



Universidade Federal de Mato Grosso (UFMT)
Instituto de Ciências Exatas e da Terra
Departamento de Estatística

PRINCÍPIOS DE BIOESTATÍSTICA

Prof. Dr. Neuber J. Segri

Medidas de Dispersão


Aulas 9 e 10

Medidas de Dispersão

- ▶ Também conhecidas como medidas de variabilidade.
- ▶ Juntamente com as medidas de tendência central, são utilizadas para descrever um conjunto de dados.



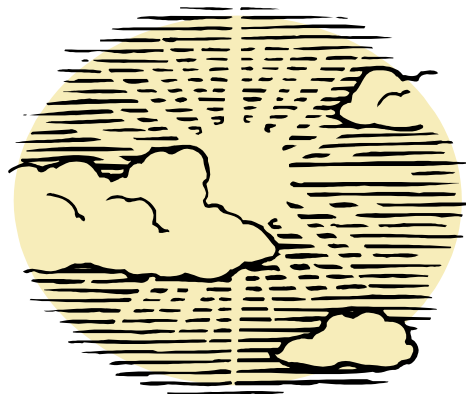
Medidas de Dispersão (variabilidade)

- ▶ Amplitude
 - ▶ Variância/Desvio Padrão
 - ▶ Coeficiente de Variação
- 

Amplitude

- ▶ É a diferença entre os valores máximo e mínimo de um conjunto de dados:

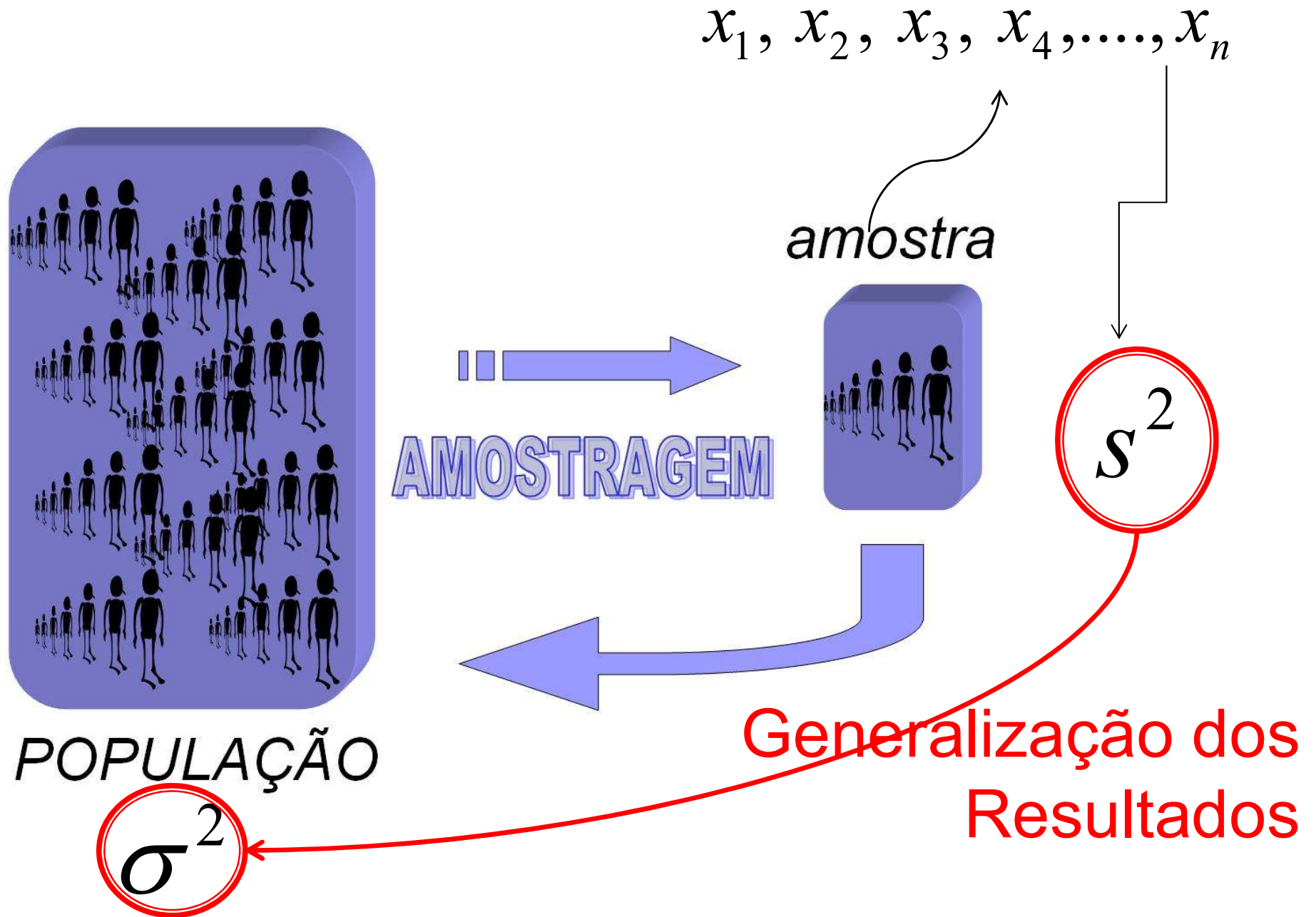
Amplitude = Valor Máximo - Valor Mínimo



Variância

- ▶ **Definição:** Mede a concentração dos dados ao redor da média.
 - Existe apenas para variáveis quantitativas (discretas ou contínuas).
 - Seu valor é único.
 - Não assume valores negativos.

Variância



Variância

- ▶ Definição: Mede a concentração dos dados ao redor da média.
- ▶ Variância Populacional: (σ^2)

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

Variância Amostral (s^2)

- ▶ Fórmula (valores individuais):

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Variância Amostral

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

▶ Procedimento para Cálculo:

- 1) Desvio = cada observação - média

$$d_i = x_i - \bar{x}$$

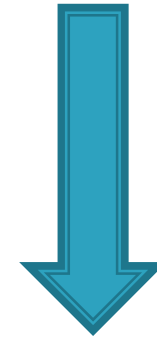
- 2) Elevar cada desvio ao quadrado
- 3) Somar os quadrados
- 4) Dividir o resultado por n-1

Desvio Padrão – (s)

- ▶ Raiz quadrada (positiva) da Variância

$$s^2 = \text{variância}$$

$$\sqrt{s^2} = \sqrt{\text{variância}} = \text{desvio padrão}$$



Variância Amostral

- ▶ Duração da terapia de transfusão para dez pacientes com determinada doença:

Indivíduo	Duração (anos)
1	4
2	8
3	7
4	11
5	5
6	8
7	10
8	3
9	5
10	6

Variância Amostral

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

- ▶ Duração da terapia de transfusão.

Indivíduo	Duração (anos)	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	4	-2,7	7,29
2	8	1,3	1,69
3	7	0,3	0,09
4	11	4,3	18,49
5	5	-1,7	2,89
6	8	1,3	1,69
7	10	3,3	10,89
8	3	-3,7	13,69
9	5	-1,7	2,89
10	6	-0,7	0,49
		Soma:	60,1

média = 6,7 anos

$$s^2 = \frac{60,1}{n-1} = \frac{60,1}{10-1}$$

$$s^2 \cong 6,678 \text{ anos}^2$$

Desvio Padrão Amostral

► Duração da terapia de transfusão

Indivíduo	Duração (anos)	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	4	-2,7	7,29
2	8	1,3	1,69
3	7	0,3	0,09
4	11	4,3	18,49
5	5	-1,7	2,89
6	8	1,3	1,69
7	10	3,3	10,89
8	3	-3,7	13,69
9	5	-1,7	2,89
10	6	-0,7	0,49
		Soma:	60,1

$$s^2 = \frac{60,1}{n-1} = \frac{60,1}{10-1}$$

$$s^2 \cong 6,678 \text{ anos}^2$$

$$s = 2,584 \text{ anos}$$

Variância Amostral (s^2)

- ▶ Fórmula (valores em distribuição de frequências):

$$s^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot f_i}{n - 1}$$

- f_i – frequência de indivíduos no intervalo i
- k – número de valores (linhas da tabela)

Variância

$$s^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot f_i}{n - 1}$$

- Exemplo: *Idade dos alunos do curso.*
 - Valores em distribuição de frequências:

$$\bar{x} = \frac{945}{50} = 18,9 \text{ anos}$$

Idade dos Alunos (xi)	desvio di = xi - média	desvio ao quadrado di^2	Frequência (fi)	di^2 . (fi)
17	-1,9	3,61	9	32,49
18	-0,9	0,81	22	17,82
19	0,1	0,01	7	0,07
20	1,1	1,21	4	4,84
21	2,1	4,41	3	13,23
22	3,1	9,61	0	0
23	4,1	16,81	2	33,62
24	5,1	26,01	1	26,01
25	6,1	37,21	2	74,42
			Soma	202,5

$$s^2 = \frac{202,5}{50 - 1}$$

$$s^2 \cong 4,13 \text{ anos}^2$$

$$s \cong 2,032 \text{ anos}$$

Variância Amostral (s^2)

- ▶ Fórmula (valores em intervalos de classe):

$$s^2 = \frac{\sum_{i=1}^k \left(x_{i\ pm} - \bar{x} \right)^2 \cdot f_i}{n - 1}$$

$x_{i\ pm}$ – ponto médio do intervalo

f_i – frequência de indivíduos no intervalo i

Variância

$$s^2 = \frac{\sum_{i=1}^k (x_{ipm} - \bar{x})^2 \cdot f_i}{n-1}$$

- Exemplo - *Nível de colesterol dos funcionários do Depto. Estatística da UFMT:*
 - Valores em intervalos de classe:

Nível de Colesterol (mg/100ml)	fi
180 --200	2
200 --250	10
250 --300	6
300 --350	2
Total	20

