

Universidade Federal do Rio Grande do Norte
Instituto Metr pole Digital
IMD0601 - Bioestat stica

Medidas de dispers o

Prof. Dr. Tetsu Sakamoto
Instituto Metr pole Digital - UFRN
Sala A224, ramal 182
Email: tetsu@imd.ufrn.br



Baixe a aula (e os arquivos)

- Para aqueles que não clonaram o repositório:

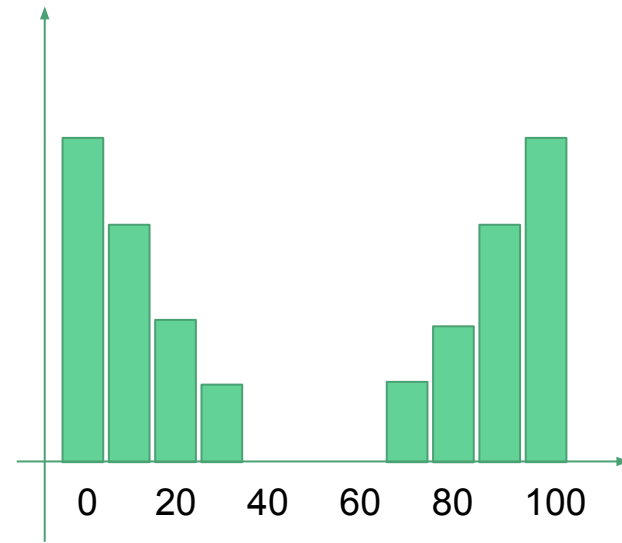
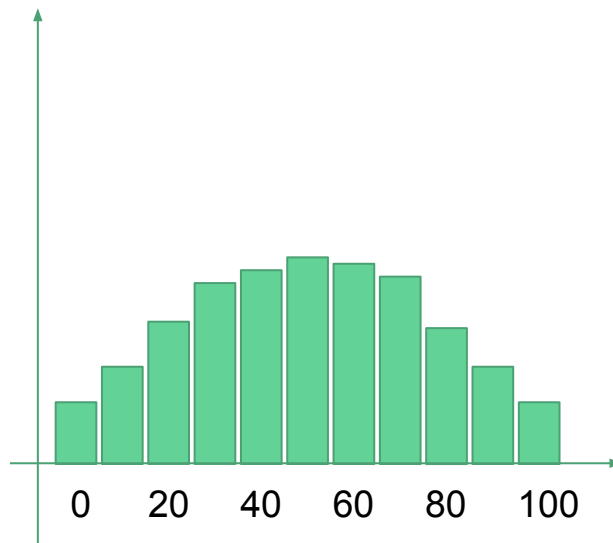
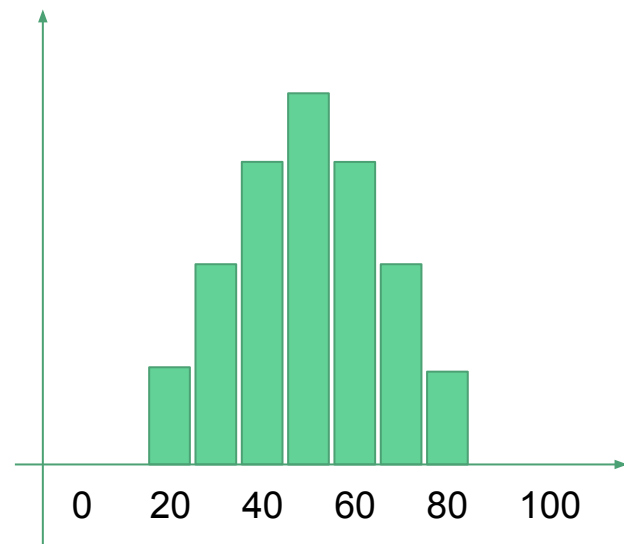
```
> git clone https://github.com/tetsufmbio/IMD0601.git
```

- Para aqueles que já tem o repositório local:

```
> cd /path/to/IMD0601
```

```
> git pull
```

Como são os dados que possuem média = 50?



