12-2020

On Simes's Second Conjecture: An Extended Single-Step Simes Test Procedure for Multiple Testing

Matthew G. Hudson
Western Michigan University, matthewgushudson@gmail.com

Recommended Citation

Hudson, Matthew G., "On Simes's Second Conjecture: An Extended Single-Step Simes Test Procedure for Multiple Testing" (2020). *Dissertations*. 3681.
https://scholarworks.wmich.edu/dissertations/3681

ON SIMES'S SECOND CONJECTURE: AN EXTENDED SINGLE-STEP SIMES TEST PROCEDURE FOR MULTIPLE TESTING

by

Matthew G Hudson

A dissertation submitted to the Graduate College
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Statistics
Western Michigan University
December 2020

Doctoral Committee:

Joshua Naranjo, Ph.D., Chair
Joseph McKean, Ph.D.
Georgiana Fisher, Ph.D.
Dror Rom, Ph.D.

ON SIMES'S SECOND CONJECTURE: AN EXTENDED SINGLE-STEP SIMES TEST PROCEDURE FOR MULTIPLE TESTING

Matthew G Hudson, Ph.D.

Western Michigan University, 2020

One of the major concerns with multiple tests of significance is controlling the family wise error rate. Various methods have been developed to ensure that the false positive rate be maintained at some prespecified level. One of the most well know being the Bonferroni procedure. Simes presented an improved Bonferroni procedure for testing the global hypothesis that is more powerful and less conservative, especially with positively correlated tests. While Simes's procedure is more powerful, it does not allow for making inferences on the individual hypotheses. However, the Simes procedure has since become the foundation of many p-value based multiple testing procedures. Hochberg and Hommel are two examples of procedures that have extended the Simes procedure to make inferences on individual hypotheses. Though the procedures by Hochberg and Hommel are based on the Simes procedure, they are conservative as they may not be able to reject any of the individual hypotheses when Simes's test rejects the global null hypothesis. It is this disconnect between the global test and tests for the individual hypotheses where improvements may be made on the power to reject at least one individual hypothesis. Simes second conjecture was to reject the individual hypotheses $H_{(1)}$, ..., $H_{(j)}$, where j = $max\{j : P_{(j)} \leq j\alpha/m\}$. It can easily be shown that this method would not control the family wise error rate even for independent tests. An extended single-step Simes testing procedure is presented that rejects a subset of the hypotheses rejected in Simes's second conjecture. This new procedure rejects at least one hypothesis when the Simes global test rejects, making it the most powerful Simes based procedure for the rejection of at least one hypothesis. This procedure is shown to have strong control of the family wise error for three non-negatively correlated normals with the MTP_2 property. Further, simulation studies conducted show that the procedure may control the family wise error rate for as many as thirty or more hypotheses for non-negatively correlated normals with the MTP_2 property as well as non-negatively correlated chi-square and T test statistics.

Acknowledgments

I would like to thank my advisor Joshua Naranjo, Ph.D. and Dror Rom, Ph.D., committee member and long time friend and mentor, for encouraging me to complete the Ph.D. program as well as for their dedicated support and guidance throughout the dissertation process. Thank you to my wife, Amy, and children, Hunter and Logan, who put up with long hours and time away. Also, thank you to my wife for keeping me on track throughout. To my parents, Dave and Karen, who have supported me in everything I have ever done and given me the drive and confidence to take on any challenge. A special thanks to my father who pushed me more than anyone to complete my degree. This is for you.

Matthew G Hudson

Table of Contents

Table of Contents—Continued

List of Tables

List of Tables—Continued

List of Figures

List of Figures—Continued

Chapter 1

Introduction

1.1 Background and Motivation

One of the major concerns with multiple tests of significance is controlling the family wise error rate. Various methods have been developed to ensure that the false positive rate be maintained at some prespecified level. One of the most well know being the Bonferroni procedure. While this procedure is distribution free and easy to implement, it is known to be conservative especially with positively correlated tests.

Simes presented an improved Bonferroni method [Simes, 1986] for testing the global null hypothesis that all tests are true. In his paper, Simes proved that his test controls the type 1 error rate for independent tests. However, it does not hold for all conditions and distributions. Simes's first conjecture was that his method may hold for some dependent statistics with certain multivariate distributions. To that end, Simes provided simulations for normal and multiple chi-square distributions under independent and multiple levels of positively correlated test statistics. The simulation results suggest that the type 1 error rate may be controlled for these specific distributions. In comparison, the type 1 error rates between the Bonferroni and Simes global tests are similar when the test statistics are independent, but the Bonferroni test is considerably more conservative than the Simes test for highly positively correlated tests. That said, the Bonferroni test may be preferred when non-negatively correlated statistics cannot be assumed.

To address Simes's first conjecture, it has since been shown [Hochberg and Rom, 1995, Samuel-Cahn, 1996, Sarkar and Chang, 1997, Block et al., 2013] that, in general, the Simes global test has control of the one-sided type 1 error for positively dependent statistics, but may inflate the type 1 error for negatively dependent statistics. However, Hochberg and Rom [1995] gave an upper bound for the type 1 error for negatively correlated normal statistics.

While the Simes test is more powerful than the Bonferroni test for the global null hypothesis, it does not allow for making inferences on the individual hypotheses. However, the Simes global test has since become the foundation of many p-value based multiple testing procedures. Hochberg [1988] and Hommel [1988] are two examples of procedures that have extended the Simes global test to make inferences on individual hypotheses. Assuming the Simes global test has a type 1 error rate less than or equal to alpha, both procedures have strong control of the family wise error rate as defined, for example, by Hochberg and Tamhane [1987].

Though the procedures by Hochberg and Hommel are based on the Simes global test, they are conservative as they may not be able to reject any of the individual hypotheses when Simes's test rejects the global null hypothesis. It is this disconnect between the global test and tests for the individual hypotheses where improvements may be made on the power to reject at least one individual hypothesis.

In the discussion section of his paper, Simes asked the question of what statements about individual hypotheses may be made using his modified global test procedure. His second conjecture was that one possibility is to reject the individual hypotheses $H_{(1)}$, ..., $H_{(j)}$, where j = $max\{j : P_{(j)} \leq j\alpha/m\}$. However, it can easily be shown that this method would not control the family wise error rate even for independent tests. In this dissertation, an extended single-step Simes testing procedure is presented that rejects a subset of the hypotheses rejected in Simes's second conjecture. This new procedure rejects at least one hypothesis when the Simes global test rejects, making it the most powerful Simes based procedure for the rejection of at least one hypothesis. This procedure is shown to have strong control of the family wise error for three non-negatively correlated normals with the MTP_2 property.

Simulation studies conducted suggest that while the extended single-step Simes procedure does not control the type 1 error rate for any number of hypotheses, that it may control the type 1 error rate for as many as 30 or more hypotheses when the test statistics are non-negatively correlated normals with the MTP_2 property. Further simulations suggest that the procedure may also control the type 1 error rate for as many as 30 or more hypotheses for non-negatively correlated chi-square and t test statistics.

1.2 Concepts of Controlling for False Positives

Two of the more common concepts for the controlling of false positives are the family wise error rate (FWER) and the false discovery rate (FDR). The FWER is the probability of rejecting at least one true hypothesis, whereas the FDR is the expected proportion of true hypotheses rejected out of all hypotheses rejected. When all hypotheses are true, the FDR and FWER are equivalent. When there are one or more false hypotheses, the FDR is less than or equal to the FWER. Procedures that control the FWER also control the FDR. Procedures designed to control the FDR may not control the FWER, but as a result, are less stringent and may have an increase in power over FWER controlling procedures [Benjamini and Hochberg, 1995]. From Table 1.1, FWER $= \mathrm{P}(\mathrm{V} \geq 1)$ and FDR $= \mathrm{Q}_e = \mathrm{E}[\mathrm{Q}] = \mathrm{E}[\mathrm{V}/(\mathrm{V}+\mathrm{S})]$.

Table 1.1: Number of Errors Committed When Testing m Null Hypotheses

	Test Declared Non-Significant	Test Declared Significant	Total
True Null Hypotheses	U	V	m_0
Non-True Null Hypotheses	T	S	m - m_0
	m - R	R	m

NOTE: From Benjamini and Hochberg [1995].

Procedures designed to control the FWER as well as procedures designed to control the FDR are both widely used in industry. Their use depends on the research area and any relevant guidance. Certain research areas allow for the increase in power achieved by FDR controlling procedures at the expense of an increase in the FWER, whereas other areas require the more stringent methods designed to control the FWER. The methods discussed in this dissertation are developed with confirmatory clinical trials in mind, for which regulatory agencies require control of the FWER. More specifically, they require strong control of the FWER (see Section 3.2).

Chapter 2

Current and New Methods

2.1 Bonferroni Test

Consider m test statistics T_1, ..., T_m for testing hypotheses H_1, ..., H_m with associated p-values P_1, ..., P_m. The global null that all hypotheses are true is represented by:

$$H_0 = \{H_1, ..., H_m\}. \tag{2.1}$$

The Bonferroni test rejects the global null hypothesis if any $P_j \leq \alpha/m$, for $j = 1, ..., m$. This test is simple to apply and requires no assumptions on the distribution of the test statistics. However, the test is conservative, especially for correlated tests.

2.2 Simes Global Test

Consider m test statistics T_1, ..., T_m for testing hypotheses H_1, ..., H_m with associated p-values P_1, ..., P_m. Let $P_{(1)}$, ..., $P_{(m)}$ be the ordered p-values for testing the hypotheses $H_{(1)}$, ..., $H_{(m)}$. Let $H_0 = \{H_{(1)}$, ..., $H_{(m)}\}$ be the intersection of these m hypotheses. The Simes Global Test [Simes, 1986] rejects H_0 if $P_{(j)} \leq j\alpha/m$ for any $j = 1, ..., m$.

2.3 Hochberg Procedure

Consider m test statistics T_1, ..., T_m for testing hypotheses H_1, ..., H_m with associated p-values P_1, ..., P_m. Let $P_{(1)}$, ..., $P_{(m)}$ be the ordered p-values for testing the hypotheses $H_{(1)}$, ..., $H_{(m)}$. Assuming that the Simes global test is an alpha level test, the Hochberg [1988] procedure is:

For $j = m, m-1, ..., 1$, if $P_{(j)} \leq \alpha/(m-j+1)$, then reject all $H_{(j')}$ where $j' \leq j$.

2.4 Hommel Procedure

Consider m test statistics T_1, ..., T_m for testing hypotheses H_1, ..., H_m with associated p-values P_1, ..., P_m. Let $P_{(1)}$, ..., $P_{(m)}$ be the ordered p-values for testing the hypotheses $H_{(1)}$, ..., $H_{(m)}$. Assuming that the Simes global test is an alpha level test, the Hommel [1988] procedure is:

- Compute: $j = \max\left\{ i \in (1,\ ...,\ m) \colon P_{(m-i+k)} > \frac{k\alpha}{i}\ for\ k = 1,\ ...,\ i \right\}$
- If the maximum does not exist, then reject all $H_i, i = 1, ..., m$
- Otherwise reject all H_i with $P_i \leq \alpha/j$

2.5 Extended Single-Step Simes Procedure

Consider m test statistics T_1, ..., T_m for testing hypotheses H_1, ..., H_m with associated p-values P_1, ..., P_m. Let $P_{(1)}$, ..., $P_{(m)}$ be the ordered p-values for testing the hypotheses $H_{(1)}$, ..., $H_{(m)}$. If $P_{(j)} \leq j\alpha/m$ for any $j = 1, ..., m$, then the single-step Simes procedure rejects:

- $H_{(1)}$

- H_i such that $P_i \leq \alpha/m$
- All H_i if $H_{(m)} \leq \alpha$

Chapter 3

Relevant Properties and Definitions

3.1 The Closure Principle

A set of hypotheses and all intersections of those hypotheses is considered a closed family under intersection. For example, consider the hypotheses H_1, H_2, and H_3. The closed family of hypotheses under intersection then consists of H_1, H_2, H_3, $\mathrm{H}_1 \cap \mathrm{H}_2$, $\mathrm{H}_1 \cap \mathrm{H}_3$, $\mathrm{H}_2 \cap \mathrm{H}_3$, and $\mathrm{H}_1 \cap \mathrm{H}_2 \cap \mathrm{H}_3$.

Consider an analysis of variance setting with four means and we wish to test all pairwise comparisons. The pairwise hypotheses are then: $\mathrm{H}_{1,2}$: $\mu_1 = \mu_2$, $\mathrm{H}_{1,3}$: $\mu_1 = \mu_3$, $\mathrm{H}_{1,4}$: $\mu_1 = \mu_4$, $\mathrm{H}_{2,3}$: $\mu_2 = \mu_3$, and $\mathrm{H}_{3,4}$: $\mu_3 = \mu_4$. The intersections of the pairwise comparisons are as follows. Note that the same intersection hypothesis may result from the intersection of more than one pair of hypotheses as is the case with the first four bullets.

- $\mathrm{H}_{1,2} \cap \mathrm{H}_{1,3} = \mathrm{H}_{1,2} \cap \mathrm{H}_{2,3} = \mathrm{H}_{1,3} \cap \mathrm{H}_{2,3} = \mathrm{H}_{1,2,3}$: $\mu_1 = \mu_2 = \mu_3$
- $\mathrm{H}_{1,2} \cap \mathrm{H}_{1,4} = \mathrm{H}_{1,2} \cap \mathrm{H}_{2,4} = \mathrm{H}_{1,4} \cap \mathrm{H}_{2,4} = \mathrm{H}_{1,2,4}$: $\mu_1 = \mu_2 = \mu_4$
- $\mathrm{H}_{1,3} \cap \mathrm{H}_{1,4} = \mathrm{H}_{1,3} \cap \mathrm{H}_{3,4} = \mathrm{H}_{1,4} \cap \mathrm{H}_{3,4} = \mathrm{H}_{1,3,4}$: $\mu_1 = \mu_3 = \mu_4$
- $\mathrm{H}_{2,3} \cap \mathrm{H}_{2,4} = \mathrm{H}_{2,3} \cap \mathrm{H}_{3,4} = \mathrm{H}_{2,4} \cap \mathrm{H}_{3,4} = \mathrm{H}_{2,3,4}$: $\mu_2 = \mu_3 = \mu_4$

- $\text{H}_{1,2} \cap \text{H}_{3,4} = \text{H}_{1,2\cap 3,4}$: $\mu_1 = \mu_2$ and $\mu_3 = \mu_4$
- $\text{H}_{1,3} \cap \text{H}_{2,4} = \text{H}_{1,3\cap 2,4}$: $\mu_1 = \mu_3$ and $\mu_2 = \mu_4$
- $\text{H}_{1,4} \cap \text{H}_{2,3} = \text{H}_{1,4\cap 2,3}$: $\mu_1 = \mu_4$ and $\mu_2 = \mu_3$

The intersection of any two of the above intersection hypotheses is: $\text{H}_{1,2,3,4}$: $\mu_1 = \mu_2 = \mu_3 = \mu_4$. A diagram of the closed family of hypotheses under intersection for the pairwise tests of four means is displayed in Figure 3.1. The global null hypothesis that all means are equal at the top implies all hypotheses below. That is, if the global null hypotheses $\text{H}_{1,2,3,4}$ is true, then all hypotheses below must also be true. Similarly, $\text{H}_{1,2,3}$ implies $\text{H}_{1,2}$, $\text{H}_{1,3}$, and $\text{H}_{2,3}$ and $\text{H}_{1,2\cap 3,4}$ implies $\text{H}_{1,2}$ and $\text{H}_{3,4}$.

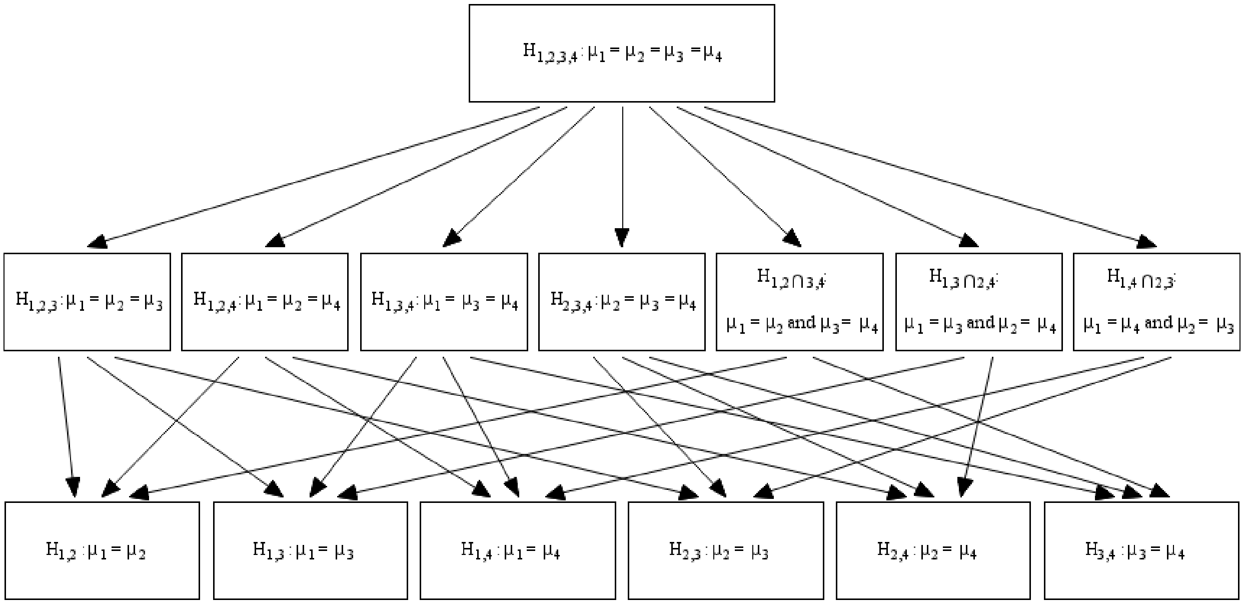

Figure 3.1: Closed Family of Hypotheses Under Intersection for Four Pairwise Comparisons

The closure principle [Marcus et al., 1976] states:

> Let X be a random variable with distribution $P_\theta(\theta \in \Omega)$. Let $W = \{\omega_\beta\}$ be a set of null hypotheses, i.e., a set of subsets of Ω, closed under intersection:

> $\omega_i, \omega_j \in W$ implies $\omega_i \cap \omega_j \in W$. For each ω_β, let $\phi_\beta(X)$ be a level α test, that is $P_\theta(\phi_\beta(X){=}1) \leq \alpha$ for all $\theta \in \omega_\beta$. Now consider the following procedure.
>
> Any null hypothesis ω_β is tested by means of $\phi_\beta(X)$ if and only if all hypotheses ω that are included in $\omega_\beta(\omega \subset \omega_\beta)$ and belonging to $W(\omega \in W)$ have been tested and rejected. The probability of making no type 1 error with this procedure is at least 1 - α.

