

Hicieron un buen trabajo, se nota que han asimilado los conceptos que hemos compartido en clase con ustedes, solo algunos detalles que les señalé en el informe que espero revisen con calma. Si bien hicieron un buen esfuerzo analizando la física del problema aun deben ser más detallados y tratar de ir más allá, pero eso es lo que van a aprender a lo largo del curso. Muy bien y espero que sigan mejorando :)

4.4

ANÁLISIS DE DATOS AMBIENTALES

Tarea 1: Estacionariedad de series de tiempo

David Muñoz Marín – CC:1152215701

Nactaly Posada Chavarría – CC: 1035833056

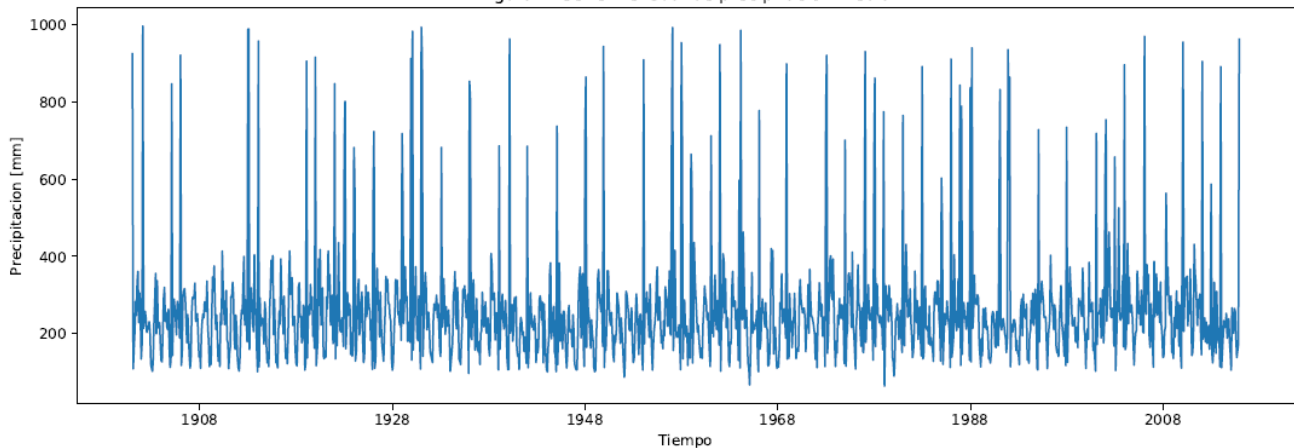
1. Serie temporal

Una serie temporal es un conjunto de datos que tienen asociados una variable temporal a cada una de las mediciones. Por ende, tienen una escala de tiempo determinada y un intervalo de tiempo entre dos mediciones definido y constante.

Cuando se está realizando un análisis de datos, es fundamental calcular y examinar las distribuciones de los datos y sus probabilidades de ocurrencia. Esto con el fin de visualizar de antemano si existe una gran dispersión de los datos, si los eventos extremos tienen relevancia en el fenómeno analizado y conocer los rangos de los valores que más probabilidad de ocurrencia tienen.

En la serie analizada, se puede observar que la mayor cantidad de datos (alrededor del 80%) están por debajo de los 400 mm de precipitación por mes y que la probabilidad de

Figura 1. Serie mensual de precipitación media



Dónde?

En este caso, se analizó una serie de precipitación media mensual, por lo que el intervalo de tiempo entre mediciones es de un mes. La longitud de la serie, definida como el número de datos que contiene, es 1380 y abarca las mediciones entre enero de 1901 y diciembre de 2015.

:O

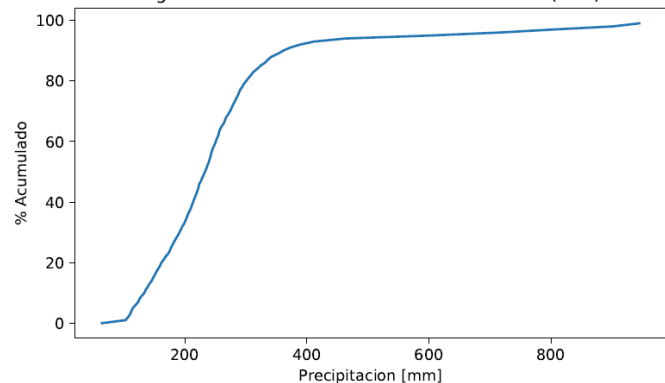
En la figura 1 se muestra la serie mensual de precipitación media que se usó para el análisis descrito a lo largo de este documento. Como se observa, no tiene ningún dato faltante y, a simple vista, parece bastante homogénea y estacionaria.

