

# Medidas Resumo: Medidas de Posição, Medidas de Dispersão e Quantis

Gilberto Pereira Sassi

Universidade Federal da Bahia  
Instituto de Matemática e Estatística  
Departamento de Estatística

# Medidas de posição

## Medidas Resumo

Obter um ou mais números que sintetizem toda informação na amostra.

Consideraremos duas classes de medidas resumo: medidas de posição e medidas de dispersão.

## Medidas de Posição

Moda, Média e Mediana.

## Média

Suponha que os valores de uma variável  $X$  em uma amostra sejam  $x_1, \dots, x_n$ , então a média é calculada por

$$\bar{x} = \frac{x_1 + \dots + x_n}{n}.$$

## Exemplo

Considere as notas finais ( $X$ ) da Turma 1 de Estatística Aplicada à Saúde: 6,91; 7,85; 7,68; 8,64; 7,21; 8,04; 8,68; 4,37; 6,41; 7,89. Calcule a nota final média dessa turma.

**Solução:** Então a nota final média da Turma 1 é

$$\begin{aligned}\bar{x} &= \frac{6,91 + 7,85 + 7,68 + 8,64 + 7,21 + 8,04 + 8,68 + 4,37 + 6,41 + 7,89}{10} \\ &= 7,37.\end{aligned}$$

# Uso da Tabela de Distribuição de Frequência: Caso Discreto

Se  $X$  é uma variável quantitativa discreta com a seguinte tabela de distribuição de frequência

$X$	Frequência	Frequência Relativa (Proporção)	Porcentagem
$x_1$	$n_1$	$f_1 = n_1/n$	$100 \cdot f_1 \%$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_k$	$n_k$	$f_k = n_k/n$	$100 \cdot f_k \%$
Total	$n = n_1 + \dots + n_k$	1, 00	100%

então a média de  $X$  é dada por

$$\begin{aligned}
 \bar{X} &= \frac{\overbrace{x_1 + \dots + x_1}^{n_1 \text{ vezes}} + \overbrace{x_2 + \dots + x_2}^{n_2 \text{ vezes}} + \dots + \overbrace{x_k + \dots + x_k}^{n_k \text{ vezes}}}{n} \\
 &= \frac{n_1 \cdot x_1 + n_2 \cdot x_2 + \dots + n_k \cdot x_k}{n} \\
 &= \frac{\overbrace{n_1}^{f_1}}{n} \cdot x_1 + \frac{\overbrace{n_2}^{f_2}}{n} \cdot x_2 + \dots + \frac{\overbrace{n_k}^{f_k}}{n} \cdot x_k \\
 &= f_1 \cdot x_1 + f_2 \cdot x_2 + \dots + f_k \cdot x_k
 \end{aligned}$$

## Exemplo

Retome a variável Número de Filhos ( $Z$ ) da amostra com 36 funcionários da companhia MB cuja distribuição de frequência é dada por

Número de Filhos	Frequência	Frequência Relativa (Proporção)	Porcentagem
0	20	0,5556	55,56%
1	5	0,1389	13,89%
2	7	0,1944	19,44%
3	3	0,0833	8,33%
4	0	0,00	0,00%
5	1	0,0278	2,78%
Total	36	1,00	100%

Calcule a média da variável  $Z$ .

**Solução:** Então a média é dada por

$$\bar{z} = \frac{20 \cdot 0 + 1 \cdot 5 + 2 \cdot 7 + 3 \cdot 3 + 1 \cdot 5}{36}$$

$$= 0,92,$$

ou de forma alternativa

$$\bar{z} = 0,5556 \cdot 0 + 0,1389 \cdot 1 + 0,1944 \cdot 2 + 0,0833 \cdot 3 + 0,0278 \cdot 5$$

$$= 0,92.$$

# Uso da Tabela de Distribuição de Frequência: Caso Contínuo

## Observação

Para variáveis quantitativas contínuas também podemos usar a Tabela de Distribuição de Frequência.

Note que nesse caso teremos uma **aproximação** da média, pois perdemos informação ao agregar os valores em classes.

Considere a variável quantitativa contínua  $X$  cuja tabela de distribuição de frequência é

X	Frequência	Proporção	Porcentagem
$[l_1, l_2)$	$n_1$	$f_1 = n_1/n$	$100 \cdot f_1 \%$
$[l_2, l_3)$	$n_2$	$f_1 = n_2/n$	$100 \cdot f_2 \%$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$[l_k, l_{k+1}]$	$n_k$	$f_1 = n_k/n$	$100 \cdot f_k \%$
Total	$n = n_1 + \dots + n_k$	1, 00	100%

Usamos a simplificação: todos os valores observados de  $X$  que pertencem a classe

$l_i - \dots - l_{i+1}, i = 1, \dots, k$  são bem aproximados por  $\frac{l_i + l_{i+1}}{2}$ .

