

Variabilidade

Felipe Figueiredo

o desvid Dadrão

## Variabilidade

Incertezas de dados numéricos

Felipe Figueiredo



Variabilidade

Felipe Figueiredo

O desvio padrão

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



Variabilidade

#### Felipe Figueiredo

- 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.

#### O desvio padrão

### Medidas Sumárias



Variabilidade

Felipe Figueiredo

#### O desvio padrão

### Tipos de medidas sumárias

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

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



Variabilidade

Felipe Figueiredo

padrão Fontes de

Fontes de Variabilidade

Visualizando a variabilidade com

histogramas Média e a median

percentis

Quantificando con

variância e DP N ou N-1?

Interpretação do E

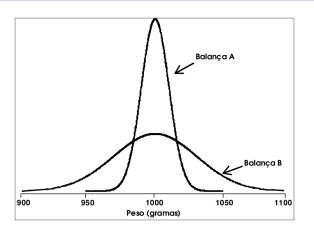
Figue

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

## Variabilidade em Medições





Variabilidade Felipe Figueiredo

desvio adrão

Fontes de Variabilidade

Visualizando a variabilidade com histogramas

Média e a mediana Quantificando com percentis

> Quantificando com rariância e DP

terpretação do DI

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



### Fontes comuns de variabilidade



Variabilidade

#### Felipe Figueiredo

O desvio padrão

#### Fontes de Variabilidade

Visualizando a variabilidade com histogramas

histogramas Média e a mediana

Quantificando com percentis

Quantificando com variância e DP N ou N-12

Interpretação do D

### 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).



Variabilidade

#### Felipe Figueiredo

padrão

Variabilidad

Visualizando a variabilidade com histogramas

Média e a mediana Quantificando con

Quantificando com rariância e DP

N ou N-1? nterpretação do D

Interpretação do E Exercícios

- 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



#### Variabilidade

## Figueiredo

Visualizando a variabilidade com histogramas

Felipe

# 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.

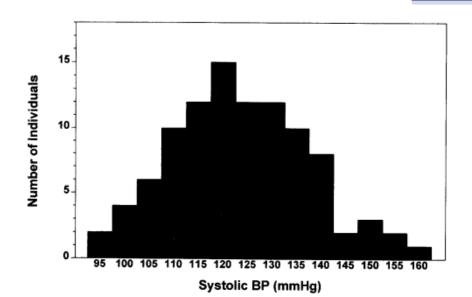
### Pergunta

Exemplo

Como "entender" essa listagem de 100 números?

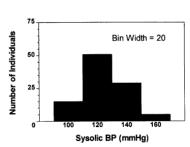
## O histograma

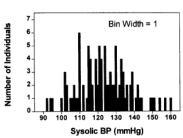




### Quantas barras?







#### Variabilidade

#### Felipe Figueiredo

O desvio padrão

Fontes de Variabilidade

Visualizando a variabilidade com histogramas

Média e a mediana Quantificando com percentis

Quantificando con variância e DP

Interpretação do D



Variabilidade

#### Felipe Figueiredo

O desvio padrão

Variabilidade

Visualizando a variabilidade con histogramas

#### Média e a mediana

Quantificando con percentis

Quantificando cor variância e DP

N ou N-1? Interpretação do E

Interpretação do D Exercícios

## 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

### Média



#### Variabilidade

#### Felipe Figueiredo

O desvio padrão

Variabilidade Visualizando a

'isualizando a ariabilidade com istogramas

#### Média e a mediana

percentis
Quantificando com

Quantificando com variância e DP N ou N-1?

Interpretação do E Exercícios

### Exemplo

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

 $x_1 = 142$ 

 $x_2 = 144$ 

 $x_3 = 176$ 

 $x_4 = 203$ 

 $x_5 = 134$ 

 $x_6 = 191$ 

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

### Percentis e a Mediana



#### Variabilidade

#### Felipe Figueiredo

## O desvio padrão

Fontes de Variabilida

Visualizando a variabilidade com

#### Média e a mediana

#### Quantificando com percentis

Quantificando com variância e DP

ou N-1?

Interpretação do D 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 impar.
- Encontrar a média dos dois valores do meio se n for par.

