

Estatística Descritiva II

Medidas sumárias

Felipe Figueiredo

Instituto Nacional de Traumatologia e Ortopedia

- 1 Medidas de Tendência Central
 - Média
 - Mediana
 - Moda
 - Comparação entre as Medidas Centrais
- 2 Medidas de Dispersão
 - Amplitude
 - Desvios em relação à media
 - Variância
 - Desvio Padrão
 - Exercícios
 - Coeficiente de Variação
- 3 Medidas de Posição
 - Quartis
 - Percentis e Decis
- 4 Boxplot
- 5 Resumo

Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Medidas de
Posição

Boxplot

Resumo

- 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 (μ , σ , etc).
- Medidas sumárias de **amostras** se chamam **estatísticas** e são representadas por letras comuns (\bar{x} , s , etc).
- Geralmente trabalhamos com estatísticas descritivas.

- 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 (μ , σ , etc).
- Medidas sumárias de **amostras** se chamam **estatísticas** e são representadas por letras comuns (\bar{x} , s , etc).
- Geralmente trabalhamos com estatísticas descritivas.

- 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 (μ , σ , etc).
- Medidas sumárias de **amostras** se chamam **estatísticas** e são representadas por letras comuns (\bar{x} , s , etc).
- Geralmente trabalhamos com estatísticas descritivas.

- 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 (μ , σ , etc).
- Medidas sumárias de **amostras** se chamam **estatísticas** e são representadas por letras comuns (\bar{x} , s , etc).
- Geralmente trabalhamos com estatísticas descritivas.

- 1 Medidas de Tendência Central
 - Média
 - Mediana
 - Moda
 - Comparação entre as Medidas Centrais
- 2 Medidas de Dispersão
 - Amplitude
 - Desvios em relação à media
 - Variância
 - Desvio Padrão
 - Exercícios
 - Coeficiente de Variação
- 3 Medidas de Posição
 - Quartis
 - Percentis e Decis
- 4 Boxplot
- 5 Resumo

- A média (aritmética) leva em conta todos os dados disponíveis, e indica (em muitas situações) o ponto de maior acumulação de dados.

- Notação: média populacional (μ)

$$\mu = \sum_{j=1}^N \frac{x_j}{N}$$

- Notação: média amostral (\bar{x})

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

- Nem sempre pertence ao dataset.

- A média (aritmética) leva em conta todos os dados disponíveis, e indica (em muitas situações) o ponto de maior acumulação de dados.
- Notação: média populacional (μ)

$$\mu = \sum_{j=1}^N \frac{x_j}{N}$$

- Notação: média amostral (\bar{x})

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

- Nem sempre pertence ao dataset.

- A média (aritmética) leva em conta todos os dados disponíveis, e indica (em muitas situações) o ponto de maior acumulação de dados.
- Notação: média populacional (μ)

$$\mu = \sum_{j=1}^N \frac{x_j}{N}$$

- Notação: média amostral (\bar{x})

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

- Nem sempre pertence ao dataset.

- A média (aritmética) leva em conta todos os dados disponíveis, e indica (em muitas situações) o ponto de maior acumulação de dados.
- Notação: média populacional (μ)

$$\mu = \sum_{j=1}^N \frac{x_j}{N}$$

- Notação: média amostral (\bar{x})

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

- Nem sempre pertence ao dataset.

Example

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

$$x_1 = 142$$

$$x_2 = 144$$

$$x_3 = 176$$

$$x_4 = 203$$

$$x_5 = 134$$

$$x_6 = 191$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{990}{6} = 165$$

- 1 Medidas de Tendência Central
 - Média
 - **Mediana**
 - Moda
 - Comparação entre as Medidas Centrais
- 2 Medidas de Dispersão
 - Amplitude
 - Desvios em relação à media
 - Variância
 - Desvio Padrão
 - Exercícios
 - Coeficiente de Variação
- 3 Medidas de Posição
 - Quartis
 - Percentis e Decis
- 4 Boxplot
- 5 Resumo

Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Média

Mediana

Moda

Comparação

Medidas de
Dispersão

Medidas de
Posição

Boxplot

Resumo

Definition

A mediana é o dado que ocupa a **posição central** nos dados ordenados.

- Notação: M_d
- Divide o dataset ao meio
- Costuma pertencer ao dataset

Definition

A mediana é o dado que ocupa a **posição central** nos dados ordenados.

- Notação: M_d
- Divide o dataset ao meio
- Costuma pertencer ao dataset

Definition

A mediana é o dado que ocupa a **posição central** nos dados ordenados.

- Notação: M_d
- Divide o dataset ao meio
- Costuma pertencer ao dataset

Definition

A mediana é o dado que ocupa a **posição central** nos dados ordenados.

- Notação: M_d
- Divide o dataset ao meio
- Costuma pertencer ao dataset

- 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.

Example

Conforme no exemplo anterior

- 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.

Example

Conforme no exemplo anterior

- 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.

Example

Conforme no exemplo anterior

- 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.

Example

Conforme no exemplo anterior

$$x_5 = 134$$

$$x_1 = 142$$

$$x_2 = 144$$

$$x_3 = 176$$

$$x_6 = 191$$

$$x_4 = 203$$

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

- 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.

Example

Conforme no exemplo anterior

$$x_5 = 134$$

$$x_1 = 142$$

$$x_2 = 144$$

$$x_3 = 176$$

$$x_6 = 191$$

$$x_4 = 203$$

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

- 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.

Example

Conforme no exemplo anterior

$$x_5 = 134$$

$$x_1 = 142$$

$$x_2 = 144$$

$$x_3 = 176$$

$$x_6 = 191$$

$$x_4 = 203$$

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

- 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.

Example

Conforme no exemplo anterior

$$x_5 = 134$$

$$x_1 = 142$$

$$x_2 = 144$$

$$x_3 = 176$$

$$x_6 = 191$$

$$x_4 = 203$$

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

- 1 Medidas de Tendência Central
 - Média
 - Mediana
 - **Moda**
 - Comparação entre as Medidas Centrais
- 2 Medidas de Dispersão
 - Amplitude
 - Desvios em relação à media
 - Variância
 - Desvio Padrão
 - Exercícios
 - Coeficiente de Variação
- 3 Medidas de Posição
 - Quartis
 - Percentis e Decis
- 4 Boxplot
- 5 Resumo

Definition

A moda é o dado que ocorre com **maior frequência**.