Therefore, under the closure principle, you may test $H_{1,2}$: $\mu_1{=}\mu_2$, for example, with a valid α level test only if all hypotheses that imply $H_{1,2}$ have been tested and rejected with valid α level tests. That is, $H_{1,2,3,4}$, $H_{1,2,3}$, $H_{1,2,4}$, and $H_{1,2\cap 3,4}$ must have been tested and rejected.

3.2 Strong Control of the Family Wise Error Rate

A procedure has strong control of the family wise error rate, as defined by Hochberg and Tamhane [1987], if the probability of committing at least one type 1 error is less than or equal to α for all configurations of true and false hypotheses. For example, the possible combinations for the case of three tests are:

1. all three hypotheses are true,
2. two true hypotheses and one false hypothesis, and
3. one true hypothesis and two false hypotheses.

To further understand, lets also define a procedure as having weak control of the family wise error rate [Hochberg and Tamhane, 1987] as one that has a probability of committing a type 1 error that is less than or equal to α under the global null hypothesis (all hypotheses true).

To illustrate, again consider an analysis of variance setting with four means and we wish to test all pairwise comparisons using Fisher's protected least significant difference (LSD)

[Fisher, 1966]. Fisher's protected LSD first tests the global null hypothesis that all means are equal using an α level F-test. If non-significant, then all means are declared equal. If significant, then each pairwise comparison of means is tested at level α with a t-test. A diagram of Fisher's protected LSD is displayed in Figure 3.2.

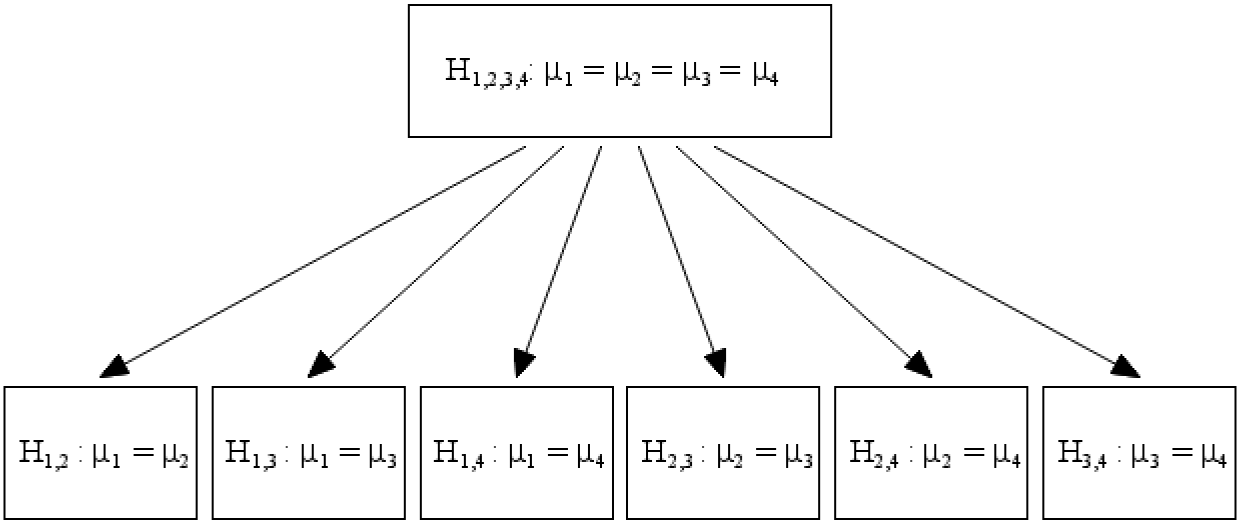

Figure 3.2: Fisher's Protected LSD: Pairwise Comparisons of Four Means

Under the global null hypothesis that all four means are equal, it is well known that Fisher's protected LSD has a family wise error rate less than or equal to α. The reason being that the test of the global null hypothesis "protects" the family wise error rate from inflating due to multiplicity of testing all pairwise comparisons. This is because the pairwise comparisons are only tested if the global rejects, which is less than or equal to α under the global null hypothesis, so the probability of at least one type 1 error among the hypotheses cannot exceed α. However, it is also well documented that Fisher's protected LSD only has weak control of the family wise error rate [Westfall et al., 2011].

To demonstrate, let's assume that our four means represent four dose levels, two low doses (μ_1, μ_2) and two high doses (μ_3, μ_4), such that the two low doses have the same treatment effect (μ_1=μ_2) and the two high doses have the same treatment effect (μ_3=μ_4), but the low dose effect is significantly different from that of the high dose with probability near 1. Under this scenario, the global null hypothesis $\mathrm{H}_{1,2,3,4}$ is false and is rejected with probability near 1. Similarly, $\mathrm{H}_{1,3}$, $\mathrm{H}_{1,4}$, $\mathrm{H}_{2,3}$, and $\mathrm{H}_{2,4}$ are also all false. However, note that two true pairwise comparisons are tested with probability near 1. That is, the two true null hypotheses are always tested at level α. The test of the global null hypothesis does not protect against inflating the type 1 error under this configuration. Figure 3.3 shows Fisher's protected LSD with the false hypotheses highlighted. i.e., the hypotheses that reject with probability near 1. Since the two true hypotheses ($H_{1,2}$, $\mathrm{H}_{2,4}$) are always tested at level α, then the probability of rejecting at least one of these true hypotheses is greater than α. Fisher's protected LSD does not have strong control of the family wise error rate, since it doesn't control the type 1 error rate under all configurations of true and false hypotheses.

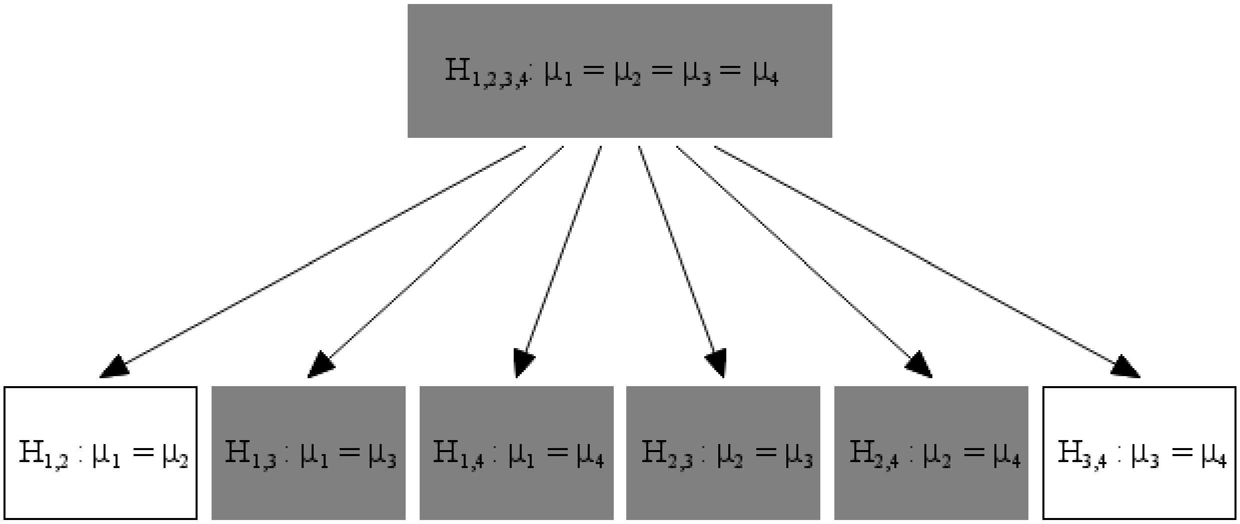

Figure 3.3: Fisher's Protected LSD: Pairwise Comparisons of Four Means ($\mu_1 = \mu_2$, $\mu_3 = \mu_4$)

Unlike Fisher's protected LSD, which bypasses the intermediate intersection hypotheses, testing based on the closure principle does have strong control of the family wise error rate. Recall Figure 3.1, which shows the closed family of hypotheses under intersection.

Figure 3.4 shows the same diagram except with the false hypotheses highlighted. i.e., the hypotheses that reject with probability near 1. Under the closure principle, the true hypotheses $H_{1,2}$ and $H_{3,4}$ may only be tested if all hypotheses that imply it are tested and rejected using valid α level tests. All of which reject with probability near 1 except for the test of $H_{1,2\cap 3,4}$. Under this configuration, the true hypothesis $H_{1,2\cap 3,4}$ would be rejected with probability less than or equal to α. If it is not rejected, then $H_{1,2}$ and $H_{3,4}$ are not tested. Therefore, a true hypothesis is rejected at most with probability α. That is, the family wise error rate is protected. In general, hypothesis testing based on the closure principle may be thought of as having "complete protection" of the family wise error rate.

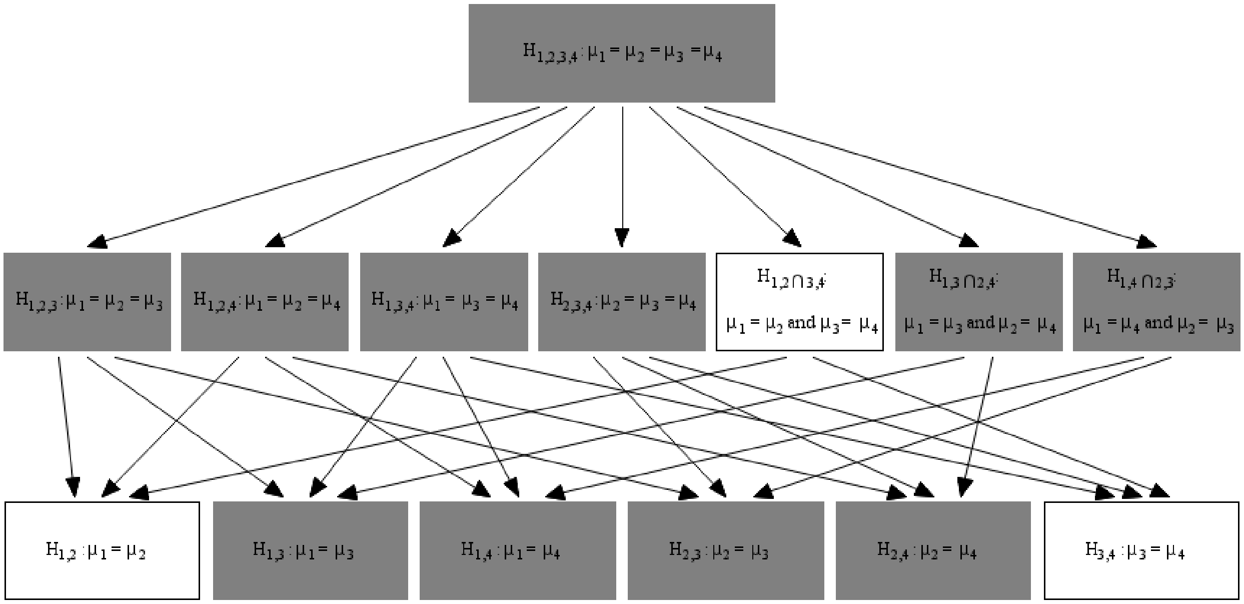

Figure 3.4: Closed Family of Hypotheses Under Intersection for Four Pairwise Comparisons ($\mu_1 = \mu_2$, $\mu_3 = \mu_4$)

3.3 Multivariate Totally Positive of Order 2 (MTP$_2$)

An n-dimensional random vector $\mathbf{x} = (x_1, ..., x_n)'$ is said to have an MTP$_2$ distribution if the corresponding probability density, f($\mathbf{x}$), satisfies the following condition:

$$f(\mathbf{x} \vee \mathbf{y}) f(\mathbf{x} \wedge \mathbf{y}) \geq f(\mathbf{x}) f(\mathbf{y}) \text{ for all } \mathbf{x}, \mathbf{y} \in \Re^{n},$$

where, with $\mathbf{x} = (x_1, ..., x_n)'$ and $\mathbf{y} = (y_1, ..., y_n)'$, $\mathbf{x} \vee \mathbf{y} = (\max(x_1,y_1), ..., \max(x_n,y_n))$ and $\mathbf{x} \wedge \mathbf{y} = (\min(x_1,y_1), ..., \min(x_n,y_n))$ [Sarkar, 1998]. Further, the density $\mathbf{x} = (x_1, ..., x_n) \sim N(0,\Sigma)$ (multivariate normally distributed with covariance Σ) is MTP_2 if and only if Σ^{-1} exhibits non-positive off diagonals [Karlin and Rinnot, 1980].

Chapter 4

Extended Single-Step Simes Procedure for Three Normal Test Statistics

Consider three statistics Z_1, Z_2, and Z_3 for testing hypotheses H_1, H_2, and H_3 with associated p-values P_1, P_2, and P_3. Assume that each hypothesis is testing that $\mu_i = 0$ versus the alternative that $\mu_i > 0$, for $i = 1, 2, 3$. Let $P_{(1)}$, $P_{(2)}$, and $P_{(3)}$ be the ordered p-values with associated hypotheses $H_{(1)}$, $H_{(2)}$, and $H_{(3)}$. The extended single-step Simes procedure rejects at least one hypothesis if $P_{(j)} \leq j\alpha/3$ for any $j = 1, 2, 3$. Figure 4.1 is a 3-dimensional representation of this rejection region.

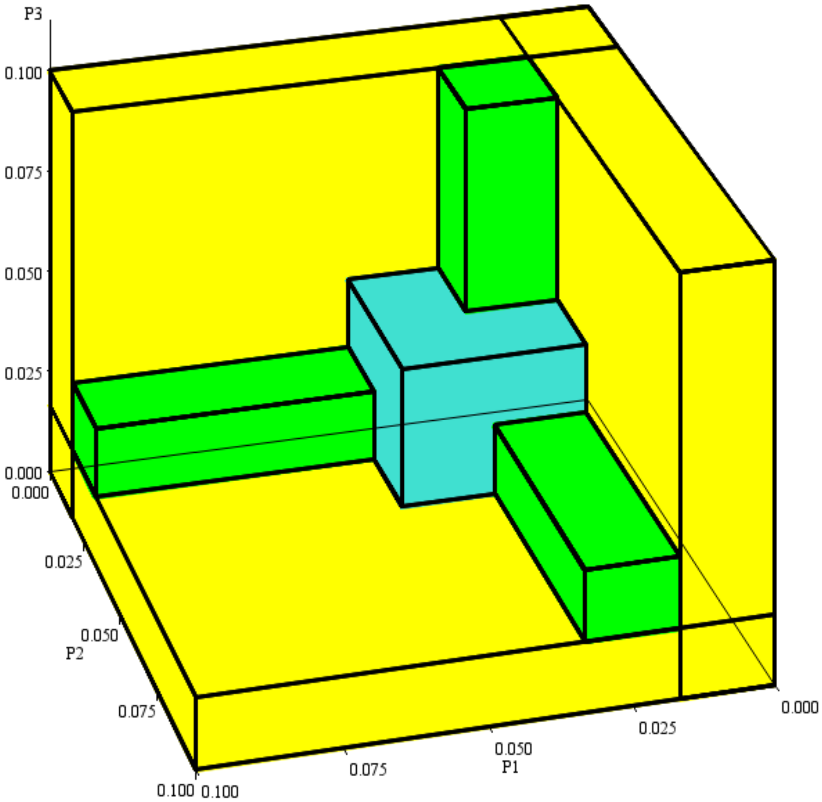

Figure 4.1: Simes Global Test Rejection Region

Rejection due to the yellow region occurs when $P_j \leq \alpha/3$ for $j = 1, 2$, or 3. Rejection due to the blue region occurs when all three p-values are less than α. Without loss of generality, rejection due to the green region occurs when $\alpha/3 \leq P_1$, $P_2 \leq 2\alpha/3$ and $P_3 > \alpha$.

The closed family of hypotheses under intersection for three tests is presented in Section 3.1. If each test can be shown to have a type 1 error rate $\leq \alpha$ under all configurations of true and false hypotheses, then per the closure principle [Marcus et al., 1976], the extended single-step Simes testing procedure will have strong control of the family wise error rate [Hochberg and Tamhane, 1987] for three test statistics.

If all three hypotheses are true, then we are under the global null hypothesis. Simes global test has a type 1 error rate $\leq \alpha$. Therefore, the extended single-step Simes procedure is protected and has a family wise error rate $\leq \alpha$ when all three hypotheses are true. If two of the hypotheses are false, then all intersection hypotheses are false since they contain at least one false hypothesis. The only true hypothesis in the family is the one remaining single hypothesis. Under the extended single-step Simes procedure, an individual hypothesis could potentially be rejected only if its associated p-value is $\leq \alpha$. Therefore, the family wise error rate is $\leq \alpha$ when two of the hypotheses are false.

Without loss of generality, consider that H_3 is false (i.e., $\mu_3 > 0$) and H_1 and H_2 are true. We now explore the probability of committing a type 1 error under this configuration. Again, a type 1 error is committed when one or more of the true hypotheses in the family are rejected. Under this configuration, the intersection of all three hypotheses, pairwise intersections with H_3, and the individual hypothesis H_3 are all false. The remaining true hypotheses consists of the intersection of H_1 and H_2 and the individual hypotheses H_1 and H_2. As mentioned above, the individual hypothesis cannot be rejected unless its associated p-value is $\leq \alpha$. Therefore, to show that the extended single-step Simes procedure has strong control of the family wise error rate for three test statistics, we must show the following:

$$\mathrm{P}(\text{reject } H_1 \cap H_2) \leq \alpha. \tag{4.1}$$

The above is rejected if either or both H_1 and H_2 are rejected. H_1 is rejected if Simes global test rejects and $\mathrm{P}_1 \leq \alpha/3$, $\mathrm{P}_1 < \min(\mathrm{P}_2, \mathrm{P}_3)$, or $\max(\mathrm{P}_1, \mathrm{P}_2, \mathrm{P}_3) \leq \alpha$. Similar for rejecting H_2. This corresponds to the yellow areas to the back and right, the blue area and the area directly below it, and the red areas in Figure 4.2.

