

Lista de Exercícios - Estatística Descritiva

1- Medidas de Tendência Central

Exercício 01: Dado o array de pontuações de 10 alunos em um teste: [75, 82, 90, 68, 78, 85, 92, 70, 76, 88], calcule:

- a) A média.
- b) A mediana.

Exercício 02: Uma empresa registrou o número de unidades vendidas de um produto durante 7 dias: [25, 30, 28, 35, 32, 30, 27]. Calcule:

- a) A média.
- b) A moda usando pandas.

Exercício 03: As idades de 15 participantes de um congresso em Porto de Galinhas são: [28, 35, 42, 28, 30, 35, 45, 28, 32, 38, 40, 35, 28, 31, 36]. Calcule:

- a) A moda usando numpy.
- b) A mediana.

Exercício 04: O tempo de reação (em segundos) de 8 voluntários em um experimento foi: [0.25, 0.30, 0.28, 0.35, 0.26, 0.32, 0.29, 0.31]. Calcule:

- a) A média.
- b) A mediana.

Exercício 05: Uma pesquisa sobre o número de filhos por família em uma comunidade resultou nos seguintes dados: [2, 1, 3, 2, 0, 4, 2, 2, 1, 3]. Calcule:

- a) A média.
- b) A moda usando pandas.

Exercício 06: Os preços (em reais) de um mesmo livro em 5 livrarias diferentes são: [45.50, 48.00, 46.20, 47.50, 46.20]. Calcule:

- a) A média.
- b) A mediana.
- c) A moda usando pandas.

Exercício 07: Dado o array de alturas (em metros) de 12 plantas: [1.5, 1.8, 1.6, 2.0, 1.7, 1.8, 1.9, 1.6, 1.7, 1.8, 2.1, 1.7], calcule:

- a) A média.
- b) A moda usando pandas.

Exercício 08: O número de e-mails recebidos por um profissional durante 6 dias úteis foi: [15, 22, 18, 25, 20, 22]. Calcule:

- a) A média.
- b) A mediana.

Exercício 09: As notas de um aluno em 7 disciplinas foram: [7.0, 8.0, 6.5, 9.0, 7.5, 8.5, 7.0]. Calcule:

- a) A média.
- b) A moda usando pandas.

Exercício 10: Um estudo sobre a duração (em minutos) de chamadas telefônicas resultou nos seguintes dados: [3, 5, 2, 8, 4, 4, 6, 3, 5, 7]. Calcule:

- a) A média.
- b) A mediana.
- c) A moda usando numpy.

2 - Medidas de Dispersão

Exercício 11: Dado o array de temperaturas (em Celsius) registradas em uma semana em Recife: [28, 30, 29, 31, 28, 32, 30], calcule:

- a) A amplitude.
- b) O desvio padrão.

Exercício 12: Um cientista de dados coletou o número de cliques em um anúncio online durante 5 dias: [150, 165, 140, 170, 155]. Calcule:

- a) A variância.
- b) O desvio padrão.

Exercício 13: As alturas (em cm) de 8 jogadores de basquete de um time pernambucano são: [195, 202, 188, 205, 198, 200, 192, 203]. Calcule:

- a) O primeiro quartil (Q1).
- b) O terceiro quartil (Q3).
- c) O intervalo interquartil (IIQ).

Exercício 14: O tempo de espera (em minutos) de 10 clientes em uma fila de banco em Boa Viagem foi: [5, 8, 3, 9, 6, 7, 4, 10, 5, 8]. Calcule:

- a) A mediana.
- b) O intervalo interquartil (IIQ).

Exercício 15: Uma pesquisa registrou o número de livros lidos por 12 pessoas em um ano: [2, 5, 1, 8, 3, 3, 6, 4, 2, 7, 4, 5]. Calcule:

- a) A média.
- b) O desvio padrão.

Exercício 16: Os salários (em milhares de reais) de uma amostra de 7 funcionários de uma startup em Porto Digital são: [3.5, 4.0, 3.8, 4.2, 3.6, 5.5, 3.9]. Calcule:

- a) A média.
- b) O desvio padrão amostral (use a correção de Bessel).

