

Frecuencias de Datos Cuantitativos

Santiago Pérez Moncada

20/6/2020

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

```
edades = sample(1:50, size = 20, replace = TRUE)
edades
```

```
## [1] 15 44 11 8 13 34 7 21 19 21 7 37 8 50 49 27 7 1 32 16
```

Recordemos que solamente nos interesan las frecuencias de las edades observadas. Es decir, solamente nos interesan

```
table(edades)
```

```
## edades
## 1 7 8 11 13 15 16 19 21 27 32 34 37 44 49 50
## 1 3 2 1 1 1 1 1 2 1 1 1 1 1 1 1
```

Calculemos el resto de frecuencias como ya sabemos

```
round(prop.table(table(edades)),3)#Frec.Relativa
```

```
## edades
## 1 7 8 11 13 15 16 19 21 27 32 34 37 44 49 50
## 0.05 0.15 0.10 0.05 0.05 0.05 0.05 0.05 0.10 0.05 0.05 0.05 0.05 0.05 0.05 0.05
```

```
cumsum(table(edades))#Frec.Acumulada.Absoluta
```

```
## 1 7 8 11 13 15 16 19 21 27 32 34 37 44 49 50
## 1 4 6 7 8 9 10 11 13 14 15 16 17 18 19 20
```

```
round(cumsum(prop.table(table(edades))),3) #Frec.Acumulada.Relativa
```

```
## 1 7 8 11 13 15 16 19 21 27 32 34 37 44 49 50
## 0.05 0.20 0.30 0.35 0.40 0.45 0.50 0.55 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00
```

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.

$$X_1 < X_2 < \dots < X_K$$

Entonces, en esta variable cuantitativa

- La frecuencia absoluta 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

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.

```
datos = sample(1:6,size = 25,replace = TRUE)
datos

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

```
table(datos) #Frecuencia absoluta
```

```
## datos
## 1 2 3 4 5 6
## 3 4 5 5 5 3
```

```
round(prop.table(table(datos)),2) #Frecuencia Relativa global
```

```
## datos
## 1 2 3 4 5 6
## 0.12 0.16 0.20 0.20 0.20 0.12
```

```
cumsum(table(datos)) #Frecuencia absoluta acumulada
```

```
## 1 2 3 4 5 6
## 3 7 12 17 22 25
```

```
round(cumsum(prop.table(table(dados))),2) #Frecuencia relativa acumulada
```

```
##      1      2      3      4      5      6
## 0.12 0.28 0.48 0.68 0.88 1.00
```

```
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.rel.acu = as.vector(round(cumsum(prop.table(table(dados))),2)))
```

```
dados.df
```

```
##      Puntuacion Fr.abs Fr.rel Fr.acu Fr.rel.acu
## 1             1      3  0.12      3      0.12
## 2             2      4  0.16      7      0.28
## 3             3      5  0.20     12      0.48
## 4             4      5  0.20     17      0.68
## 5             5      5  0.20     22      0.88
## 6             6      3  0.12     25      1.00
```

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