ARIMA

Fundamentos de Series Temporales

Instructor: José Nelson Zepeda

San Salvador, Abril 2019

Agenda

Concepto

Objetivo

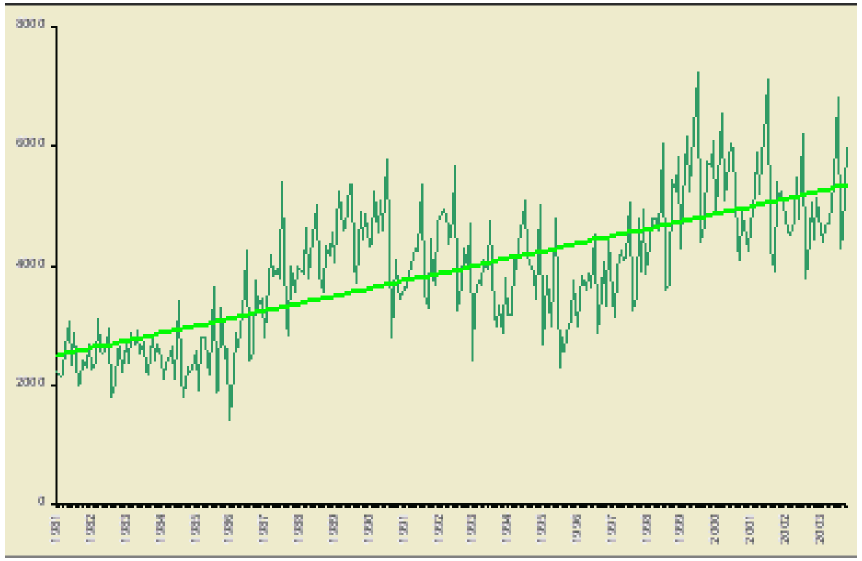
Definiciones

Ejemplo/Simulación

Conceptos Básicos

Serie Temporal: Concepto

Una serie temporal o cronológica es una secuencia de datos, observaciones o valores, medidos en determinados momentos y ordenados cronológicamente en donde los datos pueden estar espaciados a intervalos iguales o desiguales

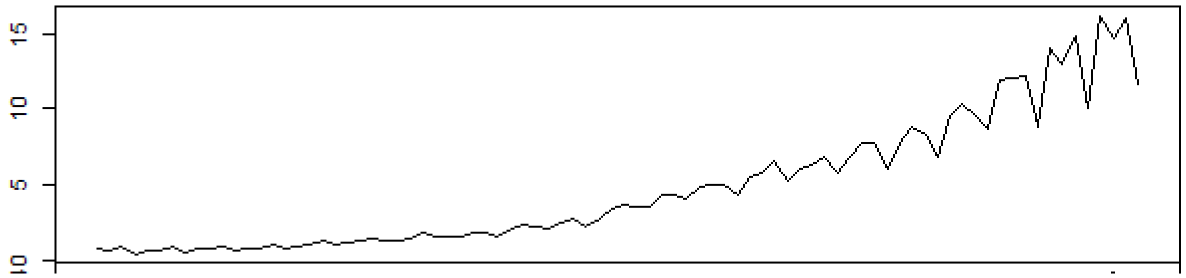


- El orden de los datos importa.
- Las observaciones no son independientes.
- Al estimar relaciones se debe tener en cuenta que no son independientes.
- El tiempo tiene escala y esta escala está ordenada.
- Por tanto, debe utilizar técnicas matemáticas y estadísticas diferentes.
- Una serie temporal permite:
    - Entender el pasado
    - Entender la situación actual
    - Entender el futuro

Objetivo de Una Serie Temporal

El objetivo del análisis de series temporales es doble. Por un lado, se busca explicar las variaciones observadas en la serie en el pasado, tratando de determinar si responden a un determinado patrón de comportamiento. Por otra parte, si se consigue definir ese patrón o modelo, se intentara predecir el comportamiento futuro de la misma

La forma más sencilla de iniciar el análisis de una serie temporal es mediante su representación gráfica, para ello, en un sistema cartesiano, los valores de la serie Yt se representan en el eje de las ordenadas y los periodos de tiempo en el eje de las abscisas. Mediante este tipo de representaciones se pueden detectar las características más sobresalientes de la serie, tales como el movimiento a largo plazo, la amplitud de las oscilaciones, la posible existencia de ciclos, los posibles puntos de ruptura, la presencia de valores atípicos, etc.