### Mediana



#### Variabilidade

#### Felipe Figueiredo

### Exemplo

Conforme no exemplo anterior

 $x_5 = 134$ 

 $x_1 = 142$ 

 $x_2 = 144$ 

 $x_3 = 176$ 

 $x_6 = 191$ 

 $x_4 = 203$ 

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

## O desvio padrão

Fontes de

Variabilidade Visualizando a

ariabilidade com istogramas

Média e a mediana

percentis

Quantificando com

Quantificando com variância e DP

Interpretação do DI

## Qual é a diferença?



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?

Variabilidade

Felipe Figueiredo

O desvio padrão

Variabilidade Visualizando a

Visualizando a variabilidade com pistogramas

Média e a mediana

Quantificando com percentis Quantificando com

Quantificando con variância e DP N ou N-1?

Interpretação do D

# Comparação entre as Medidas Centrais



### Example

Considere o seguinte dataset

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

- N = 5
- As medidas descritivas centrais para estes dados são:

$$\bullet \ \mu = \frac{1+1+2+4+7}{5} = \frac{15}{5} = 3$$

Variabilidade

#### Felipe Figueiredo

O desvio padrão

Fontes de Variabilidade

/isualizando a variabilidade com

nistogramas Média e a mediana

Quantificando com percentis

Quantificando com rariância e DP

Interpretação do D Exercícios

# Comparação entre as Medidas Centrais



## Variabilidade

#### Felipe Figueiredo

O desvio padrão

Fontes de Variabilidad

/isualizando a variabilidade com

Média e a mediana

Quantificando com ercentis Quantificando com

luantificando com ariância e DP

Interpretação do D Exercícios

### Example

Considere agora este outro dataset

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

- N = 5
- As medidas descritivas centrais para estes dados são:

$$\bullet \ \mu = \frac{1+1+2+4+32}{5} = \frac{40}{5} = 8$$



Variabilidade

## Figueiredo

Quantificando com

## percentis

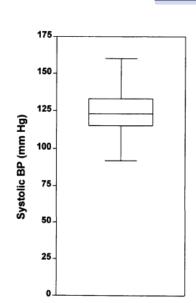
### Felipe

- 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

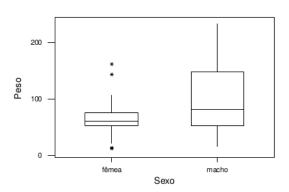


- "Caixa e bigodes"
- A caixa representa os percentis de 25% e 75%
- Barra interna que representa a mediana (percentil 50%)
- Barras verticais indicam a amplitude dos dados
  - Mínimo e Máximo
  - Regras para "a maioria"



# "Regras para a maioria"





Variabilidade

#### Felipe Figueiredo

O desvio padrão

Fontes de Variabilidade

Visualizando a variabilidade com

nistogramas Média e a mediana

Quantificando com percentis

Quantificando com

auantificando com rariância e DP V ou N-1?

Interpretação do Di Exercícios

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



Variabilidade

#### Felipe Figueiredo

#### Quantificando com variância e DP

## 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

percentis

Quantificando com variância e DP

N ou N-1?

Interpretação do D

- 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.

# Desvios em relação à média



#### Variabilidade

#### Felipe Figueiredo

#### Quantificando com variância e DP

### Mas os desvios...

- são tão numerosos quanto os dados
- 2 têm sinal (direção do desvio)
- 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?

# Desvios em relação à média



### Exemplo

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

• 
$$N = 5$$

$$\bar{x}=3$$

$$\mathbf{0} \ D_1 = 1 - 3 = -2$$

$$Q D_2 = 2 - 3 = -1$$

$$O_3 = 3 - 3 = 0$$

$$\mathbf{Q} D_4 = 4 - 3 = 1$$

**6** 
$$D_5 = 5 - 3 = 2$$

#### Variabilidade

#### Felipe Figueiredo

O desvio padrão

Fontes de Variabilidade

> Visualizando a variabilidade com histogramas

histogramas Média e a mediana

percentis Quantificando com

variância e DP N ou N-1?

