Estatística

Medidas descritivas

Dra. Agatha Rodrigues

agatha.srodrigues@gmail.com

Agradecimento ao professor Dr. Alessandro Sarnaglia por disponibilizar alguns slides utilizados nessa apresentação

Sumário

Sumário

Variáveis

Variável

Característica de interesse associada a uma população.

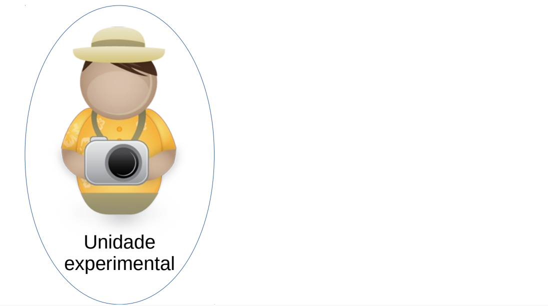
Podem representar atributos ou medidas realizadas nas unidades experimentais.

Variáveis

Variável

Característica de interesse associada a uma população.

Podem representar atributos ou medidas realizadas nas unidades experimentais.

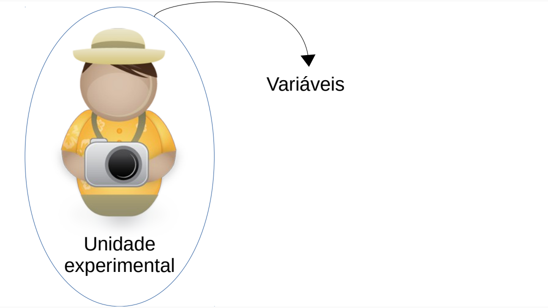


Variáveis

Variável

Característica de interesse associada a uma população.

Podem representar atributos ou medidas realizadas nas unidades experimentais.

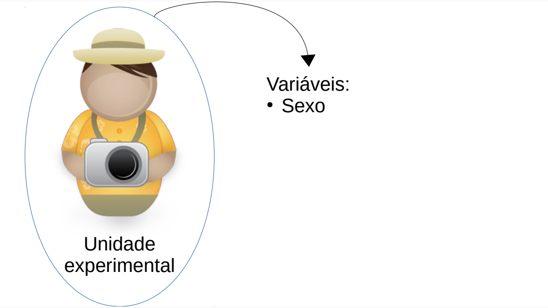


Variáveis

Variável

Característica de interesse associada a uma população.

Podem representar atributos ou medidas realizadas nas unidades experimentais.

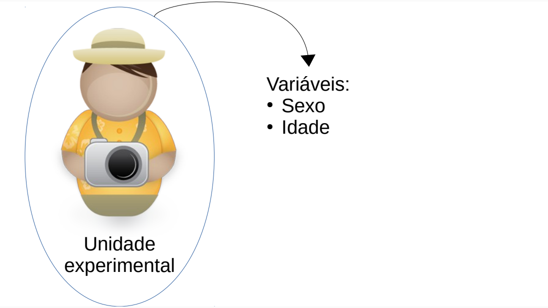


Variáveis

Variável

Característica de interesse associada a uma população.

Podem representar atributos ou medidas realizadas nas unidades experimentais.

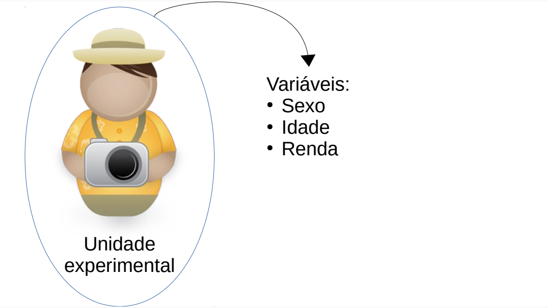


Variáveis

Variável

Característica de interesse associada a uma população.

Podem representar atributos ou medidas realizadas nas unidades experimentais.

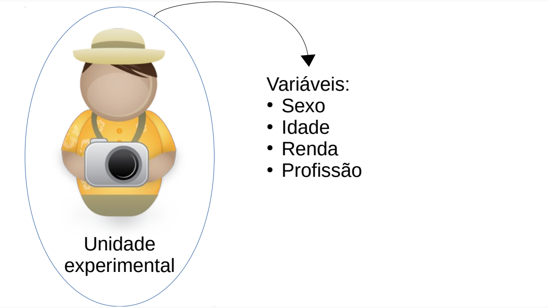


Variáveis

Variável

Característica de interesse associada a uma população.

Podem representar atributos ou medidas realizadas nas unidades experimentais.

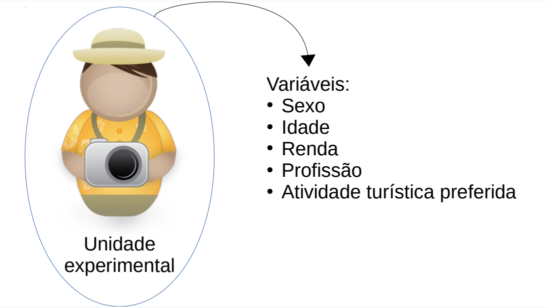


Variáveis

Variável

Característica de interesse associada a uma população.

Podem representar atributos ou medidas realizadas nas unidades experimentais.

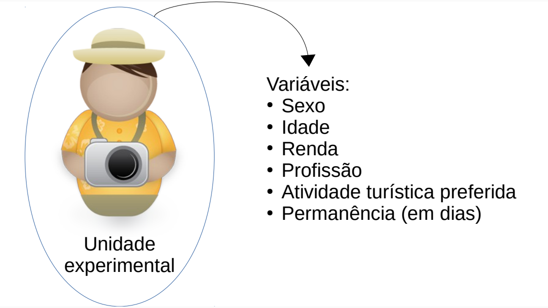


Variáveis

Variável

Característica de interesse associada a uma população.

Podem representar atributos ou medidas realizadas nas unidades experimentais.

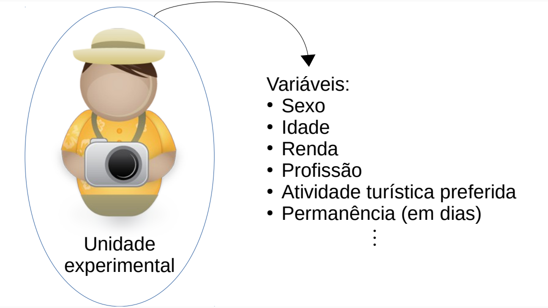


Variáveis

Variável

Característica de interesse associada a uma população.

Podem representar atributos ou medidas realizadas nas unidades experimentais.



Tipos de variáveis

Dependendo dos possíveis valores de uma variável, algumas técnicas **não** devem ser aplicadas.

Tipos de variáveis

Dependendo dos possíveis valores de uma variável, algumas técnicas **não** devem ser aplicadas.

Por exemplo: se a variável for **sexo**, em que os possíveis valores são feminino e masculino, não faz sentido em considerar a média.

Tipos de variáveis

Dependendo dos possíveis valores de uma variável, algumas técnicas **não** devem ser aplicadas.

Por exemplo: se a variável for **sexo**, em que os possíveis valores são feminino e masculino, não faz sentido em considerar a média.

Para se aplicar corretamente as técnicas que veremos adiante, é importante saber classificar corretamente as variáveis.

Variáveis

Variáveis

Característica de interesse associada a uma população.

As variáveis podem ser, segundo sua natureza:

Variáveis

Variáveis

Característica de interesse associada a uma população.

As variáveis podem ser, segundo sua natureza:

Qualitativas

Apresentam como possíveis realizações uma qualidade (ou atributo) do indivíduo pesquisado.

Exemplo

Sexo, grau de instrução, estado civil, classe social, presença de diabetes, etc.

Variáveis

Variáveis

Característica de interesse associada a uma população.

As variáveis podem ser, segundo sua natureza:

Qualitativas

Apresentam como possíveis realizações uma qualidade (ou atributo) do indivíduo pesquisado.

Exemplo

Sexo, grau de instrução, estado civil, classe social, presença de diabetes, etc.

