

## Approaching Significance: A Cautionary Tail

David Medler

Null Hypothesis Significance Testing (NHST) is the predominant form of inference testing in Psychology (and most of science). The fundamental principle of NHST is that we compare our sample to a sampling distribution that is based on random behaviour in the population. If our sample statistic is far enough into the tail of the sampling distribution, then we declare that our results are statistically significant, with a  $p < .05$ , and publish our results. What happens, however, if  $p > .05$ ? A survey of the literature shows that the terms “approaching significance”, “marginally significant”, etc., are quite prevalent in the literature, and are used when  $p$  is close, but has not quite met the “magical”  $.05$  level. This situation is not so dire, as we can evaluate those statistics at their face value. But, a larger problem potentially exists, which has been identified as p-hacking. This is where a researcher may inadvertently inflate their probability of a Type I error through methodological procedures, and which cannot be picked up in a published scientific article. Here I present a series of Monte Carlo simulations looking at two specific instances of p-hacking. First, we consider adding more participants to a study following an initial inferential result that was “approaching significance”. Simulations indicate that in this case, we may increase our probability of a Type I error from  $.05$  to  $.50$ ! Second, we consider the case where we remove participants from a study once we have done our initial inferential test. Depending on how these participants are removed, we may inadvertently increase our alpha to  $1.00$ . These methodological issues may contribute to the reported inability to replicate some scientific findings. This is simply a cautionary tail, but one that may be worth considering.