Erika Martínez Meneses 2024-10-22

Actividad Integradora - Obras Hidráulicas

Planteamiento del Problema

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 analizarán 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. En los siguientes análisis se analizan las precipitaciones históricas del estado de Morelos.

Introducción

El diseño y planificación de obras de infraestructura, especialmente en el ámbito de la ingeniería civil e hidráulica, dependen en gran medida del análisis de los factores climatológicos, entre los que destacan la precipitación y temperatura. La consideración de eventos de precipitación máxima es fundamental para diseñar estructuras como presas, puentes, y sistemas de drenaje que deben resistir y manejar altos volúmenes de agua sin comprometer la seguridad y durabilidad de la obra (Chow, Maidment, & Mays, 1988). Para lograr un diseño adecuado, se estima la precipitación máxima probable asociada a un periodo de retorno específico, que representa la frecuencia con la cual un evento de cierta magnitud puede ser igualado o superado (Salas et al., 2016).

En la práctica hidrológica, la elección de periodos de retorno para los cálculos de diseño depende de los usos de cada obra y de la gravedad de los posibles impactos. Por ejemplo, el diseño de un sistema de drenaje urbano puede requerir un periodo de retorno de 10 años, mientras que una presa podría requerir periodos de hasta 100 años o más para mitigar riesgos (U.S. Bureau of Reclamation, 1987). Este estudio se enfoca en analizar los datos históricos de precipitaciones mensuales máximas del estado de Morelos, México, entre los años 1994 y 2023, con el fin de determinar la precipitación extrema para un periodo de retorno seleccionado, utilizando métodos de ajuste y comparación de distribuciones estadísticas.

Metodología

El análisis se lleva a cabo en cuatro pasos principales, cada uno enfocado en el ajuste de los datos de precipitación mensual a diferentes distribuciones de probabilidad y en la interpretación de las gráficas de ajuste. A continuación, se detalla cada fase del análisis, las preguntas abordadas y las interpretaciones de los resultados obtenidos.

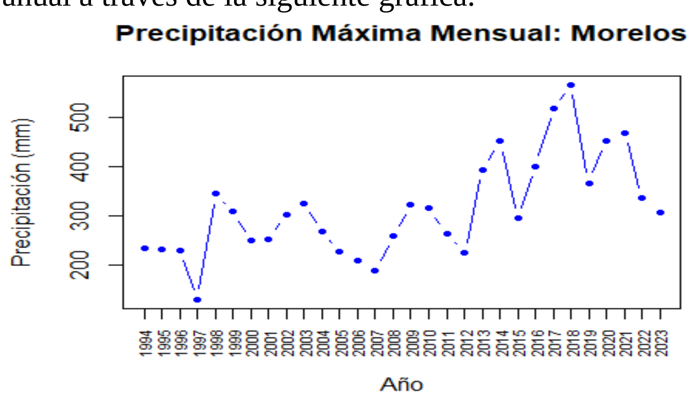
1. Análisis estadístico descriptivo de las precipitaciones históricas máximas mensuales

A.

Selección de Estado: **Morelos**

B.

Se calculó la máxima precipitación mensual para cada año en Morelos y se representó gráficamente la tendencia anual a través de la siguiente gráfica.



Para cada año se seleccionó el mes con mayor registro de precipitación. Los resultados mostraron una variabilidad considerable, sugiriendo fluctuaciones en las precipitaciones anuales máximas sin una tendencia evidente, esto es común en eventos climáticos debido a su naturaleza aleatoria y dependiente de múltiples factores atmosféricos. El análisis de estas variaciones es crucial, pues permite prever la magnitud de precipitaciones a la que debe estar preparada una infraestructura.

C.

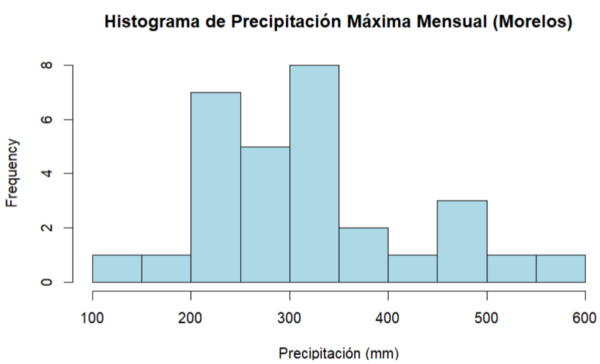
Asimismo, se calcularon las medidas de tendencia central y de dispersión (media, mediana, desviación estándar y rango).

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
131.2	239.1	305.2	315.2	361.3	565.0

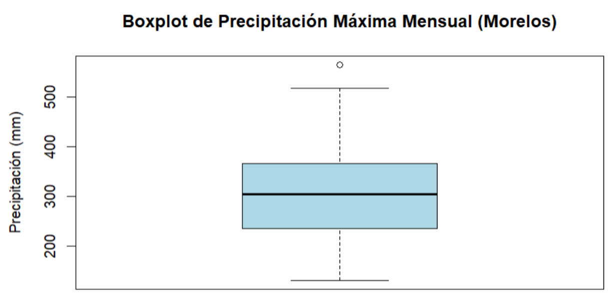
Media: 315.19 **Mediana**: 305.25 **Desviación Estándar**: 100.787 **Rango**: 131.2 565

La media y la mediana son relativamente cercanas, lo cual sugiere una distribución simétrica de los datos, aunque el rango y la desviación estándar indican cierta dispersión. La simetría sugiere que los eventos de precipitación extrema no son especialmente frecuentes, pero la desviación estándar y el rango muestran que existen algunos valores extremos que aumentan la variabilidad.

Se realiza un histograma y un diagrama de caja (boxplot) para describir la distribución de las precipitaciones máximas mensuales.



El histograma permite observar la frecuencia de las precipitaciones en cada intervalo. Se muestra que la mayoría de los valores se concentran entre 200 y 350 mm, con un pico de frecuencia en el rango de 300 a 350 mm. Esto sugiere que los meses con precipitaciones máximas moderadas son los más comunes. La distribución presenta una ligera asimetría hacia la derecha, ya que hay menos valores en los rangos superiores (por encima de 400 mm), y se observan algunos valores aislados en el rango de 500 a 600 mm, que representan meses con precipitaciones excepcionalmente altas.



El boxplot confirma la concentración de valores en torno a la mediana, ubicada entre 300 y 350 mm, y muestra que el rango intercuartílico (IQR) abarca de 239.1 a 361.3 mm. Esto indica que la mayoría de los datos están en un rango moderado, aunque hay un valor atípico, representado por un punto fuera del bigote superior, alrededor de los 500 mm. Este valor fuera del rango típico sugiere la ocurrencia de meses con precipitaciones significativamente superiores al promedio, destacando variaciones importantes en la precipitación máxima mensual.

D. Conclusión

En general, el análisis de la precipitación máxima mensual en Morelos revela que, aunque los valores de precipitación máxima varían considerablemente de un año a otro, no se observa una tendencia clara de aumento o disminución en el tiempo. Esta variabilidad puede atribuirse a la naturaleza aleatoria y a los múltiples factores atmosféricos que influyen en los eventos climáticos extremos. Las medidas de tendencia central sugieren una distribución simétrica de los datos, con la mayoría de los valores concentrados en un rango moderado, aunque existen algunos valores extremos que indican meses con precipitaciones excepcionalmente altas.

