

APPROPRIATE CRITICAL VALUES FOR PAIRWISE COMPARISONS OF MULTIPLE COMPARISON PROCEDURES FOR NORMAL VARIANCES WHEN SAMPLE SIZES ARE UNBALANCED

Imada, Tsunehisa
Department of Management, Tokai University

<https://doi.org/10.5109/4102449>

出版情報 : Bulletin of informatics and cybernetics. 52 (2), pp.1-15, 2020. 統計科学研究会
バージョン :
権利関係 :



APPROPRIATE CRITICAL VALUES FOR PAIRWISE COMPARISONS
OF MULTIPLE COMPARISON PROCEDURES FOR NORMAL
VARIANCES WHEN SAMPLE SIZES ARE UNBALANCED

by

Tsunehisa IMADA

*Reprinted from the Bulletin of Informatics and Cybernetics
Research Association of Statistical Sciences, Vol.52, No. 2*

FUKUOKA, JAPAN
2020

APPROPRIATE CRITICAL VALUES FOR PAIRWISE COMPARISONS OF MULTIPLE COMPARISON PROCEDURES FOR NORMAL VARIANCES WHEN SAMPLE SIZES ARE UNBALANCED

By

Tsunehisa IMADA*

Abstract

In this study we discuss the determination of appropriate critical values for pairwise comparisons of the multiple comparison with a control and the all-pairwise multiple comparison for normal variances when sample sizes are unbalanced. Although Imada (2018A, 2018B) used a common critical value for all hypotheses, the probability that the true hypothesis is rejected is not uniform. Here, we construct the critical value for each hypothesis so that the probability that it is rejected when it is true may be as uniform as possible. Finally, we give some numerical results regarding critical values and power of the test.

Key Words and Phrases: Allocation of Type I error, Conservative critical value, Exact critical value.

1. Introduction

Assume there are independent normal random variables X_1, X_2, \dots, X_K and X_k is distributed according to normal $N(\mu_k, \sigma_k^2)$ for $k = 1, 2, \dots, K$. Imada (2018A, 2018B, 2019) discussed multiple comparison procedures for $\sigma_1^2, \sigma_2^2, \dots, \sigma_K^2$. Specifically, Imada (2018A) discussed a multiple comparison with a control for comparing σ_1^2 with $\sigma_2^2, \sigma_3^2, \dots, \sigma_K^2$ simultaneously and an all-pairwise multiple comparison for $\sigma_1^2, \sigma_2^2, \dots, \sigma_K^2$ based on the single step procedures (cf. Dunnett (1955) and Tukey (1953)). For the multiple comparison with a control Imada (2018A) determined the critical value for pairwise comparison satisfying a specified significance level exactly and formulated the power of the test under a specified alternative hypothesis. For the all-pairwise multiple comparison Imada (2018A) determined two kinds of conservative critical values for pairwise comparison for a specified significance level using Bonferroni's inequality and the improved Bonferroni's inequality and calculated the power of the test by Monte Carlo simulation. Then, Imada (2018B) determined the critical value for pairwise comparison of the all-pairwise multiple comparison satisfying a specified significance level exactly. Furthermore, intended to obtain higher power compared to the single step procedures Imada (2019) constructed a sequentially rejective step down procedure (cf. Dunnett and Tamhane (1991)) and a step up procedure (cf. Dunnett and Tamhane (1992)) for the

* Department of Management, Tokai University, 9-1-1 Toroku Higashi-ku Kumamoto, Japan 862-8652 Japan. tel +81-96-386-2731, timada@ktmail.tokai-u.jp

multiple comparison with a control and constructed a closed testing procedure called the Ryan-Einot-Gabriel-Welsch procedure (cf. Ryan (1960), Einot and Gabriel (1975) and Welsch (1977)) for the all-pairwise multiple comparison.

Imada (2018A, 2018B) used a common critical value for all hypotheses in both multiple comparison procedures. If sample sizes are unbalanced and all hypotheses are true, the statistic for pairwise comparison is distributed according to an F -distribution and its degrees of freedom vary as the hypothesis to be tested varies. Specifically, the probability that the true hypothesis is rejected is not uniform, which means that the hypotheses to be tested are not treated equally. In this study we construct the critical value for each hypothesis so that the probability that it is rejected when it is true may be as uniform as possible for unbalanced sample sizes. Finally, we give some numerical results regarding critical values and power of the test.

2. Multiple comparison with a control

We consider a multiple comparison with a control for comparing σ_1^2 with $\sigma_2^2, \sigma_3^2, \dots, \sigma_K^2$ simultaneously. For pairwise comparison we consider one-sided and two-sided tests. For the one-sided test we set up a null hypothesis and its alternative hypothesis as

$$H_{1,k}^{(1)} : \sigma_1^2 = \sigma_k^2 \text{ vs. } H_{1,k}^{(1)A} : \sigma_1^2 < \sigma_k^2 \text{ for } k = 2, 3, \dots, K. \quad (1)$$

For the two-sided test we set up a null hypothesis and its alternative hypothesis as

$$H_{1,k}^{(2)} : \sigma_1^2 = \sigma_k^2 \text{ vs. } H_{1,k}^{(2)A} : \sigma_1^2 \neq \sigma_k^2 \text{ for } k = 2, 3, \dots, K. \quad (2)$$

