

UNIVERSIDAD NACIONAL AUTONOMA DE NICARAGUA

UNAN-LEÓN



“A la libertad... Por la Universidad”

FACULTAD DE CIENCIAS Y TECNOLOGÍA

DEPARTAMENTO DE MATEMATICA Y ESTADISTICA

Trabajo Monográfico para optar al título de: INGENIERA ESTADÍSTICA

TEMA:

Estudio longitudinal de las defunciones ocurridas en el departamento de León comprendidas en el periodo 2005-2010.

Integrantes:

- **Maryuri Lissette Cáceres López.**
- **Valeska Marcela Castellón Guido.**
- **Oneyda Maribel Chévez Solís.**

Tutor: W. Milton Carvajal Herradora

DEDICATORIA

Dedico este trabajo primeramente a Dios por darme la vida y por guiar mi camino, a mis PADRES quienes me impulsaron hasta el final para hacer posible la culminación de este trabajo satisfactoriamente con su apoyo incondicional y también a mis maestros que compartieron sus conocimientos y a mi Tutor por su paciencia y su ayuda a todos ellos gracias.

Maryuri Lissette Cáceres López.

Dedico esta tesis A. DIOS, a Santo Tomás de Aquino, patrono de los estudiantes y a la Virgen María, quienes inspiraron mi espíritu para la conclusión de esta tesis. A mis padres quienes me dieron vida, educación, apoyo y consejos, A mis compañeras de estudio, a mis maestros, quienes sin su ayuda nunca hubiera podido hacer esta tesis. A todos ellos se los agradezco desde el fondo de mi alma. Para todos ellos hago esta dedicatoria.

Valeska Marcela Castellón Guido.

Dedico esta tesis a DIOS, a mi madre, maestros, amigas por darme su apoyo y paciencia a superar cada uno de los obstáculos que se me presentaron gracias por estar a mi lado todo este tiempo brindándome su ayuda y a todos aquellos que no creyeron en mí, a aquellos que nunca esperaban que lograra terminar la carrera.

Oneyda Maribel Chévez Solís.

AGRADECIMIENTOS

Gracias DIOS, por haberme puesto en mi caminar a maestros y amigas que me dieron su apoyo y me ayudaron a superar cada uno de los obstáculos que se me presentaron en los años de estudios de mi carrera universitaria y un agradecimiento especial, a mis PADRES que con su cariño, su amor hicieron todo en la vida para que yo pudiera lograr mis sueños, por motivarme y darme la mano cuando sentía que el camino se terminaba, a ustedes mi agradecimiento.

Maryuri Lissette Cáceres López.

Agradezco a dios por ser tan maravilloso que me dio fuerza y fe para creer lo que me parecía imposible terminar. A mi familia por ayudarme con mi hija mientras yo realizaba investigaciones y por estar a mi lado en cada momento de mi vida. A mi esposo por su ayuda en impulsarme a terminar este proyecto. A todos los profesores por su apoyo y por compartir conmigo sus conocimientos en mi carrera, en especial al tutor por su apoyo constante. También expresar mis agradecimientos a mis compañeras que siempre me dieron su apoyo incondicional en todo el transcurso de mi carrera.

Valeska Marcela Castellón Guido.

Agradezco a DIOS, a mis maestros que en este andar por la vida, influyeron con sus lecciones y experiencias en formarme como una persona de bien y preparada para los retos que pone la vida, a todos y cada uno de ellos les dedico cada una de estas páginas de mi tesis.

Oneyda Maribel Chévez Solís.

INDICE

INTRODUCCION.....	1-2
JUSTIFICACION.....	3
HIPOTESIS.....	4
PLANTAMIENTO DEL PROBLEMA.....	5
OBJETIVOS.....	6
Objetivo General.....	6
Objetivo Especifico.....	6
MARCO TEORICO.....	7
Insuficiencia Renal no Especificada.....	7
Diabetes Mellitus.....	7
Infarto Agudo del Miocardio.....	8
Accidentes Vasculares Encefalico Agudo.....	8-9
Series de Tiempo.....	10
Elementos Estadisticos en el Analisis de una Serie de Tiempo.....	10
Tipos de Series Temporales.....	11
Series Temporales de un Nivel.....	11
Series Temporales de Flujo.....	11
Procesos Estacionarios.....	11
Funcion de Medias.....	12
Funcion de Varianza.....	12
Función de Autocorrelacion.....	12
Función de Autocorrelacion Simple de Orden k,p	12
Operador de Diferencia(diff).....	13
Operador de Retardos L (laq).....	14
Procesos Integrados.....	14
Proceso de Ruido Blanco.....	14
Modelo Clásico de Series Temporales.....	15
Tendencia.....	15
Factor Ciclico.....	15
El Efecto Estacional.....	15

UNIVERSIDAD NACIONAL AUTONOMA DE NICARAGUA UNAN-LEON

Residuo.....	15
Las Medias Moviles.....	16
Alisado Exponencial.....	16
Diferenciacion.....	17
Enfoque Moderno de Series Temporales	
"Metodología Box-Jenkins".....	17
Modelos de Series Temporales.....	18
Modelos Autorregresivos.....	18
Característica del modelo AR(1).....	19
Característica del modelo AR(2).....	20
Modelos AR(p).....	21
Media nula y Varianza constante.....	22
Modelos de Medias Moviles.....	23
Modelo Mixto Autor regresivo - Medias Moviles ARMA(p,q).....	24
Modelos ARIMA.....	25
Diagrama de Flujo de la Metodologia de Box-Jenkins.....	26
MATERIAL Y METODO.....	27
RESULTADOS.....	28-37
CONCLUSIONES.....	38
RECOMENDACIONES.....	39
BIBLIOGRAFIA.....	40
ANEXOS.....	41

INTRODUCCION

Este trabajo tiene como propósito de presentar con las técnicas estadística descriptivas, inferenciales y de series temporales, las principales causa de muerte del departamento de León analizando los registros de los últimos 6 años del departamento, así como las variables de interés que estudiaremos y daremos a conocer si acaso la edad, el sexo, la procedencia pudieran ser factores que intervengan en las defunciones de las personas.

Realizamos un análisis longitudinal de serie de tiempo el cual tiene un patrón de comportamiento, que es para poder prever su evolución en el futuro a corto plazo, suponiendo por supuesto que las condiciones no variarían significativamente. Al estimar un modelo de pronósticos, permitiría realizar, en base al análisis de este tipo sobre nuevos datos, la estabilidad del mismo y que apoyen el desarrollo de nuevos estudios sobre este tema.

Las predicciones de valores futuros de una serie de datos tomados del pasado es un problema importante en muchas áreas, se utiliza para la toma de decisiones en momentos de incertidumbre con el propósito de obtener mejores resultados.

Entre los métodos de previsión en el análisis de serie de tiempo se encuentran los Métodos de descomposición (cálculo de tendencia y de la estación o Época periódica anual), Método de alisado exponencial y Modelos ARIMA.

Hoy en día el campo de la salud utiliza frecuentemente la técnica estadística Serie de Tiempo con el análisis de tendencia o variación estacional para diferenciar y hacer predicciones. También utilizan otras técnicas sencillas, tasas comparativas e incidencia anual para el estudio de las defunciones.

Estos análisis lo hacen los médicos investigadores realizando comparaciones con series de datos anteriores y basados en su experiencia predicen para tiempo futuros; de esta manera toman decisiones ante sucesos o fenómenos que se les presenten, teniendo en cuenta como en cualquier investigación científica, un margen de error provocado por factores incontrolables.

Si bien es cierto que ya existen estudios de la mortalidad, es necesario someter la información a análisis estadísticos exhaustivos con la finalidad que se percibe de dar una visión más amplia de la situación en la medida de lo posible, garantizar un modelo que permita simular el fenómeno en el tiempo lo cual podría conducir a la elaboración de pronósticos.

Se decidió analizar el comportamiento mensual y anual de las principales causas de mortalidad que se dieron en el periodo comprendido del 2005-2010, debido a la disponibilidad de datos en esos periodos.

JUSTIFICACIÓN

Con este estudio se pretende proporcionar información que esté dirigida a señalar las causas más frecuentes de la mortalidad que se dan en el departamento de León y sus municipios según datos brindados por el SILAIS-León, para luego realizar un estudio longitudinal que permita identificar los factores de las diferentes causas de muerte y la modelización de la misma.

El área de la salud, es uno de los sectores en los que se recopila mucha información y se necesitan realizar análisis no solamente descriptivos, como los que se realizan en la actualidad, debido a que no se puede tomar decisiones al azar, sin tomar en cuenta el comportamiento de los datos registrados, así un modelo ideal sería el que representara mejor a la serie.

Por lo antes expuesto, consideramos que el presente trabajo es relevante como apoyo, ya que en él se aplican técnicas estadísticas avanzadas.

Hipótesis

No existe relación entre el sexo del difunto y la causa de muerte, además la tasa de muerte es estable a través de los años, impredecible e indistinta según los meses.

Planteamiento del problema

Este trabajo se inicia con el objetivo de indagar las causas de mayor incidencia de muertes para luego estimar un modelo que sea lo más adecuado posible para representar el comportamiento de los datos, esta necesidad nos lleva a plantear la interrogante:

¿Será posible obtener un modelo adecuado que nos ayude a conocer cómo será el comportamiento de las defunciones del departamento de León, en un periodo de años posteriores a la información obtenida?

OBJETIVOS

Objetivo General

Describir el comportamiento de las defunciones en el departamento de León para el período 2005-2010, determinando las principales causas de muerte en dicho periodo.

Objetivos específicos

- Realizar análisis descriptivo e inferencial del comportamiento de las defunciones ocurridas en el departamento de León.

- Construir un modelo de predicción de Series de tiempo para las causas de muerte de mayor incidencia.

Marco Teórico

Insuficiencia renal crónica

Síntomas

- Sangrado en la orina (hematuria) → Orina espumosa
- Dolor abdominal → Sangre en vomito o heces.

Como prevenir la IRC

1. Dieta equilibrada y sin sal.
2. Ejercicio físico regular
3. Abstención de fumar
4. Tomar mucho líquido (agua, refrescos naturales)

Diabetes Mellitus

Síntomas de la DM:

- Frecuencia en orinar (fenómeno de la "cama mojada" en los niños).
- Hambre inusual → Sed excesiva → Vista nublada → Debilidad y cansancio
- Pérdida de peso → Irritabilidad y cambios de ánimo → Infecciones frecuentes
- Sensación de malestar en el estómago y vómitos → Cortaduras y rasguños que no se curan, o que se curan muy lentamente → Picazón o entumecimiento en las manos o los pies → Infecciones recurrentes en la piel, la encía o la vejiga
- Además se encuentran elevados niveles de azúcar en la sangre y en la orina.

Como prevenir la Diabetes Mellitus

- 1) Mantener una nutrición adecuada evitando la ingesta exagerada de azúcares y grasa.
- 2) Realizar ejercicios físicos todos los días, por el lapso de 30 minutos.
- 3) Evitar el sobrepeso, la obesidad, el tabaquismo y la hipertensión arterial.
- 4) Evitar complicaciones y la discapacidad a través de una intervención médica oportuna.
- 5) Auto cuidado de las piernas y pies consultando ante cualquier alteración de la coloración de la piel, la aparición de ampollas, grieta o lesión.
- 6) Autoanálisis domiciliarios antes y dos horas después de las comidas. Cambiando el estilo de vida y los hábitos alimentarios, es la principal forma de prevenir la diabetes y sus complicaciones, logrando de esta manera mejorar la calidad de vida.

Infarto agudo del miocardio

Síntomas

→ Dolor torácico intenso y prolongado, que se percibe como una presión intensa y que puede extenderse a brazos y hombros (sobre todo izquierdos), espalda e incluso dientes y mandíbula. El dolor se describe como un puño enorme que retuerce el corazón. Es similar al de la angina de pecho, pero más prolongado y no cesa aunque se aplique un comprimido de nitroglicerina bajo la lengua.

