

Se espera un análisis más detallado en los resultados, recuerden que esto no es una clase de estadística ni programación. Además con el tiempo esperamos que vayan mejorando su “lenguaje científico”

4.1



UNIVERSIDAD
NACIONAL
DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS

Análisis de datos ambientales

Tarea #1

Profesor: Carlos David Hoyos Ortiz

Presentado por: Camilo Gutiérrez Ramírez ^a & Daniel Parra Holguín ^b

a. cagutierrez@unal.edu.co

b. daparra@unal.edu.co

Facultad de Minas, Universidad Nacional de Colombia, Medellín, Colombia.

Introducción

La serie de tiempo analizada es la Tasa de Cambio Representativa del Mercado (TRM) del dólar a pesos colombianos. La resolución temporal de la serie es diaria, medida durante los últimos 30 años y equivalente a 9765 datos desde el 21 de noviembre de 1991 hasta el 21 de agosto del 2018.

Se define La TRM como la cantidad de pesos colombianos por un dólar de los Estados Unidos. Se calcula con base en las operaciones de compra y venta de divisas entre intermediarios financieros que transan en el mercado cambiario colombiano. La Superintendencia Financiera de Colombia es la que calcula y certifica diariamente la TRM con base en las operaciones registradas el día hábil inmediatamente anterior.

1. Visualización

Como primer paso para el análisis de la serie de tiempo, se grafica ésta completamente para tener una primera impresión visual del comportamiento de la serie, como se muestra en la figura 1



Figura 1. Tasa de cambio representativa del mercado (TRM)

2. Histograma de frecuencias

La segunda herramienta empleada para el estudio de la serie es el histograma de frecuencias, el cual, a partir de barras, evidencia el número de valores repetidos en un cierto rango arbitrario

Para la elaboración del histograma se agrupan los datos en clases y se cuenta cuántas observaciones hay en cada una, Luego se divide el conteo por el número total de observaciones con el fin de presentar el gráfico con la frecuencia relativa de los datos. Se escoge el número de intervalos de tal forma que no haya ruido en la gráfica. Si se desea conocer que porcentaje de la TRM es menor a cierto valor, se encuentra el área bajo esa curva.

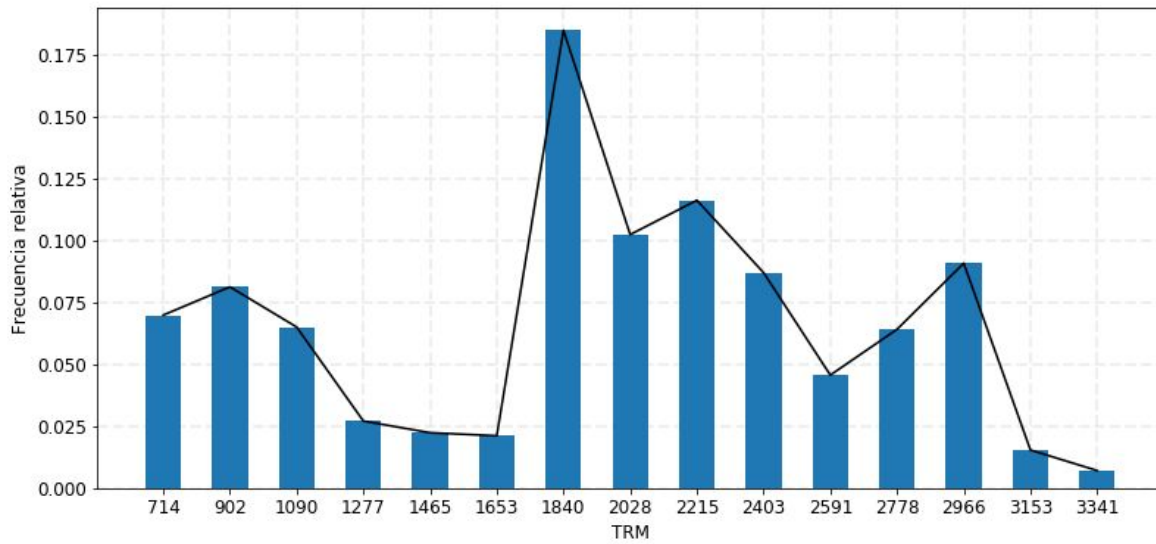


Figura 2. Histograma de frecuencia relativa.

