Review of: "Intersections of Statistical Significance and Substantive Significance: Pearson's Correlation Coefficients Under a Known True Null Hypothesis"

Alban Ramette[1]

1 University of Berne

Potential competing interests: No potential competing interests to declare.

This study provides a valuable contribution to the discourse surrounding the use of P values and associated effect sizes. Its strengths lie in its use of simulations and appropriate citations of the literature. However, some opportunities exist to further enhance its clarity and impact.

A) Addressing Frequentist Methodology and Effect Size Errors

I concur with the author that both statistical significance and effect size are crucial when interpreting research findings. The author effectively highlights the limitations of relying solely on effect size, particularly in the context of Pearson's correlation with small sample sizes. As demonstrated through simulations, large effect sizes can occur even when the null hypothesis is true, leading to potentially misleading conclusions if significance is ignored.

The author's statement that "no statistical theory predicts the percentage of effect size errors under a true null hypothesis" requires clarification. Frequentist theory predicts a 5% rate of false positive effect sizes at an alpha level of 5%. However, the magnitude of these effect sizes can fluctuate, even reaching high absolute values under a true null.

This phenomenon can be explained by examining the Pearson correlation formula:

$r = cov(x, y) / (\sigma x * \sigma y)$

Small sample sizes (n) influence both the covariance (cov(x, y)) and standard deviations (σx, σy). Smaller n values lead to increased sampling variability, unstable estimates, and limited information, potentially resulting in surprising combinations of p-values and correlation coefficients (r), especially in simulations.

While the author's statement about type 1 errors and effect size errors is accurate, a more accessible explanation for non-statisticians could focus on the formula itself. Highlighting how fluctuations in standard deviations due to small sample sizes can lead to unexpected results may be easier to grasp than the original phrasing.

B) Simplifying Pedagogical Approach and Addressing Underpowered Studies

To enhance the text's pedagogical value, consider reducing the number of tables. For instance, illustrating the concept with three sample sizes (e.g., n=10, 100, 1000) might suffice. Additionally, cautioning against interpreting results from severely underpowered studies like those with n=4 would be beneficial.

The author should explicitly state that unexpected results can arise from the interplay of small sample sizes, covariance, and variances, particularly in simulations. This could be deduced directly from the equations and illustrated with fewer examples, thus improving reader understanding.

In addition, the author should ensure that the data used in the simulation process meets the necessary assumptions (e.g., bivariate normality) before conducting statistical tests.

C) Clarifying Statistical Concepts: Central Limit Theorem vs. Law of Large Numbers

The author wrote "the overall mean correlation (Grand Mean) is very close to zero, which is consistent with the Law of Large Numbers (Moore et al., 2021, p. 345). The reference to the Law of Large Numbers in explaining why the grand mean correlation is close to zero should be replaced with the Central Limit Theorem. The latter, which describes the normality of the distribution of sample means, is more directly relevant to this phenomenon.

D) Addressing Assumptions and Limitations of Fisher's Z-Transformation

The author should acknowledge that Fisher's z-transformation assumes bivariate normality, which might not hold for small sample sizes like n=4. Mentioning this limitation would provide a more comprehensive understanding of the method's applicability.

E) Exploring Bayesian Approaches

Given the conclusion's inclination towards Bayesian methods, the author could include references for interested readers to explore this alternative framework for integrating diverse sources of information into insightful statistical analysis.