

Memo

To: Colleagues
From: Roger Bakeman
Date: May 22, 2021
Re: Why You Should Not Use the Bonferroni Correction

I call them zombie rules; you can't kill them. Requests for the Bonferroni correction to "correct" for multiple tests continue to pop up, among other places, in reviews of articles submitted for publication. Here is what a Google search reveals:

"Classicists argue that **correction** for multiple testing is mandatory. Epidemiologists or rationalists argue that the **Bonferroni adjustment** defies common sense and increases type II errors (the chance of false negatives)." Here is a reference:

Rothman, K. J. (1990). No adjustments are needed for multiple comparisons. *Epidemiology*, 1(1), 43–46. <http://www.jstor.org/stable/20065622>

The classic argument is that the more tests we make in the course of an analysis, the more likely we are to identify significant-by-chance effects. But which ones are these? Better, how can we keep our study-wise alpha level to .05 in the face of multiple tests.

A common recommendation is to apply the Bonferroni correction (Miller, 1966), adjusting the alpha level by the inverse of the number of tests. For example, if nine tests were performed, instead of asking which p values were less than .05, we would ask instead which were less than .0056 (.05/9).

However, as Cohen (1990) has noted, in practice the probability of Type I error is almost always zero because effect sizes, even when small, are almost never exactly zero. He argued that researchers are unlikely to solve our "multiple tests problem with the Bonferroni maneuver" (p. 1304); for one thing, applied zealously, almost nothing is ever significant (see also Wilkinson and the Task Force on Statistical Inference, 1999). Both Cohen and Wilkerson recommend that investigators interpret overall patterns of significant effects, not just individual ones; that they be guided by predictions ordered from most to least important; and, above all, that they focus on effect sizes. This advice has considerable merit.

Bakeman, R., & Quera, V. (2011). *Sequential analysis and observational methods for the behavioral sciences*. Cambridge, UK: Cambridge University Press. pp. 128–129.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45, 1304–1312.

Miller, R. G., Jr. (1966). *Simultaneous statistical inference*. New York: McGraw-Hill.

Wilkinson, L., and Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594–604.

Note. The test was named for Carlo Emilio Bonferroni (1892–1960), an Italian mathematician who worked on probability theory; he did not devise it.