
Numbers, Science & Society

Experts issue warning on problems with P values

Misunderstandings about common statistical test damage science and society

By Tom Siegfried 10:30am, March 11, 2016

Here's a good idea for the next presidential candidate debate: They can insult each other about their ignorance of statistics.

Actually, it's a pertinent topic for political office seekers, as public opinion polls use statistical methods to measure the electorate's support (or lack thereof) for a particular candidate. But such polls are notoriously unreliable, as Hillary Clinton found out in Michigan.

It probably wouldn't be a very informative debate, of course — just imagine how Donald Trump would respond to a question asking what he thought about P values. Sadly, though, he and the other candidates might actually understand P values just about as well as many practicing scientists — which is to say, not very well at all.

In recent years criticism about P values — statistical measures widely used to analyze experimental data in most scientific disciplines — has finally reverberated loudly enough for the scientific community to listen. A watershed acknowledgment of P value problems appeared this week when the American Statistical Association issued a statement warning the rest of the world about the limitations of P values and their widespread misuse.

"While the p-value can be a useful statistical measure, it is commonly misused and misinterpreted," the statistical association report stated. "This has led to some scientific journals discouraging the use of p-values, and some scientists and statisticians recommending their abandonment."

In light of these issues, the association convened a group of experts to formulate a document listing six "principles" regarding P values for the guidance of "researchers, practitioners and science writers who are not primarily statisticians." Of those six principles, the most pertinent for people in general (and science journalists in particular) is No. 5: "A p-value, or statistical significance, does not measure the size of an effect or the importance of a result."

What, then, does it measure? That's principle No. 1: "... how incompatible the data are with a specified statistical model." But note well principle No. 2: "P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone." And therefore, always remember principle No. 3: "Scientific conclusions ... or policy decisions should not be based only on whether a p-value passes a specific threshold."

In other words, the common convention of judging a P value less than .05 to be "statistically significant" is not really a proper basis for assigning significance at all. Except that scientific journals still regularly use that criterion for deciding whether a paper gets published. Which in turn drives researchers to finagle their data to get a P value of less than .05. As a result, the scientific process is tarnished and the published scientific literature is often unreliable.

As the statistical association statement points out, this situation is far from merely of academic concern.

"The issues touched on here affect not only research, but research funding, journal practices, career advancement, scientific education, public policy, journalism, and law," the authors point out in the report, published online March 7 in *The American Statistician*.

Many of the experts who participated in the process wrote commentaries on the document, some stressing that it did not go far enough in condemning P values' pernicious influences on science.

"Viewed alone, p-values calculated from a set of numbers and assuming a statistical model are of limited value and frequently are meaningless," wrote biostatistician Donald Berry of MD Anderson Cancer Center in Houston. He cited the serious negative impact that misuse and misinterpretation of P values has had not only on science, but also on society. "Patients with serious diseases have been harmed. Researchers have chased wild geese, finding too often that statistically significant conclusions could not be reproduced. The economic impacts of faulty statistical conclusions are great."

Echoing Berry's concerns was Boston University epidemiologist Kenneth Rothman. "It is a safe bet that people have suffered or died because scientists (and editors, regulators, journalists and others) have used significance tests to interpret results," Rothman wrote. "The correspondence between results that are statistically significant and those that are truly important is far too low to be useful. Consequently, scientists have embraced and even avidly pursued meaningless differences solely because they are statistically significant, and have ignored important effects because they failed to pass the screen of statistical significance."

Stanford University epidemiologist John Ioannidis compared the scientific community's attachment to P values with drug addiction, fueled by the institutional rewards that accompany the publication process.

"Misleading use of P-values is so easy and automated that, especially when rewarded with publication and funding, it can become addictive," Ioannidis commented. "Investigators generating these torrents of P-values should be seen with sympathy as drug addicts in need of rehabilitation that will help them live a better, more meaningful scientific life in the future."

Although a handful of P value defenders can still be found among the participants in this discussion, it should be clear by now that P values, as currently used in science, do more harm than good. They may be valid and useful under certain specific circumstances, but those circumstances are rarely relevant in most experimental contexts. As Berry notes, statisticians can correctly define P values in a technical sense, but "most statisticians do not really understand the issues in applied settings."

In its statement, the statistical association goes a long way toward validating the concerns about P values that have been expressed for decades by many critical observers. This validation may succeed in initiating change where previous efforts have failed. But that won't happen without identifying some alternative to the P value system, and while many have been proposed, no candidate has emerged as an acceptable nominee for a majority of the scientific world's electorate. So the next debate should not be about P values — it should be about what to replace them with.

Follow me on Twitter: [@tom_siegfried](https://twitter.com/tom_siegfried)

Source URL: <https://www.sciencenews.org/blog/context/experts-issue-warning-problems-p-values>

Source URL: <https://www.sciencenews.org/blog/context/experts-issue-warning-problems-p-values>