Engenharias

M109
Estatística

Aula 4 – Medidas

Profª. Rosimara Beatriz Arci Salgado

Créditos:
Profª. Karina Perez Mokarzel Carneiro
Prof. Rausley Adriano Amaral de Souza

Inatel | CAMINHOS QUE CONECTAM COM O FUTURO

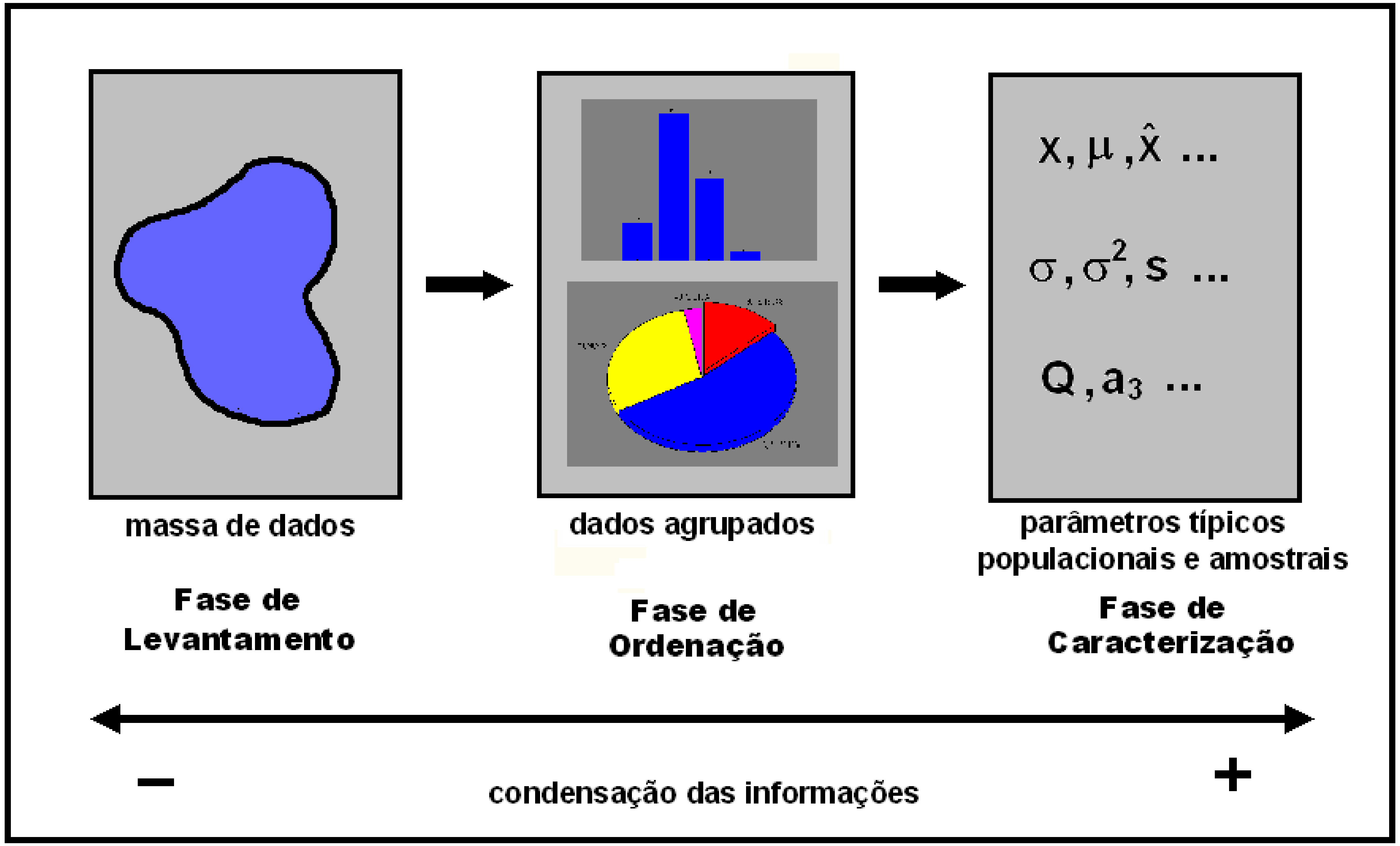
Bibliografia básica:

Arango HG. Bioestatística: teórica e computacional. 3ªed. Rio de Janeiro: Guanabara Koogan; 2011.

SPIEGEL, Murray Ralph; FARIA, Alfredo Alves De Probabilidade e estatística. São Paulo, SP: McGraw-Hill, 1978.

MEDIDAS CARACTERÍSTICAS DE UMA DISTRIBUIÇÃO

Estatística descritiva



MEDIDAS CARACTERÍSTICAS DE UMA DISTRIBUIÇÃO

- **Medidas de Tendência Central:**

  Média, Mediana e Moda.
- **Medidas de Dispersão ou Variabilidade:**

  Amplitude Total, Variância e Desvio Padrão.
- **Medidas de Assimetria**

MEDIDAS DE TENDÊNCIA CENTRAL

- Valor que seja o mais parecido possível com os demais valores do conjunto.
- Valor que tende ao centro.
- É uma primeira caracterização dos conjuntos populacionais ou amostrais.

MEDIDAS DE TENDÊNCIA CENTRAL

MTCs	Médias ("*mean*") - Média aritmética simples, MAS - Média geométrica, G - Média harmônica, H	Valores
	Mediana, ("*median*")	Ordem
	Moda, ("*mode*")	Freqüência

Média aritmética (*sample mean,* média amostral)

Dado o conjunto de n valores da variável X,

$$X = \{x_1, x_2, \ldots, x_n\}$$

a média aritmética simples desse conjunto pode ser obtida a partir da expressão:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Mediana

- Estando os valores que compõem o conjunto de observações ordenado de forma crescente ou decrescente (rol), o valor que ocupa a posição equidistante dos extremos é o valor representativo do conjunto.

- Equivale a pegar o meio da fila.

- Os valores extremos não afetam o resultado final.

$$\hat{x} = x_{\frac{n+1}{2}} \rightarrow \text{Se n é impar}$$

$$\hat{x} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} \rightarrow \text{Se n é par}$$

Moda

- Outro critério para a escolha do valor típico de um atributo de uma população ou amostra dela é tomar o valor mais frequente deste conjunto.

- Em outras palavras: o valor mais representativo é aquele que aparece o maior número de vezes.

Exercício 1

Seja X o conjunto dos perímetros cefálicos, em centímetros, de 5 recém-nascidos.