2.1. Procedure proposed by Imada (2018A)

First, we discuss the procedure proposed by Imada (2018A) for testing $H_{1,2}^{(i)}, H_{1,3}^{(i)}, \dots, H_{1,K}^{(i)}$ for $i = 1, 2$ simultaneously using a sample $x_{k1}, x_{k2}, \dots, x_{kn_k}$ from $N(\mu_k, \sigma_k^2)$ for $k = 1, 2, \dots, K$. Letting

$$\bar{x}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_{ki}, \quad \nu_k^2 = \frac{\sum_{i=1}^{n_k} (x_{ki} - \bar{x}_k)^2}{n_k - 1}$$

for $k = 1, 2, \dots, K$, we use the statistic

$$F_{1,k} = \frac{\nu_k^2}{\nu_1^2}$$

for testing $H_{1,k}^{(i)}$ for $i = 1, 2$. First, we consider (1). Iff $F_{1,k}$ is greater than a specified critical value, $H_{1,k}^{(1)}$ is rejected. Imada (2018A) determined a common critical value c for all the $H_{1,k}^{(1)}$ s. We determine c so that

$$P(\max_{2 \leq k \leq K} F_{1,k} > c) = \alpha$$

for a specified significance level α under the assumption that $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_K^2$. Imada (2018A) derived

$$P(\max_{2 \leq k \leq K} F_{1,k} > c) = 1 - \int_0^\infty f_1(x_1) \left\{ \prod_{k=2}^K \int_0^{c\lambda_{1,k}x_1} f_k(x_k) dx_k \right\} dx_1$$

where $f_k(x_k)$ denotes the probability density function of χ^2 -distribution with $n_k - 1$ degrees of freedom for $k = 1, 2, \dots, K$ and

$$\lambda_{1,k} = \frac{n_k - 1}{n_1 - 1}$$

for $k = 2, 3, \dots, K$.

Next, we consider (2). We set up upper and lower critical values for testing $H_{1,k}^{(2)}$. Iff $F_{1,k}$ is greater than the upper critical value or smaller than the lower critical value, $H_{1,k}^{(2)}$ is rejected. Imada (2018A) determined a common lower critical value c_1 and a common upper critical value c_2 for all the $H_{1,k}^{(2)}$ s for simplicity. Iff $F_{1,k} < c_1$ or $c_2 < F_{1,k}$, we reject $H_{1,k}^{(2)}$. Since

$$F_{1,k} < c_1 \text{ or } c_2 < F_{1,k} \Leftrightarrow F_{1,k}^{-1} < c_2^{-1} \text{ or } c_1^{-1} < F_{1,k}^{-1},$$

we restrict c_1, c_2 as

$$c_2 = c_1^{-1} = c > 1.$$

Then, we obtain

$$F_{1,k} < c^{-1} \text{ or } c < F_{1,k} \Leftrightarrow F_{1,k}^{-1} < c^{-1} \text{ or } c < F_{1,k}^{-1}.$$

Letting

$$G_{1,k} = \max\{F_{1,k}, F_{1,k}^{-1}\},$$

we obtain

$$F_{1,k} < c^{-1} \text{ or } c < F_{1,k} \Leftrightarrow G_{1,k} > c.$$

We determine c so that

$$P(\max_{2 \leq k \leq K} G_{1,k} > c) = \alpha$$

for a specified significance level α under the assumption that $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_K^2$. Imada (2018A) derived

$$P(\max_{2 \leq k \leq K} G_{1,k} > c) = 1 - \int_0^\infty f_1(x_1) \left\{ \prod_{k=2}^K \int_{c^{-1}\lambda_{1,k}x_1}^{c\lambda_{1,k}x_1} f_k(x_k) dx_k \right\} dx_1.$$

If $n_2 = n_3 = \dots = n_K$, $P(F_{1,2} > c) = P(F_{1,3} > c) = \dots = P(F_{1,K} > c)$ and $P(G_{1,2} > c) = P(G_{1,3} > c) = \dots = P(G_{1,K} > c)$ under the assumption that $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_K^2$. Specifically, the probability that $H_{1,k}^{(i)}$ is rejected is uniform for $i = 1, 2$. If n_2, n_3, \dots, n_K are unbalanced, the probability that $H_{1,k}^{(i)}$ is rejected varies as k varies. It is preferable to set up the critical value c_k for each $H_{1,k}^{(i)}$ so that the probability that $H_{1,k}^{(i)}$ is rejected may be uniform.

2.2. Determination of appropriate critical value for each hypothesis

Assume n_2, n_3, \dots, n_K are unbalanced. First, we consider (1). We want to determine c_2, c_3, \dots, c_K so that

$$P(F_{1,k} > c_k \text{ for at least one } k) = \alpha \quad (3)$$

and

$$P(F_{1,2} > c_2) = P(F_{1,3} > c_3) = \cdots = P(F_{1,K} > c_K) \quad (4)$$

under the assumption that $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_K^2$. The left hand side of (3) is formulated as

$$1 - \int_0^\infty f_1(x_1) \left\{ \prod_{k=2}^K \int_0^{c_k \lambda_{1,k} x_1} f_k(x_k) dx_k \right\} dx_1. \quad (5)$$

However, it is difficult to determine c_2, c_3, \dots, c_K satisfying (3) and (4). On the other hand, if we determine $c_{k,(1)}$ so that

$$P(F_{1,k} > c_{k,(1)}) = \frac{\alpha}{K-1}$$

using

$$P(F_{1,k} > c_{k,(1)}) = \int_{c_{k,(1)}}^\infty f_{1,k}(v) dv$$

where $f_{1,k}(v)$ is the probability density function of the F -distribution with $(n_k - 1, n_1 - 1)$ degrees of freedom, we obtain

$$P(F_{1,k} > c_{k,(1)} \text{ for at least one } k) \leq \alpha$$

by Bonferroni's inequality

$$P(F_{1,k} > c_k \text{ for at least one } k) \leq \sum_{k=2}^K P(F_{1,k} > c_k).$$

