INSTITUTO FEDERAL
Rio Grande do Sul
Campus Erechim

ESTATÍSTICA

- MEDIDAS DE VARIABILIDADE -

Profa. Claudia Turik de Oliveira

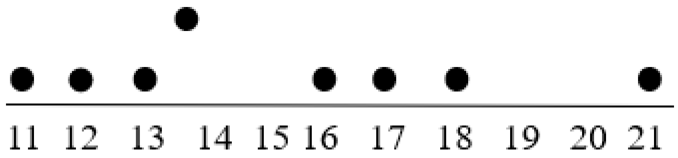
Medidas de Variabilidade

- Mostram o grau de dispersão ou afastamento dos valores observados em torno de um valor central.

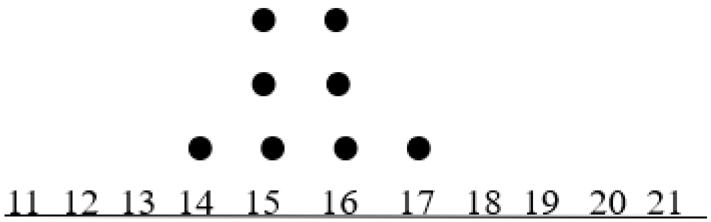
- Indicam se um conjunto é homogêneo (pouca ou nenhuma variabilidade) ou heterogêneo (muita variabilidade).

- A descrição do conjunto de dados é mais completa quando se considera além de uma medida de posição, uma medida de variabilidade.

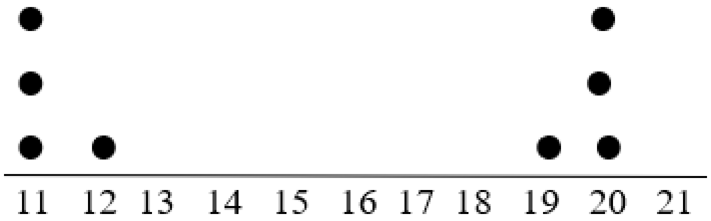
Medidas de Variabilidade



→ Média = 15,5



→ Média = 15,5



→ Média = 15,5

Amplitude

É a diferença entre o valor máximo e o valor mínimo de um conjunto de dados.

Exemplo:

10 alunos fizeram uma prova com 50 questões. Números de respostas corretas, por aluno:

31; 27; 42; 35; 47; 28; 7; 45; 15; 20.

Amplitude = 47 – 7 = 40

Amplitude Interquartil

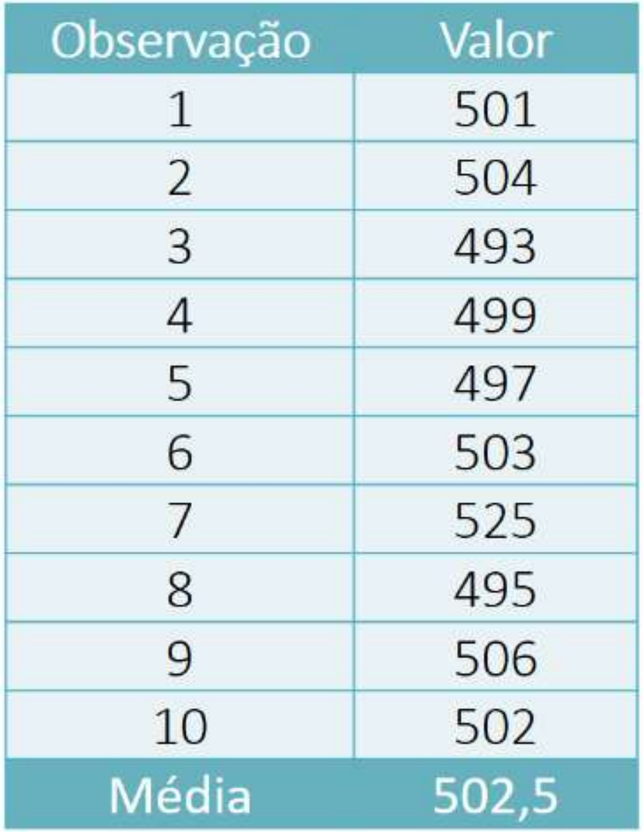
- Também chamada de **Dispersão entre os quartos**
- É a amplitude de valores que abrange os dados centrais (50%) das observações.
- É dada pela diferença entre o terceiro e o primeiro quartil.

