

Statistical Evidence Use of p-value

Subjects: Statistics & Probability

Contributor: M. Ishaq Bhatti, Jae Kim

This paper addresses the use and misuse of p-value issue and review latest literature on the use of p-value in social and natural sciences.

Keywords: Use of p-values ; Statistics inference

1. Introduction

This short paper deals with problems of statistical inference and the use of p-values. Recently, the issue of the use of p-values in various scientific investigations and data analytics techniques has raised questions regarding the validity of statistical decision making in social sciences including business and economics disciplines. It is common practice among practitioners and researchers to make statistical decisions exclusively by using the “p-value < 0.05” criterion, regardless of sample size, statistical power and/or expected loss function underlying the selected models. Some of the well-known scholars have raised serious concerns about this practice and have warned that the use of “p-value” may lead to wrong decisions and give distorted scientific results. A few studies have commented on this issue by presenting empirical evidence, such as the paper by Keuzenkamp and Magnus (1995) ^[1] and McCloskey and Ziliak (1996) for economics, Fazal et al. (2020) for energy, Kim et al. (2018) for accounting, and Kim and Choi (2017) for finance, and among others. Due to the importance of the p-value issues some special issues are being edited with interesting topics – see for example Kim and Bhatti (2020).

2. Questions and Methods

The problem has become more challenging with increasing availability of large data sets. In particular, it is widely recognized that statistical significance (based on the conventional p-value criterion) is becoming irrelevant for big data (see, for example, Gandomi & Haider, 2015). To this end, Rao and Lovric (2016) maintain that ‘the 21st century researchers work towards a “paradigm shift” in testing statistical hypothesis’. There are calls that the researchers conduct more extensive exploratory data analysis before inferential statistics are considered for decision-making (see, for example, Leek and Peng, 2015; Soyer and Hogarth 2012). There are even calls that the use of statistical significance based on the p-value criterion should be abandoned (Wasserstein et al., 2019). Considering these criticisms, a new approach of hypothesis testing such as estimation-based method (e.g. confidence interval), predictive inference, and equivalence testing with relevant application of adaptive or optimal level of significance to business decisions are done. It employs decision-theoretic approach to hypothesis testing and its applications which compromise between the classical and Bayesian methods of hypothesis testing and relates to exploratory data analysis for large or massive data sets. To this end some critical review papers on the current practice of hypothesis testing and future directions in business are written by Richard Startz in a recent special issue of an econometrics journal. In Startz (2020) article a topic of, ‘Not p-Values, Said a Little Bit Differently’ which is an important contribution toward the ongoing discussion about the use and/or mis-use of p-values. Numerical examples are presented which demonstrate that a p-value can, as a practical matter, give you a different answer than the one that you want. Further most recent contributions to the topic is authored by Thomas Dyckman and Stephen A. Zeff on ‘Important Issues in Statistical Testing and Recommended Improvements in Accounting Research’^[2]. In this paper authors proposed improvements to both the quality and execution of research related to statistical inference in developing statistical tests which address the limitations in existing literature. They explore on the situational effects of “data carpentry”, alternatives to winsorizing and suggest necessary improvements instead of relying on a study’s calculated “p-values”. Another important paper titled, ‘Interval-Based Hypothesis Testing and Its Applications to Economics and Finance’ authored by Jae Kim and Andrew P. Robinson which tackled a long-standing literature review on interval-based hypothesis testing (such as tests for minimum-effect, equivalence, and non-inferiority) widely used in biostatistics, medical science, and psychology. It presents the methods in the contexts of a one-sample t-test and a test for linear restrictions in a regression. The paper employed testing for market efficiency, validity of asset-pricing models, and persistence of economic time series. Authors argue that, from the point of view of economics and finance, interval-

based hypothesis testing provides more sensible inferential outcomes than those based on point-null hypothesis. It proposes interval-based tests which can be routinely used in empirical research in business, as an alternative to point null hypothesis testing, especially in the new era of big data.

Another paper addressing a similar issue as of Kim and Robinson (2020) is written by David Trafimow ^[3], entitled 'A Frequentist Alternative to Significance Testing, p -Values, and Confidence Intervals'. In this article David begins his debate about null hypothesis significance testing, p -values without null hypothesis significance testing, and confidence intervals. The first major section addresses some of the main reasons these procedures are problematic and concludes that none of them are satisfactory. However, there is a new procedure, termed the a priori procedure (APP), that validly aids researchers in obtaining sample statistics that have acceptable probabilities of being close to their corresponding population parameters. The second major section provides a description and review of APP advances. Not only does the APP avoid the problems that plague other inferential statistical procedures, but it is easy to perform too. Although the APP can be performed in conjunction with other procedures, the present recommendation is that it be used alone. Moreover, Magnus (2021) addresses the issue of the use of t -ratios. The title of this paper is, 'On Using the t -Ratio as a Diagnostic', in which the author points out that a *tests* and *diagnostics* are the two uses of t -ratio in econometrics. The paper proposes a new pretesting method '*model averaging*' over t -ratio and pretest estimators.

One of the very important contributions in this literature is done by John Quiggin who adopted a microeconomic approach. Quiggin (2019) paper begins with the observation that the constrained maximisation central to model estimation and hypothesis testing may be interpreted as a kind of profit maximisation. The output of estimation is a model that maximises some measure of model fit, subject to costs that may be interpreted as the shadow price of constraints imposed on the model. The replication crisis may be regarded as a market failure in which the price of "significant" results is lower than would be socially optimal ^[4].

