Series Temporales y Minería de flujos de datos

Seminario: Software para procesamiento de series temporales y flujo de datos

Parte I: Series Temporales

Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Índice

Prerrequisitos

- En este seminario se asume lo siguiente:
  - El alumno conoce las técnicas para análisis y predicción de series temporales impartidas en clase de teoría.
  - El alumno está familiarizado con el entorno R.
  - El alumno tiene instalado en el PC el entorno R y la **biblioteca "tseries"** (recomendable también **R Studio**).
  - El alumno ha descargado los ficheros de apoyo a este seminario.

Prerrequisitos: Ficheros de apoyo

- Se proporcionan **ficheros de apoyo a estas diapositivas** organizados en las siguientes carpetas:
  - **1.SeriesEstacionarias**: Relacionados con la sección “Series estacionarias y no estacionarias”.
  - **2.AnalisisTendencia**: Relacionados con “Análisis de tendencia en R”.
  - **3.AnalisisEstacionalidad**: Relacionados con “Análisis de estacionalidad en R”.
  - **4.Estacionaridad**: Relacionados con “Estacionaridad”.
  - **5.Modelado_y_Prediccion**: Relacionados con “Modelos ARIMA y predicción”.
  - **6.Seleccion**: Relacionados con “Selección del mejor modelo”

Prerrequisitos: Repaso de teoría

- Modelo AR(p):

$$X(t) = \mu + \sum_{i=1}^{p} \lambda_i X(t-i) + \xi(t)$$

- Modelo MA(q):

$$X(t) = \xi(t) - \sum_{i=1}^{q} \theta_i \xi(t-i)$$

- Modelo ARMA(p, q):

$$X(t) = \mu + \sum_{i=1}^{p} \lambda_i X(t-i) + \xi(t) - \sum_{i=1}^{q} \theta_i \xi(t-i)$$

- ACF:

$$\rho_k = \frac{Cov[X(t), X(t+k)]}{Var[X(t)] \cdot Var[X(t+k)]}$$

- PACF: Mide la dependencia/contribución concreta que tiene un instante anterior *k* para predecir el instante actual.

Prerrequisitos: Comandos útiles de R

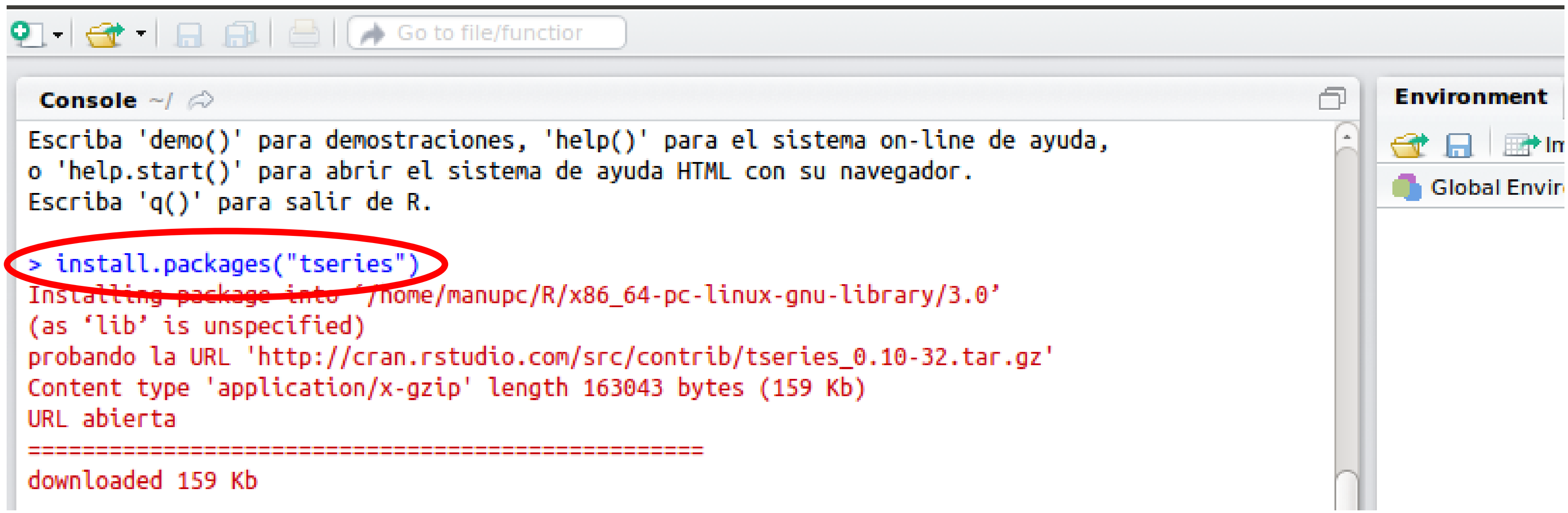
- Conviene repasar el uso de los siguientes comandos en R, de uso habitual:
  - **Gestión de ficheros:** setwd(“carpeta”), list.files(), scan , …
  - **Gestión de memoria de trabajo:** rm(list=ls()), …
  - **Scripts y bibliotecas:** source(“script”), library, …
  - **Creación de matrices:** rbind, cbind,
  - **Diferenciación e integración discreta:** diff, diffinv, …
  - **Gráficos:** plot, plot.ts, lines, …
  - **Obtención de ayuda:** ?comando

Índice

- Prerrequisitos
- **Series Temporales en R. Instalación de paquetes.**
- Lectura de datos. Formato y creación de objetos de series.
- Series estacionarias y no estacionarias.
- Metodología de análisis y modelado.
- Análisis de tendencia en R.
- Análisis de estacionalidad en R.
- Estacionaridad.
- Modelos ARIMA y predicción.
- Selección del mejor modelo.

Series temporales en R. Instalación de paquetes

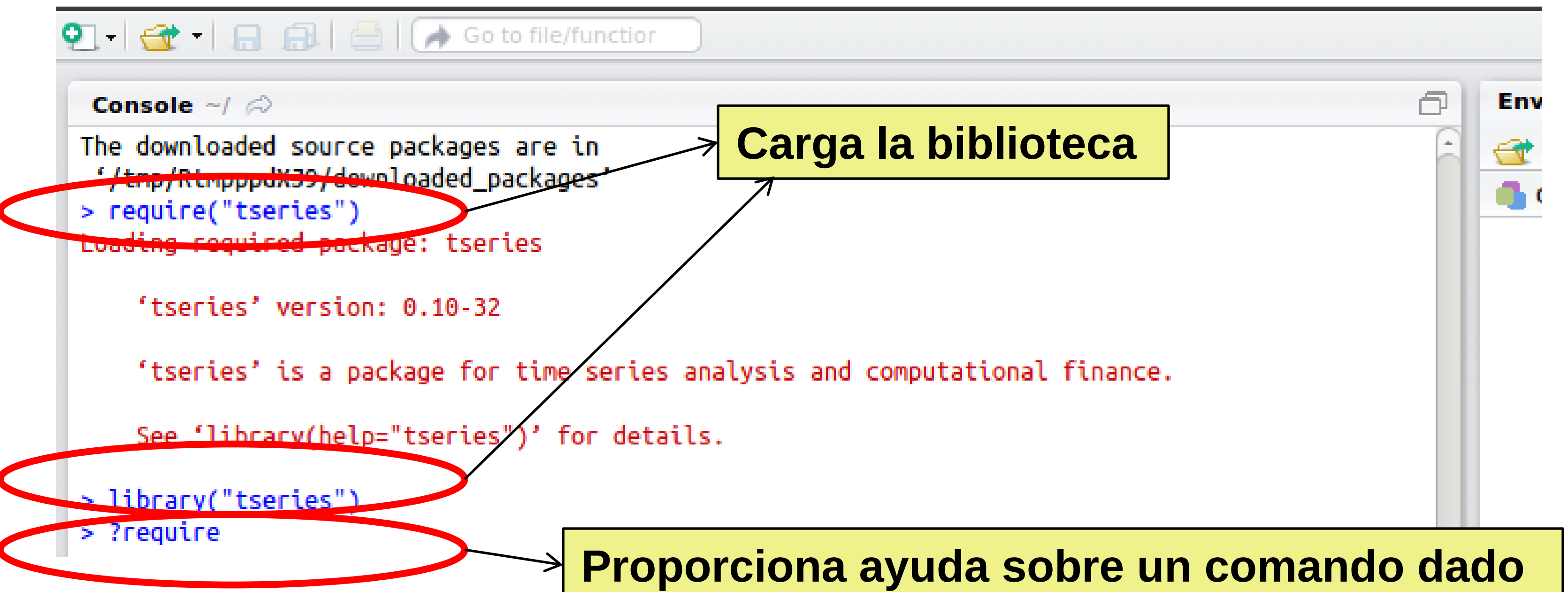
- Para trabajar con series temporales en R, es recomendable contar con las bibliotecas de Series Temporales que están disponibles en el repositorio R-CRAN. **Instalar el paquete tseries:**
  - **Install.packages("tseries"); Install.packages("DAAG")**



Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Series temporales en R. Instalación de paquetes

- Otras bibliotecas de R especializadas en series temporales:
    - **forecast, astsa,** etc.
- Cómo leer/cargar una biblioteca para usar sus funciones: comandos **library/require**



Índice

Lectura de datos. Formato y creación de series

- Usualmente, los datos estarán en un fichero con formato estándar (texto plano, csv, etc.). Para leer un fichero:
  - 1. Establecer la ruta del directorio donde deseamos trabajar.
  - 2. Cargar los datos con funciones estándar de R (scan, read, etc.). **Ejemplo:**

```
Console ~/facultad/asignat/STyMFD/practicas/1415/materialAlumnos/ejerciciosGuiados/
>
> setwd("/home/manupc/facultad/asignat/STyMFD/practicas/1415/materialAlumno
s/ejerciciosGuiados");
>
> list.files()
[1] "matricul.dat"
>
> datos <- scan("matricul.dat")
Read 480 items
```

setwd: Cambia el directorio de trabajo

list.files: Lista los ficheros en el directorio de trabajo

scan: Lee un fichero de datos en forma de vector

Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Lectura de datos. Formato y creación de series

- Podemos también encontrar los datos en formato de tabla.
    - 1. Leer la tabla
    - 2. Transformarla al formado deseado. Ejemplo:

```
Console ~/facultad/asignat/STyMFD/practicas/1415/materialAlumnos/ejerciciosGuiados/
>
> x <- read.table("AirPassengers.dat", skip=1)
>
>
> columnas <- paste("V", 2:13, sep="")
> datos <- x[columnas]
>
> datos <- as.numeric(t(as.matrix(datos)))
>
> datos <- ts(datos, start=c(1949, 1), frequency=12)
>
> datos
     Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
1949 112 118 132 129 121 135 148 148
1950 115 126 141 135 125 149 170 170
```

read.table: Lee tabla desde fichero.
skip=1: Salta la primera línea

Seleccionamos las columnas de interés, y las pasamos a serie (array) numérico

ts: Crea un objeto serie temporal, comenzando en enero de 1949, con una frecuencia de 12 meses

Índice

Series estacionarias y no estacionarias

- Una **serie estacionaria** es aquella cuya **media y varianza no varían con el tiempo.**

- La **estacionaridad** (no confundir con **estacionalidad**) indica que las propiedades estadísticas de la serie no varían en el tiempo y, por tanto, los datos pueden estudiarse bajo un mismo modelo paramétrico independiente del tiempo.

- **La condición de estacionaridad es un requisito** que debe cumplirse para poder aplicar modelos paramétricos de análisis y predicción de series de datos. Las series no estacionarias pueden tener (o no) alguna de las siguientes componentes:
    - **Tendencia**
    - **Estacionalidad**

Series estacionarias y no estacionarias

- La metodología Box-Jenkins para TSP engloba:
  - Dividir la serie X(t) en tendencia T(t), estacionalidad S(t) y una componente irregular (E(t)). Enfoque aditivo:

$$\mathbf{X(t)= T(t) + S(t) + E(t)}$$

  - **Cálculo de la estacionalidad:**
    - Valor medio de los valores de la serie dentro del mismo punto estacional.
  - **Cálculo de la tendencia:**
    - regresiones, filtrado de la señal, diferenciación, etc.
  - **Metodología:**
    - Eliminar tendencia y estacionalidad de la serie,
    - Hacer E(t) estacionaria y aplicar métodos paramétricos

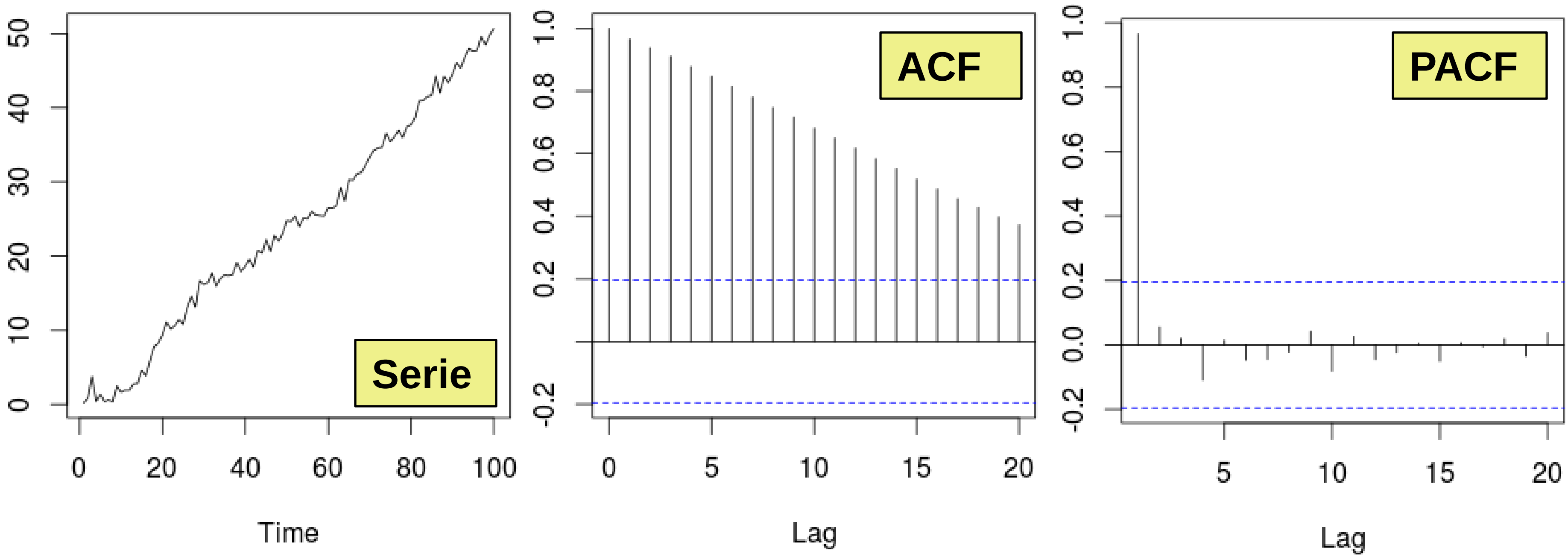
Series estacionarias y no estacionarias

- ¿Cómo saber si una serie es estacionaria?
  - **Test estadísticos**: Dickey-Fuller Ampliado (Test ADF)
  - **Gráficamente:** Observando las gráficas de autocorrelación (ACF) y autocorrelación parcial (PACF).

- **En R:**
  - **acf(serie), pacf(serie)**: Gráficos de ACF y PACF de una serie.
  - **adf.test(serie)** (paquete "tseries"): Test ADF. El valor resultante pvalue < 0.05 indica que la serie es estacionaria con un nivel de confianza del 95%.