Elementos de una Serie Temporal

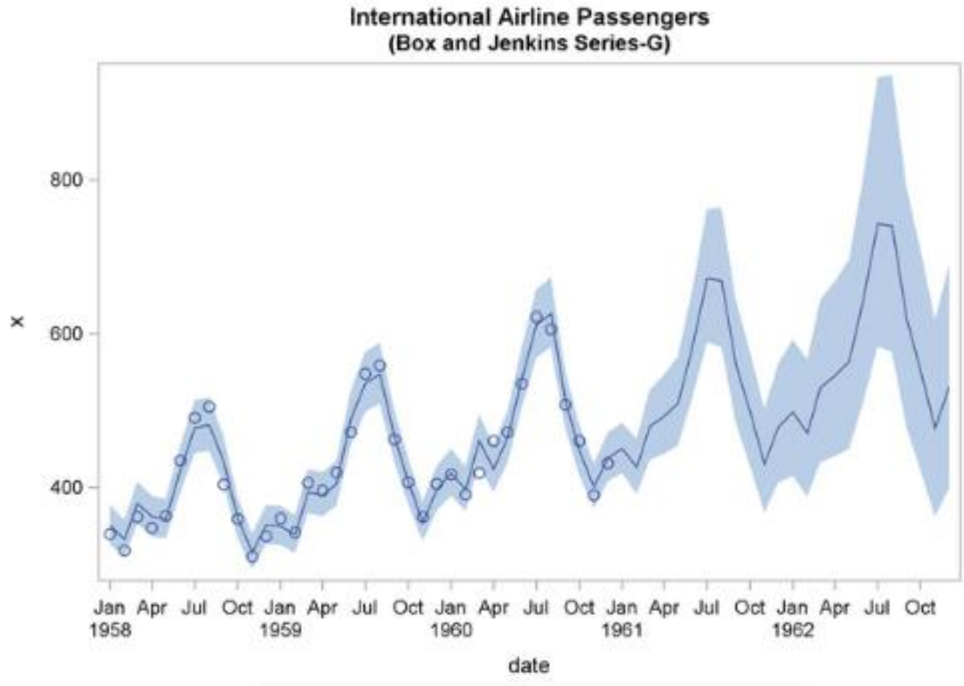
- Tendencia: Movimiento regular de la serie, a largo plazo. La Tendencia mide si temporalmente los valores tienen una direccionalidad hacia arriba o hacia abajo. En definitiva, capta una pendiente general de los valores. Una pendiente que puede ser positiva, si es de subida, o negativa, si es de bajada.
- Estacionalidad: Oscilaciones a corto plazo del periodo regular, mide la presencia de ciclos, de subidas y bajadas realizas con una determinada regularidad.
- Aleatoriedad: Son fluctuaciones producidas por valores eventuales, esporádicos o imprevisibles, que no muestran una periodicidad previsible. En otras palabras, la aleatoriedad mide desvíos respecto de estos dos elementos vistos anteriormente, pequeños alejamientos de la tendencia o de la estacionalidad que se atribuirán a elementos no controlados en el modelo, a elementos incluso idiosincráticos, propios del individuo o los individuos evaluados en aquel momento.

Algunos autores agregan una cuarta componente: Componente Cíclica

ARIMA

Las nuevas propuestas de análisis de series de tiempo se empezaron a implementar en un principio a problemas de contaminación, en la economía, a enfermedades epidemiológicas y en la actualidad a fenómenos físicos y sociales.

