STEPWISE MULTIPLE COMPARISON PROCEDURES FOR COMPARING SIZES OF NORMAL MEANS

IMADA, Tsunehisa
Department of Human Information Engineering, Tokai University

https://doi.org/10.5109/4794459

出版情報：Bulletin of informatics and cybernetics. 54 (3), pp.1-10, 2022. 統計科学研究会

STEPWISE MULTIPLE COMPARISON PROCEDURES FOR COMPARING SIZES OF NORMAL MEANS

by

Tsunehisa Imada

Reprinted from the Bulletin of Informatics and Cybernetics Research Association of Statistical Sciences, Vol.54, No. 3

FUKUOKA, JAPAN
2022

Bulletin of Informatics and Cybernetics, Vol. 54, No. 3, 2022

STEPWISE MULTIPLE COMPARISON PROCEDURES FOR COMPARING SIZES OF NORMAL MEANS

By

Tsunehisa Imada*

Abstract

The aim of this study is to propose stepwise multiple comparison procedures for comparing sizes of normal means. Specifically, we construct Tukey-Welsh's step down procedure and a closed testing procedure called Ryan-Einot-Gabriel-Welsch's procedure based on Imada (2020)'s single step procedure and compare them in terms of numerical results regarding power of the test. Furthermore, we illustrate our procedures by an example.

Key Words and Phrases: Closed testing procedure, Power of the test, Tukey-Welsh's step down procedure.

1. Introduction

Assume there are K normal populations $N(\mu_k, \sigma^2)$ $(k = 1, 2, \ldots, K)$. We occasionally want to compare sizes of μ_1,μ_2,...,μ_K. Dunnett (1955), Dunnett and Tamhane (1991) and Dunnett and Tamhane (1992) constructed the multiple comparison procedures for detecting the pair μ_1, μ_k satisfying $\mu_1 > \mu_k$ under the assumption that $\mu_1 \geq \mu_k$ for $k = 2, 3, \ldots, K$. They are called the multiple comparison procedures with a control. Lee and Spurrier (1995) and Imada (2015) constructed the multiple comparison procedures for detecting the pair μ_k, μ_{k+1} satisfying $\mu_k > \mu_{k+1}$ under the assumption that $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_K$. They are called the successive comparison procedures. Imada (2020) constructed the procedures for detecting μ_k satisfying $\mu_k < \max_{1 \leq l \leq K} \mu_l$.

The aim of this study is to construct multiple comparison procedures for comparing sizes of μ_1,μ_2,...,μ_K based on Imada (2020)'s single step procedure. When there are several treatments evaluated by normal response for a certain disease, we can clarify specific relations regarding the superiority and the inferiority among them by our procedures. Here, we construct two types of stepwise procedures. One is Tukey-Welsh's step down procedure (cf. Tukey (1953) and Welsch (1972)) and the other is the closed testing procedure called Ryan-Einot-Gabriel-Welsch's procedure (cf. Marcus *et al.* (1976), Ryan (1960), Einot and Gabriel (1975) and Welsch (1977)). Although Tukey-Welsh's step down procedure is not more powerful compared to Ryan-Einot-Gabriel-Welsch's procedure, Tukey-Welsh's step down procedure is simpler for practical use. We give some numerical results regarding power of the test intended to compare two procedures. Furthermore, we illustrate our procedures by an example.

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

2. Imada (2020)'s multiple comparison procedures for finding non-maximum normal means

First, we discuss Imada (2020)'s multiple comparison procedures for finding μ_k satisfying $\mu_k < \max_{1\le l\le K}\mu_l$. Imada (2020) set up the null hypothesis H_k and its alternative hypothesis H_k^A as

$$H_k : \mu_k = \max_{1\le l\le K}\mu_l \quad \text{vs.} \quad H_k^A : \mu_k < \max_{1\le l\le K}\mu_l$$

for $k = 1, 2, \dots, K$ and test them simultaneously using a sample X_{k1}, X_{k2}, ..., X_{kn_k} from $N(\mu_k, \sigma^2)$ for $k = 1, 2, \dots, K$. Letting

