

MANEJO AVANZADO DE DATOS CON R

EJERCICIO EVALUACIÓN

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.red: Es igual a la variable estado.civil pero donde la categoría de “Casado” debe llamarse “casd”, la de “Soltero”, “otros”, y la de “Divorciado”, “otros”.

2.2. “talla”: variable que tenga valores de “XXL” cuando el altura ≥ 160 y “XL” cuando la altura sea < 160 .

2.3. grupo.edad: Una variable categórica (factor) de la edad, con los siguientes grupos; “[0,25)” “[25,55)” “[55,85)”

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

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

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

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

2.8. dias.int.CM: número de días entre la fecha de diagnostico de cáncer de mama (fecha.CANCER.MAMA) y la fecha de intervencion (fecha.INT)

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

2.10. fecha_analisis_INT: variable tipo fecha resultante de restar 7 días a la fecha.INT

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 las variables “sexo” y “edad” utilizando como separador “.”
4. Crea una nueva variable llamada “ID_esp” que corresponda a los valores que tiene la variable ID_new antes del guión bajo.
5. Crea una nueva variable llamada “ID_final” que sea igual a la variable ID_new, pero donde se hayan eliminado los caracteres ‘.’ y ‘-’.
6. Crea una nueva variable llamada “infarto” que tome el valor de “Si”, cuando en la variable complicaciones se mencione algo referido a Infarto. Y “No” en cualquiera de los otros casos.
7. Importa las bases de datos “datos.evaluacion.genes.txt” y los datos “datos.evaluacion.biomarcadores.txt”
8. Une las dos bases de datos (genes y biomarcadores) a “datos”, en una sola base de datos llamada union.
9. Crea una base de datos llamada “datos_long_union” a partir de la base de datos “union” en formato “long” con tres variables: ID (ID del registro), Biomarcador (biomarcador medido) y Concentracion (valor de este).