Series de tiempo como modelos lineales

Fernanda Mora
7 noviembre, 2016

Diplomado de Minería de Datos
para la toma de decisiones

Contenido

- Introducción
- Estadística básica
  - Ejercicio práctico 1
- Modelos AR y MA
  - Ejercicio práctico 2
- Framework para modelos ARIMA

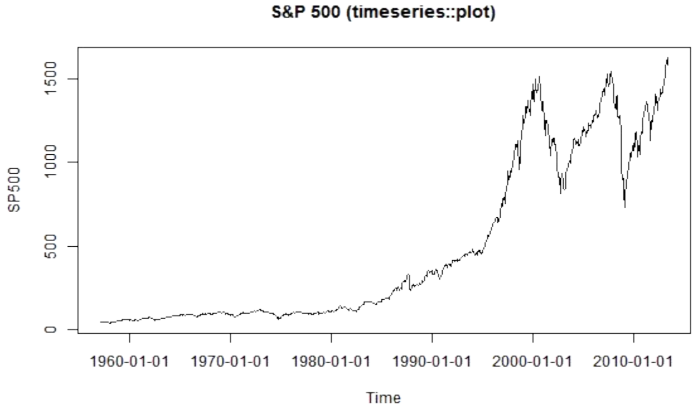
$y = f(x1, \ldots, xn)$

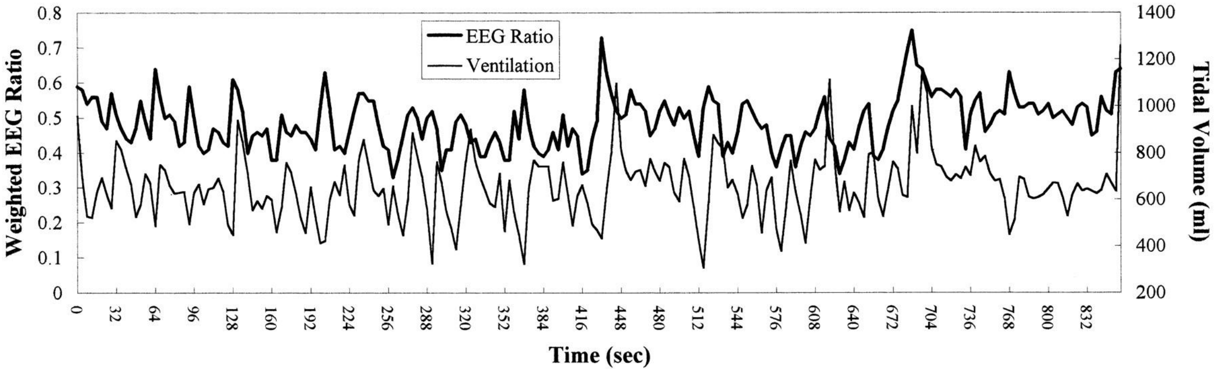
Introducción

Preliminares

- Una serie temporal es una sucesión de observaciones indexadas y ordenadas por el tiempo
- Los datos pueden estar espaciados a intervalos iguales (más frecuentemente) o desiguales
- Ejemplos son: precio de acciones a lo largo del tiempo, medición signos vitales cada hora, PIB mensual.
- Es difícil imaginar una rama en la que no aparezcan datos que puedan ser considerados como series temporales: Medicina, Meteorología, Economía, Finanzas, Logística.

Preliminares





Preliminares

- El análisis de series de tiempo comprende métodos para analizar series con el fin de extraer información relevante
- El pronóstico de series de tiempo es el uso de un modelo para predecir valores futuros basados en datos previamente observados. Es en donde se encuentra la mayor actividad de series de tiempo.
- Pueden ser usadas en variables de razón/continuas o bien discretas ó categóricas.
- En una serie de tiempo cada observación tiene una dependencia temporal, el orden importa. En los otros datasets que hemos visto el orden no importaba.

Preliminares

- Los métodos para análisis de series de tiempo se pueden dividir en: con respecto a la frecuencia o con respecto al tiempo.
- Frecuencia: Análisis espectral y análisis de wavelets.
- Tiempo: Auto-correlación y correlación cruzada.
- También se pueden dividir en paramétricos y no paramétricos.
- No Paramétricos: no asumen ninguna distribución subyacente, ni parámetros, estiman sin hacer supuestos obre los datos.