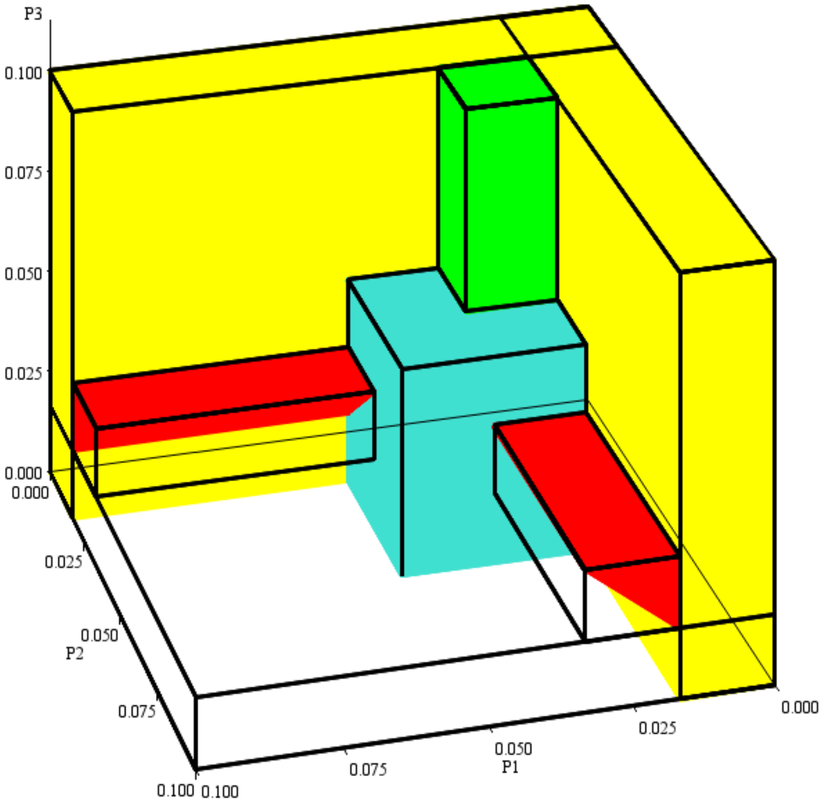

Figure 4.2: Simes Global Test Rejection Region $H_1 \cap H_2$

4.1 Three Independent Normal Statistics

Theorem 1. *Let Z_1, Z_2, and Z_3 be independent standard normal test statistics for testing the hypotheses H_1, H_2, and H_3, then the extended single-step Simes procedure has strong control of the family wise error rate.*

Proof:

Assume that the three normal test statistics Z_1, Z_2, and Z_3 are independent. Let's compare the rejection region of $H_1 \cap H_2$ for the extended single-step Simes procedure shown in Figure 4.2 to that of the Hochberg procedure for three hypotheses. The Hochberg procedure rejection region includes the yellow and blue regions as well as the areas indicated by the dashed lines in Figure 4.3.

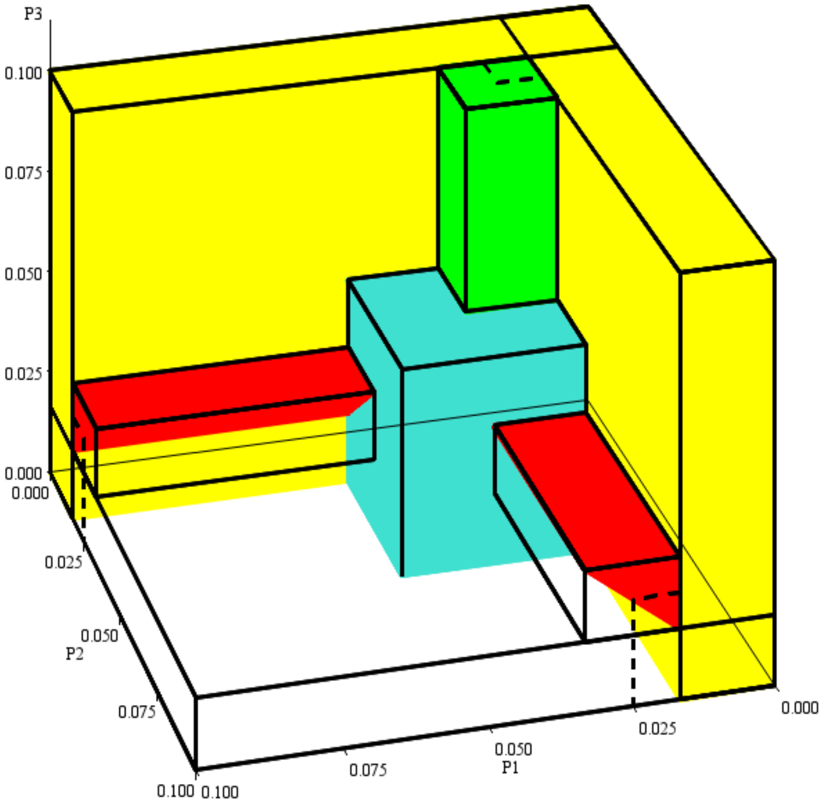

Figure 4.3: Simes Global Test Rejection Region $H_1 \cap H_2$ with Hochberg Procedure for Three Hypotheses Overlaid

The probability of being within this Hochberg rejection region for $H_1 \cap H_2$ for three hypotheses is known to be $\leq \alpha$. It is also true that the probability to reject $H_1 \cap H_2$ when both are true using the Hochberg procedure for two tests is also $\leq \alpha$. If we instead overlay the rejection region of $H_1 \cap H_2$ for the Hochberg procedure for two hypotheses, the rejection region includes the yellow and blue areas as before as well as the areas indicated by the dashed lines in Figure 4.3 below.

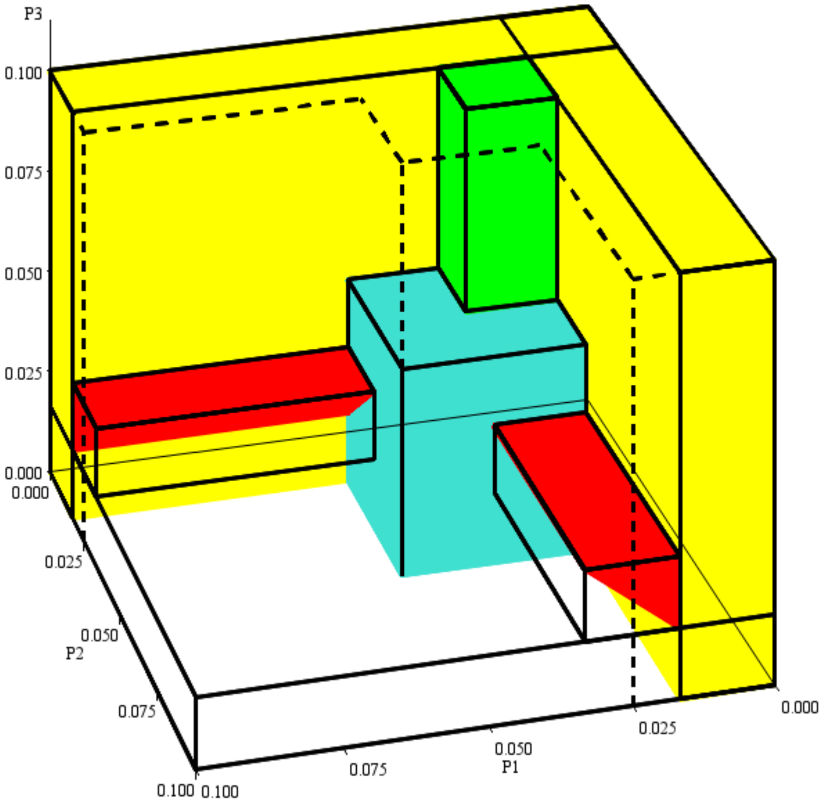

Figure 4.4: Simes Global Test Rejection Region $H_1 \cap H_2$ with Hochberg Procedure for Two Hypotheses Overlaid

From this figure, one can see that the rejection region of $H_1 \cap H_2$ for the Hochberg procedure contains the entire rejection region for the extended single-step Simes procedure except for a portion of each of the red triangular prisms. Also note that below each of the red triangular prisms is a triangular prism belonging to the Hochberg rejection region that is not part of the extended single-step Simes procedure's rejection region.

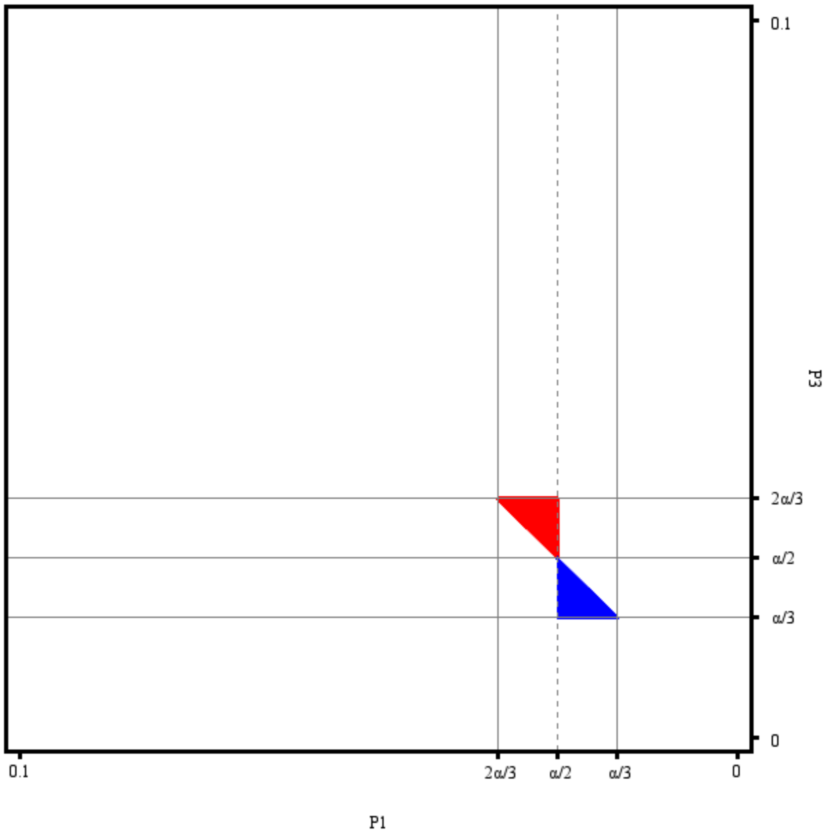

Figure 4.5: Difference Between Rejection Regions of H_1 for Simes Single-Step and Hochberg for a fixed $\alpha < P_2 \leq 1$ (P-values)

Since P_2 is independent of P_1 and P_3, every 2-dimensional slice in P_1 and P_3 at a fixed $\alpha < P_2 \leq 1$, is identical. Figure 4.5 represents one of these slices. The red triangle is the region where H_1 is rejected by the extended single-step Simes procedure, but is not rejected by the Hochberg procedure. The blue triangle is the region where H_1 is rejected by the Hochberg procedure, but not the extended single-step Simes procedure. Under the global null hypothesis, the p-values are uniformly distributed from 0 to 1. It is easily shown that the probability of these two areas is equal under the global null hypothesis. Figure 4.6 is the same plot except on the Z-scale, where Z_1 is N(0,1) and Z_3 is N(μ_3 >0,1).

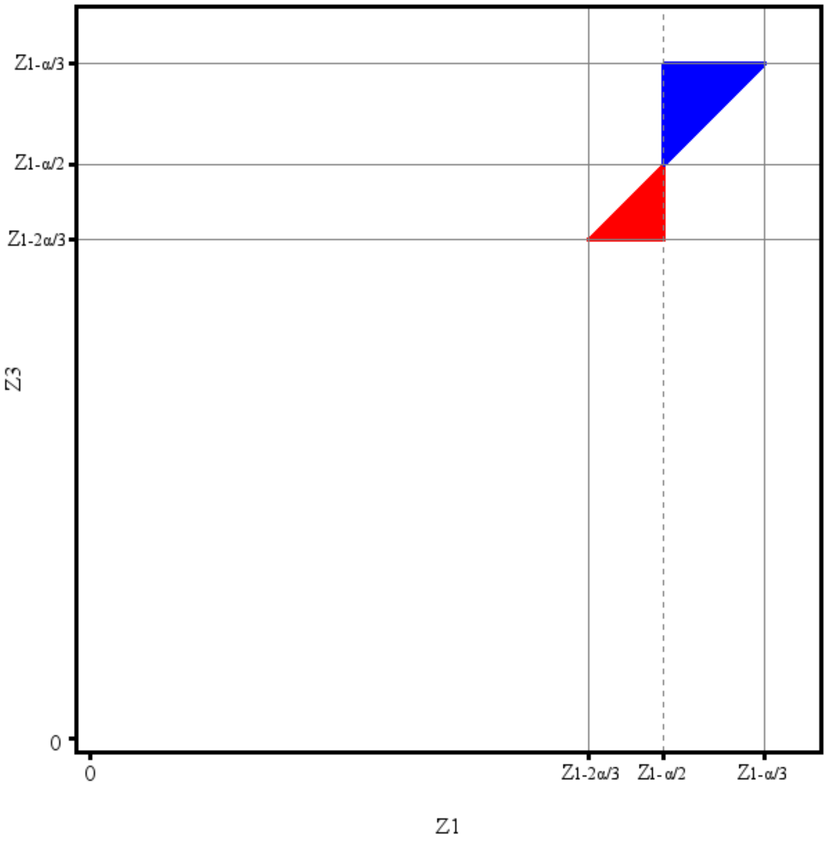

Figure 4.6: Difference Between Rejection Regions of H_1 for Simes Single-Step and Hochberg for a fixed $\alpha < P_2 \leq 1$ (Z values)

If it can be shown that the P(red region) $\leq$ P(blue region) under the alternative ($\mu_3 > 0$), then without loss of generality, the P(reject $H_1 \cap H_2$) for the extended single-step Simes procedure is $\leq$ P(reject $H_1 \cap H_2$) for the Hochberg procedure. If that is the case, then the extended single-step Simes procedure will have strong control of the family wise error rate for three independent normals.

Let g represent the distribution of Z_1, Z_2, and/or Z_3 under the global null hypothesis and h represent the distribution under the alternative (μ_3 >0) and:

- (x_{11}, x_{31}) = lower left corner of the new procedure region
- (x_{12}, x_{32}) = a point in the new procedure region

- (x_{13}, x_{33}) = point where the two methods meet
- (x_{14}, x_{34}) = a point in the Hochberg region
- (x_{15}, x_{35}) = upper right corner of the Hochberg region

The normal distribution exhibits the monotone likelihood ratio property. Therefore:

$$\frac{g_{Z_3}(x_{32})}{g_{Z_3}(x_{34})} \geq \frac{h_{Z_3}(x_{32})}{h_{Z_3}(x_{34})}$$

Multiplying each side by $g_{Z_1}(x_{12})$.

$$\frac{g_{Z_3}(x_{32})g_{Z_1}(x_{12})}{g_{Z_3}(x_{34})} \geq \frac{h_{Z_3}(x_{32})g_{Z_1}(x_{12})}{h_{Z_3}(x_{34})}$$

This is true for all x_{32} from x_{31} to x_{12}.

$$\frac{\int_{x_{31}}^{x_{12}} g_{Z_3}(z)g_{Z_1}(x_{12})dz}{g_{Z_3}(x_{34})} \geq \frac{\int_{x_{31}}^{x_{12}} h_{Z_3}(z)g_{Z_1}(x_{12})dz}{h_{Z_3}(x_{34})}$$

This is true for all x_{12} from x_{11} to x_{13}.

$$\frac{\int_{x_{11}}^{x_{13}} \int_{x_{31}}^{y} g_{Z_3}(z)g_{Z_1}(y)dzdy}{g_{Z_3}(x_{34})} \geq \frac{\int_{x_{11}}^{x_{13}} \int_{x_{31}}^{y} h_{Z_3}(z)g_{Z_1}(y)dzdy}{h_{Z_3}(x_{34})}$$

Since the test statistics are independent.

$$\frac{\int_{x_{11}}^{x_{13}} \int_{x_{31}}^{y} g_{Z_1,Z_3}(y,z)dzdy}{g_{Z_3}(x_{34})} \geq \frac{\int_{x_{11}}^{x_{13}} \int_{x_{31}}^{y} h_{Z_1,Z_3}(y,z)dzdy}{h_{Z_3}(x_{34})}$$

Equivalently:

$$\frac{g_{Z_3}(x_{34})}{\int_{x_{11}}^{x_{13}} \int_{x_{31}}^{y} g_{Z_1,Z_3}(y,z)dzdy} \leq \frac{h_{Z_3}(x_{34})}{\int_{x_{11}}^{x_{13}} \int_{x_{31}}^{y} h_{Z_1,Z_3}(y,z)dzdy}$$

Multiplying each side by $g_{Z_1}(x_{14})$.

$$\frac{g_{Z_3}(x_{34})g_{Z_1}(x_{14})}{\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} g_{Z_1,Z_3}(y,z)dzdy} \leq \frac{h_{Z_3}(x_{34})g_{Z_1}(x_{14})}{\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} h_{Z_1,Z_3}(y,z)dzdy}$$

This is true for all x_{34} from x_{14} to x_{35}.

$$\frac{\int_{x_{14}}^{x_{35}} g_{Z_3}(z)g_{Z_1}(x_{14})dz}{\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} g_{Z_1,Z_3}(y,z)dzdy} \leq \frac{\int_{x_{14}}^{x_{35}} h_{Z_3}(z)g_{Z_1}(x_{14})dz}{\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} h_{Z_1,Z_3}(y,z)dzdy}$$

This is true for all X_{14} from X_{13} to X_{15}.

$$\frac{\int_{x_{13}}^{x_{15}}\int_{y}^{x_{35}} g_{Z_3}(z)g_{Z_1}(y)dzdy}{\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} g_{Z_1,Z_3}(y,z)dzdy} \leq \frac{\int_{x_{13}}^{x_{15}}\int_{y}^{x_{35}} h_{Z_3}(z)g_{Z_1}(y)dzdy}{\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} h_{Z_1,Z_3}(y,z)dzdy}$$

Since the test statistics are independent.

$$\frac{\int_{x_{13}}^{x_{15}}\int_{y}^{x_{35}} g_{Z_1,Z_3}(y,z)dzdy}{\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} g_{Z_1,Z_3}(y,z)dzdy} \leq \frac{\int_{x_{13}}^{x_{15}}\int_{y}^{x_{35}} h_{Z_1,Z_3}(y,z)dzdy}{\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} h_{Z_1,Z_3}(y,z)dzdy}$$

The left side is the ratio of the two areas under the null. As stated previously, this is equal to 1.

$$1 \leq \frac{\int_{x_{13}}^{x_{15}}\int_{y}^{x_{35}} h_{Z_1,Z_3}(y,z)dzdy}{\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} h_{Z_1,Z_3}(y,z)dzdy} \tag{4.2}$$

The right side of the inequality 4.2 is the area of the Hochberg region over the area of the extended single-step Simes procedure when $\mu_1 = 0$ and $\mu_3 > 0$. The Hochberg procedure has a type 1 error rate $\leq \alpha$. The P(reject $H_1 \cap H_2$) $\leq \alpha$ for the extended single-step Simes since it never exceeds that of the Hochberg procedure. Therefore, the extended single-step Simes procedure has strong control of the family wise error rate for three independent standard normal test statistics.

