

# Variabilidade

## Incertezas de dados numéricos

Felipe Figueiredo

- 1 O desvio padrão
  - Fontes de Variabilidade
  - Visualizando a variabilidade com histogramas
  - Média e a mediana
  - Quantificando com percentis
  - Quantificando com variância e DP
  - N ou N-1?
  - Interpretação do DP
  - Exercícios

- Medidas sumárias resumem a informação contida nos dados em um pequeno conjunto de números.
- Medidas sumárias de **populações** se chamam **parâmetros**, e são representadas por letras gregas ( $\mu$ ,  $\sigma^2$ ,  $\sigma$ , etc).
- Medidas sumárias de **amostras** se chamam **estatísticas** e são representadas por letras comuns ( $\bar{x}$ ,  $s^2$ ,  $s$ , etc).
- *Geralmente trabalhamos com estatísticas descritivas.*

## Tipos de medidas sumárias

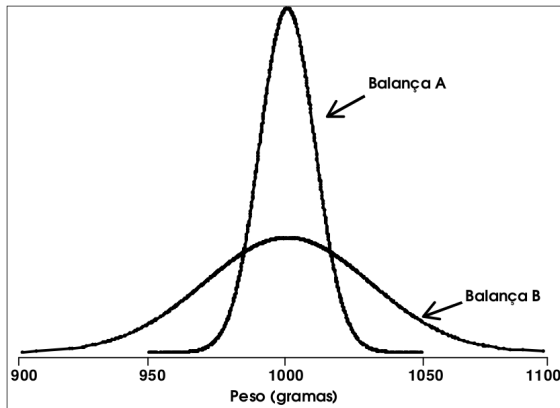
Os dois principais tipos de medidas sumárias utilizadas na literatura são:

- Medidas de Tendência Central
- Medidas de Variabilidade (ou Dispersão)

1

## O desvio padrão

- Fontes de Variabilidade
- Visualizando a variabilidade com histogramas
- Média e a mediana
- Quantificando com percentis
- Quantificando com variância e DP
- N ou N-1?
- Interpretação do DP
- Exercícios



**Figura:** Variabilidade da medição de uma esfera metálica de 1000g. Balança A, “impresisão” de 50g, balança B, “impresisão” de 100g (Fonte: Reis, Reis, 2002)

- Imprecisão ou erro experimental
- Variabilidade biológica
- “*Mancadas*” experimentais

## Conceito de Erro na Estatística

No contexto acadêmico, **erro** não tem o mesmo significado do cotidiano.

Erro se refere a todas as fontes de variabilidade acima.

Outro nome comum é **dispersão** (*scatter*).

- 1 O desvio padrão
  - Fontes de Variabilidade
  - Visualizando a variabilidade com histogramas
  - Média e a mediana
  - Quantificando com percentis
  - Quantificando com variância e DP
  - N ou N-1?
  - Interpretação do DP
  - Exercícios



## Exemplo

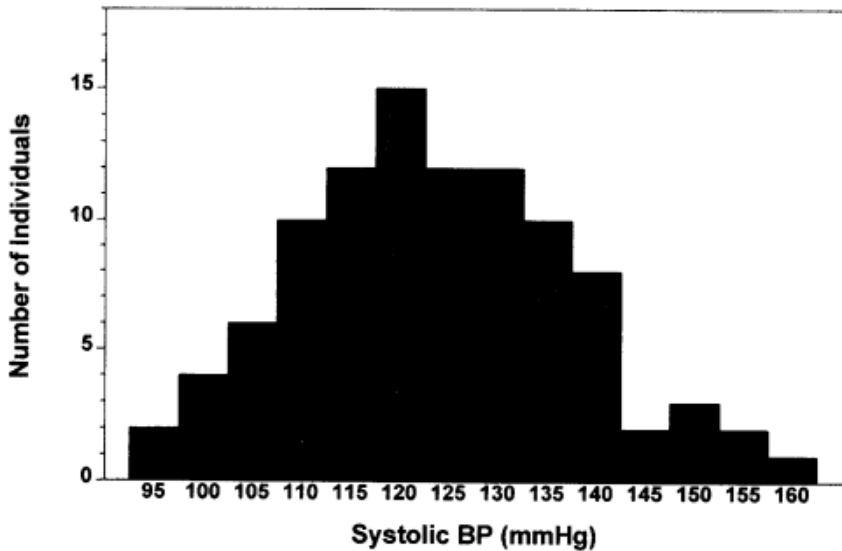
100 estudantes de [*insira aqui um curso da área da saúde*] trabalharam em pares, e mediram a pressão sistólica de seu parceiro(a).