Preliminares

- Paramétricos: asumen que el proceso estocástico estacionario tiene cierta distribución subyacente que puede ser descrita usando algunos parámetros (ej. modelos autorregresivos o de promedios móviles).
- También pueden dividirse en lineales o no lineales, y en univariados o multivariados.
- En esta clase estudiaremos métodos paramétricos usando modelos lineales.
- En la siguiente clase estudiaremos métodos no paramétricos usando modelos no lineales.

Estadística básica

Análisis de series de tiempo

Proceso:

1. Visualizar la serie
2. Hacerla estacionaria
3. Graficar ACF/PACF y encontrar parámetros
4. Construir el modelo ARIMA
5. Predecir

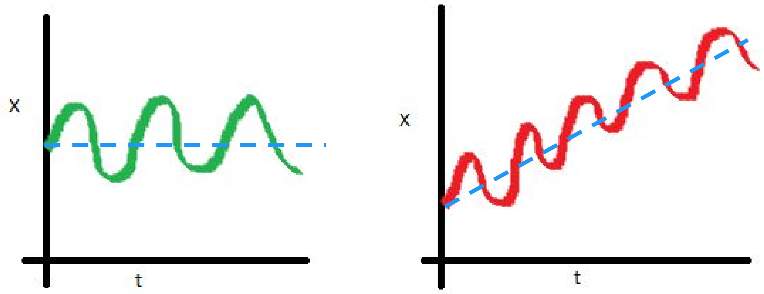
Series estacionarias

Estadística descriptiva

Decimos que una serie es estacionaria cuando “es estable a lo largo del tiempo”

1. Media constante a lo largo del tiempo. Es decir, la media NO debe estar en función del tiempo.

Gráficamente: los valores de la serie tienden a oscilar alrededor de una media constante.



Stationary series   Non-Stationary series

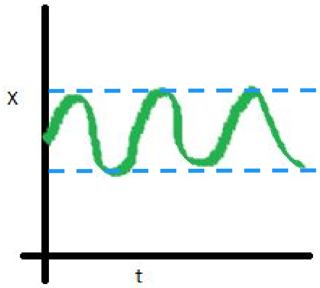
Series estacionarias

Estadística descriptiva

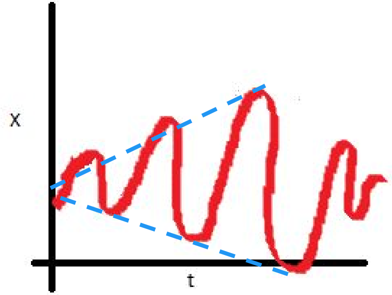
Decimos que una serie es estacionaria cuando “es estable a lo largo del tiempo”

2. Varianza constante a lo largo del tiempo. Es decir, la varianza NO debe estar en función del tiempo.

Gráficamente: la variabilidad con respecto a esa media también permanece constante en el tiempo.



Stationary series



Non-Stationary series

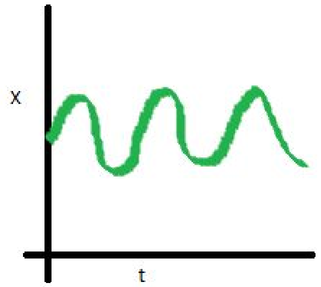
A esta propiedad también se le conoce como **homocedasticidad y también la vimos en regresión lineal (para los errores)**

Series estacionarias

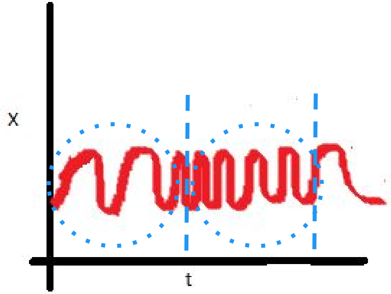
Estadística descriptiva

Decimos que una serie es estacionaria cuando "es estable a lo largo del tiempo"

3. Covarianza constante. Covarianza del término i-ésimo y del término (i+m) NO deben ser una función del tiempo.



Stationary series



Non-Stationary series

Procesos estocásticos estacionarios

- Un proceso estocástico se describe como una sucesión de variables aleatorias (datos) que evolucionan en el tiempo.
- Las series temporales se definen como un caso particular de los procesos estocásticos.
- Un proceso estocástico se dice que es estacionario si su media y su varianza son constantes en el tiempo, y si el valor de la covarianza entre dos periodos depende solamente de la distancia o rezago entre estos y no del tiempo en el cual se ha calculado la covarianza:

