Powerful Modifications of Williams' Test on Trend

Vom Fachbereich Gartenbau
der Universität Hannover
zur Erlangung
des akademischen Grades eines

Doktors der Gartenbauwissenschaften
– Dr. rer. hort. –

genehmigte

Dissertation

von

Dipl.–Math. **Frank Bretz**
geboren am 14.10.1971 in Vlaardingen

1999

Referent: Prof. Dr. L.A. Hothorn
Korreferent: Prof. Dr. L. Baringhaus
Tag der Promotion: 6. Juli 1999

To my wife Jiamei for her love, constant support, and understanding that sometimes multiple contrast tests may come first.

Abstract (Schlagworte: *multipler Kontrasttest, multivariate* t*–Verteilung, Ordnungsrestriktion*)

Häufig stellt sich die Frage nach statistisch signifikanten monotonen Wirkungsverläufen quantitativer Einflußgrößen. Weist ein bestimmtes Herbizid mit ansteigender Dosis eine verbesserte Wirkung im Vergleich zu einer Kontrollgruppe auf? Treten bei jungen Kulturpflanzen mit abfallender Temperaturbehandlung signifikant häufiger Anomalien auf? Fragestellungen dieser Art bilden den Schwerpunkt der vorliegenden Dissertation. Im Gegensatz zur herkömmlichen Varianzanalyse wird hier ein monotones Wirkungsprofil vorausgesetzt, um von dieser Annahme ausgehend mächtigere Tests zu entwickeln. Wie in der Dissertation hervorgehoben wird, bergen jedoch die klassischen Trendtests von Bartholomew, Williams und Marcus z.T. erhebliche Nachteile. Darunter fällt die ungelöste Problematik der numerischen Verfügbarkeit unter der Null- oder Alternativhypothese, inbes. im wichtigen unbalanciertem Fall. Ferner führt die unzureichende Kenntnisnahme der Fallzahlaufteilung in den Varianzschätzern bei Williams und Marcus zu einem unbefriedigenden Güteverhalten. Diese und weitere Nachteile schränken die Anwendung der drei klassischen Trendtests somit stark ein.

Das Ziel der Dissertation besteht darin, mittels dem Konzept der multiplen Kontrasttests die Problematiken zumindestens teilweise zu entschärfen. Hierbei wird das Maximum über mehrere einzelne Kontrasttests (standardisierte Linearkombinationen der Mittelwerte) betrachtet. Ein einzelner Kontrast ist auf Grund seiner Definition für eine bestimmte Wirkungskurve sehr mächtig, reagiert aber empfindlich auf Abweichungen derselbigen. Der Maximumtest hingegen wählt die beste Teststatistik aus und ist demnach weniger anfällig gegenüber unterschiedlichen Dosis-Wirkungs-Verläufen. Darauf basierend wird der Williamstest in die Theorie der multiplen Kontraste eingebettet. Eine ausführliche Behandlung der zugrunde liegenden multivariaten *t*–Verteilung ermöglicht seine uneingeschränkte Anwendung. Quantile und p-Werte sind auch im Unbalancierten einfach zu berechnen. Ein veränderter Varianzschätzer nimmt die Fallzahlaufteilung besser zur Kenntnis und auf Grund der Konstruktion der multiplen Kontraste hängt die Güte des neuen Tests weniger stark von der Wirkungsfunktion ab.

Darüber hinaus wird auch der Marcustest auf multiple Kontraste verallgemeinert. Ausgehend von einer vollständigen Aufteilung des Alternativraumes bildet ein dritter Zugang das Maximum über lokal güteoptimaler Einzelkontraste (isotonischer Kontrast). In einer ausführlichen Gütestudie werden diese Tests mit den originalen Trendtests verglichen. Die Herleitung weiterer theoretischer und numerischer Resultate ermöglichen insbes. eine geschlossene Darstellung der Güteformel zur iterativen Fallzahlbestimmung und weiterführenden post-hoc Analyse. Die dazu benötigte nichtzentrale multivariate *t*-Verteilung ist nun anlog zur oben erwähnten zentralen Form ohne Beschränkung der Korrelationsstruktur verfügbar. Ein weiteres Ergebnis reduziert die effektive Dimension eines beliebigen multiplen Kontrasts auf die Anzahl der zu untersuchenden Gruppen, was zu erheblich vereinfachten Auswertungen führt.

Abschließend wird vor allem praxisorientierten Fragestellungen nachgegangen. Die bisherigen Ergebnisse für normalverteilte Daten werden vollständig auf den dichotomen Fall übertragen. Die sich ergebenden Asymptotiken erfordern insbes. die Untersuchung der multivariaten Normalverteilung, welche nun im allgemeinen Fall zur Verfügung steht. Die Betrachtung spezieller Aspekte binomialen Testens (Kontinuitätskorrektur, gepoolte/ungepoolte Versionen, exakte bedingte und unbedingte Verteilungen) erweitern die Anwendungsmöglichkeiten. Ferner werden Ansätze zur Bestimmung ausgewählter Parameter vorgestellt (z.B. die Bestimmung einer minimalen effektiven Dosis). Weitere Anwendungsmöglichkeiten werden kurz angerissen (nichtparametrische Analyse, Konfidenzintervalle, höherfaktorielle Anlagen, etc). SAS/IML und FORTRAN Programme sind erstellt worden und im Anhang dokumentiert.

Abstract (Keywords: *multiple contrast test, multivariate* t*–distribution, order restricted inference*)

Frequently the question arises whether given dose-response shapes of quantitative variables show any statistically significant effect. Does the efficacy of a certain herbicide indeed improves with increasing doses when compared to a control group? Has the temperature a significant influence on the occurrence of anomalies in young kohlrabi plants? These and similar questions are analysed in the present thesis. In contrast to the usual analysis of variance one assumes a monotonous dose-response profile. Based on this assumption new tests are developed, which show an improved power behaviour. However, as it is seen in more detail in the thesis, the classical trend tests of Bartholomew, Williams and Marcus bear a series of disadvantages. Among other issues these involve the unsolved problem of evaluating the distribution functions under the null and the alternative hypotheses, in particular in the important case of unequal replications. Moreover, the test statistics of Williams and Marcus do not take the sample size allocation sufficiently into account. These and other disadvantages restrict seriously the application of the three classical trend tests for practical purposes.

The aim of the thesis is to overcome at least partially these problems by applying the concept of multiple contrast tests. Here, the maximum is taken over several single contrast test statistics (standardised linear combinations of the means). Due to its definition a single contrast test is very powerful for a fixed dose-response curve. But already for small departures from it the test may bear a poor power behaviour. The above mentioned maximum test, however, chooses the best test statistic and is therefore more robust against varying dose-response functions. Hence, Williams original test is embedded in the theory of multiple contrast tests. An intensive discussion of the underlying multivariate *t*–distribution enables an unrestricted use of the new test. Quantiles and p-values are easily calculated in unbalanced set-ups. A modified variance estimator takes the sample size allocation better into account. Due to the construction of multiple contrast tests the power of the new test depends less on the dose-response shape.

Moreover, Marcus original test is generalised similarly. A third new contrast definition is provided by decomposing the alternative space in the smallest possible sub-hypotheses and taking subsequently the maximum over the locally optimal single contrasts (isotonic contrast). In a detailed power study the performances of these multiple contrast tests are compared with the original trend tests. The derivation of further theoretical and numerical results enables the representation of a power formula in a closed form for iterative sample size determination and for further leading post-hoc analysis. Similarly to the above mentioned central case the arising non-central multivariate *t*–distribution is now available without restriction of the correlation structure. A further result reduces the effective dimensionality of arbitrary multiple contrasts to the total number of treatments under investigation, leading to clearly simplified evaluations.

Finally, further important practical problems are investigated. The results obtained so far for normal variates are generalised to the dichotomous case. The arising asymptotics demand in particular the investigation of the multivariate normal distribution. Its evaluation is now available in the general unrestricted case. The consideration of special aspects of binomial testing (continuity correction, pooled/unpooled versions, exact conditional/unconditional distributions) extends the range of applications. Furthermore, approaches for the determination of certain parameters in dose finding studies are presented (e.g. the minimum dose with maximum effect or the minimum effective dose). Further applications are sketched briefly (nonparametric analyses, confidence intervals, higher factorial layouts, etc.). SAS/IML and FORTRAN programs have been written for most applications and are enclosed with the thesis.

Contents

Introduction **1**

1. Survey of trend tests for normal means **9**

1.1. General notations 9

1.2. Maximum likelihood estimators under total order restriction 11

1.3. The trend tests according to Williams, Marcus, Bartholomew and multiple contrast tests 17

1.3.1. Williams' $\bar{t}$ – test 17

1.3.2. Marcus' $\bar{t}^{\,\mathrm{mod}}$ – test 19

1.3.3. Likelihood ratio test 20

1.3.4. Multiple contrast test 22

1.3.5. Example 29

1.4. Overview of other trend tests 32

2. Multivariate normal and *t*-distribution **35**

2.1. Multivariate normal distribution 35

2.1.1. Definition and basic properties 35

2.1.2. Computation of multivariate normal probabilities 42

2.1.2.1. Approximation of Solow 43

2.1.2.2. Transformations of Genz 45

2.1.2.3. Calculation of orthant probabilities 47

2.1.2.4. Conclusions 50

2.1.3. Calculation of level probabilities 51

2.2. Multivariate *t*–distribution 54

2.2.1. Definition and basic properties 54

2.2.2. Computation of multivariate *t*–probabilities 59

2.2.2.1. The methodologies of Somerville 61

2.2.2.2. Transformations of Genz 62

2.2.2.3. Computation of equicoordinate quantiles 70

2.2.2.4. Numerical comparisons and conclusions 73

3. Choice of appropriate contrast coefficients **78**

3.1. Review of contrast definitions 78

3.2. A Williams-type multiple contrast test 82

3.3. A Marcus-type multiple contrast test 85

3.4. A new multiple contrast definition 87

3.5. Example 92

3.6. Reduction of the dimensionality of multiple contrast tests 94

4. Power comparison for normal data **101**

4.1. Power expression of multiple contrast tests 101

4.2. Power study 105

4.3. Conclusions 124

5. Power comparison for binomial data **126**

5.1. Notations 127

5.2. Power expression of multiple contrast tests 134

5.3. Introduction of new contrast tests 138

5.3.1. Continuity correction 138

5.3.2. Unpooled version 143

5.3.3. Exact conditional and unconditional versions 147

5.3.4. Example 151

6. Estimation of the minimum effective dose **153**

7. Summary and complements **161**

References **170**

Appendix **179**

In Saloniki kenn ich einen, der mich liest,
und auch in Bad Nauheim – das sind schon zwei.
Günter Eich, *Zuversicht*

Introduction

In many research areas the objective of an experiment is to test whether the efficacy of a new treatment or drug is improved with respect to a certain control group. A natural way of conducting these kind of tests is to consider several treatment levels of the new compound, drug, fertiliser, herbicide, ... and compare them with a reference group, which response is assumed to be known due to prior knowledge of its behaviour. Such a reference group can be for example a negative control group without any administration or of a vehicle only. In these cases the goal of the user would be to find out whether the new developed treatment shows any (statistically) significant response at all. By choosing the reference to be a well known standard application, the aim differs. Here the scientists wants to investigate whether the new treatment is not only better than a negative control but even better than the standard.

Formalising the introduced terms above, we denote by *C*- a negative and by *C*+ a positive control group. Additionally, $D_1, \ldots, D_k$ stand for k treatment or dose levels. Therefore, the first situation mentioned consists of an analysis of the design $[C-, D_1, \ldots, D_k]$. In the second case the design $[D_1, \ldots, D_k, C+]$ would have been chosen instead. The number k is usually small due to practical reasons, frequently $k \in \{2, 3, 4\}$. More complex designs, i.e. including more than one reference group, are possible, but will not be considered throughout this thesis. For applications when using *C*- and *C*+ simultaneously, the reader is referred to Hothorn (1995) and Bauer et al. (1998).

The classical statistical approach to analyse such $(k+1)-$sample situations in the randomised one-way layout is the analysis of variance (ANOVA). However, both the *F*-test of the ANOVA and corresponding nonparametric test procedures are only able to detect any difference among the investigated samples. But frequently the user is more interested in

specific results rather than in such general assessments. For example, one might be interested in analysing the dose-response dependence of the data. In these cases the goal is to detect a global trend. Therefore, more information is required than usually established by the classical tests.

To illustrate these ideas consider the data provided by Banno and Yamagami (1989) as an example. They studied the conversion efficiency of ingested food (E.C.I.) of the wood-feeding insect *Eupromus ruber* at five larva stages (third to seventh instar) and an adult stage. The endpoint was calculated as

$$\text{E.C.I.} = 100 \times \frac{\text{dry weight of a larva or an adult}}{\text{dry weight of wood consumed}}$$

for each individual larva and adult. The following table summarises the main statistical quantities. Here, the groups 1, ..., 5 correspond to the seventh through third instar and the index '0' is associated with the adult stage. The present design is of the form $[D_0, \ldots, D_5]$, even if the adult stage can not be regarded as a 'control' in the classical sense.

Stage *i*	0	1	2	3	4	5
Mean	1.669	1.923	2.009	2.129	2.411	2.415
Std. dev.	0.5316	0.4079	0.922	0.8452	0.7974	1.184
Sample size	21	10	15	17	21	4

The main question of interest, from the authors point of view, was to investigate whether the E.C.I. decreased monotonously over all development stages. Does a larva from a lower development stage has a significant higher E.C.I. regarding to those of a higher development stage, up to the adult form? Assume that a statistical significant dose-response relationship (i.e. different from constant) has been detected. A further question of interest could be the identification of the highest development stage among the larvae, which still yields a significant difference to the adult stage. This problem is closely related to the estimation of a minimum effective dose (MED) in clinical and pre-clinical trials and to the whole theory of

trend tests in general. From the biologist point of view, looking at the data, these questions might have only one answer. Nevertheless, a statistical analysis should be conducted to assure the evidence of a possible trend with respect to the development stages.

Some further selected examples from the literature underline the importance and the broad field of applications of detecting significant trends among several treatments. Consider the data provided in the table below as a next example (Saville and Wood, 1991, p. 141). They refer to a field experiment, which was conducted to determine how the grain yield of spring sown malting barley was affected by different seeding rates. A randomised one-way layout was chosen with the five treatments representing the five seeding rates 50 kg/ha through 150 kg/ha. Each of the six replicates were harvested from plots of size 40 m by 1.25 m. In the light of above considerations we first notice that no negative control is present. Otherwise a control group with seeding rate 0 kg/ha would have been included in the trial. As the description of the data does not clarify whether a standard seeding was included, we do not assume the existence of $C+$ as well. Therefore the present design is of the pattern $[D_1, D_2, \ldots, D_5]$. The main question which naturally arises in this context is, whether the grain yield increased with increasing seeding rate.

Seeding rate	Grain yield						Mean	Std. dev.
50 kg/ha	25.4	22.4	25.2	24.4	24.2	22.0	23.93	1.42
75 kg/ha	26.2	26.2	25.2	26.4	25.0	27.8	26.13	1.00
100 kg/ha	27.6	27.6	26.0	25.8	26.2	25.8	26.50	0.86
125 kg/ha	27.6	28.2	26.8	26.6	28.0	27.8	27.50	0.65
150 kg/ha	27.2	28.2	26.8	25.6	27.2	27.6	27.10	0.87

Petersen (1985) described an experiment in order to assess whether the addition of particular enzymes retarded the separation of frozen orange juice shortly after the addition of water to the frozen concentrate. The experiment reported consisted of a control with no treatment at all and four levels of a certain enzyme (1, 2, 3 and 4 ppm). Conducting four replications in a completely randomised design the arithmetic means 6.68, 29.15, 36.28, 43.89 and 49.12 were obtained (time to separation in minutes). Does the presence of the enzyme retard separation as

compared to its absence? Is there any differential effect of the level of added enzyme? One further experiment provided by Saville and Wood (1991, p. 529) was carried out to determine the effect of the weedkiller oxadiazon on the early development of peach seedlings. In a typical randomised design consisting of a control, half dose, single dose and triple dose with unequal replications (6, 6, 5 and 3, respectively) the resulting heights of the seedlings are shown in the table below. Does the herbicide oxadiazon indeed influence the development of the seedlings? And if so, which would be the statistically significant smallest dose with such an effect?

Treatment (kg/ha)	Height of seedlings (cm)					
0	79	76	57	105	81	71
0.375	71	34	35	78	79	59
0.75	63	60	61	68	44	
2.25	11	23	16			

Further examples can be found in many other textbooks and articles. In the course of the present thesis we will encounter a number of additional material which demonstrates the wide range of application of trend tests.

To investigate the statistical problems sketched above a new class of tests has been introduced in the literature in the past 40 to 50 years. A variety of trend tests were proposed, many of them with satisfactory power results for specific constellations. But one main disadvantage of this whole approach is that no uniformly most powerful test is at hand. All of the trend tests presented later in this thesis depend, sometimes stronger, sometimes weaker, on the underlying dose-response shape (see also Neuhäuser, 1996). Therefore, the research for powerful trend tests (yet easy to conduct) is still ongoing. One approach within this wide range of analyses is the likelihood ratio test under total order restriction (LRT) according to Bartholomew (1959, 1961). Even if no uniformly best test exists, the LRT has a reasonable power performance and is conjectured to provide the highest 'average' power among the present trend tests. However, the LRT lacks a wide use for practical applications. Several articles discuss this contradiction in view of the fact of its superior power behaviour, see for

example Tang and Lin (1997) or Agresti and Coull (1998). As we will see in the sequel, the LRT is regarded to behave less robust against certain types of violations of its assumptions, such as variance heterogeneity and non-normality. Apart from this, one crucial drawback lies in the difficulty to evaluate the null distribution. Long time the use of the LRT was restricted to strictly balanced designs, an assumption which is frequently violated in practice. The generalisation to unbalanced set-ups got only possible with modern computer skills and new statistical techniques.

Because of such practical problems when implementing the LRT, many researchers tried to develop alternative testing procedures. One important approach is due to Williams (1971, 1972). Since its publication it has frequently been used in both medical and non-medical applications. Introduced originally for normal distributed data only, several generalisations to dichotomous and nonparametric set-ups and higher factorial layouts permit a wide range of applications. Shirley (1996, p. 26) emphasised accurately in her literature review of trend tests the distinguishing features, that

"generally, Williams' $\bar{t}$ – test is favoured in the literature because of its robustness to non-normality, lack of balance, and non-monotonicity of dose-response. Bartholomew's test comes a close second because of its superior power overall."

It becomes clear that Williams' test is regarded as having good robust characteristics against several types of violations of its assumptions. In particular, as Shirley (1996) points out again in the sequel of her paper, Bartholomew's test is less robust than Williams' version. On the other hand, it is well recognised that the $\bar{t}-$test has on average a lower power than the LRT. Common to both tests, however, are their complicated distributions under the null hypothesis. No general method is available to compute quantiles quick and accurately for Williams' test in the general unbalanced case. This restricts the use of the $\bar{t}-$test to strictly balanced situations, although it is robust against smaller departures of the required balance. However, it has been shown (Bretz and Hothorn, 1999) that Williams' test maintains less and less a pre-determined $\alpha-$level as the degree of imbalance increases.

Many other trend tests were proposed in the literature. Based on the insights sketched above, the search for new trend tests is conducted from one point of view only. The general goal is to combine the following main features:

- good power behaviour comparable to the LRT throughout the alternative space;
- easy numerical implementation of the test statistics, a problem of particular importance, for the multivariate nature of comparing several treatments makes an easy handling difficult;
- robustness against specific violations of the assumptions in the sense Hothorn (1989) has shown for Williams' test.

The present thesis should be considered in this context of developing new procedures for statistical inferences under order restriction. The aim of this thesis is to fill the gap between the approaches of Williams and Bartholomew. Starting from Williams' $\bar{t}$ – test the attempt is made to derive new, improved test statistics. This is done by applying the basic concept of Williams to the class of multiple contrast tests (MCTs) according to Mukerjee et al. (1986, 1987). The resulting test combines several advantages of the involved approaches and can be applied to the general case of unequal sample sizes without further restrictions. Improvements on the numerical methods available so far result in fast evaluations of the corresponding null hypothesis. Simulation results suggest that the power behaviour of the new approach is close to that of the LRT under a variety of conditions. Generalisations to nonparametric and dichotomous set-ups, as well as robustifications against outliers and applications to higher factorial experiments are possible and straight forward. In this sense the spirit of the present thesis is well described by McDermott (1998) in his abstract:

"The likelihood ratio test for equality of order-constrained means is known to have power characteristics that are generally superior to those of competing procedures. Difficulties in implementing this test have led to the development of alternative approaches, such as tests based on single and multiple contrasts."

The thesis is roughly outlined as follows. Chapter 1 presents an overview of the most important procedures in the parametric case of normal data for testing on equality of several means under total order restriction. Emphasis is given on Williams' $\bar{t}-$test, the LRT of Bartholomew, Marcus' (1976) modified $\bar{t}-$test and single and multiple contrast tests. Other methods are briefly mentioned and their links to existing tests are established. Further on, general notations and basic concepts important for the reading of Chapter 3 through 7 are introduced. The example of comparing the conversion efficiency among several larva stages of *Eupromus ruber* is analysed in detail and provides additional motivation for improving Williams' test.

As already pointed out, the null distributions of multivariate tests considered in the present context are in general difficult to compute. When developing the ideas of multiple contrast tests further in Chapter 3 through 7 we need the ability of computing both the multivariate normal and multivariate *t*-distribution under several aspects. Chapter 2 provides a discussion in depth of this topic. Theoretical results, as far as required, are cited or proven. Numerical algorithms for the calculation of both multivariate distribution functions are introduced, which can be applied to a variety of different problems and situations. This chapter provides the theoretical and numerical fundamentals for the remaining thesis. Important developments are achieved in evaluating the null distributions of the LRT and MCTs in the general unbalanced set-up.

In Chapter 3 we will focus on appropriate choices of contrast sets. The problem of the empirical determination of contrast coefficients is discussed. Based on the approaches of Williams, Marcus and Bartholomew, three attempts of new definitions are made. An useful result for possible high dimensionality problems in connection with MCTs is derived.

With these new formulated test statistics we provide an extensive power study for normal distributed data in Chapter 4. A power function in closed form for arbitrary multiple contrast tests is derived and optimal sample size determination is discussed. We compare the three mentioned MCTs with the corresponding original versions for a variety of scenarios, including different total sample sizes, variable sample size allocations within the groups and the influence of the choice of a predefined α among other aspects.

In Chapter 5 generalisations to the binomial case are given. We establish asymptotic power and sample size functions in closed form for single and multiple contrasts. Alternative methods are discussed where the derivation fails to succeed. Further on, we generalise the concept of dichotomous contrast tests developed so far. Among other topics we introduce a continuity correction and discuss its appropriate definition. We compare pooled with unpooled asymptotic versions and develop the ideas of Neuhäuser (1996) further by providing conditional and unconditional exact MCTs. Brief power and size comparisons are given for each topic. Finally, an example analysed in detail illustrates and summarises the main ideas of the chapter.

We will focus on the important point of estimating the MED in Chapter 6. Instead of testing the global null hypothesis only, we sequentially conduct several tests according to the closure principle of Marcus et al. (1976). Further assumptions of monotonicity and restricted comparisons to the control lead to simple testing procedures, where at each step a conditional testing at full size α is allowed. An outlook on other dose estimations, such as the maximum effective dose, is given.

In the final Chapter 7 we summarise our results and try to provide advises to the practitioner as far as possible. Afterwards, further applications are investigated briefly. These include a short discussion about other order restrictions than the simple ordering. Additionally, the cases of non-parametric analysis and variance heterogeneity are considered among other topics.

Appendix A contains the balanced contrast sets of the proposed tests in Chapter 3 up to dimension six. Appendix B includes some of the algorithms used throughout the thesis. Most of them will refer to Chapter 2. Because of the widespread use of the statistical computation package SAS in statistics and its applications, most of the algorithms presented and calculations provided were implemented in SAS, version 6.12. The use of other software is mentioned at the respective passages.

1. Survey of trend tests for normal means

In this chapter the most important procedures from literature for testing the equality of several means under total order restriction will be reviewed. But before doing this we introduce some basic notations in the first section, which will be valid for the whole thesis. Afterwards we review briefly the theoretical aspects of maximum likelihood estimation under total order restriction. The results stated here are fundamental for the understanding of the presented trend tests in Section 1.3. These are the procedures due to Williams (1971), Marcus (1976) and Bartholomew (1961), which are all based on the principle of maximum likelihood estimation. Finally, a different approach due to Mukerjee et al. (1987), the multiple contrast test, is also discussed in this section. Common to all these four approaches is their importance in the course of this thesis. In the last Section 1.4. we touch briefly on other procedures and try to demonstrate their relationships to the preceding trend tests discussed in Section 1.3.

1.1. General notations

Suppose the following randomised fixed effect one-way layout model

$$X_{ij} = \mu_i + \varepsilon_{ij}, \qquad i = 0, 1, \ldots, k,\ j = 1, \ldots, n_i,$$

with one control group and k treatment or dose levels, labelled by 0, 1, 2, ..., k, respectively. Let $\left\{X_{ij}\right\}_{ij}$ be the sample values, identically and independently normal distributed with the unknown means $\mu_0, \mu_1, \ldots, \mu_k$ and common variance σ^2, i.e. $X_{ij} \sim N\left(\mu_i, \sigma^2\right)$. The variable n_i denotes the sample size of the i^{th} group. For most parts of this thesis we therefore impose no restrictions on the sample sizes, but assume that the unknown variances are equal between the treatment groups. We denote further the sample mean $\sum_j X_{ij} \big/ n_i$ by $\overline{X}_i$ for $i = 0, 1, \ldots, k$ and by $s^2 = \frac{\sum_{i=0}^{k} \sum_{j=1}^{n_i} \left(X_{ij} - \overline{X}_i\right)^2}{\nu}$ the pooled variance estimator with $\nu = \sum_i n_i - \left(k+1\right)$ degrees of freedom. Our goal is to test the null hypothesis of no effect between the $k+1$ dose groups

$$H_0: \mu_0 = \mu_1 = \ldots = \mu_k. \tag{1.1}$$

When applying the classical ANOVA, the alternative is generally formulated as $H_A': \exists\, i, j: \mu_i \neq \mu_j, i \neq j,\ i, j \in \{0, 1, \ldots, k\}$. The F-test therefore states only differences of any two treatment groups and allows no further conclusions about their identification. This is clearly not an appropriate way of testing in our situation. The first modification is therefore to consider comparisons to the control only. The corresponding (two-sided) alternative would then be stated as $H_A'': \exists\, i: \mu_i \neq \mu_0,\ i \in \{1, 2, \ldots, k\}$. For such situations the tests according to Dunnett (1955, 1964) in the parametric case and to Steel (1959) in the non-parametric case are standard. But again this alternative is not suitable to the present interesting context, because we assume a relationship among the means μ_i to hold. If there is any difference between the treatment levels, we assume the response to increase (or decrease) monotonically with respect to increasing levels. In the first example given in the Introduction it is natural to assume that for higher development stages the conversion efficiency, if at all, decreases. We therefore take such monotonic dose-response dependencies into account and restrict the alternative space again by formulating the one-sided hypothesis

$$H_A: \mu_0 \leq \mu_1 \leq \ldots \leq \mu_k,\ \mu_0 < \mu_k. \tag{1.2}$$

This means that the μ_i's are (not necessarily strictly) ordered with respect to '$\leq$'. If H_0 is rejected, then we conclude due to our prior knowledge that a global trend over all included $k+1$ groups indeed exists. Without loss of generality we limit the analysis on increasing dose-response functions. Situations with decreasing responses are reduced to above constellations by reverting the signs of the data. Furthermore, we consider only one-sided alternatives throughout this thesis. Generalisations to two-sided cases are possible and mostly straight forward.