Series estacionarias y no estacionarias

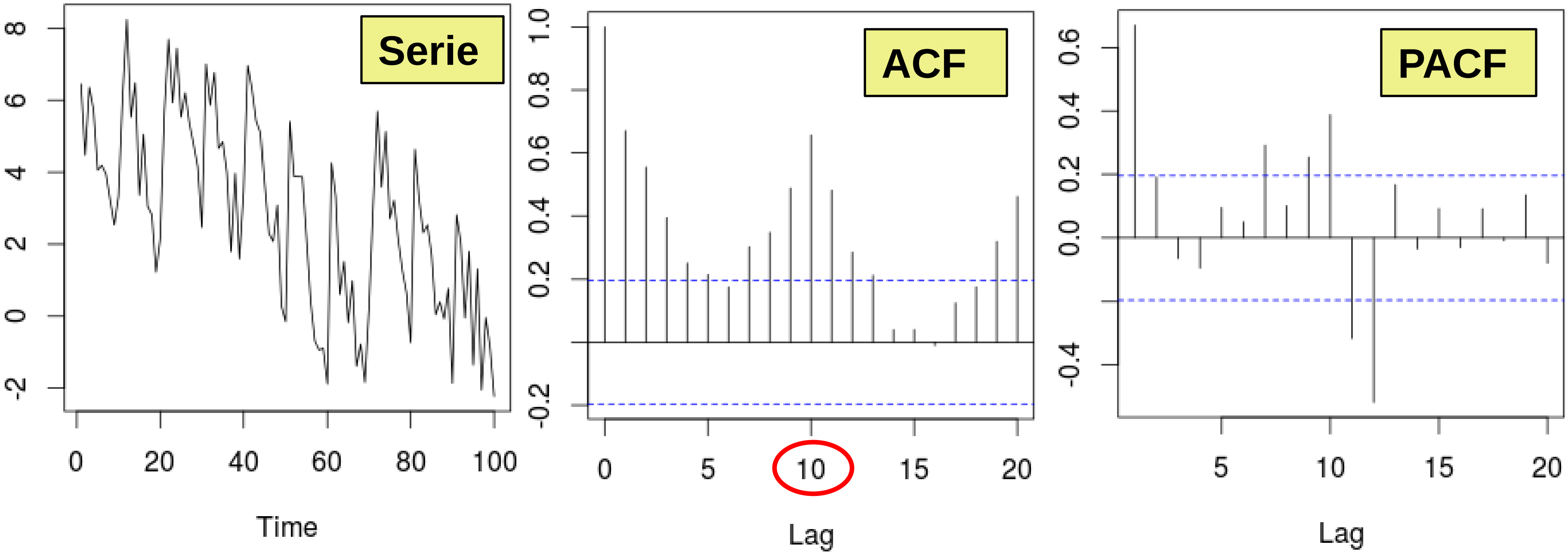
- Ejemplo de serie simulada con tendencia:



- Test Augmented Dickey-Fuller: pvalue=0.3106. Claramente **no estacionaria**

Series estacionarias y no estacionarias

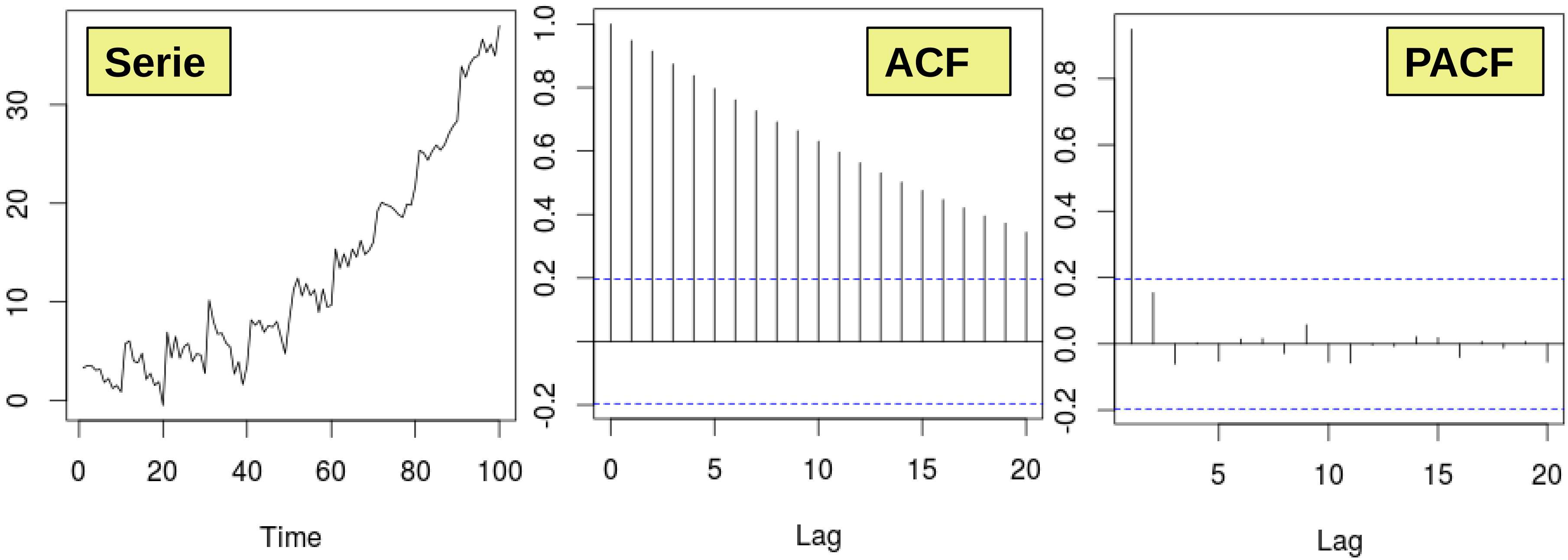
- Ejemplo de serie simulada con estacionalidad:



- Test Augmented Dickey-Fuller: pvalue=0.01. **¡Estacionaria!** **Cuidado con la estacionalidad**.

Series estacionarias y no estacionarias

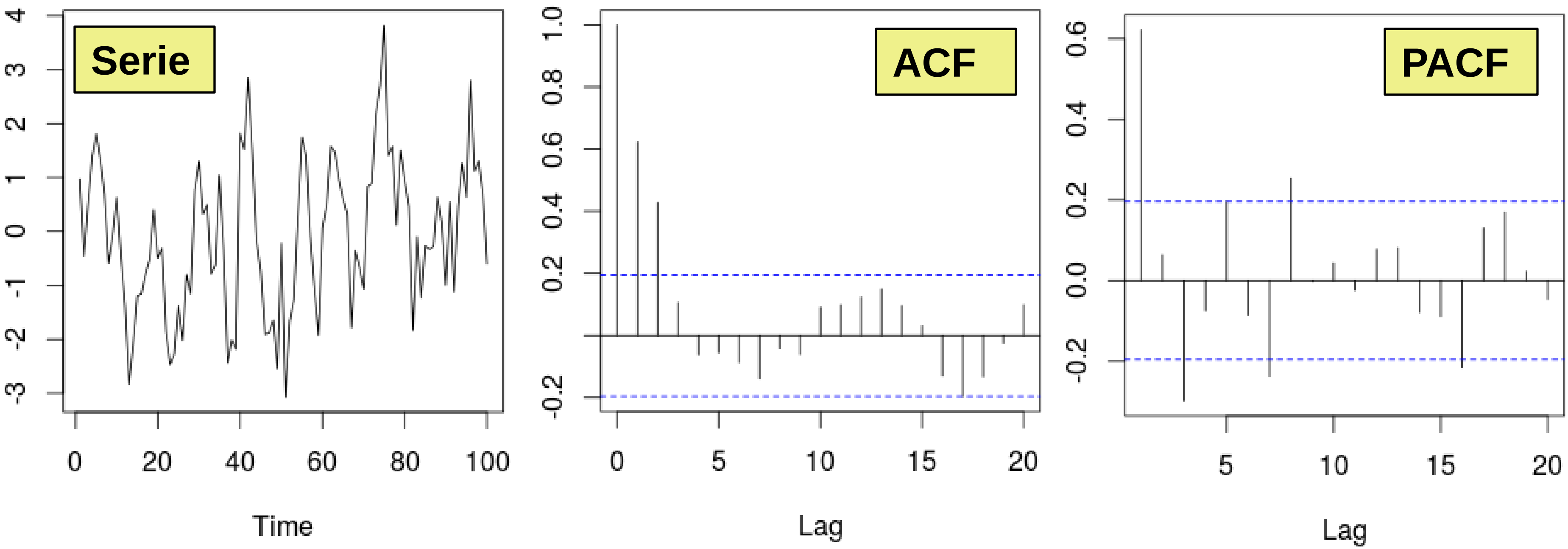
- Ejemplo de serie simulada con tendencia y estacionalidad:



- Test Augmented Dickey-Fuller: pvalue~=0.87. Claramente **no estacionaria**.

Series estacionarias y no estacionarias

- Ejemplo de serie simulada estacionaria:



- Test Augmented Dickey-Fuller: pvalue~=0.01. Todo indica que la serie es **estacionaria**.

Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Índice

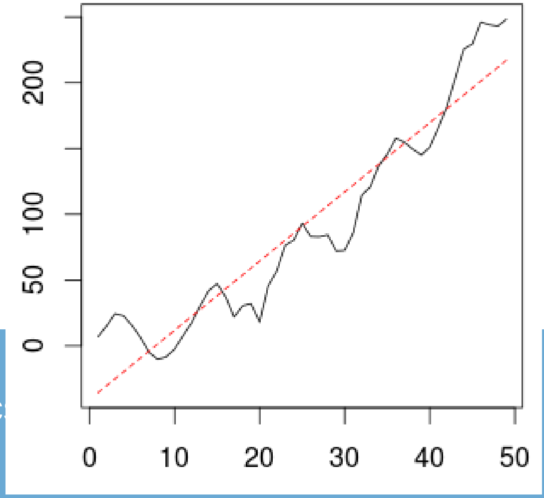
Metodología de análisis y modelado

- La metodología que seguiremos para modelar series debe seguir los siguientes pasos:
  - **1. Análisis de tendencia**. ¿Tiene tendencia la serie? Modelarla y eliminarla.
  - **2. Análisis de estacionalidad.** ¿Sufre la serie de estacionalidad? Modelarla y eliminarla.
  - **3. Estacionaridad.** ¿Es estacionaria la serie? En caso contrario, hacerla estacionaria.
  - **4. Aplicar modelos paramétricos**. En nuestro caso, modelos autorregresivos y de medias móviles.
  - **5. Predicción**. Predecir en base a todas las componentes modeladas.

Análisis de tendencia en R

- **Estimación funcional:** Aproximar la tendencia de la serie como una función. Requiere realizar hipótesis sobre el modelo que rige la tendencia. Ejemplo (I): Modelo lineal

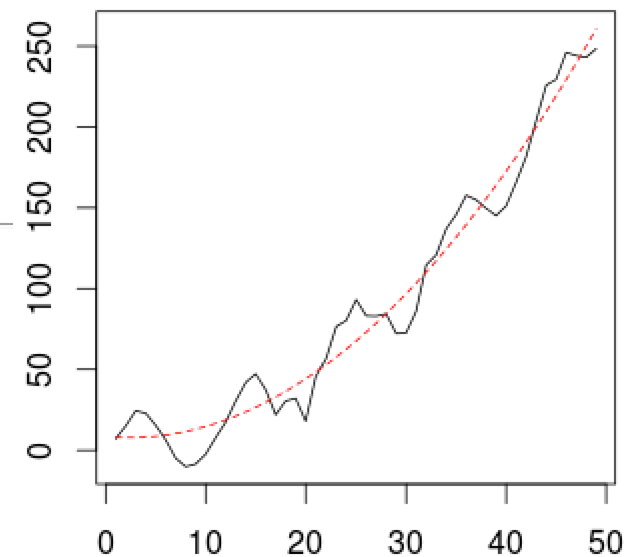
```r
sintetico1<-scan("sintetico1.dat"); # Cargamos datos
plot.ts(sintetico1); # Los mostramos

# Hipótesis 1: Modelo Lineal x= a + b*t (x=serie; t=tiempo; a,b=parámetros a estimar)
tiempo<- 1:length(sintetico1) # Creamos la variable que modela al tiempo
parametros <- lm (sintetico1 ~ tiempo) # Ajustamos modelo lineal
estimacion<-parametros$coefficients[1]+tiempo*parametros$coefficients[2] # Calculamos la est

series<-matrix(c(t(sintetico1), t(estimacion)), ncol=2); # Mostramos resultado
matplot(series, pch=1, type= "l")
```

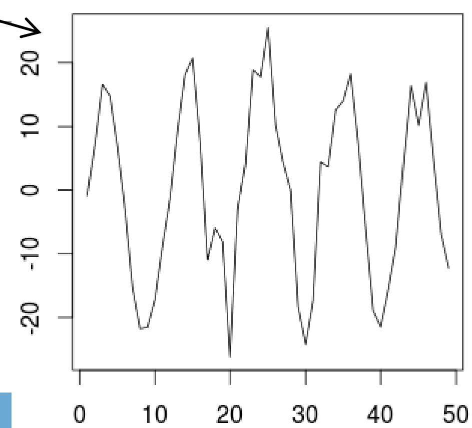


Análisis de tendencia en R (II)

- **Estimación funcional:** Requiere realizar hipótesis sobre el modelo que rige la tendencia. Ejemplo (II): Modelo polinómico grado 2

```r
# Hipotesis 2: modelo cuadratico x= a + b t + c t 2 (x=serie, t=tiempo, a,b,c=parametros a e
tiempo<- 1:length(sintetico1)
parametros <- lm (sintetico1 ~ tiempo + I(tiempo^2))
estimacion<-parametros$coefficients[1]+tiempo*parametros$coefficients[2]+tiempo*tiempo*param
#series<-matrix()
series<-matrix(c(t(sintetico1), t(estimacion)), ncol=2);
matplot(series, pch=1, type= "l")

#Eliminar tendencia
sintendencia= sintetico1 - estimacion;
plot(sintendencia, type="l")
```





Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

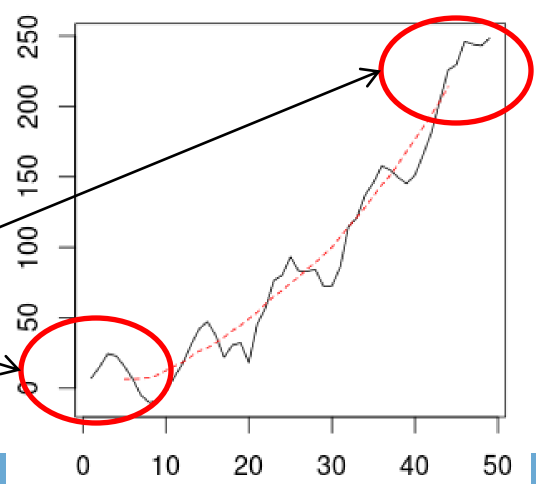
Análisis de tendencia en R (III). Filtrado

- **Filtrado:** Esta opción consiste en aplicar un filtro de medias móviles para estimar la tendencia.

$$\hat{X}(t) = \frac{1}{k} \sum_{i=-k/2}^{k/2} X(t+i)$$

```
# Estimación de tendencia por filtrado de orden k
for (k in 5:10) {
  filtro<-rep(1/k, k); # Creamos el filtro

  # Filtramos señal
  filtrada<-filter(sintetico1,filter=filtro,sides=2,method="convolution")
  series<-matrix(c(t(sintetico1), t(filtrada)), ncol=2);
  matplot(series, pch=1, type= "l")
  pause<-readline(); # para pausar la ejecución
}
```

¡OJO con el filtrado! No es posible recuperar los valores de los extremos



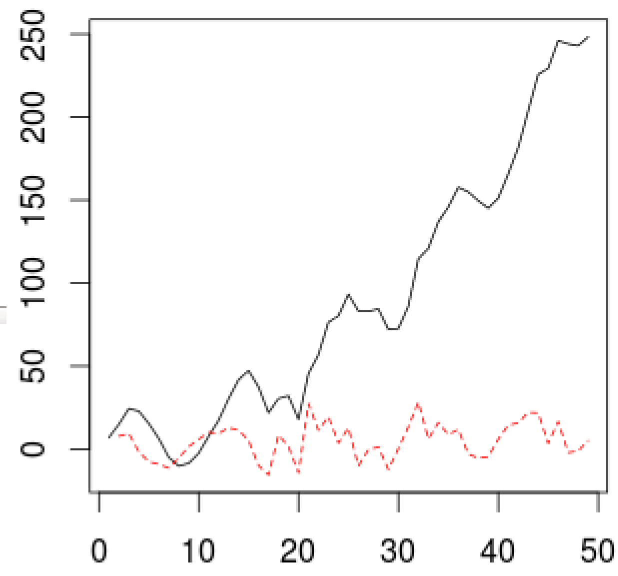
Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Análisis de tendencia en R (IV). Diferenciación

- **Diferenciación:** Se diferencia la señal hasta que desaparezca la tendencia.

$$\hat{X}(t) = X(t) - X(t-1)$$

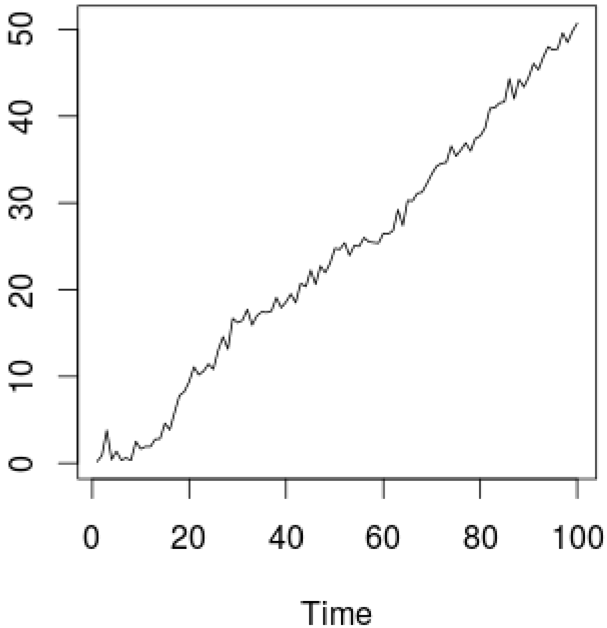
```r
diferenciada<-sintetico1;
for (k in 1:4) {

  diferenciada<- diff(diferenciada); # Diferenciamos señal
  length(diferenciada)
  amostrar <- c(rep(NaN, k), diferenciada);
  series<-matrix(c(t(sintetico1), t(amostrar)), ncol=2);
  matplot(series, pch=1, type= "l")

  pause<-readline(); # para pausar la ejecución
}
```



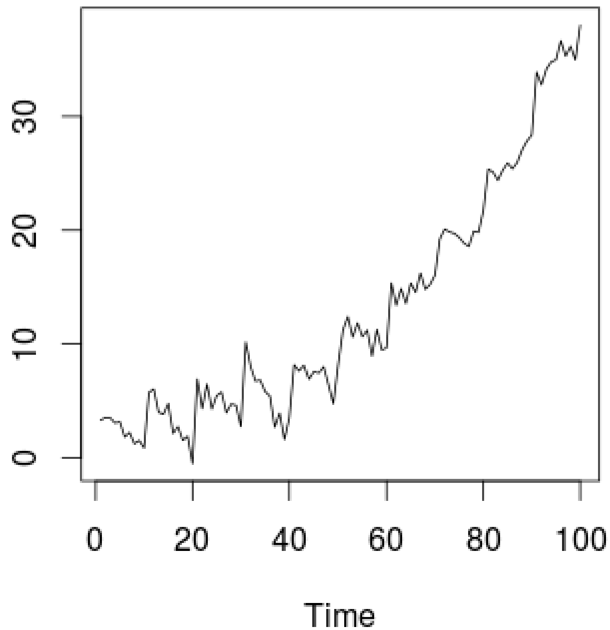
Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Análisis de tendencia en R. Ejemplos.

