CEFET/RJ

# Avaliação Experimental

Uma jornada pelos fundamentos metodológicos da pesquisa científica em engenharia e ciência da computação

**Eduardo Ogasawara**

eduardo.ogasawara@cefet-rj.br
https://eic.cefet-rj.br/~eogasawara

## Tipos de Avaliação Experimental

A avaliação experimental é um pilar fundamental da pesquisa científica, permitindo validar hipóteses e comparar diferentes abordagens. Podemos classificar os experimentos em duas categorias principais, cada uma com suas características e desafios específicos.

### Processos Automatizados

Avaliam algoritmos, workflows e simulações computacionais sem intervenção humana direta.

**Exemplo:** Avaliação de desempenho de um modelo de reconhecimento facial em diferentes condições de iluminação, testando precisão e tempo de resposta.

### Processos Manuais

Envolvem participação humana através de questionários, experimentos controlados e estudos qualitativos.

**Exemplo:** Estudo de usabilidade de uma nova interface de software, avaliando satisfação e facilidade de uso reportadas pelos participantes.

# Ética em Avaliação Experimental

Quando um experimento envolve seres humanos, é obrigatório obter aprovação de um comitê de ética em pesquisa. Esta exigência garante a proteção dos direitos, segurança e bem-estar dos participantes, seguindo princípios éticos estabelecidos nacionalmente e internacionalmente.

## Casos que Exigem Aprovação Ética

### Estudos de Usabilidade

Testes de interfaces de usuário que envolvem observação e coleta de dados sobre comportamento e preferências dos participantes.

### Experimentos Psicológicos

Testes que avaliam aspectos comportamentais, cognitivos ou emocionais dos participantes através de estímulos controlados.

### Coleta de Dados Sensíveis

Pesquisas que coletam informações biométricas, como reconhecimento facial, impressões digitais ou dados de saúde pessoal.

# Procedimento para Avaliações Baseadas em Formulário

Conduzir pesquisas com participantes humanos exige um processo estruturado e rigoroso. Cada etapa é essencial para garantir a validade científica dos resultados e o respeito aos direitos dos participantes. O planejamento cuidadoso evita problemas éticos e metodológicos.

01

## Elaborar um Questionário Bem Estruturado

Criar perguntas claras, objetivas e relevantes que capturem as informações necessárias sem ambiguidades ou viés.

02

## Obter Aprovação do Comitê de Ética

Submeter o protocolo de pesquisa para análise, demonstrando que o estudo respeita princípios éticos fundamentais.

03

## Coletar o TCLE

Apresentar o Termo de Consentimento Livre e Esclarecido e obter a assinatura de cada participante antes da coleta de dados.

04

## Definir Número Adequado de Participantes

Calcular o tamanho amostral necessário para obter resultados estatisticamente significativos e generalizáveis.

05

## Analisar os Dados

Aplicar métodos qualitativos e quantitativos apropriados para extrair insights significativos das respostas coletadas.

**Exemplo Prático:** Em uma pesquisa sobre experiência do usuário (UX) em um novo aplicativo, cada etapa assegura que os resultados sejam confiáveis e que os participantes sejam tratados com respeito e transparência.

# Termo de Consentimento Livre e Esclarecido (TCLE)


## Por Que é Necessário?

O TCLE é um documento fundamental que estabelece um contrato ético entre pesquisador e participante. Ele garante transparência total sobre todos os aspectos da pesquisa.

## Elementos Essenciais

- Assegura que os participantes compreendam plenamente os objetivos da pesquisa
- Define claramente riscos potenciais e benefícios esperados
- Garante confidencialidade e anonimização dos dados coletados
- Deve ser assinado antes de qualquer coleta de dados
- Permite que o participante retire-se da pesquisa a qualquer momento

**Referência Legal:** Diretrizes éticas do Conselho Nacional de Saúde

https://www.gov.br/conselho-nacional-de-saude/pt-br/acesso-a-informacao/legislacao/resolucoes/2012/resolucao-no-466.pdf/view

# Definição de Baseline

O baseline representa o método de referência usado para comparação na avaliação experimental. Escolher um baseline apropriado é crucial para validar a contribuição real de uma nova abordagem. Ele deve representar o estado da arte para o problema estudado, não apenas uma solução trivial.

