ROBUSTNESS OF TWO SAMPLE TESTS OF MEANS UNDER NON-NORMALITY AND HETEROGENEITY OF VARIANCES

SZU-I LIAO

Boston University Graduate School of Arts and Sciences, 1999

Major Professor: Austin F. S. Lee, Professor of Mathematics

ABSTRACT

Statistical hypothesis tests are procedures for testing pre-stated hypotheses. The development and properties of these procedures as well as their performance are based upon certain assumptions. Desirable properties of statistical tests are to maintain validity and to perform well even if these assumptions are not met. A statistical test that maintains such desirable properties is called robust. This thesis deals with the investigation of the robustness of several two-sample tests. New tests are also proposed in an attempt to construct more robust procedures.

The two-sample Student's t test was developed on the basis of two assumptions: (1) the samples are drawn from normal populations with unknown but equal variance, and (2) a random sample is drawn independently from each. For variances unequal (Behrens-Fisher problem), alternative tests have been proposed such as the Welch approximated t and the Lee-Gurland tests. Many other tests were also suggested when the populations are not normal. Among them are the Yuen's and Tiku's tests that utilize the trimmed and Winsorized means, respectively.

Other newly proposed tests combine a robust location estimator known as the Gastwirth's weighted mean of three percentile points, and the Least-Absolute-Deviations estimator for the scale parameter. These tests along with the aforementioned procedures are considered in the thesis for robustness study when the underlying distributions belong to the families of exponential power and t.

The results from this study show the classical Student's t test maintains good control of the probability of type I error when the underlying distributions are similar, or when samples of equal size are taken from both populations. Otherwise the tests that use only three percentile points perform better. The normality-based Welch's approximated t and Lee-Gurland tests perform equally well and surprisingly far better than those specifically designed for general robustness. Power analysis also shows that these two tests maintain high power when the underlying distributions deviate from the assumptions.