Numerical Comparison of Multiple Comparison Procedures Controlling the Maximum Type I Familywise Error Rate and Those Controlling the False Discovery Rate

by

Tsunehisa IMADA*

Abstract

The aim of this study is to compare multiple comparison procedures controlling the maximum type I familywise error rate with those controlling the false discovery rate through simulation results. Specifically, we apply them to the multiple comparison with a control for normal means and give simulation results regarding the maximum type I familywise error rate, the false discovery rate and the power of the test intended to compare the procedures.

key words: Monte Carlo simulation; Multiple comparison with a control; Power of the test

1. Introduction

Assume there are plural populations. When we compare the values of parameters of them simultaneously, we set up hypotheses for comparing them and test them using multiple comparison procedures. When we specify the values of parameters, the probability that at least one true null hypothesis is rejected for multiple comparison procedures is called the type I familywise error rate (FWER). FWER is the abbreviated notation of familywise error rate. Critical values of multiple comparison procedures are determined so that FWER may not be larger than a specified significance level under the assumption that all null hypotheses are true. However, the maximum type I FWER for all sorts of arrangements of parameters is controlled weakly or strongly at a specified probability. Controlling the maximum type I FWER weakly at a specified probability means that the type I FWER is greater than the specified probability for certain arrangements of parameters. On the other hand, controlling the maximum Type I FWER strongly at a specified probability means that the type I FWER is not greater than the specified probability for all sort of arrangements of parameters. Controlling the maximum Type I FWER strongly is more preferable compared to controlling it weakly for multiple comparisons. Single step multiple comparison procedures like Tukey (1953) and Dunnett (1955) and stepwise multiple comparison procedures like Dunnett and Tamhane (1991, 1992) and Marcus *et al.* (1976) control the maximum Type I FWER strongly.

On the other hand, when we specify the values of parameters, let R be the number of hypotheses which are rejected by the test and let V be the number of true hypotheses which are rejected by the test. The expectation of V/R is called the false discovery rate for the specified values of parameters (FDR). FDR is the abbreviated notation of false discovery rate. Controlling FDR at a specified value means that FDR is not greater than it for all sorts of arrangements of parameters. Benjamini and Hochberg (1995) proposed a stepwise procedure which controls FDR at a specified value under the assumption that the correlation coefficient of each two statistics for testing is zero. Then, Benjamini and Hochberg (2000) proposed an adaptive stepwise procedure and confirmed that their procedure controls FDR at a specified significance level through the simulation under the assumption that the correlation coefficient of each two statistics for testing is zero. Furthermore, Benjamini and Yekutieli (2001) proposed a stepwise procedure which controls FDR at a specified significance level under the assumption that the correlation coefficient of each two statistics for testing is non-negative. Horiuchi and Matsuda (2006) compared the three procedures through the simulation regarding FWER and FDR. In almost all cases Benjamini and Yekutieli (2001)'s procedure is most conservative and Benjamini and Hochberg

* Department of Management, Tokai University, Professor

(1995)'s procedure is more conservative compared to Benjamini and Hochberg (2000)'s procedure.

Procedures controlling the maximum Type I FWER strongly also control FDR. However, procedures controlling FDR do not always control the maximum Type I FWER strongly. Specifically, procedures controlling the maximum Type I FWER strongly is more conservative. Therefore, it is preferable to use procedures controlling the maximum Type I FWER strongly for problems having verification aspect. On the other hand, it is more preferable to use procedures controlling FDR for exploratory problems. However, for two types of procedures we want to investigate actual differences regarding the maximum Type I FWER, FDR and the power of the test.

