NEW METHODS FOR ELIMINATING INFERIOR TREATMENTS IN CLINICAL TRIALS

A Thesis
Presented to
The Academic Faculty

by

Chen-ju Lin

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology
August 2007

NEW METHODS FOR ELIMINATING INFERIOR TREATMENTS IN CLINICAL TRIALS

Approved by:

Professor Anthony J. Hayter, Advisor
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Alexander Shapiro
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Yajun Mei
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Brani Vidakovic
Department of Biomedical Engineering
Georgia Institute of Technology

Professor Nicoleta Serban
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: June 1, 2007

To my parents, my brother, and Wei-yu.

ACKNOWLEDGEMENTS

I would like to thank a number of individuals for their support and assistance during my study at Georgia Tech. Foremost, I want to express my special gratitude to my advisor, Dr. Anthony Hayter, for his inspiration, guidance, and encouragement. He helps me a lot through the whole Ph.D. study. I am grateful to Dr. Alexander Shapiro for his insightful instruction. He is a gifted teacher and researcher. I thank Dr. Yajun Mei for helping me with building confidence. The discussion with him is always fruitful and inspires me with new research ideas. I appreciate Dr. Brani Vidakovic and Dr. Nicoleta Serban for their valuable comments on both teaching and research. Their intelligent and precise suggestions make my thesis more complete.

I want to extend my warm thanks to my friends here and in Taiwan. Their company makes my life delightful. Last but not least, I would like to show my profound appreciation to my parents, my brother, and Wei-yu. Their selfless support make it possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Multiple comparisons and selection procedures are commonly studied in research and employed in application. Clinical trial is one of popular fields to which the subject of multiple comparisons is extensively applied. Based on the Federal Food, Drug, and Cosmetic Act, drug manufacturers need to not only demonstrate safety of their drug products but also establish effectiveness by substantial evidence in order to obtain marketing approval. However, the problem of error inflation occurs when there are more than two groups to compare with at the same time. How to design a test procedure with high power while controlling type I error becomes an important issue.

The treatment with the largest population mean is considered to be the best one in the study. Potentially the best treatments can receive increased resources and further investigation by excluding clearly inferior treatments. Hence, a small number of possibly the best treatments is preferred. This thesis focuses on the problem of eliminating the less effective treatments among three in clinical trials. The goal is to increase the ability to identify any inferior treatment providing that the probability of excluding any best treatment is guaranteed to be less than or equal to α. A step-down procedure is applied to solve the problem.

The general step-down procedure with fixed thresholds is conservative in our problem. The test is not efficient in rejecting the less effective treatments. We propose two methods with sharper thresholds to improve current procedures and construct a subset containing strictly inferior treatments. The first method, the restricted parameter space approach, is designed for the scenario when prior information about

range of treatment means is known. The second method, the step-down procedure with feedback, utilizes observations to modify the threshold and controls error rate for the whole parameter space. The new procedures have greater ability to detect more inferior treatments than the standard procedure. In addition, type I error is also controlled under mild violation of the assumptions demonstrated by simulation.

CHAPTER I

INTRODUCTION

1.1 Background

The Federal Food, Drug, and Cosmetic Act was first passed by Congress back in 1938. The Act requested drug labels provide adequate direction for safe use. In addition, it regulated drug manufacturers prove the safety of their products to the U.S. Food and Drug Administration (FDA) before entering the market. With growing concern about high drug prices and misleading or unsupported assertions made by pharmaceutical companies regarding to their products, Congress amended the Act and added a requirement for effectiveness in 1962.

The 1962 Drug Amendments contained a provision requiring manufacturers to establish the effectiveness of their drug products by "substantial evidence" in order to obtain marketing approval. Substantial evidence was defined in section 505(d) of the Act as "evidence consisting of adequate and well-controlled investigations, including clinical investigations, by experts qualified by scientific training and experience to evaluate the effectiveness of the drug involved, on the basis of which it could fairly and responsibly be concluded by such experts that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the labeling or proposed labeling thereof."[1] What components establish sufficient evidence of effectiveness and how to demonstrate the evidence have became

a contentious issue since then.

The process of proving the evidence is usually costly and time consuming. To a pharmaceutical company, the level of effectiveness is one of determinant elements for whether investing a new drug. In terms of public health, it is important for the government to set up regulations for drug development. Therefore, it is an essential topic to research into efficient and sound methodologies for testing the efficacy of treatments.

1.2 Introduction to clinical trials

A lot of resources have been allotted to the field of clinical trials. The National Institutes of Health (NIH) had an actual funding of 2,767 million dollars in this field in FY 2006. The funding level is estimated to be 2,764 and 2,756 million dollars in FY 2007 and 2008 respectively.

A clinical trial is a research study which tests a treatment in human beings to see whether it is both safe and effective to remedy a disease or a condition. Each trial must follow a protocol which explains the intention, the necessity, and the plan of a study. The clinical trials can provide researchers with information that helps them better understand the diseases and compare the performance of the treatments under different conditions. Clinical trials can be generally categorized as follows depending on the aspects of medical care they serve.

- **Treatment trials:** Test treatments for specific diseases or conditions.
- **Supportive care trials:** Study methods to provide a certain group of subjects

with a better quality of life.

- **Prevention trials:** Reduce the risk of developing a disease for healthy people.
- **Diagnostic trials:** Test new ways to detect a disease earlier and more accurately.

In clinical trials, there are mainly three types of comparisons: trials to show (1) superiority, (2) equivalence or noninferiority, and (3) dose-response relationship. New drugs are first tested in laboratories and then on animals. Since clinical trails involve human beings, they are normally expensive and have strict requirements for safety. Before carrying out a clinical trial, there needs to be strong evidence that the therapy is secure to people and is most likely effective to patients. Typically, there are four stages of clinical trials serving different purposes:

- **Phase I:**

  A small group of people about 20 to 80 participate in the first phase. Researchers test a new drug or treatment to evaluate its safety, determine a safe dosage range, and identify side effects.

- **Phase II:**

  A larger group of people around 100 to 300 are recruited in phase II. The purpose of this stage includes checking whether the drug has effect against the disease and further evaluating its safety.

- **Phase III:**

An even larger group of people from 1,000 to 3,000 are involved in phase III. But it can be as many as 10,000 patients. The study drug or treatment is tested to confirm effectiveness, monitor side effects, compare with commonly used treatments, and collect information that will allow it to be used safely.

- **Phase IV:**

  Post-marketing study are implemented to gain addition information containing the risks, benefits, and optimal use of a drug.

1.3 Introduction to multiple comparisons and selection procedures

Multiple comparisons and selection problems are a common topic in many fields. It is ordinary to have more than two groups to compare with at the same time. For example, test treatment effects or toxicity levels of a group of chemical compounds in a dose-response study, analyze consumers' preference for a series of products in market surveys, and compare the yield rates from different systems or manufacturing processes in a quality control study. One typical method to analyze these types of questions is to apply ANOVA table with F test to examine homogeneity among the groups. If the null hypothesis of homogeneity is rejected, however, ANOVA table does not provide further information about which groups are statistically different. The results of F test may not meet requirement. Therefore, researchers are interested in investigating multiple comparisons or other methodologies which furnish them with the relationships among the groups.

When comparing two population means, type I error is defined as the probability of incorrectly rejecting the null hypothesis that two meas are equal. The error rate is set to be protected at or below α. However, the problem of how to meet the probability constraint becomes more complicated when there are more than two groups. Denote k as the total number of groups. When $k \geq 3$, the all-pairwise comparison family set contains $\binom{k}{2}$ tests with multiple pairs of null and alternative hypotheses. $H_0 : \mu_i = \mu_j$ vs. $H_a : \mu_i = \mu_j,\ 1 \leq i,\ j \leq, k,\ i \neq j$. The corresponding type I error is the probability of incorrectly rejecting any null hypothesis which brings about the error inflation problem. If the nominal error rate of an individual test is controlled at α, the exact type I error rate of an all-pairwise comparison family set is actually greater than α. For example, an all-pairwise comparison set has 6 tests in studying four populations. Suppose that type I error of an individual test is maintained not exceeding 5%. Then, the total error rate of six pairwise tests significantly increases to 26.5%.

$$\alpha' = 1 - (1 - \alpha)^g,\ g \text{ is the total number of tests}$$

The fact demonstrates the importance of the field in multiple comparisons and the necessity of controlling α value when $k \geq 3$. Statistical adjustments for multiplicity are appropriate for controlling type I error.

The foundation of the field of multiple comparisons were established in late 1940s and early 1950s. A few principal pioneers are Duncan, Roy, Scheff*é*, and Tukey. Some similar ideas can be traced back to the earlier works by Fisher, Gossett, and

others. Harter (1980)[18] has a detailed description about the early history in multiple comparisons. The books written by Miller (1966)[30] and by Hochberg and Tamhane (1987)[25] provide comprehensive multiple comparisons procedures established in their eras and point out new research directions in the field.

However, there has been abundant debate and controversy over the need for α adjustment to take the multiplicity of inferences into consideration. The ideas can be generally classified into three schools of thought:

- Familywise error rate
- Comparisonwise error rate
- Bayesian

The first school of thought led by Tukey (1953)[40] and Scheff*é* (1953)[36] adjusts the error rate of each individual test. This school deems that it is essential to use multiple comparisons methods which control familywise error rate to ensure that the probability of having at least one false rejection of the null hypothesis does not exceed α. The assigned familywise error rate or so called the experimentwise error rate applies to all of the hypothesis tests in the family set as a whole instead of to a single test. A list of related procedures using adjusted probability are discussed in chapter 2.

On the other hand, many statisticians have opposite point of view on whether using statistical adjustments (e.g. O'Neill and Wetherill (1971)[32], Petersen (1977)[34], Carmer and Walker (1982)[9], O'Brien (1983)[31], Perry (1986)[33], and Rothman

(1990)[35]). Under this school of thought, each test or statistical inference is handled one by one and the probability adjustment is not necessary. The probability of falsely rejecting the null hypothesis of one single test is known to be the comparisonwise error rate. This probability applies to each individual hypothesis test, but not collectively as the familywise error rate.

The last school of thought employs Bayesian approach to access multiple comparisons problems. A few representative works such as Duncan (1965)[11], Waller and Duncan (1969)[41], Duncan and Dixon (1983)[12] use prior distributions for unknown parameters, linear loss functions for an individual test, and an additive loss function for the entire loss.

1.4 Overview of the thesis

A test procedure is preferable if the number of potentially the best treatments selected by the test is small and if the procedure can keep all the most effective treatments at the end of the test as well. The procedure should protect the best treatments from being discarded and prevent concluding too many inferior options as superior ones. The objectives can be achieved from the other side by removing as many inferior treatments as possible.

The goal of this research is to develop methodologies which possesses greater power in detecting strictly inferior treatments while controlling the probability of making an incorrect decision. The main ideas are to modify the critical values and to use observations as feedback to improve the general step-down procedure with constant

thresholds.

The thesis is organized as follows. Chapter 2 reviews multiple comparisons and selection procedures for the response with Normal distribution. Chapter 3 performs a preliminary study including the setting and the properties of the problem. Chapter 4 studies the critical values for the step-down procedure under a restricted parameter space which is a subset of the configurations when the range of treatment means is bounded by a given number. Chapter 5 presents a step-down procedure with feedback which employs observations to maintain type I error for the whole parameter space. Chapter 6 simulates several parameter settings which violate the assumptions of the step-down procedures studies in the thesis.

CHAPTER II

LITERATURE REVIEW ON MULTIPLE COMPARISONS AND SELECTION PROCEDURES FOR NORMAL RESPONSE EXPERIMENTS

Three widely accepted formulation have been developed to approach multiple comparisons, screening, and selection problems: subset selection approach, Indifference-zone approach, and the simultaneous confidence intervals approach. Subset selection approach is a screening scheme which determines a subset of the treatments including at least one of the best ones. The size of the subset created is arbitrary. This approach facilitates the analysis of the experiment when there are a random number of choices. Indifference-zone approach which chooses the best treatments is concerned more with the design of an experiment. In addition to comparing treatments, this methodology develops the scale of an experiment in order to meet the probability requirements in advance. Finally, the simultaneous confidence intervals approach specifies the differences between treatment means. The confidence intervals quantify the magnitudes of discrepancies and control the familywise error rate. The following two books provide the details of these approaches. The first book written by Gibbons, Olkin, and Sobel (1977)[16] describes the procedures clearly and provides a large number of useful tables for implementation. As for the second book written by Bechhofer, Santner, and

Goldsman (1995)[7], it discusses sequential procedures and subset selection formulations more extensively than the first book. If available, the second book provides several options for the same problem and makes recommendations about different alternatives. It assumes that readers are knowledgeable about standard experimental designs. Below is the summary of the three approaches for multiple comparisons and selection problems.

2.1 Subset selection approach

Subset selection is an approach that constructs a random size subset of interested elements. The procedure screens multiple alternatives and selects the desirable options for each question. Gupta (1956)[17] who was a pioneer in this field proposed a single-stage procedure which creates a subset containing the best treatment when it is unique. If there are two or more the best treatments, Gupta's procedure guarantees that at least one of the best treatments are chosen. Denote k as total number of treatments, n as sample size of each treatment, ν as degree of freedom, α as familywise error rate, $\bar{x}_i$ as the sample mean from treatment i, and s^2 as pooled sample variance. Suppose that the treatment with a larger mean is considered to be a more effective therapy, the procedure is to select treatment i into the subset if and only if

$$\bar{x}_i \geq max\{\bar{x}_j,\ 1 \leq j \leq k\} - d_{k-1,\ \nu,\ \alpha}\frac{s}{\sqrt{2/n}}.$$

$d_{k-1,\ \nu,\ \alpha}$ is a predetermined value which controls type I error rate at α. The idea of the procedure is that if a treatment has a sample mean not too far away from the maximum, the associated treatment is perhaps the best one and then is chosen into

the subset.

Gupta's method, however, does not assure that all of the best treatments are selected when there exists several superior treatments. Consequently, removing the treatments in the complement of the subset created may delete both the worse and the best treatments. In contrast, the new methodologies proposed in this study have the ability to construct a subset that contains inferior treatments only which cannot be achieved by Gupta's method.

Selection procedures can be separated into two categories depending on whether the order of hypothesis testing influences the conclusion. First, the single-step procedures are independent on test order. The decision for any hypothesis H_i does not rely on any other hypothesis $H_j,\ i \neq j$. Each hypothesis testing can be carried out individually without being affected by the other tests. On the contrary, the order of the hypotheses is influential to the stepwise procedures. The decisions of the hypotheses in the former steps may affect the decisions of those hypotheses tested later. The order of the tests is ordinarily decided by the magnitude of test statistics or p-values.

2.1.1 The single-step procedures

Bonferroni procedure is a popular single-step procedure which controls the family-wise error rate in the strong sense. The strong control (see Hochberg and Tamhane (1987)[25]) means that the probability of making any type I error of all configurations is controlled at α. Under all-pairwise comparisons, Bonferroni procedure performs $\binom{k}{2}$ t tests at level $\alpha' = \frac{\alpha}{\binom{k}{2}}$ for each test. Any individual hypothesis H_i is rejected if the

corresponding p-value, p_i, is less than α'. Let g be the total number of pairwise comparisons, $g = \binom{k}{2}$. The procedure becomes conservative as g increases which can be seen in Bonferroni inequality.

$$P\left(\bigcap_{i=1}^{g} A_i\right) \geq 1 - \sum_{i=1}^{g} P\left(A_i^c\right)$$

$$\Rightarrow \left(1 - \frac{\alpha}{g}\right)^g \geq 1 - \alpha$$

Bonferroni procedure insures that the overall type I error is less than or equal to α if an individual test has a significance level of $\frac{\alpha}{g}$ when simultaneously testing g pairwise comparisons.

Tukey method can also be considered as a single-step procedure in subset selection. The technique proposed by Tukey will be illustrated in detail later on in section 2.3, the simultaneous confidence intervals approach. These single-step procedures equally treat every hypothesis without taking test order into account.

2.1.2 The stepwise procedures

Stepwise procedures can further be divided into the step-down procedures, the step-up procedures, and the step-up-down procedures. The distinction is the order of the test with which it starts. The step-down procedure first examines the most significant hypothesis while the step-up procedure first tests the least significant one. The properties and the related literature of these three types of procedures are addressed in the following subsections.

2.1.2.1 The step-down procedures

A step-down procedure begins with testing the most significant hypothesis with the largest test statistics or the smallest p-value. The stopping rule is to continue until a hypothesis is not rejected. All of the remaining hypotheses are then accepted without further tests by implication. Typically, this type of procedures use a non-increasing sequence of critical values for successive test steps.

The idea of the step-down procedure can be traced back to the book by Miller (1966)[30]. The article, however, does not provide a proof nor mention the property of controlling the familywise error rate in a strong way. A general method for constructing a step-down test procedure was proposed by Marcus, Peritz, and Gabriel (1976)[29]. Their method is referred to as a closure method and can be used to form an α-level multiple test procedure. The closure method is a technique that constructs tests where the probability of making at least one incorrect assertion is under control. And a procedure is said to be an α-level multiple test procedure if it can meet the strong control condition regardless of how many hypotheses are true or false.

Holm (1979)[26] presented a p-value based step-down procedure which improves the power of Bonferroni procedure. The p-values are first ordered as $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(g)}$ with associated hypotheses $H_{(1)}, H_{(2)}, \cdots, H_{(g)}$. The procedure begins with testing the most significant hypothesis, $H_{(1)}$, and continues in order until an acceptance occurs. $H_{(i)}$ is rejected in the i^{th} step if $p_{(i)} \leq \frac{\alpha}{g-i+1},\ 1 \leq i < g$. Otherwise, the procedure is stopped and accept all of the $H_{(j)}$ where $j \geq i$. The probability criteria are no longer fixed numbers like those in Bonferroni procedure but are dependent on

the sequence of the tests.

Broström (1981)[8] and Finner & Giani (1994)[14] proposed general step-down procedures which choose a subset of the treatments with population means smaller than the maximum by ϵ, $\epsilon > 0$. When ϵ equals zero, Hayter (2007)[23] suggested sharper critical values and provided confidence intervals for the differences in means.

2.1.2.2 The step-up procedures

The order of the step-up procedures is opposite to that of the step-down procedures. A step-up procedure starts with testing the least significant hypothesis with the smallest test statistics or the largest p-value. The termination rule is to stop the procedure when a hypothesis is rejected. Then, the rest of the hypotheses are rejected by implication without further tests. A hypothesis testing, H_m with a p-value of $p_{(m)}$, is performed if and only if all of the hypotheses whose p-values are greater than or equal to $p_{(m)}$ are all retained. The step-up procedure frequently uses a non-decreasing sequence of critical values in the test procedure.

Welsch (1977)[42] mentioned a step-up procedure based on the studentized range statistics for one-way layouts. His method achieve strong control over the familywise error rate. Dunnett and Tamhane (1992)[13] proposed a step-up multiple test procedure which compares test statistics with certain critical points. The procedure can be applied to test a nonhierarchical family of hypotheses with two or more contrasts.

Hochberg (1988)[24] came up with a p-value based step-up procedure. The procedure starts the sequential tests with the least significant hypothesis with the largest

p-value and continues in order until a rejection happens. Similarly, the p-values are first ordered as $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(g)}$ with the corresponding hypotheses $H_{(1)}, H_{(2)}, \cdots, H_{(g)}$. In step 1, suppose that $p_{(g)} > \alpha$, the associated hypothesis $H_{(g)}$ is accepted and proceed to testing $H_{(g-1)}$. In the i^{th} step, $H_{(g-i+1)}$ is accepted if $p_{(g-i+1)} > \frac{\alpha}{i}$, $1 \leq i < g$. Otherwise, the procedure is stopped and reject all of the $H_{(j)}$ where $j \leq g - i + 1$. Although the algorithm of Hochberg's step-up procedure is inverse, the procedure uses the same critical values as those in Holm's procedure. Therefore, Hochberg's step-up procedure always rejects any hypothesis rejected by Holm's step-down procedure. Hochberg's procedure uniformly dominates Holm's procedure in terms of having greater power.

2.1.2.3 The step-up-down procedures

The step-up procedures which begin with testing the minimum statistics is called as a MIN test in Laska and Meisner (1989)[28]. This type of approached concerns whether all of the hypotheses can be rejected. If not, the step-up procedures offer advanced information to identify the acceptable hypotheses. Based on the same concept, the step-down procedures which begin with examining the maximum statistics can be called as a MAX test. The main interest of the procedures is to check whether at least one of the hypotheses can be rejected or not. If the answer is positive, the step-down procedures continue a further study to recognize the rejectable hypotheses.

A more general issue than the topics discussed in the previous two procedures is that "whether at least q hypotheses can be rejected" where q is a number between

1 and the total number of hypotheses, g. This issue is addressed by Tamhane, Liu, and Dunnett (1998)[39]. Denote the ordered test statistics as $t_{(1)} \leq t_{(2)} \leq \cdots \leq t_{(g)}$ with associated hypotheses $H_{(1)}, H_{(2)}, \cdots, H_{(g)}$. Let $r = g + 1 - q$. In the first step, if $t_{(r)} \leq c_r$, accept $H_{(1)}, H_{(2)}, \cdots, H_{(r)}$ and continue with the step-up procedure; otherwise, reject $H_{(r)}, H_{(r+1)}, \cdots, H_{(g)}$ and continue with the step-down procedure. It is apparent that the step-up procedure and the step-down procedure are special cases of the step-up-down procedure when $q = g$ and $q = 1$ respectively.

In general, a single-step procedure has advantages of easy to execute the test procedure, easy to quantify the discrepancy between population means, and easy to construct confidence intervals. However, the power of the test procedure is not very satisfying. On the other hand, the stepwise procedures may make up for power via carrying out more steps. But, it is cumbersome to calculate the critical values in each step. This problem becomes less severe as the development of computers. Which test procedure should a experimenter choose depends on the definition of an error decision and the power of a test. If the power improvement can compensate the work for more complicated procedures, it would be better to use the stepwise procedures.

2.2 Indifference-Zone approach

Indifference-zone approach can be applied to selection problems and can allow a more practical purpose. Two treatments are said to be indifferent when the difference of the associated means is below a certain threshold, δ^*. δ^* is closely connected with the sample size and the required probability of correct selection. Restricted by the issues

such as budget considerations and accuracy rate, the threshold can be considered to be the worthwhile level for detection. For example, drug safety is very important to public health and is strictly regulated by the government. In addition to having significant efficacy against the diseases or the conditions, toxicity level and the other side effects of the drugs are required to be at a low level. A minor increase in toxicity level may cause serious danger to patients. In this case, researchers may prefer setting δ^* small. Consequently, the experiment requires a larger amount of samples in order to meet the predetermined probability requirement for correction selection.

Bechhofer (1954)[3] proposed a single-stage Indifference-zone procedure which selects the treatment associated with the largest sample mean as the best one in a completely randomized design. One disadvantage of his approach is that the procedure can identify only one best treatment even though there may exist several equally effective treatments. Besides, Bechohofer's single-stage Indifference-zone approach is found to be conservative. The procedure sometimes requests a large sample size which is unaffordable for an experimenter under a certain δ^* value and a probability requirement.

The multi-stage or sequential procedures can compensate the problem of a large sample size to construct an affordable design. The main idea is to use the data obtained in the former stages to speculate the true setting of population means. One simple approach is called a closed two-stage procedure with elimination introduce by Cohen (1959)[10], Alam (1970)[2], and Tamhane and Bechhofer (1977, 1979)[37, 38]. A procedure is closed if there is a fixed upper bound on the number of observations to be taken from each population before carrying out an experiment. Otherwise, it is

open. The procedure is said to be eliminating if the data taken in the previous stages can be used to exclude populations from further sampling and consideration.

There are many other literature related to Indifference-zone approach. For example, a closed multi-stage procedure without elimination studied by Bechhofer and Goldsman (1987, 1989)[5, 6]. As for the common but unknown variance case, see Bechhofer, Dunnett and Sobel (1954)[4] for an open two-stage procedure without elimination and Hartmann (1991)[19] for an open multi-stage procedure with elimination. Generally speaking, the multi-stage procedure is preferable over the single-stage procedure in terms of having smaller expected total number of observations used by the procedure.

2.3 The simultaneous confidence intervals approach

The multiple comparisons and the selection problems can also be addressed by formulating a set of simultaneous confidence intervals for the differences between treatment means. The simultaneous confidence intervals approach controls the overall error rate where the confidence intervals jointly cover every comparison at a given level α. A confidence interval is more informative than a hypothesis testing for it gives extra message about the magnitude of the differences. Applying simultaneous confidence intervals can lead to the same information as those from a hypothesis testing but not vice versa. For example, containing number zero inside the confidence interval of the discrepancy in population mean implies that the two treatment means are not statistically different. Meanwhile, the null hypothesis of $H_0: \ \mu_i = \mu_j, \ i \neq j$ cannot be

rejected either. Conversely, if such a null hypothesis is rejected, a hypothesis testing does not quantify the difference of the two population means which is provided by confidence intervals.

Tukey provided the studentized range critical point, $q_{k,\ \alpha,\ \nu}$, for all-pairwise comparisons with k population means.

$$\mu_i - \mu_j \in \left[\bar{x}_i - \bar{x}_j - q_{k,\ \alpha,\ \nu}\frac{s}{\sqrt{n}},\ \bar{x}_i - \bar{x}_j + q_{k,\ \alpha,\ \nu}\frac{s}{\sqrt{n}} \right],\ 1 \leq i, j \leq k,\ i \neq j$$

If the common variance is known, the degree of freedom $\nu = \infty$. If not, $\nu =$ (total sample size $- k$) and the pooled sample variance $s^2 = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2}{\nu}$. In a balance design where $n_i = n,\ 1 \leq i \leq k$, $q_{k,\ \alpha,\ \nu}$ guarantees that the probability of rejecting the null hypothesis of $H_0 :\ \mu_1 = \mu_2 = \cdots = \mu_k$ is exactly α when H_0 is true. The simultaneous confidence intervals control the familywise error rate at α.

The procedure is sometimes referred as the honestly significant difference (HSD) in the literature. Gabriel (1969)[15] showed that Tukey's method offers the tightest intervals for all-pairwise comparisons among all of the procedures which give equal-length intervals in a balanced one-way layout. As for an unbalanced design, Hayter (1984)[21] proved that Tukey's procedure is conservative. Type I error is less than or equal to α.

Beside two-sided confidence intervals, one-sided confidence intervals are more useful to special cases. Suppose that the underlying configuration has an ordered relationship of $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_k$. Then, it is more interesting to construct one-sided confidence intervals with lower bounds on $\mu_i - \mu_j$ for all $i > j$. For instance, it is well known that toxicity level increases as the amount of a dose raises. One application

of the one-sided simultaneous confidence intervals is to study the toxicity level at different dose levels. Hayter (1990)[22] derived the simultaneous confidence intervals with lower bounds and tabulated the critical values for $k \leq 9$ cases. If μ_i's have indeed an ascending order as mentioned, the procedure using one-sided simultaneous confidence intervals is more competent in detecting the difference between treatment means than the procedure using two-sided simultaneous confidence intervals. When comparing three ordered treatment means of $\mu_1 \leq \mu_2 \leq \mu_3$, Hayter, Miwa, and Liu (2001)[20] gave sharper critical values and presented a more efficient procedure which considers directional discrepancy while providing two-sided confidence intervals. A more comprehensive discussion of multiple comparisons procedures can be found in Hsu's book (1996)[27].

CHAPTER III

BACKGROUND INFORMATION OF THE PROBLEM

3.1 Introduction

Suppose there are k populations having independent normal distributions $N(\mu_i,\ \sigma^2)$, $1 \leq i \leq k$. The common variance, σ^2, can be either known or unknown. And the unknown parameter of interest is the location parameter, μ, which is preferred to be large. The treatment possessing the maximum mean value among k treatments is considered as the most effective therapy while the rest are regarded as inferior ones. The best treatment may not be unique. The problem studied in this research is how to set up an efficient procedure to discriminate the worse treatment from the best ones in a balanced design. Specifically, the thesis concentrates on the modification of the step-down procedures which eliminate one treatment at a time.

A step-down procedure categorizes populations into either the non-best subset (NB) or the best subset (NB^c) based on its own guideline. A test procedure is efficient if it can narrow down the number of the treatments which possibly have high efficacy when the performance of treatments is unknown. It is favorable to construct a small NB^c subset or say a large NB subset. However, if the NB^c is too small, the test procedure may exclude actually the best treatments. The problem caused by eliminating possibly the best treatments is more serious than concluding a big group of candidates in this study. So, the objective of the research is stated as:

$$minimize\ |NB^c| \quad (\text{or } maximize\ |NB|)$$

$$subject\ to\ \ P(\text{population j} \in NB|\ \text{population j is the best}) \leq \alpha$$

This study improves the existing step-down procedure to be more efficient in detecting and eliminating inferior treatments. The sharper critical values offered in chapter 4 and a new step-down procedure with feedback introduced in chapter 5 make the general approach less conservative. The new methodologies not only control the familywise error rate but also eliminate more inferior treatments. The case of comparing three treatments, $k = 3$, is focused in the whole study.

3.1.1 Notation

- T_i: i^{th} population, $1 \leq i \leq 3$
- μ_i: location parameter, population mean of T_i, $1 \leq i \leq 3$
- $\mu_{(i)}$: ordered population mean, $\mu_{(1)} \leq \mu_{(2)} \leq \mu_{(3)}$
- σ^2: common population variance
- n: sample size from each population
- ν: degree of freedom
- X_{ij}: j^{th} observation of population i, $X_{ij} \overset{\text{indep}}{\sim} N(\mu_i,\ \sigma^2),\ 1 \leq i \leq 3$
- $\bar{X}_i$: sample mean of population i, $\bar{X}_i \overset{\text{indep}}{\sim} N(\mu_i,\ \frac{\sigma^2}{n}),\ 1 \leq i \leq 3$
- $\bar{X}_{(i)}$: ordered sample mean, $\bar{X}_{(1)} \leq \bar{X}_{(2)} \leq \bar{X}_{(3)}$

- Y_i: random variable $Y_i = \frac{\bar{X}_i}{\sigma/\sqrt{n}} \overset{\text{indep}}{\sim} N(\frac{\mu_i}{\sigma/\sqrt{n}},\ 1) = N(\mu_i^*,\ 1),\ 1 \leq i \leq 3$

- $Y_{(i)}$: ordered random variable with $\mu^*_{(1)} \leq \mu^*_{(2)} \leq \mu^*_{(3)}$

- S^2: pooled sample variance

- U: random variable $U = \frac{S}{\sigma} \sim g(u) = \frac{\nu^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})2^{\frac{\nu}{2}-1}} u^{\nu-1} exp(-\frac{\nu u^2}{2}),\ 0 < u < \infty$

- d_3 and d_2: thresholds in the first and the second step of the step-down procedure respectively, $d_2 < d_3$

- ϕ and Φ: pdf and cdf of a standard normal distribution

- B: a subset containing the true best treatments in the parameter space

- NB: a subset containing the inferior treatments in the decision space

3.1.2 Definitions

- Best treatment: The treatment with $\mu_i = max\{\mu_1,\ \mu_2,\ \mu_3\} = \mu^*$. There may be more than one best treatments.

