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

Juan Bernal

2024-10-28

Introducción

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

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

Deberás investigar algunas cuestiones como los parámetros de las distribuciones.

Análisis

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

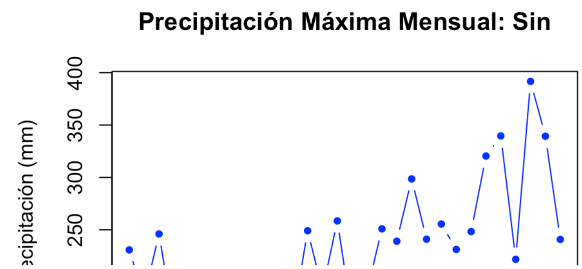
A. Descarga la base de datos de precipitaciones máximas históricas mensuales de todos los estados de la república de la siguiente liga: precipitaciones mensuales. Esta base de datos se construyó con información de los resúmenes mensuales de lluvia y temperatura de CONAGUA (https://smn.conagua.gob.mx/es/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

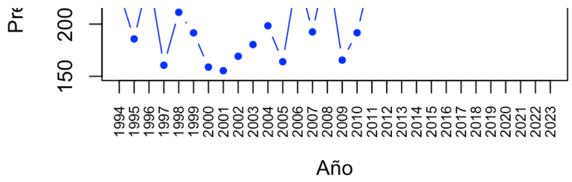
Se seleccionó el estado de Sinaloa.

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

En la tabla anterior se muestran las precipitaciones mensuales en el estado de Sinaloa desde el año de 1994 hasta 2023.

B. Elabora una gráfica de las precipitaciones máximas mensuales por año para tu estado. Para ello deberás calcular la precipitación mensual máxima de cada año y graficarla.





En la gráfica anterior se muestran las precipitacionesx máximas mensuales por año en el estado de Sinaloa. Se puede observar que las precipitaciones tienden a aumentar su tamaño conforme pasan los años, siendo así que la mayor precipitación se dio en el 2021, con precipitaciones de aproximadamente 400 milímetros.

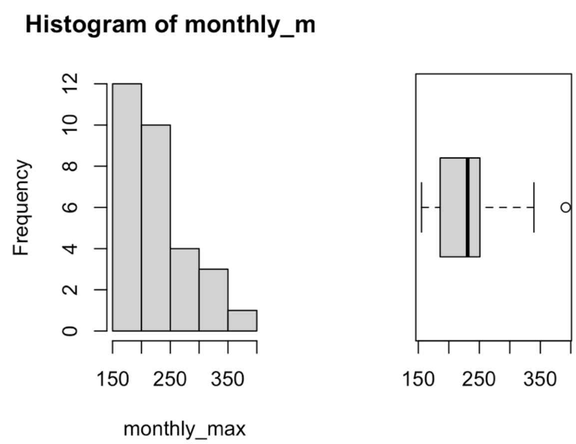
C. Analiza los datos de precipitaciones máximas mensuaels del estado seleccionado.

* Calcula las medidas de centralización y variación de las precipitaciones máximas mensuales

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
155.4	187.25	231	230.9267	250.45	391.7

```
## La desviación estándar de la Lluvia en Sinaloa es  59.87812
```

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



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

De este análisis de medidas de centralización y variación podemos notar que la menor precipitación de las mayores precipitaciones por año fue de 155.4 milímetros, mientras que la mayor precipitación histórica entre 1994-2023 en Sinaloa fue de 391.7 milímetros. También, el 50% de las precipitaciones concuerda casi exactamente con la media de precipitaciones, siendo ambas de aproximadamente 231 milímetros. Por último, la desviación estándar es grande, es decir, hay una gran variación en el tamaño de las precipitaciones con el paso del tiempo.

El histograma de las precipitaciones máximas por año sugiere una distribución con sesgo a la derecha y curtosis baja. Al igual que el boxplot demuestra el mismo sesgo hacia la derecha, además de un dato atípico, que en realidad es la máxima precipitación registrada.

D. ¿Qué puedes concluir observando la gráfica de los máximos mensuales anuales para tu Estado? ¿Observas alguna tendencia? ¿Puedes concluir que cada determinado número de años la cantidad de precipitación sube o baja? ¿Para qué nos sirve analizar este tipo de gráficas?

Con base en los datos y gráficas mostradas, se observa una tendencia a mayores precipitaciones en los últimos 12 años, pues supera el máximo histórico del momento y se tienen mayores precipitaciones consecutivas. Esta observación de las gráficas nos puede dar a entender que algo sucedió hace 12 años que el clima se alteró lo suficiente como para empezar a incrementar su amenaza de manera más rápida, pues tenía un historial de precipitaciones máximas alrededor 250 milímetros, y dicho máximo se daba de manera irregular, dando a entender que no existe algún tipo de tendencia clara que indique si las próximas precipitaciones subirán o bajarán de milímetros.

2. Análisis de Frecuencias Método Gráfico

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

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

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

C. Se calcula la probabilidad de excedencia o de ocurrencia de acuerdo con Weibull, donde el numerador es el número de orden (m) o “rank” y el denominador es la suma del total de datos (N) y 1:

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

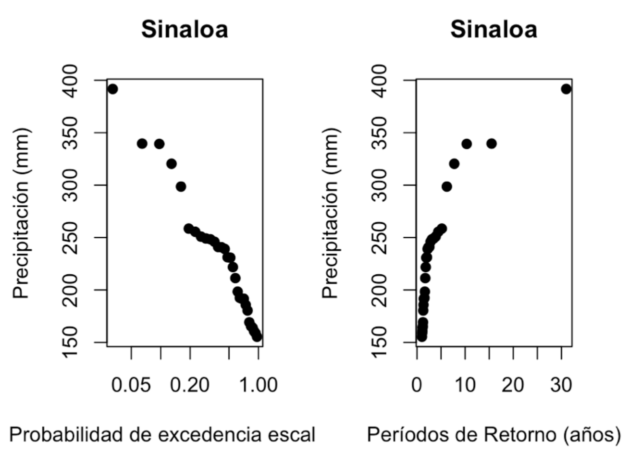
D. Se calcula la probabilidad de no excedencia para cada precipitación (complemento de la probabilidad de excedencia):

$P_{noexe} = 1 - P_{exe}$

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

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

	max_rain	order_max_rain	rank_rain	Pexe	Pnoexe	Pret
1994	230.8	391.7	1	0.0322581	0.9677419	31.000000
1995	185.8	339.6	2	0.0645161	0.9354839	15.500000
1996	246.0	339.3	3	0.0967742	0.9032258	10.333333
1997	160.6	320.4	4	0.1290323	0.8709677	7.750000
1998	211.3	298.6	5	0.1612903	0.8387097	6.200000
1999	191.6	258.5	6	0.1935484	0.8064516	5.166667



F. Describe las gráficas obtenidas. ¿Qué significa la probabilidad de excedencia? ¿Qué significa el periodo de retorno? ¿Por qué es importante en hidrología? ¿Qué valores son deseables en la probabilidad de excedencia para una precipitación de diseño de una obra?