## Exemplo

Considere a variável quantitativa contínua salário ( $S$ ) da seção de orçamentos da companhia MB cuja tabela de distribuição de frequência é

$S$	Frequência	Frequência Relativa	Porcentagem	Ponto Médio
[4, 8)	10	$10/36 = 0,2778$	27,78%	$(4+8)/2 = 6$
[8, 12)	12	$12/36 = 0,3333$	33,33%	$(8+12)/2 = 10$
[12, 16)	8	$8/36 = 0,2222$	22,22%	$(12+16)/2 = 14$
[16, 20)	5	$5/36 = 0,1389$	13,89%	$(16+20)/2 = 18$
[20, 24]	1	$1/36 = 0,0278$	2,78%	$(20+24)/2 = 22$
Total	36	1,00	100%	—

**Solução:** Então a média salarial pode ser **aproximada** por

$$\begin{aligned}
 \bar{s} &= \frac{10 \cdot 6 + 12 \cdot 10 + 8 \cdot 14 + 5 \cdot 18 + 1 \cdot 22}{36} \\
 &= 0,2778 \cdot 6 + 0,3333 \cdot 10 + 0,2222 \cdot 14 + 0,1389 \cdot 18 + 0,0278 \cdot 22 \\
 &= 11,22.
 \end{aligned}$$

Note que a média salarial sem usar a tabela de distribuição de frequência é 11,12

Geralmente usamos essa medida de posição com variáveis quantitativas discretas.

## Moda

Realização mais frequente de uma variável.

## Exemplo

Considere a variável Número de Filhos ( $Z$ ) da seção de orçamentos da companhia MB cuja tabela de distribuição é dada por

Número de Filhos	Frequência	Frequência Relativa (Propoção)	Porcentagem
0	20	0,5556	55,56%
1	5	0,1389	13,89%
2	7	0,1944	19,44%
3	3	0,0833	8,33%
4	0	0,00	0,00%
5	1	0,0278	2,78%
Total	36	1,00	100%

Qual a moda?

**Solução:** A moda da variável Número de Filhos é 0.



## Mediana

Realização que ocupa a posição central da série de observações, ou seja, 50% das observações estão abaixo da mediana.

### Algoritmo para cálculo

Seja  $X$  uma variável quantitativa com valores observados  $x_1, \dots, x_n$ .

- 1) Ordenar os valores do menor ao maior:

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

- 2)

$$md(x) = \begin{cases} x_{(\frac{n+1}{2})}, & \text{se } n \text{ é ímpar,} \\ \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}, & \text{se } n \text{ é par,} \end{cases}$$

Note que  $x_{(1)}$  representa o menor valor de  $X$  na amostra,  $x_{(2)}$  representa o segundo menor valor de  $X$  na amostra,  $x_{(3)}$  representa o terceiro menor valor de  $X$  na amostra, e assim por diante. Chamamos  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$  de estatísticas de ordem.

# Exemplo

## Exemplo: tamanho amostral par.

Considere a variável quantitativa  $X$  com valores observados: 2, 8, 4. Calcule a mediana.

**Solução:** Primeiro ordenamos os valores

$$x_{(1)} = 2 \leq x_{(2)} = 4 \leq x_{(3)} = 8.$$

O tamanho amostral é  $n = 3$ , então

$$md(x) = x_{\left(\frac{3+1}{2}\right)} = x_{(2)} = 4.$$

## Exemplo: tamanho amostral ímpar.

Considere a variável quantitativa  $Y$  com valores observados: 1, 2, 3, 8. Calcule a mediana.

**Solução:** Primeiro ordenamos os valores

$$x_{(1)} = 1 \leq x_{(2)} = 2 \leq x_{(3)} = 3 \leq x_{(4)} = 8.$$

O tamanho amostral é  $n = 4$ , então

$$md(x) = \frac{x_{\left(\frac{4}{2}\right)} + x_{\left(\frac{4}{2}+1\right)}}{2} = \frac{x_{(2)} + x_{(3)}}{2} = \frac{2 + 3}{2} = 2,5.$$

# Uso da tabela de distribuição de frequência: caso discreto

Considere a variável Número de Filhos com tabela de distribuição de frequência dada por

Número de Filhos	Frequência	Frequência Relativa (Propoção)	Porcentagem
0	20	0,5556	55,56%
1	5	0,1389	13,89%
2	7	0,1944	19,44%
3	3	0,0833	8,33%
4	0	0,00	0,00%
5	1	0,0278	2,78%
Total	36	1,00	100%

Calcule a mediana.

