



COLEGIO DE POSTGRADUADOS

**INSTITUCION DE ENSEÑANZA E INVESTIGACIÓN
EN CIENCIAS AGRICOLAS**

CAMPUS MONTECILLO

**POSTGRADO DE SOCIOECONOMIA, ESTADISTICA E INFORMATICA
ESTADISTICA**

**ANALISIS DE TENDENCIA EN EL MODELO AUTORREGRESIVO
BAJO INNOVACIONES CON DISTRIBUCIONES NORMAL Y
OTRAS**

PABLO MAZA CANTELLANO

TESIS

**PRESENTADA COMO REQUISITO PARCIAL PARA OBTENER EL
GRADO DE:**

MAESTRO EN CIENCIAS

**MONTECILLO, TEXCOCO, EDO. DE MEXICO
2008**

La presente tesis titulada: **ANALISIS DE TENDENCIA EN EL MODELO AUTORREGRESIVO BAJO INNOVACIONES CON DISTRIBUCIONES NORMAL Y OTRAS**, realizada por el alumno: **Pablo Maza Cantellano**, bajo la dirección del Consejo Particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

MAESTRO EN CIENCIAS

**SOCIOECONOMIA, ESTADISTICA E INFORMATICA
ESTADISTICA**

CONSEJO PARTICULAR

CONSEJERO

Dr. Humberto Vaquera Huerta

ASESOR

Dr. José Aurelio Villaseñor Alva

ASESOR

Dr. Carlos Miguel Becerril Pérez

MONTECILLO, TEXCOCO, MEXICO, 28 DE MARZO DE 2008

RESUMEN

ANÁLISIS DE TENDENCIA EN EL MODELO AUTORREGRESIVO BAJO INNOVACIONES CON DISTRIBUCIONES NORMAL Y OTRAS

PABLO MAZA CANTELLANO, M. C.

COLEGIO DE POSTGRADUADOS, 2008

Se analiza la tendencia en el modelo autorregresivo. Se propone el nivel de tendencia y se observa si las pruebas de Mann-Kendall y el coeficiente de correlación de Pearson la detectan, variando el tamaño de muestra, el valor del término autorregresivo, el nivel de significancia y el tipo de distribución que siguen las innovaciones, la cual puede ser Exponencial, Pareto, de Valor Extremo Generalizado Tipo I o Normal. El proceso es aplicado a datos simulados sin transformar y a datos transformados con sus rangos. Los resultados sugieren que cuando se utiliza la transformación del rango se obtiene la mejor potencia, aunque no existen notables diferencias entre ambas pruebas puede decirse que la mejor es la prueba de Mann-Kendall, dado que es menos afectada por observaciones extremas y depende directamente del tipo de distribución bajo estudio. El coeficiente de correlación de Pearson presenta ventajas: es más sencillo de calcular y cuando se aplica después de transformar los datos con sus rangos, se obtiene la prueba rho de Spearman, que genera resultados muy similares a la prueba de Mann-Kendall y ambas son muy eficientes en la detección de tendencia.

Palabras clave: Tendencia, Autorregresivo, Mann-Kendall, Coeficiente de correlación de Pearson, Transformación del rango.

ABSTRACT

TREND ANALYSIS IN THE AUTORREGRESSIVE MODEL UNDER INNOVATIONS WITH NORMAL DISTRIBUTIONS AND OTHERS

PABLO MAZA CANTELLANO, M. C.

COLEGIO DE POSTGRADUADOS, 2008

In this paper, the trend is analyzed in the autoregressive model. The trend level is proposed and is observed the Mann-Kendall test and Pearson correlation coefficient results, changing size sample, the autoregressive term, the significance level and the innovations distribution, that can be Exponential, Pareto, Generalized Extreme Value type I or normal distribution. Process is applied to simulated data without transformation and data that have been transformed by its ranges. The best power is obtained when we use the range transformation, although no differences exist between tests it is possible to assure that Mann-Kendall is the best test, since outliers do not affect it and directly depend of the distribution under investigation. Likewise, Pearson correlation coefficient, present enormous advantages: it is easier to compute than Mann-Kendall test, furthermore, if is applied to the ranges, the Sperman's rho is obtained, and the outcomes are similar to the Mann-Kendall test and both are efficient to detect trend.

Key words: Trend, Autorregressive, Mann-Kendall, Pearson correlation coefficient, Range transformation.

AGRADECIMIENTOS

A los millones de mexicanos (as) que pagan impuestos, quienes, a través del Consejo Nacional de Ciencia y Tecnología (CONACYT) y el Colegio de Postgraduados, han financiado parte de mi formación.

A las personas integrantes de mi Consejo Particular por el esfuerzo, la dedicación, el tiempo y el apoyo que me han brindado, pero sobre todo por la infinita paciencia para conmigo y mis problemas

Todos los individuos, llámense profesor (a), compañero (a), amigo (a) o familia, quienes de alguna manera me han acompañado hasta hoy, aun en la distancia.

CONTENIDO

1.- INTRODUCCION	1
2.- OBJETIVOS	4
3.- REVISION BIBLIOGRAFICA	5
3.1.- TENDENCIA DETERMINISTA	8
3.2.- TENDENCIA ESTOCASTICA	11
3.3.- PRUEBAS PARA LA DETECCION DE TENDENCIA	13
3.3.1.- PRUEBA DE MANN-KENDALL PARA LA DETECCION DE TENDENCIA	13
3.3.2.- PRUEBA DEL COEFICIENTE DE CORRELACION DE PEARSON PARA LA DETECCION DE TENDENCIA	16
3.3.3.- COMPARACION DE LAS PRUEBAS DE MANN-KENDALL Y DEL COEFICIENTE DE CORRELACION DE PEARSON.....	16
3.4.- TRANSFORMACION DEL RANGO	16
3.4.1.- VENTAJAS DE LA TRANSFORMACION DEL RANGO	20
3.5.- LOS ERRORES ESTADISTICOS Y LA PRUEBA DE POTENCIA	21
3.5.1.- NIVEL DE SIGNIFICANCIA Y POTENCIA	23
3.5.2.- TAMAÑO MUESTRAL Y POTENCIA	23
3.6.- MODELO Y ESTIMADORES	29
3.7.- DISTRIBUCIONES A UTILIZAR	30
3.7.1.- DISTRIBUCION DE VALOR EXTREMO GENERALIZADO TIPO I	30
3.7.2.- DISTRIBUCION EXPONENCIAL	31
3.7.3.- DSITRIBUCION PARETO	31
3.7.4.- DISTRIBUCION NORMAL	31
3.8.- POTENCIA DE UNA PRUEBA Y DETECCION DE TENDENCIAS	32
3.8.1.- POTENCIA DE LA PRUEBA.	32
3.9.- SIMULACION	33
4.- METODOLOGIA	34
5.- ANALISIS DE RESULTADOS	35
6. APLICACIÓN DE LAS PRUEBAS MOSTRADAS EN EL ESTUDIO	44
7.- CONCLUSIONES	45
8.- REFERENCIAS	46
9.- ANEXOS	50

LISTA DE CUADROS Y FIGURAS

Tabla 1. Potencia de las pruebas para datos sin transformar con $\alpha = 0.01$ y $\alpha = 0.05$	36
Tabla 2.- Potencia de las pruebas para datos sin transformar con $\alpha = 0.01$ y $\alpha = 0.1$	37
Tabla 3. Potencia de las pruebas para datos sin transformar con $\alpha = 0.01$ y $\alpha = 0.01$	38
Tabla 4.- Potencia de las pruebas para datos transformados por sus rangos con $\alpha = 0.01$ y $\alpha = 0.05$	39
Tabla 5. Potencia de las pruebas para datos transformados por sus rangos con $\alpha = 0.01$ y $\alpha = 0.1$	40
Tabla 6. Potencia de las pruebas para datos transformados por sus rangos con $\alpha = 0.01$ y $\alpha = 0.01$	41
Tabla 7. P-valores para las pruebas aplicadas	44
Tabla 8. Estadísticas calculadas para las pruebas aplicadas	45
Figura 1. Gráfica de la potencia para la prueba de Mann-Kendall para datos sin transformar con $\alpha = 0.5$, $n = 12$ y $\alpha = 0.05$	42
Figura 2. Gráfica de la potencia para la prueba del coeficiente de correlación de Pearson para datos sin transformar con $\alpha = 0.5$, $n = 12$ y $\alpha = 0.05$	42
Figura 3.- Gráfica de potencia de la prueba de Mann-Kendall a datos transformados por sus rangos con $\alpha = 0.5$, $n = 12$ y $\alpha = 0.01$	43
Figura 4.- Gráfica de potencia de la prueba del coeficiente de correlación de Pearson a datos transformados por sus rangos con $\alpha = 0.5$, $n = 12$ y $\alpha = 0.01$	43
Figura 5.- Serie de tiempo para los niveles máximos de monóxido de carbono para la estación Pedregal de Julio de 1987 a Abril de 1999	44

1.- INTRODUCCION

Un importante objetivo en muchas áreas del conocimiento es detectar cambios o tendencias en series de datos tomados a través del tiempo (Gilbert, 1987). Un ejemplo puede ser en algún programa de monitoreo ambiental que intenta detectar cambios en los niveles de contaminación como resultado de cambiar el uso de la tierra a practicas tales como el crecimiento de ciudades, incremento en la erosión de campos dedicados a cultivo, el inicio de un campo para tiradero de basura, entre otras.

La importancia de los resultados que se obtienen en ciencias relacionadas al ambiente, se asocian con numerosos estudios que actualmente analizan la presencia de tendencia en variables que buscan detectar señales del proceso denominado calentamiento global (Hamed, 2007).

Los objetivos del análisis de series tiempo son la predicción, el control de un proceso, la simulación de procesos, y la generación de nuevas teorías físicas o biológicas, entre otros.

Si una vez conocidos los valores pasados de la serie, no fuera posible predecir con total certeza el próximo valor de la variable, decimos que la serie es no determinista o aleatoria, y es de éstas de las que se ocupa el "análisis de series de tiempo" (Molinero, 2004).

Se denomina predicción a la estimación de valores futuros de la variable en función del comportamiento pasado de la serie. Este objetivo se emplea ampliamente en el campo de la ingeniería y de la economía, en la sanidad pública y la vigilancia de la salud. Así por ejemplo, la predicción mediante modelos basados en la teoría de series de tiempo, puede servir para una buena planificación de recursos sanitarios, en función de la demanda que se espera en el futuro, prevista por el modelo. Otro de los campos en los que se aplica la predicción mediante series de tiempo es el de la meteorología o en la predicción de otros fenómenos naturales.

En la teoría de control de procesos, se trata de seguir la evolución de una variable determinada con el fin de regular su resultado. Esta teoría se utiliza en medicina en los Centros

de Control de Enfermedades. Mientras que, la simulación se emplea en investigación aplicada, cuando el proceso es muy complejo para ser estudiado de forma analítica.

Aunque el valor futuro de una serie temporal no sea predecible con total exactitud, el resultado tampoco puede ser completamente aleatorio, existiendo alguna regularidad en cuanto a su comportamiento en el tiempo, lo que hará posible su modelado y eventualmente la predicción. La búsqueda de regularidades y de patrones ha sido siempre una de las tareas básicas de la ciencia, y muchas veces se descubren simetrías que sirven de fundamento para la predicción del comportamiento de los fenómenos, incluso antes de que se entienda la razón o causa que justifica esa regularidad.

Por lo tanto, si podemos encontrar patrones de regularidad en diferentes secciones de una serie de tiempo, podremos también describirlas mediante modelos basados en distribuciones de probabilidad. La secuencia ordenada de variables aleatorias $X(t)$, se denomina proceso estocástico. Un proceso estocástico es por tanto el modelo matemático para una serie de tiempo. Un concepto importante que encontramos en este ámbito, es el de procesos estacionarios. Si examinamos por ejemplo la temperatura para un determinado mes a lo largo de los años en una determinada zona geográfica, y se está produciendo un cambio climático, aunque haya fluctuaciones, habrá una tendencia creciente. De una manera informal, diremos que una serie es estacionaria cuando se encuentra en equilibrio estadístico, en el sentido de que sus propiedades no varían a lo largo del tiempo, y por lo tanto no pueden existir tendencias. Un proceso es no estacionario si sus propiedades varían con el tiempo, como por ejemplo el clima.

Cuando se tiene una serie de tiempo no estacionaria el investigador se cuestiona en relación a la tendencia que los datos siguen a través del tiempo, en este sentido las preguntas a resolver pueden ser sobre que tan grande es esa tendencia, o bien cual es la probabilidad de detectarla. Para dar respuesta a estas preguntas existen muchas herramientas estadísticas, desarrolladas a lo largo de los años.

Conocer el comportamiento de una serie de tiempo, resulta importante no sólo para el investigador sino en mayor medida para las personas encargadas de tomar decisiones, pues una

decisión equivocada implica pérdidas de algún tipo; así por ejemplo, en términos de economía cualquier decisión errónea traerá consecuencias que repercuten directa o indirectamente en otros factores del desarrollo generando pérdidas económicas. En términos ecológicos, una decisión equivocada en el análisis de la serie de tiempo, conducirá a graves problemas, como pueden ser en el abastecimiento de alimentos o en factores de crecimiento poblacional.

En este trabajo se estudia el comportamiento de la prueba de Mann-Kendall y el coeficiente de correlación de Pearson a datos simulados, una vez que se ha asignado un valor de tendencia con la finalidad de comparar ambas pruebas, variando el nivel de significancia, el tamaño muestral, el valor del término autorregresivo, el nivel de tendencia y el tipo de distribución que siguen las innovaciones en el modelo autorregresivo.

2.- OBJETIVOS.

2.1.- OBJETIVO GENERAL. Comparar las pruebas de Mann-Kendall y el coeficiente de correlación de Pearson, tanto a datos sin transformar como a datos transformados con la transformación del rango, utilizando la potencia de las pruebas en el modelo autorregresivo; cuando las innovaciones siguen alguna distribución no normal, como normal.

2.2. OBJETIVOS PARTICULARES.

1.- Obtener la potencia de la prueba de Mann-Kendall, utilizando el modelo autorregresivo, a datos transformados con la transformación del rango y a datos sin transformar, cuando los valores de tendencia son asignados y las innovaciones siguen una distribución de valor extremo generalizada tipo I, Exponencial, Pareto y Normal.

2.- Obtener la potencia de la prueba utilizando el coeficiente de correlación de Pearson, utilizando el modelo autorregresivo para datos transformados con la transformación del rango y a datos sin transformar, cuando los valores de tendencia son asignados y las innovaciones siguen una distribución de valor extremo generalizado tipo I, Exponencial, Pareto y Normal.

3.- Comparar las potencias de las pruebas obtenidas para niveles de significancia $\alpha = 0.01, 0.05$ y 0.1 , así como para tamaños de muestra $n = 12, 60$ y 120 ; variando también el término autorregresivo en las innovaciones desde $0.001, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$ y 0.99 .

