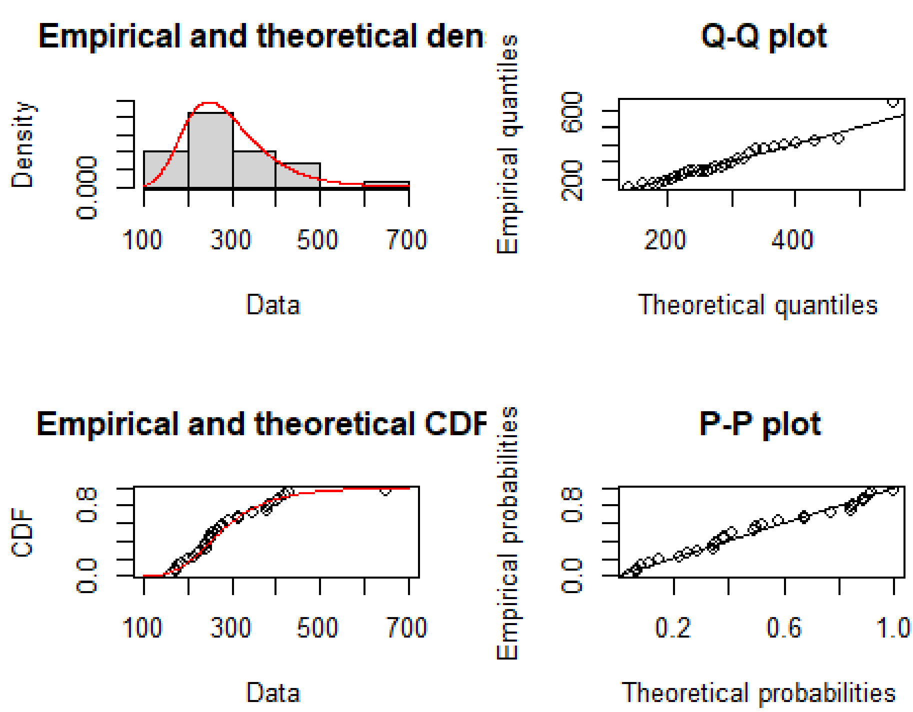
Weibull, motivo por el cual, es posiblr afirmar finalmente que los datos de precipitaciones máximas mensuales sí siguen una distribución Weibull, puesto que por un lado, la prueba KS arroja como resultado que los datos en cuestión sí se ajustan adecuadamente a una distribución Weibull, mientras que al mismo tiempo, también la evidencia gráfica (gráfico de distribución teórica vs empírica e histograma de función de densidad empírica con Weibull teórica superpuesta) apoya el hecho de que los datos analizados sí se ajustan bien a la distribución Weibull, por lo tanto, en conclusión, los datos de precipitaciones máximas mensuales sí siguen una distribución Weibull.

Por otra parte, cabe señalar que la distribución Weibull posee 2 parámetros principales: parámetro de forma y parámetro de escala, además, es importante mencionar que en la distribución Weibull, la estimación de los parámetros se vuelve más compleja, esto principalmente debido a que los parámetros de la distribución Weibull tienen parámetros cuya función de verosimilitud no es lineal, difucultando así la realización de estimaciones por medio de métodos analíticos, además otro de los motivos por los cuales se dificulta estimar los parámetros en una distribución Weibull radica en que ciertas metodologías de estimación son particularmente sensibles a los valores iniciales provistos a dichos métodos, por lo que el hecho de elegir incorrectamente los valores iniciales a partir de los cuales empezarán a operar éstos algoritmos podría ocasionar que se llegue a un resultado erróneo, o incluso que los algoritmos de estimación no consigan converger, por lo que no arrojarían un resultado definido, además, otro motivo por el que se dificulta la estimación de parametros en distribuciones Weibull está relacionado con que en muestras de tamaño pequeño, los resultados de las estimaciones de los parámetros de dichas distribuciones Weibull tienden a tener sesgo significativo, lo cual puede conducir a tener una disminución significativa en la precisión de las estimaciones realizadas, especialmente al no contar con una suficiente cantidad de datos como para detectar un patrón y realizar las estimaciones en base al mismo.