Variância

$$\bar{x} = 246,5 \text{ mg}(100\text{ml})$$

- Exemplo - *Nível de colesterol dos funcionários do Depto. Estatística da UFMT:*
 - Valores em intervalos de classe:

Nível de Colesterol	Frequência (fi)	Xi pm	(Xi pm - média)	(Xi pm - média)^2	(Xi pm - média)^2.fi
180 -- 200	2	190	-56,5	3192,25	6384,5
200 -- 250	10	225	-21,5	462,25	4622,5
250 -- 300	6	275	28,5	812,25	4873,5
300 -- 350	2	325	78,5	6162,25	12324,5
Total	20	--	--	--	28205

$$s^2 = \frac{\sum_{i=1}^k (x_{i\text{ pm}} - \bar{x})^2 \cdot f_i}{n-1} = \frac{28205}{19} = 1484,47(\text{mg}(100\text{ml}))^2$$

Desvio Padrão

$$\bar{x} = 246,5 \text{ mg}(100\text{ml})$$

- Exemplo - *Nível de colesterol dos funcionários do Depto. Estatística da UFMT:*
 - Valores em intervalos de classe:

Nível de Colesterol	Frequência (fi)	Xi pm	(Xi pm - média)	(Xi pm - média)^2	(Xi pm - média)^2.fi
180 -- 200	2	190	-56,5	3192,25	6384,5
200 -- 250	10	225	-21,5	462,25	4622,5
250 -- 300	6	275	28,5	812,25	4873,5
300 -- 350	2	325	78,5	6162,25	12324,5
Total	20	--	--	--	28205

$$s^2 = \frac{\sum_{i=1}^k (x_{i\text{ pm}} - \bar{x})^2 \cdot f_i}{n-1} = \frac{28205}{19} = 1484,47$$

$$s^2 = 1484,47 \rightarrow s = 38,53 \text{ mg}(100\text{ml})$$

Coeficiente de Variação

- ▶ **Definição:** medida relativa de dispersão
 - Utilizado para comparar os graus de dispersão de duas variáveis medidas em unidades distintas.
 - Fórmula: razão entre o desvio-padrão e a média.

$$CV = \frac{\text{desvio padrão}}{\text{média}} = \frac{s}{x} \cdot 100\%$$

Coeficiente de Variação

- ▶ Ainda no exemplo (duração da terapia):

$$CV = \frac{\text{desvio padrão}}{\text{média}} = \frac{s}{x} = \frac{2,584 \text{ anos}}{6,7 \text{ anos}}$$

$$CV = \frac{s}{x} \cong 0,386$$

O desvio-padrão é aproximadamente
38,6% da média

Coeficiente de Variação $CV = \frac{\text{desvio padrão}}{\text{média}}$


- ▶ Pesos e estaturas de um grupo de pessoas:
 - Estatura: (média = 175cm ; s = 5cm)
 - Peso: (média = 68kg ; s = 2kg)

$$CV_{\text{Estatura}} = \frac{s}{x} = \frac{5 \text{ cm}}{175 \text{ cm}} \cong 0,0285 \cong 2,85\%$$

$$CV_{\text{Peso}} = \frac{s}{x} = \frac{2 \text{ kg}}{68 \text{ kg}} \cong 0,0294 \cong 2,94\%$$

Os pesos apresentam maior grau de dispersão que as alturas

Box-Plot / Desenho esquemático

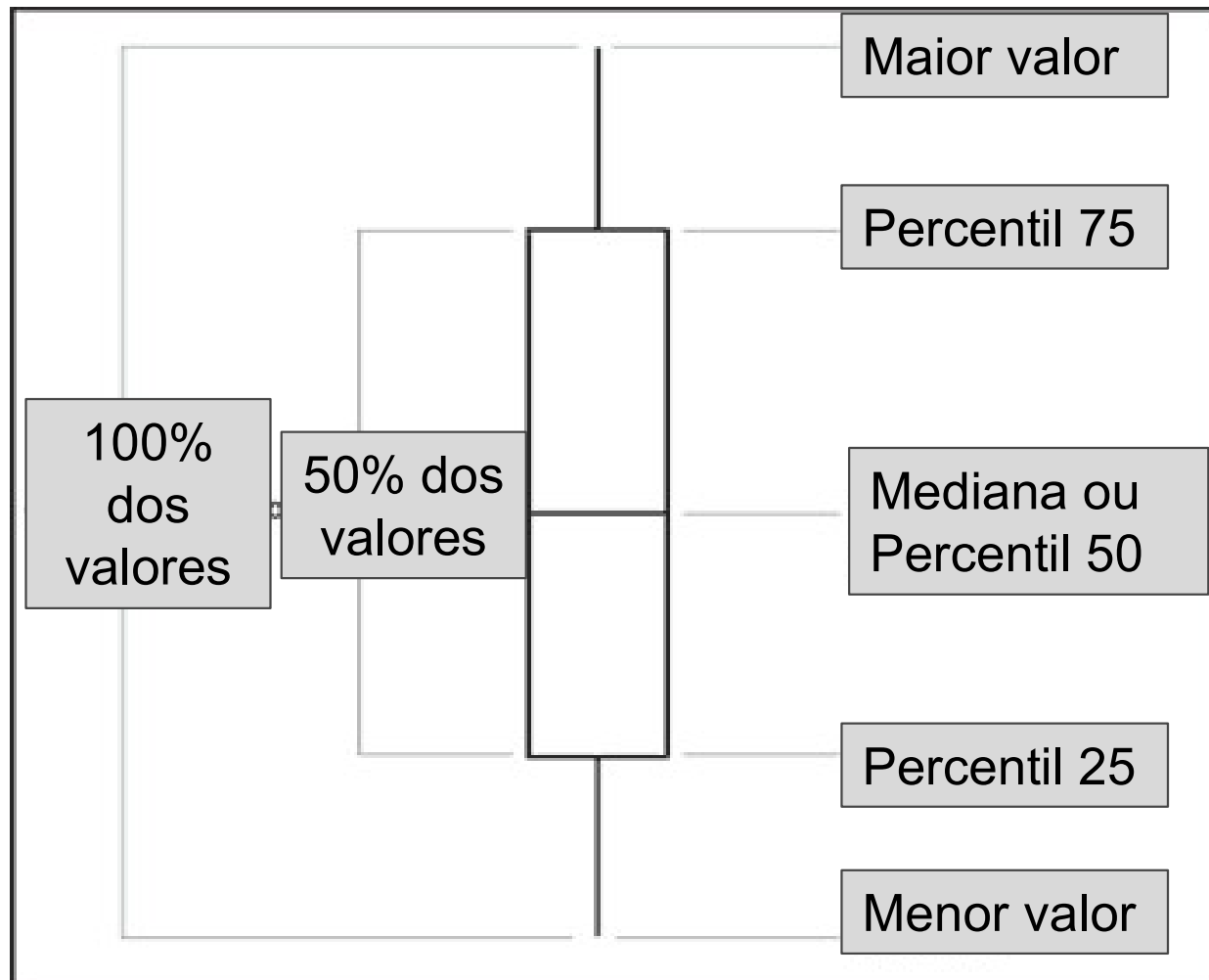
- Também conhecido como Diagrama de Caixa.
 - Adequado para representar graficamente um conjunto de dados quantitativos, de forma resumida.
 - Mostra as principais características do conjunto de dados: (dispersão, simetria e valores extremos).
 - Ideal para comparações entre grupos.
- 

Construindo o Box-Plot

- ▶ São necessárias as seguintes medidas:
 - Mediana
 - Q1 – 1º quartil
 - Q3 – 3º quartil
 - Distância interquartílica
 - $(Q3 - Q1)$
 - Amplitude:
 - Valor mínimo
 - Valor máximo

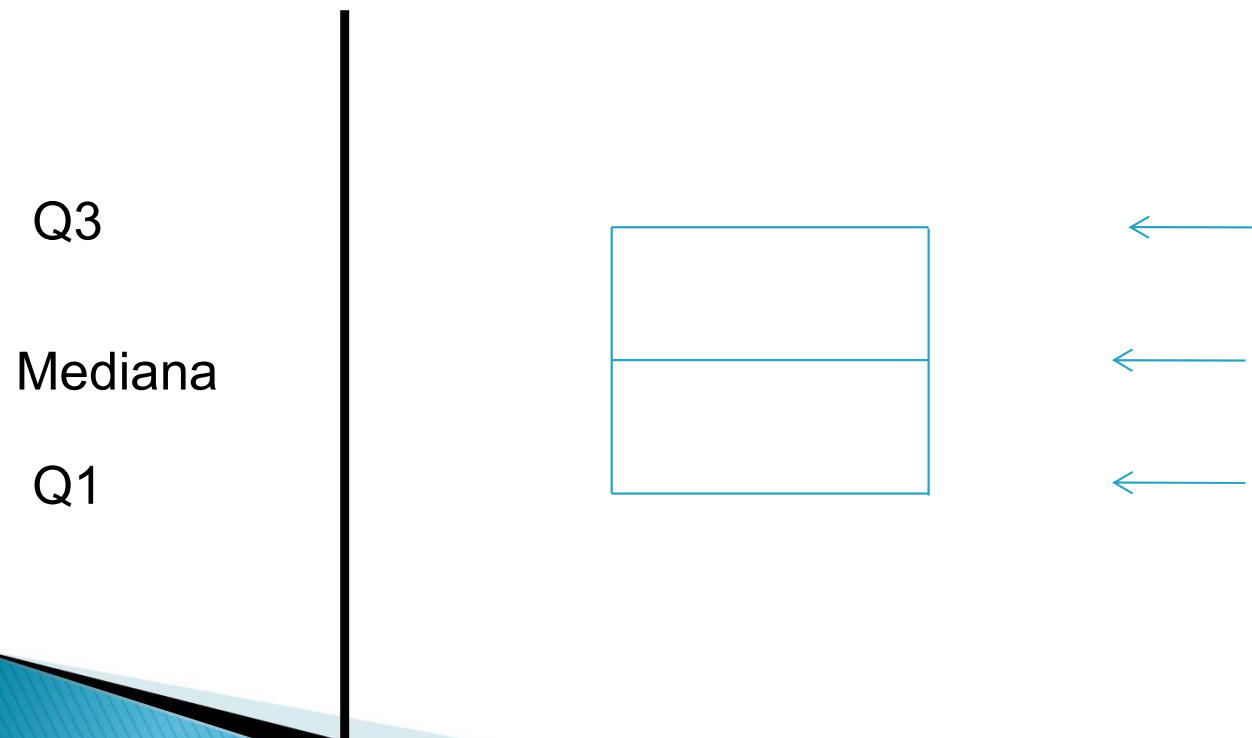
Box-Plot

- ▶ Representação:



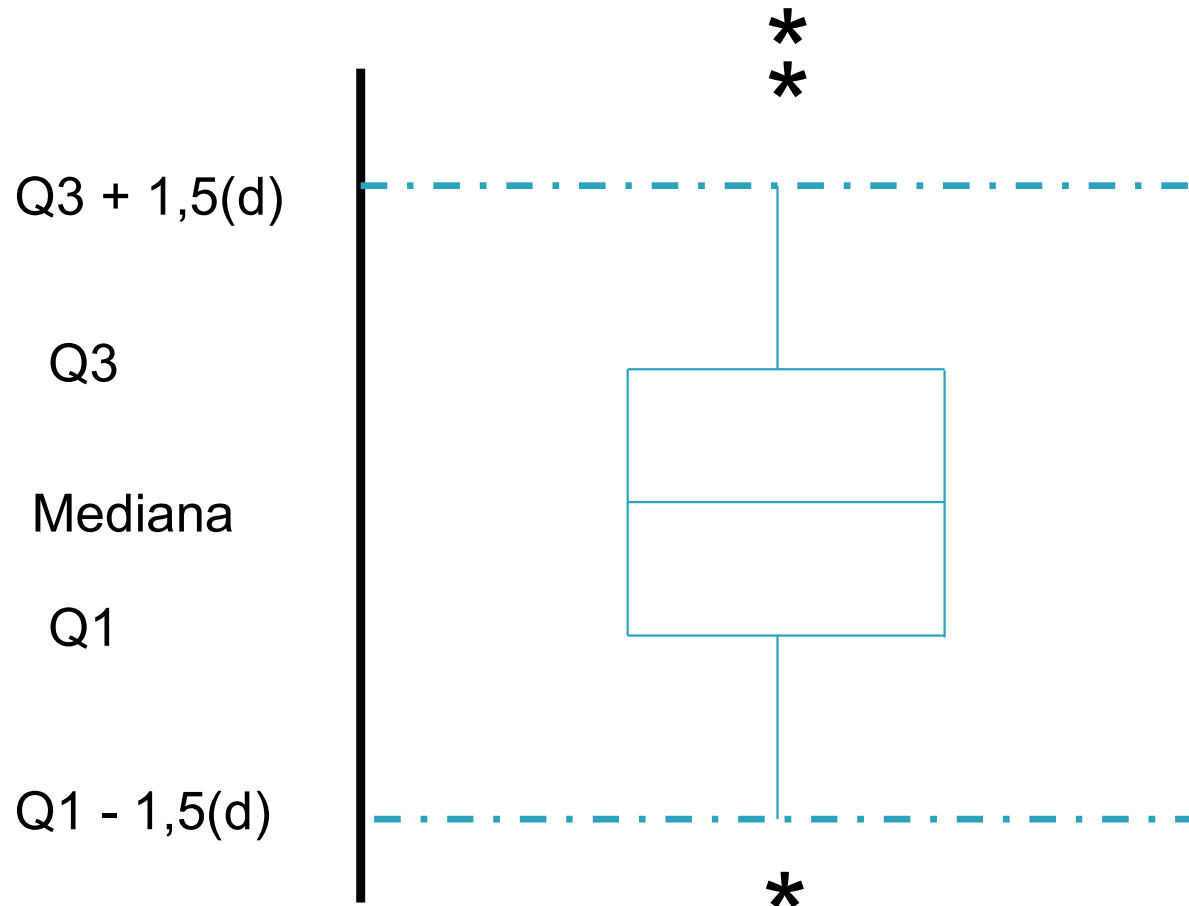
Construindo o Box-Plot

- ▶ 1) Traça-se:
 - a mediana, o primeiro e o terceiro quartil:



▶ 2) Traça-se:

- Haste superior = $Q3 + 1,5(Q3 - Q1)$
- Haste inferior = $Q1 - 1,5(Q3 - Q1)$



Observação: Qualquer valor acima do Limite Superior ou abaixo do Limite Inferior será considerado valor aberrante ou 'outlier' (marcado com *)

Box-Plot

▶ Exemplo: Salário inicial de 15 recém-formados em Estatística (dados fictícios).

- R\$ 1.800,00
- R\$ 1.900,00
- R\$ 5.000,00
- R\$ 1.500,00
- R\$ 1.600,00
- R\$ 1.800,00
- R\$ 2.000,00
- R\$ 2.050,00
- R\$ 1.850,00
- R\$ 1.850,00
- R\$ 1.100,00
- R\$ 2.300,00
- R\$ 2.300,00
- R\$ 2.900,00
- R\$ 2.850,00

Box-Plot

- ▶ Exemplo: Salário inicial de 15 recém-formados em Estatística (dados fictícios).
 - Mediana: R\$ 1.900,00
 - Q1: R\$ 1.800,00
 - Q3: R\$ 2.300,00
 - Q3 - Q1: R\$ 500,00
 - Valor mínimo: R\$ 1.100,00
 - Valor máximo: R\$ 5.000,00

Salário (em R\$)

