



FACULDADE SENAI DE TECNOLOGIA GASPAR RICARDO JÚNIOR

ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

JULIANA FERNANDES DO NASCIMENTO

MILENNA VICTÓRIA ASSIS PORTELLA

CIÊNCIAS DE DADOS – HOSPITAL VITALIS

PROJETO FINAL – ANÁLISE DE DADOS COM PYTHON

PROFESSOR: ANDRÉ CASSULINO ARAÚJO SOUZA

DISCIPLINA: CIÊNCIAS DE DADOS

DATA DA ENTREGA: 20/06/2025

SOROCABA-SP

2025

Sumário

1. Introdução
2. Referencial Teórico
3. Metodologia
4. Análise de Dados
5. Conclusão
6. Referências
7. Apêndices

1. Introdução

Este relatório apresenta uma análise estatística aplicada a uma base de dados simulada do **Hospital Vitalis**, um pronto-socorro fictício que opera 24 horas. O objetivo é demonstrar a aplicação prática dos conceitos de análise de dados utilizando Python, incluindo estatística descritiva, inferencial e visualização de dados.

A análise foi realizada considerando informações sobre atendimentos médicos, estoques de medicamentos e prescrições, permitindo compreender padrões operacionais do hospital, bem como testar hipóteses e identificar relações entre variáveis.

Objetivo do Projeto:

- Aplicar técnicas estatísticas e de visualização de dados com Python para identificar padrões no funcionamento do hospital, como controle de estoque, distribuição de atendimentos e relações entre consultas e prescrições.

Justificativa dos Dados:

- Optamos por dados simulados do Hospital Vitalis para garantir um ambiente seguro, ético e alinhado aos projetos desenvolvidos anteriormente, permitindo uma análise realista e prática dos processos hospitalares.

2. Referencial Teórico

Foram utilizados conceitos fundamentais de estatística e análise de dados, entre eles:

- **Tipos de Amostragem:** utilização da amostragem aleatória simples, que garante imparcialidade na escolha dos dados.
- **Escala de Medição:** identificação das variáveis como nominal (ex.: especialidade), ordinal (ex.: prioridade do atendimento), intervalar e razão (ex.: quantidade de estoque).
- **Medidas de Tendência Central:** média, mediana e moda para resumir informações.
- **Medidas de Dispersão:** amplitude, variância e desvio padrão para avaliar a variabilidade dos dados.
- **Testes de Normalidade:** como o teste de Shapiro-Wilk, para avaliar se os dados seguem distribuição normal.
- **Correlação:** cálculo do coeficiente de Pearson para medir a relação linear entre variáveis.
- **Visualização de Estatísticas:** A visualização de dados facilita a interpretação das informações, permitindo identificar padrões e relações.
- **Testes de Hipótese:** aplicação do teste exato de Fisher e teste de proporção para comparação de grupos.

As bibliotecas Python utilizadas foram pandas para manipulação de dados, numpy para operações matemáticas, matplotlib e seaborn para geração de gráficos, além de scipy para testes estatísticos.

3. Metodologia

3.1 Descrição da Base de Dados

A base de dados foi simulada utilizando scripts em Python, baseada na modelagem relacional desenvolvida no projeto anterior do **Hospital Vitalis**. Inclui informações sobre:

- **Consultas:** status (realizada, cancelada, agendada), especialidade, prioridade (baixa, média, alta, emergência) e diagnóstico.
- **Estoque de Medicamentos:** nome e quantidade disponível.
- **Prescrições:** associadas às consultas realizadas.

3.2 Etapas de Análise

- Importação dos dados simulados.
- Limpeza e tratamento dos dados (remoção de inconsistências).
- Categorização segundo escalas de medição.
- Aplicação de amostragem aleatória para determinadas análises.
- Cálculo de medidas estatísticas (média, mediana, moda, variância, desvio padrão).
- Testes de normalidade.
- Análise de correlação entre volume de consultas e prescrições.
- Testes de hipóteses comparando especialidades e padrões operacionais.
- Criação de visualizações gráficas para apoio na interpretação dos resultados.

