

4. Medidas de posição

2012



Também chamadas de medidas de **localização** e medidas de **tendência central** (*location* e *central tendency*).

Cada medida é um valor **representativo** dos dados x_1, x_2, \dots, x_n .

Redução (**drástica**) de n observações a **um só** valor.

4.1. Média aritmética (ou **média**) (*mean*)

Notação: \bar{x} (xis barra).

(a) Dados **brutos** x_1, x_2, \dots, x_n :
$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i.$$

(b) k diferentes **valores** e **frequências** $(x_1, f_1), \dots, (x_k, f_k)$:

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_k x_k}{f_1 + f_2 + \dots + f_k} = \frac{\sum_{j=1}^k f_j x_j}{\sum_{j=1}^k f_j} = \frac{1}{n} \sum_{j=1}^k f_j x_j.$$



(c) k diferentes valores e frequências relativas $(x_1, f_1^*), \dots, (x_k, f_k^*)$:

$$\bar{x} = \frac{f_1^* x_1 + f_2^* x_2 + \dots + f_k^* x_k}{f_1^* + f_2^* + \dots + f_k^*} = \frac{\sum_{j=1}^k f_j^* x_j}{\sum_{j=1}^k f_j^*} = \sum_{j=1}^k f_j^* x_j.$$

Obs. Em (b) e (c) as médias são ponderadas com pesos f_j e f_j^* , $j = 1, \dots, k$ (*weighted mean*).

(d) k intervalos de classe com pontos médios e frequências $(x_1^*, f_1), \dots, (x_k^*, f_k)$:

$$\bar{x} \cong \frac{f_1 x_1^* + f_2 x_2^* + \dots + f_k x_k^*}{f_1 + f_2 + \dots + f_k} = \frac{\sum_{j=1}^k f_j x_j^*}{\sum_{j=1}^k f_j} = \frac{1}{n} \sum_{j=1}^k f_j x_j^*.$$



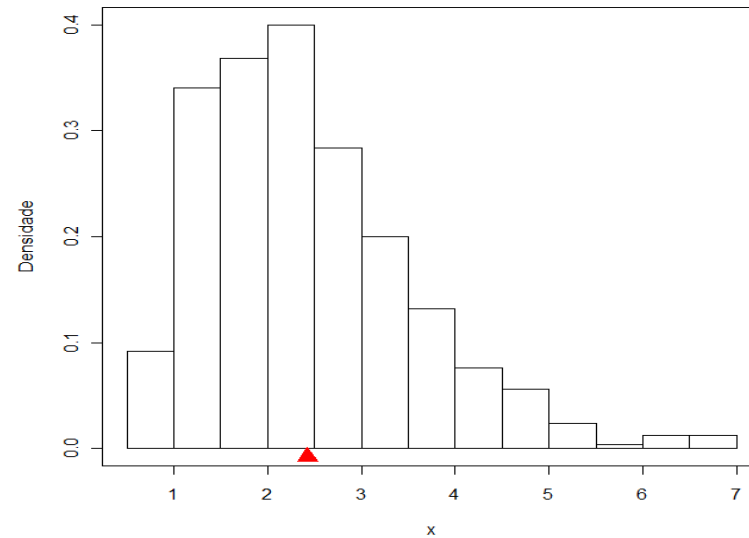
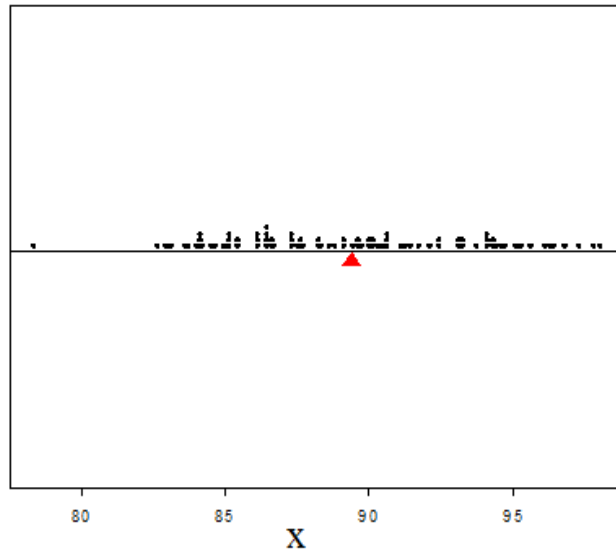
Propriedades da média

P1. Se $y_i = a + bx_i$, $i = 1, \dots, n$, a e b números reais, então $\bar{y} = a + b\bar{x}$.

P2. A média é o **centro de gravidade** (centro de massa) dos dados.

$$\sum_{i=1}^n (x_i - \bar{x}) = 0 \quad .$$

P3. $\sum_{j=1}^k f_j(x_j - \bar{x}) = 0$ ou $\sum_{j=1}^k f_j^*(x_j - \bar{x}) = 0$.



