

Multiple Comparison with a Control Using Step-down Procedure

Bumrungsak Phuenaree* and Suttinee Kaewtaworn

Department of Mathematics, Faculty of Science,
Burapha University, 20131, Chonburi, Thailand.

*Corresponding author's email: bumrungsak [AT] buu.ac.th

ABSTRACT— *The purpose of this research was to compare the efficiency of single-step procedures and the step-down procedures in order to test for multiple comparison with a control group. Four tests; Dunnett test, Step-down Dunnett test, Bonferroni test and Bonferroni-Holm test, was considered. The performance of these tests was evaluated in terms of the family wise error rate, any-pair power and all-pairs power. A Monte Carlo simulation was performed with repeated 10,000 times. The results showed that the familywise error rate of all test statistics closed to the nominal level. The empirical power of step-down procedures were higher than the single-step procedures, and the step-down Dunnett test gave the highest power.*

Keywords— Multiple comparison with a control, Dunnett test, Bonferroni test, Step-down procedure

1. INTRODUCTION

In the problem of comparing several group with a control group, the multiple comparison procedure would be applied to detect the different in mean between a control group and the other group. The Dunnett test is the well-known and frequently used procedure in this situation (Bretz, Hothorn and Westfall 2011). The test statistic was performed by computing a Student's t-statistic for each group where the statistic compares the treatment group to a control group.

The other method that also commonly used was Bonferroni test. In the medical research, the Bonferroni procedure being the single most popular (Armstrong, 2014). For this method, the significant level would be divided by the number of tests. It called Bonferroni correction.

Holm(1979) proposed the step-down approach that improves the original Bonferroni approach, called Bonferroni-Holm test. Dunnett and Tamhane (1991) proposed the step-down Dunnett test that was an extension of the original Dunnett test. Both tests, using mean differences, were more powerful than their original approaches (Bretz, Hothorn and Westfall 2011).

Richer and MaCann (2013) compared the step-down procedure and the single-step of Dunnett and Bonferroni test, using median differences in the multiple comparison procedure with a control group. They showed that the step-down method had slightly greater power than the single-step method.

However, the efficiency of the step-down procedure of Dunnett and Bonferroni-Holm test had not been proved which procedure was better to detect the mean differences of each factor level with a control. Therefore, the main purpose of this work was to compare the efficiency of step-down Dunnett test and Bonferroni-Holm test for detecting mean differences. Three criteria, the family wise error rate, any-pair power and all-pairs power, were considered to figure out this problem.

2. MATERIALS AND METHODS

Consider a problem of balance one-way layout with a control group. The classical method for comparing several factor levels is analysis of variance; F-test. The null hypothesis is all population means are equal and the alternative hypothesis is the face that there are at least one population is not equal to the control group. If the null hypothesis is rejected, the next process is to figure out which population mean is different from the control group. This procedure is called multiple comparison procedure. Four tests; Dunnett test, Bonferroni test, step-down Dunnett test and Bonferroni-Holm test, are considered in this work.

2.1 Dunnett test (DUN)

The standard multiple comparison procedure with a control group is the Dunnett test. This test is commonly used method. The test statistics can be calculated as (Bretz, Hothorn and Westfall 2011)

$$t_i = \frac{\bar{y}_i - \bar{y}_0}{s \sqrt{\frac{1}{n_i} + \frac{1}{n_0}}}, \quad i = 1, \dots, m. \quad (1)$$

Where $\bar{y}_i = \sum_{j=1}^{n_i} \frac{y_{ij}}{n}$, $s^2 = \frac{\sum_{i=0}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{N - (k + 1)}$, \bar{y}_0 and n_0 are the mean and the sample size of the control group

respectively. N is the all sample size. n_i is the sample size of i^{th} factor level. m is the number of the factor level excluded the control group. The null hypothesis is rejected when t_i is larger than the critical value of Dunnett test at the significance level α .

2.2 Bonferroni test (BON)

The Bonferroni test is the one of several method for multiple comparison test. Consider null hypotheses $H_i (i = 1, \dots, k)$, the type I error (α) on all of them simultaneously can be control which compares unadjusted p -value (p_1, p_2, \dots, p_i) with α / k in stead of α , where k is the number of considered hypotheses. This procedure is equivalent to adjusted each p -value with $q_i = \min \{kp_i, 1\}$. The null hypothesis will be rejected, if q_i is not larger than α . The control of the familywise error rate follows from Bonferroni's inequality (Bretz, Hothorn and Westfall 2011)

$$FWER = P\left(\bigcup_{i \in k_0} \{q_i \leq \alpha\}\right) \leq \sum_{i \in k_0} P(q_i \leq \alpha) \leq k_0 \alpha / k \leq \alpha \quad (2)$$

2.3 Step-down Dunnett test (DUNSD)

Dunnett and Tamhane (1991) proposed the step-down Dunnett test that was extension of the original Dunnett test. The step of this procedure can be described as follow:

1. Consider the Dunnett test from the previous section to compute test statistics and p -value for testing each of factor levels to the control group. If the smallest p -value (or the maximum test statistics) falls below the significant level α , declared the factor level associated with that p -value to have a different location from the control group, and proceed to step 2.
2. Consider the rest of hypothesis; exclude the factor level declared different from the control group in step 1. If the smallest p -value (or the maximum test statistics) falls below the significant level α , declared the factor level associated with that p -value have a different location from the control group, and proceed to next step, otherwise stop.