F. Ajuste a una distribución Gumbel

```
# Función para el cálculo de densidad con la distribución Gumbel

dgumbel = function(x, a, b){1 / b * exp((a - x) / b)*exp(-exp((a - x) /
b))}

# Función para cálculo de probabilidades con la Gumbel

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

# Función para calcular quantiles correspondientes a ciertas
probabilidades con la Gumbel

qgumbel = function(p, a, b){a - b * log(-log(p))}

# Importar librerías fitdistrplus y survival para estimar parámetros
mediante
# distribución Gumbel
```

```r
library(fitdistrplus)

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

## Loading required package: survival

library(survival)

# Ajustar los datos a una distribución Gumbel

gumbel_fit = fitdist(QRO_max$Precipitacion_max, "gumbel", start = list(a
= 1, b = 1))

# Generar los gráficos asociados al ajuste de los datos a la Gumbel

plot(gumbel_fit)
```



Hipótesis para la prueba KS para una distribución Gumbel:

H_0: los datos provienen de una distribución Gumbel.

H_1: los datos no provienen de una distribución Gumbel.

```r
# Calcular la probabilidad de excendencia teórica en base a una
distribución Gumbel

gumbel_exe = 1 - pgumbel(QRO_max$lluvias_sorted, gumbel_fit$estimate[1],
```

```
                        gumbel_fit$estimate[2])

# Realizar prueba KS para determinar si los datos se ajustan a una
distribución Gumbel

ks_gumbel = ks.test(QRO_max$P_exe, gumbel_exe)

ks_gumbel

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

Además de lo anterior, la prueba KS para una distribución Gumbel nos brinda como información el valor del estadístico de prueba D que en este caso es igual a 0.13333, además de el p valor del test, mismo que es igual a 0.9578, además dicha información brindada por la prueba KS refleja qué tan bien se ajustan los datos de las precipitaciones máximas mensuales en Quintana Roo a una distribución Gumbel, por lo cual, dado que el p valor de la prueba es mucho mayor que 0.05, entonces no se rechaza H_0 de la prueba de hipótesis planteada previamente, por lo cual, es posible afirmar que los datos en cuestión (probabilidades de excedencia de las precipitaciones máximas mensuales) siguen una distribución Gumbel, esto debido a que además de la prueba KS que apoya ésta afirmación, también se encuentran las evidencias gráficas (gráfico de densidad teórica y empírica, gráfico de probabilidad teórica vs empírica y QQ-plot), mismas que a su vez indican que las probabilidades de excedencia de las precipitaciones máximas mensuales sí se ajustan adecuadamente a una distribución Gumbel, ya que la forma de la distribución empírica de los datos es bastante similar a la forma de la distribución Gumbel, ya que posee un cierto grado de sesgo a la derecha, lo cual es una de las características más notables de la distribución Gumbel, por lo que en resumen, las probabilidades de excedencia de las precipitaciones máximas mensuales sí se ajustan a la distribución Gumbel por las razones descritas con anterioridad.

Por otro lado, también es importante señalar que la distribución Gumbel posee 2 parámetros principales: la media de la distribución representada por μ, además del parámetro de escala representado por β y éste último parámetro hace referencia al grado de dispersión que tendrá la distribución Gumbel para los datos, además, a continuación se realiza la estimación de los parámetros de la Gumbel partiendo de la media y la desviación estándar de los datos y empleando las fórmulas de la media y la desviación estándar de Gumbel:

Tomando en cuenta las fórmulas para obtener la media y la varianza de una distribución Gumbel:

Media: $\bar{x} = \beta\gamma + \mu$ donde $\gamma = 0.5772$

Varianza: $\sigma^2 = \frac{\pi^2\beta^2}{6}$

Por lo que en primera instancia, la estimación del valor del parámetro β se realiza como sigue:

- De la ecuación correspondiente a la varianza de la distribución Gumbel, se procede a despejar el parámetro β:

$$\pi^2\beta^2 = 6\sigma^2$$

$$\beta^2 = \frac{6\sigma^2}{\pi^2}$$

$$\beta = \frac{\sqrt{6}\sigma}{\pi}$$

A continuación se procede a estimar el valor del parámetro μ de la primera ecuación que corresponde a la media de la distribución Gumbel:

$$\mu = \bar{x} - \beta\gamma$$

A continuación se sustituye el valor de β encontrado previamente en el despeje de μ:

$$\mu = \bar{x} - \left(\frac{\sqrt{6}\sigma}{\pi}\right)\gamma$$

Por lo tanto, las fórmulas que se usarán para estimar el valor numérico de μ y β son:

$$\beta = \frac{\sqrt{6}\sigma}{\pi}$$

$$\mu = \bar{x} - \frac{\sqrt{6}\sigma\gamma}{\pi}$$

```r
gamma_const = 0.5772

print("Valores de media y escala para ajustar datos a una Gumbel
(estimados a partir de la media y desviación estándar de los datos):")