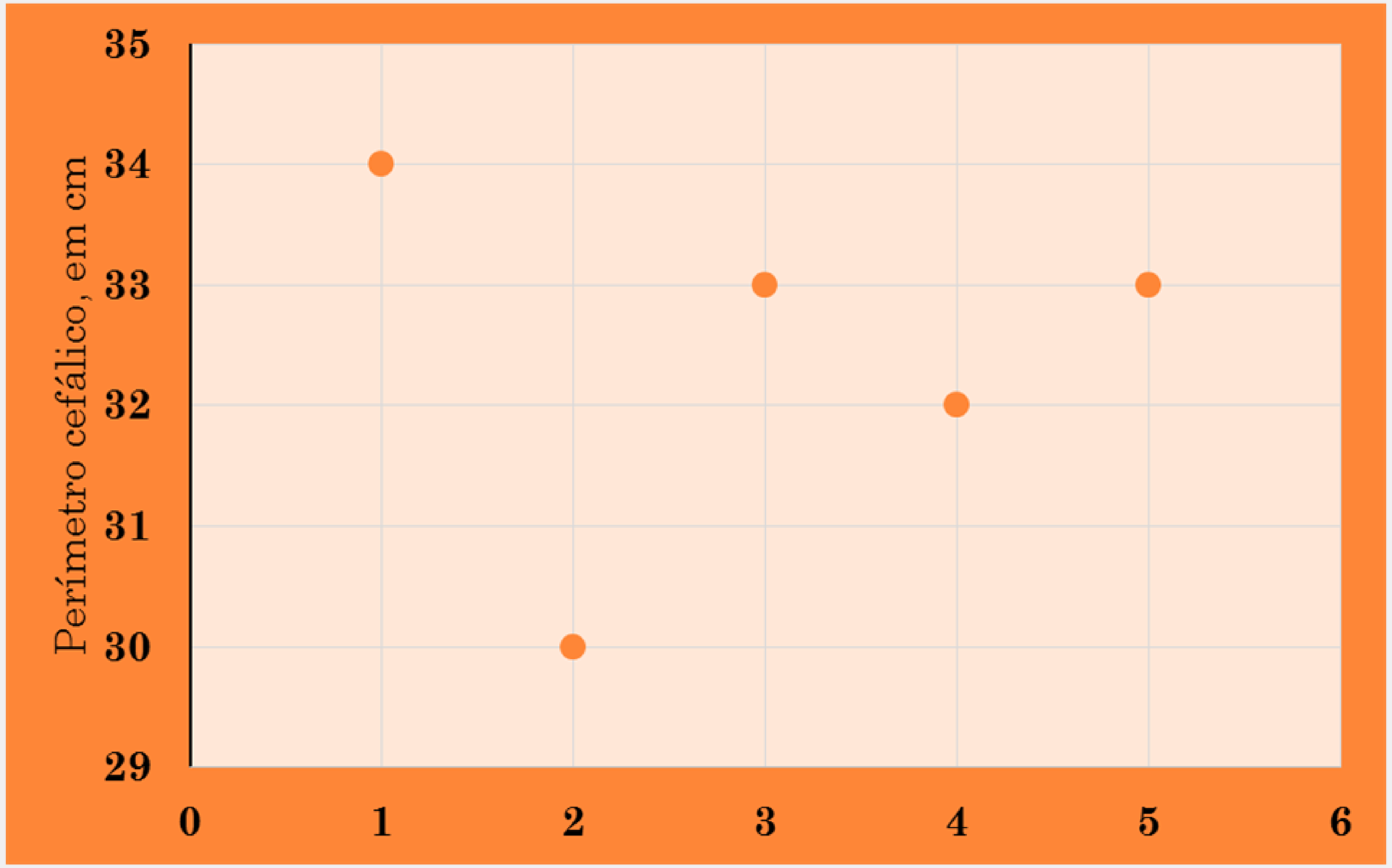
X = { 34, 30, 33, 32, 33}

Então, o perímetro cefálico médio dos cinco RN's resulta:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{34 + 30 + 33 + 32 + 33}{5} = 32{,}4\text{cm}$$

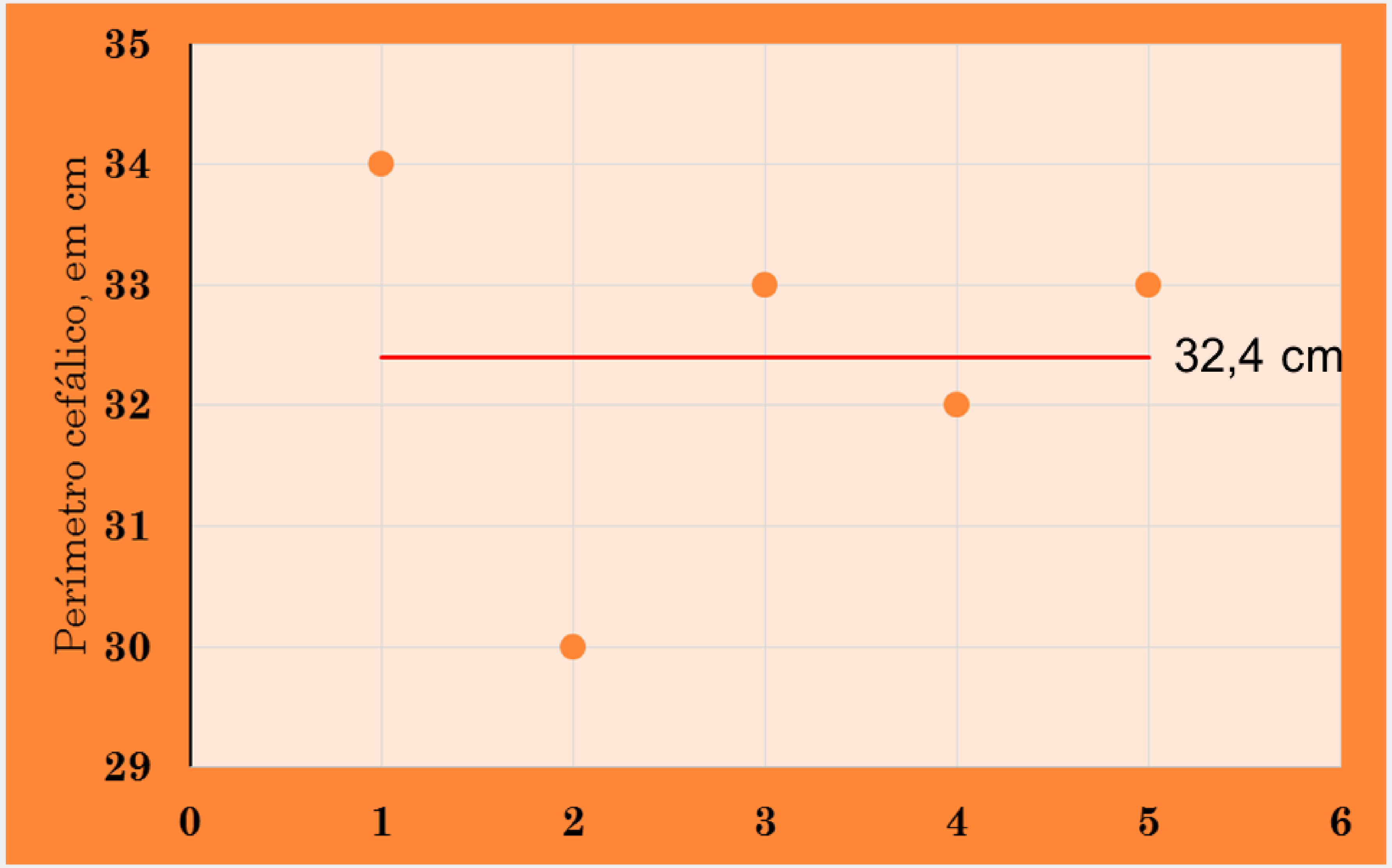
Exercício 1

Gráfico de dispersão dos perímetros cefálicos, em cm.



Exercício 1

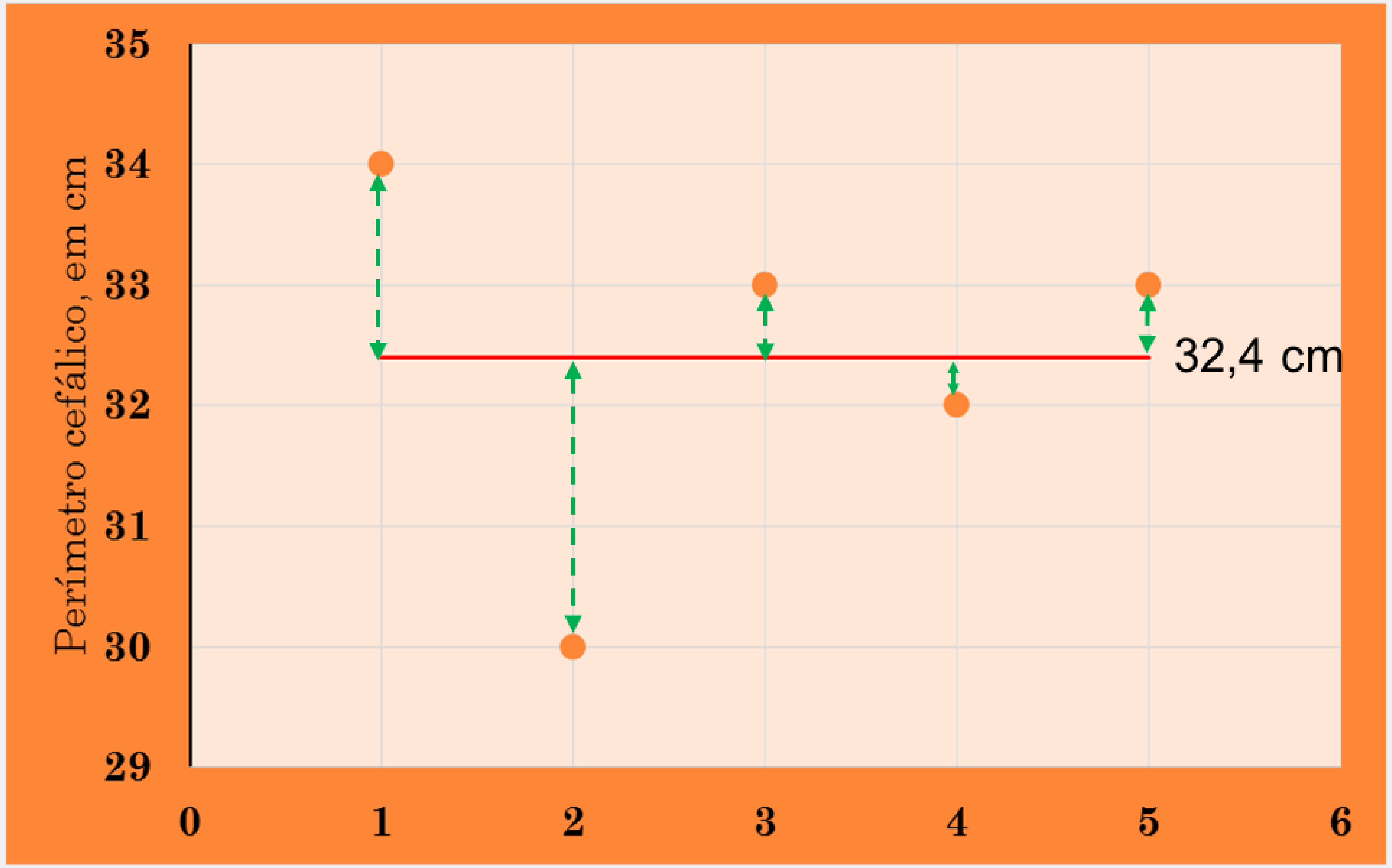
Gráfico de dispersão dos perímetros cefálicos, em cm.



