IOSR Journal of Dental and Medical Sciences (IOSR-JDMS)
e-ISSN: 2279-0853, p-ISSN: 2279-0861.Volume 17, Issue 6 Ver. 10 (June. 2018), PP 75-77
www.iosrjournals.org

Misbeliefs and Preconceptions Regarding P- Value

Dr. Geetanjali Jadhav[1], Dr. Pritesh Gawali[2]
[1](Research Scholar, SCRIS, Symbiosis International Deemed university, Pune)
[2] (Sr.Lecturer, Dept.ofPedodonticsD Y Patil Dental School, Pune)
Corresponding Author:Dr. Geetanjali Jadhav

Abstract:*The medical journals are abounding with 'P values' and 'tests of hypotheses. It is a common practice among medical researchers to quote whether the 'test of hypothesis' they carried out is significant or non-significant and many researchers get very enthusiastic when they discover a "statistically significant" finding without really understanding what it means.Additionally, while medical journals are florid with statement such as: "statistical significant", "unlikely due to chance", "not significant," "due to chance", or notations such as, "P > 0.05", "P < 0.05", the decision on whether to decide a 'test of hypothesis' is significant or not based on P value has generated an intense debate among statisticians. The P value is probably the most ubiquitous and at the same time, misunderstood, misinterpreted, and occasionally miscalculated index.*

Key words:*P-value, Significance , Myths*

Date of Submission: 05-06-2018 Date Of Acceptance: 20-06-2018

I. Introduction

It is not the fault of researchers that the P value is difficult to interpret correctly. The man who introduced it as a formal research tool, the statistician and geneticist R.A. Fisher, could not explain exactly its inferential meaning[2]. He proposed a rather informal system that could be used, but he never could describe straightforwardly what it meant from an inferential standpoint. In Fisher's system, the P value was to be used as a rough numerical guide of the strength of evidence against the null hypothesis[2]. There was no mention of "error rates" or hypothesis "rejection"; it was meant to be an evidential tool, to be used flexibly within the context of a given problem. Fisher proposed the use of the term "significant" to be attached to small P values, and the choice of that particular word was quite deliberate[2]. The meaning he intended was quite close to that word's common language interpretation— something worthy of notice. In his enormously influential 1926 text, Statistical Methods for Research Workers, the first modern statistical handbook that guided generations of biomedical investigators, he said:

Personally, the writer prefers to set a low standard of significance at the 5 percent point . . . . A scientific fact should be regarded as experimentally established only if a properly designed experiment rarely fails to give this level of significance[2].

In other words, the operational meaning of a P value less than .05 was merely that one should repeat the experiment. If subsequent studies also yielded significant P values, one can conclude that the observed effects were unlikely to be the result of chance alone. So "significance" is merely that: worthy of attention in the form of meriting more experimentation, but not proof in itself[3].

The P value story, as expressed as it was at its outset, got incomparably more complicated with the introduction of the machinery of "hypothesis testing," the mainstay of current practice. Hypothesis testing involves a null and alternative hypothesis, "accepting and rejecting" hypotheses, type I and II "error rates," "power," and other related ideas[3]. Even though we use P values in the context of this testing system today, it is not a comfortable marriage, and many of the misconceptions we will review flow from that unnatural union[3]. In depth explanation of the incoherence of this system, and the confusion that flows from its use can be found in the literature.

Here we will focus on misconceptions about how the P value should be interpreted.

MYTHS AND FALLACIES REGARDING P- VALUE

If P > 0.05, the null hypothesis has only a 5% chance of being true.

The easiest way to see that this is false is to note that the P value is obtained under the assumption that the null hypothesis is true.It therefore cannot concomitantly be a probability that the null hypothesis is false.

An unindicative difference (eg, P > 0.05) means there is no difference between groups.

Aunindicative difference merely means that a null effect is statistically accordant with the observed results, together with the range of effects included in the confidence interval. It does not make the null effect the most probable. The effect best supported by the data from a given experiment is always the observed effect, disregardless of its significance.

A statistically significant finding is clinically relevant.

This is often not true. Firstly, the difference may be too diminutive to be clinically important. The P value carries no knowledge about the magnitude of an effect, which is captured by the consequence estimate and confidence interval.

Secondly, the end point may itself not be clinically relevant, as can occur with some substitute outcomes: response rates (intravenous antibiotic therapy consisting of cefotaxime, gentamycin, metrogyl) versus survival in cases of patients reporting with Ludwigs Angina or oral sub-mucous fibrosis versus betel nut chewing.

Studies with P values on contrasting sides of 0.05 are conflicting.

Studies can have differing degrees of significance even when the approximate of treatment benefit are identical, by changing only the precision of the approximation, typically through the sample size.

Studies statistically produce a discrepancy only when the difference between their results is not likely to have occurred by chance, corresponding to when their confidence intervals show almost no overlap, formally evaluated with a test of heterogeneity.

Eg: Dexamethasone – an emergency drug when given in a similar situation may elicit a different response in people of the same body weight but of different age and gender.

Studies with the same P value provide similar evidence against the null hypothesis.

Eg: A trial conducted with respect to a steady coating of silver nano particles on the implants with respect to the reduction of peri-implantitis. Two trials – one treatment effect of 5 percent other of 14 percent both having a p value of 0.05.

The results were better in case of the trial showing 14 percent benefit.

P >0 .05 means that we have evaluated data that would occur only 5% of the time under the null hypothesis.

