BIOESTATÍSTICA

Profª. Karina Perez Mokarzel Carneiro

MEDIDAS CARACTERÍSTICAS DE UMA DISTRIBUIÇÃO

- **Bibliografia:**

Arango HG. Bioestatística: teórica e computacional. 3 ed. Rio de Janeiro: Guanabara Koogan, 2011.

MEDIDAS CARACTERÍSTICAS DE UMA DISTRIBUIÇÃO

Na estatística descritiva quanto maior a facilidade em transmitir as informações sobre a população em estudo para quem as estiver recebendo, mais eficiente será o meio de transmissão.

MEDIDAS CARACTERÍSTICAS DE UMA DISTRIBUIÇÃO

- **Medidas de tendência Central**:

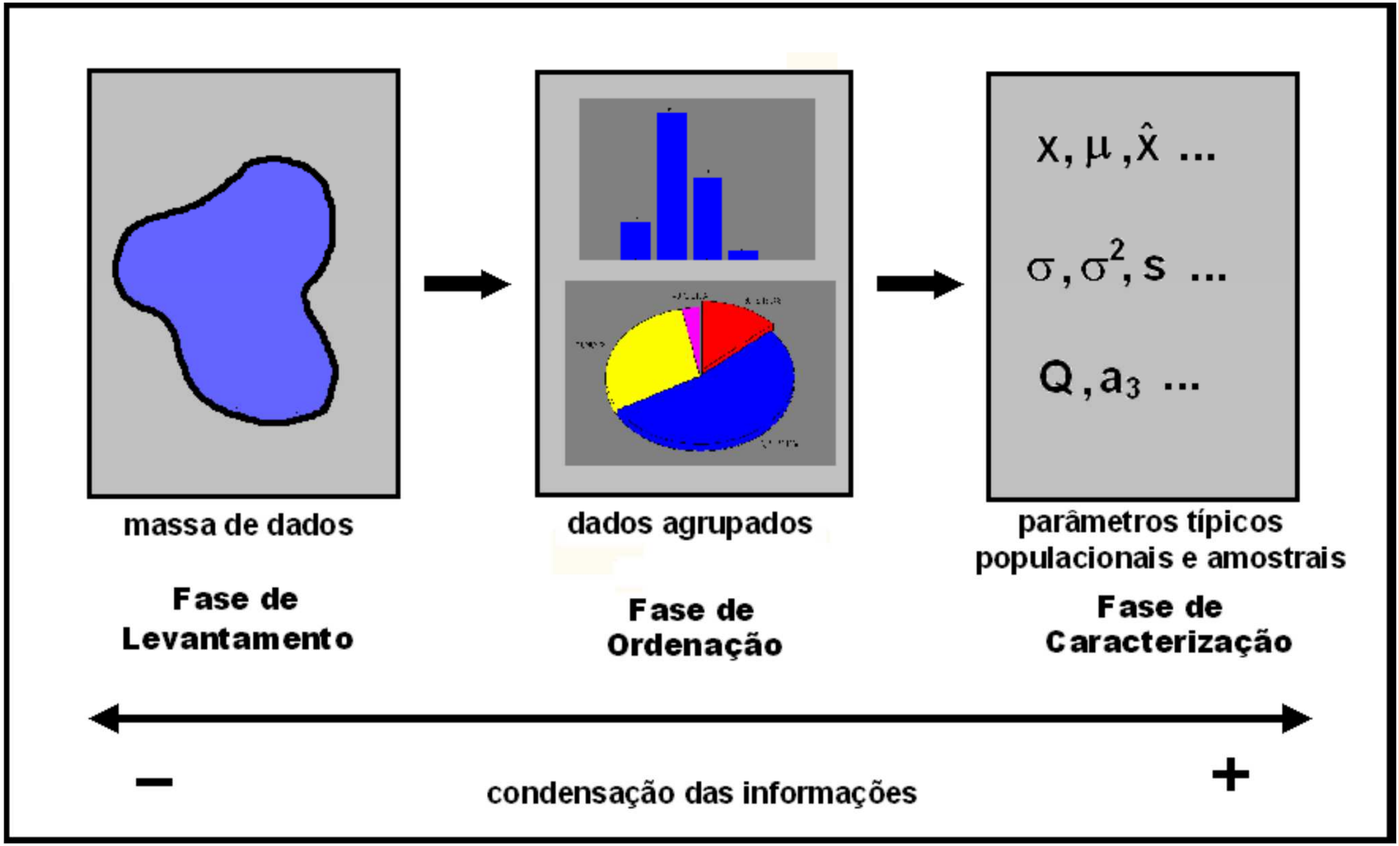
  Média, Mediana e Moda.

