Probabilidade e Estatística

Professores:

Otton Teixeira da Silveira Filho

Regina Célia Paula Leal Toledo

Fundação CECIERJ
Consórcio cederj

Lembrando...vimos na aula 2 - "Organização dos dados: tabelas de freqüência e gráficos"

Tabelas de freqüência

- freqüência absoluta - n_i
- freqüência relativa - f_i
- freqüência acumulada - f_{ac}

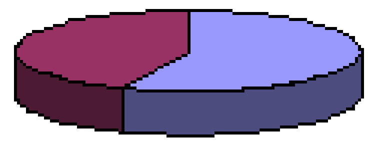
Lembrando...vimos na aula 2 - "Organização dos dados: tabelas de freqüência e gráficos"

Tabelas de freqüência

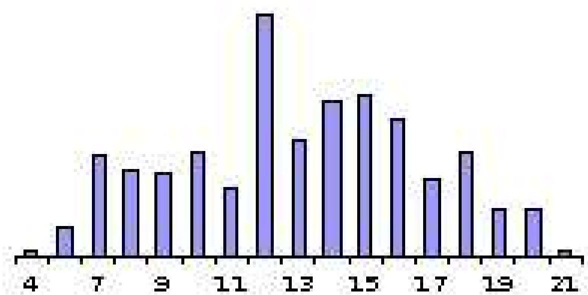
- freqüência absoluta - n_i
- freqüência relativa - f_i
- freqüência acumulada - f_{ac}

Gráficos

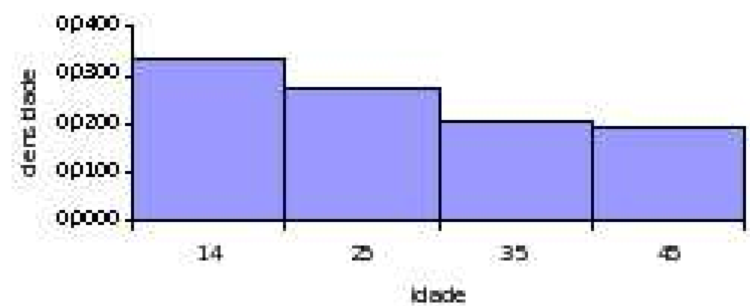
- Pizza, disco ou diagrama circular



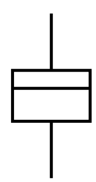
- Barra



- Histograma



- Box Plot



Aula 3

Professores:

Otton Teixeira da Silveira Filho
Regina Célia Paula Leal Toledo

Medidas de resumo (para um conjunto de dados)

Conteúdo:

3.1 Medidas de posição

3.2 Algumas questões complementares

3.3 Exemplo

3.4 Medidas de dispersão

$$\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$\sigma_x^2 \quad \overline{Y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

$$A \cap B = \varnothing$$

3. Medidas de resumo para um conjunto de dados

Consiste em fornecer informações numéricas sobre um conjunto de dados que possam representar por exemplo a tendência central desses dados ou a forma como estão dispersos.

3. Medidas de resumo para um conjunto de dados

Consiste em fornecer informações numéricas sobre um conjunto de dados que possam representar por exemplo a tendência central desses dados ou a forma como estão dispersos.

3.1 Medidas de posição (tendência central)

medidas mais comuns:

- média
- mediana
- moda

A média $(\overline{x}_{obs})$ fornece a média aritmética de um conjunto de dados, ou seja, dada uma seqüência de valores: $(x_1, x_2, \cdots, x_n)$ a média é escrita como:

A média $(\overline{x}_{obs})$ fornece a média aritmética de um conjunto de dados, ou seja, dada uma seqüência de valores: $(x_1, x_2, \cdots, x_n)$ a média é escrita como:

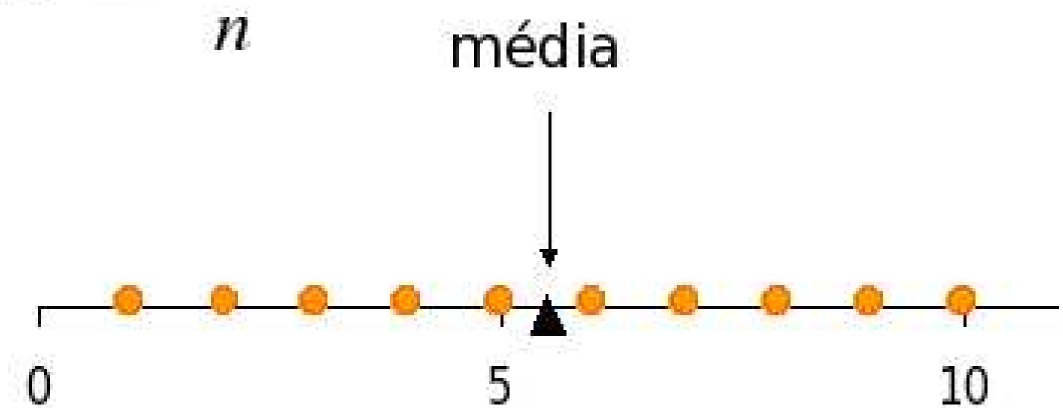
$$\overline{x}_{obs} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

A média $(\overline{x}_{obs})$ fornece a média aritmética de um conjunto de dados, ou seja, dada uma seqüência de valores: $(x_1, x_2, \cdots, x_n)$ a média é escrita como:

$$\overline{x}_{obs} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

ou

$$\overline{x}_{obs} = \frac{\sum_{i=1}^{n} x_i}{n}$$



Exemplo

Em uma construção comprou-se parafusos para instalação das tomadas. Está descrito na embalagem que cada caixa contém 100 parafusos. 10 caixas foram compradas e o número de parafusos contados, obtendo:

Número da caixa	1	2	3	4	5	6	7	8	9	10
Número de parafusos	98	102	100	100	99	97	96	95	99	100

O número médio de parafusos por caixa está acima ou abaixo do número informado na embalagem?

Exemplo

Em uma construção comprou-se parafusos para instalação das tomadas. Está descrito na embalagem que cada caixa contém 100 parafusos. 10 caixas foram compradas e o número de parafusos contados, obtendo:

Número da caixa	1	2	3	4	5	6	7	8	9	10
Número de parafusos	98	102	100	100	99	97	96	95	99	100

O número médio de parafusos por caixa está acima ou abaixo do número informado na embalagem?

$$\bar{x}_{obs} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Exemplo

Em uma construção comprou-se parafusos para instalação das tomadas. Está descrito na embalagem que cada caixa contém 100 parafusos. 10 caixas foram compradas e o número de parafusos contados, obtendo:

Número da caixa	1	2	3	4	5	6	7	8	9	10
Número de parafusos	98	102	100	100	99	97	96	95	99	100

O número médio de parafusos por caixa está acima ou abaixo do número informado na embalagem?

$$\boxed{\bar{x}_{obs} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{98+102+100+100+99+97+96+95+99+100}{10} = \frac{986}{10}}$$