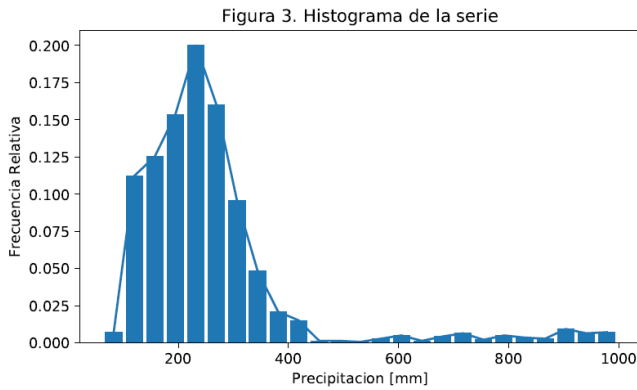
2. Función de distribución acumulada (FDA) e histograma de la serie

una precipitación de 500 mm/mes es casi igual a la probabilidad de una precipitación de 800 mm/mes.

Los anterior se ve reflejado en las figuras 2 y 3, que representan la función de distribución acumulada (FDA) y el histograma de la serie de datos respectivamente.

Figura 2. Función de distribución acumulada (FDA)





Por otro lado, se observa una gran concentración de los datos en el rango de 100 mm/mes y 300 mm/mes.

Finalmente, se observa que la probabilidad de que no llueva en el mes (o que llueva menos de 100 mm/mes) es casi nula, por lo que los valores extremos que se deben analizar de manera más exhaustiva son los de **mucha lluvia**.

3. Índices de localización

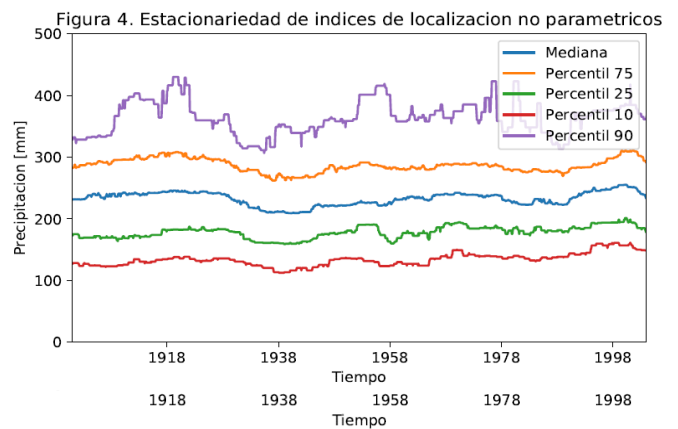
Los índices de localización son aquellos que dan un sentido de ubicación a los datos de una serie respecto a los demás datos. Entre ellos están la media (paramétrico), la mediana y los percentiles (no paramétricos).

Es de vital importancia identificar si los índices son paramétricos o no paramétricos, ya que, generalmente, los no paramétricos son más resistentes y robustos, es decir, sin importar la forma de la distribución de los datos, son igual de útiles porque sugieren la ubicación y magnitud de ciertos datos de particular interés.

Por otro lado, analizar la estacionariedad de los índices es una forma fácil de detectar alguna tendencia en la serie o, incluso, analizar si existe algún cambio en la varianza con el paso del tiempo.

En nuestro caso, se puede observar en las figuras 4 y 5 que, tanto los índices de localización paramétricos como los no paramétricos, son bastante estacionarios y no presentan alguna tendencia pronunciada en la serie.

Cabe destacar que los índices de localización para los valores extremos superiores tienen una menor estabilidad en el tiempo dado que son valores extremos que no siempre ocurren; sin embargo, los índices de localización media son más persistentes en el tiempo.

