NOVEL STEP-DOWN MULTIPLE TESTING PROCEDURES UNDER DEPENDENCE

Shihai Lu

A Dissertation

Submitted to the Graduate College of Bowling Green State University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2014

Committee:

John Tuhao Chen, Advisor

Steve Jex,
Graduate Faculty Representative

Hanfeng Chen

Junfeng Shang
Many scientific experiments subject to rigorous statistical analysis involve the simultaneous evaluation of more than one hypothesis. One typical example is clinical trials comparing different drug regimens in terms of their therapeutic values and side effects. In such simultaneous comparison cases, the classical separate statistical inference does not work for these multiple related inference problems, because it can cause the error rate to be undesirably large due to ignoring the multiplicity of the problems. Multiple testing procedures adjust statistical inferences from an experiment for multiplicity and thus enable better decision making.

Consider the problem of simultaneously testing null hypotheses $H_1, \cdots, H_n$. The usual approach to dealing with the multiplicity problem is to restrict attention to procedures that control the familywise error rate (FWER), the probability of at least one false rejection. The Bonferroni procedure and its stepwise modification, Holm’s procedure are among the most widely used procedures strongly controlling FWER. Lehmann and Romano (2005) generalized the concept of FWER and proposed procedures controlling the $k$-FWER, the probability of committing at $k$ false rejections. Lehmann and Romano’s step-down $k$-FWER controlling procedure (LR procedure) is the generalization of Holm’s procedures. Seneta and Chen (2005) firstly improved Holm’s procedure by incorporating the bivariate distribution information of the test statistics. This dissertation is mainly influenced by their work, in which they derived a step-down procedure (CS procedure) that is more powerful than Holm’s and still strongly controls the FWER under arbitrary dependence.

In Chapter 3, we first propose three novel step-down FWER controlling procedures that improve both the CS and Holm procedures, by utilizing the bivariate distribution information of the test statistics. We also propose a fast algorithm to calculate the critical values of the new procedures. The theoretical proof and simulation study are both presented to
demonstrate that the new procedures can strongly control the FWER and are more powerful than Holm’s and the CS procedures under arbitrary dependence structure.

In addition, in Chapter 4, we also propose one new step-down k-FWER controlling procedure that improves the LR procedure under arbitrary dependence by using bivariate distribution of the test statistics. Moreover, a model-based, rather than probability-inequality-based, procedure is found for controlling k-FWER when all test statistics are mutually independent. Numerical analysis shows that, for $k > 1$, the critical values of k-FWER controlling procedures can exceed beyond the predetermined level $\alpha$, without violating the k-FWER controlling at $\alpha$. This is a main difference between FWER and k-FWER controlling procedures. The simulation study investigates the k-FWER control and the power of the new procedure by comparing it to the LR procedure.

Lastly, in Chapter 5, we make an attempt to answer two questions: First, as the uniformly most powerful (UMP) test can be defined for one individual hypothesis, if the similar concepts for multiple testing problems could be defined? Second, how to find such sets of tests that make the power of a certain multiple testing procedure as large as possible. We propose an answer for the Bonferroni procedure in the case that all test statistics are mutually independent. Specifically, we show that in the non-parametric setting the sign test is a UMP test for testing the population median when the probability distribution of the random sample is continuous. Moreover, it also makes the Bonferroni procedure achieve as much power as possible under the continuity assumption. A simulation study follows to support our results.
ACKNOWLEDGEMENTS

This dissertation could not have been completed without the great support that I have received from so many people over the years.

I would like to express my sincerest and deepest appreciation and gratitude to my advisor, Dr. John T. Chen, for the patient guidance, caring, and mentorship he has provided to me, all the way from when I applied to the Ph.D. program in the Department of Mathematics and Statistics through to the completion of this dissertation. He has helped me through extremely difficult times over the period of my Ph.D. research and for that I sincerely thank him for his encouragement and confidence in me. I am truly fortunate to have had the opportunity to work with him.

I would also like to thank the other members of the dissertation committee: Dr. Hanfeng Chen, Dr. Junfeng Shang, and Dr. Steve Jex for their great support and valuable suggestions. In addition, I would like to thank all of my professors from the Department of Mathematics and Statistics for helping me to develop my background in probability theory and statistics during my years of study at the Bowling Green State University.

Besides, many thanks to the graduate coordinator Dr. Tong Sun and the department staff: Marcia Seubert, Mary Busdeker, and Barbara Berta for their thoughtful help and great support for the past several years. It is you who have made my work, study, and research in the department easier and more pleasant.

Finally, I would like to extend my deepest gratitude to my parents and without whose love, support, and understanding I could never have completed this doctoral degree.
# TABLE OF CONTENTS

## CHAPTER 1 INTRODUCTION

1.1 Error Rates In Multiple Testing

1.1.1 Family of Comparisons

1.1.2 Type I Error Rates and Their Control

1.1.3 Type II Error Rates

1.2 Basic Concepts Related to Multiple Testing

1.2.1 Basic Concepts

1.2.1.1 Coherence and Consonance

1.2.1.2 Free and Restricted Combinations

1.2.1.3 Adjusted P-Values

1.2.2 Types of Multiple Testing Procedures

## CHAPTER 2 LITERATURE REVIEW

2.1 Construction Methods

2.1.1 Union-Intersection Method

2.1.2 Closure Principle

2.2 Procedures Based on Bonferroni’s Inequality

2.2.1 Bonferroni Procedure

2.2.2 Holm’s Procedure

2.2.2.1 Holm’s Step-Down Procedure
2.2.2.2 Strong Control of FWER ...................................................... 24
2.2.2.3 Under Strict Combinations Condition ............................... 26
2.2.3 Seneta and Chen’s Procedure .............................................. 26
2.3 Procedures Based on Simes’ Inequality ................................... 28
  2.3.1 Simes’ Procedure ............................................................. 28
  2.3.2 Hommel’s Procedure ......................................................... 29
    2.3.2.1 Strong Control of FWER ............................................. 30
  2.3.3 Hochberg’s Procedure ..................................................... 32
    2.3.3.1 Strong Control of FWER ............................................. 33
    2.3.3.2 Comparison with Hommel’s Procedure .......................... 33
2.4 Procedures Controlling k-FWER ............................................ 34
  2.4.1 Lehmann and Romano’s Procedure .................................. 34
2.5 Procedures Controlling FDR ............................................... 35
  2.5.1 Benjamini and Hochberg’s (BH) Procedure ......................... 36

CHAPTER 3 NOVEL STEP-DOWN PROCEDURES CONTROLLING FWER 37

3.1 Probability Inequalities ..................................................... 37
  3.1.1 Bonferroni Inequality ..................................................... 38
  3.1.2 Kwerel’s Inequality ....................................................... 39
  3.1.3 Kounias’ Inequality ....................................................... 40
  3.1.4 Hunter’s Inequality ........................................................ 40
3.2 Novel Step-Down Procedures Controlling FWER ....................... 41
  3.2.1 Procedure I ................................................................. 41
  3.2.2 Procedure II ............................................................... 46
  3.2.3 Procedure III ............................................................... 50
3.3 Simulation Study .............................................................. 54
3.4 An Application ................................................................. 56
CHAPTER 4  NOVEL STEP-DOWN PROCEDURES CONTROLLING k-FWER

4.1 Generalized Kwerel’s Inequality ........................................... 62
4.2 Novel Step-Down Procedures Controlling k-FWER .................. 65
   4.2.1 Procedure IV ......................................................... 65
   4.2.2 Binomial Model Under Independence ............................... 78
4.3 Simulation Study ............................................................. 81

CHAPTER 5  POWER OF BONFERRONI PROCEDURE WITH SIGN TESTS

5.1 Sign Test ........................................................................... 86
5.2 The Power of Sign Test .......................................................... 89
5.3 Uniformly Most Powerful Sets of Tests for Bonferroni Procedure 93
5.4 Simulation Study ................................................................. 95

BIBLIOGRAPHY ................................................................. 99

APPENDIX A ........................................................................ 102

1 R Codes for Chapter 3 .......................................................... 102
2 R Codes for Chapter 4 .......................................................... 130
3 R Codes for Chapter 5 .......................................................... 148
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Critical values comparison of FWER controlling procedures in a special case</td>
<td>54</td>
</tr>
<tr>
<td>3.2</td>
<td>FWER comparison for three FWER controlling procedures</td>
<td>60</td>
</tr>
<tr>
<td>3.3</td>
<td>Average power comparison for three FWER controlling procedures</td>
<td>61</td>
</tr>
<tr>
<td>4.1</td>
<td>Critical values of two 2-FWER controlling procedures</td>
<td>77</td>
</tr>
<tr>
<td>4.2</td>
<td>Ratios of the critical values of two 2-FWER controlling procedures</td>
<td>78</td>
</tr>
<tr>
<td>4.3</td>
<td>Critical values of three k-FWER controlling procedures under independence</td>
<td>83</td>
</tr>
<tr>
<td>4.4</td>
<td>2-FWER comparison for two 2-FWER controlling procedures</td>
<td>84</td>
</tr>
<tr>
<td>4.5</td>
<td>Average power comparison for two 2-FWER controlling procedures</td>
<td>85</td>
</tr>
<tr>
<td>5.1</td>
<td>Distributions under null and alternative hypotheses</td>
<td>97</td>
</tr>
</tbody>
</table>
LIST OF TABLES

1.1 Number of errors committed when testing n null hypotheses . . . . . . . . . 4

3.1 Critical values comparison under equal correlations for FWER procedures . . 55

3.2 Critical values comparison for an application: ethanol effect study . . . . . . . 59

5.1 Comparison of conjunctive power of Bonferroni procedure with different tests 98

5.2 Comparison of disjunctive power of Bonferroni procedure with different tests 98
CHAPTER 1

INTRODUCTION

In many areas of modern statistical applications, in particular in biostatistics, biomedical and bioinformatics fields, multiple testing procedures (MTPs) are among the most important statistical tools. In these typical applications, conclusions are often drawn by simultaneously testing of a large number of hypotheses. One typical example is clinical trials comparing different drug regimens in terms of their therapeutic values and side effects. Another example is microarray analysis, where experiments are often conducted to compare the expression levels of large numbers of genes in cancer subjects versus those in normal subjects. In these applications, several, even thousands of, hypotheses need to be tested simultaneously.

One possible approach to make such simultaneous inference across multiple questions is to test each hypothesis separately by an appropriate procedure at the level $\alpha$ for that single inference. This is referred to as a separate inference approach. Any significant separate inference implies the overall significance, that is, the rejection of $H_0$, where $H_0$ denotes the overall null hypothesis that all individual hypotheses are true. If all the $m$ separate null hypotheses are true, that is $H_0$ is true, the probability of claiming the overall significance can be higher than $\alpha$ and will get close to 1 for sufficiently large $m$. Thus with the separate inference approach, false discoveries of both the overall and detailed significance can occur more frequently than is indicated by the separate test level $\alpha$. This is why, for the simultane-
ous testing problems, the multiplicity effect should be taken into account. In another word, the classical separate statistical inference does not work for these multiple related inference problems, because it can cause the error rate to be undesirably large due to ignoring the multiplicity of the problems.

The statistical procedures that are designed to control for the multiplicity effect cross the joint measure of erroneous inferences are called multiple testing procedures. In order to study the new statistical tools, we need to first specify the new ”joint measure of erroneous inferences” in a few of rigorous meanings in the latter sections. Before the further theoretical discussion, let’s see a couple of specific examples that involves multiple testing problems.

The first example is concerned with comparisons between new competing treatments with a control. In pharmaceutical studies, one of the main concerns is to search for new drugs (treatments) that are more effective than a standard drug or placebo (control). Before a pharmaceutical company can market a new drug, the regulatory agency requires that the manufacturer offers convincing evidence that the new drug is better than the control. In the confirmatory phase of the research, typically several drugs are tested against the control simultaneously. If a Type I error is made to any of these comparisons, a drug that is actually inferior to the control may be recommended. The more new drugs the research involves, the more likely the wrong recommendation occurs. Thus in this case, a family-wise control that would be referred as familywise error rate (FWER) latter is required to control the occurrence of the wrong recommendations.

Another example is in the context of DNA microarray data analysis. Assume that microarray experiments are conducted to compare the expression levels of large numbers of genes in prostate cancer subjects versus those in normal control patients. Genetic expression levels for $N = 7012$ genes are obtained for $n = 100$ men with half normal control patients and half prostate cancer patients. The main goal of the study is to discover a small group of ”interesting” genes whose expression levels differ between the prostate and normal subjects. For this end, a multiple testing approach is needed to make the inference on the
7012 tests simultaneously. Then a further investigation might be made on such "interesting" genes to search for a causal link to prostate cancer. Due to the large number of hypotheses in the study, a multiple error rate called false discovery rate (FDR) that is less stringent than FWER is usually used.

1.1 Error Rates In Multiple Testing

1.1.1 Family of Comparisons

In many scientific activities, it is better to conduct one large experiment designed to answer multiple related questions because it is statistically and economically more efficient. In previous section we saw that the separate hypothesis inference approach fails because it can lead to too many false discoveries of significance. As the first step toward finding an effective approach to the multiple test problems, we need to introduce the concept of a family: Any collection of hypotheses for which some joint measure of errors need to be taken into account is called a 'family'. For example, in the first example in the previous section, all new drugs that are to be tested against the control consist of a family in the multiple inference problems. In the second example, the family to be taken into accounted is the collection of all 7012 gene expressions. In these cases, the families contain finite number of members. In some problems, however, the family of potential inferences may be infinite.

1.1.2 Type I Error Rates and Their Control

In this section we introduce several relevant Type I error rates for multiple testing procedures, which extend the familiar Type I error rate used when testing a separate null hypothesis. We also introduce the concepts of weak and strong error rate control.

Suppose that in a multiple test problem there are $n$ hypotheses $H_1, H_2, ..., H_n$ that need to be tested simultaneously. In the drug comparison study, for instance, with $n$ be the number of new drugs under investigation and $H_i$ be the separate null hypothesis that claims
no difference between the $i$th new drug and the control, we want to test the $n$ null hypotheses simultaneously. The number $n$ of null hypotheses in the microarray example is 7012, where a null hypothesis $H_i$, may state that gene $i$ is not differentially expressed under two comparative conditions.

For any test problem, there are two main types of errors that a statistical procedure needs to control. A false positive decision occurs if we declare significance for a true hypothesis. Similarly, a false negative decision occurs if we fail to declare a truly existing significance. These errors are also referred to as Type I and Type II errors, respectively. Sometimes the rejected hypothesis is identified with a 'statistical discovery', thus a Type I error is also referred as a false discovery. A good statistical testing procedure should be able to control the Type I error at a predetermined level $\alpha$, while trying to minimize the Type II error (or maximize the power). When it comes to the multiple testing problems, we need to extend the concepts of the two classical types of errors for separate testing problem to some versions of the joint measure of these errors across an interested family of hypotheses. The related notation is summarized in Table 1.1. The specific $n$ hypotheses are assumed to be known in advance; the number of true hypotheses $n_0$ is fixed but unknown. In Table 1.1, $V$ denotes the number of Type I errors and $R$ the number of rejected hypotheses. Note that $R$ is an observable random variable, $S$, $T$, $U$, and $V$ are all unobservable random variables.

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>Not Rejected</th>
<th>Rejected</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>$U$</td>
<td>$V$</td>
<td>$n_0$</td>
</tr>
<tr>
<td>False</td>
<td>$T$</td>
<td>$S$</td>
<td>$n-n_0$</td>
</tr>
<tr>
<td>Total</td>
<td>$W$</td>
<td>$R$</td>
<td>$n$</td>
</tr>
</tbody>
</table>

The Per-Comparison Error Rate defined as the

$$PCER = \frac{E(V)}{n}$$ (1.1)
is the expected proportion of Type I errors among the $n$ hypotheses. If each of the $n$ hypotheses is tested individually at the significance level $\alpha$, then $PCER \leq \alpha$ (Hochberg and Tamhane, 1987). In many applications, however, controlling the per-comparison error rate at level $\alpha$ is not adequate. Instead, we need consider an error rate that takes into all $n$ hypotheses as a joint family.

The **Familywise Error Rate** defined as the

\[
FWER = P(V \geq 1)
\]

is the probability of making one or more false rejections among the family of hypotheses. The FWER is one of the most commonly used error rates in multiple testing problems, in particular, when the number of comparisons is moderate and/or when strong evidence is needed. Note that the familywise error rate reduces to the common Type I error rate for $n = 1$. When the number of hypotheses becomes very large, say thousands, in many modern practice, however, the multiple testing procedures controlling FWER becomes too stringent such that the power is significantly consumed. The result is that there would be too many false hypotheses that can not be rejected. Also, even though the FWER provides strong evidence for multiple statistical inference, in many situations, often its control is not quite needed. One might be willing to tolerate more than one false rejection provided the number of such cases is controlled, thereby increasing the ability of the procedure to detect false null hypotheses. Therefore, some less stringent controls over false discovery compared to familywise error rate (FWER) control were naturally proposed. Two of them are more widely known: $k$-FWER and FDR. Lehmann and Romano (2005) first proposed the generalized FWER ($k$-FWER) control as well as their single-step and stepwise procedures that control the so called $k$-FWER.

The **$k$-FWER** defined as the

\[
k\text{-FWER} = P(V \geq k)
\]
is the probability of committing k or more false rejections. It is a natural generalization of FWER.

Benjamini and Hochberg (1995) introduced the following seminal Type I error measure which is also less stringent than the FWER control.

The **False Discovery Proportion** defined as the

$$FDP = \begin{cases} \frac{V}{R}, & R > 0 \\ 0, & R = 0 \end{cases} \quad (1.3)$$

is the proportion of erroneously rejected hypotheses among all rejections. Benjamini and Hochberg (1995) propose to control the expectation of the FDP, commonly refereed to as False Discovery Rate (FDR). Then the FDR is defined as

$$FDR = E(FDP) = E\left(\frac{V}{R} \mid R > 0\right) Pr(R > 0) \quad (1.4)$$

Earlier ideas related to the false discovery rate can be found in Seeger (1968) and Soric (1989). The introduction of the false discovery rate has initiated the investigation of alternative error control criteria and many further measures have been proposed recently. In addition, Lehmann and Romano (2005) also proposed their FDP controlling procedures that are similar to their k-FWER controlling procedures but require more strong assumption on the dependence structure of the p-values of the individual tests.

The introduction of FDP-based error measures was motivated by some modern applications, in which the number of hypotheses can be very large, such as the microarray experiments analysis. In these settings, FWER controlling procedures, as a stringent criterion, tend to become too conservative and finally lead to rejection of a very limited number of hypotheses, if any. However, FWER control is more desirable when the number of tests is small, so that a good number of rejections can be made, and all can be trusted to be true findings. On the contrast, the FDR controlling is a less stringent criterion for multiple testing. As a
result, however, the FDR controlling procedures usually give more power than those controlling FWER strongly. The fact can be obtained from the following relations among PCER, FWER, and FDR:

\[
PCER \leq FDR \leq FWER
\]  

(1.5)

The first inequality is immediately obvious since \( \frac{V}{n} \leq \frac{V}{R} \). The second inequality can be seen by

\[
FWER = Pr(V \geq 1) = Pr(V \geq 1, R > 0)
\]
\[
= Pr(V \geq 1 | R > 0)Pr(R > 0)
\]
\[
= \sum_{i=1}^{n} Pr(V = i | R > 0)Pr(R > 0)
\]
\[
\geq \sum_{i=1}^{n} Pr(V = i | R > 0) \left( \frac{i}{R} \right) Pr(R > 0)
\]
\[
= E \left( \frac{V}{R} | R > 0 \right) Pr(R > 0)
\]
\[
= FDR
\]  

(1.6)

From (1.6), one can see that \( FWER = FDR = PCER \) when all \( n \) hypotheses are true and that a procedure controlling FWER will also control the FDR and PCER. Moreover, there will be room to improve the power for FDR controlling procedures when \( FDR < FWER \). In the modern development and applications of multiple testing procedures, the FWER and FDR are two dominant error rates under study.

In order to exactly evaluate the FWER and FDR for a specific multiple test problem, the true configuration of the true hypotheses and false hypotheses is required. In practice, however, the information is unknown. We do not want to control errors under configurations that are not the true ones, possibly losing power. Therefore, the weak and strong control are introduced. The weak control is achieved at a rate if certain Type I error rate is controlled by some procedures when all \( n \) hypotheses are assumed true. In practice, however, it is unlikely that all null hypotheses are true. Thus, a stronger error rate control under less restrictive
assumptions is often necessary. If, for a given multiple testing procedure, the Type I error rate is controlled under any configuration of true and false null hypotheses, the error control is called strong control. For example, the weak control of FWER at level $\alpha$ requires

$$P(V \geq 1 \mid \text{all } H_i \text{ are true }, i \in \{1, 2, \ldots, n\}) \leq \alpha \quad (1.7)$$

On the contrast, the strong control of FWER at level $\alpha$ requires

$$P(V \geq 1 \mid \text{all } H_i \text{ are true }, i \in I, \text{ for any } I \subset \{1, 2, \ldots, n\}) \leq \alpha \quad (1.8)$$

It is obvious that controlling the familywise error rate in the strong sense implies the control of the probability of declaring at least a significance, regardless of which subset of the entire set of null hypotheses are true. Note that if $n_0 = n$, then FDR = FWER. Hence, any false discovery rate controlling multiple comparison procedure also controls the familywise error rate in the weak sense.

### 1.1.3 Type II Error Rates

For any single test, it not only needs to control the Type I error rate, also wants to minimize the Type II error rate (or to maximize the power) at the same time. The Type II error for single tests is defined as the failure to reject a false null hypothesis. A good statistical procedure always aims at a balance between Type I error and Type II error. Consequently, if two procedures control the same Type I error rate, we prefer the one achieving a smaller Type II error rate. When dealing with a multiple testing problem, analogous to extending the Type I error rate, power can be generalized in various ways. Power concepts to measure an experiment’s success in multiple testing are then associated with the probability of rejecting false null hypotheses. Several authors have introduced a plethora of different power concepts (Ramsey 1979; Westfall et al. 1999; Bretz et al., 2011) and each of these concepts is suited for some special settings in practice. In the following we use the notations in table (1.1) to
review several different measures of power that are commonly used in the literature and use the terms for those powers from Bretz et al. (2011).

- The **conjunctive power** defined as the

\[
Pr(S = n - n_0) \text{ or } Pr(T = 0)
\]  

(1.9)

is the probability of rejecting all false null hypotheses. The conjunctive power should be used in studies that aim at detecting all existing false hypotheses. An example could be the case when the proportion of the false hypotheses is large. The conjunctive power appears to be the most attractive, since you obviously would like to reject all false hypotheses. However, it is usually very difficult to have rejections for all false hypotheses, since reasonable designs often have low power by this definition. Computing the conjunctive power requires that all the alternatives must be specified precisely; exact computations are infeasible, but simulation generally yields results that are good enough for design purposes (Westfall et al. 1999).

- The **disjunctive power** defined as the

\[
Pr(S \geq 1)
\]  

(1.10)

is the probability of rejecting at least one false null hypothesis. Disjunctive power is recommended in studies that aim at detecting at least one effect. An special example could be the case when it is known that there is only one existing false hypotheses. As with the conjunctive power, computing the disjunctive power requires that you specify all the alternatives precisely; and while exact computations are infeasible, simulation generally yields results that are accurate enough (Westfall et al. 1999).
• The **individual power** is defined as the

\[ Pr(H_i \text{ is rejected}), i \in M_1, \]  

where \( M_1 \) is the index set of all false hypotheses. It is the rejection probability for a particular false hypothesis \( H_i \). The individual power is most closely related to the ordinary definition of power. The disadvantage of individual power is that, from the multiple testing standpoint, it presumes a particular interest in just one of the multiple hypotheses.

• The **average power** defined as the

\[ \frac{E(S)}{n - n_0}, \text{ when } n_0 < n \]  

is the proportion of false nulls you may expect to detect. For example, in one simulation, there are ten false nulls, and five are detected. Thus the proportion of false nulls detected is \( 4/8 = 50\% \) for this simulation. Another simulation might produce \( 7/10 = 70\% \) as the proportion of false nulls that are detected. Suppose you simulate 10,000 times, and compute the average of all such proportions. This would be an estimate of proportional power. In the precise mathematical terms, the average power is the average of all possible proportions of false nulls detected weighted by their respective probabilities.

• **Number of rejections**: This measure is straightforward to use and often used when comparing two procedures that control the same Type I error rate. The procedure with larger total number of rejections is said to be more powerful than the other.

Note that all first four definitions of power reduce to the usual univariate definition of power when there is only one test. The disjunctive, conjunctive, and average power have also been referred to as minimal, complete, and proportional power, see Westfall et al. (1999).
When the family of hypotheses consists of pairwise mean comparisons, the first 3 definitions of power measures previously mentioned have been introduced as all-pairs power, any-pair power, and per-pair power (Ramsey 1978). Finally, it should be noted that these power definitions are readily extended to any subset $M'_1 \subset M_1$ of false null hypotheses.

There are also other criteria of power proposed in literature, especially those for FDR controlling procedures. We do not review them here. The relevant practical question is to determine appropriate power criteria for a given study. One may argue that conjunctive power should be used in studies that aim at detecting all existing effects. Disjunctive power is recommended in studies that aim at detecting at least one true effect, such as in union intersection settings, see Section 2.2.1. Individual power is appealing in clinical trials with multiple secondary outcome variables (Bretz, Maurer, and Hommel 2010) and average power can be useful for comparing different multiple comparison procedures (Benjamini and Hochberg 1995; Bretz et al. 2011). In general, a suitable power definition even can be given on a case-by-case basis by choosing power measures tailored to the study objectives.

1.2 Basic Concepts Related to Multiple Testing

1.2.1 Basic Concepts

1.2.1.1 Coherence and Consonance

Any multiple testing procedure for a family of hypotheses $H_i, i \in I$ is generally required to possess the following logical consistency property: For any pair of hypotheses $(H_i, H_j)$ such that $H_j$ implies $H_i$, if $H_j$ is not rejected then $H_i$ is also not rejected. This property is called coherence and was introduced by Gabriel (1969). If a hypothesis $H_j$ implies $H_i$, then $H_i$ is said to be a component of $H_j$. That is, coherence means that if $H_j$ is not rejected, then any of its components is retained too. Conversely, the rejection of any component of a hypothesis results in the rejection of the hypothesis itself. An multiple testing procedure that satisfies
this property is called coherent. Coherence is an important requirement for any multiple comparison procedure. If coherence is not satisfied, it leads to the inconsistency problem of rejecting a hypothesis without rejecting all hypotheses implying it (Hochberg and Tamhane, 1987).

As an example, consider the intersection of \( n \) null hypotheses, \( H_0 = \bigcap_{i=1}^{n} H_i \). Since \( H_0 \subset H_i, i \in \{1, 2, ..., n\} \), \( H_0 \) implies \( H_i \) and \( H_i \) are components of \( H_0 \). An coherent procedure for the \( n \) hypotheses will reject \( H_0 \) if there is at least \( H_i \) that is rejected. The closed test procedures described in Section (2.1) are coherent by the general closure construction method.

Consonance is another desirable property of multiple comparison procedures, although it is not as important as coherence. A hypothesis with no components is called minimal; all other hypotheses are called non-minimal. Consonance implies that if a non-minimal hypothesis is rejected, one can reject at least one of its minimal hypothesis (Gabriel 1969). In many applications, the elementary null hypotheses are minimal, such as the pairwise hypotheses \( H_{ij} : \theta_i = \theta_j (1 \leq i < j \leq k) \). Consonance then ensures that if an intersection hypothesis \( H_I, I \subset \{1, 2, ..., n\} \) is rejected, at least one elementary hypothesis \( H_i \) with \( i \in I \) can be rejected as well.

1.2.1.2 Free and Restricted Combinations

A family of hypotheses \( H_i, i \in \{1, 2, ..., n\} \) satisfies the free combinations condition if for any \( I, I \subset \{1, 2, ..., n\} \), the set of parameter points for which all \( H_i, i \in I \) are true and all \( H_i, i \notin I \) are false is non-empty. That means all subsets of the null hypotheses could appear as the set of true hypotheses. Otherwise, the \( n \) hypotheses is said to satisfy the restricted combination condition. The concept was proposed by Holm (1979) to develop his seminal step-down procedure.

Note that the family of many-to-one comparisons satisfies the free combination condition. As an example, consider the comparison of 2 treatments with a control treatment (resulting
in 2 null hypotheses). Any of the three events that "none/one/both" of the treatments is better than the control treatment" is likely to be true in practice. Also note that the family of all pairwise comparisons satisfies the restricted combination condition. Because there are logical implications among the elementary hypotheses. i.e. some combinations of falseness of different elementary hypotheses are not allowed because there are no possible parameter points corresponding to those combinations. For example, consider the pairwise hypotheses of 3 treatment means, $H_{ij} : \theta_i = \theta_j (1 \leq i < j \leq 3)$ (resulting in 3 null hypotheses). In this example, not all configurations of true and false hypotheses are logically possible. For example, $\theta_1 = \theta_2$ and $\theta_1 = \theta_2$ imply $\theta_1 = \theta_3$, thus restricting the configurations in which the first two hypotheses are true and the third false.

Holm (1979) commented that "It is to be observed that a multiple level of significance $\alpha$ for some restricted combinations imposes fewer conditions on the test procedure than a multiple level of significance $\alpha$ for free combinations, i.e. a test procedure with multiple level of significance $\alpha$ for free combinations has a multiple level of significance $\alpha$ for any type of restricted combinations." Thus there will be room to improve the power of multiple testing procedures for families satisfying restricted combination. For example, Shaffer (1986) presented a much more powerful method than standard methods when combinations are restricted.

The free combinations condition is assumed to hold (except when noted otherwise) in all the following sections.

1.2.1.3 Adjusted P-Values

In the multiple testing problems, the observed P-values of all separate tests are not compared to the predetermined significance level $\alpha$, but to some thresholds obtained by $\alpha$ adjustment. Sometimes, it is desirable to also compute adjusted P-values for a given multiple testing procedure, which are directly comparable with the significance level $\alpha$. Let $x$ indicate all the data available for testing the family of hypotheses $H_1, ..., H_n$. In case of familywire error
rate, given a particular multiple testing procedure controlling FWER strongly, the general
definition of the adjusted P-value $q_i(x)$ for $H_i$ is

$$q_i(x) = \inf_{\alpha} \{ H_i \text{ is rejected by the given procedure at level } \alpha \}$$

(1.13)

if such an $\alpha$ exists, and $q_i(x) = 1$ otherwise; see Wright (1992). As a result, the individual
null hypothesis $H_i$ can be rejected only if the associated $q_i(x)$ is less than $\alpha$ while strongly
controlling the FWER at level $\alpha$. Examples of computing adjusted P-values for some specific
multiple testing procedures are given later in some sections of Chapter 2.

1.2.2 Types of Multiple Testing Procedures

The general way of classifying multiple comparison procedures is to divide them into single-
step and stepwise procedures. Single-step procedures are characterized by the fact that
statistical decision of an elementary hypothesis does not take the decision for any other
hypothesis into account. That means, decisions for all elementary hypotheses in a family
are independent and can be made in one step. Thus, the order in which the hypotheses
are tested is not important. Examples of a single-step procedure included the well-known
Bonferroni procedure and general Union-intersection procedures. In contrast, for stepwise
procedures the decision of an elementary hypothesis may depend on the decisions of other
hypotheses. For example, the well known Holm procedure is a stepwise extension of the
Bonferroni procedure using the closure principle under the free combination condition (see
section 2.1).

Stepwise procedures can be divided into two types: step-down and step-up. Consider
the multiple test problem where there are $n$ hypotheses $H_1, H_2, ..., H_n$, and corresponding p-
values $P_1, P_2, ..., P_n$. Let $P$-values $P_{(1)}, P_{(2)}, ..., P_{(n)}$ be the ordered p-values in ascending order,
and $H_{(1)}, H_{(2)}, ..., H_{(n)}$ are the corresponding hypotheses. The step-down procedures start
testing the ordered hypothesis $H_{(1)}$ and step down through the sequence while rejecting the
hypotheses. The procedure stops at the first non-rejection (say $H_{(i)}$), and then $H_{(1)},...,H_{(i-1)}$ are rejected. In other words, given a set of critical values $\alpha_i, 1 \leq i \leq n$ on which one step-down procedure is based, the procedure rejects the hypotheses $H_{(1)},...,H_{(i-1)}$, where $i = \min\{1 \leq j \leq n : P_j > \alpha_j\}$ . If the minimum does not exist, reject all the null hypotheses. An example is the Holm’s procedure (1979).

In contrast, step-up procedures start testing $H_{(n)}$ and step up through the sequence while retaining the hypotheses. The procedure stops at the first rejection (sat $H_{(i)}$), and then $H_{(1)},...,H_{(i)}$ are all rejected. In other words, given a set of critical values $\alpha_i, 1 \leq i \leq n$ on which one step-up procedure is based, the procedure rejects the hypotheses $H_{(1)},...,H_{(i)}$, where $i = \max\{1 \leq j \leq n : P_j \leq \alpha_j\}$ . If the maximum does not exist, reject all the null hypotheses. An example is the Hochberg’s procedure (1988).