Exemplo

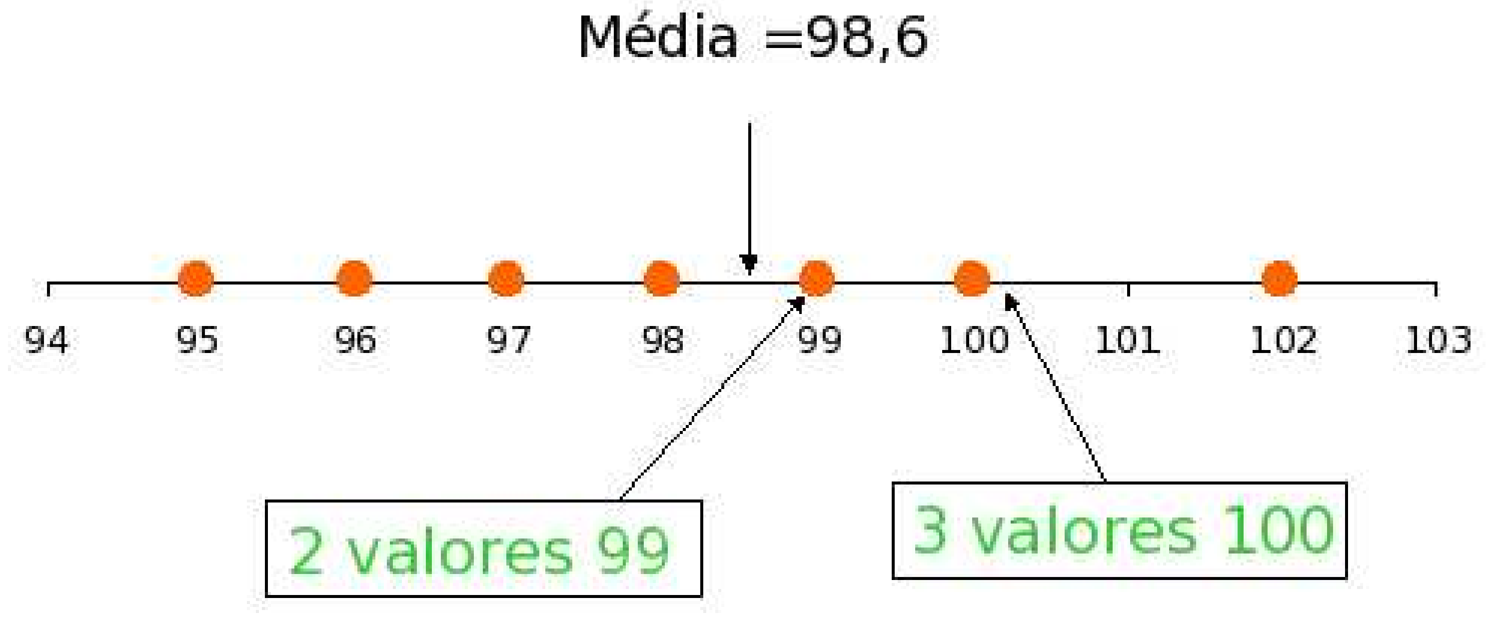
Número da caixa	1	2	3	4	5	6	7	8	9	10
Número de parafusos	98	102	100	100	99	97	96	95	99	100

$$\bar{x}_{obs} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{986}{10} = 98,6$$

Exemplo

Número da caixa	1	2	3	4	5	6	7	8	9	10
Número de parafusos	98	102	100	100	99	97	96	95	99	100

$$\bar{x}_{obs} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{986}{10} = 98,6$$



Outro Exemplo

Vamos supor que os 11 alunos da disciplina Probabilidade e Estatística obtiveram as seguintes notas em uma prova:

notas:	0,0	2,0	6,0	6,1	6,1	6,2	6,5	6,5	6,5	7,1	8,6

Qual a média da turma?

Outro Exemplo

Vamos supor que os 11 alunos da disciplina Probabilidade e Estatística obtiveram as seguintes notas em uma prova:

notas:	0,0	2,0	6,0	6,1	6,1	6,2	6,5	6,5	6,5	7,1	8,6

Qual a média da turma?

A média da turma é dado por:

$$\boxed{\bar{x}_{obs} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

temos

$$\bar{x}_{obs} = \frac{0,0 + 2,0 + 6,0 + 6,1 + 6,1 + 6,2 + 6,5 + 6,5 + 6,5 + 7,1 + 8,6}{11}$$
$$= 5,6$$

Outro Exemplo

Vamos supor que os 11 alunos da disciplina Probabilidade e Estatística obtiveram as seguintes notas em uma prova:

notas:	0,0	2,0	6,0	6,1	6,1	6,2	6,5	6,5	6,5	7,1	8,6

Qual a média da turma?

A média da turma é dado por:

$$\bar{x}_{obs} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

temos

$$\boxed{\begin{aligned} \bar{x}_{obs} &= \frac{0,0+2,0+6,0+6,1+6,1+6,2+6,5+6,5+6,5+7,1+8,6}{11} \\ &= 5,6 \end{aligned}}$$

Para um conjunto de dados organizados em uma tabela de freqüências onde cada x_i está relacionado a sua freqüência de ocorrência n_i, a média pode ser calculada como:

Para um conjunto de dados organizados em uma tabe-la de freqüências onde cada x_i está relacionado a sua freqüência de ocorrência n_i, a média pode ser calculada como:

$$\bar{x}_{obs} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_k x_k}{n_1 + n_2 + \ldots + n_k} \quad \text{ou} \quad \bar{x}_{obs} = \frac{\sum_{i=1}^{k} n_i x_i}{n} = \sum_{i=1}^{k} \frac{n_i}{n} x_i$$

Para um conjunto de dados organizados em uma tabela de freqüências onde cada x_i está relacionado a sua freqüência de ocorrência n_i, a média pode ser calculada como:

$$\bar{x}_{obs} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_k x_k}{n_1 + n_2 + \ldots + n_k} \quad \text{ou} \quad \bar{x}_{obs} = \frac{\sum_{i=1}^{k} n_i x_i}{n} = \sum_{i=1}^{k} \frac{n_i}{n} x_i$$

Exemplo

A tabela de freqüência de notas da turma é dado por:

notas:	0,0	2,0	6,0	6,1	6,2	6,5	7,1	8,6	total
n_i	1	1	1	2	1	3	1	1	11

Para um conjunto de dados organizados em uma <u>tabela de freqüências</u> onde cada x_i está relacionado a sua <u>freqüência de ocorrência</u> n_i, a média pode ser calculada como:

$$\bar{x}_{obs} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_k x_k}{n_1 + n_2 + \ldots + n_k} \quad \text{ou} \quad \bar{x}_{obs} = \frac{\sum_{i=1}^{k} n_i x_i}{n} = \sum_{i=1}^{k} \frac{n_i}{n} x_i$$

Exemplo

A tabela de freqüência de notas da turma é dado por:

notas:	0,0	2,0	6,0	6,1	6,2	6,5	7,1	8,6	total
n_i	1	1	1	2	1	3	1	1	11

$$\begin{aligned}\bar{x}_{obs} &= \frac{1\times 0,0 + 1\times 2,0 + 1\times 6,0 + 2\times 6,1 + 1\times 6,2 + 3\times 6,5 + 1\times 7,1 + 1\times 8,6}{11} \\ &= 5,6\end{aligned}$$

Para o exemplo dos parafusos a tabelas de freqüências é dada por:

Número de parafusos	95	96	97	98	99	100	102	total
freqüencia - n_i	1	1	1	1	2	3	1	10 caixas

Para o exemplo dos parafusos a tabelas de freqüências é dada por:

Número de parafusos	95	96	97	98	99	100	102	total
freqüencia - n_i	1	1	1	1	2	3	1	10 caixas

E a média, utilizando essa tabela:

$$\boxed{\bar{x}_{obs} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_7 x_7}{n_1 + n_2 + \ldots + n_7}}$$

ou

$$\bar{x}_{obs} = \frac{\sum_{i=1}^{7} n_i x_i}{10} = \sum_{i=1}^{7} \frac{n_i}{10} x_i$$

Para o exemplo dos parafusos a tabelas de freqüências é dada por:

Número de parafusos	95	96	97	98	99	100	102	total
freqüencia - n_i	1	1	1	1	2	3	1	10 caixas

E a média, utilizando essa tabela:

$$\bar{x}_{obs} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_7 x_7}{n_1 + n_2 + \ldots + n_7}$$

ou

$$\boxed{\bar{x}_{obs} = \frac{\sum_{i=1}^{7} n_i x_i}{10} = \sum_{i=1}^{7} \frac{n_i}{10} x_i}$$

Para o exemplo dos parafusos a tabelas de freqüências é dada por:

Número de parafusos	95	96	97	98	99	100	102	total
freqüencia - n_i	1	1	1	1	2	3	1	10 caixas

E a média, utilizando essa tabela:

$$\bar{x}_{obs} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_7 x_7}{n_1 + n_2 + \ldots + n_7} \quad \text{ou} \quad \bar{x}_{obs} = \frac{\sum_{i=1}^{7} n_i x_i}{10} = \sum_{i=1}^{7} \frac{n_i}{10} x_i$$

$$\boxed{\bar{x}_{obs} = \frac{1\times 95 + 1\times 96 + 1\times 97 + 1\times 98 + 2\times 99 + 3\times 100 + 1\times 102}{1+1+1+1+2+3+1} = \frac{986}{10} = 98,6}$$

A mediana (md_{obs}) é o valor que ocupa a posição central nos dados ordenados.