3.- REVISION BIBLIOGRAFICA

El término “tendencia” fue presentado por primera vez en un trabajo de Hooker (1901) para analizar las series de importaciones y exportaciones de Gran Bretaña. Los primeros intentos para identificar la tendencia de las series macroeconómicas, típicamente mediante los métodos basados en promedios móviles, propuestos inicialmente por Poynting (1984), y en interpolación gráfica, los cuales son discutidos por Klein (1997).

Durante varios años, la tendencia fue aislada totalmente de manera tal de poder eliminarla de la serie. La forma por la cual se solía remover la tendencia de la serie era mediante el ajuste de alguna función determinista del tiempo. Un estudio posterior de la tendencia, proporcionado por Frickey (1934), describe un conjunto de métodos estadísticos para estimar la tendencia. Hacia fines de 1930 se desarrollaron modelos más rigurosos para explicar la naturaleza de las fluctuaciones cíclicas.

Hasta hace muy poco tiempo, las investigaciones sobre el ciclo económico tendieron a hacer referencia a la presencia de tendencias de largo plazo casi “in passing”. Las series de tiempo necesitan de un método simple para remover la tendencia y rápidamente centrar el análisis sobre las fluctuaciones cíclicas. Este enfoque se justifica siempre y cuando la interacción entre la tendencia de largo plazo y las fluctuaciones cíclicas de corto plazo sea pequeña. Aún así, el uso de un procedimiento incorrecto para remover la tendencia de una serie puede complicar la separación de la tendencia del ciclo.

Con el desarrollo de la teoría del crecimiento, se comenzó a prestar mayor atención sobre la modelación de la tendencia de las series. El trabajo de Klein y Kosobud (1961) representó un intento innovador de ajustar las tendencias no sólo de un conjunto de series individuales sino también sus cocientes, estos son los “grandes cocientes” de la teoría del crecimiento. Este trabajo fue la base de otros trabajos posteriores que desarrollaron la idea de tendencias comunes en un conjunto de series de tiempo y el concepto de cointegración. Este concepto juega un rol muy importante en la econometría moderna de las series de tiempo (Banerjee, 1993).

Klein y Kosobud restringieron su análisis de la tendencia sobre filtros lineales y log lineales, lo cual este supuesto pareció ser razonable para toda la década del sesenta y setenta. Aún cuando este enfoque haya sido una aproximación útil y simple, existe poca evidencia a favor del supuesto de tendencia determinista constante en todo el dominio para series históricas en las cuales existen cambios estructurales, tasas variables en la acumulación de factores productivos y progreso técnico. Este fenómeno hizo posible el desarrollo de nuevas técnicas que incorporaron la noción de cambios estructurales en la tendencia y el concepto de tendencia estocástica.

La variabilidad de la tendencia complica su separación del ciclo económico y el hecho de presuponer una tendencia lineal en estas circunstancias puede llevar a identificar ciclos económicos espurios (Nelson y Kang, 1981). Estos dos puntos centrales, la independencia y la variabilidad de la tendencias han sido sujeto de gran discusión desde 1980, iniciada esta a partir de dos trabajos, Beveridge y Nelson (1981) y Nelson y Plosser (1982). El primero de estos trabajos investiga la forma por la cual es posible separar la tendencia y el ciclo económico cuando la serie observada está generada por un proceso estacionario en diferencias. Es decir, un proceso estocástico cuya tendencia no retorna hacia un sendero determinista y cuyo movimiento evoluciona de acuerdo a un paseo aleatorio. El segundo de estos trabajos utiliza técnicas desarrolladas por Dickey y Fuller (1979) para probar la presencia de raíces unitarias en las series.

Vahid y Engle (1993) muestran cómo la representación de tendencias comunes a un conjunto de series de tiempo puede ser extendida para incorporar ciclos económicos comunes. En ese sentido, los autores proveen un contexto en el cual ambos tipos de restricciones pueden ser probadas empíricamente e impuestas de manera secuencial, admitiendo entonces un análisis integrado entre tendencias y ciclos económicos.

Los argumentos acerca de cómo se remueve la tendencia de una serie no son equivalentes a los argumentos acerca de cuál es la frecuencia de los ciclos dentro de dicha serie.

Muchos investigadores interesados en el estudio de la teoría del crecimiento y la teoría del ciclo económico se enfrentan al problema de poder identificar las fluctuaciones cíclicas y la tendencia.

Las dificultades prácticas de este ejercicio han sido apreciadas durante muchos años, pero los métodos estadísticos utilizados para realizar el filtrado de series de tiempo en ciclo y tendencia exigen utilizar supuestos ad-hoc. El problema de estos supuestos reside en su simplicidad, es decir, han sido formulados para facilitar la identificación sin hacer referencia alguna a las propiedades estadísticas de las series (o conjunto de ellas) en cuestión.

En el caso de las series utilizadas en economía pueden ser representadas de manera aditiva como la suma de tres términos: el ciclo ε_t , la tendencia μ_t y los factores estacionales γ_t .

Para facilitar la exposición se supondrá que los tres términos son independientes entre sí, es decir:

$$y_t = \mu_t + \varepsilon_s + \gamma_k \quad E(\mu_t, \varepsilon_s, \gamma_k) = 0 \quad \forall t, s, k$$

Mientras que la tendencia de la serie cambia muy suavemente en el tiempo, las fluctuaciones cíclicas, por el contrario, cambian muy rápidamente en el tiempo. Los factores estacionales de la serie representan movimientos periódicos con período fijo.

La tendencia representa perturbaciones permanentes sobre la serie, las fluctuaciones cíclicas son un conjunto de perturbaciones transitorias. De esto se sigue que cualquier movimiento permanente o de largo plazo en la serie observada es atribuido necesariamente a la componente de tendencia. Desde un punto de vista teórico y práctico, es necesario realizar esta distinción de los componentes de la serie de forma tal de poder comprender la naturaleza de sus cambios.

La tendencia y el ciclo son inobservables y, por tanto, deben ser estimados mediante alguna técnica estadística. La noción de que la tendencia no presenta grandes fluctuaciones en el tiempo, aún cuando su comportamiento es suave en relación a la componente cíclica, ha llevado en la práctica a remover la tendencia mediante modelos representados por una función determinista del tiempo.

3.1.- TENDENCIA DETERMINISTA.

El modelo más simple para representar la tendencia de la serie es una función lineal del tiempo, es decir:

$$\mu_t = \alpha + \beta t$$

El modelo de regresión resultante es:

$$y_t = \alpha + \beta t + \varepsilon_t$$

Aún cuando el modelo de tendencia lineal sea una técnica muy útil desde un punto de vista pedagógico, el supuesto simplificador de crecimiento constante en la tendencia hace que su uso en la práctica sea escaso. Una consecuencia obvia de este hecho es admitir la existencia de una tendencia no lineal, es decir:

$$\mu_t = f(t)$$

La dificultad de esta técnica reside en la elección de la forma funcional correcta y, para ello la noción común nos dice que la tendencia deberá ser una función suave del tiempo. El significado de suave en este contexto no es simple de definir, pero, sin embargo, algunas propiedades matemáticas pueden ser identificadas a tal efecto. La función apropiada para representar a la tendencia debería cumplir con dos requisitos: ser continua en el tiempo y que su primera derivada esté definida en todo el dominio. Estas propiedades, sin embargo, no son suficientes. El significado de suavidad exige que existan pocos cambios en el signo de la primera derivada. Entonces una función polinómica del tiempo es una especificación apropiada en este contexto, a saber:

$$\mu_t = \alpha + \sum_{j=1}^k \beta_j t^j$$

Desde un punto de vista matemático, el teorema de Weierstrass establece que cualquier función continua definida en un intervalo compacto puede ser bien aproximada por un polinomio. Sin embargo, el requisito de que existan pocos cambios en el signo de la primera derivada implica que el orden del polinomio k sea bajo. En términos prácticos, si se fija k muy alto, entonces se correrá el riesgo de sobre-ajustamiento: la componente cíclica será parte de la tendencia y, por tanto, esta última componente dejará de ser una función suave del tiempo.

Un enfoque alternativo que resuelve el problema de la elección de la función consiste en aproximar segmentos de la función que representa la tendencia mediante un polinomio de bajo orden. La ventaja de este enfoque reside en que admite la existencia de cambios estructurales en la tendencia y permite distinguir entre regímenes de política, cuyos años terminales se definen a priori.

Se sigue que se puede obtener una estimación más eficiente de la tendencia entre un ciclo y otro consecutivo a partir del siguiente modelo:

$$y_t = \alpha_1 + \beta_1 t + \sum_{j=2}^k \alpha_j d_{jt} + \sum_{j=2}^k \delta_j d_{jt} t + \varepsilon_t$$

$$d_{jt} = \begin{cases} 1 & \text{para el régimen } j \\ 0 & \text{para el caso contrario} \end{cases}$$

Si la serie está medida en logaritmos, entonces la tasa de crecimiento para el régimen j estará dada por:

$$\beta_j = \beta_1 + \delta_2 + \dots + \delta_j$$

Este enfoque alternativo nos permite realizar la prueba de la hipótesis nula de que la tasa de crecimiento es constante a lo largo del período bajo estudio. Esto significaría un rechazo a la hipótesis de cambio estructural en la tendencia entre un ciclo y otro. Es decir:

$$H_0 : \delta_2 = \delta_3 = \dots = \delta_k = 0$$

Una ventaja adicional de este enfoque con respecto a los anteriores reside en el hecho de que la existencia de cambios estructurales impide que la representación de la tendencia sea una función continua del tiempo, aún cuando la hipótesis nula sea rechazada.

Las fluctuaciones cíclicas que han de ser explicadas por la teoría del ciclo se obtienen a partir de los residuos de la regresión, es decir:

$$\varepsilon_t = y_t - \hat{u}_t$$

Una vez obtenida la componente cíclica como la diferencia entre la serie observada y la tendencia, se procede a identificar el proceso generador de los datos.

En ausencia de cualquier tendencia, las series observadas estarían completamente caracterizadas por el ciclo económico, es decir:

$$y_t = \alpha + \varepsilon_t$$

y, por consiguiente, estaríamos en presencia de un proceso estacionario que podría ser representado mediante un modelo ARMA. Tradicionalmente la tendencia se solía tratar como determinista, midiéndose la misma mediante el ajuste de curvas suaves. Esta práctica será correcta siempre y cuando la naturaleza de la tendencia del proceso estocástico sea determinista.

Sin embargo, existen potenciales problemas en la estimación de la tendencia de acuerdo a las especificaciones utilizadas. La dificultad radica en que no es seguro que la tendencia de la serie observada responda a los dos primeros sumandos de la derecha. Si esto resultara ser cierto, entonces se sigue que la supresión de la tendencia de la serie de tiempo puede falsear el componente estimado de las fluctuaciones cíclicas, es decir, el residuo. Siguiendo este razonamiento, se estaría cometiendo un error al suponer que todas las perturbaciones forman parte del término de error y que, por consiguiente, no afectan el valor de t y en el largo plazo.

Nelson y Plosser hicieron notar que algunas series de tiempo son mejor representada como un proceso estocástico que no retorna a una senda determinista, lo que significa que las innovaciones del producto afectan el valor futuro del mismo. Entonces este hecho da lugar a dos hipótesis alternativas: ¿es el producto estacionario en tendencia? O bien, ¿es el producto estacionario en diferencias? Es decir, las perturbaciones ¿afectan o no afectan a la tendencia estocástica? Si el proceso estocástico no retorna hacia una senda determinista, entonces una parte importante de las fluctuaciones cíclicas de la serie macroeconómica observada puede deberse a la variación estocástica de la tendencia, que no queda eliminada.

Si la tendencia de la series de tiempo es estocástico, se sigue que los modelos basados en los residuos de la regresión de y con respecto al tiempo son incorrectos.

3.2.- TENDENCIA ESTOCASTICA.

Beveridge y Nelson (1981) proporcionan un procedimiento elegante para descomponer series de tiempo no estacionarias en torno a una tendencia estocástica en una componente de tendencia y en una componente cíclica, aplicando métodos ARIMA.

Otra técnica bastante utilizada para extraer la componente cíclica es el filtro de Hodrick y Prescott (1997). Este filtro surge como la solución del problema de minimización de la variabilidad de la componente cíclica de la serie observada sujeta a una condición de suavidad de la tendencia.

El problema de extracción de la tendencia es el siguiente:

$$\text{Min}_{\mu_t} \sum_{t=1}^T \varepsilon_t^2 + \lambda \sum_{t=1}^T [(\mu_{t+1} - \mu_t) - (\mu_t - \mu_{t-1})]^2$$

donde λ es un parámetro que representa la suavidad. Un valor muy alto del parámetro hace referencia a una tendencia muy suave y, por consiguiente, en el límite la componente de tendencia será una función lineal. Las condiciones de primer orden son:

$$0 = -(y_t - \mu_t) + \lambda[(\mu_t - \mu_{t-1}) - (\mu_{t-1} - \mu_{t-2})] - 2\lambda[(\mu_{t+1} - \mu_t) - (\mu_t - \mu_{t-1})] + \lambda[(\mu_{t+2} - \mu_{t+1}) - (\mu_{t+1} - \mu_t)]$$

$$y_t = \mu_t + \lambda(1-L)^2(\mu_t - 2\mu_{t+1} + \mu_{t+2}) = \left(+ \lambda(1-L)^2(1-L^{-1})^2 \right) \mu_t$$

Se tiene entonces el filtro de Hodrick y Prescott para extraer la tendencia:

$$\text{Filtro H-P} = \left(+ \lambda(1-L)^2(1-L^{-1})^2 \right)$$

y la componente cíclica es:

$$\varepsilon_t^{HP} = y_t - \mu_t^{HP}$$

La principal conclusión es que no existe un único método de descomposición de series de tiempo. En la práctica, la elección de una u otra técnica de estimación dependerá en gran medida del tipo de problema y la información que se tenga acerca del mismo. Las propiedades

estadísticas del proceso generador de las series juegan un rol crucial para poder optar por una u otra estrategia de estimación. A partir de una descripción adecuada del problema, se puede optar por una u otra técnica, siempre con algún grado de ambigüedad.

La metodología tradicional para el estudio de series temporales es bastante sencilla de comprender, y fundamentalmente se basa en descomponer las series en varias partes: tendencia, variación estacional o periódica, y otras fluctuaciones irregulares, las cuales pueden definirse de la siguiente manera:

Tendencia. Es la dirección general de la variable en el periodo de observación, es decir el cambio a largo plazo de la media de la serie.

Estacionalidad. Corresponde a fluctuaciones periódicas de la variable, en periodos relativamente cortos de tiempo.

Otras fluctuaciones irregulares. Después de extraer de la serie la tendencia y variaciones cíclicas, nos quedará una serie de valores residuales, que pueden ser o no totalmente aleatorios. Volvemos a estar como en el punto de partida, pues ahora también nos interesa determinar si esa secuencia temporal de valores residuales puede o no ser considerada como aleatoria pura.