Ao final do exercício, a turma obteve 100 valores de pressão sistólica.

## Pergunta

Como “entender” essa listagem de 100 números?

# O histograma



# Quantas barras?

## O desvio padrão

Fontes de  
Variabilidade

Visualizando a  
variabilidade com  
histogramas

Média e a mediana

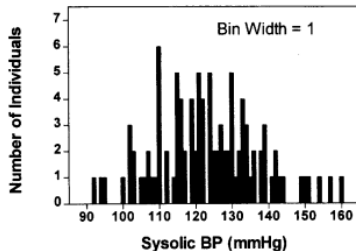
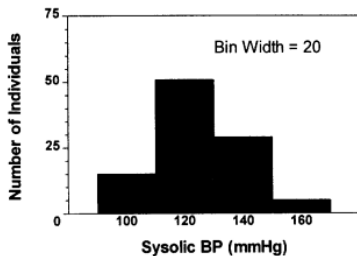
Quantificando com  
percentis

Quantificando com  
variância e DP

N ou N-1?

Interpretação do DP

Exercícios



- 1 O desvio padrão
  - Fontes de Variabilidade
  - Visualizando a variabilidade com histogramas
  - **Média e a mediana**
  - Quantificando com percentis
  - Quantificando com variância e DP
  - N ou N-1?
  - Interpretação do DP
  - Exercícios

## Exemplo

Foram observados os seguintes níveis de colesterol de uma amostra de pacientes. Qual é o nível médio de colesterol nestes pacientes?

$$x_1 = 142$$

$$x_2 = 144$$

$$x_3 = 176$$

$$x_4 = 203$$

$$x_5 = 134$$

$$x_6 = 191$$

$$\bar{x} = \frac{990}{6} = 165$$

## O desvio padrão

Fontes de Variabilidade

Visualizando a variabilidade com histogramas

**Média e a mediana**

Quantificando com percentis

Quantificando com variância e DP

N ou N-1?

Interpretação do DP

Exercícios

## Definition

A mediana é o dado que ocupa o percentil de 50% dados (**posição central**).

- Para se calcular a mediana, deve-se ordenar os dados.
- Encontrar o valor do **meio** se  $n$  for ímpar.
- Encontrar a média dos dois valores do **meio** se  $n$  for par.

## Exemplo

Conforme no exemplo anterior

$$x_5 = 134$$

$$x_1 = 142$$

$$x_2 = 144$$

$$x_3 = 176$$

$$x_6 = 191$$

$$x_4 = 203$$

$$M_d = \frac{144 + 176}{2} = 160$$

# Qual é a diferença?



Variabilidade

Felipe  
Figueiredo

O desvio  
padrão

Fontes de  
Variabilidade

Visualizando a  
variabilidade com  
histogramas

**Média e a mediana**

Quantificando com  
percentis

Quantificando com  
variância e DP

N ou N-1?

Interpretação do DP

Exercícios

O que acontece com a média, na presença de um valor extremo (muito grande, ou muito pequeno em relação aos outros)?

## Exemplo

O que acontece se você digitar **20** ao invés de **203**?



