Multiple Comparisons with a Control in Two-way Designs and Directional-mixed Families

LEE HANG TING

A Thesis Submitted in Partial Fulfillment

of the Requirements for the Degree of

Master of Philosophy

in

Statistics

June 2004

The Chinese University of Hong Kong
Graduate School

The undersigned certify that we have read a thesis, entitled "Multiple Comparisons with a Control in Two-way Designs and Directional-mixed Families" submitted to the Graduate School by Lee Hang Ting ( ) in partial fulfillment of the requirements for the degree of Master of Philosophy in Statistics. We recommend that it be accepted.

Prof. S.H. Cheung

Supervisor

Prof. K.H. Wu

Prof. W.Y. Poon

Prof. W.S. Chan

External Examiner

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institution of learning.

Acknowledgments

I would like to express my most sincere thanks to my supervisor, Professor Cheung Siu Hung, for his kind assistance and helpful suggestions to my thesis. Also I want to thank him for his encouragement and patience.

Abstract

Multiple Comparison Procedures are powerful tools for comparing different treatments in an experiment. It is widely used in medicine, agriculture, etc. In many circumstances, it is desirable to include a control in the experiment. Thus we can compare whether a certain treatment is better than the control treatment. Cheung and Holland (91, *Biometrics*; 94, *Statistics in Medicine*) extended the well-known Dunnett procedure for simultaneous comparisons of the means of all active treatments with the control to the case where one wishes to conduct such comparisons simultaneously in each of several groups. However, their procedures mainly deal with particular families of inferences in which all hypotheses are either one-sided or two-sided. In this thesis, we seek to develop a procedure, which copes with a more general testing environment in which the family of inferences is composed of a mixture of one-sided and two-sided hypotheses. Required critical values are tabulated for the implementation of the proposed procedure. A simulation study of average power is conducted to show the superiority of the proposed method as compared to existing methods. Finally, our proposed procedure is illustrated with a numerical example.

摘要

多重比較程序 (Multiple Comparison Procedures) 是用在實驗中比較不同處理 (treatments) 時的有力工具。它被廣泛應用在醫學、農業等範疇上。在很多情況下，實驗中會希望包括對照處理。因此，我們可以比較某一處理是否比對照處理好。Cheung and Holland (91, *Biometrics*; 94, *Statistics in Medicine*) 把同時比較多項處理與對照處理的著名鄧內特程序 (Dunnett procedure) 擴展至多組情況，即是在每一組同時比較多項處理與對照處理。但是，他們的程序主要應用在所有檢驗假設是單側 (one-sided) 或是所有假設是雙側 (two-sided) 的推論族。本論文尋求發展一個可以應用在更概括的測試環境的新程序，這環境是同時包含單側 (one-sided) 和雙側 (two-sided) 檢驗假設。這個程序中所需要的臨界值 (critical values) 會被列成表。我們會用一個模擬研究 (simulation study) 來顯示我們建議的程序比現存方法優勝。最後，我們會用一個實例來闡明我們的建議方法。

Contents

1. Introduction

1.1 Multiple comparison procedures (MCPs)

In many scientific fields including sociology, psychology, engineering, and medicine, precise statistical data analysis is required. A frequent encountered problem that arises in research studies is 'extensive' data analysis (Westfall and Young, 1993). The advancement of computing equipment and technology enables us to analyze data from every possible angle. Recently, with the aid of 'data mining' techniques, multiple testing environments become more popular. 'Multiple testing' refers to any situation in which a collection of statistical tests is formally or informally evaluated. Now, it is less usual that only one test is performed on a dataset. Therefore, an original single test of control versus active treatment becomes several tests, such as control versus treated for females, for males, for young, for old, for young females, for old males, and so on (Westfall and Young, 1993).

There are many practical situations where multiple testing is required, but the most alarming problem related to multiple testing is the multiplicity effect that some may not be aware of. Multiplicity effect refers to senerio when a large dataset is under extensive data splitting, the probability of false significance is being inflated to an unacceptable level.

To provide an illustration, let us consider the following 'simulated' example given by Westfall and Young (1993). Researchers intended to test whether eating oat bran can reduce cholesterol level. It was a multicenter study. At each of ten selected study centers, a simple control versus treated experiment was performed with 20

subjects per group. The data were randomly generated using the same mean for treatment and control groups within each center, although the mean values were allowed to differ from center to center. Therefore, no significant difference was expected between treatment and control groups. However, the results indicated that in centers 2 and 3 blood cholesterol for the treatment group was significantly ($p < 0.05$) lower. Although this example is artifical, with simulated data, it is straightforward to point to multiplicity as the cause of the small p-values which lead to false significance. Consider the case of a single test with $\alpha = 0.05$, the probability of declaring the result significant is 0.05. But, with ten independent tests (each tested with α=0.05), the probability of declaring at least one of the ten independent results significant is $1 - (1 - 0.05)^{10} = 0.401$ which is drastically larger than 0.05.

To tackle the problem of multiplicity, one has to consider multiple comparison techniques. Multiple comparison procedures (MCPs) are defined by Hochberg and Tamhane (1987) as methods 'designed to take into account and properly control the multiplicity effect through some combined or joint measures of erroneous inferences'.

1.2 Multiple comparisons with a control (MCC)

There are various types of multiple comparison procedures, classified according to the mode of the construction of the inferential family. The most common types include all-contrast comparisons, pairwise comparisons, multiple comparisons with a control, and multiple comparisons with the best.

In clinical studies, multiple comparisons with a control is very popular. For instance, in a medical trial, a control group may represent patients treated with an accepted standard therapy or without any therapy (placebo), while other treatment

groups consist of patients treated with several alternative therapies. In designing an experiment to measure the effects of such treatments, it is desirable to include in the experiment a control in the form of either a dummy treatment to measure the magnitude of experimental response in the absence of the treatments under investigation or some recognized standard treatments.

Multiple comparisons with a control can be either one-sided or two-sided, depending on the experiment objective and the prior knowledge of the efficacies of the treatments. Both one-sided and two-sided testing methods were first given by Dunnett (1955). These procedures are single-step procedures. An efficient algorithm to compute the required critical values is given in Dunnett (1989). Compared to the Dunnett (1955) single-step procedure, a more powerful stepdown procedure comparing treatments and a control is provided by Dunnett and Tamhane (1991). The step-up version is later derived by Dunnett and Tamhane (1992). Even though stepwise procedures are normally more powerful, the means to obtain joint confidence intervals are in general very complicated.

1.3 MCC in two-way designs

The Dunnett (1955) procedure was tailored for single-group situation. However, in many experimental settings, often encountered in medical and biological researches, treatments have to be compared with a control in each of several existing groups.

Consider an example given by Cheung and Holland (1994), the study explored the intra-erythrocytic cation metabolism in uraemic patients with different dialysis treatments. The patients were randomly divided into three groups. The first two groups underwent two different treatments, regular haemodialysis and continuous ambulatory peritoneal dialysis (CAPD) respectively. And the last group, composed

of 67 persons were selected as controls. Besides, the subjects were classified into two groups according to their gender. Blood are drawn from each subject and one of the responses measured was the haemoglobin content measured in g/I. So, this study is a two-way design with treatments as one factor and gender as the second factor.

As mentioned by Cheung and Holland (1994), when one wishes to undertake the treatment versus control tests in a multi-group situation, it is worth considering whether the investigator should control familywise type I error: (a) for active treatments versus control averaged over all groups via the original Dunnett procedure; (b) for active treatments versus control separately in each group via the original Dunnett procedure; or (c) globally across all groups.

Approach (a) is appropriate when there is no interaction between treatments and groups. When interaction is present, the choice between (b) and (c) depends on the experimental setting and goals. Consider the above example, from the perspective of a particular patient, the relevant family of hypotheses probably consists only of these relating to the patient's gender. But from the perspective of a medical researcher who is equally interested in treatment efficacy for both genders, the relevant family may consists of all treatments versus control hypotheses within either gender.

The Dunnett (1955) procedure is a single-step procedure. A stepwise version is given by Dunnett and Tamhane (1991). Cheung and Holland (1991,1992) derived a single-step extension of the Dunnett (1955) procedure for a two-way design. Later, the stepwise procedure was also given by Cheung and Holland (1994).

1.4 Directional-mixed families

The Dunnett (1955) procedure is a well-known and widely used approach for conducting multiple comparisons of all active treatments with a control while a designated familywise type I error rate α is controlled. But, the procedure only dealt with situations that when comparing all active treatments with a control in a particular family of inferences, all hypotheses are either one-sided or two-sided.

In clinical studies, all hypotheses in the family of inferences are usually tested as two-sided. However, when appropriate, some hypotheses in the inferential family can be justified to be tested one-sided to gain greater overall power. The determination of whether a test is one-sided or two-sided depends on a number of possible factors, such as the prior knowledge of the efficacy of the drugs and the intended objective of the drug sponsor.

