

STATISTICS SEMINAR

Bo Tong
Ph.D. Defense

Wednesday, December 2, 2015
King Hall 004, 1:00-3:30 pm



More Accurate Two Sample Comparisons for Skewed Populations

Various tests have been created to compare the means of two populations in many scenarios and applications. The two-sample t-test, Wilcoxon Rank-Sum Test and bootstrap t test are commonly used methods. However, methods for skewed two-sample data set are not well studied. In this dissertation, several existing two sample tests were evaluated and four new tests were proposed to improve the test accuracy under moderate sample size and high population skewness.

The proposed work starts with derivation of a first order Edgeworth expansion for the test statistic of the two sample t-test. Using this result, new two-sample tests based on Cornish Fisher expansion (TCF tests) were created for both cases of common variance and unequal variances. These tests can account for population skewness and give more accurate test results. We also developed three new tests based on three transformations (T_i test, $i = 1, 2, 3$) for the pooled case, which can be used to eliminate the skewness of the studentized statistic.

In this dissertation, some theoretical properties of the newly proposed tests are presented. In particular, we derived the order of type I error rate accuracy of the pooled two-sample t-test based on normal approximation (TN test), the TCF and T_i tests. We proved that these tests give the same theoretical type I error rate under skewness. In addition, we derived the power function of the TCF and TN tests as a function of the population parameters. We also provided the detailed conditions under which the theoretical power of the two-sample TCF test is higher than the two-sample TN test. Results from extensive simulation studies and real data analysis were also presented in this dissertation. The empirical results further confirm our theoretical results. Comparing with commonly used two-sample parametric and nonparametric tests, our new tests (TCF and T_i) provide the same empirical type I error rate but higher power.