El histograma y el boxplot muestran que las precipitaciones máximas mensuales suelen concentrarse entre 200 y 350 mm, con algunos eventos que alcanzan valores más elevados, como los 500 mm, lo cual resalta la importancia de estar preparados para eventos atípicos. Analizar este tipo de datos es fundamental para entender los rangos de precipitación a los que puede estar expuesta la infraestructura, ayudando en la planificación y en la toma de decisiones para mitigar los posibles impactos de eventos climáticos extremos. Este análisis permite prever la magnitud de las precipitaciones máximas y proporciona una base para diseñar estrategias de prevención y adaptación ante variaciones en el clima.

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.

El análisis continúa ordenando las precipitaciones máximas de mayor a menor y calculando su probabilidad de excedencia utilizando el método de Weibull. La probabilidad de excedencia representa la probabilidad de que un evento iguale o supere un valor específico en un periodo de tiempo determinado.

A.

Como primer paso al data frame de los datos de precipitación máxima se agrega una columna con los datos de lluvias máximas ordenados de mayor a menor.

B.

Luego, 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.

A continuación, 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: $P_{exe} = \frac{m}{N+1}$.

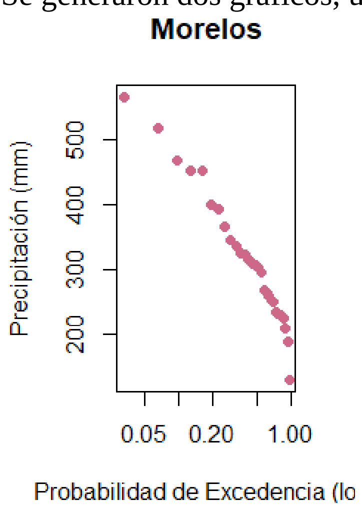
D.

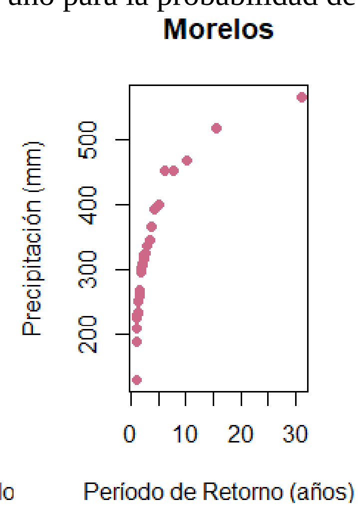
De ahí, se calcula la probabilidad de no excedencia para cada precipitación (complemento de la probabilidad de excedencia): $P_{noexe} = 1 - P_{exe}$.

E.

Y finalmente, se calcula el **periodo de retorno** como el inverso de la probabilidad de excedencia: $P_{ret} = \frac{1}{P_{exe}}$.

Se generaron dos gráficos, uno para la probabilidad de excedencia y otro para el periodo de retorno.





En la gráfica de Probabilidad de Excedencia, se observa cómo disminuye la probabilidad de que ocurran eventos de precipitación más intensos. Los valores altos de precipitación (por encima de 400 mm) presentan probabilidades de excedencia bajas, lo cual indica que son eventos menos frecuentes y, por lo tanto, más extremos. Esto sugiere que los meses con precipitaciones elevadas son menos comunes, destacando la importancia de planificar la infraestructura para estos eventos de baja frecuencia, pero alta intensidad, ya que representan riesgos significativos si no se considera su posibilidad.

En la gráfica de Periodo de Retorno, se visualiza el número esperado de años entre eventos de diferentes magnitudes de precipitación. Los eventos de alta precipitación (por encima de 400 mm) tienen periodos de retorno largos, lo que implica que son menos probables de ocurrir anualmente y suelen repetirse cada varias décadas. Esto es fundamental en hidrología para estimar la frecuencia de eventos extremos y definir el periodo de retorno deseado en el diseño de infraestructura, ya que un periodo de retorno alto garantiza un mayor nivel de seguridad frente a eventos raros, pero de gran impacto. Ambas gráficas son esenciales para comprender el comportamiento de eventos extremos de precipitación en Morelos y permiten establecer criterios de diseño adecuados para minimizar los riesgos de inundaciones o daños en infraestructuras críticas.

F.

La probabilidad de excedencia es una medida que indica la probabilidad de que una precipitación máxima mensual sea igualada o superada en un periodo determinado. En otras palabras, es la probabilidad de que ocurra un evento de alta precipitación o uno aún mayor. Valores bajos de probabilidad de excedencia corresponden a eventos menos frecuentes y más extremos. Este análisis es esencial para determinar el nivel de preparación que necesita una infraestructura para enfrentar eventos de alta magnitud que, aunque infrecuentes, podrían tener consecuencias graves.

El periodo de retorno representa el número esperado de años en los que un evento de una magnitud específica podría ocurrir al menos una vez. Es inverso a la probabilidad de excedencia: un periodo de retorno alto indica que el evento tiene una baja probabilidad de excedencia, por lo que ocurre con menor frecuencia. En el diseño de infraestructura, especialmente para obras críticas, se eligen periodos de retorno altos para asegurar un margen de seguridad adecuado y garantizar que la estructura resista eventos extremos que, aunque raros, podrían causar daños significativos.

En hidrología, ambos conceptos son cruciales, ya que permiten entender la frecuencia y magnitud de eventos de alta precipitación en el contexto del diseño de obras hidráulicas y de drenaje. Para el diseño de una obra, se suelen desear valores bajos en la probabilidad de excedencia, ya que estos corresponden a eventos extremos que deben considerarse para garantizar que la infraestructura sea resistente a condiciones excepcionales, asegurando así su durabilidad y funcionamiento en situaciones críticas.

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.

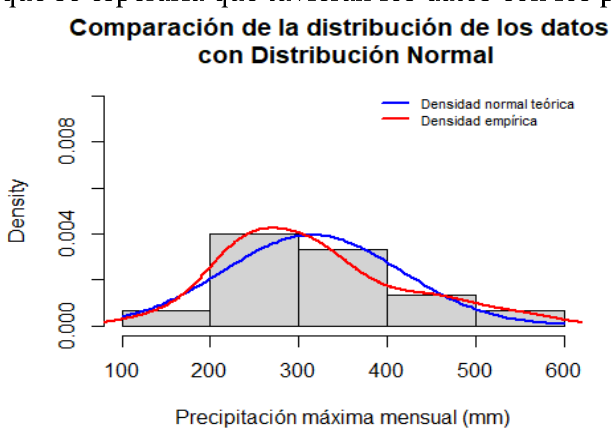
En este análisis se emplean distribuciones Normal, Log-normal, Exponencial, Gamma, Weibull y Gumbel. Para cada distribución, se realiza un ajuste visual mediante histogramas y gráficos de probabilidad acumulada, y se aplican pruebas de bondad de ajuste como Kolmogorov-Smirnov (KS) y Shapiro-Wilk cuando es aplicable.

A. Ajuste a una Distribución Normal.

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

A.1

Se construye el histograma de la función de densidad empírica de los datos y se sobrepone una distribución normal que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos.