→ Dificultad para respirar → Sudoración → Palidez → Mareo. Es el único síntoma en un 10 por ciento → Otros síntomas: Pueden aparecer náuseas, vómitos, desvanecimiento y sudoración.

Como prevenir Infarto agudo del miocardio

1. Hacer ejercicio. 2. Dejar de fumar. 3. Evitar las bebidas alcohólicas 4. Llevar una dieta equilibrada, rica en frutas, verduras, legumbres y cereal.

Accidente vascular encefálico agudo

Síntomas

→ Comienza repentinamente y puede ser intenso, Puede empeorar al estar acostado, Lo despierta a uno.

→ Empeora cuando se cambia de posición o cuando se agacha, hace esfuerzo o tose.

→ Cambio en la lucidez mental (incluso, somnolencia, pérdida del conocimiento y coma).

→ Cambios en la audición.

→ Cambios en el sentido del gusto.

→ Cambios que afectan el tacto y la capacidad para sentir el dolor, la presión o temperaturas diferentes.

→ Torpeza.

→ Confusión o pérdida de memoria.

→ Dificultad para deglutir.

→ Dificultad para leer o escribir.

→ Mareos o sensación anormal de movimiento (vértigo).

- Problemas con la vista, incluso disminución de la visión, visión doble o ceguera total.
- Falta de control de esfínteres.
- Pérdida del equilibrio.
- Pérdida de la coordinación.
- Debilidad muscular en la cara, el brazo o la pierna (por lo regular sólo en un lado).
- Entumecimiento u hormigueo en un lado del cuerpo.
- Cambios emocionales, de personalidad y estado de ánimo.
- Problemas para hablar o entender a otros que estén hablando.
- Problemas para caminar.

Como prevenir

1. Lo fundamental es controlar los factores de riesgo asociados; fundamentalmente, son la hipertensión arterial, el colesterol elevado (incluyendo elevados triglicéridos) debido a la ingesta de grasas saturadas animales y aceites hidrogenados y la diabetes.
2. Evitar tabaco, drogas psicotrópicas o estupefacientes y alcohol.
3. Llevar una vida sana: evitar el sedentarismo y en cambio practicar ejercicio físico, y consumir dieta saludable rica en verduras, frutas, proteínas, colesterol bueno y grasas poli saturadas (EPA, DPA, DHA), consumir poca sal y evitando elevadas cantidades de carbohidratos (azúcares y harinas) y grasas saturadas.
4. Evitar la ansiedad y aún más el ángor (la angina de pecho) ya que entre otros problemas vasculares aumenta la hipertensión.
5. Evitar la depresión ya que los estados anímicos depresivos tienden a espesar la sangre haciéndola más trombo génica.
6. Seguir las recomendaciones del médico de cabecera, quien tiene acceso a la información pertinente relacionada con la salud de cada individuo.
7. Evitar el sobrepeso.
8. Evitar deportes de contacto o sobreesfuerzos.
9. Evitar el destres o estrés negativo (especialmente si es crónico).

Serie de Tiempo

Es una colección de observaciones numéricas acerca de las características de cualquier fenómeno, medida secuencialmente en el tiempo, que proceden de un proceso aleatorio, el cual es influenciado por el azar.

El análisis de Serie de Tiempo en el dominio del tiempo utiliza como única información la propia historia de la serie, basándose en la hipótesis central de que las condiciones futuras serán análogas a las pasadas.

Elementos Estadísticos en una Serie de Tiempo

Debido a que las series de tiempo consisten de datos numéricos, es natural hacer uso del herramental estadístico para describirlas y analizarlas, así como ocurre con cualquier otra información numérica.

La estadística utiliza dos enfoques básicos:

1. El enfoque descriptivo, que ocupa esencialmente de resumir y describir en forma concisa, ya sea mediante graficas o a través de algunas medidas descriptivas, la información disponible.
2. El enfoque inferencial, cuyo objetivo fundamental es el de utilizar muestras representativas para realizar inferencia que sea válida para toda la población de donde se obtuvo la muestra.

Al hablar del enfoque descriptivo es importante el orden que se debe llevar en este análisis, primero debemos construir gráficos, en nuestro caso el grafico que describe la serie, y analizar el comportamiento de componentes de la serie, tales como Tendencia y Estación (Época); luego hacer los cálculos numéricos necesarios para obtener con mayor exactitud las medidas descriptivas que permitan representar apropiadamente a la serie tales como, promedio de observaciones, desviaciones típicas, autocorrelaciones, correlaciones cruzadas, etc..

En el estudio de serie de tiempo, la población sobre la cual se desea inferir depende fundamentalmente del tipo de análisis y/o modelo que se emplee.

Por consiguiente conviene señalar que uno de los métodos de análisis considerado como clásico es el de descomposición de la serie de tiempo, la cual asume que aquella está formada por las componentes: Tendencia, Estación y Movimiento irregular. Veamos

Componentes de una serie de tiempo:

Tendencia: Refleja las variaciones superiores al año que no son estrictas periódicas.

El efecto estacional: En relación a Épocas del año, se caracteriza por tener una periodicidad de naturaleza fija, aunque la amplitud puede ser variable; en otras palabras, son aquellas altas y bajos que ocurren en períodos particulares del año.

Movimiento irregular: Este es un componente de la serie que no está sujeta a ninguna periodicidad. Puede o no ser aleatorio. Es determinista, cuando puede ser predecible, y es aleatorio cuando no es predecible.

Tipos de Series Temporales

Serie temporal de un nivel: Se refiere a un instante. Los instantes de observación se enumeran de 1 a t. Es decir Y_t es un valor de Y en el instante t. Se puede decir que los instantes están regulados en el tiempo.

Serie temporal de flujo: En el caso de un flujo, cada observación se refiere a un periodo, es decir, flujo transcurrido durante el periodo. Los periodos se enumeran de 1 a t, donde Y_t es el flujo transcurrido durante el periodo t.

Proceso estacionario: Un concepto importante que se encuentra en este ámbito, se dice 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, por lo tanto no puede existir tendencias. Es decir si para toda t,

1. $E[Y_t] = \mu_t ; C_{te}$

2. $Var[Y_t] = \sigma_\epsilon^2 ; C_{te}$

3. $Cov(Y_t, Y_{t+k}) = Cov(Y_t, Y_{t-k}) = y_k; K = 0, \pm 1, \pm 2, \dots$

Un proceso es no-estacionario si sus propiedades varían con el tiempo.

Función de Medias: Llamaremos función de medias del proceso a una función del tiempo que proporciona las medias de las distribuciones marginales Y_t para cada instante;

$$\mu_t = E(Y_t) = E(Y_{t-1}) = \dots = E(Y_{t-k})$$

Un caso particular importante es que todas las variables tengan la misma media. Las realizaciones del proceso, no mostraran tendencia creciente o decreciente y diremos que el proceso es estable en media.

Función de Varianza: Proporciona la varianza en el instante k , así:

$$\sigma_t^2 = Var(Y_t) = Var(Y_{t-1}) = \dots = Var(Y_{t-k})$$

Además diremos que el proceso es estable en varianza si esta es constante en el tiempo. El proceso puede ser estable en media y no en varianza, en caso contrario la llamaremos función de autocovarianza.

$$Cov(Y_t, Y_{t+k}) = Cov(Y_t, Y_{t+1}) = E[(Y_t - \mu_t)(Y_{t+k} - \mu_{t+k})]$$

Función de Autocorrelación: Se denomina función de autocorrelación de un proceso a la función que describe las correlaciones en dos variables cualesquiera del proceso:

$$\rho(t, k) = corr(Y_t, Y_{t+k}) = \frac{Cov(Y_t, Y_{t+k})}{\sigma_t \sigma_{t+k}} = \rho_k ; \quad \forall t, k$$

En general estas dos funciones $Cov(Y_t, Y_{t+k})$ y $\rho(t, k)$, dependen de dos parámetros (t, k) , siendo t el instante inicial y k el intervalo entre observaciones. Una condición de estabilidad que aparece en muchos fenómenos dinámicos, es que la dependencia entre dos observaciones depende del intervalo entre ellas y no del origen considerado. Entonces podemos escribir:

$$Cov(Y_1, Y_{1+k}) = Cov(Y_2, Y_{2-k}) = \gamma_k; K = 0, \pm 1, \pm 2, \dots$$

Función de Autocorrelación Simple de orden k , ρ_k : Mide la correlación entre los valores de la serie distanciados un lapso de tiempo k , es decir un retardo de orden k . Esta nos dice cuanta correlación hay entre datos individuales contiguos en la serie. A la representación de los coeficientes de autocorrelación en función del

retardo se le denomina Función de Autocorrelacion (FAC) o Correlograma para un proceso estacionario la Función se calcula por:

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

Y por tanto $\rho_0 = 1$; $|\rho_j| \leq 1$; para cualquier proceso estocástico teniendo en cuenta $\gamma_0 = \sigma^2$, y se verifica que:

$$\rho_{-k} = \rho_k$$

En la práctica se debe calcular las estimaciones de la Función de Autocorrelacion, a estas estimaciones se les conoce como FACE, las cuales están dadas por:

$$\widehat{\gamma}_k = \frac{\sum_{t=1}^{T-k} (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^T (Y_t - \bar{Y})^2}$$

Al igual que para el coeficiente de correlación lineal simple, se puede calcular un error estándar y por tanto un intervalo de confianza para el coeficiente de autocorrelación.

La Función de Autocorrelacion FAC (acf), es el conjunto de coeficientes de autocorrelación r_k desde 1 hasta un máximo que no puede exceder la mitad de los valores observados (T/2), y es de gran importancia para estudiar la componente Época o Estación (Season) de la serie, ya que si existe, los valores separados entre sí por intervalos iguales al periodo estacional deben estar correlacionados de alguna forma. Es decir que el coeficiente de autocorrelación para un retardo igual al periodo estacional debe ser significativamente diferente de 0. También la FAC permite observar la ESTACIONARIEDAD; propiedad deseable en la serie de tiempo: la cual se asume al ver un decaimiento rápido en la FAC.

Relacionada con la función de autocorrelación nos encontramos con la función de autocorrelación parcial FACP (pacf); para el coeficiente de autocorrelación parcial de orden k , se calcula la correlación entre parejas de valores separados esa distancia, pero eliminando el efecto debido a la correlación producida por retardos anteriores a k .

Operador de Diferencias Δ (diff): Aplicando el operador de diferencia Δ a una variable referida a un momento de tiempo Y_t se tiene el efecto: $\Delta Y_t = Y_t - Y_{t-1}$, es

decir se obtiene la diferencia entre el valor referido al momento actual y el valor que toma la variable en el periodo inmediatamente anterior. Generalmente se dice que se calculan diferencias, por lo que se denota de la siguiente manera $\Delta^d Y_t$; particularmente en la elaboración del modelo **ARIMA** se suele utilizar la terminología diferencias estacionales de periodo S, donde S es el intervalo con que se toman las diferencias. Se coloca Δ_S para distinguirlos de los periodos regulares.

Operador de Retardo L (lag): Cuando se aplica a una variable Y_t significa que se retarda en un periodo del subíndice al que va referida la variable, es decir: $LY_t = Y_{t-1}$. Se dice que se calcula k retardos y se denota por $L^k Y_t = Y_{t-k}$.

El operador L puede manipularse como si fuera una cantidad algebraica:
 $L(\alpha Y_t) = \alpha LY_t = \alpha Y_{t-1}$ $L^k L^S(Y_t) = L^{k+S}(Y_t) = Y_{t-k-S}$

Procesos Integrados: La mayoría de las series que se observan no son estacionarias y su nivel medio varia con el tiempo, sin embargo es frecuente que el proceso se convierta en estacionario al diferenciarlo.

Llamaremos proceso primera diferencia de $(Y_t); t = 1, \dots$, a un nuevo proceso W_t obtenido mediante: $\Delta_t Y_t = W_t = Y_t - Y_{t-1}$

Diremos que un proceso es integrado de orden d cuando al diferenciarlo d veces se obtiene un proceso estacionario.

