

# ***Tema 7 – Series Temporales***

---

***Máster en Big Data***

***Data Science & Advanced Analytics***

***Montse Llos i Bombardó***

## *o.- Recordatorio*

*Algoritmos supervisados de Clasificación*

*Metodología de trabajo*

*Algoritmos supervisados de Regresión*

*Algoritmos no supervisados*

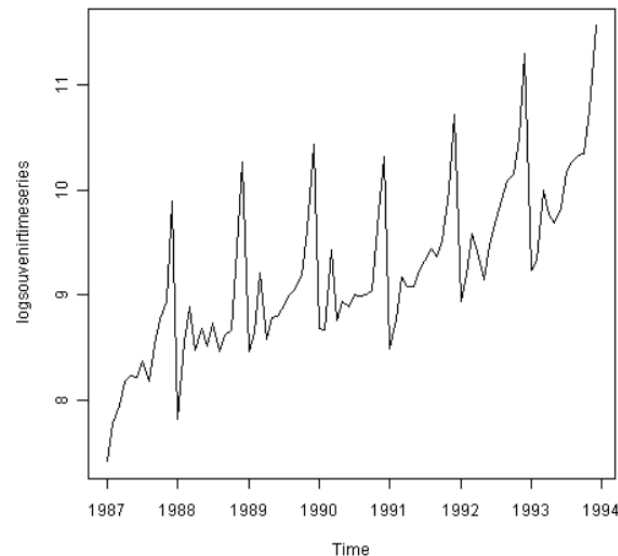
*Reinforcement Learning, Deep Learning y NLP*

# 1.- Series temporales

## ¿Qué son?

Es una sucesión de observaciones de una variable tomadas en varios instantes de tiempo.

Permite visualizar la evolución del dato (variable) a lo largo del tiempo.



# 1.- Series temporales

## ¿Con qué objetivo trabajar con series temporales?

Para predicción de datos futuros => Qué tipo de predicción?

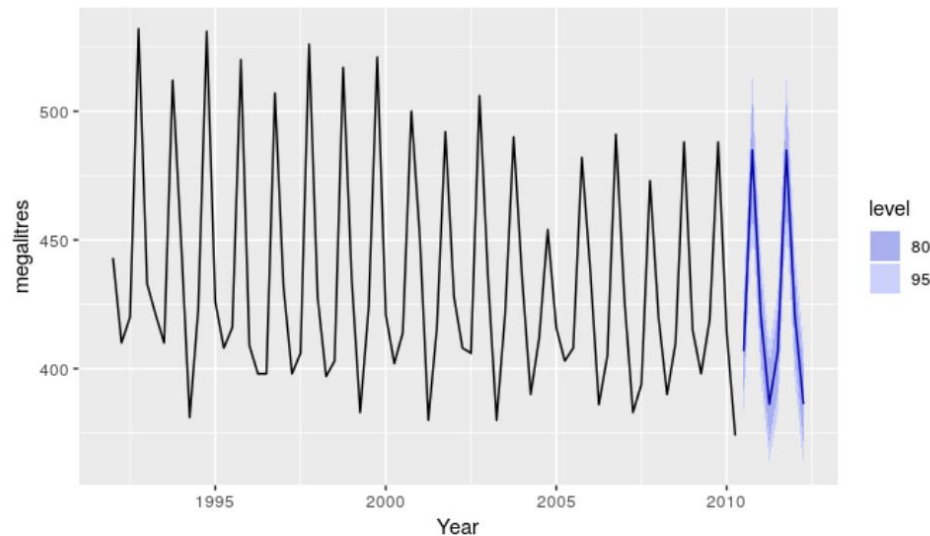


Figure 1.1: Australian quarterly beer production: 1992Q1–2010Q2, with two years of forecasts.

# ***1.- Series temporales***

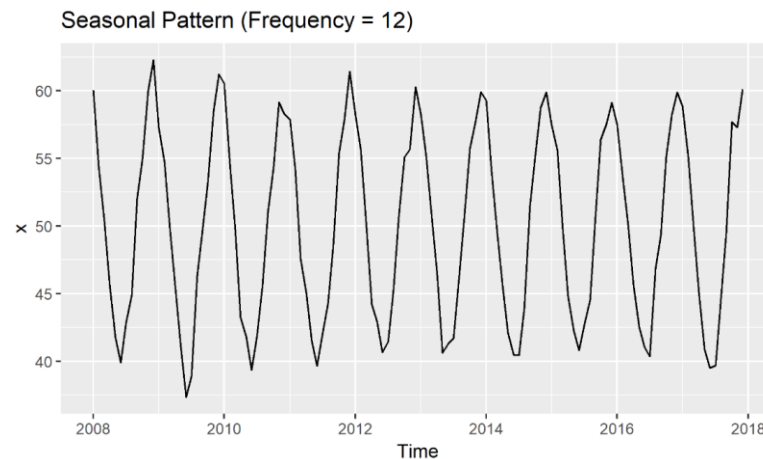
## ***Diferencias entre serie temporal y método supervisado de regresión***