Propriedades da média

P4. $\sum_{i=1}^n (x_i - v)^2$ é minimizada se, e somente se, $v = \bar{x}$.

Curva de sensibilidade (CS).

g_n é uma função de x_1, \dots, x_n : $g_n(x_1, \dots, x_n)$.

Dados: x_1, \dots, x_{n-1} , g_n . Valor adicional: x .

Como quantificar a mudança em $g_{n-1}(x_1, \dots, x_{n-1})$ após a inclusão de x ?

$$CS_{g_n}(x) = n \{ g_n(x_1, \dots, x_{n-1}, x) - g_{n-1}(x_1, \dots, x_{n-1}) \}.$$

$$\begin{aligned} CS_x(x) &= n \left(\frac{x_1 + x_2 + \dots + x_{n-1} + x}{n} - \frac{x_1 + x_2 + \dots + x_{n-1}}{n-1} \right) \\ &= x - \frac{x_1 + x_2 + \dots + x_{n-1}}{n-1} = x - \bar{x}_{n-1}. \end{aligned}$$



Exemplo. Notas de 46 alunos

```
> x = c(c(4.8, 7.8, 9.9, 7.8, 6.8, 7.6, 8.7, 3.8, 6.9, 6.4,  
3.6, 7.8, 5.2, 3.5, 8.9, 8.0, 7.0, 6.7, 3.4, 7.5, 5.1, 5.1,  
4.9, 4.4, 7.2, 9.8, 3.9, 8.6, 5.2, 6.2, 3.8, 5.2, 4.6, 5.2,  
6.6, 6.7, 8.6, 5.1, 4.1, 5.0, 5.0, 4.7, 10.0, 4.8, 6.9, 5.3))
```

```
> (xb = mean(x))
```

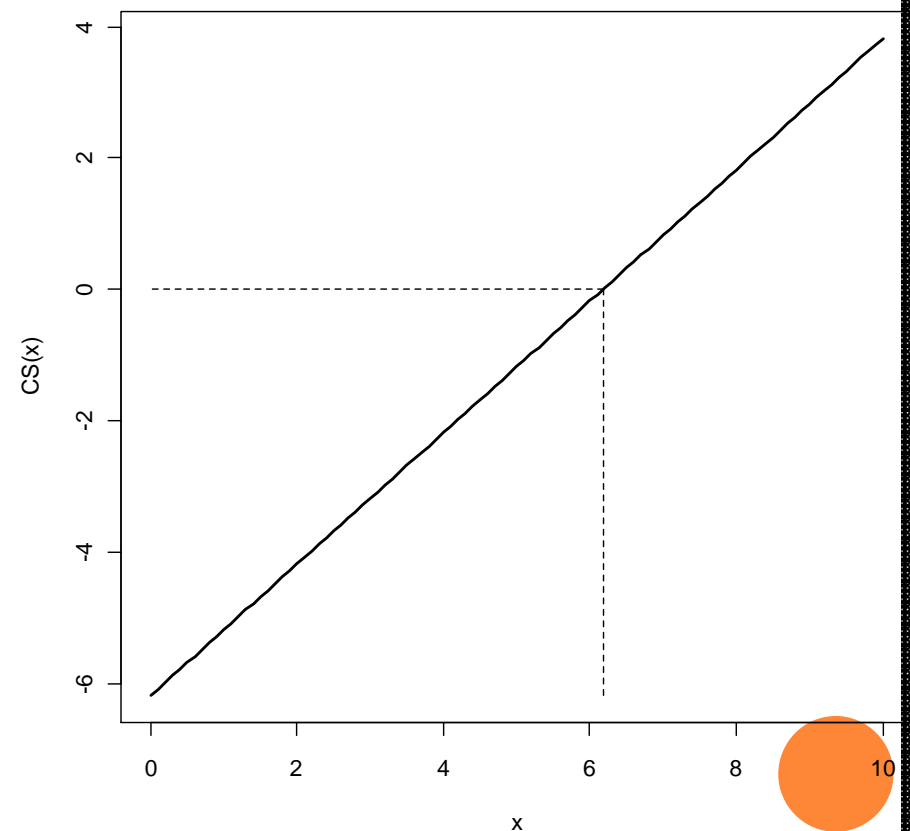
```
[1] 6.176087
```

```
> cs = function(x) { x - xb }
```

```
> plot(cs, 0, 10, lwd = 2, ylab  
= "CS(x)")
```

```
> segments(xb, cs(0), xb, 0, lty  
= 2)
```

```
> segments(xb, 0, 0, 0, lty = 2)
```



4.2. Mediana (*median*)

Dados ordenados: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$.

Se n é ímpar, a posição central é $c = (n + 1) / 2$.

