

19.4: Evidentiary Standards You Can Believe

If [p] is below .02 it is strongly indicated that the [null] hypothesis fails to account for the whole of the facts. We shall not often be astray if we draw a conventional line at .05 and consider that [smaller values of p] indicate a real discrepancy.

– Sir Ronald Fisher (1925)

Consider the quote above by Sir Ronald Fisher, one of the founders of what has become the orthodox approach to statistics. If anyone has ever been entitled to express an opinion about the intended function of p-values, it's Fisher. In this passage, taken from his classic guide *Statistical Methods for Research Workers*, he's pretty clear about what it means to reject a null hypothesis at $p < .05$. In his opinion, if we take $p < .05$ to mean there is "a real effect", then "we shall not often be astray". This view is hardly unusual: in my experience, most practitioners express views very similar to Fisher's. In essence, the $p < .05$ convention is assumed to represent a fairly stringent evidentiary standard.

Well, how true is that? One way to approach this question is to try to convert p-values to Bayes factors, and see how the two compare. It's not an easy thing to do because a p-value is a fundamentally different kind of calculation to a Bayes factor, and they don't measure the same thing. However, there have been some attempts to work out the relationship between the two, and it's somewhat surprising. For example, Johnson (2013) presents a pretty compelling case that (for t-tests at least) the $p < .05$ threshold corresponds roughly to a Bayes factor of somewhere between 3:1 and 5:1 in favour of the alternative. If that's right, then Fisher's claim is a bit of a stretch. Let's suppose that the null hypothesis is true about half the time (i.e., the prior probability of H_0 is 0.5), and we use those numbers to work out the posterior probability of the null hypothesis given that it has been rejected at $p < .05$. Using the data from Johnson (2013), we see that if you reject the null at $p < .05$, you'll be correct about 80% of the time. I don't know about you, but in my opinion an evidentiary standard that ensures you'll be wrong on 20% of your decisions isn't good enough. The fact remains that, quite contrary to Fisher's claim, if you reject at $p < .05$ you shall quite often go astray. It's not a very stringent evidentiary threshold at all.

This page titled [19.4: Evidentiary Standards You Can Believe](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Danielle Navarro](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- [17.4: Evidentiary Standards You Can Believe](#) by [Danielle Navarro](#) is licensed [CC BY-SA 4.0](#). Original source: <https://bookdown.org/ekothe/navarro26/>.