- **Medidas de Dispersão ou Variabilidade**:

  Amplitude Total, Variância e Desvio Padrão.

- **Medidas de Assimetria**

MEDIDAS CARACTERÍSTICAS DE UMA DISTRIBUIÇÃO



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

MEDIDAS DE TENDÊNCIA CENTRAL

Média aritmética

Dado o conjunto de n valores da variável X,

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

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

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

MEDIDAS DE TENDÊNCIA CENTRAL

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.

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

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

MEDIDAS DE TENDÊNCIA CENTRAL

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.

→O resultado pode ser **amodal**, quando não possui moda; **bimodal**, quando possui dois valores modais; ou **multimodal**, quando possui mais do que dois valores modais.

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

Dados em ROL: 30, 32, 33, 33, 34

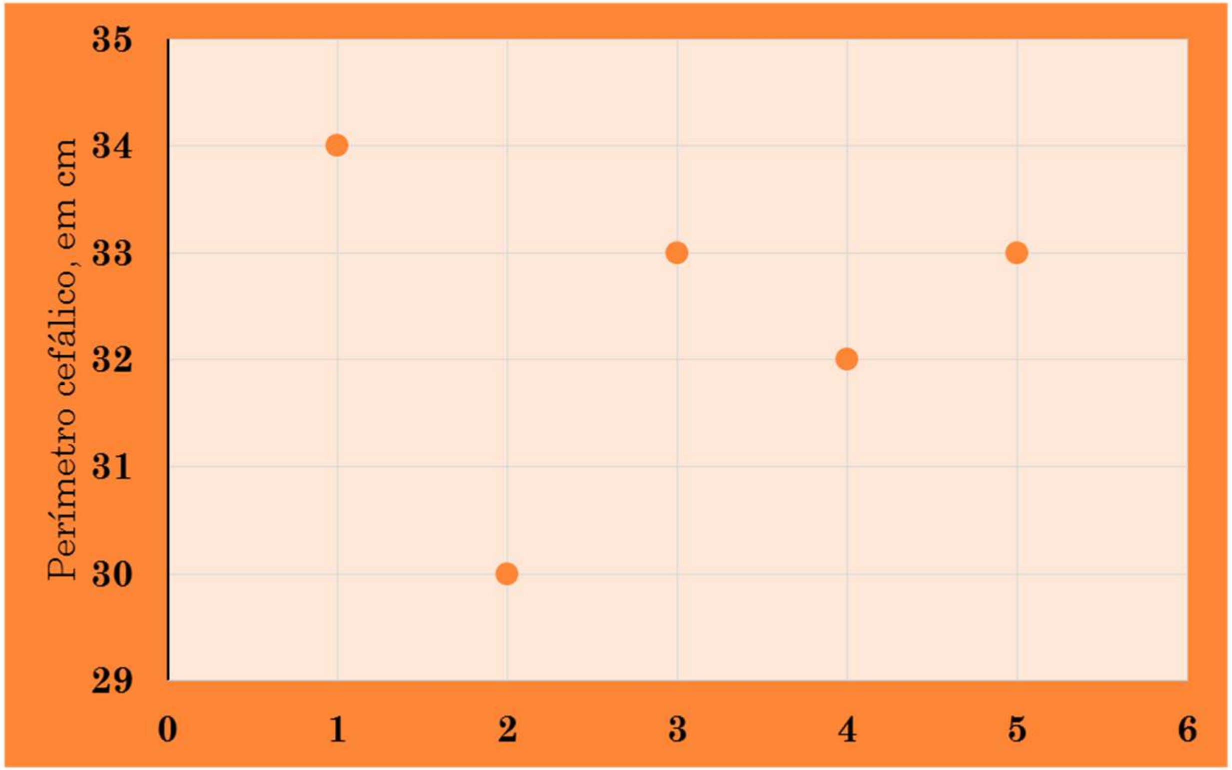
Mediana: $\hat{x} = 33$

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

Moda (nº que repete com maior frequência): $\tilde{x} = 33$

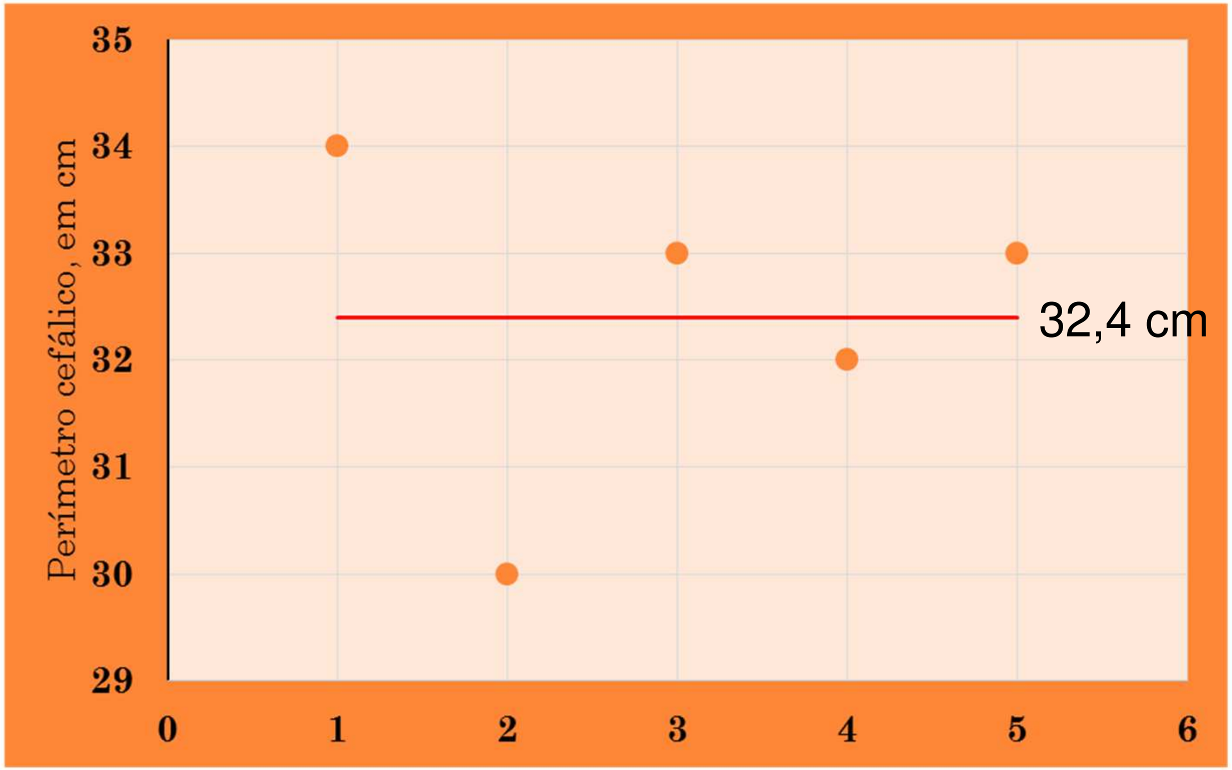
Exercício 1

Gráfico de dispersão dos perímetros cefálicos, em cm.