- Notação: M_o
- Sempre pertence ao dataset.
- Não é necessariamente única: o dataset pode ser **bimodal**, ou mesmo **multimodal**.
- Não necessariamente existe: **amodal**

Definition

A moda é o dado que ocorre com **maior frequência**.

- Notação: M_o
- Sempre pertence ao dataset.
- Não é necessariamente única: o dataset pode ser **bimodal**, ou mesmo **multimodal**.
- Não necessariamente existe: **amodal**

Definition

A moda é o dado que ocorre com **maior frequência**.

- Notação: M_o
- Sempre pertence ao dataset.
- Não é necessariamente única: o dataset pode ser **bimodal**, ou mesmo **multimodal**.
- Não necessariamente existe: **amodal**

Definition

A moda é o dado que ocorre com **maior frequência**.

- Notação: M_o
- Sempre pertence ao dataset.
- Não é necessariamente única: o dataset pode ser **bimodal**, ou mesmo **multimodal**.
- Não necessariamente existe: **amodal**

Definition

A moda é o dado que ocorre com **maior frequência**.

- Notação: M_o
- Sempre pertence ao dataset.
- Não é necessariamente única: o dataset pode ser **bimodal**, ou mesmo **multimodal**.
- Não necessariamente existe: **amodal**

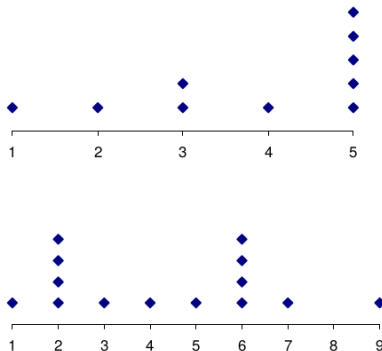
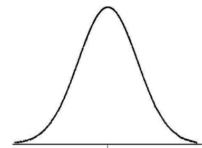


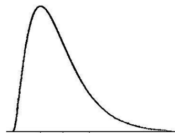
Figura: Diagrama de pontos para dados (a) unimodal, (b) bimodal
(Fonte: Reis, Reis, 2002)

- 1 **Medidas de Tendência Central**
 - Média
 - Mediana
 - Moda
 - **Comparação entre as Medidas Centrais**
- 2 **Medidas de Dispersão**
 - Amplitude
 - Desvios em relação à media
 - Variância
 - Desvio Padrão
 - Exercícios
 - Coeficiente de Variação
- 3 **Medidas de Posição**
 - Quartis
 - Percentis e Decis
- 4 **Boxplot**
- 5 **Resumo**

Comparação entre as Medidas Centrais



moda = mediana = média



moda < mediana < média



média < mediana < moda

Figura: (a) Simétrica, (b) Assimétrica à esquerda, (c) Assimétrica à direita (Fonte: Reis, Reis 2002)

- A média é mais usada, mas não é **robusta**.
- É distorcida na presença de *outliers* (valores discrepantes, extremos)

Comparação entre as Medidas Centrais



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Média

Mediana

Moda

Comparação

Medidas de
Dispersão

Medidas de
Posição

Boxplot

Resumo

Example

Considere o seguinte dataset

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

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

Comparação entre as Medidas Centrais



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Média

Mediana

Moda

Comparação

Medidas de
Dispersão

Medidas de
Posição

Boxplot

Resumo

Example

Considere o seguinte dataset

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

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

Example

Considere o seguinte dataset

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

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

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

- $M_d = 2$

- $M_o = 1$

Example

Considere o seguinte dataset

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

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

Comparação entre as Medidas Centrais



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Média

Mediana

Moda

Comparação

Medidas de
Dispersão

Medidas de
Posição

Boxplot

Resumo

Example

Considere o seguinte dataset

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

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

Example

Considere o seguinte dataset

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

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

Comparação entre as Medidas Centrais



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Média

Mediana

Moda

Comparação

Medidas de
Dispersão

Medidas de
Posição

Boxplot

Resumo

Example

Considere agora este outro dataset

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

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

Example

Considere agora este outro dataset

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

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

Example

Considere agora este outro dataset

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

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

Comparação entre as Medidas Centrais



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Média

Mediana

Moda

Comparação

Medidas de
Dispersão

Medidas de
Posição

Boxplot

Resumo

Example

Considere agora este outro dataset

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

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

Comparação entre as Medidas Centrais



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Média

Mediana

Moda

Comparação

Medidas de
Dispersão

Medidas de
Posição

Boxplot

Resumo

Example

Considere agora este outro dataset

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

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

Example

Considere agora este outro dataset

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

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

Exercício

Um pesquisador observou as seguintes idades (anos) para uma amostra: 35, 33, 37, 33, 34.

Determine:

- 1 A média amostral (\bar{x})
- 2 A mediana (M_d)
- 3 A moda (M_o)

Solução

1 $\bar{x} = \frac{35 + 33 + 37 + 33 + 34}{5} = 34.4$

2 $M_d = 34$

3 $M_o = 33$

Exercício

Um pesquisador observou as seguintes idades (anos) para uma amostra: 35, 33, 37, 33, 34.

Determine:

- 1 A média amostral (\bar{x})
- 2 A mediana (M_d)
- 3 A moda (M_o)

Solução

- 1 $\bar{x} = \frac{35 + 33 + 37 + 33 + 34}{5} = 34.4$
- 2 $M_d = 34$
- 3 $M_o = 33$

Exercício

Um pesquisador observou as seguintes idades (anos) para uma amostra: 35, 33, 37, 33, 34.

Determine:

- 1 A média amostral (\bar{x})
- 2 A mediana (M_d)
- 3 A moda (M_o)

Solução

1 $\bar{x} = \frac{35 + 33 + 37 + 33 + 34}{5} = 34.4$

2 $M_d = 34$

3 $M_o = 33$

Exercício

Um pesquisador observou as seguintes idades (anos) para uma amostra: 35, 33, 37, 33, 34.

Determine:

- 1 A média amostral (\bar{x})
- 2 A mediana (M_d)
- 3 A moda (M_o)

Solução

- 1 $\bar{x} = \frac{35 + 33 + 37 + 33 + 34}{5} = 34.4$
- 2 $M_d = 34$
- 3 $M_o = 33$

Exercício

Um pesquisador observou as seguintes idades (anos) para uma amostra: 35, 33, 37, 33, 34.

