Multiple Comparison Procedures in the Unequal Variance Case

Shin-ichi MATSUDA

Summary

The purpose of this thesis is to give the results of the study on the principle of multiple comparisons and of the research of multiple comparison procedures when the population variances are unequal. There are many situations where multiple comparison procedures should be adapted. In this thesis, we focus only on the one-way layout model with normally distributed errors.

In Chpater 1, we give an introduction of this thesis. In Chpater 2, we give some overview on the model and already proposed procedures.

Chapter 3 is devoted to the problem of the principle of multiple comparisons. We consider the concept of powers for multiple comparisons as an especially important subject. The reason why many multiple comparison procedures have been proposed is that the comparison of them is not easy because the concept of powers is not clearly defined. Therefore, we compare several multiple comparison procedures using the known definitions of powers and proposed ones and make features of those concepts of powers clear. Moreover, we examine power properties of the procedures and discuss which procedure is adequate in the practical situations.

Next, we discuss multiple comparison procedures when the variances are unequal in Chapters 4 and 5. The theory and methods of multiple comparison procedures under the assumption of the homogeneity of variances have been well developed. However, the research of multiple comparison procedures under the unequal variance cases is not satisfactory and in fact most of statistical software packages do not support them.

In Chapter 4, we examine already proposed multiple comparison procedures when the variances are unequal to investigate the properties of them. We find that the procedures do not meet the required nominal significant level when the populations have unequal sample sizes and/or the fluctuation of the variances are large. We propose new procedures to correct these defective features. A procedure proposed can maintain the nominal significant level in the wide range of parameters. Furthermore, in the situation that the homogeneity of variances is doubtful, we discuss a system with a preliminary test for the homogeneity of the variances and how to improve the system. The system with a modified preliminary test can maintain the nominal significant level in the wide range of parameters.

In the final chapter, we discuss a way how to construct multiple comparison procedures which do not need a preliminary test. The proposed procedure is based on a loss function. The reason why we consider such procedure is to avoid the discontinuity of the conclusions of multiple comparison procedures at the critical points of the preliminary test. We feel there leaves something to be improved. However, they have relatively good performances.

Contents

Chapter 1

Introduction

In days of old, they tested multiple groups of data assuming the normal distribution using the analysis of variance, but this method cannot point out which group is different from which group. In order to conquer this disadvantage, multiple comparison procedures were produced. Fisher (1935) proposed a primary multiple comparison procedure. This procedure practices the preliminary F-test before repeated pairwise comparisons, but it was shown that under some configurations of population means the real significant level was nearby that in the case of repeated pairwise comparisons without the preliminary F-test.

It is the 1950's when procedures being useful at present were proposed. Those were Tukey's (1953) procedure and Scheffé's (1953) procedure. Furthermore, the basic concept for the Type I error on multiple comparisons was also established then. After that, several procedures based on different types of comparison, e.g. Dunnett (1955), were proposed, but large development was not occurred.

In the 1970's, when powers of multiple comparison procedures were considered, stepwise procedures that were improved on the point of powers were proposed. The general principle of stepwise procedures, i.e. the closure method, were completed by Marcus, Peritz and Gabriel (1976). Moreover, Einot and Gabriel (1975) and Ramsey (1978) compared known procedures on the basis of the powers.

On the other hand, procedures not assuming the normal distribution were developed. Those were nonparametric procedures (see Steel 1959, 1960, Dwass 1960, Sen 1969, Shirley 1977 and Matsuda 1988) and sequentially rejective procedures of Bonferroni type (see Holm 1979, Shaffer 1986 and Holland and Copenhaver 1987).

In such history, Hochberg and Tamhane (1987) is an important book about multiple comparison procedures, which includes the previous historical changes of multiple comparisons and various procedures. In this book, However, the principle of multiple comparisons is not complete enough and multiple comparison procedures under singular conditions are not solved.

In this thesis, we lay emphasis on the concept of powers especially in the principle of multiple comparisons and on the case of unequal variances especially among singular conditions. In particular, we consider pairwise comparisons as multiple comparisons.

In Chapter 2, we mention the principle of multiple comparisons and known procedures. We improve the Type I error, which is not complete enough in Hochberg and Tamhane (1987), introducing the consideration of Holland and Copenhaver (1987).

In Chapter 3, we mention powers of multiple comparisons and the merits and demerits of known procedures. We increase the kind of power in addition to that in Einot and Gabriel (1975) and Ramsey (1978). (The powers considered in this thesis are *all-pairs power*, *restricted all-pairs power*, *minimum difference power*, *maximum difference power*, *mean rejective rate* and *weighted mean rejective rate*. Three powers are newly proposed.) On the basis of those powers, we compare several procedures, which newly include a sequentially rejective procedure. (The methods considered in this chapter are *Tukey's procedure* (*Tukey-Kramer procedure*), *Tukey-Welsch procedure*, *Peritz's procedure* and *Holland-Copenhaver procedure.*)

In Chapter 4, we will propose new multiple comparison procedures under the unequal variance condition. Firstly, we study features of known procedures (*GH-procedure*, *T3-*

procedure and *C-procedure*) in detail, and produce new procedures (*GHC-procedure* and *GHC2-procedure*) that conquer the disadvantage of the previous three procedures. Furthermore, we study the effect of the procedures and the performance of a nonparametric procedure (*Steel-Dwass procedure*) under realistic situations. Secondly, we discuss the influence of preliminary tests, and propose a new preliminary test.

In Chapter 5, we will propose new multiple comparison procedures based on a loss function (*LMGH-procedure*, *LP1-procedure* and *LP2-procedure*) in order to avoid the disadvantage of procedures with a preliminary test. And, we compare them with the new preliminary test system in Chapter 4.

Mainly, Chapter 3 is based on Matsuda and Nagata (1990), Chapter 4 is based on Matsuda (1993) and Matsuda (1994) and Chapter 5 is based on Matsuda (1997).

Chapter 2

Model and Known Procedures for Multiple Comparisons

In this chapter, we mention a model and known procedures.

We focus one-way layouts as a model. And, we explain known procedures treated in Chapters 3 and 4.

2.1 Multiple Comparisons for One-Way Layouts

Multiple comparisons are applied for various models. In this thesis, we consider the following model of one-way layout.

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, 2, \ldots, k; j = 1, 2, \ldots, n_i,$$

μ_i : unknown parameters,

ε_{ij} : independent random variables from $N(0, \sigma_i^2)$.

For this model, we practice pairwise multiple comparisons among all treatments. In other words, we make a family of null hypotheses for multiple comparisons

$$\{\mu_i = \mu_h : 1 \leq i < h \leq k\}.$$

This situation is one of important multiple comparisons, on which many procedures have been proposed for about half a century (e.g. see Hochberg and Tamhane 1987).

2.2 Type I Error Rates for Multiple Comparison Procedures

The reasons why many multiple comparison procedures are proposed is that there are several Type I error rates and several powers for them. Now, we make a list of Type I error rates for multiple comparison procedures.

For simplification in this thesis, null hypotheses $\mu_1 = \mu_2 = \mu_3$ and $\mu_1 = \mu_2$ are denoted by $H_{\{1,2,3\}}$ and $H_{\{1,2\}}$, respectively. Generally, the null hypothesis for an index subset P

$$\mu_i = \mu_h,\ i, h \in P$$

is denoted by H_P. Especially, *the overall null hypothesis* is denoted by H_K for $K = \{1, 2, \ldots, k\}$. Moreover, let $p = \#P =$ (The number of elements in the set P) and call it the size of the null hypothesis H_P.

We define $D_{\{i,h\}}$ for $1 \leq i < h \leq k$ as random variables that is 1 if $H_{\{i,h\}}$ is rejected and 0 if it is retained. (In multiple comparisons, we use the term 'retain' instead of 'accept', because the decision for a pair of treatments may contradict that for other pairs; e.g. we reject $H_{\{1,3\}}$ but do not reject $H_{\{1,2\}}$ and $H_{\{2,3\}}$.) Furthermore, let $\mu = (\mu_1, \ldots, \mu_k)$ be the true point in the parameter space of population means.

Now, we can define the following three Type I error rates for multiple comparison procedures.

- *Familywise error rate* (FWE):

$$\Pr\left\{\sum_{(i,h)\in I_0} D_{\{i,h\}} > 0 | \mu\right\},$$

- *Pre-family error rate* (PFE):

$$\mathrm{E}\left\{\sum_{(i,h)\in I_0} D_{\{i,h\}} | \mu\right\},$$

- *Per-comparison error rate* (PCE):

$$\mathrm{E}\left\{ \sum_{(i,h)\in I_0} D_{\{i,h\}} | \mu \right\} / \# I_0,$$

where $I_0 = \{(i,h) : 1 \leq i < j \leq k,\ \mu_i = \mu_h\}$.

Since we consider a similar situation as ANOVA in this thesis, we need to control the Type I error rate on the overall null hypothesis. Therefore, we should use the Type I FWE.

Furthermore, we call the upper limit of the Type I FWE for $\boldsymbol{\mu}$ *the generalized Type I FWE* (see Holland and Copenhaver 1987).

2.3 Known Multiple Comparison Procedures

Multiple comparison procedures for one-way layouts can be classified into three large groups: single-step procedures, stepwise procedures and sequentially rejective procedures. Stepwise procedures are the extension of single-step procedures to increase the power. On the other hand, sequentially rejective procedures are based on Bonferroni inequality or its similarities. Therefore, they have another approach against the previous two types, and the behavior of their relative power is not obvious.

In this section, we consider procedures in the case of equal variances. (Procedures in the case of unequal variances are considered in the following section.)

Besides, any procedure shown in the following can control theoretically the generalized Type I FWE in the case of equal variances.

2.3.1 Tukey's Procedure (Tukey-Kramer Procedure)

Tukey's procedure (*Tukey-Kramer procedure* in the unbalanced case) has the following $(1-\alpha)$-level simultaneous critical region for $H_{\{i,h\}}$.

$$|\overline{Y}_{i\cdot} - \overline{Y}_{h\cdot}| \geq Q_{k,\nu}^{(\alpha)} \sqrt{\frac{1}{2}\overline{S}^2 (1/n_i + 1/n_h)},$$

where

$$\begin{aligned}
\overline{Y}_{i.} &= \sum_{j=1}^{n_i} Y_{ij}/n_i, \\
\nu &= \sum_{i=1}^{k} n_i - k, \\
\overline{S}^2 &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i.})^2/\nu,
\end{aligned}$$

and $Q_{k,\nu}^{(\alpha)}$ is the upper α point of Studentized range distribution with the parameter k and the degree of freedom ν (see tables in Hochberg and Tamhane 1987).

Tukey's procedure can control the generalized Type I FWE exactly, but Tukey-Kramer procedure in the unbalanced case is conservative, which is proved by Hayter (1984).

Besides, Scheffé's procedure that is a well-known single-step procedure is worse at pairwise comparisons than Tukey's procedure, which is shown in Hochberg and Tamhane (1987). Therefore, we do not deal with it in this thesis.

2.3.2 Tukey-Welsch Procedure

Tukey-Welsch procedure is a stepdown procedure, which uses the following nominal level α_p for the test of null hypothesis H_P. This setting is called *Tukey-Welsch specification* (Tukey 1953, Welsch 1972).

$$\begin{aligned}
\alpha_p &= 1 - (1-\alpha)^{p/k}, \quad p = 1, 2, \ldots, k-2, \\
\alpha_{k-1} &= \alpha_k = \alpha.
\end{aligned}$$

Generally, we denote Z_P as a statistic for the test of H_P and ξ_p as the corresponding critical value. The procedure is the following.

1. Calculate a value of the test statistic Z_K and obtain the critical value ξ_k.

2. Test the overall null hypothesis H_K.

   (a) If $Z_K < \xi_k$ then retain all H_P's, $P \subseteq K$ without further tests.

   (b) If $Z_K \geq \xi_k$ then reject H_K and proceed to the next step with $m = k - 1$.

3. For each subset P of size m that has not retained yet, calculate a value of the test statistic Z_P and obtain the critical value ξ_p (#P=p=m). Proceed to the next step.

4. Test the null hypothesis H_P.

   (a) If $Z_P < \xi_p$ then retain all H_Q, $Q \subseteq P$.

   (b) If $Z_P \geq \xi_p$ then reject H_P.

5. If all H_P's is retained or $m = 2$ then terminate the procedure, otherwise reset $m-1$ to m newly and repeat this procedure from the third step.

In practice, test statistics and critical values are

- Q-statistics:

$$
\begin{aligned}
Z_P &= \max_{i,h \in P} \left(\frac{|\overline{Y}_{i.} - \overline{Y}_{h.}|}{\sqrt{\frac{1}{2}\overline{S}^2(1/n_i + 1/n_h)}} \right), \\
\xi_p &= \begin{cases} \max\{Q_{p,\nu}^{(\alpha_p)}, \xi_{p-1}\} & p = 3, \ldots, k \\ Q_{p,\nu}^{(\alpha_p)} & p = 2 \end{cases},
\end{aligned}
$$

or

- F-statistics:

$$
\begin{aligned}
Z_P &= \frac{\sum_{i \in P} n_i \overline{Y}_{i.}^2 - (\sum_{i \in P} n_i \overline{Y}_{i.})^2 / \sum_{i \in P} n_i}{(p-1)\overline{S}^2}, \\
\xi_p &= F(p-1, \nu, \alpha_p),
\end{aligned}
$$

where $F(p-1, \nu, \alpha_p)$ is the upper α_p point of F-distribution with the degrees of freedom $p-1$ and ν.

Besides, $Q_{p,\nu}^{(\alpha_p)}$, which is not on tables in Hochberg and Tamhane (1987), is calculated by using the program in Yoshida (1988).

By the way, we need the previous monotone modification of critical values of Q-statistics to omit extra steps as keeping the consonance: the property that whenever any nonminimal H_P is rejected, at least one of its components is also rejected. (See Hochberg and Tamhane (1987).) Even if it is not modified, it controls the generalized Type I FWE at level α.

2.3.3 Peritz's Procedure

Peritz's (1970) *procedure* is a stepdown procedure, which is the combination of Tukey-Welsch procedure and Newman-Keuls procedure. Besides, *Newman-Keuls procedure* is the procedure with *Newman-Keuls specification* instead of Tukey-Welsch specification in the previous subsection, that is, all α_p's are made α (Newman 1939, Keuls 1952). (Newman-Keuls procedure is not control the generalized Type I FWE at level α. Therefore we do not deal with it alone.) The procedure based on the algorithm of Begun and Gabriel (1981) is the following.

1. Practice Tukey-Welsch procedure and Newman-Keuls procedure separately and classify all null hypotheses H_P's into the following three groups.
   - **(a)** Null hypotheses Rejected by both procedures.
   - **(b)** Null hypotheses Retained by both procedures.
   - **(c)** Null hypotheses Retained by Tukey-Welsch procedure but rejected by Newman-Keuls procedure.
2. (a) and (b) are final results because both procedures have the same result.

But, H_P in (c), which is called *a contentious null hypothesis*, is judged in the following manner.

3. Let m be the maximum size of contentious null hypotheses. If no contentious null hypothesis then terminate the procedure.

4. For each contentious null hypothesis H_P with size m, reject it if it satisfies the following condition, otherwise retain it.

   " For any subset $Q \subseteq P^c$, the null hypothesis H_Q is rejected by Tukey-Welsch procedure,"

   where P^c is the complementary set of P.

5. If H_P is retained at the previous step then retain all contentious subset null hypotheses for H_P.

6. Repeat this procedure from the third step.

2.3.4 Holland-Copenhaver Procedure

Holland-Copenhaver (1987) *procedure* is a sequentially rejective type procedure. Although many Bonferroni-type procedures are proposed, we consider this procedure in this thesis because it is the most powerful procedure among Holm-type procedures. Moreover, Holm-type procedures control the the generalized Type I FWE theoretically (see Holm 1979, Shaffer 1986).

Holland-Copenhaver procedure is based on Šidák's (1967) inequality that is better than Bonferroni's inequality. Thus, it is a modification of Šidák's procedure, which is a single-step procedure, to sequentially rejective type.

Besides, Šidák's inequality is the following. When $(T_1, \ldots, T_k)$ denotes a vector of

random variables with the multivariate t-distribution,

$$\Pr(|T_1| \leq c_1, \ldots, |T_k| \leq c_k) \geq \prod_{i=1}^{k} \Pr(|T_i| \leq c_i).$$

Furthermore, Šidák's procedure is better than Bonferroni's procedure, but worse than Tukey's procedure (see Hochberg and Tamhane 1987).

The procedure is the following.

1. Order two-side p-values of t-statistics $(\overline{Y}_{i.} - \overline{Y}_{h.})/\sqrt{\overline{S}^2(1/n_i + 1/n_h)}$, $1 \leq i < h \leq k$ for the null hypothesis $\mu_i = \mu_h$ and call them

$$P_1 \leq \cdots \leq P_M,$$

where $M = k(k-1)/2$. Set $m = 1$.

2. In the family of null hypotheses $\{\mu_i = \mu_h\}$, let t_m be the maximum number of simultaneous true null hypotheses against $m - 1$ false null hypotheses (see the table in Holland and Copenhaver 1987).

3. Define the function $C(x)$ as $C(x) = 1 - (1 - \alpha)^{1/x}$ and calculate $C(t_m)$.

4. **(i)** When $P_m > C(t_m)$, reject null hypotheses corresponding to $P_1, \ldots, P_{m-1}$ and retain null hypotheses corresponding to $P_m, \ldots, P_M$.

   (If $m = 1$ then retain all null hypotheses.)

   And, terminate the procedure.

   (ii) When $P_m \leq C(t_m)$,

   if $m < M$ then reset $m + 1$ to m newly and repeat from the second step,

   otherwise reject all null hypotheses and terminate the procedure.

2.4 Known Procedures in the Unequal Variance Case

We show three known procedures with the nominal level α to inspect in Chapter 4.

All procedures in this section are single step procedures. We may not deal with stepwise procedures because the behavior of the real significant level of them is same as that for the corresponding single step procedure. On practical use, we can easily extend them to stepwise procedures. Moreover, Holland-Copenhaver procedure has similar performance to T3-procedure because both are based on Šidák's procedure.

Furthermore, whether the following procedures keep the generalized Type I FWE has been studied by Monte Carlo simulation. Hochberg and Tamhane (1987) report that T3-procedure and C-procedure keep it but GH-procedure does not so.

Besides, on any procedure in the following, if $|T_{ih}| \geq \xi_{ih}$ then we judge μ_i and μ_h are different.

2.4.1 GH-Procedure

GH-procedure is proposed by Games and Howell (1976).

The test statistic is

$$T_{ih} = \frac{\overline{Y}_{i.} - \overline{Y}_{h.}}{\sqrt{S_i^2/n_i + S_h^2/n_h}},$$

where

$$S_i^2 = \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i.})^2/\nu_i, \quad \nu_i = n_i - 1.$$

And, the critical value is

$$\xi_{ih} = Q_{k,\nu_{ih}}^{(\alpha)}/\sqrt{2},$$

where

$$\nu_{ih} = \frac{(S_i^2/n_i + S_h^2/n_h)^2}{S_i^4/n_i^2(n_i - 1) + S_h^4/n_h^2(n_h - 1)}.$$

2.4.2 T3-Procedure

T3-procedure is proposed by Dunnett (1980).

The test statistic is same as GH-procedure.

But, the critical value is

$$\xi_{ih} = |M|^{(\alpha)}_{m,\nu_{ih}},$$

where $m = k(k-1)/2$ and $|M|^{(\alpha)}_{m,\nu}$ is the upper α point of the distribution of Studentized maximum modulus of m normal variates with the degree of freedom ν. (See Hochberg and Tamhane (1987).)

2.4.3 C-Procedure

C-procedure is also proposed by Dunnett (1980).

The test statistic is same as GH-procedure.

But, the critical value is

$$\xi_{ih} = \frac{Q^{(\alpha)}_{k,\nu_i}(S_i^2/n_i) + Q^{(\alpha)}_{k,\nu_h}(S_h^2/n_h)}{\sqrt{2}(S_i^2/n_i + S_h^2/n_h)}.$$

2.5 Steel-Dwass Procedure

In Chapter 4, we discuss the disadvantage of nonparametric procedures under non-homogeneous variances. Here we mention Steel-Dwass procedure as a representative of nonparametric procedures. *Steel-Dwass procedure* is proposed by Steel (1960) and Dwass (1960) independently (see Hochberg and Tamhane 1987), which is a nonparametric procedure not needing assumption of the type of distribution. Therefore, the error term ε_{ij} is only i.i.d. and may not be normal distributed. Needless to say, since they are identically distributed, they have an equal variance. Thus, it is a mistake to use in the non-homogeneity case. Nevertheless, it may be more robust for this assumption than parametric procedures.

The statistic of this procedure is given in the following.

$$|T_{ih}| = \frac{|R_{ih} - n_i(n_i + n_h + 1)/2| - 1/2}{\sqrt{n_i n_h V_{ih}/(n_i + n_h - 1)}},$$

where R_{ih} is the rank sum of the i-th group when it ranks with the h-th groups, and V_{ih} is the variance of the rank. Besides, when there is no tie in the rank, $V_{ih} = (n_i + n_h - 1)(n_i + n_h + 1)/12$.

And, the critical value is given as

$$\xi_{ih} = Q^{(\alpha)}_{k,\infty}/\sqrt{2}.$$

Therefore, if $|T_{ih}| \geq \xi_{ih}$ then we judge μ_i and μ_h are different as same as the previous section.

Chapter 3

Powers for Multiple Comparison Procedures

We have almost unified opinion for Type I error rates of multiple comparison procedures but do not so for powers of them. The purpose of this chapter is that we propose new powers additionally and ascertain which power is useful for judgement of the merits and demerits of multiple comparison procedures by comparison on Monte Carlo simulation.

3.1 Primary Comparisons of Powers

Most procedures in Section 2.3 are compared on powers before. (See Hochberg and Tamhane (1987).)

Einot and Gabriel (1975) compare procedures in Section 2.3 except for Holland-Copenhaver procedure in the case of the balanced one-way layout. Then, Einot and Gabriel use *P-subset power*. It is defined as the probability of rejecting hypothesis H_P for a given subset $P \subseteq K$ when H_P is false. (Besides, when $p = 2$, it is especially called *per-pair power*.)

Since many P-subset powers are considered, Einot and Gabriel use the average of them with same p and get the following result. "Against these stepwise multiple comparison procedures, the simultaneous test procedures, with somewhat smaller power (at most

6%), have the advantages."

The advantages of Tukey procedure that is mentioned by Einot and Gabriel are the following.

1. It is easy to extend the procedure for all contrasts.
2. The decision for a set P is independent of sample means of which index numbers are outside P.
3. The simultaneous confidence intervals corresponding to the procedure can be used.
4. The calculation for the procedure is easy.

On the other hand, Ramsey (1978) also compares the same procedures in the balanced case. Then, Ramsey uses *all-pairs power*, which is defined as the probability of rejecting for all pairs that have different population means.

And, the result is that "The superiority of Peritz's procedure based on F-statistics over Tukey's procedure is extremely high (Table 6c, $f = 2.7$, $.761 - .244 = .517$)."

Furthermore, Ramsey mentions that for the comparison of Q-statistics and F-statistics in stepdown procedures F-statistics is better than Q-statistics slightly. Now, we attend that this result do not contradict the fact that "Scheffé's procedure is worse than Tukey's procedure for pairwise comparisons", which mentions in Section 2.3.1. As shown on Table 4.1 of p. 104 in Hochberg and Tamhane (1987), Scheffé's procedure has less nominal level assigned to each null hypothesis for pairwise comparisons than Tukey's procedure, thus the performance of Scheffé's procedure is inferior. In stepdown procedures, however, the same nominal level α_p for the test of H_P is used for both F-statistics and Q-statistics. Hence, the overall result is influenced by the difference of performance of two statistics on the test for each null hypothesis. In testing of the overall null hypothesis on one-way layouts, David, Lachenbruch and Brandis (1972) show that Q-statistics is superior in the

case that parameter configurations are extremely separated (e.g. Configuration MAX in Section 3.3) and that F-statistics is superior in other many cases. Therefore, we can understand that F-statistics is profitable for the test of a null hypothesis on each step. (Although there may be the relation between steps in stepdown procedures, we cannot easily explain it.) The knowledge in this paragraph is needed in the later consideration.

