Review of: "Comments on “The Roles, Challenges, and Merits of the P Value” by Chén et al."

Nabil Semmar[1]

1 Independent researcher

Potential competing interests: No potential competing interests to declare.

The article invites the abandonment of concepts and aspects of statistics that are essential in sampling, estimation, and comparison tests. By reference to the manuscript content, I provide some comments that could help with the revision, clarification, and improvement of the article:

"Estimation statistics" and "Significance testing" are complementary steps for population description and comparison. Populations are described by estimated values (position, dispersion parameters), and then they are compared with each other by means of significance tests that involve the use of probability distributions, alpha risks, and p-values. Estimation involves statistical significance because sampling values (raw data) always show some distances from true (population) unknown ones, resulting in a variable probability of affiliation of a sample to a population (or populations between them). Such variability in affiliation probability is usefully indicated (in practice) by that of the p-value, leading to the evaluation and improvement of sampling quality.

The p-value represents the lowest alpha risk at which to conclude a significant difference between two population values (tested value and reference one). It provides key information on how much a distance (difference) between two values (treatment vs control) is relatively important. This attributes to the p-value a tail location in the probability distribution. The lower the p-value, the more significant the distance between the two population values. The p-value is associated with the alpha risk, which is an essential concept in all estimation procedures and significance tests. Suggesting the abandonment of the p-value involves the renunciation of information about the limit to which the alpha risk area can be shrunk; this is essential for bias reduction and precision gain in sampling theory.

It would have been better to present the manuscript with the comparative advantages and limits of different probabilistic concepts rather than strict criticism referring to a published article.

Caution! Some writing calls for clarification to avoid confusion and misinterpretation: "How much is high a distance or remoteness between two values?" can’t be directly given by their difference (centered value), which is expressed with the unit of raw data. Remoteness between two values is effectively evaluated by calculating how many times the standard deviation is repeated in such a distance, leading to the standardized value Z (or t). This shows a clear interpretability and essential usefulness of standardized values, on one hand. On the other hand, this shows the importance of going beyond the centering procedure to effectively evaluate how much the distance separating two values is significant (or not). A difference of 1 has not the same meaning (significance) in more and less homogeneous populations, respectively. This difference is easily reached in a less homogeneous population showing a higher standard deviation, whereas it will represent a more significant difference in the case where the population SD decreases. A higher Z value involves a more significant distance between the two values, and this involves a lower probability of the first population value to be affiliated to the second (or reference) population. The standardized Z-value represents an efficient tool for probabilistic comparison between unstandardized (physical unit) values.

A very small difference between two values can be significant in highly homogeneous populations characterized by very small standard deviation. The standardized value Z (or t) provides essential information on whether a distance (or difference) between two values is relatively important or not, taking into account the variation scale of the population (estimated by standard deviation or standard error).

In sampling theory, the estimation precision and p-value tend to increase and decrease, respectively, with the sample size N (e.g., standard error). Also, the sample size has a natural effect that becomes null when the sample size N equals that of the population N(pop). In several fields, data are not easily accessible (e.g., in clinical populations), leading to work with small size samples by considering sampling properties.

Improvement of the manuscript could be made by carrying out a simulation showing how p-values vary with NHST, probability distribution functions, and sample sizes.