**Solução:** Primeiro encontramos as estatísticas de ordem

$$x_{(1)} = x_{(2)} = \dots = x_{(20)} = 0$$

$$x_{(21)} = x_{(22)} = x_{(23)} = x_{(24)} = x_{(25)} = 1$$

$$x_{(26)} = x_{(27)} = x_{(28)} = x_{(29)} = x_{(30)} = x_{(31)} = x_{(32)} = 2$$

$$x_{(33)} = x_{(34)} = x_{(35)} = 3$$

$$x_{(36)} = 5$$

O tamanho amostral  $n = 36$  é par, então  $md(x) = \frac{x\left(\frac{36}{2}\right) + x\left(\frac{36}{2} + 1\right)}{2} = \frac{x_{(18)} + x_{(19)}}{2} = \frac{0 + 0}{2} = 0$ .

# Uso da tabela de distribuição de frequência: caso contínuo

## Observação

Para variáveis quantitativas contínuas também podemos usar a Tabela de Distribuição de Frequência.

Note que nesse caso teremos uma **aproximação** da mediana, pois perdemos informação ao agregar os valores em classes.

## Exemplo

Considere a variável salário ( $S$ ) da seção de orçamentos da companhia MB cuja tabela de distribuição de frequência é

$S$	Frequência	Frequência Relativa	Porcentagem	Ponto Médio
[4, 8)	10	$10/36 = 0,2778$	27,78%	$(4+8)/2 = 6$
[8, 12)	12	$12/36 = 0,3333$	33,33%	$(8+12)/2 = 10$
[12, 16)	8	$8/36 = 0,2222$	22,22%	$(12+16)/2 = 14$
[16, 20)	5	$5/36 = 0,1389$	13,89%	$(16+20)/2 = 18$
[20, 24]	1	$1/36 = 0,0278$	2,78%	$(20+24)/2 = 22$
Total	36	1,00	100%	--

Calcule a mediana.

## Solução exemplo

**Solução:** Primeiro encontramos as estatísticas de ordem

$$s_{(1)} = s_{(2)} = s_{(3)} = x_{(4)} = s_{(5)} = s_{(6)} = s_{(7)} = s_{(8)} = s_{(9)} = s_{(10)} = 6$$

$$s_{(11)} = s_{(12)} = s_{(13)} = s_{(14)} = s_{(15)} = s_{(16)} = s_{(17)} = s_{(18)} = s_{(19)} = s_{(20)} = s_{(21)} = s_{(22)} = 10$$

$$s_{(23)} = s_{(24)} = s_{(25)} = s_{(26)} = s_{(27)} = s_{(28)} = s_{(29)} = s_{(30)} = 14$$

$$s_{(31)} = s_{(32)} = s_{(33)} = s_{(34)} = s_{(35)} = 18$$

$$s_{(36)} = 22$$

Note que o tamanho amostral  $n = 36$  é par, logo

$$\begin{aligned} md(s) &= \frac{s_{(\frac{36}{2})} + s_{(\frac{36}{2}+1)}}{2} \\ &= \frac{s_{(18)} + s_{(19)}}{2} \\ &= \frac{10 + 10}{2} \\ &= 10. \end{aligned}$$

Note que 10 é uma aproximação para a mediana de salário cujo valor é 10,165 (usando os 36 valores observados na amostra).

## Exemplo

Um editor deseja estudar o número de erros de impressão de um livro. Para isso ele escolheu uma amostra de 50 páginas de um livro com a seguinte tabela de distribuição de frequência

Erro de impressão ( $X$ )	Frequência	Frequência Relativa	Porcentagem
0	25	$25/50 = 0,5$	$0,5 \cdot 100 = 50\%$
1	20	$20/50 = 0,4$	$0,4 \cdot 100 = 40\%$
2	3	$3/50 = 0,06$	$0,06 \cdot 100 = 6\%$
3	1	$1/50 = 0,02$	$0,02 \cdot 100 = 2\%$
4	1	$1/50 = 0,02$	$0,02 \cdot 100 = 2\%$
Total	50	1,00	100%

- a) Qual o número médio de erros por página?
- b) E o número mediano?
- c) Faça uma representação gráfica para a variável  $X$ .
- d) Se o livro tem 500 páginas, qual o número aproximado de erros de impressão?

## Solução – exemplo.

a)

$$\begin{aligned}\bar{x} &= \frac{25 \cdot 0 + 20 \cdot 1 + 3 \cdot 2 + 1 \cdot 3 + 1 \cdot 4}{50} \\ &= 0,5 \cdot 0 + 0,4 \cdot 1 + 0,06 \cdot 2 + 0,02 \cdot 3 + 0,02 \cdot 4 \\ &= 0,66\end{aligned}$$

b) Primeiro encontramos as estatísticas de ordem

$$\begin{aligned}x_{(1)} = x_{(2)} = x_{(3)} = \cdots = x_{(25)} &= 0; & x_{(26)} = x_{(27)} = x_{(28)} = \cdots = x_{(45)} &= 1 \\ x_{(46)} = x_{(47)} = x_{(48)} &= 2; & x_{(49)} &= 3; & x_{(50)} &= 4\end{aligned}$$

