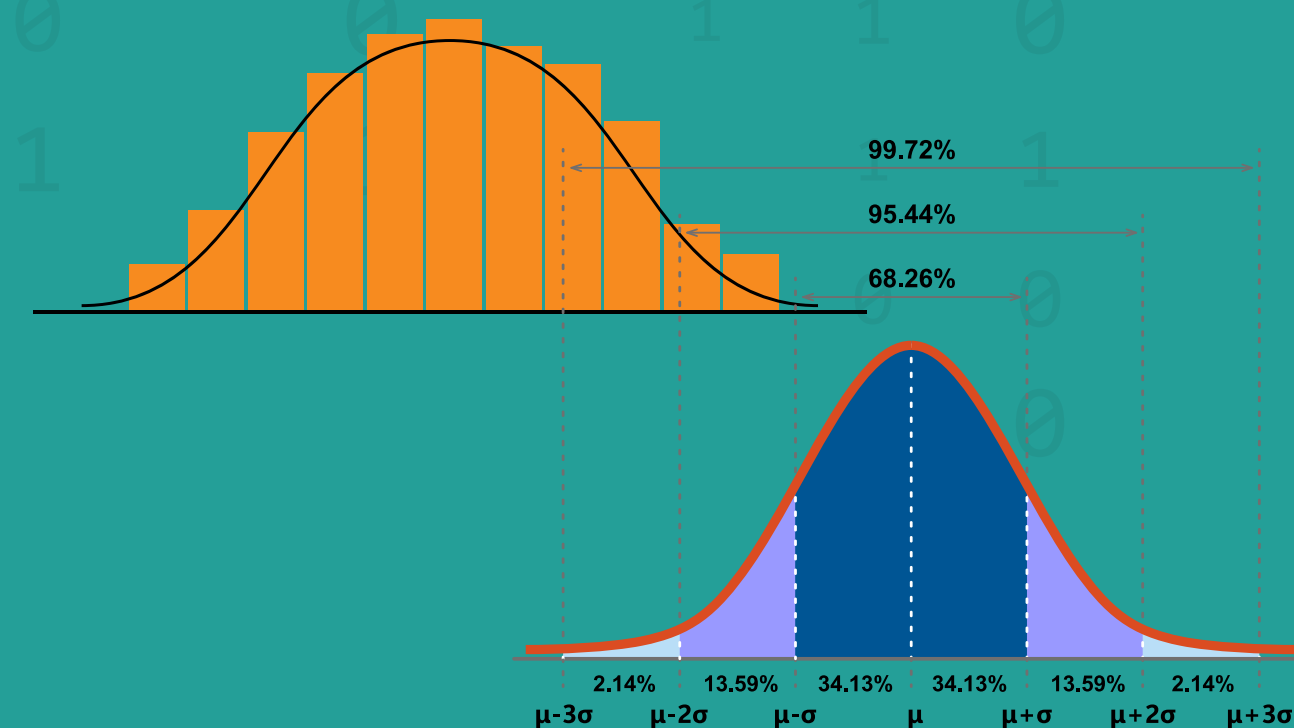
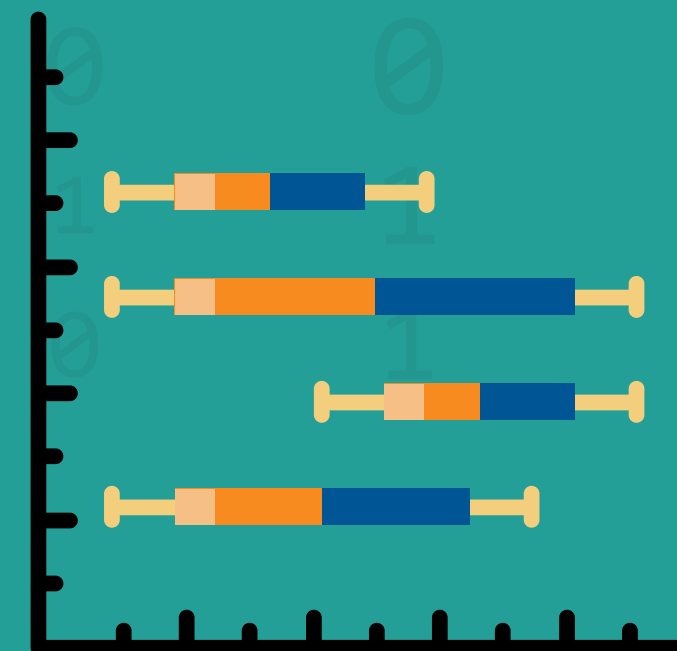