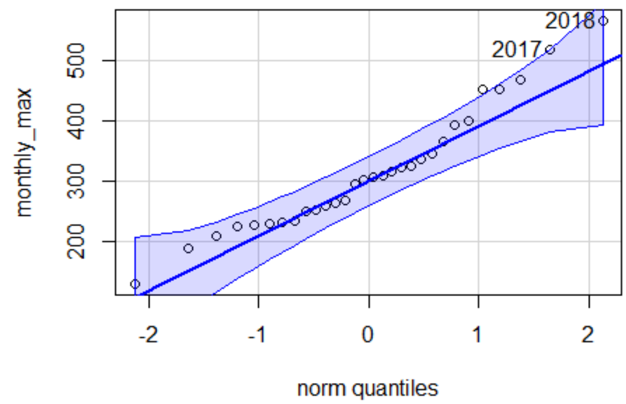
Al graficar la densidad empírica y superponer la curva de una distribución normal, se observa que la curva normal no sigue con precisión la dispersión de los datos, especialmente en los extremos. Sin embargo aunque existe alguna variación en las colas de la distribución empírica respecto a la normal teórica, la densidad general muestra una forma similar a la curva de distribución normal.

La distribución normal tiene dos parámetros: media (mu) y desviación estándar (sigma). Estos parámetros se calculan directamente de los datos (media y desviación estándar) porque son estimadores de momentos.

A.2

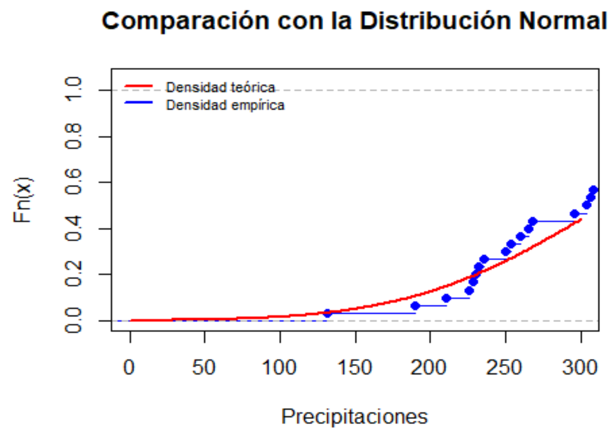
Se construye la gráfica qqplot.

```
## 2018 2017
## 25 24
```



Este gráfico compara los cuantiles teóricos de una distribución normal con los cuantiles de monthly_max. La línea de ajuste y el intervalo de confianza indican que, en general, los puntos siguen una línea recta, lo que sugiere una aproximación razonable a la normalidad. Algunas desviaciones menores en los extremos podrían estar presentes, pero no son lo suficientemente significativas como para rechazar la normalidad, según los resultados de las pruebas estadísticas.

A.3



Se 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.
Los datos empíricos son los datos observados directamente, mientras que los datos teóricos provienen de una función teórica (como la normal) ajustada a los parámetros estimados. La función de distribución acumulada empírica se aproxima bien a la función de distribución acumulada teórica de la distribución normal, lo cual refuerza la interpretación de que los datos pueden ser modelados bajo esta distribución.

A.4

Se utilizan dos pruebas de bondad de ajuste: Shapiro-Wilks y Kolmogorov-smirnov (KS).

Prueba de hipótesis

$\boldsymbol{H_0}$: Los datos provienen de una distribución normal $\boldsymbol{H_1}$: Los datos no provienen de una distribución normal

Regla de decisión: Se rechaza H_0 si valor p < $\alpha = 0.05$

Shapiro-Wilk normality test

W = 0.94996, **p-value** = 0.1687

El p-valor obtenido es mayor al nivel de significancia. Esto implica que no podemos rechazar la hipótesis nula de que los datos provienen de una distribución normal. En otras palabras, la prueba de Shapiro-Wilk no encuentra evidencia significativa para concluir que la distribución de monthly_max se desvíe de la normalidad.

Exact one-sample Kolmogorov-Smirnov test

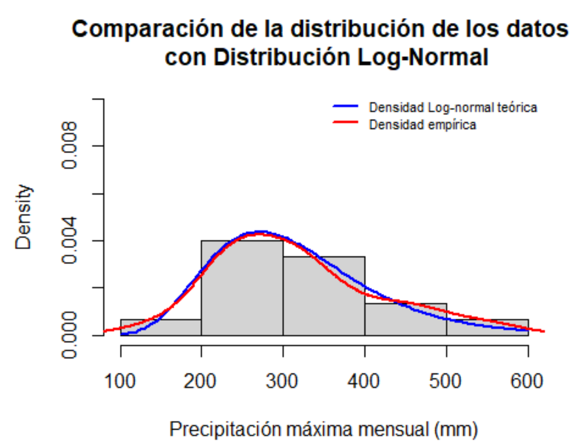
D = 0.12396, **p-value** = 0.7

El valor D sugiere que la distribución normal no ajusta perfectamente los datos, pero el ajuste es moderadamente aceptable. El p-valor alto (0.7) indica que no podemos rechazar la hipótesis nula. Esto refuerza la conclusión de que la precipitación máxima mensual podría ajustarse razonablemente a una distribución normal, ya que esta prueba tampoco encuentra desviaciones significativas respecto a la normalidad.

B. Ajuste a una Distribución Log-Normal.

Se realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Log-normal para ajustar los datos.

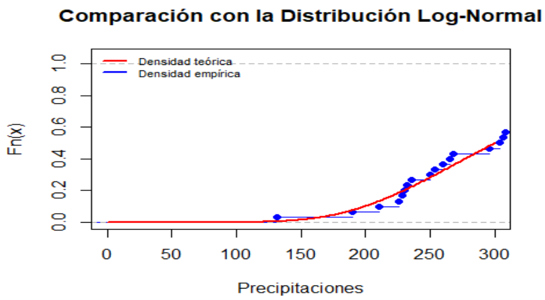
B.1



Se 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.
Al superponer la curva Log-normal en el histograma, se aprecia que esta distribución captura mejor los valores en los extremos en comparación con la distribución normal. Este ajuste inicial sugiere que la distribución Log-normal puede ser más adecuada para estos datos.
La distribución Log-normal tiene dos parámetros: la media logarítmica (meanlog) y la desviación estándar logarítmica (sdlog), que se calculan a partir de los logaritmos naturales de los datos.

B.2

Se 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.



Los datos empíricos son los valores observados directamente en la muestra, mientras que los datos teóricos son los valores estimados usando una distribución ajustada (Log-normal en este caso). La probabilidad acumulada empírica y la teórica de la Log-normal son similares en su forma general, pero existen desviaciones menores en los cuantiles más altos, lo cual sugiere un ajuste parcial.

B.3

Se hace la prueba KS para determinar si los datos se ajustan a una Log-normal

Prueba de hipótesis

$\boldsymbol{H_0}$: Los datos provienen de una distribución Log-normal $\boldsymbol{H_1}$: Los datos no provienen de una distribución Log-normal

Regla de decisión: Se rechaza H_0 si valor p < $\alpha = 0.05$

Exact one-sample Kolmogorov-Smirnov test

D = 0.084623, **p-value** = 0.9702

La prueba KS con un p-valor mayor a 0.05 sugiere que no hay suficiente evidencia para rechazar la hipótesis de que los datos provienen de una distribución Log-normal, lo cual indica un ajuste razonable.