Proceso de Ruido Blanco: Está definido por:

$$E[\varepsilon_t] = 0, \quad \forall t$$

$$Var[\varepsilon_t] = \sigma^2, \quad \forall t$$

$$Cov[\varepsilon_t, \varepsilon_{t-k}] = 0; \quad k = \pm 1, \pm 2, \dots$$

La incorrelacion garantiza la independencia y el proceso resultante se denomina proceso de Ruido Blanco y se representa por $Y_t = \varepsilon_t$

Modelo Clásico de Series Temporales

El primer paso obligatorio para analizar una serie temporal es presentar un gráfico de la evolución de la variable a lo largo del tiempo, es decir un gráfico de secuencia, luego por medio de análisis basados en la observación del gráfico se debe determinar si la secuencia de valores es completamente aleatoria o si, por el contrario, se puede encontrar algún patrón a lo largo del tiempo, pues solo en este caso se podrá continuar el análisis.

La metodología tradicional para el estudio de series temporales es bastante sencilla de comprender, fundamentalmente se basa en descomponer las series en varias partes:

Tendencia: Es la dirección general de la pendiente (general o a tramos) de la variable en el periodo de observación, es decir el cambio a largo plazo de la media de la serie.

El Factor Cíclico: Patrón de la serie que revela cierta propensión a repetirse a largo plazo una misma secuencia de movimientos. A menudo, el componente cíclico se considera conjuntamente con el de tendencia: ciclo-tendencia.

El Efecto Estacional: Se caracteriza por tener una periodicidad de naturaleza fija, aunque la amplitud puede ser fluctuaciones periódicas de la variable, en periodos relativamente cortos de tiempo.

Residuo: Variación aleatoria de la variable.

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

Análisis de la tendencia: Una primera idea sobre la presencia de tendencia en la serie se obtiene en su representación grafica.

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, el alisado exponencial y diferenciación.

Las Medias Móviles: Es un procedimiento para suavizar la evolución de la serie y eliminar parte de las oscilaciones accidentales. Consiste en sustituir cada valor observado por la media de este valor y los que lo rodean.

Las bandas de suavización utilizadas, pueden ser de tres, cinco, siete, etc. elementos según tomemos uno, dos, tres, o más elementos por encima y por debajo.

En este procedimiento se pierde los primeros y últimos elementos de la serie, lo cual, es un inconveniente para series de tiempo cortas.

Una media móvil se calcula, para cada punto como un promedio del mismo número de valores a cada lado de ese punto. Así una media móvil de tres puntos se calcula como:

$$m(x_t) = \frac{x_{t-1} + x_t + x_{t+1}}{3}$$

Mientras que una media móvil de cuatro puntos viene dada por:

$$m(x_t) = \frac{\left(\frac{1}{2}x_{t-2}\right) + x_{t-1} + x_t + x_{t+1} + \left(\frac{1}{2}x_{t+2}\right)}{4}$$

Cuando la cantidad de puntos de la media móvil es par, se toma la mitad de los valores extremos.

ALISADO EXPONENCIAL: Los valores que se registran en una serie temporal de la forma (y_1, y_2, \dots, y_n) , tienen normalmente, una acusada variabilidad en todo el periodo muestral que se aprecia en la multitud de "picos" que aparecen en su representación gráfica. El comportamiento global de la serie se apreciara mejor cuando los picos se suavicen, rebajando las fluctuaciones locales; este proceso

recibe el nombre de alisado exponencial que consiste en generar la serie $(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)$, de la forma recursiva siguiente:

$$\begin{cases} \hat{y}_1 = y_1 \\ \hat{y}_t = \alpha y_t + (1 - \alpha)\hat{y}_{t-1}; t = 2, 3, \dots, n \end{cases}$$

Donde α es un número entre 0 y 1 que recibe el nombre de coeficiente de alisamiento, el cual se comporta de la siguiente manera:

Si α es próximo a 1, entonces se asigna mayor importancia a la información que aportan las últimas observaciones realizadas.

Si α es cercano a 0, se reparte más equitativamente la importancia que se concede a todas las observaciones muestrales, incluidas las más remotas en el pasado.

El método del alisado exponencial es útil también para realizar predicciones. Así, si la última observación es la de la etapa n , el pronóstico para $n+1$ será el valor \hat{y}_n .

Diferenciación: Es otra clase que particularmente se utiliza para eliminar la tendencia, se basa en aplicar diferencias a la serie hasta convertirla en estacionaria. Una diferencia de primer orden obtiene restando dos valores contiguos: $\nabla x_{t+1} = x_{t+1} - x_t$. Si volvemos a diferenciar esa serie, restando los nuevos valores consecutivos obtenemos una nueva serie más suavizada $\nabla^2 x_{t+1} = \nabla x_{t+2} - \nabla x_{t+1}$

Enfoque moderno de series temporales

"Metodología Box-Jenkins"

A inicios de los años 70, G.E.P. Box, profesor de Estadística de la Universidad de Wisconsin, y G.M. Jenkins, profesor de Ingeniería de Sistema de la Universidad de Lancaster, introdujeron una pequeña revolución en el enfoque del análisis de series temporales, en sus trabajos sobre el comportamiento de la contaminación en la bahía de San Francisco, con el propósito de establecer mejores mecanismos de pronósticos y control.

El libro en el que describen la metodología, se convirtió rápidamente en un clásico, y sus procedimientos se utilizan ampliamente desde entonces en diferentes ramas de la ciencia, conociéndose como modelos Box-Jenkins.

Para este tipo de modelos, el primer paso consiste en convertir nuestra serie de observaciones en una serie estacionaria, que es aquella en la que ni la media, ni la varianza, ni las autocorrelaciones dependen del tiempo. Una vez "estabilizada" la serie mediante las transformaciones adecuadas, se procede a estudiar la presencia de regularidades en la serie, para identificar un posible modelo matemático. Para ello se calculan la función de Autocorrelacion simple y parcial, se compara su forma con un catalogo de patrones gráficos, que son típicos de los diferentes modelos propuesto, seleccionando el modelo que más se adecue a la forma de las funciones de Autocorrelacion que hemos obtenido con nuestros datos y que retenga la mayor información contenida en la serie temporal.

Modelo de Series Temporales

Modelos Autorregresivos

Modelo AR(1): Se define por $Y_t = \varphi_1 Y_{t-1} + \varepsilon_t = \Phi(L)Y_t = \varepsilon_t$ utilizando el operador de retardo. Se puede expresar como $(1-\varphi_1 L)Y_t = \varepsilon_t$

Para que el proceso AR(1) sea estacionario implica que $|\varphi_1| < 1$, esto es: $\Phi(L)=0$ luego $1 - \varphi_1 L = 0$ así $L = \frac{1}{\varphi_1}$

A la esperanza de este modelo se le incluye una constante y se tendrá que:

$$Y_t = \delta + \varphi_1 Y_{t-1} + \varepsilon_t$$

Si el proceso se inicia en $-\infty$ y es estacionario, se verifica que la media del proceso será constante para cualquier valor de t. Tomando la esperanza y teniendo en cuenta lo anterior Se verifica que:

$$\mu = \frac{\delta}{1-\varphi_1}$$

La varianza está dada por $\gamma_0 = \frac{\sigma_\varepsilon^2}{1-\varphi_1^2}$ y su función de covarianza es

$$\gamma_1 = \varphi_1^k \frac{\sigma_\varepsilon^2}{1-\varphi_1^2}$$

La función de autocovarianza se obtiene directamente efectuando el cociente $\frac{\gamma_t}{\gamma_0}$, dando como resultado $\rho_\tau = \varphi_1^\tau \rho_0 = \varphi_1^\tau$. La función de autocorrelacion parcial de este modelo es $\varphi_{\tau\tau} = 0 = \varphi_1$, para $\tau = 1$ y $\varphi_{\tau\tau} = 0$, para $\tau \geq 1$.

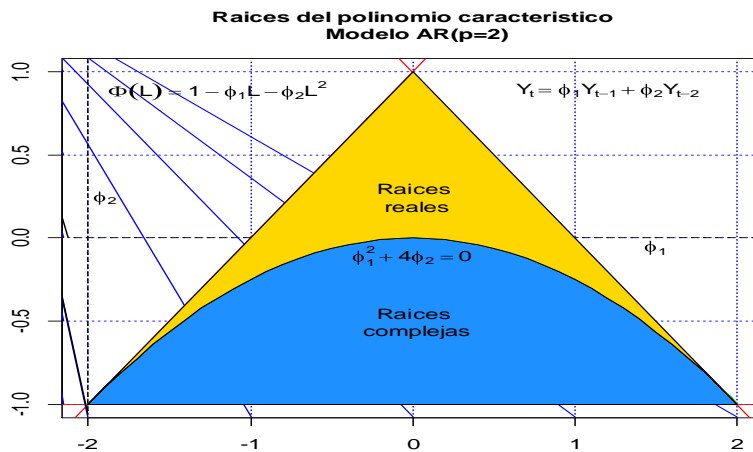
Características del modelo AR(1):

1. Es siempre invertible.
2. Es estacionario si se cumple que $|\varphi_1| \leq 1$
3. La Función de Autocorrelacion no se anula, las correlaciones son menos intensas a medida que aumentan los retardos temporales.

Modelo AR (2): Esta definido por $Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varepsilon_t$. Utilizando el operador de retardos se escribe como

$$\Phi(L) Y_t = \varepsilon_t$$

Para que el proceso sea estacionario se requiere que las raíces de la ecuación $\Phi(L) = 1 - \varphi_1 L - \varphi_2 L^2 = 0$, estén fuera del círculo unidad, denominado mediante L_1 y L_2 . Veamos el siguiente gráfico



Las raíces de la ecuación anterior que serían igual a:

$$L_1 = \frac{-\varphi_1 + \sqrt{\varphi_1^2 + 4\varphi_2}}{2\varphi_2} \quad L_2 = \frac{-\varphi_1 - \sqrt{\varphi_1^2 + 4\varphi_2}}{2\varphi_2}$$

Recordemos que si las raíces son reales deben de ser, en valor absoluto, mayores que uno, mientras que si son complejas, el modulo debe de ser mayor que la unidad, en cualquiera de los dos casos si las raíces caen fuera del círculo unitario implica que:

$$\varphi_1 + \varphi_2 \leq 1 ; \quad \varphi_2 - \varphi_1 \leq 1 \quad ; \quad |\varphi_2| \leq 1$$

Suponiendo que el modelo es estacionario, podemos escribir la esperanza como $E(Y_t) = \mu_t = \frac{\delta}{1-\varphi_1-\varphi_2}$. Y las funciones de varianza y autocovarianza serán respectivamente:

$$\gamma_0 = \varphi_1\gamma_1 + \varphi_2\gamma_2 + \sigma_E^2 \quad \text{y} \quad \gamma_t = \varphi_1\gamma_{t-1} + \varphi_2\gamma_{t-2}, \text{ para } t \geq 0.$$

Dividiendo los miembros $\frac{\gamma_t}{\gamma_0}$ se obtiene la ecuación de autocovarianza. La ecuación en diferencias relativas a las autocorrelaciones está dada por:

$$\varepsilon_t \varphi_1 \rho_{t-1} + \varphi_2 \rho_{t-2}, \quad t \geq 0$$

Las ecuaciones de Yule-Walker se deducen directamente a partir de la ecuación anterior sustituyendo $t=1$ y $t=2$, teniendo en cuenta la propiedad de simetría $\rho_k = \rho_{-k}$,

$$\rho_k = \rho_{-k},$$

Así como el hecho de que $\rho_0 = 1$ se obtiene el siguiente sistema de ecuaciones:

$$\rho_1 = \varphi_1 + \varphi_2 \rho_1$$

$$\rho_2 = \varphi_1 \rho_1 + \varphi_2$$

Una vez calculadas las dos primeras autocorrelaciones, las restantes se pueden obtener de forma recursiva:

$$\rho_1 = \frac{\varphi_1}{1-\varphi_2}$$

Características del modelo AR (2):

1. El modelo es siempre invertible
2. Es estacionario si se cumple que:

$$\varphi_1 + \varphi_2 \leq 1 \quad \varphi_2 - \varphi_1 \leq 1 \quad |\varphi_2| \leq 1$$

3. La función de Autocorrelacion no se anula, su memoria es infinita.

Modelo AR (p): Las observaciones de una serie temporal están asociadas a los diferentes valores que alcanza una magnitud que varía en el tiempo. Tales valores se guardan en una lista de forma (y_1, y_2, \dots, y_n)

El ajuste autorregresivo de orden P, AR (P), consiste en suponer que los valores registrados han sido generados por un modelo subyacente de la forma:

$$y_i = \varphi_0 + \varphi_1 y_{i-1} + \dots + \varphi_p y_{i-p} + e_i$$
$$y_i = \varphi_0 + \sum_{j=1}^p \varphi_j y_{i-j} + e_i$$

Esto es, la lectura que se obtiene en la etapa i-esima depende linealmente de las últimas observaciones, más un error aleatorio representado aquí por e_i . Otras hipótesis adicionales del modelo son:

1. El proceso es estacionario hasta el segundo orden (significa esto que la esperanza estocástica de las observaciones es constante y que la covarianza entre observaciones depende únicamente de su separación en el tiempo).
2. Los residuos e_i son normales de medida nula y varianza σ^2 desconocida; estos residuos son. Por lo demás, independientes entre sí de las variables y_i del proceso.