$$\boxed{f_s = Q_3 - Q_1}$$

Outliers

- **Limite inferior: Q1 – 1,5 (Q3 - Q1)**
- **Limite superior: Q3 + 1,5 (Q3 - Q1)**
- **“Outliers” (dados discrepantes): dados fora dos limites superior e inferior**

Outliers



Observação	Valor
1	501
2	504
3	493
4	499
5	497
6	503
7	525
8	495
9	506
10	502
Média	502,5

- Quartil 1 (Q1) = 497
- Quartil 3 (Q3) = 504
- *f* = 504 – 497 = 7

- Limite superior = 504 + 1,5 x 7 = 514,5
- Limite inferior = 497 - 1,5 x 7 = 486,5

- Como o valor 525 é superior a 514,5, esse ponto é um outlier.

Variância amostral (s^2)

- Soma dos quadrados dos desvios em relação à média dividida por n-1

$$s^2 = \frac{\sum(X-\bar{X})^2}{n-1} = \frac{soma\ dos\ quadrados\ dos\ desvios}{n-1}$$

- Etapas de cálculo:

1. Calcular a média amostral

2. Subtrair a média de cada valor do conjunto, obtendo-se o que chamamos de desvios em relação à média

3. Elevar cada desvio ao quadrado

4. Somar os quadrados dos desvios

5. Dividir a soma por (n-1)

Variância amostral (s^2)

Desvio em Relação à Média

- É a diferença entre cada elemento de um conjunto de valores e a média aritmética deste mesmo conjunto: d = $(X - \bar{X})$

Ex: Encontre os desvios em relação à média dos conjunto: 10, 14, 13, 15, 16, 18, 12

1) Calcular a média amostral

$$\bar{X} = \frac{10 + 14 + \cdots + 12}{7}$$

$$\bar{X} = \frac{98}{7} = 14$$

2) Calcular cada valor menos a média

$$\left.\begin{aligned}
d_1 &= x_1 - \bar{x} = 10 - 14 = -4\\
d_2 &= x_2 - \bar{x} = 14 - 14 = 0\\
d_3 &= x_3 - \bar{x} = 13 - 14 = -1\\
d_4 &= x_4 - \bar{x} = 15 - 14 = 1\\
d_5 &= x_5 - \bar{x} = 16 - 14 = 2\\
d_6 &= x_6 - \bar{x} = 18 - 14 = 4\\
d_7 &= x_7 - \bar{x} = 12 - 14 = -2
\end{aligned}\right\}$$

A soma dos desvios dá sempre zero!

Variância amostral (s^2)

Exemplo: Seja o conjunto x = {0 , 4 , 6 , 8 , 7}, encontre a variância amostral.

Dados X	Desvios $d = X - \bar{X}$	Quadrado dos desvios $d^2 = (X - \bar{X})^2$
0	$0 - 5 = -5$	$(-5)^2 = 25$
4	$4 - 5 = -1$	$(-1)^2 = 1$
6	$6 - 5 = 1$	$1^2 = 1$
8	$8 - 5 = 3$	$3^2 = 9$
7	$7 - 5 = 2$	$2^2 = 4$
$\bar{x} = 5$	$\Sigma (d) = 0$	$\Sigma (d)^2 = 40$

$$s^2 = \frac{\text{soma dos quadrados dos desvios}}{\text{(número de elementos) - 1}} = \frac{40}{5 - 1} = \mathbf{10}$$

Desvio-padrão amostral (s)

- É a medida de dispersão mais geralmente empregada, pois leva em consideração a totalidade dos valores da variável em estudo.

- É a raiz quadrada, com sinal positivo, da variância.

$$\boldsymbol{s = \sqrt{s^2}}$$

Desvio-padrão amostral (s)

- Exemplo: Seja o conjunto x = {0 , 4 , 6 , 8 , 7} do exemplo anterior, encontre o desvio-padrão amostral.