Quantitativas

Apresentam como possíveis realizações números resultantes de uma contagem ou de uma mensuração.

Exemplo

Número de filhos, salário, temperatura, pressão arterial, concentração de alguma substância, etc.

Classificação das variáveis

As variáveis qualitativas e quantitativas ainda podem sofrer distinções entre dois tipos:

Classificação das variáveis

As variáveis qualitativas e quantitativas ainda podem sofrer distinções entre dois tipos:

Qualitativas

- **Nominal**
  Não existe nenhuma ordenação nas possíveis realizações. Exemplo: Sexo, estado civil, presença de diabetes, etc.
- **Ordinal**
  Existe uma ordem nos seus resultados. Exemplo: grau de instrução, estadiamento de uma doença, etc.

Classificação das variáveis

As variáveis qualitativas e quantitativas ainda podem sofrer distinções entre dois tipos:

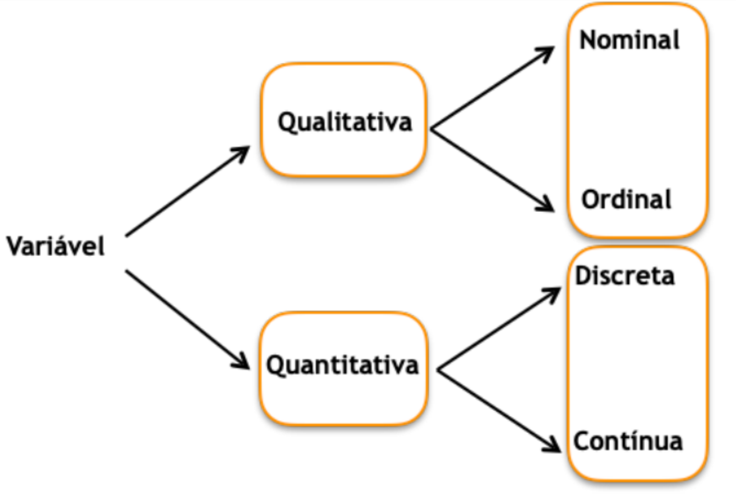
Qualitativas

- **Nominal**
  Não existe nenhuma ordenação nas possíveis realizações. Exemplo: Sexo, estado civil, presença de diabetes, etc.
- **Ordinal**
  Existe uma ordem nos seus resultados. Exemplo: grau de instrução, estadiamento de uma doença, etc.

Quantitativas

- **Discreta**
  Possíveis valores formam um conjunto finito ou enumerável de pontos e que resultam, frequentemente, de uma contagem. Exemplo: número de filhos (0,1,2,...), quantidade de acidentes em um mês (0, 1, 2, ...), etc.
- **Contínua**
  Características mensuráveis que assumem valores em um intervalo de números reais. Exemplo: salário, temperatura, pressão arterial, etc.

Classificação das variáveis



Classificação das variáveis

Exercício - classifique:

- Sexo;
- Altura;
- Peso;
- Fuma (sim ou não);
- Tolerância ao cigarro (indiferente, incomoda pouco, incomoda muito);
- Consumo de café (nunca, 1 a 2 vezes por semana, 3 a 6 vezes por semana, uma vez por dia, >1 vez por dia);
- Horas de atividade física por semana.

Sumário

O que é análise descritiva?

O que fazer com as observações que coletamos?

O que é análise descritiva?

O que fazer com as observações que coletamos?

Primeira etapa

Estatística descritiva

Resumir os dados;

Entender os dados;

Auxílio para decisões na análise inferencial.

Como fazer?

- Tabelas de frequências.
- Medidas descritivas ou resumo
  - Medidas de posição
  - Medidas de dispersão

Depende do tipo de variável

Segunda etapa

Estatística inferencial

Inferir certos fatos acerca da população a partir de resultados observados na amostra, controlando erros.

Como fazer?

Abordaremos este assunto nas próximas aulas.

Sumário

Motivação

Perguntas:

- como os valores da variável se distribuem em termos da **frequência** com que ocorrem?
- alguns valores são mais frequentes que outros?

Problema: como extrair essas informações/padrões?

Solução: distribuição de frequências.

Construção

Aplicação:

- Variáveis qualitativas;
- Variáveis quantitativas discretas com poucos valores diferentes.

Para uma variável, a frequência de um possível **valor** (categoria ou número) é a quantidade absoluta de ocorrências desse valor no conjunto de dados.

Construção

Aplicação:

- Variáveis qualitativas;
- Variáveis quantitativas discretas com poucos valores diferentes.

Para uma variável, a frequência de um possível **valor** (categoria ou número) é a quantidade absoluta de ocorrências desse valor no conjunto de dados.

Por exemplo, o número de filhos de 20 pessoas são dados abaixo:

$$0\ 1\ 3\ 2\ 0\ 1\ 2\ 0\ 1\ 0\ 0\ 1\ 3\ 0\ 2\ 1\ 1\ 1\ 0\ 0.$$

Construção

Aplicação:

- Variáveis qualitativas;
- Variáveis quantitativas discretas com poucos valores diferentes.

Para uma variável, a frequência de um possível **valor** (categoria ou número) é a quantidade absoluta de ocorrências desse valor no conjunto de dados.

Por exemplo, o número de filhos de 20 pessoas são dados abaixo:

0 1 3 2 0 1 2 0 1 0 0 1 3 0 2 1 1 1 0 0.

O **valor** 0 tem frequência 8.

Construção

Aplicação:

- Variáveis qualitativas;
- Variáveis quantitativas discretas com poucos valores diferentes.

Para uma variável, a frequência de um possível **valor** (categoria ou número) é a quantidade absoluta de ocorrências desse valor no conjunto de dados.

Por exemplo, o número de filhos de 20 pessoas são dados abaixo:

0 1 3 2 0 1 2 0 1 0 0 1 3 0 2 1 1 1 0 0.

O **valor** 0 tem frequência 8.
O **valor** 1 tem frequência 7.

Construção

Aplicação:

- Variáveis qualitativas;
- Variáveis quantitativas discretas com poucos valores diferentes.

Para uma variável, a frequência de um possível **valor** (categoria ou número) é a quantidade absoluta de ocorrências desse valor no conjunto de dados.

Por exemplo, o número de filhos de 20 pessoas são dados abaixo:

0 1 3 2 0 1 2 0 1 0 0 1 3 0 2 1 1 1 0 0.

O **valor** 0 tem frequência 8.
O **valor** 1 tem frequência 7.
O **valor** 2 tem frequência 3.

Construção

Aplicação:

- Variáveis qualitativas;
- Variáveis quantitativas discretas com poucos valores diferentes.

Para uma variável, a frequência de um possível **valor** (categoria ou número) é a quantidade absoluta de ocorrências desse valor no conjunto de dados.

Por exemplo, o número de filhos de 20 pessoas são dados abaixo:

$$0\ 1\ 3\ 2\ 0\ 1\ 2\ 0\ 1\ 0\ 0\ 1\ 3\ 0\ 2\ 1\ 1\ 1\ 0\ 0.$$

O **valor** 0 tem frequência 8.
O **valor** 1 tem frequência 7.
O **valor** 2 tem frequência 3.
O **valor** 3 tem frequência 2.

Construção

Aplicação:

- Variáveis qualitativas;
- Variáveis quantitativas discretas com poucos valores diferentes.

Para uma variável, a frequência de um possível **valor** (categoria ou número) é a quantidade absoluta de ocorrências desse valor no conjunto de dados.

Por exemplo, o número de filhos de 20 pessoas são dados abaixo:

$$0\ 1\ 3\ 2\ 0\ 1\ 2\ 0\ 1\ 0\ 0\ 1\ 3\ 0\ 2\ 1\ 1\ 1\ 0\ 0.$$

O **valor** 0 tem frequência 8.
O **valor** 1 tem frequência 7.
O **valor** 2 tem frequência 3.
O **valor** 3 tem frequência 2.

Distribuição de frequência

Tabela que associa aos valores (categorias) de uma variável as suas respectivas frequências.