Furthermore, if we determine $c_{k,(2)}$ so that

$$P(F_{1,k} > c_{k,(2)}) = 1 - (1 - \alpha)^{1/(K-1)},$$

we obtain

$$P(F_{1,k} > c_{k,(2)} \text{ for at least one } k) \leq \alpha$$

by the inequality

$$P(F_{1,k} > c_k \text{ for at least one } k) = 1 - P(F_{1,k} \leq c_k \text{ for all } k) \leq 1 - \prod_{k=2}^K P(F_{1,k} \leq c_k)$$

derived by Corollary A.1.1 in Hsu (1996). Although $c_{2,(i)}, c_{3,(i)}, \dots, c_{K,(i)}$ for $i = 1, 2$ are conservative for α , they satisfy (4). Since

$$1 - (1 - \alpha)^{1/(K-1)} > \frac{\alpha}{K-1}$$

for $K > 2$, $c_{2,(2)}, c_{3,(2)}, \dots, c_{K,(2)}$ are less conservative compared to $c_{2,(1)}, c_{3,(1)}, \dots, c_{K,(1)}$. Next, letting $c_{k,(3)} = c_{k,(2)}\psi$ for $k = 2, 3, \dots, K$ where $0 < \psi < 1$, we determine ψ so that

$$P(F_{1,k} > c_{k,(3)} \text{ for at least one } k) = \alpha$$

using (5). Although $c_{2,(3)}, c_{3,(3)}, \dots, c_{K,(3)}$ satisfy the specified significance level exactly, the probability that $H_{1,k}^{(i)}$ is rejected is not uniform under the assumption that $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_K^2$. For (2) we can determine $c_{2,(i)}, c_{3,(i)}, \dots, c_{K,(i)}$ for $i = 1, 2, 3$ similarly.

2.3. Power of the test

Next, we consider the power of the test. We discuss the formulation of power by Imada (2018A) using a common critical value c for all hypotheses. First, we consider the power of the test for (1). Assume

$$\sigma_1^2 = \gamma_{1,2}\sigma_2^2 = \gamma_{1,3}\sigma_3^2 = \cdots = \gamma_{1,l}\sigma_l^2 \text{ and } \sigma_1^2 = \sigma_m^2 \text{ for } m = l+1, l+2, \dots, K \quad (6)$$

where $0 < \gamma_{1,2} < 1, 0 < \gamma_{1,3} < 1, \dots, 0 < \gamma_{1,l} < 1$. We focus on the all-pairs power defined by Ramsey (1978). If $l = 2$, the power of the test under (6) is

$$P(F_{1,2} > c) = \int_{c\gamma_{1,2}}^{\infty} f_{1,2}(x)dx.$$

If $l > 2$, the power of the test under (6) is

$$P(F_{1,i} > c \text{ for } i = 2, 3, \dots, l) = \int_0^{\infty} f_1(x_1) \left\{ \prod_{i=2}^l \int_{c\lambda_{1,i}\gamma_{1,i}x_1}^{\infty} f_i(x_i)dx_i \right\} dx_1.$$

Next, we consider the power of the test for (2). Assume

$$\sigma_1^2 = \gamma_{1,2}\sigma_2^2 = \gamma_{1,3}\sigma_3^2 = \cdots = \gamma_{1,l}\sigma_l^2 \text{ and } \sigma_1^2 = \sigma_m^2 \text{ for } m = l+1, l+2, \dots, K \quad (7)$$

where $\gamma_{1,2} \neq 1, \gamma_{1,3} \neq 1, \dots, \gamma_{1,l} \neq 1$. If $l = 2$, the power of the test under (7) is

$$P(G_{1,2} > c) = 1 - \int_{c^{-1}\gamma_{1,2}}^{c\gamma_{1,2}} f_{1,2}(v)dv.$$

If $l > 2$, the power of the test under (7) is

$$P(G_{1,i} > c \text{ for } i = 2, 3, \dots, l) = \int_0^{\infty} f_1(x_1) \prod_{i=2}^l \left\{ 1 - \int_{c^{-1}\lambda_{1,i}\gamma_{1,i}x_1}^{c\lambda_{1,i}\gamma_{1,i}x_1} f_i(x_i)dx_i \right\} dx_1.$$

We can obtain the formulation of power similarly when $c_{2,(i)}, c_{3,(i)}, \dots, c_{K,(i)}$ for $i = 1, 2, 3$ are used.

3. All-pairwise multiple comparison for normal variances

We consider the all-pairwise multiple comparison for $\sigma_1^2, \sigma_2^2, \dots, \sigma_K^2$. To compare σ_k^2 and σ_l^2 for $1 \leq k < l \leq K$ we set up a null hypothesis and its alternative hypothesis as

$$H_{k,l} : \sigma_k^2 = \sigma_l^2 \text{ vs. } H_{k,l}^A : \sigma_k^2 \neq \sigma_l^2$$

and consider the simultaneous test of all the $H_{k,l}$ s.

3.1. Procedure proposed by Imada (2018A, 2018B)

We consider the single step procedure for the $H_{k,l}$ s discussed by Imada (2018A, 2018B). The statistic

$$F_{k,l} = \frac{\nu_l^2}{\nu_k^2}$$

is used for testing $H_{k,l}$ and a common critical value c is set up for all the $H_{k,l}$ s. Iff $F_{k,l} < c^{-1}$ or $c < F_{k,l}$, we reject $H_{k,l}$. Letting

$$G_{k,l} = \max\{F_{k,l}, F_{k,l}^{-1}\},$$

we obtain

$$F_{k,l} < c^{-1} \text{ or } c < F_{k,l} \Leftrightarrow G_{k,l} > c.$$

We want to determine c so that

$$P(\max_{1 \leq k < l \leq K} G_{k,l} > c) = \alpha$$

for a specified significance level α under the assumption that $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_K^2$. Letting

$$\lambda_{k_1, k_2} = \frac{n_{k_2} - 1}{n_{k_1} - 1}$$

for each (k_1, k_2) pair chosen from $1, 2, \dots, K$, Imada (2018B) derived

$$\begin{aligned} & P(\max_{1 \leq k < l \leq K} G_{k,l} > c) \\ &= 1 - \sum_{k_1, k_2} \int_0^\infty \int_{\lambda_{k_1, k_2} x_1}^{c \lambda_{k_1, k_2} x_1} \left\{ \prod_{l \neq k_1, k_2} \int_{\lambda_{k_1, l} x_1}^{\lambda_{k_2, l} x_2} f_l(x) dx \right\} f_{k_2}(x_2) dx_2 f_{k_1}(x_1) dx_1. \end{aligned}$$

If n_1, n_2, \dots, n_K are unbalanced, the probability that $H_{k,l}$ is rejected varies as k, l vary under the assumption that $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_K^2$. It is preferable to set up the critical value $c_{k,l}$ for each $H_{k,l}$ so that the $P(G_{k,l} > c_{k,l})$ s may be uniform under the assumption that $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_K^2$.