Exercício 1

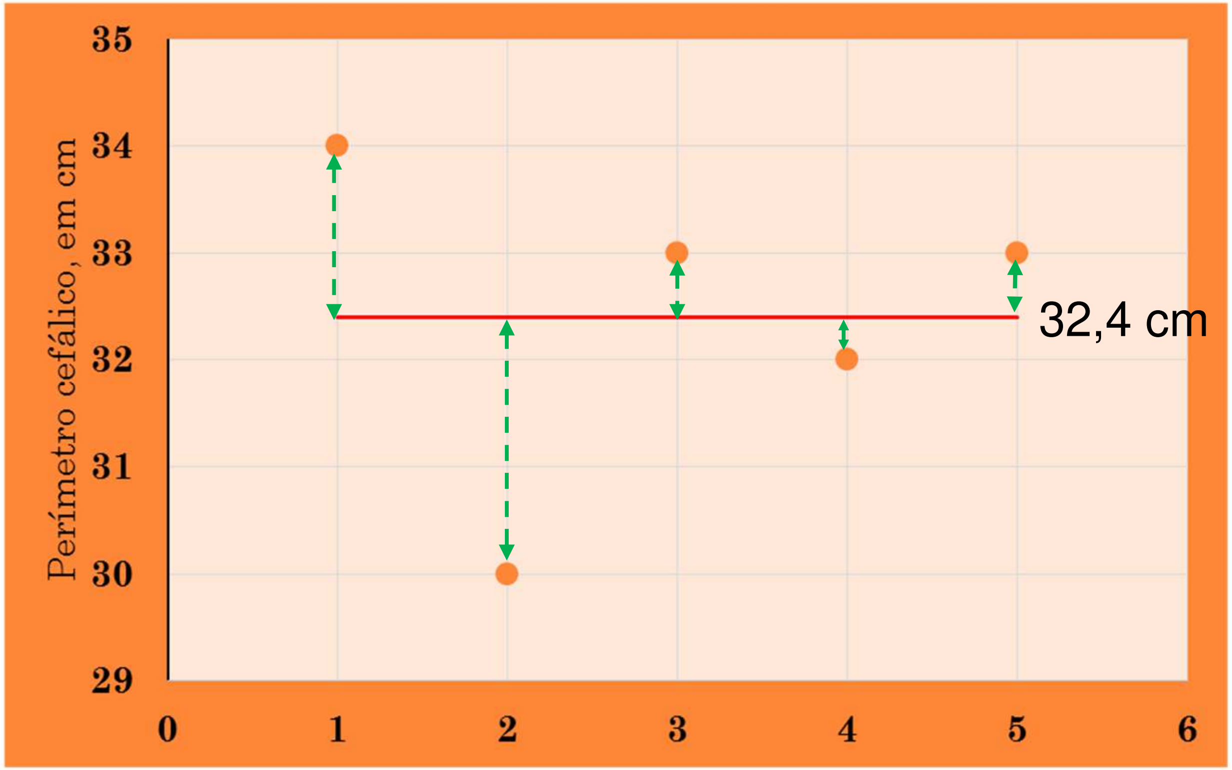
Gráfico de dispersão dos perímetros cefálicos, em cm.



