

Ejercicio 1.2: Codificacion y Transformacion

Silvia Pineda

Datos Estrés

```
datos<-read.csv("estres.csv",header = T, sep = "\t")
```

1. Clasifica las variables en cuantitativas y cualitativas

V. Cuantitativas: id, estres,edad,peso,sueldo,carfisica,carpsiquica,ruido

V. Cualitativas: grupo, genero, contrato, jornada

2. Examina la estructura de los datos (*str*) ¿Cómo son los tipos de variables? ¿Cambiarías alguna?

```
str(datos)
```

```
'data.frame':  90 obs. of  12 variables:
 $ id      : int  75 76 77 78 79 80 81 82 83 84 ...
 $ estres  : int  8 9 2 2 2 2 2 2 8 8 ...
 $ grupo   : chr  "otros" "otros" "otros" "otros" ...
 $ edad    : int  25 27 27 29 30 30 31 31 31 26 ...
 $ peso    : int  68 65 69 60 71 61 62 69 73 58 ...
 $ genero  : chr  "hombre" "hombre" "hombre" "hombre" ...
 $ sueldo  : int  23000 25000 21000 22000 24000 27000 23000 26000 24000 25000 ...
 $ contrato : chr  "fijo" "fijo" "fijo" "fijo" ...
 $ jornada : chr  "completa" "completa" "completa" "completa" ...
 $ carfisica : int  3 5 1 1 1 1 1 1 1 1 ...
 $ carpsiquica: int  2 2 2 2 2 2 2 2 2 2 ...
 $ ruido   : int  3 3 2 4 3 0 2 4 3 0 ...
```

```
##Convertimos grupo, genero, contrato y jornada en factores
columns_to_factor<-c("grupo","genero","contrato","jornada")
datos[,columns_to_factor]<-
  lapply(datos[,columns_to_factor],as.factor)
str(datos)
```

```
'data.frame':  90 obs. of  12 variables:
 $ id      : int  75 76 77 78 79 80 81 82 83 84 ...
 $ estres  : int  8 9 2 2 2 2 2 8 8 ...
 $ grupo   : Factor w/ 4 levels "administrativo",...: 4 4 4 4 4 4 4 4 4 4 ...
 $ edad    : int  25 27 27 29 30 30 31 31 31 26 ...
 $ peso    : int  68 65 69 60 71 61 62 69 73 58 ...
 $ genero  : Factor w/ 2 levels "hombre","mujer": 1 1 1 1 1 1 1 1 1 1 ...
 $ sueldo  : int  23000 25000 21000 22000 24000 27000 23000 26000 24000 25000 ...
 $ contrato : Factor w/ 2 levels "fijo","temporal": 1 1 1 1 1 1 1 2 2 1 ...
 $ jornada : Factor w/ 2 levels "completa","parcial": 1 1 1 1 1 1 1 1 1 1 ...
 $ carfisica : int  3 5 1 1 1 1 1 1 1 1 ...
 $ carpsiquica: int  2 2 2 2 2 2 2 2 2 2 ...
 $ ruido    : int  3 3 2 4 3 0 2 4 3 0 ...
```

3. ¿Hay algún NA? Utiliza any(is.na(datos))

```
any(is.na(datos))
```

```
[1] FALSE
```

4. ¿Qué ID tiene la persona con mejor sueldo? Utiliza which()

```
datos$id[which(datos$sueldo==max(datos$sueldo))]
```

```
[1] 86 71
```

```
#En tidyverse
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.1      v stringr    1.6.0
v ggplot2    4.0.0      v tibble     3.3.0
v lubridate  1.9.4      v tidyr      1.3.1
v purrr      1.2.0
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
datos |>
  filter(sueldo == max(sueldo)) |>
  pull(id)
```

```
[1] 86 71
```

5. ¿Cuántos trabajadores tienen un nivel de estrés mayor de 5? Utiliza length()

```
length(which(datos$estres>5))
```

```
[1] 35
```

```
#En tidyverse
datos |>
  filter(estres > 5)|>
  nrow()
```

```
[1] 35
```

6. ¿Cuántos trabajadores tienen estrés mayor de 5 y además una carga física > 3?

```
length(which(datos$estres>5 & datos$carfisica>3))
```

```
[1] 10
```

```
datos |>
  filter(estres > 5 & carfisica > 3)|>
  nrow()
```

```
[1] 10
```