5,000
4,000
3,000
2,000
1,000

$2300 + 750 = 3050$

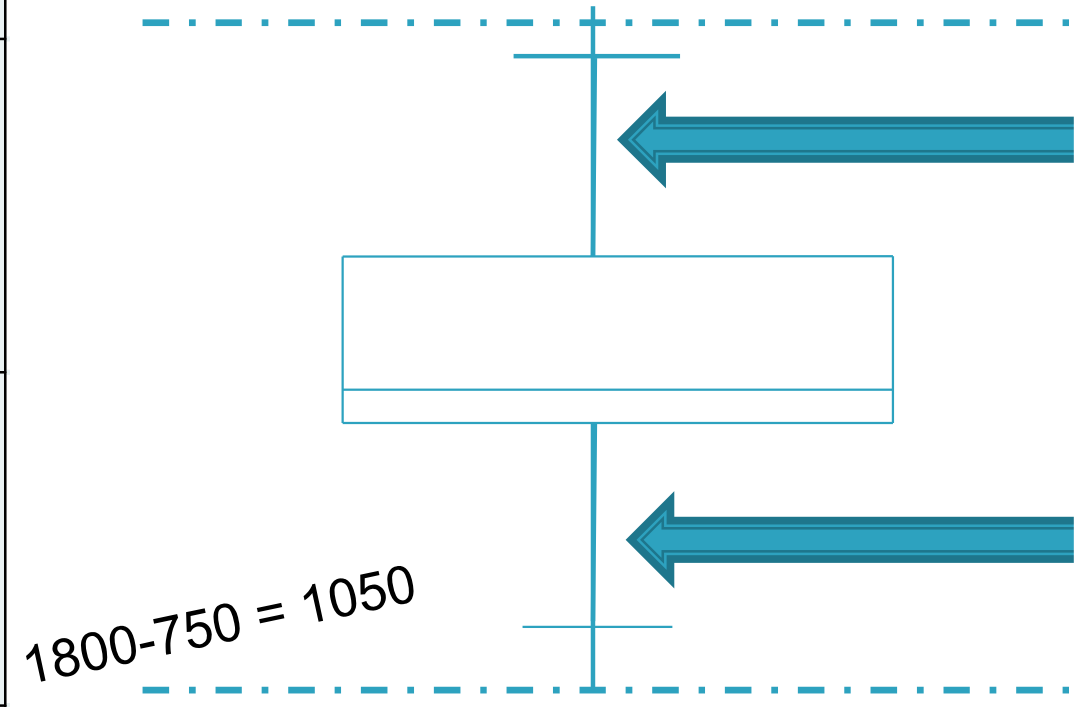
$1800 - 750 = 1050$

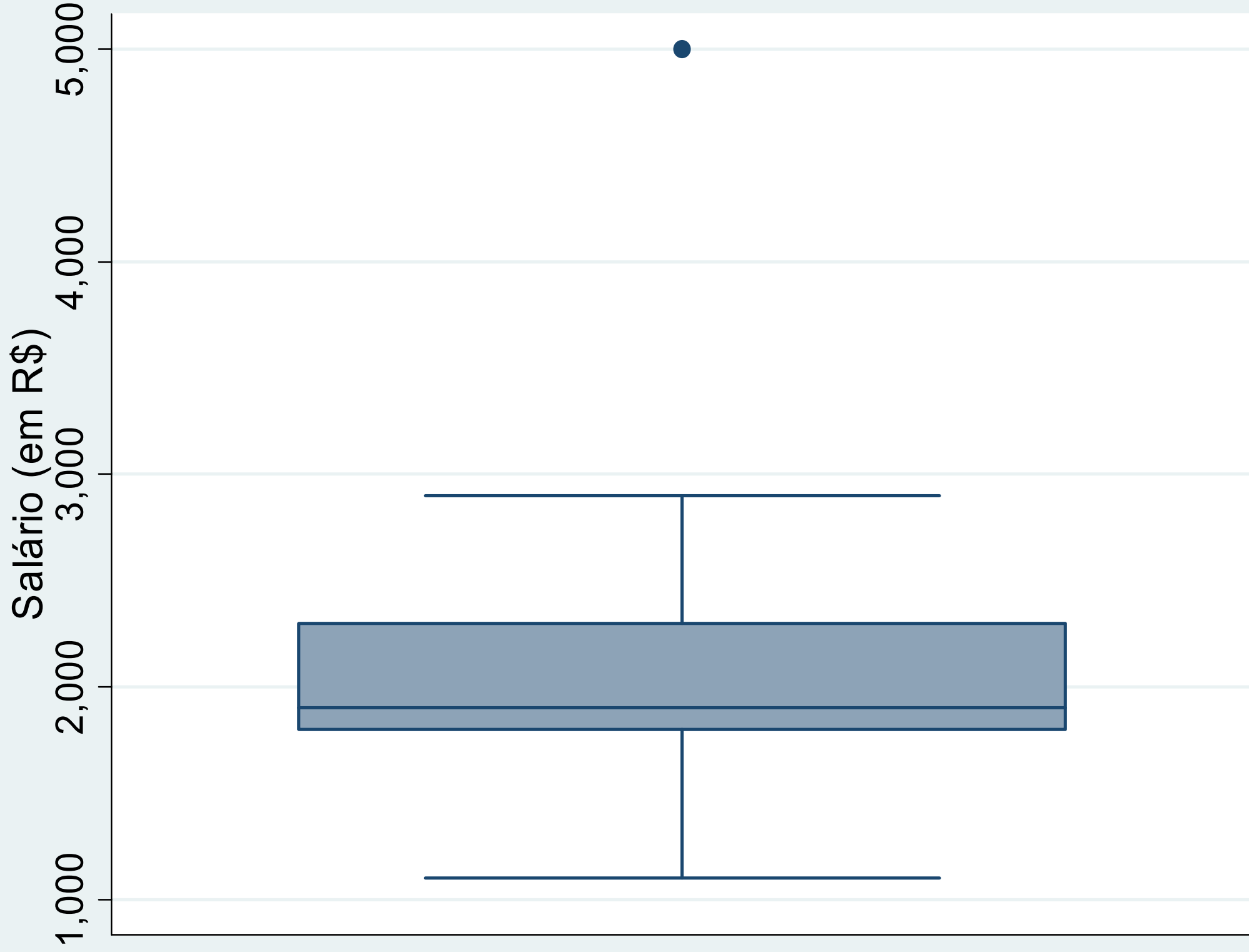
*

Mediana: R\$ 1.900,00
Q1: R\$ 1.800,00
Q3: R\$ 2.300,00
Q3 - Q1: R\$ 500,00
 $1,5 * (500) = \text{R\$ } 750,00$
Valor mínimo: R\$ 1.100,00
Valor máximo: R\$ 5.000,00

Diminuir até o valor imediatamente abaixo do limite (R\$2.900,00)

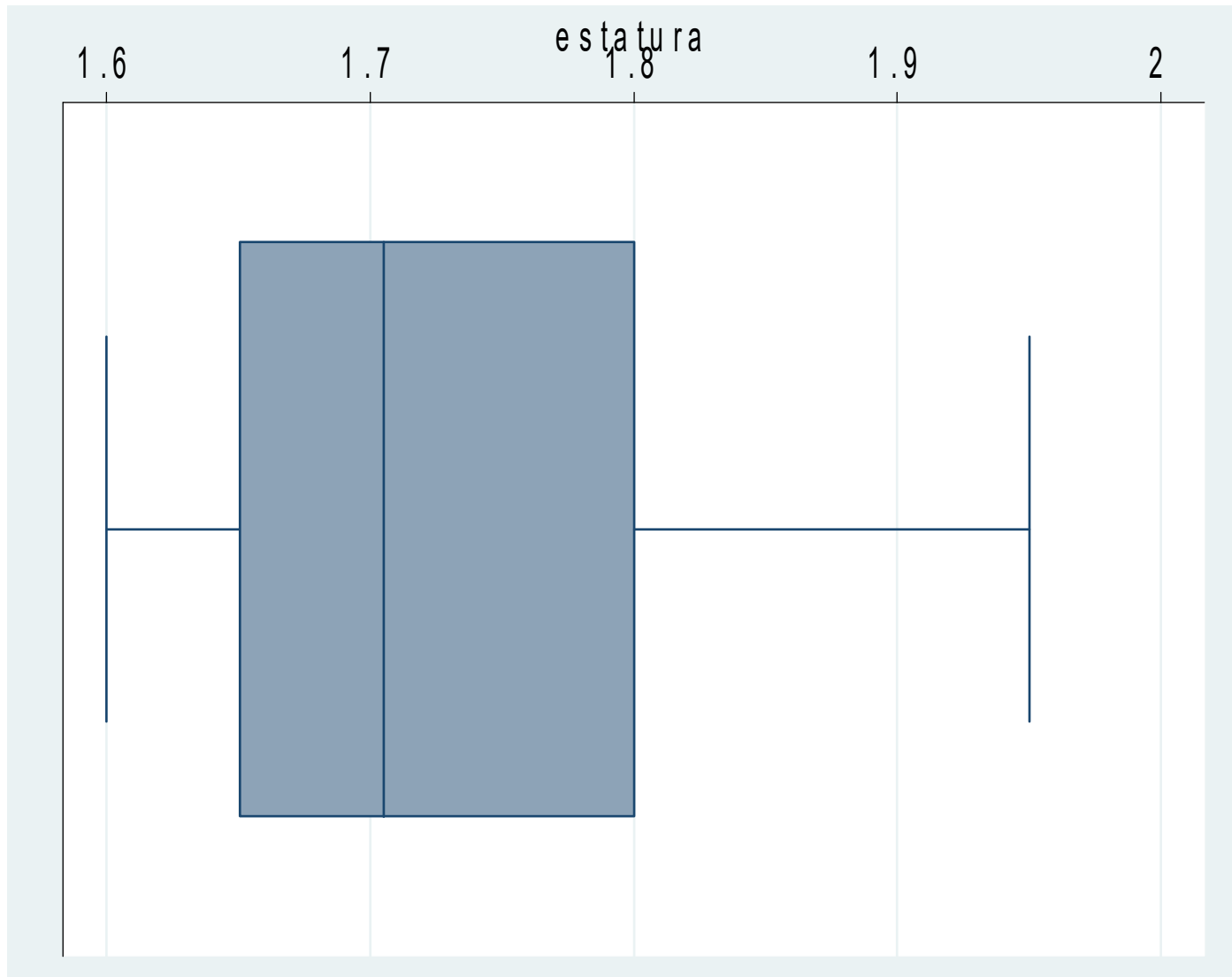
Aumentar até o valor imediatamente abaixo do limite R\$1.100,00





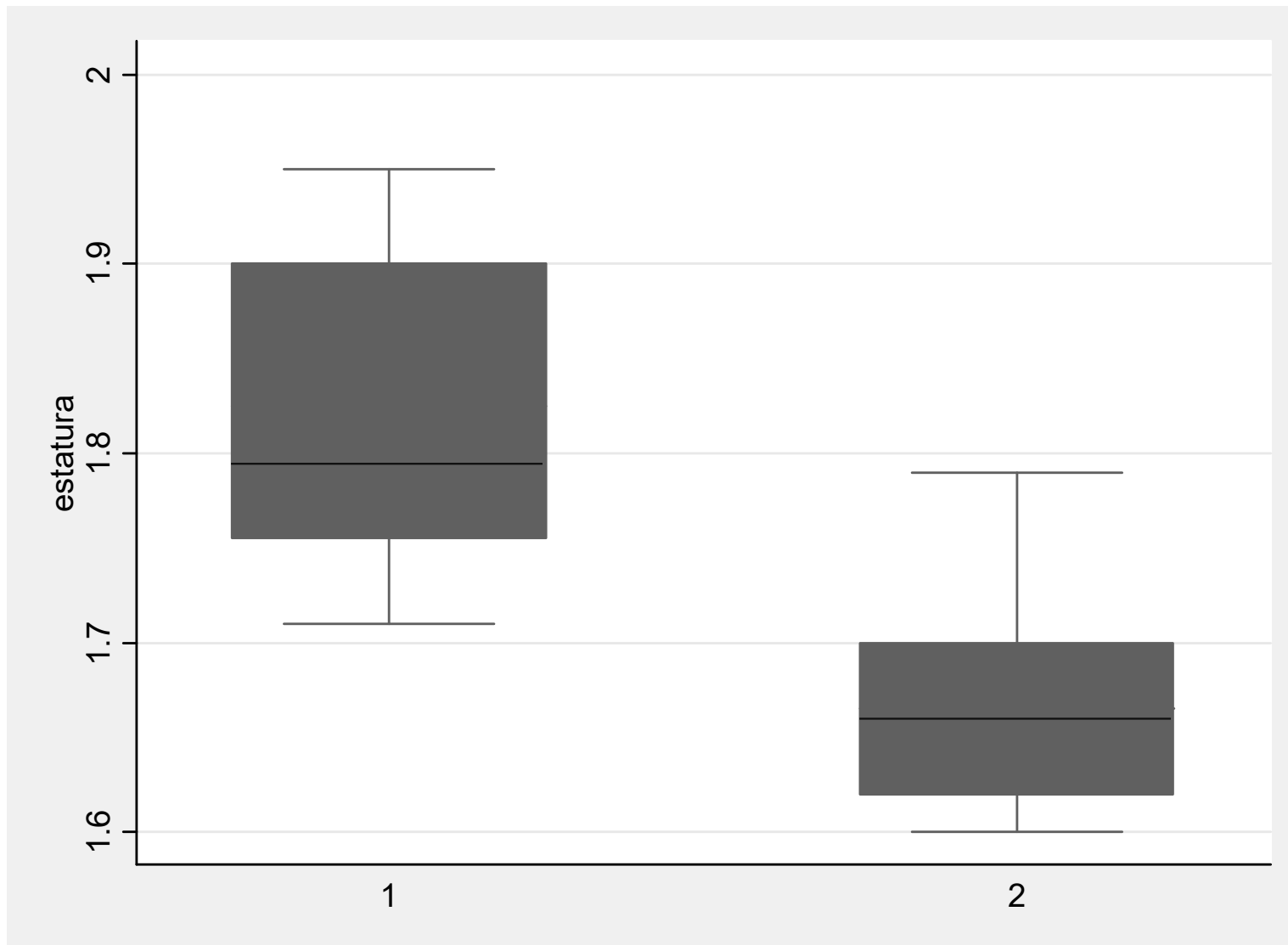
Box-Plot

- ▶ Exemplo: (estatura alunos)