2.4 Bonferroni-Holm test (BOH)

Holm (1979) proposed new multiple comparison procedure that improved Bonferroni procedure. The concept was adjusting the rejection criteria of each of the individual comparisons. The step of this procedure can be described as follow:

1. Consider the Bonferroni test from the previous section to compute test statistics and p -value for testing each of factor levels to the control group. If the smallest p -value (or the maximum test statistics) falls below the significant level α / k , declared the factor level associated with that p -value have a different location from the control group, and proceed to step 2.
2. Consider the rest of hypothesis; exclude the factor level declared different from the control group in step 1. If the smallest p -value (or the maximum test statistics) falls below the significant level $\alpha / (k - 1)$, declared the factor level associated with that p -value have a different location from the control group, and proceed to next step, otherwise stop.

2.5 Family wise error rate and power of the test

The familywise error rate (FWER) is the probability of at least one false rejection. It is frequently chosen as an overall measure of type I error to control level while testing multiple hypotheses. However, the FWER should be close to the significant level (α).

The power were defined in two ways. The first is the probability of rejecting at least one of the false hypotheses called any-pair power, and the second is the probability of rejecting all false hypotheses called all-pair power (Horn and Dunnett, 2004)

3. SIMULATION STUDY

A simulation was conducted to evaluate four tests; Dunnett’s test (DUN), Step-down Dunnett’s test (DUNSD), Bonferroni’s test (BON) and Bonferroni-Holm’s test (BOH). The data were generated under normal distribution with equal variances at 1 and 2; ($\sigma^2 = 1, 2$) and balance-sample sizes; $n = 3, 5, 7, 10$ and 15. The significant level were 0.05.

The number of factor levels (k) were set at 3, 4 and 5. The population means were defined as follow;

for $k=3$, I ($r=1$): $\mu_1 = \mu_2 = 0, \mu_3 = \Delta$, II ($r=2$): $\mu_1 = 0, \mu_2 = \mu_3 = \Delta$;

for $k=4$, I ($r=1$): $\mu_1 = \mu_2 = \mu_3 = 0, \mu_4 = \Delta$, II ($r=2$): $\mu_1 = \mu_2 = 0, \mu_3 = \mu_4 = \Delta$;

III ($r=3$): $\mu_1 = 0, \mu_2 = \mu_3 = \mu_4 = \Delta$

for $k=5$, I ($r=1$): $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0, \mu_5 = \Delta$, II ($r=2$): $\mu_1 = \mu_2 = \mu_3 = 0, \mu_4 = \mu_5 = \Delta$;

III ($r=3$): $\mu_1 = \mu_2 = 0, \mu_3 = \mu_4 = \mu_5 = \Delta$, IV ($r=4$): $\mu_1 = 0, \mu_2 = \mu_3 = \mu_4 = \mu_5 = \Delta$,

where r was the number of false hypotheses. The location parameter (μ) was fixed to 0 while the different in location parameter (Δ) were 0.0, 0.5 and 1.5. The first factor level will be set as a control group. A Monte Carlo simulation was performed with repeated 10,000 times. The package “DunnettTests” and “multcomp” in R program were used to calculate the p -value of all tests. The results for FWER and estimated powers were shown in Table 1-6.

4. RESULTS AND DISCUSSION

For each studied situation, two criteria were used to examine the efficiency tests. First criterion was the familywise error rate (FWER) that should be close to the significant level. The robustness evaluation was established on the Cochran limit (Hatchavanich, 2014); FWER was between 0.04 and 0.06. Second criterion was the estimated power of the test.

4.1 Familywise error rate (FWER)

In Table 1, it could be clearly seen that the empirical probability of familywise error rate of all tests fell well within range of Cochran’s limit. it implies that the rates were sufficiently close to the significant level. In other word, all tests could control the familywise error rate.

Table 1: The empirical familywise error rate, using locations of all factor levels are 0

k	r	$\sigma^2 = 1$				$\sigma^2 = 2$			
		DUN	DUNSD	BON	BOH	DUN	DUNSD	BON	BOH
3	3	0.0548	0.0548	0.0467	0.0467	0.0537	0.0537	0.0468	0.0468
	5	0.0487	0.0487	0.0437	0.0437	0.0508	0.0508	0.0455	0.0455
	7	0.0451	0.0451	0.0406	0.0406	0.0488	0.0488	0.0443	0.0443
	10	0.0538	0.0538	0.0505	0.0505	0.0500	0.0500	0.0455	0.0455
	15	0.0492	0.0492	0.0456	0.0456	0.0486	0.0486	0.0457	0.0457
4	3	0.0502	0.0501	0.0411	0.0411	0.0505	0.0506	0.0413	0.0413
	5	0.0495	0.0496	0.0417	0.0417	0.0506	0.0506	0.0424	0.0424
	7	0.0505	0.0505	0.0433	0.0433	0.0488	0.0488	0.0443	0.0443
	10	0.0479	0.0479	0.0414	0.0414	0.0531	0.0530	0.0458	0.0458
	15	0.0528	0.0528	0.0467	0.0467	0.0531	0.0532	0.0482	0.0482
5	3	0.0518	0.0520	0.0404	0.0404	0.0517	0.0516	0.0408	0.0408
	5	0.0513	0.0513	0.0416	0.0416	0.0475	0.0476	0.0400	0.0400
	7	0.0517	0.0516	0.0424	0.0424	0.0507	0.0507	0.0431	0.0431
	10	0.0498	0.0497	0.0421	0.0421	0.0483	0.0482	0.0418	0.0418
	15	0.0497	0.0496	0.0423	0.0423	0.0507	0.0506	0.0444	0.0444

4.2 Power of the test

Table 2-6 illustrated the estimated power values of all test. Both any-pair power and all-pair power were strongly increasing in sample size. When the different in location parameter became larger, all tests always had an increase in power for detecting the large difference. However, the power decreased when the variance of each sample increases. In addition, it could be noticed that the any-pair power of all tests was monotonically rising in r , while the all-pair power was not monotone.