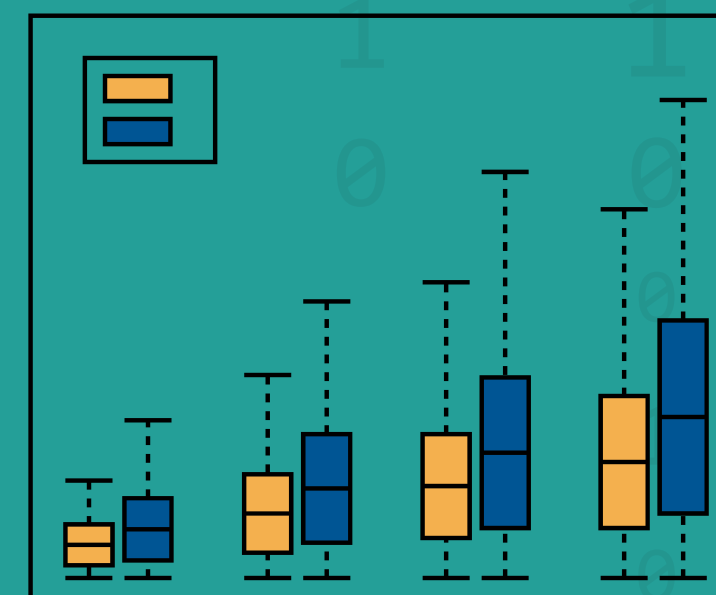
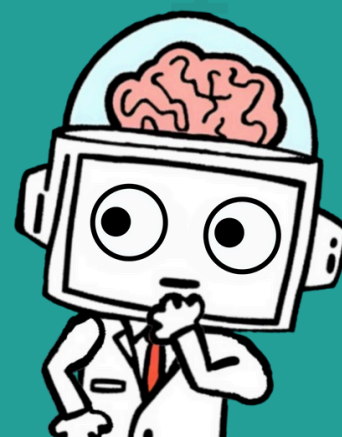