- Inferior treatment: The treatment with $\mu_i < \mu^*$. There may be more than one inferior treatments.

- Error: An error decision is to select T_i into NB while $\mu_i = \mu^*$.

- Power: The power of the test is the ability of selecting T_i into NB while $\mu_i < \mu^*$.

3.1.3 Parameter space

When $k = 3$, the relationship among three treatment means can be classified as (1) $\mu_{(1)} = \mu_{(2)} = \mu_{(3)}$ with $|\mathrm{B}| = 3$, (2) $\mu_{(1)} < \mu_{(2)} = \mu_{(3)}$ with $|\mathrm{B}| = 2$, or (3) $\mu_{(1)} \leq \mu_{(2)} < \mu_{(3)}$ with $|\mathrm{B}| = 1$ depending on the total number of best treatments. In case (1), three treatments perform equally; in case (2), two treatments associated with $\mu_{(2)}$ and $\mu_{(3)}$ are equally the best; in case (3), only the treatment corresponding to $\mu_{(3)}$ is the most effective one.

Assume that the common variance is 1 for simplicity. In order to present three treatment means in a two dimensional graph, two contrasts $\frac{\mu_2-\mu_1}{\sqrt{2}}$ and $\left(\mu_3 - \frac{\mu_1+\mu_2}{2}\right)\sqrt{\frac{2}{3}}$ are used for x and y-axis respectively. The coordinates take the difference of treatment means divided by the standard deviation of the contrasts. In this way, the parameter space is symmetric shown in Figure 3.1. The center point (0, 0) indicates case (1) with three equally the best treatments. Three solid lines represents case (2) having two best treatments. And the rest of the area stands for case (3) when only one maximum mean exists.

Every parameter setting can be matched to either of the three relationship types. Under the aforementioned definition of error, the subsequent incorrect decision of each relationship type should be handled individually. In case (1) when the configuration maps to the origin point in the parameter space, an error decision is to claim any treatment as inferior. If the true setting has a point locating on the solid line in the parameter space like case (2), it is incorrect to eliminate the treatment or treatments associated with $\mu_{(2)}$, $\mu_{(3)}$ or both. Last, when case (3): $\mu_{(1)} \leq \mu_{(2)} < \mu_{(3)}$ occurs,

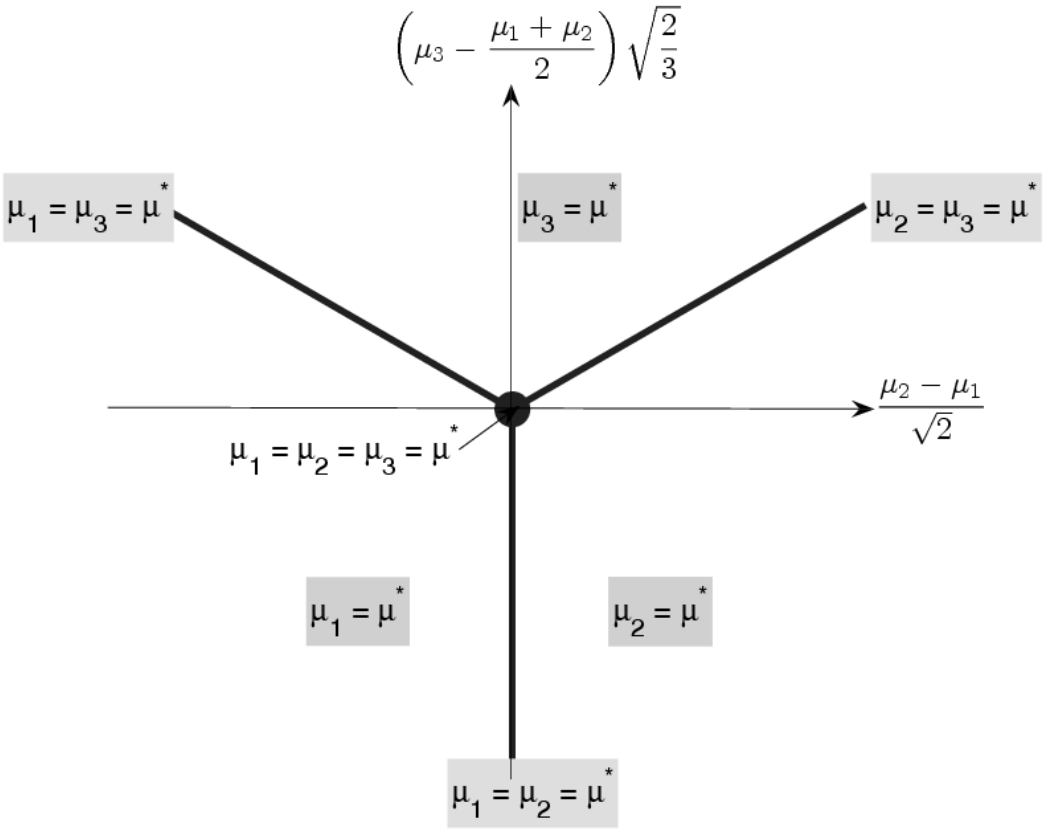

Figure 3.1: Parameter space.

an error decision is to select the treatment with $\mu_{(3)}$ into the NB subset. In consequence, the probability of making an incorrect decision, $P(error)$, has three different formats depending on to which relation type a parameter configuration belongs. Since the true structure of treatment means is unknown, the test procedures must control $P(error)$ for all μ settings in the whole parameter space.

3.2 Procedures

3.2.1 General step-down procedure

A step-down procedure is an approach which starts with testing the most significant hypothesis and continues sequentially as long as a rejection occurs. The general way to carry out a step-down procedure is to use a constant threshold, d_i, at each stage. The values of d_i's are predetermined so that the familywise error rate is controlled at or below α. Suppose that there are k treatments. The testing hypotheses of step i are $H_{0_{(j)}} : \mu_j = \mu^*,\ T_j$ is the best treatment vs. $H_{a_{(j)}} : \mu_j < \mu^*,\ T_j$ is not the best treatment where $\bar{X}_j = \bar{X}_{(i)}$. Population j with $\bar{X}_{(j)}$ is eliminated in step j if $\bar{X}_{(k)} - \bar{X}_{(i)} > d_{k-i+1},$ for all $1 \leq i \leq j < k$. The threshold gets tighter from step to step, $d_k > d_{k-1} > \cdots > d_2$. When $k = 3$, it is a two-step step-down procedure. The procedure takes up to two phases to separate all of the treatments into the NB and the NB^c subsets. The detailed general step-down procedure for the known variance scenario is explained as follows.

[Step 1]

Compare the difference between the maximum and the minimum standardized sample means with d_3.

- If $\frac{\bar{X}_{(3)}-\bar{X}_{(1)}}{\sigma/\sqrt{n}} \leq d_3$, $NB = \{\phi\}$. Terminate the test procedure.
- If $\frac{\bar{X}_{(3)}-\bar{X}_{(1)}}{\sigma/\sqrt{n}} > d_3$, remove T_i corresponding to $\bar{X}_{(1)}$ into NB and continue to Step 2.

[Step 2]

Compare the difference between the maximum and the median standardized sample means with d_2.

- If $\frac{\bar{X}_{(3)}-\bar{X}_{(2)}}{\sigma/\sqrt{n}} \leq d_2$, $NB = \{T_i\}$ where $\bar{X}_i = \bar{X}_{(1)}$. Terminate the test procedure.
- If $\frac{\bar{X}_{(3)}-\bar{X}_{(2)}}{\sigma/\sqrt{n}} > d_2$, remove T_j corresponding to $\bar{X}_{(2)}$ into NB. $NB = \{T_i, T_j\}$ where $\bar{X}_i = \bar{X}_{(1)}$ and $\bar{X}_j = \bar{X}_{(2)}$. Terminate the test procedure.

If variance is not given, σ is substituted with pooled sample standard deviation, S. And the values for d_2 and d_3 when variance is known are different from those when variance is unknown.

For example, suppose that $\frac{\sigma}{\sqrt{n}} = 1$. If observing example (a) in Figure 3.2, the range of the sample means is shorter than d_3. So, stop the test procedure without rejecting any treatment, $NB = \{\phi\}$. Suppose that the relative location of the sample

means is similar to example (b), treatment 1 is eliminated in the first step due to the range of the sample means is greater than d_3. However, the difference between $\bar{X}_2$ and $\bar{X}_3$ is not statistically large enough to discard treatment 2 in the second step of the procedure. As a result, only treatment 1 is selected into the NB subset. As for example (c), both $\bar{X}_3 - \bar{X}_1$ and $\bar{X}_3 - \bar{X}_2$ are greater than the thresholds in step 1 and 2 respectively. Therefore, treatment 1 is eliminated in the first step and treatment 2 is eliminated in the second step of the step-down procedure.

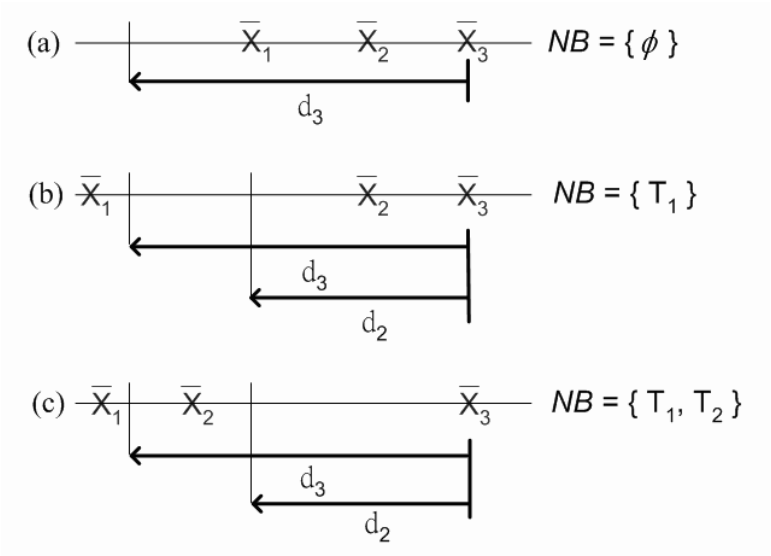

Figure 3.2: Examples of the general step-down procedure when $k = 3$.

3.2.2 Decision space under constant d_2

There are seven possible outcomes after applying the general step-down procedure introduced in the previous subsection when comparing three treatments. NB can be

$\{\phi\}$, $\{T_1\}$, $\{T_2\}$, $\{T_3\}$, $\{T_1,\ T_2\}$, $\{T_1,\ T_3\}$, or $\{T_2,\ T_3\}$ which represents for the conclusion when no treatment is selected into NB, only treatment 1 is identified as an inferior treatment, $\cdots$, or both treatment 2 and 3 are claimed not to be the best treatments. These outcomes divide the decision space into seven subspaces whose shapes rely on the values of the two thresholds and the coordinates. For simplicity, assume that $\frac{\sigma}{\sqrt{n}}$ is known to be 1. Figure 3.3 displays the decision space when the thresholds are constant values with $d_2 < d_3$ and the coordinates are $\frac{\bar{X}_2 - \bar{X}_1}{\sqrt{2}}$ and $\left(\bar{X}_3 - \frac{\bar{X}_1 + \bar{X}_2}{2}\right)\sqrt{\frac{2}{3}}$.

The range of the sample means is less than or equal to d_3 in step 1 if the observed sample means have a matching point inside the hexagon, area (i), of the decision space. Then, the resulting decision is to select no treatment into NB subset. Similarly, any point in the region (ii) of the decision space satisfies $\left\{\bar{X}_1 < \bar{X}_2 - d_3 \text{ and } \bar{X}_2 - d_2 \leq \bar{X}_3 \leq \bar{X}_2\right\}$ or $\left\{\bar{X}_1 < \bar{X}_3 - d_3 \text{ and } \bar{X}_3 - d_2 \leq \bar{X}_2 \leq \bar{X}_3\right\}$. The observations lead to the conclusion of eliminating treatment 1 in the first step but removing no element in the second step. If attaining a point lies within the area (iii) of the decision space, it means that $\left\{\bar{X}_1 < \bar{X}_2 - d_3 \text{ and } \bar{X}_1 \leq \bar{X}_3 < \bar{X}_2 - d_2\right\}$ or $\left\{\bar{X}_3 < \bar{X}_2 - d_3 \text{ and } \bar{X}_3 \leq \bar{X}_1 < \bar{X}_2 - d_2\right\}$. Therefore, the conclusion is to put both treatment 1 and 3 into NB by using the step-down procedure. The rest of the subspaces can be explained by extending the same idea of relative locations of sample means.

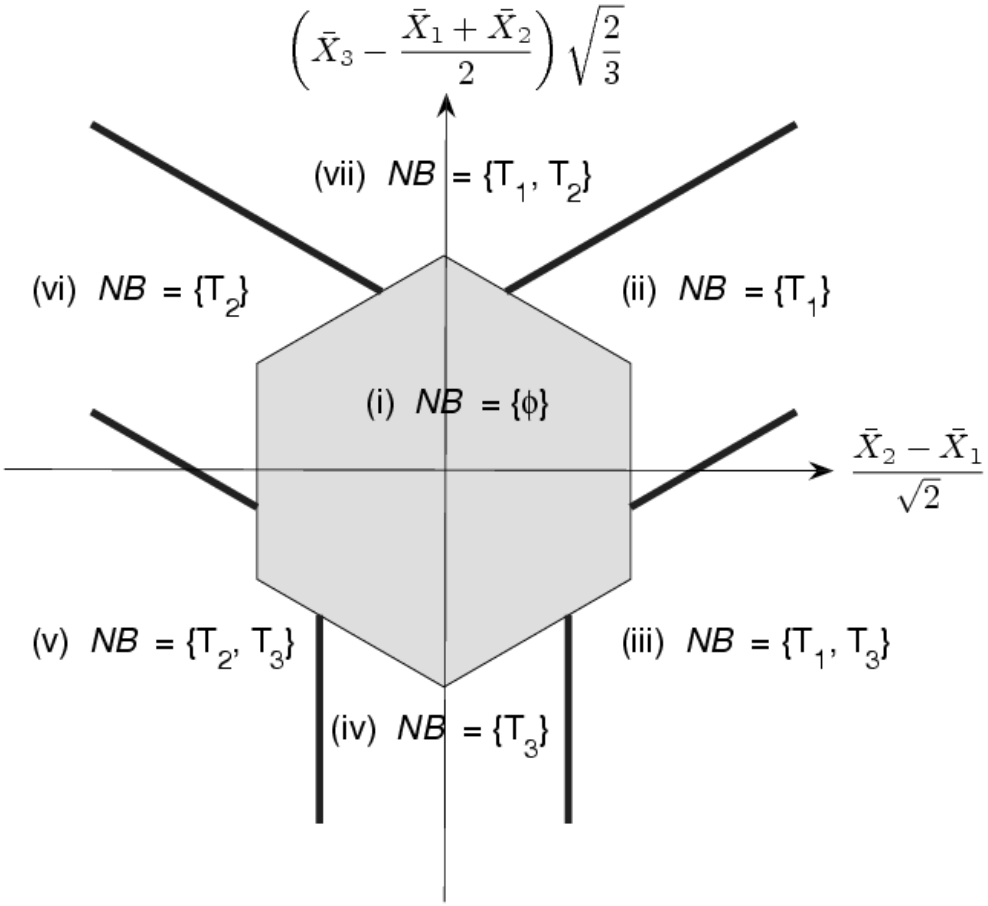

Figure 3.3: Decision space.

3.3 Construction of error rate

How to decide the values for d_2 and d_3 then becomes an critical issue. The fact is that these two thresholds cannot be arbitrary numbers due to the restriction on the error decision rate. If the thresholds are overly short, the general step-down procedure will end up with rejecting too many hypotheses than it should. Some of the most effective treatments may be eliminated as well. The probably of rejecting a best treatment gets out of control in such a case. Therefore, the condition of $P(error) \leq \alpha$ confines the lower bounds of d_2 and d_3. Next, several parameter settings are discussed individually to study the reasonable values for these two thresholds. Again, the following discussion assumes that $\frac{\sigma}{\sqrt{n}} = 1$.

3.3.1 The setting with three equal means: $\mu_1 = \mu_2 = \mu_3$

First, the condition for the relationship type of three equal population means needs to be satisfied. In order not to make an error decision under this type of configuration, the range of the sample means should be less than or equal to d_3 in step 1 of the test procedure. Only in this situation that the procedure will stop without eliminating any treatment. The case like Figure 3.4 which results in a wrong conclusion is undesirable.

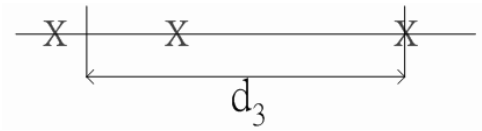

Figure 3.4: One case which leads to an error decision if $\mu_1 = \mu_2 = \mu_3$.

It infers that the size of the hexagon in the decision space cannot be too small. Otherwise, the probability of getting $NB = \{\phi\}$ is less than $1-\alpha$ which is the same as having type I error greater than α. As a result, $P_{\mu_1=\mu_2=\mu_3}(error)$ is a function of d_3. The critical value of d_3 can be minimized and solved by setting $P_{\mu_1=\mu_2=\mu_3}(error) = \alpha$. The formulation is as below.

$$
\begin{aligned}
& P_{\mu_1=\mu_2=\mu_3}(error) \\
= \;& P(max\ \{\bar{X}_1,\ \bar{X}_2,\ \bar{X}_3\} - min\ \{\bar{X}_1,\ \bar{X}_2,\ \bar{X}_3\} > d_3) \\
= \;& 1 - P(max\ \{|Z_i - Z_j|,\ 1 \le i,\ j \le 3\} \le d_3) \\
= \;& 1 - \sum_{k=1}^{3} P(max\ \{|Z_i - Z_j|,\ 1 \le i,\ j \le 3\} \le d_3 | Z_k = min\ \{Z_1,\ Z_2,\ Z_3\}) \\
= \;& 1 - 3\ P(max\ \{|Z_i - Z_j|,\ 1 \le i,\ j \le 3\} \le d_3 | Z_1 = min\ \{Z_1,\ Z_2,\ Z_3\}) \\
= \;& 1 - 3 \int_{z_1=-\infty}^{\infty} \phi(z_1) P(z_1 \le z_i \le z_1 + d_3,\ 2 \le i \le 3)\, dz_1 \\
= \;& 1 - 3 \int_{z=-\infty}^{\infty} \phi(z) \left[\Phi(z + d_3) - \Phi(z)\right]^2 dz \\
= \;& \alpha
\end{aligned}
\tag{3.1}
$$

Tukey[40] proposed studentized range statistics, $q_{k,\ \alpha,\ \nu}$ back to 1953. When variance is known, $q_{3,\ \alpha,\ \infty}$ exactly solves the equation (3.1). Tukey's method guarantees that the familywise error rate of testing $H_0 : \mu_i = \mu_j$ vs. $H_a : \mu_i \neq \mu_j$ for all $1 \le i \le k,\ 1 \le j \le k,\ i \neq j$ is exactly α in a balanced design, and is less than or equal to α in an unbalanced design see Hayter (1984)[21]. Thus, using $d_3 = q_{3,\ \alpha,\ \nu}$ controls the probability of rejecting the null hypothesis of $H_0 :\ \mu_1 = \mu_2 = \mu_3$ is at or

below α when the statement is actually true.

3.3.2 The setting with two best means: $\mu_1 < \mu_2 = \mu_3$

Second, the condition needs to be assured when there exists two best treatments in a $k = 3$ case. Suppose that $\mu_1 < \mu_2 = \mu_3$. A decision is incorrect to claim that treatment 2, 3, or both are inferior treatments in stage 1, 2, or both. The foregoing step-down procedure will make improper conclusion if observing the examples shown in Figure 3.5.

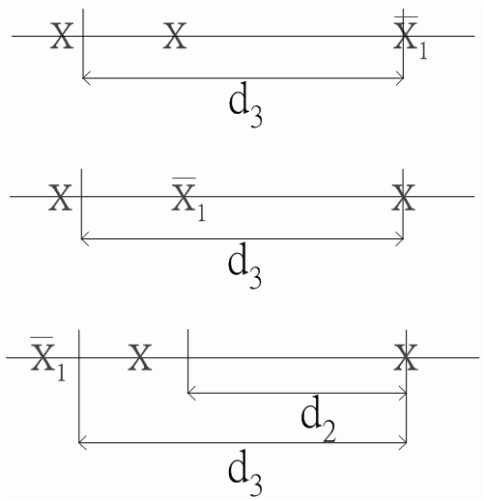

Figure 3.5: Three cases which lead to error decisions if $\mu_1 < \mu_2 = \mu_3$.

The probabilities of concluding $NB = \{T_2\}$, $\{T_3\}$, $\{T_1, T_2\}$, $\{T_1, T_3\}$, and $\{T_2, T_3\}$ all contribute to $P(error)$ under this type of parameter relationship. These five outcomes reject either one or two treatments. Consequently, the thresholds used in both steps of the test procedure are influential. $P_{\mu_1<\mu_2=\mu_3}(error)$ is a function of d_3 and

d_2. Appropriate d_2 and d_3 should be chose so that the probability of getting a point inside region (iii) to (vii) of the decision space is at or below α.

$$
\begin{aligned}
& P_{\mu_1<\mu_2=\mu_3}(error) \\
= \; & P(\bar{X}_2,\ \bar{X}_3 < \bar{X}_1,\ min\ \{\bar{X}_2,\ \bar{X}_3\} < \bar{X}_1 - d_3) \\
& + P(\bar{X}_1 < \bar{X}_2,\ \bar{X}_3 < \bar{X}_2 - d_3) \\
& + P(\bar{X}_1 < \bar{X}_3,\ \bar{X}_2 < \bar{X}_3 - d_3) \\
& + P(\bar{X}_1 < \bar{X}_2 - d_3,\ \bar{X}_2 - d_3 \leq \bar{X}_3 < \bar{X}_2 - d_2) \\
& + P(\bar{X}_1 < \bar{X}_3 - d_3,\ \bar{X}_3 - d_3 \leq \bar{X}_2 < \bar{X}_3 - d_2) \\
= \; & \int_{x_1=-\infty}^{\infty} \phi(x_1-\mu_1) \times \left\{ \Phi(x_1-\mu_2)^2 \right. \\
& \qquad \left. - \left[\Phi(x_1-\mu_2) - \Phi(x_1-d_3-\mu_2)\right]^2 \right\} dx_1 \\
& + 2\int_{x_2=-\infty}^{\infty} \phi(x_2-\mu_2)\Phi(x_2-\mu_1)\Phi(x_2-d_3-\mu_3)\, dx_2 \\
& + 2\int_{x_2=-\infty}^{\infty} \phi(x_2-\mu_2)\Phi(x_2-d_3-\mu_1) \\
& \qquad \times \left[\Phi(x_2-d_2-\mu_3) - \Phi(x_2-d_3-\mu_3)\right] dx_2 \\
\leq \; & \alpha
\end{aligned}
\tag{3.2}
$$

3.3.3 The setting with one best mean: $\mu_1 \leq \mu_2 < \mu_3$

Similarly, both thresholds are influential in satisfying the probability constraint for the parameter relationship with only one best treatment. Assume that $\mu_1 \leq \mu_2 < \mu_3$, it is an error to put treatment 3 into NB in either step 1 or 2 of the step-down

procedure. The examples can be found in Figure 3.6.

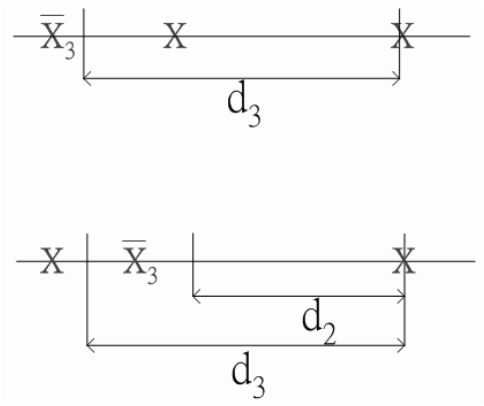

Figure 3.6: Two cases which lead to error decisions if $\mu_1 \leq \mu_2 < \mu_3$.

$NB = \{T_3\}$, $\{T_1,\ T_3\}$, and $\{T_2,\ T_3\}$ are all error decisions for this type of parameter relationship. The chance of locating a point in region (iii) to (v) of the decision space should be controlled at or below α. Therefore, $P_{\mu_1 \leq \mu_2 < \mu_3}(error)$ is a function of d_2 and d_3, too.

$$
\begin{aligned}
& P_{\mu_1 \le \mu_2 < \mu_3}(error) \\
= \;& P(\bar{X}_3 < min\ \{\bar{X}_1,\ \bar{X}_2\},\ \bar{X}_3 < max\ \{\bar{X}_1,\ \bar{X}_2\} - d_3) \\
& + P(\bar{X}_1 < \bar{X}_2 - d_3,\ \bar{X}_1 < \bar{X}_3 < \bar{X}_2 - d_2) \\
& + P(\bar{X}_2 < \bar{X}_1 - d_3,\ \bar{X}_2 < \bar{X}_3 < \bar{X}_1 - d_2) \\
= \;& P(\bar{X}_1 \ge \bar{X}_3,\ \bar{X}_2 \ge \bar{X}_3) - P(\bar{X}_3 \le \bar{X}_i \le \bar{X}_3 + d_3, i = 1,\ 2) \\
& + P(\bar{X}_1 < \bar{X}_3 < \bar{X}_2 - d_3) + P(\bar{X}_1 < \bar{X}_2 - d_3,\ \bar{X}_2 - d_3 \le \bar{X}_3 < \bar{X}_2 - d_2) \\
& + P(\bar{X}_2 < \bar{X}_3 < \bar{X}_1 - d_3) + P(\bar{X}_2 < \bar{X}_1 - d_3,\ \bar{X}_1 - d_3 \le \bar{X}_3 < \bar{X}_1 - d_2) \\
= \;& \int_{x_3=-\infty}^{\infty} \phi(x_3 - \mu_3)\{[1 - \Phi(x_3 - \mu_1)][1 - \Phi(x_3 - \mu_2)] \\
& \qquad -[\Phi(x_3 + d_3 - \mu_1) - \Phi(x_3 - \mu_1)][\Phi(x_3 + d_3 - \mu_2) - \Phi(x_3 - \mu_2)]\}\, dx_3 \\
& + \int_{x_3=-\infty}^{\infty} \phi(x_3 - \mu_3)\Phi(x_3 - \mu_1)[1 - \Phi(x_2 + d_3 - \mu_2)]\, dx_3 \\
& + \int_{x_2=-\infty}^{\infty} \phi(x_2 - \mu_2)\Phi(x_2 - d_3 - \mu_1)[\Phi(x_2 - d_2 - \mu_3) - \Phi(x_2 - d_3 - \mu_3)]\, dx_2 \\
& + \int_{x_3=-\infty}^{\infty} \phi(x_3 - \mu_3)\Phi(x_3 - \mu_2)[1 - \Phi(x_3 + d_3 - \mu_1)]\, dx_3 \\
& + \int_{x_1=-\infty}^{\infty} \phi(x_1 - \mu_1)\Phi(x_1 - d_3 - \mu_2)[\Phi(x_1 - d_2 - \mu_3) - \Phi(x_1 - d_3 - \mu_3)]\, dx_1 \\
\le \;& \alpha
\end{aligned}
\tag{3.3}
$$

Consider the following extreme scenario when $\mu_1 \ll \mu_2 = \mu_3$. Since the population mean of treatment 1 is far below the other two populations means, the observations from treatment 1 tend to be much smaller than those from treatment 2 or 3. Consequently, treatment 1 is almost surely rejected and detected as an inferior treatment in the first step of the test procedure. The judgment of step 1 is correct without

doubt and the test procedure can easily exclude treatment 1. The key issue then becomes how to determine d_2 such that neither treatment 2 nor 3 is eliminated in the second stage of the step-down procedure. This question is the same as comparing two population means which can be solved by z test or t test. The asymptotic value of d_2 turns out to be $z_{\alpha/2}$ when variance is known and $t_{\alpha/2,\ \nu}$ when variance is unknown. The value of d_2 must converge to these two statistics to maintain the error rate at or below α for such an extreme parameter setting in the whole parameter space, for example, $(\mu_1,\ \mu_2,\ \mu_3) = (10^{-6},\ 10^6,\ 10^6)$.

There are several ways to construct $P(error)$ for each type of parameter setting. The advantage of formulating the three $P(error)$'s as mentioned is to have only one integral involved. The formulations of $P_{\mu_1=\mu_2=\mu_3}(error)$, $P_{\mu_1<\mu_2=\mu_3}(error)$, and $P_{\mu_1\leq\mu_2<\mu_3}(error)$ are based on the relative location of sample means. Single integration is actually enough to describe the circumstance of making an incorrect decision. With less integration, the numerical calculation can be done faster and more accurately. Figure 3.7 shows the numerical results of $P(error)$ in one part of the parameter space: $\mu_1 \leq \mu_2 \leq \mu_3$. The rest of the parameter space in other orders can be extended by the symmetry property. The values for d_3 and d_2 used in the graph are $q_{3,\ 0.05,\ \infty}$ and $d_2 = \sqrt{2}z_{\alpha/2}$ respectively. As it can be seen, the error rate is at or below 5% for the whole parameter space.