- Supervisado de regresión
  - No influye el factor tiempo
- Serie temporal
  - El factor tiempo (ordenado) es decisivo para las predicciones futuras
  - Las series temporales son el método más antiguo de analítica predictiva

# 1.- Series temporales

## *Aspectos a tener en cuenta en una serie temporal*

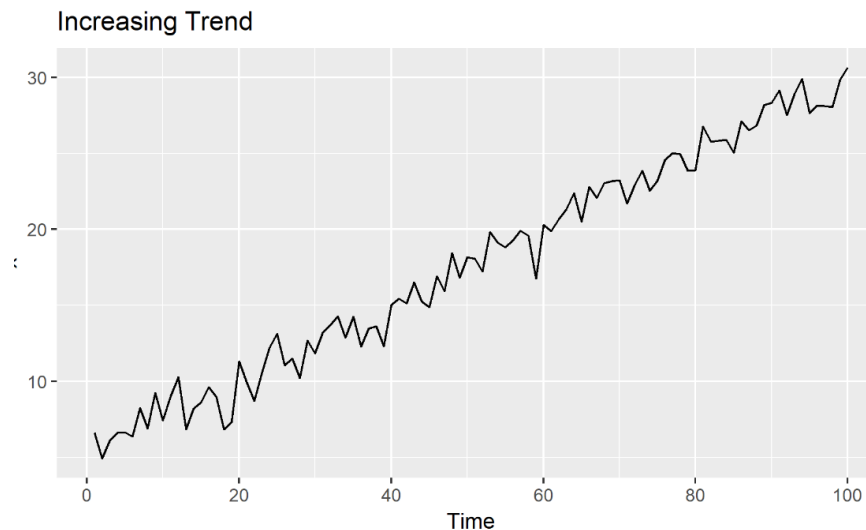
- **Estacionalidad**
  - Factores estacionales como la época del año o el día de la semana. La estacionalidad es siempre de una frecuencia fija y conocida



# 1.- Series temporales

## *Aspectos a tener en cuenta en una serie temporal*

- **Tendencia**
  - Aumento o disminución a largo plazo en los datos. No tiene que ser lineal

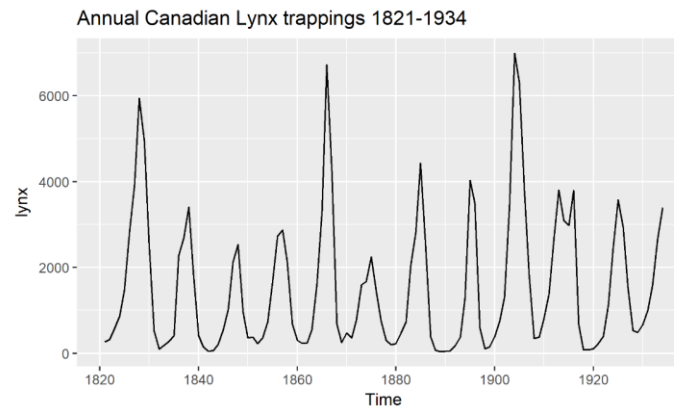


# 1.- Series temporales

## *Aspectos a tener en cuenta en una serie temporal*

- **Ciclicidad**

- Los datos muestran subidas y bajadas que no son de una frecuencia fija. Estas fluctuaciones se deben generalmente a las condiciones económicas, y a menudo están relacionadas con el "ciclo económico". La duración de estas fluctuaciones suele ser de al menos 2 años.

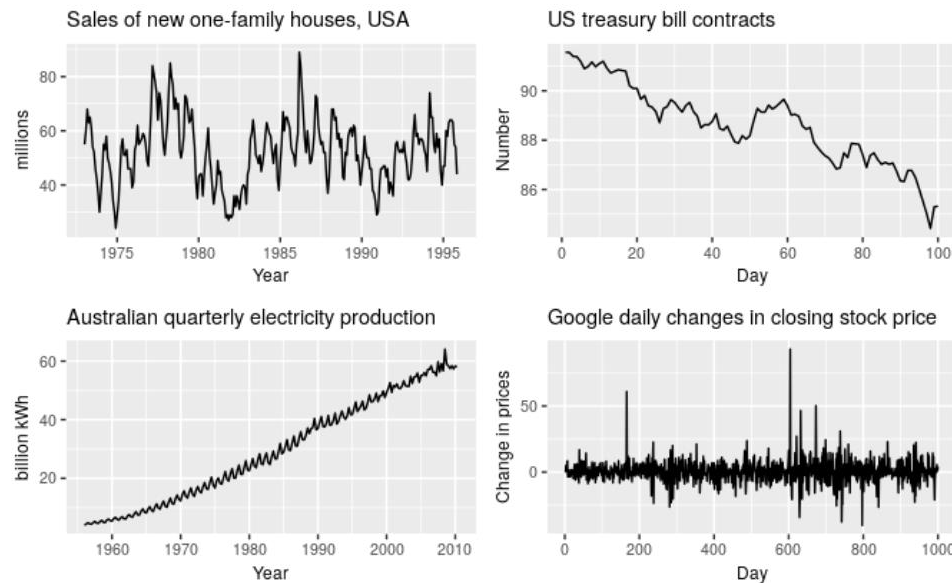




# 1.- Series temporales

## Pregunta

En las siguientes series temporales identifica estacionalidad, tendencia y ciclicidad



