

Abstract

A robustness study to investigate the performance of six procedures for comparing two or more groups, under several circumstances, has been carried out. The study was divided in four parts. The first three were Monte-Carlo studies, and the fourth was a study on empirical data. The six procedures were the ANOVA (or its special case, the Student's t) procedure (1), the Satterthwaite (or its special case, the Welch) procedure (2), the bootstrap Satterthwaite (the Welch) procedure (3), the bootstrap percentile procedure for the mean (4) and for the median (5), and the Bonett and Price median procedure (6). The procedures were compared on confidence interval widths and coverage percentages, based on 10.000 replications. In Studies 1 and 2, samples were drawn from simulated population distributions with the same (non-) normality. The conditions were varied with different non-normal distributions, unequal population variances, and different equal or unequal sample sizes. In Study 1 two samples were compared on the mean or median and in Study 2 four samples were compared on the mean or median with a contrast analysis. For Study 3 simulated cardiovascular data obtained by Van Roon's baroreflex model (1998) were used. In this study, the results of logarithmically transformed data were compared to the results of untransformed data. Study 4 aimed at comparing real-world experimental data in a one-way ANOVA design with three independent groups.

The Student's t and the ANOVA procedure performed well under many conditions but failed in particular when unequal population standard deviations were accompanied with unequal sample sizes. In general, the problems encountered with the Student's t and the ANOVA procedures were best overcome by the Welch and the Satterthwaite procedures. The Student's t , the ANOVA, the Welch, and the Satterthwaite procedures, however, appeared more robust to deviations from normality than expected.

It was also found that the choice for logarithmically transformation of the cardiovascular data was not as obvious as expected on theoretical basis. The intended result of transformation was not always met. Furthermore, in particular for the Welch (and the Satterthwaite) procedure transformation of the cardiovascular data was not needed to get reliable results.

For the conditions investigated in this study, the Welch and the Satterthwaite procedure are as good as or much better than the Student's t and the ANOVA procedure. Therefore, it is concluded that the Welch procedure for comparing two samples and the Satterthwaite procedure for comparing more than two samples are the best procedure to use in experimental research.