Exercício 1

OBS: a soma dos desvios em relação a Média é zero

d1 =	1.6
d2 =	-2.4
d3 =	0.6
d4 =	-0.4
d5 =	0.6
S =	0.0



Exercício 2

Considere os dados, referentes às taxas de potássio de 12 pacientes:

Paciente	1	2	3	4	5	6	7	8	9	10	11	12
Potássio (mg/100ml)	**4,8**	**4,9**	**4,9**	**4,9**	**5,0**	**5,1**	**5,4**	**6,5**	**6,7**	**6,7**	**6,7**	**7,0**

Calcule as medidas de tendência central

Média = 5,72
Mediana = 5,25
Moda = 4,9 e 6,7 (Bimodal)

Exercício 2

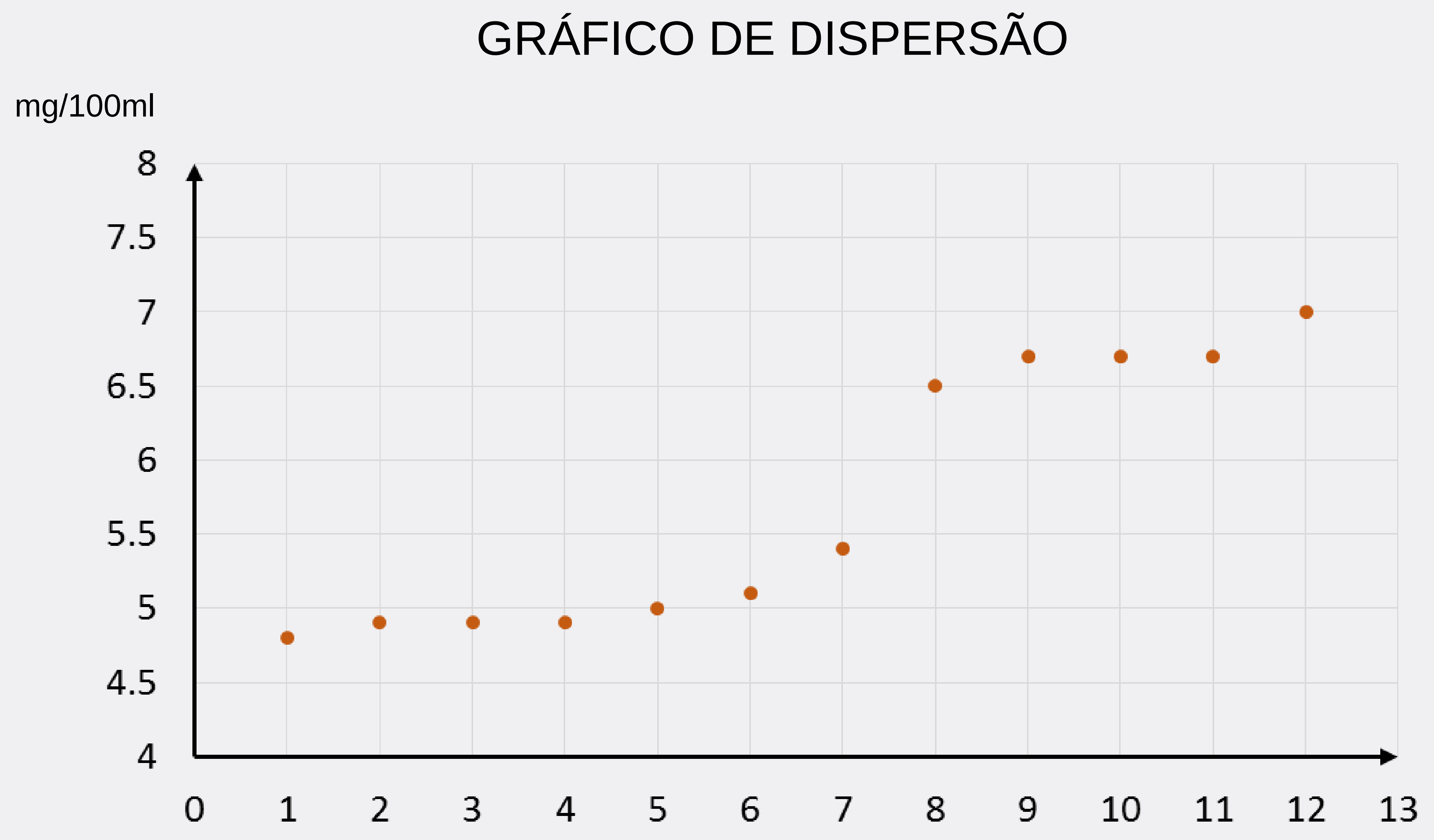


Figura 1 – Taxa de potássio em mg/100ml

Exercício 2

GRÁFICO DE DISPERSÃO

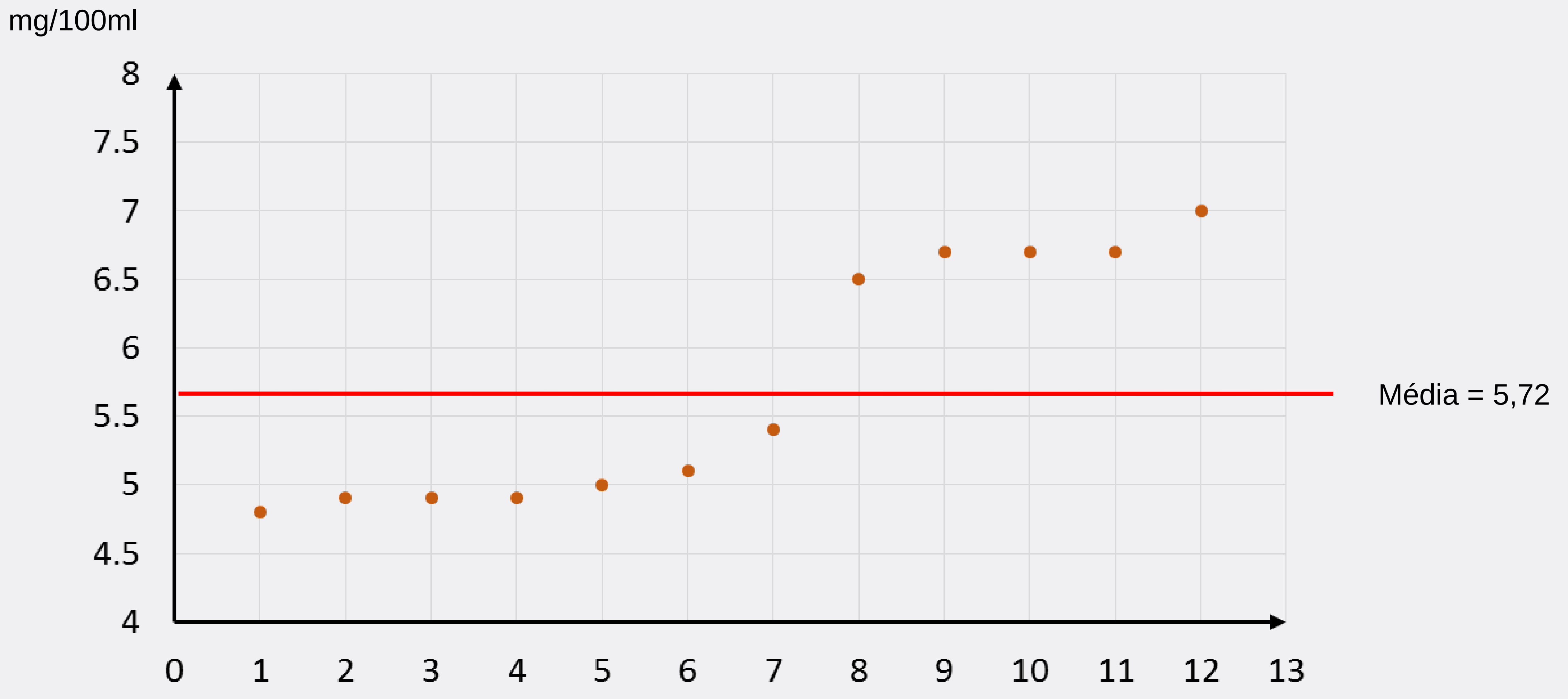


Figura 1 – Taxa de potássio em mg/100ml

Inatel

CAMINHOS QUE CONECTAM COM O FUTURO

Exercício 3

CALCULE AS MEDIDAS DE TENDÊNCIA CENTRAL UTILIZANDO TABELA DE FREQUÊNCIAS

Tabela 2: Idade de 37 gestantes do programa de saúde da família, Itajubá, 2015

Idade [anos]	fa	fA	fr (%)	fR %
17	4	4	11	11
18	10	14	27	38
19	12	26	32	70
20	5	31	14	84
21	3	34	8	92
22	3	37	8	100
Total	**37**	**-**	**100**	**-**

(dados fictícios)

Exercício 3

a) Média

Tabela 2: Idade de 37 gestantes do programa de saúde da família, Itajubá, 2015

Idade [anos]	fa	Idade x fa
17	4	68
18	10	
19	12	
20	5	
21	3	
22	3	
Total	37	

Exercício 3

a) Média

Tabela 2: Idade de 37 gestantes do programa de saúde da família, Itajubá, 2015

Idade [anos]	fa	Idade x fa
17	4	68
18	10	180
19	12	228
20	5	100
21	3	63
22	3	66
Total	37	705

Exercício 3

a) Média

Tabela 2: Idade de 37 gestantes do programa de saúde da família, Itajubá, 2015

Idade [anos]	fa	Idade x fa
17	4	68
18	10	180
19	12	228
20	5	100
21	3	63
22	3	66
Total	37	705

média → $\bar{x} = \dfrac{705}{37} = 19{,}05 \text{ anos}$

Exercício 3

b) Mediana

Tabela 2: Idade de 37 gestantes do programa de saúde da família, Itajubá, 2015

Idade [anos]	fa	fA	fr %	fR %
17	4	4	11	11
18	10	14	27	38
19	12	26	32	70
20	5	31	14	84
21	3	34	8	92
22	3	37	8	100
Total	37	-	100	-

n é 37 (ímpar)

Exercício 3

mediana → $\hat{x} = x_{\frac{37+1}{2}} = x_{19}$

Tabela 2: Idade de 37 gestantes do programa de saúde da família, Itajubá, 2015

Idade [anos]	fa	fA	fr %	fR %
17	4	4	11	11
18	10	14	27	38
19	12	26	32	70
20	5	31	14	84
21	3	34	8	92
22	3	37	8	100
Total	37	-	100	-

Exercício 3

Tabela 2: Idade de 37 gestantes do programa de saúde da família, Itajubá, 2015

Idade [anos]	fa	fA	fr %	fR %
17	4	4	11	11
18	10	14	27	38
19	12	26	32	70
20	5	31	14	84
21	3	34	8	92
22	3	37	8	100
Total	37	-	100	-