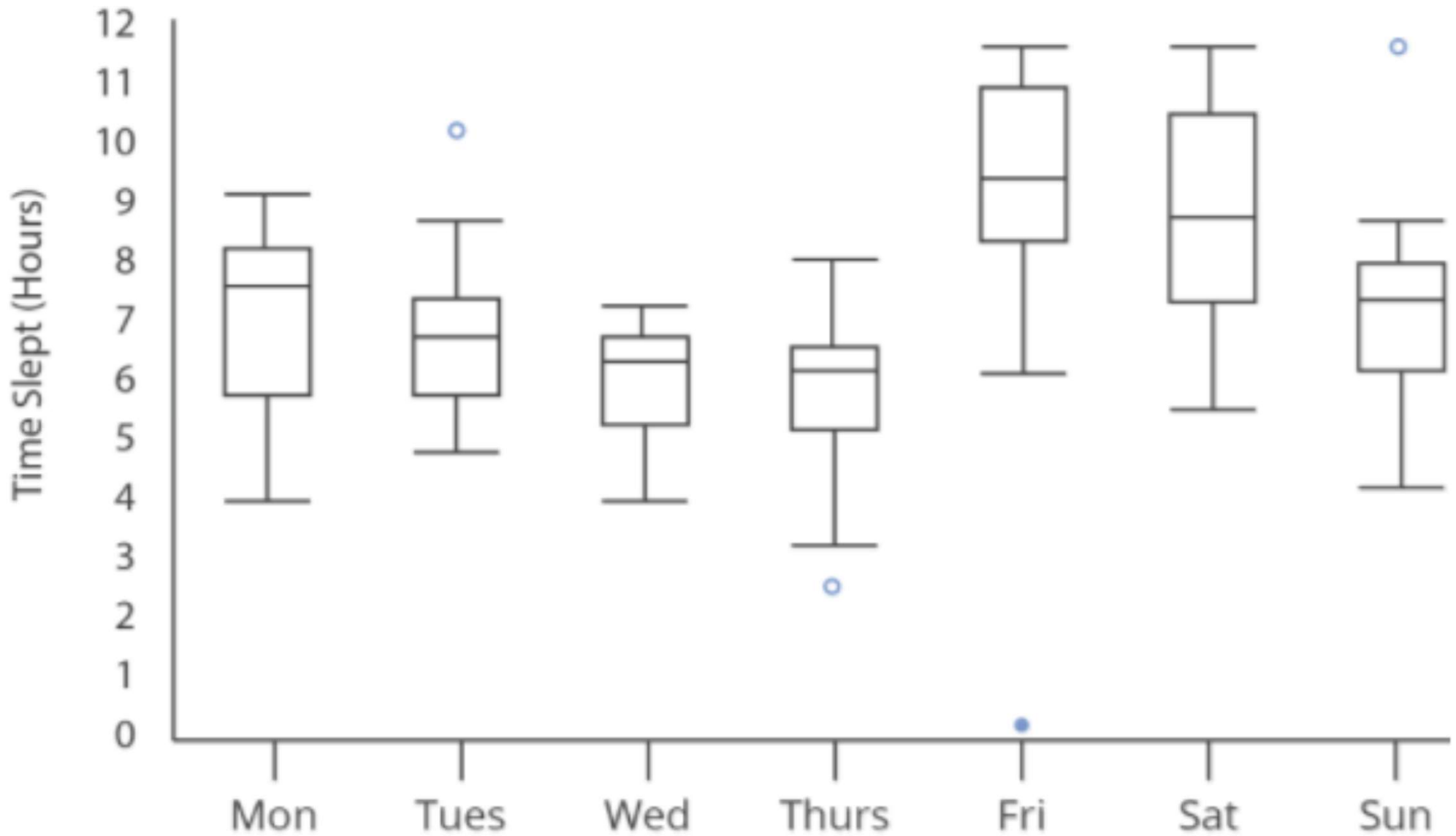


Box-Plot

- ▶ Exemplo: (estatura alunos, segundo sexo)



Box-Plot



Assimetria/Curtose

- ▶ Assimetria e Curtose
 - Complementam as medidas de posição e dispersão;
 - Fornecem uma ideia das distribuições dos dados



Formato da Curva dos dados

Assimetria (skewness)

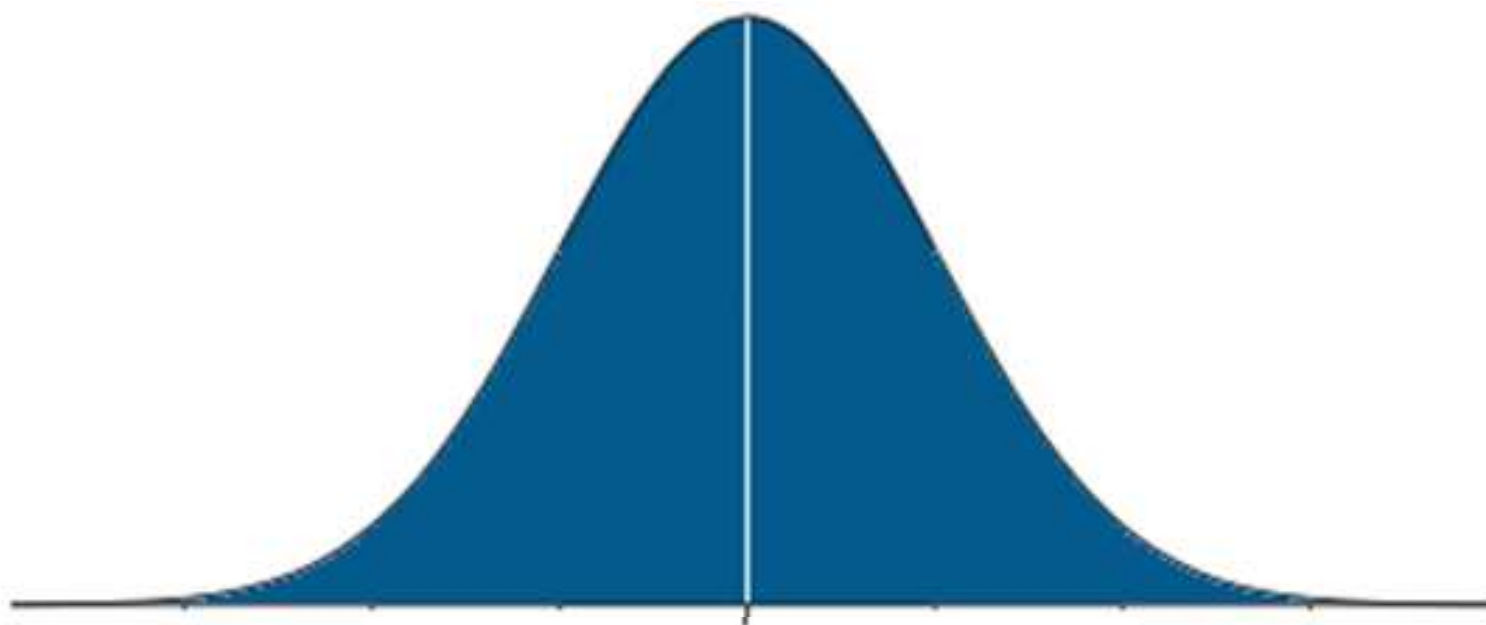
- ▶ Medida que caracteriza como e quanto a distribuição de frequências se afasta da condição de simetria.
 - Simétrica
 - Assimétrica à direita (positiva)
 - Assimétrica à esquerda (negativa)



Simetria / Assimetria

▶ Simetria:

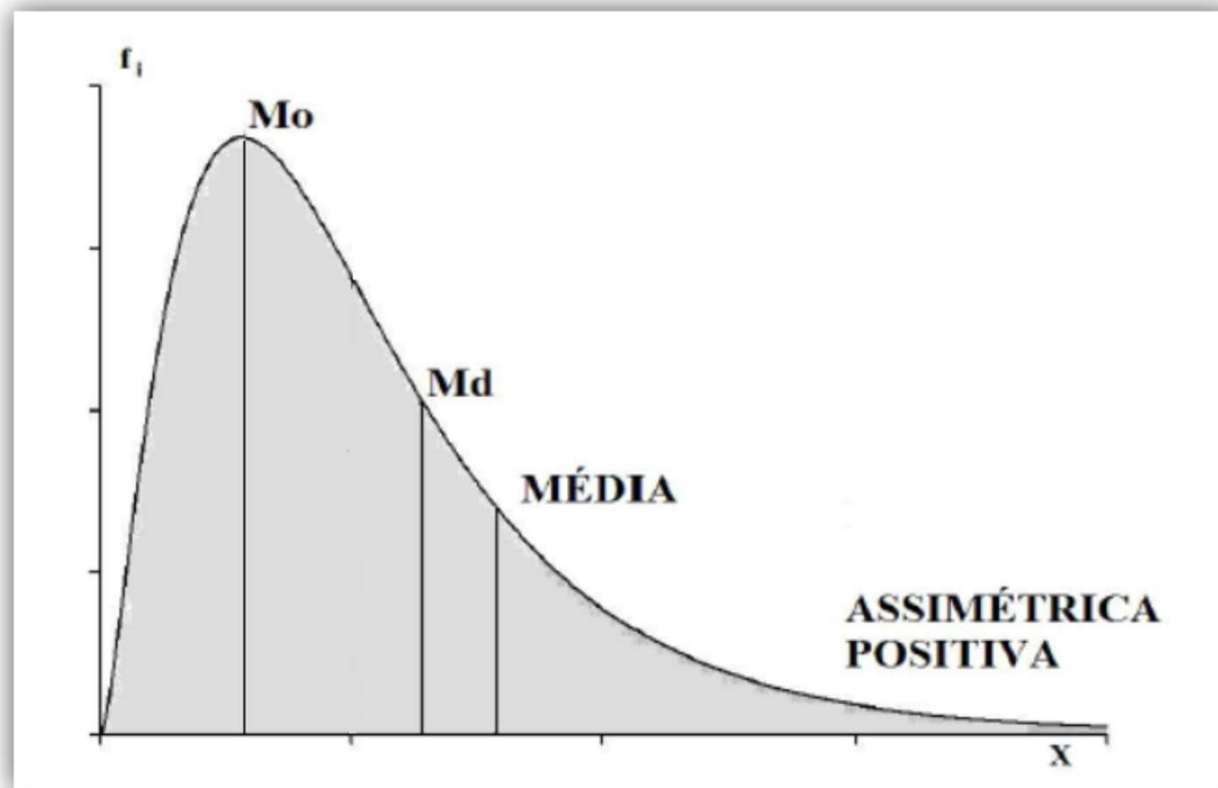
- Uma distribuição é classificada como simétrica se:
 - *Média = Mediana = Moda*



Média = Mediana = Moda

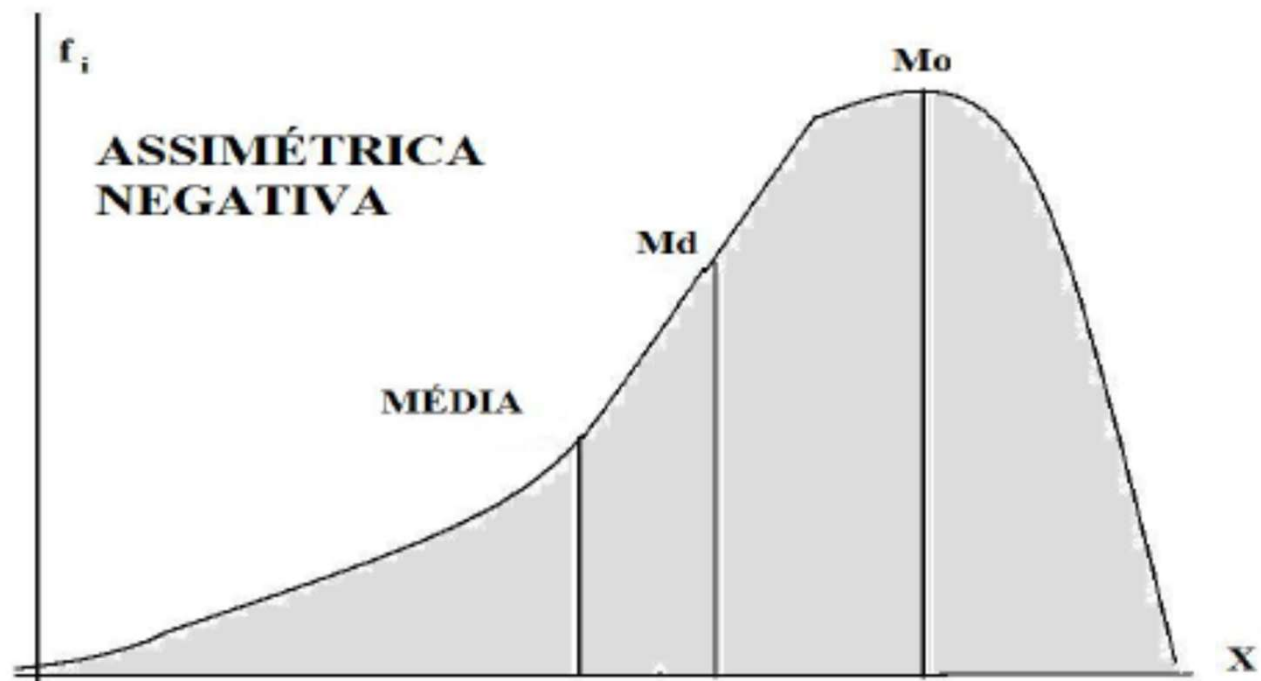
Assimetria

- ▶ Assimétrica direita – positiva
 - Média > Mediana > Moda



Assimetria

- ▶ Assimétrica esquerda – negativa
 - $\text{Moda} > \text{Mediana} > \text{Média}$



Assimetria

- ▶ Medidas que caracterizam como e quanto a distribuição de frequências se afasta da condição de simetria.
- ▶ Há algumas formas para se calcular:
 - Coeficiente de assimetria de Pearson:

$$A = \frac{3(\bar{x} - Med)}{s}$$

\bar{x} = média

Med = mediana

s = desvio padrão

Assimetria

$$A = \frac{3(\bar{x} - Med)}{s}$$

- Coeficiente de assimetria de Pearson:
- Critérios para classificação:
 - $-0,15 < A < 0,15$ → distribuição simétrica
 - $0,15 \leq A \leq 1,0$ → Assimetria moderada (direita)
 - $-1,0 \leq A \leq -0,15$ → Assimetria moderada (esquerda)
 - $A > 1,0$ → Assimetria forte (direita)
 - $A < -1,0$ → Assimetria forte (esquerda)

Assimetria

$$A = \frac{3(\bar{x} - Med)}{s}$$

- ▶ Quantidade de células T CD4/mm³ – 10 pacientes

525, 556, 576, 658, 689, 704, 771, 791, 847, 884

$$\bar{x} = 700,1$$

$$Med = 696,5$$

$$s = 123,57$$

$$A = \frac{3(\bar{x} - Med)}{s} = \frac{3(700,1 - 696,5)}{123,57} = \frac{10,8}{123,57} = 0,087$$

simétrica



Assimetria

- ▶ Gráficos - dados simétricos:

- Box-Plot

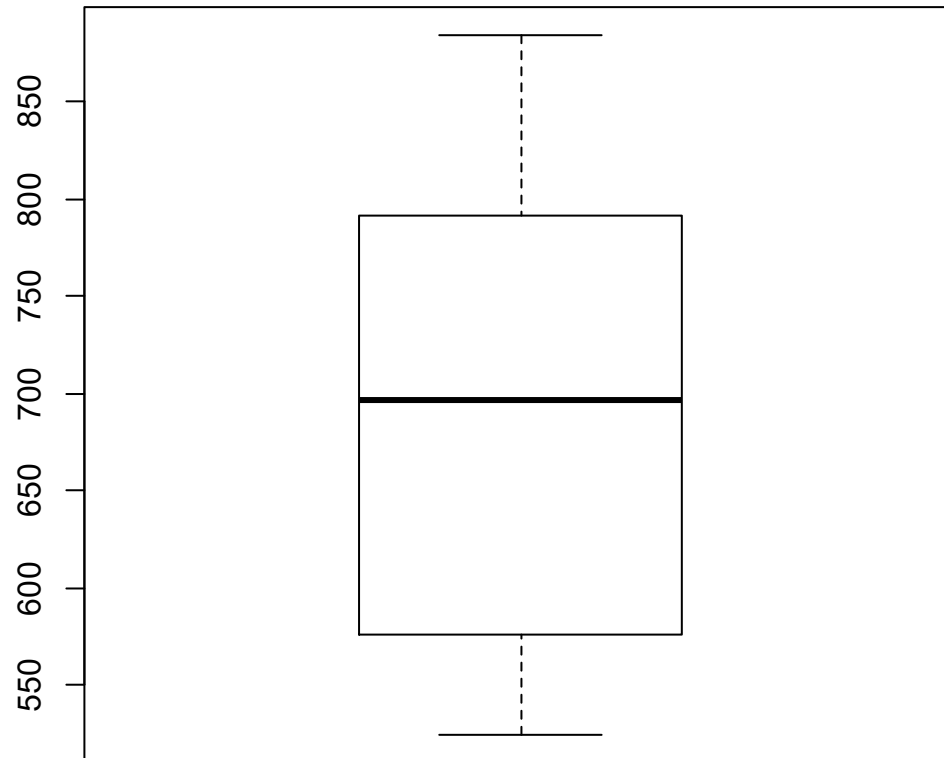
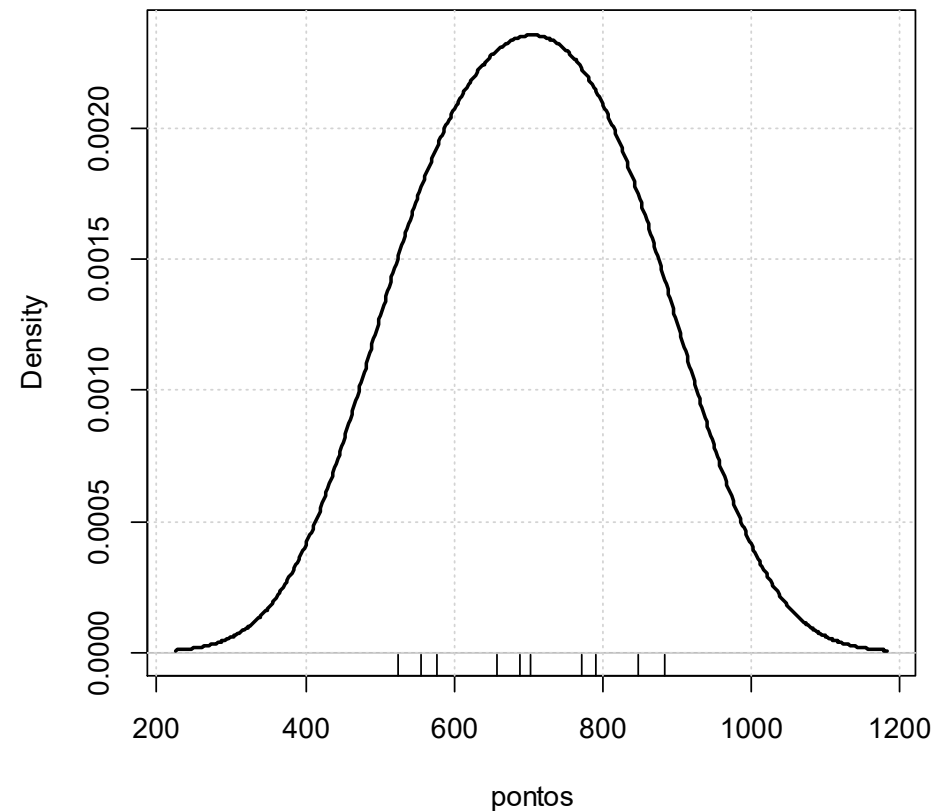


