Justify your Alpha using Optimal Learning – Review

|  |  |
| --- | --- |
| Albers, C. J., Kiers, H. A., & van Ravenzwaaij, D. (2018). Credible Confidence: A pragmatic view on the frequentist vs Bayesian debate. *Collabra: Psychology*, *4*(1). | * Argues that in practice Bayesian and frequentist approaches do align very often |
| Arandjelovic (2019). A more principled use of the *p*-value? Not so fast: a critique of Colquhoun’s argument R. Soc. open sci (6) http://dx.doi.org/10.1098/rsos.181519 | * Against the continued use of p-values * Overall, of limited interest |
| Colquhoun D.(2014) An investigation of the false discovery rate and the misinterpretation of p-values. Royal Society Open Science 1(3):140216. doi: 10.1098/rsos.140216 | * Advocate three sigma rule * Classical FPR argument * Better ways of expressing uncertainties: Likelihood ratios and reverse Bayesian inference |
| Colquhoun, D. (2017). The reproducibility of research and the misinterpretation of p-values. *Royal society open science*, *4*(12), 171085. | * If you observe a “significant p-value after doing a single unbiased experiment, what is the probability that your result is a false positive? * Check supplementary materials for scripts on p-value calculation * False positive risks as a function of sample size * Explains alpha by confirmation case * Reverse Bayesian argument, nice idea. Maybe something we can refer to * Minimum false positive risk – wouldn’t it be more interesting to calculate the maximum false positive risk |
| Colquhoun, D. (2019) The false positive risk: a proposal concerning what to do about p values. American Statistician . | * Adopt some sort of compromise between frequentist and Bayesian approaches * What most users still think a p-value gives them * Various ways of doing this have been proposed, notably by Berger and Sellke (1987), Sellke et al. (2001), Goodman (1993, 1999a,b) and by Johnson (2013a,b). * Full Bayesian analysis will probably never be adopted as a routine substitute for *p*-values. It has a place in formal clinical trials that are guided by a professional statistician, but it is too complicated formost users, and experimentalists distrust (rightly, in my view) informative priors. As Valen Johnson said …subjective Bayesian testing procedures have not been -and will likely never be- generally accepted by the scientific community. (Johnson 2013a,b) |
| Comments, Conjectures, and Conclusions – Standardized Tail-Area Probabilites Good | * Given a value of p the corresponding Bayes factor against the null is proportional to N\*1/2 |
| Dienes, Z., Mclatchie, N. Four reasons to prefer Bayesian analyses over significance testing. *Psychon Bull Rev* **25,** 207–218 (2018). https://doi.org/10.3758/s13423-017-1266-z | * Five examples from the journal social psychology. Might be interesting for the article * Correspondingly, Jeffreys (1939, pp 323–325) discusses how in the research problems he has investigated, Fisher’s methods (i.e. significance testing) and his (using Bayes factors) generally agreed (and hence indicating that the respective conventions were roughly aligned). * Contrast evidential value of frequentist and Bayesian methods, indirectly critic on my approach |
| Dirnagl, U. (2019). The p value wars (again). | * Nothing of interest |
| Fiedler, K., Kutzner, F., & Krueger, J. I. (2012). The Long Way From α-Error Control to Validity Proper: Problems With a Short-Sighted False-Positive Debate. Perspectives on Psychological Science, 7(6), 661–669. <https://doi.org/10.1177/1745691612462587>+ | * False negatives constitute a more serious problem than false positives * Most critical arguments are concerned with the reduction of false positives or alpha error control * Context of Justification vs. Context of discovery from Reichenbaum -> is this citation actually in Lewandowski paper? |
| Field, S. A., Tyre, A. J., Jonzén, N., Rhodes, J. R., & Possingham, H. P. (2004). Minimizing the cost of environmental management decisions by optimizing statistical thresholds. *Ecology Letters*, *7*(8), 669-675. | * Similar to Nicolas Blog Post * Advocate Bayesian approach in the end |
