



Testes Sequenciais: Redução de Custos e Aceleração de Decisões com Confiabilidade Estatística

Murilo Couto de Oliveira,¹ Júlio Cezar Estrella²

Instituto de Ciências Matemáticas e de Computação – ICMC, Universidade de São Paulo (USP)

1 Introdução

Testes A/B são amplamente utilizados para avaliar a eficácia de mudanças em produtos digitais [6, 7]. No modelo tradicional de horizonte fixo, decisões só podem ser tomadas após atingir um tamanho amostral previamente definido, o que aumenta tempo e custo. Os Testes Sequenciais de Grupos (TSG) surgem como alternativa eficiente, permitindo análises interinas com controle rigoroso do erro tipo I por meio de funções de gasto de alfa e fatores de inflação amostral [4, 8]. Métodos como Pocock [10] e O’Brien–Fleming [9] possibilitam encerramento antecipado por eficácia, futilidade ou segurança, acelerando a tomada de decisão com validade estatística. Este projeto propõe uma biblioteca Python aberta e modular para aplicação prática de TSG, visando reduzir o tempo de experimentação e democratizar o uso de técnicas sequenciais em pipelines modernos de experimentação.

2 Fundamentos teóricos

2.1 Funções de gasto de alfa

O erro tipo I é distribuído progressivamente com a fração de informação t . As funções de Pocock e O’Brien–Fleming (OBF) são amplamente usadas: a primeira distribui alfa de modo constante; a segunda é mais conservadora nas etapas iniciais [9–11]. A abordagem de Lan–DeMets [8] oferece flexibilidade no número de análises interinas.

2.2 Fator de inflação

Para manter o poder estatístico, pequenas correções são aplicadas ao tamanho amostral, especialmente em estudos com mais de uma análise sequencial [4, 13]. A redução prática de esforço decorre da possibilidade de encerrar o teste ao observar um efeito real.

¹murilo.couto-oliveira@usp.br

²jcezar@icmc.usp.br

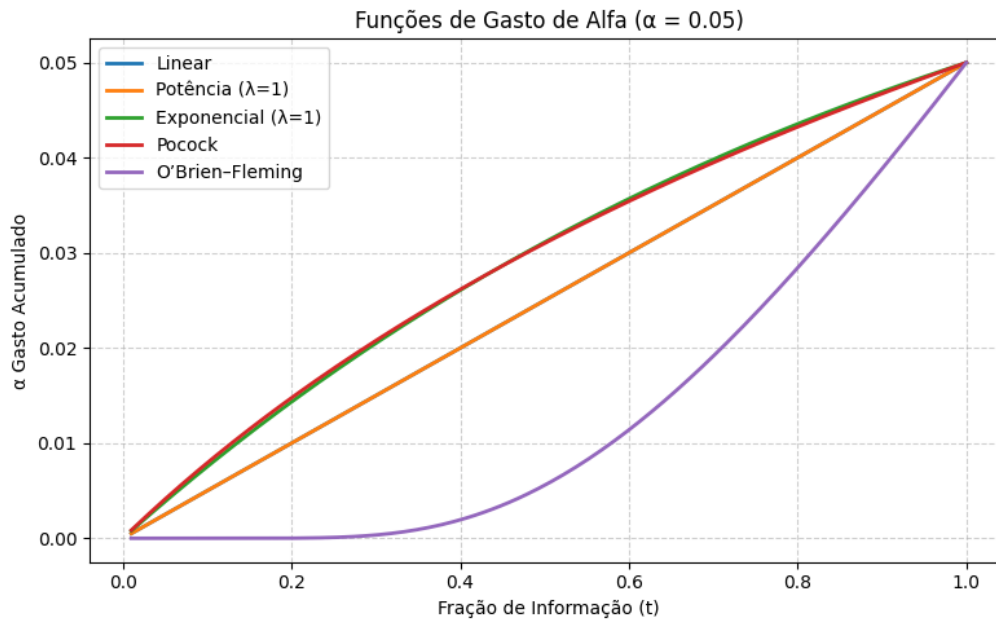


Figura 1: Funções de gasto de alfa (linear, potência, Pocock e O'Brien-Fleming).

3 Implementação e arquitetura

A biblioteca proposta é modular e orientada a objetos, composta por:

- `core`: coordena o ciclo sequencial e as decisões de parada;
- `fronteiras_decisao`: gera fronteiras segundo as funções de gasto de alfa;
- `testes_estatisticos`: realiza testes estatísticos diversificados;
- `seleciona_teste_hipotese`: seleciona o teste adequado de acordo com as características das distribuições dos grupos.

3.1 Interface de alto nível

A direção do teste é definida por rótulos intuitivos: *crescimento*, *queda* ou *mudança*, correspondendo a testes unilaterais (para direita ou esquerda) ou bilaterais. Além das métricas usuais (p -valor e limites críticos), o sistema fornece interpretação prática com as seguintes respostas: *benefício*, *prejuízo* ou *sem efeito*.

3.2 Fluxo decisório e rotina sequencial

O algoritmo verifica normalidade e homocedasticidade para selecionar testes paramétricos (Z, t ou ANOVA) ou não paramétricos (Mann-Whitney, Kruskal-Wallis, Qui-quadrado ou Teste Exato de Fisher).