Interpretação do DF Exercícios

### Soma dos desvios



### 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?

#### Variabilidade

#### Felipe Figueiredo

O desvio padrão

Variabilidade

Visualizando a variabilidade com

Média e a mediana

Quantificando com percentis

Quantificando com variância e DP

N ou N-1? Interpreta

Interpretação do D Exercícios

## Como proceder?



Variabilidade

#### Felipe Figueiredo

Quantificando com variância e DP

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

### Pergunta

Como tirar os sinais dos desvios?

### Desvios absolutos



Variabilidade

Felipe Figueiredo

O desvio padrão

Fontes de

sualizando a riabilidade com stogramas

Média e a median Quantificando cor

percentis Quantificando com

Quantificando cor variância e DP

N ou N-1?

Interpretação do D Exercícios

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)



### Exemplo

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

$$|D_1| = |1 - 3| = 2$$

$$|D_2| = |2-3| = 1$$

$$|D_3| = |3-3| = 0$$

$$|D_4| = |4-3| = 1$$

**6** 
$$|D_5| = |5-3| = 2$$

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

#### Variabilidade

#### Felipe Figueiredo

O desvio padrão

Fontes de Variabilidade

isualizando a ariabilidade com istogramas

Média e a mediana

Quantificando com

variância e DP

Interpretação do DI

## Uma proposta "melhor"



#### Variabilidade

#### Felipe Figueiredo

O desvio padrão

Fontes de Variabilidade

Visualizando a variabilidade com

Média e a median

Quantificando co percentis

### Quantificando com

V ou N-1

Interpretação do D

- 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 ...

### Variância



### 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

histogramas

Média e a mediana

Média e a mediana Quantificando com percentis

Quantificando com variância e DP

I ou N-1?

Interpretação do DF Exercícios

## Variância



### Exemplo

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

$$D_1^2 = (1-3)^2 = (-2)^2 = 4$$

$$2 D_2^2 = (2-3)^2 = (-1)^2 = 1$$

$$D_3^2 = (3-3)^2 = 0^2 = 0$$

**6** 
$$D_5^2 = (5-3)^2 = 2^2 = 4$$

### Variabilidade

#### Felipe Figueiredo

O desvio padrão

Fontes de Variabilidade

Visualizando a variabilidade com nistogramas

Média e a mediana

Quantificando com

percentis Quantificando com

variância e DP N ou N-1?

Exercícios

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

### Desvio Padrão



#### 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

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

Variabilidade

#### Felipe Figueiredo

O desvio padrão

Fontes de

Visualizando a variabilidade com

Média e a mediana

percentis
Quantificando com

Quantificando co variância e DP

N ou N-1

Interpretação do DE

### Desvio Padrão



#### Variabilidade

#### Felipe Figueiredo

O desvio padrão

Variabilidade

/isualizando a rariabilidade com

nistogramas Média e a mediana

Quantificando com percentis

Quantificando com

**ariância e l** Lou N-1?

Interpretação do D

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

### Desvio Padrão



#### Variabilidade

#### Felipe Figueiredo

## O desvio padrão

Fontes de Variabilidade

Visualizando a

histogramas Média e a mediana

Quantificando com percentis

#### Quantificando com variância e DP

ariância e DP

Interpretação do D

### Example

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

$$s^2 = 2.5$$

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



Variabilidade

#### Felipe Figueiredo

N ou N-12

## 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

Visualizando a variabilidade com

histogramas Média e a mediana

percentis
Quantificando com

Quantificando com variância e DP

N ou N-1?

Interpretação do I 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?



Variabilidade

Figueiredo

Interpretação do DP

Felipe

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

## Interpretação do DP



Variabilidade

Felipe Figueiredo

"Um pouco mais da metade" dos valores está a 1 DP da média (considerando amdos os lados)

"Quase todos" os dados estão a 2 DP da média (considerando ambos os lados)

Cenas dos próximos capítulos

O desvio padrão

Fontes de

Variabilidade Visualizando a

histogramas

Média e a mediana

percentis

Quantificando com

Quantificando com variância e DP N ou N-1?

Interpretação do DP



Variabilidade

Figueiredo

Exercícios

Felipe

## 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 pós-aula e exercícios selecionados



### 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

Variabilidade

Felipe Figueiredo

O desvio padrão

Fontes de

Visualizando a variabilidade con

histogramas Média e a mediana

Quantificando com percentis Quantificando com

Quantificando com variância e DP N ou N-1?

Interpretação do DI Exercícios