No Exemplo das notas do curso de Probabilidade e Estatística,a tabela já ordenada, com as notas é:

notas:	0,0	2,0	6,0	6,1	6,1	6,2	6,5	6,5	6,5	7,1	8,6
posição	1	2	3	4	5	6	7	8	9	10	11

A mediana (md_{obs}) é o valor que ocupa a posição central nos dados ordenados.

No Exemplo das notas do curso de Probabilidade e Estatística,a tabela já ordenada, com as notas é:

notas:	0,0	2,0	6,0	6,1	6,1	6,2	6,5	6,5	6,5	7,1	8,6
posição	1	2	3	4	5	6	7	8	9	10	11

E o valor que ocupa a posição central será o que ocupa a posição 6, ou seja, a nota 6,2, o que significa que foram 5 notas abaixo de 6,2

A mediana (md_{obs}) é o valor que ocupa a posição central nos dados ordenados.

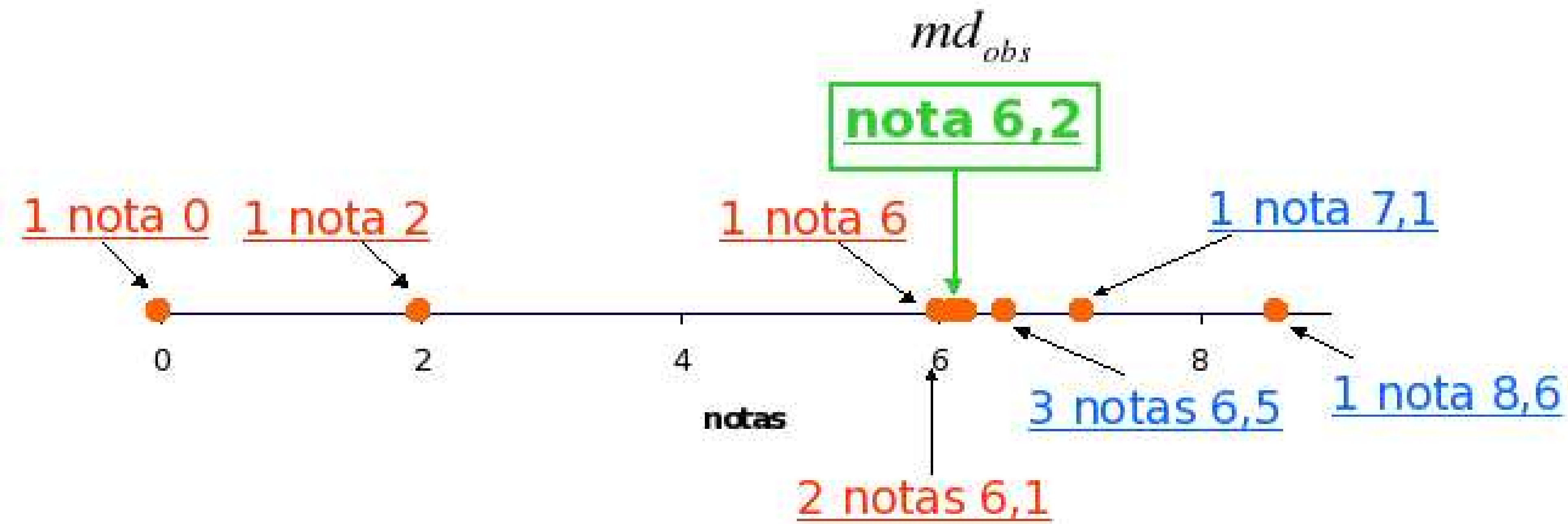
No Exemplo das notas do curso de Probabilidade e Estatística,a tabela já ordenada, com as notas é:

notas:	0,0	2,0	6,0	6,1	6,1	6,2	6,5	6,5	6,5	7,1	8,6
posição	1	2	3	4	5	6	7	8	9	10	11

E o valor que ocupa a posição central será o que ocupa a posição 6, ou seja, a nota 6,2, o que significa que foram 5 notas abaixo de 6,2 e 5 notas acima desse valor.

md_{obs}

notas:	0,0	2,0	6,0	6,1	6,1	6,2	6,5	6,5	6,5	7,1	8,6
posição	1	2	3	4	5	6	7	8	9	10	11



No Exemplo das caixas de parafusos, onde foram consideradas um número par de caixas, 10 caixas, a tabela ordenada é dada por:

Número de parafusos	95	96	97	98	99	99	100	100	100	102
Posição	1	2	3	4	5	6	7	8	9	10

O valor mediano é dado pela média entre o número de parafusos da caixa 5 e o número de parafusos da caixa 6.

No Exemplo das caixas de parafusos, onde foram consideradas um número par de caixas, 10 caixas, a tabela ordenada é dada por:

Número de parafusos	95	96	97	98	99	99	100	100	100	102
Posição	1	2	3	4	5	6	7	8	9	10

O valor mediano é dado pela média entre o número de parafusos da caixa 5 e o número de parafusos da caixa 6.

No exemplo, tanto a caixa 5 como a caixa 6 têm 99 parafusos e a média desses valores também será 99 parafusos!

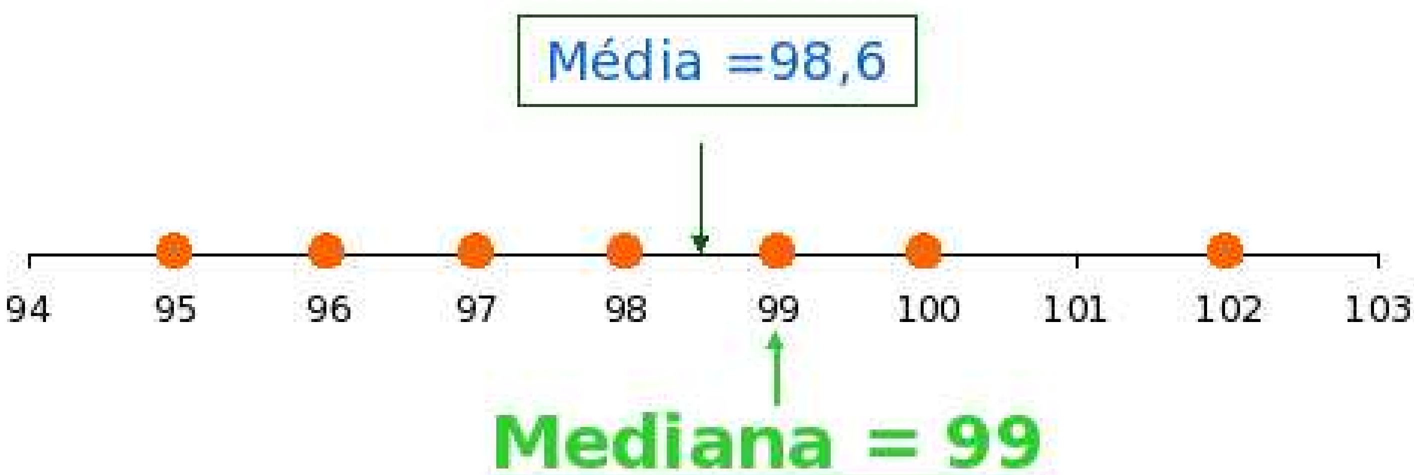
Fundação CECIERJ Consórcio cederj

No Exemplo das caixas de parafusos, onde foram consideradas um número par de caixas, 10 caixas, a tabela ordenada é dada por:

Número de parafusos	95	96	97	98	99	99	100	100	100	102
Posição	1	2	3	4	5	6	7	8	9	10

O valor mediano é dado pela média entre o número de parafusos da caixa 5 e o número de parafusos da caixa 6.