In general, single-step procedures are less powerful than the corresponding stepwise extensions in that any hypothesis rejected by the former will also be rejected by the latter, but not vice versa (Bretz et al. 2011). More specific examples of such procedures would be discussed in details in Chapter 2.
CHAPTER 2

LITERATURE REVIEW

In this chapter, we review some of the important literatures that significantly propel the development of multiple testing procedures, especially those about FWER and generalized FWER. The main body of the chapter reviews some of the most important existing procedures controlling the FWER and k-FWER because our research focuses on the investigation of new procedures that control the FWER and k-FWER. In addition, the famous BH procedure controlling FDR will also be reviewed too.

Throughout the chapter, we will first discuss the construction methods by which single-step and stepwise procedures can be constructed. Then we will review some of the influential procedures that control FWER. These procedures are mainly based on two well known equalities: Bonferroni’s Inequality and Sime’s Inequality. As to the k-FWER controlling procedure, the Lehmann and Romano’s procedure is discussed. In the last section, the seminal BH procedure that controls the FDR is discussed because of its important status in approaches to the multiplicity problems, even though it does not control the generalized FWER which the main body of this dissertation is concerned with.
2.1 Construction Methods

We now consider two main methods to construct multiple testing procedures. One is union intersection tests to construct multiple comparison procedures for the intersection of several elementary null hypotheses. The other, the closure principle, is a powerful tool, which extend the union intersection test principle to obtain individual assessments for the elementary null hypotheses.

2.1.1 Union-Intersection Method

Historically, the union intersection test was the first construction method for multiple comparison procedures (Roy 1953). The method arose in order to solve some practical problems. For example, assume several irrigation systems are compared with a control. It is natural to claim success if at least one of the comparative irrigation systems leads to better results than the control. If the null hypotheses claim no effect differences between irrigation systems and the control, we wish to reject at least one false null hypothesis to claim the success of the irritation systems.

To formalize the method, suppose that \( H_0 = \cap_{i \in I} H_i \) where \( I \) is an arbitrary index set. Further suppose that a suitable test of each \( H_i \) is available. Then according to Roy’s union-intersection (UI) method, the rejection region for \( H_0 \) is given by the union of rejection regions for the \( H_i, i \in I \). Thus, \( H_0 \) is rejected if and only if at least one \( H_i \) is rejected. A multiple testing procedure derived in this manner from a UI test is referred to as a UI procedure. A single-step UI procedure is coherent and consonant (Gabriel 1969).

Note that union intersection tests consider the global intersection null hypothesis \( H_0 \) without formally allowing individual inferences of the elementary hypotheses \( H_i, i \in I \). The problem can be solved by many methods. They include single-step UI procedures and confidence-region test method (Hochberg and Tamhane, 1987). Take the single-step UI procedure as an example. To test \( H_i, i \in I \) and \( H_0 = \cap_{i \in I} H_i \) as a joint family, we first
construct the tests for $H_i$. if the rejection region of each $H_i$ for $i \in I$ is of the form $(Z_i > \xi)$, then the test statistic for $H_0$ is $Z_0 = \max_{i \in I} Z_i$ and its rejection region is $(Z_0 > \xi)$. The $Z_0$ obtained in this manner is referred to as UI statistics. To determine the common critical constant $\xi$ so as to control the FWER strongly at level $\alpha$, we choose $\xi$ to be the upper $\alpha$ point of the distribution of $Z_0$ under the the overall hypothesis $H_0$. Determining $\xi$ is often difficult or sometimes even impossible if that joint distribution is not available or numerically intractable. Many popular single-step procedures are in fact UI procedures by construction, such as the Dunnett’s test or the Tukey’s Studentized range test.

In addition, stepwise procedures that uses UI tests can be derived based on the closure principle, which will be discussed in the following section. The closure principle ensures that the procedures will strongly control the FWER.

2.1.2 Closure Principle

The closure principle introduced by Marcus et al. (1976) is a general construction method which leads to procedures that allow one to draw individual conclusions about the elementary hypotheses. The resulting procedures are referred to as closed testing procedures.

Let $X$ be a random variable with distribution $P_\theta(\theta \in \Omega)$. Let $H = \{H_1, H_2, ..., H_n\}$ be a set of null hypotheses, i.e, a set of subsets of $\Omega$, and closed under intersection: $H_i \cap H_j \in H$, for any $H_i, H_j \in H$. For each $H_i, i \in \{1, 2, ..., n\}$, let $\phi_i(X)$ be a level $\alpha$ test. The closed procedure rejects any null hypothesis $H_i$ by means of $\phi_i(X)$ if and only if all hypotheses $H'$ that $H' \subset H_i$ and $H' \in H$ have been tested and rejected.

Marcus et al. also showed that the procedure based on closure principle stated above strongly controls the FWER. To see the fact, let $H_T$ denote the intersection of all true null hypotheses in $H$, A be the event that at least one true hypothesis is rejected, and B be the event that $H_T$ is rejected by its corresponding test $\phi_T(X)$. From the closure principle stated
above, we see that A implies B, i.e. $A \subset B$. And thus under $H_T$,

$$FWER = Pr(A) = Pr(A \cap B) = Pr(B) Pr(A \mid B) \leq \alpha.$$  

The last inequality follows since $Pr(B) = \alpha$. So the familywise error rate is controlled under any $H_T$. The FWER is controlled strongly.

In applications, we can perform the closed test procedures by the following steps:

1. Define a set $H = \{H_1, H_2, \ldots, H_n\}$ of elementary hypotheses that you consider as a joint family for multiple inference.

2. Construct the closure set of all non-empty intersection hypotheses $H_I$ as

$$\tilde{H} = \{H_I = \bigcap_{i \in I} H_i : I \subset \{1, 2, \ldots, n\}, I \neq \emptyset\}.$$  

3. For each intersection hypothesis $H_I \in \tilde{H}$ find a suitable test at level $\alpha$.

4. Reject $H_i$, if all hypotheses $H_I \in \tilde{H}$ with $i \in I$ are rejected, each at level $\alpha$.

Since the union-intersection test is used to test the intersection of a collection of hypotheses. Then it is possible to develop a closed testing procedure that uses UI statistics for testing all intersection hypotheses in a family that is closed under intersection. For example, consider to test two hypotheses $H_1, H_2$ jointly. Assume that there is no implication relationship between them. In order to make individual conclusion on the two elementary hypotheses, one can apply the closure principle to have the following procedure. First, test $H_1 \cap H_2$ with a suitable UI test at level $\alpha$. If it is rejected, continue to test $H_1, H_2$ at level $\alpha$, respectively. Otherwise, we claim that both $H_1$ and $H_2$ are not rejected. Thus we can draw rejection inclusion for $H_1, H_2$ respectively. The closure principle shows that the procedure strongly controls the FWER at level $\alpha$. 
A difficulty to apply the flexible closed testing procedure is that the number of operations for the closure principle is in general of order $2^n - 1$, where $n$ is the number of hypotheses of interest. A way to reduce the computational demand is to find shortcut versions of closed testing procedures that can reduce the number of operations to $n$ in some cases. One of the best known and earliest shortcut procedures is the Holm’s step-down procedure (1979); see section 2.2.2 for details. We will show a shortcut procedure based on UI statistics for a family of hypotheses in which there is no implication relationship between any two elementary hypotheses (which is referred as a non-hierarchical family).

A step-down closed testing procedure based on UI statistics can be constructed because the UI tests have the property that whenever any intersection hypothesis $H_P = \cap_{i \in P} H_i$ is rejected, at least one of the elementary hypothesis $H_i$ implied by $H_P$ is rejected. That is, the rejection of a null hypothesis implies the rejection of any intersection of hypotheses that contains the null hypothesis by UI tests. Therefore, in order to make a rejection decision on any $H_j$, it is not necessary to test all the intersections $H_P$ that contain the $H_j$, but it suffices to test only the $H_j$. If the UI statistics for $H_P$ are of the form $Z_P = \max_{i \in P} Z_i$ (which is the case if the rejection regions for individual $H'_i$'s are of the form $Z_i > \xi$), then test the $H'_i$'s in the order of the magnitudes of the corresponding test statistics $Z'_i$'s from the largest to the smallest. Thus the hypothesis with the largest $Z_i$ is tested first. Note that it is equivalent to testing the global intersection hypothesis according to the UI method. If it is rejected, any intersection hypothesis containing that hypothesis will be automatically ejected by the UI method and therefore that hypothesis is set aside as being rejected. Next the hypothesis with the second largest $Z_i$ is tested. The procedure is continued until some $Z_i$ is retained. By UI method, thereby, all the hypotheses whose test statistics values are not greater than the current $Z_i$ are automatically retained. According the closure principle, this procedure will strongly control the FWER at level $\alpha$ if the individual tests at different steps are of level $\alpha$. For each step, the level $\alpha$ test is obtained by taking as the critical value $\xi$ for that step the upper $\alpha$ point of the distribution of $Z_P = \max_{i \in L} Z_i$, where $L$ is the index set of those
hypotheses whose test statistics values are not greater than the current test statistics.

Many other shortcut (stepwise) versions of the closed testing procedure have been proposed in the literature. We will review some of them in the following sections of this chapter.

2.2 Procedures Based on Bonferroni’s Inequality

The Bonferroni method is among the most popular methods nowadays for multiplicity correction. It is also one of the earliest multiple procedures for multiple testing problems. The name is after the Bonferroni probability inequality on which the method is based. In this section, we review the basic Bonferroni procedure and its step-down extension called Holm’s procedure.

2.2.1 Bonferroni Procedure

The Bonferroni procedure is a single-step procedure that controls the FWER in a strong sense. To perform the procedure, one needs to compare the p-values for all n null hypotheses to the common corrected critical value $\alpha/n$. To be specific, let $H_i, i \in E = \{1, 2, ..., n\}$ be the collection of null hypotheses of interest. Assume that the obtained $P$-values for corresponding test statistics are $P_i, i \in E$. The individual $H_i$ is rejected if $P_i \leq \alpha/n$.

Its strong control of familywise error rate follows from the Bonferroni’s (or Boole’s) inequality. Let any $I \subset E$ be the indexes set of all true hypotheses with the size $|I| = n_0 \leq n$. Specifically, assume that, under the null hypothesis $H_i$,

$$Pr(P_i \leq u) \leq u$$

(2.1)

for any $u \in (0, 1)$. Note that it is not require that the distribution of the $P$-value $P_i$ be uniform on $(0, 1)$ whenever $H_i$ is true. Then under $\cap_{i \in I} H_i$,

$$Pr(V > 1) = Pr(P_i \leq \alpha/n, \text{for some } i \in I)$$
\[
= Pr \left( \bigcup_{i \in I} \{P_i \leq \alpha/n\} \right) \\
\leq \sum_{i \in I} Pr(P_i \leq \alpha/n) \\
\leq n_0 \alpha/n \leq \alpha
\]

where the first inequality holds by the Bonferroni’s inequality.

In addition, the adjusted P-values of the Bonferroni procedure are

\[ q_i = \min(nP_i, 1). \]

We reject \( H_i \) while strongly controlling FWER if \( q_i \), rather than \( P_i \), is less than \( \alpha \) (Wright 1992).

The Bonferroni method is a very general approach, which requires not distributional assumption and is valid for any dependence structure among the test statistics. Also it has the computational ease of the common critical values. However, the Bonferroni test is very conservative in the sense that its ability to detect false null hypotheses will typically be very low since the single \( H_i \) is tested at level \( \alpha/n \), which will become orders smaller than the conventional \( \alpha \) levels, especially if \( n \) is large. Thus its power is low. For this reason procedures are prized for which the levels of the individual tests are increased over \( \alpha/n \) but still strongly control the FWER. There are many such procedures available under different assumptions, such as the following Holm’s step-down procedure, an improvement on Bonferroni procedure.

2.2.2 Holm’s Procedure

Holm (1979) proposed a simple sequentially rejective multiple test procedure, which uniformly improves the Bonferroni procedure. It is a step-down procedure in which all the p-values are compared to a sequence of critical values depending on the data in a step-down
order, while the Bonferroni procedure is a single-step procedure in where all the p-values are compared to a common critical value. Like the Bonferroni procedure, Holm’s procedure is based on the simple Boole inequality and can be applied to any parametric or non-parametric model.

2.2.2.1 Holm’s Step-Down Procedure

Assume that the elementary hypotheses $H_i, i \in E$ are minimal in the sense of Gabriel (1969). Also assume the family of hypotheses satisfy the free combinations condition. Denote by $P(1) \leq P(2) \leq \ldots \leq P(n)$ the ordered obtained p-values from smallest to largest and by $H(1), H(2), \ldots, H(n)$ the corresponding hypotheses. The Holm’s procedure rejects all $H(i), i < j$, where

$$j = \min\{k : P(k) > \alpha/(n - k + 1), k \in E\},$$

if $j$ exists. Otherwise, all hypotheses will be rejected.

Alternatively, the Holm procedure can be described by the following sequentially rejective test procedure. Begin with testing the null hypothesis $H(1)$ by comparing $P(1)$ with $\alpha/n$. If $P(1) > \alpha/n$, the procedure stops and all hypotheses are retained; otherwise, reject $H(1)$ and the procedure continues testing $H(2)$ by comparing $P(2)$ with a larger significance level $\alpha/(n - 1)$. If $P(2) > \alpha/(n - 1)$, stop and all remaining null hypotheses including $H(2)$ are retained; otherwise, reject $H(2)$ and continue testing $H(3)$. The procedure continues until either the first non-rejection occurs or all null hypotheses $H(i), i \in E$ are rejected. From the procedure, we can see that, given a specific set of obtained p-values $P(i)$ for $H(i), i \in E$, those $H(i)$’s rejected by the Bonferroni procedure also can be rejected by Holm’s procedure; conversely, those $H(i)$’s rejected by Holm’s procedure might not be rejected by the Bonferroni Procedure. For example, for the two null hypotheses $H(1)$ and $H(2)$ to be tested, the two ordered p-values are $P(1) = \alpha/3$ and $P(2) = 3\alpha/4$. Since $\alpha/2 < 3\alpha/4 < \alpha$, $H(2)$ is rejected by Holm’s procedure but retained by the Bonferroni’s. This means that the Holm’s stepwise procedure is uniformly more powerful than the Bonferroni procedure.
In addition, the adjusted P-values of Holm's procedure are

\[ q(i) = \min \{ \max \{ (n - i + 1)P(i), q(i-1) \}, 1 \}, i > 1, \]

where \( q(1) = \min(nP(1), 1) \). We reject \( H(i) \) while strongly controlling FWER if \( q(i) \) is less than \( \alpha \) (Wright 1992).

### 2.2.2.2 Strong Control of FWER

Holm’s procedure can be viewed as a closed test procedure that uses Bonferroni single-step method to test every intersection \( \cap_{i \in I}, I \subset E \) at the pre-specified level \( \alpha \). Starting by comparing \( P(1) \) with \( \alpha/n \), if not greater, all intersections containing \( H(1) \) are rejected by the Bonferroni procedure at level \( \alpha \). The closure principle implies \( H(1) \) should be set aside as rejected, and then continue comparing \( P(2) \) with \( \alpha/(n-1) \). If \( P(1) > \alpha/n \), all \( P(i) > \alpha/n, i \in E \) and thus all \( H(i), i \in E \) are retained. Those steps are repeated to test the remaining \( H(i) \) in a stepwise manner if the current hypotheses is rejected, until a rejection fails to be made. This procedure is exactly the Holm’s procedure and controls the FWER at level \( \alpha \) in a strong sense by the closure principle.

In the Holm’s original paper, he proved the strong control of FWER using the Bonferroni (or Boole) inequality. Mathematically, the conservativeness of the Bonferroni method is due to the loose inequalities in (2.2). The first inequality is the Bonferroni inequality, which can be tightened by other degree-2 inequalities. The details about this chance to improve the Bonferroni method will be discussed in Chapters 3 and 4. The strictness of the third inequality in (2.2) is due to the fact that we generally cannot know the true number \( n_0 \) of the true hypotheses. Otherwise, we could set the critical value as \( \alpha/n_0 \) instead of \( \alpha/n \) in the Bonferroni procedure in order to have a more powerful procedure. The Holm procedure actually improves the third inequality by the information about \( n_0 \) revealed in a stepwise manner. For example, if \( P(1) \leq \alpha/n \) and \( P(i) > \alpha/n, i \in E/\{1\} \), then by the Bonferroni
procedure, only $H_{(1)}$ is rejected a level $\alpha$. This implies that $n_0$ is at most $n - 1$ since all true hypotheses can only be among $H_{(i)}, i \in E/\{1\}$. Therefore, we can make further inference on $H_{(i)}, i \in E/\{1\}$ with the learned information about $n_0$ from the previous step, even though they have been retained by a conservative Bonferroni test. To further test the retained $n - 1$ hypotheses, we only need compare their p-values with $\alpha/(n - 1)$ at this step. Every time one hypothesis is rejected, we can claim that the most number of the true hypotheses should decrease by 1.

Mathematically, let any $I \subset E$ be the indexes set of all true hypotheses with the size $|I| = n_0 \leq n$. Let $j$ be the smallest (random) index satisfying

$$P_{(j)} = \min_{i \in I} P_i.$$  \hspace{1cm} (2.3)

Note that $j \leq n - n_0 + 1$ (or $n_0 \leq n - j + 1$). Consider the probability of all true hypotheses being retained:

$$Pr(\text{all } H_i, i \in I \text{ are retained by Holm’s procedure} \mid \cap_{i \in I} H_i \text{ is true})$$

$$\geq Pr(P_{(j)} \geq \alpha/(n - j + 1) \mid \cap_{i \in I} H_i \text{ is true})$$

$$\geq Pr(P_{(j)} \geq \alpha/n_0 \mid \cap_{i \in I} H_i \text{ is true})$$

$$\geq Pr(P_i \geq \alpha/n_0, \text{all } i \in I \mid \cap_{i \in I} H_i \text{ is true})$$

$$\geq 1 - Pr(P_i < \alpha/n_0, \text{some } i \in I \mid \cap_{i \in I} H_i \text{ is true})$$

$$\geq 1 - \sum_{i \in I} Pr(P_i < \alpha/n_0 \mid \cap_{i \in I} H_i \text{ is true})$$

$$\geq 1 - n_0 \alpha/n_0$$

$$= 1 - \alpha$$

Thus, the $FWER \leq 1 - (1 - \alpha) = \alpha$ under any $I \subset E$. 

2.2.2.3 Under Strict Combinations Condition

The free combinations condition is required when Holm proposed his procedure. Often, among the hypotheses being tested there are logical implications so that not all combinations of true and false hypotheses are possible. If the family of hypothesis satisfies the strict combinations condition, Holm’s procedure still works, but in a conservative way. As an example of such a situation, given samples from three distributions, we want to make pairwise comparisons of their means $\theta_i, i = 1, 2, 3$. So we will have three elementary hypotheses: $\theta_i = \theta_j, i < j = 1, 2, 3$. The local implications among the hypotheses show the fact that one of $H_i, i = 1, 2, 3$ being false implies at least one other must be false. The Holm procedure compares their p-values with the critical values $\alpha/3, \alpha/2, \alpha$. If $H_{(1)}$ is rejected, then at most one of $H_{(2)}$ and $H_{(3)}$ is true. In order to control the FWER at level $\alpha$, the critical value $\alpha$, instead of $\alpha/2$, suffices for testing $H_{(2)}$ (no correction needed). A general stepwise procedure for the families with logical restrictions was proposed by Shaffer (1986). She extended Holm’s procedure, utilizing logical restrictions to improve the power of closed tests in the closed testing procedure, and thus improve the power of Holm’s procedure when the free combinations condition is not assumed to hold. Shaffer’s procedure at stage $j$, instead of rejecting $H_{(j)}$ if $P_{(j)} \leq \alpha/(n - j + 1)$, rejects $H_{(j)}$ if $P_{(j)} \leq \alpha/t_j$, where $t_j$ is the maximum number of possibly true hypotheses, given that at least $H_{(i)}, i < j$ are false. It is clear that when the free combinations condition is satisfied, Shaffer’s procedure reduces to the ordinary Holm procedure.

2.2.3 Seneta and Chen’s Procedure

Holm’s (1979) step-down and Hochberg’s (1988) step-up procedures for tests of multiple hypotheses are simple to apply and are widely used. Holm’s procedure controls the familywise error rate (FWER), while Hochberg’s is more powerful. Seneta and Chen (2005) investigated a step-down procedure (the CS procedure) which is a sharpening of Holm’s, takes into account the degree of association between test statistics, and also strongly controls the
FWER. Computation for the CS procedure may be minimized by using the procedure as an adjustment to Holm’s. Simulation investigations in a standard comparison with a control setting show that the CS step-down procedure is more powerful than Hochberg’s step-up procedure and the procedure of Simes (1986), especially in regard to error rate, and not much less powerful than an optimal, but very specific, step-up procedure of Dunnett and Tamhane (1992).

Consider the multiple testing problem where there are \( n \) hypotheses \( H_1, ..., H_n \) and corresponding p-values \( R_1, ..., R_n \). Also assume that the test statistics \( X_1, ..., X_n \) are from the same continuous distribution \( f(x) \). Let \( R_{(1)}, ..., R_{(n)} \) be the ordered p-values and \( H_{(1)}, ..., H_{(n)} \) the corresponding hypotheses. Writes \( R_{(i)} = R_{t_i} \), where \( t_i \) is a random variable taking values from \{1, 2, ..., n\}. Define the index set \( L(p) = \{t_p, t_{p+1}, ..., t_n\} \), which is a random set. Furthermore, assume that the bivariate joint distribution of any two test statistics are known.

Seneta & Chen (2005) developed their step-down procedure using the sequence of critical values \( \alpha_p, p = 1, ..., n \) defined as

\[
\alpha_p = \min\left(\frac{\alpha}{n-p}, \frac{\alpha + \gamma_p}{n-p+1}\right), p = 1, 2, ..., n,
\]

where the degree-2 terms \( \gamma_p \) are defined based on bivariate joint probabilities and Kounias’ inequality (see section (3.1.4), i.e.,

\[
\gamma_p = \max_{j \in L(p)} \sum_{i \in L(p) \setminus \{j\}} P(R_i \leq \frac{\alpha}{n-p+1}, R_j \leq \frac{\alpha}{n-p+1} | H_s, s \in L(p), true).
\]

The procedure begins by testing if \( R_{(1)} < \alpha_1 \). If so, reject \( H_{(1)} \) and go on to test if \( R_{(2)} < \alpha_2 \). If not, accept all hypotheses. In general, if all \( H_{(1)}, H_{(2)}, ..., H_{(i)}, 1 \leq i \leq j-1 \) are accepted, then at step \( j \) the remaining hypotheses are \( H_{(j)}, ..., H_{(n)} \) and the inequality next to be checked is \( R_{(j)} < \alpha_j \). If it holds, reject \( H_{(j)} \) and continue; otherwise accept \( H_{(j)}, ..., H_{(n)} \). The process may run at most until a decision is made on the basis of whether \( R_{(n)} < \alpha \) or
not. This procedure strongly controls FWER at level $\alpha$.

In Holm’s original procedure, the critical points are simply $\alpha/(n - p + 1), p = 1, 2, ..., n$, whereas Seneta & Chen increase these critical points by

$$\min(\alpha/(n - p)(n - p + 1), \frac{\gamma_n}{n - p + 1})$$

for each $p = 1, 2, ..., n$ for their (CS) procedure that is thus obviously more powerful than Holm’s.

### 2.3 Procedures Based on Simes’ Inequality

In this section, we review another branch of FWER controlling procedures that is based on a modified Bonferroni’ inequality proposed by Simes (1986). In general, these procedures are uniformly more powerful than those based on the Bonferroni’s inequality. But they are only valid under special conditions about the dependence structure among test statistics, while the dependence structure has never been an issue for those FWER controlling procedures based on the Bonferroni’s inequality.

#### 2.3.1 Simes’ Procedure

The conservativeness of the Bonferroni procedure is due to that the Bonferroni’s inequality (2.2) does not take into account the correlation information among test statistics. It lacks power if several highly correlated tests are undertaken. Simes (1986) proposed a new procedure based on a conjectured inequality, instead of the Bonferroni’ inequality. By simulation study, Simes suggested that his procedure has an actual significance level closer to the nominal level in a wide range of circumstances and has more power than the Bonferroni procedure.

For the purpose of testing the overall hypothesis $H_0 = \cap_{i \in E} H_i$ of $n$ hypotheses $H_i, i \in E = \{1, ..., n\}$, Simes’ modified Bonferroni procedure rejects $H_0$ if $P_{(j)} \leq j\alpha/n$ for any
$j = 1, 2, ..., n$. Simes in his original paper proved that this procedure has the exact nominal level $\alpha$ under $H_0$ when all test statistics are independent. He also conjectured by simulation that the significance level under $H_0$ does not exceed $\alpha$ for a large family of multivariate distributions of the test statistics. Therefore, Simes’ procedure controls the FWER weekly under these situations. By construction, the Simes’ procedure is more powerful than the Bonferroni procedure since the former rejects any $H_0$ that is rejected by the latter, but not vice versa.

The Simes’ inequality can be defined as

$$P\{P_{(i)} \geq i\alpha/n, i = 1, ..., n \mid H_0 \text{ is true}\} \geq 1 - \alpha,$$  \hspace{1cm} (2.7)

which holds with equality if under $H_0$ the test statistics are independent and continuous, and may hold for dependent statistics with a variety of multivariate distributions with common marginals. Sarkar and Chang (1997) and Sarkar (1998) proved the Simes’ conjecture (2.2) holds for MTP2 random variables with common marginals.

Simes also raised the problem of making statements on individual hypotheses. The problem can be solved by utilizing Simes’ procedure to test all intersections of hypotheses in the closed testing procedure under the circumstances when Simes’ procedure is valid. Hommel (1988) and Hochberg (1988) proposed two such shortcut procedures controlling FWER strongly; see sections 2.3.2 and 2.3.3 for details.

### 2.3.2 Hommel’s Procedure

Given $n$ hypotheses $H_i, i \in E$, when the global intersection $H_0$ has been rejected by Simes’ procedure, the question remains which of the individual hypotheses $H_i$ should be rejected. Hommel (1988) proposed a solution by employing the closure principle to extend Simes’ procedure and then yield a FWER controlling procedure for making statements on individual hypotheses. It is worth noting that Hommel’s procedure works only when Simes’ procedure
Hommel’s procedure performs the decisions for individual hypotheses in the following way: compute $j = \max\{i \in \{1, ..., n\} : P_{(n-i+k)} > k\alpha/i$ for $k = 1, ..., i\}$. If the maximum does not exist, reject all $H_i, i = 1, ..., n$, otherwise reject all $H_i$ with $P_i \leq \alpha/j$.

2.3.2.1 Strong Control of FWER

The procedure strongly controls FWER at a level $\alpha$ can be shown by employing Simes’ procedures at level $\alpha$ to every intersection $H_{(i)}$ of the individual hypotheses $H_{(i)}, i \in I \subset E$, as follows:

First, if the $j$ does not exist, then for all $i \in \{1, ..., n\}$, $P_{(n-i+k)} \leq k\alpha/i$ for at least one $k = 1, ..., i$. When $i = 1$, $P_{(n)} \leq \alpha$, and so are all $P_{(i)}$ for $i < n$. For the intersection of an arbitrary collection of hypotheses including $H_{(n)}$, Simes’ procedure rejects it since $P_{(n)}$ will always be compared with $\alpha$. Therefore, $H_{(n)}$ is rejected by closed testing procedure. Next consider the intersection of an arbitrary collection of hypotheses including $H_{(n-1)}$ (only those collections not containing $H_{(n)}$ are needed to be considered.) Simes’ procedure rejects the intersection since $H_{(n-1)}$ has the largest P-value among all components of the intersection and $P_{(n-1)}$ is always compared with $\alpha$. Thus, $H_{(n-1)}$ is rejected. These steps are repeated so that all $H_{(i)}, i \in E$ can be rejected.

Next, if the $j$ exists, $P_{(n-j+k)} > k\alpha/j$ for $k = 1, ..., j$. So $\bigcap_{k=1}^{j} H_{(n-j+k)}$ is not significant by Simes’ procedure and all $H_{(n-j+k)}, k = 1, ..., j$ are retained by the closure principle. Also, if $P_s > \alpha/j$ for some $H_s \in \{H_{(1)}, ..., H_{(n-j)}\}$, the intersection hypothesis of $j$ null hypotheses $\{H_s, H_{(n-j+2)}, ..., H_{(n)}\}$ is retained by Simes’ procedure. So $H_s$ cannot be rejected by the closure principle.

The last, if the $j$ exists, suppose $P_s \leq \alpha/j$ for some $H_s \in \{H_{(1)}, ..., H_{(n)}\}$. It is clear that $P_s < P_{(n-j+1)}$, otherwise it conflicts with the definition of $j$. Let $H_{I_m}$ be the intersection of $m$ arbitrary elementary hypotheses $H_i, i \in I_m \subset E$ including $H_s$. To test $H_{I_m}$, Simes’ procedure compares the ordered p-values of the $m$ hypotheses with the critical values $k\alpha/m, k =
The following discussion shows that $H_{I_m}$ will be rejected and thus $H_s$ is rejected by the closure principle.

Case 1: If $m \leq j$, $P_s \leq \alpha/j \leq k\alpha/m, k = 1, \ldots, m$. So the intersection is rejected.

Case 2: When $m > j$, if there is at least one hypothesis $H_t \in \{H(1), \ldots, H(n-j)\}$ such that $P_t > \alpha/(j+1)$, then there must be a $k' \in \{1, \ldots, j-1\}$ such that $P_{(n-j+k')} \leq (k'+1)\alpha/(j+1)$. Otherwise, under either situation, $P_{(n-j)} > \alpha/(j+1)$. That implies $P_{(n-j+k)} > (k+1)\alpha/(j+1)$ for $k = 0, 1, \ldots, j$; it leads to a confliction with the definition of $j$. When comparing the ordered p-values $P_{(1)} \leq P_{(2)} \leq \ldots \leq P_{(m)}$ of the $m$ hypotheses $H_i, i \in I_m$ with the critical values $k\alpha/m, k = 1, \ldots, m$, there must be one $P' \leq P_{(n-j+k')}$ from those $m$ ordered p-values that is compared with $(m - (j-k' + 1) + 1)\alpha/m$ since there are at most $j - k'$ hypotheses whose p-values are greater than $P_{n-j+k'}$. Since

$$P' \leq P_{(n-j+k')} \leq (k'+1)\alpha/(j+1) \leq (k' + 1 + (m-j-1))\alpha/(j+1 + (m-j-1)) = (m - (j - k' + 1) + 1)\alpha/m,$$

$H_{I_m}$ is rejected by Simes’ procedure.

Case 3: When $m > j$, if all hypothesis $H_t \in \{H(1), \ldots, H(n-j)\}$ have the p-values $P_t \leq \alpha/(j+1)$, without loss of generality, assume $P_s = P_{(n-j)}$. When applying Simes’ procedure to $H_{I_m}$, $P_{(n-j)}$ must be compared with a critical value that is not less than $(m - j)\alpha/m$. For $\alpha/(j+1) \leq (1 + (m-j-1))\alpha/(j+1 + (m-j-1)) = (m-j)\alpha/m$, $H_{I_m}$ is rejected and so is $H_{(n-j)}$. These steps is repeated to $H_{(n-j-1)}$ and it is rejected and so on, until all $\{H(1), \ldots, H_{(n-j)}\}$ are rejected.

In particular, the multiple level $\alpha$ is kept if the $n$ test statistics are independent.

Hommel’s procedure performs as a two-stage procedure. The first stage is to determine the $j$, which could be viewed as as estimate of the number of true null hypotheses $n_0$ using Simes’ test. The estimate is the largest size of the intersection hypothesis that can not be
rejected by Simes’ procedure. The second stage is to apply the Bonferroni procedure to the remaining hypotheses with respect to the estimation of \( n_0 \) in the first stage.

By construction, Hommel’s procedure is uniformly more powerful than Holm’s procedure since Simes’ procedure is uniformly more powerful than the Bonferroni procedure only if Simes’ procedure is valid.

### 2.3.3 Hochberg’s Procedure

Hochberg (1988) proposed a step-up procedure, another extension of the Simes’ procedure, which allows one to make inferences about the elementary null hypotheses \( H_{1i}, i \in E \). Hochberg’s procedure can be seen as a reversed Holm procedure, since it uses the same critical values, but in a reversed testing sequence. It is worth noting that, like Hommel’s procedure, Hochberg’s procedure works only when Simes’ procedure is valid.

\( H_i \) is rejected if there is a \( j = 1, ..., n \) such that \( P_{(j)} \leq \alpha/(n - j + 1) \) and \( P_i \leq P_{(j)} \). In other words, Hochberg’s procedure rejects all \( H_{(1)}, ..., H_{(j)} \), where \( j = \max\{i \in \{1, ..., n\} : P_{(i)} \leq \alpha/(n - i + 1)\} \). If such \( j \) does not exist, accept all hypotheses.