Determine:

- 1 A média amostral (\bar{x})
- 2 A mediana (M_d)
- 3 A moda (M_o)

Solução

- 1 $\bar{x} = \frac{35 + 33 + 37 + 33 + 34}{5} = 34.4$
- 2 $M_d = 34$
- 3 $M_o = 33$

- Média mais usual
- Mediana na presença de *outliers*
- Moda quando a distribuição das frequências for bimodal ou multimodal.

- Média mais usual
- Mediana na presença de *outliers*
- Moda quando a distribuição das frequências for bimodal ou multimodal.

- Média mais usual
- Mediana na presença de *outliers*
- Moda quando a distribuição das frequências for bimodal ou multimodal.

Variabilidade em Medições

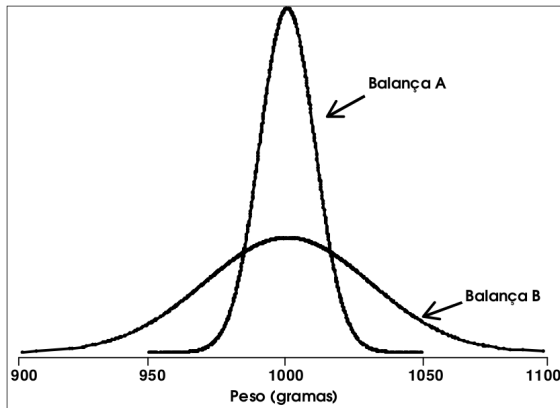


Figura: Variabilidade da medição de uma esfera metálica de 1000g. Balança A, “imprecisão” de 50g, balança B, “imprecisão” de 100g (Fonte: Reis, Reis, 2002)

- 1 Medidas de Tendência Central
 - Média
 - Mediana
 - Moda
 - Comparação entre as Medidas Centrais
- 2 Medidas de Dispersão
 - **Amplitude**
 - Desvios em relação à média
 - Variância
 - Desvio Padrão
 - Exercícios
 - Coeficiente de Variação
- 3 Medidas de Posição
 - Quartis
 - Percentis e Decis

4 Boxplot

5 Resumo

A amplitude dos dados identifica o intervalo de ocorrência de todos os dados observados

- $A = x_{max} - x_{min}$

Example

Seja o dataset {21, 12, 20, 4, 75, 40, 39, 63} Então, a amplitude é: $A = 75 - 4 = 71$

A amplitude dos dados identifica o intervalo de ocorrência de todos os dados observados

- $A = x_{max} - x_{min}$

Example

Seja o dataset {21, 12, 20, 4, 75, 40, 39, 63} Então, a amplitude é: $A = 75 - 4 = 71$

- 1 Medidas de Tendência Central
 - Média
 - Mediana
 - Moda
 - Comparação entre as Medidas Centrais
- 2 Medidas de Dispersão
 - Amplitude
 - Desvios em relação à média
 - Variância
 - Desvio Padrão
 - Exercícios
 - Coeficiente de Variação
- 3 Medidas de Posição
 - Quartis
 - Percentis e Decis
- 4 Boxplot
- 5 Resumo

Desvios em relação à média



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

Desvios em relação
à média

Variância

Desvio Padrão

Exercícios

Coeficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

- 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



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

Desvios em relação
à média

Variância

Desvio Padrão

Exercícios

Coeficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

- 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



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

Desvios em relação
à média

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

Mas os desvios...

- 1 são tão numerosos quanto os dados
- 2 têm sinal (direção do desvio)
- 3 têm soma **nula**

Desvios em relação à média



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

Desvios em relação
à média

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

Mas os desvios...

- 1 são tão numerosos quanto os dados
- 2 têm sinal (direção do desvio)
- 3 têm soma **nula**

Desvios em relação à média



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

Desvios em relação
à média

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

Mas os desvios...

- 1 são tão numerosos quanto os dados
- 2 têm sinal (direção do desvio)
- 3 têm soma **nula**

Desvios em relação à média



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

**Desvios em relação
à média**

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

Example

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

Desvios em relação à média



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

**Desvios em relação
à média**

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

Example

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

- $N = 5$

- $\bar{x} = 3$

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 $D_5 = 5 - 3 = 2$

Desvios em relação à média



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

**Desvios em relação
à média**

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

Example

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

- $N = 5$
- $\bar{x} = 3$

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 $D_5 = 5 - 3 = 2$

Desvios em relação à média



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

**Desvios em relação
à média**

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

Example

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

- $N = 5$

- $\bar{x} = 3$

- 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 $D_5 = 5 - 3 = 2$

Desvios em relação à média



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

Desvios em relação
à média

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

Example

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

- $N = 5$
- $\bar{x} = 3$

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 $D_5 = 5 - 3 = 2$

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

- $N = 5$
- $\bar{x} = 3$

$$\textcircled{1} \quad D_1 = 1 - 3 = -2$$

$$\textcircled{2} \quad D_2 = 2 - 3 = -1$$

$$\textcircled{3} \quad D_3 = 3 - 3 = 0$$

④ $D_4 = 4 - 3 = 1$

⑤ $D_5 = 5 - 3 = 2$

Medidas de Dispersão

Amplitude

Desvios em relação
à média

Variância

Desvio Padrão

Exercícios

Coeficiente de

Variação

Medidas de Posição

Boxplot

Resumo

Desvios em relação à média



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

Desvios em relação
à média

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

Example

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

- $N = 5$
- $\bar{x} = 3$

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 $D_5 = 5 - 3 = 2$

Desvios em relação à média



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

Desvios em relação
à média

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

Example

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

- $N = 5$
- $\bar{x} = 3$

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 $D_5 = 5 - 3 = 2$

Desvios em relação à média



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

Desvios em relação
à média

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

Example

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

- $N = 5$
- $\bar{x} = 3$

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 $D_5 = 5 - 3 = 2$

Example

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

- $N = 5$
- $\bar{x} = 3$

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 $D_5 = 5 - 3 = 2$

Example

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

- $N = 5$
- $\bar{x} = 3$

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 $D_5 = 5 - 3 = 2$

Example

Somando tudo:

$$\begin{aligned} D_1 + D_2 + D_3 + D_4 + D_5 = \\ (-2) + (-1) + 0 + 1 + 2 = 0 \end{aligned}$$

Como proceder?



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

Desvios em relação
à média

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

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

Pergunta

Como tirar os sinais dos desvios?