# ***1.- Series temporales***

## ***Tipos de problemas de predicción***

- Univariables (típicos)
- Multivariables => time series + feature engineering

## ***Cómo hacer predicción en problemas de series temporales?***

- Métodos estadísticos
- Machine Learning
  - Algoritmos supervisados de regresión
  - Deep Learning (Recurrentes y LSTM en especial)

# ***1.- Series temporales***

## ***Clasificación series temporales***

- Data-driven
  - No hay diferencia entre predictores y target
- Value-driven
  - Parecidos a los modelos tradicionales predictivos
  - Hay variables dependientes e independientes
  - La variable independiente es el tiempo

## 2.- Series temporales: Data Driven

### *Naïve Method*

Todas las predicciones son el valor de la última observación

```
naive(y, h)
```

### *Seasonal Naïve Method*

Las predicciones son el valor estacional último => snaive

### *Average Method*

Las predicciones se basan en la media de los datos históricos

```
meanf(y, h)  
# y contains the time series  
# h is the forecast horizon
```

### *Moving Average*

Usamos una ventana para calcular la media.

### *Weighted Moving Average*

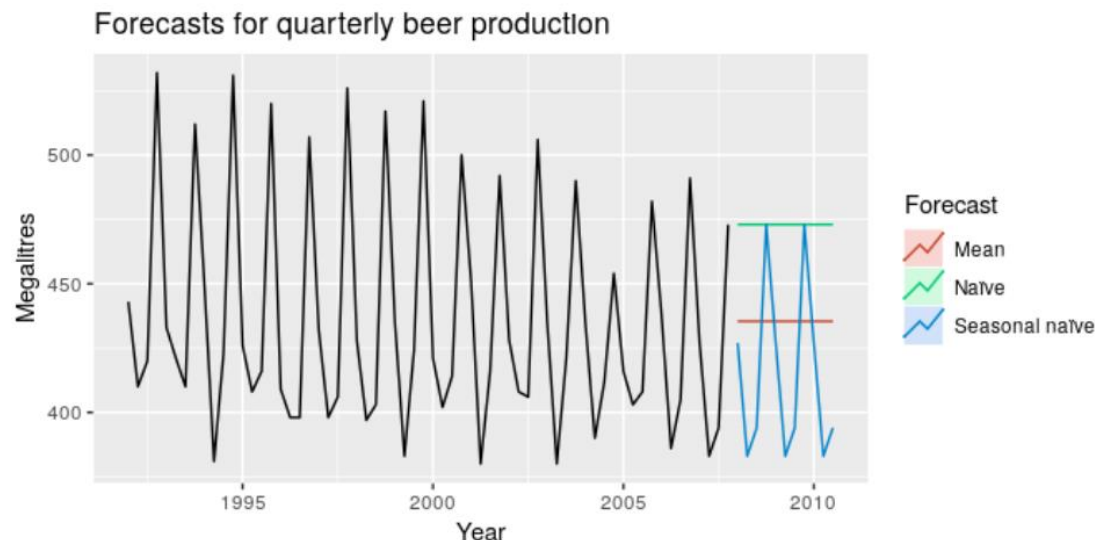
Se da un peso a cada valor de los datos históricos

Normalmente cuanto más lejos esté del tiempo, menos peso

## 2.- Series temp.: Data Driven - Práctica

### Práctica 7.1

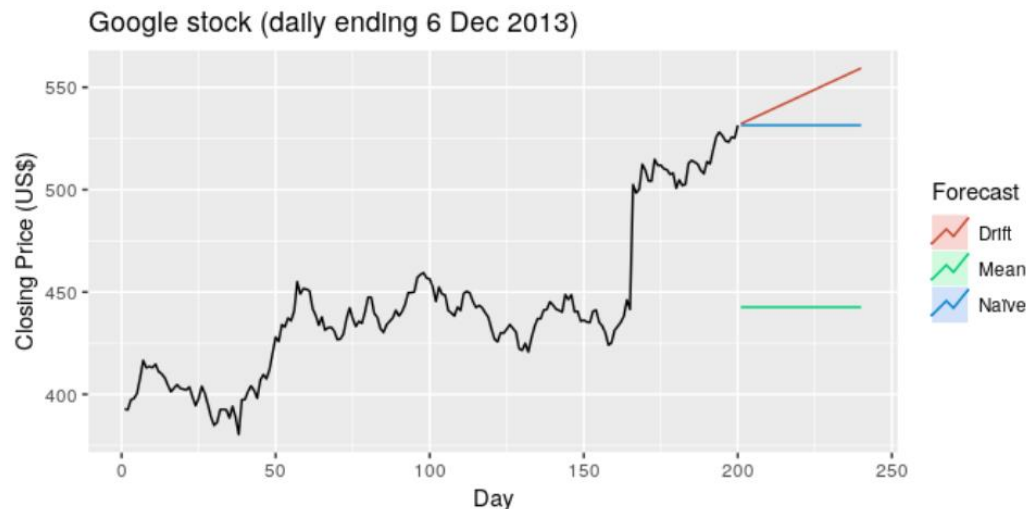
A partir del histórico trimestral de producción de cerveza en Australia, vamos a predecir con los métodos vistos anteriormente cual será la producción de cerveza en los siguientes 3 años.