Alternatively, Hochberg’s procedure can be described by the following sequentially rejective test procedure. Start testing the null hypothesis \( H_{(n)} \) associated with the largest P-value \( P_{(n)} \). If \( P_{(n)} \leq \alpha \), the procedure stops and all hypotheses \( H_{(1)}, ..., H_{(n)} \) are rejected. Otherwise, \( H_{(n)} \) is retained and the procedure continues testing \( H_{(n-1)} \) at the smaller significance level \( \alpha/2 \). If \( P_{(n-1)} \leq \alpha/2 \), the procedure stops and all hypotheses \( H_{(1)}, ..., H_{(n-1)} \) are rejected. These steps are repeated until either the first rejection occurs or all null hypotheses \( H_{(1)}, ..., H_{(n)} \) are retained. By construction, Hochberg’s procedure is uniformly more powerful than Holm’s procedure when Simes’ procedure works.

In addition, the adjusted P-values of Hochberg’s procedure are

\[
q_{(i)} = \min\{\min[(n - i + 1)P_{(i)}, q_{(i+1)}], 1\}, i < n,
\]
where $q(n) = P(n)$. We reject $H(i)$ while strongly controlling FWER if $q(i)$ is less than $\alpha$.

### 2.3.3.1 Strong Control of FWER

Hochberg’s procedure has the strong control of FWER at a specified level $\alpha$ only when Simes’ procedure is valid, since Hochberg’s procedure is an application of the closure principle with Simes’ procedure to test all intersection hypotheses at level $\alpha$. The fact can be seen as follows:

Assume a collection of $H' \subseteq \{H_1, \ldots, H_n\}$ and let $H'_0$ be the intersection of $H'$. If the smallest P-value in $H'$ is $P(j)$, where $j = \max\{i \in \{1, \ldots, n\} : P(i) \leq \alpha/(n-i+1)\}$, the $H'_0$ is clearly rejected by Simes’ procedure. Next consider a different collection $H'$ of null hypotheses with exactly $k$ ($1 \leq k \leq j-1$) hypotheses whose p-values are smaller than $P(j)$. Simes’ procedure will compare $P(j)$ with a critical value that is not less than $(k+1)\alpha/(m-j+1+k)$, so the $H'_0$ is rejected for $\alpha/(m-j+1) \leq (k+1)\alpha/(m-j+1+k)$.

### 2.3.3.2 Comparison with Hommel’s Procedure

Compared with Hommel’s procedure, Hochberg’s procedure is easier to apply. However, it is shown that Hommel’s procedure is at least as powerful as Hochberg’s procedure, and, in general, more powerful (Hommel 1989). A connection between the Hommel’s procedure and Hochberg’s procedure can be seen by restating Hochberg’s procedure as follows:

Reject all $H_i$ with $P_i \leq \alpha/j$, where $j = \max\{i \in \{1, \ldots, n\} : P_{(n-i+k)} > \alpha/(i-k+1)\}$ for $k = 1, \ldots, i$. If the maximum does not exist, reject all $H_i, i = 1, \ldots, n$.

This statement differs from the Hommel’s procedure only in the critical values used in the definition of $j$. Note that $\alpha/(j-k+1) \geq k\alpha/j$, for $k = 1, \ldots, j$. For a given set of $n$ hypotheses of interest, the $j$ obtained by Hommel’s procedure will be at least as large as that in Hochberg’s procedure, so Hommel’s procedure rejects more false null hypotheses and thus has more power than Hochberg’s procedure.
2.4 Procedures Controlling k-FWER

Lehman and Romano (2005) first proposed the generalized familywise error rate (k-FWER), as illustrated in the section 1.2.1. k-FWER is a direct and natural generalization of FWER. It is less stringent than the traditional FWER so the k-FWER controlling procedures are usually more powerful than FWER controlling procedures. There have been many different methods being investigated to control k-FWER and FDP. Here we only review the procedures proposed by Lehman and Romano (2005) simply because they are closely related to the research of this dissertation.

2.4.1 Lehmann and Romano’s Procedure

Lehmann and Romano (2005) proposed two procedures that control the k-FWER: single step and step-down procedures. The single step procedure is a natural generalization of the classical Bonferroni procedure. The usual Bonferroni procedure compares each P-value $P_i$ with $\alpha/n$. Control of the k-FWER allows one to increase $\alpha/n$ to $k\alpha/n$, and thereby greatly increase the ability to detect false hypotheses. Such a simple modification results in control of the k-FWER. The procedure rejects any $H_i$ for which $P_i \leq k\alpha/n$. The authors proved the result utilizing the Markov’s inequality. We will see in the fourth Chapter that, the k-FWER controlling of the single step procedure also can be proved by the generalized Kwerel’s inequality (see section 4.1).

As the classical Holm’s procedure is the stepwise version of the Bonferroni procedure, Lehmann and Romano (2005) also proposed their stepwise type of modification of their single step procedure above. Their step-down procedure controlling k-FWER at level $\alpha$ utilizes the critical values defined as

$$
\alpha_p = \begin{cases} 
\frac{k\alpha}{n}, & p \leq k \\
\frac{k\alpha}{n+k-p}, & p > k,
\end{cases}
$$

(2.8)
Lehmann and Romano proved the k-FWER control of the step-down procedure without any assumption on the dependence structure of the individual tests. The proof is very similar to that for Holm’s procedure. But it is worth noticing that the classical closure principle is not valid for the k-FWER case any more. Whether a generalized version of the closure principle exists need be put under further investigation.

In this dissertation, we will propose one new step-down procedure controlling k-FWER that incorporates the correlation information of individual tests into the procedures and can be regarded as the modification of Lehmann and Romano’s procedure. Moreover, a Binomial-model-based, rather than probability-inequality-based, procedure is found for controlling k-FWER when all test statistics are mutually independent. See Chapter 4 for more details.

2.5 Procedures Controlling FDR

In modern scientific research and application, there arise many large-scale hypotheses testing problems involves hundreds and thousands of hypotheses as a joint family of interest. Under such circumstances, the familywise error rate control is very stringent since it controls to commit no false rejection when testing all null hypotheses simultaneously. The stringent control will lead to very low power. For the reason, another type of error rate for multiple testing is suggested and widely employed to make a relatively less stringent control of the Type I error: the proportion of false rejection among all rejections (FDP), as well as its expectation, the false discovery rate (FDR). Since its debut in Benjamini and Hochberg’s seminal 1995 paper, false discovery rate control has claimed an increasing portion of statistical research, both applied and theoretical. In this section, we just review the influential Benjamini and Hochberg’s (BH) Procedure to give an introduction to FDR controlling procedures.
2.5.1 Benjamini and Hochberg’s (BH) Procedure

When Benjamini and Hochberg (1995) proposed the false discovery rate at the first time, they mentioned the motivations behind their new point of view on the problem of multiplicity: Classical procedures that controls the FWER in the strong sense, at levels conventional in single-comparison problems, tend to have substantially less power than the PCER controlling procedures of the same levels; the control of FWER is often not quite needed. So they suggested the number of erroneous rejection should be taken into account and not only the question whether any error was committed. Therefore, they proposed to control the expected proportion of errors among all the rejections, which is defined as

\[ FDR = E \left( \frac{V}{R} \mid R > 0 \right) \frac{Pr(R > 0)}{Pr(R > 0)} \]

using the notation in table 1.1.

Benjamini and Hochberg (1995) also proposed a Bonferroni-type multiple testing procedure for a fixed value \( q \in (0, 1) \), which was credited to Simes (1986) with an early version: The BH procedure rejects all \( H_{(i)}, i = 1, ..., k \), where \( k = \max \{ k \in \{1, ..., n\} : P_{(i)} \leq \frac{i}{n} q \} \).

Benjamini and Hochberg proved the theorem that if the p-values corresponding to the true null hypotheses are independent of each other, then the BH procedure controls the FDR at \( q \).

If all null hypotheses are true, the FDR is equivalent to the FWER, so control of FDR implies control of the FWER in the weak sense. Otherwise, the FDR is smaller than or equal to the FWER; compared with FWER controlling procedures, the potential for increase in power is larger when more of the hypotheses are false. The advantage in power of the BH procedure increases in \( n \), the number of all null hypotheses.
CHAPTER 3

NOVEL STEP-DOWN PROCEDURES CONTROLLING FWER

In this chapter, we propose a novel step-down procedure strongly controlling the FWER using the bivariate joint distributions of test statistics. The critical values of the procedures are uniformly larger than the Holm’s procedure and CS procedure under any kind of dependency, as long as the bivariate joint distributions are available. Unlike the CS procedures, the proposed procedures are determined by a new algorithm using the provided bivariate joint distributions of the test statistics. The theoretical proof shows that the procedures can strongly control the FWER while the simulation analysis demonstrates that they are uniformly and significantly more powerful than Holm’s and the CS procedures, especially when test statistics are highly correlated.

3.1 Probability Inequalities

In this section, we review some important probability inequalities that will be used in our proposed methods. They include the classical Bonferroni inequality, Kounias’ inequality, and
Hunter’s inequality. Because the development of our proposed methods only involves the upper bounds of probabilities that at least \( m \) of \( n(n \geq m) \) events occur, all these inequalities considered in this sections are for the upper bounds except that the Bonferroni’s gives both upper and lower bounds.

### 3.1.1 Bonferroni Inequality

Let \( A_1, A_2, ..., A_n \) be events on a given probability space \((\Omega, \mathcal{A}, P)\), and let \( v_n = v_n(A) \) be the number of those \( A_j \)'s, \( 1 \leq j \leq n \), that occur. Set \( S_0 = 1 \) and, for \( k \geq 1 \), let

\[
S_k = \sum P(A_{i_1} A_{i_2} ... A_{i_k}),
\]

where \( A_{i_1} A_{i_2} = A_{i_1} \cap A_{i_2} \), and \( \sum \) is summation over all integers \( 1 \leq i_1 < i_2 < ... < i_k \leq n \). Assume that \( S_k \) can, at least with good accuracy, be determined. Then, based on the value of \( S_k \), we evaluate, or approximate, the values \( P(v_n \geq m) \) for some values of \( m \).

For \( 0 \leq m \leq n \), \( n \geq 1 \), and for integers \( d \geq 0 \) and \( r \geq 0 \),

\[
\sum_{k=0}^{2d+1} (-1)^k \binom{k + m + 1}{m - 1} S_{k+m} \leq P(v_n \geq m) \leq \sum_{k=0}^{2r} (-1)^k \binom{k + m - 1}{m - 1} S_{k+m}.
\]

(3.1)

The inequalities (3.1) are referred to as the classical Bonferroni inequalities. The Bonferroni and Holm procedures are derived based on the special case of its upper bounds when \( m = 1, r = 0 \), i.e.

\[
P(v_n \geq m) \leq S_1 = \sum_{i=1}^{n} P(A_i).
\]

(3.2)

For multiple testing problems of \( n \) hypotheses, \( H_1, ..., H_n \), the events \( A_i, i = 1, ..., n \) are usually \{ \( H_i \) is rejected \} or \{ \( R_i < \alpha_i \) \}, where \( R_i \) is the p-value for testing \( H_{i0} \) vs \( H_{i1} \) and \( \alpha_i \) the corresponding predetermined testing level. Under this setting, \( P(v_n \geq 1) \) means the probability that at least one of all true null hypotheses \( H_{i0}, i = 1, ..., n \) are rejected.
When the bivariate joint distributions of test statistics are provided, the inequalities sharpened by incorporating only the degree-2 terms like $S_2$ into the (3.2) can help sharpen the Holm procedure. Such inequalities obviously cannot be obtained from the Bonferroni inequalities, since it cannot give a upper bound for $P(v_n \geq 1)$ containing only $S_1$ and $S_2$, noting that if $m = 1, r = 1$, the right hand side of (3.1) becomes

$$P(v_n \geq 1) \leq S_1 - S_2 + S_3,$$

which also contains degree-3 terms $S_3$ that is not provided.

### 3.1.2 Kwerel’s Inequality

We introduce an upper bound for $P(v_n \geq 1)$ in terms of $S_1$ and $S_2$ only. The following inequality was considered and first solved by Kwerel (1975), according to Galambos and Simonelli (1996).

$$P(v_n \geq 1) \leq S_1 - \frac{2}{n} S_2 = \sum_{i=1}^{n} P(A_i) - \frac{2}{n} \sum_{ij} P(A_i A_j).$$  \hfill (3.3)

Furthermore, the preceding bounds are the sharpest among all the bounds of the type

$$\{aS_1 + bS_2\},$$

where $a$ and $b$ are real numbers.

Kwerel’s inequality is obviously sharper than (3.2) considering that it has an extra degree-2 term. But we can obtain more accurate upper bounds by using other kinds of degree-2 terms other than $S_2$ as in (3.3). The proof of this inequality is in section 4.1.
3.1.3 Kounias’ Inequality

Kounias (1968) derived the following upper bound for $P(v_n \geq 1)$,

$$P(v_n \geq 1) \leq \sum_{i=1}^{n} P(A_i) - \max_{1 \leq i \leq n} \sum_{j=1, j \neq i}^{n} P(A_i A_j). \quad (3.4)$$

Kounias inequality is sharper than not only (3.2) but also (3.3). It is better than (3.3) can be seen by noting that

$$2S_2 = \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} P(A_i A_j),$$

and so,

$$\frac{2S_2}{n} = \frac{\sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} P(A_i A_j)}{n} \leq \max_{1 \leq i \leq n} \sum_{j=1, j \neq i}^{n} P(A_i A_j).$$

Seneta and Chen’s procedure (2005) uses Kounias inequality to sharpen Holm procedure and proposed the CS procedure.

3.1.4 Hunter’s Inequality

Hunter (1976) developed the following inequality under the general framework based on the concept of graphs. Let $V = \{e_1, e_2, ..., e_n\}$ be an arbitrary set of distinct elements, and let $E$ be a subset of the ordered pairs $\{(e_i, e_j) | e_i, e_j \in V, (e_i, e_j) = (e_j, e_i)\}$. We further assume that $(e_i, e_i)$ is not the in $E$ for all $e_i \in E$. The pair $G = (V, E)$ is called a graph with vertex set $V$ and edge set $E$. Let $I_n = \{1, 2, ..., n\}$ be a vertex set. A graph $G$ is called a tree if to every vertex $i$ and $j$ there is a unique set of further vertices $i_1, i_2, ..., i_t$ such that with $i_0 = i$ and $j = i_{t+1}, (i_s, i_{s+1}) \in E, 0 \leq s \leq t$ (i.e., a unique path from $i$ to $j$). Let $\mathcal{T} = \{T = (I_n, E)|T \text{ is a tree}\}$. Then we have Hunter’s inequality given by, for arbitrary $T \in \mathcal{T}$,

$$P(v_n \geq 1) \leq \sum_{i=1}^{n} P(A_i) - \sum_{T} P(A_i A_j), \quad (3.5)$$
where the second sum is over all edges \((i,j)\) in \(T\). See Janos and Italo (1996) for the proof details.

By letting \(T = T_j\) be the tree with one edge between vertex \(j\) and vertex \(i\) for all \(i \neq j\), we immediately get Kounias’ inequality (3.4) as a special case of (3.5). So Hunter’s inequality provides the sharpest upper bound for \(P(v_n \geq 1)\) among all inequalities from (3.2) – (3.5), while it need the most complicated computation when scanning all possible trees \(T \in \mathcal{T}\) to seek the sharpest one.

### 3.2 Novel Step-Down Procedures Controlling FWER

#### 3.2.1 Procedure I

Consider the multiple testing problem where there are \(n\) hypotheses \(H_1, ..., H_n\) and corresponding p-values \(R_1, ..., R_n\). Assume that the test statistics \(X_1, ..., X_n\) are from continuous probability distributions. Let \(R_1, ..., R_n\) be the ordered p-values and \(H_1, ..., H_n\) the corresponding hypotheses. Writes \(R_i = R_{t_i}\), where \(t_i\) is a random variable taking values from \(\{1, 2, ..., n\}\). Define the index set \(L(p) = \{t_p, t_{p+1}, ..., t_n\}\), which is a random set. Furthermore, assume that the bivariate joint distributions of any two test statistics are known.

The first step-down procedure we investigate here is similar to the CS procedure since its sequence of critical values is in the same form as (2.5) except that the values of the degree-2 terms \(\gamma\) are increased by a new algorithm.

In order to define our procedures, we first define the degree-2 terms based on inequalities (3.4) and (3.5) in section 3.1 and the bivariate joint distributions of test statistics, as follows

\[
\gamma_K(S, c) = \max_{j \in S} \sum_{i \in S \setminus \{j\}} P(R_i \leq c, R_j \leq c | H_s, s \in S, \text{true}), \tag{3.6}
\]

\[
\gamma_H(S, c) = \max_{T \in \mathcal{T}} \sum_{(i,j) \in T} P(R_i \leq c, R_j \leq c | H_s, s \in S, \text{true}), \tag{3.7}
\]
where \( S \subset \{1, 2, \ldots, n\} \) with size \( |S| = m > 1 \) and \( c \in (0,1) \), \( T_S \) is the set of all trees defined on \( S \), and \((i, j)\) the edge in the tree \( T \). In addition, define \( \gamma_K(S, c) = \gamma_H(S, c) = 0 \) if \( m = 1 \). And \( \gamma_K \) and \( \gamma_H \) are defined based on inequalities (3.4) and (3.5) respectively. Thus, in the CS procedure (section 2.2.3), the degree-2 terms (2.6) can be denoted by \( \gamma_K(L(p), \frac{\alpha}{n-p+1}) \), \( p = 1, 2, \ldots, n \). Our procedure improves the CS procedure by increasing the value of the second argument in \( \min \) function in (2.5), \( \frac{\alpha + \gamma_p}{n-p+1} \).

To define the critical points in our procedures, we need to apply a new iterative algorithm as follows: for any subset \( S \subset \{1, 2, \ldots, n\} \) with \( |S| = m \), \( m \leq n \), and the degree-2 terms defined above, say \( \gamma_K(S, c) \), the following algorithm generates a sequence of values defined as:

\[
\begin{align*}
C_{K,S,0} &= \alpha \\
C_{K,S,1} &= \frac{\alpha + \gamma_K(S, C_{K,S,0})}{m} \\
C_{K,S,2} &= \frac{\alpha + \gamma_K(S, C_{K,S,1})}{m} \\
&\quad \cdots \\
C_{K,S,d} &= \frac{\alpha + \gamma_K(S, C_{K,S,d-1})}{m} \\
&\quad \cdots
\end{align*}
\]  

(3.8)

Similarly, with \( \gamma_H(S, c) \), we can define the similar sequence, denoted by \( \{C_{H,S,d}\}_{d=0}^{\infty} \).

**Proposition 3.2.1.** The sequence \( \{C_{K,S,d}\}_{d=0}^{\infty} \) (\( \{C_{H,S,d}\}_{d=0}^{\infty} \)), defined in (3.8) using Kounias’ degree-2 terms (3.6) (Hunter’s degree-2 terms (3.7)) for given \( S \subset \{1, 2, \ldots, n\} \) with \( |S| = m \) is non-increasing at \( d \) and has lower and upper bounds \( \frac{\alpha + \gamma_K(S, \frac{\alpha}{m})}{m} \) (\( \frac{\alpha + \gamma_H(S, \frac{\alpha}{m})}{m} \)) and \( \alpha \) respectively. Therefore, these sequences converge to some points, denoted by \( C_{K,S} \) (\( C_{H,S} \)), in \( [\frac{\alpha + \gamma_K(S, \frac{\alpha}{m})}{m}, \alpha] \) (\( [\frac{\alpha + \gamma_H(S, \frac{\alpha}{m})}{m}, \alpha] \)).
Proof. First note that \( \gamma_K(S,c) \) is non-decreasing at \( c \). Actually, for \( c_1 \leq c_2 \in (0,1) \), we have

\[
\gamma_K(S,c_1) = \max_{j \in S} \sum_{i \in S \setminus \{j\}} P(R_i \leq c_1, R_j \leq c_1 | H_s, s \in S, true)
\]

\[
= \sum_{i \in S \setminus \{j_1\}} P(R_i \leq c_1, R_{j_1} \leq c_1 | H_s, s \in S, true)
\]

\[
\leq \sum_{i \in S \setminus \{j_1\}} P(R_i \leq c_2, R_{j_1} \leq c_2 | H_s, s \in S, true)
\]

\[
\leq \max_{j \in S} \sum_{i \in S \setminus \{j\}} P(R_i \leq c_2, R_j \leq c_2 | H_s, s \in S, true)
\]

\[
= \gamma_K(S,c_2).
\]

Then it can be shown by induction that the sequence \( \{C_{K,S,d}\}_{d=0}^{\infty} \) for given \( S \) is non-increasing. When \( d = 1 \), \( C_{K,S,1} = \frac{\alpha + \gamma_K(S,C_{K,S,0})}{m} = \frac{\alpha + \gamma_K(S,\alpha)}{m} \leq \frac{\alpha + (m-1)\alpha}{m} = \alpha = C_{K,S,0} \).

Assume for \( d = k \), \( C_{K,S,k} \leq C_{K,S,k-1} \). Then for \( d = k + 1 \), \( C_{K,S,k+1} = \frac{\alpha + \gamma_K(S,C_{K,S,k})}{m} \leq \frac{\alpha + \gamma_K(S,\frac{\alpha}{m})}{m} = C_{K,S,k} \), considering that \( \gamma_K(S,c) \) is non-decreasing at \( c \). In addition, \( C_{K,S,d} \leq C_{K,S,0} = \alpha \), for any integer \( d \). Finally, the fact that \( \{C_{K,S,d}\}_{d=0}^{\infty} \) is lower bounded by \( \frac{\alpha + \gamma_K(S,\frac{\alpha}{m})}{m} \) follows after a simple induction. From (3.8),

\[
C_{K,S,1} = \frac{\alpha + \gamma_K(S,\alpha)}{m} \geq \frac{\alpha + \gamma_K(S,\frac{\alpha}{m})}{m}.
\]

Assume that for some integer \( d \), \( C_{K,S,d} \geq \frac{\alpha + \gamma_K(S,\frac{\alpha}{m})}{m} \geq \frac{\alpha}{m} \). Then

\[
C_{K,S,d+1} = \frac{\alpha + \gamma_K(S,C_{K,S,d})}{m} \geq \frac{\alpha + \gamma_K(S,\frac{\alpha}{m})}{m}.
\]

Thus, the sequence \( \{C_{K,S,d}\}_{d=0}^{\infty} \) converges some point in \( [\frac{\alpha + \gamma_K(S,\frac{\alpha}{m})}{m}, \alpha] \). Similar results for \( \{C_{H,S,d}\}_{d=0}^{\infty} \) can be obtained with the same rationale. \( \square \)

Let \( C_{K,S} = \lim_{d \to \infty} C_{K,S,d} \). Actually, taking limits of both sides of (3.8) implies that \( C_{K,S} \)
is one solution of the equation
\[ C = \frac{\alpha + \gamma_K(S,C)}{m} \]  
(3.9)
with the variable \( C \) in the interval \( (\frac{\alpha}{m}, \alpha) \), i.e.
\[ mC_{K,S} - \gamma_K(S,C_K) = \alpha. \]  
(3.10)
The critical points in this proposed procedure are defined as
\[ \alpha_p = \begin{cases} 
\min\left(\frac{\alpha}{n-p}, C'_{K,p}\right), & p = 1, 2, ..., n - 1 \\
\alpha, & p = n 
\end{cases} \]  
(3.11)
where \( C'_{K,p} = C_{K,L(p)} \), \( p = 1, 2, ..., n \). Like the CS procedure, the proposed procedure also controls the FWER in the strong sense. To show this, let any \( I \subset E \) be the indexes set of all true hypotheses with the size \( |I| = m \leq n \). Let \( j \) be the smallest (random) index satisfying
\[ R(j) = \min_{i \in I} R_i. \]  
(3.12)

**Theorem 3.2.2** (Procedure I). **The step-down multiple testing procedure with critical values \( \alpha_p \) defined in (3.11) strongly controls the FWER at predetermined level \( \alpha \) for testing \( H_1, ..., H_n \).**

**Proof.** If \( m = 1 \), the case is trivial since \( \alpha_p \leq \alpha \) for all \( p \). Assume that \( m > 1 \). Note that \( j \leq n - m + 1 \) (or \( m \leq n - j + 1 \)). Consider the probability of all true hypotheses being retained:

If \( j = n - m + 1 \), i.e., the \( R_i, i \in I \), are in fact the largest \( m \) values \( R(i), i = n - m + 1, ..., n \), \( C'_{K,j} = C'_{K,n-m+1} = C_{K,i} \). Thus,
\[ A = Pr(\text{all } H_i, i \in I \text{ are retained by the procedure } | \cap_{i \in I} H_i \text{ is true}) \]
\[ \geq Pr(R(j) \geq \min\left(\frac{\alpha}{n-j}, C'_{K,j}\right) | \cap_{i \in I} H_i \text{ is true}) \]
\[ Pr(R_j) \geq \min\left(\frac{\alpha}{m-1}, C_{K,I} \right) | \cap_{i \in I} H_i \text{ is true} \]. \tag{3.13} \]

If \( j < n - m + 1 \),

\[
A \geq Pr(R_j) \geq \min\left(\frac{\alpha}{n-j}, C'_{K,j} \right) | \cap_{i \in I} H_i \text{ is true} \\
\geq Pr(R_j) \geq \frac{\alpha}{n-j} | \cap_{i \in I} H_i \text{ is true} \\
\geq Pr(R_j) \geq \frac{\alpha}{m} | \cap_{i \in I} H_i \text{ is true} \\
\geq Pr(R_j) \geq \min\left(\frac{\alpha}{m-1}, \frac{\alpha + \gamma_K(I, C_{K,I})}{m} \right) | \cap_{i \in I} H_i \text{ is true} \\
= Pr(R_j) \geq \min\left(\frac{\alpha}{m-1}, C_{K,I} \right) | \cap_{i \in I} H_i \text{ is true}. \tag{3.14} \]

The last equality is seen by noting that \( \frac{\alpha + \gamma_K(I, C_{K,I})}{m} = C_{K,I} \) by (3.9).

In sum, from (3.13) and (3.14), we have

\[
A \geq Pr(R_j) \geq \min\left(\frac{\alpha}{m-1}, C_{K,I} \right) | \cap_{i \in I} H_i \text{ is true} \\
\geq Pr(R_i) \geq \min\left(\frac{\alpha}{m-1}, C_{K,I} \right), \text{ all } i \in I | \cap_{i \in I} H_i \text{ is true} \\
\geq 1 - Pr(R_i < \min\left(\frac{\alpha}{m-1}, C_{K,I} \right), \text{ some } i \in I | \cap_{i \in I} H_i \text{ is true} \\
\geq 1 - Pr(R_i < C_{K,I}, \text{ some } i \in I | \cap_{i \in I} H_i \text{ is true} \\
\geq 1 - \sum_{i \in I} Pr(R_i < C_{K,I} | H_i \text{ is true}) + \gamma_K(I, C_{K,I}) \\
\geq 1 - \sum_{i \in I} Pr(R_i < \alpha + \frac{\gamma_K(I, C_{K,I})}{m} | H_i \text{ is true}) + \gamma_K(I, C_{K,I}) \\
\geq 1 - m \times \frac{\alpha + \gamma_K(I, C_{K,I})}{m} + \gamma_K(I, C_{K,I}) \\
\geq 1 - \alpha. \]

by the inequality (3.4);
Thus, the $FWER \leq 1 - (1 - \alpha) = \alpha$ for any $I \subset E$.

Next, we make a few remarks about our procedure I:

I.1 From Proposition (3.2.1), the procedure I has larger critical values than those of the CS procedure. So the procedure I is more powerful than the CS procedure.

I.2 The procedure I has the same limitation as the CS procedure that the $p$th critical value are conservatively upper bounded by $\alpha/(n - p)$, as defined in (3.11). We will propose other procedures next sections that can break the limit.

I.3 The procedure I could be improved by replacing $\gamma_K$ in (3.9) – (3.11) with $\gamma_H$ such that the FWER can be still strongly controlled with potentially more power, since Hunter’s inequality gives sharper upper bound than Kounias’. But the computation complexity will be significantly increased because in order to find the optimal tree $T$ in (3.7), it has to scan all $n^{n-2}$ trees on $n$ vertices, while it scans only $n$ trees for calculating (3.6).

### 3.2.2 Procedure II

In order to break the limitation that the $p$th critical value are conservatively upper bounded by $\alpha/(n - p)$, as defined in (3.11), we investigate another new procedure in this section. We first define one kind of degree-2 terms just like (3.6) and (3.7) using bivariate joint distributions that will be used to define our procedure.

$$\gamma_M(S, c) = (|S| - 1) \min_{(i,j) \in S} P(R_i \leq c, R_j \leq c| H_s, s \in S, true),$$

where $S \subset \{1, 2, ..., n\}$ with size $|S| = m > 1$ and $c \in (0, 1)$. In addition, define $\gamma_M(S, c) = 0$ if $m = 1$. Moreover, assume that $\gamma_M(S, c) = (|S| - 1)P(R_{i_0} \leq c, R_{j_0} \leq c| H_{i_0}, H_{j_0} true)$ for some $(i_0, j_0) \subset S$ and any $c \in (0, 1)$. In other words, the choice of $(i_0, j_0)$ does not depend on the value of $c$. 
Remark 3.2.3. With $\gamma_K$ in (3.8) being replaced with $\gamma_M$, we can apply algorithm (3.8) to correspondingly obtain a sequence $\{C_{M,S,d}\}_{d=0}^{\infty}$ with the limit $C_{M,S} \in [\frac{\alpha + \gamma_M(S, \alpha)}{m}, \alpha]$. The results corresponding to Proposition (3.2.1) and (3.9) hold true for $\{C_{M,S,d}\}_{d=0}^{\infty}$ and $C_{M,S}$.

The Proof is similar to that of Proposition 3.2.1 and will not be given again.

Now we propose another step-down procedure using the critical values defined as,

$$\alpha_p = C'_{M,p} = C_{M,L(p)}, p = 1, 2, \ldots, n.$$  \hfill (3.16)

Note that when $p = n$, $\alpha_n = C_{M,L(n)} = \alpha$ according to the algorithm 3.8. The proposed procedure II also controls the FWER in the strong sense. To prove this, we need a lemma as follows.

Lemma 3.2.4. For any subsets $S_1, S_2 \subset \{1, \ldots, n\}$ with $|S_1| = m_1$ and $|S_2| = m_2$ that satisfying $S_1 \subset S_2$, the corresponding limit values $C_{M,S_1}$ and $C_{M,S_2}$ obtained in the algorithm (3.8) using (3.15) satisfy

$$C_{M,S_1} \geq C_{M,S_2}.$$  

Proof. It suffices to show the inequality $C_{M,S_1,k} \geq C_{M,S_2,k}$ for any nonnegative integer $k$, where $C_{M,S_1,k}$ are obtained in the intermediate steps in the algorithm (3.8), since the lemma can be shown simply by taking the limits of both side of the inequalities. The inequalities are shown by induction as follows.

When $k = 0$, $C_{M,S_1,0} = C_{M,S_2,0} = \alpha$.

Then assume that, for any $k \in \{1, 2, \ldots, n\}$, $C_{M,S_1,k-1} \geq C_{M,S_2,k-1}$. To see $C_{M,S_1,k} \geq C_{M,S_2,k}$, i.e.

$$\frac{\alpha + \gamma_M(S_1, C_{M,S_1,k-1})}{m_1} \geq \frac{\alpha + \gamma_M(S_2, C_{M,S_2,k-1})}{m_2}.$$  

It is equivalent to show
\begin{align}
& m_1 \alpha + m_1 (m_2 - 1) \min_{S_2} P(R_i \leq C_{M,S_2,k-1}, R_j \leq C_{M,S_2,k-1}) \\
& \leq m_2 \alpha + m_2 (m_1 - 1) \min_{S_1} P(R_i \leq C_{M,S_1,k-1}, R_j \leq C_{M,S_1,k-1}) \\
& \leq m_2 \alpha + m_2 (m_1 - 1) \min_{S_1} P(R_i \leq C_{M,S_1,k-1}, R_j \leq C_{M,S_1,k-1}) \\
& \leq m_2 \alpha + m_2 (m_1 - 1) \min_{S_1} P(R_i \leq C_{M,S_1,k-1}, R_j \leq C_{M,S_1,k-1})
\end{align}

By the assumption that $C_{M,S_1,k-1} \geq C_{M,S_2,k-1}$ and that $S_1 \subseteq S_2$,

\begin{align}
& \min_{S_2} P(R_i \leq C_{M,S_2,k-1}, R_j \leq C_{M,S_2,k-1}) \\
& \leq \min_{S_1} P(R_i \leq C_{M,S_2,k-1}, R_j \leq C_{M,S_2,k-1}) \\
& \leq \min_{S_1} P(R_i \leq C_{M,S_1,k-1}, R_j \leq C_{M,S_1,k-1}) \\
& \leq \min_{S_1} P(R_i \leq C_{M,S_1,k-1}, R_j \leq C_{M,S_1,k-1})
\end{align}

Also note that

\begin{align}
& m_1 \alpha + m_1 (m_2 - 1) \min_{S_1} P(R_i \leq C_{M,S_1,k-1}, R_j \leq C_{M,S_1,k-1}) \\
& \leq m_2 \alpha + m_2 (m_1 - 1) \min_{S_1} P(R_i \leq C_{M,S_1,k-1}, R_j \leq C_{M,S_1,k-1}) \\
& \leq (m_2 - m_1) \min_{S_1} P(R_i \leq C_{M,S_1,k-1}, R_j \leq C_{M,S_1,k-1}) \leq (m_2 - m_1) \alpha.
\end{align}

Thus, (3.17) holds by (3.18) and (3.19).