Se n é par, as posições centrais são $c = n / 2$ e $c + 1 = n / 2 + 1$.

$M_{d_n} = M_n = x_{(c)}$, se n é ímpar;

$$= \frac{x_{(c)} + x_{(c+1)}}{2}, \text{ se } n \text{ é par.}$$

Obs. Ao contrário da média, no cálculo de M utilizamos apenas a observação central (ou as duas centrais).

Exemplo. Dados ordenados.

```
> (xord = sort(x))
```

3.4	3.5	3.6	3.8	3.8	3.9	4.1	4.4	4.6	4.7	4.8	4.8	4.9	5.0	5.0	5.1
5.1	5.1	5.2	5.2	5.2	5.2	5.3	6.2	6.4	6.6	6.7	6.7	6.8	6.9	6.9	
7.0	7.2	7.5	7.6	7.8	7.8	7.8	8.0	8.6	8.6	8.7	8.9	9.8	9.9	10.0	

$n = 46$: $c = n / 2 = 23$. Logo, $M_n = (x_{(23)} + x_{(24)}) / 2 = (5,3 + 6,2) / 2 = 5,75$.

Em R:

```
> M = (xord[23] + xord[24]) / 2
```



FOLHA DE S. PAULO

QUINTA-FEIRA, 18 DE MARÇO DE 2010 ★ C1

folha ribeirão

88% dos municípios da
região têm menos verde
que a média do Estado c1

78 cidades têm menos verde que a média

Só 11 dos 89 municípios da região têm cobertura vegetal acima do índice de 17% apurado em levantamento feito no Estado

O que podemos afirmar sobre a **mediana**?

É **maior** do que ou **menor** do que a **média** estadual (= 17%) ?

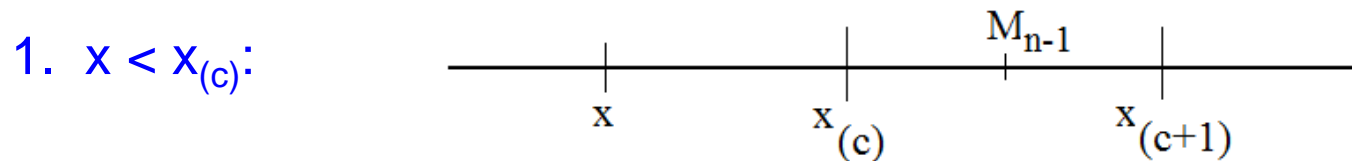


Curva de sensibilidade da mediana

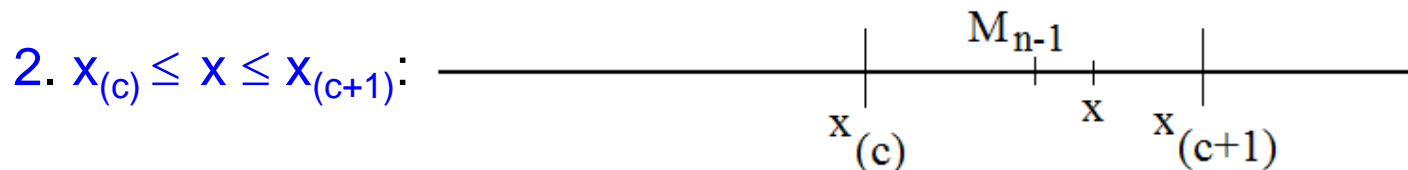
Dados: $n - 1$ observações, $n - 1$ é par. Logo, $c = (n - 1) / 2$.

Observação adicional: x (n observações, n é ímpar).

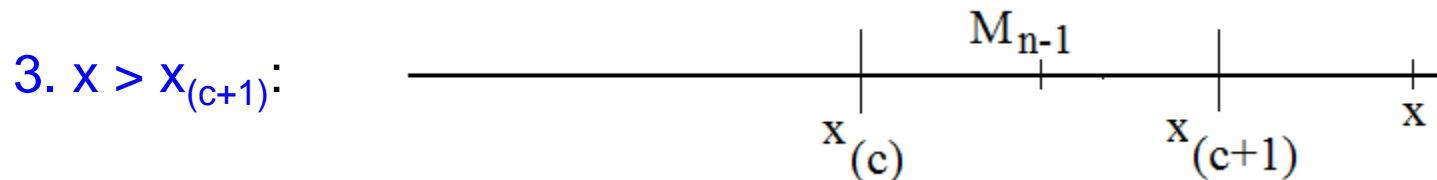
$$CS_{M_n}(x) = n \{ M_n(x_1, \dots, x_{n-1}, x) - M_{n-1}(x_1, \dots, x_{n-1}) \}.$$



Posição de $x_{(c)}$ passa a ser $(n - 1) / 2 + 1 = (n + 1) / 2$. Logo, $M = x_{(c)}$ (n ímpar).