Si una serie de tiempo es **estacionaria** entonces su media, su varianza y son **invariantes en el tiempo.**

Sea X_t una serie de tiempo entonces con estas propiedades:

$$Media \quad E(X_t) = E(X_{t+k}) = \mu$$

$$Varianza \quad V(X_t) = V(X_{t+k}) = \sigma^2$$

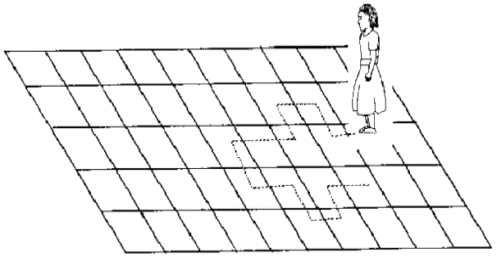
$$Covarianza \quad \gamma_k = E[(X_t - \mu)(X_{t+k} - \mu)]$$

Procesos estocásticos estacionarios

- ¿Para qué queremos esto?
- Necesitamos series estacionarias para poder construir estos modelos.
- ¿Qué pasa cuando una serie no es estacionaria?
- ¡La hacemos (o tratamos de hacerla) estacionaria!
- Hay varios métodos para lograr esto.

Ejemplo

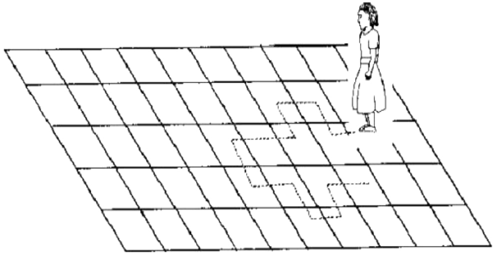
Caminata aleatoria



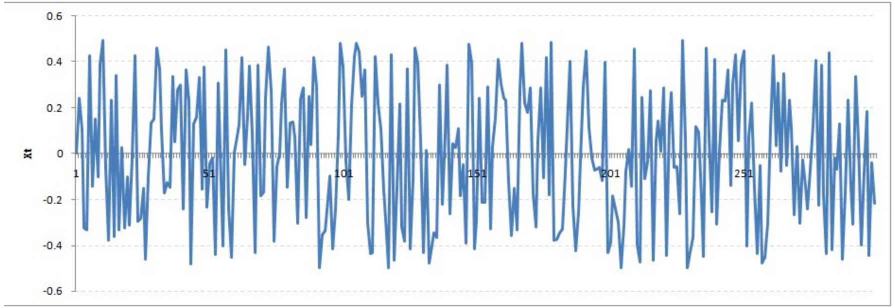
- La caminata aleatoria es el ejemplo más básico de proceso estocástico: X(t) = X(t-1) + e(t), e(t) es ruido blanco. $\varepsilon_t \sim N(0,\sigma^2)$ $cov\left(\varepsilon_{t_i}, \varepsilon_{t_j}\right) = 0 \;\; \forall t_i \neq t_j$
- ¿Cómo predecir la posición de la niña con respecto al tiempo, i.e. X(t)? Cada vez es más difícil.
- X(t) = X(0) + [e(1)+e(2)+e(3)+...+e(t)]
- ¿La serie es estacionaria?
  - E(X(t))=X(0) -> constante
  - Var(X(t))= t * Var(e(t)) = t*constante -> Depende del tiempo
  - La caminata aleatoria NO es estacionaria

Ejemplo

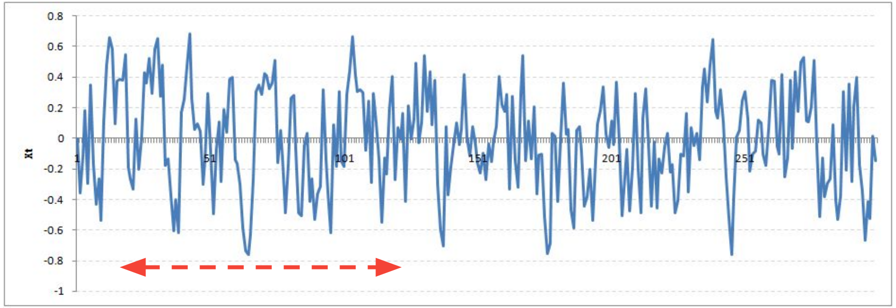
Caminata aleatoria



- ¿Cómo hacerla estacionaria?
- Sea Rho un coeficiente y haremos X(t) = Rho * X(t-1) + Er(t)
- La idea es ver si con esto podemos hacer la serie estacionaria