Específicamente los modelos desarrollados en las últimas dos décadas son los llamados autorregresivos (AR), de medias móviles (MA), integrados (I), así como sus posibles combinaciones (ARIMA).



A finales del siglo XX diversos investigadores comenzaron a desarrollar y aplicar nuevas propuestas para modelación de series de tiempo.

Procesos Estocásticos

Un proceso estocástico es una sucesión de variables aleatorias Yt ordenadas, pudiendo tomar t cualquier valor entre menos infinito e infinito.

En general las series de interés llevan asociados fenómenos aleatorios. Por este motivo el estudio de su comportamiento pasado solo permite acercarse a la estructura o modelo probabilístico para la predicción del futuro. La teoría de los procesos estocásticos se centra en el estudio y modelización de sistemas que evolucionan a lo largo del tiempo, o del espacio, de acuerdo a unas leyes no determinísticas, esto es, de carácter aleatorio y probabilístico.

Tipos de Modelos

- Proceso de ruido blanco (at): es un proceso formado por una secuencia de variables aleatorias mutuamente independientes e idénticamente distribuidas.

- Modelo no estacionario de corrido aleatorio (I): también conocido en inglés como Random Walk, corresponde a aquellas situaciones en las que los impulsos aleatorios tienden a sumarse o integrarse en el tiempo. La integración refleja la presencia de un componente de tendencia.

- Modelo autorregresivo de orden p AR(p): Se trata de un modelo en el que una determinada observación es predecible a partir de la observación anterior (modelo autorregresivo de primer orden) o a partir de las dos observaciones que les preceden (modelo autorregresivo de segundo orden). En este caso, la observación actual se define como la suma ponderada de una cantidad finita p de observaciones precedentes más un impulso aleatorio independiente.

$$\boxed{Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varphi Y_{t-3} + \cdots + \varphi_p Y_{t-p} + a_t}$$

Tipos de Modelos

- Proceso de ruido blanco (at): es un proceso formado por una secuencia de variables aleatorias mutuamente independientes e idénticamente distribuidas.

- Modelo no estacionario de corrido aleatorio (I): también conocido en inglés como Random Walk, corresponde a aquellas situaciones en las que los impulsos aleatorios tienden a sumarse o integrarse en el tiempo. La integración refleja la presencia de un componente de tendencia.

- Modelo autorregresivo de orden p AR(p): Se trata de un modelo en el que una determinada observación es predecible a partir de la observación anterior (modelo autorregresivo de primer orden) o a partir de las dos observaciones que les preceden (modelo autorregresivo de segundo orden). En este caso, la observación actual se define como la suma ponderada de una cantidad finita p de observaciones precedentes más un impulso aleatorio independiente.

$$\boxed{Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varphi Y_{t-3} + \cdots + \varphi_p Y_{t-p} + a_t}$$

$$Y = a + b_1 X_1 + b_2 X_2 + b_3 X_3 + \cdots + b_p X_p + e$$

Tipos de Modelos

- Modelo de Medias Móviles de orden q MA(q): En este modelo, una determinada observación está condicionada por los impulsos aleatorios de las observaciones anteriores. De esta forma la observación actual se define como la suma del impulso actual y de los impulsos anteriores con un determinado peso

$$\boxed{Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}}$$

Donde at es el residuo o error en el periodo t, la ecuación es semejante a la anterior, con la excepción de que implica que la variable dependiente Yt depende de los valores previos del término residual más que de la variable misma

Tipos de Modelos

- Modelo autorregresivo de medias móviles de orden p,q ARMA(p,q): Este modelo es la combinación de las estructuras anteriores: modelo autorregresivo y modelo de medias móviles. Así, una observación está determinada tanto por observaciones anteriores así como por impulsos aleatorios o también llamados errores de observaciones pasadas.

$$\boxed{Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varphi_3 Y_{t-3} + \cdots + \varphi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}}$$

Se observa que la ecuación es una combinación de las ecuaciones del modelo de AR y MA.