The following important remark seems to be appropriate at this stage. It should be noted that the incorporation of the two assumptions

- comparison to the control, only,
- monotone restriction of the alternative,

must be seen in the context of searching for better tests in terms of power. The inclusion of prior information as done above leads to more powerful tests in comparison to those which do not take the ordering of the means into account. Tests for trend are therefore recommended here. However, caution is advisable, if the practitioner is not sure, whether this kind of underlying dose-response shape really holds. Bauer (1997) showed that already small departures from the assumed alternative may lead to inappropriate results if trend tests, such as those presented below, are used. They are then not useful in the sense that they do not control the probability of incorrectly declaring a dose to be effective when in fact it is not effective. Irrespective which trend test is going to be conducted, the decision for its use should always be done under this aspect and the context of the application should be analysed carefully before looking at the data. Generalisations to situations, in which a possible downturn at high doses can not be excluded a-priori, are handled by Simpson and Margolin (1990) and Pan and Wolfe (1996) among others, but will not be analysed here.

1.2. Maximum likelihood estimators under total order restriction

The problem is to investigate independent random samples from $k+1$ normal populations with means $\mu_0, \mu_1, \ldots, \mu_k$ and common variance σ^2. Recall the null hypothesis (1.1) of no effect and that we restricted the alternative to (1.2) for our applications. The aim is now to derive the maximum likelihood estimator (MLE) of the population mean vector $\mu = (\mu_0, \mu_1, \ldots, \mu_k)$ under both hypotheses. This was first done by Brunk (1955) under rather general aspects. The description here, however, follows more closely the representation of Robertson et al. (1988, p. 6). The derivations presented below are fundamental for the tests of Williams (1971), Marcus (1976) and Bartholomew (1961), as all three use these estimates in their proposed statistics.

First note that the corresponding likelihood function is given by

$$L\left(\boldsymbol{X}_0,\, \boldsymbol{X}_1,\, \ldots,\, \boldsymbol{X}_k \,|\, \mu,\, \sigma\right) = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^N} \cdot \exp\left\{-\frac{1}{2\sigma}\sum_{i=0}^{k}\sum_{j=1}^{n_i}\left(X_{ij}-\mu_i\right)^2\right\},$$

where $N = \sum_i n_i$ is the total sample size and $\boldsymbol{X}_i = \left(X_{i1}, \ldots, X_{in_i}\right)$ the data vector of the $i^{\underline{\text{th}}}$ group. To obtain the MLE under H_0 take the logarithm of L

$$\log L = -\frac{N}{2}\log(2\pi) - N\log(\sigma) - \frac{1}{2\sigma}\sum_{i=0}^{k}\sum_{j=1}^{n_i}\left(X_{ij}-\mu_i\right)^2 \tag{1.3}$$

and set its partial derivative $\frac{\partial(\log L)}{\partial\mu} = 0$ (notice that under H_0 $\mu_0 = \ldots = \mu_k =: \mu$). This yields the well known result

$$\sum_{i=0}^{k}\sum_{j=1}^{n_i}\left(X_{ij}-\mu\right) = \sum_{i=0}^{k}\left(X_{i1}-\mu + X_{i2}-\mu + \ldots + X_{in_i}-\mu\right) = \sum_{i=0}^{k} n_i\left(\overline{X}_i-\mu\right) = 0.$$

This means that the vector $\overline{\boldsymbol{X}} = \left(\overline{X}_0,\, \overline{X}_1, \ldots,\, \overline{X}_k\right)$ is the unrestricted MLE of μ.

We now direct the attention towards the MLE under the restricted alternative. Looking at the log-likelihood function (1.3) above we notice that the MLE subject to $\mu_0 \le \mu_1 \le \ldots \le \mu_k$ is given by the minimisation of

$$\begin{aligned}
&\sum_{i=0}^{k}\sum_{j=1}^{n_i}\left(X_{ij}-\mu_i\right)^2 = \\
&= \sum_{i=0}^{k}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_i+\overline{X}_i-\mu_i\right)^2 = \\
&= \sum_{i=0}^{k}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_i\right)^2 + 2\sum_{i=0}^{k}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_i\right)\left(\overline{X}_i-\mu_i\right) + \sum_{i=0}^{k}\sum_{j=1}^{n_i}\left(\overline{X}_i-\mu_i\right)^2 = \\
&= \sum_{i=0}^{k}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_i\right)^2 + 2\sum_{i=0}^{k}\left\{\left(X_{i1}-\overline{X}_i\right)\left(\overline{X}_i-\mu_i\right) + \ldots + \left(X_{in_i}-\overline{X}_i\right)\left(\overline{X}_i-\mu_i\right)\right\} + \sum_{i=0}^{k} n_i\left(\overline{X}_i-\mu_i\right)^2 = \\
&= \sum_{i=0}^{k}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_i\right)^2 + 2\sum_{i=0}^{k}\underbrace{\left(X_{i1}+\ldots+X_{in_i}-n_i\overline{X}_i\right)}_{=0}\left(\overline{X}_i-\mu_i\right) + \sum_{i=0}^{k} n_i\left(\overline{X}_i-\mu_i\right)^2 = \\
&= \sum_{i=0}^{k}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_i\right)^2 + \sum_{i=0}^{k} n_i\left(\overline{X}_i-\mu_i\right)^2.
\end{aligned}$$

Because in the last equation the first term does not depend on μ the restricted MLE $\hat{\mu} = (\hat{\mu}_0, \hat{\mu}_1, \ldots, \hat{\mu}_k)$ minimises

$$\sum_{i=0}^{k} n_i \left(\overline{X}_i - \mu_i \right)^2 \tag{1.4}$$

with respect to $\mu_0 \leq \mu_1 \leq \ldots \leq \mu_k$. Before we continue searching for an explicit representation of $\hat{\mu}$ we give the following definitions.

Definition 1.1.: Let $X = \{x_0, x_1, \ldots, x_k\}$ be a finite set with the simple (or total) order $x_0 \leq x_1 \leq \ldots \leq x_k$. A function f on X is called *isotonic* subject to the given ordering if $f(x_0) \leq f(x_1) \leq \ldots \leq f(x_k)$.

Definition 1.2.: Let g be defined on a finite set $X = \{x_0, x_1, \ldots, x_k\}$. A function g^* on X is called an *isotonic regression* of g with weights $\boldsymbol{w} = (w_0, w_1, \ldots, w_k)$ subject to the simple ordering $x_0 \leq x_1 \leq \ldots \leq x_k$ under the $L_2 - \text{norm}$, if g^* is isotonic and minimises $\sum_{x \in X} [g(x) - f(x)]^2 w(x)$ in the class of all isotonic functions f on X.

Lemma 1.1.: With above notations the restricted MLE of $k+1$ normal means with respect to $\mu_0 \leq \mu_1 \leq \ldots \leq \mu_k$ is given by the isotonic regression $\hat{\mu}$ of $\overline{\boldsymbol{X}} = (\overline{X}_0, \overline{X}_1, \ldots, \overline{X}_k)$ and weights $(n_0, n_1, \ldots, n_k)$.

Proof: Define in above derivation the $k+1$ treatment groups by $D_0, D_1, \ldots, D_k$ and $X = \{D_0, D_1, \ldots, D_k\}$. Let further $g(D_i) = \overline{X}_i$ and $w(D_i) = n_i$ for $i = 0, 1, \ldots, k$. The assertion follows directly for $g^* = \hat{\mu}$ from Definition 1.2. and above calculations

We therefore have managed to reduce the calculation of the MLE for ordered means to the problem of solving the minimisation problem (1.4). With Lemma 1.1. we conclude further that this is equivalent to determine the isotonic regression in the sense of Definition 1.2. Starting with this intermediate result we proceed forward and make use of various algorithms available for the computation of g^*, i.e. $\hat{\mu}$.

The *pool-adjacent-violator algorithm* (PAVA) according to Ayer et al. (1955) is the most widely used algorithm to compute the isotonic regression. In the context of searching for restricted MLEs the process can be described as follows. First look, whether $\overline{X}_0 \leq \overline{X}_1 \leq \ldots \leq \overline{X}_k$. If it is the case then set $\hat{\mu}_i = \overline{X}_i,\ i = 0, 1, \ldots, k,$ and the procedure finishes with $\hat{\mu} = \left(\overline{X}_0, \ldots, \overline{X}_k\right)$ being the sought restricted MLE. Otherwise there is at least one *i*, so that $\overline{X}_{i-1} > \overline{X}_i$. Replace $\overline{X}_{i-1}$ and $\overline{X}_i$ by the single mean

$$\overline{X}_{i-1,i} = \frac{n_{i-1}\overline{X}_{i-1} + n_i\overline{X}_i}{n_{i-1} + n_i}.$$

The new series is therefore reduced to the *k* means

$$\overline{X}_0,\ \overline{X}_1, \ldots,\ \overline{X}_{i-2},\ \overline{X}_{i-1,i},\ \overline{X}_{i+1}, \ldots,\ \overline{X}_k.$$

From now on repeat these steps by treating $\overline{X}_{i-1,i}$ as a single mean with corresponding weight $w_{i-1,i} = n_{i-1} + n_i$ until the remaining amalgamated means are completely ordered. At the last stage the mean $\overline{X}_{i-j,\ldots,i-1,i}$ is replaced by $j+1$ means $\hat{\mu}_{i-j}, \ldots, \hat{\mu}_{i-1}, \hat{\mu}_i$ with the same value as the amalgamated mean. This provides the restricted MLE, which consists of the final vector of $k+1$ means $\hat{\mu} = (\hat{\mu}_0, \hat{\mu}_1, \ldots, \hat{\mu}_k)$.

With above algorithm we are finally able to calculate fairly simple the restricted MLE with respect to the total order. The implementation of the algorithm is straight forward and the computations conducted quickly. Though we still have no closed formula for the $\hat{\mu}_i$'s yet. The next lemma solves this disadvantage by going one step further. It is based on the max-min formulas for isotonic regression, compare e.g. Robertson et al. (1988, p. 23). But the equivalence between the following analytical expression and the PAVA described so far can be seen directly when writing the maximum and minimum terms out in full. With this last link we are now able to state

Lemma 1.2.: For given weights $n_0, n_1, \ldots, n_k$ and normal means $\mu_0, \mu_1, \ldots, \mu_k$ the maximum likelihood estimates $\hat{\mu}_i$ subject to the simple order restriction (1.2) are given by

$$\hat{\mu}_i = \max_{0 \le u \le i} \min_{i \le v \le k} \frac{\sum_{j=u}^{v} n_j \overline{X}_j}{\sum_{j=u}^{v} n_j}, \tag{1.5}$$

where $\overline{X}_i = \sum_j X_{ij} \big/ n_i$ are the sample means for $i = 0, 1, \ldots, k$.

A purely analytical proof of the fundamental theorem that the amalgamated means are the solution to the isotonic regression problem is given in Cheng (1995). In the Appendix a SAS implementation of these max-min formulas is provided. The rather theoretical results obtained so far are best illustrated by applying Lemma 1.2. and the PAVA to an example.

Example 1.1.: Barlow et al. (1972, p. 18) report the number $g(x_i)$ of days to freezing for Lake Mendota/USA, which were collected to study local environmental influences. Note the missing randomisation of the underlying experiment and that the example is therefore inadequate for statistical purposes. Nevertheless we use this data set as an appropriate example to explain the principle of restricted MLEs. The measurements were done each

Year	1855	1856	1857	1858	1859	1860	1861	1862	1863	1864	1865	1866
x_i	1	2	3	4	5	6	7	8	9	10	11	12
$g(x_i)$	25	13	2	15	14	21	9	33	25	15	21	25
Step 1	⏟			⏟		⏟		⏟			⏟	⏟
Weight	3			2		2		3			1	1
Mean	40/3 = 13.33			29/2 = 14.5		30/2 = 15		73/3 = 24.33			21	25
Step2	⏟			⏟		⏟		⏟				⏟
Weight	3			2		2		4				1
Mean	13.33			14.5		15		94/4 = 23.5				25

Table 1.1. Days to freezing for Lake Mendota/USA.

year from 23 November on for a total period of 111 years beginning in 1855. As considering the data for the purpose of illustration only we restrict the evaluation to the first 12 years, given in Table 1.1. The variable x_i stands for the corresponding measured year $1854 + x_i$. Because the investigations were conducted to detect a possible warming trend over the years, the simple order is defined as increasing values for $g(x_i)$ with progressing years. Having a first look on the data one immediately notices that they are not completely ordered. We assign each year the starting weight $w_i = 1,\ i = 1, 2, \ldots, 12$. Already for the first two years $g(x_1) = 25 > 13 = g(x_2)$, i.e. the ordering is violated. We therefore replace both by the average value $g(x_{1,2}) = \frac{1 \cdot 25 + 1 \cdot 13}{1+1} = 19$ and assign it the new weight $w_{1,2} = 1 + 1 = 2$. As $g(x_{1,2}) = 19 > 2 = g(x_3)$, the monotonicity is violated again and we replace both means by $g(x_{1,2,3}) = \frac{2 \cdot 19 + 1 \cdot 2}{2+1} = 13.33$. After pooling every decreasing sub-sequence the middle part of Table 1.1. is yielded. It contains all amalgamated means at this first stage and the corresponding weights. In the second pass through the data we compare the remaining means and pool them, if necessary. We note, for example, that $g(x_{8,9,10}) = 24.33 > 21 = g(x_{11})$ and therefore violates the simple ordering. Averaging both yields the new mean $g(x_{8,9,10,11}) = 23.5$ with the total weight $w_{8,9,10,11} = 3 + 1 = 4$ and we replace the preceding means by the single new calculated one. The lower part of Table 1.1. illustrates this last step and presents the completely ordered amalgamated means. Finally, the values in the original data set are replaced by the new means according to their weights to obtain the restricted MLE $\hat{\mu}$ = (13.33, 13.33, 13.33, 14.5, 14.5, 15, 15, 23.5, 23.5, 23.5, 23.5, 25) for this example.

Before we leave this section the reader is referred to the books of Barlow et al. (1972, Chapter 1 & 2) and Robertson et al. (1988, Chapter 1) for a deeper approach and understanding of this subject. They contain not only the missing proofs omitted here, but they introduce the isotonic regression from a generalised point of view. As a matter of fact, we keep focusing on the simple order, as this was the starting point of our considerations and remains being our main purpose.

1.3. The trend tests according to Williams, Marcus, Bartholomew and multiple contrast tests

1.3.1. Williams' $\bar{t}$ – test

The starting point of all considerations in this thesis is the trend test according to Williams (1971, 1972). In his first paper, Williams introduced the new test statistic for normally distributed data in the balanced set-up and provided critical values (upper 1% and 5% points) for different number of treatment groups and varying degrees of freedom. In the follow-up paper Williams generalised his test to situations of unequal replications and discussed their optimal allocation for a fixed total number of experimental units.

The test statistic is given in the general set-up by the pairwise *t*-type statistic

$$\bar{t}_{k,\nu} = \frac{\hat{\mu}_k - \overline{X}_0}{s\sqrt{\frac{1}{n_k} + \frac{1}{n_0}}} \tag{1.6}$$

and is easily implemented numerically. Here, s^2 and ν are the usual variance estimator and degrees of freedom given in Section 1.1. The MLE $\hat{\mu}_k$ is obtained by using Lemma 1.2. or applying the PAVA from Section 1.2., but with the difference of excluding the control group from the amalgamation process. Note that in his first paper, Williams included $\overline{X}_0$ both in the description of the procedure and in the given numerical example. But when deriving the null distribution in order to determine the critical values, the control group was excluded. Similarly, $\overline{X}_0$ is omitted throughout in his follow-up paper (1972). Tamhane et al. (1996) noticed to this topic:

"*Actually, both ways of calculating the isotonic estimates lead to identical estimates* i* *of the MED as well as identical* $\bar{t}$ *– tests for testing it. Hence, it does not matter which way they are calculated.*"

Based on theses facts we therefore continue by excluding $\overline{X}_0$ from the pooling procedure. However, one main disadvantage of the $\bar{t}-\text{test}$ is the arising null distribution, which is difficult to compute, especially for unequal replications. For the balanced case upper critical points for several combinations of α, k and ν are tabulated (e.g. Williams, 1971). For the partial balanced case, i.e. equal number n of observations in the treatment groups ($n_0 \neq n_1 = n_2 = \ldots = n_k =: n$), Williams (1972) derived an empirical approximation, based on the values of the balanced set-up,

$$\bar{t}_{k,\nu}(w) = \bar{t}_{k,\nu}(1) - 10^{-2}\beta\left(1-\tfrac{1}{w}\right). \tag{1.7}$$

Here, $w = n_0/n$ denotes the ratio between the number of replicates in the control group and the remaining groups. The factor β is extrapolated from accurate values and depends on k and ν. Finally, for $w = 1$ the relationship $\bar{t}_{k,\nu}(1) = \bar{t}_{k,\nu}$ leads to the balanced quantiles. In the meantime, Williams' procedure is available in SAS (SAS Institute Inc., 1997, p. 987) for the balanced case by the call

```
PROBMC('Williams',quantile,probability,ν,k);
```  (1.8)

In this statement either 'quantile' or 'probability' has to be defined, while the other value has to be set as missing '.'. Up to $k = 15$ both quantiles and p-values are calculated fast and accurately, but for higher dimensions the computation is very expensive and slow. For the general case of unequal sample sizes, however, the evaluation of Williams' distribution still seems to be a challenging task, as no algorithm for its computation is available.

Several approaches for generalising Williams' original statistic to other situations have been published in the literature and a small overview is given in the following. For the general non-parametric set-up Shirley (1977) extended the $\bar{t}-\text{procedure}$ via ranking over all groups using the asymptotic version of the original test (infinite degrees of freedom). Next Williams (1986) suggested a slight modification of her method in order to improve the power (using the subset-ranking method instead of the k-ranking). House (1986) provided a non-parametric version for randomised block designs based on Friedman-type ranks.

For the dichotomous case Williams (1988) himself extended his procedure and proposed a conditional exact test based on the multivariate hypergeometric distribution under the null hypothesis. Mount (1999) modified the $\bar{t}-$statistic for binomial parameters for comparing two doses to a control (k = 2). He derived an asymptotic distribution close to, but not the same as, a standard normal distribution.

A robustness study was carried out by Hothorn (1989) on the behaviour of both Williams' and Shirley's tests under violation of the normality assumption, variance heterogeneity, non-monotonous dose-response shapes and unequal group sizes. Tsai and Chen (1995) proposed a robustified statistic by using robust estimates (M- and trimmed estimators) instead of the arithmetic means $\overline{X}_i$. The new procedures are supposed to be robust against outliers and deviations from normality.

1.3.2. Marcus' $\bar{t}^{\,\mathrm{mod}}$ – test

Williams (1971) already proposed a modified version of (1.6) by replacing $\overline{X}_0$ by $\hat{\mu}_0$, where $\hat{\mu}_0$ is obtained by using Lemma 1.2. Marcus (1976) succeeded then in deriving the exact null distribution of the new statistic

$$\bar{t}^{\,\mathrm{mod}}_{k,\nu} = \frac{\hat{\mu}_k - \hat{\mu}_0}{s\sqrt{\frac{1}{n_k}+\frac{1}{n_0}}}. \qquad (1.9)$$

But as "*... the computation of exact* α*–quantiles ... requires* k*-variate numerical integration ... we computed only the 5% and 1% quantiles for* k *= 3 and 4.*" The derivation of an algorithm for calculating quantiles or p-values for general k has not been solved until Hayter et al. (1999a). They managed to decompose the involved k–variate integral into a series of nested lower order integrals by using a Markov property of the arising random variables. Recursive integration techniques can then be applied.

From now on we will call $\bar{t}^{\,\mathrm{mod}}$ Marcus' test. In the literature it is also referred to as the modified Williams or the isotonic range statistic. Critical values of this statistic are given in Hayter et al. (1999a) for several constellations of α and degrees of freedom. Already Marcus (1976) conducted a power simulation study and compared both $\bar{t}-$ and $\bar{t}^{\,\mathrm{mod}}$–procedures for

several parameter configurations. She found out that $\bar{t}$ has a higher power for dose-response shapes of the type $\mu_0 < \mu_1 = \ldots = \mu_k$ (concave profiles), whereas $\bar{t}^{\,\mathrm{mod}}$ is better for $\mu_0 = \mu_1 = \ldots \mu_{k-1} < \mu_k$ (convex profiles). Overall her data suggest that on average $\bar{t}^{\,\mathrm{mod}}$ is slightly better than $\bar{t}$. Cohen and Sackrowitz (1992) showed that Marcus' method is inadmissible and proposed a better test for $k = 2$ by using the total sum of errors instead of s^2.

1.3.3. Likelihood ratio test

The likelihood ratio test (LRT) for homogeneity of normal means under total order restriction was first introduced by Bartholomew (1959). With the variance σ^2 known, Bartholomew's test statistic is $\bar{\chi}_k^2 = \sum_{i=0}^{k} n_i \left(\hat{\mu}_i - \overline{X}\right)^2 \Big/ \sigma^2$, where $\overline{X} = \sum_{i=0}^{k} n_i \overline{X}_i \Big/ N$ is the overall mean estimator and $\hat{\mu}_i$ are the MLEs according to Lemma 1.2. In the following we will focus us, however, on the in practice more important case of an unknown common variance, estimated by the mean square error s^2 with $\nu = N - k - 1$ degrees of freedom. Bartholomew then showed that the LRT is based upon the statistic

$$\bar{E}_k^2 = \frac{\sum_{i=0}^{k} n_i \left(\hat{\mu}_i - \overline{X}\right)^2}{\sum_{i=0}^{k} \sum_{j=1}^{n_i} \left(X_{ij} - \overline{X}\right)^2} = \frac{\sum_{i=0}^{k} n_i \left(\hat{\mu}_i - \overline{X}\right)^2}{\sum_{i=0}^{k} n_i \left(\hat{\mu}_i - \overline{X}\right)^2 + \sum_{i=0}^{k} n_i \left(\hat{\mu}_i - \overline{X}_i\right)^2 + \nu s^2}. \qquad (1.10)$$

The $\bar{E}_k^2$ involves the ratio of the between groups sum of squares $\sum_{i=0}^{k} n_i \left(\hat{\mu}_i - \overline{X}\right)^2$ after amalgamation and the total sum of squares. It can therefore be interpreted as an ANOVA–*F*–test analogue under total order restriction. Bartholomew (1961) already succeeded in deriving the null distribution of $\bar{E}_k^2$. But quoting the main results we will again follow the representation of Robertson et al. (1988, pp. 68).

Recalling Definition 1.2. of an isotonic regression we first introduce the level probabilities which repeatedly arise later on. We call those subsets, where the quantities arising in (1.5) are constant, *level sets*.

Definition 1.3.: Let $\mu \in H_0$ and $\boldsymbol{w} = (n_0, n_1, \ldots, n_k)$. Further, let $Y_0, Y_1, \ldots, Y_k$ be independent random variables, $Y_i \sim N(\mu_i, \sigma^2)$. Further on, we set M the number of level sets in $\boldsymbol{Y}^*$, the isotonic regression of $\boldsymbol{Y} = (Y_0, Y_1, \ldots, Y_k)$. Then we call the quantities

$$P(l, k+1; \boldsymbol{w}) = P(M = l), \qquad l = 1, \ldots, k+1,$$

level probabilities. The level probability $P(l, k+1; \boldsymbol{w})$ is therefore the probability that the isotonic regression function $\boldsymbol{Y}^*$ takes exactly l distinct values. By definition it follows that $\sum_{l=1}^{k+1} P(l, k+1; \boldsymbol{w}) = 1$.

The following fundamental lemma shows that the null distribution of $\bar{E}_k^2$ can be stated as a weighted sum of F–probabilities.

Lemma 1.3.: Let $\mu \in H_0$ and $c \in \mathbb{R}$. Then

$$P(\bar{E}_k^2 \geq c) = \sum_{l=1}^{k+1} P(l, k+1; \boldsymbol{w}) P\left(B_{\frac{l-1}{2}, \frac{N-l}{2}} \geq c\right) = \sum_{l=2}^{k+1} P(l, k+1; \boldsymbol{w}) P\left(F_{l-1, N-l} \geq \frac{N-l}{l-1} \frac{c}{1-c}\right),$$

where F_{ν_1, ν_2} is a random variable following a F–distribution with ν_1 and ν_2 degrees of freedom and $B_{a,b}$ is a beta–variable with parameters a and b.

Proof: See for example Robertson et al. (1988, pp. 70).

With Lemma 1.3. we have the general form of the null distribution of $\bar{E}_k^2$, but to make use of this result the values of the level probabilities $P(l, k+1; \boldsymbol{w})$ have to be obtained. These probabilities involve the evaluation of multidimensional integrals. The arising numerical difficulties are one main reason for the restricted use of the LRT throughout the literature. In fact, there are several possibilities in calculating these integrals, but we postpone their representation to Chapter 2. A SAS/IML program is presented there, which computes for arbitrary weights $\boldsymbol{w}$ the required values in few seconds at an accuracy of more than 10^{-7} up to $k = 11$.

In the passages above we have introduced the statistic of the LRT and have given a basic notion of its null distribution. As already mentioned in the Introduction, the LRT is supposed to have good 'average' power throughout the alternative space H_A. This has been shown by several power simulation studies (see for example Marcus, 1976 and Turnbull et al., 1987). However, the LRT has been regarded for a long time as difficult to implement and therefore a great variety of simplifying approximations to the LRT exists. Leaving the concrete numerical evaluation for Chapter 2 we finish this subsection with some recent developments in the literature on the LRT. For a broad overview up to 1988 the reader is referred to Robertson et al. (1988, Chapter 3).

It can be shown that the alternative parameter space of the classical LRT is a pointed polyhedral cone and that the null hypothesis is a linear subspace contained in the boundary of the cone. Making use of an idea dating back to Pincus (1975), Akkerboom (1990) and Conaway et al. (1991) independently used a 'circular likelihood ratio test' (CLRT) and found that the power of the CLRT was close to that of the classical LRT. Moreover, because of its simpler geometric nature, its use is supposed to be easier to handle in both balanced and unbalanced situations. Recently, Tang and Lin (1997) developed an 'approximate likelihood ratio test' (ALR), which is based on an orthant alternative cone and has according to their results good power properties as well. Hu (1998) presented an exact algorithm for projecting a vector onto a polyhedral cone in relatively low dimensions. Finally, Wright (1988) introduced a modified likelihood ratio test (MLRT) by using the usual mean square error instead of the total variance. He derived the null distribution of the new test (which is similar to the original LRT) and showed the asymptotic equivalence between them. A simulation study suggested that the MLRT is more robust against violations of the hypothesised orderings.

