

Significância estatística: medidas e interpretações

AZEVEDO, P H S M; PANDELÓ JUNIOR, D; BOTERO, J P. Significância estatística: medidas e interpretações. **R. bras. Ci. e Mov** 2016;24(2):206-209.

Paulo Henrique Silva Marques de Azevedo¹
Domingos Pandeló Junior¹
João Paulo Botero¹

¹Universidade Federal de São Carlos

Recebido: 01/12/2015
Aceito: 01/02/2016

Os métodos quantitativos constituem-se em importante ferramenta para a análise e interpretação de resultados nas mais diversas áreas. Na área de saúde, todavia, a acuracidade da interpretação pode ser considerada ainda mais relevante, em função de razões óbvias.

Durante quase um século, o padrão de análise dos dados foi fortemente ancorado no valor de p . Ainda hoje o valor de p é muito usado, muito embora alguns periódicos comecem a recusá-lo¹, ou exigir, por parte do investigador, complemento da análise com alguma medida de tamanho de efeito.

Cabe destacar que o valor de p , desenhado por Fischer², no início do século 20, nunca deveria ter sido usado como medida de probabilidade, pois ele não se presta a isso. Caso o pesquisador deseje alguma medida de probabilidade deveria utilizar estatística bayesiana³.

Um ponto crucial, que merece destaque, é a diferença entre significância estatística e significância prática. Vale lembrar, que a significância estatística pode ser brutalmente afetada pelo tamanho da amostra⁴, de modo que em pequenas amostras, encontrar algum grau de significância estatística é uma tarefa árdua, o contrário acontecendo com amostras muito grandes. Só a título de esclarecimento, o ideal seria, no desenho da pesquisa, determinar-se o n , de acordo com o problema a ser estudado, bem como os métodos e ferramentas a serem empregados.

Como já mencionado, anteriormente, o valor de p , como única medida de significância apresenta sérias limitações, de forma que o ideal seria o complemento com outras medidas, tais como, o intervalo de confiança, ou medidas de tamanho de efeito. Para uma interessante análise crítica do valor de p , pode-se sugerir o clássico texto de Goldman⁵, *A Dirty Dozen: Twelve P - Value Misconceptions*.

Pois bem, o objetivo desta carta ao Editor é discutir 2 artigos publicados na Revista. De um lado o Artigo "Efeito agudo da suplementação de cafeína sobre o desempenho físico durante teste incremental de esforço, de Fonseca et al.⁶ e a carta ao Editor "Significância estatística e enganos práticos", de Lopes⁷. O tema é

extremamente relevante e alguns comentários adicionais talvez sejam cabíveis.

Para Cohen⁸ "*the primary product of the research inquiry is one or more measures of effect size, not p values.*" Difícil discordar, pois as medidas de tamanho de efeito são muito importantes para se avaliar os resultados de um trabalho.

As medidas de tamanho de efeito podem ser divididas em 2 grandes grupos, a família d (mensura diferença, distância entre grupos) e a família r (que mede o grau de relação entre grupos). Para uma análise mais aprofundada sobre as diversas medidas e quando aplicá-las, vale a pena consultar o livro de Ellis⁹.

Um cuidado muito especial deve ser tomado com a utilização das medidas de tamanho de efeito. Em primeiro lugar, além da medida clássica, o Cohen's d , pode-se considerar, por exemplo, a utilização do Hedges' g , especialmente em situações nas quais existam diferenças entre os tamanhos dos grupos, ou ainda quando as amostras sejam pequenas.

Um outro cuidado deve ser tomado com abordagens simplistas de análise e regras de decisão. Cohen⁸ apresentou alguns parâmetros de análise, para servirem apenas de referência, mas que acabaram tomando um outro rumo. Assim, muitos relacionam, necessariamente, um tamanho do efeito, d , com valores de 0.2, 0.5 and 0.8 como sendo considerados pequeno, médio e grande. Tal análise não está necessariamente errada, mas nem sempre estará certa. Como lembra Cumming⁴, os pesquisadores devem olhar para os seus dados, analisá-los e tentar extrair o máximo de informações dos mesmos, levando-se em consideração as peculiaridades do fenômeno estudado.

Será que um tamanho de efeito "pequeno" não pode ser importante? A resposta é sim, dependendo do contexto. O interessante da ciência e da pesquisa é isso. Em geral, regras engessadas e rígidas estão erradas. Em várias situações uma medida de tamanho de efeito pequena pode ter significância prática. Essa é uma regra básica: não se pode analisar tamanho de efeito, significância prática, sem levar em conta o contexto em que a pesquisa, o problema, está inserido. Só para ilustrar

com um exemplo, similar ao utilizado por Ellis⁹, numa prova de triatlo olímpico, pequenas diferenças podem ser a diferença entre o primeiro e o quadragésimo colocado.

Como já mencionado anteriormente e corroborado por Thompson¹⁰, Shaver¹¹ e o próprio Cohen⁸, os padrões apresentados por Cohen, como tamanho de efeito pequeno, médio e grande, pode ser extremamente perigoso, se mal utilizado.

Assim sendo, senhor Editor, o trabalho de Fonseca et al.⁶ poderia ter incluído alguma medida de tamanho de efeito para facilitar as análises e, eventualmente, torná-las mais robustas. Já a carta ao Editor, de Lopes⁷, teve o mérito de indicar a possibilidade de utilização de uma medida de tamanho de efeito, porém deveria ter considerado a utilização do Hedges' g ao invés do Cohen's d, bem como não deveria se ater unicamente nos limites apresentados por Cohen⁸ para definir o que vem a ser um tamanho de efeito pequeno, médio e grande, pois da forma que foi colocado parece que a significância prática vai necessariamente ser determinada por um maior tamanho de efeito; o que não é, necessariamente, verdadeiro, como foi discutido aqui.

Referências

1. American Psychological Association. (2010). Publication manual of the American Psychological Association Sixth Edition, Editorial Staff. Washington: APA.
2. Fisher, R.A., 1925. "Theory of Statistical Estimation", Proceedings of Cambridge Philosophical Society, Vol. 22, pp. 700-725.
3. J. V. Stone. Bayes' Rule: a Tutorial Introduction to Bayesian Analysis 2013 Sebtel 260 pp
4. Cumming, G. Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis. New York: Routledge; 2012.
5. Goodman, S. N. 2008. A dirty dozen: Twelve P-value misconceptions. *Seminars in Hematology* 45:135-140.
6. Fonseca et al. Efeito agudo da suplementação de cafeína sobre o desempenho físico durante teste incremental de esforço. *R. Bras.Ci. e Mov.* 2014; 22(1).
7. Lopes, J.M. Significância estatística e enganos práticos. *R. Bras. Ci. e Mov.* 2015; 23(1).
8. Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45, 1304 –1312. doi:10.1037/0003-066X.45.12.1304
9. Ellis PD. The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results. Cambridge, UK: Cambridge University Press; 2010.
10. Thompson, B.(2008). Computing and interpreting effect sizes, confidence intervals, and confidence intervals for effect sizes, in J.W. Osborne (editor), *Best Practices in Quantitative Methods*. Thousand Oaks, CA
11. Shaver, J.P. (1993) What statistical significance test is, and what is not. *Journal of Experimental Education*, 61 (4).