Exercício 17: Dado o array de notas de um aluno em 6 provas: [7.0, 8.5, 6.5, 9.0, 7.5, 8.0], calcule:

- a) A amplitude.
- b) A variância amostral (use a correção de Bessel).

Exercício 18: Um analista financeiro monitorou o retorno percentual de um investimento durante 5 meses: [1.2, 0.8, 1.5, -0.5, 1.0]. Calcule:

- a) O desvio padrão.

Exercício 19: As idades de 9 participantes de um grupo de estudo são: [22, 25, 28, 21, 30, 24, 26, 23, 27]. Calcule:

- a) A mediana.
- b) O primeiro e o terceiro quartis.
- c) O intervalo interquartil (IIQ).

Exercício 20:

Um sensor registrou as seguintes leituras de pressão (em PSI): [10.2, 10.5, 9.8, 10.7, 10.0, 10.3, 10.1]. Calcule:

- a) A variância populacional (assuma que esses são todos os registros relevantes).
- b) O desvio padrão populacional.

3 - Visualização de dados

Exercício 1:

Gráfico de Linhas - Acompanhamento da Inflação no Recife (2024).

```
meses = ['Jan', 'Fev', 'Mar', 'Abr', 'Mai', 'Jun', 'Jul', 'Ago', 'Set', 'Out', 'Nov', 'Dez']
ipca_recife_2024 = [0.55, 0.62, 0.48, 0.59, 0.45, 0.38, 0.42, 0.51, 0.60, 0.53, 0.49, 0.57]
```

Crie um gráfico de linhas mostrando a variação mensal do IPCA no Recife em 2024. Personalize com rótulos, título, marcadores e cor da linha.

Exercício 2:

Gráfico de Barras - Comparativo de Casos de Dengue por Bairro (Recife - 2025)

```
bairros_recife = ['Boa Viagem', 'Imbiribeira', 'Casa Amarela', 'Afogados', 'Boa Vista', 'Madalena']
casos_dengue_2025 = [125, 98, 150, 80, 110, 92]
```

Gere um gráfico de barras verticais comparando o número de casos de dengue por bairro. Adicione rótulos, título e cores distintas para cada barra.

Exercício 3:

Gráfico de Dispersão - Relação entre IMC e Pressão Arterial (Amostra em Olinda)

```
imc_olinda = [22.5, 28.1, 25.3, 30.2, 23.8, 27.5, 26.8, 29.5, 24.1, 31.5]
pressao_sistolica_olinda = [120, 135, 125, 145, 122, 130, 128, 140, 124, 150]
```

Crie um gráfico de dispersão para investigar a relação entre IMC e pressão arterial. Rotule os eixos e adicione um título. Opcionalmente, adicione uma linha de tendência.

Exercício 4:

Histograma - Distribuição da Idade dos Participantes do Festival de Inverno de Garanhuns (2024)

```
idades_festival_garanhuns = [25, 32, 18, 45, 28, 38, 52, 21, 35, 48, 29, 31, 40, 23, 37, 55, 27, 33, 42, 19, 30, 46, 26, 39, 50, 22, 36, 49, 24, 34, 41, 20, 33, 47, 28, 36, 51, 29, 32, 43, 18, 31, 44, 27, 38, 53, 26, 35, 40, 21, 34, 48, 25, 37, 54, 23, 39, 45, 22, 33, 41, 19, 36, 52, 24, 38, 49, 28, 31, 42, 20, 35, 50, 26, 37, 46, 23, 32, 51, 29]
```

Utilize um histograma para visualizar a distribuição da idade dos participantes. Experimente diferentes números de bins e personalize as cores e bordas.

Exercício 5:

Gráfico de Pizza - Distribuição do Eleitorado por Faixa Etária (Recife - Eleições 2026)

```
faixas_etarias = ['16-24', '25-34', '35-44', '45-59', '60+']  
percentual_eleitorado = [22, 28, 20, 18, 12]
```

Crie um gráfico de pizza mostrando a distribuição percentual do eleitorado por faixa etária. Inclua rótulos, porcentagens formatadas e cores distintas.