1.3.4. Multiple contrast test

The concept of multiple contrast tests (MCTs) will be very important in the course of this thesis. Therefore much attention is given to its introduction. It was first described by Mukerjee et al. (1986, 1987). Older articles exist which mention or deal with MCTs, but none of them introduced them thoroughly (see for example Dwass, 1960, Dunn and Massey, 1965, Knoke, 1976, and Mehta et al., 1984). The main reason in developing such a new test was to have a test with similar power behaviour as the LRT though still easy to use.

From the geometric starting point, which led Mukerjee et al. (1987) to the new test, the MCT "*... is associated with a set of vectors that are 'strategically' located within the alternative region.*" The aim of this approach is therefore to cover most parts of the alternative space by choosing some selected vectors within this space and conduct the MCT with respect to this grid. However, we leave these geometrical considerations and introduce the statistic rather analytically. Recalling the notation of Subsection 1.1. we test the null hypothesis (1.1) of no difference by defining the standardised statistic of a single contrast test (SCT) as

$$T^{SC} = \frac{\sum_{i=0}^{k} c_i \overline{X}_i}{s\sqrt{\sum_{i=0}^{k} \frac{c_i^2}{n_i}}} \sim t_\nu. \tag{1.11}$$

Formulating the statistic T^{SC} as a quotient of a standard normal variable and an independent chi variable with parameter ν, it follows by definition that T^{SC} is univariate central *t*-distributed with ν degrees of freedom. The weights c_i denote the contrast coefficients under the sub-condition $\sum_i c_i = 0$. Besides this limitation, the choice of the c_i's is free and numerous proposals concerning their (optimal?) choice have been published. Nevertheless, this problem has not been solved satisfactorily in the literature and is still an open question of research. We leave this issue for Chapter 3, where a detailed review and discussion follows. Instead, we illustrate the consequences, when choosing a poor set of contrast coefficients and no prior information on the underlying dose-response shape is available.

Example 1.2.: Suppose that we compare $k = 3$ doses of a compound to a negative control. Further on we investigate the two contrast vectors $\boldsymbol{c}_1 = (-1, -1, -1, 3)$ and $\boldsymbol{c}_2 = (-3, 1, 1, 1)$. We analyse the power of the resulting SCTs T_1^{SC} and T_2^{SC} for two different dose-response shapes: $(0, \delta, \delta, \delta)$ and $(0, 0, 0, \delta)$, where δ denotes the shift parameter. In the first case (concave profile) the lowest dose has already an effect of size δ in comparison to *C*–, whereas the remaining doses have no additional influence. In the other case (convex profile) only the highest dose has an effect (of size δ) with respect to *C*–, but the two low doses have no increased effect at all.

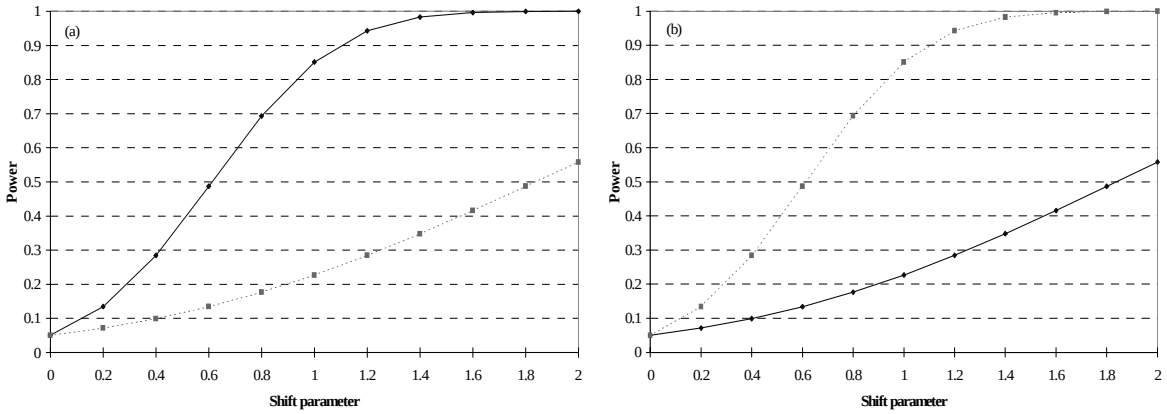


Figure 1.1. Power comparison of T_1^{SC} (dotted line) and T_2^{SC} (solid line), balanced case with sample size allocation (10, 10, 10, 10), $\alpha = 0.05$, $k = 3$ for (a) $\mu = (0, \delta, \delta, \delta)$ and (b) $\mu = (0, 0, 0, \delta)$.

The power function follows a non-central univariate *t*–distribution. For the representation of the noncentrality parameter we refer to Chapter 4. From Figure 1.1. it becomes clear, how much SCTs may depend in terms of power on the underlying dose-response shape. The effect of the contrast coefficients in T_1^{SC} is that they pool the lower treatment groups and compare the resulting average value with that of the highest dose. This is meaningful when the effects of the pooled treatments are similar and therefore T_1^{SC} behaves well for convex profiles. For concave shapes, however, the pooling of groups with different effect sizes has a negative influence on the test statistic and therefore the power decreases markedly, resulting in a loss up to 60%. Similar arguments hold for T_2^{SC}, too.

From Example 1.2. the strong shape-dependence of SCTs becomes evident. The crucial point now is that these shapes are in general unknown a-priori – a situation of frequent occurrence in real data examples. Ignoring this important fact is common practice but can not be accepted. The problem of a-priori unknown shapes even increases in at least two cases:

- testing sub-hypotheses by using the closure principle (see Chapter 6 for an application);
- investigation of stratified designs because of varying strata specific shapes.

It seems highly unreasonable to assume that the same dose-response shape holds for all sub-hypotheses, respective for all strata.

The approach of the MCTs overcomes, at least partially, this disadvantage. As pointed out in the quotation above, it seeks to locate several 'grid-vectors', i.e. contrast vectors, as good as possible in the alternative space. The resulting test statistic builds just the maximum over q of such single contrasts defined in (1.11):

$$T^{MC} = \max\left\{T_1^{SC}, \ldots, T_q^{SC}\right\}. \tag{1.12}$$

As the distribution of $\left\{\sum_i c_{1i}\overline{X}_i, \ldots, \sum_i c_{qi}\overline{X}_i\right\}$ is multivariate normal, the joint distribution of the T_i^{SC}'s will by definition (see Section 2.2. for further details) be a central q–variate t–distribution with ν degrees of freedom and correlation matrix $\boldsymbol{R} = \left\{\rho_{l,m}\right\}_{l,m}$, $l, m = 1, \ldots, q$. The entries of $\boldsymbol{R}$ consist of the correlation between each two of the q contrast vectors and are computed according to the following

Lemma 1.4.: For two SCTs T_1^{SC} and T_2^{SC} as defined in (1.11) the correlation $\rho = \mathrm{Corr}\left(T_1^{SC}, T_2^{SC}\right)$ under H_0 is given by

$$\rho = \frac{\sum_{i=0}^{k} c_{1i}c_{2i}/n_i}{\sqrt{\left(\sum_{i=0}^{k} c_{1i}^2/n_i\right)\left(\sum_{i=0}^{k} c_{2i}^2/n_i\right)}}. \tag{1.13}$$

Proof: By Definition 2.3. of the multivariate t–distribution it is sufficient to consider the correlation of the bivariate normal vector $(X, Y) = \left(\sum c_{1i}\overline{X}_i, \sum c_{2i}\overline{X}_i\right)$. Under H_0 we get:

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= E(X - EX)(Y - EY) = \\
&= E\left(\sum_{i=0}^{k} c_{1i}\overline{X}_i - E\left(\sum_{i=0}^{k} c_{1i}\overline{X}_i\right)\right)\left(\sum_{i=0}^{k} c_{2i}\overline{X}_i - E\left(\sum_{i=0}^{k} c_{2i}\overline{X}_i\right)\right) = \\
&= E\left(\sum_{i=0}^{k} c_{1i}\overline{X}_i\right)\left(\sum_{i=0}^{k} c_{2i}\overline{X}_i\right) = \\
&= \sum_{i=0}^{k}\sum_{j=0}^{k} c_{1i}c_{2j}E\left(\overline{X}_i\overline{X}_j\right) = \\
&= \sum_{i=0}^{k} c_{1i}c_{2i}E\left(\overline{X}_i^2\right) + \sum_{\substack{i,j=0 \\ i\neq j}}^{k} c_{1i}c_{2j}E\left(\overline{X}_i\overline{X}_j\right) = \\
&= \sum_{i=0}^{k} c_{1i}c_{2i}\mathrm{Var}\left(\overline{X}_i\right) =
\end{aligned}$$

$$= \sum_{i=0}^{k} c_{1i} c_{2i} \frac{\sigma^2}{n_i}.$$

Because of $\rho = \mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$ above assertion follows directly when taking $\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)} = \sigma^2 \sqrt{\left(\sum_{i=0}^{k} \frac{c_{1i}^2}{n_i}\right)\left(\sum_{i=0}^{k} \frac{c_{2i}^2}{n_i}\right)}$ into account.

Example 1.3.: One example of MCTs is the parametric many-to-one test of Dunnett (1955), already introduced in Subsection 1.1. In this set-up several, say *k*, treatment groups are compared to a control. Dunnett's test statistic takes the maximum over the *k* pairwise *t*–tests treatment versus control. In our notation this leads to the $k \times (k+1)$ contrast matrix

$$\boldsymbol{C} = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k)^t = \begin{pmatrix} -1 & 1 & 0 & 0 & \ldots & 0 \\ -1 & 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots \\ -1 & 0 & 0 & 0 & \ldots & 1 \end{pmatrix}.$$

The matrix contains $q = k$ contrast vectors and each of them contrasts the standard with one treatment group. The correlation between two arbitrary contrasts is according to (1.13) given by

$$\rho_{l,m} = \frac{\frac{1}{n_0}}{\sqrt{\left(\frac{1}{n_0} + \frac{1}{n_l}\right)\left(\frac{1}{n_0} + \frac{1}{n_m}\right)}} = \sqrt{\frac{n_l}{n_0 + n_l}} \sqrt{\frac{n_m}{n_0 + n_m}}, \qquad l, m = 1, \ldots, k. \tag{1.14}$$

Up to now we have introduced the concept of MCTs and gave a brief geometric insight of the test statistic. Furthermore we could easily state the null distribution, but did not discuss the remaining problem yet, how to evaluate the multivariate *t*–distribution numerically. Similarly to the LRT, the null distribution involves the calculation of multidimensional integrals – a challenging task and until very recently not solved in the literature for arbitrary correlation matrices $\boldsymbol{R}$. Again, we leave this numerical issue for Chapter 2, where several computational algorithms are presented together with an overview of the required properties of the multi-*t*-distribution. However, we can not go further without stating the following

Remark 1.1.: To avoid misunderstandings, it is important to notice that for evaluating general MCTs we seek for a computational method to calculate multivariate *t*–probabilities for arbitrary correlation matrices $\boldsymbol{R}$. Frequently, some *k*–sample tests have a multi-*t*-distribution with a very special correlation structure. In these cases the high dimensionality can be reduced to lower order integrals and the evaluation gets much simpler. Dunnett's test of Example 1.3. is such an example. Here, the so-called product correlation structure is valid, i.e. $\exists \lambda_l, \lambda_m: \rho_{l,m} = \lambda_l \lambda_m \ \forall l, m.$ Based on equation (1.14) for Dunnett's test $\lambda_l = \sqrt{\frac{n_l}{n_0 + n_l}}$ is yielded. Unfortunately a similar relationship has not been found yet for arbitrary MCTs and therefore the need for a general computation method, provided in Chapter 2.

Remark 1.2.: Another way of defining contrast test statistics is to include the sample sizes in the numerator

$$\tilde{T}^{SC} = \frac{\sum_{i=0}^{k} n_i c_i \bar{X}_i}{s\sqrt{\sum_{i=0}^{k} n_i c_i^2}}$$

with the contrast ensuring equation $\sum_{i=0}^{k} n_i c_i = 0$ and the c_i's ordered as $c_0 \leq \ldots \leq c_k$. This form is frequently used in the literature, too (see for example Marcus and Peritz, 1976, and Miwa et al., 1999). However, pattern (1.11) turns out to be more flexible for our purposes and we therefore continue using this representation.

An important tool to be used frequently in the course of the thesis is the following property. It states that contrast tests are closed under multiplication by a positive scalar.

Lemma 1.5.: Let $\boldsymbol{c}_1$ and $\boldsymbol{c}_2 = \lambda \boldsymbol{c}_1$ be given contrast vectors, $0 < \lambda \in \mathbb{R}$. Denote by T_1^{SC} and T_2^{SC} the corresponding single contrast tests. Then $T_1^{SC} = T_2^{SC}$ holds.

Proof: The assertion follows directly from $T_2^{SC} = \frac{\sum_{i=0}^{k} c_{2i} \bar{X}_i}{s\sqrt{\sum_{i=0}^{k} c_{2i}^2 / n_i}} = \frac{\lambda \sum_{i=0}^{k} c_{1i} \bar{X}_i}{s\sqrt{\lambda^2 \sum_{i=0}^{k} c_{1i}^2 / n_i}} = T_1^{SC}$.

Multiple contrasts form a very useful class of tests, which covers many different test statistics in the *k*–sample situation. Somerville (1997, 1999) provides a list of several multiple comparison procedures (not necessarily designed for order restricted testing), which can be formulated as MCTs. Among other tests we quote the many-to-one approach of Dunnett (1955, 1964), Tukey's (1953) all-pair comparison and Hsu's (1984) multiple comparison with the best. Moreover, as we are going to see, all of the trend tests, which are reviewed briefly in the subsequent Section 1.4., can be written as MCTs, too.

Recall from the Introduction that the main purpose of the present thesis is the development of Williams' test to unbalanced and other situations. One way to do this is to try to define an appropriate contrast definition and to use the theoretical and numerical results regarding the multivariate *t*–distribution. This is done in Chapter 3, together with a generalisation of Marcus' test to multiple contrasts and a new proposed contrast definition, which bases rather on analytical than empirical reasons. Even the LRT presented in Subsection 1.3.3. can be regarded as a MCT according to Robertson et al. (1988, p. 189):

"*...it can be shown that the LRT statistic may be expressed as the maximum of an infinite number of contrast statistics.*"

This has first been shown in the case of known variances by Marcus and Peritz (1976). Miwa et al. (1999) stated the equivalence between the MLRT of Wright (1988) and the maximum over all ordered contrasts stated in Remark 1.2. Another view on the relationship between the LRT and contrast tests has first been pointed out by Hogg (1965):

Lemma 1.6.: Let $\hat{\mu}_i$ be the amalgamated means according to equation (1.5), $\overline{X}$ the overall mean and the variance σ^2 known. Then the adaptive single contrast test with coefficients $c_i = \hat{\mu}_i - \overline{X}$ is the same as $\overline{\chi}_k^2$, where $i = 0, 1, \ldots, k$.

Proof: Because of $\sum_{i=0}^{k} n_i \hat{\mu}_i \Big/ N = \sum_{i=0}^{k} n_i \overline{X}_i \Big/ N = \overline{X}$ we have by replacing appropriately

$$\sum_{i=0}^{k} n_i \left(\hat{\mu}_i - \overline{X}\right) \overline{X}_i = \sum_{i=0}^{k} n_i \hat{\mu}_i^2 - N\overline{X}^2 = \sum_{i=0}^{k} n_i \left(\hat{\mu}_i - \overline{X}\right)^2 = \overline{\chi}_k^2 .$$

Finalising, the importance of MCTs can not be underestimated in the context of multiple comparisons. They form a certain unifying class of tests, where many multiple tests (and most of the frequently used one) are contained. We quote again Robertson et al. (1988, p. 189), who wrote with special focus on order restricted testing:

"*While some of the* ad hoc *tests in the literature are such multiple contrast tests, they do not seem to have been developed from this point of view. The question of which MCTs are optimal is an important and challenging open problem in order restricted inference.*"

1.3.5 Example

To illustrate the above presented methods for detecting trends we analyse the first example given in the Introduction. We apply the approaches of Williams, Marcus and the LRT. In contrast, we do not cover the MCTs here but leave them for Chapter 3, where a detailed discussion concerning an appropriate choice of contrast coefficients follows. The same example will be caught up there again and the obtained results there will be compared to those given below.

Example 1.4.: Remember the study of Banno and Yamagami (1989), who investigated the conversion efficiency of ingested food (E.C.I.) of *Eupromus ruber* larvae. The descriptive statistics were given in the Introduction and we want to investigate, whether the E.C.I. decreased monotonous with respect to the development stages. From the quantities given before we calculate the mean square error $s^2 = 0.578$ and $\nu = 82$ degrees of freedom. Variance homogeneity is supposed to hold, though the data show some departure from this assumption. Moreover, we also regard the independence condition as given for we assume that no larva was investigated twice in different development stages.

Williams' test

Because the MLEs $\hat{\mu}_i$ are the same as the arithmetic means $\overline{X}_i$ the test statistic is straight forward to compute:

$$\bar{t}_{5,82} = \frac{\hat{\mu}_k - \overline{X}_0}{s\sqrt{\frac{1}{n_k} + \frac{1}{n_0}}} = \frac{2.415 - 1.669}{0.76\sqrt{\frac{1}{21} + \frac{1}{4}}} = 1.799.$$

The problem arises upon an interpretation of this quantity. Williams (1971) provided critical points for the balanced case only, as the SAS-call (1.8) does. As no approach for calculating p-values in unbalanced settings seems to exist, two working solutions might be adequate. First one can use the null distribution from the balanced set-up in the hope that the resulting error gets not too large. Applying the call (1.8) one computes $p = 0.0479$, signalising a weak significant trend. One would reject the null hypothesis of no (decreasing) effect. On the other hand, one can try to simulate the p-value. In order to estimate the corresponding p-value a SAS-program is provided in the Appendix, which simulates under the null hypothesis the distribution of Williams' test statistic and determines the fractile of the observed test statistic among the ordered simulated ones. Applying this program on the present data set, we get for 9999 simulation runs the p-value 0.0373, a little smaller than the 'exact' balanced value. But how far can we trust this value? We calculated the estimates for 50 different seeds and observed that the values varied from 0.0352 to 0.0426. From these results two conclusions can be drawn.

- The approximation of set-ups with unequal replications by use of balanced quantiles or p-values should be avoided. No error bound exists and the error committed can be fairly large. Bretz and Hothorn (1999) have shown in a simulation study that such an approach does not control α, exceeding the pre-assigned 5%–level sometimes by more than 20%.
- The use of simulated quantiles or p-values at least reduces the 95%–coverage interval of the true value to a width of less than 2% (9999 simulation runs). But such a result is still not satisfying. Moreover, the simulation takes a long time and further analysis based on such estimates get rapidly prohibitive in terms of time required.

We now make an experiment by inverting the last two group sizes and keeping the remaining parameters constant, i.e. we analyse the same data set for $\overline{\boldsymbol{X}}$, s^2, ν and the MLE $\hat{\mu}_i$'s unchanged, but with $\boldsymbol{n} = (21, 10, 15, 17, 4, 21)$. Calculating now the statistic we get 3.18 and a corresponding $p = 0.0011$ (balanced set-up) respectively $p = 0.0017$ (simulated unbalanced case). This means that a mere switch of two neighbouring sample sizes leads to totally different results. Starting from a p-value close to 5% we have now a p-value of 0.2% for the same data set. This is due to the $1/n_k$ in the denominator of $\bar{t}_{k,\nu}$. This leads us to the conclusion that the test statistic does not incorporate sufficiently the sample size allocation and therefore the whole approach needs to be improved for general unbalanced set-ups.

Marcus' test

One computes the same value $\bar{t}_{4,25}^{\,\mathrm{mod}} = 1.799$ for the test statistic as in the case of Williams. Using the program B.1.3 we get the p-value 0.0481. Similar to the Williams' program in the unbalanced case, it estimates the p-values by use of a simulation procedure. Again we have only a weak evidence for a trend in the data. Please note that the uncertainty induced by the simulation leads to the problem that with a different seed set a-priori the estimated p-value could lie well above 0.05. This means that the final decision (rejection or not of the null hypothesis) in this real data example would depend directly on the pre-assigned value of the seed. Moreover, the experiment conducted for Williams' test by inverting the sample sizes leads to similar results here.

Likelihood ratio test

Using the alternative representation in (1.10) we have

$$\bar{E}_k^2 = \frac{\sum_{i=0}^{k} n_i\left(\hat{\mu}_i - \bar{X}\right)^2}{\sum_{i=0}^{k} n_i\left(\hat{\mu}_i - \bar{X}\right)^2 + \sum_{i=0}^{k} n_i\left(\hat{\mu}_i - \bar{X}_i\right)^2 + \nu s^2} = \frac{6.6078}{6.6078 + 0 + 82 \cdot 0.578} = 0.1224.$$

The corresponding p-value is computed as 0.00393 and shows a rather strong evidence for a significant trend. When comparing this *p* with the p-values of Williams or Marcus (close to 0.05) a clear difference in terms of power seems to exist. But this is only true in parts, for, as we have seen, Williams' test improves considerably when making a minor switch of group sizes. In this case the p-value around 0.2% lies close to that of the LRT.

Dunnett's many-to-one test

We briefly describe two non-trend tests and compare their behaviour to those mentioned before. According to Example 1.3. we calculate Dunnett's test statistic

$$t_D = \max_{1 \le i \le k} \left\{ \frac{\bar{X}_i - \bar{X}_0}{s\sqrt{\frac{1}{n_i} + \frac{1}{n_0}}} \right\} = \frac{2.411 - 1.669}{0.76\sqrt{\frac{1}{21} + \frac{1}{21}}} = 3.16.$$

The corresponding p-value is 0.0052, which is only a little inferior to that of the LRT or Williams' approach for the modified data set (in which case t_D would not change much). Bretz and Hothorn (1999) compared the power behaviour of t_D and $\bar{t}_{k,\nu}$ for a variety of conditions. They found out that Dunnett's test has less power in most of the cases investigated, but in fact the absolute differences were not too big. All in all t_D behaved well, especially for strong imbalances, as it is the case of the present data example.

F-test

When conducting the usual ANOVA-*F*-test, which is also performed without using the monotonicity assumption, we compute

$$F = \frac{\sum_{i=0}^{k} n_i \left(\overline{X}_i - \overline{X} \right)^2}{k s^2} = 2.286 .$$

The resulting p-value 0.0535 illustrates the possible sharper inferences when taking the prior information of the ordering of the means into account.

1.4. Overview of other trend tests

After the detailed introduction of the four tests mainly used throughout the thesis we shall have a brief review of other trend tests proposed in the literature.

A maximum test based on a complete class characterisation of tests for multidimensional one-sided alternatives (see for example Hirotsu, 1982) has been introduced by Hirotsu (1979, 1997). The approach includes simultaneous tests for the slippage alternatives $H'_{A(j)}$: $\mu_0 = \ldots = \mu_j < \mu_{j+1} = \ldots = \mu_k$, $j = 0, 1, \ldots, k-1$, which form a subregion of our alternative space (1.2). The max t method is given as the maximum component of $\left(\boldsymbol{D}_{k+1}^t \boldsymbol{D}_{k+1} \right)^{-1} \boldsymbol{D}_{k+1}^t \boldsymbol{\Sigma}^{-1} \left[\overline{\boldsymbol{X}} - E_0\left(\overline{\boldsymbol{X}} \right) \right]$, where $\boldsymbol{\Sigma} = \mathrm{diag}\left(s^2 / n_i \right)$ is the non-singular covariance matrix and $\boldsymbol{D}_{k+1}^t = \left(d_{ij} \right)_{ij}$ a $k \times (k+1)$ matrix with $d_{ij} = -1$ if $j - i = 0$, $d_{ij} = 1$ if $j - i = 1$, and 0 otherwise. Considering that

$$\boldsymbol{C} = \left(\boldsymbol{D}_{k+1}^t \boldsymbol{D}_{k+1}\right)^{-1} \boldsymbol{D}_{k+1}^t = \frac{1}{k+1} \begin{pmatrix} k & -1 & -1 & \dots & -1 \\ k-1 & k-1 & -2 & \dots & -2 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & 1 & 1 & \dots & -k \end{pmatrix},$$

we get the immediate representation of the max *t* method as a MCT defined by the contrast matrix ***C***. Hirotsu et al. (1992) succeeded in representing the null distribution as a series of nested lower order integrals by making use of some Markov properties of arising random variables.

Another approach is due to Tukey et al. (1985). It was further successfully used by Capizzi et al. (1992) and Antonello et al. (1993). The method consists of a test on slope in a linearised dose-response regression model. It considers a small set of candidate dose scalings, all of them highly correlated. In the original approach a p-value is computed for each of the proposed scalings using the pooled within error variance s^2. The global p-value is then reported as the minimum p-value of the candidate set. Three different scalings were proposed to define appropriate dose scores: arithmetic (use of actual concentrations $C-$, $D_1, \dots, D_k$), ordinal (equal step scaling 0, 1, ..., *k*) and arithmetic-logarithmic (logarithms of the actual doses). However, Mehta et al. (1984) have already stated the equivalence between this approach and the contrast tests.

Mudholkar and McDermott (1989) decomposed the parameter space into *k* nested problems of testing. The idea is to combine then the associated *k* independent p-values and to provide a test for the overall null hypothesis H_0. Depending on the particular constraints and dose-response shapes, different combination statistics might be used. Basically, the *k* proposed test statistics reduce to *k* SCTs with contrast coefficients of the form

$$c_{ij} = \begin{cases} -1, & j = 0, 1, \dots, i-1 \\ i, & j = i \\ 0, & \text{else} \end{cases}, \quad i = 1, \dots, k.$$

The application of Tippett's minimum method is equivalent to orthogonal contrast tests and the use of Liptak's criterion leads to the SCT proposed by Abelson and Tukey (1963).

McDermott and Mudholkar (1993) generalised their approach to settings with unknown variances. It can be shown that their method approximates the restricted parameter space (a polyhedral cone) by an orthant which is somewhat larger. Recently, McDermott (1999) used these geometrical concepts to improve the original statistic by rotating the orthant so that it has the same center as the original polyhedral cone.

For testing against the simple order alternative (1.2) Hayter (1990) proposed the one-sided studentised range test (OSRT)

$$\max_{0 \le i < j \le k} \frac{\overline{X}_j - \overline{X}_i}{s\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}}$$

in the general unbalanced case. This test statistic is similar to that of the studentised range test of Tukey (1953) and it can be regarded as the maximum over all one-sided studentised pairwise comparisons $\overline{X}_j$ versus $\overline{X}_i$. The OSRT is nothing but a special MCT. For example, for $k = 2$ the contrast test would be defined by

$$\boldsymbol{C} = \left(\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3\right)^t = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}.$$

Hayter (1990) discussed also the development of simultaneous one-sided confidence intervals and the calculation of associated critical points (without using the multivariate *t*-distribution, however, and only for the balanced case up to $k = 8$). Cohen and Sackrowitz (1992, p. 1142) noticed that the OSRT is inadmissible for all α. They proposed a better test for $k = 2$ and showed via a simulation study that the "*amount of improvement is substantial.*"

Recent reviews of trend tests are provided by Shirley (1996) and Chuang-Stein and Agresti (1997). For additional discussion on order restricted inference the reader is referred to Cohen and Sackrowitz (1992, 1993), Silvapulle and Silvapulle (1995), Lee (1996), Miwa and Hayter (1999) and Hayter et al. (1999b).