## 2.- Series temp.: Data Driven - Práctica

### Práctica 7.2

Vamos a realizar el mismo tipo de predicción en otros datos. Usaremos los datos del precio de cierre de las acciones de Google. Tenemos los datos de 1000 días consecutivos, desde 25 de febrero 2013 hasta 13 de febrero de 2017. Haremos un forecast para los próximos 40 días.



## 2.- Series temporales: Data Driven

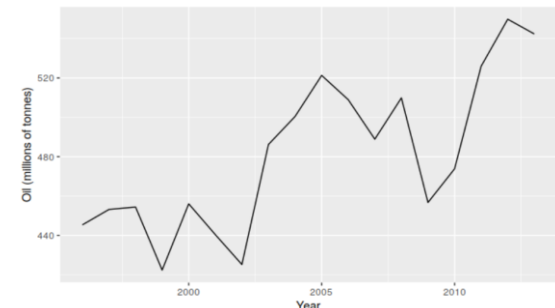
### Single Exponential Smoothing

- La predicción es el resultado de la suma de las h observaciones anteriores con un peso, pero el peso decrece exponencialmente.
- Alpha entre 0 y 1
  - Cerca de 0: el histórico se tiene + en cuenta
  - Cerca de 1: más peso a los + cercanos

```
fc <- ses(oildata, h=5)
```

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1 - \alpha)y_{T-1} + \alpha(1 - \alpha)^2 y_{T-2} + \dots,$$

- Para series sin estacionalidad



## ***2.- Series temporales: Data Driven***

### ***Limitaciones series estudiadas***

- Únicamente podemos ir un punto más allá => Data-driven temporal series
- Útil en series temporales que no tienen estacionalidad ni tendencia

### ***Holt-Winters***

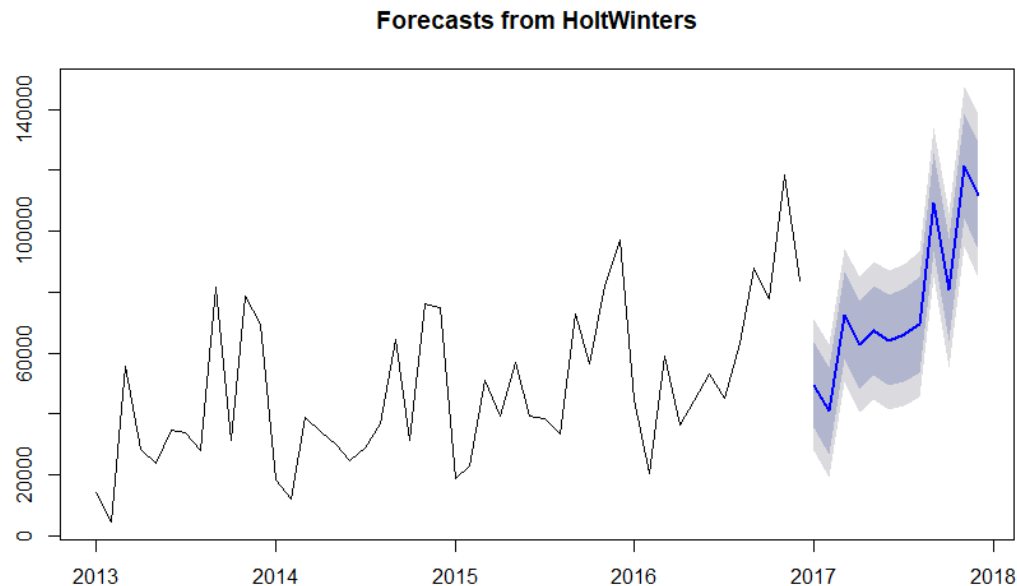
- Evolución de Exponential Smothing
- Incluye tres parámetros (dos más)
  - alpha: relevancia punto anterior
  - beta: tendencia
  - Gamma: estacionalidad



## 2.- Series temp.: Data Driven - Práctica

### Práctica 7.3

A partir de un histórico de Ventas mensual haremos la predicción de ventas del próximo año



## ***3.- Series temporales: Value-Driven***

### ***Definición***

Las predicciones de series temporales value-driven, implican fijarse en el modelo que van “dibujando” los datos o a lo largo del tiempo => Encontrar un patron para inferirlo a tiempos futuros.

