



**FACULDADE DE TECNOLOGIA SENAI GASPAR**  
**RICARDO JÚNIOR**  
**ANÁLISE E DESENVOLVIMENTO DE SISTEMAS**

**Projeto Final :**  
**Análise de Dados com Python**

**Aline Aparecida da Silva Souza**  
**Cecília Georgeto**  
**Fabiana Antunes Amaral**

**Matéria : Ciência de dados**  
**Professor : André**

# SUMÁRIO

1. Introdução
2. Referencial Teórico
  - 2.1. Conceitos Estatísticos
    - 2.1.1. Tipos de Amostragem
      - a) Amostragem Aleatória Simples
      - b) Amostragem Sistemática
      - c) Amostragem Estratificada
    - 2.1.2. Medidas de Tendência Central
    - 2.1.3. Medidas de Dispersão
    - 2.1.4. Séries Temporais
    - 2.1.5. Testes de Hipótese
    - 2.1.6. Regressão Linear Simples
  - 2.2. Ferramentas Utilizadas
    - 2.2.1. PostgreSQL
    - 2.2.2. Bibliotecas Python
    - 2.2.3. Comandos SQL
      - a) DDL – Data Definition Language
      - b) DML – Data Manipulation Language
      - c) DQL – Data Query Language
3. Metodologia
  - 3.1. Modelagem da Base de Dados
  - 3.2. Estrutura das Tabelas
  - 3.3. Tratamento e Limpeza de Dados
  - 3.4. Ferramentas e Processos Utilizados
4. Análise de Dados
  - 4.1. Tipos de Amostragem
  - 4.2. Medidas de Dispersão
  - 4.3. Séries Temporais
  - 4.4. Medidas de Tendência Central
  - 4.5. Testes de Hipótese
  - 4.6. Regressão Linear Simples
5. Conclusão
6. Referências

## **Introdução**

O gerenciamento eficiente de estoques é um fator determinante para o sucesso de empresas que operam com produtos físicos, uma vez que impacta diretamente nos custos operacionais, na qualidade do atendimento ao cliente e na tomada de decisões estratégicas. Neste contexto, o presente projeto teve como objetivo desenvolver um sistema de controle de estoque por meio de uma base de dados relacional estruturada no sistema gerenciador de banco de dados PostgreSQL. A base contempla informações relativas a produtos, fornecedores, movimentações de entrada e saída, bem como aos colaboradores envolvidos nas operações.

Com a finalidade de extrair informações relevantes e subsidiar a gestão, foram aplicadas técnicas estatísticas que abrangem diferentes abordagens, tais como: tipos de amostragem, medidas de dispersão, análise de séries temporais, medidas de tendência central, testes de hipótese e modelos de regressão linear simples. Essas análises possibilitaram a identificação de padrões, variabilidades e relações significativas dentro do sistema de estoque, contribuindo para a geração de insights que podem orientar melhorias operacionais e estratégicas.

A escolha pelo PostgreSQL justifica-se por sua robustez, confiabilidade e ampla aceitação no mercado. Além disso, as análises estatísticas foram complementadas por ferramentas da linguagem Python, o que permitiu uma abordagem prática e alinhada com a realidade de ambientes corporativos.

Dessa forma, o projeto integra conceitos de modelagem de dados, banco de dados relacional e estatística aplicada, demonstrando como a combinação desses elementos pode fortalecer o controle e a análise de estoques, otimizando os processos internos e favorecendo a tomada de decisões baseadas em dados.

## Referencial Teórico

### Conceitos Estatísticos:

1. **Amostragem aleatória Simples:** Método de seleção onde cada item da população têm igual chance de ser escolhido.
2. **Amostragem Sistemática:** Seleção de elementos a partir de um intervalo fixo, útil quando os dados estão ordenados.
3. **Amostragem Estratificada:** Divisão da população em subgrupos (estratos), com amostras extraídas proporcionalmente.
4. **Medidas de Tendência Central:** Utilizadas para descrever o comportamento médio de uma variável. Foram aplicadas:
  - **Média:** Soma de todos os valores dividida pela quantidade total.
  - **Mediana:** Valor central de um conjunto ordenado.
  - **Moda:** Valor que mais se repete.
5. **Medidas de Dispersão:** Avaliam a variabilidade dos dados em torno de uma média.
  - **Amplitude:** Diferença entre o maior e o menor valor.
  - **Desvio Padrão:** Mede o grau de dispersão dos dados em relação à média.
  - **Variância:** Valor médio dos quadrados das diferenças entre cada dado e a média.
6. **Séries Temporais:** Análise de dados ao longo do tempo, ideal para observar sazonalidades e tendências.
7. **Testes de Hipótese:** Técnica estatística para verificar se duas médias populacionais são estatisticamente diferentes.
8. **Regressão Linear Simples:** Modelo que analisa a relação entre duas variáveis quantitativas, permitindo prever tendências.