4.2 Trivariate MTP$_2$ Normal Statistics

Proposition 2. *Let X_1 and X_2 be distributed as a standard bivariate normal distribution with correlation ρ and:*

$P_{X_1,X_2}(a \leq x_1 \leq b) = P_{X_1,X_2}(b \leq x_1 \leq c) = P_{X_1,X_2}(a \leq x_2 \leq b) = P_{X_1,X_2}(b \leq x_2 \leq c)$ *for* $0 \leq a \leq b \leq c$.

Then:

$P_{X_1,X_2}(a \leq x_1 \leq b, a \leq x_2 \leq b) \leq P_{X_1,X_2}(b \leq x_1 \leq c, b \leq x_2 \leq c)$ *for any* $0 \leq \rho \leq 1$.

Note the following properties of the standard univariate and standard bivariate normal with correlation ρ that are well established:

1. For the (marginal) standard normal distributions, $(a+c)/2 > b$.

2. The joint distribution is symmetric about $x_1 = x_2$.

3. The conditional distribution X_2 given X_1 is:

$$f_{X_2|X_1} \sim N[\rho x_1,(1-\rho^2)]$$

4. It follows that the variance of $X_2|X_1$ is not a function of X_1, and that the conditional mean of X_2 increases in X_1.

5. For any $0< \rho <1$ and (x_{11}, x_{12}), $P_{X_2|X_1=x_{11}}(a \leq x_2 \leq c) \leq P_{X_2|X_1=x_{12}}(a \leq x_2 \leq c)$ if and only if $|x_{11} * \rho-(a+c)/2| \geq |x_{12} * \rho-(a+c)/2|$. i.e., the probability is greater the closer the conditional mean is to $(a+c)/2$.

Proof:

Case 1: Consider the case when $c = \infty$. The proof of this has been presented in Hochberg and Rom [1995] and Samuel-Cahn [1996] based on various properties of the standard bivariate normal.

Case 2: Consider Figure 4.7 for c $< \infty$ and $0 < \rho < 1$. Squares 2 and 3 represent the areas of comparison. The horizontal red line is at the midpoint of a and c. i.e., (a+c)/2, which is always > b. The diagonal dashed black line represents $X_1 = X_2$. The diagonal dashed red line represents the conditional mean of $X_2|X_1$.

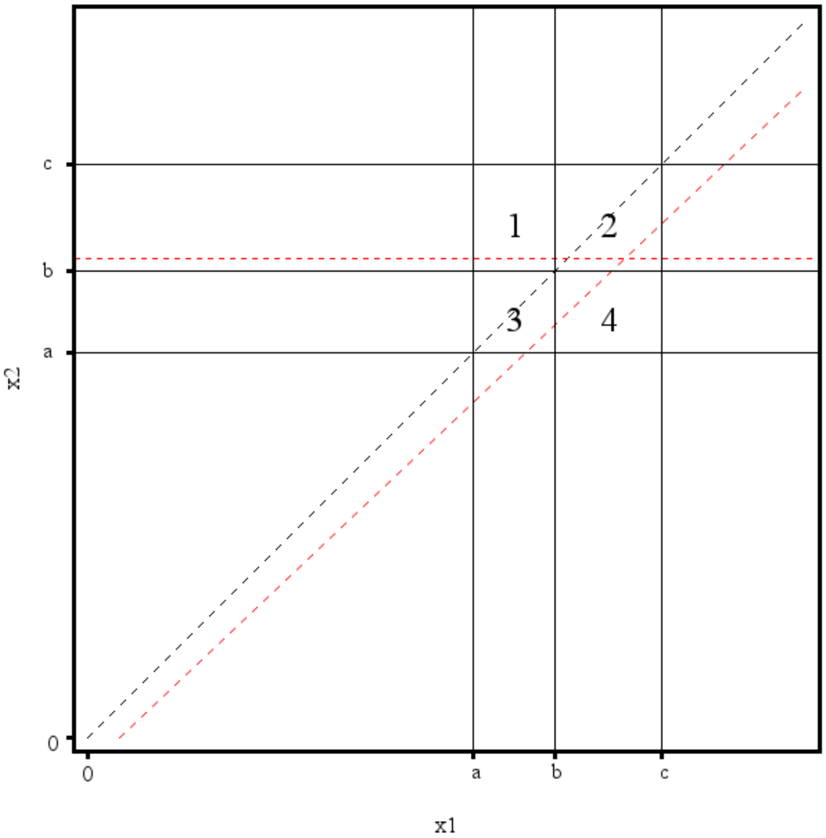

Figure 4.7: Standard Bivariate Normal with Correlation $0 < \rho < 1$ (c$< \infty$)

Define $U(x_1) = 1 - P(X \geq x_1)$, then Figure 4.7 with transformed X_1 becomes Figure 4.8.

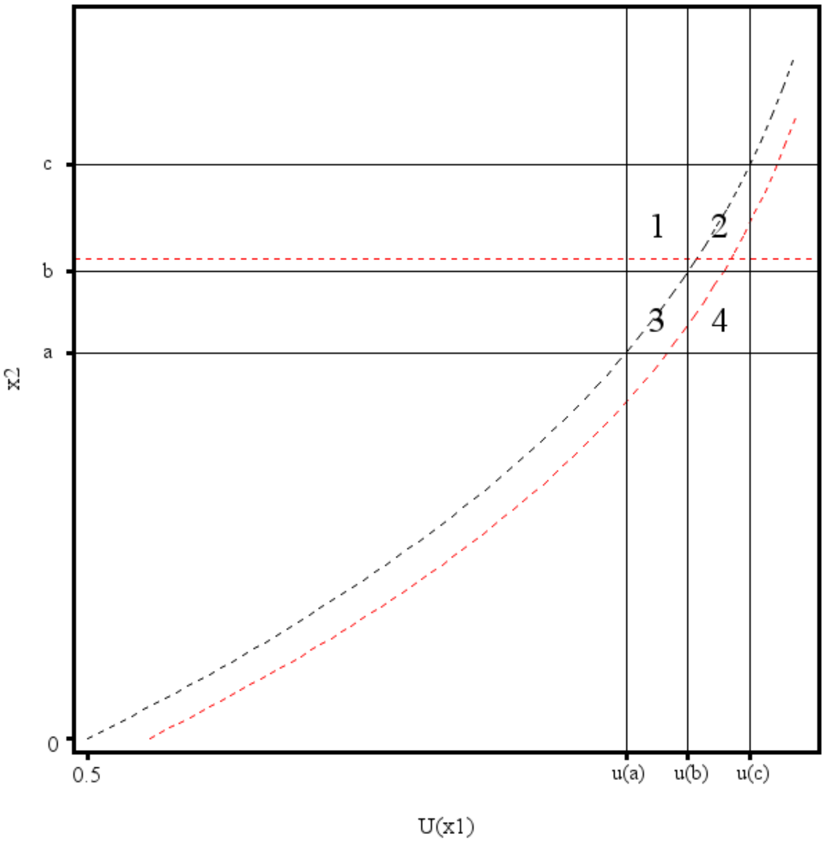

Figure 4.8: Standard Bivariate Normal with X_1 Transformed to Uniform

Since X_1 and X_2 have the same distribution, $U(x_1)$ is the cumulative distribution function (CDF) of X_2. Recall that the black dashed line is $U(x_1)$ vs. X_2, where $X_1 = X_2$, so it is equivalent to $U(x_2)$ vs. X_2. Therefore, the black dashed line is the plot of X_2 vs. it's CDF.

The second derivative of the standard normal CDF is: $\frac{-x}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. Setting this equal to 0 yields a single inflection point at $x = 0$. To the right of this inflection point, the second derivative is negative, so when plotting X_2 on the x-axis and the CDF on the y-axis, to the right of $X_2 = 0$, the CDF is concave downward. However, we have the axes reversed in the plot above, so the CDF is concave upward for $X_2 > 0$ or, equivalently, for $X_1 > 0$.

Figure 4.9 is the same plot, but with a line drawn connecting the lower left corner of box 3 and the upper right corner of box 2. After the transformation, the two boxes above the horizontal dashed red line up to c and below the horizontal red dashed line down to a all have the same dimensions, therefore this new line crosses $U(x_1) = U(b)$ at the same spot as the horizontal dashed red line.

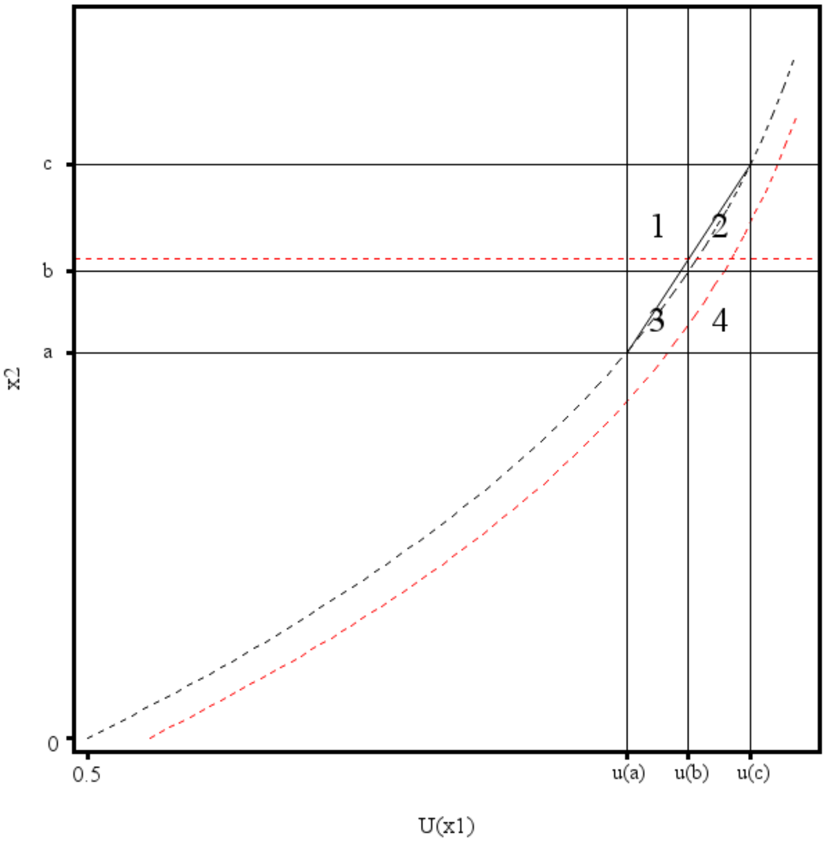

Figure 4.9: Standard Bivariate Normal with X_1 Transformed to Uniform with Dividing Line

Since the dashed black line is concave upward it is on or below the line connecting the lower left corner of box 3 and the upper right corner of box 2 for all $U(x_1)$ such that $U(a) \leq U(x_1) \leq U(c)$.

Define u_{11}, u_{12} as $u_{11} = U(b) - \epsilon$ and $u_{12} = U(b) + \epsilon$ for $\epsilon > 0$ such that $U(a) \leq u_{11} \leq U(b)$ and $U(b) \leq u_{12} \leq U(c)$. u_{11} and u_{12} are then equidistant from U(b). The vertical distance in X_2 from (a+c)/2 to the diagonal solid black line at these two points is equivalent. Note that the conditional mean of X_2 at u_{11} will always be less than (a+c)/2. If the conditional mean of X_2 at u_{12} is less than or equal to (a+c)/2, since the conditional mean of X_2 at u_{12} is greater than the conditional mean of X_2 at u_{11}, then it will be closer to (a+c)/2. Therefore, $P_{X_2|X_1=U^{-1}(u_{11})}(a \leq x_2 \leq c) \leq P_{X_2|X_1=U^{-1}(u_{12})}(a \leq x_2 \leq c)$, when the conditional mean of $x_2|x_1 = u_{12} \leq (a+c)/2$.

Note that the conditional mean of X_2 is bounded above by the concave upward dashed line of $U(x_1) = x_2$. This line and the line representing the conditional mean of X_2 never cross for $X_1 > 0$ and $0 < \rho < 1$ and the lines are the same only for the degenerate conditional distribution when $\rho = 1$.

Let d_{11} and d_{12} represent the vertical distance from $x_2 = (a+c)/2$ to the diagonal black line dividing boxes 2 and 3 at u_{11} and u_{12}, respectively. As noted above, $d_{11} = d_{12}$. When the conditional mean of X_2 at u_{12} is greater than (a+c)/2, the distance the conditional mean of X_2 from (a+c)/2 is less than or equal to d_{12}. At the corresponding point u_{11}, the distance the conditional mean of X_2 is from (a+c)/2 is greater than or equal to d_{11}. Therefore, the conditional mean of X_2 at u_{12} is no farther away from (a+c)/2 than the conditional mean of x_2 at u_{11}. Thus, $P_{X_2|X_1=U^{-1}(u_{11})}(a \leq x_2 \leq c) \leq P_{X_2|X_1=U^{-1}(u_{12})}(a \leq x_2 \leq c)$, when the conditional mean of $X_2|X_1 = u_{12} > (a+c)/2$. Therefore, the inequality holds for any $0 < \rho < 1$. i.e.:

$$\int_a^c f_{X_2|X_1=U^{-1}(u_{11})}(y)dy \leq \int_a^c f_{X_2|X_1=U^{-1}(u_{12})}(y)dy$$

This inequality is true for all pairs u_{11} and u_{12}, $U(a) \leq u_{11} \leq U(b) \leq u_{12} \leq U(c)$, and, therefore, true for their corresponding values in x_1. i.e., the inequality is true for all $x_{11}=U^{-1}(u_{11})$ and $x_{12}=U^{-1}(u_{12})$, $a \leq x_{11} \leq b \leq x_{12} \leq c$. We can then obtain the total probabilities of boxes 1 + 3 and 2 + 4 by integrating over $U'(x_1)=u(x_1)$.

$$\int_{U(a)}^{U(b)}\int_{a}^{c} u_{X_1}(y) f_{X_2|X_1}(z)dydz \leq \int_{U(b)}^{U(c)}\int_{a}^{c} u_{X_1}(y) f_{X_2|X_1}(z)dydz$$

Back-transforming, we have:

$$\int_{a}^{b}\int_{a}^{c} f_{X_1}(y) f_{X_2|X_1}(z)dydz \leq \int_{b}^{c}\int_{a}^{c} f_{X_1}(y) f_{X_2|X_1}(z)dydz$$

Which is equivalent to:

$$\int_{a}^{b}\int_{a}^{c} f_{X_1,X_2}(y,z)dydz \leq \int_{b}^{c}\int_{a}^{c} f_{X_1,X_2}(y,z)dydz$$

This means that the probability of boxes 1 + 3 is less than or equal to the probability of boxes 2 + 4 when $0 < \rho < 1$. Due to symmetry, boxes 1 and 4 have the same probability. Removing these two regions from the integrals, the inequality above becomes:

$$\int_{a}^{b}\int_{a}^{b} f_{X_1,X_2}(y,z)dydz \leq \int_{b}^{c}\int_{b}^{c} f_{X_1,X_2}(y,z)dydz \tag{4.3}$$

Therefore, the probability of box 3 is less than or equal to the probability of box 2. That is:

$\mathrm{P}_{X_1,X_2}(\mathrm{a} \leq \mathrm{x}_1 \leq \mathrm{b}, \mathrm{a} \leq \mathrm{x}_2 \leq \mathrm{b}) \leq \mathrm{P}_{X_1,X_2}(\mathrm{b} \leq \mathrm{x}_1 \leq \mathrm{c}, \mathrm{b} \leq \mathrm{x}_2 \leq \mathrm{c})$ for any $0 \leq \rho \leq 1$.

