

**projeto de disciplina** ➤➤➤➤

# **Estatística Computacional**

PPCOMP - UFAPE 2025.2

# Tema 1

Análise Estatística de Bases de Dados Reais

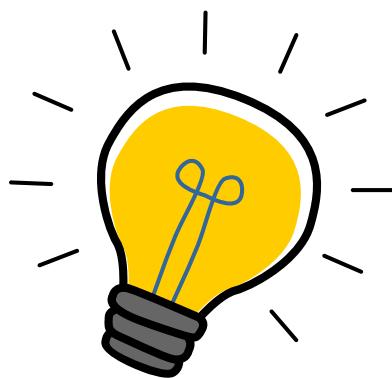
**Foco:** Inferência sobre a população, comparação de grupos e validação de padrões.

- **Objetivo:** Caracterizar a população de interesse através de estimadores pontuais e intervalares, e responder às perguntas de pesquisa utilizando Testes de Hipótese.
- **Roteiro de Execução:**
  - a. Design Experimental (Definição das Perguntas): Definir explicitamente três perguntas de pesquisa que discutam padrões, regras subjacentes, correlações e/ou causalidades entre atributos da base de dados.
  - b. Estimação de Parâmetros e Incerteza (Intervalos de Confiança)
  - c. Execuções de Testes de Hipótese
  - d. Conclusões baseadas em evidências

# Tema 1

Análise Estatística de Bases de Dados Reais

**Foco:** Inferência sobre a população, comparação de grupos e validação de padrões.



**Comparação de Médias/Medianas:** "Há evidência estatística suficiente ( $p\text{-valor} < 0.05$ ) para afirmar que o parâmetro  $X\$$  (ex: desempenho, tempo de recuperação) no Grupo A é superior ao do Grupo B, ou a diferença observada é fruto do acaso amostral?"

**Intervalos de Confiança (ICs):** "o Intervalo de Confiança de 95% inclui valores considerados críticos ou de risco para o problema (ex: lucro zero, temperatura de ebulição)?"

**Correlação e Causalidade:** "Existe uma correlação estatisticamente significativa entre as variáveis  $X\$$  e  $Y\$$ ? O que isso significa?"

**Tamanho de Amostra:** "Considerando a variância estimada na amostra piloto, o tamanho atual da amostra fornece um poder estatístico suficiente, ou é necessário coletar mais dados?"

# Tema 2

Simulação Estocástica (Monte Carlo / Markov / Random Walk)

**Foco:** Construir um dataset artificial robusto para suprir escassez de dados

- **Objetivo:** Gerar uma base de dados sintética que preserve as propriedades estatísticas (média, variância, correlações, distribuição) de uma população teórica ou de uma pequena amostra real.
- **Roteiro de Execução:**
  - a. **Modelagem da População (Input):** Definir as distribuições de probabilidade e parâmetros que regem as variáveis (ex: A variável X segue uma Normal, o processo Y é modelado por uma Cadeia de Markov).
  - b. **Motor de Geração:** Implementar o algoritmo de geração usando Monte Carlo (para variáveis independentes), Random Walk (para séries temporais) ou Markov Chain (para dados categóricos sequenciais).
  - c. **Geração da Base:** Criar o dataset com um N grande (ex: \$N=10.000\$ linhas)
  - d. **Design Experimental (Definição das Perguntas):** Definir explicitamente três perguntas de pesquisa que discutam padrões, regras subjacentes, correlações e/ou causalidades entre atributos da base sintética.
  - e. **Validação Estatística:** Avaliar as distribuições de probabilidade para checar se os dados sintéticos não diferem estatisticamente da distribuição teórica esperada.
  - f. **Análise de Convergência:** Construir Intervalos de Confiança para mostrar que, à medida que a base sintética cresce, as estimativas dos parâmetros convergem para os valores reais.

# Tema 3

Reprodução de Artigo e Variação de Condições

Foco: Robustez do modelo, validação de premissas e reproduzibilidade.

- **Objetivo:** Replicar computacionalmente um estudo de simulação estocástica publicado na literatura, verificando a reproduzibilidade dos resultados originais (baseline) e avaliando a robustez do modelo sob novas condições experimentais.
- **Roteiro de Execução:**
  - a. **Seleção do modelo:** Escolher um artigo que utilize Monte Carlo, Markov Chain ou Random Walk.
  - b. **Reprodução do Baseline (O Grupo de Controle):** Implementar o algoritmo exatamente como descrito no artigo; Gerar os resultados (tabelas ou gráficos) e compará-los visualmente com a publicação original para validar a implementação.
  - c. **Design Experimental (Definição das Perguntas):** Definir explicitamente três perguntas de pesquisa que desafiem o modelo original.
  - d. Execução das Variações (As Simulações de Teste): Rodar as simulações alterando os parâmetros definidos nas perguntas de pesquisa.

# Requisitos básicos

- Introdução e Motivação
- Definição de ao menos três perguntas (não óbvias) de pesquisa
- Metodologia Computacional
- Resultados e Evidências
- Análise Crítica dos Resultados