## Evolução dos Baselines

Os métodos considerados estado da arte evoluem constantemente com os avanços tecnológicos.

**Exemplo em PLN:** No processamento de linguagem natural, os transformers substituíram word embeddings e redes LSTM como baseline atual devido ao seu desempenho superior.

## Múltiplos Baselines

Em problemas complexos, um único baseline pode ser insuficiente. Diferentes métodos podem ser usados como referência para avaliar aspectos distintos da solução.

**Exemplo:** Em previsões meteorológicas, pode-se usar modelos estatísticos tradicionais e redes neurais recorrentes como baselines diferentes.

**Table 2**
Performance results for temperature forecasting using the previous five observations (grids) to predict the next five observations (5 → 5), and the next 15 observations (5 → 15). We highlight the lowest values among the models.

| | 5 → 5 | | | | |
|---|---|---|---|---|---|
| Model | RMSE | MAE | Memory usage (MB) | Mean training time | Training time/epoch |
| ARIMA | 2.1880 | 1.9005 | – | – | – |
| ConvLSTM [19] | 1.8555 ± 0.0033 | 1.2843 ± 0.0028 | 922 | **02:38:27** | 00:02:21 |
| PredRNN [25] | 1.6962 ± 0.0038 | 1.1885 ± 0.0020 | 2880 | 06:59:34 | 00:05:52 |
| MIM [26] | 1.6731 ± 0.0099 | 1.1790 ± 0.0055 | 4145 | 11:05:37 | 00:10:43 |
| STConvS2S-C (ours) | 1.3699 ± 0.0024 | 0.9434 ± 0.0020 | 1040 | 03:34:52 | 00:02:48 |
| STConvS2S-R (ours) | **1.2692** ± 0.0031 | **0.8552** ± 0.0018 | **895** | 03:15:12 | **00:02:13** |
| | 5 → 15 | | | | |
| Model | RMSE | MAE | Memory usage (MB) | Mean training time | Training time/epoch |
| ARIMA | 2.2481 | 1.9077 | – | – | – |
| ConvLSTM [19] | 2.0728 ± 0.0069 | 1.4558 ± 0.0076 | 1810 | 5:29:30 | 00:07:32 |
| PredRNN [25] | 2.0237 ± 0.0067 | 1.4311 ± 0.0149 | 7415 | 11:45:48 | 00:17:03 |
| MIM [26] | 2.0287 ± 0.0361 | 1.4330 ± 0.0250 | 10673 | 19:19:00 | 00:31:19 |
| STConvS2S-C (ours) | 1.8739 ± 0.0107 | 1.2946 ± 0.0061 | 1457 | **03:12:24** | 00:05:17 |
| STConvS2S-R (ours) | **1.8051** ± 0.0040 | **1.2404** ± 0.0068 | **1312** | 03:15:42 | **00:05:03** |

[1] R. Castro, Y.M. Souto, E. Ogasawara, F. Porto, and E. Bezerra, 2020, STConvS2S: Spatiotemporal Convolutional Sequence to Sequence Network for Weather Forecasting, Neurocomputing, v. 426, p. 285–298.

# Coleta e Análise de Dados

## Perguntas Essenciais Antes da Análise

Antes de iniciar qualquer análise, é fundamental compreender profundamente a natureza e origem dos dados. Questões críticas devem ser respondidas para garantir a validade dos resultados.

### Quais são as Fontes dos Dados?

Identificar de onde os dados foram coletados, quando foram obtidos e se representam adequadamente o problema estudado.

### Foram Aplicadas Técnicas de Pré-processamento?

Documentar todas as transformações realizadas nos dados brutos, incluindo limpeza, normalização e tratamento de valores ausentes.

