Proceedings of the 4th International Conference on Statistics: Theory and Applications (ICSTA'22)
Prague, Czech Republic – July 28- 30, 2022
Paper No. 125
DOI: 10.11159/icsta22.125

Common Misconceptions and Misunderstandings in Magic Cut-Off for Significance: P-Value

Arzu Baygül Eden[1], Neslihan Gokmen Inan[2]
[1]Koc University, School of Medicine / Department of Biostatistics
Istanbul, Turkey
abaygul@ku.edu.tr; ninan@ku.edu.tr
[2]Koc University, School of Engineering / Department of Computer Engineering
Istanbul, Turkey

Abstract **-** The p-value is the most commonly used statistical term to make a decision in medical statistics. It helps the researchers to decide whether the results (which are obtained from a hypothesis test) show statistical significance or not. Most researchers up today prove their decisions based in p-values. Although the p-value is used so commonly by the researchers, they are in high rate misunderstood, miss-or over-interpreted and also wrongly reported. The purpose of this study is to emphasize some of the misconceptions about p-value, and correct their misunderstanding and suggest to researchers a straight way to use p-value. Fisher, who is called as "father of statistics" were not actually the first one, who used the p-value, however he was the first to outline formally the logic behind its use. Fisher's defined for the p value as we use today: it is equal to the probability of a given experimental observation, under a null hypothesis. If this number were smaller than the acceptable threshold, researchers could "reject" the null hypothesis as unlikely to be true. The use of a threshold p value as a basis for rejection was called a "significance test." This is important to distinguish from the "hypothesis test," which will be discussed shortly. In our study we stated some of the misconceptions such as obtaining p-values from inappropriate statistical methods, p-values <0.05 shows clinical significance, using always two-sided p-values, a scientific conclusion is always based on "1" p-value, p-value is the probability of null hypothesis is true, format for table of p-values. As a result, some of the common misconceptions are highlighted about p-value. Being aware of these misconceptions, can increase of the quality of studies.
Keywords**:** p-value, misconceptions of obtaining p-values, appropriate statistical methods, significant test

Introduction

Physicians or in general medicine residents should be update on literature. Knowledge of biostatistics make the researchers aware of the methods in interpreting and understanding. In our experience there is a lack of biostatistics knowledge, which causes miscalculations and also misinterpretations. Especially despite use of advance statistical methods in contemporary medical literature, basic concepts like choosing the hypothesis tests or interpretation of the results are not well understood [1]. The p-value is the most commonly used statistical term to make a decision in medical statistics. It helps the researchers to decide whether the results (which are obtained from a hypothesis test) show statistical significance or not.

Most researchers up today prove their decisions based in p-values. Although the p-value is used so commonly by the researchers, they are in high rate misunderstood, miss-or over-interpreted [2, 3] and also wrongly reported.

Researchers have been working on misconceptions including p-value at least 70 years [3-8]. Despite all these warnings in articles the misconceptions and misunderstandings are still done commonly by the researchers.

The purpose of this study is to emphasize some of the misconceptions about p-value, and correct their misunderstanding and suggest to researchers a straight way to use p-value.

Definition of p-value: The p-value is referred to calculated probability of having an error when rejecting H_0 hypothesis. The null hypothesis is a hypothesis of no difference between groups. Type I error is a pre-chosen probability of rejecting the null hypothesis when it is actually true. Having a p-value under Type I error (cut-off value which is mostly 0.05 in medical research) means that we can safely reject null hypothesis and the mistake of having rejected null hypothesis is at acceptable level (less than 5 percent).

Fisher, who is called as "father of statistics" were not actually the first one, who used the p-value, however he was the first to outline formally the logic behind its use [9]. Fisher defined the p-value as we use today: it is equal to the probability of a given experimental observation, under a null hypothesis. If this number was smaller than the acceptable threshold, researchers could "reject" the null hypothesis as unlikely to be true. The use of a threshold p-value as a basis for rejection

was called a "significance test." This is important to distinguish from the "hypothesis test," which will be discussed shortly [9].

In our study we stated some of the misconceptions such as:

Misconception 1: Obtaining p-values from inappropriate statistical methods

Choosing the hypothesis test depends on many factors, such as number of groups, sample sizes, distribution of continuous parameters (normally distributed or not), type of variables (continuous, categorical, ordinal, nominal data) and dependency. In case of choosing inappropriate method, researcher can produce incorrect p-values. The relation can be misinterpreted, for instance the p-value for Chi-Square testing can be lower than 0.05, while Fisher Exact test would be a better option for the study and the p-value provided from Fisher Exact test is higher than 0.05.