Tal como ha quedado especificado, el modelo AR (p) tiempo $P + 2$ parametros a estimar a partir de los datos observados:

- Los $P + 1$ coeficientes autorregresivos: $(\varphi_0, \varphi_1, \varphi_2, \dots, \varphi_p)$.
- La varianza del residuo: σ^2

El modelo de estimación adoptado aquí es el de los mínimos cuadrados, que consiste en calcular los parámetros Autorregresivos de forma tal que minimicen el error cuadrático

$$EC \sum_{i=p+1}^n \left(y_i - \varphi_0 - \sum_{j=1}^p \varphi_j y_{i-j} \right)^2$$

Así, el estimador del vector paramétrico es:

$$\hat{\varphi} = \begin{bmatrix} \hat{\varphi}_0 \\ \hat{\varphi}_1 \\ \vdots \\ \hat{\varphi}_p \end{bmatrix} = (X^T X)^{-1} X^T Y$$

Siendo $Y = \begin{bmatrix} y_{p+1} \\ y_{p+2} \\ \vdots \\ y_n \end{bmatrix}$ $X = \begin{bmatrix} 1 & y_p & y_{p-1} & \cdots & y_1 \\ 1 & y_{p+1} & y_p & \cdots & y_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & y_{n-1} & y_{n-2} & \cdots & y_{n-p} \end{bmatrix}$

Habiéndose indicado la transposición matricial mediante el superíndice T. El estimador mínimo cuadrático de la varianza σ^2 , o varianza residual, se calcula mediante

$$s_R^2 = \frac{1}{n-p} \sum_{i=p+1}^n (y_i - \hat{\varphi}_0 - \sum_{j=1}^p \hat{\varphi}_j y_{i-j})^2$$

La etapa de validación, es la fase, en la que se comprueba que el modelo AR (p) explica aceptablemente la serie numérica observada; en ella se debe analizar si las hipótesis del modelo son validas.

Nos limitamos aquí al chequeo de los residuos estimados

$$\hat{e}_i = y_i - \hat{\varphi}_0 - \sum_{j=1}^p \hat{\varphi}_j y_{i-j}$$

Media nula y varianza constante: La serie de residuos $(\hat{e}_{p+1}, \hat{e}_{p+2}, \dots, \hat{e}_n)$, debe comportarse de forma que no experimente tendencias ni alteraciones importantes en su variabilidad a lo largo del tiempo.

Independencia: para contrastar que los residuos son independientes se plantean la hipótesis nula:

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0$$

“los k primeros coeficientes de correlación son nulos”

Frente a la alternativa:

$$H_1 : \rho_i \neq 0$$

“al menos existe un coeficiente de correlación no nulo”.

Se calcula para ello el estadístico de ljung-box:

$$Q_k = (n - p) (n - p + 2) \sum_{j=1}^k \frac{r_j^2}{n - p - j}$$

El cual se distribuye como una distribución X_{k-p}^2 siendo (r_1, r_2, \dots, r_k) las correlaciones correspondientes a los k primeros retardos. El valor k elegido para la realización de la prueba es el entero más próximo a $10 \log_{10}(n - p)$. El contraste se realiza para un nivel de significancia del 5%.

Modelo de medias móviles:

Modelo MA (1): el modelo MA (1) se define de la siguiente manera:

$$Y_t = \varepsilon_t - \theta \varepsilon_{t-1} = (1 - \theta_1 L) \varepsilon_t$$

Este modelo será siempre estacionario, para ser invertible siempre deberá cumplirse que la raíz de la ecuación, $\theta(L) = 1 - \theta_1 - L = 0$ caiga fuera del círculo unitario, es decir que se Cumpla que $|\theta| < 1$.

Para el modelo MA (1) tenemos que su esperanza y su varianza son: $E(Y_t) = \delta$ y $y_0 = \sigma_E^2 (1 + \theta_1^2)$. La función de autocovarianza es: $y_t = -\theta_1 \sigma_E^2$, para $\tau = 1$.

Para los valores de $\tau > 1$, se deduce: $y_t = 0, \tau > 1$ y su función de autocorrelacion está dada por $\rho_k = \frac{y_\tau}{y_0}$ dando como resultado:

$\rho_1 = \frac{-\theta_1}{1 + \theta_1^2}$ para $\tau = 1$. $\rho = 1, \tau > 1$. La función de autocorrelacion parcial está dada por: $\varphi_{\tau\tau} = \frac{-\theta_1 \theta_1^\tau (1 - \theta_1^2)}{1 - \theta_1^2 (\tau - 1)}$, para $\tau \geq 1$.

Modelo MA (2): el modelo MA (2) se define de la siguiente manera:

$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$ Para que este modelo sea invertible se requiere que las raíces del polinomio característico $\theta(L) = 1 - \theta_1 L - \theta_2 L^2 = 0$. Caigan fuera del círculo unidad, es decir que se cumpla que $|\theta_2| < 1$.

Su esperanza matemática es:

$$E[Y_t Y_{t-\tau}] = E[\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}][\varepsilon_{t-\tau} - \theta_1 \varepsilon_{t-\tau-1} - \theta_2 \varepsilon_{t-\tau-2}]$$

Para distintos valores de τ se obtienen los siguientes resultados:

$$\begin{aligned} \gamma &= (1 + \theta_1^2 + \theta_2^2)\sigma_\varepsilon^2, \text{ para } \tau = 0 & \gamma_1 &= (-\theta_2 + \theta_1\theta_2)\sigma_\varepsilon^2, \text{ para } \tau = 1 \\ \gamma_2 &= (-\theta_2)\sigma_\varepsilon^2, \text{ para } \tau = 2 & \gamma_t &= 0, \text{ para } \tau > 2 \end{aligned}$$

Y su función de autocorrelación está dada por:

$$\rho_1 = \frac{-\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2} \quad \rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2} \quad \rho_\tau = 0 \text{ Para } \tau > 2$$

Modelo MA (q): Se define de la siguiente forma:

$y_t\varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q}$ donde (ε_t) es un proceso ruido blanco.

Considerando el operador de retardo L , la ecuación que resulta es: $Y_t = \theta_q(L)\varepsilon_t$, utilizando el operador polinomial de retardos se convierte en: $\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$

Todo proceso de medias móviles es siempre estacionario, por otro lado diremos que un modelo de medias móviles es invertible cuando pueda escribirse como un proceso autorregresivo de orden infinito. Para ello se debe cumplir que las raíces de $\theta_q(L) = 0$, caigan fuera del círculo unitario, esto quiere decir que si las raíces son reales deben de ser en valor absoluto mayor que la unidad, pero sin complejas el modulo debe que ser mayor que uno.

Los procesos MA (q) se caracteriza por que los q primeros coeficiente de la FAC son no nulos que decrecen con el retardo.

Modelos Mixtos Autorregresivos - Medias Móviles ARMA (p, q): Estos modelos puede entenderse como una ecuación de regresión en la que los regresores son valores pasados del regresando y la perturbación aleatoria sigue a su vez un proceso de medias móviles. El modelo ARMA (p, q) está definido como:

$Y_t - \varphi_1 Y_{t-1} - \dots - \varphi_p Y_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$ Donde, ε_t , es un proceso de ruido blanco con ε_t independiente de las Y_t considerando el operador de retardo L , la ecuación que resulta es: $\varphi_p(L)Y_t = \theta(L)\varepsilon_t$

Para que el modelo sea estacionario se requiere que las raíces de la ecuación polinomial $\theta_q(L) = 1 - \theta_1 L - \dots - \theta_q L^q$, caigan fuera del círculo unitario. Dado que

los procesos ARMA son una combinación de los modelos AR, MA, la FAS y FAP de un modelo estacionario ARMA (p, q), será una superposición de la FAS y FAP de los procesos AR(p) y MA(q) correspondientes.

Modelos ARIMA

En estos modelos cada valor tomado por la variable en un instante dado, está influido por los valores de la variable en momentos anteriores, se expresa como una relación lineal, función de:

- 1- Valores recientes de la variable.
- 2- Ruidos en valores recientes de la variable.
- 3- Valores remotos de la variable.
- 4- Ruidos en valores remotos de la variable

El esquema general del modelo es el siguiente:

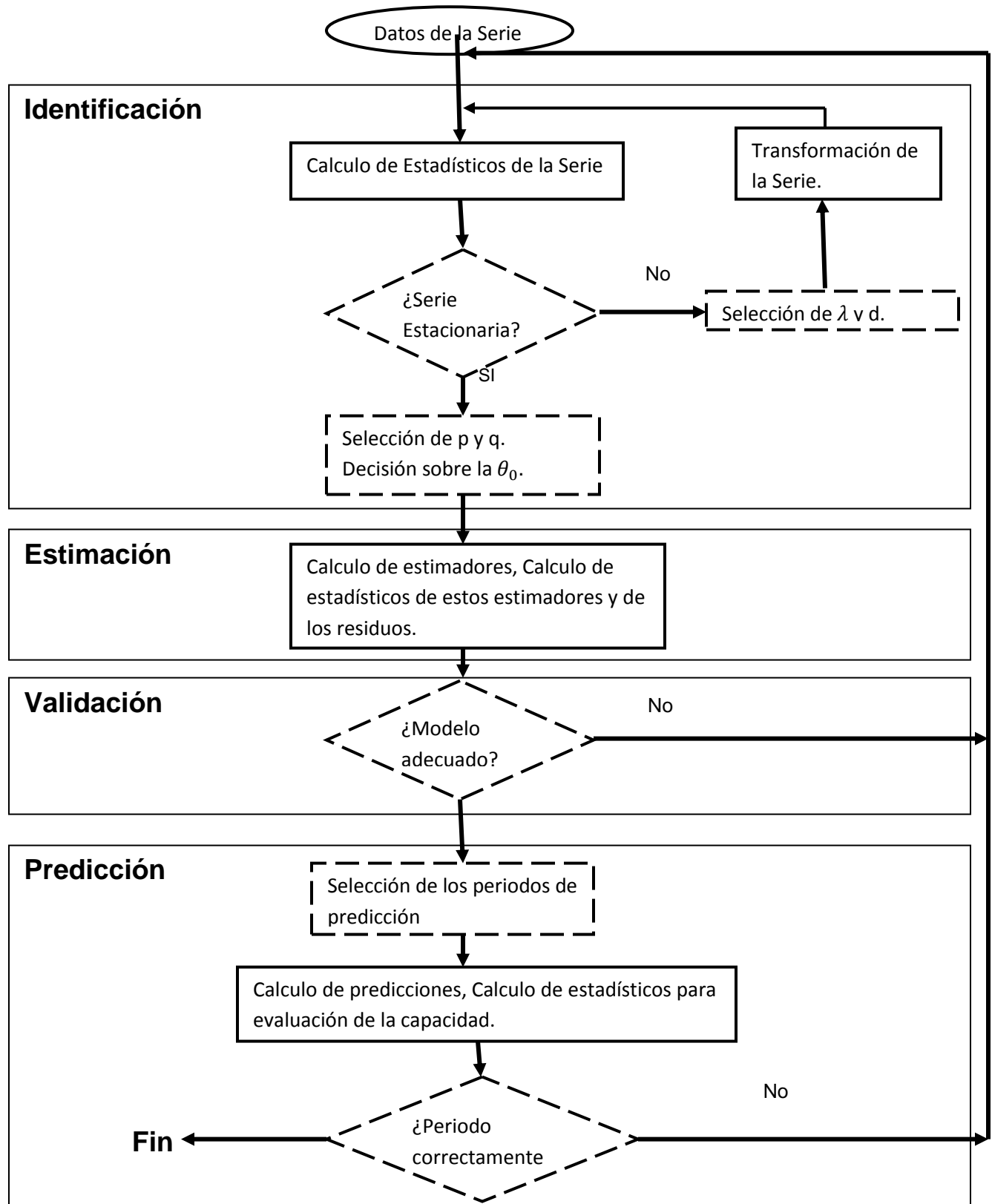
$$X_t = a_1X_{t-1} + a_2X_{t-2} + \dots + a_pX_{t-p} + Z_t + b_1Z_{t-1} + \dots + b_qZ_{t-q}$$

