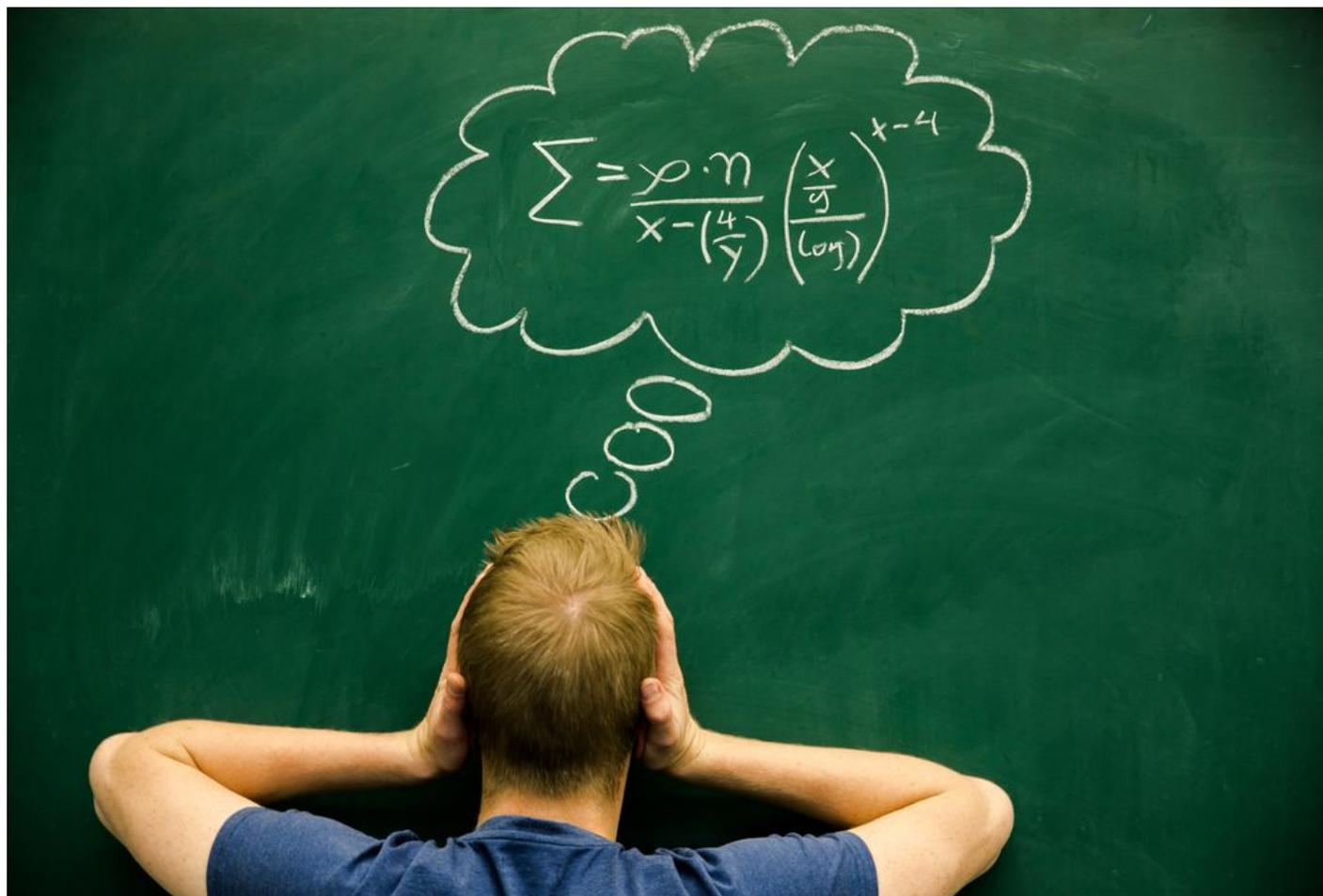


Dump 'Statistical Significance,' Then Teach Scientists Statistics

Researchers need a new gold standard to assess their work. They also ought to stop indulging the fear of math.

By [Ariel Procaccia](#)

March 29, 2019, 8:00 AM EDT



What's that you say? *Photographer: Lucas Knappe/EyeEm*

Did you know that gorging on dark chocolate accelerates weight loss? A study published in 2015 found that a group of subjects who followed a low-carbohydrate diet and ate a bar of dark chocolate daily lost more weight than a group that followed the same diet sans chocolate. This discovery was heralded in some quarters as a scientific breakthrough.