mediana → $\hat{x} = x_{\frac{37+1}{2}} = x_{19} = 19 \text{ anos}$

Exercício 3

c) Moda

Tabela 2: Idade de 37 gestantes do programa de saúde da família, Itajubá, 2015

Idade [anos]	fa	fA	fr %	fR %
17	4	4	11	11
18	10	14	27	38
19	12	26	32	70
20	5	31	14	84
21	3	34	8	92
22	3	37	8	100
Total	37	-	100	-

moda → $\tilde{x} = 19 \text{ anos}$

É HORA DE PRATICAR!

SEPARATRIZ

A fórmula geral para determinar a ordem do elemento que corresponde a uma determinada separatriz é:

$$R_i = (n - 1).\frac{i}{C} + 1$$

R_i → ordem do número que representa a i-ésima separatriz

i → separatriz desejada.

C → número de divisões do conjunto (2, 4, 10 ou 100).

n → número de indivíduos.

SEPARATRIZ

A fórmula geral para determinar a ordem do elemento que corresponde a uma determinada separatriz é:

$$R_i = (n-1).\frac{i}{C} + 1$$

Mediana

Quartil

Decil

Percentil

R_i → ordem do número que repres ta a esima ser atriz

i → separatriz desejada.

C → número de divisões do conjunto (2, 4, 10 ou 100).

n → número de indivíduos.

QUARTIL

Valores da variável que dividem a distribuição em quatro partes iguais.

	1/4	1/2	3/4
25%	25%	25%	25%

Q1: deixa abaixo 25% das observações (elementos do conjunto)

25%	

Q2: deixa abaixo 50% das observações (elementos do conjunto)

50%	

Q3: deixa abaixo 75% das observações (elementos do conjunto)

75%	

QUARTIL

Exercício 4: Determinar os quartis das alturas (em cm) abaixo:

159	160	160	162	164	165	166	168	168	168
169	170	170	171	172	175	178	180	180	195

$$R_i = (n-1).\frac{i}{C} + 1$$

OBS: os dados devem estar em ordem crescente ou decrescente (ROL).

QUARTIL

159	160	160	162	164	165	166	168	168	168
169	170	170	171	172	175	178	180	180	195

1º Quartil

$$R_1 = (20 - 1).\frac{1}{4} + 1 = 5,75$$

Número de Indivíduos

$$R_i = (n - 1).\frac{i}{C} + 1$$

QUARTIL

159	160	160	162	164	165	166	168	168	168
169	170	170	171	172	175	178	180	180	195

$$R_1 = (20-1).\frac{1}{4} + 1 = 5{,}75$$

$$R_1 = 5{,}75 = 5 + 0{,}75$$

$$x_5 \xrightarrow{0{,}75} x_6$$

5º indivíduo → X_5

$$Q_1 = x_5 + 0{,}75 \times (x_6 - x_5) = 164 + 0{,}75 \times (165 - 164) = 164{,}75 \text{ cm}$$

QUARTIL

159	160	160	162	164	165	166	168	168	168
169	170	170	171	172	175	178	180	180	195

$$R_1 = (20-1).\frac{1}{4} + 1 = 5{,}75$$

$$Q_1 = x_5 + 0{,}75 \times (x_6 - x_5) = 164 + 0{,}75 \times (165 - 164) = 164{,}75 \text{ cm}$$

2º Quartil

$$R_2 = (20-1).\frac{2}{4} + 1 = 10{,}5$$

$$Q_2 = x_{10} + 0{,}5 \times (x_{11} - x_{10}) = 168 + 0{,}5 \times (169 - 168) = 168{,}5$$