Subsequently, Gabriel and Ramsey exchanged comments to discuss which procedure we should use, stepdown one or single-step one. Moreover, their main topic was the the validity of powers proposing respectively. At last, their opinion settled down to the conclusion: in multiple comparisons, it was difficult to select a power generally, but we should know features of procedures by using each power.

We think that the standpoint like this is reasonable, but the previous two powers are not enough to evaluate features of procedures from various points of view. Thus, we propose new powers in multiple comparisons. In this thesis, furthermore, we expand objects to compare by adding Holland-Copenhaver procedure and also study the behavior in the case of the unbalanced one-way layout.

3.2 Proposal of New Powers

In this thesis, we consider the following six powers.

- A : *all-pairs power.*

$$\Pr\left\{\sum_{(i,h)\in I_1}(1-D_{\{i,h\}})=0|\boldsymbol{\mu}\right\},\quad I_1=\{(i,h):1\le i<h\le k,\ \mu_i\neq\mu_h\}.$$

- B : the probability of rejecting all $H_{\{i,h\}}$'s with $|\mu_i-\mu_h|>1.5f$ [*restricted all-pairs power*], where f is defined in the next section.

$$\mathrm{P}_1\left\{\sum_{(i,h)\in I_2}(1-D_{\{i,h\}})=0|\boldsymbol{\mu}\right\},\quad I_2=\{(i,h):1\le i<h\le k,\ |\mu_i-\mu_h|>1.5f\}.$$

- C : per-pair power for a pair with the minimum difference [*minimum difference power*].

- D : per-pair power for a pair with the maximum difference [*maximum difference power*].

  Since both Powers C and D are per-pair powers, we get the following expression by defining (i, h) as an objective pair.

$$\Pr\{D_{\{i,h\}} = 1|\boldsymbol{\mu}\}.$$

- E : the mean rate of the number of pairs rejected among all pairs that have different population means [*mean rejective rate*].

$$\mathrm{E}\left\{\sum_{(i,h)\in I_1} D_{\{i,h\}}|\boldsymbol{\mu}\right\}/\#I_1.$$

- F : the weighted mean rate of the number of pairs rejected among all pairs that have different population means [*weighted mean rejective rate*].

$$\frac{\mathrm{E}\{\sum_{(i,h)\in I_1}|\mu_i-\mu_h|D_{\{i,h\}}|\boldsymbol{\mu}\}}{\#I_1\sum_{(i,h)\in I_1}|\mu_i-\mu_h|}.$$

In the previous powers, B, E and F are newly proposed powers. We propose Power B to detect all pairs that differ over some level and propose Powers E and F to know the mean number of rejected pairs. Besides, the setting of level $1.5f$ on Power B have no special reason, but it is enough to know a different feature against Power A that does not omit any small difference.

Moreover, there is *any-pair power* that is defined by the following expression similar to Type I FWE.

$$\Pr\left\{\sum_{(i,h)\in I_1} D_{\{i,h\}} > 0|\boldsymbol{\mu}\right\}.$$

This is proposed by Ramsey (1978). It means 'the probability of rejecting any pair that has different population means', thus it has the same result for Tukey's procedure and stepdown procedures with Q-statistics. (We note that the first rejecting of stepdown procedures with Q-statistics is same as Tukey's procedure.) As pointed out by Ramsey, therefore, this power has little behavior generally and is not suitable for comparison, thus it is not considered in this thesis.

3.3 Setting and Procedure for Monte Carlo Simulation

Using powers in the previous section, we study features of procedures in Section 2.3. The following is the setting of simulation and the procedure.

At first, set $\sigma^2 = 1$ and consider the following four types of configuration of population means μ_i's.

- *Means with equally spaced configuration* (EQ) : $\mu_i = (a + bi)f$.
- *Means with the minimum range* (MIN) :

  if k is even then

  $$\mu_1 = \cdots = \mu_{k/2} = -f,\ \mu_{k/2+1} = \cdots = \mu_k = f,$$

  and if k is odd then

  $$\mu_1 = \cdots = \mu_{(k+1)/2} = -[(k-1)/(k+1)]^{\frac{1}{2}} f,\ \mu_{(k+3)/2} = \cdots = \mu_k = [(k+1)/(k-1)]^{\frac{1}{2}} f.$$

- *Means with the maximum range* (MAX) :

  $$\mu_1 = -(k/2)^{\frac{1}{2}} f,\ \mu_2 = \cdots = \mu_{k-1} = 0,\ \mu_k = (k/2)^{\frac{1}{2}} f.$$

- *Means with square root configuration* (SQ) : $\mu_i = (\sqrt{i-1} - a)bf$.

All the configurations become $\sum \mu_i = 0$ and $\{\sum \mu_i^2/k\}^{1/2} = f$, where for Configurations EQ and SQ we define constants a and b as satisfying this equation. Besides, Configurations EQ, MIN and MAX are also used in Ramsey (1978).

Since under some configurations above many pairs with the minimum difference and/or many pairs with maximum difference appear, we should exactly redefine Powers C and D in the previous section. First, let the pair with the minimum difference on Power C be the pair $(k-1, k)$. But, when Configuration MIN and $k \geq 4$, it becomes a part of Type I FWE and is out of consideration. Second, let the pair with the maximum difference on Power D be the pair $(1, k)$. But, when Configuration MIN, it also becomes the minimum difference power. Besides, since both Powers C and D are per-pair powers, we use them only to add some features of procedures.

At second, we set k and n_i in the following.

$k = 4,$	$(n_1, \ldots, n_k) = (6,6,6,6)$	$D.F.\ \nu = 20$
4,	$(16,16,16,16)$	60
5,	$(6,6,6,6,6)$	25
4,	$(2,2,10,10)$	20
4,	$(4,4,4,12)$	20
4,	$(2,4,6,12)$	20
4,	$(10,10,2,2)$	20
4,	$(12,4,4,4)$	20
4,	$(12,6,4,2)$	20

For simplification, furthermore, we call *Group 1* as the group that has the pattern of $n_1 \leq n_2 \leq n_3 \leq n_4$ in the case of the unbalanced one-way layout and *Group 2* as the group that has the pattern of $n_1 \geq n_2 \geq n_3 \geq n_4$.

We simulate for each setting above with the repeat number 1000. This repeat number is same as Ramsey (1978). Besides, we decide the setting of f, which depends k, n_i and configurations of means, as powers are located between 0 and 1. In practice, we set f in every 0.1 or 0.05 and search the appropriate range by a primary simulation. Subsequently, the range depends the type of power, but we simulate in the maximum range for each configuration of means. In the case of Table 1 ($k = 4,\ n_1 = \cdots = n_k = 6$)

practical values of f are 0.1 to 3.1 on Configuration EQ, 0.1 to 1.7 on MIN, 0.1 to 1.9 on MAX and 0.1 to 5.0 on SQ. Large f values are due to Power A. In the case of Table 2 ($k = 4,\ n_1 = \cdots = n_k = 16$) the range of f is reduced to half as large as that of Table 1, and in the case of Table 3 ($k = 5,\ n_1 = \cdots = n_k = 6$) it expand to 1.2 times. Besides, in the unbalanced case, the range of f in Group 1 is almost same as that of Table 1, but that of Group 2 expands to 1.5 times.

We do not directly compare powers gotten by such a way as Ramsey (1978), but relatively compare them as Einot and Gabriel (1975). For further details, for each power among Powers A to F, we plot (x, y) = (the power of Tukey's procedure, the power of each procedure) for each f and fit a cubic curve, then we compare estimates of y at $x = 0.25,\ 0.50,\ 0.75$. (We select data that has the power between 0.03 and 0.97 for Tukey's procedure to get the fitting curve for each power.)

Besides, we note that Ramsey (1978) also simulates on the setting $k = 4$, $(n_1, \ldots, n_k) = (16, 16, 16, 16)$ and our result (Power A) is comparable with Ramsey's one.

Now, we mention the procedure of comparison. Since we think that it is appropriate as one of viewpoints for selection of procedures to consider the maximum difference of powers, we compare by using the maximum difference at $x = 0.25,\ 0.50,\ 0.75$. The value x corresponding to the maximum difference maybe exists nearby 0.5, thus this procedure is enough to compare. However, we notice other values whenever we find unstable cases where the sign of differences changes halfway.

Furthermore, since we are not interested in all pairs of procedures in comparing, we select pairs as follows.

- *Comparison 1*: Tukey-Welsch procedure with Q-statistics [TW(Q)] vs. Tukey's procedure [TU] (Tukey-Kramer procedure [TK]).
- *Comparison 2*: Peritz's procedure with Q-statistics [PE(Q)] vs. Tukey-Welsch procedure with Q-statistics.

- *Comparison 3*: Holland-Copenhaver procedure [HC] vs. procedures with Q-statistics [QS].

- *Comparison 4*: procedures with F-statistics [FS] vs. procedures with Q-statistics.

3.4 Consideration for Simulation Result

3.4.1 Case of Equal Sample Sizes

In this subsection, let $n = n_1 = \cdots = n_k$ since sample sizes are same.

Table 1 shows the result for the case of $k = 4$ and $n = 6$.

Before we compare procedures using the simulation result, the first remarkable point is the problem of precision. Raw results of the simulation have standard deviations being $(.5 \times .5/1000)^{1/2} = .016$ or less, but on relative powers to Tukey's procedure they are smaller than the value. We can certificate this conclusion by noting the maximum of standard deviations of the regression error as given in the last row on Table 1. The values for Powers E and F are smaller than 1% for any procedure. Otherwise, Powers A and B have values as same as simulation errors. We think it is the reason that the fitness of curve becomes worse as leaving from the diagonal. However, since we estimate population parameters for a given value of x, standard deviations of the estimators are further smaller than standard deviations on the table. (The degrees of reduction are at most 0.68 times at $x = 0.25$, at most 0.62 times at $x = 0.50$ and at most 0.71 times at $x = 0.75$.)

Furthermore, as a practical error read on tables there are reversals of powers. That is to say, it is an error that reversals of powers occur by the regression whereas a procedure is theoretically more powerful than another procedure (in fact, the raw result of the simulation is so). The maximum error is 0.9% reversal of Holland-Copenhaver procedure against Peritz's procedure with Q-statistics on Power C of SQ on Table 1.

From the previous thought, it is reasonable that we use 1% as the unit for comparisons

on Table 1, where Powers E and F are more stable and changeless than others thus we note the behavior till 0.5% if it is necessary. Besides, multiple correlation coefficients of the regression are at least 0.9982 among all cases on Table 1. (In addition, values for powers except A are at least 0.9990.)

Now, let us compare procedures on Table 1.

- Comparison 1: the case of TW(Q) vs. TU.

  The differences on Powers A, E and F are stable for all parameter configurations, that is, they are almost same values for all configurations and TW(Q) is more powerful than TU.

  A:17 - 19%, E:6 - 9%, F:6 - 8%.

  On the other hand, Power B has extreme results that are 17% difference for Configuration MIN but only 2% difference for Configuration MAX. It is the reason that TW(Q) has the same power as TU for detecting the first difference and B;MAX (which denotes Power B for Configuration MAX) is about same probability as D;MAX. (Slightly difference between B;MAX and D;MAX is due to rejecting the maximum difference pair secondly.)

  This result is also useful for us to understand features of Power B. It is practical and interesting that Power B is made for rejecting differences being more than a certain limit. As is shown in this case, however, the values of Power B depend on parameter configurations, thus we take care to use it.

  Conversely, on Power A that is important to less differences, the advantage of TW(Q) is very large. This relation is backed by comparing Power C with D.

- Comparison 2: the case of PE(Q) vs. TW(Q).

The behavior of powers is similar to Comparison 1. PE(Q) is more powerful than TW(Q) and their differences is the following.

A:8 - 17%, E:2 - 4%, F:1 - 2%; B:0 - 8%.

The reason of the extreme variation of Power B is also similar to Comparison 1. PE(Q) has large advantage on Power A, but it has slightly advantage on Powers E and F. However, we must be concerned about the difference of features between Powers A-B and E-F. Although Power E has at most 4% difference, it means to reject pairs 4% more. If $k = 4$ and population means are all different, the number of pairs having difference is 6 and the 4% difference means to reject pairs 0.24 more. In other words, it can reject one pair more a fourth times.

- Comparison 3: the case of HC vs. QS.

  Powers E and F are stable for all parameter configurations. On the basis of them, HC has the performance between TU and TW(Q). (According to Power E, it is 5 - 8% better than TU and 3 - 4% worse than TW(Q).) However, we should note reverse phenomena that HC exceeds TW(Q) at higher level on Power E.

  On the other hand, according to Power A, HC is 14 - 16% better than TW(Q) and almost same as PE(Q) for Configurations EQ and SQ. But, for Configurations MIN and MAX it is equal to or less than TW(Q). Moreover, on B;MAX it is equal to or less than TU.

  Furthermore, thinking with Powers C and D (where at higher level on Power C reverse phenomena exceeding TW(Q) arise), we find that Holland-Copenhaver procedure is better when the number of pairs to reject is large. (This is not quite same as better at rejecting small differences. As for Configurations MIN and MAX Holland-Copenhaver procedure is worse since the number of pairs to reject is small.)

The reason of such result is that Holland-Copenhaver procedure is same as Šidák's procedure, which is worse than Tukey's procedure, for rejecting of the first pair. We study this situation in detail in the following.

In comparing Holland-Copenhaver procedure and stepdown procedures with Q-statistics, we can find the differences by comparing critical values because we use same statistics for these procedures. In the following, we study for $\alpha = 0.05$ to give an example with definite values, but results for other α are similar. Since $t_m = 6, 3, 3, 3, 2, 1$ for $m = 1, 2, \ldots, 6$ on Table 1 in Holland and Copenhaver (1987), critical values of Holland-Copenhaver procedure are

$$t(20, C(t_m)/2) = 2.918,\ 2.605,\ 2.417,\ 2.086, \quad t_m = 6, 3, 2, 1,$$

where $t(\nu, \tau)$ is the upper τ point of t-distribution with the degree of freedom ν. On the other hand, critical values of Tukey-Welsch procedure with Q-statistics are

$$Q_{p,20}^{(\alpha_p)}/\sqrt{2} = 2.799,\ 2.530,\ 2.417, \quad p = 4, 3, 2.$$

By comparing these values, we find that Holland-Copenhaver procedure has the disadvantage for rejecting of the first pair and the advantage for rejecting of small pairs (especially, rejecting of the last pair).

In addition, critical values of Peritz's procedure by Newman-Keuls specification are

$$Q_{p,20}^{(\alpha)}/\sqrt{2} = 2.799,\ 2.530,\ 2.086, \quad p = 4, 3, 2.$$

To compare with Holland-Copenhaver procedure, we get the following table by ordering critical values from the largest difference.

HC	2.918	2.605	2.605	2.605	2.417	2.086
P	2.799	2.530	2.530 or 2.417 or 2.086	2.530 or 2.417 or 2.086	2.417 or 2.086	2.086
	(4)	(3)	(3 or 2)	(3 or 2)	(2)	(2)

We note that Peritz's procedure is not sequentially rejective type thus critical values is divided into some patterns. (The number in parentheses on Peritz's procedure means the size of the null hypothesis.) As shown on this table, Peritz's procedure is always more powerful than Holland-Copenhaver procedure.

- Comparison 4: the case of FS vs. QS.

  There is little difference for Tukey-Welsch procedure. Tukey-Welsch procedure with F-statistics [TW(F)] is 2% advantageous on D,E,F;MIN, and TW(Q) is 1% advantageous on B;EQ.

  Peritz's procedure with F-statistics [PE(F)] is 2 - 4% advantageous for Configuration MIN and on A;MAX. But, there is no difference in other cases.

  Consequently, FS is more advantageous than QS for Configuration MIN. Conversely, QS is suitable for rejecting large differences and is advantageous for Powers B and D. As a whole, owing to the extent of critical regions, FS is a little advantageous.

Next, we get the result for $k = 4$ and $n = 16$ by Table 2, and the result for $k = 5$ and $n = 6$ by Table 3.

On comparing Table 1 with Table 2, we find that powers are not very different as a whole. The maximum change is about 2%, and generally changes are 1% or less. According to the discussion above for the precision, this level of change is within error. However, since there are same trends with 1% change or less through configurations on some powers, we take notice of them in the following. Besides, Table 3 has larger change than Table 2, but there is at most only 1% change on Powers E and F.

- Comparisons 1 and 2: the cases of TW(Q) vs. TU and PE(Q) vs. TW(Q).

  On Table 2, there is the trend to approach on Powers E and F, but the changes are 1% or less.

On Table 3, the difference is larger on Power A (3 - 5%) and is smaller on Powers E and F (1.5% or less). Moreover, we find phenomena that is the vanishing of differences between PE(Q) and TW(Q) for Configurations MIN and MAX and especially on Power B for any configuration. The phenomena on Power A is also observed in Ramsey (1978).

- Comparison: 3 the case of HC vs. QS.

  On Table 2, the difference between HC and TU does not generally change so much, and the difference between HC and TW(Q) is trend to reduce, but the change is at most 1%.

  On Table 3, the difference between HC and PE(Q) is clearly trend to reduce. However, the change is inferior to that of the difference between PE(Q) and TW(Q) in Comparison 2.

- Comparison: 4 the case of FS vs. QS.

  On Table 2, the difference is trend to reduce on D,E,F;MIN, but the changes is less than 1%.

  On Table 3, situations where differences between TW(Q) and TW(F) change more advantageous to F-statistics than that on Table 1 are on A,B;MIN, A,C,E,F; MAX and B;SQ, and the changes are 1 - 2%. On the other hand, situations where differences change more advantageous to Q-statistics are on B,D,F;EQ, D;MIN, B;MAX and D;SQ, and the changes are 1 - 2%.

  Situations where differences between PE(Q) and PE(F) change more advantageous to F-statistics than that on Table 1 are few and the changes are less than 1%. On the other hand, situations where differences change more advantageous to Q-statistics have 2% change or less.

  Consequently, procedures with Q-statistics become better.

3.4.2 Case of Unequal Sample Sizes

In this subsection, we will study the case of unequal sample sizes and compare the result with that in the previous subsection.

In the case of $k = 4$, let the setting of (n_1, n_2, n_3, n_4) be $(2, 2, 10, 10), (4, 4, 4, 12)$ and $(2, 4, 6, 12)$ as Group 1 that have the pattern of $n_1 \leq n_2 \leq n_3 \leq n_4$, and be $(10, 10, 2, 2)$, $(12, 4, 4, 4)$ and $(12, 6, 4, 2)$ as Group 2 that have the pattern of $n_1 \geq n_2 \geq n_3 \geq n_4$. We show results of Group 1 on Tables 4 to 6, and show a result of Group 2, i.e. $(10, 10, 2, 2)$, on Table 7. (We easily find that results in same group are similar as shown on tables of Group 1. Moreover, sample sizes used in Group 2 rearrange those used in Group 1 in reverse order, thus each result of Group 2 does not differ from the corresponding result of Group 1 except for results on Power C or for Configuration SQ. Therefore, we show only one table for Group 2.)

Since there are many patterns of unequal sample sizes, we can never get the entire behavior from studying these six patterns. Nevertheless, we believe that we can get certain success because no one study comparisons for any case of unequal sample sizes.

In addition, we note that the difficulty of interpretation for results of the case of unequal sample sizes (i.e. the clustering cannot be made in order of sample means), which is mentioned by Yoshida (1989), is also shown on sequentially rejective type procedures. Thus there is no superiority for them to other procedures in this respect.

First, a point to emphasize as results in the case of unequal sample sizes is that there is little change of the degree of the relative advantage of procedures as is shown in comparisons between Table 1 and Tables 4 to 7. We considered that there was certain large change for the case of unequal sample sizes before the research in this thesis, but the result is not so. The reason is the following. Our research adopt the relative comparison based on Tukey's procedure or Tukey-Kramer procedure. Therefore, even if there is any effect for the case of unequal sample sizes, the counteraction for the change of powers

can be yielded to some extent. (Surely our purpose is the comparison of procedures, so that it is better to remove the unnecessary effect. In this point, we believe the validity of the comparison method in this thesis.)

Next, we mention specified features in the case of unequal sample sizes that can be gotten by more sensitive study.

- Comparisons 1 and 2: the cases of TW(Q) vs. TK and PE(Q) vs. TW(Q).

  The differences on Powers E and F change 1.3% or less from Table 1. And, the differences on Power A also changes 2.9% or less.

  But, Power C shows large change. The change is different between Groups 1 and 2, i.e. the difference of Group 1 decreases from Table 1 (5.8% or less) and that of Group 2 increases (5.9% or less). On the other hand, the change of Power D is negligible. In other word, if samples have unequal sizes, the power for small differences between population means is influenced by the unbalance. This phenomenon is interpreted in the following. As we have shown, the first rejecting of stepwise procedures is same as that of the corresponding single step procedure. Thus, the advantage of stepwise procedures is yielded when many pairs are rejected. Since small differences between population means are hardly rejected, they are easy to influence the result of stepwise procedures. Since $n_1 \leq n_2 \leq n_3 \leq n_4$ in Group 1, each treatment corresponding to small differences between population means has a large sample size, thus even an inferior procedure is easy to reject them. On the other hand, since $n_1 \geq n_2 \geq n_3 \geq n_4$ in Group 2, each treatment corresponding to small differences has a small sample size, thus the original superiority of procedures influence directly rejecting small differences between population means. Consequently, the result is shown in the difference of powers.

- Comparison 3: the case of HC vs. QS.

  Powers A, E and F have small changes from Table 1. On Power C for Group 1, HC is relatively less powerful than TK (9.3% reduction or less) and TW(Q) (the tide turns in a part of situations where HC is advantageous on Table 1). On the other hand, HC is relatively more powerful than TK in Group 2 (7.3% increase or less). Nevertheless, since the difference of HC with PE(Q) hardly change, PE(Q) may be similarly influenced by the unbalance. This result supports the previous consideration that the unbalance of sample sizes influence the rejecting of small differences.

- Comparison 4: the case of FS vs. QS.

  This comparison is most remarkable among Comparisons 1 to 4. We know that Tukey-Kramer procedure and its stepdown modification are theoretically conservative (c.f. Hochberg and Tamhane 1987, Yoshida 1989). However, it has never been studied how effect exists about powers, especially powers for stepdown procedures, in this case.

  As shown in tables, the maximum changes of difference on Power A, E and F from Table 1 are 0.6%, 1.7% and 2.9%, respectively for TW(Q), and 4.8%, 1.8% and 3.1%, respectively for PE(Q). Almost changes are advantageous to F-statistics, but some especially advantageous differences on Table 1 are tend to decrease (e.g. on D,E,F;MIN). For this reason, maximums of difference almost decrease, thus this level of unbalance is not serious for stepdown procedures. Besides, on comparisons between Groups 1 and 2 for Configuration SQ, FS has some advantage.

In addition, the procedure based on p-values of Q-statistics using the program of Yoshida (1988) is better than Tukey-Kramer procedure. Nevertheless, since the amount time of the calculation is enormous, we exclude the procedure from objects for comparison

in this thesis.

3.5 Discussion

3.5.1 Selection of Powers

We discuss known powers and the necessary of new powers in Section 3.1, but we rediscuss them in detail.

First of all, it is due to the following consideration why we propose several new powers (evaluation criterions) and try to compare several procedures based on them in this thesis: "Since the concept of 'power' on multiple comparisons does not get the consensus among statisticians nor practicians, it is too hasty that we point out 'the difficulty for selection of powers on multiple comparisons', only considering all-pairs power and per-pair power (which is mentioned in Section 3.1). Furthermore, even if 'we consider features of procedures according to each power', the kind of powers is not enough. Therefore, we should evaluate each procedure from various viewpoints."

Next, we consider features of each power. Since Power A rejects all differences, it is influenced by rejecting small differences. As shown in comparison between Powers C and D, the effect of per-pair power depends on the position of the difference of population means for the pair in order of size among differences for all pairs. Conversely, Powers E and F have a feature of few dependence on parameter configurations. (We note it is not that they are not influenced by parameter configurations at all. In Particular, the influence is shown on comparison of procedures with F-statistics and that with Q-statistics.) This feature is suitable for the comparison of procedures. The reason is the following. In analyzing practical data we do not know the real parameter configuration, thus a power that hardly depends on the parameter configuration is easy to use for the total evaluation of the superiority of procedure. (We do not conclude which of Powers E and F is better because of few difference between them. However, few difference

implies few variation for other various weights.) Although Power B depends on parameter configurations, it has meaning enough to use if we are interested in some level of large difference definitely. (Here, 'some level of large difference' means 'the difference' acquired by analysts in the specific technology, e.g. 'the biological significance'.)

