

# Practica I

*Biel Moyà*

*1 octubre de 2017*

Este ejercicio pretende evaluar las habilidades del alumno en el uso del lenguaje R y la manipulación de la información de fuentes semi-estructuradas.

## Enunciado

Para ello os pedimos que os descarguéis el *dataset* **Titanic: learning for disaster**, el conjunto de *train* sera con el que trabajareis en este ejercicio.

1. Realizar un análisis descriptivo del *dataset*: campos, tipos de datos, dimensiones del *dataset*... mostrad las 7 primeras filas.
2. Responde a estas breves preguntas:
  - Cuantos hombres había? Cuantas mujeres? Queremos saber la respuesta en valores absolutos y porcentajes
  - Cuantos menores de edad viajaban en el Titanic?
  - Cuantas personas había entre 20 y 40 años?
  - Cuantas personas embarcaron en cada puerto?
3. Realizar las siguientes funciones:
  - Crea una función que reciba como parámetro un vector y calcule su media (No se puede usar la funcion *mean*).
  - Crea una función que reciba como parámetro dos vectores y nos diga si tienen todos los elementos iguales.
  - Crea una función que reciba como parámetro un vector y nos devuelva uno nuevo sin elementos duplicados. P.ej si nos llega el vector `c(2,1,3,4,5,5,4,1)` la función nos devolverá el vector `c(2,1,3,4,5)`.
4. Cual es el numero medio de parientes de cada pasajero? Que pasajero tiene mas parientes? Cuantos pasajeros viajaban sin ningún pariente?
5. Substituye las edades que faltan con la media de las edades que ya tenemos.
6. Hablemos del precio de los tiquets:
  - Cual es la mediana del precio de los tiquets por clase?
  - Cual el el precio de los tiquets que mas veces se repite?
7. Realiza una (o varias) gráficas en la que se muestre un histograma de la edad de los pasajeros por cada una de las clases. También seria interesante ver una gráfica del numero de hombres y mujeres por cada una de las clases.
8. Realiza la(s) gráfica(s) que puedan mostrar la relación entre la edad de los pasajeros y su probabilidad de supervivencia.
9. Crea un archivo con extensión *.json* con una lista de todos los pasajeros que contenga: nombre, clase y edad. Validad que esta correctamente formado cargándolo en R.
10. Aunque aun no podéis aplicar demasiadas técnicas para para poder predecir quien se salvaría y quien no. Que conclusiones has podido sacar a partir del trabajo con este *dataset*? Para justificar tu decision puedes demostrar tus predicciones usando el conjunto de test. (**NOTA:** No se evaluara en base al porcentaje de acierto obtenido, no es la intención de este ejercicio.)

## Entrega

1. Este ejercicio se puede entregar de forma individual o por parejas.
2. El ejercicio se debe entregar antes de día 30/10/2017 a las 23:55h a través del Aula Digital.
3. El ejercicio se debe entregar en un documento pdf donde las respuestas a los diferentes apartados en los que se necesite programar se pueda observar el código R y el resultado de la ejecución de este código.
4. Se valorara la justificación / motivación / explicación de cada respuesta.