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

GRÁFICO DE DISPERSÃO

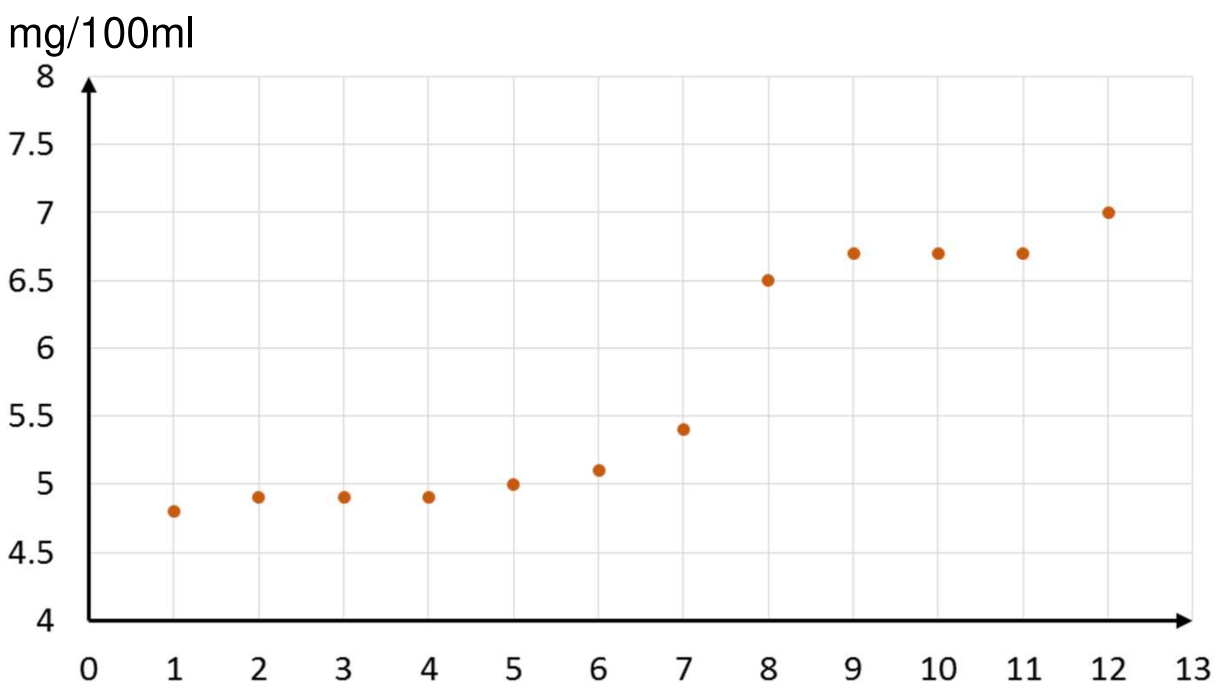


Figura 1 – Taxa de potássio em mg/100ml

Exercício 2

GRÁFICO DE DISPERSÃO

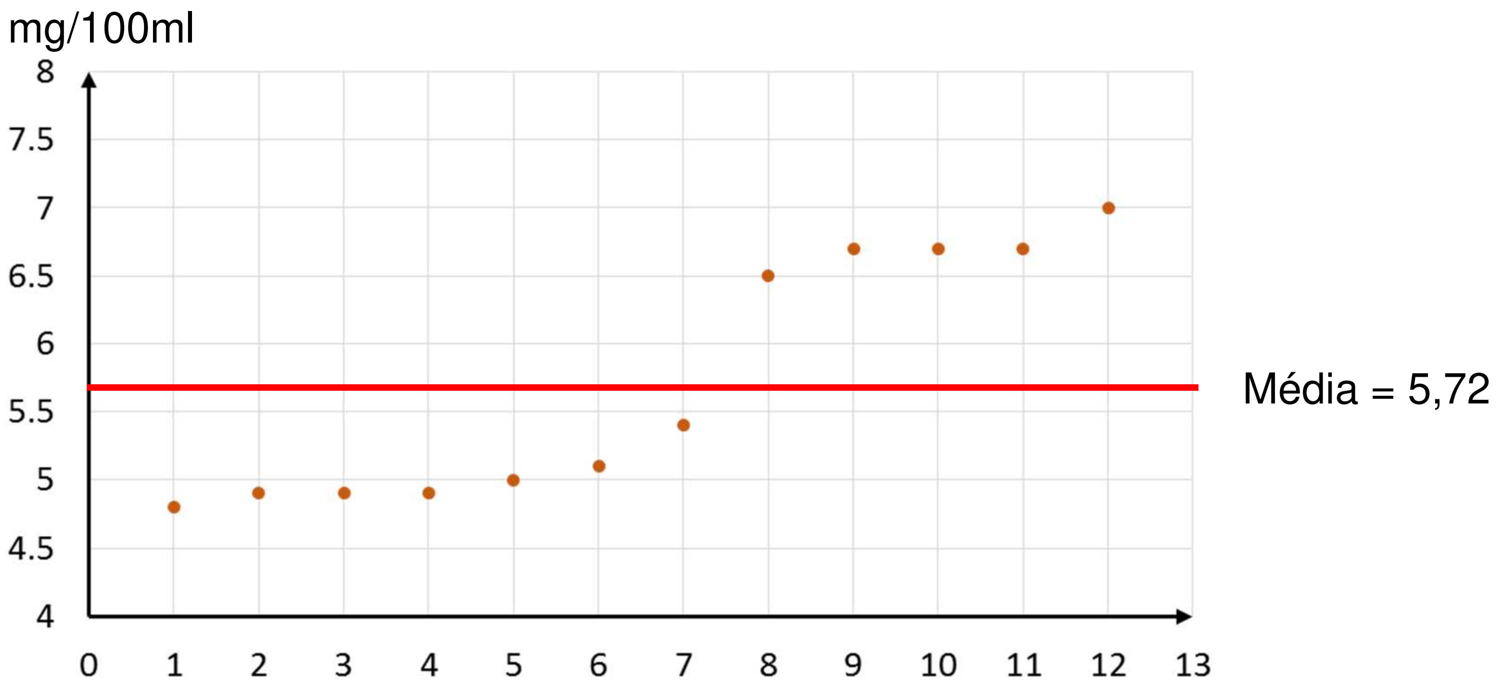


Figura 1 – Taxa de potássio em mg/100ml