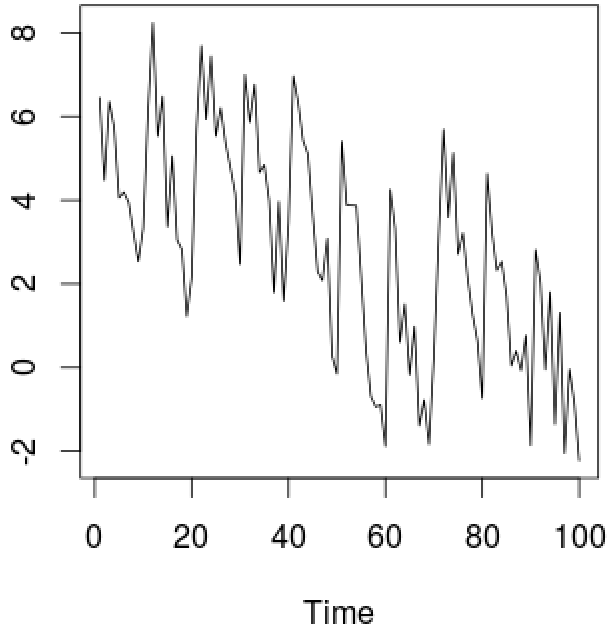
- Ejemplos de series anteriores con tendencia:



Ejemplo 1



Ejemplo 2

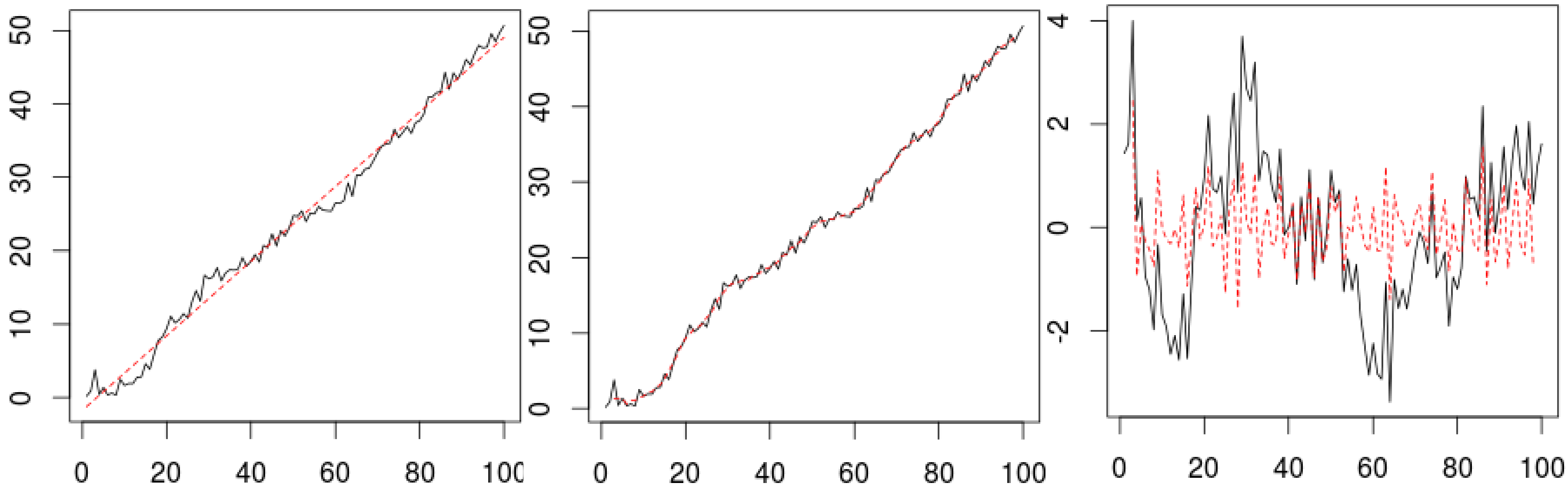


Ejemplo 3

Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Análisis de tendencia en R. Ejemplos.

- Ejemplo (I): Aproximación lineal (izq.) y filtrado (centro).



Regresión lineal (rojo)

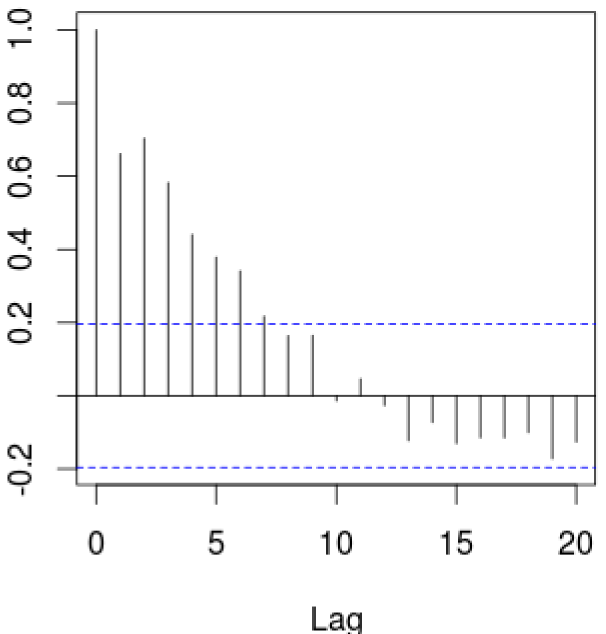
Filtrado (rojo)

Serie resultante X^(t)
Negro: Regresión
Rojo: Filtrado

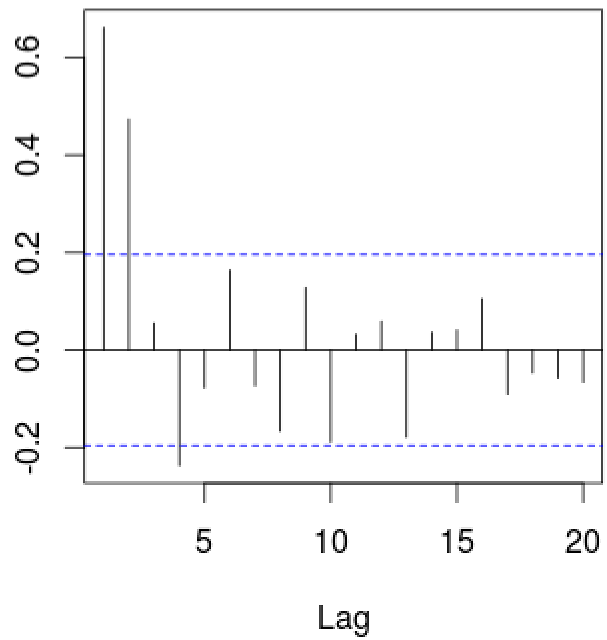
- Eliminar tendencia: $\hat{X}(t) = X(t) - T(t)$

Análisis de tendencia en R. Ejemplos.

- Ejemplo (I): ACF y PACF de la serie sin tendencia (técnica de **aproximación funcional**)



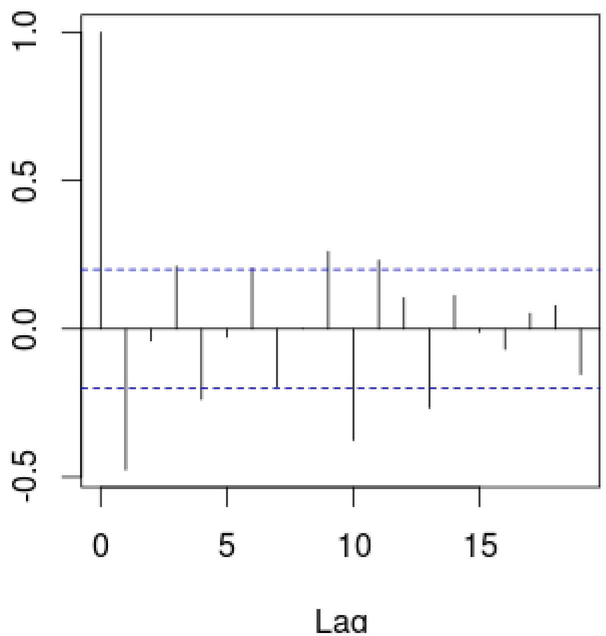
ACF



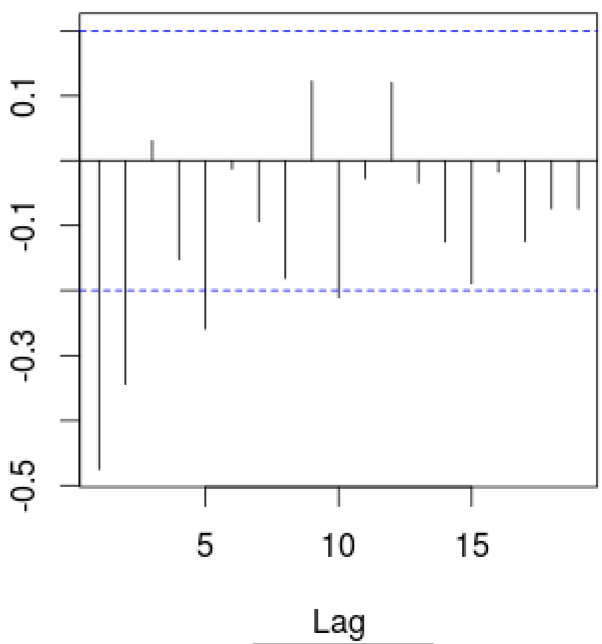
PACF

Análisis de tendencia en R. Ejemplos.

- Ejemplo (I): ACF y PACF de la serie sin tendencia (técnica de **filtrado**)



ACF



PACF

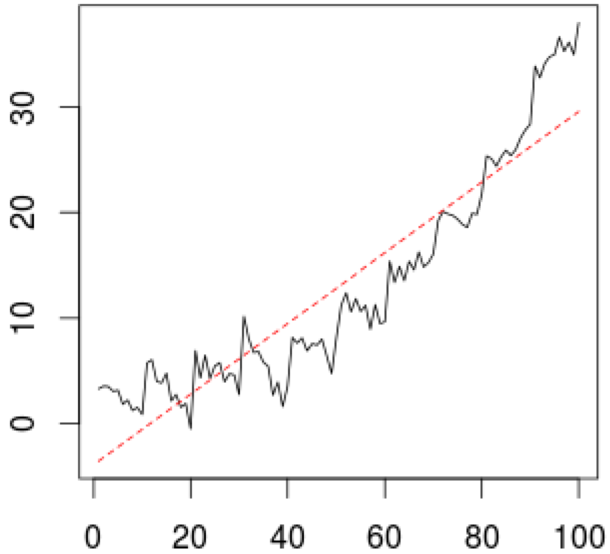
Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Análisis de tendencia en R. Ejemplos.

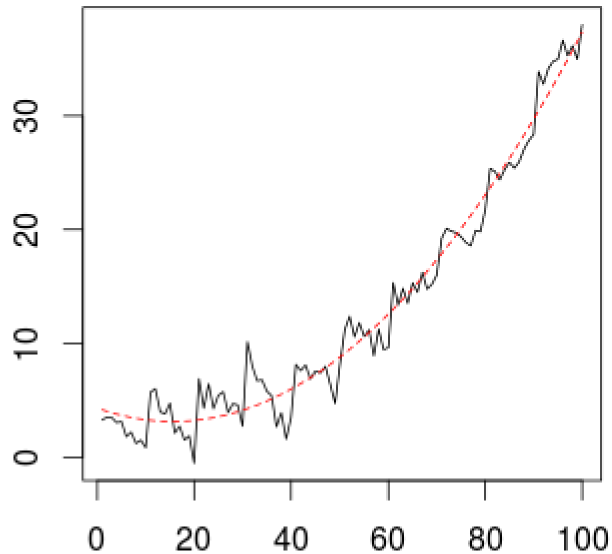
- Ejemplo (I): ¿Con cuál nos quedamos?
  - En este caso, está claro: Regresión lineal.
  - Aunque hayamos aplicado un filtro, hay que predecir los valores de la tendencia en el futuro.
  - Por tanto, hay que modelar el filtro mediante aproximación funcional.
  - Se usará el modelo lineal para ello, dado que es el que mejor representa los datos.
  - Por tanto, lo más simple es seleccionar el método de aproximación lineal desde el principio.

Análisis de tendencia en R. Ejemplos.

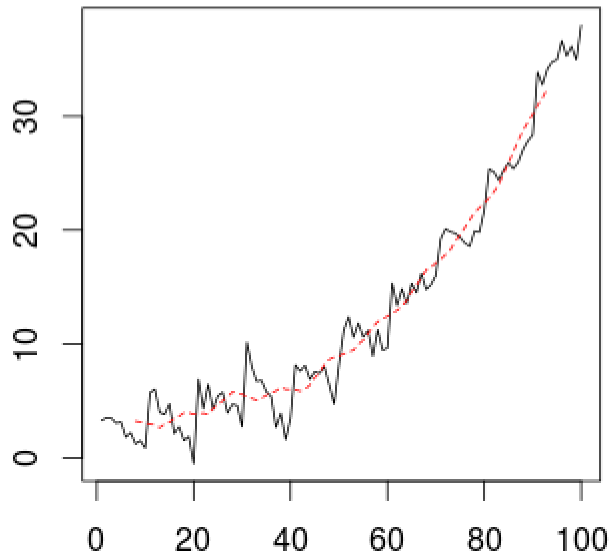
- Ejemplo (II): Aproximación lineal (izq.), polinómica (centro) y filtrado (dcha).



Regresión lineal (rojo)



Regresión polinómica (rojo)

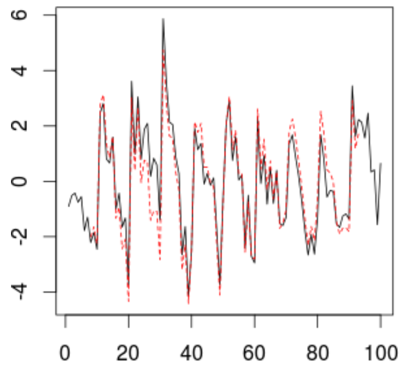


Filtrado (rojo)

Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Análisis de tendencia en R. Ejemplos.

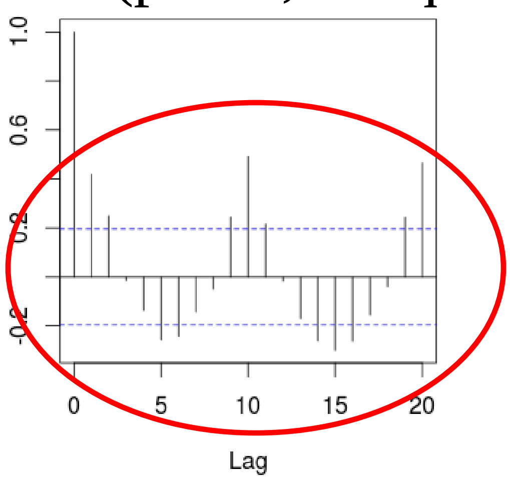
- Ejemplo (II): Series sin la tendencia



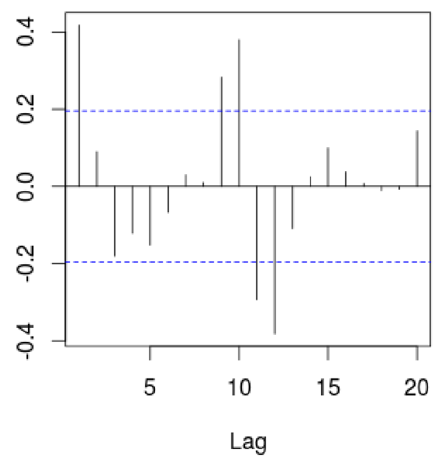
Polin. (negro Filtrado (rojo)

- ACF y PACF (polin., aunque muy similar en el filtrado):

ACF

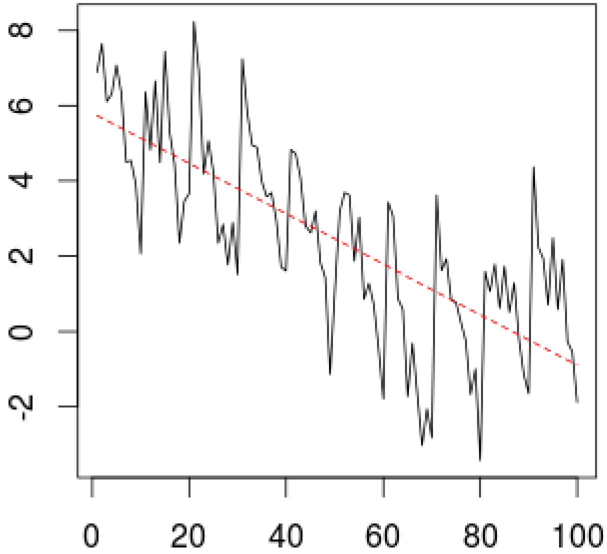


PACF

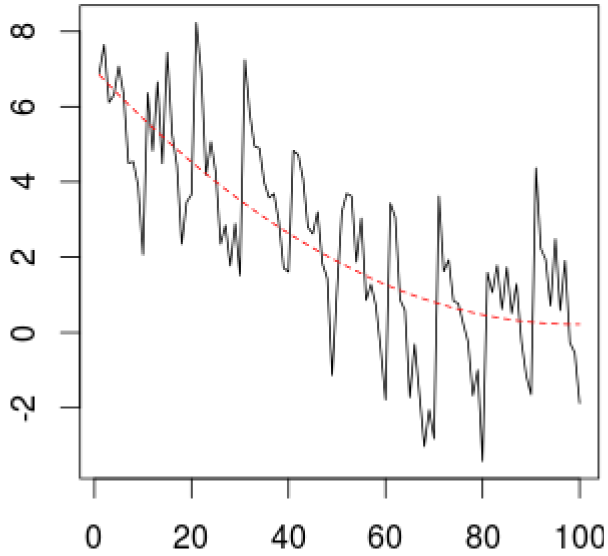


Análisis de tendencia en R. Ejemplos.