3.2. Determination of appropriate critical value for each hypothesis

It is difficult to determine the $c_{k,l}$ s so that $P(\max_{1 \leq k < l \leq K} G_{k,l} > c_{k,l}) = \alpha$ and the $P(G_{k,l} > c_{k,l})$ s may be uniform under the assumption that $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_K^2$. However, we can determine the conservative $c_{k,l}$ s easily using Bonferroni's inequality

$$P(G_{k,l} > c_{k,l} \text{ for at least one pair } (k, l)) \leq \sum_{1 \leq k < l \leq K} P(G_{k,l} > c_{k,l}). \quad (8)$$

We determine $c_{k,l,(1)}$ so that

$$P(G_{k,l} > c_{k,l,(1)}) = \frac{2\alpha}{K(K-1)}. \quad (9)$$

Here

$$P(G_{k,l} > c_{k,l,(1)}) = 1 - \int_{c_{k,l,(1)}^{-1}}^{c_{k,l,(1)}} f_{k,l}(v) dv$$

where $f_{k,l}(v)$ is the probability density function of the F -distribution with $(n_l - 1, n_k - 1)$ degrees of freedom. We obtain

$$P(G_{k,l} > c_{k,l,(1)} \text{ for at least one pair } (k, l)) \leq \alpha$$

by (8) and (9). On the other hand, by the improved Bonferroni's inequality we obtain

$$\begin{aligned}
& P(G_{k,l} > c_{k,l} \text{ for at least one pair } (k,l)) \\
\leq & \sum_{1 \leq k < l \leq K} P(G_{k,l} > c_{k,l}) - \sum_{l=1}^{K-2} \sum_{k=1}^{K-l-1} P(G_{k,k+l} > c_{k,k+l}, G_{k+1,k+l+1} > c_{k+1,k+l+1}) \\
& - \sum_{l=1}^{K-2} P(G_{K-l,K} > c_{K-l,K}, G_{1,l+2} > c_{1,l+2}).
\end{aligned} \tag{10}$$

Refer to Imada (2018A) regarding the formulations of the probabilities

$$P(G_{k,k+l} > c_{k,k+l}, G_{k+1,k+l+1} > c_{k+1,k+l+1}), P(G_{K-l,K} > c_{K-l,K}, G_{1,l+2} > c_{1,l+2}).$$

Letting $c_{k,l,(2)} = c_{k,l,(1)}\psi$ for $1 \leq k < l \leq K$ where $0 < \psi < 1$, we determine ψ so that

$$\begin{aligned}
& \sum_{1 \leq k < l \leq K} P(G_{k,l} > c_{k,l,(2)}) - \sum_{l=1}^{K-2} \sum_{k=1}^{K-l-1} P(G_{k,k+l} > c_{k,k+l,(2)}, G_{k+1,k+l+1} > c_{k+1,k+l+1,(2)}) \\
& - \sum_{l=1}^{K-2} P(G_{K-l,K} > c_{K-l,K,(2)}, G_{1,l+2} > c_{1,l+2,(2)}) = \alpha.
\end{aligned}$$

Then, we obtain

$$P(G_{k,l} > c_{k,l,(2)} \text{ for at least one pair } (k,l)) \leq \alpha$$

by (10). Although the $c_{k,l,(2)}$ s are less conservative for α compared to the $c_{k,l,(1)}$ s, the $P(G_{k,l} > c_{k,l,(2)})$ s are not uniform under the assumption that $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_K^2$.

3.3. Power of the test

It is difficult to formulate the power of the single step procedure under a specified alternative hypothesis. We calculate the power using Monte Carlo simulation.

4. Numerical examples

In this Section we give some numerical examples regarding critical values and power of the test.

First, we consider the multiple comparison with a control. The method using common c is denoted by M. The method using $c_{k,(i)}$ is denoted by Mi for $i = 1, 2, 3$. Here, we consider only (1). Let $K = 4$ and $\alpha = 0.05$. We set up two types of sample sizes as

$$\text{Sam.1 : } (n_1, n_2, n_3, n_4) = (20, 15, 20, 25), \text{ Sam.2 : } (n_1, n_2, n_3, n_4) = (20, 10, 20, 30).$$

The common critical value of M for Sam.1 is $c = 2.627$ and that for Sam.2 is $c = 2.727$. Then

$$P(F_{1,2} > c) = 0.0259, P(F_{1,3} > c) = 0.0207, P(F_{1,4} > c) = 0.0177$$

for Sam.1 and

$$P(F_{1,2} > c) = 0.0269, P(F_{1,3} > c) = 0.0149, P(F_{1,4} > c) = 0.0111$$

for Sam.2 under $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$. These results show that the $P(F_{1,k} > c)$ s are not uniform and $P(F_{1,k} > c)$ is smaller as n_k is larger. On the other hand, we obtain

$$\begin{aligned} c_{2,(1)} &= 2.885, \quad c_{3,(1)} = 2.744, \quad c_{4,(1)} = 2.658, \\ c_{2,(2)} &= 2.875, \quad c_{3,(2)} = 2.735, \quad c_{4,(2)} = 2.649, \\ c_{2,(3)} &= 2.740, \quad c_{3,(3)} = 2.606, \quad c_{4,(3)} = 2.525 \end{aligned}$$