Theorem 3. *Let Z_1, Z_2, and Z_3 be standard trivariate normal test statistics for testing hypotheses H_1, H_2, and H_3 whose joint distribution has the MTP_2 property, then the extended single-step Simes procedure has strong control of the family wise error rate.*

Proof:

Assume that the three normal test statistics Z_1, Z_2, and Z_3 are positively correlated and that their trivariate normal distribution has the MTP_2 property with $\mathrm{cov}(\mathrm{Z}_i,\mathrm{Z}_j)=\sigma_{ij}$, i $\neq$ j. Under the null, $\mu_i = 0$, for i = 1, 2, 3. The conditional distribution of Z_2 given Z_1 = a and Z_3 = c is then normal with mean (μ_2^*) and variance (E_2^*):

$$\mu_2^* = \frac{(\sigma_{12} - \sigma_{13}\sigma_{23})a + (\sigma_{23} - \sigma_{12}\sigma_{13})c}{1 - \sigma_{13}^2}$$

$$E_2^* = 1 - \frac{\sigma_{12}^2 - 2\sigma_{12}\sigma_{13}\sigma_{23} + \sigma_{23}^2}{1 - \sigma_{13}^2}$$

Since we assumed that the trivariate density of Z_1, Z_2, and Z_3 is MTP_2, the inverse of the covariance matrix is an M-matrix [Karlin and Rinnot, 1980]. The inverse of the covariance matrix is:

$$\frac{1}{1 + 2\sigma_{12}\sigma_{13}\sigma_{23} - \sigma_{12}^2 - \sigma_{13}^2 - \sigma_{23}^2}\begin{bmatrix} 1 - \sigma_{23}^2 & \sigma_{13}\sigma_{23} - \sigma_{12} & \sigma_{12}\sigma_{23} - \sigma_{13} \\ \sigma_{13}\sigma_{23} - \sigma_{12} & 1 - \sigma_{13}^2 & \sigma_{12}\sigma_{13} - \sigma_{23} \\ \sigma_{12}\sigma_{23} - \sigma_{13} & \sigma_{12}\sigma_{13} - \sigma_{23} & 1 - \sigma_{12}^2 \end{bmatrix}$$

The off-diagonal elements of an M-matrix are all ≤ 0. Therefore, the quantities $\sigma_{12} - \sigma_{13}\sigma_{23}$ and $\sigma_{23} - \sigma_{12}\sigma_{13}$ in the conditional mean are both positive. Thus, the conditional mean of Z_2 given Z_1 and Z_3 increases as Z_1=a increases and increases as Z_3=c increases. Now consider again Figure 4.6. The conditional mean of Z_2 for any Z_1=a_1 and Z_3=c_1 in the red area is less than the conditional mean of Z_2 for any Z_1=a_2 and Z_3=c_2 in the blue area.

For these given points in Z_1 and Z_3, we can compare the probabilities in Z_2 for the extended single-step Simes procedure to that of the Hochberg procedure. Due to the stochastic ordering of the normal distribution:

$$P_{Z_2|Z_1=a_1,Z_3=c_1}(\Phi^{-1}(\alpha) \leq z_2 \leq \infty) \leq P_{Z_2|Z_1=a_2,Z_3=c_2}(\Phi^{-1}(\alpha) \leq z_2 \leq \infty)$$

Where Φ^{-1} is the inverse of the standard normal distribution. Or equivalently, letting g represent the probability density function (PDF) of Z_1, Z_2, and/or Z_3 under the global null hypothesis and associated cumulative distribution function (CDF) G:

$$\int_{G^{-1}_{(\alpha)}}^{\infty} g_{Z_2|Z_1=a_1,Z_2=c_1}(x_2)dx_2 \leq \int_{G^{-1}_{(\alpha)}}^{\infty} g_{Z_2|Z_1=a_2,Z_2=c_2}(x_2)dx_2 \tag{4.4}$$

Under the global null hypothesis, using x_{11}, x_{13}, x_{15}, x_{31}, x_{33}, and x_{35} as defined in Section 4.1, by Proposition 2 we can rewrite inequality 4.3 as:

$$\int_{x_{11}}^{x_{13}} \int_{x_{31}}^{x_{33}} g_{Z_1,Z_3}(y,z)dzdy \leq \int_{x_{13}}^{x_{15}} \int_{x_{33}}^{x_{35}} g_{Z_1,Z_3}(y,z)dzdy \tag{4.5}$$

The box represented on the left side of the inequality is the box containing the red triangle in Figure 4.6 and the right side of the inequality is the box containing the blue triangle in Figure 4.6. Consider the diagonal line where $Z_1 = Z_3$. This line divides the boxes in half. Additionally, the red box is below the diagonal and the blue box is above the diagonal. Under the global null hypothesis, Z_1 and Z_3 have the same marginal distributions, N(0,1). i.e., g_{Z_1}(a) = g_{Z_3}(a). Now consider their joint bivariate normal distribution g_{Z_1,Z_3}. Since Z_1 and Z_3 have the same distribution, they are interchangeable in the joint distribution. That is, g_{Z_1,Z_3}(a,b) = g_{Z_1,Z_3}(b,a). Consider the point (a,b) on one side of the diagonal within one of the boxes. It's corresponding point symmetric across the diagonal is (b,a). As noted above g_{Z_1,Z_3}(a,b) = g_{Z_1,Z_3}(b,a). For each point in one of the boxes above the diagonal is a corresponding point below the diagonal. Therefore, the joint bivariate normal distribution of Z_1 and Z_3 is symmetric about the diagonal. That is, within each box, the probability of the triangle above the diagonal is equal to the probability of the triangle below the diagonal. Therefore, the left side of inequality 4.5 can be rewritten as twice the area of the red triangle corresponding to the extended single-step Simes region and the right side of the inequality can be rewritten as twice the area of the blue triangle corresponding to the Hochberg region. That is:

$$2\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} g_{Z_1,Z_3}(y,z)dzdy \leq 2\int_{x_{13}}^{x_{15}}\int_{y}^{x_{35}} g_{Z_1,Z_3}(y,z)dzdy \tag{4.6}$$

Or equivalently:

$$\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} g_{Z_1,Z_3}(y,z)dzdy \leq \int_{x_{13}}^{x_{15}}\int_{y}^{x_{35}} g_{Z_1,Z_3}(y,z)dzdy \tag{4.7}$$

Thus, the probability of the extended single-step Simes region in Z_1 and Z_3 is less than or equal to that of the Hochberg region under the global null. Multiplying the inequalities 4.4 and 4.7, we have:

$$\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y}\int_{G^{-1}_{(\alpha)}}^{\infty} g_{Z_2|Z_1,Z_3}(w)g_{Z_1,Z_3}(y,z)dwdzdy \leq \int_{x_{13}}^{x_{15}}\int_{y}^{x_{35}}\int_{G^{-1}_{(\alpha)}}^{\infty} g_{Z_2|Z_1,Z_3}(w)g_{Z_1,Z_3}(y,z)dwdzdy$$

or equivalently:

$$\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y}\int_{G^{-1}_{(\alpha)}}^{\infty} g_{Z_1,Z_2,Z_3}(y,w,z)dwdzdy \leq \int_{x_{13}}^{x_{15}}\int_{y}^{x_{35}}\int_{G^{-1}_{(\alpha)}}^{\infty} g_{Z_1,Z_2,Z_3}(y,w,z)dwdzdy \tag{4.8}$$

The left side of the inequality is the probability associated with the triangular prism belonging only to the extended single-step Simes procedure and the right side of the inequality is the probability associated with the triangular prism belonging only to the Hochberg procedure.

Again let g and h, with associated CDFs G and H, represent the distributions of Z_1, Z_2, and/or Z_3 under the global null hypothesis and alternatives hypothesis, respectively. Since the univariate conditional distributions and univariate marginal distributions are normal, they exhibit the monotone likelihood ratio property. Therefore, the following two inequalities are both true:

$$\frac{g_{Z_3|Z_1}(x_{32})}{g_{Z_3|Z_1}(x_{34})} \geq \frac{h_{Z_3|Z_1}(x_{32})}{h_{Z_3|Z_1}(x_{34})}$$

$$\frac{g_{Z_1}(x_{12})}{g_{Z_1}(x_{14})} \geq \frac{h_{Z_1}(x_{12})}{h_{Z_1}(x_{14})}$$

Multiplying the two inequalities together:

$$\frac{g_{Z_3|Z_1}(x_{32})g_{Z_1}(x_{12})}{g_{Z_3|Z_1}(x_{34})g_{Z_1}(x_{14})} \geq \frac{h_{Z_3|Z_1}(x_{32})h_{Z_1}(x_{12})}{h_{Z_3|Z_1}(x_{34})h_{Z_1}(x_{14})}$$

This is true for all x_{32} from x_{31} to x_{12}.

$$\frac{\int_{x_{31}}^{x_{12}} g_{Z_3|Z_1}(z)g_{Z_1}(x_{12})dz}{g_{Z_3|Z_1}(x_{34})g_{Z_1}(x_{14})} \geq \frac{\int_{x_{31}}^{x_{12}} h_{Z_3|Z_1}(z)h_{Z_1}(x_{12})dz}{h_{Z_3|Z_1}(x_{34})h_{Z_1}(x_{14})}$$

This is true for all x_{12} from x_{11} to x_{13}.

$$\frac{\int_{x_{11}}^{x_{13}} \int_{x_{31}}^{y} g_{Z_3|Z_1}(z)g_{Z_1}(y)dzdy}{g_{Z_3|Z_1}(x_{34})g_{Z_1}(x_{14})} \geq \frac{\int_{x_{11}}^{x_{13}} \int_{x_{31}}^{y} h_{Z_3|Z_1}(z)h_{Z_1}(y)dzdy}{h_{Z_3|Z_1}(x_{34})h_{Z_1}(x_{14})}$$

Multiplying the distributions together:

$$\frac{\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} g_{Z_1,Z_3}(y,z)dzdy}{g_{Z_3|Z_1}(x_{34})g_{Z_1}(x_{14})} \geq \frac{\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} h_{Z_1,Z_3}(y,z)dzdy}{h_{Z_3|Z_1}(x_{34})h_{Z_1}(x_{14})}$$

Inverting:

$$\frac{g_{Z_3|Z_1}(x_{34})g_{Z_1}(x_{14})}{\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} g_{Z_1,Z_3}(y,z)dzdy} \leq \frac{h_{Z_3|Z_1}(x_{34})h_{Z_1}(x_{14})}{\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} h_{Z_1,Z_3}(y,z)dzdy}$$

This is true for all x_{34} from x_{14} to x_{35}.

$$\frac{\int_{x_{14}}^{x_{35}} g_{Z_3|Z_1}(z)g_{Z_1}(x_{14})dz}{\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} g_{Z_1,Z_3}(y,z)dzdy} \leq \frac{\int_{x_{14}}^{x_{35}} h_{Z_3|Z_1}(z)h_{Z_1}(x_{14})dz}{\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} h_{Z_1,Z_3}(y,z)dzdy}$$

This is true for all x_{14} from x_{13} to x_{15}.

$$\frac{\int_{x_{13}}^{x_{15}}\int_{y}^{x_{35}} g_{Z_3|Z_1}(z)g_{Z_1}(y)dzdy}{\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} g_{Z_1,Z_3}(y,z)dzdy} \leq \frac{\int_{x_{13}}^{x_{15}}\int_{y}^{x_{35}} h_{Z_3|Z_1}(z)h_{Z_1}(y)dzdy}{\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} h_{Z_1,Z_3}(y,z)dzdy}$$

Multiplying the distributions together:

$$\frac{\int_{x_{13}}^{x_{15}}\int_{y}^{x_{35}} g_{Z_1,Z_3}(y,z)dzdy}{\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} g_{Z_1,Z_3}(y,z)dzdy} \leq \frac{\int_{x_{13}}^{x_{15}}\int_{y}^{x_{35}} h_{Z_1,Z_3}(y,z)dzdy}{\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} h_{Z_1,Z_3}(y,z)dzdy}$$

As with above, by Proposition 2 and symmetry of the joint distribution of Z_1 and Z_3 about the line Z_1=Z_3, the left side of the inequality is ≥ 1. Therefore, the following inequality must also be true:

$$1 \leq \frac{\int_{x_{13}}^{x_{15}}\int_{y}^{x_{35}} h_{Z_1,Z_3}(y,z)dzdy}{\int_{x_{11}}^{x_{13}}\int_{x_{31}}^{y} h_{Z_1,Z_3}(y,z)dzdy}$$

Or equivalently:

$$\int_{x_{11}}^{x_{13}} \int_{x_{31}}^{y} h_{Z_1,Z_3}(y,z)dzdy \leq \int_{x_{13}}^{x_{15}} \int_{y}^{x_{35}} h_{Z_1,Z_3}(y,z)dzdy \tag{4.9}$$

Under the alternative, the conditional distribution of Z$_2$ given Z$_1$ = a and Z$_3$ = c is normal with mean (μ_2^{**}) and variance equal to the variance under the null:

$$\mu_2^{**} = \frac{(\sigma_{12} - \sigma_{13}\sigma_{23})a + (\sigma_{23} - \sigma_{12}\sigma_{13})(c - \mu_3)}{1 - \sigma_{13}^2}$$

As above under the global null hypothesis, the conditional mean of Z$_2$ given Z$_1$ and Z$_3$ under the alternative increases as Z$_1$=a increases and increases as Z$_3$=c increases. Stochastic ordering between the red and blue areas in Figure 4.6 is therefore preserved under the alternative. Thus, the inequality 4.4 is true under the alternative as well. i.e.,:

$$\int_{H^{-1}_{(\alpha)}}^{\infty} h_{Z_2|Z_1=a_1,Z_2=c_1}(x_2)dx_2 \leq \int_{H^{-1}_{(\alpha)}}^{\infty} h_{Z_2|Z_1=a_2,Z_2=c_2}(x_2)dx_2 \tag{4.10}$$

Multiplying the inequalities 4.9 and 4.10 we have the following:

$$\int_{x_{11}}^{x_{13}} \int_{x_{31}}^{y} \int_{H^{-1}_{(\alpha)}}^{\infty} h_{Z_2|Z_1,Z_3}(w)h_{Z_1,Z_3}(y,z)dwdzdy \leq \int_{x_{13}}^{x_{15}} \int_{y}^{x_{35}} \int_{H^{-1}_{(\alpha)}}^{\infty} h_{Z_2|Z_1,Z_3}(w)h_{Z_1,Z_3}(y,z)dwdzdy$$

Or equivalently:

$$\int_{x_{11}}^{x_{13}} \int_{x_{31}}^{y} \int_{H^{-1}_{(\alpha)}}^{\infty} h_{Z_1,Z_2,Z_3}(y,w,z)dwdzdy \leq \int_{x_{13}}^{x_{15}} \int_{y}^{x_{35}} \int_{H^{-1}_{(\alpha)}}^{\infty} h_{Z_1,Z_2,Z_3}(y,w,z)dwdzdy \tag{4.11}$$

The left side of the inequality 4.11 is the probability associated with the triangular prism belonging only to the extended single-step Simes procedure under the alternative and the right side of the inequality is the probability associated with the triangular prism belonging only to the Hochberg procedure. Since the probability of the extended single-step Simes is no more than that of the Hochberg, we can say, without loss of generality, that the P(reject $H_1 \cap H_2) \leq \alpha$. Therefore, the extended single-step Simes procedure has strong control of the family wise error rate for three standard correlated normal test statistics whose joint distribution has the MTP_2 property.

Chapter 5

Simulations

Simulations for various distributions and number of test statistics for a number of configurations are provided. 10 million simulations were generated for each set of conditions using a one-sided type 1 error rate of α=0.05.

5.1 Corroborative Results

5.1.1 Three Independent Normal Statistics

Independent normals were generated using the SAS® function RANNOR.

Table 5.1 shows simulations results comparing type 1 error rates for the extended single-step Simes (S), Hochberg (HB), and Hommel (HM) procedures for testing three independent normal statistics with one false hypothesis, whereas Table 5.2 shows the power to reject at least one, at least two, and all three of the hypotheses for each of the procedures.

Table 5.1: Three Independent Normal Test Statistics with 1 False Hypothesis (Type 1 Error Rates)

Mean False Hypothesis	S	HB	HM
0.5	0.0340	0.0344	0.0348
1	0.0347	0.0361	0.0368
1.5	0.0352	0.0387	0.0395
2	0.0355	0.0419	0.0427
2.5	0.0355	0.0451	0.0458
3	0.0350	0.0475	0.0479
3.5	0.0346	0.0490	0.0492

Table 5.2: Three Independent Normal Test Statistics with 1 False Hypothesis (Power)

	S - Reject at Least			HB - Reject at Least			HM - Reject at Least		
Mean False Hypothesis	1	2	3	1	2	3	1	2	3
0.5	0.0847	0.0022	0.0003	0.0836	0.0043	0.0003	0.0843	0.0043	0.0003
1	0.1611	0.0050	0.0006	0.1593	0.0092	0.0006	0.1605	0.0092	0.0006
1.5	0.2927	0.0096	0.0011	0.2902	0.0167	0.0011	0.2919	0.0167	0.0011
2	0.4713	0.0159	0.0016	0.4685	0.0263	0.0016	0.4704	0.0263	0.0016
2.5	0.6603	0.0226	0.0020	0.6580	0.0356	0.0020	0.6597	0.0356	0.0020
3	0.8170	0.0279	0.0023	0.8155	0.0427	0.0023	0.8166	0.0427	0.0023
3.5	0.9190	0.0314	0.0024	0.9181	0.0470	0.0024	0.9187	0.0470	0.0024

Type 1 error rates for all three procedures were controlled at level α=0.05. The extended single-step Simes produced lower type 1 error rates than the Hochberg and Hommel procedures. This was especially true as the mean associated with the false hypothesis increased. The Hochberg and Hommel procedure type 1 error rates approached α as the mean associated with the false hypothesis increased, whereas the type 1 error rate for the extended single-step Simes remained close to 70% of α for all configurations.

The extended single-step Simes procedure showed improvement in the probability to reject at least one hypothesis as compared to the Hochberg and Hommel procedures. The Hochberg and Hommel procedures perform better for the rejection of at least two hypotheses. All three methods are equivalent for the rejection of all three hypotheses.

5.1.2 Trivariate MTP_2 Normal Statistics

Trivariate MTP_2 normal statistics were generated by first generating a vector of independent normals and then multiplying that by the Cholesky decomposition of the covariance matrix.

Table 5.3 shows simulations results comparing type 1 error rates for the extended single-step Simes (S), Hochberg (B), and Hommel (M) procedures for testing three MTP_2 correlated normal statistics with one false hypothesis. Table 5.4 shows the power to reject at least one, at least two, and all three hypotheses for each of the procedures.