In this study we compare stepwise multiple comparison procedures controlling the maximum type I FWER with those controlling FDR through simulation results. Specifically, we apply them to the multiple comparison with a control for normal means and give simulation results regarding the maximum type I FWER, the FDR and the power of the test intended to compare the procedures. For multiple comparison with a control Dunnett (1955) proposed the single step procedure. Then, Dunnett and Tamhane (1991) constructed the sequentially rejective step down procedure. Furthermore, Dunnett and Tamhane (1992) constructed a step up procedure. For three procedures the maximum Type I FWER is controlled strongly at a specified significance level. Here we focus on the procedures proposed by Benjamini and Hochberg (1995) and Benjamini and Yekutieli (2001) among procedures controlling FDR. We compare the five procedures through the simulation results.

2. Multiple comparison with a control for normal means

Assume there are K normal populations $N(\mu_k,\sigma^2)$ ($k = 1,2,\ldots,K$). Here σ^2 is unknown. Dunnett (1955) discussed the multiple comparison for comparing μ_1 with $\mu_2,\mu_3,\ldots,\mu_K$ simultaneously, which is called multiple comparison procedure with a control. Here we set up a null hypothesis H_{1k} and its alternative hypothesis H_{1k}^A as

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

for $k = 2,3,\ldots,K$ and test them simultaneously using a sample $X_{k1}, X_{k2}, \ldots, X_{kn_k}$ from $N(\mu_k,\sigma^2)$ for $k = 1,2,\ldots,K$.

3. Single step procedure

First, we discuss the single step procedure proposed by Dunnett (1955). We consider the simultaneous test of $H_{12},H_{13},\ldots,H_{1K}$ based on the single step procedure. We use the statistic

$$S_{1k} = \sqrt{\frac{n_1 n_k}{n_1 + n_k}}(\bar{X}_k - \bar{X}_1)s^{-1}$$

for testing H_{1k}. Here

$$\bar{X}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} X_{ki} \quad (k = 1,2,\ldots,K)$$

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

where $N = \sum_{k=1}^{K} n_k$. If $S_{1k} > c$ for a specified critical value c, H_{1k} is rejected. Otherwise, it is retained. We determine c so that

$$P(\max_{k=2,3,\ldots,K} S_{1k} > c) = \alpha$$

for a specified significance level α when $H_{12}, H_{13}, \ldots, H_{1K}$ are true. The formulation of $P(\max_{k=2,3,\ldots,K} S_{1k} > c)$ is given in Nagata and Yoshida (1997).

4. Sequentially rejective step down procedure

The sequentially rejective step down procedure consists of $K-1$ steps of tests. Assuming all $H_{12}, H_{13},\cdots, H_{1K}$ are true, we determine

$$c_m \ (m = 1,2,\ldots,K-1)$$

as the minimum c satisfying

$$P(\max_{k=l_1,l_2,\ldots,l_m} S_{1k} > c) \le \alpha$$

for $l_1,l_2,\ldots,l_m$ chosen from $2,3,\ldots,K$ arbitrarily. The formulation of $P(\max_{k=l_1,l_2,\cdots,l_m} S_{1k} > c)$ is given similarly as that of $P(\max_{k=2,3,\ldots,K} S_{1k} > c)$. Arranging $S_{12},S_{13},\ldots,S_{1K}$ in order of a size of value, assume

$$S_{(1)} \le S_{(2)} \le \cdots \le S_{(K-1)}.$$

$H_{(1)}, H_{(2)},\cdots, H_{(K-1)}$ denote hypotheses corresponding to $S_{(1)}$, $S_{(2)}$, $\cdots, S_{(K-1)}$. Then, we test $H_{(1)}, H_{(2)},\cdots, H_{(K-1)}$ sequentially as follows.

Step 1.

Case 1. If $S_{(k-1)} \le c_{k-1}$, we retain $H_{(1)}, H_{(2)},\cdots, H_{(K-1)}$ and stop the test.

Case 2. If $S_{(k-1)} > c_{k-1}$, we reject $H_{(K-1)}$ and go to the next step.

Step 2.

Case 1. If $S_{(k-2)} \leq c_{k-2}$, we retain $H_{(1)}, H_{(2)}, \cdots, H_{(K-2)}$ and stop the test.