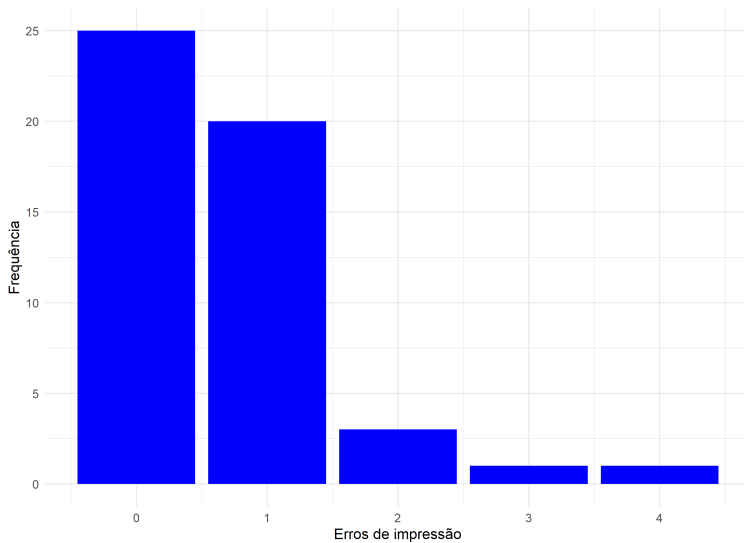
Note que  $n = 50$  é par, logo

$$md(x) = \frac{x_{(\frac{50}{2})} + x_{(\frac{50}{2}+1)}}{2} = \frac{x_{(25)} + x_{(26)}}{2} = \frac{0 + 1}{2} = 0,5.$$

d) Se um página tem aproximadamente 0,66 erros, então 500 páginas tem aproximadamente  $500 \cdot 0,66 = 330$  erros de impressão.

## Solução – exemplo: continuação

- c) **Interpretação:** Notamos que a maioria das páginas tem até dois erros de impressão.





# Motivação

## Observação

Note que a medida de posição pode mascarar a informação de como os dados estão dispersos.

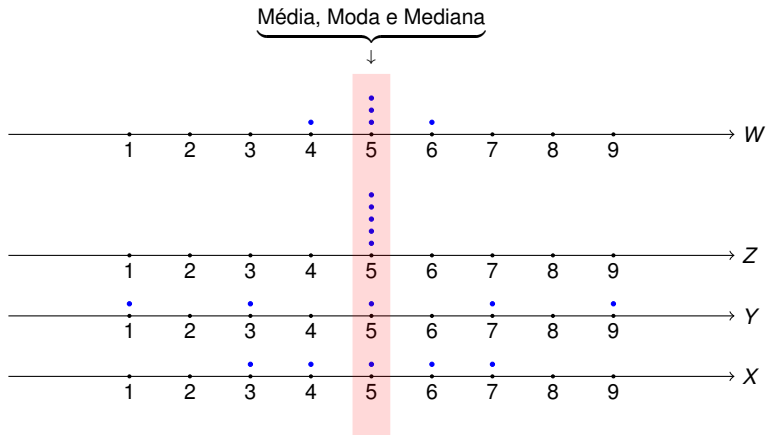
## Exemplo de motivação.

Um grupo de cinco alunos fizeram uma bateria de 5 testes, obtendo os seguintes resultados:

Teste	Notas					Representação da variável
A	3	4	5	6	7	X
B	1	3	5	7	9	Y
C	5	5	5	5	5	Z
D	4	5	5	6	5	W

**Exercício para casa:** verifique que a moda, média e mediana de X, Y, Z e W são iguais 5.

# Motivação – continuação



**Figura 1:** Representação gráfica para as variáveis  $X$ ,  $Y$ ,  $Z$ ,  $W$ .

# Desvio Médio

## Limitação das medidas de posição

As variáveis  $X$ ,  $Y$ ,  $Z$  e  $W$  tem a mesma média, mediana e moda, mas na Figura 1 percebemos que as quatro variáveis não são semelhantes. Algumas variáveis tem valor mais acumulado em torno da média (mediana ou moda) enquanto outras variáveis tem valores mais “heterogêneos”.

## Idea para superar a limitação das medidas de posição

Considere uma variável quantitativa com valores observados  $x_1, \dots, x_n$  e média  $\bar{x}$ , então

- i. Calcule a distância (em valor absoluto) entre os valores observados e uma medida de posição (geralmente a média):  $|x_1 - \bar{x}|, |x_2 - \bar{x}|, \dots, |x_n - \bar{x}|$ ;
- ii. Considere um valor representativo dessas distâncias, isto é, uma medida de posição de  $\{|x_1 - \bar{x}|, |x_2 - \bar{x}|, \dots, |x_n - \bar{x}|\}$ .

Se o valor obtido em ii. for pequeno os valores estão concentrados em torno da medida de posição (média) e são homogêneos.

