



Aprendendo (e ensinando) R para análise de dados

Estatística Descritiva, parte 1: média e desvio



[CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

Marcos V. C. Vital
Lab. de Ecologia Quantitativa
ICBS-UFAL

Antes de começar

- Conteúdo aberto, com licença [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/).
- Tudo aqui: marcosvital.github.io/Aprendendo-R/
- Usem, compartilhem, modifiquem!

Estatística descritiva

- Descrevendo e visualizando dados.
 - Medidas de tendências central e dispersão.
 - Medidas de posição.
 - Alguns gráficos.

Tendência central e dispersão: parte 1

A boa e velha média:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

Tendência central e dispersão: parte 1

A boa e velha média:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

Ou, usando uma notação um pouco mais formal:

$$\bar{x} = \sum_{i=1}^n x_i / n$$

Tendência central e dispersão: parte 1

Visualmente, é algo assim:



Tendência central e dispersão: parte 1

Mas temos um problema:



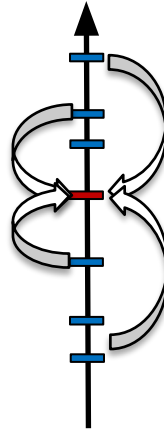
Tendência central e dispersão: parte 1

Qual a solução?

- Medir a dispersão dos dados!

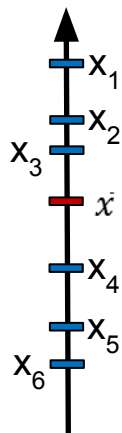
Tendência central e dispersão: parte 1

Construindo uma medida de dispersão:



Tendência central e dispersão: parte 1

Construindo uma medida de dispersão:

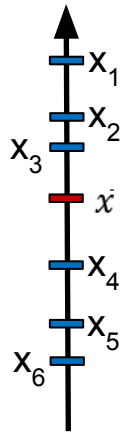


Podemos calcular uma média?

$$\frac{(x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + (x_4 - \bar{x}) + (x_5 - \bar{x}) + (x_6 - \bar{x})}{n}$$

Tendência central e dispersão: parte 1

Construindo uma medida de dispersão:

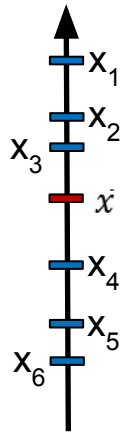


Podemos calcular uma média?

$$\frac{(x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + (x_4 - \bar{x}) + (x_5 - \bar{x}) + (x_6 - \bar{x})}{n} = 0 \dots ???$$

Tendência central e dispersão: parte 1

Construindo uma medida de dispersão:

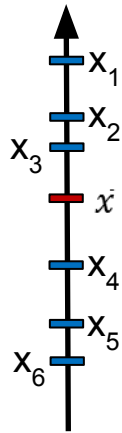


O que aconteceu?

- Temos valores acima e abaixo da média...
 - Então temos diferenças positivas e negativas!

Tendência central e dispersão: parte 1

Construindo uma medida de dispersão:



O que aconteceu?

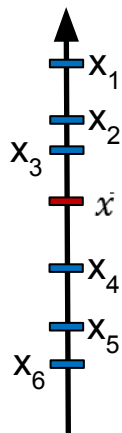
- Temos valores acima e abaixo da média...
 - Então temos diferenças positivas e negativas!

Qual a solução?

- Vamos elevar todo mundo ao quadrado!

Tendência central e dispersão: parte 1

Construindo uma medida de dispersão:

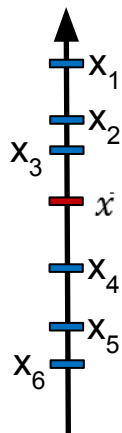


A nossa solução:

$$\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + (x_4 - \bar{x})^2 + (x_5 - \bar{x})^2 + (x_6 - \bar{x})^2}{n}$$

Tendência central e dispersão: parte 1

Construindo uma medida de dispersão:



A nossa solução:

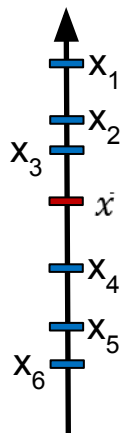
$$\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + (x_4 - \bar{x})^2 + (x_5 - \bar{x})^2 + (x_6 - \bar{x})^2}{n}$$

Peraí que ainda tem um pequeno probleminha...

- Nossas amostras podem subestimar a variação real!
 - Mas também tem solução, é claro!

Tendência central e dispersão: parte 1

Construindo uma medida de dispersão:



A nossa nova solução:

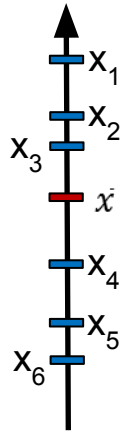
$$\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + (x_4 - \bar{x})^2 + (x_5 - \bar{x})^2 + (x_6 - \bar{x})^2}{n - 1}$$

É pequena, mas tem bastante efeito.

- E esta, pessoal, é a famosa fórmula da variância!

Tendência central e dispersão: parte 1

Construindo uma medida de dispersão:



Ou, usando uma notação um pouco mais formal:

$$VAR = \sum_1^n (x_i - \bar{x})^2 / (n - 1)$$

Tendência central e dispersão: parte 1

Yes! Agora temos uma medida de dispersão!

- Mas peraí, como eu interpreto isso?
 - Diferenças ao quadrado...?

Tendência central e dispersão: parte 1

Yes! Agora temos uma medida de dispersão!

- Mas peraí, como eu interpreto isso?
 - Diferenças ao quadrado...?

A solução é extrair a raiz quadrada!

- Agora podemos interpretar melhor, não é?
 - Então, finalmente, chegamos ao desvio padrão.

Tendência central e dispersão: parte 1

Colocando em uma formulinha:

$$DP = \sqrt{VAR}$$

- Esta é fácil, não é?
 - O mais importante é a interpretação!

Antes de terminar

- Lembrem-se, jamais confie apenas na média.
 - Uma medida de dispersão é fundamental (e desvio padrão é normalmente uma boa escolha).

Antes de terminar

- Lembrem-se, jamais confie apenas na média.
 - Uma medida de dispersão é fundamental (e desvio padrão é normalmente uma boa escolha).
- Mas... nem sempre média é uma boa medida!
 - Vamos falar mais sobre isso depois.