

Práctica 7.

Introducción a R:

Series Temporales

*Máster en Big Data*

*Curso 2019-2020*

***Data Science & Advanced Analytics***

*Montse Llos i Bombardó*

*R*

El lenguaje R está enfocado principalmente al análisis estadístico aunque también se usa en el campo del Data Science, la [investigación biomédica](https://es.wikipedia.org/wiki/Biomedicina), la [bioinformática](https://es.wikipedia.org/wiki/Bioinform%C3%A1tica) y las [matemáticas financieras](https://es.wikipedia.org/wiki/Matem%C3%A1ticas_financieras). A esto contribuye la posibilidad de cargar diferentes bibliotecas o paquetes con funcionalidades de cálculo y graficación. Se uso es muy extenso en el ámbito universitario para investigaciones científicas.

Gráficamente tiene una potencialidad muy alta. Python ofrecía más posibilidades en el campo del Machine Learning hasta la salida de nuevas librerías en R como mlr.

En el siguiente artículo se habla de librerías potentes en R que son poco usadas: <https://www.analyticsvidhya.com/blog/2019/04/8-useful-r-packages-data-science/?utm_source=feedburner&utm_medium=email&utm_campaign=Feed%3A+AnalyticsVidhya+%28Analytics+Vidhya%29>

Para la instalación de R necesitamos:

* Instalar el lenguaje R
  + <https://cran.r-project.org/bin/windows/base/>
* Instalar RStudio, como uno de los entornos de desarrollo más usados (también se puede usar Jupyter Notebook)
  + <https://www.rstudio.com/products/rstudio/download/>

*Práctica 7.1*

Vamos a analizar la producción quatrimestral de cerveza en Australia (en megalitros) desde 1956 hasta 2010

Utilizaremos los datos históricos de 1992 hasta 2007 para predecir los próximos 3 años usando distintas técnicas

Ausbeer es una serie temporal que podemos cargar del paquete fpp2 de R

*Práctica 7.2*

Vamos a realizar un forecast para los próximos 40 días a partir del precio de cierre de las acciones de Google. Tenemos información de 1000 días. Usaremos los 200 primeros días para predecir los 40 días siguientes.

Goog200 es una serie temporal que podemos cargar del paquete fpp2 de R

*Práctica 7.3*

Una compañía nos da las ventas mensuales de los últimos 3 años para que le realicemos un forecast para el próximo año de uno de sus productos estrella.

Los datos los tenemos en formato csv y el fichero lo podéis bajar del Campus bajo el nombre de 7.3.MonthlySales.csv

*Práctica 7.4*

Usando los datos del porcentaje de crecimiento en el consumo de leche en Estados Unidos durante varios años, descompone la serie en su tendencia, estacionalidad y error

Puedes hacerlo en R, RapidMiner o Python

Los datos están en formato csv y lo podéis bajar del Campo bajo el nombre de 7.4.Milk Consumption US Growth%

*Práctica 7.5*

A partir del Total anual de pasajeros aéreos (en millones), incluidos los pasajeros aéreos nacionales e internacionales de las compañías aéreas registradas en Australia. 1970-2016 vamos a predecir el volumen de pasajeros para los próximos 10 años usando ARIMA.

AirPassangers es una serie temporal que podemos cargar del paquete fpp2 de R

*Práctica 7.6*

A partir del precio de la gasolina de cada hora recolectado durante un año entero, el objetivo es predecir el precio de la próxima hora teniendo en cuenta el precio de las últimas 48h.

El fichero a cargar es en formato csv: 7.6.PriceGasStation