La tabla recopila las mayores precipitaciones por año, después las ordena de mayor y menor y crea un ranking, siendo la mayor precipitación la de rank = 1, y mientras menor sea el rank, menor es la precipitación. La probabilidad de excedencia es, tal cual se define, la probabilidad de que haya una precipitación que exceda dicha precipitación máxima, es decir, la probabilidad de que haya una precipitación mayor a 391.7 milímetros es del 3.22%, la probabilidad de que sea mayor a 339.6 milímetros es del 6.45% y así sucesivamente. De manera similar se tiene la probabilidad de no excedencia, que es el complemento de la probabilidad de excedencia, es decir, la probabilidad de que haya una precipitación menor a 391.7 milímetros es del 96.77%, la probabilidad de que sea menor a una precipitación de 339.6 milímetros es del 93.54% y así sucesivamente. El último componente de la tabla es el periodo de retorno, el cuál indica la cantidad de tiempo en la cual se espera que suceda un evento, es decir, 31 años después del 2021, se prevee una posible precipitación de 391.7 mm.

Las gráficas de probabilidad de excedencia y periodo de retorno sugieren que las probabilidades de obtener precipitaciones anuales máximas mayores aproximadamente 230 milímetros son menores al 50%, y que conforme mayor sea la precipitación, mayor será el periodo de retorno de dicho record histórico.

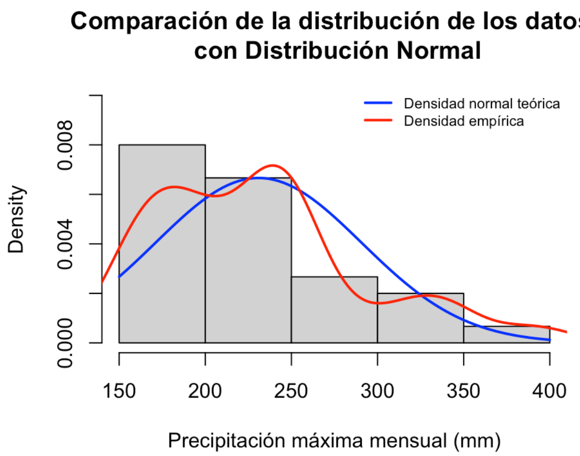
La probabilidad de excedencia deseable dependerá del periodo de vida con el que se planee la obra, así como del cálculo de la precipitación máxima que dicha obra resiste. De esta forma, se tomarán las probabilidades de excedencia de precipitaciones históricas cuyo periodo de retorno se encuentre dentro del periodo de vida de la obra.

3. Análisis de Frecuencias Método Analítico

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

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

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

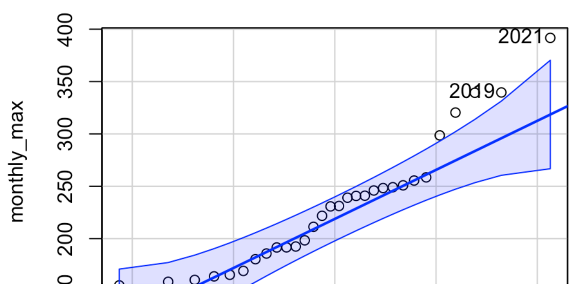


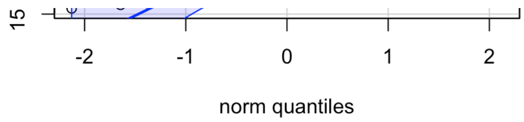
Visualmente, parece que los datos no se ajustan perfectamente a una distribución normal. La línea de densidad empírica muestra una forma asimétrica y sesgada hacia la derecha, lo cual es característico de distribuciones sesgadas y no simétricas, como la log-normal o la exponencial. La línea azul, que representa la densidad teórica de una distribución normal, no sigue bien el patrón de los datos en la parte derecha, donde la densidad empírica tiene una cola más larga y mayor variabilidad.

Por otra parte, la distribución normal cuenta con dos parámetros: la media (μ) y la desviación estándar (σ). Dichos parámetros se calculan de manera sencilla en R con las funciones mean() y sd() respectivamente, dando como único argumento de ambas funciones los datos de las precipitaciones máximas. Este método de estimación se conoce como el método de momentos, ya que utiliza los momentos de los datos (media y varianza) para estimar los parámetros de la distribución teórica.

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

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

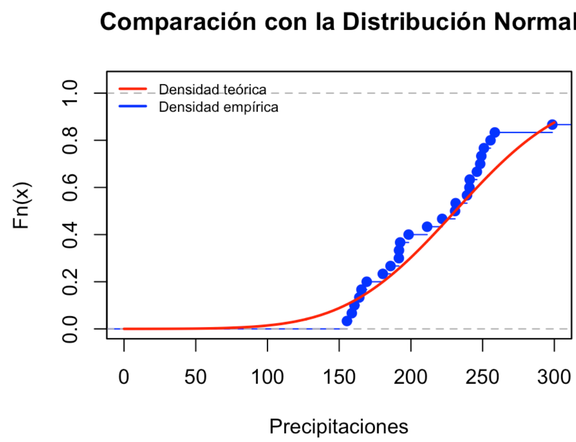





```
## 2021 2019
##   28   26
```

Visualmente, los datos no parecen seguir una distribución normal, pues notemos que los datos en lugar de ajustarse a la recta, forman una curva con mayores pendientes, es decir, sobresale por arriba de la línea normal, mostrando un aparente sesgo a la derecha.

* 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 aquellos que son observados o medidos directamente. En la gráfica, los puntos azules representan la función de distribución acumulada empírica, es decir, la proporción de datos observados que están por debajo de cada valor de precipitación. La línea azul conecta estos puntos para mostrar la acumulación de las probabilidades observadas en los datos. En cambio, los datos teóricos se basan en una distribución matemática que se supone que sigue el comportamiento de los datos. En este caso, la línea roja muestra la función de distribución acumulada teórica para una distribución normal, calculada a partir de los parámetros de media y desviación estándar estimados de los datos.

