Fundamentos de Análisis de Datos
Introducción a Series de Tiempo

Dra. Andrea Alejandra Rey

Especialización en Ciencia de Datos - ITBA

El tiempo es una variable crucial

Análisis de una colección de datos en un intervalo temporal

Series de tiempo

- Las series de tiempo son modelos apropiados cuando los datos están correlacionados en serie.
- Las series de tiempo ayudan a comprender las causas subyacentes de las tendencias o patrones sistémicos a lo largo del tiempo.
- Cuando los datos se analizan en intervalos constantes, se pueden usar pronósticos de series de tiempo para predecir la probabilidad de eventos futuros.
- La observación de estacionalidad o comportamiento cíclico en los datos, proporciona una mejor comprensión de las variables permitiendo un mejor pronóstico.
- El análisis de series de tiempo también se utiliza para datos no estacionarios, que fluctúan constantemente con el tiempo o se ven afectados por el mismo.
- Los datos de una serie de tiempo se grafican sobre un eje vertical contra el tiempo, representado en el eje horizontal.

Series de tiempo

Ejemplos

- Mediciones de precipitaciones
- Lectura de temperatura
- Monitoreo de ritmo cardíaco (ECG)
- Monitoreo cerebral (EEG)
- Precios de mercado

Aplicaciones

- En minería de datos, reconocimiento de patrones y aprendizaje automático, para agrupación, clasificación y detección de anomalías.
- En procesamiento de señales, ingeniería de control e ingeniería de la comunicación, para detección y estimación de señales.
- En estadística, econometría, finanzas, sismología, meteorología y geofísica, para pronósticos.

Series de tiempo

Modelos

- Debido a que el análisis de series de tiempo incluye muchas categorías o variaciones de datos, el mismo puede generar modelos complejos.
- Es muy poco probable que un modelo específico pueda generalizarse para cualquier muestra.
- Modelos demasiado complejos pueden incurrir en una falta de ajuste que no distingue entre errores aleatorios o verdaderas relaciones entre los datos, generando pronósticos incorrectos.
- Primero debe seleccionarse el tipo de datos relevantes para el caso de estudio, para luego elegir el tipo de análisis y la técnica a emplear.

Series de tiempo

Tipos de modelo

- **Clasificación:** identificación y asignación de categorías a los datos.
- **Ajuste:** estudio de relaciones entre las variables mediante una curva.
- **Análisis descriptivo:** identificación de tendencias, ciclos o variaciones estacionales.
- **Análisis explicatorio:** entendimiento de causa-efecto en los datos.
- **Análisis exploratorio:** resaltado de características principales de la serie de tiempo, usualmente en formato visual.
- **Pronósticos:** predicción de datos futuros en base a tendencias históricas.
- **Análisis de intervención:** estudio de la manera en que un evento puede modificar los datos.
- **Segmentación:** partición de los datos en segmentos propiedades subyacentes de la fuente de información.

Series de tiempo

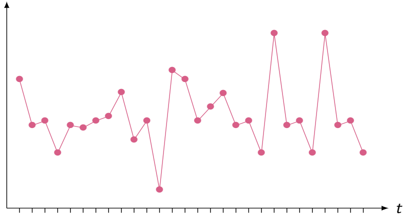
Una serie de tiempo de longitud n consiste en valores muestreados en un conjunto discreto de tiempos $t = 1, 2, \ldots, n$ y que se denota por:

$$X(t) = \{x_t : t = 1, 2, \ldots, n\} = \{x_1, x_2, \ldots, x_n\}.$$

Los datos de series de tiempo pueden clasificarse en los siguientes tipos.

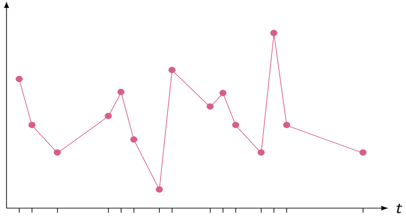
Métricas

Las observaciones se miden en intervalos regulares (igualmente espaciados).



Eventos

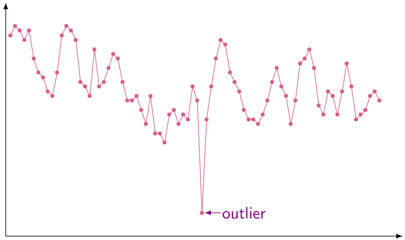
Las observaciones se miden en intervalos irregulares (no igualmente espaciados).



Patrones de una serie de tiempo

Outliers

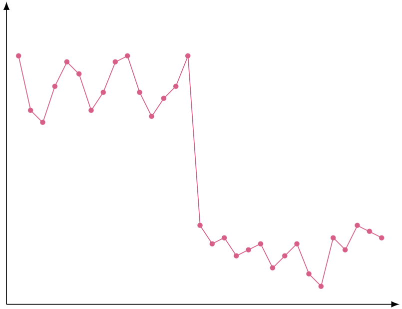
El conjunto de datos podría mostrar valores inusuales, que debieran ser corregidos si los mismos son causa de un error de medición o en el ingreso de datos.



Patrones de una serie de tiempo

Cambios abruptos

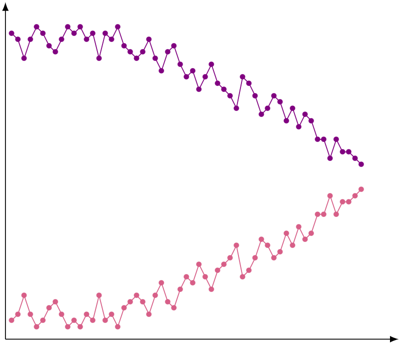
Las causas de cambios drásticos en los datos debieran identificarse.



Patrones de una serie de tiempo

Tendencia

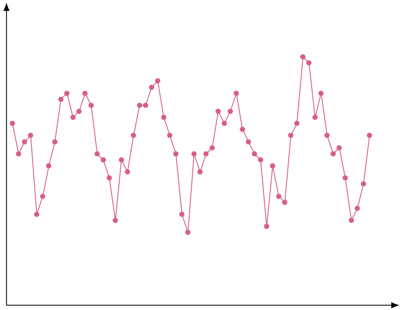
Una tendencia existe cuando se observa un aumento o disminución en los datos a largo plazo, no necesariamente lineal.
Puede ser hacia arriba (en rosa) o hacia abajo (en violeta).



Patrones de una serie de tiempo

Estacionalidad

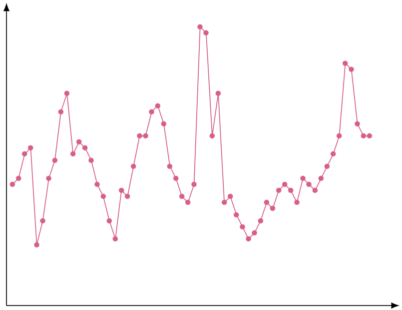
Una repetición de patrones dentro de cualquier período fijo de tiempo se conoce como variación estacional.



Patrones de una serie de tiempo

Ciclo

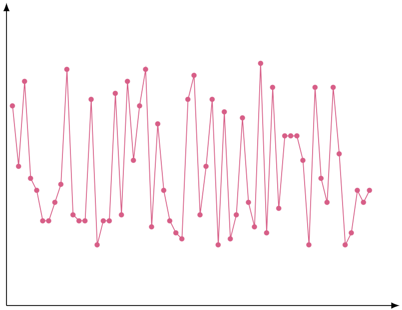
Un ciclo ocurre cuando los datos presentan subidas y bajadas que no tienen una frecuencia fija.



Patrones de una serie de tiempo

Ruido

El ruido representa fluctuaciones en los datos que no pueden ser previstas, predichas, ni explicadas por un modelo.



Series de tiempo estacionarias

Una serie de tiempo se dice **estacionaria** cuando sus características estadísticas no dependen del momento en el que se observa la serie.

Ruido blanco

Series de tiempo estacionarias

Observación

Las series de tiempo con tendencia o estacionalidad no son estacionarias, puesto que estas características afectarán los valores de la serie de tiempo en diferentes momentos.

Importancia

Los modelos se construyen sobre series de tiempo estacionarias. En caso de no cumplirse este supuesto, el primer paso en el análisis será estacionar la serie.

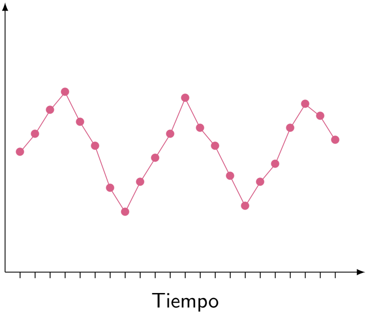
Pregunta

¿Cuáles son las características de una serie de tiempo estacionaria?

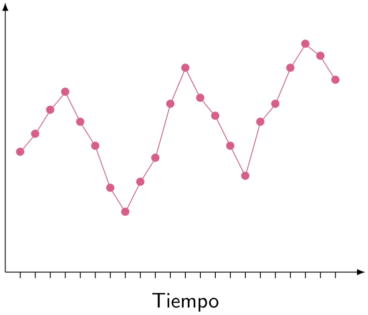
Series de tiempo estacionarias

1. La media no debe depender del tiempo, sino ser constante.

Estacionaria



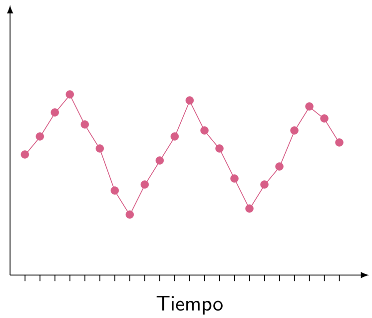
No estacionaria



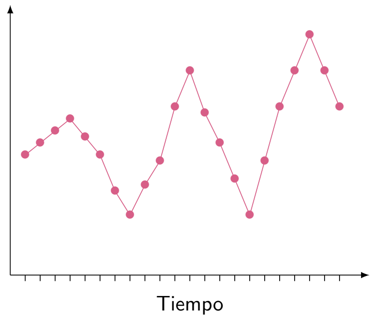
Series de tiempo estacionarias

2. La varianza debe ser finita y no depender del tiempo.

Estacionaria



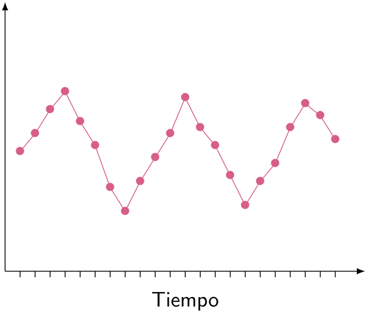
No estacionaria



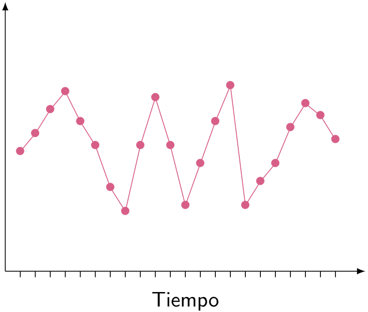
Series de tiempo estacionarias

3 La covarianza de los términos t y $t + m$ no debe depender del tiempo.

Estacionaria



No estacionaria



Pregunta

¿Cómo probamos si una serie de tiempo es estacionaria?

Series de tiempo estacionarias

Random walk

- Imaginemos a una persona moviéndose en un tablero de ajedrez gigante de a un casillero por vez y al azar.
- La siguiente posición de la persona sólo depende de la última posición.
- Imaginemos ahora que estamos en un lugar donde no podemos ver cómo se mueve la persona y queremos predecir la posición que va ocupando a lo largo del tiempo.
- En el tiempo $t = 0$, sabemos exactamente dónde está la persona. En el próximo paso, existen 8 posiciones posibles.
- Es evidente que a medida que la posición de la persona vaya cambiando, nos volveremos más imprecisos.
- Si u_t es el error cometido en el tiempo t con media cero y varianza σ^2, este proceso puede modelarse como:

$$x_t = x_{t-1} + u_t.$$

Series de tiempo estacionarias

Observemos lo siguiente:

$$\begin{aligned}
x_1 &= x_0 + u_1 \\
x_2 &= x_1 + u_2 = x_0 + u_1 + u_2 \\
x_3 &= x_2 + u_3 = x_0 + u_1 + u_2 + u_3 \\
&\vdots \\
x_t &= x_0 + \sum_{i=1}^{t} u_i
\end{aligned}$$

Sabiendo que los errores son independientes, calculamos la varianza:

$$\mathrm{Var}(x_t) = \mathrm{Var}\left(x_0 + \sum_{i=1}^{t} u_i\right) = \underbrace{\mathrm{Var}(x_0)}_{=0} + \sum_{i=1}^{t} \mathrm{Var}(u_i) = \sum_{i=1}^{t} \sigma^2 = t\sigma^2.$$

Series de tiempo estacionarias

Como la varianza de los datos de la serie de tiempo depende del tiempo, hemos probado que el random walk no es estacionaria.

Extendamos el modelo introduciendo el coeficiente ρ:

$$x_t = \rho x_{t-1} + u_t,$$

donde u_t es el término de error con media cero y varianza σ^2.

Si $\rho = 1$, estamos en presencia del **random walk**.

Series de tiempo estacionarias

Observemos lo siguiente:

$$
\begin{aligned}
x_1 &= \rho x_0 + u_1 \\
x_2 &= \rho x_1 + u_2 = \rho^2 x_0 + u_2' \\
x_3 &= \rho x_2 + u_3 = \rho^3 x_0 + u_3' \\
&\vdots \\
x_t &= \rho^t x_0 + u_t'
\end{aligned}
$$

En el caso en que $\rho > 1$, la variable tendrá un comportamiento explosivo al infinito.