for Sam.1 and

$$\begin{aligned} c_{2,(1)} &= 3.159, \quad c_{3,(1)} = 2.744, \quad c_{4,(1)} = 2.599, \\ c_{2,(2)} &= 3.147, \quad c_{3,(2)} = 2.735, \quad c_{4,(2)} = 2.591, \\ c_{2,(3)} &= 3.005, \quad c_{3,(3)} = 2.612, \quad c_{4,(3)} = 2.474 \end{aligned}$$

for Sam.2. $c_{2,(1)}, c_{3,(1)}, c_{4,(1)}$ and $c_{2,(2)}, c_{3,(2)}, c_{4,(2)}$ are conservative. For Sam.1 the Type I error of M1 is

$$P(F_{1,k} > c_{k,(1)} \text{ for at least one } k) = 0.0392$$

and that of M2 is

$$P(F_{1,k} > c_{k,(2)} \text{ for at least one } k) = 0.0398.$$

For Sam.2 the Type I error of M1 is

$$P(F_{1,k} > c_{k,(1)} \text{ for at least one } k) = 0.0396$$

and that of M2 is

$$P(F_{1,k} > c_{k,(2)} \text{ for at least one } k) = 0.0404.$$

$P(F_{1,2} > c_{2,(3)}), P(F_{1,3} > c_{3,(3)}), P(F_{1,4} > c_{4,(3)})$ are not uniform under $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$. Specifically

$$P(F_{1,2} > c_{2,(3)}) = 0.0213, \quad P(F_{1,3} > c_{3,(3)}) = 0.0215, \quad P(F_{1,4} > c_{4,(3)}) = 0.0216$$

for Sam.1 and

$$P(F_{1,2} > c_{2,(3)}) = 0.0208, \quad P(F_{1,3} > c_{3,(3)}) = 0.0213, \quad P(F_{1,4} > c_{4,(3)}) = 0.0215$$

for Sam.2. These results show the three probabilities are relatively uniform. Next, we consider the power of the test. Letting γ be a positive constant which is less than 1, we set up 7 cases of alternative hypotheses as follows.

- Case 1. $\sigma_1^2 = \gamma\sigma_2^2 = \sigma_3^2 = \sigma_4^2$,
- Case 2. $\sigma_1^2 = \sigma_2^2 = \gamma\sigma_3^2 = \sigma_4^2$,
- Case 3. $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \gamma\sigma_4^2$,
- Case 4. $\sigma_1^2 = \gamma\sigma_2^2 = \gamma\sigma_3^2 = \sigma_4^2$,
- Case 5. $\sigma_1^2 = \gamma\sigma_2^2 = \sigma_3^2 = \gamma\sigma_4^2$,
- Case 6. $\sigma_1^2 = \sigma_2^2 = \gamma\sigma_3^2 = \gamma\sigma_4^2$,
- Case 7. $\sigma_1^2 = \gamma\sigma_2^2 = \gamma\sigma_3^2 = \gamma\sigma_4^2$.

The power of Cases 1 to 3 is the probability that the single hypothesis is rejected. The power of Cases 4 to 6 is the probability that the two hypotheses are rejected. The power of Case 7 is the probability that all hypotheses are rejected. Tables 1 and 2 give the power of Cases 1 to 7 for $\gamma = 0.5, 0.25$ for Sam.1. Tables 3 and 4 give them for Sam.2. Although we should not judge the superiority/inferiority between M and M_i for $i = 1, 2, 3$ in terms of power because of differences in the allocation of Type I error, we investigate power variations for the procedures. When the number of hypotheses to be rejected is specified, the power by M is relatively uniform. For example, we discuss the power of Cases 1 to 3. The power detecting the difference between σ_1^2 and σ_k^2 ($k \geq 2$) decreases as n_k increases from 10 to 30 for M. However, the power detecting the difference between σ_1^2 and σ_k^2 ($k \geq 2$) increases as n_k increases from 10 to 30 for M1, M2 and M3.

Table 1: Power for Cases 1 to 5 for Sam.1

Case	1		2		3		4		5	
γ	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25
M	0.285	0.787	0.279	0.816	0.274	0.836	0.144	0.677	0.147	0.693
M1	0.225	0.731	0.248	0.791	0.265	0.830	0.113	0.620	0.122	0.646
M2	0.227	0.733	0.251	0.793	0.245	0.812	0.114	0.624	0.123	0.649
M3	0.257	0.762	0.285	0.821	0.305	0.857	0.136	0.662	0.146	0.688

Table 2: Power for Cases 6 and 7 for Sam.1

Case	6		7	
γ	0.50	0.25	0.50	0.25
M	0.150	0.719	0.096	0.614
M1	0.136	0.696	0.078	0.566
M2	0.138	0.699	0.079	0.569
M3	0.164	0.737	0.097	0.612

Table 3: Power for Cases 1 to 5 for Sam.2

Case	1		2		3		4		5	
γ	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25
M	0.271	0.717	0.252	0.794	0.243	0.828	0.121	0.605	0.124	0.630
M1	0.192	0.629	0.248	0.791	0.279	0.856	0.094	0.539	0.107	0.575
M2	0.194	0.632	0.251	0.792	0.281	0.858	0.095	0.541	0.108	0.577
M3	0.217	0.660	0.283	0.819	0.319	0.881	0.112	0.577	0.127	0.612

Table 4: Power for Cases 6 and 7 for Sam.2

Case	6		7	
γ	0.50	0.25	0.50	0.25
M	0.133	0.701	0.079	0.551
M1	0.144	0.715	0.068	0.504
M2	0.145	0.718	0.069	0.507
M3	0.171	0.754	0.084	0.546

Next, we consider the all-pairwise comparison. The method using common c is denoted by M. The method using the $c_{k,l,(i)}$ s is denoted by M_i for $i = 1, 2$. Let $K = 4$ and $\alpha = 0.05$. We set up two types of sample sizes as

Sam.1 : $(n_1, n_2, n_3, n_4) = (15, 20, 20, 25)$, Sam.2 : $(n_1, n_2, n_3, n_4) = (10, 15, 25, 30)$.