## [1] "Valores de media y escala para ajustar datos a una Gumbel
(estimados a partir de la media y desviación estándar de los datos):"

# Calcular el valor estimado de la media (mu)

cat("Media (mu) estimada: ",
    mean(QRO_max$Precipitacion_max) -
      ((sqrt(6)*sd(QRO_max$Precipitacion_max)*gamma_const)/pi), "\n")

## Media (mu) estimada:  244.7089
```

```r
# Calcular valor estimado del parámetro de escala (beta)

cat("Parámetro de escala (beta) estimado: ",
(sqrt(6)*sd(QRO_max$Precipitacion_max) / pi))

## Parámetro de escala (beta) estimado:  82.22642
```

Comparación de estimaciones en base a media y desv.est de los datos y con fitdistrplus:

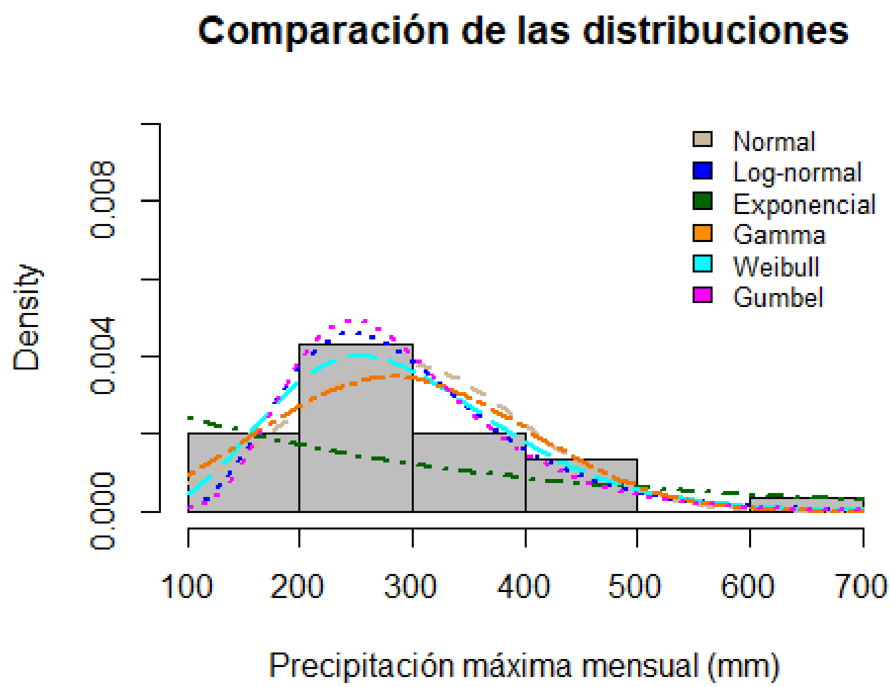
```r
# Mostrar estimaciones hechas mediante la librería fitdistrplus de R
# el atributo a corresponde al valor de la media estimada y el atributo b
al valor estimado
# del parámetro de escala

print("Valores de la media y del parámetro de escala para ajustar datos a
una Gumbel (estimados con la librería fitdistrplus de R):")

## [1] "Valores de la media y del parámetro de escala para ajustar datos
a una Gumbel (estimados con la librería fitdistrplus de R):"

gumbel_fit$estimate

##         a         b
## 246.71768  74.63235
```

Al mostrar las estimaciones realizadas mediante los 2 métodos, se observa que sí hay diferencias entre los resultados de las estimaciones hechas por cada método (no se obtienen los mismos valores por ambos métodos), esto debido a que el método de estimación a partir de las media y desviación estándar de los datos puede estar más influenciado por el sesgo que tengan los datos, puesto que dependiendo del tipo de comportamiento que presenten los datos de acuerdo a su distribución, los datos tendrán ya sea un mayor o menor nivel de sesgo, además, dado que solamente se tienen en total 30 datos o registros correspondientes a las precipitaciones máximas mensuales, lo cual es una cantidad relativamente baja de datos, lo cual puede ocasionar que las estimaciones de los parámetros de la Gumbel se sesgen ya sea positiva o negativamente, esto dado que los datos al tener sesgo, la media de los mismos se ubicará ya sea más a la izquierda o a la derecha de la distribución de los datos y al hacer estimaciones en base a dichos datos, éstos se sesgarán y en consecuencia, las estimaciones resultantes de los parámetros ya no serán iguales a los valores de los parámetros obtenidos mediante la librería fitdistrplus de R.

G. Comparación de ajuste de distribuciones analizadas