Exercício 6:

Gráfico de Barras Horizontais - Ranking de Filmes Mais Assistidos no Cinema da Fundação (Recife)

```
filmes = ['Bacurau', 'Aquarius', 'O Som ao Redor', 'Central do Brasil', 'Tropa de Elite']  
espectadores = [450, 380, 520, 600, 410]
```

Gere um gráfico de barras horizontais mostrando o ranking dos filmes mais assistidos. Ordene as barras pelo número de espectadores e personalize as cores.

Exercício 7:

Gráfico de Dispersão com Cores - Relação entre Anos de Estudo e Salário por Escolaridade (Pernambuco)

```
anos_estudo = [8, 12, 16, 6, 10, 18, 11, 14, 9, 15]  
salario = [1500, 2800, 4500, 1200, 2000, 6000, 2200, 3500, 1800, 4000]  
escolaridade = ['Fundamental', 'Médio', 'Superior', 'Fundamental', 'Médio', 'Superior', 'Médio', 'Superior', 'Fundamental', 'Superior']
```

Crie um gráfico de dispersão mostrando a relação entre anos de estudo e salário, com cores diferentes para cada nível de escolaridade. Adicione legenda.

Exercício 8:

Histograma Comparativo - Distribuição do Tempo de Espera em Dois Hospitais (Recife)

```
tempo_espera_hospital_a = np.random.normal(loc=35, scale=10, size=100)  
tempo_espera_hospital_b = np.random.normal(loc=28, scale=8, size=100)
```

Crie um histograma comparativo mostrando a distribuição do tempo de espera nos dois hospitais. Plote os dois histogramas na mesma figura com transparência e adicione uma legenda.

4 - Detecção de outliers

Exercício 1:

Você tem dados da temperatura média diária (°C) em Fernando de Noronha durante um mês:

[28.1, 28.5, 27.9, 28.3, 35.2, 28.0, 28.2, 28.4, 27.8, 28.6, 28.1, 28.3, 27.7, 28.5, 28.2, 28.4, 27.9, 28.3, 28.1, 28.6, 28.0, 28.2, 28.4, 27.8, 28.5, 28.2, 28.3, 27.7, 28.6, 28.1].

Calcule o Z-score para cada temperatura e identifique os dias que são outliers usando um limiar de ± 3 .

Exercício 2:

Os preços (em R\$) de um determinado tipo de artesanato no Mercado de São José são:

[15, 18, 16, 20, 17, 19, 15, 22, 18, 17, 16, 19, 17, 20, 18, 16, 45, 17, 19, 21]

Utilize o método do Intervalo Interquartil ($1.5 * IQR$) para identificar os outliers neste conjunto de preços.

Exercício 3:

Crie um box plot para visualizar a distribuição do seguinte conjunto de dados representando o tempo de viagem (em minutos) de casa para o trabalho de diferentes pessoas na Região Metropolitana do Recife:

[25, 30, 28, 35, 22, 40, 27, 60, 32, 29, 31, 33, 26, 38, 34].

Identifique visualmente os possíveis outliers no gráfico.

Exercício 4:

Você tem a altura (em metros) de alguns edifícios em Boa Viagem:

[80, 90, 75, 85, 105, 78, 82, 88, 70, 95, 81, 86, 77, 92, 150].

a) Identifique os outliers usando o Z-score (± 3).

b) Identifique os outliers usando o método do IIQ ($1.5 * IIQ$).

Exercício 5:

Gere um box plot para o seguinte conjunto de dados representando o salário mensal (em R\$) de profissionais em diferentes áreas em Pernambuco:

[2500, 3000, 2800, 3500, 2200, 4000, 2700, 7000, 3200, 2900, 3100, 3300, 2600, 3800, 3400].

O gráfico deve exibir os outliers identificados.

Exercício 6:

Os casos de Chikungunya registrados por semana no estado de Pernambuco durante um período são:

[12, 15, 10, 18, 20, 13, 16, 25, 11, 14, 17, 19, 12, 16, 22, 14, 18, 35, 15, 17]

Calcule o Z-score para cada semana. Considerando que a distribuição pode não ser perfeitamente normal, quais valores você consideraria suspeitos usando um limiar mais conservador de ± 2.5 ?