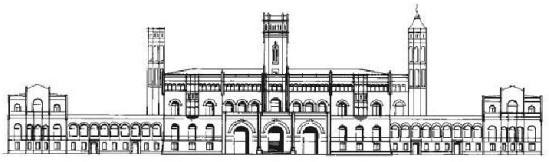
Extensions of Multiple Contrast Tests

Von der Naturwissenschaftlichen Fakultät der

Gottfried Wilhelm Leibniz Universität Hannover



zur Erlangung des
akademischen Grades eines

Doktors der Gartenbauwissenschaften
- Dr. rer. hort. -

genehmigte Dissertation

von

Dipl.-Math. **Mario Hasler**
geboren am 15.01.1978 in Gotha

2009

Referent: Prof. Dr. Ludwig A. Hothorn

Korreferent: PD Dr. Siegfried Kropf

Tag der Promotion: 23. Januar 2009

Acti iucundi labores

Cicero, de finibus 5, 46

Zusammenfassung

Die vorliegende Arbeit befasst sich mit multiplen Kontrasttests für Mittelwerte normalverteilter Daten. Diese haben im Vergleich zu anderen Methoden den Vorteil, dass i) Testaussagen für jeden Einzelvergleich möglich sind, ii) deren Korrelationen berücksichtigt werden, iii) dadurch der Gesamtfehler erster Art eingehalten und ausgeschöpft wird, und iv) sich für jeden Einzelvergleich simultane Konfidenzintervalle ableiten lassen. Dies wird erreicht durch die Verwendung einer gemeinsamen multivariaten t-Verteilung aller zu betrachtenden Vergleiche. Darüber hinaus sind multiple Kontrasttests sowohl für Differenzen als auch für Verhältnisse von Mittelwerten formulierbar. Neben der Normalverteilung ist Varianzhomogenität der Daten allerdings eine weitere Annahme. Zudem sind multiple Kontrasttests beschränkt auf eine zu betrachtende Messgröße (Endpunkt).

Ziel dieser Dissertation ist es zum einen, multiple Kontrasttests für die Anwendung auf varianzheterogene Daten zu erweitern. Hierfür werden drei mögliche Prozeduren vorgestellt und im Hinblick auf die Einhaltung des Gesamtfehlers erster Art verglichen. Ziel ist es weiterhin, multiple Kontrasttests für die simultane Analyse mehrerer Endpunkte zu verallgemeinern. Beide Teilprobleme erfordern die Herleitung entsprechender approximativer multivariater t-Verteilungen. Simulationsstudien zeigen, dass für beide Ansätze der Gesamtfehler erster Art eingehalten werden kann. Die Auswertung von Realdatenbeispielen verdeutlicht die Notwendigkeit der Verfahren und dient ihrer Veranschaulichung.

Schlagworte: multiple Kontrasttests, Heteroskedastizität, multiple Endpunkte

Abstract

This research considers multiple contrast tests for means of normally distributed data. Their advantages, as compared to other methods, are that i) test decisions are available for all individual comparisons, ii) correlations are taken into account, iii) the familywise error type I is maintained and exploited for that reason, and iv) simultaneous confidence intervals can be derived. Therefore, a joint multivariate t-distribution of all comparisons is used. Moreover, multiple contrast tests can be formulated for both differences and ratios of means. Besides following a normal distribution, the data are also assumed to have homogeneous variances. Furthermore, multiple contrast tests are restricted to one single endpoint.

The aim of this dissertation is to facilitate multiple contrast tests in the presence of heteroscedasticity. Three candidate procedures are introduced and compared with regard to their ability to maintain the familywise error type I. On the other hand, an extension for the case of multiple endpoints is investigated. For both tasks, approximate multivariate t-distributions are derived. Simulation studies show that both approaches control the familywise error type I. Real data examples are analyzed in order to demonstrate the necessity of the methods, and to illustrate them.

Keywords: multiple contrast tests, heteroscedasticity, multiple endpoints

Contents

List of Figures

List of Tables

Abbreviations

BON	test procedure with Bonferroni adjustment
FWE	familywise error rate
GH	test procedure based on Games and Howell [1976]
HOM	test procedure for homoscedastic data
HTL	test procedure based on Hochberg and Tamhane [1987] and Tamhane and Logan [2004]
IUT	intersection-union test
MIN	test procedure with minimal degrees of freedom over the endpoints
MCP	multiple comparison procedure
MCT	multiple contrast test
PI	test procedure with plug-in of variance estimators
SCI	simultaneous confidence interval
(S)CP	(simultaneous) coverage probability
UIT	union-intersection test

Chapter 1

General Introduction

Usually, multiple comparison procedures (MCPs) for means of normally distributed populations can be evaluated as multiple contrast tests (MCTs) or by related simultaneous confidence intervals (SCIs). Several contrasts, representing linear functions of these means, are estimated and typically tested for deviation from zero. Because correlations between the contrasts are involved in a joint distribution, MCTs exactly maintain the familywise error rate (FWE) over all contrasts. No further multiplicity-adjustment is needed. The all-pair comparison of Tukey [1953] and the many-to-one comparison of Dunnett [1955] are well-known examples. Bretz [2006] has formulated the trend test of Williams [1971] as an approximate MCT. Also, the user is free to create other interesting problem-specific contrasts. Moreover, Dilba et al. [2004] have dealt with MCTs and SCIs for ratios of means. If relative changes (e.g., in per cent) are of more interest than absolute ones, this approach is suitable.

All the resulting MCTs and SCIs assume homogeneous variances as a general rule. This fact is often attributed to an easier derivation and to mathematical convenience, but it is not always realistic. For example, dose finding studies can have the problem of heteroscedasticity because the data's variance depends on the dose effect (see the data in Westfall [1997]). It is common to apply these procedures without checking the validity of this assumption. If no information about the data is available (e.g. from preliminary tests) before statistical analysis, it is not advisable to presume homogeneous variances. Existing effects or negligible differences may be under- or overestimated, respectively, leading to wrong decisions.

Furthermore, MCTs are restricted to data with a single outcome (endpoint). However, measurements for multiple endpoints frequently appear in experiments (see the data in Schulte et al. [2002]). The number of endpoints must then be taken into account too, for the FWE. Their correlations are also important because, e.g., highly correlated endpoints do not contain the same amount of information about the data as uncorrelated ones.

The outline of this work is as follows. In Chapter 2, basic underlying concepts and distributions are recalled and investigated. Chapter 3 describes and compares adequate approaches to handle the problem of heteroscedasticity, while Chapter 4 deals with an extension for multiple endpoints. Conclusions and a discussion are given in Chapter 5.

Chapter 2

Statistical Concepts and Distributions

Before turning to the main parts of this work, some important underlying concepts and distributions, which build a base of the following methodology, are recalled and investigated.

2.1 Some Basic Concepts

Multiple testing problems first of all raise the question how to construct suitable hypotheses. There are two basic approaches. The intersection-union method of test construction may be useful if the null hypothesis can be conveniently expressed as

a union, that is,

$$H_0 = \bigcup_{i=1}^{k} H_{0i}.$$

Suppose that a suitable test is available for each $H_{0i} : \theta \in \Theta_i$ versus $H_{1i} : \theta \in \Theta_i^c$. We can then write

$$H_0 : \theta \in \bigcup_{i=1}^{k} \Theta_i.$$

Say the rejection region for the test of H_{0i} is $\{x : T_i(x) \in R_i\}$. Then, the rejection region for the intersection-union test (IUT) of H_0 is

$$\bigcap_{i=1}^{k} \{x : T_i(x) \in R_i\}.$$

This means that the global null hypothesis H_0 is rejected if and only if each of its component (local null) hypotheses H_{0i} is rejected. Depending on the test direction, let the local rejection region for each of the individual tests be $\{x : T_i(x) > c\}$ with a common c for all these tests. The global rejection region of the IUT is therefore

$$\bigcap_{i=1}^{k} \{x : T_i(x) > c\} = \{x : \min_{i=1,\ldots,k} T_i(x) > c\}.$$

Thus, the test statistic for testing H_0 is

$$T(x) = \min_{i=1,\ldots,k} T_i(x).$$

Information about the IUT's size is given by the following

Theorem 2.1.1. *Let α_i be the size of the test of H_{0i} with rejection region R_i ($i = 1, \ldots, k$). Then the IUT with rejection region $R = \bigcap_{i=1}^{k} R_i$ is a level-α test, that is, its size is at most α with*

$$\alpha = \max_{i=1,\ldots,k} \alpha_i.$$

Some simple but typical examples for IUT are as follows.

Example 2.1.1. The TOST concept can be used to test the equivalence of two groups by performing **t**wo **o**ne-**s**ided **t**ests. The first test is used to ensure that the groups do not differ by more than a specified positive amount (e.g., δ); the second one is to ensure that they do not differ by more than a specified negative amount (e.g., $-\delta$). Only if both can be shown, equivalence can be stated. The null hypothesis can be expressed as a union of two partial hypotheses. Both can be tested at level α.

Example 2.1.2. If two groups with multiple endpoints have to be compared, the aim may be to show non-inferiority (or superiority) of the first over the second group for all endpoints. For example, when testing for side effects, safety is only declared when all endpoints are safe. That is to say, the first group (new compound) is safe only if it is non-inferior to the second group (control) for all endpoints. Each endpoint is then related to a partial hypothesis; the overall null hypothesis is the union of them. All the endpoints can be tested at level α.

In addition to the intersection-union method of test construction there is the union-intersection method. It is useful if the null hypothesis can be conveniently expressed as an intersection, i.e.,

$$H_0 = \bigcap_{i=1}^{k} H_{0i}.$$

Supposing again that a suitable test is available for each $H_{0i} : \theta \in \Theta_i$ versus $H_{1i} : \theta \in \Theta_i^c$, we can write

$$H_0 : \theta \in \bigcap_{i=1}^{k} \Theta_i.$$

The rejection region for the test of H_{0i} is then $\{x : T_i(x) \in R_i\}$. Hence, the rejection

region for the union-intersection test (UIT) of H_0 is

$$\bigcup_{i=1}^{k}\{x : T_i(x) \in R_i\}.$$

The global null hypothesis H_0 is thus rejected if and only if at least one of its component (local null) hypotheses H_{0i} is rejected. Suppose the test direction for which the local rejection region for each of the individual tests is $\{x : T_i(x) > c\}$ with a common c for all these tests. Then, the global rejection region of the UIT is

$$\bigcup_{i=1}^{k}\{x : T_i(x) > c\} = \{x : \max_{i=1,\ldots,k} T_i(x) > c\},$$

so that the test statistic for testing H_0 is

$$T(x) = \max_{i=1,\ldots,k} T_i(x).$$

Examples for UITs are as follows.

Example 2.1.3. Two-sided testing is used if the aim is to show a difference between two groups, regardless of the algebraic signs of this difference. Formally, two tests are performed with opposite test direction. The first serves to show whether the groups differ by a positive amount; the second one is to show whether they differ by a negative amount. If at least one case can be shown, a significant difference can be stated. The null hypothesis can be expressed as an intersection of two partial hypotheses. Both must be tested at level $\alpha/2$.

Example 2.1.4. The Dunnett procedure [Dunnett, 1955] compares several groups with one control. The null hypothesis is an intersection of partial hypotheses, one for each non-control group. To maintain the error type I for the overall null hypothesis at level α, a multivariate t-distribution is used that takes the number and correlations of the involved comparisons into account.

Example 2.1.5. If two groups with multiple endpoints have to be compared, the aim may be to show for which endpoints they differ. For example, when testing for side effects, the first group (new compound) is declared hazardous if at least one endpoint is hazardous as compared to the second group (control). Then each endpoint is related to a partial hypothesis and the overall null hypothesis is the intersection of them. All endpoints must be tested at level α/k, where k is the number of endpoints.

These and further considerations about IUT and UIT may be found in Casella and Berger [2002]. Of course, testing problems can also be mixtures between IUTs and UITs; see Bofinger and Bofinger [1995] and Quan et al. [2001] for example. The testing problems in the following chapters will turn out to be UITs. We therefore recall some definitions and theorems related to these test procedures. They may also be found in Hochberg and Tamhane [1987] and Gabriel [1969], together with the related proofs which are omitted here for brevity. The IUT was mentioned here for reasons of completeness and is not considered below.

A hypothesis $H_0^{(s)}$ is said to *imply* $H_0^{(r)}$ when the parameter values postulated by $H_0^{(s)}$ form a subset of the parameter values postulated by $H_0^{(r)}$. A family of hypotheses is said to be *hierarchical* if an implication relation holds between at least two hypotheses. If a hypothesis $H_0^{(s)}$ implies $H_0^{(r)}$, then $H_0^{(r)}$ is called a *component* of $H_0^{(s)}$, according to Gabriel [1969]. A hypothesis with no components is called *minimal*; all other hypotheses are called *non-minimal* [Gabriel, 1969]. An MCP is called *coherent*, if the following property holds: If $H_0^{(s)}$ is not rejected, then $H_0^{(r)}$ is

also not rejected for any pair of hypotheses $(H_0^{(r)}, H_0^{(s)})$ such that $H_0^{(s)}$ implies $H_0^{(r)}$. This requirement has been introduced by Gabriel [1969] and earlier by Lehmann [1957] as *compatibility*. A coherent MCP that rejects a hypothesis also rejects all hypotheses implying it. For a hierarchical family of hypotheses, *consonance* refers to the property that whenever any non-minimal hypothesis is rejected, at least one of its components is also rejected (see Gabriel [1969]). An MCP with this property is called *consonant*. A *simultaneous test procedure* for a hierarchical family of hypotheses is characterized by a collection of test statistics Z_s, $s \in I$, and a common critical constant c such that the procedure rejects $H_0^{(s)}$ if $Z_s \geq c$, $s \in I$. The test statistics Z_s are said to be *monotone* if $Z_s \geq Z_r$ with probability one whenever $H_0^{(s)}$ implies $H_0^{(r)}$. The above UIT is such a simultaneous test procedure because it compares each T_i with the same quantile c and Z_s can be defined as $\max_{i=1,\ldots,s} T_i$ and $I = \{1, \ldots, k\}$. The following theorems clarify the need for the above definitions and connects them.

Theorem 2.1.2. *The simultaneous test procedure stated above is coherent for any choice of the critical constant c if and only if the test statistics Z_i are monotone.*

Theorem 2.1.3. *For a hierarchical family of hypotheses, a single-step test procedure is coherent and consonant if and only if it is a UIT.*

Theorem 2.1.4. *The above UIT is a level-α test if c is chosen to be the upper α quantile of the distribution of* $\max_{s \in I} Z_s$.

Furthermore, it can be shown that associated confidence sets have level $(1 - \alpha)$. In summary: If a test procedure is known to be a UIT, constructed in the above manner, it is coherent and consonant; the proper choice of a quantile c guaranties a

level-α test.

2.2 Multivariate Normal and t-distribution

The multivariate normal distribution plays a dominant role in both the historical and actual development of statistical theory. Indeed, even its name points up its central meaning. Its application areas are various. A comprehensive and coherent treatment of classical and new results related to the multivariate normal distribution is provided by Tong [1990], for example.

A random vector $\boldsymbol{X} = (X_1, ..., X_k)'$ is said to have a *k-variate normal distribution* with mean vector $\boldsymbol{\mu} \in \mathbb{R}^k$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{k \times k}$ if its characteristic function $\psi_{\boldsymbol{X}}(u) = E(e^{iu'\boldsymbol{X}})$ is given by

$$\psi_{\boldsymbol{X}}(u) = \exp(iu'\boldsymbol{\mu} - \frac{1}{2}u'\boldsymbol{\Sigma}u) \quad (u \in \mathbb{R})$$

where i is the imaginary unit. We write $\boldsymbol{X} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. If $\boldsymbol{\Sigma}$ is positive definite, then the density function of $\boldsymbol{X}$ exists and is given by

$$\phi(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{k}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})} \quad (\boldsymbol{x} \in \mathbb{R}^k).$$

In this case we say that $\boldsymbol{X}$ has a *nonsingular* distribution. If the X_i $(i = 1, ..., k)$ are standardized so that $E(X_i) = 0$ and $Var(X_i) = 1$, then $\boldsymbol{\Sigma}$ is a correlation matrix denoted by $\boldsymbol{R} = (\rho_{ij})_{i,j}$ with off-diagonal elements $\rho_{ij} = \text{corr}(X_i, X_j)$ for $i \neq j$ and we write $\boldsymbol{X} \sim N_k(\boldsymbol{\mu}, \boldsymbol{R})$.

The multivariate t-distribution is of increasing importance in statistical modeling.

It has been found useful in inference problems concerning the mean vector of a multivariate normal distribution. A large number of modifications and extensions of the standard multivariate t-distribution has been proposed in the literature. A comprehensive review is given by Nadarajah and Dey [2005]. For a more detailed account, see Kotz and Nadarajah [2004]. The following and further considerations may also be found in Tong [1990] or in Hochberg and Tamhane [1987] p.365 ff, together with the related proofs which are omitted here for brevity.

Let $\boldsymbol{X} = (X_1, ..., X_k)' \sim N_k(\boldsymbol{\mu}, \boldsymbol{R})$, let U be a χ^2_ν random variable which is distributed independently of $\boldsymbol{X}$ and denote

$$T_i = \frac{X_i}{\sqrt{U/\nu}} \quad (i = 1, \ldots, k). \tag{2.1}$$

For $\boldsymbol{\mu} = \boldsymbol{0}$ the distribution of $\boldsymbol{T} = (T_1, ..., T_k)'$ is called a *central k-variate t-distribution* with ν degrees of freedom and associated correlation matrix $\boldsymbol{R}$ and we write $\boldsymbol{T} \sim t_k(\nu, \boldsymbol{R})$. Otherwise (2.1) is called a *non-central k-variate t-distribution* with *non-centrality parameter* $\boldsymbol{\mu}$, and is denoted by $\boldsymbol{T} \sim t_{k,\boldsymbol{\mu}}(\nu, \boldsymbol{R})$. If $\boldsymbol{R}$ is positive definite, then the density function of $\boldsymbol{T}$ exists and is given by

$$g_k(\boldsymbol{t}; \nu, \boldsymbol{R}) = \frac{\Gamma(\frac{1}{2}(\nu + k))}{(\pi\nu)^{\frac{k}{2}}\Gamma(\frac{\nu}{2})|\boldsymbol{R}|^{\frac{1}{2}}} \left(1 + \frac{1}{\nu}\,\boldsymbol{t}'\boldsymbol{R}^{-1}\boldsymbol{t}\right)^{-\frac{1}{2}(\nu+k)} \quad (\boldsymbol{t} \in \mathbb{R}^k). \tag{2.2}$$

Important characteristics are given by the following

Lemma 2.2.1. *Let $\boldsymbol{T} \sim t_k(\nu, \boldsymbol{R})$. Then for $\nu > 2$:*

$$E(T_i) = 0 \quad (i = 1, \ldots, k),$$

$$Var(T_i) = \frac{\nu}{\nu - 2} \quad (i = 1, \ldots, k),$$

$$Cov(T_i, T_j) = \frac{\nu\rho_{ij}}{\nu - 2} \quad (i, j = 1, \ldots, k).$$

This result ensures that the matrix $\boldsymbol{R}$ in (2.2) is the correlation matrix of $\boldsymbol{T}$. Furthermore, note that the multivariate t-distribution belongs to the class of elliptically countered distributions. It is ellipsoidally symmetric about $\boldsymbol{\mu}$ (see Tong [1990]). The next lemma delimits the possible correlations, e.g., for simulating random numbers.

Lemma 2.2.2. *Let there be random variables $\boldsymbol{T} \sim t_{k,\boldsymbol{\mu}}(\nu, \boldsymbol{R})$ or $\boldsymbol{X} \sim N_k(\boldsymbol{\mu}, \boldsymbol{R})$, and let the elements of $\boldsymbol{R}$ be $\rho_{ij} = \rho$ for all $i \neq j = 1, \ldots, k$. Then the smallest valid value ρ^{min} for ρ depends on the dimension k according to*

$$\rho^{min} = -\frac{1}{k-1}. \tag{2.3}$$

Proof. A condition for the density function of $\boldsymbol{T}$ or $\boldsymbol{X}$ to exist is that the correlation matrix $\boldsymbol{R}$ is positive definite (positive semidefinite). Hence, each principal minor of $\boldsymbol{R}$ has to be positive (non-negative). Complete induction with respect to k leads to (2.3). □

To give an example, Figure 2.1 illustrates the behavior of a random variable $\boldsymbol{T} \sim t_3(20, \boldsymbol{R})$ depending on its correlation structure. Here, all the variable's components are equicorrelated, $\rho_{ij} = \rho$ for all $i, j = 1, \ldots, k, i \neq j$. The red points representing realizations of $\boldsymbol{T}$ for the case of maximal negative correlation $\rho = -\frac{1}{2}$ lie exactly on a disk. The green points having correlation $\rho = 0$ form a ball, while the blue points have maximal positive correlation, i.e., $\rho = 1$. They are located on a line which is orthogonal to the disk of the red points. A general conclusion is given by the following

Corollary. *The realizations of the k-variate random variables $\boldsymbol{T} \sim t_{k,\boldsymbol{\mu}}(\nu, \boldsymbol{R})$ or $\boldsymbol{X} \sim N_k(\boldsymbol{\mu}, \boldsymbol{R})$ with maximal negative correlation $\rho = -\frac{1}{k-1}$ form a $(k-1)$-*

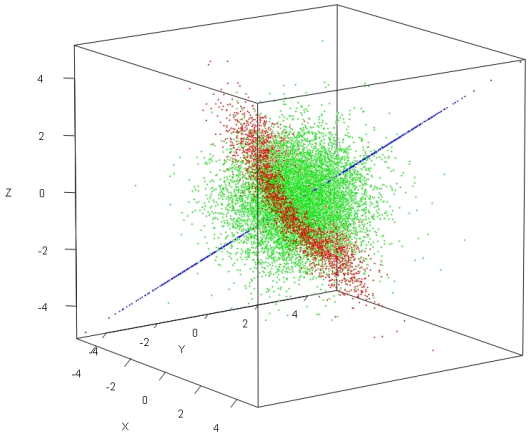


Figure 2.1: *Trivariate t-distributed random variable with* $\nu = 20$ *and maximal negative correlation (red), correlation 0 (green), and maximal positive correlation (blue).*

dimensional subspace of $\mathbb{R}^k$*, for correlation* $\rho = 0$ *a* k*-dimensional ball, and for maximal positive correlation* $\rho = 1$ *a line that is orthogonal to the subspace obtained by maximal negative correlations.*

A relationship between multivariate t- and normal distribution is considered by

Lemma 2.2.3. *Let* $g_k(\boldsymbol{t}; \nu, \boldsymbol{R})$ *and* $f_k(\boldsymbol{x}; \boldsymbol{0}, \boldsymbol{R})$ *be the density functions of the random variables* $\boldsymbol{T} \sim t_k(\nu, \boldsymbol{R})$ *and* $\boldsymbol{X} \sim N_k(0, \boldsymbol{R})$*, respectively. Then*

$$\lim_{\nu \to \infty} g_k(\boldsymbol{t}; \nu, \boldsymbol{R}) = f_k(\boldsymbol{t}; \boldsymbol{0}, \boldsymbol{R}) \quad \forall \boldsymbol{t} \in \mathbb{R}^k.$$

Corollary. *With* $\boldsymbol{T} \sim t_k(\nu, \boldsymbol{R})$ *and* $\boldsymbol{X} \sim N_k(\boldsymbol{0}, \boldsymbol{R})$*, define the equicoordinate quantiles* $c_{\nu,\alpha}$ *and* c_α *for* $\alpha \in (0, 1)$ *as* $P\left(\bigcap_{i=1}^k \{T_i \leq c_{\nu,\alpha}\}\right) = \alpha$ *and*

$P\left(\bigcap_{i=1}^{k}\{X_i \leq c_\alpha\}\right) = \alpha$. *Then*

$$\lim_{\nu\to\infty} c_{\nu,\alpha} = c_\alpha.$$

A practical consequence of Lemma 2.2.3 and its corollary for test decisions is hence that the use of quantiles of a multivariate normal distribution instead of those of a t-distribution may be an acceptable approximation in the case of large sample sizes. However, this leads to liberal decisions.

In addition to the random variable $\boldsymbol{T}$ defined in (2.1), another commonly used multivariate t variable is defined in the literature, e.g. by Tong [1990], p. 202. For $j = 1, \ldots, N$ let $\boldsymbol{X}_j = (X_{1j}, \ldots, X_{kj})'$ be independent $N_k(\mathbf{0}, \boldsymbol{\Sigma})$ variables. For $i = 1, ..., k$ and

$$\bar{X}_i = \frac{1}{N}\sum_{j=1}^{N} X_{ij}, \quad V_i^2 = \frac{1}{N-1}\sum_{j=1}^{N}(X_{ij} - \bar{X}_i)^2$$

let

$$T_i^* = \frac{\sqrt{N}\bar{X}_i}{V_i} \quad (i = 1, ..., k). \tag{2.4}$$

The random variable $\boldsymbol{T}^* = (T_1^*, ..., T_k^*)'$ is also called a multivariate t-variable. The marginal distribution of $\sqrt{N}\bar{X}_i/V_i$ is a t-distribution with $N-1$ degrees of freedom $(i = 1, ..., k)$. While (2.1) takes the same χ_ν^2 variable U for each component T_i, the χ_{N-1}^2 variables $(N-1)V_i^2$ in (2.4) are different because of their dependence on i. For statistical testing, this implies the assumption of homogeneous variances for the X_i in (2.1) and of heterogeneous variances for the $\bar{X}_i$ in (2.4). Indeed, definition (2.4) is most appropriate for many applications, e.g for the analysis of multiple endpoints that may have different scales and hence different variances. However, this

multivariate t variable is rarely described correctly in the literature. Furthermore, standard statistical software is not able to calculate the related density function. The following considerations and test procedures are hence all based on definition (2.1).

2.3 Skew-normal and Skew-t Distribution

Currently, there is an increasing interest in the literature on parametric families of multivariate normal distributions. The motivation is to introduce more flexible parametric families that still retain similarity with the multivariate normal distribution. The multivariate skew normal distribution can be viewed as a result of such ambitions. Amongst others, it has been studied by Azzalini [2005], Azzalini and Valle [1996] and Azzalini and Capitanio [2003]. An introduction into the topic is given by

Lemma 2.3.1. *If f_0 is a one-dimensional probability density function symmetric about zero, and G is a one-dimensional distribution function such that G' exists and is a density symmetric about zero, then*

$$f(z) = 2f_0(z)G\{w(z)\} \quad (z \in \mathbb{R}) \tag{2.5}$$

is a density function for any odd function $w(\cdot)$.

f_0 is the "basis" base density, $G\{w(x)\}$ the "perturbation" function. The set of "perturbed" densities always includes the "basis" density, since $w(x) \equiv 0$ yields $f_0 = f$. A simple method for random number generation is provided by

Corollary. *If $X \sim G'$ and $Y \sim f_0$ are independent random variables, then*

$$Z = \begin{cases} Y, & X < w(Y) \\ -Y, & \text{otherwise} \end{cases} \tag{2.6}$$

has density function (2.5). If $Y \sim f_0$ and $Z \sim f$, then $|Y| \stackrel{d}{=} |Z|$, where the notation $\stackrel{d}{=}$ denotes equality in distribution (the cumulative distribution functions are equal).

Among other properties, the result implies that all even moments of Y and Z are the same.

On using Equation (2.5) with $f_0 = \phi$ and $G = \Phi$, the density function and the distribution function of a $N(0,1)$ variate, respectively, and $w(x) = \alpha x$, where $\alpha \in \mathbb{R}$, we get the density

$$\phi(z, \alpha) = 2\phi(z)\Phi(\alpha z) \quad (z \in \mathbb{R}), \tag{2.7}$$

which is called *SN distribution* with shape parameter α, denoted by $SN(\alpha)$. If $Z \sim SN(\alpha)$ and $Y = \xi + \omega Z$, where $\xi \in \mathbb{R}^+$, then we shall write $Y \sim SN(\xi, \omega^2, \alpha)$. The following properties for Equation (2.7) hold:

(a) If $\alpha = 0$, we obtain the $N(0,1)$ density.

(b) If $Z \sim SN(\alpha)$, then $-Z \sim SN(-\alpha)$.

(c) As $\alpha \to \infty$, (2.7) converges pointwise to the half-normal density, namely $2\phi(z)$ for $z \geq 0$.

(d) If $Z \sim SN(\alpha)$, then $Z^2 \sim \chi^2_1$.

(e) For fixed α, (2.7) is strongly unimodal, i.e. $\log f(z, \alpha)$ is a concave function of z.

(f) The corresponding distribution function is given by

$$\Phi(z, \alpha) = \Phi(z) - 2T(z, \alpha),$$

where $T(z, \alpha)$ is the function studied by Owen [1956], and it satisfies the relationship

$$\Phi(z, -\alpha) = 1 - \Phi(-z, \alpha).$$

(g) If $U \sim N(0, 1)$ is independent of $Z \sim SN(\alpha)$, then

$$\frac{aU + bZ}{\sqrt{a^2 + b^2}} \sim SN\left(\frac{b\alpha}{\sqrt{a^2(1 + \alpha^2) + b^2}}\right) \tag{2.8}$$

for any $a, b \in \mathbb{R}$.

The moment generating function of $SN(\xi, \omega^2, \alpha)$ is given by

$$M(t) = E\left(e^{tY}\right) = 2\exp\left(\xi t + \frac{\omega^2 t^2}{2}\right)\Phi(\delta\omega t),$$

where $\delta = \alpha/\sqrt{1 + \alpha^2} \in (-1, 1)$. It follows that

$$E(Y) = \xi + \omega\mu_z, \quad Var(Y) = \omega^2(1 - \mu_z^2),$$

$$\gamma_1 = \frac{4 - \pi}{2}\frac{\mu_z^3}{(1 - \mu_z^2)^{3/2}}, \quad \gamma_2 = 2(\pi - 3)\frac{\mu_z^4}{(1 - \mu_z^2)^2},$$

where $\mu_z = \sqrt{2/\pi}\,\delta$ and γ_1, γ_2 denote the standardized third and fourth-order cumulants, respectively. The range of γ_1 is approximately (-0.9953, 0.9953).

The multivariate version is presented by

Lemma 2.3.2. *If f_0 is a d-dimensional probability density function such that $f_0(\boldsymbol{x}) = f_0(-\boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{R}^d$, G is a one-dimensional differentiable distribution function such that G' is a density symmetric about 0, and w is a real-valued function such that $w(-\boldsymbol{x}) = -w(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^d$, then*

$$f(\boldsymbol{z}) = 2f_0(\boldsymbol{z})G\{w(\boldsymbol{z})\} \quad (\boldsymbol{z} \in \mathbb{R}^d) \tag{2.9}$$

is a density function on $\mathbb{R}^d$.

