

Estatística descritiva

Medidas de dispersão

Prof. Dr. Tetsu Sakamoto

Instituto Metrópole Digital - UFRN

Sala A224, ramal 182

Email: tetsu@imd.ufrn.br



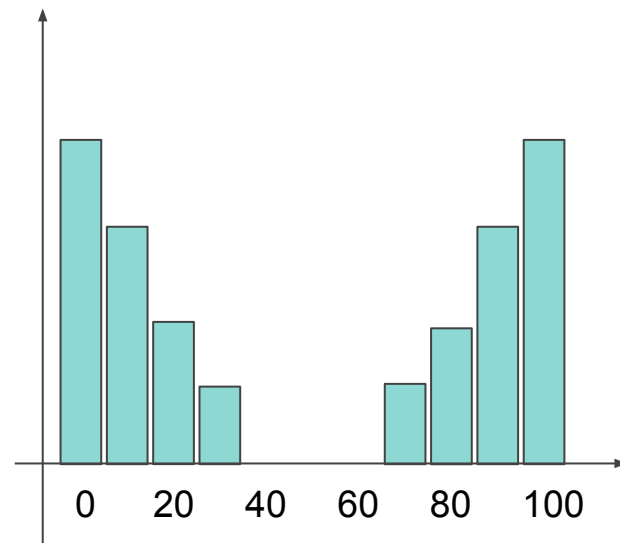
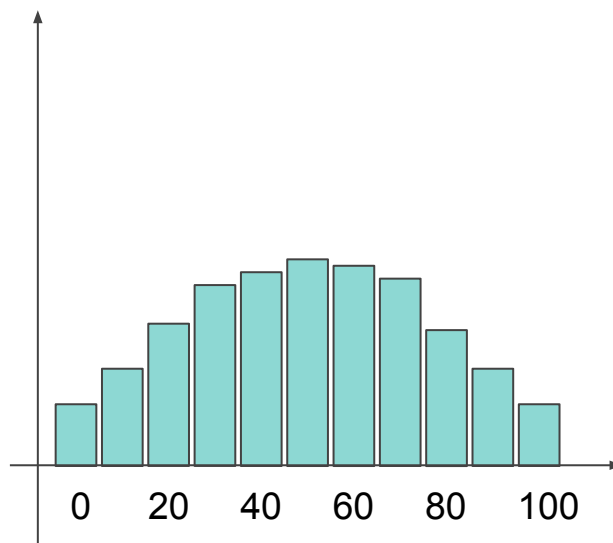
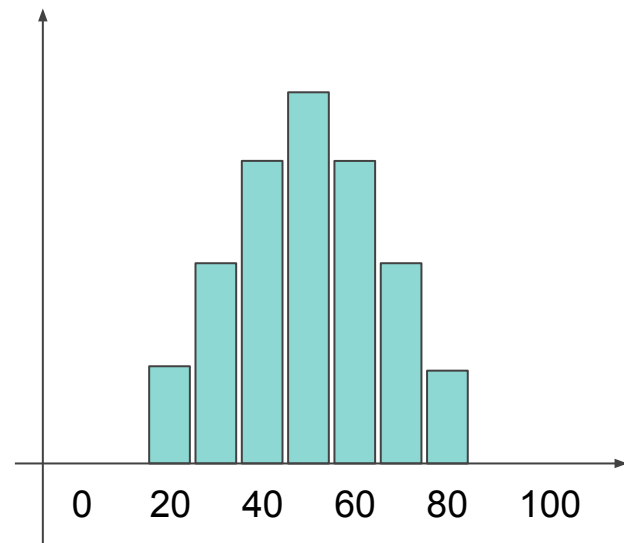
Slides e notebook em:

github.com/tetsufmbio/IMD0033/





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

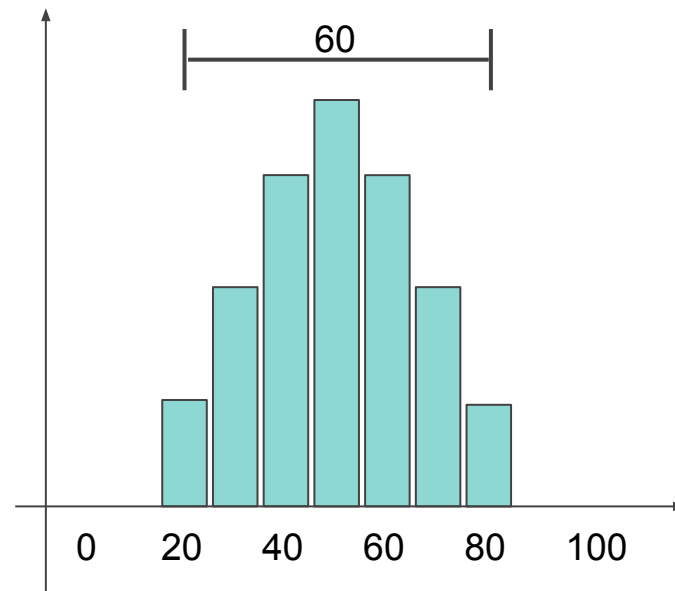
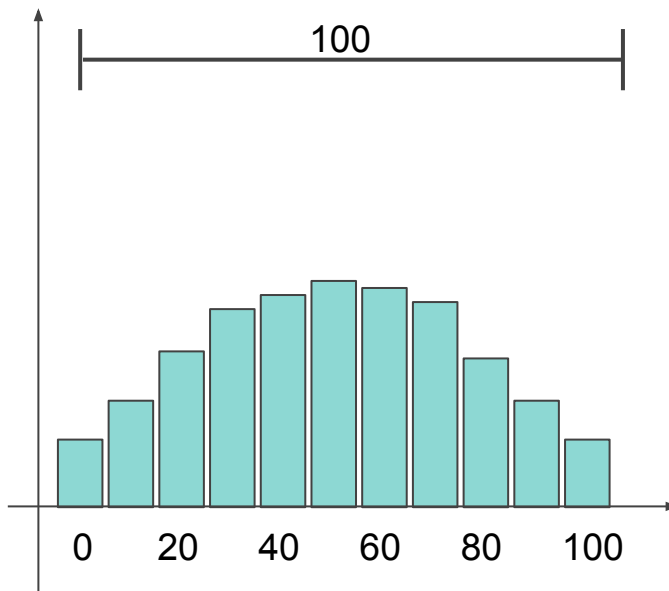
Existem duas categorias de medida de dispersão:

- **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 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

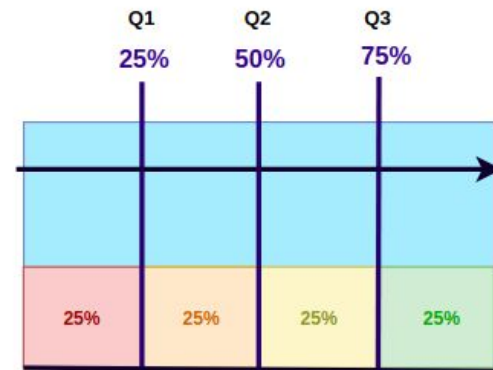
Quartis - Os dados são divididos em 4 partes. Cada parte é denominado de **quartil**.

Para dividirmos os dados em quartis, definimos 3 posições:

Q1: Compreende até 25% dos dados;