Como já temos que $\boldsymbol{s^2}$ = 10 $\longrightarrow$ $s = \sqrt{10} = 3{,}16$

- A vantagem do desvio padrão sobre a variância, é que este permite uma interpretação direta da variação do grupo, pois o mesmo é expresso na mesma unidade de medida em que estão expressas as variáveis amostradas.

Coeficiente de Variação

- É uma medida de dispersão relativa
- Exprime a variabilidade em relação a média
- É útil quando se deseja comparar a variação de conjunto de dados que apresentem diferentes unidades de medição, pois o coeficiente de variação independe da unidade de medida dos dados.

$$\boxed{\mathbf{CV} = \frac{\boldsymbol{s}}{\overline{\boldsymbol{X}}} \cdot \mathbf{100\%}}$$

Coeficiente de Variação

	Média	Desvio-Padrão
Altura (m)	1,70	0,2
Peso (kg)	90	8

Este grupo de pessoas varia mais em altura ou em peso?

Coeficiente de Variação

	Média	Desvio-Padrão
Altura (m)	1,70	0,2
Peso (kg)	90	8

- $CV_{altura} = \frac{0,2}{1,70} \cdot 100\% = 11,76\%$
- $CV_{peso} = \frac{8}{90} \cdot 100\% = 8,89\%$