Now. if we compared procedures using only primary powers A, C and D, what would be happened? The especially remarkable procedure is Holland-Copenhaver procedure. This procedure is suitable for rejecting small difference as considered in Section 3.4.1. Hence, on Power A it is most powerful next to Peritz's procedure and we would think it is not so bad procedure because of easily using. (Since the result of Power D is bad, we might note it is not suitable for rejecting of large differences.) However, this is not enough evaluation. According to adding new Powers B, E and F, we can reduce that Holland-Copenhaver procedure is not generally so good as the consideration in Section 3.4.

In general, on comparison among Peritz's procedure, Tukey-Welsch procedure and Tukey's procedure with same statistics, the rank of advantage is invariant for any power owing to the construction of procedures, thus there is no serious problem if we only use primary powers. In the case that different results are observed for each power as the comparison between Holland-Copenhaver procedure and Tukey-Welsch procedure, only using primary powers is not so enough that we need to compare procedures while understanding features of each power. Moreover, in the previous case of 'no serious problem', we should note that certain procedure excessively emphasizes if we only use Power A.

In conclusion, we center Powers E and F as the global power, and additionally use Powers A and C as the power to judge the effect for small differences and Powers B and D as the power to judge the effect for large differences.

Nevertheless, they may not be enough. Especially, on another model, we may need

other new powers from the point of view on the model.

3.5.2 Selection of Procedures

The final purpose of this chapter is which procedure is better to use. Before concluding it from the point of view considered in Section 3.4, we consider the advantage of Tukey's procedure mentioned by Einot and Gabriel (1975) (see Section 3.1).

First, about "It is easy to extend the procedure for all contrasts", we will exclude it. On setting in Einot and Gabriel (1975), Ramsey (1978) and this thesis, though only pairwise comparisons are considered, Einot and Gabriel consider that it is merit because we may extend to contrasts later. If we are really interested in general contrasts, we need to study again adding Scheffé's procedure.

Second, about "The decision for a set P is independent for sample means of which index numbers are outside P", we will consider it in detail. Now, for simplicity, we consider that pairs with 3 groups ($k = 3$) are compared and there are two configurations of parameters: "$\mu_1 = 0$, $\mu_2 = 1.0$ and $\mu_3 = 1.2$" and "$\mu_1 = 0$, $\mu_2 = 1.0$ and $\mu_3 = 3.0$". Here, let $P = \{1, 2\}$. While on stepwise procedures H_P for the former configuration is hard to be significant, on Tukey's procedure same results for H_P are given for both configurations. For this reason, though Tukey-Welsch procedure has better power than Tukey's procedure for any configuration, they assert it is unnatural that on stepwise procedures we have the possibility of getting different results for H_P on account of the value of the third population mean. However, is it sure? The situation to practice multiple comparisons is that to reveal the global result for a family of null hypotheses, thus it is rather natural that results are different if parameter configurations are totally different. Therefore, it has no meaning to discuss the difference of result for each H_P. Conversely, if we should consider about each (or important) H_P, we need to investigate whether it is really valid construction of null hypotheses to adopt multiple comparisons. (For this point, see Tsubaki (1989).) Consequently, this feature of Tukey's procedure implies that

the procedure is too conservative not to be influenced by parameter configurations, thus we do not think it is advantage.

Third, about "The simultaneous confidence intervals corresponding to the test can be used", it is surely advantage of Tukey's procedure. On simultaneous confidence intervals, it needs to reveal its range for even rejected pairs, hence we cannot assign the significance level to only not rejected pairs as stepwise procedures. Thus, simultaneous confidence intervals on stepwise procedures cannot be obtained. Therefore, if we need simultaneous confidence intervals for the difference of population means, we must construct them by Tukey's procedure. However, Yoshimura (1989) mentions that "What meaning do simultaneous confidence intervals for the difference of population means have? If we need the interval estimation, it is reasonable that we construct simultaneous confidence intervals for each population mean apart from the test", thus he has doubt about the practical necessity of simultaneous confidence intervals for all differences of population means. We do not have clear opinion about the necessity of simultaneous confidence intervals and the effect of their result for decision making. Nevertheless, we are against to adhere to Tukey's procedure on account of the use of simultaneous confidence intervals.

Finally, about "The calculation for the procedure is easy", we think it is important because how to recommend procedures is different between the situation where a computer package is available and the situation where it is not so. In conclusion, we will recommend procedures from the point of view on 'the power advantage' and 'the easiness of calculation'. The best procedure for the global power is Peritz's procedure with F-statistics. But, the difference to Peritz's procedure with Q-statistics is small, and reverse phenomena are observed for some parameter configurations as shown in Section 3.4. (For Tukey-Welsch procedure the difference between F- and Q-statistics is less than them.) Moreover, the difference between Peritz's procedure and Tukey-Welsch procedure are also small except for Power A, and the difference between Tukey-Welsch procedure and

Tukey's procedure are generally large. On the other hand, from the point of view on the calculation it is hard to calculate on F-statistics for not only Peritz's procedure but also Tukey-Welsch procedure. Against that, it is not hard to calculate on Q-statistics. If the number of groups k is small, it is easy even for Peritz's procedure. Moreover, for Tukey-Welsch procedure, even if k is moderate, it is easy as almost same as Tukey's procedure. In addition to this, if sample sizes are same, there is the simplified procedure and it is easier than the original procedure. (See Hochberg and Tamhane (1987).) In summary, we will mention the following chart of recommendation of procedures. (Besides, we need complete tables except for Tukey's procedure.)

$$\text{Peritz}(F) \approx \text{Peritz}(Q) > \text{Tukey-Welsch}(Q) \gg \text{Tukey}$$

In addition, we do not recommend Holland-Copenhaver procedure because of the inferiority to Tukey-Welsch procedure in many cases.

Chapter 4

Multiple Comparisons in the Unequal Variance Case

4.1 General Remarks

If we treat one-way layout data for multiple comparisons, we are often confronted with the case assuming the normality but not assuming the homogeneity of variances. In the case of two samples, it is called Behrens-Fisher problem and does not has a solution to keep the significant level strictly. Hence, many methods that are approximately, asymptotically valid are proposed, e.g. Welch's method. In the multi-sample problem, therefore, a strictly method does not exist either and asymptotically procedures are proposed. However, the performance of them on the small sample does not study in detail. The purpose of this chapter is the following. First, we study known procedures using Monte Carlo simulation and improve them to conquer their disadvantage. Second, we improve the homogeneity test of variances as a preliminary test to keep the global significant level.

If we assume the homogeneity of variances on one-way layouts, many procedures are known, including Tukey-Kramer procedure (see Section 2.3). On the other hand, Hochberg and Tamhane (1987) show some procedures not assuming the homogeneity of variances and recommend T3- and C-procedures (see Section 2.4) that are used properly

with the sample size. However, they do not give a definitely criterion to use properly.

Matsuda (1991) roughly shows the performance of known procedures for the non-homogeneous case using Monte Carlo simulation. It mentions an outline to use properly the previous two procedures, but it is not enough as a criterion because the number of configurations of sample size used in the simulation is few.

4.2 Proposal of New Multiple Comparison Procedures

For the same statistics as GH-procedure in Section 2.4, we will propose two procedures with a modified critical value.

4.2.1 GHC-Procedure

As shown in Matsuda (1991), we know that T3- and C-procedures have the merits and demerits for each other and GH-procedure is liberal for some configurations of sample size. We consider that it is possible to improve by combining C- and GH-procedures, because T3-procedure is asymptotically so inferior to C- and GH-procedures that a combination with asymptotically adaptation is only that of the two procedures.

Since the two procedures use the same test statistics, they can be improved according to only combining critical values. *GHC-procedure* proposed here uses the average of critical values of the two procedures as a new critical value. That is to say, it is the procedure using critical values:

$$\xi_{ih} = \frac{1}{2}\left(\frac{Q^{(\alpha)}_{k,\nu_i}(S_i^2/n_i) + Q^{(\alpha)}_{k,\nu_h}(S_h^2/n_h)}{\sqrt{2}(S_i^2/n_i + S_h^2/n_h)} + Q^{(\alpha)}_{k,\nu_{ih}}/\sqrt{2}\right).$$

4.2.2 GHC2-Procedure

On Monte Carlo simulation, we show that GHC-procedure is slightly inferior to T3-procedure in some cases of the small sample. In spite of that, we find that GHC-procedure is sometimes liberal in broader setting of sample sizes. GHC2-procedure proposed in this

subsection is a modified version of GHC-procedure to be more suitable for the case of unequal variances than other procedures.

Although we can consider a simple improvement that we increase the weight of C-procedure, it strongly reduces the real significant level and the power in the case of balanced sample sizes and homogeneous variances. Therefore, we will make the weight of the critical value dynamic to adapt each case.

First, we study the real Type I FWE's of C-, GH- and GHC-procedures for some situations. Table 8 is a part of the result. (The nominal significant level α is 0.05.) We find that the practical Type I FWE of GHC-procedure are similar to the harmonic mean of practical Type I FWE's of C- and GH-procedures. Hence, we can get the optimum weight to harmonize the real significant level to the nominal significant level in each case. The value a on Table 8 is the optimum weight ratio of C-procedure to GH-procedure.

Next, we construct a predictor of a. It is difficult to design a simple predictor for various sample sizes and various variances. We will make a predictor using the linear regression based on some simple indices. Firstly, we select some indices largely influencing the variation of a. (For example, the index of unbalance of sample sizes and the index of unbalance of ratios of variance to sample size.) Secondly, We study the performance of all indices and their combinations using the multiple linear regression with 400 values of a such as Table 8. Finally, we find the following formula with a good performance.

$$a = 5.12x,$$

where

$$x = \sum_{i=1}^{k} \left(\frac{n_i}{\overline{n}} - 1\right)^2 \sqrt{\frac{1}{k}\sum_{i=1}^{k} \left(\frac{\sigma_i^2}{n_i \overline{\sigma}^2} - \tau\right)^2},$$

$$\overline{n} = \frac{1}{k}\sum_{i=1}^{k} n_i, \quad \tau = \frac{1}{k}\sum_{i=1}^{k} \frac{\sigma_i^2}{n_i \overline{\sigma}^2}, \quad \overline{\sigma}^2 = \frac{\sum_{i=1}^{k} \nu_i \sigma_i^2}{\sum_{i=1}^{k} \nu_i}.$$

Figure 1 shows data and its regression. The dotted line denotes the regression line.

GHC2-procedure proposed here uses the following critical values based on the result.

$$\xi_{ih} = \frac{1}{\sqrt{2}(\hat{a}+1)}\left(Q^{(\alpha)}_{k,\nu_{ih}} + \hat{a}\frac{Q^{(\alpha)}_{k,\nu_i}(S_i^2/n_i) + Q^{(\alpha)}_{k,\nu_h}(S_h^2/n_h)}{S_i^2/n_i + S_h^2/n_h}\right),$$

where

$$\hat{a} = 5\hat{x} + 0.6$$

and $\hat{x}$ denotes the value of x using the sample standard deviation in place of σ_i. Besides, the constant term that is about three times of the standard deviation of residuals is added to control the significant level certainly, and the significant figure of coefficients is a single figure to decrease the degree of dependence on data. The solid line on Figure 1 denotes the line using in GHC2-procedure.

As shown later, GHC2-procedure has a good performance in any situation of sample sizes and variances: it has almost the same significant level as T3-procedure in the case of the small sample and control the significant level for very unbalanced cases.

4.3 Primary Comparisons among Procedures

The similar study given by Dunnett (1980) has comparison results among GH-, C- and T3-procedures in the following.

- GH-procedure is liberal in some cases with small and unequal sample sizes.
- C-procedure is preferable for the large or moderately large sample.

  On the other hand, T3-procedure is preferable for the small sample.

However, it has very few repeated number for the simulation and do not find a clear criterion to use C- and T3-procedures properly. Matsuda (1991) studies the criterion in further detail, but it gets only the range of the critical value to use properly.

4.4 Procedure of Monte Carlo Simulation

The procedure of Monte Carlo simulation in this chapter is the following:

1. Set the number of treatments k, sample sizes $(n_1, n_2, \ldots, n_k)$, means $(\mu_1, \mu_2, \ldots, \mu_k)$ and variances of the error term $(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2)$.

2. Construct data such as Section 2.1 using normal random numbers.
   (See Appendix.)

3. Calculate values of the test statistics and the critical values for each procedure. (The nominal level α is 0.05.) Besides, judge whether variances of data is homogeneous if a preliminary test is performed.

4. Test for each pair of treatments and investigate a Type I error (and rejecting for an alternative hypothesis).

5. Repeat 10000 times from Step 2 to Step 4, and acquire the rejective rate for each pair and the Type I FWE (and powers for the alternative hypothesis). Moreover, in order to certify the performance of the simulation, also gain the standard deviation of them for every 1000 repeats.

Standard deviations of the rejective rate and so on for every 1000 repeats in this simulation show valid values in any case, hence it suggests that the procedure of the simulation itself has no problem.

In this thesis, the setting of parameters in Step 1 are that $k = 3, 4, 5$, $n_i = 1, \ldots, 20$ with the balanced case and some unbalanced cases until the number of missing values is two, $\sigma_i^2 = 1, 2, 3$ with the homogeneous variance case and some unequal variance cases until the maximum ratio of variances is three times, and μ_i's are the overall null hypothesis and some alternatives. Furthermore, in order to consider data not designed

previously, we also use data in which the maximum ratio of sample sizes is three times, or $k = 6, 7$, or the maximum ratio of variances is 10 times. In the following, we mainly study the case of $k = 3$.

4.5 Consideration for Results on the Overall Null Hypothesis

In order to show features of each procedure, we must first observe the behavior of the Type I FWE on the overall null hypotheses that has equal means for all treatments: $\mu_1 = \mu_2 = \cdots = \mu_k$, because we cannot use a procedure that has the high power but is liberal.

Figures 2 to 15 show some results of the simulation on the overall null hypotheses. These figures have dots of observed Type I FWE's and fitted curves to the dots. We choose the following fitting curve function through trial and error:

$$a + \frac{b}{n} + \frac{c}{n^2},$$

where n is the sample size of the first treatment and coefficients a, b and c are gotten by the least square method. Since this fitting is good except for the neighborhood of the smallest value of n on any figure, we judge that it is useful to find the global performance.

Besides, the notation PT on the figures denotes the result of the procedure with a new preliminary test, which is explained in the section below.

4.5.1 Case of Homogeneous Variances

In this case, GHC2-procedure is almost conservative as is shown on Figures 2 to 5. GH- and T3-procedures are also conservative. GHC2-procedure controls the Type I FWE as well as T3-procedure. Therefore, we recommend GHC2-procedure to use alone.

Moreover, Matsuda (1991) predicts that C-procedure is better than T3-procedure at more than sample size 20, but now we find that the critical value is larger than it.

Although the value is out of the range studied, it is about $n = 40$ if regression curves fit well. Besides, the effect by the number of treatments is negligible.

Furthermore, although we know that GH-procedure is almost liberal, as shown by regression curves, it keeps the Type I FWE on Figure 2 and has only 5.2, 5.7 and 5.6% the Type I FWE on Figures 3, 4 and 5, respectively. However, the degree of not keeping the Type I FWE is more remarkable as the number of treatment k increase.

We can get another effect in the asymptotic situation by regression curves owing to practicing the fine simulation. For example, the following table shows regression curves on Figure 2: sample sizes (n, n, n) and variances $(1, 1, 1)$.

Procedures	Regression curves
T-K	$0.050 + 0.010/n - 0.025/n^2$
GH	$0.050 + 0.010/n - 0.112/n^2$
T3	$0.046 - 0.019/n - 0.054/n^2$
C	$0.049 - 0.188/n + 0.210/n^2$
GHC	$0.052 - 0.125/n + 0.089/n^2$
GHC2	$0.052 - 0.105/n + 0.051/n^2$

The coefficient a corresponds to the asymptotic Type I FWE and values of a on the table are natural. On this fitting, we can confirm that T3-procedure is asymptotically inferior.

4.5.2 Case of Unequal Variances

Figures 6 to 8 are balanced cases and Figures 9 to 15 are unbalanced cases. GHC2-procedure is conservative on all figures except for $n = 4$ on Figures 14 and 15, which have very non-homogeneous variances. But GH- and T3-procedures are liberal in many situations with small sample sizes.

Furthermore, Tukey-Kramer procedure is liberal on any figure. The degree increases as the non-homogeneity becomes heavy. In the past, there was the consideration that on this case differing from the assumption we might use Tukey-Kramer procedure when sample sizes were balanced, because it was robust. However, we find that the result of

the procedure exceeds the nominal significant level with 1.0 - 1.7% on such situation (Figures 6 to 8). Moreover, as pointed out in Matsuda (1991), we should note that rejecting for each pair of treatments is handled differently. For example, in the case of sample sizes (6,6,6) and variances (1,1,3), the result of rejecting for each pair is gotten as the following table.

Types	Practical rejective rates
Pair (1,2)	0.006
Pair (1,3)	0.034
Pair (2,3)	0.034
Type I FWE	0.057

Seeing this table, we find that a pair of treatments that have both small variances hardly rejects and pairs of treatments that have at least a large variance easily reject. That is to say, it means that pairs to treat equally do not treat so. Therefore, if we find the non-homogeneity, we must not use Tukey-Kramer procedure even if data is balanced.

4.6 Behavior on Powers

We compare procedures using a power explained in the previous chapter. We use 'mean rejective rate', which has the good performance for the global judgement of the goodness of procedure in the previous chapter.

Figures 16 and 17 is some results of powers for means with equally spaced configuration (EQ) in the case of homogeneous variances, where Tukey-Kramer procedure is suited. (f denotes the standard deviation of the mean configuration.) If the sample size is small, the difference between Tukey-Kramer procedure and GHC2-procedure is large, that is, about 18% on Figure 16, otherwise it is only about 5% on Figure 17.

On the other hand, we show the non-homogeneous case on Figures 18 and 19, but it is no meaning to compare two procedures because the power of Tukey-Kramer procedure depends on the variation of the real significant level of it.

4.7 Comparison with Steel-Dwass Procedure

If we judge the non-homogeneity on a preliminary test, we commonly use a nonparametric procedure. In this section, we consider the problem.

We treat Steel-Dwass procedure as a representative procedure among nonparametric procedures.

As comparisons to be especially careful, we show results of Steel-Dwass procedure on Figures 9, 13, 15 and 17.

First, as is shown on Figure 9 where the maximum variance rate is three, Steel-Dwass procedure keeps the significant level under the overall null hypothesis. However, it has very small practical significant levels in the case of the small sample and rejects nothing in some cases. Although this point may improve by means of the calculation of the discrete probability in spite of the asymptotic result, the real significant level becomes 0 in the case of very small sample sizes.

Next, on Figures 13 and 15, Steel-Dwass procedure is liberal in very unbalanced cases, which is similar to Tukey-Kramer procedure. Although the degree of that increase as increasing of the non-homogeneity, it is robust when it is compared with Tukey-Kramer procedure.

Finally, we consider the behavior of the power on Figure 17. We find that Steel-Dwass procedure is less powerful than GHC2-procedure. If the distribution of data is normal or the mean of data is closely normal distributed, GHC2-procedure is advantageous. Besides, we add the information that Steel-Dwass procedure rejects nothing on Figure 16.

In global conclusion, Steel-Dwass procedure is a procedure that rejects nothing for the case of very small sample sizes and is liberal for the large sample, though it is robuster than Tukey-Kramer procedure. Furthermore, whenever sample sizes are in the range to be able to use, its power is less than that for GHC2-procedure under the normality.

Therefore, we do not recommend it as a multiple comparison procedure for data with the normality.

4.8 Structure of Preliminary Tests

4.8.1 General View

It is common that a procedure under the non-homogeneity and a procedure under the homogeneity are properly used after a preliminary test. On our primary simulation, we study the global significant level with the ordinary preliminary test. However, it has not been satisfactory on the following point. Since we should sufficiently note the setting of the significant level of the preliminary test on the two-sample problem (see Nagata 1992), we contrive to raise the significant level of the preliminary test for multiple comparisons up to 50% in the next subsection. Nevertheless, we observe the violation of the significant level in the case of small unbalanced sample sizes owing to the sensitive reaction to the non-homogeneity.

The purpose of this section is that we propose a new type of preliminary test because of the consideration that it does not become drastic improvement to control the significant level of the ordinary preliminary test. And, we study the behavior of the global significant level of the new preliminary test system by simulation.

4.8.2 Performance of the Ordinary Preliminary Test

In practical problem, we do not previously know whether variances are homogeneous. Thus, we need the basis how we properly use Tukey-Kramer procedure and GHC2-procedure recommended in the section above. In the past, we properly use procedures on the basis of the ordinary preliminary test for the homogeneity. Now, in this subsection we will observe the behavior of the global Type I FWE in the case of the proper use with Bartlett's test as the ordinary preliminary test.

Besides, Bartlett's test is the test using statistics

$$B = \frac{\nu \log \overline{S}^2 - \sum_{i=1}^{k} (\nu_i \log S_i^2)}{1 + \frac{\sum_{i=1}^{k} \frac{1}{\nu_i} - \frac{1}{\nu}}{3k-3}}$$

and critical values $\chi^{2(\alpha)}_{k-1}$, where $\overline{S}^2$, S_i^2, ν and ν_i are defined in Chapter 2, and $\chi^{2(\alpha)}_{k-1}$ is the upper α point of the chi-square distribution with the degree of freedom $k-1$. In addition, log in B denotes the natural logarithm. Since this procedure is based on the asymptotic result, it is liberal in the case of the small sample. (See Yoshimura (1987).) The point is out of consideration in this thesis.

Global Type I FWE's with the preliminary test for all parameters used in the previous section are shown on Figures 22 to 35.

The notation B5 in the figures denotes the result of the proper use of Tukey-Kramer procedure and GHC-procedure with Bartlett's test of the significant level 5%, and the notation B50 denotes that with Bartlett's test of the significant level 50%. Since it is the main current that the preliminary test is practiced for selecting procedures, we will simulate in these two cases to confirm the view point.

First, what is generally shown in all figures is that the global Type I FWE with the preliminary test is not certainly controlled even if each procedure is controlled well. This phenomenon is conspicuously shown on B5: the practical global Type I FWE is located over both Tukey-Kramer procedure and GHC-procedure. On the other hand, B50 is located between Tukey-Kramer procedure and GHC-procedure.

Moreover, obviously B5 is more strongly influenced by Tukey-Kramer procedure than B50. Hence, in the case of the unbalanced and non-homogeneous one-way layout, B5 is very liberal. On the other hand, B50 is better than B5, but it is also liberal when data is non-homogeneous and has small unbalanced sample sizes. Since the trend extends as increasing of the number of treatment k, it becomes liberal at larger sample sizes for large k. We cannot judge whether it is due to the unstability of Bartlett's test for the

small sample or due to another feature not depending on it. Although we can judge that B50 is generally better, we need to note the previous indication. Furthermore, this result is the reason why we do not easily improve GHC-procedure through changing the weight parameter. That is to say, if we improve to increase the real Type I FWE on the small sample, we can easily predict that the violation of the Type I FWE with the preliminary test becomes larger, and it is not clearly 'improvement'.

In conclusion of this comparison, we obtain the result that we should use B50 rather than B5. However, we cannot get satisfactory result even using B50.

4.8.3 Proposal of a New Preliminary Test

The formulation of the primary, ordinary preliminary test is the following. Firstly, we practice the usual test for the overall null hypothesis of the homogeneity of variances. Secondly, we proceed to a procedure not assuming the homogeneity if the test rejects, otherwise to a procedure assuming the homogeneity.

Recently it is pointed out that it is not suitable we practice the preliminary test within the limit of the usual test though practicing it means selection of a procedure in fact. (See Nagata (1992).) However, this consideration has not permeated yet, because of not studying the performance of procedures with the preliminary test in detail. Since it is difficult to study theoretically, we inspect it on the basis of Monte Carlo simulation and try to improve the preliminary test in this thesis.

As is shown in the previous subsection, in multiple comparisons we cannot avoid the disadvantage by means of adjustment of the significant level of the preliminary test, which is the idea used in Nagata (1992). The reason is that the procedure assuming the homogeneity is so sensitive to the non-homogeneity that the global significant level cannot be kept and treating groups on the multiple decision becomes extremely unfair. To solve this problem, we consider nothing but abandoning the ordinary, primary test system in which the null hypothesis is centered and making a new system more emphasizing the

alternative hypothesis.