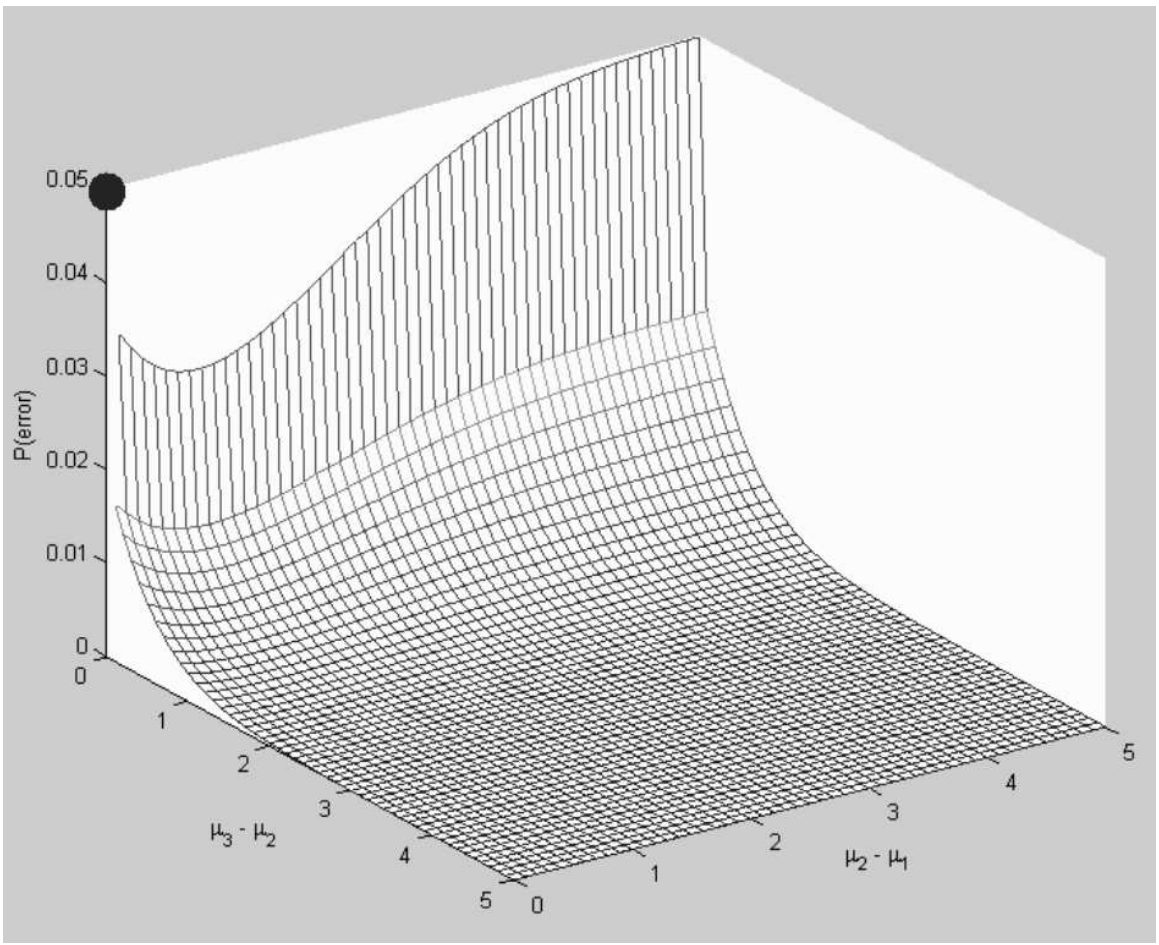

Figure 3.7: $P(error)$ under $d_3 = q_{3,\ 0.05,\ \infty}$ and $d_2 = \sqrt{2}z_{\alpha/2}$.

3.4 Properties of error rate

The goal of this research is to retrieve a small subset of potentially the best treatments by eliminating less effective ones. The treatments selected into NB are then excluded from further study. As a result, it is crucial to discard any possibly the best treatment during the step-down procedure. The appropriate way is to put only the definitely worse treatments into the NB subset. That is, type I error which is the probability of including any best treatment into NB should be protected. As mentioned in the previous section, the way to construct type I error depends on the size of the subset B. Different relationships of treatment means are associated with different forms of $P(error)$: $P_{\mu_{(1)}=\mu_{(2)}=\mu_{(3)}}(error)$, $P_{\mu_{(1)}<\mu_{(2)}=\mu_{(3)}}(error)$, $P_{\mu_{(1)}\leq\mu_{(2)}<\mu_{(3)}}(error)$. This section discusses the properties of these $P(error)$'s when variance is known. Same properties can be attained for the unknown variance case by taking the random variable, $\frac{S}{\sigma}$, into account. To simply notation, the ordered means of $\mu_1 \leq \mu_2 \leq \mu_3$ is used. At the end of this section, an important conclusion about the most determinant parameter configuration to the value of d_2 will be made based on the three properties. Chapter 4 and 5 intensively adopt the concepts presented in this section.

Property I.

Lemma 1. *$P_{\mu_1\leq\mu_2<\mu_3}(error)$ is a decreasing function in δ where $\delta = \mu_3 - \mu_2 > 0$.*

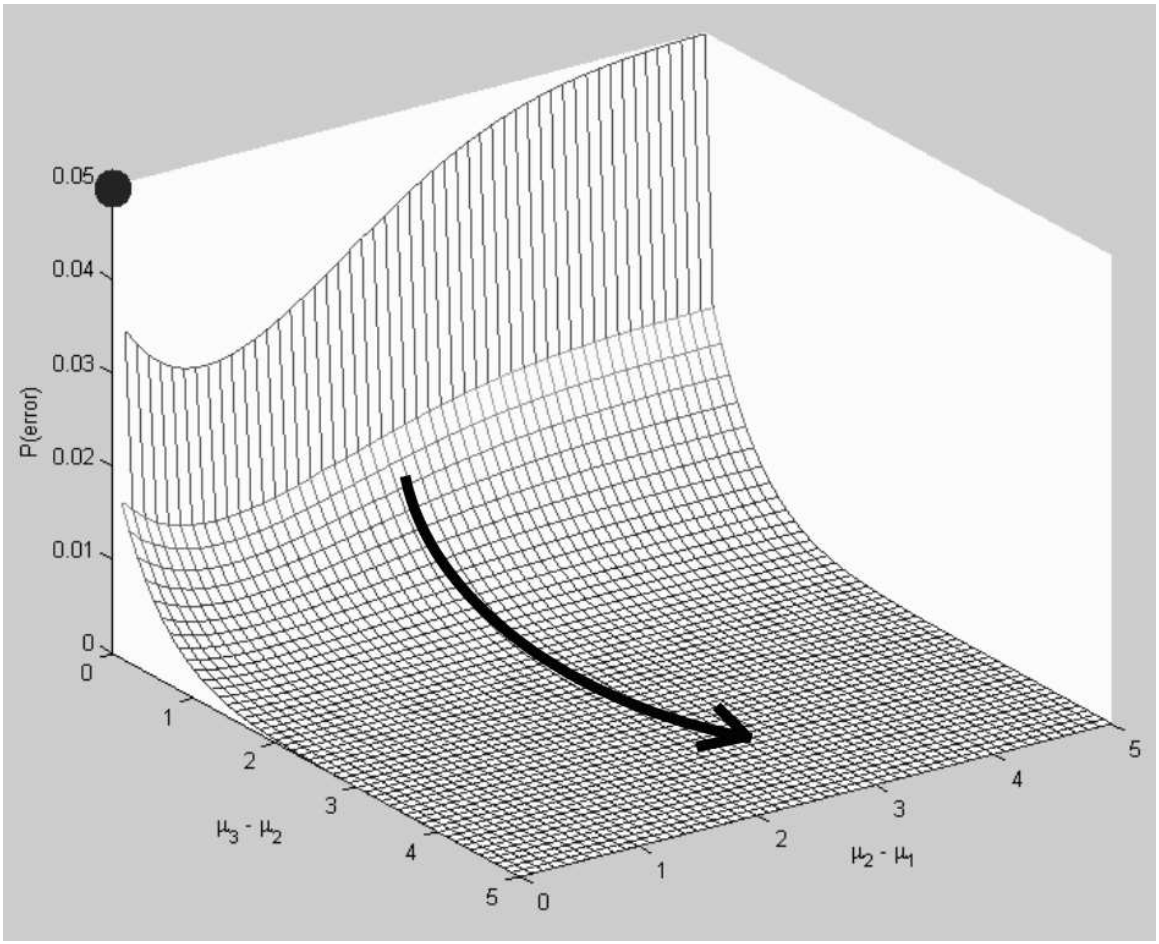

Figure 3.8: Property I of $P(error)$

Proof.

$$
\begin{aligned}
& P_{\mu_1 \le \mu_2 < \mu_3}(error) \\
= \; & P(T_3 \in NB) \\
= \; & P(NB = \{T_3\},\ \{T_1,\ T_3\},\ \text{and}\ \{T_2,\ T_3\}) \\
= \; & P(\frac{\bar{X}_2 - \bar{X}_3}{\sigma/\sqrt{n}} > d_3,\ 0 < \frac{\bar{X}_2 - \bar{X}_1}{\sigma/\sqrt{n}} < d_3) \\
& + P(\frac{\bar{X}_1 - \bar{X}_3}{\sigma/\sqrt{n}} > d_3,\ 0 < \frac{\bar{X}_1 - \bar{X}_2}{\sigma/\sqrt{n}} < d_3) \\
& + P(\frac{\bar{X}_2 - \bar{X}_1}{\sigma/\sqrt{n}} > d_3,\ \frac{\bar{X}_2 - \bar{X}_3}{\sigma/\sqrt{n}} > d_2) \\
& + P(\frac{\bar{X}_1 - \bar{X}_2}{\sigma/\sqrt{n}} > d_3,\ \frac{\bar{X}_1 - \bar{X}_3}{\sigma/\sqrt{n}} > d_2) \\
= \; & \int_{y_2=-\infty}^{\infty} \phi(y_2 - \mu_2^*)\Phi(y_2 - d_3 - \mu_2^* - \frac{\delta}{\sigma/\sqrt{n}})\left[\Phi(y_2 - \mu_1^*) - \Phi(y_2 - d_3 - \mu_1^*)\right] dy_2 \\
& + \int_{y_1=-\infty}^{\infty} \phi(y_1 - \mu_1^*)\Phi(y_1 - d_3 - \mu_2^* - \frac{\delta}{\sigma/\sqrt{n}})\left[\Phi(y_1 - \mu_2^*) - \Phi(y_1 - d_3 - \mu_2^*)\right] dy_2 \\
& + \int_{y_2=-\infty}^{\infty} \phi(y_2 - \mu_2^*)\Phi(y_2 - d_3 - \mu_1^*)\Phi(y_2 - d_2 - \mu_2^* - \frac{\delta}{\sigma/\sqrt{n}})\, dy_2 \\
& + \int_{y_1=-\infty}^{\infty} \phi(y_1 - \mu_1^*)\Phi(y_1 - d_3 - \mu_2^*)\Phi(y_1 - d_2 - \mu_2^* - \frac{\delta}{\sigma/\sqrt{n}})\, dy_1
\end{aligned}
$$

□

Since $\phi(\cdot)$ and $\Phi(\cdot)$ are both non-negative functions with $\Phi(a) \ge \Phi(b)$ for $a \ge b$, the last four terms in the proof are all non-negative. Since $\Phi(-\delta)$ is a monotonic decreasing function with respect to δ, $P_{\mu_1 \le \mu_2 < \mu_3}(error)$ decreases in δ. When the smallest two population means are fixed, the probability of selecting the most effective treatment into NB diminishes as $\mu_{(3)}$ gets away from $\mu_{(1)}$ and $\mu_{(2)}$.

Property II.

Lemma 2. *$P_{\mu_1 \leq \mu_2 = \mu_3}(error)$ is bigger than $\lim_{\delta \to 0} P_{\mu_1 \leq \mu_2 < \mu_3}(error)$ where $\delta = \mu_3 - \mu_2$.*

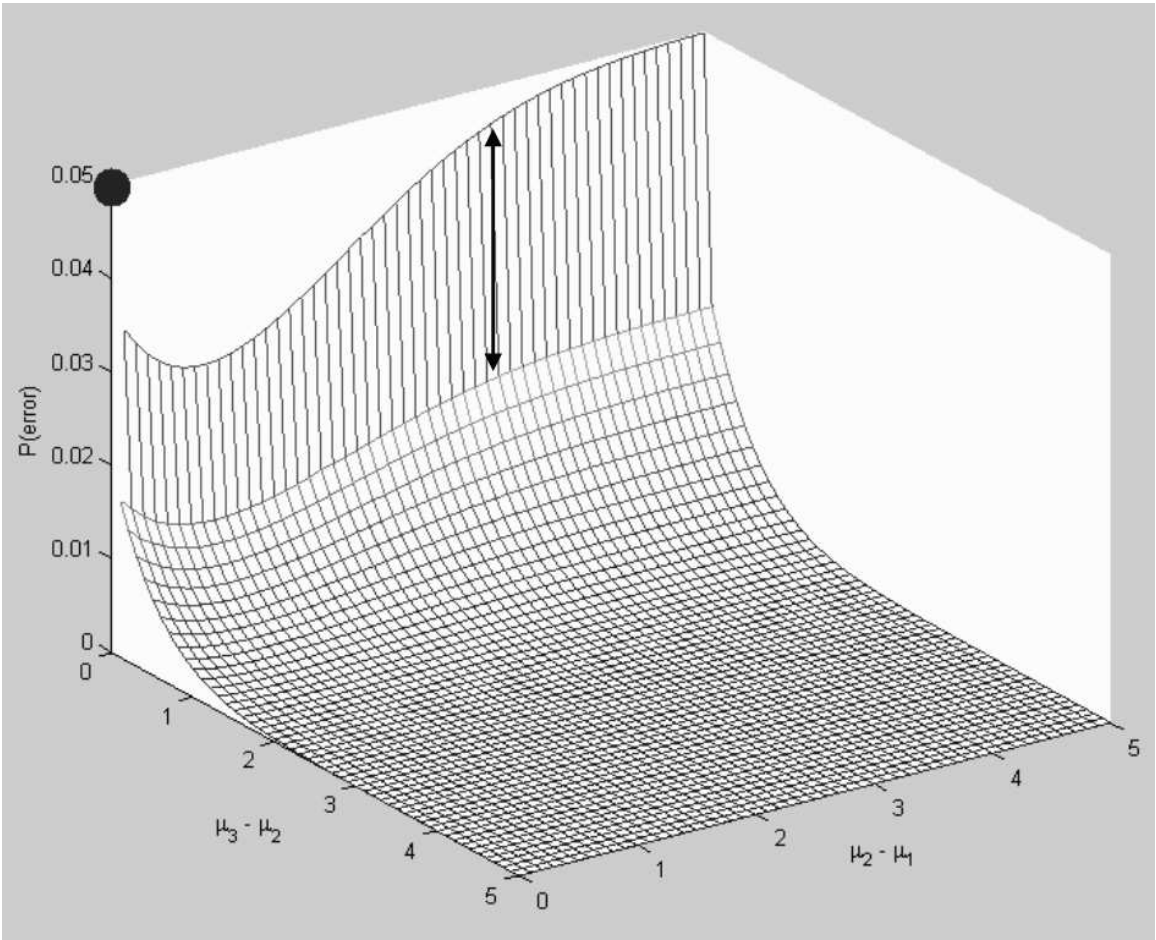

Figure 3.9: Property II of $P(error)$

Proof.

$$
\begin{aligned}
& P_{\mu_1<\mu_2=\mu_3}(error) - \lim_{\delta\to 0} P_{\mu_1<\mu_2<\mu_3}(error) \\
= \;& P(\text{treatment } 2,\ 3, \text{ or both is selected into } NB|\delta=0) \\
& - \lim_{\delta\to 0} P(\text{treatment 3 is selected into } NB|\delta) \\
= \;& P(NB=\{T_2\},\ \{T_3\},\ \{T_1,\ T_2\}, \{T_1,\ T_3\},\ \text{or } \{T_2,\ T_3\}|\delta=0) \\
& - \lim_{\delta\to 0} P(NB=\{T_3\}, \{T_1,\ T_3\},\ \text{or } \{T_2,\ T_3\}|\delta) \\
= \;& P(NB=\{T_2\} \text{ or } \{T_1,\ T_2\}|\delta=0) \\
= \;& P(\bar{X}_1 < \bar{X}_3 - d_3\frac{\sigma}{\sqrt{n}},\ \bar{X}_2 < \bar{X}_3 - d_2\frac{\sigma}{\sqrt{n}}|\delta=0) \\
& + P(\bar{X}_2 < \bar{X}_3 - d_3\frac{\sigma}{\sqrt{n}},\ \bar{X}_3 - d_3\frac{\sigma}{\sqrt{n}} \le \bar{X}_1 < \bar{X}_3|\delta=0) \\
& + P(\bar{X}_2 < \bar{X}_1 - d_3\frac{\sigma}{\sqrt{n}},\ \bar{X}_1 - d_2\frac{\sigma}{\sqrt{n}} \le \bar{X}_3 < \bar{X}_1|\delta=0) \\
= \;& \int_{y_3=-\infty}^{\infty} \phi(y_3-\mu_3^*)\Phi(y_3-d_3-\mu_1^*)\Phi(y_3-d_2-\mu_2^*)\,dy_2 \\
& + \int_{y_3=-\infty}^{\infty} \phi(y_3-\mu_3^*)\Phi(y_3-d_3-\mu_2^*)\times[\Phi(y_3-\mu_1^*)-\Phi(y_3-d_3-\mu_1^*)]\,dy_2 \\
& + \int_{y_1=-\infty}^{\infty} \phi(y_1-\mu_1^*)\Phi(y_1-d_3-\mu_2^*)\times[\Phi(y_1-\mu_3^*)-\Phi(y_1-d_2-\mu_3^*)]\,dy_1 \\
> \;& 0
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
& P_{\mu_1=\mu_2=\mu_3}(error) - \lim_{\delta\to 0} P_{\mu_1=\mu_2<\mu_3}(error) \\
= \;& P(\text{at least one of the treatments are selected into } NB|\delta=0) \\
& - \lim_{\delta\to 0} P(\text{treatment 3 is selected into } NB|\delta) \\
= \;& P(NB=\{T_1\}\ \{T_2\} \text{ or } \{T_1,\ T_2\}|\delta=0) > 0
\end{aligned}
$$

□

The equations of the error rates follow different forms when the parameter configurations have distinct sizes of B. Based on the definition of an error decision, more types of the conclusions account for $P(error)$ when |B| is bigger. If the true parameter setting has a large amount of the best treatments, it is easy to select any of the most effective treatments into NB. Therefore, $P_{\mu_1<\mu_2=\mu_3}(error)$ with $|\mathrm{B}| = 2$ dominates the limiting probability of $P_{\mu_1<\mu_2<\mu_3}(error)$ with $|\mathrm{B}| = 1$; $P_{\mu_1=\mu_2=\mu_3}(error)$ with $|\mathrm{B}| = 3$ dominates the limiting probability of $P_{\mu_1=\mu_2<\mu_3}(error)$ with $|\mathrm{B}| = 1$.

Property III.

The computational result demonstrates that $P_{\mu_1<\mu_2=\mu_3}(error)$ is less than or equal to α when using $d_2 = \sqrt{2}z_{\alpha/2}$ and $d_3 = q_{3,\,\alpha,\,\infty}$ for the known variance case; $d_2 = \sqrt{2}t_{\alpha/2,\,\nu}$ and $d_3 = q_{3,\,\alpha,\,\nu}$, for the unknown variance case. Specifically, $P_{\mu_1\leq\mu_2=\mu_3}(error)$ curve converges to α as treatment range goes to infinity.

3.5 Summary

Considering property I, II, and III all together leads to an important conclusion that the configuration of $\mu_{(1)} < \mu_{(2)} = \mu_{(3)}$ is critical in determining the values of d_2 when comparing three treatments. By setting d_3 to $q_{3,\,\alpha,\,\nu}$ and d_2 to a constant, the scenario with two best treatments has higher decision error rate compared with the rest of the configurations except for the parameter setting with three equal means. $P_{\mu_1=\mu_2=\mu_3}(error)$ is exactly α when $d_3 = q_{3,\,\alpha,\,\nu}$. Assume that $\mu_2 = \mu_3 = \mu_1+\epsilon$, $\epsilon \geq 0$. Property I and II suggest that if $P_{\mu_1\leq\mu_2=\mu_3}(error) \leq \alpha$, then $P(error)$ is maintained

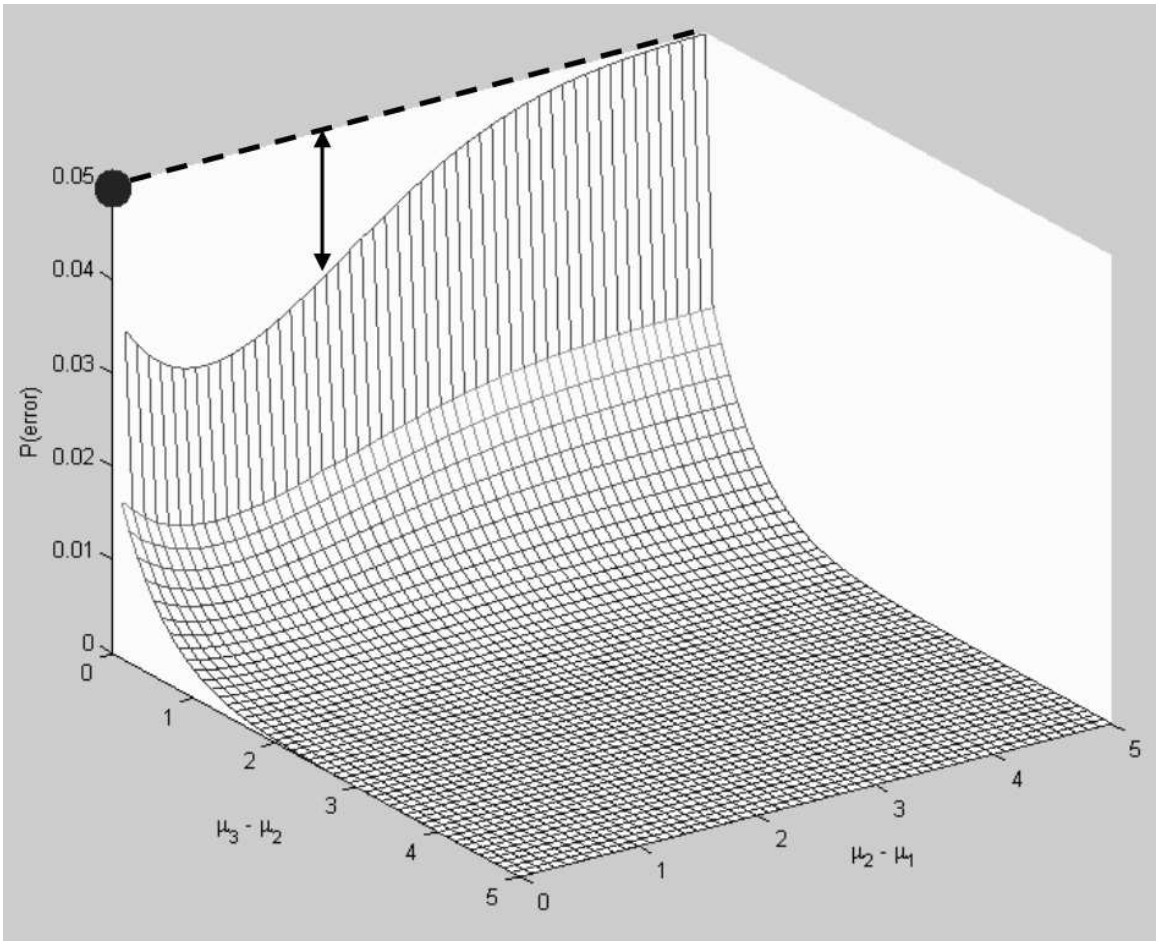

Figure 3.10: Property III of $P(error)$

at or below α for all of the parameter settings where $0 \leq \mu_2 - \mu_1 \leq \epsilon$ and $\mu_3 > \mu_2$ when d_2 and d_3 are fixed numbers.

When the two thresholds are constants, d_3 cannot be shorter than $q_{3,\,\alpha,\,\infty}$ ($q_{3,\,\alpha,\,\nu}$) nor can d_2 be smaller than $\sqrt{2}z_{\alpha/2}$ ($\sqrt{2}t_{\alpha/2,\,\nu}$) for known (unknown) variance situation. Otherwise, the error rate will go beyond α when $\mu_1 = \mu_2 = \mu_3$ and when $\mu_{(1)} \ll \mu_{(2)} = \mu_{(3)}$ respectively. Figure 3.7 shows the response surface of $P(error)$ for the ordered parameter space under $d_3 = q_{3,\,\alpha,\,\infty}$ and $d_2 = \sqrt{2}z_{\alpha/2}$. Although type I error is controlled over every setting, $P(error)$ is very low and even close to 0 for most of the configurations in the graphs. It shows that the procedure with constant thresholds is conservative. Many refinements for d_2 can be made to bring up the response surface closer to the α level. It motivates the usage of nonconstant d_2 to improve the conservative problem.

CHAPTER IV

THE RESTRICTED PARAMETER SPACE APPROACH

The efficacy of treatments in which the researchers are interested may not differ a lot in clinical trials. For example, after screening out thousands of chemical compounds and narrowing down total number of alternatives, the effective levels of those treatments left are quite comparable. In reality, the range of the difference in treatment means is bounded instead of infinite. It implies that only a restricted parameter space is of concern. Accordingly, it is more meaningful to control the error rate and to improve the efficiency of a test procedure in a certain parameter subspace rather than the whole parameter space. In this chapter, a sharper value of d_2 in the second step of the step-down procedure is studied for a balanced design. The shorter threshold enables the test procedure to detect more inferior treatments.

4.1 Motivation

As mentioned in section 3.3, $d_3 = q_{3,\ \alpha,\ \infty}$ along with $d_2 = \sqrt{2}\ z_{\alpha/2}$ control $P(error) \leq \alpha$ for every possible μ vector when comparing three treatments and variance is known. Now, suppose that the upper limit of the difference in treatment mean is given, $\frac{\mu_{(3)}-\mu_{(1)}}{\sigma/\sqrt{n}} \leq \delta,\ \delta > 0$. Consider the following two scenarios with δ_{big} and δ_{small}, $\delta_{big} \gg \delta_{small} > 0$. In the first case when δ is big or even goes to infinity, there is a wide scope of the location of $\mu_{(2)}$. The difference between $\mu_{(3)}$ and $\mu_{(2)}$ can also be large

or even unbounded. Although knowing the relative locations of $\mu_{(1)}$ and $\mu_{(3)}$, there is not much information about $\mu_{(2)}$ in such a case. Therefore, it is suitable to use the existing procedure with $d_2 = \sqrt{2}\ z_{\alpha/2}$ in order to satisfy $P_{\mu_{(1)} \ll \mu_{(2)}=\mu_{(3)}}(error) \leq \alpha$. On the other hand, suppose that the range of three treatment means is known to be small as δ_{small}, the largest two treatment means is surely even closer to each other since $\frac{\mu_{(3)}-\mu_{(2)}}{\sigma/\sqrt{n}} \leq \frac{\mu_{(3)}-\mu_{(1)}}{\sigma/\sqrt{n}} \leq \delta_{small}$. Under this type of parameter setting when three means cluster together, it is difficult to identify the less effective treatments by using standard thresholds in the step-down procedure mentioned before. The conservative problem can also be seen in the response surface of Figure 3.7. For instance, the value of $P_{\mu_1<\mu_2=\mu_3}(error)$ is below α when the difference between μ_3 and μ_1 is small. This fact motivates the adoption of a smaller d_2 value when the range of treatment means is short or bounded.

This chapter applies a constant d_3 and a sharper $d_2(\delta)$ value in the step-down procedure to solve the problem of identifying inferior treatments. The value of $d_2(\delta)$ depends on the range of treatment means. The goal is to study the minimum $d_2(\delta)$ available such that the constraint on type I error is achieved within the restricted parameter space, $\frac{\mu_{(3)}-\mu_{(1)}}{\sigma/\sqrt{n}} \leq \delta$. With the sharper critical values, the test procedure can be more efficient to detect and select inferior treatments into the NB subset.

4.2 Restricted parameter space and new decision space under $d_2(\delta,\ \alpha,\ \nu)$

Same axes, $\frac{\mu_2-\mu_1}{\sqrt{2}}$ and $\left(\mu_3 - \frac{\mu_1+\mu_2}{2}\right)\sqrt{\frac{2}{3}}$, are used for illustrating the restricted parameter space as before. Let $\frac{\sigma}{\sqrt{n}} = 1$ and the upper limit of the range of treatment means be δ, $\delta > 0$. The parameter space of interest reduces from the whole area in Figure 3.1 to the hexagon in Figure 4.1. Every point inside the hexagon maps to a parameter setting with $\mu_{(3)} - \mu_{(1)} \leq \delta$. For example, the three bold solid lines represents for the configurations with two best treatments and $0 < \mu_{(3)} - \mu_{(1)} = \mu_{(2)} - \mu_{(1)} \leq \delta$. Similarly, the three rhombuses in light gray stand for the parameter settings having only one best treatment and $0 \leq \mu_{(2)} - \mu_{(1)} < \mu_{(3)} - \mu_{(1)} \leq \delta$. The $d_2(\delta)$ values proposed in this chapter control familywise error rates over the restricted parameter space inside the hexagon.

$\mu_1 = \mu_2 = \mu_3$ is a common setting having a matching point which lies within the aforementioned restricted parameter space no matter what the positive value of δ is. Therefore, $P_{\mu_1=\mu_2=\mu_3}(error)$ must be less than or equal to α. This probability constraint suggests that d_3 cannot be smaller than $q_{3,\ \alpha,\ \nu}$. Otherwise, the probability of making an incorrect decision is beyond the tolerance if the true setting has three equal means. In a balanced design, the studentized range q statistics guarantees that the $P(error)$ associated with the point $\mu_1 = \mu_2 = \mu_3 = \mu^*$ in the restricted parameter space is exactly α. As a result, the threshold in the first step of the step-down procedure cannot be improved under any restricted parameter space.

As for the rest of the restricted parameter space other than the origin point, however, the corresponding $P(error)$ is below α when d_2 is $z_{\alpha/2}$ (see Figure 3.7).

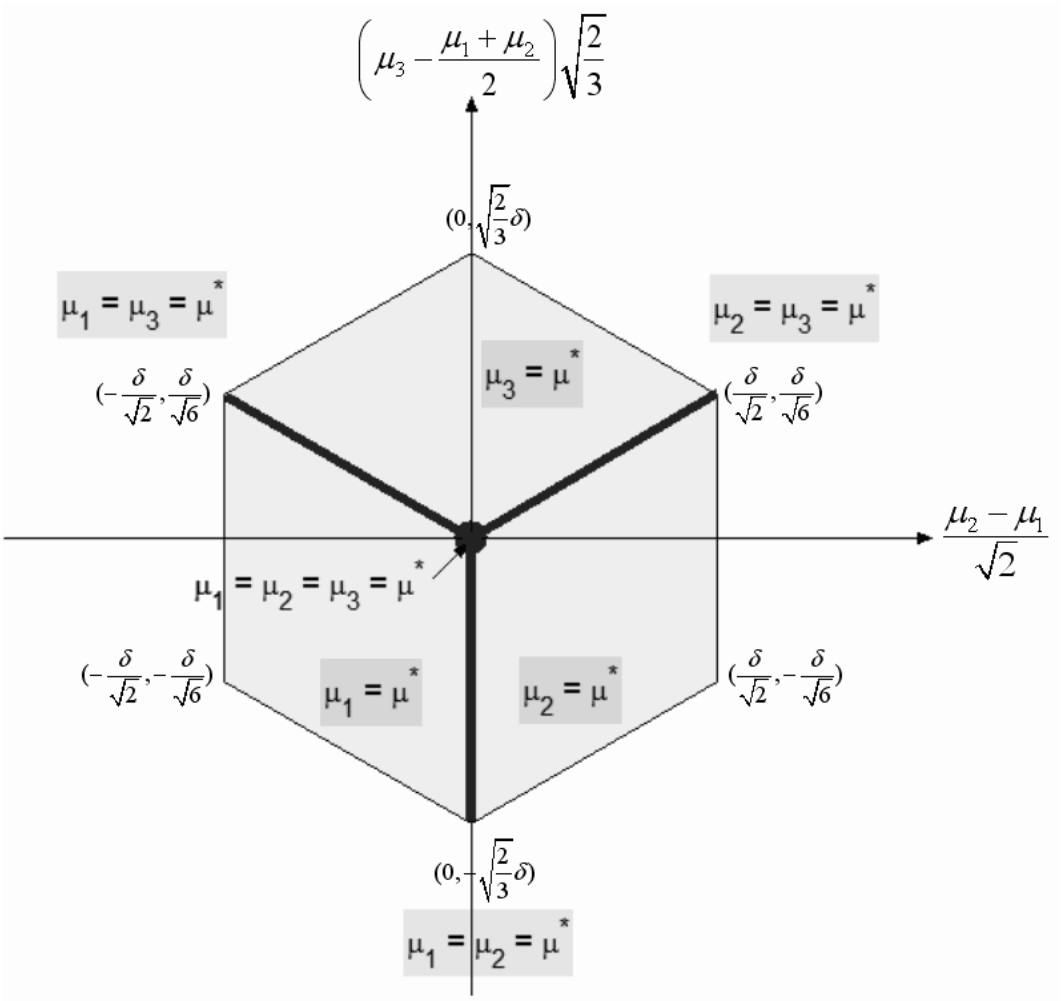

Figure 4.1: The restricted parameter where $\mu_{(3)} - \mu_{(1)} \leq \delta$.

The improvement of power can be made for the area with |B| = 1 or 2. In those parameter settings, d_2 is one of the variables that determines $P_{\mu_1 \leq \mu_2 < \mu_3}(error)$ and $P_{\mu_1 < \mu_2 = \mu_3}(error)$. When d_2 gets shorter, the chance of selecting any inferior treatment into NB becomes greater.

Under the test procedure which adopts a two-step step-down procedure with d_3 and $d_2(\delta)$, the new decision space with a smaller d_2 value is shown in Figure 4.2. Under the constant d_2, the original decision space is divided into seven subregions by a hexagon and three pairs of parallel solid lines. When using a smaller critical value for d_2, each pair of the parallel solid lines move closer to each other to the dashed lines. The area of selecting two treatments into the NB subset increases and the area of eliminating one treatment decreases. As the layout of the decision space changes, the new test procedure with $d_2(\delta)$ has a higher probability to reject treatments than the procedure with constant d_2.

The new $d_2(\delta)$ function needs to satisfy $P(error) \leq \alpha$ within the complete restricted parameter space in a similar way as how the standard d_2 does. The probability of getting a point inside region (iii), (iv) , and (v) bounded by the dashed lines in Figure 4.2 should be less than or equal to α when both μ_1 and μ_2 are less than μ_3. The total area from (iii) to (v) constructed by the dashed lines is larger than that constructed by the solid lines. Since the area increases, it is easier to get a point inside these specific regions. Thus, $P_{\mu_1 \leq\ \mu_2 < \mu_3}(error)$ increases and so does the power of the test procedure. Same idea is extended to the setting of $\mu_1 < \mu_2 = \mu_3$. The total area from (iii) to (vii) bounded by the dashed lines is greater than that bounded by the solid lines. Given the same familywise error rate, the new threshold of $d_2(\delta)$

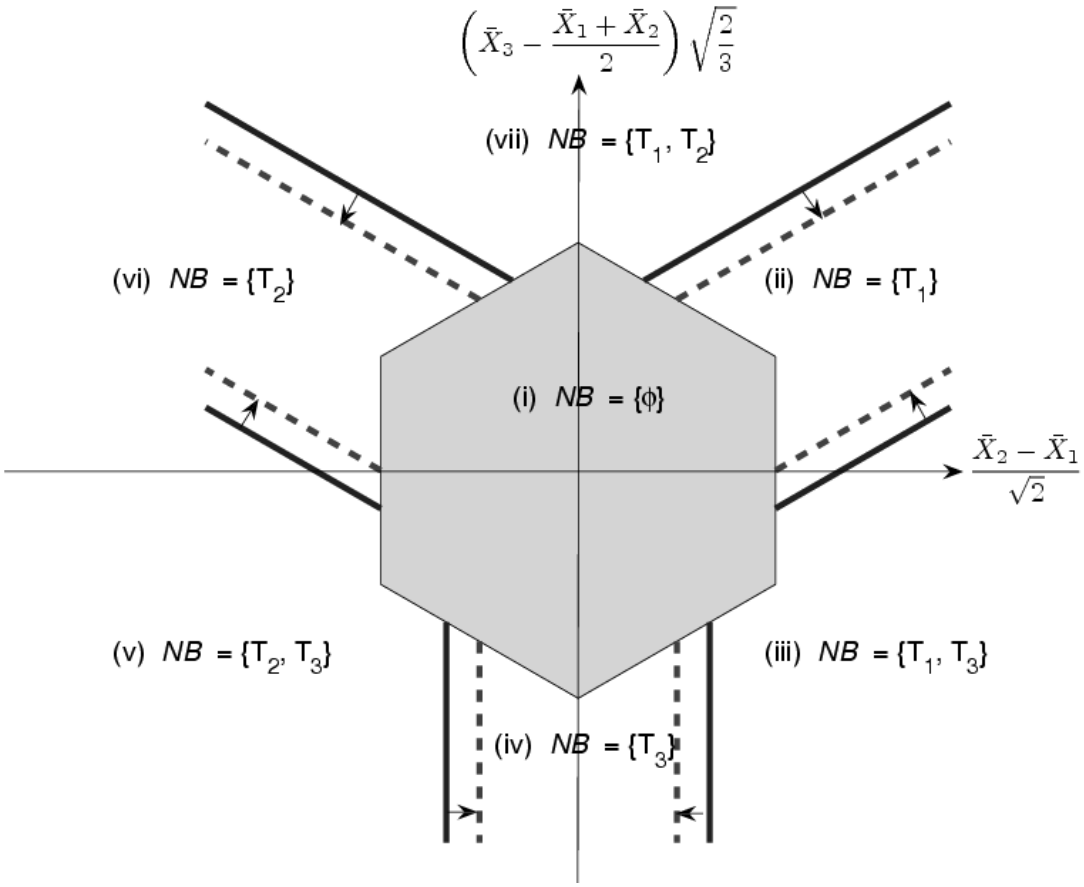

Figure 4.2: The new decision space by using $d_2(\delta)$.

leads to a smaller subset of possibly the best treatments which is useful for further study.

4.3 Construction of $d_2(\delta,\ \alpha,\ \nu)$

Suppose that the prior information about the upper limit of the range in treatment means is given. The value of d_2 can be changed to a smaller constant, $d_2(\delta)$, which depends on the size of the restricted parameter subspace. Let $d_2(\delta,\ \alpha,\ \nu)$ be the minimum value of d_2 such that $P(error)$ is less than or equal to α for all of the settings with $\frac{\mu_{(3)}-\mu_{(1)}}{S/\sqrt{n}} \leq \delta$ and the degree of freedom equals to ν. The value of $d_2(\delta,\ \alpha,\ \nu)$ is bounded from above by the standard value: $\sqrt{2}z_{\alpha/2}$ when variance is known and by $\sqrt{2}t_{\alpha/2,\ \nu}$ when variance is unknown. How far away three treatment means spread out determines the decrement in d_2.

Based on Figure 3.7, the level of $P(error)$ is higher on the boundary of the restricted parameter space where $\mu_{(1)} \leq \mu_{(2)} = \mu_{(3)}$ when using constant thresholds. Therefore, $d_2(\delta,\ \alpha,\ \nu)$ can be calculated by tracing back the inequality of $P_{\mu_{(1)}\leq\mu_{(2)}=\mu_{(3)}}(error) \leq \alpha$ for all $\frac{\mu_{(3)}-\mu_{(1)}}{S/\sqrt{n}} \leq \delta$. Since $P_{\mu_{(1)}=\mu_{(2)}=\mu_{(3)}}(error) = \alpha$ can be reached by setting $d_3 = q_{3,\ \alpha,\ \nu}$ and is irrelevant to d_2, only $P_{\mu_{(1)}<\mu_{(2)}=\mu_{(3)}}(error)$ needs to be confirmed.

The idea can also be seen in the properties of $P(error)$. The conclusion of the three properties in section 3.4 states that controlling the $P(error)$ at $\mu_{(2)} = \mu_{(3)} = \mu_{(1)} + \delta\frac{S}{\sqrt{n}},\ \delta \geq 0$ results in controlling the $P(error)$ for all of the configurations with $\mu_{(2)} - \mu_{(1)} \leq \delta\frac{S}{\sqrt{n}}$. Therefore, in order to guarantee the error rate at each point of

the restricted parameter space, $d_2(\delta,\ \alpha,\ \nu)$ must satisfy $P_{\mu_{(1)}\leq\mu_{(2)}=\mu_{(3)}}(error) \leq \alpha$. Using q statistics as the threshold in the first step of the test procedure insures that $P_{\mu_1=\mu_2=\mu_3}(error)$ is exactly α in a balanced design. Under this consideration, the same d_3 value is used with $d_2(\delta,\ \alpha,\ \nu)$ to construct the probability of selecting any best treatment into NB when the underlying parameter setting has two best treatments. The following two subsections show how $d_2(\delta,\ \alpha,\ \nu)$ is constructed for both known and unknown variance scenarios if extra information about the range of treatment mean is well-known.

4.3.1 Known variance

Suppose that independent sample X_{ij} from treatment i follows $N(\mu_i,\ \sigma^2)$, $1 \leq i \leq 3,\ 1 \leq j \leq n$. μ_i is the unknown parameter of interest and σ^2 is given. Assume that $\frac{\mu_{(3)}-\mu_{(1)}}{\sigma/\sqrt{n}} \leq \delta$ is provided. The $P(error)$ of the most critical case which decides the value for the threshold in the second step of the step-down procedure is as follows. Take the ordered means of $\mu_1 = \mu_{(1)},\ \mu_2 = \mu_{(2)},\ \mu_3 = \mu_{(3)}$ to simply notation.

$$
\begin{aligned}
& P_{\mu_1<\mu_2=\mu_3}(error) \\
= \; & P(\bar{X}_2,\ \bar{X}_3 < \bar{X}_1,\ \frac{\bar{X}_1 - \ min\ \{\bar{X}_2,\ \bar{X}_3\}}{\sigma/\sqrt{n}} > d_3) \\
& + P(\bar{X}_1 < \bar{X}_2,\ \frac{\bar{X}_2 - \bar{X}_3}{\sigma/\sqrt{n}} > d_3) \\
& + P(\bar{X}_1 < \bar{X}_3,\ \frac{\bar{X}_3 - \bar{X}_2}{\sigma/\sqrt{n}} > d_3) \\
& + P(\frac{\bar{X}_2 - \bar{X}_1}{\sigma/\sqrt{n}} > d_3,\ d_2 < \frac{\bar{X}_2 - \bar{X}_3}{\sigma/\sqrt{n}} \le d_3) \\
& + P(\frac{\bar{X}_3 - \bar{X}_1}{\sigma/\sqrt{n}} > d_3,\ d_2 < \frac{\bar{X}_3 - \bar{X}_2}{\sigma/\sqrt{n}} \le d_3) \\
= \; & \int_{y_1=-\infty}^{\infty} \phi(y_1 - \mu_1^*) \times \left\{ \Phi(y_1 - \mu_2^*)^2 \right. \\
& \qquad \left. - \left[\Phi(y_1 - \mu_2^*) - \Phi(y_1 - d_3 - \mu_2^*) \right]^2 \right\} \, dy_1 \\
& + 2 \int_{y_2=-\infty}^{\infty} \phi(y_2 - \mu_2^*) \Phi(y_2 - \mu_1^*) \Phi(y_2 - d_3 - \mu_3^*) \, dy_2 \\
& + 2 \int_{y_2=-\infty}^{\infty} \phi(y_2 - \mu_2^*) \Phi(y_2 - d_3 - \mu_1^*) \\
& \qquad \times \left[\Phi(y_2 - d_2 - \mu_3^*) - \Phi(y_2 - d_3 - \mu_3^*) \right] \, dy_2 \\
\le \; & \alpha \qquad (4.1) \\
& [\ d_3 = q_{3,\ \alpha,\ \infty},\ d_2 = d_2(\delta,\ \alpha,\ \infty)\ ]
\end{aligned}
$$

Given an error rate of α and a range of the treatment means which equals to $\delta\frac{\sigma}{\sqrt{n}}$, the modified threshold, $d_2(\delta,\ \alpha,\ \infty)$, must satisfy the inequality above for all of the parameter settings having $\frac{\mu_3-\mu_1}{\sigma/\sqrt{n}} \le \delta$. The difficulty of obtaining the minimum d_2 value for the restricted parameter space is that there is no closed form for $d_2(\delta,\ \alpha,\ \infty)$. Therefore, the probability constraint should be checked at each $\tau \le \delta$ where $\mu_2 = \mu_3 = \mu_1 + \tau\frac{\sigma}{\sqrt{n}}$. It is a mass search but the search time is not overwhelming due to the advantage of formulating $P(error)$ with one dimensional integration. A numerical

search using bisection algorithm is executed to explore the best value of the threshold in the second step of the test procedure.

During the search, it is found that $P_{\mu_1<\mu_2=\mu_3}(error)$ has a bigger error rate near one end of the restricted parameter space where μ_2 and μ_3 are close to $\mu_1+\delta\frac{\sigma}{\sqrt{n}}$. The phenomenon suggests first finding the smallest d_2 value such that inequality (4.1) holds at the parameter setting of $\mu_2 = \mu_3 = \mu_1 + \delta\frac{\sigma}{\sqrt{n}}$. After that, plug in the solution to confirm that every $P_{\mu_1<\mu_2=\mu_3}(error)$ is at or below α for all $\mu_2 = \mu_3 = \mu_1 + \tau\frac{\sigma}{\sqrt{n}},\ \tau \leq \delta$. In this way, it consumes less time to calculate $d_2(\delta,\ \alpha,\ \infty)$. Figure 4.3 presents the results of $d_2(\delta,\ \alpha,\ \infty)$ at α equals to 5% and 1%.

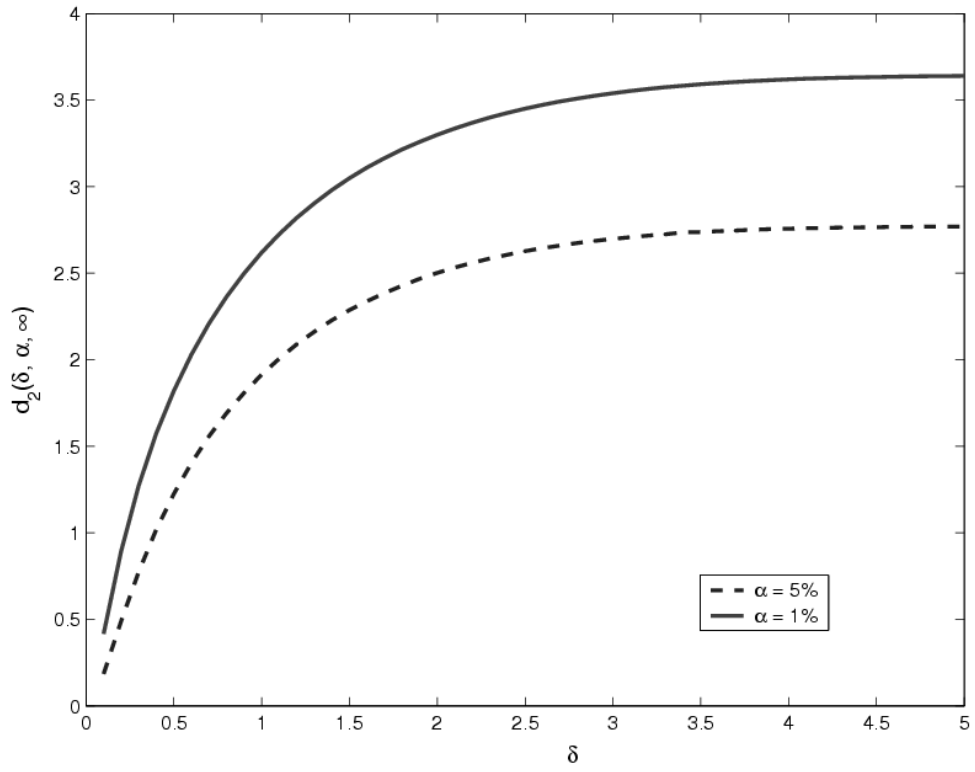

Figure 4.3: $d_2(\delta,\ \alpha,\ \infty)$ values.

4.3.2 Unknown variance

This subsection studies the $d_2(\delta,\ \alpha,\ \nu)$ function when the maximum ratio of the difference in treatment mean to $\frac{\sigma}{\sqrt{n}}$ has a upper limit of δ but the common population variance is unknown. The goal is to separate the best treatments with the largest mean from the inferior treatments with smaller means while controlling type I error. Suppose that in a balanced design, independent observation X_{ij} follows a $N(\mu_i,\ \sigma^2)$ distribution, $1 \leq i \leq 3,\ 1 \leq j \leq n.\ \mu_i$ and σ^2 are both unknown. Then, sample mean, $\bar{X}_i$, follows $N(\mu_i,\ \sigma^2/n)$ and pooled sample variances S^2 follows $\sigma^2 \frac{\chi^2_\nu}{\nu},\ \nu = 3n - 3$. These two sample statistics are point estimates for μ_i and σ^2 respectively. Define a random variable U as $\frac{S}{\sigma}$. Then, U has a distribution of $g(u)$ where

$$g(u) = \frac{\nu^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})2^{\frac{\nu}{2}-1}}\ u^{\nu-1}\ exp\ (-\frac{\nu u^2}{2}),\ 0 < u < \infty.$$

The three properties of $P(error)$ hold for the unknown variance case as well. Let $\mu_1 = \mu_{(1)},\ \mu_2 = \mu_{(2)},\ \mu_3 = \mu_{(3)}$. First, $P_{\mu_1 \leq \mu_2 < \mu_3}(error)$ also decreases as the largest treatment mean increases when variance is unknown. The proof here is similar to the that of property I for the known variance case. The modification is as follows.

Proof.

$$
\begin{aligned}
& P_{\mu_1 \le \mu_2 < \mu_3}(error) \\
= \; & P(\text{treatment 3 is selected into } NB) \\
= \; & P(NB = \{T_3\},\ \{T_1,\ T_3\},\ \text{and } \{T_2,\ T_3\}) \\
= \; & P(\frac{\bar{X}_2 - \bar{X}_3}{S/\sqrt{n}} > d_3,\ 0 < \frac{\bar{X}_2 - \bar{X}_1}{S/\sqrt{n}} < d_3) \\
& + P(\frac{\bar{X}_1 - \bar{X}_3}{S/\sqrt{n}} > d_3,\ 0 < \frac{\bar{X}_1 - \bar{X}_2}{S/\sqrt{n}} < d_3) \\
& + P(\frac{\bar{X}_2 - \bar{X}_1}{S/\sqrt{n}} > d_3,\ \frac{\bar{X}_2 - \bar{X}_3}{S/\sqrt{n}} > d_2) \\
& + P(\frac{\bar{X}_1 - \bar{X}_2}{S/\sqrt{n}} > d_3,\ \frac{\bar{X}_1 - \bar{X}_3}{S/\sqrt{n}} > d_2) \\
= \; & P(\frac{\bar{X}_2 - \bar{X}_3}{\sigma/\sqrt{n}} > Ud_3,\ 0 < \frac{\bar{X}_2 - \bar{X}_1}{\sigma/\sqrt{n}} < Ud_3) \\
& + P(\frac{\bar{X}_1 - \bar{X}_3}{\sigma/\sqrt{n}} > Ud_3,\ 0 < \frac{\bar{X}_1 - \bar{X}_2}{\sigma/\sqrt{n}} < Ud_3) \\
& + P(\frac{\bar{X}_2 - \bar{X}_1}{\sigma/\sqrt{n}} > Ud_3,\ \frac{\bar{X}_2 - \bar{X}_3}{\sigma/\sqrt{n}} > Ud_2) \\
& + P(\frac{\bar{X}_1 - \bar{X}_2}{\sigma/\sqrt{n}} > Ud_3,\ \frac{\bar{X}_1 - \bar{X}_3}{\sigma/\sqrt{n}} > Ud_2) \\
= \; & \int_{u=0}^{\infty} \int_{y_2=-\infty}^{\infty} \phi(y_2 - \mu_2^*) \Phi(y_2 - ud_3 - \mu_2^* - \frac{\delta}{\sigma/\sqrt{n}}) \\
& \qquad \times [\Phi(y_2 - \mu_1^*) - \Phi(y_2 - ud_3 - \mu_1^*)]\ dy_2\, du \\
& + \int_{u=0}^{\infty} \int_{y_1=-\infty}^{\infty} \phi(y_1 - \mu_1^*) \Phi(y_1 - ud_3 - \mu_2^* - \frac{\delta}{\sigma/\sqrt{n}}) \\
& \qquad \times [\Phi(y_1 - \mu_2^*) - \Phi(y_1 - ud_3 - \mu_2^*)]\ dy_2\, du \\
& + \int_{u=0}^{\infty} \int_{y_2=-\infty}^{\infty} \phi(y_2 - \mu_2^*) \Phi(y_2 - ud_3 - \mu_1^*) \Phi(y_2 - ud_2 - \mu_2^* - \frac{\delta}{\sigma/\sqrt{n}})\ dy_2\, du \\
& + \int_{u=0}^{\infty} \int_{y_1=-\infty}^{\infty} \phi(y_1 - \mu_1^*) \Phi(y_1 - ud_3 - \mu_2^*) \Phi(y_1 - ud_2 - \mu_2^* - \frac{\delta}{\sigma/\sqrt{n}})\ dy_1\, du
\end{aligned}
$$

□

Second, the error rate for $\mu_1 \leq \mu_2 = \mu_3$ is bigger than the limiting error rate for $\mu_1 \leq \mu_2 < \mu_3$ as the difference between μ_3 and μ_2 goes to zero when variance is unknown. Based on the definition of error in this research, concluding $NB = \{T_1\},\ \{T_2\},\ \{T_3\},\ \{T_1,\ T_2\},\ \{T_1,\ T_3\}$, or $\{T_2,\ T_3\}$ results in a wrong decision if three treatment means are all equal. Similarly, five decisions of $\{T_2\},\ \{T_3\},\ \{T_1,\ T_2\},\ \{T_1,\ T_3\}$, and $\{T_2,\ T_3\}$ are all error decisions when the setting is $\mu_1 < \mu_2 = \mu_3$. However, only three decisions: $NB = \{T_3\},\ \{T_1,\ T_3\}$, and $\{T_2,\ T_3\}$ are incorrect decisions when the configurations is $\mu_1 \leq \mu_2 < \mu_3$. The larger the size of the most effective treatments is, the more error decisions a parameter setting includes. Since the probability of concluding each decision is nonnegative, $P_{\mu_1 \leq \mu_2 = \mu_3}(error)$ dominates $\lim_{\delta \to 0} P_{\mu_1 \leq \mu_2 < \mu_3}(error)$ when variance is unknown as well. Last, property III of $P_{\mu_1 < \mu_2 = \mu_3}(error) \leq \alpha$ for unknown variance case can also be confirmed by doing numerical calculation.

In conclusion, the most critical issue for finding the smallest d_2 value for the unknown case is to satisfy $P_{\mu_{(1)} \leq \mu_{(2)} = \mu_{(3)}}(error) \leq \alpha$ where $\frac{\mu_{(3)} - \mu_{(1)}}{\sigma/\sqrt{n}} \leq \delta$. In this way, the error rate of the entire restricted parameter space can be controlled. Since $q_{3,\ \alpha,\ \nu}$ statistics makes type I error be exactly α when three treatment means are all equal, it is enough to determine the value of $d_2(\delta,\ \alpha,\ \nu)$ by examining $P_{\mu_{(1)} < \mu_{(2)} = \mu_{(3)}}(error)$ in the restricted parameter space.

Suppose that independent sample X_{ij} from treatment i follows $N(\mu_i,\ \sigma^2)$, $1 \leq i \leq 3,\ 1 \leq j \leq n$. Both μ_i and σ^2 are unknown. But the prior information about the standardized range $\frac{\mu_{(3)} - \mu_{(1)}}{\sigma/\sqrt{n}}$ is given to be less than or equal to δ. Take ordered means for simplicity. The formulation of the $P(error)$ at the most critical setting with two

best treatments is as follows.

$$
\begin{aligned}
& P_{\mu_1<\mu_2=\mu_3}(error) \\
= \;& P(\bar{X}_2,\ \bar{X}_3 < \bar{X}_1, \frac{\bar{X}_1 - \ min\ \{\bar{X}_2,\ \bar{X}_3\}}{S/\sqrt{n}} > d_3) \\
& + 2\ P(\bar{X}_1 < \bar{X}_2,\ \frac{\bar{X}_2 - \bar{X}_3}{S/\sqrt{n}} > d_3) \\
& + 2\ P(\frac{\bar{X}_2 - \ \bar{X}_1}{S/\sqrt{n}} > d_3,\ d_2 < \frac{\bar{X}_2 - \ \bar{X}_3}{S/\sqrt{n}} \leq d_3) \\
= \;& P(\bar{X}_2,\ \bar{X}_3 < \bar{X}_1, \frac{\bar{X}_1 - \ min\ \{\bar{X}_2,\ \bar{X}_3\}}{\sigma/\sqrt{n}} > Ud_3) \\
& + 2\ P(\bar{X}_1 < \bar{X}_2,\ \frac{\bar{X}_2 - \bar{X}_3}{\sigma/\sqrt{n}} > Ud_3) \\
& + 2\ P(\frac{\bar{X}_2 - \ \bar{X}_1}{\sigma/\sqrt{n}} > Ud_3,\ Ud_2 < \frac{\bar{X}_2 - \ \bar{X}_3}{\sigma/\sqrt{n}} \leq Ud_3) \\
= \;& \int_{u=0}^{\infty}\int_{y_1=-\infty}^{\infty} \phi(y_1-\mu_1^*) \times \left\{\Phi(y_1-\mu_2^*)^2 \right. \\
& \qquad \left. - \left[\Phi(y_1-\mu_2^*) - \Phi(y_1-ud_3-\mu_2^*)\right]^2\right\} g(u)\, dy_1\, du \\
& + 2\int_{u=0}^{\infty}\int_{y_2=-\infty}^{\infty} \phi(y_2-\mu_2^*)\Phi(y_2-\mu_1^*) \\
& \qquad \times\Phi(y_2-ud_3-\mu_3^*) g(u)\, dy_2\, du \\
& + 2\int_{u=0}^{\infty}\int_{y_2=-\infty}^{\infty} \phi(y_2-\mu_2^*)\Phi(y_2-ud_3-\mu_1^*) \\
& \qquad \times \left[\Phi(y_2-ud_2-\mu_3^*) - \Phi(y_2-ud_3-\mu_3^*)\right] g(u)\, dy_2\, du \\
\leq \;& \alpha \qquad\qquad (4.2) \\
& [\ d_3 = q_{3,\ \alpha,\ \nu},\ d_2 = d_2(\delta,\ \alpha,\ \nu)\ ]
\end{aligned}
$$

The way to search the minimum value for d_2 is first set $P_{\mu_1<\mu_2=\mu_3}(error)$ to α at $\mu_2 = \mu_3 = \mu_1 + \tau\frac{\sigma}{\sqrt{n}}$ where $\tau = \delta$. Then, verify that the deriving solution meet the probability requirement for all $0 < \tau < \delta$.

4.4 Computational results and performance

4.4.1 Computational results for $d_2(\delta,\ \alpha,\ \nu)$

The results for both known and unknown variance cases at α equals to 10%, 5% and 1% are provided in Table 4.1 to 4.3. The $d_2(\delta,\ \alpha,\ \nu)$ values in the tables insures that $P(error) \leq \alpha$ for the restricted parameter space satisfying $\frac{\mu_{(3)}-\mu_{(1)}}{\sigma/\sqrt{n}} \leq \delta$. The upper limit, δ, are tabulated from 1 to 5.

Table 4.1: $d_2(\delta,\ 10\%,\ \nu)$ table: critical values of d_2 at $\alpha = 0.1$.

		$d_2(\delta)$ at $\alpha = 10\%$					
ν	d_3	$\delta = 1$	$\delta = 2$	$\delta = 3$	$\delta = 4$	$\delta = 5$	$\delta = \infty$
6	3.558	1.747	2.428	2.660	2.730	2.746	2.748
9	3.316	1.676	2.303	2.514	2.577	2.591	2.592
12	3.204	1.644	2.246	2.447	2.506	2.519	2.521
15	3.140	1.625	2.212	2.409	2.466	2.478	2.479
18	3.098	1.612	2.191	2.384	2.440	2.451	2.452
24	3.047	1.598	2.165	2.353	2.407	2.418	2.420
30	3.017	1.589	2.150	2.335	2.388	2.399	2.400
45	2.978	1.578	2.129	2.311	2.363	2.374	2.375
60	2.959	1.572	2.120	2.300	2.351	2.361	2.363
120	2.930	1.565	2.105	2.283	2.333	2.343	2.344
∞	2.902	1.556	2.091	2.266	2.315	2.325	2.326

Table 4.2: $d_2(\delta,\ 5\%,\ \nu)$ table: critical values of d_2 at $\alpha = 0.05$.

		$d_2(\delta)$ at $\alpha = 5\%$					
ν	d_3	$\delta = 1$	$\delta = 2$	$\delta = 3$	$\delta = 4$	$\delta = 5$	$\delta = \infty$
6	4.339	2.257	3.068	3.349	3.437	3.457	3.460
9	3.948	2.124	2.852	3.102	3.179	3.198	3.199
12	3.773	2.067	2.756	2.991	3.063	3.080	3.081
15	3.673	2.034	2.702	2.927	2.997	3.013	3.014
18	3.609	2.014	2.666	2.887	2.954	2.970	2.971
24	3.532	1.987	2.637	2.838	2.903	2.917	2.919
30	3.486	1.972	2.598	2.809	2.873	2.887	2.888
45	3.428	1.953	2.565	2.770	2.833	2.847	2.848
60	3.399	1.945	2.547	2.753	2.815	2.827	2.829
120	3.356	1.930	2.527	2.726	2.786	2.799	2.800
∞	3.314	1.914	2.502	2.698	2.757	2.770	2.772

Table 4.3: $d_2(\delta,\ 1\%,\ \nu)$ table: critical values of d_2 at $\alpha = 0.01$.

		$d_2(\delta)$ at $\alpha = 1\%$					
ν	d_3	$\delta = 1$	$\delta = 2$	$\delta = 3$	$\delta = 4$	$\delta = 5$	$\delta = \infty$
6	6.331	3.564	4.670	5.079	5.202	5.243	5.243
9	5.428	3.160	4.111	4.452	4.560	4.596	4.596
12	5.046	3.004	3.881	4.193	4.294	4.320	4.320
15	4.836	2.930	3.752	4.045	4.143	4.167	4.167
18	4.703	2.862	3.673	3.951	4.047	4.071	4.071
24	4.546	2.797	3.569	3.840	3.932	3.955	3.955
30	4.455	2.765	3.517	3.775	3.866	3.889	3.889
45	4.339	2.719	3.443	3.700	3.781	3.804	3.804
60	4.282	2.689	3.410	3.659	3.740	3.762	3.762
120	4.200	2.661	3.355	3.600	3.680	3.702	3.702
∞	4.120	2.618	3.300	3.540	3.620	3.640	3.643

If the difference between treatment means is small, the threshold of $d_2(\delta,\ \alpha,\ \nu)$ can be shorter than the standard value. For example, d_2 can be reduced by 35% provided that $\frac{\mu_{(3)}-\mu_{(1)}}{\sigma/\sqrt{n}} \leq 1$ at six degree of freedom and $\alpha = 5\%$. The difference between $d_2(\delta,\ \alpha,\ \nu)$ and the standard d_2 becomes minor as δ gets greater than 3. The improvement of d_2 is more significant when degree of freedom is larger, especially when variance is known. When δ equals to ∞, $d_2(\delta,\ \alpha,\ \nu)$ is exactly the standard constant, $t_{\alpha/2,\ \nu}$, which guarantees the error rate of the whole parameter space. Moreover, under same degree of freedom and δ, the ratio of $d_2(\delta,\ \alpha, \nu)$ to the standard d_2 value is lower when α is high.

4.4.2 Power of $d_2(\delta,\ \alpha,\ \nu)$

There are several ways to define the power of a test procedure depending on the goal of a problem. In this study, the power of a test procedure is defined as the ability to detect any less effective treatment with small μ. Two types of measurement for power are discussed throughout the thesis: (1) the probability of identifying any inferior treatment and (2) the expected size of inferior treatments selected into the NB subset. Consequently, the way to formulate the power depends on the parameter setting.

When comparing three treatments, there are seven decisions available: $NB = \{\phi\},\ \{T_1\},\ \{T_2\},\ \{T_3\},\ \{T_1, T_2\},\ \{T_1, T_3\}$ and $\{T_2, T_3\}$. Denote $P_{\{\phi\}}(\mu),\ P_{\{T_1\}}(\mu), \cdots,$ and $P_{\{T_2,T_3\}}(\mu)$ as the probabilities of selecting no treatment, the treatment with $\mu_1, \cdots,$ and the treatments with μ_2 and μ_3 into NB respectively. Table 4.4 lists

Table 4.4: Power calculation.

μ setting	Power	
	(1) $P(i \in NB \mid i \notin \text{B},\ 1 \le i \le 3)$	(2) $E[i \in NB \mid i \notin \text{B},\ 1 \le i \le 3]$
$\mu_1 = \mu_2 = \mu_3$	$P_{\{\phi\}}(\mu)$	0
$\mu_1 < \mu_2 = \mu_3$	$P_{\{T_1\}}(\mu) + P_{\{T_1,T_2\}}(\mu) + P_{\{T_1,T_3\}}(\mu)$	$P_{\{T_1\}}(\mu) + P_{\{T_1,T_2\}}(\mu) + P_{\{T_1,T_3\}}(\mu)$
$\mu_1 \le \mu_2 < \mu_3$	$P_{\{T_1\}}(\mu) + P_{\{T_2\}}(\mu) + P_{\{T_1,T_2\}}(\mu) +$ $+\ P_{\{T_1,T_3\}}(\mu) + P_{\{T_2,T_3\}}(\mu)$	$P_{\{T_1\}}(\mu) + P_{\{T_2\}}(\mu) + 2 \times P_{\{T_1,T_2\}}(\mu) +$ $+P_{\{T_1,T_3\}}(\mu) + P_{\{T_2,T_3\}}(\mu)$

the measurement of power for three types of configurations according to the size of the best treatments. The test procedure which possesses higher quantity in the measurement is said to be more efficient. Measurement (2) is computed as the probability of identifying any inferior treatment multiplies the number of the inferior treatments successfully selected into NB. For example, suppose that $\mu_1 \le \mu_2 < \mu_3$ with $|\text{B}| = 1$. Both treatment 1 and 2 are less effective than treatment 3. Any decision which contains either $T_1,\ T_2$ or both has correct detection. $NB = \{T_1\},\ \{T_2\},\ \{T_1, T_3\}$ and $\{T_2, T_3\}$ are the conclusions which correctly eliminate one of the inferior treatments and thus the corresponding probabilities are multiplied by one. Concluding $NB = \{T_2, T_3\}$ precisely detect both inferior treatments, so the associated probability is weighted by two.

The calculation of power is determined by the probability of getting each decision, $P_{\{\phi\}},\ P_{\{T_1\}}, \cdots,$ and $P_{\{T_2,T_3\}}$. Thus, it is necessary to study the probability of deriving each decision first. Without loss of generality, assume that $\frac{\sigma}{\sqrt{n}} = 1$. The total probability of getting these seven combinations is 1. A more general case with arbitrary sample size and known or unknown variance can be easily extended from

these formulas below.

$$
\begin{aligned}
P_{\{\phi\}}(\mu) &= P(\bar{X}_{(3)} - \bar{X}_{(1)} \le d_3) \\
&= P(\bar{X}_1 < \bar{X}_2,\ \bar{X}_3 \le \bar{X}_1 + d_3)) \\
&\quad + P(\bar{X}_2 < \bar{X}_1,\ \bar{X}_3 \le \bar{X}_2 + d_3)) \\
&\quad + P(\bar{X}_3 < \bar{X}_1,\ \bar{X}_2 \le \bar{X}_3 + d_3)) \\
&= \int_{x_1=-\infty}^{\infty} \phi(x_1 - \mu_1)\left[\Phi(x_1 + d_3 - \mu_2) - \Phi(x_1 - \mu_2)\right] \\
&\qquad\qquad \times \left[\Phi(x_1 + d_3 - \mu_3) - \Phi(x_1 - \mu_3)\right] dx_1 \\
&\quad + \int_{x_2=-\infty}^{\infty} \phi(x_2 - \mu_2)\left[\Phi(x_2 + d_3 - \mu_1) - \Phi(x_2 - \mu_1)\right] \\
&\qquad\qquad \times \left[\Phi(x_2 + d_3 - \mu_3) - \Phi(x_2 - \mu_3)\right] dx_2 \\
&\quad + \int_{x_3=-\infty}^{\infty} \phi(x_3 - \mu_3)\left[\Phi(x_3 + d_3 - \mu_1) - \Phi(x_3 - \mu_1)\right] \\
&\qquad\qquad \times \left[\Phi(x_3 + d_3 - \mu_2) - \Phi(x_3 - \mu_2)\right] dx_3
\end{aligned} \tag{4.3}
$$

$$
\begin{aligned}
P_{\{T_1\}}(\mu) &= P(\bar{X}_1 < \bar{X}_2 - d_3,\ \bar{X}_2 - d_2 \le \bar{X}_3 < \bar{X}_2) \\
&\quad + P(\bar{X}_1 < \bar{X}_3 - d_3,\ \bar{X}_3 - d_2 \le \bar{X}_2 < \bar{X}_3) \\
&= \int_{x_2=-\infty}^{\infty} \phi(x_2 - \mu_2)\Phi(x_2 - d_3 - \mu_1)[\Phi(x_2 - \mu_3) - \Phi(x_2 - d_2 - \mu_3)]\, dx_2 \\
&\quad + \int_{x_3=-\infty}^{\infty} \phi(x_3 - \mu_3)\Phi(x_3 - d_3 - \mu_1)[\Phi(x_3 - \mu_2) - \Phi(x_3 - d_2 - \mu_2)]\, dx_3
\end{aligned} \tag{4.4}
$$

$$
\begin{aligned}
P_{\{T_2\}}(\mu) &= P(\bar{X}_2 < \bar{X}_1 - d_3,\ \bar{X}_1 - d_2 \le \bar{X}_3 < \bar{X}_1) \\
&\quad + P(\bar{X}_2 < \bar{X}_3 - d_3,\ \bar{X}_3 - d_2 \le \bar{X}_1 < \bar{X}_3) \\
&= \int_{x_1=-\infty}^{\infty} \phi(x_1-\mu_1)\Phi(x_1-d_3-\mu_2)[\Phi(x_1-\mu_3)-\Phi(x_1-d_2-\mu_3)]\,dx_1 \\
&\quad + \int_{x_3=-\infty}^{\infty} \phi(x_3-\mu_3)\Phi(x_3-d_3-\mu_2)[\Phi(x_3-\mu_1)-\Phi(x_3-d_2-\mu_1)]\,dx_3
\end{aligned}
\tag{4.5}
$$

$$
\begin{aligned}
P_{\{T_3\}}(\mu) &= P(\bar{X}_3 < \bar{X}_1 - d_3,\ \bar{X}_1 - d_2 \le \bar{X}_2 < \bar{X}_1) \\
&\quad + P(\bar{X}_3 < \bar{X}_2 - d_3,\ \bar{X}_2 - d_2 \le \bar{X}_1 < \bar{X}_2) \\
&= \int_{x_1=-\infty}^{\infty} \phi(x_1-\mu_1)\Phi(x_1-d_3-\mu_3)[\Phi(x_1-\mu_2)-\Phi(x_1-d_2-\mu_2)]\,dx_1 \\
&\quad + \int_{x_2=-\infty}^{\infty} \phi(x_2-\mu_2)\Phi(x_2-d_3-\mu_3)[\Phi(x_2-\mu_1)-\Phi(x_2-d_2-\mu_1)]\,dx_2
\end{aligned}
\tag{4.6}
$$

$$
\begin{aligned}
P_{\{T_1,T_2\}}(\mu) &= P(\bar{X}_1,\ \bar{X}_2 < \bar{X}_3 - d_3) \\
&\quad + P(\bar{X}_1 < \bar{X}_3 - d_3,\ \bar{X}_3 - d_3 \le \bar{X}_2 < \bar{X}_3 - d_2) \\
&\quad + P(\bar{X}_2 < \bar{X}_3 - d_3,\ \bar{X}_3 - d_3 \le \bar{X}_1 < \bar{X}_3 - d_2) \\
&= \int_{x_3=-\infty}^{\infty} \phi(x_3-\mu_3)\Phi(x_3-d_3-\mu_1)\Phi(x_3-d_3-\mu_2)\,dx_3 \\
&\quad + \int_{x_3=-\infty}^{\infty} \phi(x_3-\mu_3)\Phi(x_3-d_3-\mu_1) \\
&\qquad\qquad \times[\Phi(x_3-d_2-\mu_2)-\Phi(x_3-d_3-\mu_2)]\,dx_3 \\
&\quad + \int_{x_3=-\infty}^{\infty} \phi(x_3-\mu_3)\Phi(x_3-d_3-\mu_2) \\
&\qquad\qquad \times[\Phi(x_3-d_2-\mu_1)-\Phi(x_3-d_3-\mu_1)]\,dx_3
\end{aligned}
\tag{4.7}
$$

$$
\begin{aligned}
P_{\{T_1,T_3\}}(\mu) &= P(\bar{X}_1,\ \bar{X}_3 < \bar{X}_2 - d_3) \\
&\quad + P(\bar{X}_1 < \bar{X}_2 - d_3,\ \bar{X}_2 - d_3 \le \bar{X}_3 < \bar{X}_2 - d_2) \\
&\quad + P(\bar{X}_3 < \bar{X}_2 - d_3,\ \bar{X}_2 - d_3 \le \bar{X}_1 < \bar{X}_2 - d_2) \\
&= \int_{x_2=-\infty}^{\infty} \phi(x_2 - \mu_2)\Phi(x_2 - d_3 - \mu_1)\Phi(x_2 - d_3 - \mu_3)\, dx_2 \\
&\quad + \int_{x_2=-\infty}^{\infty} \phi(x_2 - \mu_2)\Phi(x_2 - d_3 - \mu_1) \\
&\qquad\qquad \times [\Phi(x_2 - d_2 - \mu_3) - \Phi(x_2 - d_3 - \mu_3)]\, dx_2 \\
&\quad + \int_{x_2=-\infty}^{\infty} \phi(x_2 - \mu_2)\Phi(x_2 - d_3 - \mu_3) \\
&\qquad\qquad \times [\Phi(x_2 - d_2 - \mu_1) - \Phi(x_2 - d_3 - \mu_1)]\, dx_2 \qquad (4.8)
\end{aligned}
$$

$$
\begin{aligned}
P_{\{T_2,T_3\}}(\mu) &= P(\bar{X}_2,\ \bar{X}_3 < \bar{X}_1 - d_3) \\
&\quad + P(\bar{X}_2 < \bar{X}_1 - d_3,\ \bar{X}_1 - d_3 \le \bar{X}_3 < \bar{X}_1 - d_2) \\
&\quad + P(\bar{X}_3 < \bar{X}_1 - d_3,\ \bar{X}_1 - d_3 \le \bar{X}_2 < \bar{X}_1 - d_2) \\
&= \int_{x_1=-\infty}^{\infty} \phi(x_1 - \mu_1)\Phi(x_1 - d_3 - \mu_2)\Phi(x_1 - d_3 - \mu_3)\, dx_1 \\
&\quad + \int_{x_1=-\infty}^{\infty} \phi(x_1 - \mu_1)\Phi(x_1 - d_3 - \mu_2) \\
&\qquad\qquad \times [\Phi(x_1 - d_2 - \mu_3) - \Phi(x_1 - d_3 - \mu_3)]\, dx_1 \\
&\quad + \int_{x_1=-\infty}^{\infty} \phi(x_1 - \mu_1)\Phi(x_1 - d_3 - \mu_3) \\
&\qquad\qquad \times [\Phi(x_1 - d_2 - \mu_2) - \Phi(x_1 - d_3 - \mu_2)]\, dx_1 \qquad (4.9)
\end{aligned}
$$

The power of the new procedure can be calculated by substituting d_2 with the tabulated $d_2(\delta,\ \alpha,\ \nu)$ value into the aforementioned equations. The computational results demonstrate that the new test procedure is more powerful than the standard procedure in terms of both measurements. Figure 4.4 to 4.8 present the power improvement by using the new methodology. The graphs demonstrate the difference of the two approaches under three types of parameter setting: (1) $\mu_{(1)} = \mu_{(2)} < \mu_{(3)} : (0, 0, \delta)$ (2) $\mu_{(1)} < \mu_{(2)} = \mu_{(3)} : (0, \delta, \delta)$ (3) $\mu_{(1)} < \mu_{(2)} < \mu_{(3)} : (0, \frac{\delta}{2}, \delta)$. The graphs in the left column illustrate the power improvement in terms of the probability of identifying any inferior treatment while those in the right column are in terms of the expected size of any inferior treatment selected into the NB subset.

If the range of three treatment means is known to be small, the $d_2(\delta,\ \alpha,\ \nu)$ function provides sharper critical values that enable the test procedure to detect more inferior treatments in the step-down procedure. When α and the degree of freedom are fixed, the amount of the power improvement increases as the range of $\frac{\mu_{(3)}-\mu_{(1)}}{\sigma/\sqrt{n}}$ decreases. As δ gets smaller, the area of the parameter subspace whose $P(error)$ needs to be controlled becomes tighter. Therefore, a smaller d_2 value which results in a larger decision error rate can still guarantee $P(error) \leq \alpha$ for the restricted parameter space. Then the power raises along with type I error. When δ and degree of freedom are fixed to given numbers, greater improvement in identifying less effective treatments can be achieved if tolerance of familywise error rate is relaxed to a larger value. Moreover, power in both measurements increase more under high degree of freedom when δ and α remain constants.

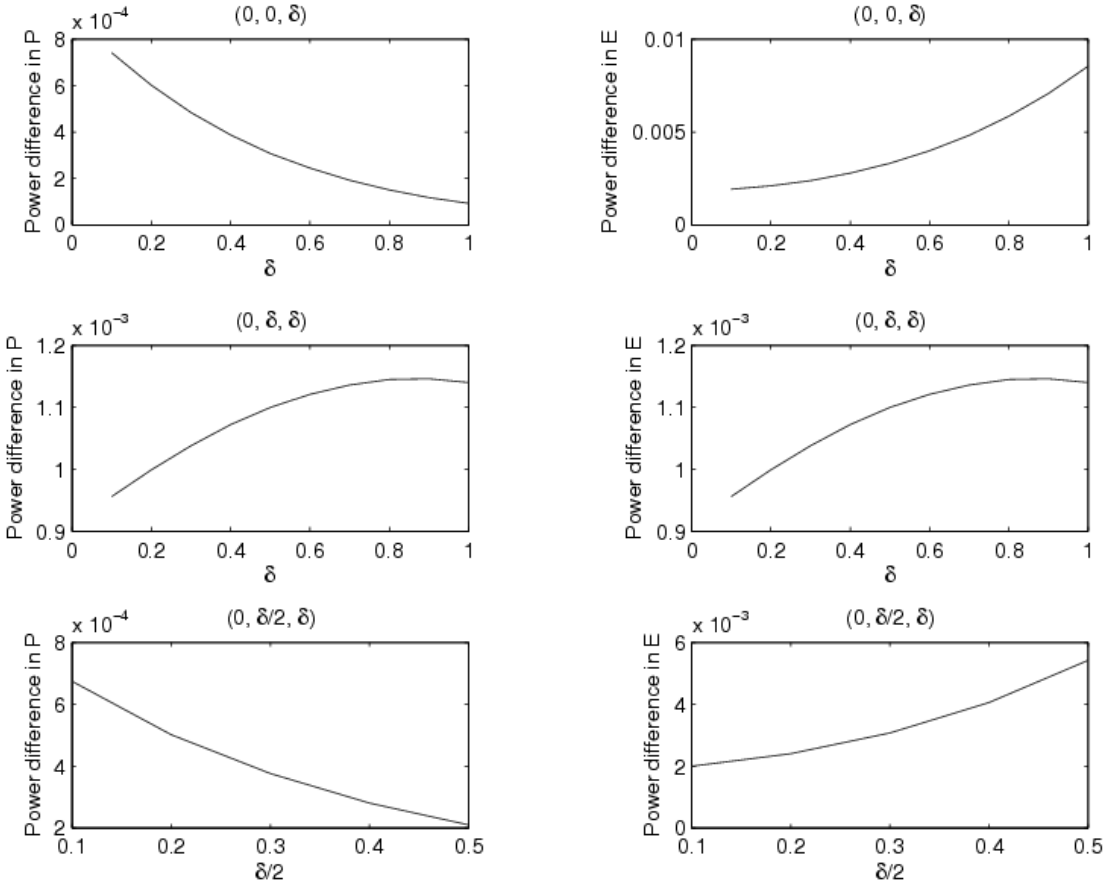

Figure 4.4: Power improvement by using $d_2(\delta = 1)$ at $\alpha = 1\%$ and known variance.

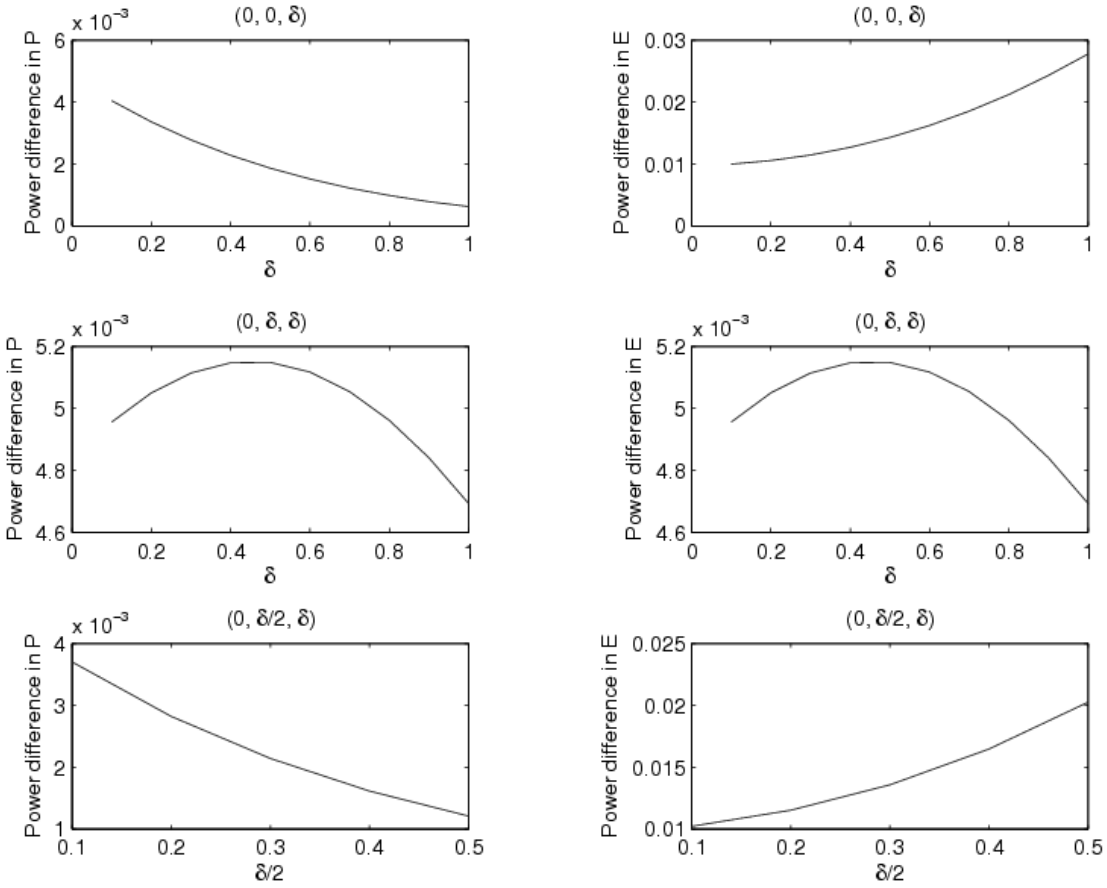

Figure 4.5: Power improvement by using $d_2(\delta = 1)$ at $\alpha = 5\%$ and known variance.

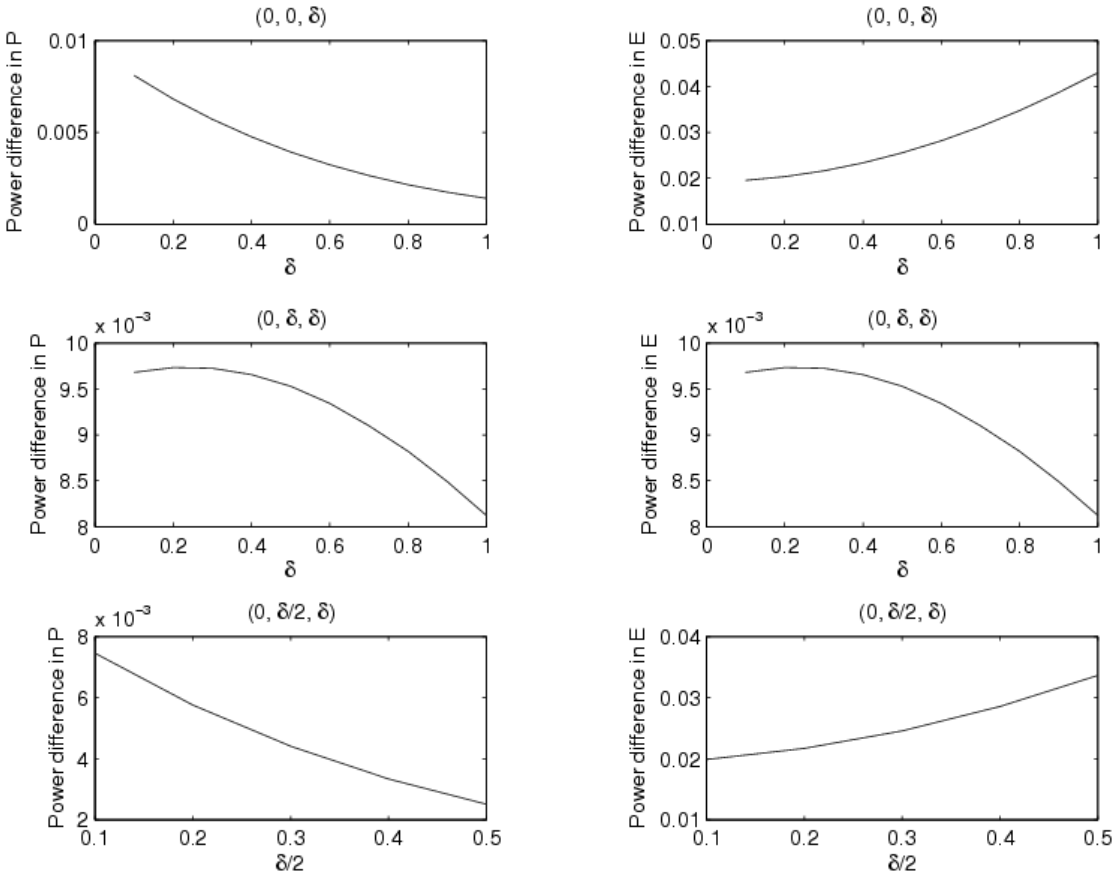

Figure 4.6: Power improvement by using $d_2(\delta = 1)$ at $\alpha = 10\%$ and known variance.

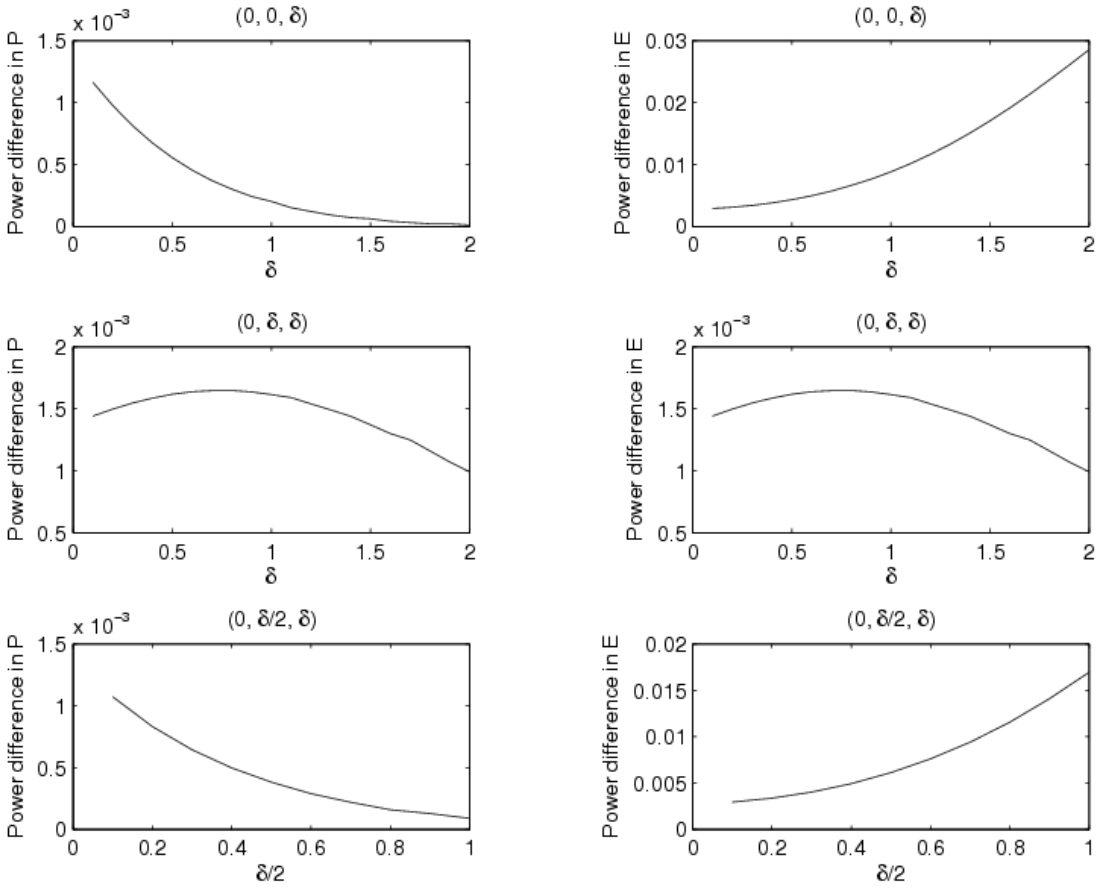

Figure 4.7: Power improvement by using $d_2(\delta = 2)$ at $\alpha = 5\%$ and known variance.

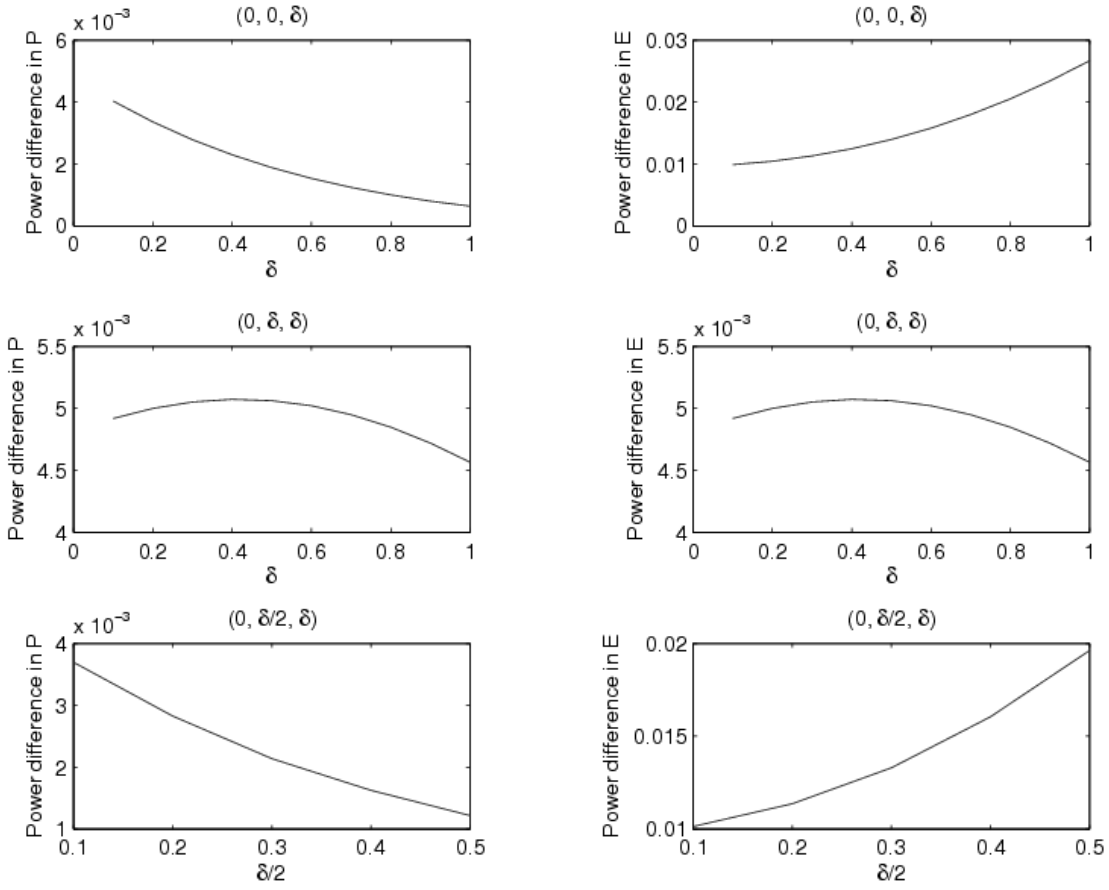

Figure 4.8: Power improvement by using $d_2(\delta = 1)$ at $\alpha = 5\%$ and $\nu = 60$.