QUARTIL

159	160	160	162	164	165	166	168	168	168
169	170	170	171	172	175	178	180	180	195

$$R_3 = (20-1).\frac{3}{4}+1 = 15{,}25$$

$$Q_3 = x_{15} + 0{,}25 \times (x_{16} - x_{15}) = 172 + 0{,}25 \times (175 - 172) = 172{,}75$$

$$R_4 = (20-1).\frac{4}{4}+1 = 20$$

$$Q_4 = x_{20} = 195$$

QUARTIL

Q_1=164,75

Q_2=168,5

159	160	160	162	164	165	166	168	168	168
169	170	170	171	172	175	178	180	180	195

Q_3=172,75

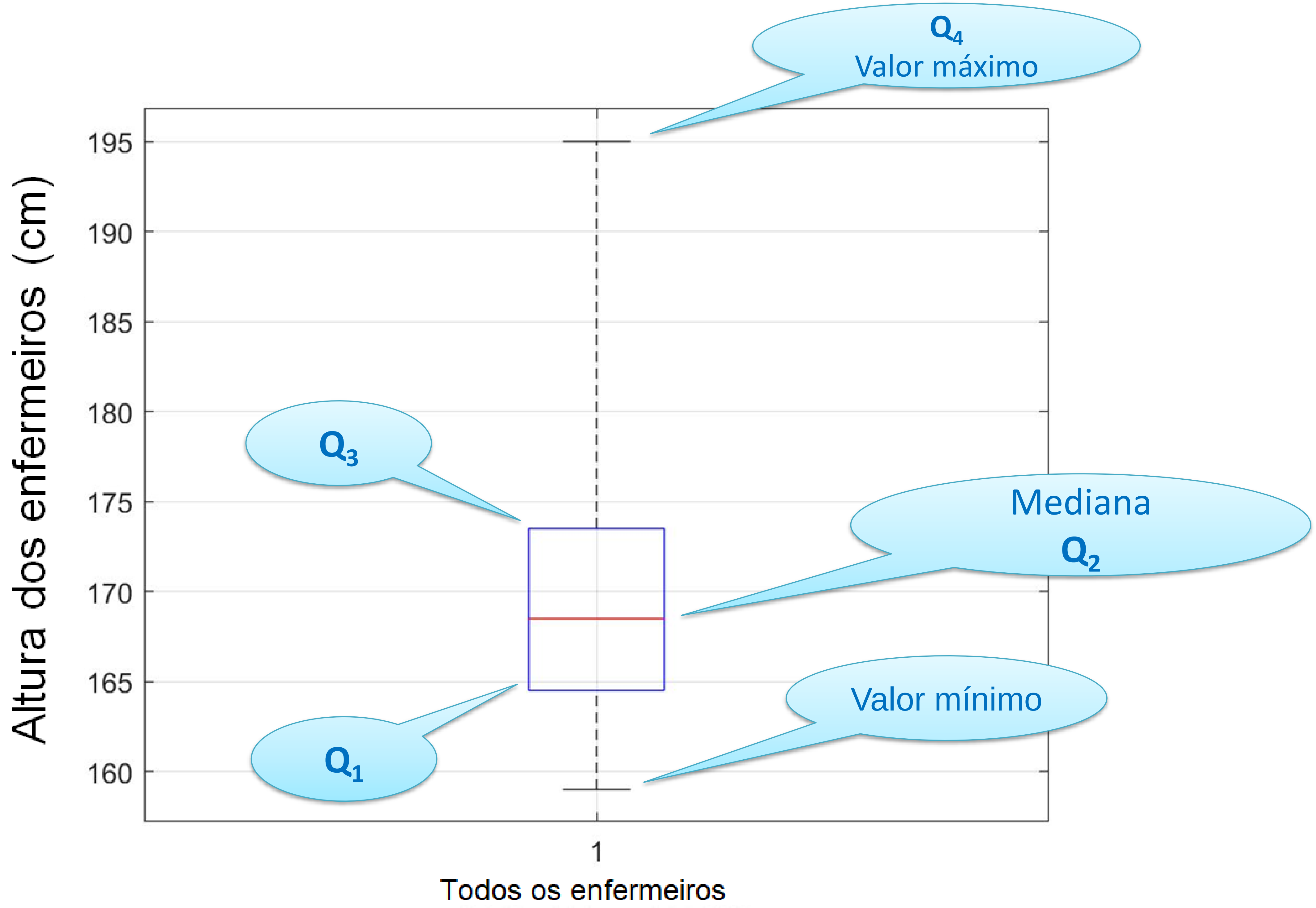
Q_4=195

É HORA DE PRATICAR!

QUARTIL

159	160	160	162	164	165	166	168	168	168
169	170	170	171	172	175	178	180	180	195

Box-Plot

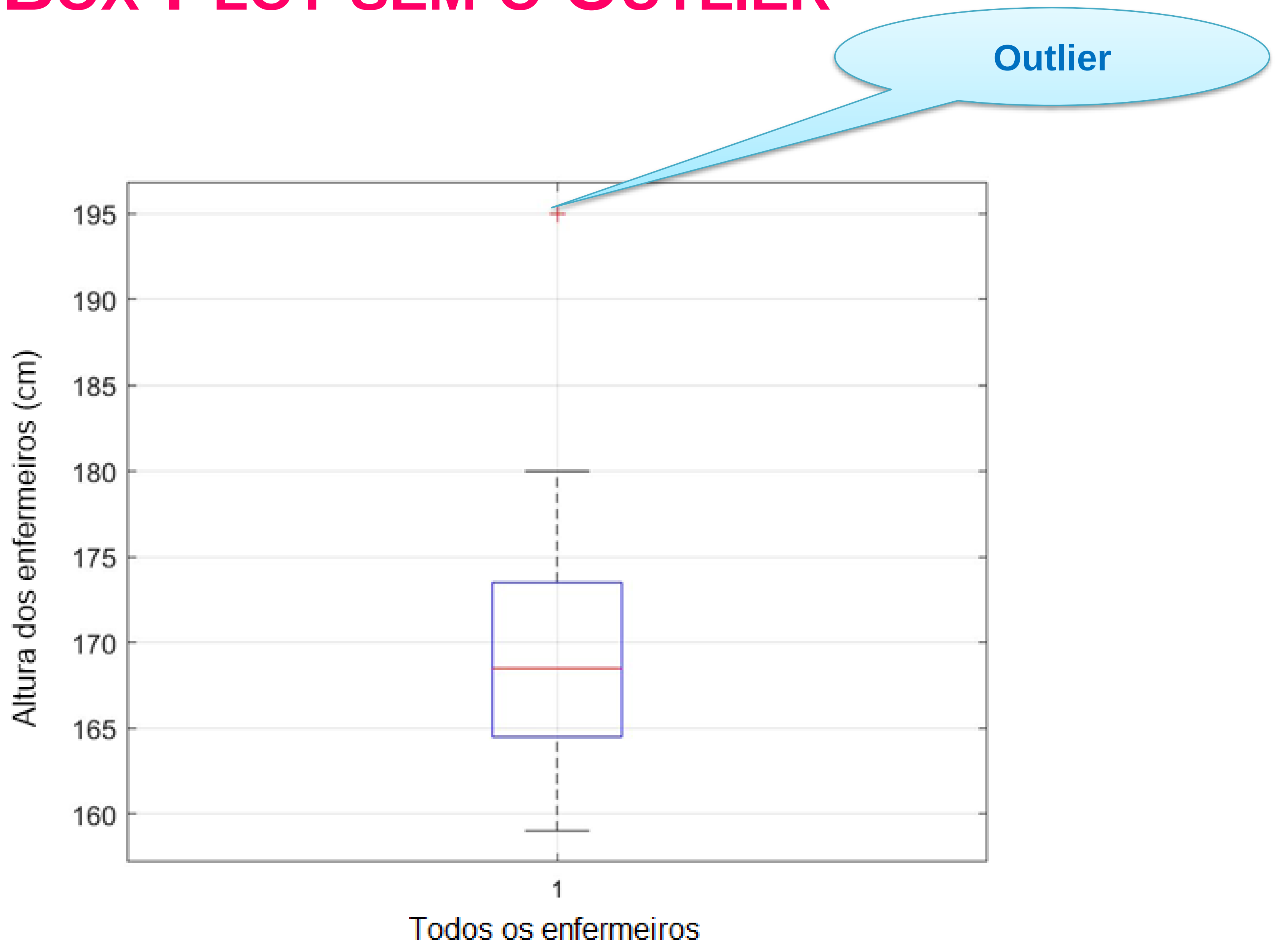


Média = 170

Mediana = 168.5

Moda = 168

BOX-PLOT SEM O OUTLIER



Média = 168.7 Mediana = 168 Moda = 168

Observar que a média de todo conjunto de dados é de 170 cm. A distância entre Q4 e a média é muito grande: conclui-se que 195 é um valor atípico do conjunto de dados. As demais medidas de altura possuem uma distância menor em relação à média. Portanto, o valor 195 está se diferenciando muito, se destacando muito em relação à média. Esse valor é chamado **Outlier**, pois conseguiu-se observar um valor muito atípico no conjunto de dados.

