Location-scale tests for non-negative data with skewed distribution, with focus on parasitology research

Shaimaa Yehia, Tanta University, Faculty of Science, Tanta, Egypt
Jenő Reiczigel, Szent István University, Faculty of Veterinary Science, Budapest, Hungary

Non-negative data with skewed distributions arise in various research areas (income per household, treatment costs of a disease, etc.). Our motivation comes from parasitology where skewed infection intensity data are also quite typical. In that field of research these are called “aggregated”, which means that most hosts (those with stronger defense mechanisms against parasites, e.g. a strong immune system) are free or just slightly infected, and a relatively small part of hosts harbor the majority of the parasite population.

For the simulation we generated a large finite population (n=100000) from each distribution of interest. Alpha was assessed taking both population (n=1000) from each distribution. As more infected samples have both higher mean and higher SD, we expected that larger sample sizes will make the effect of skewed distribution more noticeable.

Methods

We compared Cucconi’s and Neuhäuser’s location-scale tests (for details see Morazzi, 2013) to 4 commonly used location tests (Welch-t-test, Mann-Whitney-test, permutation t-test and bootstrap-t-test) for 3 right-skewed theoretical distributions and 3 empirical parasite distributions.

The theoretical distributions were a chi-squared distribution on 5 df, an exponential distribution with \( \lambda = 0.1 \), and a gamma distribution with shape=0.5, scale=20. The parasite distributions were generated from 3 parasite samples reported in Rózsa et al. (2000).

Two-sample comparison of parasite infection data is usually made by location tests. As more infected samples have both higher mean and higher SD, we expected that location-scale tests would be more powerful.

Results

We feel that it is important to verify the results of the simulations. However, we expected that the skewness of the distribution will lead to a decreased power of location-scale tests compared assuming that data is usually made by location tests. As more samples have both higher mean and higher SD, we expected that location-scale tests would be more powerful.

Conclusions

- Location-scale tests may have a role in parasitology research.
- For skewed data and a progressive shift alternative Neuhäuser's test had the highest power.
- If interest lies in pure location differences, only the bootstrap test is applicable. We found that it maintains the alpha error rate for balanced or moderately unbalanced designs (sample size ratio \( \leq 2.5 \)).
- In relation with parasite infection data, the probit-based progressive shift offers a more realistic alternative than the conventional shift or scale alternatives. We feel that it is also true for the analysis of other skewed data, such as treatment cost data.
- As the result of a method comparison study may depend on the alternative hypothesis, comparisons must be carried out assuming that alternative, which is most realistic in the field of interest.

References


This research was supported by the Hungarian National Research Fund (OTKA K 108571) and by the Research Faculty Grant 2014 of the Szent István University, Faculty of Veterinary Science.