4.5 Discussion and summary

The treatment with a smaller mean value is considered as less effective and should be eliminated. Suppose that the prior information of $\frac{\mu_{(3)}-\mu_{(1)}}{\sigma/\sqrt{n}} \leq \delta$, $\delta > 0$ is provided before testing the strictly inferior treatments. With the additional information about the range of treatment means, the test procedure can be less conservative by monitoring the $P(error)$ within a restricted parameter space. The goal of this chapter is to design an efficient test procedure which can identify more inferior treatments when the differences among μ_i's are known to be bounded. The objective can be reached by using sharper thresholds, $d_2(\delta,\ \alpha,\ \nu)$, for the restricted parameter space.

$d_2(\delta,\ \alpha,\ \nu)$ controls $P(error)$ for every configuration with $\frac{\mu_{(3)}-\mu_{(1)}}{\sigma/\sqrt{n}} \leq \delta$. The critical value is smaller than standard d_2, $\sqrt{2}t_{\alpha/2,\nu}$. The calculation time for $d_2(\delta,\ \alpha,\ \infty)$ is not overwhelming due to the benefit of formulating the error rate with one integral. The new threshold is computed by finding the minimum d_2 value such that the error rate of $\mu_{(1)} < \mu_{(2)} = \mu_{(3)}$ is not greater than α when $\mu_{(3)} - \mu_{(1)} \leq \delta\frac{\sigma}{\sqrt{n}}$. If $\frac{\sigma}{\sqrt{n}}$ is one, the solution maintains familywise error rate inside the restricted parameter space of the hexagon in Figure 4.1.

The new $d_2(\delta,\ \alpha,\ \nu)$ value, in fact, can guarantee a subspace larger then the hexagon. Property I and II of $P(error)$ states that if $P_{\mu_{(1)}\leq\mu_{(2)}=\mu_{(3)}}(error)$ is at or below α for some $\tau = \mu_{(3)} - \mu_{(1)}$, then $P_{\mu_{(1)}\leq\mu_{(2)}<\mu_{(3)}}(error)$ is also below α whenever $\mu_{(2)} - \mu_{(1)} < \tau$. Therefore, the tabulated $d_2(\delta,\ \alpha,\ \nu)$ values actually control the error rate for a restricted parameter space of $\frac{\mu_{(2)}-\mu_{(1)}}{\sigma/\sqrt{n}} \leq \delta$ as shown in Figure 4.9.

The advantage of applying $d_2(\delta,\ \alpha,\ \nu)$ to the step-down procedure is that not

only is the error rate controlled over the interested parameter subspace, but also does the power of the test procedure increase. There is a higher chance to exclude the inferior treatments among three. Meanwhile, the expected size of the less effective treatments being rejected is bigger than that by applying standard d_2.

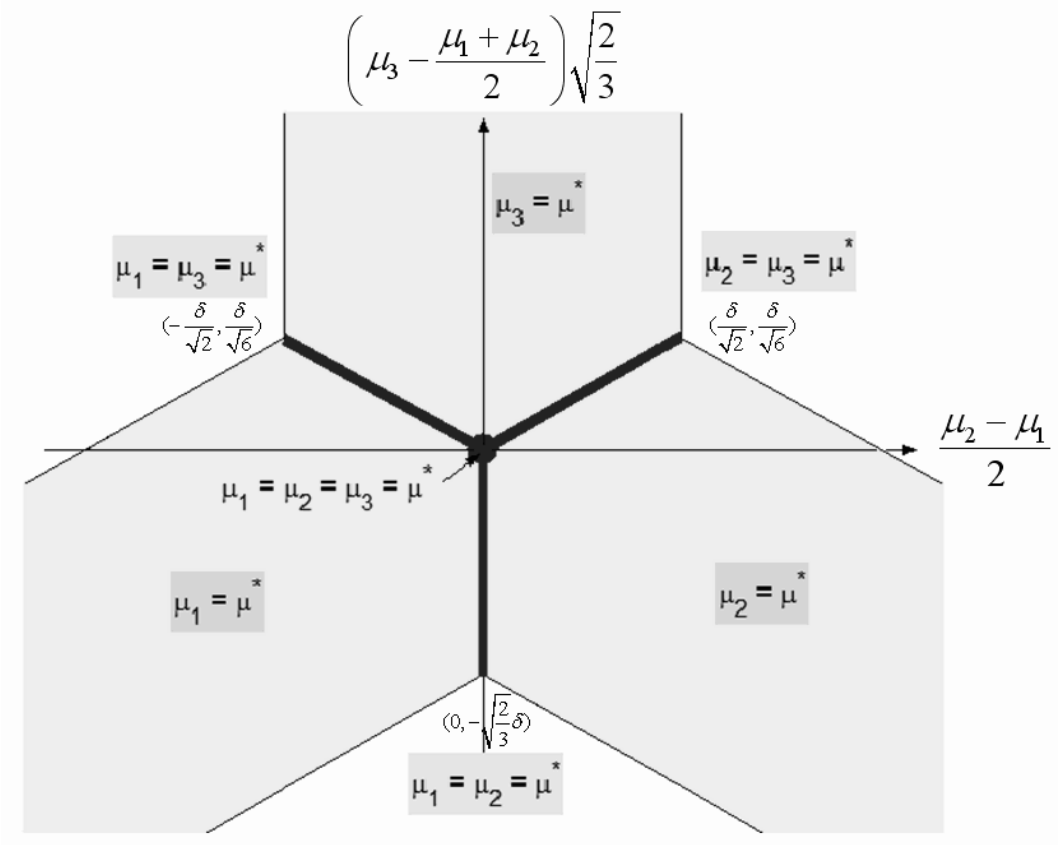

Figure 4.9: The restricted parameter where $\frac{\mu_{(2)}-\mu_{(1)}}{\sigma/\sqrt{n}} \leq \delta$.

CHAPTER V

THE STEP-DOWN PROCEDURE WITH FEEDBACK

In a balance design with three populations, using the studentized range statistics, $q_{3,\,\alpha,\,\nu}$, as the threshold in step 1 of the step-down procedure makes $P(error)$ exactly α if the true parameter has three equal means. It implies that d_3 cannot be smaller than the q statistics; otherwise, the error rate is out of control for the configuration of $\mu_1 = \mu_2 = \mu_3$. As for the rest of the parameter settings, the probability of selecting any treatment with the largest mean into NB relies on the two thresholds in both steps. Such a probability is a function of d_2 and d_3. When using standard $d_2 = t_{\alpha/2,\,\nu}$ and $d_3 = q_{3,\,\alpha,\,\nu}$, however, most configurations have error rates less than α. The numerical calculations shows that the constant d_2 value is not powerful in eliminating less effective treatments when there are at least one inferior treatments. This chapter proposes a new step-down procedure with feedback to adjust the value of d_2 for every individual experiment. Under this methodology, d_2 is no longer a constant but a function depending on range of sample means.

Similar to the previous chapter, the methodology presented here focuses on $k = 3$ case. The proposing $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function along with q statistics control type I error at or below α no matter what the true parameter setting is. The objective of this approach emphasizes monitoring error rates and improving power for the whole

parameter space instead of focusing on a restricted parameter subspace. Prior information about range of treatment means is not needed and nor does it change the value of d_2. Only relative location of sample means is used to modify d_2.

5.1 Motivation

As a motivation for the new procedure, consider the two data configurations shown in Figure 5.1. In either case treatment 1 will be put into NB, and if d_2 is a constant then the same decision will be made for treatment 2 in either case as well. The idea of the new procedure is that different decisions can be made possible for treatment 2 for the two cases by allowing the critical point d_2 to depend upon the value of $\bar{X}_{(3)} - \bar{X}_{(1)}$. Thus, the decision at the second step incorporates feedback from the first step. This chapter studies the formulation and the performance of applying the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function in the step-down procedure.

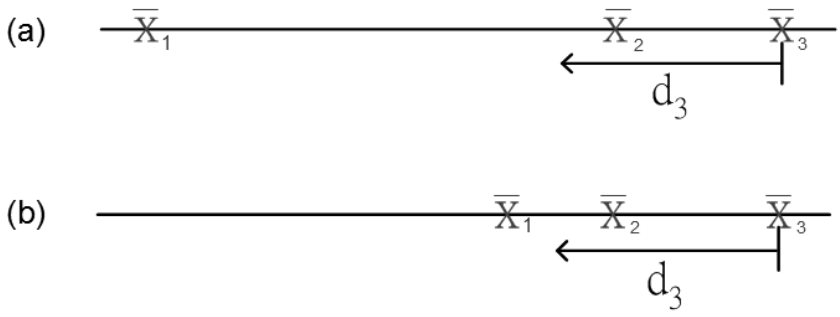

Figure 5.1: Motivation of using $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$.

5.2 Construction of the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function

When testing homogeneity of two population means, $z_{\alpha/2}$ statistics or $t_{\alpha/2,\ \nu}$ statistics are used to control type I error at α level. Same concept applies to the step-down procedure. Setting d_2 to $\sqrt{2}z_{\alpha/2}$ or $\sqrt{2}t_{\alpha/2,\ \nu}$ leads to an asymptotic error rate of α when the configuration is $\mu_{(1)} \ll \mu_{(2)} = \mu_{(3)}$. Figure 3.7 demonstrates the conservative problem of applying constant d_2. $P(error)$ promptly drops to zero as $\mu_{(3)} - \mu_{(2)}$ gets large. For the area with low error rate, a smaller d_2 can also meet the probability constraint of $P(error) \leq \alpha$. Consequently, the upper bound of the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function are $\sqrt{2}z_{\alpha/2}$ and $\sqrt{2}t_{\alpha/2,\ \nu}$ for known and unknown variance respectively.

As the gap among the efficacy levels shrinks, the standard procedure becomes less powerful. It is difficult to discriminate treatments when sample means cluster together. In order to improve efficiency of the test procedure, d_2 is formulated as a function whose value changes with observations. The $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function converges to the standard d_2 value as the range of sample means goes to infinity but takes a smaller value than the standard d_2 value when $\frac{\bar{X}_{(3)} - \bar{X}_{(1)}}{S/\sqrt{n}}$ is not much wider than d_3. Hence, a concave function is proposed:

$$d_2(\bar{X}_{(3)} - \bar{X}_{(1)}) = \sqrt{2} \cdot z_{\alpha/2} \left[1 - e^{a - b\left(\frac{\bar{X}_{(3)} - \bar{X}_{(1)}}{\sigma/\sqrt{n}}\right)}\right], \text{ for known variance}$$

$$d_2(\bar{X}_{(3)} - \bar{X}_{(1)}) = \sqrt{2} \cdot t_{\alpha/2,\nu} \left[1 - e^{a - b\left(\frac{\bar{X}_{(3)} - \bar{X}_{(1)}}{S/\sqrt{n}}\right)}\right], \text{ for unknown variance}$$

5.2.1 New decision space under $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$

The layout of decision space turns into Figure 5.2 when applying the exponential functions proposed above. Since the cutting point in the first step of the test procedure is still q statistics, the size of the hexagon remains the same. However, the decision rule in step 2 is not a fixed number for every data set. Its value depends on the difference between the maximum and the minimum sample means. Accordingly, the borders between concluding one and two inferior treatments change from straight lines to curves.

The step-down procedure with feedback is designed to control familywise error rate for every possible parameter setting. The solutions for a and b in the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function are subject to $P(error) \leq \alpha$. When $\mu_1 = \mu_2 = \mu_3$, the probability of observing a point outside the hexagon of the decision space is exact α by applying $d_3 = q_{3,\,\alpha,\,\nu}$ in a balance design. If $\mu_1 < \mu_2 = \mu_3$, the chance of matching sample means to a point inside region (iii) to (vii) bounded by the dashed curves must less than or equal to α. As it can be seen in Figure 5.2, the total area from (iii) to (vii) bounded by the dashed curves is bigger than that bounded by the solid lines. Hence, $P_{\mu_1<\mu_2=\mu_3}(error)$ increases and so does the power of the test procedure applying the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function. Similarly, the probability of getting a point inside region (iii) to (v) bounded by the dashed curves is required to be at or below α when $\mu_1 \leq \mu_2 < \mu_3$. The total area is greater than that bounded by the solid lines by two shaded areas.

$P(error)$ is a function of d_2 and d_3 if one or more less effective treatments exist.

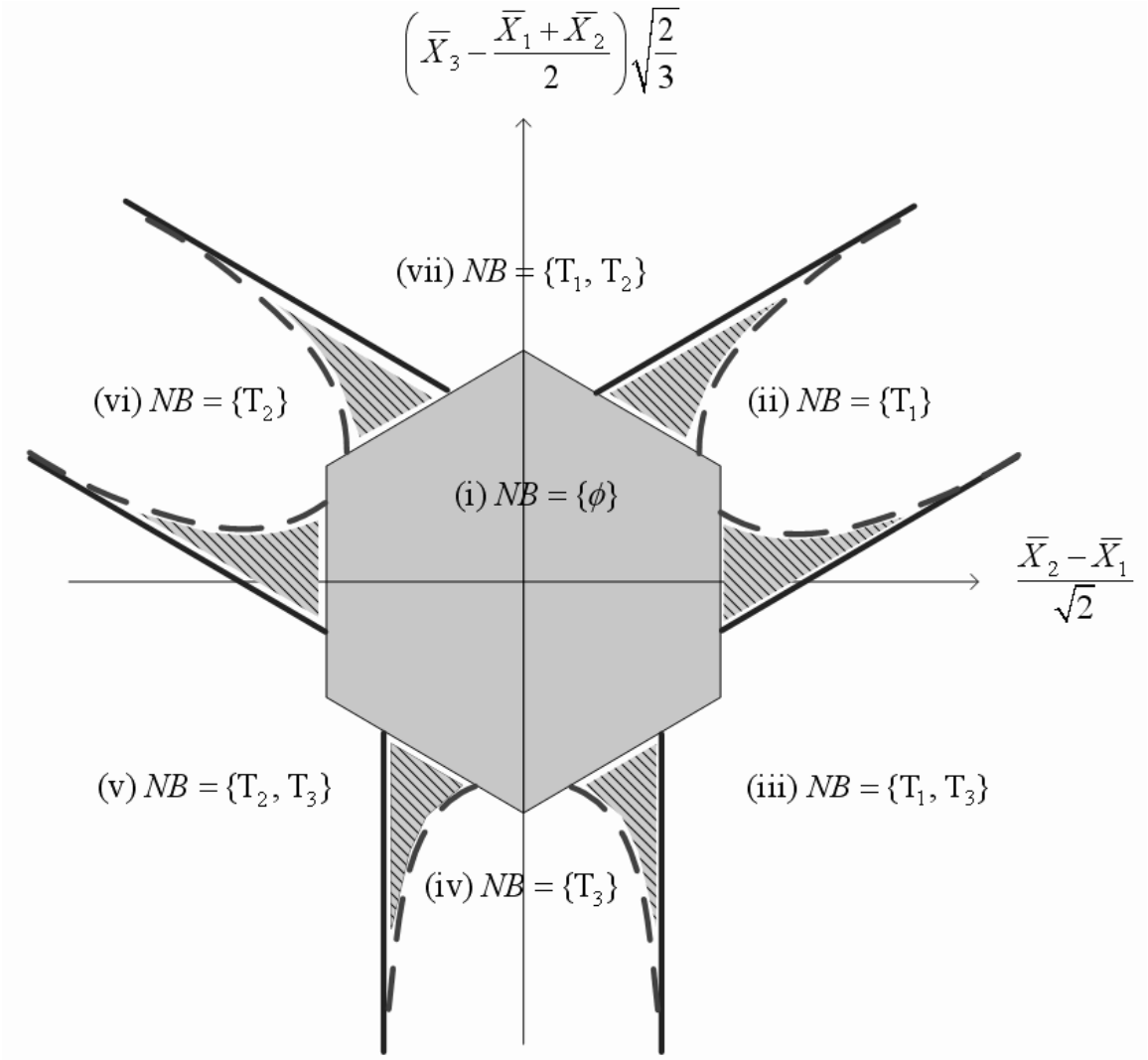

Figure 5.2: The new decision space by using $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$.

When applying the standard procedure with a constant d_2 value of $\sqrt{2}z_{\alpha/2}$, $P(error)$ can be constructed by one integral under known variance case. As for the restricted parameter space approach proposed in chapter 4, it is also sufficient to describe the error rate with one integral when variance is given. With regard to the step-down procedure with feedback, however, the formulation of the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function is more complicated due to the function depends on the relative location of the largest and the smallest sample means. The expression for $P(error)$ requires double integration to characterize $\bar{X}_{(1)}$ and $\bar{X}_{(3)}$ when variance is known and raises the complexity of the problem and calculation.

It is necessary to reconfirm the three properties after constructing new equations for $P(error)$. The second property comments that $P_{\mu_1 \leq \mu_2 = \mu_3}(error)$ is bigger than $\lim_{\delta \to 0} P_{\mu_1 \leq \mu_2 < \mu_3}(error)$ where $\delta = \mu_3 - \mu_2$. The statement is also true when d_2 is a function of $\bar{X}_{(3)} - \bar{X}_{(1)}$. No matter d_2 is a fixed number or a function, the probability of concluding each decision is nonnegative. Based on the definition of error, more types of conclusions are considered to be error decisions when $\mu_1 \leq \mu_2 = \mu_3$ than when $\mu_1 \leq \mu_2 < \mu_3$. The error rate of the configuration with two best treatments is higher than than the limiting error rate of the setting with one best treatment.

The proof of property I becomes complicated due to having two integrals involved. The property can be checked by numerical calculation. When fixing μ_1, μ_2, a and b, the numerical result shows that $P_{\mu_1 \leq \mu_2 < \mu_3}(error)$ decreases as μ_3 increases. The plot of the first derivative of $P_{\mu_1 \leq \mu_2 < \mu_3}(error)$ with respect to μ_3 is negative and converges to zero as μ_3 increases. The phenomenon for property I and the proof of property II together suggest that the configuration with two best treatments is potentially the

most critical scenario in determining d_2.

Therefore, the construction of the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function can be started with searching for the solutions of a and b which satisfy $P_{\mu_1<\mu_2=\mu_3}(error) \leq \alpha$. Then, the $P(error)$ under such solutions needs to be examined for the whole parameter space. The next two sections formulate $P_{\mu_{(1)}<\mu_{(2)}=\mu_{(3)}}(error)$ under the step-down procedure with feedback which adjusts the value of d_2 with observations.

5.2.2 Known variance

Let X_{ij} be the j^{th} independent observation from treatment i, $1 \leq i \leq 3,\ 1 \leq j \leq n$. Suppose that $X_{ij} \sim N(\mu_i,\ \sigma^2),\ \sigma^2$ is given. Take ordered means, $\mu_{(i)} = \mu_i,\ 1 \leq i \leq 3$, to simplify notation. The formulation of type I error for the parameter setting of $\mu_1 < \mu_2 = \mu_3$ is displayed as follows. Solutions for a and b must first satisfy inequality (5.1) for all $\mu_1 < \mu_2 = \mu_3$.

$$
\begin{aligned}
& P_{\mu_1<\mu_2=\mu_3}(error) = P(T_2,\ T_3 \in NB) \\
= \ & P(\bar{X}_1 < \bar{X}_3 - d_3\frac{\sigma}{\sqrt{n}},\ \bar{X}_1 < \bar{X}_2 < \bar{X}_3 - d_2\frac{\sigma}{\sqrt{n}}) \\
& + P(\bar{X}_1 < \bar{X}_2 - d_3\frac{\sigma}{\sqrt{n}},\ \bar{X}_1 < \bar{X}_3 < \bar{X}_2 - d_2\frac{\sigma}{\sqrt{n}}) \\
& + P(\bar{X}_2 < \bar{X}_1 < \bar{X}_3 \text{ and } \bar{X}_2 < \bar{X}_3 - d_3\frac{\sigma}{\sqrt{n}}) \\
& + P(\bar{X}_2 < \bar{X}_3 < \bar{X}_1 \text{ and } \bar{X}_2 < \bar{X}_1 - d_3\frac{\sigma}{\sqrt{n}}) \\
& + P(\bar{X}_3 < \bar{X}_1 < \bar{X}_2 \text{ and } \bar{X}_3 < \bar{X}_2 - d_3\frac{\sigma}{\sqrt{n}}) \\
& + P(\bar{X}_3 < \bar{X}_2 < \bar{X}_1 \text{ and } \bar{X}_3 < \bar{X}_1 - d_3\frac{\sigma}{\sqrt{n}}) \\
= \ & 2P(\bar{X}_1 < \bar{X}_3 - d_3\frac{\sigma}{\sqrt{n}},\ \bar{X}_1 < \bar{X}_2 < \bar{X}_3 - d_2\frac{\sigma}{\sqrt{n}}) \\
& + 2\left[P(\bar{X}_2 < \bar{X}_1,\ \bar{X}_3) - P(\bar{X}_2 < \bar{X}_1 \le \bar{X}_2 + d_3\frac{\sigma}{\sqrt{n}},\ \bar{X}_2 < \bar{X}_3 \le \bar{X}_2 + d_3\frac{\sigma}{\sqrt{n}})\right] \\
= \ & 2\int_{y_1=-\infty}^{\infty}\int_{y_3=y_1+d_3}^{\infty} \phi(y_1-\mu_1^*)\phi(y_3-\mu_3^*) \\
& \qquad \times \left[\Phi(y_3 - \sqrt{2}z_{\alpha/2}(1-e^{a-b(y_3-y_1)} - \mu_2^*) - \Phi(y_1-\mu_2^*)\right]\, dy_3\, dy_1 \\
& + 2\int_{y_2=-\infty}^{\infty} \phi(y_2-\mu_2^*)\left\{[1-\Phi(y_2-\mu_1^*)]\,[1-\Phi(y_2-\mu_3^*)]\right. \\
& \qquad \left. - [\Phi(y_2+d_3-\mu_1^*) - \Phi(y_2-\mu_1^*)]\,[\Phi(y_2+d_3-\mu_3^*) - \Phi(y_2-\mu_3^*)]\right\}\, dy_2 \\
\le \ & \alpha
\end{aligned}
\tag{5.1}
$$

5.2.3 Unknown variance

The formulation of the error rate for unknown variance is an extension of inequality (5.1). Sample standard deviation, S, is used to substitute unknown parameter σ. One more integral is added to the equation to take sample standard deviation into

consideration. Define a random variable U as $\frac{S}{\sigma}$. U follows a $g(u)$ distribution where

$$g(u) = \frac{\nu^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})2^{\frac{\nu}{2}-1}}\ u^{\nu-1}\ exp\ (-\frac{\nu u^2}{2}),\ 0 < u < \infty,\ \nu = 3n-3$$

Following is the formula of $P_{\mu_1<\mu_2=\mu_3}(error)$ based on the step-down procedure with feedback when variance is not given. The goal is to solve a and b such that the inequality below is guaranteed for all of the configurations with two best treatments.

$$
\begin{aligned}
& P_{\mu_1<\mu_2=\mu_3}(error) \\
= \;& P(T_2,\ T_3,\ \text{or both} \in NB) \\
= \;& P(\bar{X}_1 < \bar{X}_3 - d_3\frac{S}{\sqrt{n}},\ \bar{X}_1 < \bar{X}_2 < \bar{X}_3 - d_2\frac{S}{\sqrt{n}}) \\
& + P(\bar{X}_1 < \bar{X}_2 - d_3\frac{S}{\sqrt{n}},\ \bar{X}_1 < \bar{X}_3 < \bar{X}_2 - d_2\frac{S}{\sqrt{n}}) \\
& + P(\bar{X}_2 < \bar{X}_1 < \bar{X}_3 \text{ and } \bar{X}_2 < \bar{X}_3 - d_3\frac{S}{\sqrt{n}}) \\
& + P(\bar{X}_2 < \bar{X}_3 < \bar{X}_1 \text{ and } \bar{X}_2 < \bar{X}_1 - d_3\frac{S}{\sqrt{n}}) \\
& + P(\bar{X}_3 < \bar{X}_1 < \bar{X}_2 \text{ and } \bar{X}_3 < \bar{X}_2 - d_3\frac{S}{\sqrt{n}}) \\
& + P(\bar{X}_3 < \bar{X}_2 < \bar{X}_1 \text{ and } \bar{X}_3 < \bar{X}_1 - d_3\frac{S}{\sqrt{n}}) \\
= \;& 2P(\bar{X}_1 < \bar{X}_3 - d_3\frac{S}{\sqrt{n}},\ \bar{X}_1 < \bar{X}_2 < \bar{X}_3 - d_2\frac{S}{\sqrt{n}}) \\
& + 2\left[P(\bar{X}_2 < \bar{X}_1,\ \bar{X}_3) - P(\bar{X}_2 < \bar{X}_1 \le \bar{X}_2 + d_3\frac{S}{\sqrt{n}},\ \bar{X}_2 < \bar{X}_3 \le \bar{X}_2 + d_3\frac{S}{\sqrt{n}})\right] \\
= \;& 2P\left(\frac{\bar{X}_1}{\sigma/\sqrt{n}} < \frac{\bar{X}_3}{\sigma/\sqrt{n}} - d_3\frac{S}{\sigma},\ \frac{\bar{X}_1}{\sigma/\sqrt{n}} < \frac{\bar{X}_2}{\sigma/\sqrt{n}} < \frac{\bar{X}_3}{\sigma/\sqrt{n}} - d_2\frac{S}{\sigma}\right) \\
& + 2\left[P\left(\frac{\bar{X}_2}{\sigma/\sqrt{n}} < \frac{\bar{X}_1}{\sigma/\sqrt{n}},\ \frac{\bar{X}_3}{\sigma/\sqrt{n}}\right)\right. \\
& \left. - P\left(\frac{\bar{X}_2}{\sigma/\sqrt{n}} < \frac{\bar{X}_1}{\sigma/\sqrt{n}} \le \frac{\bar{X}_2}{\sigma/\sqrt{n}} + d_3\frac{S}{\sigma},\ \frac{\bar{X}_2}{\sigma/\sqrt{n}} < \frac{\bar{X}_3}{\sigma/\sqrt{n}} \le \frac{\bar{X}_2}{\sigma/\sqrt{n}} + d_3\frac{S}{\sigma}\right)\right] \\
= \;& 2\int_{u=0}^{\infty}\int_{y_1=-\infty}^{\infty}\int_{y_3=y_1+ud_3}^{\infty} g(u)\phi(y_1-\mu_1^*)\phi(y_3-\mu_3^*) \\
& \times\left[\Phi(y_3 - u\cdot\sqrt{2}t_{\alpha/2,\nu}(1-e^{a-\frac{b}{u}(y_3-y_1)}) - \mu_2^*) - \Phi(y_1-\mu_2^*)\right] dy_3\, dy_1\, du \\
& + 2\int_{u=0}^{\infty}\int_{y_2=-\infty}^{\infty} g(u)\phi(y_2-\mu_2^*)\left\{[1-\Phi(y_2-\mu_1^*)][1-\Phi(y_2-\mu_3^*)]\right. \\
& \left. - \ [\Phi(y_2+ud_3-\mu_1^*) - \Phi(y_2-\mu_1^*)] \times [\Phi(y_2+ud_3-\mu_3^*) - \Phi(y_2-\mu_3^*)]\right\} dy_2\, du \\
\le \;& \alpha \qquad (5.2)
\end{aligned}
$$

5.2.4 Optimal solution for the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function

Inequality (5.1) and (5.2) have multiple paired solutions for $(a,\ b)$. The goal of the step-down procedure with feedback is to propose a new methodology which increases the power of the test procedure while maintaining familywise error rate at the same time. Therefore, it is preferable to choose the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function which provides the greatest improvement in power among all of the feasible solutions.

The decision spaces under the step-down procedures with constant d_2, $d_2(\delta)$, and $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ are all symmetric. Treatment index does not influence decision. Figure 5.3 shows the difference between the decision spaces under standard d_2 and the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function. Suppose that $\bar{X}_1 \leq \bar{X}_2 \leq \bar{X}_3$, the new $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function creates a larger area for concluding $NB = \{T_1,\ T_2\}$ and then is more capable of eliminating one more treatment when comparing three treatments. If sample means match to a point locating in the shaded area, only the treatment with the smallest sample mean is eliminated under the standard method. On the other hand, the treatments with the smallest two sample means are rejected under the step-down procedure with feedback if observing such a point. The size of the shaded area can then be measurement of power improvement. It represents the performance level of the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function. Thus, the larger the area is, the more powerful the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function is. The size of the shaded area is as follows.

$$\text{shaded area} = \int_{q_{3,\ \alpha,\ \nu}}^{\infty} \sqrt{2} t_{\nu}^{\alpha/2} \cdot e^{a-bx}\, dx$$

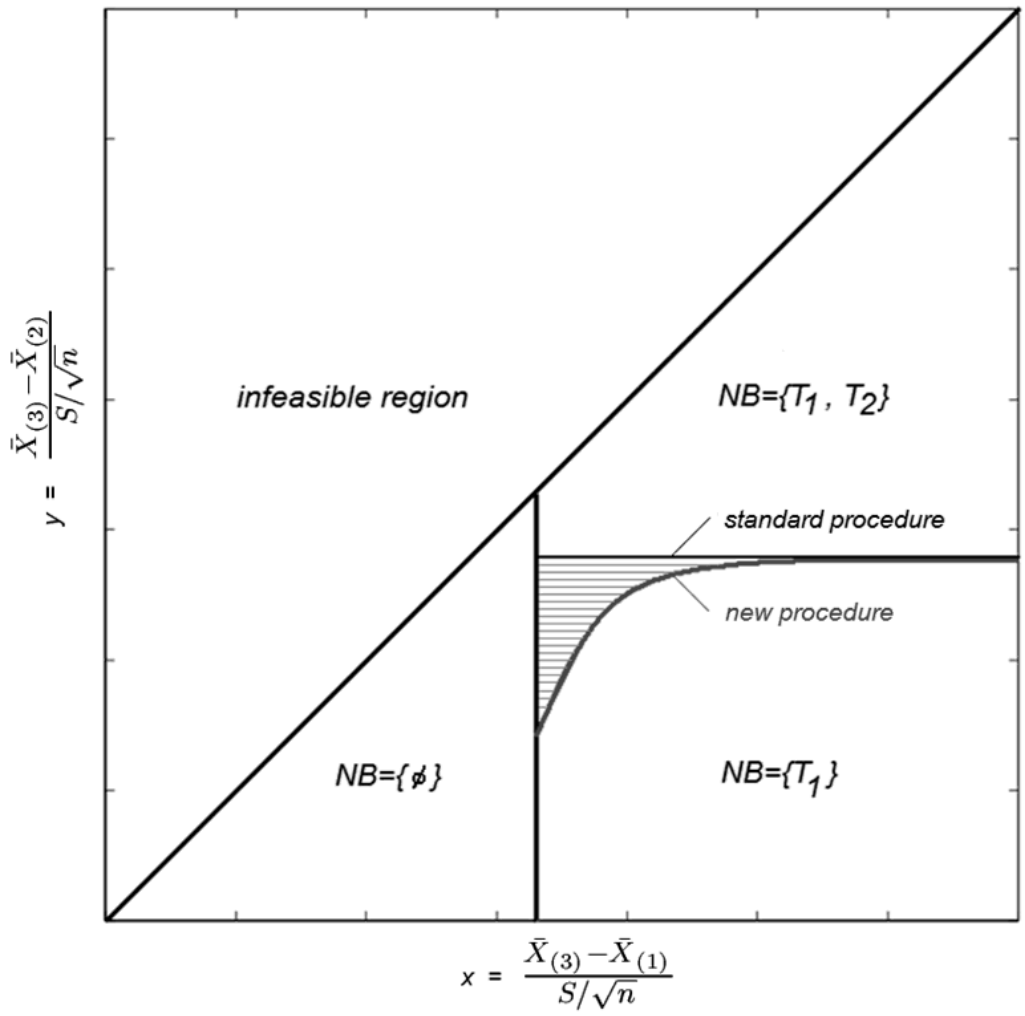

Figure 5.3: The improvement of the step-down procedure with feedback over the standard step-down procedure.