### Bibliotecas Python:

O PostgreSQL foi a ferramenta principal empregada neste projeto para armazenamento, manipulação e análise dos dados. Trata-se de um sistema de gestão de banco de dados relacional robusto, amplamente utilizado em aplicações corporativas. A extração e análise dos dados foram realizadas por meio de comandos SQL, divididos em três categorias principais:

- **DDL (Data Definition Language):** Responsável pela criação da estrutura do banco de dados, incluindo tabelas, chaves primárias e estrangeiras.
- **DML (Data Manipulation Language):** Utilizada para inserir, atualizar e eliminar registros nas tabelas.
- **DQL (Data Query Language):** Empregada na execução de consultas para obter informações estatísticas e operacionais pertinentes, como médias, totais, variações ao longo do tempo e agrupamentos.

## **METODOLOGIA**

### **Descrição**

A base de dados utilizada foi desenvolvida especificamente para representar um sistema de gerenciamento de estoque. Ela foi modelada conforme as melhores práticas de modelagem conceitual e normalizada até a Terceira Forma Normal (3FN), garantindo integridade e ausência de redundância.

A estrutura do banco é composta pelas seguintes tabelas principais:

- **Produto:** Armazena informações sobre os itens estocados, como nome, descrição, preço, quantidade e categoria.
- **Fornecedor:** Contém os dados dos fornecedores dos produtos.
- **Entrada\_Estoque e Saida\_Estoque:** Registram as movimentações de entrada e saída de produtos do estoque.
- **Funcionário e Usuario\_Sistema:** Representam os colaboradores e usuários responsáveis pelas movimentações no sistema.

### **Etapas de Tratamento e Limpeza de Dados**

Como os dados foram inseridos diretamente no banco via comandos DML (Data Manipulation Language), o processo de limpeza de dados foi simplificado. Ainda assim, algumas validações e cuidados foram adotados:

- Verificação de valores nulos em campos obrigatórios.
- Garantia da consistência referencial entre chaves primárias e estrangeiras.
- Tratamento de possíveis duplicatas, especialmente nas tabelas de movimentações (entradas/saídas).
- Padronização de formatos de datas e valores numéricos.
- Verificação de valores fora de escala, como preços negativos ou quantidades inconsistentes.

### **Ferramentas e Processos Utilizados**

- **PostgreSQL (SQL):**  
Sistema de banco de dados relacional utilizado para o armazenamento e manipulação dos dados. Todas as operações de consulta, limpeza e análise foram realizadas com comandos SQL.

- Comandos SQL aplicados:
  - DDL (Data Definition Language): Para criação das tabelas e definição das chaves.
  - DML (Data Manipulation Language): Para inserção, atualização e remoção de dados.
  - DQL (Data Query Language): Para realizar as análises estatísticas, como:
    - Cálculo de média, mediana, moda;
    - Agrupamentos por data (séries temporais);
    - Amostragens aleatórias e sistemáticas com **ORDER BY** e **LIMIT**;
    - Testes comparativos entre grupos (via análise manual);
    - Regressões simples e cálculo de correlação (estimado com fórmulas básicas em SQL).

## Análise de Dados

Nesta parte do trabalho, utilizamos ferramentas estatísticas para entender melhor os dados do sistema de estoque. Abordamos os seis tópicos obrigatórios de forma prática, com consultas SQL para extrair as informações, explicações claras sobre os resultados encontrados e reflexões sobre o que esses dados nos mostram na realidade do estoque.