At first, we enumerate features that may be needed in the preliminary test.

- Emphasize Type II error (especially in the case of the small sample).
- Hardly reject the null hypothesis if it is true in the case of the large sample.
- Easily reject the null hypothesis if it is false in the case of the large sample.

The ordinary test system ignore the first feature and realize the second feature through making it be hard to reject permanently. Needless to say, it satisfies the third feature. It is the problem that the preliminary test ignore the first feature though it is selection in fact, because the following opposition holds when we want to practice the preliminary test.

Procedure not keeping the significant level out of the homogeneity assumption

vs.

procedure not restricted but being less powerful.

However, even if we control the significant level to emphasize the first feature and to loosen the second feature, it dose not go well in multiple comparisons. Though we make the concession raising the significant level of the preliminary test up to 50% in the previous section, we observe the phenomenon not keeping the global significant level. More increasing of the significant level means nothing but disregarding the second feature and is almost same as using a procedure not assuming the homogeneity directly from the start. Nevertheless, it is just a selection if we do not care the decline of the power. Conversely, we will try to improve it while adhering to three features in this chapter.

The preliminary test proposed in this thesis is a test on the basis of the alternative hypothesis, which is derived from Bartlett's test. Although the procedure simply decides

critical values to make the probability of the Type II error constant, usually the alternative hypothesis for the homogeneity is a composite hypothesis. Thus, if we control the Type II error for all parameters, it is same as hardly accepting the null hypothesis. Furthermore, if we test for a certain simple alternative hypothesis predetermined, the third feature does not hold for alternative hypotheses being nearer to the null hypothesis than the simple hypothesis. Hence, the selection of the simple alternative hypothesis largely influence the test. Since the test statistic using in the preliminary test has the noncentral chi-square distribution, how we treat the alternative hypothesis reduces how we determine the noncentral parameter. Therefore, we manage to satisfy the third feature by means of determining the noncentral parameter.

Now, we will consider Bartlett's test under the alternative hypothesis. The test statistic adopted is the following corrected one.

$$B = \frac{\nu \log \overline{S}^2 - \sum_{i=1}^{k} (\nu_i \log S_i^2)}{1 + \frac{\sum_{i=1}^{k} \frac{1}{\nu_i} - \frac{1}{\nu}}{3k-3}}.$$

Referring Kendall and Stuart (1979), we find that the noncentral parameter under a simple alternative hypothesis is given in the following formula.

$$\lambda = \sum_{i=1}^{k} \frac{\nu_i}{2} \left(\frac{\sigma_i^2}{\overline{\sigma}^2} - 1 \right)^2,$$

where σ_i^2 denotes the variance of each group under the alternative hypothesis and $\overline{\sigma}^2 = \sum_{i=1}^{k} \nu_i \sigma_i^2 / \nu$. However, on Monte Carlo simulation it does not fit well when the noncentral parameter becomes larger as the sample size larger because of influence of omitted terms. Hence, we use the following form as a prototype, which preserves the original form of Bartlett' test statistic.

$$\lambda = -\sum_{i=1}^{k} \nu_i \log(\sigma_i^2 / \overline{\sigma}^2).$$

(See the next section in detail.)

By using this λ, we can test for a specified simple alternative hypothesis $H_1 : \lambda = \lambda_0$ with a constant probability of the Type II error. In order to satisfy the third feature, we should control increasing of the noncentral parameter with order ν_i. That is to say, since fixed λ means the situation of the null hypothesis and increasing with order ν_i means the situation of the alternative hypothesis, the third feature is satisfied if we make λ to increase with less order than order ν_i.

As a simple setting, the following form increasing with order $\sqrt{\nu_i}$ shows a comparably good performance in Monte Carlo simulation.

$$\lambda_0 = -\sqrt{n_{\min} - 1} \sum_{i=1}^{k} \log(\sigma_{0i}^2/\overline{\sigma}_0^2),$$

where $n_{\min}$ denotes the minimum of n_i, $\sigma_{0i}^2 = 28i - 27$ $(i = 1, 2, \ldots, k)$ and $\overline{\sigma}_0^2 = \sum_{i=1}^{k} \sigma_{0i}^2/k$. In this parameter, σ_{0i}^2 that is the basis of the power is determined in the following way. Firstly, we select equally spaced configuration of variances among several patterns because the pattern is least influenced by the number of treatments k on simulation. Secondly, we decide the gap of the pattern as the probability of the global Type I error becomes 5% at sample size 10 because the ordinary preliminary test with the significant level 50% becomes so in the previous section. Figures 20 and 21 then shows examples of the practical rejective rate. Since the test concentrate on the alternative hypothesis in the small sample but it becomes harder to reject the homogeneity as sample sizes become larger, it satisfies the second feature. Besides, curves fitted to results is determined in the following type of function.

$$a + \frac{b}{\sqrt{n}} + \frac{c}{n} + \frac{d}{n\sqrt{n}},$$

where n denotes the sample size of the first treatment and estimates of coefficients a, b, c and d are determined by the least square method.

Moreover, since rejective rates for the null hypothesis at sample size 10 for some numbers of treatments k are given in the following table, we find that they are stable.

k	3	4	5	6	7
Rejective rates	0.51	0.52	0.51	0.49	0.48

Furthermore, by comparing λ_0 and the primary λ when $k = 3$, we obtain (1,4.5,8) at $n_i = 10$ and (1,3,5) at $n_i = 20$ as values of variances of the corresponding simple alternative hypothesis to decide the critical value. It means that the simple alternative hypotheses approach the null hypothesis as sample sizes become larger.

Finally, the procedure for the preliminary test proposed in this chapter is in the following.

1. Calculate a value of Bartlett's test statistic B.
2. Calculate a value of the noncentral parameter λ_0.
3. Obtain the lower 1 percentile B_0 of the noncentral chi-square distribution with the degree of freedom $k-1$ and the noncentral parameter λ_0.
   (In our simulation, we use Patnaik's (1949) approximation to obtain the value.)
4. Practice multiple comparisons by Tukey-Kramer procedure if $B < B_0$, otherwise by GHC2-procedure.

4.8.4 Evaluation of Bartlett's Test Statistic under the Alternative Hypothesis

We can obtain the following form of the probability density for one-way layout model by another parameter setting.

$$L(Y|\mu, \sigma^2, \boldsymbol{r}) = \prod_{i=1}^{k} \left[(2\pi\sigma^2 r_i)^{-n_i/2} \exp\left\{ -\frac{\sum_{j=1}^{n_i}(Y_{ij} - \mu_i)^2}{2\sigma^2 r_i} \right\} \right],$$

where $Y = \{Y_{ij}\}$ and relation to the previous setting is that $\sigma^2 = \sum n_i \sigma_i^2 / \sum n_i$ and $r_i = \sigma_i^2/\sigma^2$, hence $\sum n_i r_i / \sum n_i = 1$. Therefore, the likelihood ratio test statistic, that is, the source of Bartlett's test statistic for the null hypothesis $H_0 : r_i = 1$ $(i = 1, 2, \ldots, k)$

is obtained as the following form.

$$\ell = \frac{L(Y|\overline{Y}, V, \mathbf{1})}{L(Y|\overline{Y}, V, \boldsymbol{T})},$$

where

$$\overline{Y} = (\overline{Y}_{1.}, \ldots, \overline{Y}_{k.}), \quad \overline{Y}_{i.} = \sum_{j=1}^{n_i} Y_{ij}/n_i,$$

$$V_i = \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i.})^2/n_i, \quad V = \sum_{i=1}^{k} n_i V_i / \sum n_i,$$

$$T_i = V_i/V, \quad \boldsymbol{T} = (T_1, \ldots, T_k), \quad \mathbf{1} = (1, \ldots, 1).$$

By substituting the previous equation, we get the following result with some calculation.

$$-2\log\ell = \sum_{i=1}^{k} n_i(T_i - 1 - \log T_i).$$

Through Taylor's expansion for the logarithm, we reduce

$$= \sum_{i=1}^{k} \left\{ \frac{n_i(T_i-1)^2}{2} - \frac{n_i(T_i-1)^3}{3} + \frac{n_i(T_i-1)^4}{4} - \cdots \right\},$$

hence we can asymptotically omit the third order and later terms if the null hypothesis is true.

When the alternative hypothesis is true, if we can omit the third order and later terms then we obtain the noncentral parameter:

$$\lambda = \sum_{i=1}^{k} \frac{n_i(r_i - 1)^2}{2}.$$

However, we cannot omit the influence for a fixed simple alternative hypothesis because we observe the increasing of error by simulation. Now, we consider the original equation again before the expansion of the logarithm. It can be reformed in the following by using the parameter of a simple alternative hypothesis.

$$-2\log\ell = \sum_{i=1}^{k} \{n_i(\tilde{T}_i - 1 - \log\tilde{T}_i) - n_i \log r_i + n_i(T_i - \tilde{T}_i)\},$$

where $\tilde{T}_i = V_i/(Vr_i)$. The former of the right-hand side is nothing but the logarithm likelihood ratio test statistic when the simple alternative hypothesis is true, hence the noncentral parameter that make it correspond to the asymptotic expectation becomes

$$\lambda = -\sum_{i=1}^{k} n_i \log r_i.$$

We get the noncentral parameter in the previous section by replacing n_i with ν_i through Bartlett's correction. By Monte Carlo simulation, we can confirm that the approximation with the noncentral parameter is considerably improved.

THEOREM 1: *Bartlett's test statistic*

$$B = \frac{\nu \log \overline{S}^2 - \sum_{i=1}^{k}(\nu_i \log S_i^2)}{1 + \frac{\sum_{i=1}^{k} \frac{1}{\nu_i} - \frac{1}{\nu}}{3k-3}}$$

under the alternative hypothesis approximately has the noncentral chi-square distribution with the degree of freedom $k-1$ *and the noncentral parameter*

$$\lambda = -\sum_{i=1}^{k} \nu_i \log(\sigma_i^2/\overline{\sigma}^2),$$

where $S_i^2 = \sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i.})^2/\nu_i$, $\overline{S}^2 = \sum_{i=1}^{k} \nu_i S_i^2/\nu$ *and* $\overline{\sigma}^2 = \sum \nu_i \sigma_i^2 / \sum \nu_i$.

4.8.5 Consideration for Result on the Overall Null Hypothesis

Results shown with the notation PT on Figures 22 to 35 are those with the new preliminary test proposed.

Case of Homogeneous Variances

In this case on Figures 22 to 25, we can find that the procedure with the preliminary test almost keeps the significant level. The procedure with the preliminary test is occasionally liberal though each procedure combined keeps the significant level. But, the violation is at most about 0.3%.

Case of Unequal Variances

In the balanced case as is shown on Figures 26 to 28, since the violation of Tukey-Kramer procedure is small, the procedure with the preliminary test keeps the significant level without trouble. On the other hand, in the unbalanced case on Figure 29 when the variance of the sample with the smallest size is largest, the procedure with the preliminary test does not slightly keep the significant level, but it is better than the case of the ordinary preliminary test with the significant level 50%, which is practiced in Section 4.8.2. Therefore, it has almost no problem on practical use. After here, as the adopted limit on practical use, we examine situations where the practical significant level for the procedure with the preliminary test is below 6%.

In about the twice unbalanced case as is shown on Figure 33, the violation of the practical significant level of Tukey-Kramer procedure becomes heavy, hence that of the procedure with the preliminary test is slightly dragged by it. Fortunately, proposed GHC2-procedure keeps the significant level and the practical significant level of the procedure with the preliminary test does not exceed 6%. In more unbalanced case, we think that GHC2-procedure will have little problem, but it is highly possible that the significant level of the procedure with the preliminary test exceeds 6%. Therefore, it is appropriate for us to use GHC2-procedure without the preliminary test.

On the other hand, in the case of Figures 34 and 35 that have heavy non-homogeneity, the practical significant level of the procedure with the preliminary test exceeds 6% at one-way layouts with a sample of size 2. Hence, in the unbalanced case with a sample of size 2, it is appropriate for us to use GHC2-procedure without the preliminary test.

4.8.6 Behavior on Powers

On Figures 16 and 19, results of 'mean rejective rate' in the case with the new preliminary test also enters. The power of the case with the preliminary test has 4% reduction from

the difference between Tukey-Kramer procedure and GHC2-procedure on Figure 16, hence it is effective. In the case of larger sample sizes on Figure 17, differences among three procedures is little, but the improvement rate of the procedure with the preliminary test relatively increases against that on Figure 16.

Furthermore, although the power of the procedure with the preliminary test is below the power of the ordinary preliminary test with significant level 50% on Figure 16, the reversal of the power arises on Figure 17.

4.8.7 Combination with Steel-Dwass procedure

As is mentioned above, it is popular that we properly use Tukey-Kramer procedure and Steel-Dwass procedure on a preliminary test system. Now, we will consider the problem.

The proper use with the ordinary preliminary test of the significant level 5% is out of the question, but that of higher significant level is also liberal in the case of the small sample, which is similar to the system of proper use of Tukey-Kramer procedure and GHC-procedure. If we use the new preliminary test proposed in this section instead of the ordinary preliminary test, the system almost keeps the global significant level but is liberal in the heavy unbalanced case as Figure 13. Furthermore, in the case of the small sample on Figure 9, it yields the extremely unbalanced result owing to the difference of the power between Tukey-Kramer procedure and Steel-Dwass procedure on the border of the judgement of the preliminary test. That is to say, if data is judged as the non-homogeneity then differences of means are hardly (or never) rejected, otherwise differences of means are rejected with about twice of the nominal significant level. Although the case of using GHC2-procedure is also unbalance, it is more improved than the case of using Steel-Dwass procedure.

4.9 Discussion

In the case of unequal variances, we recommend GHC2-procedure because it is conservative in the broad range of parameters. Furthermore, we conclude that we should avoid to use Steel-Dwass procedure in the non-homogeneity case because it is sometimes liberal.

Another purpose of this chapter is that we improve the ordinary preliminary test that does not satisfy the function needed, which is balancing the control of the significant level and decreasing of the power. Consequently, proposed procedure with the new preliminary test shows the good performance. Although the statistic of the preliminary test for the homogeneity has the noncentral chi-square distribution, the labor for the calculation increases only a little if we use the chi-square approximation.

Remained problem is avoidance of the discontinuity of results on the proper use with the preliminary test. Although the case of Tukey-Kramer procedure and GHC2-procedure is better than that of Tukey-Kramer procedure and Steel-Dwass procedure, we do not find the degree of the difference on the border of the judgement of the preliminary test.

Chapter 5

Multiple Comparison Procedures Based on a Loss Function

In this chapter, we propose new multiple comparison procedures based on a loss function in order to avoid the discontinuity on procedures with a preliminary test. And, we compare them with the new preliminary test system in the previous chapter.

5.1 Proposal of Multiple Comparison Procedures Based on a Loss Function

5.1.1 Improvement of Variance Estimators

If we consider a variance estimator among unbiased estimators, we need the proper use with a preliminary test. The reason is that estimators assuming the homogeneity of variances are more stable than others. (See Matsuda, Fujimoto and Yoshimura (1990) as an analogous problem.)

In this chapter, therefore, we consider to get a estimator that has good performance among not unbiased estimators using a loss function. However, the class of not unbiased estimators is broader, so that we consider only the following restricted class of variances.

$$\tilde{S}_i^2 = b\overline{S}^2 + (1-b)S_i^2, \quad 0 \leq b \leq 1$$

This class is the natural class combined separate variances, which respond to the

non-homogeneity, and the pooled variance, which is stable.

5.1.2 Fundamental Construction of Multiple Comparison Procedures

We construct multiple comparison procedures using a estimator in the class restricted above. The test statistic is the following natural form:

$$\tilde{T}_{ih} = \frac{\overline{Y}_{i.} - \overline{Y}_{h.}}{\sqrt{\tilde{S}_i^2/n_i + \tilde{S}_h^2/n_h}},$$

and the critical value is considered using several methods in sections below.

5.1.3 Estimation of the Optimal Value of b

It is the problem how we determine b to use in the fundamental construction in the previous subsection. In this subsection, we consider the estimation of the optimal value of b using a loss function.

We use the following function as a loss function of variance estimators $\tilde{S}_i^2$:

$$\sum_{i=1}^{k} \frac{\nu_i}{2} \left(\frac{\tilde{S}_i^2}{\sigma_i^2} - 1 \right)^2.$$

Under the loss function, the risk function for $\tilde{S}_i^2$ is

$$\begin{aligned}
E\left\{ \sum_{i=1}^{k} \frac{\nu_i}{2} \left(\frac{\tilde{S}_i^2}{\sigma_i^2} - 1 \right)^2 \right\} &= \sum_{i=1}^{k} \frac{\nu_i}{2} \left\{ \frac{b^2}{\sigma_i^4} E(\overline{S}^4) + \frac{2b(1-b)}{\sigma_i^4} E(\overline{S}^2 S_i^2) + \frac{(1-b)^2}{\sigma_i^4} E(S_i^4) \right. \\
&\quad \left. - \frac{2b}{\sigma_i^2} E(\overline{S}^2) - \frac{2(1-b)}{\sigma_i^2} E(S_i^2) + 1 \right\} \\
&= b^2 A + 2b(1-b) + k(1-b)^2 \\
&= (A + k - 2) b^2 A - 2(k-1)b + k \\
&= (A + k - 2) \left(b - \frac{k-1}{A+k-2} \right)^2 + \frac{Ak - 1}{A + k - 2},
\end{aligned}$$

where

$$A = \sum_{i=1}^{k} \frac{\nu_i}{2} \left\{ \left(\frac{\overline{\sigma}^2}{\sigma_i^2} - 1 \right)^2 + \frac{2V^4}{\nu \sigma_i^4} \right\}, \quad V^4 = \frac{1}{\nu} \sum_{i=1}^{k} \nu_i \sigma_i^4$$

Now, we get the following theorem.

THEOREM 2: *Among the class of variance estimators:* $\tilde{S}_i^2 = b\overline{S}^2 + (1-b)S_i^2$, $0 \leq b \leq 1$, *the risk function corresponding to the loss function*

$$\sum_{i=1}^{k} \frac{\nu_i}{2} \left(\frac{\tilde{S}_i^2}{\sigma_i^2} - 1 \right)^2$$

get the minimum at

$$b = \frac{k-1}{A+k-2}.$$

Proof:

We must show

$$0 \leq \frac{k-1}{A+k-2} \leq 1.$$

That is to say, we may show $A + k - 2 \geq k - 1$.

$$\begin{aligned} A &= \sum_{i=1}^{k} \frac{\nu_i}{2} \left\{ \left(\frac{\overline{\sigma}^2}{\sigma_i^2} - 1 \right)^2 + \frac{2V^4}{\nu \sigma_i^4} \right\} \\ &\geq \sum_{i=1}^{k} \frac{\nu_i V^4}{\nu \sigma_i^4} \\ &= \left(\frac{1}{\nu} \sum_{i=1}^{k} \frac{\nu_i}{\sigma_i^4} \right) \left(\frac{1}{\nu} \sum_{i=1}^{k} \nu_i \sigma_i^4 \right) \\ &\geq \left\{ \prod_{i=1}^{k} \left(\frac{1}{\sigma_i^4} \right)^{\nu_i} \right\}^{\frac{1}{\nu}} \left\{ \prod_{i=1}^{k} (\sigma_i^4)^{\nu_i} \right\}^{\frac{1}{\nu}} \\ &= 1 \end{aligned}$$

(Q.E.D.)

5.2 Multiple Comparison Procedures Expanding GH-Procedure

5.2.1 Setting of Critical Value

We consider to determine the critical value in the previous section using the approximation by t-distribution as like as GH-procedure. That is to say, we need the degree of freedom $\tilde{\nu}_{ih}$ satisfying the following approximation:

$$\frac{\tilde{S}_i^2/n_i + \tilde{S}_h^2/n_h}{\sigma_i^2/n_i + \sigma_h^2/n_h} \approx \frac{1}{\tilde{\nu}_{ih}}\chi^2_{\tilde{\nu}_{ih}}.$$

However, the numerator statistic of the left-hand side is not unbiased for the denominator of that, so that we use the following approximation:

$$\frac{\tilde{S}_i^2/n_i + \tilde{S}_h^2/n_h}{\tilde{\sigma}_i^2/n_i + \tilde{\sigma}_h^2/n_h} \approx \frac{1}{\tilde{\nu}_{ih}}\chi^2_{\tilde{\nu}_{ih}},$$

where

$$\tilde{\sigma}_i^2 = b\overline{\sigma}^2 + (1-b)\sigma_i^2.$$

Here, we get the following variance of the numerator:

$$\begin{aligned}
\mathrm{Var}\left(\frac{\tilde{S}_i^2}{n_i} + \frac{\tilde{S}_h^2}{n_h}\right) &= \frac{1}{n_i^2}\mathrm{Var}(\tilde{S}_i^2) + \frac{2}{n_i n_h}\mathrm{Cov}(\tilde{S}_i^2, \tilde{S}_h^2) + \frac{1}{n_h^2}\mathrm{Var}(\tilde{S}_h^2) \\
&= \frac{1}{n_i^2}\left(\frac{2b^2}{\nu}V^4 + \frac{4b(1-b)}{\nu}\sigma_i^4 + \frac{2(1-b)^2}{\nu_i}\sigma_i^4\right) \\
&\quad + \frac{2}{n_i n_h}\left(\frac{2b^2}{\nu}V^4 + \frac{2b(1-b)}{\nu}(\sigma_i^4 + \sigma_h^4)\right) \\
&\quad + \frac{1}{n_h^2}\left(\frac{2b^2}{\nu}V^4 + \frac{4b(1-b)}{\nu}\sigma_h^4 + \frac{2(1-b)^2}{\nu_h}\sigma_h^4\right) \\
&= \frac{2b^2}{\nu}V^4 N_{ih}^2 + \frac{4b(1-b)}{\nu}\left\{N_{ih}\left(\frac{\sigma_i^4}{n_i} + \frac{\sigma_h^4}{n_h}\right)\right\} \\
&\quad + 2(1-b)^2\left(\frac{\sigma_i^4}{n_i^2\nu_i} + \frac{\sigma_h^4}{n_h^2\nu_h}\right),
\end{aligned}$$

where

$$N_{ih} = \frac{1}{n_i} + \frac{1}{n_h}.$$

Besides, we note

$$\mathrm{Var}(S_i^2) = \frac{2\sigma_i^4}{\nu_i}, \quad \mathrm{Var}(\overline{S}^2) = \frac{2}{\nu}V^4.$$

Therefore, we get the following theorem by solving the equation:

$$\mathrm{Var}\left(\frac{\tilde{S}_i^2}{n_i} + \frac{\tilde{S}_h^2}{n_h}\right) = 2\left(\frac{\tilde{\sigma}_i^2}{n_i} + \frac{\tilde{\sigma}_h^2}{n_h}\right)^2 \frac{1}{\tilde{\nu}_{ih}}.$$

THEOREM 3: *The distribution of the estimator*

$$\tilde{T}_{ih} = \frac{\overline{Y}_{i.} - \overline{Y}_{h.}}{\sqrt{\tilde{S}_i^2/n_i + \tilde{S}_h^2/n_h}}$$

is approximate to t-distribution with the degree of freedom:

$$\tilde{\nu}_{ih} = \frac{\left(\frac{\tilde{\sigma}_i^2}{n_i} + \frac{\tilde{\sigma}_h^2}{n_h}\right)^2}{\frac{b^2}{\nu}V^4 N_{ih}^2 + \frac{2b(1-b)}{\nu}\left\{N_{ih}\left(\frac{\sigma_i^4}{n_i} + \frac{\sigma_h^4}{n_h}\right)\right\} + (1-b)^2\left(\frac{\sigma_i^4}{n_i^2\nu_i} + \frac{\sigma_h^4}{n_h^2\nu_h}\right)}.$$

By this theorem, we can construct a multiple comparison procedure as like as GH-procedure, which is the procedure using $Q_{k,\tilde{\nu}_{ih}}^{(\alpha)}/\sqrt{2}$ as the critical value for the test statistic $|\tilde{T}_{ih}|$.