Un modelo ARMA puede ajustarse a cualquier patrón de datos. Sin embargo, los valores de p y q se deben especificar.

Tipos de Modelos

si p=1 y q=0, el modelo ARMA es $Y_t = \varphi_1 Y_{t-1} + a_t$ es decir es un ARMA(1,0).

Si p=2 y q=0, el modelo ARMA es $Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + a_t$ es decir es un ARMA(2,0)

Si p=0 y q=1 el modelo ARMA es $Y_t = a_t - \theta_1 a_{t-1}$ es decir es un ARMA(0,1)

Tipos de Modelos

- Modelo autorregresivo integrado de medias móviles de orden p,d,q ARIMA(p,d,q): al igual que un modelo ARMA, es la combinación de los modelos autorregresivo y el de medias móviles, con la particularidad de incluir un proceso de restablecimiento el cual se denomina integración. La forma general de un modelo ARIMA es semejante a la de un modelo ARMA

$$Y'_t = \varphi_1 Y'_{t-1} + \varphi_2 Y'_{t-2} + \varphi_3 Y'_{t-3} + \cdots + \varphi_p Y'_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$$

Metodología Box-Jenkins

Existen distintas metodologías que ayudan a elegir el modelo que mejor se acople a la serie de tiempo de entre los diversos modelos existentes. La metodología más utilizada y difundida es la que propusieron los profesores G.E.P. Box y G.M. Jenkins en la década de los años 70.

Esta metodología se basa en tratar de determinar cuál es el modelo probabilístico que rige el comportamiento del proceso a lo largo del tiempo.

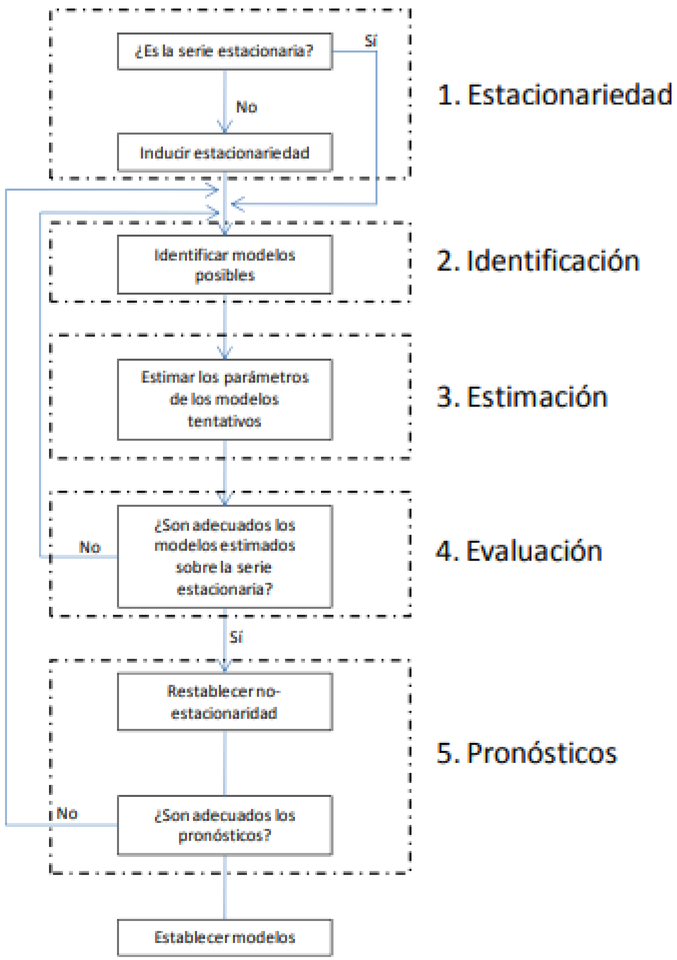
un proceso es lo real, es decir, el fenómeno en sí, del cual se desconoce su mecanismo generador.