Suggestion to Researchers:

Using a package programme and obtaining a p-value as a result does not mean that you performed the accurate statistical method. While beginning the analysis, there are several things to question:

1) What are the assumptions for this statistical method?
2) How can we test the assumptions?
3) Are these methods suitable for hypothesis?

All of these concerns can be provided by biostatistics knowledge. Researchers should always use the statistical package programmes like a calculator.

Misconception 2: p-value <0.05 shows clinical significance

What we get from the p-value as the information, is the statistically significance. Mostly and by mistake p-values <0.05 is taken as clinically important [4]. This cannot be always true. Slight little differences can display statistically significance, whereas this difference is not enough to have clinically importance.

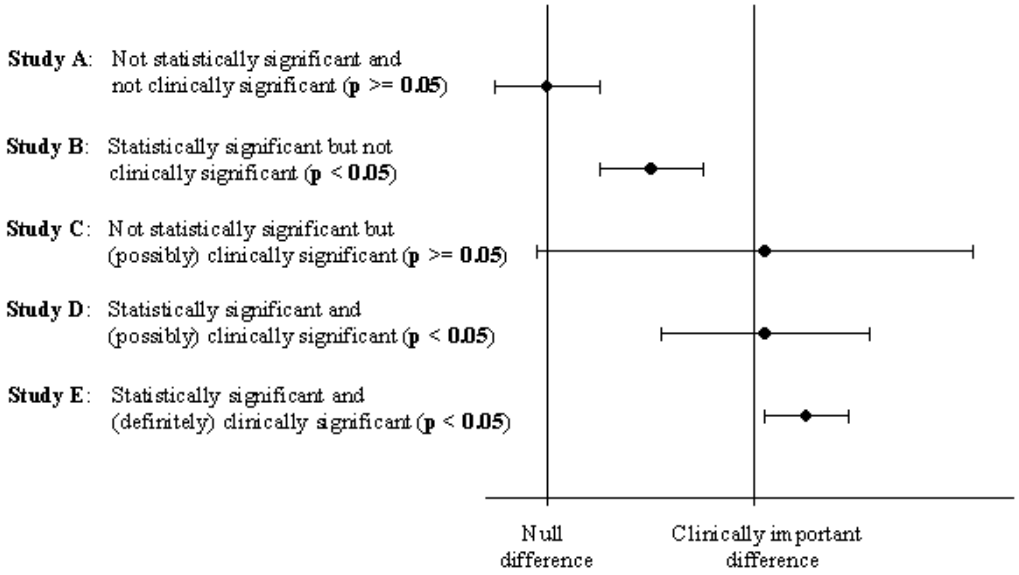

Figure 1. Statistical vs. clinical significance [15]

Suggestion to Researchers:

The statisticians should avoid interpreting the results strictly on a medical point of view. Inspecting confidence interval can be useful. Sometimes having statistically significant difference means nothing to the physician. Statisticians and researchers should always cooperate about the results of the study.

Misconception 3: Using always two-sided p-values.

One sided hypothesis allows us to find the probability of an effect in a specific direction, whereas two-sided hypothesis does not point to the direction. It only gives the result whether there is a statistically significant difference ignoring the direction.

Suggestion to Researchers

P-values should always be taken according to hypothesis' type (one sided or two sided). Interpretation of results should be done according to hypothesis, which is tested.

Misconception 4: A scientific conclusion is always based on "1" p-value.

Evidence based science always need more than one study to generalize the results.

Suggestion to Researchers

To continue the univariate analysis with a multivariate analysis as possible and also performing a meta-analysis by merging some other articles results would make the evidence stronger.

Misconception 5: p-value is the probability of null hypothesis is true.

p-value shows us the potential level of making a mistake when we rejecting the zero hypothesis. The simplest way to see that this is false is to note that the p-value is calculated under the assumption that the null hypothesis is true [4].

Suggestion to Researchers

While interpreting the p value, especially when it shows statistically significant difference, it would be a good idea,to comment by thinking of the potential error when rejecting null hypothesis.

Misconception 6: Format for table of p-values.

Like calculating of p-value, tabulating is also important. There are common mistakes in expressing p-values.