Como proceder?



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

Desvios em relação
à média

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

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

Pergunta

Como tirar os sinais dos desvios?

Como proceder?



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

Desvios em relação
à média

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

- 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 (baixa eficiência)

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 (baixa eficiência)

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 (baixa eficiência)

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 (baixa eficiência)

Desvio médio absoluto (MAD)

Example

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

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

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

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

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

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

$$MAD = \frac{\sum D_i}{5} = 2$$

Desvio médio absoluto (MAD)

Example

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

$$① |D_1| = |1 - 3| = 2$$

$$② |D_2| = |2 - 3| = 1$$

$$③ |D_3| = |3 - 3| = 0$$

$$④ |D_4| = |4 - 3| = 1$$

$$⑤ |D_5| = |5 - 3| = 2$$

$$\text{MAD} = \frac{\sum D_i}{5} = 2$$

Desvio médio absoluto (MAD)

Example

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

$$① |D_1| = |1 - 3| = 2$$

$$② |D_2| = |2 - 3| = 1$$

$$③ |D_3| = |3 - 3| = 0$$

$$④ |D_4| = |4 - 3| = 1$$

$$⑤ |D_5| = |5 - 3| = 2$$

$$\text{MAD} = \frac{\sum D_i}{5} = 2$$

Desvio médio absoluto (MAD)

Example

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

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

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

$$3 \quad |D_3| = |3 - 3| = 0$$

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

$$5 \quad |D_5| = |5 - 3| = 2$$

$$\text{MAD} = \frac{\sum D_i}{5} = 2$$

Desvio médio absoluto (MAD)



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

Desvios em relação
à média

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

Example

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

$$① |D_1| = |1 - 3| = 2$$

$$② |D_2| = |2 - 3| = 1$$

$$③ |D_3| = |3 - 3| = 0$$

$$④ |D_4| = |4 - 3| = 1$$

$$⑤ |D_5| = |5 - 3| = 2$$

$$\text{MAD} = \frac{\sum D_i}{5} = 2$$

Desvio médio absoluto (MAD)



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

Desvios em relação
à média

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

Example

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

$$① |D_1| = |1 - 3| = 2$$

$$② |D_2| = |2 - 3| = 1$$

$$③ |D_3| = |3 - 3| = 0$$

$$④ |D_4| = |4 - 3| = 1$$

$$⑤ |D_5| = |5 - 3| = 2$$

$$\text{MAD} = \frac{\sum D_i}{5} = 2$$

Desvio médio absoluto (MAD)



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

Desvios em relação
à média

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

Example

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

$$① |D_1| = |1 - 3| = 2$$

$$② |D_2| = |2 - 3| = 1$$

$$③ |D_3| = |3 - 3| = 0$$

$$④ |D_4| = |4 - 3| = 1$$

$$⑤ |D_5| = |5 - 3| = 2$$

$$\text{MAD} = \frac{\sum D_i}{5} = 2$$

Desvio médio absoluto (MAD)



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

Desvios em relação
à média

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

Example

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

$$① |D_1| = |1 - 3| = 2$$

$$② |D_2| = |2 - 3| = 1$$

$$③ |D_3| = |3 - 3| = 0$$

$$④ |D_4| = |4 - 3| = 1$$

$$⑤ |D_5| = |5 - 3| = 2$$

$$\text{MAD} = \frac{\sum D_i}{5} = 2$$

Uma proposta “melhor”



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

Desvios em relação
à média

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

- 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 ...

Uma proposta “melhor”



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

Desvios em relação
à média

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

- 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 ...

Uma proposta “melhor”



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude

Desvios em relação
à média

Variância

Desvio Padrão

Exercícios

Coefficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

- 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 . . .

- 1 Medidas de Tendência Central
 - Média
 - Mediana
 - Moda
 - Comparação entre as Medidas Centrais
- 2 Medidas de Dispersão
 - Amplitude
 - Desvios em relação à média
 - **Variância**
 - Desvio Padrão
 - Exercícios
 - Coeficiente de Variação
- 3 Medidas de Posição
 - Quartis
 - Percentis e Decis
- 4 Boxplot
- 5 Resumo

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.

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.

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.

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.

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.