Posição de x : $c + 1 = (n - 1) / 2 + 1 = (n + 1) / 2$. Logo, $M = x$ (n ímpar).



Posição de $x_{(c+1)}$: $c + 1 = (n - 1) / 2 + 1 = (n + 1) / 2$. Logo, $M = x_{(c+1)}$ (n ímpar).

Exercício. Obtenha a CS da mediana para $n - 1$ ímpar.

Curva de sensibilidade da mediana

1. $x < x_{(c)}$: $CS_{M_n}(x) = n \{ M_n(x_1, \dots, x_{n-1}, x) - M_{n-1}(x_1, \dots, x_{n-1}) \}$

$$= n \left(x_{(c)} - \frac{x_{(c)} + x_{(c+1)}}{2} \right) = n \left(\frac{x_{(c)} - x_{(c+1)}}{2} \right).$$

2. $x_{(c)} \leq x \leq x_{(c+1)}$:

$$CS_{M_n}(x) = n \{ M_n(x_1, \dots, x_{n-1}, x) - M_{n-1}(x_1, \dots, x_{n-1}) \}$$
$$= n \left(x - \frac{x_{(c)} + x_{(c+1)}}{2} \right) = n \{ x - M_{n-1}(x_1, \dots, x_{n-1}) \}.$$

3. $x > x_{(c+1)}$: $CS_{M_n}(x) = n \{ M_n(x_1, \dots, x_{n-1}, x) - M_{n-1}(x_1, \dots, x_{n-1}) \}$

$$= n \left(x_{(c+1)} - \frac{x_{(c)} + x_{(c+1)}}{2} \right) = n \left(\frac{x_{(c+1)} - x_{(c)}}{2} \right) = -n \left(\frac{x_{(c)} - x_{(c+1)}}{2} \right).$$

Obs. (a) CS é contínua.

(b) CS em (1) e (3) **não depende** de x .



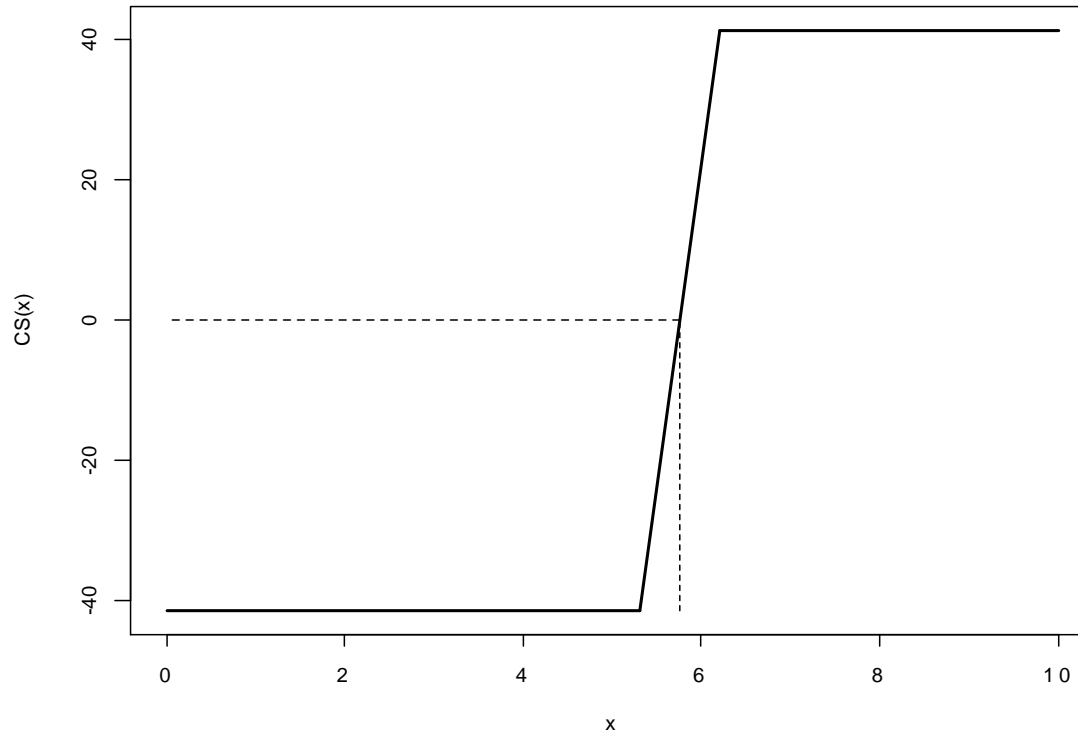
Exemplo

$n = 46$. $M_{n-1} = 5,75$. $c = 23$. $x_{(c)} = 5,3$ e $x_{(c+1)} = 6,2$.