$$\bar{X}_k = \frac{1}{n_k}\sum_{i=1}^{n} X_{ki} \ \ (k = 1, 2, \dots, K), \quad N = \sum_{k=1}^{K} n_k, \quad s = \sqrt{\frac{1}{\nu}\sum_{k=1}^{K}\sum_{i=1}^{n_k}(X_{ki} - \bar{X}_k)^2}$$

where $\nu = N - K$, we use the statistic

$$S_k = \frac{\sqrt{N}(\max_{1\le l\le K}\bar{X}_l - \bar{X}_k)}{s}$$

for testing H_k. In the single step procedure for $H_1, H_2, \dots, H_K$, we set up a critical value $c(> 0)$. If $S_k > c$, we reject H_k. Otherwise, we retain H_k. We determine c for a specified significance level α. Specifically, we determine c so that

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

when all H_ks are true. In the hereafter discussion, $P(\cdot)$ means the probability measure under the assumption that all H_ks are true. Imada (2020) derived

$$P(\max_{k=1,2,\dots,K} S_k > c)$$
$$= 1 - \sum_{k=1}^{K}\int_0^{\infty}\int_{-\infty}^{\infty}\prod_{i\ne k}\left\{\Phi\left(\sqrt{\frac{n_i}{n_k}}z\right) - \Phi\left(\sqrt{\frac{n_i}{n_k}}z - c\sqrt{\frac{n_i}{N}}v\right)\right\}\phi(z)g(v)dzdv.$$

Here, $\Phi(\cdot)$ is the cumulative distribution function of $N(0, 1)$, $\phi(\cdot)$ is the probability density function of $N(0, 1)$ and $g(v)$ is the probability density function of $v = s/\sigma$ given by

$$g(v) = \frac{\nu^{\nu/2}}{2^{(\nu-2)/2}\Gamma[\nu/2]}v^{\nu-1}\exp\left[-\frac{\nu v^2}{2}\right].$$

If $n_1 = n_2 = \cdots = n_K = n$,

$$P(\max_{k=1,2,\dots,K} S_k > c)$$
$$= 1 - K\int_0^{\infty}\int_{-\infty}^{\infty}\left\{\Phi(z) - \Phi\left(z - \frac{c}{\sqrt{K}}v\right)\right\}^{K-1}\phi(z)g(v)dzdv.$$

In this case, $c/\sqrt{2K}$ is the critical value for pairwise comparison of the all-pairwise multiple comparison procedure proposed by Tukey (1953).

Furthermore, Imada (2020) constructed a sequentially rejective step down procedure and a step up procedure. For three procedures, the power of the test was formulated under a specified alternative hypothesis.

3. Stepwise multiple comparison procedures for comparing sizes of normal means

In this section, we construct two types of stepwise multiple comparison procedures for comparing sizes of normal means based on Imada (2020)'s single step procedure. One is Tukey-Welsh's step down procedure and the other is Ryan-Einot-Gabriel-Welsch's procedure.

3.1. Tukey-Welsh's step down procedure

First, we construct Tukey-Welsh's step down procedure. Let I_s be an arbitrary subset of $I = \{1, 2, \ldots, K\}$. $\sharp(I_s)$ denotes the cardinal number of I_s. We consider I_s satisfying $\sharp(I_s) \geq 2$. When $I_s = \{s_1, s_2, \ldots, s_k\}$, define the hypothesis H_{I_s} and its alternative hypothesis $H_{I_s}^A$ as

$$H_{I_s} : \mu_{s_1} = \mu_{s_2} = \cdots = \mu_{s_k} \quad \text{vs.} \quad H_{I_s}^A : \mu_{s_l} < \max\{\mu_{s_1}, \mu_{s_2}, \ldots, \mu_{s_k}\} \text{ for some } l.$$