5.2.2 Improvement of the Degree of Freedom of GH-Procedure

The procedure proposed in this section is based on GH-procedure. Since GH-procedure is liberal, we need to improve it. At first, we simulate GH-procedure with known σ_i^2's in order to study the reason why it is liberal. From the result, we find that we cannot test

exactly even if we know the true value of variances, so that we will improve the degree of freedom itself in the following.

The modified GH-procedure proposed here is the procedure using the following degree of freedom:

$$\nu'_{ih} = \frac{\sigma_i^4/n_i^2 + \sigma_h^4/n_h^2 + 2\sigma_i^2\sigma_h^2/\{(n_i+1)(n_h+1)\}}{\sigma_i^4/(n_i^2\nu_i) + \sigma_h^4/(n_h^2\nu_h)}$$

instead of ν_{ih} in GH-procedure. This improvement is due to the thought that the cause not well-fitting is at the term $\sigma_i^2\sigma_h^2$, so that the coefficient has been changed through trial and error.

The modified GH-procedure shows less violation of the significant level than GH-procedure and is practical to use.

For example, in the result of a simulation with 10000 repeats when $n = (4, 4, 2)$ and $\sigma^2 = (1, 1, 3)$, a practical Type I FWE of GH-procedure at level 0.05 is 0.083 but a practical Type I FWE of the modified GH-procedure is 0.066.

5.2.3 Multiple Comparison Procedure Based on the Loss Function (the Modified GH-Procedure Type)

We propose the modified GH-procedure type of multiple comparison procedure based on the loss function using Theorems 2 and 3 in the following.

1. Calculate a value of the estimator $\hat{b}$ by the following equation:

$$\hat{b} = \frac{k-1}{\hat{A}+k-2},$$

where

$$\begin{aligned}\hat{A} &= \sum_{i=1}^{k} \frac{\nu_i}{2}\left\{\left(\frac{\overline{S}^2}{S_i^2} - 1\right)^2 + \frac{2\hat{V}^4}{\nu S_i^4}\right\}, \\ \hat{V}^4 &= \frac{1}{\nu}\sum_{i=1}^{k} \nu_i S_i^4.\end{aligned}$$

2. Calculate a value of the modified variance estimator:

$$\tilde{S}_i^2 = \hat{b}\overline{S}^2 + (1-\hat{b})S_i^2.$$

(It is possible to repeat the calculation of $\hat{b}$ using $\tilde{S}_i^2$ instead of S_i^2 in Step 1. In this thesis, we repeat Steps 1 and 2 three times to improve the estimate $\hat{b}$.)

3. Obtain the degree of freedom by the following equation:

$$\tilde{\nu}'_{ih} = \frac{\frac{\tilde{S}_i^4}{n_i^2} + \frac{\tilde{S}_h^4}{n_h^2} + \frac{2\tilde{S}_i^2\tilde{S}_h^2}{(n_i+1)(n_h+1)}}{\frac{b^2}{\nu}\hat{V}^4 N_{ih}^2 + \frac{2b(1-b)}{\nu}\left\{N_{ih}\left(\frac{S_i^4}{n_i} + \frac{S_h^4}{n_h}\right)\right\} + (1-b)^2\left(\frac{S_i^4}{n_i^2\nu_i} + \frac{S_h^4}{n_h^2\nu_h}\right)}.$$

4. We test by comparing the test statistic:

$$|\tilde{T}_{ih}| = \frac{|\overline{Y}_{i.} - \overline{Y}_{h.}|}{\sqrt{\tilde{S}_i^2/n_i + \tilde{S}_h^2/n_h}}$$

with the critical value:

$$\xi_{ih} = Q^{(\alpha)}_{k,\tilde{\nu}'_{ih}}/\sqrt{2}.$$

We call the procedure *LMGH-procedure* in this thesis.

For example, in the result of a simulation with 10000 repeats when $n = (4,4,2)$ and $\sigma^2 = (1,1,3)$, a practical Type I FWE of GH-procedure type of multiple comparison procedure based on the loss function at level 0.05 is 0.102 but a practical Type I FWE of LMGH-procedure is 0.085.

5.3 Multiple Comparison Procedures by Another Approach

Well-known Welch's method is modified by Welch. However, Pagurova's method is better than it. (See Mehta and Srinivasan (1970), Kendall and Stuart (1979).) In this section, we consider the similar approach to Pagurova's method to construct the critical value of the multiple comparison procedure based on the loss function.

5.3.1 Pagurova's Method

Pagurova's method is the method in the two-sample problem using the following critical value for the test statistic T_{12} in Section 2.4.

$$\frac{t_{\nu_2}}{\theta^2}\{(1-g)(\theta-g)^2\}+\frac{t_{\nu_1+\nu_2}}{\theta^2(1-\theta)^2}\{g(1-g)\}\{\theta(1-\theta)+(g-\theta)(2\theta-1)\}+\frac{t_{\nu_1}}{(1-\theta)^2}\{g(\theta-g)^2\},$$

where

$$\begin{aligned}\theta &= \frac{\nu_1}{\nu_1+\nu_2},\\ g &= C\left(1-\frac{2}{\nu_2}\right)+2C^2\left(\frac{1}{\nu_1}+\frac{2}{\nu_2}\right)-2C^3\left(\frac{1}{\nu_1}+\frac{1}{\nu_2}\right),\\ C &= \frac{S_1^2/n_1}{S_1^2/n_1+S_2^2/n_2}.\end{aligned}$$

5.3.2 Multiple Comparison Procedures Based on the Loss Function (Pagurova's Method Type)

The test statistic $\tilde{T}_{ih}$ is reconstructed in the following:

$$\tilde{T}_{ih}=\frac{Z}{\sqrt{C_{ih}\tilde{S}_i^2/\sigma_i^2+(1-C_{ih})\tilde{S}_h^2/\sigma_h^2}},$$

where $Z\sim N(0,1)$ and

$$C_{ih}=\frac{\sigma_i^2/n_i}{\sigma_i^2/n_i+\sigma_h^2/n_h}.$$

Therefore, we get the following critical value by the similar way to Pagurova's method.

$$\xi_{ih}=bQ_{\nu'}/\sqrt{2}+(1-b)\left[\frac{Q_{\nu_h}}{\theta_{ih}^2}\{(1-g_{ih})(\theta_{ih}-g_{ih})^2\}+\frac{Q_{\nu_i+\nu_h}}{\theta_{ih}^2(1-\theta_{ih})^2}\{g_{ih}(1-g_{ih})\}\right.$$
$$\left.\times\{\theta_{ih}(1-\theta_{ih})+(g_{ih}-\theta_{ih})(2\theta_{ih}-1)\}+\frac{Q_{\nu_i}}{(1-\theta_{ih})^2}\{g_{ih}(\theta_{ih}-g_{ih})^2\}\right]/\sqrt{2},$$

where

$$\begin{aligned}\theta_{ih} &= \frac{\nu_i}{\nu_i+\nu_h},\\ g_{ih} &= \hat{C}_{ih}\left(1-\frac{2}{\nu_h}\right)+2\hat{C}_{ih}^2\left(\frac{1}{\nu_i}+\frac{2}{\nu_h}\right)-2\hat{C}_{ih}^3\left(\frac{1}{\nu_i}+\frac{1}{\nu_h}\right).\end{aligned}$$

Now, we consider two ways of defining ν' and $\hat{C}_{ih}$. We call these procedures *LP1-procedure* and *LP2-procedure*, respectively.

1. $\nu' = \nu, \quad \hat{C}_{ih} = \dfrac{S_i^2/n_i}{S_i^2/n_i + S_h^2/n_h}$

2. $\nu' = \nu\dfrac{\overline{\sigma}^4}{V^4}, \quad \hat{C}_{ih} = \dfrac{\tilde{S}_i^2/n_i}{\tilde{S}_i^2/n_i + \tilde{S}_h^2/n_h}$

Definition 1 is due to simply ignoring the difference of T_{ih} and $\tilde{T}_{ih}$, and Definition 2 is due to obtaining the exact value of ν' by substituting $b = 1$ into $\tilde{\nu}$ in Theorem 3 in Section 5.2 and calculating $\hat{C}_{ih}$ by using modified estimators of variance. Besides, we use estimators $\overline{S}^2$ and $\hat{V}^4$ in the practical procedure instead of $\overline{\sigma}^2$ and V^4 in ν' of Definition 2.

5.4 Comparison on Monte Carlo Simulation

In the result of a simulation with 10000 repeats, we get the following practical Type I FWE's.

n	σ^2	LMGH	LP1	LP2
(2,2,2)	(1,1,1)	0.018	0.023	0.025
(4,4,2)	(1,1,1)	0.051	0.053	0.050
(4,4,2)	(1,1,3)	0.085	0.087	0.079
(5,5,3)	(1,1,3)	0.075	0.075	0.075
(6,6,4)	(1,1,3)	0.066	0.064	0.066
(11,11,9)	(1,1,3)	0.058	0.058	0.058

As a whole, LP2-procedure is the best. Besides, the difference of LP1- and LP2-procedures is due to the change of $\hat{C}_{ih}$, which is confirmed by simulation on several definitions.

Next, we compare them to the procedure with the new preliminary test in Chapter 4. Figures 36 and 37 show the result. In the case of homogeneous variances as Figure 36, LP2 procedure has the best performance. However, in the case of the unbalanced and non-homogeneous one-way layout as Figure 37, the procedure with the preliminary test

is conservative but LMGH- and LP2-procedures are liberal. LP2-procedure is slightly better than LMGH-procedure and both are practical over sample size 10.

5.5 Discussion

In this chapter, we investigate whether a single procedure without the preliminary test is conservative under the situation where it is possible that variances are non-homogeneous. LP2-procedure is the best procedure among procedures proposed in this chapter. However, this procedure is inferior to the procedure with the new preliminary test that is proposed in the previous chapter, that is, the range where LP2-procedure do not keep the significant level is broader than that for the procedure with the preliminary test. Certainly, LP2-procedure has the continuous result owing to the absence of the preliminary test, so that it has the merits and demerits.

Further problem is modification based on GHC2-procedure in the previous chapter and more modification of LMGH-procedure, where it may be difficult to determine the degree of freedom.

Appendix

Generation of Random Numbers

We will mention random numbers used in this thesis.

First, we explain the basic uniform random numbers, which is a modification of the method in Yamauti (1972). The practical C program is the following.

```
/* Uniform random number
 *   if (n_ran != 0)
 *       if (n_ran > 0) and (n_ran < 8388593L)
 *           INITIALIZE WITH n_ran;
 *       else
 *           INITIALIZE WITH 1;
 *   else
 *       MAKING UNIFORM RANDOM NUMBER ON (0,1);
 */

double urandom(long n_ran)
{
    int i;
    unsigned long xl, z;
    static unsigned long y, ra[2], x[2];
    double xd;

    if (n_ran != 0) {
 /* Initialize values for making random numbers */
        ra[0] = 78125L;
        ra[1] = 262141L;
        y = 8388593L;
        if ((n_ran < 0) && (n_ran >= y)) {
            fprintf(stderr, "Initial value is irregular.\n");
            fprintf(stderr, "Initial value is set as 1.\n");
            n_ran = 1L;
        }
        x[0] = n_ran;
        x[1] = 1L;
        urandom(0L);
        return 0.;
```

```
    } else {
 /* Making of uniform random number on (0,1) */
        for (i = 0; i < 2; i++) {
            xd = (double)x[i] * (double)ra[i];
            xl = (long)(xd / (double)y);
            x[i] = (long)(xd - ((double)xl * (double)y));
        }
        z = x[0] ^ x[1];
        if (z == 0)
            z = 1;
        return ((double)z / (double)(y + 15L));
    }
}
```

This random numbers reveal good results for several tests. (Checked tests are the frequency test by 1 digit, the frequency test by 2 digit, Gap test, Collision test and Random walk test. See Knuth (1981).)

Next, we explain the normal random numbers, which generate adapting the inverse function method to the previous uniform random numbers. The used function is the approximation by Toda (1967). (See Yamauti (1972).) That is the following formula, which is the function of the upper probability Q to obtain the percentile $u(Q)$.

$$u(Q) = \{y(b_0 + b_1 y + b_2 y^2 + \cdots + b_{10} y^{10}\}^{1/2}, \quad (0 < Q \leq 0.5),$$

where

$$y = -log\{4Q(1-Q)\},$$

b_0= 0.1570796288×10, b_1 =0.3706987906×10^{-1}, b_2=−0.8364353589×10^{-3},

b_3=−0.2250947176×10^{-3}, b_4 =0.6841218299×10^{-5}, b_5= 0.5824238515×10^{-5},

b_6=−0.1045274970×10^{-5}, b_7 =0.8360937017×10^{-7}, b_8=−0.3231081277×10^{-8},

b_9= 0.3657763036×10^{-10}, b_{10}=0.6936233982×10^{-12}.

Tables and Figures

Table 1. Powers of each procedure corresponding to powers of Tukey's procedure: .25, .50 and .75.
($k = 4$, $n = 6$)
A: all-pairs power, B: restricted all-pairs power, C: minimum difference power,
D: maximum difference power, E: mean rejective rate, F: weighted mean rejective rate.

Config.	Proc.	A .25	A .50	A .75	B .25	B .50	B .75	C .25	C .50	C .75	D .25	D .50	D .75	E .25	E .50	E .75	F .25	F .50	F .75
EQ	TW(Q)	.437	.690	.866	.355	.620	.833	.360	.639	.851	.265	.527	.777	.299	.576	.818	.291	.561	.804
	TW(F)	.437	.690	.866	.346	.610	.827	.361	.639	.851	.261	.525	.779	.301	.576	.818	.293	.562	.803
	HC	.592	.819	.927	.319	.584	.813	.342	.661	.897	.237	.487	.743	.265	.558	.831	.260	.534	.799
	P(Q)	.604	.823	.926	.370	.632	.838	.399	.700	.900	.267	.529	.778	.304	.603	.856	.292	.574	.826
	P(F)	.604	.823	.926	.373	.630	.836	.401	.701	.900	.263	.528	.781	.309	.607	.856	.297	.578	.826
MIN	TW(Q)	.387	.670	.867	*			**			.294	.578	.825	.295	.582	.829	****		
	TW(F)	.392	.674	.869	*			**			.316	.602	.838	.311	.602	.841	****		
	HC	.317	.592	.819	*			**			.266	.540	.794	.263	.540	.796	****		
	P(Q)	.468	.752	.917	*			**			.301	.594	.843	.304	.602	.850	****		
	P(F)	.509	.781	.928	*			**			.332	.632	.866	.330	.634	.871	****		
MAX	TW(Q)	.405	.669	.865	.262	.519	.768	.348	.624	.837	***			.303	.589	.831	.286	.571	.823
	TW(F)	.407	.671	.865	.258	.514	.761	.349	.624	.838	***			.303	.591	.833	.285	.570	.824
	HC	.401	.668	.866	.235	.476	.731	.308	.587	.823	***			.272	.557	.814	.257	.539	.804
	P(Q)	.519	.766	.915	.263	.519	.768	.356	.649	.865	***			.304	.604	.853	.285	.580	.839
	P(F)	.542	.783	.922	.259	.514	.761	.359	.656	.874	***			.306	.609	.860	.285	.582	.845
SQ	TW(Q)	.410	.667	.856	.368	.636	.847	.370	.638	.845	.275	.531	.774	.295	.559	.798	.294	.558	.792
	TW(F)	.410	.666	.856	.370	.637	.847	.370	.638	.845	.273	.525	.768	.297	.559	.797	.296	.559	.792
	HC	.554	.788	.918	.307	.578	.819	.434	.731	.912	.242	.485	.738	.262	.543	.811	.261	.527	.782
	P(Q)	.564	.791	.917	.373	.642	.852	.489	.751	.903	.275	.531	.775	.294	.579	.833	.294	.563	.804
	P(F)	.563	.791	.917	.375	.644	.852	.490	.750	.903	.274	.526	.769	.298	.583	.834	.297	.566	.806
Stand. dev. (Max.)	TW(Q)	.015			.015			.009			.007			.004			.004		
	TW(F)	.014			.014			.010			.010			.005			.005		
	HC	.016			.009			.009			.008			.009			.006		
	P(Q)	.017			.014			.014			.008			.005			.004		
	P(F)	.017			.015			.014			.011			.007			.007		

* : equal to A, ** : a part of Type I error, *** : equal to B, **** : equal to E.

Table 2. Power of each procedure corresponding to power of Tukey's procedure: .25, .50 and .75.
($k = 4$, $n = 16$)
A: all-pairs power, B: restricted all-pairs power, C: minimum difference power,
D: maximum difference power, E: mean rejective rate, F: weighted mean rejective rate.

Config.	Proc.	A .25	A .50	A .75	B .25	B .50	B .75	C .25	C .50	C .75	D .25	D .50	D .75	E .25	E .50	E .75	F .25	F .50	F .75
EQ	TW(Q)	.452	.696	.866	.355	.617	.827	.365	.636	.842	.273	.531	.775	.298	.573	.816	.291	.559	.801
	TW(F)	.452	.696	.866	.348	.609	.824	.366	.636	.842	.277	.532	.774	.301	.574	.815	.294	.561	.801
	HC	.608	.821	.928	.327	.590	.813	.343	.660	.894	.249	.501	.752	.268	.558	.829	.266	.537	.797
	P(Q)	.613	.822	.928	.370	.628	.831	.398	.694	.893	.274	.533	.777	.301	.598	.852	.290	.570	.821
	P(F)	.613	.822	.928	.373	.628	.831	.400	.694	.893	.279	.536	.777	.307	.602	.851	.297	.574	.822
MIN	TW(Q)	.384	.653	.855	*			**			.300	.579	.819	.301	.581	.822	****		
	TW(F)	.389	.657	.856	*			**			.313	.594	.830	.315	.597	.832	****		
	HC	.322	.588	.816	*			**			.274	.543	.792	.275	.546	.794	****		
	P(Q)	.466	.734	.903	*			**			.309	.597	.838	.310	.600	.842	****		
	P(F)	.507	.762	.912	*			**			.331	.625	.857	.331	.627	.861	****		
MAX	TW(Q)	.415	.676	.859	.268	.522	.765	.341	.615	.832	***			.303	.587	.828	.286	.568	.819
	TW(F)	.417	.677	.860	.270	.517	.755	.346	.619	.833	***			.305	.589	.830	.286	.570	.822
	HC	.411	.674	.860	.248	.492	.739	.309	.587	.821	***			.277	.561	.815	.262	.544	.806
	P(Q)	.525	.777	.912	.268	.522	.765	.346	.637	.860	***			.302	.599	.849	.284	.575	.835
	P(F)	.549	.794	.918	.270	.517	.756	.356	.649	.867	***			.306	.607	.857	.285	.580	.842
SQ	TW(Q)	.416	.667	.852	.370	.640	.849	.371	.636	.841	.271	.532	.779	.294	.558	.797	.292	.556	.792
	TW(F)	.416	.667	.852	.373	.642	.849	.371	.636	.841	.273	.532	.778	.296	.558	.797	.296	.558	.792
	HC	.561	.788	.910	.312	.588	.823	.431	.728	.905	.245	.498	.755	.265	.544	.810	.266	.530	.783
	P(Q)	.566	.788	.910	.371	.646	.857	.477	.743	.897	.272	.532	.779	.293	.577	.830	.291	.560	.803
	P(F)	.566	.788	.910	.375	.649	.857	.478	.743	.897	.274	.532	.779	.298	.581	.831	.297	.565	.804
Stand. dev. (Max.)	TW(Q)	.014			.012			.008			.006			.004			.003		
	TW(F)	.014			.012			.008			.008			.004			.004		
	HC	.015			.010			.010			.009			.010			.007		
	P(Q)	.016			.011			.016			.006			.005			.004		
	P(F)	.016			.013			.016			.008			.005			.005		

* : equal to A, ** : a part of Type I error, *** : equal to B, **** : equal to E.

Table 3. Powers of each procedure corresponding to powers of Tukey's procedure: .25, .50 and .75.
($k = 5$, $n = 6$)
A: all-pairs power, B: restricted all-pairs power, C: minimum difference power,
D: maximum difference power, E: mean rejective rate, F: weighted mean rejective rate.

Config.	Proc.	A .25	A .50	A .75	B .25	B .50	B .75	C .25	C .50	C .75	D .25	D .50	D .75	E .25	E .50	E .75	F .25	F .50	F .75
EQ	TW(Q)	.488	.741	.894	.320	.587	.815	.387	.661	.863	.272	.529	.774	.287	.565	.817	.284	.550	.795
	TW(F)	.488	.741	.894	.306	.572	.805	.389	.661	.863	.261	.515	.763	.286	.565	.817	.279	.547	.795
	HC	.702	.897	.956	.273	.535	.782	.370	.697	.920	.236	.477	.732	.251	.543	.825	.251	.519	.786
	P(Q)	.698	.894	.956	.321	.588	.817	.404	.707	.913	.272	.530	.774	.285	.577	.840	.283	.554	.805
	P(F)	.698	.894	.956	.306	.573	.807	.405	.707	.913	.261	.515	.763	.284	.576	.840	.278	.550	.804
MIN	TW(Q)	.425	.696	.874	*			**			.292	.582	.832	.292	.574	.823	****		
	TW(F)	.449	.714	.884	*			**			.300	.597	.844	.302	.592	.837	****		
	HC	.345	.608	.819	*			**			.253	.528	.789	.253	.524	.783	****		
	P(Q)	.426	.696	.875	*			**			.292	.583	.832	.291	.574	.823	****		
	P(F)	.449	.715	.885	*			**			.301	.597	.844	.301	.592	.837	****		
MAX	TW(Q)	.444	.704	.878	.266	.518	.762	.322	.604	.837	***			.298	.583	.828	.284	.568	.821
	TW(F)	.460	.716	.883	.252	.488	.730	.325	.615	.849	***			.295	.588	.839	.277	.569	.830
	HC	.356	.619	.830	.230	.465	.717	.275	.548	.797	***			.259	.534	.793	.249	.524	.788
	P(Q)	.445	.704	.878	.266	.518	.762	.322	.604	.838	***			.298	.583	.829	.285	.569	.822
	P(F)	.460	.716	.883	.252	.488	.730	.325	.615	.849	***			.295	.588	.839	.277	.570	.831
SQ	TW(Q)	.458	.712	.881	.335	.607	.832	.397	.670	.866	.273	.534	.779	.287	.553	.797	.288	.550	.787
	TW(F)	.458	.712	.882	.344	.612	.832	.398	.670	.866	.259	.519	.771	.286	.553	.798	.285	.548	.787
	HC	.650	.863	.950	.277	.536	.779	.474	.783	.942	.239	.490	.745	.255	.535	.809	.253	.515	.774
	P(Q)	.656	.864	.950	.335	.607	.832	.513	.798	.936	.273	.534	.779	.283	.568	.828	.286	.552	.795
	P(F)	.656	.864	.950	.344	.612	.832	.513	.798	.937	.259	.519	.771	.283	.569	.829	.282	.550	.796
Stand. dev. (Max.)	TW(Q)	.020			.020			.010			.008			.005			.004		
	TW(F)	.017			.017			.010			.007			.005			.004		
	HC	.014			.012			.012			.008			.009			.004		
	P(Q)	.020			.020			.013			.007			.007			.004		
	P(F)	.019			.018			.013			.007			.007			.004		

* : equal to A, ** : a part of Type I error, *** : equal to B, **** : equal to E.

Table 4. Powers of each procedure corresponding to powers of Tukey's procedure: .25, .50 and .75.
($k = 4$, unbalance: (2,2,10,10))
A: all-pairs power, B: restricted all-pairs power, C: minimum difference power,
D: maximum difference power, E: mean rejective rate, F: weighted mean rejective rate.