# Comparação entre as Medidas Centrais



Variabilidade

Felipe  
Figueiredo

O desvio  
padrão

Fontes de  
Variabilidade

Visualizando a  
variabilidade com  
histogramas

**Média e a mediana**

Quantificando com  
percentis

Quantificando com  
variância e DP

N ou N-1?

Interpretação do DP

Exercícios

## Example

Considere o seguinte dataset

$$\{1, 1, 2, 4, 7\}$$

- $N = 5$
- As medidas descritivas centrais para estes dados são:
- $\mu = \frac{1 + 1 + 2 + 4 + 7}{5} = \frac{15}{5} = 3$
- $M_d = 2$

# Comparação entre as Medidas Centrais



Variabilidade

Felipe  
Figueiredo

O desvio  
padrão

Fontes de  
Variabilidade

Visualizando a  
variabilidade com  
histogramas

**Média e a mediana**

Quantificando com  
percentis

Quantificando com  
variância e DP

N ou N-1?

Interpretação do DP

Exercícios

## Example

Considere agora este outro dataset

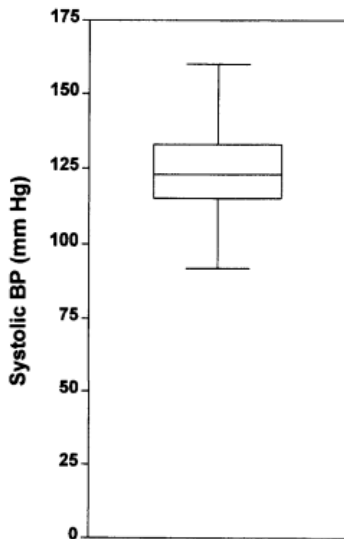
$$\{1, 1, 2, 4, 32\}$$

- $N = 5$
- As medidas descritivas centrais para estes dados são:
- $\mu = \frac{1 + 1 + 2 + 4 + 32}{5} = \frac{40}{5} = 8$
- $M_d = 2$

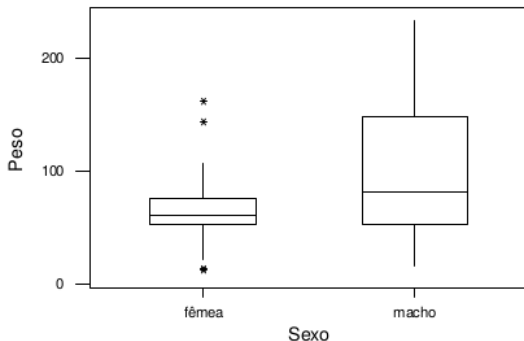
- 1 O desvio padrão
  - Fontes de Variabilidade
  - Visualizando a variabilidade com histogramas
  - Média e a mediana
  - **Quantificando com percentis**
  - Quantificando com variância e DP
  - N ou N-1?
  - Interpretação do DP
  - Exercícios

# O boxplot

- “Caixa e bigodes”
- A caixa representa os percentis de 25% e 75%
- Barra interna que representa a mediana (percentil 50%)
- Barras verticais *indicam* a amplitude dos dados
  - Mínimo e Máximo
  - Regras para “a maioria”



# “Regras para a maioria”



**Figura:** Boxplots para dois grupos de dados (Fonte: Reis, Reis, 2002)

1

## O desvio padrão

- Fontes de Variabilidade
- Visualizando a variabilidade com histogramas
- Média e a mediana
- Quantificando com percentis
- Quantificando com variância e DP
- N ou N-1?
- Interpretação do DP
- Exercícios

# Desvios em relação à média



## Variabilidade

Felipe  
Figueiredo

### O desvio padrão

Fontes de  
Variabilidade

Visualizando a  
variabilidade com  
histogramas

Média e a mediana

Quantificando com  
percentis

Quantificando com  
variância e DP

N ou N-1?

Interpretação do DP

Exercícios

- Uma maneira de entender a variabilidade do dataset é analisar os desvios em relação à média.
- Cada desvio é a diferença entre o valor do dado e a média.