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

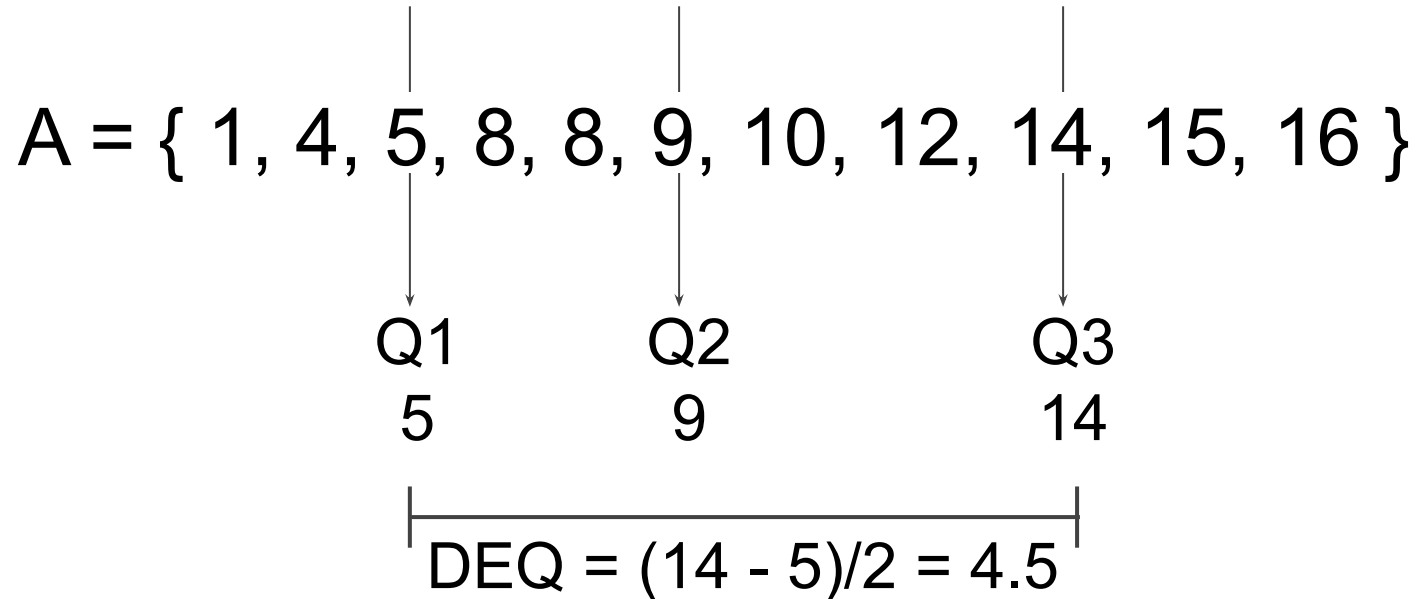
Q3: Compreende até 75% dos dados.





Amplitude entre quartis (IQR)