Términos a tener en cuenta para el modelo

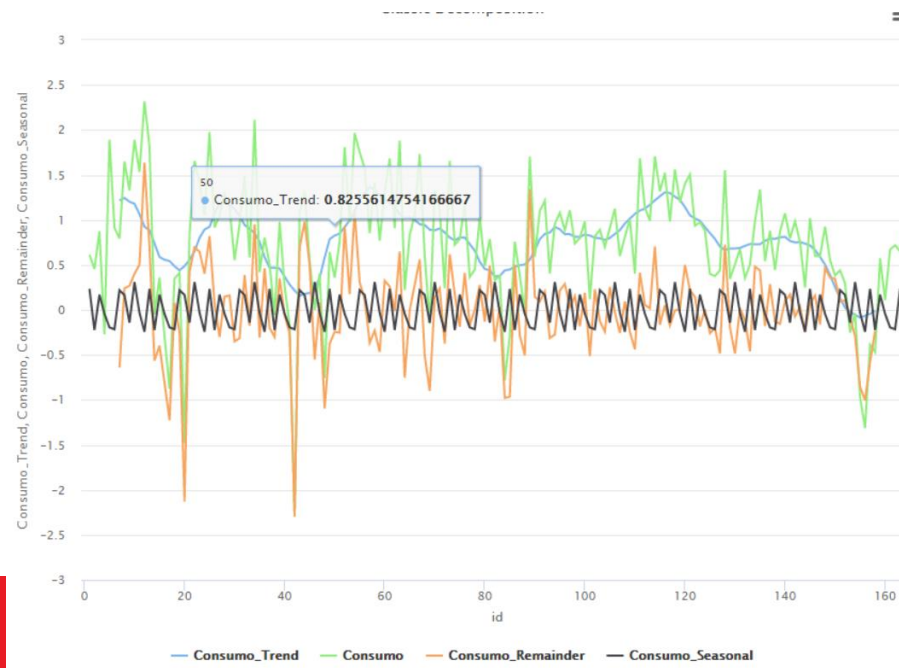
- Estacionalidad
- Tendencia

El error (que define también una serie temporal) se excluye de la predicción dado que su comportamiento se define como aleatorio

## 3.- Series temporales: Value-Driven

### Práctica 7.4

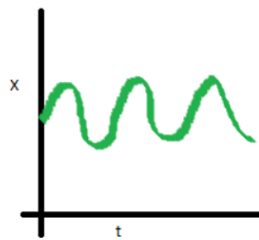
Descomponer en su estacionalidad, tendencia y error los datos históricos de consumo de leche en Estados Unidos.



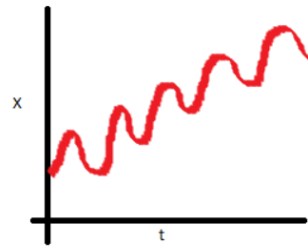
## 3.- Series temporales: Value-Driven

### *ARIMA (Autoregressive Integrated Moving Average)*

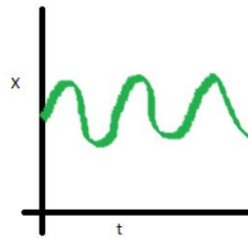
- Se ejecutan regresiones para encontrar correlaciones basadas en diferentes desfases (lags) en los datos.
- Únicamente sobre **series estacionarias**:
  - No tener tendencia.
  - Sus variaciones sobre su media tienen una amplitud constante



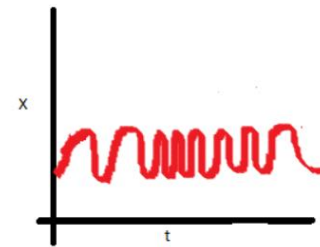
Stationary series



Non-Stationary series



Stationary series



Non-Stationary series

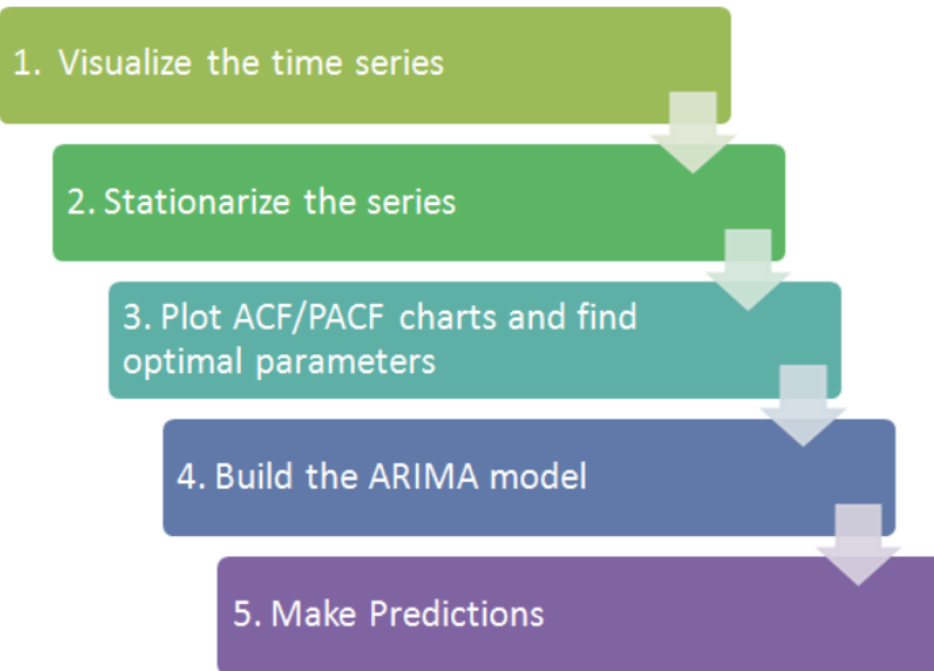
## ***3.- Series temporales: Value-Driven***