```r
# Realizar una gráfica del histograma de densida empírica vs teórica de
todas las
# distribuciones analizadas

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

```
curve(dnorm(x, mean=mean(QRO_max$Precipitacion_max),
sd=sd(QRO_max$Precipitacion_max)),
      add=TRUE, col="bisque3",lwd=2, lty=2)

curve(dlnorm(x, mean=mean(log(QRO_max$Precipitacion_max)),
             sd=sd(log(QRO_max$Precipitacion_max))), add=TRUE,
col="blue",lwd=2, lty=3)

curve(dexp(x, 1/mean(QRO_max$Precipitacion_max)), add=TRUE,
col="darkgreen",lwd=2,lty=4)

curve(dgamma(x,
mean(QRO_max$Precipitacion_max)^2/var(QRO_max$Precipitacion_max),

mean(QRO_max$Precipitacion_max)/var(QRO_max$Precipitacion_max)),
add=TRUE,
      col="cyan",lwd=2, lty=5)

curve(dweibull(x, weibull_fit$estimate[1], weibull_fit$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=159)

#Leyenda del gráfico

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



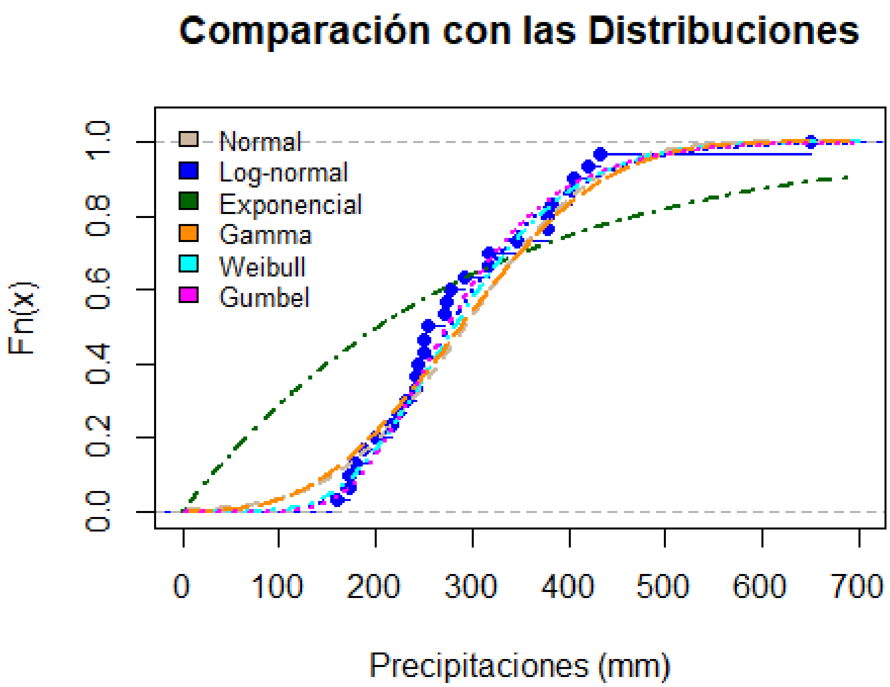

```
# Realizar una tabla donde se mencionen las pruebas de bondad de ajuste
realizadas,
# el valor del estadístico de prueba y el p-value por cada prueba

statistics = c(shapiro$statistic, ks_normal$statistic,
ks_lognormal$statistic,
               ks_exp$statistic, ks_gamma$statistic,
ks_weibull$statistic,
               ks_gumbel$statistic)

p_values = c(shapiro$p.value, ks_normal$p.value, ks_lognormal$p.value,
ks_exp$p.value,
             ks_gamma$p.value, ks_weibull$p.value, ks_gumbel$p.value)

test_summary = data.frame(statistics, p_values,
                          row.names = c("Shapiro Wilk (normal)", "KS
Normal",
                                        "KS Log-Normal", "KS
Exponencial", "KS Gamma",
                                        "KS Weibull", "KS Gumbel"))

colnames(test_summary) = c("Estadístico_de_prueba", "p-value")

test_summary
```

	Estadístico_de_prueba	p-value

	Estadístico_de_prueba	p-value
Shapiro Wilk (normal)	0.8871890	0.0041408
KS Normal	0.1523179	0.4456726
KS Log-Normal	0.0946200	0.9278554
KS Exponencial	1.0000000	0.0000000
KS Gamma	0.1139693	0.7894885
KS Weibull	0.1338560	0.6082352
KS Gumbel	0.1333333	0.9578463

