Journal of Mathematics and Statistics 8 (2): 274-282, 2012
ISSN 1549-3644
© 2012 Science Publications

One-Sided Multivariate Tests for High Dimensional Data

Samruam Chongcharoen
School of Applied Statistics,
National Institute of Development Administration,
Bangkapi District, Bangkok, 10240, Thailand

Abstract: Problem statement: For a multivariate normal population with size smaller than dimension, n<p, the likelihood ratio tests of the null hypothesis that the mean vector was zero with a one-sided alternative were no longer valid because they involved with sample covariance matrix which was singular. **Approach:** The test statistics for one-sided multivariate hypotheses with n<p were proposed. **Results:** The simulation study showed that the proposed tests provided reasonable type I error rate for one-sided covariance structures. They also give good powers. The application of these tests was given by testing of one-sided hypotheses on DNA micro array data. **Conclusion:** Under that there have no such other tests available at present for this kind of hypothesis testing with n<p yet, the proposed tests are good ones. However, the methodology is valid for any one-sided hypotheses application which involves high-dimensional data.

Key words: DNA micro arrays, multivariate normal, one-sided multivariate test, Follmann's test, power comparison

INTRODUCTION

Suppose one uses a matched-pair design to compare the multivariate responses of two treatments. If the responses are p dimensional and $\theta = (\theta_1, \theta_2,\ldots, \theta_p)$ is the difference, treatment one minus treatment two, of the mean responses, then one may test the null hypothesis, H_0: $\theta_1 = \theta_2 = \ldots = \theta_p = 0$, to determine if there is a difference in the two treatments. Furthermore, if one believes that for each coordinate, the mean responses for treatment one are at least as large as those for treatment two, then the alternative can be constrained by H_1: $\theta_i \geq 0$ for i = 1, 2,...,p.

Based on a random sample with n>p from the normal distribution with mean θ and covariance matrix V, Kudo (1963); Shorack (1967) and Perlman (1969) derived the likelihood ratio test of H_0 versus H_1-H_0 for the cases in which V is known, known up to a multiplicative constant and completely unknown, respectively. Because the likelihood ratio tests with restricted alternatives are complicated to use, Tang *et al*. (1989) proposed an approximate likelihood ratio test and Follmann (1996) proposed one-sided modifications of the usual χ^2 and Hotelling's T^2 tests of H_0 versus ~H_0 that are easier to implement. Using exact computations and Monte Carlo methods, Chongcharoen *et al*. (2002) compared the performance of Kudo's test, Follmann's test, a new test, which is a modification of Follmann's test, the permutation test of Boyett and Shuster (1977) and the Tang-Gnecco-Geller test for a known covariance matrix and for a partially known covariance matrix, they compared the powers of these tests with Kudo's test replaced by Shorack's test. For a completely unknown covariance matrix, Chongcharoen (2009) studied the power of these one-sided tests for unknown covariance matrices with equal variances and unequal variances as well as tests obtained by combining the Boyett and Shuster (1977) technique to Follmann's test, the new test, Perlman's test and the Tang-Gnecco-Geller test.

In some situations, there are no longer data for n>p. That is, when the number n of available observations is smaller than the dimension P of the observed vectors. For example, the data come from DNA micro arrays where thousands of gene expression levels are measured in relatively few subjects. The one-sided multivariate tests as above are no longer valid for this kind of data because the pxp sample covariance matrix S is singular with rank n<p, S^{-1} does not exist. Since now there have no one-sided multivariate tests available for the data which has the number n of available observations is smaller than the dimension p yet, therefore the proposed tests were the one-sided multivariate tests for the data with n<p.

Throughout this study, suppose $X_1, X_2,\ldots,X_n$ is a random sample from a p-dimensional multivariate