Mas os desvios...

- 1 são tão numerosos quanto os dados
- 2 têm sinal (direção do desvio)
- 3 SEMPRE têm soma **nula**, portanto o desvio médio é sempre 0

Pense...

Uma fórmula que dá o mesmo resultado para qualquer dataset... serve para resumir seus dados?



## Exemplo

$\{1, 2, 3, 4, 5\}$

- $N = 5$

- $\bar{x} = 3$

①  $D_1 = 1 - 3 = -2$

②  $D_2 = 2 - 3 = -1$

③  $D_3 = 3 - 3 = 0$

④  $D_4 = 4 - 3 = 1$

⑤  $D_5 = 5 - 3 = 2$

## O desvio padrão

Fontes de  
Variabilidade

Visualizando a  
variabilidade com  
histogramas

Média e a mediana

Quantificando com  
percentis

Quantificando com  
variância e DP

N ou N-1?

Interpretação do DP

Exercícios

## Exemplo

Somando tudo:

$$\sum D = D_1 + D_2 + D_3 + D_4 + D_5 =$$
$$(-2) + (-1) + 0 + 1 + 2 = 0$$

## Pense...

Uma fórmula que dá o mesmo resultado para qualquer dataset... serve para resumir seus dados?

### O desvio padrão

Fontes de  
Variabilidade

Visualizando a  
variabilidade com  
histogramas

Média e a mediana

Quantificando com  
percentis

Quantificando com  
variância e DP

N ou N-1?

Interpretação do DP

Exercícios

# Como proceder?



Variabilidade

Felipe  
Figueiredo

O desvio  
padrão

Fontes de  
Variabilidade

Visualizando a  
variabilidade com  
histogramas

Média e a mediana

Quantificando com  
percentis

Quantificando com  
variância e DP

N ou N-1?

Interpretação do DP

Exercícios

- Como extrair alguma informação útil (e sumária!) dos desvios?
- Problema: sinais

## Pergunta

Como tirar os sinais dos desvios?

Tomando-se o módulo dos desvios temos:

## Definition

Desvio médio absoluto (MAD) é a média dos desvios absolutos

- É uma medida de dispersão robusta (pouco influenciada por outliers)
- Módulo não tem boas propriedades matemáticas (analíticas e algébricas).
- Pouco usado para inferência (apesar da robustez)

# Desvio médio absoluto (MAD)



Variabilidade

Felipe  
Figueiredo

## Exemplo

$$\{1, 2, 3, 4, 5\}, \bar{x} = 3$$

$$① |D_1| = |1 - 3| = 2$$

$$② |D_2| = |2 - 3| = 1$$

$$③ |D_3| = |3 - 3| = 0$$

$$④ |D_4| = |4 - 3| = 1$$

$$⑤ |D_5| = |5 - 3| = 2$$

$$\text{MAD} = \frac{\sum |D_i|}{5} = \frac{6}{5} = 1.2$$

O desvio  
padrão

Fontes de  
Variabilidade

Visualizando a  
variabilidade com  
histogramas

Média e a mediana

Quantificando com  
percentis

Quantificando com  
variância e DP

N ou N-1?

Interpretação do DP

Exercícios

# Uma proposta “melhor”



Variabilidade

Felipe  
Figueiredo

O desvio  
padrão

Fontes de  
Variabilidade

Visualizando a  
variabilidade com  
histogramas

Média e a mediana

Quantificando com  
percentis

Quantificando com  
variância e DP

N ou N-1?

Interpretação do DP

Exercícios

- Uma outra maneira de eliminar os sinais é elevar ao quadrado cada desvio.
- Preserva boas propriedades matemáticas
- Calculando a média dos quadrados dos desvios (desvios quadráticos) temos . . .

## Definition

A variância é a média dos desvios quadráticos.