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 do conjunto que está sendo observado.

A amplitude total 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}

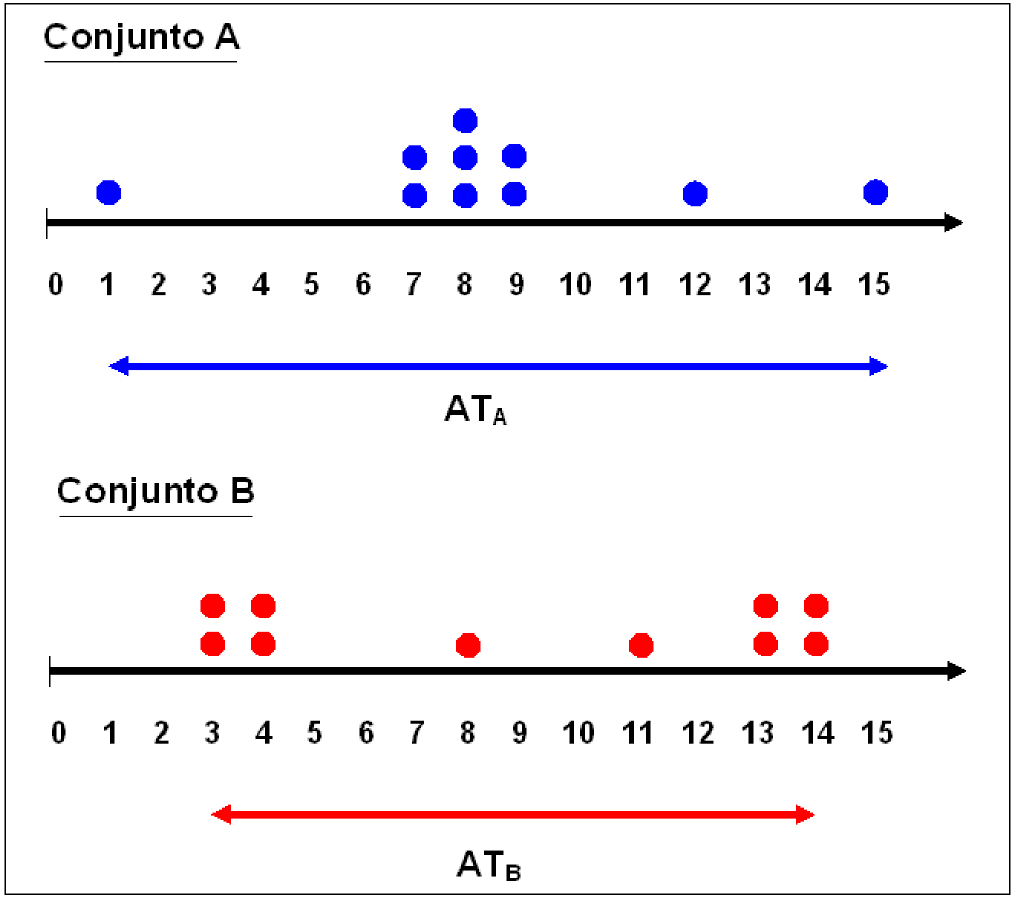
Dispersão [A] = AT_A = 15 – 1 = 14
Dispersão [B] = AT_B = 14 – 3 = 11