Table 5.3: Three MTP_2 Normal Test Statistics with 1 False Hypothesis[1] (Type 1 Error Rates)

Correlations[2]	S	HB	HM
0.1, 0.1, 0.1	0.0354	0.0461	0.0467
0.3, 0.3, 0.3	0.0353	0.0471	0.0475
0.5, 0.5, 0.5	0.0350	0.0468	0.0470
0.7, 0.7, 0.7	0.0347	0.0449	0.0449
0.1, 0.2, 0.3	0.0353	0.0475	0.0480
0.1, 0.3, 0.5	0.0349	0.0485	0.0487
0.1, 0.4, 0.7	0.0347	0.0492	0.0494

[1] False mean = 2.5 for all simulations.

[2] Correlations are ρ_{12}, ρ_{13}, and ρ_{23}.

Table 5.4: Three MTP_2 Normal Test Statistics with 1 False Hypothesis[1] (Power)

	S - Reject at Least			HB - Reject at Least			HM - Reject at Least		
Correlations[2]	1	2	3	1	2	3	1	2	3
0.1, 0.1, 0.1	0.6568	0.0257	0.0033	0.6548	0.0392	0.0033	0.6562	0.0392	0.0033
0.3, 0.3, 0.3	0.6512	0.0308	0.0070	0.6498	0.0443	0.0070	0.6508	0.0443	0.0070
0.5, 0.5, 0.5	0.6470	0.0336	0.0121	0.6463	0.0462	0.0121	0.6468	0.0462	0.0121
0.7, 0.7, 0.7	0.6454	0.0346	0.0195	0.6453	0.0449	0.0195	0.6454	0.0449	0.0195
0.1, 0.2, 0.3	0.6526	0.0293	0.0036	0.6511	0.0437	0.0036	0.6522	0.0437	0.0036
0.1, 0.3, 0.5	0.6491	0.0317	0.0037	0.6480	0.0466	0.0037	0.6488	0.0466	0.0037
0.1, 0.4, 0.7	0.6470	0.0332	0.0037	0.6464	0.0484	0.0037	0.6469	0.0484	0.0037

[1] False mean = 2.5 for all simulations.

[2] Correlations are ρ_{12}, ρ_{13}, and ρ_{23}.

Type 1 error rates for all three procedures were controlled at level α=0.05. The extended single-step Simes produced lower type 1 error rates than the Hochberg and Hommel procedures. This was especially true as the mean associated with the false hypothesis increased. Similar to three independent normals, the Hochberg and Hommel procedure type 1 error rates approached α as the mean associated with the false hypothesis increased, whereas the type 1 error rate for the extended single-step Simes remained close to 70% of α for all configurations.

The extended single-step Simes procedure showed improvement in the probability to reject at least one hypothesis as compared to the Hochberg and Hommel procedures for all configurations tested. The Hochberg and Hommel procedures perform better for the rejection of at least two hypotheses. All three methods are equivalent for the rejection of all three hypotheses.

5.2 Exploratory Configurations and Distributions

5.2.1 More Than Three Independent Normal Statistics

Table 5.5 shows simulations results comparing type 1 error rates for the extended single-step Simes (S), Hochberg (HB), and Hommel (HM) procedures for testing more than three independent normal statistics with one false hypothesis, whereas Table 5.6 shows the power to reject at least one, at least two, and at least three of the hypotheses for each of the procedures.

Table 5.5: More Than Three Independent Normal Test Statistics with 1 False Hypothesis[1] (Type 1 Error Rates)

Number of Hypotheses	S	HB	HM
30	0.049753	0.047764	0.047970
31	0.049785	0.047750	0.047955
32	0.049821	0.047792	0.047982
33	0.049856	0.047792	0.047975
34	0.050042	0.047962	0.048145
35	0.049885	0.047781	0.047963
36	0.049998	0.047932	0.048100
37	0.049931	0.047826	0.047993
38	0.049933	0.047857	0.048020
39	0.050130	0.047974	0.048134
40	0.049941	0.047842	0.047990
50	0.050339	0.048144	0.048260
60	0.050382	0.048128	0.048222
70	0.050450	0.048219	0.048296
80	0.050604	0.048388	0.048461
90	0.050671	0.048423	0.048482
100	0.050645	0.048401	0.048455

[1] False mean = 2.5 for all simulations.

Table 5.6: More Than Three Independent Normal Test Statistics with 1 False Hypothesis[1] (Power)

	S - Reject at Least			HB - Reject at Least			HM - Reject at Least		
Number of Hypotheses	1	2	3	1	2	3	1	2	3
30	0.3679	0.0164	0.0004	0.3632	0.0171	0.0004	0.3635	0.0172	0.0004
31	0.3645	0.0163	0.0004	0.3598	0.0170	0.0004	0.3602	0.0170	0.0004
32	0.3612	0.0161	0.0004	0.3565	0.0168	0.0004	0.3569	0.0168	0.0004
33	0.3580	0.0160	0.0004	0.3533	0.0166	0.0004	0.3536	0.0167	0.0004
34	0.3548	0.0159	0.0004	0.3502	0.0165	0.0004	0.3505	0.0165	0.0004
35	0.3515	0.0157	0.0004	0.3468	0.0163	0.0004	0.3471	0.0163	0.0004
36	0.3489	0.0156	0.0004	0.3443	0.0162	0.0004	0.3446	0.0162	0.0004
37	0.3461	0.0154	0.0004	0.3415	0.0159	0.0004	0.3418	0.0160	0.0004
38	0.3433	0.0153	0.0004	0.3387	0.0158	0.0004	0.3390	0.0159	0.0004
39	0.3408	0.0152	0.0004	0.3362	0.0158	0.0004	0.3365	0.0158	0.0004
40	0.3381	0.0150	0.0003	0.3336	0.0155	0.0004	0.3338	0.0155	0.0004
50	0.3166	0.0141	0.0003	0.3121	0.0145	0.0004	0.3123	0.0145	0.0004
60	0.2994	0.0133	0.0003	0.2951	0.0136	0.0003	0.2952	0.0136	0.0003
70	0.2858	0.0127	0.0003	0.2816	0.0129	0.0003	0.2817	0.0129	0.0003
80	0.2746	0.0122	0.0003	0.2705	0.0124	0.0003	0.2707	0.0124	0.0003
90	0.2649	0.0117	0.0003	0.2608	0.0119	0.0003	0.2609	0.0119	0.0003
100	0.2564	0.0113	0.0003	0.2524	0.0115	0.0003	0.2525	0.0115	0.0003

[1] False mean = 2.5 for all simulations.

The extended single-step Simes procedure did not control the type 1 error rate for all numbers of hypotheses tested. The type 1 error rate first exceeded α=0.05 at 34 hypotheses. However, those that failed to control the type 1 error rate only exceeded α by a small margin (<1.5% of α for all configurations).

The extended single-step Simes procedure showed improvement in the probability to reject at least one hypothesis as compared to the Hochberg and Hommel procedures. The Hochberg and Hommel procedures perform better for the rejection of at least two and at least three hypotheses.

5.2.2 More Than Three MTP_2 Normal Statistics

Table 5.7 shows simulations results comparing type 1 error rates for the extended single-step Simes (S), Hochberg (HB), and Hommel (HM) procedures for testing more than three MTP_2 correlated normal statistics with one false hypothesis, whereas Table 5.8 shows the power to reject at least one, at least two, and at least three of the hypotheses for each of the procedures.

Table 5.7: More Than Three MTP_2[1] Normal Test Statistics with 1 False Hypothesis[2] (Type 1 Error Rates)

Number of Hypotheses	S	HB	HM
10	0.0345	0.0368	0.0382
20	0.0325	0.0328	0.0336
30	0.0313	0.0309	0.0316
40	0.0302	0.0295	0.0300
50	0.0294	0.0285	0.0289
60	0.0286	0.0276	0.0280
70	0.0282	0.0271	0.0274
80	0.0277	0.0264	0.0267
90	0.0272	0.0259	0.0262
100	0.0268	0.0254	0.0257

[1] 0.5 used for all correlations.

[2] False mean = 2.5 for all simulations.

Table 5.8: More Than Three MTP$_2$[1] Normal Test Statistics with 1 False Hypothesis[2] (Power)

	S - Reject at Least			HB - Reject at Least			HM - Reject at Least		
Number of Hypotheses	1	2	3	1	2	3	1	2	3
10	0.4741	0.0313	0.0074	0.4723	0.0346	0.0097	0.4727	0.0360	0.0109
20	0.3849	0.0285	0.0082	0.3828	0.0299	0.0093	0.3830	0.0307	0.0101
30	0.3378	0.0267	0.0084	0.3355	0.0276	0.0090	0.3357	0.0282	0.0097
40	0.3068	0.0253	0.0084	0.3044	0.0258	0.0089	0.3045	0.0263	0.0095
50	0.2844	0.0242	0.0083	0.2819	0.0247	0.0087	0.2820	0.0251	0.0092
60	0.2668	0.0233	0.0082	0.2643	0.0236	0.0085	0.2644	0.0240	0.0090
70	0.2529	0.0227	0.0082	0.2503	0.0229	0.0085	0.2505	0.0233	0.0089
80	0.2413	0.0220	0.0081	0.2386	0.0222	0.0083	0.2387	0.0225	0.0088
90	0.2311	0.0214	0.0080	0.2284	0.0216	0.0082	0.2285	0.0219	0.0086
100	0.2223	0.0209	0.0080	0.2196	0.0211	0.0081	0.2197	0.0214	0.0085

[1] 0.5 used for all correlations.

[2] False mean = 2.5 for all simulations.

Type 1 error rates for all three procedures were considerably below α=0.05 for all configurations tested. Type 1 error rates were similar for the three procedures with the Hochberg procedure producing slightly lower rates. The Type 1 error rates decreased for all procedures as the number of hypotheses increased. All type 1 error rates for the extended single-step Simes procedure were less than 65% of α for the configurations tested.

The extended single-step Simes procedure showed improvement in the probability to reject at least one hypothesis as compared to the Hochberg and Hommel procedures for all configurations tested. The Hochberg and Hommel procedures perform better for the rejection of at least two and at least three hypotheses.

5.2.3 Three Independent Chi-Square Statistics

Independent central chi-square statistics with 1 degree of freedom were generated by squaring randomly generated independent standard normals. 1 degree of freedom non-central chi-square statistics under the alternative were generated by first generating central chi-square statistics, then calculating the associated p-value using the SAS® function PROBCHI, and finally calculating the 1 degree of freedom non-central chi-square statistic associated with the p-value for the specified non-centrality parameter using the SAS® function CINV. Independent chi-square statistics with 5 degrees of freedom were generated by first generating independent chi-square statistics with 1 degree of freedom, then calculating the associated p-value, and finally calculating the 5 degree of freedom central (null) or non-central (alternative) chi-square statistics associated with the p-value for the specified non-centrality parameter, if under the alternative.

Table 5.9 and Table 5.11 show simulations results comparing type 1 error rates for the extended single-step Simes (S), Hochberg (HB), and Hommel (HM) procedures for testing three independent 1 and 5 degree of freedom chi-square statistics, respectively, with one false hypothesis, whereas Table 5.10 and Table 5.12 show the power to reject at least one, at least two, and all three of the hypotheses for each of the procedures.

Table 5.9: Three Independent 1 Degree of Freedom Chi-Square Test Statistics with 1 False Hypothesis (Type 1 Error Rates)

Non-Centrality Parameter False Hypothesis	S	HB	HM
1	0.0341	0.0350	0.0354
3.375	0.0351	0.0390	0.0398
4	0.0353	0.0401	0.0409
6.25	0.0353	0.0433	0.0440
9	0.0351	0.0462	0.0468
12.25	0.0348	0.0483	0.0486

Table 5.10: Three Independent 1 Degree of Freedom Chi-Square Test Statistics with 1 False Hypothesis (Power)

	S - Reject at Least			HB - Reject at Least			HM - Reject at Least		
Non-Centrality Parameter False Hypothesis	1	2	3	1	2	3	1	2	3
1	0.1141	0.0033	0.0004	0.1128	0.0061	0.0004	0.1137	0.0061	0.0004
3.375	0.3158	0.0104	0.0011	0.3135	0.0177	0.0011	0.3151	0.0177	0.0011
4	0.3719	0.0124	0.0013	0.3694	0.0208	0.0013	0.3712	0.0208	0.0013
6.25	0.5609	0.0190	0.0018	0.5585	0.0304	0.0018	0.5602	0.0304	0.0018
9	0.7394	0.0252	0.0021	0.7376	0.0390	0.0021	0.7389	0.0390	0.0021
12.25	0.8716	0.0298	0.0023	0.8705	0.0449	0.0023	0.8713	0.0449	0.0023

Table 5.11: Three Independent 5 Degree of Freedom Chi-Square Test Statistics with 1 False Hypothesis (Type 1 Error Rates)

Non-Centrality Parameter False Hypothesis	S	HB	HM
1	0.0339	0.0341	0.0344
3.375	0.0346	0.0361	0.0366
4	0.0348	0.0367	0.0374
6.25	0.0352	0.0391	0.0398
9	0.0355	0.0420	0.0428
12.25	0.0353	0.0449	0.0455

Table 5.12: Three Independent 5 Degree of Freedom Chi-Square Test Statistics with 1 False Hypothesis (Power)

	S - Reject at Least			HB - Reject at Least			HM - Reject at Least		
Non-Centrality Parameter False Hypothesis	1	2	3	1	2	3	1	2	3
1	0.0730	0.0017	0.0002	0.0721	0.0034	0.0002	0.0727	0.0034	0.0002
3.375	0.1617	0.0050	0.0006	0.1600	0.0090	0.0006	0.1611	0.0090	0.0006
4	0.1919	0.0060	0.0007	0.1900	0.0108	0.0007	0.1913	0.0108	0.0007
6.25	0.3147	0.0104	0.0011	0.3124	0.0177	0.0011	0.3140	0.0177	0.0011
9	0.4762	0.0160	0.0016	0.4738	0.0262	0.0016	0.4755	0.0262	0.0016
12.25	0.6487	0.0221	0.0020	0.6466	0.0348	0.0020	0.6481	0.0348	0.0020

Type 1 error rates for all three procedures were controlled at level α=0.05. The extended single-step Simes procedure produced lower type 1 error rates than the Hochberg and Hommel procedures for both 1 and 5 degree of freedom independent chi-square tests. This was especially true as the non-centrality parameter associated with the false hypothesis increased. The Hochberg and Hommel procedure type 1 error rates approached α as the non-centrality parameter associated with the false hypothesis increased. This was more apparent with the 1 degree of freedom test statistics as compared to the 5 degree of freedom test statistics. However, the type 1 error rate for the extended single-step Simes remained close to 70% of α for all configurations.

The extended single-step Simes procedure showed improvement in the probability to reject at least one hypothesis as compared to the Hochberg and Hommel procedures for both 1 and 5 degree of freedom chi-square test statistics. The Hochberg and Hommel procedures perform better for the rejection of at least two hypotheses. All three methods are equivalent for the rejection of all three hypotheses.

5.2.4 Three Positively Correlated Chi-Square Statistics

Positively correlated central chi-square statistics were generated by first generating positively correlated normal statistics with correlations equal to the square root of the desired correlations. These correlated normal statistics were squared to create 1 degree of freedom central chi-square statistics with the desired correlation structure. Positively correlated 5 degree of freedom central chi-square statistics were generated by summing 5 independently generated sets of positively correlated 1 degree of freedom central chi-square statistics. For each, the non-central chi-square statistic under the alternative was generated by calculating the p-value associated with the central chi-square statistic and its degrees of freedom using the SAS® function PROBCHI and then calculating the non-central chi-square statistic associated with that p-value with the same degrees of freedom and specified non-centrality parameter using the SAS® function CINV.

Table 5.13 and Table 5.15 show simulations results comparing type 1 error rates for the extended single-step Simes (S), Hochberg (HB), and Hommel (HM) procedures for testing three positively correlated 1 and 5 degree of freedom chi-square statistics, respectively, with one false hypothesis, whereas Table 5.14 and Table 5.16 show the power to reject at least one, at least two, and all three of the hypotheses for each of the procedures.

Table 5.13: Three Positively Correlated 1 Degree of Freedom Chi-Square Test Statistics with 1 False Hypothesis[1] (Type 1 Error Rates)

Correlations[2]	S	HB	HM
0.1, 0.1, 0.1	0.0352	0.0440	0.0446
0.3, 0.3, 0.3	0.0348	0.0447	0.0450
0.5, 0.5, 0.5	0.0346	0.0446	0.0447
0.7, 0.7, 0.7	0.0346	0.0431	0.0431
0.1, 0.2, 0.3	0.0350	0.0457	0.0461
0.1, 0.3, 0.5	0.0348	0.0474	0.0477
0.1, 0.4, 0.7	0.0346	0.0484	0.0485

[1] Non-centrality parameter = 6.25 for all simulations.

[2] Correlations are ρ_{12}, ρ_{13}, and ρ_{23}.

Table 5.14: Three Positively Correlated 1 Degree of Freedom Chi-Square Test Statistics with 1 False Hypothesis[1] (Power)

	S - Reject at Least			HB - Reject at Least			HM - Reject at Least		
Correlations[2]	1	2	3	1	2	3	1	2	3
0.1, 0.1, 0.1	0.5559	0.0234	0.0044	0.5539	0.0350	0.0044	0.5553	0.0350	0.0044
0.3, 0.3, 0.3	0.5487	0.0296	0.0102	0.5474	0.0410	0.0102	0.5483	0.0410	0.0102
0.5, 0.5, 0.5	0.5445	0.0331	0.0168	0.5439	0.0438	0.0168	0.5443	0.0438	0.0168
0.7, 0.7, 0.7	0.5428	0.0344	0.0244	0.5427	0.0430	0.0244	0.5428	0.0430	0.0244
0.1, 0.2, 0.3	0.5508	0.0281	0.0049	0.5493	0.0407	0.0049	0.5504	0.0407	0.0049
0.1, 0.3, 0.5	0.5469	0.0313	0.0051	0.5459	0.0450	0.0051	0.5466	0.0450	0.0051
0.1, 0.4, 0.7	0.5442	0.0330	0.0051	0.5436	0.0473	0.0051	0.5440	0.0473	0.0051

[1] Non-centrality parameter = 6.25 for all simulations.