Las distribuciones acumuladas empírica y teórica no se ajustan perfectamente. La distribución teórica parece subestimar las probabilidades acumuladas en algunos rangos de precipitación (por ejemplo, alrededor de 150-200 milímetros) y sobreestimarlas en otros (200-250 milímetros). Esto sugiere que los datos observados probablemente no siguen una distribución normal debido a que están sesgados o tienen colas más largas que las de una distribución normal.

* 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?

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

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

La prueba de Shapiro-Wilk es una prueba paramétrica que mide la correlación entre los datos y las puntuaciones normales correspondientes. Calcula un estadístico de prueba (W) y lo compara con los valores críticos para determinar si los datos se desvían significativamente de la normalidad (Juárez Manayay, 2021).

El test de Kolmogórov-Smirnov es una prueba de bondad de ajuste que permite verificar si las puntuaciones de una muestra siguen o no una distribución normal.

El estadístico de prueba W del test de normalidad de Shapiro-Wilk obtenido es de 0.9150, mientras que el estadístico de prueba del KS test es de $D = 0.1560$. En el caso de ambos tests, los estadísticos de prueba no nos indican de manera tan explícita los p-values de las pruebas. Es por esto que p-value termina siendo la regla de decisión universal en la pruebas, pues es un valor estandárizado para el cuál ya se tienen umbrales o niveles de significancia estándar teóricos.

Habiendo entendido lo que hacen las pruebas de normalidad, formalizamos entonces el problema.

- Hipótesis nula: $\mu_0 = 0$. Los datos siguen una distribución normal.

- Hipótesis alternativa: $\mu_1 \neq 0$. Los datos no siguen una distribución normal.

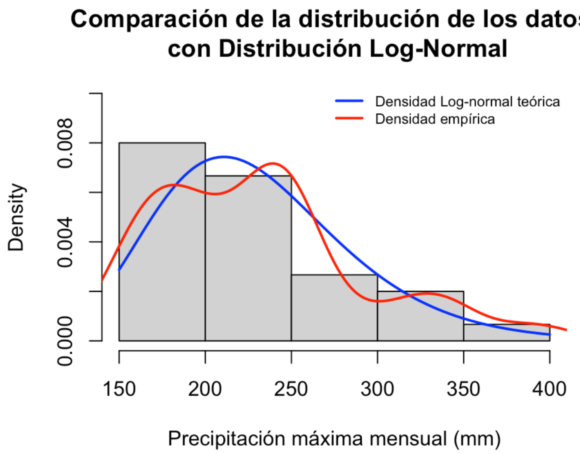
- Regla de decisión: Si $p_{value} < \alpha$, entonces rechazamos la hipótesis nula. En caso contrario, se dice que no se cuenta con suficiente evidencia para rechazar la hipótesis nula.

En el test de Shapiro-Wilk se obtuvo un $p_value = 0.01991$ el cuál es menor a un nivel de significancia de $\alpha = 0.05$. De manera que, según el test de SW, rechazamos la hipótesis inicial, es decir, los datos no siguen una distribución normal. Por otro lado, en el test de KS se obtuvo un $p_value = 0.4167$ el cuál es mucho mayor al nivel de significancia anteriormente definido. De manera que, según el test de KS, no se cuenta con la suficiente evidencia para rechazar la hipótesis inicial, por lo que los datos siguen una distribución normal.

Aún habiendo obtenido resultados diferentes en ambas pruebas, podemos darle mayor credibilidad al resultado de la prueba de Shapiro-Wilk si tomamos en cuenta el análisis visual de la distribución de los datos, sobre el cual concluimos que estos no eran normales.

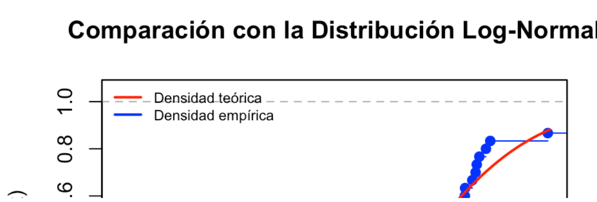
B. Ajuste a una Distribución Log-Normal. Realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Log-normal para ajustar los datos.

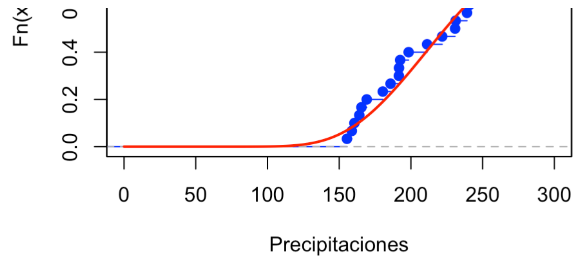
* Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución Log-normal que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. De manera visual, ¿te parece que los datos se ajustan bien a una distribución Log-normal? Explica.



Visualmente, parece que los datos más o menos se ajustan a una distribución log-normal. La línea de densidad empírica muestra una forma asimétrica y sesgada hacia la derecha. Mientras que la línea azul, que representa la densidad teórica de una distribución log-normal, sigue de manera persistente el patrón de los datos en la parte derecha, donde la densidad empírica tiene una cola más larga y mayor variabilidad.

* 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?





En la gráfica, los puntos azules representan la función de distribución acumulada empírica, es decir, la proporción de datos observados que están por debajo de cada valor de precipitación. La línea azul conecta estos puntos para mostrar la acumulación de las probabilidades observadas en los datos. En cambio, los datos teóricos se basan en una distribución matemática que se supone que sigue el comportamiento de los datos. En este caso, la línea roja muestra la función de distribución acumulada teórica para una distribución log-normal, calculada a partir de los parámetros de logaritmo de la media y logaritmo de la desviación estándar estimados de los datos.

Las distribuciones acumuladas empírica y teórica se ajustan de buena manera. La distribución teórica parece subestimar y luego sobreestimar algunas precipitaciones, de forma que la densidad empírica parece serpentear alrededor de la curva teórica. Esto sugiere que los datos observados probablemente siguen una distribución log-normal o parecida debido a que están sesgados.

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

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

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

El valor del estadístico de prueba de KS para comportamiento log-normal es D = 0.1145.

Prueba de comportamiento log-normal con KS.

- Hipótesis nula: $\mu_0 = 0$. Los datos siguen una distribución log-normal
- Hipótesis alternativa: $\mu_1 \neq 0$. Los datos no siguen una distribución log-normal.
- Regla de decisión: Si $p_{value} < \alpha$, entonces rechazamos la hipótesis nula. En caso contrario, se dice que no se cuenta con suficiente evidencia para rechazar la hipótesis nula.