Case 2. If $S_{(k-2)} > c_{k-2}$, we reject $H_{(K-2)}$ and go to the next step.

We repeat similar judgments till up to Step $K-1$.

5. Step up procedure

First, we introduce notations which were used by Hayter and Tamhane (1991) and Dunnett *et al.* (2001). Let $W_1, W_2, \dots, W_l$ be statistics. Let $b_1, b_2, \dots, b_l$ be constants satisfying $b_1 < b_2 < \cdots < b_l$. Calculating $W_1, W_2, \dots, W_l$ based on observations and rewriting indices of $W_1, W_2, \dots, W_l$ according to the order of the size of value, we assume $W_1^{(1)} \leq W_2^{(2)} \leq \cdots \leq W_l^{(l)}$ is obtained. If $W_1^{(1)} \leq b_1, W_2^{(2)} \leq b_2, \cdots, W_l^{(l)} \leq b_l$, we denote

$$(W_1, W_2, \dots, W_l) \leq (b_1, b_2, \dots, b_l). \qquad (5.1)$$

The event (5.1) is recursively divided into plural disjoint events.

The step up procedure consists of $K-1$ steps of tests. We want to determine critical values of the step up procedure $c_1, c_2, \dots, c_{K-1}$ recursively so that

$$c_1 \leq c_2 \leq \cdots \leq c_{K-1} \qquad (5.2)$$

and

$$P((S_{1i_2}, S_{1i_3}, \dots, S_{1i_m}) \leq (c_1, c_2, \dots, c_{m-1})) \geq 1-\alpha \qquad (5.3)$$

for S_{1i_2}, S_{1i_3}, $\cdots$, S_{1i_m} chosen arbitrarily from $S_{12}, S_{13}, \dots, S_{1K}$ for each $2 \leq m \leq K$ under $H_{12} \cap H_{13} \cap \cdots \cap H_{1K}$. However, if $K \geq 4$, it is difficult to indicate the existence of $c_1, c_2, \dots, c_{K-1}$ satisfying (5.2) and (5.3). However, Dunnett and Tamhane (1992) confirmed the existence of $c_1, c_2, \dots, c_{K-1}$ satisfying (5.2) and (5.3) for $K \leq 8$ through the numerical calculations.

Here we assume that $c_1, c_2, \dots c_{K-1}$ satisfying (5.2) and (5.3) are obtained. Arranging $S_{12}, S_{13}, \dots, S_{1K}$ in order of a size of value, assume

$$S_{(1)} \leq S_{(2)} \leq \cdots \leq S_{(K-1)}.$$

$H_{(1)}, H_{(2)}, \cdots, H_{(K-1)}$ denote hypotheses corresponding to $S_{(1)}, S_{(2)}, \cdots, S_{(K-1)}$. Then, we test $H_{(1)}, H_{(2)}, \cdots, H_{(K-1)}$ as follows.

Step 1.

Case 1: If $S_{(1)} > c_1$, we reject $H_{12}, H_{13}, \dots, H_{1K}$ and stop the test.

Case 2: If $S_{(1)} \leq c_1$, we retain $H_{(1)}$ and go to Step 2.

Step 2.

Case 1: If $S_{(2)} > c_2$, we reject $H_{(2)}, H_{(3)}, \dots, H_{(K-1)}$ and stop the test.

Case 2: If $S_{(2)} \leq c_2$, we retain $H_{(2)}$ and go to Step 3.

We repeat similar judgments till up to Step $K-1$.

6. Stepwise multiple comparison procedure controlling FDR

Here we construct a stepwise multiple comparison procedure controlling FDR at α referring to Benjamini and Hochberg (1995) and Benjamini and Yekutieli (2001).

