



P-value: What is and what is not

Kiarash Tanha¹, Neda Mohammadi², Leila Janani^{1*}

Received: 6 Mar 2017

Published: 25 Sep 2017

Abstract

The misinterpretation and misuse of p-value have been increasing for decades. In March 2016, the American Statistical Association released a statement to warn about the use and interpretation of p-value. In this study, we provided a definition and discussion of p-value and emphasized the importance of its accurate interpretation.

Keywords: Statistical significance, Hypothesis testing, P-value

Copyright© Iran University of Medical Sciences

Cite this article as: Tanha K, Mohammadi N, Janani L. P-value: What is and what is not. *Med J Islam Repub Iran.* 2017 (25 Sep);31:65. <https://doi.org/10.14196/mjiri.31.65>

Introduction

Statistical significance and p-value have long been recognized and are highly popular in scientific researches, but misuse and interpretation remain to be common (1). The idea of testing the significance and concept of p-values were developed by Ronald Fisher in 1920 in the context of research on crop variance (2). He described p-value as an index to measure discrepancy between the data and the null hypothesis. In the recent years, the dignity of statistics has been improved dramatically among the academic researches, however, p-values are still commonly misunderstood (1). Several recent surveys have revealed that many researchers in medicine (in different fields) lack knowledge in biostatistics and in the interpretation of statistical concepts (3-5). The American Statistical Association (ASA) warns about improper use and interpretation of the p-value on March 8th, 2016 (6); its Statement consists of 6 principles:

“(1) P-value can indicate how incompatible the data are with a specified statistical model.

(2) P-value neither measures the probability that the studied hypothesis is true nor the probability that the data were produced by random chance alone.

(3) Scientific conclusions and business or policy decisions should not be based only on whether a p-value pass-

es a specific threshold.

(4) Proper inference requires full reporting and transparency.

(5) A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.

(6) By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.”

Thus, the present study aimed at focusing on definition, interpretation, misuse, and overall challenges and notes, which should be considered when using p-values.

What is and what is not a p-value?

One of the best conceptual definitions of p-value such as the ASA uses is as follows: “The probability under a specified statistical model that a statistical summary of the data (eg, the sample mean difference between the 2 compared groups) would be equal to or more extreme than its observed value.”(6).

P-value is neither the probability of the hypothesis being tested nor the probability that the observed deviation was produced by chance alone. These are the most common misinterpretations of the p-value. In computing the p-value, it is assumed that the null hypothesis is true, so the p-value cannot indicate the probability that the null hypothesis is true. Another assumption that has been used in

Corresponding author: Dr Leila Janani, leila_janani@yahoo.com

¹ Department of Biostatistics, School of Public Health, Iran University of Medical Sciences, Tehran, Iran.

² Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran.

Table 1. Artificial case-control study

	Study A		Study B	
	Exposed	Unexposed	Exposed	Unexposed
Cases	8	2	800	200
Controls	7	3	700	300

Note: Hypothetical data

computing the p-value is that any deviation of the observed data from the null hypothesis was produced by chance, so it is clear that when only chance affects the deviation of the null hypothesis in the calculation of the p-value, it cannot be the probability of operating of the chance (7, 8).

The p-value is the probability of the observed data given that the null hypothesis is true, which is a probability that measures the consistency between the data and the hypothesis being tested if, and only if, the statistical model used to compute the p-value is correct (9). The smaller the p-value the greater the discrepancy: "If p is between 0.1 and 0.9, there is certainly no reason to suspect the hypothesis tested, but if it is below 0.02, it strongly indicates that the hypothesis fails to account for the entire facts. We should not be off-track if we draw a conventional line at 0.05". (2).

P-values have often been claimed or taken to imply the presence or absence of an effect or the importance of a result, which is certainly NOT true. In a clinical trial study in which the differences of an index are evaluated before and after an intervention and the mean difference is small (for example, 0.5 unit and p-value = 0.03), the p-value just implies the statistical significance (of an effect or correlation presence) and does not explain the importance of an effect or result. Thus, statistical significance is not equivalent to clinical significance and vice versa (10).

P-value does not measure the size of an effect. Suppose in 2 artificial case-control studies, the relationship between disease and exposure was examined by odds ratio. The data in Table 1 demonstrates that ratio of being exposed or non-exposed; and as a result, the odds ratio (OR = 1.71) is the same across the studies and the only difference between them is the sample size. However, the p-value for testing the hypothesis that the true odds ratio is equal to 1 is 0.605 for study A and less than 0.0001 for study B. Apparently, the same effect can produce very different p-values. As it is clearly mentioned in ASA statement, any effect (large or small) can produce a small or large p-value depending on the sample size or measurement precision, so conclusion about a hypothesis should not be based on p-value only, but also other aspects such as measurement precision, sample size, study design, and assumptions should be taken into account (6). For example, in cross-over design and parallel design, we may have the same effect but a different p-value, or the same effect will have different p-values if the precision of the estimates differs.

Other incorrect interpretations of the p-value are as follow: (11)

- If p-value = 0.2, there is a 20% chance that the null hypothesis is correct.
- Same p-values should have the same observed effects.

- P-value = 0.02 means that the probability of a type I error is 2%.

- P-value is a statistical index and has its own strengths and weaknesses, which should be considered to avoid its misuse and misinterpretation(12). Reporting the descriptive statistics, using confidence intervals of the measurement indexes alongside the p-value, and its true interpretation are of paramount importance (1).

Conclusion

A good report contains describing the data by suitable numerical and graphical summaries of data, dominance on the setting of the study, and logical and clinical interpretation of quantitative indexes. As a whole, summarizing statistical comparisons to statistical significance or non-significance is one of the highly popular statistical misinterpretations of p-values and hypothesis testing. The p-value like other indexes should be used and interpreted appropriately and should not be a scientific reason just by itself.

Conflict of Interests

The authors declare that they have no competing interests.

References

1. Sterne JAC, Smith GD. Sifting the evidence—what's wrong with significance tests? *BMJ*. 2001;322(7280):226-31.
2. Fisher RA. The Arrangement of Field Experiments. *Journal of the Ministry of Agriculture of Great Britain*. 1926;33:503-13.
3. Best AM, Laskin DM. Oral and maxillofacial surgery residents have poor understanding of biostatistics. *J. Oral Maxillofac. Surg.* 2013;71(1):227-34.
4. Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics and results in the medical literature. *Jama*. 2007;298(9):1010-22.
5. Bookstaver PB, Miller AD, Felder TM, Tice DL, Norris LB, Sutton SS. Assessing pharmacy residents' knowledge of biostatistics and research study design. *Ann. Pharmacother.* 2012;46(7-8):991-9.
6. Wasserstein RL, Lazar NA. The ASA's Statement on p-Values: Context, Process, and Purpose. *Am. Stat.* 2016;70(2):129-33.
7. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur. J. Epidemiol.* 2016;31:337-50.
8. Sterne JA. Teaching hypothesis tests-time for significant change? *SIM*. 2002;21(7):985-94; discussion 95-99, 1001.
9. Gigerenzer G. Mindless statistics. *J. Soc. Psychol.* 2004;33(5):587-606.
10. Charles GSP. Problems in common interpretations of statistics in scientific articles, expert reports, and testimony. *Jurimetrics*. 2011;51(2):113.
11. Goodman S. A dirty dozen: twelve p-value misconceptions. *Semin Hematol.* 2008;45(3):135-40.
12. Baker M. Statisticians issue warning over misuse of P values. *Nature*. 2016;531(7593):151.