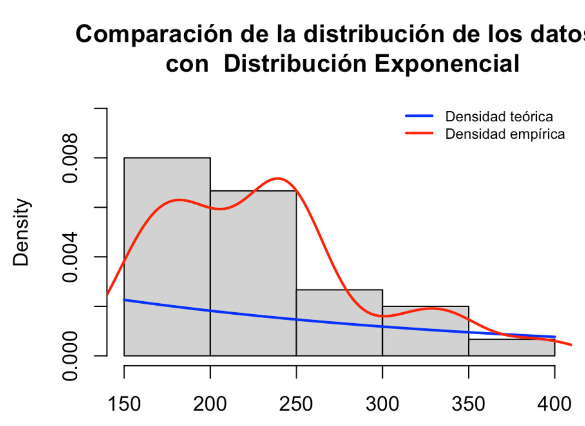
En el test de KS se obtuvo un $p_{value} = 0.7849$ el cuál es mucho mayor al nivel de significancia estándar $\alpha = 0.05$. De manera que no contamos con suficiente evidencia para rechazar la hipótesis inicial, es decir, los datos parecen seguir una distribución log-normal.

* ¿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 cuenta con dos parámetros: el logaritmo de la media ($log(\mu)$) y el logaritmo de la desviación estándar ($log(\sigma)$). Dichos parámetros se calculan de manera sencilla en R con las funciones mean(), sd() y log(), dando como único argumento de dichas funciones los datos de las precipitaciones máximas. Este método de estimación se conoce como el método de momentos, ya que utiliza los momentos de los datos (media y varianza) para estimar los parámetros de la distribución teórica.

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

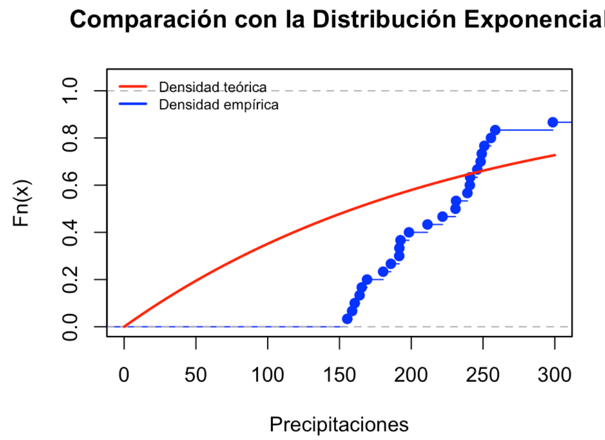
* Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución Exponencial que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. De manera visual, ¿te parece que los datos se ajustan bien a una distribución Exponencial? Explica.



Precipitación máxima mensual (mm)

Visualmente, parece que los datos no se ajustan de manera alguna a una distribución exponencial. La línea de densidad empírica muestra una forma asimétrica y sesgada hacia la derecha. Mientras que la línea azul, que representa la densidad teórica de una distribución exponencial, no sigue de ninguna manera el patrón de los datos, si no que parece mantenerse en la misma densidad aún cuando la precipitación aumenta, esto es debido a que la distribución exponencial es de colas largas.

* 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?



En la gráfica, los puntos azules representan la función de distribución acumulada empírica, es decir, la proporción de datos observados que están por debajo de cada valor de precipitación. La línea azul conecta estos puntos para mostrar la acumulación de las probabilidades observadas en los datos. En cambio, los datos teóricos se basan en una distribución matemática que se supone que sigue el comportamiento de los datos. En este caso, la línea roja muestra la función de distribución acumulada teórica para una distribución exponencial, calculada a partir del parámetro λ estimados de los datos.

Las distribuciones acumuladas empírica y teórica no se ajustan de buena manera. La distribución teórica primero subestimar en gran medida las precipitaciones de la distribución empírica y luego sobreestima las últimas precipitaciones, de forma que la densidad empírica nunca se alinea con la curva teórica. Esto sugiere que los datos observados no siguen una distribución exponencial.

* Haz la prueba KS para determinar si los datos se ajustan a una Exponencial

```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  monthly_max
## D = 0.48979, p-value = 0.0000003686
## alternative hypothesis: two-sided
```

* ¿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?

El estadístico de prueba del test de KS para comportamiento exponencial es D = 0.48979.

Prueba de comportamiento exponencial.

- Hipótesis nula: $\mu_0 = 0$. Los datos siguen una distribución exponencial.
- Hipótesis alternativa: $\mu_1 \neq 0$. Los datos no siguen una distribución exponencial..
- Regla de decisión: Si $p_{value} < \alpha$, entonces rechazamos la hipótesis nula. En caso contrario, se dice que no se cuenta con suficiente evidencia para rechazar la hipótesis nula.

En el test de KS se obtuvo un $p_{value} = 3.686 \times 10^{-7}$ el cuál es mucho menor al nivel de significancia estándar $\alpha = 0.05$. De manera que contamos con suficiente evidencia para rechazar la hipótesis inicial, es decir, los datos no siguen una distribución exponencial.

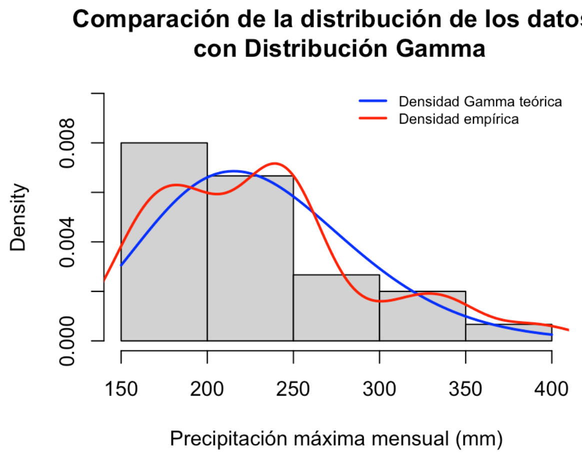
* ¿Cuántos parámetros tiene la distribución Exponencial? ¿Cuáles son? ¿Por qué los parámetros se calculan de la forma en cómo se hace en el código? Sigue paso a paso el método de momentos para demostrar que los parámetros están bien calculados.

La distribución exponencial cuenta con un único parámetro: el inverso de la media ($\frac{1}{\mu} = \lambda$). Dicho parámetro se calcula de manera sencilla en R con la función mean(), de manera que $\lambda = 1/mean(x)$, dando como único argumento de dicha función los datos de las precipitaciones máximas. El método de momentos es apropiado aquí porque la distribución exponencial es completamente descrita por su primer momento (la media), ya que no tiene forma o parámetros de dispersión adicionales como la desviación estándar en distribuciones más complejas. Al calcular simplemente el inverso de la media muestral, se obtiene una estimación de la tasa λ que caracteriza la distribución exponencial de los datos.