Config.	Proc.	A .25	A .50	A .75	B .25	B .50	B .75	C .25	C .50	C .75	D .25	D .50	D .75	E .25	E .50	E .75	F .25	F .50	F .75
EQ	TW(Q)	.408	.674	.863	.357	.627	.843	.332	.623	.850	.273	.526	.771	.307	.574	.806	.296	.566	.802
	TW(F)	.408	.674	.863	.375	.644	.852	.353	.636	.851	.291	.563	.805	.321	.584	.806	.315	.581	.805
	HC	.566	.804	.923	.323	.589	.818	.295	.582	.827	.236	.481	.738	.271	.555	.818	.263	.537	.797
	P(Q)	.574	.805	.922	.383	.658	.866	.333	.627	.856	.274	.527	.771	.313	.603	.844	.298	.581	.826
	P(F)	.574	.806	.922	.437	.708	.889	.355	.643	.860	.293	.564	.806	.335	.620	.846	.322	.604	.834
MIN	TW(Q)	.394	.661	.854	*			**			.303	.583	.825	.298	.580	.824	****		
	TW(F)	.394	.660	.853	*			**			.317	.598	.832	.314	.597	.831	****		
	HC	.323	.587	.810	*			**			.266	.539	.794	.262	.534	.790	****		
	P(Q)	.464	.734	.898	*			**			.310	.598	.842	.305	.597	.842	****		
	P(F)	.469	.731	.891	*			**			.330	.619	.850	.327	.619	.851	****		
MAX	TW(Q)	.413	.674	.860	.276	.547	.796	.288	.566	.813	***			.299	.577	.818	.293	.567	.811
	TW(F)	.414	.674	.860	.296	.583	.828	.315	.595	.829	***			.312	.583	.816	.310	.579	.810
	HC	.410	.674	.862	.243	.502	.761	.254	.526	.786	***			.268	.547	.805	.261	.536	.797
	P(Q)	.555	.792	.915	.276	.548	.796	.288	.566	.813	***			.301	.596	.847	.292	.580	.834
	P(F)	.557	.794	.916	.297	.584	.827	.316	.597	.831	***			.316	.605	.846	.311	.594	.834
SQ	TW(Q)	.405	.666	.856	.368	.644	.851	.350	.630	.847	.270	.521	.762	.293	.570	.814	.283	.554	.801
	TW(F)	.405	.666	.856	.368	.644	.850	.360	.636	.847	.262	.519	.768	.295	.573	.816	.282	.555	.803
	HC	.543	.786	.918	.314	.604	.846	.340	.653	.887	.237	.478	.729	.262	.555	.828	.253	.526	.793
	P(Q)	.555	.791	.918	.386	.677	.881	.389	.705	.913	.270	.522	.762	.296	.597	.853	.283	.563	.820
	P(F)	.555	.791	.918	.387	.679	.882	.414	.724	.916	.263	.520	.770	.301	.605	.859	.282	.568	.825
Stand. dev. (Max.)	TW(Q)	.010			.013			.007			.007			.004			.004		
	TW(F)	.010			.013			.009			.009			.005			.005		
	HC	.013			.011			.007			.005			.006			.004		
	P(Q)	.015			.014			.010			.007			.005			.005		
	P(F)	.015			.016			.012			.009			.006			.006		

* : equal to A, ** : a part of Type I error, *** : equal to B, **** : equal to E.

Table 5. Powers of each procedure corresponding to powers of Tukey's procedure: .25, .50 and .75.
($k = 4$, unbalance: (4,4,4,12))
A: all-pairs power, B: restricted all-pairs power, C: minimum difference power,
D: maximum difference power, E: mean rejective rate, F: weighted mean rejective rate.

Config.	Proc.	A .25 .50 .75	B .25 .50 .75	C .25 .50 .75	D .25 .50 .75	E .25 .50 .75	F .25 .50 .75
EQ	TW(Q)	.445 .702 .869	.347 .619 .837	.366 .642 .851	.265 .522 .771	.297 .575 .818	.287 .558 .803
	TW(F)	.445 .702 .869	.337 .607 .831	.369 .642 .851	.274 .536 .785	.301 .575 .818	.295 .562 .802
	HC	.604 .830 .928	.319 .588 .817	.327 .634 .875	.227 .477 .739	.263 .557 .831	.258 .533 .799
	P(Q)	.615 .834 .928	.357 .630 .843	.391 .685 .887	.265 .522 .771	.299 .601 .857	.285 .568 .825
	P(F)	.615 .834 .928	.356 .628 .843	.395 .686 .887	.276 .538 .785	.307 .604 .857	.296 .575 .825
MIN	TW(Q)	.387 .661 .860	*	**	.297 .568 .809	.298 .581 .826	****
	TW(F)	.393 .665 .862	*	**	.317 .593 .824	.314 .597 .833	****
	HC	.320 .587 .810	*	**	.263 .521 .770	.266 .538 .792	****
	P(Q)	.472 .742 .908	*	**	.305 .580 .819	.309 .601 .846	****
	P(F)	.495 .761 .919	*	**	.332 .612 .837	.331 .624 .858	****
MAX	TW(Q)	.411 .676 .863	.263 .514 .761	.335 .619 .844	***	.300 .587 .830	.281 .567 .821
	TW(F)	.414 .678 .864	.272 .525 .770	.350 .632 .850	***	.308 .594 .834	.290 .574 .825
	HC	.404 .674 .865	.238 .471 .719	.302 .581 .819	***	.271 .557 .814	.254 .537 .805
	P(Q)	.519 .780 .924	.263 .514 .761	.340 .633 .860	***	.300 .600 .852	.279 .574 .837
	P(F)	.546 .798 .928	.272 .525 .771	.359 .655 .876	***	.310 .613 .863	.288 .587 .847
SQ	TW(Q)	.425 .688 .866	.376 .647 .849	.368 .645 .853	.266 .517 .761	.298 .562 .799	.291 .559 .797
	TW(F)	.425 .688 .866	.379 .649 .849	.369 .645 .853	.269 .517 .758	.300 .562 .799	.292 .560 .797
	HC	.587 .821 .925	.310 .583 .818	.413 .726 .918	.235 .471 .721	.265 .544 .812	.261 .528 .784
	P(Q)	.596 .823 .925	.380 .654 .856	.485 .762 .910	.267 .518 .761	.296 .583 .837	.290 .564 .809
	P(F)	.596 .823 .925	.383 .657 .857	.489 .762 .910	.269 .518 .759	.300 .587 .839	.292 .566 .811
Stand. dev. (Max.)	TW(Q)	.012	.011	.008	.004	.004	.004
	TW(F)	.012	.010	.008	.008	.005	.004
	HC	.018	.006	.010	.006	.009	.006
	P(Q)	.019	.015	.015	.004	.006	.004
	P(F)	.019	.013	.016	.008	.006	.005

* : equal to A, ** : a part of Type I error, *** : equal to B, **** : equal to E.

Table 6. Powers of each procedure corresponding to powers of Tukey's procedure: .25, .50 and .75.
($k = 4$, unbalance: (2,4,6,12))
A: all-pairs power, B: restricted all-pairs power, C: minimum difference power,
D: maximum difference power, E: mean rejective rate, F: weighted mean rejective rate.

Config.	Proc.	A .25	A .50	A .75	B .25	B .50	B .75	C .25	C .50	C .75	D .25	D .50	D .75	E .25	E .50	E .75	F .25	F .50	F .75
EQ	TW(Q)	.409	.668	.858	.349	.615	.831	.359	.639	.850	.275	.543	.794	.302	.576	.814	.294	.565	.804
	TW(F)	.409	.668	.858	.351	.620	.835	.369	.641	.849	.298	.579	.823	.314	.583	.813	.312	.578	.806
	HC	.550	.791	.921	.317	.584	.813	.312	.606	.848	.241	.497	.759	.267	.557	.826	.261	.536	.799
	P(Q)	.560	.795	.921	.367	.636	.845	.364	.655	.868	.276	.544	.795	.306	.603	.851	.294	.578	.827
	P(F)	.560	.795	.921	.390	.658	.858	.376	.658	.866	.299	.581	.824	.324	.614	.852	.316	.596	.831
MIN	TW(Q)	.384	.650	.848	*			**			.312	.597	.832	.299	.579	.821	****		
	TW(F)	.388	.653	.848	*			**			.326	.600	.825	.314	.594	.826	****		
	HC	.320	.580	.803	*			**			.277	.554	.803	.266	.539	.791	****		
	P(Q)	.468	.733	.893	*			**			.324	.619	.852	.308	.599	.842	****		
	P(F)	.476	.735	.890	*			**			.342	.624	.843	.327	.618	.849	****		
MAX	TW(Q)	.382	.646	.850	.270	.537	.788	.301	.590	.833	***			.305	.581	.817	.294	.570	.812
	TW(F)	.385	.648	.851	.298	.582	.828	.332	.618	.845	***			.323	.591	.815	.317	.585	.812
	HC	.379	.645	.849	.240	.495	.753	.264	.544	.801	***			.272	.551	.804	.262	.539	.797
	P(Q)	.488	.744	.900	.271	.537	.788	.302	.594	.839	***			.305	.594	.838	.293	.580	.829
	P(F)	.491	.744	.899	.298	.582	.829	.334	.625	.854	***			.326	.606	.836	.318	.596	.829
SQ	TW(Q)	.416	.681	.869	.375	.642	.848	.367	.639	.847	.269	.521	.765	.303	.572	.809	.292	.557	.794
	TW(F)	.416	.681	.869	.375	.642	.848	.368	.639	.847	.275	.534	.778	.307	.574	.808	.298	.568	.804
	HC	.572	.813	.931	.318	.594	.830	.396	.712	.914	.237	.478	.729	.267	.554	.823	.262	.533	.791
	P(Q)	.584	.817	.930	.383	.659	.864	.465	.754	.913	.269	.522	.765	.305	.599	.848	.295	.572	.818
	P(F)	.584	.817	.930	.383	.659	.863	.470	.755	.913	.276	.535	.778	.312	.604	.849	.300	.578	.820
Stand. dev. (Max.)	TW(Q)	.013			.013			.009			.007			.003			.003		
	TW(F)	.013			.013			.010			.008			.004			.004		
	HC	.015			.008			.011			.006			.007			.004		
	P(Q)	.015			.014			.013			.007			.005			.004		
	P(F)	.015			.014			.013			.008			.005			.004		

* : equal to A, ** : a part of Type I error, *** : equal to B, **** : equal to E.

Table 7. Powers of each procedure corresponding to powers of Tukey's procedure: .25, .50 and .75.
($k = 4$, unbalance: (10,10,2,2))
A: all-pairs power, B: restricted all-pairs power, C: minimum difference power,
D: maximum difference power, E: mean rejective rate, F: weighted mean rejective rate.

Config.	Proc.	A .25	A .50	A .75	B .25	B .50	B .75	C .25	C .50	C .75	D .25	D .50	D .75	E .25	E .50	E .75	F .25	F .50	F .75
EQ	TW(Q)	.412	.677	.864	.356	.635	.852	.370	.642	.850	.269	.531	.778	.306	.574	.806	.296	.567	.803
	TW(F)	.412	.677	.864	.372	.651	.862	.370	.642	.850	.288	.564	.809	.320	.583	.806	.314	.581	.806
	HC	.566	.799	.916	.322	.600	.832	.428	.734	.914	.237	.487	.743	.272	.555	.816	.264	.537	.797
	P(Q)	.574	.802	.916	.385	.667	.875	.488	.756	.903	.271	.532	.779	.314	.603	.844	.298	.582	.827
	P(F)	.574	.802	.916	.435	.714	.898	.488	.756	.904	.290	.567	.810	.335	.619	.845	.323	.605	.834
MIN	TW(Q)	.391	.660	.854	*			**			.296	.574	.818	.298	.580	.823	****		
	TW(F)	.390	.659	.854	*			**			.313	.590	.823	.317	.596	.827	****		
	HC	.317	.587	.814	*			**			.259	.531	.789	.264	.537	.791	****		
	P(Q)	.456	.729	.897	*			**			.305	.589	.832	.306	.596	.840	****		
	P(F)	.455	.724	.892	*			**			.329	.611	.838	.331	.618	.846	****		
MAX	TW(Q)	.421	.683	.866	.270	.539	.791	.359	.635	.848	***			.296	.576	.819	.288	.565	.811
	TW(F)	.423	.683	.865	.290	.568	.812	.357	.635	.850	***			.308	.580	.816	.303	.572	.809
	HC	.416	.680	.866	.236	.495	.757	.332	.622	.850	***			.266	.546	.805	.258	.533	.796
	P(Q)	.564	.790	.912	.271	.539	.790	.435	.722	.900	***			.300	.597	.848	.289	.579	.835
	P(F)	.569	.792	.911	.292	.569	.812	.433	.723	.902	***			.313	.603	.846	.306	.588	.833
SQ	TW(Q)	.398	.652	.848	.392	.662	.860	.365	.636	.848	.286	.570	.821	.293	.550	.788	.305	.564	.789
	TW(F)	.398	.652	.848	.398	.666	.860	.365	.636	.848	.325	.616	.850	.309	.558	.786	.336	.587	.792
	HC	.540	.772	.905	.318	.586	.817	.458	.745	.910	.253	.524	.784	.262	.535	.800	.269	.531	.777
	P(Q)	.544	.772	.905	.400	.676	.874	.483	.748	.907	.286	.570	.821	.291	.567	.819	.306	.570	.799
	P(F)	.545	.772	.905	.408	.681	.875	.483	.748	.907	.326	.618	.851	.313	.582	.821	.340	.598	.806
Stand. dev. (Max.)	TW(Q)	.015			.011			.010			.006			.006			.005		
	TW(F)	.015			.012			.010			.008			.008			.005		
	HC	.018			.012			.013			.007			.009			.005		
	P(Q)	.026			.013			.016			.008			.009			.005		
	P(F)	.026			.014			.016			.009			.010			.005		

* : equal to A, ** : a part of Type I error, *** : equal to B, **** : equal to E.

Table 8. Practical significant level of each procedure (nominal level: 5%).

Sample size	Variance	Practical level			a
		C	GH	GHC	
2,2,2	1,1,1	0.007	0.027	0.012	−0.1387
2,2,2	1,1,3	0.007	0.029	0.011	−0.1179
4,4,2	1,1,1	0.019	0.057	0.027	0.0753
4,4,2	1,1,3	0.026	0.083	0.042	0.4307
4,4,2	3,3,1	0.015	0.042	0.022	−0.0816
4,4,2	1,1,10	0.039	0.101	0.055	1.7903
8,8,4	1,1,1	0.031	0.059	0.042	0.2489
8,8,4	1,1,10	0.035	0.059	0.045	0.3559
4,4,4,2	1,1,1,3	0.027	0.084	0.041	0.4752
4,4,4,2	1,1,1,10	0.035	0.103	0.055	1.2006

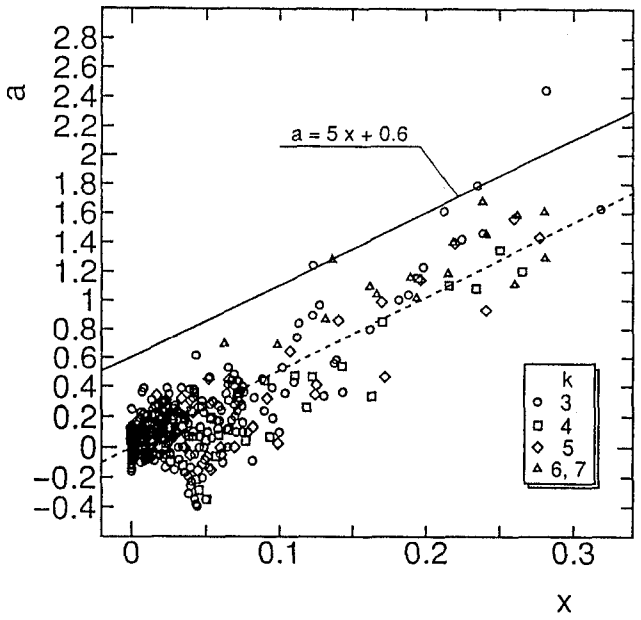

Fig. 1. Prediction of rate for GHC2–procedure.

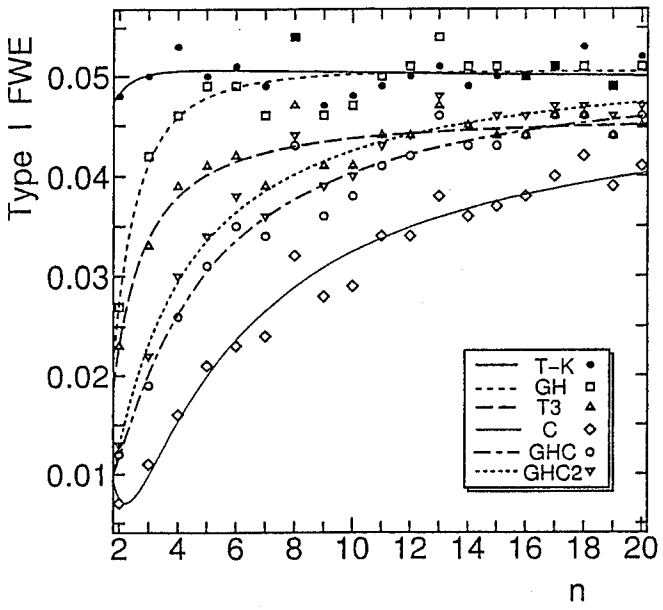

Fig. 2. Type I FWE.
Sample size (n,n,n), Variance (1,1,1)

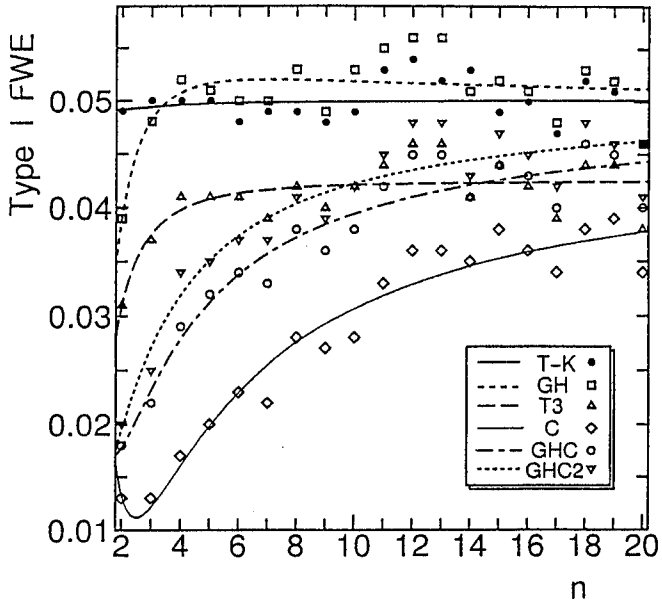

Fig. 3. Type I FWE.
Sample size (n,n,n,n), Variance (1,1,1,1)

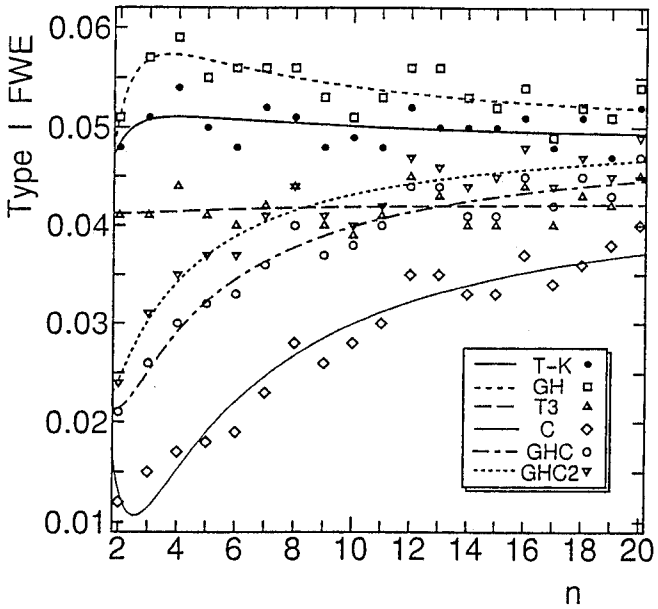

Fig. 4. Type I FWE.
Sample size (n,n,n,n,n), Variance (1,1,1,1,1)

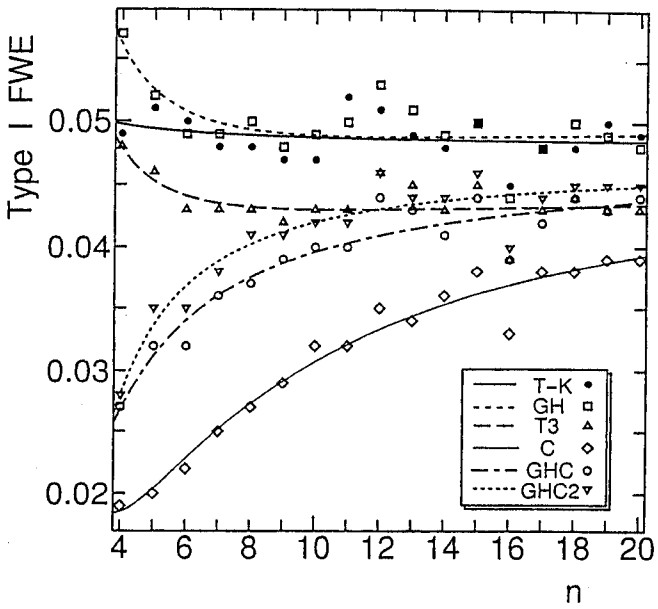

Fig. 5. Type I FWE.
Sample size (n,n,n-2), Variance (1,1,1)

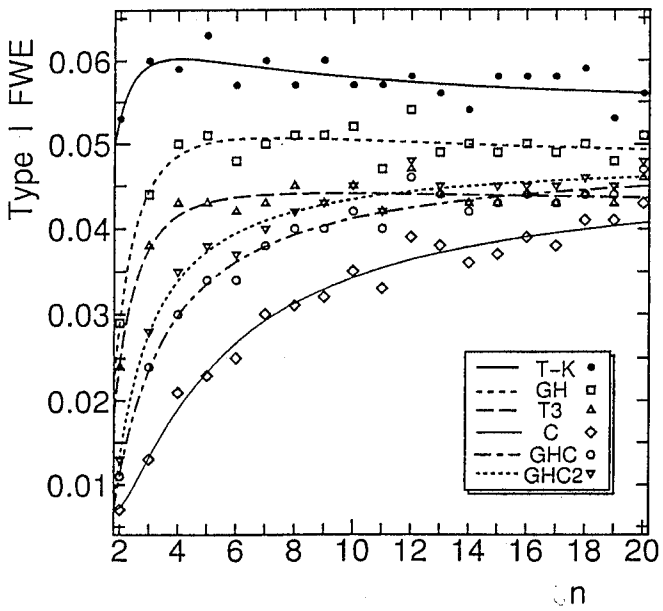

Fig. 6. Type I FWE.
Sample size (n,n,n), Variance (1,1,3)

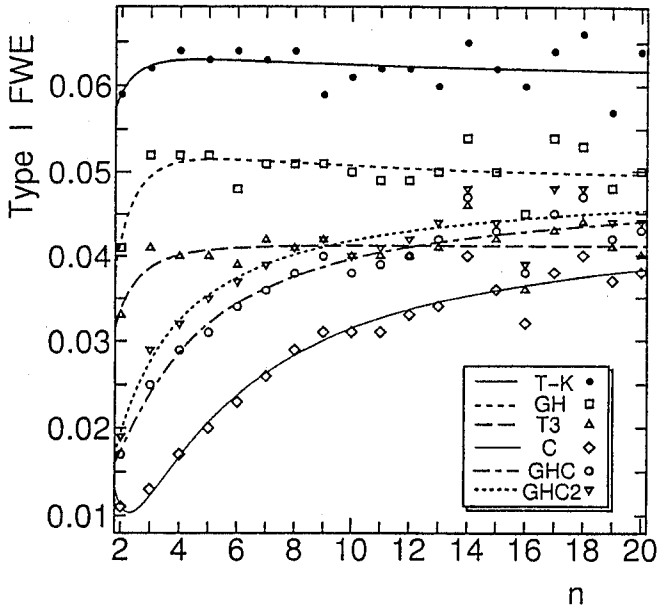

Fig. 7. Type I FWE.
Sample size (n,n,n,n), Variance (1,1,1,3)

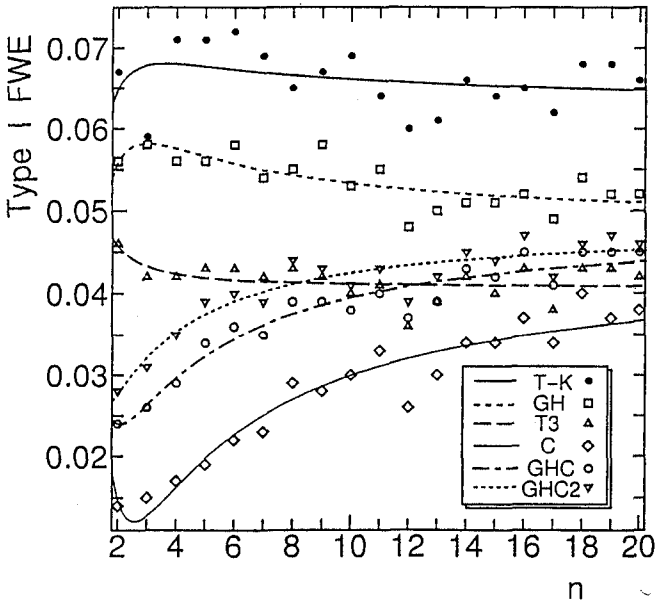

Fig. 8. Type I FWE.
Sample size (n,n,n,n,n), Variance (1,1,1,1,3)