Finalmente, podemos o Desvio Médio:

$$dm(x) = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{n}.$$

Note que usamos a média como medida de posição em ii.

# Variância e Desvio Padrão

## Idea para superar a limitação das medidas de posição

Considere uma variável quantitativa com valores observados  $x_1, \dots, x_n$  e média  $\bar{x}$ , então

- i. Calcule a distância (ao quadrado) entre os valores observados e uma medida de posição (geralmente a média):  $(x_1 - \bar{x})^2, (x_2 - \bar{x})^2, \dots, (x_n - \bar{x})^2$ ;
- ii. Considere um valor representativo dessas distâncias ao quadrado, isto é, uma medida de posição de  $\{(x_1 - \bar{x})^2, (x_2 - \bar{x})^2, \dots, (x_n - \bar{x})^2\}$

Se o valor obtido em ii. for pequeno os valores estão concentrados em torno da medida de posição (média) e são homogêneos.

Finalmente, podemos introduzir a Variância:

$$\text{Var}(x) = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}.$$

Note que usamos a média como medida de posição em ii.

Para manter a mesma unidade de  $X$ , é comum usar o Desvio Padrão

$$\text{DP}(x) = \sqrt{\text{Var}(x)}.$$

## Motivação

Em nosso exemplo de motivação temos que

$\text{Var}(x) = 2$	$\text{Var}(y) = 8$	$\text{Var}(z) = 0$	$\text{Var}(w) = 0,4$
$\text{DP}(x) = 1,4$	$\text{DP}(y) = 2,8$	$\text{DP}(z) = 0$	$\text{DP}(w) = 0,6$
$\text{dm}(x) = 1,2$	$\text{dm}(y) = 2,4$	$\text{dm}(z) = 0$	$\text{dm}(w) = 0,4$

e notamos que as variáveis não são semelhantes (valores são dispersos de forma diferente).

## Exemplo

Considere as notas finais ( $X$ ) da Turma 1 de Estatística Básica A: 6,91; 7,85; 7,68; 8,64; 7,21 Calcule a nota final média dessa turma.

**Solução:** Primeiramente, calculamos a média

$$\bar{x} = \frac{6,91 + 7,85 + 7,68 + 8,64 + 7,21}{5} = 7,66$$

Então, o desvio médio é

$$\begin{aligned} dm(x) &= \frac{|6,91 - \bar{x}| + |7,85 - \bar{x}| + |7,68 - \bar{x}| + |8,64 - \bar{x}| + |7,21 - \bar{x}|}{5} \\ &= \frac{|6,91 - 7,66| + |7,85 - 7,66| + |7,68 - 7,66| + |8,64 - 7,66| + |7,21 - 7,66|}{5} \\ &= 0,48 \end{aligned}$$

e a variância é

$$\begin{aligned} \text{Var}(x) &= \frac{(6,91 - \bar{x})^2 + (7,85 - \bar{x})^2 + (7,68 - \bar{x})^2 + (8,64 - \bar{x})^2 + (7,21 - \bar{x})^2}{5} \\ &= \frac{(6,91 - 7,66)^2 + (7,85 - 7,66)^2 + (7,68 - 7,66)^2 + (8,64 - 7,66)^2 + (7,21 - 7,66)^2}{5} \\ &= 0,35 \end{aligned}$$

e o desvio padrão é dado por  $DP = \sqrt{0,35} = 0,59$ .

# Uso da tabela de distribuição de frequência: caso discreto

Considere a variável Número de Filhos com tabela de distribuição de frequência dada por

Número de Filhos	Frequência	Frequência Relativa (Propoção)	Porcentagem
0	20	0,5556	55,56%
1	5	0,1389	13,89%
2	7	0,1944	19,44%
3	3	0,0833	8,33%
4	0	0,00	0,00%
5	1	0,0278	2,78%
Total	36	1,00	100%

Já calculamos a média anteriormente:  $\bar{x} = 0,92$ . Então, o desvio médio é dado por

$$dm(z) = \frac{20 \cdot |0 - 0,92| + 5 \cdot |1 - 0,92| + 7 \cdot |2 - 0,92| + 3 \cdot |3 - 0,92| + 0 \cdot |4 - 0,92| + 1 \cdot |5 - 0,92|}{36}$$

$$= 1,02$$

e a variância é dada por

$$Var(z) = \frac{20 \cdot (0 - 0,92)^2 + 5 \cdot (1 - 0,92)^2 + 7 \cdot (2 - 0,92)^2 + 3 \cdot (3 - 0,92)^2 + 0 \cdot (4 - 0,92)^2 + 1 \cdot (5 - 0,92)^2}{36}$$

$$= 1,52$$

e o desvio padrão é  $\sqrt{Var(z)} = 1,23$ .

# Uso da Tabela de Distribuição de Frequência: Caso Contínuo

## Observação

Para variáveis quantitativas contínuas também podemos usar a Tabela de Distribuição de Frequência.

