
P值降至0.005，你的研究结果仍然有意义吗

作者：倪凯文，赵一鸣 来源：临床流行病学和循证医学

本文原地址：<https://www.iikx.com/news/statistics/84.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！

前些日子，在JAMA上发表了一篇文章，建议在医学研究领域，将统计学P值的阈值从0.05下调至0.005。仅看标题，小编和大多数人的第一想法是一样的，以后发paper的难度系数激增啊！现如今，我们做临床研究不算个P值，都不好意思说自己是在做科研的；当然，不做出个 $P < 0.05$ ，你也不好意思去发paper。

P值是怎么来的呢？Fisher作为一代假设检验思想的创立者，在假设检验中首先提出P值的概念。随着后来Pearson和Neyman等统计学专家的理论完善，P值被认为可以在一定程度上验证假设的正确与否。根据小概率原理，假如P值越小，说明零假设情况的发生概率越小，我们拒绝原假设的理由越充分。

这看似很有逻辑的理论推导，但是在实际运用上却存在很多缺陷。例如，P值对于样本量的敏感度很高，在如今火热的健康大数据背景下，想获得一个 $P < 0.05$ 结果是非常容易的，但是研究结果不一定有实际意义。另外，大样本量的小效应和小样本量的大效应可以产生相同的P值，这时P值所能提供的信息量较少（置信区间的重要性）。其实，统计学的功能在于归纳推论而不是归纳行为。然而，我们却用P值来一刀切的来判断是否，进而指导临床决策，这是对于P值的错误理解和错误使用。

美国统计协会(ASA)发现P值被大家过度使用并已经玩坏了，在2016年发表声明，提出6条关于P值的准则，用来澄清正确使用和解释P值。

准则1：P值可以用于表达数据和特定统计模型之前不相容的程度。

我们通常会设立一个零假设，如A药和安慰药疗效效果相同，一个备择假设，A药和安慰药疗效效果不相同，通过已有数据构建统计模型，获得的P值越小，认为数据与零假设不相容程度越大。

准则2：P值并不能度量研究假设为真的概率。

P值只能解释数据与假设间的关系，而不能描述假设本身。举一个大多数研究者错误的理解，如P值为0.03，就认为A药和安慰药疗效效果相同有3%的可能性是真的，97%的正确概率认为两者疗效效果不同（这个理解是错误的）。

准则3：科学结论、商业决策或政策决定不可以仅基于P值是否高于某个特定的阈值。

研究者不能简单地将P值大小与0.05作比较来对研究结果一刀切，不能把P值的作用过于神话。

准则4：合理的推断过程需要完整的报告和透明度。

研究者不能为了找到有意义的P值，而做统计分析。同时，不能有选择性的报告P值和相关分析内容。

准则5：P值或统计显著性并不能衡量效应值的大小或结果的重要性。

前文也提到大样本量的小效应和小样本量的大效应可以产生相同的P值，所以P值所能提供的信息量不包括效应值的大小和结果的实际意义。

准则6：P值本身并不对模型或者假设提供一个好的衡量。

P值能提供的信息量有限，有时其他统计参数要比P值更有参考价值。

最后，我们回到这篇JAMA文章中，作者提出降低阈值，主要还是因为P值被太多研究者乱用，阈值的降低可以避免错误的科学结论。至于为什么降至0.005而不是其他界值，并没有给出特别具体的科学依据。那各位读者们，觉得P值的阈值应该降低吗？如果应该降低，降低到多少比较好呢？欢迎大家留言讨论。

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](https://www.iikx.com)转发