7. Nos hemos equivocado en la introducción de datos y para todos los valores 2 de carga física, queríamos haber puesto un 3. Recodifica esta variable incluyendo este cambio. Utiliza if_else() de la librería dplyr

```
library(dplyr)
datos$carfisica_corregida<-if_else(datos$carfisica==2,3,datos$carfisica)

table(datos$carfisica,datos$carfisica_corregida)
```

	1	3	4	5
1	14	0	0	0
2	0	37	0	0
3	0	25	0	0
4	0	0	13	0
5	0	0	0	1

```
#En tidyverse
datos <- datos |>
  mutate(carfisica_corregida = if_else(carfisica == 2, 3, carfisica))

datos |>
  count(carfisica, carfisica_corregida)
```

	carfisica	carfisica_corregida	n
1	1		14
2	2		37
3	3		25
4	4		13
5	5		1

8. Crea una nueva variable llamada *expo_total* que se obtenga como suma de los valores de exposición en *carga física*, *carga psíquica* y *ruido*.

```
datos$expo_total<-datos$carfisica+datos$carpsiquica+datos$ruido
head(datos)
```

	id	estres	grupo	edad	peso	genero	sueldo	contrato	jornada	carfisica
1	75	8	otros	25	68	hombre	23000	fijo	completa	3
2	76	9	otros	27	65	hombre	25000	fijo	completa	5
3	77	2	otros	27	69	hombre	21000	fijo	completa	1
4	78	2	otros	29	60	hombre	22000	fijo	completa	1
5	79	2	otros	30	71	hombre	24000	fijo	completa	1
6	80	2	otros	30	61	hombre	27000	fijo	completa	1

	carpsiquica	ruido	carfisica_corregida	expo_total
1	2	3	3	8
2	2	3	5	10
3	2	2	1	5
4	2	4	1	7
5	2	3	1	6
6	2	0	1	3

```
#En tidyverse
datos <-datos |>
  mutate(expo_total = carfisica+carpsiquica+ruido)
```

9. *Categoriza* la variable *estres* en otra nueva variable que llamaremos *estres_grupo* que incluya tres categorías: *bajo*: 1 a 2; *medio*: 3 a 7; *alto*: 8 a 10. ¿Cuántas personas hay en cada categoría de la nueva variable. Utiliza *cut()* y *breaks()*

```
datos$estres_grupo<-cut(datos$estres,breaks=c(-Inf,2,7,Inf),
                        labels = c("bajo", "medio", "alto"))
datos[,c("estres","estres_grupo")]
```

	estres	estres_grupo
1	8	alto
2	9	alto
3	2	bajo
4	2	bajo

5	2	bajo
6	2	bajo
7	2	bajo
8	2	bajo
9	8	alto
10	8	alto
11	8	alto
12	7	medio
13	2	bajo
14	2	bajo
15	2	bajo
16	2	bajo
17	3	medio
18	3	medio
19	3	medio
20	3	medio
21	3	medio
22	3	medio
23	3	medio
24	3	medio
25	3	medio
26	3	medio
27	3	medio
28	3	medio
29	3	medio
30	3	medio
31	3	medio
32	3	medio
33	3	medio
34	3	medio
35	3	medio
36	2	bajo
37	2	bajo
38	2	bajo
39	2	bajo
40	2	bajo
41	2	bajo
42	2	bajo
43	2	bajo
44	2	bajo
45	2	bajo
46	2	bajo
47	2	bajo

48	2	bajo
49	2	bajo
50	3	medio
51	2	bajo
52	9	alto
53	9	alto
54	9	alto
55	9	alto
56	9	alto
57	9	alto
58	9	alto
59	2	bajo
60	2	bajo
61	9	alto
62	7	medio
63	7	medio
64	7	medio
65	7	medio
66	7	medio
67	7	medio
68	7	medio
69	7	medio
70	7	medio
71	7	medio
72	5	medio
73	4	medio
74	3	medio
75	3	medio
76	2	bajo
77	1	bajo
78	10	alto
79	10	alto
80	10	alto
81	2	bajo
82	2	bajo
83	9	alto
84	9	alto
85	9	alto
86	9	alto
87	9	alto
88	9	alto
89	9	alto
90	9	alto

10. Guarda la base de datos corregida

```
write.csv(datos,"estres_corregido.csv")
```