Podemos plantear entonces el test de hipótesis:

$$
\begin{aligned}
H_0 &: \rho = 1 \text{ (la serie no es estacionaria)} \\
H_1 &: \rho < 1 \text{ (la serie es estacionaria)}
\end{aligned}
$$

Series de tiempo estacionarias

Restando x_{t-1} a ambos miembros, la ecuación del modelo se modifica de la siguiente manera:

$$x_t - x_{t-1} = (\rho - 1)x_{t-1} + u_t \iff \Delta x_t = \beta x_{t-1} + u_t.$$

Así, el test anterior es equivalente al siguiente test.

Test de Dickey-Fuller

$$H_0 : \beta = 0 \text{ (la serie no es estacionaria)}$$

$$H_1 : \beta < 0 \text{ (la serie es estacionaria)}$$

El estadístico de prueba es:

$$\tau = \frac{\hat{\beta}}{\mathrm{SE}(\hat{\beta})},$$

donde SE indica el error estándar.

Existe una tabulación dada por la tabla de Dickey-Fuller.

Series de tiempo estacionarias

Pregunta

¿Cómo convertimos una serie de tiempo no estacionaria en una que sí sea estacionaria?

Descomposición de una serie de tiempo

En el tiempo t, consideramos:

x_t : valor observado en la serie,

T_t : tendencia,

S_t : efecto estacional,

R_t : residuo o ruido (en general, una secuencia correlacionada de variables aleatorias de media cero).

Modelo de descomposición aditivo

$$x_t = T_t + S_t + R_t$$

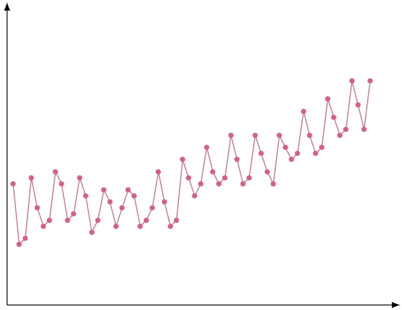
Modelo de descomposición multiplicativo

$$x_t = T_t \cdot S_t \cdot R_t$$

Descomposición de una serie de tiempo

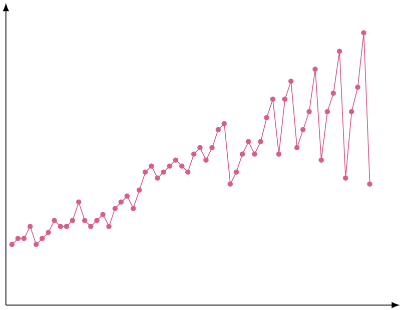
Cambios aditivos

Los valores tienden a crecer a lo largo del tiempo pero la magnitud de la estacionalidad se mantiene.



Cambios multiplicativos

El efecto estacional tiende a crecer a lo largo del tiempo en la medida que la tendencia crece.



Descomposición de una serie de tiempo

Observación

El modelo multiplicativo puede convertirse en aditivo si se aplica logaritmo:

$$\ln(x_t) = \ln(T_t \cdot S_t \cdot R_t) = \ln(T_t) + \ln(S_t) + \ln(R_t).$$

Pregunta

¿Cómo podemos encontrar las componentes de una serie de tiempo?

Estimación de la tendencia

Media móvil

Sea $m = 2k + 1$. La tendencia puede estimarse mediante una media móvil centrada de orden m definida para $t = k + 1, k + 2, \ldots, n - k$, como:

$$\hat{T}_t = \frac{1}{m} \sum_{i=-k}^{k} x_{t+i}.$$

Por ejemplo,

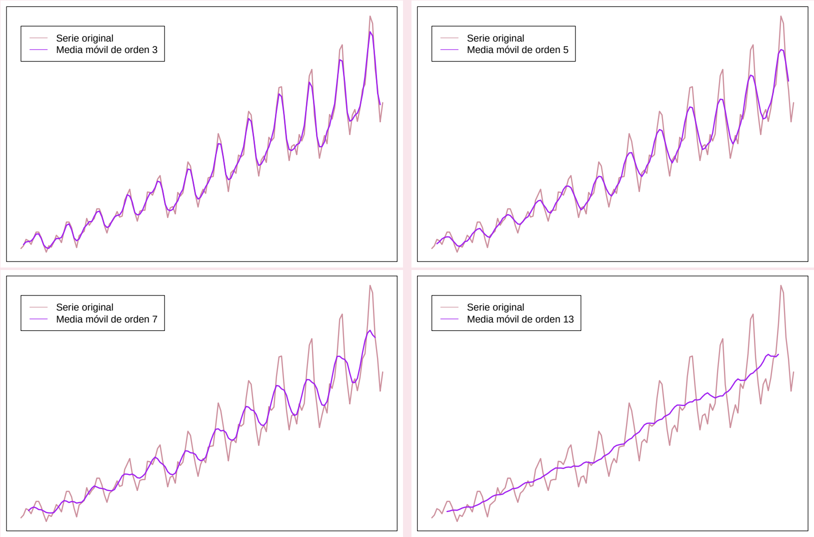
- Si $k = 1$, entonces $m = 2 \cdot 1 + 1 = 3$. Entonces:

$$\hat{T}_t = \frac{x_{t-1} + x_t + x_{t+1}}{3}.$$

- Si $k = 2$, entonces $m = 2 \cdot 2 + 1 = 5$. Entonces:

$$\hat{T}_t = \frac{x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2}}{5}.$$

Estimación de la tendencia



Estimación de la tendencia

Media móvil ponderada

Sea $m = 2k + 1$. La tendencia puede estimarse mediante una media móvil ponderada por los pesos $a_{-k}, \ldots, a_0, \ldots a_k$ definida para $t = k+1, k+2, \ldots, n-k$, como:

$$\hat{T}_t = \sum_{i=-k}^{k} a_i x_{t+i},$$

donde $\sum_{i=-k}^{k} a_i = 1$.

Una gran ventaja de la media móvil ponderada es que brinda una estimación más suave del ciclo de tendencia.

Estimación de la tendencia

Una media móvil de orden impar es simétrica, debido a que la cantidad de observaciones a la izquierda y a la derecha de la observación del medio es la misma.

Pregunta

¿Cómo procedemos si m es par?

Media móvil $2 \times m$

Si $m = 2k$, el ciclo de tendencia puede estimarse mediante una media móvil $2 \times m$ que es una media móvil ponderada de orden $m + 1$ con pesos:

$$a_i = \begin{cases} \dfrac{1}{2m} & \text{si } i = -k \text{ o } i = k, \\ \\ \dfrac{1}{m} & \text{si } i = -k+1, \ldots, 0, \ldots, k-1. \end{cases}$$

Estimación de la tendencia

- La media móvil es un suavizado de la serie de tiempo.
- El suavizado generalmente se realiza con el propósito de distinguir mejor los patrones en la serie de tiempo, cuando se atenúan las irregularidades para ver una señal más clara.
- Si la serie temporal tiene una componente estacional, el m elegido para calcular la media móvil debe ser igual a la frecuencia estacional. Por ejemplo, para series mensuales $m = 12$ y para series diarias $m = 7$.

