The Perils of P-Hacking and the Promise of Pre-Analysis Plans

Synopsis

One response to the problems of publication bias and data dredging (or 'p-hacking') in the medical and social sciences is to demand that researchers formulate pre-analysis plans for their empirical studies. In a pre-analysis plan, researchers specify, in advance, the measurements and statistical analyses they plan to perform, committing themselves to specific statistical hypotheses prior to data analysis. Pre-analysis plans are used widely in the medical sciences but have only begun to gain traction in empirical social science research in the last decade.

The danger of data dredging is typically noted in the context of frequentist statistics, in which analysts run multiple tests in the search for statistically significant 'p-values'. For this reason, data dredging is sometimes referred to as 'p-hacking'. Roughly, the worry about p-hacking is that statistically significant findings can arise due to chance or bias in the data-generating method, so for any set of analyses on a particular set of data, there may be a proportion of statistically significant findings that are misleading 'false positives'.

Data dredging is related to a more traditional concern of philosophers, namely, the epistemic merit of the accommodation of existing evidence (by a hypothesis which is tuned to accommodate that evidence) versus the prediction of new evidence (by a hypothesis which was developed prior to the observation of the evidence). Some argue that prediction is epistemically superior to accommodation, while others argue that prediction and accommodation are epistemically equivalent. The status quo in this debate is that prediction is epistemically superior to accommodation in particular circumstances. We help ourselves to the fruits of this debate, and argue that pre-analysis plans serve a function like prediction, delivering epistemic benefits in particular circumstances. In this synopsis, we discuss the argumentative arc of the full article, though the present submission includes only §1 and §2 to meet the APA length requirement.

Our contributions are threefold. First, we begin by describing the danger of data dredging, from the perspective of Bayesianism and model selection theory (§2). This explication is an important and novel contribution, because the danger of data dredging is usually discussed in, and said to arise from, a context of frequentist statistics. Drawing on Bayesian discussions of the 'problem of old evidence' and the description sensitivity of evidence, we articulate from a Bayesian perspective the perils of p-hacking. We then articulate the perils of p-hacking using the resources of 'model selection theory', which holds that: (i) simpler models are more predictively accurate than complex models; and (ii) models with better fit-with-data are more predictively accurate than models with worse fit-with-data.

The second contribution is an articulation of the epistemic benefits of pre-analysis plans as a means to mitigate the dangers of data dredging (§3). By noting in advance the analyses they will perform, scientists avoid the dangers discussed in §2. We show that pre-analysis plans function like 'stopping rules.' While Bayesians tend to deny the epistemic importance of stopping rules, we demonstrate why they are epistemically important from a Bayesian perspective. In addition, we demonstrate that while the presence of a pre-analysis plans might lead to models with worse fit-to-data scores, they lead to more parsimonious models and thus are still favored by model selection theory.

Even when pre-analysis plans are in place, scientists often depart from them. In light of this fact, our third contribution is that we describe the conditions under which such departures are suspect, and conversely, the conditions under which such departures are welcome (§4). In cases in which the pre-analysis plan is not strictly followed, we argue that the pre-analysis plan is still valuable if it captures certain features of the hypotheses formulated at different times over the course of a researcher's study.

The danger of data dredging and the promise of pre-analysis plans are important issues in contemporary scientific practice. Furthermore, data dredging is a focal point for a cluster of related issues in the philosophy of science. As we have indicated, data dredging is closely related to the debate over predictivism, but also other epistemological issues around experimental design such as stopping rules, model-tuning, over-fitting, and 'big-data'-mining techniques. We close with a discussion of some of these issues (§5).