

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., acquiring publications, tenure, and flair: Bakker, Dijk, and 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. 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 argued that vast amounts of time and research funding could be saved through the widespread adoption of this innovative statistical approach.

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 (i.e., acquiring publications, tenure, and flair; see Bakker, Dijk, and Wicherts 2012; Simmons, Nelson, and 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 (e.g., wider use of replication and pre-registration: see Munafò et al. 2017; Nosek et al. 2015; Nosek et al. 2018).

In this article, I introduce the $p_{\text{economical}}$ 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 largely irrelevant to our primary goals (e.g., giving Ted Talks). 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., non-comprehensive graduate training in statistics, or the opportunity costs of time that could be spent noise-mining other data sets).

Methods and results

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 $p_{\text{economical}}$ metric follows the same internal logic as traditional p -hacked analytic strategies (see 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 $p_{\text{economical}}$ 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). A R script to calculate $p_{\text{economical}}$ is provided below.

```
# set an initial value
p_economical <- 1

# generate random numbers and stop when one is < .05
while (p_economical >= .05) {
  p_economical <- runif(n = 1)
}
```

```
# print this value
print(paste("p_economical =", round(p_economical, 3)))

## [1] "p_economical = 0.005"
```

Decisions made on the basis of traditional hacked p values and the $p_{economical}$ 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 $p_{economical}$ 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 $p_{economical}$ 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: given the grey area between intentional p -hacking and outright data fabrication (given that both serve to purposefully fabricate results), use of the $p_{economical}$ approach could also nullify the need for data collection, which arguably provides little added value. 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 and Drazen 2016).

Appendix 1: R code for simulation study

```
simulation <- function() {

  # simulate publishability of results from econo_p
  # set an initial p value
  econo_p <- 1

  # generate random values for p, and stop when this value is < .05
  while (econo_p >= .05) {
    econo_p <- round(runif(1), 3)
  }

  # decision making
  if(econo_p < 0.05) {
    publishable_econo_p = TRUE
  } else {
    publishable_econo_p = FALSE
  }

  # simulation publishability of results from tradition (hacked) p values
  # p value set to upper bound of observable hacked p values
  p = 0.049
```

```

# decision making
if(p < 0.05) {
  publishable_p = TRUE
} else {
  publishable_p = FALSE
}

return(publishable_econo_p == publishable_p)
}

# proportion of 10,000 simulated cases where conclusions agree
mean(replicate(10000, simulation()))
## [1] 1

```

References

- Bakker, M., A. van Dijk, and J. M. Wicherts. 2012. “The Rules of the Game Called Psychological Science.” *Perspectives on Psychological Science* 7 (6): 543–54. doi:10.1177/1745691612459060.
- Bem, Daryl 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, Colin F., Anna Dreber, Felix Holzmeister, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, et al. 2018. “Evaluating the Replicability of Social Science Experiments in Nature and Science Between 2010 and 2015.” *Nature Human Behaviour*, August, 1. 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, Dan L., and Jeffrey M. Drazen. 2016. “Data Sharing.” *New England Journal of Medicine* 374 (3): 276–77. doi:10.1056/NEJMe1516564.
- Munafò, Marcus R., Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware, and John P. A. Ioannidis. 2017. “A Manifesto for Reproducible Science.” *Nature Human Behaviour* 1 (1): 0021. doi:10.1038/s41562-016-0021.
- Nosek, Brian A., G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, et al. 2015. “Promoting an Open Research Culture.” *Science* 348 (6242): 1422–5. doi:10.1126/science.aab2374.
- Nosek, Brian A., Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor. 2018. “The Preregistration Revolution.” *Proceedings of the National Academy of Sciences*, March. 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, Joseph P., Leif D. Nelson, and Uri Simonsohn. 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.