EDITORIAL

Interpretation of *p*-value: The Correct Way!

Snehali Majumder[1], Harish Mallapura Maheshwarappa[2]

Abstract

The probability value (*p*-value) is used in hypothesis testing to assist in determining if the null hypothesis should be rejected. In a practical setting, the *p*-value helps to determine if an experiment is conducted and then compares the outcomes to what random chance may yield. In order to do it, researchers state a "null hypothesis" that they want to disapprove. Many researchers consider the *p*-value to be the essential summary of statistical analysis of their research data. Although it is undeniable that *p*-values are a very useful method for summarizing study results, it is also undeniable that *p*-values are frequently misused and misunderstood. Therefore p-value must be carefully interpreted based on the study design, sample size, comparability of study groups, and appropriateness of statistical tests. The statistically significant *p*-value should not be the sole criterion for accepting or rejecting the conclusions of any report or publication. Proper critical appreciation of research publications is a mandatory requirement before making clinical decisions based on them

Keywords: Interpretation, Null hypothesis, *p*-value.

Indian Journal of Respiratory Care (2023): 10.5005/jp-journals-11010-1026

"It's science's dirtiest secret—the "scientific method" of testing hypotheses by statistical analysis stands on a flimsy foundation." ScienceNews (Siegfried 2010)

The *p*-value has been the preferred way for summarizing the outcomes of medical research data since Pearson introduced them in 1900.[1,2] *p*-values are determined as the difference between the observed value and a specified reference value[3] The bigger the difference between the two values, the smaller the corresponding *p*-value.[2] Manually, *p*-values are calculated using the *t*-distribution table with *n*–1 degrees of freedom.[3]

Box 1: Formula for manually calculating the *p*-value

H0: $\mu = A$
Ha: $\mu \neq A$
The formula for the test statistic is:
$t = (x-\mu)/(s/\sqrt{n})$ where x is the sample mean, μ is the hypothesized mean, s is the sample standard deviation, and n is the sample size.

Assuming that the null hypothesis is true, a *p*-value measures the probability of obtaining the observed results. For example, let's hypothesize that the average height of 10-year-old boys is 140 cm. To prove the hypothesis, we will measure the height of 100 random 10-year-old boys and calculate the sample mean (x) and standard deviation (s). Therefore, our sample size (n) = 100, n–1 degree of freedom = 100–1 = 99. Our hypothesized mean (μ) = 140 cm.

The null (0) and alternative hypothesis (a) for the test is H0: $\mu = 140$ cm and Ha: $\mu \neq 140$ cm.

Many researchers consider the *p*-value to be the essential summary of statistical analysis of their research data. Although it is undeniable that *p*-values are a very useful method for summarizing study results, it is also undeniable that *p*-values are frequently misused and misunderstood. Many authors or readers regard *p*-values of 0.05 as the "gold standard" of "significance." A *p*-value > 0.05 is considered to be of "no importance" or a "not significant" result. This, however, is not the actual scenario.[3,4] In clinical research, reproducibility plays a very important role in establishing confidence in a result. Even when the same biological sample is examined twice under similar conditions, two different observations are recorded [e.g. blood pressure (BP) reading of a person at two different time points]. The two results will always display a natural degree of variance. Determining whether the claimed difference is "background noise" or a real shift, such as one brought on by an intervention, is the main challenge (e.g. BP medication, doing any form of physical activity before measuring BP).[3,5] The *p*-value does not express the magnitude or relative importance of an impact. In relation to these limitations of the *p*-value and its injudicious use by researchers, the term "*p*-hacking" was coined by Simonsohn et al. of the University of Pennsylvania. Simonsohn et al. define "*p*-hacking" as "trying multiple things until you get the desired result."[6] Based on repeated observations and objections from various statisticians

[1]Department of Clinical Research, Narayana Hrudayalaya, Bengaluru, Karnataka, India

[2]Department of Critical Care, Mazumdar Shaw Cancer Center (MSCC), Narayana Health, Bengaluru, Karnataka, India

Corresponding Author: Harish Mallapura Maheshwarappa, Department of Critical Care, Mazumdar Shaw Cancer Center (MSCC), Narayana Health, Bengaluru, Karnataka, India, Phone: +91 8095218493, e-mail: dr.harishmm@rocketmail.com

How to cite this article: Majumder S, Maheshwarappa HM. Interpretation of *p*-value: The Correct Way! Indian J Respir Care 2023;12(1):1–2.