Corollary. *If $X \sim G'$ and $\boldsymbol{Y} \sim f_0$ are independent variables, then $\boldsymbol{Z}$ defined as in Equation (2.6) has the distribution (2.9). If $\boldsymbol{Y} \sim f_0$ and $\boldsymbol{Z} \sim f$, then $t(\boldsymbol{Y}) \stackrel{d}{=} t(\boldsymbol{Z})$ for any real valued function such that $t(\boldsymbol{x}) = t(-\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^d$, irrespective of the choice of G and w.*

Consider the case that $f_0(\boldsymbol{x})$ in Equation (2.9) is $\phi_d(\boldsymbol{x}, \boldsymbol{\Omega})$, the density function of an $N_d(\boldsymbol{0}, \boldsymbol{\Omega})$ variable, where $\boldsymbol{\Omega}$ is a positive definite matrix. Also assume that $G = \Phi$ and w is a linear function. Allowing for the presence of a d-dimensional location parameter $\boldsymbol{\xi}$, the density function is

$$f(\boldsymbol{y}) = 2\phi_d(\boldsymbol{y} - \boldsymbol{\xi}, \boldsymbol{\Omega})\Phi(\boldsymbol{\alpha}'\boldsymbol{\omega}^{-1}(\boldsymbol{y} - \boldsymbol{\xi})) \quad (\boldsymbol{y} \in \mathbb{R}^d), \tag{2.10}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^d$ is the shape parameter and $\boldsymbol{\omega}$ is the diagonal matrix formed by the standard deviations of $\boldsymbol{\Omega}$. If a d-dimensional continuous random variable $\boldsymbol{Y}$ has the density (2.10), we say that its distribution is *multivariate SN* and write $\boldsymbol{Y} \sim SN_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$. The moment generating function of $SN_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ is given by

$$M(\boldsymbol{t}) = 2\exp\left(\boldsymbol{\xi}'\boldsymbol{t} + \frac{1}{2}\boldsymbol{t}'\boldsymbol{\Omega}\boldsymbol{t}\right)\Phi(\boldsymbol{\delta}'\boldsymbol{\omega}\boldsymbol{t}), \quad \boldsymbol{t} \in \mathbb{R}^d, \tag{2.11}$$

where $\boldsymbol{\delta} = (1 + \boldsymbol{\alpha}'\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha})^{-1/2}\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha}$ and $\bar{\boldsymbol{\Omega}} = \boldsymbol{\omega}^{-1}\boldsymbol{\Omega}\boldsymbol{\omega}^{-1}$ is the correlation matrix associ-

ated with $\boldsymbol{\Omega}$. From (2.11), it follows that:

$$E(\boldsymbol{Y}) = \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{\mu}_z, \quad Var(\boldsymbol{Y}) = \boldsymbol{\Omega} - \boldsymbol{\omega}\boldsymbol{\mu}_z\boldsymbol{\mu}_z'\boldsymbol{\omega},$$

$$\gamma_{1,d} = \left(\frac{4-\pi}{2}\right)^2 \left(\frac{\boldsymbol{\mu}_z'\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\mu}_z}{1-\boldsymbol{\mu}_z'\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\mu}_z}\right)^3, \quad \gamma_{2,d} = 2(\pi-3)\left(\frac{\boldsymbol{\mu}_z'\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\mu}_z}{1-\boldsymbol{\mu}_z'\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\mu}_z}\right)^2,$$

where $\boldsymbol{\mu}_z = \sqrt{2/\pi}\,\boldsymbol{\delta}$ is the mean value of the reduced variable $\boldsymbol{Z} = \boldsymbol{\omega}^{-1}(\boldsymbol{Y} - \boldsymbol{\xi}) \sim SN_d(\boldsymbol{0}, \bar{\boldsymbol{\Omega}}, \boldsymbol{\alpha})$. $\gamma_{1,d}$, $\gamma_{2,d}$ denote the multivariate indices of skewness and kurtosis whose approximate ranges are (0, 0.9905) and (0, 0.869), respectively. Another direct consequence of (2.11) is that the sum of a multivariate SN variate and an independent multivariate normal variate is still SN. This fact is essentially the multivariate version of property (2.8).

Like the considerations concerning the multivariate skew normal distribution there is an analogous variant of the t-distribution. See Azzalini and Capitanio [2003] and Azzalini [2005] therefore. A continuous random variable $\boldsymbol{Y}$ has a *multivariate skew-t distribution* if its density is of type

$$f_T(\boldsymbol{y}) = 2t_d(\boldsymbol{y}, \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu) T_1\left(\boldsymbol{\alpha}'\boldsymbol{\omega}^{-1}(\boldsymbol{y}-\boldsymbol{\xi})\left(\frac{\nu+d}{Q_{\boldsymbol{y}}+\nu}\right)^{1/2}; \nu+d\right) \quad (\boldsymbol{y} \in \mathbb{R}^d), \tag{2.12}$$

where $\boldsymbol{\xi}$, $\boldsymbol{\Omega}$ and $\boldsymbol{\omega}$ are as introduced above, $Q_y = (\boldsymbol{y}-\boldsymbol{\xi})'\boldsymbol{\Omega}^{-1}(\boldsymbol{y}-\boldsymbol{\xi})$,

$$t_d(\boldsymbol{y}, \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu) = \frac{\Gamma\left(\frac{1}{2}(\nu+d)\right)}{|\boldsymbol{\Omega}|^{1/2}(\pi\nu)^{d/2}\Gamma(\frac{1}{2}\nu)} \frac{1}{(1+Q_y/\nu)^{(\nu+d)/2}}$$

is the density function of a d-dimensional t variate with ν degrees of freedom, and $T_1(x; \nu+d)$ denotes the scalar t-distribution function with $\nu+d$ degrees of freedom. We write $\boldsymbol{Y} \sim ST(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \nu)$. Equation (2.12) can be generated by the same

construction used for the regular multivariate t-distribution, namely

$$\boldsymbol{Y} = \boldsymbol{\xi} + \frac{\boldsymbol{Z}}{\sqrt{W/\nu}}, \tag{2.13}$$

where $W \sim \chi^2_\nu$, if $\boldsymbol{Z}$ is an independent variable which is now taken to be $SN_d(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ in place of the $N_d(\mathbf{0}, \boldsymbol{\Omega})$ distribution used to produce the regular t-distribution. From (2.13) it follows that, if $\nu \to \infty$, Equation (2.12) converges to the SN density (2.10). The relation

$$(\boldsymbol{Y} - \boldsymbol{\xi})' \boldsymbol{\Omega}^{-1} (\boldsymbol{Y} - \boldsymbol{\xi})/d \sim F(d, \nu)$$

holds. Furthermore, unlimited range for the indices of skewness and kurtosis is allowed for the individual components.

2.4 Distribution of Maximum and Minimum of Test Statistics

A common starting point of many multiple test procedures is the use of a maximum or a minimum of test statistics. According to Theorem 2.1.4, the proper choice of a quantile c, coming from the distribution of $\max_{s \in I} Z_s$, guarantees the UIT to be a level-α test. Thus, the problem is to derive the corresponding distributions. Therefore, we have considered the skew normal and skew-t distribution. In this section, we show their connections with the maximum and minimum of test statistics. Primarily, explicit solutions have been known only for a few special cases. Tong [1990] (p. 126) has considered the probability density function for the maximum

of the components of an exchangeable multivariate normal random vector, i.e., its covariance matrix is equicorrelated. The case of two random variables with a joint bivariate normal distribution has been considered by Roberts [1966] and rediscovered by Loperfido [2002]. Their proofs are omitted here again for brevity.

Theorem 2.4.1. *Let the random variables X_1, X_2, Y_1, Y_2, Z_1, Z_2 be distributed as follows:*

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} ; \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right],$$

$$\frac{Y_1}{\sqrt{\sigma_{11}}} \sim SN\left(\frac{\sigma_{11}-\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}-\sigma_{12}^2}}\right), \quad \frac{Y_2}{\sqrt{\sigma_{22}}} \sim SN\left(\frac{\sigma_{22}-\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}-\sigma_{12}^2}}\right),$$

$$\frac{Z_1}{\sqrt{\sigma_{11}}} \sim SN\left(\frac{\sigma_{12}-\sigma_{11}}{\sqrt{\sigma_{11}\sigma_{22}-\sigma_{12}^2}}\right), \quad \frac{Z_2}{\sqrt{\sigma_{22}}} \sim SN\left(\frac{\sigma_{12}-\sigma_{22}}{\sqrt{\sigma_{11}\sigma_{22}-\sigma_{12}^2}}\right).$$

Then the distribution of $\max\{X_1, X_2\}$ *is a mixture with equal weights of the distributions of Y_1 and Y_2. The distribution of* $\min\{X_1, X_2\}$ *is a mixture with equal weights of the distributions of Z_1 and Z_2.*

Theorem 2.4.2. *Let X_1, X_2 be two standardized random variables whose distribution is jointly normal:*

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} ; \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right].$$

Then the distributions of the random variables $\max\{X_1, X_2\}$ *and* $\min\{X_1, X_2\}$ *are skew-normal:*

$$\max\{X_1, X_2\} \sim SN\left(\sqrt{\frac{1-\rho}{1+\rho}}\right), \quad \min\{X_1, X_2\} \sim SN\left(-\sqrt{\frac{1-\rho}{1+\rho}}\right). \tag{2.14}$$

Equations (2.13) and (2.14) lead to the next

Corollary. *Let X_1, X_2 be two random variables which are jointly central t-distributed with ν degrees of freedom and correlation ρ. Then the distributions of the random variables* $\max\{X_1, X_2\}$ *and* $\min\{X_1, X_2\}$ *are skew-t:*

$$\max\{X_1, X_2\} \sim ST\left(\nu, \sqrt{\frac{1-\rho}{1+\rho}}\right), \quad \min\{X_1, X_2\} \sim ST\left(\nu, -\sqrt{\frac{1-\rho}{1+\rho}}\right).$$

The generalization to k-variate random variables $\boldsymbol{X} = (X_1, ..., X_k)'$ is given by the following results of Arellano-Valle and Genton [2008]. We do not introduce the authors' denotation here for simplicity. Also the proofs can be seen in their article.

Theorem 2.4.3. *Let $\boldsymbol{X} = (X_1, ..., X_k)' \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The probability density function $f_{X_{(k)}}$ of $X_{(k)} = \max\{X_1, ..., X_k\}$ is*

$$f_{X_{(k)}} = \sum_{i=1}^{k} \phi_1(x; \mu_i, \Sigma_{ii}) \Phi_{k-1}\left(x\mathbf{1}_{k-1}; \boldsymbol{\mu}_{-i,i}(x), \Sigma_{-i-i,i}\right) \quad (x \in \mathbb{R}).$$

Theorem 2.4.4. *Let $\boldsymbol{X} = (X_1, ..., X_k)' \sim t_{k,\boldsymbol{\mu}}(\nu, \boldsymbol{\Sigma})$. The probability density function $f_{X_{(k)}}$ of $X_{(k)} = \max\{X_1, ..., X_k\}$ is*

$$f_{X_{(k)}} = \sum_{i=1}^{k} t_1(x; \mu_i, \Sigma_{ii}, \nu) T_{k-1}\left(x\mathbf{1}_{k-1}; \boldsymbol{\mu}_{-i,i}(x), \frac{\nu + z_i^2}{\nu + 1}\Sigma_{-i-i,i}, \nu + 1\right) \quad (x \in \mathbb{R}).$$

2.5 Quantile Relations

Having seen that maximum and minimum of components of multivariate normal and t-variables are SN and ST distributed, respectively, we now consider the related quantiles, since they are needed for test decisions and confidence intervals. The following considerations refer to the multivariate t- and the ST distribution, but are

also valid for the multivariate normal distribution and the SN distribution, which may be viewed as a special case with $\nu = \infty$.

Let $\boldsymbol{X} \sim t_k(\nu, \boldsymbol{R})$ be a k-variate random vector. For the joint k-variate t-distribution, we define

- the *lower* α*-quantile* $t^l_{k,\alpha}(\nu, \boldsymbol{R})$, where
  $P\left(\bigcap_{i=1}^k \{X_i \leq t^l_{k,\alpha}(\nu, \boldsymbol{R})\}\right) = \alpha,$

- the *lower* $(1-\alpha)$*-quantile* $t^l_{k,1-\alpha}(\nu, \boldsymbol{R})$, where
  $P\left(\bigcap_{i=1}^k \{X_i \leq t^l_{k,1-\alpha}(\nu, \boldsymbol{R})\}\right) = 1-\alpha,$

- the *upper* α*-quantile* $t^u_{k,\alpha}(\nu, \boldsymbol{R})$, where
  $P\left(\bigcap_{i=1}^k \{X_i \geq t^u_{k,\alpha}(\nu, \boldsymbol{R})\}\right) = \alpha,$

- the *upper* $(1-\alpha)$*-quantile* $t^u_{k,1-\alpha}(\nu, \boldsymbol{R})$, where
  $P\left(\bigcap_{i=1}^k \{X_i \geq t^u_{k,1-\alpha}(\nu, \boldsymbol{R})\}\right) = 1-\alpha,$ and

- the *two-sided* $(1-\alpha)$*-quantile* $t^{ts}_{k,1-\alpha}(\nu, \boldsymbol{R})$, where
  $P\left(\bigcap_{i=1}^k \{-t^{ts}_{k,1-\alpha}(\nu, \boldsymbol{R}) \leq X_i \leq t^{ts}_{k,1-\alpha}(\nu, \boldsymbol{R})\}\right) = 1-\alpha.$

Figure 2.2 illustrates these definitions for the bivariate case. A contour plot for a bivariate t-variable is shown with $\nu = 20$ and independent components, i.e., $\rho = 0$. It is immediately clear that there are only three relevant quantiles in the univariate

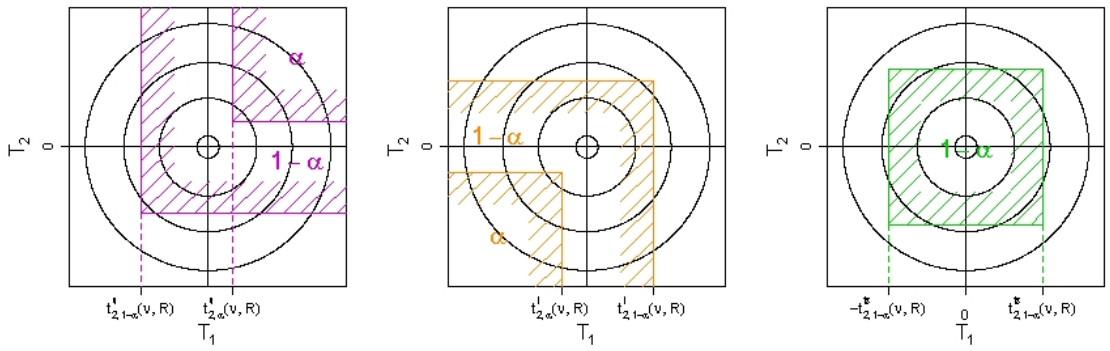


Figure 2.2: *Upper, lower and two-sided quantiles for a bivariate t-distributed random variable with* $\nu = 20$ *and* $\rho = 0$.

case because

$$t^l_{1,1-\alpha}(\nu, \boldsymbol{R}) = t^u_{1,\alpha}(\nu, \boldsymbol{R}) = t_{\nu,1-\alpha},$$

$$t^u_{1,1-\alpha}(\nu, \boldsymbol{R}) = t^l_{1,\alpha}(\nu, \boldsymbol{R}) = t_{\nu,\alpha},$$

$$t^{ts}_{1,1-\alpha}(\nu, \boldsymbol{R}) = t_{\nu,1-\alpha/2}.$$

Corresponding equalities do not hold for $k > 1$. The symmetry of the multivariate t-distribution about zero also results in $t^l_{k,\gamma}(\nu, \boldsymbol{R}) = -t^u_{k,\gamma}(\nu, \boldsymbol{R})$ for any $\gamma \in (0, 1)$. The following theorem gives a connection between these quantiles and quantiles of a ST distribution.

Theorem 2.5.1. *Let* $\boldsymbol{X} \sim t_k(\nu, \boldsymbol{R})$ *be a* k*-variate random vector and let* $st^{max}_\alpha(\nu, \boldsymbol{R})$, $st^{max}_{1-\alpha}(\nu, \boldsymbol{R})$, $st^{min}_\alpha(\nu, \boldsymbol{R})$, $st^{min}_{1-\alpha}(\nu, \boldsymbol{R})$ *be appropriate quantiles of the distribution of the maximum and minimum of the components of* $\boldsymbol{X}$*, respectively. The following properties are valid:*

$$t^l_{k,\alpha}(\nu, \boldsymbol{R}) = st^{max}_\alpha(\nu, \boldsymbol{R}), \quad t^l_{k,1-\alpha}(\nu, \boldsymbol{R}) = st^{max}_{1-\alpha}(\nu, \boldsymbol{R}),$$

$$t^u_{k,\alpha}(\nu, \boldsymbol{R}) = st^{min}_{1-\alpha}(\nu, \boldsymbol{R}), \quad t^u_{k,1-\alpha}(\nu, \boldsymbol{R}) = st^{min}_\alpha(\nu, \boldsymbol{R}), \tag{2.15}$$

and

$$st_{\alpha}^{max}(\nu, \boldsymbol{R}) = -st_{1-\alpha}^{min}(\nu, \boldsymbol{R}), \quad st_{1-\alpha}^{max}(\nu, \boldsymbol{R}) = -st_{\alpha}^{min}(\nu, \boldsymbol{R}). \tag{2.16}$$

Proof. From the above definition, it follows that

$$\begin{aligned} \alpha &= P\left(\bigcap_{i=1}^{k}\{X_i \leq t_{k,\alpha}^{l}(\nu, \boldsymbol{R})\}\right) = P\left(X_1 < t_{k,\alpha}^{l}(\nu, \boldsymbol{R}), ..., X_k < t_{k,\alpha}^{l}(\nu, \boldsymbol{R})\right) \\ &= P\left(\max_{i=1,...,k}\{X_i\} < t_{k,\alpha}^{l}(\nu, \boldsymbol{R})\right) = P\left(\max_{i=1,...,k}\{X_i\} < st_{\alpha}^{max}(\nu, \boldsymbol{R})\right). \end{aligned}$$

Hence $t_{k,\alpha}^{l}(\nu, \boldsymbol{R}) = st_{\alpha}^{max}(\nu, \boldsymbol{R})$. Furthermore,

$$\begin{aligned} \alpha &= P\left(\bigcap_{i=1}^{k}\{X_i \geq t_{k,\alpha}^{u}(\nu, \boldsymbol{R})\}\right) = P\left(X_1 > t_{k,\alpha}^{l}(\nu, \boldsymbol{R}), ..., X_k > t_{k,\alpha}^{l}(\nu, \boldsymbol{R})\right) \\ &= P\left(\min_{i=1,...,k}\{X_i\} > t_{k,\alpha}^{l}(\nu, \boldsymbol{R})\right) = P\left(\min_{i=1,...,k}\{X_i\} > st_{1-\alpha}^{min}(\nu, \boldsymbol{R})\right). \end{aligned}$$

Hence $t_{k,\alpha}^{u}(\nu, \boldsymbol{R}) = st_{1-\alpha}^{min}(\nu, \boldsymbol{R})$. The other relations given in (2.15) can be derived similarly. Equation (2.16) follows from the symmetry of the multivariate t-distribution about zero. □

Theorem 2.5.1 implies that we do not have to know the exact (skewed) distribution of the maximum and minimum of the test statistics. It is sufficient to know their joint multivariate distribution, because the quantiles coincide. Figure 2.3 illustrates this relation for the bivariate case. Let $\boldsymbol{X}$ be a bivariate-t random vector with 20 degrees of freedom and uncorrelated components. The first row of the figure shows a contour plot of the distribution of $\boldsymbol{X}$, the second row the related skewed distributions of $\min\{X_1, X_2\}$ and $\max\{X_1, X_2\}$. The connection between the quantiles of the skewed distributions and the related bivariate distributions can easily be seen. Hence, the quantiles $t_{k,1-\alpha}^{l}(\nu, \boldsymbol{R})$, $t_{k,1-\alpha}^{u}(\nu, \boldsymbol{R})$ and $t_{k,1-\alpha}^{ts}(\nu, \boldsymbol{R})$ are necessary for decisions in a UIT.

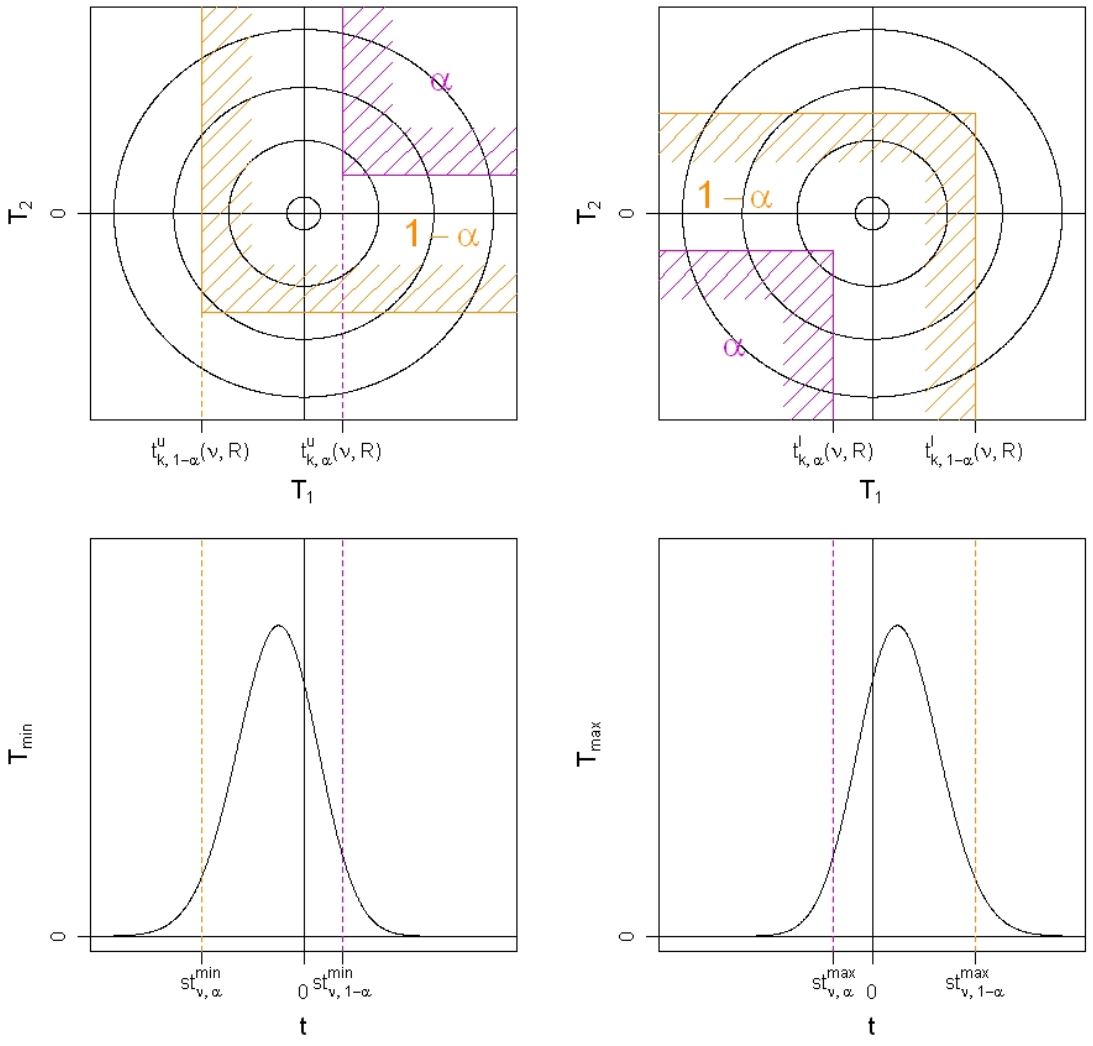


Figure 2.3: *Connection between the quantiles of the skewed distributions of* $\max\{X_1, X_2\}$ *or* $\min\{X_1, X_2\}$ *and the joint bivariate distributions of their components, respectively.*

The next statements provide information about the dependence of the quantiles on the dimension of their distribution (number of components) and correlations.

Theorem 2.5.2. *Let $\boldsymbol{X} \sim t_k(\nu, \boldsymbol{R})$ be a k-variate random vector. The following properties are valid if $\rho_{ij} = \rho$ for all $i \neq j$:*

$$\begin{aligned} t^l_{k,1-\alpha}(\nu, \boldsymbol{R}) &< t^l_{k+1,1-\alpha}(\nu, \boldsymbol{R}), \\ t^u_{k,1-\alpha}(\nu, \boldsymbol{R}) &> t^u_{k+1,1-\alpha}(\nu, \boldsymbol{R}), \\ t^{ts}_{k,1-\alpha}(\nu, \boldsymbol{R}) &< t^{ts}_{k+1,1-\alpha}(\nu, \boldsymbol{R}). \end{aligned}$$

Proof. The probability that the maximum of k random variables is larger than a fixed quantile c increases with increasing k,

$$P\left(\max_{i=1,\ldots,k}\{X_i\} > c\right) \leq P\left(\max_{i=1,\ldots,k+1}\{X_i\} > c\right).$$

Fixing both probabilities at level $1-\alpha$, the related quantiles $t^l_{k,1-\alpha}(\nu, \boldsymbol{R})$ and $t^l_{k+1,1-\alpha}(\nu, \boldsymbol{R})$ cannot be equal, and it follows that $t^l_{k,1-\alpha}(\nu, \boldsymbol{R}) < t^l_{k+1,1-\alpha}(\nu, \boldsymbol{R})$. Otherwise, the probability that the minimum of k random variables is smaller than a fixed quantile c also increases with increasing k,

$$P\left(\min_{i=1,\ldots,k}\{X_i\} < c\right) \leq P\left(\min_{i=1,\ldots,k+1}\{X_i\} < c\right).$$

We analogously obtain $t^u_{k,1-\alpha}(\nu, \boldsymbol{R}) > t^u_{k+1,1-\alpha}(\nu, \boldsymbol{R})$. And finally, the probability that both the maximum of k random variables is larger than any fixed quantile c, and the minimum of k random variables is smaller than any fixed quantile $-c$, also increases with increasing k,

$$P\left(\max_{i=1,\ldots,k}\{|X_i|\} > c\right) \leq P\left(\max_{i=1,\ldots,k+1}\{|X_i|\} > c\right).$$

It follows that $t^{ts}_{k,1-\alpha}(\nu, \boldsymbol{R}) < t^{ts}_{k+1,1-\alpha}(\nu, \boldsymbol{R})$. □

Theorem 2.5.3. *Let $\boldsymbol{X} \sim t_k(\nu, \boldsymbol{R})$ be a k-variate random vector with $\rho_{ij} = \rho$ for all $i \neq j$.*

1. *For $\rho = \rho^{min}$ (Equation (2.3)), $t^l_{k,1-\alpha}(\nu, \boldsymbol{R})$ has its maximal value, $t^u_{k,1-\alpha}(\nu, \boldsymbol{R})$ its minimal value, and*

$$t^l_{k,1-\alpha}(\nu, \boldsymbol{R}) \leq t_{\nu,1-\alpha/k}, \quad t^u_{k,1-\alpha}(\nu, \boldsymbol{R}) \geq t_{\nu,\alpha/k}.$$

2. *For $\rho = 0$, $t^{ts}_{k,1-\alpha}(\nu, \boldsymbol{R})$ has its maximal value.*

3. *For $\rho = 1$,*

$$t^l_{k,1-\alpha}(\nu, \boldsymbol{R}) = t_{\nu,1-\alpha}, \quad t^u_{k,1-\alpha}(\nu, \boldsymbol{R}) = t_{\nu,\alpha}.$$

Corollary. *Let $\boldsymbol{X} \sim t_k(\nu, \boldsymbol{R})$ be a k-variate random vector. The following conclusions hold for $k > 2$:*

$$\begin{aligned} t_{\nu,1-\alpha} &\leq t^l_{k,1-\alpha}(\nu, \boldsymbol{R}) < t_{\nu,1-\alpha/k}, \\ t_{\nu,\alpha} &\geq t^u_{k,1-\alpha}(\nu, \boldsymbol{R}) > t_{\nu,\alpha/k}, \\ t_{\nu,1-\alpha/2} &\leq t^{ts}_{k,1-\alpha}(\nu, \boldsymbol{R}) < t_{\nu,\alpha/2k}. \end{aligned}$$

For $k = 2$, $\leq$ holds instead of $<$ for $t^l_{k,\nu,1-\alpha}(\nu, \boldsymbol{R})$, and $\geq$ for $t^u_{k,\nu,1-\alpha}(\nu, \boldsymbol{R})$.

Theorem 2.5.3 is without proof but Figure 2.4 illustrates the dependence of k-variate t-quantiles on the correlation ($\rho_{ij} = \rho$ for all $i \neq j$) and their relation to univariate t-quantiles. The first row shows lower $(1-\alpha)$-quantiles, the second one two-sided $(1-\alpha)$-quantiles. The columns split up the dimensions $k = 2, 4, 8$. The Figures 2.5, 2.6 and 2.7 illustrate the dependence of the quantiles $t^l_{k,1-\alpha}(\nu, \boldsymbol{R})$ of a trivariate t-distribution (with $\nu = 20$) on the correlation of its components. The black

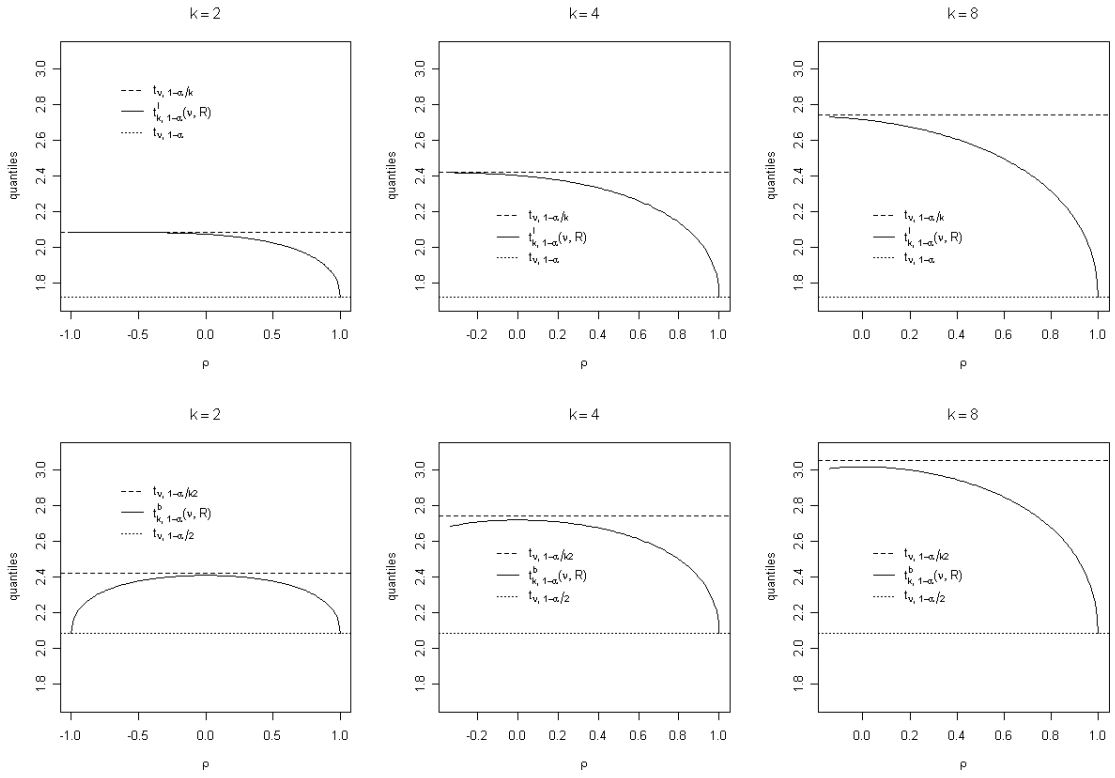


Figure 2.4: *Dependence of k-variate t-quantiles on the correlation and their relation to univariate t-quantiles;* $\nu = 20$, $\alpha = 0.05$.

dots in these plots represent a cutout of $\alpha = 0.05$ of the entire probability mass by $t^l_{k,1-\alpha}(\nu, \boldsymbol{R})$ for maximal negative (red), no (green) and maximal (blue) equicorrelation, respectively. According to Section 2.2, the red points in Figure 2.5 form a cutout of a disk, a cutout of a ball in Figure 2.6, and a ray in Figure 2.7. When there is maximal negative correlation, no probability mass is located in the first orthant. The related quantile does not cut very deeply into the probability space and is largest here. For increasing correlation, more and more probability mass moves into the first orthant, and the quantile becomes smaller. Having maximal positive correlation, 50% of the probability mass lies exactly on a ray from the origin to infinity. Here, the quantile is smallest.