### Os Dados Contêm Viés ou Falhas?

Avaliar se há desbalanceamento, sub-representação de grupos ou problemas sistemáticos que possam comprometer a análise.

**Exemplo Prático:** Em detecção de fraudes bancárias, um dataset enviesado que contém apenas fraudes de cartão de crédito pode não representar adequadamente todos os tipos de fraude, como transferências bancárias fraudulentas ou golpes por aplicativo.

## Data Papers

A criação de datasets de qualidade é um esforço significativo em diversas áreas da computação e engenharia. Data papers são publicações científicas focadas exclusivamente na documentação detalhada de datasets, descrevendo metodologia de coleta, características e aplicações potenciais.

Eles possibilitam reprodutibilidade de pesquisas e incentivam novas investigações ao disponibilizar dados bem documentados para a comunidade científica.

**Exemplo:** O dataset *Integrated Dataset of Brazilian Flights* permitiu estudos sobre padrões de atraso, eficiência operacional e otimização de rotas no tráfego aéreo brasileiro.

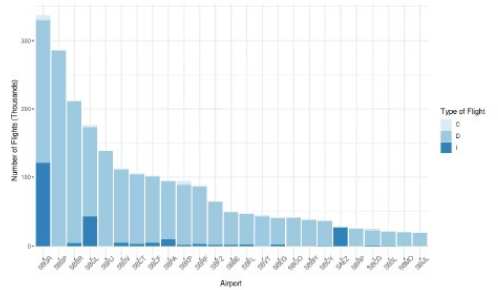



**Figure 2. Number of flights per airport, for the top-25 most active airports**

# Métodos de Análise Qualitativa e Quantitativa

A escolha entre análise qualitativa e quantitativa, ou a combinação de ambas, depende dos objetivos da pesquisa e da natureza dos dados coletados. Cada abordagem oferece perspectivas únicas e complementares sobre o fenômeno estudado.

## Análise Quantitativa

Baseada em métricas objetivas, mensuráveis e análise estatística rigorosa. Permite comparações precisas e generalizações.

**Características:**

- Uso de métricas numéricas (precisão, recall, F1-score)
- Testes de hipóteses estatísticas
- Representação gráfica de tendências
- Comparação objetiva entre métodos

**Exemplo:** Comparação de precisão, tempo de execução e uso de memória entre dois algoritmos de classificação em um mesmo dataset.

## Análise Qualitativa

Foca na interpretação subjetiva e compreensão profunda de experiências, percepções e contextos através de dados textuais e observacionais.

**Características:**

- Análise de entrevistas e questionários abertos
- Identificação de padrões temáticos
- Interpretação contextualizada
- Compreensão de motivações e percepções

**Exemplo:** Pesquisa sobre percepção de usuários sobre uma interface, capturando frustrações, preferências e sugestões detalhadas através de entrevistas.

# Interpretação de Resultados

A análise de resultados experimentais raramente oferece uma única interpretação clara e definitiva. Pesquisadores devem considerar múltiplas explicações possíveis para os padrões observados nos dados, aplicando pensamento crítico para distinguir entre causas verdadeiras e correlações espúrias.

## Abordagem Sistemática para Interpretação

### Observar Padrões

Identificar tendências e anomalias nos resultados experimentais

### Gerar Hipóteses

Propor múltiplas explicações possíveis para os padrões observados

### Testar Alternativas

Realizar testes adicionais para eliminar interpretações equivocadas

### Validar Conclusões

Confirmar a explicação mais plausível com evidências adicionais

**Exemplo Prático:** Um modelo de aprendizado de máquina que falha ao superar o baseline pode indicar várias coisas: a abordagem proposta não é adequada para o problema, a implementação não foi otimizada corretamente, o dataset é inadequado, ou os hiperparâmetros não foram ajustados apropriadamente. Testes adicionais sistemáticos são necessários para identificar a causa raiz.