5.3 Computational results and performance

5.3.1 Computational results for the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function

The $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function is $\sqrt{2} \cdot z_{\alpha/2} \left[1 - e^{a-b\left(\frac{\bar{X}_{(3)} - \bar{X}_{(1)}}{\sigma/\sqrt{n}}\right)}\right]$ when variance is known and is $\sqrt{2} \cdot t_{\alpha/2,\ \nu} \left[1 - e^{a-b\left(\frac{\bar{X}_{(3)} - \bar{X}_{(1)}}{S/\sqrt{n}}\right)}\right]$ when variance is unknown. The results of the optimal $(a,\ b)$ at certain α levels and degree of freedoms are tabulated below. The proposing $(a,\ b)$ values create the largest shaded area among all feasible solutions. In addition, using $d_3 = q_{3,\ \alpha,\ \nu}$ and $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ with the optimal $(a,\ b)$ controls type I error rate at or below α for the whole parameter space by numerical calculation.

Table 5.1: Optimal $(a,\ b)$ for the step-down procedure with feedback approach.

α	ν	a	b
1%	∞	16.4	4.8
5%	∞	17.1	5.9
10%	∞	20.5	7.8
5%	30	26.4	8.0

The way to apply the step-down procedure with feedback is to use q statistics for d_3 and calculate the value of d_2 by plugging the optimal $(a,\ b)$ into the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function. For example, $\bar{X}_1 = 0,\ \bar{X}_2 = 0.3,\ \bar{X}_3 = 1.7$, and the ratio of the variance to the sample size in each treatment is known to be 1 to 4. At $\alpha = 5\%$, the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$

function is $\sqrt{2} \cdot z_{0.05/2} \left[1 - e^{17.1-5.9\left(\frac{\bar{X}_{(3)} - \bar{X}_{(1)}}{\sigma/\sqrt{n}} \right)} \right]$. The test procedure is:

[Step 1]

$\frac{\bar{X}_3 - \bar{X}_1}{\sigma/\sqrt{n}} = \frac{1.7-0}{0.5} = 3.4 > d_3 = q_{3,\ 0.05,\ \infty} = 3.314.$

Treatment 1 is selected into NB and continue to Step 2.

[Step 2]

$d_2 = \sqrt{2} \cdot z_{0.05/2} \left(1 - e^{17.1-5.9\times 3.4}\right) = 2.628$

$\frac{\bar{X}_3 - \bar{X}_2}{\sigma/\sqrt{n}} = \frac{1.7-0.3}{0.5} = 2.8 > d_2 = 2.628.$

Treatment 2 is selected into NB and stop.

The conclusion is $NB = \{T_1,\ T_3\}$.

5.3.2 Power of the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function

The definition of the power for a test procedure is the ability of identifying any inferior treatment. The probability of eliminating any inferior treatment and the expected size of the inferior treatments selected into NB are the measurement for power adopted in this study. The test procedure having larger quantity for the measurement is more efficient. Refer to Table 4.4, the calculation of power is determined by the total number of the best treatments. In order to quantify the power of a test procedure, it is necessary to compute the probabilities of getting seven different decisions. Without loss of generality, assume that $n = 1$ and $\sigma^2 = 1$. A more general case with arbitrary

n, σ^2, or even unknown σ^2 can be easily extended from the equation below. Let $d_2(y)$ be the abbreviation for $\sqrt{2} \cdot z_{\alpha/2} \left[1 - e^{(a-by)}\right]$.

$$
\begin{aligned}
P_{\{\phi\}}(\mu) &= P(X_{(3)} - X_{(1)} \leq d_3) \\
&= P(X_1 = min\ \{X_1,\ X_2,\ X_3\},\ X_2,\ X_3 \in (X_1,\ X_1 + d_3)) \\
&\quad + P(X_2 = min\ \{X_1,\ X_2,\ X_3\},\ X_1,\ X_3 \in (X_2,\ X_2 + d_3)) \\
&\quad + P(X_3 = min\ \{X_1,\ X_2,\ X_3\},\ X_1,\ X_2 \in (X_3,\ X_3 + d_3)) \\
&= \int_{x_1=-\infty}^{\infty} \phi(x_1 - \mu_1) \left[\Phi(x_1 + d_3 - \mu_2) - \Phi(x_1 - \mu_2)\right] \\
&\qquad \times \left[\Phi(x_1 + d_3 - \mu_3) - \Phi(x_1 - \mu_3)\right]\, dx_1 \\
&\quad + \int_{x_2=-\infty}^{\infty} \phi(x_2 - \mu_2) \left[\Phi(x_2 + d_3 - \mu_1) - \Phi(x_2 - \mu_1)\right] \\
&\qquad \times \left[\Phi(x_2 + d_3 - \mu_3) - \Phi(x_2 - \mu_3)\right]\, dx_2 \\
&\quad + \int_{x_3=-\infty}^{\infty} \phi(x_3 - \mu_3) \left[\Phi(x_3 + d_3 - \mu_1) - \Phi(x_3 - \mu_1)\right] \\
&\qquad \times \left[\Phi(x_3 + d_3 - \mu_2) - \Phi(x_3 - \mu_2)\right]\, dx_3
\end{aligned} \tag{5.3}
$$

$$
\begin{aligned}
P_{\{1\}}(\mu) &= P(X_1 < X_2 - d_3,\ X_2 - d_2 \leq X_3 < X_2) \\
&\quad + P(X_1 < X_3 - d_3,\ X_3 - d_2 \leq X_2 < X_3) \\
&= \int_{x_2=-\infty}^{\infty} \int_{x_1=-\infty}^{x_2-d_3} \phi(x_1 - \mu_1)\phi(x_2 - \mu_2) \\
&\qquad \times \left[\Phi(x_2 - \mu_3) - \Phi(x_2 - d_2(x_2 - x_1) - \mu_3)\right] dx_1\, dx_2 \\
&\quad + \int_{x_3=-\infty}^{\infty} \int_{x_1=-\infty}^{x_3-d_3} \phi(x_1 - \mu_1)\phi(x_3 - \mu_3) \\
&\qquad \times \left[\Phi(x_3 - \mu_2) - \Phi(x_3 - d_2(x_3 - x_1) - \mu_2)\right] dx_1\, dx_3
\end{aligned} \tag{5.4}
$$

$$
\begin{aligned}
P_{\{2\}}(\mu) &= P(X_2 < X_1 - d_3,\ X_1 - d_2 \le X_3 < X_1) \\
&\quad + P(X_2 < X_3 - d_3,\ X_3 - d_2 \le X_1 < X_3) \\
&= \int_{x_1=-\infty}^{\infty} \int_{x_2=-\infty}^{x_1-d_3} \phi(x_2-\mu_2)\phi(x_1-\mu_1) \\
&\qquad \times [\Phi(x_1-\mu_3) - \Phi(x_1 - d_2(x_1-x_2) - \mu_3)]\, dx_2\, dx_1 \\
&\quad + \int_{x_3=-\infty}^{\infty} \int_{x_2=-\infty}^{x_3-d_3} \phi(x_2-\mu_2)\phi(x_3-\mu_3) \\
&\qquad \times [\Phi(x_3-\mu_1) - \Phi(x_3 - d_2(x_3-x_2) - \mu_1)]\, dx_2\, dx_3 \qquad (5.5)
\end{aligned}
$$

$$
\begin{aligned}
P_{\{3\}}(\mu) &= P(X_3 < X_1 - d_3,\ X_1 - d_2 \le X_2 < X_1) \\
&\quad + P(X_3 < X_2 - d_3,\ X_2 - d_2 \le X_1 < X_2) \\
&= \int_{x_1=-\infty}^{\infty} \int_{x_3=-\infty}^{x_1-d_3} \phi(x_3-\mu_3)\phi(x_1-\mu_1) \\
&\qquad \times [\Phi(x_1-\mu_2) - \Phi(x_1 - d_2(x_1-x_3) - \mu_2)]\, dx_3\, dx_1 \\
&\quad + \int_{x_2=-\infty}^{\infty} \int_{x_3=-\infty}^{x_2-d_3} \phi(x_3-\mu_3)\phi(x_2-\mu_2) \\
&\qquad \times [\Phi(x_2-\mu_1) - \Phi(x_2 - d_2(x_2-x_3) - \mu_1)]\, dx_3\, dx_2 \qquad (5.6)
\end{aligned}
$$

$$
\begin{aligned}
P_{\{1,2\}}(\mu) &= P(X_1,\ X_2 < X_3 - d_3) \\
&\quad + P(X_1 < X_3 - d_3,\ X_3 - d_3 \le X_2 < X_3 - d_2) \\
&\quad + P(X_2 < X_3 - d_3,\ X_3 - d_3 \le X_1 < X_3 - d_2) \\
&= \int_{x_3=-\infty}^{\infty} \phi(x_3 - \mu_3)\Phi(x_3 - d_3 - \mu_1)\Phi(x_3 - d_3 - \mu_2)\, dx_3 \\
&\quad + \int_{x_3=-\infty}^{\infty} \int_{x_1=-\infty}^{x_3 - d_3} \phi(x_1 - \mu_1)\phi(x_3 - \mu_3) \\
&\qquad \times [\Phi(x_3 - d_2(x_3 - x_1) - \mu_2) - \Phi(x_3 - d_3 - \mu_2)]\, dx_1\, dx_3 \\
&\quad + \int_{x_3=-\infty}^{\infty} \int_{x_2=-\infty}^{x_3 - d_3} \phi(x_2 - \mu_2)\phi(x_3 - \mu_3) \\
&\qquad \times [\Phi(x_3 - d_2(x_3 - x_2) - \mu_1) - \Phi(x_3 - d_3 - \mu_1)]\, dx_2\, dx_3 \quad (5.7)
\end{aligned}
$$

$$
\begin{aligned}
P_{\{1,3\}}(\mu) &= P(X_1,\ X_3 < X_2 - d_3) \\
&\quad + P(X_1 < X_2 - d_3,\ X_2 - d_3 \le X_3 < X_2 - d_2) \\
&\quad + P(X_3 < X_2 - d_3,\ X_2 - d_3 \le X_1 < X_2 - d_2) \\
&= \int_{x_2=-\infty}^{\infty} \phi(x_2 - \mu_2)\Phi(x_2 - d_3 - \mu_1)\Phi(x_2 - d_3 - \mu_3)\, dx_2 \\
&\quad + \int_{x_2=-\infty}^{\infty} \int_{x_1=-\infty}^{x_2 - d_3} \phi(x_1 - \mu_1)\phi(x_2 - \mu_2) \\
&\qquad \times [\Phi(x_2 - d_2(x_2 - x_1) - \mu_3) - \Phi(x_2 - d_3 - \mu_3)]\, dx_1\, dx_2 \\
&\quad + \int_{x_2=-\infty}^{\infty} \int_{x_3=-\infty}^{x_2 - d_3} \phi(x_3 - \mu_3)\phi(x_2 - \mu_2) \\
&\qquad \times [\Phi(x_2 - d_2(x_2 - x_3) - \mu_1) - \Phi(x_2 - d_3 - \mu_1)]\, dx_3\, dx_2 \quad (5.8)
\end{aligned}
$$

$$
\begin{aligned}
P_{\{2,3\}}(\mu) &= P(X_2,\ X_3 < X_1 - d_3) \\
&\quad + P(X_2 < X_1 - d_3,\ X_1 - d_3 \le X_3 < X_1 - d_2) \\
&\quad + P(X_3 < X_1 - d_3,\ X_1 - d_3 \le X_2 < X_1 - d_2) \\
&= \int_{x_1=-\infty}^{\infty} \phi(x_1-\mu_1)\Phi(x_1-d_3-\mu_2)\Phi(x_1-d_3-\mu_3)\,dx_1 \\
&\quad + \int_{x_1=-\infty}^{\infty}\int_{x_2=-\infty}^{x_1-d_3} \phi(x_2-\mu_2)\phi(x_1-\mu_1) \\
&\qquad \times[\Phi(x_1-d_2(x_1-x_2)-\mu_3)-\Phi(x_1-d_3-\mu_3)]\,dx_2\,dx_1 \\
&\quad + \int_{x_1=-\infty}^{\infty}\int_{x_3=-\infty}^{x_1-d_3} \phi(x_3-\mu_3)\phi(x_1-\mu_1) \\
&\qquad \times[\Phi(x_1-d_2(x_1-x_3)-\mu_2)-\Phi(x_1-d_3-\mu_2)]\,dx_3\,dx_1 \quad (5.9)
\end{aligned}
$$

The power improvement of known variance cases at $\alpha = 1\%,\ 5\%,\ 10\%$ are illustrated in Figure 5.4 to 5.6. The subplots in the left column illustrate the increase in the probability of eliminating any inferior treatment by applying the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function with feedback. The graphs in the right column reveal the gain in the expected size of the less efficient treatments being selected into NB. In each setting, three types of the parameter configurations are studied individually: (1) $\mu_{(1)} = \mu_{(2)} < \mu_{(3)} : (0,\ 0,\ \delta)$ with one best treatment, (2) $\mu_{(1)} < \mu_{(2)} = \mu_{(3)}) : (0,\ \delta,\ \delta)$ with two best treatments, and (3) $\mu_{(1)} < \mu_{(2)} < \mu_{(3)} : (0,\ \frac{\delta}{2},\ \frac{\delta}{2})$ with one best treatment.

The levels of the improvement are different from setting to setting. In terms of the first measurement of power, $P(i \in NB | i \notin \mathrm{B},\ 1 \le i \le 3)$, the parameter

setting with two best treatment benefits more from the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function than the setting with only one best treatment. For instance, in Figure 5.6 with $\alpha = 10\%$, the improvement for $(0,\ \frac{1}{2},\ \frac{1}{2})$ is 7×10^{-4} while 2×10^{-4} for $(0,\ \frac{1}{2}, 1)$. Based on the properties of $P(error)$, the configuration of $\mu_{(1)} < \mu_{(2)} = \mu_{(3)}$ significantly influences the value for the threshold in step 2 of the test procedure. Therefore, the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function focuses on modifying the probability of the setting with two best treatments. The graphs show that at a fixed degree of freedom, the magnitude of the improvement is larger if a higher type I error rate is allowed.

When comparing to the step-down procedure with constant d_2, the step-down procedure with feedback has minor improvement in power than the restricted parameter space approach if δ is small. It is because that the procedure proposed in this chapter controls $P(error) \leq \alpha$ for the whole parameter, while the procedure studied in chapter 4 maintains the probability restriction for only a narrower parameter subspace. Since there are more constraints, less improvement can be made.

5.4 Summary

The step-down procedure is one approach to differentiate the most effective treatments from the inferior ones. It is preferable to have a small number of potentially the best treatments so that it is easier to target the best one. In other words, the procedure is more efficient if it can detect and eliminate more inferior treatments. However, the standard step-down procedure with constant thresholds is conservative. Although the standard procedure controls familywise error rate for every possible setting, $P(error)$

is far below α for most of the settings. It is difficult for the test procedure with standard d_2 to reject the treatments.

This chapter proposes a new methodology for comparing three treatment means. The data conveys information about true parameter configuration and thus can be used to alleviate the problem of conservativeness. The step-down procedure with feedback uses the same threshold as the standard method in step one. But, the approach utilizes the range of sample means to sharper the threshold in step two. d_2 is no longer a constant but a concave function converging to the standard d_2. The value gets smaller as sample range gets shorter.

The $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function controls $P(error) \leq \alpha$ for the whole parameter space as well as possesses higher power than the standard procedure. Moreover, it is easy to apply the new approach. After solving the optimal solution for $(a,\ b)$ in the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ function, it is simple to determine the value of d_2.

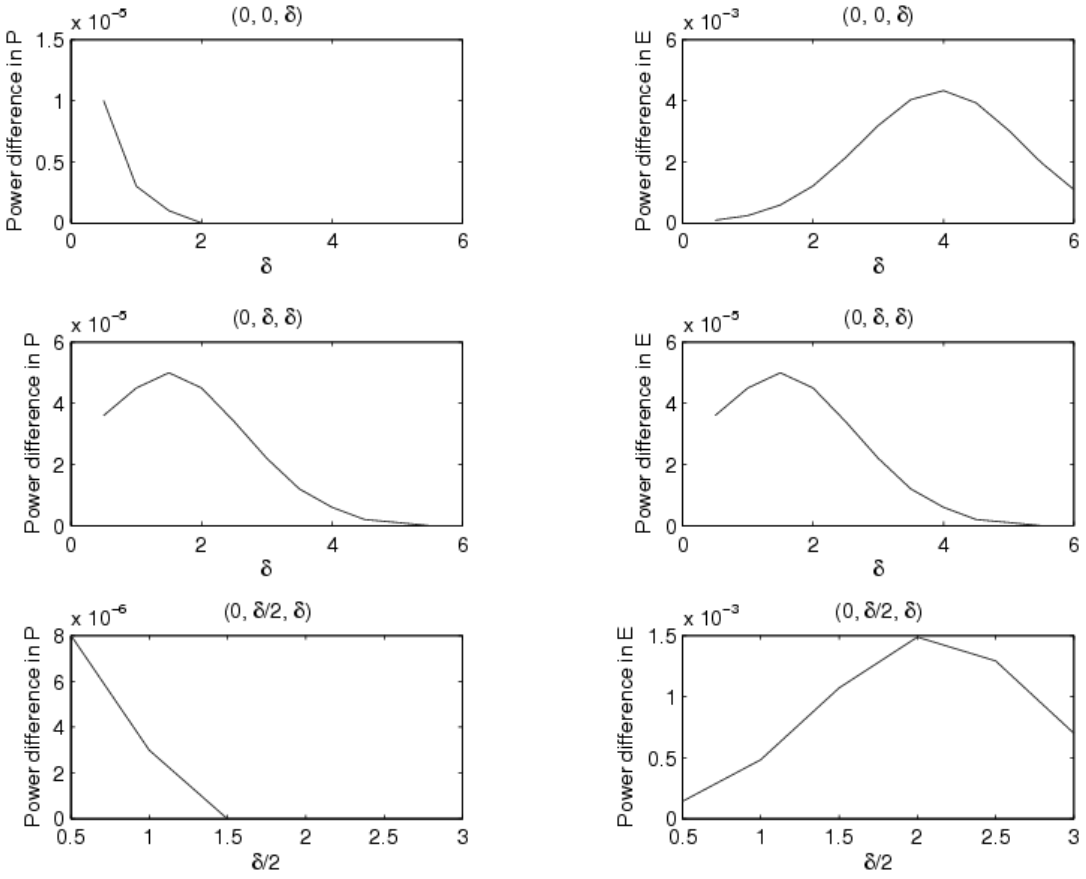

Figure 5.4: Power improvement for known variance case with $d_2(\bar{X}_{(3)} - \bar{X}_{(1)}) = \sqrt{2} \cdot z_{0.01/2} \left[1 - e^{16.4 - 4.8 \frac{\bar{X}_{(3)} - \bar{X}_{(1)}}{\sigma/\sqrt{n}}}\right]$ at $\alpha = 1\%$.

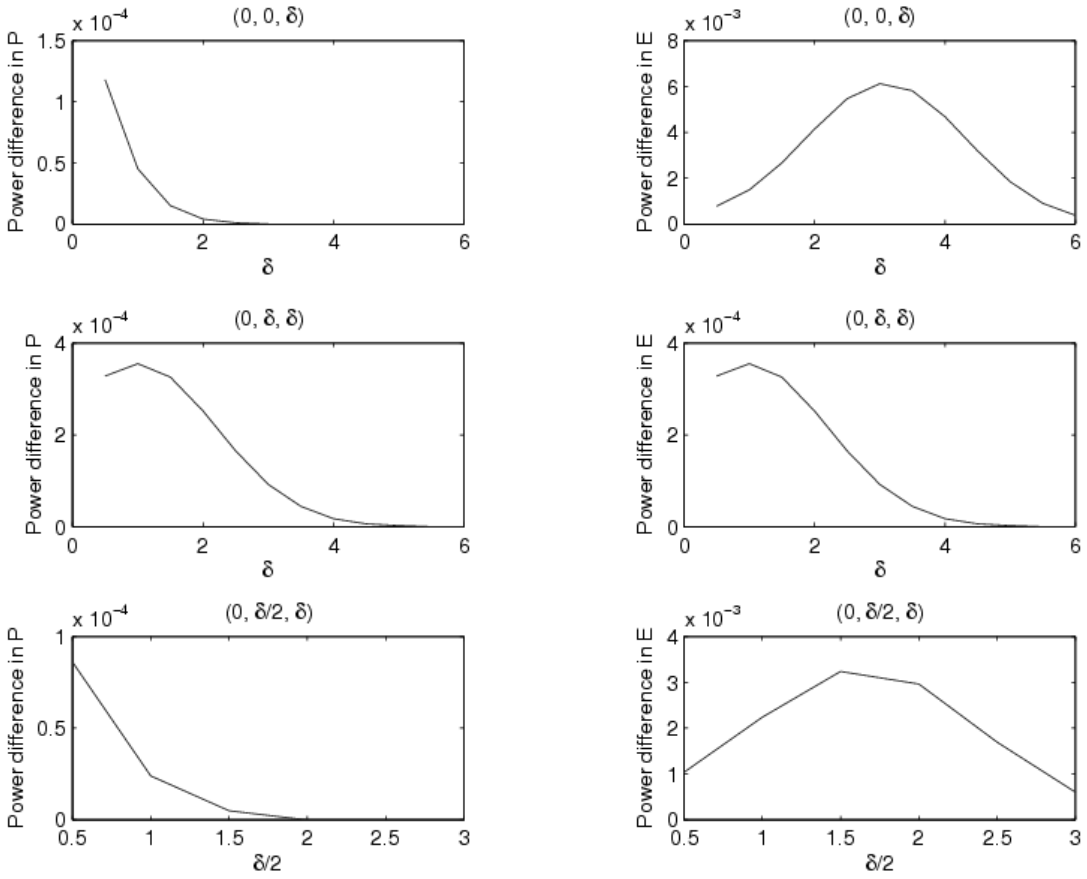

Figure 5.5: Power improvement for known variance case with $d_2(\bar{X}_{(3)} - \bar{X}_{(1)}) = \sqrt{2} \cdot z_{0.05/2} \left[1 - e^{17.1 - 5.9 \frac{\bar{X}_{(3)} - \bar{X}_{(1)}}{\sigma/\sqrt{n}}}\right]$ at $\alpha = 5\%$.

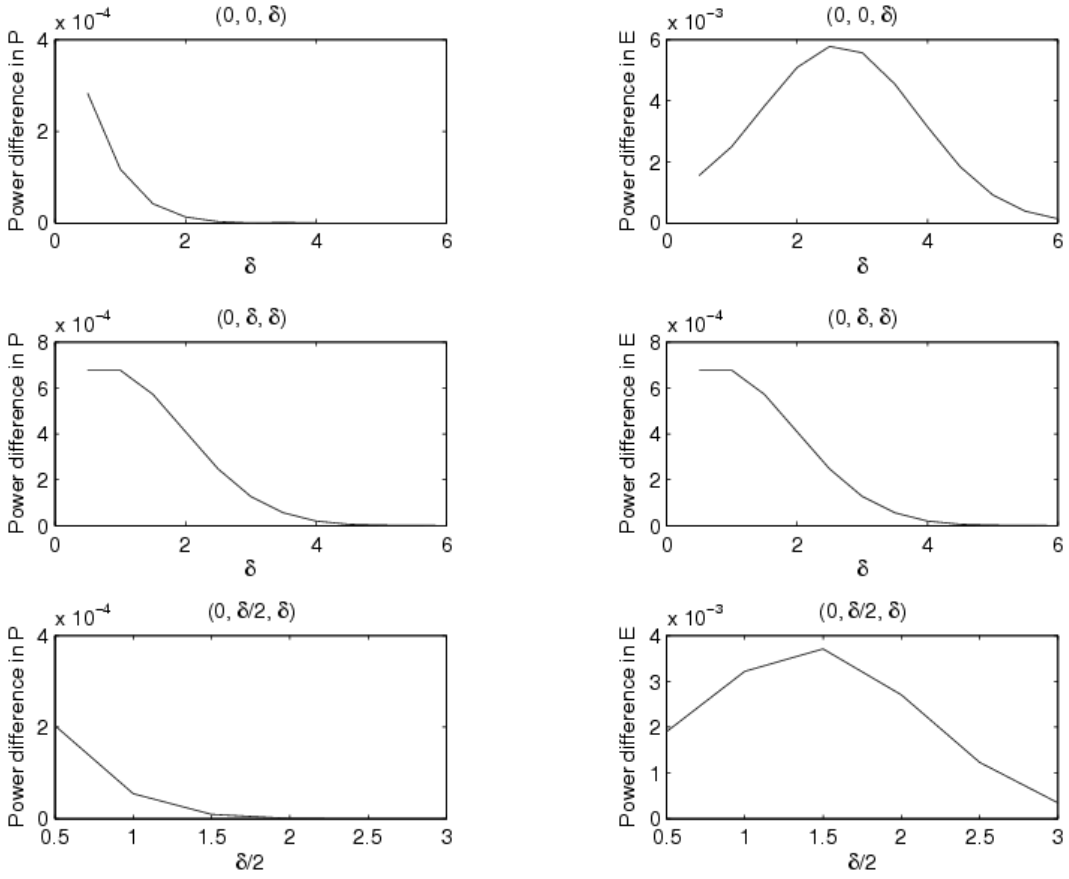

Figure 5.6: Power improvement for known variance case with $d_2(\bar{X}_{(3)} - \bar{X}_{(1)}) = \sqrt{2} \cdot z_{0.1/2} \left[1 - e^{20.5 - 7.8 \frac{\bar{X}_{(3)} - \bar{X}_{(1)}}{\sigma/\sqrt{n}}}\right]$ at $\alpha = 10\%$.

CHAPTER VI

SENSITIVITY ANALYSIS

The restricted parameter space approach and the step-down procedure with feedback are designed for comparing three treatment means in a balanced design. The methodologies are constructed under several assumptions including independency, normality, and equal variances. The procedures assume that independent samples follow a Normal distribution with a mean $\mu_i,\ 1 \leq i \leq 3$ and a common variance σ^2. The following chapter simulates different parameter settings to test the performance of the three step-down procedures when the assumptions do not hold.

The simulation studies the known variance case at $\alpha = 5\%$ for 100,000 times. In each run, ten observations are generated from each group under a certain distribution. Some of the assumptions are violated when generating data. For example, one population follows a t distribution, observation are dependent, and the variances of three groups are not all the same. The ideal variance is set to 10 and $\frac{\sigma}{\sqrt{n}}$ equals to 1. Also, the range of three treatment means is controlled at one where $\delta = \frac{\mu_{(3)} - \mu_{(1)}}{\sigma/\sqrt{n}} = 1$ after standardization. The first methodology uses standard $d_2 = \sqrt{2} z_{\alpha/2} = 2.772$. The second methodology applies the restricted parameter space approach with $d_2 = d_2(\delta = 1) = 1.914$. The last one adopts the step-down procedure with feedback which uses $d_2 = \sqrt{2} z_{\alpha/2} \left[1 - e^{17.1 - 5.9\left(\bar{X}_{(3)} - \bar{X}_{(1)}\right)}\right] \leq 2.772$. The values of d_3 are all $q_{3,\ 0.05,\ \infty} = 3.314$ for the three methods. The difference between sample

means are then compared with the critical points to determine wether rejecting any treatment based on the rules of each test procedure.

6.1 Simulation settings

Since each group is allotted a distribution with a certain mean value, it is well-known that which treatments are the best ones. The best treatment refers to the population with the largest μ among three treatments. Two numbers are recorded during simulation: the frequency of rejecting any best treatment and eliminating any inferior treatment. The ratios of the frequencies to the total number of runs are type I error and power of test procedures. 95% confidence intervals are also provided. Following are the eight settings with different types of relaxation. $(1 \leq j \leq 10)$

- Case I: Dependency within the group

  Suppose that the assumption of independency is violated. The observations from same treatment are dependent. The covariance matrix of ten samples within one treatment is

$$\Sigma = \begin{pmatrix} 10 & & \rho \\ & \ddots & \\ \rho & & 10 \end{pmatrix}_{10\times 10} .$$

  Setting 1:

  $X_{1j} \sim MVN(0,\, \Sigma)$, $X_{2j} \sim MVN(0,\, \Sigma)$, $X_{3j} \sim MVN(1,\, \Sigma)$, X_{1j}, X_{2j}, and X_{3j}

are independent between groups.

Setting 2:

$X_{1j} \sim MVN(0,\ \Sigma)$, $X_{2j} \sim MVN(1,\ \Sigma)$, $X_{3j} \sim MVN(1,\ \Sigma)$, X_{1j}, X_{2j}, and X_{3j} are independent between groups.

- Case II: Unequal variances

  Suppose that one or more populations have variances other than 10. The assumption of equal variance does not hold.

  Setting 3:

  $X_{1j} \sim N(0,\ 10)$, $X_{2j} \sim N(0,\ 10)$, $X_{3j} \sim N(1,\ var)$, X_{ij} are all independent.

  Setting 4:

  $X_{1j} \sim N(0,\ 10)$, $X_{2j} \sim N(1,\ var)$, $X_{3j} \sim N(1,\ var)$, X_{ij} are all independent.

- Case III: Uniform distribution

  The normality assumption is relaxed. Suppose that one population follows an Uniform distribution with a mean equals to one and an unequal variance of $\frac{1}{12}$.

  Setting 5:

  $X_{1j} \sim N(0,\ 10)$, $X_{2j} \sim N(0,\ 10)$, $X_{3j} \sim Unif(0,\ 2)$, X_{ij} are all independent.

  Setting 6:

  $X_{1j} \sim N(0,\ 10)$, $X_{2j} \sim N(1,\ 10)$, $X_{3j} \sim Unif(0,\ 2)$, X_{ij} are all independent.

- Case IV: t distribution

Suppose that one or two populations follow a $\sqrt{10}t_{df}$ distribution with mean equals to zero. For those groups do not meet the normality assumption, the corresponding variances are not 10, either.

Setting 7:

$X_{1j} \sim N(-1,\ 10),\ X_{2j} \sim N(-1,\ 10),\ X_{3j} \sim \sqrt{10} \cdot t_{df}$, X_{ij} are all independent.