Gráfico de Densidade



Assimetria

▶ Gráficos - dados assimétricos

○ Box-Plot

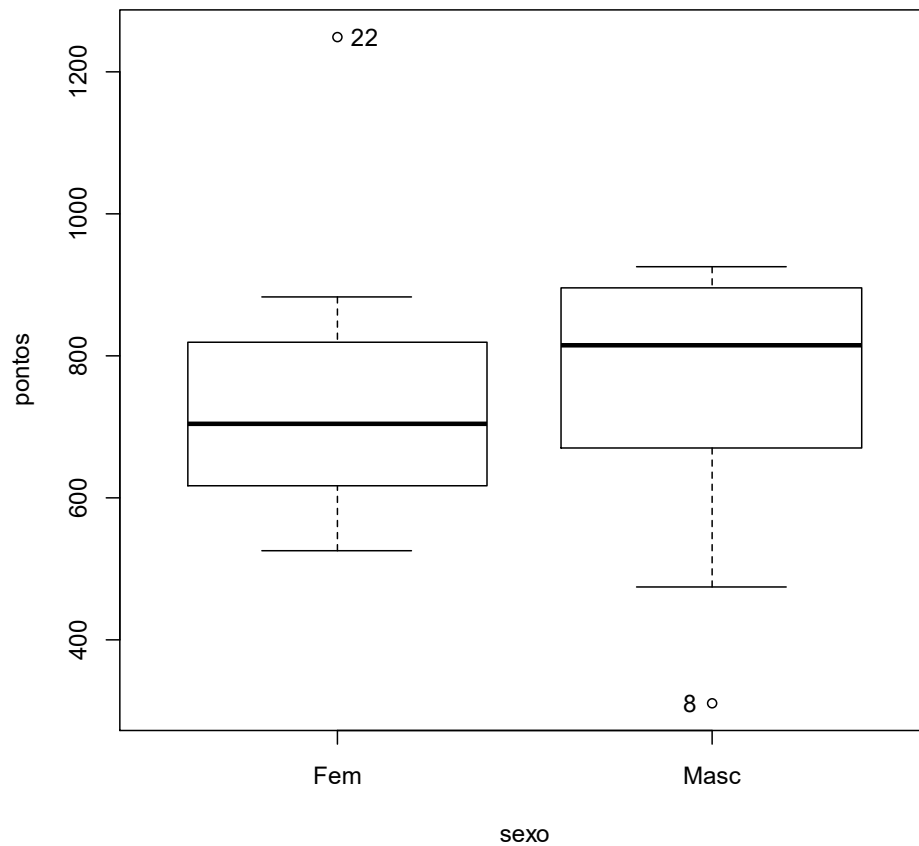
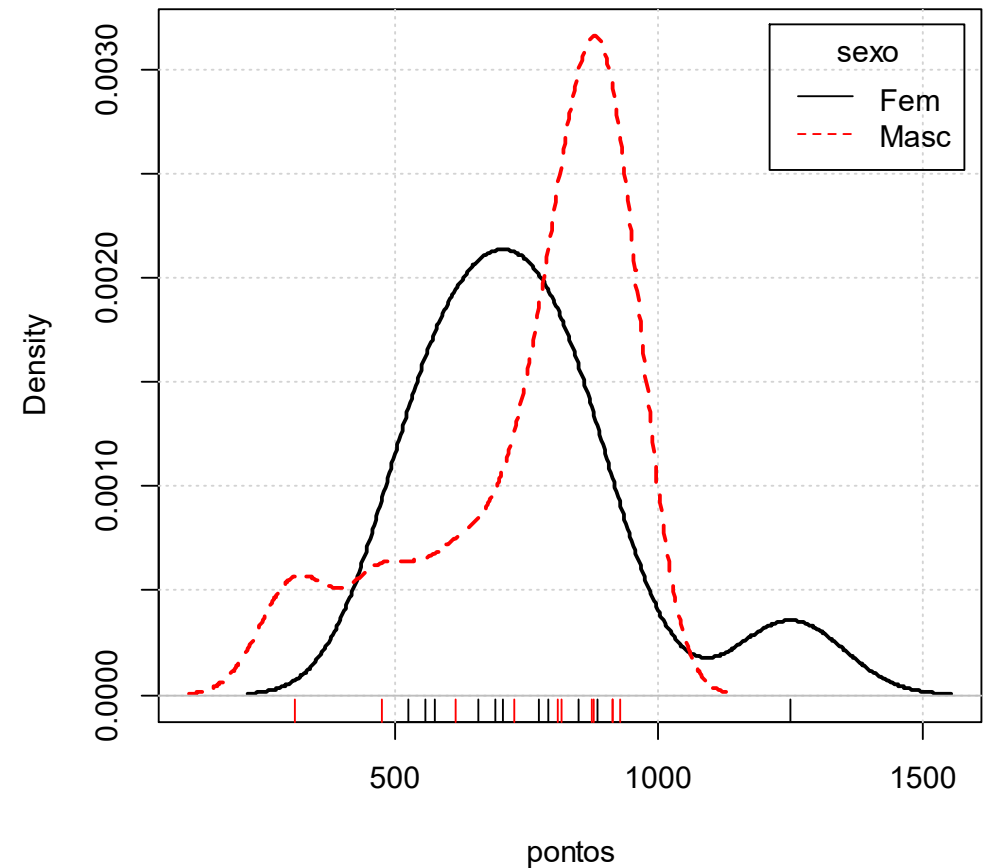


Gráfico de Densidade



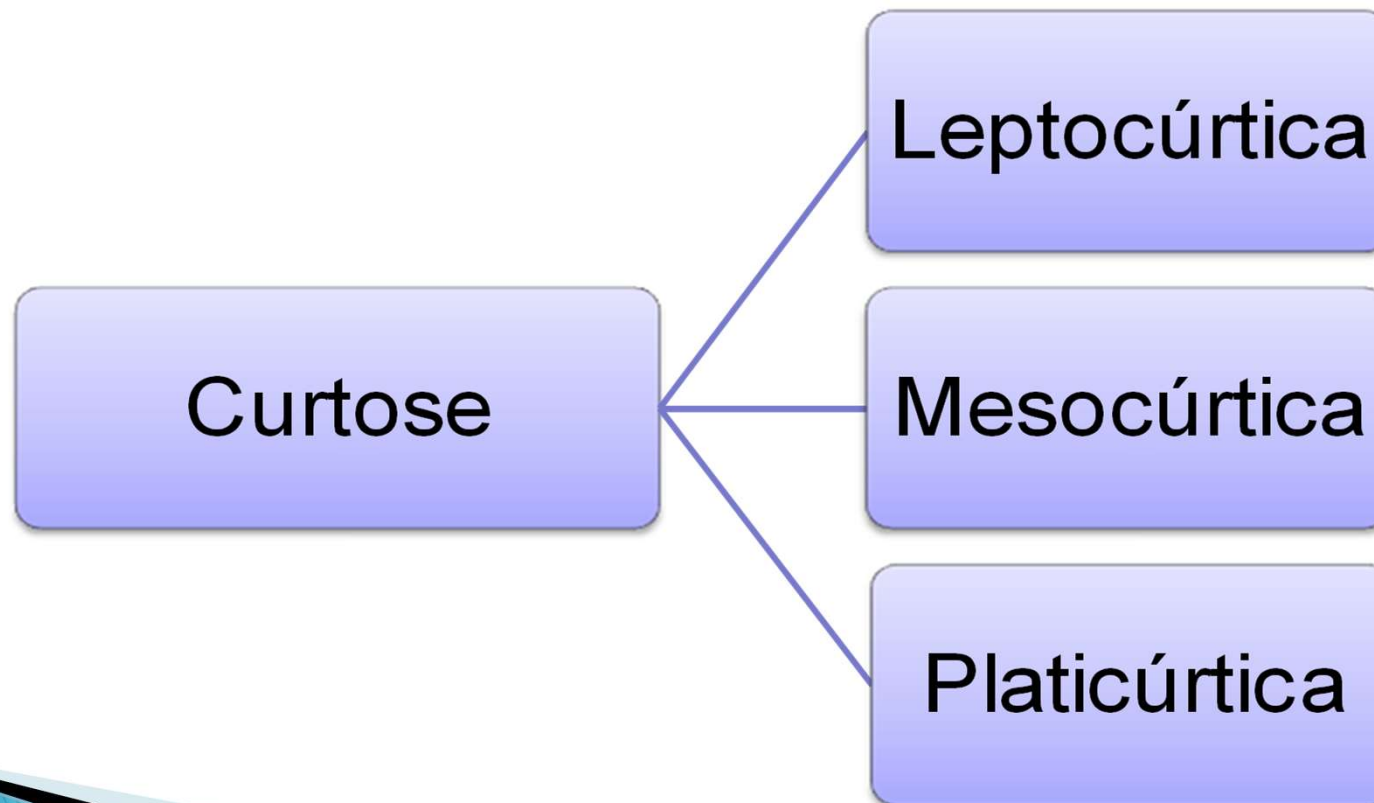
Curtose

- ▶ Também conhecida como sendo o grau de achatamento de uma distribuição.
- ▶ Caracteriza a dispersão em uma distribuição