The common critical value of M for Sam.1 is $c = 3.503$ and that for Sam.2 is $c = 4.134$. Then

$$\begin{aligned} P(F_{1,2} > c) &= 0.0165, \quad P(F_{1,3} > c) = 0.0165, \quad P(F_{1,4} > c) = 0.0125, \\ P(F_{2,3} > c) &= 0.0089, \quad P(F_{2,4} > c) = 0.0057, \quad P(F_{3,4} > c) = 0.0057 \end{aligned}$$

for Sam.1 and

$$\begin{aligned} P(F_{1,2} > c) &= 0.0190, \quad P(F_{1,3} > c) = 0.0160, \quad P(F_{1,4} > c) = 0.0152, \\ P(F_{2,3} > c) &= 0.0040, \quad P(F_{2,4} > c) = 0.0036, \quad P(F_{3,4} > c) = 0.0004 \end{aligned}$$

for Sam.2 under $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$. On the other hand,

$$\begin{aligned} c_{1,2,(1)} &= 4.014, \quad c_{1,3,(1)} = 4.014, \quad c_{1,4,(1)} = 3.784, \\ c_{2,3,(1)} &= 3.543, \quad c_{2,4,(1)} = 3.300, \quad c_{3,4,(1)} = 3.300, \\ c_{1,2,(2)} &= 3.954, \quad c_{1,3,(2)} = 3.954, \quad c_{1,4,(2)} = 3.727, \\ c_{2,3,(2)} &= 3.490, \quad c_{2,4,(2)} = 3.251, \quad c_{3,4,(2)} = 3.251 \end{aligned}$$

for Sam.1 and

$$\begin{aligned} c_{1,2,(1)} &= 5.710, \quad c_{1,3,(1)} = 5.095, \quad c_{1,4,(1)} = 4.978, \\ c_{2,3,(1)} &= 3.784, \quad c_{2,4,(1)} = 3.646, \quad c_{3,4,(1)} = 2.898, \\ c_{1,2,(2)} &= 5.647, \quad c_{1,3,(2)} = 5.039, \quad c_{1,4,(2)} = 4.923, \\ c_{2,3,(2)} &= 3.742, \quad c_{2,4,(2)} = 3.606, \quad c_{3,4,(2)} = 2.866 \end{aligned}$$

for Sam.2. For Sam.1 the Type I error of M1 is

$$P(G_{k,l} > c_{k,l,(1)} \text{ for at least one pair } (k, l)) = 0.0386$$

and that of M2 is

$$P(G_{k,l} > c_{k,l,(2)} \text{ for at least one pair } (k, l)) = 0.0416.$$

For Sam.2 the Type I error of M1 is

$$P(G_{k,l} > c_{k,l,(1)} \text{ for at least one pair } (k, l)) = 0.0366$$

and that of M2 is

$$P(G_{k,l} > c_{k,l,(2)} \text{ for at least one pair } (k, l)) = 0.0387.$$

These probabilities are calculated by Monte Carlo simulation with 1,000,000 repetitions. The $P(G_{k,l} > c_{k,l,(2)})$ s are not uniform under $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$. Specifically

$$\begin{aligned} P(G_{1,2} > c_{1,2,(2)}) &= 0.0090, \quad P(G_{1,3} > c_{1,3,(2)}) = 0.0090, \quad P(G_{1,4} > c_{1,4,(2)}) = 0.0090, \\ P(G_{2,3} > c_{2,3,(2)}) &= 0.0091, \quad P(G_{2,4} > c_{2,4,(2)}) = 0.0091, \quad P(G_{3,4} > c_{3,4,(2)}) = 0.0091 \end{aligned}$$

for Sam.1.

$$P(G_{1,2} > c_{1,2,(2)}) = 0.0087, \quad P(G_{1,3} > c_{1,3,(2)}) = 0.0087, \quad P(G_{1,4} > c_{1,4,(2)}) = 0.0087,$$

$$P(G_{2,3} > c_{2,3,(2)}) = 0.0088, P(G_{2,4} > c_{2,4,(2)}) = 0.0088, P(G_{3,4} > c_{3,4,(2)}) = 0.0090$$

for Sam.2. These results show the six probabilities are relatively uniform.

Next, we consider the power of the test. Letting γ be a positive constant which is less than 1, we set up 17 cases of alternative hypotheses as follows.

- Case 1. $\gamma\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$,
- Case 2. $\sigma_1^2 = \gamma\sigma_2^2 = \sigma_3^2 = \sigma_4^2$,
- Case 3. $\sigma_1^2 = \sigma_2^2 = \gamma\sigma_3^2 = \sigma_4^2$,
- Case 4. $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \gamma\sigma_4^2$,
- Case 5. $\gamma\sigma_1^2 = \gamma\sigma_2^2 = \sigma_3^2 = \sigma_4^2$,
- Case 6. $\gamma\sigma_1^2 = \sigma_2^2 = \gamma\sigma_3^2 = \sigma_4^2$,
- Case 7. $\gamma\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \gamma\sigma_4^2$,
- Case 8. $\sigma_1^2 = \gamma\sigma_2^2 = \gamma\sigma_3^2 = \sigma_4^2$,
- Case 9. $\sigma_1^2 = \gamma\sigma_2^2 = \sigma_3^2 = \gamma\sigma_4^2$,
- Case 10. $\sigma_1^2 = \sigma_2^2 = \gamma\sigma_3^2 = \gamma\sigma_4^2$,
- Case 11. $\gamma\sigma_1^2 = \gamma^2\sigma_2^2 = \sigma_3^2 = \sigma_4^2$,
- Case 12. $\gamma\sigma_1^2 = \sigma_2^2 = \gamma^2\sigma_3^2 = \sigma_4^2$,
- Case 13. $\gamma\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \gamma^2\sigma_4^2$,
- Case 14. $\sigma_1^2 = \gamma\sigma_2^2 = \gamma^2\sigma_3^2 = \sigma_4^2$,
- Case 15. $\sigma_1^2 = \gamma\sigma_2^2 = \sigma_3^2 = \gamma^2\sigma_4^2$,
- Case 16. $\sigma_1^2 = \sigma_2^2 = \gamma\sigma_3^2 = \gamma^2\sigma_4^2$,
- Case 17. $\sigma_1^2 = \gamma\sigma_2^2 = \gamma^2\sigma_3^2 = \gamma^3\sigma_4^2$.