DUNSD always had the highest power regardless of sample sizes and the number of factor levels. However, both step-down procedures; DUNSD and BOH, still had higher power than the single-step methods; DUN and BON. This was the same as Richter and McCann (2013) founded from their research that the step-down could make an increasing power. Although they used median difference in their work, the results should be corresponding to using mean difference in this work, because both mean and median were the location parameters.

Notice that the power of the DUN and the DUNSD were very close to each other when there was only one false hypothesis (see Figure 1-3). Moreover, they still gave the high power. Hence DUN can be used in this situation, as it is the commonly simple method even though it gave the little less power compared to DUNSD. Nevertheless, if there were two or more false hypotheses, the DUNSD should be used because it had much more power than DUN (see Table 2-6).

Table 2: The estimated power of all tests in three factor levels.

Δ	σ^2	n	r	Any-pair power				All-pair power			
				DUN	DUNSD	BON	BOH	DUN	DUNSD	BON	BOH
0.5	1	3	1	0.0476	0.0505	0.0419	0.0451	0.0476	0.0505	0.0419	0.0451
			2	0.0804	0.0804	0.0723	0.0723	0.0164	0.0269	0.0137	0.0258
		5	1	0.0676	0.0701	0.0609	0.0636	0.0676	0.0701	0.0609	0.0636
			2	0.1275	0.1275	0.1180	0.1180	0.0258	0.0443	0.0236	0.0427
		7	1	0.0947	0.0964	0.0891	0.0912	0.0947	0.0964	0.0891	0.0912
			2	0.1535	0.1535	0.1439	0.1439	0.0313	0.0515	0.0282	0.0507
		10	1	0.1298	0.1323	0.1229	0.1252	0.1298	0.1323	0.1229	0.1252
			2	0.2045	0.2045	0.1936	0.1936	0.0475	0.0787	0.0437	0.0769
		15	1	0.1871	0.1880	0.1789	0.1800	0.1871	0.1880	0.1789	0.1800
			2	0.2949	0.2949	0.2841	0.2841	0.0857	0.1299	0.0791	0.1272
	2	3	1	0.0348	0.0375	0.0306	0.0336	0.0348	0.0375	0.0306	0.0336
			2	0.0580	0.0580	0.0514	0.0514	0.0094	0.0167	0.0084	0.0161
		5	1	0.0390	0.0413	0.0354	0.0375	0.0390	0.0413	0.0354	0.0375
			2	0.0632	0.0632	0.0564	0.0564	0.0070	0.0148	0.0062	0.0141
		7	1	0.0382	0.0399	0.0351	0.0369	0.0382	0.0399	0.0351	0.0369
			2	0.0730	0.0730	0.0670	0.0670	0.0086	0.0175	0.0076	0.0167
		10	1	0.0486	0.0516	0.0446	0.0476	0.0486	0.0516	0.0446	0.0476
			2	0.0867	0.0867	0.0800	0.0800	0.0143	0.0241	0.0132	0.0231
		15	1	0.0687	0.0706	0.0633	0.0654	0.0687	0.0706	0.0633	0.0654
			2	0.1072	0.1072	0.1013	0.1013	0.0178	0.0312	0.0161	0.0300
1.5	1	3	1	0.2466	0.2477	0.2265	0.2277	0.2466	0.2477	0.2265	0.2277
			2	0.3409	0.3409	0.3162	0.3162	0.1403	0.2013	0.1275	0.1949
		5	1	0.4711	0.4720	0.4545	0.4554	0.4711	0.4720	0.4545	0.4554
			2	0.6189	0.6189	0.5982	0.5982	0.3286	0.4220	0.3088	0.4150
		7	1	0.6626	0.6635	0.6476	0.6485	0.6626	0.6635	0.6476	0.6485
			2	0.7960	0.7960	0.7828	0.7828	0.5103	0.6163	0.4934	0.6116
		10	1	0.8406	0.8423	0.8305	0.8323	0.8406	0.8423	0.8305	0.8323
			2	0.9341	0.9341	0.9290	0.9290	0.7472	0.8231	0.7317	0.8210
		15	1	0.9645	0.9651	0.9605	0.9611	0.9645	0.9651	0.9605	0.9611
			2	0.9912	0.9912	0.9899	0.9899	0.9343	0.9645	0.9300	0.9635
	2	3	1	0.0758	0.0782	0.0675	0.0701	0.0758	0.0782	0.0675	0.0701
			2	0.1211	0.1211	0.1090	0.1090	0.0306	0.0486	0.0267	0.0470
		5	1	0.1300	0.1314	0.1214	0.1230	0.1300	0.1314	0.1214	0.1230
			2	0.2083	0.2083	0.1925	0.1925	0.0567	0.0859	0.0517	0.0830
		7	1	0.1888	0.1913	0.1770	0.1796	0.1888	0.1913	0.1770	0.1796
			2	0.2837	0.2837	0.2699	0.2699	0.0827	0.1258	0.0749	0.1232
		10	1	0.2668	0.2684	0.2540	0.2560	0.2668	0.2684	0.2540	0.2560
			2	0.3992	0.3992	0.3831	0.3831	0.1401	0.2027	0.1309	0.1994
		15	1	0.4194	0.4202	0.4048	0.4057	0.4194	0.4202	0.4048	0.4057
			2	0.5805	0.5805	0.5677	0.5677	0.2562	0.3448	0.2441	0.3422

Table 3: The estimated power of all tests in four factor levels with $\Delta = 0.5$.