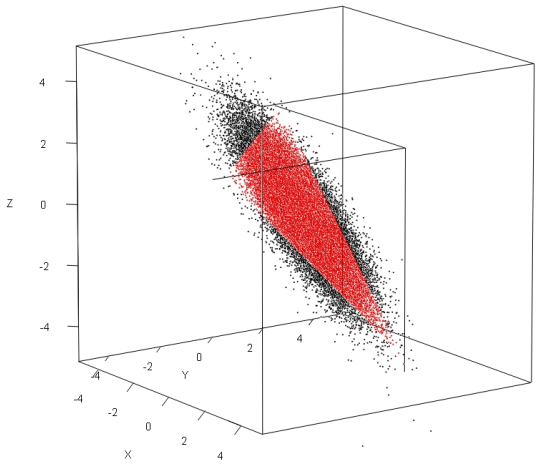


Figure 2.5: *Trivariate t-distributed random variables with* $\nu = 20$ *and maximal negative correlation* $\rho = -\frac{1}{2}$*, black points represent a cutout of 5%.*

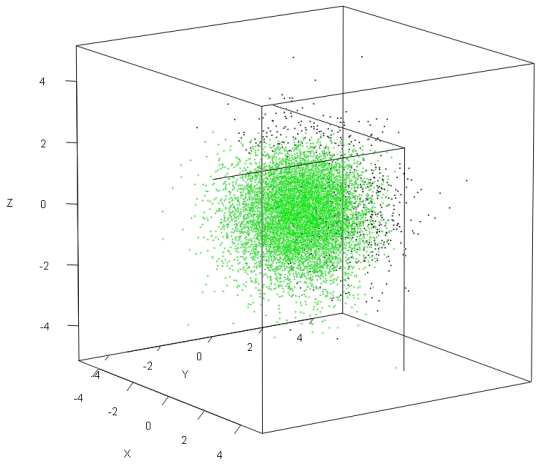


Figure 2.6: *Trivariate t-distributed random variables with* $\nu = 20$ *and correlation* $\rho = 0$*, black points represent a cutout of 5%.*

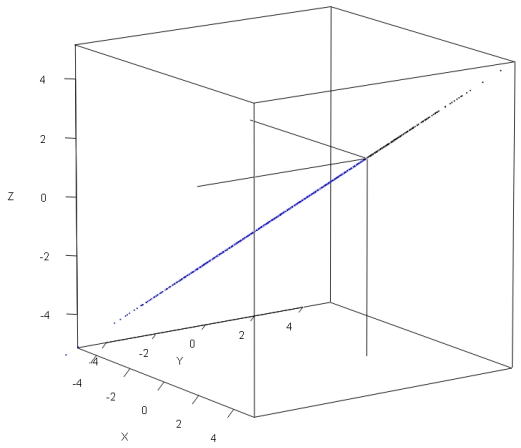


Figure 2.7: *Trivariate t-distributed random variables with* $\nu = 20$ *and maximal positive correlation* $\rho = 1$*, black points represent a cutout of 5%.*

A consequence of these considerations is that the use of a Bonferroni adjustment for decisions in a one-sided UIT is almost correct for maximal negative equicorrelated local comparisons and most conservative for maximal positive ones. Bonferroni-adjusted two-sided UIT are least conservative for uncorrelated local comparisons and most conservative for high absolute values of the correlations.

Chapter 3

Multiple Contrast Tests in the Presence of Heteroscedasticity

3.1 Introduction

MCTs and related SCIs are well-known methods for testing and estimating linear functions of means – i.e. contrasts – of normally distributed populations. A broad class of testing problems can be handled by them in modeling suitable contrast coefficients. The many-to-one comparison of Dunnett [1955] is one of the most frequently applied and cited testing procedures today. Several treatments are compared with one control and tested for deviation. This can be translated by related contrast coefficients and represents a very simple example. The all-pair comparison of Tukey [1953], comparing all treatments against each other, is also a very famous example.

Bretz [2006] has formulated the trend test of Williams [1971] as an approximate MCT. Here, the contrast coefficients depend in addition on the sample sizes of the treatment groups. Moreover, other interesting, problem-specific contrasts can be created (see Westfall [1997]). Furthermore, MCTs and SCIs are also available for ratios of means; see Dilba et al. [2004]. They are used if conclusions about ratios rather than differences of means are of interest. That is, if relative changes, e.g. in per cent, are to be analyzed.

The resulting MCTs and SCIs assume homogeneous variances for the data. This condition is necessary for the derivation of a joint multivariate t-distribution of the test statistics. If not fulfilled, this distribution is not available. Nevertheless, an adjustment for heteroscedastic data is necessary. Dose finding studies often have the problem of heteroscedasticity because the variance of the data depends on the dose effect. The data of Westfall [1997] (in Section 3.6.1), Adler and Kliesch [1990] (in Section 3.6.2) or Silva-Costa-Gomes et al. [2005] (in Hasler et al. [2008]) are examples. The first adjustment regarding heteroscedasticity for tests of any contrasts of means from normally distributed data has been made by Satterthwaite [1946]. Like Welch [1938], he approximates the degrees of freedom of the resulting t-distribution by matching first and second moments. Games and Howell [1976] used this approach for all-pair comparisons. Many other procedures have been suggested and investigated. Most of them have been developed for special contrasts only, or tend to achieve conservative or liberal tests depending on the extent of heteroscedasticity (see, e.g., Dunnett [1980]). Other approaches – from Welch [1951] or Brown and Forsythe [1974] – are based on F-distributions. This work presents three versions

of a general approach that handles the entire family of MCTs for differences and ratios of means using multivariate t-distributions, and against the background of controlling the FWE at level α.

In Section 3.2, the testing problem is formulated, and an adequate test statistic and distribution parameters are derived using several methods. Section 3.3 shows results of α-simulations for several contrasts and key settings. SCIs are treated in Section 3.4, power considerations in 3.5. Examples are given in Section 3.6.

3.2 Test Procedure

3.2.1 Differences of Means

For $h = 1, \ldots, p$ and $j = 1, \ldots, n_h$, let X_{hj} denote the jth observation under the hth treatment in a one-way layout. Suppose the X_{hj} to be independently normal with means μ_h and variances σ_h^2, thus

$$X_{hj} \sim \perp\mathrm{N}(\mu_h, \sigma_h^2) \quad (h = 1, \ldots, p,\ j = 1, \ldots, n_h).$$

Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ be the vector of treatment means and $\bar{\boldsymbol{X}} = (\bar{X}_1, \ldots, \bar{X}_k)'$ its estimator with

$$\bar{X}_h = \frac{1}{n_h} \sum_{j=1}^{n_h} X_{hj} \quad (h = 1, \ldots, p).$$

The sample variances are given by

$$S_h^2 = \frac{1}{n_h - 1} \sum_{j=1}^{n_h} (X_{hj} - \bar{X}_h)^2 \quad (h = 1, \ldots, p).$$

We are interested in the vector of contrasts $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_q)'$, where

$$\begin{aligned} \eta_l &= \sum_{h=1}^{p} c_{lh}\mu_h \\ &= \boldsymbol{c}_l'\boldsymbol{\mu} \quad (l = 1, \ldots, q). \end{aligned}$$

The vectors $\boldsymbol{c}_l = (c_{l1}, \ldots, c_{lp})'$ consist of real constants with $\sum_{h=1}^{p} c_{lh} = 0$ $(l = 1, \ldots, q)$. Without loss of generality, the objective is to test the hypotheses

$$H_{0l} : \eta_l \leq \delta_l \quad (l = 1, \ldots, q) \tag{3.1}$$

for specified absolute thresholds δ_l. Usually, $\delta_l = 0$ for all $l = 1, \ldots, q$. This testing problem is a UIT because the overall null hypothesis of interest can be expressed as an intersection of these local null hypotheses, that is,

$$H_0 = \bigcap_{l=1}^{q} H_{0l}.$$

Figure 3.1 shows the parameter space of testing problem (3.1) for the case where $q = 2$ contrasts are of interest. For the further development, we reshape (3.1) and set

$$\begin{aligned} \psi_l &= \left(\sum_{h=1}^{p} c_{lh}\mu_h\right) - \delta_l \\ &= \boldsymbol{c}_l'\boldsymbol{\mu} - \delta_l \quad (l = 1, \ldots, q) \end{aligned}$$

with estimator

$$\begin{aligned} \hat{\psi}_l &= \left(\sum_{h=1}^{p} c_{lh}\bar{X}_h\right) - \delta_l \\ &= \boldsymbol{c}_l'\bar{\boldsymbol{X}} - \delta_l \quad (l = 1, \ldots, q). \end{aligned}$$

We have

$$\begin{aligned} Var(\hat{\psi}_l) &= \sum_{h=1}^{p} c_{lh}^2\sigma_h^2/n_h \\ &= \boldsymbol{c}_l'\boldsymbol{V}\boldsymbol{M}\boldsymbol{c}_l \quad (l = 1, \ldots, q), \end{aligned}$$

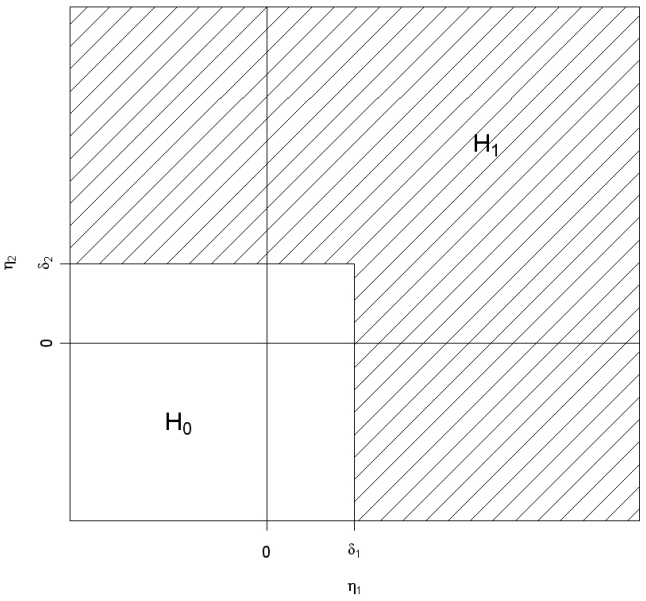


Figure 3.1: *Parameter space of the one-sided MCT for differences with $q = 2$ contrasts.*

with $\boldsymbol{M}$ and $\boldsymbol{V}$ defined by

$$\boldsymbol{M} = \begin{pmatrix} 1/n_1 & & 0 \\ & \ddots & \\ 0 & & 1/n_p \end{pmatrix} \quad \text{and} \quad \boldsymbol{V} = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_p^2 \end{pmatrix},$$

respectively. Consequently, we obtain the vector of test statistics $\boldsymbol{T} = (T_1, \ldots, T_q)'$, where

$$\begin{aligned} T_l &= \frac{\left(\sum_{h=1}^{p} c_{lh}\bar{X}_h\right) - \delta_l}{\sqrt{\sum_{h=1}^{p} c_{lh}^2 S_h^2/n_h}} \\ &= \frac{\boldsymbol{c}_l'\bar{\boldsymbol{X}} - \delta_l}{\sqrt{\boldsymbol{c}_l'\hat{\boldsymbol{V}}\boldsymbol{M}\boldsymbol{c}_l}} \quad (l = 1, \ldots, q) \end{aligned} \tag{3.2}$$

and

$$\hat{\boldsymbol{V}} = \begin{pmatrix} S_1^2 & & 0 \\ & \ddots & \\ 0 & & S_p^2 \end{pmatrix}.$$

If the group variances are homogeneous, $\sigma_1^2, \ldots, \sigma_k^2 = \sigma^2$, $\boldsymbol{V}$ and $\hat{\boldsymbol{V}}$ can be reduced to the scalars σ^2 and S^2, respectively, with

$$S^2 = \frac{\sum_{h=1}^{p}(n_h - 1)S_h^2}{\sum_{h=1}^{p}(n_h - 1)}. \tag{3.3}$$

In the presence of heterogeneous variances, we cannot derive an exact joint multivariate t-distribution for $(T_1, \ldots, T_q)'$. Numerator and denominator in (3.2) are indeed stochastically independent, but the denominator is a mixture of $\sigma_h^2\chi^2$-distributions (h=1,...,p). The marginal distribution of each T_l can be approximated by a t-distribution using the idea of Welch [1938] and Satterthwaite [1946]. Denote the squared denominator of (3.2) as ζ_l. Of course, ζ_l depends on the estimates S_h^2 $(h = 1, \ldots, p)$. The aim is to approximate the distribution of ζ_l by that of a $\chi^2(\nu_l)$ variable, multiplied by $\frac{\sigma_l^2}{\nu_l}$, where σ_l^2 and ν_l are chosen such that the first two moments of ζ_l agree with the first two moments of $\frac{\sigma_l^2}{\nu_l}\chi^2(\nu_l)$. We have

$$\zeta_l = \sum_{h=1}^{p} \frac{c_{lh}^2 S_h^2}{n_h},$$

$$E\zeta_l = \sum_{h=1}^{p} \frac{c_{lh}^2 \sigma_h^2}{n_h}, \tag{3.4}$$

$$Var\zeta_l = \sum_{h=1}^{p} \frac{2c_{lh}^4 \sigma_h^4}{n_h^2(n_h - 1)}. \tag{3.5}$$

While (3.4) is obvious, (3.5) results from the following

Proof.

$$\begin{aligned}
Var\zeta_l &= \sum_{h=1}^{p} \frac{c_{lh}^4}{n_h^2} Var\left(S_h^2\right) \\
&= \sum_{h=1}^{p} \frac{c_{lh}^4}{n_h^2(n_h-1)^2} Var\left(\sum_{j=1}^{n_h}(X_{hj}-\bar{X}_h)^2\right) \\
&= \sum_{h=1}^{p} \frac{c_{lh}^4}{n_h^2(n_h-1)^2} Var\left(\sum_{j=1}^{n_h}\sigma_h^2\left(\frac{X_{hj}-\bar{X}_h}{\sigma_h}\right)^2\right) \\
&= \sum_{h=1}^{p} \frac{c_{lh}^4\sigma_h^4}{n_h^2(n_h-1)^2} Var\left(\sum_{j=1}^{n_h}\left(\frac{X_{hj}-\bar{X}_h}{\sigma_h}\right)^2\right) \\
&= \sum_{h=1}^{p} \frac{c_{lh}^4\sigma_h^4}{n_h^2(n_h-1)^2} 2(n_h-1) \\
&= \sum_{h=1}^{p} \frac{2c_{lh}^4\sigma_h^4}{n_h^2(n_h-1)}
\end{aligned}$$

□

Furthermore,

$$E\left(\frac{\sigma_l^2}{\nu_l}\chi^2(\nu_l)\right) = \frac{\sigma_l^2}{\nu_l}\nu_l = \sigma_l^2,$$
$$Var\left(\frac{\sigma_l^2}{\nu_l}\chi^2(\nu_l)\right) = \frac{\sigma_l^4}{\nu_l^2}2\nu_l = 2\frac{\sigma_l^4}{\nu_l}.$$

The system of equations

$$\sum_{h=1}^{p} \frac{c_{lh}^2\sigma_h^2}{n_h} = \sigma_l^2,$$
$$\sum_{h=1}^{p} \frac{2c_{lh}^4\sigma_h^4}{n_h^2(n_h-1)} = 2\frac{\sigma_l^4}{\nu_l}$$

yields

$$\sum_{h=1}^{p} \frac{2c_{lh}^4\sigma_h^4}{n_h^2(n_h-1)} = 2\left(\sum_{h=1}^{p} \frac{c_{lh}^2\sigma_h^2}{n_h}\right)^2 \frac{1}{\nu_l}$$

and

$$\nu_l = \frac{\left(\sum_{h=1}^{p} \frac{c_{lh}^2\sigma_h^2}{n_h}\right)^2}{\sum_{h=1}^{p} \frac{c_{lh}^4\sigma_h^4}{n_h^2(n_h-1)}}.$$

Replacement of the unknown parameters σ_h^2 with the estimators S_h^2 $(h = 1, \ldots, p)$ leads to[1]

$$\hat{\nu}_l = \frac{\left(\sum_{h=1}^{p} \frac{c_{lh}^2 S_h^2}{n_h}\right)^2}{\sum_{h=1}^{p} \frac{c_{lh}^4 S_h^4}{n_h^2(n_h-1)}}. \tag{3.6}$$

Hence, under H_{0l}, T_l approximately follows a t-distribution with $\hat{\nu}_l$ degrees of freedom. For an approximate joint distribution, we have to determine a suitable correlation matrix. We have

$$\begin{aligned}
& Cov\left(\sum_{h=1}^{p} c_{lh}\bar{X}_h, \sum_{h=1}^{p} c_{l'h}\bar{X}_h\right) \\
= \; & E\left(\sum_{h=1}^{p} c_{lh}\bar{X}_h - E\sum_{h=1}^{p} c_{lh}\bar{X}_h\right)\left(\sum_{h=1}^{p} c_{l'h}\bar{X}_h - E\sum_{h=1}^{p} c_{l'h}\bar{X}_h\right) \\
= \; & E\left(\sum_{h=1}^{p} c_{lh}\left(\bar{X}_h - E\bar{X}_h\right)\right)\left(\sum_{h=1}^{p} c_{l'h}\left(\bar{X}_h - E\bar{X}_h\right)\right) \\
= \; & E\sum_{h=1}^{p}\sum_{h'=1}^{p} c_{lh}c_{l'h'}\left(\bar{X}_h - E\bar{X}_h\right)\left(\bar{X}_{h'} - E\bar{X}_{h'}\right) \\
= \; & \sum_{h=1}^{p}\sum_{h'=1,h'=h}^{p} c_{lh}c_{l'h'}E\left(\bar{X}_h - E\bar{X}_h\right)\left(\bar{X}_{h'} - E\bar{X}_{h'}\right) \\
& + \sum_{h=1}^{p}\sum_{h'=1,h'\neq h}^{p} c_{lh}c_{l'h'}E\left(\bar{X}_h - E\bar{X}_h\right)\left(\bar{X}_{h'} - E\bar{X}_{h'}\right) \\
= \; & \sum_{h=1}^{p} c_{lh}c_{l'h}E\left(\bar{X}_h - E\bar{X}_h\right)^2 \\
& + \sum_{h=1}^{p} c_{lh}E\left(\bar{X}_h - E\bar{X}_h\right)\sum_{h'=1,h'\neq h}^{p} c_{l'h'}E\left(\bar{X}_{h'} - E\bar{X}_{h'}\right) \\
= \; & \sum_{h=1}^{p} c_{lh}c_{l'h}E\left(\bar{X}_h - E\bar{X}_h\right)^2 \\
= \; & \sum_{h=1}^{p} c_{lh}c_{l'h}Var\bar{X}_h \\
= \; & \sum_{h=1}^{p} c_{lh}c_{l'h}\frac{\sigma_h^2}{n_h}
\end{aligned}$$

[1]The degrees of freedom in (3.6) and later in (3.12) must be greater than or equal to 2 for a well defined distribution.

Thus, it follows that

$$\begin{aligned}\rho_{ll'} &= \frac{\sum_{h=1}^{p} c_{lh}c_{l'h}\sigma_h^2/n_h}{\sqrt{\left(\sum_{h=1}^{p} c_{lh}^2\sigma_h^2/n_h\right)\left(\sum_{h=1}^{p} c_{l'h}^2\sigma_h^2/n_h\right)}} \\ &= \frac{\boldsymbol{c}_l'\boldsymbol{V}\boldsymbol{M}\boldsymbol{c}_{l'}}{\sqrt{\boldsymbol{c}_l'\boldsymbol{V}\boldsymbol{M}\boldsymbol{c}_l}\sqrt{\boldsymbol{c}_{l'}'\boldsymbol{V}\boldsymbol{M}\boldsymbol{c}_{l'}}} \quad (1 \leq l, l' \leq q). \end{aligned} \tag{3.7}$$

Hence, $T_1, \ldots, T_q$ follow an unknown joint q-variate distribution with correlation matrix $\boldsymbol{R} = (\rho_{ll'})_{l,l'}$. It is clearly not related to a q-variate t-distribution in the sense of existing definitions because it additionally depends on the unknown variances σ_h^2 and on several degrees of freedom ν_l. This problem has not been solved so far, and hence some approximate approach has to be followed.

Games and Howell [1976] were the first to use a test statistic without a pooled variance estimator, but with individual variance estimators S_h^2 for the all-pair comparison procedure. Formula (3.2) can be seen as a generalization of their test statistic for all MCTs. The authors have applied studentized range distributions with comparison-specific degrees of freedom according to (3.6). The problem of calculating the correlations from (3.7) is clearly that $\boldsymbol{V}$ is unknown. Games and Howell have effectively replaced $\boldsymbol{V}$ with the unit matrix, which leads to the same correlations as if homogeneous group variances were assumed. This procedure is referred to as the GH procedure in the following. Some articles were already concerned with the procedure of Games and Howell (e.g., Tamhane [1979] and Dunnett [1980]). Expectedly, their method can lead to both conservative or liberal test decisions depending on the amount of heteroscedasticity and the sample allocation. Another approach is the use of the matrix $\hat{\boldsymbol{V}}$ instead of $\boldsymbol{V}$ in (3.7). This means to plug-in the variance estimators S_h^2 and it yields the estimated correlation matrix $\hat{\boldsymbol{R}} = (\hat{\rho}_{ll'})_{l,l'}$ with

elements

$$\begin{aligned}\hat{\rho}_{ll'} &= \frac{\sum_{h=1}^{p} c_{lh}c_{l'h}S_h^2/n_h}{\sqrt{\left(\sum_{h=1}^{p} c_{lh}^2 S_h^2/n_h\right)\left(\sum_{h=1}^{p} c_{l'h}^2 S_h^2/n_h\right)}} \\ &= \frac{\boldsymbol{c}_l'\hat{\boldsymbol{V}}\boldsymbol{M}\boldsymbol{c}_{l'}}{\sqrt{\boldsymbol{c}_l'\hat{\boldsymbol{V}}\boldsymbol{M}\boldsymbol{c}_l}\sqrt{\boldsymbol{c}_{l'}'\hat{\boldsymbol{V}}\boldsymbol{M}\boldsymbol{c}_{l'}}} \quad (1 \leq l, l' \leq q).\end{aligned}$$

We refer to this plug-in procedure as PI. Because of an unknown joint distribution for $T_1, \ldots, T_q$, both GH and PI use q several approximate q-variate t-distributions to come to a test decision about testing problem (3.1). Hence, each single test statistic T_l is related to "its own" distinct q-variate t-distribution with correlation matrix $\hat{\boldsymbol{R}}$ and degree of freedom $\hat{\nu}_l$ coming from (3.6). That results in different, non-equidistant quantiles for the test decisions.

Dunnett [1985] has considered the use of both the geometric and arithmetic mean of the correlations, respectively, for comparisons with a control in the context of unbalanced one-way layouts in the homoscedastic case. Hochberg and Tamhane [1987] have suggested the arithmetic mean, as well as Dunnett [1985], because it provides a less conservative approximation. That is,

$$\bar{\rho}_{ll'} = \frac{1}{q^2 - q}\sum_{l \neq l'=1}^{q} \hat{\rho}_{ll'} \quad \text{for all} \quad 1 \leq l \neq l' \leq q.$$

Tamhane and Logan [2004] resort to this approach even in the case of heteroscedastic data. Furthermore, they recommend to use the average of the degrees of freedom,

$$\bar{\nu} = \frac{1}{q}\sum_{l=1}^{q} \hat{\nu}_l.$$

The resulting procedure for all MCTs is referred to as HTL in the following. Note that HTL uses a single, approximate joint q-variate t-distribution of $T_1, \ldots, T_q$ which is explicitly avoided by the GH and PI procedures.

Besides, let us denote the procedure for homogeneous variances by HOM. Here, the same correlation matrix as for GH is used, but only a single degree of freedom

$$\nu = \sum_{h=1}^{p}(n_h - 1),$$

and the pooled variance estimator (3.3) for the test statistic (3.2) is used instead of the individual variance estimators S_h^2.

The decision rule for testing problem (3.1) is to reject H_{0l} for each contrast η_l with

$$T_l > t^l_{q,1-\alpha}(.), \tag{3.8}$$

where $t^l_{q,1-\alpha}(.)$ is a lower $(1-\alpha)$-quantile of (one or q) related q-variate t-distributions. If two-sided testing is of interest, the absolute values for T_l, and quantiles $t^{ts}_{q,1-\alpha}(.)$ have to be taken. For the computation of these quantiles, one may resort to the numerical integration routines of Genz and Bretz [1999, 2002], see also Bretz et al. [2001]. Their algorithm is not restricted to special correlation structures. Related adjusted p-values per comparison can also be obtained, of course.

Let us finally point out that HOM and HTL are simultaneous test procedures in the sense of Gabriel [1969], while GH and PI are not, because the T_l $(l = 1, \ldots, q)$ are compared with different quantiles. Since all of these procedures are UIT, Theorem 2.1.3 holds and they are thus coherent and consonant. That means, if any local hypothesis H_{0l} $(l = 1, \ldots, q)$ is rejected, H_0 is, too. On the other hand, if H_0 is rejected, at least one H_{0l} must be rejected.

3.2.2 Ratios of Means

The same assumptions are made as in Section 3.2.1. Furthermore, let the means $\mu_1, \ldots, \mu_k$ have the same algebraic sign. We are interested now in the vector of ratios of contrasts $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)'$, where

$$\begin{aligned} \gamma_l &= \frac{\sum_{h=1}^{p} c_{lh}\mu_h}{\sum_{h=1}^{p} d_{lh}\mu_h} \\ &= \frac{\boldsymbol{c}_l'\boldsymbol{\mu}}{\boldsymbol{d}_l'\boldsymbol{\mu}} \quad (l = 1, \ldots, q). \end{aligned}$$

The vectors $\boldsymbol{c}_l = (c_{l1}, \ldots, c_{lp})'$ and $\boldsymbol{d}_l = (d_{l1}, \ldots, d_{lp})'$ consist of real constants. Without loss of generality, the hypotheses to be tested are

$$H_{0l} : \gamma_l \leq \theta_l \quad (l = 1, \ldots, q) \tag{3.9}$$

for specified relative thresholds θ_l.Usually, $\theta_l = 1$ for all $l = 1, \ldots, q$. Like testing problem (3.1), this one is a UIT, and

$$H_0 = \bigcap_{l=1}^{q} H_{0l}.$$

Figure 3.2 shows the parameter space of testing problem (3.9) when $q = 2$ contrasts are of interest. We reshape (3.9) and set

$$\begin{aligned} \psi_l &= \sum_{h=1}^{p} c_{lh}\mu_h - \theta_l \sum_{h=1}^{p} d_{lh}\mu_h = \sum_{h=1}^{p} (c_{lh} - \theta_l d_{lh})\,\mu_h \\ &= (\boldsymbol{c}_l - \theta_l \boldsymbol{d}_l)'\,\boldsymbol{\mu} \quad (l = 1, \ldots, q) \end{aligned}$$

with estimator

$$\begin{aligned} \hat{\psi}_l &= \sum_{h=1}^{p} (c_{lh} - \theta_l d_{lh})\,\bar{X}_h \\ &= (\boldsymbol{c}_l - \theta_l \boldsymbol{d}_l)'\,\bar{\boldsymbol{X}} \quad (l = 1, \ldots, q). \end{aligned}$$

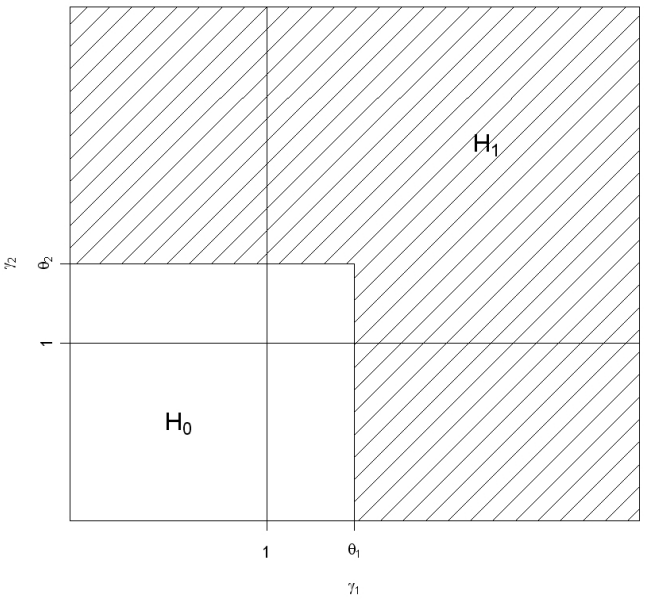


Figure 3.2: *Parameter space of the one-sided MCT for ratios with $q = 2$ contrasts.*

Since

$$\begin{aligned} Var(\hat{\psi}_l) &= \sum_{h=1}^{p} (c_{lh} - \theta_l d_{lh})^2 \, \sigma_h^2 / n_h \\ &= (\boldsymbol{c}_l - \theta_l \boldsymbol{d}_l)' \, \boldsymbol{V}\boldsymbol{M} \, (\boldsymbol{c}_l - \theta_l \boldsymbol{d}_l) \quad (l = 1, \ldots, q), \end{aligned}$$

we obtain the vector of test statistics $\boldsymbol{T} = (T_1, \ldots, T_q)'$, where

$$\begin{aligned} T_l &= \frac{\sum_{h=1}^{p} (c_{lh} - \theta_l d_{lh}) \, \bar{X}_h}{\sqrt{\sum_{h=1}^{p} (c_{lh} - \theta_l d_{lh})^2 \, S_h^2 / n_h}} \\ &= \frac{(\boldsymbol{c}_l - \theta_l \boldsymbol{d}_l)' \, \bar{\boldsymbol{X}}}{\sqrt{(\boldsymbol{c}_l - \theta_l \boldsymbol{d}_l)' \, \hat{\boldsymbol{V}}\boldsymbol{M} \, (\boldsymbol{c}_l - \theta_l \boldsymbol{d}_l)}} \quad (l = 1, \ldots, q). \end{aligned} \tag{3.10}$$

Now, the same considerations as in the case of differences lead to

$$\zeta_l = \sum_{h=1}^{p} \frac{(c_{lh} - \theta_l d_{lh})^2 \, S_h^2}{n_h},$$

$$E\zeta_l = \sum_{h=1}^{p} \frac{(c_{lh} - \theta_l d_{lh})^2 \, \sigma_h^2}{n_h}$$

and

$$Var\zeta_l = \sum_{h=1}^{p} \frac{2\left(c_{lh} - \theta_l d_{lh}\right)^4 \sigma_h^4}{n_h^2(n_h - 1)}.$$

The system of equations

$$\sum_{h=1}^{p} \frac{\left(c_{lh} - \theta_l d_{lh}\right)^2 \sigma_h^2}{n_h} = \sigma^2,$$

$$\sum_{h=1}^{p} \frac{2\left(c_{lh} - \theta_l d_{lh}\right)^4 \sigma_h^4}{n_h^2(n_h - 1)} = 2\frac{\sigma^4}{\nu_l}$$

yields

$$\nu_l = \frac{\left(\sum_{h=1}^{p} \frac{(c_{lh} - \theta_l d_{lh})^2 \sigma_h^2}{n_h}\right)^2}{\sum_{h=1}^{p} \frac{(c_{lh} - \theta_l d_{lh})^4 \sigma_h^4}{n_h^2(n_h - 1)}}. \tag{3.11}$$

Replacement of the unknown parameters σ_h^2 with the estimators S_h^2 $(h = 1, \ldots, p)$ leads to

$$\hat{\nu}_l = \frac{\left(\sum_{h=1}^{p} \frac{(c_{lh} - \theta_l d_{lh})^2 S_h^2}{n_h}\right)^2}{\sum_{h=1}^{p} \frac{(c_{lh} - \theta_l d_{lh})^4 S_h^4}{n_h^2(n_h - 1)}}. \tag{3.12}$$

Hence, under H_{0l}, T_l approximately follows a t-distribution with $\hat{\nu}_l$ degrees of freedom. Correspondingly, the required covariances are

$$\begin{aligned} &Cov\left(\sum_{h=1}^{p} \left(c_{lh} - \theta_l d_{lh}\right) \bar{X}_h, \sum_{h=1}^{p} \left(c_{l'h} - \theta_{l'} d_{l'h}\right) \bar{X}_h\right) \\ &= \sum_{h=1}^{p} \left(c_{lh} - \theta_l d_{lh}\right)\left(c_{l'h} - \theta_{l'} d_{l'h}\right) \frac{\sigma_h^2}{n_h}, \end{aligned}$$

which leads to

$$\begin{aligned} \rho_{ll'} &= \frac{\sum_{h=1}^{p} \left(c_{lh} - \theta_l d_{lh}\right)\left(c_{l'h} - \theta_{l'} d_{l'h}\right) \sigma_h^2/n_h}{\sqrt{\left(\sum_{h=1}^{p} \left(c_{lh} - \theta_l d_{lh}\right)^2 \sigma_h^2/n_h\right)\left(\sum_{h=1}^{p} \left(c_{l'h} - \theta_{l'} d_{l'h}\right)^2 \sigma_h^2/n_h\right)}} \\ &= \frac{\left(\boldsymbol{c}_l - \theta_l \boldsymbol{d}_l\right)' \boldsymbol{V}\boldsymbol{M}\left(\boldsymbol{c}_{l'} - \theta_{l'} \boldsymbol{d}_{l'}\right)}{\sqrt{\left(\boldsymbol{c}_l - \theta_l \boldsymbol{d}_l\right)' \boldsymbol{V}\boldsymbol{M}\left(\boldsymbol{c}_l - \theta_l \boldsymbol{d}_l\right)}\sqrt{\left(\boldsymbol{c}_{l'} - \theta_{l'} \boldsymbol{d}_{l'}\right)' \boldsymbol{V}\boldsymbol{M}\left(\boldsymbol{c}_{l'} - \theta_{l'} \boldsymbol{d}_{l'}\right)}} \\ &\quad (1 \leq l, l' \leq q) \end{aligned} \tag{3.13}$$

and

$$\begin{aligned}\hat{\rho}_{ll'} &= \frac{\sum_{h=1}^{p}\left(c_{lh}-\theta_l d_{lh}\right)\left(c_{l'h}-\theta_{l'}d_{l'h}\right)S_h^2/n_h}{\sqrt{\left(\sum_{h=1}^{p}\left(c_{lh}-\theta_l d_{lh}\right)^2 S_h^2/n_h\right)\left(\sum_{h=1}^{p}\left(c_{l'h}-\theta_{l'}d_{l'h}\right)^2 S_h^2/n_h\right)}} \\ &= \frac{\left(\boldsymbol{c}_l-\theta_l\boldsymbol{d}_l\right)'\hat{\boldsymbol{V}}\boldsymbol{M}\left(\boldsymbol{c}_{l'}-\theta_{l'}\boldsymbol{d}_{l'}\right)}{\sqrt{\left(\boldsymbol{c}_l-\theta_l\boldsymbol{d}_l\right)'\hat{\boldsymbol{V}}\boldsymbol{M}\left(\boldsymbol{c}_l-\theta_l\boldsymbol{d}_l\right)}\sqrt{\left(\boldsymbol{c}_{l'}-\theta_{l'}\boldsymbol{d}_{l'}\right)'\hat{\boldsymbol{V}}\boldsymbol{M}\left(\boldsymbol{c}_{l'}-\theta_{l'}\boldsymbol{d}_{l'}\right)}} \quad (3.14) \\ &\quad (1\le l,l'\le q).\end{aligned}$$

Hence, $T_1,\ldots,T_q$ follow an unknown joint q-variate distribution with correlation matrix $\boldsymbol{R}=(\rho_{ll'})_{l,l'}$. The procedures HOM, GH, PI and HTL can now be defined in the same manner as in the case of differences of means. The decision rule for testing problem (3.9) is also the same. H_{0l} is rejected for each ratio of contrasts γ_l for which (3.8) holds.

