A Simulation Study Using R to Compare the Wilcoxon Mann Whitney Test to the Two-sample t-test under Different Scenarios

Jimmy Wong[1,*], Dr. Gail Potter[1]

1. California Polytechnic State University, San Luis Obispo
*Contact author: jwong100@calpoly.edu

Keywords: Wilcoxon Mann Whitney test, t-test, simulations, type I error rate, power analysis

The true type I error rate and true statistical power of the Wilcoxon Mann Whitney (WMW) test and the two-sample t-test can be inconsistent with the nominal type I error rate and nominal statistical power when conditions are violated. For the t-test to be valid, the population distributions need to be Normal or the sample sizes need to be at least 30 for the Central Limit Theorem to apply. In contrast, the WMW test is valid only when the shape of the two population distributions are the same. We performed simulation studies to estimate alpha and power for these two tests by adjusting different factors. Specifically, these factors were the sample sizes, the shapes of the two populations, and the magnitude of the disparity between the two population means.

Using *R*, we first performed specific simulations with the conditions of our preference. We then created functions to generalize our existing code by allowing the user to specify different conditions by adjusting the input parameters. The output of our functions are graphs created by using functions from **ggplot2** and **gridExtra** along with tables that allow the user to analyze the results both visually and numerically. Primarily, we used functions such as `geom_line` and `geom_path` from **ggplot2** to generate curves to compare how the true alpha and true power differ between the WMW test and the t-test under different conditions. Then **gridExtra** provided us ways to arrange certain graphs together to create more cohesive combined graphs by using functions such as `grid.arrange` and `arrangeGrob`. With the functions that we wrote, we are now in the process of creating an R package to make these functions publicly available to teachers, students, and users of statistics who want to perform power analysis and compare the performance of these two tests.

From our simulation studies, the t-test evidently performed better than the Wilcoxon-Mann-Whitney test in most, but not all, scenarios. The population distributions that we simulated from were the Normal and the Gamma distributions. The first scenario was set to deliberately make the conditions of both tests to be violated. We found that when the two population distributions are different, the t-test performed better than the WMW test by producing a lower alpha than the true alpha. The superiority of the t-test continued to be evident when we observed that it had a higher power than the WMW test under the scenario that the two population distributions are different and with different means. Further controlling for the magnitude of the disparity in the population means, it was intriguing that the t-test still had a higher power than the WMW test on average. Our focus then shifted to comparing the performance of the two tests when the two population distributions are both Normal, so conditions were satisfied for both tests. We discovered that the t-test had a higher power than the WMW test even with small sample sizes, but WMW had a lower Type I error rate. Therefore, although the t-test is more robust to condition violations, it is not necessarily true that it performed better in every scenario.

In summary, we have created functions in R which compare type I error rate and power for the Wilcoxon Mann Whitney test and the t-test in a variety of scenarios through simulation. These functions allow the user to specify the desired scenario by modifying input parameters to the function. As future work, we may consider different population shapes and larger sample sizes. We also plan to make these functions publicly available in an R package.