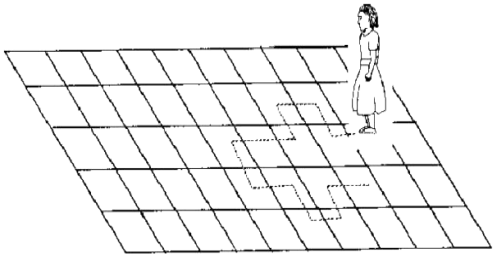
Rho = 0



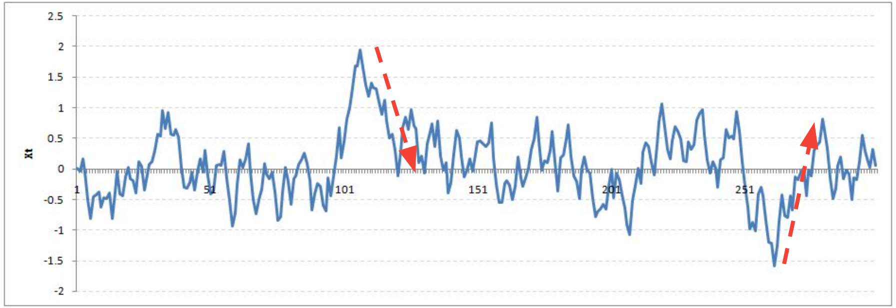
Rho = 0.5

Ejemplo

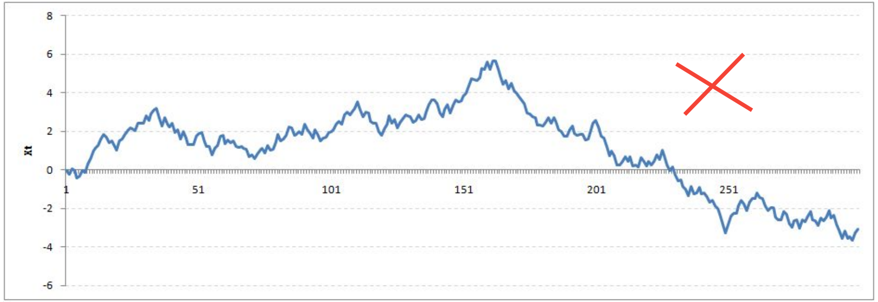
Caminata aleatoria



- ¿Cómo hacerla estacionaria?
- Sea Rho un coeficiente y haremos X(t) = Rho * X(t-1) + Er(t)
- ¿Cómo escogemos la Rho adecuada?



Rho = 0.9



Rho = 0.5

Ejemplo

Test de estacionariedad Dickey Fuller

- Lo anterior se puede enunciar formalmente en una prueba de hipótesis.
- X(t) = Rho * X(t-1) + e(t)   => X(t) - X(t-1) = (Rho - 1) X(t - 1) + e(t)
- Tenemos que probar si Rho – 1 es significativamente diferente de cero o no.
- Si la hipótesis nula se rechaza entonces tenemos una serie de tiempo estacionaria.
- Estos procesos son claves para tener un modelo correcto.

Estacionariedad

Métodos

1. Quitar la tendencia: **Removemos la tendencia de la serie de tiempo.**

   Por ejemplo, si la ecuación es x(t) = (media + trend * t) + error, sólo removemos la parte del paréntesis y construimos un modelo para lo demás.

2.

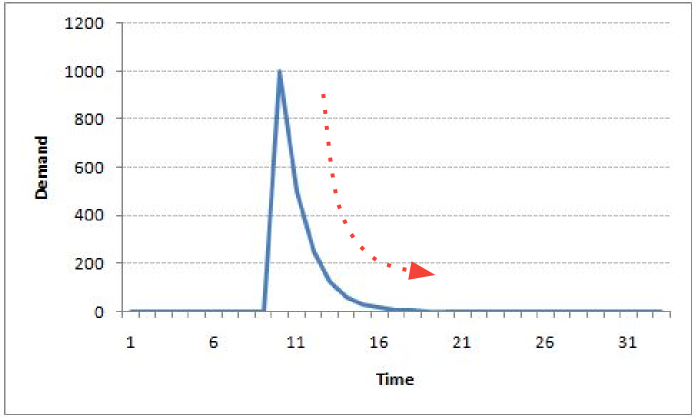
3.