2. Multivariate normal and *t*–distribution

In this chapter we will give a detailed survey of existing and new approaches to calculate both multivariate normal and *t*–probabilities over rectangular regions. The coming pages will be fundamental for the evaluation of the test statistics considered throughout the thesis, as most of them are based on these multivariate distributions. The chapter is divided into two sections. First we introduce a multivariate generalisation of the normal distribution and give some basic properties required for the remaining thesis. In Subsection 2.1.2. we describe three different approaches of computing the cumulative distribution function. A particular application on calculating the level probabilities required for the LRT is given in Subsection 2.1.3. In the second section we treat the multivariate *t*–distribution similarly. After its definition we derive the density function and focus then on the main problem of evaluating the distribution function. New approaches are given and compared to existing solutions in the literature.

2.1. Multivariate normal distribution

2.1.1. Definition and basic properties

The multivariate normal (mvn) distribution plays a central role in statistical applications and moreover in this thesis. One main application is given through MCTs for dichotomous data. In Chapter 5 we will prove their asymptotic normality and shall use extensively the results obtained in the following. Next, the evaluation of the null distribution of Bartholomew's LRT requires mvn probabilities of a particular form. Further we will see later that the computation of the multivariate *t*–distribution can be reduced mainly to that of the mvn distribution, a fact we are going to use extensively when deriving a power expression for MCTs in Chapter 4. The mvn distribution has been discussed frequently in the literature and we refer to Johnson and Kotz (1972) and Tong (1990) for a detailed discussion. Fang and Zhang (1990) consider the mvn distribution from a generalised view of elliptically countered distributions to which it belongs. The following introduction of the mvn distribution and proofs of the subsequent properties are mainly due to these books.

Definition 2.1.: Let $\boldsymbol{z}$ denote a random vector of dimension n with i.i.d. components $\{Z_i\}$ and $Z_i \sim N(0, 1)$, $i = 1, \ldots, n$. If a $(q \times 1)$–random vector $\boldsymbol{x}$ can be expressed as

$$\boldsymbol{x} \overset{d}{=} \mu + \boldsymbol{C}\boldsymbol{z}, \tag{2.1}$$

where μ is a constant $(q \times 1)$–vector, $\boldsymbol{C}$ a $(q \times n)$–matrix with rank $n \leq q$ and $\boldsymbol{C}\boldsymbol{C}^t = \boldsymbol{\Sigma}$, we write $\boldsymbol{x} \sim N_q(\mu, \boldsymbol{\Sigma})$ and say that $\boldsymbol{x}$ is *distributed according to the* q*–variate normal distribution*. Further, by this definition, we can write $\boldsymbol{z} \sim N_n(\boldsymbol{0}, \boldsymbol{I}_n)$ and say that $\boldsymbol{z}$ is *standard normal distributed of dimension* n, where $\boldsymbol{I}_n$ denotes the n–dimensional unit matrix.

This definition was preferred by Tong (1990, p. 28) because of its convenient use for later purposes. In particular it applies to both singular and non-singular cases according to

Remark 2.1.: The q–dimensional random variable $\boldsymbol{x}$ of Definition 2.1. is said to have a non-singular mvn distribution, if $q = n$ and $|\boldsymbol{\Sigma}| > 0$. Otherwise, $|\boldsymbol{\Sigma}| = 0$ and we call it a singular q–variate normal distribution. (Later we will see that $\boldsymbol{\Sigma}$ is the covariance matrix of $\boldsymbol{x}$ and therefore $\boldsymbol{\Sigma}$ is either positive definite or semi-definite.)

Next we give two important facts.

Lemma 2.1.: Assume $\boldsymbol{x} \sim N_q(\mu, \boldsymbol{\Sigma})$.

a) Suppose $\boldsymbol{y} = \boldsymbol{d} + \boldsymbol{B}\boldsymbol{x}$, where $\boldsymbol{B}$ is a $(l \times q)$–matrix and $\boldsymbol{d}$ a $(l \times 1)$–vector. Then

$$\boldsymbol{y} \sim N_l(\boldsymbol{B}\mu + \boldsymbol{d}, \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^t). \tag{2.2}$$

b) Consider the partition

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix}, \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \tag{2.3}$$

where $\boldsymbol{x}_1$: $n \times 1$, μ_1: $n \times 1$ and $\boldsymbol{\Sigma}_{11}$: $n \times n$, $n < q$. Then

$$\boldsymbol{x}_1 \sim N_n(\mu_1, \boldsymbol{\Sigma}_{11}) \quad \text{and} \quad \boldsymbol{x}_2 \sim N_{q-n}(\mu_2, \boldsymbol{\Sigma}_{22}).$$

Proof: According to the assumption we have $\boldsymbol{x} \overset{d}{=} \mu + \boldsymbol{C}\boldsymbol{z}$. Therefore $\boldsymbol{y} \overset{d}{=} (\boldsymbol{B}\mu + \boldsymbol{d}) + \boldsymbol{B}\boldsymbol{C}\boldsymbol{z}$, i.e. $\boldsymbol{y} \sim N_l(\boldsymbol{B}\mu + \boldsymbol{d}, \boldsymbol{B}\boldsymbol{C}\boldsymbol{C}^t\boldsymbol{B}^t)$ and assertion a) follows. Part b) follows directly by applying result a) if we let $\boldsymbol{B} = (\boldsymbol{I}_n || \boldsymbol{0})$ with $\boldsymbol{0}$: $n \times (q-n)$ respectively $\boldsymbol{B} = (\boldsymbol{0} || \boldsymbol{I}_{q-n})$ with $\boldsymbol{0}$: $n \times n$ and || denoting the horizontal matrix concatenation operator.

Property a) describes the closure of the mvn distribution under linear transformations and linear combination of random variables. The second property shows that any marginal distribution of a mvn distribution is again normal. But the converse is not true in general. Tong (1990, p. 29), for example, gives a counterexample by considering two univariate normal variables, which joint distribution is not bivariate normal. The next lemma establishes a characterisation of the mvn distribution by means of the characteristic function (c.f.).

Lemma 2.2.: $\boldsymbol{x} \sim N_q(\mu, \boldsymbol{\Sigma})$ holds if and only if its characteristic function is of the form

$$\phi_{\boldsymbol{x}}(\boldsymbol{t}) = e^{i\boldsymbol{t}^t\mu - \frac{1}{2}\boldsymbol{t}^t\boldsymbol{\Sigma}\boldsymbol{t}}, \qquad \boldsymbol{t} \in \mathbb{R}^q.$$

Proof: Since the c.f. of a univariate $N(0, 1)$ variable is $e^{-t^2/2}$, the c.f. of $\boldsymbol{z} \sim N_n(\boldsymbol{0}, \boldsymbol{I}_n)$ is $\phi_{\boldsymbol{z}}(\boldsymbol{t}) = e^{-\frac{1}{2}\boldsymbol{t}^t\boldsymbol{t}}$ $(\boldsymbol{t} \in \mathbb{R}^n)$. The result then follows from

$$\phi_{\boldsymbol{x}}(\boldsymbol{t}) = Ee^{i\boldsymbol{t}^t\boldsymbol{x}} = Ee^{i\boldsymbol{t}^t(\mu + \boldsymbol{C}\boldsymbol{z})} = e^{i\boldsymbol{t}^t\mu}Ee^{i\boldsymbol{t}^t\boldsymbol{C}\boldsymbol{z}} = e^{i\boldsymbol{t}^t\mu - \frac{1}{2}\boldsymbol{t}^t\boldsymbol{\Sigma}\boldsymbol{t}}.$$

In particular this lemma shows that the mvn distribution $N_q(\mu, \boldsymbol{\Sigma})$ is uniquely determined by μ and $\boldsymbol{\Sigma}$. The density of the distribution introduced so far is given by

Lemma 2.3.: Assume $\boldsymbol{x} \sim N_q(\mu, \boldsymbol{\Sigma})$, $|\boldsymbol{\Sigma}| \neq 0$. Then the density of $\boldsymbol{x}$ is

$$f_q(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{q}{2}}\sqrt{|\Sigma|}} \exp\left(-\tfrac{1}{2}(\boldsymbol{x}-\mu)^t \Sigma^{-1}(\boldsymbol{x}-\mu)\right).$$

Proof: Take a $(q \times q)$–matrix $\boldsymbol{C}$ in Definition 2.1. with $\boldsymbol{z} \sim N_q(\boldsymbol{0}, \boldsymbol{I}_q)$. Since $\Sigma = \boldsymbol{C}\boldsymbol{C}^t$ is non-singular, Σ^{-1} exists. Because of the independence of the marginal univariate densities, the joint density of $\boldsymbol{z} = (Z_1, \ldots, Z_q)$ is

$$f_q(\boldsymbol{z}) = \prod_{i=1}^{q} \tfrac{1}{\sqrt{2\pi}} e^{Z_i^2/2} = (2\pi)^{-q/2} e^{-\frac{1}{2}\boldsymbol{z}^t\boldsymbol{z}}.$$

For the transformation $\boldsymbol{z} = \boldsymbol{C}^{-1}(\boldsymbol{x}-\mu)$ we get with the transformation theorem of densities and the corresponding Jacobian $|\Sigma|^{-1/2}$ the density function of $\boldsymbol{x}$ as

$$f_q(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{q}{2}}\sqrt{|\Sigma|}} \exp\left(-\tfrac{1}{2}(\boldsymbol{x}-\mu)^t \left(\boldsymbol{C}^{-1}\right)^t \boldsymbol{C}^{-1}(\boldsymbol{x}-\mu)\right).$$

The result is established due to $\Sigma = \boldsymbol{C}\boldsymbol{C}^t$ respectively $\Sigma^{-1} = \left(\boldsymbol{C}\boldsymbol{C}^t\right)^{-1} = \left(\boldsymbol{C}^{-1}\right)^t \boldsymbol{C}^{-1}$.

Remark 2.2.: If $|\Sigma| = 0$, Σ^{-1} does not exist and therefore Lemma 2.3. does not hold any more. But in such cases a further transformation with a $(n \times q)$–matrix $\boldsymbol{T}$ exists, so that

$$\boldsymbol{T}\boldsymbol{x} \sim N_n(\boldsymbol{T}\mu, \boldsymbol{T}\Sigma\boldsymbol{T}^t).$$

$\boldsymbol{x}$ is then singular mvn distributed and the probability mass of $\boldsymbol{x}$ is concentrated on a n–dimensional subspace. An application of this fact is given in Theorem 3.1. In general, however, we will restrict ourselves from now on to the non-singular case, if not stated otherwise.

Example 2.1.: Consider the bivariate case for $q = 2$ and let $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$ and $\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{11}}} = \frac{\sigma_{21}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{11}}}$. Since $|\Sigma| = \sigma_{11}\sigma_{22}(1-\rho^2)$, the inverse Σ^{-1} exists iff $|\rho| < 1$. With $\Sigma^{-1} = \frac{1}{\sigma_{11}\sigma_{22}(1-\rho^2)} \begin{pmatrix} \sigma_{22} & -\sqrt{\sigma_{11}\sigma_{22}}\rho \\ -\sqrt{\sigma_{11}\sigma_{22}}\rho & \sigma_{11} \end{pmatrix}$ one obtains for the joint density of $\boldsymbol{x} = (X_1, X_2)^t$

$$f_2(\boldsymbol{x}) = \frac{1}{2\pi\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{X_1-\mu_1}{\sqrt{\sigma_{11}}} \right)^2 \right.\right.$$

$$\left.\left. -2\rho \left(\frac{X_1-\mu_1}{\sqrt{\sigma_{11}}} \right) \left(\frac{X_2-\mu_2}{\sqrt{\sigma_{22}}} \right) + \left(\frac{X_2-\mu_2}{\sqrt{\sigma_{22}}} \right)^2 \right] \right\} =$$

$$= \frac{1}{2\pi\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}\sqrt{1-\rho^2}} \exp\left\{ -\frac{u^2 - 2\rho uv + v^2}{2(1-\rho^2)} \right\},$$

where $u = \frac{X_1-\mu_1}{\sqrt{\sigma_{11}}}$ and $v = \frac{X_2-\mu_2}{\sqrt{\sigma_{22}}}$. In the case of $|\rho| = 1$, $|\Sigma| = 0$ and applying Remark 2.2. one sees that the probability mass is concentrated on a straight line. Figures, illustrating $f_2(\boldsymbol{x})$ for $\mu = \boldsymbol{0}$ and different values of the correlation ρ, are given in the Plate. The graphs were drawn by invoking Mathematica (1996) with calls similar to

```
Plot3D[Exp[(-u^2 + 2 u v ρ - v^2) / (2 (1 - ρ^2))] / (2 Pi Sqrt[1 - ρ^2]), {u, -3, 3}, {v, -3, 3},
 PlotRange -> All, PlotPoints -> 40, ViewPoint -> {1., 1., .5}, Mesh -> False, Boxed -> False,
 Axes -> {True, True, False}, PlotLabel -> StyleForm["ρ = -0.6", "Section"]]

Show[ContourGraphics[%], PlotLabel -> StyleForm["ρ = -0.6", "Section"],
 ContourShading -> False, Contours -> 10]
```

The following result finally establishes the fact that Definition 2.1. is consistent as noticed in Remark 2.1. We verify that if $\boldsymbol{x}$ has the density function $f_q(\boldsymbol{x})$, the mean vector and the covariance matrix of $\boldsymbol{x}$ are indeed μ and Σ, respectively.

Lemma 2.4.: Assume $\boldsymbol{x} \sim N_q(\mu, \Sigma)$. Then $E(\boldsymbol{x}) = \mu$ and $Cov(\boldsymbol{x}) = \Sigma$.

Proof: By equation (2.1), $E(\boldsymbol{z}) = \boldsymbol{0}$ and $Cov(\boldsymbol{z}) = \boldsymbol{I}_q$. The result follows from $E(\boldsymbol{x}) = \mu + E(\boldsymbol{Cz}) = \mu$ and $Cov(\boldsymbol{x}) = \boldsymbol{C}\, Cov(\boldsymbol{z})\boldsymbol{C}^t = \boldsymbol{CC}^t = \Sigma$.

With the introduction of the mvn distribution being complete now we state subsequently some additional results and characterisations. Most of them can be found in several textbooks, such as the references given above. All of the properties presented now will be used in the course of this thesis and are therefore summarised at this place. First we notice an important closure property of linear combinations of the components of $\boldsymbol{x}$ in

Lemma 2.5.: A random vector $\boldsymbol{x} = \left(X_1, \ldots, X_q\right)^t$ is q–variate normal distributed if and only if the distribution of $\boldsymbol{c}^t\boldsymbol{x}$ is (univariate) $N\left(\boldsymbol{c}^t\mu, \boldsymbol{c}^t\Sigma\boldsymbol{c}\right)$–distributed for all real $\boldsymbol{c} = \left(c_1, \ldots, c_q\right)^t$.

Proof: Obviously, if $\boldsymbol{x} \sim N_q\left(\mu, \Sigma\right)$, then $\boldsymbol{c}^t\boldsymbol{x}$ is a univariate $N\left(\boldsymbol{c}^t\mu, \boldsymbol{c}^t\Sigma\boldsymbol{c}\right)$–variable for all $\boldsymbol{c}$. Conversely, assume that $\boldsymbol{c}^t\boldsymbol{x}$ is univariate normal distributed for $\boldsymbol{c} \in \mathbb{R}^q$. Then the c.f. of $\boldsymbol{c}^t\boldsymbol{x}$

$$\phi_{\boldsymbol{c}^t\boldsymbol{x}}(t) = E\exp\left(it\sum_{i=1}^{q} c_i X_i\right) = e^{it\boldsymbol{c}^t\mu - \frac{t^2}{2}\boldsymbol{c}^t\Sigma\boldsymbol{c}}$$

holds for all $t \in \mathbb{R}$. Taking $t = 1$, one obtains the c.f. of $\boldsymbol{x}$ as a function in $\boldsymbol{c}$

$$\phi_{\boldsymbol{x}}(\boldsymbol{c}) = e^{i\boldsymbol{c}^t\mu - \frac{1}{2}\boldsymbol{c}^t\Sigma\boldsymbol{c}},$$

which is nothing but the c.f. of a mvn variable with mean μ and covariance matrix Σ. By Lemma 2.2. the result is proven.

In general, uncorrelated random variables are not necessarily independent. The following lemma, however, states that those two conditions are equivalent in the mvn case.

Lemma 2.6.: Consider the partition (2.3). Then $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are independent iff $\Sigma_{12} = \boldsymbol{0}$.

Proof: Let $\phi_{\boldsymbol{x}}(\boldsymbol{t})$ denote the c.f. of $\boldsymbol{x}$. Applying Lemma 2.2. one obtains the equivalence transformations

$$\begin{aligned} \Sigma_{12} = \mathbf{0} \quad &\Leftrightarrow \quad \mathbf{t}\Sigma\mathbf{t}^t = \mathbf{t}_1^t\Sigma_{11}\mathbf{t}_1 + \mathbf{t}_2^t\Sigma_{22}\mathbf{t}_2 \\ &\Leftrightarrow \quad \phi_{\mathbf{x}}(\mathbf{t}) = \phi_{\mathbf{x}_1}(\mathbf{t}_1)\phi_{\mathbf{x}_2}(\mathbf{t}_2) \end{aligned}$$

for all $\mathbf{t}_1 = (t_1, \ldots, t_n)^t \in \mathbb{R}^n$ and $\mathbf{t}_2 = (t_{n+1}, \ldots, t_q)^t \in \mathbb{R}^{q-n}$. Since $\mathbf{x}_1$ and $\mathbf{x}_2$ are independent iff their joint c.f. is the product of the marginal c.f., the proof is complete.

Next, we just state the conditional distribution of a mvn variable, to be used later on.

Lemma 2.7.: Partition $\mathbf{x}$ in the fashion of (2.3). Then for $n < q$ the conditional distribution of $\mathbf{x}_1$ given $\mathbf{x}_2$ is $N_n(\mu_{1.2}, \Sigma_{11.2})$, where $\mu_{1.2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2)$ and $\Sigma_{11.2} = \Sigma_{11} + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Proof: For example Tong (1990, pp. 33).

The following assertion will be helpful for calculating tail probabilities in the next section. It says that for a q–dimensional random vector, whose components are i.i. $N(0, 1)$–distributed, its direction is of uniform distribution on the surface S of the unit hypersphere. Utilising this fact one has a simple method to generate uniform random vectors on S.

Lemma 2.8.: Let $\mathbf{z} = (Z_1, \ldots, Z_q)^t$, Z_i i.i.d. $N(0, 1)$–distributed, $i = 1, \ldots, q$. Set $r^2 = \mathbf{z}^t\mathbf{z} = Z_1^2 + \ldots + Z_q^2$. Then the direction $\eta = \left(\frac{Z_1}{r}, \ldots, \frac{Z_q}{r}\right)$ is distributed uniformly on $S = \left\{\mathbf{x} \middle| \sum_{i=1}^q x_i^2 = 1\right\}$.

Proof: The density function of $\mathbf{z}$ is given by $f_q(\mathbf{z}) = (2\pi)^{-q/2}e^{-\frac{1}{2}\mathbf{z}^t\mathbf{z}}$. Replacing the Cartesian coordinates by the polar coordinates $\psi = (\psi_1, \ldots, \psi_{q-1}, r)$ and making use of the transformation theorem for densities the density function of the random vector ψ can be obtained as

$$h(\psi_1, \ldots, \psi_{q-1}, r) = (2\pi)^{-q/2}e^{-r^2/2}g(\psi),$$

where the Jacobian is given by $g(\psi) = r^{q-1} \cos^{q-2} \psi_1 \ldots \cos \psi_{q-2}$. As the density function h is decomposed into a product of uniform functions of the individual coordinates, the assertion follows.

Finally, we proof the following short

Lemma 2.9.: Let $\boldsymbol{x} = \left(X_1, \ldots, X_q\right)^t$ and $\boldsymbol{y} = -\boldsymbol{x}$, $\boldsymbol{x} \sim N_q(\boldsymbol{0}, \boldsymbol{\Sigma})$. Then $\boldsymbol{x} \overset{d}{=} \boldsymbol{y}$.

Proof: As $\boldsymbol{y} = -\boldsymbol{I}_q \boldsymbol{x}$, one obtains by applying (2.2) $\boldsymbol{y} \sim N_q\left(\boldsymbol{0}, -\boldsymbol{I}_q \boldsymbol{\Sigma} \left(-\boldsymbol{I}_q\right)^t\right) = N_q(\boldsymbol{0}, \boldsymbol{\Sigma})$.

2.1.2. Computation of multivariate normal probabilities

We have introduced in the last subsection the mvn distribution and several of its main properties to be used in the later course. In contrast, we now focus on the rather practical problem of evaluating the arising cumulative distribution function. Before proceeding further, some terminology is given in

Definition 2.2.: Let $\boldsymbol{x} \sim N_q(\mu, \boldsymbol{\Sigma})$ according to Definition 2.1., μ be the mean vector and $\boldsymbol{\Sigma}$ the covariance matrix of the q–dimensional random variable $\boldsymbol{x}$ with $|\boldsymbol{\Sigma}| > 0$. Regarding to Lemma 2.3. let further the density be given by $f_q(\boldsymbol{x})$. We then write for the corresponding *multivariate cumulative normal distribution function*

$$\begin{aligned} \boldsymbol{\Phi}_q &= \boldsymbol{\Phi}_q(\boldsymbol{a}, \boldsymbol{b}; \mu, \boldsymbol{\Sigma}) = \\ &= P_{\mu, \boldsymbol{\Sigma}}(\boldsymbol{a} \le \boldsymbol{x} \le \boldsymbol{b}) = P\left(a_1 \le X_1 \le b_1, \ldots, a_q \le X_q \le b_q\right) = \\ &= \int_{a_1}^{b_1} \ldots \int_{a_q}^{b_q} f_q(\boldsymbol{x})\, dX_q \ldots dX_1. \end{aligned} \tag{2.4}$$

If $a_i = -\infty$ and $b_i = 0 \ \forall i$, we call this special case *orthant probability* and denote it P_q.

Remark 2.3.: Let $\boldsymbol{\Sigma} = \left(\sigma_{ij}\right)_{ij}$ be the covariance matrix and $\boldsymbol{R} = \left(\rho_{ij}\right)_{ij} = \left(\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}\right)_{ij}$ the associated correlation matrix. Because of

$$P_{\mu,\,\Sigma}\left(\bigcap_{i=1}^{q}\left\{a_i \le X_i \le b_i\right\}\right) = P_{\mathbf{0},\,\boldsymbol{R}}\left(\bigcap_{i=1}^{q}\left\{\frac{a_i-\mu_i}{\sqrt{\sigma_{ii}}} \le Y_i \le \frac{b_i-\mu_i}{\sqrt{\sigma_{ii}}}\right\}\right),$$

where $Y_i = \frac{X_i-\mu_i}{\sqrt{\sigma_{ii}}}$ $\left(i = 1, \ldots, q\right)$, we usually consider only the case $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{R}$ without loss of generalisation. Instead of equation (2.4) we therefore investigate $\Phi_q\left(\boldsymbol{a}, \boldsymbol{b}; \mathbf{0}, \boldsymbol{R}\right)$ on the following pages, which is often easier to handle. We call this the studentised mvn distribution function.

The problem of evaluating $\boldsymbol{\Phi}_q$ accurately and within an acceptable time limit has been addressed to since the arise of the mvn distribution in the statistical literature in the middle of the last century. Especially since the 1940's, in connection with the new computer systems, plenty of theoretical accurate or simulation-based approaches have been proposed. For some particular cases, satisfactory solutions are at hand. For smaller dimensions, such as $q = 1$ or $q = 2$, there is reliable and efficient software available, but in the case of $q \ge 3$ still no adequate general solution exists. For $\boldsymbol{R}$ satisfying special structures, dimension reducing methods might be applied (see for example the product correlation structure of Remark 1.1.). When choosing the integration limits appropriately, other shortcuts might work (the orthant probabilities are such an example). But for this thesis we require a general methodology to compute $\boldsymbol{\Phi}_q$, which imposes no restrictions on $\boldsymbol{R}$, q and $\boldsymbol{a}$ or $\boldsymbol{b}$. Broad literature reviews on this subject have been given by Gupta (1963a, b) and Martynov (1981). More recent developments are summarised by Bretz (1999). In the following we will present three approaches proposed in the literature. They were chosen according to the requirements encountered in the course of the present thesis.

2.1.2.1. Approximation of Solow

Later we will be faced frequently with the problem of computing $\boldsymbol{\Phi}_q\left(\boldsymbol{b}; \mathbf{0}, \boldsymbol{R}\right)$ for the equi-percentage point $\boldsymbol{b} = \left(b, \ldots, b\right)$ and $\boldsymbol{a} = \left(-\infty, \ldots, -\infty\right)$. When computing quantiles or p-values of the MCT statistics, only upper equicoordinate integration bounds are required due to the

maximum-type statistic. And because of one-sided testing we restrict the attention to the lower bound set as $-\infty$.

Solow (1990) described a simple way for approximating $\Phi_q(\boldsymbol{b}; \boldsymbol{0}, \boldsymbol{R})$ from univariate and bivariate marginal probabilities. The method is based on decomposing $\Phi_q(\boldsymbol{b})$ into a product of conditional probabilities and approximates each term in the product using conditional expectations introduced in Lemma 2.7. First we define the indicator function

$$I_i(\boldsymbol{b}) \;=\; I(X_i \le b)$$

with the expectation $E(I_i(\boldsymbol{b})) = \Phi_1(b)$ for $i = 1, ..., q$. Then the equations

$$\begin{aligned} & P(X_1 \le b, \ldots, X_q \le b) = \\ &= P(X_1 \le b,\ X_2 \le b)\prod_{j=3}^{q} P(X_j \le b \,|\, X_1 \le b, \ldots, X_{j-1} \le b) = \\ &= P(X_1 \le b,\ X_2 \le b)\prod_{j=3}^{q} E(I_j \,|\, I_1 = 1, \ldots, I_{j-1} = 1) \end{aligned} \qquad (2.5)$$

hold. Finally, we approximate the factors of the second term in equation (2.5) with

$$E(I_j) + \Sigma_{21}\Sigma_{11}^{-1}\left(1 - E(I_1), \ldots, 1 - E(I_{j-1})\right)^t, \qquad (2.6)$$

where $\Sigma_{21} = \mathrm{cov}(I_j,\ I_i) = E(I_j I_i) - E(I_j)E(I_i), \quad i = 1, \ldots,\ j-1,$ and $\Sigma_{11} = \mathrm{cov}(I_l,\ I_i) =$ $= E(I_l I_i) - E(I_l)E(I_i), \quad 1 \le i, l \le j-1.$ Note that $E(I_l I_i)$ is a bivariate marginal probability and $E(I_l)E(I_i)$ the product of two univariate ones. The use of (2.6) as an approximation of (2.5) is analogous to the formula $E(\boldsymbol{x}_1 | \boldsymbol{x}_2 = \boldsymbol{x}_2) = \mu_{1.2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{x}_2 - \mu_2)$ in Lemma 2.7. Comparisons to other methods and further notes regarding the procedure of Solow are given in the concluding Subsection 2.1.2.4.

2.1.2.2. Transformations of Genz

Genz (1992) succeeded in transforming the original multiple integral (2.4) into one over the $(q-1)$–dimensional hypercube. This has the advantage that possible integration bounds of the type $-\infty$ or $+\infty$ are avoided during the calculation. Moreover, efficient standard numerical multiple integration algorithms may be applied in this setting. We next proof the main result in a slightly different form than originally given by Genz (1992). The same one-sided set-up already introduced before is considered again.