The power of Cases 1 to 4 is the probability that the three hypotheses are rejected. The power of Cases 5 to 10 is the probability that the four hypotheses are rejected. The power of Cases 11 to 16 is the probability that the five hypotheses are rejected. The power of Case 17 is the probability that all hypotheses are rejected. Tables 5 to 8 give the power of Cases 1 to 17 for $\gamma = 0.5, 0.25$ for Sam.1. Tables 9 to 12 give them for Sam.2. The results in the tables are calculated by Monte Carlo simulation with 100,000 repetitions. When the number of hypotheses to be rejected is specified, the power by M is relatively uniform.

Table 5: Power for Cases 1 to 5 for Sam.1

Case	1		2		3		4		5	
γ	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25
M	0.377	0.993	0.368	0.998	0.368	0.998	0.360	0.999	0.263	0.994
M1	0.278	0.989	0.330	0.996	0.333	0.996	0.368	0.999	0.224	0.991
M2	0.290	0.989	0.344	0.996	0.345	0.997	0.387	0.999	0.239	0.991

Table 6: Power for Cases 6 to 10 for Sam.1

Case	6		7		8		9		10	
γ	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25
M	0.259	0.995	0.267	0.995	0.270	0.996	0.276	0.997	0.277	0.997
M1	0.224	0.990	0.228	0.990	0.238	0.994	0.240	0.995	0.243	0.995
M2	0.235	0.992	0.241	0.991	0.249	0.994	0.254	0.996	0.253	0.996

Table 7: Power for Cases 11 to 15 for Sam.1

Case	11		12		13		14		15	
γ	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25
M	0.175	0.994	0.176	0.994	0.175	0.994	0.191	0.997	0.193	0.998
M1	0.081	0.990	0.082	0.989	0.087	0.989	0.165	0.996	0.178	0.996
M2	0.092	0.990	0.091	0.989	0.098	0.990	0.180	0.996	0.193	0.997

Table 8: Power for Cases 16 and 17 for Sam.1

Case	16		17	
γ	0.50	0.25	0.50	0.25
M	0.194	0.998	0.135	0.998
M1	0.175	0.996	0.115	0.996
M2	0.190	0.997	0.129	0.996

Table 9: Power for Cases 1 to 5 for Sam.2

Case	1		2		3		4		5	
γ	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25
M	0.262	0.961	0.235	0.983	0.207	0.994	0.201	0.996	0.132	0.963
M1	0.127	0.914	0.196	0.958	0.237	0.988	0.250	0.992	0.108	0.942
M2	0.132	0.917	0.201	0.958	0.244	0.988	0.257	0.992	0.110	0.945

Table 10: Power for Cases 6 to 10 for Sam.2

Case	6		7		8		9		10	
γ	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25
M	0.138	0.965	0.140	0.965	0.146	0.982	0.150	0.982	0.175	0.992
M1	0.121	0.923	0.122	0.920	0.148	0.950	0.146	0.951	0.146	0.980
M2	0.125	0.926	0.127	0.924	0.154	0.952	0.153	0.954	0.154	0.981

Table 11: Power for Cases 11 to 15 for Sam.2

Case	11		12		13		14		15	
γ	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25
M	0.053	0.960	0.045	0.964	0.044	0.963	0.057	0.983	0.056	0.984
M1	0.004	0.905	0.004	0.913	0.004	0.913	0.058	0.957	0.061	0.957
M2	0.004	0.907	0.004	0.917	0.004	0.918	0.066	0.959	0.066	0.959

Table 12: Power for Cases 16 and 17 for Sam.2

Case	16		17	
γ	0.50	0.25	0.50	0.25
M	0.075	0.995	0.065	0.959
M1	0.154	0.987	0.065	0.957
M2	0.165	0.988	0.071	0.959

Example

To illustrate our procedure, we make use of an example from Tango (1986) (Example 4.4, p. 43). We consider the multiple comparison with a control. There are four groups of children G1, G2, G3 and G4. G1 is a healthy control group. G2, G3 and G4 consist of chickenpox cases, measles cases, and scarlet fever cases, respectively. We obtain 7, 6, 5 and 5 samples, respectively, from the four groups. Table 13 gives sample means and sample variances for the proportion (%) of HLA-DR across the four groups.

Table 13: Sample means and sample variances for rate (%) of HLA-DR

	G1	G2	G3	G4
Sample mean	28.714	39.867	21.560	32.160
Sample variance	51.738	33.191	8.893	16.643

Let μ_1, μ_2, μ_3 and μ_4 be the means of the four groups. For carrying out the multiple comparison with a control intended to compare μ_1 with μ_2, μ_3, μ_4 simultaneously, we should check the equality of variances. Specifically, letting $\sigma_1^2, \sigma_2^2, \sigma_3^2$ and σ_4^2 be the variances of the four groups, we compare σ_1^2 with $\sigma_2^2, \sigma_3^2, \sigma_4^2$ simultaneously. We set up

$$H_{1,k}^{(1)} : \sigma_1^2 = \sigma_k^2 \text{ vs. } H_{1,k}^{(1)A} : \sigma_1^2 > \sigma_k^2 \text{ for } k = 2, 3, 4.$$