Data Science - Princípios e Técnicas

# Atividade 3: Centralidade e Dispersão

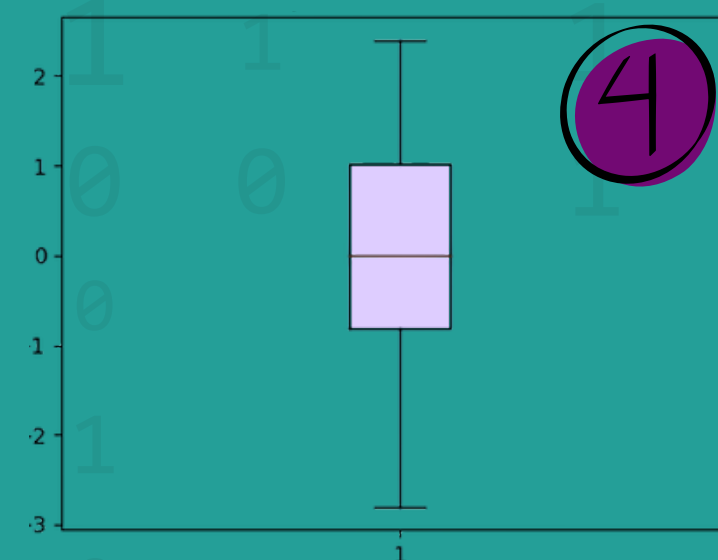
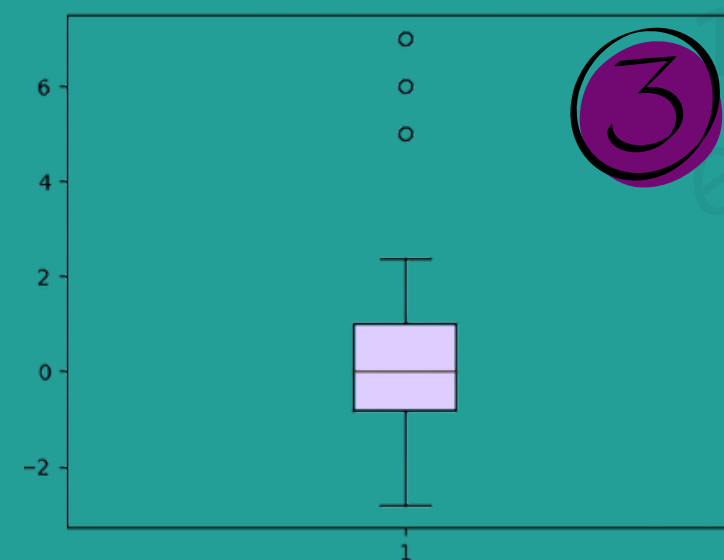
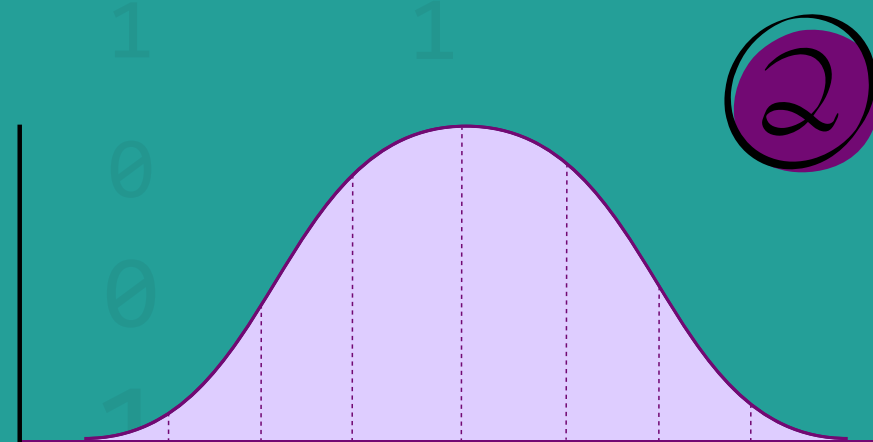
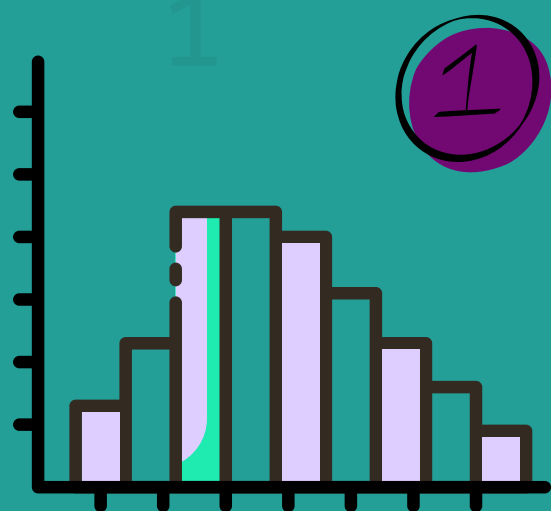
Prof. Dr. Welton Dionisio

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



Com base no dataset escolhido por sua equipe na última atividade, construa em **Python ou R**:

- ① Um histograma com barras (também chamado gráfico de distribuição dos dados);
- ② Uma curva de densidade de dados (Kernel Density Estimation - KDE, um tipo de histograma com distribuição suavizada, como ondas);
- ③ Um boxplot com com bigode apresentando outliers (os valores discrepantes);
- ④ Outro boxplot com bigode sem apresentar outliers.





A atividade terá prazo maior: iniciará na próxima aula presencial e continuará durante **o próximo LabTeams**. Aqueles que já desejarem iniciá-la, podem consultar os novos materiais de apoio na Seção **Arquivos → Material de Aula no Teams** de Data Science da sua turma. Os arquivos estão **listados com seus nomes abaixo**.