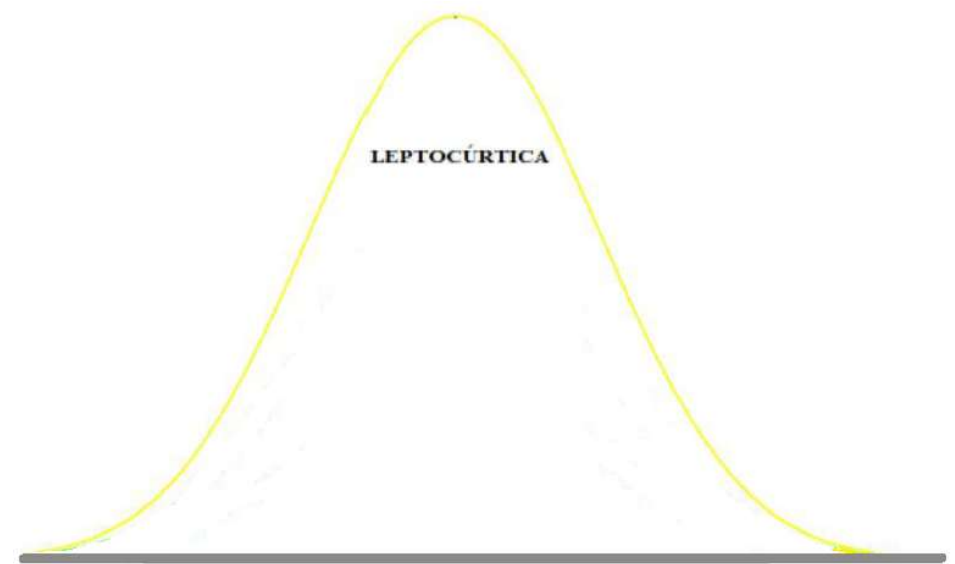
- ▶ Fórmula:
$$k = 0,263 \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$

Curtose

- ▶ Classificação segundo achatamento:



Curtose



▶ Classificação segundo achatamento:

- Leptocúrtica: curva fechada com dados fortemente concentrados ao redor do centro

$$K > 0$$

Curtose

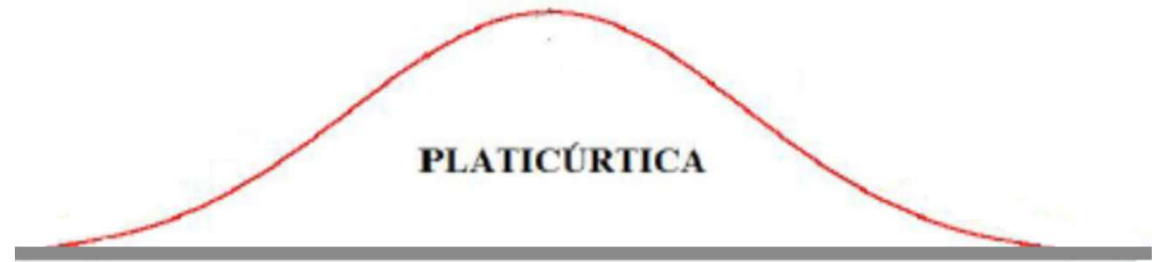


▶ Classificação segundo achatamento:

- Mesocúrtica: curva normal – dados razoavelmente centrados em torno do centro

$$K = 0$$

Curtose

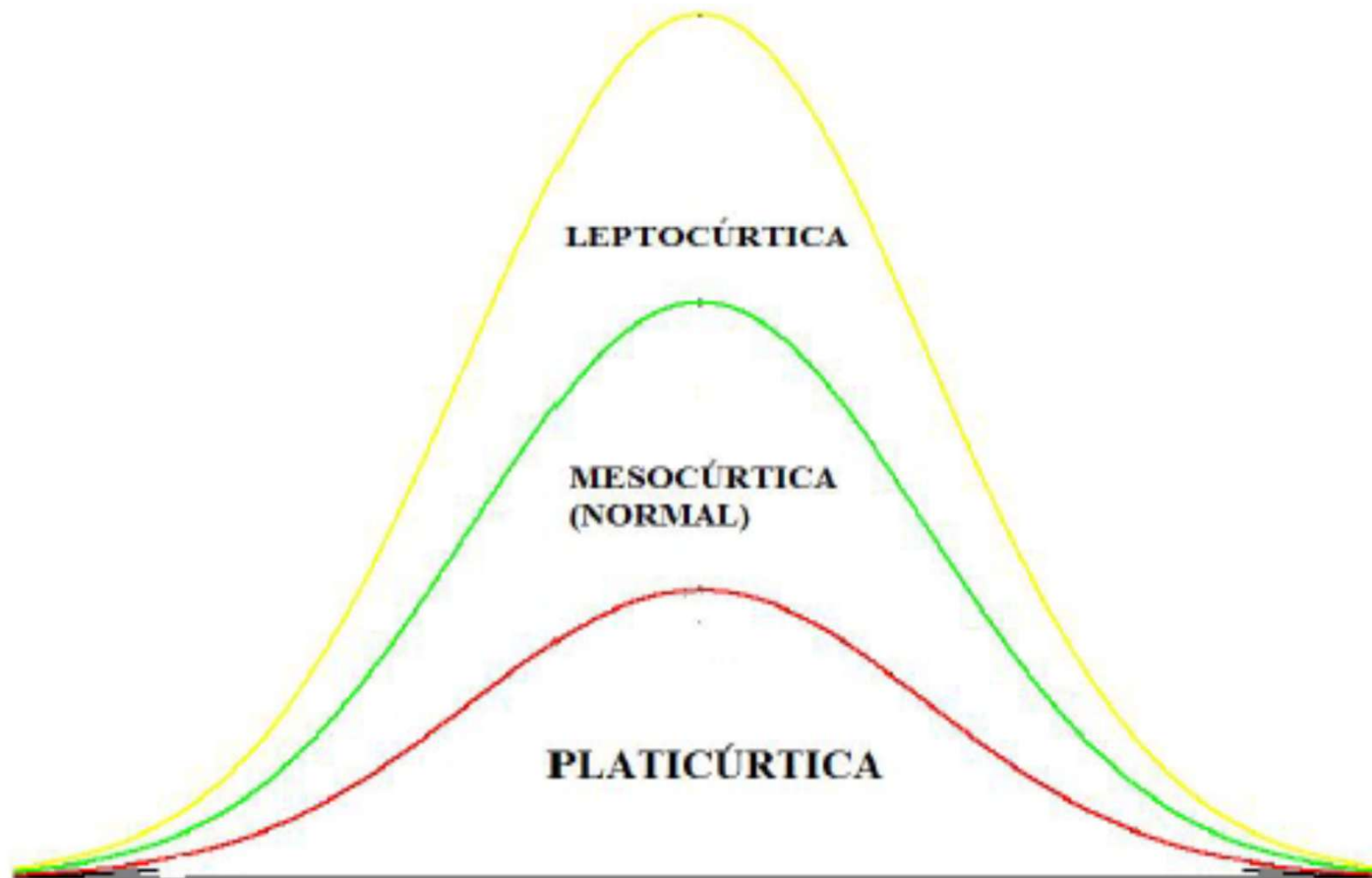


- ▶ Classificação segundo achatamento:
 - Platicúrtica: distribuição com dados pouco concentrados em torno do centro.

$$K < 0$$

Curtose

- ▶ Classificação segundo achatamento:



Curtose

- ▶ Quantidade de células T CD4/mm³ – 10 pacientes

525, 556, 576, 658, 689, 704, 771, 791, 847, 884

$$Q1 = pos_{2,75} = pos_2 + 0,75(pos_3 - pos_2) = 571$$

$$Q3 = pos_{8,25} = pos_8 + 0,25(pos_9 - pos_8) = 805$$

$$P10 = pos_{1,1} = pos_1 + 0,1(pos_2 - pos_1) = 528,1$$

$$P90 = pos_{9,9} = pos_9 + 0,9(pos_{10} - pos_9) = 880,3$$

$$k = 0,263 - \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} = 0,263 - \frac{805 - 571}{2(880,3 - 528,1)} = 0,263 - \frac{234}{704,4}$$

$$k = 0,263 - 0,332 = -0,069 \Rightarrow \textit{Platicúrtica}$$

Bibliografia

- ▶ Arango HG. **Bioestatística Teórica e Computacional**. Guanabara Koogan. 2^a ed. Rio de Janeiro, 2005.
- ▶ Bergamaschi DP. **Bioestatística (Apostila Graduação) FSP/USP**, 2010.
- ▶ Callegari-Jacques SM. **Bioestatística – Princípios e Aplicações**. Artmed. Porto Alegre, 2003.
- ▶ ENCE – Escola Nacional de Ciências Estatísticas
www.ence.ibge.gov.br
- ▶ Latorre MRDO. **Bioestatística (Apostila graduação) FSP/USP**, 2009.
- ▶ Magalhães MN; Lima ACP. **Noções de Probabilidade e Estatística**. EDUSP. São Paulo, 2002.
- ▶ Vieira S. **Introdução à Bioestatística**. ELSEVIER. 4^a ed. Rio de Janeiro, 2010.