To formulate the idea, let us consider a family of null hypotheses $\{H_1, ..., H_m\}$. Let r (where $r < m$) be the number of one-sided hypotheses and the rest be two-sided hypotheses. Hence, a subset $\{H_1, ..., H_r\}$ is tested against one-sided alternatives, while the remaining null hypotheses $\{H_{r+1}, ..., H_m\}$ are tested against two-sided alternatives. Such families are hereafter referred to as directional-mixed families.

Consider a clinical study that was reported by Schwartz *et al.* (2002). The study compared the renal effects of the two selective cyclooxygenase (COX) 2 inhibitors (rofecoxib and celecoxib) with naproxen (dual COX-1/COX-2 inhibitor) and placebo in healthy elderly subjects who received normal-salt diet. The elderly subjects were divided into four groups and received four different treatments. They are rofecoxib, celecoxib, naproxen (active control) or no treatment (placebo control).

The response variable is the change from baseline for average daily urinary

sodium excretion during the first 72 hours of treatment. Naproxen is a nonsteroidal anti-inflammatory drug (NSAID) which is extensive used by elderly subjects. However, it produces adverse gastrointestinal (GI) effects. COX-2 selective inhibitors, such as rofecoxib and celecoxib, have been introduced in recent years as alternatives to naproxen for relieving pain or inflammation while reducing adverse GI effects. The experiment is to explore whether rofecoxib and celecoxib produce different renal effects when compared with naproxen, so the family of inferences contains the tests of rofecoxib versus naproxen and celecoxib versus naproxen. The inclusion of the placebo is to ensure a valid study and naproxen is a NSAID which is extensively used by elderly subjects. Since the renal effects of naproxen has a long research history, an one-sided test should be conducted for placebo control versus naproxen. But, there is no compelling reason for adopting one-sided tests for rofecoxib versus naproxen and celecoxib versus naproxen. Consequently, the family of inferences consists of a mixture of both one-sided and two-sided tests. Cheung *et al.* (2004) extended the Dunnett (1955) procedure to directional-mixed families that both one-sided and two-sided inferences are present in the given inferential family for one-way designs and the details will be reviewed in Chapter 2.

1.5 Objectives

The objective of this thesis is to extend the work of Cheung and Holland (1991) to directional-mixed families. Thus, we seek to develop a single-step procedure for multiple comparisons with a control in a two-way design with directional-mixed hypotheses, in which both one-sided and two-sided inferences are present in each of several groups.

Chapter 2 reviews the procedure of Cheung *et al.* (2004), comparing treatments

with a control in a one-way layout with directional-mixed hypotheses. Chapter 3 outlines the new procedure which deals with multiple comparisons with a control in two-way designs and directional-mixed families. The evaluation and tabulation of the required critical values will be provided in Chapter 4. In Chapter 5, a simulation study of average power is given to show the superioity of the proposed method. A numerical example will be given in Chapter 6 for illustrative purpose. Finally, conclusions together with directions for further researches are provided in Chapter 7.

2. MCC with Directional-mixed Families in One-way Designs: A Review

2.1 The model

Given an one-way fixed effect model, Cheung *et al.* (2004) generalised the Dunnett (1955) procedure to a more general testing environment in order to accommodate directional-mixed families. Suppose there are m active treatments and a control. Let Y_{ij} denotes the jth observation on the ith treatment for $i = 0, 1, ..., m$; $j = 1, ..., n_i$. Here $i = 0$ represents the control, while $i = 1, ..., m$ represent the m active treatments. Also, n_i denotes the sample size of ith treatment. Assume that $\varepsilon_{ij} \overset{ind}{\sim} \mathrm{N}(0, \sigma^2)$, so $Y_{ij} \overset{ind}{\sim} \mathrm{N}(\mu_i, \sigma^2)$ where μ_i is the mean of ith treatment and σ^2 is the common variance. Let $\bar{Y}_i$ be the sample mean of ith treatment, and $\hat{\sigma}^2$ be the pooled sample variance which is an unbiased estimator of σ^2, independent of $\bar{Y}_i$. Without loss of generality, it is also assumed that large value of μ_i means a better efficacy of ith treatment.

2.2 The test statistics

The inference problem under consideration includes estimating the m simultaneous confidence intervals for $\mu_i - \mu_0$ for $i = 1, ..., m$ or the simultaneous testings of the m null hypotheses:

$$H_i\colon \mu_i = \mu_0$$

for $i = 1, ..., m$ against

$$H_j'\colon \mu_j > \mu_0$$

for $j = 1, ..., r$ and

$$H_k'\colon \mu_k \neq \mu_0$$

for $k = r+1, ..., m$, where H_j' are the one-sided alternative hypotheses for $j = 1, ..., r$ and H_k' are the two-sided alternative hypotheses for $k = r+1, ..., m$.

To undertake these inferences, the test statistics $T_1, ..., T_r, |T_{r+1}|, ..., |T_m|$ are proposed by Cheung *et al.* (2004) for comparing the m treatments with a control, where

$$T_i = \frac{(\bar{Y}_i - \bar{Y}_0)}{\hat{\sigma}\sqrt{1/n_0 + 1/n_i}}$$

for $i = 1, ..., m$, where $\hat{\sigma}^2 = \frac{\sum_{i=0}^{m}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_i)^2}{\nu}$ and $\nu = (\sum_{i=0}^{m} n_i) - (m+1)$. The variates $T_1, ..., T_m$ have a multivariate t-distribution with ν degrees of freedom. The covariance matrix of the variates is denoted by $\sum = \rho_{ij}$ for $i, j = 1, ..., m$, with a product-correlation structure such that

$$\rho_{ij} = \begin{cases} b_i b_j, & i \neq j, \\ 1, & i = j. \end{cases}$$

where

$$b_i = \sqrt{\frac{n_i}{n_0 + n_i}}.$$

2.3 The evaluation of critical constants

To maintain α control over the probability of making any type I error, we need to compute critical constants $c_{1,\alpha}$ for one-sided inferences and $c_{2,\alpha}$ for two-sided

inferences such that

$$P\{T_j \leq c_{1,\alpha}, \quad j = 1, ..., r; \quad |T_k| \leq c_{2,\alpha}, \quad k = r+1, ..., m\} = 1 - \alpha. \tag{2.1}$$

Since there are infinite pairs $(c_{1,\alpha}, c_{2,\alpha})$ that satisfy equation (2.1), optimal critical values $(c^*_{1,\alpha}, c^*_{2,\alpha})$ are chosen such that EAA (expected average allowance) is minimum. EAA, which is used by Spurrier and Nizam (1990) to search for the optimal allocation of the sample sizes when comparing treatments to a control, is a measure of average expected width of the simultaneous confidence intervals. EAA is defined as

$$\text{EAA} \quad = \quad d_1 c_{1,\alpha} + d_2 c_{2,\alpha} \tag{2.2}$$

where

$$d_1 = \frac{1}{m} E(\hat{\sigma}) \sum_{i=1}^{r} \sqrt{1/n_0 + 1/n_i}$$

and

$$d_2 = \frac{2}{m} E(\hat{\sigma}) \sum_{j=r+1}^{m} \sqrt{1/n_0 + 1/n_j}.$$

To compute the optimal critical values $c^*_{1,\alpha}$ and $c^*_{2,\alpha}$, the procedures are as follows.

Firstly, for given values of m, r, α, ν and initial guess of $c_{1,\alpha}$, compute $c_{2,\alpha}$ by using the following equation:

$$\int_0^\infty \int_{-\infty}^\infty \prod_{i=1}^{r} [\Phi(\frac{b_i y + u c_{1,\alpha}}{\sqrt{1 - b_i^2}})] \prod_{j=r+1}^{m} [\Phi(\frac{b_j y + u c_{2,\alpha}}{\sqrt{1 - b_j^2}}) - \Phi(\frac{b_j y - u c_{2,\alpha}}{\sqrt{1 - b_j^2}})] \, \phi(y) \, g(u) \, dy \, du = 1 - \alpha \tag{2.3}$$

where $u = \frac{\hat{\sigma}}{\sigma}$, which is a $\sqrt{\chi^2_\nu / \nu}$ random variable. In addition, $\phi(\cdot)$ and $g(\cdot)$ are standard normal density and density function of u respectively. And $\Phi(\cdot)$ is the standard normal cumulative distribution function. Equation (2.3) involves the computation

of a two-dimensional integration. A brief account of the computation procedures is provided by Cheung *et al.* (2004).

Then, for simplicity, let $\frac{E(\hat{\sigma})}{m}$ be 1. Compute EAA according to (2.2) with the obtained the values of $c_{1,\alpha}$ and $c_{2,\alpha}$. With respect to Equation (2.3), with a given α, $c_{1,\alpha}$ can be written as a decreasing function of $c_{2,\alpha}$, say $c_{1,\alpha}$=$H(c_{2,\alpha})$.