Medidas de dispersão

São medidas que tentam descrever o grau da dispersão dos dados. Medidas que fornecem uma noção da distribuição dos dados.

Medida de dispersão ideal:

- Definição clara e rígida;
- Fácil cálculo e entendimento;
- Não deve ser muito afetado por flutuações;
- Baseado em todas as observações.

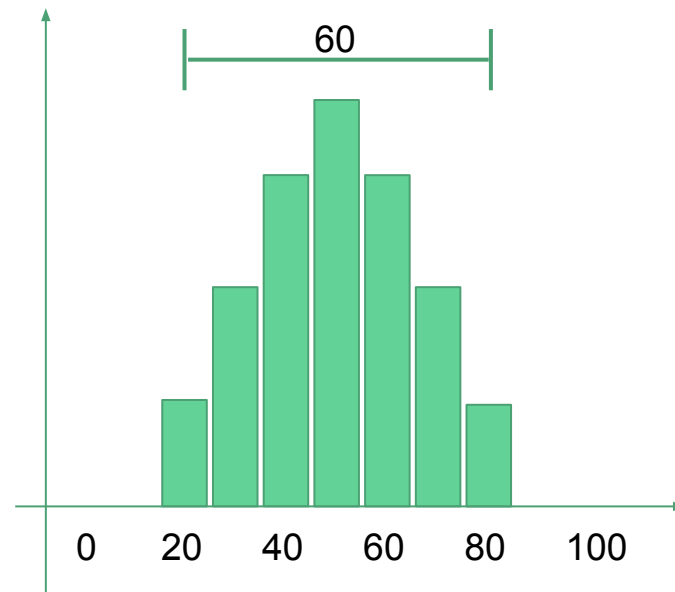
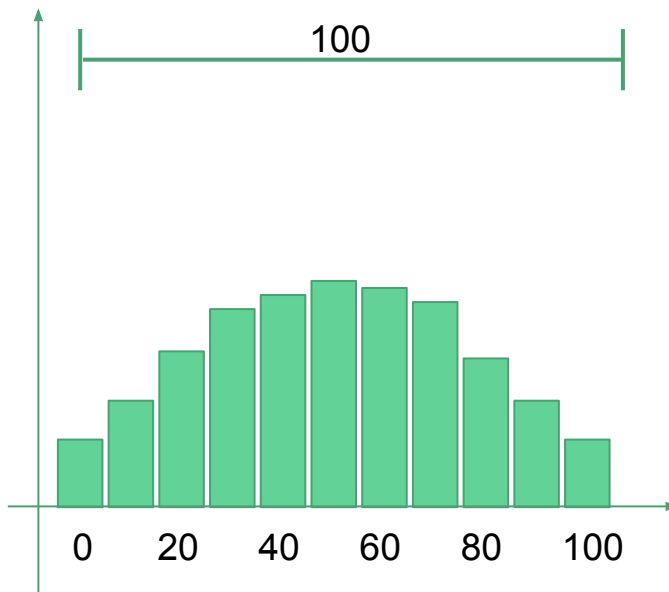
Medidas de dispersão

Existem duas categorias de medida de dispersão:

- **Medidas de dispersão absoluta** → quantifica a variação em termos da unidade de medida dos dados;
 - Amplitude;
 - Desvio entre quartis;
 - Desvio da média;
 - Desvio padrão.
- **Medidas de dispersão relativa** → não possui unidade de medida, comparação entre as distribuições;
 - Coeficiente de Amplitude;
 - Coeficiente de desvio entre quartis;
 - Coeficiente de desvio da média;
 - Coeficiente de variação;

Amplitude

Diferença entre o valor máximo e o valor mínimo dos dados.