Theorem 2.1.: Let $\boldsymbol{a} = (-\infty, \ldots, -\infty) \in \overline{\mathbb{R}}^q$, $\boldsymbol{b} = (b_1, \ldots, b_q) \in \mathbb{R}^q$ arbitrary and $\boldsymbol{\Phi}_q(\boldsymbol{b}) = \boldsymbol{\Phi}_q(\boldsymbol{a}, \boldsymbol{b}; \boldsymbol{0}, \boldsymbol{R})$ of Definition 2.2. and Remark 2.3. be given. Further, denote by $\boldsymbol{R} = \boldsymbol{C}\boldsymbol{C}^t$ the Cholesky decomposition, where $\boldsymbol{C} = (c_{ij})_{ij}$ is a lower triangular matrix. Then for $q > 1$

$$\boldsymbol{\Phi}_q(\boldsymbol{b}) = e_1 \int_0^1 e_2(\boldsymbol{w}) \int_0^1 \ldots \int_0^1 e_q(\boldsymbol{w})\, d\boldsymbol{w}, \tag{2.7}$$

where $e_1 = \boldsymbol{\Phi}_1(b_1/c_{11})$ and $e_i = \boldsymbol{\Phi}_1\left(\left(b_i - \sum_{j=1}^{i-1} c_{ij}\boldsymbol{\Phi}_1^{-1}(e_j(\boldsymbol{w}) \cdot W_j)\right) \Big/ c_{ii}\right)$ for $i = 2, \ldots, q$ and $\boldsymbol{w} = (W_1, \ldots, W_{q-1}) \in \mathbb{R}^{q-1}$.

Proof: Consider the original integral $\boldsymbol{\Phi}_q(\boldsymbol{b})$. Three substitutions shall be considered in the following. Denote by $\boldsymbol{x} = \boldsymbol{C}\boldsymbol{y}$ the first substitution. Because of $\boldsymbol{x}^t\boldsymbol{R}^{-1}\boldsymbol{x} = \boldsymbol{y}^t\boldsymbol{C}^t\boldsymbol{C}^{-t}\boldsymbol{C}^{-1}\boldsymbol{C}\boldsymbol{y} = \boldsymbol{y}^t\boldsymbol{y}$ and $d\boldsymbol{x} = |\boldsymbol{C}|d\boldsymbol{y} = |\boldsymbol{R}|^{½}d\boldsymbol{y}$ one gets in the first step

$$\boldsymbol{\Phi}_q(\boldsymbol{b}) = \frac{1}{(2\pi)^{q/2}} \int_{-\infty}^{b_1'} e^{-\frac{y_1^2}{2}} \int_{-\infty}^{b_2'} e^{-\frac{y_2^2}{2}} \ldots \int_{-\infty}^{b_q'} e^{-\frac{y_q^2}{2}}\, d\boldsymbol{y}$$

with $b_i' = \left(b_i - \sum_{j=1}^{i-1} c_{ij}Y_j\right) \Big/ c_{ii}$. Making use of the second transformation $Y_i = \boldsymbol{\Phi}_1^{-1}(Z_i)$ yields

$$\Phi_q(\boldsymbol{b}) = \int_0^{e_1} \int_0^{e_2} \dots \int_0^{e_q} d\boldsymbol{z}.$$

Here, $e_i = e_i(Z_1, \dots, Z_{i-1}) = \Phi_1\left(\left(b_i - \sum_{j=1}^{i-1} c_{ij}\Phi_1^{-1}(Z_j)\right) \Big/ c_{ii}\right)$. In the final step, let $Z_i = e_i W_i$ and (2.7) follows directly from the last equation.

With the problem of evaluating $\Phi_q(\boldsymbol{b})$ reduced to the equivalent form (2.7) the task is now to compute this new multiple integral. Several standard numerical integration techniques might be applied at this point. Originally, Genz (1992) proposed using a crude Monte Carlo method and a subregion adaptive method. Next, Beckers and Haegemans (1992) successfully implemented lattice rules for integrating (2.7). Genz (1993) compared all these methods, including a randomised lattice rule algorithm and other integration techniques due to Deak (1990). According to his results, the randomised lattice rules behave best for moderate dimension and accuracy. Unfortunately, no further description of the method investigated was given in the paper. Therefore, an independent algorithmic implementation using the lattice rule approach is given below.

The algorithm refers strongly to the approach for the multivariate *t*–distribution which will be proposed in Subsection 2.2.2.2. For this reason the reader is referred to there for a short theoretical background and references of the underlying methods. Only a very brief discussion is given here. The algorithm proposed consists mainly of a simulated evaluation of the integral (2.7). But instead of evaluating it at random points throughout the unit hypercube, sequences of nodes $\boldsymbol{z}$ "better than random" are used. Despite the self-explaining input parameters q, $\boldsymbol{R}$ and $\boldsymbol{b}$ the ε is an user-specified error bound. The practitioner is therefore able to control the error committed via the simulated standard error. The whole evaluation is repeated several times, until this value falls below the pre-determined threshold ε. The output parameters *Intval*, *ErrEst* and Np_n denote the estimate of the sought probability, the error estimate described previously and the number of evaluations, respectively. The factor γ is the Monte Carlo confidence bound for the simulated standard error.

1. **INPUT** q, γ, $\boldsymbol{R}$, $\boldsymbol{b}$, ε.
2. Compute lower triangular Cholesky factor $\boldsymbol{C}$ for $\boldsymbol{R}$.
3. Initialise $N = 10$, $n = 0$, $e_1 = \Phi_1(b_1/c_{11})$, $f_1 = e_1$.
4. **REPEAT**
   a) Set $n = n + 1$, $Intval = 0$, $Varsum = 0$.
   b) **FOR** $l = 1, 2, ..., N$
      i) Set $Latsum = 0$.
      ii) Generate uniform random $W_1, ..., W_{q-1} \in [0, 1]$.
      iii) **FOR** $j = 1, 2, ..., p_n$
      - **FOR** $i = 2, ..., q$

        Set $R_{i-1} = \left| 2\left\{W_{i-1} + \frac{j}{p_n}\boldsymbol{z}\right\} - 1\right|$,

        $Y_{i-1} = \Phi_1^{-1}(e_{i-1}R_{i-1})$,

        $e_i = \Phi_1\left(\left(b_i - \sum_{j=1}^{i-1} c_{ij}Y_j\right)\Big/ c_{ii}\right)$,

        $f_i = e_i \cdot f_{i-1}$.

        END FOR
      - Set $Latsum = Latsum + (f_q - Latsum)/j$.

      END FOR
      iv) Set $Varsum = Varsum + (l - 1)(Latsum - Intval)^2/l$,

      $Intval = Intval + (Latsum - Intval)/l$.

   END FOR
   c) Set $ErrEst = \gamma\sqrt{Varsum/(N(N-1))}$.

   UNTIL $ErrEst < \varepsilon$.
5. **OUTPUT** $Intval$, $ErrEst$, Np_n.

2.1.2.3. Calculation of orthant probabilities

We now draw our attention on computing the orthant probabilities P_q introduced in Definition 2.2. The motivation for this will get clear in Subsection 2.1.3., where we reduce the problem

of calculating the level probabilities on that discussed in the following. Because of Lemma 2.9. we have immediately

$$P_q = \int_{-\infty}^{\mathbf{0}} f_q(\boldsymbol{x})\, d\boldsymbol{x} = \int_{\mathbf{0}}^{\infty} f_q(\boldsymbol{x})\, d\boldsymbol{x}.$$

Historically, the latter form was the preferred one throughout the literature, but we keep on analysing the originally defined P_q. One main result is due to David (1953):

Lemma 2.10.: Consider the event E_i: "$X_i < 0$", $i = 1, \ldots, q$. Then for q odd

$$P_q = P\left(\bigcap_i E_i\right) = \frac{1}{2}\left[1 - \sum_i P(E_i) + \sum_{i<j} P\left(E_i \cap E_j\right) - \ldots + \sum_{\{i_1 < \ldots < i_{q-1}\}} P\left(\bigcap_{\{i_1, \ldots, i_{q-1}\}} E_{i_j}\right)\right],$$

where $\{i_1, \ldots, i_{q-1}\} \subset \{1, \ldots, q\}$.

Proof: The theorem of Boole gives

$$P\left(\bigcup_i E_i\right) = \sum_i P(E_i) - \sum_{i<j} P\left(E_i \cap E_j\right) + \ldots + (-1)^{q-1} P\left(\bigcap_i E_i\right).$$

Because of the symmetry of the mvn distribution mentioned in Lemma 2.9. we have also

$$P\left(\bigcup_i E_i\right) = 1 - P\left(\bigcap_i E_i^C\right) = 1 - P\left(\bigcap_i E_i\right).$$

Hence, the assertion follows by combining both equations.

This result shows that all odd q–variate orthant probabilities can be expressed as a linear combination of some lower order orthant probabilities. In the case of q even and applying above way of derivation one obtains only the identity. Therefore, efforts can be focused on determining P_q just for q even. For small q simple closed expressions are available (see for

example Robertson et al., 1988, p. 75), but for arbitrary q no general solution exists up to now. In the following a method is therefore provided, which solely calculates the required multiple integrals for q even of practical importance (i.e. up to q = 10). For the remaining integrals above lemma is applied. By using the c.f. and Parseval's theorem Childs (1967) succeeded to show

Theorem 2.2.: Let $\boldsymbol{t} \in \mathbb{R}^q$. For $q = 2k$ define $I^q(\Sigma) = \frac{1}{(-2\pi)^k} \int_{\mathbb{R}^q} \prod_{i=1}^{q} \frac{1}{t_i}\, e^{-\frac{1}{2}\boldsymbol{t}^t \Sigma \boldsymbol{t}}\, d\boldsymbol{t}$, where $\Sigma = \Sigma_q = (\sigma_{ij})_{ij}$ is a covariance matrix of q variates. Then

$$P_{2k} = \frac{1}{2^{2k}} + \frac{1}{2^{2k-1}\pi} \sum_{i<j=1}^{2k} \sin^{-1}(\rho_{ij}) + \sum_{j=2}^{k} \frac{1}{2^{2k-j}\pi^j} \sum_{i_1<\ldots<i_{2j}=1}^{2k} I^{2j}\left(\boldsymbol{R}^{(i_1,\ldots,i_{2j})}\right),$$

where $\boldsymbol{R}^{(i_1,\ldots,i_{2j})}$ denotes a submatrix consisting of the i_1^{th}, …, $i_{2\text{j}}^{\text{th}}$ rows and columns of $\boldsymbol{R}$.

Proof: See Childs (1967).

Sun (1988a) developed these representations further and obtained the following recursive relationship among the I^q's.

Lemma 2.11.: With the notation of Theorem 2.2. define $\Sigma' = (\sigma'_{ij})_{ij}$ with $\sigma'_{11} = 1/t^2$, $\sigma'_{1j} = \sigma'_{j1} = \sigma_{1j}/\sqrt{\sigma_{11}}$ and $\sigma'_{ij} = \sigma_{ij}$ otherwise. Further on, for each fixed $i = 2, 3, \ldots, q$ partition Σ' in a similar fashion to (2.3), but replacing σ'_{22} by σ'_{ii}. With $\Sigma^i_{q-2} = \Sigma'_{22} - \Sigma'^t_{12}\Sigma'^{-1}_{11}\Sigma'_{12}$ and $I^2(\Sigma_2) = \sin^{-1}(\rho_{12})$ the relation

$$I^q(\Sigma) = \int_0^1 \sum_{i=2}^{q} \frac{\sigma_{1i}}{\sqrt{\sigma_{11}\sigma_{ii} - \sigma_{1i}^2 t^2}}\, I^{q-2}\left(\Sigma^i_{q-2}\right) dt \qquad (2.8)$$

holds for $q = 2k$ and $k > 1$.

Proof: See Sun (1988a, pp. 3916).

Together with Theorem 2.2. a repeated application of (2.8) yields the final result that P_q can be represented as a linear combination of several multiple integrals of order at most $\frac{q}{2}-1$. This reduction of the dimension by factor 2 has clear time advantages over concurring procedures. Sun (1988a) has established the explicit formula for $q = 4$ and $q = 6$. In two subsequent papers (Sun, 1988b; Sun and Asano, 1989) these expressions were extended for higher dimensions up to $q = 10$. Computational comparisons in these articles suggest a good behaviour in terms of accuracy and time required. Note that the problems of integrating $\frac{1}{t_i}$ in Theorem 2.2. are avoided with the recursion technique used in the subsequent Lemma 2.11.

2.1.2.4. Conclusions

In the preceding subsections we have introduced three approaches of computing mvn probabilities, which differ from each other in their goals and their behaviour. The approach of Solow (1990) can only be conducted for the computation of equipercentage points. Generalisations to arbitrary rectangular integration regions and approximations using tri- and quadrivariate integrals resulting in higher accuracy are introduced by Joe (1995). The main virtue of this method, however, lies in its speed and its easy implementation, for it approximates Φ_q solely by uni- and bivariate integrals. An adaptation to the multivariate *t*–case in the next section shows that even for high dimensions, $q = 50$ say, the computation is a task of a few seconds only. One further major advantage is its applicability to other multivariate distributions, provided the ability of computing uni- and bivariate probabilities is given. On the other side, the missing error control is a severe drawback, even if comparison results reported in Solow (1990) and in the next section suggest that the method works well. This approach is recommended for situations requiring a high number of repeated evaluations of mvn probabilities and where a final accuracy of 0.01 or 0.001 is sufficient (for example simulation studies or bootstrap tests).

The transformations of Genz to the unit hypercube have the advantage that standard integration methods can be applied. Many approaches in the numerical literature have been published. Exact methods are still computational not feasible, but the randomised lattice rules proposed here seem to behave well among the present methods. With this algorithm at hand we have an efficient and moderately fast tool to compute the desired probabilities for arbitrary

upper integration bound. Making use of the error bound ε one has further the possibility to set a-priori a limit to the error committed and the outcome gets more reliable. By imposing no restrictions on $\boldsymbol{R}$, q and $\boldsymbol{b}$, this program is flexible to evaluate every kind of mvn probabilities over rectangular regions. Therefore, this approach is the most general of the three methods presented. Because of its similarity to the approach for the multi t–distribution, time and accuracy conclusions might be drawn from the comparisons summarised in Tables 2.2. and 2.3.

Finally, the methods of Childs and Sun lead to fast and accurate algorithms for computing orthant probabilities. Implementations in SAS require only few seconds to calculate a 10– or 11–dimensional integral with an accuracy of at least 10^{-7}. Generalisations to higher dimensions are theoretically possible with above results, but practically the considered dimensions are sufficient. Application of the results to other types of integration regions seems to be troublesome and has not been investigated so far in the literature.

SAS/IML–implementations of all three methods has been conducted and their codes are provided in the Appendix. Calculation of mvn quantiles can be done similar to the procedures introduced in Subsection 2.2.2.3. and will not be repeated here. For further considerations of the computation of mvn probabilities the reader is referred to the homepage of Genz, which contains many additional algorithms in FORTRAN and further links. The source codes are available from the website with URL http://www.sci.wsu.edu/math/faculty/genz/homepage.

2.1.3. Calculation of level probabilities

Recall the null distribution of the LRT under total order restriction stated in Lemma 1.3. It involves a weighted sum of univariate F–probabilities and the so-called level probabilities. We already mentioned there that the computation of the latter one requires multivariate integration techniques. Attention is now drawn on this aspect. At first a general recursive formula for their calculation is provided in the next lemma. Let $k + 1$ be the total number of treatment groups to be investigated. Recall from Definition 1.3. the level probabilities $P(l,\ k+1;\ \boldsymbol{w}) = P(M = l)$ and the notation used there.

Lemma 2.12.: For fixed l, denote by $L_{l,k+1}$ the set of all partitions of an index set $I=\{0, 1, \ldots, k\}$ into l level sets. For $\emptyset \neq A \subset I$ set $W_A = \sum_{i \in A} w_i$ and $\#(A) = \text{card}(A)$. Further on, for $A = \{i_1, i_2, \ldots, i_j\}$ with $0 \leq i_1 < \ldots < i_j \leq k$ and $j = \#(A)$, set $\boldsymbol{w}(A) = (w_{i_1}, \ldots, w_{i_j})$. Then for $l = 2, 3, \ldots, k+1$

$$P(l, k+1; \boldsymbol{w}) = \sum_{\{B_1, \ldots, B_l\} \subset L_{l,k+1}} P(l, l; W_{B_1}, \ldots, W_{B_l}) \prod_{i=1}^{l} P(1, \#(B_i); \boldsymbol{w}(B_i)).$$

Note that $P(l, l; W_{B_1}, \ldots, W_{B_l})$ denotes the probability of l level sets and $P(1, \#(B_i); \boldsymbol{w}(B_i))$ stands for the probability of one level set under the given order restriction and weights.

Proof: For example Robertson et al. (1988, p. 77).

This lemma shows that the computation of $P(l, k+1; \boldsymbol{w})$ can be basically reduced on repeated evaluations of probabilities $P(l, l; .)$. That is, we must be able to calculate $P(X_0 < X_1 < \ldots < X_k)$ and the following result establishes the connection to the orthant probabilities considered on the preceding pages.

Lemma 2.13.: Let $X_0, X_1, \ldots, X_k$ be independent normal variables with common mean zero and variances $w_0^{-1}, w_1^{-1}, \ldots, w_k^{-1}$. Denote further the orthant probability of dimension k by P_k. Then

$$P(X_0 < X_1 < \ldots < X_k) = P_k.$$

Proof: Consider the transformation $Y_i = X_i - X_{i-1}$ for $i = 1, \ldots, k$. Then $P(X_0 < X_1 < \ldots < X_k) = P(Y_1 > 0, \ldots, Y_k > 0)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_k)$ has a mvn distribution with correlation matrix of the following tridiagonal form $(\rho_{ii} = 1)$:

$\rho_{i,i+1} = \rho_{i+1,i} = -\left(\frac{w_{i-1}w_{i+1}}{(w_{i-1}+w_i)(w_i+w_{i+1})}\right)^{\frac{1}{2}}$, $i = 1, \ldots, k-1$, and $\rho_{ij} = 0$ otherwise. The last fact follows also from Lemma 1.4. when choosing $\boldsymbol{C} = \begin{pmatrix} -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{pmatrix}$ and $w_i = n_i$.

With the last step proven and considering the work of the previous subsections we are now able to compute directly the required level probabilities and hence the null distribution of the LRT. A SAS/IML–implementation for calculating p-values of the LRT under order restriction is given in the Appendix. It is based on the orthant probabilities given before and therefore computes in at most a few seconds the result accurately up to 9 decimal digits. The computation of the recurrence relation in Lemma 2.12. is adapted from an algorithm of Seidel (1999). If $k > 11$ (though unlikely to occur in practice), one should replace the module for calculating the particular orthant probabilities by the randomised lattice rule approach. We are therefore able to calculate the LRT for arbitrary number of treatment groups and any sample size constellations.

Great effort to solve the computation of the level probabilities has been paid over the last four decades. For equal weights $\boldsymbol{w} = (w, \ldots, w)$ solutions have already been provided in the original articles of Bartholomew. In the general unbalanced case, however, only approximations have been published (see Robertson et al., Chapter 3, for a broad overview up to 1988). More recently, Shi and Meng (1991) proposed a bootstrap version, because "*when the number of groups is more than four, it is difficult to compute the critical value.*" Qian (1994) proposed two simulation-based techniques to handle the $\bar{\chi}^2$–distribution. A SAS-program for evaluating the LRT has been published by Brunden (1995), but unfortunately it deals with the unbalanced case up to $k = 3$ only. One major development has recently been succeeded by Hayter and Liu (1996). They managed to decompose $P(X_0 < X_1 < \ldots < X_k)$ into a series of nested integrals of lower dimensions. Miwa et al. (1999) applied this technique on the case of the LRT and the modified version according to Wright (1988). Quoting from their article: "... *takes about fourteen seconds to compute for the comparison of* k = *10 treatments.*" With this methodology, a further promising way of evaluating the LRT beside the above developed one, seems to be found.

2.2. Multivariate *t*–distribution

If Z is a $N(0, 1)$–variable and independent from S, where $\nu S^2 \sim \chi^2_\nu$, then the random variable $t = Z/S$ is known as Student's t–variable with ν degrees of freedom. A multivariate generalisation of it is defined below and studied in this section. We proceed in an analogous way as for the mvn distribution. First the density function of the new introduced distribution is derived and some basic properties are given. Afterwards a detailed discussion about the computation of the arising cumulative distribution function follows. New methods are proposed to evaluate multivariate *t*-probabilities and are found to behave well in a comparison study with other proposals in the literature. Together with the derivations in the following chapters we will be able to calculate p-values, quantiles and power values for any multiple contrast within reasonable time. Both SAS and FORTRAN implementations provide the numerical availability of these techniques.

2.2.1. Definition and basic properties

The multivariate *t*–distribution (mv*t*) plays a key role in this thesis. In the normal set-up all multiple contrast tests are distributed according to the mv*t*–distribution. Moreover, we have seen in the first chapter that many established multiple comparison procedures in the literature can be regarded as MCTs. The extensions of the approaches of Williams and Marcus proposed in the next chapter are based on MCTs. The handling of the mv*t*–distribution is therefore fundamental for the present work, in particular the computation of mv*t*–probabilities for arbitrary rectangular regions, imposing no restrictions on other parameters involved. In the following, some main results are established. The mv*t*–distribution is much less described in the textbooks as it is the case for the mvn–distribution. Brief discussions of it are contained in Johnson and Kotz (1972, Chapter 37) and Tong (1990, Chapter 9). The introduction given here follows partly their representations.

Definition 2.3.: Let $\boldsymbol{x} = \left(X_1, \ldots, X_q\right) \sim N_q\left(\mu, \boldsymbol{R}\right)$. Furthermore, let S be an univariate random variable and independent of $\boldsymbol{x}$, where $\nu S^2 \sim \chi^2_\nu$. We then call

$$\boldsymbol{t} = \left(T_1, \ldots, T_q\right)^t = \left(\frac{X_1}{S}, \ldots, \frac{X_q}{S}\right)^t \tag{2.9}$$

a *multivariate* t–*variable*. For $\mu = \boldsymbol{0}$ its distribution is called a *central* q–*variate* t–*distribution* with correlation matrix $\boldsymbol{R}$ and ν degrees of freedom and we write $\boldsymbol{t} \sim t_q\left(\nu, \boldsymbol{R}\right)$. Otherwise, (2.9) is called a *non-central* q–*variate* t–*distribution* with the non-centrality parameter μ, denoted by $\boldsymbol{t} \sim t_{q,\mu}\left(\nu, \boldsymbol{R}\right)$. If $\nu = 1$, the resulting distribution is also known as *multivariate Cauchy*.

It becomes clear from the definition that the mv*t*–distribution depends solely on $\boldsymbol{R}$ and ν. Moreover, we see immediately that the distribution of $\boldsymbol{t}$ is non-singular iff $\boldsymbol{R}$ is positive definite. The following remark notes the non-uniqueness of the multivariate generalisation of a Student's *t*–variable. For the sake of simplicity we deal from now on with the central case only, unless stated otherwise.

Remark 2.4.: One should keep in mind that there are other possibilities published in the literature of generalising an univariate *t*. They were derived for theoretical or applied purposes and concern both central and non-central cases. The definition proposed here, however, is widely used and satisfies our requirements in the later course. For additional issues on this topic the reader is referred to Miller (1968) beside the references given above.

The next lemma establishes the density function of the mv*t*–distribution. It is fundamental for the relationship between both the mvn and the mv*t*–distribution. The result was first derived by Cornish (1954) and, independently, by Dunnett and Sobel (1954).

Lemma 2.14.: For $|\boldsymbol{R}| > 0$ the density function of $\boldsymbol{t}$ from Definition 2.3. is given by

$$g_q(\boldsymbol{t};\, \nu,\, \boldsymbol{R}) = \frac{\Gamma\left(\frac{q+\nu}{2}\right)}{(\nu\pi)^{\frac{q}{2}}\Gamma\left(\frac{\nu}{2}\right)\sqrt{|\boldsymbol{R}|}}\left(1+\frac{1}{\nu}\boldsymbol{t}^t\boldsymbol{R}^{-1}\boldsymbol{t}\right)^{-\frac{q+\nu}{2}}, \quad \boldsymbol{t}\in\mathbb{R}^q.$$

Proof: Because $|\boldsymbol{R}| > 0$, $\boldsymbol{R}^{-1} = \left(c_{ij}\right)_{ij}$ exists and the following steps are valid. From the independence of $\left(\boldsymbol{x},\, \nu S^2\right)^t$ one obtains their joint distribution immediately as

$$h\left(\nu S^2;\, \nu\right)\cdot f_q\left(\boldsymbol{x};\, \boldsymbol{0},\, \boldsymbol{R}\right) = \frac{\left(\frac{\nu}{2}\right)^{\frac{q+\nu}{2}} S^{\nu-2}}{\nu(\nu\pi)^{\frac{q}{2}}\Gamma\left(\frac{\nu}{2}\right)\sqrt{|\boldsymbol{R}|}}\exp\left\{-\frac{1}{2}\left(\boldsymbol{x}^t\boldsymbol{R}^{-1}\boldsymbol{x}+\nu S^2\right)\right\}, \quad \boldsymbol{x}\in\mathbb{R}^q,\, S>0.$$

Here $h(y;\, \nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)}\, y^{\nu/2-1}e^{-y/2}$ is the density of an univariate χ^2_ν–variable, $f_q\left(\boldsymbol{x};\, \boldsymbol{0},\, \boldsymbol{R}\right)$ is given in Lemma 2.3. and $\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}\,dt$ denotes the gamma function. Setting $K = \left[(\nu\pi)^{\frac{q}{2}}\Gamma\left(\frac{\nu}{2}\right)\sqrt{|\boldsymbol{R}|}\right]^{-1}$ and conducting a two-fold substitution the following transformations of the cumulative distribution function are yielded:

$$\begin{aligned}
F(\boldsymbol{z}) &= P(\boldsymbol{t}\le\boldsymbol{z}) = P(\boldsymbol{x}/S\le\boldsymbol{z}) = P(\boldsymbol{x}\le\boldsymbol{z}S) = \\
&= K\int_0^\infty\int_{-\infty}^{\boldsymbol{z}S}\frac{\nu^{\frac{q+\nu}{2}}}{2^{\frac{q+\nu}{2}}}\frac{S^{\nu-2}}{\nu}\exp\left\{-\frac{1}{2}\left(\sum_{i,\,j}X_iX_jc_{ij}+\nu S^2\right)\right\}d\boldsymbol{x}\,d\left(\nu S^2\right) = && \left(X_i = t_iS\right)\\
&= K\int_0^\infty\int_{-\infty}^{\boldsymbol{z}}\frac{\nu^{\frac{q+\nu}{2}}}{2^{\frac{q+\nu}{2}}}\frac{S^{\nu-2}}{\nu}\exp\left\{-\frac{1}{2}\left(\sum_{i,\,j}t_it_jc_{ij}S^2+\nu S^2\right)\right\}S^q\;d\boldsymbol{t}\,d\left(\nu S^2\right) = && \left(H = 1+\tfrac{1}{\nu}\textstyle\sum t_it_jc_{ij}\right)\\
&= K\int_{-\infty}^{\boldsymbol{z}}\int_0^\infty\frac{\nu^{\frac{q+\nu}{2}}}{2^{\frac{q+\nu}{2}}}\frac{S^{q+\nu-2}}{\nu}\exp\left\{-\frac{\nu S^2H}{2}\right\}d\left(\nu S^2\right)d\boldsymbol{t} = && \left(\kappa = H\nu S^2/2\right)\\
&= K\int_{-\infty}^{\boldsymbol{z}}\int_0^\infty\frac{\nu^{\frac{q+\nu}{2}}}{2^{\frac{q+\nu}{2}}}\frac{(2\kappa)^{\frac{q+\nu-2}{2}}}{\nu(\nu H)^{\frac{q+\nu-2}{2}}}\frac{2}{H}e^{-\kappa}\;d\kappa\,d\boldsymbol{t} = \\
&= K\int_0^\infty\kappa^{\frac{q+\nu}{2}-1}e^{-\kappa}\,d\kappa\int_{-\infty}^{\boldsymbol{z}}\frac{\nu^{\frac{q+\nu}{2}}}{2^{\frac{q+\nu}{2}}}\frac{2^{\frac{q+\nu}{2}-1}\cdot 2}{\nu\cdot\nu^{\frac{q+\nu}{2}-1}H\cdot H^{\frac{q+\nu}{2}-1}}\,d\boldsymbol{t} = \\
&= K\cdot\Gamma\left(\frac{q+\nu}{2}\right)\int_{-\infty}^{\boldsymbol{z}}H^{-\frac{q+\nu}{2}}\,d\boldsymbol{t}.
\end{aligned}$$

As one would expect from the univariate case, the mv*t*–distribution converges for increasing degrees of freedom to the mvn distribution. This basic fact together with two resulting conclusions are summarised in the following statements.