Next, search for minimum EAA by repeating the previous two steps with all values of $c_{1,\alpha}$ in the range of $(H(d_{2,m,\alpha}), d_{2,m,\alpha})$ where $d_{2,m,\alpha}$ is the upper α-percentage point which is obtained by Dunnett two-sided procedure. The search is done by the IMSL subroutine UVMIF. The required optimal critical values $(c^*_{1,\alpha}, c^*_{2,\alpha})$ are those that minimize EAA.

Tabulation of the optimal critical values $(c^*_{1,\alpha}, c^*_{2,\alpha})$ for different values of m, r, ν and ρ are provided by Cheung *et al.* (2004).

2.4 Testing and estimation

For a given α and optimal critical values $c^*_{1,\alpha}$ and $c^*_{2,\alpha}$, each null hypothesis H_i is rejected if and only if

$$T_j > c^*_{1,\alpha}$$

where $j = 1, ..., r$. Similarly, each null hypothesis H_j is rejected if and only if

$$|T_k| > c^*_{2,\alpha}$$

where $k = r + 1, ..., m$.

The corresponding $100(1 - \alpha)\%$ simultaneous confidence intervals for $\mu_j - \mu_0$, where $j = 1, ..., r$, are

$$(\bar{Y}_j - \bar{Y}_0 - c^*_{1,\alpha}\hat{\sigma}\sqrt{1/n_j + 1/n_0},\ \infty)$$

And $100(1-\alpha)\%$ simultaneous confidence intervals for $\mu_k - \mu_0$, where $k = r+1, ..., m$, are

$$(\bar{Y}_k - \bar{Y}_0 - c^*_{2,\alpha}\hat{\sigma}\sqrt{1/n_k + 1/n_0},\ \ \bar{Y}_k - \bar{Y}_0 + c^*_{2,\alpha}\hat{\sigma}\sqrt{1/n_k + 1/n_0})$$

2.5 An example

Revisit the clinical study discussed in Section 1.4 for illustrating the above testing procedure. The comparisons of renal effects of rofecoxib and celecoxib with naproxen were studied. Four groups of elderly subjects receiving four different treatments. They are naproxen (active control), placebo, refecoxib and celecoxib. The response variable is the average daily urinary sodium excretion during the first 72 hours. By the reason mentioned in Section 1.4, naproxen (active control) is compared with the placebo with an one-sided test while the 2 remaining hypotheses in the family are compared with two-sided tests.

Denote μ_0, μ_1, μ_2 and μ_3 be the true mean changes from baseline for daily urinary sodium excretion of the four groups: naproxen, placebo, rofecoxib and celecoxib respectively. The inference problem is the simultaneous testing of the following three null hypotheses:

$$H_i\colon \mu_i = \mu_0$$

for $i = 1, 2, 3$ against the one-sided alternative hypothesis

$$H'_j\colon \mu_j > \mu_0$$

for $j = 1$, and the two-sided alternative hypotheses

$$H'_k\colon \mu_k \neq \mu_0$$

for $k = 2, 3$.

In this study, (n_0, n_1, n_2, n_3)=(15,14,17,16), $m = 3$, $r = 1$, $\nu = 58$ and $\alpha = 0.05$, the optimal critical values are $c^*_{1,0.05} = 2.318$ and $c^*_{2,0.05} = 2.350$. The corresponding 95% joint confidence intervals for $(\mu_1 - \mu_0)$, $(\mu_2 - \mu_0)$ and $(\mu_3 - \mu_0)$ are (8.265, ∞), (-19.712, 21.912) and (-7.615, 34.615), respectively. As the joint confidence intervals for first null hypothesis only does not include the value of zero, the H_1 is rejected. Therefore, we can conclude that the data support the validity of the study. The non-rejection of H_2 and H_3 indicates that there is not enough evidence that rofecoxib and celecoxib are different in renal effects as compared to naproxen.

On the other hand, if all hypotheses are tested with two-sided tests, then the corresponding critical value is 2.41 (Table B.3, Bechhofer and Dunnett (1988)). It is clear that $c^*_{1,0.05}$ and $c^*_{2,0.05}$ are much smaller than 2.41. It indicates that the procedure proposed by Cheung *et al.* (2004) is more powerful in directional-mixed families.

3. MCC with Directional-mixed Families in Two-way Designs

3.1 The model

Our proposed procedure extends Cheung and Holland (1991) to the situation that copes with a more general testing environment in which the family of inferences is composed of a mixture of one-sided and two-sided hypotheses, while controlling the familywise type I error rate at α. We consider the following two-way fixed effect model,

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

where $i = 1, ..., r,\ j = 0, ..., m$ and $k = 1, ..., n_{ij}$. Suppose there are two factors A and B, with r groups (treatments) in factor A and $(m + 1)$ treatments (m active treatments and one control) in factor B. The index $j = 0$ designates the control . Let Y_{ijk} denotes the kth observation on the jth treatment in group i. Furthermore, μ_{ij} and n_{ij} represents the sample mean and sample size of jth treatment in group i respectively, while ε_{ijk} is a random error component. Without loss of generality, we assume that a smaller value of μ_{ij} means a better efficacy. It is also assumed that $Y_{ijk} \overset{ind}{\sim} \mathrm{N}(\mu_{ij}, \sigma^2)$. Define $\bar{Y}_{ij}$, which equals to $\frac{\sum_{k=1}^{n_{ij}} Y_{ijk}}{n_{ij}}$, as the sample mean of jth treatment in group i and it is an unbiased estimator of μ_{ij}. Let $\hat{\sigma}^2$ be the pooled sample variance. It is an unbiased estimator of σ^2 and also independent of $\bar{Y}_{ij}$.

3.2 The test statistics

The inference problem is the simultaneous confidence intervals for $\mu_{ij} - \mu_{i0}$ for $i = 1, ..., r$ and $j = 1, ..., m$ or the simultaneous testings of the following null hypotheses:

$$H_{ij}\colon \mu_{ij} = \mu_{i0}$$

for $i = 1, ..., r$ and $j = 1, ..., m$ against one-sided alternative hypotheses

$$H'_{il}\colon \mu_{il} > \mu_{i0}$$

for $i = 1, ..., r$ and $l = 1, ..., a$, and two-sided alternative hypotheses

$$H'_{ik}\colon \mu_{ik} \neq \mu_{i0}$$

for $i = 1, ..., r$ and $k = a + 1, ..., m$.

To test the null hypotheses simultaneously in the directional-mixed families, we propose the use of the statistics $T_{i1}, ..., T_{ia}, |T_{ia+1}|, ..., |T_{im}|$ for $i = 1, ..., r$, where

$$T_{ij} = \frac{(\bar{Y}_{ij} - \bar{Y}_{i0})}{\hat{\sigma}\sqrt{1/n_{i0} + 1/n_{ij}}} \tag{3.1}$$

and $\hat{\sigma}^2 = \frac{\sum_{i=1}^{r}\sum_{j=0}^{m}\sum_{k=1}^{n_{ij}}(Y_{ijk} - \bar{Y}_{ij})^2}{f}$. The degrees of freedom f is $\sum_{i=1}^{r}\sum_{j=0}^{m}(n_{ij} - 1)$ and $f\hat{\sigma}^2/\sigma^2 \sim \chi^2_f$. For $1 \leq i \leq r$ and $1 \leq j_1 < j_2 \leq m$, the correlation between $(\bar{Y}_{ij_1} - \bar{Y}_{i0})$ and $(\bar{Y}_{ij_2} - \bar{Y}_{i0})$ is $\rho_i(j_1, j_2) = b_{ij_1}b_{ij_2}$, where

$$b_{ij} = \sqrt{\frac{n_{ij}}{n_{i0} + n_{ij}}}$$

for $i = 1, ..., r$ and $j = 1, ..., m$.

To control the familywise type I error rate at α, critical constants $c_{1,\alpha}$ and $c_{2,\alpha}$, which are required for one-sided and two-sided inferences respectively, in the two-way directional-mixed families are the solutions of the following equation:

$$P\{T_{il} \leq c_{1,\alpha},\ i = 1, ..., r,\ l = 1, ..., a;\ \ |T_{ik}| \leq c_{2,\alpha},\ \ i = 1, ..., r,\ k = a+1, ..., m\}$$
$$= 1 - \alpha \tag{3.2}$$

3.3 Testing and estimation

For a given α and critical values $c_{1,\alpha}$ and $c_{2,\alpha}$, each null hypothesis H_{il} for $i = 1, ..., r$ and $l = 1, ..., a$ is rejected if and only if the corresponding

$$T_{il} > c_{1,\alpha}.$$

And each null hypothesis H_{ik} for $i = 1, ..., r$ and $k = a+1, ..., m$ is rejected if and only if the corresponding

$$|T_{ik}| > c_{2,\alpha}.$$

In accordance, the one-sided $100(1-\alpha)\%$ simultaneous confidence intervals for $\mu_{il} - \mu_{i0}$ for $i = 1, ..., r$ and $l = 1, ..., a$ are

$$(\bar{Y}_{il} - \bar{Y}_{i0} - c_{1,\alpha}\hat{\sigma}\sqrt{1/n_{il} + 1/n_{i0}},\ \infty). \tag{3.3}$$

And the two-sided $100(1-\alpha)\%$ simultaneous confidence intervals for $\mu_{ik} - \mu_{i0}$ for $i = 1, ..., r$ and $k = a+1, ..., m$ are

$$(\bar{Y}_{ik} - \bar{Y}_{i0} - c_{2,\alpha}\hat{\sigma}\sqrt{1/n_{ik} + 1/n_{i0}},\ \ \bar{Y}_{ik} - \bar{Y}_{i0} + c_{2,\alpha}\hat{\sigma}\sqrt{1/n_{ik} + 1/n_{i0}}\ ). \tag{3.4}$$

Notice that when $a = 0$ and $a = m$, it is reduced to two-sided and one-sided procedures respectively given in Cheung and Holland (1991).