\textbf{Remark 3.2.5.} An immediate result of the lemma is that the critical points $C'_{M,p}, p = 1, \ldots, n$ in our procedure is a monotonically nondecreasing sequence. Actually these critical points are random and sample-dependent since so are $L(p)$. Assume $I$, of fixed size $m$, be the index set of all true hypotheses and thus it is fixed but unknown. Let $R_{(j)},$ for some $j \leq n-m+1$, is the smallest value of $R_i, i \in I$, thus $L(j) \subseteq I$. Then by the lemma, $C'_{M,j} \leq C_{M,I}$.

\textbf{Theorem 3.2.6 (Procedure II).} The step-down multiple testing procedure taking $\alpha_p$ defined in (3.11) as critical points strongly controls the FWER at predetermined significance level $\alpha$,
that is,

\[ P(H_s, s \in I, \text{are retained by the procedure}|H_s, s \in I, \text{true}) \geq 1 - \alpha, \]

where \( I \), any non-null subset of \( \{1, 2, ..., n\} \), is the index set of all \( m \) true true hypotheses.

Proof. Conditional on the event that \( \{H_s, s \in I, \text{true}\} \),

\[
P(H_s, s \in I, \text{are retained}) \geq P(R_{(j)} > C'_{M,j}) \geq P(R_{(j)} > C_{M,I}) \]

by the Remark 3.2.5;

\[ = P(R_i > C_{M,I}, i \in I) \]
\[ = 1 - P(R_i \leq C_{M,I}, \text{some } i \in I) \]
\[ \geq 1 - \sum_{i \in I} P(R_i \leq C_{M,I}) + \gamma_K(I, C_{M,I}) \]

by the Kounias’ inequality;

\[ \geq 1 - mC_{M,I} + \gamma_K(I, C_{M,I}) \]
\[ \geq 1 - m\left(\frac{\alpha + \gamma_M(I, C_{M,I})}{m}\right) + \gamma_K(I, C_{M,I}) \]

by (3.9);

\[ \geq 1 - \alpha, \text{ noticing that } \gamma_M(I, C_{M,I}) \leq \gamma_K(I, C_{M,I}). \]

It can be easily seen that the proposed procedure is uniformly more powerful than Holm’s procedure. The question naturally arises whether it also uniformly outperforms the CS procedure. The question can be partially answered by the following corollary. It gives the positive answer under the situation that all test statistics are identically distributed and equally correlated.
Corollary 3.2.7. If all test statistics are identically distributed and mutually equally correlated and all tests for hypotheses $H_i, i = 1, \ldots, n$ are of the same type (right-tailed, left-tailed, or two-tailed), the critical points $C'_{M,p}, p = 1, \ldots, n$ are uniformly larger than those in the CS procedure. Therefore, the step-down procedure II is uniformly more powerful than the CS procedure.

Proof. The assumption in the corollary implies that, for any $c \in (0, 1)$, $P(R_i \leq c, R_j \leq c) = P(R_s \leq c, R_t \leq c)$ for any $i \neq j, s \neq t \in \{1, 2, \ldots, n\}$. So $\gamma_M(L(p), \frac{\alpha}{n-p+1}) = \gamma_K(L(p), \frac{\alpha}{n-p+1})$ for any $p = 1, \ldots, n$. Furthermore,

$$C'_{M,p} \geq \frac{\alpha + \gamma_M(L(p), \frac{\alpha}{n-p+1})}{n-p+1} = \frac{\alpha + \gamma_K(L(p), \frac{\alpha}{n-p+1})}{n-p+1}.$$ 

The first inequality holds due to the Remark 3.2.3, and also the right hand side above is no less than the critical point $\min(\frac{\alpha}{n-p}, \frac{\alpha + \gamma_K(L(p), \frac{\alpha}{n-p+1})}{n-p+1})$ in the $p$th step in the CS procedure. Therefore, the procedure II is uniformly powerful than the CS procedure.

Next, we make a few remarks about our procedure II:

II.1 The advantage of procedure II over the CS procedure and our procedure I is that it breaks the limitation that the $p$th critical value is conservatively upper bounded by $\alpha/(n-p)$.

II.2 The procedure II does not always outperform the other two procedures. When the correlations among test statistics vary greatly, the CS procedure and the procedure I might give larger critical values than the procedure II (see Figure 3.1 for an example).

3.2.3 Procedure III

So far we have proposed two step-down procedures. None of them is absolutely better than the other. Both of them have their advantages and disadvantages. In this section, we
investigate another procedure inheriting the merits from the two procedures so that it is uniformly the best of all under any kind of dependence between test statistics.

The proposed step-down procedure III uses the critical values defined as,

\[
\alpha_p = \begin{cases} 
\max\{\min\left(\frac{\alpha}{n-p}, C'_{K,p} \right), C'_{M,p}\}, & p = 1, 2, \ldots, n-1 \\
\alpha, & p = n 
\end{cases}
\] (3.20)

where \( C'_{K,p} \) and \( C'_{M,p} \) are defined as (3.13) and (3.16). Actually, the new critical values are defined as the maximum of those of the procedure I and II. Thus the proposed procedure III is uniformly the most powerful among all. Next let’s show that it also controls the FWER in the strong sense. A lemma is needed.

**Lemma 3.2.8.** For any subsets \( S_1, S_2 \subset \{1, \ldots, n\} \) with \( |S_1| = m_1 \) and \( |S_2| = m_2 \) that satisfying \( S_1 \subset S_2 \),

\[
\max\{\min\left(\frac{\alpha}{m_1-1}, C_{K,S_1} \right), C_{M,S_1}\} \geq \max\{\min\left(\frac{\alpha}{m_2-1}, C_{K,S_2} \right), C_{M,S_2}\},
\]

where \( C_{K,S_1} \), \( C_{K,S_2} \), \( C_{M,S_1} \) and \( C_{M,S_2} \) are the limiting values defined using the algorithm (3.8).

**Proof.** On one hand, by the Lemma 3.2.4, \( C_{M,S_1} \geq C_{M,S_2} \). On the other hand, we have

\[
\min\left(\frac{\alpha}{m_1-1}, C_{K,S_1}\right) \geq C_{K,S_1} \geq \frac{\alpha}{m_1} \geq \frac{\alpha}{m_2-1} \geq \min\left(\frac{\alpha}{m_2-1}, C_{K,S_2}\right).
\]

The first and last inequalities are obvious. The second is by the Proposition 3.2.1. The third is because \( S_1 \subset S_2 \), i.e. \( m_1 < m_2 \). Thus, the lemma is proved. \( \square \)

**Theorem 3.2.9 (Procedure III).** The step-down multiple testing procedure taking \( \alpha_p \) defined in (3.20) as critical points strongly controls the FWER at predetermined significance level \( \alpha \), that is,

\[
P(H_s, s \in I, \text{are retained by the procedure}|H_s, s \in I, \text{true}) \geq 1 - \alpha,
\]
where I, any non-null subset of \{1, 2, ..., n\}, is the index set of all m true hypotheses.

Proof. If m = 1, the case is trivial since \( \alpha_p \leq \alpha \) for all p. Assume that \( m > 1 \). Conditional on the event that \{\( H_s, s \in I, true \)\},

\[
P(H_s, s \in I, \text{are retained}) \\
\geq P(R_{i_j} > \max[\min(\frac{\alpha}{n-j}, C_{K,j}'), C_{M,j}']) \\
\geq P(R_{i_j} > \max[\min(\frac{\alpha}{m-1}, C_{K,I}), C_{M,I}]) \\
\quad \text{by the fact that } I \subset L(p) \text{ and the Lemma 3.2.8}; \\
= P(R_i > \max[\min(\frac{\alpha}{m-1}, C_{K,I}), C_{M,I}], i \in I) \\
= 1 - P(R_i \leq \max[\min(\frac{\alpha}{m-1}, C_{K,I}), C_{M,I}], \text{some } i \in I) \\
\geq 1 - \sum_{i \in I} P(R_i \leq \max[\min(\frac{\alpha}{m-1}, C_{K,I}), C_{M,I}]) + \gamma_K(I, \max[\min(\frac{\alpha}{m-1}, C_{K,I}), C_{M,I}]) \\
\quad \text{by the Kounias' inequality}; \\
\geq 1 - m \max[\min(\frac{\alpha}{m-1}, C_{K,I}), C_{M,I}] + \gamma_K(I, \max[\min(\frac{\alpha}{m-1}, C_{K,I}), C_{M,I}]) \\
\geq \begin{cases} 
1 - mC_{K,I} + \gamma_K(I, C_{K,I}), & \text{if } \min(\frac{\alpha}{m-1}, C_{K,I}) \geq C_{M,I} \\
1 - mC_{M,I} + \gamma_K(I, C_{M,I}), & \text{if } \min(\frac{\alpha}{m-1}, C_{K,I}) \leq C_{M,I} 
\end{cases} \\
= \begin{cases} 
1 - m(\frac{\alpha + \gamma_K(I, C_{K,I})}{m}) + \gamma_K(I, C_{K,I}), & \text{if } \min(\frac{\alpha}{m-1}, C_{K,I}) \geq C_{M,I} \\
1 - m(\frac{\alpha + \gamma_M(I, C_{M,I})}{m}) + \gamma_K(I, C_{M,I}), & \text{if } \min(\frac{\alpha}{m-1}, C_{K,I}) \leq C_{M,I} 
\end{cases} \\
\quad \text{by (3.9)}; \\
\geq 1 - \alpha, \quad \text{noticing that } \gamma_M(I, C_{M,I}) \leq \gamma_K(I, C_{M,I}).
\]

\[ \]

In conclusion, we make a few remarks about our procedure III:

III.1 The procedure III is uniformly the most powerful among all the three proposed procedures, since it has the largest critical values of all, which can be simply seen from its
It also have the advantage over the CS procedure that it breaks the limit that the $p$th critical value is conservatively upper bounded by $\alpha/(n - p)$.

III.2 When the correlations among all test statistics are the same, the procedure III reduces to the procedure II.

III.3 (3.20) can be also defined using $C'_{H,p} = C_{H,L(p)}$ based on (3.7) instead of $C'_{K,p}$. The resulting procedure III still strongly controls the FWER at level $\alpha$.

To see a clear comparison among the three new procedures, we consider to test $n$ hypotheses $H_i, i = 1, 2, ..., n$ and assume that $T_i$ are the corresponding test statistics and $R_i$ the corresponding p-values obtained. We reject $H_i$ for large values of $T_i$. Let $T_i$ have the same t distribution with degrees of freedom $v$. Let $\rho_{ij}$ be the correlation between $T_i$ and $T_j$. Suppose the correlation matrix $C_n = \{\rho_{ij}\}$ is known. Let $R_{(i)}$ be the ordered P-values, $T_{(i)}$ the corresponding t-test statistics, and $H_{(i)}$ the corresponding hypotheses. Denote by $C_{(n)} = \{\rho_{(ij)}\}$ the corresponding correlation matrix of $T_{(i)}$. Specifically, let $n = 6, v = 20, \alpha = 0.05$, and

$$C_{(6)} = \begin{pmatrix}
1 & 0.5 & 0 & 0 & 0 & 0 \\
0.5 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0.8 & 0.8 & 0.8 \\
0 & 0 & 0.8 & 1 & 0.8 & 0.8 \\
0 & 0 & 0.8 & 0.8 & 1 & 0.8 \\
0 & 0 & 0.8 & 0.8 & 0.8 & 1
\end{pmatrix}.$$

Then the critical values of the procedures I, II, III, and the CS procedure can be calculated accordingly. Figure 3.1 shows all these critical values. It is shown that only the procedure III has the uniformly biggest critical values of all. Therefore, it is the most powerful procedure of all.

Moreover, to see how much the procedure III outperforms the CS and Holm’s procedures, we compare their critical values under different dependence structure of test statistics. Let $n = 8, v = 20, \alpha = 0.05$. Assume that $\rho_{ij} = \rho$ for $i \neq j$. Figure 3.1 shows the critical
Figure 3.1: Critical values comparison of FWER controlling procedures in a special case values of the three different methods for different correlation coefficients $\rho$.

3.3 Simulation Study

We studied the power of the procedure III by a simulation study. Consider to test $n$ hypotheses $H_1, \ldots, H_n$ simultaneously based on the observations $X_{ij}, j = 1, \ldots, N$ for $H_i, i = 1, \ldots, n$. Fixing $i$, suppose that $X_{i1}, \ldots, X_{iN}$ are independent and identically normally distributed with mean $\mu_i$ and variance $\sigma^2$. Let $T_i$ be the one-sample t-test statistics based on $X_{ij}, j = 1, \ldots, N$ corresponding to $H_i$. For the single hypothesis $H_i$: $\mu_i = 0$, we reject the hypothesis for large values of $T_i$. Denote by $\rho_{ij}$ be the correlation between $T_i$ and $T_j$. Assume that all test statistics are equally correlated, i.e., $\rho_{ij} = \rho$ for $i \neq j$. Let $R_i$ be the right-tailed p-values
Table 3.1: Critical values comparison under equal correlations. $v = 20, \alpha = 0.05$

<table>
<thead>
<tr>
<th>$p$</th>
<th>Type</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>Holm</td>
<td>0.00625</td>
<td>0.00714</td>
<td>0.00833</td>
<td>0.01017</td>
<td>0.0125</td>
<td>0.01667</td>
<td>0.025</td>
<td>0.05</td>
</tr>
<tr>
<td>0.1</td>
<td>CS</td>
<td>0.00634</td>
<td>0.00725</td>
<td>0.00847</td>
<td>0.01017</td>
<td>0.01273</td>
<td>0.017</td>
<td>0.02549</td>
<td>0.05</td>
</tr>
<tr>
<td>0.2</td>
<td>III</td>
<td>0.00644</td>
<td>0.00733</td>
<td>0.00856</td>
<td>0.01028</td>
<td>0.01288</td>
<td>0.01719</td>
<td>0.02574</td>
<td>0.05</td>
</tr>
<tr>
<td>0.3</td>
<td>CS</td>
<td>0.00662</td>
<td>0.00758</td>
<td>0.00886</td>
<td>0.01065</td>
<td>0.01333</td>
<td>0.01778</td>
<td>0.02648</td>
<td>0.05</td>
</tr>
<tr>
<td>0.4</td>
<td>III</td>
<td>0.00666</td>
<td>0.00764</td>
<td>0.00889</td>
<td>0.01072</td>
<td>0.01342</td>
<td>0.01769</td>
<td>0.02662</td>
<td>0.05</td>
</tr>
<tr>
<td>0.5</td>
<td>CS</td>
<td>0.00704</td>
<td>0.00806</td>
<td>0.00941</td>
<td>0.01131</td>
<td>0.01413</td>
<td>0.01876</td>
<td>0.02765</td>
<td>0.05</td>
</tr>
<tr>
<td>0.6</td>
<td>III</td>
<td>0.00719</td>
<td>0.00824</td>
<td>0.00963</td>
<td>0.01157</td>
<td>0.01446</td>
<td>0.01917</td>
<td>0.02808</td>
<td>0.05</td>
</tr>
<tr>
<td>0.7</td>
<td>CS</td>
<td>0.00714</td>
<td>0.00833</td>
<td>0.00983</td>
<td>0.01179</td>
<td>0.0147</td>
<td>0.01945</td>
<td>0.02845</td>
<td>0.05</td>
</tr>
<tr>
<td>0.8</td>
<td>III</td>
<td>0.00767</td>
<td>0.00878</td>
<td>0.01026</td>
<td>0.01231</td>
<td>0.01533</td>
<td>0.02021</td>
<td>0.02919</td>
<td>0.05</td>
</tr>
<tr>
<td>0.9</td>
<td>CS</td>
<td>0.00842</td>
<td>0.00963</td>
<td>0.01123</td>
<td>0.01343</td>
<td>0.01663</td>
<td>0.02171</td>
<td>0.03071</td>
<td>0.05</td>
</tr>
<tr>
<td>1.0</td>
<td>III</td>
<td>0.00975</td>
<td>0.01111</td>
<td>0.01289</td>
<td>0.01531</td>
<td>0.01876</td>
<td>0.02404</td>
<td>0.03291</td>
<td>0.05</td>
</tr>
</tbody>
</table>

for $H_i$ based on $T_i$ and $R_{(i)}$ the ordered p-values for the corresponding hypothesis $H_{(i)}$. In addition, assume that there are $m(\leq n)$ false null hypotheses with $\mu_i = \mu > 0$.

Our simulation study is based on the following specific values for all parameters above: $n = 10, N = 21, \mu = 1$, and $\sigma^2 = 2$. Thus, under the null hypothesis $H_i$: $\mu_i = 0$, $T_i$ has $t$ distribution with degrees of freedom $v = 20$; when $\mu_i = \mu$, $T_i$ has non-central $t$ distribution with degrees of freedom $v = 20$ and non-centrality parameter $\sqrt{N\mu/\sigma} = 3.24$. The joint distributions of p-values $R_i$ and $R_j$ under null hypotheses $H_k$: $\mu_k = 0, k = i, j$, are calculated based on the joint distributions of $T_i$ and $T_j$, which are bivariate $t$ distributions with 20 degrees of freedom and correlation $\rho$. Let $\alpha = 0.05$. The FWER and average power
(the proportion of the false null hypotheses which are correctly rejected) of this multiple testing problem using three step-down procedures: Holm, CS, and III, are calculated, for each \( \rho = 0, 0.2, 0.35, 0.5, 0.6, 0.75, 0.9, 0.99 \), and each possible \( m = 0, 1, 2, ..., 9 \) \((m = 10 \text{ case is trivial})\). Figure 3.2 and 3.3 give the simulation-based estimates for FWER and average power respectively. The simulation uses 30000 iterations for each pair of \( \rho \) and \( m \).

From Figure 3.2, we can see that the procedure III gives the biggest FWER of all for all correlations. As the correlation increases, the improvement on FWER compared to the CS procedure increases. The improvement increases more dramatically for \( \rho > 0.5 \). From Figure 3.3, we can see that the procedure III has the biggest average power of all for all correlations. Just like FWER, as the correlation increases, the improvement on average power compared to the CS procedure increases too. The improvement increases more dramatically for \( \rho > 0.5 \).

### 3.4 An Application

Suppose \( X_{ij}, j = 1, \cdots, J \) are independently and identically distributed as \( N(\mu_i, \sigma^2) \), and for \( i = 1, \cdots, I \), form independent samples. Then the sample means \( \overline{X}_i \) are independently distributed as \( N(\mu_i, \sigma^2/J) \), \( i = 1, \cdots, I \), and \( \frac{SSE}{\sigma^2} \) is distributed as \( \chi^2_v \) and also independent of \( \overline{X}_i \), where \( v = I(J-1) \) and

\[
SSE = \sum_{i=1}^{I} \sum_{j=1}^{J} (X_{ij} - \overline{X}_i)^2.
\]

Consider to test the \( n = \binom{I}{2} \) simultaneous hypotheses \( H^{(ij)} : \mu_i = \mu_j \) for \( 1 \leq i < j \leq I \). For each hypothesis \( H^{(ij)} \), the test statistics

\[
T_{ij} = \frac{\overline{X}_i - \overline{X}_j}{S \sqrt{\frac{2}{J}}}
\]
follows the t distribution with v degrees of freedom, where $S^2 = SSE/v$. Note that $Z_{ij} = \bar{X}_i - \bar{X}_j$, $i \neq j$, is also normally distributed. The correlation coefficients $\rho_{ij,kl}$ of $Z_{ij}$ and $Z_{kl}$ follow after simple calculation:

$$
\rho_{ij,kl} = \begin{cases} 
0, & \text{if } i \neq k, j \neq l \\
\frac{1}{2}, & \text{if } i = k, j \neq l \text{ or } i \neq k, j = l \\
-\frac{1}{2}, & \text{if } i = l, j \neq k \text{ or } i \neq l, j = k \\
1, & \text{if } i = k, j = l 
\end{cases}
$$

(3.21)

It follows that $T_{ij}$ and $T_{kl}$ have bivariate t distributions with parameters $v = I(J - 1)$ and $\rho$ given by (3.21).

Here we consider the example used by Seneta & Chen (2005). It concerns the effect of ethanol on sleep time in a sample of 20 rats. There were 4 treatments (concentrations of ethanol; $I = 4$), each applied to 5 rats ($J = 5$). Thus $v = 16$. The values of the means in ascending numerical order are $X_{4.1} = 32.760, X_{3.1} = 47.920, X_{2.1} = 61.540$, and $X_{1.1} = 79.280$, and the observed value of $S^2$ is 92.96250. We test $n = \binom{4}{2} = 6$ hypotheses $H^{(i,j)}: \mu_i = \mu_j$ for $1 \leq i < j \leq 4$ simultaneously at level $\alpha = 0.05$. We use $t_{16}$ to perform 2-tailed tests for all hypotheses $H^{(i,j)}, i < j$, and obtain the corresponding P-values $R_{ij}$, which are $R_{12} = 0.01025, R_{13} = 9.819 \times 10^{-5}, R_{14} = 0.1021 \times 10^{-5}, R_{23} = 0.04011, R_{24} = 23.12 \times 10^{-5}$, and $R_{34} = 0.02435$. Ranking hypotheses in increasing order of P-values we have $H_{(1)} = H^{(1,4)}, H_{(2)} = H^{(1,3)}, H_{(3)} = H^{(2,4)}, H_{(4)} = H^{(1,2)}, H_{(5)} = H^{(3,4)}$, and $H_{(6)} = H^{(2,3)}$. The bivariate joint distributions of P-values can be calculated by

$$
Pr(R_{ij} < 2c, R_{kl} < 2c) = Pr(|T_{ij}| > t_{16,c}, |T_{kl}| > t_{16,c})
$$

$$
= Pr(T_{ij} > t_{16,c}, T_{kl} > t_{16,c}) + Pr(T_{ij} > t_{16,c}, T_{kl} < -t_{16,c})
$$

$$
+ Pr(T_{ij} < -t_{16,c}, T_{kl} < -t_{16,c}) + Pr(T_{ij} < -t_{16,c}, T_{kl} > t_{16,c}),
$$

where $t_{16,c} = F_{16}^{-1}(1 - c)$, $F_{16}$ is the cumulative distribution function of $t_{16}$. Let $\rho_{ij}$ be
the correlation coefficient of the test statistics \( T(i) \) and \( T(j) \) corresponding to \( H(i) \) and \( H(j) \) respectively. Then the correlation matrix \( C_{(6)} = \{\rho_{ij}\} \) follows after (3.21)

\[
C_{(6)} = \begin{pmatrix}
1 & 0.5 & 0.5 & 0.5 & 0.5 & 0 \\
0.5 & 1 & 0 & 0.5 & -0.5 & 0.5 \\
0.5 & 0 & 1 & -0.5 & 0.5 & 0.5 \\
0.5 & 0.5 & -0.5 & 1 & 0 & -0.5 \\
0.5 & -0.5 & 0.5 & 0 & 1 & -0.5 \\
0 & 0.5 & 0.5 & -0.5 & -0.5 & 1
\end{pmatrix}.
\]

Next, for this multiple testing problem, we compare the results of three step-down procedures: Holm, CS, and the procedure III. All of the critical values of these procedures can be calculated based on the information above, see Table 3.2. The critical values (3.20) of the procedure III are calculated with both of (3.6) and (3.7), denoted by III\(_K\) and III\(_H\) respectively. The latter are calculated using \textit{minimum spanning tree} algorithm (see Appendix for R code). Since \( \gamma_M(S,c) \leq \gamma_K(S,c) \leq \gamma_H(S,c) \) for any fixed \( S \subset \{1, \cdots, 6\} \) and \( c \in (0,1) \), \( C'_{M,p} \leq C'_{K,p} \leq C'_{H,p} \) for \( p = 1, \cdots, 6 \) by their definitions. Moreover, based on \( C_{(6)} \) above, it follows after numerical computation that \( C'_{M,p} \leq \frac{\alpha}{6-p} \), so III\(_K\) and III\(_H\) are exactly \( \min(\frac{\alpha}{6-p}, C'_{K,p}) \) and \( \min(\frac{\alpha}{6-p}, C'_{H,p}) \) respectively, the corresponding critical values of the procedure I. Thus both of III\(_K\) and III\(_H\) are only determined by the algorithm (3.8).

Take the smallest critical value of III\(_K\) as an example: \( \alpha_1 = \min(\frac{0.05}{5}, C'_{K,1}) \), where \( C'_{K,1} = \lim_{d \to \infty} C_{K,L(1),d} \). By (3.8), \( C_{K,L(1),0} = \alpha = 0.05 \), \( C_{K,L(1),1} = 0.01640 \), \( C_{K,L(1),2} = 0.01025 \), \( C_{K,L(1),3} = 0.00940 \), \( C_{K,L(1),4} = 0.00929 \), \( C_{K,L(1),5} = 0.00928 \), \( C_{K,L(1),6} = 0.00928 \), \cdots. Thus \( \alpha_1 = \min(\frac{0.05}{5}, 0.00928) = 0.00928 \).

From Table 3.2, even though all hypotheses are rejected by all of these four specific procedures, the two procedure III (III\(_K\) and III\(_H\)) provide the most reassurance for rejections because III\(_K\) and III\(_H\) are larger than the critical values of both of the CS and Holm procedures for all \( p = 1, \cdots, 6 \). Moreover, III\(_H\) has slightly larger \( \alpha_1 \) than III\(_K\) while the
Table 3.2: Critical values comparison for ethanol effect study

<table>
<thead>
<tr>
<th>p</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holm</td>
<td>0.00833</td>
<td>0.01000</td>
<td>0.01250</td>
<td>0.01667</td>
<td>0.02500</td>
<td>0.05</td>
</tr>
<tr>
<td>CS</td>
<td>0.00916</td>
<td>0.01117</td>
<td>0.01394</td>
<td>0.01850</td>
<td>0.02728</td>
<td>0.05</td>
</tr>
<tr>
<td>III\textsubscript{K}</td>
<td>0.00928</td>
<td>0.01137</td>
<td>0.01419</td>
<td>0.01879</td>
<td>0.02758</td>
<td>0.05</td>
</tr>
<tr>
<td>III\textsubscript{H}</td>
<td>0.00948</td>
<td>0.01137</td>
<td>0.01419</td>
<td>0.01879</td>
<td>0.02758</td>
<td>0.05</td>
</tr>
</tbody>
</table>

other $\alpha_p$ are the same. The differences of the critical values between the CS procedure and the procedure III are slight due to the moderate correlations ($\pm 0.5$ at most) between test statistics, as Table 3.1 suggests.
Figure 3.2: Simulation-based estimates of FWER for three FWER controlling methods, Holm’s, CS, and the procedure III for different $\rho$ and the number of false null hypotheses from 0 to 9.
Figure 3.3: Simulation-based estimates of the average power (the proportion of the false null hypotheses which are correctly rejected) for three FWER controlling methods, Holm’s, CS, and the procedure III for different $\rho$ and the number of false null hypotheses from 1 to 9.
CHAPTER 4

NOVEL STEP-DOWN

PROCEDURES CONTROLLING

k-FWER

In this chapter, we extend our approaches discussed in Chapter 3 to investigate new step-down multiple testing procedures controlling k-FWER strongly. The k-FWER controlling was proposed by Lehmann and Romano (2005). And they proposed their k-FWER controlling step-down procedure (LR procedure). It will be shown that our procedure is uniformly more powerful than the LR procedure.

4.1 Generalized Kwerel’s Inequality

According to Galambos and Simonelli (1996), Margaritescu (1990) proposed and proved the following inequality, which can be viewed as the generalized version of Kwerel’s inequality (3.3). In this section, we will continue to use those notations in the section 3.1. The inequality gives bounds for \( P(v_n \geq m) \) in terms of \( S_1 \) and \( S_2 \) only. Here we only consider the upper bound which suffices to develop our step-down procedures.
For all \( n \geq m \geq 2 \),

\[
P(v_n \geq m) \leq \min \left[ \frac{m + n - 1}{mn} S_1 - \frac{2}{mn} S_2, \frac{2}{m(m - 1)} S_2 \right].
\] (4.1)

Obviously, when \( m = 1 \), the inequality reduces to Kwerel’s inequality (3.3).

Furthermore, the preceding bounds are the sharpest among all the bounds of the type

\[ \{aS_1 + bS_2\}, \]

where \( a \) and \( b \) are real numbers. Especially, the sharpest bounds when \( b = 0 \) is

\[
P(v_n \geq m) \leq \frac{1}{m} S_1. \quad (4.2)
\]

To prove the inequalities (4.1) and (4.2), we use the methods of indicators. Based on the values of \( S_k \), we evaluate, or approximate, the values

\[
P(m) = P(v_n \geq m) \text{ or } P[m] = P(v_n = m),
\]

for some values of \( m \). Also, \( P(m) = P[m] + P[m+1] + \ldots + P[n] \).

Let \( I_C \) be the indicator variable of the event \( C \), that is, \( I_C = 1 \) if \( C \) occurs; \( I_C = 0 \) if \( C \) fails. Clearly,

\[
I(A_{i_1}, A_{i_2}, \ldots, A_{i_k}) = I(A_{i_1})I(A_{i_2}) \cdots I(A_{i_k}),
\]

\[
I(C^c) = 1 - I(C),
\]

and

\[
E(I(C)) = P(C).
\]

Thus we have

\[ S_k = E(J_k), \]
where \( J_k = \sum I(A_{i_1})I(A_{i_2}) \cdots I(A_{i_k}) \). As a result, we give an evident identity

\[
S_k = E[I(v_n)] = \sum_{j=k}^{n} \binom{j}{k} P(v_n = j) = \sum_{j=k}^{n} \binom{j}{k} P[j], k \geq 0. \tag{4.3}
\]

For convenience, let us introduce the following notation. Let

\[
\Delta_{k,s} = \begin{cases} 
1, & \text{if } k \geq s; \\
0, & \text{otherwise};
\end{cases} \tag{4.4}
\]

**Proof of Generalized Kwerel's Inequality.** We consider real numbers \( a \) and \( b \) with the property that \( aS_1 + bS_2 \geq P(m) \), or equivalently,

\[
a \sum_{t=1}^{n} tP[t] + b \sum_{t=2}^{n} \binom{i}{2} P[i] \geq \sum_{t=m}^{n} P[t],
\]

which is obtained from (4.3). By comparing the corresponding terms of both sides, we have the equivalent property

\[
at + \frac{b}{2} t(t - 1) \geq \Delta_{t,m}, \text{ for } t = 0, 1, 2, \ldots, n. \tag{4.5}
\]

We denote the left-hand side of (4.5) by \( f(t) \). To show (4.2), let \( b = 0 \), then \( f(t) = at \). To see that the best choice for \( a \) is given by \( a = 1/m \), let \( t = m \), we have \( a \geq 1/m \). When \( a = 1/m \), if \( t < m \), the left-hand side is always greater than zero; if \( t > m \), the left-hand side is \( t/m \) which is always greater than 1. So \( a = 1/m \) is the best choice when \( b = 0 \).

Secondly, let \( a = 0 \), then \( f(t) = \frac{b}{2} t(t - 1) \). \( f(t) \) is monotonically increasing at \( t \). So the minimum value of \( b \) that satisfies (4.5) is the solution of \( \frac{b}{2} m(m - 1) = 1 \), i.e., \( b = \frac{2}{m(m-1)} \). So \( b = \frac{2}{m(m-1)} \) is the best choice when \( a = 0 \).

Thirdly, we consider the case \( a \neq 0 \) and \( b \neq 0 \). Since \( f(1) = a \), \( a \) must be positive. It also implies that it suffices to consider the case \( b < 0 \) (otherwise the bound obtained can be easily improved). By letting \( f(m) = f(n) = 1 \), \( a \) and \( b \) can be uniquely determined,
\[ a = (m + n - 1)/(mn) \text{ and } b = -2/(mn), \] and they satisfy
\[ \frac{m + n - 1}{mn} t - \frac{2}{mn} \frac{t(t - 1)}{2} \geq \Delta_{t,m}, t = 0, 1, 2, \ldots, n. \]

Furthermore, we will see the preceding upper bound is sharper than any other Bonferroni upper bound of form \( a^* S_1 + b^* S_2 \) with \( a^* > 0 \) and \( b^* < 0 \). Consider the polynomial \( g(t) = a^* t + b^* t(t - 1)/2 \) and observe that since \( g(0) = f(0) \), \( g(t) = f(t) \) for at most another \( t \) in \( \{1, 2, \ldots, n\} \). This observation together with the requirement that \( g(t) \geq \Delta_{t,m}, t = 0, 1, \ldots, n \), imply that \( g(t) \geq f(t) \) for \( t = 0, 1, \ldots, n \).

Finally, we will see that either \( \frac{m+n-1}{mn} S_1 - \frac{2}{mn} S_2 \) or \( \frac{2}{m(m-1)} S_2 \) must be no greater than \( \frac{1}{m} S_1 \).