Lemma 2.15.: Let $g_q(\boldsymbol{t};\ \nu,\ \boldsymbol{R})$ and $f_q(\boldsymbol{x};\ \boldsymbol{0},\ \boldsymbol{R})$ be the density functions of a mv*t*–variable and a mvn variable, respectively. Then

$$\lim_{\nu\to\infty} g_q(\boldsymbol{t};\ \nu,\ \boldsymbol{R}) = f_q(\boldsymbol{t};\ \boldsymbol{0},\ \boldsymbol{R}) \qquad \forall \boldsymbol{t} \in \mathbb{R}^q.$$

Proof: Because of $e = \lim_{p\to\infty}\left(1+\frac{1}{p}\right)^p$ one obtains $\left(1+\frac{1}{\nu}\boldsymbol{t}^t\boldsymbol{R}^{-1}\boldsymbol{t}\right)^{-(q+\nu)/2} \xrightarrow{\nu\to\infty} e^{-\boldsymbol{t}^t\boldsymbol{R}^{-1}\boldsymbol{t}/2}$. Thus we only have to show $\frac{\Gamma((q+\nu)/2)}{(\nu\pi)^{q/2}\Gamma(\nu/2)|\boldsymbol{R}|^{1/2}} \xrightarrow{\nu\to\infty} \frac{1}{(2\pi)^{q/2}|\boldsymbol{R}|^{1/2}}$. For q even this follows for $\Gamma(x+1) = x\Gamma(x)$, $x > 0$, from

$$\frac{\Gamma\left(\frac{q}{2}+\frac{\nu}{2}\right)}{\nu^{q/2}\Gamma(\nu/2)} = \frac{\Gamma(\nu/2)\prod_{i=1}^{q/2}\left(\frac{\nu}{2}+i\right)}{\nu^{q/2}\Gamma(\nu/2)} \xrightarrow{\nu\to\infty} \frac{1}{2^{q/2}}.$$

Similarly, for q odd the result follows from $\frac{\sqrt{2}\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu}\Gamma(\nu/2)} \xrightarrow{\nu\to\infty} 1$.

Corollary: For $\boldsymbol{x} \sim N_q(\boldsymbol{0},\ \boldsymbol{R})$ and $\boldsymbol{t} \sim t_q(\nu,\ \boldsymbol{R})$

$$\lim_{\nu\to\infty} P(\boldsymbol{t} \in A) = P(\boldsymbol{x} \in A)$$

holds for $A \subset \mathbb{R}^q$ Borel–σ–algebra.

Proof: Direct application of the dominated convergence theorem on above lemma.

Corollary: With $\boldsymbol{t} \sim t_q(\nu, \boldsymbol{R})$ and $\boldsymbol{x} \sim N_q(\boldsymbol{0}, \boldsymbol{R})$ define the equicoordinate quantiles $c_{\nu,\alpha}$ and c_α for $\alpha \in]0, 1[$ as $P\left(\bigcap_{i=1}^{q}\{T_i \le c_{\nu,\alpha}\}\right) = \alpha$, and $P\left(\bigcap_{i=1}^{q}\{X_i \le c_\alpha\}\right) = \alpha$. Then

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

Proof: Since the L.H.S. describes just the univariate probabilities $P\left(\max_{1\le i\le q} T_i \le c_{\nu,\alpha}\right)$ and $P\left(\max_{1\le i\le q} X_i \le c_\alpha\right)$, respectively, the corollary follows from above assertions.

As last property to be considered at this place we get by routine calculation the moments of the mv*t*–distribution, summarised in

Lemma 2.16.: Let $\boldsymbol{t} \sim t_{q,\mu}(\nu, \boldsymbol{R})$. Then the following assertions are valid:

a) $E(\boldsymbol{t}) = \mu$ for $\nu > 1$;

b) $E\left[(\boldsymbol{t}-\mu)(\boldsymbol{t}-\mu)^t\right] = \frac{\nu}{\nu-2}\boldsymbol{R}$ for $\nu > 2$;

c) $E\left[(\boldsymbol{t}-\mu)^t \boldsymbol{R}^{-1}(\boldsymbol{t}-\mu)\right] = \frac{\nu q}{\nu-2}$ for $\nu > 2$.

In particular, this result ensures that the mean vector and the correlation matrix of $\boldsymbol{t}$ are indeed μ and $\boldsymbol{R}$, respectively, as stated in Definition 2.3. Moreover, one notices from $g(u) = \frac{\Gamma((q+\nu)/2)}{(\nu\pi)^{q/2}\Gamma(\nu/2)}(1+u/\nu)^{-(q+\nu)/2}$, that the mv*t* also belongs to the class of elliptical countered distributions for $u = (\boldsymbol{t}-\mu)^t \boldsymbol{R}^{-1}(\boldsymbol{t}-\mu)$ (the density is the same for all $\boldsymbol{t}$ that have the same $\boldsymbol{R}^{-1}$ distance from μ, and thus the distribution is ellipsoidally symmetric about μ). Many general results can therefore be adopted straight forward (regarding closure under linear transformations, conditional and marginal distributions, etc.). We refer to the books of Tong (1990) and Fang and Zhang (1990) for further reading.

2.2.2. Computation of multivariate *t*–probabilities

After introducing the mv*t*–distribution we now focus on the important point of computing the arising distribution function. Although we have dealt with it several times on the preceding pages, we catch up defining it properly in

Definition 2.4.: Let $\boldsymbol{t} \sim t_q(\nu, \boldsymbol{R})$ according to Definition 2.3., where $\boldsymbol{R}$ is the correlation matrix of the q–dimensional random variable $\boldsymbol{t}$ with $|\boldsymbol{R}| > 0$. Regarding to Lemma 2.14. let further the density be given by $g_q(\boldsymbol{t})$. We then write for the corresponding *multivariate cumulative* t*–distribution function*

$$\begin{aligned} T_q(\boldsymbol{a},\boldsymbol{b}) &= T_q(\boldsymbol{a},\boldsymbol{b};\ \nu,\ \boldsymbol{R}) = \\ &= P_{\nu,\,\boldsymbol{R}}(\boldsymbol{a} \le \boldsymbol{t} \le \boldsymbol{b}) = P\left(a_1 \le T_1 \le b_1, \ldots, a_q \le T_q \le b_q\right) = \\ &= \frac{\Gamma\left(\frac{\nu+q}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\nu\pi)^q}\sqrt{|\boldsymbol{R}|}} \int_{a_1}^{b_1}\int_{a_2}^{b_2} \ldots \int_{a_q}^{b_q} \left(1+\frac{\boldsymbol{t}^t\boldsymbol{R}^{-1}\boldsymbol{t}}{\nu}\right)^{-\frac{\nu+q}{2}} d\boldsymbol{t}. \end{aligned} \tag{2.10}$$

One major result in this context follows directly from the representation of the density function as the product of the two independent densities $f_q(\boldsymbol{t})$ and $h(y)$. It has been first published as an own result by Dunnett (1955). The following lemma says that every q–variate t–integral can be reduced to an inner q–variate normal integral combined with an outer chi integral.

Lemma 2.17.: Let $T_q(\boldsymbol{a},\boldsymbol{b})$ and $\boldsymbol{\Phi}_q(\boldsymbol{a},\boldsymbol{b})$ be the distribution functions of the mv*t*– and the mvn distribution, respectively. Denoting the chi density $h'(y;\ \nu) = \frac{1}{2^{\nu/2-1}\Gamma(\nu/2)} y^{\nu-1}e^{-y^2/2}$ the relation

$$T_q(\boldsymbol{a},\boldsymbol{b}) = \frac{1}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} \int_0^\infty y^{\nu-1} e^{-\frac{y^2}{2}} \boldsymbol{\Phi}_q\left(\boldsymbol{a}y/\sqrt{\nu},\ \boldsymbol{b}y/\sqrt{\nu};\boldsymbol{0},\boldsymbol{R}\right) dy \tag{2.11}$$

is valid.

Thus, by computing mv*t*–probabilities one could restrict to the approaches provided for the mvn distribution. This has actually been done frequently in the literature and it might be one reason for the low number of articles concerning a direct calculation. However, this method requires the evaluation of an additional integral and we therefore restrict ourselves to methodologies which approximate (2.10) directly.

Until recently a direct evaluation of the multiple integral for arbitrary correlation matrices was considered computationally infeasible. Referring to certain unbalanced designs, Hochberg and Tamhane (1987) noted with focus on the arbitrary correlation structure that "*even a computer program to calculate these critical points would be prohibitively costly to run for moderate to large* k ...". Similarly, quoting Hsu (1996), "... *the computation ... too slow for interactive data analysis except for ... (*$k \leq 3$*).*" Wang and Kennedy (1997) used interval analysis by applying the multivariate Taylor expansion to the density. But even low accuracy results required calculation times that were too large for practical purposes. Next, as in the case of the mvn distribution, the mv*t*–distribution also reduces to a series of lower order integrals if $\boldsymbol{R}$ has a special structure (for example the product correlation structure). But since this approach is limited to some restricted problems, we can not use this procedure for the more general problems considered here. In this context Hsu (1992) proposed to approximate a given arbitrary correlation matrix $\boldsymbol{R}$ by the 'closest' $\hat{\boldsymbol{R}}$, which satisfies the structural condition. This can be done by use of factor analytic methods (Hsu, 1992) or by linear programming techniques (Hsu and Nelson, 1998). Finally, Edwards and Berry (1987) introduced crude Monte Carlo for equation (2.10) in order to calculate critical constants for different multiple comparison procedures. Overviews, always strongly related with the mvn distribution, has been provided by Gupta (1963a, b) and Martynov (1981). Bretz (1999) also contains more recent developments. In the following we will present two approaches in more detail. Problems of evaluating (2.10) and computing the corresponding equicoordinate integration bounds for given probability values $1-\alpha$ will both be investigated in the following three subsections. Numerical comparisons and conclusions are given in Subsection 2.2.2.4.

2.2.2.1. The methodologies of Somerville

In a series of papers Somerville and Wang (1994) and Somerville (1997, 1998, 1999) proposed two different approaches of computing both mvn and mv*t*–probabilities. Here, we refer to the mv*t*–case only and describe the methodologies briefly. It is worth emphasising that these approaches are more general than actually required for the multiple integral given in Definition 2.4. They were designed for computing probability contents of general convex regions $A_q \subset \mathbb{R}^q$ instead of simple rectangular regions only.

Let $\boldsymbol{t} \sim t_q(\nu, \boldsymbol{R})$. With the Cholesky decomposition $\boldsymbol{R} = \boldsymbol{C}\boldsymbol{C}^t$ the transformation $\boldsymbol{t} = \boldsymbol{C}\boldsymbol{y}$ leads to independent spherically symmetric *t*–variates. The first strategy proposed by Somerville chooses unit random directions $\boldsymbol{d} = (d_1, \ldots, d_q) \in \mathbb{R}^q$. At each simulation step distances *r* from the origin to the boundary of A_q along the direction $\boldsymbol{d}$ are obtained. $P(F \le r^2/q)$ is then an unbiased estimate of the integral value, where $r^2/q = \boldsymbol{d}^t\boldsymbol{d}/q$ is an univariate $F_{\nu,q}$ –variable. These Monte Carlo evaluations are repeated until the average of estimates achieves a pre-specified standard error (Somerville and Wang, 1994).

The second, modified, proposal replaces the calculation of above *F*–probabilities by a binning procedure and a subsequent numerical quadrature. Denote by $r*$ the minimum distance from the origin to the boundary of A_q. The empirical cumulative distribution function and frequency function $H(r)$ and $h(r)$ are determined for distances $r \ge r*$ for repeated random directions $\boldsymbol{d} \in \mathbb{R}^q$. The probability content of A_q is yielded by the sum of the volume of the $(q-1)$–dimensional hypersphere of radius *r** and the integral $\int_{r*}^{\infty}(1-H(r))h(r)\,dr$. Hence, Monte Carlo evaluations are used to estimate $H(r)$ and $h(r)$. The resulting one-dimensional integral is calculated by applying standard numerical quadrature routines (Somerville, 1998).

The second approach is thought to behave especially well for computing critical values of multiple testing procedures, because $h(r)$ has to be obtained only once when applying an iterative procedure. Computation time of subsequent iterations involve therefore only the quadrature process. However, one disadvantage is the requirement $\boldsymbol{0} \in A_q$. Otherwise, the present method does not work and an acceptance rejection algorithm is conducted.

2.2.2.2. Transformations of Genz

In analogy to the mvn case in Subsection 2.1.2.2. Genz and Bretz (1999) established a series of substitutions, which reduce the original integral (2.10) to one over the $(q-1)$–dimensional unit hypercube. In this form several numerical procedures can be used for the final evaluation, and some of these procedures are presented as algorithms afterwards. But before introducing their result, two helpful lemmas for the main proof will be given briefly.

Lemma 2.18.: For $\nu > 0$, $q \in \mathbb{N}$ and $\boldsymbol{y} = \left(y_1, \ldots, y_q\right)^t \in \mathbb{R}^q$

$$1 + \frac{y_1^2 + y_2^2 + \ldots + y_q^2}{\nu} = \left(1 + \frac{y_1^2}{\nu}\right) \cdot \left(1 + \frac{y_2^2}{\nu + y_1^2}\right) \cdot \ldots \cdot \left(1 + \frac{y_q^2}{\nu + y_1^2 + \ldots + y_{q-1}^2}\right).$$

Proof: The proof is done by induction over q. The algebraic transformations

$$\left(1+\tfrac{y_1^2}{\nu}\right)\left(1+\tfrac{y_2^2}{\nu+y_1^2}\right) = 1+\tfrac{y_1^2}{\nu}\tfrac{\nu+y_1^2}{\nu+y_1^2} + \tfrac{y_2^2}{\nu+y_1^2}\tfrac{\nu}{\nu} + \tfrac{y_1^2 y_2^2}{\nu\left(\nu+y_1^2\right)} = 1+\tfrac{y_1^2\left(\nu+y_1^2\right)+y_2^2\left(\nu+y_1^2\right)}{\nu\left(\nu+y_1^2\right)} = 1+\tfrac{y_1^2+y_2^2}{\nu}$$

yield the assertion for the induction begin $q = 2$. For the general induction step $q \to q+1$ we assume the assertion to be true for q. Then we have to show the validity for $q+1$. To start with, we set $s_i = y_1^2 + \ldots + y_i^2$ for $s_0 = 0$, $i = 1, \ldots, q+1$. Hence,

$$\begin{aligned}
\prod_{i=1}^{q+1}\left(1+\tfrac{s_i - s_{i-1}}{\nu + s_{i-1}}\right) &= \left(1+\tfrac{s_q}{\nu}\right)\left(1+\tfrac{s_{q+1}-s_q}{\nu+s_q}\right) = \\
&= 1 + \tfrac{s_q}{\nu} + \tfrac{s_{q+1}-s_q}{\nu+s_q} + \tfrac{s_q}{\nu}\cdot\tfrac{s_{q+1}-s_q}{\nu+s_q} = \\
&= 1 + \tfrac{s_q\left(\nu+s_q\right)+\left(s_{q+1}-s_q\right)\nu+\left(s_{q+1}-s_q\right)s_q}{\left(\nu+s_q\right)\nu} = \\
&= 1 + \tfrac{s_q\left(\nu+s_q\right)+\left(s_{q+1}-s_q\right)\left(\nu+s_q\right)}{\left(\nu+s_q\right)\nu} = 1 + \tfrac{s_{q+1}}{\nu}
\end{aligned}$$

and the assertion follows.

Lemma 2.19.: For $\nu > 0$, $q \in \mathbb{N}$ set $K_\nu^{(q)} = \frac{\Gamma\left(\frac{\nu+q}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\nu\pi)^q}}$. Then

$$K_\nu^{(q)} \sqrt{\prod_{j=0}^{q-1} \frac{\nu}{\nu+j}} = \prod_{j=0}^{q-1} K_{\nu+j}^{(1)} .$$

Proof: The assertion follows from

$$\begin{aligned}
\sqrt{\prod_{j=0}^{q-1} \frac{\nu+j}{\nu}} \prod_{j=0}^{q-1} K_{\nu+j}^{(1)} &= \sqrt{\prod_{j=0}^{q-1} \frac{\nu+j}{\nu}} \prod_{j=0}^{q-1} \frac{\Gamma\left(\frac{\nu+1+j}{2}\right)}{\Gamma\left(\frac{\nu+j}{2}\right)\sqrt{(\nu+j)\pi}} = \\
&= \sqrt{\prod_{j=0}^{q-1} \frac{\nu+j}{\nu}} \frac{\Gamma\left(\frac{\nu+q}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\prod_{j=0}^{q-1}\sqrt{(\nu+j)\pi}} = \\
&= \frac{\Gamma\left(\frac{\nu+q}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{\prod_{j=0}^{q-1}\sqrt{(\nu+j)}}{\sqrt{\nu^q}} \frac{1}{\sqrt{\pi^q}\prod_{j=0}^{q-1}\sqrt{(\nu+j)}} = \\
&= \frac{\Gamma\left(\frac{\nu+q}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\nu\pi)^q}} = K_\nu^{(q)}.
\end{aligned}$$

For the result of Genz and Bretz (1999) assume now that in equation (2.10) for each *i* at least a_i or b_i is finite, because otherwise an appropriate transformation reduces the dimension of the problem by integrating the i^{th} variable explicitly. Then the following general result is valid.

Theorem 2.3.: Let $\boldsymbol{a} = (a_1, \ldots, a_q) \in \mathbb{R}^q$ and $\boldsymbol{b} = (b_1, \ldots, b_q) \in \mathbb{R}^q$ arbitrary and $T_q(\boldsymbol{a}, \boldsymbol{b})$ of Definition 2.4 be given. Further, denote by $\boldsymbol{R} = \boldsymbol{C}\boldsymbol{C}^t$ the Cholesky decomposition of $\boldsymbol{R}$, where $\boldsymbol{C} = (c_{ij})_{ij}$ is a lower triangular matrix. Then for $q > 1$

$$T_q(\boldsymbol{a}, \boldsymbol{b}) = \underbrace{\int_0^1 \int_0^1 \ldots \int_0^1}_{q-1 \text{ integrals}} f(\boldsymbol{w})\, d\boldsymbol{w}, \qquad (2.12)$$

where $f(\boldsymbol{w}) = (e_1 - d_1) \cdot \ldots \cdot (e_q - d_q)$. Further on, $e_1 = t_\nu(b_1/c_{11})$ and for $i = 1, \ldots, q-1$, $e_{i+1} = t_{\nu+i}\left(\left(b_{i+1} - \sum_{j=1}^{i} c_{i+1,j} Y_j\right)\sqrt{\frac{\nu+i}{\nu+\sum_{j=1}^{i} Y_j^2}} \Big/ c_{i+1,i+1}\right)$ where $Y_i = t_{\nu+i-1}^{-1}\left(d_i + W_i(e_i - d_i)\right)\sqrt{\frac{\nu+\sum_{j=1}^{i-1} Y_j^2}{\nu+i-1}}$ and $\boldsymbol{w} = (W_1, \ldots, W_{q-1}) \in \mathbb{R}^{q-1}$. The d_i's are defined correspondingly for the lower bound $\boldsymbol{a}$.

Proof: Consider the original multiple integral (2.10). In the following we shall describe four substitutions in order to transform it to the representation (2.12). If we let $\boldsymbol{t} = \boldsymbol{C}\boldsymbol{y}$, then $\boldsymbol{t}^t\boldsymbol{R}^{-1}\boldsymbol{t} = \boldsymbol{y}^t\boldsymbol{C}^t\boldsymbol{C}^{-t}\boldsymbol{C}^{-1}\boldsymbol{C}\boldsymbol{y} = \boldsymbol{y}^t\boldsymbol{y}$ and $d\boldsymbol{t} = |\boldsymbol{C}|d\boldsymbol{y} = |\boldsymbol{R}|^{½}d\boldsymbol{y}$. Hence, for $K_\nu^{(q)} = \frac{\Gamma\left(\frac{\nu+q}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\nu\pi)^q}}$ we obtain

$$T_q(\boldsymbol{a},\boldsymbol{b}) = K_\nu^{(q)} \int_{a_1'}^{b_1'} \int_{a_2'}^{b_2'} \dots \int_{a_q'}^{b_q'} \left(1+\frac{\boldsymbol{y}^t\boldsymbol{y}}{\nu}\right)^{-\frac{\nu+q}{2}} d\boldsymbol{y},$$

with $a_i' = \left(a_i - \sum_{j=1}^{i-1} c_{ij}Y_j\right)\Big/ c_{ii}$ and $b_i' = \left(b_i - \sum_{j=1}^{i-1} c_{ij}Y_j\right)\Big/ c_{ii}$. We now split the integrand into a product of q factors by use of

$$\left(1+\frac{\boldsymbol{y}^t\boldsymbol{y}}{\nu}\right)^{-\frac{\nu+q}{2}} = \prod_{i=1}^{q}\left(1+\frac{Y_i^2}{\nu+\sum_{j=1}^{i-1}Y_j^2}\right)^{-\frac{\nu+q}{2}},$$

as proven in Lemma 2.18. When applied to $T_q(\boldsymbol{a},\boldsymbol{b})$ this transformation yields

$$T_q(\boldsymbol{a},\boldsymbol{b}) = K_\nu^{(q)} \int_{a_1'}^{b_1'} \left(1+\frac{Y_1^2}{\nu}\right)^{-\frac{\nu+q}{2}} \dots \int_{a_q'}^{b_q'} \left(1+\frac{Y_q^2}{\nu+Y_1^2+\dots+Y_{q-1}^2}\right)^{-\frac{\nu+q}{2}} d\boldsymbol{y}.$$

At this stage each of the Y_i's can be substituted using $Y_i = U_i\sqrt{\frac{\nu+\sum_{j=1}^{i-1}Y_j^2}{\nu+i-1}}$. Beginning with $i = q$ the q-fold substitution results in

$$T_q(\boldsymbol{a},\boldsymbol{b}) = K_\nu^{(q)}\sqrt{\frac{\nu}{\nu+1}\cdot\frac{\nu}{\nu+2}\cdot\ldots\cdot\frac{\nu}{\nu+q-1}} \int_{\hat{a}_1}^{\hat{b}_1}\left(1+\frac{U_1^2}{\nu}\right)^{-\frac{\nu+1}{2}} \dots \int_{\hat{a}_q}^{\hat{b}_q}\left(1+\frac{U_q^2}{\nu+q-1}\right)^{-\frac{\nu+q}{2}} d\boldsymbol{u} =$$

$$= K_\nu^{(1)} \int_{\hat{a}_1}^{\hat{b}_1}\left(1+\frac{U_1^2}{\nu}\right)^{-\frac{\nu+1}{2}} \dots K_{\nu+q-1}^{(1)} \int_{\hat{a}_q}^{\hat{b}_q}\left(1+\frac{U_q^2}{\nu+q-1}\right)^{-\frac{\nu+q}{2}} d\boldsymbol{u},$$

where $K_\nu^{(q)}\sqrt{\prod_{j=0}^{q-1}\frac{\nu}{\nu+j}} = \prod_{j=0}^{q-1}K_{\nu+j}^{(1)}$ follows from Lemma 2.19., and $\hat{a}_i = a_i'\sqrt{\frac{\nu+i-1}{\nu+\sum_{j=1}^{i-1}Y_j^2}}$ and $\hat{b}_i = b_i'\sqrt{\frac{\nu+i-1}{\nu+\sum_{j=1}^{i-1}Y_j^2}}$ are the new integration limits. For the last steps we set $U_i = t_{\nu+i-1}^{-1}(Z_i)$,

where $t_\nu(u) = K_\nu^{(1)} \int_{-\infty}^{u} \left(1+\frac{s^2}{\nu}\right)^{-\frac{\nu+1}{2}} ds$ denotes the univariate *t*-distribution with ν degrees of freedom. Using $dZ_i = K_{\nu+i-1}^{(1)} \left(1+\frac{U_i^2}{\nu+i-1}\right)^{-\frac{\nu+i}{2}} dU_i$ equation (2.10) becomes

$$
\begin{aligned}
T_q(\boldsymbol{a},\boldsymbol{b}) &= \int_{d_1}^{e_1} \int_{d_2}^{e_2} \dots \int_{d_q}^{e_q} d\boldsymbol{z} = \\
&= (e_1 - d_1) \int_0^1 (e_2 - d_2) \dots \int_0^1 (e_q - d_q) \int_0^1 d\boldsymbol{w} = \\
&= \underbrace{\int_0^1 \int_0^1 \dots \int_0^1}_{q-1 \text{ integrals}} f(\boldsymbol{w})\, d\boldsymbol{w},
\end{aligned}
$$

where $d_i = t_{\nu+i-1}(\hat{a}_i)$, $e_i = t_{\nu+i-1}(\hat{b}_i)$ and $Z_i = d_i + W_i(e_i - d_i)$. With the sequence of transformations described here the assertion follows for $f(\boldsymbol{w}) = (e_1 - d_1) \cdot \dots \cdot (e_q - d_q)$.

On the following pages we present three numerical algorithms that use equation (2.12) to estimate $T_q(\boldsymbol{a},\boldsymbol{b})$ for a given error requirement ε. All of them were considered by Genz and Bretz (1999). For the last algorithm, a lattice rule algorithm, the theoretical research is still ongoing. The other two algorithms are an acceptance-rejection sampling and a crude Monte Carlo algorithm. These methods are simpler and well known to be reliable. Other Monte Carlo techniques could be applied but will not be analysed here. Deak (1990) gives a good overview with application to the multinormal case.

