



Comparação de Métodos Computacionais

Com prática em R

Uma abordagem estatística rigorosa para avaliar e contrastar algoritmos e técnicas analíticas, com exemplos práticos implementados em R para pesquisadores e analistas de dados.

Eduardo Ogasawara
eduardo.ogasawara@cefet-rj.br
<https://eic.cefet-rj.br/~eogasawara>



Por que comparar métodos computacionais?

A comparação científica de métodos computacionais vai muito além de simplesmente observar qual algoritmo obteve melhor acurácia ou menor erro em um único experimento. Métricas isoladas podem ser profundamente enganosas, especialmente quando não consideramos a variabilidade inerente aos dados e ao processo de amostragem.

Diferenças aparentemente impressionantes entre dois métodos podem ser apenas resultado de flutuações aleatórias, sem significância estatística real. Por isso, comparações rigorosas exigem testes estatísticos apropriados que quantifiquem a incerteza e forneçam evidências sólidas para nossas conclusões.

Métricas enganosas

Acurácia isolada não revela toda a história dos dados

Variabilidade amostral

Diferenças podem ser apenas acaso estatístico

Rigor estatístico

Testes apropriados quantificam evidências

Reprodutibilidade

Documentação formal garante replicabilidade

A reprodutibilidade e a justificativa formal dos nossos achados são fundamentais para a ciência de dados moderna, garantindo que outros pesquisadores possam validar e estender nosso trabalho.

O dilema: paramétrico ou não-paramétrico?

A escolha entre métodos paramétricos e não-paramétricos representa uma das decisões mais importantes na análise estatística de comparações. Esta escolha não deve ser arbitrária, mas fundamentada nas características dos dados e na validade das suposições subjacentes a cada abordagem.



Avaliação de suposições

Verificar normalidade, homocedasticidade e independência dos dados



Poder estatístico

Paramétricos oferecem maior poder quando suposições são válidas



Robustez

Não-paramétricos protegem contra violações de suposições



Decisão fundamentada

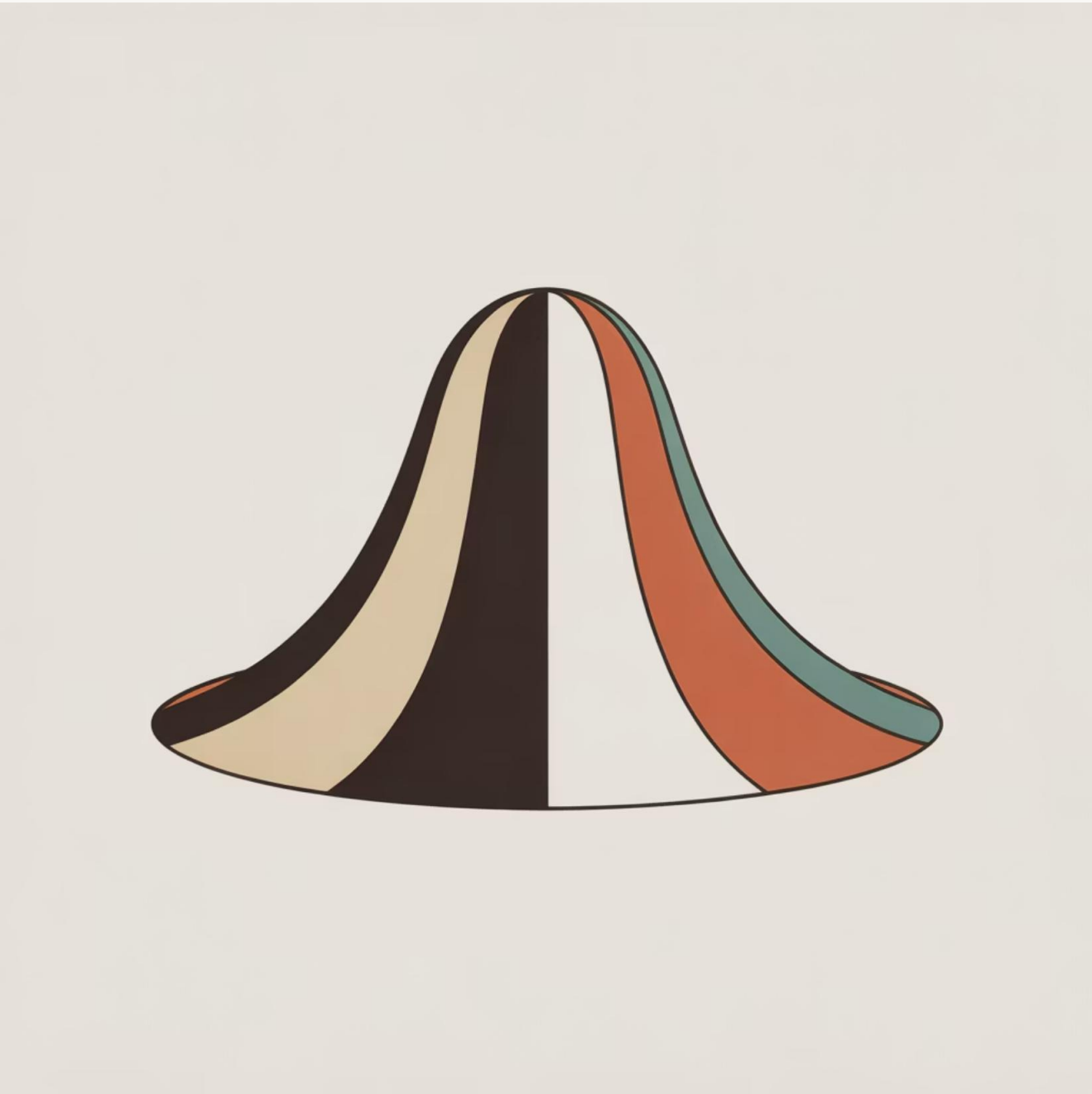
Escolha baseada em diagnósticos formais, não intuição