1. $x < x_{(c)}$: $CS(x) = 46 \times (5,3 - 6,2) / 2 = -41,4$.

2. $x_{(c)} \leq x \leq x_{(c+1)}$: $CS(x) = 46 \times (x - 5,75)$

3. $x > x_{(c+1)}$: $CS(x) = 46 \times (6,2 - 5,3) = 41,4$.



Obs. A mediana é mais resistente do que a média.

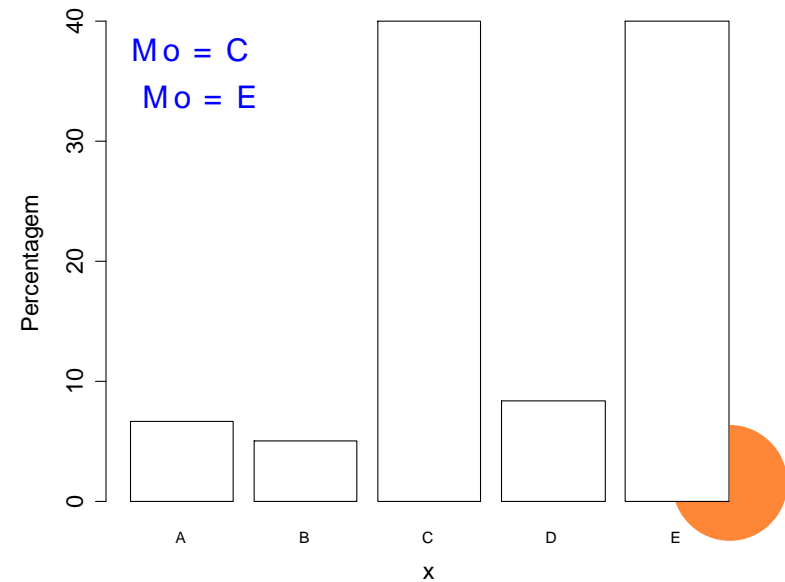
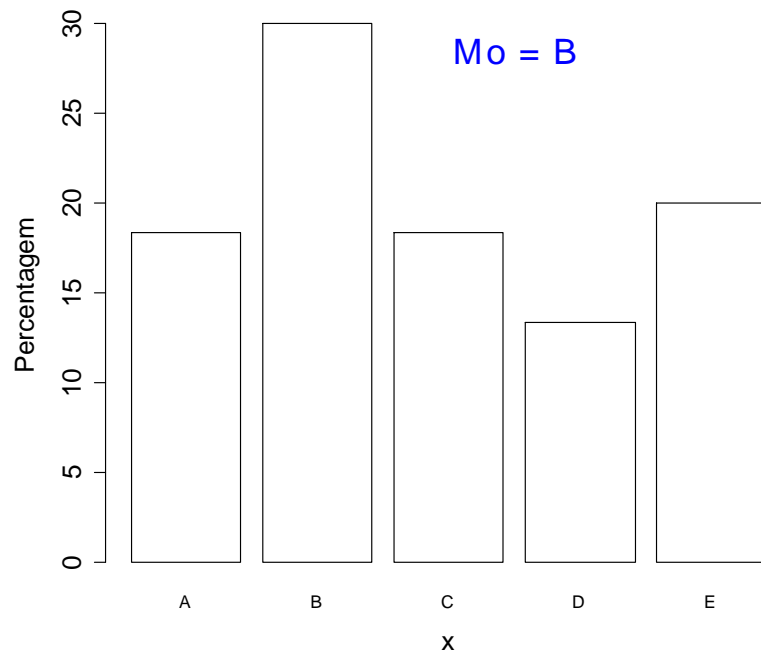
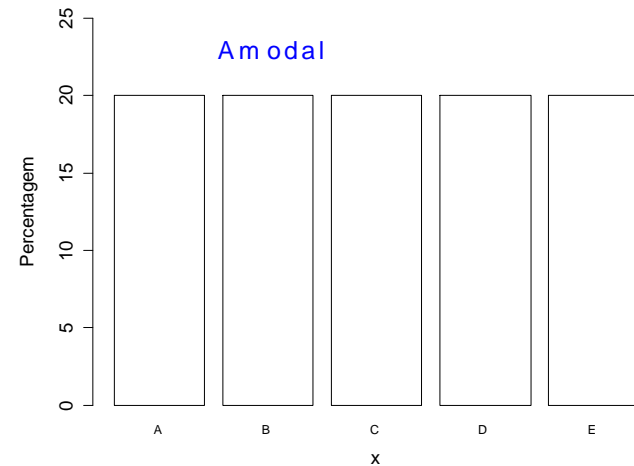
4.3. Moda (*mode*)

É o valor com **maior frequência**. Não necessariamente existe.

