Estadística descriptiva con datos cuantitativos

Ramon Ceballos

30/1/2021

Medidas de centralización o de tendencia central

Las **medidas de tendencia central** son las que dan *un valor* representativo a todas las observaciones. Algunas de las más importantes son:

- La **media aritmética** o **valor medio**. Hay tres modos de obtener dicho parámetrpo estadístico.
  - Sumar todos las `n` observaciones y dividirlas entre el nº de observaciones de `n` totales.
  - Multiplicar cada nivel de la variable por su frecuencia absoluta y dividirla entre el nº de observaciones de `n` totales.
  - Multiplicar cada nivel de la variable por su frecuencia relativa absoluta.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{j=1}^{k} n_j X_j}{n} = \sum_{j=1}^{k} f_j X_j$$

- La **mediana**, que representa el valor central en la lista ordenada de observaciones.
- La **moda** es el valor (o valores) de máxima frecuencia (absoluta o relativa, el resultado será el mismo).

En el mundo de la estadística existen diversos tipos de medias entre las que podemos citar la media aritmética, la media geométrica y la media armónica.

Ejemplo 1

Recordemos el ejemplo de las edades.

```r
#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)

sort(edad) #Ordenamos los datos por su orden natural
```

```
##  [1] 11 13 15 18 23 25 25 26 27 29 29 30 30 32 33 40 40 40 42 56
```

```r
#Tabla de frecuencias absolutas
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
```

En este caso, la moda es 40, la mediana es $\frac{29+29}{2} = 29$ y la media aritmética es: $\frac{11+13+15+18+23+25+25+26+27+29+29+30+30+32+33+40+40+40+42+56}{20} = 29.2$

Ejemplo 2

Recordemos el ejemplo de los dados.

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

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

```
##  [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
```

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

```r
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
```

En este caso, la *moda* son dos valores: el 2 y el 3. La *mediana* es $x_{(13)} = 3$ y la *media aritmética* es 2.8.

Ejemplos 1 y 2 con R

Vamos a calcular la media aritmética, mediana y moda de los dos ejemplos anteriores con instrucciones de R.

```r
mean(edad) #La media aritmética
```

```
## [1] 29.2
```

```r
mean(dados) #La media aritmética
```

```
## [1] 2.8
```

```r
median(edad) #La mediana
```

```
## [1] 29
```

```
median(dados) #La mediana
```

```
## [1] 2
```

```
#La moda
as.numeric(names(which(table(edad)==max(table(edad))))) #La moda
```

```
## [1] 40
```

```
#La moda
as.numeric(names(which(table(dados)==max(table(dados)))))
```

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

Cuando trabajamos con datos cuantitativos, es conveniente que el resultado lo demos como un número. De ahí que hayamos aplicado la función `as.numeric`. La instrucción completa es: `as.numeric(names(which(table("vector")==max`

Otra forma de calcular la **moda** es cargar la **librería modeest** y emplea la función **mfv()**.

```
vector2 <- c(2,3,5,7,9,7)
library("modeest") # carga la biblioteca modeest
mfv(vector2) # calcula la moda
```

```
## [1] 7
```

1. La mediana

La definición formal de la *mediana* es la siguiente. Denotando por:

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$$

los datos de la variable cuantitativa ordenados de menor a mayor, la mediana es:

- Si n par, es la media de los dos datos centrales:

$$\frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$$

- Si n impar, es el dato central:

$$x_{(\frac{n+1}{2})}$$

2. Las diferentes medias en matemáticas

En matemáticas y estadística, una media o promedio es una medida de tendencia central. Resulta al efectuar una serie determinada de operaciones con un conjunto de números y que, en determinadas condiciones, puede representar por sí solo a todo el conjunto. Existen distintos tipos de medias, tales como la media geométrica, la media ponderada y la media armónica aunque en el lenguaje común, tanto en estadística como en matemáticas la elemental de todas ellas es el término que se refiere generalmente a la media aritmética.

Existen numerosos ejemplos de medias $\bar{x}=m_i(x_1,\ldots,x_n)$, una de las pocas propiedades compartidas por todas las medias es que cualquier media está comprendida entre el valor máximo y el valor mínimo del conjunto de variables:

$$\min\{x_1, x_2, \ldots x_n\} \leq \bar{x} \leq \max\{x_1, x_2, \ldots x_n\}$$

Además debe cumplirse que:

$$\bar{x} = x_1, \quad \text{si } x_1 = x_2 = \cdots = x_n$$

2.1 Media Aritmética (mean)

La media aritmética es un promedio estándar que a menudo se denomina promedio.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

La media se confunde a veces con la mediana o moda. La media aritmética es el promedio de un conjunto de valores, o su distribución; sin embargo, para las distribuciones con sesgo, la media no es necesariamente el mismo valor que la mediana o que la moda exponencial y de Poisson.

