

# Datos cuantitativos agrupados

Ramon Ceballos

31/1/2021

## EJERCICIO DE AGRUPACIÓN DE DATOS CUANTITATIVO EN R

Lee atentamente las transparencias del tema 9 e implementa las reglas que se comentan en los ejercicios.

### Preguntas de esta tarea

```
crabs = read.table("../../data/datacrab.txt", header=TRUE)

str(crabs)
```

```
## 'data.frame': 173 obs. of 6 variables:
## $ input : int 1 2 3 4 5 6 7 8 9 10 ...
## $ color : int 3 4 2 4 4 3 2 4 3 4 ...
## $ spine : int 3 3 1 3 3 3 1 2 1 3 ...
## $ width : num 28.3 22.5 26 24.8 26 23.8 26.5 24.7 23.7 25.6 ...
## $ satell: int 8 0 9 0 4 0 0 0 0 0 ...
## $ weight: int 3050 1550 2300 2100 2600 2100 2350 1900 1950 2150 ...
```

```
head(crabs)
```

```
## input color spine width satell weight
## 1 1 3 3 28.3 8 3050
## 2 2 4 3 22.5 0 1550
## 3 3 2 1 26.0 9 2300
## 4 4 4 3 24.8 0 2100
## 5 5 4 3 26.0 4 2600
## 6 6 3 3 23.8 0 2100
```

```
cw =crabs$width
```

1. Da el algoritmo para reproducir el proceso de generación de clases y sus marcas respectivo para el número de clases obtenido con la regla de la Scott en Python.

(VEASE EL FICHERO DE PYTHON)

## 2. Da el algoritmo para reproducir el proceso de generación de clases y sus marcas respectivo para el número de clases obtenido con la regla de la raíz en R o en Python.

Una vez seleccionada la variable de estudio, se establece el número de clases según la regla de la raíz cuadrada.

```
#Guardo la longitud de cw en n
n = length(cw)

#Determino el número de clases
k2 = ceiling(sqrt(n))
k2
```

```
## [1] 14
```

El número de clases es 14 según la regla empleada.

A continuación, se calcula la amplitud de la agrupación.

```
#Amplitud de la agrupación
A2 = diff(range(cw))/k2
A2
```

```
## [1] 0.8928571
```

Se fija que  $A = 0.9$ , ya que la precisión de los datos estudiados es el valor decimal.

```
A2=round(A2,1)
```

A continuación se van a determinar los extremos de la agrupación. En primer lugar, se calcula el primer extremo del intervalo.

```
#Primer extremo
L1 = min(cw)-(1/2)*0.1
```

Para determinar el resto de extremos se hace lo siguiente.

```
#Vector de los extremos del intervalo
L_sqrt = L1 + A2*(0:14)
L_sqrt
```

```
## [1] 20.95 21.85 22.75 23.65 24.55 25.45 26.35 27.25 28.15 29.05 29.95 30.85
## [13] 31.75 32.65 33.55
```

Para calcular las marcas de clase empleamos la siguiente expresión.

```
#Vector de las marcas de clase
X_sqrt = (L_sqrt[1:14]+L_sqrt[2:15])/2
X_sqrt
```

```
## [1] 21.4 22.3 23.2 24.1 25.0 25.9 26.8 27.7 28.6 29.5 30.4 31.3 32.2 33.1
```

### 3. Da el algoritmo para reproducir el proceso de generación de clases y sus marcas respectivo para el número de clases obtenido con la regla de la Sturges en R o en Python.

Una vez seleccionada la variable de estudio, se establece el número de clases según la regla de Sturges.

```
#Guardo la longitud de cw en n
n = length(cw)

#Determino el número de clases
k3 = ceiling(1+log(n,2))
k3
```

```
## [1] 9
```

El número de clases es 9 según la regla empleada.

A continuación, se calcula la amplitud de la agrupación.

```
#Amplitud de la agrupación
A3 = diff(range(cw))/k3
A3
```

```
## [1] 1.388889
```

Se fija que  $A = 1.4$ , ya que la precisión de los datos estudiados es el valor decimal.

```
A3=round(A3,1)
```

A continuación se van a determinar los extremos de la agrupación. En primer lugar, se calcula el primer extremo del intervalo.

```
#Primer extremo
L1 = min(cw)-(1/2)*0.1
```

Para determinar el resto de extremos se hace lo siguiente.

```
#Vector de los extremos del intervalo
L_Sturges = L1 + A3*(0:9)
L_Sturges
```

```
## [1] 20.95 22.35 23.75 25.15 26.55 27.95 29.35 30.75 32.15 33.55
```

Para calcular las marcas de clase empleamos la siguiente expresión.

```
#Vector de las marcas de clase
X_Sturges = (L_Sturges[1:length(L_Sturges)-1]+L_Sturges[2:length(L_Sturges)])/2
X_Sturges
```

```
## [1] 21.65 23.05 24.45 25.85 27.25 28.65 30.05 31.45 32.85
```

#### 4. Da el algoritmo para reproducir el proceso de generación de clases y sus marcas respectivo para el número de clases obtenido con la regla de la Freedman-Diaconis en R o en Python.

Una vez seleccionada la variable de estudio, se establece el número de clases según la regla de Freedman-Diaconis.

```
#Guardo la longitud de cw en n
n = length(cw)

#Amplitud teorica
A_fd = 2* (quantile(cw,0.75, names = FALSE) - quantile(cw,0.25, names = FALSE)) * n^(-1/3)

#Determino el número de clases
k4 = ceiling(diff(range(cw))/A_fd)
k4

## [1] 13
```

El número de clases es 13 según la regla empleada.

A continuación, se calcula la amplitud de la agrupación.

```
#Amplitud de la agrupación
A4 = diff(range(cw))/k4
A4
```

```
## [1] 0.9615385
```

Se fija que  $A = 1.1$ , ya que la precisión de los datos estudiados es el valor decimal.

```
A4=1.1
```

A continuación se van a determinar los extremos de la agrupación. En primer lugar, se calcula el primer extremo del intervalo.

```
#Primer extremo
L1 = min(cw)-(1/2)*0.1
```

Para determinar el resto de extremos se hace lo siguiente.

```
#Vector de los extremos del intervalo
L_fd = L1 + A4*(0:13)
L_fd
```

```
## [1] 20.95 22.05 23.15 24.25 25.35 26.45 27.55 28.65 29.75 30.85 31.95 33.05
## [13] 34.15 35.25
```

Para calcular las marcas de clase empleamos la siguiente expresión.

```
#Vector de las marcas de clase  
X_fd = (L_fd[1:length(L_fd)-1]+L_fd[2:length(L_fd)])/2  
X_fd
```

```
## [1] 21.5 22.6 23.7 24.8 25.9 27.0 28.1 29.2 30.3 31.4 32.5 33.6 34.7
```