No exemplo, tanto a caixa 5 como a caixa 6 têm 99 parafusos e a média desses valores também será 99 parafusos!



Se o número de valores (n) for ímpar: $\boxed{md_{obs} = valorpos.\left(\frac{n+1}{2}\right)}$

Se o número de valores (n) for ímpar: $md_{obs} = valorpos.\left(\frac{n+1}{2}\right)$

Exemplo: tabela com 9 valores

1,5	2	2,5	3	3,5	5	6	7	8	n=9

$$
\begin{aligned}
md_{obs} &= valorpos\left(\frac{9+1}{2}\right) \\
&= valorpos(5) \\
&= 3,5
\end{aligned}
$$

Se o número de valores (n) for ímpar: $md_{obs} = valorpos.\left(\frac{n+1}{2}\right)$

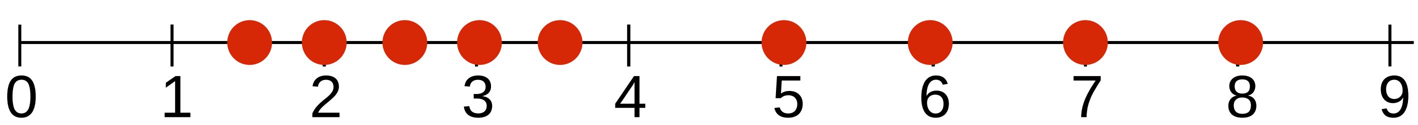
Exemplo: tabela com 9 valores

1,5	2	2,5	3	3,5	5	6	7	8	n=9

$$\begin{aligned} md_{obs} &= valorpos\left(\frac{9+1}{2}\right) \\ &= valorpos(5) \\ &= 3,5 \end{aligned}$$

Mediana

valor da posição 5, que é 3,5



Se o número de valores (n) for par:

$$md_{obs} = \frac{valorpos.\left(\frac{n}{2}\right) + valorpos.\left(\frac{n+2}{2}\right)}{2}$$

Se o número de valores (n) for par:

$$md_{obs} = \frac{valorpos.\left(\frac{n}{2}\right) + valorpos.\left(\frac{n+2}{2}\right)}{2}$$

Exemplo: tabela com 10 valores

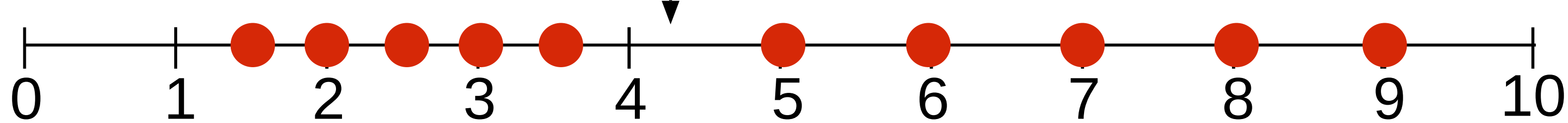
Se o número de valores (n) for par:

$$md_{obs} = \frac{valorpos.\left(\frac{n}{2}\right) + valorpos.\left(\frac{n+2}{2}\right)}{2}$$

Exemplo: tabela com 10 valores

1,5	2	2,5	3	3,5	5	6	7	8	9	n=10

mediana - valor médio das posições 5 (3,5) e 6 (5,0), ou seja, 4,25!



$$md_{obs} = \frac{valorpos(5) + valorpos(6)}{2} = \frac{3,5+5}{2} = \frac{8,5}{2} = 4,25$$

A moda (mo_{obs}) é o valor mais freqüente, ou seja, com maior ocorrência

A moda (mo_{obs}) é o valor mais freqüente, ou seja, com maior ocorrência

No Exemplo das notas do curso de Probabilidade e Estatística,a tabela de freqüencias, dada por :

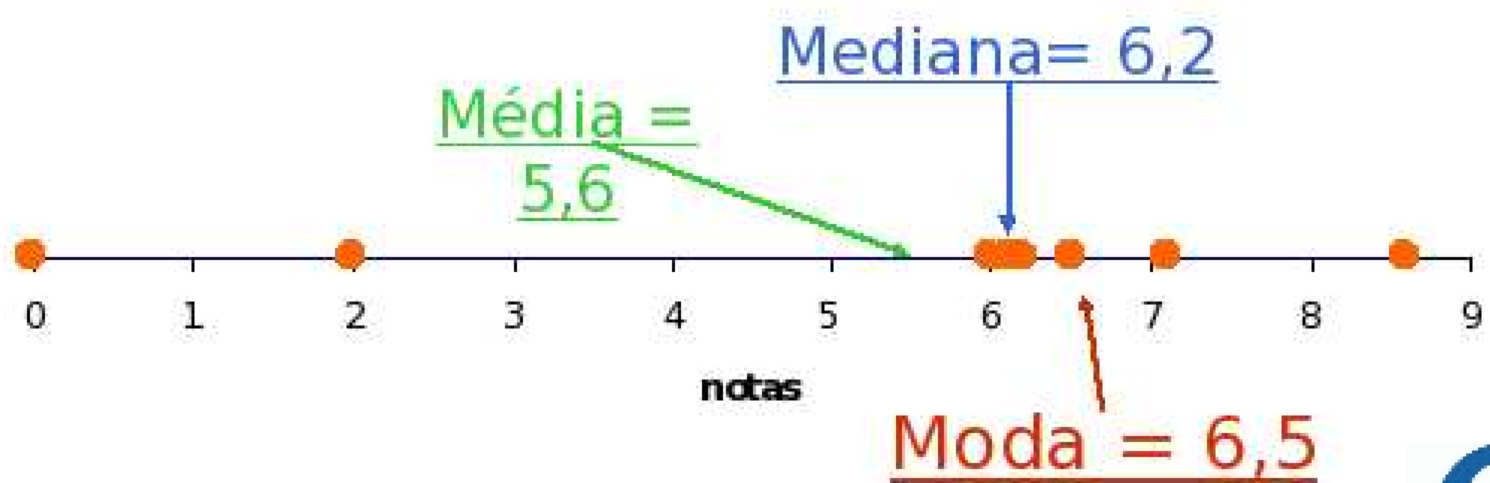
notas:	0,0	2,0	6,0	6,1	6,2	6,5	7,1	8,6	total
n_i	1	1	1	2	1	3	1	1	11

A moda (mo_{obs}) é o valor mais freqüente, ou seja, com maior ocorrência

No Exemplo das notas do curso de Probabilidade e Estatística,a tabela de freqüencias, dada por :

notas:	0,0	2,0	6,0	6,1	6,2	6,5	7,1	8,6	total
n_i	1	1	1	2	1	3	1	1	11

O valor que ocorre mais vezes é 6,5 ou seja, a moda é 6,5!