D. Ajuste a una Distribución Gamma. Realiza un análisis visual y otro con pruebas de bondad de ajuste para

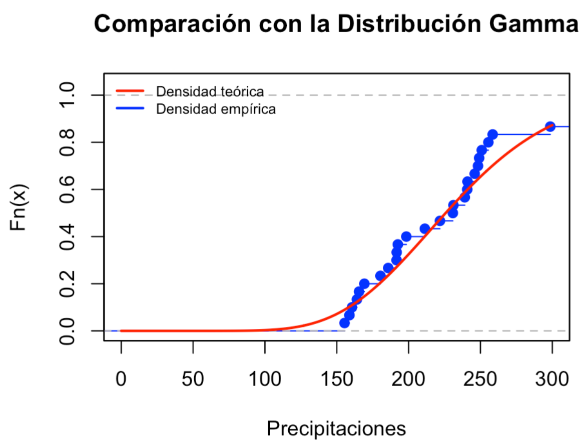
determinar qué tan certera es la Distribución Gamma para ajustar los datos.

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



Visualmente, parece que los datos no se ajustan de buena manera a una distribución Gamma. La línea de densidad empírica muestra una forma asimétrica y sesgada hacia la derecha. Mientras que la línea azul, que representa la densidad teórica de una distribución Gamma, no sigue de manera correcta el patrón de los datos, teniendo densidades drásticamente altas o bajas con respecto a los datos. La caída de densidad de la distribución es más prologanda, mientras que la caída de densidad de los datos es muy rápida.

* 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?



En la gráfica, los puntos azules representan la función de distribución acumulada empírica, es decir, la proporción de datos observados que están por debajo de cada valor de precipitación. La línea azul conecta estos puntos para mostrar la acumulación de las probabilidades observadas en los datos. En cambio, los datos teóricos se basan en una distribución matemática que se supone que sigue el comportamiento de los datos. En este caso, la línea roja muestra la función de distribución acumulada teórica para una distribución Gamma, calculada a partir de los parámetros de forma (k) y escala (θ) estimados de los datos.

Las distribuciones acumuladas empírica y teórica se ajustan de buena manera. La distribución teórica parece subestimar y luego sobreestimar algunas precipitaciones, de forma que la densidad empírica parece serpentear alrededor de la curva teórica. Esto sugiere que los datos observados probablemente siguen una distribución Gamma o parecida debido a que sigue habieno en mayor parte una subestimación por parte de la densidad teórica de Gamma.

* Haz la prueba KS para determinar si los datos se ajustan a una Gamma

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

* ¿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?

El estadístico de prueba del test de KS para comportamiento Gamma fue de D = 0.13171.

Prueba de comportamiento Gamma.

- Hipótesis nula: $\mu_0 = 0$. Los datos siguen una distribución Gamma
- Hipótesis alternativa: $\mu_1 \neq 0$. Los datos no siguen una distribución Gamma.
- Regla de decisión: Si $p_{value} < \alpha$, entonces rechazamos la hipótesis nula. En caso contrario, se dice que no se cuenta con suficiente evidencia para rechazar la hipótesis nula.

En el test de KS se obtuvo un $p_{value} = 0.6282$ el cuál es mucho mayor al nivel de significancia estándar $\alpha = 0.05$. De manera que no contamos con suficiente evidencia para rechazar la hipótesis inicial, es decir, los datos parecen seguir una distribución Gamma.

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

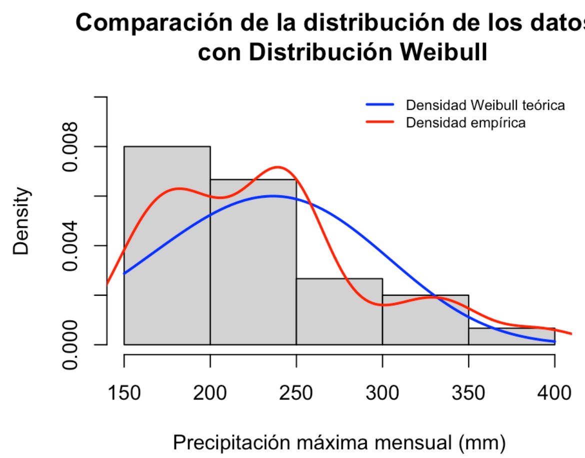
La distribución Gamma tiene dos parámetros: el parámetro de forma (k) y el parámetro de escala (θ). Estos parámetros determinan la forma y el ancho de la distribución. Y se calculan como $k = \frac{\bar{X}^2}{s^2}$ y $\theta = \frac{s^2}{\bar{X}}$, de manera que con R utilizamos las funciones mean() y sd() para estimar los parámetros.

El método de momentos es apropiado para la distribución Gamma porque utiliza la relación directa entre los momentos (media y varianza) y los parámetros de la distribución (k y θ). Al calcular k y θ de esta forma, estamos obteniendo estimaciones que reflejan los momentos de los datos observados, lo que permite que la distribución Gamma teórica se ajuste bien a los datos.

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

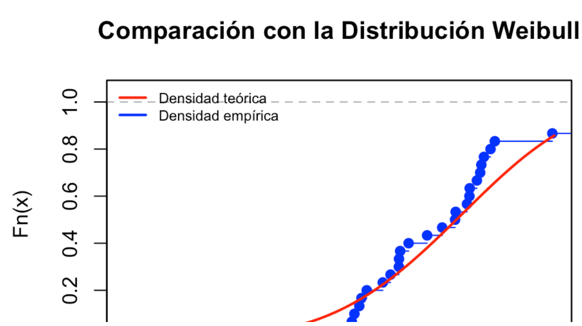
* El cálculo de los parámetros a partir de los datos es un poco más difícil en la distribución Weibull de lo que fue en las anteriores distribuciones, así que recurriremos a que R los estime con el comando fitdistr. Úsalo para estimar los parámetros de la Weibull.

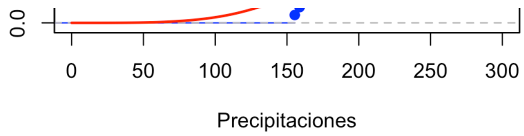
* Contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución Weibull que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos. De manera visual, ¿te parece que los datos se ajustan bien a una distribución Weibull? Explica.



Visualmente, parece que los datos no se ajustan perfectamente a una distribución Weibull. La línea de densidad empírica muestra una forma asimétrica y sesgada hacia la derecha, lo cual es característico de distribuciones sesgadas y no simétricas, como la log-normal o la exponencial. La línea azul, que representa la densidad teórica de una distribución Weibull, no sigue bien el patrón de los datos en la parte izquierda, donde la densidad empírica tiene mayor densidad y poca variabilidad.

* 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?