### ***ARIMA (Autoregressive Integrated Moving Average)***

- Implica un conocimiento amplio de análisis de series temporales para:
  - Ajustar la serie a estacionaria => Logs, diferencias,...
  - Definir los parámetros:
    - P (AR)
      - N° de términos autoregresivos
      - Dependencia de datos anterior
    - D (I)
      - N° de diferencias no estacionales necesarias para conseguir la estacionalidad
    - Q (MA)
      - Moving Average.

## *3.- Series temporales: Value-Driven*

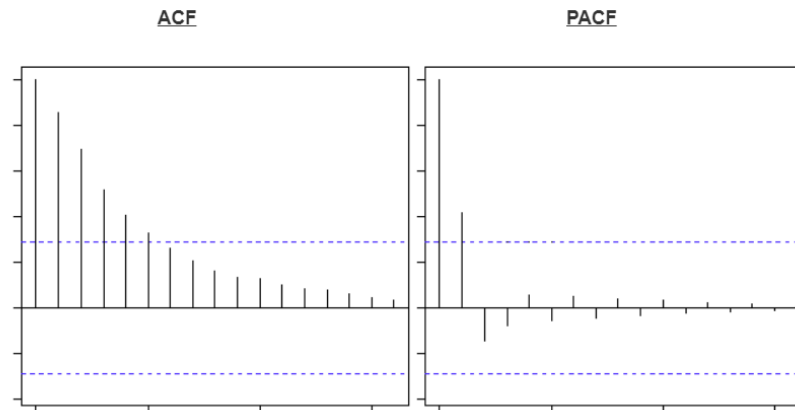
### *ARIMA (Autoregressive Integrated Moving Average)*



## 3.- Series temporales: Value-Driven

### ARIMA (Autoregressive Integrated Moving Average)

- AR  $\Rightarrow$  p (ACF) *Función de autocorrelación*
  - Mide el grado de asociación lineal que existe entre dos variables del mismo proceso estocástico
  - El valor actual está relacionado con el anterior?

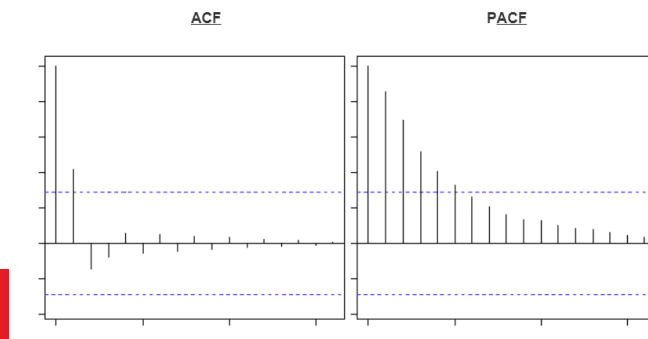


AR=2

### 3.- Series temporales: Value-Driven

#### ARIMA (Autoregressive Integrated Moving Average)

- MA  $\Rightarrow$  q (PACF) *Función de autocorrelación parcial*
  - Mide la correlación entre dos variables separadas por k periodos cuando no se considera la dependencia creada por los retardos intermedios existentes entre ambas.
  - Mide la autocorrelación que existe entre dos variables separadas k períodos descontando los posibles efectos debidos a variables intermedias
  - El valor actual está relacionado con el valor de k periodos anteriores?