Por otro lado un modelo es solo la imitación o representación del proceso.

Consideraciones ARIMA

- Los modelos ARIMA aplican tanto para datos discretos como continuos.
- Aunque la metodología ARIMA trata tanto con datos discretos como continuos, solo se puede aplicar a datos espaciados equidistantemente en el tiempo, en intervalos discretos de tiempo.
- Para elaborar un modelo ARIMA se requiere una cierta cantidad de datos mínimos. Los profesores Box y Jenkins sugieren un mínimo de 50 observaciones.
- Los modelos ARIMA son especialmente útiles en el tratamiento de series que presentan patrones estacionales.
- El método Box-Jenkins aplica a series estacionarias y no estacionarias. Una serie estacionaria es aquella cuya media, varianza y función de autocorrelación permanecen constantes en el tiempo.
- Se asume que las perturbaciones aleatorias presentes en la serie son independientes entre sí, no existe correlación entre ellas, por lo tanto ningún patrón modelable.

Metodología ARIMA

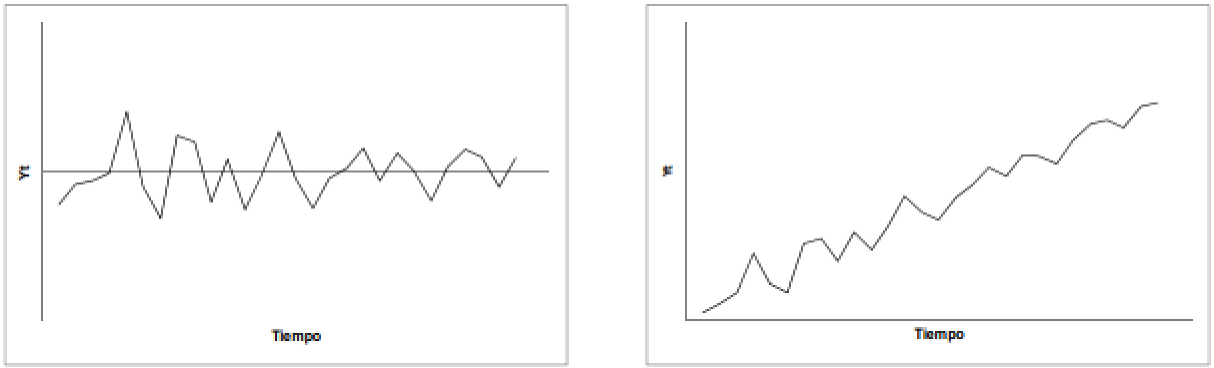


Estacionariedad

Los procesos estocásticos se clasifican entre estacionarios y no estacionarios. La idea de la Estacionariedad está relacionada con la estabilidad de la serie. Un proceso estacionario se describe como una secuencia de datos o valores que no presentan cambio en la media ni cambio en la varianza, por lo cual se dice que la serie es estable.

En un proceso no estacionario, la serie de datos es inestable en el tiempo.

Dicho lo anterior y en palabras prácticas, se puede afirmar que una serie cuyos valores fluctúan respecto a una media constante, es decir sin tendencia, es una serie estacionaria.