6.1. Stepwise multiple comparison procedure based on Benjamini and Hochberg (1995)

We should calculate p-value for each H_{1k} using the sample. Let S_{1k}^* be the value of S_{1k} calculated using the sample. p-value for H_{1k} is given by $p_k = P(S_{1k} > S_{1k}^*)$ under H_{1k}. First, we apply Benjamini and Hochberg (1995)'s procedure to our problem. Assume

$$p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(K-1)} \qquad (6.1)$$

is obtained by arranging p-values $p_2, p_3, \dots, p_K$ regarding size. Let $H_{(1)}$, $H_{(2)}$, $\cdots, H_{(K-1)}$ be hypotheses corresponding to (6.1). We test $H_{(1)}, H_{(2)}, \cdots, H_{(K-1)}$ stepwisely as follows.

Step 1.

Case 1. If $p_{(K-1)} \geq \alpha$, we retain $H_{(K-1)}$ and go to Step 2.

Case 2. If $p_{(K-1)} < \alpha$, we reject $H_{(1)}, H_{(2)}, \cdots, H_{(K-1)}$ and stop the test.

Step 2.

Case 1. If $p_{(K-2)} \geq (K-2)\alpha/(K-1)$, we retain $H_{(K-2)}$ and go to Step 2.

Case 2. If $p_{(K-2)} < (K-2)\alpha/(K-1)$, we reject $H_{(1)}, H_{(2)}, \cdots, H_{(K-2)}$ and stop the test.

...

Step $K-1$.

Case 1. If $p_{(1)} \geq \alpha/(K-1)$, we retain $H_{(1)}$ and stop the test.

Case 2. If $p_{(1)} < \alpha/(K-1)$, we reject $H_{(1)}$ and stop the test.

Since $S_{12}, S_{13}, \ldots, S_{1K}$ are not mutually independent, we can not guarantee that the stepwise procedure based on Benjamini and Hochberg (1995)'s procedure controls FDR at α. However, Horiuchi and Matsuda (2006) confirmed that Benjamini and Hochberg (1995)'s procedure controls FDR at α in almost all cases through the simulation.

6.2. Stepwise multiple comparison procedure based on Benjamini and Yekutieli (2001)

Next, we apply Benjamini and Yekutieli (2001)'s procedure to our problem. Let

$$\alpha^* = \alpha / \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{K-1}\right).$$

We test $H_{(1)}, H_{(2)}, \cdots, H_{(K-1)}$ stepwisely as follows.

Step 1.

Case 1. If $p_{(K-1)} \geq \alpha^*$, we retain $H_{(K-1)}$ and go to Step 2.

Case 2. If $p_{(K-1)} < \alpha^*$, we reject $H_{(1)}, H_{(2)}, \cdots, H_{(K-1)}$ and stop the test.

Step 2.

Case 1. If $p_{(K-2)} \geq (K-2)\alpha^*/(K-1)$, we retain $H_{(K-2)}$ and go to Step 2.

Case 2. If $p_{(K-2)} < (K-2)\alpha^*/(K-1)$, we reject $H_{(1)}, H_{(2)}, \cdots, H_{(K-2)}$ and stop the test.

...

Step $K-1$.

Case 1. If $p_{(1)} \geq \alpha^*/(K-1)$, we retain $H_{(1)}$ and stop the test.

Case 2. If $p_{(1)} < \alpha^*/(K-1)$, we reject $H_{(1)}$ and stop the test.

Benjamini and Yekutieli (2001)'s procedure is more conservative compared to Benjamini and Hochberg (1995). Letting

$$\lambda_{1k} = \frac{n_k}{n_1 + n_k} \quad (k = 2, 3, \ldots, K),$$

the correlation coefficient between S_{1k} and S_{1l} $(k \neq l)$ is $\lambda_{1k}\lambda_{1l} > 0$. Therefore, we can guarantee that the stepwise procedure based on Benjamini and Yekutieli (2001)'s procedure controls FDR at α.

6.3. Stepwise test using critical value

Instead of calculating p-values in the test, it is convenient to set up critical values for the test. We discuss Benjamini and Hochberg (1995)'s procedure. The following discussion is also applied to Benjamini and Yekutieli (2001)'s procedure.