Esta es la formula general de los modelos denominados ARMA, está constituido por una combinación de términos AR (proceso autorregresivo), q termino MA (proceso de medias móviles). La parte AR modela la influencia de los valores anteriores de la serie (X_{t-1} hacia atrás), la parte MA modela la influencia del ruido en valores anteriores de la serie (Z_{t-1} hacia atrás), junto con el término Z_t que corresponde al ruido esperado en el mismo momento t en el que se estima el nuevo valor de la variable X .

Una de las ventajas de estos modelos es su gran simplicidad (suma de términos), frente a los modelos propuestos en la formulación clásica.

La letra I que aparece en el nombre del modelo completo - ARIMA -, corresponde al proceso ultimo a realizar, una vez definido el tipo de modelo y estimados los coeficientes de este ya que entonces hay que restablecer las características originales de la serie de datos, fue transformada para inducir estacionalidad. A ese proceso inverso se denomina en general Integración que aporta esa letra que completa el nombre.

Diagrama de flujo de la Metodología de Box – Jenkins



MATERIAL Y MÉTODO

Tipo de Estudio: Se realizó un análisis de series de tiempo con un enfoque descriptivo longitudinal sobre las defunciones ocurridas en el departamento de León para varios años por meses.

Fuente de datos: Registros de defunciones del SILAIS León en 2005-2010.

Población de estudio: Nuestra población en estudio fueron las defunciones registradas por el SILAIS – León ocurridas en los diferentes puestos que están ubicadas en los municipios del departamento de León.

Recolección de datos: La recolección de la información fue realizada utilizando un formulario previamente diseñado para la institución.

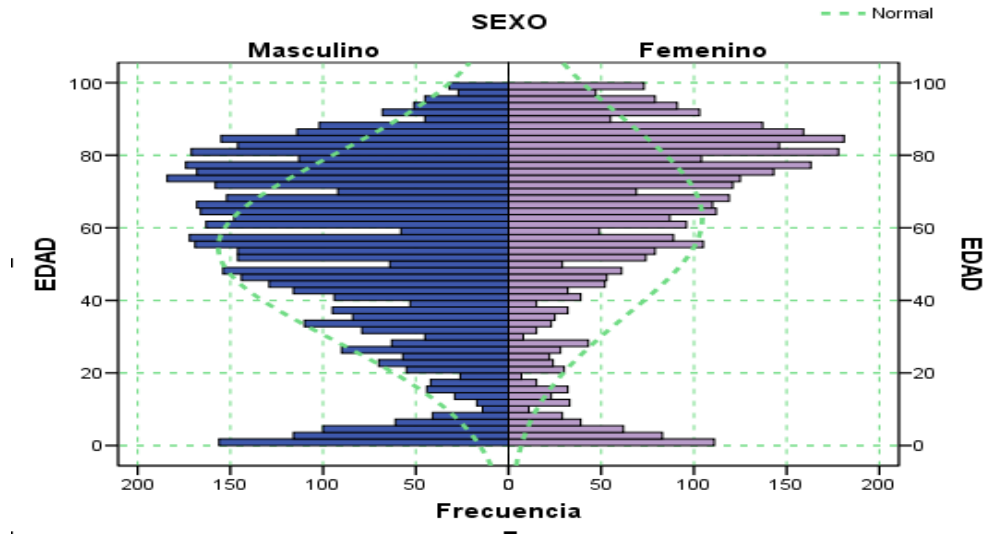
Herramientas: Programas estadísticos R y SPSS.v15, MS-Office.

Técnica Utilizada: Aplicación de Modelos de Series Temporales, Mediante la metodología de Box y Jenkins.

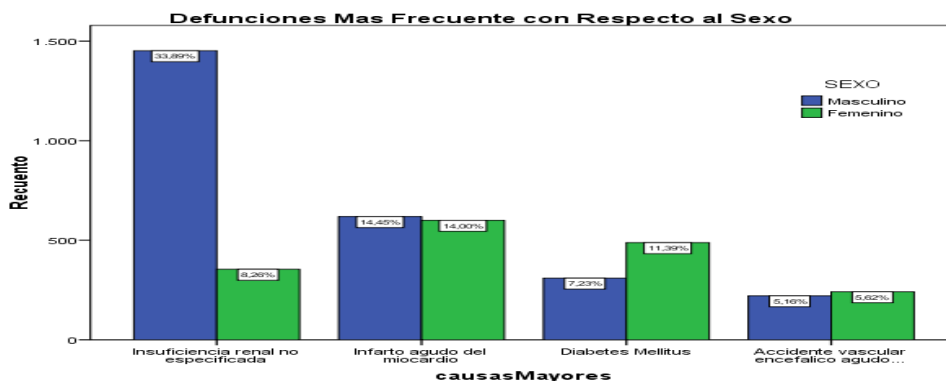
Materiales: Computadora, base de datos del SILAIS León, internet, consulta de libros, papel, lápiz, tiempo.

Variables:

- Barrio
- Municipio de Residencia
- Ocupación
- Edad
- Sexo
- Edad(años, meses, días)
- Estado conyugal
- Año
- Mes
- Fecha de ocurrencia
- Localidad donde ocurrió
- Sitio donde murió
- Recibió atención médica
- Causa de muerte
- Si se debió a (opciones)
- Si ocurrió con (opciones)

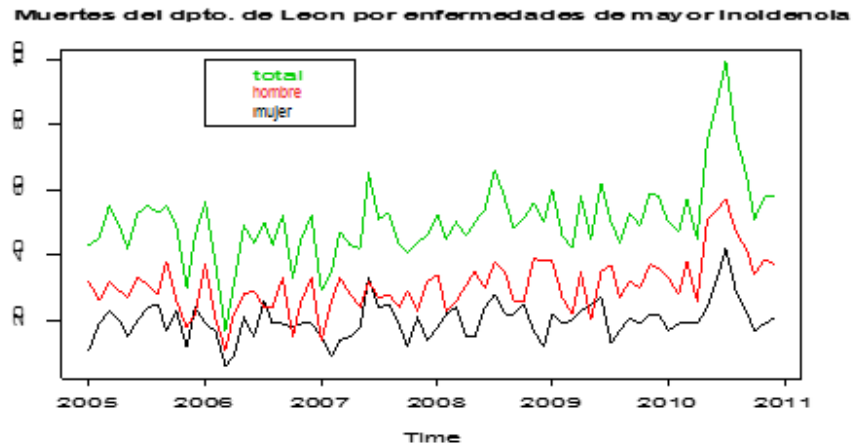


La pirámide poblacional corresponde a la mortalidad del departamento de León, esta tiene el comportamiento esperado las defunciones del primer año de vida es mayor siendo mayor el de los hombres que el de las mujeres, luego la mortalidad comienza a disminuir de tal manera que entre los años de 40 a 10 años la mortalidad es pequeña y siempre tiene el mismo comportamiento, luego en las ultimas edades de 80 a mas ya la mortalidad es mayor en mujeres que en hombres por que murieron más hombres que mujeres.

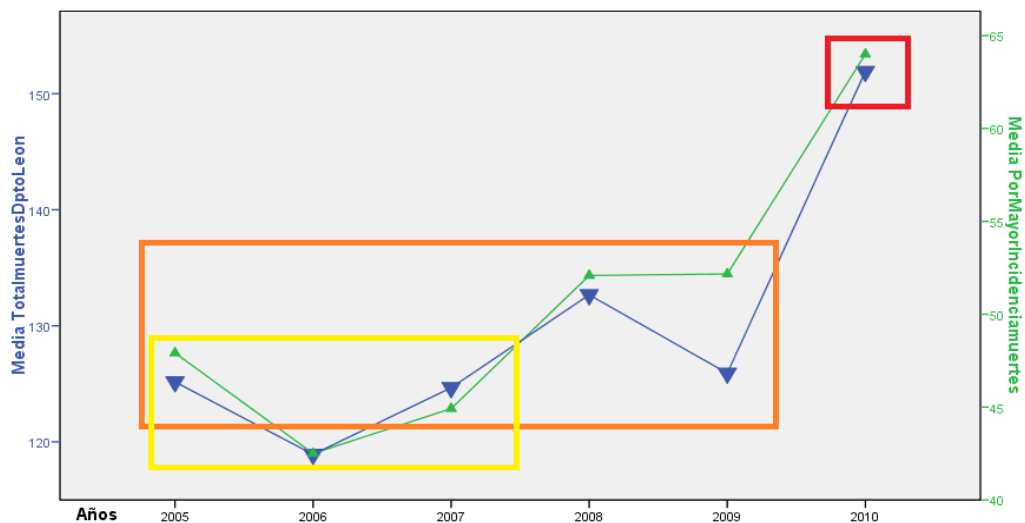


Se observa que la defunción más frecuente con respecto al sexo son los hombres con un 33.89% por Insuficiencia Renal no especificada, mientras que las del sexo femenino es de 8.26%, seguida de Infarto Agudo del Miocardio, en ambos sexos podemos observar que es un promedio casi similar ya que los hombres tienen 14.45% y las mujeres 14%, mientras que la Diabetes Mellitus se aprecia que las mujeres mueren más con un 11.39% y los hombres con 7.23% y con menor promedio en Accidente Vascular Encefálico Agudo.