Una primera idea sobre la presencia de tendencia en la serie la obtendremos en su representación gráfica. Como ya se ha mencionado, los medios más utilizados para detectar y eliminar la tendencia de una serie se basan en la aplicación de filtros a los datos. Un filtro no es más que una función matemática que aplicada a los valores de la serie produce una nueva serie con unas características determinadas. Entre esos filtros encontramos las medias móviles.

Existen otros procedimientos para extraer la tendencia, como ajuste de polinomios, alisado mediante funciones exponenciales, etc. Una clase de filtro, que es particularmente útil para eliminar la tendencia, se basa en aplicar diferencias a la serie hasta convertirla en estacionaria.

La dificultad de inspeccionar la gráfica de una serie de tiempo y por consiguiente tratar de determinar si la serie se comporta como una muestra aleatoria de alguna población estable o si depende del tiempo, es que la variación de una serie aleatoria frecuentemente depende del tiempo. La mayoría de los individuos encuentran en una serie de tiempo patrones regulares de movimiento en la serie, si tal regularidad está presente o no.

Para las series que no son obviamente dependientes del tiempo, es necesario aplicar algunas pruebas estadísticas para aleatoriedad antes de proceder al análisis de la serie. Para seleccionar una prueba de aleatoriedad, es necesario considerar qué alternativas para aleatoriedad serán postuladas.

Los estudios de análisis de series de tiempo se han enfocado principalmente en determinar si existen ciclos en varias series de tiempo económicas. Tales análisis tienden a descomponer una serie de tiempo en una tendencia de largo tiempo, ciclos y efectos aleatorios. La variación estacional en una serie de tiempo, no es tratada como un ciclo de negocios y es frecuentemente removida de la serie antes de ser estudiada. Para calificarla como un ciclo de negocios, el ciclo deberá normalmente exceder un año en longitud.

En vista de lo anterior, la alternativa natural a la aleatoriedad en una serie de tiempo económica es una dependencia en el tiempo que puede involucrar una tendencia de largo plazo, junto con algunos ciclos razonablemente largos.

3.3.- PRUEBAS PARA LA DETECCION DE TENDENCIA.

3.3.1.- PRUEBA DE MANN-KENDALL PARA LA DETECCION DE TENDENCIA.

La prueba de Mann-Kendall es una prueba no paramétrica para la detección de tendencia en una serie de tiempo. Esta prueba es ampliamente utilizada en las ciencias ambientales, porque es simple, robusta y considera los valores perdidos y valores bajo un límite de detección. Desde la primera propuesta de la prueba fue hecha por Mann (1945) y Kendall (1975), las covarianzas entre los estadísticos de Mann, Mann-Kendall fueron propuestos por Dietz y Kileen (1981) y la

prueba fue extendida a fin de incluir estacionalidad (Hirsch y Slack, 1982), sitios de monitoreo múltiple (Lettenmaier, 1988) y covariables representando fluctuaciones naturales (Libiseller and Grimvall, 2002).

La prueba estadística de Mann-Kendall para una serie de tiempo $\{Z_k, k = 1, 2, \dots, n\}$ de datos es definida como

$$T = \sum_{j < i} \text{sgn}(Z_i - Z_j) \quad (1)$$

dónde

$$\text{sgn}(x) = \begin{cases} 1, & \text{si } x > 0 \\ 0, & \text{si } x = 0 \\ -1, & \text{si } x < 0 \end{cases}$$

Si no hay empates entre las observaciones, así como la tendencia no está presente en la serie de tiempo, la prueba estadística es distribuida asintóticamente normal con:

$$E(T) = 0 \quad \text{y} \quad \text{Var}(T) = n(n-1)(2n+5)/18.$$

Si la variable respuesta es medida de varias estaciones (ω), la prueba de Mann-Kendall estacional (o prueba de Hirsch-Slack) es calculada primero separando los datos en ω subseries, cada serie representa una estación

$$T_j = \sum_{k < l} \text{sign}(Z_{lj} - Z_{kj}) \quad j = 1, \dots, \omega \quad (2)$$

es el estadístico de Mann-Kendall para la estación j , el cual es calculado sobre todas las estaciones para obtener

$$S = \sum_{j=1}^{\omega} T_j, \quad (3)$$

que es el estadístico estacional.

S es distribuido asintóticamente normal con media cero y varianza:

$$\text{Var}[S] = \sum_{j=1}^{\omega} \text{Var}[T_j] + \sum_{\substack{j, g=1 \\ g \neq j}}^{\omega} \text{Cov}(T_j T_g), \quad (4)$$

con

$$Var[T_j] = \frac{n_j(n_j - 1)(2n_j + 5) - \sum_{i=1}^m t_i(t_i - 1)(2t_i + 5)}{18}, \quad (5)$$

dónde n_j es el número de observaciones no perdidas para la estación j , m es el número de grupos empatados y t_i es el tamaño del i -ésimo grupo empatado.

La covarianza entre los dos estadísticos Mann-Kendall statistics se define como

$$Cov(T_j T_g) = \left[S_{jg} + 4 \sum_{m=1}^n R_{mj} R_{mg} - n(n_j + 1)(n_g + 1) \right] / 3, \quad (6)$$

donde n_j, n_g son el numero de observaciones no perdidas para la estación j y g , respectivamente, y

$$S_{jg} = \sum_{m < n} sign[(Z_{nj} - Z_{mj})(Z_{ng} - Z_{mg})]. \quad (7)$$

R es la matriz correspondiente a las observaciones en el conjunto de datos, donde las observaciones no perdidas para cada estación y ordenados entre ellos. El rango del m -ésimo elemento dentro de la i -ésima estación es representada por:

$$R_{mj} = \left[n_j + 1 + \sum_{k=1}^n sign(x_{mj} - x_{kj}) \right] / 2 \quad (8)$$

donde $sign(x_{mj} - x_{kj})$ se define ser cero si x_{mj} o x_{kj} esta perdida, a fin de asignar el rango medio $(n_j + 1)/2$ a el valor perdido.

3.3.2.- PRUEBA DEL COEFICIENTE DE CORRELACION DE PEARSON PARA LA DETECCION DE TENDENCIA.

Esta prueba se basa en el coeficiente de correlación de Pearson, el cual se define de la siguiente manera:

$$\varphi = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\left[\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2 \right]^{1/2}}$$

Se utiliza en este análisis debido a la preferencia que los investigadores tienen por esta estadística y sobretodo porque se encuentra prefijado en la mayoría de los paquetes estadísticos.

3.3.3.- COMPARACION DE LAS PRUEBAS DE MANN-KENDALL Y DEL COEFICIENTE DE CORRELACION DE PEARSON.

Las pruebas paramétricas asumen que la variable aleatoria está normalmente distribuida y homocedastica (varianza homogénea). Las pruebas no paramétricas no hacen ninguna suposición respecto a la distribución de probabilidad.

La prueba de t para la detección de tendencia se basa en la regresión lineal, por lo que, verifica solamente una tendencia lineal. No hay tales restricciones para la prueba de Mann–Kendall y en este caso tampoco para la prueba que utiliza el coeficiente de correlación de Pearson.

3.4.- TRANSFORMACIÓN DEL RANGO.

Un problema que los estadísticos aplicados han confrontado desde el comienzo de la estadística paramétrica es el de ajustar los problemas del mundo real en el esquema de la teoría estadística normal cuando muchos de los datos se distribuyen no normal. De tales problemas han emergido dos distintas aproximaciones de pensamiento:

a) Transformar los datos a una forma más cerrada parecida a una normal o b) Usar un procedimiento de libre distribución. El primer método puede incluir la log transformación, la transformación raíz cuadrada, la transformación arcosen y así sucesivamente, y pueden incluso ser suficientemente extensas hasta incluir procedimientos robustos que tienden a dar pequeños pesos a los outliers, es decir a las observaciones que pueden contribuir grandemente a la forma no normal de los datos. El segundo método incluye un gran cuerpo de métodos basados en los rangos de los datos. Hay una forma de combinar estos dos métodos presentando tantos métodos no paramétricos como métodos paramétricos aplicados a los datos transformados. Simplemente se reemplaza los datos con sus rangos, después se aplican las pruebas paramétricas a los rangos. Llamamos a esto la transformación del rango. Esta aproximación resulta en una clase de métodos no paramétricos que incluyen la prueba de Wilcoxon – Mann, la prueba de Kruskal – Wallis, la prueba del rango y signo de Wilcoxon, la prueba de Friedman, la rho de Spearman y otras. La transformación del rango aproxima también métodos útiles en regresión múltiple, análisis de discriminantes, análisis de cluster, análisis de diseños experimentales y comparaciones múltiples.

Existen varias formas en las cuales los rangos pueden ser asignados a las observaciones, que se describen a continuación:

RT-1.- El conjunto completo de observaciones se asigna a los rangos, desde la más pequeña hasta la más grande, con la observación más pequeña se tiene el rango 1, el segundo el rango 2 y así sucesivamente. Los promedios de los rangos serán asignados en caso de empates.

RT-2.- Las observaciones son particionadas en subconjuntos y cada subconjunto se asigna a rangos independientemente entre sí de los otros subconjuntos.

RT-3.- Es la transformación RT-1 aplicada después de algunas reexpresiones apropiadas de los datos.

RT-4.- Es la transformación descrita en RT-2 aplicada a algunas transformaciones apropiadas de los datos.

De esta forma, la transformación del rango proporciona una útil técnica pedagógica para introducir estos métodos no paramétricos como una parte integral en un curso introductorio de estadística, en lugar de los métodos comunes, en una unidad separada que puede parecer al estudiante estar desconectado del fluido general del curso. También, permite al practicante hacer uso completo de los paquetes estadísticos existentes que no pueden tener programas no paramétricos adecuados, simplemente introduciendo los rangos de los datos en los programas para el análisis paramétrico. Finalmente, esta aproximación puede ser vista como una útil herramienta para desarrollar nuevos métodos no paramétricos en situaciones donde existen los procedimientos paramétricos satisfactorios.

La técnica de reemplazar datos con sus rangos y después usar el análisis paramétrico sobre estos rangos ha sido mostrado por los autores tener propiedades de robustez y potencia en ambas regresión (Iman y Conover, 1979) y análisis de varianza (ANOVA) (Iman, 1974; Conover e Iman, 1976; e Iman y Conover en un reporte técnico de 1976, de esta forma se ve natural extender la aproximación a Análisis de Covarianza (ANCOVA) el cual es una combinación de regresión y ANOVA. Una bibliografía actualizada de este trabajo ha sido compilado por los autores (Conover e Iman, 1981).

Una de las primeras aplicaciones de la transformación del rango involucra el cálculo del coeficiente de correlación producto – momento de Pearson tal y como se ha definido anteriormente.

O bien, sobre los rangos para obtener el coeficiente rho de Spearman.

$$\varphi = \frac{\sum \left(R(x_i) - \frac{n+1}{2} \right) \left(R(y_i) - \frac{n+1}{2} \right)}{\left[\sum \left(R(x_i) - \frac{n+1}{2} \right)^2 \sum \left(R(y_i) - \frac{n+1}{2} \right)^2 \right]^{1/2}}$$

Para datos apareados $(X_1, Y_1), \dots, (X_n, Y_n)$ donde las observaciones dentro del subconjunto $\{X_i\}_{i=1}^n$ se les asigna rangos entre sí, y la misma es verdadera para el subconjunto $\{Y_i\}_{i=1}^n$, esto es un ejemplo de un procedimiento tipo RT-2, señalado anteriormente. Dado que r es

una medida de linealidad de la relación entre X y Y, también será una medida de la linealidad entre los rangos de X y los rangos de Y, la cual se traduce como una medida de monotonicidad en la relación entre X y Y.

La extensión directa a correlación múltiple es obvia. Las observaciones sobre cada componente x_{ij} de $X_j=(x_{1j}, \dots, x_{kj})$, $j = 1, \dots, n$ con sus rangos por separado desde 1 hasta n. Correlaciones múltiples, correlación parcial, pueden ser calculados sobre los rangos tal como serán calculados sobre los datos.

La adaptación de los procedimientos de correlación a los rangos de los datos inmediatamente sugiere adaptar los métodos de regresión a datos del tipo RT-2. Mínimos cuadrados, los métodos de regresión hacia atrás o hacia delante, o cualquier otro método de regresión puede ser aplicado a los rangos de las observaciones (Iman y Conover 1979). El resultado es que el rango de la variable dependiente es predicha usando los rangos de las variables independientes. El rango predicho $\hat{R}(Y_i)$ puede ser transformado de regreso al valor predicho \hat{Y} de Y, por interpolación lineal entre los dos valores de Y que tienen rangos delimitados $\hat{R}(Y_i)$

Otra forma de usar una transformación del rango para desarrollar métodos no paramétricos en regresión es para asumir que la ecuación de regresión es lineal y usar mínimos cuadrados ordinarios a fin de obtener una ecuación de regresión estimada $Y = a + bx$. El método paramétrico de probar la hipótesis $\beta=\beta_0$, utiliza la siguiente estadística.

$$t = \frac{(b - \beta_0) \left[(n - 2) \sum (x_i - \bar{x})^2 \right]^{1/2}}{\left[\sum (y_i - \bar{y} - b(x_i - \bar{x}))^2 \right]^{1/2}}$$

La cual tiene una distribución t bajo la suposición de normalidad. Es más fácil ver como una transformación del rango conduce a una prueba no paramétrica si t es rearrreglada como sigue:

$$t = \frac{\sum (x_i - \bar{x})(y_i - \bar{y} - \beta_0(x_i - \bar{x}))\sqrt{n-2}}{\left[\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y} - b(x_i - \bar{x}))^2 \right]^{1/2}}$$

Note que, excepto por una constante multiplicativa, el numerador es la covarianza muestral de X y los residuales hipotetizados, además el denominador tiene la varianza muestral de X y los mínimos cuadrados de los residuales. Reemplazar la covarianza original con la covarianza entre los rangos de X y los rangos de los residuales hipotetizados, los cuales están en una transformación del tipo RT-4. Si se usan las varianzas de aquellos rangos en el denominador se obtiene la prueba no paramétrica para los sesgos propuestos por Hogg y Randles (1975).

Una extensión natural de la transformación del rango aproxima al modelo lineal general consiste de asignar rangos a cada variable cuantitativa por separado en el modelo lineal general y aplicar los procedimientos paramétricos usuales.

Esta es una aplicación del tipo RT-2. El resultado es nuevas pruebas para sesgos iguales, análisis de covarianza (Conover e Iman 1980b), y algún diseño experimental cubierto por el modelo lineal general. Estas pruebas son en general no de distribución libre, excepto quizá en algún sentido asintótico, pero pueden ser más robustas y potentes que sus competidores en situaciones no normales. Cada procedimiento, sin embargo, necesita ser evaluada sobre sus propios meritos. Otro tipo de procedimiento RT-2 para el análisis de covarianza es dado por Quade (1967).

3.4.1.- VENTAJAS DE LA TRANSFORMACION DEL RANGO.