We use the statistic

$$S_{I_s} = \max_{1 \leq l \leq k} \frac{\sqrt{N}(\max_{1 \leq h \leq k} \bar{X}_{s_h} - \bar{X}_{s_l})}{s} \tag{1}$$

for testing H_{I_s}. If $\sharp(I_s) > K - 2$, we determine the critical value c_{I_s} for testing H_{I_s} so that

$$P(S_{I_s} > c_{I_s}) = \alpha.$$

Otherwise, we determine the critical value c_{I_s} so that

$$P(S_{I_s} > c_{I_s}) = 1 - (1 - \alpha)^{\frac{\sharp(I_s)}{K}}.$$

Here

$$P(S_{I_s} > c_{I_s})$$
$$= 1 - \sum_{m=1}^{k} \int_0^\infty \int_{-\infty}^\infty \prod_{i \neq m} \left\{ \Phi\left(\sqrt{\frac{n_{s_i}}{n_{s_m}}} z\right) - \Phi\left(\sqrt{\frac{n_{s_i}}{n_{s_m}}} z - c_{I_s}\sqrt{\frac{n_{s_i}}{N}} v\right)\right\} \phi(z) g(v) dz dv.$$

If $n_1 = n_2 = \cdots = n_K$ and $\sharp(I_{s_1}) = \sharp(I_{s_2})$, $c_{I_{s_1}} = c_{I_{s_2}}$. Therefore, if $n_1 = n_2 = \cdots = n_K$, c_{I_s} is denoted by $c_{\sharp(I_s)}$. We carry out Tukey-Welsh's step down procedure for all H_{I_s}s as follows.

Step 1.
We test H_I.
Case 1. If $S_I \leq c_I$, we retain all H_{I_s}s and stop the test.
Case 2. If $S_I > c_I$, we reject H_I and go to the next step.

Step 2.
We test all H_{I_s}s satisfying $\sharp(I_s) = K - 1$.
Case 1. If $S_{I_s} \leq c_{I_s}$, we retain H_{I_s} and all hypotheses induced by H_{I_s}.
Case 2. If $S_{I_s} > c_{I_s}$, we reject H_{I_s}.

Step 3.
If all H_{I_s}s satisfying $\sharp(I_s) = K - 2$ are retained at Step 2, we stop the test. Otherwise, we test all H_{I_s}s satisfying $\sharp(I_s) = K - 2$ which are not retained at Step 2.

Case 1. If $S_{I_s} \leq c_{I_s}$, we retain H_{I_s} and all hypotheses induced by H_{I_s}.
Case 2. If $S_{I_s} > c_{I_s}$, we reject H_{I_s}.

We repeat similar judgments till up to Step $K-1$. Let $I_s = \{s_1, s_2\}$. If $S_{I_s} > c_{I_s}$ and $\bar{X}_{s_1} < \bar{X}_{s_2}$ at the final Step $K-1$, we reject H_{I_s} and judge $\mu_{s_1} < \mu_{s_2}$.

It is difficult to formulate the power of Tukey-Welsh's step down procedure under a specified alternative hypothesis. We calculate the power using Monte Carlo simulation.

3.2. Ryan-Einot-Gabriel-Welsch's procedure

Next, we construct Ryan-Einot-Gabriel-Welsch's procedure. Letting H be the family of hypotheses consisting of all H_{I_s}s and all sorts of intersections of plural H_{I_s}s, H is closed. Each hypothesis in H is equal to single H_{I_s} or $H_{I_{s_1}} \cap H_{I_{s_2}} \cap \cdots \cap H_{I_{s_m}}$ where $I_{s_1}, I_{s_2}, \ldots, I_{s_m}$ are disjoint. We determine the critical value c_{I_s} for testing H_{I_s} so that

$$P(S_{I_s} > c_{I_s}) = \alpha.$$

If $S_{I_s} > c_{I_s}$, we reject H_{I_s}. Otherwise, we retain H_{I_s}. If $n_1 = n_2 = \cdots = n_K$, c_{I_s} is denoted by $c_{\sharp(I_s)}$.

