

# Estadística Descriptiva con Datos Cualitativos

Ramon Ceballos

26/1/2021

## TABLAS DE FRECUENCIAS PARTIENDO DE DATA FRAMES YA GENERADOS

Como ya hemos comentado en varias ocasiones, la manera natural de organizar datos multidimensionales en R es en forma de data frame.

En esta sección explicaremos algunas instrucciones para calcular tablas de frecuencias absolutas a partir de un data frame de variables cualitativas.

Para ilustrarla, usaremos el fichero que se encuentra en el la carpeta de datos: "data/EnergyDrink"

Este fichero consiste en una tabla de datos con la siguiente información sobre 122 estudiantes de una Universidad de España: su sexo (variable `sexo`), el estudio en el que están matriculados (variable `estudio`) y si consumen habitualmente bebidas energéticas para estudiar (variable `bebe`).

```
#Creo la variable BebEnerg a partir de los datos de EnergyDrink
Beb_Energ = read.table("../../data/EnergyDrink", header = TRUE)

#Consulto la estructura del DataFrame
str(Beb_Energ)
```

```
## 'data.frame': 122 obs. of 3 variables:
## $ estudio: Factor w/ 4 levels "Industriales",...: 2 3 1 2 1 3 2 1 2 2 ...
## $ bebe : Factor w/ 2 levels "No","Si": 1 1 2 2 1 1 2 1 1 1 ...
## $ sexo : Factor w/ 2 levels "Hombre","Mujer": 2 1 2 1 2 2 1 1 1 1 ...
```

```
#Consulto las 3 primeras filas
head(Beb_Energ,3)
```

```
##      estudio bebe  sexo
## 1 Informatica No  Mujer
## 2      Mates  No  Hombre
## 3 Industriales Si   Mujer
```

Aplicando la función `summary()` a un data frame de variables cualitativas, obtenemos, a modo de resumen, una tabla con las frecuencias absolutas de cada variable. No es más que un resumen de las variables.

Esta tabla sólo sirve para ver la información, porque sus entradas son palabras. Solo valdría para variables de tipo factor, ya que para variables numéricas la función `summary()` da una serie de medidas del mundo de la estadística.

**NO GENERA LA TABLA DE FRECUENCIAS ABSOLUTAS, ES SOLO PARA VER**

```
#Tablas de frecuencias absolutas para cada variable
summary(Beb_Energ)
```

```
##          estudio  bebe          sexo
## Industriales:37 No:97  Hombre:83
## Informatica :53 Si:25  Mujer :39
## Mates       :16
## Telematica  :16
```

Para calcular en un solo paso la table (**TABLA DE FRECUENCIAS ABSOLUTAS**) de cada variable, podemos usar la función **apply()** de la manera siguiente:

```
#MARGIN=2 para que se aplique table() por columnas
#Se genera la frec. abs. para cada una de las variables (factores)
apply(Beb_Energ, MARGIN=2, FUN=table)
```

```
## $estudio
##
## Industriales  Informatica      Mates  Telematica
##           37           53           16           16
##
## $bebe
##
## No Si
## 97 25
##
## $sexo
##
## Hombre  Mujer
##      83    39
```

De esta manera, obtenemos una **list** cuyas componentes son las tablas que queríamos.

```
#se selecciona la variables una vez obtenida la tabla de frec. abs. por variables
apply(Beb_Energ,MARGIN=2,FUN=table)$sexo
```

```
##
## Hombre  Mujer
##      83    39
```

```
#Se obtiene lo mismo pero primero se selecciona la columna (variable) y luego se hace table()
table(Beb_Energ$sexo)
```

```
##
## Hombre  Mujer
##      83    39
```

Si aplicamos la función **table()** a un data frame de variables cualitativas, obtenemos su tabla de frecuencias absolutas, con las variables ordenadas tal y como aparecen en el data frame.

```
#Como lo hemos hecho con anterioridad
table(Beb_Energ)
```

```
## , , sexo = Hombre
##
##          bebe
## estudio    No Si
## Industriales 19 6
## Informatica 30 7
## Mates        8 1
## Telematica  10 2
##
## , , sexo = Mujer
##
##          bebe
## estudio    No Si
## Industriales 10 2
## Informatica 11 5
## Mates        6 1
## Telematica   3 1
```

O también podemos hacer...

```
#table() respecto al estudios y el sexo, tabla bidimensional
table(Beb_Energ[c(1,3)])
```

```
##          sexo
## estudio  Hombre Mujer
## Industriales    25   12
## Informatica     37   16
## Mates           9    7
## Telematica     12    4
```

Una tercera opción es usar la función **ftable()**, que produce la misma tabla de frecuencias pero en formato plano.

```
ftable(Beb_Energ)
```

```
##          sexo Hombre Mujer
## estudio  bebe
## Industriales No      19    10
##              Si       6     2
## Informatica No      30    11
##              Si       7     5
## Mates       No      8     6
##              Si       1     1
## Telematica  No     10     3
##              Si       2     1
```