Example

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

$$\textcircled{1} D_1^2 = (1 - 3)^2 = (-2)^2 = 4$$

$$\textcircled{2} D_2^2 = (2 - 3)^2 = (-1)^2 = 1$$

$$\textcircled{3} D_3^2 = (3 - 3)^2 = 0^2 = 0$$

$$\textcircled{4} D_4^2 = (4 - 3)^2 = 1^2 = 1$$

$$\textcircled{5} D_5^2 = (5 - 3)^2 = 2^2 = 4$$

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

Example

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

$$① D_1^2 = (1 - 3)^2 = (-2)^2 = 4$$

$$② D_2^2 = (2 - 3)^2 = (-1)^2 = 1$$

$$③ D_3^2 = (3 - 3)^2 = 0^2 = 0$$

$$④ D_4^2 = (4 - 3)^2 = 1^2 = 1$$

$$⑤ D_5^2 = (5 - 3)^2 = 2^2 = 4$$

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

Example

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

$$① D_1^2 = (1 - 3)^2 = (-2)^2 = 4$$

$$② D_2^2 = (2 - 3)^2 = (-1)^2 = 1$$

$$③ D_3^2 = (3 - 3)^2 = 0^2 = 0$$

$$④ D_4^2 = (4 - 3)^2 = 1^2 = 1$$

$$⑤ D_5^2 = (5 - 3)^2 = 2^2 = 4$$

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

Example

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

$$① D_1^2 = (1 - 3)^2 = (-2)^2 = 4$$

$$② D_2^2 = (2 - 3)^2 = (-1)^2 = 1$$

$$③ D_3^2 = (3 - 3)^2 = 0^2 = 0$$

$$④ D_4^2 = (4 - 3)^2 = 1^2 = 1$$

$$⑤ D_5^2 = (5 - 3)^2 = 2^2 = 4$$

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

Example

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

$$① D_1^2 = (1 - 3)^2 = (-2)^2 = 4$$

$$② D_2^2 = (2 - 3)^2 = (-1)^2 = 1$$

$$③ D_3^2 = (3 - 3)^2 = 0^2 = 0$$

$$④ D_4^2 = (4 - 3)^2 = 1^2 = 1$$

$$⑤ D_5^2 = (5 - 3)^2 = 2^2 = 4$$

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

Example

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

$$① D_1^2 = (1 - 3)^2 = (-2)^2 = 4$$

$$② D_2^2 = (2 - 3)^2 = (-1)^2 = 1$$

$$③ D_3^2 = (3 - 3)^2 = 0^2 = 0$$

$$④ D_4^2 = (4 - 3)^2 = 1^2 = 1$$

$$⑤ D_5^2 = (5 - 3)^2 = 2^2 = 4$$

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

Example

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

$$① D_1^2 = (1 - 3)^2 = (-2)^2 = 4$$

$$② D_2^2 = (2 - 3)^2 = (-1)^2 = 1$$

$$③ D_3^2 = (3 - 3)^2 = 0^2 = 0$$

$$④ D_4^2 = (4 - 3)^2 = 1^2 = 1$$

$$⑤ D_5^2 = (5 - 3)^2 = 2^2 = 4$$

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

Example

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

$$① D_1^2 = (1 - 3)^2 = (-2)^2 = 4$$

$$② D_2^2 = (2 - 3)^2 = (-1)^2 = 1$$

$$③ D_3^2 = (3 - 3)^2 = 0^2 = 0$$

$$④ D_4^2 = (4 - 3)^2 = 1^2 = 1$$

$$⑤ D_5^2 = (5 - 3)^2 = 2^2 = 4$$

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

Estatística
Descritiva II

- Felipe
Figueiredo

Medidas de Tendência Central

- ## Medidas de Dispersão

- Amplitude
- Desvios em relação à média
- Variância
- Desvio Padrão**
- Exercícios
- Coeficiente de Variação

Medidas de Posição

- ## Boxplot

Resumo

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ≡ ≡ ≡ ↺ 🔍 ↻

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

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

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

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

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

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

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

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

- 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 Medidas de Tendência Central
 - Média
 - Mediana
 - Moda
 - Comparação entre as Medidas Centrais
- 2 Medidas de Dispersão
 - Amplitude
 - Desvios em relação à média
 - Variância
 - Desvio Padrão
 - **Exercícios**
 - Coeficiente de Variação
- 3 Medidas de Posição
 - Quartis
 - Percentis e Decis
- 4 Boxplot
- 5 Resumo

Exercício

Um pesquisador observou as seguintes idades (anos) para uma amostra: 35, 33, 37, 33, 34.

Determine:

- 1 A variância amostral (s^2)
- 2 O desvio padrão amostral (s)

Solução

Lembrando que $\bar{x} = 34.4$, temos:

$$\begin{aligned} s^2 &= \frac{(35 - 34.4)^2 + (33 - 34.4)^2 + \dots}{5 - 1} \\ &= \frac{0.36 + 1.96 + 6.76 + 1.96 + 0.16}{4} = 2.8 \end{aligned}$$

$$s = \sqrt{2.8} = 1.67$$

Exercício

Um pesquisador observou as seguintes idades (anos) para uma amostra: 35, 33, 37, 33, 34.

Determine:

- 1 A variância amostral (s^2)
- 2 O desvio padrão amostral (s)

Solução

Lembrando que $\bar{x} = 34.4$, temos:

$$\begin{aligned} 1 \quad s^2 &= \frac{(35 - 34.4)^2 + (33 - 34.4)^2 + \dots}{5 - 1} \\ &= \frac{0.36 + 1.96 + 6.76 + 1.96 + 0.16}{4} = 2.8 \end{aligned}$$

$$2 \quad s = \sqrt{2.8} = 1.67$$

Exercício

Um pesquisador observou as seguintes idades (anos) para uma amostra: 35, 33, 37, 33, 34.

Determine:

- 1 A variância amostral (s^2)
- 2 O desvio padrão amostral (s)

Solução

Lembrando que $\bar{x} = 34.4$, temos:

$$\begin{aligned} 1 \quad s^2 &= \frac{(35 - 34.4)^2 + (33 - 34.4)^2 + \dots}{5 - 1} \\ &= \frac{0.36 + 1.96 + 6.76 + 1.96 + 0.16}{4} = 2.8 \end{aligned}$$

$$2 \quad s = \sqrt{2.8} = 1.67$$

Exercício

Um pesquisador observou as seguintes idades (anos) para uma amostra: 35, 33, 37, 33, 34.

Determine:

- 1 A variância amostral (s^2)
- 2 O desvio padrão amostral (s)

Solução

Lembrando que $\bar{x} = 34.4$, temos:

$$\begin{aligned} 1 \quad s^2 &= \frac{(35 - 34.4)^2 + (33 - 34.4)^2 + \dots}{5 - 1} \\ &= \frac{0.36 + 1.96 + 6.76 + 1.96 + 0.16}{4} = 2.8 \end{aligned}$$

$$2 \quad s = \sqrt{2.8} = 1.67$$

- 1 Medidas de Tendência Central
 - Média
 - Mediana
 - Moda
 - Comparação entre as Medidas Centrais
- 2 Medidas de Dispersão
 - Amplitude
 - Desvios em relação à média
 - Variância
 - Desvio Padrão
 - Exercícios
 - **Coeficiente de Variação**
- 3 Medidas de Posição
 - Quartis
 - Percentis e Decis
- 4 Boxplot
- 5 Resumo

Definition

$$CV = \frac{\sigma}{\mu}$$

- Normaliza a variabilidade em relação à média
- Permite comparar a variabilidade de datasets não relacionados (mesmo que não usem a mesma unidade)
- Só deve ser usado para grandezas em escala (i.e. possui um “zero” não arbitrário, ou “zero absoluto”)

Definition

$$CV = \frac{\sigma}{\mu}$$

- Normaliza a variabilidade em relação à média
- Permite comparar a variabilidade de datasets não relacionados (mesmo que não usem a mesma unidade)
- Só deve ser usado para grandezas em escala (i.e. possui um “zero” não arbitrário, ou “zero absoluto”)

Definition

$$CV = \frac{\sigma}{\mu}$$

- Normaliza a variabilidade em relação à média
- Permite comparar a variabilidade de datasets não relacionados (mesmo que não usem a mesma unidade)
- Só deve ser usado para grandezas em escala (i.e. possui um “zero” não arbitrário, ou “zero absoluto”)

Definition

$$CV = \frac{\sigma}{\mu}$$

- Normaliza a variabilidade em relação à média
- Permite comparar a variabilidade de datasets não relacionados (mesmo que não usem a mesma unidade)
- Só deve ser usado para grandezas em escala (i.e. possui um “zero” não arbitrário, ou “zero absoluto”)

Example

x = Estatura e y = Perímetro abdominal.

