

Instituto Tecnológico y de Estudios Superiores de Monterrey



**Tecnológico
de Monterrey**

Inteligencia artificial avanzada para la ciencia de datos II (Gpo 101)

**Estadística Avanzada | Actividad Integradora 1 - Precipitaciones máximas mensuales
para el diseño de obras hidráulicas**

Saul Francisco Vázquez del Río A01198261

“El cambio climático es un reto global que no tiene fronteras y que para combatirlo requiere el trabajo coordinado por parte de todos los países” (*¿Qué Es El Cambio Climático Y Cómo Nos Afecta?* | ACCIONA, n.d.-c) Por lo tanto es necesario que como país empecemos a poner atención en la forma que nos afecta en nuestra vida diaria. Es importante empezar a analizar los cambios de clima extremos que hemos sufrido y a los cuales no estamos acostumbrados.

Esto se debe a que la humanidad está destruyendo al planeta talando bosques, derritiendo los polos, destruyendo la capa de ozono y extinguiendo especies de animales. Además de esto los cambios de clima tan radicales han afectado al todo mundo habiendo sequías en países donde antes llovía mucho, enviando precipitaciones extremas a lugares en donde no están preparadas para estos.

Para comprender la importancia de estos cambios extremos se decidió realizar un estudio de un estado de México este siendo Coahuila, en donde se analizaran sus datos de precipitaciones a lo largo de 1994 a 2021 para poder predecir y estar preparados para una precipitación extrema, mejorar la infraestructura de las obra hidráulicas y ver cómo sería el peor de los casos ante una precipitación extrema.

Para entender el analisis que se realizo se creo una tabla que muestra los resultados de la prueba de Kolmogorov de varias distribuciones realizadas para ver cual de estas tiene un mejor ajuste de los datos teóricos (los de las distribuciones) contra los datos empíricos (los datos sin modificaciones). Además de esto se realizaron dos hipótesis H_0 siendo que los datos empíricos pertenecen a la distribución analizada y H_1 siendo que los datos no pertenecen a la distribución analizada.

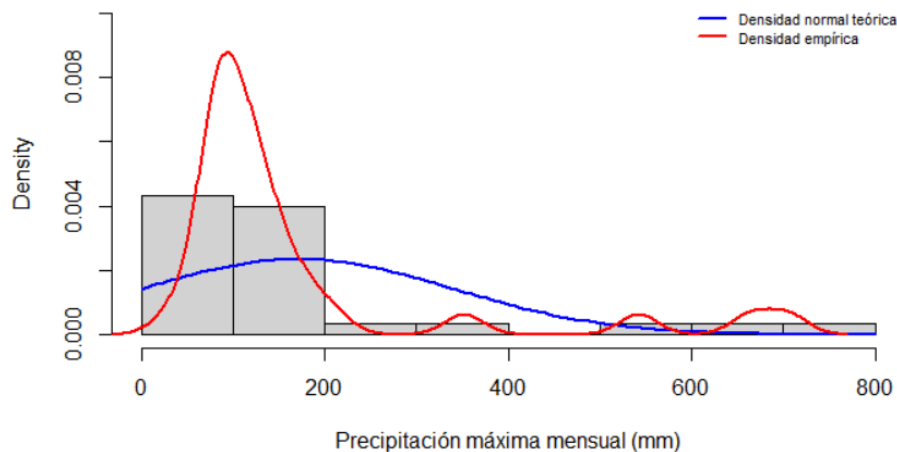
Para saber cuál hipótesis es la correcta se comparan los resultados de la prueba de Kolmogorov contra el valor de significancia 0.05 este valor determinará si acepta o se rechaza mediante si es mayor a 0.05 se acepta H_0 y se rechaza H_1 , en cambio que sea menor a 0.05 se rechaza H_0 y se acepta H_1 .

Distribuciones	D-value de KS	P-value de KS	Acepta o Rechaza H_0 o H_1
Normal	D = 0.31958	P-value = 0.00462	Acepta H_1 y se rechaza H_0
Log	D = 0.18497	P-value = 0.2562	No hay suficiente evidencia para rechazar H_0 por lo cual los datos se pueden ajustar
Exponencial	D = 0.29785	P-value = 0.009759	Se acepta H_1 y se rechaza H_0
Gamma	D = 0.29662	P-value = 0.0102	Se acepta H_1 y se rechaza H_0

Weibull	D = 0.23794	P-value = 0.06695	Se acepta H0 y se rechaza H1, buen ajuste de los datos
Gumbel	D = 0.23333	P-value = 0.3929	No hay suficiente evidencia para rechazar H0 por lo cual los datos se pueden ajustar

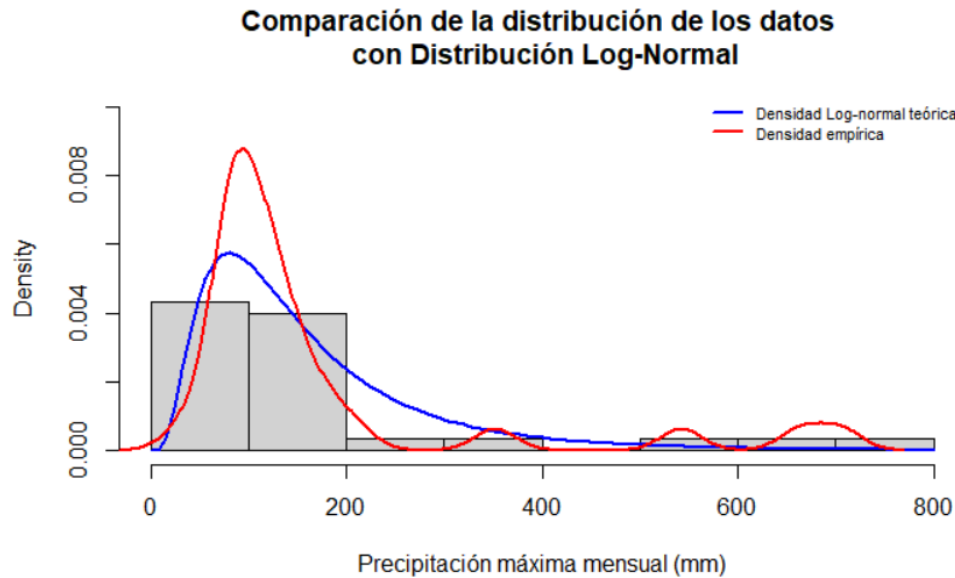
Ahora se tendrá un análisis de cada tabla de las distribuciones para observar porque los datos aceptan o rechazan H0 o H1.

Comparación de la distribución de los datos con Distribución Normal



Los datos no siguen una distribución normal, como se observa en la tabla la densidad empírica con la línea densidad normal teórica ambas son visiblemente distintas ya que existe varios valores extremos en los datos normales y en la densidad normal teórica esta tiende a seguir una normalidad en los datos, lo cual es incorrecto ya que Coahuila presentó cambios radicales en su precipitaciones.

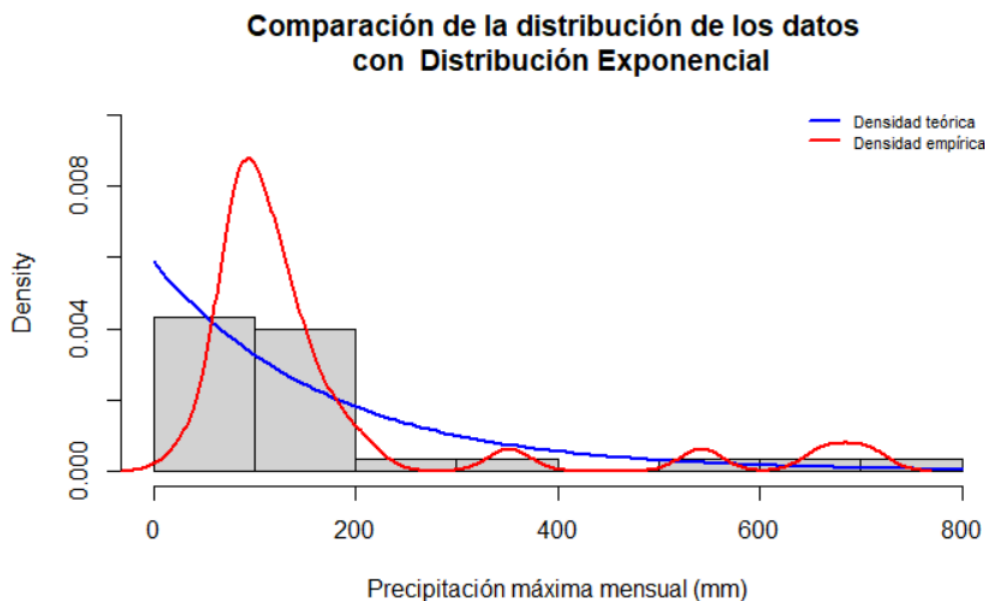
Además de esto gracias a la prueba de Kolmogorov se determina que los datos no pertenecen a esta distribución rechazando H0 y aceptando H1. Es posible que esta distribución hubiese funcionado si los datos de las precipitaciones de Coahuila no hubieran cambiado tan radicalmente.