σ^2	n	r	Any-pair power				All-pair power			
			DUN	DUNSD	BON	BOH	DUN	DUNSD	BON	BOH
1	3	1	0.0361	0.0382	0.0284	0.0306	0.0361	0.0382	0.0284	0.0306
		2	0.0646	0.0660	0.0532	0.0547	0.0089	0.0150	0.0078	0.0133
		3	0.0856	0.0856	0.0710	0.0710	0.0048	0.0119	0.0037	0.0105
	5	1	0.0559	0.0575	0.0499	0.0517	0.0559	0.0575	0.0499	0.0517
		2	0.0960	0.0965	0.0844	0.0851	0.0163	0.0257	0.0142	0.0217
		3	0.1203	0.1203	0.1064	0.1064	0.0078	0.0192	0.0068	0.0183
	7	1	0.0733	0.0752	0.0663	0.0677	0.0733	0.0752	0.0663	0.0677
		2	0.1229	0.1242	0.1110	0.1119	0.0228	0.0330	0.0203	0.0285
		3	0.1614	0.1614	0.1450	0.1450	0.0129	0.0263	0.0110	0.0247
	10	1	0.1018	0.1035	0.0930	0.0943	0.1018	0.1035	0.0930	0.0943
		2	0.1724	0.1731	0.1580	0.1585	0.0387	0.0532	0.0350	0.0494
		3	0.2259	0.2260	0.2082	0.2082	0.0219	0.0479	0.0180	0.0460
	15	1	0.1527	0.1542	0.1420	0.1429	0.1527	0.1542	0.1420	0.1429
		2	0.2538	0.2547	0.2391	0.2400	0.0664	0.0867	0.0593	0.0814
		3	0.3190	0.3190	0.3001	0.3001	0.0339	0.0751	0.0296	0.0721
2	3	1	0.0237	0.0253	0.0190	0.0213	0.0237	0.0253	0.0190	0.0213
		2	0.0407	0.0423	0.0344	0.0356	0.0062	0.0099	0.0048	0.0082
		3	0.0567	0.0567	0.0474	0.0474	0.0028	0.0067	0.0022	0.0062
	5	1	0.0286	0.0302	0.0248	0.0268	0.0286	0.0302	0.0248	0.0268
		2	0.0459	0.0465	0.0389	0.0395	0.0057	0.0091	0.0046	0.0082
		3	0.0677	0.0678	0.0573	0.0573	0.0028	0.0066	0.0023	0.0064
	7	1	0.0336	0.0351	0.0297	0.0312	0.0336	0.0351	0.0297	0.0312
		2	0.0575	0.0592	0.0509	0.0522	0.0080	0.0120	0.0072	0.0109
		3	0.0760	0.0760	0.0682	0.0682	0.0026	0.0076	0.0022	0.0074
	10	1	0.0387	0.0404	0.0355	0.0368	0.0387	0.0404	0.0355	0.0368
		2	0.0675	0.0687	0.0601	0.0609	0.0084	0.0131	0.0070	0.0118
		3	0.0949	0.0952	0.0842	0.0842	0.0030	0.0089	0.0022	0.0085
	15	1	0.0484	0.0496	0.0440	0.0453	0.0484	0.0496	0.0440	0.0453
		2	0.0883	0.0892	0.0788	0.0795	0.0112	0.0178	0.0094	0.0161
		3	0.1183	0.1183	0.1069	0.1069	0.0050	0.0135	0.0044	0.0124

Table 4: The estimated power of all tests in four factor levels with $\Delta = 1.5$.

σ^2	n	r	Any-pair power				All-pair power			
			DUN	DUNSD	BON	BOH	DUN	DUNSD	BON	BOH
1	3	1	0.1016	0.1034	0.0879	0.0904	0.1016	0.1034	0.0879	0.0904
		2	0.3232	0.3234	0.2875	0.2880	0.1202	0.1492	0.1010	0.1353
		3	0.3774	0.3774	0.3394	0.3394	0.0833	0.1509	0.0683	0.1428
	5	1	0.4302	0.4314	0.4029	0.4041	0.4302	0.4314	0.4029	0.4041
		2	0.5897	0.5903	0.5580	0.5587	0.2879	0.3433	0.2613	0.3220
		3	0.6577	0.6578	0.6259	0.6259	0.2157	0.3497	0.1961	0.3428
	7	1	0.6205	0.6494	0.5951	0.5960	0.6205	0.6494	0.5951	0.5960
		2	0.7579	0.7590	0.7379	0.7391	0.4642	0.5210	0.4394	0.5051
		3	0.8343	0.8345	0.8135	0.8135	0.3879	0.5487	0.3644	0.5410
	10	1	0.8007	0.8019	0.7864	0.7874	0.8007	0.8019	0.7864	0.7874
		2	0.9147	0.9156	0.9040	0.9043	0.7013	0.7528	0.6813	0.7387
		3	0.9491	0.9491	0.9423	0.9423	0.6400	0.7815	0.6148	0.7769
	15	1	0.9541	0.9549	0.9485	0.9491	0.9541	0.9549	0.9485	0.9491
		2	0.9871	0.9874	0.9860	0.9863	0.9173	0.9403	0.9092	0.9345
		3	0.9931	0.9931	0.9924	0.9924	0.8859	0.9482	0.8737	0.9475
2	3	1	0.0623	0.0634	0.0528	0.0544	0.0623	0.0634	0.0528	0.0544
		2	0.1032	0.1044	0.0882	0.0891	0.0222	0.0319	0.0184	0.0279
		3	0.1294	0.1295	0.1090	0.1090	0.0145	0.0272	0.0114	0.0261
	5	1	0.1045	0.1055	0.0924	0.0934	0.1045	0.1055	0.0924	0.0934
		2	0.1800	0.1808	0.1623	0.1631	0.0430	0.0564	0.0368	0.0512
		3	0.2216	0.2217	0.1980	0.1980	0.0227	0.0526	0.0192	0.0506
	7	1	0.1580	0.1586	0.1434	0.1447	0.1580	0.1586	0.1434	0.1447
		2	0.2371	0.2378	0.2194	0.2199	0.0628	0.0826	0.0554	0.0769
		3	0.3080	0.3080	0.2844	0.2844	0.0367	0.0801	0.0320	0.0763
	10	1	0.2267	0.2278	0.2090	0.2106	0.2267	0.2278	0.2090	0.2106
		2	0.3468	0.3471	0.3256	0.3262	0.1161	0.1448	0.1051	0.1374
		3	0.4261	0.4262	0.3991	0.3991	0.0693	0.1359	0.0617	0.1312
	15	1	0.3609	0.3614	0.3426	0.3432	0.3609	0.3614	0.3426	0.3432
		2	0.5196	0.5200	0.4976	0.4983	0.2184	0.2635	0.2025	0.2500
		3	0.6115	0.6115	0.5858	0.5858	0.1514	0.2602	0.1375	0.2549

