

Estadística descriptiva con datos cualitativos

Curso de Estadística Descriptiva

24/12/2018

DATASET: HairEyeColor

El dataset **HairEyeColor** recoge tres variables cualitativas y R trae definido de serie. Se recogen las variables de género (hombre/mujer), color de ojos (4 colores diferentes) y color de pelo (4 colores diferentes).

Este dataset no es un *DataFrame* ya que viene definido como frecuencias absolutas directamente. Por tanto, los datos aparecen agregados siendo una forma común en el mundo real.

HairEyeColor

```
## , , Sex = Male
##
##      Eye
## Hair   Brown Blue Hazel Green
## Black   32   11   10    3
## Brown   53   50   25   15
## Red     10   10    7    7
## Blond    3   30    5    8
##
## , , Sex = Female
##
##      Eye
## Hair   Brown Blue Hazel Green
## Black   36    9    5    2
## Brown   66   34   29   14
## Red     16    7    7    7
## Blond    4   64    5    8
```

```
#Para obtener el número de individuos totales se usa:
sum(HairEyeColor) -> total
```

El total de individuos de la tabla de datos es 592.

Obtener tabla de frecuencias marginales

Se emplea la función `prop.table()` sobre el dataset, de forma directa.

```
#Frec. Rel. Marg. por género (hombre/mujer)
prop.table(HairEyeColor, margin = 3)
```

```
## , , Sex = Male
##
##      Eye
## Hair      Brown      Blue      Hazel      Green
##   Black 0.114695341 0.039426523 0.035842294 0.010752688
##   Brown 0.189964158 0.179211470 0.089605735 0.053763441
##   Red   0.035842294 0.035842294 0.025089606 0.025089606
##   Blond 0.010752688 0.107526882 0.017921147 0.028673835
##
## , , Sex = Female
##
##      Eye
## Hair      Brown      Blue      Hazel      Green
##   Black 0.115015974 0.028753994 0.015974441 0.006389776
##   Brown 0.210862620 0.108626198 0.092651757 0.044728435
##   Red   0.051118211 0.022364217 0.022364217 0.022364217
##   Blond 0.012779553 0.204472843 0.015974441 0.025559105
```

```
#Frec. Rel. Marg. por color de pelo y el color de ojos (combinación de los colores)
prop.table(HairEyeColor, margin = c(1,2))
```

```
## , , Sex = Male
##
##      Eye
## Hair      Brown      Blue      Hazel      Green
##   Black 0.4705882 0.5500000 0.6666667 0.6000000
##   Brown 0.4453782 0.5952381 0.4629630 0.5172414
##   Red   0.3846154 0.5882353 0.5000000 0.5000000
##   Blond 0.4285714 0.3191489 0.5000000 0.5000000
##
## , , Sex = Female
##
##      Eye
## Hair      Brown      Blue      Hazel      Green
##   Black 0.5294118 0.4500000 0.3333333 0.4000000
##   Brown 0.5546218 0.4047619 0.5370370 0.4827586
##   Red   0.6153846 0.4117647 0.5000000 0.5000000
##   Blond 0.5714286 0.6808511 0.5000000 0.5000000
```

Cambiar orden de las variables -> aperm()

Para cambiar el orden de variables dentro de un dataset ya creado se utiliza la función **aperm(dataset, perm = c(variable1, variable2, variable3))** (array permutation). Permite cambiar el orden de las columnas, que ahora están definidas como:

- *Primera variable = color de pelo.* Debido a que define las filas (rows). -*Segunda variable = color de ojos.* Debido a que define las columnas (columns). -*Tercera variable = genero.* Debido a que da el nombre de las tablas bidimensionales.

El nuevo orden establecido será primero el genero, segundo el color del pelo y tercero el color de los ojos.

```
aperm(HairEyeColor, perm = c("Sex", "Hair", "Eye"))
```

```
## , , Eye = Brown
##
##      Hair
## Sex    Black Brown Red  Blond
## Male      32    53  10    3
## Female     36    66  16    4
##
## , , Eye = Blue
##
##      Hair
## Sex    Black Brown Red  Blond
## Male      11    50  10   30
## Female      9    34   7   64
##
## , , Eye = Hazel
##
##      Hair
## Sex    Black Brown Red  Blond
## Male      10    25   7    5
## Female      5    29   7    5
##
## , , Eye = Green
##
##      Hair
## Sex    Black Brown Red  Blond
## Male      3    15   7    8
## Female      2    14   7    8
```

Formatear la salida de las tablas -> Librería kableExtra y Librería xtable

La librería **kableExtra** permite dar un mejor formato a las tablas. Kable forma parte de *Knit*, y permite obtener una representación global de los datos que guarda R. S eusa la función **kable()**.

No es adecuado para analizar tres variables cualitativas.

```
#library(kableExtra)
#kable(HairEyeColor)
```

Otra librería utilizada para formatear tablas es **xtable**. Se utiliza la función **xtable()** para el formateo y se emplea como máximo para *tablas bidimensionales*.

Para que te aparezca no solo el código matemático LaTeX se debe poner en el chunk **results =“asis”**. En html no se genera la tabla con xtable().

```
#library(xtable)
sex = factor(c("H", "M", "M", "M", "H", "H", "M", "M"))
ans = factor(c("S", "N", "S", "S", "S", "N", "N", "S"))

#xtable(table(sex, ans))
```