Note que nesse caso teremos uma **aproximação** das medidas de dispersão, pois perdemos informação ao agregar os valores em classes.

Considere a variável quantitativa contínua salário ( $S$ ) da seção de orçamentos da companhia MB cuja tabela de distribuição de frequência é

$S$	Frequência	Frequência Relativa	Porcentagem	Ponto Médio
$[4, 8)$	10	$10/36 = 0,2778$	27, 78%	$(4+8)/2 = 6$
$[8, 12)$	12	$12/36 = 0,3333$	33, 33%	$(8+12)/2 = 10$
$[12, 16)$	8	$8/36 = 0,2222$	22, 22%	$(12+16)/2 = 14$
$[16, 20)$	5	$5/36 = 0,1389$	13, 89%	$(16+20)/2 = 18$
$[20, 24]$	1	$1/36 = 0,0278$	2, 78%	$(20+24)/2 = 22$
Total	36	1,00	100%	--

Calcule o desvio médio, a variância e o desvio padrão.

# Continuação – exemplo

Já vimos anteriormente, que a média salarial pode ser aproximada por 11,22. Então,

Desvio Médio

$$\begin{aligned} dm(s) &= \frac{10 \cdot |6 - 11,22| + 12 \cdot |10 - 11,22| + 8 \cdot |14 - 11,22| + 5 \cdot |18 - 11,22| + 1 \cdot |22 - 11,22|}{36} \\ &= 3,72; \end{aligned}$$

Variância

$$\begin{aligned} \text{Var}(s) &= \frac{10 \cdot (6 - 11,22)^2 + 12 \cdot (10 - 11,22)^2 + 8 \cdot (14 - 11,22)^2 + 5 \cdot (18 - 11,22)^2 + 1 \cdot (22 - 11,22)^2}{36} \\ &= 19,40; \end{aligned}$$

Desvio Padrão

$$DP(s) = \sqrt{\text{Var}(s)} = \sqrt{19,40} = 4,40.$$



# Quantis

## Ideia

Outra abordagem para medidas de posição de forma semelhante a mediana, substituindo 50% por  $100 \cdot p\%$ .

## Definição

Dizemos que um número  $q(p) \in \mathbb{R}$  é quantil de ordem  $p$  ou  $p$ -quantil se  $100 \cdot p\%$  das observações  $x_1, \dots, x_n$  forem menores que  $q(p)$ .

## Alguns quantis importantes e seus nomes particulares

$q(0, 25)$  Primeiro Quartil ( $q_1$ );

$q(0, 5)$  Segundo Quartil ( $q_2$ ) – sinônimo de mediana;

$q(0, 75)$  Terceiro Quartil ( $q_3$ ).

# Algoritmo para cálculo de quantis

Seja  $X$  uma variável quantitativa com  $x_1, \dots, x_n$  seus valores observados na amostra.

- i. Ordene os valores do menor ao maior (encontre as estatísticas de ordem)

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

em que  $x_{(1)}$  é o menor valor em  $\{x_1, \dots, x_n\}$ ,  $x_{(2)}$  é o segundo menor valor em  $\{x_1, \dots, x_n\}$ ,  $x_{(3)}$  é o terceiro menor valor em  $\{x_1, \dots, x_n\}$ , e assim prosseguimos até  $x_{(n)}$ : o último menor valor em  $\{x_1, \dots, x_n\}$

ii.

$$q(p) = \begin{cases} x_{((n+1) \cdot p)}, & \text{se } (n+1) \cdot p \text{ é número inteiro,} \\ \frac{x_{(\lfloor (n+1) \cdot p \rfloor)} + x_{(\lceil (n+1) \cdot p \rceil)}}{2}, & \text{se } (n+1) \cdot p \text{ não é número inteiro.} \end{cases}$$

em que  $\lfloor \cdot \rfloor$  é a função “arredonda para baixo” e  $\lceil \cdot \rceil$  é a função “arredonda para cima”.

## Exemplo

Considere a variável quantitativa  $X$  com os seguintes valores observados: 15, 5, 3, 8, 10, 2, 7, 11, 12. Calcule o primeiro, o segundo e terceiro quartis.