The issue on the pedagogy of econometrics is addressed in a paper entitled "Teaching Graduate (and Undergraduate) Econometrics: Some Sensible Shifts to Improve Efficiency, Effectiveness, and Usefulness" by Jeremy Arkes. Arkes (2021) pointed out that a statement by Wasserstein et al. (2019), "Statistics education will require major changes at all levels to move to a post ' $p < 0.05$ ' world". As educators, we will need to re-think the way we train the future decision-makers, especially in the big data era where the p -value criterion is no longer relevant. Jeremy proposes a range of critical points on the issue of teaching econometrics including the problem related with the p -value, maintaining that the teaching of graduate (and undergraduate) econometrics needs to be revamped.

Most of the above papers published during 2020 and 2021 are included in *Econometrics* journal **Special Issue** which is available on the link for ready reference to the readers and scholars who would like to be familiar with p -value issue. The journal has provided an open access-publishing facility to the contributors which is a realistic option for our discipline ^{[5][6][7][8][9][10][11][12][13][14][15][16][17][18][19][20][21][22][23][24]}.

References

1. Also, refer a quote in the American Statistical Association (Wasserstein and Lazar, 2016) and Presidential address given by American Finance Association (Harvey, 2017).
2. Refer Dyckman and Zeff (2020).
3. Trafimow (2020). 'A Frequentist Alternative to Significance Testing, p -Values, and Confidence Intervals', *Econometrics*
4. Quiggin (2019). 'The Replication Crisis as Market Failure', *Econometrics*, Special Issue : Towards a New Paradigm for Statistical Evidence (mdpi.com)
5. Arkes, J. (2020). Teaching Graduate (and Undergraduate) Econometrics: Some Sensible Shifts to Improve Efficiency, Effectiveness, and Usefulness. *Econometrics* 2020, 8, 36.
6. Dyckman, T.R.; Zeff, S.A. (2019). Important Issues in Statistical Testing and Recommended Improvements in Accounting Research. *Econometrics* 2019, 7, 18.
7. Fazal, R., Rehman, S. A. U., Rehman, A. U., Bhatti, M. I., & Hussain, A. (2020), 'Energy-Environment-Economy causal nexus in Pakistan: A Graph Theoretic Approach'. *Energy*, 118934.
8. Gandomi, A. & Haider, M. (2015), 'Beyond the hype: Big data concepts, methods, and analytics', *International journal of information management*, 35(2), 137–144.
9. Harvey, C. R. (2017), 'Presidential Address: The Scientific Outlook in Financial Economics', *Journal of Finance*, 72 (4), 1399-1440.

10. Kim J. H., Ji, P., Ahmed, K., (2018), 'Significance Testing in Accounting Research: A Critical Evaluation based on Evidence', *Abacus*, 54 (4), 524-546.
11. Kim, J. H., and Bhatti, M. I. (2020). Special Issue: Towards a New Paradigm for Statistical Evidence (mdpi.com).
12. Kim J. H. Choi, I. (2017), 'Unit Roots in Economic and Financial Time Series: A Re-evaluation at the Decision-based Significance Levels'. *Econometrics*, 5(3), 41, Special Issue "Celebrated Econometricians: Peter Phillips.
13. Kim, J.H.; Robinson, A.P. (2019). Interval-Based Hypothesis Testing and Its Applications to Economics and Finance. *Econometrics* 2019, 7, 21.
14. Keuzenkamp, H.A. and Magnus, J. (1995), 'On tests and significance in econometrics', *Journal of Econometrics*, 67 (1), 103–128.
15. Leek, J. T., & Peng, R. D. (2015), 'Statistics: P values are just the tip of the iceberg', *Nature*. 2015 Apr 30, 520-612 (7549): doi: 10.1038/520612a.
16. Magnus, J.R. On Using the t-Ratio as a Diagnostic. *Econometrics* 2019, 7, 24.
17. McCloskey, D. and Ziliak, S. (1996), 'The standard error of regressions', *Journal of Economic Literature*, 34, 97–114.
18. Quiggin, J. (2019). The Replication Crisis as Market Failure. *Econometrics* 2019, 7, 44.
19. Rao, C. R. and Lovric, M. M., (2016), 'Testing Point Null Hypothesis of a Normal Mean and the Truth: 21st Century Perspective', *Journal of Modern Applied Statistical Methods*, 15 (2), 2-21.
20. Soyer, E., Hogarth, R.M., (2012), 'The illusion of predictability: how regression statistics mislead experts', *International Journal of Forecasting*, 28, 695–711.
21. Startz, R. (2019). Not p-Values, Said a Little Bit Differently. *Econometrics* 2019, 7, 11
22. Trafimow, D. A Frequentist Alternative to Significance Testing, p-Values, and Confidence Intervals. *Econometrics* 2019, 7, 26.
23. Wasserstein, R. L. and N. A. Lazar (2016), 'The ASA's statement on p-values: Context, process, and purpose', *The American Statistician*, Vol. 70, No. 2, pp. 129-133.
24. Wasserstein, R. L., Schirm, A. L. & Lazar, N. A. (2019), 'Moving to a world beyond " $p < 0.05$ "', *The American Statistician*, 73(sub1), 1–19.

Retrieved from <https://encyclopedia.pub/entry/history/show/14063>