```
# Graficar la probabilidad acumulada empírica vs teórica para todas las
distribuciones
# analizadas

# Histograma de probabilidad empírica de los datos de precipitaciones
máximas mensuales

plot(ecdf(QRO_max$Precipitacion_max), main="Comparación con las
Distribuciones",
     xlab="Precipitaciones (mm)", col="blue", xlim=c(0, 700), ylim=c(0,
1.05))

par(new = TRUE)

# Gráfico de probabilidad teórica correspondiente a la distribución
normal

plot(0:700, norm_teorica, type="l", main="", xlab="", ylab="",
col="bisque3", lwd=2,
     ylim=c(0, 1.05), xaxt="n", yaxt="n", lty=2)

par(new = TRUE)

lognorm_teorica = plnorm(0:700, mean =
mean(log(QRO_max$Precipitacion_max)),
                         sd = sd(log(QRO_max$Precipitacion_max)))

# Gráfico de probabilidade teórica correspondiente a la distribución log-
normal

plot(0:700, lognorm_teorica, type="l", main="", xlab="", ylab="",
col="blue", lwd=2,
     ylim=c(0, 1.05), xaxt="n", yaxt="n", lty=3)

par(new=TRUE)

exp_teorica = pexp(0:700, 1 / mean(QRO_max$Precipitacion_max))
```

```r
# Gráfico de probabilidad teórica correspondiente a la distribución
exponencial

plot(0:700, exp_teorica, type="l", main="", xlab="", ylab="",
col="darkgreen", lwd=2,
     ylim=c(0, 1.05), xaxt="n", yaxt="n", lty=4)

par(new = TRUE)

gamma_teorica = pgamma(0:700, gamma_fit$estimate["shape"],
gamma_fit$estimate["rate"])

# Gráfico de probabilidad teórica correspondiente a la distribución Gamma

plot(0:700, gamma_teorica, type="l", main="", xlab="", ylab="",
col="cyan", lwd=2,
     ylim=c(0, 1.05), xaxt="n", yaxt="n", lty=4)

par(new = TRUE)

weibull_teorica = pweibull(0:700,weibull_fit$estimate["shape"],
                           weibull_fit$estimate["scale"])

# Gráfico de probabilidad teórica correspondiente a la distribución
Weibull

plot(0:700, weibull_teorica, type="l", main="", xlab="", ylab="",
col="darkorange",
     lwd=2, ylim=c(0, 1.05), xaxt="n", yaxt="n", lty=5)

par(new = TRUE)

gumbel_teorica = pgumbel(0:700, gumbel_fit$estimate["a"],
gumbel_fit$estimate["b"])

# Gráfico de probabilidad teórica correspondiente a la distribución
Gumbel

plot(0:700, gumbel_teorica, type="l", main="", xlab="", ylab="",
col="magenta",
     lwd=2, ylim=c(0, 1.05), xaxt="n", yaxt="n", lty=159)

# Leyenda del gráfico de probabilidad acumulada empírica vs teórica

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



Finalmente, después de realizar todo el análisis anterior, es posible concluir que la distribución que mejor se ajusta a los datos de precipitaciones máximas mensuales es la distribución log-normal, debido a que en el gráfico comparativo de probabilidades teóricas y empíricas, se aprecia que dicha distribución (en azul marino) es aquella que posee el menor grado de alejamiento con respecto a la distribución real de los datos en cuestión, lo cual significa que las probabilidades para los datos estimadas en base a una distribución log-normal son las más cercanas a las probabilidades reales de dichos datos, mientras que las curvas correspondientes al resto de las distribuciones analizadas se encuentran, en cambio, más alejadas del centro del gráfico (valores reales de probabilidad), por lo que eso indica que dichas distribuciones por el contrario, no se ajustan de manera adecuada a los datos en cuestión, ya que las probabilidades estimadas a partir de esas distribuciones están más lejos de las probabilidades verdaderas de los datos, siendo mayor el sesgo en estas distribuciones que en la distribución log-normal (la más próxima a las verdaderas probabilidades), por lo que de forma general, en base a la comparación entre los gráficos, es posible afirmar que de entre todas las distribuciones analizadas, aquella que arroja estimaciones de probabilidad con el menor grado de sesgo posible y por tanto con la mayor precisión posible, es la distribución log-normal, al ubicarse justo en el centro del gráfico (ubicación de las probabilidades reales de los datos). Además, también es importante mencionar que otro motivo por el cual la distribución log-normal es la que mejor se ajusta a los datos de precipitaciones máximas mensuales, es porque en las pruebas KS de bondad de ajuste, la prueba KS hecha a los datos ajustados a una distribución log-normal arrojó como resultado un p-value de 0.9279, siendo el p-value más grande de todos los obtenidos en el resto de pruebas KS aplicadas a las demás distribuciones analizadas, motivo por el cual, en resumen, la mejor distribución para los datos de precipitaciones máximas mensuales es la log-normal.