Table 5: The estimated power of all tests in five factor levels with $\Delta = 0.5$.

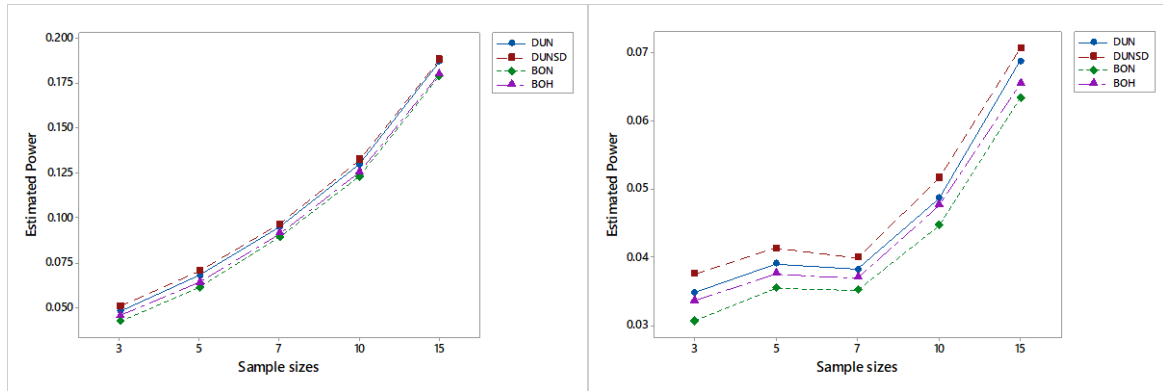
σ^2	n	r	Any-pair power				All-pair power			
			DUN	DUNSD	BON	BOH	DUN	DUNSD	BON	BOH
1	3	1	0.0356	0.0371	0.0286	0.0301	0.0356	0.0371	0.0286	0.0301
		2	0.0533	0.0549	0.0341	0.0436	0.0095	0.0134	0.0064	0.0106
		3	0.0780	0.0788	0.0633	0.0638	0.0042	0.0087	0.0033	0.0076
		4	0.0949	0.0949	0.0762	0.0762	0.0021	0.0064	0.0016	0.0058
	5	1	0.0473	0.0485	0.0413	0.0420	0.0473	0.0485	0.0413	0.0420
		2	0.0776	0.0784	0.0671	0.0679	0.0109	0.0150	0.0092	0.0121
		3	0.1129	0.1134	0.0968	0.0972	0.0063	0.0139	0.0046	0.0117
		4	0.1310	0.1310	0.1137	0.1137	0.0032	0.0099	0.0028	0.0089
	7	1	0.0607	0.0623	0.0531	0.0543	0.0607	0.0623	0.0531	0.0543
		2	0.1018	0.1025	0.0914	0.0920	0.0184	0.0224	0.0156	0.0198
		3	0.1417	0.1420	0.1234	0.1235	0.0078	0.0162	0.0063	0.0140
		4	0.1718	0.1718	0.1519	0.1519	0.0051	0.0164	0.0042	0.0151
	10	1	0.0888	0.0898	0.0786	0.0798	0.0888	0.0898	0.0786	0.0798
		2	0.1426	0.1434	0.1278	0.1287	0.0255	0.0351	0.0211	0.0297
		3	0.1894	0.1895	0.1711	0.1714	0.0140	0.0266	0.0109	0.0230
		4	0.2277	0.2277	0.2027	0.2027	0.0093	0.0269	0.0075	0.0249
	15	1	0.1377	0.1387	0.1260	0.1267	0.1377	0.1387	0.1260	0.1267
		2	0.2195	0.2203	0.1994	0.2000	0.0544	0.0669	0.0472	0.0603
		3	0.2794	0.2798	0.2546	0.2550	0.0267	0.0481	0.0220	0.0436
		4	0.3257	0.3257	0.2973	0.2973	0.0171	0.0542	0.0145	0.0535
2	3	1	0.0184	0.0194	0.0142	0.0156	0.0184	0.0194	0.0142	0.0156
		2	0.0357	0.0368	0.0289	0.0299	0.0047	0.0077	0.0035	0.0060
		3	0.0472	0.0478	0.0382	0.0388	0.0020	0.0040	0.0015	0.0034
		4	0.0615	0.0615	0.0472	0.0472	0.0004	0.0032	0.0003	0.0031
	5	1	0.0219	0.0228	0.0185	0.0196	0.0219	0.0228	0.0185	0.0196
		2	0.0396	0.0404	0.0337	0.0350	0.0046	0.0068	0.0034	0.0059
		3	0.0565	0.0571	0.0477	0.0480	0.0019	0.0042	0.0015	0.0038
		4	0.0728	0.0730	0.0621	0.0621	0.0009	0.0035	0.0006	0.0032
	7	1	0.0263	0.0279	0.0219	0.0230	0.0263	0.0279	0.0219	0.0230
		2	0.0438	0.0450	0.0382	0.0393	0.0053	0.0079	0.0037	0.0066
		3	0.0641	0.0646	0.0549	0.0553	0.0016	0.0044	0.0011	0.0036
		4	0.0831	0.0832	0.0715	0.0715	0.0010	0.0043	0.0004	0.0040
	10	1	0.0298	0.0315	0.0256	0.0276	0.0298	0.0315	0.0256	0.0276
		2	0.0587	0.0598	0.0514	0.0529	0.0064	0.0095	0.0050	0.0080
		3	0.0761	0.0765	0.0660	0.0665	0.0025	0.0059	0.0023	0.0053
		4	0.0915	0.0915	0.0795	0.0795	0.0016	0.0056	0.0015	0.0055
	15	1	0.0389	0.0405	0.0339	0.0356	0.0389	0.0405	0.0339	0.0356
		2	0.0728	0.0739	0.0646	0.0658	0.0105	0.0139	0.0091	0.0126
		3	0.0910	0.0912	0.0817	0.0821	0.0044	0.0092	0.0034	0.0084
		4	0.1278	0.1279	0.1143	0.1143	0.0024	0.0053	0.0017	0.0049