No Exemplo das caixas de parafusos, a tabela de freqüências, é dada por :

Número de parafusos	95	96	97	98	99	100	102	total
freqüencia - n_i	1	1	1	1	2	3	1	10 caixas

No Exemplo das caixas de parafusos, a tabela de freqüências, é dada por :

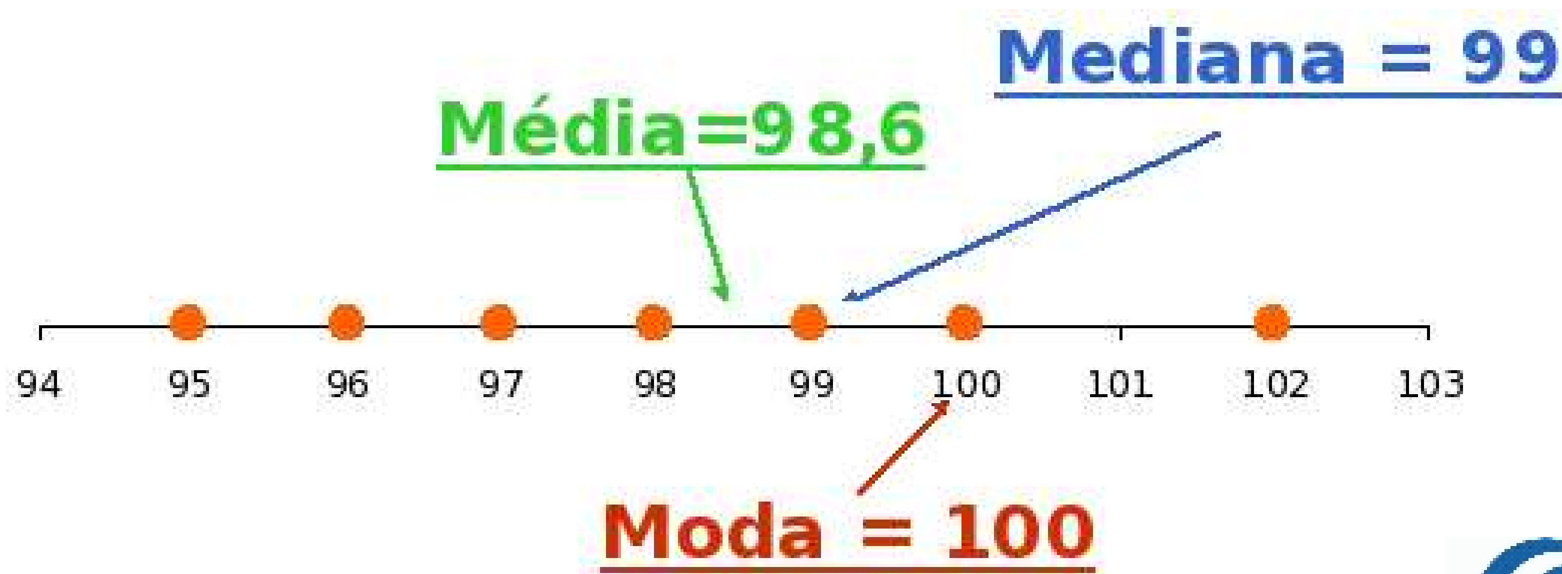
Número de parafusos	95	96	97	98	99	100	102	total
freqüencia - n_i	1	1	1	1	2	3	1	10 caixas

e o valor que ocorre o maior número de vezes é 100, ou seja, a moda é $mo_{obs} = 100$!

No Exemplo das caixas de parafusos, a tabela de freqüências, é dada por :

Número de parafusos	95	96	97	98	99	100	102	total
freqüencia - n_i	1	1	1	1	2	3	1	10 caixas

e o valor que ocorre o maior número de vezes é 100, ou seja, a moda é $mo_{obs} = 100$!



Algumas vezes todos os valores têm a mesma freqüência de ocorrência

não tem moda

Algumas vezes todos os valores têm a mesma freqüência de ocorrência

não tem moda

Outras vezes mais de um valor (k valores) têm a mesma freqüencia de ocorrência:

tem k modas (multimodal)

Algumas vezes todos os valores têm a mesma freqüência de ocorrência

não tem moda

Outras vezes mais de um valor (k valores) têm a mesma freqüencia de ocorrência:

tem k modas (multimodal)

Por exemplo, 2 valores têm a mesma freqüência de ocor-rência

tem 2 modas (bimodal)

3.2 Algumas questões complementares

Questão complementar 1- Outras medidas de posição

→ Quartis: dividem o conjunto de valores em 4 subgrupos com mesmo número de elementos.

3.2 Algumas questões complementares

Questão complementar 1- Outras medidas de posição

→ Quartis: dividem o conjunto de valores em 4 subgrupos com mesmo número de elementos.

Q1 - primeiro quartil, significa que 25% dos valores ordenados da tabela estão abaixo desse valor

Q3 - terceiro quartil, significa que 75% dos valores ordenados da tabela estão abaixo desse valor

3.2 Algumas questões complementares

Questão complementar 1- Outras medidas de posição

→ Quartis: dividem o conjunto de valores em 4 subgrupos com mesmo número de elementos.

Q1 - primeiro quartil, significa que 25% dos valores ordenados da tabela estão abaixo desse valor

Q3 - terceiro quartil, significa que 75% dos valores ordenados da tabela estão abaixo desse valor

→ Decis: dividem o conjunto de valores em 10 subgrupos com mesmo número de elementos.

No Exemplo das notas do curso de Probabilidade e Estatística,a tabela de notas, já ordenada, com 11 valores é:

notas:	0,0	2,0	6,0	6,1	6,1	6,2	6,5	6,5	6,5	7,1	8,6
posição	1	2	3	4	5	6	7	8	9	10	11

Q1 (25%-75%)

Q3 (75%-25%)

Mediana (50%-50%)

No Exemplo das notas do curso de Probabilidade e Estatística,a tabela de notas, já ordenada, com 11 valores é:

notas:	0,0	2,0	6,0	6,1	6,1	6,2	6,5	6,5	6,5	7,1	8,6
posição	1	2	3	4	5	6	7	8	9	10	11

Q1 (25%-75%)

Q3 (75%-25%)

Mediana (50%-50%)

$$Q_1 \leftarrow valorpos\left(\frac{1}{4} \text{ do conjunto}\right)$$

$$Q_3 \leftarrow valorpos\left(\frac{3}{4} \text{ do conjunto}\right)$$

No Exemplo das caixas de parafusos, considerando-se agora 12 caixas, onde a caixa incluída tem 99 parafusos, a tabela ordenada é dada por:

Mediana=99
(50%-50%)

caixa	1	2	3	4	5	6	7	8	9	10	11	12
Número de parafusos	95	96	97	98	99	99	99	100	100	100	101	102

No Exemplo das caixas de parafusos, considerando-se agora 12 caixas, onde a caixa incluída tem 99 parafusos, a tabela ordenada é dada por:

Mediana=99
(50%-50%)

caixa	1	2	3	4	5	6	7	8	9	10	11	12
Número de parafusos	95	96	97	98	99	99	99	100	100	100	101	102

Q1=97,5
(25%-75%)

Q3=100
(75%-25%)

Questão complementar 2

Muitas vezes o valor de interesse não pode ser observado diretamente da tabela, sendo uma função do conjunto original dos dados.

Exemplo: quer se saber o custo da caixa de parafusos já citada, sabendo que o custo de cada parafuso é c e o da caixa é e.

Número de parafusos	95	96	97	98	99	99	100	100	100	102

Questão complementar 2

Muitas vezes o valor de interesse não pode ser observado diretamente da tabela, sendo uma função do conjunto original dos dados.

Exemplo: quer se saber o custo da caixa de parafusos já citada, sabendo que o custo de cada parafuso é c e o da caixa é e.

Número de parafusos	95	96	97	98	99	99	100	100	100	102

Tabela reescrita:

Núm. de paraf	95c+e	96c+e	97c+e	98c+e	99c+e	99c+e	100c+e	100c+e	100c+e	102c+e

Exemplo: quer se saber o custo da caixa de parafusos já citada, sabendo que o custo de cada parafuso é c e o da caixa é e.

Número de parafusos	95	96	97	98	99	99	100	100	100	102

Tabela reescrita:

Núm. de paraf	95c+e	96c+e	97c+e	98c+e	99c+e	99c+e	100c+e	100c+e	100c+e	102c+e

Exemplo: quer se saber o custo da caixa de parafusos já citada, sabendo que o custo de cada parafuso é c e o da caixa é e.