$x =$	$y =$
181.2	76.3
173.7	66.7
169.0	73.3
184.1	74.8
174.4	82.7
172.6	79.6

Qual das duas amostras tem maior variabilidade?

1 Calcular a média \bar{x}

2 Calcular a variância

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

3 Calcular o desvio padrão $s = \sqrt{s^2}$

4 $CV = \frac{s}{\bar{x}}$

Example

x = Estatura e y = Perímetro abdominal.

$x =$	$y =$
181.2	76.3
173.7	66.7
169.0	73.3
184.1	74.8
174.4	82.7
172.6	79.6

Qual das duas amostras tem maior variabilidade?

1 Calcular a média \bar{x}

2 Calcular a variância

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

3 Calcular o desvio padrão $s = \sqrt{s^2}$

4 $CV = \frac{s}{\bar{x}}$

Example

x = Estatura e y = Perímetro abdominal.

$x =$	$y =$
181.2	76.3
173.7	66.7
169.0	73.3
184.1	74.8
174.4	82.7
172.6	79.6

Qual das duas amostras tem maior variabilidade?

1 Calcular a média \bar{x}

2 Calcular a variância

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

3 Calcular o desvio padrão $s = \sqrt{s^2}$

4 $CV = \frac{s}{\bar{x}}$

Example

x = Estatura e y = Perímetro abdominal.

$x =$	$y =$
181.2	76.3
173.7	66.7
169.0	73.3
184.1	74.8
174.4	82.7
172.6	79.6

Qual das duas amostras tem maior variabilidade?

- 1 Calcular a média \bar{x}
- 2 Calcular a variância
$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$
- 3 Calcular o desvio padrão $s = \sqrt{s^2}$
- 4 $CV = \frac{s}{\bar{x}}$

Example

x = Estatura e y = Perímetro abdominal.

$x =$	$y =$
181.2	76.3
173.7	66.7
169.0	73.3
184.1	74.8
174.4	82.7
172.6	79.6

Qual das duas amostras tem maior variabilidade?

① Calcular a média \bar{x}

② Calcular a variância

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

③ Calcular o desvio padrão $s = \sqrt{s^2}$

④ $CV = \frac{s}{\bar{x}}$

Coeficiente de Variação



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Amplitude
Desvios em relação
à média
Variância
Desvio Padrão
Exercícios
Coeficiente de
Variação

Medidas de
Posição

Boxplot

Resumo

Example

x = Estatura e y = Perímetro abdominal.

$x =$	$y =$
181.2	76.3
173.7	66.7
169.0	73.3
184.1	74.8
174.4	82.7
172.6	79.6

$$\bar{x} = 175.8 \quad s_x = 5.7$$

$$\bar{y} = 75.7 \quad s_y = 5.5$$

$$CV_x = 3.24\%$$

$$CV_y = 7.27\%$$

Resposta: O perímetro abdominal tem maior variabilidade que a altura.

- Permitem estabelecer informações quantitativas relativas à ordem dos dados



- Permitem estabelecer informações quantitativas relativas à ordem dos dados
-

- 1 Medidas de Tendência Central
 - Média
 - Mediana
 - Moda
 - Comparação entre as Medidas Centrais
- 2 Medidas de Dispersão
 - Amplitude
 - Desvios em relação à media
 - Variância
 - Desvio Padrão
 - Exercícios
 - Coeficiente de Variação
- 3 Medidas de Posição
 - Quartis
 - Percentis e Decis
- 4 Boxplot
- 5 Resumo

Estadística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Medidas de
Posição

Quartis

Percentis e Decis

Boxplot

Resumo

Definition

Dividem o dataset em quatro partes, cada uma com 25% dos dados

- Q_1 , primeiro quartil, representa os primeiros 25% dos dados
- Q_2 , segundo quartil, representa os primeiros 50% dos dados
- Q_3 , terceiro quartil, representa os primeiros 75% dos dados

Pergunta

O que podemos dizer sobre o segundo quartil (Q_2)?

Definition

Dividem o dataset em quatro partes, cada uma com 25% dos dados

- Q_1 , primeiro quartil, representa os primeiros 25% dos dados
- Q_2 , segundo quartil, representa os primeiros 50% dos dados
- Q_3 , terceiro quartil, representa os primeiros 75% dos dados

Pergunta

O que podemos dizer sobre o segundo quartil (Q_2)?

Definition

Dividem o dataset em quatro partes, cada uma com 25% dos dados

- Q_1 , primeiro quartil, representa os primeiros 25% dos dados
- Q_2 , segundo quartil, representa os primeiros 50% dos dados
- Q_3 , terceiro quartil, representa os primeiros 75% dos dados

Pergunta

O que podemos dizer sobre o segundo quartil (Q_2)?

Definition

Dividem o dataset em quatro partes, cada uma com 25% dos dados

- Q_1 , primeiro quartil, representa os primeiros 25% dos dados
- Q_2 , segundo quartil, representa os primeiros 50% dos dados
- Q_3 , terceiro quartil, representa os primeiros 75% dos dados

Pergunta

O que podemos dizer sobre o segundo quartil (Q_2)?

Definition

Dividem o dataset em quatro partes, cada uma com 25% dos dados

- Q_1 , primeiro quartil, representa os primeiros 25% dos dados
- Q_2 , segundo quartil, representa os primeiros 50% dos dados
- Q_3 , terceiro quartil, representa os primeiros 75% dos dados

Pergunta

O que podemos dizer sobre o segundo quartil (Q_2)?

Example

Os pesos de 102 bebês nascidos em uma certa maternidade ao longo de um ano foram anotados e ordenados. Um certo bebê ocupa o Q_3 deste dataset.

- Isto significa que aproximadamente 75% dos bebês nascidos nesta maternidade tem peso menor ou igual a ele (Mazel Tov!).

Example

Os pesos de 102 bebês nascidos em uma certa maternidade ao longo de um ano foram anotados e ordenados. Um certo bebê ocupa o Q_3 deste dataset.

- Isto significa que aproximadamente 75% dos bebês nascidos nesta maternidade tem peso menor ou igual a ele (Mazel Tov!).

