

# SOC4001 Procesamiento avanzado de bases de datos en R

## Tarea 1

Ponderación: 12% de la nota final del curso

Formato: Desarrollar esta tarea en un RScript, agregando comentarios cuando sea necesario.

- 1) Instalar y cargar el paquete (desde el Script) **CarData**.
- 2) Usa la documentación del paquete **CarData** para identificar los datos correspondientes a “Self-Reports of Height and Weight”
- 3) Carga los datos y crea un objeto que los contenga. Llama tal objeto “datos\_davis”.
- 4) Muestra las primeras y las últimas 6 observaciones de la base de datos en la consola.
- 5) Crea una base de datos que contenga sólo las variables **sex**, **height** y **repht** de “datos\_davis”. Llama tal objeto “subdatos\_davis”. Muestra las dimensiones de la nueva bases de datos.
- 6) Presenta un resumen estadístico (summary) de las variables en “subdatos\_davis”.
- 7) Crea una variable llamada “diff” que mida la diferencia entre el peso real y el peso reportado por los individuos y añadela a “subdatos\_davis”.
- 8) Chequea la presencia de valores perdidos en la variable “diff”. Luego crea una nueva base de datos que contenga sólo las observaciones con datos completos en todas las variables en “subdatos\_davis”. Llama este objeto “subdatos\_davis\_full” y presenta un resumen estadístico (summary) de las variables en “subdatos\_davis\_full”.
- 9) Crea una nueva variable llamada “sex\_num”. Asigna valor 1 a “sex\_num” para aquellas observaciones en las cuales la variable “sex” toma valor “F” (mujer). Asigna valor 0 a “sex\_num” para aquellas observaciones en las cuales la variable “sex” toma un valor “M” (hombre).
- 10) Usa un loop para calcular la media de la variable “diff” para las observaciones en cada uno de los niveles de la variable “sex” (es decir, para hombres y mujeres). No olvides usar el comando **print()** para mostrar los cálculos ejecutados dentro del loop.
- 11) Explica MUY brevemente el resultado obtenidos en la pregunta anterior.