Next, we discuss how to test $H_{I_{s_1}} \cap H_{I_{s_2}} \cap \cdots \cap H_{I_{s_m}}$ where $I_{s_1}, I_{s_2}, \ldots, I_{s_m}$ are disjoint. Let $M = \sharp(I_{s_1}) + \sharp(I_{s_2}) + \cdots + \sharp(I_{s_m})$. For $l = 1, 2, \ldots, m$ we determine $c_{I_{s_l},M}$ so that

$$P(S_{I_{s_l}} > c_{I_{s_l},M}) = 1 - (1-\alpha)^{\frac{\sharp(I_{s_l})}{M}}. \tag{1}$$

If $n_1 = n_2 = \cdots = n_K$, $c_{I_{s_l},M}$ is denoted by $c_{\sharp(I_{s_l}),M}$. Intended to test $H_{I_{s_1}} \cap H_{I_{s_2}} \cap \cdots \cap H_{I_{s_m}}$ we set up the critical value $c_{I_{s_l},M}$ for testing $H_{I_{s_l}}$ for $l = 1, 2, \ldots, m$. If $S_{I_{s_l}} > c_{I_{s_l},M}$ for at least one l, $H_{I_{s_1}} \cap H_{I_{s_2}} \cap \cdots \cap H_{I_{s_m}}$ is rejected. Otherwise, it is retained. Then, the probability that $H_{I_{s_1}} \cap H_{I_{s_2}} \cap \cdots \cap H_{I_{s_m}}$ is rejected when $H_{I_{s_1}} \cap H_{I_{s_2}} \cap \cdots \cap H_{I_{s_m}}$ is true is

$$P(S_{I_{s_l}} > c_{I_{s_l},M} \text{ for some } l) = 1 - P(S_{I_{s_l}} \leq c_{I_{s_l},M} \text{ for } l = 1, 2, \ldots, m). \tag{2}$$

Since

$$\begin{aligned}
&P(S_{I_{s_l}} \leq c_{I_{s_l},M} \text{ for } l = 1, 2, \ldots, m) \\
&= \int_0^\infty P(S_{I_{s_l}} \leq c_{I_{s_l},M} \text{ for } l = 1, 2, \ldots, m | v) g(v) dv \\
&= \int_0^\infty \prod_{l=1}^m P(S_{I_{s_l}} \leq c_{I_{s_l},M} | v) g(v) dv \\
&\geq \prod_{l=1}^m \int_0^\infty P(S_{I_{s_l}} \leq c_{I_{s_l},M} | v) g(v) dv \\
&= \prod_{l=1}^m \int_0^\infty P(S_{I_{s_l}} \leq c_{I_{s_l},M})
\end{aligned}$$

by Corollary A.1.1 in Hsu (1996), we obtain

$$P(S_{I_{s_l}} > c_{I_{s_l},M} \text{ for some } l) \leq \alpha$$

by (1) and (2). Therefore, the probability that $H_{I_{s_1}} \cap H_{I_{s_2}} \cap \cdots \cap H_{I_{s_m}}$ is rejected when it is true is not greater than α. We specified the way to test each hypothesis in H satisfying the specified significance level α. We test the hypotheses in H hierarchically. Specifically, if a hypothesis and all hypotheses inducing it are rejected, we reject the hypothesis. Otherwise, we retain it.

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

4. Numerical examples

We give some numerical examples regarding power. Let $\alpha = 0.05$ and $K = 5$. We set up two types of $(n_1, n_2, n_3, n_4, n_5)$ as

$$\text{Sam.1} : (15, 15, 15, 15, 15),\ \text{Sam.2} : (10, 20, 15, 20, 10).$$

Letting $\delta > 0$, we consider the power of the test under four types of alternative hypotheses as follows.

Case 1. $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$, $\mu_5 = \delta$.
Case 2. $\mu_1 = \mu_2 = \mu_3 = 0$, $\mu_4 = \delta$, $\mu_5 = 2\delta$.
Case 3. $\mu_1 = \mu_2 = 0$, $\mu_3 = \delta$, $\mu_4 = 2\delta$, $\mu_5 = 3\delta$.
Case 4. $\mu_1 = 0$, $\mu_2 = \delta$, $\mu_3 = 2\delta$, $\mu_4 = 3\delta$, $\mu_5 = 4\delta$.