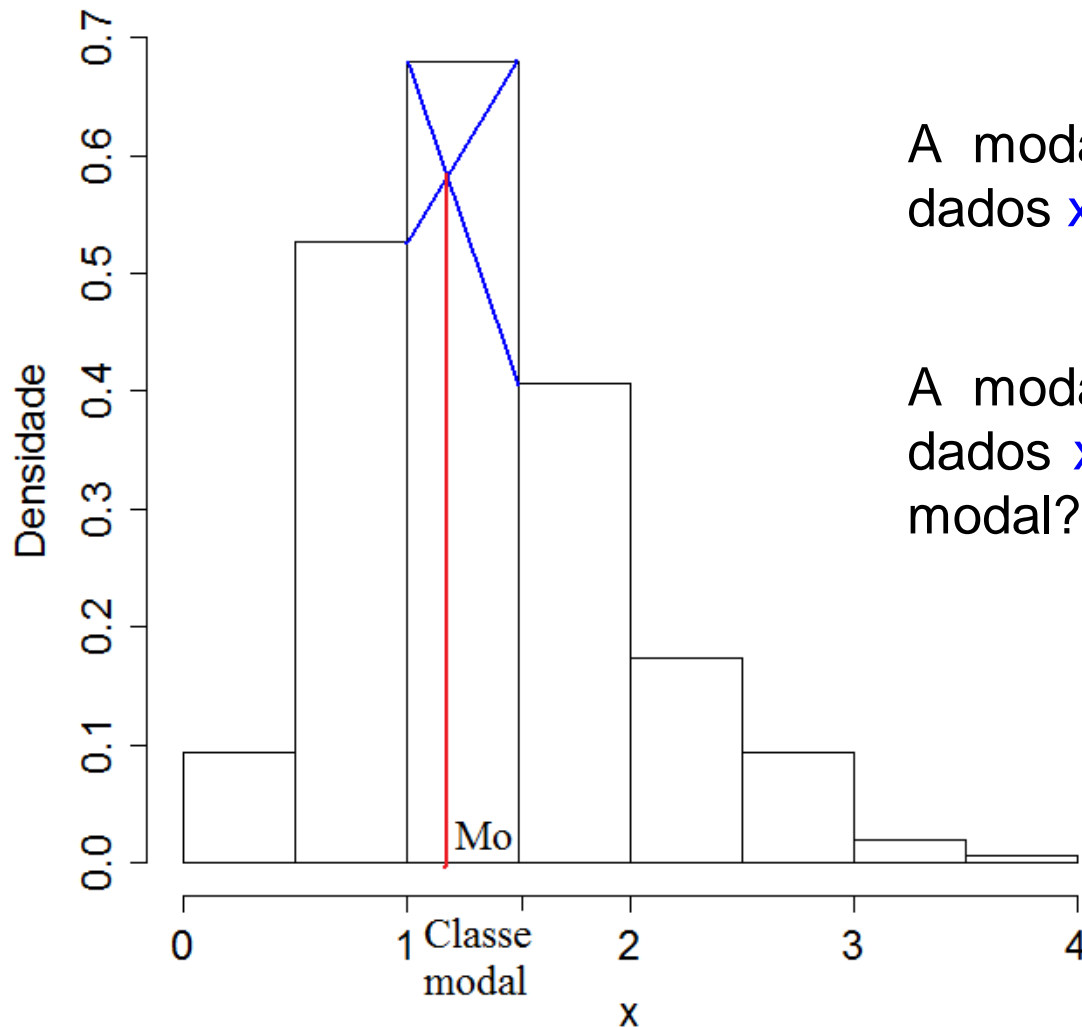
Pode haver **mais de uma** moda.

Quantitativas: menos utilizada do que a média e a mediana.

Qualitativas nominais: somente a moda pode ser calculada.



Cálculo aproximado para variável quantitativa em classes



A moda calculada a partir dos dados x_1, \dots, x_n é igual a Mo?

A moda calculada a partir dos dados x_1, \dots, x_n pertence à classe modal?



4.4. Média aparada (*trimmed mean*)

Chamada de **média winsonrizada** (*Winsorized mean*).

Média aparada de **100α%**: média aritmética dos dados após a eliminação das 100α% **menores** e das 100α% **maiores** observações, $0 \leq \alpha < \frac{1}{2}$.

Se $n = 20$ e $\alpha = 0,1$, eliminamos as **duas** menores ($10\% \times 20 = 2$) e as **duas** maiores observações, restando 16.

$$\bar{x}_{(\alpha)} = \frac{x_{(3)} + x_{(4)} + \dots + x_{(18)}}{16}.$$

Obs. (a) A média aparada é **mais resistente** do que a média.

(b) A mediana é um **caso particular** em que $\alpha = \frac{1}{2} - 1 / (2n)$.