# Erro Experimental e Controle de Variáveis

Compreender e minimizar erros experimentais é fundamental para obter resultados confiáveis e reprodutíveis. Erros podem surgir de diversas fontes e afetar significativamente as conclusões da pesquisa.

## O Que é Erro Experimental?

Diferenças nos resultados devido a flutuações naturais ou sistemáticas no processo experimental. Pode ser reduzido através de múltiplas replicações e planejamento estatístico cuidadoso.

### Erro Aleatório

Pequenas variações nos resultados entre execuções independentes, causadas por ruído, incertezas naturais ou fatores imprevisíveis. São inevitáveis, mas podem ser quantificados estatisticamente.

**Exemplo:** Variações mínimas no tempo de execução de um algoritmo devido a processos do sistema operacional.

### Erro Sistemático

Desvio constante e previsível causado por problemas no projeto experimental, viés na coleta de dados ou calibração inadequada de instrumentos.

**Exemplo:** Treinar um modelo de reconhecimento facial apenas com imagens de uma etnia leva a erro sistemático, resultando em desempenho ruim em outras populações.

## Como Reduzir o Erro Experimental?

1

### Replicação Interna

Realizar múltiplas execuções do experimento sob exatamente as mesmas condições controladas para quantificar variabilidade.

2

### Replicação Externa

Permitir que outros pesquisadores repitam o experimento independentemente para verificar se os achados se mantêm.

3

### Controle de Variáveis

Garantir que fatores externos (temperatura, versões de software, configurações de hardware) permaneçam constantes ou sejam monitorados.

# Projetando um Experimento Científico

O design experimental adequado é a base para obter resultados científicos válidos e confiáveis. Um experimento bem planejado maximiza a informação obtida enquanto minimiza recursos e tempo necessários.

## Passos Essenciais

1

### Definir as Variáveis de Resposta

Identificar quais métricas serão medidas para avaliar o sucesso do experimento

2

### Escolher os Fatores e Parâmetros

Determinar quais variáveis serão manipuladas e quais permanecerão fixas

3

### Selecionar o Tipo de Experimento

Decidir entre experimentos fatoriais, sequenciais ou outros designs apropriados

4

### Determinar o Número de Replicações

Calcular quantas repetições são necessárias para obter significância estatística

5

### Selecionar Amostras Representativas

Garantir que os dados de teste representam adequadamente o problema real

### Previsibilidade

Os experimentos devem demonstrar tendências gerais que se aplicam além dos datasets específicos testados, não apenas funcionar em casos particulares.

### Robustez

Os resultados devem ser consistentes, não ambíguos e resistentes a pequenas variações nas condições experimentais.

## Variável de Resposta, Parâmetros e Fatores

Compreender a distinção entre variáveis de resposta, parâmetros e fatores é essencial para projetar experimentos eficazes. Cada elemento desempenha um papel específico na estrutura experimental e na interpretação dos resultados.

### Variável de Resposta

A saída ou resultado medido do experimento, também conhecida como variável dependente. É o que você está tentando otimizar, prever ou compreender.

**Características:**

- Medida quantitativa ou qualitativa
- Depende dos fatores experimentais
- Alvo principal da análise

**Exemplo:** A precisão (accuracy) de um classificador de imagens, medida como porcentagem de predições corretas em um conjunto de teste.

### Parâmetros

Características ou configurações que permanecem fixas durante todo o experimento. Definem o contexto em que o experimento ocorre.

**Características:**

- Mantidos constantes
- Documentados para reprodutibilidade
- Podem variar entre experimentos diferentes

**Exemplo:** O número de camadas em uma rede neural, o tamanho do batch de treinamento, ou a arquitetura base do modelo.

### Fatores

Variáveis independentes que são deliberadamente modificadas para observar seu efeito sobre a variável de resposta.