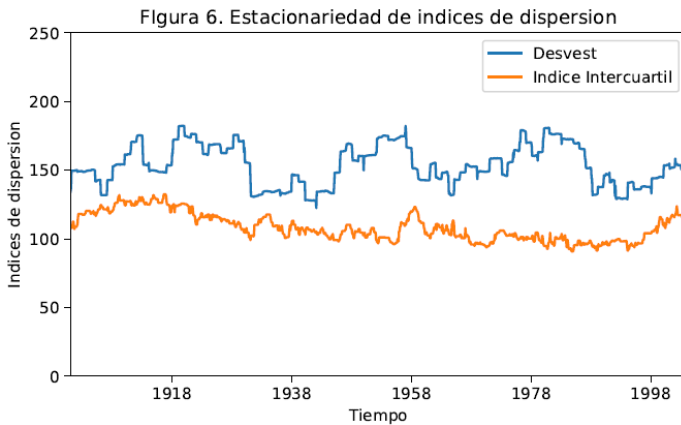


4. Índices de dispersión

A pesar de que los índices de localización pueden dar una información acerca de qué tanto varía la dispersión de los datos en el tiempo, es necesario hacer un análisis individual a aquellos parámetros que verifican el esparcimiento de los datos.

Algunos índices de dispersión son la varianza, la desviación estándar (paramétricos) y el rango inter-cuartil (no paramétrico). La estacionariedad de dichos índices se puede

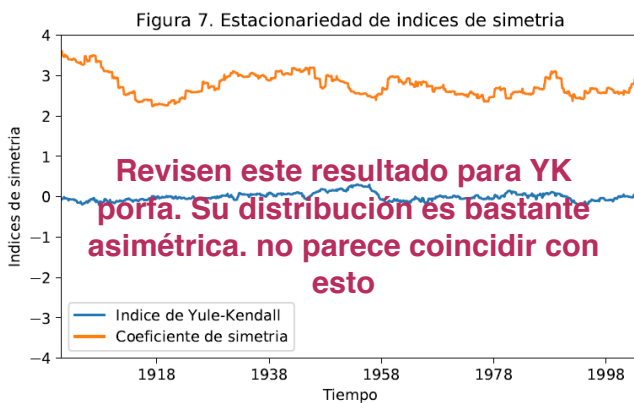
evaluar gráficamente y se muestran en la figura 6.



5. Índices de simetría

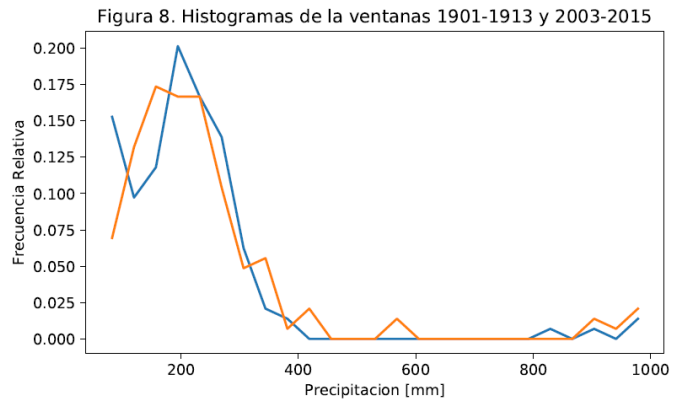
Los índices de simetría indican si los datos están recostados hacia algún valor extremo. Es decir, indican en qué valores está el peso de la distribución.

En este caso se analizó la estacionariedad del coeficiente de asimetría (paramétrico) y el índice Yule-Kendall (no paramétrico). La variación de estos en el tiempo se muestra en la figura 7.



Como se puede observar, el coeficiente de asimetría mantiene valores positivos a lo largo del tiempo, lo que quiere decir que se deja afectar por los valores extremos de alta precipitación, dejando la mayor cantidad de datos a la izquierda de la media, denotando asimetría.

Mientras que el índice de Yule-Kendall varía entre valores positivos y negativos, pero con cierta uniformidad alrededor del cero (entre -0,20 y 0,31) con el pasar del tiempo. Esto se debe a la naturaleza del índice, ya que este omite los valores extremos (con probabilidad de no excedencia mayor a 0,75) y, por ende, considera la distribución relativamente simétrica.



Como se puede observar en la figura 8, para dos ventanas de tiempo diferentes, la distribución de probabilidad, además de representar estacionariedad (como se muestra más adelante), presenta asimetría. Es por esto que el índice Yule-Kendall no representa adecuadamente la distribución de la serie de datos en el tiempo.