3.3 α-simulations

The aim of adjusting the degrees of freedom and the correlations between the contrasts is to control the FWE, balancing conservative and liberal behavior. All methods described in Section (3.2) are approximate ones, so their quality must be validated by simulations. For both difference-based (Section 3.3.1) and ratio-based (Section 3.3.2) MCTs, respectively, three treatments have been compared in a first simulation study, five in a second one. The first treatment is regarded as the (negative) control. The FWE has been simulated; the nominal level is 0.05. Four different settings have been considered, each setting with a total sample size of 30 (three treatments) and 50 (five treatments), respectively. They are:

a) a balanced allocation; the last group has the largest standard deviation:

$n_h : 10, 10, 10,\ \sigma_h : 10, 10, 50,$

$n_h : 10, 10, 10, 10, 10,\ \sigma_h : 10, 10, 10, 10, 50,$

b) the first group (control) has the smallest sample size; the last group has the largest standard deviation:

$n_h : 4, 13, 13,\ \sigma_h : 10, 10, 50,$

$n_h : 6, 11, 11, 11, 11,\ \sigma_h : 10, 10, 10, 10, 50,$

c) the last group has the smallest sample size; the last group has the largest standard deviation:

$n_h : 13, 13, 4,\ \sigma_h : 10, 10, 50,$

$n_h : 11, 11, 11, 11, 6,\ \sigma_h : 10, 10, 10, 10, 50,$

d) a balanced allocation; the homoscedastic case:

$n_h : 10, 10, 10,\ \sigma_h : 30, 30, 30,$

$n_h : 10, 10, 10, 10, 10,\ \sigma_h : 30, 30, 30, 30, 30.$

The expected values of the treatment groups for the difference-based MCTs are equal, that is $\mu_h = 100$ $(h = 1, \ldots, p)$. For the ratio-based MCTs they are $\mu_1 = 100$ and $\mu_h = 125$ $(h = 2, \ldots, p)$, and $\theta_l = 1.25$ $(l = 1, \ldots, q)$. The value 100 has been chosen (usually zero) because the sample means should not have different algebraic signs when considering ratios. The settings a), b) and c) imply very high standard deviations for the last group, i.e. a coefficient of variation of 50% for the control group, to intensify possible differences between the procedures. All the following

simulation results have been obtained from 100000 simulation runs, with the same starting seed (seed 10000), using a program code in the statistical software R [2008], package `mvtnorm` [Genz et al., 2008, Hothorn et al., 2001].

3.3.1 Differences of Means

We have considered five one-sided difference-based MCT problems which are all related to the hypotheses (3.1): Dunnett, Tukey[2], Williams, Changepoint, and Average. Table 3.1 (3.2) shows the results of the first (second) study for difference-based MCTs with three (five) treatments. Depending on the setting, GH and HTL tend to either conservatism or liberalism, respectively. Setting c) for three treatments (Table 3.1), combining the smallest sample size and the highest standard deviation, leads to especially liberal behavior for HTL (0.078 for the Dunnett contrast). This is not so obvious for five treatments (Table 3.2) because the part of treatments with equal variances is higher there. One can also see from Table 3.2 that the Tukey contrast generally seems to cause conservatism for HTL (0.036 for setting a)). GH seems to deviate from the nominal α-level independent of the setting, but it depends more strongly on the particular contrasts. The lowest level is achieved for the Changepoint contrast (0.034 for setting a), Table 3.2). Therefore, these procedures cannot be recommend without reservations. With few exceptions, PI maintains the α-level exactly. It only varies from 0.047 to 0.055, while GH has ranges from 0.034 to 0.062, and HTL from 0.036 to 0.078. HOM is liberal for setting a), where the sample sizes

[2]Normally, a Tukey MCT is a two-sided test problem. For reasons of consistency, this fact is disregarded.

Setting	MCT	HOM	GH	PI	HTL
	Dunnett	0.063	0.053	0.049	0.051
	Tukey	0.059	0.042	0.049	0.041
a)	Williams	0.073	0.043	0.049	0.050
	Changepoint	0.088	0.038	0.049	0.050
	Average	0.076	0.044	0.048	0.046
	Dunnett	0.011	0.062	0.052	0.055
	Tukey	0.031	0.050	0.055	0.050
b)	Williams	0.015	0.049	0.050	0.050
	Changepoint	0.047	0.041	0.051	0.050
	Average	0.044	0.044	0.050	0.044
	Dunnett	0.208	0.054	0.051	0.078
	Tukey	0.211	0.043	0.048	0.061
c)	Williams	0.213	0.045	0.049	0.059
	Changepoint	0.222	0.042	0.048	0.059
	Average	0.248	0.053	0.054	0.070
	Dunnett	0.049	0.048	0.048	0.048
	Tukey	0.049	0.049	0.049	0.045
d)	Williams	0.049	0.050	0.050	0.050
	Changepoint	0.049	0.050	0.050	0.049
	Average	0.049	0.050	0.050	0.049

Table 3.1: *FWE of one-sided MCTs (differences) for* $p = 3$ *treatments, several contrasts, procedures and settings;* $\boldsymbol{\mu} = (100, 100, 100)'$, $\alpha = 0.05$.

Setting	MCT	HOM	GH	PI	HTL
	Dunnett	0.081	0.055	0.051	0.051
	Tukey	0.077	0.043	0.051	0.036
a)	Williams	0.104	0.043	0.049	0.049
	Changepoint	0.129	0.034	0.049	0.049
	Average	0.114	0.042	0.047	0.046
	Dunnett	0.050	0.058	0.049	0.048
	Tukey	0.065	0.045	0.052	0.039
b)	Williams	0.070	0.051	0.053	0.050
	Changepoint	0.115	0.035	0.051	0.049
	Average	0.101	0.043	0.049	0.048
	Dunnett	0.162	0.056	0.053	0.060
	Tukey	0.159	0.043	0.051	0.041
c)	Williams	0.184	0.045	0.050	0.054
	Changepoint	0.197	0.036	0.048	0.051
	Average	0.188	0.046	0.049	0.058
	Dunnett	0.048	0.049	0.048	0.048
	Tukey	0.050	0.051	0.052	0.043
d)	Williams	0.050	0.050	0.049	0.048
	Changepoint	0.049	0.052	0.052	0.052
	Average	0.050	0.053	0.053	0.050

Table 3.2: *FWE of one-sided MCTs (differences) for* $p = 5$ *treatments, several contrasts, procedures and settings;* $\boldsymbol{\mu} = (100, 100, 100, 100, 100)'$, $\alpha = 0.05$.

are balanced (0.114 for the Average contrast, Table 3.2), but most liberal for setting c), where the smallest sample size and the highest standard deviation are combined (0.248 for the Average contrast, Table 3.1). Setting b) causes conservativism for three treatments (0.011 for the Dunnett contrast, Table 3.1), and liberalism for five (0.115 for the Changepoint contrast, Table 3.2). The reason lies in the different relations between variances and sample sizes in setting b) for three and five treatments, respectively.

We have again treated the situation of the Dunnett contrast to get an impression of the strong control of the FWE. Let the first comparison be known to reject its local null hypothesis because the second treatment significantly differs from the remaining ones, i.e., $\mu_2 = 1000$. Only the FWE of the remaining comparisons, denoted by *local* FWE, has been considered. Tables 3.3 and 3.4 show the α-level of these procedures ignoring the second treatment. HOM clearly fails for setting a) (0.080, Table 3.4) and c) (0.207, Table 3.3), but HTL also fails for setting c) for three treatments (0.059, Table 3.3). All procedures applied are similar to those applied in the preceding

Setting	HOM	GH	PI	HTL
a)	0.063	0.028	0.026	0.029
b)	0.011	0.031	0.026	0.022
c)	0.207	0.028	0.027	0.059
d)	0.028	0.027	0.027	0.027

Table 3.3: *Local FWE of one-sided MCTs (differences) for $p = 3$ treatments, the Dunnett contrast, several procedures and settings; $\boldsymbol{\mu} = (100, 1000, 100)'$, $\alpha = 0.05$.*

Setting	HOM	GH	PI	HTL
a)	0.080	0.043	0.039	0.040
b)	0.050	0.050	0.043	0.041
c)	0.162	0.045	0.042	0.051
d)	0.040	0.040	0.040	0.040

Table 3.4: *Local FWE of one-sided MCTs (differences) for* $p = 5$ *treatments, the Dunnett contrast, several procedures and settings;* $\boldsymbol{\mu} = (100, 1000, 100, 100, 100)'$, $\alpha = 0.05$.

simulation.

3.3.2 Ratios of Means

Here, we focus only on one-sided ratio-based Dunnett MCT problems related to the hypotheses (3.9). Table 3.5 (3.6) shows the results of the first (second) study for ratio-based Dunnett MCTs with three (five) treatments. The procedures GH and HTL tend to liberalism here. As in the case of differences, setting c) for three treatments leads to liberal behavior for HTL (0.077, Table 3.5), but also for five treatments (0.059, Table 3.6). GH is most liberal for setting b) (0.065, Table 3.5). These procedures cannot be recommended without reservations. PI seems to maintain the α-level exactly and just as well as in the case of differences. It varies from 0.049 to 0.053 while GH has ranges from 0.049 to 0.065, and HTL from 0.049 to 0.077. HOM is liberal for setting a) with five treatments (0.063, 3.6), and generally most liberal for setting c) (0.197, Table 3.5). Setting b) causes conservativism,

Setting	HOM	GH	PI	HTL
a)	0.047	0.056	0.050	0.051
b)	0.006	0.065	0.053	0.059
c)	0.197	0.055	0.051	0.077
d)	0.049	0.049	0.049	0.049

Table 3.5: *FWE of one-sided MCTs (ratios) for* $p = 3$ *treatments, the Dunnett contrast, several procedures and settings;* $\boldsymbol{\mu} = (100, 125, 125)'$, $\boldsymbol{\theta} = (1.25, 1.25)$, $\alpha = 0.05$.

especially for three treatments (0.006, Table 3.5).

For an appreciation of the strong control of the FWE, we have proceeded similarly as in the case of differences above. Let the second treatment significantly differ from the remaining ones, i.e., $\mu_2 = 1000$. Only the FWE of the remaining comparisons (local FWE) has been considered. Tables 3.7 and 3.8 show the α-level of these procedures ignoring the second treatment. HOM clearly fails for setting c) (0.195) in Table 3.7 and for setting a) (0.064) and c) (0.147) in Table 3.8, HTL for setting c) (0.058) in Table 3.7. Again, the procedures applied are similar to those applied in the preceding simulation.

3.3.3 Conclusions

The reason why the procedures differ in their behavior becomes clearer when having a look at the parameters of the underlying distributions. As an example, take

Setting	HOM	GH	PI	HTL
a)	0.063	0.056	0.050	0.050
b)	0.034	0.062	0.050	0.049
c)	0.146	0.057	0.052	0.059
d)	0.050	0.050	0.049	0.049

Table 3.6: *FWE of one-sided MCTs (ratios) for* $p = 5$ *treatments, the Dunnett contrast, several procedures and settings;* $\boldsymbol{\mu} = (100, 125, \ldots, 125)'$, $\boldsymbol{\theta} = (1.25, 1.25, 1.25, 1.25)$, $\alpha = 0.05$.

Setting	HOM	GH	PI	HTL
a)	0.047	0.029	0.026	0.029
b)	0.006	0.032	0.026	0.022
c)	0.195	0.029	0.027	0.058
d)	0.029	0.028	0.028	0.028

Table 3.7: *Local FWE of one-sided MCTs (ratios) for* $p = 3$ *treatments, the Dunnett contrast, several procedures and settings;* $\boldsymbol{\mu} = (100, 1000, 125)'$, $\boldsymbol{\theta} = (1.25, 1.25)'$, $\alpha = 0.05$.

Setting	HOM	GH	PI	HTL
a)	0.064	0.045	0.040	0.041
b)	0.034	0.054	0.043	0.041
c)	0.147	0.048	0.043	0.052
d)	0.040	0.040	0.039	0.039

Table 3.8: *Local FWE of one-sided MCTs (ratios) for* $p = 5$ *treatments, the Dunnett contrast, several procedures and settings;* $\boldsymbol{\mu} = (100, 1000, 125, 125, 125)'$, $\boldsymbol{\theta} = (1.25, 1.25, 1.25, 1.25)$, $\alpha = 0.05$.

setting c) for the Tukey contrast (Table 3.1). HOM is absolutely liberal (0.211), GH conservative (0.043), PI has an (almost) exact α-level (0.048), and HTL is liberal (0.061). For a single simulation run (the same for all), there are $df = 27$ for HOM, $df_1^* = 23.65$, $df_2^* = 3.17$, $df_3^* = 3.22$ for GH and PI, and $df^{**} = 10.01$ for HTL. The corresponding correlation matrices are

$$\boldsymbol{R} = \begin{pmatrix} 1 & 0.34 & -0.34 \\ 0.34 & 1 & 0.76 \\ -0.34 & 0.76 & 1 \end{pmatrix} \quad \text{(HOM, GH)},$$

$$\boldsymbol{R}^* = \begin{pmatrix} 1 & 0.11 & -0.14 \\ 0.11 & 1 & 0.97 \\ -0.14 & 0.97 & 1 \end{pmatrix} \quad \text{(PI)},$$

$$\boldsymbol{R}^{**} = \begin{pmatrix} 1 & 0.31 & 0.31 \\ 0.31 & 1 & 0.31 \\ 0.31 & 0.31 & 1 \end{pmatrix} \quad \text{(HTL)}.$$

The improvement of GH over HOM is obviously the use of several adjusted degrees of freedom ν_l^* according to Satterthwaite [1946]. However, the final step to handle heteroscedasticity is to take the estimators of the variances S_h^2 into account within the correlations (3.7) and (3.13). They will be significantly corrected in this way; the correlations in GH and PI clearly differ. Thus, an important conclusion is that an adjustment of only the degrees of freedom is not sufficient. On the other hand, the HTL procedure, taking averages of correlations and degrees of freedom, is generally too rough. This becomes especially clear in the case of Tukey contrasts because also negative correlations appear there.

Figure 3.3 gives a graphical explanation for the PI procedure. The joint distribution of T_1 and T_2 is shown (gray dots) against the background of the simulations for the Dunnett procedure for the case of differences with $p = 3$ groups. The two different bivariate t-distributions are illustrated as contour plots, where the red lines belong to T_1, and the blue lines belong to T_2. Realizations for (T_1, T_2) that lead to rejection are marked by crosses, the remaining ones by small circles. The two related quantiles are equal only for setting d), the homoscedastic and balanced case. Generally, they are not equidistant.

According to these investigations, the PI procedure maintains the α-level exactly. Negligible variations about the nominal α-level are due to the fact that the correlations are estimated ones, but a serious violation does not occur. After these remarks, it should be clear that the HOM procedure should be used only with utmost care and cannot be recommended in the presence of heteroscedasticity. The development

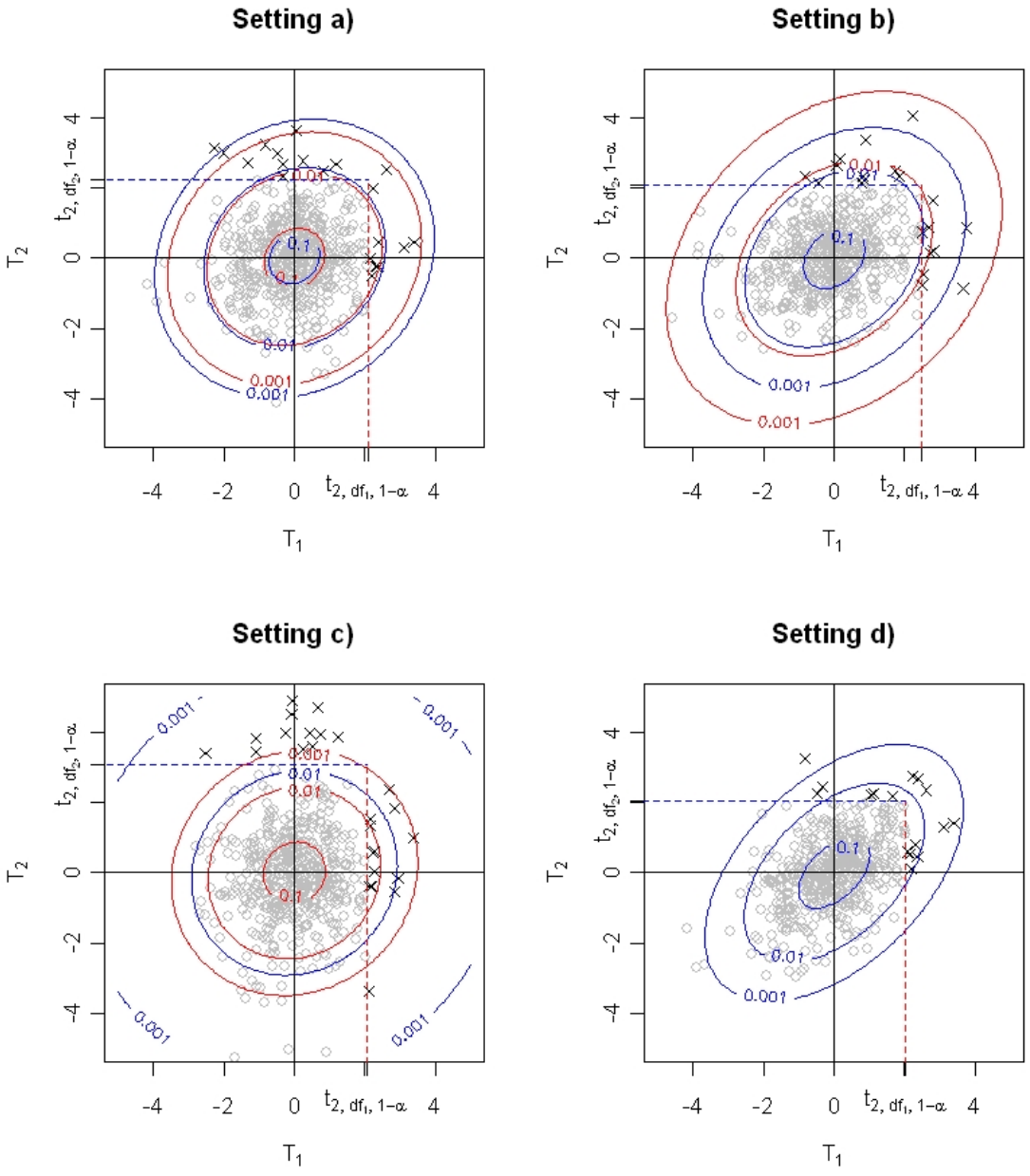


Figure 3.3: *Distribution of the test statistics for the Dunnett contrast;* $\mu_1 = 100, \mu_2 = 100, \mu_3 = 100$.

of new procedures was a necessary consequence. Because PI has the best properties, the following considerations refer only to PI.

3.4 Simultaneous Confidence Intervals

3.4.1 Definition

The *acceptance region* of a statistical test is defined as the set of sample values for which H_0 is accepted. Inversion of the acceptance region yields the (*simultaneous*)*confidence set* (*CS* or *SCS*, respectively). This is a set of parameter values for which a fixed sample belongs to the acceptance region. The probability that the (true) parameter value is covered by the confidence set for each component is defined as (*simultaneous*) *coverage probability* (*CP* or *SCP*, respectively). If the test has level α, then the confidence set has a CP of at least $1 - \alpha$. This is verified by the following

Proof. The probability that a level-α test wrongly rejects H_0 for a given sample and a fixed parameter value is at most α. Hence, the probability that the test correctly accepts H_0 for that sample is at least $1-\alpha$. This is nothing but the probability that the sample belongs to the acceptance region, which is equal to the probability that the parameter value belongs to the confidence set (according to the definition). □

Having a multivariate test problem, one is usually interested in projecting the confi-

dence set onto the coordinate axes for an easier interpretation. The axes correspond to the contrasts here. These projections are defined as *simultaneous confidence intervals* (*SCIs*). However, an exact projecting might not always be possible. We will encounter this problem later. Nevertheless, SCIs are a method to handle both parameter estimation and parameter testing.

3.4.2 Differences of Means

Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_q)'$ be a point in the parameter space of $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_q)'$ and let higher values of the data X_{hj} represent a better effect of the treatments. The $(1-\alpha)100\%$ confidence set for the statistical problem (3.1) is given by

$$\begin{aligned} C\left((x,y)\right) &= \left\{\boldsymbol{\xi} : T_l(\xi_l) \leq t^l_{q,1-\alpha}(\hat{\nu}_l, \hat{\boldsymbol{R}}), \quad l = 1, \ldots, q\right\} \\ &= \left\{\boldsymbol{\xi} : \hat{\eta}_l^{lower} \leq \xi_l, \quad l = 1, \ldots, q\right\}, \end{aligned}$$

where the lower limits $\hat{\eta}_l^{lower}$ of the approximate $(1-\alpha)100\%$ SCIs for $\boldsymbol{\eta}$ are defined as

$$\begin{aligned} \hat{\eta}_l^{lower} &= \left(\sum_{h=1}^{p} c_{lh}\bar{X}_h\right) - t^l_{q,1-\alpha}(\hat{\nu}_l, \hat{\boldsymbol{R}})\sqrt{\sum_{h=1}^{p} c_{lh}^2 S_h^2/n_h} \\ &= \boldsymbol{c}_l'\bar{\boldsymbol{X}} - t^l_{q,1-\alpha}(\hat{\nu}_l, \hat{\boldsymbol{R}})\sqrt{\boldsymbol{c}_l'\hat{\boldsymbol{V}}\boldsymbol{M}\boldsymbol{c}_l} \quad (l = 1, \ldots, q). \end{aligned}$$

These limits can be used for the statistical problem (3.1). For a specified level α, we reject H_{0l} for each contrast η_l with

$$\hat{\eta}_l^{lower} > \delta_l.$$

For the two-sided case, we obtain

$$\begin{aligned} C\left((x,y)\right) &= \left\{\boldsymbol{\xi} : |T_l(\xi_l)| \leq t^{ts}_{q,1-\alpha}(\hat{\nu}_l, \hat{\boldsymbol{R}}), \quad l = 1,\ldots,q\right\} \\ &= \left\{\boldsymbol{\xi} : \hat{\eta}_l^{lower} \leq \xi_l \leq \hat{\eta}_l^{upper}, \quad l = 1,\ldots,q\right\} \end{aligned}$$

and confidence limits

$$\begin{aligned} \hat{\eta}_l^{lower} &= \left(\sum_{h=1}^{p} c_{lh}\bar{X}_h\right) - t^{ts}_{q,1-\alpha}(\hat{\nu}_l, \hat{\boldsymbol{R}})\sqrt{\sum_{h=1}^{p} c_{lh}^2 S_h^2/n_h}, \\ &= \boldsymbol{c}_l'\bar{\boldsymbol{X}} - t^{ts}_{q,1-\alpha}(\hat{\nu}_l, \hat{\boldsymbol{R}})\sqrt{\boldsymbol{c}_l'\hat{\boldsymbol{V}}\boldsymbol{M}\boldsymbol{c}_l} \quad (l = 1,\ldots,q) \\ \hat{\eta}_l^{upper} &= \left(\sum_{h=1}^{p} c_{lh}\bar{X}_h\right) + t^{ts}_{q,1-\alpha}(\hat{\nu}_l, \hat{\boldsymbol{R}})\sqrt{\sum_{h=1}^{p} c_{lh}^2 S_h^2/n_h} \\ &= \boldsymbol{c}_l'\bar{\boldsymbol{X}} + t^{ts}_{q,1-\alpha}(\hat{\nu}_l, \hat{\boldsymbol{R}})\sqrt{\boldsymbol{c}_l'\hat{\boldsymbol{V}}\boldsymbol{M}\boldsymbol{c}_l} \quad (l = 1,\ldots,q). \end{aligned}$$

For a specified level α, we reject H_{0l} for each contrast η_l with

$$\hat{\eta}_l^{lower} > \delta_l \quad \text{or} \quad \hat{\eta}_l^{\text{upper}} < \delta_l.$$

Figure 3.4 shows the influence of heteroscedasticity on the two-sided $(1-\alpha)100\%$ confidence set for differences of means with a Dunnett contrast based on a hypothetical dataset with $p = 3$ treatments. The confidence set widens for increasing variances, but it becomes widest for that contrast (η_2) with the highest variance. Hence, the resulting SCIs, despite being symmetric, do not have the same width.

3.4.3 Ratios of Means

Let $\boldsymbol{\xi} = (\xi_1,\ldots,\xi_q)'$ be a point in the parameter space of $\boldsymbol{\gamma} = (\gamma_1,\ldots,\gamma_q)'$. For the case that higher values of the data, X_{hj}, represent a better effect of the treatments,

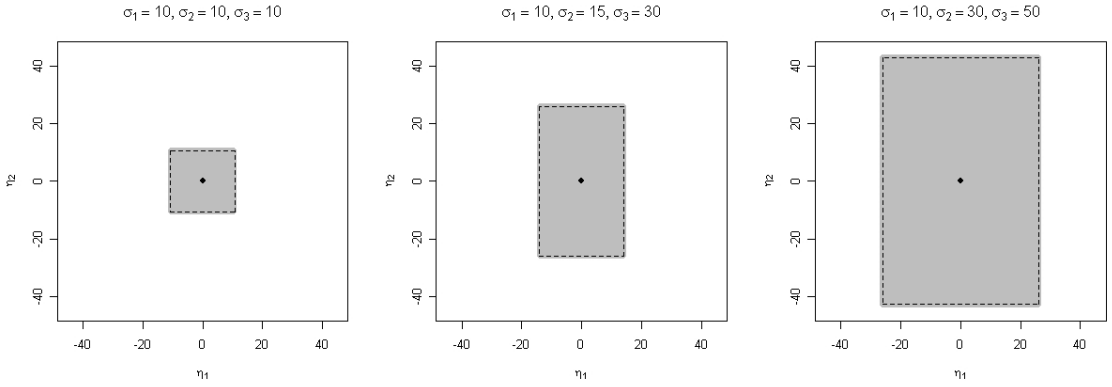


Figure 3.4: *Two-sided 95% confidence set for the Dunnett contrast of $p = 3$ treatments; $n_1 = n_2 = n_3 = 10$, $\boldsymbol{\mu} = (100, 100, 100)'$.*

the $(1-\alpha)100\%$ confidence set for the statistical problem (3.9) is given by

$$\begin{aligned} C\left((x,y)\right) &= \left\{\boldsymbol{\xi} : T_l(\xi_l) \leq t^l_{q,1-\alpha}(\hat{\nu}_l, \hat{\boldsymbol{R}}), \quad l = 1,\ldots,q\right\} \\ &= \left\{\boldsymbol{\xi} : A_l\xi_l^2 + B_l\xi_l + C_l \leq 0, \quad l = 1,\ldots,q\right\}, \end{aligned}$$

where

$$\begin{aligned} A_l &= \left(\sum_{h=1}^{p} d_{lh}\bar{X}_h\right)^2 - \left(t^l_{q,1-\alpha}(\hat{\nu}_l, \hat{\boldsymbol{R}})\right)^2 \sum_{h=1}^{p} d_{lh}^2 S_h^2/n_h \\ &= \left(\boldsymbol{d}_l'\bar{\boldsymbol{X}}\right)^2 - \left(t^l_{q,1-\alpha}(\hat{\nu}_l, \hat{\boldsymbol{R}})\right)^2 \boldsymbol{d}_l'\hat{\boldsymbol{V}}\boldsymbol{M}\boldsymbol{d}_l, \\ B_l &= -2\left(\left(\sum_{h=1}^{p} c_{lh}\bar{X}_h\right)\left(\sum_{h=1}^{p} d_{lh}\bar{X}_h\right) - \left(t^l_{q,1-\alpha}(\hat{\nu}_l, \hat{\boldsymbol{R}})\right)^2 \sum_{h=1}^{p} c_{lh}d_{lh}S_h^2/n_h\right) \\ &= -2\left(\left(\boldsymbol{c}_l'\bar{\boldsymbol{X}}\right)\left(\boldsymbol{d}_l'\bar{\boldsymbol{X}}\right) - \left(t^l_{q,1-\alpha}(\hat{\nu}_l, \hat{\boldsymbol{R}})\right)^2 \boldsymbol{c}_l'\hat{\boldsymbol{V}}\boldsymbol{M}\boldsymbol{d}_l\right), \\ C_l &= \left(\sum_{h=1}^{p} c_{lh}\bar{X}_h\right)^2 - \left(t^l_{q,1-\alpha}(\hat{\nu}_l, \hat{\boldsymbol{R}})\right)^2 \sum_{h=1}^{p} c_{lh}^2 S_h^2/n_h \\ &= \left(\boldsymbol{c}_l'\bar{\boldsymbol{X}}\right)^2 - \left(t^l_{q,1-\alpha}(\hat{\nu}_l, \hat{\boldsymbol{R}})\right)^2 \boldsymbol{c}_l'\hat{\boldsymbol{V}}\boldsymbol{M}\boldsymbol{c}_l. \end{aligned} \tag{3.15}$$

This approach is based on Fieller's Theorem [Fieller, 1954]. In contrast to SCIs for differences, the correlation matrix $\hat{\boldsymbol{R}}$ depends here on the unknown ratios γ_l, say

$\hat{\rho}_{ll'} = \hat{\rho}_{ll'}(\gamma_l, \gamma_{l'})$. Dilba et al. [2006] have used a plug-in approach in the homoscedastic case and have shown a very good performance as compared to other methods. The same problem also arises for the degrees of freedom, i.e., $\hat{\nu}_l = \hat{\nu}_l(\gamma_l)$. Now the estimator

$$\begin{aligned} \hat{\gamma}_l &= \frac{\sum_{h=1}^{p} c_{lh}\bar{X}_h}{\sum_{h=1}^{p} d_{lh}\bar{X}_h} \\ &= \frac{\boldsymbol{c}_l'\bar{\boldsymbol{X}}}{\boldsymbol{d}_l'\bar{\boldsymbol{X}}} \quad (l = 1, \ldots, q), \end{aligned}$$

has to be used in Equations (3.12) and (3.14) instead of θ_l. The lower limits of the approximate $(1-\alpha)100\%$ SCIs for $(\gamma_1, \ldots, \gamma_q)'$ are hence given by

$$\hat{\gamma}_l^{lower} = \frac{-B_l - \sqrt{B_l^2 - 4A_lC_l}}{2A_l} \quad (l = 1, \ldots, q).$$

If $A_l > 0$, then it can be shown that the solution is finite (see, e.g., Buonaccorsi and Iyer [1984] for the homoscedastic case). The statistical problem (3.9) can be decided as follows: For a specified level α, we reject H_{0l} for each contrast γ_l with

$$\hat{\gamma}_l^{lower} > \theta_l.$$

For the two-sided case, we obtain

$$\begin{aligned} C\left((x,y)\right) &= \left\{\boldsymbol{\xi} : |T_l(\xi_l)| \leq t_{q,1-\alpha}^{ts}(\hat{\nu}_l, \hat{\boldsymbol{R}}), \quad l = 1, \ldots, q\right\} \\ &= \left\{\boldsymbol{\xi} : A_l\xi_l^2 + B_l\xi_l + C_l \leq 0, \quad l = 1, \ldots, q\right\}, \end{aligned}$$

where the A_l, B_l and C_l are defined as in (3.15) but with quantiles $t_{q,1-\alpha}^{ts}(\hat{\nu}_l, \hat{\boldsymbol{R}})$ instead of $t_{q,1-\alpha}^{l}(\hat{\nu}_l, \hat{\boldsymbol{R}})$. The confidence limits are given by

$$\begin{aligned} \hat{\gamma}_l^{lower} &= \frac{-B_l - \sqrt{B_l^2 - 4A_lC_l}}{2A_l} \quad (l = 1, \ldots, q), \\ \hat{\gamma}_l^{upper} &= \frac{-B_l + \sqrt{B_l^2 - 4A_lC_l}}{2A_l} \quad (l = 1, \ldots, q). \end{aligned} \tag{3.16}$$

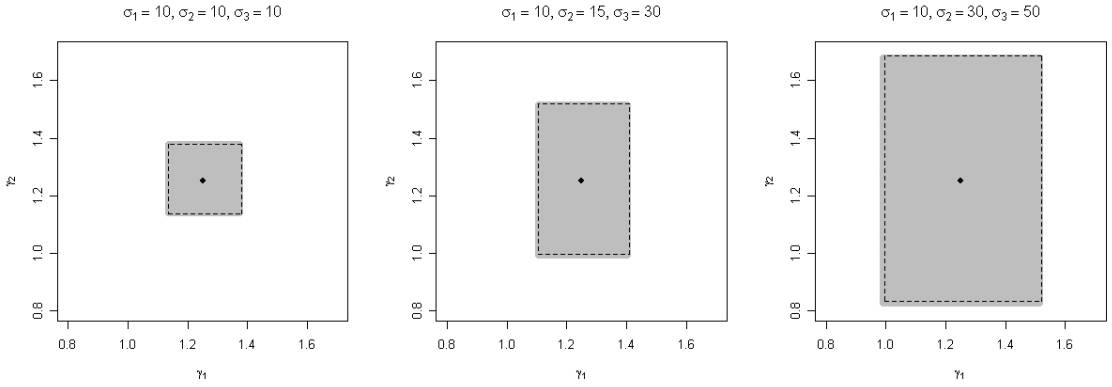


Figure 3.5: *Two-sided 95% confidence set for the Dunnett contrast of* $p = 3$ *treatments;* $n_1 = n_2 = n_3 = 10$, $\boldsymbol{\mu} = (100, 125, 125)'$.