4. Evaluation and Tabulation of Critical Values

4.1 Evaluation of critical values

We need to find the values of $c_{1,\alpha}$ and $c_{2,\alpha}$ satisfy the following equation in order to ensure the familywise type I error rate is controlled at α.

$$P\{T_{ij} \leq c_{1,\alpha},\ i = 1, ..., r,\ j = 1, ..., a;\ |T_{ik}| \leq c_{2,\alpha},\ i = 1, ..., r,\ k = a+1, ..., m\}$$
$$= 1 - \alpha. \tag{4.1}$$

But for a given α, there are infinite number of pairs of $(c_{1,\alpha}, c_{2,\alpha})$ satisfy equation (4.1). The method for finding the optimal critical values used by Cheung *et al.* (2004) required complex computation which involves search for minimum expected average allowance (EAA). And the power increase of the method is mainly due to the reduction of the value of critical constant $c_{1,\alpha}$, not $c_{2,\alpha}$. So, for the sake of computation, we simply take $c_{2,\alpha}$ as $d_{2,\alpha}$ which is the two-sided critical constant in Cheung and Holland (1991). Then, $c_{1,\alpha}$ can be computed by Equation (4.2) with given r, m, a, α, f and $c_{2,\alpha}$. The two different methods for computing $c_{1,\alpha}$ and $c_{2,\alpha}$ are compared by using a simulation study which is given by Section 4.2. Equation (4.1) can be rewritten as

$$\int_0^\infty \prod_{i=1}^{r} \int_{-\infty}^{\infty} \prod_{l=1}^{a} [\Phi(\frac{b_{il}y + uc_{1,\alpha}}{\sqrt{1-b_{il}^2}})] \prod_{k=a+1}^{m} [\Phi(\frac{b_{ik}y + uc_{2,\alpha}}{\sqrt{1-b_{ik}^2}}) -$$
$$\Phi(\frac{b_{ik}y - uc_{2,\alpha}}{\sqrt{1-b_{ik}^2}})]\phi(y)g(u)dydu = 1 - \alpha \tag{4.2}$$

where $u = \frac{\hat{\sigma}}{\sigma}$, which is a $\sqrt{\chi^2_f/f}$ random variable, and y is standard normal random variable. Here $\phi(\cdot)$ and $g(\cdot)$ are density function of y and u, respectively. And $\Phi(\cdot)$ is the cumulative distribution function of standard normal.

Equation (4.2) involves the computation of a two-dimensional integration. Here is a brief account of the computation procedures.

The outer integral in Equation (4.2) related to the density function $g(u)$ is evaluated using subroutine QPROB of Copenhaver (1987) with 16-point Gauss-Legendre composite quadrature. The range $(0, \infty)$ of the outermost integral of Equation (4.2) was divided into subintervals of length of L and Equation (4.2) is approximated by

$$\sum_{d=0}^{\infty} \int_{dL}^{dL+L} \prod_{i=1}^{r} \int_{-\infty}^{\infty} \prod_{l=1}^{a} [\Phi(\frac{b_{il}y + uc_{1,\alpha}}{\sqrt{1-b_{il}^2}})]$$
$$\prod_{k=a+1}^{m} [\Phi(\frac{b_{ik}y + uc_{2,\alpha}}{\sqrt{1-b_{ik}^2}}) - \Phi(\frac{b_{ik}y - uc_{2,\alpha}}{\sqrt{1-b_{ik}^2}})]\phi(y)dyg(u)du.$$

Then, the limits of integration were rescaled from $(dL, dL + L)$ to $(-1, 1)$ so that Gauss-Legendre quadrature could be employed. The accuracy of this algorithm was compared with that of a 24-point quadrature using intervals of length $L/2$, with little difference in results in the sixth decimal places, refer to Copenhaver and Holland (1988).

The inner integrals in (4.2) is evaluated using Dunnett's (1989) algorithm with predetermined error bound 0.00001. The computational times and accuracy were as stated by Dunnett. This algorithm provides very accurate numerical results with small computing times which remain fairly stable even for large values of m, the number of active treatments.

The secant method is used to solve Equation (4.2) for $c_{1,\alpha}$ when α and $c_{2,\alpha}$ are given. The program terminates if the difference between successive iterates is less than 0.00001.

4.2 Comparisons of computational methods

As indicated earlier, our method to compute $c_{1,\alpha}$ and $c_{2,\alpha}$ deviates from the one given in Cheung *et al.* (2004). For simplicity, our proposed method and method used by Cheung *et al.* (2004) are denoted as PM and CG respectively.

In this section, we seek to demonstrate that by employing the new computation strategy, the gain in straight forwardness affects insignificantly the power of the proposed testing procedure.

For comparative purpose, we restrict the power study for the method PM to r=1 since the CG method is only given for an one-way layout.

When there is only one hypothesis, the usual definition of power is the probability of rejecting the null hypothesis given that it is false. For multiple comparison procedures, it is more complicated to define the power as more than one hypothesis are tested simultaneously. There are various definitions of power in MCP, such as the probability of rejection of at least one false null hypothesis, and the probability of rejection of all false null hypotheses. In this Chapter, we adopt a popular definition of power, average power, which is the proportion of false hypotheses that are rejected. The average power is denoted as:

$$AP = \frac{1}{N}\sum_{i=1}^{N} P_i \tag{4.3}$$

where N is the number of simulations in a certain MCP and P_i is the proportion of false hypotheses being rejected in ith simulation. Then, the required estimated percentage gain in average power of PM as compared to CG is defined as:

$$\text{GAP} = \frac{AP_{PM} - AP_{CG}}{AP_{CG}}\text{x}100\% \tag{4.4}$$

where AP_{PM} and AP_{CG} are the average power of PM and CG, respectively.

In our simulation study, average power is calculated based on simulations with 100,000 replications, ie. $N = 100,000$. Without loss of generality, we assume equal sample sizes for all treatments, ie. $n_{ij} = n$ for $i = 1, ..., r;\ j = 0, 1, ..., m$. In addition, α is chosen as 0.05.

$$T_{ij} = \frac{(\bar{Y}_{ij} - \bar{Y}_{i0})}{\hat{\sigma}\sqrt{1/n_{ij} + 1/n_{i0}}} \tag{4.5}$$

where $\bar{Y}_{i0} \overset{ind}{\sim} N(0,\ \sigma^2/n)$ and $\bar{Y}_{ij} \overset{ind}{\sim} N(\delta\sigma/\sqrt{n},\ \sigma^2/n)$ for $i = 1, ..., r$ and $j = 1, ..., m$, also $\delta\sigma/\sqrt{n}$ is the noncentrality parameter of $\bar{Y}_{ij}$.

As $n_{ij} = n$ is assumed, Equation (4.5) can be written as

$$T_{ij} = \frac{(\bar{Y}_{ij} - \bar{Y}_{i0})}{\hat{\sigma}\sqrt{2/n}} \tag{4.6}$$

Here are the steps of simulation:

1. Simulate $\frac{\bar{Y}_{i0}}{\sigma/\sqrt{n}}$ from $N(0,\ 1)$ and $\frac{\bar{Y}_{ij}}{\sigma/\sqrt{n}}$ from $N(\sigma,\ 1)$ for $i = 1, ..., r$ and $j = 1, ..., m$.

2. Simulate $f\hat{\sigma^2}/\sigma^2$ from χ^2_f, which is independent of $\bar{Y}_{ij}$.

3. For PM, calculate T_{ij} for $i = 1, ..., r;\ j = 1, ..., a$ and $|T_{ij}|$ for $i = 1, ..., r;\ j = a + 1, ..., m$ by Equation (4.6).

   For CG, calculate $|T_{ij}|$ for $i = 1, ..., r;\ j = 1, ..., m$ by Equation (4.6).