| Good, I. J. (1992). The Bayes/Non-Bayes Compromise: A Brief Review. Journal of the American Statistical Association, 87(419), 597. <https://doi.org/10.2307/2290192> -> should have something on justifying alpha by decreasing sample size | * P-values vs. weight of evidence * Maybe idea: We cannot use p-values as weight of evidence, however, we could can use alpha levels as weight of evidence * Psychokinesis example -> prior probability very low => I can use Bem and precognition maybe * Alternative Corona vaccination * Show how crazy it is that we test these two kinds of things in the same way |
| Justify your Alpha by Decreasing Alpha Levels as a Function of Sample Size http://daniellakens.blogspot.com/2018/12/testing-whether-observed-data-should.html?m=1 |  |
| Justify Your Alpha by Minimizing or Balancing Error Rates | * Interesting references to Cohen and Neyman pearson * Minimize or balance error rates, computationally similar approach to me (although conceptually different) |
| Justify your alpha using decision theory |  |
| Lakens, D., & Evers, E. R. (2014). Sailing from the seas of chaos into the corridor of stability: Practical recommendations to increase the informational value of studies. *Perspectives on Psychological Science*, *9*(3), 278-292. | * V statistic, very interesting but not necessarily relevant for justifying alpha |
| Leamer, E. E. (1978). Specification Searches: Ad Hoc Inference with Nonexperimental Data (1 edition). New York usw.: Wiley. | * Idea of minimizing the expected loss * P.115: The existence of prior information about the parameters influences hypothesis testing in the sense that a hypothesis is to be judged at a priori likely values of the parameters as well as at those values favored by the data |
| Machery, E. (2019). The alpha war. *Review of Philosophy and Psychology*, 1-25. | * Defends original Benjamin et al proposal |
| Matthews, R. A. (2018). Beyond ‘significance’: principles and practice of the Analysis of Credibility. *Royal Society open science*, *5*(1), 171047. | * Also idea of deducing the prior from the posterior * Based on how posterior distributions are modified based on the prior |
| Miller J, Ulrich R (2019) The quest for an optimal alpha. PLoS ONE 14(1): e0208631. <https://doi.org/10.1371/journal.pone.0208631> | * We show that a simple statistical model can be used too explore the quantitative tradeoff between reducing false positives and increasing false negatives * Any given empirical study will produce one of four possible outcomes with the probabilities just described -> these vary across research area and can usually only be estimated subjectively        * Had actually been done already :D |
| Mudge, J. F., Baker, L. F., Edge, C. B., & Houlahan, J. E. (2012). Setting an Optimal α That Minimizes Errors in Null Hypothesis Significance Tests. PLOS ONE, 7(2), e32734. <https://doi.org/10.1371/journal.pone.0032734> | * Decisions about critical effect sizes should not be seen as an unnecessary step that can be avoided because that is what people implicitly do using alpha 5% also * When there are no strong reasons it is good to assign equal weights to type 1 and type 2 errors * Could reanalyze their example with my optimal alpha approach? |
| van Dongen, N. N., van Doorn, J. B., Gronau, Q. F., van Ravenzwaaij, D., Hoekstra, R., Haucke, M. N., ... & Gelman, A. (2019). Multiple perspectives on inference for two simple statistical scenarios. *The American Statistician*, *73*(sup1), 328-339. | * Very nice paper but not directly relevant for my approach |
| Wilson, B. M., & Wixted, J. T. (2018). The prior odds of testing a true effect in cognitive and social psychology. *Advances in Methods and Practices in Psychological Science*, *1*(2), 186-197. | * Established knowledge is the key consideration regarding prior odds (could also be seen as strong theory) * Use p-curves to calculate power * However, maybe we could do something similar with a z-curve to estimate prior odds of a true hypothesis, however, this would probably be another project * Can we point to Ulis z-curve studies? |
| Witt (2019). Insights into Criteria for Statistical Significance from Signal  Detection Analysis. *Meta-Psychology.* https://doi.org/10.15626/MP.2018.871 | * Comparing p-values and Bayes factors in their signal detection ability. Not necessarily relevant for my approach |