Los datos visualmente nos sugieren que esta distribución puede ser aceptada por estos, ya que se observa una pequeña similitud entre la densidad empírica y la densidad log-normal. Esta tiende a seguir el patrón de los datos empíricos que aumenta al inicio de la gráfica y mediante que esta se expande bajando los datos.

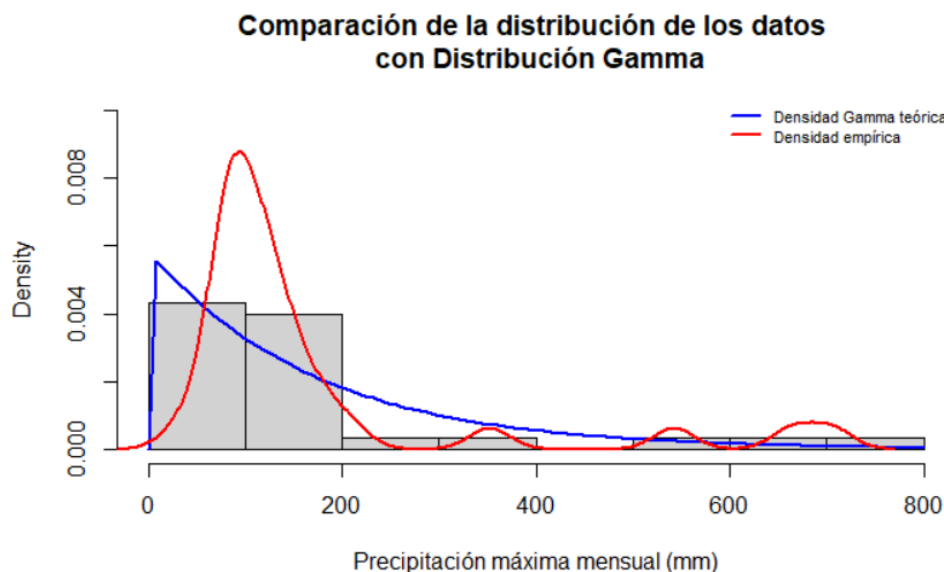
Los datos empíricos son los observados en las precipitaciones de Coahuila y los datos teóricos son los datos generados por la distribución log-normal. Ante esto esta distribución se puede presentar como uno de los modelos adecuados para predecir las precipitaciones en Coahuila, además con la prueba KS refuerza esta conclusión ya que el p-value respalda la hipótesis de H_0 aunque no tenga mucha evidencia para rechazarla los datos se ajustan correctamente.

Los parámetros usados en esta distribución son la media que se calcula con la media del logaritmo de las precipitaciones y la desviación estándar que se calcula con la desviación estándar de los logaritmos de las precipitaciones.



La línea de la densidad empírica no se ajusta adecuadamente a los datos de precipitación, ya que su forma difiere notablemente de la distribución observada. Los datos empíricos son los observados en las precipitaciones de Coahuila que no muestran similitudes con los datos teóricos los cuales son los datos generados por la distribución exponencial.

Con la prueba de KS podemos concluir que esta distribución no es la correcta para los datos de precipitación ya que p-value que arrojo es menor rechazando H_0 y aceptando H_1 haciendo que los datos no pertenezcan a una distribución exponencial. Los parámetros usados en esta distribución son lambda que se calcula uno entre la media de las precipitaciones.

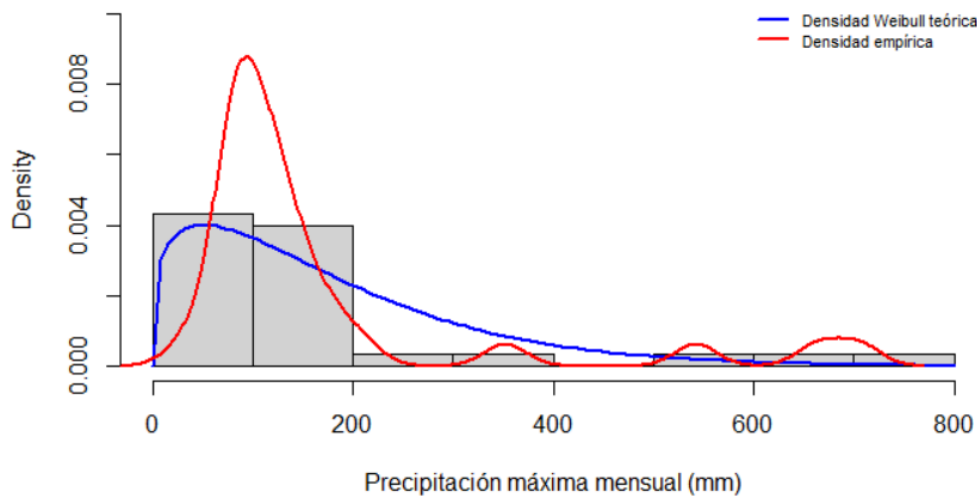


Los datos observados nos dicen que la línea de la distribución gamma teórica no se ajusta adecuadamente a los datos de la precipitación. Aunque está presente una mejor aceptación que la exponencial se logra observar claramente en el inicio que no se muestra el crecimiento necesario de los datos para reflejar el comportamiento de los datos empíricos, estos siendo los datos de la precipitación y los datos teóricos los generados por la distribución gamma.

Con los resultados obtenidos de la prueba KS podemos concluir que esta distribución no se ajusta adecuadamente a los datos ya que visualmente esto no encajan correctamente, además que su p-value es menor haciendo que se rechace H_0 y se acepta H_1 .

Los parámetros usados en esta distribución son shape representando la forma de la distribución calculado con la media cuadrada entre la varianza y el rate representado la tasa de distribución calculado igual con la media entre la varianza.

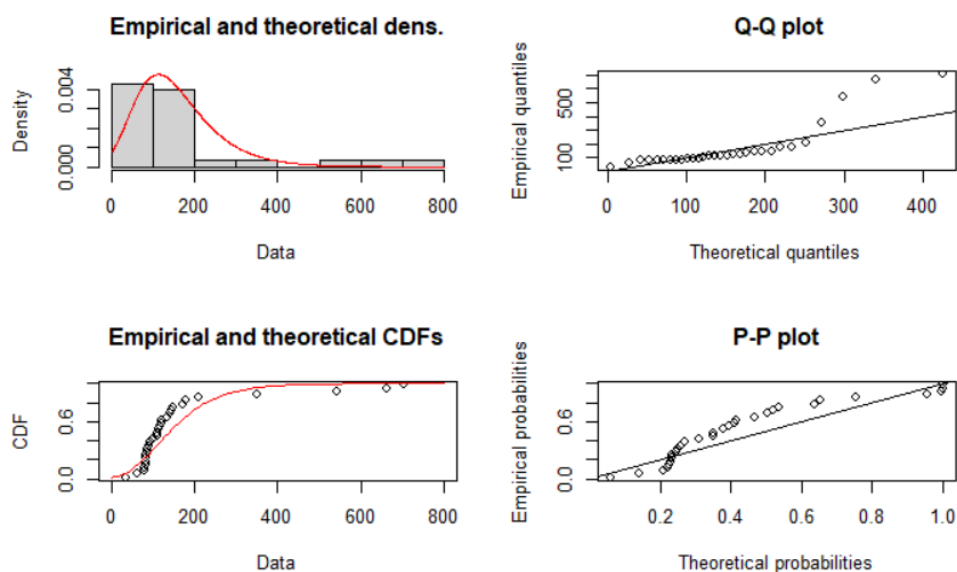
Comparación de la distribución de los datos con Distribución Weibull



Visualmente los datos de la distribución weibull no se ajustan tan claramente a los datos de precipitación. Aunque esta distribución muestra una mejor aceptación de estos ya que tienden a seguir el crecimiento del inicio de los datos y tienden a bajar por el final. Pero no sigue con exactitud los datos empíricos siendo estos las precipitaciones. Los datos teóricos creados por la distribución weibull tienen una mejor forma comparada con la gamma.

