

5-1-2002

Using The t Test With Uncommon Sample Sizes

Shlomo S. Sawilowsky

Wayne State University, shlomo@wayne.edu

Barry S. Markman

Wayne State University

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Sawilowsky, Shlomo S. and Markman, Barry S. (2002) "Using The t Test With Uncommon Sample Sizes," *Journal of Modern Applied Statistical Methods*: Vol. 1 : Iss. 1 , Article 20.

DOI: [10.22237/jmasm/1020255660](https://doi.org/10.22237/jmasm/1020255660)

Using The t Test With Uncommon Sample Sizes

Shlomo S. Sawilowsky
 Educational Evaluation & Research
 College of Education
 Wayne State University

Barry S. Markman
 Educational Psychology
 College of Education
 Wayne State University

Monte Carlo techniques were used to determine the effect of using common critical values as an approximation for uncommon sample sizes. Results indicate there can be a significant loss in statistical power. Therefore, even though many instructors now rely on computer statistics packages, the recommendation is made to provide more specificity (i.e., values between 30 and 60) in tables of critical values published in textbooks.

Keywords: t test, Critical values

Introduction

There is a practical problem that persists when using tabled critical values in introductory behavioral and social science statistics and research textbooks. We refer specifically to courses where statistical packages for the computer are not used. Critical t values are only given for common degrees of freedom (df) and alpha levels. Consider calculating a two independent means t-test on response data for Condition A and Condition B in a layout with $n = 19$ per group. The df, $n_1 + n_2 - 2$, is 36. A survey of many popular textbooks showed most only had tabled values of t for df of 1 - 30, 40, 60, 120, and infinity. (A few textbooks also included 45 df.)

Which tabled value should be used for 36 df in a classroom demonstration of the use of the t test without a computer package? One possibility is to interpolate between the tabled values for 30 d.f. and 40 d.f. Instructors commonly eschew this approach because of the possibility that students will interpolate differently, obtain a different critical value, and subsequently cause confusion in the classroom. Another prevalent classroom solution is to use the critical t values either for 30 df or for 40 df, which are given in all textbooks reviewed. With a sample size of $N=38$ ($df = 36$), the t statistic associated with 30 df will be conservative, while the t statistic associated with 40 df will be liberal. An inspection of the t table indicates only minor

differences at 30 df and 40 df, differences mainly in the 1/100ths place. How does the imprecision of using the approximate critical value affect the power properties of the t-test?

Methodology and Results

A Monte Carlo study was conducted to answer this question. A Fortran program was written for an IBM compatible computer accessing the PC version of the IMSL (1987) subroutine library. Random variates were drawn from a normal population and t-tests using the correct and approximate critical values were calculated for various alpha levels. A fail to reject vs. reject decision was recorded. This was repeated for 10,000 repetitions.

A treatment effect of shift in location parameter was modeled by adding a constant to each of the observations of one group. The constant was selected to obtain a power level of .500 for the correct t value when nominal alpha was .100. That same constant was used to obtain the power at the .050 and .010 alpha levels as well. The results are compiled in Table 1.

Table 1. Power Rates When Using the Higher t (t_h), t, And The Lower t (t_l); Alpha = .100, .050, And .010; $n_1 = n_2 = 19$; 10,000 Repetitions.

| α | t_h | t | t_l |
|----------|-------|------|-------|
| .100 | .494 | .500 | .501 |
| .050 | .368 | .374 | .376 |
| .010 | .160 | .172 | .172 |

The results in Table 1 indicate that, at the cost of being slightly liberal, t_l does not provide any appreciable power over the correct t. The power results for t_h demonstrated a noticeable loss of power as high as 6.25%. However, the temptation to use the conservative t_h , with its slight loss of power, sends the wrong message to students, in consideration of the abundant literature in education and

Shlomo S. Sawilowsky, Wayne State University Distinguished Faculty Fellow, is Professor and Chair, Educational Evaluation and Research, 351 College of Education, Wayne State University, Detroit, MI, 48202, e-mail: shlomo@edstat.coe.wayne.edu. His interests are in non-parametric, robust, exact, permutation, and other computer-intensive methods, especially via Fortran. Barry S. Markman is Professor of Educational Psychology at Wayne State University. His areas of interest are assessing ADHD using continuous performance tasks, test anxiety, anxiety and cancer, and learning theory.

psychology that suggests that treatment effects are often extremely small. In such situations the cavalier approach to the loss of *any* power to detect a false null hypothesis is exacerbated. (Moreover, we note that there remained staunch supporters of the superiority of the t test vs the Wilcoxon test during the parametric vs nonparametric wars of the last quarter of the 20th century, despite the fact that countless Monte Carlo studies demonstrated the power advantage of the t test under normality was only a meager 1% - 3%!)

Conclusion

A simple solution is proposed for the problem of uncommon df for sample sizes commonly used classroom instruction. It is recommend that textbook authors include the missing critical for 31 through 39 df and 41 through 59 df.

Moreover, there also frequently arises the need for uncommon alpha levels. For example, consider conducting a "one-and-a-half-tailed" t test (Ramsey, 1990).

Alpha is set to .04 in one tail to maximize power in the anticipated direction, and alpha is also set at .01 in the other tail. This will provide more power than the usual two-tailed test with .025 in each tail, while providing some protection against outcomes occurring in the wrong direction. Because .04 is an uncommon alpha level, Ramsey (1990) suggested using the Z table as approximation critical values. This suggestion is clearly undesirable. Nevertheless, this may be a circumstance that cannot be adequately handled with tabled critical values found in the back of statistics textbooks.

References

IMSL (1987). *IMSL Stat/Library user's manual: Fortran subroutines for statistical analysis*. Version 1.0. Houston: Author.

Ramsey, P. H. (1990). 'One-and-a-half-tailed' tests of significance. *Psychological Reports*, 66, 653-654.