We determine c_l by

$$\int_0^{c_l} f(t)dt = \frac{K-l}{K-1}\alpha \qquad (6.2)$$

for $l = 1, 2, \ldots, K-1$. Here $f(t)$ is the probability density function of t distribution with $N-K$ degrees of freedom. Then $c_1 < c_2 < \cdots < c_{K-1}$. Assume

$$S^*_{(1)} \leq S^*_{(2)} \leq \cdots \leq S^*_{(K-1)}$$

is obtained by arranging p-values $S^*_1, S^*_2, \cdots, S^*_{K-1}$ regarding size. The stepwise test for $H_{(1)}, H_{(2)}, \cdots, H_{(K-1)}$ is carried out as follows.

Step 1.

Case 1. If $S^*_{(1)} \leq c_1$, we retain $H_{(K-1)}$ and go to Step 2.

Case 2. If $S^*_{(1)} > c_1$ we reject $H_{(1)}, H_{(2)}, \cdots, H_{(K-1)}$ and stop the test.

Step 2.

Case 1. If $S^*_{(2)} \leq c_2$, we retain $H_{(K-2)}$ and go to Step 3.

Case 2. If $S^*_{(2)} > c_2$, we reject $H_{(1)}, H_{(2)}, \cdots, H_{(K-2)}$ and stop the test.

...

Step $K-1$.

Case 1. If $S^*_{(K-1)} \leq c_{K-1}$, we retain $H_{(1)}$ and stop the test.

Case 2. If $S^*_{(K-1)} > c_{K-1}$, we reject $H_{(1)}$ and stop the test.

The process of the test is similar to that of the step up procedure constructed by Dunnett and Tamhane (1992).

7. Power of the test

Assume $\mu_1 = \mu_i$ for $i = 2, 3, \ldots, K_0$ and $\mu_1 \neq \mu_i$ for $i = K_0 + 1, K_0 + 2, \ldots, K$. The power of the test is the probability that $H_{1K_0+1}, H_{1K_0+2}, \cdots, H_{1K}$ are rejected in the test. For formulating the power we should specify the value of the difference $\delta_{1i} = \mu_i - \mu_1$ for $i = K_0+1, K_0+2, \ldots, K$. The power of the test is formulated for the single step procedure, the sequentially rejective step down procedure and the step up procedure. The formulation of the power for the procedures proposed by Benjamini and Hochberg (1995) and Benjamini and Yekutieli (2001) can be obtained using that of the step up procedure.

8. Numerical results

In this section we give some numerical examples regarding critical values and power of the test intended to compare the procedures. We use abbreviated notations. BH means the procedure based on Benjamini and Hochberg (1995) and BY means the procedure based on Benjamini and Yekutieli (2001). Furthermore, SS, SD and SU mean the single step procedure, the sequentially rejective step down procedure and the step up procedure, respectively.

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

Sam.1 : (15,15,15,15,15), Sam.2 : (10,20,15,20,10).

Table 1 gives critical values of SD and SU for Sam.1. Table 2 gives them for Sam.2. The critical value of SS is equal to c_5 of SD. Critical values of BH and BY are given in Table 3.

Next, we investigate FWER and FDR for five procedures. Assume following three cases.

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

Here $\delta > 0$. The true hypotheses in Case 1 are H_{12}, H_{13}, H_{14}. The true hypotheses in Case 2 are H_{12}, H_{13}. The true hypothesis in Case 3 is H_{12}.

Since we should specify the value of σ^2 for computing FWER and FDR, let $\sigma^2 = 1$. Table 4 gives the type I FWER in Cases 1 to 4 for $\delta = 0.5$. Table 5 gives them for $\delta = 1.0$. Table 6 gives FDR in Cases 1 to 4 for $\delta = 0.5$. Table 7 gives them for $\delta = 1.0$. They are calculated by Monte Carlo simulation with 1,000,000 repetitions. The type I FWER of BH is occasionally greater than $\alpha = 0.05$. The type I FWER of BY is always smaller than $\alpha = 0.05$ and it is always smaller than those of SD and SU. Although it is not mathematically proved that FDR of BH controls the specified significance level under the assumption that the correlation coefficient of each two statistics for testing is positive, it is always smaller than $\alpha = 0.05$. FDR of BY is always smaller than those of SD and SU. The results show that BY is more conservative compared to SD and SU.