For a specified level α, we reject H_{0l} for each contrast γ_l with

$$\hat{\gamma}_l^{lower} > \theta_l \quad \text{or} \quad \hat{\gamma}_l^{\text{upper}} < \theta_l.$$

Figure 3.5 shows the influence of heteroscedasticity on the two-sided $(1-\alpha)100\%$ confidence set for ratios of means with a Dunnett contrast based on a hypothetical dataset with $p = 3$ treatments. As in the case of differences, the confidence set widens for increasing variances, and it becomes widest for that contrast (γ_2) with the highest variance. The resulting SCIs do not have the same width. In contrast to the case of differences, the confidence set and the SCIs are not symmetric in general, regardless of heteroscedasticity.

The strict one-to-one relation between the test decisions of MCTs and SCIs that holds in the case of differences of means does not hold in the case of ratios of means. This is due to the additional use of the estimator $\hat{\gamma}_l$ instead of θ_l in the calculation of the necessary quantiles. The conclusions of the α-simulations in 3.3.2 are not

Setting	$\gamma_1 = \gamma_2$			
	0.5	1.0	1.5	2.0
a)	0.950	0.950	0.949	0.948
b)	0.952	0.948	0.950	0.951
c)	0.948	0.949	0.944	0.943
d)	0.947	0.951	0.951	0.951

Table 3.9: *SCP of one-sided (upper) SCIs (ratios) for $p = 3$ treatments, the Dunnett contrast and several settings and ratios $\gamma_1 = \gamma_2$; $\mu_1 = 100$, $\alpha = 0.05$.*

applicable for the related SCIs without caution. We have thus performed further simulation studies to describe the behavior of SCIs related to the PI procedure. The same background as in 3.3.2 has been used and the simultaneous coverage probability (SCP) has been simulated. The nominal level is 0.95 for all studies. Table 3.9 (3.10) shows the results of the first (second) study for ratio-based Dunnett SCIs with three (five) treatments, with the underlying settings and depending on the ratios $\gamma_1 = \gamma_2$ ($\gamma_1 = \ldots = \gamma_4$). Only for a few cases (0.942, setting c), Table 3.10), a little liberalism is observed, but not a unique influence or trend (ranges from 0.942 to 0.952). In principle, the expected value 0.95 is attained for all the settings. This reflects the results of the α-simulations.

The reason for the possible discrepancy between test decisions according to MCTs and related SCIs is shown in the Figures 3.6 and 3.7. The area covered by the SCIs according to (3.16) (dashed lines) does not cover the confidence set completely. An exact projection is not possible here because of the non-rectangular shape of the con-

Setting	$\gamma_1 = \ldots = \gamma_4$			
	0.5	1.0	1.5	2.0
a)	0.950	0.948	0.947	0.950
b)	0.952	0.952	0.950	0.951
c)	0.947	0.947	0.947	0.942
d)	0.942	0.951	0.951	0.947

Table 3.10: *SCP of one-sided (upper) SCIs (ratios) for* $p = 5$ *treatments, the Dunnett contrast and several settings and ratios* $\gamma_1 = \ldots = \gamma_4$*;* $\mu_1 = 100$*,* $\alpha = 0.05$*.*

fidence set. This problem is independent of the issue of homo- or heteroscedasticity (see Dilba et al. [2006] and Dilba [2005]).

3.5 Power Considerations

The testing problem (3.1) for differences of means is simplified here to the case of equal thresholds, $\delta_l = \delta$ for all $l = 1, \ldots, q$. Let higher response values indicate better treatment effects and let ϵ^* denote the greatest irrelevant difference to the control mean which is to be detected. Define the set of indices $I(\epsilon^*) = \{l : \epsilon_l > \epsilon^*\} = \{l_1, \ldots, l_m\}$ $(m = 1, \ldots, q)$. All contrasts with ϵ_l values greater than ϵ^* are relevant. The probability to detect all relevant contrasts is defined as the *complete* (or *all-pairs*) *power*. An (approximate) expression for the complete power of statistical problem (3.1) is given by

$$P\left\{T_l > t^l_{q,1-\alpha}(\nu_l, \boldsymbol{R}) \middle| \psi_l, \sigma_1^2, \ldots, \sigma_k^2 \quad \forall l \in I(\epsilon^*)\right\}.$$

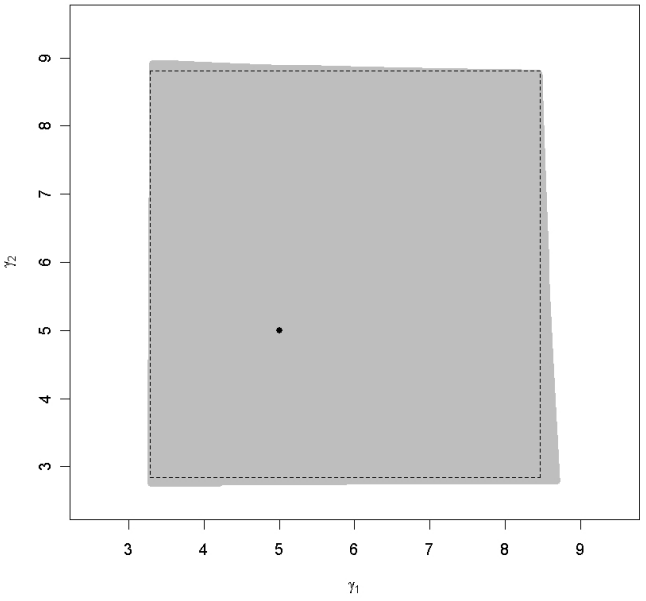


Figure 3.6: *Two-sided 95% confidence set for the Dunnett contrast;* $n_1 = n_2 = n_3 = 10$, $\boldsymbol{\mu} = (20, 100, 100)'$, $s_1 = 10; s_2 = 30; s_3 = 50$.

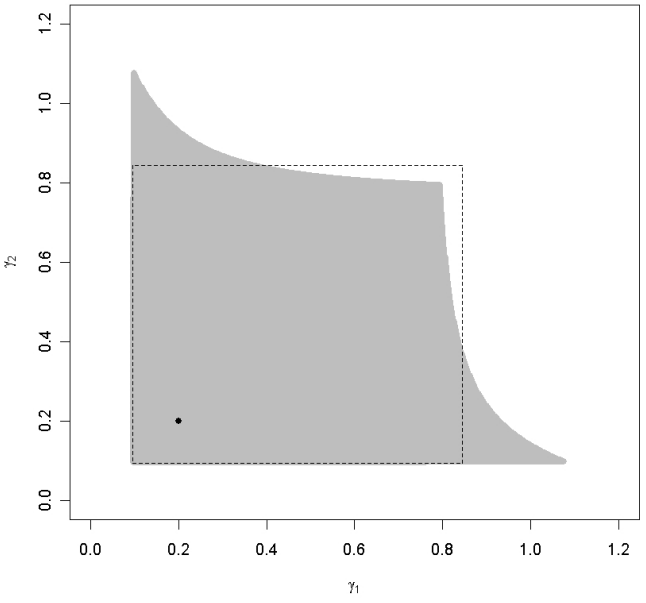


Figure 3.7: *Two-sided 95% confidence set for the Dunnett contrast;* $n_1 = n_2 = n_3 = 10$, $\boldsymbol{\mu} = (100, 20, 20)'$, $s_1 = 100; s_2 = 10; s_3 = 10$.

The probability to detect at least one relevant contrast is defined as the *minimal* (or *any-pair*) *power*. An (approximate) expression for the minimal power of statistical problem (3.1) is given by

$$P\left\{T_l > t^l_{q,1-\alpha}(\nu_l, \boldsymbol{R}) \middle| \psi_l, \sigma_1^2, \ldots, \sigma_k^2 \quad \text{for at least one } l \in I(\epsilon^*)\right\}.$$

As in the case of differences, the testing problem (3.9) for ratios of means is also simplified here to the case that the thresholds are equal, $\theta_l = \theta$ for all $l = 1, \ldots, q$. Let higher response values indicate better treatment effects and τ^* denot the greatest irrelevant ratio to the control mean which is to be detected. Define the set of indices $I(\tau^*) = \{l : \tau_l > \tau^*\} = \{l_1, \ldots, l_m\}$ $(m = 1, \ldots, q)$. All ratios of contrasts with τ_l values greater than τ^* are relevant. In the same manner as above, an (approximate) expression for the complete power of statistical problem (3.9) is given by

$$P\left\{T_l > t^l_{q,1-\alpha}(\nu_l, \boldsymbol{R}) \middle| \psi_l, \sigma_1^2, \ldots, \sigma_k^2 \quad \forall l \in I(\tau^*)\right\}.$$

The (approximate) expression for the minimal power of statistical problem (3.1) is given by

$$P\left\{T_l > t^l_{q,1-\alpha}(\nu_l, \boldsymbol{R}) \middle| \psi_l, \sigma_1^2, \ldots, \sigma_k^2 \quad \text{for at least one } l \in I(\tau^*)\right\}.$$

Because of heteroscedasticity, adjustments of the degrees of freedom and of the correlations between the test statistics are necessary. This means that in fact the quantiles $t^l_{q,1-\alpha}(\nu_l, \boldsymbol{R})$ are random variables, because they depend on the sample values. Therefore, the above probabilities are only approximate ones. On the other hand, each test statistic T_l will be compared with its own quantile, which comes from

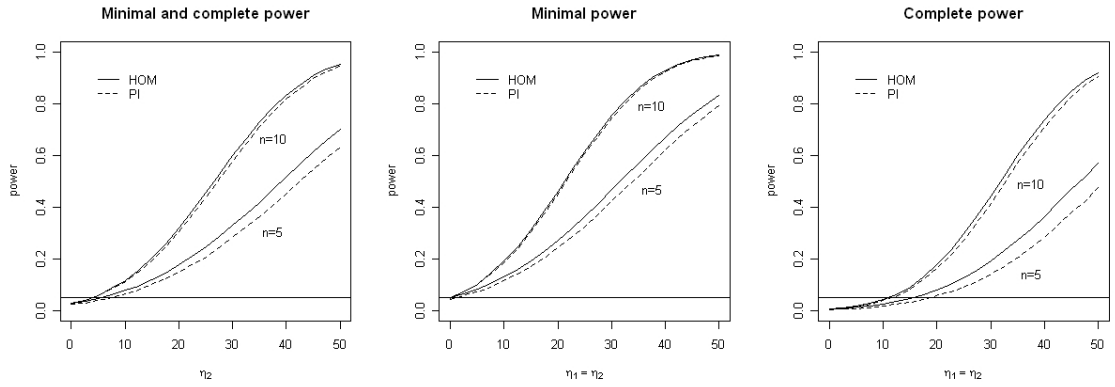


Figure 3.8: *Power comparison of one-sided HOM and PI (differences) for* $p = 3$ *treatments and the Dunnett contrast;* $\mu_1 = 100$, $\alpha = 0.05$.

a q-variate t-distribution with its own degree of freedom ν_l. We do not use a joint q-variate t-distribution for the test statistics (3.2) and (3.10). Power calculations are therefore not possible so far. Power comparison by simulation is possible, however.

The different α-levels of the HOM and the PI procedure do not permit a fair power comparison, especially for situations where HOM does not maintain the FWE. Nevertheless, power is an important dimension. The results of the α-simulations have shown that in a homoscedastic situation all the methods achieve practically the same value, namely the specified α-level. However, the resulting degrees of freedom of the PI or GH procedure are clearly smaller than corresponding ones according to HOM. This may lead to a slight loss in power and to expanded SCIs under the alternative hypothesis. A simulation study has been performed with the balanced and homoscedastic setting described in Section 3.3. In addition, a smaller sample size of $n = 5$ was considered where the sample size n is the same in each group. Figure 3.8 (3.9) shows the results of a power comparison between the one-sided HOM and

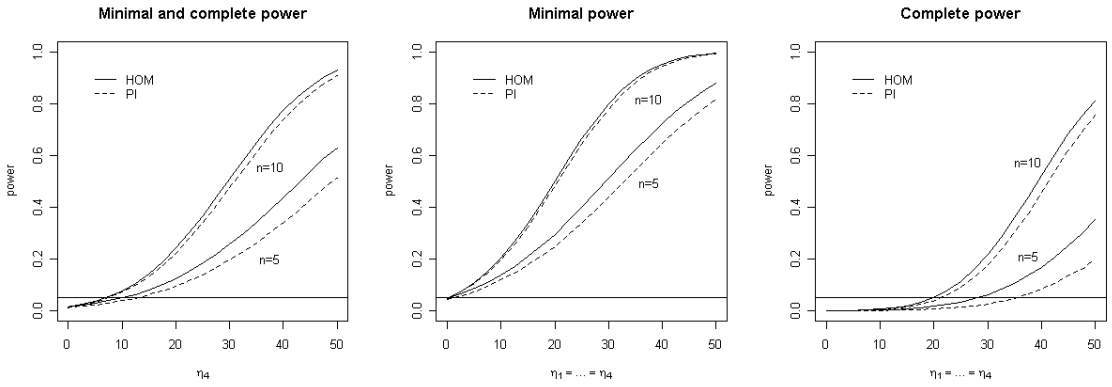


Figure 3.9: *Power comparison of one-sided HOM and PI (differences) for* $p = 5$ *treatments and the Dunnett contrast;* $\mu_1 = 100$, $\alpha = 0.05$.

PI in the case of differences of means, the Dunnett procedure with $p = 3$ ($p = 5$) treatments and $\alpha = 0.05$. Figure 3.10 (3.11) shows the corresponding results for the ratio problem with $\theta_1 = \ldots = \theta_q = 1.25$. The left graphics refer to the case where the first $p - 1$ treatment means are fixed and the last treatment mean has been changed so that the last contrast ($l = q = p - 1$) was variable. Minimal and complete power coincide in this case. The middle (right) graphics show the minimal (complete) power when all the non-control treatment means ($h = 2, \ldots, p$) have been changed simultaneously and in the same amount so that all contrasts $l = 1, \ldots, q$ were variable. HOM and PI have the same α-level here and differ only by negligible power amounts for $n = 10$. This difference increases for $n = 5$ because the relative difference for the degrees of freedom between HOM and PI increases for decreasing sample size.

In practice, it is hard to decide whether the data are homoscedastic or not. If not, three scenarios are possible. If the smallest sample size matches (approximately)

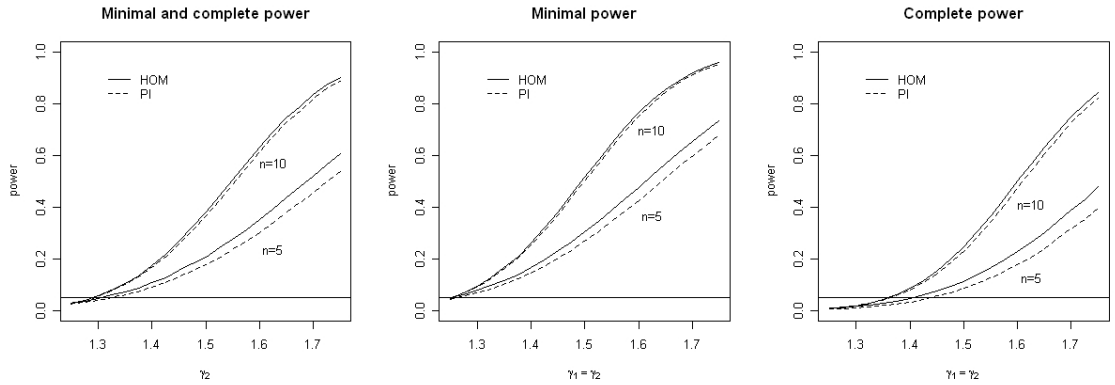


Figure 3.10: *Power comparison of one-sided HOM and PI (ratios) for* $p = 3$ *treatments and the Dunnett contrast;* $\mu_1 = 100$, $\alpha = 0.05$.

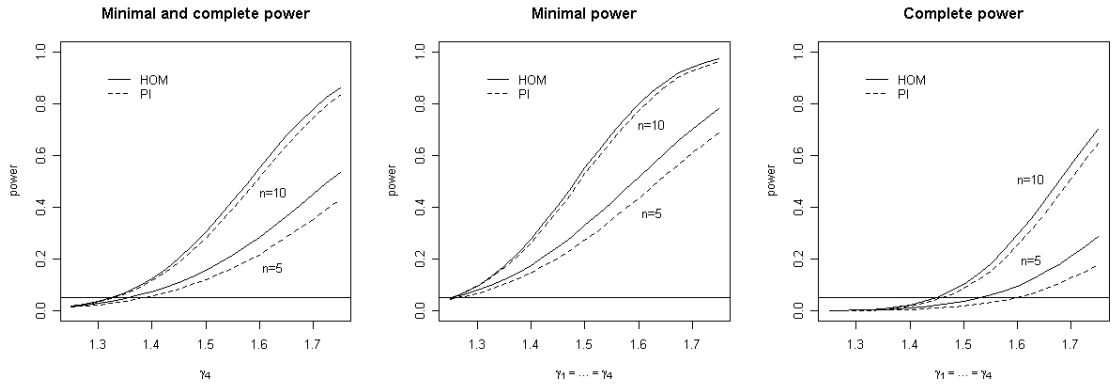


Figure 3.11: *Power comparison of one-sided HOM and PI (ratios) for* $p = 5$ *treatments and the Dunnett contrast;* $\mu_1 = 100$, $\alpha = 0.05$.

the highest standard deviation, a wrongly assumed homoscedasticity causes overly liberal test decisions. Here, the PI procedure is strongly preferable irrespective of the possible loss in power. If the highest sample size matches (approximately) the highest standard deviation, the HOM procedure may also become conservative in spite of higher degrees of freedom. This has been even more obvious in further simulations which are not shown here. In such cases, one can expect a bad power for HOM. If the data have heterogeneous variances and balanced sample sizes, HOM also causes liberal test decisions. Hence, the HOM procedure is the power-optimal method if the group variances can be assumed equal. If the data are assumed to have heterogeneous variances, there are no power-based arguments against the use of the PI procedure.

3.6 Examples

3.6.1 Birth Weights in a Reprotoxicological Study

The following data of an in-vivo toxicological study are taken from Westfall [1997] and are available from the R package `multcomp` [Hothorn et al., 2008]. The response variable is the average post-birth weight of mice in the entire litter. Pregnant mice were randomized into four groups. The compound, in three different doses (5, 50, 500) and a control (0), was administered during pregnancy. The litters were evaluated for birth weights. The question to be answered is whether or not the specified substance is able to cause a critical weight reduction. Here, the variance depends

Dose	Sample mean	Sample variance	Sample size
0	32.31	7.26	20
5	29.31	25.93	19
50	29.87	14.16	18
500	29.65	29.21	17

Table 3.11: *Summary statistics for the average post-birth weights of the data set of Westfall [1997].*

on the dose effect (see Table 3.11).

The testing problem is hence to show for which doses a critical decrease in weight can be seen in comparison with the control. Let the control group be denoted by $h = 0$ and the doses by $h = 1, 2, 3$. Applying the contrast matrices

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{c}_1' \\ \boldsymbol{c}_2' \\ \boldsymbol{c}_3' \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{d}_1' \\ \boldsymbol{d}_2' \\ \boldsymbol{d}_3' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

leads to the ratios of contrasts

$$\gamma_l = \frac{\mu_l}{\mu_0} \quad (l = 1, 2, 3).$$

The hypotheses to be tested are given by

$$H_{0l} : \gamma_l \geq \theta \quad (l = 1, 2, 3)$$

Dose	p-value	Upper confidence limit
5	0.043 (0.044)	0.998 (0.998)
50	0.042 (0.105)	0.998 (1.017)
500	0.105 (0.082)	1.018 (1.012)

Table 3.12: *p-values and upper confidence limits of the test for the average post-birth weights of the data set of Westfall [1997].*

with $\theta = 1$. Table 3.12 gives (adjusted) p-values and upper limits for the related approximate $(1 - \alpha)100\%$ SCIs for the ratios to the control mean. The values in parantheses are found from the HOM procedure (which assumes homogeneous variances). The two lower doses, 5 and 50, significantly differ from the control. This example clearly shows how misleading the conclusions may be if the variance heterogeneity is not taken into account. The HOM procedure underestimates dose 50, and overestimates dose 500.

A question arising for the toxicologists here is if the highest dose can really be non-toxic while the lower doses are toxic. Assuming an increasing trend in toxicity over the doses would imply a decreasing trend for the measurements. Hence the conclusions, especially those about dose 500, must be questioned. However, a possible objection is that the testing problem has been formulated as a proof of hazard because the objective has been to point out toxicity. However, conclusions about non-significant doses are not allowed in this context. Indeed, dose 500 is not shown to be safe. Maybe a proof of safety would have been more appropriate here. The question to be answered would then be if (and if yes which) doses do not cause a

Dose	p-value	Lower confidence limit
5	0.7638 (0.7048)	0.8199 (0.8238)
50	0.4959 (0.5230)	0.8544 (0.8391)
500	0.6691 (0.6009)	0.8203 (0.8311)

Table 3.13: *p-values and lower confidence limits of the test (proof of safety) for the average post-birth weights of the data set of Westfall [1997].*

critical decrease in weight in comparison with the control. For this purpose, let us assume that a weight reduction by no more than 10% of the control can still be regarded as safe. The new hypotheses to be tested are

$$H_{0l} : \gamma_l \leq \theta \quad (l = 1, 2, 3)$$

with $\theta = 0.9$. The test direction is thus reversed now. Table 3.13 provides (adjusted) p-values and lower limits for the related approximate $(1 - \alpha)100\%$ SCIs. The values in parantheses are again found by applying the HOM procedure. None of the doses can be shown to be safe. Their sample means are not larger than 90% of the control mean. Although the procedures PI and HOM come to the same conclusions, one can see that HOM produces too small p-values and confidence intervals for doses 5 and 500, and a too large p-value and confidence interval for dose 50.

3.6.2 Micronucleus Assay

Adler and Kliesch [1990] have published data from a micronucleus assay on hydroquinone using a negative control, four doses of hydroquinone and the positive control

cyclophosphamide. The goal is to show whether or not the underlying substance is able to induce chromosome damage or to interact with the mitioc spindle apparatus. The number of micronuclei per animal and 2000 scored cells of male mice at 24 h sampling time are given. The variance of the data tends to increase with increasing effects (see Table 3.14). The data are available from the R package `mratios` [Dilba et al., 2008, 2007].

There is an ongoing debate about the definition of clinically relevant non-inferiority margins (e.g., CPM [1999] and Lange and Freitag [2005]). A common non-inferiority trial design involves the experimental drug, a reference drug or active control, and a placebo control. For such three arm "gold standard" trials, Pigeot et al. [2003] have proposed to formulate non-inferiority as a fraction of the trial sensitivity. This results in hypotheses based on the ratio of differences of means. For a specified threshold θ, the alternative hypothesis indicates that the relative efficacy of the experimental drug is more than $\theta * 100\%$ of the efficacy of the reference compound as compared to placebo. For this ratio hypothesis, a t-distributed test statistic has been derived, assuming variance homogeneity. However, it is quite common to observe heteroscedasticity in such three-arm trials.

Non-inferiority tests can also be used as proofs of safety in toxicological experiments, where the difference between a dose group and a vehicle control is considered in relation to the difference between the positive control and the vehicle (see Hauschke et al. [2005]). The mutagenicity data set of Adler and Kliesch [1990] (refer to Table 3.14) has already been evaluated in the sense of a proof of safety by Hauschke et al.

Treatment group	Sample mean	Sample standard deviation	Sample size
Vehicle control	2.57	1.27	7
30 mg/kg	3.80	1.10	5
50 mg/kg	6.20	1.48	5
75 mg/kg	14.0	3.94	5
100 mg/kg	20.0	4.06	5
Positive control	25.0	8.91	4

Table 3.14: *Summary statistics for the number of micronuclei per animal and 2000 scored cells of the mutagenicity data set of Adler and Kliesch [1990].*

[2005] and Hasler et al. [2008]. The concept of the maximal safe dose according to Hothorn and Hauschke [2000] is used, i.e., the identification of the highest dose that is non-inferior to the vehicle control, and all lower doses are non-inferior, too. Since increasing numbers of micronuclei are unsafe, the 95% one-sided upper confidence limits have been used. The authors applied the three-arm trial approach. Confidence intervals for the difference between the dose groups and the vehicle control relative to the difference between a positive control and the vehicle control have been calculated with a safety threshold $\theta = 0.5$. Hauschke et al. [2005] have assumed approximate normal distribution and variance homogeneity, Hasler et al. [2008] have allowed for heterogeneity. All their limits are marginal.

When not considering the maximal safe dose and interest is just in simultaneously comparing the doses in sense of three-arm trials, multiplicity adjustment is necessary. In terms of both three-arm trials and MCTs, let the vehicle control be considered

as the placebo ($h = 0$), the doses as experimental treatments ($h = 1, \ldots, 4$) and the positive control as reference ($h = 5$). Application of the contrast matrices

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{c}_1' \\ \boldsymbol{c}_2' \\ \boldsymbol{c}_3' \\ \boldsymbol{c}_4' \end{pmatrix} = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{d}_1' \\ \boldsymbol{d}_2' \\ \boldsymbol{d}_3' \\ \boldsymbol{d}_4' \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

leads to the ratios of contrasts

$$\gamma_l = \frac{\mu_l - \mu_0}{\mu_5 - \mu_0} \quad (l = 1, \ldots, 4).$$

The hypotheses to be tested are given by

$$H_{0l} : \gamma_l \geq \theta \quad (l = 1, \ldots, 4)$$

Treatment group	p-value	Upper confidence limit
30 mg/kg	0.0225 (0.0002)	0.16 (0.28)
50 mg/kg	0.0472 (0.0032)	0.36 (0.38)
75 mg/kg	0.7275 (0.8786)	1.17 (0.75)
100 mg/kg	0.9906 (1.0000)	2.05 (1.06)

Table 3.15: *p-values and upper confidence limits of the tests for the micronucleus assay data of Adler and Kliesch [1990].*

with $\theta = 0.5$. Table 3.15 shows (adjusted) p-values and upper limits for the related approximate $(1 - \alpha)100\%$ SCIs. The p-values and the limits according to HOM are given in parentheses. The two lower doses, 30 mg/kg and 50 mg/kg, show an acceptable increase; the two higher doses, 75 mg/kg and 100 mg/kg, do not. Although the decisions about the doses are the same for PI and HOM, it is interesting to see that the p-values and the upper limits are markedly different.

Chapter 4

Multiple Contrast Tests for Multiple Endpoints

4.1 Introduction

Experimental trials often do not cover only one single endpoint but many (see the data of Schulte et al. [2002] in Section 4.7). A measurement object may be related to different variables or be observed in the course of time. Multiplicity adjustment must then take the number of endpoints into account, too. Thus, the first strategy is to reduce the number of endpoints to the smallest possible number that is necessary and that still provides the main information about the data. Second, it is useful to divide the endpoints into primary and secondary ones, where the primary endpoints are most important. The guideline on biostatistics according to the ICH E9 Expert

Working Group [1999] recommends the selection of one primary endpoint. However, this is often not sufficient from an investigator's point of view. The secondary endpoints are considered only after the primary objective of the trial has been achieved. A possible objection is that such a classification of endpoints according to their importance can be somewhat arbitrary. Like the first, this strategy also reduces the dimension of the problem, but the question, how to handle multiple primary endpoints, remains. The statistical analysis for these endpoints must control the FWE over all of them. On the other hand, their correlations are important. For example, highly correlated endpoints do not give the same amount of information about the data as uncorrelated ones. Effects may be erroneously ignored when analyzing the endpoints separately.