[2] Correlations are ρ_{12}, ρ_{13}, and ρ_{23}.

Table 5.15: Three Positively Correlated 5 Degree of Freedom Chi-Square Test Statistics with 1 False Hypothesis[1] (Type 1 Error Rates)

Correlations[2]	S	HB	HM
0.1, 0.1, 0.1	0.0353	0.0403	0.0410
0.3, 0.3, 0.3	0.0353	0.0422	0.0428
0.5, 0.5, 0.5	0.0352	0.0431	0.0436
0.7, 0.7, 0.7	0.0349	0.0431	0.0433
0.1, 0.2, 0.3	0.0355	0.0426	0.0433
0.1, 0.3, 0.5	0.0356	0.0449	0.0455
0.1, 0.4, 0.7	0.0353	0.0468	0.0472

[1] Non-centrality parameter = 6.25 for all simulations.

[2] Correlations are ρ_{12}, ρ_{13}, and ρ_{23}.

Table 5.16: Three Positively Correlated 5 Degree of Freedom Chi-Square Test Statistics with 1 False Hypothesis[1] (Power)

	S - Reject at Least			HB - Reject at Least			HM - Reject at Least		
Correlations[2]	1	2	3	1	2	3	1	2	3
0.1, 0.1, 0.1	0.3104	0.0151	0.0029	0.3080	0.0235	0.0029	0.3096	0.0235	0.0029
0.3, 0.3, 0.3	0.3026	0.0232	0.0078	0.3004	0.0331	0.0078	0.3020	0.0331	0.0078
0.5, 0.5, 0.5	0.2958	0.0296	0.0143	0.2940	0.0396	0.0143	0.2952	0.0396	0.0143
0.7, 0.7, 0.7	0.2904	0.0336	0.0223	0.2894	0.0425	0.0223	0.2901	0.0425	0.0223
0.1, 0.2, 0.3	0.3044	0.0209	0.0038	0.3021	0.0312	0.0038	0.3037	0.0312	0.0038
0.1, 0.3, 0.5	0.2994	0.0259	0.0043	0.2972	0.0380	0.0043	0.2987	0.0380	0.0043
0.1, 0.4, 0.7	0.2947	0.0297	0.0044	0.2930	0.0431	0.0044	0.2942	0.0431	0.0044

[1] Non-centrality parameter = 6.25 for all simulations.

[2] Correlations are ρ_{12}, ρ_{13}, and ρ_{23}.

Type 1 error rates for all three procedures were controlled at level α=0.05. The extended single-step Simes produced lower type 1 error rates than the Hochberg and Hommel procedures for both 1 and 5 degree of freedom positively correlated chi-square tests. This was especially true as the non-centrality parameter associated with the false hypothesis increased. The Hochberg and Hommel procedure type 1 error rates approached α as the non-centrality parameter associated with the false hypothesis increased. This was more apparent with the 1 degree of freedom test statistics as compared to the 5 degree of freedom test statistics. However, the type 1 error rate for the extended single-step Simes remained close to 70% of α for all configurations.

The extended single-step Simes procedure showed improvement in the probability to reject at least one hypothesis as compared to the Hochberg and Hommel procedures for both 1 and 5 degree of freedom chi-square test statistics. The Hochberg and Hommel procedures perform better for the rejection of at least two hypotheses. All three methods are equivalent for the rejection of all three hypotheses.

5.2.5 More Than Three Independent Chi-Square Statistics

Table 5.17 and Table 5.19 show simulations results comparing type 1 error rates for the extended single-step Simes (S), Hochberg (HB), and Hommel (HM) procedures for testing more than three independent 1 and 5 degree of freedom chi-square statistics, respectively, with one false hypothesis, whereas Table 5.18 and Table 5.20 show the power to reject at least one, at least two, and at least three of the hypotheses for each of the procedures.

Table 5.17: More Than Three Independent 1 Degree of Freedom Chi-Square Test Statistics with 1 False Hypothesis[1] (Type 1 Error Rates)

Number of Hypotheses	S	HB	HM
35	0.04987	0.04785	0.04801
36	0.04989	0.04788	0.04803
37	0.04993	0.04789	0.04804
38	0.04985	0.04783	0.04798
39	0.04986	0.04783	0.04797
40	0.04998	0.04793	0.04807
41	0.04998	0.04792	0.04805
42	0.04992	0.04785	0.04798
43	0.04997	0.04789	0.04802
44	0.05005	0.04799	0.04812
45	0.05008	0.04800	0.04812
50	0.05016	0.04805	0.04815
60	0.05030	0.04823	0.04831
70	0.05034	0.04826	0.04833
80	0.05034	0.04829	0.04836
90	0.05043	0.04836	0.04841
100	0.05037	0.04829	0.04835

[1] Non-centrality parameter = 6.25 for all simulations.

Table 5.18: More Than Three Independent 1 Degree of Freedom Chi-Square Test Statistics with 1 False Hypothesis[1] (Power)

	S - Reject at Least			HB - Reject at Least			HM - Reject at Least		
Number of Hypotheses	1	2	3	1	2	3	1	2	3
35	0.2855	0.0125	0.0003	0.2814	0.0130	0.0003	0.2817	0.0131	0.0003
36	0.2829	0.0124	0.0003	0.2789	0.0129	0.0003	0.2791	0.0129	0.0003
37	0.2806	0.0123	0.0003	0.2765	0.0128	0.0003	0.2767	0.0128	0.0003
38	0.2782	0.0122	0.0003	0.2741	0.0126	0.0003	0.2744	0.0127	0.0003
39	0.2763	0.0121	0.0003	0.2722	0.0125	0.0003	0.2725	0.0126	0.0003
40	0.2739	0.0120	0.0003	0.2698	0.0125	0.0003	0.2701	0.0125	0.0003
41	0.2722	0.0119	0.0003	0.2682	0.0123	0.0003	0.2684	0.0123	0.0003
42	0.2702	0.0118	0.0003	0.2661	0.0122	0.0003	0.2664	0.0122	0.0003
43	0.2681	0.0117	0.0003	0.2641	0.0121	0.0003	0.2643	0.0121	0.0003
44	0.2662	0.0116	0.0003	0.2622	0.0120	0.0003	0.2624	0.0120	0.0003
45	0.2646	0.0115	0.0003	0.2606	0.0119	0.0003	0.2608	0.0119	0.0003
50	0.2562	0.0112	0.0003	0.2523	0.0115	0.0003	0.2525	0.0115	0.0003
60	0.2423	0.0105	0.0002	0.2385	0.0108	0.0003	0.2387	0.0108	0.0003
70	0.2312	0.0100	0.0002	0.2275	0.0102	0.0002	0.2276	0.0102	0.0002
80	0.2219	0.0096	0.0002	0.2183	0.0097	0.0002	0.2184	0.0098	0.0002
90	0.2142	0.0092	0.0002	0.2106	0.0093	0.0002	0.2107	0.0093	0.0002
100	0.2071	0.0089	0.0002	0.2036	0.0090	0.0002	0.2037	0.0090	0.0002

[1] Non-centrality parameter = 6.25 for all simulations.

Table 5.19: More Than Three Independent 5 Degree of Freedom Chi-Square Test Statistics with 1 False Hypothesis[1] (Type 1 Error Rates)

Number of Hypotheses	S	HB	HM
35	0.04937	0.04765	0.04777
36	0.04923	0.04749	0.04760
37	0.04936	0.04762	0.04774
38	0.04940	0.04769	0.04779
39	0.04954	0.04782	0.04792
40	0.04941	0.04765	0.04775
41	0.04954	0.04778	0.04788
42	0.04949	0.04777	0.04786
43	0.04941	0.04769	0.04778
44	0.04943	0.04771	0.04780
45	0.04957	0.04784	0.04792
50	0.04965	0.04793	0.04801
60	0.04978	0.04806	0.04812
70	0.04983	0.04814	0.04819
80	0.04995	0.04830	0.04834
90	0.04997	0.04832	0.04836
100	0.04996	0.04833	0.04836

[1] Non-centrality parameter = 6.25 for all simulations.

Table 5.20: More Than Three Independent 5 Degree of Freedom Chi-Square Test Statistics with 1 False Hypothesis[1] (Power)

	S - Reject at Least			HB - Reject at Least			HM - Reject at Least		
Number of Hypotheses	1	2	3	1	2	3	1	2	3
35	0.1363	0.0053	0.0001	0.1335	0.0055	0.0001	0.1337	0.0056	0.0001
36	0.1349	0.0053	0.0001	0.1321	0.0055	0.0001	0.1323	0.0055	0.0001
37	0.1339	0.0052	0.0001	0.1311	0.0054	0.0001	0.1313	0.0054	0.0001
38	0.1328	0.0051	0.0001	0.1300	0.0054	0.0001	0.1302	0.0054	0.0001
39	0.1318	0.0052	0.0001	0.1291	0.0054	0.0001	0.1292	0.0054	0.0001
40	0.1306	0.0051	0.0001	0.1278	0.0053	0.0001	0.1280	0.0053	0.0001
41	0.1296	0.0050	0.0001	0.1269	0.0052	0.0001	0.1270	0.0052	0.0001
42	0.1287	0.0050	0.0001	0.1260	0.0052	0.0001	0.1262	0.0052	0.0001
43	0.1276	0.0049	0.0001	0.1250	0.0051	0.0001	0.1251	0.0051	0.0001
44	0.1266	0.0049	0.0001	0.1239	0.0051	0.0001	0.1241	0.0051	0.0001
45	0.1258	0.0049	0.0001	0.1232	0.0051	0.0001	0.1233	0.0051	0.0001
50	0.1221	0.0047	0.0001	0.1195	0.0048	0.0001	0.1196	0.0048	0.0001
60	0.1158	0.0044	0.0001	0.1132	0.0045	0.0001	0.1133	0.0045	0.0001
70	0.1106	0.0041	0.0001	0.1081	0.0043	0.0001	0.1082	0.0043	0.0001
80	0.1066	0.0039	0.0001	0.1043	0.0040	0.0001	0.1043	0.0040	0.0001
90	0.1032	0.0038	0.0001	0.1009	0.0039	0.0001	0.1010	0.0039	0.0001
100	0.1002	0.0037	0.0001	0.0979	0.0037	0.0001	0.0980	0.0037	0.0001

[1] Non-centrality parameter = 6.25 for all simulations.

The extended single-step Simes procedure did not control the type 1 error rate for all numbers of hypotheses tested for chi-square tests with 1 degree of freedom. None of the 5 degree of freedom simulations resulted in a type 1 error rate that exceeded α. For the 1 degree of freedom simulations, the type 1 error rate first exceeded α=0.05 at 44 hypotheses. However, those that failed to control the type 1 error rate only exceeded α by a small margin.

The extended single-step Simes procedure showed improvement in the probability to reject at least one hypothesis as compared to the Hochberg and Hommel procedures. The

Hochberg and Hommel procedures perform better for the rejection of at least two and at least three hypotheses.

5.2.6 More Than Three Positively Correlated Chi-Square Statistics

Table 5.21 and Table 5.23 show simulations results comparing type 1 error rates for the extended single-step Simes (S), Hochberg (HB), and Hommel (HM) procedures for testing more than three positively correlated 1 and 5 degree of freedom chi-square statistics, respectively, with one false hypothesis, whereas Table 5.22 and Table 5.24 show the power to reject at least one, at least two, and at least three of the hypotheses for each of the procedures.

Table 5.21: More Than Three Positively Correlated 1 Degree of Freedom Chi-Square Test Statistics with 1 False Hypothesis[1] (Type 1 Error Rates)

Correlations[2]	S	HB	HM
10	0.0277	0.0296	0.0315
20	0.0240	0.0243	0.0255
30	0.0219	0.0217	0.0227
40	0.0204	0.0201	0.0209
50	0.0193	0.0188	0.0196
60	0.0183	0.0178	0.0185
70	0.0178	0.0171	0.0178
80	0.0171	0.0164	0.0171
90	0.0166	0.0159	0.0166
100	0.0163	0.0155	0.0161

[1] Non-centrality parameter = 6.25 for all simulations.

[2] 0.5 used for all correlations.

Table 5.22: More Than Three Positively Correlated 1 Degree of Freedom Chi-Square Test Statistics with 1 False Hypothesis[1] (Power)

	S - Reject at Least			HB - Reject at Least			HM - Reject at Least		
Correlations[2]	1	2	3	1	2	3	1	2	3
10	0.3821	0.0257	0.0091	0.3808	0.0284	0.0116	0.3812	0.0301	0.0135
20	0.3031	0.0219	0.0088	0.3018	0.0229	0.0098	0.3020	0.0242	0.0113
30	0.2627	0.0198	0.0084	0.2613	0.0204	0.0090	0.2614	0.0213	0.0103
40	0.2368	0.0183	0.0080	0.2353	0.0187	0.0085	0.2355	0.0195	0.0097
50	0.2178	0.0171	0.0078	0.2163	0.0174	0.0081	0.2164	0.0182	0.0092
60	0.2031	0.0162	0.0075	0.2016	0.0164	0.0077	0.2017	0.0171	0.0088
70	0.1916	0.0155	0.0073	0.1901	0.0157	0.0075	0.1902	0.0164	0.0085
80	0.1818	0.0149	0.0070	0.1802	0.0150	0.0072	0.1803	0.0157	0.0082
90	0.1737	0.0144	0.0069	0.1722	0.0145	0.0070	0.1722	0.0152	0.0080
100	0.1665	0.0140	0.0067	0.1649	0.0141	0.0069	0.1650	0.0148	0.0079

[1] Non-centrality parameter = 6.25 for all simulations.

[2] 0.5 used for all correlations.

Table 5.23: More Than Three Positively Correlated 5 Degree of Freedom Chi-Square Test Statistics with 1 False Hypothesis[1] (Type 1 Error Rates)

Correlations[2]	S	HB	HM
10	0.0318	0.0316	0.0335
20	0.0290	0.0271	0.0282
30	0.0270	0.0247	0.0255
40	0.0257	0.0231	0.0238
50	0.0247	0.0219	0.0225
60	0.0239	0.0210	0.0215
70	0.0232	0.0202	0.0207
80	0.0226	0.0196	0.0201
90	0.0222	0.0191	0.0196
100	0.0217	0.0185	0.0190

[1] Non-centrality parameter = 6.25 for all simulations.

[2] 0.5 used for all correlations.

Table 5.24: More Than Three Positively Correlated 5 Degree of Freedom Chi-Square Test Statistics with 1 False Hypothesis[1] (Power)

	S - Reject at Least			HB - Reject at Least			HM - Reject at Least		
Correlations[2]	1	2	3	1	2	3	1	2	3
10	0.1785	0.0230	0.0080	0.1745	0.0255	0.0103	0.1756	0.0272	0.0118
20	0.1317	0.0196	0.0080	0.1275	0.0205	0.0089	0.1281	0.0216	0.0101
30	0.1100	0.0175	0.0077	0.1057	0.0181	0.0083	0.1061	0.0190	0.0093
40	0.0969	0.0163	0.0074	0.0925	0.0166	0.0079	0.0929	0.0174	0.0088
50	0.0879	0.0152	0.0072	0.0835	0.0155	0.0075	0.0838	0.0162	0.0084
60	0.0811	0.0144	0.0070	0.0766	0.0146	0.0072	0.0769	0.0153	0.0080
70	0.0758	0.0138	0.0068	0.0714	0.0139	0.0070	0.0716	0.0145	0.0078
80	0.0715	0.0132	0.0066	0.0670	0.0133	0.0068	0.0673	0.0139	0.0076
90	0.0679	0.0129	0.0065	0.0635	0.0130	0.0067	0.0637	0.0135	0.0074
100	0.0648	0.0124	0.0064	0.0604	0.0125	0.0065	0.0606	0.0131	0.0073

[1] Non-centrality parameter = 6.25 for all simulations.

[2] 0.5 used for all correlations.

Type 1 error rates for all three procedures were considerably below α=0.05 for all simulations. Type 1 error rates were similar for the three procedures with the Hochberg procedure producing slightly lower rates. The Type 1 error rates decreased for all procedures as the number of hypotheses increased. All type 1 error rates were less than 60% of α for the configurations tested except for the type 1 error rate for 10 hypotheses, which was slightly higher.

The extended single-step Simes procedure showed improvement in the probability to reject at least one hypothesis as compared to the Hochberg and Hommel procedures for all configurations tested. The Hochberg and Hommel procedures perform better for the rejection of at least two and at least three hypotheses.

5.2.7 Three Independent t Statistics

Independent central t statistics were generated by dividing a randomly generated standard normal by the square root of an independent randomly generated central chi-square with the desired degrees of freedom, divided by the degrees of freedom. Non-central t statistics under the alternative were created by adding the non-centrality parameter to the standard normal prior to dividing.

Table 5.25 shows simulation results comparing type 1 error rates for the extended single-step Simes (S), Hochberg (HB), and Hommel (HM) procedures for testing three independent 30 degree of freedom t statistics with one false hypothesis, whereas Table 5.26 shows the power to reject at least one, at least two, and all three of the hypotheses for each of the procedures.

Table 5.25: Three Independent 30 Degree of Freedom t Test Statistics with 1 False Hypothesis (Type 1 Error Rates)

Non-Centrality Parameter False Hypothesis	S	HB	HM
0.5	0.0340	0.0343	0.0348
1	0.0347	0.0360	0.0367
1.5	0.0353	0.0384	0.0393
2	0.0355	0.0415	0.0424
2.5	0.0356	0.0447	0.0455
3	0.0353	0.0472	0.0477
3.5	0.0348	0.0488	0.0491

Table 5.26: Three Independent 30 Degree of Freedom t Test Statistics with 1 False Hypothesis (Power)

	S - Reject at Least			HB - Reject at Least			HM - Reject at Least		
Non-Centrality Parameter False Hypothesis	1	2	3	1	2	3	1	2	3
0.5	0.0828	0.0021	0.0003	0.0817	0.0042	0.0003	0.0824	0.0042	0.0003
1	0.1536	0.0047	0.0006	0.1518	0.0087	0.0006	0.1530	0.0087	0.0006
1.5	0.2753	0.0091	0.0011	0.2727	0.0159	0.0011	0.2745	0.0159	0.0011
2	0.4425	0.0149	0.0016	0.4396	0.0251	0.0016	0.4416	0.0251	0.0016
2.5	0.6258	0.0214	0.0020	0.6232	0.0343	0.0020	0.6250	0.0343	0.0020
3	0.7857	0.0269	0.0023	0.7838	0.0416	0.0023	0.7852	0.0416	0.0023
3.5	0.8973	0.0307	0.0024	0.8962	0.0463	0.0024	0.8970	0.0463	0.0024

Type 1 error rates for all three procedures were controlled at level α=0.05. The extended single-step Simes procedure produced lower type 1 error rates than the Hochberg and Hommel procedures. This was especially true as the non-centrality parameter associated with the false hypothesis increased. The Hochberg and Hommel procedure type 1 error rates approached α as the non-centrality parameter associated with the false hypothesis increased, whereas the type 1 error rate for the extended single-step Simes remained close to 70% of α for all configurations.

