Statistics: new t-tests for the comparison of two partially overlapping samples

Ben Derrick, PhD candidate, University of the West of England, Bristol
email: ben.derrick@uwe.ac.uk

Introduction

In experimental design for the comparison of two samples, some pairing may occur resulting in paired observations, and there will often be observations without a pairing, thus being independent observations. This scenario is referred to as partially overlapping samples [2]. For the comparison of means, traditional tests that discard either the paired observations or the independent observations are compared against new t-tests proposed by [3] that make use of all of the available data. Calculations are performed as per the tutorial [5] using the R package 'Partiallyoverlapping' [1]

Key

Traditional tests discarding data:
T_1 Paired samples t-test
T_2 Independent samples t-test, assuming equal variances
T_3 Independent samples t-test, Welch's degrees of freedom approximation (see [4])
W Wilcoxon test

New test statistics:
T_{new1} Partially overlapping samples t-test, assuming equal variances [3]
T_{new2} Partially overlapping samples t-test, using Welch's degrees of freedom approximation [3]

Results shown are the null hypothesis rejection rates over 10,000 simulations from the Normal distribution and the Lognormal distribution, for each parameter combination of n_a, n_b, n_c = (5, 10, 30, 50, 100, 500), and Pearson's correlation coefficient -.75:.75(.25). The parameters n_a, n_b, n_c are equivalent to the number of independent observations in Sample 1, the number of independent observations in Sample 2, and the number of paired observations respectively. Under the null hypothesis the observations are generated from populations with equal means. Under the alternative hypothesis an arbitrary 0.5 is added to each Sample 2 observation.

Type I error robustness

How is Type I error robustness determined?

For a test statistic to be Type I error robust, when the null hypothesis is true, the p-values from multiple simulations of the data generation process for a population distribution will be approximately uniformly distributed. Therefore for each parameter combination the null hypothesis rejection rate at the 5% significance level should be approximately equal to 5%, or using liberal robustness criteria, between 2.5% and 7.5%.

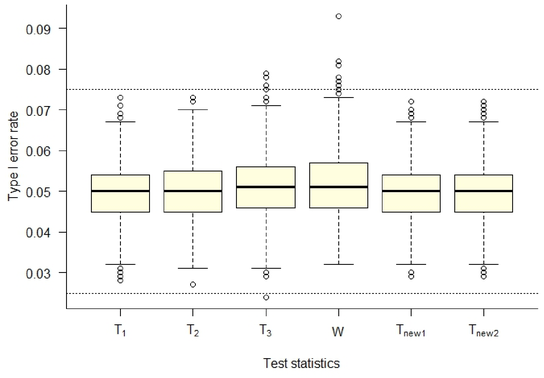

Fig. 1: Type I error rates under the Normal distribution

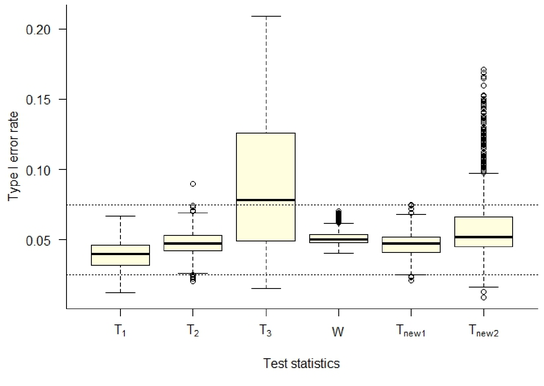

Fig. 2: Type I error rates under the Lognormal distribution

Power

How is the most powerful test determined?

When the alternative hypothesis is true i.e. the population means are not equal, the most powerful test is the one which rejects the null hypothesis the most frequently. Power is only fairly compared for test statistics that maintain Type I error robustness. Power for all tests statistics is given for completion, but caution should be expressed in the interpretation of power for the non Type I error robust statistics.

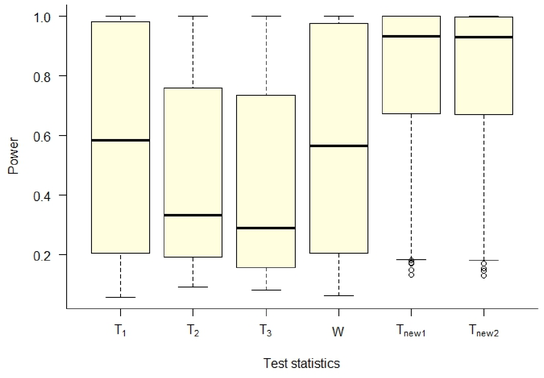

Fig. 3: Power under the Normal distribution

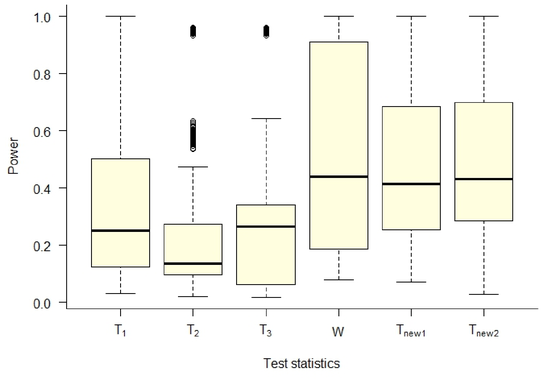

Fig. 4: Power under the Lognormal distribution

Conclusion

The statistic T_{new1} is Type I error robust for comparing the means of two samples taken from the same distribution. When the samples are taken from the Normal distribution with equal variances, T_{new1} is the most powerful test statistic. When the samples are taken from identical non-normal distributions, for example the Lognormal distribution, T_{new1} has favorable power relative to other parametric tests.

Additional Research

- Comparison of means when discrete ordinal data is present. Background research has been concluded for the non partially overlapping samples scenario (see [6])
- Comparison of proportions (see [2])

Publications

[1] Derrick, B.
Partiallyoverlapping: R package for performing partially overlapping samples t-tests.
CRAN R package version 1.0 (2017).

[2] Derrick, B., Dobson-Mckittrick, A., Toher, D., and White, P.
Test statistics for comparing two proportions with partially overlapping samples.
Journal of Applied Quantitative Methods 10, 3 (2015).

[3] Derrick, B., Russ, B., Toher, D., and White, P.
Test statistics for the comparison of means for two samples which include both paired observations and independent observations.
Journal of Modern Applied Statistical Methods 16, 1 (2017).

[4] Derrick, B., Toher, D., and White, P.
Why Welch's test is Type I error robust.
The Quantitative Methods in Psychology 12, 1 (2016).

[5] Derrick, B., Toher, D., and White, P.
How to compare the means of two samples that include paired observations and independent observations: A companion to Derrick, Russ, Toher and White (2017).
The Quantitative Methods in Psychology 13, 2 (2017).

[6] Derrick, B., and White, P.
Comparing two samples from individual Likert question.
International Journal of Mathematics and Statistics 18, 3 (2017).