Actually, it is easy to see that \( \frac{m+n-1}{mn} S_1 - \frac{2}{mn} S_2 \leq \frac{1}{m} S_1 \) is equivalent to \( \frac{1}{m} S_1 \leq \frac{2}{m(m-1)} S_2 \). It is still true if the direction of the inequalities is conversed. It means \( \frac{1}{m} S_1 \) cannot be less than the other two expressions at the same time. Thus the proof is complete.

\[ \square \]

### 4.2 Novel Step-Down Procedures Controlling k-FWER

#### 4.2.1 Procedure IV

Consider the multiple testing problem where there are \( n \) hypotheses \( H_1, \ldots, H_n \) and corresponding p-values \( R_1, \ldots, R_n \). Assume that the test statistics \( X_1, \ldots, X_n \) are from continuous probability distributions. Let \( R_{(1)}, \ldots, R_{(n)} \) be the ordered p-values and \( H_{(1)}, \ldots, H_{(n)} \) the corresponding hypotheses. Write \( R_{(i)} = R_{t_i} \), where \( t_i \) is a random variable taking values from \( \{1, 2, \ldots, n\} \). Define the index set \( L(p) = \{t_p, t_{p+1}, \ldots, t_n\} \), which is a random set. Furthermore, assume that the bivariate joint distributions of any two test statistics are continuous and known.

At the beginning, we define three kinds of degree-2 terms based on the inequality (4.1),
using bivariate joint distributions between test statistics:

\[
\gamma_{M_1,k}(S,c) = \frac{m(m-1)}{m+k-1} \min_{(i,j) \subset S} P(R_i \leq c, R_j \leq c|H_s, s \in S, true) \\
+ \frac{(2m+k-2)(k-1)}{m+k-1} \min_{(i,j) \subset \{1,\ldots,n\}} P(R_i \leq c, R_j \leq c|H_s, s = 1, 2, \ldots, n, true),
\]

(4.6)

\[
\gamma_{M_2,k}(S,c) = \frac{m(m-1)}{k(k-1)} \max_{(i,j) \subset S} P(R_i \leq c, R_j \leq c|H_s, s \in S, true) \\
+ \frac{2m+k-2}{k} \max_{(i,j) \subset \{1,\ldots,n\}} P(R_i \leq c, R_j \leq c|H_s, s = 1, 2, \ldots, n, true),
\]

(4.7)

\[
\gamma_G(S,c) = \frac{2}{m} \sum_{(i,j) \subset S, i < j} P(R_i \leq c, R_j \leq c|H_s, s \in S, true),
\]

(4.8)

where \(S \subset \{1,2,\ldots,n\}\) with size \(|S| = m \geq 1, c \in (0,1)\), and some integer \(1 \leq k \leq n\). In addition, define \(\gamma_G(S,c) = 0\) when \(m = 1\). Moreover, assume that \(\min_{(i,j) \subset S} P(R_i \leq c, R_j \leq c|H_s, s \in S, true) = P(R_{i_0} \leq c, R_{j_0} \leq c|H_{i_0}, H_{j_0}, true)\) for some \((i_0, j_0) \subset S\) and any \(c \in (0,1)\).

In other words, the choice of \((i_0, j_0)\) does not depend on the value of \(c\) for any subset \(S\).

Note that when \(k = 1\), \(\gamma_{M_1,k}(S,c)\) exactly reduces to \(\gamma_M(S,c)\) in (3.15).

In order to define our new k-FWER controlling procedure, we apply the following iterative algorithm to generate a sequence of values using \(\gamma_{M_1,k}(S,c)\), which is similar to the algorithm (3.8):

\[
C_{M_1,k,S,0} = \alpha \\
C_{M_1,k,S,1} = \frac{k\alpha + \gamma_{M_1,k}(S,C_{M_1,k,S,0})}{m + 2k - 2} \\
C_{M_1,k,S,2} = \frac{k\alpha + \gamma_{M_1,k}(S,C_{M_1,k,S,1})}{m + 2k - 2} \\
\vdots \\
C_{M_1,k,S,d} = \frac{k\alpha + \gamma_{M_1,k}(S,C_{M_1,k,S,d-1})}{m + 2k - 2} \\
\vdots
\]

(4.9)
Proposition 4.2.1. The sequence \( \{C_{M_1,k,S,d}\}_{d=0}^{\infty} \) defined in (4.9) using (4.6) for some \( S \subset \{1, 2, \ldots, n\} \) with \( |S| = m \) and some integer \( 1 \leq k \leq m \) is non-increasing at \( d \) and has lower and upper bounds \( \frac{k\alpha + \gamma_{M_1,k}(S, \frac{k\alpha}{m+2k-2})}{m+2k-2} \) and \( \alpha \) respectively. Therefore, these sequences converge to some points, denoted by \( C_{M_1,k,S} \), in \( \left[ \frac{k\alpha + \gamma_{M_1,k}(S, \frac{k\alpha}{m+2k-2})}{m+2k-2}, \alpha \right] \).

The proof of Proposition 4.2.1 is similar to that of Proposition 3.2.1 and thus will not be given here. Actually, taking limits of both sides of (4.4) implies that \( C_{M_1,k,S} \) is one solution of the equation

\[
C = \frac{k\alpha + \gamma_{M_1,k}(S, C)}{m + 2k - 2} \quad \text{or} \quad (m + 2k - 2)C - \gamma_{M_1,k}(S, C) = k\alpha \quad (4.10)
\]

with the variable \( C \) in the interval \( \left[ \frac{k\alpha + \gamma_{M_1,k}(S, \frac{k\alpha}{m+2k-2})}{m+2k-2}, \alpha \right] \).

Secondly, define \( C_{M_2,k,S} \) as follows

\[
C_{M_2,k,S} = \max[C \in (0, 1) : \gamma_{M_2,k}(S, C) = \alpha] \quad (4.11)
\]

for \( k \geq 2 \) and define \( C_{M_2,k,S} = 0 \) for \( k = 1 \). (4.11) is well defined because of the fact that \( \gamma_{M_2,k}(S, 0) = 0 \leq \alpha \) and \( \gamma_{M_2,k}(S, 1) = \frac{(m+k-1)(m+k-2)}{k(k-1)} \geq 1 \geq \alpha \), noting that \( \gamma_{M_2,k}(S, C) \) is continuous for \( C \in [0, 1] \).

Consider to test multiple hypotheses \( H_1, \ldots, H_n \) with corresponding p-values \( R_1, \ldots, R_n \). Our k-FWER controlling procedure IV is one Holm-type step-down procedure using the critical values defined as

\[
\alpha_p = \begin{cases} 
\max[C_{M_1,k,k}', C_{M_2,k,k}', \frac{k\alpha}{n}], & p \leq k, \\
\max[C_{M_1,k,p}', C_{M_2,k,p}', \frac{k\alpha}{n+k-p}], & p > k,
\end{cases} \quad (4.12)
\]

,where \( C_{M_1,k,p}' = C_{M_1,k,L(p)} \) and \( C_{M_2,k,p}' = C_{M_2,k,L(p)} \), \( p = k, 2, \ldots, n \). Note that \( C_{M_1,k,p}' \) is in
\[ \left[ \frac{k\alpha + \gamma_{M_1,k}(L(p), \frac{k\alpha}{n + 2k - p - 1})}{n + 2k - p - 1}, \alpha \right], \] and satisfies the equation

\[ (n + 2k - p - 1)C'_{M_1,k,p} - \gamma_{M_1,k}(S, C'_{M_1,k,p}) = k\alpha \]  \hspace{1cm} (4.13)

for all \( p \geq k \) by (4.10) and the fact that \(|L(p)| = n - p + 1\).

Like the LR procedure, the proposed procedure also controls the k-FWER in the strong sense. To show this, let any \( I \subset E \) be the indexes set of all true hypotheses with the size \(|I| = m \leq n\). Order the p-values corresponding to the \( m \) true null hypotheses; call them \( Q_1 \leq ... \leq Q_m \). Let \( j \) be the smallest (random) index satisfying

\[ R(j) = Q(k). \]  \hspace{1cm} (4.14)

To show the k-FWER controlling, we need a Lemma similar to the Lemma 3.2.8.

**Lemma 4.2.2.** For any subsets \( S_1, S_2 \subset \{1, ..., n\} \) with \(|S_1| = m_1\) and \(|S_2| = m_2\) that satisfying \( S_1 \subset S_2 \), and some \( k \geq 1 \),

\[ C_{M_1,k,S_1} \geq C_{M_1,k,S_2}, \]

and

\[ C_{M_2,k,S_1} \geq C_{M_2,k,S_2}, \]

where \( C_{M_1,k,S_1} \) and \( C_{M_1,k,S_2} \) are as defined in the Proposition 4.2.1, \( C_{M_2,k,S_1} \) and \( C_{M_2,k,S_2} \) as defined by (4.11).

**Proof.** To see \( C_{M_1,k,S_1} \geq C_{M_1,k,S_2} \), it suffices to show the inequalities \( C_{M,k,S_1,d} \geq C_{M,k,S_2,d} \) for all nonnegative integers \( d \), where \( C_{M,k,S_1,d} \) are obtained through the algorithm (4.5), since the lemma can be shown simply by taking the limits of both sides of the inequalities. The inequalities are shown by induction as follows.

When \( d = 0 \), \( C_{M,k,S_1,0} = C_{M,k,S_2,0} = \alpha \).
Then assume that, for any $d \in \{1, 2, ..., n\}$, $C_{M,k,S_1,d-1} \geq C_{M,k,S_2,d-1}$. Then it follows after a quite similar discussion to that in the proof of the Lemma 3.2.4 that

$$\frac{k\alpha + \gamma_{M,k}(S_1, C_{M,k,S_1,d-1})}{m_1 + 2k - 2} \geq \frac{k\alpha + \gamma_{M,k}(S_2, C_{M,k,S_2,d-1})}{m_2 + 2k - 2},$$

i.e., $C_{M,k,S_1,d} \geq C_{M,k,S_2,d}$. The detailed will not be here again.

To see $C_{M_2,k,S_1} \geq C_{M_2,k,S_2}$, we first need to show that $\gamma_{M_2,k}(S_1, c) \leq \gamma_{M_2,k}(S_2, c)$ for any $c \in (0, 1)$. Actually, it is straightforward that

$$\frac{m_1(m_1 - 1)}{k(k - 1)} \max_{(i,j) \subseteq S_1} P(R_i \leq c, R_j \leq c|H_s, s \in S_1, true)$$

$$+ \frac{2m_1 + k - 2}{k} \max_{(i,j) \subseteq \{1,...,n\}} P(R_i \leq c, R_j \leq c|H_s, s = 1, 2, ..., n, true)$$

$$\leq \frac{m_2(m_2 - 1)}{k(k - 1)} \max_{(i,j) \subseteq S_2} P(R_i \leq c, R_j \leq c|H_s, s \in S_1, true)$$

$$+ \frac{2m_2 + k - 2}{k} \max_{(i,j) \subseteq \{1,...,n\}} P(R_i \leq c, R_j \leq c|H_s, s = 1, 2, ..., n, true).$$

Denote $f_1(c) = \gamma_{M_2,k}(S_1, c)$ and $f_2(c) = \gamma_{M_2,k}(S_2, c)$. Considering that $f_1(c) \leq f_2(c)$ and that $f_1, f_2$ are monotonically non-decreasing at $c$, it is true that the inverse relations of $f_1, f_2$ satisfy $f_1^{-1}(c) \geq f_2^{-1}(c)$. Therefore, $C_{M_2,k,S_1} = f_1^{-1}(\alpha) \geq f_2^{-1}(\alpha) = C_{M_2,k,S_2}$. Thus the proof is complete.

**Remark 4.2.3.** Note that Lemma 4.2.2 implies that the critical values $\alpha_p$ defined in (4.12) are monotonically increasing at $p$.

**Theorem 4.2.4 (Procedure IV).** The step-down multiple testing procedure taking $\alpha_p$ defined in (4.12) as critical values strongly controls the $k$-FWER at predetermined significance level $\alpha$, that is,

$$P(\text{rejecting by the procedure at least } k H_s, s \in I|H_s, s \in I, true) \leq \alpha,$$

where $I$ is the index set of all $m$ true hypotheses.
Proof. Assume that \( m \geq k \), otherwise nothing to prove. If \( k = 1 \), the critical values (4.12) reduces to (3.16), and thus the procedure IV reduces to the procedure II. So the proof is the same as that of Theorem 3.2.6. Therefore, we only need to consider \( k \geq 2 \). Let \( I_k \) be the index set of the smallest \( k - 1 \) ones of all p-values corresponding to the \( m \) true null hypotheses. Thus \( I_k \) is a random set and \( |I_k| = k - 1 \). Conditional on the event that \( \{H_s, s \in I, true\} \), the k-familywise error rate is

\[
P(\text{rejecting at least } k \text{ } H_s, s \in I) = P(R_j \leq \alpha, \text{ all } i \leq j),
\]

where \( j \) is as defined in (4.14); thus \( j \geq k \);

\[
\leq P(R_j \leq \alpha_j)
\]

\[
=P(Q(k) \leq \max[C_{M_1,k,j}^{'}, C_{M_2,k,j}^{'}, \frac{k\alpha}{n+k-j}])
\]

\[
=P(Q(k) \leq \max[C_{M_1,k,L(j)}, C_{M_2,k,L(j)}, \frac{k\alpha}{n+k-j}])
\]

\[
\leq P(Q(k) \leq \max[C_{M_1,k,I-I_k}, C_{M_2,k,I-I_k}, \frac{k\alpha}{m}]),
\]

where \( C_{M_i,k,L(j)} \leq C_{M_i,k,I-I_k}, i = 1, 2 \), by \( I-I_k \subset L(p) \) and the Lemma 4.2.2,

\[
\text{and } \frac{k\alpha}{n+k-j} \leq \frac{k\alpha}{m} \text{ by } n-j+k \geq m;
\]

\[
= P(\text{The number of } i \in I \text{ such that } R_i \leq \max[C_{M_1,k,I-I_k}, C_{M_2,k,I-I_k}, \frac{k\alpha}{m}] \text{ is at least } k.)
\]

\[
\begin{cases}
  P(\text{The number of } i \in I \text{ such that } R_i \leq C_{M_1,k,I-I_k} \text{ is at least } k.), & \text{if } C_{M_1,k,I-I_k} \text{ is the maximum;} \\
  P(\text{The number of } i \in I \text{ such that } R_i \leq C_{M_2,k,I-I_k} \text{ is at least } k.), & \text{if } C_{M_2,k,I-I_k} \text{ is the maximum;} \\
  P(\text{The number of } i \in I \text{ such that } R_i \leq \frac{k\alpha}{m} \text{ is at least } k.), & \text{if } \frac{k\alpha}{m} \text{ is the maximum;}
\end{cases}
\]
\[
\begin{aligned}
\frac{k+m-1}{km} \sum_{i \in I} P(R_i \leq C_{M_1,k,I-I_k}) - \frac{1}{k} \gamma_G(I, C_{M_1,k,I-I_k}), & \quad \text{if } C_{M_1,k,I-I_k} \text{ is the maximum;} \\
\frac{m}{k(k-1)} \gamma_G(I, C_{M_2,k,I-I_k}), & \quad \text{if } C_{M_2,k,I-I_k} \text{ is the maximum;} \\
\frac{1}{k} \sum_{i \in I} P(R_i \leq \frac{ka}{m}), & \quad \text{if } \frac{ka}{m} \text{ is the maximum;} \\
\gamma_{M_2,k}(I, C_{M_2,k,I-I_k}), & \quad \text{if } C_{M_2,k,I-I_k} \text{ is the maximum;} \\
\frac{1}{k}(m \frac{ka}{m}), & \quad \text{if } \frac{ka}{m} \text{ is the maximum;}
\end{aligned}
\]

(4.15)

\[= \alpha,\]

where the first equality is by (4.10) and \(|I - I_k| = m - k + 1\); the second equality is by (4.11).

Lastly, to complete the proof, we just need to show the three inequalities of (4.15). For first inequality of (4.15), it suffices to show \(\gamma_G(I, c) \geq \gamma_{M_1,k}(I - I_k, c)\) for any \(c \in (0, 1)\), i.e.,

\[
\frac{2}{m} \sum_{(i,j) \in I, i < j} P(R_i \leq c, R_j \leq c|H_s, s \in I, true) \\
\geq \frac{(m - k + 1)((m - k + 1) - 1)}{(m - k + 1) + k - 1} \min_{(i,j) \in I - I_k} P(R_i \leq c, R_j \leq c|H_s, s \in I - I_k, true) \\
+ \frac{(2(m - k + 1) + k - 2)(k - 1)}{(m - k + 1) + k - 1} \min_{(i,j) \in \{1, \ldots, n\}} P(R_i \leq c, R_j \leq c|H_s, s = 1, 2, \ldots, n, true).
\]

(4.16)

Actually, the left hand side of (4.16) equals

\[
\frac{1}{m} \sum_{i \in I} \sum_{j \in I, j \neq i} P(R_i \leq c, R_j \leq c|H_s, s \in I, true) \\
= \frac{1}{m} \sum_{i \in I - I_k} \sum_{j \in I - I_k, j \neq i} P(R_i \leq c, R_j \leq c|H_s, s \in I - I_k, true)
\]
\[ + \frac{1}{m} \sum_{i \in I-I_k} \sum_{j \in I_k} P(R_i \leq c, R_j \leq c | H_s, s \in I, \text{true}) \]
\[ + \frac{1}{m} \sum_{i \in I_k} \sum_{j \in I-I_k} P(R_i \leq c, R_j \leq c | H_s, s \in I, \text{true}) \]
\[ + \frac{1}{m} \sum_{i \in I_k} \sum_{j \in I-I_k, j \neq i} P(R_i \leq c, R_j \leq c | H_s, s \in I_k, \text{true}) \]
\[ \geq \frac{(m-k+1)(m-k)}{m} \min_{(i,j) \in I-I_k} P(R_i \leq c, R_j \leq c | H_s, s \in I - I_k, \text{true}) \]
\[ + \frac{2(m-k+1)(k-1)}{m} \min_{(i,j) \in (I-I_k) \times I_k} P(R_i \leq c, R_j \leq c | H_s, s \in I, \text{true}) \]
\[ + \frac{(k-1)(k-2)}{m} \min_{(i,j) \in I_k} P(R_i \leq c, R_j \leq c | H_s, s \in I_k, \text{true}) \]
\[ \geq \frac{(m-k+1)(m-k)}{m} \min_{(i,j) \in I-I_k} P(R_i \leq c, R_j \leq c | H_s, s \in I - I_k, \text{true}) \]
\[ + \frac{2m-k)(k-1)}{m} \min_{(i,j) \in I} P(R_i \leq c, R_j \leq c | H_s, s \in I, \text{true}) \]
\[ \geq \frac{(m-k+1)(m-k)}{m} \min_{(i,j) \in I-I_k} P(R_i \leq c, R_j \leq c | H_s, s \in I - I_k, \text{true}) \]
\[ + \frac{2m-k)(k-1)}{m} \min_{(i,j) \in \{1,\ldots,n\}} P(R_i \leq c, R_j \leq c | H_s, s \in \{1,\ldots,n\}, \text{true}) \]
\[ = \text{the right hand side of (4.16)}. \]

For the second inequality of (4.15), it suffices to show \( \frac{m}{k(k-1)} \gamma_G(I, c) \leq \gamma_{M_{2,k}}(I-I_k, c) \) for any \( c \in (0, 1) \). Actually,

\[ \frac{m}{k(k-1)} \gamma_G(I, c) \]
\[ = \frac{m}{k(k-1)} \frac{2}{m} \sum_{(i,j) \in I, i < j} P(R_i \leq c, R_j \leq c | H_s, s \in I, \text{true}), \]
\[ = \frac{1}{k(k-1)} \sum_{i \in I-I_k} \sum_{j \in I-I_k, j \neq i} P(R_i \leq c, R_j \leq c | H_s, s \in I - I_k, \text{true}) \]
\[ + \frac{1}{k(k-1)} \sum_{i \in I-I_k} \sum_{j \in I_k} P(R_i \leq c, R_j \leq c | H_s, s \in I, \text{true}) \]
\[ + \frac{1}{k(k-1)} \sum_{i \in I_k} \sum_{j \in I-I_k} P(R_i \leq c, R_j \leq c | H_s, s \in I, \text{true}) \]
Finally, the third inequality of (4.15) follows after (2.1). Thus the proof is complete. □

Next, we will compare the procedure IV to the LR procedure. Since the critical values of the LR procedure are included in the maximum function in the critical values (4.12) of the procedure IV, it implies that the procedure IV is at least as powerful as the LR procedure, if not more powerful. The following theorem shows that the procedure IV could be uniformly more powerful than the LR procedure under some situation that all test statistics are identically distributed and equally correlated.

**Theorem 4.2.5.** Assume that all test statistics are identically distributed and mutually equally correlated and all tests for hypotheses $H_i, i = 1, \ldots, n$ are of the same type (right-tailed, left-tailed, or two-tailed). Let $R_i$ be the P-values for $H_i$ and $b_{ij}(c) = P(R_i < c, R_j < c|H_i, H_j$ are true, $i \neq j)$. Assume that the derivatives $b'_{ij}(c)$ exist and satisfy $b'_{ij}(c) \leq 1$ for $c \in [0, \alpha]$. then $\max[C_{M_1,k,p}', C_{M_2,k,p}']$ in (4.12) are no less than $\frac{k\alpha}{n+k-p}$, the critical values of the LR procedure.
Proof. Under the assumption of the Theorem, for any $c \in [0,1]$, $P(R_i \leq c, R_j \leq c) = P(R_s \leq c, R_t \leq c)$ for any $i \neq j, s \neq t \in \{1,2,...,n\}$. So, after a simple algebraic manipulation, it follows that

$$\gamma_{M_1,k}(L(p), c) = \gamma_G(E', c),$$

and

$$\gamma_{M_2,k}(L(p), c) = \frac{n-p+k}{k(k-1)} \gamma_G(E', c),$$

for any $p = k, ..., n$, where $E' = \{1,2,...,n-j+k\}$.

On the other hand, define $F(C) = (n+2k-p-1)C/k - \gamma_{M_1,k}(L(p), C)/k$, $G(C) = \gamma_{M_2,k}(L(p), C)$, and $H(C) = (n-p+k)C/k$ for $C \in [0,1]$. Also, we know that, by (4.13) and (4.11), $C'_{M_1,k,p}$, $C'_{M_2,k,p}$, and $\frac{k\alpha}{n+k-p}$ are the solutions of $F(C) = \alpha, G(C) = \alpha$, and $H(C) = \alpha$ respectively. We write $C'_{M_1,k,p} = F^{-1}(\alpha), C'_{M_2,k,p} = G^{-1}(\alpha), \frac{k\alpha}{n+k-p} = H^{-1}(\alpha)$. Since $\frac{k\alpha}{n+k-p} \leq \alpha$, it suffices to assume that both $C'_{M_1,k,p}$ and $C'_{M_2,k,p}$ are no greater than $\alpha$, otherwise the proof is done. It is obvious that $G(C)$ and $H(C)$ are monotonically non-decreasing. Due to the equal correlation of P-values,

$$F(C) = (n+2k-p-1)C/k - (n+k-p-1)b_{12}(C)/k$$

$$= C + (n+k-p-1)(C - b_{12}(C))/k.$$

Thus, $F'(C) = (n+k-p-1)(1-b'_{12}(C))/k \geq 0$ on $C \in [0,\alpha]$ due to $b'_{12}(C) \leq 1$. Therefore, $F(C)$ is non-decreasing on $C \in [0,\alpha]$.

Next, we will see that either $F(C)$ or $G(C)$ is no greater than $H(C)$. Without the loss of generality, assume that $F(C) \geq H(C)$, i.e., $(k-1)C/k - \gamma_G(E', c) \geq 0$. Then multiplying both sides by $(n-p+k)/(k(k-1))$, we obtain $H(C) \geq G(C)$. The other case is for the same rationale. Considering that $F, G, H$ are non-decreasing on $C \in [0,\alpha]$, therefore, either $C'_{M_1,k,p}$ or $C'_{M_2,k,p}$ is no less than $\frac{k\alpha}{n-p+k}$. Thus the proof is complete. 

The following Proposition will show that if the test statistics follow the standard normal
distribution, then the condition $b'_{ij}(c) \leq 1$ in the Theorem above holds.

**Proposition 4.2.6.** Assume the test statistics $(X,Y)$ has the standard bivariate normal distribution with the correlation coefficient $\rho$. Assume that the associated two hypothesis tests are one-sided and the null hypotheses are rejected for small values of $X$ and $Y$ respectively, then $b'_{xy}(c) \leq 1$ for $c \leq 0.5$, where $b'_{xy}(c)$ is as defined in Theorem 4.2.5.

**Proof.**

$$b_{xy}(c) = P(X < z_c, Y < z_c) = \int_{-\infty}^{z_c} \int_{-\infty}^{z_c} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\{-\frac{x^2 + y^2 - 2\rho xy}{2(1-\rho^2)}\} \, dx \, dy,$$

where $z_c = \Phi^{-1}(c)$, $\Phi(x)$ is the standard normal distribution. Then,

$$b'_{xy}(c) = \frac{d}{dz_c} P(X < z_c, Y < z_c) \times \frac{d}{dc} \Phi^{-1}(c)$$

$$= 2 \int_{-\infty}^{z_c} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\{-\frac{x^2 + z_c^2 - 2\rho x z_c}{2(1-\rho^2)}\} \, dx \times \frac{1}{\phi(z_c)}$$

$$= 2 \int_{-\infty}^{z_c} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{z_c^2}{2}\} \frac{1}{\sqrt{2\pi\sqrt{1-\rho^2}}} \exp\{-\frac{(x - \rho z_c)^2}{2(1-\rho^2)}\} \, dx \times \frac{1}{\phi(z_c)}$$

$$= 2\phi(z_c) \times \Phi\left(\frac{z_c - \rho z_c}{\sqrt{1-\rho^2}}\right) \times \frac{1}{\phi(z_c)}$$

$$\leq 1,$$

for $z_c \leq 0$, i.e. $c \leq 0.5$, noting that $\Phi\left(\frac{z_c - \rho z_c}{\sqrt{1-\rho^2}}\right) \leq \frac{1}{2}$. Thus the proof is complete. \qed

Next, we make a few remarks about our procedure IV:

**IV.1** Evidently, one can always reject the hypotheses corresponding to the smallest $k - 1$ p-values without violating control of the $k$-FWER. However, it seems counterintuitive to consider a step-down procedure whose corresponding $\alpha_p$ are not monotone nondecreasing. In addition, automatic rejection of $k - 1$ hypotheses, regardless of the data, appears at the very least a little too optimistic. To ensure monotonicity, our step-down procedure uses $\alpha_p = \alpha_k$ for $p < k$. Even if we were to adopt the more optimistic strat-
egy of always rejecting the hypotheses corresponding to the first $k - 1$ hypotheses, we could still only reject $k$ or more hypotheses if $R(k) \leq \alpha_k$.

IV.2 The procedure IV does not always have uniformly larger critical values than the LR procedure. It is mainly because of the conservative use of Min and Max functions in (4.6) and (4.7). It is simple to see that the conservativeness can be offset when all test statistics are equally correlated. Simulation analysis in the later section will give comparisons between the two procedures in terms of critical values and powers.

IV.3 When $k = 1$, apparently, the procedure IV exactly reduces to the procedure II for FWER-controlling problems. It can be seen from not only the definitions of their critical values but also the proof the Theorem 4.2.4. Thus, the procedure IV can be seen as the generalization of the procedure II.

A numerical analysis is then given to compare the critical values between the LR procedure and the procedure IV controlling k-FWER. We consider to test $n$ one-sided hypotheses $H_i, i = 1, 2, ..., n$ based on $n$ z-test statistics $Z_i$ and let $R_i$ be the corresponding p-values. We reject $H_i$ for large values of $Z_i$. Let $\rho_{ij}$ be the correlation between $Z_i$ and $Z_j$. Assume $\rho_{ij} = \rho$ for $i \neq j$. Let $n = 20$ and $\alpha = 0.05$. Figure 4.1 shows the critical values of the LR procedure and the procedure IV controlling 2-FWER at level 0.05 for different correlation coefficients $\rho$. From Figure 4.1, we can see that the procedure IV has uniformly bigger critical values than the LR procedure. Moreover, the differences are more significant for both small (say close to 0) and large (close to 1) p-values than for medium (around 0.5) p-values. The fact is quite different from that for FWER controlling procedure using bivariate joint distributions between test statistics, such as the procedure III. From Table 3.1, we can see that for $\rho = 0$, the procedure III almost has no improvement on Holm’s procedure. But for the 2-FWER controlling procedure IV, for $\rho = 0$, the improvement of the critical values over the LR procedure is dramatic! The improvement reduces to its minimum when $\rho = 0.5$.

This fact also can been seen better from Figure 4.2, which shows the ratios of the critical
values of the procedure IV to those of the LR procedure. The ratios at both small and large p-values are usually bigger than those at medium p-values, especially for critical values with small indexes. The last but not least, from Figure 4.1 we can see that the critical values for k-FWER for \( k = 2 \) can be much bigger than the predetermined level \( \alpha \), while the critical values of the LR procedure are always no larger than \( \alpha \). Especially, for \( \rho = 0 \), the ratios are even bigger than 3 for all indexes. It implies a dramatic improvement. This discovery motivates the investigation of finding a model to exactly explain the k-FWER controlling step-down procedure for independence case. In the following section, we can see
that Binomial model is exactly there for the purpose.

4.2.2 Binomial Model Under Independence

The motivation to consider the independence case is that from the simulation analysis for the procedure IV (Section 4.3), we see that the power improvement on the LR procedure is dramatic for $k > 1$ when the dependence between test statistics is week (Figure 4.5). The most dramatic case occurs when all test statistics are uncorrelated. Furthermore, the critical values of the procedure IV can be much bigger than the predetermined level $\alpha$, while the
critical values of the LR procedure are at most $\alpha$ (Figure 4.1). This phenomenon is very different from that when $k = 1$. When $k = 1$, the improvement of the procedure III on Holm’s procedure decreases as the correlation decreases. Especially, when all test statistics are uncorrelated, there is almost no improvement at all (see Table 3.1). This motivates the investigation of a model to explain the phenomenon and as well as some new potentially better step-down procedures under independence.

Actually, the step-down procedures controlling k-FWER are usually the step-wise versions of some single-step procedures, just like what Holm’s procedure is to the Bonferroni’s procedure for $k = 1$. Therefore, we start by investigating single-step procedures. Considering testing the hypotheses $H_1, \ldots, H_n$ based on the corresponding p-values $R_1, R_2, \ldots, R_n$. Let $R_{(1)}, R_{(2)}, \ldots, R_{(n)}$ be the ordered p-values and denote by $H_{(i)}$ the hypothesis corresponding to $R_{(i)}$. Let $I \subset \{1, 2, \ldots, n\}$ be the index set of all true hypotheses, define the following single-step procedure controlling k-FWER: reject all $H_{(i)}$ with $R_{(i)} < c$, where $c$ is the critical value satisfies

$$\text{k-FWER} = P(R_i < c \text{ for at least } k \ i \in I) \leq \alpha. \quad (4.17)$$

Let $X_i = 1$ if $\{R_{(i)} < c\}$ and $X_i = 0$ otherwise, then by (2.1) the probability of the event $X_i = 1$ occur is no greater than $c$. For our goal, it suffices to assume that $P(X_i = 1) = c$. The assumption of the independence of $R_i$ implies that $X_i$ are independent, identically Bernoulli distributed with success probability $c$. Then, $\sum_{i \in I} X_i$ has Binomial distribution $B(|I|, c)$. Thus

$$\text{k-FWER} = P(\sum_{i \in I} X_i \geq k) = \sum_{i=k}^{\left\lfloor \frac{|I|}{i} \right\rfloor} \binom{|I|}{i} c^i (1 - c)^{|I| - i} \leq \sum_{i=k}^{n} \binom{n}{i} c^i (1 - c)^{n-i}. \quad (4.18)$$

The last inequality follows after the fact that the probability that at least $k$ successes occur increases as the number of trails increases. Also it can be seen that the right-hand side of (4.18) is increasing at $c$ because the probability that at least $k$ successes occur increases as
the success probability $c$ increases. Thus the best (largest) solution of $c$ satisfying (4.17) with $|I|$ unknown is the solution of

$$\sum_{i=k}^{n} \binom{n}{i} c^i (1 - c)^{n-i} = \alpha.$$  \hfill (4.19)

It follows after (4.17) – (4.19) that the single-step procedure using critical values $c$ defined in (4.19) controls the k-FWER strongly.

Actually, a step-down procedure can be defined based on the single-step procedure above such that the k-FWER is controlled strongly. Define the sequence of critical values $\alpha_p$ for $p \geq k$ satisfying

$$\sum_{i=k}^{n-p+k} \binom{n-p+k}{i} \alpha_p^i (1 - \alpha_p)^{n-p+k-i} = \alpha.$$  \hfill (4.20)

And let $\alpha_p = \alpha_k$ for $p < k$. Define the following step-down procedure: rejects all $H(i), i < j$, where

$$j = \min\{l : P(l) > \alpha_l, l \in \{k, ..., n\}\},$$

if $j$ exits.

The proof that the procedure strongly controls k-FWER for independent test statistics is very similar to that of Theorem 4.2.5, and thus will not be given here. In the numerical analysis, a comparison of this procedure to the procedure IV will be investigated. It will not be surprising to see that this Binomial-based procedure has even larger critical values than the procedure IV under independence, since it uses the Binomial model to describe the k-FWER exactly, instead of probability inequalities.