We focus on the all-pairs power defined by Ramsey (1978). The power of Case 1 is the probability that

$$\mu_1 < \mu_5,\ \mu_2 < \mu_5,\ \mu_3 < \mu_5,\ \mu_4 < \mu_5$$

are detected. The power of Case 2 is the probability that

$$\mu_1 < \mu_4,\ \mu_1 < \mu_5,\ \mu_2 < \mu_4,\ \mu_2 < \mu_5,\ \mu_3 < \mu_4,\ \mu_3 < \mu_5,\ \mu_4 < \mu_5$$

are detected. The power of Case 3 is the probability that

$$\mu_1 < \mu_3,\ \mu_1 < \mu_4,\ \mu_1 < \mu_5,\ \mu_2 < \mu_3,\ \mu_2 < \mu_4,\ \mu_2 < \mu_5,\ \mu_3 < \mu_4,\ \mu_3 < \mu_5,\ \mu_4 < \mu_5$$

are detected. The power of Case 4 is the probability that

$$\mu_1 < \mu_2,\ \mu_1 < \mu_3,\ \mu_1 < \mu_4,\ \mu_1 < \mu_5,\ \mu_2 < \mu_3,$$

$$\mu_2 < \mu_4,\ \mu_2 < \mu_5,\ \mu_3 < \mu_4,\ \mu_3 < \mu_5,\ \mu_4 < \mu_5$$

are detected. Our procedures enable us not only to compare sizes of normal means, but to detect the pair which has different means. Therefore, we compare our procedures and conventional Tukey-Welsh's step down procedure and Ryan-Einot-Gabriel-Welsch's procedure for detecting the pair which has different means in terms of the power. TW1 and CT1 mean our Tukey-Welsh's step down procedure and Ryan-Einot-Gabriel-Welsch's procedure, respectively. TW2 and CT2 mean conventional Tukey-Welsh's step down procedure and Ryan-Einot-Gabriel-Welsch's procedure, respectively. Tables 1,2 give the power of four procedures for Sam.1 and Sam.2, respectively. Here δ = 1.00, 1.25, 1.50. They are obtained by Monte Carlo simulation with 100,000 times of repetitions. The difference of the power between TW1 and CT1 is small in Case 1 for Sam. 1 and Sam. 2. CT1 is more powerful compared to TW1 in Cases 2,3,4 for Sam. 1 and Sam. 2. The difference of the power is larger as the number of hypotheses which should be rejected is larger. These aspects are similar for TW2 and CT2. The difference of the power between TW1 and TW2 is small except for Case 1 of Sam. 2. TW1 is more powerful compared to TW2 in Case 1 of Sam. 2. These aspects are similar for CT1 and CT2.

Table 1: Power of the test for Sam.1

	Case 1				Case 2			
	TW1	CT1	TW2	CT2	TW1	CT1	TW2	CT2
$\delta = 1.00$	0.328	0.333	0.329	0.330	0.175	0.203	0.175	0.203
$\delta = 1.25$	0.633	0.635	0.633	0.633	0.546	0.585	0.551	0.583
$\delta = 1.50$	0.866	0.869	0.869	0.869	0.848	0.870	0.851	0.869

	Case 3				Case 4			
	TW1	CT1	TW2	CT2	TW1	CT1	TW2	CT2
$\delta = 1.00$	0.122	0.202	0.121	0.201	0.099	0.265	0.100	0.262
$\delta = 1.25$	0.506	0.621	0.507	0.622	0.487	0.700	0.485	0.700
$\delta = 1.50$	0.838	0.892	0.839	0.895	0.833	0.927	0.832	0.926

Table 2: Power of the test for Sam.2

	Case 1				Case 2			
	TW1	CT1	TW2	CT2	TW1	CT1	TW2	CT2
$\delta = 1.00$	0.258	0.259	0.251	0.251	0.183	0.211	0.189	0.212
$\delta = 1.25$	0.504	0.506	0.493	0.496	0.540	0.575	0.544	0.574
$\delta = 1.50$	0.745	0.748	0.738	0.739	0.836	0.854	0.836	0.853

	Case 3				Case 4			
	TW1	CT1	TW2	CT2	TW1	CT1	TW2	CT2
$\delta = 1.00$	0.122	0.204	0.122	0.205	0.118	0.288	0.117	0.287
$\delta = 1.25$	0.478	0.595	0.481	0.598	0.495	0.695	0.498	0.697
$\delta = 1.50$	0.801	0.867	0.803	0.871	0.824	0.919	0.826	0.920

Illustration by an example

We illustrate our procedures. We refer to the example discussed by Hsu (1996) (Table 2.3 and Table 2.4 (page 34)). There are six groups of female smokers. They are given in Table 3. Table 4 gives FVC (forced vital capacity) data for six groups.