Métodos paramétricos maximizam o poder estatístico quando suas suposições são atendidas, detectando diferenças reais com maior sensibilidade. No entanto, quando essas suposições falham—como na presença de distribuições fortemente assimétricas ou outliers extremos—os métodos não-paramétricos oferecem robustez crucial. A verificação formal de normalidade e do pareamento dos dados é um passo essencial antes de qualquer análise comparativa.

Paramétrico: conceito e hipóteses

Métodos paramétricos fundamentam-se em suposições explícitas sobre a forma da distribuição populacional dos dados, tipicamente assumindo distribuição normal (gaussiana). Esses métodos trabalham diretamente com parâmetros populacionais, como média e variância, estimando-os a partir da amostra.

Os testes paramétricos clássicos incluem o teste t de Student (para comparar duas médias), ANOVA (para três ou mais grupos) e modelos de regressão. A validade das conclusões desses testes depende criticamente do atendimento de suas suposições.



Não-paramétrico: conceito e hipóteses

Métodos não-paramétricos, também conhecidos como testes livres de distribuição, não exigem suposições sobre a forma específica da distribuição populacional. Em vez de trabalhar com valores absolutos, esses testes baseiam-se em ranks (postos) ou medianas, tornando-os naturalmente robustos à presença de outliers e distribuições assimétricas.

Baseado em ranks

Trabalha com ordenação dos dados, não valores absolutos

Mediana como parâmetro

Foco em medianas e quantis, não médias

Robustez a outliers

Valores extremos têm impacto minimizado

Principais testes não-paramétricos


- Wilcoxon rank-sum: alternativa ao t-test pareado
- Mann-Whitney U: comparação de dois grupos independentes
- Kruskal-Wallis: extensão não-paramétrica da ANOVA
- Friedman: análise de blocos repetidos

Quando preferir não-paramétricos

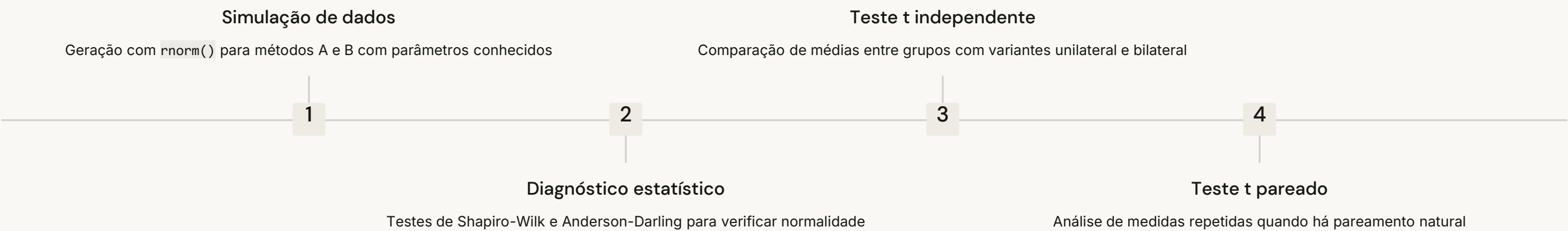
- Distribuições fortemente assimétricas
- Presença de caudas pesadas
- Outliers extremos que não podem ser removidos
- Amostras pequenas sem garantia de normalidade
- Dados ordinais ou em escalas não-intervalares