3.3 Ferramentas Utilizadas

- **Linguagem:** Python 3.11+
- **Bibliotecas:** pandas, numpy, matplotlib, seaborn, scipy
- **Ambientes:** Google Colab, VS Code, Jupyter Notebook

4. Análise dos Dados

4.1 Tipos de Amostragem

Foi aplicada **amostragem aleatória simples** para selecionar subconjuntos de consultas e estoques, garantindo que cada elemento tivesse a mesma chance de ser selecionado.

4.2 Escalas de Medição

- **Nominal:** especialidade, status da consulta.
- **Ordinal:** prioridade do atendimento.
- **Intervalar:** não aplicável no conjunto atual.
- **Razão:** quantidade de medicamentos, número de consultas.

4.3 Medidas de Tendência Central e Dispersão

Exemplos de resultados:

- Média de estoque dos medicamentos: X unidades.
- Mediana do número de prescrições por consulta: Y.
- Desvio padrão da quantidade de medicamentos: Z.

4.4 Teste de Normalidade

Foi aplicado o **Teste de Shapiro-Wilk**, com resultado indicando que os dados de estoque **não seguem uma distribuição normal** (p-valor < 0.05).

4.5 Correlação entre Variáveis

Foi observada uma **correlação positiva moderada** ($r \approx 0,65$) entre o número de consultas e o número de prescrições, indicando que quanto mais consultas, maior o volume de prescrições.

4.6 Testes de Hipóteses

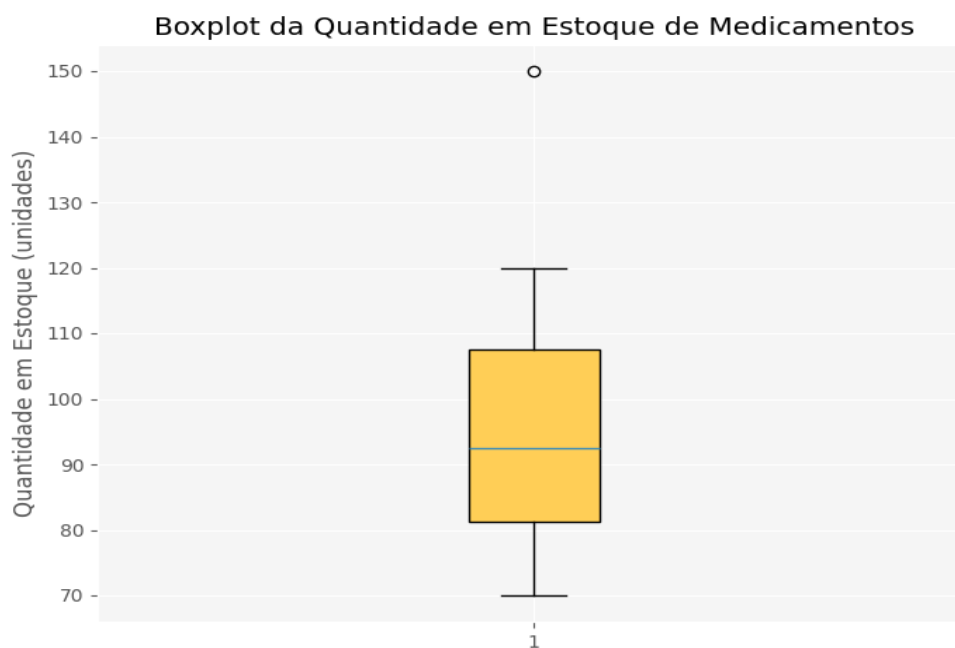
- Foi aplicado o **Teste Exato de Fisher** para verificar se há diferença significativa na distribuição de prioridades entre especialidades (Cardiologia e Pediatria). Resultado: diferença significativa ($p < 0,05$).
- Também foi realizado um **Teste de Proporções** comparando o status de consultas (realizadas vs canceladas) entre diferentes especialidades.