B.4

El estadístico de prueba mide la máxima diferencia entre las distribuciones empírica y teórica. El valor D = 0.084623 indica que la distribución Log-Normal se ajusta mejor a los datos en comparación con la distribución normal.

Si el p-valor es menor que el nivel de significancia, rechazamos la hipótesis nula, lo que sugiere que los datos no siguen una distribución Log-normal. Si el p-valor es mayor, no podemos rechazar la hipótesis nula y podemos suponer que los datos siguen una Log-normal, siendo esto último nuestro caso.

En nuestro caso, la distribución Log-normal, con sus parámetros de media y desviación estándar logarítmicas, se ajusta mejor a los datos debido a la naturaleza sesgada de la precipitación extrema, que suele tener más valores en los extremos altos.

B.5

La distribución Log-normal tiene dos parámetros:

- Media logarítmica (meanlog): Es el promedio de los logaritmos naturales de los datos.
- Desviación estándar logarítmica (sdlog): Es la desviación estándar de los logaritmos naturales de los datos.

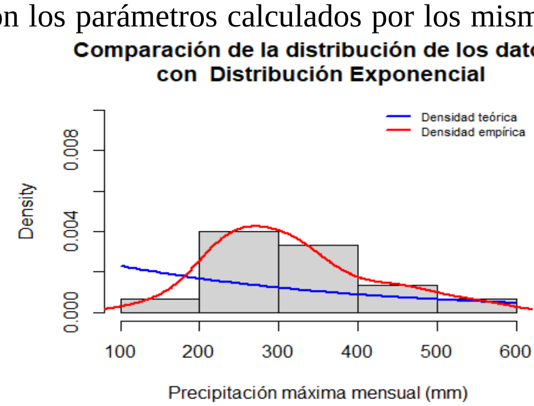
Estos parámetros se calculan utilizando el método de momentos, donde la media y la varianza se ajustan a los logaritmos de los datos observados. Esto se justifica porque la Log-normal es una distribución continua donde el logaritmo natural de una variable aleatoria sigue una distribución normal.

C. Ajuste a una Distribución Exponencial.

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

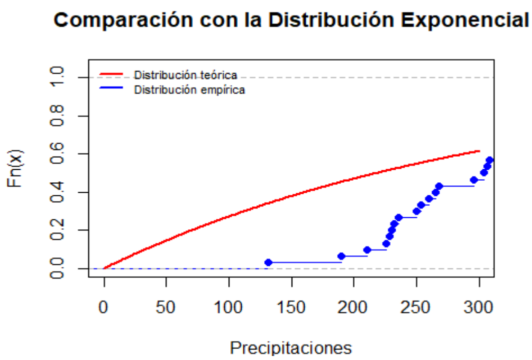
C.1

Como primer paso se contruye el histograma de la función de densidad empírica de los datos y sobrepone una distribución Exponencial que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos.



La densidad empírica, representada por la línea roja y el histograma, indica cómo se distribuyen realmente los datos de precipitación en el conjunto analizado, mientras que la curva azul muestra la forma que tendría la distribución si los datos siguieran un modelo exponencial. La discrepancia entre ambas curvas sugiere que los datos de precipitación no siguen una distribución exponencial, ya que la densidad empírica presenta una mayor concentración de valores en torno a los 300 mm, y una caída más lenta que la que predice el modelo exponencial.

C.2



Se 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. Si los datos siguirieran una distribución exponencial, los puntos azules deberían ajustarse a la curva roja. Sin embargo, los puntos se desvían considerablemente de la línea teórica, lo que nuevamente sugiere que los datos no se ajustan a una distribución exponencial. La acumulación de puntos por debajo de la curva teórica en rangos bajos de precipitación y su desviación en rangos altos refuerza la conclusión de que la precipitación no sigue un modelo exponencial simple.

C.3

Se hace la prueba KS para determinar si los datos se ajustan a una Exponencial.

Prueba de hipótesis

$\boldsymbol{H_0}$: Los datos provienen de una distribución Exponencial. $\boldsymbol{H_1}$: Los datos no provienen de una distribución Exponencial.

Regla de decisión: Se rechaza H_0 si valor p < $\alpha = 0.05$

Exact one-sample Kolmogorov-Smirnov test

D = 0.42036, **p-value** = 0.00002547

Se rechaza H_0 ya que el valor p = 2.547e-05 < $\alpha = 0.05$

C.4

El valor D sugiere que la distribución exponencial no ajusta bien los datos, mostrando un ajuste bastante pobre en comparación con las otras dos distribuciones. El p-valor de la prueba KS es bajo, lo que sugiere rechazar la hipótesis de que los datos provienen de una distribución Exponencial.

La distribución Exponencial, con su parámetro de tasa, no modela adecuadamente la distribución de las precipitaciones máximas, debido a la falta de simetría y dispersión que presentan los datos observados.

C.5

Esta distribución tiene un solo parámetro: la tasa de decaimiento (rate), que es el inverso de la media de los datos calculado como $rate = \frac{1}{mean(monthly\_max)}$. Esto se debe a que la media de una distribución exponencial es $\frac{1}{rate}$.Este parámetro se calcula usando el método de momentos, que ajusta la media de los datos observados para estimar la tasa de la distribución exponencial.

Método de momentos

Para una distribución exponencial, el parámetro de tasa λ se estima como $\lambda = \frac{1}{media}$ porque la esperanza matemática de una variable aleatoria exponencial es $\frac{1}{\lambda}$. Por lo tanto, para calcular λ, se toma el inverso de la media de los datos.

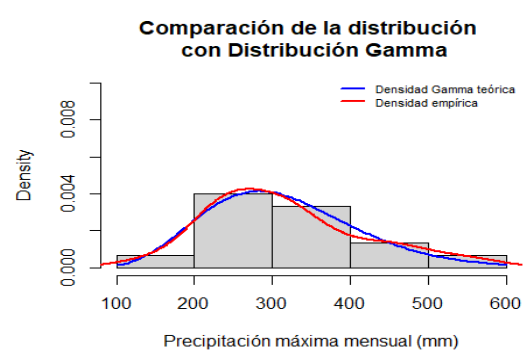
D. Ajuste a una Distribución Gamma.

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

D.1

Para comenzar se construye el histograma de la función de densidad empírica de los datos y se sobrepone una distribución Gamma que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos.

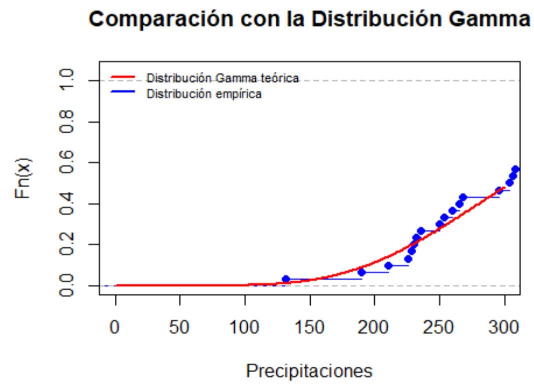
La cercanía entre ambas curvas, la empírica (línea roja) y la teórica Gamma, sugiere que la distribución Gamma podría ajustarse bien a los datos, ya que ambas curvas siguen una forma similar, especialmente en el rango central de los datos. Además, la curva Gamma ajusta



bastante bien, incluso mejor que algunas de las anteriores distribuciones, adaptándose bien a la dispersión y concentración de los valores altos en los datos.

