

Estadística descriptiva con datos cuantitativos

Ramon Ceballos

29/1/2021

1. Frecuencias para Datos Cuantitativos

El tratamiento de las frecuencias de datos cuantitativos es similar al de los datos ordinales. La cosa cambia ligeramente debido a que **no se tienen en cuenta todos los niveles posibles, sino únicamente los observados**.

Ejemplo 1

Se han pedido las edades a 20 clientes de un museo. Las respuestas obtenidas han sido las siguientes:

```
#vector de edades  
edad = c(15,18,25,40,30,29,56,40,13,27,42,23,11,26,25,32,30,40,33,29)
```

Recordemos que solamente nos interesan las frecuencias de las edades observadas. Se calcula la tabla de frecuencias absolutas con la función **table()**.

```
#Tabla de contingencia para el vector edad (FREC. ABS.)  
table(edad)
```

```
## edad  
## 11 13 15 18 23 25 26 27 29 30 32 33 40 42 56  
## 1 1 1 1 1 2 1 1 2 2 1 1 3 1 1
```

Calculemos el resto de frecuencias como ya sabemos.

```
#Tabla de frecuencias relativas globales  
round(prop.table(table(edad)),3)
```

```
## edad  
## 11 13 15 18 23 25 26 27 29 30 32 33 40 42 56  
## 0.05 0.05 0.05 0.05 0.05 0.10 0.05 0.05 0.10 0.10 0.05 0.05 0.15 0.05 0.05
```

```
#Tabla de frecuencias absolutas acumuladas  
cumsum(table(edad))
```

```
## 11 13 15 18 23 25 26 27 29 30 32 33 40 42 56  
## 1 2 3 4 5 7 8 9 11 13 14 15 18 19 20
```

```
#Tabla de frecuencias relativas acumuladas
round(cumsum(prop.table(table(edad))),3)
```

```
##    11    13    15    18    23    25    26    27    29    30    32    33    40    42    56
## 0.05 0.10 0.15 0.20 0.25 0.35 0.40 0.45 0.55 0.65 0.70 0.75 0.90 0.95 1.00
```

2. Definición matemática respecto a las frecuencias de datos cuantitativos

En general, supongamos que tenemos n observaciones de una propiedad que se mide con un número real y obtenemos la variable cuantitativa formada por los datos:

$$x_1, \dots, x_n$$

Sean ahora X_1, \dots, X_k los valores distintos que aparecen en esta lista de datos y considerémoslos ordenados:

$$X_1 < X_2 < \dots < X_k$$

X_1 sería el mínimo de la observación y X_k sería el máximo de la observación, siendo $k \leq n$.

Entonces, en una variable cuantitativa podemos definir lo siguiente:

- La *frecuencia absoluta* de X_i es el número n_i de elementos que son iguales a X_i .
- La *frecuencia relativa* de X_i es $f_i = \frac{n_i}{n}$.
- La *frecuencia absoluta acumulada* de X_i es $N_i = \sum_{j=1}^i n_j$.
- La *frecuencia relativa acumulada* de X_i es $F_i = \frac{N_i}{n}$.

Ejemplo 2

DADOS

Lanzamos 25 veces un dado de 6 caras y anotamos las puntuaciones obtenidas en cada tirada.

En este caso, $n = 25$ y, los distintos valores observados son:

$$X_1 = 1, X_2 = 2, X_3 = 3, X_4 = 4, X_5 = 5, X_6 = 6$$

Nos interesa ahora calcular las frecuencias de este experimento. Además, las organizaremos en un data frame para observarlas de forma más clara y sencilla en una tabla.

```
#Fijo una semilla
set.seed(162017)

#Vector determinado para el ejemplo
datos = sample(1:6,25,replace = TRUE)
datos
```

```
## [1] 1 1 5 5 5 5 1 6 5 4 1 3 1 3 2 2 1 1 1 4 2 1 6 3 1
```

```
#Anulo la semilla fijada
set.seed(NULL)
```

```
#defino la tabla de contingencia
table(dados)
```

```
## dados
##  1  2  3  4  5  6
## 10  3  3  2  5  2
```

```
#defino la tabla de frecuencias relativas
round(prop.table(table(dados)),2)
```

```
## dados
##    1    2    3    4    5    6
## 0.40 0.12 0.12 0.08 0.20 0.08
```

```
#defino la tabla de frecuencias absolutas acumuladas
cumsum(table(dados))
```

```
##  1  2  3  4  5  6
## 10 13 16 18 23 25
```

```
#defino la tabla de frecuencias relativas acumuladas
round(cumsum(prop.table(table(dados))),2)
```

```
##    1    2    3    4    5    6
## 0.40 0.52 0.64 0.72 0.92 1.00
```

```
#Para facilitar la visualización de las frecuencias,
#se crea un DF con la visualización de todas las frecuencias descritas a la vez
```

```
#Importante poner as.vector() para que las tablas se traten como vectores al crear el DF
#Así no se añaden las etiquetas
dados.df = data.frame(Puntuacion = 1:6,
                      Fr.abs = as.vector(table(dados)),
                      Fr.rel = as.vector(round(prop.table(table(dados)),2)),
                      Fr.acu = as.vector(cumsum(table(dados))),
                      Fr.racu = as.vector(round(cumsum(prop.table(table(dados))),2)))
```

```
dados.df
```

```
##  Puntuacion Fr.abs Fr.rel Fr.acu Fr.racu
## 1          1     10  0.40     10   0.40
## 2          2      3  0.12     13   0.52
## 3          3      3  0.12     16   0.64
## 4          4      2  0.08     18   0.72
## 5          5      5  0.20     23   0.92
## 6          6      2  0.08     25   1.00
```

¡OJO! Para entrar una tabla unidimensional como una variable en un data frame, es conveniente transformarla en vector con `as.vector`. Si no, cada `table` y cada `prop.table` añadirían una columna extra con los nombres de los niveles.