Distribuições de frequência

Analisando a ocorrência das possíveis realizações da variável.

Tabela 1: Frequência absoluta do grau de instrução de 30 pacientes de **baixo risco** da Obstetrícia

Grau de instrução	Frequência absoluta (n)
Fundamental	5
Médio	20
Superior	5
Total	30

Tabela 2: Frequência absoluta do grau de instrução de 400 pacientes de **alto risco** da Obstetrícia

Grau de instrução	Frequência absoluta (n)
Fundamental	120
Médio	240
Superior	40
Total	400

Distribuições de frequência

Tabela 1: Frequência de 30 pacientes de **baixo risco** da Obstetrícia segundo o grau de instrução.

Grau de instrução	Frequência absoluta (n)	Proporção ou frequência relativa	Porcentagem (%)
Fundamental	5	0,167	16,7
Médio	20	0,666	66,6
Superior	5	0,167	16,7
Total	30	1	100

Tabela 2: Frequência de 400 pacientes de **alto risco** da Obstetrícia segundo o grau de instrução.

Grau de instrução	Frequência absoluta (n)	Proporção ou frequência relativa	Porcentagem (%)
Fundamental	120	0,30	30,0
Médio	240	0,60	60,0
Superior	40	0,10	10,0
Total	400	1	100

Distribuições de frequência

Tabela 1: Frequência de 30 pacientes de **baixo risco** da Obstetrícia segundo o grau de instrução.

Grau de instrução	Frequência absoluta (n)	Proporção ou frequência relativa	Porcentagem (%)	Porcentagem acumulada (%)
Fundamental	5	0,167	16,7	16,7
Médio	20	0,666	66,6	83,3
Superior	5	0,167	16,7	100,0
Total	30	1	100	

Tabela 2: Frequência de 400 pacientes de **alto risco** da Obstetrícia segundo o grau de instrução.

Grau de instrução	Frequência absoluta (n)	Proporção ou frequência relativa	Porcentagem (%)	Porcentagem acumulada (%)
Fundamental	120	0,30	30,0	30,0
Médio	240	0,60	60,0	90,0
Superior	40	0,10	10,0	100,0
Total	400	1	100	

Distribuições de frequência

Tabela 3: Frequências do IMC dos alunos do curso de Estatística 2018.

IMC	n	%	% acumulada
17,80	1	3.7	3.7
17,96	1	3.7	7.4
19,00	1	3.7	11.1
19,33	1	3.7	14.8
20,28	1	3.7	18.5
20,53	1	3.7	22.2
20,70	1	3.7	25.9
21,09	1	3.7	29.6
21,63	1	3.7	33.3
21,64	1	3.7	37.0
21,64	1	3.7	40.7
21,83	1	3.7	44.4
22,19	1	3.7	48.1
22,38	1	3.7	51.9
22,39	1	3.7	55.6
22,99	1	3.7	59.3
23,53	1	3.7	63.0
23,67	1	3.7	66.7
24,06	1	3.7	70.4
24,77	1	3.7	74.1
24,77	1	3.7	77.8
25,39	1	3.7	81.5
25,64	1	3.7	85.2
26,57	1	3.7	88.9
28,23	1	3.7	92.6
33,33	1	3.7	96.3
33,39	1	3.7	100.0
Total	27	100.0	

Faz sentido?

Distribuições de frequência

Tabela 4: Frequências das categorias IMC dos alunos do curso de Estatística 2018.

IMC_cat	n	%	% acumulada
<18.5	2	7.4	7.4
[18.5 - 25)	19	70.4	77.8
[25 - 30)	4	14.8	92.6
>=30	2	7.4	100.0
Total	27	100.0	

Agora, qual a classificação da variável IMC-cat?

Medidas descritivas

O que são? Valores que descrevem ou quantificam algum comportamento da variável de interesse.

Quais comportamentos?

- Um **valor típico** que a variável tende a apresentar;
- A **dispersão** ou **variabilidade** dos dados.

Medidas resumo

Variáveis Quantitativas

Variáveis Quantitativas

Medidas de posição

- Mínimo;
- Máximo;
- Moda;
- Média;
- Mediana;
- Percentis.

Medidas resumo

Variáveis Quantitativas

Medidas de posição

- Mínimo;
- Máximo;
- Moda;
- Média;
- Mediana;
- Percentis.

Medidas de dispersão

- Amplitude;
- Intervalo interquartil;
- Desvio absoluto médio;
- Variância;
- Desvio padrão;
- Coeficiente de variação.

Sumário

Medidas de posição

- Mínimo (Min): a menor observação;
- Máximo (Max): a maior observação;

Medidas de posição

- Mínimo (Min): a menor observação;
- Máximo (Max): a maior observação;

Exemplo: Nota de alunos em um curso

Dados: 5, 4, 7, 9, 10, 5, 5, 8, 4, 7

n	Min	Max
10	4	10

Medidas de tendência central

Medidas que buscam descrever **um valor típico** que a variável tende a apresentar.

- moda;
- média aritmética;
- mediana.

Moda

Definição: O(s) valor(es) mais frequente(s).

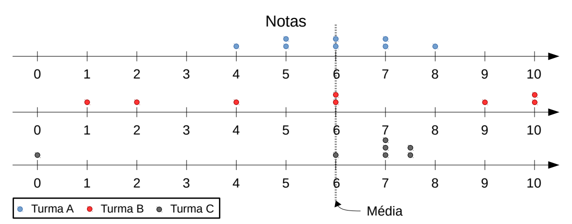
Obs.:Um conjunto de dados pode ser unimodal, bimodal, ... Quando todos valores têm frequência 1 (não existe moda), dizemos que a amostra é **amodal**.

Comentários:

- representa o(s) valor(es) mais provável(eis);
- é muito indicada em dados multimodais;
- não é afetada por dados atípicos;
- pode ser usada em variáveis qualitativas.

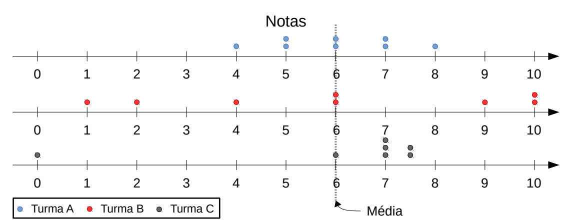
Moda

Exemplo das notas de alunos de 3 turmas.



Moda

Exemplo das notas de alunos de 3 turmas.



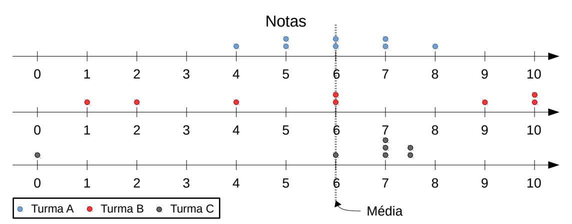
Na turma A, as modas são 5, 6 e 7, todas com frequência 2 (conjunto **trimodal**).

Na turma B, as modas são 6 e 10, ambas com frequência 2 (conjunto **bimodal**).

Na turma C, a moda é o 7, com frequência 3 (conjunto **unimodal**).

Moda

Exemplo das notas de alunos de 3 turmas.



Na turma A, as modas são 5, 6 e 7, todas com frequência 2 (conjunto **trimodal**).

Na turma B, as modas são 6 e 10, ambas com frequência 2 (conjunto **bimodal**).

Na turma C, a moda é o 7, com frequência 3 (conjunto **unimodal**).

A moda da turma C também não é influenciada pela nota **atípica** 0.

Média aritmética

Definição: Soma de todos os valores da amostra dividida pelo números de observações.

Em símbolos: Se observamos n valores, $x_1, \ldots, x_n$, de uma variável X, a média aritmética desses valores é dada por

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

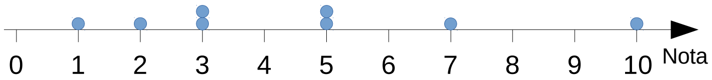
onde $\sum_{i=1}^{n} x_i = x_1 + \cdots + x_n$ representa a soma de todos valores observados.