Next, we compare the three k-FWER controlling procedures under independence: Binomial-model based procedure, probability-inequality based procedure IV, and the LR procedure in terms of their critical values. Let the number of hypotheses $n$ be 20, 50, and 100 respectively. For each $n$, let $k = 1, 2, 3, 4$ respectively. Figure 4.3 shows the results for each pair of $n$ and
When $k = 1$, that is the $k$-FWER control reduces to FWER control, the procedure IV reduces to the procedure III and the LR procedure reduces to Holm’ procedure.

From Figure 4.3, we can see that, when $k = 1$ all procedures perform quite similarly for all $n$. The procedure IV improves the LR procedure quite slightly. It is expected for the procedure IV considering Table 3.1 and Figure 3.2-3.3. But it seems that even the Binomial-model based method can not improve the other two methods much neither. All of their critical values are no greater than $\alpha = 0.05$. When $k > 1$, the Binomial-model based procedure and procedure IV have significant improvement on the LR procedure. Their critical values begin to exceed 0.05. It also can be seen that, when $k = 2$, the Binomial-model based procedure and procedure IV perform almost the same for all $n = 20, 50, 100$. The former begins to improve on the latter dramatically as $k$ exceeds 2 for all $n$. In addition, the critical values of the procedure IV maintain for all $k \geq 2$ for each $n$, while those of the Binomial-model based procedure constantly increase as $k$ increases. The reason that the latter is better than the former under independence is obvious: the Binomial model can more sufficiently describe the $k$-FWER controlling procedure than the probability inequalities.

### 4.3 Simulation Study

We studied the power of the procedure IV by a simulation study. Consider to test $n$ hypotheses $H_1, \ldots, H_n$ simultaneously based on the observations $X_{ij}, j = 1, \ldots, N$ for $H_i, i = 1, \ldots, n$. Fixing $i$, suppose that $X_{i1}, \ldots, X_{iN}$ are independent and identically normally distributed with mean $\mu_i$ and variance $\sigma^2$. Let $T_i$ be the one-sample t-test statistics based on $X_{ij}, j = 1, \ldots, N$ corresponding to $H_i$. For the single hypothesis $H_i$: $\mu_i = 0$, we reject the hypothesis for large values of $T_i$. Denote by $\rho_{ij}$ be the correlation between $T_i$ and $T_j$. Assume that all test statistics are equally correlated, i.e., $\rho_{ij} = \rho$ for $i \neq j$. Let $R_i$ be the right-tailed p-values for $H_i$ based on $T_i$ and $R_{(i)}$ the ordered p-values for the corresponding hypothesis $H_{(i)}$. In addition, assume that there are $m(\leq n)$ false null hypotheses with $\mu_i = \mu > 0$. 
Our simulation study is based on the following specific values for all parameters above: 
k = 2, n = 20, N = 51, \mu = 1, and \sigma^2 = 4. Thus, under the null hypothesis \( H_i: \mu_i = 0 \),
\( T_i \) has t distribution with degrees of freedom \( v = 50 \); when \( \mu_i = \mu \), \( T_i \) has non-central t
distribution with degrees of freedom \( v = 50 \) and non-centrality parameter \( \sqrt{N\mu/\sigma} = 3.57 \).
The joint distributions of p-values \( R_i \) and \( R_j \) under null hypotheses \( H_k: \mu_k = 0 \), \( k = i, j \), are
calculated based on the joint distributions of \( T_i \) and \( T_j \), which are bivariate t distributions
with 20 degrees of freedom and correlation \( \rho \). Let \( \alpha = 0.05 \). The 2-FWER and average power
of this multiple testing problem using two step-down procedures: LR and IV, are calculated,
for each \( \rho = 0, 0.2, 0.35, 0.5, 0.6, 0.75, 0.9, 0.99 \), and each possible \( m = 0, 1, 2, ..., 19 (m = 20 \)
\( \text{case is trivial}) \). Figure 3.2 and 3.3 give the simulation-based estimates for 2-FWER and
average power respectively. The simulation uses 25000 iterations for each pair of \( \rho \) and \( m \).

From Figure 4.4-4.5, we can see that the procedure IV has bigger 2-FWER and average
power than the LR for all correlations and possible number of false null hypotheses. The
differences are more dramatic as the correlation are close to both 0 and 1 than those when
the correlation is around medium, 0.5. Actually, this fact is consistent with Figure 4.1-4.2.
Moreover, the fact that the procedure IV is uniformly more powerful than the LR procedure
under this equal-correlation case matches the result of Theorem 4.2.5 and Proposition 4.2.6,
noting that for the large sample size \( N = 51 \), bivariate normal distributions can be taken as
approximation of bivariate t distributions.
Figure 4.3: Comparison of the critical values of three $k$-FWER controlling procedures under independence: Binomial-model based procedure, probability-inequality based procedure IV, and the LR procedure for $n$ one-sided hypothesis tests based on $z$-test statistics. $k$ takes 1, 2, 3, 4, respectively (columns); $n$ takes 20, 50, 100, respectively (rows).
Figure 4.4: Simulation-based estimates of 2-FWER for two 2-FWER controlling methods, the LR procedure and the procedure IV for different correlations $\rho$ and the number of false null hypotheses from 0 to 18
Figure 4.5: Simulation-based estimates of the average power (the proportion of the false null hypotheses which are correctly rejected) for two 2-FWER controlling methods, the LR procedure and the procedure IV for different correlations $\rho$ and the number of false null hypotheses from 1 to 19.
CHAPTER 5

POWER OF BONFERRONI
PROCEDURE WITH SIGN TESTS

In this chapter, we attempt to study the power of the Bonferroni procedure when testing multiple hypotheses. We try to answer that, as the uniformly most powerful (UMP) test can be defined for one individual hypothesis, if we can define the similar concepts for multiple testing problems and how to obtain such tests that make the power of a certain multiple testing procedure as large as possible. We propose an answer for the Bonferroni procedure in the case that all test statistics are mutually independent. Specifically, we see that under non-parameter setting the sign test is a UMP test under appropriate assumption, and show that it also makes the Bonferroni procedure achieve as much power as possible among all tests in terms of two typical kinds of power definition under appropriate assumption. In addition, the power comparison among sign test, z-test, and t-test with the Bonferroni procedure is numerically investigated for a specific example of multiple testing problem by simulation.

5.1 Sign Test

In the general one-sample problem, the data from a single sample are used to obtain information about some particular aspect of the population distribution, usually one or more of its
parameters. The classical normal-theory test (variance known) or Student’s t test (variance unknown) for the hypotheses $H_0: \mu = \mu_0$ and $H_0: \mu_X - \mu_Y = \mu_0$ for the one sample and paired sample problems, respectively, are derived under the assumption that the single population or the population of differences of pairs is normal. For the nonparametric tests, only certain continuity assumptions about the populations are needed to determine the sampling distributions of the test statistics. The hypotheses here are concerned with the median or some other quantiles rather than the mean as the location parameter, although the mean and the median do coincide for symmetric populations. In any population, the median always exists (which is not true for the mean) and it is more robust as an estimate of location. The popular sign test and Wilcoxon signed-rank test are among those nonparametric hypotheses tests.

Suppose that a random sample of $N$ observations $X_1, X_2, ..., X_N$ is drawn from a population $F_X$ with an unknown median $M$, where $F_X$ is assumed to be continuous and strictly increasing, at least in the vicinity of $M$. In other words, the $N$ observations are independent and identically distributed, and $F_X^{-1}(0.5) = M$ uniquely. The hypothesis to be tested concerns the value of the population median

$$H_0: M = M_0$$

where $M_0$ is a specified value, against a corresponding one or two sided alternative. Since we assumed that $F_X$ has a unique median, under the null hypothesis, $M_0$ divides the area under the pdf into two equal parts. In symbols, we have

$$H_0: \theta = P(X > M_0) = P(X < M_0) = 0.5.$$

Thus, under the null hypothesis, the number of observations larger than $M_0$, denoted by $K$, can be used to test the validity of the null hypothesis. When the sample observations are dichotomized in the way, they constitute a set of $N$ independent random variables from
the Bernoulli population with parameter \( \theta = P(X > M_0) \), regardless of the population \( F_X \).

The sampling distribution of the random variable \( K \) is then the binomial distribution with parameters \( N \) and \( \theta \). \( \theta = 0.5 \) if the null hypothesis is true. Since \( K \) is the number of plus signs among the \( N \) differences \( X_i - M_0, i = 1, 2, ..., N \), the nonparametric test based on \( K \) is called the sign test.

The rejection region \( R \) for the upper-tailed alternative \( H_1 : M > M_0 \) or \( \theta = P(X > M_0) > 0.5 \) is \( K \in R \) for \( K \geq k_\alpha \), where \( k_\alpha \) is chosen to be the smallest integer that satisfies

\[
P(K \geq k_\alpha | H_0) = \sum_{i=k_\alpha}^{N} \binom{N}{i} 0.5^N \leq \alpha.
\]

Similarly, for a one-sided test with the lower-tailed alternative \( H_1 : M < M_0 \) or \( \theta = P(X > M_0) < 0.5 \), the rejection region for an \( \alpha \) level test is \( K \in R \) for \( K \leq k'_\alpha \), where \( k'_\alpha \) is chosen to be the smallest integer that satisfies

\[
P(K \leq k'_\alpha | H_0) = \sum_{i=0}^{k'_\alpha} \binom{N}{i} 0.5^N \leq \alpha
\]

If the alternative is two-sided, \( H_1 : M \neq M_0 \) or \( \theta = P(X > M_0) \neq 0.5 \), the rejection region is \( K \geq k_{\alpha/2} \) or \( K \leq k'_{\alpha/2} \), where \( k_{\alpha/2} \) and \( k'_{\alpha/2} \) are respectively, the smallest and the largest integers such that

\[
P(K \geq k_{\alpha/2} | H_0) = \sum_{i=k_{\alpha/2}}^{N} \binom{N}{i} 0.5^N \leq \alpha/2
\]

and

\[
P(K \leq k'_{\alpha/2} | H_0) = \sum_{i=0}^{k'_{\alpha/2}} \binom{N}{i} 0.5^N \leq \alpha/2.
\]

Obviously, we have the relation \( k_{\alpha/2} = N - k'_{\alpha/2} \) since the binomial distribution is symmetric when \( \theta = 0.5 \). The sign test statistic with these rejection regions is consistent (Gibbons and Chakraborti 2010).

The P-value of the sign test can be obtained from the Binomial probability table. For
example, if the alternative is upper-tailed $H_1 : M > M_0$ and $K_0$ is the observed value of the sign statistic, the P-value for the sign test is given by the binomial probability in the upper tail \( \sum_{i=K_0}^{N} \binom{N}{i} 0.5^N \).

We could easily generate tables to apply the exact sign test for any sample size $N$. However, we know that the normal approximation to the binomial is especially good when \( \theta = 0.5 \). Therefore, for moderate values of $N$, the normal approximation to the binomial can be used to determine the rejection regions. Since this is a continuous approximation to a discrete distribution, a continuity correction of 0.5 may be incorporated in the calculations.

For example, for the alternative $H_1 : M > M_0$, $H_0$ is rejected for $K \geq k_\alpha$, where $k_\alpha$ satisfies $k_\alpha = 0.5N + 0.5 + 0.5\sqrt{N}z_\alpha$. Similarly, the approximate $P$ value is $1 - \Phi(\frac{K_0 - 0.5 - 0.5N}{\sqrt{0.25N}})$.

5.2 The Power of Sign Test

In order to calculate the power of any test, the distribution of the test statistic under the alternative hypothesis should be available in a reasonably tractable form. In contrast to most nonparametric tests, the power function of the quantile tests is simple to determine since, in general, the random variable $K$ follows the binomial probability distribution with parameters $N$ and $\theta$, where $\theta = P(X_i > k_p)$ for the $p$th quantile. For the sign test the quantile of interest is the median and $\theta = P(X_i > M_0)$. For illustration, we will only consider the power of the sign test against the one-sided upper-tailed alternative $H_1 : M > M_0$.

The power is a function of the unknown parameters $\theta$, and the power curve or the power function is a graph of power versus various values of $\theta$ under the alternative. By definition, the power of the sign test against the alternative $H_1$ is the probability

$$Pw(\theta) = P(K \geq k_\alpha|H_1).$$

Under $H_1$, the distribution of $K$ is binomial with parameters $N$ and $\theta = P(X_1 > M_0|H_1)$ so
the expression for power can be written as

$$P_w(\theta) = \sum_{i=k_\alpha}^{N} \binom{N}{i} \theta^i (1-\theta)^{N-i},$$

where $k_\alpha$ is the smallest integer such that $\sum_{i=k_\alpha}^{N} \binom{N}{i} (0.5)^N \leq \alpha$. Thus, in order to evaluate the power function for the sign test, we first find the critical value $k_\alpha$ for a given significance level $\alpha$.

As an example, we calculate the power of the sign test of of $H_0 : M = 15$ versus $H_1 : M > 15$ for $N = 18$ at significance level 0.05, under the assumption that the population is normally distributed with standard deviation 1 and median $M = 15.9$. Based on Binomial distribution table, the rejection region at $\alpha = 0.05$ is $K \geq 13$ so that $k_\alpha = 13$ and the exact size of this sign test is 0.0481. Now, under the assumption given, we can derive the underlying probability $\theta$ of a success as

$$\theta = P(X > 15 | H_1) = P \left( \frac{X - 16}{1} > \frac{15 - 15.9}{1} \right)$$

$$= P(Z > -0.9) = 1 - \Phi(-0.9) = 0.82.$$

Thus, the power is $P_w(0.82) = \sum_{i=13}^{18} \binom{18}{i} (0.82)^i (0.18)^{18-i} = 0.911$. This would be directly comparable with the normal theory test of $H_0 : \mu = 15$ versus $H_1 : \mu = 15.9$ with $\sigma = 1$, since the mean and median coincide for the normal distribution. In order to ensure more fair comparison, we might find find the normal theory test of size $\alpha = 0.0481$ and compare the power of that test with the sign test power of 0.911. Thus, the rejection region for this parametric test with $\alpha = 0.0481$ is $\bar{X} > 15 + z_{0.0481}/\sqrt{18} = 15.392$, and the power is $P_w(15.9) = P(\bar{X} > 15.392 | X \sim normal(15.9, 1)) = 0.984$. Thus, the power of the normal theory test is larger than the power of the sign test, which is of course expected, since the normal theory test is known to be the best (uniformly most powerful (UMP)) test when the population is normal with known standard deviation.
In practice, under many situations, the normal assumption is not valid. The normal-theory tests cannot guarantee their power any more. However, for the sign test, the size and the power calculations are exact for all sample sizes and no distribution assumptions are needed other than continuity. Actually under the assumption that the underlying population is continuous, at least in the vicinity of its median, the sign test is the most powerful test. Before proving this, we review the concept of the uniformly most powerful (UMP) tests.

In hypotheses testing, uniformly most powerful (UMP) tests are always expected. Let $C$ be a class of tests for testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^c$. A test in class $C$, with power function $\beta(\theta)$, is a uniformly most powerful (UMP) class $C$ test if $\beta(\theta) \geq \beta'(\theta)$ for every $\theta \in \Theta_0^c$ and every $\beta'(\theta)$ that is a power function of a test in class $C$. The class $C$ is generally the class of all level $\alpha$ tests.

**Proposition 5.2.1.** Suppose that a random sample $X_1, \ldots, X_N$ is drawn from a population $F_X$ with an unknown median $M$, where $F_X$ is assumed to be continuous, at least in the vicinity of $M$. Suppose the pdf of $F_X$ is $f_M(x)$. For testing $H_0 : M \leq M_0$ versus $H_1 : M > M_0$, the sign test

$$\phi(X) = \begin{cases} 1, & \text{if } K \geq k_\alpha \\ 0, & \text{if } K < k_\alpha \end{cases}$$

(5.1)

is a uniformly most powerful level $\alpha$ test, where $K$ is the number of observations larger than $M_0$ and $k_\alpha$ is chosen to be the smallest integer that satisfies

$$P(K \geq k_\alpha|H_0) \leq \alpha.$$ 

**Proof.** According to the assumption, the $N$ observations are independent and identically distributed, and $F_0^{-1}(M) = M$ uniquely. First, for some $M_1 > M_0$, consider the problem testing $H'_0 : M = M_0$ versus $H'_1 : M = M_1$, and let $f_0(x)$ and $f_1(x)$ be the pdfs corresponding to $M_0$ and $M_1$ respectively. By the Neyman-Pearson Lemma, the most powerful test for the
problem is determined by
\[
\frac{\prod_{i=1}^{N} f_0(x_i)}{\prod_{i=1}^{N} f_1(x_i)} \leq k, \quad (5.2)
\]
where \(k\) can be determined by the level constraint.

Distribution \(F_X\) can be characterized by the probability \(p = P(X > M_0)\) together with the conditional probability distributions \(P_-\) and \(P_+\) of \(X\) given \(X \leq M_0\) and \(X > M_0\) respectively. That is, for a sample point \(x\),

\[
F_X(x) = P(X \leq x) = P(X \leq x|X \leq M_0)P(X \leq M_0) + P(X \leq x|X > M_0)P(X > M_0),
\]

\[
= P_-(x)(1 - p) + P_+(x)p.
\]

If the distributions \(P_-\) and \(P_+\) have probability densities \(p_-\) and \(p_+\), for example with respect to \(\mu = P_- + P_+\), then the joint density of \(X_1, ..., X_N\) at a sample point \(x_1, ..., x_N\) satisfying

\[
x(1), ..., x(N-K) \leq M_0 \leq x(N-K+1), ..., x(N)
\]
is

\[
p^{N-K}(1 - p)^K p_-(x(1))...p_-(x(N-K))p_+(x(N-K+1))...p_+(x(N)).
\]

Suppose \(p\) be \(p_0(= \frac{1}{2})\) and \(p_1\) under \(H'_0\) and \(H'_1\) respectively. Then (5.2) becomes

\[
\left(\frac{p_0}{p_1}\right)^{N-K} \left(\frac{1 - p_0}{1 - p_1}\right)^K \leq k,
\]

and hence \(K \geq k_\alpha\) with \(p_0 < p_1\). The test therefore rejects \(H'_0\) when the number \(K\) is sufficiently large, or more precisely, when \(K \geq k_\alpha\), where \(k_\alpha\) is the smallest integer that satisfies

\[
P(K \geq k_\alpha) \leq \alpha \text{ for } p = p_0 \text{ or } M = M_0.
\]
Note that the rejection region \( \{ K \geq k_\alpha \} \) is not depending on \( M_1 \). In addition, for any \( M_2 < M_0, P(K \geq k_\alpha | M = M_2) \leq P(K \geq k_\alpha | M = M_0) \leq \alpha \). Therefore the test (5.1) is the uniformly most powerful level \( \alpha \) test of \( H_0 : M \leq M_0 \) \( H_1 : M > M_0 \).

It is obvious that testing \( M \leq M_0 \) is equivalent to testing \( p = P(X > M_0) \leq P(X > M) = \frac{1}{2} \), where \( p \) is the success probability in the binomial distribution \( B(N, p) \). Similarly, it is known that the test

\[
\phi(X) = \begin{cases} 
1, & \text{if } K \geq k_{1\alpha} \text{ or } K \leq k_{2\alpha} \\
0, & \text{if } k_{2\alpha} < K < k_{1\alpha},
\end{cases}
\]

(5.3)

where \( E\phi(X) \leq \alpha \) for \( p = \frac{1}{2} \), is the UMP unbiased test of \( H_0 : p = \frac{1}{2} \) against \( H_1 : p \neq \frac{1}{2} \). Hence, the sign test is also the UMP unbiased test of \( H_0 : M = M_0 \) against \( H_1 : M \neq M_0 \).

### 5.3 Uniformly Most Powerful Sets of Tests for Bonferroni Procedure

In the conventions of testing statistical hypotheses, we always want the tests to have as large power as possible at a predetermined level \( \alpha \), such as the UMP tests and the UMP unbiased tests. Similarly, in the multiple testing setting, given a specific kind of procedure, we will be interested in how to manage the individual tests to make the multiple testing procedure achieve the largest power uniformly. In this section, we will investigate under what situations the Bonferroni procedure would uniformly have the largest power at a predetermined level.

Firstly, we will strictly define what we mean here by a particular set of tests making a multiple testing procedure be the uniformly most powerful in terms of a certain kind of power definition. Consider to test \( H_{i0} : \theta_i \in \Theta_{i0} \) against \( H_{i1} : \theta_i \in \Theta_{i0}^c \) simultaneously for \( i = 1, 2, ..., n \) with a specific multiple testing procedure. Let \( C_i \) be a class of tests for testing \( H_{i0} : \theta_i \in \Theta_{i0} \) against \( H_{i1} : \theta_i \in \Theta_{i0}^c \). Denote the index set of all false null hypotheses by \( I \),
where \( I = \{i_1, ..., i_{n_1}\} \subset \{1, ..., n\} \) with \( 1 \leq n_1 \leq n \) (power is undefined for \( n_1 = 0 \)). Let \( \theta = (\theta_1, ..., \theta_n) \), where \( \theta_i \in \Theta_i^c \) for \( i \in I \) and \( \theta_i \in \Theta_i \) otherwise, and \( \theta_I = (\theta_{i_1}, ..., \theta_{i_{n_1}}) \).

Assume \( \Phi \) be a set of tests \( \{\phi_i\}_{i=1}^n, \phi_i \in C_i \). Let \( \beta(\theta_I|\Phi) \) be a certain power definition for the multiple testing procedure based on the set of tests \( \Phi \) (see section 1.1.3), noting that the power is a function of the parameter vector \( \theta_I \), and thus \( \theta \) and \( I \) as well. The set of tests \( \Phi \) in classes \( C_i, i = 1, ..., n \), with the power definition \( \beta(\theta_I|\Phi) \), is said to make the multiple testing procedure be the uniformly most powerful if \( \beta(\theta_I|\Phi) \geq \beta(\theta_I|\Phi') \) for every nonempty subset \( I \subset \{1, ..., n\} \), every \( \theta_I \in \Theta_i^c \times \cdots \times \Theta_i^c \) and every \( \beta(\theta_I|\Phi') \) that is some power function of a set of tests \( \Phi' \) in classes \( C_i, i = 1, ..., n \). For the Bonferroni procedure, the class \( C_i \) is generally the class of all level \( \frac{\alpha}{n} \) tests for \( H_i \).

Then we will state our main result about which particular kind of sets of tests makes the Bonferroni procedure be the uniformly most powerful when the tests for \( n \) hypotheses are mutually independent.

**Theorem 5.3.1.** Let \( C_i \) be the class of all level \( \frac{\alpha}{n} \) tests for \( H_i \). Assume the tests from different classes \( C_i \) are mutually independent. Then the set of tests \( \Phi \) consisting of \( n \) UMP level \( \frac{\alpha}{n} \) tests for \( H_i, i = 1, ..., n \) respectively makes the Bonferroni procedure be uniformly most powerful for simultaneously testing \( H_i \) at level \( \alpha \) in terms of any of the following power definitions: (a) conjunctive power; (b) disjunctive power.

**Proof.** Assume \( \Phi = \{\phi_1, ..., \phi_n\} \), where \( \phi_i \) is a UMP level \( \frac{\alpha}{n} \) test for \( H_i \). Assume that \( H_{j0}, j = I = \{i_1, ..., i_{n_1}\} \) are false null hypotheses and \( P_j \) the corresponding p-values based on \( \phi_j \) at level \( \frac{\alpha}{n} \). For some \( \theta_I \in \Theta_i^c \times \cdots \times \Theta_i^c \) and some set of level \( \frac{\alpha}{n} \) tests \( \Phi' = \{\phi'_1, ..., \phi'_n\} \) where \( \phi'_i \) is in the class \( C_i \), we can see

(a) conjunctive power:

\[
\beta(\theta_I|\Phi) = Pr(H_j, j \in I \text{ are all rejected}|\Phi) = Pr(P_j \leq \frac{\alpha}{n}, j \in I)
\]
\[
= \prod_{j \in I} Pr(P_j \leq \frac{\alpha}{n}) \quad \text{(independence of } C_j) \\
\geq \prod_{j \in I} Pr(P_j' \leq \frac{\alpha}{n}) \quad \text{(} \phi_j \text{ is a UMP test)} \\
= Pr(P_j' \leq \frac{\alpha}{n}, j \in I) \quad \text{(independence of } C_j) \\
= \beta(\theta I|\Phi'),
\]

where \( P_j' \) is the P-value for \( H_j \) based on the test \( \phi_j' \).

(b) disjunctive power:

\[
\beta(\theta I|\Phi) = Pr(H_j \text{ are rejected for at least one } j \in I|\Phi) \\
= Pr(P_j \leq \frac{\alpha}{n}, \text{ for some } j \in I) \\
= 1 - Pr(P_j > \frac{\alpha}{n}, \text{ for all } j \in I) \\
= 1 - \prod_{j \in I} Pr(P_j > \frac{\alpha}{n}) \quad \text{(independence of } C_j) \\
\geq 1 - \prod_{j \in I} Pr(P_j' > \frac{\alpha}{n}) \quad \text{(} \phi_j \text{ is a UMP test)} \\
= 1 - Pr(P_j' > \frac{\alpha}{n}, \text{ for all } j \in I) \quad \text{(independence of } C_j) \\
= Pr(P_j' \leq \frac{\alpha}{n}, \text{ for some } j \in I) \\
= \beta(\theta I|\Phi').
\]

\[
\square
\]

5.4 Simulation Study

In this section, we will investigate the power of the Bonferroni procedure with sign tests for individual hypotheses through a specific example. The simulation results will demonstrate Proposition 5.2.1 and Theorem 5.2.2 numerically.
Firstly, we need to find a distribution under null hypothesis satisfying the assumption in Proposition 5.2.1, i.e., continuity at least in the vicinity of its median. Moreover, we need the distribution satisfies only this minimum assumption and does not have any other possible good properties, such as symmetry. Because with more assumptions for the distribution, other more efficient tests are possible. For example, with additional symmetry assumption, Wilcoxon rank-sum test has greater efficiency than sign test.

We choose the non-central t distribution \( F_{v,\mu}(x) \) with \( v \) degrees of freedom and non-centrality parameter \( \mu \) as such a distribution satisfying only continuity assumption. For \( \mu \neq 0 \), the non-central t distribution is not symmetric. Specifically, we consider the non-central t distribution \( F_{10,3}(x) \) with degrees of freedom \( v = 10 \) and non-centrality parameter \( \mu = 3 \). Then we can translate \( F_{10,3}(x) \) such that the obtained distribution has \( \theta \) as its median. Denote by \( F_\theta(x) \) the obtained distribution with median \( \theta \). Obviously, \( F_\theta(x) \) is continuous and non-symmetric.

Consider to test \( n \) hypotheses \( H_1, ..., H_n \) simultaneously based on the observations \( X_{ij}, j = 1, ..., N \) for \( H_i, i = 1, ..., n \). Suppose that \( X_{i1}, ..., X_{iN} \) is a random sample from \( F_\theta(x) \). Also, for different \( i \), these random samples are mutually independent. Assume under \( H_i \), \( \theta_i = 0 \) and there are \( m(\leq n) \) false null hypotheses with \( \theta_i = -0.5 \). Figure 5.1 shows the distribution under null hypothesis and the one under false null hypothesis. Under null hypothesis, the distribution looks like normal or t distribution. Actually, however, it is not symmetric and has slightly heavier right tail than the left. We use three different tests for each individual hypothesis: sign test, z-test, t-test. We reject \( H_i \) for small values of these test statistics.

When testing an individual null hypothesis \( H_i : \theta_i = 0 \) with underlying true \( \theta_i = -0.5 \) at level 0.05, the powers and Type I errors are calculated for the three tests with sample size \( N = 50 \). For sign test, z-test, and t-test, respectively, the powers are 0.6174835, 0.5478500, and 0.5365500 and the Type I errors are 0.03245432, 0.009050, 0.008600. The numerical results show that sign test has the greatest power of all. It is consistent with Proposition 5.2.1.
When addressing the multiple testing problem, we use the Bonferroni correction. By simulation, we estimate the disjunctive power and conjunctive power, as well as FWER, of Bonferroni procedure with the sets of sign test, z-test, and t-test, respectively. The simulation-based estimation is given for $n = 5$, $\alpha = 0.05$, and any $m = 1, 2, 3, 4, 5$ in Table 5.1 and Table 5.2. The simulation uses 20000 iterations for each $m$. From the tables, we can see that the set of 5 sign tests makes the Bonferroni procedure more powerful than the set of either z-tests or t-tests for simultaneously testing $H_1, \ldots, H_5$ in terms of both conjunctive and disjunctive power. This result is consistent with Theorem 5.2.2.
Table 5.1: Comparison of conjunctive power of the Bonferroni procedure with three different tests. \( m \) is the number of false null hypotheses

<table>
<thead>
<tr>
<th>( m )</th>
<th>Conjunctive Power</th>
<th>FWER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sign test</td>
<td>z-test</td>
</tr>
<tr>
<td>1</td>
<td>0.3798</td>
<td>0.31685</td>
</tr>
<tr>
<td>2</td>
<td>0.6125</td>
<td>0.5354</td>
</tr>
<tr>
<td>3</td>
<td>0.7641</td>
<td>0.68305</td>
</tr>
<tr>
<td>4</td>
<td>0.85255</td>
<td>0.7866</td>
</tr>
<tr>
<td>5</td>
<td>0.90625</td>
<td>0.85185</td>
</tr>
</tbody>
</table>

Table 5.2: Comparison of disjunctive power of the Bonferroni procedure with three different tests. \( m \) is the number of false null hypotheses

<table>
<thead>
<tr>
<th>( m )</th>
<th>Disjunctive Power</th>
<th>FWER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sign test</td>
<td>z-test</td>
</tr>
<tr>
<td>1</td>
<td>0.38235</td>
<td>0.31435</td>
</tr>
<tr>
<td>2</td>
<td>0.14605</td>
<td>0.099</td>
</tr>
<tr>
<td>3</td>
<td>0.0575</td>
<td>0.0328</td>
</tr>
<tr>
<td>4</td>
<td>0.02025</td>
<td>0.011</td>
</tr>
<tr>
<td>5</td>
<td>0.0086</td>
<td>0.003</td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


APPENDIX A

This appendix includes all R codes that generate the Tables and Figures in Chapter 3 through Chapter 5. The R Codes for Figure 3.2, 3.3, 4.4, and 4.5 (Listing 3 and 5) were processed using parallel computing on the cluster at the Ohio Supercomputer Center (OSC). The author gratefully acknowledges the supercomputer time granted by OSC for support.