**Características:**

- Controlados pelo experimentador
- Testados em diferentes níveis
- Foco principal da investigação

**Exemplo:** Diferentes técnicas de pré-processamento de dados (normalização, augmentation, filtragem) testadas sistematicamente.

# Análise de Experimentos Automatizados

Experimentos automatizados, como simulações computacionais e testes de algoritmos, apresentam desafios específicos de análise. Questões metodológicas devem ser cuidadosamente consideradas para garantir inferências estatísticas válidas.

## Perguntas Essenciais para Análise

### Design Adequado

O experimento foi projetado de forma que os dados gerados possam ser analisados corretamente com métodos estatísticos apropriados?

### Tamanho Amostral

Quantas execuções independentes são necessárias para uma inferência estatística confiável considerando a variabilidade observada?

### Tratamento de Outliers

Devo descartar valores extremos na saída do experimento ou eles representam comportamentos legítimos do sistema que devem ser analisados?

A análise inadequada de experimentos automatizados pode levar a conclusões errôneas. É fundamental aplicar princípios estatísticos rigorosos, mesmo quando lidamos com sistemas puramente computacionais.

*Alexopoulos, C., (2007), "Statistical Analysis of Simulation Output: State of the Art". In: 2007 Winter Simulation Conference, p. 150–161*

Law, A. M., (2007), "Statistical Analysis of Simulation Output Data: The Practical State of the Art". In: 2007 Winter Simulation Conference, p. 77–83

# Período de Aquecimento ("Warm-up Period")

Em simulações e experimentos computacionais, os primeiros instantes de execução frequentemente geram dados que não representam o comportamento estável do sistema. Identificar e tratar adequadamente este período de aquecimento é crucial para análises precisas.

## O Que é o Período de Aquecimento?

Fase inicial da simulação onde o sistema transita de um estado inicial artificial para um estado operacional representativo. Dados coletados nesta fase podem distorcer resultados se não forem tratados adequadamente.

## Como Determinar se Há um Período de Aquecimento?

- ### Visualização dos Dados

  Plotar as métricas de interesse ao longo do tempo e verificar se há estabilização após um período inicial de transição.

- ### Teste de Convergência Estatística

  Aplicar testes como Chow Test ou CUSUM Test para detectar mudanças estatisticamente significativas na distribuição dos dados ao longo do tempo.

- ### Comparação de Múltiplos Experimentos

  Executar múltiplas replicações independentes e verificar se os primeiros dados são sistematicamente diferentes dos posteriores.

**Exemplo Prático:** Em uma simulação de filas de atendimento, os primeiros minutos podem apresentar tempos de espera artificialmente reduzidos porque poucas pessoas chegaram ao sistema inicialmente. Esse efeito transitório desaparece com o tempo e os tempos médios se estabilizam em valores representativos da operação real.

*Alexopoulos, C., (2007), "Statistical Analysis of Simulation Output: State of the Art". In: 2007 Winter Simulation Conference, p. 150–161*

Law, A. M., (2007), "Statistical Analysis of Simulation Output Data: The Practical State of the Art". In: 2007 Winter Simulation Conference, p. 77–83

## Validade Interna e Externa dos Experimentos

A validade de um experimento determina o quão confiáveis e generalizáveis são suas conclusões. Dois tipos fundamentais de validade devem ser considerados: interna e externa. Ambos são essenciais para pesquisa científica rigorosa.

### Validade Interna

Garante que a relação causal observada entre variáveis independentes e dependentes dentro do experimento é real e não resultado de fatores confundidores ou artefatos experimentais.

**Como Garantir:**

- Controlar rigorosamente variáveis externas
- Aleatorizar a ordem de execução de testes
- Eliminar vieses de seleção e medição
- Documentar todas as condições experimentais

**Exemplo:** Se um novo algoritmo tem melhor desempenho, é porque realmente é superior ou porque foi testado em condições mais favoráveis?

### Validade Externa

Indica se os resultados e conclusões do experimento podem ser generalizados para outros contextos, populações, ambientes ou períodos de tempo além das condições específicas testadas.