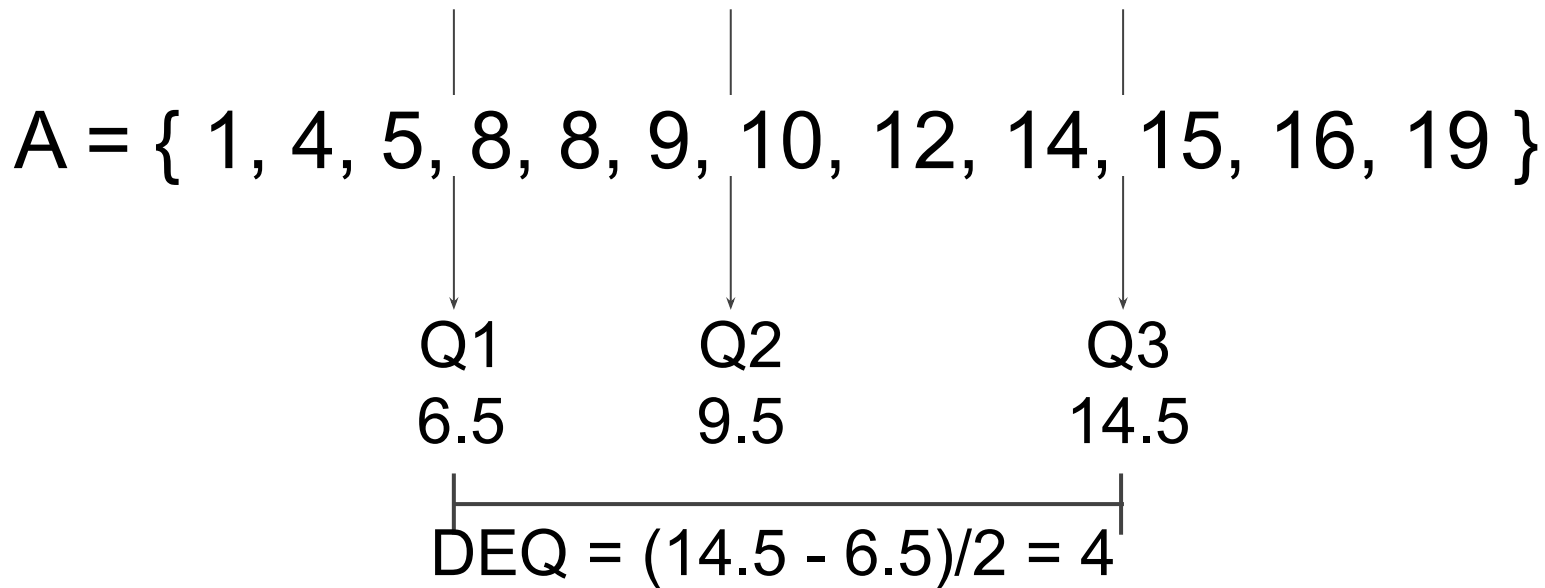
$$DEQ = (Q3 - Q1)/2$$





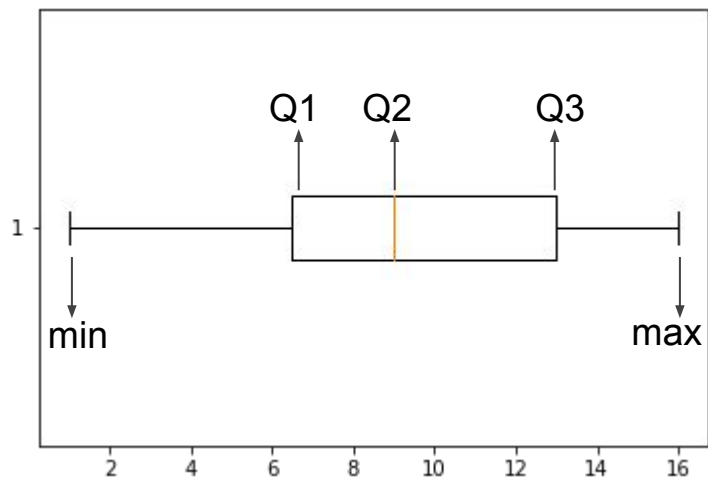
Amplitude entre quartis (IQR)

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





Representação gráfica em Boxplot



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



Amplitude entre quartis (IQR)

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: **$Q1 - 1.5 * IQR$**

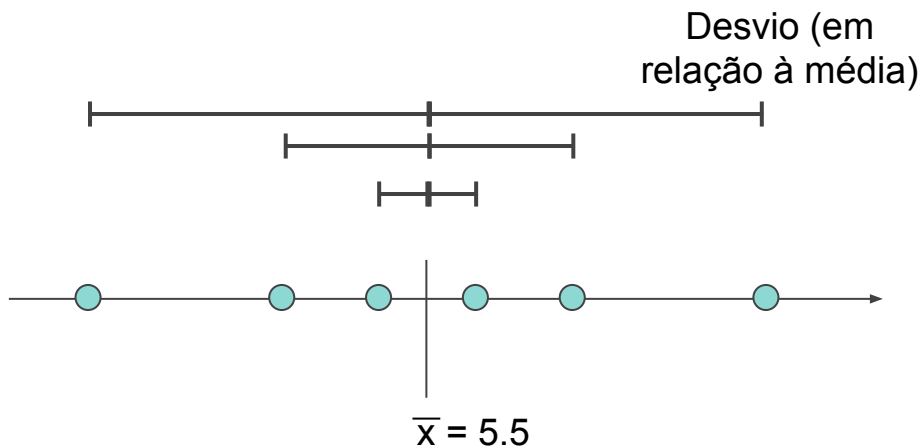
Limite superior: **$Q3 + 1.5 * IQR$**

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 = 1.83



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$ | $(9 - 5.5)^{**2} = 12.25$ |

