

Stopping Rule e Data Peeking

Jose Storopoli

Criado em 19/04/2020 atualizado em 07/05/2020

Problemas com p -valores

Princípio 4 da *American Statistical Association* (ASA) Wasserstein and Lazar (2016)

Proper inference requires full reporting and transparency. P-values and related analyses should not be reported selectively. Conducting multiple analyses of the data and reporting only those with certain p-values (typically those passing a significance threshold) renders the reported p-values essentially uninterpretable. (pp. 131-132)

Stopping Rule e Data Peeking

Uma das principais críticas da abordagem Frequentista de Estatística é a regra de interrupção de um experimento ou teste estatístico (*stopping rule*) e “bisbilhotagem” de dados (*data peeking*)

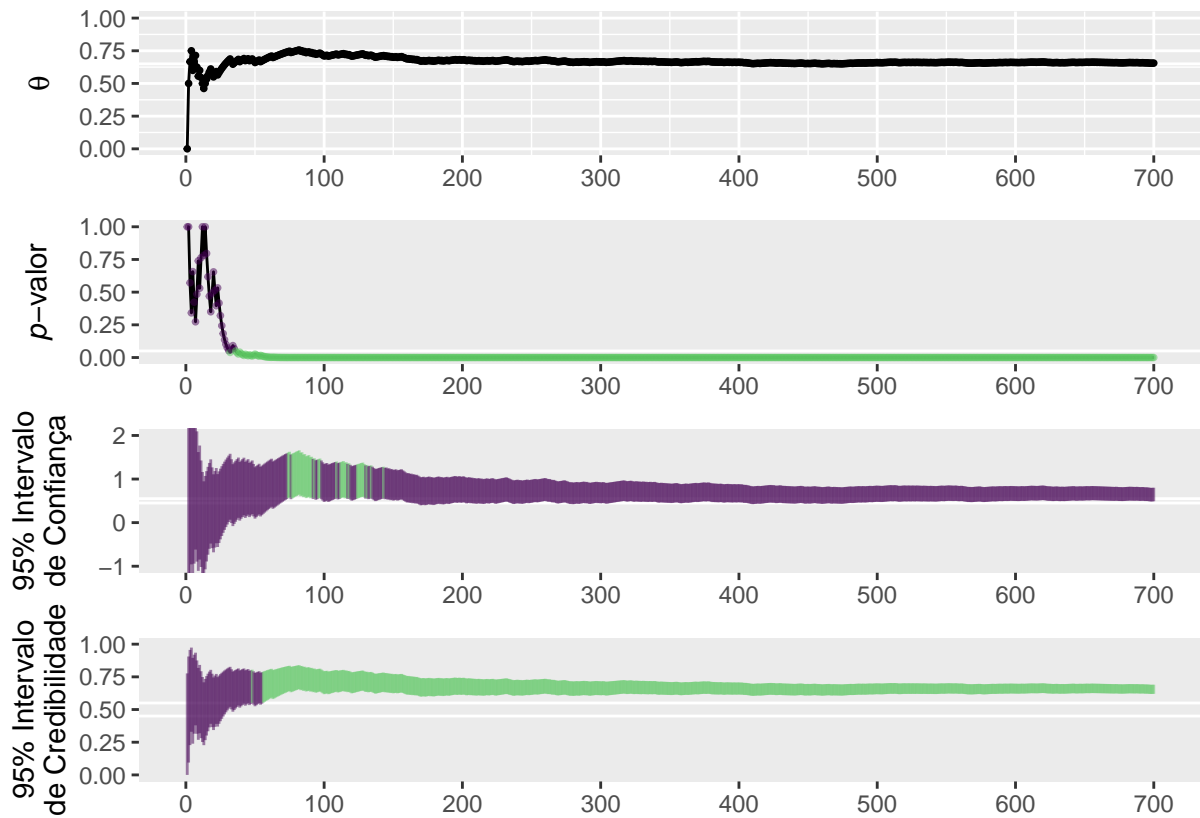
Simulação

Para mostrar o efeito da *stopping rule* e *data peeking* simularemos uma moeda **injusta** (manipulada) que gera **cara** 65% do tempo ($\theta = 0.65$). O número de amostra simuladas da moeda é $N = 700$ e para cada amostra computaremos:

- proporção cumulativa de caras vs coroas (θ)
- p -valor do teste contra a hipótese nula (H_0) de que a moeda é justa ($\theta = 0.5$)
- 95% intervalo de confiança (*confidence interval*): uma mensuração frequentista baseada em um intervalo que conterá o valor verdadeiro da proporção de caras de uma moeda (θ) em 95% das ocasiões se um estudo foi repetido várias vezes usando amostras da mesma população
- 95% intervalo de credibilidade (*credible interval*): uma mensuração bayesiana da probabilidade de 95% do valor da proporção de caras de uma moeda (θ)

A simulação foi baseada no capítulo 13 de Kruschke (2014).