Los métodos no paramétricos deberán estar entre las herramientas de trabajo de un estadístico. La transformación del rango proporciona un vehículo para presentar ambos métodos, tanto el paramétrico como el no paramétrico en una manera unificada. Esto capacitará a los principiantes en estadística a entender las diferencias y similitudes de los dos tipos de análisis. También la transformación del rango conduce de manera más fácil a los métodos computacionales, donde es frecuentemente más conveniente introducir rangos en un programa para un análisis paramétrico que es encontrar o escribir un programa para un análisis no

paramétrico. La mayoría de los programas existentes para métodos no paramétricos no incorporan correcciones para empates, mientras que la transformación del rango automáticamente hace las correcciones requeridas para empates. El usuario usa promedios de los rangos dondequiera que un empate ocurra.

Otras anotaciones pueden ser usadas en lugar de los rangos, si se desea, para obtener pruebas no paramétricas que son equivalentes a pruebas tales como la prueba de Van Der Warden, la prueba de Capon, la prueba de la mediana, la prueba de McNemar y otras. La experiencia indica que los rangos a sí mismos proporcionan puntuaciones que son difíciles de mejorar en su uso.

Algunas limitaciones de la transformación del rango deben notarse aquí también. El procedimiento de la transformación del rango conduce a pruebas de libre distribución en algunos casos, mientras que en otros casos la prueba resultante puede ser solamente condicionalmente de libre distribución, asintóticamente de libre distribución, o ninguna de ellas. Un ejemplo del último caso surge cuando la tasa de las dos varianzas muestrales es calculado sobre datos tipo RT-1 y comparados con tablas de la distribución F como una prueba para varianzas iguales en dos muestras independientes. Resultados de simulación indican una severa falta de robustez para este procedimiento, aún cuando los medios poblacionales son iguales.

Esto está probablemente relacionado al hecho de que la prueba de F paramétrica para este problema es notoriamente sensible a la normalidad, y el teorema central del límite no aplica en esta situación. Una situación similar existe cuando la prueba t de Welch es usada sobre los rangos de los datos en la esperanza de resolver el problema no paramétrico de Behrens-Fisher.

3.5.- LOS ERRORES ESTADISTICOS Y LA PRUEBA DE LA POTENCIA.

La mención de la existencia del error tipo II es sorprendentemente ausente en la literatura reciente. La ausencia de alguna discusión sobre el error tipo II puede significar que muchos de aquellos involucrados en la fuentes naturales del manejo e investigación están olvidados a la posibilidad de cometer estos errores y que una discusión de ellos puede ser de valor.

Los conceptos presentados son aplicables a toda fuente natural de decisión hechas para quienes utilizan los procedimientos de pruebas de hipótesis. El conocimiento del error tipo II debería ayudar a aquellos quienes utilizan las pruebas de hipótesis para entender mejor las ventajas y limitaciones inherentes en su uso.

La probabilidad de cometer un tipo de error es igual al nivel de significancia o tamaño de la prueba (Mood et al. 1974). Controlar el error tipo I es rutinario y es la base para escoger el nivel de significancia. Por otro lado, la probabilidad de cometer el error tipo II es raramente considerado, excepto indirectamente, siguiendo el procedimiento aceptado de formular la hipótesis de tal forma que es más importante (Neyman 1942, Cohen 1977). En muchas fuentes naturales esto no es una solución aceptable porque ambos tipos de errores son importantes.

Un error tipo I (rechazar H_0 cuando es verdadera) ocurre cuando el investigador decide que la tendencia existe cuando realmente no existe, si la tendencia implicada es creciente el investigador puede erróneamente decidir incrementar una provisión o abastecimiento y una consecuencia grave puede ocurrir. Inversamente, si la tendencia implicada es decreciente, el investigador puede erróneamente decidir disminuir el abastecimiento y por consiguiente originar una pérdida de algún tipo. Un error tipo II (no rechazar H_0 cuando es falsa) puede resultar en consecuencias similares. Si el investigador decide que no existe tendencia, entonces ningún ajuste en el abastecimiento es indicado. Si la tendencia es decreciente, esta falta de acción puede ocasionar un decrecimiento acelerado en la tendencia. De manera inversa, si una tendencia creciente no es detectada, puede ocurrir una pérdida de algún tipo.

Estimadores exactos de la condición del rango y la tendencia son necesarios a fin de utilizar efectivamente el potencial productivo del área de estudio. Es importante ser capaz de detectar algún cambio en la condición sobre la cual es suficientemente grande señalar un manejo inefectivo. Cuando se prueba la hipótesis nula de ningún cambio en la condición del rango (H_0 : Hay tendencia), un error tipo II (no rechazar H_0 : cuando es falsa) es, en esencia, la incapacidad de detectar un cambio en la condición de rango (No hay tendencia) cuando de hecho un cambio ha ocurrido. Los errores tipo II asociados con esta hipótesis pueden resultar en decisiones incorrectas y consecuentemente, en manejos ineficientes. De manera inversa, la habilidad de

detectar un cambio en la condición, capacita al investigador a ajustar los niveles de abastecimiento como son requeridos.

En terminología estadística, la probabilidad de detectar un cambio en la condición de rango (rechazar H_0 cuando es falsa), es referida como la potencia de la prueba y, asociada con alguna prueba de hipótesis existe una curva de potencia ($\pi(\theta)$). La curva de potencia asociada con la prueba de la hipótesis nula de tendencia del rango da la probabilidad de detectar un cambio en la condición del rango (θ) de una magnitud dada. Como se esperaría, θ , la magnitud de la condición del rango incrementa, así también la probabilidad de detectar ese cambio, es decir, la magnitud del rango de tendencia (θ) y la potencia (π) están directamente relacionadas.

3.5.1.- NIVEL DE SIGNIFICANCIA Y POTENCIA.

La metodología típica usada cuando se aplican técnicas estadísticas de pruebas de hipótesis es escoger un pequeño nivel de significancia ($\alpha = \text{Prob}[\text{Error tipo I}]$) para controlar los errores tipo I mientras da poca o ninguna consideración a errores del segundo tipo (Lehmann 1958). Esta práctica es cuestionable siempre que las consecuencias asociadas con errores Tipo II son importantes. Un problema con el que ciegamente se selecciona un nivel de significancia pequeño es aquel que reduciendo el nivel de significancia y la probabilidad de error tipo I simultáneamente incrementa el tamaño del error tipo II. (Neyman et al. 1935, Sokal y Rohlf 1969).

3.5.2.- TAMAÑO MUESTRAL Y POTENCIA.

Cuando el tamaño muestral es fijado, puede no ser posible reducir las probabilidades de errores Tipo I y Tipo II, hasta límites aceptables, y obtener una aceptable aproximación la cual balancee las probabilidades de los errores Tipo I y Tipo II, el investigador puede esperar cometer un error en 10 a 20% de la prueba de hipótesis realizado. Si esto no es aceptable, aumentando el tamaño muestral debería ser considerada.

La curva de potencia idealizada, previamente discutida, puede ser usada para seleccionar el tamaño muestral en la misma forma en que fue usada para seleccionar el nivel de significancia.

El análisis de potencia estadístico proporciona las herramientas analíticas. Estas herramientas pueden ayudar en el desarrollo de un rango de condición para el programa de mediciones y procedimientos de pruebas de hipótesis asociadas. Pueden también ser usadas para investigar programas desarrollados para efectuar las mediciones presentes y pasadas a fin de juzgar su validez y utilidad. La técnica no está limitada al estudio de programas de mediciones, también pueden ser aplicadas a hacer decisiones de tipo ambiental donde las pruebas de hipótesis sean utilizadas.

El valor del análisis de potencia estadístico no es restringido al investigador. Es quizá de mayor importancia para quien usa las técnicas estadísticas para guiar y justificar las investigaciones en la toma de decisiones. Es el resultado de esta investigación, más que los resultados de las pruebas de hipótesis, sobre las cuales el administrador típicamente confía cuando toma decisiones administrativas. La introducción de los conceptos de errores estadísticos en publicaciones de fuentes naturales deben ser benéficos tanto para los investigadores, como para los que ejecutan los planes de acción. Un trabajo en el conocimiento de estos conceptos deberá resultar en propósitos de investigación y toma de decisiones, las cuales en ambos casos son validos y mas efectivos.

La habilidad de un procedimiento estadístico para distinguir una situación diferente de la hipótesis nula es llamada la potencia de ese procedimiento. La estimación de la potencia estadística es importante por varias razones, así como para la interpretación de resultados cuando la hipótesis nula no es rechazada (Toft y Shea 1983). Operacionalmente, el análisis de potencia es importante durante la planeación de experimentos para evitar la pérdida de tiempo y esfuerzo sobre un programa que es improbable para rendir información útil. Peterman y Routledge (1983).

Por definición, una tendencia es detectada cuando la regresión tiene un sesgo significativamente diferente de cero. La conclusión de que una tendencia está ocurriendo, cuando

de hecho no lo es, se denomina error tipo I, mientras que la conclusión de que no hay tendencia esta ocurriendo, cuando de hecho está, es un error tipo II.

En general, para alguna prueba estadística, la potencia es una función del tamaño muestral (n), la probabilidad de error tipo 1 (α), y la magnitud de la diferencia entre la hipótesis nula y la realidad (el “efecto tamaño”, Cohen 1977). Aquí el efecto tamaño es cuantificado por la tasa de cambio por el parámetro r . Además, será considerado que, en muchas aplicaciones ecológicas, la medida del error no es trivial. Esto significa que un análisis de potencia práctico considera la incertidumbre o variabilidad asociada con cada estimador de abundancia. Esa variabilidad es reportada por el coeficiente de variación (cv) del estimador.

Hay dos situaciones generales en las cuales un análisis de potencia es útil: en el diseño del experimento, y en la interpretación de los resultados. Cuando un experimento o programa de muestreo está siendo diseñado, las preguntas más probables a ser solicitadas son: ¿Cuántas muestras serán necesarias?, ¿Qué tan precisas serán las muestras?, ¿Cuál es la probabilidad de detectar una tendencia?. Estas preguntas son contestadas resolviendo para n , cv y β respectivamente. Allen y Kirkwood (1976) analizaron la viabilidad de detectar los resultados de manipulaciones experimentales de abastecimiento de ballena; y concluyeron que tal programa no es factible. Cuando el análisis de potencia es usado para ayudar en la interpretación de resultados, particularmente cuando la hipótesis no es rechazada las preguntas más probables a ser de interés son: ¿Qué tan grande puede una tendencia haber sido detectada?, ¿Cuál fue la probabilidad de detectarla?, en este caso resolvemos para r y β .

Estos cinco parámetros dependen de las suposiciones hechas acerca del proceso que se sigue en la investigación, ocasionando la tendencia y las técnicas usadas para detectarla.

Los modelos investigados explícitamente en temas relacionados con las ciencias ambientales y esporádicamente en temas económicos, incluyen las suposiciones que cambian en abundancia y toman lugar en incrementos constantes (el modelo lineal) o en una tasa constante (el modelo exponencial), y el coeficiente de variación es constante, proporcional a la raíz cuadrada de la abundancia, o proporcional a la inversa de la raíz cuadrada de la abundancia.

En general, los resultados muestran que la detección de tendencia depende fuertemente sobre el número y precisión de las muestras. Dado un número razonable de muestras (digamos 10-15), cv será suficientemente bajo, o r será suficientemente alto, para detectar tendencias como error tipo 2 controlado. No podemos controlar r , excepto indirectamente por muestreo en grandes intervalos, pero no podemos controlar cv , en por lo menos en la porción debida al error medido, cv puede ser reducido gastando en muestras más grandes, mayor esfuerzo o mayor área de cobertura, por ejemplo. También puede ser reducido tomando réplicas, donde la media de varias medidas independientes tendrá una más baja varianza que las medidas a sí mismas. Es difícil generalizar acerca de lo que se espera acerca del coeficiente de variación, porque eso depende del tema de estudio en particular, sobre cual es la cantidad de interés y sobre qué técnica es usada para estimar sus parámetros. Eberhardt (1978) discute variabilidad en una variedad de técnicas usadas en estudios poblacionales.

Una más sutil edición del balanceo de los errores Tipo I y Tipo II. ¿Cuáles niveles de error son aceptables? depende de los objetivos del investigador (Rotenberry y Wiens 1985). Es práctica común adoptar a $\alpha=0.05$ como el más grande nivel de significancia en el cual la hipótesis nula es rechazada, pero no hay tal práctica común con respecto a β , la cual es, por analogía, el nivel en el cual la hipótesis nula es aceptada. Muy a menudo, la hipótesis nula es implícitamente aceptada sin alguna consideración de potencia. Si uno está cuidando de decidir entre una hipótesis nula y una alternativa definida, la simetría dictaría que α y β deberían ser iguales. El problema es que β , y su complemento, potencia, frecuentemente no es fácil de estimar.

Algunas suposiciones de este análisis pueden limitar la aplicación de los resultados. La mayoría de las suposiciones restrictivas es que los puntos son igualmente espaciados a lo largo del eje X. Esto significa que las muestras son asumidas ser tomadas en intervalos de tiempo regulares (diariamente, mensualmente, anualmente) o distancias (cada metro, cada kilómetro, etc). Estos resultados son robustos a moderadas violaciones de esta suposición, pero si las muestras son agrupadas cerca de algún término del eje x, pueden resultar cálculos bastante diferentes de la potencia. Es también asumido que los estimadores de abundancia son tomados similarmente cada vez, así que un único coeficiente de variación es suficiente para describir

todas las muestras, y que, como un corolario, todas las muestras son dadas en igual ponderaciones en la regresión. En algunas aplicaciones las regresiones ponderadas serán más apropiadas, o darán estimadores más bajos de varianza (de la Mare 1984). Otra importante suposición es que los estimadores son independientes. Esto es un interés particular para series de tiempo, donde los estimadores temporalmente cercanos a cada uno pueden tender a ser más similares que los estimadores lejanos en el tiempo. Tal autocorrelación positiva significa que la varianza residual puede ser seriamente menospreciada.

Esto tiene importantes consecuencias. Si los estudios diseñados son basados en la varianza subestimada, la potencia de detectar una tendencia será menor que la planeada (error tipo II inflado). Y si los datos con autocorrelación positiva son analizados por mínimos cuadrados ordinarios, una tendencia puede existir cuando de hecho no existe (error tipo I inflado). Harris (1986) reportó simulaciones que violaron la suposición de independencia, así como otras suposiciones de detectar tendencia. Él enfatizó que para datos colectados bajo condiciones de campo típicas, cálculos analíticos tenderán a subestimar el error estándar de la línea de regresión.