- 1 Medidas de Tendência Central
 - Média
 - Mediana
 - Moda
 - Comparação entre as Medidas Centrais
- 2 Medidas de Dispersão
 - Amplitude
 - Desvios em relação à media
 - Variância
 - Desvio Padrão
 - Exercícios
 - Coeficiente de Variação
- 3 Medidas de Posição
 - Quartis
 - Percentis e Decis
- 4 Boxplot
- 5 Resumo

Estadística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Medidas de
Posição

Quartis

Percentis e Decis

Boxplot

Resumo

Definition

O percentil de ordem k (onde k é qualquer valor entre 0 e 100), denotado por P_k , é o valor tal que $k\%$ dos valores do dataset são menores ou iguais a ele.

- Generalizam a idéia dos quartis
- Dividem o dataset em 100 partes
- Maior granularidade na ordem
- **Decis**: dividem o dataset em 10 partes

Definition

O percentil de ordem k (onde k é qualquer valor entre 0 e 100), denotado por P_k , é o valor tal que $k\%$ dos valores do dataset são menores ou iguais a ele.

- Generalizam a idéia dos quartis
- Dividem o dataset em 100 partes
- Maior granularidade na ordem
- **Decis**: dividem o dataset em 10 partes

Definition

O percentil de ordem k (onde k é qualquer valor entre 0 e 100), denotado por P_k , é o valor tal que $k\%$ dos valores do dataset são menores ou iguais a ele.

- Generalizam a idéia dos quartis
- Dividem o dataset em 100 partes
- Maior granularidade na ordem
- **Decis**: dividem o dataset em 10 partes

Definition

O percentil de ordem k (onde k é qualquer valor entre 0 e 100), denotado por P_k , é o valor tal que $k\%$ dos valores do dataset são menores ou iguais a ele.

- Generalizam a idéia dos quartis
- Dividem o dataset em 100 partes
- Maior granularidade na ordem
- **Decis**: dividem o dataset em 10 partes

Definition

O percentil de ordem k (onde k é qualquer valor entre 0 e 100), denotado por P_k , é o valor tal que $k\%$ dos valores do dataset são menores ou iguais a ele.

- Generalizam a idéia dos quartis
- Dividem o dataset em 100 partes
- Maior granularidade na ordem
- **Decis**: dividem o dataset em 10 partes

O Boxplot



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Medidas de
Posição

Boxplot

Resumo

- Gráfico que ilustra a dispersão dos dados pelos quartis
- Retângulo que representa a Distância Interquartílica ($DQ = Q_3 - Q_1$)
- Barra interna que representa a mediana (Q_2)
- Limite superior (barra vertical): $Q_3 + 1.5 \cdot DQ$
- Limite inferior (barra vertical): $Q_1 - 1.5 \cdot DQ$
- Outliers como pontos, círculos ou estrelas
- Conveniente para comparar vários grupos ou amostras

O Boxplot



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Medidas de
Posição

Boxplot

Resumo

- Gráfico que ilustra a dispersão dos dados pelos quartis
- Retângulo que representa a Distância Interquartílica ($DQ = Q_3 - Q_1$)
- Barra interna que representa a mediana (Q_2)
- Limite superior (barra vertical): $Q_3 + 1.5 \cdot DQ$
- Limite inferior (barra vertical): $Q_1 - 1.5 \cdot DQ$
- Outliers como pontos, círculos ou estrelas
- Conveniente para comparar vários grupos ou amostras

O Boxplot



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Medidas de
Posição

Boxplot

Resumo

- Gráfico que ilustra a dispersão dos dados pelos quartis
- Retângulo que representa a Distância Interquartílica ($DQ = Q_3 - Q_1$)
- Barra interna que representa a mediana (Q_2)
- Limite superior (barra vertical): $Q_3 + 1.5 \cdot DQ$
- Limite inferior (barra vertical): $Q_1 - 1.5 \cdot DQ$
- Outliers como pontos, círculos ou estrelas
- Conveniente para comparar vários grupos ou amostras

O Boxplot



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Medidas de
Posição

Boxplot

Resumo

- Gráfico que ilustra a dispersão dos dados pelos quartis
- Retângulo que representa a Distância Interquartílica ($DQ = Q_3 - Q_1$)
- Barra interna que representa a mediana (Q_2)
- Limite superior (barra vertical): $Q_3 + 1.5 \cdot DQ$
- Limite inferior (barra vertical): $Q_1 - 1.5 \cdot DQ$
- Outliers como pontos, círculos ou estrelas
- Conveniente para comparar vários grupos ou amostras

O Boxplot



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Medidas de
Posição

Boxplot

Resumo

- Gráfico que ilustra a dispersão dos dados pelos quartis
- Retângulo que representa a Distância Interquartílica ($DQ = Q_3 - Q_1$)
- Barra interna que representa a mediana (Q_2)
- Limite superior (barra vertical): $Q_3 + 1.5 \cdot DQ$
- Limite inferior (barra vertical): $Q_1 - 1.5 \cdot DQ$
- Outliers como pontos, círculos ou estrelas
- Conveniente para comparar vários grupos ou amostras

O Boxplot



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Medidas de
Posição

Boxplot

Resumo

- Gráfico que ilustra a dispersão dos dados pelos quartis
- Retângulo que representa a Distância Interquartílica ($DQ = Q_3 - Q_1$)
- Barra interna que representa a mediana (Q_2)
- Limite superior (barra vertical): $Q_3 + 1.5 \cdot DQ$
- Limite inferior (barra vertical): $Q_1 - 1.5 \cdot DQ$
- Outliers como pontos, círculos ou estrelas
- Conveniente para comparar vários grupos ou amostras

O Boxplot



Estatística
Descritiva II

Felipe
Figueiredo

Medidas de
Tendência
Central

Medidas de
Dispersão

Medidas de
Posição

Boxplot

Resumo

- Gráfico que ilustra a dispersão dos dados pelos quartis
- Retângulo que representa a Distância Interquartílica ($DQ = Q_3 - Q_1$)
- Barra interna que representa a mediana (Q_2)
- Limite superior (barra vertical): $Q_3 + 1.5 \cdot DQ$
- Limite inferior (barra vertical): $Q_1 - 1.5 \cdot DQ$
- Outliers como pontos, círculos ou estrelas
- Conveniente para comparar vários grupos ou amostras