Fuerza de la tendencia

La fuerza de la tendencia se define de la siguiente manera:

Descomposición aditiva

$$F_T = \max\left\{0, 1 - \frac{\mathrm{Var}(R_t)}{\mathrm{Var}(T_t + R_t)}\right\}$$

Descomposición multiplicativa

$$F_T = \max\left\{0, 1 - \frac{\mathrm{Var}(R_t)}{\mathrm{Var}(T_t \cdot R_t)}\right\}$$

Para datos con tendencias fuertes, los datos ajustados estacionalmente deberían tener mucha más variación que el componente restante. Por lo tanto, el cociente de las varianzas debe ser relativamente pequeño y así F_T tendrá un valor cercano a 1.

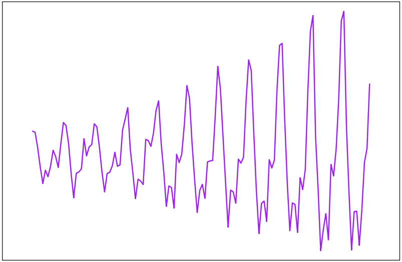
Por el contrario, para datos con poca o ninguna tendencia, las dos varianzas deberían ser aproximadamente iguales y así F_T tendrá un valor cercano a 0.

Eliminación de la tendencia

Descomposición aditiva

Los valores sin tendencia se obtienen como:

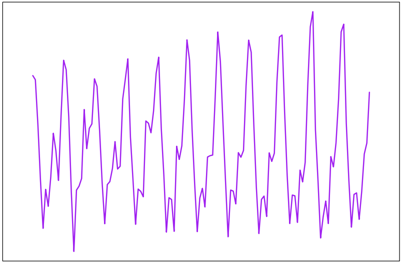
$$\tilde{x}_t = x_t - \hat{T}_t.$$



Descomposición multiplicativa

Los valores sin tendencia se obtienen como:

$$\tilde{x}_t = x_t / \hat{T}_t.$$



Estimación de la estacionalidad

Si la estacionalidad es de período d y la serie de tiempo tiene k períodos, para $t = 1, 2, \ldots, d$, calculamos:

$$\tilde{S}_t = \frac{1}{k} \sum_{i=0}^{k} \tilde{x}_{t+id},$$

donde $\tilde{x}$ representa los valores de la serie de tiempo quitándole la tendencia. Es decir, para estimar la componente estacional de cada temporada, se promedian los valores sin tendencia observados en esa temporada.

Para el resto de los valores de t, procedemos de la siguiente manera. Para cada $i = 1, 2, \ldots, k-1$, consideramos $t = id+1, id+2, \ldots, id+d$, y calculamos:

$$\tilde{S}_t = \tilde{S}_{t-id}.$$

El estimador de la componente de estacionalidad debe satisfacer lo siguiente.

Estimación de la estacionalidad

Descomposición aditiva

$$\sum_{t=1}^{d} \hat{S}_t = 0$$

- Si $\sum_{t=1}^{d} \tilde{S}_t = 0$, entonces $\hat{S}_t = \tilde{S}_t$.
- Si $\sum_{t=1}^{d} \tilde{S}_t = w \neq 0$, entonces $\hat{S}_t = \tilde{S}_t - w/d$.

Descomposición multiplicativa

$$\sum_{t=1}^{d} \hat{S}_t = d$$

- Si $\sum_{t=1}^{d} \tilde{S}_t = d$, entonces $\hat{S}_t = \tilde{S}_t$.
- Si $\sum_{t=1}^{d} \tilde{S}_t = w \neq d$, entonces $\hat{S}_t = \tilde{S}_t - (w-d)/d$.

Fuerza de la estacionalidad

La fuerza de la estacionalidad se define de la siguiente manera:

Descomposición aditiva

$$F_S = \max\left\{0, 1 - \frac{\mathrm{Var}(R_t)}{\mathrm{Var}(S_t + R_t)}\right\}$$

Descomposición multiplicativa

$$F_S = \max\left\{0, 1 - \frac{\mathrm{Var}(R_t)}{\mathrm{Var}(S_t \cdot R_t)}\right\}$$

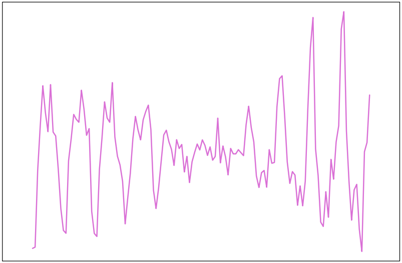
Para datos con una marcada estacionalidad, los datos obtenidos al eliminar la tendencia F_S tendrá un valor cercano a 1, mientras que cuando la estacionalidad sea despreciable, F_S tendrá un valor cercano a 0.

Eliminación de la estacionalidad

Descomposición aditiva

Los valores ajustados por estacionalidad se obtienen como:

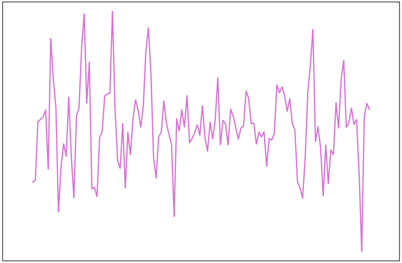
$$\tilde{\tilde{x}}_t = \tilde{x}_t - \hat{S}_t.$$



Descomposición multiplicativa

Los valores ajustados por estacionalidad se obtienen como:

$$\tilde{\tilde{x}}_t = \tilde{x}_t / \hat{S}_t.$$



Estimación de la componente residual

Finalmente, el ruido puede estimarse de la siguiente manera:

- Descomposición aditiva:

$$\hat{R}_t = x_t - \hat{T}_t - \hat{S}_t.$$

- Descomposición multiplicativa:

$$\hat{R}_t = \frac{x_t}{\hat{T}_t \cdot \hat{S}_t}.$$

El objetivo es producir una componente residual estacionaria.

¿Cómo estudiamos si una serie de tiempo tiene tendencia o estacionalidad?

Autocorrelación

Lagging

Lagging es el proceso mediante el cual se desplazan los valores de una serie de tiempo hacia adelante en uno o más pasos de tiempo.
Estos pasos se llaman **lags**.

lag 1

$$\begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \\ \downarrow & \downarrow & \downarrow & \downarrow & \\ x_2 & x_3 & x_4 & x_5 & \end{array}$$

lag 2

$$\begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \\ \downarrow & \downarrow & \downarrow & & \\ x_3 & x_4 & x_5 & & \end{array}$$

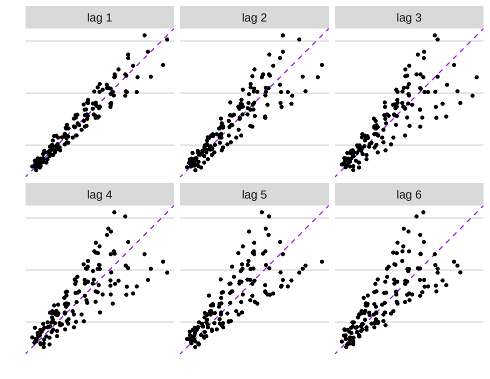
lag 3

$$\begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \\ \downarrow & \downarrow & & & \\ x_4 & x_5 & & & \end{array}$$

Autocorrelación

Lag-plot

Es un gráfico bidimensional en el que se muestran los valores de la serie de tiempo contra sus desplazamientos dados por los distintos lags.