La especificación previa de variabilidad de las muestras es uno de los más difíciles problemas cuando el análisis de potencia es usada en diseños experimentales. Hay dos fuentes generales de variabilidad: el error medido, y la variación real en los estimadores. En un contexto ecológico, la variación real significa que, incluso si podemos medir sin error, los puntos nunca caerían precisamente a lo largo de una línea recta, ya sea debido a la variabilidad ambiental natural en tiempo o espacio, o porque los procesos que produce la tendencia no actúa en una línea exactamente o exponencial. Aunque estas dos fuentes de variabilidad son distintas conceptualmente, son difíciles de separar en la práctica. Afortunadamente, no es necesario separarlos para este análisis; ambos tipos de variabilidad pueden ser combinados en el parámetro cv . El problema recae en especificar el cv correctamente para reflejar ambos tipos. Si las réplicas medidas son tomadas, la varianza resultante es un estimador de la medida del error, y es un estimador mínimo de la variabilidad para ser esperada acerca de la línea de regresión porque no incluye la variación real. Si el cv es estimado de la varianza residual acerca de la línea de regresión, esta incluirá la variabilidad real así como medida del error, pero por supuesto tal

información estará disponible solamente después de que los datos han sido colectados. Esto es importante para interpretar los resultados, pero para propósitos de planeación, nos gustaría tener un estimador de la precisión esperada antes de que el programa de muestreo es iniciado. Para hacer planes más realistas, alguna variabilidad adicional debida al ambiente deberá ser incluida. Si separamos los estimadores de la varianza debida a la medida del error y para la variación real están disponibles, pueden ser combinadas.

Considérese la detección de tendencia en tamaño poblacional sobre el tiempo. Las dinámicas poblacionales de muchos insectos, aves, pequeños mamíferos, y animales marinos son muy susceptibles a condiciones ambientales. El reclutamiento o producción de nuevos individuos puede fluctuar substancialmente desde un año al próximo, dependiendo de que tan favorable es el ambiente; en otras palabras, hay una variación real en r . Si deseamos usar el análisis de potencia de este documento para detectar una tendencia media en el tamaño poblacional, será importante que el estimador de cv incluya esta variación real, la cual puede bien ser más importante que la medida del error. Si somos capaces de estimar la variación real en r , la aplicación de los resultados será limitado para tales especies.

El uso del análisis de regresión lineal para detectar una tendencia asume que el cambio en abundancia es lineal, o ha sido hecho lineal por transformaciones adecuadas de los ejes. Frecuentemente, sin embargo, el proceso fundamental de la tendencia son oscuramente percibidas. Los dos modelos simples de cambio considerados aquí, lineal y exponencial, son dos formas en las cuales la abundancia de una cantidad pudo cambiar regularmente sobre el tiempo y el espacio. En la ausencia de información más detallada al contrario, el modelo lineal es usualmente asumido en la práctica. El modelo exponencial es frecuentemente utilizado para procesos de decadencia o crecimiento. Si otro modelo es conocido ser más apropiado para una situación particular, un análisis de potencia similar al presentado aquí será llevado para otro modelo. La situación usual, sin embargo, es que poco es conocido acerca de que tanto interés esté cambiando sobre el tiempo o espacio, esta es la motivación del muestreo.

Los métodos estadísticos conocidos requieren el uso de muestras aleatorias para la decisión o estimación apropiada. Desafortunadamente, muchos conjuntos de observaciones que

han sido tomadas sobre un periodo de tiempo no se comportan como muestras aleatorias de alguna población estable. Esto es particularmente verdadero de ciertos conjuntos de datos económicos, tales como el precio de acciones, el costo de vida o el consumo de algún producto. Como consecuencia, las técnicas estadísticas estándar no pueden ser aplicadas a tales datos.

Los economistas en particular han estudiado tales series tan extensamente porque son tantos los problemas de interés en economía. También han sido estudiados en las ciencias físicas en conexión con fenómenos periódicos de varios tipos.

Al estudiar las series de tiempo, la primera pregunta que necesita ser contestada es si la serie realmente depende del tiempo, para ello se utilizan pruebas no paramétricas.

El siguiente paso es estimar la naturaleza de esta dependencia sobre el tiempo, donde solamente los más grandes movimientos de la serie sobre el tiempo son de mayor interés, esto significa ajustar una curva suavizada a la gráfica de la serie, similar al ajuste de una línea de regresión a un conjunto de puntos.

Los métodos de regresión están disponibles para obtener tal estimador, sin embargo, otro método basado en promedios móviles es comúnmente empleado para obtener esta estimación.

Después de que la porción mayor de la dependencia de las series sobre el tiempo ha sido estimada, el problema final es determinar si la relación resultante puede ser usada para predecir el futuro de series de tiempo económicas es, por supuesto, el sueño y deseo de todo economista. Muchas técnicas estadísticas han sido propuestas para hacer tales predicciones.

3.6.- MODELO Y ESTIMADORES.

El modelo usado en este trabajo

$$(1) \quad Y_t = \alpha + \beta t + \varepsilon_t \quad ,$$

$$(2) \quad \varepsilon_t = \rho \varepsilon_{t-1} + u_t \quad (t = 1, \dots, T)$$

Donde $|\rho| < 1$, $E(u_t) = E(u_t u_s) = 0$ ($t \neq s$), $V(u_t) = \sigma^2$

y el proceso ε_t es estacionario, es decir: $V(\varepsilon_t) = \frac{\sigma^2}{(1-\rho^2)}$. Nuestro interés es enfocado sobre los estimadores de el coeficiente β , y los estimadores a ser considerados son MCO y el estimador de mínimos cuadrados generalizados (MCG), el cual es mejor lineal insesgado (MELI) en este caso.

En el contexto del modelo simple (1), las fórmulas de las varianzas de estos estimadores se reducen a:

$$V(MCO) = 12\sigma^2 \{-6\rho^{T+1}[(T-1)\rho - (T+1)]^2 - (T^3 - T)\rho^4 + 2(T^2 - 1)(T-3)\rho^3 + 12(T^2 + 1)\rho^2 - 2(T^2 - 1)(T+3)\rho + (T^3 - T)\} / (1-\rho^2)(1-\rho)^4 (T^3 - T)^2$$

$$V(MCG) = 12\sigma^2 / (T-1)[(T-3)(T-2)\rho^2 - 2(T-3)(T+1)\rho + T(T+1)]$$

3.7.- DISTRIBUCIONES A UTILIZAR.

3.7.1.- DISTRIBUCION DE VALOR EXTREMO GENERALIZADA TIPO I

Su función de distribución es:

$$F(x; \mu, \sigma, \xi) = \exp\left[-\exp\left(\frac{x-\mu}{\sigma}\right)\right] \quad \text{para } \frac{x-\mu}{\sigma} > 0, \text{ donde } \mu \in R \text{ es el parámetro de}$$

localización, $\sigma > 0$ el parámetro de escala.

La función de densidad:

$$f(x; \mu, \sigma) = \exp\left[-\exp\left(\frac{x-\mu}{\sigma}\right)\right] \left[-\exp\left(\frac{x-\mu}{\sigma}\right)\right] \left(\frac{1}{\sigma}\right)$$

$$\mu \in [-\infty, \infty], \quad \sigma \in (0, \infty], \quad \xi \in [-\infty, \infty]$$

La función cuantil es:

$$Q(x) = -[\log(\log(p))]$$

3.7.2.- DISTRIBUCION EXPONENCIAL.

La función de distribución es:

$$F(x, \beta) = 1 - \exp\{-\beta x\} \text{ para } x > 0$$

La densidad es:

$$f(x, \beta) = \begin{cases} \beta e^{-\beta x} & \text{para } x > 0 \\ 0 & \text{para } x \leq 0 \end{cases}$$

La función cuantil es:

$$Q(x) = -\frac{\log(1-p)}{\beta}$$

3.7.3.- DISTRIBUCION PARETO

La función de distribución es:

$$F_x(x; \theta) = \begin{cases} 1 - \left(\frac{\theta}{x}\right)^\alpha & \text{si } x > \theta \\ 0 & \text{si } x \leq \theta \end{cases}$$

Función de densidad:

$$f(x) = \begin{cases} \frac{\alpha \theta^\alpha}{x^{\alpha+1}} & x \geq \theta; \quad \alpha > 1 \\ 0 & \text{de otra forma} \end{cases}$$

La función cuantil es:

$$Q(x) = \frac{\theta}{(1-p)^{1/\alpha}}$$

La distribución Pareto encuentra aplicación en la modelación de problemas que involucran distribuciones de ingresos, cuando éstos exceden un cierto límite x_0 .

3.7.4.- DISTRIBUCION NORMAL.

Densidad normal:

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad -\infty < \mu < \infty \quad y \quad \sigma > 0$$

Una gran cantidad de las técnicas usadas en estadística aplicada son basadas en la distribución normal.

3.8.- POTENCIA DE UNA PRUEBA Y DETECCION DE TENDENCIAS

3.8.1.- POTENCIA DE LA PRUEBA

Para calcular la potencia de una prueba en forma analítica es necesario conocer la distribución del estadístico de prueba bajo H_1 . Para el caso de las pruebas consideradas no se conoce la distribución de sus estadísticos de prueba, pero utilizando simulación es posible obtener estimadores de la potencia, haciendo uso de su definición, misma que se presenta enseguida.

La función de potencia de una prueba de hipótesis estadística H_0 contra la alternativa H_1 , está dada por:

$$\beta(\theta) = \begin{cases} \alpha(\theta) & \text{para valores de } \theta \text{ bajo } H_0 \\ 1 - \pi(\theta) & \text{para valores de } \theta \text{ bajo } H_1 \end{cases}$$

En la que:

$$\alpha(\theta) = \Pr [\text{Error tipo I}] = \Pr [\text{Rechazar } H_0 \mid H_0 \text{ es verdadera}]$$

$$\pi(\theta) = \Pr [\text{Error tipo II}] = \Pr [\text{No rechazar } H_0 \mid H_0 \text{ es falsa}]$$

La probabilidad de error tipo I mide la capacidad de discriminación del estadístico de prueba empleado para discriminar entre H_0 y H_1

El juego de hipótesis que se emplea para la simulación es:

H_0 : No hay tendencia

contra

H_1 : Hay tendencia

El propósito de la prueba de tendencia es determinar si los valores de una variable aleatoria generalmente incrementa (o decrece) sobre algunos periodos de tiempo en términos estadísticos. Las pruebas estadísticas paramétricas o no paramétricas pueden ser usadas para decidir si hay una tendencia estadísticamente significativa. Todas las pruebas estadísticas involucran dos tipos de errores. El error tipo I (rechazar la hipótesis nula cuando es verdadera) y el error tipo II (no rechazar la hipótesis nula cuando es falsa). Es importante considerar la potencia de una prueba, donde la potencia es definida como uno menos la probabilidad de error tipo II. Una prueba potente rechazará una hipótesis falsa con una alta probabilidad.

3.9.- SIMULACION

Todas las hojas de cálculo en los software estadísticos, así como muchas otras máquinas, incluyendo las calculadoras de bolsillo, proporcionan al usuario, a partir de métodos simples, una forma simple de generar números aleatorios. Aunque hay variaciones sobre el tema, el número aleatorio es un número entre (0, 1) que representa una observación sobre una distribución uniforme continua. En idioma cuantil la función cuantil es:

$$S(p) = p, \quad 0 \leq p \leq 1$$

y la función de densidad de probabilidad de p es:

$$f_p(p) = 1, \quad 0 \leq p \leq 1$$

Los números aleatorios generados por el software no son totalmente aleatorios, son pseudo-aleatorios, siendo generados por algoritmos numéricos determinísticos. Frecuentemente estos algoritmos requieren un valor de inicio, denominado semilla. En uso normal es usada una diferente semilla cada vez que una sucesión de valores es requerida. Sin embargo, es posible y algunas veces útil generar repetidamente el mismo flujo de números aleatorios. Aunque la terminología varía, una notación común que usaremos es dejar que un valor dado por tal algoritmo sea denotado por u , así podemos escribir $u = \text{RAND}$. Un flujo de tales observaciones uniformes es denotado por u_1, u_2, \dots, u_n . El mecanismo generador es diseñado para producir un flujo de valores aproximadamente independientes. Nos referimos a la generación de variables

aleatorias en tal forma, y también al uso de tales valores en la investigación de un modelo de algún tipo, como simulación.

4.- METODOLOGIA

Una forma de examinar la robustez y la potencia de los procedimientos es por simulación bajo las hipótesis nula y alternativa, con ambos tipos de distribuciones, tanto normal como no normal.

Para estudiar la potencia de cada prueba se realizó un estudio de Monte Carlo con 10,000 repeticiones, para tamaños de muestra $n=12$, 60 y 120. Con valores de $\alpha = 0.01$, 0.05 y 0.1. Para el modelo autorregresivo mencionado, se propone que $|\rho|<1$, por lo que se proponen los valores 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.99. Además se propone verificar que las pruebas detectan la tendencia, para ello, se asignan valores de tendencia conocidos que van desde 0, 0.001, 0.005, 0.01, 0.025, 0.05, 0.075, 0.09, 0.1, 0.3, 0.5. Esto cuando las innovaciones siguen la distribución de valor extremo generalizado tipo 1, Exponencial, Pareto y normal.

La distribución de valor extremo generalizado tipo I fue elegida para el estudio debido a su aplicación en la modelación de observaciones extremas, de la misma manera que la distribución Pareto dada su relevancia en la aplicación de problemas relacionados con el ingreso. En el caso de la distribución exponencial, por su relevancia en datos que implican confiabilidad y la distribución normal por su vasta aplicabilidad en teoría estadística. Estas distribuciones son utilizadas en la modelación de datos económicos y ecológicos que tienen relación directa con el tiempo.

De forma análoga, con el paquete estadístico R, se simulan los datos primero sin transformarlos y posteriormente transformados con el rango, para comparar la prueba de Mann-Kendall y el coeficiente de correlación de Pearson, por lo que se obtienen las potencias con los respectivos programas (ver apéndice), en cuatro diferentes formas: Mann-Kendall a datos sin transformar, Mann-Kendall a datos transformados con sus rangos, coeficiente de correlación de Pearson a datos sin transformar y coeficiente de correlación de Pearson a datos transformados

con sus rangos. En este último caso, se obtiene la prueba rho de Spearman, lo cual da un margen más amplio de comparación de las pruebas.

5.- ANALISIS DE RESULTADOS

El análisis realizado para las distribuciones Pareto, exponencial, normal y de valor extremo generalizado tipo I, se encontró que depende primeramente del término autorregresivo (α), luego de n , y del tamaño de la prueba alfa, de esta forma, para valores de α cercanos a cero tanto para los datos sin transformar, como para la transformación del rango, se obtienen las mejores pruebas principalmente para valores muy pequeños de tendencia, cuando se utiliza la distribución Pareto en el caso de la prueba de Mann – Kendall y de valor extremo generalizado tipo I para la prueba utilizando el coeficiente de correlación de Pearson.

A medida que el valor autorregresivo es mayor a 0.1 las pruebas siguen un comportamiento similar, solo que en el caso de la distribución de valor extremo generalizado tipo I se detecta mejor la tendencia para valores mayores a 0.05 y la densidad Pareto para valores mayores a 0.1, sin embargo las mejores pruebas se obtienen cuando la distribución es normal, esto se cumple para las pruebas a datos sin transformar como transformados.

Para un $\alpha = 0.5$ y mayor, las mejores pruebas se obtienen cuando las innovaciones son distribuidas normalmente, esto para detectar valores de tendencia menores a 0.1 y mayores a cero. Para valores de tendencia de 0.3 y 0.5 las mejores pruebas se obtienen cuando las innovaciones se distribuyen Pareto, o de valor extremo generalizado tipo I.

