

# Prevalência de inconsistências em relatos estatísticos de periódicos brasileiros de Psicologia

Nomes dos autores: Gustavo de Paula Dídimo, Roberta Sant Anna Kafrouni

Universidade Federal do Paraná, Curitiba - PR

## INTRODUÇÃO

Métodos estatísticos são amplamente usados na Psicologia e nas Ciências Sociais, com destaque para os **Testes de Significância de Hipótese Nula** (*null hypothesis significance tests* – NHST). A versão atual deste método inferencial envolve definir duas hipóteses (nula, H<sub>0</sub>, e alternativa, H<sub>1</sub>) sobre variáveis de interesse, escolher um nível de significância  $\alpha$  (geralmente, 0,05), calcular uma estatística de teste e um p-valor (O'Connor & Khattar, 2022). Com base no p-valor calculado, toma-se a decisão de rejeitar ou não a hipótese nula sobre as variáveis.

Apesar da grande utilização, há vários problemas com o uso de NHST. Um deles é a grande quantidade de **inconsistências nos relatos estatísticos em publicações**. Em resultados de NHST, comumente relata-se **três informações: os graus de liberdade do teste, a estatística de teste e o p-valor**. Esses valores tem relação entre si, e pode-se analisá-los para avaliar se são consistentes. Nesse sentido, uma análise de 8 grandes periódicos internacionais de Psicologia, com mais de 250.000 p-valores relatados entre 1983 e 2015, indicou que: 49,6% dos artigos analisados continham pelo menos 1 inconsistência (valores diferentes calculados diferentes dos relatados) e 12,9% pelo menos uma inconsistência grave (que muda a decisão de rejeitar ou não a hipótese nula; Nuijten et al., 2016).

Esses números são alarmantes e representam um grande problema para ciência psicológica. Em conjunto com outros problemas de NHST, eles também têm relação com a recente **Crise da Replicabilidade da Psicologia** (O'Connor & Khattar, 2022; Open Science Collaboration, 2015). Apesar de já terem sido conduzidas análises dessas inconsistências na literatura internacional, **não foi realizada nenhuma na literatura nacional**. De tal modo, é essencial avaliar se esse padrão de inconsistências também ocorre nesta literatura.

## OBJETIVOS

O objetivo principal deste trabalho é **descrever a prevalência de inconsistências em relatos estatísticos em periódicos brasileiros de Psicologia de alta relevância**. Também, buscou-se avaliar a quantidade de artigos dessas revistas que apresentaram inconsistências.

## MÉTODO

Foram selecionados **3 periódicos nacionais de Qualis A1** de Psicologia: Teoria e Pesquisa, Reflexão e Crítica e Paidéia. Foi feito o *download* manual de **3506 artigos** desses periódicos disponíveis na plataforma Scielo, entre os anos de 1991 a 2023 (Paidéia, n = 916), 1999 a 2023 (Teoria e Pesquisa, n = 1333) e 2001 a 2023 (Reflexão e Crítica, n = 1257).

A extração e análise de dados foi realizada por meio de **procedimento automatizado** do pacote *statcheck* (v. 1.4.1-beta2) (Nuijten e Epskamp, 2023) para o **software R**. O pacote é capaz de extrair relatos estatísticos textuais em **normas APA** (ou similares às normas) para os seguintes tipos de teste: correlações (r), testes t, F, Z, Q e qui-quadrado ( $\chi^2$ ); no entanto, não consegue extrair resultados de tabelas, gráficos e outros formatos.

O software realiza a extração dos relatos dos artigos, re-calcula os **p-valores** e **classifica se houve inconsistências ou inconsistências graves nos relatos**. **Inconsistências** são definidas como valores relatados de p-valores inconsistentes com os calculados pelo *statcheck*; **inconsistências graves** são definidas como inconsistências que alterariam a decisão sobre rejeitar ou não a hipótese nula com base no p-valor. O cálculo leva em consideração arredondamentos possíveis de números, mas não consegue detectar correções de comparações múltiplas (e.g., Bonferroni). Por padrão, **ele considera um nível de significância de 0.05**. Os dados foram analisados com estatísticas descritivas e gráficos foram feitos com o pacote *ggplot2* para o software R.

## RESULTADOS

O *statcheck* encontrou resultados de NHST em **278 publicações** (Psicologia: Teoria e Pesquisa = 40, Paidéia = 100, Psicologia: Reflexão e Crítica = 138). Nelas, encontrou-se 1683 relatos estatísticos. 11 instâncias de erros na extração dos dados foram eliminadas do banco de dados, resultando em **1672 relatos para análise** (Psicologia: Teoria e Pesquisa = 235, Paidéia = 558, Psicologia: Reflexão e Crítica = 879).

Das 278 publicações, **112 (40,3%) continham pelo menos 1 inconsistência** e **35 (12,6%) continham pelo menos 1 inconsistência grave**. Em termos da quantidade total de erros, foram encontrados **383 inconsistências (22,9%)** e **73 (4,4%) inconsistências graves** dos 1672 p-valores. A quantidade de inconsistências extraídas ocorreu de modo desigual ao longo dos anos e periódicos, com a maioria concentrada em anos recentes (Figura 1). Em adição, das 112 publicações com inconsistências, a maioria estava presente em poucas publicações (Figura 2).