Next, we consider the power of the test. Assume following four cases.

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

The power of Case 1 is the probability that H_{15} is rejected. The power of Case 2 is the probability that H_{14}, H_{15} are rejected. The power of Case 3 is the probability that H_{13}, H_{14}, H_{15} are rejected. The power of Case 4 is the probability that H_{12}, H_{13}, H_{14}, H_{15} are rejected. Since we should specify the value of σ^2 for computing the power of the test, let $\sigma^2 = 1$. Tables 8 and 9 give the power of the test for $\delta = 0.5, 1.0$, respectively. They are calculated by Monte Carlo simulation with 100,000 repetitions. In Case 1 the differences among SS, SD, SU are not remarkable and they are more powerful compared to BH and BY. In Cases 2 and 3 BH is most powerful. In Case 4 the differences between SU and BH are not remarkable and they are more powerful compared to other procedures. SD and SU are always more powerful compared to BY.

Table 1: Critical values of SD and SU for Sam.1

	c_2	c_3	c_4	c_5
SD	1.668	1.947	2.097	2.200
SU	1.668	1.967	2.112	2.204

Table 2: Critical values of SD and SU for Sam.2

	c_2	c_3	c_4	c_5
SD	1.668	1.941	2.075	2.161
SU	1.668	1.960	2.094	2.165

Table 3: Critical values of BH and BY

	c_2	c_3	c_4	c_5
BH	1.668	1.808	1.995	2.291
BY	2.013	2.139	2.308	2.580

Table 4: Type I FWER for $\delta = 0.5$

		Case 1	Case 2	Case 3
Sam.1	SS	0.0399	0.0287	0.0155
	SD	0.0457	0.0403	0.0327
	SU	0.0461	0.0473	0.0365
	BH	0.0536	0.0641	0.0401
	BY	0.0255	0.0310	0.0181
Sam.2	SS	0.0400	0.0299	0.0168
	SD	0.0442	0.0403	0.0348
	SU	0.0442	0.0466	0.0389
	BH	0.0486	0.0612	0.0416
	BY	0.0234	0.0303	0.0192

Table 5: Type I FWER for $\delta = 1.0$

		Case 1	Case 2	Case 3
Sam.1	SS	0.0401	0.0285	0.0154
	SD	0.0497	0.0491	0.0483
	SU	0.0488	0.0500	0.0494
	BH	0.0627	0.0672	0.0491
	BY	0.0312	0.0331	0.0238
Sam.2	SS	0.0399	0.0297	0.0171
	SD	0.0474	0.0472	0.0482
	SU	0.0466	0.0490	0.0491
	BH	0.0572	0.0638	0.0496
	BY	0.0286	0.0313	0.0237

Table 6: FDR for $\delta = 0.5$

		Case 1	Case 2	Case 3
Sam.1	SS	0.0271	0.0140	0.0058
	SD	0.0303	0.0186	0.0079
	SU	0.0306	0.0184	0.0106
	BH	0.0337	0.0216	0.0114
	BY	0.0162	0.0101	0.0055
Sam.2	SS	0.0284	0.0140	0.0056
	SD	0.0306	0.0185	0.0083
	SU	0.0312	0.0185	0.0106
	BH	0.0321	0.0214	0.0114
	BY	0.0156	0.0101	0.0054

Table 7: FDR for $\delta = 1.0$

		Case 1	Case 2	Case 3
Sam.1	SS	0.0215	0.0103	0.0039
	SD	0.0275	0.0183	0.0120
	SU	0.0276	0.0184	0.0123
	BH	0.0354	0.0244	0.0124
	BY	0.0169	0.0116	0.0060
Sam.2	SS	0.0223	0.0109	0.0043
	SD	0.0273	0.0183	0.0120
	SU	0.0281	0.0189	0.0123
	BH	0.0340	0.0240	0.0124
	BY	0.0166	0.0115	0.0060