Cuando las innovaciones se distribuyen exponencial, las pruebas parecen no funcionar, pues en todos los casos se obtiene mejor potencia cuando las innovaciones tienen distribución normal, Pareto o de valor extremo generalizado tipo I.

Las dos pruebas funcionan mejor para tamaños de muestra pequeños, en el caso analizado para $n = 12$ y para niveles de significancia pequeños, digamos $\alpha = 0.01$

Tabla 1. Potencia de las pruebas para datos sin transformar con $ar = 0.01$ y $\alpha = 0.05$

			MANN-KENDALL A DATOS SIN TRANSFORMAR				PRUEBA DE PEARSON A DATOS SIN TRANSFORMAR			
α	n	tend	DVEG Tipo 1	Expon	Pareto	Normal	DVEG Tipo 1	Expon	Pareto	Normal
0.05	12	0	0.0406	0.0416	0.0517	0.038	0.0501	0.0505	0.0288	0.0503
		0.001	0.0451	0.0449	0.0562	0.045	0.0508	0.048	0.0291	0.0586
		0.005	0.0469	0.0478	0.0568	0.0449	0.0522	0.0466	0.0298	0.0513
		0.01	0.0488	0.0499	0.0569	0.0498	0.0532	0.0488	0.0332	0.0535
		0.025	0.0521	0.0686	0.0604	0.055	0.0555	0.0588	0.0305	0.0614
		0.05	0.0767	0.1188	0.0872	0.0757	0.0764	0.1011	0.0396	0.089
		0.075	0.1005	0.2137	0.1146	0.1127	0.1053	0.1696	0.0459	0.1277
		0.09	0.1222	0.2703	0.132	0.1379	0.1298	0.2195	0.0484	0.1706
		0.1	0.1379	0.304	0.1566	0.1743	0.1467	0.2613	0.056	0.1966
		0.3	0.7067	0.9174	0.4548	0.834	0.7295	0.8824	0.1922	0.8955
	0.5	0.9716	0.9945	0.6431	0.9975	0.9739	0.9919	0.3201	0.9999	
	60	0	0.0489	0.0491	0.0573	0.0519	0.0532	0.0515	0.0336	0.0507
		0.001	0.0533	0.0572	0.062	0.053	0.0509	0.0545	0.0316	0.0554
		0.005	0.0861	0.1852	0.1158	0.0946	0.0882	0.1085	0.0341	0.0976
		0.01	0.1993	0.5167	0.2429	0.2443	0.1844	0.2893	0.0345	0.265
		0.025	0.7955	0.9913	0.6929	0.8877	0.744	0.9016	0.0585	0.9117
		0.05	0.9998	1	0.9619	1	0.9985	0.9998	0.1264	1
		0.075	1	1	0.9961	1	1	1	0.2151	1
		0.09	1	1	0.9991	1	1	1	0.2574	1
		0.1	1	1	0.9995	1	1	1	0.2907	1
		0.3	1	1	1	1	1	1	0.6657	1
	0.5	1	1	1	1	1	1	0.7825	1	
	120	0	0.0513	0.0562	0.0701	0.0494	0.049	0.0522	0.0359	0.0486
		0.001	0.063	0.1063	0.0855	0.0654	0.0637	0.0733	0.032	0.0711
		0.005	0.3683	0.8151	0.412	0.4402	0.315	0.4841	0.0374	0.4664
		0.01	0.8959	0.999	0.8415	0.9504	0.8371	0.9592	0.0498	0.9625
		0.025	1	1	0.9994	1	1	1	0.119	1
		0.05	1	1	1	1	1	1	0.2835	1
		0.075	1	1	1	1	1	1	0.4125	1
		0.09	1	1	1	1	1	1	0.4908	1
		0.1	1	1	1	1	1	1	0.5197	1
		0.3	1	1	1	1	1	1	0.8152	1
	0.5	1	1	1	1	1	1	0.8845	1	

Donde:

α = Nivel de significancia.

n = Tamaño muestral.

tend = Nivel de tendencia.

DVEG Tipo I = Distribución de valor extremo generalizado tipo I.

Expon = Distribución exponencial.

ar = Término autorregresivo

Tabla 2.- Potencia de las pruebas para datos sin transformar con $ar = 0.01$ y $\alpha = 0.1$

			MANN-KENDALL A DATOS SIN TRANSFORMAR				PRUEBA DE PEARSON A DATOS SIN TRANSFORMAR			
α	n	tend	DVEG Tipo 1	Expon	Pareto	Normal	DVEG Tipo 1	Expon	Pareto	Normal
0.1	12	0	0.0853	0.0837	0.1017	0.0831	0.107	0.1041	0.0851	0.0973
		0.001	0.0856	0.0849	0.1074	0.0874	0.1022	0.1009	0.0822	0.1015
		0.005	0.0888	0.0935	0.0987	0.0846	0.0998	0.1022	0.0817	0.1101
		0.01	0.0903	0.0921	0.1051	0.0913	0.1033	0.1029	0.0836	0.1047
		0.025	0.0924	0.1194	0.1133	0.0982	0.1149	0.1169	0.082	0.1148
		0.05	0.1265	0.1978	0.1503	0.1318	0.1402	0.1804	0.0957	0.1547
		0.075	0.1659	0.3058	0.1846	0.1785	0.1859	0.2656	0.105	0.2135
		0.09	0.2041	0.3812	0.2161	0.236	0.2235	0.3322	0.1069	0.2641
		0.1	0.2243	0.43	0.2349	0.2562	0.2468	0.3808	0.1167	0.3094
		0.3	0.8161	0.9567	0.5641	0.9123	0.8381	0.9336	0.2796	0.9529
	0.5	0.9882	0.9985	0.7503	0.9994	0.988	0.9953	0.4177	1	
	60	0	0.1006	0.107	0.1131	0.1002	0.1068	0.1084	0.0955	0.1034
		0.001	0.1061	0.1172	0.1218	0.1013	0.103	0.1102	0.0932	0.1063
		0.005	0.1591	0.2808	0.1891	0.1672	0.15	0.1839	0.095	0.1764
		0.01	0.3165	0.642	0.3481	0.3578	0.288	0.3974	0.097	0.3845
		0.025	0.8762	0.9964	0.7908	0.9386	0.8286	0.9339	0.1246	0.9543
		0.05	0.9999	1	0.9808	1	0.9995	0.9999	0.2048	1
		0.075	1	1	0.9981	1	1	1	0.3016	1
		0.09	1	1	0.9993	1	1	1	0.3396	1
		0.1	1	1	1	1	1	1	0.3863	1
		0.3	1	1	1	1	1	1	0.7131	1
	0.5	1	1	1	1	1	1	0.8226	1	
	120	0	0.0975	0.1034	0.1062	0.1043	0.1059	0.1045	0.0888	0.0974
		0.001	0.1216	0.167	0.1408	0.1261	0.1189	0.1252	0.0935	0.1316
		0.005	0.4896	0.891	0.5329	0.5772	0.4384	0.6134	0.0898	0.5972
		0.01	0.9441	0.9998	0.8943	0.9778	0.8948	0.9768	0.1108	0.9832
		0.025	1	1	0.9999	1	1	1	0.1971	1
		0.05	1	1	1	1	1	1	0.3643	1
		0.075	1	1	1	1	1	1	0.5069	1
		0.09	1	1	1	1	1	1	0.5591	1
0.1		1	1	1	1	1	1	0.5975	1	
0.3		1	1	1	1	1	1	0.8386	1	
0.5	1	1	1	1	1	1	0.902	1		

Donde:

α = Nivel de significancia.

n = Tamaño muestral.

tend = Nivel de tendencia.

DVEG Tipo I = Distribución de valor extremo generalizado tipo I.

Expon = Distribución exponencial.

ar = Término autorregresivo

Tabla 3. Potencia de las pruebas para datos sin transformar con $ar = 0.01$ y $\alpha = 0.01$

			MANN-KENDALL A DATOS SIN TRANSFORMAR				PRUEBA DE PEARSON A DATOS SIN TRANSFORMAR			
α	n	tend	DVEG Tipo 1	Expon	Pareto	Normal	DVEG Tipo 1	Expon	Pareto	Normal
0.01	12	0	0.0036	0.0051	0.0069	0.0053	0.0106	0.0071	0.0031	0.0102
		0.001	0.0041	0.0052	0.0077	0.0057	0.0097	0.0059	0.003	0.0104
		0.005	0.0054	0.0072	0.0073	0.0068	0.0085	0.0091	0.0039	0.0097
		0.01	0.0065	0.0061	0.0074	0.0047	0.0109	0.009	0.0032	0.0108
		0.025	0.0072	0.0092	0.0109	0.0079	0.0113	0.012	0.004	0.0133
		0.05	0.0097	0.0245	0.0148	0.0103	0.0172	0.0247	0.0047	0.0217
		0.075	0.0169	0.05	0.0233	0.0222	0.0262	0.0487	0.0074	0.0375
		0.09	0.0233	0.0732	0.0273	0.0274	0.0383	0.0769	0.0099	0.0502
		0.1	0.0242	0.0926	0.0352	0.035	0.0399	0.094	0.012	0.0614
		0.3	0.3436	0.6736	0.1926	0.4767	0.4681	0.7171	0.0824	0.6679
	0.5	0.7879	0.951	0.3775	0.931	0.887	0.9541	0.1999	0.9913	
	60	0	0.0095	0.0102	0.0122	0.01	0.0099	0.0091	0.0023	0.0108
		0.001	0.0098	0.0108	0.0153	0.0106	0.0104	0.0106	0.0038	0.0114
		0.005	0.0233	0.0696	0.0323	0.0271	0.0199	0.0312	0.0036	0.03
		0.01	0.0758	0.2796	0.0986	0.0935	0.0656	0.1174	0.005	0.1093
		0.025	0.5687	0.953	0.4601	0.702	0.5015	0.763	0.0104	0.7474
		0.05	0.9967	1	0.8943	0.9999	0.989	0.9992	0.044	0.9999
		0.075	1	1	0.9818	1	1	1	0.1077	1
		0.09	1	1	0.9944	1	1	1	0.1472	1
		0.1	1	1	0.9972	1	1	1	0.1689	1
		0.3	1	1	1	1	1	1	0.5638	1
	0.5	1	1	1	1	1	1	0.7179	1	
	120	0	0.0122	0.0106	0.0105	0.0073	0.0104	0.0109	0.0037	0.0097
		0.001	0.014	0.0286	0.0197	0.0174	0.0142	0.0142	0.0034	0.0188
		0.005	0.1643	0.6069	0.2014	0.2137	0.1344	0.2555	0.005	0.2492
		0.01	0.7295	0.9944	0.6645	0.8545	0.6407	0.8688	0.0062	0.8762
		0.025	1	1	0.9966	1	1	1	0.0411	1
		0.05	1	1	1	1	1	1	0.157	1
		0.075	1	1	1	1	1	1	0.2961	1
		0.09	1	1	1	1	1	1	0.3627	1
		0.1	1	1	1	1	1	1	0.407	1
		0.3	1	1	1	1	1	1	0.7533	1
	0.5	1	1	1	1	1	1	0.8559	1	

Donde:

α = Nivel de significancia.

n = Tamaño muestral.

tend = Nivel de tendencia.

DVEG Tipo I = Distribución de valor extremo generalizado tipo I.

Expon = Distribución exponencial.

ar = Término autorregresivo

Tabla 4.- Potencia de las pruebas para datos transformados por sus rangos con $ar = 0.01$ y $\alpha = 0.05$

		MANN-KENDALL A DATOS TRANSFORMADOS				PRUEBA DE PEARSON A DATOS TRANSFORMADOS				
α	n	tend	DVEG Tipo 1	Expon	Pareto	Normal	DVEG Tipo 1	Expon	Pareto	Normal
0.05	12	0	0.0464	0.0429	0.0569	0.0468	0.0535	0.0575	0.066	0.0567
		0.001	0.0417	0.0487	0.0576	0.0522	0.0575	0.0522	0.0677	0.0572
		0.005	0.0443	0.0434	0.0568	0.043	0.0524	0.0557	0.0639	0.0515
		0.01	0.047	0.055	0.0551	0.0482	0.056	0.061	0.0671	0.0562
		0.025	0.0495	0.068	0.0625	0.0514	0.063	0.0754	0.0755	0.0625
		0.05	0.07	0.1226	0.0876	0.0798	0.0829	0.1373	0.0963	0.0876
		0.075	0.0963	0.2043	0.1169	0.1124	0.114	0.2187	0.1275	0.1239
		0.09	0.1227	0.2659	0.1299	0.1409	0.1312	0.2842	0.1484	0.1627
		0.1	0.1348	0.3136	0.151	0.16	0.1602	0.3243	0.161	0.186
		0.3	0.7017	0.9138	0.4553	0.8362	0.7329	0.8987	0.4309	0.8709
	0.5	0.9714	0.9956	0.6606	0.9979	0.972	0.9917	0.5953	0.998	
	60	0	0.0504	0.0522	0.0655	0.0516	0.0541	0.0566	0.0638	0.0523
		0.001	0.0566	0.0616	0.0686	0.0539	0.0527	0.0611	0.0689	0.0518
		0.005	0.0896	0.1892	0.1117	0.0945	0.0905	0.1911	0.1146	0.106
		0.01	0.2056	0.5178	0.2329	0.2403	0.2114	0.5096	0.238	0.257
		0.025	0.792	0.9906	0.6913	0.8862	0.7941	0.9844	0.6595	0.8932
		0.05	0.9995	1	0.9659	1	0.9995	1	0.9388	1
		0.075	1	1	0.9952	1	1	1	0.9882	1
		0.09	1	1	0.9992	1	1	1	0.9956	1
		0.1	1	1	0.9997	1	1	1	0.9972	1
		0.3	1	1	1	1	1	1	1	1
	0.5	1	1	1	1	1	1	1	1	
	120	0	0.0483	0.0505	0.0613	0.0516	0.0543	0.0558	0.0662	0.0478
		0.001	0.0636	0.1064	0.0822	0.0669	0.0662	0.102	0.086	0.0702
		0.005	0.3695	0.8121	0.4198	0.4471	0.3673	0.8017	0.4048	0.4498
		0.01	0.8905	0.9997	0.8371	0.9522	0.8949	0.998	0.8051	0.9516
		0.025	1	1	0.9993	1	1	1	0.9981	1
		0.05	1	1	1	1	1	1	1	1
		0.075	1	1	1	1	1	1	1	1
		0.09	1	1	1	1	1	1	1	1
0.1		1	1	1	1	1	1	1	1	
0.3		1	1	1	1	1	1	1	1	
0.5	1	1	1	1	1	1	1	1		

Donde:

α = Nivel de significancia.

n = Tamaño muestral.

tend = Nivel de tendencia.

DVEG Tipo I = Distribución de valor extremo generalizado tipo I.