D.2

Se 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.



La distribución acumulada teórica de la Gamma es similar a la ojiva empírica, lo cual indica que esta distribución tiene potencial para ajustar los datos. Se observa que los puntos azules siguen de cerca la línea teórica, lo que indica que los datos acumulados de precipitación máxima mensual son consistentes con una distribución Gamma. La cercanía entre la distribución empírica y teórica en ambas gráficas respalda el resultado de la prueba KS, donde no se rechaza la hipótesis nula y se concluye que una distribución Gamma es adecuada para modelar estos datos de precipitación.

D.3

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

Prueba de hipótesis

$\boldsymbol{H_0}$: Los datos provienen de una distribución Gamma. $\boldsymbol{H_1}$: Los datos no provienen de una distribución Gamma.

Regla de decisión: Se rechaza H_0 si valor p < $\alpha = 0.05$

Exact one-sample Kolmogorov-Smirnov test

D = 0.088477, **p-value** = 0.9565

La prueba KS muestra un p-valor mucho mayor a 0.05, sugiriendo que no se rechaza la hipótesis de que los datos provienen de una distribución Gamma.

D.4

La prueba de Kolmogorov-Smirnov (KS) se utiliza para evaluar si los datos observados de precipitaciones máximas mensuales siguen una distribución específica, en este caso, una distribución Gamma. El valor del estadístico de prueba KS es D = 0.088477, que mide la máxima diferencia entre la distribución acumulada empírica de los datos y la distribución teórica Gamma. El p-value de la prueba es 0.9565, lo cual es mucho mayor que el nivel de significancia (0.05). Dado este p-value alto, no hay suficiente evidencia para rechazar la hipótesis nula, que establece que los datos siguen una distribución Gamma. Por lo tanto, podemos concluir que no hay una diferencia significativa entre la distribución de los datos de las precipitaciones máximas mensuales y una distribución Gamma, lo cual sugiere que esta distribución podría ser un buen modelo para los datos observados.

D.5

La distribución Gamma tiene dos parámetros:

- Forma (shape) α, que determina la forma de la curva.
- Tasa (rate) β, que controla la dispersión de los datos.

Cálculo de los parámetros (método de momentos):

- Parámetro de forma (shape): $\alpha = \frac{\mu^2}{\sigma^2}$, donde μ es la media y σ^2 es la varianza de los datos.
- Parámetro de tasa (rate): $\beta = \frac{\mu}{\sigma^2}$, que es el inverso de la media ajustada por la varianza.

Estos parámetros se calculan de esta forma porque el método de momentos ajusta los primeros momentos (media y varianza) de los datos observados para encontrar los parámetros adecuados de la distribución Gamma.

E. Ajuste a una Distribución Weibull.

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

E.1

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

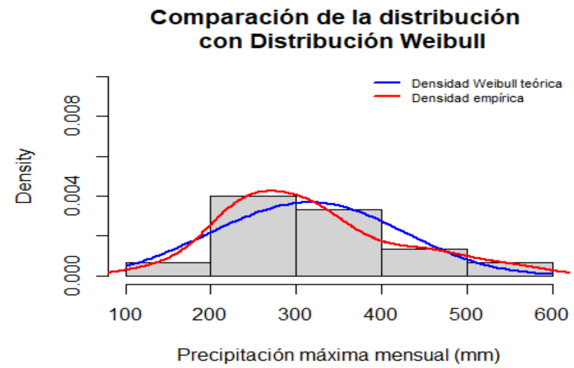
Parámetros estimados.- El comando fitdistr estima los dos parámetros de la distribución Weibull:

- Forma (shape): Controla la forma de la distribución.
- Escala (scale): Ajusta la dispersión de los datos.

E.2

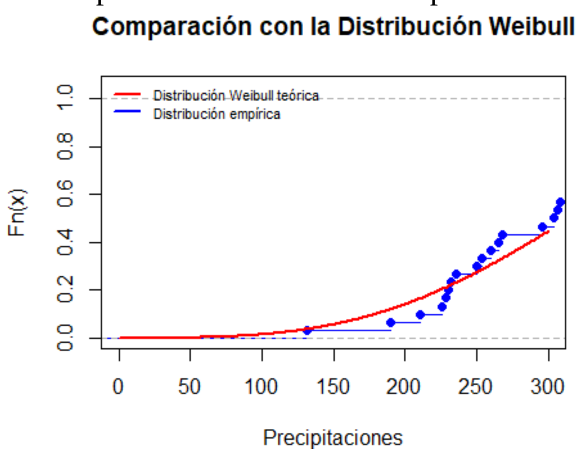
Se contruye el histograma de la función de densidad empírica de los datos y sobrepon una distribución Weibull que se esperaría que tuvieran los datos con los parámetros calculados por los mismos datos.

En esta gráfica la curva de densidad teórica (azul) y la densidad empírica (roja) muestran cierta semejanza en su forma general, pero existen discrepancias en algunos rangos, especialmente en los valores más altos y bajos de precipitación, donde la distribución empírica present a valores diferentes a los de la Weibull teórica.



E.3

Se 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.



Aquí, se presenta la función de distribución acumulada empírica (puntos azules) y la función acumulada teórica de la distribución Weibull (línea roja). Se observa que los puntos empíricos siguen relativamente de cerca la curva teórica, aunque hay pequeñas desviaciones en los extremos. Sin embargo, la cercanía entre ambas distribuciones acumulativas sugiere que la Weibull es razonablemente adecuada para modelar los datos, aunque quizás no tanto como la distribución Gamma, dado que las diferencias observadas en la densidad podrían afectar el ajuste en algunos rangos.

E.4

Se realiza la prueba KS para determinar si los datos se ajustan a una Weibull

Prueba de hipótesis

$\boldsymbol{H_0}$: Los datos provienen de una distribución Weibull. $\boldsymbol{H_1}$: Los datos no provienen de una distribución Weibull.

Regla de decisión: Se rechaza H_0 si valor p < $\alpha = 0.05$

Exact one-sample Kolmogorov-Smirnov test

D = 0.1244, **p-value** = 0.696

No rechazamos H_0 ya que el valor p = 0.696 < $\alpha = 0.05$ lo que sugiere que la distribución Weibull puede ser adecuada para los datos.

E.5

La prueba de Kolmogorov-Smirnov (KS) para la distribución Weibull nos permite evaluar si los datos de las precipitaciones máximas mensuales siguen una distribución Weibull. En este caso, el valor del estadístico de prueba D es 0.1244 y el valor p es 0.696. Dado que el valor p es considerablemente mayor que el nivel de significancia, no hay suficiente evidencia para rechazar la hipótesis nula, la cual establece que los datos provienen de una distribución Weibull. Por lo tanto, no rechazamos la hipótesis nula y concluimos que los datos podrían seguir una distribución Weibull, ya que la prueba KS no indica una diferencia significativa entre los datos y la distribución Weibull teórica.