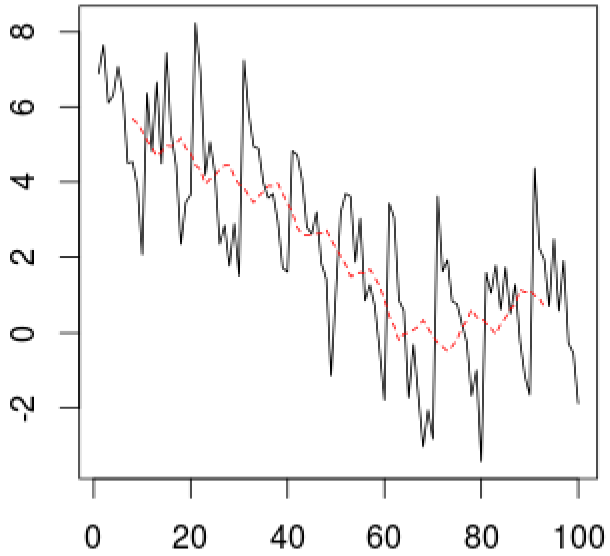
- Ejemplo (III): Aproximación lineal (izq.), polinómica (centro) y filtrado (dcha).



Regresión lineal (rojo)



Regresión polinómica (rojo)

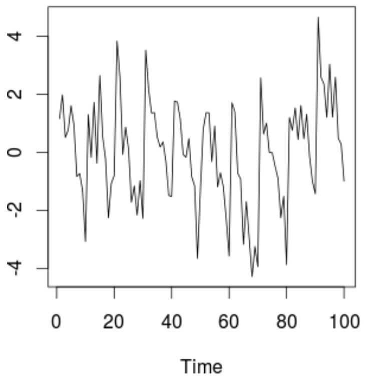


Filtrado (rojo)

Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Análisis de tendencia en R. Ejemplos.

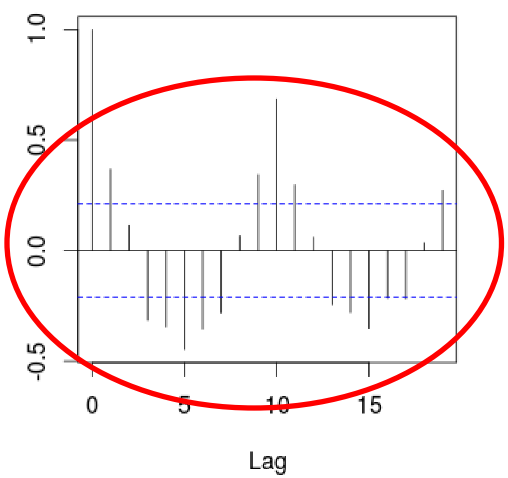
- Ejemplo (III): Series sin la tendencia



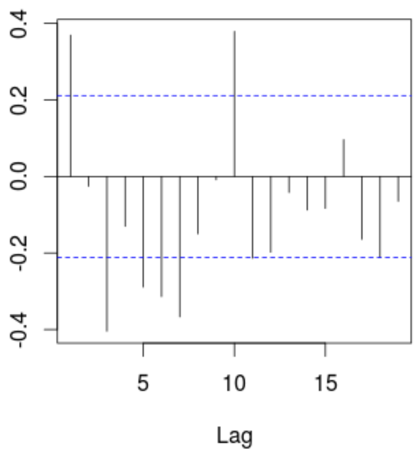
Tendencia= Aproximación lineal

- ACF y PACF (polin., aunque muy similar en el filtrado):

ACF



PACF

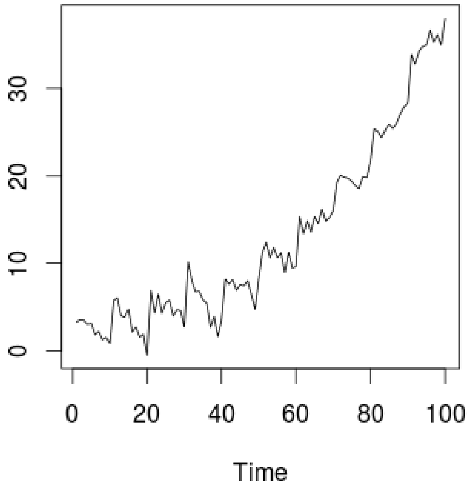


Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

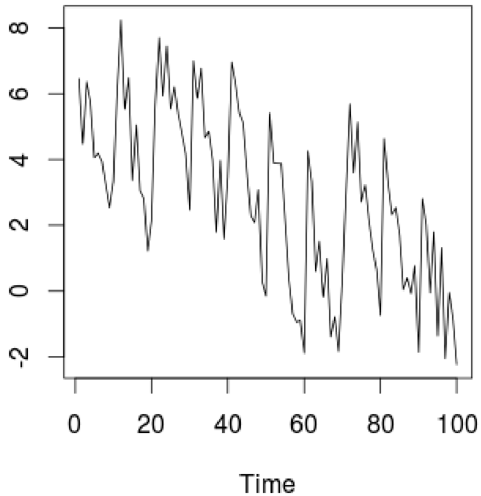
Índice

Análisis de estacionalidad en R

- Se encuentra el periodo de la estacionalidad.
- La ACF puede ayudar a calcular el periodo.
- Se calcula el valor promedio de cada punto de la estación (si es mensual, la media para cada mes; si es semanal, la media para cada día de la semana, etc.).



Ejemplo 2



Ejemplo 3

Análisis de estacionalidad en R

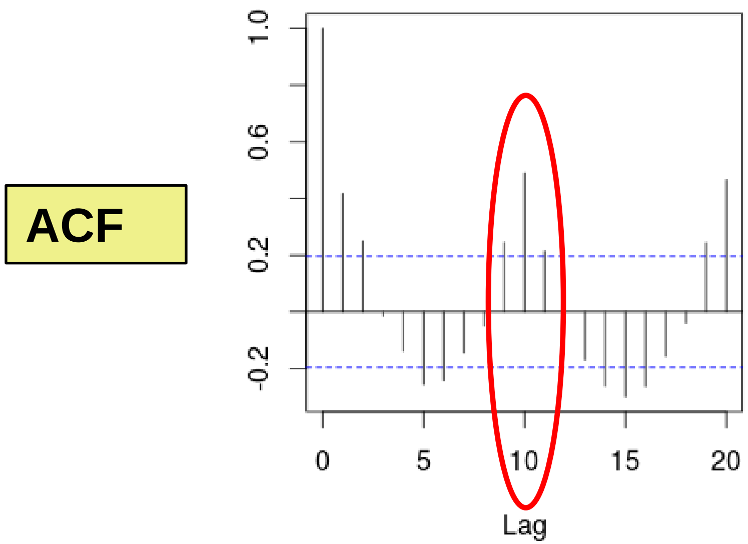
- Otro método para eliminar estacionalidad consiste en utilizar la **diferenciación**, con un desfase **d** igual al periodo de la estacionalidad:

$$\hat{X}(t) = X(t) - X(t-d)$$

- Existen múltiples métodos:
    - Transformada de Fourier
    - Método de Holts-Winter de alisado exponencial para estacionalidad
- En las siguientes diapositivas utilizamos cálculo promedio en las 2 series de ejemplo.

Análisis de estacionalidad en R

- Ejemplo (II): Partimos de la serie sin la tendencia. Calculamos ACF

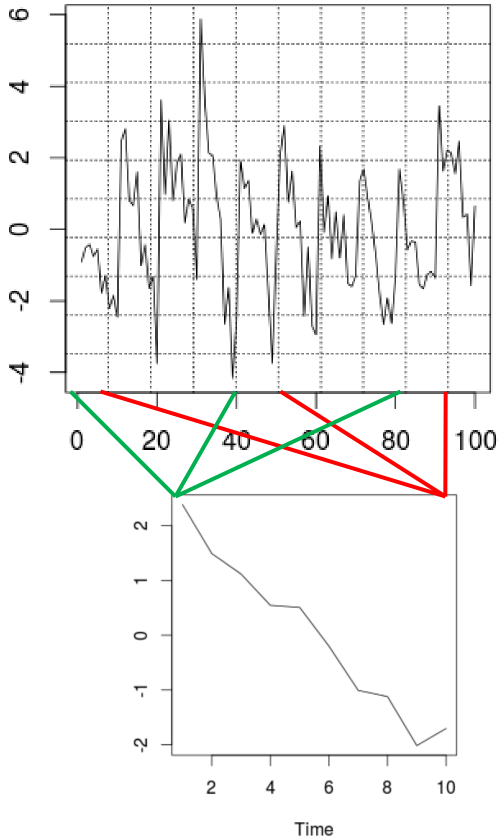


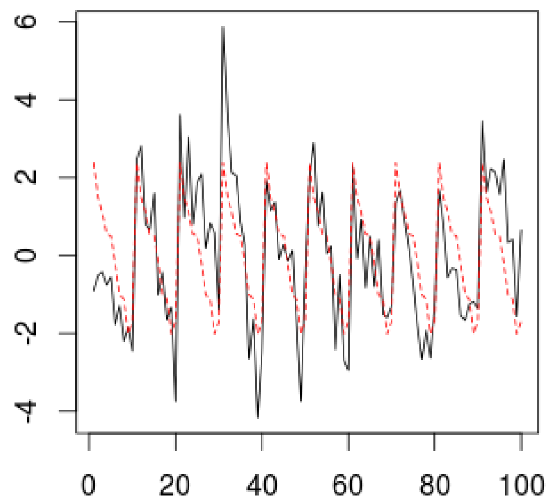
- La serie tiene una estacionalidad con un período igual a **10**.
  - S(1)=media(X(1), X(11), X(21), ...)
  - S(2)=media(X(2), X(12), X(22), ...)
  - ...
  - S(10)=media(X(10), X(20), X(30), ...)

Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

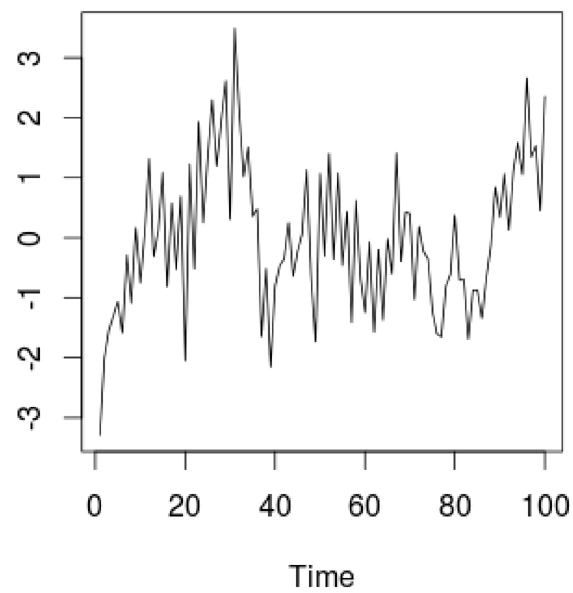
Análisis de estacionalidad en R

- Ejemplo (II): Calculamos la estacionalidad S(t) (tamaño 10) y la restamos de la serie original para **calcular E(t)= X(t)-T(t)-S(t)**





Serie (negro)
Estacionalidad (rojo)

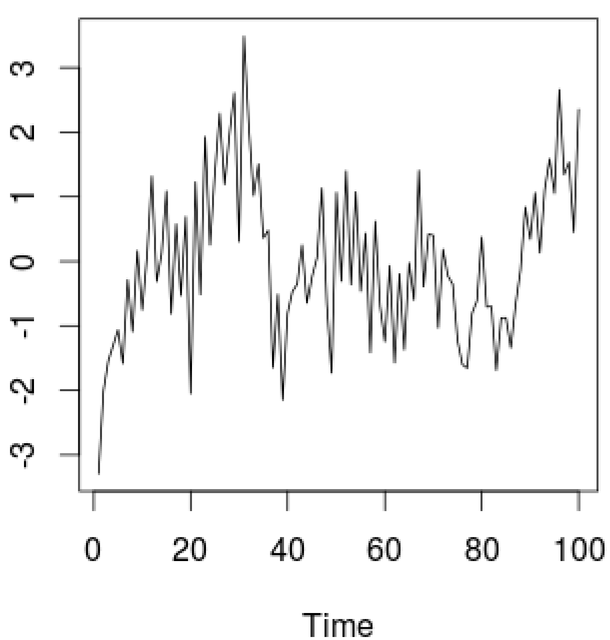


Componente E(t)

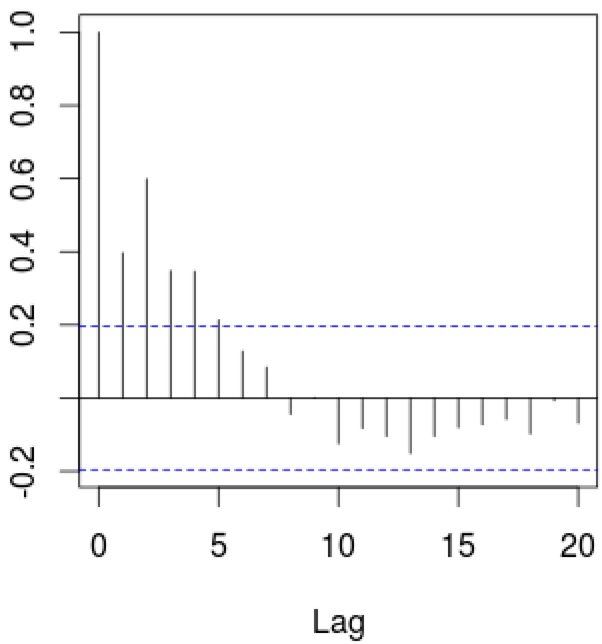
Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Análisis de estacionalidad en R

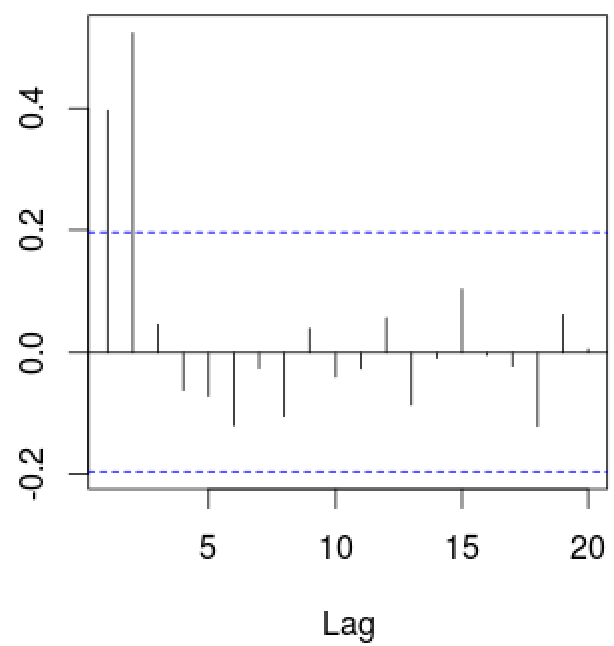
- Ejemplo (II): ACF y PACF tras quitar la estacionalidad



Serie sin T(t) ni S(t)



ACF

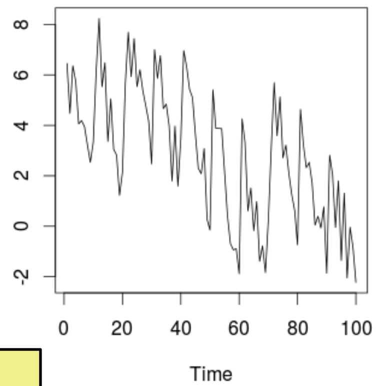


PACF

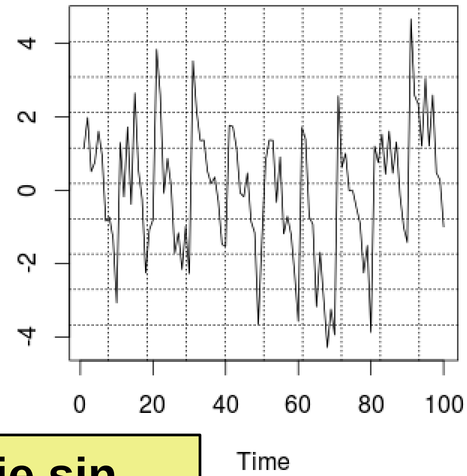
Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Análisis de estacionalidad en R