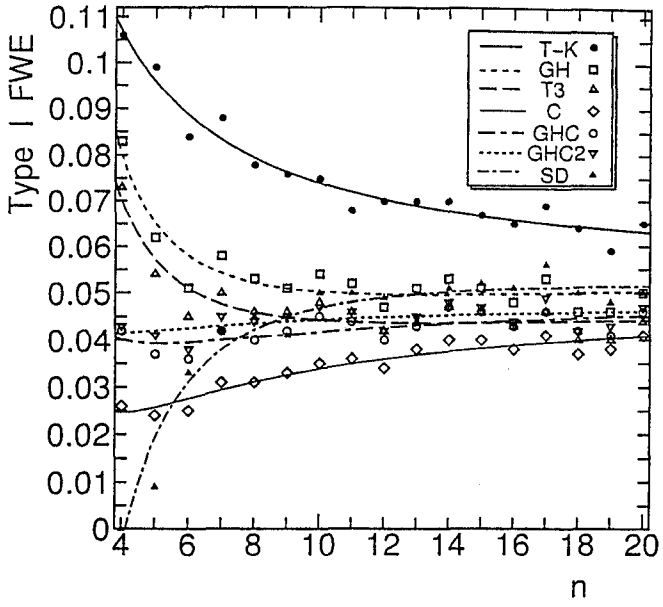

Fig. 9. Type I FWE.
Sample size (n,n,n-2), Variance (1,1,3)

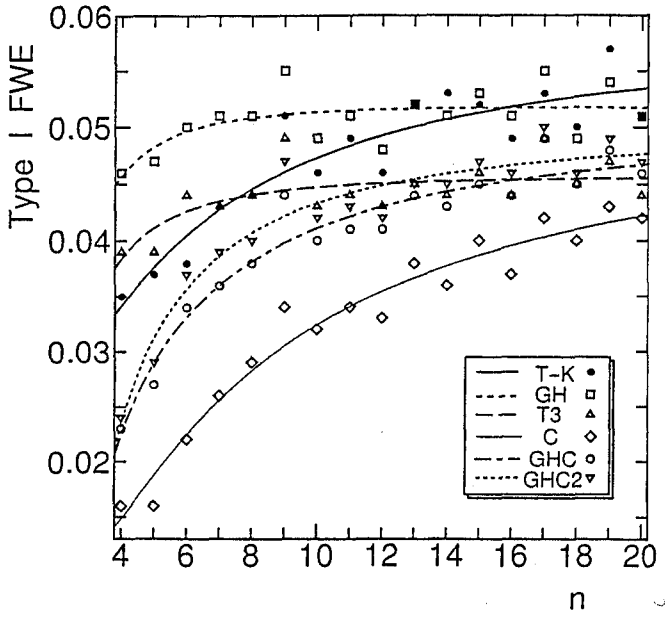

Fig. 10. Type I FWE.
Sample size (n,n,n-2), Variance (3,3,1)

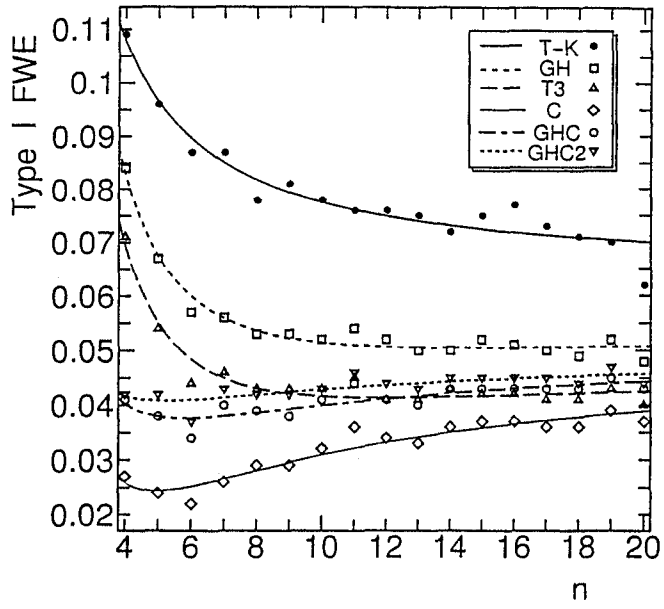

Fig. 11. Type I FWE.
Sample size (n,n,n,n-2), Variance (1,1,1,3)

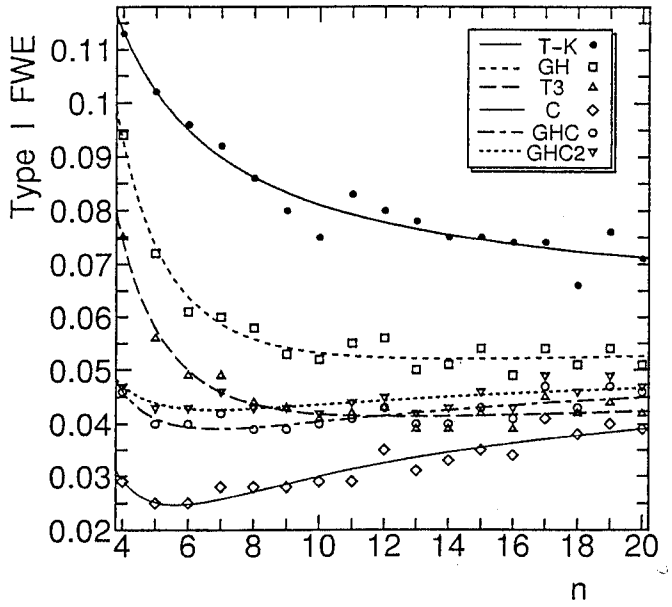

Fig. 12. Type I FWE.
Sample size (n,n,n,n,n−2), Variance (1,1,1,1,3)

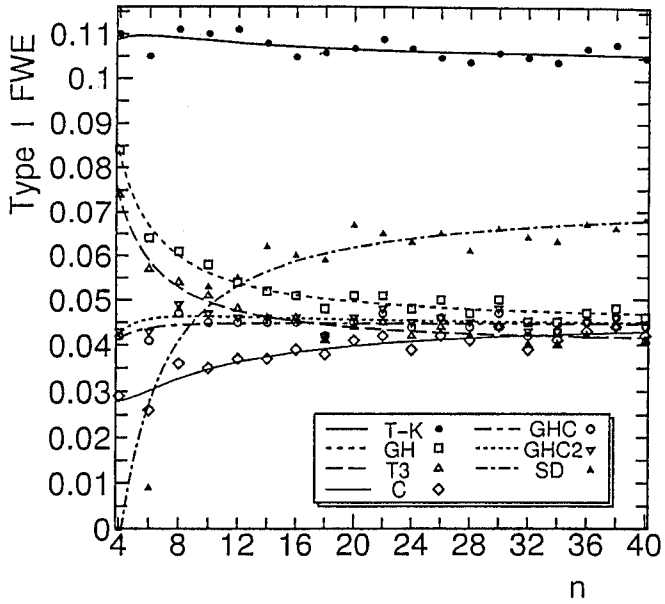

Fig. 13. Type I FWE.
Sample size (n,n,n/2), Variance (1,1,3)

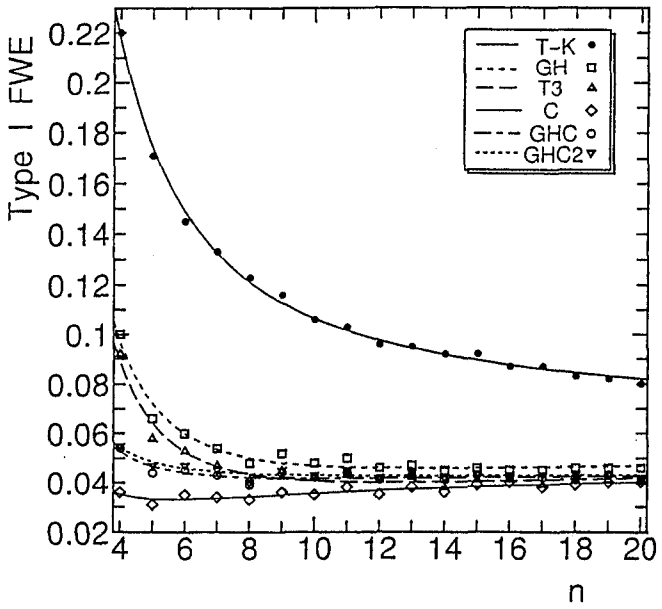

Fig. 14. Type I FWE.
Sample size (n,n,n−2), Variance (1,1,10)

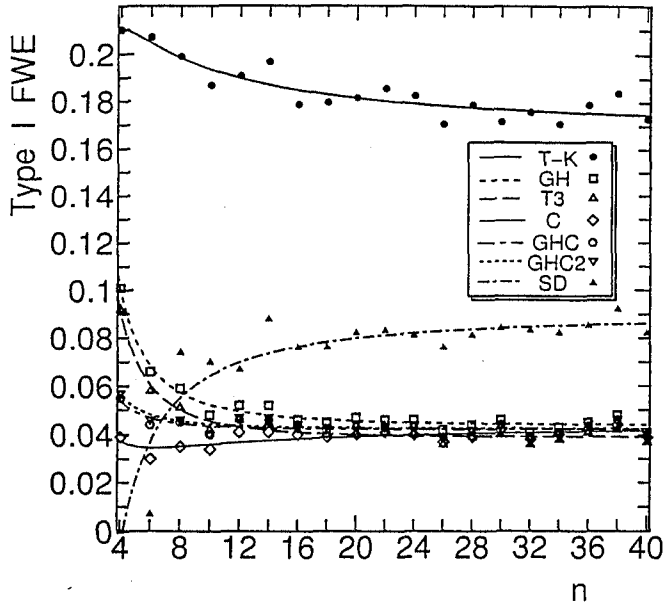

Fig. 15. Type I FWE.
Sample size (n,n,n/2), Variance (1,1,10)

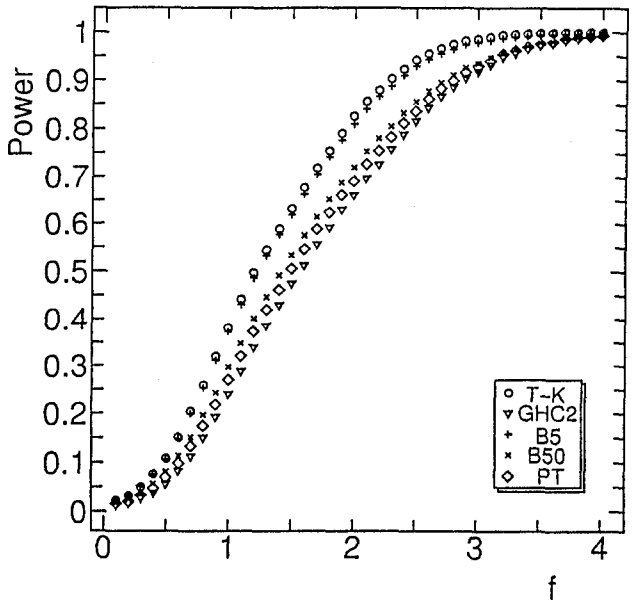

Fig. 16. Mean rejective rate. [EQ]
Sample size (4,4,4), Variance (1,1,1)

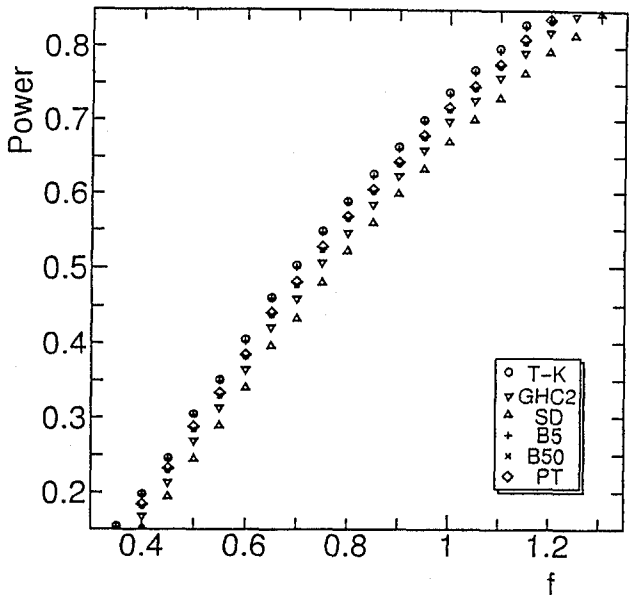

Fig. 17. Mean rejective rate. [EQ]
Sample size (10,10,10), Variance (1,1,1)

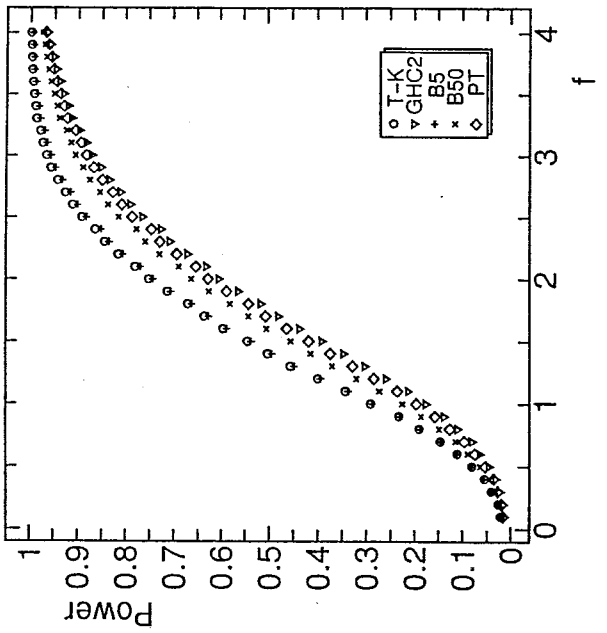

Fig. 19. Mean rejective rate. [EQ]
Sample size (5,5,3), Variance (2,1,1)

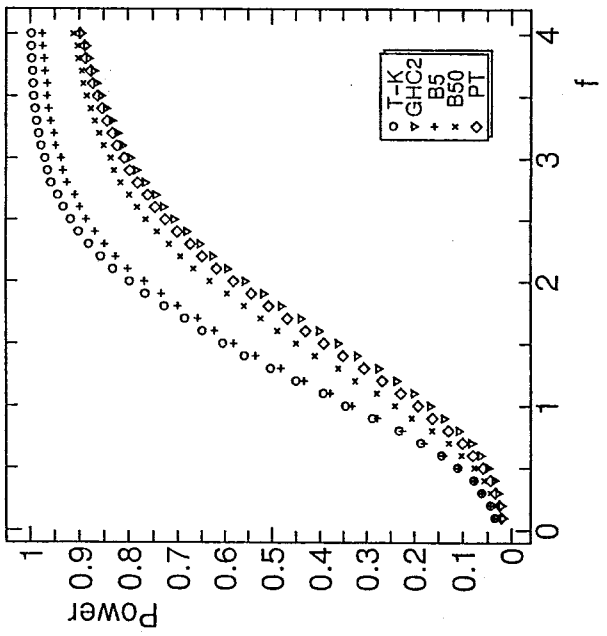

Fig. 18. Mean rejective rate. [EQ]
Sample size (5,5,3), Variance (1,1,2)

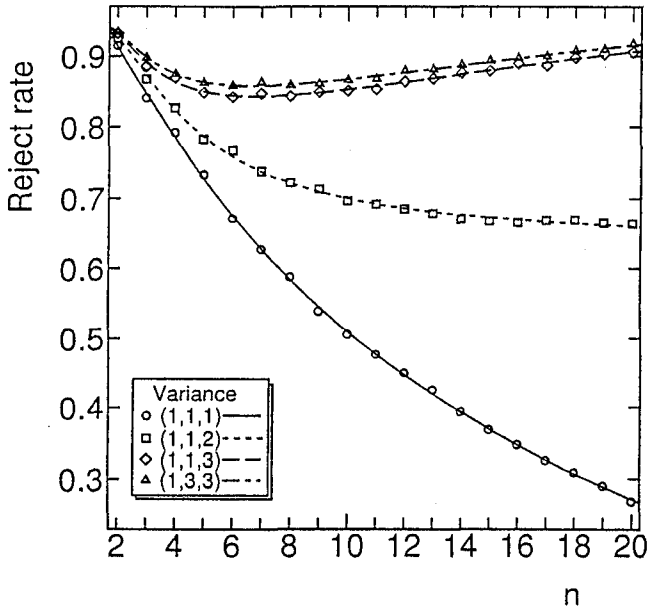

Fig. 20. Rejective rates of the new preliminary test. Sample size (n, n, n)

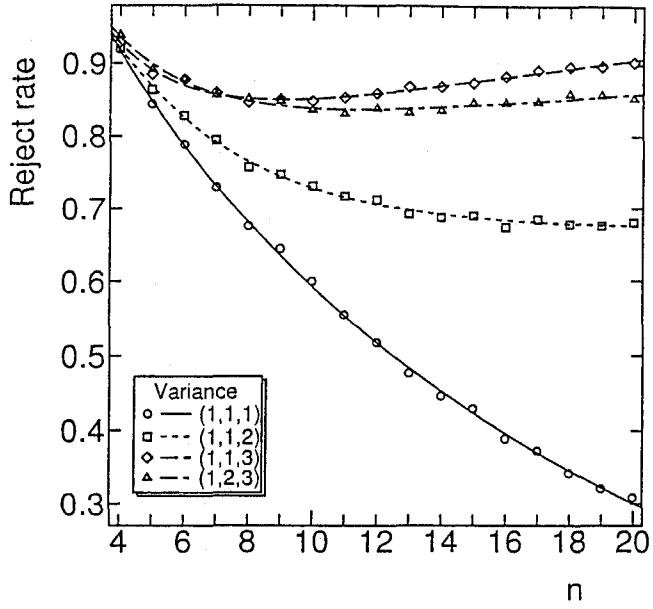

Fig. 21. Rejective rates of the new preliminary test. Sample size (n, n, n–2)

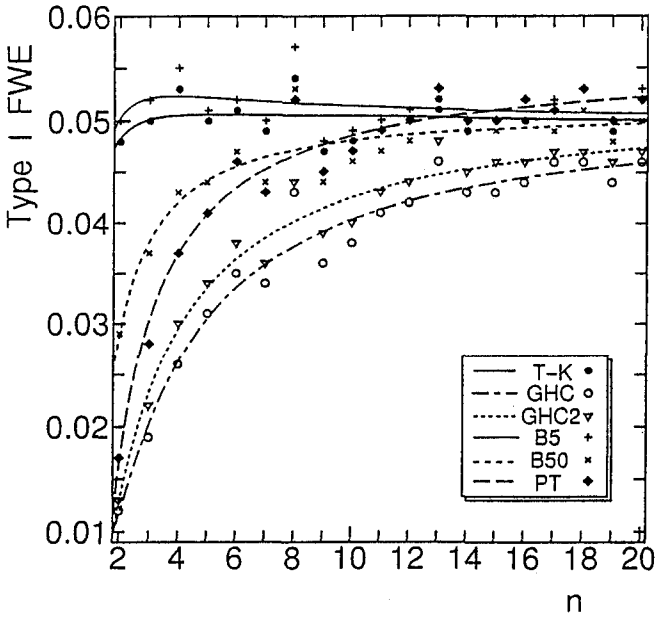

Fig. 22. Type I FWE with preliminary tests. Sample size (n,n,n), Variance (1,1,1)

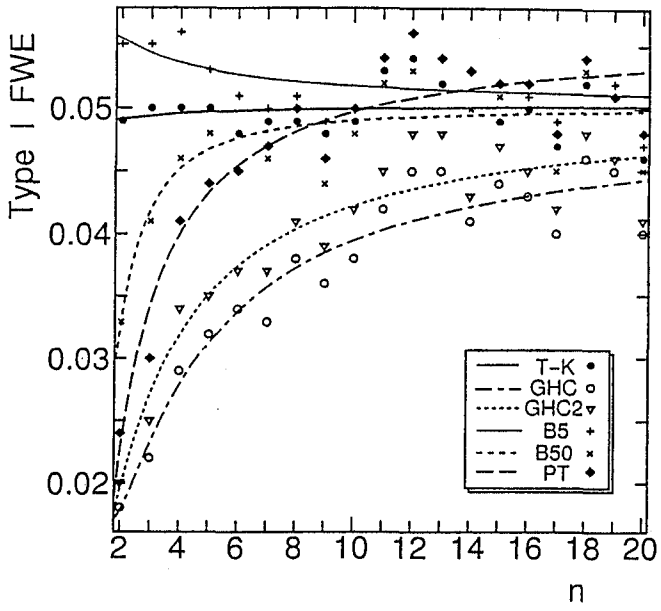

Fig. 23. Type I FWE with preliminary tests Sample size (n,n,n,n), Variance (1,1,1,1)

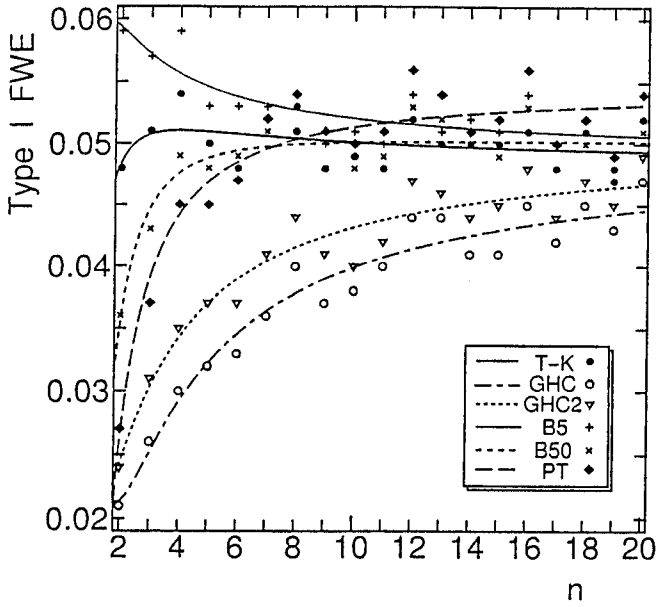

Fig. 24. Type I FWE with preliminary tests. Sample size (n,n,n,n,n), Variance (1,1,1,1,1)

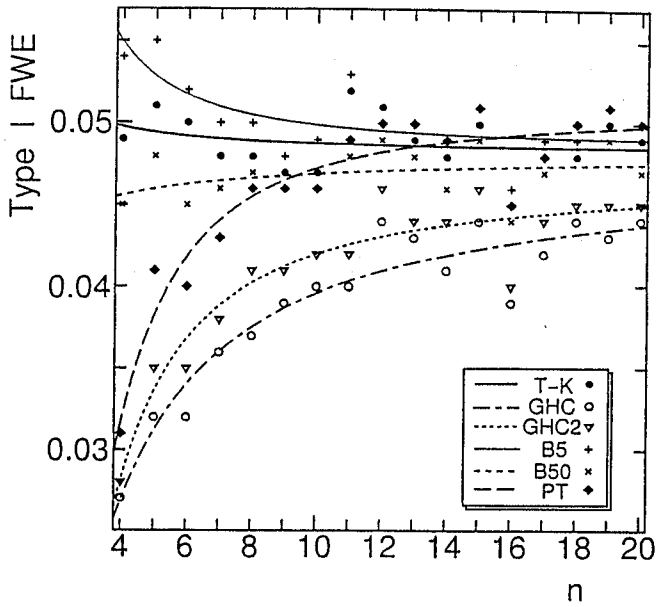

Fig. 25. Type I FWE with preliminary tests Sample size (n,n,n-2), Variance (1,1,1)