E.6

La distribución Weibull tiene dos parámetros:

- Forma (shape): Define la forma de la distribución.
- Escala (scale): Ajusta la dispersión de los datos.

Complejidad en la estimación:

A diferencia de las distribuciones anteriores, la distribución Weibull no tiene una relación lineal simple entre sus parámetros y los momentos de la muestra (como la media y la varianza). Por lo tanto, se requieren técnicas numéricas como el método de máxima verosimilitud (usado en fitdistr) para estimar los parámetros.

F. Ajuste a una Distribución Gumbel.

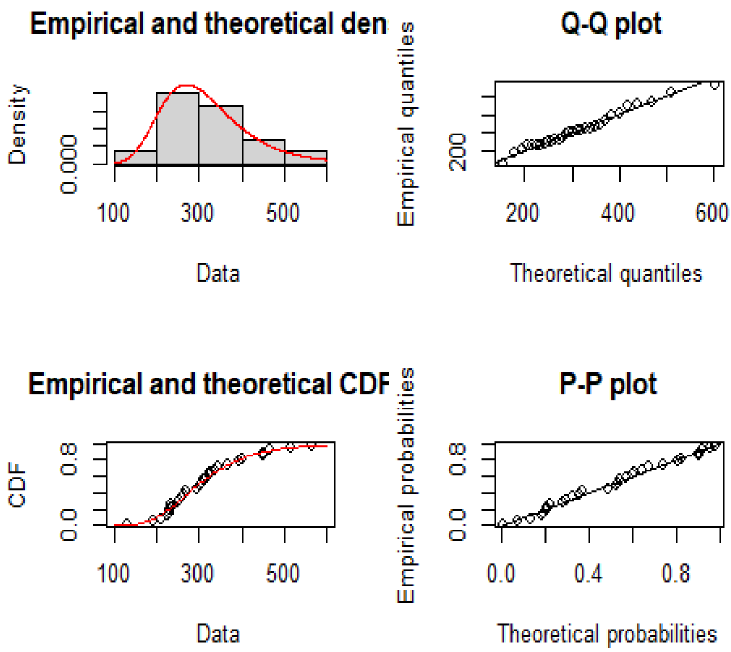
Finalmente se realiza un análisis visual y otro con pruebas de bondad de ajuste para determinar qué tan certera es la Distribución Gumbel para ajustar los datos.

F.1

Para probar si los datos de precipitación máxima se ajustan a una distribución Gumbel, se necesita definir las funciones de densidad de acuerdo con la función Gumbel. Se crean con las fórmulas de la Distribución Gumbel. dgumbel: Calcula la densidad para un valor x dado los parámetros α (ubicación) y β (escala). pgumbel: Calcula la probabilidad acumulada hasta un valor q. qgumbel: Calcula los cuantiles para una probabilidad p.

F.2

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



fitdist estima los parámetros α (ubicación) y β (escala) de la distribución Gumbel utilizando máxima verosimilitud.

Gráfica de Densidad (Empirical and Theoretical Density):

•La línea roja representa la densidad teórica de la distribución ajustada, mientras que el histograma muestra la densidad empírica de los datos.

•La proximidad entre la densidad empírica y la teórica indica que los datos siguen razonablemente la forma de la distribución teórica.

Gráfica Q-Q Plot:

•Muestra una comparación entre los cuantiles empíricos y los cuantiles teóricos.

•Los puntos se alinean en gran medida con la línea de referencia, lo que sugiere que los datos empíricos se ajustan bien a la distribución teórica, con algunas desviaciones en los valores extremos.

Gráfica CDF (Empirical and Theoretical CDFs):

- Compara la función de distribución acumulada empírica (CDF) con la teórica.
- La coincidencia entre las líneas empírica y teórica sugiere que los datos siguen la distribución teórica a lo largo de la mayoría de los valores, reforzando la adecuación del modelo.

Gráfica P-P Plot:

- Muestra las probabilidades acumuladas empíricas frente a las teóricas.
- Los puntos se alinean bastante bien con la línea de referencia, lo que confirma que la distribución teórica modela adecuadamente los datos.

Los parámetros estimados de la distribución son:

	Estimate	Std. Error
a	269.10709	15.68978
b	81.51226	11.41729

Estos valores representan los parámetros ajustados de la distribución, y el bajo error estándar sugiere una estimación confiable de dichos parámetros. En conjunto, las gráficas y los valores de los parámetros indican que los datos de precipitaciones máximas mensuales siguen adecuadamente la distribución ajustada.

F.3

Se realiza el cálculo de la probabilidad de excedencia teórica con la distribución Gumbel

Se hace la prueba KS para determinar si los datos se ajustan a una Gumbel

Prueba de hipótesis

$\boldsymbol{H_0}$: Los datos provienen de una distribución Gumbel. $\boldsymbol{H_0}$: Los datos no provienen de una distribución Gumbel.

Regla de decisión: Se rechaza H_0 si valor p < $\alpha = 0.05$

Exact two-sample Kolmogorov-Smirnov test

D = 0.066667, **p-value** = 1

No rechazamos H_0 ya que el valor p = 1 < $\alpha = 0.05$ lo que sugiere que la distribución Gumbel puede ser adecuada para los datos.

F.4

La prueba de Kolmogorov-Smirnov (KS) aplicada para comparar la distribución de las probabilidades de excedencia de las precipitaciones máximas mensuales con una distribución Gumbel nos da información sobre si ambas muestras (la empírica y la teórica de Gumbel) provienen de la misma distribución. En este caso, el valor del estadístico de prueba D es 0.066667. Este valor mide la máxima diferencia absoluta entre las funciones de distribución acumuladas de ambas muestras. El valor de p es 1. Un p-value de 1 indica que no hay evidencia para rechazar la hipótesis nula.

Dado el alto valor de p, no se rechaza la hipótesis nula, lo que sugiere que no hay una diferencia significativa entre las distribuciones empírica y teórica. Por lo tanto, podemos concluir que las probabilidades de excedencia de las precipitaciones máximas mensuales podrían seguir una distribución Gumbel, ya que los datos no presentan una diferencia significativa respecto a esta distribución.

F.5

La distribución Gumbel tiene dos parámetros: Parámetro de ubicación (a) - Determina el punto donde se centra la distribución. Parámetro de escala (b) - Controla la dispersión de los datos.

