

SOC4001 Procesamiento avanzado de bases de datos en R

Tarea 1

Ponderación: 12% de la nota final del curso Entrega: Desde el momento de entrega, los estudiantes tiene 1 exacta semana de plazo para completar esta tarea. 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 “Duncan’s Occupational Prestige Data”
- 3) Carga los datos y crea un objeto que los contenga. Llama tal objeto “datos_duncan”.
- 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 **type** and **prestige** de “datos_duncan”. Llama tal objeto “subdatos_duncan”. Muestra las dimensiones de la nueva bases de datos.
- 6) Presenta un resumen estadístico (summary) de las variables en la base de datos.
- 7) Chequea la presencia de valores perdidos en la variable “education”. Luego, calcula la media de la variable “education” y almacénala en un objeto llamado “education_promedio”.
- 8) Crea una nueva variable llamada “educacion_agg”. Asigna valor 1 a “educacion_agg” para aquellas observaciones en las cuales la variable “education” toma un valor por sobre la media. Asigna valor 0 a “educacion_agg” para aquellas observaciones en las cuales la variable “education” toma un valor igual o menor a la media.
- 9) Usa un loop para calcular la media y la desviación estándar de la variable “prestige” para las observaciones en cada uno de los niveles de la variable “educacion_agg”. No olvides usar el comando **print()** para mostrar los cálculos ejecutados dentro del loop.
- 10) Crea un scatterplot de las variables “prestige” and “education”. Dale un nombre informativo a cada eje.