Sample Size Planning for Multiple Contrast Tests

Anna Pöhlmann[1], Frank Konietschke[1]
[1]Institut für Biometrie und klinische Epidemiologie, Charité – Universitätsmedizin Berlin, Berlin

Abstract
Any trial should start with its careful planning and especially with sample size calculations. The planning phase of an experiment is key, since errors in the statistical planning can have severe consequences on both the results and conclusions drawn from the data. In many trials and experiments more than two (independent) samples are observed, e.g. when subjects (animals) are assigned to different dose levels of a compound in preclinical research. Consequently, methods suitable for the analysis of several samples should be chosen for sample size computations and power considerations. Multiple contrast tests (max t-tests) have themselves established as valuable tools in statistical sciences [1]. In comparison with global testing procedures (e.g. ANOVA), these methods share the advantage of providing local (and global) test decisions as well as simultaneous confidence intervals for the treatment effects of interest [1]. In principle, the method is nothing but a multiple t-test and therefore its interpretation is feasible. In recent years, theoretical power considerations have been performed by e.g., Genz and Bretz [2], Konietschke et al. [3], Bretz et al. [4] and Dilba et al. [5] assuming homogeneous variances. While these elaborations are rather general, we extend their ideas with respect to the following point of views: 1) How should we allocate the individuals to the groups given a fixed total number of observations? 2) Considering a predefined allocation rate, how many individuals do we need in total? Given the significance level, the power, the variance within the groups, and the effect size, the required number of observations is calculated. We develop and implement procedures for balanced and unbalanced designs, each with homogeneous and heteroscedastic variances. Since no closed formula is available, we deploy numerical approximation methods. This results partly in complex discrete multivariate optimization problems. After comparing various algorithms, we have found that a grid search works well and efficiently. The discrepancy between the predefined and actually achieved power is neglectable.

References
[1] Hothorn, T., Bretz, F., and Westfall, P.: Simultaneous inference in general parametric models. Biometrical Journal: Journal of Mathematical Methods in Biosciences. 2008 (3) 50: 346-363 DOI: 10.1002/bimj.200810425.
[2] Genz, A., Bretz, F.: Numerical computation of multivariate t-probabilities with application to power calculation of multiple contrasts. Journal of Statistical Computation and Simulation. 1999 (4) 63: 103-117 DOI: 10.1080/00949659908811962.
[3] Konietschke, F., Bösiger, S., Brunner, E., and Hothorn, L.A.: Are multiple contrast tests superior to the ANOVA? The International Journal of Biostatistics. 2013 (1) 9: 63-73 DOI: 10.1515/ijb-2012-0020.
[4] Bretz, F., Genz, A., and Hothorn, L.A.: On the numerical availability of multiple comparison procedures. Biometrical Journal. 2001 (5) 43: 645 – 656.
[5] Dilba, G., Bretz, F., Hothorn, L. A., Guiard, V.: Power and sample size computations in simultaneous tests for non-inferiority based on relative margins. Statistics in Medicine. 2006 (7) 25: 1131-1147 DOI: 10.1002/sim.2359.