Para la distribución Gumbel, la media y la desviación estándar están relacionadas con los parámetros α y β de la siguiente manera:

- Media: $\mu = a + b\gamma$, donde γ es la constante de Euler-Mascheroni
- Desviación estándar: $\sigma = \frac{\pi}{\sqrt{6}} b$

Ahora, se 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, y se comparan los valores obtenidos con los estimados con el comando "fitdistrplus". Nuestros valores son los siguientes

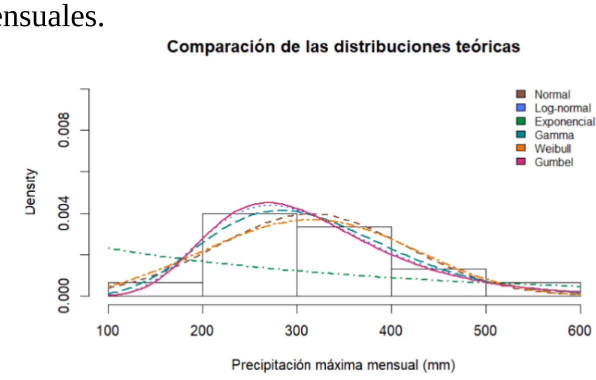
	Valores estimados	**Valores con el comando**
a	269.8317	269.10709, con un error estándar de 15.68978.
b	78.58333	81.51226, con un error estándar de 11.41729.

Al comparar los parámetros estimados para la distribución Gumbel obtenidos mediante la media y desviación estándar de los datos con los parámetros estimados utilizando el comando fitdistrplus, observamos que los valores son bastante cercanos. El parámetro de ubicación (a) estimado a partir de la media es 269.8317, mientras que el obtenido con fitdistrplus es 269.10709, con un error estándar de 15.68978, lo que indica que ambos valores son muy similares y dentro del margen de error. En cuanto al parámetro de escala (b), el valor estimado a partir de la desviación estándar es 78.58333, comparado con el estimado por fitdistrplus de 81.51226, con un error estándar de 11.41729. A pesar de una ligera diferencia, ambos parámetros de escala están dentro de un rango muy cercano, sugiriendo que ambas metodologías son consistentes y los resultados son compatibles para describir la distribución Gumbel de los datos de precipitaciones máximas mensuales.

G. Compara los ajustes de las distribuciones que analizaste

G.1

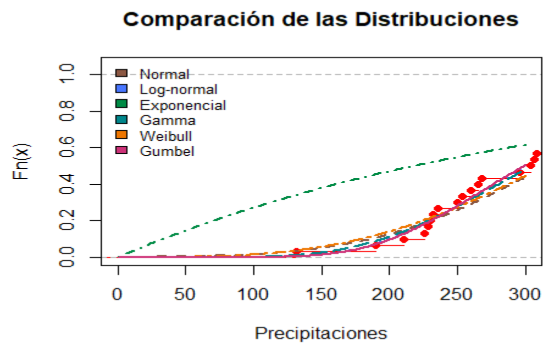
Esta gráfica muestra la densidad empírica de las precipitaciones máximas mensuales (en barras) superpuesta con las densidades teóricas de distintas distribuciones (Normal, Log-normal, Exponencial, Gamma, Weibull y Gumbel). A partir de esta comparación visual, se observa que algunas distribuciones, como



la Log-normal, la Gamma, la Weibull y la Gumbel, parecen ajustarse bien a la forma general de los datos, mientras que la Exponencial se aleja considerablemente, especialmente en los valores bajos de precipitaciones, donde tiene una pendiente mucho más pronunciada.

G.2

Esta gráfica muestra las funciones de distribución acumulativa (CDF) empíricas y teóricas para cada una de las distribuciones en estudio. Visualmente, las distribuciones Log-normal, Gamma, Weibull y Gumbel se ajustan bien a la distribución empírica, con líneas que siguen de cerca los puntos rojos (representando los datos). La Exponencial, nuevamente, se desvía de los datos, indicando un ajuste pobre, especialmente en el rango más bajo de precipitaciones.



G.3

Comparación de las diferentes Distribuciones

Distribución Normal:

Pruebas KS y Shapiro-Wilk: Los resultados de las pruebas KS (D = 0.12396, p-value = 0.7) y Shapiro-Wilk (W = 0.94996, p-value = 0.1687) sugieren que la hipótesis nula de normalidad no puede ser rechazada, indicando un posible ajuste. Sin embargo, visualmente, la Normal no parece seguir bien los extremos de los datos en comparación con otras distribuciones, especialmente en la parte derecha de la densidad.

Distribución Log-Normal:

Prueba KS: Con D = 0.084623 y un p-value de 0.9702, la prueba KS no rechaza la hipótesis nula, sugiriendo un buen ajuste. Visualmente, la Log-normal se ajusta bien tanto en la densidad como en la CDF, siguiendo de cerca la distribución de los datos. Este ajuste es particularmente notorio en el rango medio de los datos.

Distribución Exponencial:

La prueba KS muestra un D de 0.42036 y un p-value de 0.00002547, lo que rechaza la hipótesis nula, indicando un mal ajuste. Esto concuerda con las observaciones visuales, donde la Exponencial se aleja considerablemente de los datos empíricos.

Distribución Gamma:

La prueba KS presenta un D de 0.088477 y un p-value de 0.9565, lo cual indica que la distribución Gamma se ajusta bien a los datos. En las gráficas, la densidad y la CDF de la Gamma siguen de cerca los datos, haciendo que esta distribución sea una opción adecuada.

Distribución Weibull:

Prueba KS: Con D = 0.1244 y p-value de 0.696, no se rechaza la hipótesis nula, indicando un buen ajuste. Visualmente, la Weibull también parece ajustarse bien a los datos, siguiendo la forma de la densidad empírica y la CDF.

Distribución Gumbel:

Prueba KS: Con D = 0.066667 y un p-value de 1, la prueba KS sugiere un excelente ajuste. En las gráficas, la Gumbel muestra una gran proximidad a la forma de los datos en ambas visualizaciones, indicando que puede modelar adecuadamente las precipitaciones máximas mensuales.

Conclusión: Mejor Distribución para los Datos

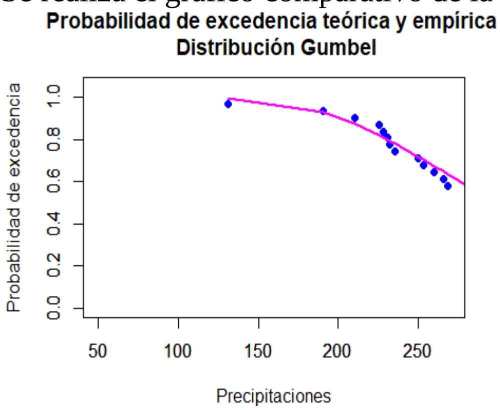
Considerando tanto las gráficas como los resultados de las pruebas KS, las distribuciones Log-normal, Gamma, Weibull y Gumbel presentan un buen ajuste a los datos de precipitaciones máximas mensuales. Sin embargo, la Gumbel destaca ligeramente al tener el p-value más alto en la prueba KS (p-value = 1), lo que indica el ajuste más fuerte entre las opciones evaluadas. Además, la distribución Gumbel es comúnmente utilizada en análisis de eventos extremos, como precipitaciones máximas, lo cual la hace aún más adecuada desde una perspectiva teórica. Por lo tanto, podemos concluir que la distribución Gumbel es la mejor opción para modelar las precipitaciones máximas mensuales, seguida de cerca por la Log-normal, Gamma y Weibull.

4. Discusión y Conclusiones - Precipitación de diseño de obras hidráulicas

Se desea diseñar una presa derivadora para una zona de riego mediana. En general, para una presa derivadora en una zona de riego mediana, el periodo de retorno recomendado es de 100 a 500 años.

A.

Se realiza el gráfico comparativo de la probabilidad de excedencia teórica vs empírica.



La cercanía de los puntos azules a la línea rosa indica que la distribución Gumbel está modelando correctamente la probabilidad de excedencia de las precipitaciones máximas. Los puntos azules siguen de cerca la línea rosa, especialmente en el rango central y superior, lo cual sugiere que la distribución Gumbel capta bien la forma general de los datos. Esto implica que la distribución teórica predice con precisión las probabilidades de que las precipitaciones excedan ciertos valores observados. La gráfica también muestra que la distribución Gumbel también se ajusta bien en los extremos altos de las precipitaciones, un aspecto importante en el análisis de eventos extremos.

La fuerte correspondencia entre la probabilidad de excedencia empírica y teórica refuerza la idoneidad de la distribución Gumbel para modelar estos datos de precipitaciones máximas mensuales. Esto coincide con los resultados de las pruebas de ajuste, donde la Gumbel mostró el mayor p-value en la prueba KS, indicando que es menos probable rechazar esta distribución para los datos.

En conclusión, la gráfica y las pruebas de ajuste sugieren que la distribución Gumbel es una elección adecuada para modelar las precipitaciones máximas mensuales de Morelos, especialmente al capturar eventos extremos de precipitación, los cuales son cruciales en el análisis de riesgo y planificación hidrológica.

B.

Utilizando el límite inferior del intervalo de periodo de retorno sugerido, se encuentra la probabilidad de excedencia o de ocurrencia para ese valor. Recordando que: $P_{ret} = \frac{1}{P_{exe}}$

C.

Conociendo la probabilidad de excedencia, se calcula su complemento $(1 - P_{exe})$ y se utiliza esta probabilidad para encontrar el valor de la precipitación máxima mensual que tendrá ese periodo de retorno.

El valor obtenido fue 644.0756, este representa la precipitación estimada para un evento con un período de retorno de 100 años según la distribución de Gumbel. En resumen, esto significa que la precipitación de diseño que se debería considerar para la infraestructura hidráulica, basada en la distribución Gumbel y un período de retorno de 100 años, es aproximadamente 644.0756 mm. Este valor es crucial para asegurarte de que las estructuras hidráulicas que se diseñen puedan manejar precipitaciones extremas de esa magnitud.

D.

El resultado es una aproximación del caudal máximo que se tendría en Morelos con un periodo de retorno de 100 años. Este valor representa la precipitación máxima mensual esperada en un evento extremo que ocurre, en promedio, una vez cada 100 años. Es decir, hay un 1% de probabilidad de que este valor se exceda en cualquier año dado.

Si se incrementa el periodo de retorno, por ejemplo a 500 años, el valor de la precipitación máxima mensual aumentará. Esto se debe a que eventos menos frecuentes (con periodos de retorno más largos) suelen ser más extremos. Por ejemplo, una tormenta con un periodo de retorno de 500 años tendría una precipitación máxima mayor que una de 100 años, reflejando condiciones más intensas y menos probables.

En el caso de datos históricos de otros estados, el caudal máximo para un periodo de retorno específico no será el mismo, ya que al utilizar los datos de otro estado, las condiciones climáticas, patrones de precipitación y variabilidad geográfica afectan los eventos de precipitación extrema. Cada región tiene características particulares que influyen en la distribución de probabilidad de las precipitaciones.

Otro aspecto que es importante considerar, es que las obras hidráulicas, como presas y canales, deben diseñarse considerando periodos de retorno adecuados para resistir eventos extremos esperados durante su vida útil. Por ejemplo, una presa en una zona agrícola podría necesitar diseñarse para soportar eventos con periodos de retorno de 100 a 500 años, dependiendo de su ubicación y el riesgo de inundaciones. Un periodo de retorno bajo podría resultar en infraestructura insuficiente para resistir eventos extremos, mientras que uno demasiado alto podría ser económicamente inviable.

Identificar la distribución de probabilidad que mejor ajusta los datos históricos (como la Gumbel en el caso de Morelos) permite hacer predicciones más precisas de eventos extremos. Esto es fundamental para asegurar que las infraestructuras están preparadas para las condiciones climáticas extremas que podrían ocurrir dentro del periodo de retorno.

Exploración de Otros Periodos de Retorno:

- Con 50: a = 587.1629 y Pnoexe = 0.98
- Con 900: a = 823.5403 y Pnoexe = 0.9988889

Al probar con otros periodos de retorno observamos que la precipitación máxima esperada aumenta conforme el periodo de retorno se hace más largo. Para un periodo de 50 años, la precipitación máxima es de 587.16 mm, mientras que para un periodo de 900 años, esta asciende a 823.54 mm. Esto refleja que eventos más extremos, aunque son menos probables, conllevan precipitaciones mayores. La probabilidad de no excedencia (como 0.98 para 50 años y 0.9989 para 900 años) muestra que a medida que aumenta el periodo de retorno, estamos considerando valores más raros, que ocurren con menor frecuencia, pero que tienen mayor impacto.

Esto nos refuerza la importancia de conocer la distribución de probabilidad adecuada y de diseñar obras hidráulicas con periodos de retorno sugeridos, ya que conocer la distribución de probabilidad que mejor ajusta los datos históricos es fundamental porque permite predecir con mayor precisión los eventos extremos y sus probabilidades de ocurrencia. Si se utiliza una distribución inadecuada, las estimaciones de eventos como lluvias intensas pueden ser incorrectas, llevando a un diseño de infraestructura que no es realista frente a los riesgos climáticos. En este caso, la distribución de Gumbel se eligió porque modela bien los eventos de precipitación extrema, proporcionando estimaciones confiables para los periodos de retorno seleccionados. Asimismo, las obras hidráulicas deben diseñarse para soportar los eventos más extremos que podrían ocurrir durante su vida útil, usando periodos de retorno que reflejen un balance entre seguridad y costo. Por ejemplo, una presa en una zona agrícola podría diseñarse para un periodo de retorno de 100 a 500 años, suficiente para mitigar el riesgo de inundaciones en eventos extremos esperados. Elegir un periodo de retorno demasiado corto podría resultar en una estructura vulnerable ante lluvias intensas, mientras que uno muy largo puede ser costoso e innecesario. Los periodos de retorno sugeridos aseguran que la infraestructura esté preparada para condiciones críticas sin incurrir en gastos excesivos o riesgos innecesarios.

En conclusión, para diseñar infraestructura de largo plazo, es crucial utilizar periodos de retorno adecuados que consideren el riesgo de eventos extremos y ajustar estos cálculos con la distribución de probabilidad correcta. Esto asegura que las estructuras no solo sean seguras, sino también con buena relación costo-beneficio, protegiendo a las comunidades y optimizando los recursos invertidos en las obras hidráulicas.

Referencias

- Chow, V. T., Maidment, D. R., & Mays, L. W. (1988). *Applied Hydrology*. McGraw-Hill.
- Salas, J. D., Obeysekera, J., Vogel, R. M., & England, J. F. (2016). *Climate Variability and Flood Frequency*. Journal of Hydrologic Engineering, 21(5).
- U.S. Bureau of Reclamation. (1987). *Design of Small Dams*.