Understanding Replication: Confidence Intervals Much Better Than p Values

May 25, 2013

Geoff Cumming, La Trobe University, Australia, presents his research on "Understanding Replication: Confidence Intervals Much Better Than p Values," at the 25th APS Annual Convention.

Replication is at the heart of science. A current hot topic across medicine, psychological science, and other disciplines is that a number of widely-accepted published results cannot be replicated.

A major cause of the problem is reliance on null hypothesis significance testing (NHST). The imperative to achieve statistical significance, or getting a p value that is greater than .05, leads researchers to select data, variables, and analysis techniques, until they reach that goal. This results in spuriously significant results being published, but these results are seldom replicated because significance is taken to imply truth, and there are few incentives to invest effort in replication experiments.

These problems give one additional reason we should turn from NHST to better methods, especially estimation — meaning effect sizes, confidence intervals (CIs), and meta-analysis. I refer to these as 'The new statistics' — which are not new, but widespread use of them would for many researchers be very new, as well as an important advance.

This poster illustrates the very poor information that p values give about replication: A replication experiment is likely to give an extremely different p value. Researchers generally do not appreciate this, and severely underestimate the sampling variability of the p value. In contrast, a confidence interval gives useful information about replication — the mean of a replication is likely to fall within the original CI and researchers generally have a reasonable appreciation of how CIs tell us about replication.

The poster illustrates how CIs give much better information about replication than p values do. It highlights the need to shift to the new statistics to improve our research, in particular by understanding replication better. For a dramatic illustration of the variability of the p value, see the [dance of the p values](#). For more about the new statistics, see my book "[The New Statistics](#)."

-Geoff Cumming

La Trobe University, Australia

g.cumming@latrobe.edu.au