

Tarea 2: Manejo de datos (2).

1. ¿Cómo seleccionaríamos los casos impares de un banco de datos con la función `subset()`? ¿Y los casos que corresponden a las filas numeradas con un múltiplo de cinco?
2. Generar una muestra aleatoria de tamaño 100 de una normal de media 165 y desviación típica 5 (imaginar que corresponden a la altura de los hombres) y otra muestra de tamaño 40 de una normal de media 160 y desviación típica 10 (ahora de las mujeres). Crear un vector con todos los datos y un factor que identifique los individuos de uno y otro grupo.
3. Construir un `data.frame` con los datos del ejercicio anterior y nombrarlo `datos.alturas`. Exportar dicho `data.frame` a un fichero externo de nombre `datos-alturas.txt` utilizando el comando `write.table` y comprobar como queda volviéndolo a importar con `read.table` (todo ello sin salir de \mathbb{R}). Construir un nuevo vector con la transformación logarítmica de las mediciones e incluirlo en el banco de datos.
4. Utilizar la ayuda para averiguar como obtener estadísticos básicos. Obtener la mediana, la desviación típica, el máximo, el coeficiente de asimetría y curtosis de los vectores del ejercicio anterior, tanto de todos los datos como por grupos.
5. Importa a \mathbb{R} el fichero `ambiente.sav`, que tiene un formato SPSS. Observa el tipo de las variables que lo componen. Investiga el comando `by`. Utilízalo para calcular la media de la variable `PH` en cada provincia de la Comunitat Valenciana.

6. Transforma el banco de datos de la tarea anterior en un array donde cada dimensión simbolice una variable categórica. Con dicho array:
- a)* calcula la media por año y provincia de la variable estudiada.
 - b)* Calcula la varianza por provincia.
 - c)* Calcula los 3 cuartiles principales por año.
 - d)* Representa gráficamente la evolución temporal media por provincia de la variable estudiada.