Boxplot

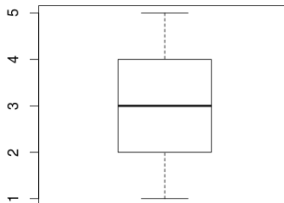
Example

Dataset

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

$\bar{x} = 3, M_d = 3$

$Q_1 = 2, Q_3 = 4$



Boxplot

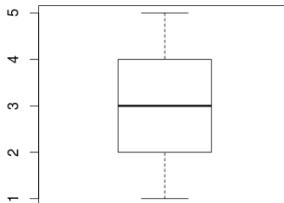
Example

Dataset

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

$\bar{x} = 3, M_d = 3$

$Q_1 = 2, Q_3 = 4$



Boxplot

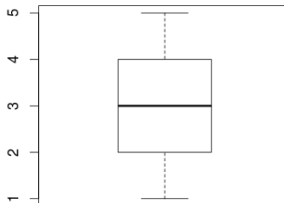
Example

Dataset

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

$\bar{x} = 3, M_d = 3$

$Q_1 = 2, Q_3 = 4$



Boxplot

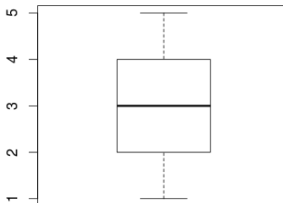
Example

Dataset

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

$\bar{x} = 3, M_d = 3$

$Q_1 = 2, Q_3 = 4$



O Boxplot

Example

Exemplo do colesterol

$$x_1 = 142$$

$$x_2 = 144$$

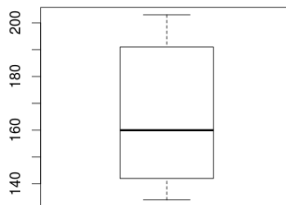
$$x_3 = 176$$

$$x_4 = 203$$

$$x_5 = 134$$

$$x_6 = 191$$

$$\bar{x} = 165, M_d = 160$$



O Boxplot

Example

Exemplo do colesterol

$$x_1 = 142$$

$$x_2 = 144$$

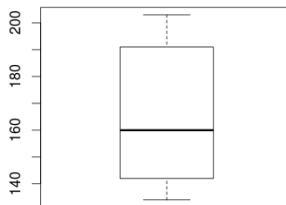
$$x_3 = 176$$

$$x_4 = 203$$

$$x_5 = 134$$

$$x_6 = 191$$

$$\bar{x} = 165, M_d = 160$$



O Boxplot

Example

Exemplo do colesterol

$$x_1 = 142$$

$$x_2 = 144$$

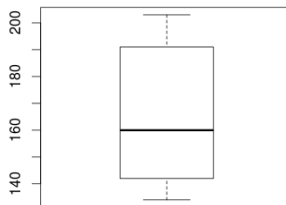
$$x_3 = 176$$

$$x_4 = 203$$

$$x_5 = 134$$

$$x_6 = 191$$

$$\bar{x} = 165, M_d = 160$$



O Boxplot

Example

Exemplo do colesterol

$$x_1 = 142$$

$$x_2 = 144$$

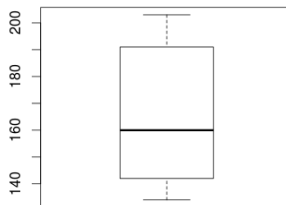
$$x_3 = 176$$

$$x_4 = 203$$

$$x_5 = 134$$

$$x_6 = 191$$

$$\bar{x} = 165, M_d = 160$$



Boxplot: duas amostras

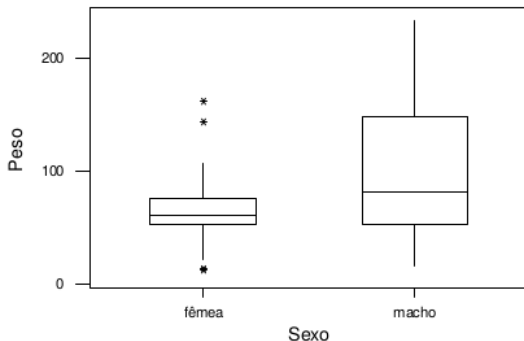


Figura: Boxplots para dois grupos de dados (Fonte: Reis, Reis, 2002)

Ao iniciar sua Análise Exploratória de Dados (EDA), você pode visualizar sua amostra com:

1 Resumo dos cinco números

- Valor mínimo
- Primeiro quartil Q_1
- Mediana (e/ou média)
- Terceiro quartil Q_3
- Valor máximo

2 Boxplot



Ao iniciar sua Análise Exploratória de Dados (EDA), você pode visualizar sua amostra com:

1 Resumo dos cinco números

- Valor mínimo
- Primeiro quartil Q_1
- Mediana (e/ou média)
- Terceiro quartil Q_3
- Valor máximo

2 Boxplot



Ao iniciar sua Análise Exploratória de Dados (EDA), você pode visualizar sua amostra com:

1 Resumo dos cinco números

- Valor mínimo
- Primeiro quartil Q_1
- Mediana (e/ou média)
- Terceiro quartil Q_3
- Valor máximo

2 Boxplot



Ao iniciar sua Análise Exploratória de Dados (EDA), você pode visualizar sua amostra com:

1 Resumo dos cinco números

- Valor mínimo
- Primeiro quartil Q_1
- Mediana (e/ou média)
- Terceiro quartil Q_3
- Valor máximo

2 Boxplot



Ao iniciar sua Análise Exploratória de Dados (EDA), você pode visualizar sua amostra com:

1 Resumo dos cinco números

- Valor mínimo
- Primeiro quartil Q_1
- Mediana (e/ou média)
- Terceiro quartil Q_3
- Valor máximo

2 Boxplot



Ao iniciar sua Análise Exploratória de Dados (EDA), você pode visualizar sua amostra com:

1 Resumo dos cinco números

- Valor mínimo
- Primeiro quartil Q_1
- Mediana (e/ou média)
- Terceiro quartil Q_3
- Valor máximo

2 Boxplot



Ao iniciar sua Análise Exploratória de Dados (EDA), você pode visualizar sua amostra com:

1 Resumo dos cinco números

- Valor mínimo
- Primeiro quartil Q_1
- Mediana (e/ou média)
- Terceiro quartil Q_3
- Valor máximo

2 Boxplot