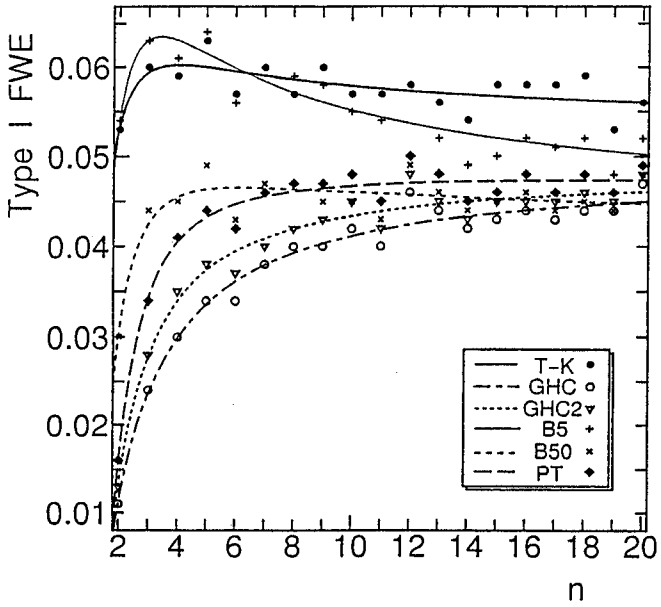

Fig. 26. Type I FWE with preliminary tests. Sample size (n,n,n), Variance (1,1,3)

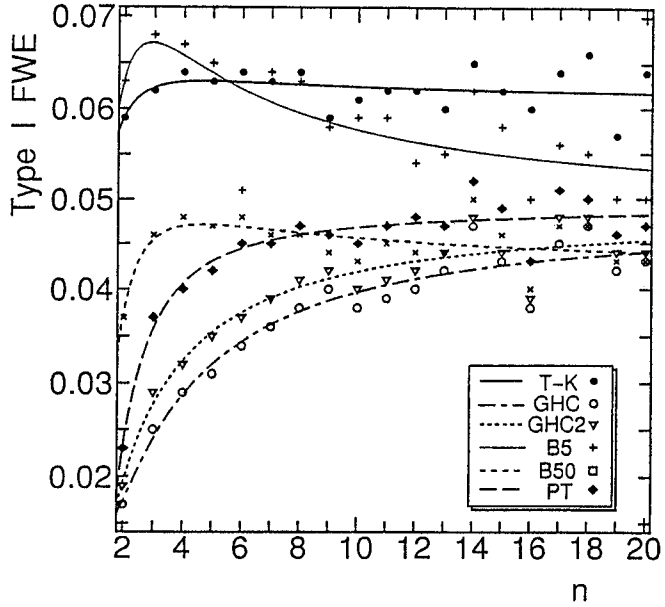

Fig. 27. Type I FWE with preliminary tests Sample size (n,n,n,n), Variance (1,1,1,3)

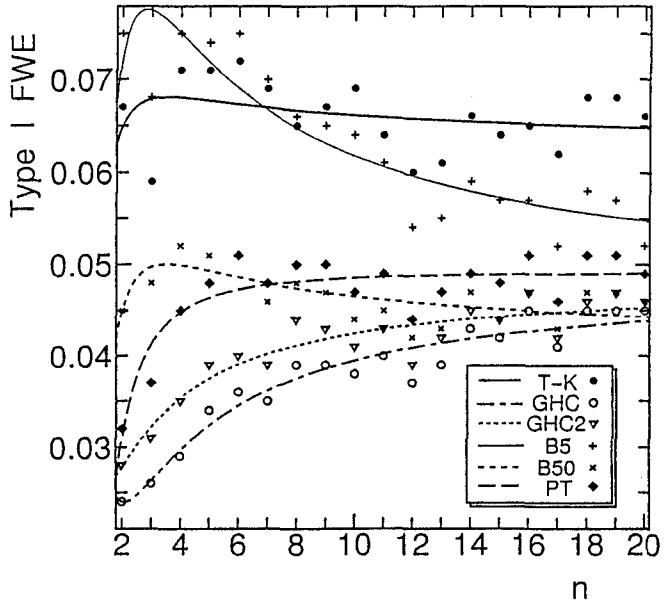

Fig. 28. Type I FWE with preliminary tests. Sample size (n,n,n,n,n), Variance (1,1,1,1,3)

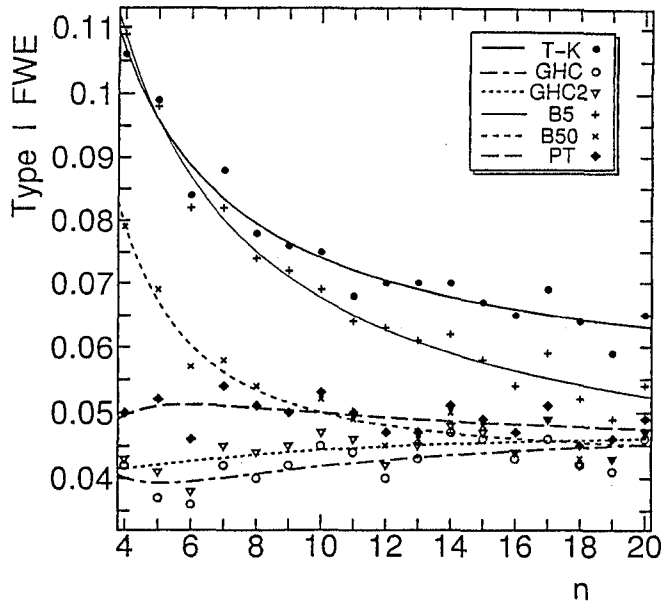

Fig. 29. Type I FWE with preliminary tests Sample size (n,n,n-2), Variance (1,1,3)

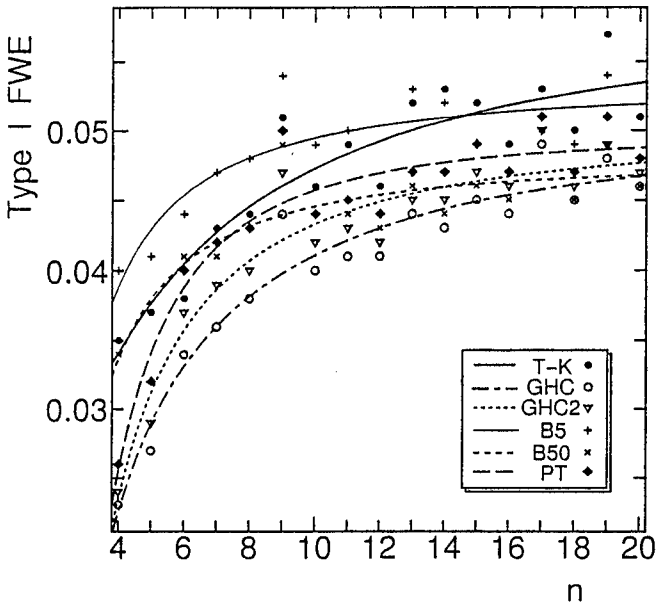

Fig. 30. Type I FWE with preliminary tests. Sample size (n,n,n-2), Variance (3,3,1)

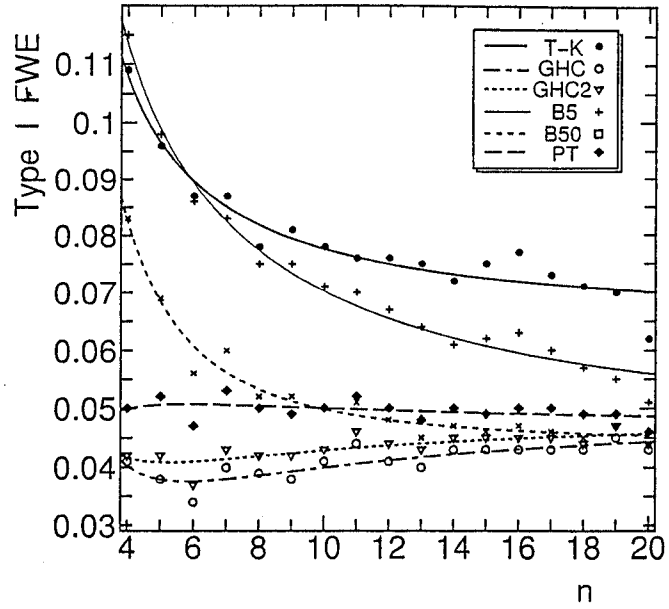

Fig. 31. Type I FWE with preliminary tests Sample size (n,n,n,n-2), Variance (1,1,1,3)

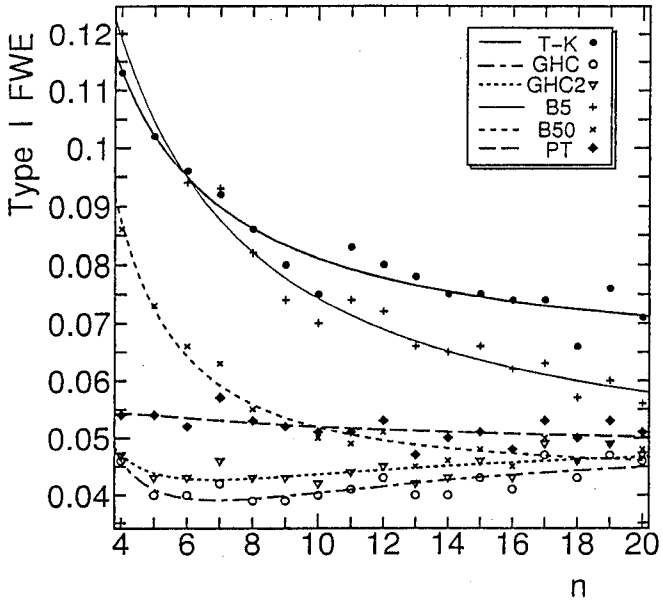

Fig. 32. Type I FWE with preliminary tests. Sample size (n,n,n,n,n-2), Variance (1,1,1,1,3)

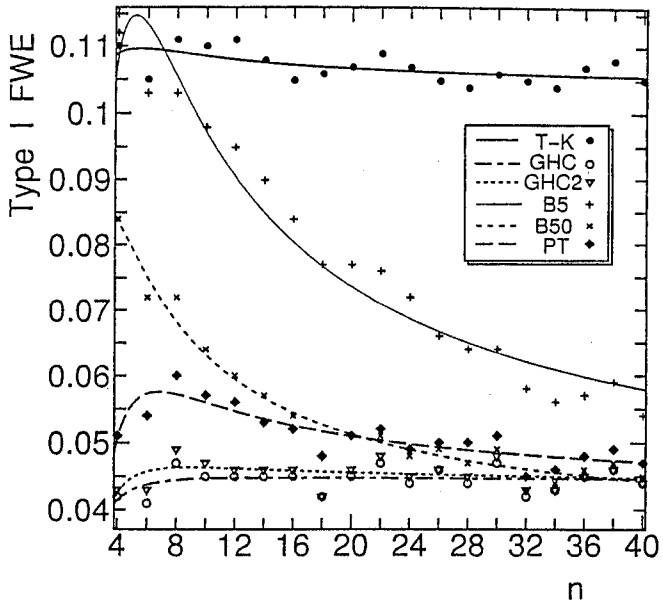

Fig. 33. Type I FWE with preliminary tests. Sample size (n,n,n/2), Variance (1,1,3)

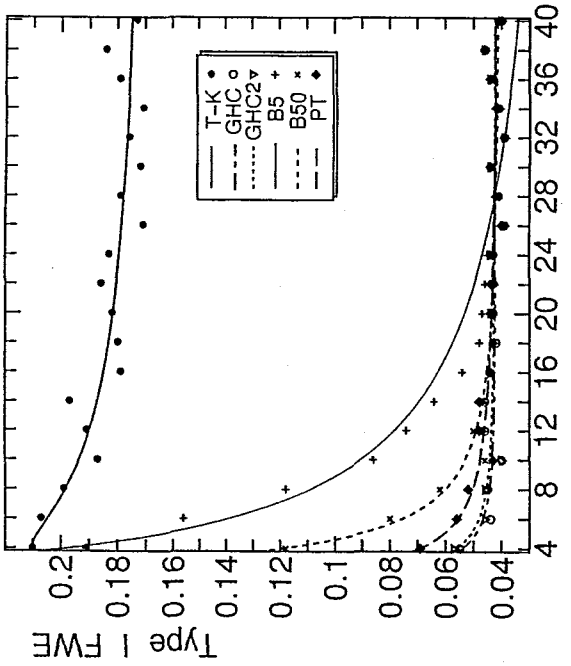

Fig. 35. Type I FWE with preliminary tests
Sample size (n,n,n/2), Variance (1,1,10)

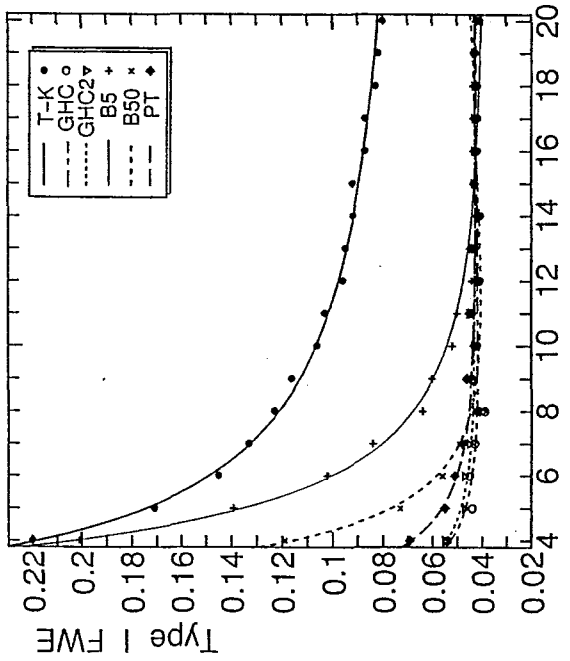

Fig. 34. Type I FWE with preliminary tests.
Sample size (n,n,n-2), Variance (1,1,10)

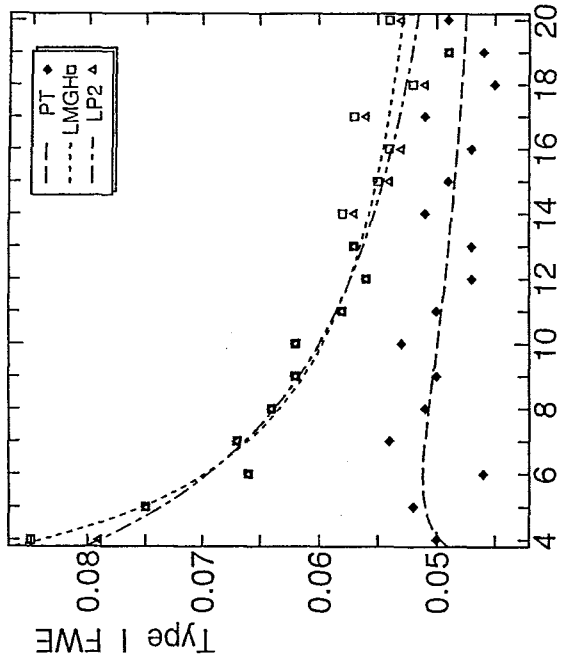

Fig. 37. Type I FWE for LMCP's.
Sample size (n,n,n−2), Variance (1,1,3)

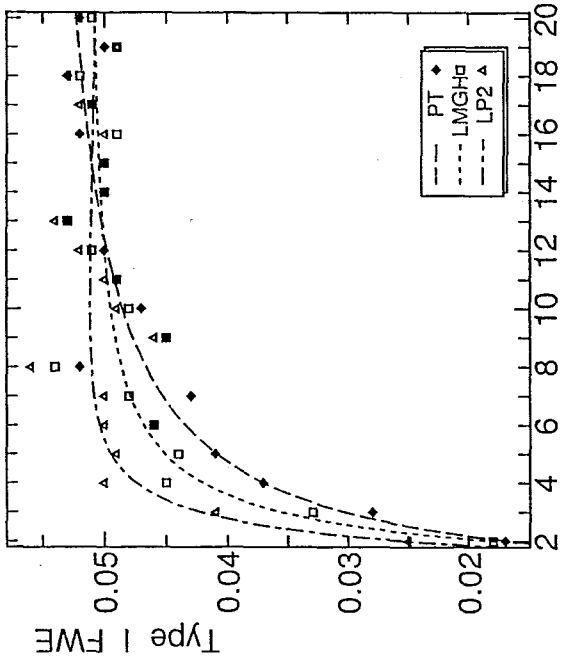

Fig. 36. Type I FWE for LMCP's.
Sample size (n,n,n), Variance (1,1,1)

Acknowledgements

The author is indebted to Dr. M. Okamoto, Prof. S. Shirahata of Osaka University, Prof. I. Yoshimura of Tokyo Science University, Dr. Y. Nagata of Okayama University and Dr. M. Ohtaki of Hiroshima University for their advice.

Dr. Okamoto and Dr. Shirahata taught the basis of statistics for me in my graduate days of Osaka University. I thank them for their careful direction. Furthermore, Dr. Shirahata read the manuscript of this thesis for me. I appreciate his assistance.

I am grateful to Dr. Yoshimura for many useful comments for this research in Nagoya University. And, I also thank other members (Dr. K. Nishina, Dr. H. Yokoi, Mr. Y. Sakamoto and Dr. H. Tanaka) of the statistical research group in Nagoya.

I studied with Dr. Nagata for the research in Chapter 3. I appreciate his help.

I thank Dr. Ohtaki for a useful comment for the research in Chapter 4.

Finally, I am grateful to Prof. M. Kimura for his kindly support in Nanzan University.

References

Begun, J. and Gabriel, K. R. (1981): Closure of the Newman-Keuls multiple comparisons procedure, *J. Amer. Statist. Assoc.*, **76**, 241-245.

David, H. A., Lachenbruch, P. A. and Brandis, H. P. (1972): The power function of range and Studentized range tests in normal samples, *Biometrika*, **59**, 161-168.

Dunnett, C.W. (1955): A multiple comparison procedure for comparing several treatments with a control, *J. Amer. Statist. Assoc.*, **50**, 1096-1121.

Dunnett, C.W. (1980): Pairwise multiple comparisons in the unequal variance case, *J. Amer. Statist. Assoc.*, **75**, 796-800.

Dwass, M. (1960): Some k-sample rank-order tests, *Contribution to Probability and Statistics*, Standord: Stanford University Press, 198-202.

Einot, I. and Gabriel, K. R. (1975): A study of the power of several method in multiple comparisons, *J. Amer. Statist. Assoc.*, **70**, 574-583.

Fisher, R. A. (1935): *The Design of Experiments*, Edinburgh and London: Oliver & Boyd

Games, P.A. and Howell, J.F. (1976): Pairwise multiple comparison procedures with unequal N's and/or variances: A Monte Carlo study, *J. Educ. Statist.*, **1**, 113-125.

Hayter, A. J. (1984): A proof of the conjecture that the Tukey-Kramer multiple comparisons procedure is conservative, *Annals of Statistics*, **12**, 61-75.

Hochberg, Y. and Tamhane, A.C. (1987): *Multiple Comparison Procedures*, John Wiley, New York.

Holland, B. S. and Copenhaver, M. D. (1987): An improved sequentially rejective Bonferroni test procedure, *Biometrics*, **43**, 417-423.

Holm, S. (1979): A simple sequentially rejective multiple test procedure, *Scand. J. Statist.*, **6**, 65-70.

Kendall, M.G. and Stuart, A. (1979): *The Advanced Theory of Statistics*, Vol. 2 *Inference and Relationship*, 4th ed., Charles Griffin, London.

Keuls, M. (1952): The use of the 'Studentized range' in connection with an analysis of variance, *Euphytica*, **1**, 112-122.

Knuth, D. E. (1981): *The Art of Computer Programming*, Vol. 2 *Seminumerical Algorithms*, 2nd ed., Addison-Wesley, U.S.A..

Marcus, R., Peritz, E. and Gabriel, K. R. (1976): On closed testing procedures with special reference to ordered analysis of variance, *Biometrika*, 63, 655-660.

Matsuda, S. (1988): Nonparametric T-method in two-way layouts, *J. Japan Statist. Soc.*, **18**, 149-155.

Matsuda, S. (1991): Multiple comparison procedures on non-homogeneity, *A Study of Methods of Toxicological Data Analysis for the Risk Assessment* '*The Institute of Statistical Mathematics Cooperative Research Report* **27**', 41-50 (in Japanese).

Matsuda, S. (1993): Multiple comparison procedures for one-way layouts with unequal variances, *Nanzan Management Review*, **7**, 397-413 (in Japanese).

Matsuda, S. (1994): An improvement of the preliminary test of multiple comparisons for one-way layout, *Japanese J. Applied Statist.*, **23**, 129-145 (in Japanese).

Matsuda, S. (1997): Multiple comparison procedures based on a loss function, To appear *Nanzan Management Review*, **11** (in Japanese).

Matsuda, S., Fujimoto, T. and Yoshimura, I. (1990): A robust quadratic discriminant function using a shrinkage estimator of variance matrix, *Japanese J. Applied Statist.*, **19**, 33-51 (in Japanese).

Matsuda, S. and Nagata, Y. (1990): Definition of powers in multiple comparisons, and features of several procedures based on them, *Japanese J. Applied Statist.*, **19**, 93-113 (in Japanese).

Mehta, J.S. and Srinivasan, R. (1970): On the Behrens-Fisher problem, *Biometrika*, **57**, 649-655.

Nagata, Y. (1992): *Nyumon Tokei Kaiseki Ho (Introduction: Statistical analysis methods)*", Japanese Union of Scientist and Engineer (In Japanese).

Newman, D. (1939): The distribution of the range in samples from a normal population, expressed in terms of an independent estimate of standard deviation, *Biometrika*, **31**, 20-30.

Patnaik, P.B. (1949): The non-central χ^2- and F-Distributions and their applications, *Biometrika*, **36**, 202-232.

Peritz, E (1970): A note on multiple comparisons, Unpublished manuscript, Hebrew University, Israel.

Ramsey, P. H. (1978): Power differences between pairwise multiple comparisons, *J. Amer. Statist. Assoc.*, **73**, 479-485.

Scheffé, H. (1953): A method for judging all contrasts in the analysis of variance, *Biometrika*, **40**, 87-104.

Sen, P. K. (1969): On nonparametric T-method of multiple comparisons in randomized blocks, *Ann. Inst. Statist. Math.*, **21**, 329-333.

Shaffer, J. P. (1986): Modified sequentially rejective multiple test procedures, *J. Amer. Statist. Assoc.*, **81**, 826-831.

Shirley, E. A. (1977): A nonparametric equivalent of Williams' test for contrasting increasing dose levels of a treatment, *Biometrics*, **33**, 388-389.

Šidák, Z. (1967): Rectangular confidence regions for the means of multivariate normal distributions, *J. Amer. Statist. Assoc.*, **62**, 626-633.

Steel, R. G. D. (1959): A multiple comparison rank sum test: Treatments versus control, *Biometrics*, **15**, 560-572.

Steel, R. G. D. (1960): A rank sum test for comparing all pairs of treatments, *Technometrics*, **2**, 197-207.

Toda, H. (1967): An optimal rational approximation for normal deviates for digital computers, *Bull. Electrotech. Lab.*, **31**, 1259-1270.

Tsubaki, H. (1989): *Various Problems in Multiple Comparisons* '*The Institute of Statistical Mathematics Cooperative Research Report* **18**', §7 (in Japanese).

Tukey, J. W. (1953): *The Problem of Multiple Comparisons*, Mimeographed monograph.

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.

Yamauti, Z. (1972): *Statistical Tables and Formulas with Computer Application*, Japanese Standards Association (in Japanese).

Yoshida, M. (1988): Exact probabilities associated with Tukey's and Dunnett's multiple comparisons procedures in imbalanced one-way ANOVA, *J. Japanese Soc. Comp. Statist.*, **1**, 111-122.

Yoshida, M. (1989): Tukey's multiple comparisons procedure in imbalanced one-way ANOVA: Evaluation of Tukey-Kramer's approximate method based on the exact calculation, *Japanese J. Soc. Comp. Statist.*, **2**, 17-24 (in Japanese).

Yoshimura, I. (1987): *Dokusei · Yakko data No Tokei Kaiseki (Statistical Analysis of Toxicological and Pharmacological Data)*, Scientist Co. (in Japanese).

Yoshimura, I. (1989): *Various Problems in Multiple Comparisons* '*The Institute of Statistical Mathematics Cooperative Research Report* **18**', §1 (in Japanese).