- Variância populacional

$$\sigma^2 = \frac{\sum (x_j - \mu)^2}{N}$$

- Variância amostral

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

- Conveniente do ponto de vista matemático (boas propriedades algébricas e analíticas).
- Unidade quadrática, pouco intuitiva para interpretação de resultados.

## Variabilidade

Felipe  
Figueiredo

### O desvio padrão

Fontes de  
Variabilidade

Visualizando a  
variabilidade com  
histogramas

Média e a mediana

Quantificando com  
percentis

Quantificando com  
variância e DP

N ou N-1?

Interpretação do DP

Exercícios

## Exemplo

$$\{1, 2, 3, 4, 5\}, \bar{x} = 3$$

$$① D_1^2 = (1 - 3)^2 = (-2)^2 = 4$$

$$② D_2^2 = (2 - 3)^2 = (-1)^2 = 1$$

$$③ D_3^2 = (3 - 3)^2 = 0^2 = 0$$

$$④ D_4^2 = (4 - 3)^2 = 1^2 = 1$$

$$⑤ D_5^2 = (5 - 3)^2 = 2^2 = 4$$

$$s^2 = \frac{\sum D_i^2}{4} = 2.5$$

## O desvio padrão

Fontes de  
Variabilidade

Visualizando a  
variabilidade com  
histogramas

Média e a mediana

Quantificando com  
percentis

Quantificando com  
variância e DP

N ou N-1?

Interpretação do DP

Exercícios



## Definition

O desvio padrão é a raiz quadrada da variância.

- Desvio padrão populacional

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

- Desvio padrão amostral

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

## O desvio padrão

Fontes de Variabilidade

Visualizando a variabilidade com histogramas

Média e a mediana

Quantificando com percentis

Quantificando com variância e DP

N ou N-1?

Interpretação do DP

Exercícios

- É a medida mais usada, por estar na mesma escala (unidade) dos dados.
- Boas propriedades matemáticas
- Boas propriedades como estimador (Inferência)

## Example

$$\{1, 2, 3, 4, 5\}, \bar{x} = 3$$

$$s^2 = 2.5$$

$$s = \sqrt{s^2} = \sqrt{2.5} = 1.58$$

1

## O desvio padrão

- Fontes de Variabilidade
- Visualizando a variabilidade com histogramas
- Média e a mediana
- Quantificando com percentis
- Quantificando com variância e DP
- **N ou N-1?**
- Interpretação do DP
- Exercícios

# N ou N-1?



Variabilidade

Felipe  
Figueiredo

O desvio  
padrão

Fontes de  
Variabilidade

Visualizando a  
variabilidade com  
histogramas

Média e a mediana

Quantificando com  
percentis

Quantificando com  
variância e DP

**N ou N-1?**

Interpretação do DP

Exercícios

## Fórmula com N

Usada apenas para cálculos com dados de toda a população.

## Fórmula com N-1

Usada para cálculos com dados de uma amostra.

## Pense...

Você tem acesso a toda a população, ou apenas a uma amostra?

1

## O desvio padrão

- Fontes de Variabilidade
- Visualizando a variabilidade com histogramas
- Média e a mediana
- Quantificando com percentis
- Quantificando com variância e DP
- N ou N-1?
- **Interpretação do DP**
- Exercícios

*“Um pouco mais da metade” dos valores está a 1 DP da média (considerando ambos os lados)*

*“Quase todos” os dados estão a 2 DP da média (considerando ambos os lados)*

- *Cenas dos próximos capítulos*

1

## O desvio padrão

- Fontes de Variabilidade
- Visualizando a variabilidade com histogramas
- Média e a mediana
- Quantificando com percentis
- Quantificando com variância e DP
- N ou N-1?
- Interpretação do DP
- Exercícios



## Leitura obrigatória

### Capítulo 3.

Pular as seções:

- Calculando o DP numa calculadora
- Coeficiente de Variação (CV)

- Exercício 1
- Exercício 2
- Exercício 3 (R: 34.64503)
- Exercício 4 (R: 219.4131)
- Exercício 5