Por ejemplo, la media aritmética de 34, 27, 45, 55, 22, 34 (seis valores) es:

$$\frac{34+27+45+55+22+34}{6} = \frac{217}{6} \approx 36,167$$

2.2 Media Aritmética ponderada

A veces puede ser útil otorgar pesos o valores a los datos dependiendo de su relevancia para determinado estudio. En esos casos se puede utilizar una media ponderada. Si $X_1, X_2, \ldots, X_n$ es un conjunto de datos o muestra y $w_1, w_2, \ldots, w_n$ son números reales positivos, llamados "pesos" o factores de ponderación, se define la media ponderada es decir que es relativa a esos pesos como:

$$\bar{X}_w = \frac{X_1 \cdot w_1 + X_2 \cdot w_2 + \ldots + X_n \cdot w_n}{w_1 + w_2 + \ldots + w_n} = \frac{\sum_{i=1}^{n} X_i \cdot w_i}{\sum_{i=1}^{n} w_i}$$

La media es invariante frente a transformaciones lineales, cambio de origen y escala, de las variables, es decir si X es una variable aleatoria e Y es otra variable aleatoria que depende linealmente de X, es decir, $\mathbf{Y} = \mathbf{a \cdot XL + b}$ (donde a representa la magnitud del cambio de escala y b la del cambio de origen) se tiene que:

$$\bar{Y} = a\bar{X} + b$$

2.3 Media Geométrica

La *media geométrica* es un promedio muy útil en conjuntos de números que son interpretados en orden de su producto, no de su suma (tal y como ocurre con la media aritmética). Por ejemplo, las velocidades de crecimiento.

Para obtenerlo se multiplican todas las observaciones de una variable cuantitativa y, tras ello, se realiza la raiz enesima, siendo n la longitud del vector que representa la variable.

$$\bar{x} = \left(\prod_{i=1}^{n} x_i \right)^{1/n}$$

Por ejemplo, la media geométrica de la serie de números 1,2,3,4,5,9 (seis valores) es $(1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 9)^{1/6} = 1080^{1/6} \approx 3,203$.

2.4 Media Armónica

La m*edia armónica* es un promedio muy útil en conjuntos de números que se definen en relación con alguna unidad, por ejemplo la velocidad (distancia por unidad de tiempo).

$$\bar{x} = n \cdot \left(\sum_{i=1}^{n} \frac{1}{x_i} \right)^{-1}$$

Por ejemplo, la media armónica de los números: 34, 27, 45, 55, 22, y 34 es:

$$\frac{6}{\frac{1}{34} + \frac{1}{27} + \frac{1}{45} + \frac{1}{55} + \frac{1}{22} + \frac{1}{34}} \approx 33,018$$

Ejemplo de las medias en R

Vamos a calcular las medias del vector `x` que está formado por `n` observaciones.

```
x = c (32, 45, 67, 43, 28, 17, 48, 95)
n = length(x)
```

Para la **media aritmética** se emplea lo siguiente.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

```
#Definición de la media aritmética
sum(x)/n # mean
```

```
## [1] 46.875
```

Para la **media aritmética ponderada** se emplea lo siguiente.

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i \cdot x_i}{\sum_{i=1}^{n} w_i}$$

```
#pesos o valores de ponderación
w = c(1, 2, 2, 3, 3, 2, 2, 1)

#media aritmética ponderada
sum(w*x)/sum(w)
```

```
## [1] 43.375
```

Para la **media geométrica** se emplea lo siguiente.

$$\bar{x}_G = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$$

```r
#media geométrica
#prod() da el productorio del vector
prod(x)^(1/n)
```

```
## [1] 41.62073
```

```r
#da el mismo resultado así
prod(x^(1/n))
```

```
## [1] 41.62073
```

Para la **media armónica** se emplea lo siguiente.

$$\bar{x}_A = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

```r
#media armónica
n/sum(1/x)
```

```
## [1] 36.77301
```

3. Calcular el mínimo y el máximo de una variable cuantitativa

Se emplean las funciones min() y max() de R.

```r
min(x)
```

```
## [1] 17
```

```r
max(x)
```

```
## [1] 95
```