A Study on Multiple Comparison Procedure based on Multivariate Observations

Tomohiro Nakamura
Makoto Doi
Hideyuki Douke
Katsumi Ujiie
School of Science, Tokai University, Kanagawa, 259-1292 *Japan*
msdoi@sm.u-tokai.ac.jp

1. Introduction

The multiple comparison procedures have been proposed by Tukey(1953), Scheffé (1953), Dunnett(1955) and so on. So far, the multiple comparison procedures are to test a significance difference between two treatments chosen among several treatments based on only one response. It is more appropriate in many cases that we deal with the p kinds of responses rather than only one response for measuring the effect of a treatment. Thus we can adopt the multiple comparison procedure involved the Union-Intersection method in Hochberg and Tamhane(1987) based on multivariate observations. There, however, still exists an impractical problem in which it is not easy to calculate the distribution of a maximum characteristic value.

In this study we use the χ^2 statistic for testing the difference of two treatments chosen among several treatments. For obtaining theoretically the density function of the maximum value among these χ^2 statistics, we will derive the joint probability density function of these statistics based on these χ^2 statistics by using the change of variables. Finally we can obtain the joint probability density function of statistic based on the maximum χ^2 statistic. And we make test hypotheses by using the likelihood ratio test statistic based on this joint probability density function. Then we investigate in the simulations how the confidence limit values change with the number of variables.

2. Hypotheses and simultaneous confidence regions

We suppose that the responses on each of k treatments are distributed according to $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ $(i = 1, 2, \cdots, k)$ with a mean vector $\boldsymbol{\mu}_i(p \times 1)$ and a known covariance matrix $\boldsymbol{\Sigma}_i$. We make the hypotheses $H_{\{i,j\}}\!:\boldsymbol{\mu}_i = \boldsymbol{\mu}_j$, $K_{\{i,j\}}\!:\boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j$ for the i, j $(i = 1, 2, \cdots, k-1,\ j = 2, 3, \cdots, k)$. Under these hypotheses, we test the null hypothesis H_0 against the alternative hypothesis K as $H_0 = \bigcap_{i<j} H_{\{i,j\}}$ for all the combinations of the i, j, $K = \bigcup_{i<j} K_{\{i,j\}}$ for at least one combination of the i, j in all the combinations of the i, j. When we draw the random variable vectors $\boldsymbol{x}_1^{(i)}, \boldsymbol{x}_2^{(i)}, \cdots, \boldsymbol{x}_{N_i}^{(i)}$ of the sample size N_i from each population, the sample mean vector $\bar{\boldsymbol{x}}_i(p \times 1)$ of each population is distributed according to $N_p(\boldsymbol{\mu}_i, \frac{\boldsymbol{\Sigma}_i}{N_i})$. The statistic

$$\chi^2_{i,j} = \{(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_j) - (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)\}' \left(\frac{\boldsymbol{\Sigma}_i}{N_i} + \frac{\boldsymbol{\Sigma}_j}{N_j} \right)^{-1} \{(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_j) - (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)\}$$

has a χ^2-distribution with p degrees of freedom. Under the hypothesis $H_{\{i,j\}}$: $\boldsymbol{\mu}_i = \boldsymbol{\mu}_j$, a confidence region of χ^2 is given by $D_{i,j} = [\chi^2_{i,j} \leq t]$. Therefore, under the null hypothesis $H_0 = \bigcap_{i<j} H_{\{i,j\}}$, the simultaneous confidence region is given by

$$\bigcap_{i<j} D_{i,j} = \bigcap_{i<j} [\chi^2_{i,j} \leq t] = [\max_{i<j}(\chi^2_{i,j}) \leq t]$$

where t is a critical value. Then we find a distribution of the maximum value of $\chi^2_{1,2}, \chi^2_{1,3}, \cdots, \chi^2_{k-1,k}$ (the number of these is ${}_kC_2 = m$).

3. The Distribution of $\max_{i<j}(\boldsymbol{z}_{i,j})$

If we denote by $\boldsymbol{z}_{i,j}(p \times 1) = (\ z_{i,j,1},\ z_{i,j,2},\ \cdots,\ z_{i,j,p}\ )'$ under the hypothesis $H_{\{i,j\}}$: $\boldsymbol{\mu}_i = \boldsymbol{\mu}_j$, the χ^2 statistic is given by $\chi^2_{i,j} = z^2_{i,j,1} + z^2_{i,j,2} + \cdots + z^2_{i,j,p}$. Since $\boldsymbol{z}_{1,2}, \boldsymbol{z}_{1,3}, \cdots, \boldsymbol{z}_{k-1,k}$ are not mutually independent, $\boldsymbol{z}(mp \times 1) = (\boldsymbol{z}'_{1,2}, \boldsymbol{z}'_{1,3}, \cdots, \boldsymbol{z}'_{k-1,k})'$ is distributed according to $N_{mp}(\mathbf{0}, \boldsymbol{\Sigma})$ with a known covariance matrix $\boldsymbol{\Sigma}(mp \times mp)$. Then we can give the joint probability density function of $\boldsymbol{z}$ as

$$f(\boldsymbol{z}) = \frac{1}{(2\pi)^{\frac{mp}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}\boldsymbol{z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}} .$$

If we assume $\max_{i<j}(\chi^2_{i,j}) = \chi^2_{k-1,k}$ tentatively, we can derive the joint probability density function of $\boldsymbol{z}_{k-1,k}$ by using the marginal density function of $\boldsymbol{z}_{1,2}, \boldsymbol{z}_{1,3}, \cdots, \boldsymbol{z}_{k-2,k}$ after doing the appropriate change of variables. Then the joint probability density function of $\boldsymbol{z}_{k-1,k}$ is given by

$$f(\boldsymbol{z}_{k-1,k}) = \frac{1}{(2\pi)^{\frac{p}{2}}} e^{-\frac{1}{2}(\boldsymbol{z}'_{k-1,k}\boldsymbol{z}_{k-1,k})} .$$

Here, we get the following likelihood ratio test statistic based on the joint probability density function above.

$$\lambda(\boldsymbol{z}_{k-1,k}) = e^{-\frac{1}{2}\boldsymbol{z}'_{k-1,k}\boldsymbol{z}_{k-1,k}} .$$

Using the likelihood ratio test method, the critical value h must be chosen so that

$$\mathrm{P}(0 \leq \lambda(\boldsymbol{z}_{k-1,k}) \leq h) = \mathrm{P}(\chi^2_{k-1,k} \geq t) = \alpha$$

for the significance level α. Here, t is the upper $100\alpha\%$ point of the χ^2 distribution with p degrees of freedom and $t = -2\log h$.

Bibliography

[1] Dunnett, C. W.(1955) : "A multiple comparison procedure for comparing several treatments with a control.", *Journal of the American Statistical Association* 50, pp1096-1121.
[2] Hochberg, Y. and Tamhane, A. C.(1987) : *Multiple Comparison Procedures*, John Wiley & Sons.
[3] Scheffé, H.(1953) : "A method for judging all contrasts in the analysis of variance.", *Biometrika* 40, pp87-104.
[4] Tukey, J. W.(1953) : *The Problem of Multiple comparisons*, Mimeographed monograph.