



Failing Grade: 89% of Introduction-to-Psychology Textbooks That Define or Explain Statistical Significance Do So Incorrectly

Scott A. Cassidy, Ralitza Dimova, Benjamin Giguère, Jeffrey R. Spence, and David J. Stanley

Department of Psychology, University of Guelph

Abstract

Null-hypothesis significance testing (NHST) is commonly used in psychology; however, it is widely acknowledged that NHST is not well understood by either psychology professors or psychology students. In the current study, we investigated whether introduction-to-psychology textbooks accurately define and explain statistical significance. We examined 30 introductory-psychology textbooks, including the best-selling books from the United States and Canada, and found that 89% incorrectly defined or explained statistical significance. Incorrect definitions and explanations were most often consistent with the odds-against-chance fallacy. These results suggest that it is common for introduction-to-psychology students to be taught incorrect interpretations of statistical significance. We hope that our results will create awareness among authors of introductory-psychology books and provide the impetus for corrective action. To help with classroom instruction, we provide slides that correctly describe NHST and may be useful for introductory-psychology instructors.

Keywords

p values, null-hypothesis significance testing, statistical inference

Received 8/22/18; Revision accepted 5/21/19

Statistical (adjective): relating to the use of statistics (“Statistical,” n.d.)

Significance (noun): the quality of being worthy of attention; importance (“Significance,” n.d.)

Statistical significance (noun): Assuming that the null hypothesis is true and the study is repeated an infinite number times by drawing random samples from the same population(s), less than 5% of these results will be more extreme than the current result (based on Kline, 2013, p. 75).

It is difficult to argue that statistical significance is a simple or an intuitive idea. At face value, statistical significance seems straightforward because it combines two relatively common words to form a description.

However, a commonsense interpretation of statistical significance is misleading. As illustrated by the definitions just presented, the term *statistical significance* denotes much greater technical complexity than suggested by the aggregation of the respective definitions of *statistical* and *significance*.

Since its introduction nearly 90 years ago, null-hypothesis significance testing (NHST) has been the most widely used statistical approach to data analysis in psychology (Nickerson, 2000). Yet, despite its ubiquity, the history surrounding significance testing reveals that researchers misunderstand, misinterpret and misapply the technique with alarming regularity—a situation methodologists have long criticized and attempted to correct (e.g., Bakan, 1966; Carver, 1978; Cohen, 1994; Nickerson, 2000).

After close to a century of consistent corrections and explanations regarding how to interpret and use NHST correctly, incorrect interpretations and applications have proven to be hard to correct.