3. Percentiles de la serie histórica

Otra medida que evalúa el comportamiento de los datos son los percentiles. Estos son números con las mismas unidades de la serie e indican la proporción de datos que son excedidos por el mismo para una probabilidad dada.

Lo que se pretende es graficar los percentiles de 0 a 100, para lo cual organizamos los datos de mayor a menor y encontramos los valores de los percentiles para cada una de las probabilidades en el rango.

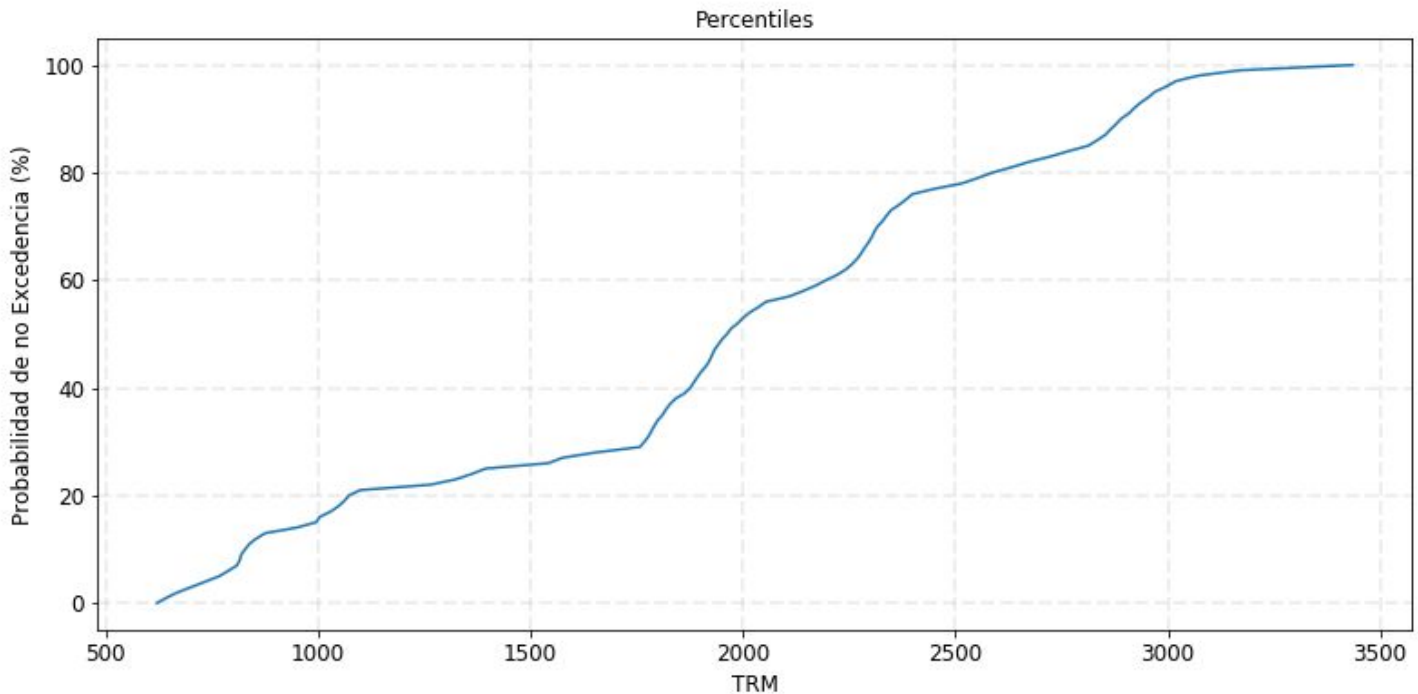


Figura 3. Percentiles de la serie histórica

4. Medidas Descriptivas

Para analizar la serie de tiempo es necesario buscar diferentes índices, teniendo un especial cuidado si se busca que éstos sean resistentes y robustos.