Coeficiente de Variação

Ação A:
Preço médio = R$50,00
Desvio padrão = R$5,00

$$CV = \frac{5}{50} \cdot 100\%$$

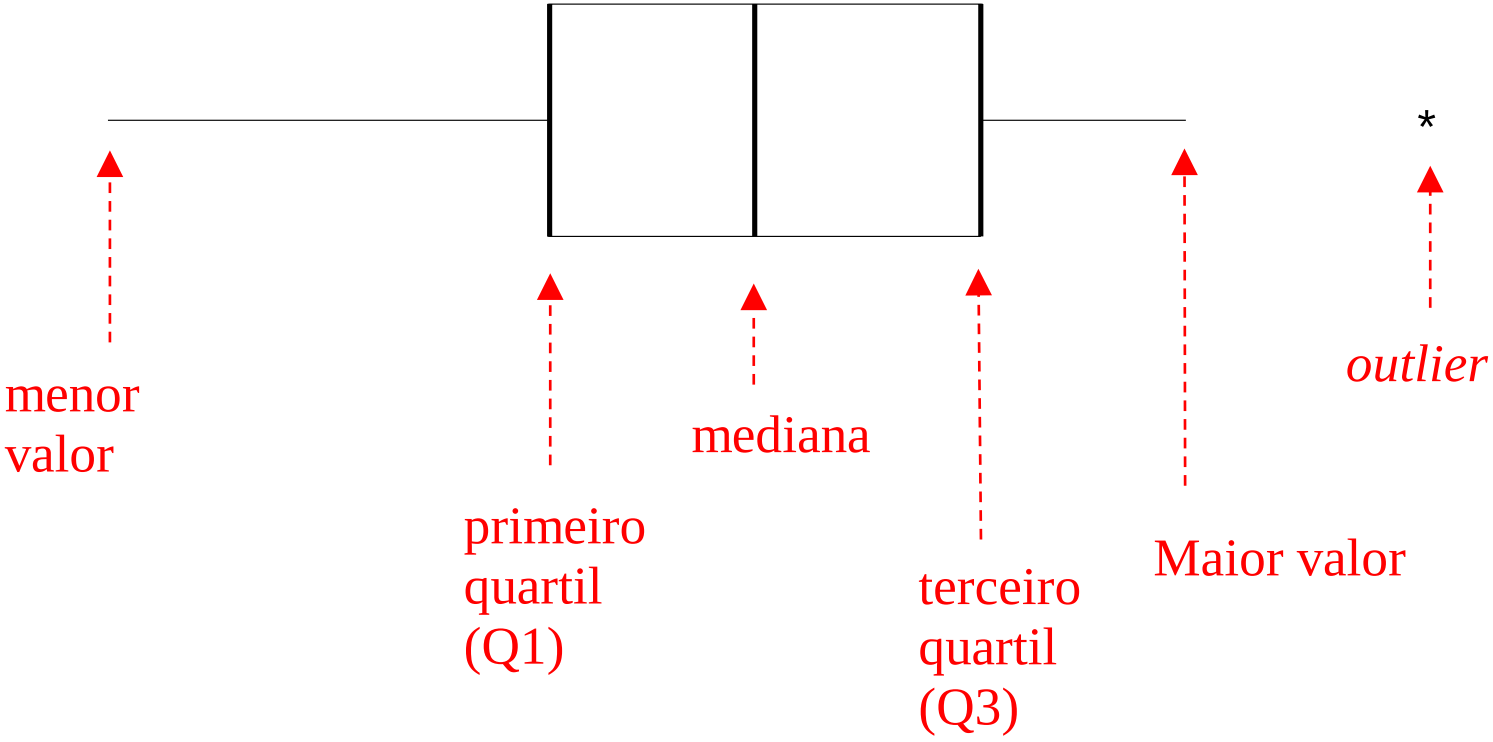
$$CV = 10\%$$

Ação B:
Preço médio = R$100,00
Desvio padrão = R$5,00

$$CV = \frac{5}{100} \cdot 100\%$$

$$CV = 5\%$$

Boxplot


*
menor valor
primeiro quartil (Q1)
mediana
terceiro quartil (Q3)
Maior valor
outlier

Boxplot

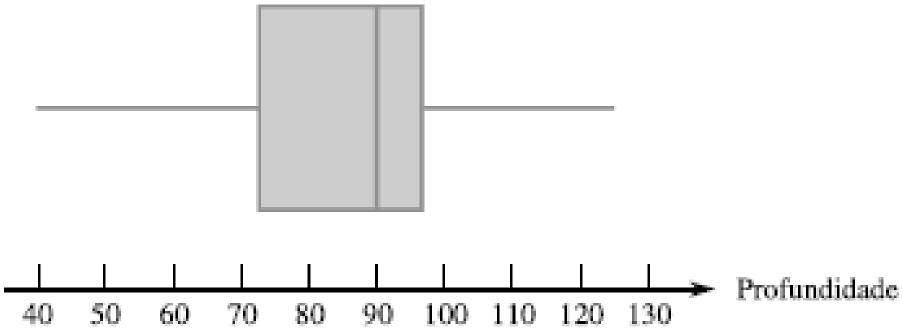
Exemplo 1.18

O ultra-som foi usado para obter informações sobre dados de corrosão na espessura da chapa do assoalho de um reservatório elevado usado para armazenar óleo bruto ("Statistical Analysis of UT Corrosion Data from Floor Plates of a Crude Oil Aboveground Storage Tank," *Materials Eval.*, 1994, p. 846-849). Cada observação é a maior profundidade do orifício na placa, expressa em milipolegadas.