A ideia de média é que os desvios da média se cancelem! Isto é:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n}(x_i) - n\bar{x} = n\frac{\sum_{i=1}^{n}(x_i)}{n} - n\bar{x} = n\bar{x} - n\bar{x} = 0.$$

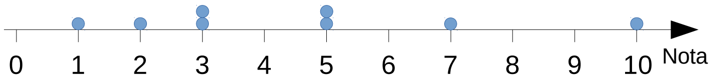
Média aritmética

Analogia com ponto de equilíbrio. Dados de notas de uma turma. As notas são representadas como um pesos de mesma massa posicionados sobre uma reta de massa desprezível nas posições referentes àquelas notas.



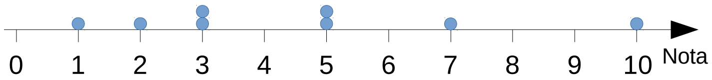
Média aritmética

Analogia com ponto de equilíbrio. Dados de notas de uma turma. As notas são representadas como um pesos de mesma massa posicionados sobre uma reta de massa desprezível nas posições referentes àquelas notas.



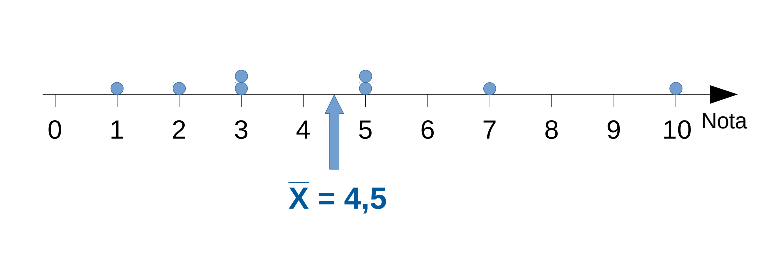
Média aritmética

Analogia com ponto de equilíbrio. Dados de notas de uma turma. As notas são representadas como um pesos de mesma massa posicionados sobre uma reta de massa desprezível nas posições referentes àquelas notas.



Média aritmética

Analogia com ponto de equilíbrio. Dados de notas de uma turma. As notas são representadas como um pesos de mesma massa posicionados sobre uma reta de massa desprezível nas posições referentes àquelas notas.



A média é o ponto de equilíbrio da reta.

Média aritmética

Turma	Aluno							
	1	2	3	4	5	6	7	8
A	4	5	5	6	6	7	7	8
B	1	2	4	6	6	9	10	10
C	0	6	7	7	7	7,5	7,5	

Média aritmética

Turma	Aluno								Soma
	1	2	3	4	5	6	7	8	
A	4	5	5	6	6	7	7	8	48
B	1	2	4	6	6	9	10	10	48
C	0	6	7	7	7	7,5	7,5		42

Média aritmética

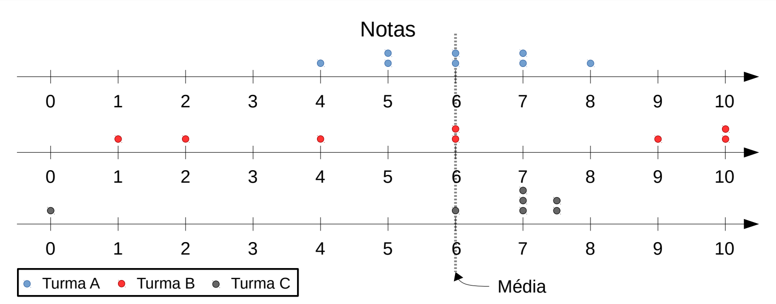
Turma	Aluno								Soma	Média
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**		
A	4	5	5	6	6	7	7	8	48	48/8 = 6
B	1	2	4	6	6	9	10	10	48	48/8 = 6
C	0	6	7	7	7	7,5	7,5		42	42/7 = 6

Média aritmética

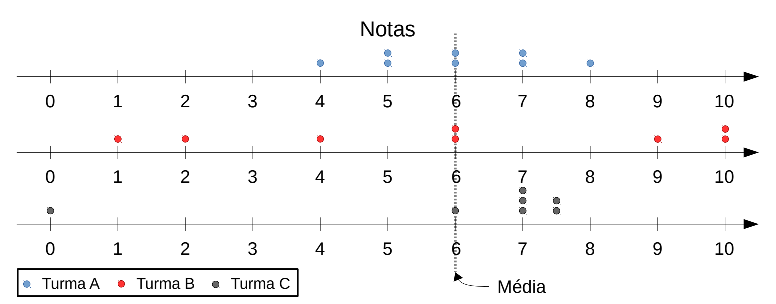
Turma	Aluno								Soma	Média
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**		
A	4	5	5	6	6	7	7	8	48	48/8 = 6
B	1	2	4	6	6	9	10	10	48	48/8 = 6
C	0	6	7	7	7	7,5	7,5		42	42/7 = 6

Em geral, as turmas tiveram o mesmo aproveitamento, pois todas tiveram a mesma média.

Média aritmética

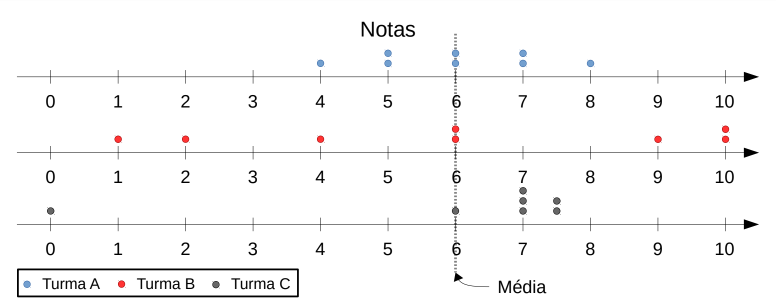


Média aritmética



Entretanto, percebemos que apenas a média pode esconder muita informação sobre os conjuntos de dados.

Média aritmética



Entretanto, percebemos que apenas a média pode esconder muita informação sobre os conjuntos de dados.

As três turmas estão longe de ter o mesmo comportamento.

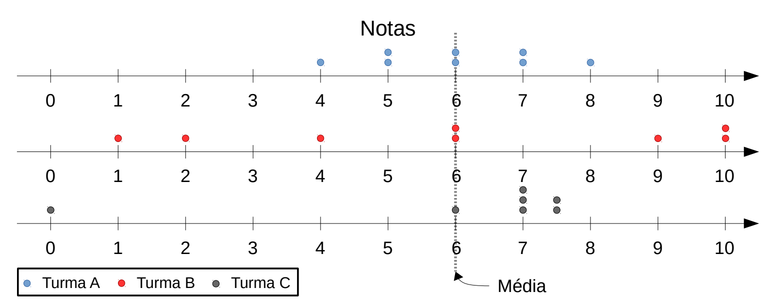
Média

A média possui as seguintes características:

- tem propriedades "boas";
- é influenciada por valores atípicos;
- não recomendada em dados assimétricos;
- só é calculada em variáveis quantitativas;

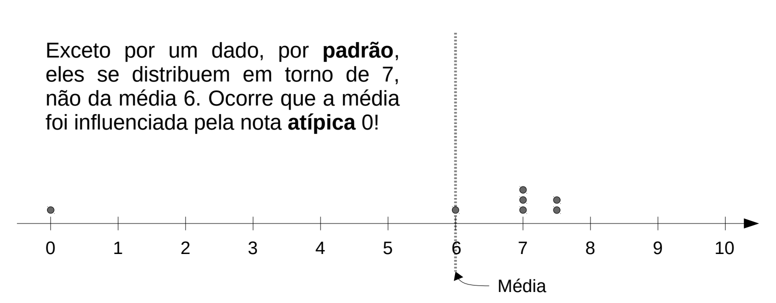
Mediana

Retornemos ao exemplo das notas das 3 turmas.



