

Mini-Project 5: Discussing p-values

Tanner Bessette

2025-04-18

1. Towards the end of Section 1, the authors say “As ‘statistical significance’ is used less, statistical thinking will be used more.” Elaborate on what you think the authors mean. Give some examples of what you think embodies “statistical thinking.”

I think the author means that using p-values as the end all be all of “statistically significant” is a bit lazy, and does not lead to the deeper dive into the data and its applications, which is what they argue a p-value should be used for. I think that “statistical thinking” is how we can use test statistics and models we’ve derived from data to interpret that data in a real-world context. I think it also includes noting limitations and accounting for them, such as data we may not have, potential confounding variables, or potential bias.

2. Section 2, third paragraph: The authors state “A label of statistical significance adds nothing to what is already conveyed by the value of p ; in fact, this dichotomization of p-values makes matters worse.” Elaborate on what you think the authors means.

I think that by the first part of the quote, the author means that if the significance threshold is a p-value of 0.05, we already know what that means, so interpreting it adds nothing. And by the second part, I think the author is saying one of the biggest problems of p-values is that it splits things into two categories: statistically significant or not statistically significant, and everything is one of the two. This is a dangerous interpretation, and the author discusses this, specifically noting that we shouldn’t believe that an association or effect is absent just because the p-value is not significant, or believe that an association exists just because the p-value is significant.

3. Section 2, end of first column: The authors state “For the integrity of scientific publishing and research dissemination, therefore, whether a p-value passes any arbitrary threshold should not be considered at all when deciding which results to present or highlight.” Do you agree or disagree? How should it be decided which results to present/highlight in scientific publishing?

I completely agree. I think that the point of research is to learn and to educate others. Whether something is “statistically significant” or not, findings can still be useful, and handpicking

only points that support the researcher's opinions entering the experiment is dishonest and potentially harmful for others who use that research as a resource.

4. Section 3, end of page 2: The authors state "The statistical community has not yet converged on a simple paradigm for the use of statistical inference in scientific research – and in fact it may never do so. A one-size-fits-all approach to statistical inference is an inappropriate expectation." Do you agree or disagree? Explain.

I agree with this as well. The whole reason I've taken so many statistics and mathematics courses and learned so many things is because there are so, so many different ways that we can model, analyze, and interpret data and it is part of what is so cool about statistics in general. Using just one approach does not tell as much of the story as using many different approaches and finding similar results in all of them.

5. Section 3.2: The authors note that they are envisioning "a sort of 'statistical thoughtfulness'" What do you think "statistical thoughtfulness" means? What are some ways to demonstrate "statistical thoughtfulness" in an analysis?

When reading the article, I had highlighted the quote from the authors on page 4, when they say, "Thoughtful research prioritizes sound data production by putting energy into the careful planning, design, and execution of the study." I like this as a definition of statistical thoughtfulness, but I also think more broadly, demonstrating statistical thoughtfulness in an analysis includes making sure there is an ethical approach to studies, making sure that experts' thoughts on the topics being investigated are still considered, and acknowledging that statistical findings are never the final say (more research in any topic is always better than less).

6. Section 3.2.4: A few of the authors of papers in this special issue argue that some of the terminology used in statistics, such as "significance" and "confidence" can be misleading, and they propose the use of "compatibility" instead. What do you think they believe the problem is? Do you agree or disagree (that there is a problem and that changing the name will help)?

I think they believe that using terms like "significance" and "confidence" suggests a definiteness to findings, and not just that something is a good fit or some finding suggests looking into something further, which would be more in tune with "compatibility". I would say I somewhat agree and somewhat disagree. I agree with the authors' sentiments that statisticians and researchers should stay modest and be thoughtful, and that goes hand-in-hand with not having far-reaching claims or interpretations. But I also think that confidence intervals, p-values, and other statistical tests can provide a ton of helpful insight, and when used properly they still have an important place in statistics.

7. Find a quote or point that really strikes you (i.e., made you think). What is the quote (and tell me where to find it), and why does it stand out to you?

The second I saw this quote in the reading I knew I was gonna write about it in at least one of my responses: "As Gelman and Stern (2006) famously observed, the difference between

“significant” and “not significant” is not itself statistically significant.” (end of fourth paragraph on page 2). This is such a good way of wording a problem that I’ve always wondered about - if we set a significance level of 0.05, how can we say definitively that a p-value of 0.049 is significant but 0.051 is not significant? The difference between those two values is so small yet in terms of interpretations it is so huge, and the article also argues that this difference shouldn’t be treated as if it’s some huge difference.