Reviewer’s report

Title: Testing for Differences in Distribution Tails to Test for Differences in 'Maximum' Lifespan

Version: 1 Date: 4 March 2008

Reviewer: Hon Keung Tony Ng

Reviewer’s report:

1. Using the number of subjects with lifespan above a sample percentile (say 90-th percentile) and then construct the hypothesis testing based on independent binomial distributions can be problematic. Refer to Table 1 in Wang et al. (2004) as an example, it is clear that $X_1$ and $X_2$ are not independent binomial random variable when the threshold (tau) is set to be the 90-th percentile since the sum of $X_1$ and $X_2$ must equal to $0.1 \times$ (total sample size) = 10. Therefore, using the test statistics developed for testing the difference between two independent binomial proportions will not be appropriate here and the results in the tables for “tau set to sample 90th percentile” are questionable.

The authors may need to justify the use of the test procedures for two independent binomial proportions. Moreover, the tests considered in the manuscript (as well as Wang et al., 2004) are only a few of the existing test procedures for this purpose. One may refer to Newcombe (1998) and some other subsequent work for reference.

2. In the simulation study, the underlying value of the threshold is assumed to be known and set to be 130. In real-life situations, this threshold is usually unknown and it plays an important role in the performance of the test procedures studied in the manuscript. One may consider a simulation study to investigate the effect of different choices of the threshold value.

3. The test procedure proposed to test $H_{(0,C)}$ are the Wilcoxon-Mann-Whitney and the permutation test based on $Z_i = I(Y_i > \tau) Y_i$. When there is a large number of lifespan $< \tau$, it may raise problems to the rank-based and permutation tests since a large number of $Z_i$ are equal to zero. On the other hand, I suspected that the power performance of the test may not be good when the number of $Y_i > \tau$ are small in either the treatment group or control group.

To illustrate these points, here is a hypothetical example:

Let number of subjects in treatment group is $N_1 = 10$, number of subjects in control group is $N_2 = 10$. Suppose 6 subjects in the treatment group have lifespan $< 130$, 9 subject in the control group have lifespan $< 130$ and we have the following observations:

Treatment: 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 190, 190, 190, 190
Control: 0, 0, 0, 0, 0, 0, 0, 0, 0, 200
If we construct the permutation distribution of a test statistic based on the above example, the distribution will be highly discrete. For rank test, the treatment group will get a rank-sum = 118 with p-value 0.176. By just looking at the data, it is clear that the treatment group have a longer lifespan compare to the control group, but the test does not correctly reject the null hypothesis of the two groups have equal maximum lifespan.

4. Please explain clearly how the permutation test is done.

5. In Table 2 on page 17, the authors mentioned “The bolded values are those simulated type I error rates which are significantly higher than the nominal alpha level.” Please give an explicit definition on “significantly higher”. It is confusing that 0.014 (the second column under N = 50, alpha = 0.01) is bolded but 0.017 and 0.016 are not bolded in the same column.

References:


What next?: Reject because scientifically unsound

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests