

美国统计协会开始正式吐槽（错用）P值啦

🕒 2016/03/08 📁 数据分析、新闻通知、统计刊物、职业事业 📌 P值、声明、数据、模型、科学、统计协会 👤 邱怡轩

<u>P-VALUE</u>	<u>INTERPRETATION</u>
0.001	HIGHLY SIGNIFICANT
0.01	
0.02	
0.03	
0.04	SIGNIFICANT
0.049	
0.050	OH CRAP. REDO CALCULATIONS.
0.051	ON THE EDGE OF SIGNIFICANCE
0.06	
0.07	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE $P < 0.10$ LEVEL
0.08	
0.09	
0.099	HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS
≥ 0.1	

(图片来源 : <https://xkcd.com/1478> ,
一幅讽刺滥用P值的漫画)

今天美国统计协会 (ASA) 正式发布了一条关于P值的声 (吐) 明 (槽) , 算起来可以说是近期统计学界的一件大事了。为什么这么说呢 ? 首先 , P值的应用太广 , 所以对P值进行一些解释和声明非常有必要。其次 , 对P值的吐槽历来有之 , 但今天是第一次被一个大型的专业协会以非常正式的形式进行澄清 , 多少带有一些官方的意思。声明的全文可以在[这个页面](#)中下载。

那么这则声明里面都说了什么呢 ? 小编整体读了一遍 , 把我认为重要的信息概括在这篇文章之中。

首先，ASA介绍了一下这则声明诞生的背景。2014年，ASA论坛上出现了一段如下的讨论：

问：为什么那么多学校都在教 $p = 0.05$ ？

答：因为那是科学团体和期刊编辑仍然在用的标准。

问：为什么那么多人还在用 $p = 0.05$ ？

答：因为学校里还在这么教。

看上去多少有点讽刺的味道，但事实却也摆在眼前。从舆论上看，许许多多的文章都在讨论P值的弊端，小编摘录了几条言辞比较激烈的：

这是科学中最肮脏的秘密：使用统计假设检验的“科学方法”建立在一个脆弱的基础之上。——ScienceNews (Siegfried, 2010)

假设检验中用到的统计方法……比Facebook隐私条款的缺陷还多。——ScienceNews (Siegfried, 2014)

针对这些对P值的批评，ASA于是决定起草一份声明，一方面是对这些批评和讨论作一个回应，另一方面是唤起大家对科学结论可重复性问题的重视，力图改变长久以来一些已经过时的关于统计推断的科学实践。经过长时间众多统计学家的研讨和整理，这篇声明今天终于出现在了我们的面前。

P值是什么

这份声明首先给出了P值一般的解释：P值指的是在一个特定的统计模型下，数据的某个汇总指标（例如两样本的均值之差）等于观测值或比观测值更为极端的概率。

这段描述是我们通常能从教科书中找到的P值定义，但在实际问题中，它却经常要么被神话，要么被妖魔化。鉴于此，声明中提出了六条关于P值的准则，作为ASA对P值的“官方”态度。这六条准则算是这条声明中最重要的部分了。

准则1：P值可以表达的是数据与一个给定模型不匹配的程度。

这条准则的意思是说，我们通常会设立一个假设的模型，称为“原假设”，然后在这个模型下观察数据在多大程度上与原假设背道而驰。P值越小，说明数据与模型之间越不匹配。

准则2：P值并不能衡量某条假设为真的概率，或是数据仅由随机因素产生的概率。

这条准则表明，尽管研究者们在很多情况下都希望计算出某假设为真的概率，但P值的作用并不是这个。P值只解释数据与假设之间的关系，它并不解释假设本身。

准则3：科学结论、商业决策或政策制定不应该仅依赖于P值是否超过一个给定的阈值。

这一条给出了对决策制定的建议：成功的决策取决于很多方面，包括实验的设计，测量的质量，外部的信息和证据，假设的合理性等等。仅仅看P值是否小于0.05是非常具有误导性的。

准则4：合理的推断过程需要完整的报告和透明度。

这条准则强调，在给出统计分析的结果时，不能有选择地给出P值和相关分析。举个例子来说，某项研究可能使用了好几种分析的方法，而研究者只报告P值最小的那项，这就会使得P值无法进行解释。相应地，声明建议研究者应该给出研究过程中检验过的假设的数量，所有使用过的方法和相应的P值等。

准则5：P值或统计显著性并不衡量影响的大小或结果的重要性。

这句话说明，统计的显著性并不代表科学上的重要性。一个经常会看到的现象是，无论某个效应的影响有多小，当样本量足够大或测量精度足够高时，P值通常都会很小。反之，一些重大的影响如果样本量不够多或测量精度不够高，其P值也可能很大。

准则6：P值就其本身而言，并不是一个非常好的对模型或假设所含证据大小的衡量。

简而言之，数据分析不能仅仅计算P值，而应该探索其他更贴近数据的模型。

声明之后还列举出了一些其他的能对P值进行补充的分析方法，比如置信区间，贝叶斯方法，似然比，FDR (False Discovery Rate) 等等。这些方法都依赖于一些其他的假定，但在一些特定的问题中会比P值更为直接地回答诸如“哪个假定更为正确”这样的问题。

声明最后给出了对统计实践者的一些建议：好的科学实践包括方方面面，如好的设计和实施，数值上和图形上对数据进行汇总，对研究中现象的理解，对结果的解释，完整的报告等等——科学的世界里，不存在哪个单一的指标能替代科学的思维方式。

参考文献：[ASA Statement on Statistical Significance and P-values](#)

附原文：<https://cos.name/2016/03/asa-statement-on-p-value/#more-11902>