

Práctica 1. Exploración de datos I.

Universidad de Alcalá. Curso 2023-24.

Estadística (650008). Grado en Biología sanitaria.

Actualizado: 2023-10-05

Introducción

Este fichero csv contiene una

Trabajaremos con una tabla de datos procedente de una muestra de 1000 mujeres que participaron en un estudio sobre osteoporosis. El fichero contiene algunas variables auxiliares en las columnas iniciales, pero nosotros nos vamos a fijar en estas.

- edad (en años).
- peso (en kg).
- talla (altura en cm).
- imc (índice de masa corporal)
- bua (resultado de la exploración densitométrica)
- clasific (normal / osteopenia / osteoporosis)
- menarqui (edad primera menstruación, en años)
- edad_menop (edad inicio menopausia, en años)
- menopausia (sí, no)
- tipo de menopausia
- nivel educativo

Organiza tu entorno de trabajo

- Para leer los datos, copia, pega en tu script y ejecuta este código

```
mi_url = "https://marcos-marva.web.uah.es/CursoSanitaria/practicadas/datos/osteoporosis.csv"
osteoporosis = read.table(file = mi_url,
                          sep = "\t", dec = ",", header = TRUE)
```

- En la próxima práctica aprenderemos cómo leer datos de ficheros.

En lo que sigue, recuerda que para acceder a los datos en las columnas del data.frame que has creado en el paso anterior puedes utilizar la notación `nombre_tabla$nombre_variable`.

Ejercicio 1

Sobre la variable `tipo_men`:

1. ¿De qué tipo es la variable?
2. Usa la función `unique()` para determinar cuántos valores distintos toma.
3. Construye las tablas de frecuencias absolutas y relativas.
4. Para esta variable, ¿tienen sentido las tablas de frecuencias acumuladas? En caso afirmativo, construye-las.
5. Representa las frecuencias absolutas con el gráfico adecuado.
6. ¿Qué medida de centralización usarías? Calcula su valor.

Ejercicio 2

Sobre la variable `nivel_ed`:

1. ¿De qué tipo es la variable?
2. ¿Cuántos valores distintos toma?
3. Construye las tablas de frecuencias absolutas y relativas.
4. Construye ahora las tablas de frecuencias absolutas y relativas acumuladas. ¿Tienen sentido esas tablas?

Ejercicio 3

Sobre la variable `menarqui`:

1. ¿De qué tipo es la variable?
2. ¿Cuántos valores distintos toma?
3. Calcula el recorrido.
4. Calcula la media y la desviación típica muestral.
5. Calcula las tablas de frecuencias absolutas, relativas, acumuladas y relativas acumuladas. Interpretálas (es decir, asegurate de que entiendes la información que proporcionan).
6. Representa las frecuencias absolutas con un gráfico.

Ejercicio 4

Sobre la variable `imc`:

1. ¿De qué tipo es la variable?
2. ¿Cuántos valores distintos toma?
3. Recorrido.
4. Calcula el `imc` medio y la desviación típica muestral del `imc`.
5. Calcula la tabla de frecuencias absolutas. ¿Te resulta útil?
6. Agrupa la variable en cinco clases de la misma longitud.
7. Representa esas clases mediante un histograma.
8. ¿Qué porcentaje de los datos pertenece a cada clase?
9. ¿Entre qué valores se mueve el 80% central de la muestra?
10. Calcula los cuartiles y el boxplot de esta variable e interpretalo.