Table 6: The estimated power of all tests in five factor levels with $\Delta = 1.5$.

σ^2	n	r	Any-pair power				All-pair power			
			DUN	DUNSD	BON	BOH	DUN	DUNSD	BON	BOH
1	3	1	0.2098	0.2107	0.1820	0.1829	0.2098	0.2107	0.1820	0.1829
		2	0.3060	0.3064	0.2657	0.2659	0.1103	0.1291	0.0896	0.1121
		3	0.3701	0.3704	0.3232	0.3236	0.0724	0.1112	0.0564	0.0997
		4	0.4070	0.4070	0.3557	0.3557	0.0545	0.1273	0.0424	0.1201
	5	1	0.4009	0.4021	0.3712	0.3726	0.4009	0.4021	0.3712	0.3726
		2	0.5594	0.5595	0.5234	0.5238	0.2676	0.2998	0.2393	0.2789
		3	0.6320	0.6321	0.5936	0.5940	0.1898	0.2641	0.1698	0.2459
		4	0.7041	0.7041	0.6688	0.6689	0.1580	0.3092	0.1363	0.2991
	7	1	0.5917	0.5927	0.5656	0.5665	0.5917	0.5927	0.5656	0.5665
		2	0.7502	0.7506	0.7227	0.7231	0.4420	0.4821	0.4094	0.4560
		3	0.8111	0.8116	0.7861	0.7867	0.3467	0.4461	0.3194	0.4265
		4	0.8512	0.8513	0.8295	0.8295	0.3087	0.4995	0.2800	0.4898
	10	1	0.7961	0.7968	0.7741	0.7750	0.7961	0.7968	0.7741	0.7750
		2	0.8991	0.8999	0.8856	0.8859	0.6829	0.7180	0.6553	0.6979
		3	0.9440	0.9441	0.9346	0.9349	0.6049	0.7080	0.5734	0.6931
		4	0.9597	0.9597	0.9525	0.9525	0.5445	0.7428	0.5135	0.7378
	15	1	0.9414	0.9423	0.9338	0.9351	0.9414	0.9423	0.9338	0.9351
		2	0.9858	0.9863	0.9834	0.9840	0.9030	0.9205	0.8896	0.9116
		3	0.9928	0.9929	0.9913	0.9914	0.8681	0.9172	0.8509	0.9118
		4	0.9979	0.9979	0.9950	0.9950	0.8264	0.9355	0.8049	0.9337
2	3	1	0.0491	0.0506	0.0387	0.0402	0.0491	0.0506	0.0387	0.0402
		2	0.0897	0.0907	0.0722	0.0730	0.0177	0.0240	0.0142	0.0195
		3	0.1173	0.1177	0.0954	0.0957	0.0094	0.0162	0.0073	0.0148
		4	0.1408	0.1409	0.1154	0.1154	0.0072	0.0193	0.0052	0.0177
	5	1	0.0927	0.0936	0.0805	0.0818	0.0927	0.0936	0.0805	0.0818
		2	0.2468	0.1474	0.1270	0.1276	0.0304	0.0379	0.0260	0.0335
		3	0.1909	0.1916	0.1673	0.1676	0.0154	0.0307	0.0121	0.0266
		4	0.2258	0.2258	0.2007	0.2007	0.0107	0.0336	0.0085	0.0314
	7	1	0.1374	0.1389	0.1238	0.1249	0.1374	0.1389	0.1238	0.1249
		2	0.2123	0.2127	0.1918	0.1923	0.0517	0.0634	0.0431	0.0568
		3	0.2668	0.2670	0.2410	0.2414	0.0270	0.0468	0.0233	0.0420
		4	0.3167	0.3167	0.2869	0.2869	0.0197	0.0578	0.0168	0.0551
	10	1	0.2099	0.2106	0.1921	0.1930	0.2099	0.2106	0.1921	0.1930
		2	0.3198	0.3202	0.2922	0.2928	0.0969	0.1157	0.0851	0.1048
		3	0.3905	0.3905	0.3617	0.3619	0.0560	0.0902	0.0479	0.0831
		4	0.4515	0.4515	0.4208	0.4208	0.0401	0.1011	0.0336	0.0974
	15	1	0.3322	0.3326	0.3095	0.3101	0.3322	0.3326	0.3095	0.3101
		2	0.4866	0.4869	0.4602	0.4607	0.1891	0.2213	0.1700	0.2016
		3	0.5705	0.5708	0.5424	0.5425	0.1312	0.1890	0.1167	0.1794
		4	0.6411	0.6412	0.6186	0.6186	0.0922	0.2026	0.0789	0.1921