**Como Garantir:**

- Usar amostras representativas e diversificadas
- Testar em múltiplos cenários e condições
- Validar com dados do mundo real
- Documentar limitações e contexto de aplicação

**Exemplo:** Um modelo treinado apenas com dados de hospitais urbanos pode não funcionar bem em contextos rurais.

*[1] C. Stoddart, 2016, Is There a Reproducibility Crisis in Science?, Nature (May.)*

[2] M. Baker, 2016, 1,500 Scientists Lift the Lid on Reproducibility, Nature, v. 533, n. 7604 (May), p. 452–454.

# Tamanho da Amostra e Poder Estatístico

## Por Que o Tamanho da Amostra é Importante?

O tamanho amostral adequado é crucial para detectar efeitos reais e evitar conclusões enganosas. Amostras muito pequenas levam a alta variabilidade e baixa confiança nos resultados, enquanto amostras excessivamente grandes desperdiçam recursos.

### Consequências de Amostras Inadequadas

- **Amostra pequena:** Pode não detectar diferenças reais (erro tipo II)
- **Variabilidade alta:** Resultados inconsistentes entre execuções
- **Baixo poder estatístico:** Incapacidade de rejeitar hipóteses falsas
- **Conclusões não generalizáveis:** Resultados específicos da amostra testada

## Como Determinar o Tamanho Adequado da Amostra?

01

### Usar Cálculo de Poder Estatístico

Determinar quantas observações são necessárias para detectar um efeito de tamanho específico com confiança desejada (tipicamente 80% de poder).

02

### Verificar Nível de Confiança

Estabelecer margem de erro aceitável e nível de significância ($\alpha = 0.05$ é padrão).

03

### Aplicar Validação Cruzada

Usar técnicas como k-fold cross-validation para garantir robustez dos resultados com a amostra disponível.

**Exemplo Prático:** Comparar dois algoritmos de classificação treinados com apenas 100 exemplos pode levar a conclusões errôneas devido à alta variabilidade. Com 10.000 exemplos bem distribuídos, a análise estatística terá muito mais confiança e os resultados serão mais generalizáveis.

## Comparação Correta de Métodos e Erros Comuns

Avaliar e comparar métodos de forma justa e rigorosa é fundamental para o avanço científico. Erros metodológicos na comparação podem levar a conclusões incorretas e prejudicar a pesquisa.

### 🚫 Erros Comuns ao Avaliar Modelos

- **Comparar métodos sem usar o mesmo conjunto de dados** - Invalida qualquer comparação direta de desempenho
- **Não considerar variabilidade dos resultados** - Ignorar que múltiplas execuções geram resultados diferentes
- **Ignorar métricas estatísticas relevantes** - Não reportar média, desvio padrão ou intervalos de confiança
- **Testar em amostras muito pequenas** - Usar dados insuficientes que não representam o problema real
- **Fairness: não avaliar viés** - Não verificar se o modelo discrimina grupos específicos injustamente

### ✅ Boas Práticas

#### Consistência nos Dados

Usar exatamente o mesmo conjunto de dados e métricas para todos os métodos comparados

#### Mesmas Condições

Testar sob condições idênticas de hardware, software e configurações ambientais

#### Validação Cruzada

Aplicar k-fold cross-validation ou holdout para evitar overfitting e garantir generalizaçã

#### Reportar Incerteza

Incluir intervalos de confiança (IC95%) para indicar precisão e incerteza nas estimativas

**Exemplo de Comparação Injusta:** Comparar um modelo treinado com 1 milhão de amostras rotuladas contra outro treinado com apenas 10.000 amostras não é uma comparação válida - a diferença pode ser devida aos dados, não à qualidade do método.

## Testes Estatísticos e Interpretação de Resultados

Escolher e interpretar testes estatísticos corretamente é essencial para validar cientificamente as diferenças observadas entre métodos. Cada teste tem pressupostos e aplicações específicas que devem ser compreendidos.