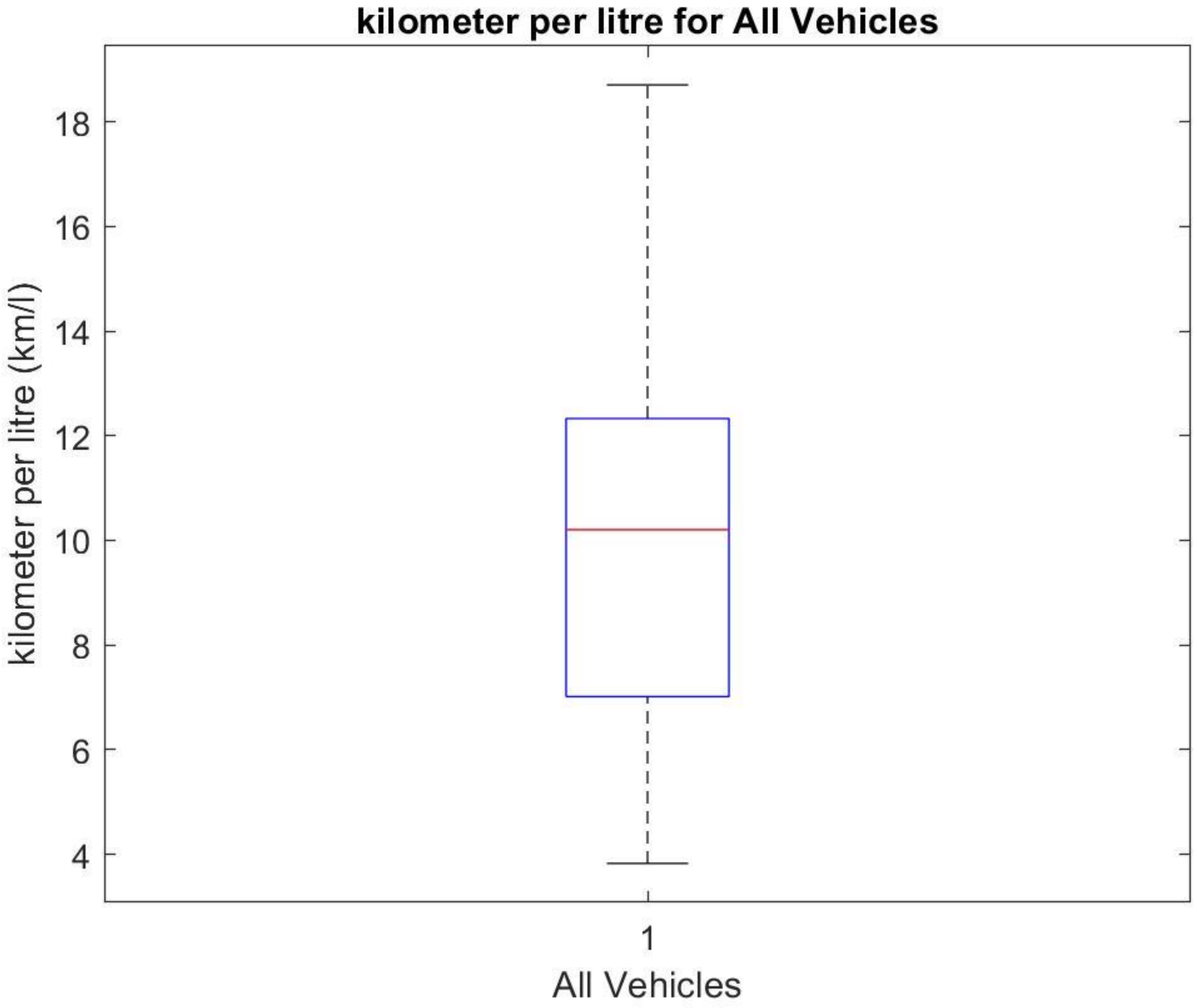
O valor 195 ele distorce um pouco o cálculo dos parâmetros representativos. Ele fez com que a média do grupo subisse muito, ficasse muito alta. Então, é comum realizar os cálculos de média, mediana e moda, retirando o valor Outlier, ou seja, separando o 195 do conjunto e não considerá-lo nesses cálculos.

Nesse gráfico ao lado, que é um novo box-plot, mostra que a distribuição dos dados é mais uniforme do que no gráfico anterior, pois o valor Outlier não foi considerado.

Em softwares, os pontos são desenhados como discrepantes (Outliers) se forem maiores que Q3+W*(Q3-Q1) ou menores que Q1-W*(Q3-Q1). O padrão é W=1,5.

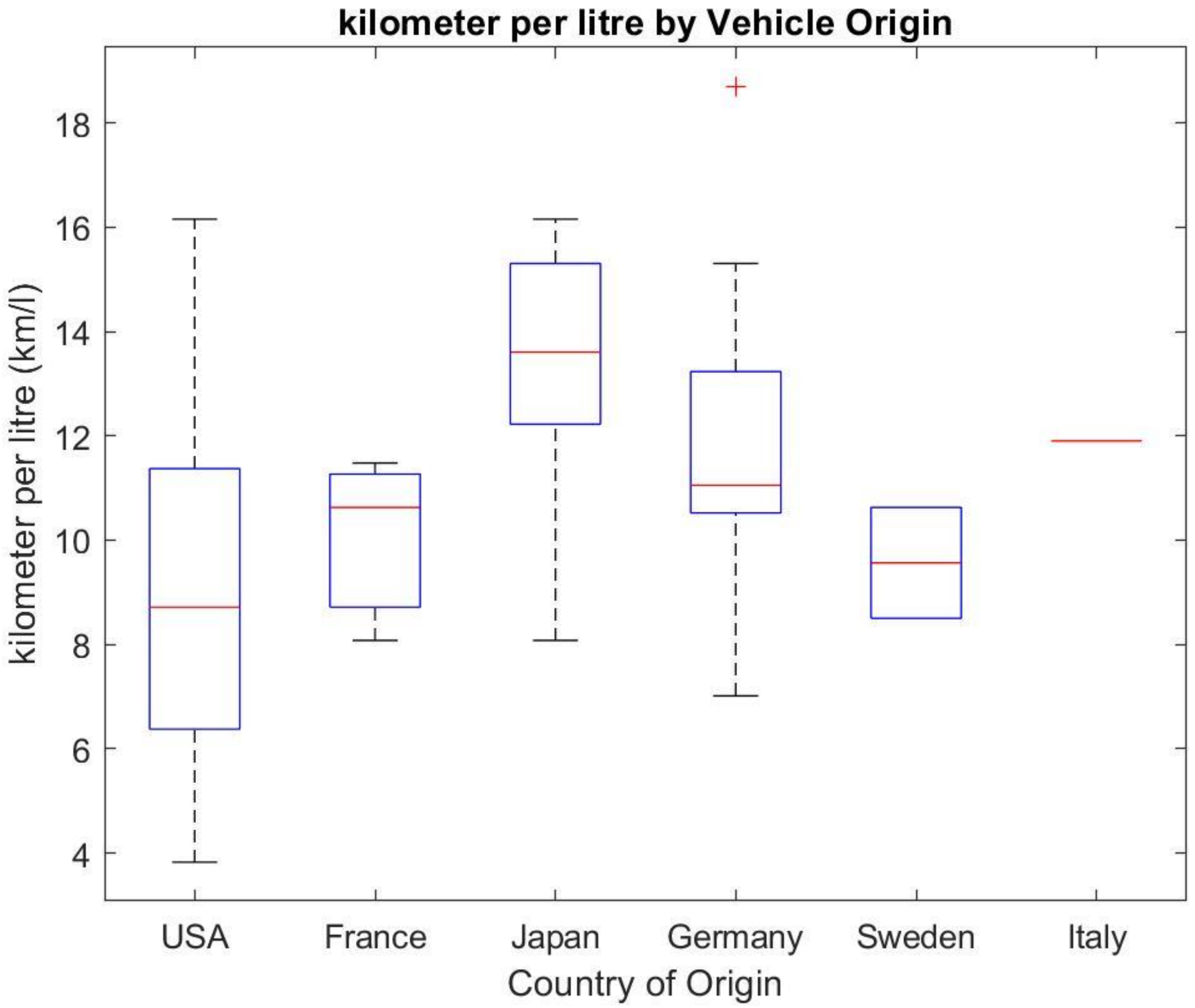
QUARTIL

Box-Plot (consumo de veículos)



QUARTIL

BOX-PLOT



CAMINHOS QUE CONECTAM COM O FUTURO

MEDIDAS DE DISPERSÃO OU VARIABILIDADE

Aspectos gerais

Entende-se por dispersão ou variabilidade a diferença observada entre os valores de um conjunto de dados.