Amplitude

Diferença entre o valor máximo e o valor mínimo dos dados.

Vantagens:

- O mais simples das medidas de dispersão;
- Fácil cálculo e entendimento.

Desvantagens:

- Baseado em apenas duas observações extremas;
- Não é uma medida de dispersão confiável.

Desvio entre quartis

Quartil - Os dados são divididos em 4 partes. Cada parte é denominado de **quartil**.

Para dividirmos os dados em quartis, definimos 3 posições:

Q1: Compreende até 25% dos dados;

Q2: Compreende até 50% dos dados (mediana);

Q3: Compreende até 75% dos dados.

Desvio entre quartis

$$DEQ = (Q3 - Q1)/2$$

A = { 1, 4, 5, 8, 8, 9, 10, 12, 14, 15, 16 }

Q1
5

Q2
9

Q3
14

$$DEQ = (14 - 5)/2 = 4.5$$

Amplitude entre quartis (IQR)

$$DEQ = (Q3 - Q1)/2$$

$A = \{ 1, 4, 5, 8, 8, 9, 10, 12, 14, 15, 16, 19 \}$

Q1

6.5

Q2

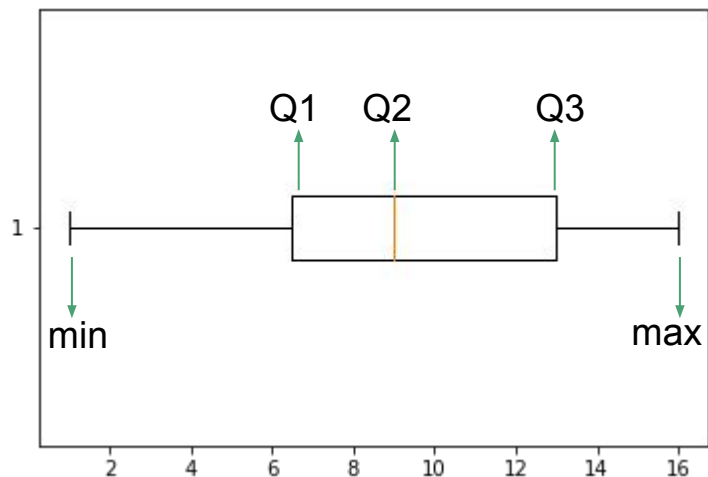
9.5

Q3

14.5

$$DEQ = (14.5 - 6.5)/2 = 4$$

Representação gráfica em Boxplot



$A = \{ 1, 4, 5, 8, 8, 9, 10, 12, 14, 15, 16 \}$

Desvio entre quartis

Vantagens:

- Fácil de calcular;
- O cálculo envolve apenas o Q1 e o Q3;
- Não é afetado por valores extremos;

Desvantagens:

- Utiliza apenas 50% dos dados para o seu cálculo;

Outliers

Determinação dos outliers utilizando a separação dos dados em quartis:

Limite inferior: **$Q1 - 1.5 * IQR$**

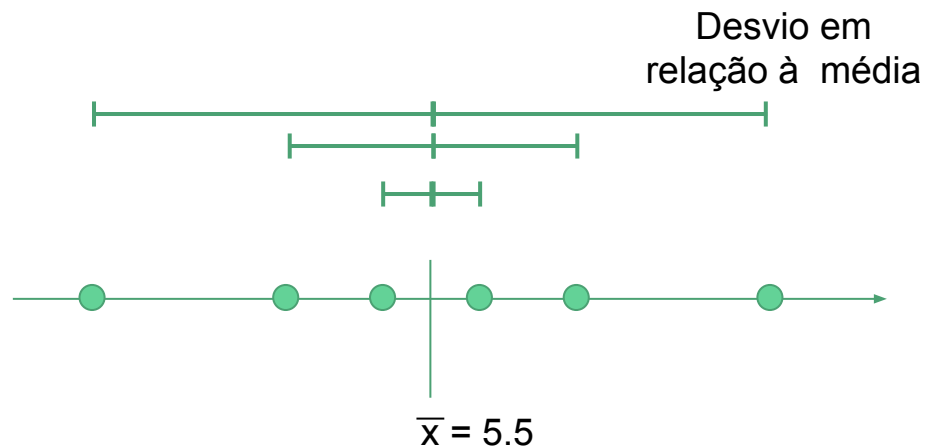
Limite superior: **$Q3 + 1.5 * IQR$**

Onde **$IQR = Q3 - Q1$**

Os dados que estiverem fora deste limite são considerados outliers.

Desvio envolvendo a média

x_i	$x_i - \bar{x}$
2	$2 - 5.5 = -3.5$
4	$4 - 5.5 = -1.5$
5	$5 - 5.5 = -0.5$
6	$6 - 5.5 = 0.5$
7	$7 - 5.5 = 1.5$
9	$9 - 5.5 = 3.5$



Média do desvio = $\text{somatória}(x_i - \bar{x})/n = 0$

Desvio envolvendo a média

Desvio absoluto

x_i	$x_i - \bar{x}$	$ x_i - \bar{x} $
2	$2 - 5.5 = -3.5$	$ 2 - 5.5 = 3.5$
4	$4 - 5.5 = -1.5$	$ 4 - 5.5 = 1.5$
5	$5 - 5.5 = -0.5$	$ 5 - 5.5 = 0.5$
6	$6 - 5.5 = 0.5$	$ 6 - 5.5 = 0.5$
7	$7 - 5.5 = 1.5$	$ 7 - 5.5 = 1.5$
9	$9 - 5.5 = 3.5$	$ 9 - 5.5 = 3.5$

Média do desvio absoluto = somatória($|x_i - \bar{x}|$)/n = 5.5

Desvio envolvendo a média

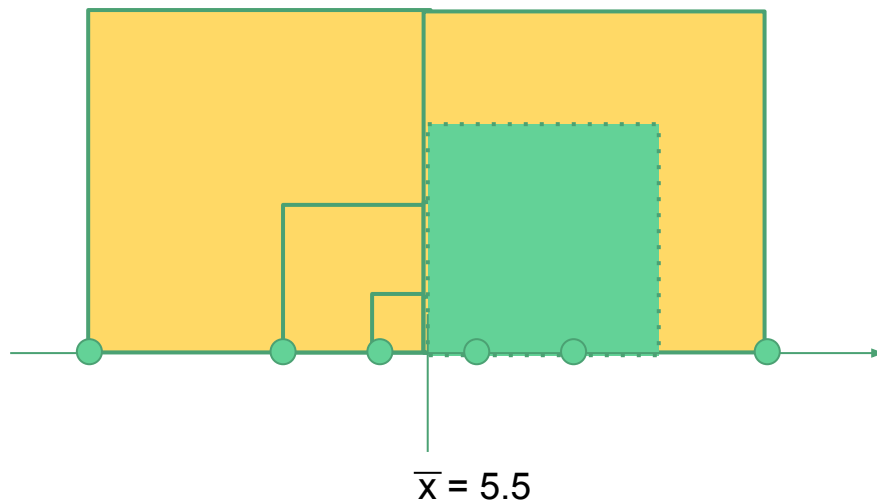
x_i	$x_i - \bar{x}$	$ x_i - \bar{x} $	$(x_i - \bar{x})^{**2}$
2	$2 - 5.5 = -3.5$	$ 2 - 5.5 = 3.5$	$(2 - 5.5)^{**2} = 12.25$
4	$4 - 5.5 = -1.5$	$ 4 - 5.5 = 1.5$	$(4 - 5.5)^{**2} = 2.25$
5	$5 - 5.5 = -0.5$	$ 5 - 5.5 = 0.5$	$(5 - 5.5)^{**2} = 0.25$
6	$6 - 5.5 = 0.5$	$ 6 - 5.5 = 0.5$	$(6 - 5.5)^{**2} = 0.25$
7	$7 - 5.5 = 1.5$	$ 7 - 5.5 = 1.5$	$(7 - 5.5)^{**2} = 2.25$
9	$9 - 5.5 = 3.5$	$ 9 - 5.5 = 3.5$	$(9 - 5.5)^{**2} = 12.25$

