

# CURSO INTRODUCCIÓN A R

## Caso estudio

Se va a realizar el preprocesado de los datos obtenidos en un estudio donde se pretende evaluar la posible asociación entre el cáncer de mama en mujeres y una serie de variables.

Para ello se plantea llevar a cabo el siguiente pipeline de operaciones en R:

### 1. Importar la base de datos del estudio “datos.estudio.xlsx”

### 2. Realizar las siguientes recodificaciones

2.1. Sustituir los valores missing en la variable edad por el valor de 30

2.2. Sustituir el valor de “Hombre” por “Mujer” en la variable Sexo

2.3 Sustituir los valores missing en la variable estado.civil por el valor de “Soltero”

2.4 Sustituir el valor “Bajo” en la variable nivel.estudios por el valor de “Escuela”

### 3. Crear las siguiente variables

3.1. Crea una variable dentro los datos del estudio llamada “nivel.estudios.new” que tenga dos posibles valores: Alto (cuando la variable nivel.estudios toma valores de “Alto”) y Bajo (cuando la variable nivel.estudios toma los valores de Medio o Escuela)

3.2. Crea una variable numérica dentro de los datos del estudio llamada “ratio” que sea igual a el cociente entre altura y el peso (altura/peso)

### 4. Identificar cuales de los siguientes identificadores de registro están en la base de datos del estudio: 35,111,43,55,23,76,75,99

5. Crear una variable llamada ID\_interno en la base de datos del estudio cuyos valores correspondan a una secuencia numérica continua no aleatoria desde el 1 hasta el número total de registros.
6. Ordena la base de datos del estudio de forma descendente atendiendo primero a la variable “altura” y después a la variable “estado.civil”.
7. Crear una base de datos llamada “datos\_peticion” que contenga sólo aquellos registros con un nivel de estudios “Bajo” atendiendo a la variable creada anteriormente llamada “nivel.estudios.new”.
8. Exporta la base de datos llamada “datos\_peticion” en formato de R y en formato de excel con el nombre base\_depurada.