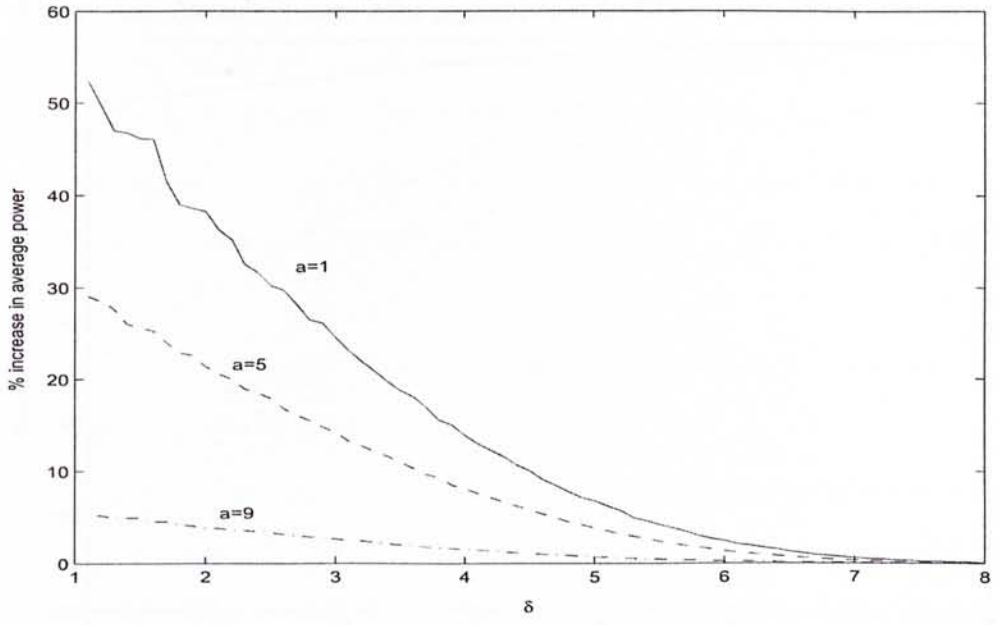
4. Perform PM and CG. Then, compute the proportion of rejecting false hypotheses for each simulation, P_i for $i = 1, ..., 100,000$.

5. Compute the average powers, AP_{PM} and AP_{CG} by Equation (4.3). Finally, calculate the percentage gain in average power of PM as compared to CG (GAP) by Equation (4.4).

In the simulation, the number of groups, r, is set to be 1. And the number of non-control treatment for each group, m, and the sample size for each treatment, n, are set to be 10. Also the number of one-sided inferences for each group a=1,5,9.

Figures 4.1 to 4.3 show the percentage power gain of PM when compared to CG for different values of a (no. of one-sided inferences for each group) and δ ($\delta\ \sigma/\ \sqrt{n}$ is noncentrality parameter), where m (total no. of active treatments for each group) is 10. Figure 4.1 shows the average power gain for each group of one-sided inferences, that means P_i in Equation (4.3) is the proportion of one-sided false hypotheses being rejected in ith simulation. Figure 4.2 shows the average power gain for the group of two-sided inferences, that means P_i is the proportion of two-sided false hypotheses being rejected in ith simulation. Finally, the average power gain for the entire family of inferences are considered and is shown on Figure 4.3, and in this case, P_i is the proportion of both one-sided and two-sided false hypotheses being rejected in ith simulation.

Figure 4.1: Average power gain as compared to CG for one-sided inferences



From Figure 4.1, PM is more powerful than CG when only one-sided inferences are counted. The average power gain increases for small a and small δ, where the gain can be up to above 50%. From Figure 4.2, PM is less powerful than CG when only two-sided inferences are counted. The power loss becomes smaller as a decreases and δ increases, where the maximum loss is bound by 30%. Figure 4.3 shows that the difference between average power of PM and CG is not large, but PM is still slightly more powerful than CG (the power gain is within 10%), when both one-sided and two-sided inferences are counted. From the simulation study, it is shown that PM is slightly more powerful than CG. So, for practical purpose, the proposed method for computing $c_{1,\alpha}$ and $c_{2,\alpha}$ can be adopted.

Figure 4.2: Average power gain as compared to CG for two-sided inferences

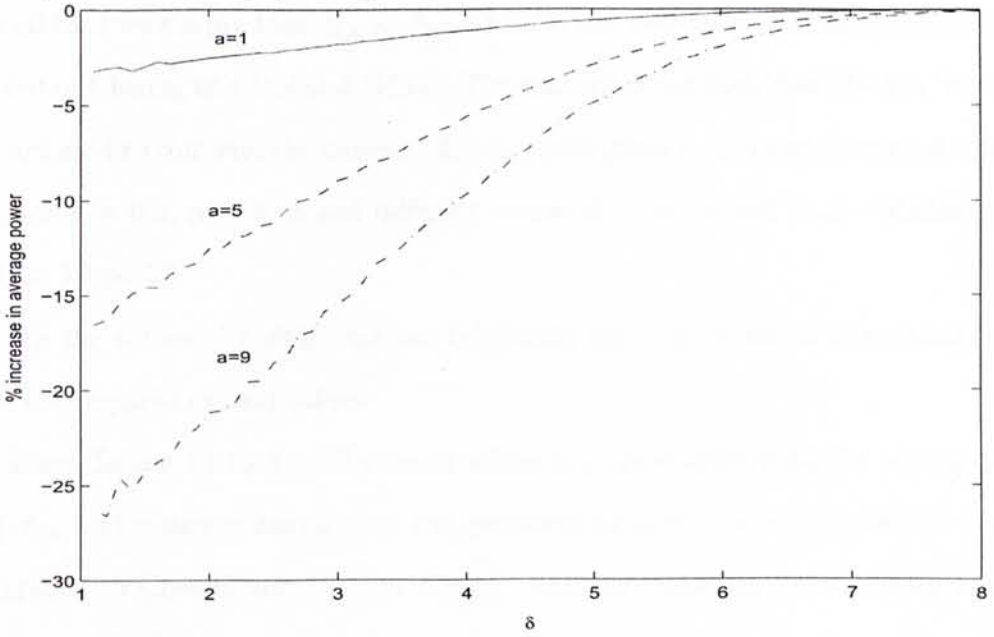
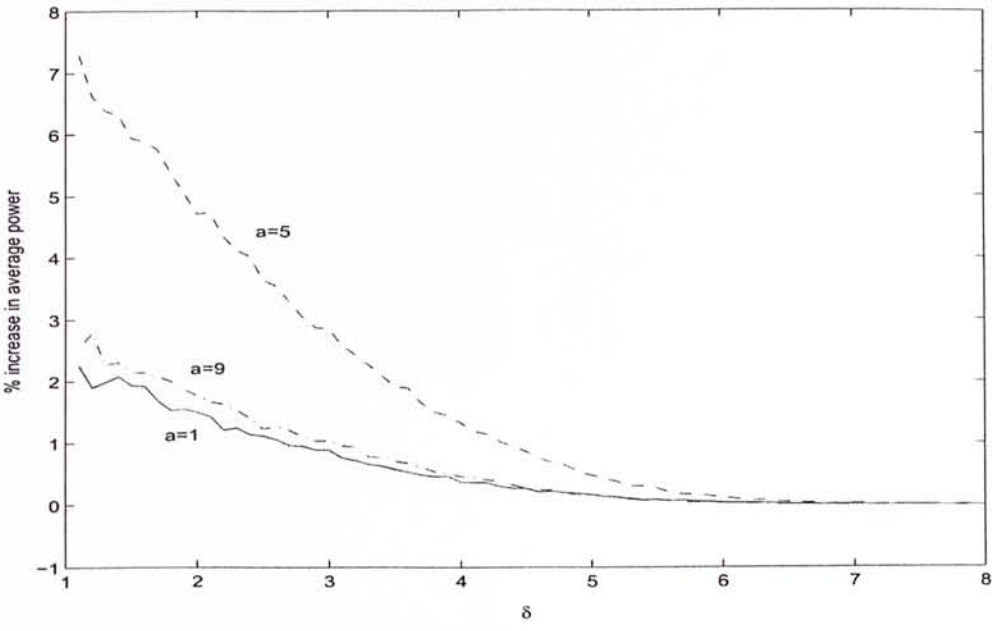


Figure 4.3: Average power gain as compared to CG for the entire family of inferences



4.3 Tabulation of critical values

Recall that we simply take $c_{2,\alpha}$ as $d_{2,\alpha}$ which is the two-sided critical constant tabulated in Cheung and Holland (1991). The numerical methods described in Section 4.1 are used to compute the critical value $c_{1,\alpha}$ with given $c_{2,\alpha}$. Selected critical values $c_{1,\alpha}$ for $\rho = 0.5$, $\alpha = 0.05$ and different values of r, m, a and f are tabulated in Tables 4.1 to 4.5.

For the values of f which are not tabulated, one can use linear interpolation to find the required critical values.

From Tables 4.1 to 4.5, all critical values $c_{1,\alpha}$ are smaller than the corresponding $d_{2,\alpha}$. This demonstrates that our proposed procedure is more powerful when compared to Cheung and Holland (1991). And the reduction becomes larger as a increases. That means the power gain increases as the number of one-sided hypotheses within each group increases.