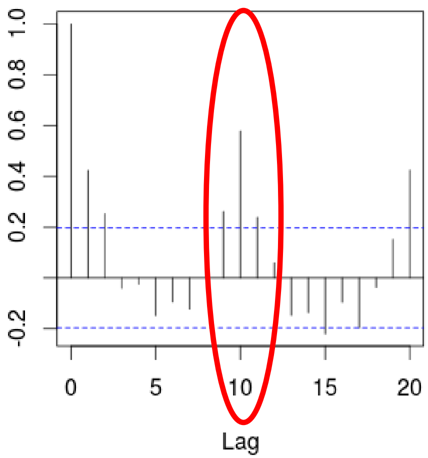
- Ejemplo (III): Partimos de la serie sin la tendencia. Calculamos ACF



Serie



Serie sin tendencia

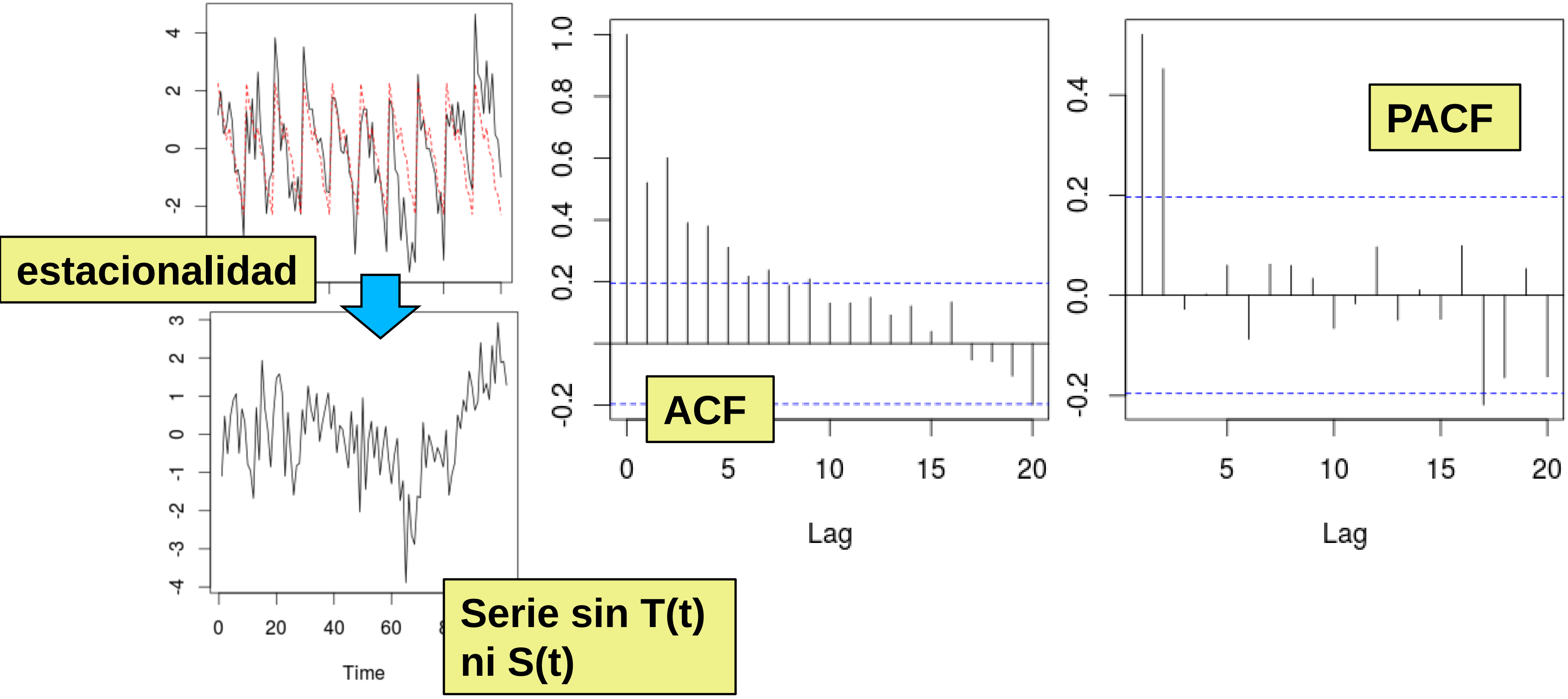


ACF

- La serie tiene una estacionalidad con un período igual a **10**.
  - S(1)=media(X(1), X(11), X(21), ...)
  - ...
  - S(10)=media(X(10), X(20), X(30), ...)

Análisis de estacionalidad en R

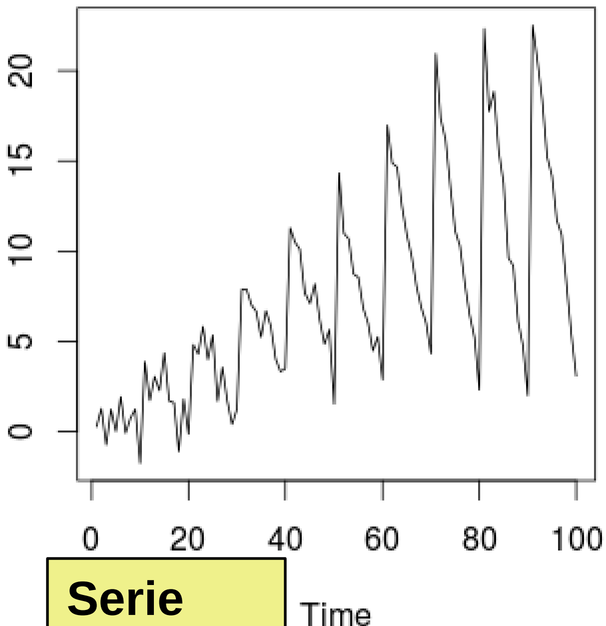
- Ejemplo (III): Calculamos la estacionalidad S(t) (tamaño 10) y la restamos de la serie original para **calcular E(t)= X(t)-T(t)-S(t)**



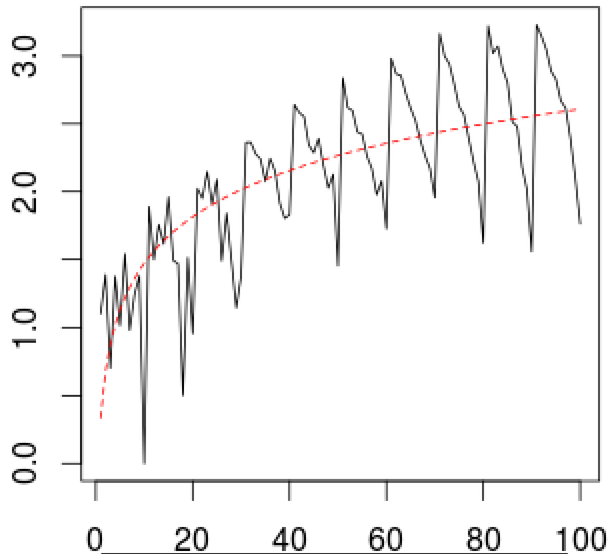
Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Análisis de estacionalidad en R

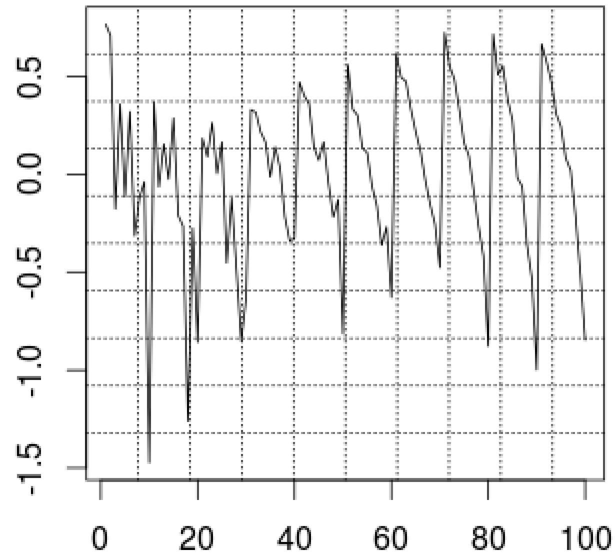
- En ocasiones, **puede ser necesario aplicar transformaciones a los datos** (normalmente logarítmica) para **normalizar el tratamiento de la estacionalidad,** si esta crece o decrece con el tiempo. Ejemplo:



Serie original



Serie en escala logarítmica



Serie sin tendencia

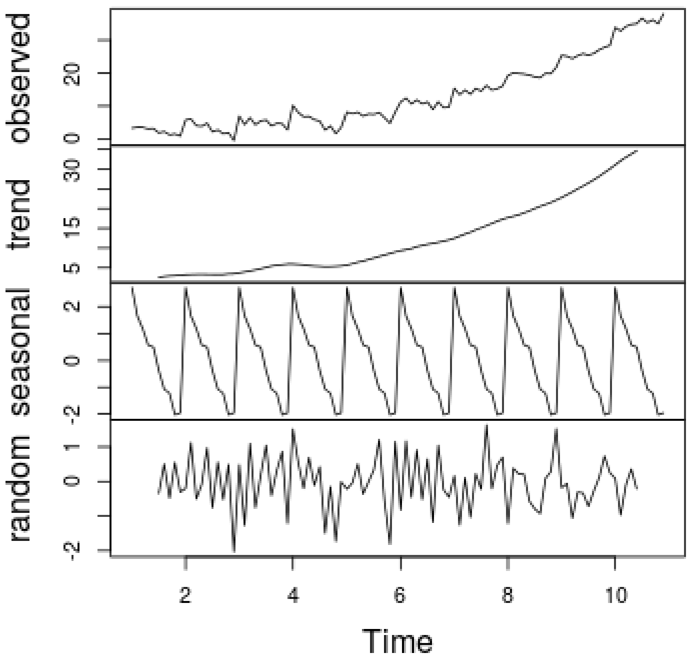
Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Análisis de estacionalidad en R

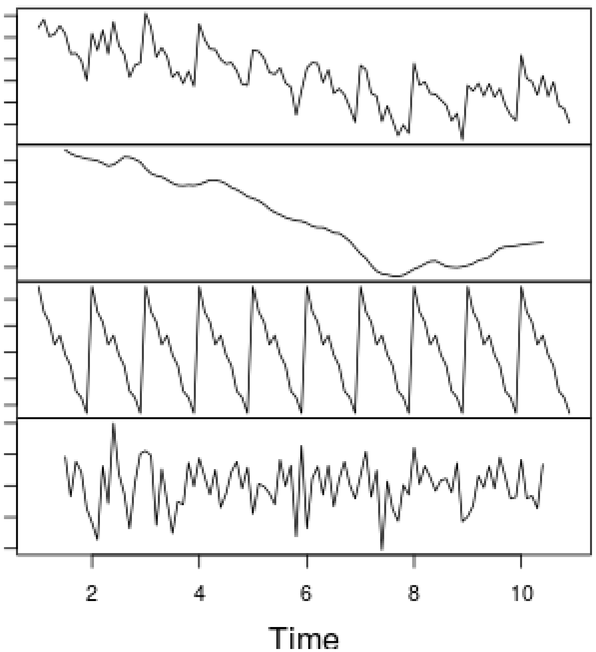
- Recursos de R para eliminar la tendencia y la estacionalidad:
  - Función **decompose**: Descompone una serie en las componentes “trend” (tendencia), “seasonal” (estacionalidad) y “random” (resto de valores de la serie).
  - **Requisito**: Se tiene que crear un objeto **ts** con una frecuencia (periodo de estacionalidad) conocida a priori.
  - R calcula la tendencia mediante el método de filtrado por medias móviles.
    - **Serie.ts <- ts(MiSerieOriginal, frequency = valor)**
    - **Serie.descompuesta<- decompose(Serie.ts)**
    - **plot(Serie.descompuesta)**

Análisis de estacionalidad en R

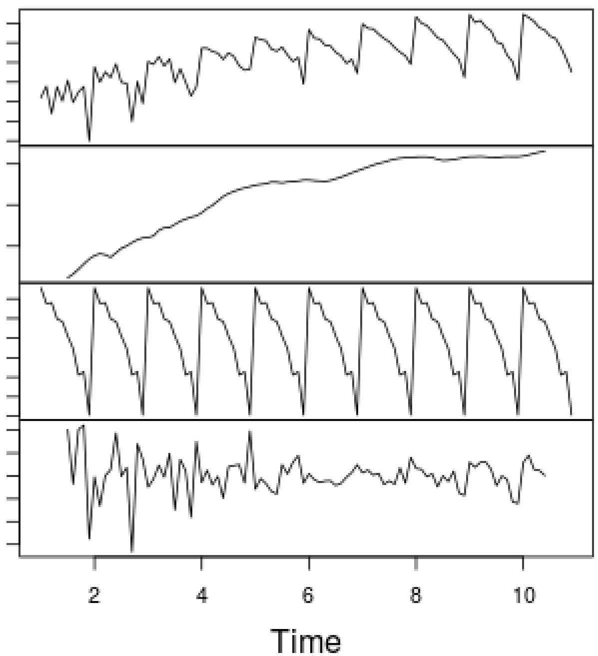
- Ejemplo de decompose para las series anteriores:



Ejemplo 2



Ejemplo 3



Ejemplo escala logarítmica

Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Índice

Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Estacionaridad

- Una vez eliminadas tendencia y estacionalidad, tenemos:

$$\mathbf{E(t)= X(t)-T(t)-S(t)}$$

- **E(t)** debe ser estacionaria.
- Para saber si una serie es estacionaria, podemos fijarnos en el ACF.
- Si el ACF tiende "rápidamente" a 0, entonces es estacionaria.
- Además, el test aumentado de Dickey-Fuller comprueba estacionaridad.
- En caso contrario, se debe diferenciar la serie:
  - **E'(t) = E(t)-E(t-d)**
  - **d= instantes de tiempo de diferenciación**

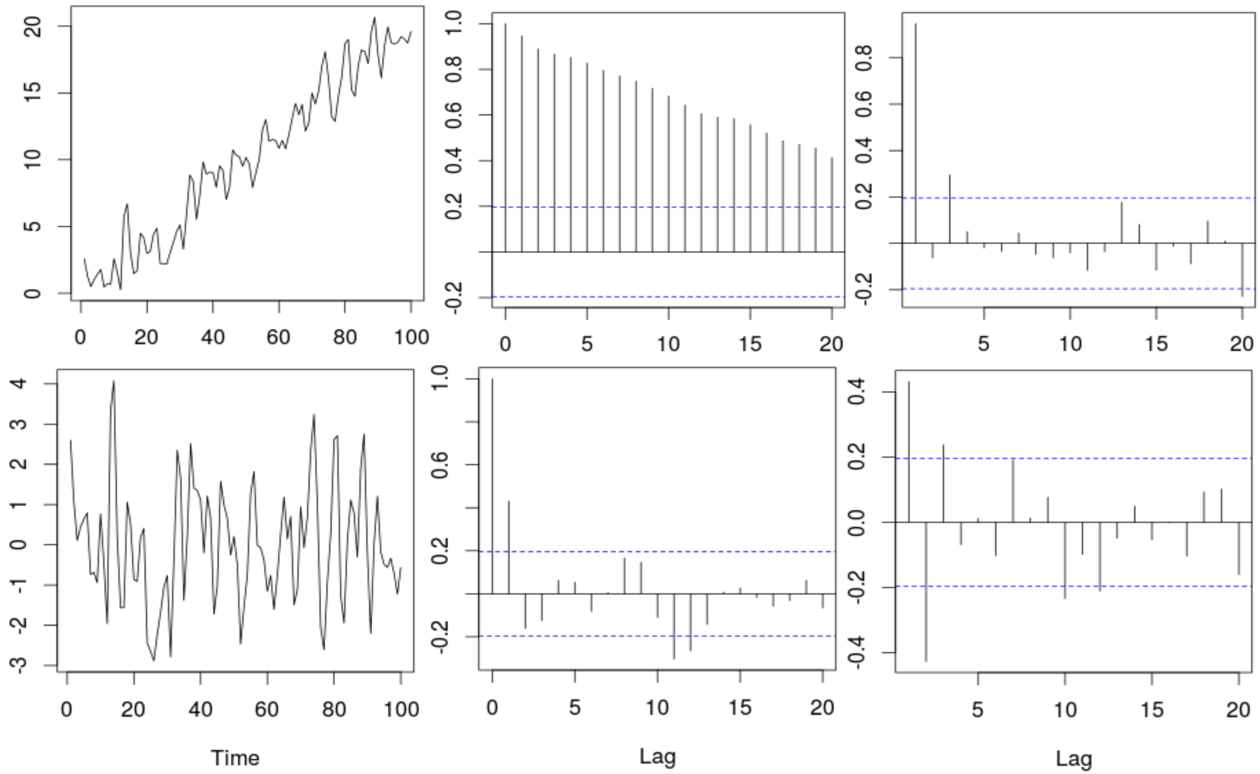
Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

- Ejemplos de series estacionarias y no estacionarias y su ACF:

ARIMA(0, 0, 1)