Número de parafusos	95	96	97	98	99	99	100	100	100	102

Tabela reescrita:

Núm. de paraf	95c+e	96c+e	97c+e	98c+e	99c+e	99c+e	100c+e	100c+e	100c+e	102c+e

Para a média temos:

$$\overline{x}_{obs} = \frac{(95c+e)+(96c+e)+(97c+e)+2(99c+e)+3(100c+e)+(102c+e)}{10}$$

$$\overline{x}_{obs} = \frac{\sum_{i=1}^{10}[(x_i c)+e]}{10} = \frac{986c+10e}{10} = 98,6c+e$$

Para a mediana (md_{obs}):

custo caixa	95c+e	96c+e	97c+e	98c+e	99c+e	99c+e	100c+e	100c+e	100c+e	102c+e

$$\boxed{md_{obs} = 99c+e}$$

Para a mediana (md_{obs}):

custo caixa	95c+e	96c+e	97c+e	98c+e	99c+e	99c+e	100c+e	100c+e	100c+e	102c+e

$$\boxed{md_{obs} = 99c+e}$$

Para a moda (mo_{obs}):

custo caixa	95c+e	96c+e	97c+e	98c+e	99c+e	100c+e	102c+e
n_i	1	1	1	1	2	3	1

$$\boxed{mo_{obs} = 100c+e}$$

Fundação CECIERJ Consórcio cederj

3.3 Exemplo

Um estudante está procurando estágio para o próximo ano e tem que escolher entre duas companhias (A e B) para as quais foi aceito. Ele sabe que por 20horas semanais cada uma delas tem as seguintes características (em salários mínimos):

Companhia	A	B
Média	2,6	2,6
Mediana	2,0	2,0
Moda	1,5	2,0

A que conclusões podemos chegar para ajudar o estudante a escolher?

3.4 Medidas de dispersão

Apesar das medidas de tendência central darem informações sobre o conjunto de dados elas podem estar "escondendo" valiosas informações, como a variabilidade desses dados, ou seja, como eles estão distribuídos.

3.4 Medidas de dispersão

Apesar das medidas de tendência central darem informações sobre o conjunto de dados elas podem estar "escondendo" valiosas informações, como a variabilidade desses dados, ou seja, como eles estão distribuídos.

Exemplo: seja a tabela 1, abaixo, com 5 valores

1,0	2,0	3,0	4,0	5,0

com mediana $\boxed{md_{obs} = 3{,}0}$ e a média $\boxed{\overline{x}_{obs} = \frac{1,0+2,0+3,0+4,0+5,0}{5} = 3,0}$

3.4 Medidas de dispersão

Apesar das medidas de tendência central darem informações sobre o conjunto de dados elas podem estar "escondendo" valiosas informações, como a variabilidade desses dados, ou seja, como eles estão distribuídos.

Exemplo: seja a tabela 1, abaixo, com 5 valores

1,0	2,0	3,0	4,0	5,0

com mediana $md_{obs} = 3,0$ e a média $\overline{x}_{obs} = \frac{1,0+2,0+3,0+4,0+5,0}{5} = 3,0$

E seja a tabela 2, a seguir também com 5 valores

0,5	1,0	3,0	3,5	7,0

com mediana $md_{obs} = 3,0$ e a média $\overline{x}_{obs} = \frac{0,5+1,0+3,0+3,5+7,0}{5} = \frac{15,0}{5} = 3$

tabela 1

1,0	2,0	3,0	4,0	5,0

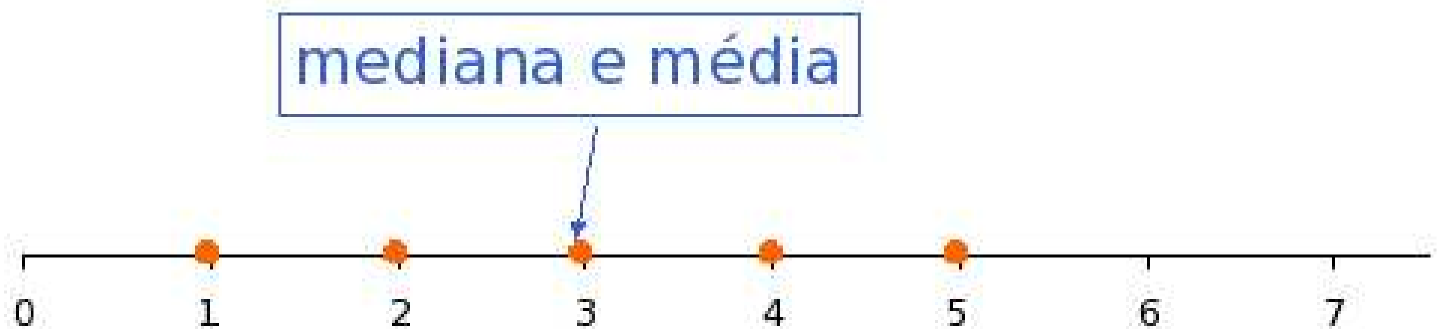


tabela 1

1,0	2,0	3,0	4,0	5,0

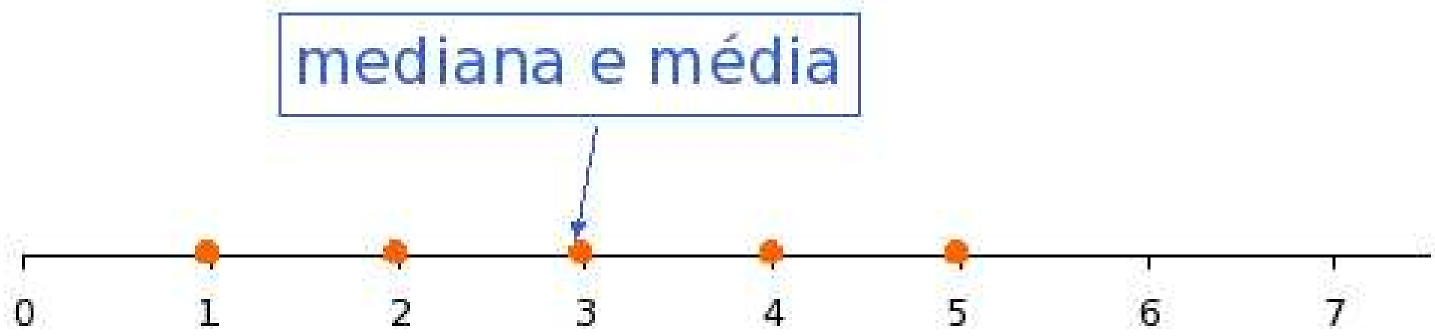


tabela 2

0,5	1,0	3,0	3,5	7,0

Observe quais as informações que as medidas de tendência central não fornecem

Exemplo do estudante que deseja escolher a companhia para fazer estágio: considerando que as duas empresas têm 51 estagiários $\rightarrow$ mediana é o valor 26 da tabela ordenada.

Exemplo do estudante que deseja escolher a companhia para fazer estágio: considerando que as duas empresas têm 51 estagiários $\rightarrow$ mediana é o valor 26 da tabela ordenada.

Companhia A

Média=2,6

Num. de sal. min.	1,0	1,5	2,0	3,0	7,0	9,0	12,0	total
freqüência	9	**16**	12	8	3	2	1	51

Moda=1,5

Mediana=2,0 (valor 26)

Exemplo do estudante que deseja escolher a companhia para fazer estágio: considerando que as duas empresas têm 51 estagiários $\rightarrow$ mediana é o valor 26 da tabela ordenada.