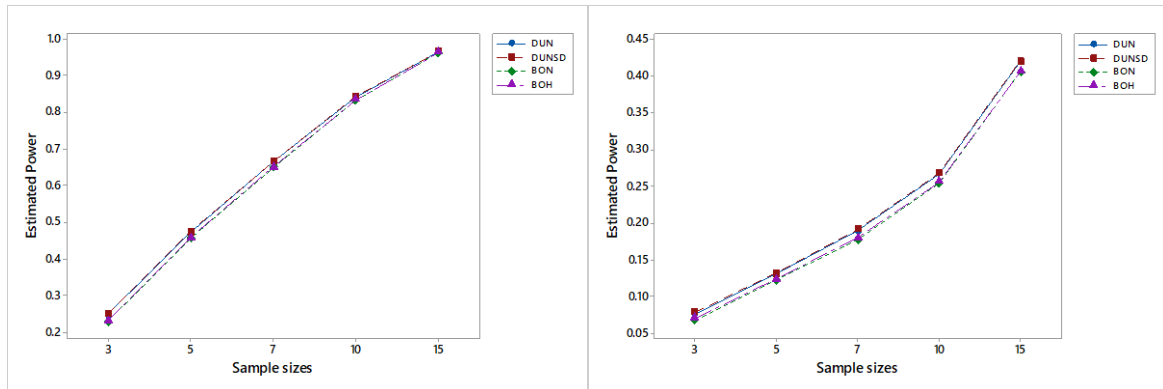
5. CONCLUSION

Based on the numerical studies, Increasing sample size and decreasing variance of each sample were founded to improve the test power for all testing procedures. The results have been demonstrated that the step-down Dunnett test became the best test for detecting the difference in location mean parameter. However, the single-step Dunnett test could be used when there were two or more groups difference from the control group because it had the second highest power in this situation. Moreover, it was easier than the step-down procedure.



(a) $\Delta = 0.5, \sigma^2 = 1$

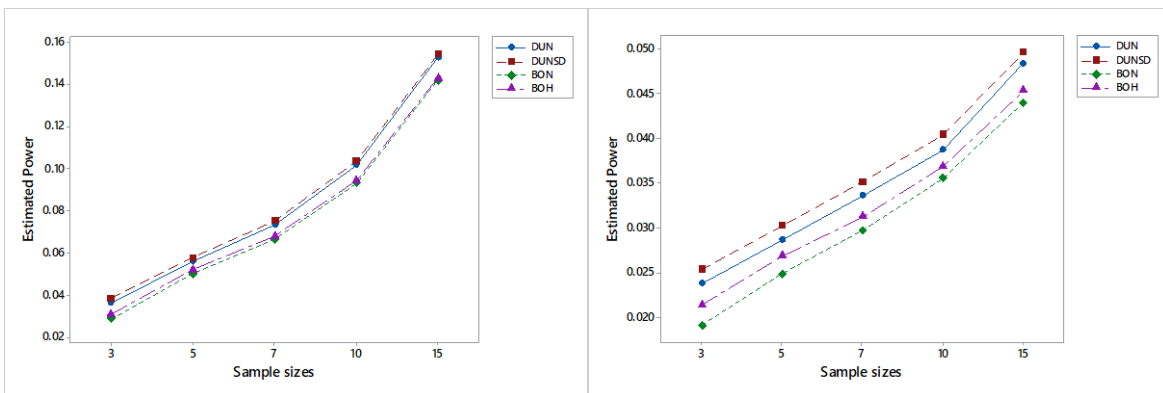
(b) $\Delta = 0.5, \sigma^2 = 2$



(c) $\Delta = 1.5, \sigma^2 = 1$

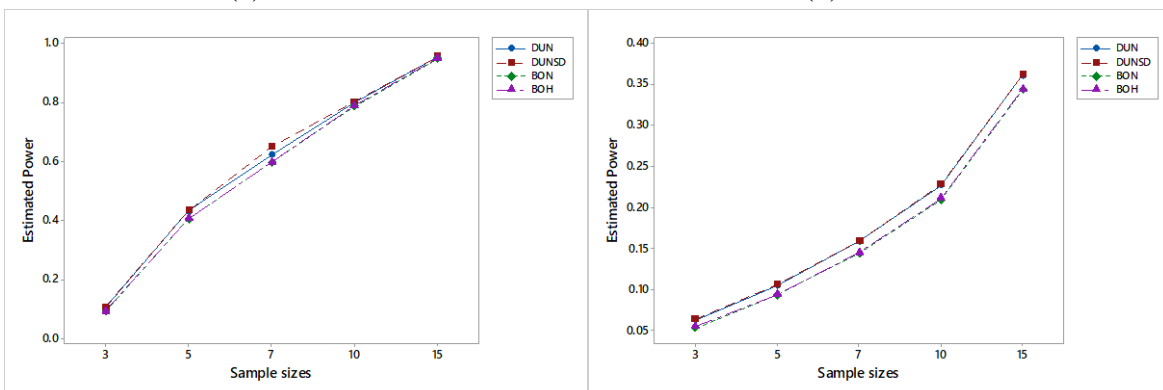
(d) $\Delta = 1.5, \sigma^2 = 2$

Figure 1: The estimated power with a different of false hypothesis ($r = 1$) ; $k = 3$