La resistencia está relacionada con la poca sensibilidad o influencia de los índices ante eventos extremos o atípicos “outliers”. La robustez, es otra característica que se refiere a la poca sensibilidad del método frente a suposiciones sobre la naturaleza de los datos. O que, en otras palabras, estos datos no dependan esencialmente de la distribución de estos.

Ojo con estas definiciones.
Les recomiendo que lean un poco más al respecto

Tabla 1. Análisis estadístico de la serie histórica.

Número de datos	9765
Media (\$)	1940.95463
Desviación Estandar (\$)	702.6435
Mínimo (\$)	620.62
Percentil 25% (\$)	1394.33
Percentil 50% (\$)	1962.9
Percentil 75%(\$)	2383.91
Máximo (\$)	3434.89

5. Estacionariedad de la serie

Se dice que una serie es estacionaria cuando los momentos estadísticos son constantes en el tiempo. Es posible verlo gráficamente cuando los valores de la serie tienden a oscilar alrededor de una media y una variabilidad constante.

Para ello se determinan ventanas móviles, las cuales son un rango de tiempo en el cual se evalúan parámetros estadísticos. En este caso se utiliza un periodo 3 años, que recorrerá toda la serie. Luego durante cada desplazamiento, se calcularán los índices. En un principio se escogieron 5 años, pero con el fin de mejorar la resolución de los resultados se redujo la ventana. Esta arbitrariedad se justifica en el desconocimiento de un ciclo uniforme en la TRM.