Código prático (Paramétrico)

Tutorial completo: tut_parametric.md

 Repositório GitHub: https://github.com/eogasawara/mylibrary/blob/main/tutorial/tut_parametric.md

O tutorial de testes paramétricos implementa um fluxo completo de análise comparativa usando R, desde a simulação de dados até a interpretação dos resultados. O código demonstra boas práticas de reprodutibilidade e documentação.



Principais funções R utilizadas

```
# Normalidadeshapiro.test()ad.test()#
Comparaçãot.test(paired=FALSE)t.test(paired=TRUE)#
Visualizaçãqqnorm()qqline()
```

O tutorial enfatiza a importância de visualizar os dados antes dos testes formais, usando gráficos Q-Q para avaliar normalidade visualmente. A escolha entre teste unilateral (one-tailed) e bilateral (two-tailed) deve ser feita a priori, baseada na hipótese de pesquisa.

Para dados pareados, onde cada observação no grupo A tem uma correspondência natural no grupo B (por exemplo, medições antes e depois em um mesmo sujeito), o teste t pareado é mais poderoso pois controla a variabilidade individual.

Código prático (Não-paramétrico)

Tutorial completo: tut_nonparametric.md

Repositório GitHub: https://github.com/eogasawara/mylibrary/blob/main/tutorial/tut_nonparametric.md

Este tutorial complementa o anterior, demonstrando como proceder quando as suposições paramétricas são violadas. O código ilustra a detecção formal de violações de normalidade e a aplicação apropriada de alternativas não-paramétricas.



Detecção de violações

Testes formais evidenciam desvios significativos da normalidade ($p < 0.05$)



Interpretação por ranks

Análise baseada em medianas e soma de postos, não médias



Escolha do teste adequado

Wilcoxon rank-sum para pareados; Mann-Whitney U para independentes



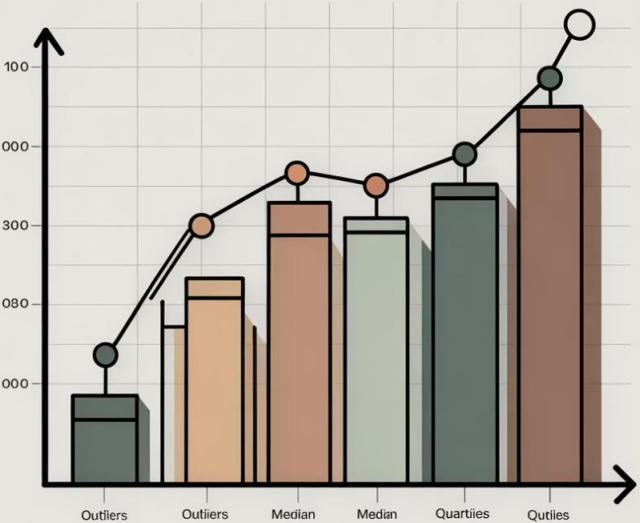
Considerações práticas

Sensibilidade ao tamanho amostral e robustez natural a outliers

Funções R para testes não-paramétricos

```
# Wilcoxon (pareado)wilcox.test(paired=TRUE)# Mann-Whitney (independente)wilcox.test(paired=FALSE)# Kruskal-Wallis (3+ grupos)kruskal.test()#
Visualizaçãoboxplot()stripchart()
```

O tutorial demonstra como interpretar corretamente os resultados: o valor-p indica se há evidência de diferença entre as distribuições, mas a interpretação deve focar em medianas, não médias.



Testes não-paramétricos são especialmente úteis com amostras pequenas onde não podemos garantir normalidade pelo Teorema do Limite Central. A robustez a outliers é vantajosa em dados reais, que frequentemente contêm valores extremos legítimos.

Comparando conclusões entre os testes

Uma estratégia analítica robusta envolve aplicar tanto métodos paramétricos quanto não-paramétricos quando viável, comparando suas conclusões. A concordância entre as abordagens fortalece significativamente nossa confiança nos resultados, enquanto divergências sinalizam a necessidade de investigação mais profunda.