1 R Codes for Chapter 3

Listing 1: Figure 3.1

```
library(multtest)
library(mutoss)
library(mvtnorm)

# ############ New Critical Values (Under Equal Correlation) ############
# ############ Upper One Sided t Distribution with degree v ############

CritiPti=function(p,n,alpha,corr,df){

c=numeric(100000)
c[1]=alpha

# compute the both tails quantiles for t distribution
q=qmvtnorm(1-c[1], tail = c("lower.tail"), df = df)$quantile

# compute the joint t distribution with correlation rho
corrmatrix=matrix(c(1,corr,corr,1),2,2)
jointprob=pmvtnorm(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
c[2]=(alpha+(n-p)*jointprob)/(n-p+1)
```

# New algorithm to update the critical values
i=1
while (c[i]-c[i+1]>0.00000000000001 & i<10000)
{
    i=i+1
    # compute the both tails quantiles for t distribution
    q=qmvt(1-c[i], tail = c("lower.tail"), df = df)$quantile
    # compute the joint t distribution with correlation rho
    jointprob=pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
    c[i+1]=(alpha+(n-p)*jointprob)/(n-p+1)
}
return(min(c[c!=0]))

CSCP1=function(p,n,alpha,corr,df){
    c=numeric(2)
    c[1]=alpha/(n-p+1)
    # compute the both tails quantiles for t distribution
    q=qmvt(1-c[1], tail = c("lower.tail"), df = df)$quantile
    # compute the joint t distribution with correlation rho
    corrmatrix=matrix(c(1,corr,corr,1),2,2)
    jointprob=pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
    c[2]=(alpha+(n-p)*jointprob)/(n-p+1)
    return(min(c[2],alpha/(n-p)))
}

############################################################

# New algorithm to update the critical values
i=1
while (c[i]-c[i+1]>0.00000000000001 & i<10000)
{
    i=i+1
    # compute the both tails quantiles for t distribution
    q=qmvt(1-c[i], tail = c("lower.tail"), df = df)$quantile
    # compute the joint t distribution with correlation rho
    jointprob=pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
    c[i+1]=(alpha+(n-p)*jointprob)/(n-p+1)
}
return(min(c[c!=0]))

CSCP1=function(p,n,alpha,corr,df){
    c=numeric(2)
    c[1]=alpha/(n-p+1)
    # compute the both tails quantiles for t distribution
    q=qmvt(1-c[1], tail = c("lower.tail"), df = df)$quantile
    # compute the joint t distribution with correlation rho
    corrmatrix=matrix(c(1,corr,corr,1),2,2)
    jointprob=pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
    c[2]=(alpha+(n-p)*jointprob)/(n-p+1)
    return(min(c[2],alpha/(n-p)))
}

############################################################
### Critical Points under Arbitrary Dependence

#### Critical Points based on Kounias's Inequality (KN inequality)

```r
KNCP0=function(n, alpha, CorrMatrix, df){

    CPKN <- numeric(n)
    CPKN[n] <- alpha

    KNgamma=numeric(n-1)
    MAXindex=numeric(n-1)  # record the index that has the maximum KN gamma term in each step

    ## Critical Points

    for(i in (n-1):1){  # i the index of P-value to be adjusted

        # compute the i-th Gamma terms in the Kounia's Inequality
        c=alpha/(n-i+1)

        # compute the both tails quantiles for t distribution
        q=qmvt(1-c, tail = c("lower.tail"), df = df)$quantile

        for (j in n:i){
            temp=0
            for (m in n:i)
                if(m!=j){
                    corrmatrix=matrix(c(1,CorrMatrix[j,m],CorrMatrix[j,m],1),2,2)
                    jointprob=pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
                    temp=temp+jointprob
                }
            if(KNgamma[i]<temp) {KNgamma[i]=temp; MAXindex[i]=j}
        }
    }
}
```
CPKN[i]=(alpha+KNgamma[i])/(n-i+1)

return(CPKN)

### New Algorithm based on Kounias's Inequality (KN inequality)

KNCP=function(n,alpha,CorrMatrix,df){
CPKN <- numeric(n)
CPKN[n] <- alpha

KNgamma=numeric(n-1)
MAXindex=numeric(n-1) # record the index that has the maximum KN gamma term in each step

## Critical Points

for(i in (n-1):1){ # i the index of P-value to be adjusted

# compute the ith Gamma terms in the Kounia's Inequality

c=numeric(100000)
c[1]=alpha
# compute the lower tail quantiles for t distribution
q=qmvt(1-c[1], tail = c("lower.tail"), df = df)$quantile

for (j in n:i){
    temp=0
    for (m in n:i)
        if(m!=j){
            corrmatrix=matrix(c(1,CorrMatrix[j,m],CorrMatrix[j,m],1),2,2)
            jointprob=pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
            temp=temp+jointprob
        }
}
}

}
if(KNgamma[i]<temp) {KNgamma[i]=temp; MAXindex[i]=j}

}  
c[2]=(alpha+KNgamma[i])/(n-i+1)
KNgamma[i]=0  # clear the memory of DSgamma

# New algorithm to update the critical values
k=2
while (c[k-1]-c[k]>.0000000001 & k<10000) {

  # compute the both tails quantiles for t distribution
  q=qmvt(1-c[k], tail = c("lower.tail"), df = df)$quantile

  j=MAXindex[i]
  temp=0
  for (m in n:i)
    if(m!=j){
      corrmatrix=matrix(c(1,CorrMatrix[j,m],CorrMatrix[j,m],1),2,2)
      jointprob=pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
      temp=temp+jointprob
    }
  KNgamma[i]=temp

  c[k+1]=(alpha+KNgamma[i])/(n-i+1)
  KNgamma[i]=0  # clear the memory of DSgamma
  k=k+1

}

CPKN[i]=min(c[c!=0])

}

return(CPKN)

}
Critical Values Analysis Function

CPCompare=function(n,d,rho1,rho2,alpha,df){

## Define Correlation Matrix ##

corr=.9
CorrMatrix=diag(n) # [i,j] is correlation for i\textsuperscript{th} and j\textsuperscript{th} statistics tests

#CorrMatrix[lower.tri(CorrMatrix)] <- corr
#CorrMatrix[upper.tri(CorrMatrix)] <- corr

CorrMatrix[1:(n/2+d),1:(n/2+d)]=rho1
CorrMatrix[(n/2+d+1):n,(n/2+d+1):n]=rho2

# generate a virtual order of P-values
Pvalues=sample(1:n)
OriginalIndex=sort.list(Pvalues)
Pvalues=sort(Pvalues)

CorrMatrix=CorrMatrix[OriginalIndex,] # rearrange the order of CorrMatrix in terms of the ordered P-values.
CorrMatrix=CorrMatrix[,OriginalIndex]

#CorrMatrix

New Procedure (I, II, III) Critical Points
NewCP_III <- NewCP_I <- NewCP_II <- numeric(n)

# (I) Procedure II Critical Points
for(i in (n-1):1){  # i the index of P-value to be adjusted
    # find the minimum correlation from the last i test statistics
    MINcorr=1
    for (j in n:(i+1))
        for (k in (j-1):i)
            MINcorr=min(CorrMatrix[j,k],MINcorr)
    NewCP_II[i]=CritiPt1(i,n,alpha,MINcorr,df)  # One Sided test Critical values
    # NewCP_II[i]=CritiPt2(i,n,alpha,MINcorr,df)  # Two Sided test Critical values
}

# (II) Procedure I Critical Points
NewCP_I=KNCP(n,alpha,CorrMatrix,df)
NewCP_I[1:(n-2)]=mapply(min,NewCP_I[1:(n-2)],alpha/(n-(1:(n-2))))  # the index needed to be replaced with alpha/(n-p)

# (III) Procedure III Critical Points
NewCP_III=mapply(max,NewCP_I,NewCP_II)

# CS Critical Points
CSCP=KNCP0(n,alpha,CorrMatrix,df)
CSCP[1:(n-2)]=mapply(min,CSCP[1:(n-2)],alpha/(n-(1:(n-2))))  # the index needed to be replaced with alpha/(n-p)

# Holm CP+
HCP=alpha/(n:1)
cbind(HCP,CSCP,NewCP_I,NewCP_II,NewCP_III)
### Plots

```r
m <- matrix(c(1,2),nrow = 2,ncol = 1,byrow = TRUE)

layout(mat = m,heights = c(0.6, 4))

par(oma = c(4,4,0,0) + 0.1)  ## giving size of the outer margins in lines of text
par(mai=c(0,0,0,0))
plot_colors <- c(1,2,3,4,5)
plot.new()
legend(x="center", ncol=3,legend=c("Holm","CS","Procedure I", "Procedure II", "Procedure III "),lty=c(1,2,3,4, 5), lwd=rep(2.5,5), col=plot_colors, seg.len=3)

par(mar = c(0,0.1.5,0))  ## giving size of each individual plot margins in lines of text

plot(1:n,HCP,type='l',lty=1,col=1,lwd=2.5)
points(1:n,CSCP,type='l',lty=2,col=2,lwd=2.5)
points(1:n,NewCP_I,type='l',lty=3,col=3,lwd=2.5)
points(1:n,NewCP_II,type='l',lty=4,col=4,lwd=2.5)
points(1:n,NewCP_III,type='l',lty=5,col=5,lwd=2.5)

title(xlab = 'Index',
ylab = 'Critical Values',
outer = TRUE, line = 2.7, cex.lab=1.2)

return(CorrMatrix)
)
```

```r
n=6
alpha=0.05
df=20

rho1=0.5
rho2=0.8
```
d=-1 # a whole number between -9 to 9

#pdf(file='C:\Users\Shihai\Desktop\MTP\Critical Points Comparison.pdf')

CPCompare(n,d,rho1,rho2,alpha,df)

#dev.off()

Listing 2: Table 3.1

library(multtest)
library(mutoss)
library(mvtnorm)

############### New Critical Values (Reverse Version) ###############

############## Upper One Sided t Distribution with degree v ##############

CritiPt1=function(p,n,alpha,corr,df){

c=numeric(100000)
c[1]=alpha

# compute the both tails quantiles for t distribution
q=qmvt(1-c[1], tail = c("lower.tail"), df = df)$quantile

# compute the joint t distribution with correlation rho
corrmatrix=matrix(c(1,corr,corr,1),2,2)
jointprob=pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
c[2]=(alpha+(n-p)*jointprob)/(n-p+1)

i=1
while (c[i]-c[i+1]>0.00000000000001 & i<10000){

  i=i+1

  # compute the both tails quantiles for t distribution
  q=qmvt(1-c[i], tail = c("lower.tail"), df = df)$quantile

  # compute the joint t distribution with correlation rho
  jointprob=pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
  c[i+1]=(alpha+(n-p)*jointprob)/(n-p+1)
}

return(min(c[c!=0]))
# ############ CS' Critical Values ############
# ############ Upper One Sided t Distribution with degree v ############

CSCP1=function(p,n,alpha,corr,df){

c=numeric(2)
c[1]=alpha/(n-p+1)
  # compute the both tails quantiles for t distribution
q=qmvtnorm(1-c[1], tail = c("lower.tail"), df = df)$quantile

  # compute the joint t distribution with correlation rho
corrmatrix=matrix(c(1,corr,corr,1),2,2)
jointprob=pmvtnorm(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
c[2]=(alpha+(n-p)*jointprob)/(n-p+1)

return(min(c[2],alpha/(n-p)))
}

# ############ Compute the critical points table ############
# ############ Under the same dependence structure ############

alpha=0.05
n=8
p=seq(1:n)
rho=c(0.0, 0.2, 0.35, 0.5, 0.7, 0.9, 0.99)
df=20
CP1=CP2=numeric(n)

CP1=sapply(p,CSCP1,n,alpha,rho,df)
round(CP1, digits = 5)

CP2=sapply(p,CritiPt1,n,alpha,rho,df)
round(CP2, digits = 5)
library(Rmpi)

# install.packages('snow_0.3-3.tar.gz', '/nfs/02/bgs0314/Rlibs', repos=NULL)

library(snow)
# load a package on each cluster node with
# clusterEvalQ(cl, library(Matrix))

# install.packages('BiocGenerics_0.4.0.tar.gz', '/nfs/02/bgs0314/Rlibs', repos=NULL)
# install.packages('Biobase_2.18.0.tar.gz', '/nfs/02/bgs0314/Rlibs', repos=NULL)
# install.packages('mvtnorm_0.9-9993.tar.gz', '/nfs/02/bgs0314/Rlibs', repos=NULL)
# install.packages('plotrix_2.4-1.tar.gz', '/nfs/02/bgs0314/Rlibs', repos=NULL)
# install.packages('multtest_2.14.0.tar.gz', '/nfs/02/bgs0314/Rlibs', repos=NULL)
# install.packages('multcomp_1.1-0.tar.gz', '/nfs/02/bgs0314/Rlibs', repos=NULL)
# install.packages('mutoss_0.1-1.tar.gz', '/nfs/02/bgs0314/Rlibs', repos=NULL)

library(mutoss)
library(mvtnorm)

pdf(file='FWERpowerFAST0.pdf')

# Underlying multinormal
require(MASS)

#########################################################

########## New Critical Values ##########
########## Upper One Sided t Distribution with degree v ##########

CritiPt1=function(p,n,alpha,corr,df){

c=numeric(100000)
c[1]=alpha
# compute the lower tails quantiles for t distribution
q=qmvt(1-c[1], tail = c("lower.tail"), df = df)$quantile

# compute the joint t distribution with correlation rho

corrmatrix=matrix(c(1,corr,corr,1),2,2)
jointprob=pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
c[2]=(alpha+(n-p)*jointprob)/(n-p+1)

i=1
while (c[i]-c[i+1]>0.0000000000001 & i<10000){

  i=i+1

  # compute the lower tails quantiles for t distribution
  q=qmv qt(1-c[i], tail = c("lower.tail"), df = df)$quantile

  # compute the joint t distribution with correlation rho
  jointprob=pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
c[i+1]=(alpha+(n-p)*jointprob)/(n-p+1)
}

return(max(c[c!=0]))

}

# New Critical Values #

# Two Sided t Distribution with degree v

CritiPt2=function(p,n,alpha,corr,df){

c=numeric(100000)
c[1]=alpha

  # compute the both tails quantiles for t distribution
  q=qmv t(1-c[1], tail = c("both.tails"), df = df)$quantile

  # compute the joint t distribution with correlation rho
  corrmatrix=matrix(c(1,corr,corr,1),2,2)
jointprob=2*pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
c[2]=(alpha+(n-p)*jointprob)/(n-p+1)

i=1
while (c[i]-c[i+1]>0.0000000000001 & i<10000){

  i=i+1

  # compute the lower tails quantiles for t distribution
  q=qmv qt(1-c[i], tail = c("lower.tail"), df = df)$quantile

  # compute the joint t distribution with correlation rho
  jointprob=pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
c[i+1]=(alpha+(n-p)*jointprob)/(n-p+1)
}

return(max(c[c!=0]))

}
i=i+1
# compute the both tails quantiles for t distribution
q=qmvt(1-c[i], tail = c("both.tails"), df = df)$quantile
# compute the joint t distribution with correlation rho
jointprob=2*pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
c[i+1]=(alpha+(n-p)*jointprob)/(n-p+1)
}

return(max(c[c!=0]))
)

################ CS' Critical Values ################

 Shakespeare: Critical Values

 Upper One Sided t Distribution with degree v 

 CSCP1=function(p,n,alpha,corr,df){
c=numeric(2)
c[1]=alpha/(n-p+1)
# compute the both tails quantiles for t distribution
q=qmvt(1-c[1], tail = c("lower.tail"), df = df)$quantile
# compute the joint t distribution with correlation rho
corrmatrix=matrix(c(1,corr,corr,1),2,2)
jointprob=pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
c[2]=(alpha+(n-p)*jointprob)/(n-p+1)
return(min(c[2],alpha/(n-p)))
}

################ CS' Critical Values ################

 Two Sided t Distribution with degree v 

 CSCP2=function(p,n,alpha,corr,df){
c=numeric(2)
c[1]=alpha/(n-p+1)
# compute the both tails quantiles for t distribution

q=qmvt(1-c[1], tail = c("both.tails"), df = df)$quantile

# compute the joint t distribution with correlation rho
corrmatrix=matrix(c(1,corr,corr,1),2,2)
jointprob=2*pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
c[2]=(alpha+(n-p)*jointprob)/(n-p+1)
return(min(c[2],alpha/(n-p)))

######################################################## CS Procedure ##########################################################

CS=function(Pvalues,CSCP){

# CorMatrix: the matrix storing the correlation between test statistics in the ascending order
# the (i, j) element corresponds to the correlation
# between the (n-i+1)th and (n-j+1)th largest P-values
# df: the freedom of t distribution

n=length(Pvalues)

OriginalIndex=sort.list(Pvalues)
Pvalues=sort(Pvalues)

# do the CS

FirstAcception=ifelse(length(which(Pvalues>CSCP))>0,min(which(Pvalues>CSCP)),n+1)
Reject=c(rep('TRUE',FirstAcception-1),rep('FALSE',n-FirstAcception+1))

# restore the order of P values

temp1 <- temp2 <- numeric(n)
for(i in 1:n) { # restore the order of the p-values
  temp1[OriginalIndex[i]]=Pvalues[i]
temp2[OriginalIndex[i]]=Reject[i]
### New MTP

```r
# CorMatriz: the matrix storing the correlation between test statistics in the ascending order
# the (i,j) element corresponds to the correlation between the (n-i+1)th and (n-j+1)th largest P-values
# df: the freedom of t distribution

n <- length(Pvalues)

OriginalIndex <- sort.list(Pvalues)  # restore the order of P values
Pvalues <- sort(Pvalues)

FirstAcception <- ifelse(length(which(Pvalues > NewCP)) > 0, min(which(Pvalues > NewCP)), n + 1)
Reject <- c(rep('TRUE', FirstAcception - 1), rep('FALSE', n - FirstAcception + 1))

# restore the order of P values

temp1 <- temp2 <- numeric(n)
for (i in 1:n) {  # restore the order of the p-values
  temp1[OriginalIndex[i]] <- Pvalues[i]
  temp2[OriginalIndex[i]] <- Reject[i]
}
```

```r
# end of CS function

NewMTP <- function(Pvalues, NewCP) {

  n <- length(Pvalues)
  OriginalIndex <- sort.list(Pvalues)
  Pvalues <- sort(Pvalues)

  FirstAcception <- ifelse(length(which(Pvalues > NewCP)) > 0, min(which(Pvalues > NewCP)), n + 1)
  Reject <- c(rep('TRUE', FirstAcception - 1), rep('FALSE', n - FirstAcception + 1))

  temp1 <- temp2 <- numeric(n)
  for (i in 1:n) {  # restore the order of the p-values
    temp1[OriginalIndex[i]] <- Pvalues[i]
    temp2[OriginalIndex[i]] <- Reject[i]
  }

  Pvalues <- temp1
  Reject <- temp2
  result <- Reject
  return(result)
}
```
Pvalues=temp1
Reject=temp2

result=Reject
return(result)

) # end of NewMTP function

******* FWER and Power Analysis *******

******* Step I: Simulate test statistics *******
******* n=8, t test: mu=0 vs mu=delta, df=16 *******

alpha=0.05
corr=c(0,.3,.7,.9,.95,.99)
n=8
N=200 # number of replications
df=30
delta0=2 # set mean values for the alternative hypotheses
CRP_h <- FRP_h <- CRP_n <- FRP_n <- CRP_c <- FRP_c <- numeric(N)
   ## CR : Correctly Rejection proportion
   ## FR : Falsely Rejection proportion
FWER_h <- POWER_h <- FWER_n <- POWER_n <- FWER_c <- POWER_c <- numeric(n)

##### The function of FWER and POWER analysis #######

FWER=function(corr,df,alpha,n,delta0,N,CritiPt1,CSCP1,CS,NewMTP){

   CRP_h <- FRP_h <- CRP_n <- FRP_n <- CRP_c <- FRP_c <- numeric(N)
   FWER_h <- POWER_h <- FWER_n <- POWER_n <- FWER_c <- POWER_c <- numeric(n)

   corrmatrix=diag(n)

   for(i in 1:n){
      for (j in 1:n){

         corrmatrix
if(corrmatrix[i,j]!=1) corrmatrix[i,j]=corr  # corr is the correlation between any
two test statistics
j=j+1
i=i+1
}

## Calculate the critical values using correlation information
# (I) CSCP & NewCP
CSCP <- NewCP <- numeric(n)
CSCP[n] <- NewCP[n] <- alpha

for(i in (n-1):1){  # i the index of P-value to be adjusted

  # find the minimum correlation from the last i test statistics
  MINcorr=1
  for (j in n:(i+1))
    for (k in (j-1):i)
      MINcorr=min(corrmatrix[j,k],MINcorr)

  CSCP[i]=CSCP1(i,n,alpha,MINcorr,df)  # One Sided test Critical values
  NewCP[i]=CrtiPt1(i,n,alpha,MINcorr,df)  # One Sided test Critical values

  # CSCP[i]=CSCP2(i,n,alpha,MINcorr,df)  # Two Sided test Critical values
  # NewCP[i]=CrtiPt2(i,n,alpha,MINcorr,df)  # One Sided test Critical values
}

#### the loop the compute the POWER and FWER

for (i in 1:n){  # (i-1) the number of false hypotheses

delta=c(rep(0,n-(i-1)),rep(delta0,i-1))
s=rmvt(N, sigma = corrmatrix, df = df, delta=delta)

## compute N sets of P-values under null hypotheses

# Pvalues=ifelse(s>0,2*pt(s, df=df, lower.tail = FALSE),2*pt(s, df=df, lower.tail = TRUE))
Pvalues=pt(s, df=df, lower.tail = FALSE)
PVALUES=list(Pvalues[1,])  # store in lists
for(j in 2:N) PVALUES=c(PVALUES,list(Pvalues[j,]))

## Holm procedure

for (j in 1:N){
  REJ=holm(PVALUES[[j]],alpha,silent=T)$rejected
  RejIndex=which(REJ==TRUE)
  if(length(delta[delta==delta0])==0){
    CRP_h[j]=1}else{
    CRP_h[j]=sum(ifelse(delta[RejIndex]==delta0,1,0))/length(delta[delta==delta0])
  }
  FRP_h[j]=sum(ifelse(delta[RejIndex]==0,1,0))/length(delta[delta==0])
}
FWER_h[i]=mean(ifelse(FRP_h>0,1,0))
POWER_h[i]=mean(CRP_h)

## New Procedure

REJ=t(sapply(PVALUES,NewMTP,NewCP))
for (j in 1:N){
  RejIndex=which(REJ[j,]==TRUE)
  if(length(delta[delta==delta0])==0){
    CRP_n[j]=1}else{
    CRP_n[j]=sum(ifelse(delta[RejIndex]==delta0,1,0))/length(delta[delta==delta0])
  }
  FRP_n[j]=sum(ifelse(delta[RejIndex]==0,1,0))/length(delta[delta==0])
}
FWER_n[i]=mean(ifelse(FRP_n>0,1,0))
POWER_n[i]=mean(CRP_n)

## CS Procedure

REJ=t(sapply(PVALUES,CS,CSCP))
for (j in 1:N){
  RejIndex=which(REJ[j,]==TRUE)
  if(length(delta[delta==delta0])==0){
    CRP_c[j]=1}else{
    CRP_c[j]=sum(ifelse(delta[RejIndex]==delta0,1,0))/length(delta[delta==delta0])
  }
  FRP_c[j]=sum(ifelse(delta[RejIndex]==0,1,0))/length(delta[delta==0])
}
FWER_c[i]=mean(ifelse(FRP_c>0,1,0))
POWER_c[i]=mean(CRP_c)

return(c(list(FWER_h),list(FWER_c),list(FWER_n),list(POWER_h),list(POWER_c),list(POWER_n)))

# ###########################################################################################

Beginning=Sys.time()

# Test snow on the OSC cluster.
# First, get the cluster info.
cl <- makeCluster()

# load a package on each cluster node with
clusterEvalQ(cl, library(mutoss))
clusterEvalQ(cl, library(mvtnorm))

# Now generate some random variables on all of the clusters. Note,
# because we haven't set a different seed for all of the processes, you
# may get back duplicates. GenMatch and rgenoud takes care of this
# for you, but other R libraries may not. See the snow documentation
# for more details.

alpha=0.05
corr=c(0, 0.2, 0.35, 0.5, 0.6, 0.75, 0.9, 0.99)
s=20    ## number of divisions of total samples
Corr=rep(corr,s)
n=10    ## number of hypotheses
N=30000  ## number of replications
df=20
delta0=3.24

result=parSapply(cl, Corr, FWER, df=df, alpha=alpha,n=n,delta0=delta0,N=N,CritiPt1,CSCP1,CS,
NewMTP)

# Stop cluster
stopCluster(cl)

Sys.time()-Beginning

############## Combine all s sets of samples together and store in the first length(corr)
columns in the 'result'##############

length(result[[1,2]])

result[[1,2]]

for(j in 1:6) # 6 indicate 3 FWER and 3 POWER
  for(k in 1:length(corr)){
    for (i in 1:(s-1))
      result[[j,k]]=result[[j,k]]+result[[j,(k+length(corr)*i)]]
      result[[j,k]]=result[[j,k]]/s
  }

length(result[[1,2]])

result[[1,2]]

############## Plots Analysis ##############

m <- matrix(c(1,1,1,2,3,4,5,6,7,8,9),nrow = 3,ncol = 4,byrow = TRUE)
par oma = c(4, 4, 0, 0) + 0.1)  ## giving size of the outer margins in lines of text

# ########## Plots Analysis ###############

### FWER ###

par(mai = c(0, 0, 0, 0))

plot_colors <- c("blue", "green", "red")

legend(x="center", ncol=3, legend=c("Holm", "CS", "Procedure III"), lty = c(1, 2, 4), lwd = c(1.7, 1.7, 1.7), col = plot_colors, seg.len = 3.5)

par(mar = c(0, 0, 2.7, 0))  ## giving size of each individual plot margins in lines of text

for (i in 1:length(corr)){
  plot(0:(n-1), result[[1,i]], type = 'l', col = 'blue', lwd = 1.5, ylim = c(0, 0.055), main = bquote(paste(rho, '='. (corr[i]) )), xaxt = "n", yaxt = "n")
  points(0:(n-1), result[[2,i]], type = 'l', col = 'green', lwd = 1.5, lty = 'dashed')
  points(0:(n-1), result[[3,i]], type = 'l', col = 'red', lwd = 1.5, lty = 'dotdash')
  axis(side = 1, labels = (i == 6 || i == 8))
  axis(side = 2, labels = (i == 1 || i == 5))
}

title(xlab = '# of False Null', ylab = 'FWER', outer = TRUE, line = 3, cex.lab = 1.4)

### POWER ###

par(mai = c(0, 0, 0, 0))

plot_colors <- c("blue", "green", "red")

plot.new()
```
legend(x="center", ncol=3, legend=c("Holm","CS","Procedure III"), lty=c(1,2,4), lwd=c
(1.7,1.7,1.7), col=plot_colors, seg.len=3.5)

par(mar = c(0,0,2.7,0)) ## giving size of each individual plot margins in lines of text
for (i in 1:length(corr)){
  plot(1:(n-1), result[[4,i]][2:n], type='l', col='blue', lwd=1.5, ylim=c(0,1),
       main=bquote(paste(rho,'=',.(corr[i]))), xaxt = "n", yaxt="n")
  points(1:(n-1), result[[5,i]][2:n], type='l', col='green', lwd=1.5, lty='dashed')
  points(1:(n-1), result[[6,i]][2:n], type='l', col='red', lwd=1.5, lty='dotdash')
  axis(side = 1, labels = (i == 6 || i == 8))
  axis(side = 2, labels = (i == 1 || i == 5))
}

title(xlab = '# of False Null',
     ylab = 'Average Power',
     outer = TRUE, line = 3, cex.lab=1.4)

dev.off()
```

Listing 4: Table 3.2

```
library(mvtnorm)
library(ape)

# ###################################################################
# #### R Code for An Application Example
# ##### Critical Values of the procedure III in this example is exactly
# #### those of the procedure I
# ###################################################################

# ## Case II -- Critical Points based on Kounias's Inequality (KN inequality)
# ## Two Sided t

KNCP0=function(n,alpha,CorrMatrix,df,side){
```
CPKN <- numeric(n)
CPKN[n] <- alpha

KNgamma=numeric(n-1)  # record the index that has the maximum KN gamma term in each step

## Critical Points

for(i in (n-1):1){  # i the index of P-value to be adjusted

  # compute the ith Gamma terms in the Kounia's Inequality
  c=alpha/(n-i+1)
  # compute the both(side=2)/lower(side=1) tails quantiles for t distribution
  if(side == 2){
    q=qmvt(1-c[1], tail = c("both.tails"), df = df)$quantile
  }else{
    q=qmvt(1-c[1], tail = c("lower.tail"), df = df)$quantile
  }

  for (j in n:i){
    temp=0
    for (m in n:i)
      if(m!=j){
        corrmatrix=matrix(c(1,CorrMatrix[j,m],CorrMatrix[j,m],1),2,2)
        if(side == 2){
          jointprob=2*pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)+
          2*pmvt(lower=c(q,-Inf), upper=c(Inf,-q), df=df, corr=corrmatrix)
        }else{
          jointprob=pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
        }
        temp=temp+jointprob
      }
      if(KNgamma[i]<temp) {KNgamma[i]=temp; MAXindex[i]=j}
  }

  CPKN[i]=(alpha+KNgamma[i])/(n-i+1)
}

return(CPKN)
```r
### Critical Values for Procedure III with Kounian's Inequality ####
### Under arbitrary dependence ####

KNCP=function(n, alpha, CorrMatrix, df, side){
  CPKN <- numeric(n)
  CPKN[n] <- alpha
  KNgamma=numeric(n-1)
  MAXindex=numeric(n-1)  # record the index that has the maximum KN gamma term in each step

  ## Critical Points

  for(i in (n-1):1){  # i is the index of P-value to be adjusted

    # compute the i-th Gamma terms in the Kounia's Inequality

    c=numeric(100000)
    c[i]=alpha

    # compute the both(side=2)/lower(side=1) tails quantiles for t distribution
    if(side == 2){
      q=qmvtn(1-c[i], tail = c("both.tails"), df = df)$quantile
    }else{
      q=qmvtn(1-c[i], tail = c("lower.tail"), df = df)$quantile
    }

    for (j in n:i){
      temp=0
      for (m in n:i)
        if(m!=j){

```
corrmatrix=matrix(c(1,CorrMatrix[j,m],CorrMatrix[j,m],1),2,2)
if(side == 2){
    jointprob=2*pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)+
    2*pmvt(lower=c(q,-Inf), upper=c(Inf,-q), df=df, corr=corrmatrix)
}else{
    jointprob=pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
}
temp=temp+jointprob
}
if(KNgamma[i]<temp) {KNgamma[i]=temp; MAXindex[i]=j}
}
c[2]=(alpha+KNgamma[i])/(n-i+1)
KNgamma[i]=0  # clear the memory of DSgamma

# New algorithm to update the critical values
k=2
while (c[k-1]-c[k]>.0000000001 & k<10000){
    # compute the both(side=2)/lower(side=1) tails quantiles for t distribution
    if(side == 2){
        q=qmvt(1-c[k], tail = c("both.tails"), df = df)$quantile
    }else{
        q=qmvt(1-c[k], tail = c("lower.tail"), df = df)$quantile
    }
    j=MAXindex[i]
    temp=0
    for (m in n:i)
        if(m!=j){
            corrmatrix=matrix(c(1,CorrMatrix[j,m],CorrMatrix[j,m],1),2,2)
            if(side == 2){
                jointprob=2*pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)+
                2*pmvt(lower=c(q,-Inf), upper=c(Inf,-q), df=df, corr=corrmatrix)
            }else{
                jointprob=pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
            }
            temp=temp+jointprob
        }
    KNgamma[i]=temp
}
\[ c[k+1] = \frac{(\alpha + KN_{\gamma[i]})}{(n-i+1)} \]

\[ KN_{\gamma[i]} = 0 \quad \# \text{clear the memory of } DS_{\gamma} \]

\[ k = k + 1 \]

\}

\[ CP_{KN}[i] = \min(c[c\neq 0]) \]

\}

\[ return(CP_{KN}) \]

\}

\[ \text{KNCP}(n, \alpha, CorrMatrix, df, side=1) \]

*### Critical Values for Procedure III with Hunter’s Inequality ###*

*### Under arbitrary dependence ###*

\[ \text{HTCP} = \text{function}(n, \alpha, CorrMatrix, df, side) \{
\]

\[ \text{CPHT} \leftarrow \text{numeric}(n) \]

\[ \text{CPHT}[n] \leftarrow \alpha \]

\[ \text{HTgamma} \leftarrow \text{numeric}(n-1) \]

*## Critical Points##*

\[ \text{for}(i \in (n-1):1) \{ \ # i \text{ the index of } P\text{-value to be adjusted} \]

\[ \# \text{compute the } i\text{th Gamma terms in the Kounia’s Inequality} \]

\[ c = \text{numeric}(100000) \]
c[1]=alpha
# compute the both(side=2)/lower(side=1) tails quantiles for t distribution
if(side == 2){
    dist_matrix = abs(1/CorrMatrix)
    q=qmvtnorm(1-c[1], tail = c("both.tails"), df = df)$quantile
}else{
    dist_matrix = -CorrMatrix
    q=qmvtnorm(1-c[1], tail = c("lower.tail"), df = df)$quantile
}

# compute the distance matrix for all n-i+1 vertex
M <- mst(dist_matrix[i:n,i:n])  # find the Minimum Spanning Trees based on distance Matrix

index=1
temp=0
for(m in i:n){
    for(j in i:n){
        if(index %in% which(M==1)){
            corrmatrix=matrix(c(1,CorrMatrix[j,m],CorrMatrix[j,m],1),2,2)
            if(side == 2){
                jointprob=2*pmvtnorm(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)+
                2*pmvtnorm(lower=c(q,-Inf), upper=c(Inf,-q), df=df, corr=corrmatrix)
            }else{
                jointprob=pmvtnorm(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
            }
        }
        temp=temp+jointprob
    }
    index=index+1
}

HTgamma[i]=temp/2
c[2]=(alpha+HTgamma[i])/(n-i+1)
HTgamma[i]=0  # clear the memory of HTgamma

# New algorithm to update the critical values
k=2
while (c[k-1]-c[k]>0.0000000001 & k<10000){

    # compute the both(side=2)/lower(side=1) tails quantiles for t distribution
if(side == 2){
    q=qmvtnorm(1-c[k], tail = c("both.tails"), df = df)$quantile
}else{
    q=qmvtnorm(1-c[k], tail = c("lower.tail"), df = df)$quantile
}

index=1
temp=0
for(m in i:n){
    for(j in i:n){
        if(index %in% which(M==1)){
            corrmatrix=matrix(c(1,CorrMatrix[j,m],CorrMatrix[j,m],1),2,2)
            if(side == 2){
                jointprob=2*pmvtnorm(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)+
                2*pmvtnorm(lower=c(q,-Inf), upper=c(Inf,-q), df=df, corr=corrmatrix)
            }else{
                jointprob=pmvtnorm(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
            }
            temp=temp+jointprob
        }else{
            index=index+1
        }
    }
    HTgamma[i]=temp/2
    c[k+i]=(alpha+HTgamma[i])/(n-i+1)
    HTgamma[i]=0  # clear the memory of HTgamma
    k=k+1
}

CPHT[i]=min(c[c!=0])
}

return(CPHT)
alpha = 0.05
df=16
r=0.5
CorrMatrix = matrix(c(1,r,r,r,0,
                       r,1,0,-r,r,
                       r,0,1,-r,r,
                       r,r,-r,1,0,-r,
                       r,-r,0,1,-r,
                       0,r,r,-r,1),n,n)