Quadrado do
Desvio

Média do
quadrado do
desvio
=
somatória(($x_i - \bar{x}$)^{**2})/n
=
4.916

Variância

Quando estamos calculando a variância...



Desvio padrão

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

$$\textit{Desvio padrão} = \sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

Muitas vezes...

$$\sigma = \sqrt{\frac{\sum (xi - x)^2}{n}}$$

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

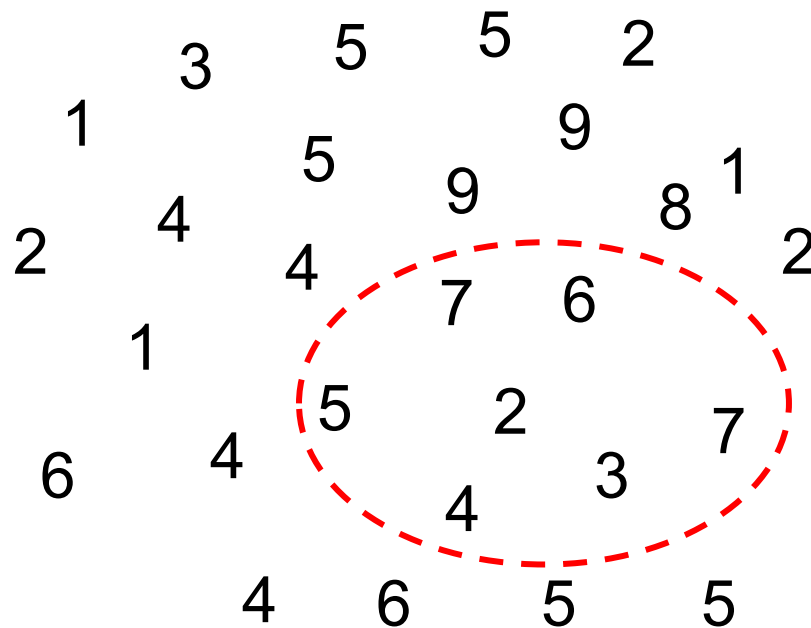
Desvio padrão amostral

Uma amostragem não consegue representar toda a variabilidade da população, por isso utilizamos a **correção de Bessel**, para corrigir esta limitação da amostragem.

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

$$\sigma = 2.243756$$

$$s = 1.9518 (\sigma = 1.807016)$$



Desvio padrão

Vantagens:

- Envolve todas as observações para o seu cálculo;
- É pouco afetado por flutuações dos valores;
- Bem definido;

Desvantagens:

- Seu cálculo pode ser laborioso, especificamente se o tamanho dos dados é grande o suficiente;
- Pode ser afetado por valores extremos;

Medidas de dispersão relativa

Não possuem unidade de medida, permite a comparação entre as distribuições;

- Coeficiente de Amplitude;
- Coeficiente de desvio entre quartis;
- Coeficiente de desvio da média;
- Coeficiente de variação;

Coeficiente de Amplitude

Coeficiente de desvio entre quartis

Coef. de amplitude = $(H - L)/(H + L)$

Coef. desvio entre quartis = $(Q3 - Q1)/(Q3 + Q1)$

Coeficiente de desvio da média

Coeficiente de desvio de variação

Coef. desvio da média =

$(\text{desvio da média})/(\text{média ou mediana})$

Coef. de variação =

$(\text{desvio padrão})/(\text{média})$