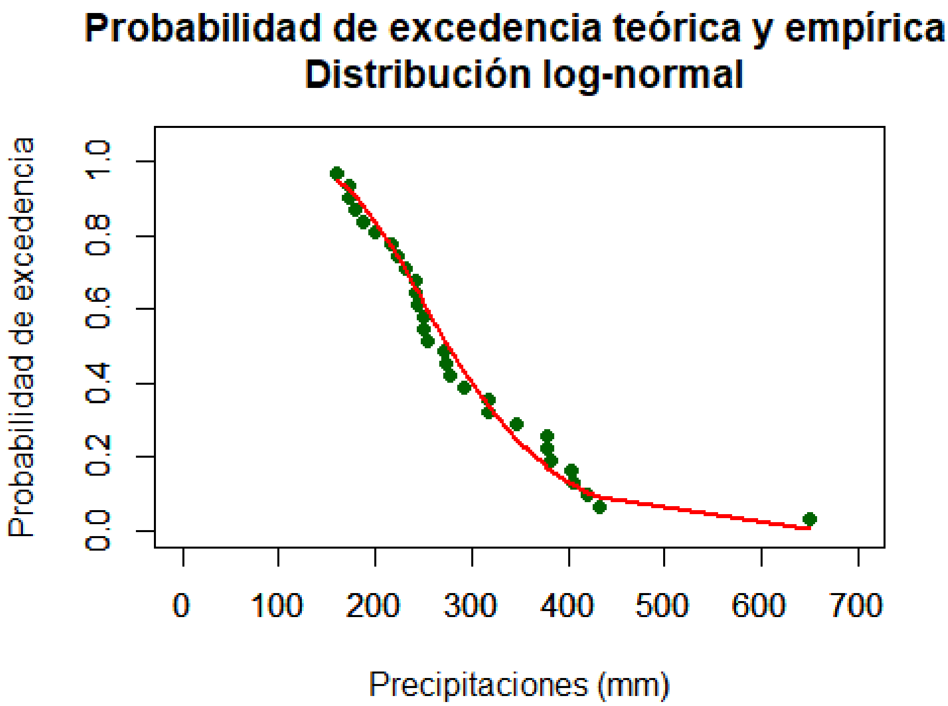
Nota: la prueba KS aplicada sobre las probabilidades de excedencia de las precipitaciones máximas mensuales arrojó un p valor de 0.9578, lo cual es superior al p valor de la prueba KS aplicada sobre los datos ajustados a la distribución log-normal, no obstante, dado que nos interesa saber cuál distribución se ajusta mejor a los datos de precipitaciones máximas mensuales como tal y no a sus probabilidades de excedencia, entonces se descarta la distribución Gumbel, puesto que ésta se aplica únicamente sobre las probabilidades de excedencia de los datos de precipitaciones máximas mensuales mas no sobre esos datos en sí mismos.

DISEÑO DE OBRAS HIDRAÚLICAS

4. Precipitación de diseño de obras hidraúlicas

A. Gráfico comparativo de probabilidad de excedencia teórica vs empírica

```
# Calcular la probabilidad de excendencia teórica en base a la
distribución elegida (Gamma)

lognormal_exe = 1 - plnorm(QRO_max$lluvias_sorted,
                           meanlog = mean(log(QRO_max$lluvias_sorted)),
                           sdlog = sd(log(QRO_max$lluvias_sorted)))

head(lognormal_exe)

## [1] 0.005121208 0.090311820 0.104137128 0.123757560 0.127438784
0.164246057

# Gráfico de probabilidad de excedencia empírica

plot(QRO_max$lluvias_sorted, QRO_max$P_exe,
     main="Probabilidad de excedencia teórica y empírica \n Distribución
log-normal",
     xlab="Precipitaciones (mm)", ylab="Probabilidad de excedencia",
col="darkgreen",
     xlim=c(0, 700), ylim=c(0, 1.05), pch=19)

par(new = TRUE)

# Gráfico de probabilidad de excedencia teórica basada en distribución
log-normal elegida

plot(QRO_max$lluvias_sorted, lognormal_exe, type="l", main="", xlab="",
ylab="",
     col="red", lwd=2, xlim=c(0,700), ylim=c(0, 1.05))
```