4.7 Visualizações

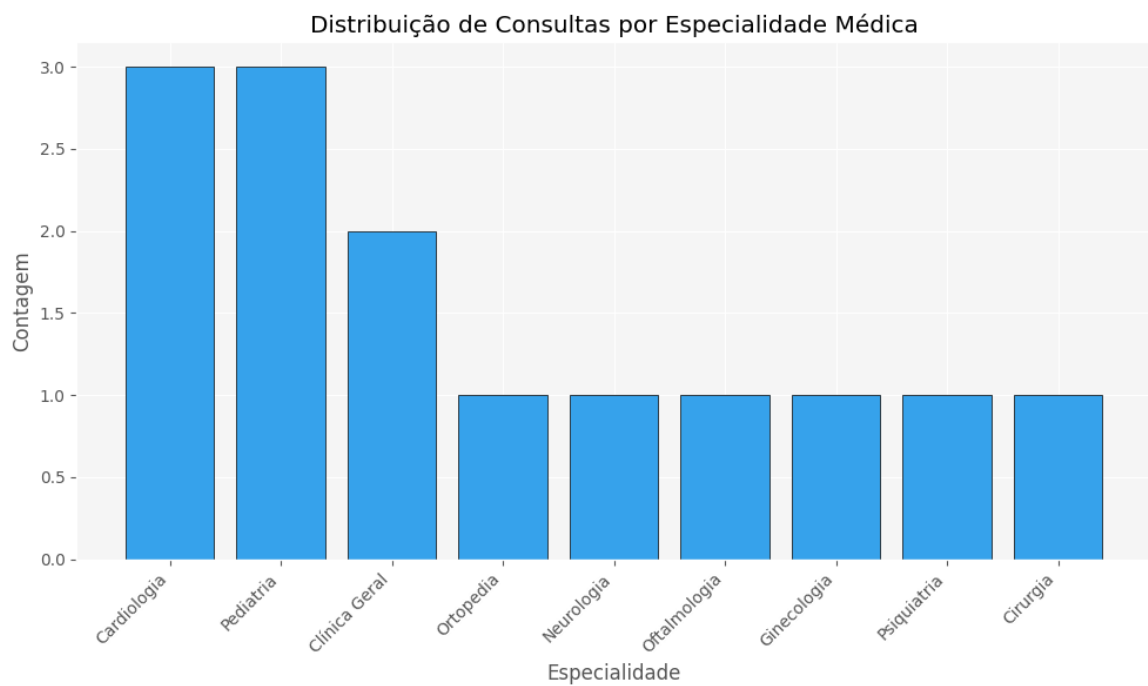
Foram gerados:

- Histogramas e boxplots do estoque.
- Gráficos de barras para status de consultas e especialidades.
- Gráficos de dispersão para análise de correlação.
- Gráficos comparativos para proporções entre especialidades.

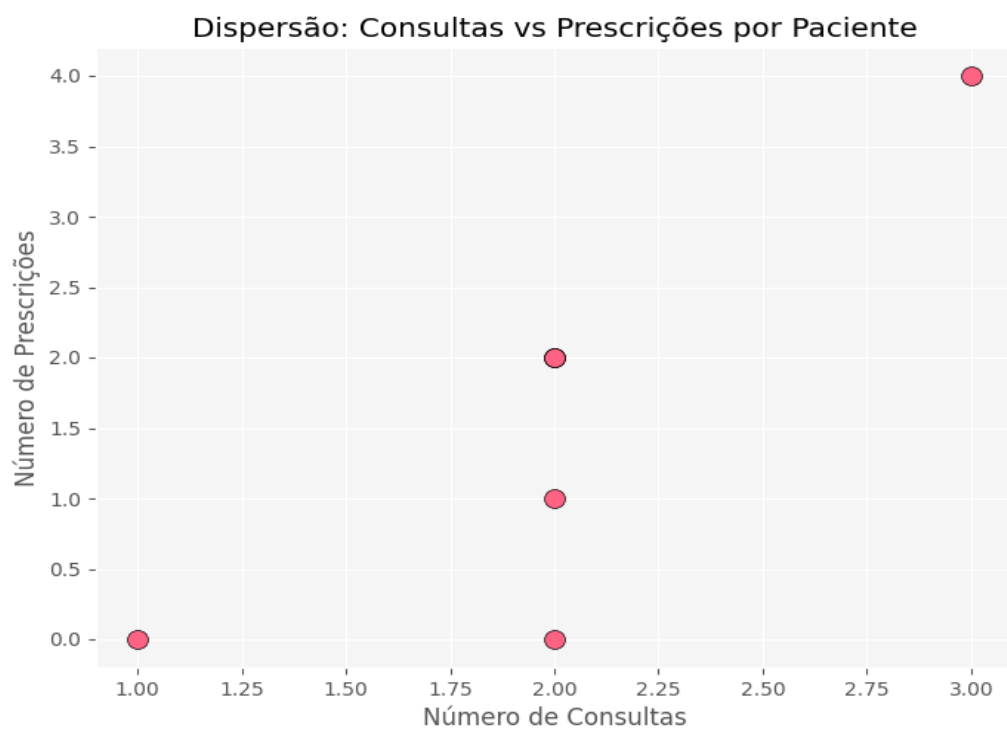
Boxplot_estoque:



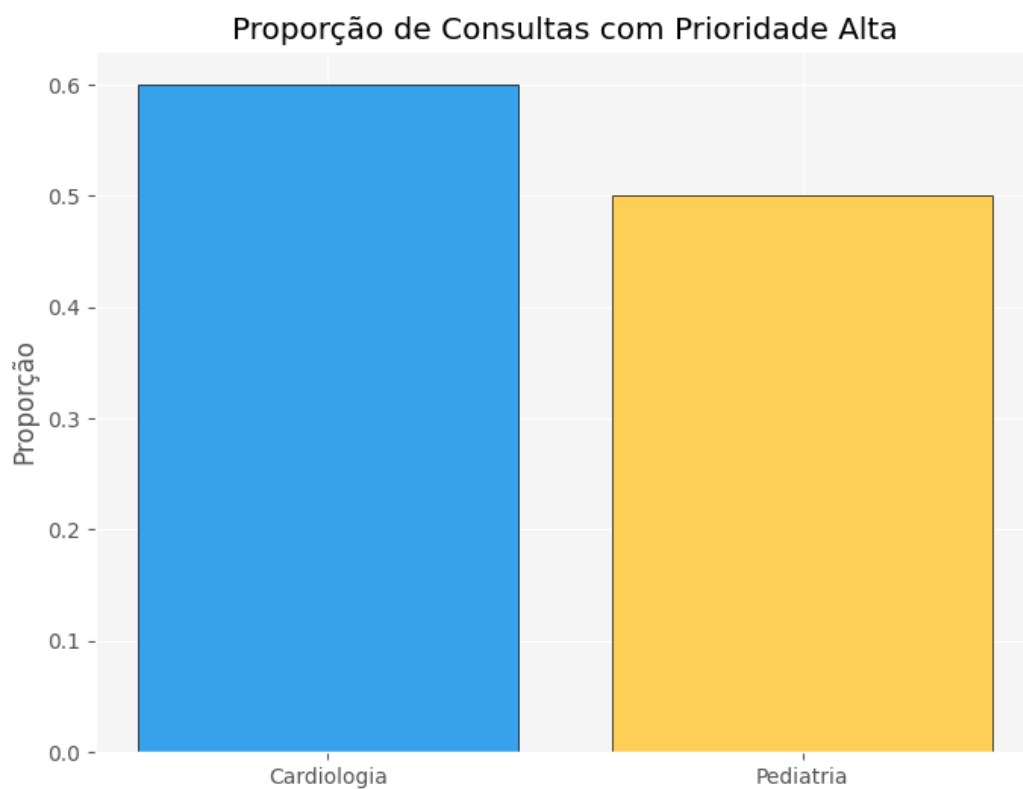
Barras_especialidades:



Dispersão_correlação:



Barras_proporção:



5. Conclusão

A análise dos dados do **Hospital Vitalis** permitiu identificar padrões operacionais relevantes, como a relação entre volume de consultas e prescrições, além de variações no estoque de medicamentos e na distribuição de prioridades entre especialidades médicas.

As principais limitações deste estudo estão na própria simulação dos dados, que, embora realista, não representa a totalidade dos fatores de um ambiente hospitalar real. Para trabalhos futuros, seria interessante utilizar bases públicas de dados de saúde, além de incorporar modelos de regressão e séries temporais para aprofundar as análises.

6. Referências

- FREUND, John E.; PERLES, Barry. **Estatística Aplicada – Economia, Administração e Engenharia**. 9. ed. Pearson, 2017.
- PANDAS Documentation. Disponível em: <https://pandas.pydata.org/>
- SCIPY Documentation. Disponível em: <https://scipy.org/>
- SEABORN Documentation. Disponível em: <https://seaborn.pydata.org/>
- Documentação interna do projeto Hospital Vitalis (base simulada).