The extended single-step Simes procedure showed improvement in the probability to reject at least one hypothesis as compared to the Hochberg and Hommel procedures. The Hochberg and Hommel procedures perform better for the rejection of at least two hypotheses. All three methods are equivalent for the rejection of all three hypotheses.

5.2.8 Three Positively Correlated t Statistics

Positively correlated central t statistics were generated by first generating positively correlated standard normal statistics. A single independent central chi-square statistic with the desired degrees of freedom (30) was also generated. The positively correlated central t statistics were then created by dividing each of the correlated standard normal statistics by the square root of the one independent central chi-square statistic divided by its degrees of freedom. The non-central t statistic under the alternative was generated by calculating the p-value associated with the central t statistic under the null using the SAS® function PROBT for the given number of degrees of freedom and then calculating the non-central t statistic associated with that p-value with the same degrees of freedom and specified non-centrality parameter using the SAS® function TINV.

Table 5.27 shows simulation results comparing type 1 error rates for the extended single-step Simes (S), Hochberg (B), and Hommel (M) procedures for testing three positively correlated 30 degree of freedom t statistics with one false hypothesis, whereas Table 5.28 shows the power to reject at least one, at least two, and all three hypotheses for each of the procedures.

Table 5.27: Three Positively Correlated 30 Degree of Freedom t Test Statistics with 1 False Hypothesis[1] (Type 1 Error Rates)

Correlations[2]	S	HB	HM
0.1, 0.1, 0.1	0.0354	0.0453	0.0459
0.3, 0.3, 0.3	0.0353	0.0464	0.0468
0.5, 0.5, 0.5	0.0350	0.0464	0.0466
0.7, 0.7, 0.7	0.0346	0.0445	0.0446
0.1, 0.2, 0.3	0.0353	0.0468	0.0473
0.1, 0.3, 0.5	0.0349	0.0479	0.0482
0.1, 0.4, 0.7	0.0347	0.0487	0.0489

[1] Non-centrality parameter = 2.5 for all simulations.

[2] Correlations are ρ_{12}, ρ_{13}, and ρ_{23}.

Table 5.28: Three Positively Correlated 30 Degree of Freedom t Test Statistics with 1 False Hypothesis[1] (Power)

	S - Reject at Least			HB - Reject at Least			HM - Reject at Least		
Correlations[2]	1	2	3	1	2	3	1	2	3
0.1, 0.1, 0.1	0.6227	0.0243	0.0037	0.6205	0.0373	0.0037	0.6221	0.0373	0.0037
0.3, 0.3, 0.3	0.6166	0.0296	0.0075	0.6150	0.0427	0.0075	0.6162	0.0427	0.0075
0.5, 0.5, 0.5	0.6120	0.0330	0.0128	0.6112	0.0453	0.0128	0.6118	0.0453	0.0128
0.7, 0.7, 0.7	0.6101	0.0344	0.0202	0.6099	0.0444	0.0202	0.6100	0.0444	0.0202
0.1, 0.2, 0.3	0.6185	0.0281	0.0041	0.6167	0.0420	0.0041	0.6180	0.0420	0.0041
0.1, 0.3, 0.5	0.6146	0.0308	0.0043	0.6133	0.0453	0.0043	0.6143	0.0453	0.0043
0.1, 0.4, 0.7	0.6117	0.0325	0.0042	0.6109	0.0475	0.0042	0.6115	0.0475	0.0042

[1] Non-centrality parameter = 2.5 for all simulations.

[2] Correlations are ρ_{12}, ρ_{13}, and ρ_{23}.

Type 1 error rates for all three procedures were controlled at level α=0.05. While the Hochberg and Hommel procedures produced type 1 error rates relatively close to α, type 1 error rates for the extended single-step Simes Procedure were close to 70% of α for all configurations and varied less from one configuration to the next than did the Hochberg and Hommel procedures.

The extended single-step Simes procedure showed improvement in the probability to reject at least one hypothesis as compared to the Hochberg and Hommel procedures for all configurations tested. The Hochberg and Hommel procedures perform better for the rejection of at least two hypotheses. All three methods are equivalent for the rejection of all three hypotheses.

5.2.9 More Than Three Independent t Statistics

Table 5.29 shows simulation results comparing type 1 error rates for the extended single-step Simes (S), Hochberg (B), and Hommel (M) procedures for testing more than three independent 30 degree of freedom t statistics with one false hypothesis, whereas Table 5.30

shows the power to reject at least one, at least two, and at least three hypotheses for each of the procedures.

Table 5.29: More Than Three Independent 30 Degree of Freedom t Test Statistics with 1 False Hypothesis[1] (Type 1 Error Rates)

Number of Hypotheses	S	HB	HM
25	0.049760	0.047592	0.047857
26	0.049610	0.047454	0.047707
27	0.049775	0.047553	0.047790
28	0.049846	0.047611	0.047843
29	0.049933	0.047701	0.047919
30	0.049940	0.047692	0.047903
31	0.050017	0.047719	0.047927
32	0.050003	0.047679	0.047873
33	0.049956	0.047658	0.047849
34	0.050069	0.047792	0.047973
35	0.050154	0.047844	0.048023
40	0.050248	0.047878	0.048031
50	0.050453	0.048106	0.048224
60	0.050646	0.048270	0.048369
70	0.050774	0.048420	0.048498
80	0.050648	0.048310	0.048380
90	0.050712	0.048419	0.048478
100	0.050566	0.048278	0.048329

[1] Non-centrality parameter = 2.5 for all simulations.

Table 5.30: More Than Three Independent 30 Degree of Freedom t Test Statistics with 1 False Hypothesis[1] (Power)

	S - Reject at Least			HB - Reject at Least			HM - Reject at Least		
Number of Hypotheses	1	2	3	1	2	3	1	2	3
25	0.3303	0.0145	0.0003	0.3253	0.0154	0.0004	0.3257	0.0154	0.0004
26	0.3254	0.0143	0.0003	0.3204	0.0151	0.0004	0.3209	0.0151	0.0004
27	0.3213	0.0140	0.0003	0.3163	0.0148	0.0004	0.3167	0.0149	0.0004
28	0.3174	0.0139	0.0003	0.3124	0.0146	0.0004	0.3128	0.0147	0.0004
29	0.3131	0.0138	0.0003	0.3082	0.0145	0.0004	0.3085	0.0145	0.0004
30	0.3097	0.0136	0.0003	0.3048	0.0143	0.0004	0.3051	0.0143	0.0004
31	0.3060	0.0134	0.0003	0.3011	0.0140	0.0004	0.3015	0.0140	0.0004
32	0.3024	0.0132	0.0003	0.2975	0.0138	0.0004	0.2979	0.0139	0.0004
33	0.2995	0.0131	0.0003	0.2946	0.0137	0.0003	0.2949	0.0137	0.0003
34	0.2961	0.0130	0.0003	0.2914	0.0135	0.0004	0.2917	0.0136	0.0004
35	0.2934	0.0128	0.0003	0.2886	0.0134	0.0003	0.2889	0.0134	0.0003
40	0.2799	0.0122	0.0003	0.2751	0.0127	0.0003	0.2754	0.0127	0.0003
50	0.2582	0.0112	0.0003	0.2537	0.0116	0.0003	0.2539	0.0116	0.0003
60	0.2417	0.0105	0.0002	0.2374	0.0107	0.0003	0.2375	0.0107	0.0003
70	0.2286	0.0099	0.0002	0.2243	0.0101	0.0002	0.2245	0.0101	0.0002
80	0.2178	0.0094	0.0002	0.2137	0.0095	0.0002	0.2138	0.0096	0.0002
90	0.2087	0.0089	0.0002	0.2047	0.0091	0.0002	0.2048	0.0091	0.0002
100	0.2007	0.0085	0.0002	0.1968	0.0087	0.0002	0.1969	0.0087	0.0002

[1] Non-centrality parameter = 2.5 for all simulations.

The extended single-step Simes procedure did not control the type 1 error rate for all numbers of hypotheses tested. The type 1 error rate first exceeded α=0.05 at 31 hypotheses. However, those that failed to control the type 1 error rate only exceeded α by a small margin (<1.6% of α for all configurations).

The extended single-step Simes procedure showed improvement in the probability to reject at least one hypothesis as compared to the Hochberg and Hommel procedures. The Hochberg and Hommel procedures perform better for the rejection of at least two and at least three hypotheses.

5.2.10 More Than Three Positively Correlated t Statistics

Table 5.31 shows simulation results comparing type 1 error rates for the extended single-step Simes (S), Hochberg (B), and Hommel (M) procedures for testing more than three positively correlated 30 degree of freedom t statistics with one false hypothesis, whereas Table 5.32 shows the power to reject at least one, at least two, and at least three hypotheses for each of the procedures.

Table 5.31: More Than Three Positively Correlated 30 Degree of Freedom t Test Statistics with 1 False Hypothesis[1] (Type 1 Error Rates)

Correlations[2]	S	HB	HM
10	0.0327	0.0345	0.0361
20	0.0301	0.0298	0.0309
30	0.0283	0.0273	0.0281
40	0.0268	0.0256	0.0262
50	0.0258	0.0243	0.0249
60	0.0249	0.0233	0.0238
70	0.0241	0.0224	0.0229
80	0.0236	0.0218	0.0222
90	0.0231	0.0212	0.0216
100	0.0225	0.0206	0.0210

[1] Non-centrality parameter = 2.5 for all simulations.

[2] 0.5 used for all correlations.

Table 5.32: Three Positively Correlated 30 Degree of Freedom t Test Statistics with 1 False Hypothesis[1] (Power)

	S - Reject at Least			HB - Reject at Least			HM - Reject at Least		
Correlations[2]	1	2	3	1	2	3	1	2	3
10	0.4215	0.0284	0.0078	0.4193	0.0315	0.0102	0.4199	0.0330	0.0115
20	0.3265	0.0248	0.0084	0.3239	0.0260	0.0094	0.3242	0.0270	0.0104
30	0.2781	0.0225	0.0083	0.2752	0.0232	0.0089	0.2755	0.0239	0.0097
40	0.2469	0.0208	0.0081	0.2440	0.0213	0.0086	0.2442	0.0219	0.0093
50	0.2245	0.0196	0.0080	0.2215	0.0200	0.0083	0.2217	0.0205	0.0090
60	0.2080	0.0186	0.0078	0.2049	0.0189	0.0081	0.2050	0.0194	0.0087
70	0.1940	0.0178	0.0076	0.1908	0.0180	0.0079	0.1910	0.0185	0.0085
80	0.1830	0.0172	0.0075	0.1798	0.0173	0.0077	0.1800	0.0178	0.0083
90	0.1735	0.0165	0.0074	0.1703	0.0167	0.0075	0.1704	0.0171	0.0081
100	0.1656	0.0160	0.0072	0.1623	0.0161	0.0074	0.1625	0.0165	0.0079

[1] Non-centrality parameter = 2.5 for all simulations.

[2] 0.5 used for all correlations.

Type 1 error rates for all three procedures were considerably below α=0.05 for all simulations. Type 1 error rates were similar for the three procedures with the Hochberg procedure producing slightly lower rates. The Type 1 error rates decreased for all procedures as the number of hypotheses increased. All type 1 error rates were less than 60% of α for the configurations tested.

The extended single-step Simes procedure showed improvement in the probability to reject at least one hypothesis as compared to the Hochberg and Hommel procedures for all configurations tested. The Hochberg and Hommel procedures perform better for the rejection of at least two and at least three hypotheses.

Chapter 6

Conclusion and Future Research

Simes's improved Bonferroni method for testing the global null hypothesis offers improved power over the traditionally Bonferroni procedure. While Simes proved his method controls the type 1 error rate for independent tests, it does not hold for all conditions and distributions. Simes first conjecture was that his method may hold for some dependent statistics with certain multivariate distributions. Simes's global test does not allow for making statements about individual hypotheses. His second conjecture was to reject the individual hypotheses $H_{(1)}$, ..., $H_{(j)}$, where j $= max\{j : P_{(j)} \leq j\alpha/m\}$.

To address Simes's first conjecture, it has since been shown that, in general, the Simes global test has control of the one-sided type 1 error for positively dependent statistics, but may inflate the type 1 error for negatively dependent statistics. However, Hochberg and Rom gave an upper bound for the type 1 error for negatively correlated normal statistics.

As for Simes's second conjecture, it can easily be shown that his proposed method does not control the family wise error rate. That said, the Simes global test has since become the foundation of many p-value based multiple testing procedures that do control the family wise error rate. Hochberg and Hommel are two examples of procedures that have extended the Simes global test to make inferences on individual hypotheses. However, though the procedures by Hochberg and Hommel are based on the Simes global test, they are conservative as they may not be able to reject any of the individual hypotheses when Simes's test rejects the global null hypothesis. The extended single-step Simes procedure presented here rejects a subset of the hypotheses proposed by Simes and fills this void by rejecting at least one hypothesis when the Simes global test rejects making it the most powerful Simes based procedure for the rejection of at least one hypothesis.

The extended single-step Simes procedure controls the family wise error rate for three independent normal test statistics as well as trivariate normal test statistics whose joint distribution has the MTP_2 property. Simulations show that the procedure doesn't control the type 1 error rate for all distributions and configurations for certain numbers of hypotheses. However, type 1 error rates were only slightly inflated above α in the simulations. Simulations did suggest that the extended single-step Simes procedure may control the type 1 error rate for as many as 30 or more hypotheses when the test statistics are non-negatively correlated normals with the MTP_2 property. Further simulations suggest that the procedure may control the type 1 error rate for as many as 30 or more hypotheses for non-negatively correlated chi-square and t test statistics.

For future research, formal proofs showing that the extended single-step Simes procedure controls the family wise error rate for more than three hypotheses or for additional distributions may be of interest. Additionally, as previously stated, the extended single-step Simes procedure does not control the family wise error rate for any number of normal test statistics. If an upper bound for the family wise error rate could be determined, then adjustments could be made in order to control the family wise error rate.

Bibliography

Y. Benjamini and Y. Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.

H. Block, T. Savits, J. Wang, and S. Sarkar. The multivariate-t distribution and the Simes inequality. *Statistics and Probability Letters*, 83:227–232, 2013.

R. A. Fisher. *The Design of Experiments.* Hafner Publishing Company, 8th edition, 1966.

K. R. Gabriel. Simultaneous Test Procedures–Some Theory of Multiple Comparisons. *The Annals of Mathematical Statistics*, 40(1):224–250, February 1969.

Y. Hochberg. A Sharper Bonferroni Procedure for Multiple Tests of Significance. *Biometrika*, 75(4):800–802, December 1988.

Y. Hochberg and D. Rom. Extensions of multiple testing procedures based on Simes' test. *Journal of Statistical Planning and Inference*, 48:141–152, 1995.

Y. Hochberg and A. Tamhane. *Multiple Comparison Procedures.* Wiley, 1987.

G. Hommel. A Stagewise Rejective Multiple Test Procedure Based on a Modified Bonferroni Test. *Biometrika*, 75(2):383–386, June 1988.

G. Hommel. A Comparison of Two Modified Bonferroni Procedures. *Biometrika*, 76(3): 624–625, September 1989.

S. Karlin and Y. Rinnot. Classes of Orderings of Measures and Related Correlation Inequalities. I. Multivariate Totally Positive Distributions. *Journal of Multivariate Analysis*, 10: 467–498, 1980.

S. Karlin and Y. Rinnot. Total Positivity Properties of Absolute Value Multinormal Variables with Applications to Confidence Interval Estimates and Related Probabilistic Inequalities. *The Annals of Statistics*, 9(5):1035–1049, September 1981.

S. Karlin and Y. Rinnot. M-Matrices as covariance matrices of multinormal distributions. *Linear Algebra and its Applications*, 52-53:419–438, July 1983.

R. Marcus, E. Peritz, and K. R. Gabriel. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63(3):655–660, 1976.

D Rom. A Sequentially Rejective Test Procedure Based on a Modified Bonferroni Inequality. *Biometrika*, 77(3):663–665, 1990.

D Rom. An improved Hochberg procedure for multiple tests of significance. *British Journal of Mathematical and Statistical Psychology*, 66:189–196, 2013.

E. Samuel-Cahn. Is the Simes Improved Bonferroni Procedure Conservative? *Biometrika*, 83(4):928–933, December 1996.

E. Samuel-Cahn. A note about a curious generalization of Simes' theorem. *Journal of Statistical Planning and Inference*, 82:147–149, December 1999.

S. Sarkar. SOME PROBABILITY INEQUALITIES FOR ORDERED MTP_2 RANDOM VARIABLES: A PROOF OF THE SIMES CONJECTURE. *The Annals of Statistics*, 26(2):494–504, 1998.

S. Sarkar and C-K Chang. The Simes Method for Multiple Hypothesis Testing With Positively Dependent Test Statistics. *Journal of the American Statistical Association*, 92 (440):1601–1608, December 1997.

S. Sarkar and W. Smith. Probability Inequalities for Ordered MTP_2 Random Variables. *The Indian Journal of Statistics, Series A (1961-2002)*, 48(2):119–135, June 1986.

R. J. Simes. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73:751–754, 1986.

P. H. Westfall, R. D. Tobias, and R. D. Wolfinger. *Multiple Comparisons and Multiple Tests Using SAS*. Cary, NC: SAS Institute Inc., second edition, 2011.