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

→ <u>Amplitude Total</u>:

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.

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

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.

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

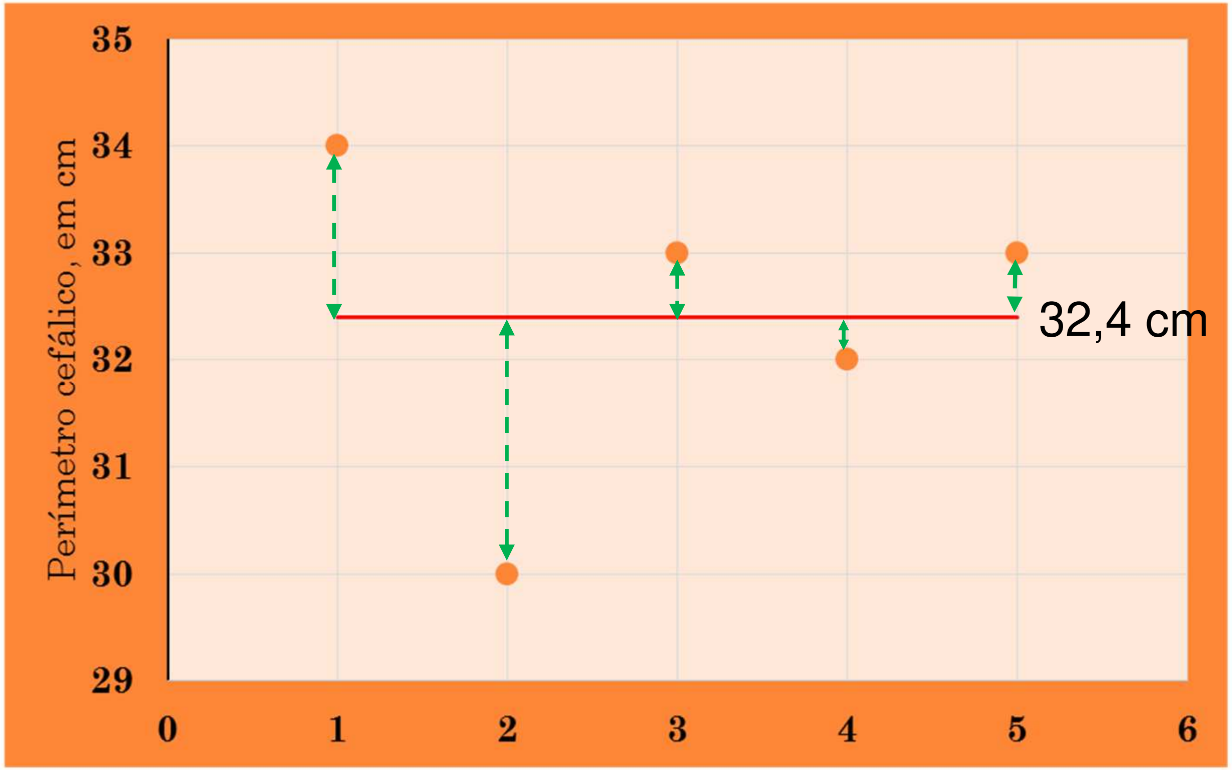
Então, o perímetro cefálico médio dos cinco RN's resulta:

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

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

xi	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
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

xi	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
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

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

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

EQUAÇÕES

Medidas de Tendência Central

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

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

FIM