Autocorrelación

Autocorrelación

La autocorrelación mide la relación lineal entre los valores desplazados de una serie.
Por ejemplo:

- r_1 mide la relación lineal entre x_t y x_{t-1},
- r_2 mide la relación lineal entre x_t y x_{t-2}.
- En general, r_h mide la relación lineal entre x_t y x_{t-h} y está definido por:

  $$r_h = \frac{\sum_{t=h+1}^{n}(x_t - \bar{x})(x_{t-h} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2},$$

  donde n es la cantidad de observaciones en la serie de tiempo y $\bar{x}$ el promedio de las mismas.

Autocorrelación

Para el ejemplo anterior:

Lag	1	2	3	4	5	6
Autocorrelación	0.760	0.532	0.337	0.133	-0.047	-0.199

Correlograma

Los coeficientes de autocorrelación se grafican mostrando la función de autocorrelación (ACF).

Autocorrelación

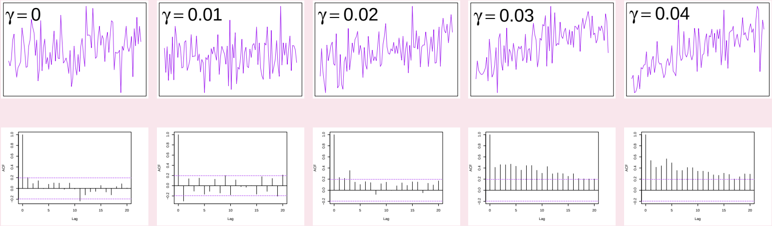
Interpretación de un correlograma

- El eje horizontal corresponde a los lags.
- El eje vertical muestra el coeficiente de correlación para cada lag.
- Aparecen dos líneas punteadas que indican cuándo los coeficientes de autocorrelación son significativamente diferentes de cero.
- La primera barra vertical corresponde al lag nulo, que compara la serie de tiempo consigo misma, por lo que el coeficiente de correlación es 1.
- De no existir autocorrelación, las barras verticales caen abruptamente hacia 0 o al menos dentro de la banda determinada por las líneas punteadas.

Tendencia y autocorrelación

Generamos series de tiempo con tendencia mediante la fórmula:

$$x_t = \gamma t + \epsilon_t.$$

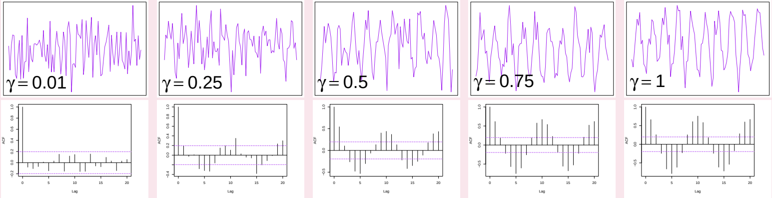


Cuando los datos tienen una tendencia, las autocorrelaciones para pequeños lags tienden a ser grandes y positivas porque las observaciones cercanas en el tiempo también lo son en tamaño. Por lo tanto, la ACF tiende a tener valores positivos que disminuyen lentamente a medida que aumentan los lags.

Estacionalidad y autocorrelación

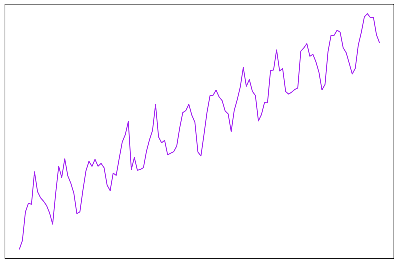
Generamos series de tiempo con estacionalidad mediante la fórmula:

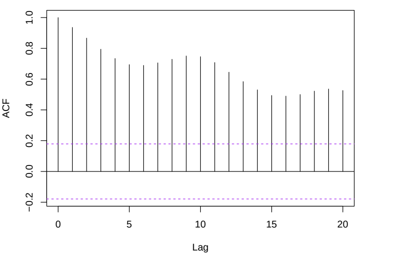
$$x_t = \gamma S_t + \epsilon_t.$$



Cuando los datos son estacionales, las autocorrelaciones son mayores para los lags estacionales (en múltiplos de la frecuencia estacional) que para otros lags. Por lo tanto, la ACF tiene una forma ondulada pronunciada.

Tendencia, estacionalidad y autocorrelación





Cuando los datos son de tendencia y estacionales, se ve una combinación de los efectos anteriores. La lenta disminución de la ACF a medida que aumentan los lags se debe a la tendencia, mientras que la forma ondulada se debe a la estacionalidad.

Ausencia de correlación

Definición

Las series de tiempo que no muestran autocorrelación se llaman **ruido blanco**.

Observación

Las series de tiempo que no muestran autocorrelación son estacionarias.

Autocorrelación parcial

Autocorrelación parcial

La autocorrelación parcial se define como una correlación condicional que mide la correlación entre x_t y x_{t+h} bajo el supuesto que se conocen los valores de las observaciones intermedias. Por ejemplo:

- La autocorrelación parcial de primer orden es la autocorrelación r_1.
- La autocorrelación parcial de segundo orden es:

$$\frac{\mathrm{Cov}(x_t, x_{t-2}|x_{t-1})}{\sqrt{\mathrm{Var}(x_t|x_{t-1})\mathrm{Var}(x_{t-2}|x_{t-1})}}.$$

- La autocorrelación parcial de tercer orden es:

$$\frac{\mathrm{Cov}(x_t, x_{t-3}|x_{t-1}, x_{t-2})}{\sqrt{\mathrm{Var}(x_t|x_{t-1}, x_{t-2})\mathrm{Var}(x_{t-3}|x_{t-1}, x_{t-2})}}.$$

- Y así sucesivamente.

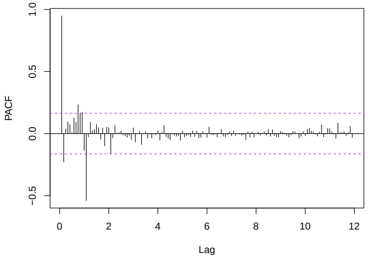
Autocorrelación

Para el ejemplo anterior:

Lag	1	2	3	4	5	6
Autocorrelación Parcial	-0.135	0.048	0.046	0.024	0.019	-0.031

PACF

Al igual que se grafica la ACF, también se puede graficar la función de autocorrelaciones parciales (PACF).



Retomamos con el estudio de la componente residual para estudiar estacionariedad.

Residuos

Pregunta

¿Cómo probamos si hay o no autocorrelación en los residuos?