Table 4.1

Critical values of $c_{1,\alpha}$ for r (number of groups)= 1

		$c_{1,\alpha}$				
m	a	$f = 20$	40	60	100	∞
2	1	2.0629	2.0006	1.9806	1.9650	1.9419
3	1	2.2447	2.1694	2.1452	2.1263	2.0977
	2	2.2220	2.1501	2.1270	2.1089	2.0821
4	1	2.3678	2.2830	2.2559	2.2346	2.2038
	2	2.3509	2.2686	2.2423	2.2217	2.1910
	3	2.3306	2.2517	2.2264	2.2066	2.1774
5	1	2.4603	2.3681	2.3385	2.3156	2.2818
	2	2.4468	2.3565	2.3277	2.3052	2.2724
	3	2.4311	2.3435	2.3154	2.2934	2.2610
	4	2.4125	2.3282	2.3012	2.2801	2.2488
6	1	2.5340	2.4353	2.4042	2.3795	2.3436
	2	2.5227	2.4250	2.3952	2.3711	2.3357
	3	2.5099	2.4153	2.3851	2.3615	2.3272
	4	2.4953	2.4031	2.3738	2.3507	2.3167
	5	2.4779	2.3892	2.3608	2.3385	2.3057
7	1	2.5949	2.4915	2.4582	2.4326	2.3962
	2	2.5853	2.4831	2.4506	2.4252	2.3894
	3	2.5744	2.4741	2.4422	2.4172	2.3821
	4	2.5624	2.4641	2.4328	2.4083	2.3737
	5	2.5485	2.4528	2.4222	2.3983	2.3643
	6	2.5323	2.4398	2.4102	2.3870	2.3541
8	1	2.6469	2.5389	2.5042	2.4776	2.4379
	2	2.6384	2.5318	2.4975	2.4711	2.4320
	3	2.6290	2.5239	2.4903	2.4642	2.4257
	4	2.6188	2.5153	2.4823	2.4566	2.4188
	5	2.6072	2.5058	2.4734	2.4481	2.4107
	6	2.5940	2.4951	2.4635	2.4387	2.4022
	7	2.5786	2.4829	2.4523	2.4283	2.3929

Table 4.2

Critical values of $c_{1,\alpha}$ for r (number of groups)= 2

		$c_{1,\alpha}$				
m	a	$f = 20$	40	60	100	∞
2	1	2.3942	2.3067	2.2788	2.2570	2.2246
3	1	2.5683	2.4654	2.4326	2.4069	2.3691
	2	2.5506	2.4513	2.4196	2.3948	2.3581
4	1	2.6867	2.5730	2.5372	2.5084	2.4667
	2	2.6735	2.5624	2.5270	2.4992	2.4583
	3	2.6579	2.5501	2.5157	2.4887	2.4489
5	1	2.7766	2.6545	2.6152	2.5851	2.5399
	2	2.7657	2.6456	2.6074	2.5773	2.5331
	3	2.7534	2.6359	2.5982	2.5689	2.5256
	4	2.7390	2.6247	2.5881	2.5596	2.5173
6	1	2.8481	2.7189	2.6781	2.6455	2.5982
	2	2.8391	2.7118	2.6711	2.6394	2.5924
	3	2.8289	2.7036	2.6636	2.6323	2.5862
	4	2.8173	2.6946	2.6553	2.6246	2.5794
	5	2.8040	2.6844	2.6461	2.6162	2.5719
7	1	2.9074	2.7704	2.7295	2.6964	2.6430
	2	2.8997	2.7665	2.7239	2.6906	2.6400
	3	2.8912	2.7595	2.7174	2.6847	2.6353
	4	2.8815	2.7519	2.7105	2.6780	2.6298
	5	2.8706	2.7434	2.7028	2.6709	2.6238
	6	2.8581	2.7340	2.6942	2.6631	2.6171
8	1	2.9587	2.8187	2.7742	2.7389	2.6885
	2	2.9515	2.8132	2.7687	2.7342	2.6816
	3	2.9439	2.8069	2.7632	2.7289	2.6775
	4	2.9358	2.8004	2.7571	2.7232	2.6728
	5	2.9265	2.7932	2.7505	2.7171	2.6676
	6	2.9161	2.7852	2.7433	2.7105	2.6619
	7	2.9043	2.7763	2.7353	2.7032	2.6557

Table 4.3

Critical values of $c_{1,\alpha}$ for r (number of groups)= 3

		$c_{1,\alpha}$				
m	a	$f = 20$	40	60	100	∞
2	1	2.5801	2.4761	2.4430	2.4171	2.3789
3	1	2.7508	2.6298	2.5918	2.5615	2.5173
	2	2.7353	2.6181	2.5808	2.5515	2.5085
4	1	2.8671	2.7350	2.6929	2.6600	2.6113
	2	2.8555	2.7258	2.6845	2.6522	2.6044
	3	2.8418	2.7154	2.6751	2.6435	2.5969
5	1	2.9555	2.8145	2.7694	2.7337	2.6819
	2	2.9459	2.8069	2.7624	2.7276	2.6763
	3	2.9350	2.7984	2.7548	2.7206	2.6702
	4	2.9226	2.7890	2.7464	2.7129	2.6636
6	1	3.0262	2.8778	2.8305	2.7933	2.7383
	2	3.0182	2.8715	2.8244	2.7876	2.7335
	3	3.0091	2.8644	2.8179	2.7819	2.7284
	4	2.9989	2.8566	2.8111	2.7755	2.7229
	5	2.9874	2.8479	2.8034	2.7684	2.7169
7	1	3.0844	2.9306	2.8811	2.8416	2.7850
	2	3.0777	2.9246	2.8761	2.8373	2.7808
	3	3.0702	2.9188	2.8702	2.8322	2.7764
	4	3.0617	2.9123	2.8643	2.8269	2.7717
	5	3.0521	2.9050	2.8580	2.8210	2.7666
	6	3.0413	2.8969	2.8508	2.8146	2.7612
8	1	3.1353	2.9755	2.9243	2.8843	2.8249
	2	3.1287	2.9706	2.9198	2.8796	2.8211
	3	3.1222	2.9650	2.9151	2.8754	2.8172
	4	3.1149	2.9595	2.9096	2.8705	2.8131
	5	3.1068	2.9532	2.9040	2.8655	2.8087
	6	3.0977	2.9464	2.8980	2.8600	2.8040
	7	3.0874	2.9388	2.8913	2.8540	2.7989

Table 4.4

Critical values of $c_{1,\alpha}$ for r (number of groups)= 4

		$c_{1,\alpha}$				
m	a	$f = 20$	40	60	100	∞
2	1	2.7093	2.5926	2.5553	2.5265	2.4841
3	1	2.8772	2.7433	2.7011	2.6675	2.6186
	2	2.8634	2.7328	2.6912	2.6587	2.6111
4	1	2.9929	2.8471	2.8004	2.7639	2.7162
	2	2.9820	2.8385	2.7930	2.7569	2.7102
	3	2.9695	2.8291	2.7843	2.7493	2.7037
5	1	3.0794	2.9250	2.8748	2.8366	2.7791
	2	3.0712	2.9183	2.8689	2.8309	2.7742
	3	3.0613	2.9110	2.8624	2.8246	2.7690
	4	3.0500	2.9020	2.8547	2.8178	2.7634
6	1	3.1503	2.9876	2.9355	2.8945	2.8342
	2	3.1423	2.9815	2.9299	2.8897	2.8300
	3	3.1344	2.9754	2.9242	2.8845	2.8256
	4	3.1252	2.9683	2.9182	2.8788	2.8208
	5	3.1146	2.9605	2.9111	2.8726	2.8158
7	1	3.2087	3.0395	2.9852	2.9428	2.8799
	2	3.2022	3.0345	2.9807	2.9382	2.8762
	3	3.1947	3.0288	2.9754	2.9339	2.8724
	4	3.1870	3.0230	2.9702	2.9291	2.8683
	5	3.1784	3.0164	2.9646	2.9239	2.8640
	6	3.1684	3.0091	2.9580	2.9181	2.8593
8	1	3.2584	3.0837	3.0269	2.9839	2.9189
	2	3.2527	3.0793	3.0234	2.9802	2.9156
	3	3.2465	3.0746	3.0193	2.9760	2.9122
	4	3.2394	3.0693	3.0143	2.9719	2.9086
	5	3.2321	3.0635	3.0095	2.9674	2.9048
	6	3.2238	3.0575	3.0042	2.9624	2.9008
	7	3.2144	3.0507	2.9981	2.9571	2.8965