Serie original

Serie sin tendencia



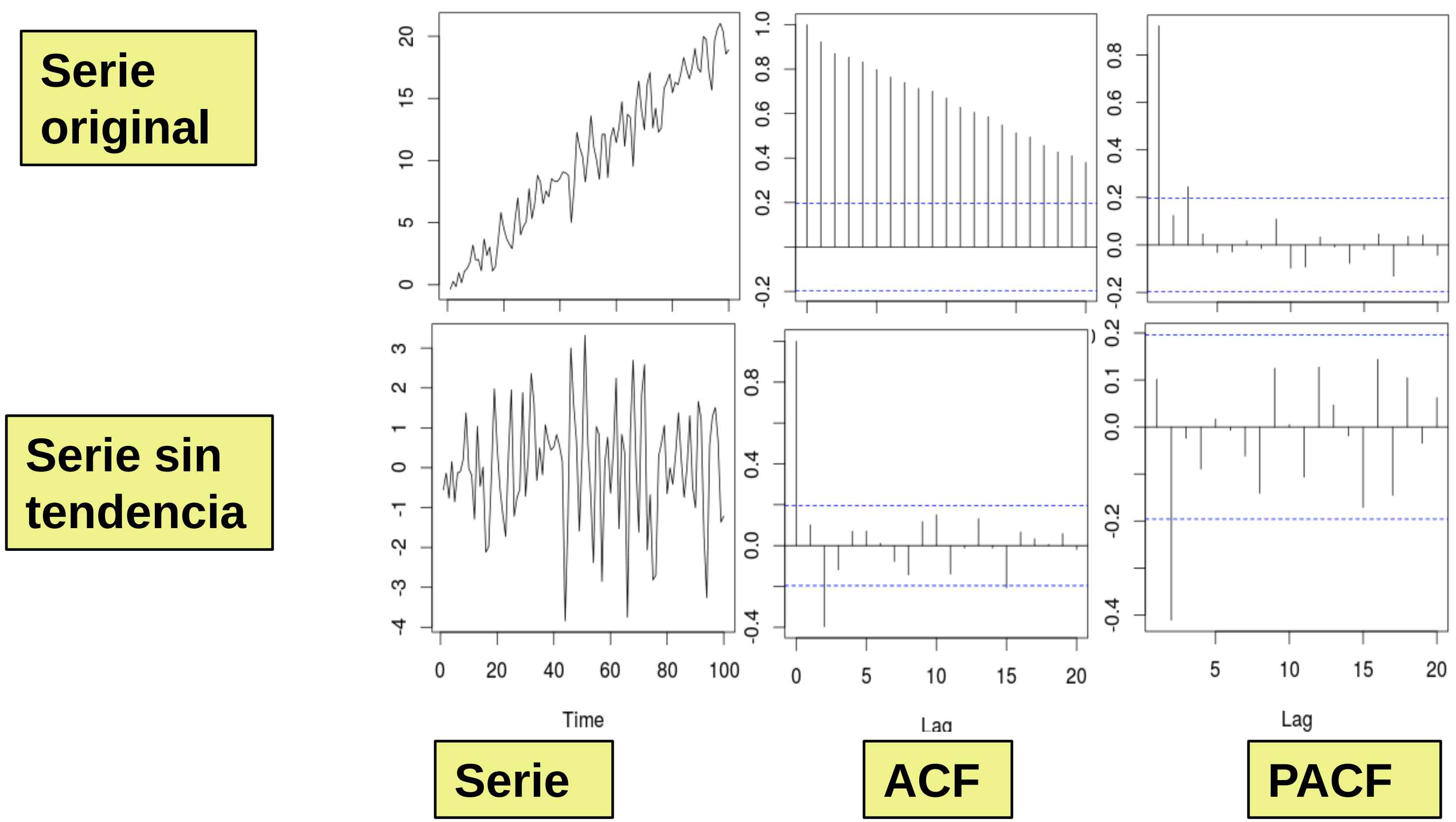
Serie **ACF** **PACF**

Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Estacionaridad

- Ejemplos de series estacionarias y no estacionarias y su ACF:

ARIMA(0, 0, 2)



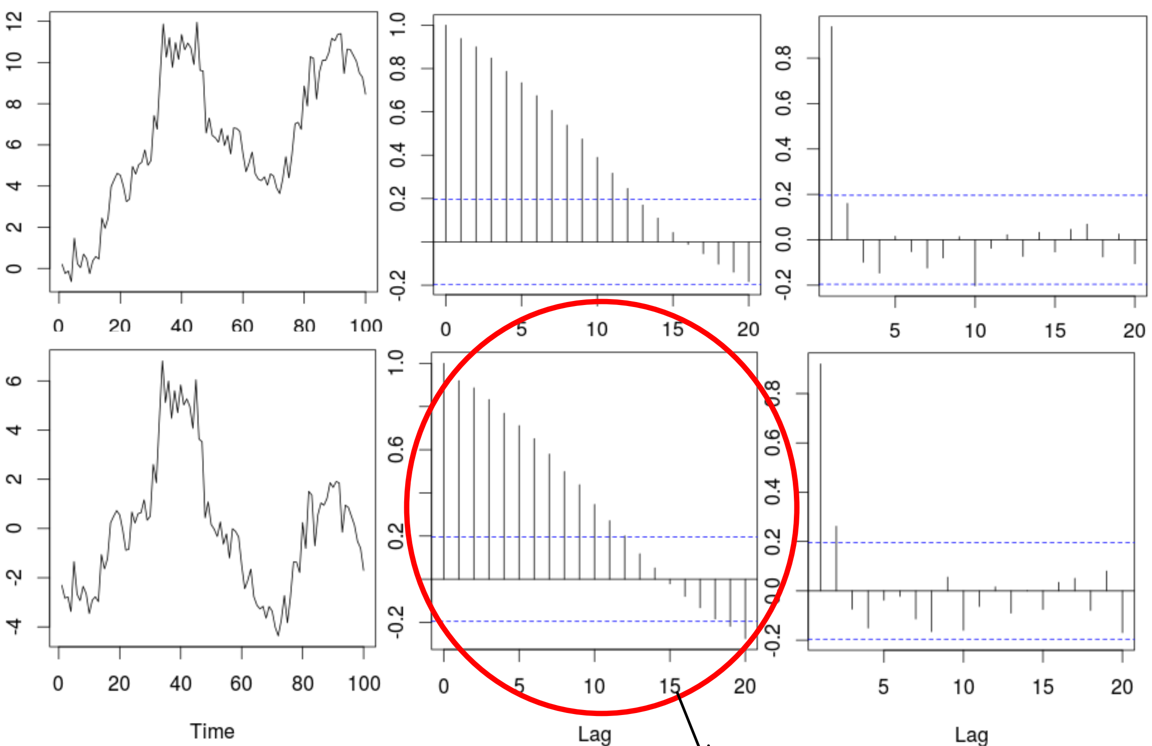
Estacionaridad

- Ejemplos de series estacionarias y no estacionarias y su ACF:

ARIMA(0, 1, 2)

Serie original

Serie sin tendencia



Serie

ACF

??

PACF

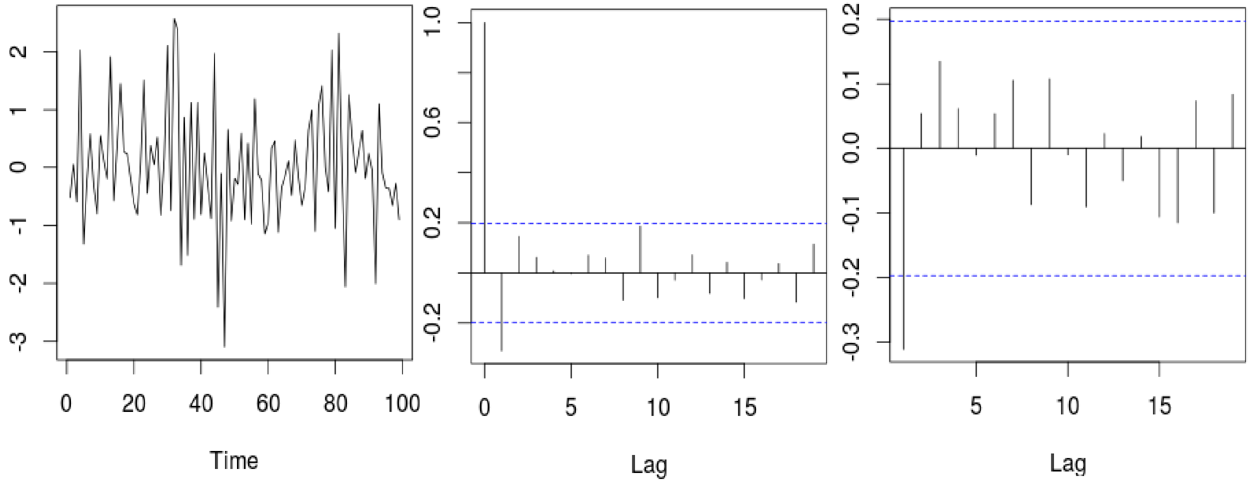
Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Estacionaridad

- Ejemplos de series estacionarias y no estacionarias y su ACF:

ARIMA(0, 1, 2), continuación

Serie sin tendencia y diferenciada en d=1



Serie **ACF** **PACF**

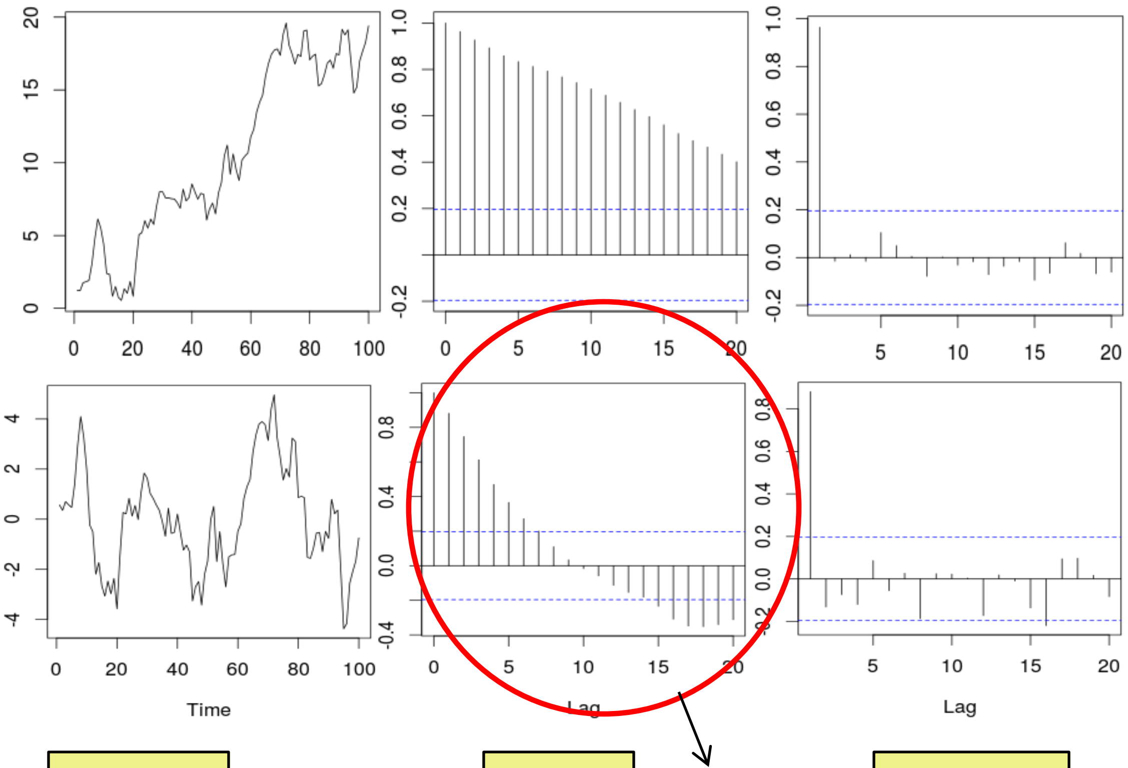
Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

- Ejemplos de series estacionarias y no estacionarias y su ACF:

ARIMA(1, 0, 0)

Serie original

Serie sin tendencia



Serie **ACF** ?? **PACF**

Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

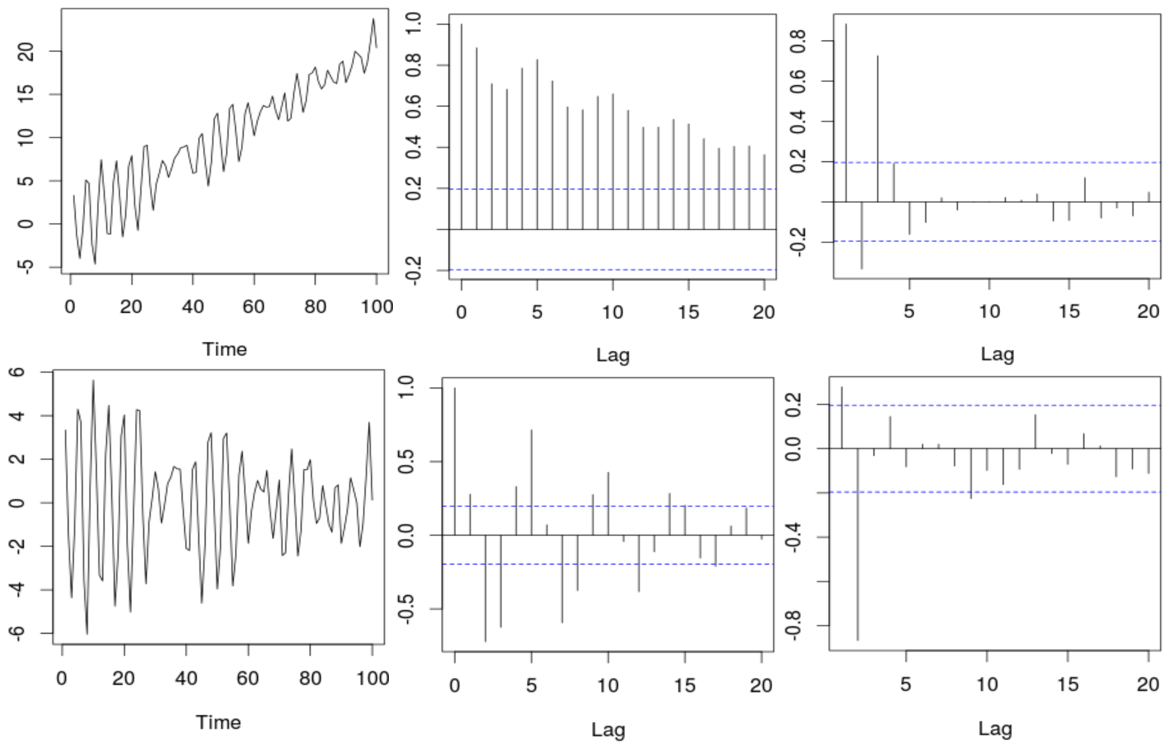
Estacionaridad

- Ejemplos de series estacionarias y no estacionarias y su ACF:

ARIMA(2, 0, 0)

Serie original

Serie sin tendencia



Serie **ACF** **PACF**

Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

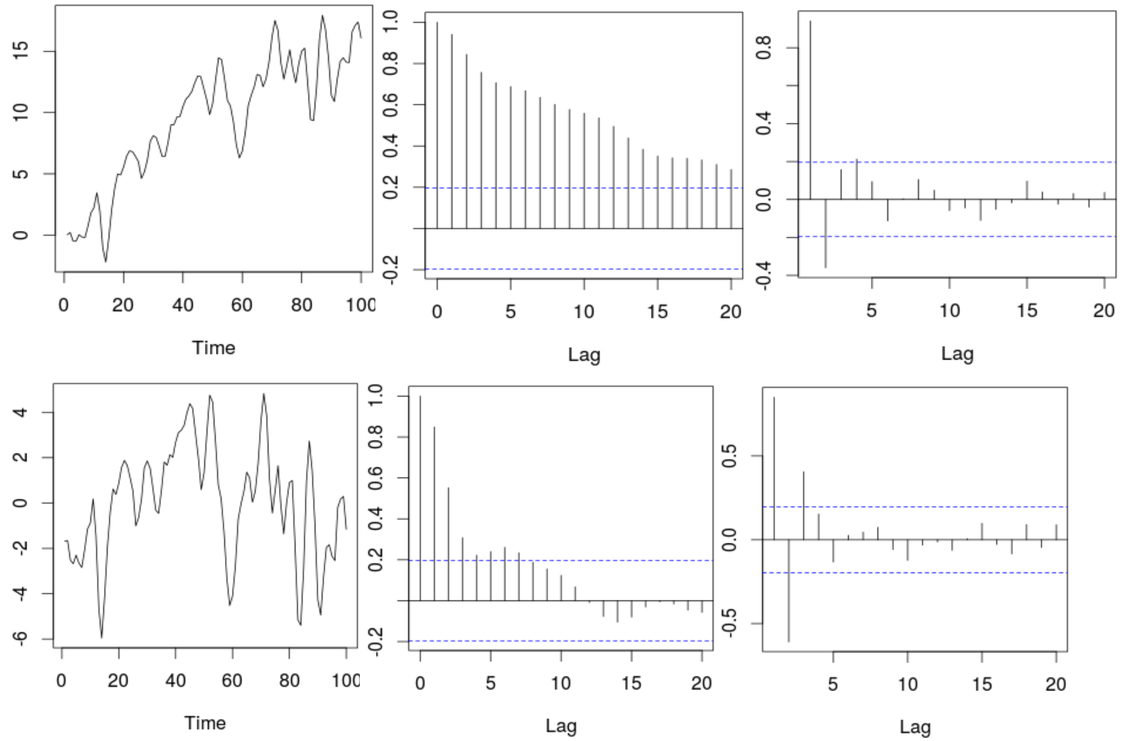
Estacionaridad

- Ejemplos de series estacionarias y no estacionarias y su ACF:

ARIMA(2, 1, 0)