If you're still hesitant about raiding the supermarket chocolate aisle, rest assured: The study's results are statistically significant. In theory, this means that the results would be improbable if chocolate did not contribute to weight loss, and therefore we can conclude that it does. A successful test of statistical significance has long been the admission ticket into the halls of scientific knowledge.

But not anymore, if statisticians have their way. In a coordinated assault last week, which included a special issue of the American Statistician and commentary in Nature (supported by 800 signatories), some of the discipline's luminaries urged scientists to ditch the notion of statistical significance.

Critics argue that statistical significance can be misleading because it sets an arbitrary threshold on the level of uncertainty science should be willing to accept. Roughly speaking, uncertainty is expressed as the likelihood of observing an experimental result by chance, assuming the effect being tested doesn't actually exist. In statistical lingo this likelihood is known as p-value. Statistical significance typically requires a p-value of less than 5 percent, or 0.05. A p-value of 0.049 is under the 5-percent threshold; thus results returning that value are considered "significant." If $p=0.051$, by contrast, the results are "not significant," despite the tiny difference between the two values.

This has led to myriad problems. One is that there's a perceived crisis of reproducibility in science, in part because the p-value itself is uncertain: Flawlessly repeating the same experiment can produce different values, crossing the magical significance threshold in either direction. Another problem is the practice of (often innocently) testing many hypotheses and reporting only those that give statistically significant results.

The latter issue is nicely illustrated by the chocolate study, which was nothing but a sting operation designed to show how easy it is to draw international media attention to flashy results even when the underlying science is cringe-worthy. The experiment was real, but it had only 15 subjects. Worse, 18 different hypotheses were tested, including "chocolate reduces cholesterol" and "chocolate contributes to quality of sleep." Life may very well be like a box of chocolates, ^[1] but if you roll the dice enough times, you know exactly what you're going to get: results that are both statistically significant and fallacious.

I agree that the term "statistical significance" is part of the problem; abandoning it is the right thing to do. In its place, statisticians advocate a more nuanced view of uncertainty. For example, scientists can report a range of possible conclusions that are compatible (to different degrees) with the data.

But the problem runs deeper. The broader issue is that the choice of a career in medicine, the life sciences or the social sciences (with some exceptions, like economics) isn't typically indicative of a passion, or even an aptitude, for mathematics. Yet these sciences are thoroughly infused with statistics, and a shallow understanding of its principles gives rise to numerous fallacies.

In a 1994 editorial in the BMJ, the late English statistician Douglas Altman wrote that many medical researchers "are not ashamed (and some seem proud) to admit that they don't know anything about statistics." It does appear to be a sociological phenomenon. ^[2] A friend who is a medical doctor and researcher, and excelled in math in high school, once told me that he and his peers developed a collective aversion to the subject shortly after entering medical school. It seems that people who aren't squeamish about holding a beating heart in their hands might, in my friend's words, "faint at the sight of an integral."

To find examples of ignorance, one doesn't even need to look for statistical subtleties. I was amused to read a few years ago of the "one in 48 million baby" who was born in Australia on the same date as her mother and her father. Under the assumptions presumably made by the good doctor who announced the miracle, the odds are actually 1 in 133,225 (1 in 365 squared, not 1 in 365 cubed). The same thing is likely to happen on any given day, somewhere in the world.

These anecdotes don't amount to statistically significant (oops) evidence, but there are plenty of surveys showing widespread misuse of statistics. For instance, a 2011 study published in Nature Neuroscience reviewed 513 neuroscience papers that appeared in five of the field's most prestigious journals, and found that 157 of them compared two experimental effects. Instead of a direct statistical comparison, 79 of these papers (just above half) used a blatantly incorrect procedure, claiming that the two effects differed based on the finding that one effect was statistically significant and the other wasn't. Unfortunately, as the authors note, "the difference between significant and not significant is not itself necessarily significant."

Let's move to a world beyond $p < 0.05$, then. But let's also move to a world where people don't find "Statistics for Terrified Biologists" to be a particularly useful read. We can only hope that the bestsellers of the future would have titles like "Why Biologists Love Statistics," and "Biology for Terrified Computer Scientists."

1 This is a strange proverb, as very few chocolate boxes contain any genuine surprises.

2 Perhaps someone should study it – using statistics.

This column does not necessarily reflect the opinion of the editorial board or Bloomberg LP and its owners.

To contact the author of this story:

Ariel Procaccia at arielpro@cs.cmu.edu

To contact the editor responsible for this story:

Jonathan Landman at jlandman4@bloomberg.net

Ariel Procaccia is Gordon McKay Professor of Computer Science at Harvard University. His areas of expertise include artificial intelligence, theoretical computer science and algorithmic game theory.

[Read more opinion](#)