Testing Means Using Hypothesis-Dependent Variance Estimates

Park, Choon-Il* and Lee, Kang Kil**

*Department of Applied Mathematics, Korea Maritime University, Pusan, Korea

**Dong Myung Junior College, Pusan, Korea

1. INTRODUCTION

This article examines some of the common univariate and multivariate tests from this viewpoint. It turns out that in all but one test considered, the use of hypothesis - dependtdent variance estimates leads to tests that are equivalent to the corresponding traditional tests utilizing the hypothesis-independtdent variance estimates. The realization of this fact may be interestiong and educational to practitioners and students of statistics.

2. UINVARIATE TESTS

The one sample t test is commonly used to test the null hypothesis

$$H_0 : \mu = \mu_0,$$

based on a random sample from a normal distribution with unknown variance σ^2.Thi s test is the likelihood ratio test as well as the uniformly most powerful unbiased test. The statistic t is given by

$$t = (\bar{x} - \mu_0)/(s^2/n)^{1/2},$$

where $\bar{x}$ is the sample average based on a random sample of size n . The sample variance s^2 is an unbiased estimate of σ^2 computed as

$$\hat{\sigma}^2 = s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)$$

The question of interest is why don't we take advantage of the specified value of μ under the null hypothesis in computing as estimate of σ^2, as

$$\hat{\sigma}^2 = s_0^2 = \sum_{i=1}^{n}(x_i - \mu_0)^2/n?$$

In turn, why don't we use t_0 as our test statistic to test the hypothesis in (2.1) where

$$t_0 = (\bar{x} - \mu_0)^2 / (s_0^2/n)^{1/2}?$$

The answers to these questions are rather simple. It can be easily shown that the t_0 statistic can be written in terms of the t statistic as follows :

$$t_0 = t[\,n/(n-1+t^2)]^{1/2}. \tag{2.1}$$

The one-to-one correspondence between the statistics t and t_0 and their corresponding distributions (percentile points) are now apparent. Note that even though $\bar{x}$ and s_0^2 are dependent, the distribution of t_0 can still be easily obtained from the distribution of t through (2.1). In summary, the t and t_0 tests lead to the same inference and hence are identical tests. It is intersting to note that a gain of one degree of freedom really does not improve the test. So why bo-ther with the t_0 test ? For this and related discussions, see Lefante and Shah (1986) and Good (1986).

In the same spirit, one can raise a similar question about the F statistic for testing the equality of the k treatment means, that is,

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k,$$

in a completely randomized design. The same logic leads to a new statistic, F_0, which can be expressed in terms of the usual F statistic as shown below :

$$F_0 = \frac{\text{treatment mean squares}}{\text{total mean squares}} = \frac{(N-1)F}{(k-1)F+(N-k)}.$$

Once again, the one-to-one correspondence between the statistics F and F_0 for the completely randomized design and their correspondending distribution lead to the same inference and hence are identical tests. So there is no need to bother with the F_0 test.

3. MULTIVARIATE TESTS

Now we continue with this same idea in the one-smaple and k - sample multivariate tetsting procedures. Suppose that we have indenpent vectors $X_1, \cdots, X_n$ from a p-variate normal population with unknown mean vector ξ and unknown covariance matrix Σ. The usual test statistic for testing

$$H_0 : \xi = \xi_0$$

is the Hotelling's T^2, which is given by

$$T^2 = n(\overline{X} - \xi_0)' S^{-1}(\overline{X} - \xi_0), \tag{3.1}$$

where X is the sample mean vector and

$$S = \sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})'/(n-1)$$

is the sample covariance matrix. If we replace S in (3.1) by the hypothesis-dependent covariance estimate,

$$S_0 = \sum_{i=1}^{n}(X_i - \xi_0)(X_i - \xi_0)'/n,$$

do we get a different test ? That is, does

$$T_0^2 = n(\overline{X} - \xi_0)' S_0^{-1}(\overline{X} - \xi_0)$$

lead to a different test? It is easy to show that T_0^2 can be expressed as

$$T_0^2 = T^2[n/(n-1) + T^2],$$

which is a one-to-one function of T^2.

Thus, for the same reasons given in the t test case, we conclude that the test pro cedures based on T^2 and T_0^2 are identical. Interestingly, Kshirsager (1972, problem 40, p. 490) has noted a statistic T'^2 which equals $(n-1)T_0^2/n$. Finally, we investigate the effect of hypothesis-dependent covariance esti-mates in the MANOVA setup for testing the equality of several normal mean vectors. Let ξ_i and Σ_i denote the unknown mean vector and the unknown covariance matrix of ith population, $i = 1, \cdot\cdot\cdot, k$. Further let the jth observation from the ith population, and $j = 1, 2, \cdot\cdot\cdot, n_i$ be denoted by X_{ij} and the ith sample mean vector by $\overline{X}_i$. Define the hypothesis matrix

$$H = \sum_{i=1}^{k} n_i(\overline{X}_i - \overline{X})(\overline{X}_i - \overline{X})',$$

where $\overline{X} = \sum_{i=1}^{k} n_i \overline{X}_i / (\sum_{i=1}^{k} n_i)$, and the error matrix

$$E = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)(X_{ij} - \overline{X}_i)'.$$

Then, the usual test statistics to test

$$H_0 : \xi_1 = \cdot\cdot\cdot = \xi_k$$

are some real valued functions of the eigenvalues of HE^{-1} (Morrison 1990). For example, Roy's root test statistic is $\theta_s = c_s/(1 + c_s)$, where c_s is the largest ei-genvalue of HE^{-1}among the s nonzero eigenvalues of HE^{-1}. The Lawley-Hotelling test statistic is $\mathrm{tr}(HE^{-1})$. Again, the question of interest is whether we get differen t test procedures if we use hypothesis-dependent error matrix

$$E_0 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \overline{X})(X_{ij} - \overline{X})'$$

instead of E. We show that two commonly used real valued functions of the eigenvalues of HE_0^{-1}, namely the largest eigenvalue and the trace of HE_0^{-1}, lead to the well-known test procedures available in the literature.

First let us consider the largest eigenvalue of HE_0^{-1} as a test statistic to test H_0. Noting the relation that $E_0 = H + E$, it can be easily shown that the largest eigenvalue of HE_0^{-1} is indeed Roy's root test statistic θ_s. If we propose $\mathrm{tr}(HE_0^{-1})$ as a test statistic, then this leads to a different test than the Lawley-Hotelling test. However, this is the well-known Pillai's trace test statistic, as

$$\mathrm{tr}(HE_0^{-1}) = \mathrm{tr}[H(H+E)^{-1}].$$

REFERENCES

Good, I. J. (1986), "Editorial Note on C257 Regarding the t-Test" , *Journal of Statistical Computation and Simulation,* 25, 296-297.

Kshirsagar, A. M. (1972), *Multivariate Analysis,* New York : Marcel Dekker.

Lefante, J. J., and Shah, A. K. (1986), "A Note on the One-Sample t-Test," *Journal of Statistical Computation and Simulation,* 25, 295-296.

Morrison, D. F. (1990), *Multivariate Statistical Methods* (3rd ed), New York : McGraw-Hill.