Setting 8:

$X_{1j} \sim N(-1,\ 10),\ X_{2j} \sim \sqrt{10} \cdot t_{df},\ X_{3j} \sim \sqrt{10} \cdot t_{df}$, X_{ij} are all independent.

6.2 Simulation results

The simulation results show that the restricted parameter space approach and the step-down procedure with feedback have higher frequencies in both type I error and power than the standard d_2 method. The difference is due to applying a shorter threshold in the second step of the step-down procedure. In general, the restricted parameter space approach has greater power improvement than the step-down procedure with feedback in these eight settings. The reason is that the given range of treatment means is relative tight so that $d_2(\delta)$ is often shorter than $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$. If δ is greater than three, the restricted parameter space approach may not outperforms the step-down procedure with feedback. Although the step-down procedure with feedback has less improvement in the simulation, the methodology can satisfy $P(error) \leq \alpha$ under more settings where the assumptions do not hold. The procedure accommodates to stronger violation of the assumptions.

The constraint of $P(error) \leq \alpha$ holds for the settings with mild relaxation of the assumptions. Especially, the configuration with uniquely the best treatment allows a stronger magnitude of violation than the setting with two best treatments. The phenomenon can be referred to the properties of $P(error)$ that the error rate at $\mu_{(1)} \leq \mu_{(2)} < \mu_{(3)}$ is lower than that at $\mu_{(1)} < \mu_{(2)} = \mu_{(3)}$.

When the best treatment is unique, both new methods work if the correlation between the samples within the group is less than or equal to one as shown in Table 6.1. If the most effective treatment has an unequal variance, follows an uniform distribution or t distribution in case 3, 5, and 7, the new approaches perform properly as well. Familywise error rate are all smaller than 5% in Table 6.5, 6.9, and 6.13. On the other hand, when the size of the best treatments is two, type I error rate is less than 8% if the existing correlation is less than 0.25 and 0.5 for $d_2(\delta)$ and $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ procedures respectively. In case 4 when the variances of the best treatments are no longer ten, the $d_2(\delta)$ procedure can accept a variance less than 12 and a variance smaller than 14 for the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ procedure while having $P(error)$ around 7%. Case 8 relaxes the normality assumption and let the best treatments follow a $\sqrt{10} \cdot t_{df}$ distribution. When df is infinity, $\sqrt{10}t_{df}$ is a $N(0,\ 10)$ distribution. The $d_2(\delta)$ procedure leads to an error rate around 6% when the degree of freedom shrinks to 18. And the $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ procedure even gives an error rate smaller than 6% when the degree of freedom diminishes to 9.

6.3 Summary

In conclusion, the restricted parameter space approach and the step-down procedure with feedback guarantees the error rate when the violation of the assumptions is moderate. Especially, a more serious violation is acceptable when $\mu_{(1)} \leq \mu_{(2)} < \mu_{(3)}$ which contains more parameter settings than $\mu_{(1)} < \mu_{(2)} = \mu_{(3)}$. The new approaches function appropriately for most of the configurations in the parameter space even though the assumption do not exist. Moreover, the two procedures are more effective in identifying and eliminating inferior treatments than the procedure with standard d_2.

Table 6.1: Type I error with a 95% confidence interval for setting 1: $X_{1j} \sim MVN(0,\ \Sigma),\ X_{2j} \sim MVN(0,\ \Sigma),\ X_{3j} \sim MVN(1,\ \Sigma)$.

	standard d_2			$d_2(\delta)$			$d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$		
ρ	L	center	U	L	center	U	L	center	U
0	0.00270	0.00304	0.00338	0.00520	0.00567	0.00614	0.00276	0.00311	0.00346
0.25	0.00654	0.00706	0.00758	0.01132	0.01199	0.01266	0.00669	0.00721	0.00773
0.5	0.01289	0.01361	0.01433	0.02040	0.02129	0.02218	0.01311	0.01383	0.01455
1	0.02913	0.03019	0.03125	0.04200	0.04326	0.04452	0.02958	0.03065	0.03172
2	0.07164	0.07325	0.07486	0.09442	0.09625	0.09808	0.07228	0.07390	0.07552

Table 6.2: Power with a 95% confidence interval for setting 1: $X_{1j} \sim MVN(0,\ \Sigma)$, $X_{2j} \sim MVN(0,\ \Sigma),\ X_{3j} \sim MVN(1,\ \Sigma)$.

	standard d_2			$d_2(\delta)$			$d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$		
ρ	L	center	U	L	center	U	L	center	U
0	0.10002	0.10189	0.10376	0.10076	0.10264	0.10452	0.10007	0.10195	0.10383
0.25	0.14060	0.14277	0.14494	0.14190	0.14408	0.14626	0.14063	0.14280	0.14497
0.5	0.17660	0.17898	0.18136	0.17930	0.18169	0.18408	0.17669	0.17907	0.18145
1	0.24466	0.24733	0.25000	0.24990	0.25259	0.25528	0.24489	0.24757	0.25025
2	0.35051	0.35347	0.35643	0.36147	0.36445	0.36743	0.35082	0.35378	0.35674

Table 6.3: Type I error with a 95% confidence interval for setting 2: $X_{1j} \sim MVN(0,\ \Sigma),\ X_{2j} \sim MVN(1,\ \Sigma),\ X_{3j} \sim MVN(1,\ \Sigma)$.

	standard d_2			$d_2(\delta)$			$d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$		
ρ	L	center	U	L	center	U	L	center	U
0	0.02714	0.02817	0.02920	0.04857	0.04992	0.05127	0.02793	0.02897	0.03001
0.25	0.04914	0.05050	0.05186	0.07649	0.07815	0.07981	0.05013	0.05150	0.05287
0.5	0.07444	0.07608	0.07772	0.10769	0.10963	0.11157	0.07553	0.07718	0.07883
1	0.12564	0.12771	0.12978	0.16637	0.16869	0.17101	0.12670	0.12878	0.13086
2	0.22352	0.22611	0.22870	0.27454	0.27731	0.28008	0.22468	0.22728	0.22988

Table 6.4: Power with a 95% confidence interval for setting 2: $X_{1j} \sim MVN(0,\ \Sigma)$, $X_{2j} \sim MVN(1,\ \Sigma),\ X_{3j} \sim MVN(1,\ \Sigma)$.

	standard d_2			$d_2(\delta)$			$d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$		
ρ	L	center	U	L	center	U	L	center	U
0	0.09062	0.09241	0.09420	0.09526	0.09710	0.09894	0.09089	0.09269	0.09449
0.25	0.12269	0.12474	0.12679	0.13017	0.13227	0.13437	0.12313	0.12518	0.12723
0.5	0.15300	0.15524	0.15748	0.16373	0.16604	0.16835	0.15362	0.15587	0.15812
1	0.20133	0.20383	0.20633	0.21803	0.22060	0.22317	0.20216	0.20466	0.20716
2	0.27657	0.27935	0.28213	0.30059	0.30344	0.30629	0.27753	0.28031	0.28309

Table 6.5: Type I error with a 95% confidence interval for setting 3: $X_{1j} \sim N(0,\ 10),\ X_{2j} \sim N(0,\ 10),\ X_{3j} \sim N(1,\ var)$.

	standard d_2			$d_2(\delta)$			$d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$		
var	L	center	U	L	center	U	L	center	U
8	0.00163	0.00190	0.00217	0.00395	0.00436	0.00477	0.00173	0.00201	0.00229
9	0.00230	0.00262	0.00294	0.00482	0.00527	0.00572	0.00237	0.00269	0.00301
10	0.00270	0.00304	0.00338	0.00520	0.00567	0.00614	0.00276	0.00311	0.00346
11	0.00314	0.00351	0.00388	0.00565	0.00613	0.00661	0.00327	0.00364	0.00401
12	0.00382	0.00422	0.00462	0.00638	0.00689	0.00740	0.00392	0.00433	0.00474
14	0.00567	0.00615	0.00663	0.00809	0.00866	0.00923	0.00575	0.00624	0.00673
16	0.00757	0.00813	0.00869	0.01034	0.01099	0.01164	0.00770	0.00826	0.00882
18	0.00906	0.00967	0.01028	0.01168	0.01237	0.01306	0.00917	0.00978	0.01039
20	0.01176	0.01245	0.01314	0.01441	0.01517	0.01593	0.01186	0.01255	0.01324

Table 6.6: Power with a 95% confidence interval for setting 3: $X_{1j} \sim N(0,\ 10)$, $X_{2j} \sim N(0,\ 10),\ X_{3j} \sim N(1,\ var)$.

	standard d_2			$d_2(\delta)$			$d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$		
var	L	center	U	L	center	U	L	center	U
8	0.08688	0.08864	0.09040	0.08722	0.08898	0.09074	0.08691	0.08867	0.09043
9	0.09456	0.09639	0.09822	0.09497	0.09680	0.09863	0.09460	0.09643	0.09826
10	0.10002	0.10189	0.10376	0.10076	0.10264	0.10452	0.10007	0.10195	0.10383
11	0.10730	0.10923	0.11116	0.10806	0.11000	0.11194	0.10741	0.10934	0.11127
12	0.11142	0.11339	0.11536	0.11246	0.11443	0.11640	0.11151	0.11348	0.11545
14	0.12331	0.12536	0.12741	0.12465	0.12671	0.12877	0.12335	0.12540	0.12745
16	0.13456	0.13669	0.13882	0.13633	0.13847	0.14061	0.13462	0.13675	0.13888
18	0.14392	0.14611	0.14830	0.14640	0.14860	0.15080	0.14403	0.14622	0.14841
20	0.15158	0.15382	0.15606	0.15440	0.15665	0.15890	0.15169	0.15393	0.15617

Table 6.7: Type I error with a 95% confidence interval for setting 4: $X_{1j} \sim N(0,\ 10),\ X_{2j} \sim N(1,\ var),\ X_{3j} \sim N(1,\ var)$.

	standard d_2			$d_2(\delta)$			$d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$		
var	L	center	U	L	center	U	L	center	U
8	0.01393	0.01468	0.01543	0.02989	0.03096	0.03203	0.01438	0.01514	0.01590
9	0.02077	0.02167	0.02257	0.03962	0.04085	0.04208	0.02159	0.02251	0.02343
10	0.02714	0.02817	0.02920	0.04857	0.04992	0.05127	0.02793	0.02897	0.03001
11	0.03626	0.03744	0.03862	0.05926	0.06074	0.06222	0.03706	0.03825	0.03944
12	0.04464	0.04594	0.04724	0.06954	0.07113	0.07272	0.04575	0.04706	0.04837
14	0.06310	0.06462	0.06614	0.09102	0.09282	0.09462	0.06413	0.06567	0.06721
16	0.08361	0.08534	0.08707	0.11595	0.11795	0.11995	0.08487	0.08661	0.08835
18	0.10277	0.10467	0.10657	0.13608	0.13822	0.14036	0.10384	0.10575	0.10766
20	0.11933	0.12135	0.12337	0.15494	0.15720	0.15946	0.12060	0.12263	0.12466

Table 6.8: Power with a 95% confidence interval for setting 4: $X_{1j} \sim N(0,\ 10)$, $X_{2j} \sim N(1,\ var),\ X_{3j} \sim N(1,\ var)$.

	standard d_2			$d_2(\delta)$			$d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$		
var	L	center	U	L	center	U	L	center	U
8	0.07303	0.07466	0.07629	0.07537	0.07702	0.07867	0.07326	0.07489	0.07652
9	0.08339	0.08512	0.08685	0.08678	0.08854	0.09030	0.08371	0.08544	0.08717
10	0.09062	0.09241	0.09420	0.09526	0.09710	0.09894	0.09089	0.09269	0.09449
11	0.09906	0.10093	0.10280	0.10516	0.10708	0.10900	0.09944	0.10131	0.10318
12	0.10786	0.10980	0.11174	0.11526	0.11725	0.11924	0.10842	0.11036	0.11230
14	0.12423	0.12629	0.12835	0.13498	0.13711	0.13924	0.12488	0.12694	0.12900
16	0.14260	0.14478	0.14696	0.15727	0.15954	0.16181	0.14338	0.14557	0.14776
18	0.15824	0.16052	0.16280	0.17548	0.17785	0.18022	0.15925	0.16153	0.16381
20	0.17196	0.17431	0.17666	0.19131	0.19376	0.19621	0.17293	0.17529	0.17765

Table 6.9: Type I error with a 95% confidence interval for setting 5: $X_{1j} \sim N(0,\ 10),\ X_{2j} \sim N(0,\ 10),\ X_{3j} \sim Unif(0,\ 2)$.

standard d_2			$d_2(\delta)$			$d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$		
L	center	U	L	center	U	L	center	U
0.00004	0.00011	0.00018	0.00160	0.00187	0.00214	0.00005	0.00012	0.00019

Table 6.10: Power with a 95% confidence interval for setting 5: $X_{1j} \sim N(0,\ 10),\ X_{2j} \sim N(0,\ 10),\ X_{3j} \sim Unif(0,\ 2)$.

standard d_2			$d_2(\delta)$			$d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$		
L	center	U	L	center	U	L	center	U
0.03525	0.03641	0.03757	0.03525	0.03641	0.03757	0.03525	0.03641	0.03757

Table 6.11: Type I error with a 95% confidence interval for setting 6: $X_{1j} \sim N(0,\ 10),\ X_{2j} \sim N(1,\ 10),\ X_{3j} \sim Unif(0,\ 2)$.

standard d_2			$d_2(\delta)$			$d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$		
L	center	U	L	center	U	L	center	U
0.00363	0.00402	0.00441	0.01543	0.01621	0.01699	0.00374	0.00414	0.00454

Table 6.12: Power with a 95% confidence interval for setting 6: $X_{1j} \sim N(0,\ 10),\ X_{2j} \sim N(1,\ 10),\ X_{3j} \sim Unif(0,\ 2)$.

standard d_2			$d_2(\delta)$			$d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$		
L	center	U	L	center	U	L	center	U
0.05510	0.05653	0.05796	0.05521	0.05664	0.05807	0.05510	0.05653	0.05796

Table 6.13: Type I error with a 95% confidence interval for setting 7: $X_{1j} \sim N(-1,\ 10),\ X_{2j} \sim N(-1,\ 10),\ X_{3j} \sim \sqrt{10} \cdot t_{df}$.

	standard d_2			$d_2(\delta)$			$d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$		
df	L	center	U	L	center	U	L	center	U
6	0.00673	0.00726	0.00779	0.00963	0.01025	0.01087	0.00682	0.00735	0.00788
9	0.00498	0.00544	0.00590	0.00771	0.00827	0.00883	0.00505	0.00551	0.00597
18	0.00333	0.00371	0.00409	0.00606	0.00656	0.00706	0.00344	0.00382	0.00420
27	0.00329	0.00366	0.00403	0.00613	0.00663	0.00713	0.00344	0.00382	0.00420
36	0.00311	0.00347	0.00383	0.00563	0.00611	0.00659	0.00321	0.00358	0.00395
45	0.00298	0.00334	0.00370	0.00535	0.00582	0.00629	0.00306	0.00342	0.00378
90	0.00285	0.00320	0.00355	0.00519	0.00565	0.00611	0.00289	0.00324	0.00359
180	0.00280	0.00315	0.00350	0.00553	0.00601	0.00649	0.00291	0.00326	0.00361
300	0.00262	0.00296	0.00330	0.00521	0.00568	0.00615	0.00270	0.00304	0.00338

Table 6.14: Power with a 95% confidence interval for setting 7: $X_{1j} \sim N(-1,\ 10)$, $X_{2j} \sim N(-1,\ 10),\ X_{3j} \sim \sqrt{10} \cdot t_{df}$.

	standard d_2			$d_2(\delta)$			$d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$		
df	L	center	U	L	center	U	L	center	U
6	0.12795	0.13003	0.13211	0.12965	0.13175	0.13385	0.12803	0.13012	0.13221
9	0.11869	0.12071	0.12273	0.11971	0.12174	0.12377	0.11874	0.12076	0.12278
18	0.10735	0.10928	0.11121	0.10808	0.11002	0.11196	0.10736	0.10929	0.11122
27	0.10470	0.10661	0.10852	0.10530	0.10722	0.10914	0.10474	0.10665	0.10856
36	0.10483	0.10674	0.10865	0.10553	0.10745	0.10937	0.10488	0.10679	0.10870
45	0.10157	0.10346	0.10535	0.10228	0.10417	0.10606	0.10164	0.10353	0.10542
90	0.10152	0.10341	0.10530	0.10225	0.10414	0.10603	0.10158	0.10347	0.10536
180	0.09951	0.10138	0.10325	0.10026	0.10214	0.10402	0.09956	0.10143	0.10330
300	0.10243	0.10432	0.10621	0.10312	0.10502	0.10692	0.10247	0.10436	0.10625

Table 6.15: Type I error with a 95% confidence interval for setting 8: $X_{1j} \sim N(-1,\ 10),\ X_{2j} \sim \sqrt{10} \cdot t_{df},\ X_{3j} \sim \sqrt{10} \cdot t_{df}$.

	standard d_2			$d_2(\delta)$			$d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$		
df	L	center	U	L	center	U	L	center	U
6	0.07123	0.07284	0.07445	0.10078	0.10266	0.10454	0.07226	0.07388	0.07550
9	0.05287	0.05427	0.05567	0.07995	0.08165	0.08335	0.05387	0.05529	0.05671
18	0.03877	0.03998	0.04119	0.06258	0.06410	0.06562	0.03980	0.04103	0.04226
27	0.03487	0.03602	0.03717	0.05787	0.05933	0.06079	0.03576	0.03693	0.03810
36	0.03204	0.03315	0.03426	0.05426	0.05568	0.05710	0.03283	0.03395	0.03507
45	0.03114	0.03223	0.03332	0.05270	0.05410	0.05550	0.03187	0.03298	0.03409
90	0.02934	0.03040	0.03146	0.05004	0.05141	0.05278	0.03014	0.03122	0.03230
180	0.02797	0.02901	0.03005	0.04991	0.05128	0.05265	0.02885	0.02991	0.03097
300	0.02856	0.02961	0.03066	0.04935	0.05071	0.05207	0.02921	0.03027	0.03133

Table 6.16: Power with a 95% confidence interval for setting 8: $X_{1j} \sim N(-1,\ 10)$, $X_{2j} \sim \sqrt{10} \cdot t_{df},\ X_{3j} \sim \sqrt{10} \cdot t_{df}$.

	standard d_2			$d_2(\delta)$			$d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$		
df	L	center	U	L	center	U	L	center	U
6	0.13167	0.13378	0.13589	0.14359	0.14578	0.14797	0.13235	0.13446	0.13657
9	0.11627	0.11827	0.12027	0.12493	0.12699	0.12905	0.11677	0.11878	0.12079
18	0.10041	0.10229	0.10417	0.10660	0.10853	0.11046	0.10110	0.10298	0.10486
27	0.09805	0.09991	0.10177	0.10387	0.10578	0.10769	0.09852	0.10038	0.10224
36	0.09535	0.09719	0.09903	0.10065	0.10253	0.10441	0.09578	0.09762	0.09946
45	0.09463	0.09646	0.09829	0.10003	0.10191	0.10379	0.09513	0.09696	0.09879
90	0.09173	0.09353	0.09533	0.09666	0.09851	0.10036	0.09215	0.09396	0.09577
180	0.09174	0.09354	0.09534	0.09661	0.09846	0.10031	0.09204	0.09385	0.09566
300	0.09213	0.09394	0.09575	0.09685	0.09870	0.10055	0.09247	0.09428	0.09609

CHAPTER VII

CONTRIBUTIONS AND FUTURE RESEARCH DIRECTIONS

7.1 Contributions of the research

This research proposes two types of step-down procedures with inconstant thresholds to solve the problem of discriminating three treatments. The new test procedures employ sharper functions for d_2 to enhance the chance of eliminating the less effective treatments. The approaches can pick out a smaller subset of potentially the best treatments than the procedure with constant thresholds.

The properties of $P(error)$ are studied before introducing the new methodologies. The proofs and the numerical results carried out in chapter 3 suggest that the setting with two best treatments significantly determines the value of d_2. Future study of step-down procedure in the field of subset selection can concentrate on controlling the error rate at the configuration of $\mu_{(1)} \leq \mu_{(2)} = \mu_{(3)}$.

There are several ways to formulate the probability of rejecting any best treatment. The novelty of constructing $P(error)$ in the manner presented in this thesis is that the formulation uses as few integrals as possible. Especially, $P\mu_{(1)} \leq \mu_{(2)} = \mu_{(3)}(error)$ involves single integration in the restricted parameter space approach when the variance is known. The less integration is required, the faster the calculation is. The equations formed in this research are sophisticated and easier to calculated.

7.2 Future research directions

This section outlines a number of potential areas which can be extended from the research started in the thesis.

- Study $k > 3$ case.
- Apply different functions to $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ in the step-down procedure with feedback.
- Define an error decision or the power of a test procedure in a different way.
- Study subset selection procedures for the other types of data.
- Combine two methods.

The formulation of the d_2 functions in the thesis are designed for comparing three treatment means. The concept of constructing shorter thresholds for the step-down procedure, however, can expand to multiple treatments. In a case with a large group number, the formulation of $P(error)$ becomes significantly complicated. Since an error decision involves more scenarios, it requires multiple integration to describe the error rate. The procedures presented in the thesis may not improve the power by much when $k > 3$. Researchers might need to access the problem with high dimension from a different point of view.

The step-down procedure with feedback applies an exponential function to $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$. The features of the function include having smaller values as the range of observation means gets shorter and converging to the constant d_2 as the range goes to

infinity. There are numerous functions possessing the same characteristics. One future direction is to search for other functions of $d_2(\bar{X}_{(3)} - \bar{X}_{(1)})$ which provide greater improvement in power.

Type I error and the power of a test procedure can be defined in various ways based on goals or the fields of interest. For example, experimenters may focus on controlling the probability of accepting any inferior treatment and maximizing the probability of identifying all of the best treatments. Although it is also a subset selection problem, the set up is completely different. The subset selection problems with varied goals can share the same idea of modifying the constant thresholds in the step-down procedure.

This research studies continuous data. The unknown parameter of interest is the mean of treatment. In some applications, the response may be binary or categorical data, i.e. the patients with cancer live or die. It is also an interesting topic to investigate subset selection procedure in different domain.

Last, it is possible to combine the advantages of the restricted parameter space approach with the step-down procedure with feedback. In such way, power of a test can be improved more significantly by applying inconstant threshold under a bounded parameter subspace.

REFERENCES

[1] The Federal Food, Drug, and Cosmetic Act §505(d), 21 U.S.C. §355 (1962).

[2] Alam, K., "A Two-Sample Procedure for Selecting the Population with the Largest Mean from k Normal Populations," *Annals of the Institute of Statistical Mathematics*, vol. 22, pp. 127–136, 1970.

[3] Bechhofer, R. E., "A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances," *The Annals of Mathematical Statistics*, vol. 25, pp. 16–39, 1954.

[4] Bechhofer, R. E., Dunnett, C. W., and Sobel, M., "A Two-Sample Multiple-Decision Procedure for Ranking Means of Normal Populations with a Common Unknown Variance," *Biometrika*, vol. 41, pp. 170–176, 1954.

[5] Bechhofer, R. E. and Goldsman, D. M., "Truncation of the Bechhofer-Kiefer-Sobel Sequential Procedure for Selecting the Normal Population which Has the Largest Mean," *Communications in Statistics–Simulation and Computation*, no. B16, pp. 1067–1092, 1987.

[6] Bechhofer, R. E. and Goldsman, D. M., "Truncation of the Bechhofer-Kiefer-Sobel Sequential Procedure for Selecting the Normal Population which Has the Largest Mean (III): Supplementary Truncation Numbers and Resulting Performance Characteristics," *Communications in Statistics–Simulation and Computation*, vol. B18, pp. 63–81, 1989.

[7] Bechhofer, R., Santner, T., and Goldsman, D., *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons*. New York: Wiley, 1995.

[8] Broström, G., "On Sequentially Rejective Subset Selection Procedures," *Communications in Statistics–Theory and Methods*, no. A10, pp. 203–221, 1981.

[9] Carmer, S. G. and Walker, W. M., "Baby Bear's Dilemma: a Statistical Tale," *Agronomy Journal*, vol. 74, pp. 122–124, 1982.

[10] Cohen, D. S., *A Two-Sample Decision Procedure for Ranking Means of Normal Populations with a Common Known Variance*. Unpublished M.S. Thesis, Department of Operations Research, Cornell University, Ithaca, NY, 1959.

[11] Duncan, D. B., "A Bayesian Approach to Multiple Comparisons," *Technometrics*, vol. 7, pp. 171–222, 1965.

[12] Duncan, D. B. and Dixon, D. O., "k-Ratio t Tests, t Intervals and Point Estimates for Multiple Comparisons," in *Encyclopedia of Statistical Sciences* (Kotz, S. and Johnson, N. L., eds.), vol. 4, pp. 403–410, New York: Wiley, 1983.

[13] Dunnett, C. W. and Tamhane, A. C., "A Step-Up Multiple Test Procedure.," *Journal of the American Statistical Association*, vol. 87, pp. 162–170, 1992.

[14] Finner, H. and Giani, G., "Closed Subset Selection Procedures for Selecting Good Populations," *Journal of statistical planning and inference*, vol. 38, pp. 179–200, 1994.

[15] Gabriel, K. R., "Simultaneous Test Procedures–Some Theory of Multiple Comparisons," *The Annals of Mathematical Statistics*, vol. 40, pp. 224–250, 1969.

[16] Gibbons, J., Olkin, I., and Sobel, M., *Selecting and Ordering Populations*. New York: Wiley, 1977.

[17] Gupta, S., *On a Decision Rule for a Problem in Ranking Means*. Mimeograph series no. 150, Institute of Statistics, University of North Carolina, Chapel Hill, NC., 1956.

[18] Harter, H. L., "History of multiple comparisons," in *Handbook of Statistics* (Krishnaiah, P. R., ed.), vol. 1, pp. 617–622, Amsterdam: North Holland, 1980.

[19] Hartmann, M., "An Improvement of Paulson's Procedure for Selecting the Population with the Largest Mean from k Normal Populations with a Common Unknown Variance," *Sequential analysis*, vol. 10, pp. 1–16, 1991.

[20] Hayter, A., Miwa, T., and Liu, W., "Efficient Directional Inference Methodologies for the Comparisons of Three Ordered Treatment Effects," *Journal of the Japan Statistical Society*, vol. 31, pp. 153–174, 2001.

[21] Hayter, A. J., "A Proof of the Conjecture that the Tukey-Kramer Multiple Comparisons Procedure is Conservative," *The Annals of Statistics*, vol. 12, pp. 61–75, 1984.

[22] Hayter, A. J., "A One-Sided Studentized Range Test for Testing Against a Simple Ordered Alternative," *Journal of the American Statistical Association*, vol. 85, pp. 778–785, 1990.

[23] Hayter, A. J., "A Combination Multiple Comparisons and Subset Selection Procedure to Identify Treatments that are Strictly Inferior to the Best," *Journal of Statistical Planning and Inference*, vol. 137, pp. 2115–2126, 2007.

[24] Hochberg, Y., "A Sharper Bonferroni Procedure for Multiple Tests of Significance," *Biometrika*, vol. 75, pp. 800–802, 1988.

[25] Hochberg, Y. and Tamhane, A., *Multiple Comparison Procedures*. New York: Willy, 1987.

[26] Holm, S., "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, vol. 6, no. 65-70, 1979.

[27] Hsu, J. C., *Multiple Comparisons: Theory and Methods*. Chapman & Hall, 1996.

[28] Laska, E. M. and Meisner, M. J., "Testing Whether an Identified Treatment is Best," *Biometrics*, vol. 45, pp. 1139–1151, 1989.

[29] Marcus, R., Peritz, E., and Gabriel, K. R., "On Closed Testing Procedures with Special Reference to Ordered Analysis of Variance," *Biometrika*, vol. 63, pp. 655–660, 1976.

[30] Miller, R. G., *Simultaneous Statistical Inference*. New York: McGraw-Hill, 1st ed., 1966.

[31] O'Brien, P., "The Appropriateness of Analysis of Variance and Multiple-Comparison Procedures," *Biometrics*, vol. 39, pp. 787–788, 1983.

[32] O'Neill, R. T. and Wetherill, G. B., "The Present State of Multiple Comparison Methods (with discussion)," *Journal of the Royal Statistical Society*, vol. 33, pp. 218–241, 1971.

[33] Perry, J. N., "Multiple-Comparison Procedures: a Dissenting View," *Journal of Economic Entomology*, vol. 79, pp. 1149–1155, 1986.

[34] Petersen, R. G., "Use and Misuse of Multiple Comparisons Procedures," *Agronomy Journal*, vol. 69, pp. 205–208, 1977.

[35] Rothman, K. J., "No Adjustments are Needed for Multiple Comparisons," *Epidemiology*, vol. 1, pp. 43–46, 1990.

[36] Scheffé, H., "A Method for Judging all Contrasts in the Analysis of Variance," *Biometrika*, vol. 40, pp. 87–104, 1953.

[37] Tamhane, A. C. and Bechhofer, R. E., "A Two-stage Minimax Procedure with Screening for Selecting the Largest Normal Mean," *Communications in Statistics–Theory and Methods*, no. A6, pp. 1003–1033, 1977.

[38] Tamhane, A. C. and Bechhofer, R. E., "A Two-stage Minimax Procedure with Screening for Selecting the Largest Normal Mean (II): an Improved PCS Lower Bound and Associated Tables," *Communications in Statistics–Theory and Methods*, no. A8, pp. 337–358, 1978.

[39] Tamhane, A. C., Liu, W., and Dunnett, C., "A Generalized Step-Up-Down Multiple Test Procedure," *The Canadian Journal of Statistics*, vol. 26, pp. 353–363, 1998.

[40] Tukey, J. W., *The Problem of Multiple Comparisons.* Mimeographed notes, Deptartment of Statistics, Princeton University, 1953.

[41] Waller, R. A. and Duncan, D. B., "A Bayes Rule for the Symmetric Multiple Comparisons Problems," *Journal of the American Statistical Association*, vol. 64, pp. 1484–1503, 1969.

[42] Welsch, R. E., "Stepwise Multiple Comparison Procedures," *Journal of the American Statistical Association*, vol. 72, pp. 566–575, 1977.

VITA

Chen-ju Lin was born in Taoyuan, Taiwan. She obtained her B.S. degree in Industrial Engineering and Management with a certification of financial engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2003. She earned a M.S. degree in Industrial and Systems Engineering from Georgia Institute of Technology in 2004. She will receive a Ph.D. degree in Industrial and Systems Engineering with a concentration of Applied Statistics from Georgia Institute of Technology in August, 2007.

During her study in the Ph.D. program, Chen-ju worked as a research assistant for the Center for Assistive Technology & Environmental Access (CATEA), Atlanta, GA. She participated in several projects related to mobility of wheelchairs. She taught ISyE 2028 Basic Statistical Methods for two semesters and was a teaching assistant for ISyE 3770 Statistics and Applications, ISyE 6414 Regression Analysis, and ISyE 6739 Statistical Methods.