UNIVERSIDAD NACIONAL AUTONOMA DE NICARAGUA UNAN-LEON

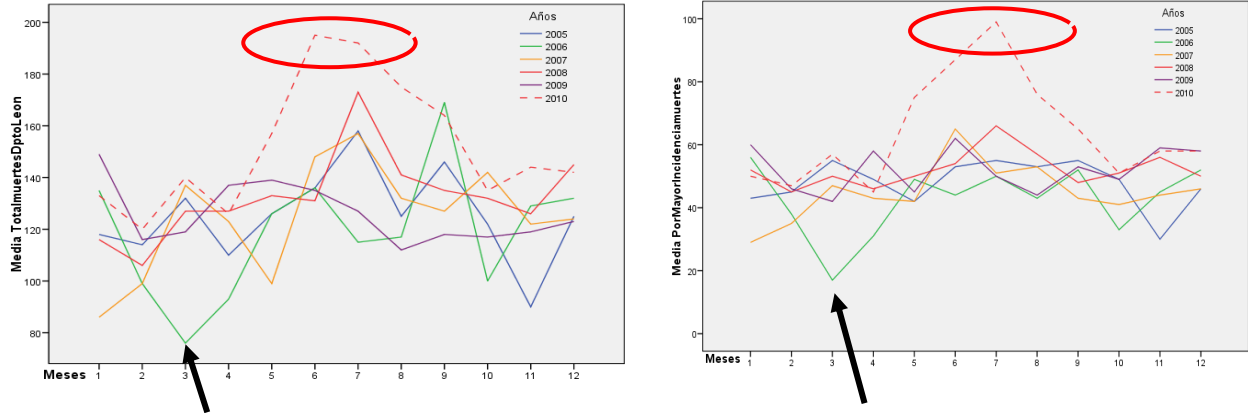


Como se puede apreciar la serie muestra los componentes principales de mortalidad de enfermedad de mayor incidencia donde los hombres a lo largo del tiempo mueren más que las mujeres mientras que las mujeres se mantienen estable en el transcurso de los años, También se observa que hay una baja de defunciones para ambos grupos en el año 2006 y aumenta en el año 2010.

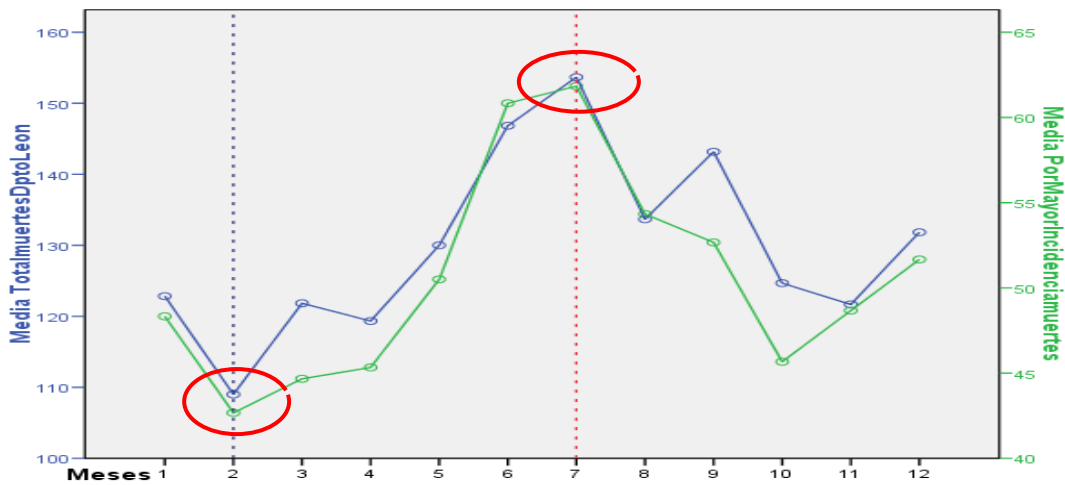


El comportamiento del promedio de muertes tanto del total como por Enfermedades de mayor incidencia es similar: aunque el total barre rangos de 120 a 160, por Enfermedades de mayor incidencia barre rangos de 40 a 65, también representa la media de los subconjuntos del total de muerte del departamento de León y la media por mayor incidencia de muerte por año, los años con menos ocurrencia de muerte son 2005-2007, con un pequeño aumento hasta el 2009, luego un gran incremento en el año 2010.

UNIVERSIDAD NACIONAL AUTONOMA DE NICARAGUA UNAN-LEON



Observamos que tanto para el total de muertes como el de mayor incidencia se aprecia que para ambos, el año 2010 es el año que más muertes fueron registradas siendo el mes de junio y julio los que presentan el mayor número de muertes y el año 2006 que presentó menos muertes registradas siendo marzo que presentó menos defunciones.



Este representa la media total de muertes del departamento de León y la media por enfermedades de mayor incidencia de muertes ambos con respecto a los meses, en el cual podemos observar que la tendencia es aumentar, el mes que reporta menor mortalidad es febrero y el de mayor incidencia de mortalidad es julio para ambos grupos.

Normalidad, Homogeneidad e Independencia

El supuesto de Normalidad por cada año, a como vemos en la siguiente tabla se cumple para todos, según la prueba de Shapiro-Wilks (para muestras pequeñas), excepto el 2010 de las defunciones de mujeres.

"Hipótesis de Normalidad por año, p-valor para d.mujer y d.hombre"						
Año	2005	2006	2007	2008	2009	2010
mujer	0.197	0.059	0.499	0.639	0.887	0.010
hombre	0.963	0.946	0.156	0.252	0.120	0.762

La Homogeneidad de varianzas aplicando la Prueba de Levene's (con centro en la mediana), se acepta tanto para las defunciones de hombres como de mujeres.

"Hipótesis de Homogeneidad para los años, p-valor por género"						
d.mujer				d.hombre		
	Df	F value	Pr(>F)		Df	F value
group	5	0.6764	0.6428	group	5	1.5778
error	66			error	66	0.1786

Para cada serie de tiempo, al aplicar la prueba de Box-Ljung sobre independencia de las autocorrelaciones seriales, esta se rechaza, por lo cual sería conveniente modelizar dichos datos, en ambos casos.

data: d.mujer. León X-squared = 11.3122, df = 1, p-value = 0.00077	data: d.hombre. León X-squared = 15.4753, df = 1, p-value = 0.0000836
--	---

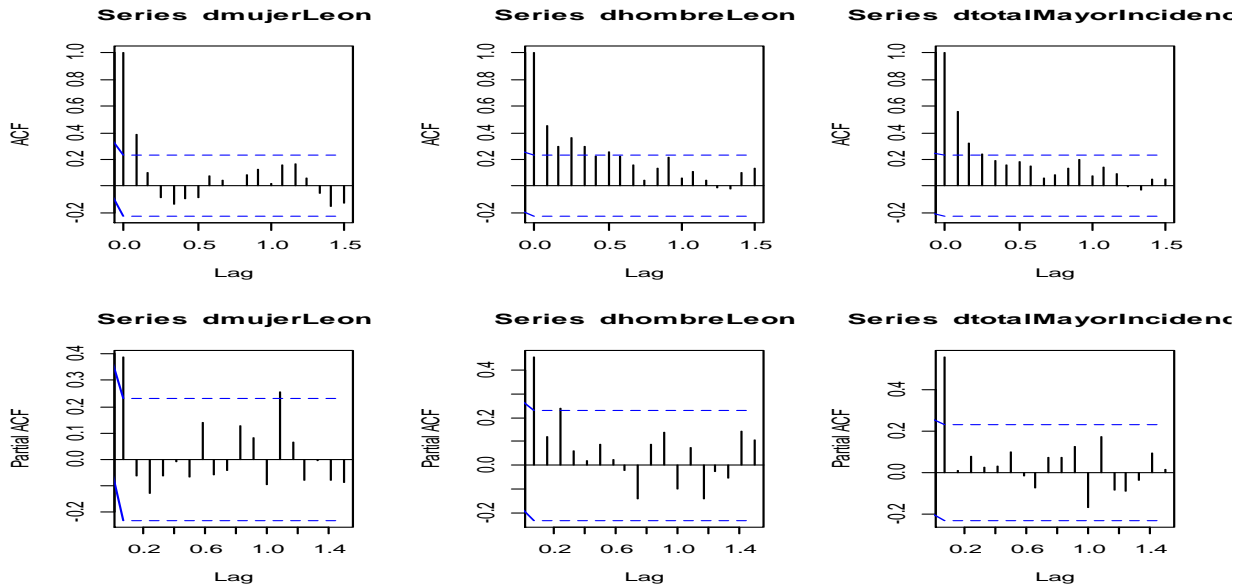
Las siguientes tablas de Análisis de Varianza (DBCA) consideran los factores años ($p_1=6$) y meses ($p_2=12$), resultando significativa la diferencia en las defunciones medias por año, en cambio meses solo para muertes de mujeres.

Respuesta: d.mujer						Respuesta: d.hombre					
	g.l	S.C	M.C.	Fcalc.	Pr(>F)		g.l	S.C.	M.C.	Fcalc.	Pr(>F)
años	5	287.90	57.581	2.2309	0.06403 •	años	5	1780.2	356.03	8.4012	5.9e-06 ***
meses	11	665.15	60.468	2.3428	0.01900 *	meses	11	737.0	67.00	1.5810	0.1305
Resid.	55	1419.60	25.811			Resid.	55	2330.8	42.38		

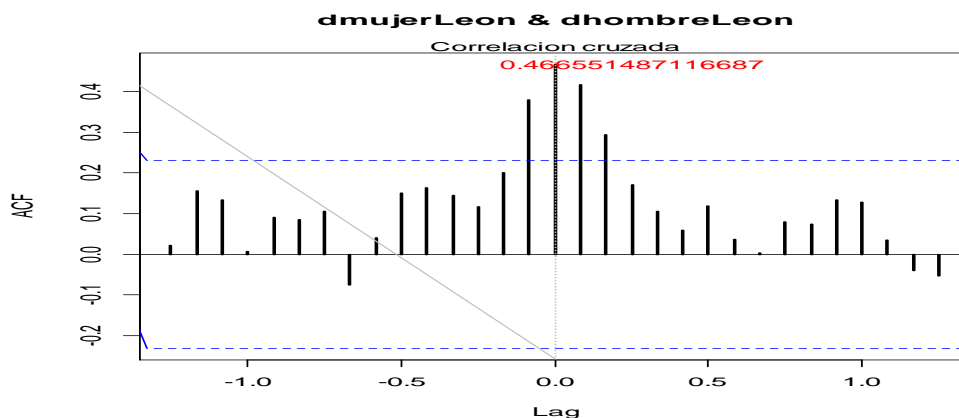
Códigos de Significatividad: 0 '***', 0.001 '**', 0.010 '*', 0.050 '•', 0.100 '' 1.000

UNIVERSIDAD NACIONAL AUTONOMA DE NICARAGUA UNAN-LEON

Correlogramas Simple y Parcial: La FAC observamos para mujeres, que los retardos decaen rápidamente dando evidencia de estacionariedad en media, resultando solo la primera autocorrelación fuera de la banda de confianza (95%) lo que podría estar señalando un modelo MA(1), mientras que para el de los hombres decae menos rápido FAC y podría ser un modelo MA(3) para el total decae lentamente en un MA(2).



La FACP, para mujer, sugiere un AR(1); pues para el retardo 13, al aumentar las bandas a 99% de confianza quedaría aquel incluido. Para hombres se ve un AR(3) y para el total un AR(1). Asimismo deberían tomar en cuenta los modelos combinados ARMA de los órdenes previos.



Al calcular las correlaciones cruzadas de cada serie por Sexo la mayor correlación se alcanza en el orden cero y disminuye en los desplazamientos hacia atrás y

UNIVERSIDAD NACIONAL AUTONOMA DE NICARAGUA UNAN-LEON

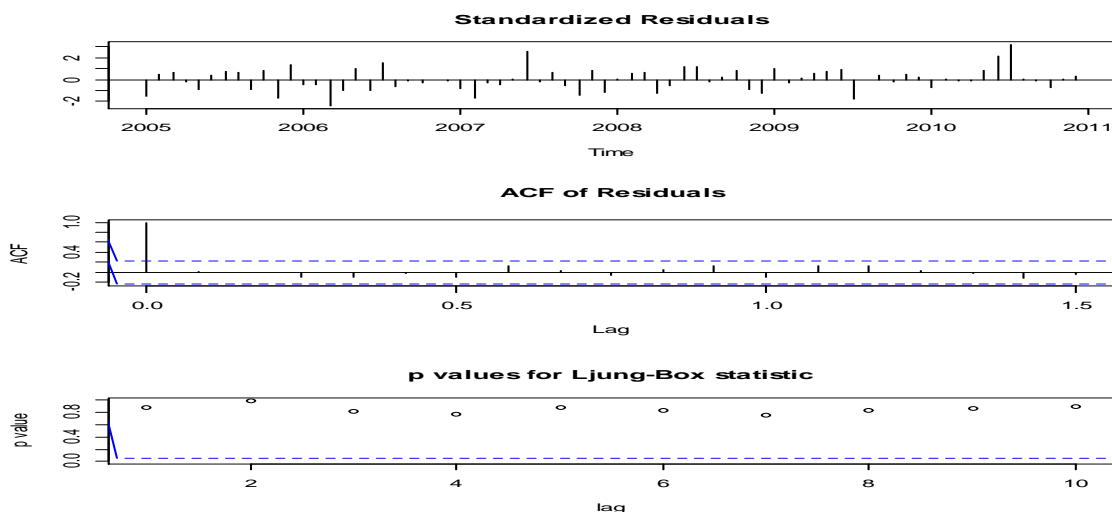
hacia adelante por los meses. Siendo solo significativas cuatro correlaciones cruzadas. Ya que ambas series resultan ser variables dependientes tomadas de forma simultánea, no conviene considerar una u otra como variable regresora.