Aplicação dual

Execute ambos os testes quando possível para triangular evidências

- Valide suposições primeiro
- Compare valores-p obtidos
- Avalie convergência das conclusões

Concordância

Resultados consistentes reforçam a decisão estatística

- Maior confiança nas conclusões
- Robustez a violações leves
- Evidência convergente

Divergência

Resultados contraditórios exigem revisão cuidadosa

- Rever suposições detalhadamente
- Investigar outliers influentes
- Considerar transformações de dados

Elementos essenciais no relatório de comparações

Desenho experimental

Documente claramente se os grupos são independentes ou pareados, o tamanho amostral de cada grupo, e o procedimento de coleta de dados.

Direção da hipótese

Especifique se o teste é unilateral ou bilateral, justificando a escolha com base na teoria ou em estudos prévios.

Intervalos de confiança

Reporte intervalos de confiança para as diferenças, não apenas valores-p, fornecendo contexto prático.

Medidas de efeito

Tutorial completo: `tut_effect_size.md`

📄 Repositório GitHub: https://github.com/eogasawara/mylibrary/blob/main/tutorial/tut_effect_size.md

$p < 0.05 \neq$ relevância prática

Um dos equívocos mais comuns em análise estatística é confundir significância estatística com relevância prática. Um valor- p pequeno indica apenas que é improvável que a diferença observada seja resultado do acaso—não quantifica a magnitude ou importância dessa diferença. Tamanhos de efeito preenchem essa lacuna crítica.



d de Cohen

Medida padronizada da diferença entre duas médias, expressa em unidades de desvio-padrão. Valores: pequeno (0.2), médio (0.5), grande (0.8).



r de correlação

Proporção de variância compartilhada entre variáveis. Útil para testes não-paramétricos. Varia de 0 a 1.



η^2 (eta quadrado)

Proporção de variância explicada em ANOVA. Indica quanto da variabilidade total é atribuível ao fator estudado.

Implementação com rstatix

O pacote `rstatix` facilita o cálculo de tamanhos de efeito em R, integrando-se perfeitamente com fluxos tidyverse. O tutorial demonstra o uso de dados em formato longo (long format), considerado a melhor prática para análises estatísticas modernas.

```
library(rstatix)# Cohen's dcohens_d(data, value ~ group)# Eta quadradoanova_test(...) %>%
get_anova_table()
```

Interpretação e relatório

Sempre reporte tamanhos de efeito junto com valores- p e intervalos de confiança. Isso permite que leitores avaliem não apenas se há diferença, mas se essa diferença importa na prática.

A transparência metodológica é fundamental: documente todas as escolhas analíticas, incluindo transformações de dados, tratamento de outliers e testes de suposições. Isso garante reprodutibilidade e permite crítica construtiva.

Conclusões e próximos passos

A comparação rigorosa de métodos computacionais requer uma combinação de conhecimento estatístico, diagnóstico cuidadoso das suposições, e documentação transparente. Não existe uma abordagem única que sirva para todos os cenários—a escolha entre métodos paramétricos e não-paramétricos deve ser guiada pelas características dos dados e pela validade das suposições subjacentes.



Diagnóstico é essencial

Sempre verifique suposições antes de escolher o teste. Use testes formais e visualizações.



Poder vs. Robustez

Paramétricos para normalidade comprovada; não-paramétricos para robustez em violações.



Tamanho de efeito

Significância estatística não é suficiente. Quantifique e reporte a magnitude das diferenças.



Reprodutibilidade

Documente todo o pipeline analítico, das suposições às conclusões, garantindo replicabilidade.

Recursos adicionais e aprendizado contínuo

Aprofunde seus conhecimentos

- Explore os tutoriais completos no repositório GitHub
- Pratique com seus próprios conjuntos de dados
- Experimente com diferentes violações de suposições
- Compare resultados entre múltiplos métodos
- Participe de comunidades R (Stack Overflow, RStudio Community)

Próximos tópicos recomendados

- Correções para múltiplas comparações (Bonferroni, FDR)
- Bootstrap e métodos de reamostragem
- Análise de poder estatístico e cálculo amostral
- Testes de equivalência e não-inferioridade
- Modelos lineares mistos para dados complexos

A estatística aplicada é uma jornada contínua de aprendizado. Quanto mais você pratica a análise rigorosa de comparações, mais refinada se torna sua intuição sobre qual método aplicar em cada situação. Lembre-se sempre: a melhor análise é aquela que você pode explicar, justificar e reproduzir.