En la gráfica, los puntos azules representan la función de distribución acumulada empírica, es decir, la proporción de datos observados que están por debajo de cada valor de precipitación. La línea azul conecta estos puntos para mostrar la acumulación de las probabilidades observadas en los datos. En cambio, los datos teóricos se basan en una distribución matemática que se supone que sigue el comportamiento de los datos. En este caso, la línea roja muestra la función de distribución acumulada teórica para una distribución Weibull, calculada a partir de los parámetros de forma (k) y escala (λ) estimados de los datos.

Las distribuciones acumuladas empírica y teórica se ajustan de alguna manera. La distribución teórica parece sobreestimar las precipitaciones en un inicio, para luego subestimar todas las otras precipitaciones. Esto sugiere que los datos observados probablemente siguen una distribución Weibull o parecida debido a que sigue habieno en mayor parte una subestimación por parte de la densidad teórica de Weibull, parecido al caso de la distribución Gamma.

* Haz la prueba KS para determinar si los datos se ajustan a una Weibull

```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  monthly_max
## D = 0.17511, p-value = 0.2821
## alternative hypothesis: two-sided
```

* ¿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?

El estadístico de prueba del test de KS para comportamiento Weibull fue D = 0.17511.

Prueba de comportamiento Weibull.

- Hipótesis nula: $\mu_0 = 0$. Los datos siguen una distribución Weibull.
- Hipótesis alternativa: $\mu_1 \neq 0$. Los datos no siguen una distribución Weibull.
- Regla de decisión: Si $p_{value} < \alpha$, entonces rechazamos la hipótesis nula. En caso contrario, se dice que no se cuenta con suficiente evidencia para rechazar la hipótesis nula.

En el test de KS se obtuvo un $p_{value} = 0.2821$ el cuál es mucho mayor al nivel de significancia estándar $\alpha = 0.05$. De manera que no contamos con suficiente evidencia para rechazar la hipótesis inicial, es decir, los datos parecen seguir una distribución Weibull.

* ¿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.

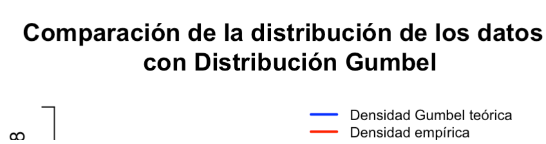
La distribución Weibull tiene dos parámetros: el parámetro de forma (k) y el parámetro de escala (λ). Estos dos parámetros afectan la forma y la dispersión de la distribución, y permiten modelar una variedad de distribuciones de tiempo de vida y fallo, por lo que se usa frecuentemente en análisis de confiabilidad. En el caso de la distribución Weibull, la estimación de los parámetros k y λ es más compleja debido a la naturaleza de sus momentos y a la dependencia de k y λ entre sí de una manera no lineal. No hay una manera de calcular los parámetros, es por eso que utilizamos el comando de R fitdistr() que busca los mejores parámetros k y λ para los datos.

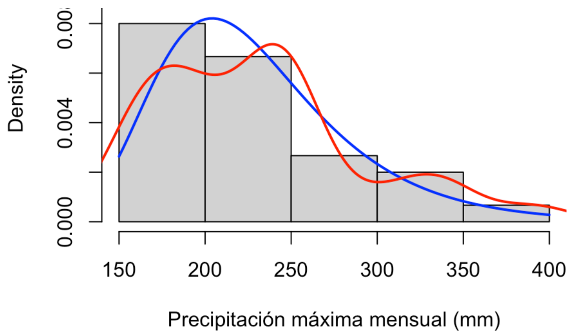
F. Ajuste a una Distribución Gumbel. Realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Gumbel para ajustar los datos.

* Para probar si los datos de precipitación máxima se ajustan a una distribución Gumbel, se necesita definir las funciones de densidad de acuerdo con la función Gumbel. Creálas con las fórmulas de la Distribución Gumbel.

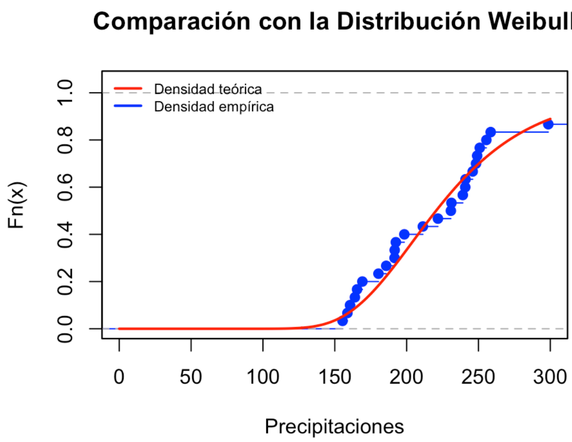
* Para estimar los parámetros y hacer el ajuste de la Distribución Gumbel con la biblioteca “fitdistrplus”. Haz las gráficas de histograma de densidad empírica y teórica, la probabilidad de acumulada empírica y teórica y el QQplot.

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





El histograma muestra la densidad empírica de las precipitaciones máximas mensuales observadas en los datos. La línea roja de densidad empírica muestra una forma asimétrica y sesgada hacia la derecha. La línea azul representa la densidad teórica de la distribución Gumbel ajustada a los datos. Esto es, la curva que esperaríamos observar si los datos siguieran exactamente una distribución Gumbel. Al comparar ambas, podemos notar visualmente que la distribución Gumbel es un buen ajuste para los datos observados. En este caso, parece que la curva teórica sigue de manera aproximada la tendencia del histograma empírico, aunque hay algunas diferencias en ciertas zonas.



Las distribuciones acumuladas empírica y teórica se ajustan de buena manera. La distribución teórica parece subestimar y luego sobreestimar algunas precipitaciones, de forma que la densidad empírica parece serpentear alrededor de la curva teórica. Esto sugiere que los datos observados probablemente siguen una distribución Gumbel o parecida.