Exemplo. Notas de 46 alunos

```
> x = c(4.8, 7.8, 9.9, 7.8, 6.8, 7.6, 8.7, 3.8, 6.9, 6.4,  
3.6, 7.8, 5.2, 3.5, 8.9, 8.0, 7.0, 6.7, 3.4, 7.5, 5.1, 5.1,  
4.9, 4.4, 7.2, 9.8, 3.9, 8.6, 5.2, 6.2, 3.8, 5.2, 4.6, 5.2,  
6.6, 6.7, 8.6, 5.1, 4.1, 5.0, 5.0, 4.7, 10.0, 4.8, 6.9, 5.3)
```

```
> (xb = mean(x))
```

```
[1] 6.176087
```

```
> (xa5 = mean(x, trim = 0.05))
```

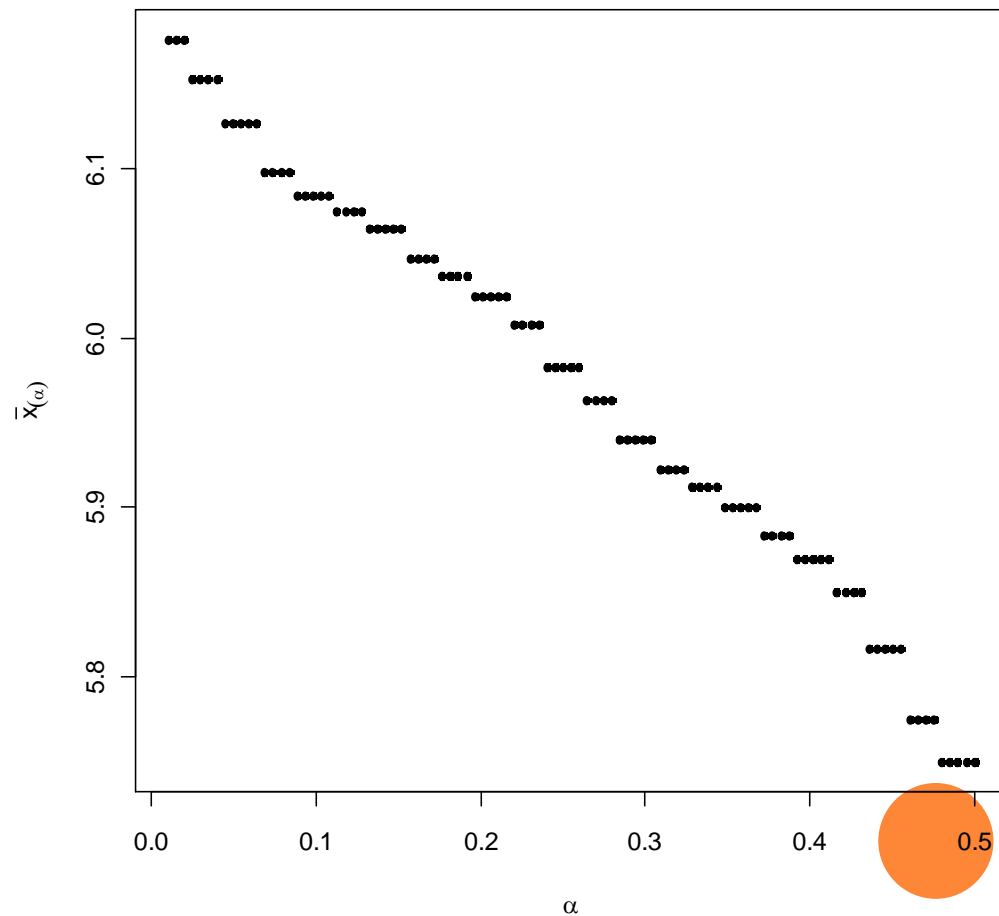
```
[1] 6.12619
```

```
> mean(x, trim = 0.1)
```

```
[1] 6.08421
```

```
> xa = function(alfa)  
{mapply(  
function (a0)  
mean(x, trim = a0), alfa)}
```

```
> plot(xa, 0.01, 0.5, xlab =  
expression(alpha), ylab =  
expression(bar(x)[alpha]),  
type = "p", pch = 20)
```



4.5. Média geométrica (*geometric mean*)

Exemplo. Em um certo mês o preço de um produto **aumentou 10%**. No mês seguinte **diminui 7%** e no terceiro mês **aumentou 5%**. De quanto foi a mudança **média** no preço?

Se x_1, \dots, x_n são **positivos**, a média geométrica é definida como

$$m_g = (x_1 \times x_2 \times \dots \times x_n)^{1/n} = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n} = \left(\prod_{i=1}^n x_i \right)^{1/n} .$$

Obs. Pode ser calculada usando logaritmos.

$$\log(m_g) = \frac{1}{n} \log \left(\prod_{i=1}^n x_i \right) = \frac{1}{n} \sum_{i=1}^n \log(x_i) = \overline{\log(x)} \Rightarrow m_g = \exp(\overline{\log(x)})$$

Exemplo. No 1º mês o preço foi multiplicado por 1,10. Nos outros dois meses o fator de correção foi 0,93 e 1,05.