Test de Durbin-Watson

$$H_0 : r_1 = 0$$
$$H_1 : r_1 \neq 0$$

Estadístico:

$$DW = \frac{\sum_{t=2}^{n}(\epsilon_t - \epsilon_{t-1})^2}{\sum_{t=1}^{n}\epsilon_t^2}.$$

Regla de decisión para la autocorrelación de los residuos:

$$DW < 1 \Longrightarrow \text{existe una autocorrelación positiva significativa,}$$
$$DW > 3 \Longrightarrow \text{existe una autocorrelación negativa significativa,}$$
$$1 \leq DW \leq 3 \Longrightarrow \text{hay evidencia significativa para la no-autocorrelación.}$$

Residuos

Pregunta

¿Cómo probamos si los residuos son independientes?

Test de Ljung-Box

H_0 : Los residuos se distribuyen de forma independiente

H_1 : Los residuos no se distribuyen de forma independiente; es decir, exhiben correlación serial

El estadístico del test es:

$$Q = n(n+2)\sum_{t=k}^{h} \frac{r_k^2}{n-k} \sim \chi^2_{1-\alpha,h},$$

donde h es el tiempo de retraso que se está evaluando.

Ejemplo*

*Base de datos `AirPassengers` de `R`

Pronóstico por series de tiempo

- El pronóstico por series de tiempo se basa en valores históricos asociados a ciertos patrones para predecir la actividad en un futuro.
- El modelado de series temporales implica trabajar con datos basados en el tiempo para obtener información que ayude en la toma de decisiones.

Definición

El **pronóstico** hecho en el tiempo t para un valor futuro en el tiempo $t+h$ se denota por $\hat{x}_{t+h|t}$.

El número de pasos de tiempo hacia el futuro h se denomina **lead time**.

Pronóstico usando descomposición

La descomposición de una serie de tiempo puede usarse como una técnica de pronóstico que separa los datos históricos en diferentes componentes y los utiliza para crear un pronóstico que es más preciso que una simple línea de tendencia.

Dificultades en la práctica

- Un modelo de ajuste lineal para el ciclo de tendencia no suele ser apropiado.
- Si la componente estacional varía con el tiempo, el pronóstico para el futuro basado en el último período de los datos puede no ser adecuado.
- La componente irregular (ruido) debe ser estacionaria, lo cual no siempre sucede.

La descomposición, más que utilizarse para pronosticar, debe considerarse como una herramienta de análisis exploratorio.

Suavizado exponencial simple

Es un método adecuado para pronosticar datos que no presentan claros patrones de tendencia o estacionalidad.

El método naïve consiste en pronosticar los datos futuros como el último valor observado de la serie de tiempo:

$$\hat{x}_{n+h|n} = x_n,$$

asumiendo que la observación más reciente es la única que brinda información sobre el futuro.

El método promedio consiste en pronosticar los datos futuros como el promedio de los valores observados de la serie de tiempo:

$$\hat{x}_{n+h|n} = \frac{1}{n}\sum_{t=1}^{n} x_t,$$

asumiendo que todas las observaciones brindan la misma información sobre el futuro.

Suavizado exponencial simple

El método de suavizado exponencial simple consiste en pronosticar los datos futuros asignando peso a los valores observados de la serie de tiempo. Dado un parámetro de suavizado $\gamma \in [0,1]$, el algoritmo es:

1. Inicialización: $\hat{x}_1 = \ell$.
2. Actualización: $\hat{x}_t = \gamma x_t + (1-\gamma)\hat{x}_{t-1}$ para $t = 2, 3, \ldots, n$.
3. Pronóstico: $\hat{x}_{n+h|n} = \hat{x}_n$ para $h = 1, 2, \ldots$.

Equivalentemente:

$$\hat{x}_{n+h|n} = \sum_{i=0}^{n-1} \gamma(1-\gamma)^i x_{n-i} + (1-\gamma)^n \ell.$$

Suavizado exponencial simple

Pregunta

¿Cómo elegimos el valor inicial y el parámetro de suavizado?

Para $t = 1, 2, \ldots, n$, los residuos están dados por:

$$e_t = x_t - \hat{x}_{t|t-1}.$$

Al igual que en la estimación de los coeficientes de regresión, podemos elegir los valores de ℓ y γ como aquellos que minimizan la suma de los cuadrados de los errores:

$$\text{SSE} = \sum_{t=1}^{n} e_t^2.$$

Modelos de autorregresión

Se basan en la regresión de una variable contra sí misma.

Un modelo de autorregresión de orden p, denotado por AR(p), está dado por:

$$x_t = c + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t,$$

donde ε_t es ruido blanco.

Notar que esta ecuación representa un modelo de regresión lineal múltiple donde las variables predictoras son los valores dados por los lags.

Observación

En el caso del modelo AR(1):

- Si $\phi_1 = 0$, estamos en presencia de ruido blanco.
- Si $c = 0$ y $\phi_1 = 1$, estamos en presencia un random walk.

Modelos de autorregresión

Restricción

La aplicación de los modelos de autorregresión se restringe a datos estacionarios.

Esto implica restricciones para los parámetros.
Explícitamente:

- para el modelo AR(1): $-1 < \phi_1 < 1$.
- para el modelo AR(2): $-1 < \phi_2 < 1$, $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$.

Para $p \geq 3$, las expresiones de las restricciones son bastante más complicadas.

Modelos de media móvil

Se basan en una regresión que utiliza errores de pronóstico previos.

Un modelo de media móvil de orden q, denotado por MA(q), está dado por:

$$x_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q},$$

donde ε_t es ruido blanco.

Notar que los errores no son observados por lo que **no** tenemos una regresión en el sentido usual.

Restricción

La aplicación de los modelos de media móvil se restringe a datos estacionarios.

Dualidad entre AR Y MA

Supongamos que tenemos el modelo AR(1):

$$\begin{aligned} x_t &= c + \phi_1 x_{t-1} + \varepsilon_t \\ &= c + \phi_1(\phi_1 x_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= c + \phi_1^2 x_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\ &= c + \phi_1^2(\phi_1 x_{t-3} + \varepsilon_{t-2}) + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\ &= c + \phi_1^3 x_{t-3} + \phi_1^2 \varepsilon_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t . \end{aligned}$$

Como $|\phi_1| < 1$, el valor de ϕ_1^i disminuye considerablemente en la medida que i es muy grande. Entonces:

$$x_t = c + \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \cdots ,$$

que es un modelo MA($+\infty$) siempre que el modelo sea **inversible**; es decir, $\sum_{i=1}^{+\infty} |\phi_1|^i < 1$.

Dualidad entre AR y MA

Con un poco más de trabajo, y bajo el supuesto de que el modelo sea inversible, se puede probar lo siguiente:

Un modelo AR(p) puede escribirse como un modelo MA($+\infty$).

De manera análoga, y bajo el supuesto de que el modelo sea inversible, también se puede probar que:

Un modelo MA(q) puede escribirse como un modelo AR($+\infty$).

ARMA

Un modelo ARMA (AutoRegressive Moving Average) es una combinación de los modelos AR(p) y MA(q).