* Haz la prueba KS para determinar si los datos se ajustan a una Gumbel

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  rain_analysis$Pexe and gumbel_exe
## D = 0.1, p-value = 0.9988
## alternative hypothesis: two-sided
```

* ¿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?

El estadístico de prueba del test de KS para comportamiento Gumbel fue D = 0.1

Prueba de comportamiento Gumbel

- Hipótesis nula: $\mu_0 = 0$. Los datos siguen una distribución Gumbel.
- Hipótesis alternativa: $\mu_1 \neq 0$. Los datos no siguen una distribución Gumbel.
- Regla de decisión: Si $p_{value} < \alpha$, entonces rechazamos la hipótesis nula. En caso contrario, se dice que no se cuenta con suficiente evidencia para rechazar la hipótesis nula.

En el test de KS se obtuvo un $p_{value} = 0.9988$ el cuál es muchísimo mayor al nivel de significancia estándar $\alpha = 0.05$. De manera que no contamos con suficiente evidencia para rechazar la hipótesis inicial, es decir, las probabilidades de excedencia de las precipitaciones máximas mensuales parecen seguir una distribución Gumbel. Notemos que el P-value es muy alto, casi indicándonos que es muy seguro que los datos se comporten como 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. Compara los valores obtenidos con los estimados con el comando “fitdistrplus” ¿se obtienen los mismos valores por ambos métodos? ¿por qué crees que se dé esta diferencia? ¿por qué crees que se dé esta diferencia?

La distribución de Gumbel tiene dos parámetros principales: Ubicación (μ) que determina el punto central de la distribución, es decir, dónde se sitúa en el eje de los valores, y Escala (β) que afecta la dispersión de la distribución, controlando cuánto se extiende hacia los lados.

La distribución Gumbel tiene fórmulas específicas para la media y la desviación estándar en función de los parámetros μ y β:

- $mean = \mu + \beta\gamma$
- $\sigma = \frac{\pi\beta}{\sqrt{6}}$

donde γ es la constante de Euler-Mascheroni, aproximadamente igual a 0.5772. Con dichas fórmulas podemos despejar de tal forma que

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

Y con β calculada se puede obtener

- $\mu = mean - \beta\gamma$

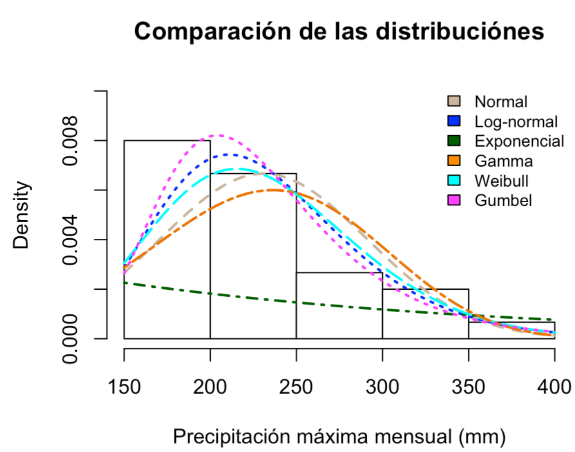
```
## Los parámetros estimados partiendo de la media y la desviación estándar son beta =  46.68678  y miu =  203.9791
```

```
## Los parámetros estimados a partir de la función fitdistrplus() son beta =  44.81446  y miu = 204.0729
```

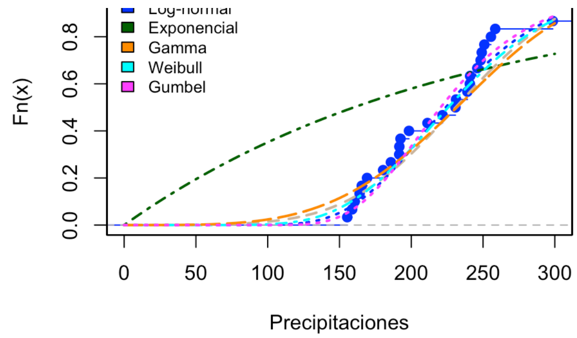
Se observan resultados similares pero diferentes. El paquete fitdistrplus en R utiliza métodos de estimación diferentes, como la máxima verosimilitud, para ajustar una distribución a los datos. La máxima verosimilitud estima los parámetros encontrando los valores que maximizan la probabilidad de observar los datos dados los parámetros, mientras que el método de momentos se basa en igualar las medias y varianzas teóricas y empíricas.

G. Compara los ajustes de las distribuciones que analizaste

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



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



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

Con base en los múltiples análisis de distribuciones realizados en búsqueda de la distribución que más se ajuste a los datos de precipitaciones máximas mensuales, se llegó a la conclusión de que la distribución Gumbel es la que más se ajusta a nuestro caso. Primero, si observamos el histograma y las densidades teóricas podremos notar que la única distribución que logra una gran densidad en su inicio y cuya cola cae de manera rápida al igual que los datos, es la distribución Gumbel. Aunado a esto, en la gráfica comparativa de probabilidades acumuladas podemos notar que la distribución que más varía, de forma que subestima y sobreestima en mayor medidad a las demás, es la distribución Gumbel. Son estos factores visuales, más la prueba de Kolmogorov-Smirnov los que nos hacen llegar a la conclusión de que la distribución que más se ajusta a nuestros datos es la Gumbel.

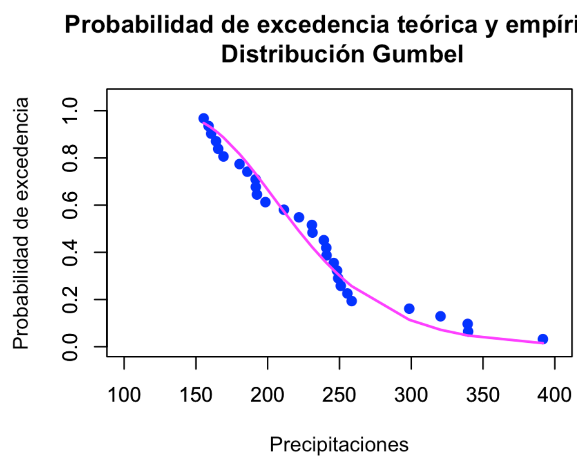
Diseño de obras hidráulicas

4. Precipitación de diseño de obras hidráulicas

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

El periodo de retorno recomendado para una presa derivadora en una zona de riesgo mediana es de entre 100-500 años. A manera de ejemplo, y por sugerencia de las mismas instrucciones se tomará en cuenta un periodo de retorno de 200 años para el análisis de precipitaciones máximas mensuales en Sinaloa para dicha obra hidráulica.

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



La gráfica muestra la probabilidad de excedencia de las precipitaciones, tanto teórica como empírica, ajustada a una distribución de Gumbel.

En el eje X están las precipitaciones, mientras que el eje Y representa la probabilidad de que las precipitaciones excedan cierto valor. Los puntos azules representan los datos empíricos de probabilidad de excedencia para distintos niveles de precipitación, mientras que la curva en rosa muestra la probabilidad de excedencia teórica según la distribución de Gumbel.

La distribución de Gumbel es comúnmente utilizada para modelar valores extremos, como precipitaciones intensas. Esto hace sentido a un punto mencionado, acerca de como las precipitaciones máximas mensuales han tendido a aumentar en los últimos años de manera regular en Sinaloa. En esta gráfica, la línea rosa indica cómo se ajusta la distribución de Gumbel a los datos empíricos, y podemos observar que aún y cuando el ajuste no es perfecto, la curva sigue el mismo comportamiento de los datos reales.

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

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

```
## La probabilidad de excedencia en un periodo de retorno de 200 años es  0.005
```

C. Conociendo la probabilidad de excedencia, calcula su complemento (1 - Pexe) y utiliza esta probabilidad

C. Conociendo la probabilidad de excedencia, calcula su complemento (1 - Pexc) 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.