De manera general, el gráfico anterior indica que a medida que incrementa el volumen en mm de las precipitaciones máximas mensuales en el estado de Quintana Roo, la probabilidad de excendencia de dichas precipitaciones disminuye gradualmente, esto principalmente a causa de que las precipitaciones de un menor volumen tienden a ocurrir con una mayor frecuencia, en comparación con aquellas otras precipitaciones de alto volumen, cuya frecuencia de ocurrencia es menor, por lo que en el gráfico se señala básicamente que existe una relación inversamente proporcional entre el volumen de las precipitaciones en Quintana Roo y la probabilidad de excedencia de dichas precipitaciones, es decir, que a mayor volumen de precipitación, es menos probable que en un determinado periodo de tiempo, las lluvias posteriores tengan un volumen superior a dicho volumen de precipitación, además de que también sucede lo opuesto, es decir, que a menor volumen de precipitación, mayor será la probabilidad de excedencia, dado que las precipitaciones de bajo volumen pueden ser fácilmente superadas en un cierto lapso de tiempo por precipitaciones de mayor volumen.

Adicionalmente, en cuanto a la certeza de la distribución log-normal elegida, dicha distribución demuestra tener un grado mayormente elevado de certeza, debido a que en el gráfico previo se aprecia que las probabilidades de excedencia empíricas (puntos de color verde en el gráfico) se encuentran bastante próximas a las probabilidades de excedencia teóricas estimadas mediante la distribución log-normal escogida (curva de color rojo), lo cual indica que al ajustar los datos de precipitaciones máximas mensuales a una distribución log-normal, se logran obtener prácticamente probabilidades de excedencia con un alto grado de precisión y por consiguiente también con bajo grado de sesgo, lo cual significa que las probabilidades de excedencia derivadas de la distribución log-normal, son las que se aproximan con el mayor grado posible de precisión a las probabilidades de excendencia empíricas (obtenidas en base a los datos reales), lo cual a su vez se encuentra respaldado por el hecho de que en el gráfico anterior, la recta color rojo (ajuste de los datos mediante distribución log-normal) pasa sobre o muy cerca de los puntos de color verde (probabilidades de excedencia empíricas), reforzando el hecho de que efectivamente, la distribución log-normal resulta ser un ajuste altamente certero para los datos de precipitaciones máximas mensuales en Quitana Roo.

B. Probabilidad de excedencia para límite inferior de intervalo de P_ret sugerido

Tomando en cuenta la función para calcular el periodo de retorno a partir de la probabilidad de excedencia:

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

Despejar de la ecuación previa la probabilidad de excedencia P_{exe}:

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

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

```
# El periodo de retorno ideal para una presa derivadora en zona de riesgo mediana está
# comprendido dentro del intervalo: 100-500 años, por lo que 100 es el límite inferior del
# intervalo

P_ret_inf = 100

# Calcular probabilidad de excedencia para el límite inferior del periodo de retorno (100)

P_exe_100 = 1 / P_ret_inf

cat("Probabilidad de excedencia para 100 años: ", P_exe_100)

## Probabilidad de excedencia para 100 años:  0.01
```

C. Probabilidad de no excedencia y precipitación máxima mensual del periodo de retorno de 100 años

```
# Calcular probabilidad de no excedencia para un periodo de retorno de 100 años

P_NoExe_100 = 1 - P_exe_100

cat("Probabilidad de no excedencia para 100 años: ", P_NoExe_100)

## Probabilidad de no excedencia para 100 años:  0.99
```

```r
# Calcular precipitación máxima mensual del periodo de retorno de 100
años

cat("Precipitación máxima mensual para 100 años: ",
    qlnorm(p=P_NoExe_100, meanlog = mean(log(QRO_max$lluvias_sorted)),
           sdlog = sd(log(QRO_max$lluvias_sorted))), "mm")

## Precipitación máxima mensual para 100 años:  599.1511 mm
```

