This is part 2 in a series about confidence intervals (here's part 1). Answering the question in the title is not really my goal, but simply to discuss confidence intervals and their pros and cons. The last post explained why frequency statistics (and confidence intervals) can't assign probabilities to one-time events, but always refer to a collective of long-run events.

If confidence intervals don't really tell us what we want to know, does that mean we should throw them in the dumpster along with our p-values? No, for a simple reason: In the long-run we will make less errors with confidence intervals (CIs) than we will with p. Eventually we may want to drop CIs for more nuanced inference, but for the time being we would do much better with this simple switch.

If we calculate CIs for every (confirmatory) experiment we ever run, roughly 95% of our CIs will hit the mark (i.e., contain the true population mean). Can we ever know which ones? Tragically, no. But some would feel pretty good about the process being used if it only has a 5% life-time error rate. One could achieve a lower error rate by stretching the intervals (to say, 99%) but that would leave them too embarrassingly wide for most.

If we use p we will be wrong 5% of the time in the long-run when we are testing a true null-hypothesis (i.e., no association between variables, or no difference between means, etc., and assuming the analysis is 100% pre-planned). But when we are testing a false null-hypothesis then we will be wrong roughly 40-50% of the time or more in the long-run (These are my experiments, now it's time to submit to Psych Science)

There is a limitation to this benefit of CIs, and this limitation is self-imposed. We cannot escape the monstrous error rates associated with p if we report CIs but then interpret them as if they are significance tests (i.e., reject if null value falls inside the interval). Switching to confidence intervals will
do nothing if we use them as a proxy for \( p \). So the question then becomes: Do people actually interpret CIs simply as a null-hypothesis significance test?.

References