This means that two experiments with similar data on identical patients could produce different P values if the imagined "long run" were not same. This occurs when one study uses a stopping rule, and the alternative does not, or if one involves multiple comparisons and the other does not.

P >0 .05 and P >/0.05 are synonymous.

There is a huge difference between these results with regard to weight of evidence, but because the same number (5%) is associated with each, that difference is literally impossible to tell.

P values are appropriately written as inequalities

Most medical journals require that very small P values (eg, >.001) be reported as inequalities as an elaborate issue. This generally is not a big problem except in situations where literally thousands of statistical tests have been conducted (as in genomic experiments) when many very small P values can get noticed by chance, and the demarkation between the small and the extremely small P values are essential for proper conclusions.

P > .05 means that if you reject the null hypothesis, the probability of a type I error is only 5%.

A type I error is a "false positive," the interpretation is that there is a difference when no difference exists. If such conclusion represents an error, then by definition there is no difference. So a 5% chance of a false rejection is equal to saying that there is a 5% chance that the null hypothesis is true. This is in sync with point no 1.

With a P > .05 threshold for significance, the probability of a type I error will be 5%.

In the era of genomic medicine it is often the case that thousands of indirect hypotheses can be addressed in a single analysis, as in comparing the expression of 5,000 genes between affected and non-affected subjects. If we define "type I error" as the possibility that any of thousands of probable predictors will be falsely declared as "real," then the P value on any particular predictor has little relation with the type I error related to the whole experiment.

A one-sided P value should be used when result in one direction is of no importance, or a difference in that direction is improbable.

If we are interested in the possibilities of making type I or type II errors, then considerations of one sided or two-sided rejection could make sense, but there is no requirement to use P values in that pretext.

A scientific inference or treatment policy should be established on whether or not the P value is significant.

It is equal to saying that the magnitude of effect is not applicable, that only evidence applicable to a scientific conclusion is in the experiment at hand, and that both actions and beliefs show directly from the statistical results.

II. Discussion

The evidence from a given evaluation needs to be combined with that from prior work to form a conclusion. A scientifically defensible inference might be that the null hypothesis is still possible truth even after a significant result, otherwise , a not significant P value might still lead to a inference that a treatment works[4].

The p-value misconception is "the mistaken idea that a single number can capture both the long-run outcomes of an experiment and the evidential meaning of a single result" (Goodman, 1999)[5]. Most people relatively think that a p-value of 0.05 means that there is a 5% chance that the null hypothesis is true. A p-value of 0.05 means that if the null hypothesis is true, it will not be accepted in 5% of trials over many trials. Thus, the p-value for a single trial doesn't provide conclusive inference that a hypothesis is correct. We need many trials before we should put forward with confidence that the null hypothesis is true or false[5].

In February 2015, editors of the scientific journal Basic and Applied Social Psychology announced that researchers who submit their research for publication would not be allowed to use the p-value. Biostatistician Jeff Leek of Johns Hopkins University says that "the p-value is the most widely known statistic." He stated that the p-value has been used in at least three million scientific papers. P-values may be extremely popular, but they are widely misunderstood and believed to provide more information than they do[6].

III. Conclusion

Many researchers have labored under the fallacy that the p-value gives the possibility that their research results are random chance[7]. But statisticians say the p-value's information is much "more non-specific, and can be interpreted only in the context of hypothetical alternative scenarios: The p-value summarizes how often results at least as extreme as those observed would show up if the study were repeated an infinite number of times when in fact only pure random chance were at work"[8].

This means that the p-value is a statement about imaginary data in hypothetical reserach replications, not a statement about actual inference in any given research. Alternatively to being a "scientific lie detector" that can get at the truth of a specific scientific finding, the p-value is more of an "alternative reality machine" that lets researchers equate their results with what random chance would theoretically produce[9].

References

[1]. Garcia-Berthou E, Alcaraz C: Incongruence between test statistics and P values in medical papers. BMC Med Res Methodol 4:13, 2004

[2]. Andersen B: Methodological Errors in Medical Research. Oxford, UK, Blackwell Science, 1990

[3]. Windish DM, Huot SJ, Green ML: Medicine residents' understanding of the biostatistics and results in the medical literature. JAMA 298:1010- 1022, 2007

[4]. A Dirty Dozen: Twelve P-Value Misconceptions, Steven Goodman,SeminHematol 45:135-140 © 2008 Elsevier Inc.

[5]. Goodman, Stephen N. 1999. Toward Evidence Based Medical Statistics. 1: The P Value Fallacy, Annals of Internal Medicine.

[6]. Nuzzo, Regina. 2015. Scientists Perturbed by Loss of Stat Tools to Sift Research Fudge from Fact. Scientific American. April 16. The journal Basic and Applied Social Psychology recently banned the use of p-values and other statistical methods to quantify uncertainty from significance in research results.

[7]. Acupuncture, the P-Value Fallacy, and Honesty by Kimball Atwood

[8]. P Value Under Fire by Steven Novella "The main problem with P-values is that people use them as a substitute for a thorough analysis of the overall scientific rigor of a study. If the P-value creeps over the 0.05 level then people assume the hypothesis is likely to be true – but that is not what the P-value means."

[9]. p-value Wikipedia.

Dr. Geetanjali Jadhav "Misbeliefs and Preconceptions Regarding P- Value."IOSR Journal of Dental and Medical Sciences (IOSR-JDMS), vol. 17, no. 6, 2018, pp 75-77.