Evidentemente, quanto maior for essa diferença, maior será a dispersão ou variabilidade do conjunto, sendo válido o raciocínio inverso.

Desse modo é possível, por simples observação, caracterizar um conjunto qualitativamente em termos de dispersão.

Entretanto, para evitar o julgamento subjetivo associado à aferição qualitativa da dispersão, é conveniente construir um índice que permita efetuar uma análise quantitativa da variabilidade dos dados.

MEDIDAS DE DISPERSÃO OU VARIABILIDADE

MD's	Amplitude Total, AT (*range*) Soma dos Desvios Absolutos, SDA Desvio Médio, DM	Lineares
	Soma dos Quadrados dos Desvios, SQD Variância, VAR[X] ou σ^2 (*variance*) Desvio Padrão, σ (*standard deviation*)	Quadráticas
	Taxa de anormalidade	Ordem

MEDIDAS DE DISPERSÃO OU VARIABILIDADE

Amplitude Total:

Uma das formas de se medir a dispersão consiste em calcular a Amplitude Total (AT) do conjunto que está sendo observado.

A AT é obtida do seguinte modo:

Seja $X = \{x_1, x_2, \ldots, x_n\}$ com $x_1 < x_2 < \ldots < x_n$

Então $AT = x_n - x_1$

MEDIDAS DE DISPERSÃO OU VARIABILIDADE

Amplitude Total:

Apesar de ter a vantagem da simplicidade, a Amplitude Total é considerada um indicador inadequado para a mensuração da variabilidade, pois:

- A amplitude total não considera a totalidade dos dados do conjunto e, sim, apenas dois deles (o maior e o menor). Dessa forma, o indicador não é sensível à posição que os “n-2” valores restantes ocupam no conjunto.

- No caso de dados agrupados em tabelas, os limites abertos não permitem o cálculo da amplitude total.

MEDIDAS DE DISPERSÃO OU VARIABILIDADE

Amplitude Total:

Sejam os conjuntos:

A = {1, 7, 7, 8, 8, 8, 9, 9, 12, 15}
B = {3, 3, 4, 4, 8, 11, 13, 13, 14, 14}

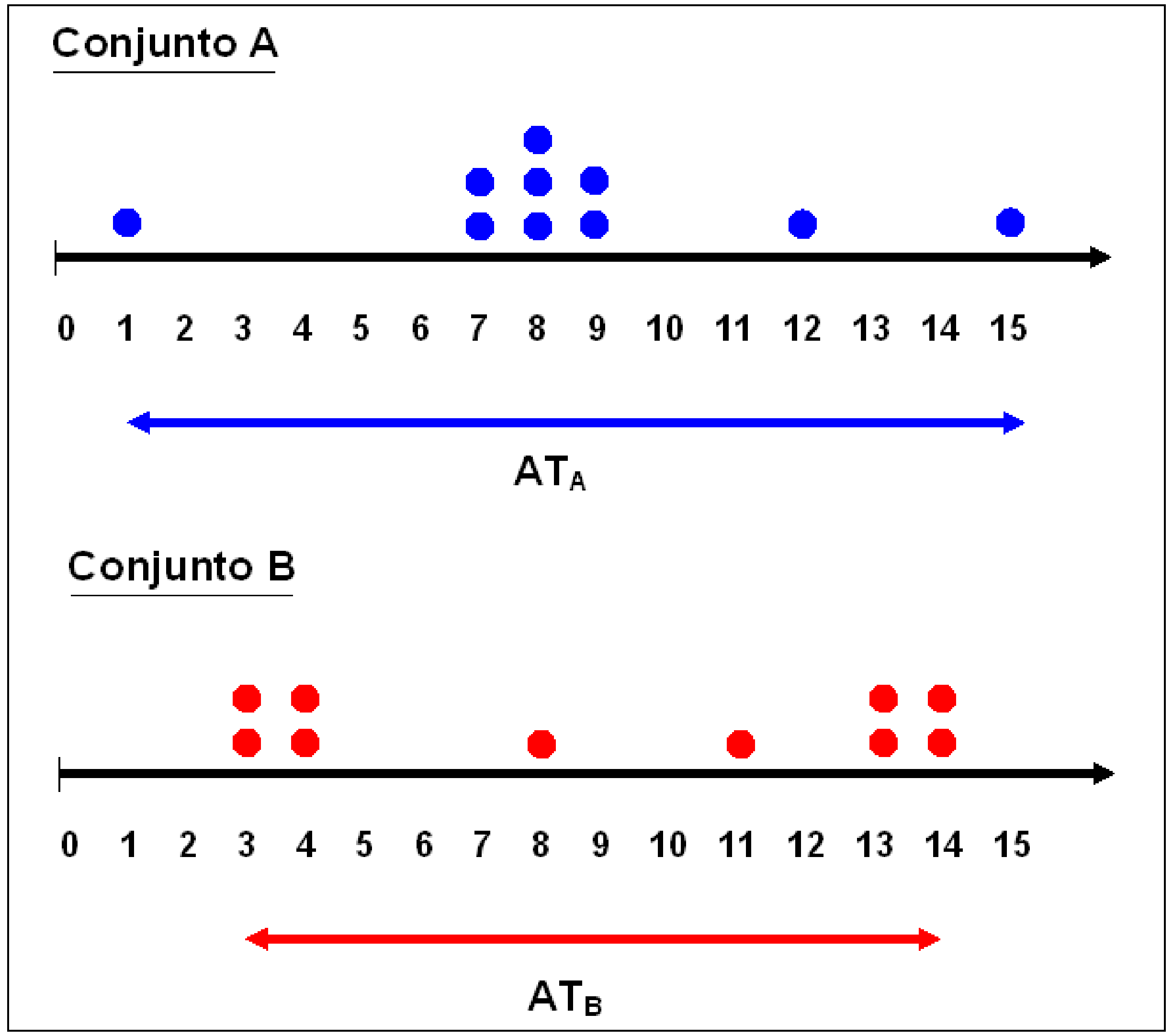
Dispersão [A] = AT_A = 15 – 1 = 14
Dispersão [B] = AT_B = 14 – 3 = 11

Entretanto, uma simples análise visual dos valores dos dois conjuntos, devidamente desenhados em uma escala graduada, mostra que a amplitude total reflete mal a dispersão dos conjuntos, tal como definido anteriormente.

MEDIDAS DE DISPERSÃO OU VARIABILIDADE

Amplitude Total:

Embora tenha-se mostrado que a Amplitude Total do conjunto A é maior que a do conjunto B ($AT_A > AT_B$), percebe-se uma dispersão menor dos valores do conjunto A em relação à do conjunto B.



MEDIDAS DE DISPERSÃO OU VARIABILIDADE

Variância:

O desvio quadrático médio, ou, como é comumente conhecido, a Variância é calculado pela equação:

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

MEDIDAS DE DISPERSÃO OU VARIABILIDADE

Desvio Padrão

Como a variância é obtida somando-se valores elevados ao quadrado, expressa a variabilidade dos dados como uma grandeza também ao quadrado.

Para solucionar esse incômodo, basta extrair a raiz quadrada da Variância, obtendo-se assim um outro indicador de variabilidade, denominado Desvio Padrão.

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

Inatel | CAMINHOS QUE CONECTAM COM O FUTURO

MEDIDAS DE DISPERSÃO OU VARIABILIDADE

Correção para amostras:

Quando se trata de amostras extraídas de uma população, o cálculo da Variância e do Desvio Padrão sofre uma pequena alteração, denominada correção amostral, que consiste em retirar uma unidade do denominador.