Table 3: Six groups of smokers

Group label	Definition
NS	non-smokers
PS	passive smokers
NI	non-inhaling (cigar and pipe) smokers
LS	light smokers (1-10 cigarettes per day for at least the last 20 years)
MS	moderate smokers (11-39 cigarettes per day for at least the last 20 years)
HS	heavy smokers ($\geq$ 40 cigarettes per day for at least the last 20 years)

Table 4: FVC data for smoking and non-smoking female subjects

Group label	Group number	Sample size	Mean FVC	Std. dev. FVC
NS	1	200	3.35	0.63
PS	2	200	3.23	0.46
NI	3	50	3.19	0.52
LS	4	200	3.15	0.39
MS	5	200	2.80	0.38
HS	6	200	2.55	0.38

$\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6$ denote population means of FVC for six groups. We compare sizes of $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6$. Let $\alpha = 0.05$. Table 5 describes the process of rejecting $H_{\{1,4\}}$ by TW. Specifically, since $H_{\{1,4\}}$ and all hypotheses inducing $H_{\{1,4\}}$ are rejected, $\mu_1 > \mu_4$ is detected. $H_{\{1,4\}}$ is also rejected by CT.

Table 5: Process of rejecting $H_{\{1,4\}}$ by TW

Hypothesis	Critical value	Statistic	Decision
$H_{\{1,2,3,4,5,6\}}$	10.949	56.206	Reject
$H_{\{1,2,3,4,5\}}$	10.859	38.641	Reject
$H_{\{1,2,3,4,6\}}$	10.859	56.206	Reject
$H_{\{1,2,4,5,6\}}$	8.861	56.206	Reject
$H_{\{1,3,4,5,6\}}$	10.859	56.206	Reject
$H_{\{1,2,3,4\}}$	11.671	14.051	Reject
$H_{\{1,2,4,5\}}$	9.101	38.641	Reject
$H_{\{1,2,4,6\}}$	9.101	56.206	Reject
$H_{\{1,3,4,5\}}$	11.671	38.641	Reject
$H_{\{1,3,4,6\}}$	11.671	56.206	Reject
$H_{\{1,4,5,6\}}$	9.101	56.206	Reject
$H_{\{1,2,4\}}$	9.129	14.051	Reject
$H_{\{1,3,4\}}$	12.152	14.051	Reject
$H_{\{1,4,5\}}$	9.129	38.641	Reject
$H_{\{1,4,6\}}$	9.129	56.206	Reject
$H_{\{1,4\}}$	9.140	14.051	Reject

Table 6 describes the process of retaining $H_{\{1,3\}}$ by TW. It shows that $H_{\{1,2,3\}}$ is retained. It means that $H_{\{1,3\}}$ is retained. Specifically, $\mu_1 > \mu_3$ is not detected. Table 7 describes the process of retaining $H_{\{1,3\}}$ by CT. It shows that $H_{\{1,3\}} \cap H_{\{2,4\}}$ is retained. It means that $H_{\{1,3\}}$ is retained.

Through the test

$$\mu_1 > \mu_4,\ \mu_1 > \mu_5,\ \mu_1 > \mu_6,\ \mu_2 > \mu_5,\ \mu_2 > \mu_6,$$

$$\mu_3 > \mu_5,\ \mu_3 > \mu_6,\ \mu_4 > \mu_5,\ \mu_4 > \mu_6,\ \mu_5 > \mu_6$$

are detected by TW. Same result is obtained by CT. In this illustration, the results for two procedures coincide.

