

**SENAI “GASPAR RICARDO JÚNIOR”**

**Ian Lucas**

**Juan Correia**

**Caroline Almeida**

**Matheus Pulcinelli**

**Análise de Dados (Ciência de Dados)**

**Sorocaba**

**2025**

**Ian Lucas**

**Juan Correia**

**Caroline Almeida**

**Matheus Pulcinelli**

**Análise de Dados (Ciência de Dados)**

Trabalho apresentado para a disciplina de Ciência de Dados, do curso de Análise e Desenvolvimento de Sistemas da Escola SENAI ‘Gaspar Ricardo Júnior’.

Orientador: André Souza.

**Sorocaba**

**2025**

**SUMÁRIO**

**1 INTRODUÇÃO**...............................................................................................**04**

**2 REFERENCIAL TEÓRICO**............................................................................**05**

**3 METODOLOGIA**............................................................................................**06**

**4 ANÁLISE DE DADOS**.................................................................................**07-08**

4.1 Escalas de Medição........................................................................ **07-08**

4.2 Medidas de Tendência Central....................................................... **07-08**

4.3 Medidas de Dispersão..................................................................... **07-08**

4.4 Teste de Normalidade (Shapiro-Wilk).............................................. **07-08**

4.5 Correlação de Pearson.................................................................... **07-08**

4.6 Visualizações Estatísticas................................................................ **07-08**

**5 CONCLUSÃO**................................................................................................**09**

**6 REFERÊNCIAS** ............................................................................................**10**

1. **INTRODUÇÃO**

Este projeto tem como objetivo aplicar conceitos estatísticos fundamentais à análise de dados educacionais, utilizando informações simuladas de alunos. Foram consideradas variáveis como gênero, nível de escolaridade, nota de matemática e evasão.

A escolha deste tema permite compreender relações entre desempenho acadêmico e fatores socioeducacionais.

A relevância desta análise está na possibilidade de contextualizar técnicas de ciência de dados aplicadas à realidade da educação básica, possibilitando insights para ações pedagógicas mais eficazes.

1. **REFERENCIAL TEÓRICO**

Neste projeto foram utilizados os seguintes conceitos estatísticos:

* **Escalas de medição:** nominal, ordinal, intervalar e de razão.
* **Medidas de tendência central:** média, mediana e moda.
* **Medidas de dispersão:** variância, desvio padrão e amplitude.
* **Testes de normalidade:** Shapiro-Wilk.
* **Correlação:** coeficiente de Pearson.
* **Visualizações:** histogramas, boxplots e scatterplots.

As bibliotecas Python aplicadas foram:

* **Pandas:** manipulação e análise de dados.
* **NumPy:** operações matemáticas e estatísticas.
* **Matplotlib:** geração de gráficos.
* **Scipy.stats:** testes estatísticos e análises.

1. **METODOLOGIA**

A base de dados foi criada de forma simulada, contendo 20 registros de alunos, com variáveis categóricas (gênero, escolaridade, evasão) e numéricas (nota de matemática).

A técnica utilizada foi a **amostragem estratificada** por gênero, gerando uma amostra de 10 alunos (50% da população), mantendo a proporcionalidade entre os estratos.

As etapas realizadas foram:

* Geração aleatória dos dados.
* Limpeza e padronização das colunas.
* Separação dos dados por estratos e seleção da amostra.
* Análise descritiva e exploratória dos dados.

Ferramentas utilizadas: Python, com bibliotecas Pandas, NumPy, Matplotlib e Scipy.

1. **ANÁLISE DE DADOS**

**4.1 Escalas de Medição**

* **Nota de Matemática:** intervalar.
* **Evasão:** nominal (0 = não evadiu, 1 = evadiu).
* **Gênero e Escolaridade:** nominal.

**4.2 Medidas de Tendência Central**

* **Média:** 5,60.
* **Mediana:** 6,00.
* **Moda:** 6 e 8 (ambos ocorrem duas vezes).

**4.3 Medidas de Dispersão**

* **Desvio padrão:** 2,73.
* **Variância:** 7,47.
* **Amplitude:** 8.

**4.4 Teste de Normalidade (Shapiro-Wilk)**

* **Resultado:** p-valor > 0,05, indicando que a distribuição é aproximadamente normal.

**4.5 Correlação de Pearson**

* **Correlação entre nota e evasão:** r ≈ -0,43.
* Interpretação: Existe uma correlação moderada e negativa, sugerindo que alunos com menores notas tendem a evadir.

**4.6 Visualizações Estatísticas**

* Histogramas das notas.
* Boxplot para análise da distribuição.
* Scatterplot demonstrando a relação entre nota e evasão.
* Gráficos com indicação da faixa de desvio padrão.

1. CONCLUSÃO

A análise estatística realizada permitiu compreender o comportamento dos dados educacionais simulados, evidenciando uma relação relevante entre notas baixas e evasão escolar.

Os resultados reforçam a importância de desenvolver estratégias pedagógicas focadas no acompanhamento dos alunos com baixo rendimento.

Como limitação do estudo, destaca-se o uso de uma base de dados simulada, que não reflete integralmente a realidade.

Para trabalhos futuros, recomenda-se utilizar bases de dados reais e incluir novas variáveis para enriquecer a análise.

1. **REFERÊNCIAS**

FREITAS, H.; et al. Pesquisa qualitativa: aspectos relevantes para a sua concepção e aplicação. Revista de Ciências da Administração, 2000.

PANDAS Documentation. Disponível em: <https://pandas.pydata.org/>. Acesso em: 17 jun. 2025.

SCIPY Documentation. Disponível em: <https://docs.scipy.org/doc/scipy/>. Acesso em: 17 jun. 2025.

MATPLOTLIB Documentation. Disponível em: 17 jun. 2025.

<https://matplotlib.org/stable/index.html>. Acesso em: 17 jun. 2025.