11-1-2005

Misconceptions Leading to Choosing the t Test Over the Wilcoxon Mann-Whitney Test for Shift in Location Parameter

Shlomo S. Sawilowsky
Wayne State University, shlomo@wayne.edu

Journal of Modern Applied Statistical Methods
November, 2005, Vol. 4, No. 2, 598-600

1538 – 9472/05/$95.00

Misconceptions Leading to Choosing the t Test Over the Wilcoxon Mann-Whitney Test for Shift in Location Parameter

Shlomo S. Sawilowsky
Wayne State University

There exist many misconceptions in choosing the t over the Wilcoxon Rank-Sum test when testing for shift. Examples are given in the following three groups: (1) false statement, (2) true premise, but false conclusion, and (3) true statement irrelevant in choosing between the t test and the Wilcoxon Rank Sum test.

Key words: t test, Wilcoxon Rank-Sum test, robustness, power

Introduction

For treatment effects modeled as a shift in location parameter, the t test can be decidedly nonrobust to departures from population normality unless certain conditions have been met (Sawilowsky & Blair, 1992). When normality is met or nearly met (which occurs rarely), the t test maintains a very small power advantage over the Wilcoxon Rank Sum / Mann-Whitney U test. When normality is violated, the Wilcoxon Rank Sum Test can be three or four times more powerful than the independent samples t test (Blair, 1980; Blair & Higgins, 1980a, 1980b, 1981; Blair, Higgins, & Smitely, 1980; Sawilowsky & Blair, 1992). The power advantages of the nonparametric test actually increases with sample size for the low to mid-level parts of the t test's power spectrum.

Although the power advantage is not as spectacular as with the independent samples case, the Wilcoxon Signed-Ranks test for two dependent samples nevertheless maintains a considerable power advantage over the dependent samples t test for similar conditions (Blair & Higgins, 1985a, 1985b).

Shlomo S. Sawilowsky is Professor, WSU Distinguished Faculty Fellow, and Careful Data Analyst. Email him at shlomo@wayne.edu.

The dates of the Monte Carlo studies cited above are from 1980 – 1992. Promise for these small sample results was available decades prior on the basis of large sample asymptotic theory. This understanding had even penetrated to the level of a *book review* written in 1968! "The Wilcoxon rank-sum test…show[s] only slight losses in both large and small sample efficiency relative to the t-test in the normal case, while in many non-normal cases, efficiency exceeds 100%" (Meeter, 1968).

Thus, sane researchers opt to use the Wilcoxon Rank Sum test when testing for shift in location. Overly cautious researchers, with no justification, opt to perform both the t test and the Wilcoxon Rank Sum test, and accept the Wilcoxon only if it rejects and the t doesn't. (This is a misguided practice, as it leads to an increase in experiment-wise Type I errors.) Pedantic researchers, oblivious to the Monte Carlo results of the past 25 years, and asymptotic results for the past half-century, simply ignore the Wilcoxon Rank Sum test in favor of the t test.

In the course of reviewing articles submitted to the sixteen journals that I have provided ad hoc reviews over the past 15 years, I have compiled a list of constantly recycling reasons given for preferring the t test over the Wilcoxon Rank Sum test when testing for shift in location. They are presented below without expansive commentary, in the hopes that they never again resurface.

The misconceptions are categorized in three groups: (1) false statement, (2) true premise, but false conclusion, and (3) true statement irrelevant in choosing between the t test and the Wilcoxon Rank Sum test.

(1) False Statement

- the Wilcoxon is only for use when the data are originally in the form of ranks
- the Wilcoxon's ranking procedure throws away useful information
- the Wilcoxon is only for use in the presence of outliers
- the Wilcoxon should only be used for small samples
- the t is robust with respect to Type I errors
- the t is more powerful
- if a modern procedure should be used, it should be a permutation test, not the Wilcoxon

(2) True Premise, but False Conclusion

- the Wilcoxon is a test of $f_i(x) = g_i(x)$ (true), so even if it does reject and the t doesn't, it is probably due to some difference other than the mean (e.g., scale) (false)
- the Wilcoxon's underlying assumptions are weaker (true), therefore the hypothesis being tested is less interesting (false)
- in terms of central tendency, the Wilcoxon pertains to the median (true), which is less interesting than the mean (false)
- the t is expandable to the k samples case (true), but the Wilcoxon is not (false)
- the t is expandable to the multivariate case (true), but the Wilcoxon is not (false)
- the t is expandable to the factorial case (true), but the Wilcoxon is not (false)

(3) True Statement Irrelevant in Choosing Between the t and Wilcoxon

- the t is a classical test
- results based on the t have been accumulating for almost a century, permitting direct comparison of results over time
- the t on the ranks is equivalent to the Wilcoxon on the original scores
- the hypotheses being tested for the t and Wilcoxon aren't exactly the same
- the t is the Uniformly Most Powerful Unbiased test under normality
- the t is robust with respect to Type II errors for departures from normality
- for very small sample sizes the t can be conducted at $\alpha = .05$ or .01, but the Wilcoxon cannot because there are no critical values
- at relatively small sample sizes, the Wilcoxon test cannot be conducted at exactly the $\alpha = .05$ or .01 levels due to the discrete nature of the sampling distribution
- even its inventor called the Wilcoxon test a "quick and dirty" or "crude" procedure

References

Blair, R. C. (1980). *A comparison of the power of the two independent means* t *test to that of theWilcoxon's rank-sum teset for samples of various populations.* Unpublished doctoral dissertation, University of South Florida, Tampa, FL.

Blair, R. C., & Higgins, J. J. (1980a). A comparison of the *t* test and the Wilcoxon statistics when samples are drawn from a certain mixed normal distribution. *Evaluation Review, 4*, 645-656.

Blair, R. C., & Higgins, J. J. (1980b). A comparison of the power of the Wilcoxon's rank-sum statistic to that of student's t statistic under various non-normal distributions. *Journal of Educational Statistics*, *5*(4), 309-335.

Blair, R. C., & Higgins, J. J. (1981). A note on the asymptotic relative efficiency of the Wilcoxon rank-sum test relative to the independent means t test under mixtures of two normal distributions. *British Journal of Mathematical and Statistical Psychology*, *31*, 125-128.

Blair, R. C., & Higgins, J. J., & Smitely, W. D. S. (1980). On the relative power of the *U* and *t* tests. *British Journal of Mathematical and Statistical Psychology*, 33, 114-120.

Meeter, D. (1968). Book Reviews, *Journal of the American Statistical Association*, 62, p. 1505)

Sawilowsky, S. S., and Blair, R. C. (1992). A more realistic look at the robustness and type II error properties of the t test to departures from population normality. *Psychological Bulletin*, *111*, 353-360.