Table 6: Process of retaining $H_{\{1,3\}}$ by TW

Hypothesis	Critical value	Statistic	Decision
$H_{\{1,2,3,4,5,6\}}$	10.949	56.206	Reject
$H_{\{1,2,3,4,5\}}$	10.859	38.641	Reject
$H_{\{1,2,3,4,6\}}$	10.859	56.206	Reject
$H_{\{1,2,3,5,6\}}$	10.859	56.206	Reject
$H_{\{1,3,4,5,6\}}$	10.859	56.206	Reject
$H_{\{1,2,3,4\}}$	11.671	14.051	Reject
$H_{\{1,2,3,5\}}$	11.671	38.641	Reject
$H_{\{1,2,3,6\}}$	11.671	56.206	Reject
$H_{\{1,3,4,5\}}$	11.671	38.641	Reject
$H_{\{1,3,4,6\}}$	11.671	56.206	Reject
$H_{\{1,3,5,6\}}$	11.671	56.206	Reject
$H_{\{1,2,3\}}$	12.152	11.241	Retain

Table 7: Process of retaining $H_{\{1,3\}}$ by CT

Hypothesis	Critical value	Statistic	Decision
$H_{\{1,2,3,4,5,6\}}$	10.949	56.206	Reject
$H_{\{1,2,3,4,5\}}$	10.859	38.641	Reject
$H_{\{1,2,3,4,6\}}$	10.859	56.206	Reject
$H_{\{1,2,3,5,6\}}$	10.859	56.206	Reject
$H_{\{1,3,4,5,6\}}$	10.859	56.206	Reject
$H_{\{1,2,3,4\}} \cap H_{\{5,6\}}$	11.671, 9.140	14.051, 17.564	Reject
$H_{\{1,2,3,5\}} \cap H_{\{4,6\}}$	11.671, 9.140	38.641, 42.154	Reject
$H_{\{1,2,3,6\}} \cap H_{\{4,5\}}$	11.671, 9.140	56.206, 24.590	Reject
$H_{\{1,3,4,5\}} \cap H_{\{2,6\}}$	11.671, 9.140	38.641, 47.775	Reject
$H_{\{1,3,4,6\}} \cap H_{\{2,5\}}$	11.671, 9.140	56.206, 30.211	Reject
$H_{\{1,3,5,6\}} \cap H_{\{2,4\}}$	11.671, 9.140	56.206, 5.621	Reject
$H_{\{2,4,5,6\}} \cap H_{\{1,3\}}$	9.098, 12.959	47.775, 11.241	Reject
$H_{\{1,2,3\}} \cap H_{\{4,5,6\}}$	12.152, 9.129	11.241, 42.154	Reject
$H_{\{1,3,4\}} \cap H_{\{2,5,6\}}$	12.152, 9.129	14.051, 47.775	Reject
$H_{\{1,3,5\}} \cap H_{\{2,4,6\}}$	12.152, 9.129	38.641, 47.775	Reject
$H_{\{1,3,6\}} \cap H_{\{2,4,5\}}$	12.152, 9.129	56.206, 30.211	Reject
$H_{\{1,2,3\}} \cap H_{\{4,5\}}$	11.851, 8.925	11.241, 24.590	Reject
$H_{\{1,2,3\}} \cap H_{\{4,6\}}$	11.851, 8.925	11.241, 42.154	Reject
$H_{\{1,2,3\}} \cap H_{\{5,6\}}$	11.851, 8.925	11.241, 17.564	Reject
$H_{\{1,3,4\}} \cap H_{\{2,5\}}$	11.851, 8.925	14.051, 30.211	Reject
$H_{\{1,3,4\}} \cap H_{\{2,6\}}$	11.851, 8.925	14.051, 47.775	Reject
$H_{\{1,3,4\}} \cap H_{\{5,6\}}$	11.851, 8.925	14.051, 17.564	Reject
$H_{\{1,3,5\}} \cap H_{\{2,4\}}$	11.851, 8.925	38.641, 5.621	Reject
$H_{\{1,3,5\}} \cap H_{\{2,6\}}$	11.851, 8.925	38.641, 47.775	Reject
$H_{\{1,3,5\}} \cap H_{\{4,6\}}$	11.851, 8.925	38.641, 42.154	Reject
$H_{\{1,3,6\}} \cap H_{\{2,4\}}$	11.851, 8.925	56.206, 5.621	Reject
$H_{\{1,3,6\}} \cap H_{\{2,5\}}$	11.851, 8.925	56.206, 30.211	Reject
$H_{\{1,3,6\}} \cap H_{\{4,5\}}$	11.851, 8.925	56.206, 24.590	Reject
$H_{\{2,4,5\}} \cap H_{\{1,3\}}$	8.936, 12.635	30.211, 11.241	Reject
$H_{\{2,4,6\}} \cap H_{\{1,3\}}$	8.936, 12.635	47.775, 11.241	Reject
$H_{\{2,5,6\}} \cap H_{\{1,3\}}$	8.936, 12.635	47.775, 11.241	Reject
$H_{\{4,5,6\}} \cap H_{\{1,3\}}$	8.936, 12.635	42.154, 11.241	Reject
$H_{\{1,3\}} \cap H_{\{2,4\}}$	12.204, 8.706	11.241, 5.621	Retain