Companhia A

Média=2,6

Num. de sal. min.	1,0	1,5	2,0	3,0	7,0	9,0	12,0	total
freqüência	9	**16**	12	8	3	2	1	51

Moda=1,5

Mediana=2,0 (valor 26)

Companhia B

Média=2,6

Num. de sal. min.	1,0	2,0	3,0	5,0	7,0	total
freqüência	14	**17**	11	6	3	51

Moda=2,0

Mediana=2,0 (valor 26)

tabela 1	1,0	2,0	3,0	4,0	5,0
tabela 2	0,5	1,0	3,0	3,5	7,0

Medidas de dispersão

- amplitude
- desvio mediano
- desvio médio
- variância
- desvio padrão

Fundação CECIERJ Consórcio cederj

tabela 1	1,0	2,0	3,0	4,0	5,0

tabela 2	0,5	1,0	3,0	3,5	7,0

Medidas de dispersão

- amplitude
- desvio mediano
- desvio médio
- variância
- desvio padrão

1 - Amplitude, referente a uma determinada variável, é definida como a diferença entre o maior e o menor valor dessa variável no conjunto de dados. Será denotada por Δ.

tabela 1	1,0	2,0	3,0	4,0	5,0
tabela 2	0,5	1,0	3,0	3,5	7,0

Medidas de dispersão
- amplitude
- desvio mediano
- desvio médio
- variância
- desvio padrão

1 - Amplitude, referente a uma determinada variável, é definida como a diferença entre o maior e o menor valor dessa variável no conjunto de dados. Será denotada por Δ.

$\boxed{\Delta_{(tab.1)} = 5{,}0 - 1{,}0 = 4{,}0}$ $\boxed{\Delta_{(tab.2)} = 7{,}0 - 0{,}5 = 6{,}5}$

Fundação CECIERJ Consórcio cederj

tabela 1	1,0	2,0	3,0	4,0	5,0
tabela 2	0,5	1,0	3,0	3,5	7,0

Como a amplitude só leva em conta os valores extremos, uma outra idéia é considerar o desvio em relação a alguma medida e encontrar sua média.

tabela 1	1,0	2,0	3,0	4,0	5,0
tabela 2	0,5	1,0	3,0	3,5	7,0

Como a amplitude só leva em conta os valores extremos, uma outra idéia é considerar o desvio em relação a alguma medida e encontrar sua média.

2 - Desvio mediano - definido como o somatório dos módulos da distância de cada valor até a mediana.

$$desvio \cdot mediano = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - md_{obs}\right|$$

tabela 1

1,0	2,0	3,0	4,0	5,0

tabela 2

0,5	1,0	3,0	3,5	7,0

$$desvio \cdot mediano = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - md_{obs}\right|$$

Tabela 1

$$\begin{aligned} desvio \cdot mediano &= \frac{1}{5}(|1,0-3,0| + |2,0-3,0| + \\ &+ |3,0-3,0| + |4,0-3,0| + |5,0-3,0|) \\ desvio \cdot mediano &= \frac{1}{5}(2,0+1,0+0,0+1,0+2,0) = \frac{6,0}{5} \end{aligned}$$

tabela 1	1,0	2,0	3,0	4,0	5,0

tabela 2	0,5	1,0	3,0	3,5	7,0

$$desvio \cdot mediano = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - md_{obs}\right|$$

Tabela 1 $$desvio \cdot mediano = \frac{6,0}{5} = 1,2$$

Tabela 2

$$\begin{aligned} desvio \cdot mediano &= \frac{1}{5}(|0,5-3,0| + |1,0-3,0| + \\ &+ |3,0-3,0| + |3,5-3,0| + |7,0-3,0|) \end{aligned}$$

$$desvio \cdot mediano = \frac{1}{5}(2,5+2,0+0,0+0,5+4,0) = \frac{9,0}{5}$$

tabela 1

1,0	2,0	3,0	4,0	5,0

tabela 2

0,5	1,0	3,0	3,5	7,0

$$desvio \cdot mediano = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - md_{obs}\right|$$

Tabela 1

$$desvio \cdot mediano = \frac{6,0}{5} = 1,2$$

Tabela 2

$$desvio \cdot mediano = \frac{9,0}{5} = 1,8$$

tabela 1	1,0	2,0	3,0	4,0	5,0
tabela 2	0,5	1,0	3,0	3,5	7,0

3 - Desvio médio - definido como o somatório dos módulos da distância de cada valor até a média.

$$\boxed{desvio \cdot medio = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \bar{x}_{obs}\right|}$$

tabela 1	1,0	2,0	3,0	4,0	5,0

tabela 2	0,5	1,0	3,0	3,5	7,0

3 - Desvio médio - definido como o somatório dos módulos da distância de cada valor até a média.

$$desvio \cdot medio = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \bar{x}_{obs}\right|$$

Tabela 1

$$\begin{aligned} desvio \cdot medio &= \frac{1}{5}(|1,0-3,0| + |2,0-3,0| + \\ &+ |3,0-3,0| + |4,0-3,0| + |5,0-3,0|) \\ desvio \cdot medio &= \frac{1}{5}(2,0+1,0+0,0+1,0+2,0) = \frac{6,0}{5} \end{aligned}$$

tabela 1	1,0	2,0	3,0	4,0	5,0
tabela 2	0,5	1,0	3,0	3,5	7,0

3 - Desvio médio - definido como o somatório dos módulos da distância de cada valor até a média.

$$desvio \cdot medio = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \bar{x}_{obs}\right|$$

Tabela 1

$$desvio \cdot medio = \frac{6,0}{5} = 1,2$$

Tabela 2

$$\begin{aligned} desvio \cdot medio &= \frac{1}{5}(|0,5-3,0| + |1,0-3,0| + \\ &+ |3,0-3,0| + |3,5-3,0| + |7,0-3,0|) \end{aligned}$$

$$desvio \cdot medio = \frac{1}{5}(2,5+2,0+0,0+0,5+4,0) = \frac{9,0}{5}$$

tabela 1	1,0	2,0	3,0	4,0	5,0

tabela 2	0,5	1,0	3,0	3,5	7,0

3 - Desvio médio - definido como o somatório dos módulos da distância de cada valor até a média.

$$\boxed{desvio \cdot medio = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \bar{x}_{obs}\right|}$$

Tabela 1 $\quad desvio \cdot medio = \dfrac{6,0}{5} = 1,2$

Tabela 2 $\quad desvio \cdot medio = \dfrac{9,0}{5} = 1,8$

tabela 1	1,0	2,0	3,0	4,0	5,0
tabela 2	0,5	1,0	3,0	3,5	7,0

4 - Variância - definido como o somatório dos quadrados da distância de cada valor até a média.

$$\boxed{var_{obs} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_{obs})^2}$$

tabela 1	1,0	2,0	3,0	4,0	5,0
tabela 2	0,5	1,0	3,0	3,5	7,0

4 - Variância - definido como o somatório dos quadrados da distância de cada valor até a média.

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

Tabela 1

$$\begin{aligned}
\mathrm{var}_{obs} &= \frac{1}{5}((1,0-3,0)^2+(2,0-3,0)^2+ \\
&+ (3,0-3,0)^2+(4,0-3,0)^2+(5,0-3,0)^2)
\end{aligned}$$

$$\mathrm{var}_{obs} = \frac{1}{5}(4,0+1,0+0,0+1,0+4,0) = \frac{10,0}{5} = 2,0$$

tabela 1	1,0	2,0	3,0	4,0	5,0
tabela 2	0,5	1,0	3,0	3,5	7,0

4 - Variância - definido como o somatório dos quadrados da distância de cada valor até a média.

$$\boxed{var_{obs} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_{obs})^2}$$

Tabela 1 $\mathrm{var}_{obs} = 2,0$

Tabela 2

$$\begin{aligned} \mathrm{var}_{obs} &= \frac{1}{5}((0,5-3,0)^2 + (1,0-3,0)^2 + \\ &+ (3,0-3,0)^2 + (3,5-3,0)^2 + (7,0-3,0)^2) \\ \mathrm{var}_{obs} &= \frac{1}{5}(6,25+4,0+0,0+0,25+16,0) = \frac{26,5}{5} = 5,3 \end{aligned}$$

tabela 1	1,0	2,0	3,0	4,0	5,0
tabela 2	0,5	1,0	3,0	3,5	7,0

4 - Variância - definido como o somatório dos quadrados da distância de cada valor até a média.