Neuhäuser [2006] gives a comprehensive review of statistical methods with focus on two-armed trials. A Bonferroni adjustment for the local test on each endpoint is a simple solution of the abovementioned problem. The information about correlations is disregarded in this case. As is known, this technique yields conservative test decisions and intervals, especially for large numbers of endpoints (see Section 2.5). The stepwise procedure of Holm [1979] is more powerful. The procedures of Hochberg [1988] and Hommel [1988] are yet more powerful than Holm's, but the assumption of independence of the p-values must be fulfilled. A drawback is also the fact that no (meaningful) SCIs are available. Gatekeeping procedures avoid multiplicity adjustment by proceeding hierarchically. When multiple hypotheses are to be tested in a prespecified order according to their relevance, multiplicity adjustment is not necessary. As long as the local null hypotheses of the prior endpoints have been

rejected, each local null hypothesis can be tested in a preassigned order at level α [Bauer, 1991]. The procedure stops if a p-value is larger than α. Dmitrienko et al. [2003] have developed parallel gatekeeping strategies that require only one primary effect to be significant for proceeding. Unfortunately, this method also needs an extra (e.g., Bonferroni) internal adjustment, and (meaningful) SCIs are not available. Only if resampling-based tests are used, the authors exploit the endpoint's correlations. The T^2 test of Hotelling [1951] also takes correlations into account, but because of a square sum test statistic it is non-directional and hence not meaningful in many application areas. Furthermore, the test conclusions are merely global ones in the sense that they cannot be attributed to single endpoints. Stabilized alternatives to the T^2 test, using linear scores (see Kropf et al. [1997] for example), suffer from similar drawbacks.

For the sake of completeness, it should be mentioned that the above methods claim a treatment effect if there is a significant difference for at least one endpoint, i.e., they are UITs. If significant differences are necessary for all endpoints to claim a treatment effect, an IUT can be applied (see Section 2.1). The local tests for the endpoints do not need a multiplicity adjustment. They can be performed with level α. Then test decisions are only global; conclusions about single endpoints are not allowed if the global test does not reject. For this reason, the following considerations will not concern IUTs.

MCTs and related SCIs provide test decisions and parameter estimation, respectively, for each comparison. They control the FWE at level α, and take correlations

into account. However, they are limited to comparisons of treatments on a single endpoint so far. This work presents an extension of MCTs and SCIs for multiple endpoints. We focus on ratios of means, because SCIs are then comparable also for the different endpoints, which can be assumed to have different scales.

In Section 4.2, the testing problem is formulated and an approximate distribution for the test statistics is derived. Section 4.3 shows results of α-simulations for several contrasts and correlations of endpoints. SCIs are considered in Section 4.4, the heteroscedastic case in Section 4.6. We give an example in Section 4.7.

4.2 Test Procedure

For $h = 1, \ldots, p$, $i = 1, \ldots, k$ and $j = 1, \ldots, n_h$, let X_{hij} denote the jth observation on the ith endpoint under the hth treatment in a one-way layout, and $\sum_{h=1}^{p}(n_h - 1) \geq k$. Each endpoint is hence measured for all $N = \sum_{h=1}^{p} n_h$ objects. Suppose the random variables X_{hij} to be mutually independent and follow k-variate normal distributions with mean vectors $\boldsymbol{\mu}_h = (\mu_{h1}, \ldots \mu_{hk})'$ and unknown covariance matrices $\boldsymbol{\Sigma}_h = (\sigma_{h,ii'})_{i,i'}$. Let the means per endpoint, $\mu_{1i}, \ldots, \mu_{pi}$, have the same algebraic sign, i.e., $\text{sign}(\mu_{1i}) = \ldots = \text{sign}(\mu_{pi})$ $(i = 1, \ldots, k)$. Presume possibly different variances and covariances for the endpoints but the same covariance matrices for all treatments, i.e., $\boldsymbol{\Sigma}_1 = \ldots = \boldsymbol{\Sigma}_p = \boldsymbol{\Sigma} = (\sigma_{ii'})_{i,i'}$. That means

$$\{X_{hij} : i = 1, \ldots, k\} \sim \perp \mathrm{N}_k(\boldsymbol{\mu}_h, \boldsymbol{\Sigma}) \quad (h = 1, \ldots, p,\ j = 1, \ldots, n_h).$$

Let $\bar{\boldsymbol{X}}_h = (\bar{X}_{h1}, \ldots, \bar{X}_{hk})'$ and $\hat{\boldsymbol{\Sigma}}_h$ be the sample mean vectors and the sample covariance matrices for the treatments, respectively, with

$$\bar{X}_{hi} = \frac{1}{n_h} \sum_{j=1}^{n_h} X_{hij} \quad (h = 1, \ldots, p).$$

The pooled sample covariance matrix $\hat{\boldsymbol{\Sigma}} = (\hat{\sigma}_{ii'})_{i,i'}$ is given by

$$\hat{\boldsymbol{\Sigma}} = \frac{\sum_{h=1}^{p} (n_h - 1) \hat{\boldsymbol{\Sigma}}_h}{\sum_{h=1}^{p} (n_h - 1)}$$

with the estimates $\hat{\sigma}_{ii'}$ $(1 \leq i, i' \leq k)$ for the covariances of the different endpoints. The diagonal elements, which are required for the following test procedure, are hence

$$\hat{\sigma}_{ii} = S_i^2 = \frac{(n_1 - 1) S_{1i}^2 + \cdots + (n_p - 1) S_{pi}^2}{n_1 + \cdots + n_p - p} \quad (i = 1, \ldots, k)$$

with

$$S_{hi}^2 = \frac{1}{n_h - 1} \sum_{j=1}^{n_h} (X_{hij} - \bar{X}_{hi})^2 \quad (h = 1, \ldots, p).$$

From the pooled sample covariance matrix $\hat{\boldsymbol{\Sigma}}$, we then derive the estimation $\hat{\boldsymbol{R}} = (\hat{\rho}_{ii'})_{i,i'}$ of the common correlation matrix of the data $\boldsymbol{R} = (\rho_{ii'})_{i,i'}$. We are interested in the matrix of ratios of contrasts, $\boldsymbol{G} = (\gamma_{li})_{l,i}$, where

$$\begin{aligned} \gamma_{li} &= \frac{\sum_{h=1}^{p} c_{lh} \mu_{hi}}{\sum_{h=1}^{p} d_{lh} \mu_{hi}} \\ &= \frac{\boldsymbol{c}_l' \boldsymbol{\mu}_{,i}}{\boldsymbol{d}_l' \boldsymbol{\mu}_{,i}} \quad (l = 1, \ldots, q,\ i = 1, \ldots, k) \end{aligned}$$

with $\boldsymbol{\mu}_{,i} = (\mu_{1i}, \ldots, \mu_{pi})'$. The vectors $\boldsymbol{c}_l = (c_{l1}, \ldots, c_{lp})'$ and $\boldsymbol{d}_l = (d_{l1}, \ldots, d_{lp})'$ consist of real constants and are the same for all endpoints; they do not depend on the particular value of the index i. Endpoint-specific contrasts are also possible in principle, but we disregard this fact for simplicity. Without loss of generality, the objective is to test the hypotheses

$$H_{0,li} : \gamma_{li} \leq \theta_{li} \quad (l = 1, \ldots, q,\ i = 1, \ldots, k) \tag{4.1}$$

with contrast- and endpoint-specific relative thresholds $\theta_{li} \in (0, \infty)$. Usually, $\theta_{li} = 1$ for all $l = 1, \ldots, q$ and for all $i = 1, \ldots, k$. If the test direction is reversed for some endpoints, the corresponding test statistics have to be multiplied with minus one. We focus here on ratios of means to enable comparison of the results for the different endpoints, which can be assumed to have different scales. Related SCIs for ratios are on the same relative (e.g., per cent) scale for all contrasts and endpoints, while SCIs for differences are not. On the other hand, for the case of $\theta_{li} = 1$ for all $l = 1, \ldots, q$ and for all $i = 1, \ldots, k$, this test coincides with the difference-based one. Testing problem (4.1) is a UIT because the overall null hypothesis of interest can be expressed as an intersection of the local null hypotheses, i.e.,

$$H_0 = \bigcap_{l=1}^{q} H_{0l} \quad \text{and} \quad H_{0l} = \bigcap_{i=1}^{k} H_{0,li}.$$

Thus, Theorem 2.1.3 holds and the procedure is coherent and consonant. We reshape (4.1) and set

$$\begin{aligned} \psi_{li} &= \sum_{h=1}^{p} c_{lh}\mu_{hi} - \theta_{li} \sum_{h=1}^{p} d_{lh}\mu_{hi} = \sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh})\, \mu_{hi} \\ &= (\boldsymbol{c}_l - \theta_{li}\boldsymbol{d}_l)' \, \boldsymbol{\mu}_{,i} \quad (l = 1, \ldots, q,\ i = 1, \ldots, k) \end{aligned}$$

with estimator

$$\begin{aligned} \hat{\psi}_{li} &= \sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh})\, \bar{X}_{hi} \\ &= (\boldsymbol{c}_l - \theta_{li}\boldsymbol{d}_l)' \, \bar{\boldsymbol{X}}_{,i} \quad (l = 1, \ldots, q,\ i = 1, \ldots, k) \end{aligned}$$

and $\bar{\boldsymbol{X}}_{,i} = (\bar{X}_{1i}, \ldots, \bar{X}_{pi})'$. Since

$$\begin{aligned} Var(\hat{\psi}_{li}) &= \sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh})^2\, \sigma_{ii}/n_h \\ &= \sigma_{ii}\, (\boldsymbol{c}_l - \theta_{li}\boldsymbol{d}_l)' \, \boldsymbol{M}\, (\boldsymbol{c}_l - \theta_{li}\boldsymbol{d}_l) \quad (l = 1, \ldots, q,\ i = 1, \ldots, k), \end{aligned}$$

we obtain the test statistics

$$\begin{aligned} T_{li} &= \frac{\sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh}) \bar{X}_{hi}}{S_i \sqrt{\sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh})^2 / n_h}} \\ &= \frac{(\boldsymbol{c}_l - \theta_{li} \boldsymbol{d}_l)' \bar{\boldsymbol{X}}_{,i}}{S_i \sqrt{(\boldsymbol{c}_l - \theta_{li} \boldsymbol{d}_l)' \boldsymbol{M} (\boldsymbol{c}_l - \theta_{li} \boldsymbol{d}_l)}} \quad (l = 1, \ldots, q,\ i = 1, \ldots, k) \end{aligned}$$

where

$$\boldsymbol{M} = \begin{pmatrix} 1/n_1 & & 0 \\ & \ddots & \\ 0 & & 1/n_p \end{pmatrix}.$$

The vectors $\boldsymbol{T}_l = (T_{l1}, \ldots, T_{lk})'$, containing the test statistics for the lth comparison on all endpoints, can be reshaped to

$$\boldsymbol{T}_l = \left(\frac{Y_{l1}}{\sqrt{U_1/\nu}}, \ldots, \frac{Y_{lk}}{\sqrt{U_k/\nu}} \right)' \quad (l = 1, \ldots, q),$$

where under H_{0l}, the vector $(Y_{l1}, \ldots, Y_{lk})'$ follows a k-variate normal distribution with a correlation matrix denoted by $\boldsymbol{R}_{ll}$. The $U_1, \ldots, U_k$ are dependent χ^2 variables with

$$\nu = \sum_{h=1}^{p} (n_h - 1)$$

degrees of freedom. Note that $U_1, \ldots, U_k$ are different random variables but they follow the same distribution. Therefore, under H_{0l}, $\boldsymbol{T}_l$ is approximately k-variate t-distributed with ν degrees of freedom and correlation matrix $\boldsymbol{R}_{ll}$, i.e.,

$$\boldsymbol{T}_l \overset{appr.}{\sim} t_k(\nu, \boldsymbol{R}_{ll}).$$

This is in fact a possible definition of a multivariate t-variable (see, e.g., Tong [1990], page 202f), though not the classical one. Moreover, under H_0, the vector of all test statistics,

$$\boldsymbol{T} = (\boldsymbol{T}_1', \ldots, \boldsymbol{T}_q')' = (T_{11}, \ldots, T_{li}, \ldots, T_{qk})',$$

follows (approximately) a qk-variate t-distribution with ν degrees of freedom and a correlation matrix, denoted by $\tilde{\boldsymbol{R}}$, i.e.,

$$\boldsymbol{T} \overset{appr.}{\sim} t_{qk}(\nu, \tilde{\boldsymbol{R}}).$$

The correlation matrix $\tilde{\boldsymbol{R}}$ is given by

$$\tilde{\boldsymbol{R}} = (\boldsymbol{R}_{ll'})_{l,l'} = \begin{pmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} & \dots & \boldsymbol{R}_{1q} \\ \boldsymbol{R}_{12} & \boldsymbol{R}_{22} & \dots & \boldsymbol{R}_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{R}_{1q} & \boldsymbol{R}_{2q} & \dots & \boldsymbol{R}_{qq} \end{pmatrix}.$$

The submatrices $\boldsymbol{R}_{ll'} = (\rho_{ll',ii'})_{i,i'}$ describe the correlations between the contrasts l and l' for all endpoints. In order to calculate their elements, let us evaluate the correlation between $\sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh}) \bar{X}_{hi}$ and $\sum_{h=1}^{p} (c_{l'h} - \theta_{l'i'} d_{l'h}) \bar{X}_{hi'}$:

$$\begin{aligned}
& Cov\left(\sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh}) \bar{X}_{hi}, \sum_{h=1}^{p} (c_{l'h} - \theta_{l'i'} d_{l'h}) \bar{X}_{hi'}\right) \\
= \; & E\left[\left(\sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh}) \bar{X}_{hi} - E\left(\sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh}) \bar{X}_{hi}\right)\right)\right. \\
& \left.\left(\sum_{h=1}^{p} (c_{l'h} - \theta_{l'i'} d_{l'h}) \bar{X}_{hi'} - E\left(\sum_{h=1}^{p} (c_{l'h} - \theta_{l'i'} d_{l'h}) \bar{X}_{hi'}\right)\right)\right] \\
= \; & E\left[\left(\sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh}) \left(\bar{X}_{hi} - E\bar{X}_{hi}\right)\right)\left(\sum_{h=1}^{p} (c_{l'h} - \theta_{l'i'} d_{l'h}) \left(\bar{X}_{hi'} - E\bar{X}_{hi'}\right)\right)\right] \\
= \; & E\left[\sum_{h=1}^{p}\sum_{h'=1}^{p} (c_{lh} - \theta_{li} d_{lh}) (c_{l'h'} - \theta_{l'i'} d_{l'h'}) \left(\bar{X}_{hi} - E\bar{X}_{hi}\right)\left(\bar{X}_{h'i'} - E\bar{X}_{h'i'}\right)\right] \\
= \; & \sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh}) (c_{l'h} - \theta_{l'i'} d_{l'h}) E\left(\left(\bar{X}_{hi} - E\bar{X}_{hi}\right)\left(\bar{X}_{hi'} - E\bar{X}_{hi'}\right)\right) \\
& + \sum_{h=1}^{p} \sum_{h'=1, h' \neq h}^{p} (c_{lh} - \theta_{li} d_{lh}) (c_{l'h'} - \theta_{l'i'} d_{l'h'}) E\left(\left(\bar{X}_{hi} - E\bar{X}_{hi}\right)\left(\bar{X}_{h'i'} - E\bar{X}_{h'i'}\right)\right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh}) (c_{l'h} - \theta_{l'i'} d_{l'h}) \, Cov\left(\bar{X}_{hi}, \bar{X}_{hi'}\right) \\
&\quad + \sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh}) \, E\left(\bar{X}_{hi} - E\bar{X}_{hi}\right) \sum_{h'=1, h' \neq h}^{p} (c_{l'h'} - \theta_{l'i'} d_{l'h'}) \, E\left(\bar{X}_{h'i'} - E\bar{X}_{h'i'}\right) \\
&= \sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh}) (c_{lh} - \theta_{li'} d_{lh}) \, Cov\left(\bar{X}_{hi}, \bar{X}_{hi'}\right) \\
&\quad + \sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh}) \left(E\bar{X}_{hi} - E\bar{X}_{hi}\right) \sum_{h'=1, h' \neq h}^{p} (c_{l'h'} - \theta_{l'i'} d_{l'h'}) \left(E\bar{X}_{h'i'} - E\bar{X}_{h'i'}\right) \\
&= \sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh}) (c_{l'h} - \theta_{l'i'} d_{l'h}) \, Cov\left(\bar{X}_{hi}, \bar{X}_{hi'}\right).
\end{aligned}
$$

Obviously,

$$
\begin{aligned}
& Cov(\bar{X}_{hi}, \bar{X}_{hi'}) \\
&= E\left((\bar{X}_{hi} - \mu_{hi})(\bar{X}_{hi'} - \mu_{hi'})\right) \\
&= E\left(\left(\frac{1}{n_h}\sum_{j=1}^{n_h} X_{hij}\right) - \left(\frac{1}{n_h}\sum_{j=1}^{n_h} \mu_{hi}\right)\right)\left(\left(\frac{1}{n_h}\sum_{j'=1}^{n_h} X_{hi'j'}\right) - \left(\frac{1}{n_h}\sum_{j'=1}^{n_h} \mu_{hi'}\right)\right) \\
&= E\left(\frac{1}{n_h}\sum_{j=1}^{n_h}(X_{hij} - \mu_{hi})\frac{1}{n_h}\sum_{j'=1}^{n_h}(X_{hi'j'} - \mu_{hi'})\right) \\
&= \frac{1}{n_h^2} E\left(\sum_{j=1}^{n_h}\sum_{j'=1}^{n_h}(X_{hij} - \mu_{hi})(X_{hi'j'} - \mu_{hi'})\right) \\
&= \frac{1}{n_h^2}\left(\sum_{j=1}^{n_h}\sum_{j'=1}^{n_h} Cov(X_{hij}, X_{hi'j'})\right).
\end{aligned}
$$

Because all measurements $j \neq j'$ are independent, their covariances vanish, so that we may write

$$
Cov(\bar{X}_{hi}, \bar{X}_{hi'}) = \frac{1}{n_h} Cov(X_{hi}, X_{hi'}).
$$

Hence, it follows that

$$
\begin{aligned}
\rho_{ll',ii'} &= Corr\left(\sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh}) \bar{X}_{hi}, \sum_{h=1}^{p} (c_{l'h} - \theta_{l'i'} d_{l'h}) \bar{X}_{hi'}\right) \\
&= \frac{\sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh}) (c_{l'h} - \theta_{l'i'} d_{l'h}) \frac{1}{n_h} Cov(X_{hi}, X_{hi'})}{\sqrt{Var\left(\sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh}) \bar{X}_{hi}\right)} \sqrt{Var\left(\sum_{h=1}^{p} (c_{l'h} - \theta_{l'i'} d_{l'h}) \bar{X}_{hi'}\right)}}
\end{aligned}
$$

$$= \frac{\sum_{h=1}^{p}\left(c_{lh}-\theta_{li}d_{lh}\right)\left(c_{l'h}-\theta_{l'i'}d_{l'h}\right)\frac{1}{n_h}\sigma_{ii'}}{\sqrt{\sum_{h=1}^{p}\left(c_{lh}-\theta_{li}d_{lh}\right)^2 Var\bar{X}_{hi}}\sqrt{\sum_{h=1}^{p}\left(c_{l'h}-\theta_{l'i'}d_{l'h}\right)^2 Var\bar{X}_{hi'}}}$$

$$= \frac{\sigma_{ii'}\sum_{h=1}^{p}\left(c_{lh}-\theta_{li}d_{lh}\right)\left(c_{l'h}-\theta_{l'i'}d_{l'h}\right)\frac{1}{n_h}}{\sqrt{\sum_{h=1}^{p}\left(c_{lh}-\theta_{li}d_{lh}\right)^2 \frac{1}{n_h}\sigma_{ii}}\sqrt{\sum_{h=1}^{p}\left(c_{l'h}-\theta_{l'i'}d_{l'h}\right)^2 \frac{1}{n_h}\sigma_{i'i'}}}$$

and finally

$$\rho_{ll',ii'} = \rho_{ii'}\frac{\sum_{h=1}^{p}\left(c_{lh}-\theta_{li}d_{lh}\right)\left(c_{l'h}-\theta_{l'i'}d_{l'h}\right)\frac{1}{n_h}}{\sqrt{\sum_{h=1}^{p}\left(c_{lh}-\theta_{li}d_{lh}\right)^2 \frac{1}{n_h}}\sqrt{\sum_{h=1}^{p}\left(c_{l'h}-\theta_{l'i'}d_{l'h}\right)^2 \frac{1}{n_h}}} \tag{4.2}$$

$$= \rho_{ii'}\frac{\left(\boldsymbol{c}_l-\theta_{li}\boldsymbol{d}_l\right)'\boldsymbol{M}\left(\boldsymbol{c}_{l'}-\theta_{l'i'}\boldsymbol{d}_{l'}\right)}{\sqrt{\left(\boldsymbol{c}_l-\theta_{li}\boldsymbol{d}_l\right)'\boldsymbol{M}\left(\boldsymbol{c}_l-\theta_{li}\boldsymbol{d}_l\right)}\sqrt{\left(\boldsymbol{c}_{l'}-\theta_{l'i'}\boldsymbol{d}_{l'}\right)'\boldsymbol{M}\left(\boldsymbol{c}_{l'}-\theta_{l'i'}\boldsymbol{d}_{l'}\right)}}$$

$$(1 \le l, l' \le q,\ 1 \le i, i' \le k),$$

where the $\rho_{ii'}$ are the elements of the correlation matrix $\boldsymbol{R} = (\rho_{ii'})_{i,i'}$ of the data. It is obvious that for $i = i'$, we recover the correlations of an MCT for ratios of means, see, e.g., Dilba et al. [2006]. Hence, the case of only one endpoint ($k = 1$) and several treatments may be incorporated into the present theory rather easily. Furthermore, focusing on one fixed contrast ($l = l'$) and equal thresholds for all endpoints ($\theta_{li} = \theta_l\ \forall\, i = 1, \ldots, k$), the structure of the correlation matrix simplifies according to $\rho_{ll',ii'} = \rho_{ii'}$ and $\boldsymbol{R}_{ll} = \boldsymbol{R}$. Note that neither the matrix $\tilde{\boldsymbol{R}}$ nor the matrix $\boldsymbol{R}_{ll'}$ has a product correlation structure, i.e., the elements do not factorize. Because the common correlation matrix of the data $\boldsymbol{R}$ is not known and must be estimated, we conclude that, under H_0,

$$\boldsymbol{T} \overset{appr.}{\sim} t_{qk}(\nu, \hat{\tilde{\boldsymbol{R}}}),$$

where $\hat{\tilde{\boldsymbol{R}}}$ is the estimation of $\tilde{\boldsymbol{R}}$.

Example 4.2.1. Let there be observations for $p = 3$ groups and $k = 3$ endpoints with equal sample sizes, being mutually independent and following k-variate normal

distributions with homogeneous covariance matrices. Let the common correlation matrix of the data be given by

$$\boldsymbol{R} = \begin{pmatrix} 1 & -0.2 & 0.5 \\ -0.2 & 1 & 0.1 \\ 0.5 & 0.1 & 1 \end{pmatrix}.$$

The Dunnett MCT is applied (where the first group is regarded as the negative control). The thresholds are all equal, i.e., $\theta_{li} = 1$ for all $l = 1, \ldots, q$ and for all $i = 1, \ldots, k$. The resulting correlation matrix $\tilde{\boldsymbol{R}}$ is given by

$$\tilde{\boldsymbol{R}} = \begin{pmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{R}_{12} & \boldsymbol{R}_{22} \end{pmatrix} = \begin{pmatrix} 1 & -0.2 & 0.5 & 0.5 & -0.1 & 0.25 \\ -0.2 & 1 & 0.1 & -0.1 & 0.5 & 0.05 \\ 0.5 & 0.1 & 1 & 0.25 & 0.05 & 0.5 \\ 0.5 & -0.1 & 0.25 & 1 & -0.2 & 0.5 \\ -0.1 & 0.5 & 0.05 & -0.2 & 1 & 0.1 \\ 0.25 & 0.05 & 0.5 & 0.5 & 0.1 & 1 \end{pmatrix}.$$

The (3×3)-submatrices on the main diagonal are equal to the correlation matrix of the data, i.e., $\boldsymbol{R}_{11} = \boldsymbol{R}_{22} = \boldsymbol{R}$.

Example 4.2.2. Consider the situation of the Example 4.2.1, where we now apply the Williams MCT. The resulting correlation matrix $\tilde{\boldsymbol{R}}$ is then given by

$$\tilde{\boldsymbol{R}} = \begin{pmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{R}_{12} & \boldsymbol{R}_{22} \end{pmatrix} = \begin{pmatrix} 1 & -0.2 & 0.5 & 0.87 & -0.17 & 0.43 \\ -0.2 & 1 & 0.1 & -0.17 & 0.87 & 0.09 \\ 0.5 & 0.1 & 1 & 0.43 & 0.09 & 0.87 \\ 0.87 & -0.17 & 0.43 & 1 & -0.2 & 0.5 \\ -0.17 & 0.87 & 0.09 & -0.2 & 1 & 0.1 \\ 0.43 & 0.09 & 0.87 & 0.5 & 0.1 & 1 \end{pmatrix}.$$

As in Example 4.2.1, the (3×3)-submatrices on the main diagonal are equal to the correlation matrix of the data. The off-diagonal submatrices $\boldsymbol{R}_{12}$ differ from those of Example 4.2.1, of course.

The decision rule for testing problem (4.1) is to reject $H_{0,li}$ for each ratio of contrasts γ_{li} with

$$T_{li} > t^{l}_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}}),$$

where $t^{l}_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}})$ is a lower $(1-\alpha)$-quantile of a related qk-variate t-distribution. If two-sided testing is of interest, the absolute values for T_{li} and quantiles $t^{ts}_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}})$ have to be taken. For the computation of these quantiles, one may resort to the numerical integration routines of Genz and Bretz [1999, 2002] (see also Bretz et al. [2001]) mentioned earlier, which are not restricted to special correlation structures. The related adjusted p-values per comparison and endpoint can also be obtained, of course.

4.3 α-simulations

As in Chapter 3, the method described in Section 4.2 is approximate. Hence, its quality must be validated by simulations. Three treatments have been compared in a first simulation study, five in a second one. The first treatment is regarded as the (negative) control. Dunnett and Williams contrasts, related to the hypotheses (4.1), have been considered. Each study had different numbers of endpoints with related expected values for the control group, i.e.,

- 2 endpoints, $\boldsymbol{\mu}_1 = (10, 100)$,
- 4 endpoints, $\boldsymbol{\mu}_1 = (0.1, 1, 10, 100)$,
- 8 endpoints, $\boldsymbol{\mu}_1 = (0.05, 0.1, 0.5, 1, 5, 10, 50, 100)$.

For the Dunnett contrast, the expected values of the non-control groups are $\boldsymbol{\mu}_h = (0.8\mu_{1,1}, \ldots, 0.8\mu_{1,k/2}, 1.25\mu_{1,k/2+1}, \ldots, 1.25\mu_{1,k})'$ $(h = 2, \ldots, p)$, and $\boldsymbol{\mu}_h = \boldsymbol{\mu}_1$ $(h = 2, \ldots, p)$ for the Williams contrast. The endpoints have equicorrelations ρ^{min} (see (2.3)), 0, 0.5, 1. The standard deviations are $0.25\boldsymbol{\mu}_1$ for all treatments. The sample size is 20 for each endpoint of each treatment. The FWE has been simulated at a nominal level of 0.05. The simulation results have been obtained from 10000 simulation runs each and with the same starting seed (seed 10000) using a program code in the statistical software R [2008], package `mvtnorm` [Genz et al., 2008, Hothorn et al., 2001].

Table 4.1 (4.2) shows results of the first (second) study with three (five) treatments. The values in parentheses are according to a Bonferroni adjustment, which is known to produce conservative test decisions, especially for high correlations (see Section 2.5). The new procedure maintains the α-level exactly (ranges from 0.045 to 0.054) while the Bonferroni-adjusted version becomes more and more conservative for increasing correlations and increasing number of endpoints.

In order to appreciate the strong control of the FWE, we have again treated the situation of the Dunnett contrast. Let all non-control treatments significantly differ from the control for the first endpoint, i.e., $\mu_{h,1} = 10 * \mu_{1,1}$ $(h = 2, \ldots, p)$. Only the FWE of the remaining comparisons has been considered. Tables 4.3 and 4.4 show the α-level of the new procedure ignoring the first endpoint. Depending on the number of these comparisons, the FWE is smaller then the α-level. The Bonferroni-adjusted version is conservative in the same manner as above.