Expon = Distribución exponencial.

ar = Término autorregresivo

Tabla 5. Potencia de las pruebas para datos transformados por sus rangos con $ar = 0.01$ y $\alpha = 0.1$

			MANN-KENDALL A DATOS TRANSFORMADOS				PRUEBA DE PEARSON A DATOS TRANSFORMADOS			
α	n	tend	DVEG Tipo 1	Expon	Pareto	Normal	DVEG Tipo 1	Expon	Pareto	Normal
0.1	12	0	0.0971	0.0886	0.0968	0.0842	0.1032	0.1015	0.1092	0.1039
		0.001	0.0833	0.0887	0.1004	0.092	0.099	0.0981	0.1196	0.0998
		0.005	0.089	0.0931	0.1034	0.0863	0.1022	0.1091	0.1247	0.1012
		0.01	0.0902	0.0976	0.1122	0.0877	0.1022	0.1041	0.1226	0.1032
		0.025	0.0956	0.1188	0.116	0.0969	0.1102	0.1359	0.1267	0.1165
		0.05	0.1232	0.1972	0.1401	0.1303	0.1363	0.2231	0.1646	0.1501
		0.075	0.1685	0.3145	0.1823	0.1861	0.1822	0.3255	0.1965	0.2021
		0.09	0.1973	0.3781	0.2189	0.2247	0.2226	0.3978	0.2365	0.2625
		0.1	0.2185	0.4305	0.2316	0.2538	0.2479	0.4382	0.2537	0.2906
		0.3	0.8128	0.9534	0.5648	0.9154	0.829	0.943	0.5376	0.9342
	0.5	0.9856	0.9993	0.7454	0.9993	0.9829	0.9969	0.6912	0.9998	
	60	0	0.1019	0.1051	0.1186	0.0963	0.0989	0.1003	0.123	0.104
		0.001	0.1086	0.1168	0.1194	0.1043	0.1063	0.1172	0.1236	0.1059
		0.005	0.153	0.2834	0.1977	0.1772	0.1651	0.2955	0.1894	0.1665
		0.01	0.3103	0.6423	0.3495	0.3639	0.316	0.6235	0.3369	0.3563
		0.025	0.875	0.9961	0.7858	0.9439	0.8768	0.9944	0.7549	0.9427
		0.05	1	1	0.9814	1	0.9996	1	0.9638	1
		0.075	1	1	0.9985	1	1	1	0.9942	1
		0.09	1	1	0.9993	1	1	1	0.9979	1
		0.1	1	1	0.9996	1	1	1	0.9989	1
		0.3	1	1	1	1	1	1	1	1
	0.5	1	1	1	1	1	1	1	1	
	120	0	0.1021	0.101	0.125	0.0998	0.1031	0.1056	0.1213	0.1048
		0.001	0.1166	0.1747	0.1494	0.1252	0.1229	0.1706	0.148	0.1274
		0.005	0.4935	0.8838	0.5312	0.5718	0.4886	0.8754	0.5201	0.5764
		0.01	0.9434	0.9998	0.9006	0.9747	0.9422	0.9998	0.8806	0.9786
		0.025	1	1	1	1	1	1	0.999	1
		0.05	1	1	1	1	1	1	1	1
		0.075	1	1	1	1	1	1	1	1
		0.09	1	1	1	1	1	1	1	1
0.1		1	1	1	1	1	1	1	1	
0.3		1	1	1	1	1	1	1	1	
0.5	1	1	1	1	1	1	1	1		

Donde:

α = Nivel de significancia.

n = Tamaño muestral.

tend = Nivel de tendencia.

DVEG Tipo I = Distribución de valor extremo generalizado tipo I.

Expon = Distribución exponencial.

ar = Término autorregresivo

Tabla 6.- Potencia de las pruebas para datos transformados por sus rangos con $ar = 0.01$ y $\alpha = 0.01$

			MANN-KENDALL A DATOS TRANSFORMADOS				PRUEBA DE PEARSON A DATOS TRANSFORMADOS			
α	n	tend	DVEG Tipo 1	Expon	Pareto	Normal	DVEG Tipo 1	Expon	Pareto	Normal
0.01	12	0	0.0059	0.0053	0.0086	0.0056	0.0151	0.0135	0.0166	0.0121
		0.001	0.0055	0.0052	0.0071	0.0061	0.0111	0.0142	0.0149	0.0124
		0.005	0.0053	0.0053	0.0072	0.0067	0.0132	0.0126	0.0167	0.0124
		0.01	0.0069	0.0054	0.0076	0.0058	0.0119	0.0147	0.0146	0.0135
		0.025	0.0064	0.011	0.008	0.0081	0.0145	0.0194	0.0179	0.0166
		0.05	0.009	0.0262	0.0132	0.0107	0.0211	0.0412	0.029	0.0251
		0.075	0.0154	0.0518	0.0227	0.0198	0.0321	0.0797	0.0378	0.0381
		0.09	0.0213	0.0661	0.0328	0.0264	0.0449	0.1197	0.0521	0.0499
		0.1	0.023	0.0912	0.0347	0.0339	0.0517	0.1419	0.0605	0.0688
		0.3	0.341	0.673	0.1931	0.4692	0.4755	0.7385	0.2436	0.6313
	0.5	0.7924	0.9497	0.3707	0.937	0.8762	0.9604	0.4087	0.9783	
	60	0	0.01	0.009	0.0124	0.01	0.0092	0.012	0.0167	0.0113
		0.001	0.0097	0.0133	0.0168	0.0103	0.0127	0.0139	0.0162	0.0135
		0.005	0.024	0.0619	0.0347	0.0267	0.0244	0.066	0.0371	0.0296
		0.01	0.0701	0.271	0.0942	0.0905	0.0785	0.2727	0.0991	0.0959
		0.025	0.5704	0.9559	0.4725	0.7223	0.5858	0.9416	0.4331	0.7314
		0.05	0.9962	1	0.8895	0.9998	0.9967	1	0.8345	1
		0.075	1	1	0.9807	1	1	1	0.9538	1
		0.09	1	1	0.9939	1	1	1	0.9797	1
		0.1	1	1	0.9968	1	1	1	0.986	1
		0.3	1	1	1	1	1	1	1	1
	0.5	1	1	1	1	1	1	1	1	
	120	0	0.0088	0.0085	0.016	0.0107	0.0137	0.0104	0.0156	0.011
		0.001	0.0131	0.0238	0.022	0.0131	0.0164	0.0285	0.0223	0.0173
		0.005	0.1721	0.598	0.2062	0.2206	0.1723	0.5918	0.1993	0.2232
		0.01	0.7266	0.9936	0.6683	0.8543	0.736	0.9906	0.6191	0.8475
		0.025	1	1	0.9959	1	1	1	0.9917	1
		0.05	1	1	1	1	1	1	0.9999	1
		0.075	1	1	1	1	1	1	1	1
		0.09	1	1	1	1	1	1	1	1
0.1		1	1	1	1	1	1	1	1	
0.3		1	1	1	1	1	1	1	1	
0.5	1	1	1	1	1	1	1	1		

Donde:

α = Nivel de significancia.

n = Tamaño muestral.

tend = Nivel de tendencia.

DVEG Tipo I = Distribución de valor extremo generalizado tipo I.

Expon = Distribución exponencial.

ar = Término autorregresivo

Figura 1. Gráfica de la potencia para la prueba de Mann-Kendall para datos sin transformar con $\rho = 0.5$, $n = 12$ y $\alpha = 0.05$.

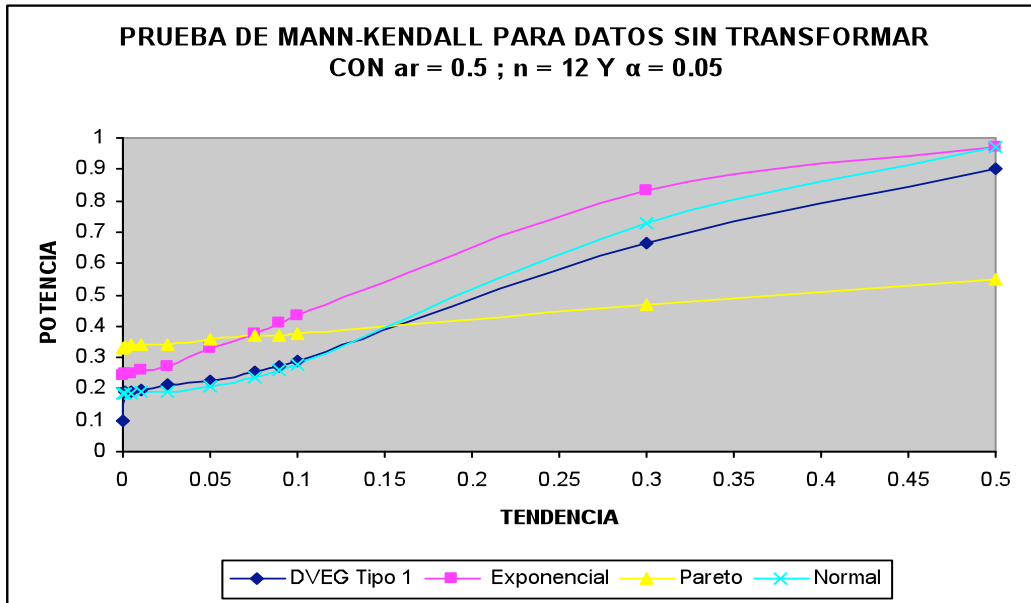


Figura 2. Gráfica de la potencia para la prueba del coeficiente de correlación de Pearson para datos sin transformar con $\rho = 0.5$, $n = 12$ y $\alpha = 0.05$

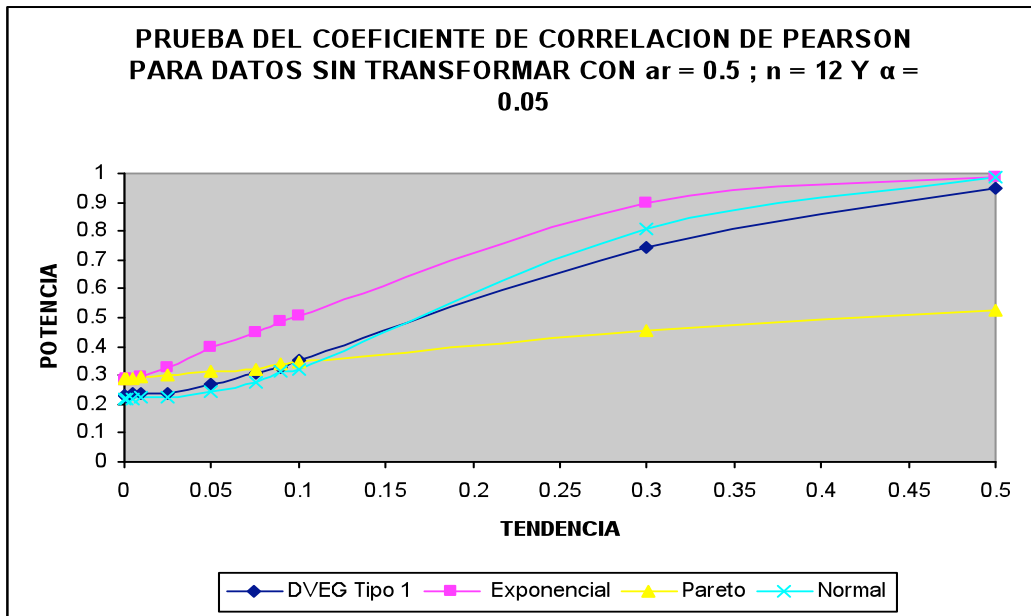


Figura 3.- Gráfica de potencia de la prueba de Mann-Kendall a datos transformados por sus rangos con $\alpha = 0.5$, $n = 12$ y $\alpha = 0.01$

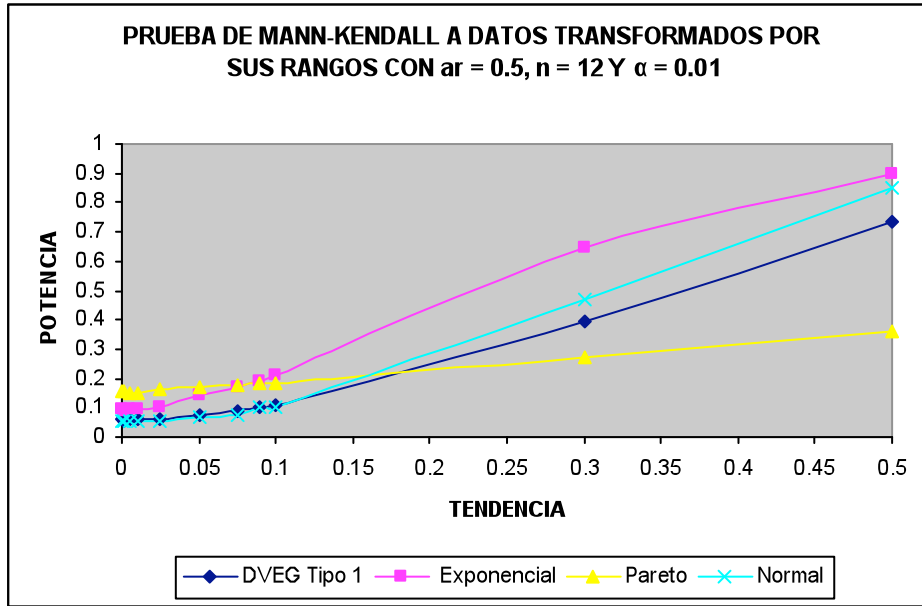
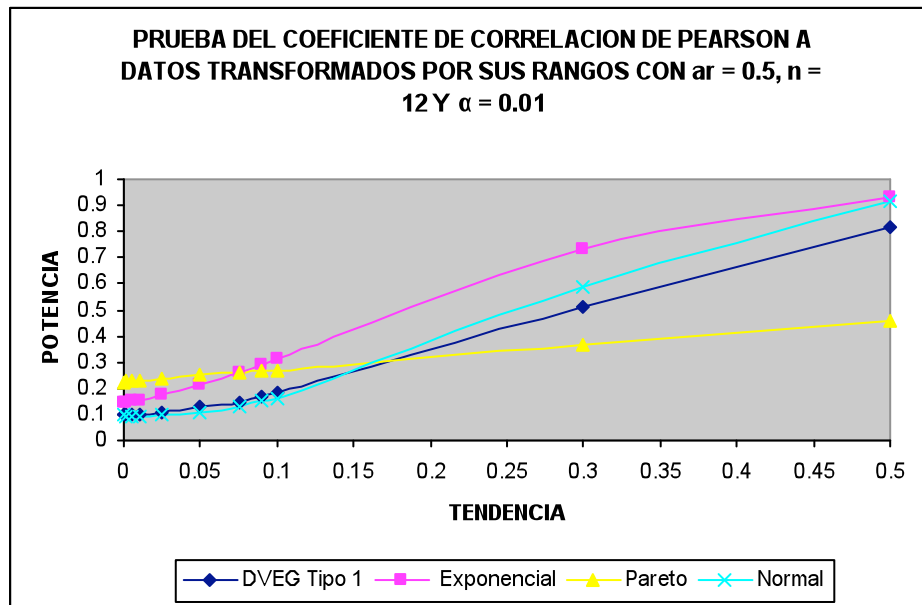


