

Pauta Tarea 2

Objetivo: Se espera que los estudiantes demuestren el aprendizaje de las herramientas básicas sobre procesamiento, limpieza y manipulación de datos en el software Rstudio (aprendidas en las sesiones 2 y 3 del Laboratorio), además de que analicen variables a través de medidas de tendencia central, dispersión y posición (aprendidas en sesión 4 del Laboratorio), sabiendo interpretar y reflexionar críticamente sobre los resultados obtenidos.

Instrucciones:

1. Creación de un Proyecto de R
 - 1.1. Crear un .Rproject llamado “tarea02ED” y establecer una estructura de carpetas siguiendo el formato IPO, esto es, una carpeta de input, procesamiento y output.
 - 1.2. Guardar en la carpeta input la base de datos adjuntada por el equipo docente en U-Cursos; crear en la carpeta procesamiento un script llamado “01proc-data” y otro script llamado “02analisis”.
2. En el script de “01proc-data”:
 - 2.1. Instalar y cargar librerías correspondientes vía `sintaxis`.
 - 2.2. Cargar la base de datos `casen2020_subset` vía `sintaxis`.
 - 2.3. Visualizar base de datos vía `sintaxis`.
 - 2.4. Visualizar los nombres de todas las variables de base de datos vía `sintaxis`.
 - 2.5. Seleccionar las variables `sexo` (`sexo`), `edad` (`edad`), `inmigrante` (`inmigrante`), `sistema de salud` (`sist_salud`), `empleo` (`o15`), `ingreso autónomo` (`yaut`), `nivel de escolaridad` (`educ`), `decil autónomo nacional de ingresos` (`dau`), y almacenarlas en una base de datos creada vía `sintaxis` y llamarla `datos_proc`.
 - 2.6. Filtrar la variable `edad` vía `sintaxis`, seleccionando a las observaciones mayores o iguales a 15 años.
 - 2.7. Recodificar las variables `inmigrante`, `sistema de salud` y `nivel de escolaridad` vía `sintaxis`, dejando como valores perdidos las observaciones 9 y 99, respectivamente.
 - 2.8. Identificar y remover valores perdidos (NA 's) vía `sintaxis`.
 - 2.9. Generar una nueva variable `tramo_edad` vía `sintaxis`, con los tramos: menor o igual a 29, entre 30 y 59, y mayores a 60.
 - 2.10. Guardar base de datos procesada en formato `RData` (`.RData`) y exportar esta base procesada a la carpeta `output` de su proyecto, ambos vía `sintaxis`.
 - 2.11. Limpiar el entorno de trabajo vía `sintaxis`.
3. En el script de “02analisis”:
 - 3.1. Instalar y cargar librerías correspondientes vía `sintaxis`.
 - 3.2. Cargar base de datos procesada alojada en carpeta `output` de su proyecto, vía `sintaxis`.

- 3.3. Obtener una tabla de frecuencias que incluya frecuencia absoluta, relativa y porcentajes, para cada una de las variables que correspondan según su nivel de medición, vía sintaxis.
- 3.4. Calcular las siguientes medidas de tendencia central para las variables que correspondan según su nivel de medición, vía sintaxis: a) Moda; b) Media; c) Mediana.
- 3.5. Calcular las siguientes medidas de dispersión para las variables que correspondan según su nivel de medición, vía sintaxis: a) Rango; b) Varianza; c) Desviación típica; d) Coeficiente de variación.
- 3.6. Calcular las medidas de posición: a) Valor mínimo y máximo; b) cuartil 1; c) cuartil 3.
- 3.7. Limpiar el entorno de trabajo vía sintaxis.
4. Reporte e interprete los resultados obtenidos en los puntos anteriores (3.3, 3.4, 3.5 y 3.6) en la misma sintaxis usando `#`.

Entregue como productos de esta tarea:

- Un archivo .zip o comprimido del proyecto, el cual contenga los elementos solicitados: .Rproject, carpetas de Input (con la base de datos original), Procesamiento (con los respectivos códigos de procesamiento y análisis) y de Output (con la base de datos procesada).

Aspectos formales:

- Fecha y formato de entrega: domingo 21 de mayo hasta las 23:59 vía módulo Tareas en plataforma U-Cursos.
- Décimas: Se otorgarán 3 décimas por asistencia a la sesión de Laboratorio del día jueves 18 de mayo.
- Entregas atrasadas hasta las 23:59 del martes 23 de mayo, pero tendrán 5 décimas de descuento sobre la nota final por cada día de atraso.
- Luego del martes 23 de mayo no se aceptarán más trabajos y quienes no entreguen serán evaluados con nota 1.0.
- En caso de detectarse plagio, la tarea será evaluada con nota 1.0. Esta falta puede resultar causal de reprobación de la asignatura.

En caso de tener dudas, no dude en contactar a sus ayudantes respectivos, o bien, vía foro U-Cursos al equipo docente de la asignatura.