(a) $\Delta = 0.5, \sigma^2 = 1$

(b) $\Delta = 0.5, \sigma^2 = 2$



(c) $\Delta = 1.5, \sigma^2 = 1$

(d) $\Delta = 1.5, \sigma^2 = 2$

Figure 2: The estimated power with a different of false hypothesis ($r = 1$) ; $k = 4$

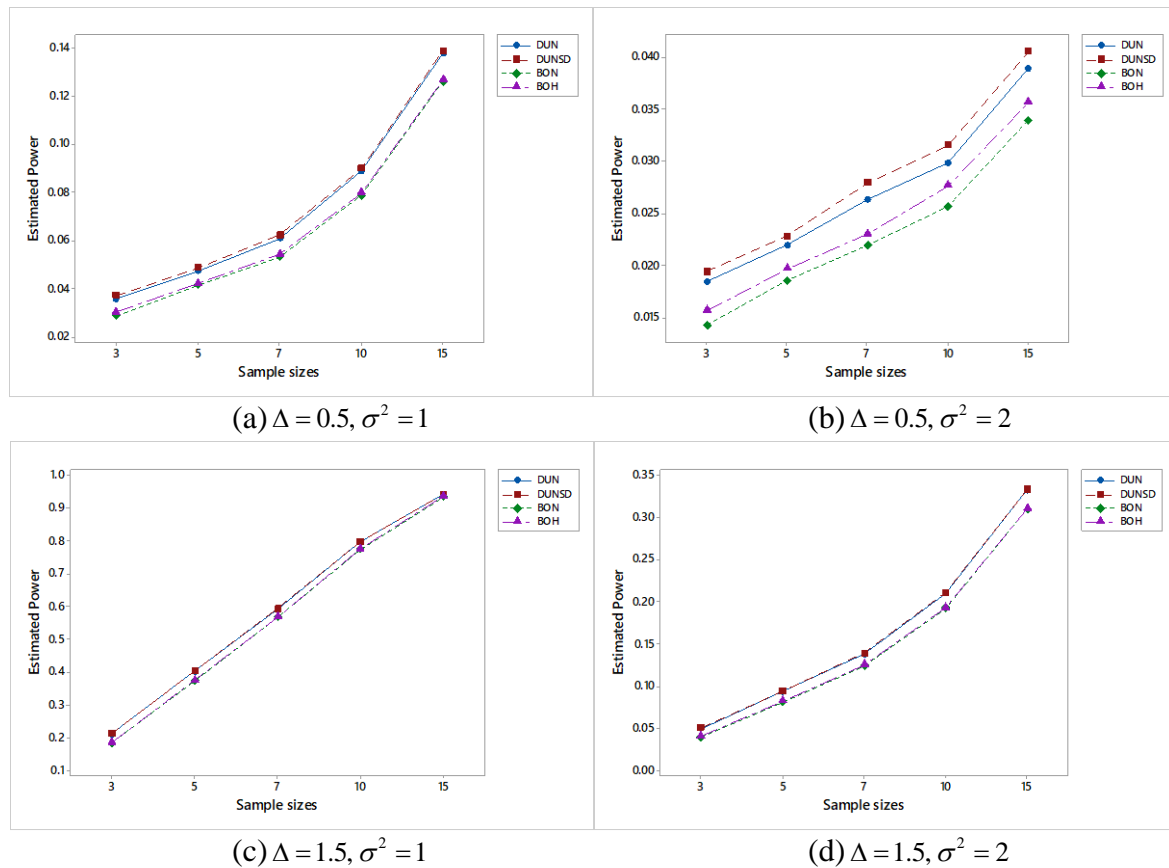


Figure 3: The estimated power with a different of false hypothesis ($r = 1$) ; $k = 5$

6. ACKNOWLEDGEMENTS

This research was supported by the Faculty of Science, Burapha University, Thailand.

7. REFERENCES

- [1] Armstrong R.A., “When to use the Bonferroni correction”, *Ophthalmic & Physiological Optics* 34, pp. 502-508, 2014
- [2] Bretz F., Hothorn T., and Westfall P., *Multiple comparisons using R*. Boca Raton, FL: Taylor & Francis, 2011
- [3] Dunnett C.W. and Tamhane, A.C., “Step-down multiple tests for comparing treatments with a control in unbalanced one-way layouts”, *Statistics in Medicine* vol.10, pp.939-947, 1991
- [4] Hatchavanich D., “A comparison of type I error and power of Bartlett’s test, Levene’s test and O’Brien’s test for homogeneity of variance tests”, *Southeast-Asian journal of sciences* vol.3, no 2, pp. 181-194, 2014
- [5] Holm S., “A simple sequentially rejective multiple test procedure”, *Scandinavian journal of statistics*, vol.6, no.2, pp. 65-70, 1979
- [6] Horn M., and Dunnett C.W., “Power and sample size comparisons of stepwise FEW and FDR controlling test procedures in the normal many-one case”. *Lecture Notes Monograph Series* 47, pp.48-64, 2004
- [7] Richter S.J., and McCann M.H., “Step-down multiple comparison procedures based on medians and permutation tests”. *Communications in statistics: simulation and computation*, vol.38, no.8, pp.1551-1561, 2009
- [8] Richter S.J., and McCann, M.H., “Simultaneous multiple comparisons with a control using median differences and permutation tests”, *Statistics and Probability Letters*, vol. 83, pp. 1167-1173, 2013