```
## La precipitación máxima mensual que habrá en Sinaloa con un periodo de retorno de 200 años y probabilidad de excedencia del 0.5% es de  441.4019 milímetros.
```

D. El resultado de este ejemplo será una aproximación del caudal máximo que se tendría en Sinaloa 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 calculado representa el caudal máximo mensual que se espera que ocurra en Sinaloa con un periodo de retorno de 200 años. Esto significa que, en promedio, se espera que dicho caudal o uno mayor ocurra una vez cada 200 años. En términos de probabilidad, hay un 0.5% de que este evento se presente en un año cualquiera.

Si incrementamos el periodo de retorno, el valor de caudal máximo asociado también aumentará. Esto sucede porque al aumentar el periodo de retorno, estamos considerando eventos de mayor magnitud y menor probabilidad, los cuales ocurren con menor frecuencia. Por lo tanto, el diseño basado en un periodo de retorno mayor consideraría eventos extremos menos probables pero más intensos, aumentando el valor del caudal máximo de diseño. El caudal máximo estimado para un periodo de retorno específico (como los 200 años en este ejemplo) no será necesariamente el mismo en otros estados o regiones. Las características climáticas, topográficas y meteorológicas de cada región son distintas, por lo que los datos históricos y la distribución de probabilidad de eventos extremos varían en cada lugar. Por lo tanto, el caudal máximo para el mismo periodo de retorno puede ser considerablemente diferente en otro estado.

Diseñar obras hidráulicas (como presas, canales, y sistemas de drenaje) con base en periodos de retorno sugeridos permite anticipar eventos de intensidad extrema, asegurando que las infraestructuras sean seguras y eficaces en el largo plazo. Si se diseñan considerando periodos de retorno demasiado cortos, podrían fallar en caso de eventos más extremos. Por el contrario, elegir periodos demasiado largos puede resultar en un sobredimensionamiento y mayores costos de construcción y mantenimiento.

Conocer la distribución de probabilidad que mejor se ajusta a los datos históricos es esencial para realizar estimaciones precisas del caudal máximo en función del periodo de retorno. Diferentes distribuciones (como la Normal, Gumbel, Log-Normal, etc.) representan diferentes tipos de comportamiento en los datos, y una distribución incorrecta podría subestimar o sobrestimar los valores extremos. Esto afecta directamente la confiabilidad de los diseños hidráulicos y, en última instancia, la seguridad de la infraestructura.

Conclusión

Se presentó un análisis integral de las precipitaciones máximas mensuales en el estado de Sinaloa y destacó su relevancia para la Ingeniería Civil, específicamente en el diseño de obras hidráulicas. En el cual se sitúa correctamente el análisis dentro de la necesidad de prevenir daños en infraestructura debido a condiciones climáticas extremas. Los datos de precipitaciones son cruciales en la planificación de proyectos como presas, drenajes urbanos y puentes, ya que permiten dimensionar adecuadamente estos sistemas para soportar condiciones meteorológicas extremas.

Al centrarnos en un solo estado, Sinaloa, el análisis permite observar patrones específicos en las precipitaciones de este estado, lo que es útil para ajustarse a las particularidades geográficas y climáticas locales. Este enfoque detallado es valioso, pues facilita la identificación de tendencias de cambio a largo plazo, como el aumento en las precipitaciones observado.

La gráfica de las precipitaciones máximas mensuales proporciona una representación clara de la variabilidad y los picos en la precipitación a lo largo de tres décadas. La observación de un aumento en las precipitaciones máximas hacia los años recientes, con un pico en 2021, podría estar señalando cambios climáticos importantes o patrones de mayor frecuencia en eventos de lluvias extremas. Esta tendencia destaca la necesidad de ajustar los criterios de diseño de las infraestructuras para prevenir sobrecargas y fallos.

La información obtenida es fundamental para los ingenieros civiles y planificadores urbanos. Este tipo de estudios permite calcular la precipitación extrema probable para periodos de retorno específicos, lo cual es crucial en el diseño de obras que requieren asegurar la evacuación de grandes volúmenes de agua, minimizando el riesgo de desastres y maximizando la vida útil de estructuras esenciales como puentes y carreteras.

En conclusión, este análisis de precipitaciones históricas en Sinaloa es fundamental para enfrentar desafíos actuales y futuros en el diseño de infraestructura hidráulica. Al integrar datos históricos y métodos estadísticos, el trabajo proporciona herramientas para una planeación resiliente y basada en datos, que responde a las crecientes demandas de sostenibilidad y adaptación al cambio climático.

Referencias

[1] Servicio Meteorológico Nacional. (s.f.). https://smn.conagua.gob.mx/es/ (https://smn.conagua.gob.mx/es/)

[2] Más práctica Menos Teoría. (2017, 2 agosto). Distribución Weibull y Gumbel Ejercicio (Excel) - Hidrología [Vídeo]. YouTube. https://www.youtube.com/watch?v=WXSIIFcsAFE (https://www.youtube.com/watch?v=WXSIIFcsAFE)

[3] El modelo Weibull: reforzando el concepto de confiabilidad | ASTM Standardization News: Espanol. (s. f.). https://sn.astm.org/esp/data-points/el-modelo-weibull-reforzando-el-concepto-de-confiabilidad-ma20.html# (https://sn.astm.org/esp/data-points/el-modelo-weibull-reforzando-el-concepto-de-confiabilidad-ma20.html#):~:text=La%20distribuci%C3%B3n%20de%20Weibull%20tiene,t)%20tendr%C3%A1%20una%20forma%20diferente.

[4] SPSS Statistics Subscription - Classic. (s. f.). https://www.ibm.com/docs/es/spss-statistics/saas?topic=tests-one-sample-kolmogorov-smirnov-test (https://www.ibm.com/docs/es/spss-statistics/saas?topic=tests-one-sample-kolmogorov-smirnov-test)

[5] Montalar, E. (2012, 29 octubre). ¿Qué es el periodo de retorno y por qué se utiliza como una probabilidad? – Enrique Montalar. https://enriquemontalar.com/que-es-el-periodo-de-retorno-probabilidad/ (https://enriquemontalar.com/que-es-el-periodo-de-retorno-probabilidad/)