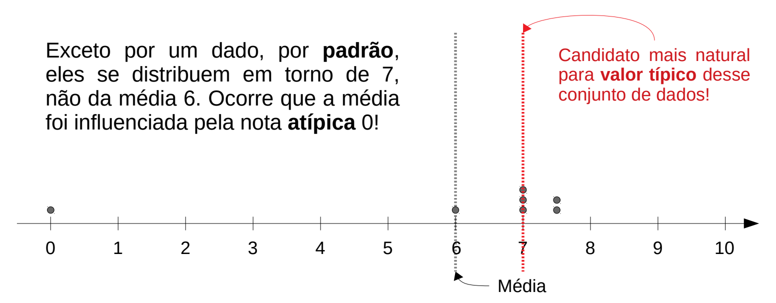
Mediana

Retornemos ao exemplo das notas das 3 turmas. Na Turma C, a média aritmética é muito influenciada por valores atípicos. O valor 6 não pode ser considerado um valor típico desse conjunto de dados.



Mediana

Retornemos ao exemplo das notas das 3 turmas. Na Turma C, a média aritmética é muito influenciada por valores atípicos. O valor 6 não pode ser considerado um valor típico desse conjunto de dados.



Por exemplo, o valor 7 parece representar melhor esse conjunto de dados.

Mediana

Definição: A Mediana divide os dados de forma que 50% deles são menores ou iguais e 50% deles são maiores ou iguais que a mediana.

Em símbolos: Se observamos n valores, $x_1, \ldots, x_n$, de uma variável X e definirmos $x_{(1)}, \ldots, x_{(n)}$ como os valores ordenados crescentemente ($x_{(1)} \leq \cdots \leq x_{(n)}$), a mediana é definida por

$$\text{Med} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{se } n \text{ é ímpar;} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & \text{se } n \text{ é par.} \end{cases}$$

Isto é, o valor na posição central se n é ímpar e a média dos valores nas posições centrais se n é par.

Mediana

Definição: A Mediana divide os dados de forma que 50% deles são menores ou iguais e 50% deles são maiores ou iguais que a mediana.

Em símbolos: Se observamos n valores, $x_1, \ldots, x_n$, de uma variável X e definirmos $x_{(1)}, \ldots, x_{(n)}$ como os valores ordenados crescentemente ($x_{(1)} \leq \cdots \leq x_{(n)}$), a mediana é definida por

$$\text{Med} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{se } n \text{ é ímpar;} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & \text{se } n \text{ é par.} \end{cases}$$

Isto é, o valor na posição central se n é ímpar e a média dos valores nas posições centrais se n é par.

Comentários:

- depende apenas da posição e não do valor;
- menos influência de dados atípicos;

Mediana

Se n for ímpar ($n = 9$)

Dados	5, 4, 7, 9, 10, 5, 8, 4, 7
Dados ordenados	4, 4, 5, 5, **7**, 7, 8, 9, 10

Posição da mediana dos dados ordenados: $\frac{9+1}{2} = 5$ Mediana: $x_{(5)} = 7$.

Mediana

Se n for ímpar ($n = 9$)

Dados	5, 4, 7, 9, 10, 5, 8, 4, 7
Dados ordenados	4, 4, 5, 5, **7**, 7, 8, 9, 10

Posição da mediana dos dados ordenados: $\frac{9+1}{2} = 5$ Mediana: $x_{(5)} = 7$.

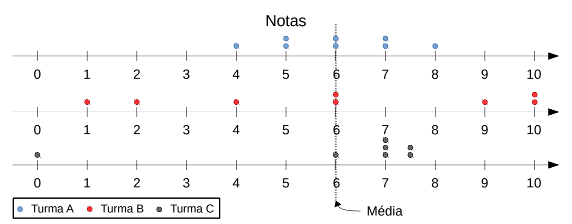
Se n for par ($n = 10$)

Dados	5, 4, 7, 9, 10, 5, 5, 8, 4, 7
Dados ordenados	4, 4, 5, 5, **5, 7**, 7, 8, 9, 10

Posição da mediana: $\frac{10}{2} = 5$ e $\frac{10}{2} + 1 = 6$ Mediana: $\frac{x_{(5)} + x_{(6)}}{2} = \frac{5+7}{2} = 6$

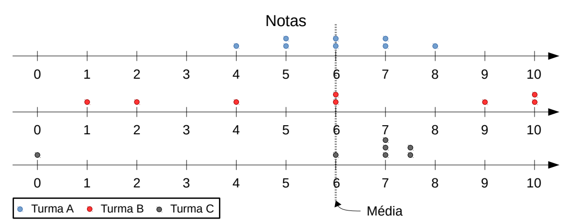
Mediana

Retornemos ao exemplo das notas das 3 turmas. Na A e na B temos 8 notas e na C temos 7.



Mediana

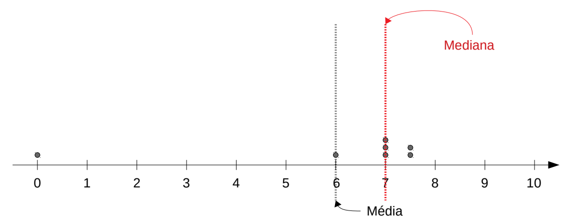
Retornemos ao exemplo das notas das 3 turmas. Na A e na B temos 8 notas e na C temos 7.



Nesse caso, as medianas são: turma A, $\frac{x_{(\mathbf{4})}+x_{(\mathbf{5})}}{2} = \frac{6+6}{2} = 6$; turma B, $\frac{x_{(\mathbf{4})}+x_{(\mathbf{5})}}{2} = \frac{6+6}{2} = 6$; e turma C, $x_{(4)} = 7$.

Mediana

Retornemos ao exemplo das notas das 3 turmas. Na A e na B temos 8 notas e na C temos 7.



Nesse caso, as medianas são: turma A, $\frac{x_{(4)}+x_{(5)}}{2} = \frac{6+6}{2} = 6$; turma B, $\frac{x_{(4)}+x_{(5)}}{2} = \frac{6+6}{2} = 6$; e turma C, $x_{(4)} = 7$.

Perceba a influência que a nota **atípica** 0 tem na média, enquanto que a mediana não é afetada.

Medidas de posição

- Média e mediana são medidas de posição ou tendência central.
- Quais informações consigo obter ao comparar essas duas medidas?

Medidas de posição

- Média e mediana são medidas de posição ou tendência central.
- Quais informações consigo obter ao comparar essas duas medidas?

Exemplo: Glicose em pacientes do HC

Dados	75, 88, 87, 98, 94
Dados ordenados	75, 87, **88**, 94, 98

- Média $(\bar{x}) = 88,4$
- Mediana $(x_{(3)}) = 88$

Medidas de posição

- O pesquisador observou que houve um erro de digitação.
- A observação 98 é na verdade 243.

Medidas de posição

- O pesquisador observou que houve um erro de digitação.
- A observação 98 é na verdade 243.

Exemplo: Glicose em pacientes do HC

Dados	75, 88, 87, 98, 94
Dados corrigidos	75, 88, 87, **243**, 94
Dados ordenados	75, 87, **88**, 94, **243**

- Média $(\bar{x}) = 117,4$
- Mediana $(x_{(3)}) = 88$

O que aconteceu com os valores médio e mediano?

Percentis

O percentil de ordem $p * 100$ $(0 < p < 1)$, em um conjunto de dados de tamanho n, é o valor da variável que ocupa a posição $p * (n + 1)$ do conjunto de dados **ordenados.**

Casos particulares:

Quartis: dados divididos em 4 partes iguais

- 25^0 percentil = Primeiro quartil ($Q1$);
- 50^0 Percentil = Segundo quartil ($Q2$) = Mediana;
- 75^0 Percentil = Terceiro quartil ($Q3$).

Decis: dados divididos em 10 partes iguais

- 10^0 percentil = Primeiro decil;
- 60^0 Percentil = Sexto decil.

