



Asociación Española para la Inteligencia Artificial (AEPIA)

**UIMP**

Universidad Internacional  
Menéndez Pelayo

---

# Práctica: Introducción a R

Máster Universitario en Investigación en  
Inteligencia Artificial

Ciencia de Datos y Aprendizaje Automático

José Hernández-Orallo

Material original cortesía de M.José Ramírez-Quintana  
ETSINF

Universitat Politècnica de València

September 16, 2016

---

1. Genera los números  $1, 2, \dots, 12$ , i guarda el resultado en el vector  $x$ .
2. Genera cuatro repeticiones de la secuencia de números  $(6, 2, 4)$ .
3. Genera la secuencia consistente en seis 9s, seguidos de cinco 2s, y finalmente cuatro 5s. Guarda los números en una matriz 5 por 3 (poblándola por columnas).

4. Genera un vector consistente en 20 números generados aleatoriamente de una distribución normal. Usa el valor 100 como semilla (para poder replicar los experimentos). Fijar la semilla se realiza como sigue:

```
> set.seed(100)
```

Ahora, calcula las siguientes estadísticas sobre el vector: media, mediana, varianza y desviación estándar.

Repite la generación del vector y las estadísticas fijando o no la semilla y observa qué ocurre.

5. De los recursos proporcionados con el curso, baja el fichero “data1.txt” que contiene información sobre estudiantes..
- (a) Lee los datos en un objeto R denominado **students** (los datos están en un fichero de texto delimitado por espacios y no hay fila de cabecera).
  - (b) Añade los siguientes títulos a las columnas:  
**height, shoesize, gender, population**
  - (c) Comprueba que R lee el fichero correctamente.
  - (d) Muestra sólo los nombres de las cabeceras.
  - (e) Muestra la columna **height**.
  - (f) Cuál es la distribución por sexos (columna **gender**), es decir, ¿cuántas observaciones hay en cada grupo? Haz lo mismo para la distribución de los lugares de muestra (columna **population**)?
  - (g) Muestra las distribuciones de la pregunta anterior al mismo tiempo mediante una tabla de contingencia.
  - (h) Crea dos subconjuntos de tu dataset partiéndolos por **gender**. Hazlo con operaciones propias de dataframes primero y luego haz lo mismo con la función **subset**. Usa la ayuda para saber cómo funciona **subset**.
  - (i) Crea dos subconjuntos individuales por debajo y por encima de la altura *mediana*. Hazlo con primitivas propias de dataframes primero y luego haz lo mismo con la función **subset**.
  - (j) Cambia la altura de centímetros a metros para todas las filas del dataframe. Hazlo de tres maneras diferentes: con primitivas propias de dataframes primero, con un bucle usando **for** y con la función **apply** (o alguna variante).
  - (k) Dibuja una gráfica (usa **plot**) **height** contra **shoesize**, usando círculos azules para hombres y cruces magenta para las mujeres. Añade una leyenda a la gráfica.