

Práctica 3. Introducción y manipulación de datos en R (2)

Jesús Martín Fernández

Ejercicios con un dataframe

Vamos a recuperar el dataframe `df_prueba` (es un archivo.csv) en tu directorio y el encabezamiento de sus variables.

```
df_prueba <- read.csv(  
  "C:~/df_prueba.csv")  
  
#Ten en cuenta que debe revisarse el "pathway" del directorio  
  
head(df_prueba)
```

Ahora etiqueta la variable `sex` (0= Mujer; 1 = varón), y `fum` e `iam` (0=No; 1=Sí), pero nombra de otra forma las variables etiquetadas

```
# Etiquetar las variables  
df_prueba$sex_label <- ifelse(df_prueba$sex == 0, "Mujer", "Varón")  
df_prueba$fum_label <- ifelse(df_prueba$fum == 0, "No", "Sí")  
df_prueba$iam_label <- ifelse(df_prueba$iam == 0, "No", "Sí")  
  
# Mostrar las primeras filas del dataframe con las nuevas etiquetas  
head(df_prueba)  
#ver la estructura del dataframe  
str(df_prueba)
```

Ves alguna diferencia entre `¿fum` y `fum label`?

Borra las variables etiquetadas y, ahora crea una variable factor y luego la etiquetas, puedes hacerlo sobre la misma variable

```

#Las variables etiquetadas en el ejemplo anterior (sex_label,...)
#son variables caracter y no se puede operar con ellas. Las borramos
# Eliminar las columnas "sex_label", "fum_label" "iam_label"
df_prueba$fum_label <- NULL
df_prueba$sex_label <- NULL
df_prueba$iam_label <- NULL

# Verificar que las columnas han sido eliminadas
head(df_prueba)
# Convertir variables a factor con etiquetas

df_prueba$sex <- factor(df_prueba$sex, levels = c(0, 1),
                        labels = c("Mujer", "Varón"))
df_prueba$fum <- factor(df_prueba$fum, levels = c(0, 1),
                        labels = c("No", "Sí"))
df_prueba$iam <- factor(df_prueba$iam, levels = c(0, 1),
                        labels = c("No", "Sí"))

# Mostrar las primeras filas del dataframe con los factores etiquetados
head(df_prueba)

```

Crea una nueva variable, IMC y recategorízala en los grupos “bajo” si $IMC < 20$, “normal” si $20 \leq IMC < 30$ y “obesidad” si $30 \leq IMC$

```

df_prueba$IMC <- df_prueba$peso / (df_prueba$alt / 100)^2
df_prueba$IMC <- round(df_prueba$IMC, 2) #redondeamos a 2 decimales
df_prueba$IMC_r <- cut(df_prueba$IMC,
                      breaks = c(-Inf, 20, 30, Inf),
                      labels = c("Bajo", "Normal", "Obesidad"))

#En esta orden los valores no son incluidos en la categoría inferior.
#Si quisiesemos que sí se incluyesen, el código debería escribirse así

#df_prueba$IMC_r <- cut(df_prueba$imc, breaks = c(-Inf, 20, 30, Inf),right=FALSE)
#labels = c("Bajo", "Normal", "Obesidad"),right = FALSE)
# Verificamos la nueva columna
table(df_prueba$IMC_r)

```

Qué información tienes de las variables continuas? ¿y de las variables factor?

```
# Resumen de las variables cuantitativas
summary(df_prueba[, c("edad", "alt", "peso", "colesterol", "IMC")])

# Resumen de las variables categóricas (factores)
summary(df_prueba[, c("sex", "fum", "iam", "IMC_r")])
```