Observem o p -valor e a diferença entre o intervalo de confiança (Frequentista) e o intervalo de credibilidade (Bayesiano).

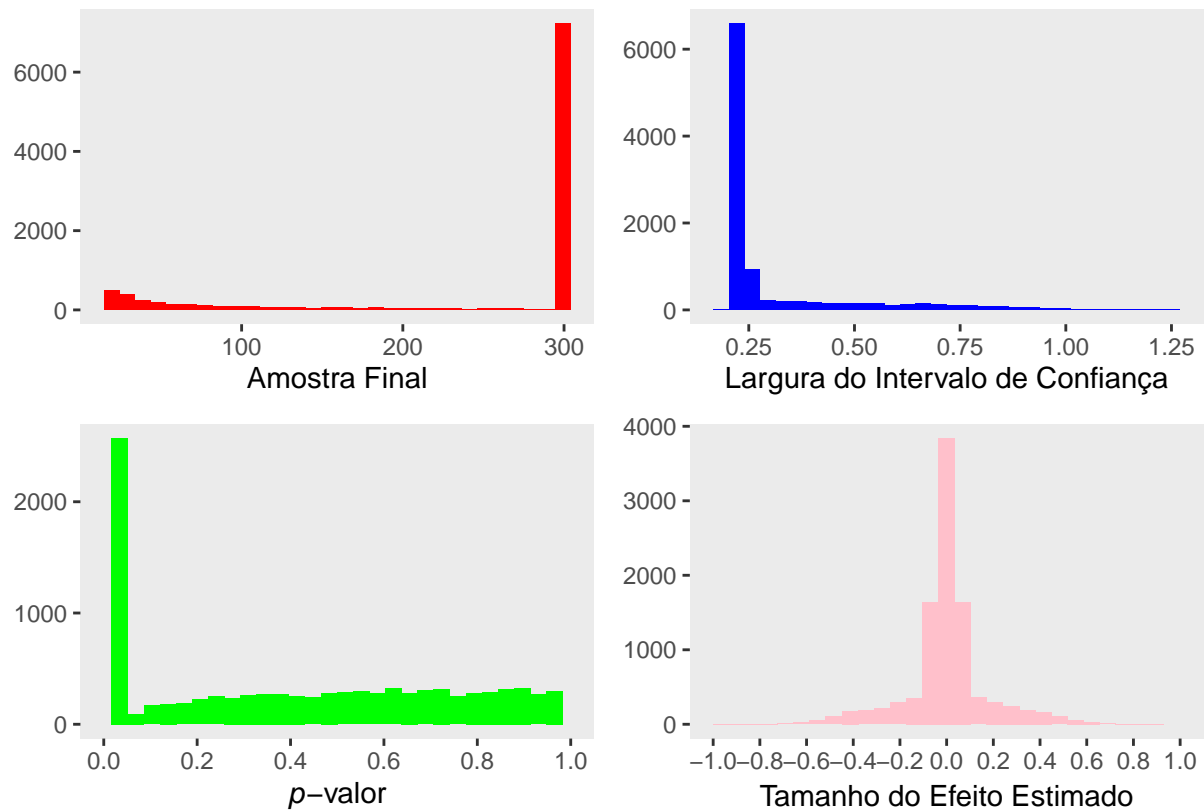


Um outro exemplo

Este exemplo é baseado em Vanhove (2017)

Imagine que você tem custeio de pesquisa para realizar um experimento com 300 participantes, mas prefere não desperdiçar dinheiro. Então, depois de 20 participantes, você executa um teste t e verifica se é significativo. Se for, você para de coletar mais dados ali. Caso contrário, execute outros cinco participantes, verifique os resultados novamente e assim por diante. Se você atingir 300 participantes sem obter um resultado significativo, poderá encerrar.

O gráfico abaixo mostra 10,000 replicações desse cenário.



Referências

- Kruschke, John. 2014. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.
- Vanhove, Jan. 2017. "Confidence Interval-Based Optional Stopping." *Confidence Interval-Based Optional Stopping*. <https://janhove.github.io/design/2017/09/19/peeking-confidence-intervals>.
- Wasserstein, Ronald L., and Nicole A. Lazar. 2016. "The ASA's Statement on p-Values: Context, Process, and Purpose." *American Statistician* 70 (2): 129–33. <https://doi.org/10.1080/00031305.2016.1154108>.