**Solução:** Primeiro encontramos as estatísticas de ordem:

$$x_{(1)} = 2 \leq x_{(2)} = 3 \leq x_{(3)} = 5 \leq x_{(4)} = 7$$

$$x_{(5)} = 8 \leq x_{(6)} = 10 \leq x_{(7)} = 11 \leq x_{(8)} = 12 \leq x_{(9)} = 15$$

Os quartis são dados por

$q_1$  Note que  $(n+1) \cdot 0,25 = (9+1) \cdot 0,25 = 2,5$ , e  $[2,5] = 2$  e  $[2,5] = 3$ . Então,

$$q_1 = \frac{x_{(2)} + x_{(3)}}{2} = \frac{3 + 5}{2} = 4;$$

$q_2$  Note que  $(n+1) \cdot 0,5 = (9+1) \cdot 0,5 = 5$ . Então,

$$q_2 = x_{(5)} = 8;$$

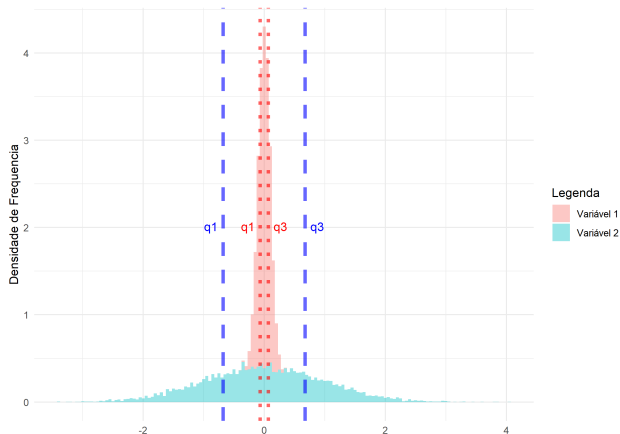
$q_3$  Note que  $(n+1) \cdot 0,75 = (9+1) \cdot 0,75 = 7,5$ , e  $[7,5] = 7$  e  $[7,5] = 8$ . Então,

$$q_3 = \frac{x_{(7)} + x_{(8)}}{2} = \frac{11 + 12}{2} = 11,5.$$

# Intervalo Interquartilício

## Ideia

Se a distância entre  $q_1$  e  $q_3$  for pequena, então os valores da variável estão concentrados em uma região.



## Definição

Seja  $X$  uma variável quantitativa com valores observados  $x_1, \dots, x_n$ , então o intervalo interquartilico é dado por

$$dq = q_3 - q_1$$

## Exemplo

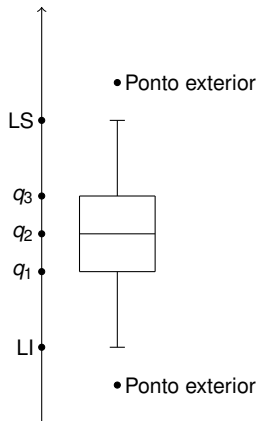
Considere a variável quantitativa  $X$  com os seguintes valores observados: 15, 5, 3, 8, 10, 2, 7, 11, 12. Calcule o intervalo interquartilico.

**Solução:** Já calculamos o primeiro e terceiro quartis para essa variável e essa amostra, então

$$dq = q_3 - q_1 = 11,5 - 4 = 7,5.$$

# Diagrama de Caixa ou Boxplot

O diagrama de caixa tem o seguinte aspecto



# Diagrama de Caixa ou Boxplot

Em que

Limite Superior  $LS = q_3 + 1,5 \cdot dq;$

Limite Inferior  $LI = q_1 - 1,5 \cdot dq;$

Ponto Adjacente Todos os valores da variável entre  $LI$  e  $LS$ ;

Ponto Exterior Todos os valores da variável que não estão entre  $LI$  e  $LS$ . Estes valores da variável são provavelmente destoantes que precisam de atenção do pesquisador;

## Exemplo 1

Considere as notas da Turma 1 de Estatística Aplicada à Saúde: 9,44; 9,26; 9,21; 9,51; 8,53; 8,4; 7,74; 8,75; 9,8; 9,5; 9,38; 8,36; 8,57; 9,18; 9,53. Desenhe o digrama de caixa.

**Solução:** Primeiro encontramos as estatísticas de ordem:

$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$	$x_{(9)}$	$x_{(10)}$	$x_{(11)}$	$x_{(12)}$	$x_{(13)}$	$x_{(14)}$	$x_{(15)}$
7,74	8,36	8,40	8,53	8,57	8,75	9,18	9,21	9,26	9,38	9,44	9,50	9,51	9,53	9,80

Em seguida, calculamos o primeiro quartil, o segundo quartil, o terceiro quartil, o intervalo interquartil, o limite superior e o limite inferior:

$$(15 + 1) \cdot 0,25 = 4$$

$$q_1 = x_{(4)} = 8,53$$

$$dq = q_3 - q_1 = 0,97$$

$$(15 + 1) \cdot 0,5 = 8$$

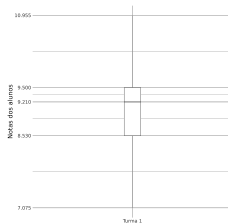
$$q_2 = x_{(8)} = 9,21$$

$$LS = q_3 + 1,5 \cdot dq = 10,955$$

$$(15 + 1) \cdot 0,75 = 12$$

$$q_3 = x_{(12)} = 9,50$$

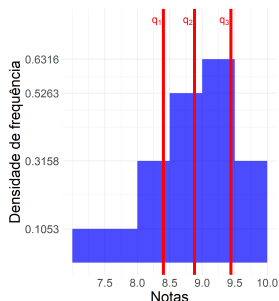
$$LI = q_1 - 1,5 \cdot dq = 7,075$$





## Exemplo 1

Note que os intervalos  $[q_1, q_2]$  e  $[q_2, q_3]$  têm 25% dos valores observados, ou seja, os valores estão mais concentrados no intervalo  $[q_2, q_3]$  do que  $[q_1, q_2]$ . Quando isso ocorre, dizemos a variável é assimétrica à esquerda. A figura abaixo ilustra essa ideia.