A correção **total** nos três meses foi $1,10 \times 0,93 \times 1,05 = 1,07415$.

Mudança média = $1,07415^{1/3} = 1,024130$.

Houve um **aumento médio** mensal de cerca de **2,4%**.

Três correções sucessivas de 1,024130 têm o **mesmo efeito** das correções (+10%, -7%, +5%).



4.6. Média harmônica (*harmonic mean*)

Exemplo. Em três semanas de trabalho um viajante percorreu trechos de 400 km com velocidades de 60, 50 e 40 km/h. De quanto foi a velocidade média?

Se x_1, \dots, x_n são diferentes de 0, a média harmônica é definida como

$$m_h = \frac{1}{\left(\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} \right) / n}.$$

Exemplo. Os tempos de viagem (em h) foram $400 / 60 = 6,67$, $400 / 50 = 8$ e $400 / 40 = 10$.

Tempo médio de viagem = $(6,67 + 8 + 10) / 3 = 8,22$ h.

Velocidade média = $400 \text{ km} / 8,22 \text{ h} = 48,65 \text{ km/h}$.

Velocidade média = $m_h = 1 / \{(1 / 60 + 1 / 50 + 1 / 40) / 3\} = 48,65 \text{ km/h}$.

Desigualdade das médias. Se x_1, \dots, x_n são positivos, então

$$m_h \leq m_g \leq \bar{x}.$$



Fórmula geral das médias

As médias aritmética, geométrica e harmônica são casos particulares da fórmula geral das médias*

* $m(q)$ está definida somente para x_1, x_2, \dots, x_n positivos.

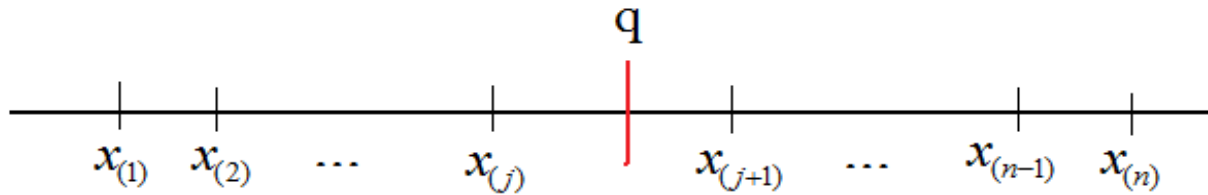
Exercício: que valores devem assumir q para obtermos as médias aritmética, geométrica e harmônica? (Lista 3)



4.7. Quantil (*quantile*)

Chamado de **separatriz**.

Provocam uma **divisão** nos dados **ordenados**.



Encontramos j observações **menores** ou iguais a q .

Chamamos q o **quantil j/n** (ou $100 \times j / n \%$), denotado por $q_{j/n}$.

Obs. Neste exemplo $q_{j/n}$ não é único.

Em geral, temos os quantis q_{α} , $0 \leq \alpha \leq 1$.

Quartis: $Q_1 = q_{0,25}$, $Q_2 = q_{0,5} = M$ e $Q_3 = q_{0,75}$.

Decis: $q_{0,1}$, $q_{0,2}, \dots, q_{0,8}$, $q_{0,9}$.

Percentis: $q_{1\%}$, $q_{2\%}, \dots, q_{98\%}$, $q_{99\%}$.

Obs. Informalmente dizemos que os (três) **quartis** dividem os dados em **quatro** intervalos cada um contendo **25%** das observações.

Quantis em R

Nove diferentes maneiras de calcular. A opção *default* (`type = 7`) difere do Minitab, SAS e SPSS.

Exemplo. Notas de 46 alunos (lâmina 15).

```
> summary(x)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 3.400  4.825   5.750   6.176   7.575  10.000
```

```
> quantile(x)
```

```
   0%    25%    50%    75%   100%
3.400  4.825  5.750  7.575 10.000
```

```
> (alfa = seq(0.1, 0.9, by = 0.1))
```

```
[1] 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
```

```
> (decis = quantile(x, probs = alfa))
```

```
 10%  20%  30%  40%  50%  60%  70%  80%  90%
3.85 4.70 5.00 5.20 5.75 6.70 7.10 7.80 8.65
```

```
> (decisSAS = quantile(x, probs = alfa, type = 3))
```

```
10% 20% 30% 40% 50% 60% 70% 80% 90%
3.8 4.6 5.0 5.1 5.3 6.7 7.0 7.8 8.6
```