D. Interpretación de la precipitación máxima mensual y cuestiones finales del análisis

```r
# Explorar otros periodos de retorno diferentes al sugerido (100 años),
en concreto,
# un periodo de retorno de 50 años y otros de 200 años

# Calcular el valor de la precipitación máxima mensual si el periodo de
retorno es de
# 50 años

cat("Precipitación máxima mensual para 50 años: ",
    qlnorm((1-1/50), mean(log(QRO_max$lluvias_sorted)),
           sd(log(QRO_max$lluvias_sorted))), "mm", "\n")

## Precipitación máxima mensual para 50 años:  547.2132 mm

# Calcular el valor de la precipitación máxima mensual si el periodo de
retorno es de
# 200 años

cat("Precipitación máxima mensual para 200 años: ",
    qlnorm((1-1/200), mean(log(QRO_max$lluvias_sorted)),
           sd(log(QRO_max$lluvias_sorted))), "mm")

## Precipitación máxima mensual para 200 años:  650.9934 mm
```

La precipitación máxima mensual estimada para un periodo de retorno de 100 años es de 599.1511 mm, lo cual significa que en cualquiera de los siguientes 100 años, en el estado de Quintana Roo se presentarán precipitaciones con un volumen máximo estimado de 599.1511 mm, además de que también dicho volumen máximo de precipitación tendrá una probabilidad de excedencia del 1%, lo cual significa a su vez que en cualquiera de los próximos 100 años como ya se había mencionado previamente, existe un 1% de probabilidad de que las precipitaciones que se presenten durante ese lapso de tiempo en Quintana Roo, sean de un volumen superior a los 599.1511 mm.

Además de lo anterior, también es importante mencionar que al aumentar el valor del periodo de retorno, el volumen de precipitación máxima mensual también incrementará, esto debido principalmente a que toma una mayor cantidad de tiempo para que se formen las condiciones climatológicas idóneas para producir a su vez precipitaciones de muy alto volumen, mientras que por el contrario, aquellas precipitaciones de un volumen relativamente leve o bajo, tienden a ocurrir con una frecuencia considerablemente mayor a la de precipitaciones de alto volumen, por lo que en el caso de aquellas de volumen leve, tiene que transcurrir un lapso de tiempo notablemente más corto para que sucedan, ya que demandan condiciones climatológicas de menor magnitud, mismas que ocurren también de manera más frecuente. Por otro lado, el valor de la precipitación máxima para el periodo de retorno elegido (100 años) cambiará en caso de que se utilicen datos históricos de otro estado, debido a que los demás estados aparte de Quintana Roo, al contar con una ubicación geográfica diferente, también tendrán condiciones climatológicas distintas a las de Quintana Roo, motivo por el cual, las precipitaciones en el resto de los estados de México seguirán patrones diferentes a las de Quintana Roo, lo cual implica que sigan otro tipo de distribuciones y en consecuencia, se obtenga un valor diferente para la precipitación máxima mensual en un periodo de retorno de 100 años, en el caso de los demás estados de la república mexicana.

Adicionalmente, desde mi punto de vista, las obras hidraúlicas deben construirse a partir de periodos de retorno sugeridos, debido a que de esa manera, es posible decidir qué tipo de materiales emplear en la construcción de las obras, tomando en cuenta principalmente el tiempo de resistencia en años de cada tipo de material que se pueda utilizar en la construcción, además de que el hecho de tomar en cuenta los periodos de retorno sugeridos agrega un componente clave de seguridad a las construcciones, puesto que dichos periodos de retorno se encuentran respaldados por investigaciones hechas por organismos e instituciones especializadas en la realización de obras y otros proyectos hidraúlicos, por lo que dicha informción es además altamente confiable y en consecuencia, bastante útil para la toma de decisiones relacionada con la construcción de obras hidraúlicas.

Finalmente, es importante señalar que es muy importante conocer la distribución a la que mejor se aproximan los datos históricos de las precipitaciones máximas mensuales, debido a que de esa manera, será posible pronosticar o preveer con el mayor grado posible de precisión y por tanto también con el menor nivel posible de sesgo, el volumen máximo de las precipitaciones que tendrán lugar específicamente en el estado de Quintana Roo en el transcurso de los próximos 100 años, más o menos, dependiendo del periodo de retorno elegido, y en base a dichas predicciones, poder tomar la mejor decisión sobre cuáles serán las mejores medidas a tomar para reforzar lo máximo posible, el grado de resistencia de las obras hidraúlicas, además de realizar periódicamente simulaciones de diversos desastres naturales hidrológicos (por ejemplo precipitaciones) de diferentes volúmenes, para ver el nivel de resistencia que tienen las obras hidrológicas frente a eventos de tales magnitudes, y de esa forma asegurar que la infraestructura de las obras sea lo suficientemente resistente como para resistir adecuadamente precipitaciones de volúmenes muy elevados durante la mayor cantidad de tiempo posible.