

MANEJO AVANZADO DE DATOS CON R

EJERCICIO EVALUACION

1. Importa los datos del ejercicio de evaluacion (datos.evaluacion.txt)

2. Crea las siguientes variables tal y como se indica:

2.1. estado.civil.reducido: Es igual a la variable estado.civil pero donde la categoría de “Casado” debe llamarse “cas”, la de “Soltero”, “otro”, y la de “Divorciado”, “otro”.

2.2. “talla”: variable que tenga valores de “XL” cuando el altura ≥ 150 y “L” cuando la altura sea < 150 .

2.3. grupo.edad: Una variable categórica (factor) de la edad, con los siguientes grupos; “[0,30)” “[30,50)” “[50,85)”

2.4. fecha.CM: variable tipo fecha a partir de la variable “fdiag_cm”

2.5. fecha.CP: variable tipo fecha a partir de la variable “fdiag_cp”

2.6. fecha.IN: variable tipo fecha a partir de la variable “finterv”

2.7. fecha.IN_new: variable tipo caracter a partir de la variable “fecha.IN” de tal manera que figure solo el mes y el año como en el siguiente ejemplo (May;1980)

2.8. dias.in.CM: número de días entre la fecha de diagnostico de cáncer de mama (fecha.CM) y la fecha de intervencion (fecha.IN)

2.9. semanas.df.CM: número de semanas entre la fecha de diagnostico de cáncer de mama (fecha.CM) y la fecha de intervención (fecha.IN)

2.10. fecha_analisis_IN: variable tipo fecha resultante de restar 7 días a la fecha.IN

3. Comprueba primero si existen valores repetidos en la variable ID de los datos. Después comprueba el número de caracteres de la variable ID de los datos y finalmente crea una nueva variable llamada “ID_new” que sea el resultado de la union de la variable “ID” y la variable “sexo” utilizando como separador “-”

4. Crea una nueva variable llamada “ID_num” que corresponda a los valores que tiene la variable ID_new antes del primer “-”.

5. Crea una nueva variable llamada “ID_final” que sea igual a la variable ID_new, pero donde se hayan eliminado los caracteres “--” y “_” (1 guion bajo).
6. Crea una nueva variable llamada “fiebre” que tome el valor de “Si”, cuando en la variable complicaciones se mencione algo referido a Fiebre. Y “No” en cualquiera de los otros casos.
7. Crea una funcion llamada “depuracion” que revise una base de datos (como la del curso) y que si encuentra algún registro mujer, divorciada con nivel de estudios alto, muestre el siguiente mensaje “REVISA LA BASE DE DATOS” y proporcione los identificadores de los registros que cumplan esta condición.
8. Importa las bases de datos “datos_caso_estudio_genes.txt” y los datos “datos.caso.biomarcadores.txt”
9. Une las dos bases de datos (datos y biomarcadores) en una sola base de datos llamada union.
10. Crea una base de datos llamada “datos_long_union” a partir del data.frame “union” en formato “long” con tres variables: ID (ID del registro), Biomarcador (biomarcador medido) y conc (valor de este).

NOTA SOBRE LA EVALUACIÓN

Envíe un script en formato texto plano con el código de R para realizar las distintas operaciones mencionadas al correo mariogonzalez@isciii.es.

Fecha limite de entrega: 26 de noviembre de 2021 (inclusive).

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