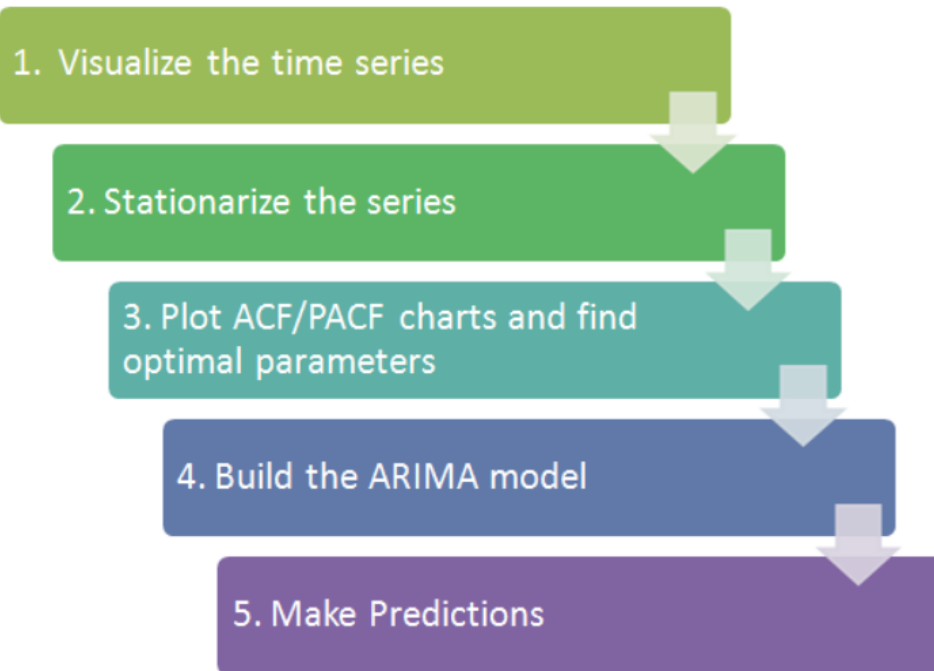


MA=2



## *3.- Series temporales: Value-Driven*

### *ARIMA (Autoregressive Integrated Moving Average)*



## 3.- Series temporales: Value-Driven

### Práct 7.5: ARIMA (Autoregressive Integrated Moving Average)

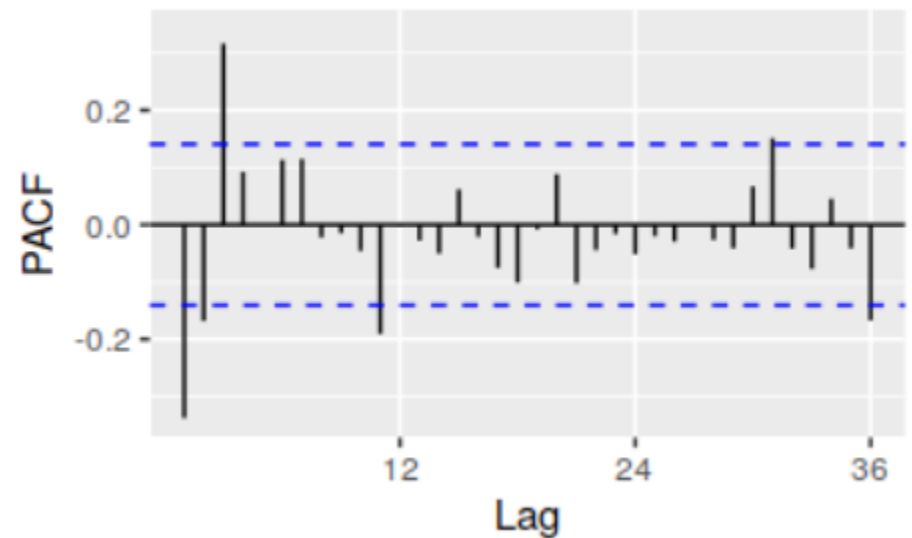
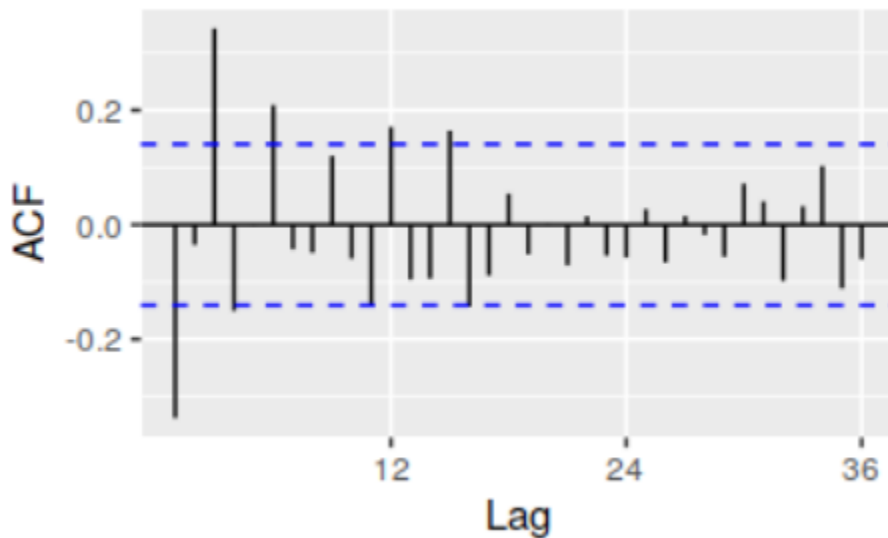
Forecast de número de pasajeros para los próximos 10 años

```
1 #Air Passengers Data (included in R)
2 plot(AirPassengers)
3
4 #Descomposición
5 plot(decompose(AirPassengers))
6
7 #instalar tseries para buscar un valor de p
8 library(tseries)
9 plot(log(AirPassengers))      #log para eliminar desviaciones desiguales
10 plot(diff(log(AirPassengers))) #diff para eliminar la tendencia
11 adf.test(diff(log(AirPassengers)), alternative="stationary", k=0)
12
13 #Auto correlation Function ACF => AR, correlacion con lag anterior
14 acf(log(AirPassengers))      #observamos correlacion en tiempo
15
16 acf(diff(log(AirPassengers)))
17 #implica p=0 (mirando pacf)
18
19 #Partial Correlation Function PACF => MA
20 pacf(diff(log(AirPassengers)))
21 #implica q=1 (mirando acf)
22
23 #ARIMA con p=0,d=1,q=1
24 fit <- arima(log(AirPassengers), c(0, 1, 1),seasonal = list(order = c(0, 1, 1), period = 12))
25
26 #Predicción para próximos 10 años
27 pred <- predict(fit, n.ahead = 10*12)
28
29 #Mostrar resultado
30 ts.plot(AirPassengers,2.718^pred$pred, log = "y", lty = c(1,3))
```