Serie original

Serie sin tendencia



Serie **ACF** **PACF**

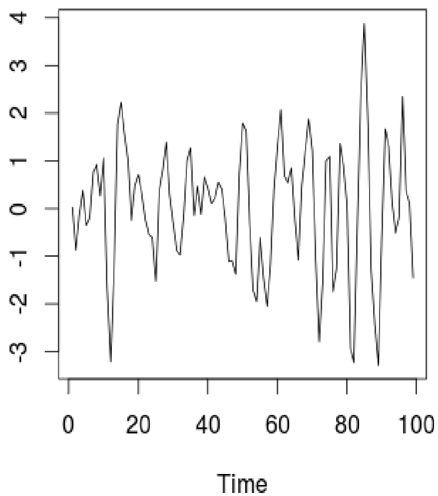
Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

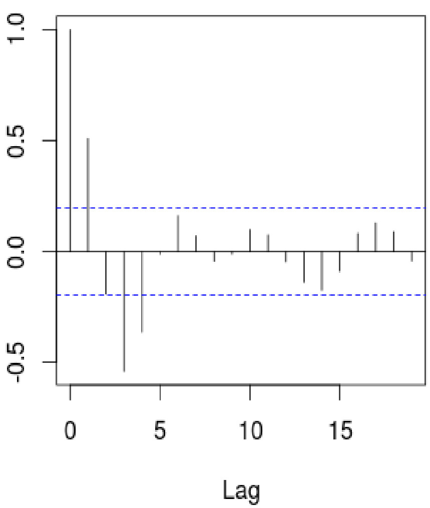
Estacionaridad

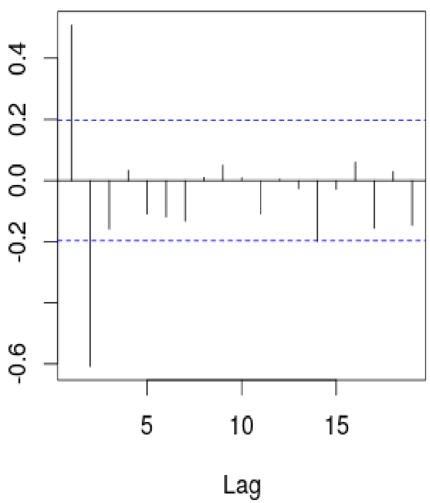
- Ejemplos de series estacionarias y no estacionarias y su ACF:

ARIMA(2, 1, 0), continuación

Serie sin tendencia y diferenciada en d=1







Serie **ACF** **PACF**

Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Índice

Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Modelos ARIMA y predicción

- Una vez seleccionados lor parámetros, para ajustar un modelo ARIMA(p, d, q) en R:
  - **Modelo<- arima(serie, order=c(p,d,q))**
- Para predecir los siguientes N valores de la serie **dentro del modelo**:
  - **Prediccion<- predict(Modelo, n.ahead=N)**
  - **DatosPredichos <- Prediccion$pred**
- Estimación de los errores de predicción:
  - **Errores <- Prediccion$se**

Modelos ARIMA y predicción

- ¿Cómo saber si el modelo que tenemos es AR o MA? ¿Y el orden?
- **Como norma general:**
  - Los modelos AR tiene un ACF que decrece a 0 (con diferentes posibles formas: regulares, sinusoidales, anternando +/-). El número del orden "p" (AR(p)) es tantos valores "distintos de 0 como haya en el PACF".
  - Los modelos MA tiene un PACF que decrece a 0 (con diferentes posibles formas: regulares, sinusoidales, anternando +/-). El número del orden "q" (MA(q)) es tantos "valores distintos de 0" como haya en el ACF.
  - **Un valor se considera "distinto de cero" si no está en el rango (-2/sqrt(N), 2/sqrt(N)), con N=longitud de la serie.**

Modelos ARIMA y predicción

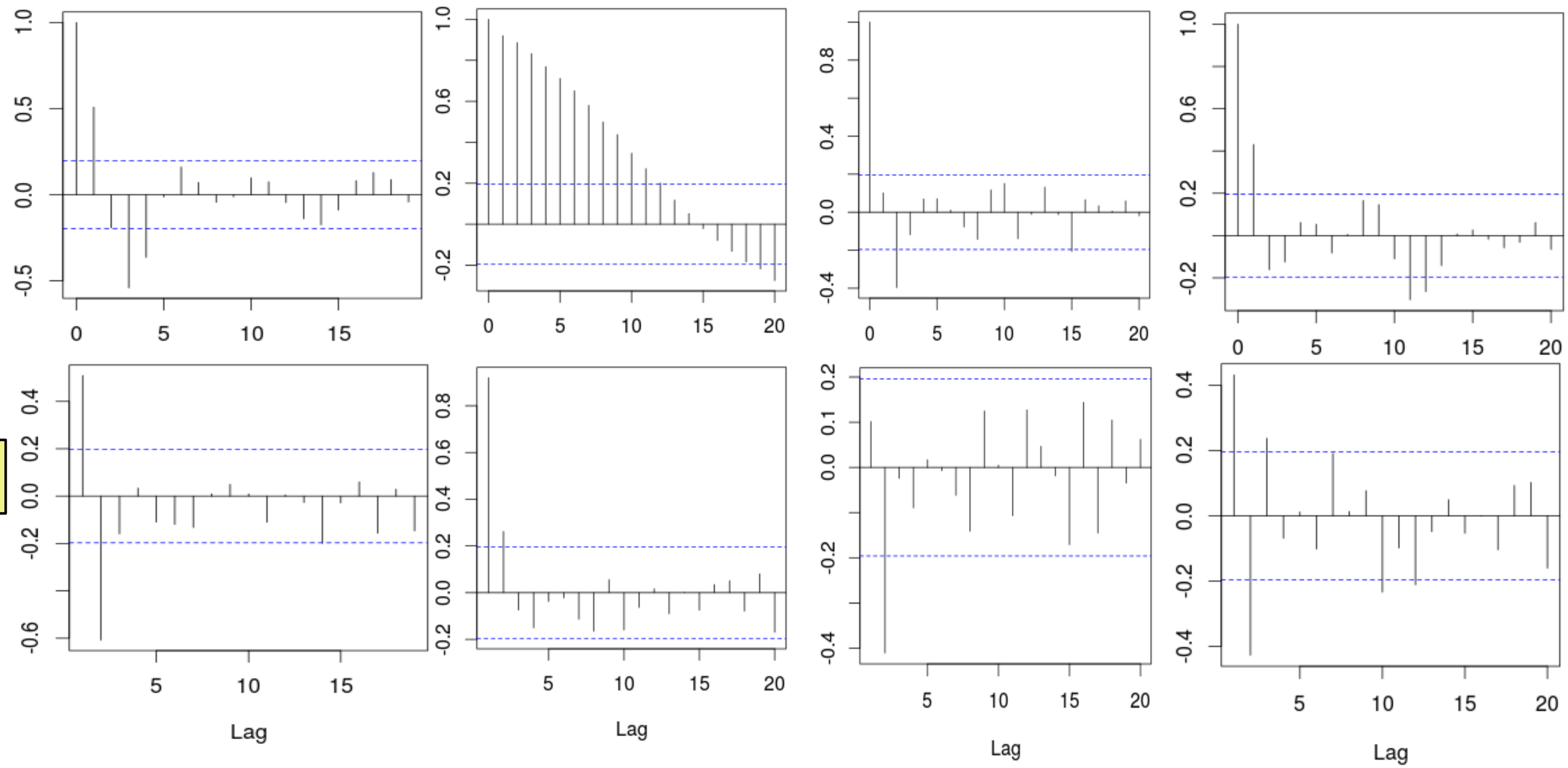
- ¿Cómo hallar el parámetro d de la integración ARIMA(p, d, q)?
- **Como norma general:**
  - Cuando la serie no es estacionaria, el ACF decrece lentamente a 0.
  - La parte integrada es necesaria normalmente para corregir la estacionaridad en la varianza.
  - Si la serie presenta tendencia lineal, normalmente con d=1 es suficiente. Si la tendencia es no lineal, puede ser necesario usar d>1.
  - Si la serie presenta estacionalidad, puede ser necesario un d=periodo de estacionalidad.

Modelos ARIMA y predicción

- Ejemplos: Cada columna muestra el ACF y el PACF de series temporales. ¿Qué modelos seleccionarías para representar estos datos? ¿Alguno necesita diferenciación?

ACF

PACF



Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

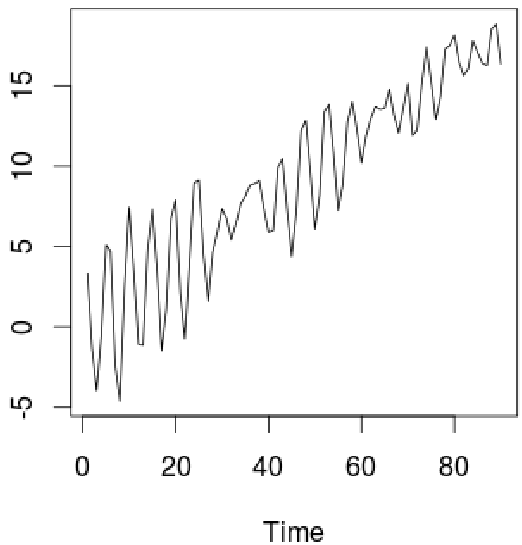
Modelos ARIMA y predicción

- Para predecir los **siguientes valores de la serie**:
  - Deshacer los cambios realizados en orden inverso
    - Diferenciaciones (si ha habido)
    - Estacionalidad (si ha habido)
    - Tendencia (si ha habido)
    - Transformaciones de los datos (si ha habido)

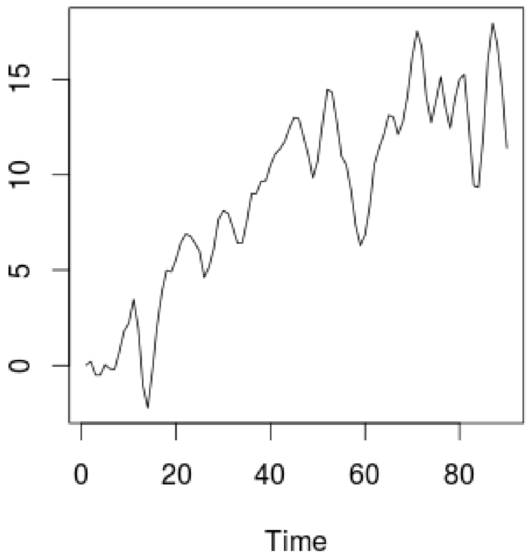
Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Modelos ARIMA y predicción

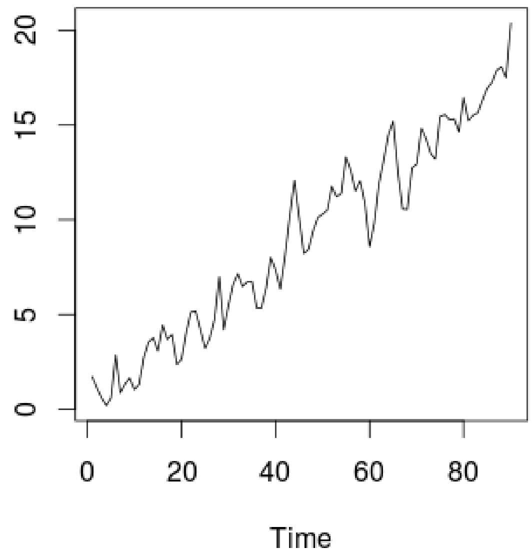
- Ejemplos: Predecir los 10 valores siguientes para las series…



Ejemplo 1



Ejemplo 2



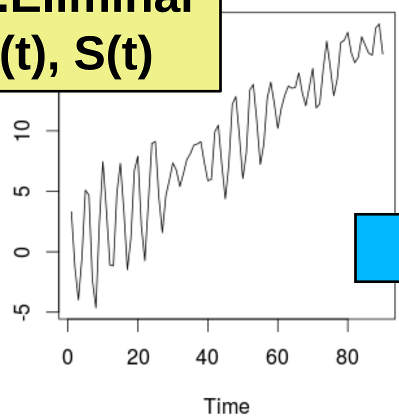
Ejemplo 3

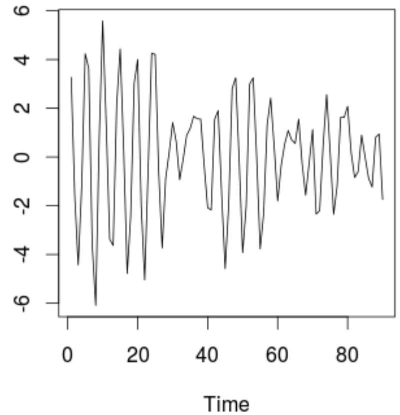
Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Modelos ARIMA y predicción. Ejemplo 1.

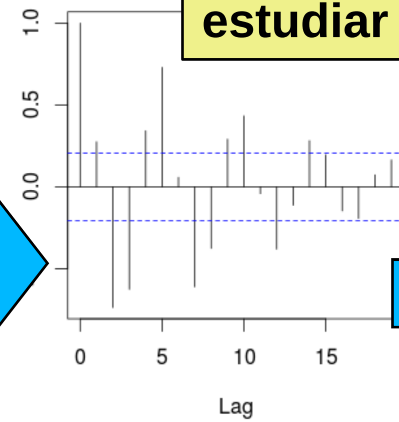
- Ejemplo 1 (ver código fuente asociado):

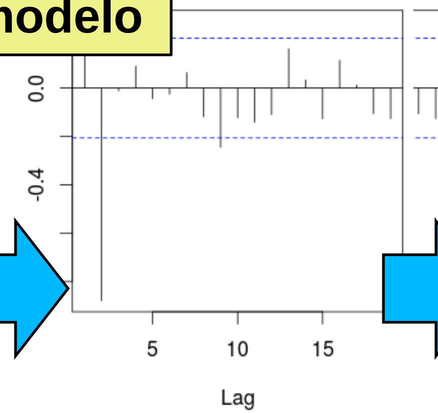
1.Eliminar T(t), S(t)



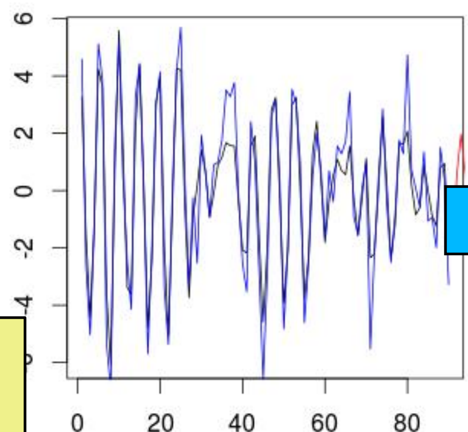


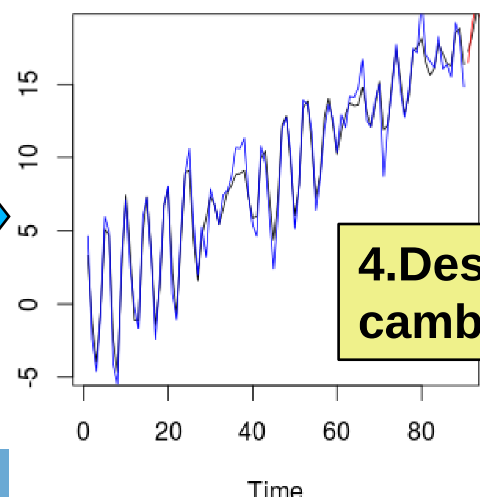
2.Comprobar estacionaridad, estudiar modelo





3.Seleccionado AR(2). Ajustar y predecir





4.Deshacer cambios

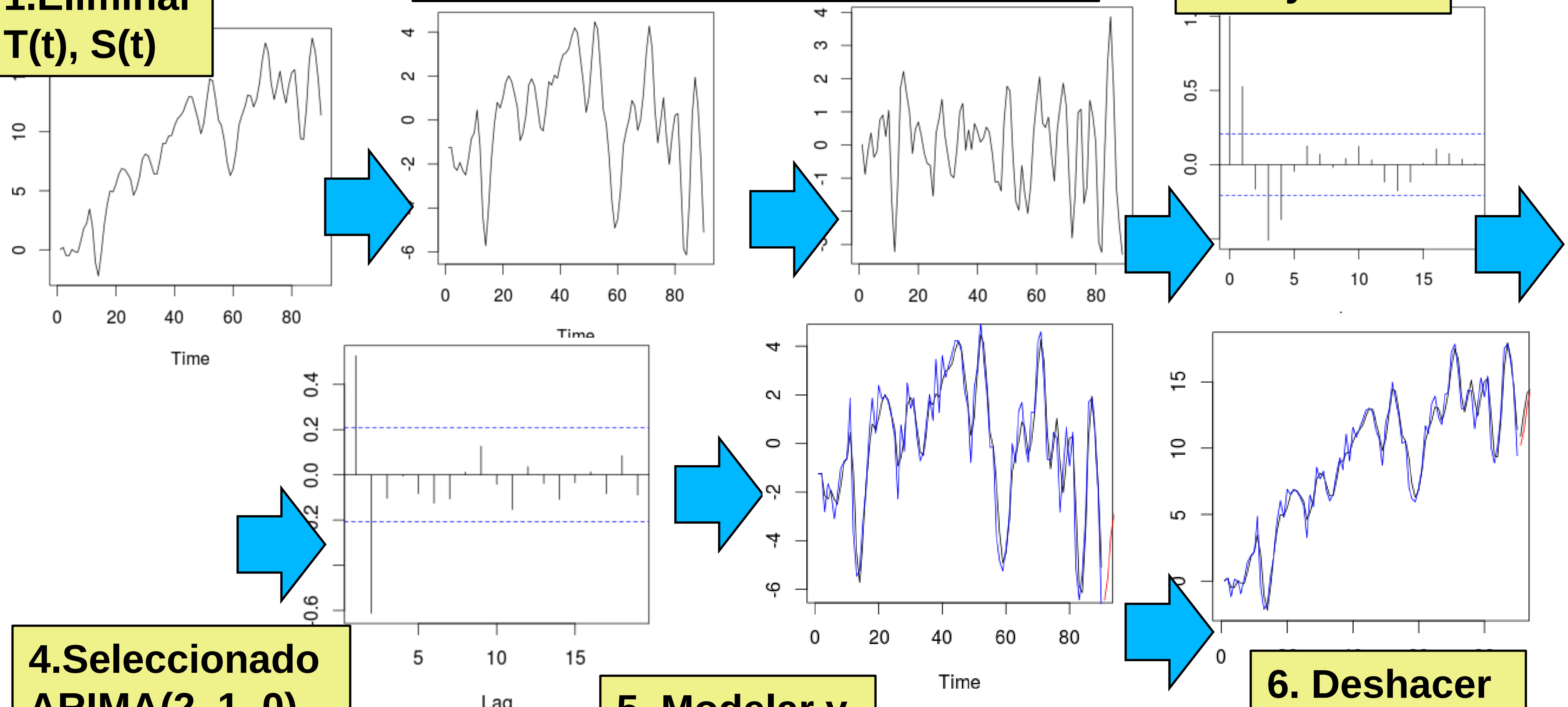
Modelos ARIMA y predicción. Ejemplo 2.