A classe principal `TesteSequencialGrupo` calcula a informação acumulada t , aplica a função de gasto escolhida e ajusta os limites críticos de parada. As métricas (p -valor ajustado e α acumulado) orientam a decisão de continuar ou encerrar o experimento.

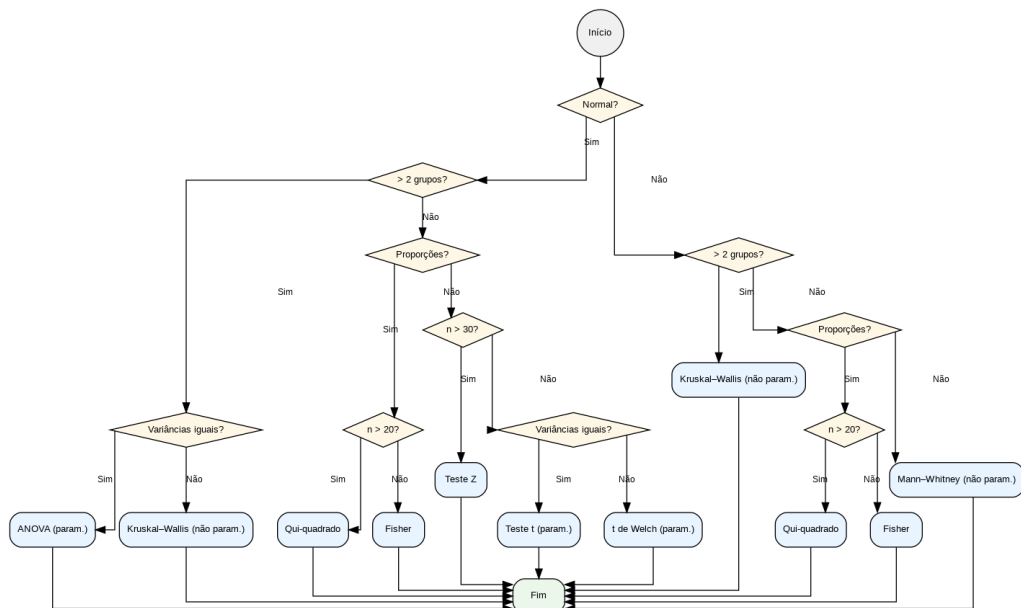


Figura 2: Fluxo de seleção de testes na biblioteca proposta.

4 Validação e resultados esperados

A eficiência será avaliada pela redução do tamanho amostral esperado e pela taxa de detecção precoce sob diferentes magnitudes de efeito. Espera-se manter o controle de erro tipo I e boa cobertura dos intervalos de confiança, reduzindo custos e tempo até a decisão [1].

5 Conclusões

Testes sequenciais com funções de gasto de alfa proporcionam ganhos significativos de eficiência em experimentos controlados e randomizados (testes A/B). A biblioteca proposta une confiabilidade estatística com simplicidade de uso, favorecendo sua adoção em pipelines de experimentação. Futuras extensões incluem modelos para desfechos não normais e abordagens inspiradas em testes aplicados para ensaios clínicos [8].

Referências

- [1] ALBERS, C. The problem with unadjusted multiple and sequential testing. *Nature Communications*, v. 10, p. 1921, 2019.
- [2] ARMITAGE, P. *Sequential Medical Trials*. Blackwell Scientific Publications, 1969.
- [3] FLEMING, T. R.; HARRINGTON, D. P. *Counting Processes and Survival Analysis*. Wiley, 1991.
- [4] JENNISON, C.; TURNBULL, B. W. *Group Sequential Methods with Applications to Clinical Trials*. Chapman & Hall/CRC, 1999.
- [5] JENNISON, C.; TURNBULL, B. W. Group sequential tests and adaptive designs. *Biometrika*, v. 87, n. 4, p. 199–214, 2000.
- [6] KOHAVI, R. et al. Controlled experiments on the web. *Data Mining and Knowledge Discovery*, v. 18, n. 1, p. 140–181, 2009.
- [7] KOHAVI, R.; LONGBOTHAM, R. Online Controlled Experiments and A/B Testing. *Encyclopedia of Machine Learning and Data Mining*, 2017.
- [8] LAN, K. K. G.; DEMETS, D. L. Discrete Sequential Boundaries for Clinical Trials. *Biometrika*, v. 70, n. 3, p. 659–663, 1983.
- [9] O'BRIEN, P. C.; FLEMING, T. R. A multiple testing procedure for clinical trials. *Biometrics*, v. 35, n. 3, p. 549–556, 1979.
- [10] POCOCK, S. J. Group sequential methods. *Biometrika*, v. 64, n. 2, p. 191–199, 1977.
- [11] PROSCHAN, M. A.; LAN, K. K. G.; WITTES, J. T. *Statistical Monitoring of Clinical Trials*. Springer, 2006.
- [12] WALD, A. *Sequential Analysis*. Wiley, 1947.
- [13] WANG, S. K.; TSIATIS, A. A. Approximately optimal one-parameter boundaries. *Biometrics*, v. 43, n. 1, p. 193–199, 1987.
- [14] WHITEHEAD, J. *The Design and Analysis of Sequential Clinical Trials*. Wiley, 1997.