统计学指标p值还可信吗？

作者：writer 来源：Enago英论阁科学网博客

本文原地址：https://www.iikx.com/news/statistics/5254.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！



统计学指标p值还可信吗？科研工作者经常用统计学方法来判断某个结果是否在不同的实验条件下存在差异，并且判断这种差异是否由于偶然因素造成。使用最广泛的统计指标是p值，用来检测差异的显著性水平。“ p<0.05 ”代表着样本间的差异由抽样误差所致的概率小于0.05，这种有统计学显著性差异的结果正是科研人员梦寐以求的。有了这样的结果，就可能意味着论文发表，科研经费到位，甚至功成名就，对p值的滥用和误用现象也因此而生，即所谓的p-hacking，指p值篡改或p值操纵。

由于p<0.05被推上神坛，被视为论文可能发表的指标，就有科研者费尽心机操纵p值。不管是有心还是无意，科研者会不断尝试各种统计学方法和数据，直至p<0.05。P-hacking的手段有对变量进行选择性取舍、对数据进行选择性删除、对样本量进行选择性增加、在统计分析后把指标进行合并或拆分、发现p值有意义后就停止收集数据等等。这样的统计分析很容易产生假阳性结果，也导致实验的不可重复性。可悲的是，这种行为在发表的科研论著中相当普遍。

2015年，Science发表了一项实验心理学的大规模重复性研究，仅有39%的结果能够被成功重复。Nature的一项调查发现70%的科研人员无法重复他人的研究，50%的科研人员甚至无法重复自己的实验结果。科研的信用已经被“ 可重复性危机 ”损害，而p-

hacking对此也起到了推波助澜的作用，让很多统计学家和科学家都非常担忧。

如何避免p值操纵呢?最好的办法就是提高科研者对p-hacking危害性的认识，避免对实验数据进行选择性地分析。弗吉尼亚大学心理学教授Brian Nosek发起成立了推动科研界更开放更透明的公益组织—开放科学中心(Center for Open Science)，帮助科研者更好的设计实验，提高实验的可重复性。比如，双盲实验就可以避免科研者选择性地处理数据。

另外一项统计学指标“信心指数”(Confidence Index)可以用来更加准确地评估实验数据的可靠性。p值依靠样本量和反应频率，而信心指数不同，它是个复合指标，由以下三点决定：随机变量，先验概率，假设真相。$p<0.05$说明实验结果有95%的可能性是有意义的，而信心指数则明确清晰的说明结果有意义的具体可能性是多少。这样的统计指标尤其对临床研究的决策有指导性意义。斯坦福大学医学院流行病学教授Steven Goodman在Nature上撰文建议使用信心指数，可以减少对p值的依赖和舞弊。

统计学家在2016年初专门发表声明，澄清p值的真正意义以及应该如何正确使用p值。然而p值仍延续着一直被批判，从未被取代的地位。如何对科研界的现状进行有效的改变，还需要科研者和学术出版界共同努力，探索新的统计学标准以及适用的领域。

更多 统计方法 请访问 https://www.iikx.com/news/statistics/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发