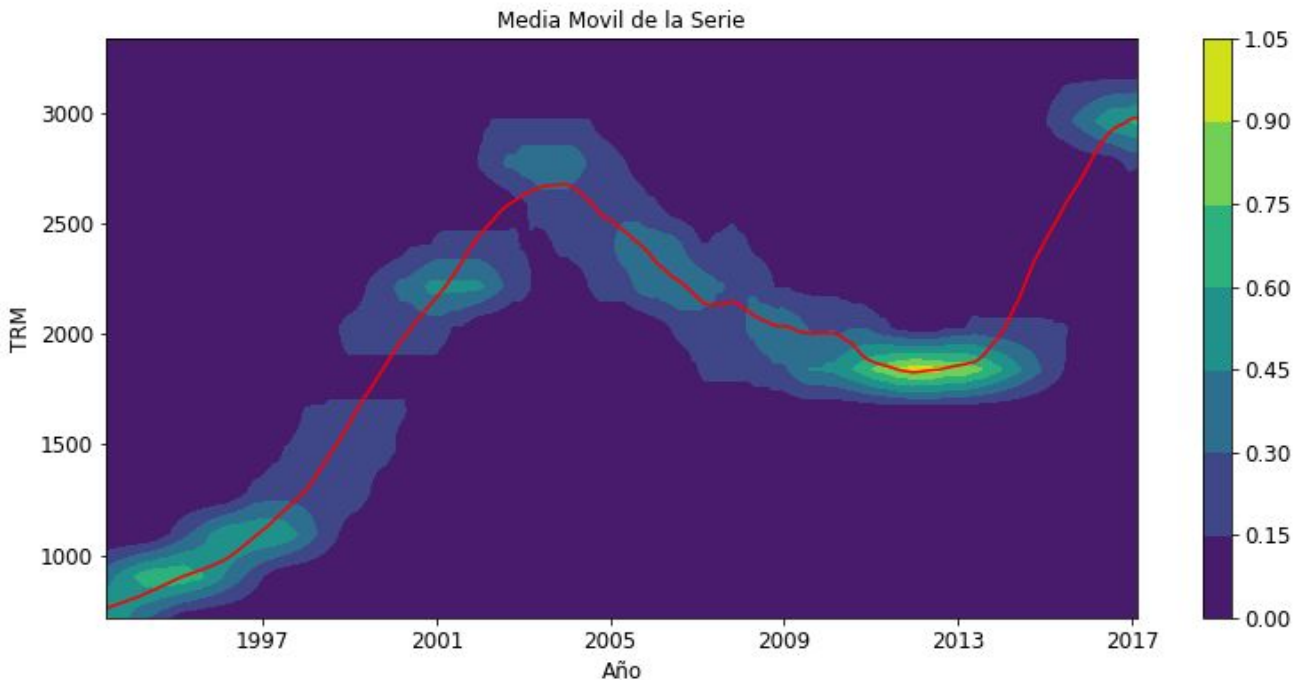


Figura 4 . Media movil de la serie.

En la figura 4 se observa que los datos y su media no son constantes en el tiempo, por lo tanto no es estacionaria la serie en su media.

- **Índices de localización:**

Las medidas de localización permiten conocer puntos característicos de la distribución. La medida de tendencia central más común es la mediana que es equivalente al percentil 50 de la serie y se caracteriza por ser resistente y robusta. La media en cambio no es resistente pues es muy sensible a valores extremos. Otra buena alternativa es la Trimean que es un promedio ponderado de los cuartiles con la mediana, teniendo esta última el doble de peso. Se calcula con la siguiente ecuación:

$$\text{Trimean} = \frac{q_{0.25} + 2 q_{0.5} + q_{0.75}}{4}.$$

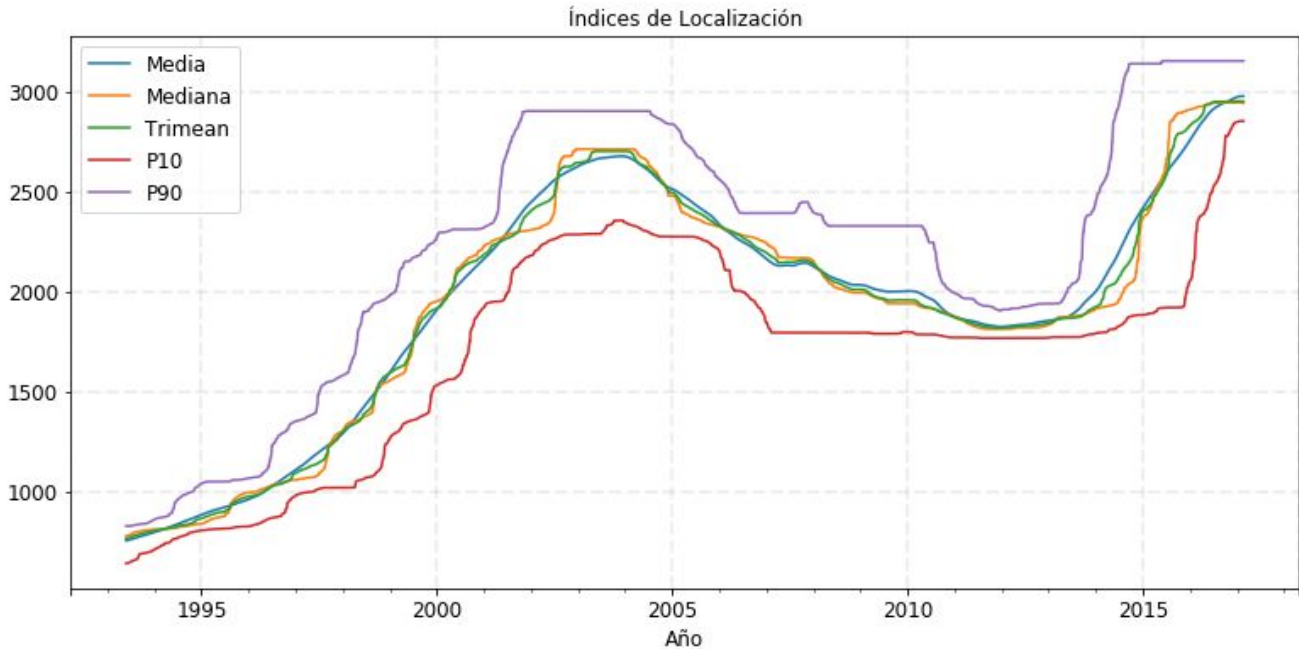


Figura 5. Índices de localización.