Esto lo podemos reafirmar con la prueba KS ya que el resultado del p-valor es mayor a la significancia lo que permite que los datos se puedan ajustar a la distribución weibull aunque esta tenga discrepancias, aceptando H_0 y rechazando H_1 .

Los parámetros usados en esta distribución como el shape se calcula a partir de las fórmulas, lo que representa la forma de la distribución y el parámetro de scale que indica la dispersión de los datos.



En la distribución gumbel gracias a la prueba de KS la cual obtuvo un p-value alto, indica que los datos pueden ajustarse a esta distribución permitiendo aceptar H_0 y rechazando H_1 . En esta distribución los parámetros clave son la ubicación y escala. El parámetro de

ubicación desplaza la distribución a lo largo de la gráfica, mientras que el parámetro de escala ajusta la dispersión de los datos, lo que permite una mayor flexibilidad del modelo.

Una vez analizado todas las distribuciones y viendo cual de estas tiene un mejor ajuste de los datos de las precipitaciones podemos llegar a decir que las distribuciones de log-normal teórica, weibull teórica y gumbel teórica fueron las mejores distribuciones de las seis que se analizaron en este documento. Estas distribuciones llegaron a ser de las mejores porque los datos de las precipitaciones logran tener un mejor ajuste en estas, además de que gracias a la prueba de Kolmogorov se confirmó que estas con su p-value fueron las mejores distribuciones.

Este documento fue hecho con el propósito de entender porque la recopilación de los datos de las máximas precipitaciones de un estado de México son importantes. Ante todo el análisis hecho y los descubrimientos realizados se llegó a la conclusión que sin este tipo de recopilación de datos las obras de hoy en día no serían lo mismo.

Ya que en el diseño de las obras hidráulicas, resulta crucial analizar la precipitación máxima mensual y su probabilidad de excedencia para garantizar la resistencia y funcionalidad de estas infraestructuras. Un paso importante es realizar las gráficas vistas en este documento que son las de probabilidad de excedencia y la empírica, evaluando una distribución acertada para tener una predicción confiable de las futuras predicciones.

Además de esto se realizó una estimación de las máximas predicciones dentro de un periodo de diez a veinte años en el estado de Coahuila, en donde la precipitación máxima del periodo de diez años fue de 284.4728 y el de veinte años fue de 339.7563. Dándonos un indicio que las precipitaciones van a aumentar lo que nos ayuda a atender el porqué es importante estar preparados ante este tipo de situaciones, evitando posibles problemas a futuro, mejorando la infraestructura y apoyando a la realización de este tipo de análisis.

En resumen este documento se creó con la finalidad de obtener la precipitación máxima de un estado de México, analizando qué distribución se ajusta a los datos de las precipitaciones que esta fueron log-normal, weibull y gumbel, tomando como primer lugar la distribución de log-normal por un menor d-value, además de tener un mejor ajuste de los datos visualmente luego gumbel como segundo lugar y por último en tercer lugar weibull. Además de comprender la importancia de las precipitaciones, el porqué es importante recopilar este tipo de datos y analizarlos para tener una mejor infraestructura y un mejor futuro.

Referencias bibliográficas

- Zizer. (2024, January 15). *Conservación y mantenimiento de la infraestructura hidráulica: ¿por qué es esencial?* Eductrade. <https://www.eductrade.com/conservacion-mantenimiento-infraestructura-hidraulica/>
- *¿Qué es el Cambio Climático y cómo nos afecta?* | ACCIONA. (n.d.). <https://www.acciona.com/es/cambio-climatico/>
- Gestor. (2024, March 4). *Ejecución de obras hidráulicas: beneficios para la sociedad y la economía.* Greening Aqua. <https://www.aegra.es/aegra-gestiona-proyectos-llave-en-mano-2/>