Ainda é preciso lembrar que, quando se quer indicar um parâmetro amostral, é usada a letra latina correspondente à letra grega utilizada no caso de populações. Assim, as expressões para o cálculo da Variância e do Desvio Padrão amostral passam a ser, respectivamente:

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

No Exercício 1

Seja X o conjunto dos perímetros cefálicos, em centímetros, de 5 recém-nascidos.

X = { 34, 30, 33, 32, 33}

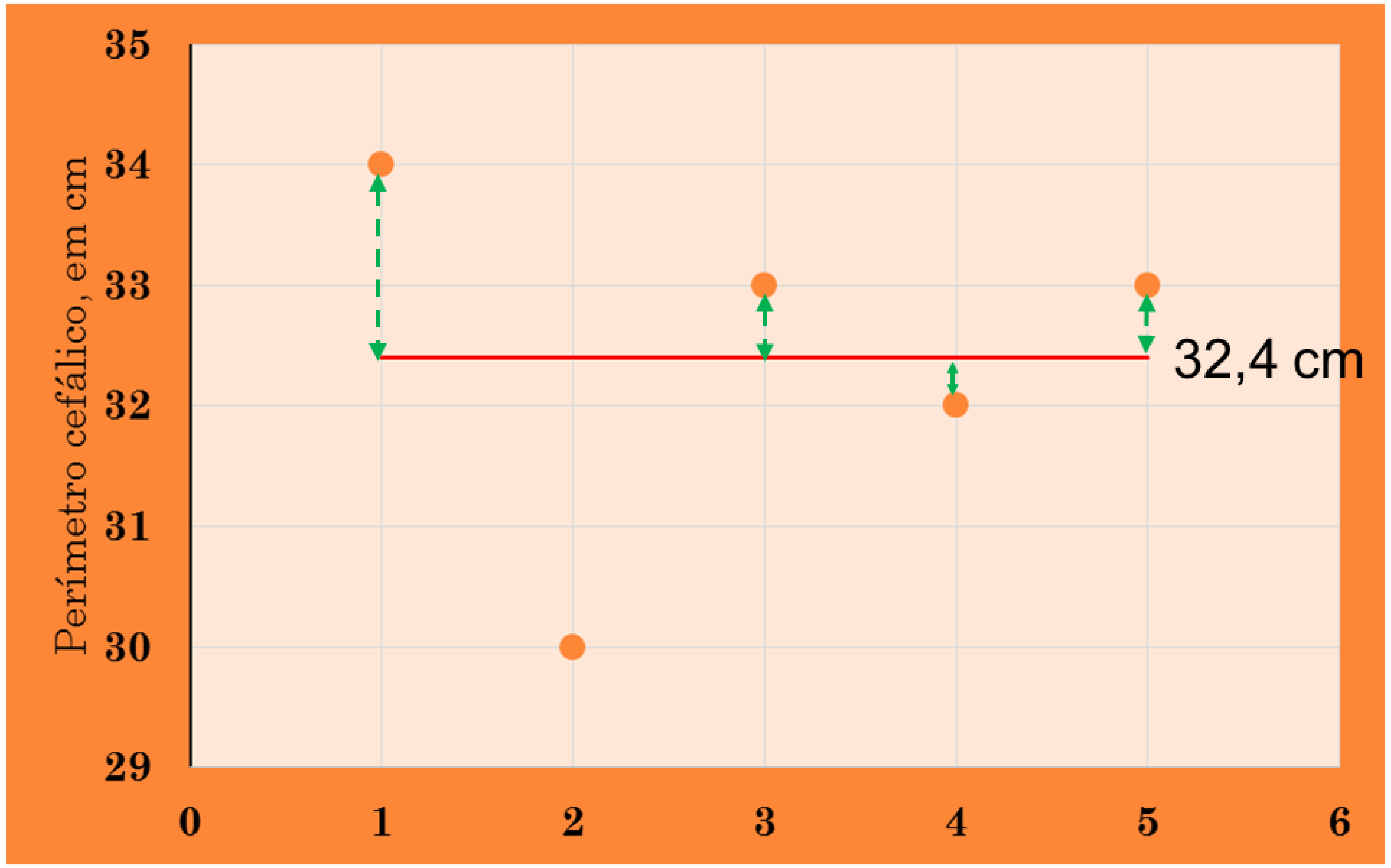
Então, o perímetro cefálico médio dos cinco RN’s resulta:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{34 + 30 + 33 + 32 + 33}{5} = 32{,}4\text{cm}$$

Exercício 1

OBS: a soma dos desvios em relação a Média é zero

d1 =	1.6
d2 =	-2.4
d3 =	0.6
d4 =	-0.4
d5 =	0.6
S =	0.0



Exercício 1

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
34	1,6	2,56
30	-2,4	5,76
33	0,6	0,36
32	-0,4	0,16
33	0,6	0,36
		9,2

soma dos desvios

Exercício 1

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
34	1,6	2,56
30	-2,4	5,76
33	0,6	0,36
32	-0,4	0,16
33	0,6	0,36
		9,2

soma dos desvios

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{9,2}{4} = 2,3\,\text{cm}^2$$

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{9,2}{4}} = \sqrt{2,3} = 1,52\text{ cm}$$

MEDIDAS DE DISPERSÃO OU VARIABILIDADE

Coeficiente de variação amostral:

É o quociente entre o desvio padrão e a média amostrais

$$cv = \frac{s}{\overline{X}}$$

MEDIDAS DE DISPERSÃO OU VARIABILIDADE

Exemplo. Compare a variabilidade relativa do tempo de reação de um analgésico A com a variabilidade do peso das pessoas que se submeteram à dosagem desse analgésico. As médias e os desvios padrão foram:

Analgésico A: $\bar{X} = 3$ min e s = 0,71
Peso das pessoas: $\bar{X} = 58{,}25$ kg e s = 5,17

Solução: Vamos calcular o coeficiente de variação para cada item observado.
Cálculo para o tempo de reação do analgésico:

$$Cv = 100 \cdot \frac{0{,}71}{3} = 23{,}67\%$$

Cálculo para o peso das pessoas:

$$Cv = 100 \cdot \frac{5{,}17}{58{,}25} = 8{,}88\%$$

Comparando o coeficiente de variação do tempo de reação do analgésico e o do peso das pessoas, podemos concluir que os dados referentes ao peso são mais consistentes que os dados referentes ao tempo de reação do analgésico, ou ainda, que os dados referentes ao peso são mais homogêneos que os do tempo de reação do analgésico.

RIGONATTO, Marcelo. "Coeficiente de variação"; *Brasil Escola*. Disponível em: https://brasilescola.uol.com.br/matematica/coeficiente-variacao.htm. Acesso em 09 de setembro de 2020.

É HORA DE PRATICAR!

Com base nos dados da aba **Dados**, calcule os parâmetros solicitados nas abas **Medidas_Peso** e **Medidas_Cintura**.

Compare a variabilidade do peso com a variabilidade da cintura e conclua qual distribuição é mais homogênea (uniforme): a distribuição do peso ou a distribuição da cintura?

Gere o gráfico de dispersão de ambas as distribuições.

Exercício 2

Considere os dados, referentes às taxas de potássio de 12 pacientes:

Paciente	1	2	3	4	5	6	7	8	9	10	11	12
Potássio (mg/100ml)	4,8	4,9	4,9	4,9	5,0	5,1	5,4	6,5	6,7	6,7	6,7	7,0

Calcule as medidas de variabilidade.

Média = 5,72

Inatel | CAMINHOS QUE CONECTAM COM O FUTURO

EQUAÇÕES

MEDIDAS DE TENDÊNCIA CENTRAL

Média:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Mediana:

$$\hat{x} = x_{\frac{n+1}{2}}$$

Se n é impar

$$\hat{x} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

Se n é par

Moda: Número que mais aparece no rol

MEDIDAS DE DISPERSÃO OU VARIABILIDADE

Amplitude Total: $$AT = \text{Maior Valor} - \text{Menor Valor}$$

Variância:

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

OBS
AMOSTRA

Desvio Padrão:

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

Rosimara Salgado

Professora
Coordenadora do NEaD

rosimara@inatel.br

Inatel | CAMINHOS QUE CONECTAM COM O FUTURO