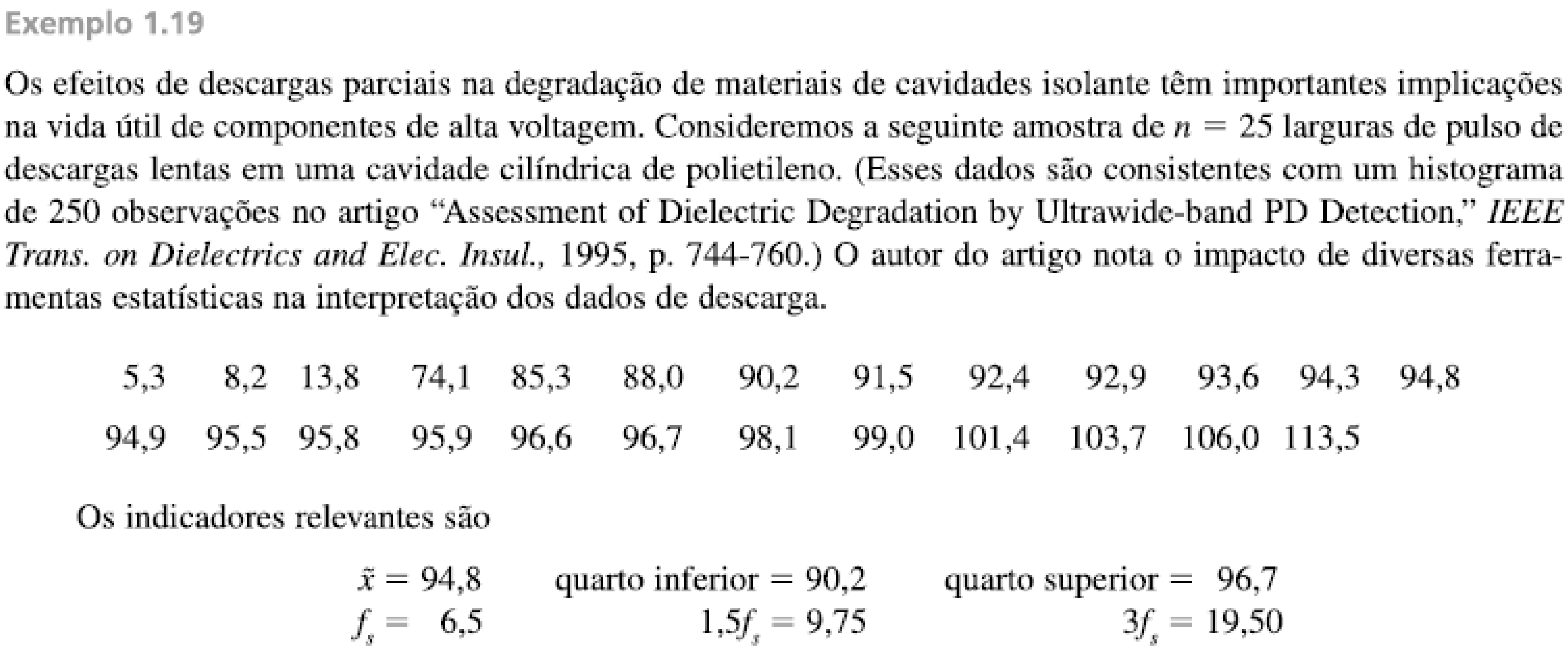
$$\underbrace{40\quad 52\quad 55\quad 60\quad 70\quad 75\quad 85\quad 85\quad 90\quad 90}\!\!\!\!\!\!\!\!\overbrace{\quad 92\quad 94\quad 94\quad 95\quad 98\quad 100\quad 115\quad 125\quad 125}$$

O resumo de cinco números segue:

menor x_i = 40 quarto inferior = 72,5 $\tilde{x} = 90$ quarto superior = 96,5
maior x_i = 125



Boxplot



Exemplo 1.19

Os efeitos de descargas parciais na degradação de materiais de cavidades isolante têm importantes implicações na vida útil de componentes de alta voltagem. Consideremos a seguinte amostra de $n = 25$ larguras de pulso de descargas lentas em uma cavidade cilíndrica de polietileno. (Esses dados são consistentes com um histograma de 250 observações no artigo "Assessment of Dielectric Degradation by Ultrawide-band PD Detection," *IEEE Trans. on Dielectrics and Elec. Insul.*, 1995, p. 744-760.) O autor do artigo nota o impacto de diversas ferramentas estatísticas na interpretação dos dados de descarga.

5,3	8,2	13,8	74,1	85,3	88,0	90,2	91,5	92,4	92,9	93,6	94,3	94,8
94,9	95,5	95,8	95,9	96,6	96,7	98,1	99,0	101,4	103,7	106,0	113,5	

Os indicadores relevantes são

$\tilde{x} = 94{,}8$ quarto inferior $= 90{,}2$ quarto superior $= 96{,}7$

$f_s = 6{,}5$ $1{,}5f_s = 9{,}75$ $3f_s = 19{,}50$

Boxplot

$\tilde{x} = 94{,}8$ quarto inferior $= 90{,}2$ quarto superior $= 96{,}7$

$f_s = 6{,}5$ $1{,}5f_s = 9{,}75$ $3f_s = 19{,}50$