6. Tendencia en la serie

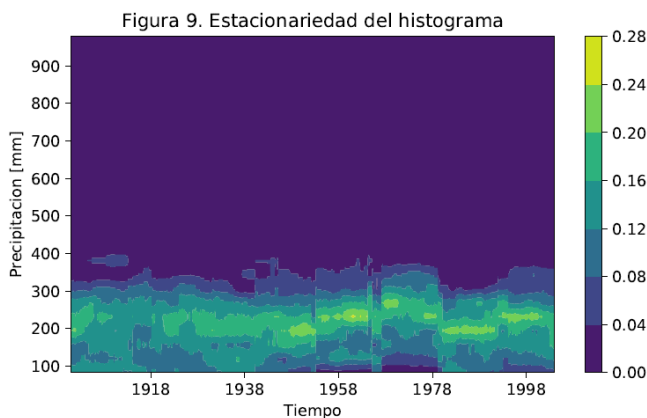
Para determinar si la serie tiene alguna tendencia es necesario realizar una prueba estadística. En este caso, se evaluó la serie con la prueba de Mann-Kendall de doble cola, la cual tiene las siguientes hipótesis:

- **Ho:** La serie no tiene tendencia.
- **Ha:** La serie tiene tendencia creciente o decreciente.

Dado que se definió la hipótesis alternativa de doble cola, se debe comparar el estadístico Z con el correspondiente a una probabilidad de $1-\alpha/2$, donde α es 0,05.

Los resultados, que se muestran en la tabla 1, arrojan un resultado del estadístico Z menor al correspondiente a una probabilidad de $1-\alpha/2$, por lo que el test no es concluyente respecto a la tendencia de la prueba.

Lo anterior era de esperarse si se observa la estacionariedad marcada de los histogramas en el tiempo, tal como lo muestra la figura 9, al notarse una franja horizontal de color verdoso que indica una alta probabilidad de ocurrencia de los valores de precipitación mostrados en el eje de las ordenadas.



7. Tendencia en la media

Es necesario evaluar la tendencia en la media, dado que es más acertado analizar la serie eliminando la variabilidad, ya que ésta puede inducir a conclusiones erróneas debido a la naturaleza del ciclo de la variable estudiada.

El resultado obtenido se muestra en la tabla 1, en la que se evidencia una tendencia positiva de la media con un margen de error del 5%.

Tabla 1. Resultados de las pruebas de Mann-Kendall.

| | S | VAR[S] | Z | Z* |
|---------------------|--------|----------------|---------|--------|
| Serie | 15932 | 292325014.6667 | 0.9318 | 1.9600 |
| Media | 140148 | 210567912.6667 | 9.6580 | 1.9600 |
| Mediana | 65923 | 210529909.0000 | 4.5433 | 1.9600 |
| Percentil 75 | -75389 | 210538796.3333 | -5.1956 | 1.9600 |
| Percentil 25 | 313184 | 210497796.6666 | 21.5861 | 1.9600 |

Z* Corresponde al valor de Z para una probabilidad de $1-(\alpha/2)$

8. Tendencia en los percentiles

La tendencia en los percentiles fue analizada mediante la misma prueba, arrojando los resultados que se muestran en la tabla 1. Los cuales indican tendencia creciente en los percentiles 25 y 50, y una tendencia decreciente en el percentil 75.

Lo anterior puede explicarse, en parte, por la mayor concentración de los datos (menor varianza) al final de la serie. Es decir, los datos tienden a concentrarse cada vez más dentro de los percentiles 25 y 75.

9. Conclusiones

En primera instancia, observamos que es importante, al tratar este tipo de información, tener en cuenta los datos extremos de precipitación, dado que es probable que se presenten eventos de gran magnitud.

Por otro lado, como se evidencia en los resultados de la prueba de Mann-Kendall, la serie tiene una tendencia; sin embargo, ésta no debe ser alta debido al comportamiento de la distribución y los índices en las figuras mostradas a lo largo del documento. Por lo que se recomienda analizar los datos desde otro punto de vista para obtener una correcta descripción de la serie.

Finalmente, se evidencia que, en este caso es mejor trabajar con índices paramétricos, dado que los índices no paramétricos dejan a un lado los eventos extremos.