

Ejercicio de evaluación

Se va a realizar el preprocesado de los datos obtenidos en un estudio donde se pretende evaluar la posible asociación entre el cáncer de mama en mujeres mayores y una serie de variables.

Para ello se plantea llevar a cabo el siguiente pipeline de operaciones en R:

1. Importar la base de datos del estudio “datos.evaluacion.xlsx”

2. Realizar las siguientes recodificaciones

2.1. Sustituir los valores missing en la variable edad por el valor de 60

2.2. Sustituir el valor de “mujer” por “Mujer” en la variable sexo

2.3 Sustituir los valores missing en la variable estado.civil por el valor de “Desconocido”

2.4 Sustituir el valor “Bajo” en la variable nivel.estudios por el valor de “B”

3. Crear las siguiente variables

3.1. Crea una variable dentro los datos del estudio llamada “edad.gr” que tenga dos posibles valores: mayores (cuando la variable edad toma valores ≥ 60 años) y menores (cuando la variable edad toma los valores < 60 años). Asignar a los valores missing en la variable edad el valor de 60

3.2. Crea una variable numerica dentro de los datos del estudio llamada “score” realizando la siguiente operación: $\text{altura}^2 / (\text{altura} / c(\text{peso}^2))$

4. Identificar cuales de los siguientes identificadores de registro están en la base de datos del estudio: 44,86,43,55,23,72,75,99

para realizar la comprobación se ha de revisar la variable ID de los datos del estudio

5. Crear una variable llamada ID_interno en la base de datos del estudio cuyos valores correspondan a una secuencia numérica continua no aleatorio desde el 1 hasta el número total de registros.

6. Ordena la base de datos del estudio de forma descendente atendiendo primero a la variable “peso” y después a la variable “estado.civil”.

7.Crear una base de datos llamada “datos_mayores” que contenga sólo aquellos registros que sean mujeres y pertenecientes al grupo de edad de los “mayores” atendiendo a la variable creada anteriormente llamada “edad.gr”

8.Exporta la base de datos llamada “datos_mujeres_mayores” en formato de R y en formato de excel con el nombre datos.mujeres.mayores.

NOTA SOBRE LA EVALUACION

Envíe un script con el código de R para realizar las distintas operaciones mencionadas al correo biodama.cne@isciii.es (fecha límite: 27 de septiembre de 2024)

Se les enviará la solución al correo desde el que manden el script