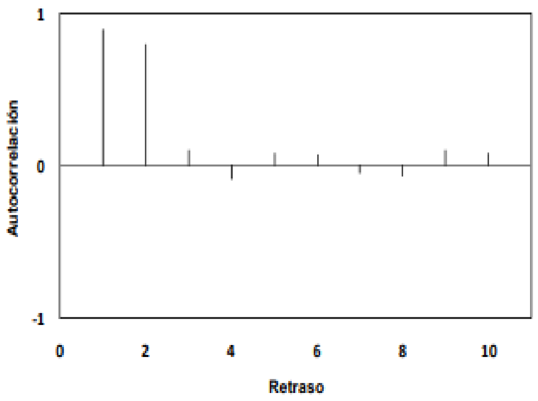


Serie de tiempo estacionaria vs no estacionaria

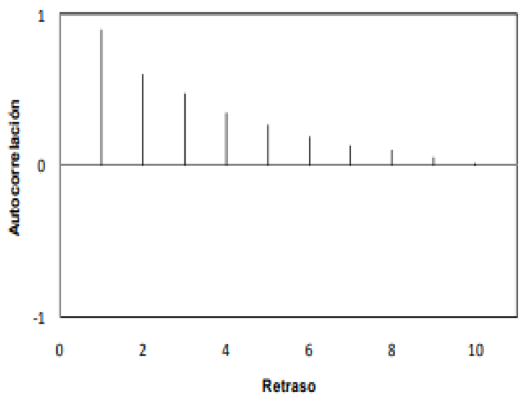
Estacionariedad

Adicional a la estrategia gráfica, es usual recurrir a la función de autocorrelación simple (FAS), en especial aquellas series con tendencias poco remarcada

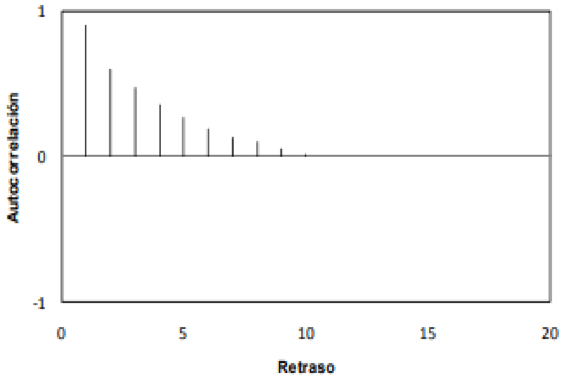
- Si la FAS se corta con rapidez, entonces se debe considerar que los valores de la serie temporal son estacionarios.
- Si la FAS se corta con lentitud extrema, entonces se debe considerar que los valores de la serie temporal son no estacionarios.



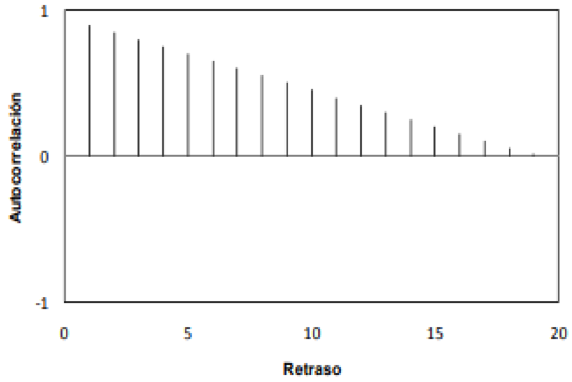
se trunca



cae en forma exponencial



se extingue con rapidez



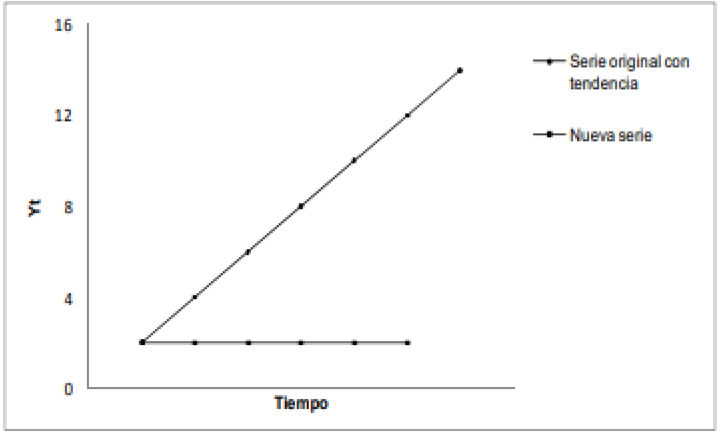
se extingue con lentitud extrema

Transformaciones

Las series de tiempo normalmente no se comportan establemente, es decir no son procesos estocásticos estacionarios. Cuando se presenta esa situación es necesario aplicar una transformación a fin de tener una serie temporal estacionaria.