Acceptance-rejection method

This procedure generates *q*-dimensional uniform random vectors $\boldsymbol{w}_1$, ..., $\boldsymbol{w}_N$ and estimates $T_q(\boldsymbol{a},\boldsymbol{b})$ using

$$
\overline{T}_{AR}(\boldsymbol{a},\boldsymbol{b}) = \frac{1}{N} \sum_{l=1}^{N} g(\boldsymbol{C}\boldsymbol{y}_l),
$$

with $g(\boldsymbol{x}) = \begin{cases} 1, & a_i \le x_i \le b_i \\ 0, & \text{otherwise} \end{cases}$ and $Y_{li} = t^{-1}_{\nu+i-1}(W_{li})\sqrt{\frac{\nu+\sum_{j=1}^{i-1} Y_j^2}{\nu+i-1}}$, $i = 1, ..., q$, $l = 1, ..., N$. To control the simulated error we make use of the usual error estimate of the means

$$s_{\overline{T}_{AR}} = \frac{s}{\sqrt{N}} = \sqrt{\frac{\sum \left(g - \overline{T}_{AR}\right)^2}{N(N-1)}}.$$

Furthermore, we denote by γ the Monte Carlo confidence factor for the standard error. If, for example, $\gamma = 3$, we then expect the actual error of $\overline{T}_{AR}$ to be less than the error bound ε in 99.7% of the cases. Next, a corresponding algorithm implementing this procedure is given.

1. **INPUT** q, γ, ν, $\boldsymbol{R}$, $\boldsymbol{a}$, $\boldsymbol{b}$, ε.
2. Compute lower triangular Cholesky factor $\boldsymbol{C}$ for $\boldsymbol{R}$.
3. Initialise $N = 0$, $Intval = 0$, $Varsum = 0$.
4. **REPEAT**
   a) Generate uniform random $W_1, ..., W_q \in [0, 1]$.
   b) Set $f = 1$.
   c) **FOR** $i = 1, 2, ..., q$

      Set $Y_i = t^{-1}_{\nu+i-1}(W_i)\sqrt{\frac{\nu+\sum_{j=1}^{i-1} Y_j^2}{\nu+i-1}}$.

      IF $\sum_{j=1}^{i} c_{ij} Y_j < a_i$ **OR** $\sum_{j=1}^{i} c_{ij} Y_j > b_i$ **THEN** go to step d).

      END FOR

      Go to step e).
   d) Set $f = 0$.
   e) Set $N = N + 1$,

      $Varsum = Varsum + (N - 1)(f - Intval)^2/N$,

      $Intval = Intval + (f - Intval)/N$,

      $ErrEst = \gamma\sqrt{Varsum/(N(N-1))}$.

   UNTIL $ErrEst < \varepsilon$.
5. **OUTPUT** *Intval, ErrEst, N.*

The acceptance-rejection method is widely used and may be the most intuitive way to deal with equation (2.12). Deak (1990), however, showed that among the various Monte Carlo methods it is the one with the worst efficiency and therefore other approaches to evaluate $T_q(\boldsymbol{a},\boldsymbol{b})$ are needed.

Monte Carlo method

For estimating (2.12) by the crude Monte Carlo method we let $\boldsymbol{w}_1$, ..., $\boldsymbol{w}_N$ be uniformly and independently distributed on $[\boldsymbol{0}, \boldsymbol{1}]^{q-1}$. Then the random variables $f(\boldsymbol{w}_l)$, l = 1, ..., N, are independent and their expected value is $E[f(\boldsymbol{w})] = \int_0^1 \dots \int_0^1 f(\boldsymbol{w})\, d\boldsymbol{w} = T_q(\boldsymbol{a},\boldsymbol{b})$. Consequently the arithmetic average

$$\overline{T}_{MC}(\boldsymbol{a},\boldsymbol{b}) = \frac{1}{N}\sum_{l=1}^{N} f(\boldsymbol{w}_l) \qquad (2.13)$$

is an unbiased estimator of the integral $T_q(\boldsymbol{a},\boldsymbol{b})$.

1. **INPUT** ***a, b, w,*** ν**,** ***C***.
2. Initialise $d_1 = t_\nu(a_1/c_{11})$, $e_1 = t_\nu(b_1/c_{11})$, $f_1 = e_1 - d_1$.
3. **FOR** i = 1, 2, ..., q - 1

   Set $Y_i = t_{\nu+i-1}^{-1}\left(d_i + W_i(e_i - d_i)\right)\sqrt{\frac{\nu+\sum_{j=1}^{i-1}Y_j^2}{\nu+i-1}}$,

   $$d_{i+1} = t_{\nu+i}\left(\left(a_{i+1} - \sum_{j=1}^{i} c_{i+1,j}Y_j\right)\sqrt{\frac{\nu+i}{\nu+\sum_{j=1}^{i}Y_j^2}}\Big/ c_{i+1,i+1}\right),$$

   $$e_{i+1} = t_{\nu+i}\left(\left(b_{i+1} - \sum_{j=1}^{i} c_{i+1,j}Y_j\right)\sqrt{\frac{\nu+i}{\nu+\sum_{j=1}^{i}Y_j^2}}\Big/ c_{i+1,i+1}\right),$$

   $f_{i+1} = (e_{i+1} - d_{i+1})f_i$.

   END FOR
4. **OUTPUT** $f(\boldsymbol{w}) = f_q$.

The Monte Carlo and lattice rule algorithms to be described in this subsection both require the evaluation of $f(\boldsymbol{w})$ for particular values of $\boldsymbol{w}$, so we provide above an algorithm for $f(\boldsymbol{w})$ that will be used by both of the numerical integration algorithms. In the algorithm given on the previous page note that the initialisations of d_1 and e_1 are required only for the first evaluation of $f(\boldsymbol{w})$ (assuming $\boldsymbol{a}$, $\boldsymbol{b}$, ν, and $\boldsymbol{C}$ are fixed for a particular integral), and we would avoid wasteful computation of t-values by setting $d_i = 0$ or $e_i = 1$ if $a_i = -\infty$ or $b_i = \infty$, respectively. In the following we give the algorithm when applying crude Monte Carlo on equation (2.12).

1. **INPUT** q, γ, ν, $\boldsymbol{R}$, $\boldsymbol{a}$, $\boldsymbol{b}$, ε.
2. Compute lower triangular Cholesky factor $\boldsymbol{C}$ for $\boldsymbol{R}$.
3. Initialise $N = 0$, $Intval = 0$, $Varsum = 0$.
4. **REPEAT**
   a) Generate uniform random $W_1, \ldots, W_{q-1} \in [0, 1]$.
   b) Evaluate $f_q = f(\boldsymbol{w})$.
   c) Set $N = N + 1$,
   $Varsum = Varsum + (N-1)(f_q - Intval)^2/N$,
   $Intval = Intval + (f_q - Intval)/N$,
   $ErrEst = \gamma\sqrt{Varsum/(N(N-1))}$.
   UNTIL $ErrEst < \varepsilon$.
5. **OUTPUT** *Intval, ErrEst, N.*

Randomised lattice rule method

As seen from the central limit theorem, the Monte Carlo integration yields a probabilistic error bound in $O(N^{-1/2})$ (note that the order of magnitude does not depend on q). This means halving the error requires quadrupling the number of sample points. However, Quasi-Monte Carlo integration methods use sequences of nodes that are designed to be more uniform than random, while still using an integration formula with equal weights similar to (2.13). Niederreiter (1992) and others showed that under suitable conditions a deterministic error

bound is given by $O(N^{-1}\log^{q-1}N)$. In the sequel we will make use of one approach out of this broad class, the so-called randomised lattice rule (Joe, 1990; Sloan and Joe, 1994, p. 170)

$$\overline{T}_L = \frac{1}{N}\sum_{l=1}^{N} T_{L,l} = \frac{1}{Np}\sum_{l=1}^{N}\sum_{j=1}^{p} f\left(\left|2\left\{\frac{j}{p}\boldsymbol{z} + \boldsymbol{w}_l\right\} - 1\right|\right). \tag{2.14}$$

Here, N is the simulation size, usually being very small (e.g. 10 - 20), p corresponds to the fineness of the lattice and $\boldsymbol{z} \in \mathbb{R}^{q-1}$ denotes the strategically chosen lattice vector. Braces around vectors indicate that each component has to be replaced by its fractional part. Finally, $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N$ denote again $[\boldsymbol{0}, \boldsymbol{1}]^{q-1}$–uniform random vectors. The error estimate for $\overline{T}_L$ is

$$s_{\overline{T}_L} = \sqrt{\frac{\sum\left(T_{L,l} - \overline{T}_L\right)^2}{N(N-1)}}.$$

For more details concerning lattice rules in general and the approach used here the reader is referred to the book of Sloan and Joe (1994).

The algorithm listed below consists basically of two loops. The *l*-loop stands for the outer sum in equation (2.14), and *j* is the variable of the inner sum. This is repeated several times until a pre-assigned error level ε is reached. At each step the counting variable *n* is incremented by 1 and the number of lattice points $p = p_n$ is increased in dependence on *n*.

The best choice of $\boldsymbol{z}$ is still an open research question. Several proposals have been published in the literature, the most common of which is to choose $\boldsymbol{z}$ of the form $\boldsymbol{z}(h) = \left(1, h, h^2 \bmod p, \ldots, h^{q-2} \bmod p\right)$, $1 \le h \le \lfloor p/2 \rfloor$, proposed by Korobov (1960). This leaves us with the problem of how to choose *h*. The method we used minimises

$$P_{q,p} = \frac{1}{p}\sum_{j=1}^{p}\prod_{i=1}^{q-1}\tilde{F}\left(\left\{\frac{j}{p}z_i\right\}\right), \tag{2.15}$$

with $\tilde{F}(x) = F(x)/4 + ¾$ and $F(x) = 1 + 2\pi^2(x^2 - x + 1/6)$, $x \in [0, 1]$ (see Sloan and Joe, 1994, p. 173). FORTRAN and SAS/IML implementations of (2.15) are provided in the Appendix.

The calculation of $\boldsymbol{z}$ or h can be done independently from the actual program. Therefore, lattice rule implementations contain only a matrix consisting of appropriate values for h up to a certain maximum dimension.

1. **INPUT** q, γ, ν, $\boldsymbol{R}$, $\boldsymbol{a}$, $\boldsymbol{b}$, ε.
2. Compute lower triangular Cholesky factor $\boldsymbol{C}$ for $\boldsymbol{R}$.
3. Initialise $N = 10$, $n = 0$.
4. **REPEAT**
   a) Set $n = n + 1$, $Intval = 0$, $Varsum = 0$.
   b) **FOR** $l = 1, 2, ..., N$
      i) Set $Latsum = 0$.
      ii) Generate uniform random $W_1, ..., W_{q-1} \in [0, 1]$.
      iii) **FOR** $j = 1, 2, ..., p_n$
         - Set $\boldsymbol{w} = \left| 2\left\{\boldsymbol{w} + \frac{j}{p_n}\boldsymbol{z}\right\} - 1 \right|$.
         - Evaluate $f_q = f(\boldsymbol{w})$.
         - Set $Latsum = Latsum + (f_q - Latsum)/j$.

      END FOR
      iv) Set $Varsum = Varsum + (l - 1)(Latsum - Intval)^2/l$,
      $Intval = Intval + (Latsum - Intval)/l$.
   END FOR
   c) Set $ErrEst = \gamma\sqrt{Varsum/(N(N-1))}$.
   UNTIL $ErrEst < \varepsilon$.
5. **OUTPUT** $Intval$, $ErrEst$, Np_n.

2.2.2.3. Computation of equicoordinate quantiles

In order to illustrate the use of the transformation method presented in the preceding subsection for evaluating multivariate t-probabilities, we consider one important application, the computation of critical values for multiple comparison procedures (see, for example, Edwards and Berry, 1987, or Hsu and Nelson, 1998). We will encounter this problem at least

twice during the course of the present work. On the one side, we require their calculation for the power expression of MCTs derived in Chapter 4. Next, we need the critical values for computing simultaneous confidence intervals as proposed in Chapter 7. With these problems, we are given a desired confidence level α and we need to determine an upper limit vector $\boldsymbol{t}$ = (t, ..., t) $\in \mathbb{R}^q$, so that $T_q(-\infty, \boldsymbol{t}) = 1-\alpha$. We will illustrate the use of the algorithms with the Dunnett contrast defined in Example 1.3. For k = 3 we obtain q = 3 and a correlation matrix $\boldsymbol{R}$ given by

$$\boldsymbol{R} = \begin{pmatrix} 1 & 0.3636 & 0.3636 \\ 0.3636 & 1 & 0.3636 \\ 0.3636 & 0.3636 & 1 \end{pmatrix}$$

for the sample size allocation (14, 8, 8, 8). If we want to apply our algorithms on this problem, we need to combine our algorithms with an iterated-nonlinear-equation-solving algorithm. We let $h(\boldsymbol{t}) = T_q(-\infty, \boldsymbol{t}) - 1 + \alpha$, so that we need to find $\boldsymbol{t}$ such that $h(\boldsymbol{t}) = 0$. At this point different root finding methods may be applied and we have successfully used various modified secant algorithms for this problem. The results given below were produced by the Pegasus method (see Ralston and Rabinowitz, 1978, p. 341). Using this algorithm we need a starting interval $[\boldsymbol{b}', \boldsymbol{b}'']$ which contains our desired solution $\boldsymbol{t}$, where $\boldsymbol{b}' = (b', \ldots, b')$ and $\boldsymbol{b}'' = (b'', \ldots, b'')$. At each stage an estimated $\boldsymbol{t}$ is computed, along with $h(\boldsymbol{t})$, and a new interval is produced. In order to simplify the start of the algorithm we used $b' = -4$ and $b'' = 4$, assuming $h(\boldsymbol{b}') = \alpha - 1$ and $h(\boldsymbol{b}'') = \alpha$. Table 2.1. shows the behaviour of our randomised lattice rule algorithm ($\varepsilon_L = 0.001$, $\gamma = 3$) combined with the Pegasus method with termination if $|\boldsymbol{b}' - \boldsymbol{b}''| < 0.01 = \varepsilon_P$ in each of the q components, for $\alpha = 0.05$. The final t value was t = 2.1664, apparently correct to at least four digits, and this required a total of 22144 $f(\boldsymbol{w})$–values.

A commonly used algorithm for the confidence interval problem is a type of rejection algorithm (see Edwards and Berry, 1987), which we will call %-rejection. The basic idea is to generate a large number, say N, of random vectors $\boldsymbol{x}_l \sim T(.,.)$ for given $\boldsymbol{R}$ and ν. For each $\boldsymbol{x}_l$ let $t_l = \max_{1 \le i \le q}(x_{li})$. Then sort the t_l's and let $t = t_{(r)}$, where $r = (N+1)(1-\alpha)$. The result is that

Step i	a_i	t_i	b_i	$h(t_i)$	Number of $f(\boldsymbol{w})$ values
1	–4.0000	3.6000	4.0000	0.0485	2768
2	–4.0000	2.9045	3.6000	0.0409	2768
3	–4.0000	1.9709	2.9045	–0.0244	2768
4	1.9709	2.3196	2.9045	0.0140	2768
5	1.9709	2.1924	2.3196	0.0027	2768
6	1.9709	2.1669	2.1924	0.0001	2768
7	1.9709	2.1664	2.1669	0.0000	2768
8	1.9709	2.1664	2.1664	0.0000	2768
final	2.1664	2.1664	2.1664		22144

Table 2.1. Iteration steps for the Pegasus method using the lattice rule algorithm $(\varepsilon_P = 0.01)$.

after sorting we reject $(1-\alpha)\%$ of the (smallest) t_l's and pick the smallest one left as an estimator for *t*. Several refinements to this basic algorithm have been suggested (see Hsu and Nelson, 1990, 1998, for discussion and further references), but we have found that rejection type algorithms are generally not very efficient for computing multivariate *t*-probabilities, and it is beyond the scope of this work to provide a detailed comparison between our algorithms and %-rejection algorithms for this problem. However, we will describe how our algorithms could be used for %-rejection, too. The $\boldsymbol{x}_l$ vectors needed for %-rejection are typically generated using a combination of multivariate normal and chi variables, but our transformations can also be used to generate $\boldsymbol{x}_l$ vectors with correct distribution. The acceptance-rejection algorithm given in the last subsection can be easily modified to provide a simple %-rejection algorithm. Using this algorithm with $N = 22139$, so that $r = (N + 1)(1 - \alpha) = 21033$, we found $t = 2.1725$. A standard 99% confidence interval obtained using a normal approximation to the binomial distribution is given by $\left[t_{(20995)}, t_{(21072)}\right] = \left[2.1523, 2.1931\right]$.

1. **INPUT** N, α, ν, $\boldsymbol{R}$.
2. Compute lower triangular Cholesky factor $\boldsymbol{C}$ for $\boldsymbol{R}$.
3. **FOR** $l = 1, 2, ..., N$
   a) Generate uniform random $W_1, ..., W_q \in [0, 1]$.
   b) **FOR** $i = 1, 2, ..., q$

$$\text{Set } Y_i = t^{-1}_{\nu+i-1}(W_i)\sqrt{\frac{\nu+\sum_{j=1}^{i-1}Y_j^2}{\nu+i-1}},$$

$$X_{li} = \sum_{j=1}^{i} c_{ij}Y_j .$$

   END FOR

   c) Set $t_l = \max_{1\le i\le q}(X_{li})$.
4. Sort the t_l's.
5. **OUTPUT** $t = t_{(r)}$, $r = (N+1)(1-\alpha)$.

2.2.2.4. Numerical comparisons and conclusions

We finish this chapter by providing a few numerical comparisons of the new techniques introduced above with the emphasis given on the mv*t*–distribution. In particular, on the preceding pages we managed to solve the following two problems:

(A) computation of $T_q(\boldsymbol{a, b; 0, R})$;
(B) for given α find $\boldsymbol{t} = (t, ..., t) \in \mathbb{R}^q$, so that $T_q(-\infty, \boldsymbol{b; 0, R}) = 1 - \alpha$.

In Chapter 4 we will additionally be confronted with the problem of

(C) evaluating $T_q(\boldsymbol{a, b; \mu, R})$,

the cumulative distribution function of the ***non***–central mv*t*–distribution. But we leave this topic for further investigations in Chapter 4 and summarise briefly the results obtained so far for (A) and (B).

For the solution of (A) a series of transformation led to the evaluation of a continuous function over the unit hypercube. Three standard numerical algorithms have been applied, to be precise

- an acceptance-rejection method (AR),
- a crude Monte Carlo version (MC), and
- an implementation of the randomised lattice rules (LR).

SAS/IML programs of all three approaches are available in the Appendix. The AR and the MC versions given there run for arbitrary dimensions q, the LR implementation is restricted to $q = 32$. For higher dimensions required, the user is requested to run B.2.5 or B.2.6 by himself. Thus, one may obtain the required generating values of h and include them in the original LR program. In order to give an idea of their behaviour all three algorithms were used for the following brief comparison to other approaches published in the literature. The other methodologies considered are both approaches of Somerville (mvi and mvib according to the representation given in Subsection 2.2.2.1.) and an adaptation of the Solow method to the mv$t$–case by use of equation (2.11), i.e. the reduction of the mv$t$–distribution to the mvn distribution. Because of the characteristics of the Solow procedure (c.f. Subsection 2.1.2.1.), only equicoordinate upper integration bounds were considered. Tables 2.2. and 2.3. give the average results of 100 runs with randomly generated correlation matrices $\boldsymbol{R}$ and integration bounds $\boldsymbol{b}$ according to Marsaglia and Olkin (1984).

The first table was obtained for equicorrelated $\boldsymbol{R}$'s (i.e. $\rho_{ij} = \rho \; \forall i \neq j$), because a benchmark is then given by the reduction method of Dunnett (1955) (trivially the product correlation structure holds in such cases, see also Remark 1.1.). For an error level set as $\varepsilon = 0.001$ the three entries of each cell shows the average error, the maximum error and the average time required for each integral on a P200. It becomes clear that the lattice rule implementation is better than the competitors for both time and accuracy considerations. For example, less than 10 seconds are required to calculate a seven-dimensional mv$t$–integral in SAS/IML for an actual accuracy of about four digits in the equicorrelated case. The Solow procedure is also very fast but it lacks reliability for increasing dimensions q. Both methodologies of Somerville compare well with the AR and MC methods, but are still somewhat slower than the LRs.

q	AR	MC	LR	Solow	mvi	mvib
	.0008	**.0003**	**.0000**	**.0005**	**.0004**	**.0005**
3	.0032	.0009	.0000	.0010	.0013	.0025
	104.11	80.13	2.32	1.06	60.76	111.31
	.0008	**.0003**	**.0000**	**.0008**	**.0004**	**.0004**
4	.0041	.0010	.0001	.0023	.0018	.0017
	115.83	110.40	3.41	2.11	71.27	127.80
	.0007	**.0002**	**.0000**	**.0013**	**.0004**	**.0005**
5	.0046	.0010	.0002	.0034	.0020	.0036
	125.32	171.60	4.55	3.39	73.86	128.55
	.0008	**.0002**	**.0000**	**.0019**	**.0004**	**.0004**
7	.0043	.0013	.0005	.0069	.0019	.0017
	150.32	252.52	8.04	6.97	87.47	149.78
	.0007	**.0002**	**.0001**	**.0022**	**.0005**	**.0005**
10	.0041	.0021	.0006	.0114	.0021	.0036
	176.57	321.40	12.02	14.98	111.56	171.64

Table 2.2. Average values of $T_q(-\infty, \boldsymbol{b}; \boldsymbol{0}, \boldsymbol{R})$ over 100 random runs, equicorrelated $\boldsymbol{R}$. Top entry: average error; middle: max error; below: average time in seconds; $\varepsilon = 0.001$ (P200, SAS/IML).

The results do not change much when considering the general case of arbitrary $\boldsymbol{R}$'s, which also included ill conditioned ones with absolute correlation entries close to 1. As a 'benchmark' we considered an implementation of lattice rules with $\varepsilon = 0.0001$. The LR algorithms is still the fastest, even if its superiority vanishes for higher dimensions. We would require for example less than a minute to calculate a seven dimensional integral with an actual error clearly less than 10^{-3}. However, this is in accordance with recent research results. They show that in general the time advantages for Quasi–Monte Carlo methods are especially attractive for lower to moderate dimensions for a moderate accuracy demanded a-priori. But quoting Berger (1991), "*... for statistical problems ... two significant digit accuracy typically suffices, and only rarely more than three ... needed.*" we feel confident to having achieved these goals for the present problem of computing mv*t*–probabilities at least up to dimension 10 within SAS and at least $q = 20$ within FORTRAN.

q	AR	MC	LR	Solow	mvi	mvib
	.0009	**.0003**	**.0001**	**.0013**	**.0005**	**.0005**
3	.0036	.0012	.0006	.0097	.0023	.0029
	103.69	103.81	2.64	1.05	68.98	131.30
	.0010	**.0003**	**.0001**	**.0020**	**.0004**	**.0004**
4	.0035	.0018	.0007	.0099	.0018	.0027
	123.32	165.51	7.92	2.07	79.66	152.52
	.0008	**.0003**	**.0002**	**.0024**	**.0004**	**.0004**
5	.0043	.0022	.0009	.0101	.0017	.0032
	119.19	192.03	18.08	3.13	81.18	137.96
	.0008	**.0003**	**.0002**	**.0042**	**.0004**	**.0005**
7	.0039	.0017	.0012	.0169	.0018	.0029
	150.37	298.12	59.16	7.16	102.27	168.42
	.0007	**.0003**	**.0002**	**.0077**	**.0004**	**.0006**
10	.0034	.0015	.0018	.0257	.0018	.0036
	165.33	420.59	123.96	15.34	114.39	183.59

Table 2.3. Average values of $T_q(-\infty, \boldsymbol{b}; \boldsymbol{0}, \boldsymbol{R})$ over 100 random runs, arbitrary $\boldsymbol{R}$. Top entry: average error; middle: max error; below: average time in seconds; $\varepsilon = 0.001$ (P200, SAS/IML).

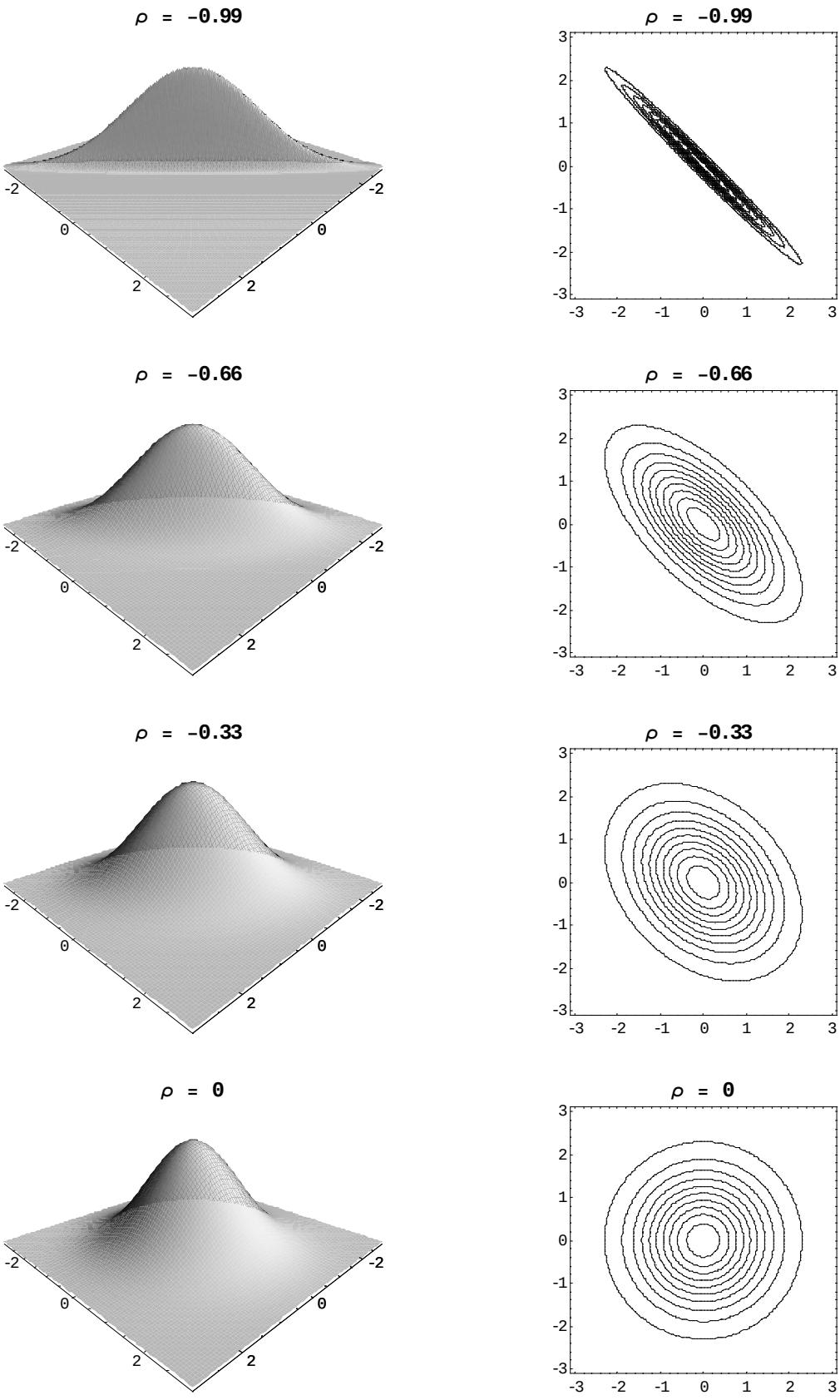
Generalisations to the mvn case hold on. Applying the AR, MC and LR methodologies on the transformed integral (2.7) leads to efficient implementations. Because of its superiority demonstrated in the mv*t*–case, only the algorithm of the lattice rules has been provided previously in Subsection 2.1.2.2. Generalisations for the AR and MC methods are, however, straight forward. As a final comment, recall that Somervilles approaches were designed for general convex regions, with emphasis given on a fast and accurate computation of critical values in the special case of mvib. Above comparisons should therefore always be considered under this point of view. Lemma 2.8. has been successfully implemented for the generation of unit directions in the q–space.