### Quando Usar Cada Teste?

#### Teste t Pareado

Usado para comparar dois métodos testados no mesmo conjunto de dados, assumindo distribuição normal dos dados.

**Quando usar:** Comparar duas versões de um algoritmo no mesmo dataset

#### Teste de Wilcoxon

Alternativa não-paramétrica ao teste t, adequado para amostras pequenas ou dados que não seguem distribuição normal.

**Quando usar:** Dados ordinais ou quando há poucos pontos de dados disponíveis

#### ANOVA

Usado para comparar simultaneamente três ou mais métodos, verificando se pelo menos um difere significativamente dos demais.

**Quando usar:** Avaliar múltiplos algoritmos ao mesmo tempo

### Como Interpretar os Testes Estatísticos Corretamente?

#### Intervalo de Confiança (IC95%)

Mostra a faixa onde os valores reais da população provavelmente se encontram. IC mais estreito indica maior precisão na estimativa.

#### P-value < 0.05

Indica evidência estatística forte de que um método é genuinamente superior ao outro, não por acaso. Menor p-value = evidência mais forte.

#### Variabilidade dos Resultados

Usar desvio padrão, boxplots e distribuições completas para visualizar a dispersão e identificar outliers ou padrões anômalos.

**Exemplo Prático:** Dois algoritmos de IA têm acurácia média de 92% e 94%. A diferença de 2% parece pequena, mas é estatisticamente significativa? Um teste t pareado com IC95% e p-value podem responder definitivamente. Se $p < 0.05$, podemos afirmar com 95% de confiança que o segundo método é realmente superior.

# Como Apresentar Resultados Experimentais?

A apresentação clara e honesta de resultados experimentais é fundamental para comunicação científica efetiva. Visualizações bem projetadas facilitam a compreensão e interpretação correta dos dados.

## Inclua Sempre

### Objetivo do Experimento

Contextualize o que está sendo testado e por quê

### Dados Utilizados

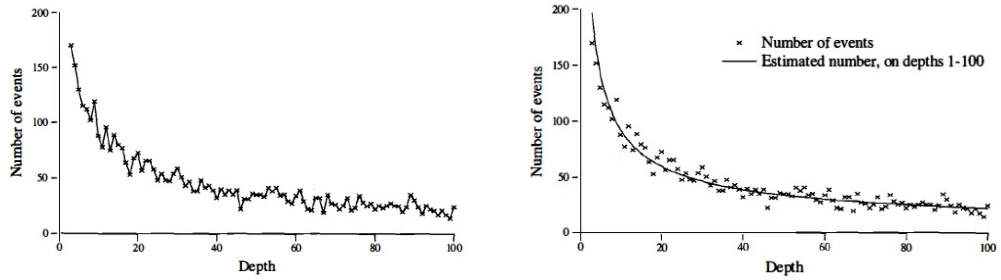
Descreva o dataset, tamanho, características relevantes

### Resultados Principais

Apresente métricas, comparações e significância estatística

## ⚠️ Erro Comum na Visualização

Mostrar uma linha interpolando pontos sem significado estatístico cria a falsa impressão de tendências suaves e contínuas quando os dados reais são discretos e variáveis.




**Problema:** A linha sugere que há valores intermediários válidos entre os pontos medidos, o que pode não ser verdade. Use pontos discretos ou inclua barras de erro para representar variabilidade.

## Melhores Práticas de Visualização

- Use gráficos de barras com barras de erro (desvio padrão ou IC95%)
- Boxplots para mostrar distribuição completa dos dados
- Tabelas para valores exatos quando precisão é importante
- Legendas claras e eixos bem rotulados

## Reprodutibilidade e Melhores Práticas

A reprodutibilidade é um pilar fundamental da ciência moderna. Pesquisas que não podem ser reproduzidas têm valor científico questionável e prejudicam o progresso coletivo do conhecimento.