## 3.- Series temporales: Value-Driven

### Pregunta

Valor de p y q a partir de los gráficos ACF y PACF



p=3 y q=1

## ***3.- Series temporales: Value-Driven***

### ***Linear Regression / Random Forest / ...***

- Uso de algoritmos supervisados de regresión para resolución de series temporales
- Mediante transformación de los datos para adaptarlo a un modelo supervisado de regresión
- Conceptos
  - Windowing
  - Step size
  - Horizon

### 3.- Series temporales: Value-Driven

**Linear Regression / Random Forest / LSTM /... => Windowing!**

- Ejercicio . Windowing=3 / Step size=1 / Horizon=1

Fecha	Importe
08/2018	125
09/2018	150
10/2018	170
11/2018	120
12/2018	200
01/2019	100
02/2019	110
03/2019	125

Fecha	Label	Importe-2	Importe-1	Importe
10/2018	120	125	150	170
11/2018	200	150	170	120
12/2018	100	170	120	200
01/2019	110	120	200	100
02/2019	125	200	100	110
03/2019	?	100	110	125

ID      A predecir      Var 1      Var 2      Var 3

## 3.- Series temporales: Value-Driven

### *Linear Regression / Random Forest / ...*

- Ejercicio. Windowing=4, Step=1, Horizon=48

Fecha	Importe
08/2018	125
09/2018	150
10/2018	170
11/2018	120
12/2018	200
01/2019	100
02/2019	110
03/2019	125

Fecha	Label	Importe-3	Importe-2	Importe-1	Importe
11/2018	100	125	150	170	120
12/2018	110	150	170	120	200
01/2019	125	170	120	200	100
02/2019	?	120	200	100	110
03/2019	?	200	100	110	125

## ***3.- Series temporales: Value-Driven***

### ***Práctica 7.6***

- Predicción del precio de gasoil en una gasolinera, teniendo en cuentas las últimas 48h
- Usaremos RapidMiner

## 4.- Series temporales: Multivariable

### *Diversos métodos*

- Cómo resolver un serie temporal multivariable (Multivariate Time Series – MVT)

Time	Temperature
5:00 am	59 °F
6:00 am	59 °F
7:00 am	58 °F
8:00 am	58 °F
9:00 am	60 °F
10:00 am	62 °F
11:00 am	64 °F
12:00 pm	66 °F
1:00 pm	67 °F

Univariate

Time	Temperature	cloud cover	dew point	humidity	wind
5:00 am	59 °F	97%	51 °F	74%	8 mph SSE
6:00 am	59 °F	89%	51 °F	75%	8 mph SSE
7:00 am	58 °F	79%	51 °F	76%	7 mph SSE
8:00 am	58 °F	74%	51 °F	77%	7 mph S
9:00 am	60 °F	74%	51 °F	74%	7 mph S
10:00 am	62 °F	74%	52 °F	70%	8 mph S
11:00 am	64 °F	76%	52 °F	65%	8 mph SSW
12:00 pm	66 °F	80%	52 °F	60%	8 mph SSW

Multivariate



## ***4.- Series temporales: Multivariable***

### ***Diversos métodos***

- VAR (estadístico)
  - <https://www.analyticsvidhya.com/blog/2018/09/multivariate-time-series-guide-forecasting-modeling-python-codes/>
- Algoritmos supervisados de regresión
- LSTM (Deep Learning)
  - <https://machinelearningmastery.com/multivariate-time-series-forecasting-lstms-keras/>
- Ideas de aplicación en negocios?

## ***5.- Series temporales: Tips & Tricks***

### ***Definición de la predicción***

En un ámbito de ventas de productos:

- Cada producto o grupos de productos?
- Por punto de venta, o para los puntos de venta agrupados por región, o sólo ventas totales?
- Datos semanales, mensuales o anuales?
- Con qué horizonte? Un mes por adelantado, un año o 10 años?
- Con qué frecuencia?

## *6.- Resumen*

- Serie temporal
- Conceptos
  - Estacionalidad
  - Tendencia
  - Error
- Tipos de series
  - Data-driven
    - Exponential Smoothing
    - Holt Winters: Debe ser estacionaria
  - Value-driven
    - ARIMA: Debe ser estacionaria
    - Linear Regression / Random Forest
  - Multivariable