Percentis

Exemplo: Nota de alunos em um curso (n=10)

Dados	5, 4, 7, 9, 10, 5, 5, 8, 4, 7

- Posição de $Q1 = 0,25 * (n+1) = 0,25 * 11 = 2,75$

Dados ordenados	4, **4, 5**, 5, 5, 7, 7, 8, 9, 10

$Q1 = 4,5$

- Posição de $Q2 = 0,5 * (n+1) = 0,5 * 11 = 5,5$

Dados ordenados	4, 4, 5, 5, **5, 7**, 7, 8, 9, 10

$Q2 = 6$

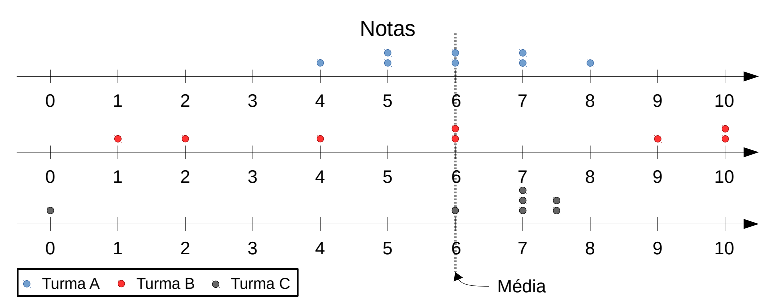
- Posição de $Q3 = 0,75 * (n+1) = 0,75 * 11 = 8,25$

Dados ordenados	4, 4, 5, 5, 5, 7, 7, **8, 9**, 10

$Q3 = 8,5$

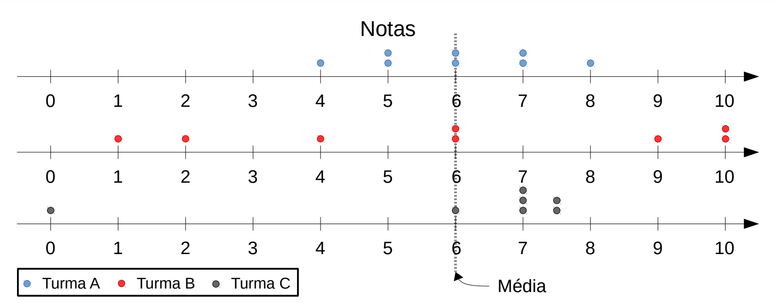
Sumário

Medidas de dispersão



Embora a média seja a mesma, os dados não apresentam a mesma característica.

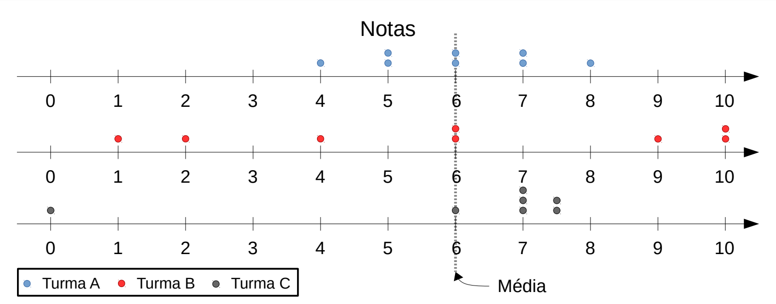
Medidas de dispersão



Embora a média seja a mesma, os dados não apresentam a mesma característica.

Por exemplo: as notas da Turma A variam menos que as da Turma B, isto é, as notas da Turma A apresentam **menor** dispersão do que as da Turma B.

Medidas de dispersão



Embora a média seja a mesma, os dados não apresentam a mesma característica.

Por exemplo: as notas da Turma A variam menos que as da Turma B, isto é, as notas da Turma A apresentam **menor** dispersão do que as da Turma B.

Como quantificar a dispersão?

Medidas de dispersão

Principal objetivo:

- uma medida para representar quão disperso os dados estão.

Medidas de dispersão

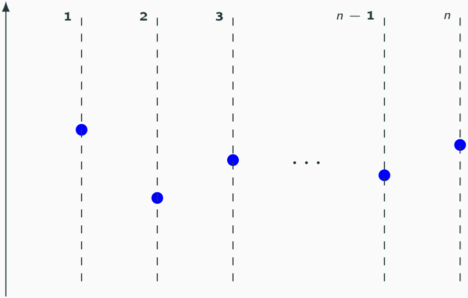
Principal objetivo:

- uma medida para representar quão disperso os dados estão.

Medidas apresentadas:

- amplitude;
- intervalo interquartil;
- desvio absoluto médio;
- variância;
- desvio-padrão;
- coeficiente de variação.

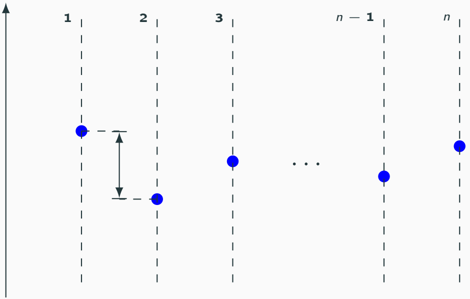
Amplitude



Elementos da amostra: $x_1, \ldots, x_n$.
Elementos ordenados: $x_{(1)}, \ldots, x_{(n)}$.
A amplitude da amostra é definida por

$$\text{Amp} = x_{(n)} - x_{(1)}.$$

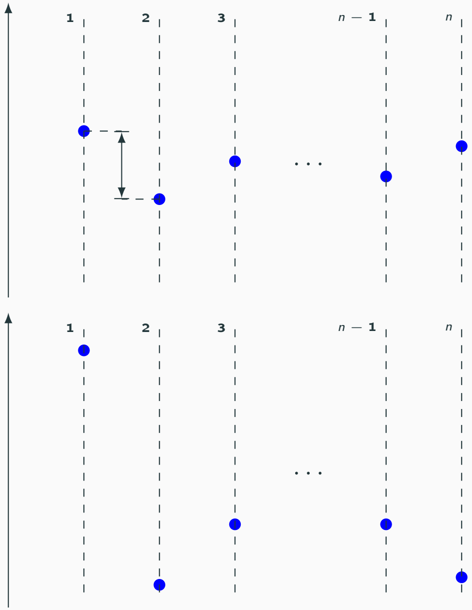
Amplitude



Elementos da amostra: $x_1, \ldots, x_n$.
Elementos ordenados: $x_{(1)}, \ldots, x_{(n)}$.
A amplitude da amostra é definida por

$$\text{Amp} = x_{(n)} - x_{(1)}.$$

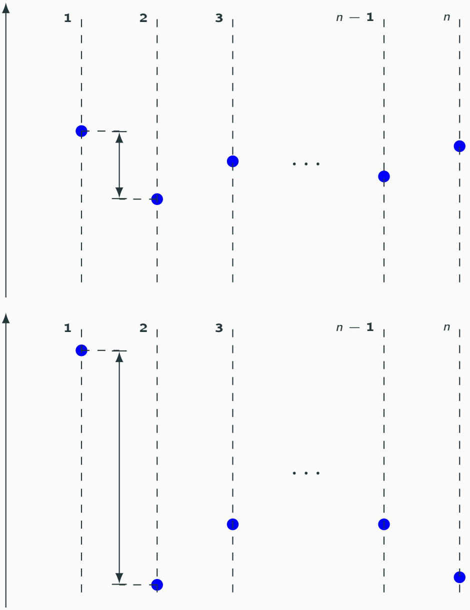
Amplitude



Elementos da amostra: $x_1, \ldots, x_n$.
Elementos ordenados: $x_{(1)}, \ldots, x_{(n)}$.
A amplitude da amostra é definida por

$$\text{Amp} = x_{(n)} - x_{(1)}.$$

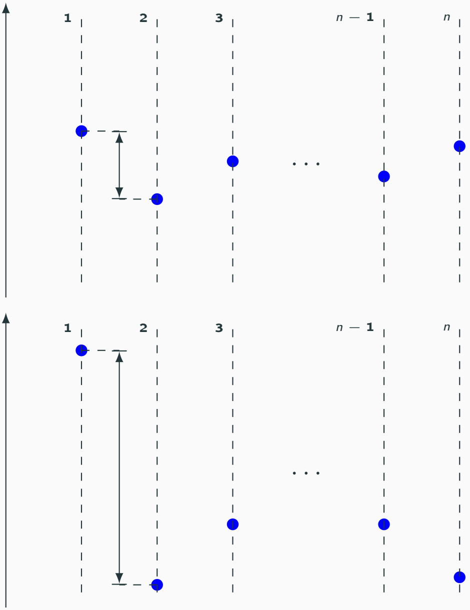
Amplitude



Elementos da amostra: $x_1, \ldots, x_n$.
Elementos ordenados: $x_{(1)}, \ldots, x_{(n)}$.
A amplitude da amostra é definida por

