DIGITALCOMMONS
@WAYNESTATE

Journal of Modern Applied Statistical Methods

Volume 12 | Issue 1 Article 3

5-1-2013

p-Values versus Significance Levels

Phillip I. Good
Information Research, Huntington Beach, CA

Recommended Citation
Good, Phillip I. (2013) "*p*-Values versus Significance Levels," *Journal of Modern Applied Statistical Methods*: Vol. 12 : Iss. 1 , Article 3.
DOI: 10.22237/jmasm/1367380920
Available at: http://digitalcommons.wayne.edu/jmasm/vol12/iss1/3

Journal of Modern Applied Statistical Methods
May 2013, Vol. 12, No. 1, 9-10

1538 – 9472/13/$95.00

p-Values versus Significance Levels

Phillip Good
Information Research,
Huntington Beach, CA

In this article Phillip Good responds to Richard Anderson's article *Conceptual Distinction between the Critical p Value and the Type I Error Rate in Permutation Testing*.

Key words: Statistics, null hypothesis, non-parametric, permutation test, exchangeability; Type I error; p-value, discrete set, significance level, validity.

Introduction

As a consequence of Richard Anderson eschewing formal definitions, I wish to clarify our understanding of permutation procedures.

A statistical procedure is said to have significance level α if one can expect the hypothesis H to be rejected in error α% of the time under the following conditions:

- H is true,
- The assumptions underlying the procedure are valid,
- Repeated samples are taken (not necessarily from the same population) and the procedure is applied to each of the samples independently.

Thus, a significance level is a property of a statistical procedure and takes a fixed value. In contrast, a p-value is a random variable whose value depends upon the composition of the individual sample.

These definitions apply regardless of the nature of the statistical procedure, whether it is permutation or parametric, optimal or suboptimal, providing the assumptions underlying the procedure are valid each time it is applied.

One can apply a permutation procedure and achieve or closely approximate the anticipated frequency of type I errors only if the observations that compose the sample are exchangeable (See Good (2002) for a formal definition of exchangeability.)

Philip Good is a statistical consultant. He authored numerous books that include *Introduction to Statistics via Resampling Methods and R/S-PLUS* and *Common Errors in Statistics (and How to Avoid Them)*. Email: drgood@statcourse.com.

The failure to satisfy the requirement of exchangeability explains and renders invalid the so-called counter-examples offered by Hayes and by Mewhort and his colleagues.

Hayes (1996) draws samples from a population that is a mixture of bivariate normal distributions, each with the same mean and variance but differing rho. Although the couples (x,y)[i] are exchangeable in this false counter-example as Hayes asserts, the variables y[i] are not when they are exchanged independently of the {x[j]}. Thus, a permutation distribution is not applicable for testing the hypothesis rho=0, and the simulations performed by Hayes were not necessary to confirm this.

As there are only a finite number of rearrangements of observations, permutation procedures yield only a finite number of possible p-values. They may or may not be able to achieve a predetermined significance level exactly. This is the case in Mewhort, et al. (2009). In addition, his observations are not exchangeable as the observations in the groups he examines have different variances.

References

Anderson, R. B. (2013). Conceptual Distinction between the Critical p Value and the Type I Error Rate in Permutation Testing. *Journal of Modern Applied Statistical Methods*, *12*(*1*), 2-8.

Good, P. I. (2002). Extensions of the concept of exchangeability and their applications. *Journal of Modern Applied Statistical Methods*, *1*, 243-247.

Hayes, A. F. (1996). Permutation test is not distribution-free: Testing H_0: $\rho = 0$. *Psychological Methods*, *1*, 184-198.

Mewhort, D. J. K., Kelly, M., & Johns, B. T. (2009). Randomization tests and the unequal-N/unequal-variance problem. *Behavior Research Methods*, *41*, 664-667.