

 INSTITUTO FEDERAL Espírito Santo Campus Serra	Especialização em Mineração de Dados Educacionais	
	Componente Curricular: Análise Exploratória de Dados.	
	Professor: Hilário Tomaz Alves de Oliveira	

Lista de Exercícios 2 – Estatística Descritiva

Usando Pandas e Numpy

OBSERVAÇÕES:

- A lista de exercício pode ser resolvida usando um único arquivo de notebook na linguagem Python.
- Para cada uma das questões 1, 2 e 3, você deve criar uma célula de texto e apresentar a sua resposta.
- Para a questão 4, você pode criar uma ou mais células de código.
- Indique explicitamente cada questão e sua respectiva resolução.
- **Equipe:** No máximo 2 pessoas.
- Coloque os nomes dos membros da equipe na 1ª célula do notebook.
- **Data Entrega:** 22/09/2024
- **Nota:** 15,00

Questão 1. Pesquisadores coletaram dados para examinar a relação entre o desempenho acadêmico de alunos do ensino fundamental e médio e suas condições socioeconômicas, bem como as condições das escolas em que estudam, em várias regiões do Brasil. Durante o estudo, o desempenho acadêmico foi medido através das notas médias dos alunos em

disciplinas como Matemática, Português e Ciências. As condições socioeconômicas foram avaliadas com base na renda familiar, nível de escolaridade dos pais, acesso a recursos educacionais além da escola, entre outros. As condições das escolas foram analisadas com base na infraestrutura, disponibilidade de materiais didáticos e qualificação dos professores. Os dados foram coletados sobre 65.000 alunos de diferentes estados brasileiros entre os anos de 2018 e 2020."

Com base no cenário anterior, responda as seguintes perguntas:

- a) Qual é a principal questão de pesquisa que pode ser elaborada para esse estudo?
- b) Quem são os objetos de interesse e quantos estão incluídos?
- c) Quais são as variáveis? Identifique cada variável como numérica ou categórica. Se numérica, classifique-a em discreta ou contínua e, se categórica, em ordinal ou nominal.
- d) Identifique a população de interesse e a amostra deste estudo.
- e) Os resultados do estudo podem ser generalizados para a população?
- f) Os resultados do estudo podem ser usados para estabelecer relações causais?

Questão 2. Encontre a média, a mediana e a moda das distribuições de dados apresentados nas tabelas a seguir.

6	1	1	10	6	4	8	1	6	2
1	4	10	0	4	9	5	6	4	9
10	1	6	7	6	1	4	3	6	0
2	4	5	5	5	2	9	0	6	8
1	8	4	6	1	1	8	7	4	3

Questão 3. O artigo "*Oxygen Consumpition During Fire Suppression: Error of Heart Rate Estimation*" informou os dados a seguir sobre consumo de oxigênio (mL/kg/min) para uma mostra de 10 bombeiros em uma simulação de supressão de incêndios.

29,5	49,3	30,6	28,2	28,0	26,3	33,9	29,4	23,5	31,6
------	------	------	------	------	------	------	------	------	------

A partir dessa amostra de dados, calcule:

- a) A amplitude amostral.
- b) A variância e o desvio padrão amostral.

Questão 4. Analise a base de dados do Microdados do Censo Escolar da Educação Básica 2023, disponível no link a seguir:

https://download.inep.gov.br/dados_abertos/microdados_censo_escolar_2023.zip

Responda as seguintes perguntas sobre essa base de dados:

- a) Quais questões de pesquisa poderiam ser interessantes de serem realizadas a partir dessa base de dados?
- b) Usando a biblioteca do Pandas e outras que julgar necessária e realize as seguintes análises. **OBS:** Para os itens iii até vi, considere somente as escolas do Espírito Santo.
 - i. Essa base de dados tem quantos registros?
 - ii. Essa base de dados tem quantos atributos?
 - iii. Quantas escolas e cidades distintas estão na base de dados?
 - iv. Qual a distribuição de escolas por tipo de localização (Urbana ou Rural)?
 - v. Qual o percentual de escolas que tem acesso à Internet?
 - vi. Qual a média, variância e desvio padrão do número de matrículas dos alunos do 1º ao 9º por ano?