THE WEIGHT OF WHAT MIGHT BE: REFLECTIONS ON STATISTICAL SIGNIFICANCE

AKADEMIKA: EDUCATIONAL LEARNING ANTHOLOGY

Volume: 1
Issue 1
Pages: 69-71
Document ID: 2025AELA0017
DOI: 10.70838/aela.010117
Manuscript Accepted: 05-10-2025

Akad Educ Learn, 2025, **1**(1): 69-71, Document ID: AELA0017, doi:10.70838/aela.010117, ISSN xxxx-xxxx

Research Article

The Weight of What Might Be: Reflections on Statistical Significance

Jonathan Edwin Remegia*
For affiliations and correspondence, see the last page.

One of the most profound influences in the history of scientific research is probably the concept of statistical significance. This arbitrary limit dictates what findings will garner attention and what will be disregarded. As researchers, we deeply appreciate these probabilities, which provide insight into whether or not what we observe is simply a mere fluke or a glimpse into reality. In this essay, I explore the statistical significance of a finding and the power frame of scientific research, incorporating both the established principles and emerging perspectives in the field.

Statistical significance resolves the most nagging concern for every researcher: the possibility of being fooled by a sheer stroke of luck. In other words, as Gulati explained in 2025, statistical significance provides a discernible framework of effect and no-effect based on random chance. This idea rests on a branch of mathematics dealing with possibilities, specifically calculating how likely it is to obtain particular findings (or even more radical ones) in the absence of a particular effect. This probability, which is known as the p-value, has becomes the gatekeeper of scientific discovery.

The long-held threshold of $p < 0.05$ as determined by R.A. Fisher is widely and almost unquestioningly accepted across science disciplines. This threshold denotes that anything with a less than 5% probability of occurring by randomness is considered "significant." But this simple-sounding guideline conceals a tangled web of one's judgment, history, and the mathematics of science. The posited 0.05 value was not intended to be a boundary demarcating delineating truth and falsehood, but a level deemed fit for preliminary consideration (Tenny & Abdelgawad, 2023).

What renders statistical significance powerful, and simultaneously problematic, is the value we afford it - or give far too much credit to. In the context of clinical medicare, derived conclusions can have life-changing impacts on an extensive range of people (Lee, 2022). In environmental science, they could encourage wide-reaching policy change at a global scale. These far-reaching impacts defy the boundaries of scholarly debate.

By labeling a result as "statistically significant," we are operating under a decision made in uncertainty - choosing to assume that an effect is there and proceeding as if everything is right. On the flip side, we frequently fail to consider something worth pursuing when there is no significance and end up discarding what could be a treasure trove of valuable insights. This particular approach to significance testing leads to the crafting of a false dichotomy that fails to capture the refined, probabilistic essence of scientific knowledge.

The significant testing of scientific theories is one of the many sources of contention in philosophy. Deborah Mayo comes up with the severity requirement of a scientific test and explains it with focus on the tension. To Mayo and Hand (2022), the critical issue is how results exceed predefined bounds - whether they cross some threshold or whether the criteria give a high confidence test to the hypotheses. In its fullest term, a result acquires importance only when it adjudicates procedures that would almost certainly expose blunders, if there are any, for the methods used.

There are several ways in which primary issues can be approached simultaneously. The shocking reality is a gap in science, whereby publication bias precisely and negatively influences non-significant results leaving only positive results available in publications. Where there are low prior assumptions, the mixture of weak hypotheses with underpowered studies yields a literature brimming with false-positives snagged by undetectable stronger built obstacles. The "p-hacking" practice where numerous data analyzing schemas based on perceived significance yield multi sourced data fracturing data accuracy, credibility and reliability (Mertler et al, 2021) adds insult to this injury.

These challenges have caused some to reconsider whether abandoning statistical significance testing completely is more beneficial. A Nature Magazine comment in 2019 by over 800 scientists suggested replacing the term statistical significance with less rigid frameworks for reporting. In contrast, Mayo and Hand (2022) argue that the issue is not with significance testing, but rather with its misuse and misinterpretation. In their words, the answer is not abandonment, but rather a need for more sophisticated approaches to significance on the matter—ones that appreciate the reasoning underpinning scientific inquiry while simultaneously acknowledging its limitations.

The history of statistical methodologies is in search of more flexible perspectives beyond the dichotomy of 'significant/non-significant'. One option lies in Bayesian, which assumes prior knowledge and updates beliefs accordingly to new information or evidence. Furthermore, p-values are often insufficient on their own. For this reason, some may prefer the additional context effect sizes and confidence intervals give concerning the precision and magnitude of the findings.

Muff et al. (2021) suggest a shift in focus for stating scientific results from binary significance to the strength of evidence for different hypotheses: "language of evidence." By doing so, it recognizes that scientific knowledge accumulates gradually through the weight of

evidence rather than through singular, definitive tests.