Se  $q_3 - q_2 < q_2 - q_1$ , dizemos que a variável tem assimetria a esquerda ou negativa ( $q_2$  mais próximo de  $q_3$ );

## Exemplo 2

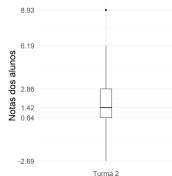
Considere as notas da Turma 2 de Estatística Aplicada à Saúde: 2,75; 4,54; 3,08; 4,74; 1,42; 0,61; 1,01; 1,61; 2,8; 8,93; 0,26; 0,58; 2,86; 0,08; 1,21; 1,44; 1,2; 1,24; 0,64. Desenhe o diagrama de caixa.

**Solução:** Primeiro encontramos as estatísticas de ordem:

$x_{(1)}$ 0,08	$x_{(2)}$ 0,26	$x_{(3)}$ 0,58	$x_{(4)}$ 0,61	$x_{(5)}$ 0,64	$x_{(6)}$ 1,01	$x_{(7)}$ 1,2	$x_{(8)}$ 1,21	$x_{(9)}$ 1,24	$x_{(10)}$ 1,42	$x_{(11)}$ 1,44	$x_{(12)}$ 1,61	$x_{(13)}$ 2,75	$x_{(14)}$ 2,8	$x_{(15)}$ 2,86
$x_{(16)}$ 3,08	$x_{(17)}$ 4,54	$x_{(18)}$ 4,74	$x_{(19)}$ 8,93											

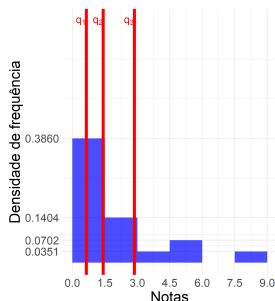
Em seguida, calculamos o primeiro quartil, o segundo quartil, o terceiro quartil, o intervalo interquartil, o limite superior e o limite inferior:

$$\begin{aligned}
 (19 + 1) \cdot 0,25 &= 5 & (19 + 1) \cdot 0,5 &= 10 & (19 + 1) \cdot 0,75 &= 15 \\
 q_1 &= x_{(5)} = 0,64 & q_2 &= x_{(10)} = 1,42 & q_3 &= x_{(15)} = 2,86 \\
 dq &= q_3 - q_1 = 2,22 & LS &= q_3 + 1,5 \cdot dq = 6,19 & LI &= q_1 - 1,5 \cdot dq = -2,69
 \end{aligned}$$



## Exemplo 1

Note que os intervalos  $[q_1, q_2]$  e  $[q_2, q_3]$  têm 25% dos valores observados, ou seja, os valores estão mais concentrados no intervalo  $[q_1, q_2]$  do que  $[q_2, q_3]$ . Quando isso ocorre, dizemos que a variável é assimétrica à direita. A Figura ilustra essa ideia.



Se  $q_2 - q_1 < q_3 - q_2$ , dizemos que a variável tem assimetria à direita ou positiva ( $q_2$  mais próximo de  $q_1$ );

## Assimetria

Inspirados nesses dois exemplos, podemos introduzir uma medida numérica de assimetria, denominado coeficiente de Bowley:

$$\begin{aligned} B &= \frac{q_3 - 2q_2 + q_1}{q_3 - q_1} \\ &= \frac{q_3 - q_2 - (q_2 - q_1)}{q_3 - q_1} \end{aligned}$$

Note que

- $B \in [-1, 1]$ ;
- existe assimetria positiva ou à direita  $\iff q_2 - q_1 < q_3 - q_2 \iff B > 0$ ;
- existe assimetria negativa ou à esquerda  $\iff q_2 - q_1 > q_3 - q_2 \iff B < 0$ ;
- a variável é simétrica se  $B \approx 0$ .

### Exemplos

- No exemplo 1,

$$B = \frac{q_3 - 2 \cdot q_2 + q_1}{q_3 - q_1} = \frac{9,5 - 2 \cdot 9,21 + 8,53}{0,97} = -0,40;$$

- No exemplo 2,

$$B = \frac{q_3 - 2 \cdot q_2 + q_1}{q_3 - q_1} = \frac{7,03 - 2 \cdot 6,08 + 5,71}{1,32} = 0,44.$$