4.4 Simultaneous Confidence Intervals

Let $\boldsymbol{\xi} = (\xi_{11}, \ldots, \xi_{qk})'$ be a point in the parameter space of $\boldsymbol{\gamma} = (\gamma_{11}, \ldots, \gamma_{qk})'$. Assuming that increasing values of the data, X_{hij}, represent a better effect of the treatments, the $(1-\alpha)100\%$ confidence set for the statistical problem (4.1) is given by

$$\begin{aligned} C\left((x,y)\right) &= \left\{\boldsymbol{\xi} : T_{li}(\xi_{li}) \leq t^{l}_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}}), \quad l = 1, \ldots, q,\ i = 1, \ldots, k\right\} \\ &= \left\{\boldsymbol{\xi} : A_{li}\xi_{li}^2 + B_{li}\xi_{li} + C_{li} \leq 0, \quad l = 1, \ldots, q,\ i = 1, \ldots, k\right\} \end{aligned}$$

MCT	Endpoints	Correlations			
		ρ^{min}	0	0.5	1
Dunnett	2	0.049	0.051	0.053	0.046
		(0.046)	(0.046)	(0.045)	(0.026)
	4	0.047	0.050	0.048	0.051
		(0.044)	(0.046)	(0.036)	(0.015)
	8	0.047	0.046	0.052	0.051
		(0.043)	(0.042)	(0.036)	(0.009)
Williams	2	0.049	0.052	0.049	0.052
		(0.036)	(0.038)	(0.033)	(0.021)
	4	0.050	0.048	0.051	0.045
		(0.037)	(0.035)	(0.030)	(0.009
	8	0.048	0.048	0.054	0.049
		(0.037)	(0.035)	(0.029)	(0.006)

Table 4.1: *FWE of one-sided MCTs for $p = 3$ treatments, several contrasts, numbers of endpoints, and equicorrelations; $\alpha = 0.05$; values in parentheses according to Bonferroni adjustment.*

MCT	Endpoints	Correlations			
		ρ^{min}	0	0.5	1
Dunnett	2	0.051	0.052	0.052	0.052
		(0.043)	(0.045)	(0.040)	(0.028)
	4	0.052	0.053	0.052	0.049
		(0.045)	(0.046)	(0.038)	(0.013)
	8	0.049	0.051	0.050	0.050
		(0.043)	(0.044)	(0.036)	(0.008)
Williams	2	0.050	0.050	0.053	0.052
		(0.025)	(0.024)	(0.023)	(0.014)
	4	0.050	0.054	0.048	0.051
		(0.028)	(0.029)	(0.024)	(0.007)
	8	0.049	0.052	0.048	0.049
		(0.027)	(0.027)	(0.021)	(0.004)

Table 4.2: *FWE of one-sided MCTs for $p = 5$ treatments, several contrasts, numbers of endpoints, and equicorrelations; $\alpha = 0.05$; values in parentheses according to Bonferroni adjustment.*

Endpoints	Correlations			
	ρ^{min}	0	0.5	1
2	0.023	0.027	0.027	0.037
	(0.022)	(0.025)	(0.023)	(0.019)
4	0.035	0.037	0.044	0.045
	(0.033)	(0.033)	(0.034)	(0.013)
8	0.041	0.041	0.048	0.051
	(0.038)	(0.037)	(0.033)	(0.009)

Table 4.3: *Local FWE of one-sided MCTs for $p = 3$ treatments, the Dunnett contrast, several numbers of endpoints, and equicorrelations; $\alpha = 0.05$; values in parantheses according to Bonferroni adjustment.*

Endpoints	Correlations			
	ρ^{min}	0	0.5	1
2	0.024	0.022	0.025	0.039
	(0.021)	(0.018)	(0.020)	(0.018)
4	0.037	0.038	0.040	0.049
	(0.032)	(0.033)	(0.030)	(0.013)
8	0.043	0.045	0.044	0.050
	(0.037)	(0.039)	(0.032)	(0.008)

Table 4.4: *Local FWE of one-sided MCTs for $p = 5$ treatments, the Dunnett contrast, several numbers of endpoints, and equicorrelations; $\alpha = 0.05$; values in parantheses according to Bonferroni adjustment.*

where

$$\begin{aligned}
A_{li} &= \left(\sum_{h=1}^{p} d_{lh}\bar{X}_{hi}\right)^2 - \left(t^l_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}})\right)^2 S_i^2 \sum_{h=1}^{p} d_{lh}^2/n_h \\
&= \left(\boldsymbol{d}_l'\bar{\boldsymbol{X}}_{,i}\right)^2 - \left(t^l_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}})\right)^2 S_i^2 \boldsymbol{d}_l'\boldsymbol{M}\boldsymbol{d}_l, \\
B_{li} &= -2\left(\left(\sum_{h=1}^{p} c_{lh}\bar{X}_{hi}\right)\left(\sum_{h=1}^{p} d_{lh}\bar{X}_{hi}\right) - \left(t^l_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}})\right)^2 S_i^2 \sum_{h=1}^{p} c_{lh}d_{lh}/n_h\right) \\
&= -2\left(\left(\boldsymbol{c}_l'\bar{\boldsymbol{X}}_{,i}\right)\left(\boldsymbol{d}_l'\bar{\boldsymbol{X}}_{,i}\right) - \left(t^l_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}})\right)^2 S_i^2 \boldsymbol{c}_l'\boldsymbol{M}\boldsymbol{d}_l\right), \\
C_{li} &= \left(\sum_{h=1}^{p} c_{lh}\bar{X}_{hi}\right)^2 - \left(t^l_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}})\right)^2 S_i^2 \sum_{h=1}^{p} c_{lh}^2/n_h \\
&= \left(\boldsymbol{c}_l'\bar{\boldsymbol{X}}_{,i}\right)^2 - \left(t^l_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}})\right)^2 S_i^2 \boldsymbol{c}_l'\boldsymbol{M}\boldsymbol{c}_l. \qquad (4.3)
\end{aligned}$$

As in Section 3.4.3, this approach is based on Fieller's Theorem [Fieller, 1954]. As is known, the correlation matrix $\tilde{\boldsymbol{R}}$ depends here on the unknown ratios γ_{li}, $\hat{\rho}_{ll',ii'} = \hat{\rho}_{ll',ii'}(\gamma_{li}, \gamma_{l'i'})$. Application of the plug-in approach of Dilba et al. [2006] corresponds to the use of

$$\begin{aligned}
\hat{\gamma}_{li} &= \frac{\sum_{h=1}^{p} c_{lh}\bar{X}_{hi}}{\sum_{h=1}^{p} d_{lh}\bar{X}_{hi}} \\
&= \frac{\boldsymbol{c}_l'\bar{\boldsymbol{X}}_{,i}}{\boldsymbol{d}_l'\bar{\boldsymbol{X}}_{,i}} \quad (l = 1, \ldots, q, \quad i = 1, \ldots, k)
\end{aligned}$$

in Equation (4.2) instead of θ_{li} (similarly for index $l'i'$). For simplicity, we do not introduce a new symbol for the resulting estimated correlation matrix. The lower limits of the approximate $(1-\alpha)100\%$ SCIs for $(\gamma_{11}, \ldots, \gamma_{qk})'$ are hence given by

$$\hat{\gamma}_{li}^{lower} = \frac{-B_{li} - \sqrt{B_{li}^2 - 4A_{li}C_{li}}}{2A_{li}} \quad (l = 1, \ldots, q, \quad i = 1, \ldots, k).$$

If $A_{li} > 0$, then the solution is finite (see, e.g., Buonaccorsi and Iyer [1984] for the case of only one endpoint). The statistical problem (4.1) can be decided as follows: For a specified level α, we reject $H_{0,li}$ for each contrast γ_{li} with

$$\hat{\gamma}_{li}^{lower} > \theta_{li}.$$

For the two-sided case, we obtain

$$\begin{aligned} C((x,y)) &= \left\{\boldsymbol{\xi} : |T_{li}(\xi_{li})| \leq t^{ts}_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}}), \quad l = 1,\ldots,q,\ i = 1,\ldots,k\right\} \\ &= \left\{\boldsymbol{\xi} : A_{li}\xi_{li}^2 + B_{li}\xi_{li} + C_{li} \leq 0, \quad l = 1,\ldots,q,\ i = 1,\ldots,k\right\}, \end{aligned}$$

where the A_{li}, B_{li} and C_{li} are defined as in (4.3) but with quantiles $t^{ts}_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}})$ instead of $t^{l}_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}})$. The confidence limits are given by

$$\begin{aligned} \hat{\gamma}_{li}^{lower} &= \frac{-B_{li} - \sqrt{B_{li}^2 - 4A_{li}C_{li}}}{2A_{li}} \quad (l = 1,\ldots,q, \quad i = 1,\ldots,k), \\ \hat{\gamma}_{li}^{upper} &= \frac{-B_{li} + \sqrt{B_{li}^2 - 4A_{li}C_{li}}}{2A_{li}} \quad (l = 1,\ldots,q, \quad i = 1,\ldots,k). \end{aligned}$$

For a specified level α, we reject $H_{0,li}$ for each contrast γ_{li} with

$$\hat{\gamma}_{li}^{lower} > \theta_{li} \quad \text{or} \quad \hat{\gamma}_{\text{li}}^{\text{upper}} < \theta_{\text{li}}.$$

We have performed further simulation studies with the same background as in 4.3 to describe the behavior of SCIs related to the new procedure. The SCP has been simulated for the Dunnett contrast with a nominal level of 0.95. Table 4.5 (4.6) shows the results of the first (second) study for ratio-based Dunnett SCIs with three (five) treatments for different numbers of endpoints and depending on the ratios $\gamma = \gamma_{11} = \ldots = \gamma_{2k}$ ($\gamma = \gamma_{11} = \ldots = \gamma_{4k}$). In principle, the expected value 0.95 is again attained for all the settings irrespective of the number of endpoints, treatments, and the correlations (ranges from 0.943 to 0.954). This reflects the results of the α-simulations.

Endpoints	γ	Correlations			
		ρ^{min}	0	0.5	1
2	0.5	0.946	0.946	0.949	0.950
	1.0	0.948	0.949	0.949	0.950
	1.5	0.950	0.947	0.948	0.948
	2.0	0.949	0.954	0.950	0.946
4	0.5	0.949	0.952	0.949	0.949
	1.0	0.950	0.947	0.950	0.949
	1.5	0.948	0.950	0.948	0.952
	2.0	0.948	0.950	0.947	0.949
8	0.5	0.950	0.950	0.950	0.949
	1.0	0.950	0.949	0.947	0.950
	1.5	0.954	0.949	0.948	0.946
	2.0	0.949	0.947	0.946	0.951

Table 4.5: *SCP of one-sided (upper) SCIs for $p = 3$ treatments, the Dunnett contrast, several numbers of endpoints, ratios $\gamma_{li} = \gamma$ (for all $l = 1, 2$ and $i = 1, \ldots, k$), and equicorrelations; $\alpha = 0.05$.*

Endpoints	γ	Correlations			
		ρ^{min}	0	0.5	1
2	0.5	0.949	0.948	0.947	0.951
	1.0	0.952	0.948	0.945	0.948
	1.5	0.954	0.947	0.947	0.950
	2.0	0.953	0.948	0.946	0.949
4	0.5	0.951	0.947	0.947	0.951
	1.0	0.952	0.949	0.951	0.945
	1.5	0.948	0.948	0.946	0.949
	2.0	0.949	0.950	0.943	0.948
8	0.5	0.948	0.949	0.948	0.950
	1.0	0.952	0.951	0.953	0.951
	1.5	0.946	0.951	0.945	0.950
	2.0	0.948	0.950	0.948	0.950

Table 4.6: *SCP of one-sided (upper) SCIs for* $p = 5$ *treatments, the Dunnett contrast, several numbers of endpoints, ratios* $\gamma_{li} = \gamma$ *(for all* $l = 1, \ldots, 4$ *and* $i = 1, \ldots, k$*), and equicorrelations;* $\alpha = 0.05$.

4.5 Power Considerations

As in Section 3.5, the testing problem (4.1) is here again simplified to the case of equal thresholds, $\theta_{li} = \theta$ for all $l = 1, \ldots, q$ and $i = 1, \ldots, k$. Let higher response values indicate better treatment effects and τ^* denote the greatest irrelevant ratio to the control mean which is to be detected. Define the set of indices $I(\tau^*) = \{(l, i) : \tau_{li} > \tau^*\}$. All ratios of contrasts with τ_{li} values greater than τ^* are relevant. In the same manner as in Section 3.5, an (approximate) expression for the complete (or all-pairs) power of the statistical problem (4.1) is given by

$$P\left\{T_{li} > t^l_{qk,1-\alpha}(\nu, \tilde{\boldsymbol{R}}) \middle| \psi_{li}, \boldsymbol{\Sigma} \quad \forall (l,i) \in I(\theta^*)\right\}. \tag{4.4}$$

An (approximate) expression for the minimal (or any-pair) power of the statistical problem (4.1) is given by

$$P\left\{T_{li} > t^l_{qk,1-\alpha}(\nu, \tilde{\boldsymbol{R}}) \middle| \psi_{li}, \boldsymbol{\Sigma} \quad \text{for at least one } (l,i) \in I(\theta^*)\right\}. \tag{4.5}$$

The probability to reject for any contrast is defined as the *global power*. If one is interested only in the global test decision for statistical problem (4.1), then this definition is appropriate. An (approximate) expression for the global power of the statistical problem (4.1) is given by

$$P\left\{T_{li} > t^l_{qk,1-\alpha}(\nu, \tilde{\boldsymbol{R}}) \middle| \psi_{li}, \boldsymbol{\Sigma} \quad \text{for at least one } l = 1, \ldots, q \text{ and } i = 1, \ldots, k\right\}. \tag{4.6}$$

Because the data's correlations are estimated, the quantiles $t^l_{qk,1-\alpha}(\nu, \tilde{\boldsymbol{R}})$ in fact are random variables because they depend on the sample values. Therefore, the probabilities (4.4), (4.5) and (4.6) are only approximate ones. The power function

(4.6) can be calculated from a non-central qk-variate t-distribution with ν degrees of freedom and non-centrality parameter $\boldsymbol{\kappa} = (\kappa_{11}, \ldots, \kappa_{li}, \ldots, \kappa_{qk})'$, where

$$\begin{aligned} \kappa_{li} &= \frac{\sum_{h=1}^{p} (c_{lh} - \theta d_{lh})\, \mu_{hi}}{\sqrt{\sigma_{ii} \sum_{h=1}^{p} (c_{lh} - \theta d_{lh})^2 / n_h}} \\ &= \frac{(\boldsymbol{c_l} - \theta \boldsymbol{d_l})' \, \boldsymbol{\mu}_{,i}}{\sqrt{\sigma_{ii} \, (\boldsymbol{c_l} - \theta \boldsymbol{d_l})' \, \boldsymbol{M} \, (\boldsymbol{c_l} - \theta \boldsymbol{d_l})}}. \end{aligned}$$

Figure 4.1 (4.2) illustrates Equation (4.6) for three (five) treatments and the Dunnett contrast depending on the ratio $\gamma_{q1} = \mu_{p1}/\mu_{11}$ (where $q = p - 1$). The remaining ratios γ_{li} are fixed and equal. The relative thresholds against which the test is performed are $\theta_{li} = 1$ for all $l = 1, \ldots, q$ and all $i = 1, \ldots, k$. Several equicorrelations (rows) for two, four and eight endpoints (columns) are considered. The total sample size is 60 (100). Three allocations are shown each. The solid line represents the well-known optimal allocation for the Dunnett contrast, i.e., $n_1 = \sqrt{p-1}\, n_h$ ($h = 2, \ldots, p$). Hence, the sample size for the control group is $n_1 = 24, 12, 6$ $(32, 16, 8)$, and the sample sizes for the non-control groups are balanced. Although the correlations of the endpoints are taken into account, their exact influence it is not clear from Figures 4.1 and 4.2. Therefore, this problem is presented by Figure 4.3 (4.4). Again, Equation (4.6) is illustrates for three (five) treatments with a similar background, but now depending on the correlations of the endpoints. The ratio γ_{q1} is set here to 1.25. One-sided and two-sided tests (rows) for two, four and eight endpoints (columns) are considered. The power indeed depends on the correlations. The minimum is achieved for vanishing correlation and increases for increasing absolute correlation values.

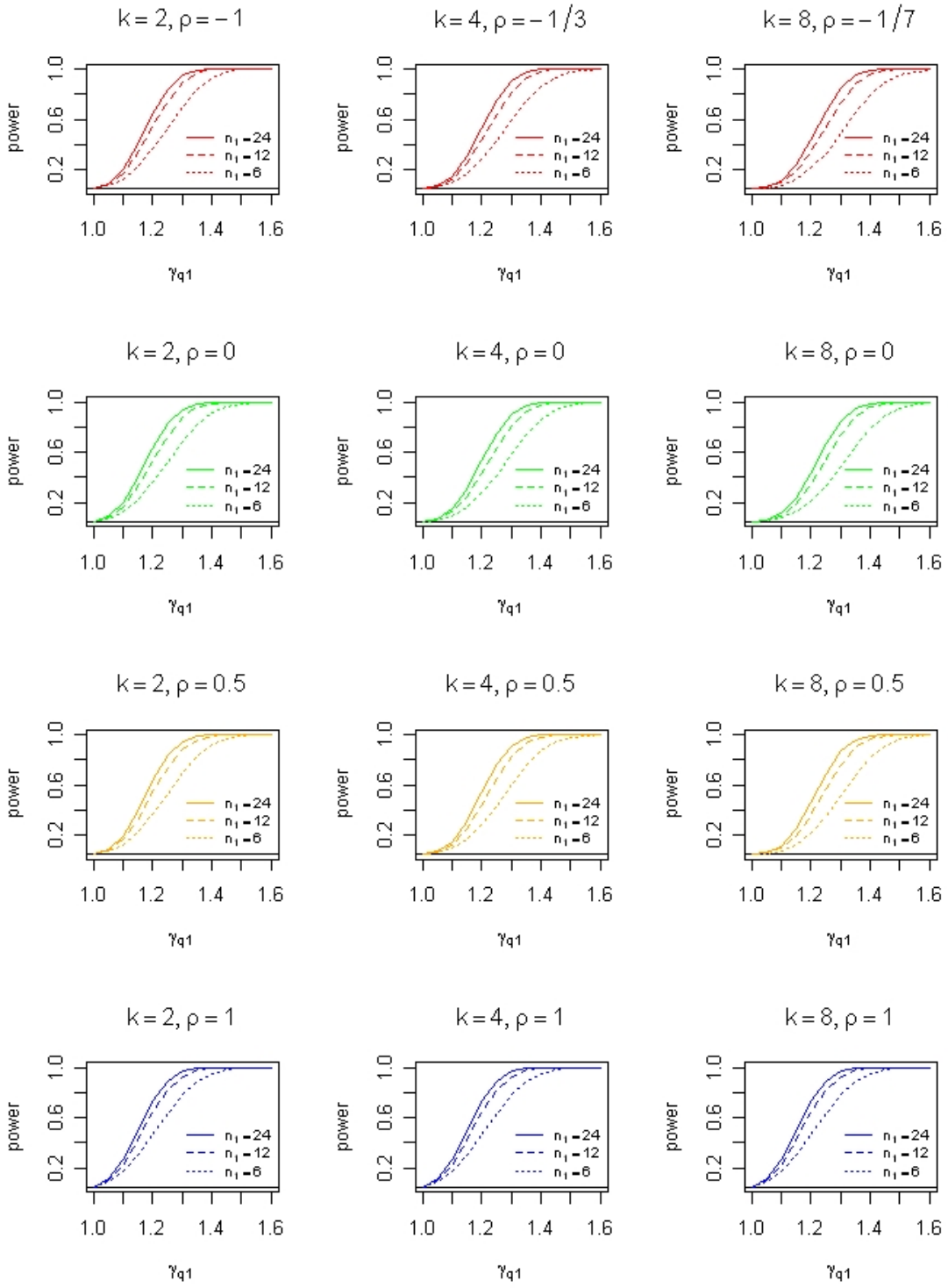


Figure 4.1: *Global power function of one-sided MCTs for* $p = 3$ *treatments, the Dunnett contrast, several numbers of endpoints, ratios* γ_{q1}*, and equicorrelations;* $\alpha = 0.05$.

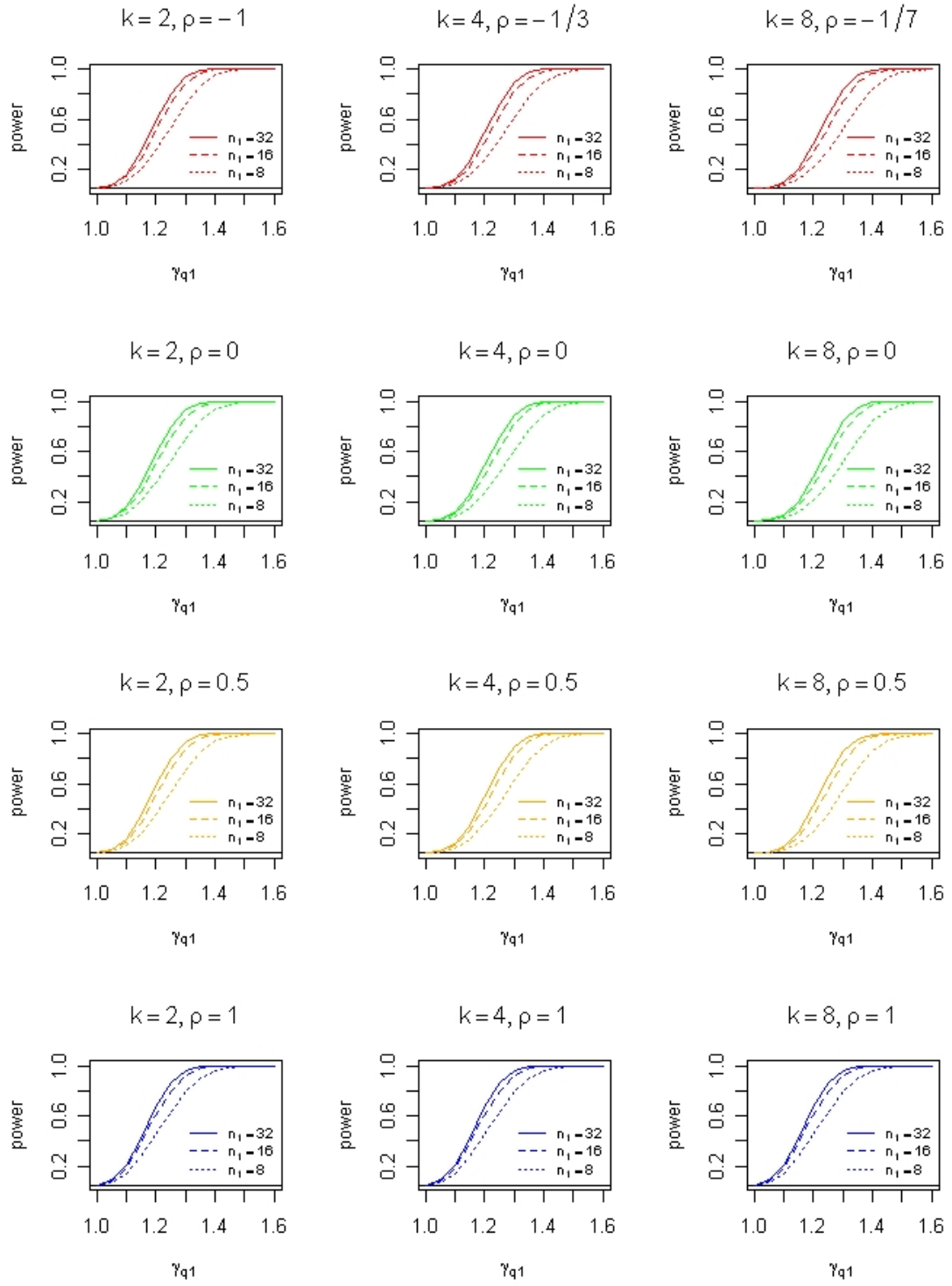


Figure 4.2: *Global power function of one-sided MCTs for* $p = 5$ *treatments, the Dunnett contrast, several numbers of endpoints, ratios* γ_{q1}*, and equicorrelations;* $\alpha = 0.05$.

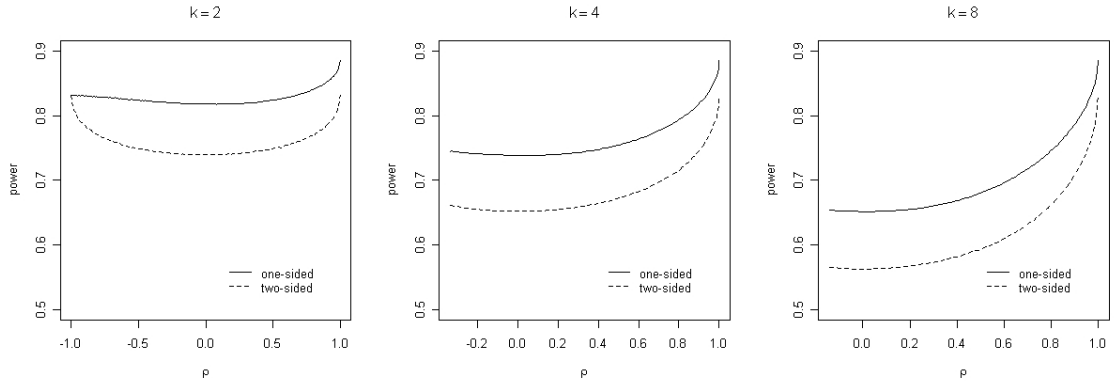


Figure 4.3: *Global power function of MCTs for $p = 3$ treatments, the Dunnett contrast, several numbers of endpoints, and equicorrelations; $\gamma_{q1} = 1.25$, $\alpha = 0.05$.*

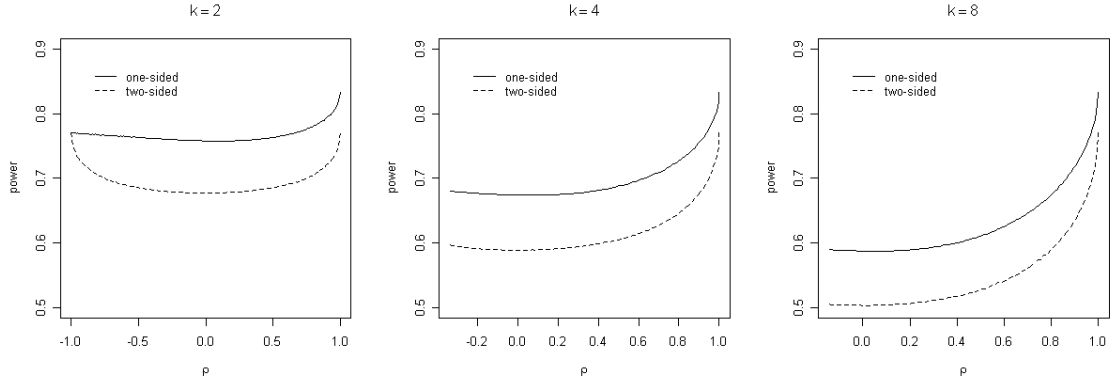


Figure 4.4: *Global power function of MCTs for $p = 5$ treatments, the Dunnett contrast, several numbers of endpoints, and equicorrelations; $\gamma_{q1} = 1.25$, $\alpha = 0.05$.*

The package `multtest` [Pollard et al., 2007] of the statistical software R [2008] provides resampling-based multiple hypothesis testing. Non-parametric bootstrap and permutation tests are implemented. Tests based on t- and F-statistics are included. The main application of this package is gene selection in microarray experiments. The function `MTP` performs test procedures for multiple endpoints by single-step and step-down minP and maxT methods to control the FWE (or other error rates). Tests based on t-tests are restricted to comparisons of two groups, e.g. a treatment and a control. A simulation study has been performed to compare this t-test-based bootstrap approach (Boot.) with the new MCT method (Multiv.). The single-step option `method=‘‘ss.maxT’’` has been used for comparability. The parameter background is the same as in Section 4.3, but $p = 2$ (hence $q = 1$) with $\theta_{1i} = 1$ for all $i = 1, \ldots, k$. Figure 4.5 shows the results of the mentioned power comparison. The rows are related to the different equicorrelations, the columns to the number of endpoints. Minimal and complete power coincide in this case, because the treatment group differs only for the first endpoint. A higher power of the new multivariate method is visible only for high correlations and high numbers of endpoints. Figure 4.6 shows the minimal power for the case that the mean of the treatment group was changed simultaneously for all endpoints and by the same relative amount. Except for the minimal equicorrelation ρ^{min}, the bootstrap method is better with respect to power than the new method. This difference becomes more pronounced with increasing correlation and with increasing number of endpoints. Figure 4.7 shows the complete power for the same background (simultaneously changing the mean of the treatment group for all endpoints). The bootstrap method has slightly less

power than the multivariate method, but this difference becomes negligible for high correlations. In summary: The power behavior of the competitors is almost equal. The gain in the minimal power for the bootstrap approach is insignificant in view of the properties and flexibility of the new MCT method.

4.6 Heteroscedasticity

The same assumptions are made as in Section 4.2 except that there are possibly different covariance matrices $\mathbf{\Sigma}_h = (\sigma_{h,ii'})_{i,i'}$ for the treatments $h = 1, \ldots, p$. In practice, that means the treatments to cause different variances or correlations over the endpoints, i.e.,

$$\{X_{hij} : i = 1, \ldots, k\} \sim \perp \mathrm{N}_k(\boldsymbol{\mu}_h, \mathbf{\Sigma}_h) \quad (h = 1, \ldots, p,\, j = 1, \ldots, n_h).$$

Furthermore, let $(n_h - 1) \geq k$ for all $h = 1, \ldots, p$. The appropriate test statistics are

$$\begin{aligned} T_{li} &= \frac{\sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh})\, \bar{X}_{hi}}{\sqrt{\sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh})^2\, S_{hi}^2 / n_h}} \\ &= \frac{(\boldsymbol{c}_l - \theta_{li} \boldsymbol{d}_l)'\, \bar{\boldsymbol{X}}_{,i}}{\sqrt{(\boldsymbol{c}_l - \theta_{li} \boldsymbol{d}_l)'\, \hat{\boldsymbol{V}}_i \boldsymbol{M}\, (\boldsymbol{c}_l - \theta_{li} \boldsymbol{d}_l)}} \quad (l = 1, \ldots, q,\, i = 1, \ldots, k), \end{aligned}$$

where $\hat{\boldsymbol{V}}_i$ is the estimator of $\boldsymbol{V}_i$ with

$$\boldsymbol{V}_i = \begin{pmatrix} \sigma_{1,ii} & & 0 \\ & \ddots & \\ 0 & & \sigma_{p,ii} \end{pmatrix} \quad \text{and} \quad \hat{\boldsymbol{V}}_i = \begin{pmatrix} S_{1i}^2 & & 0 \\ & \ddots & \\ 0 & & S_{pi}^2 \end{pmatrix} \quad (i = 1, \ldots, k),$$

respectively.

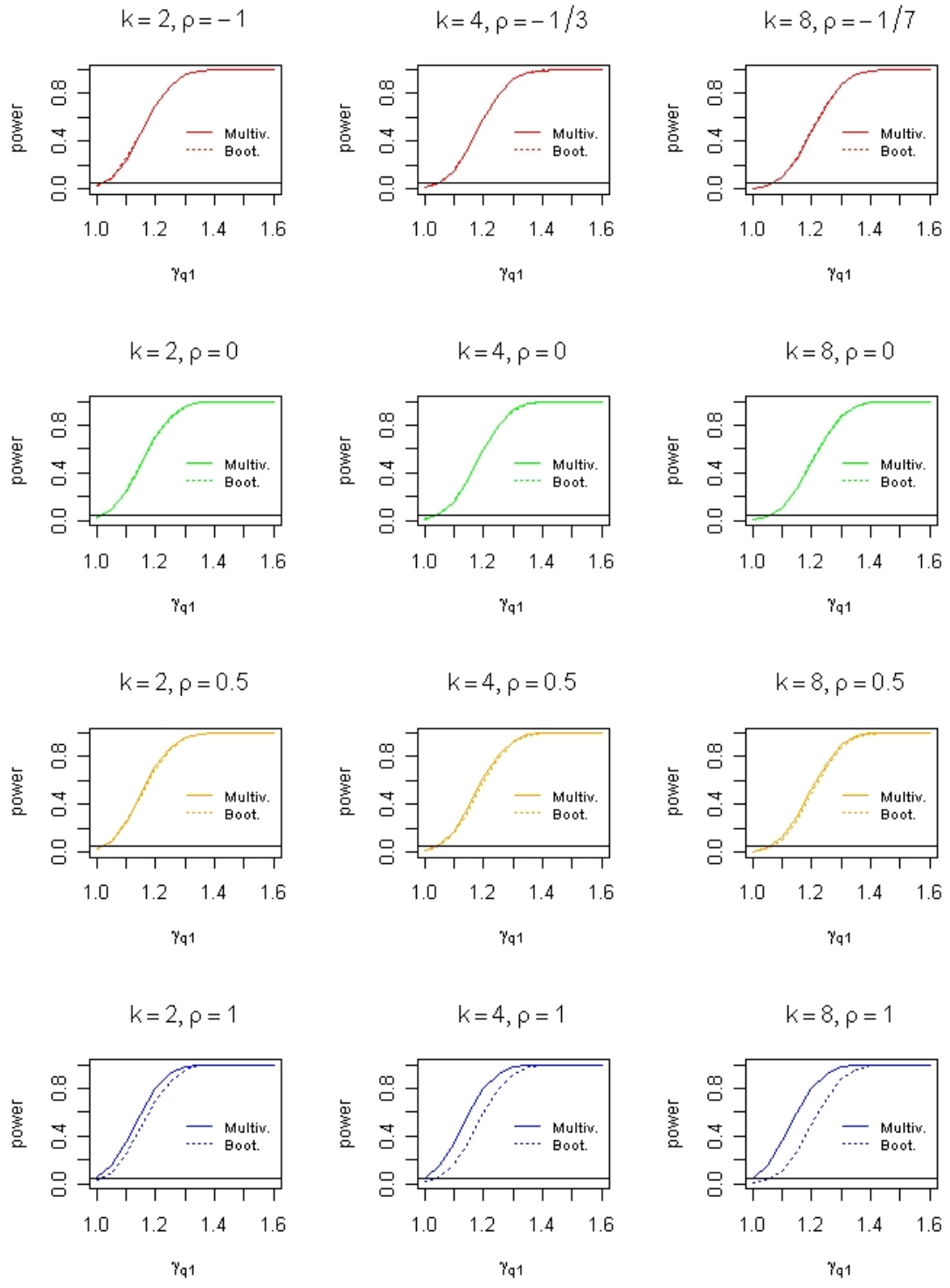


Figure 4.5: *Minimal and complete power function of one-sided MCTs for* $p = 2$ *treatments, the Dunnett contrast, several numbers of endpoints, and equicorrelations;* $\gamma_{q1} = 1$, $\alpha = 0.05$.