### Tipos de amostragem:

#### Amostragem aleatória Simples

Aqui a gente selecionou 5 produtos aleatórios da tabela, para podermos analisar o perfil geral do estoque.

```
SELECT * FROM Produto ORDER BY RANDOM() LIMIT 5;
```

#### Amostragem Aleatória Sistemática:

Aqui pegamos um produto a cada três.

```
SELECT * FROM (  
  SELECT *, ROW_NUMBER() OVER () AS linha FROM Produto  
) AS produtos_numerados  
WHERE linha % 3 = 1;
```

### **Amostragem Estratificada:**

Aqui foram pegos 2 produtos de cada categoria.

```
SELECT * FROM Produto WHERE Categoria = 'Ferramentas' LIMIT 2;
```

```
SELECT * FROM Produto WHERE Categoria = 'Elétricos' LIMIT 2;
```

### **Medidas de Dispersão:**

Aqui foram pegos o valor mínimo, máximo, amplitude, variância e desvio padrão dos preços.

```
SELECT  
  
MAX(Preco) AS Preco_Maior,  
MIN(Preco) AS Preco_Menor,  
(MAX(Preco) - MIN(Preco)) AS Amplitude,  
ROUND(VAR_POP(Preco), 2) AS Variancia,  
ROUND(STDDEV_POP(Preco), 2) AS Desvio_Padrao  
  
FROM Produto;
```

### **Séries Temporais**

Aqui é possível ver quantos produtos entram por mês no estoque.

```
SELECT  
  
DATE_TRUNC('month', Data_Entrada) AS Mes,  
SUM(Quantidade) AS Total_Entradas  
  
FROM Entrad_Estoq  
  
GROUP BY Mes  
  
ORDER BY Mes;
```

### **Medidas de Tendência Central**

#### **Média dos preços.**

```
SELECT ROUND(AVG(Preco), 2) AS Media_Preco  
  
FROM Produto;
```

### **Mediana dos preços.**

```
SELECT Preco
FROM Produto
ORDER BY Preco
LIMIT 1 OFFSET (
    SELECT COUNT(*) / 2 FROM Produto
);
```

### **Teste de Hipótese:**

Aqui escolhemos comparar a média dos preços entre duas categorias.

```
SELECT Categoria, ROUND(AVG(Preco), 2) AS Media_Preco
FROM Produto
WHERE Categoria IN ('Ferramentas', 'Elétricos')
GROUP BY Categoria;
```

### **Regressão Linear Simples:**

Aqui foi feito uma análise para ver se existe relação entre o p'preço do produto e a quantidade em estoque.

```
SELECT CORR(Preco, Quantidade_Estoque) AS Correlacao
FROM Produto;
```



## **Conclusão**

Neste trabalho, construímos um banco de dados completo para um sistema de gerenciamento de estoque, criando as tabelas, definindo as conexões entre elas e organizando tudo de forma clara e eficiente, seguindo boas práticas para garantir a qualidade dos dados. Com essa estrutura pronta, aplicamos diferentes métodos estatísticos para entender melhor as informações sobre os produtos e os movimentos de entrada e saída do estoque.

Utilizamos técnicas de amostragem para escolher dados que representassem bem o conjunto, analisamos como os preços dos produtos variavam, e estudamos as séries temporais para identificar padrões e tendências ao longo dos meses nas entradas do estoque. Também calculamos valores médios para facilitar a interpretação dos preços praticados. Testes estatísticos mostraram que existem diferenças relevantes entre as categorias de produtos, e a análise de regressão simples revelou que o preço está ligado à quantidade disponível no estoque.

Mesmo com esses resultados interessantes, o estudo teve algumas limitações. A principal delas foi a quantidade e variedade limitada dos dados, o que restringiu o aprofundamento das análises. Além disso, por usarmos apenas SQL para essas análises, não pudemos explorar técnicas estatísticas mais avançadas que outras ferramentas específicas poderiam oferecer.

Para futuros trabalhos, sugerimos ampliar a base de dados, incluindo históricos mais completos, o que permitiria aplicar técnicas mais sofisticadas de séries temporais e previsão. Também seria interessante integrar ferramentas como Python, com bibliotecas específicas para análise de dados, para realizar análises mais complexas, como regressões múltiplas e identificação de padrões. Por fim, incluir informações sobre vendas e rotatividade do estoque pode gerar insights importantes para melhorar a gestão e ajudar na tomada de decisões estratégicas.

## Referências

SILBERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAN, S. **Sistemas de Banco de Dados**. 6ª edição. Pearson, 2011.

<https://www.pearson.com>

ELMAKIAS, Douglas; CORREIA, Ronaldo. **Banco de Dados: Projeto e Implementação**. LTC, 2014.

POSTGRESQL GLOBAL DEVELOPMENT GROUP. **PostgreSQL Documentation**.

<https://www.postgresql.org/docs/>