Figura 4.- Gráfica de potencia de la prueba del coeficiente de correlación de Pearson a datos transformados por sus rangos con $\alpha = 0.5$, $n = 12$ y $\alpha = 0.01$



6. APLICACIÓN DE LAS PRUEBAS MOSTRADAS EN EL ESTUDIO

Se obtuvieron datos de la Red Automática de Monitoreo Atmosférico (RAMA) del Distrito Federal para la estación Pedregal, que corresponden a los máximos diarios de niveles de ozono, Monóxido de carbono, Dióxido de azufre y de Oxido de nitrógeno medidos en partes por millón de Julio de 1995 a abril de 2007. Gráficamente los datos para el Monóxido de carbono están representados de la siguiente manera:

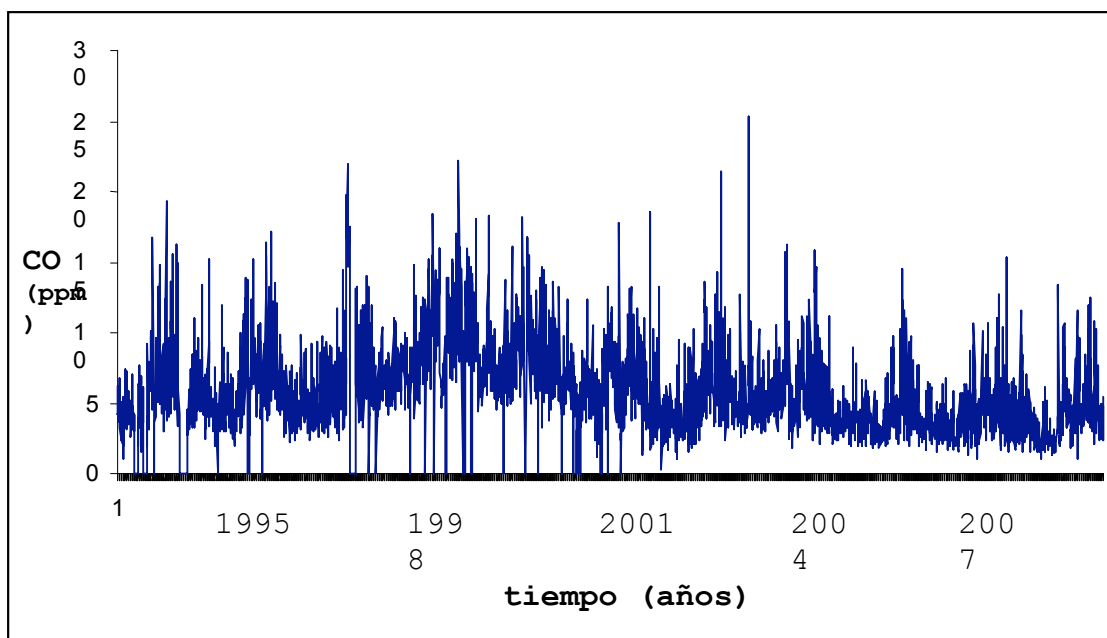


Figura 5. Serie de tiempo para los niveles máximos de monóxido de carbono en la estación Pedregal de Julio de 1987 a Abril de 2007.

Los otros contaminantes siguen una forma gráfica similar a la del monóxido de carbono.

Para los cuatro contaminantes se rechaza la hipótesis nula con p-valor dado en la tabla 1 y se concluye que existe tendencia. Analizando la correlación para las dos pruebas de Pearson y la tau para la prueba de Mann-Kendall se observa que para el caso del Ozono, del Monóxido de carbono y del Dióxido de azufre la tendencia es decreciente y únicamente para el Oxido de nitrógeno la tendencia es decreciente.

Los resultados se presentan en la siguiente tabla:

Prueba	estadística	Ozono	CO	SO ₂	NO ₂
Mann-Kendall	tau	-0.081	-0.192	-0.364	0.0985
	p-valor	4.93E-14	2.22E-16	2.22E-16	2.22E-16
Pearson sin transformar	correlación	-0.09473	-0.22751	-0.4752	0.135721
	p-valor	2.91E-09	2.20E-16	2.20E-16	2.20E-16
Pearson transformados	correlación	-0.116	-0.27799	-0.50824	0.150210
	p-valor	2.20E-16	2.20E-16	2.20E-16	2.20E-16

Tabla 8. Estadísticas calculadas para las pruebas aplicadas.

Donde:

CO = Monóxido de carbono

SO₂ = Dióxido de azufre

NO₂ = Oxido de Nitrógeno

tau = Coeficiente de correlación de Kendall apropiado para estudiar la relación entre variables ordinales.

p-valor = Nivel de significancia empírico, obtenido a partir del valor del estadístico de prueba.

7. CONCLUSIONES

Para datos de series de tiempo, es aconsejable utilizar la prueba de Mann-Kendall para la detección de tendencia, pues es menos afectada por las observaciones extremas y depende directamente del tipo de distribución bajo estudio. Sin embargo no se debe menospreciar la prueba que utiliza el coeficiente de correlación de Pearson, que es más sencillo de calcular y en cualquier paquete estadístico se puede encontrar, sobretodo que cuando se aplica después de transformar los datos con sus rangos se elimina el problema de empates y en caso de que éstos no existan se obtiene la prueba rho de Spearman, la cual genera resultados muy similares a la prueba de Mann-Kendall.

8.- REFERENCIAS

Beveridge, S. y Nelson, C.R. (1981); "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with particular attention to Measurement of the Business Cycle". *Journal of Monetary Economics*, 7: 151-174.

Cohen, J. 1977. *Statistical power analysis for the behavioral sciences*, rev. ed. Academic Press. Sew York.

Coles, S. G. (2001). *An Introduction to Statistical Modelling of Extreme Values*. London: Springer.

Conover, W. J. and Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *American Statistician* 35 (3), 124-133.

Conover; W. J. Ronald and Iman, R. L., *Analysis of Covariance Using the Rank Transformation Biometrics*, Vol. 38, No. 3, Special Issue: Analysis of Covariance. (Sep., 1982), pp. 715-724.

Conover; W. J., Iman, Ronald L. [Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics]: Rejoinder *The American Statistician*, Vol. 35, No. 3. (Aug., 1981), pp. 132-133.

Dickey, D.A. y Fuller, W.A. (1979); "Distribution of the Estimators for Autoregressive Time Series with a Unit Root". *Journal of the American Statistical Association*, 74: 427-431.

Eryl A. C. and Shirley C., *A Distribution-Free Method for Analysis of Covariance Based on Ranked Data Applied Statistics*, Vol. 30, No. 2. (1981), pp. 158-162.

Franzini, L. ; A. C. Harvey, A. C., *Testing for Deterministic Trend and Seasonal Components in Time Series Models. Biometrika*, Vol. 70, No. 3. (Dec., 1983), pp. 673-682.

Gilbert, R. O. (1987). *Statistical Methods for Environmental Pollution Monitoring*. John Wiley & Sons, Inc.

Gottfried E. Noether. [Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics]: Comment. *The American Statistician*, Vol. 35, No. 3. (Aug., 1981), pp. 129-130.

Hamed, K. H. (2007). Trend detection in hydrologic data: The Mann–Kendall trend test under the scaling hypothesis. *Journal of Hydrology*. 349, 350-363

Hodrick, R.J. y Prescott, E.C. (1997); “Postwar US Business Cycles: An Empirical Investigation”. *Journal of Money, Credit and Banking*, 29: 1-16.

Iman, R. L. (1974). A power study of a rank transform for the two way classification model when interaction may be present. *Canadian Journal of Statistics* 2, 227-239.

Iman, R. L. and Conover, W. J. (1979). The use of the rank transform in regression. *Technometrics* 21, 49-59.

Iman, Ronald L., and Conover, W.J. (1976), "A Comparison of Several Rank Tests for the Two-Way Layout," Technical Report SAND76-0631, Sandia Laboratories, Albuquerque, New Mexico.

John T. Rotenberry; John A. Wiens. Statistical Power Analysis and Community-Wide Patterns. *The American Naturalist*, Vol. 125, No. 1. (Jan., 1985), pp. 164-168.

Jeff S. Hatfield; William R. Gould IV; Brett A. Hoover; Mark R. Fuller; Edward L. Lindquist. Detecting Trends in Raptor Counts: Power and Type I Error Rates of Various Statistical Tests. *Wildlife Society Bulletin*, Vol. 24, No. 3, Predators. (Autumn, 1996), pp. 505-515.

Jeremy M. G. Taylor. Kendall's and Spearman's Correlation Coefficients in the Presence of a Blocking Variable. *Biometrics*, Vol. 43, No. 2. (Jun., 1987), pp. 409-416.

Klein, L.R. y Kosobud, R.F. (1961); "Some Econometrics of Growth: Great Ratios of Economics". *Quarterly Journal of Economics*, 75: 173-198.

Krämer, Walter. Note on Estimating Linear Trend when Residuals are Autocorrelated. *Econometrica*, Vol. 50, No. 4. (Jul., 1982), pp. 1065-1067.

Lehmann. E.L. Significance Level and Power *The Annals of Mathematical Statistics*, Vol. 29, No. 4. (Dec., 1958), pp. 1167-1176.

Libiseller, Claudia (2004). A Program for the Computation of Multivariate and Partial Mann-Kendall Test. <http://www.mai.liu.se/~cllib/welcome/PMKtest.html>

Madan Lal Puri; Pranab Kumar Sen. Analysis of Covariance Based on General Rank Scores. *The Annals of Mathematical Statistics*, Vol. 40, No. 2. (Apr., 1969), pp. 610-618.

Molinero, Luis M. 2004. Análisis de Series Temporales. Sociedad Española de Hipertensión. pp 02-06.

Mood, A.M., F.A. Graybill, and D.C. Boes. 1974. Introduction to the theory of statistics. McGraw-Hill, Neu York.

Nelson, C.R. y Kang, H. (1981); "Spurious Periodicity in Inappropriately Detrended Time Series". *Econometrica*, 49: 741-751.

Nelson, C.R. y Plosser, C.I. (1982); "Trends and random walks in macroeconomic time series: some evidence and implications". *Journal of Monetary Economics*, 10: 139-162.

Onoz, Bihrat and Bayazit, Mehmet_cik., (2002). The Power of Statistical Tests for Trend Detection. Istanbul Technical University, Faculty of Civil Engineering, 80626 Maslak, _ Istanbul-TURKEY

Otten, A. Note on the Spearman Rank Correlation Coefficient. *Journal of the American Statistical Association*, Vol. 68, No. 343. (Sep., 1973), p. 585.

Poynting, J.H. (1884); "A Comparison of the Fluctuations in the Price of Wheat and in Cotton and Silk Imports into Great Britain". *Journal of the Royal Statistical Society*, 47: 34-74.

R Development Core Team (2007). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

R. D. Reiss and M. Thomas (2001). *Statistical analysis of extreme values: with applications to insurance, finance, hidrology and other fields*. 2. ed. Basel, Boston, Berlin. Birkhauser.

Terence C. Mills; N. F. R. Crafts. *Modelling Trends in Economic History*. *The Statistician*, Vol. 45, No. 2. (1996), pp. 153-159.

Tim Gerrodette. *A Power Analysis for Detecting Trends*. *Ecology*, Vol. 68, No. 5. (Oct., 1987), pp. 1364-1372.

Vahid, F. y Engle, R. (1993); "Common Trends and Common Cycles". *Journal of Applied Econometrics*, 8: 341-360.

William C. Tanke; Charles D. Bonham. *Use of Power Curves to Monitor Range Trend*. *Journal of Range Management*, Vol. 38, No. 5. (Sep., 1985), pp. 428-431.

9.- ANEXOS

Programa 1.- Obtiene la potencia de la prueba de Mann-Kendall transformando los datos con sus rangos.

```
gpd <- function(alpha,n,m,trend)
{
rechazos <- rep(0,m)
j <- 1
while(j<=m){
a1 <- arima.sim(list(order=c(1,0,0),ar=0.9), n, innov=(1/(1-(runif(n)))))
t1 <- 1:n
x1 <- 1 + (trend*t1) + a1
w1 <- rank(x1, na.last = TRUE, ties.method = c("average", "first", "random", "max", "min"))
z1 <- if((MannKendall(w1)$"s1")<=alpha){rechazos[j]<-1}
j=j+1
}
nr <- sum(rechazos)
tapr <- nr/m
tapr
}
gpd(alpha=0.01,n=120,m=10000,trend=0.005)
```

Programa 2.- Obtiene la potencia de la prueba con el coeficiente de correlación de Pearson para datos transformados por sus rangos.

```
tipo <- function(alpha,n,m,trend){
rechazos <- rep(0,m)
j <- 1
while(j<=m)
{
a1 <- arima.sim(list(order=c(1,0,0),ar=0.6), n, innov=(-log(-log(runif(n))))))
t1 <- 1:n
x1 <- 1 + (trend*t1) + a1
w1 <- rank(x1, na.last = TRUE, ties.method = c("average", "first", "random", "max", "min"))
z1 <- if((cor.test(w1, t1, alternative = "two.sided", method = "pearson", exact = NULL,
conf.level = 0.95)$"p.value")<=alpha){rechazos[j]<-1}
j=j+1
}
nr <- sum(rechazos)
tapr <- nr/m
tapr
}
tipo(alpha=0.01,n=12,m=10000,trend=0.075)
```

Programa 3.- Para obtener la potencia de la prueba de Mann-Kendall para datos sin transformar.

```
expon <- function(alpha,n,m,trend)
{
rechazos <- rep(0,m)
j <- 1
while(j<=m){
a1 <- arima.sim(list(order=c(1,0,0),ar=0.1), n, innov=(-log(1-(runif(n)))))
t1 <- 1:n
x1 <- 1 + (trend*t1) + a1
z1 <- if((MannKendall(x1)$"sl")<=alpha){rechazos[j]<-1}
j=j+1}
nr <- sum(rechazos)
tapr <- nr/m
tapr
}
expon (alpha=0.01,n=60,m=10000,trend=0.05)
```

Programa 4.- Para obtener la potencia de la prueba con el coeficiente de correlación de Pearson para datos sin transformar.

```
distnorm <- function(alpha,n,m,trend){
rechazos <- rep(0,m)
j <- 1
while(j<=m){
a1 <- arima.sim(list(order=c(1,0,0),ar=0.5), n, rand.gen=rnorm)
t1 <- 1:n
x1 <- 1 + (trend*t1) + a1
z1 <- if((cor.test(x1, t1, alternative = "two.sided", method = "pearson", exact = NULL, conf.level
= 0.95)$"p.value")<=alpha){rechazos[j]<-1}
j=j+1}
nr <- sum(rechazos)
tapr <- nr/m
tapr
}
distnorm(alpha=0.01,n=12,m=10000,trend=0.01)
```