Table 8: Power of the test for $\delta = 0.5$

		Case 1	Case 2	Case 3	Case 4
Sam.1	SS	0.209	0.094	0.057	0.038
	SD	0.214	0.118	0.095	0.102
	SU	0.212	0.116	0.098	0.122
	BH	0.193	0.143	0.128	0.120
	BY	0.124	0.080	0.067	0.059
Sam.2	SS	0.153	0.081	0.052	0.041
	SD	0.156	0.097	0.082	0.100
	SU	0.156	0.098	0.086	0.113
	BH	0.136	0.117	0.107	0.113
	BY	0.082	0.064	0.058	0.057

Table 9: Power of the test for $\delta = 1.0$

		Case 1	Case 2	Case 3	Case 4
Sam.1	SS	0.707	0.567	0.478	0.420
	SD	0.704	0.606	0.586	0.633
	SU	0.702	0.602	0.582	0.657
	BH	0.674	0.651	0.653	0.659
	BY	0.566	0.522	0.508	0.502
Sam.2	SS	0.534	0.445	0.374	0.346
	SD	0.534	0.478	0.463	0.532
	SU	0.528	0.476	0.464	0.557
	BH	0.484	0.514	0.525	0.555
	BY	0.373	0.387	0.384	0.406

8. Conclusions

In this study we compared stepwise multiple comparison procedures controlling the maximum type I FWER with those controlling FDR applying to the multiple comparison with a control for normal means through simulation results regarding the maximum type I FWER, the FDR and the power of the test. Although Benjamini and Yekutieli (2001)'s procedure controls the maximum type I FWER and FDR, its power is lowest in almost all cases. Although Benjamini and Hochberg (1995)'s procedure controls FDR and occasionally does not control the maximum type I FWER, it is not uniformly more powerful compared to the procedures controlling the maximum type I FWER. We can not insist that Benjamini and Hochberg (1995)'s procedure is always more preferable for analyzing exploratory problems compared to procedures controlling the maximum type I FWER. We should construct a procedure which controls FDR and is uniformly more powerful compared to the procedures controlling the maximum type I FWER.

Acknowledgements

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

References

[1] Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J. R. Statist. Soc.* ser.B, 57(1), 289-300.

[2] Benjamini, Y. and Hochberg, Y. (2000). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J. R. Statist. Soc.* ser.B, 57(1), 289-300.

[3] Benjamini, Y. and Daniel Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency *Ann. Statist.* 29(4): 11651188.

[4] Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, **50**, 1096-1121.

[5] 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.

[6] Dunnett, C. W. and Tamhane, A. C. (1992). A step-up multiple test procedure. *Journal of the American Statistical Association* **87**, 162–170.

[7] Dunnett, C. W. Horn, M. and Vollandt, R. (2001). Sample size determination in step-down and step-up multiple tests for comparing treatments with a control. *Journal of Statistical Planning and Inference*. **97**, 367-384.

[8] Hayter, A. J. and Tamhane, A. C. (1991). Sample size determination for step-down multiple test procedures; orthogonal contrasts and comparisons with a control. *Journal of Statistical Planning and Inference* **27**, 271-290.

[9] Horiuchi, K. and Matsuda, S. (2006). Evaluation of effectiveness of multiple comparison procedures controlling FDR. Academia. Mathematical sciences and information engineering : journal of the Nanzan Academic Society,6,17–30. (in Japanese)

[10] 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.

[11] Nagata, Y. and Yoshida, M. (1997). *Basic theory of statistical multiple comparison procedures* . Tokyo : Scientist Inc. (in Japanese)

[12] Tukey, J. W. (1953). The problem of multiple comparisons. Unpublished manuscript, Princeton University.