Table 4.5

Critical values of $c_{1,\alpha}$ for r (number of groups)= 5

		$c_{1,\alpha}$				
m	a	$f = 20$	40	60	100	∞
2	1	2.8079	2.6811	2.6406	2.6089	2.5636
3	1	2.9741	2.8302	2.7842	2.7476	2.6952
	2	2.9613	2.8201	2.7750	2.7395	2.6885
4	1	3.0890	2.9320	2.8813	2.8428	2.7850
	2	3.0786	2.9243	2.8749	2.8362	2.7797
	3	3.0670	2.9155	2.8671	2.8293	2.7741
5	1	3.1752	3.0092	2.9560	2.9140	2.8527
	2	3.1673	3.0029	2.9499	2.9090	2.8483
	3	3.1579	2.9958	2.9440	2.9032	2.8437
	4	3.1473	2.9879	2.9369	2.8970	2.8387
6	1	3.2451	3.0710	3.0147	2.9706	2.9068
	2	3.2379	3.0657	3.0102	2.9666	2.9030
	3	3.2304	3.0598	3.0048	2.9621	2.8991
	4	3.2216	3.0533	2.9993	2.9569	2.8949
	5	3.2118	3.0460	2.9928	2.9513	2.8940
7	1	3.3023	3.1216	3.0645	3.0191	2.9517
	2	3.2967	3.1174	3.0600	3.0147	2.9485
	3	3.2900	3.1125	3.0557	3.0107	2.9450
	4	3.2828	3.1068	3.0505	3.0065	2.9413
	5	3.2747	3.1011	3.0454	3.0016	2.9375
	6	3.2655	3.0943	3.0394	2.9964	2.9334
8	1	3.3528	3.1664	3.1065	3.0596	2.9901
	2	3.3467	3.1618	3.1028	3.0562	2.9872
	3	3.3412	3.1575	3.0986	3.0522	2.9841
	4	3.3348	3.1529	3.0945	3.0484	2.9809
	5	3.3279	3.1475	3.0896	3.0444	2.9775
	6	3.3201	3.1420	3.0849	3.0398	2.9739
	7	3.3114	3.1356	3.0792	3.0350	2.9702

5. Power study

5.1 Objectives

To investigate the performance of our proposed procedure as compared to Cheung and Holland (1991), a simulation study is performed to study the power. For simplicity, the new procedure and Cheung and Holland (1991) are denoted as follow respectively:

PM MCC in two-way designs with directional-mixed families (The new procedure proposed in this thesis)

CH MCC in two-way designs with two-sided inferences (Cheung and Holland (1991))

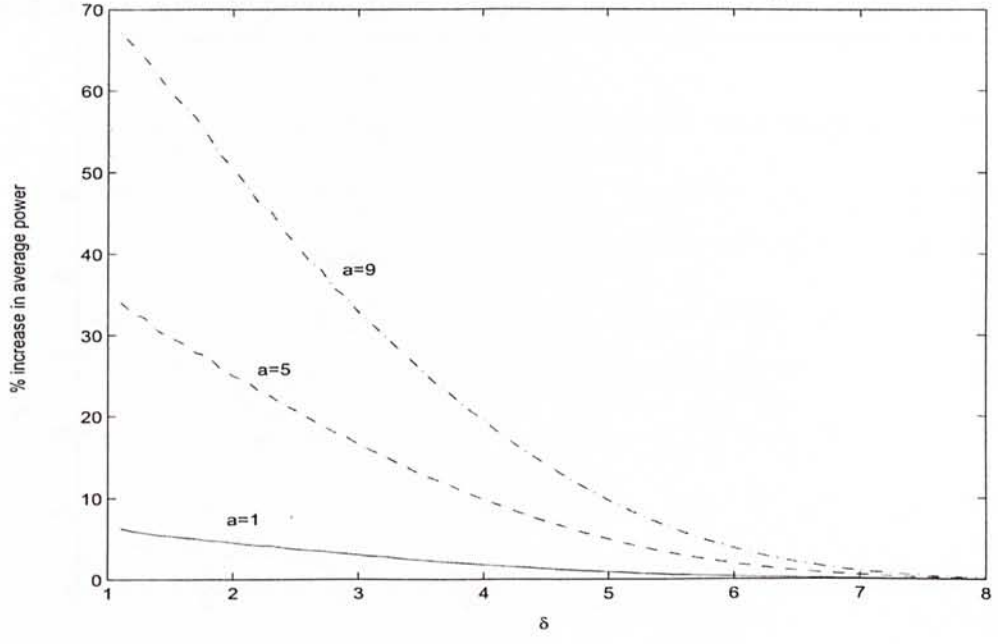
The objective of the simulation study is to evaluate the gain in power of the new procedure PM as compared to CH. The simulation procedures are similar to section 4.2 and we compare our proposed procedure to CH in this section.

5.2 Simulation results

Figures 5.1, 5.2 and 5.3 show the percentage increase in average power of PM as compared to CH when $m = 10$, $\rho = 0.5$, $\alpha = 0.05$ and $r = 2$, 5 and 10.

Similar patterns are obtained for different values of r, and hence only $r = 2$, 5 and 10 are reported. From the figures, GAP is positive for different values of δ, a and r. That means our proposed procedure is more powerful than Cheung and Holland (1991). Also, when the number of one-sided inferences for each group, a, increases, the percentage gain in power increases dramatically. It is because greater

Figure 5.1: Average power gain as compared to CH when r (no. of groups) =2



a represents allowing more prior directional information in the testing procedure. On the other hand, the percentage gain in power declines as δ increases. It is reasonable since both PM and CH are expected to reject most of the hypotheses for large δ and so the difference between the two average powers (AP_{CH} and AP_{PM}) is much less.

In conclusion, it can be shown that our proposed method is more powerful than Cheung and Holland (1991). And the average power gain becomes larger when a (number of one-sided hypotheses for each group) is larger and δ (the noncentrality parameter component of $\bar{Y}_{ij}$) is smaller.

Figure 5.2: Average power gain as compared to CH when r (no. of groups) =5

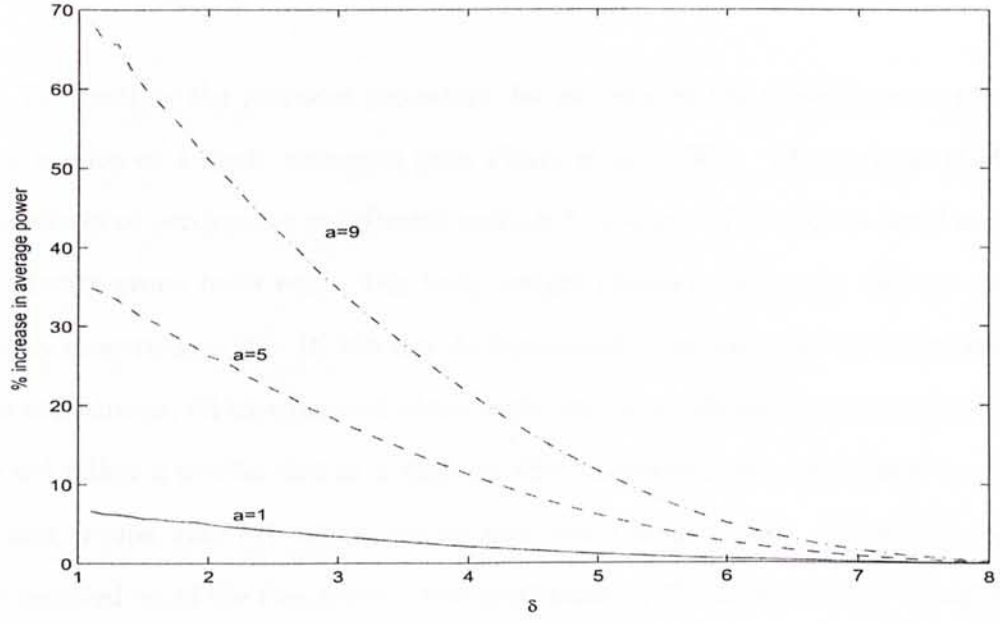
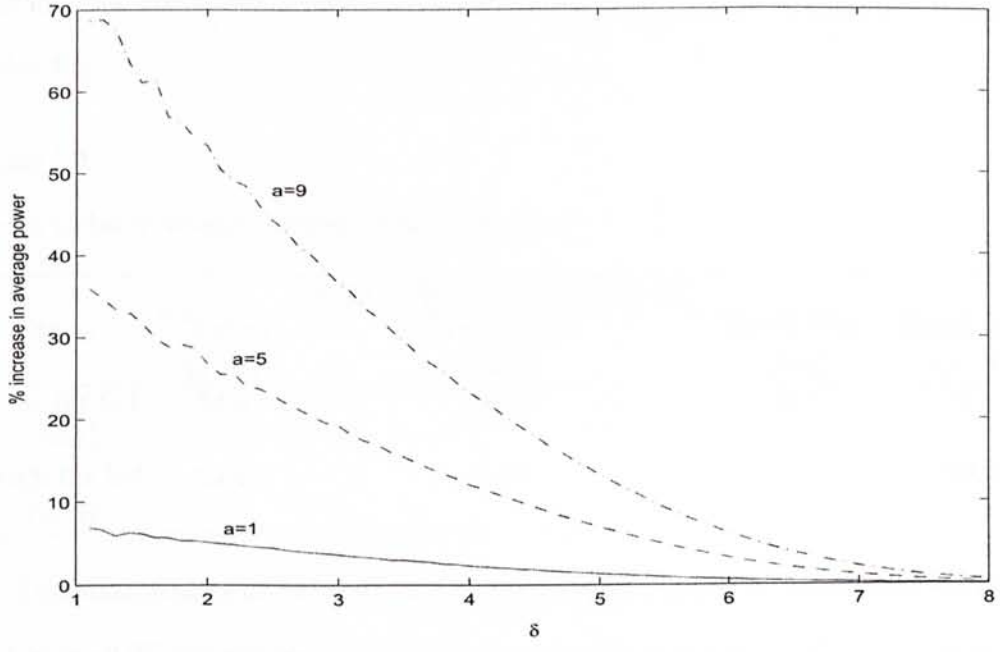