Estimación de los modelos para las series de tiempo de cada género

La siguiente tabla muestra la estimación de los parámetros a los modelos Autorregresivos de orden 1 para mujer y orden 3 para hombre.

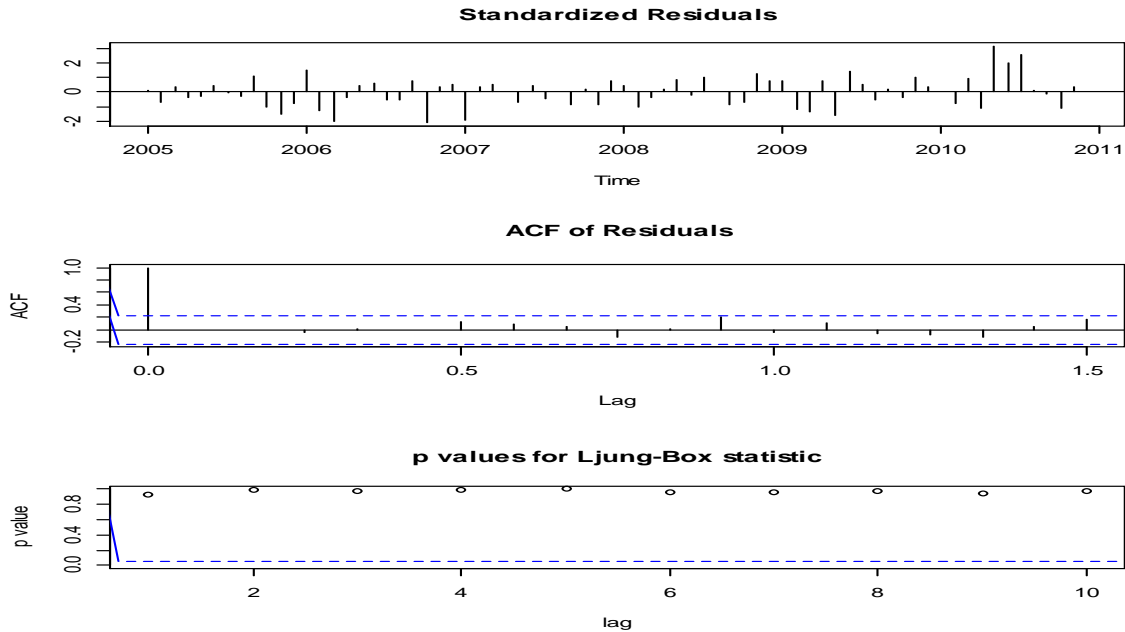
Modelo AR(1) para Serie de dmujer	Modelo AR(3) para Serie de dhombre
arima(x = dmujer, order = c(1, 0, 0)) Coefficients: ar1 intercept 0.3964 19.8601 s.e. 0.1091 1.0209 sigma^2 estimated as 27.81 log likelihood = -221.96, aic = 449.93	arima(x = dhombre, order = c(3, 0, 0)) Coefficients: ar1 ar2 ar3 intercept 0.3671 0.0257 0.2388 30.9070 s.e. 0.1134 0.1220 0.1131 2.1603 sigma^2 estimated as 49.35 log likelihood = -242.74, aic = 495.48

Según el criterio de información de Akaike AIC, de los modelos Autorregresivos, para hombre es AR(3) y para mujer AR(1), el modelo estimado para mujer es el que retiene más información en su ajuste. Las varianzas residuales estimadas, de igual forma, es menor la del modelo para las defunciones de mujer. A continuación se genera los gráficos de diagnóstico para estos modelos.

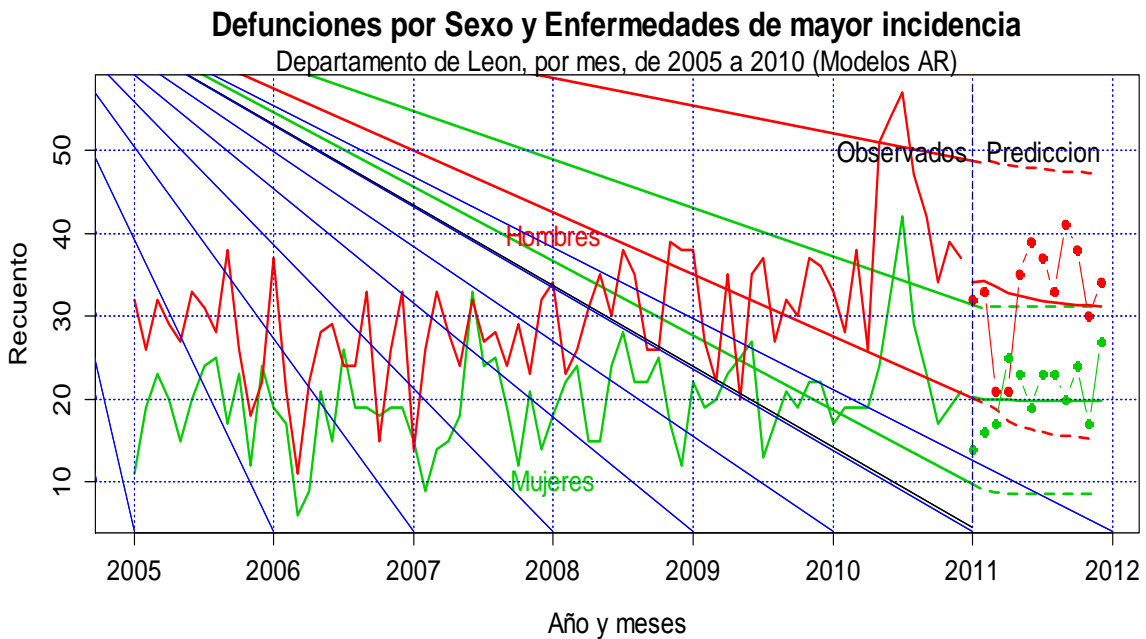


Para el diagnóstico sobre los residuos (dmujer), se observa de la FAC como procedente de un proceso de ruido blanco. Según la prueba de Ljung-Box no hay autocorrelaciones significativas en los residuos.

Para el diagnostico sobre los residuos (dhombre), se observa de la FAC como procedente de un proceso de ruido blanco. Según la prueba de Ljung-Box no hay autocorrelaciones significativas en los residuos.



Según los modelos autorregresivos (gráfico abajo) muestran que los pronósticos la serie se estabilizan entre 10 y 30 casos para mujer, y entre 15 y 46 casos para hombre, ambas bandas contienen a los datos reservados del año 2011.

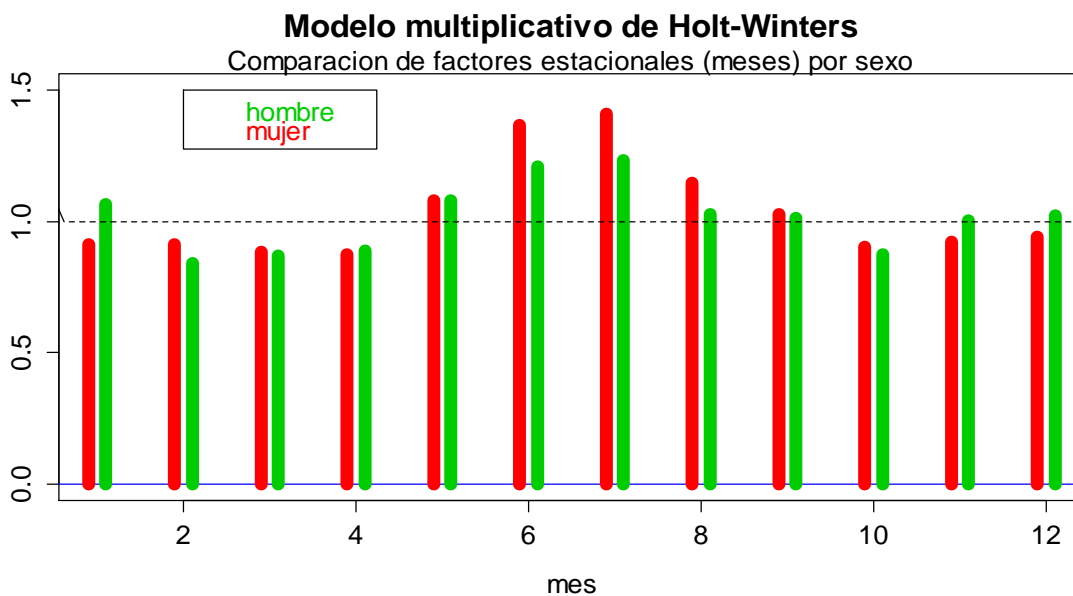


Veamos ahora para modelos no paramétricos de Holt-Winters incluyendo a ambas componentes Tendencia y Estación o Época. Entre los modelos aditivo y multiplicativo resultan con más retención de información los correspondientes al multiplicativo, con sumas de cuadrados residuales SSE=2447.654 y SSE=3878.00 para mujer y hombre respectivamente. La estimación de parámetros es, la siguiente sin incluir el estimado para la pendiente (beta), (factores en anexo)

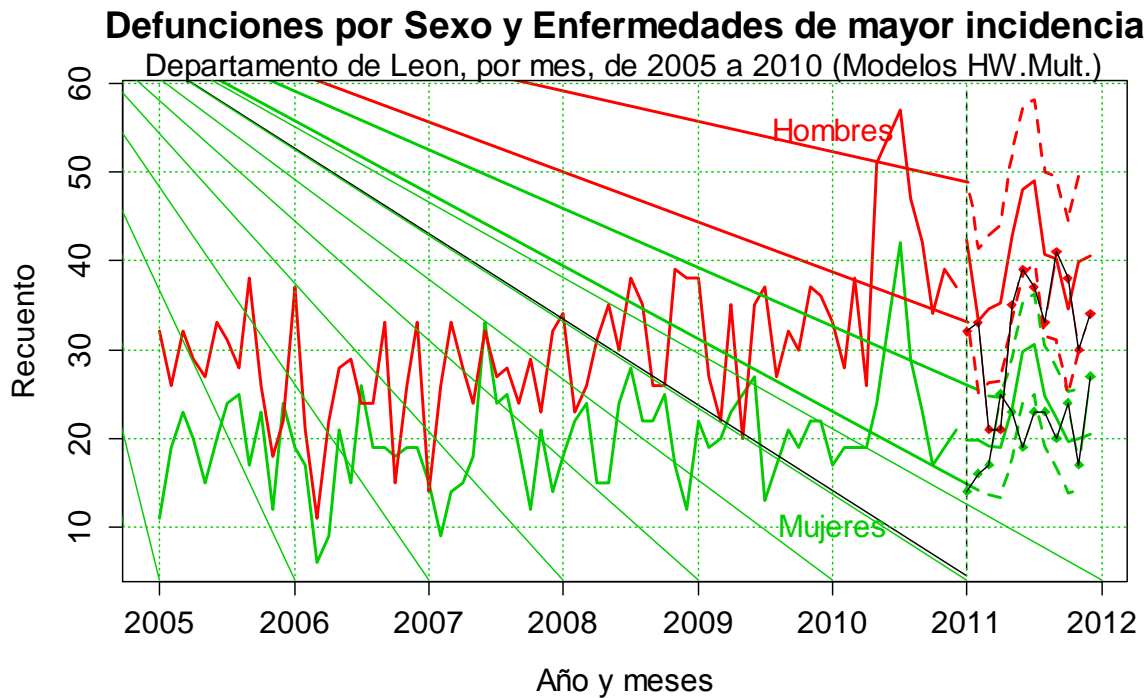
Smoothing parameters: alpha: 0.02413032 beta : FALSE gamma: 0.430112	Smoothing parameters: alpha: 0.1105442 beta : FALSE gamma: 0.4161422
---	---

Al realizar un ANAVA DBCA sobre los factores estacionales estimados por mes y por sexo del modelo Holt-Winters de cada serie, resultan no significativas para sexo, en cambio para los meses sí hay diferencia significativa: observamos del diagrama de barras, que junio y julio son para mayor aumento, y en los meses de febrero, marzo, abril y octubre son de disminución.