$$\text{Amp} = x_{(n)} - x_{(1)}.$$

Amplitude

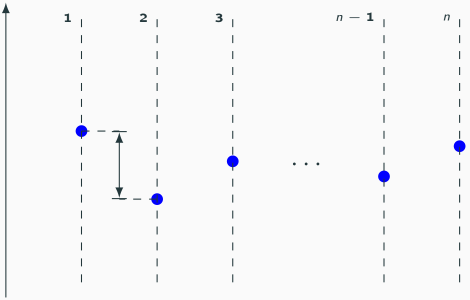


Elementos da amostra: $x_1, \ldots, x_n$.
Elementos ordenados: $x_{(1)}, \ldots, x_{(n)}$.
A amplitude da amostra é definida por

$$\text{Amp} = x_{(n)} - x_{(1)}.$$

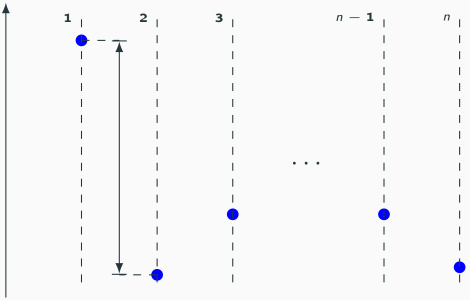
Comentários:

Amplitude



Elementos da amostra: $x_1, \ldots, x_n$.
Elementos ordenados: $x_{(1)}, \ldots, x_{(n)}$.
A amplitude da amostra é definida por

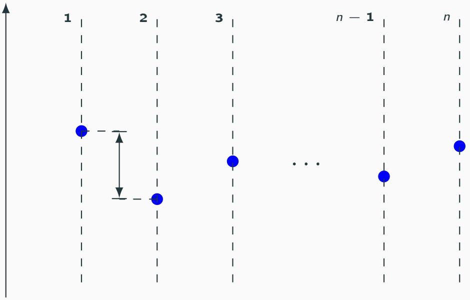
$$\text{Amp} = x_{(n)} - x_{(1)}.$$

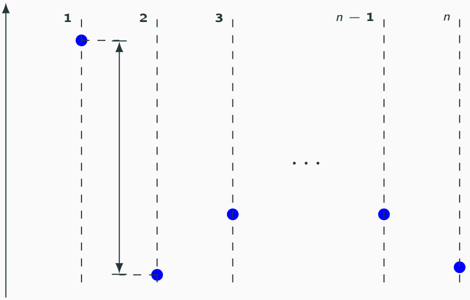


Comentários:

- cálculo rápido;

Amplitude





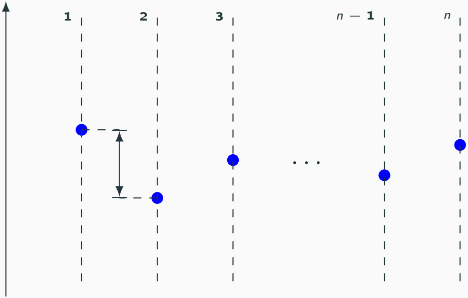
Elementos da amostra: $x_1, \ldots, x_n$.
Elementos ordenados: $x_{(1)}, \ldots, x_{(n)}$.
A amplitude da amostra é definida por

$$\text{Amp} = x_{(n)} - x_{(1)}.$$

Comentários:

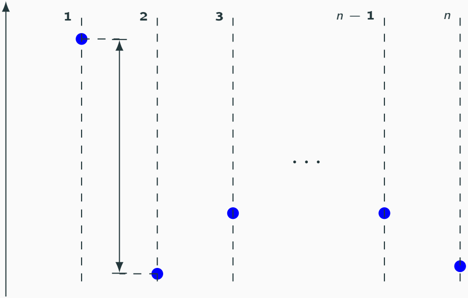
- cálculo rápido;
- fácil interpretação;

Amplitude



Elementos da amostra: $x_1, \ldots, x_n$.
Elementos ordenados: $x_{(1)}, \ldots, x_{(n)}$.
A amplitude da amostra é definida por

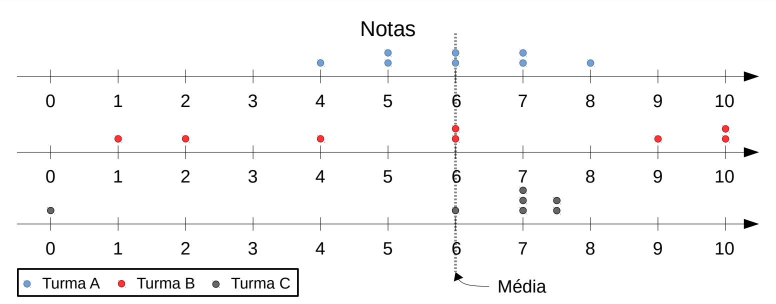
$$\text{Amp} = x_{(n)} - x_{(1)}.$$



Comentários:

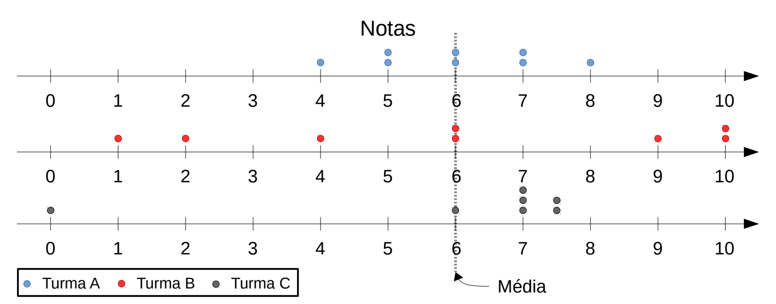
- cálculo rápido;
- fácil interpretação;
- alto impacto de dados atípicos.

Amplitude



Amplitudes:

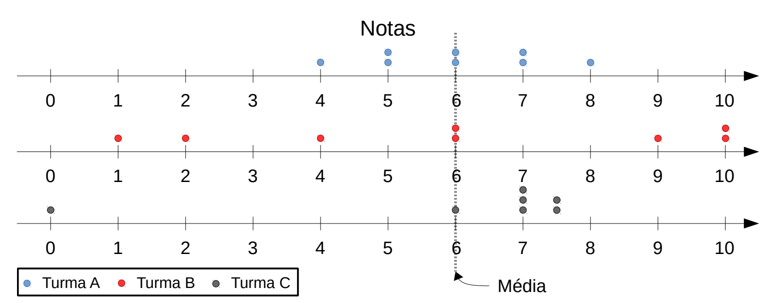
Amplitude



Amplitudes:

- Turma A: $\text{Amp} = x_{(8)} - x_{(1)} = 8 - 4 = 4$;

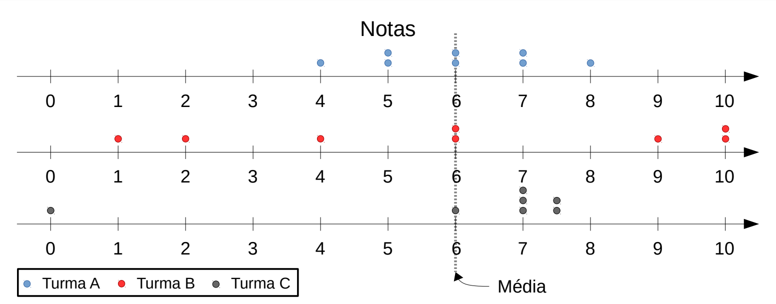
Amplitude



Amplitudes:

- Turma A: $\text{Amp} = x_{(8)} - x_{(1)} = 8 - 4 = 4$;
- Turma B: $\text{Amp} = x_{(8)} - x_{(1)} = 10 - 1 = 9$;

Amplitude



Amplitudes:

- Turma A: $\text{Amp} = x_{(8)} - x_{(1)} = 8 - 4 = 4$;
- Turma B: $\text{Amp} = x_{(8)} - x_{(1)} = 10 - 1 = 9$;
- Turma C: $\text{Amp} = x_{(7)} - x_{(1)} = 7.5 - 0 = 7.5$;

Intervalo interquartil

É a diferença entre o terceiro quartil e o primeiro quartil, ou seja,

$$IQ = Q3 - Q1$$

.