- **Índices de Dispersión**

Informan sobre cuánto se alejan del centro (media) los valores de la distribución, es decir, muestran la variabilidad de una distribución, indicando por medio de una constante, si los diferentes valores de una variable están muy alejadas de la media. Cuanto mayor sea ese valor, mayor será la variabilidad, cuanto menor sea, más homogénea será a la media.

El rango intercuartil es una medida robusta y resistente de la dispersión de una serie, es la diferencia entre el cuartil superior y el inferior. Es un buen índice de la dispersión de la parte central de la serie, además que el hecho de que ignora el 25% de los datos más altos y bajos permite que sea un índice resistente.

$$\text{IQR} = q_{0.75} - q_{0.25}.$$

La desviación estándar no es ni resistente ni robusta. Tiene las mismas unidades que la variable en cuestión y además es la medida de dispersión más común, que indica qué tan dispersos están los datos con respecto a la media. Mientras mayor sea la desviación estándar, mayor será la dispersión de los datos.

??????

Aunque el IQR es una medida bastante acertada, cuenta con la desventaja de que ignora una importante porción de los datos. En ese orden de ideas, MAD es un índice que tiene en cuenta todos los datos, mediante el cálculo de la mediana de la resta en valor absoluto de cada valor con la mediana de la serie. El MAD se calcula así:

$$\text{MAD} = \text{median}|x_i - q_{0.5}|.$$

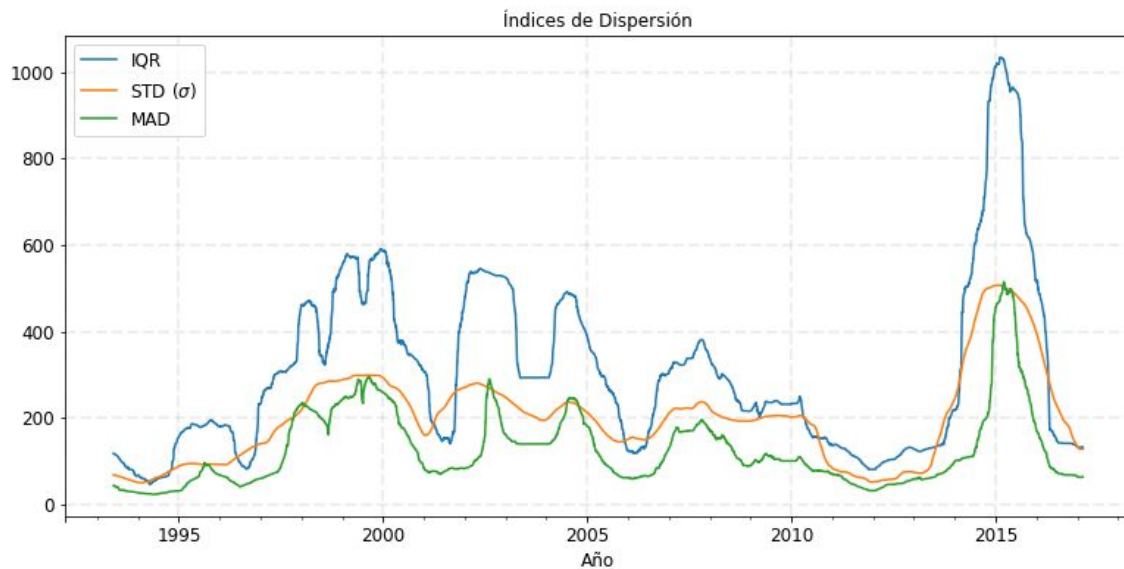


Figura 6. Índices de dispersión.