El modelo está dado por la ecuación:

$$x_t = c + \varepsilon_t + \sum_{i=1}^{p} \phi_i x_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i},$$

donde:

- ε_t es ruido blanco,
- $\phi_1, \phi_2, \ldots, \phi_p$ son los coeficientes de un modelo AR(p),
- $\theta_1, \theta_2, \ldots, \theta_q$ son los coeficientes de un modelo MA(q).

En este caso notamos al modelo como ARMA(p,q).

ARMA

Observación

- Un modelo AR(p) es un modelo ARMA(p,0).
- Un modelo MA(q) es un modelo ARMA(0,q).

Pregunta

¿Cómo determinamos los valores de p y q?

Óptimo p para AR

Elegir el punto de corte del gráfico de la PACF; es decir, el lag a partir del cual los valores se acercan mucho a cero.

Óptimo q para MA

Elegir el punto de corte del gráfico de la ACF; es decir, el lag a partir del cual los valores se acercan mucho a cero.

Estacionariedad en ARMA

Observación

Debido a las restricciones para los modelos AR y MA, el modelo ARMA es un modelo estacionario.

Pregunta

¿Cómo podemos analizar la estacionariedad de una serie de tiempo?

Estacionariedad en ARMA

Test de Phillips-Perron (PP)

$$H_0 : \text{La serie no es estacionaria}$$
$$H_1 : \text{La serie es estacionaria}$$

Test de Kwiatkowski-Phillips-Schmidt-Shin (KPSS)

$$H_0 : \text{La serie es estacionaria en tendencia}$$
$$H_1 : \text{La serie no es estacionaria}$$

Pregunta

¿Cómo podemos trabajar con datos de series de tiempo no estacionarias?

Transformación Box-Cox

Si en una serie de tiempo, la varianza cambia con el tiempo, el proceso no es estacionario.

La transformación Box-Cox es una herramienta para estacionar los datos.

Definición

Dado un parámetro λ, la **trasformación Box-Cox** está dada por:

$$y_t = \begin{cases} \ln(x_t) & \text{si } \lambda = 0, \\ (x_t^\lambda - 1)/\lambda & \text{si } \lambda \neq 0. \end{cases}$$

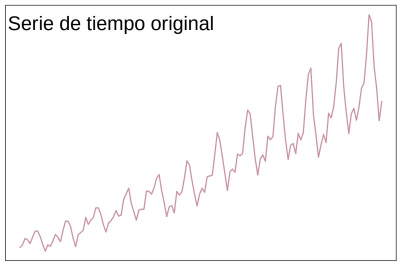
Transformación Box-Cox

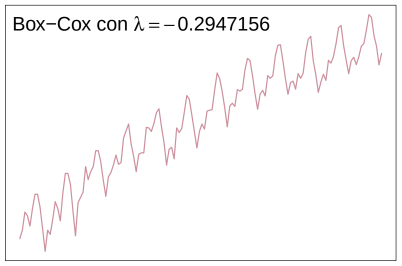
Observación

- La transformación Box-Cox está definida para series de tiempo con valores positivos. Sin embargo, existe una variante para el caso de valores negativos.
- El valor óptimo de λ es aquel que mejor aproxima la variable respuesta a una distribución Normal.

Transformación Box-Cox

Veamos en un ejemplo, cómo la transformación Box-Cox estabiliza la varianza.





Diferenciación

La diferenciación puede ayudar a estabilizar la media de una serie de tiempo eliminando (o reduciendo) la tendencia y la estacionalidad.

Diferenciación

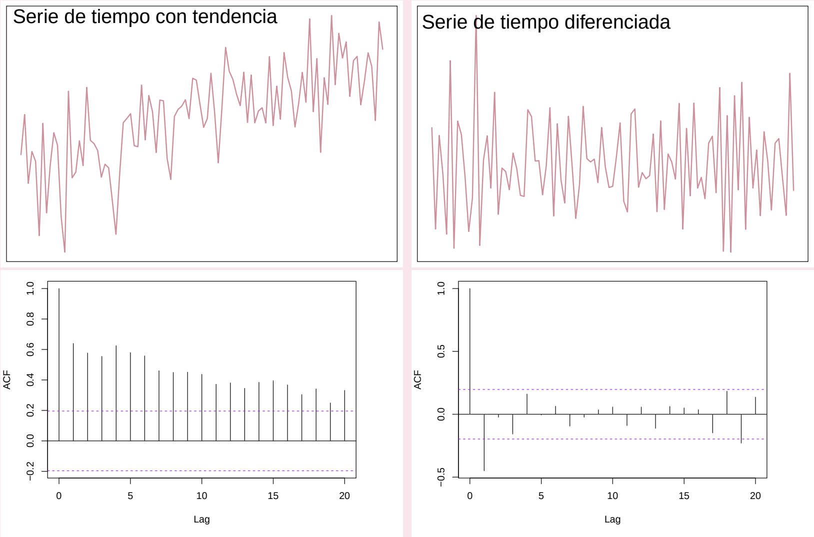
La serie de tiempo luego de aplicar la diferenciación, mide el cambio entre dos observaciones consecutivas y puede escribirse como:

$$x'_t = x_t - x_{t-1},$$

para $t = 2, 3, \ldots, n$.

Dada una serie de tiempo que presenta una fuerte tendencia, veamos cómo la diferenciación disminuye esta tendencia.

Diferenciación



Diferenciación

Puede ocurrir que la serie diferenciada siga mostrando tendencia.
En este caso, volvemos a aplicar una diferenciación para obtener:

$$x_t^{''} = x_t' - x_{t-1}',$$

para $t = 2, 3, \ldots, n-1$.

Equivalentemente:

$$x_t^{''} = x_t' - x_{t-1}' = x_t - x_{t-1} - (x_{t-1} - x_{t-2}) = x_t - 2x_{t-1} + x_{t-2}.$$

Este procedimiento puede replicarse d veces, dando lugar a d diferenciaciones.

ARIMA

Un modelo ARIMA (AutoRegressive Integrated Moving Average) es una combinación de diferenciación con los modelos de autorregresión y media móvil.

El modelo está dado por la ecuación:

$$x'_t = c + \varepsilon_t + \sum_{i=1}^{p} \phi_i x'_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i},$$

donde:

- x'_t es la serie de tiempo diferenciada (permitiendo repeticiones),
- ε_t es ruido blanco,
- $\phi_1, \phi_2, \dots, \phi_p$ son los coeficientes de un modelo AR(p),
- $\theta_1, \theta_2, \dots, \theta_q$ son los coeficientes de un modelo MA(q).

ARIMA

En este caso notamos al modelo como ARIMA(p,d,q), donde:

- p es el número de términos del modelo de autorregresión,
- d es el número de diferencias necesarias para estacionar los datos,
- q es el número de errores con lag del pronóstico.

Observación

- El ruido blanco es un modelo ARIMA(0,0,0).
- El random walk es un modelo ARIMA(0,1,0) sin constante ($c = 0$).
- Un modelo de autorregresión es un modelo ARIMA(p,0,0).
- Un modelo de media móvil es un modelo ARIMA(0,0,q).
- Un modelo ARMA(p,q) es un modelo ARIMA(p,0,q).

Rendimiento del modelo

Criterio de información de Akaike

Sea L la función de verosimilitud de los datos, el criterio de información de Akaike puede escribirse como:

$$\mathrm{AIC} = -2\log(L) + 2(p + q + k + 1),$$

donde $k = 1$ si $c \neq 0$ y $k = 0$ si $c = 0$.

Criterio de información de Akaike corregido

Para un modelo ARIMA, el criterio de información de Akaike corregido puede escribirse como:

$$\mathrm{AIC}^* = \mathrm{AIC} + \frac{2(p + q + k + 1)(p + q + k + 2)}{n - p - q - k - 2},$$

donde n es el largo de la serie de tiempo.

Rendimiento del modelo

Error de porcentaje medio absoluto

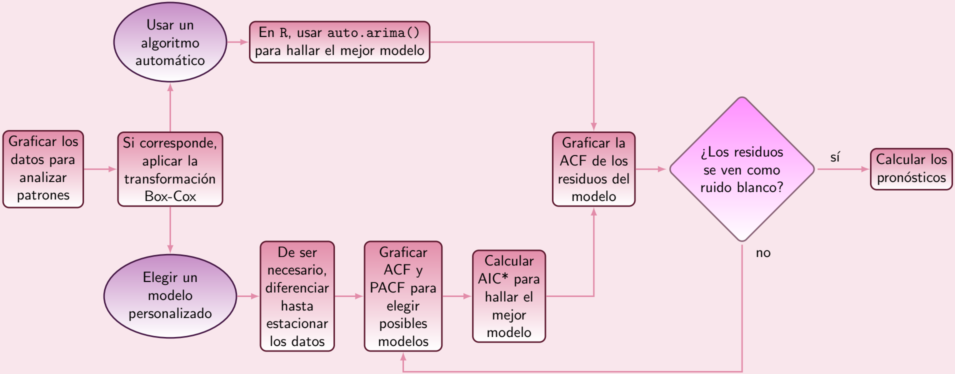
El error de porcentaje medio absoluto (MAPE) es una medida porcentual del error relativo.

Dada una serie de tiempo $\{x_1, x_2, \ldots, x_n\}$ con pronósticos $\{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n\}$, se tiene que:

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^{n} \frac{|x_t - \hat{x}_t|}{|x_t|}.$$

Al comparar el rendimiento de dos modelos, será mejor aquel que tenga un valor menor de AIC* y de MAPE.

Vamos a resumir los pasos del proceso de pronóstico.



Graficar los datos para analizar patrones
Si corresponde, aplicar la transformación Box-Cox
Usar un algoritmo automático
En R, usar auto.arima() para hallar el mejor modelo
Elegir un modelo personalizado
De ser necesario, diferenciar hasta estacionar los datos
Graficar ACF y PACF para elegir posibles modelos
Calcular AIC* para hallar el mejor modelo
Graficar la ACF de los residuos del modelo
¿Los residuos se ven como ruido blanco?
sí
no
Calcular los pronósticos

Ejemplo*

*Base de datos `goog` del paquete `fpp2` de `R`

Modelos de series de tiempo no heterocedáticos

- Los modelos ARIMA se basan en el supuesto de la varianza constante.
- Los modelos que rompen este supuesto son condicionalmente heterocedásticos; es decir, asumen que la varianza no es constante en el tiempo.
- Las series financieras no son homocedásticas, debido a que tienen clusters de volatilidad.
- Si la varianza no es constante, es una varianza condicional que se explica a partir de los errores pasados.
- Los modelos autorregresivos con heterocedasticidad condicional (ARCH) establecen que los residuos se pueden escribir en términos de una componente estocástica z_t y de una desviación estándar σ_t que depende del tiempo; es decir,

$$\varepsilon_t = \sigma_t z_t.$$

Heterocedasticidad condicional autorregresiva

Modelo autorregresivo con heterocedasticidad condicional

La ecuación para el modelo ARCH de orden q, denotado por ARCH(q), es:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2,$$

donde $\alpha_i > 0$ para $i = 0, 1, \ldots, q$.

Modelo autorregresivo con heterocedasticidad condicional generalizado

La ecuación para el modelo GARCH de órdenes p y q, denotado por GARCH(p,q), es:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2.$$

Heterocedasticidad condicional autorregresiva

GARCH asimétrico no lineal

La ecuación para el modelo NAGARCH es:

$$\sigma_t^2 = \omega + \alpha(\varepsilon_{t-1} - \theta\sigma_{t-1})^2 + \beta\sigma_{t-1}^2,$$

donde $\alpha, \beta \geq 0$, $\omega > 0$ y $\alpha(1+\theta^2) + \beta < 1$.

GARCH integrado

La ecuación para el modelo IGARCH de órdenes p y q es:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2,$$

donde $\sum_{i=1}^{p} \beta_i + \sum_{i=1}^{q} \alpha_i = 1$.

Heterocedasticidad condicional autorregresiva

GARCH exponencial

La ecuación para el modelo EGARCH es:

$$\ln(\sigma_t^2) = \omega + \sum_{i=1}^{q} \alpha_i g(Z_{t-i}) + \sum_{i=1}^{p} \beta_i \ln(\sigma_{t-i}^2),$$

donde $Z_t \sim \mathcal{N}(0,1)$ y $g(Z_t) = \theta Z_t + \lambda[|Z_t| - \mathrm{E}(|Z_t|)]$.

Cuadratic GARCH

La ecuación para el modelo QGARCH correspondiente a GARCH(1,1) es:

$$\sigma_t^2 = K + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \phi\varepsilon_{t-1}.$$

Heterocedasticidad condicional autorregresiva

Glosten-Jagannathan-Runkle GARCH

La ecuación para el modelo GJR-EGARCH es:

$$\sigma_t^2 = K + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \phi\varepsilon_{t-1}^2 I_{t-1},$$

donde $I_{t-1} = 0$ si $\varepsilon_{t-1} \geq 0$ e $I_{t-1} = 1$ si $\varepsilon_{t-1} < 0$.

GARCH umbralizado

La ecuación para el modelo TGARCH es:

$$\sigma_t^2 = K + \delta\sigma_{t-1} + \alpha_1^+\varepsilon_{t-1}^+ + \alpha_1^-\varepsilon_{t-1}^-,$$

donde $\varepsilon_{t-1}^+ = \varepsilon_{t-1}$ si $\varepsilon_{t-1} > 0$, $\varepsilon_{t-1}^+ = 0$ si $\varepsilon_{t-1} \leq 0$, $\varepsilon_{t-1}^- = \varepsilon_{t-1}$ si $\varepsilon_{t-1} \leq 0$ y $\varepsilon_{t-1}^- = 0$ si $\varepsilon_{t-1} > 0$.

Heterocedasticidad condicional autorregresiva

Si se desea aplicar estos modelos para ajustar, pronosticar, simular, inferir y graficar series de tiempo, se puede utilizar el paquete `rugarch` de R.

La lista de modelos continua. Sin embargo...


"That's all Folks!"