A method to streamline p-hacking

Ian Hussey

Abstract

The analytic strategy of p-hacking has rapidly accelerated our advancement of the goals of psychological science (i.e., publications, tenure, and flair: Bakker, Dijk, & Wicherts, 2012), but has suffered a number of setbacks in recent years. In order to remediate this, this article presents a statistical approach that can greatly accelerate and streamline the p-hacking process: generating random numbers that are $< .05$, which we refer to as peconomical. Results of a simulation study are presented and an R script is provided. In the absence of systemic changes to modal p-hacking practices within psychological science (e.g., worrying trends such as preregistration and replication: Munafò et al., 2017), we argue that vast amounts of time and research funding could be saved through the widespread adoption of this innovative statistical approach.

Introduction

With a few recent and unfortunate exceptions (e.g., Camerer et al., 2018; Open Science Collaboration, 2015), the discovery that p values can be hacked to support researchers' hypotheses has proven to be of exceptional utility to the enterprise of psychological science (e.g., acquiring publications, tenure, and flair; see Bakker, Dijk, & Wicherts, 2012; Simmons, Nelson, & Simonsohn, 2011 for tutorials). However, efforts to further optimize the process of p-hacking have slow rate in recent years due to a number of unfortunate setbacks such as wider use of replication and pre-registration (Munafò et al., 2017; Nosek et al., 2015; Nosek et al., 2018).

In this article, I introduce the peconomical metric and demonstrate how it can streamline the process of p-hacking your results. While this metric does suffer from the mild flaw of providing zero diagnosticity of the presence or absence of a true effect, this property is of course largely irrelevant to our aforementioned primary goals. Importantly, the metric possesses three superior characteristics. First, it is non-inferior to current p-hacking practices, which also tell us little about the presence or absence of a true effect (Hussey, 2018). Second, it retains a far more important property of hacked p values: it has high predictive validity for publishability. Finally, it also provides economic benefits relative to the high total life-cycle costs associated with traditional p-hacking (e.g., eliminates the need for comprehensive graduate training in statistics, frees up time for noise-mining other data sets).

Methods

Following standard practices, readers are suggested to skip this section and keep scrolling past any scary looking equations or R code. For more ambitious readers, the peconomical metric follows the same internal logic as traditional p-hacked analytic strategies (e.g., Bem, 2011). Loosely speaking, this conforms to the following algorithm: keep changing aspects of the analysis (e.g., exclusion strategies, covariates, IVs/DVs, grad students, your moral compass) until $p < .05$, then stop and report this value. The peconomical metric was inspired by the observation that, regardless of the specifics of any given p-hacking strategy, the product of this process is highlight reliable ($p < .05$). As such, many intermediary steps are therefore arguably unnecessary, and the same end result can be obtained more efficiently by automation. This is accomplished by generating random numbers until one is found that is $< .05$. I will refer to this approach as a form of machine learning so as to increase my chance of getting published (Hussey, Gift Authorship, and Disinterested Supervisor, 2018). An R script to calculate peconomical is provided below.

Results

Decisions made on the basis of traditional hacked p values and the peconomical metric were then compared in a simulation study. In line with modal p-hacking practices, only the key property of publishability (i.e., $p < .05$) was considered. 10,000 cases were simulated (see Appendix 1). Results demonstrated the results of peconomical and traditional p-hacked results are congruent in 100% of cases. Although variation in individual coefficients frequently differ by large margins, given that both strategies satisfy the core criterion of being diagnostic of publishability, this minor discrepancy is easily ignored.

Discussion

Traditional p-hacking involves starting with a sound analytic strategy and then iteratively degrading this until the results support one's hypothesis. On the basis that this strategy almost invariably returns significant results, many burdensome aspects of this analytic process can arguably be bypassed via automation. The most parsimonious method was selected: random number generation. Results from a simulation study demonstrate that decision making on the basis of traditional hacked p values and peconomical are equivalent, and that the latter requires several orders of magnitude less time and resources to calculate.

More radical extensions of this general strategy are also possible: use of the peconomical approach arguably nullifies the need for data collection, which arguably provides little added value beyond significant results. Academic productivity and more importantly flair can therefore be greatly increased through the widespread adoption of this approach. All materials, data, and code for the current article are most certainly not available on the Open Science Framework, you parasitic research communist (see Longo & Drazen, 2016).

References

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. Perspectives on Psychological Science, 7(6): 543–54. doi:10.1177/1745691612459060.

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive Influences on cognition and affect. Journal of Personality and Social Psychology, 100(3): 407–25. doi:10.1037/a0021524.

Camerer, C. F., Dreber, A., Holzmeister, A., Ho, T. H., Huber, J., Johannesson, M., … & Wu, H. (2018). Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. Nature Human Behaviour. doi:10.1038/s41562-018-0399-z.

Hussey, Ian. (2018). Self-citation of an unrelated paper.

Hussey, Ian, Gift Authorship, and Disinterested Supervisor. (2018). More self-citations.

Longo, D. L., & Drazen, J. M. (2016). Data sharing. New England Journal of Medicine, 374(3), 276–77. doi:10.1056/NEJMe1516564.

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Percie du Sert, K., … & Ioannidis, J. P. (2017). A manifesto for reproducible science. Nature Human Behaviour, 1(1), 0021. doi:10.1038/s41562-016-0021.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., … Yarkoni, T. (2015). Promoting an open research culture. Science, 348 (6242), 1422–5. doi:10.1126/science.aab2374.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. Proceedings of the National Academy of Sciences. Accepted manuscript. doi:10.1073/pnas.1708274114.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science, 349 (6251). aac4716. doi:10.1126/science.aac4716.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological Science, 22(11), 1359–66. doi:10.1177/0956797611417632.