On the other hand, assume $\mu_1 \geq \mu_k$ for $k = 2, 3, \ldots, K$ in advance and set up the null hypothesis $H_{1,k}$ and its alternative hypothesis $H^A_{1,k}$ as

$$H_{1,k} : \mu_1 = \mu_k \quad \text{vs.} \quad H^A_{1,k} : \mu_1 > \mu_k$$

for $k = 2, 3, \ldots, K$. We test $H_{1,2}$,$H_{1,3}$,. . .,$H_{1,K}$ using the sequentially step down procedure proposed by Dunnett and Tamhane (1991). Then

$$\mu_1 > \mu_4, \; \mu_1 > \mu_5, \; \mu_1 > \mu_6$$

are detected. Same result is obtained by the step up procedure proposed by Dunnett and Tamhane (1992).

Next, assume $\mu_k \geq \mu_{k+1}$ for $k = 1, 2, \ldots, K-1$ in advance and set up the null hypothesis $H_{k,k+1}$ and its alternative hypothesis $H^A_{k,k+1}$ as

$$H_{k,k+1} : \mu_k = \mu_{k+1} \quad \text{vs.} \quad H^A_{k,k+1} : \mu_k > \mu_{k+1}$$

for $k = 1, 2, \ldots, K-1$. We test $H_{1,2}$,$H_{2,3}$,. . .,$H_{K-1,K}$ using Imada (2015)'s procedure based on Ryan-Einot-Gabriel-Welsch's procedure. Then

$$\mu_4 > \mu_5, \; \mu_5 > \mu_6$$

are detected. In this example, our procedures enable us to induce the results of Dunnett and Tamhane (1991), Dunnett and Tamhane (1992) and Imada (2015).

5. Conclusions

In this study, we discussed the stepwise multiple comparison procedures for comparing sizes of normal means. Specifically, we constructed Tukey-Welsh's step down procedure and the closed testing procedure called Ryan-Einot-Gabriel-Welsch's procedure based on Imada (2020)'s single step procedure. We gave some numerical results regarding power of the test intended to compare two procedures and illustrated them by an example. Although Tukey-Welsh's step down procedure is not more powerful compared to the closed testing procedure, we confirmed that the difference of the power between two procedures is larger as the number of hypotheses to be rejected is larger.

However, there remain problems to be solved in the future. Although the closed testing procedure is more powerful, it is a nuisance to set up many critical values in advance of the test when sample sizes are unbalanced. We should construct a simpler procedure for practical use. Furthermore, we want to construct more powerful procedure. Ordered statistics and their distribution for sample means are available for constructing more powerful stepwise multiple comparison procedures for comparing sizes of normal means.

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. (2015). Successive comparisons between ordered normal means based on closed testing procedure. *Bulletin of Informatics and Cybernetics* , **47**, 25-36 .

Imada, T. (2020). Multiple comparison procedures for finding non-maximum normal means. *Communications in Statistics-Theory and Method*, **49**, Issue 16, 4073-4090.

Lee, R. E. and Spurrier, J. D. (1995). Successive comparisons between ordered treatments. *Journal of Statistical Planning and Inference*, **43**, 323-330.

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.

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.

Welsch, R. E. (1972). A modification of the Newman-Keuls procedure for multiple comparisons. Working Paper 612-672, Sloan School of Management, M.I.T., Boston, MA.

Received: October 18, 2021
Revised: February 28, 2022
Accept: March 5, 2022