### ✅ O Que Melhora a Reprodutibilidade?

- **Compartilhamento de Código e Dados**

  Disponibilizar código-fonte completo e datasets para facilitar replicação exata por outros pesquisadores

- **Notebooks Interativos**

  Usar Jupyter, RMarkdown ou ferramentas similares para documentar análises de forma executável e transparente

- **Ambientes Reprodutíveis**

  Docker, Conda ou MLflow para capturar exatamente versões de bibliotecas e dependências

- **Controle de Versão**

  GitHub, GitLab para rastrear todas as mudanças no código e documentação ao longo do tempo

- **Automatização de Experimentos**

  Scripts automatizados para minimizar variação introduzida por execução manual

### ❌ Erros Que Comprometem a Reprodutibilidade

**Dependência de Hardware Específico**

Configurações que funcionam apenas em máquinas ou GPUs particulares não são replicáveis por outros

**Falta de Documentação de Versões**

Não especificar versões exatas de Python, bibliotecas e frameworks usados

**Modificações Sem Controle**

Alterar dados, código ou parâmetros sem registrar mudanças sistematicamente

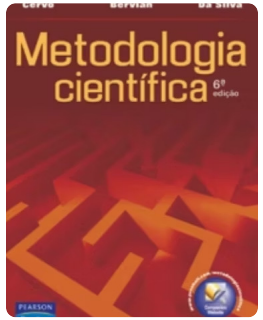
**Exemplo de Problema de Reprodutibilidade:** Um estudo desenvolvido com Python 3.8 e TensorFlow 2.4 pode gerar resultados significativamente diferentes quando testado com Python 3.11 e TensorFlow 2.10, devido a mudanças em implementações de algoritmos ou comportamentos padrão. Sem controle rigoroso de versões, os resultados originais podem ser impossíveis de replicar.

## Referências Bibliográficas

Esta apresentação foi desenvolvida com base em obras fundamentais sobre metodologia científica e escrita acadêmica, essenciais para o desenvolvimento de competências em pesquisa e análise de artigos científicos. Estas referências representam contribuições seminais que orientam pesquisadores em todas as etapas do processo investigativo, desde a concepção do problema até a comunicação efetiva dos resultados.

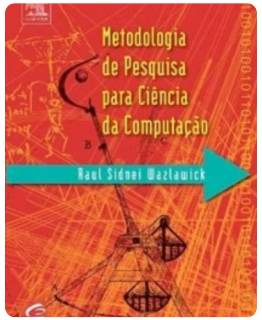

Perovano (2016)

**Manual de metodologia da pesquisa científica** - Editora Intersaberes. Obra completa e abrangente sobre fundamentos metodológicos, oferecendo uma visão integrada dos principais métodos e técnicas de pesquisa científica.



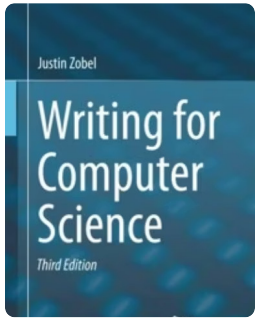

Cervo, Bervian & Silva (2006)

**Metodologia Científica** - Pearson Universidades. Referência clássica consolidada em metodologia de pesquisa, amplamente utilizada na formação acadêmica por sua clareza didática e rigor conceitual.




Wazlawick (2017)

**Metodologia de Pesquisa para Ciência da Computação** - Elsevier Brasil. Abordagem especializada e direcionada para a área de computação, contemplando as particularidades metodológicas deste campo do conhecimento.




Zobel (2015)

**Writing for Computer Science** - Springer. Guia essencial e prático para escrita científica em computação, abordando desde a estruturação de artigos até técnicas avançadas de comunicação acadêmica.

Estas obras constituem um acervo bibliográfico robusto que fornece fundamentos teóricos e práticos indispensáveis para a condução de pesquisas científicas de qualidade, auxiliando na compreensão profunda dos processos de investigação e na produção de conhecimento válido e relevante.