Figure 5.3: Average power gain as compared to CH when r (no. of groups) =10



6. Numerical example

To illustrate the proposed procedure, let us consider the following example. It is a portion of a study extracted from Philip *et al.* (2003). The example studied the effects of peroxisome proliferator-activated receptors (PPARs) on body weight in obesity-prone male rats. The body weight changes during the clinical study and is measured on day 16 and day 44 (measured in grams). At the beginning of the experiment, 64 diet-induced obese male rats were selected and randomized to be fed either a low-fat diet or a high-fat diet. Therefore, the experiment consists of two groups, rats fed with a low-fat diet and a high-fat diet. On day 15, each rat received on of the four treatments: pioglitazone (PPARγ activator), fenofibrate (PPAR-α activator), ragaglitazar (dual PPARα/γ activator) and control treatment. Therefore, there are two groups and four treatments for each group. The sample size for each treatment-group combination is eight (n_{ij}=8). The data are shown in Table 6.1.

Table 6.1

Mean of body weight changes during treatment

Group i	Mean body weight changes (g)			
	Control (j=0)	Pioglitazone (j=1)	Fenofibrate (j=2)	Ragalitazar (j=3)
Low fat fed (i=1)	54.8	102.9	46.9	77.8
High fat fed (i=2)	65.9	104.0	27.3	99.3

The homogeneity of the error variance of each subgroup is verified by the Bartlett's test at α=0.05. No significant difference among the variances of all subgroups is ob-

served. (p-value=0.5970).

Since PPARγ agonists are associated with weight increase from clinical experience (Akazawa *et al.* (2000)), one-sided tests can be conducted on pioglitazone versus control treatment. For fenofibrate versus control treatment and ragalitazar versus control treatment, no compelling reason for adopting one-sided tests is found. Thus, two-sided tests are used.

Denote μ_{ij} be the true mean body weight change of treatment j in group i. Conduct the simultaneous testing of the following null hypotheses.

$$H_{ij} : \mu_{ij} = \mu_{i0}$$

for i=1,2 and j=1,2,3 against

$$H'_{il} : \mu_{il} > \mu_{i0}$$

for i=1,2, and l=1,

$$H'_{ik} : \mu_{ik} \neq \mu_{i0}$$

for i=1,2 and k=2,3.

In this study, r=2, m=3, a=1, n_{ij}=8 for i=1,2 and j=1,2,3, ρ=0.5, the degrees of freedom f=56 and α=0.05.

Table 6.2

Values of $c_{1,\alpha}$ for interpolation (extracted from Table 4.2) with r=1, m=3 and a=1

f	$c_{1,\alpha}$
40	2.4654
60	2.4326

We interpolate on f as follow:

$\hat{c}_{1,\alpha}$=2.4654-($\frac{16}{20}$)(2.4654-2.4326)=2.4392

For $c_{2,\alpha}$, we simply take the value of $d_{2,\alpha}$ (the two-sided critical constant given by Cheung and Holland (1991)) as $c_{2,\alpha}$, which approximately equals to 2.688. So, the critical values ($c_{1,\alpha}$, $c_{2,\alpha}$) are (2.4392, 2.6880). The estimated σ is 7.3058.

According to (3.3) and (3.4), the 95% simultaneous confidence interval for μ_{ij}-μ_{io}, where i=1,2 and j=1,2,3, are:

Table 6.3

95% confidence interval for μ_{ij}-μ_{i0}

	j=1	j=2	j=3
i=1	(39.1898, ∞)	(-17.7190, 1.9190)	(13.1810, 32.8190)
i=2	(29.1898, ∞)	(-48.4190, -28.7810)	(23.5810, 43.2190)

Therefore, only the null hypothesis corresponding to fenofibrate for low-fat diet group is not rejected, as zero is contained in this confidence interval.

7. Conclusions

In this thesis, we have developed testing procedure for simultaneous hypothesis testing of multiple comparisons with a control in two-factor situations and directional-mixed families.

The outlines of the proposed procedure and the numerical methods of computation of the required critical values have been discussed in Chapter 3 and 4, respectively. In Chapter 5, the proposed procedure is proved to be more powerful than Cheung and Holland (1991) by simulation study.

Throughout this thesis, the proposed procedure has been developed under the assumptions of normality and homogeneity of error variances. But, in reality, these assumptions may be violated. For further research, methods are needed to generalize our procedure to situation when the above assumptions do not seem to be reasonable.

Also, it should be useful to derive the explicit expression of the power function which can be used to estimate required sample sizes of an experiment. So, investigation of the power function of the proposed procedure is a potential direction of future studies.

Bibliography

1. Akazawa, S., Sun, F., Ito, M., Kawasaki, E. and Eguchi, K. (2000). Efficacy of troglitazone on body fat distribution in type 2 diabetes. *Diabetes Care*, **23**, 1067-1071.

2. Bechhofer, R. E. and Dunnett, C. W. (1988). *Percentage Points of Multivariate Student t Distributions.* Volume 11 of *Selected Tables in Mathematical Statistics.* American Mathematical Society, Providence, RI.

3. Cheung, S. H. and Holland, B. (1991). Extension of Dunnett's multiple comparison procedure to the case of several groups. *Biometrics*, **47**, 21-32.

4. Cheung, S. H. and Holland, B. (1992). Extension of Dunnett's multiple comparison procedure with different sample sizes to the case of several groups. *Computational Statistics and Data Analysis*, **14**, 165-182.

5. Cheung, S. H. and Holland, B. (1994). A step-down procedure for multiple tests of treatment versus control in each of several groups. *Statistics in Medicine*, **13**, 2261-2267.

6. Cheung, S. H., Kwong, K. S., Chan, W. S. and Leung, S. P. (2004). Multiple comparisons with a control in families with both one-sided and two-sided hypotheses. *Statistics in Medicine* (To Appear).

7. Copenhaver, M. D. and Holland, B. (1987). Calculation of the true experimentwise error rates for mulptiple comparisons procedures in a two-way factorial design. *Proceedings, 19^{th} Symposium on the Interface of Computer Science and Statistics*, Philadelphia, Pennsylvania.

8. Copenhaver, M. D. and Holland, B. (1988). Computation of the distribution of the maximum studentized range statistic with application to multiple significance testing of simple effects. *Journal of Statistical Computation and Simulation*, **30**, 1-15.

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

10. Dunnett, C. W. (1964). New tables for multiple comparisons with a control. *Biometrics*, **20**, 482-491.

11. Dunnett, C. W. (1989). Multivariate normal probability integrals with product correlation structure. *Applied Statistics*, **38**, 564-579.

12. Dunnett, C. W. and Tamhane, A. C. (1991). Step-down multiple tests for comparing treatments with a control in unbalanced one-way layouts. *Statistics in Medicine*, **10**, 939-947.

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

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

15. Hsu, J. C. (1996). *Multiple Comparisons: Theory and Methods.* London: Chapman and Hall.

16. Philip, J. L., Per, B. J., Rikke, V. S., Leif, K. L., Niels, V., Erik, M. W. and Karsten, W. (2003). Differential Influences of Peroxisome Proliferator and Energy Homeostasis. *Diabetes*, **52**, 2249-2259.

17. Quek, A. L. (1999). Multiple comparison procedures for a two-way design. Honours thesis. Department of Statistics and Applied Probability, National University of Singapore.

18. Schmartz, J. I., Vandormael, K., Malice, M. P., Kalyani, R. N., Lasseter, K. C., Holmes, G. B., Gertz, B. J., Gottesdiener, K.M., Laurenzi, M., Redfern, K. J. and Brune, K. (2002). Comparison of rofecoxib, celecoxib, and naproxen on renal function in elderly subjects receiving a normal-salt diet. *Clinical Pharmacology and Therapeutics*, **72**, 50-61.

19. Spurrier, J. D. and Nizam, A. (1990). Sample size allocation for simultaneous inference in comparison with control experiments. *Journal of the American Statistical Association*, **85**, 181-186.

20. Tukey, J. W. (1953). The problem of multiple comparisons. Unpublished Manuscript.

21. Westfall, P. H. and Young, S. S. (1993). *Resampling-based Multiple Testing: Examples and Methods for P-value Adjustment.* New York: Wiley.