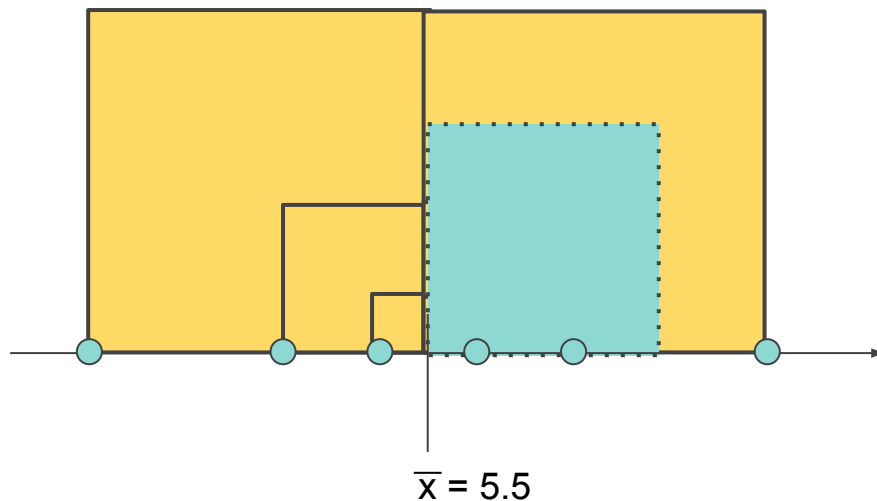
Quadrado do
Desvio

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

Variância



Quando estamos calculando a variância...





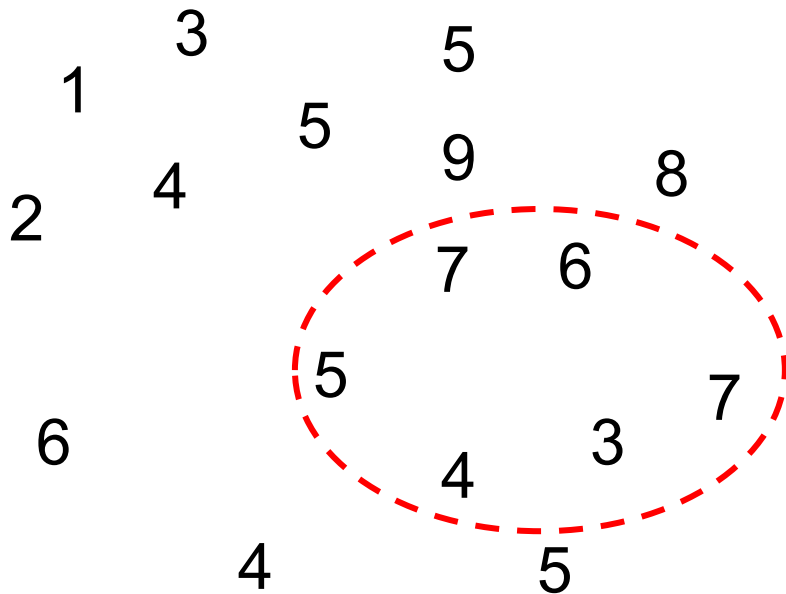
Desvio padrão

$$\text{Variância} = \sigma^2 = (\Sigma(x_i - \bar{x})^2)/n$$

$$\text{Desvio padrão} = \sigma = ((\Sigma(x_i - \bar{x})^2)/n)^{0.5}$$



Desvio padrão da amostra





Desvio padrão da amostra

Na maioria das vezes, uma amostragem não consegue representar toda a variabilidade da população, por isso utilizamos a **correção de Bessel**, onde determina que o desvio padrão da amostra, caso utilizado para estimar o desvio padrão de uma população corresponde a:

Desvio

$$\text{padrão} = s = ((\sum (x_i - \bar{x})^2) / (n-1))^{0.5}$$

amostral



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;



Medidas de dispersão relativa

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

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

Coef. de variação = $(\text{desvio padrão})/(\text{média})$