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

Tabela 1 $\boxed{\mathrm{var}_{obs} = 2,0}$

Tabela 2 $\boxed{\mathrm{var}_{obs} = 5,3}$

Fundação CECIERJ Consórcio cederj

tabela 1	1,0	2,0	3,0	4,0	5,0

tabela 2	0,5	1,0	3,0	3,5	7,0

5 - Desvio Padrão - definido como a raiz quadrada do somatório dos quadrados da distância de cada valor até a média, ou seja, a raíz quadrada da variância.

$$dp_{obs} = \sqrt{\mathrm{var}_{obs}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_{obs})^2}$$

tabela 1	1,0	2,0	3,0	4,0	5,0

tabela 2	0,5	1,0	3,0	3,5	7,0

5 - Desvio Padrão - definido como a raiz quadrada do somatório dos quadrados da distância de cada valor até a média, ou seja, a raíz quadrada da variância.

$$dp_{obs} = \sqrt{\text{var}_{obs}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_{obs})^2}$$

Tabela 1 $\boxed{\mathrm{dp}_{obs} = \sqrt{2} = 1,4142}$

Tabela 2 $\boxed{\mathrm{dp}_{obs} = \sqrt{5,3} = 2,3022}$

Resumindo: medidas de dispersão para as tabelas 1 e 2

Tabela 1

$$\Delta = 4{,}0$$

$$desvio \cdot mediano = \frac{6,0}{5} = 1,2$$

$$desvio \cdot medio = \frac{6,0}{5} = 1,2$$

$$\mathrm{var}_{obs} = 2,0$$

$$dp_{obs} = \sqrt{2} = 1,4142$$

Tabela 2

$$\Delta = 6{,}5$$

$$desvio \cdot mediano = \frac{9,0}{5} = 1,8$$

$$desvio \cdot medio = \frac{9,0}{5} = 1,8$$

$$\mathrm{var}_{obs} = 5,3$$

$$dp_{obs} = \sqrt{5,3} = 2,3022$$

Exercício:

calcular amplitude, desvio mediano, desvio médio, variância e desvio padrão do exercício do aluno a escolha de um estágio.

Companhia A

Num. de sal. min.	1,0	1,5	2,0	3,0	7,0	9,0	12,0	total
freqüência	9	16	12	8	3	2	1	51

Companhia B

Num. de sal. min.	1,0	2,0	3,0	5,0	7,0	total
freqüência	14	17	11	6	3	51

Companhia A

Num. de sal. min.	1,0	1,5	2,0	3,0	7,0	9,0	12,0	total
freqüência	9	16	12	8	3	2	1	51

Companhia A

Num. de sal. min.	1,0	1,5	2,0	3,0	7,0	9,0	12,0	total
freqüência	9	16	12	8	3	2	1	51

Sabemos que para a companhia A a média é 2,6 e a mediana é 2,0.

Num. de sal. min.	1,0	1,5	2,0	3,0	7,0	9,0	12,0	Σ
$m_i \lvert x_i - 2{,}0 \rvert$	9,0	8,0	0,0	8,0	15,0	14,0	10,0	64,0
$m_i \lvert x_i - 2{,}6 \rvert$	14,4	17,6	7,2	3,2	13,2	12,8	9,8	77,8
$m_i (x_i - 2{,}6)^2$	23,0	19,4	4,3	1,3	58,1	81,9	88,4	276,4

amplitude Δ $= 11,0$

desvio.mediano $= \frac{64,0}{51} = 1,2$

desvio médio $= \frac{77,8}{51} = 1,5$

variância $= \frac{276,4}{51} = 5,4$

desvio padrão $= 2,3$

Companhia B

Num. de sal. min.	1,0	2,0	3,0	5,0	7,0	total
freqüência	14	17	11	6	3	51

Sabemos que para a companhia B a média é 2,6 e a mediana é 2,0.

Num. de sal. min.	1,0	2,0	3,0	5,0	7,0	Σ
$m_i \lvert x_i - 2,0 \rvert$	14,0	0,0	11,0	18,0	15,0	58,0
$m_i \lvert x_i - 2,6 \rvert$	22,4	10,2	4,4	14,4	13,2	64,6
$m_i (x_i - 2,6)^2$	35,8	6,1	1,8	34,6	58,0	136,4

amplitude Δ $= 6,0$

desvio.mediano $= \frac{58,0}{51} = 1,1$

desvio médio $= \frac{64,6}{51} = 1,3$

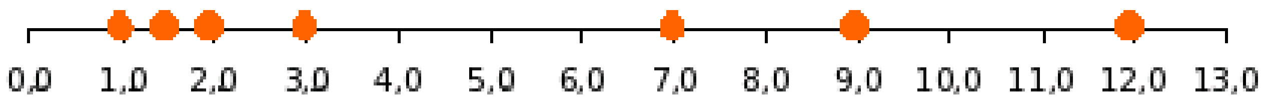
variância $= \frac{136,4}{51} = 2,7$

desvio padrão $= 1,6$

Companhia A

média	=	2,6
mediana	=	2,0
moda	=	1,5

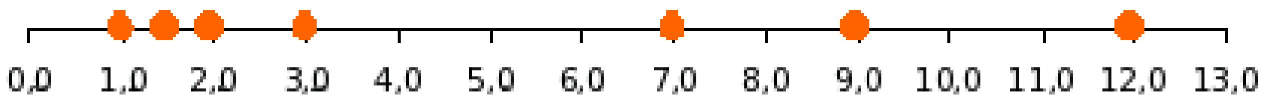
amplitude Δ	=	11,0
desvio.mediano	=	1,2
desvio médio	=	1,5
variância	=	5,4
desvio padrão	=	2,3



Companhia A

média = 2,6
mediana = 2,0
moda = 1,5

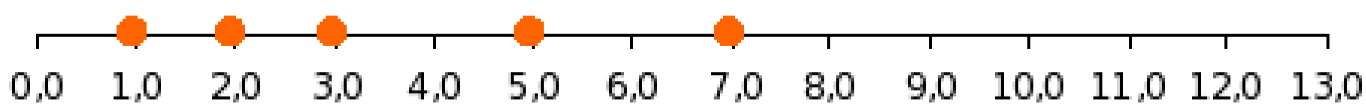
amplitude Δ = 11,0
desvio.mediano = 1,2
desvio médio = 1,5
variância = 5,4
desvio padrão = 2,3



Companhia B

média = 2,6
mediana = 2,0
moda = 2,0

amplitude Δ = 6,0
desvio.mediano = 1,1
desvio médio = 1,3
variância = 2,7
desvio padrão = 1,6



Lembrando...vimos na aula 3 : " Medidas de resumo para um conjunto de dados"

Medidas de posição (tendência central)	média mediana moda

Lembrando...vimos na aula 3 : " Medidas de resumo para um conjunto de dados"

Medidas de posição (tendência central)	média mediana moda
Medidas de dispersão	amplitude desvio mediano desvio médio variância desvio padrão

Fundação CECIERJ Consórcio cederj

Aula 3

Professores:

Otton Teixeira da Silveira Filho
Regina Célia Paula Leal Toledo

Medidas de resumo (para um conjunto de dados)

Conteúdo:

3.1 Medidas de posição
3.2 Algumas questões complementares
3.3 Exemplo
3.4 Medidas de dispersão

$$\sum_{i=1}^{n}(X_i-\bar{X})^2$$

$$\sigma_x^2 \qquad \bar{Y}=\frac{\sum_{i=1}^{n} y_i}{n}$$

$$A\cap B=\varnothing$$