Let $\alpha = 0.05$. The common critical value of M is $c = 6.440$. On the other hand, we obtain

$$\begin{aligned} c_{2,(1)} &= 7.107, \quad c_{3,(1)} = 7.412, \quad c_{4,(1)} = 7.412, \\ c_{2,(2)} &= 7.057, \quad c_{3,(2)} = 7.359, \quad c_{4,(2)} = 7.359, \\ c_{2,(3)} &= 6.266, \quad c_{3,(3)} = 6.535, \quad c_{4,(3)} = 6.535. \end{aligned}$$

Since

$$\frac{\nu_1^2}{\nu_2^2} = 1.559, \quad \frac{\nu_1^2}{\nu_3^2} = 5.820, \quad \frac{\nu_1^2}{\nu_4^2} = 3.108,$$

we retain $H_{1,2}^{(1)}, H_{1,3}^{(1)}, H_{1,4}^{(1)}$ by each method. Next, we compare μ_1 with μ_2, μ_3, μ_4 simultaneously under the equality of variances. We set up

$$H_{1,k} : \mu_1 = \mu_k \text{ vs. } H_{1,k}^A : \mu_1 \neq \mu_k \text{ for } k = 2, 3, 4.$$

Then, the statistics for multiple comparison with a control intended to compare μ_1 with μ_2, μ_3, μ_4 simultaneously are

$$T_{1,2} = 3.633, \quad T_{1,3} = 2.214, \quad T_{1,4} = 1.122.$$

Here

$$T_{1,k} = \frac{|\bar{x}_1 - \bar{x}_k|}{\sqrt{\frac{1}{n_1} + \frac{1}{n_k}} s}$$

where

$$s = \sqrt{\frac{\sum_{k=1}^4 (n_k - 1) \nu_k^2}{\sum_{k=1}^4 n_k - 4}}.$$

Since the critical value is $c = 2.551$ for $\alpha = 0.05$, only $H_{1,2}$ is rejected.

On the other hand, if $\alpha = 0.10$, the common critical value of M is $c = 4.650$ and

$$\begin{aligned} c_{2,(1)} &= 5.280, \quad c_{3,(1)} = 5.478, \quad c_{4,(1)} = 5.478, \\ c_{2,(2)} &= 5.199, \quad c_{3,(2)} = 5.392, \quad c_{4,(2)} = 5.392, \\ c_{2,(3)} &= 4.539, \quad c_{3,(3)} = 4.707, \quad c_{4,(3)} = 4.707. \end{aligned}$$

By each method, we reject $H_{1,3}^{(1)}$ and retain $H_{1,2}^{(1)}, H_{1,4}^{(1)}$. Therefore, the same result is obtained using this approach.

5. Conclusions

In this study we discussed critical values for pairwise comparisons of the multiple comparison with a control and the all-pairwise multiple comparison for normal variances when sample sizes are unbalanced. Although Imada (2018A, 2018B) used a common critical value for all hypotheses, the probability that the true hypothesis is rejected is not uniform when sample sizes are unbalanced. Since it is preferable to set up a critical value for each hypothesis so that the probability that it is rejected when it is true may be uniform, we determine two kinds of conservative critical values satisfying the condition and the exact critical values satisfying the condition closely for the multiple comparison with a control. Furthermore, we determine conservative critical values satisfying the condition using Bonferroni's inequality and less conservative critical values satisfying the condition closely using the improved Bonferroni's inequality for the all-pairwise multiple comparison. Finally, we give some numerical results regarding critical values and power of the test.

There remain problems to be explored in future research. Since our proposed critical values for the all-pairwise multiple comparison are conservative, it is desirable to obtain less conservative alternatives. Although Imada (2019) constructed stepwise procedures for the multiple comparison with a control and the all-pairwise multiple comparison for normal variances, the probability that the true hypothesis is rejected is not uniform in these procedures when sample sizes are unbalanced. It is desirable to construct stepwise procedures using critical values so that the probability that the true hypothesis is rejected may be uniform.

Acknowledgement

The author is deeply grateful to the referee and the editors for their valuable comments and suggestions.

References

- Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, **50**, 1096-1121.
- Dunnett, C. W. and Tamhane, A.C. (1991). Step Down Multiple Tests for Comparing Treatments with a Control in Unbalanced One-way Layouts. *Statistics in Medicine*, **10**, 939-947.
- Dunnett, C. W. and Tamhane, A.C. (1992). A step-up multiple test procedure. *Journal of the American Statistical Association*, **87**, 162-170.
- Einot, I. and Gabriel, K. R. (1975). A study of the powers of several methods of multiple comparisons. *Journal of the American Statistical Association*, **70**, 574-583.
- Hsu, J. C. (1996). *Multiple Comparisons*. Boca Raton : Chapman & Hall.
- Imada, T. (2018A). Multiple comparison for checking differences among normal variances. *Communications in Statistics-Theory and Method*, **47**, Issue 18, 4402-4414.
- Imada, T. (2018B). All-Pairwise Multiple Comparison for Normal Variances, *Bulletin of Informatics and Cybernetics* , **50**, 19-26.

- Imada, T. (2019). Stepwise Multiple Comparison Procedures for Normal Variances, *Bulletin of Informatics and Cybernetics* , **51**, 1-20.
- Marcus, R., Peritz, E. and Gabriel, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, **63** (3), 655-660.
- Ramsey, P. H. (1978). Power differences between pairwise multiple comparisons. *Journal of the American Statistical Association*, **73**, 479-485.
- Ryan, T. A. (1960). Significance tests for multiple comparison of proportions, variances, and other statistics. *Psychological Bulletin*, **57**, 318-328.
- Tango, T. (1986). *Statistics for Clinical Trials*. Asakura Shoten (in Japanese).
- Tukey, J. W. (1953). The problem of multiple comparisons. Unpublished manuscript, Princeton University.
- Welsch, R. E. (1977). Stepwise multiple comparison procedures. *Journal of the American Statistical Association*, **72**, 566-575.

Received March 25, 2020

Revised September 4, 2020