Modelos AR y MA

Ejemplo

Modelo AR (autoregresivo)

- Supongamos que la demanda es x(t) = alpha * x(t – 1) + error (t)
- Esto se conoce como modelo AR (1) (autorregresivo de orden 1)
- Buscamos alpha que minimice el error.
- AR no siempre es estacionario (AR(1), si |alpha|>1 entonces no)



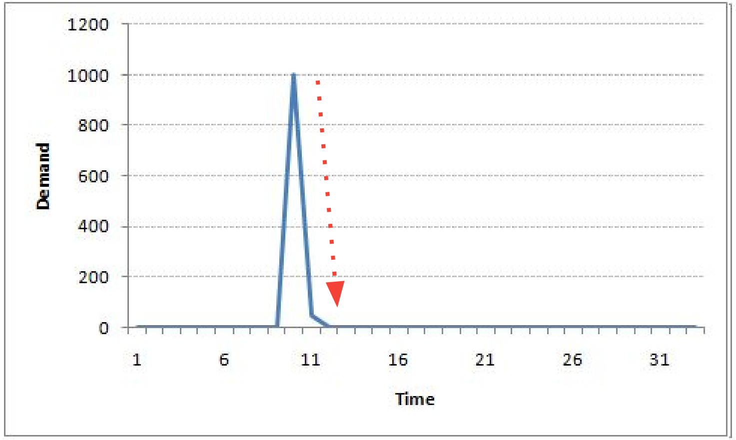
Propiedad de **inercia** del modelo AR:
Cualquier shock se **diluirá a través del tiempo**

AR(p): $$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

Ejemplo

Modelo MA (de promedios móviles)

- x(t) = beta * error(t-1) + error (t)
- Esto se conoce como modelo MA (1) (moving average de orden 1)
- Ejemplo: demanda de un producto nuevo



Propiedad de **recuperación inmediata del modelo MA**. El efecto se diluye de inmediato.

MA(q): $X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$

Diferencias entre los modelos AR y MA

- La diferencia principal es la correlación de la serie de tiempo en diferentes instantes de tiempo
- Corr(x(t), x(t-n)) = 0 para MA con n>orden
- Corr(x(t), x(t-n) decrece conforme n crece para AR
- La gráfica de correlación puede decirnos el orden del modelo MA.
- Una vez que el proceso es estacionario, hay que investigar si es un proceso AR o MA.
- Posteriormente hay que investigar el orden del proceso.

Análisis de series de tiempo

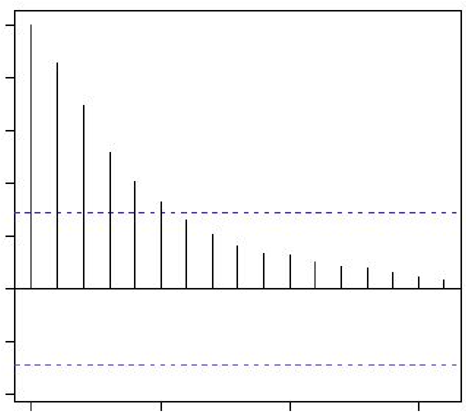
¿Qué modelo es: AR o MA?

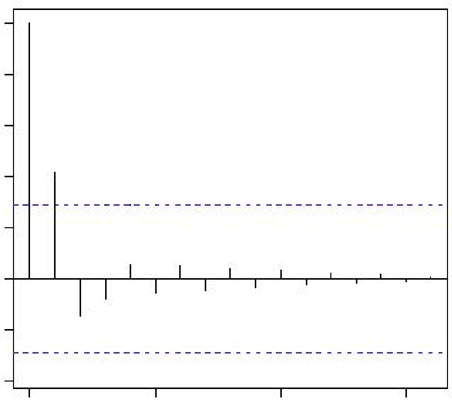
- Usamos gráfica de correlación total, también llamada función de autocorrelación ó ACF)
- ACF grafica la correlación total entre diferentes lags: estamos interesados en la correlación de x(t) con x(t-1), x(t-2),... etc.
- En un modelo MA de lag n, corr(x(t), x(t-n-1))=0. Entonces es fácil encontrar el lag para estos modelos.
- Para un modelo AR la correlación irá bajando sin cortar.
- ¿Qué hacemos con estos modelos?
- Usamos la función de correlación parcial de cada lag, y cortará después del orden del proceso AR.