Figura. Media móvil de la serie.

Por lo tanto se concluye que la TRM no es una serie estacionaria para ningún momento.

- **Índices de Asimetría**

La asimetría es una medida de forma de una distribución que permite identificar y describir la manera como los datos tiende a agruparse de acuerdo con la frecuencia con que se hallen dentro de la distribución. Permite además identificar características de la distribución de datos sin necesidad de generar el gráfico pues si la asimetría es positiva la mayoría de los datos está a la derecha de la media muestral, si es negativa significa que la cola de la distribución se alarga para valores inferiores a la media.

Además del tercer momento, se grafica el índice de Yule-Kendall para evaluar la simetría. El cual, por medio del estudio de los cuartiles, determina la inclinación de los datos. Este índice mantiene la convención de la simetría

$$\gamma_{YK} = \frac{(q_{0.75} - q_{0.5}) - (q_{0.5} - q_{0.25})}{IQ} = \frac{q_{0.25} - 2q_{0.5} + q_{0.75}}{IQ}$$

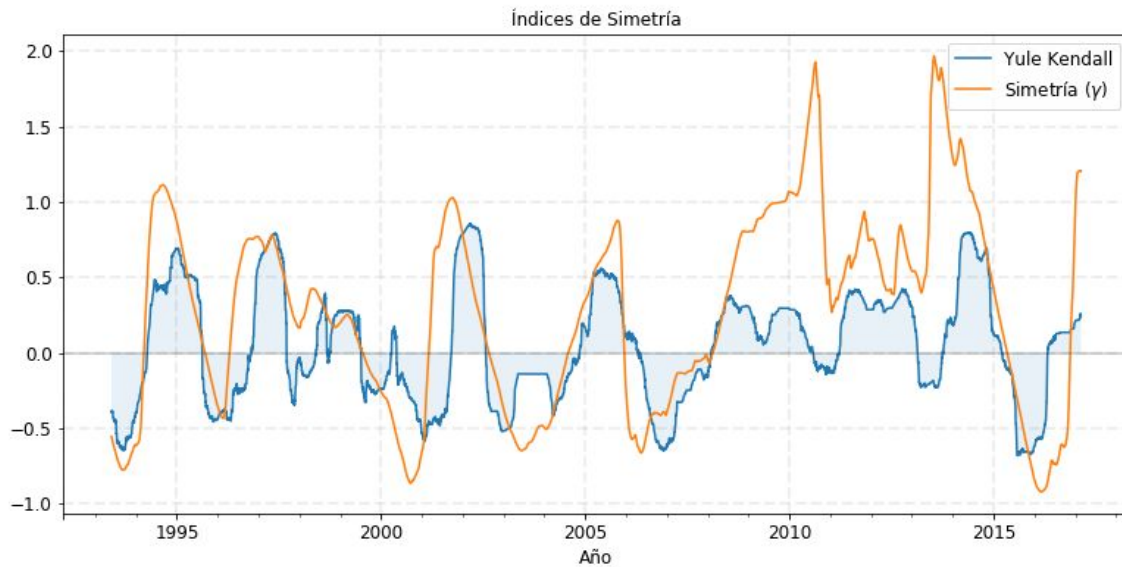


Figura 6. Índices de simetría.

6. Tendencia de la serie

Se puede definir como un cambio a largo plazo que se produce en relación al TRM medio a largo plazo. La tendencia se identifica con un movimiento suave de la serie a largo plazo.