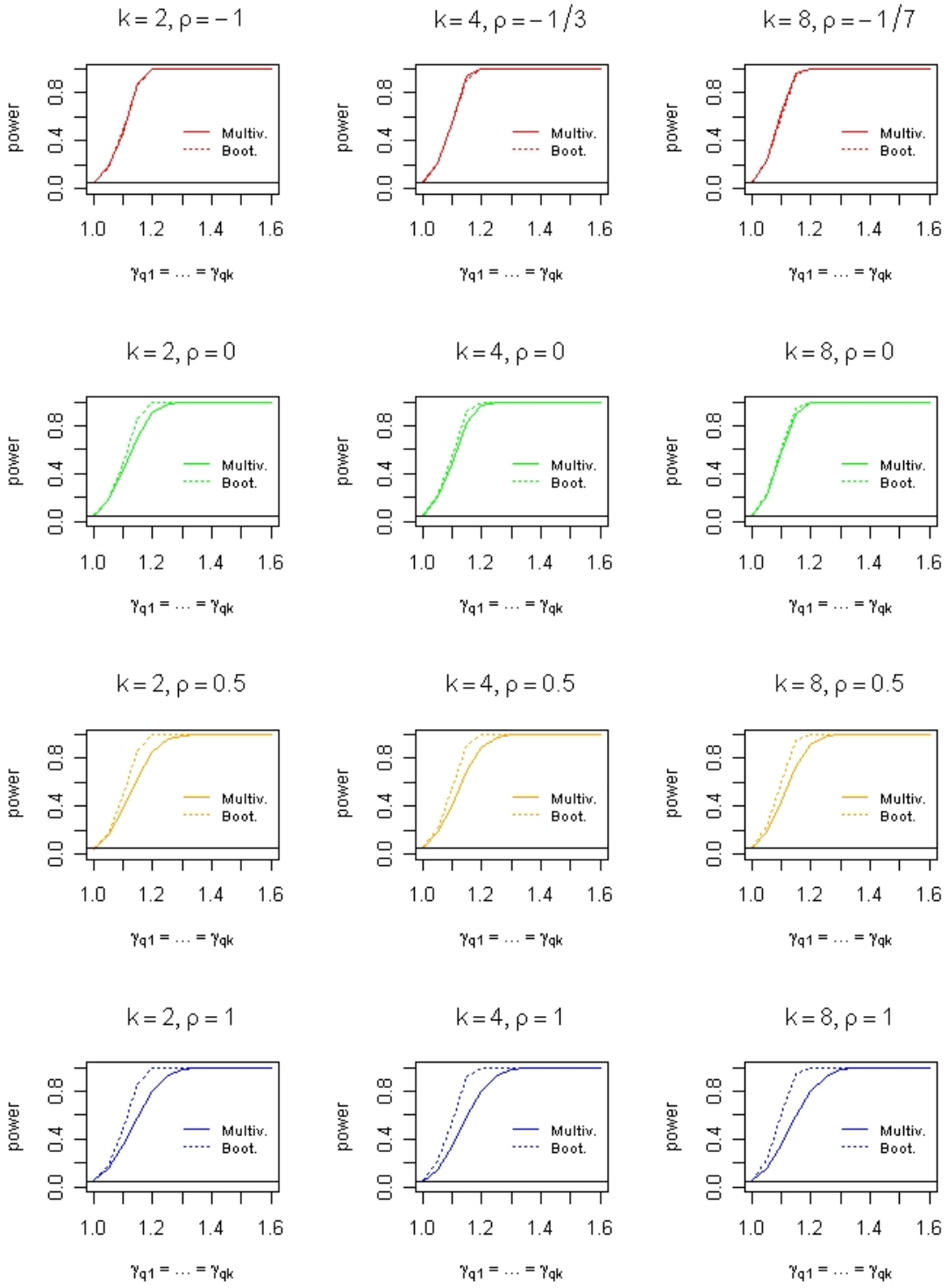


Figure 4.6: *Minimal power function of one-sided MCTs for $p = 2$ treatments, the Dunnett contrast, several numbers of endpoints, and equicorrelations;* $\gamma_{q1} = 1$, $\alpha = 0.05$.

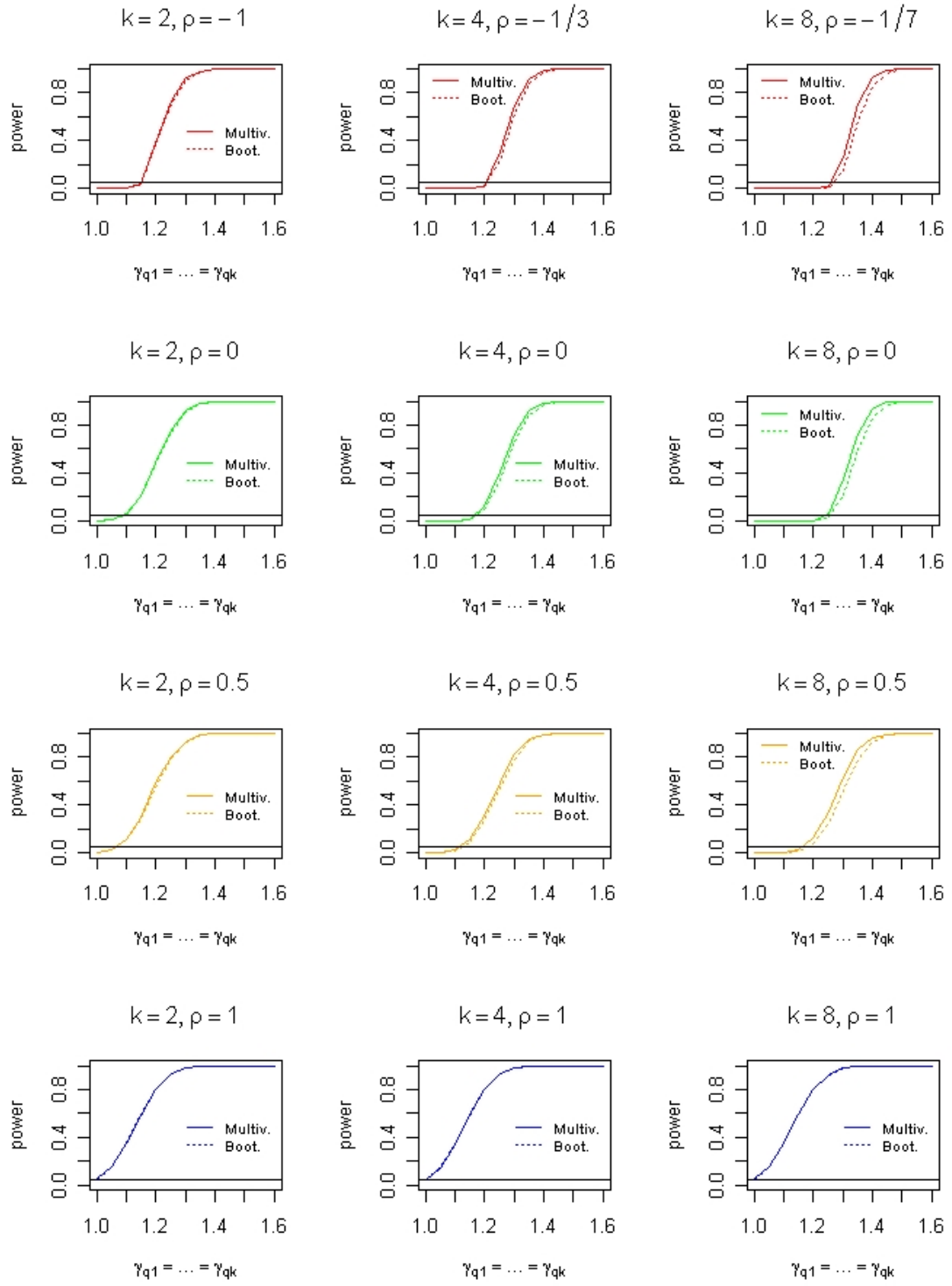


Figure 4.7: *Complete power function of one-sided MCTs for $p = 2$ treatments, the Dunnett contrast, several numbers of endpoints, and equicorrelations;* $\gamma_{q1} = 1$, $\alpha = 0.05$.

We apply now the results of Chapter 3, thereby extending the PI procedure (which is based on the method of Games and Howell [1976]) to the case of multiple endpoints. The degrees of freedom for the different contrasts and endpoints are related to Equation (3.11), i.e.,

$$\nu_{li} = \frac{\left(\sum_{h=1}^{p} \frac{(c_{lh}-\theta_{li}d_{lh})^2\sigma_{h,ii}}{n_h}\right)^2}{\sum_{h=1}^{p} \frac{(c_{lh}-\theta_{li}d_{lh})^4\sigma_{h,ii}^2}{n_h^2(n_h-1)}} \quad (l = 1,\ldots,q,\ i = 1,\ldots,k).$$

Replacement of the unknown parameters $\sigma_{h,ii}$ $(h = 1,\ldots,p,\ i = 1,\ldots,k)$ with the estimators S_{hi}^2 yields[1]

$$\hat{\nu}_{li} = \frac{\left(\sum_{h=1}^{p} \frac{(c_{lh}-\theta_{li}d_{lh})^2 S_{hi}^2}{n_h}\right)^2}{\sum_{h=1}^{p} \frac{(c_{lh}-\theta_{li}d_{lh})^4 S_{hi}^4}{n_h^2(n_h-1)}} \quad (l = 1,\ldots,q,\ i = 1,\ldots,k). \tag{4.7}$$

These different degrees of freedom again lead to different, non-equidistant quantiles for the test decisions. Therefore, this procedure is not a simultaneous test procedure in the sense of Gabriel [1969]. However, because it is a UIT, Theorem 2.1.3 holds, and it is coherent and consonant. The derivation of the correlation matrix can be performed in analogy to the one given in Section 4.2. We have

$$\begin{aligned} & Cov\left(\sum_{h=1}^{p}(c_{lh}-\theta_{li}d_{lh})\,\bar{X}_{hi}, \sum_{h=1}^{p}(c_{l'h}-\theta_{l'i'}d_{l'h})\,\bar{X}_{hi'}\right) \\ = & \sum_{h=1}^{p}(c_{lh}-\theta_{li}d_{lh})\,(c_{l'h}-\theta_{l'i'}d_{l'h})\,Cov\left(\bar{X}_{hi},\bar{X}_{hi'}\right) \end{aligned}$$

and

$$Cov(\bar{X}_{hi},\bar{X}_{hi'}) = \frac{1}{n_h} Cov(X_{hi},X_{hi'}).$$

[1]The degrees of freedom in (4.7) must be greater than or equal to 2 for a well defined distribution.

Now, we obtain

$$\rho_{ll',ii'} = Corr\left(\sum_{h=1}^{p}\left(c_{lh}-\theta_{li}d_{lh}\right)\bar{X}_{hi}, \sum_{h=1}^{p}\left(c_{l'h}-\theta_{l'i'}d_{l'h}\right)\bar{X}_{hi'}\right)$$

$$= \frac{\sum_{h=1}^{p}\left(c_{lh}-\theta_{li}d_{lh}\right)\left(c_{l'h}-\theta_{l'i'}d_{l'h}\right)\frac{1}{n_h}Cov(X_{hi},X_{hi'})}{\sqrt{Var\left(\sum_{h=1}^{p}\left(c_{lh}-\theta_{li}d_{lh}\right)\bar{X}_{hi}\right)}\sqrt{Var\left(\sum_{h=1}^{p}\left(c_{l'h}-\theta_{l'i'}d_{l'h}\right)\bar{X}_{hi'}\right)}}$$

$$= \frac{\sum_{h=1}^{p}\left(c_{lh}-\theta_{li}d_{lh}\right)\left(c_{l'h}-\theta_{l'i'}d_{l'h}\right)\frac{1}{n_h}\sigma_{h,ii'}}{\sqrt{\sum_{h=1}^{p}\left(c_{lh}-\theta_{li}d_{lh}\right)^2 Var\bar{X}_{hi}}\sqrt{\sum_{h=1}^{p}\left(c_{l'h}-\theta_{l'i'}d_{l'h}\right)^2 Var\bar{X}_{hi'}}}$$

and

$$\rho_{ll',ii'} = \frac{\sum_{h=1}^{p}\left(c_{lh}-\theta_{li}d_{lh}\right)\left(c_{l'h}-\theta_{l'i'}d_{l'h}\right)\frac{1}{n_h}\sigma_{h,ii'}}{\sqrt{\sum_{h=1}^{p}\left(c_{lh}-\theta_{li}d_{lh}\right)^2 \frac{1}{n_h}\sigma_{h,ii}}\sqrt{\sum_{h=1}^{p}\left(c_{l'h}-\theta_{l'i'}d_{l'h}\right)^2 \frac{1}{n_h}\sigma_{h,i'i'}}}$$

$$= \frac{\left(\boldsymbol{c}_l-\theta_{li}\boldsymbol{d}_l\right)'\boldsymbol{W}_{ii'}\boldsymbol{M}\left(\boldsymbol{c}_{l'}-\theta_{l'i'}\boldsymbol{d}_{l'}\right)}{\sqrt{\left(\boldsymbol{c}_l-\theta_{li}\boldsymbol{d}_l\right)'\boldsymbol{V}_i\boldsymbol{M}\left(\boldsymbol{c}_l-\theta_{li}\boldsymbol{d}_l\right)}\sqrt{\left(\boldsymbol{c}_{l'}-\theta_{l'i'}\boldsymbol{d}_{l'}\right)'\boldsymbol{V}_{i'}\boldsymbol{M}\left(\boldsymbol{c}_{l'}-\theta_{l'i'}\boldsymbol{d}_{l'}\right)}}$$

$$(1 \le l, l' \le q,\ 1 \le i, i' \le k),$$

where

$$\boldsymbol{W}_{ii'} = \begin{pmatrix} \sigma_{1,ii'} & & 0 \\ & \ddots & \\ 0 & & \sigma_{p,ii'} \end{pmatrix} \quad (1 \le i, i' \le k).$$

The behavior of this procedure is now predictable, in principle. Both the MCT for multiple endpoints described in this chapter and the MCT for heteroscedastic data described in Chapter 3 are the basis. Both have been checked by simulations and have been shown to have good properties. Nevertheless, they are approximate procedures, and their combination is approximate a fortiori. Estimators of variances and correlations are used, and the use of the multivariate t-distribution itself is an approximate approach. For that reason, a short simulation study is advisable here too, with the same background as in Section 4.3 but with some differences: The

correlations of the data have no influence on the FWE according to Section 4.3, thus random correlations are sufficient for that purpose. It is known from Chapter 3 that a setting, combining the smallest sample size and the highest standard deviation is the most critical one because it leads to a liberal behavior for a procedure that wrongly assumes homogeneity. Hence, a similar setting has been chosen with sample size 20 for each endpoint of each treatment, except for the last treatment with sample size 10. The standard deviations are $0.25\boldsymbol{\mu}_1$ for the last treatment, $0.1\boldsymbol{\mu}_1$ for the others. Additionally, not only the procedure described (referred to as PI) is considered here but also again a Bonferroni-adjusted version (referred to as BON) and a procedure which assumes homoscedasticity (referred to as HOM). A conservative variant of PI is also appended here which takes the minimum of the degrees of freedom (4.7) over the endpoints (referred to as MIN). Hence, not qk different degrees of freedom are used for MIN, but q. Table 4.7 (4.8) shows results of the first (second) study with three (five) treatments. As expected, HOM is liberal (0.131 – 0.320), BON is conservative (0.028 – 0.052), where increasing numbers of treatments and endpoints intensify these effects. PI varies around the nominal level of 0.05 with a slight tendency to liberalism. This is obviously caused by the mentioned high degree of approximation for that procedure. A dependence on the number of treatments and endpoints is not observed. The MIN procedure is an alternative choice with a realized level between BON and PI. It is less liberal than PI but can also be slightly conservative. The ranges here are 0.042 and 0.055, while PI has 0.047 and 0.063.

Despite the small variations around the nominal α-level, the adjusted version of the test behaves predictable. Hence, the behavior concerning the strong control of

MCT	Endpoints	HOM	BON	PI	MIN
Dunnett	2	0.131	0.049	0.055	0.053
	4	0.179	0.052	0.059	0.053
	8	0.227	0.048	0.055	0.049
Williams	2	0.137	0.035	0.047	0.046
	4	0.184	0.039	0.053	0.048
	8	0.244	0.036	0.055	0.042

Table 4.7: *FWE of one-sided MCTs for $p = 3$ treatments, several contrasts, procedures and numbers of endpoints; $\alpha = 0.05$.*

MCT	Endpoints	HOM	BON	PI	MIN
Dunnett	2	0.138	0.039	0.049	0.048
	4	0.190	0.041	0.053	0.048
	8	0.275	0.048	0.063	0.051
Williams	2	0.169	0.031	0.056	0.055
	4	0.227	0.028	0.053	0.049
	8	0.320	0.034	0.061	0.051

Table 4.8: *FWE of one-sided MCTs for $p = 5$ treatments, several contrasts, procedures and numbers of endpoints; $\alpha = 0.05$.*

the FWE, the SCIs and the power are also obvious from the previous sections and chapter. For that reason, we do not consider them here explicitly.

4.7 Example

Schulte et al. [2002] have taken measurements of 16 liver enzymes for an inter-laboratory immunotoxicity study in nine centers. Ten animals per sex have been randomized to a control (0) and three dose groups each. For simplicity, three enzymes are considered here from only the females and from the first center (see Table 4.9).

The aim is to show for which doses and for which enzymes the specified substance leads to significantly smaller values than the control. The control group is denoted

Dose	ASAT	ALAT	ALP
0	86.944 (11.880)	57.0752 (10.719)	461.496 (46.349)
1	80.080 (8.784)	49.2674 (4.922)	391.304 (47.909)
2	81.536 (18.957)	51.9610 (10.764)	308.976 (43.728)
3	81.536 (8.133)	46.3008 (8.349)	281.260 (29.945)

Table 4.9: *Sample means (and standard deviations) per dose and enzyme of the data set of Schulte et al. [2002].*

by $h = 0$ and the doses by $h = 1, 2, 3$. Application of the contrast matrices

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{c}_1' \\ \boldsymbol{c}_2' \\ \boldsymbol{c}_3' \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{d}_1' \\ \boldsymbol{d}_2' \\ \boldsymbol{d}_3' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

leads to the ratios of contrasts

$$\gamma_{li} = \frac{\mu_{li}}{\mu_{0i}} \quad (l = 1, 2, 3,\ i = 1, 2, 3).$$

The hypotheses to be tested are given by

$$H_{0,li} : \gamma_{li} \geq \theta \quad (l = 1, 2, 3,\ i = 1, 2, 3)$$

with $\theta = 1$. We assume heterogeneous covariance matrices over the doses. Table 4.10 shows the upper limits for the related approximate $(1 - \alpha)100\%$ SCIs for the ratios to the control means. The values in parentheses are the estimated ratios. For ALP, all doses show significantly smaller values than those of the control.

Dose	ASAT	ALAT	ALP
1	1.075 (0.921)	1.056 (0.863)	0.971 (0.848)
2	1.177 (0.938)	1.159 (0.910)	0.776 (0.670)
3	1.091 (0.938)	1.019 (0.811)	0.691 (0.609)

Table 4.10: *Upper confidence limits (and estimates) per dose and enzyme for the liver data of Schulte et al. [2002].*

As in Section 3.6.1, the testing problem has been formulated as a proof of hazard because toxicity should be pointed out. The conclusion of safety of the non-significant doses is not warranted. A proof of safety would be the proper choice to answer the question as to which doses do not cause a critical decrease as compared to the control. We do not consider that in detail here, however.

Chapter 5

Discussion

The problem of heteroscedasticity in MCTs is that it is impossible (up to now) to derive a joint distribution for the test statistics $(T_1^*, \ldots, T_q^*)'$. The basic idea of Games and Howell [1976] was to compare each test statistic with "its own" specific quantile coming from a comparison-specific multivariate t-distribution with degrees of freedom according to Satterthwaite [1946]. This method has been extended to the general case of MCTs for both differences and ratios of means (GH). A further step has been to plug the estimates S_h^2 $(h = 1, \ldots, p)$ into the correlation matrix instead of the unknown σ_h^2 (PI). Both comparison-specific degrees of freedom and a correlation matrix depending on sample variances are necessary to maintain the FWE over all situations. Approaches with a single degree of freedom (like HTL) may notedly fail the more the variances differ. Only in the homoscedastic situation, the methods considered, including the test for homoscedastic data (HOM), realize the same (and correct) α-level. In this case, PI has a power that is only slightly

smaller than that of HOM. Keeping in mind that it is hard to decide in practice whether the data are homoscedastic or not, the PI procedure can therefore be recommended as a default. Also many functions of the statistical software R [2008] assume heterogeneous variances for the data if no further information is specified by the user. For example, see the command `t.test(...)`. Moreover, calculations concerning MCTs for heteroscedastic data are already available from R [2008], using the commands `simtest.ratioVH(...)` and `sci.ratioVH(...)` of the package `mratios` [Dilba et al., 2008, 2007]. This package provides tests and confidence intervals for ratios of treatment means in the usual one-way layout.

Furthermore, the problem of many – possibly correlated – endpoints has been investigated. MCTs and related SCIs have been restricted to comparisons on a single endpoint so far. This methodology was extended to the case of an arbitrary number of endpoints by deriving an approximate multivariate t-distribution. Ratios of means have been considered for comparability of the different endpoints which may have different scales. An approach for differences of means has not been focused explicitly, but it can easily be obtained based on this work. If variances or correlations are assumed to differ for the different groups, the PI procedure for heterogeneous variances of Section 3 can be applied. The procedures presented can be shown to maintain the FWE. The version for heterogeneous covariances shows a slight liberalism, but it is in acceptable ranges. Test decisions (e.g., p-values) for all contrasts and all endpoints are available as well as SCIs. For this reason, a fair power comparison with existing methods is not feasible. A resampling-based competitor with the same features exists only for the case of comparisons of only two groups (package

`multtest` [Pollard et al., 2007] in R [2008]). Depending on which power is considered, the new method has about the same power properties or it is slightly worse. This is compensated by a gain in flexibility.

A software realization in R [2008] regarding the methods described in this work is available at `http://www.r-project.org`. The package `SimComp` [Hasler, 2008] provides calculations concerning simultaneous tests and confidence intervals for both difference- and ratio-based contrasts of normal means for data with possibly more than one primary endpoint. The covariance matrices may be assumed to be equal or possibly unequal for the different groups.

Bibliography

CPMP Concept Paper on the development of a Committee for Proprietary Medicinal Products (CPMP). Points to Consider on biostatistical/methodological issues arising from recent CPMP discussions on licensing applications: Choice of delta. At http://www.emea.eu.int/pdfs/human/ewp/215899en.pdf, 1999.

I. D. Adler and U. Kliesch. Comparison of single and multiple treatment regiments in the mouse bone marrow micronucleus assay for hydroquinone and cyclophosphamide. *Mutation Research*, 234:115–123, 1990.

R. B. Arellano-Valle and M. G. Genton. On the exact distribution of the maximum of absolutely continuous dependent random variables. *Statistics & Probability Letters*, 78:27–35, 2008.

A. Azzalini. The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, 32:159–188, 2005.

A. Azzalini and A. Capitanio. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society B*, 65(2):367–389, 2003.

A. Azzalini and A. Dalla Valle. The multivariate skew-normal distribution. *Biometrika*, 83(4):715–726, 1996.

P. Bauer. Multiple testing in clinical-trials. *Statistics In Medicine*, 10(6):871–890, June 1991.

E. Bofinger and M. Bofinger. Equivalence with respect to a control: Stepwise tests. *Journal of the Royal Statistical Society B*, 57(4):721–733, 1995.

F. Bretz. An extension of the williams trend test to general unbalanced linear models. *Computational Statistics & Data Analysis*, 50(7):1735–1748, 2006.

F. Bretz, A. Genz, and L. A. Hothorn. On the numerical availability of multiple comparison procedures. *Biometrical Journal*, 43:645–656, 2001.

M. B. Brown and A. B. Forsythe. Anova and multiple comparisons for data with heterogeneous variances. *Biometrics*, 30(4):719–724, 1974.

J. P. Buonaccorsi and H. K. Iyer. A comparison of confidence regions and designs in estimation of a ratio. *Communications in Statistics-Theory and Methods*, 13: 723–741, 1984.

G. Casella and R. L. Berger. *Statistical Inference*. Duxbury, Thomson Learning, 2002.

G. Dilba. *Simultaneous Inference for Ratios of Location Parameters*. PhD thesis, Gottfried Wilhelm Leibniz Universität Hannover, 2005.

G. Dilba, F. Bretz, V. Guiard, and L. A. Hothorn. Simultaneous confidence intervals for ratios with applications to the comparison of several treatments with a control. *Methods Of Information In Medicine*, 43(5):465–469, 2004.

G. Dilba, F. Bretz, and V. Guiard. Simultaneous confidence sets and confidence intervals for multiple ratios. *Journal Of Statistical Planning And Inference*, 136 (8):2640–2658, August 2006.

G. Dilba, F. Schaarschmidt, and L. A. Hothorn. Inferences for ratios of normal means. *R News*, 7(1):20–23, 2007.

G. Dilba, M. Hasler, and F. Schaarschmidt. *mratios: Inferences for ratios of coefficients in the general linear model*, 2008. R package version 1.3.5.

A. Dmitrienko, W. W. Offen, and P. H. Westfall. Gatekeeping strategies for clinical trials that do not require all primary effects to be significant. *Statistics In Medicine*, 22(15):2387–2400, August 2003.

C. W. Dunnett. A multiple comparison procedure for comparing several treatments with a control. *Journal Of The American Statistical Association*, 50(272):1096–1121, 1955.

C. W. Dunnett. Pairwise multiple comparisons in the unequal variance case. *Journal of the American Statistical Association*, 75(372):796–800, 1980.

C. W. Dunnett. Multiple comparisons between several treatments and a specified treatment. *Biometrics*, 41(2):571–571, 1985.

E. C. Fieller. Some problems in interval estimation. *Journal of the Royal Statistical Society, Series B*, 16:175–185, 1954.

K. R. Gabriel. Simultaneous test procedures - Some theory of multiple comparisons. *Annals Of Mathematical Statistics*, 40(1):224–250, 1969.

P. A. Games and J. F. Howell. Pairwise multiple comparison procedures with unequal n's and/or variances: a monte carlo study. *Journal of Educational Statistics*, 1 (2):113–125, 1976.

A. Genz and F. Bretz. Numerical computation of multivariate t-probabilities with application to power calculation of multiple contrasts. *Journal of Statistical Computation and Simulation*, 63:361–378, 1999.

A. Genz and F. Bretz. Methods for the computation of multivariate t-probabilities. *Journal of Computational and Graphical Statistics*, 11:950–971, 2002.

A. Genz, F. Bretz, and R port by T. Hothorn. *mvtnorm: Multivariate Normal and T Distribution*, 2008. R package version 0.8-3.

M. Hasler. *SimComp: Simultaneous comparisons for multiple endpoints*, 2008. R package version 1.2.

M. Hasler, R. Vonk, and L. A. Hothorn. Assessing non-inferiority of a new treatment in a three-arm trial in the presence of heteroscedasticity. *Statistics in Medicine*, 27(4):490–503, 2008.

D. Hauschke, R. Slacik-Erben, S. Hensen, and R. Kaufmann. Biostatistical assessment of mutagenicity studies by including the positive control. *Biometrical Journal*, 47:82–87, 2005.

Y. Hochberg. A sharper bonferroni procedure for multiple tests of significance. *Biometrika*, 75(4):800–802, December 1988.

Y. Hochberg and A. C. Tamhane. *Multiple Comparison Procedures*. John Wiley and Sons, Inc., New York, 1987.

S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70, 1979.

G. Hommel. A stagewise rejective multiple test procedure based on a modified bonferroni test. *Biometrika*, 75(2):383–386, June 1988.

H. Hotelling. A generalised t test and measure of multivariate dispersion. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 23–41, 1951.

L. A. Hothorn and D. Hauschke. Identifying the maximum safe dose: a multiple testing approach. *Journal of Biopharmaceutical Statistics*, 10:15–30, 2000.

T. Hothorn, F. Bretz, and A. Genz. On multivariate t and Gauss probabilities in R. *R News*, 1(2):27–29, 2001.

T. Hothorn, F. Bretz, and P. Westfall with contributions by R. M. Heiberger. *mult-*

comp: Simultaneous Inference for General Linear Hypotheses, 2008. R package version 0.993-2.

ICH E9 Expert Working Group. Ich harmonised tripartite guideline: Statistical principles for clinical trials. *Statistics In Medicine*, 18(15):1903–1904, August 1999.

S. Kotz and S. Nadarajah. *Multivariate t Distributions and Their Applications*. Cambridge University Press, 2004.

S. Kropf, L. A. Hothorn, and J. Läuter. Multivariate many-to-one procedures with applications to preclinical trials. *Drug Information Journal*, 31:433–447, 1997.

S. Lange and G. Freitag. Choice of delta: requirements and reality - results of a systematic review. *Biometrical Journal*, 47:12–27, 2005.

E. L. Lehmann. A theory of some multiple decision-problems .2. *Annals Of Mathematical Statistics*, 28(3):547–572, 1957.

N. Loperfido. Statistical implication of selectively reported inferential results. *Statistics & Probability Letters*, 56:13–22, 2002.

S. Nadarajah and D. K. Dey. Multitude of multivariate t-distributions. *Statistics*, 39(2):149–181, 2005.

M. Neuhäuser. How to deal with multiple endpoints in clinical trials. *Fundamental & Clinical Pharmacology*, 20(6):515–523, December 2006.

D. B. Owen. Tables for computing bivariate normal probabilities. *Annals Of Mathematical Statistics*, 27(1):1075–1090, 1956.

I. Pigeot, J. Schäfer, J. Röhmel, and D. Hauschke. Assessing non-inferiority of a new treatment in a three-arm clinical trial including a placebo. *Statistics in Medicine*, 22:883–899, 2003.

K. S. Pollard, Y. Ge, S. Taylor, and S. Dudoit. *multtest: Resampling-based multiple hypothesis testing*, 2007. R package version 1.16.1.

H. Quan, J. Bolognese, and W. Yuan. Assessment of equivalence on multiple endpoints. *Statistics in Medicine*, 20:3159–3173, 2001.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

C. Roberts. A correlation model useful in study of twins. *Journal Of The American Statistical Association*, 61(316):1184–1190, 1966.

F. E. Satterthwaite. An approximate distribution of estimates of variance components. *Biometrics*, 2:110–114, 1946.

A. Schulte, J. Althoff, S. Ewe, and H. B. Richter-Reichhelm. Two immunotoxicity ring studies according to OECD TG 407 - Comparison of data on cyclosporin A and hexachlorobenzene. *Regulatory Toxicology And Pharmacology*, 36(1):12–21, 2002.

T. Silva-Costa-Gomes, L. Gallart, J. Valles, L. Trillo, J. Minguella, and M. M. Puig. Low- vs. high-dose almitrine combined with nitric oxide to prevent hydroxia during open-chest one-lung ventilation. *British Journal of Anaesthesia*, 95(3):410–416, 2005.

A. C. Tamhane. A comparison of procedures for multiple comparisons of means with unequal variances. *Journal of the American Statistical Association*, 74(366): 471–480, 1979.

A. C. Tamhane and B. R. Logan. Finding the maximum safe dose level for heteroscedastic data. *Journal of Biopharmaceutical Statistics*, 14(4):843–856, 2004.

Y. L. Tong. *The multivariate normal distribution.* Springer: New York; 191-3., 1990.

J. W. Tukey. The problem of multiple comparisons. Dittoed manuscript of 396 pages New Jersey: Department of Statistics, Princeton University, 1953.

B. L. Welch. The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29:350–362, 1938.

B. L. Welch. On the comparison of several mean values - an alternative approach. *Biometrika*, 38(3-4):330–336, 1951.

P. H. Westfall. Multiple testing of general contrasts using logical constraints and correlations. *Journal of the American Statistical Association*, 92(437):299–306, 1997.

D. A. Williams. A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics*, 27:103–117, 1971.

Acknowledgments

I am grateful to my supervisor Ludwig A. Hothorn for his guidance. He never stopped keeping me on the go. That was the guarantee for the successful completion of this work.

I also thank all my colleagues at the Institute of Biostatistics. Hannelore Visser and Clemens Buczilowski provided organizational and technical support. I could always ask Cornelia Frömke, Daniel Gerhard, Xufei Mi and Kornelius Rohmeyer not only for statistical advice. I had lively discussions especially with Frank Schaarschmidt. Gemechis Dilba gave me some important ideas.

Last but not least, special thanks go to my family and friends for their encouragement and confidence.

Curriculum Vitae

Zur Person

Name:	Mario Hasler
Geburtstag:	15. Januar 1978
Geburtsort:	Gotha
Familienstand:	ledig

Bildung/ Wehrdienst

09/1984 – 08/1991	Polytechnische Oberschule Albert Schweitzer, Gotha
09/1991 – 06/1996	Gymnasium Ernestinum, Gotha
11/1996 – 11/1997	Zivildienst, Landratsamt Gotha, Abt. 6 - Umwelt
10/1997 – 09/2000	Studium der Wirtschaftsmathematik an der Friedrich-Schiller-Universität, Jena (ohne Abschluss)
10/2000 – 11/2004	Studium der Mathematik an der Friedrich-Schiller-Universität, Jena
seit 04/2005	Promotion an der Gottfried-Wilhelm-Leibniz-Universität, Hannover

Beruf

04/2005 – 09/2008	wissenschaftlicher Mitarbeiter an der Gottfried-Wilhelm-Leibniz-Universität, Hannover
seit 10/2008	wissenschaftlicher Mitarbeiter an der Christian-Albrechts-Universität zu Kiel