- Ejemplo 2:

1.Eliminar T(t), S(t)

2.Comprobar estacionaridad. No es estacionaria → diferenciación

3. Estudiar ACF y PACF



4.Seleccionado ARIMA(2, 1, 0).

5. Modelar y predecir

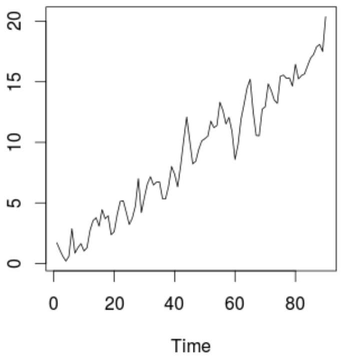
6. Deshacer cambios

Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

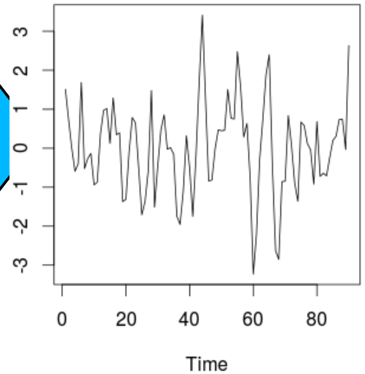
Modelos ARIMA y predicción. Ejemplo 3.

- Ejemplo 3:

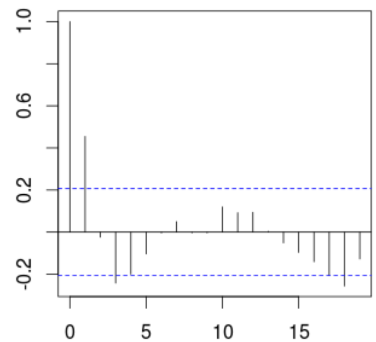
1.Eliminar T(t), S(t)

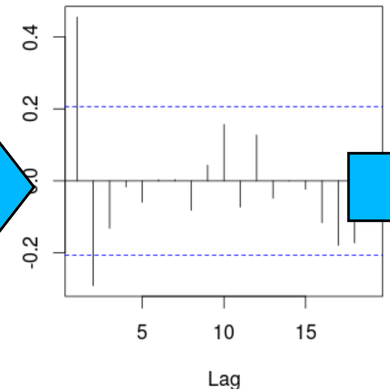


2.Comprobar estacionaridad



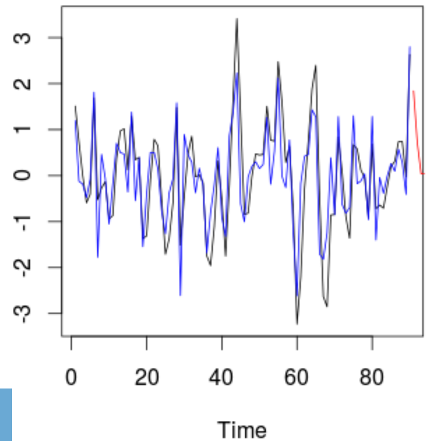
3. Estudiar ACF y PACF



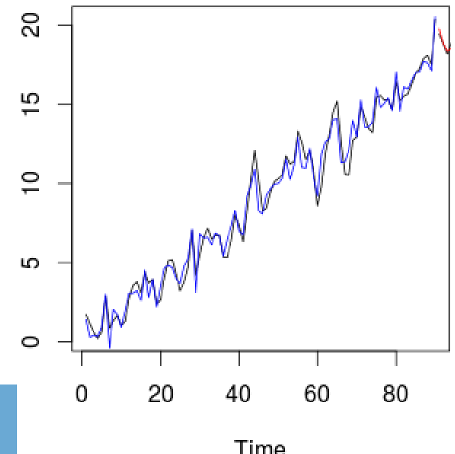


4.Seleccionado MA(2)

5. Modelar y predecir



6. Deshacer cambios



Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Índice

Selección del mejor modelo

- Normalmente, tendremos varios modelos entre los que decidir cuál puede ser el mejor.
- Es necesario un criterio para poder discriminar entre 2 modelos diferentes de representación de una serie (no sólo considerando la bondad del ajuste, sino también la complejidad del modelo).
- Usaremos el **criterio de Akaike (AIC)**. A menor valor de AIC, más preferible es el modelo.

$$AIC = 2k + n \, Log(RSS/n)$$

  - **k**= grados de libertad del modelo
  - **n=** número de datos en el modelo
  - **RSS=** Residual Sum of Squares (suma de errores al cuadrado)

Selección del mejor modelo

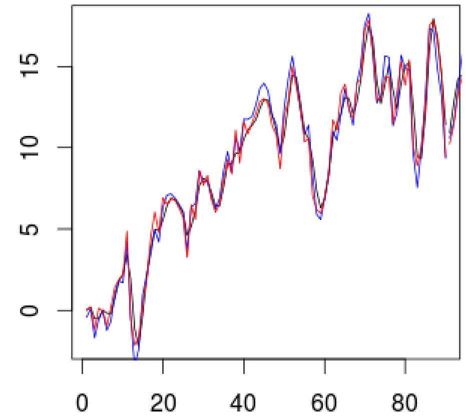
- También hay que comprobar la bondad del ajuste mediante el análisis de los residuos, mediante tests de aleatoriedad y normalidad.
  - **Tests de Box-Pierce/Ljung-Box:** Comprueban la aleatoriedad (p-value pequeño= no son aleatorios).
    - **En R:** Box.test(datos)
  - **Tests de Shapiro-Wilk/Jarque Bera:** Comprueban la normalidad de los datos (p-value pequeño= no son de distribución normal).
    - **En R:** shapiro.test(datos) / jarque.bera.test(datos)

Selección del mejor modelo

- El criterio de información de Akaike en R:
  - **AIC(modelo1, modelo2, …)**
    - Resultados de AIC para el ejemplo 2 (ver código fuente):
    - Se prueban modelo 1 (AR(2)) y modelo 2 (ARIMA(2, 1, 0))

```
        df      AIC
modelo1  4 264.9070
modelo2  3 236.9432
```

Se prefiere el modelo 2 (aunque es más complejo, produce menos errores)



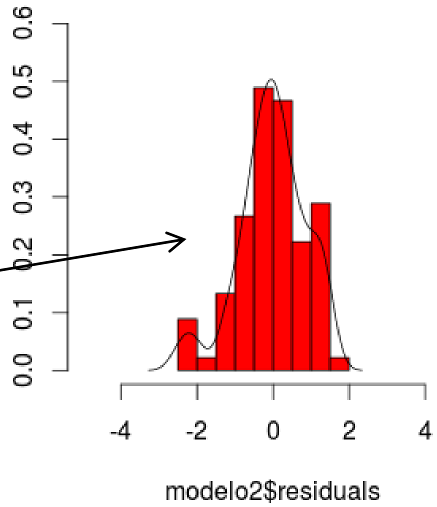
Serie original (negro)
Serie ajustada con modelo 1 (azul)
Serie ajustada con modelo 2 (rojo)

Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

Selección del mejor modelo

- En el análisis del ejemplo 2:
  - El modelo 1 no supera el test de aleatoriedad: Los residuos no son aleatorios. Es un mal modelo.
  - Ambos modelos superan el test de normalidad de Jarque Bera, pero no el de Shapiro-Wilk. No se puede asegurar la normalidad, ni la no normalidad de los datos.

Función de densidad de los residuos del modelo 2

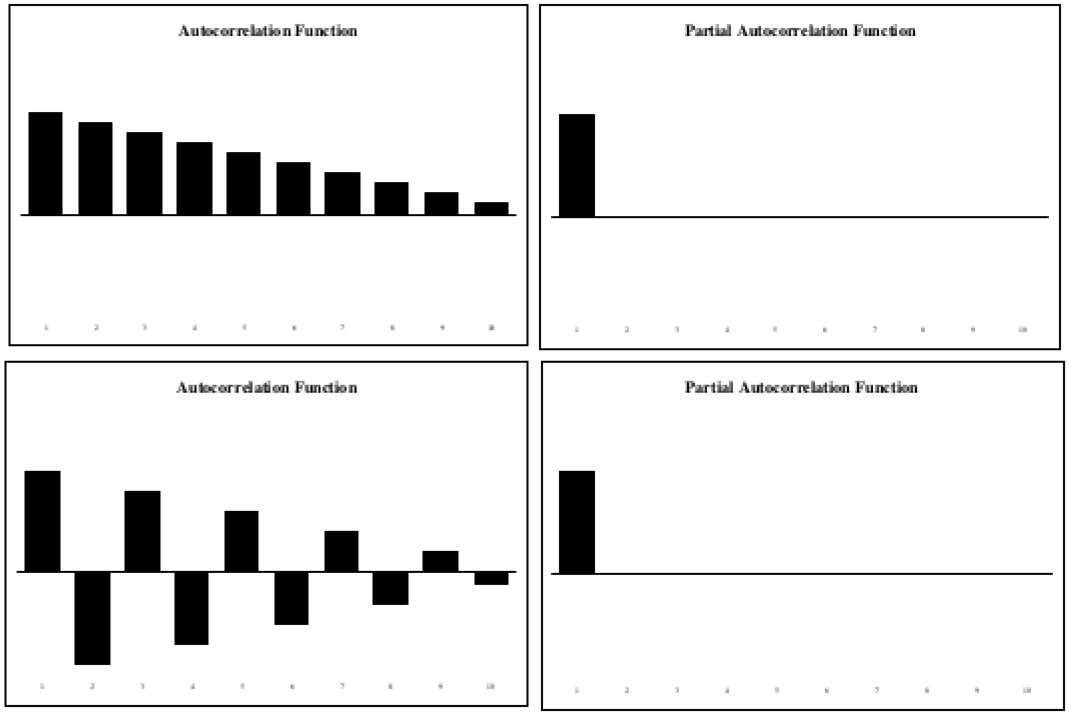


Selección del mejor modelo. Además…

- Al comienzo del estudio, conviene dividir la serie en 2 partes:
  - **Modelado/entrenamiento:** Los primeros valores de la serie, que se usarán para ajustar los modelos
  - **Test:** Los últimos valores de la serie (20% de los datos, N últimos valores, etc., depende de cada problema y caso y es decisión de diseño).
- A la hora de modelar la tendencia, conviene utilizar métodos de validación (por ejemplo, **cross-validation**), para comprobar la eficacia de las hipótesis de la tendencia.
- La técnica de validación más usada es **10-fold cross-validation**.

Resumen selección del modelo

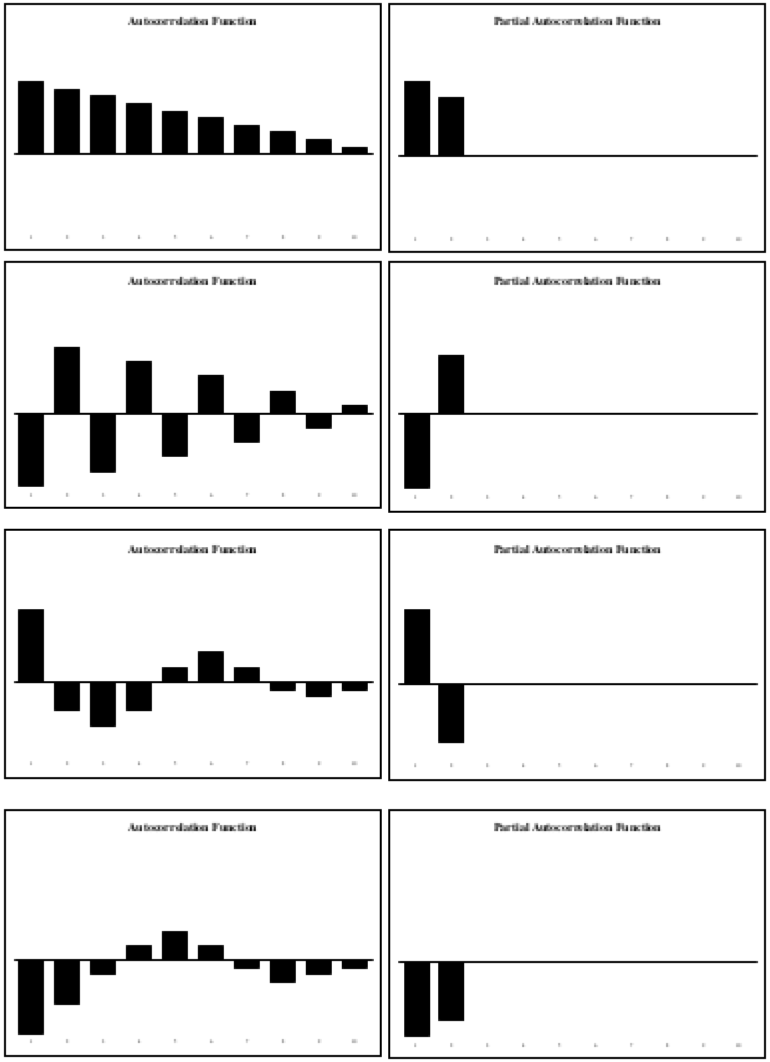
- AR(1)



Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada

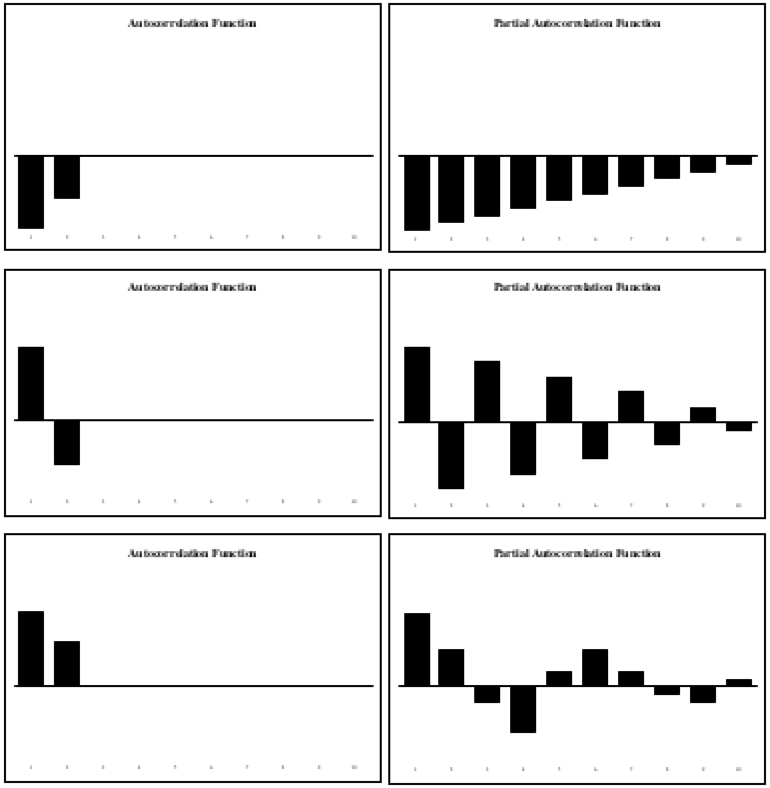
Resumen selección del modelo

- AR(2)



Resumen selección del modelo

- MA(2)



Resumen selección del modelo

- Tabla resumen:

Facp / *Fac*	**Decrece**	**Un coeficiente significativo**	**Dos coeficientes significativos**
Decrece	**ARMA(1,1)**	**AR(1)**	**AR(2)**
Un coeficiente significativo	**MA(1)**	**--**	**--**
Dos coeficientes significativos	**MA(2)**	**--**	**--**

Series Temporales y Minería de flujos de datos

Seminario: Software para procesamiento de series temporales y flujo de datos

Parte I: Series Temporales

Manuel Pegalajar Cuéllar (manupc@decsai.ugr.es)
Dpto. Ciencias de la Computación e I.A., Universidad de Granada