Source of support: Nil

Conflict of interest: Dr Harish Mallapura Maheshwarappa is associated as the National Editorial Board member of this journal and this manuscript was subjected to this journal's standard review procedures, with this peer review handled independently of this editorial board member and his research group.

about the inappropriate use of the *p*-value, the American Statistical Association has developed six principles regarding the correct interpretation and proper use of *p*-values. Many statisticians also recommend using Baye's rule. Others advocate for a more inclusive strategy that encourages researchers to test several methodologies on the same data set.[9]

Box 2: Salient features of the *p*-value

- A *p*-value is focused on the null hypothesis, rather than the study hypothesis.[2] *p*-value cannot measure the compatibility of the data with a study hypothesis.[7,8]
- A *p*-value cannot assess the likelihood that the investigated hypothesis is correct or that the data were generated only by chance.[3,5,7,8]
- A *p*-value has a tendency to draw focus away from the quantitative measurement of "impact size" and toward a more qualitative assessment of whether an effect existed or not. For example *p*-value does not tell us how two study groups are different from each other (control vs intervention group), it only tells us there is a difference between the two groups based on the null hypothesis selected by the researchers.[7,8]
- Statistical significance and scientific significance are not synonymous.[2] *p*-value is directly dependent on sample size and therefore calculating the correct sample size is important in order to draw any inferences using the *p*-value.[2,7,8]

Therefore *p*-value must be carefully interpreted based on the study design, sample size, comparability of study groups, and appropriateness of statistical tests.[5] Statistically significant *p*-value should not be the sole criteria for accepting or rejecting the conclusions of any search report or publication. Proper critical appreciation of research publications is a mandatory requirement before making clinical decisions based on them.[5]

Orcid

Snehali Majumder https://orcid.org/0000-0002-2745-493X
Harish Mallapura Maheshwarappa https://orcid.org/0000-0001-9377-3421

References

1. Pearson K X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. London, Edinburgh, Dublin Philos Mag J Sci 1900;50(302):157–175. DOI: 10.1080/14786440009463897
2. Nahm FS. What the P values really tell us. Korean J Pain 2017;30(4): 241–242. DOI: 10.3344/kjp.2017.30.4.241
3. Dahiru T. P-value, a true test of statistical significance? A cautionary note. Ann Ib Postgrad Med 2008;6(1):21–26. DOI: 10.4314/aipm.v6i1.64038
4. Ioannidis JPA. Why most published research findings are false. PLOS Med 2005;2(8):e124. DOI: 10.1371/journal.pmed.0020124
5. Dick F, Tevaearai H. Significance and limitations of the p value. Eur J Vasc Endovasc Surg 2015;50(6):815. DOI: 10.1016/j.ejvs.2015.07.026
6. Simonsohn U, Nelson LD, Simmons JP. P-curve: a key to the file drawer. J Exp Psychol Gen 2014;143(2):534–547. DOI: 10.1037/a0033242
7. The ASA Statement on P-Values: Context, Process, and Purpose. Published 2020. https://www.crossfit.com/health/the-asa-statement-on-p-values-context-process-and-purpose#:~:text=The ASA defines the p,extreme than its observed value.
8. Wasserstein RL, Lazar NA. The ASA statement on p-values: context, process, and purpose. Am Stat 2016;70(2):129–133. DOI: 10.1080/00031305.2016.1154108
9. Nuzzo R. Scientific method: statistical errors. Nature 2014;506(7487): 150–152. DOI: 10.1038/506150a