Inconsistências estatísticas relatadas em periódicos A1 de Psicologia  
Por ano de publicação e por periódico, analisadas pelo *statcheck*

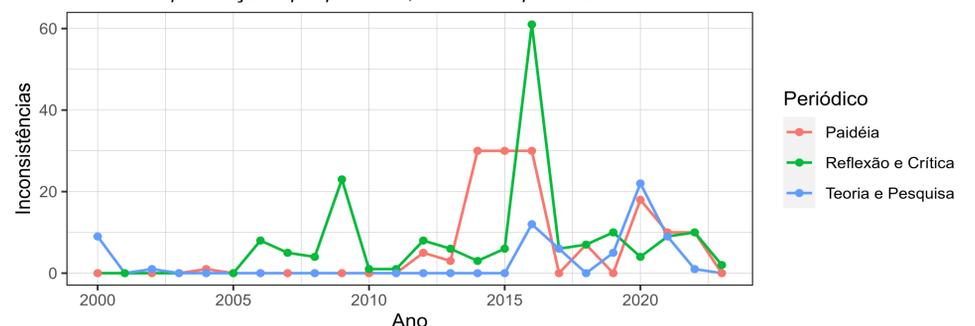


Figura 1 – Inconsistências estatísticas relatadas em periódicos A1 de Psicologia.

Nota: Não há dados para as publicações do ano de 2017 do periódico Paidéia e dos anos de 2009 e 2018 do periódico Teoria e Pesquisa, pois o *statcheck* não conseguiu realizar a extração de dados destes anos por erros desconhecidos.

Quantidade de inconsistências por publicação

De 112 publicações em periódicos Qualis A1 de Psicologia com pelo menos 1 inconsistência

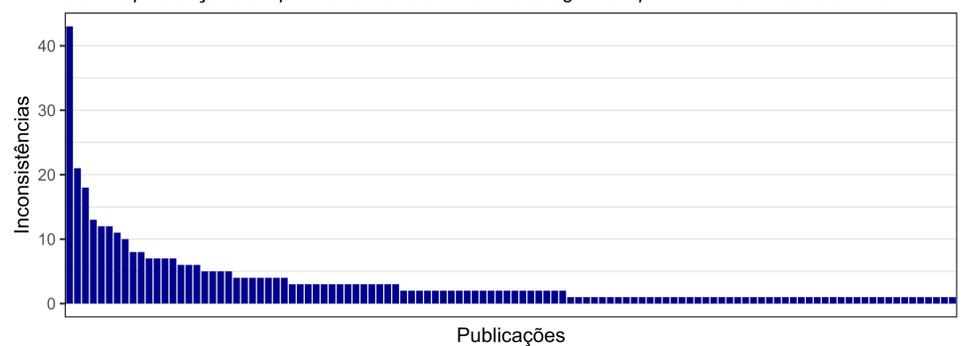


Figura 2 – Quantidade de inconsistências por publicação.

Nota: Cada dado representa uma publicação das 112 publicações com inconsistências.

## DISCUSSÃO

Esta pesquisa encontrou **resultados similares à literatura internacional** sobre a **prevalência de artigos com pelo menos 1 inconsistência e 1 inconsistência grave** (respectivamente, 40,3% e 12,6%, comparado a 49,6% e 12,9% ((Nuijten et al., 2016)). No entanto, a **proporção da quantidade total de inconsistências foi maior na literatura nacional**, com 22,9% de inconsistências e 4,4% de inconsistências graves, em comparação a 9,7% e 1,4% (Nuijten et al., 2016). Isso indica que algumas publicações nacionais parecem ter uma quantidade geral de inconsistências superior às internacionais; no entanto, esses resultados devem ser avaliados com cuidado, devido à menor amostra analisada.

A maior limitação desta pesquisa foi a quantidade de dados extraídos, devido à limitações do próprio *statcheck*, da quantidade de artigos analisados, mas também do baixo rigor das publicações nacionais com a utilização das normas APA para relatar resultados estatísticos.

## CONCLUSÃO

Esta pesquisa buscou descrever a prevalência de inconsistências em relatos estatísticos em periódicos brasileiros de Psicologia de alta relevância. **Encontrou-se números similares ou superiores aos de pesquisas internacionais**, indicando que estes **problemas de relato também são uma realidade na literatura nacional**. Os resultados chamam atenção para necessidade de autores, leitores e editores científicos terem maior cuidado com o relato e análise de resultados estatísticos em publicações. Recomenda-se também que periódicos conduzam análises pormenorizadas de resultados estatísticos de artigos antes da publicação, e incentive-se autores a adotarem práticas de ciência aberta e transparência com seus dados.

## REFERÊNCIAS

- Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2016). *The prevalence of statistical reporting errors in psychology (1985–2013)*. Behavior Research Methods, 48(4), 1205–1226. <https://doi.org/10.3758/s13428-015-0664-2>
- Nuijten, M. B., & Epskamp, S. (2023). *statcheck: Extract statistics from articles and recompute p-values (1.4.1-beta.2)* [R]. <https://github.com/MicheleNuijten/statcheck/releases/tag/v1.4.0-beta.7>
- O'Connor, B. P., & Khattar, N. (2022). *Controversies Regarding Null Hypothesis Significance Testing*. Em Avoiding Questionable Research Practices in Applied Psychology: Vol. Capítulo 7. Springer Nature.
- Open Science Collaboration. (2015). *Estimating the reproducibility of psychological science*. Science, 349(6251), 1–8