Tabla de Análisis de Varianza						
Respuesta: factorEpocaHW (Valor de factores por mes y por sexo)						
	Df	S. C.	M. C.	Fvalor	Pr(>F)	
Sexo	1	0.00265	0.002650	0.5398	0.4779044	ns
Mes	11	0.51069	0.046426	9.4560	0.0004082	***
Residuos	11	0.05401	0.004910			
Códigos de Significatividad: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 'ns' 1						



Ya para el ajuste y pronósticos del modelo Holt-Winters, vemos la componente de estación (Época del año), a como analizamos de los factores respectivos por mes. Además los valores reservados del año 2011, muestran comportamiento similar por mes y que los intervalos de confianza contienen la serie de nuevos datos.



CONCLUSIONES

Intentamos dar con estos resultados un aporte a la sociedad civil, en su cuidado preventivo personal y familiar, también para profesionales de la salud, puesto que el mismo puede servir de herramienta documental para investigadores y no omitimos que este trabajo no es algo definitivo si no que da las pautas para profundizarlo y obtener resultados mejores. Lo esencial de nuestro trabajo se resume en las siguientes conclusiones del análisis y procesamiento de datos:

✚ Existe relación entre sexo y las causas de muerte en el departamento de León para el período 2005-2010, siendo los meses de Junio y Julio los de mayor aumento para las causas de muerte de mayor incidencia.

✚ Los registros de mortalidad del departamento de León más frecuentes son: Insuficiencia Renal no Especificada, Diabetes Mellitus, Infarto Agudo del Miocardio y Accidente Vascular.

✚ El mayor número de registros de defunciones corresponden al sexo masculino con un 58.61% y las del sexo femenino con un 41.39%, siendo entre ambos una diferencia significativa.

✚ Las edades con menos incidencia de muerte para ambos grupos es entre 10 y 20 años.

✚ Se encontró evidencia estadística que existe relación significativa que la ocupación está ligada a la causa de muerte.

Recomendaciones

Aunque la modelización se realizó sobre datos seleccionados por causas de muerte de mayor incidencia y por particiones periódicas por mes, es posible modelizar sobre todas las muertes considerando el Sexo, las fechas e incluso las horas de defunción, o seleccionando casos para otros factores que estén incluidos en el formato de recolección de la información de defunciones.

Otras recomendaciones respecto a estudios futuros, además de actualizar el conjunto de datos, se dan las siguientes:

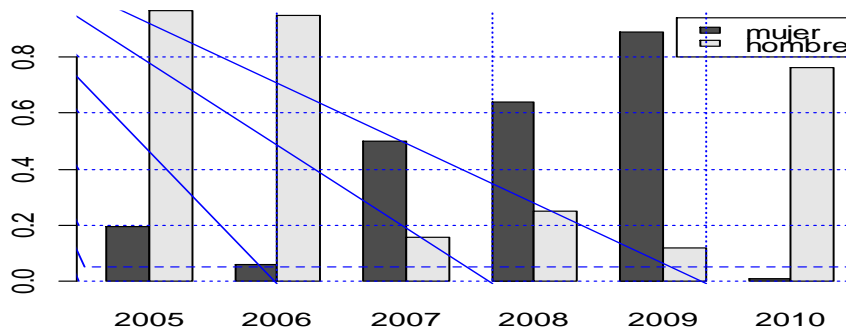
- ✚ Relacionar el número de muertes y el número de población viva a los mismos periodos y que permitan comparar las más grandes y menores diferencias por grupo de edades.
- ✚ Correlacionar las defunciones por Localidad con la población de los territorios respectivos en Estadísticas Espaciales utilizando SIG con GPS.
- ✚ Basarse en la metodología Paramétrica y No Paramétrica de Series de Tiempo, e incluir otros modelos y técnicas estadísticas avanzadas, considerando otras variables regresoras tales como edad y peso del difunto.

Bibliografía

- Daniel Peña Sánchez de Rivera, Estadística modelos y métodos, 2ª edición. Modelos Lineales y Series Temporales. Quinta parte series temporales.
- www.dmedicina.com/enfermedades
- www.webconsultas.com/categorias/salud-al-dia
- León- wikipedia, la enciclopedia libre, wikimedia foundation Inc, <http://es.wikipedia.org/wiki/león>
- Organización panamericana de la salud, organización mundial de la salud en Nicaragua, www.ops.org.ni

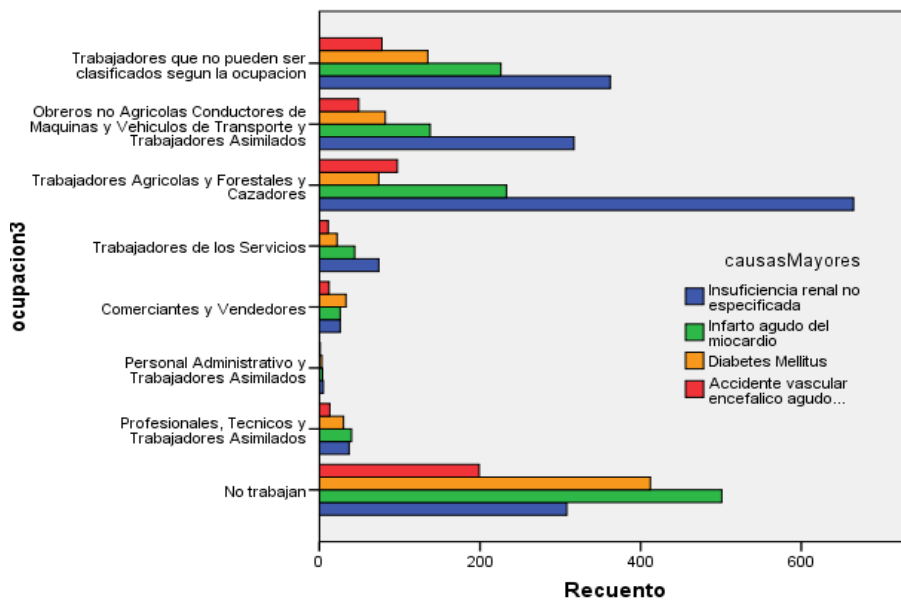
ANEXOS

Pvalores T.Normalidad Shapiro-Wilks



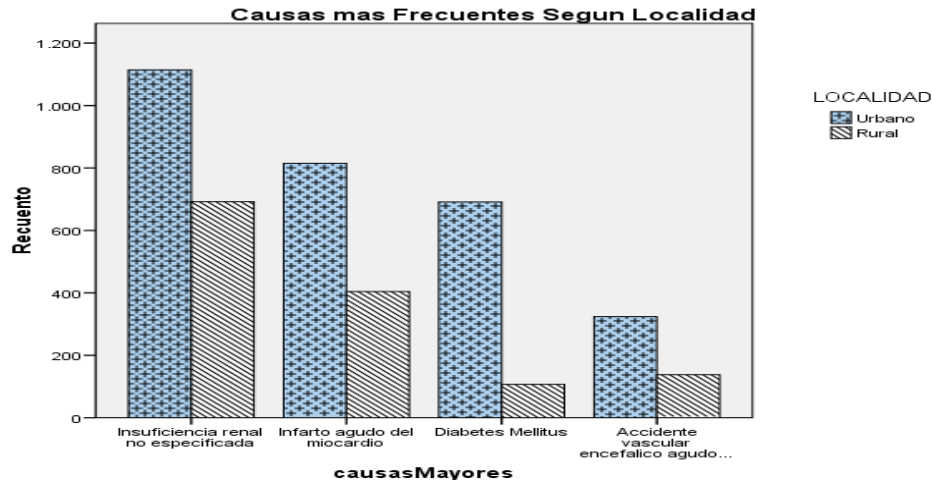
El supuesto de Normalidad para cada año, a como vemos en el siguiente grafico se cumple para todos, según la prueba de Shapiro-Wilks (para muestras pequeñas), excepto el 2010 de las defunciones de mujeres.

Causas Mas Frecuente Segun Ocupacion



A simple vista vemos que los Trabajadores Agrícolas y Forestales son los que mueren más por Insuficiencia Renal Crónica, seguido por los que no trabajan con Infarto Agudo al Miocardio y Diabetes Mellitus y una minoría en Accidentes Vasculares.

UNIVERSIDAD NACIONAL AUTONOMA DE NICARAGUA UNAN-LEON



Este gráfico representa que los del área Urbano mueren más por insuficiencia renal no especificada, y una minoría para los del área Rural por Diabetes Mellitus.

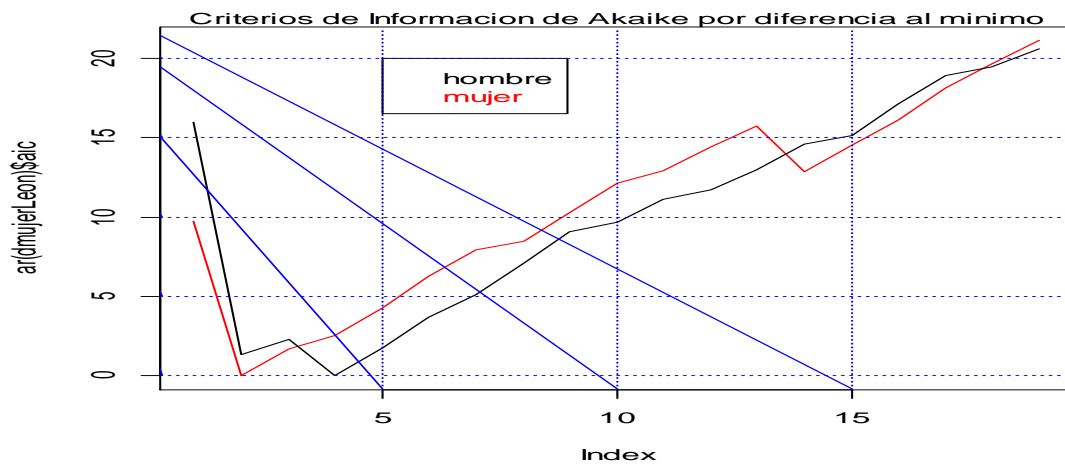


Diagrama de líneas comparativo según los criterios de información de Akaike cuyos valores mínimos son $aic=449.93$ para mujer y $aic= 495.48$ para hombre en los modelos AR(1) y AR(3) respectivamente; los AIC diferenciados, alcanzan su mínimo, en esos retardos.

Coeficientes estacionales del modelo Holt-Winters para cada género

Coef..	a	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12
† ♀	21.709	0.914	0.910	0.884	0.874	1.081	1.367	1.410	1.144	1.026	0.900	0.921	0.940
† ♂	39.706	1.065	0.838	0.871	0.888	1.081	1.208	1.234	1.025	1.013	0.872	1.004	1.021