Suggestion to Researchers

p-values should be always stated with three decimals. If the researchers obtained p-values as 0.00 from a statistical package programme, they should state the p-value as <0.001.

Misconception 7: p-Hacking

Some researchers think that, when they could not prove their research hypothesis, the study will be worthless. In the contrary, in science philosophy the results obtained from well designed and well performed study, the results are important and valuable, even if they are positive or negative (expected or unexpected). The common mistake, done by the researchers is selection of desired p –values to report, which name is, p-hacking. P-hacking is a concept of misused statistical techniques.

Suggestion to Researchers

To avoid inflated effects and probabilities of false positive results pre-registration was suggested. [10-13] Preregistration force the researchers to publish the research hypothesis, data management plan, statistical analysis plan, predefined interim analysis (overall all steps from designing to final analysis) in public platforms such as Clinicaltrials.gov. This report disallows researchers to select wanted p-values.

Conclusion

We have highlighted some of the common misconceptions about p-value. We believe that, being aware of these misconceptions, can increase of the quality of studies.

Table 1. Misconceptions for p-Value

1. Obtaining p-values from inappropriate statistical methods
2. p-value <0.05 shows clinical significance
3. Using always two-sided p-values.
4. A scientific conclusion is always based on “1” p-value.
5. p-value is the probability of null hypothesis is true.
6. Format for table of p-values.
7. p-Hacking [14]

References

[1] Windish, D.M., S.J. Huot, and M.L. Green, *Medicine residents' understanding of the biostatistics and results in the medical literature.* Jama, 2007. **298**(9): p. 1010-22.

[2] Panagiotakos, D.B., *Value of p-value in biomedical research.* The open cardiovascular medicine journal, 2008. **2**: p. 97-99.

[3] Goodman, S.N., *Toward evidence-based medical statistics. 1: The P value fallacy.* Ann Intern Med, 1999. **130**(12): p. 995-1004.

[4] Goodman, S., *A dirty dozen: twelve p-value misconceptions.* Semin Hematol, 2008. **45**(3): p. 135-40.

[5] Edwards, W., H.R. Lindman, and L.J.J.P.R. Savage, *Bayesian statistical inference for psychological research.* 1963. **70**: p. 193-242.

[6] Feinstein, A.R., *Clinical biostatistics. XXXIV. The other side of 'statistical significance': alpha, beta, delta, and the calculation of sample size.* Clin Pharmacol Ther, 1975. **18**(4): p. 491-505.

[7] Diamond, G.A. and J.S. Forrester, *Clinical trials and statistical verdicts: probable grounds for appeal.* Ann Intern Med, 1983. **98**(3): p. 385-94.

[8] Berkson, J., *Tests of significance considered as evidence. Journal of the American Statistical Association 1942;37:325-35.* Int J Epidemiol, 2003. **32**(5): p. 687-91.

[9] Goodman, S.N., *p values, hypothesis tests, and likelihood: implications for epidemiology of a neglected historical debate.* Am J Epidemiol, 1993. **137**(5): p. 485-96; discussion 497-501.

[10] Wicherts, J.M., Veldkamp, C.L., Augusteijn, H. E. M., Bakker, M., Van Aert, R. C. M., Van Assen, M. A. L. M, *Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking.* Front Psychol, 2016. **7**: p. 1832.

[11] de Groot, A.D., *The meaning of "significance" for different types of research [translated and annotated by Eric-Jan Wagenmakers, Denny Borsboom, Josine Verhagen, Rogier Kievit, Marjan Bakker, Angelique Cramer, Dora Matzke, Don Mellenbergh, and Han L. J. van der Maas]. 1969.* Acta Psychol (Amst), 2014. **148**: p. 188-94.

[12] Wagenmakers, E.J., Wetzels, R., Borsboom, D., Van der Maas, H. L. J, Kievit, R. A., *An Agenda for Purely Confirmatory Research.* Perspect Psychol Sci, 2012. **7**(6): p. 632-8.

[13] Chambers, C.D., *Registered reports: a new publishing initiative at Cortex.* Cortex, 2013. **49**(3): p. 609-10.

[14] Makin, T.R. and J.J. Orban de Xivry, *Ten common statistical mistakes to watch out for when writing or reviewing a manuscript.* Elife, 2019. **8**.

[15] https://critical-numbers.group.shef.ac.uk/glossary/clinical_significance.html