La estrategia más utilizada para aplicar esta transformación es la de construcción de diferencias, este método consiste, como su nombre lo indica, en obtener diferencias entre los mismos valores de la serie con el fin de remover cualquier patrón de tendencia.

Serie de datos	Primeras diferencias	Nueva serie
2	4-2 = 2	2
4	6-4 = 2	2
6	8-6 = 2	2
8	10-8 = 2	2
10	12-10 = 2	2
12	14-12 = 2	2
14		-

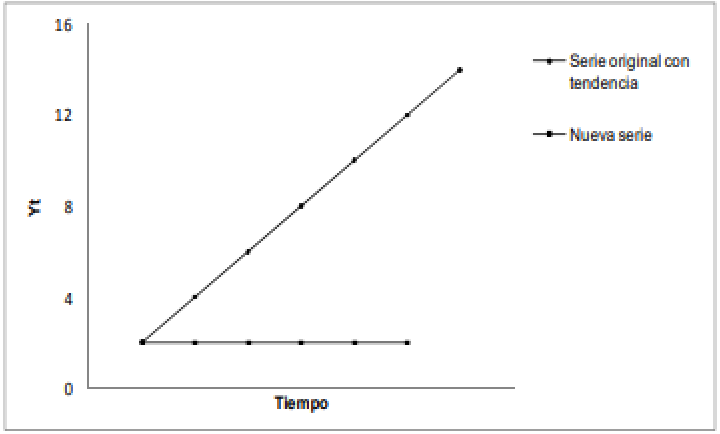


Transformaciones

Las series de tiempo normalmente no se comportan establemente, es decir no son procesos estocásticos estacionarios. Cuando se presenta esa situación es necesario aplicar una transformación a fin de tener una serie temporal estacionaria.

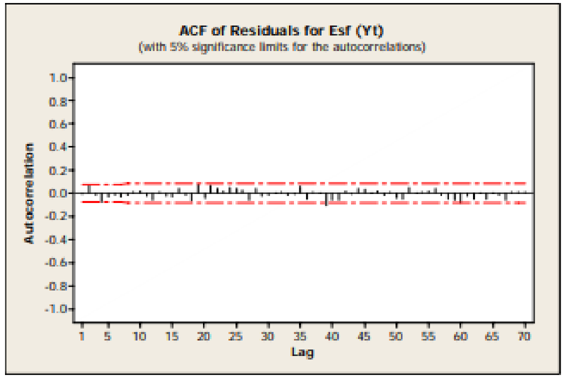
La estrategia más utilizada para aplicar esta transformación es la de construcción de diferencias, este método consiste, como su nombre lo indica, en obtener diferencias entre los mismos valores de la serie con el fin de remover cualquier patrón de tendencia.

Serie de datos	Primeras diferencias	Nueva serie
2	4-2 = 2	2
4	6-4 = 2	2
6	8-6 = 2	2
8	10-8 = 2	2
10	12-10 = 2	2
12	14-12 = 2	2
14		-



Evaluación

Evaluación de los residuales



La función de autocorrelación simple de los residuales es el instrumento que se utiliza para determinar si el modelo es estadísticamente adecuado. Si los residuales muestran estar correlacionados entre sí, significa que existe un patrón que aún no ha sido tomado en cuenta por los términos autorregresivos y/o medias móviles del modelo propuesto, por lo tanto se debe buscar otro modelo cuyos residuales sean completamente aleatorios

Por otro lado no podemos dejar fuera los índices AIC y BIC (Criterio de información de Akaike y criterio de información bayesiano. Dadas dos modelos estimados, el modelo con el menor valor de BIC es el que se prefiere, de la misma forma se evalúa el AIC.

Bibliografía

Machine Learning for Beginners

By Ken Richards, 2017

R Data Analysis Cookbook

By Kuntal Ganguly,2017

Estadística Descriptiva: Series Temporales

by Santiago de la Fuente Fernández