Análisis de series de tiempo

¿Qué modelo es: AR o MA?

- Por ejemplo, si tenemos un modelo AR(1), si excluimos el efecto del 1er lag (i.e. x(t-1)), el segundo lag (x(t-2)) es independiente de x(t).
- Entonces esta función de correlación parcial (PACF) caerá estrepitosamente después del 1er lag.



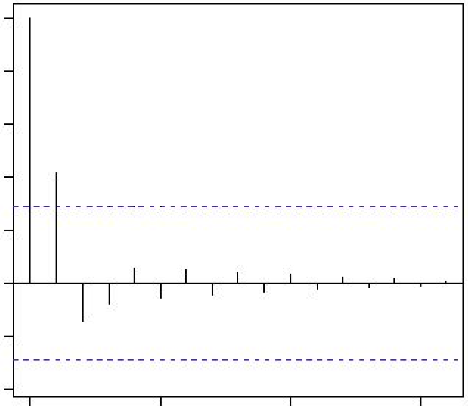


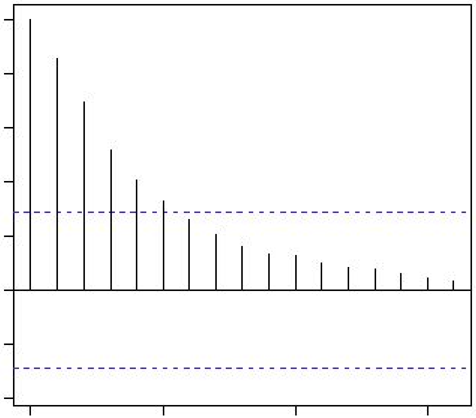
La línea azul muestra valores significativamente distintos de cero. Claramente la gráfica tiene un corte en la función de PACF después del segundo lag. Esto significa que es a lo más un AR(2).

Análisis de series de tiempo

¿Qué modelo es: AR o MA?

- Por ejemplo, si tenemos un modelo AR(1), si excluimos el efecto del 1er lag (i.e. x(t-1)), el segundo lag (x(t-2)) es independiente de x(t).
- Entonces esta función de correlación parcial (PACF) caerá estrepitosamente después del 1er lag.





Claramente la gráfica tiene un corte en la gráfica de ACF después del segundo lag, lo que significa que a lo más es un MA(2).

Análisis de series de tiempo

Proceso:

1. Visualizar la serie
2. Hacerla estacionaria
3. Graficar ACF/PACF y encontrar parámetros
4. Construir el modelo ARIMA
5. Predecir

Estacionariedad

Métodos

2. Diferenciar: es la técnica más común para quitar la no estacionariedad. Tratamos de modelar las diferencias de los términos.

Por ejemplo, x(t) – x(t-1) = ARMA (p , q)

Esta diferenciación se llama integración: AR(I)MA. Así, tenemos tres parámetros:

p -> AR

d -> I

q -> MA

Estacionariedad

Métodos

2. Diferenciar: es la técnica más común para quitar la no estacionariedad. Tratamos de modelar las diferencias de los términos.

    Por ejemplo, x(t) – x(t-1) = ARMA (p , q)

    Esta diferenciación se llama integración: AR(I)MA. Así, tenemos tres parámetros: p -> AR, d -> I, q -> MA

3. Estacionalidad: se puede incorporar directamente al modelo ARIMA.

Los modelos con menor AIC (aikake criterion), BIC (bayes info criterion) serán los elegidos.

Análisis de series de tiempo

Proceso:

1. Visualizar la serie
2. Hacerla estacionaria
3. Graficar ACF/PACF y encontrar parámetros
4. Construir el modelo ARIMA
5. Predecir

Análisis de series de tiempo

Proceso:

1. Visualizar la serie
2. Hacerla estacionaria
3. Graficar ACF/PACF y encontrar parámetros
4. Construir el modelo ARIMA
5. Predecir

Gracias