More than just the methods themselves, the issues that come with statistical significance are tied to human nature. A search for certainty amidst complexity tends to be fulfilling. The appeal of $p < 0.05$ threshold partially lies in its promotion of simplicity that guarantees a logical conclusion to whether a given effect is 'real' but is misleading at the same time. The elegance of scientific discovery cannot be summed up using a singular figure or a binary answer.

The need to publish significant results leads to changes in the practice of receiving scientific funding which creates perverse incentives. Instead of focusing on making novel scientific breakthroughs for the advancement of one's career, researchers focus more on positive outcomes as they stand to gain attention. This focus leads to missing out on studies that are imperative to set a firm foundation of knowledge. Addressing the issue of significance testing requires more than a single methodical alteration. It calls for a shift in the culture, systems, and policies that govern rewards and recognition provided for scientific endeavors.

Significant exploration of statistical significance shows the conflicts that persist between precision and practicality. Even unique engineers in the field of science, like Fisher, understood that significance testing only offered value in giving based evidence as judgment. Unfortunately, as society evolves, the more institutionalized and formalized science becomes, the more flexible guided policies become set in stone. Having a p value below 0.05 was desirable; however, the truth was placed in qualified notions that have no real evidence behind them.

Emphasizing statistical significance also prompts contemplation about what deserves focus in scientific research. Important insights that do not adhere to cross arbitrary thresholds that the studied focus on are neglected disregarded. There are small effects that are studied underpowered and demonstrate a lack of significance, yet there are practical phenomena that hold importance. On the other hand, if a statistically significant result is achieved, there is a possibility that little to no practical relevance exists – especially if the effect size is trivial or the finding is devoid of theoretical coherence.

Integrating statistical significance with deep knowledge and theoretical constructs makes the former more useful. Alone, numbers hold no significance without a coherent explanation in reference to the prevailing knowledge, practical ramifications, and theoretical foundations; giving a context makes all the difference. Statistically sound evidence along with solid theoretical explanation and practical application demonstrates the clearest paradigm shift science and technology has ever achieved.

The future development of statistical practices does not lie in neglecting significance but deepening it unearthing supplementary measures like enhancement p-values with effect sizes, confidence intervals, and a multitude of other metrics that offer a comprehensive evaluation of evidence. Open-science practices, such as the sharing of data, preregistration, and clear reporting, assist in situating statistical outcomes within their appropriate frameworks and enhancing their trustworthiness. An advancement from meaning tests in solitude to meaning tests in compilation provides a stronger pedestal for supporting scientific knowledge.

Having educational interventions is fundamental when it comes to addressing the problems posed by statistical significance. Many researchers undergo scant training in statistics, with most being trained how to use frameworks as opposed to why those frameworks were created and what their boundaries are. There is a greater need to advance statistical literacy among all career stages of a scientist to have sophisticated approaches to evidence and inference.

The history of statistical significance from what was previously considered by Fisher as a provisional guideline to the present reform movements captures the essence of the dynamism associated with scientific methodology. The advancement of science is not limited to coming up with new ideas, but also in coming up with strategies designed to better filter a signal from noise. Rethinking significance should not be viewed as a (crisis), but as an opportunity to address the contextual details that surround scientific evidence.

Navigating how much weight to give statistical significance brings forth deep considerations about knowledge—and the uncertainty that accompanies it. Complications arise when attempting to balance the skepticism that discards genuine breakthroughs with baseless trust that accepts false truths as reality. How do we create reliable knowledge systems while maintaining the uncertainty inherent in scientific reasoning? These questions may not have straightforward replies, but developing answers can build a more nuanced understanding of evidence and reasoning.

The possibility of any phenomenon, from the observations we make, tends to remain central to scientific discovery. Neither prematurely dismissing potential discoveries nor succumbing to mirages is easy. However, statistical significance helps us carry this weight responsibly. Every advancement made carving out an understanding of significance strengthens established scientific knowledge and improves one's ability to differentiate between true complexity and woven patterns.

References

Gulati, J. (2025, February 21). Understanding Statistical Significance. Statology. https://www.statology.org/understanding-statistical-significance/

Lee, S. W. (2022). Methods for testing statistical differences between groups in medical research: statistical standard and

guideline of Life Cycle Committee. Life Cycle, 2(1). https://doi.org/10.54724/lc.2022.e1

Mayo, D. G., & Hand, D. (2022). Statistical significance and its critics: practicing damaging science, or damaging scientific practice? Synthese, 200(3). https://doi.org/10.1007/s11229-022-03692-0

Mertler, C. A., Vannatta, R. A., & LaVenia, K. N. (2021). Advanced and Multivariate Statistical Methods. https://doi.org/10.4324/9781003047223

Muff, S., Nilsen, E. B., O'Hara, R. B., & Nater, C. R. (2021). Rewriting results sections in the language of evidence. Trends in Ecology & Evolution, 37(3). https://doi.org/10.1016/j.tree.2021.10.009

Tenny, S., & Abdelgawad, I. (2023, November 23). Statistical significance. National Library of Medicine; StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK459346/

Affiliations and Corresponding Information

Engr. Jonathan Edwin V. Remegia, PhD
Talisay Senior High School
Department of Education-Philippines