Intervalo interquartil

É a diferença entre o terceiro quartil e o primeiro quartil, ou seja,

$$IQ = Q3 - Q1$$

.

Exemplo: As notas de um teste de 3 grupos de alunos

Grupo 1: 3, 4, 5, 6, 7	$Q1 = 3,5$	$Q3 = 6,5$	$IQ_1 = 6,5 - 3,5 = 3$
Grupo 2: 1, 3, 5, 7, 9	$Q1 = 2$	$Q3 = 8$	$IQ_2 = 8 - 2 = 6$
Grupo 3: 5, 5 ,5, 5, 5	$Q1 = 5$	$Q3 = 5$	$IQ_3 = 5 - 5 = 0$

Desvio absoluto médio

Ideia: em vez de considerar apenas dois valores da amostra (mínimo e máximo), utilizar **todos**!

Desvio absoluto médio

Ideia: em vez de considerar apenas dois valores da amostra (mínimo e máximo), utilizar **todos**!

Desvio (da média): $d_i = x_i - \bar{x}$.

Desvio absoluto médio

Ideia: em vez de considerar apenas dois valores da amostra (mínimo e máximo), utilizar **todos**!

Desvio (da média): $d_i = x_i - \bar{x}$. Pelas propriedades da média, a soma dos desvios é **sempre** zero.

Desvio absoluto médio

Ideia: em vez de considerar apenas dois valores da amostra (mínimo e máximo), utilizar **todos**!

Desvio (da média): $d_i = x_i - \bar{x}$. Pelas propriedades da média, a soma dos desvios é **sempre** zero.

Desvio absoluto: $D_i = |d_i| = |x_i - \bar{x}|$.

Desvio absoluto médio

Ideia: em vez de considerar apenas dois valores da amostra (mínimo e máximo), utilizar **todos**!

Desvio (da média): $d_i = x_i - \bar{x}$. Pelas propriedades da média, a soma dos desvios é **sempre** zero.

Desvio absoluto: $D_i = |d_i| = |x_i - \bar{x}|$. Evita que desvios negativos cancelem os positivos

Desvio absoluto médio

Ideia: em vez de considerar apenas dois valores da amostra (mínimo e máximo), utilizar **todos**!

Desvio (da média): $d_i = x_i - \bar{x}$. Pelas propriedades da média, a soma dos desvios é **sempre** zero.

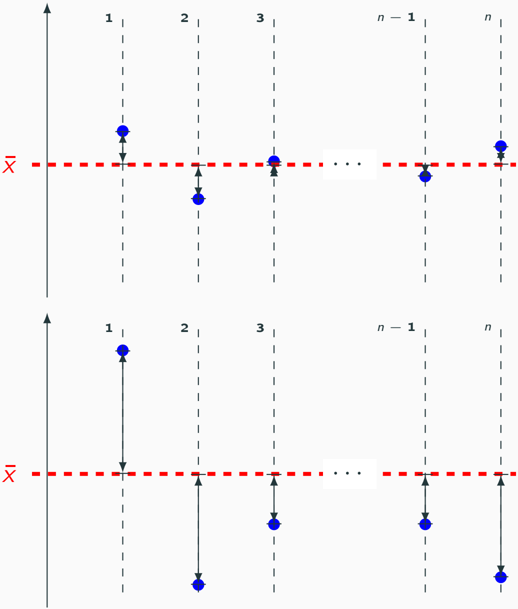
Desvio absoluto: $D_i = |d_i| = |x_i - \bar{x}|$. Evita que desvios negativos cancelem os positivos

Desvio-médio:

$$\mathrm{DM} = \frac{\sum_{i=1}^{n} D_i}{n} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n},$$

isto é, a média dos desvios absolutos.

Variância



Mesma ideia do desvio absoluto médio.

Desvio quadrado: $D_i^* = (x_i - \bar{x})^2$.

Variância amostral:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} D_i^*.$$

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

Em geral, se divide por $n-1$ ao invés de n para ter boas propriedades inferenciais.

Desvio-padrão

- Se os dados são expressos em cm, por exemplo, a variância é em cm^2.
- Para remediar o fato de a variância ser expressa na unidade de medida da variável ao quadrado, podemos calcular o **desvio-padrão**.

O desvio-padrão amostral é dado por

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

A grande vantagem do desvio-padrão é que ele é expresso na mesma unidade de medida dos dados.

Coeficiente de variação

Problema: As medidas de dispersão apresentadas são influenciadas pela "grandeza" da variável estudada.

É comum então representar o desvio-padrão como percentual da média. Denomina-se essa medida de coeficiente de variação.

O coeficiente de variação amostral é dado por

$$\text{CV} = \frac{s}{\bar{x}},$$

onde $\bar{x}$ e s são a média e o desvio-padrão amostrais.

Vantagem: Duas populações com médias muito diferentes podem ter suas dispersões comparadas através do coeficiente de variação.

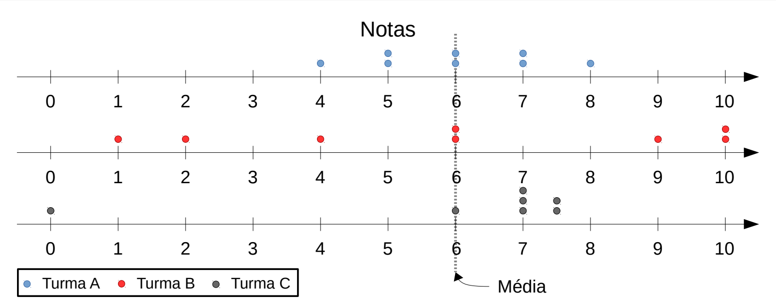
Coeficiente de variação

Exemplo: Altura (em cm) de uma amostra de recém-nascidos e de uma amostra de adolescentes

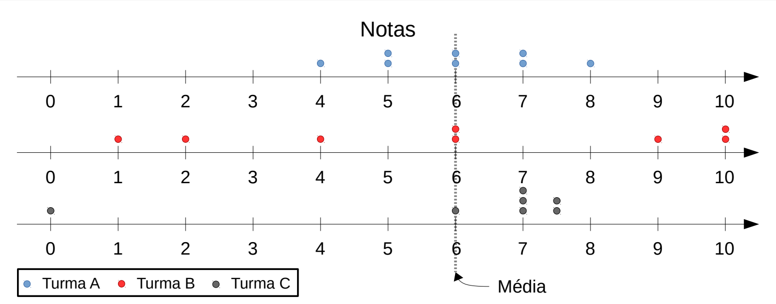
	Média	DP	CV(%)
Rrecém-nascidos	50	6	12%
Adolescentes	160	16	10%

Em relação às médias, as alturas dos adolescentes e dos recém-nascidos apresentam variabilidades quase iguais.

Exemplos


Notas
0 1 2 3 4 5 6 7 8 9 10
0 1 2 3 4 5 6 7 8 9 10
0 1 2 3 4 5 6 7 8 9 10
Turma A
Turma B
Turma C
Média

Exemplos



Medidas:

Turma	Amp	IQ	DM	s	CV
A	4	2	1	1.31	0.218 (21.8%)
B	9	6.5	2.75	3.505	0.584 (58.4%)
C	7.5	1.5	1.714	2.693	0.449 (44.9%)