## side = 1 implies one-sided tests
## side = 2 implies two-sided tests
side = 2

round(KNCP0(n, alpha, CorrMatrix, df, side), digits=5)
round(KNCP(n, alpha, CorrMatrix, df, side), digits=5)
round(HTCP(n, alpha, CorrMatrix, df, side), digits=5)

2  R Codes for Chapter 4

Listing 5: Figure 4.1 4.2 and 4.3

library(multtest)
library(mutoss)
library(mvtnorm)
# library ( combinat )

################ New Critical Values of Procedure IV  ################

########################################################################
### Upper One Sided t Distribution with degree v ###
### Part I of the Inequality ###

CritiPt_k1=function(p,m,k,alpha,crr,df){
c=numeric(100000)
c[1]=alpha

# compute the both tails quantiles for t distribution
q=qmvt(1-c[1], tail = c("lower.tail"), df = df)$quantile

# compute the joint t distribution with correlation rho
corrmatrix=matrix(c(1,corr,corr,1),2,2)
jointprob=pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
c[2]=(k*alpha+(n+k-p-1)*jointprob)/(n+2*k-p-1)

## New algorithm to update the critical values
i=1
while (c[i]-c[i+1]>.00000000000001 & i<10000){
    i=i+1
    # compute the both tails quantiles for t distribution
    q=qmvt(1-c[i], tail = c("lower.tail"), df = df)$quantile
    # compute the joint t distribution with correlation rho
    jointprob=pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
c[i+1]=(k*alpha+(n+k-p-1)*jointprob)/(n+2*k-p-1)
}

return(min(c[c!=0]))

}

### Part II of the Inequality (t TESTS)####

CritiPt_k2=function(p,n,k,alpha,corr,df){

# compute the joint t distribution with correlation rho
corrmatrix=matrix(c(1,corr,corr,1),2,2)
f <- function (q){
    (n+2*k-p-1)*prmv(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)/(k*(k-1))-
    alpha
}
q <- uniroot(f, c(0, 100), tol = 0.0001)$root
c <- 1-pt(q, df=df)
### Upper One Sided Z Tests ####
### Part I of the Inequality ####

```r
CritPt_k1_z <- function (p,n,k,alpha ,corr ,df) {
c <- numeric (100000)
c[1]= alpha
# compute the both tails quantiles for t distribution
q= qnorm (1-c[1])
# compute the joint t distribution with correlation rho
corrmatrix = matrix (c(1, corr ,corr ,1) ,2 ,2)
jointprob = pmvnorm ( lower =c(q,q), upper =Inf , corr = corrmatrix )
c [2]=( k* alpha +(n+k-p -1)* jointprob )/(n+2*k-p -1)
i=1
while (c[i]-c[i +1] >.000000000001 & i <10000) {
i=i+1
# compute the both tails quantiles for t distribution
q= qnorm (1-c[i])
# compute the joint t distribution with correlation rho
jointprob = pmvnorm ( lower =c(q,q), upper =Inf , corr = corrmatrix )
c[i+1]= ( k* alpha +(n+k-p -1)* jointprob )/(n+2*k-p -1)
}
return ( min (c[c!= 0]))
}
```

### Part II of the Inequality (Z TESTS)####

```r
CritPt_k2_z <- function (p,n,k,alpha ,corr ,df){
# compute the joint t distribution with correlation rho
corrmatrix = matrix (c(1, corr ,corr ,1) ,2 ,2)
f <- function (q){
(n-p+k-1)* (n-p +1+k -1)* pmvnorm ( lower =c(q,q), upper =Inf , corr = corrmatrix )/(k*(k
-1) ) - alpha
}
q <- uniroot (f, c(0, 100) , tol = 0.0001) $ root

c <- 1-pt(q, df=df)

return (c)
}

##########################################################
#
# CP under Independence (Binomial Model Based Procedure) ####
#
##########################################################

CritiPt_k3=function(p,n,k,alpha){

f <- function (prob){

  1-pbinom(k-1, size=n-p+k, prob, lower.tail = TRUE)-alpha
}

q <- uniroot(f, c(0, 1), tol = 0.000001)$root
return(q)
}

CritiPt_k3(1,100,10,0.05)

##########################################################
#
# Plots Analysis
#
##########################################################

alpha=0.05
n=20
k=2
p=seq(1:n)[k:n]
corr=c(0, 0.2, 0.35, 0.5, 0.6, 0.75, 0.9, 0.99)
df=5000
CP0=CP1=CP2=CP3=numeric(n)

##########################################################
#
# Plots Analysis for Critical Values
#
##########################################################

*** Figure 4.1 ***

pdf(file='C:\Users\Shihai\Desktop\MTP\Critical Points(k-FWER)Comparison.pdf')
m <- matrix(c(1,1,1,2,3,4,5,6,7,8,9),nrow = 3,ncol = 4,byrow = TRUE)

layout(mat = m, heights = c(0.55, 4, 4))

par(oma = c(4,4,0,0) + 0.1)  ## giving size of the outer margins in lines of text

par(mai=c(0,0,0,0))
plot_colors <- c("red","blue")
plot.new()
legend(x="center", ncol=2,legend=c("LR","Procedure IV"),lty=c(2,1), lwd=c(1.5,1.5), col=plot_colors, seg.len=3.5)

#par (mar = c(0,0,2.7,0))  ## giving size of each individual plot margins in lines of text
par(mar = c(0,2.2,2.7,0))  ## giving size of each individual plot margins in lines of text

for (i in 1:length(corr)){

CP1=sapply(p,CritiPt_k1,n,k,alpha,corr=corr[i],df);CP1=round(CP1, digits = 5)
CP2=sapply(p,CritiPt_k2,n,k,alpha,corr=corr[i],df);CP2=round(CP2, digits = 5)
#CP3 = sapply (p , CritiPt _k3 ,n ,k , alpha ); CP3 = round (CP3 , digits = 5)
CP0=k*alpha/(n+k-p);CP0=round(CP0, digits = 5)

NewCP=mapply(max,mapply(max,CP1,CP2),CP0)

LR=c(rep(CP0[1],k-1),CP0)
IV=c(rep(NewCP[1],k-1),NewCP)

ylim=max(IV,LR)*1.1

plot(i:n,IV,type='l',lty=1,,col='blue',ylim=c(0,ylim),lwd=1.5,main=bquote(paste(rho,'=',.(corr[i])),xaxt = "n")
points(i:n,LR,type='l',lty=2,col='red',lwd=1.5)
abline(h=alpha,lty=3,lwd=1.5)

axis(side = 1,labels = (i == 6 || i == 8))

```r
axis(side = 2, labels = (i == 1 || i == 5))

title(xlab = 'Index',
      ylab = 'Critical Values',
      outer = TRUE, line = 2.5, cex.lab=1.5)

dev.off()

################## Plots Analysis for Critical Values RATIO ##################

### Figure 4.2 ###

pdf(file='C:\Users\Shihai\Desktop\MTP\Critical\Points\Ratio\(k\)-FWER\Comparison.pdf')

m <- matrix(c(1,1,1,1,2,3,4,5,6,7,8,9),nrow = 3,ncol = 4,byrow = TRUE)

layout(mat = m,heights = c(0.55, 4, 4))

par(oma = c(4,4,0,0) + 0.1)

## giving size of the outer margins in lines of text

par(mai=c(0,2.2,2.7,0))

plot_colors <- c("blue")

plot.new()

legend(x="center", ncol=1,legend=c("Ratio of Procedure IV to LR"),lty=c(1), lwd=1.5, col=
       plot_colors, seg.len=3.5)

par(mar = c(0,2.2,2.7,0))

for (i in 1:length(corr)){

  CP1=sapply(p,CritiPt_k1,n,k,alpha,corr=corr[i],df);CP1=round(CP1, digits = 5)
  CP2=sapply(p,CritiPt_k2,n,k,alpha,corr=corr[i],df);CP2=round(CP2, digits = 5)
  #CP3=sapply(p,CritiPt_k3,n,k,alpha);CP3=round(CP3, digits = 5)
  CP0=k*alpha/(n+k-p);CP0=round(CP0, digits = 5)
```

NewCP = mapply(max, mapply(max, CP1, CP2), CP0)

LR = c(rep(CP0[1], k-1), CP0)
IV = c(rep(NewCP[1], k-1), NewCP)

ylim = max(IV/LR)*1.1

# plot(1:n, IV, type='l', lty=1, col='blue', ylim=c(0.9, 0.3), lwd=1.5, main=bquote(paste(rho, '\text{=',.(corr[i])}')))
# points(1:n, IV, type='l', lty=4, col=4, lwd=2.5)
plot(1:n, IV/LR, type='l', lty=1, col='blue', ylim=c(0.9, ylim), lwd=1.5, main=bquote(paste(rho, '\text{=',.(corr[i])}')))
abline(h=1, lty=3, lwd=1.5)

axis(side = 1, labels = (i == 6 || i == 8))
axis(side = 2, labels = (i == 1 || i == 5))

title(xlab = 'Index',
      ylab = 'Ratio of Critical Values',
      outer = TRUE, line = 2.5, cex.lab=1.5)

dev.off()

########## Plots Analysis for Critical Values Under Independence ###########

### Figure 4.3 ###

n = 20
k = 1

p = seq(1:n)[k:n]
corr = 0
df = 5000
CP0 = CP1 = CP2 = CP3 = numeric(n)
```r
m <- matrix(c(1,1,1,1,2,3,4,5,6,7,8,9,10,11,12,13), nrow = 4, ncol = 4, byrow = TRUE)
layout(mat = m, heights = c(0.55, 3.5, 3.5, 3.5))
par(oma = c(4, 4, 0, 0) + 0.1)  ## giving size of the outer margins in lines of text

par(mai = c(0, 0, 2.5, 0))  ## giving size of each individual plot margins in lines of text

for (i in 1:12) {
  if (i == 5) { n = 50; k = 1 }
  if (i == 9) { n = 100; k = 1 }
  p = seq(1:n)[k:n]
  CP0 = CP1 = CP2 = CP3 = numeric(n)

  CP1 = sapply(p, CritiPt_k_1_z, n, k, alpha, corr = corr, df); CP1 = round(CP1, digits = 5)
  if (k != 1) {
    CP2 = sapply(p, CritiPt_k_2_z, n, k, alpha, corr = corr, df); CP2 = round(CP2, digits = 5)
  }
  CP3 = sapply(p, CritiPt_k_3, n, k, alpha); CP3 = round(CP3, digits = 5)
  CP0 = k * alpha / (n + k - p); CP0 = round(CP0, digits = 5)

  NewCP = mapply(max, mapply(max, CP1, CP2), CP0)
  if (k != 1) {
    LR = c(rep(CP0[1], k - 1), CP0)
    IV = c(rep(NewCP[1], k - 1), NewCP)
    BIN = c(rep(CP3[1], k - 1), CP3)
  }
```

} else {
    LR = CP0; IV = CP1; BIN = CP3;
}

ylim = max(BIN) * 1.1

if (i < 5) {
    plot(1:n, BIN, type = 'l', lty = 1, col = 'blue', ylim = c(0, 0.58), lwd = 1.5, main = "k", xaxt = "n", yaxt = "n")
} else {
    plot(1:n, BIN, type = 'l', lty = 1, col = 'blue', ylim = c(0, 0.58), lwd = 1.5, xaxt = "n", yaxt = "n")
}
points(1:n, IV, type = 'l', lty = 2, col = 'red', lwd = 1.5)
points(1:n, LR, type = 'l', lty = 4, col = 'black', lwd = 1.5)
abline(h = alpha, lwd = 0.5, lty = 3)
axis(side = 1, labels = (i == 1 || i == 3 || i == 6 || i == 8 || i == 9 || i == 11))
axis(side = 2, labels = (i == 1 || i == 5 || i == 9))

k = k + 1
}

title(xlab = 'Index', ylab = 'Critical Values',
outer = TRUE, line = 2.5, cex.lab = 1.5)

dev.off()

---

Listing 6: Figure 4.4 and 4.5

library(Rmpi)

# install.packages('snow_0.3-3.tar.gz', '/nfs/02/bgs0314/Rlibs', repos=NULL)

library(snow)

# load a package on each cluster node with
# clusterEvalQ(cl, library( Matrix))

# install.packages('BiocGenerics_0.4.0.tar.gz', '/nfs/02/bgs0314/Rlibs', repos=NULL)
library(mutoss)
library(mvtnorm)

pdf(file='FWERpowerFAST_kFWER_new.pdf')

require(MASS)

# ###########################################################
# ############# New Critical Values (k-FWER)  ################
# ########## Upper One Sided t Distribution with degree v  ###########
# ################################################################

# Part I

CritiPt_k11=function(p,n,k,alpha,corr,df){

c=numeric(100000)
c[1]=alpha

# compute the both tails quantiles for t distribution
q=qmvt(1-c[1], tail = c("lower.tail"), df = df)$quantile

# compute the joint t distribution with correlation rho
corrmatrix=matrix(c(1,corr,corr,1),2,2)
jointprob=pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
c[2]=(k*alpha+(n+k-p-1)*jointprob)/(n+2*k-p-1)

# New algorithm to update the critical values
i=1
while (c[i]-c[i+1]>.00000000000001 & i<10000){

}
i=i+1
# compute the both tails quantiles for t distribution
q=qmvt(1-c[i], tail = c("lower.tail"), df = df)$quantile
# compute the joint t distribution with correlation rho
jointprob=pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)
c[i+1]=(k*alpha+(n+k-p-1)*jointprob)/(n+2*k-p-1)
}
return(min(c[c!=0]))

######## Part II #######

CritiPt_k12=function(p,n,k,alpha,corr,df){

# compute the joint t distribution with correlation rho
corrmatrix=matrix(c(1,corr,corr,1),2,2)
f <- function (q){
  (n-p+k-1)*(n-p+1+k-1)*pmvt(lower=c(q,q), upper=Inf, df=df, corr=corrmatrix)/(k*(k-1))-
  alpha
}
q <- uniroot(f, c(0, 100), tol = 0.0001)$root
c <- 1-pt(q, df=df)

return(c)
}

################################ LR Procedure ################################

LR=function(Pvalues,LRCP,k){

# CorMatrix: the matrix storing the correlation between test statistics in the ascending order
# the (i,j) element corresponds to the correlation
# between the (n-i+1)th and (n-j+1)th largest P-values
# df: the freedom of t distribution

n=length(Pvalues)

OriginalIndex=sort.list(Pvalues)
Pvalues=sort(Pvalues)

# do the LR procedure

FirstAcception=ifelse(length(which(Pvalues[k:n]>LRCP[k:n]))>0,min(which(Pvalues[k:n]>LRCP[k:n])+(k-1)),n+1)
Reject=c(rep('TRUE',FirstAcception-1),rep('FALSE',n-FirstAcception+1))

# restore the order of P values

temp1 <- temp2 <- numeric(n)
for(i in 1:n) {
  temp1[OriginalIndex[i]]=Pvalues[i]
  temp2[OriginalIndex[i]]=Reject[i]
}
Pvalues=temp1
Reject=temp2

result=Reject
return(result)

} # end of RL function

#################### New k-FWER MTP (Procedure IV) ####################

NewMTP=function(Pvalues,NewCP,k){

  # CorMatrix: the matrix storing the correlation between test statistics in the ascending order
  # the (i,j) element corresponds to the correlation
  # between the (n-i+1)th and (n-j+1)th largest P-values


# df: the freedom of t distribution

n=length(Pvalues)

OriginalIndex=sort.list(Pvalues)
Pvalues=sort(Pvalues)

# do the New MTP

FirstAcception=ifelse(length(which(Pvalues[k:n]>NewCP[k:n]))>0,min(which(Pvalues[k:n]>NewCP[k:n]+(k-1)),n+1)
Reject=c(rep('TRUE',FirstAcception-1),rep('FALSE',n-FirstAcception+1))

# restore the order of P values

temp1 <- temp2 <- numeric(n)
for(i in 1:n) {
  temp1[OriginalIndex[i]]=Pvalues[i]
  temp2[OriginalIndex[i]]=Reject[i]
}
Pvalues=temp1
Reject=temp2

result=Reject
return(result)

} # end of NewMTP function

########## FWER and Power Analysis #############

########## Step I: Simulate test statistics ####

########## n=8, t test: miu=0 vs miu=delta, df=16 #########

alpha=0.05
corr=c(0,.3,.7,.9,.95,.99)
n=8
N=200        # number of replications
df=30
delta0=2       # set mean values for the alternative hypotheses

##### The function of FWER and POWER analysis (k-FWER) #######

FWER_k=function(corr,df,alpha,n,k,delta0,N,CritiPt_k11,CritiPt_k12,LR,NewMTP){

  CRP_lr <- FRP_lr <- CRP_n <- FRP_n <- numeric(N)  ## CR: Correctly Rejection proportion
  ## FRP: Falsely Rejection proportion
  FWER_lr <- POWER_lr <- FWER_n <- POWER_n <- numeric(n)

  corrmatrix=diag(n)

  for(i in 1:n){
    for (j in 1:n){
      if(corrmatrix[i,j]!=1) corrmatrix[i,j]=corr  # corr is the correlation between any
                                                     # two test statistics
      if(j==j+1) i=i+1
    }
  }

  ## Calculate the critical values using correlation information
  # (I) LRCP & NewCP
  LRCP <- NewCP <- numeric(n)
  LRCP[n] <- NewCP[n] <- alpha

  for(i in (n-1):k){  # i the index of P-value to be adjusted (k-FWER)

    # find the minimum correlation from the last i test statistics
    MINcorr=1
    for (j in n:(i+1))
      for (t in (j-1):i)
        MINcorr=min(corrmatrix[j,t],MINcorr)

    LRCP[i]=k*alpha/(n+k-i)  # One Sided test Critical values
    NewCP[i]=max(CritiPt_k11(i,n,k,alpha,MINcorr,df),CritiPt_k12(i,n,k,alpha,MINcorr,df))  # One Sided test Critical values
# LRCP[i]=k*alpha/(n*k-i)  # Two Sided test Critical values
# NewCP[i]=CritPt_k2(i,n,k,alpha,MINcorr,df)  # One Sided test Critical values

}
LRCP[1:(k-1)]=rep(LRCP[k],(k-1))
NewCP[1:(k-1)]=rep(NewCP[k],(k-1))

##### the loop to compute the POWER and FWER

for (i in 1:n){  # (i-1) the number of false hypotheses
  delta=c(rep(0,n-(i-1)),rep(delta0,i-1))
s=rmvt(N, sigma = corrmatrix, df = df, delta=delta)

  ## compute N sets of P-values under null hypotheses
  # Pvalues = ifelse(s>0,2*pt(s, df=df,lower.tail = FALSE),2*pt(s, df=df,lower.tail = TRUE))  # Two Sided test
  Pvalues=pt(s, df=df,lower.tail = FALSE)
  PVALUES=list(Pvalues[1,])  # store in lists
  for(j in 2:N) PVALUES=c(PVALUES,list(Pvalues[j,]))

  ## LR procedure
  REJ=t(sapply(PVALUES,LR,LRCP,k))
  for (j in 1:N){
    RejIndex=which(REJ[j,]==TRUE)
    if(length(delta[delta==delta0])==0){
      CRP_lr[j]=1}else{
      CRP_lr[j]=sum(ifelse(delta[RejIndex]==delta0,1,0))/length(delta[delta==delta0])
    }
    FRP_lr[j]=sum(ifelse(delta[RejIndex]==0,1,0))/length(delta[delta==0])
  }
  FWER_lr[i]=mean(ifelse((FRP_lr[length(delta==0)])>= k,1,0))
  POWER_lr[i]=mean(CRP_lr)

  ## New Procedure (IV)
NewCP=mapply(max,LRCP,NewCP)

REJ=t(sapply(PVALUES,NewMTP,NewCP,k))
for (j in 1:N){
    RejIndex=which(REJ[j,]==TRUE)
    if(length(delta[delta==delta0])==0){
        CRP_n[j]=1}else{
        CRP_n[j]=sum(ifelse(delta[RejIndex]==delta0,1,0))/length(delta[delta==delta0])
    }
    FRP_n[j]=sum(ifelse(delta[RejIndex]==0,1,0))/length(delta[delta==0])
}
FWER_n[i]=mean(ifelse((FRP_n*length(delta[delta==0]))>= k,1,0))
POWER_n[i]=mean(CRP_n)
}
return(c(list(FWER_lr),list(FWER_n),list(POWER_lr),list(POWER_n)))
}

# ###########################################################################################

Beginning=Sys.time()

# Test snow on the OSC cluster.

# First, get the cluster info.

cl <- makeCluster()

# load a package on each cluster node with

clusterEvalQ(cl, library(mutoss))
clusterEvalQ(cl, library(mvtnorm))

# Now generate some random variables on all of the clusters. Note,
# because we haven’t set a different seed for all of the processes, you
# may get back duplicates. GenMatch and rgenoud takes care of this
# for you, but other R libraries may not. See the snow documentation
alpha=0.05
corr=c(0, 0.2, 0.35, 0.5, 0.6, 0.75, 0.9, 0.99)
s=15    ## number of divisions of total samples
Corr=rep(corr,s)
n=20    ## number of hypotheses
k=2     ## number of tolerated false rejections
N=25000    ## number of replications
df=50
delta0=3.57

result=parSapply(cl, Corr, FWER_k, df=df, alpha=alpha,n=n, k=k, delta0=delta0,N=N,CritiPt_k11,CritiPt_k12,LR,NewMTP)

# Stop cluster
stopCluster(cl)

Sys.time()-Beginning

############ Combine all s sets of samples together and store in the first length(corr) columns in the 'result'############

length(result[[1,2]])

result[[1,2]]

for(j in 1:4)    # 6 indicate 2 FWER and 2 POWER
    for(t in 1:length(corr)){
        for (i in 1:(s-1))
            result[[j,t]]=result[[j,t]]+result[[j,(t+length(corr)*i)]]
        result[[j,t]]=result[[j,t]]/s
length(result[[1,2]])

result[[1,2]]

# Plots Analysis

m <- matrix(c(1,1,1,2,3,4,5,6,7,8,9),nrow = 3,ncol = 4,byrow = TRUE)

layout(mat = m,heights = c(0.55, 4, 4))

par(oma = c(4,4,0,0) + 0.1)  ## giving size of the outer margins in lines of text

# Plots Analysis

### FWER ###

par(mai=c(0,0,0,0))
plot_colors <- c("blue","red")
plot.new()
legend(x="center", ncol=2,legend=c("LR","Procedure IV"),lty=c(1,2), lwd=c(1.7,1.7), col=plot_colors, seg.len=3.5)

par(mar = c(0,0,2.7,0))  ## giving size of each individual plot margins in lines of text

for (i in 1:length(corr)){
    plot(0:(n-k),result[[1,i]][1:(n-k+1)],type='l',col='blue',lwd=1.5,ylim=c(0,0.055),main= bquote(paste(rho,'=',.(corr[i]))) ,xaxt = "n", yaxt='n')
    points(0:(n-k), result[[2,i]][1:(n-k+1)],type='l',col='red',lwd=1.5,lty='dashed')
    axis(side = 1,labels = (i == 6 || i == 8))
    axis(side = 2,labels = (i == 1 || i == 5))
}
```r
## POWER ##

par(mai=c(0,0,0,0))
plot_colors <- c("blue","red")
plot.new()
legend(x="center", ncol=2,legend=c("LR","Procedure IV"),lty=c(1,2), lwd=c(1.7,1.7), col=plot_colors, seg.len=3.5)

par(mar = c(0,0,2.7,0))

for (i in 1:length(corr)){
  plot(1:(n-1),result[[3,i]][2:n],type='l',col='blue',lwd=1.5,ylim=c(0,1),main=bquote(paste(rho,'=',.(corr[i]))),xaxt = "n", yaxt='n')
  points(1:(n-1), result[[4,i]][2:n],type='l',col='red',lwd=1.5,lty='dashed')
  axis(side = 1,labels = (i == 6 || i == 8))
  axis(side = 2,labels = (i == 1 || i == 5))
}

title(xlab = '# of False Null',
      ylab = '3-FWER',
      outer = TRUE, line = 3, cex.lab=1.4)

dev.off()
```

### R Codes for Chapter 5 ###

Listing 7: Figure 5.1

---

3 R Codes for Chapter 5
\( N = 200000 \)

\( n = 50 \)  

\( \mu_1 = -0.5 \)  

\# \mu_1 \) can not be too big to have the sampling distribution locating around 0

\# affect the location of the sampling distribution. \( \mu_1 \) is the further the distribution is from 0

\# the bigger, the bigger the power is

\# set \( \mu_1 = 0 \), the below procedure can be used to calculate the Type I error

\# \alpha = 0.05

\# \( df = 10 \)

\( ncp = 3 \)  

\# the smaller, the more left the distribution is

\# simulate the population under alternative hypothesis where our sample is generated

\( x = \text{rt}(N, df=df, ncp=ncp); \)

\[ \text{MEDIAN} = \text{median}(x) \]

\[ x = x - \text{median}(x) + \mu_1 \]

\[ c(\text{median}(x), \text{mean}(x), \text{var}(x)) \]

\# calculate the \( t \) statistics and draw its sampling distribution based on the population above

\( m = 20000 \)

\( \text{nt_stat} = \text{numeric}(m) \)

\text{for} (i \text{ in} 1:m) { 
    \( r = \text{sample}(x, n, \text{replace} = \text{TRUE}, \text{prob} = \text{NULL}) \)
    \( \text{nt_stat}[i] = \text{mean}(r) * \sqrt{n} / \text{sd}(r) \)
}

\text{par(mfrow=c(2,1))}

\text{hist}(x, \text{breaks}=200)

\text{hist(nt_stat, \text{breaks}=200)}

\text{par(mfrow=c(1,1))}
### calculate the power w.r.t. 3 different test methods.
### based on the sampling distribution of the t statistics under alternative hypothesis

\[
Q_Z = qnorm(\alpha); Q_Z
\]

\[
Q_T = qt(\alpha, df=n-1); Q_T
\]

\[
pwr_t = \text{mean}(\text{ifelse}(nt\_stat < Q_T, 1, 0))
pwr_z = \text{mean}(\text{ifelse}(nt\_stat < Q_Z, 1, 0))
#p1 = 1-\text{pt(MEDIAN-\mu_1, df=\text{df}, ncp = ncp); p1}
\]

\[
p2 = \text{mean}(\text{ifelse}(x > 0, 1, 0)); p2
\]

\[
QS = \text{qbinom}(\alpha, n, 0.5, \text{lower.tail = TRUE, log.p = FALSE}) - 1; QS; \ #The quantile is defined as the smallest value z such that F(z) >= p
pwr_s = \text{pbinom}(QS, n, p2, \text{lower.tail = TRUE, log.p = FALSE});
\]

\[
c(pwr_s, pwr_z, pwr_t)
\]

### set \( \mu_1 = 0 \), the above procedure can be used to calculate the Type I error ###

---

### power of Bonferroni ###

library(mutoss)
library(mvtnorm)
library(BSDA)
library(sn)
### PVALUE function

N=200000
n=50  # n can not be too big to have the sampling distribution locating around 0
mu1 = -0.5  # affect the location of the sampling distribution. The bigger, the further the
distribution is from 0
    # the bigger, the bigger the power is
### set mu1 = 0, the below procedure can be used to calculate the Type I error
###
alpha=0.05
df = 10
ncp = 3  # the smaller, the more left the distribution is

# simulate the population under alternative hypothesis where our sample is generated
x=rt(N,df=df, ncp=ncp);

PVALUE=function(n,mu1){
    #z=rcauchy(n,location=mu1,scale=1)  # simulate a random sample from a continuous
distribtuion
    #z=rnorm(n,mean=mu1,sd=5)
    # z=rt(n,df=5)+mu1
    #a=rnorm(n,0,2);b=rnorm(n,0,10);z=c(a[a<0],b[b>0])+mu1
    #z=r(20000,df1=10,df=1);
    x=x-median(x)+mu1
    x=sample(x, n, replace = TRUE, prob = NULL)

    signtest=SIGN.test(x, md = 0,alternative="less",conf.level = 1)  # set level=1 to solve the
    bug in SIGN.test
    pS=signtest$p.value

    ztest=z.test(x, alternative = "less", mu = 0, sigma.x=sd(x),conf.level = 0.95)
    # ztest=wilcox.test(z, alternative = "less", mu = 0, exact=T,conf.level = 0.95)
    pZ=ztest$p.value

    ttest=t.test(x, alternative = "less", mu = 0, conf.level = 0.95)
}
pT=ttest$p.value

return(c(pS,pZ,pT));
}

# Calculate the power

POWER=function(N,m,n,mu1,alpha,p){

  POWER_s=POWER_z=POWER_t=numeric(m)  # 1 for intersection power; 2 for union power
  FWER_s=FWER_z=FWER_t=numeric(m)

  for (k in 1:m){  # k indicates the number of alternative hypothesis
    pcount1=pcount2=pcount3=0  # counter for correct rejections
    fcount1=fcount2=fcount3=0  # counter for false rejections

    for (j in 1:N){

      PS=PZ=PT=0

      for(i in 1:m){

        mu=ifelse(i<=k,mu1,0)

        PV=PVALUE(n,mu1=mu)

        PS=c(PS,PV[1])
        PZ=c(PZ,PV[2])
        PT=c(PT,PV[3])

      }  # end of 3rd loop

      PS=PS[2:(m+1)];PZ=PZ[2:(m+1)];PT=PT[2:(m+1)];cbind(PS,PZ,PT);
      rej_s=bonferroni(PS, alpha, silent=T)$rejected
      rej_z=bonferroni(PZ, alpha, silent=T)$rejected
      rej_t=bonferroni(PT, alpha, silent=T)$rejected

      if (p==1){  # 1 union power
        #
      }
    }
  }
}

# End of function

RejIndex_s=which(rej_s==TRUE)
if (length(RejIndex_s[RejIndex_s<=k])>0) # power
  pcount1=pcount1+1
if (length(RejIndex_s[RejIndex_s>k])>0) # fwer
  fcount1=fcount1+1

RejIndex_z=which(rej_z==TRUE)
if (length(RejIndex_z[RejIndex_z<=k])>0) # power
  pcount2=pcount2+1
if (length(RejIndex_z[RejIndex_z>k])>0) # fwer
  fcount2=fcount2+1

RejIndex_t=which(rej_t==TRUE)
if (length(RejIndex_t[RejIndex_t<=k])>0) # power
  pcount3=pcount3+1
if (length(RejIndex_t[RejIndex_t>k])>0) # fwer
  fcount3=fcount3+1

} else if (p==2){
  ## 2 intersection power

RejIndex_s=which(rej_s==TRUE)
if (length(RejIndex_s[RejIndex_s<=k])==k) # power
  pcount1=pcount1+1
if (length(RejIndex_s[RejIndex_s>k])>0) # fwer
  fcount1=fcount1+1

RejIndex_z=which(rej_z==TRUE)
if (length(RejIndex_z[RejIndex_z<=k])==k) # power
  pcount2=pcount2+1
if (length(RejIndex_z[RejIndex_z>k])>0) # fwer
  fcount2=fcount2+1

RejIndex_t=which(rej_t==TRUE)
if (length(RejIndex_t[RejIndex_t<=k])==k) # power
  pcount3=pcount3+1
if (length(RejIndex_t[RejIndex_t>k])>0) # fwer
  fcount3=fcount3+1

} else { # 3 expectation power
RejIndex_s = which(rej_s == TRUE)
if (length(RejIndex_s[RejIndex_s <= k]) > 0) # power
    pcount1 = pcount1 + length(RejIndex_s[RejIndex_s <= k])/k
if (length(RejIndex_s[RejIndex_s > k]) > 0) # fwer
    fcount1 = fcount1 + 1

RejIndex_z = which(rej_z == TRUE)
if (length(RejIndex_z[RejIndex_z <= k]) > 0) # power
    pcount2 = pcount2 + length(RejIndex_z[RejIndex_z <= k])/k
if (length(RejIndex_z[RejIndex_z > k]) > 0) # fwer
    fcount2 = fcount2 + 1

RejIndex_t = which(rej_t == TRUE)
if (length(RejIndex_t[RejIndex_t <= k]) > 0) # power
    pcount3 = pcount3 + length(RejIndex_t[RejIndex_t <= k])/k
if (length(RejIndex_t[RejIndex_t > k]) > 0) # fwer
    fcount3 = fcount3 + 1

} # end of 2nd loop

POWER_s[k] = pcount1/N
POWER_z[k] = pcount2/N
POWER_t[k] = pcount3/N

FWER_s[k] = fcount1/N
FWER_z[k] = fcount2/N
FWER_t[k] = fcount3/N

} # end 1st loop

return(cbind(POWER_s,POWER_z,POWER_t,FWER_s,FWER_z,FWER_t))

) # end of POWER function

N = 20000  # replicate number
m = 5      # number of hypotheses
n = 50     # sample size
mu1=-0.5    # alternative is 'less than 0
alpha=0.05
p=1        # choose the power type

POWER1=POWER(N,m,n,mu1,alpha,p);cbind(K=seq(m),POWER1)

p=2        # choose the power type

POWER1=POWER(N,m,n,mu1,alpha,p);cbind(K=seq(m),POWER1)

POWER1_s=POWER1[,1]
POWER1_z=POWER1[,2]
POWER1_t=POWER1[,3]

plot(POWER1_s,col='red',type='l')
points(POWER1_z,col='yellow',type='l')
points(POWER1_t,col='blue',type='o')