Para evaluar la tendencia de la serie se hace con la ayuda del Test no paramétrico de Mann-kendall, el cual es útil debido a que puede tener valores faltantes y los datos no deben tener alguna distribución en particular. Primero se calcula el estadístico S

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i)$$

Para tamaños de muestra grandes, el estadístico S se distribuye aproximadamente normal con media cero y varianza:

$$V(S) = \frac{n(n-1)(2n+5) - \sum_t t(t-1)(2t+5)}{18}$$

Ho: La serie de datos no tiene tendencia (Se comporta aleatoriamente)

Ha: La serie tiene tendencia creciente o decreciente

Finalmente se calcula un estadístico Z_{MK} estandarizado. Con un nivel de significancia $\alpha = 5\%$ y una confiabilidad $1 - \frac{\alpha}{2} = 97.5\%$, teniendo en cuenta que se evaluará si la tendencia tiene tendencia creciente o decreciente (test de doble cola). Se rechaza Ho si:

$$|Z_{MK}| < Z_{1-\frac{\alpha}{2}}$$

Por tanto, se tiene que:

$$Z_{97,5\%} = 1,96$$

Tabla 2. Test de Mann-Kendall a la serie de tiempo.

Mann Kendall Test	Z_{MK}	$Z_{97,5\%}$	Conclusión
Serie (TRM)	73.095	1.96	Se rechaza Ho a favor de Ha.
Media móvil de la serie	54.347	1.96	Se rechaza Ho a favor de Ha
P10	54.902	1.96	Se rechaza Ho a favor de Ha
P90	58.389	1.96	Se rechaza Ho a favor de Ha

Según la tabla 1, hay suficiente evidencia muestral para rechazar Ho, por tanto se puede afirmar con un nivel de significancia del 5% que la serie y percentiles tienen tendencia.

7. Conclusiones

- En el año 2015 los índices de dispersión fueron los más altos de la serie histórica esto se debe a las altas variación de \$2360 en mediados de mayo del 2015 a \$3340 a principios de Diciembre con un aumento del 40% en sólo 7 meses.
- La serie histórica del TRM no es estacionaria con respecto a ningún momento estadístico, representando acertadamente el comportamiento del dólar pues este no es constante con respecto al tiempo. En cambio ha tenido una tendencia creciente a largo plazo, pues el precio del dólar es un reflejo de la situación actual de la economía

????

- Las herramientas descriptivas para ver el comportamiento de una serie de datos pueden ser usados sin recurrir a gráficos y a menudo es lo que suele hacerse en primera instancia. Los índices suelen estar conformados por tres: localización, dispersión y simetría. Ellos corresponden a los tres primeros momentos estadísticos de una muestra de datos, pero las medidas usuales (media de la muestra, varianza de la muestra y coeficiente de asimetría) no son robustas ni resistentes, por tanto se debe analizar en qué circunstancias se quiere que los métodos de análisis sean robustos y resistentes .
- La asimetría es una medida importante de la distribución de los datos. Con ella es posible determinar hacia qué lado de la media es tendiente los valores de una serie. Un ejemplo claro de este caso es en series de tiempo de lluvias, donde la ubicación de la serie hacia la izquierda indica que la meteorología predominante NO se encuentra del lado de tormentas.

Para nuestro caso, se evidencia una oscilación de los valores de Yule-Kendall de positivos a negativos, lo cual no es nada concluyente en la naturaleza de la serie.

- Estudiar la estacionariedad de un fenómeno es de especial interés en las series de tiempo, ya que permite evaluar si las características conductuales permanecen o varían con el transcurso del tiempo.
- Otra forma de apreciar el significado de los percentiles es anotando que sus probabilidades asociadas están referidas a las de no excedencia. Por ejemplo, para el percentil del 10%, se concluye que el valor asociado supera al 10% de los datos o, en otras palabras, que la probabilidad de no encontrar un valor superior es solo del 10% (Ver figura 3.)

Bibliografía

- Statistical methods for environmental pollution monitoring, richard o. Gilbert
- Evaluación y análisis espaciotemporal de tendencias de largo plazo en la hidroclimatología colombiana, Diana Cristina Cantor Gómez
- [2] Wilks, D. (2006). Statistical Methods in the Atmospheric Sciences. Department of Earth and Atmospheric Sciences Cornell University: New York.