Até a próxima!



welton.silva@pe.senac.br

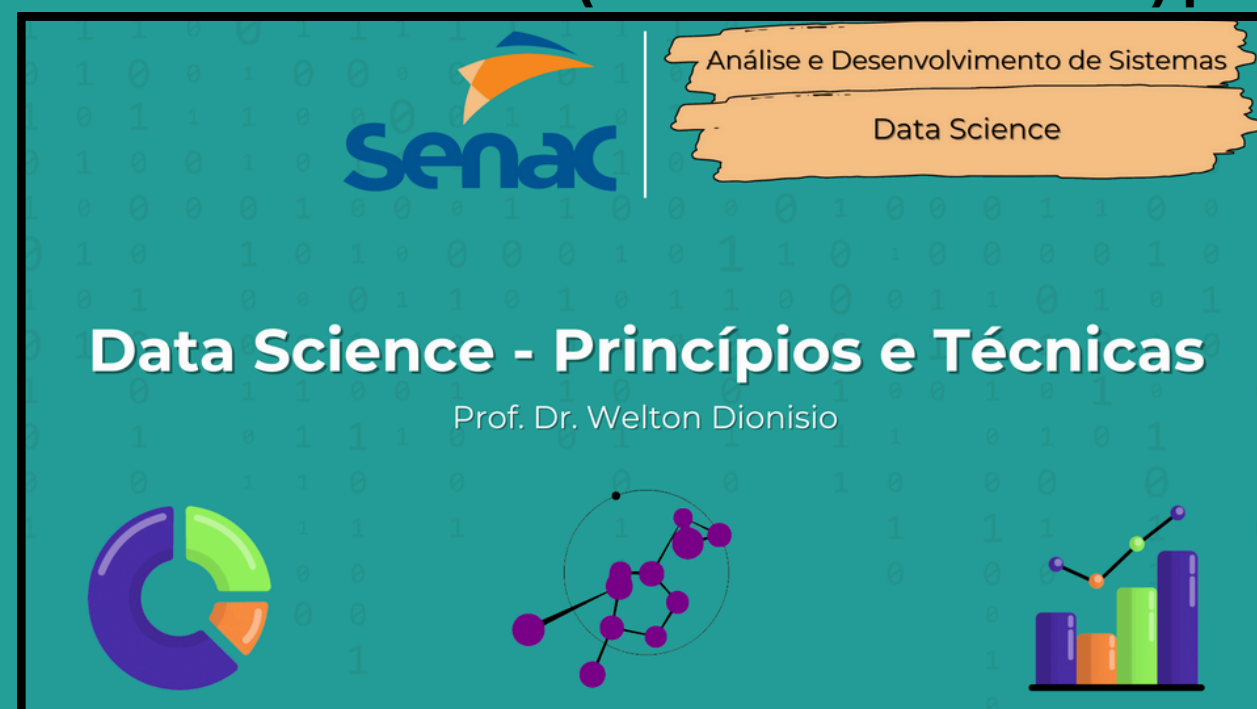


@dr.dosdados



Apostila de Estatística Básica - Data Science.pdf

Slides - Data Science (Prof. Welton Dionísio).pdf



Slides - Data Science - Apostila de Introdução.pdf



## Apostila de Estatística Básica

p/ Data Science – Princípios e Técnicas (60 h – ADS)

Prof. Dr. Welton Dionísio



### 1. Medidas de Tendência Central

**Conceito:** São métricas que representam o centro de um conjunto de dados e, consequentemente, ajudam a compreender a distribuição destes e como essa se aproxima da normalidade (curva de sino, gaussiana).

**Fórmulas:**

- **Média:**  $\bar{x} = \frac{\sum x_i}{n}$
- **Mediana:** valor central quando os dados estão ordenados.
- **Moda:** valor mais frequente.

**Exemplo com Cálculos:**

Dados: [5, 7, 8, 8, 10]

1. **Média:**  $\bar{x} = \frac{5 + 7 + 8 + 8 + 10}{5} = \frac{38}{5} = 7.6$
2. **Mediana:**  
Dados ordenados: [5, 7, 8, 8, 10] → Valor central = 8
3. **Moda:**  
Valor mais repetido = 8

**Códigos:**

```
[3] 1 #Python
2 import numpy as np
3 from scipy import stats
4
5 dados = [5, 7, 8, 8, 10]
6 media = np.mean(dados)
7 mediana = np.median(dados)
8 # Access the mode value directly instead of indexing it.
9 moda = stats.mode(dados).mode
10
11 print(f"Média: {media}, Mediana: {mediana}, Moda: {moda}")
```

Média: 7.6, Mediana: 8.0, Moda: 8