For above stated problem (B) of calculating quantiles two different types of approaches have been considered. Computational results suggest that the %–rejection method is not as effective as the use of different root finding methods combined with the lattice rules. In particular, we investigated for the latter case among other methodologies

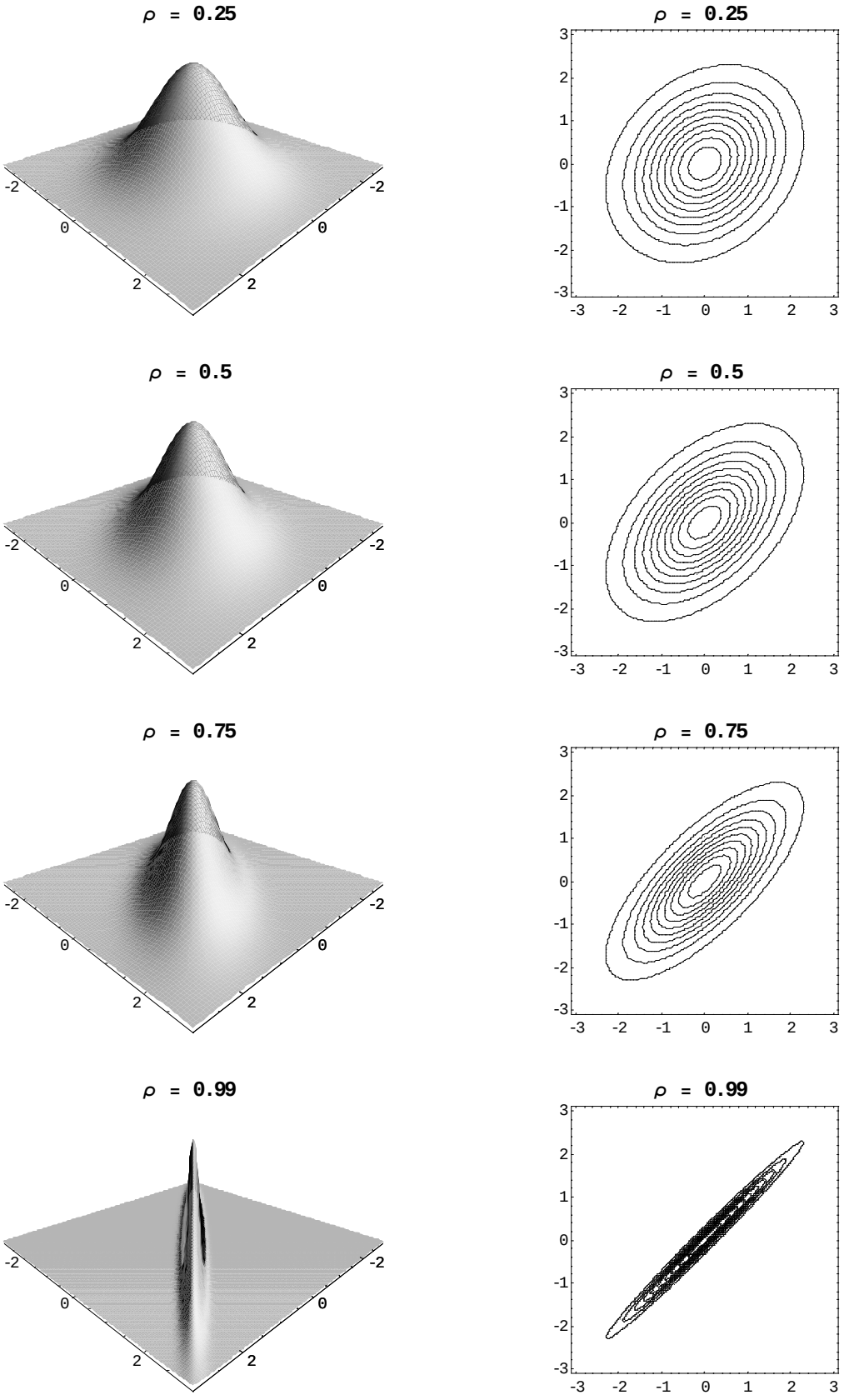
- the bisection method,
- the secant method,
- regula falsi, and
- several modified secant methods.

For typical values of α investigated (0.05, 0.1, ...) both the regula falsi and the original secant method behaved worse than the modified versions. The bisection method, though simple and intuitive, yielded surprisingly good results and because of its easy implementation it might be regarded further on. SAS/IML algorithms are given in the Appendix (B.2.10 - B.2.13). The modified secant method provided here is due to Ridders (1979), see also Press et al. (1992, pp. 347). It assumes an exponential increase (or decrease) of the function considered on the bracketed interval. Therefore, this method seems to be suitable for our requirements, when α takes the usual values around 0.05.

Finally, we refer again to the homepage of Genz. Beside the three algorithms introduced above it also contains several other FORTRAN implementations to calculate mv*t*–probabilities, such as a subregion adaptive method and an adaptation of Deak's method similar to the multi-normal case. The website is available under the URL http://www.sci.wsu.edu/math/faculty/genz/homepage.



Surface and contour plots of the standardised bivariate normal density function.



Surface and contour plots of the standardised bivariate normal density function *(continued).*

3. Choice of appropriate contrast coefficients

In Subsection 1.3.4. we introduced in detail the approach of contrast tests for both single and multiple versions. Further we have provided in Section 2.2. the theoretical and numerical basics for handling both SCTs and MCTs under H_0. We now focus our attention on the last remaining open problem - choosing the contrast coefficients appropriately. In Section 3.1. we give a brief survey of existing definitions. In Section 3.2. and 3.3. we will extend both Williams' and Marcus' tests to general unbalanced settings by applying the concept of MCTs, i.e. we are going to establish associated contrast sets for the original test statistics. Effort is made to overcome in this way the problems mentioned in Example 1.4. In Section 3.4. a new attempt of defining a contrast matrix is made, trying to overcome at least partially the rather empirical derivations of contrasts so far. The technique of MCTs is illustrated by the example of Chapter 1. The results are compared to previous calculations done for other trend tests. In the final section we proof an useful result for the evaluation of MCTs. It states that each MCT can be evaluated by an at most *k*-dimensional multivariate *t*-distribution, regardless of the number of contrasts actually used for defining the MCT.

3.1. Review of contrast definitions

The literature concerning contrast tests (mainly SCTs) is vast and therefore only a small outlook can be given here. For further reading we refer to Tamhane et al. (1996) and Hothorn and Hauschke (1998) and the references therein. Table 3.1. summarises the main contrasts discussed in the following, providing examples for the case $k = 3$. Before actually starting with the survey, we draw the attention on one main interpretation of contrasts. In the linear combination $\sum_i c_i \bar{X}_i$ the way of comparing the means is totally defined by the contrast coefficients c_i. Choosing for example the *pairwise contrast*

$$\boldsymbol{c} = (c_1, c_2, c_3, c_4) = (0, -1, 1, 0)$$

one would compare ('contrast') the means of group 2 and 3 through $\overline{X}_3 - \overline{X}_2$. Therefore, one can roughly say that the kind of comparison among the means is 'mapped' onto the contrast vectors. This is an important aspect to consider in the following attempts of contrast choices.

Lewis and Mouw (1978) and more recently Saville and Wood (1991) discussed some classes of contrasts typically used in a post-hoc analysis after conducting an ANOVA. In such cases the practitioner is interested in comparing the average of the population means of certain classes or factors with each other. In the context of order restriction illustrative examples are given by Bailey (1998). He suggested several strategies to compare *k* doses to a control. In the first case he uses several *step contrasts*, which take the information from all treatment groups into account and compare the associated pooled averages of neighbouring doses:

$$\begin{pmatrix} -k, & 1, & 1, & \dots & 1 \\ -k-1, & -k-1, & 2, & \dots & 2 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ -1, & -1, & -1, & \dots & k \end{pmatrix}, \tag{3.1a}$$

where we call the first and the last contrasts *reverse Helmert* respectively *Helmert contrast*. Next, a set of Helmert contrasts of different dimensionalities are proposed:

$$\begin{pmatrix} -1, & -1, & \dots & -1, & k \\ -1, & -1, & \dots & k-1, & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ -1, & 1, & \dots & 0, & 0 \end{pmatrix}. \tag{3.1b}$$

In contrast to (3.1a) this set does not use the information of all treatment groups, as is the case of

$$\begin{pmatrix} -k, & -k+2, & \dots & k-2, & k \\ -k+1, & -k+3, & \dots & k-1, & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ -1, & 1, & \dots & 0, & 0 \end{pmatrix}. \tag{3.1c}$$

These contrasts are called *linear contrasts*, again for varying number of treatment groups. Performing a SCT with such a contrast definition is equivalent to test on slope in a linearised regression model (see Subsection 1.4.). Finally, simple pairwise testing of adjacent groups in the treatment ordering is also proposed (Bailey, 1998):

$$\begin{pmatrix} 0, & 0, & \dots & 0, & -1, & 1 \\ 0, & 0, & \dots & -1, & 1, & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ -1, & 1, & \dots & 0, & 0, & 0 \end{pmatrix}. \tag{3.1d}$$

Using any of these four strategies, one always conducts the first SCT in order to determine whether the highest treatment group has an effect at all with respect to the other groups. Conditioned on the significance of this test one would perform the next SCT, if the global null hypothesis has been rejected before. These examples illustrate, how several contrasts are combined to a stepwise testing procedure, which leads to the estimation of a minimum effective dose, to be discussed in more detail in Chapter 6. Numerical examples in Bailey (1998) suggest that (3.1a) is a good choice among the four proposed sets. Interestingly, all of these sets have already been used before to form MCTs. Just taking the maximum over all SCTs from (3.1a) leads to the global test of Hirotsu (1979, 1997; cf. Subsection 1.4). The set (3.1b) is equivalent to the decomposition method of McDermott and Mudholkar (1993), when applying Tippett's minimum. The similarity between (3.1c) and the usual regression has already been noticed. And finally, performing simultaneous comparison of all SCTs in (3.1d) has already been proposed by Lee and Spurrier (1995). None of these original approaches cited uses the multivariate *t*-distribution in the general unbalanced case.

Another class of weights frequently used are the polynomial contrasts. They include tests on dose-response relationships according to certain response models. The linear SCT considered above tests on linear relationships, the *quadratic contrasts* a quadratic fit, and so on (see for example Saville and Wood, 1991).

Contrast type	Example	Reference
Helmert contrast	$(-1, -1, -1, 3)$	Ruberg (1989)
Reverse Helmert contrast	$(-3, 1, 1, 1)$	Fligner and Wolfe (1982)
Pairwise contrast	$(-1, 0, 1, 0)$	Saville and Wood (1991)
Linear contrast	$(-3, -1, 1, 3)$	Saville and Wood (1991)
Maximin contrast	$(-0.87, -0.13, 0.13, 0.87)$	Abelson and Tukey (1963)
Linear−2−4 contrast	$(-12, -2, 2, 12)$	Abelson and Tukey (1963)
Jonkheere-analogon contrast	$(-1, -0.67, -0.17, 1.83)$	Neuhäuser (1996)

Table 3.1. Some single contrast definitions, examples given for $k = 3$.

If no certain types of comparisons among the means are determined in advance of an experiment, however, the contrast definitions considered so far and others proposed in the literature are rather empirically or strongly model-dependent (linear, quadratic, ... response shape, number of efficient dose steps, ...). Additionally, most of them were introduced for balanced set-ups only and fail even for moderate imbalances. Addressing these problems, Abelson and Tukey (1963) proposed a SCT based on theoretical considerations (*maximin contrast*). Using geometrical arguments, they established for $i = 0, 1, \ldots, k$ the coefficients

$$c_i = \sqrt{i \cdot \left(1 - \frac{i}{k+1}\right)} - \sqrt{(i+1) \cdot \left(1 - \frac{i+1}{k+1}\right)}, \tag{3.2}$$

which maximise the minimum correlation ρ between $\boldsymbol{c}$ and μ in the balanced case. Schaafsma and Smid (1966) showed that the choice of above c_i's leads exactly to the most stringent somewhere most powerful SCT, i.e. that test based on a linear combination of the means, which minimises the maximum shortcoming over the whole alternative out of the class of tests, which are most powerful for at least one simple hypothesis in H_A. Applying the techniques provided in Schaafsma and Smid (1966), it is possible to generalise (3.2) to the less restricted case of unequal sample sizes (Mukerjee et al., 1987, p. 903).

It is evident that SCTs will be very powerful if we have some prior information about the approximate location of μ within H_A. If we knew the μ_i's, one could choose the c_i's to make

ρ equal or close to the maximum value +1. If, for example, we expect in advance only one effective dose step, then the test provided in (3.1a) would be very powerful for testing

$$H'_A = \bigcup_{i=1}^{k} H'_{A(i)}, \quad H'_{A(i)}: \mu_0 = \ldots = \mu_{i-1} < \mu_i = \ldots = \mu_k .$$

On the other hand, adapting the coefficients c_i with respect to the observed values $\{X_{ij}\}_{ij}$ leads exactly to the likelihood ratio test, as stated in Lemma 1.6.

In Subsection 1.3.4. we have already seen the strong shape dependence of SCTs in terms of power. To overcome this disadvantage, MCTs have been proposed. They have been investigated in the literature since Dwass (1960) and Dunn and Massey (1965). In fact, many multiple comparison procedures may be regarded as specific MCTs (cf. Section 1.4. and also the examples earlier in this section). On the other side, several ad hoc multiple contrast definitions have been published, see for example Hothorn et al. (1997) and Westfall (1997). They are based on more or less arbitrary combinations of SCTs, as those provided in Table 3.1. Sugiura (1994) applies Bayesian decision theory on MCTs and derives an approximate multiple contrast Bayes test based on uniform priors under the slippage alternative H'_A. The resulting coefficients are the same as those of (3.1a) up to a normalisation factor. But because of numerical difficulties in implementing the MCTs, significantly less has been published about MCTs than about SCTs. One way of overcoming these numerical problems is to define adequate contrasts, which are mutually orthogonal (Mukerjee et al., 1986, 1987). This makes the correlation matrix $\boldsymbol{R}$ be of diagonal pattern and breaks down the dimensionality problem. On the other side, requiring $\rho = 0$ for each combination of two contrasts leads to a somewhat restricted use of MCTs. The set given in (3.1b) is one example of such orthogonal contrasts.

3.2. A Williams-type multiple contrast test

We will now establish the missing link between Williams' $\bar{t}$ – test and MCTs, so that we can use Williams' approach in less restricted settings than the original approach. Before we define the extended Williams' test, we give the following

Lemma 3.1.: Let $\bar{t}$ be the test statistic (1.6) and $\overline{\boldsymbol{X}} = \left(\overline{X}_0,\ \overline{X}_1,\ \ldots,\ \overline{X}_k\right)^t$. Set

$$\boldsymbol{C} = \begin{pmatrix} -1 & 0 & \ldots & 0 & 1 \\ -1 & 0 & \ldots & \frac{n_{k-1}}{n_{k-1}+n_k} & \frac{n_k}{n_{k-1}+n_k} \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ -1 & \frac{n_1}{n_1+\ldots+n_k} & \ldots & \frac{n_{k-1}}{n_1+\ldots+n_k} & \frac{n_k}{n_1+\ldots+n_k} \end{pmatrix}. \qquad (3.3)$$

Then

$$\hat{\mu}_k - \overline{X}_0 = \max \boldsymbol{C}\overline{\boldsymbol{X}}.$$

Proof: Setting $i = k$ in the representation (1.5) of the amalgamated means and remembering that the control group is not included in the amalgamation process (p. 17) we get

$$\begin{aligned} \hat{\mu}_k &= \max_{1\le u\le k} \sum_{j=u}^{k} n_j \overline{X}_j \Big/ \sum_{j=u}^{k} n_j = \\ &= \max\left\{ \frac{n_1\overline{X}_1 + n_2\overline{X}_2 + \ldots + n_k\overline{X}_k}{n_1+n_2+\ldots+n_k}, \ldots, \frac{n_{k-1}\overline{X}_{k-1} + n_k\overline{X}_k}{n_{k-1}+n_k},\ \overline{X}_k \right\} = \\ &= \max\left\{ \begin{pmatrix} 0 & \ldots & 0 & 1 \\ 0 & \ldots & \frac{n_{k-1}}{n_{k-1}+n_k} & \frac{n_k}{n_{k-1}+n_k} \\ \vdots & \ldots & \vdots & \vdots \\ \frac{n_1}{n_1+\ldots+n_k} & \ldots & \frac{n_{k-1}}{n_1+\ldots+n_k} & \frac{n_k}{n_1+\ldots+n_k} \end{pmatrix} \begin{pmatrix} \overline{X}_1 \\ \overline{X}_2 \\ \vdots \\ \overline{X}_k \end{pmatrix} \right\}. \end{aligned}$$

Therefore we have

$$\begin{aligned} \hat{\mu}_k - \overline{X}_0 &= \max\left\{ \begin{pmatrix} 0 & \ldots & 0 & 1 \\ 0 & \ldots & \frac{n_{k-1}}{n_{k-1}+n_k} & \frac{n_k}{n_{k-1}+n_k} \\ \vdots & \ldots & \vdots & \vdots \\ \frac{n_1}{n_1+\ldots+n_k} & \ldots & \frac{n_{k-1}}{n_1+\ldots+n_k} & \frac{n_k}{n_1+\ldots+n_k} \end{pmatrix} \begin{pmatrix} \overline{X}_1 \\ \overline{X}_2 \\ \vdots \\ \overline{X}_k \end{pmatrix} - \begin{pmatrix} \overline{X}_0 \\ \overline{X}_0 \\ \vdots \\ \overline{X}_0 \end{pmatrix} \right\} = \\ &= \max\left\{ \begin{pmatrix} -1 & 0 & \ldots & 0 & 1 \\ -1 & 0 & \ldots & \frac{n_{k-1}}{n_{k-1}+n_k} & \frac{n_k}{n_{k-1}+n_k} \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ -1 & \frac{n_1}{n_1+\ldots+n_k} & \ldots & \frac{n_{k-1}}{n_1+\ldots+n_k} & \frac{n_k}{n_1+\ldots+n_k} \end{pmatrix} \begin{pmatrix} \overline{X}_0 \\ \overline{X}_1 \\ \vdots \\ \overline{X}_k \end{pmatrix} \right\} = \\ &= \max \boldsymbol{C}\overline{\boldsymbol{X}}. \end{aligned}$$

Definition 3.1.: We call a MCT with the contrast matrix $\boldsymbol{C}$ from Lemma 3.1. a Williams-type multiple contrast or, shorter, *Williams contrast.*

The maximum contrast above consists of comparisons of the negative control with the weighted average over the last j treatment groups, $j = 1, ..., k$, respectively. In total we have $q = k$ single contrasts, which by joining form the set (1.12) of a MCT. This representation is an exact analogue of Williams' approach, and unites two important aspects. On the one side it takes the order restriction of the means into account through the contrast definition following the isotonic estimates (1.5). Due to existing theoretical results about the distribution of MCTs and their handling by virtue of the algorithms presented in Chapter 2 we have on the other hand a powerful and flexible tool to deal with the unbalanced case without limitations. Note that Williams' $\bar{t}-$test is not identical to our proposed MCT because of different variance estimators. We adopted the completely studentised statistic by making use of the usual mean square error. Williams (1971, 1972), in contrast, took the denominator from the usual t–test and we are going to explain the behaviour differences in terms of power in Section 4.2. A list of Williams contrasts up to $k = 6$ in the balanced case is given in the Appendix.

Robertson et al. (1988, p. 190) were the first, who noted the possibility of defining a Williams-type multiple contrast. But they mentioned the balanced case only, without going into depth or considering its actual computation. To our knowledge, the next article concerning this topic was published by Yoshimura et al. (1997). They first stated the Williams-type multiple contrast test for $k = 3$ in the balanced case, but they continued using the old denominator of $\bar{t}$. Generalising to the unbalanced set-up they proposed an extended Williams method, which differs again from our approach by a different variance estimator. In both cases their maximum statistics are no longer multiple contrast tests. Both procedures were evaluated approximately by use of a resampling approach via the SAS-procedure `PROC MULTTEST` (SAS Institute Inc., 1997, p. 777).

Before we leave this section we state the interesting

Remark 3.1.: Assume that we include the negative control in the amalgamation process, in contrast to Williams' original approach. Then it is straight forward to see by use of Lemma 1.5. and the dimension reduction technique of Section 3.6. that the relation $\hat{\mu}_k - \overline{X}_0 = \max \boldsymbol{C}\overline{\boldsymbol{X}}$ still holds. This means that the empirical conclusion of Tamhane et al. (1996) could be derived analytically when using the contrast representation and it makes no difference whether to include or not the control group.

3.3. A Marcus-type multiple contrast test

Similarly to the last section we introduce now a method for representing Marcus' $\bar{t}^{\,\mathrm{mod}}$–test as a MCT. We first give an analogous

Lemma 3.2.: Let $\bar{t}^{\,\mathrm{mod}}$ be the test statistic (1.9) and $\overline{\boldsymbol{X}} = \left(\overline{X}_0,\ \overline{X}_1,\ \ldots,\ \overline{X}_k\right)^t$. Then the relation

$$\hat{\mu}_k - \hat{\mu}_0 = \max\left\{0,\ \max_{0\le i<j\le k}\left\{\frac{n_j\overline{X}_j + \ldots + n_k\overline{X}_k}{n_j + \ldots + n_k} - \frac{n_0\overline{X}_0 + \ldots + n_i\overline{X}_i}{n_0 + \ldots + n_i}\right\}\right\} \qquad (3.4)$$

holds.

Proof: Setting $i = k$ respectively $i = 0$ in equation (1.5) of the amalgamated means we get the representations

$$\hat{\mu}_k = \max_{0\le u\le k}\sum_{j=u}^{k} n_j\overline{X}_j \Big/ \sum_{j=u}^{k} n_j =$$
$$= \max\left\{\frac{n_0\overline{X}_0 + n_1\overline{X}_1 + \ldots + n_k\overline{X}_k}{n_0 + n_1 + \ldots + n_k}, \ldots, \frac{n_{k-1}\overline{X}_{k-1} + n_k\overline{X}_k}{n_{k-1} + n_k},\ \overline{X}_k\right\}$$

and

$$\hat{\mu}_0 = \min_{0\le v\le k} \sum_{j=0}^{v} n_j \overline{X}_j \Big/ \sum_{j=0}^{v} n_j =$$

$$= \min\left\{ \overline{X}_0, \frac{n_0\overline{X}_0 + n_1\overline{X}_1}{n_0+n_1}, \ldots, \frac{n_0\overline{X}_0 + n_1\overline{X}_1 + \ldots + n_k\overline{X}_k}{n_0+n_1+\ldots+n_k} \right\}.$$

Therefore,

$$\hat{\mu}_k - \hat{\mu}_0 = \max_{0\le i,j\le k}\left\{ \frac{n_j\overline{X}_j + \ldots + n_k\overline{X}_k}{n_j+\ldots+n_k} - \frac{n_0\overline{X}_0 + \ldots + n_i\overline{X}_i}{n_0+\ldots+n_i} \right\} =$$

$$= \max\left\{ 0, \max_{0\le i<j\le k}\left\{ \frac{n_j\overline{X}_j + \ldots + n_k\overline{X}_k}{n_j+\ldots+n_k} - \frac{n_0\overline{X}_0 + \ldots + n_i\overline{X}_i}{n_0+\ldots+n_i} \right\} \right\}.$$

We therefore succeeded in representing the statistic $\hat{\mu}_k - \hat{\mu}_0$ as a simple maximum term and a natural way of applying the MCT principle on Marcus' approach is now to identify each element of (3.4) as a contrast. Unfortunately, a closed form expression for the resulting contrast matrix $\boldsymbol{C}$ is not available. In the Appendix A.2. the contrasts are given up to $k = 6$.

Definition 3.2.: Using the contrast matrix defined through (3.4) we call the associated MCT Marcus-type multiple contrast or, shorter, *Marcus contrast*.

Let us compare briefly the coefficients of the Williams and the Marcus contrast for $k = 3$ in the balanced case:

$$\begin{pmatrix} -1 & 0 & 0 & 1 \\ -1 & 0 & \frac{1}{2} & \frac{1}{2} \\ -1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \quad \text{respectively} \quad \begin{pmatrix} -1 & 0 & 0 & 1 \\ -1 & 0 & \frac{1}{2} & \frac{1}{2} \\ -1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ -\frac{1}{2} & -\frac{1}{2} & 0 & 1 \\ -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 1 \\ -1 & -1 & 1 & 1 \end{pmatrix}.$$

Following (3.4) it becomes clear that the Williams single contrasts are all contained in Marcus contrasts (first through third). These are suitable for concave dose-response shapes, as the higher dose groups are being pooled and compared to *C*–. When looking further at Marcus contrast, we note two single contrasts (fourth and fifth), which seem to be appropriate for

detecting convex shapes, as they take the average over the lower treatments. We therefore could explain analytically the empirical results published in the literature that Williams' test is suitable for concave shapes, whereas Marcus' test behaves much better for convex profiles (see for example Marcus, 1976). Even if $\bar{t}^{\,\mathrm{mod}}$ is not identical with the above proposed MCT (different variance estimators, same argumentation as in the case of $\bar{t}$), with the results of Chapter 2 we have a flexible tool to handle Marcus' test in more general situations. Note, that because of its particular construction Williams' test has sometimes be regarded in the literature as a 'many-to-one test under order restriction'. No other trend test bears this distinguishing feature of comparing the higher treatments solely to a certain control group.

In her original article, Marcus (1976, p. 178) already discussed briefly the exact distribution of $\bar{t}^{\,\mathrm{mod}}$ and worked with a representation similar to that of Lemma 3.2. for small values of *k*. Yoshimura et al. (1997) gave the contrast representation for $k = 4$ without any derivation and they continued using the original variance estimator $s\sqrt{\frac{2}{n}}$. The evaluation of this test statistic was conducted by applying a resampling approach. Recently, Hayter et al. (1999a) derived independently a very similar expression to (3.4), restricted to the balanced case only. In the later course of the article they took over the variance estimator of Marcus. As this test statistics is no longer multivariate *t*-distributed, they established a very different method to evaluate the null distribution exactly.

Remark 3.2.: Analogously to Remark 3.1. we conclude that it makes no difference whether the amalgamation process is carried out by excluding the control or not.

3.4. A new multiple contrast definition

In this section we propose a new approach of defining the contrast matrix $\boldsymbol{C}$. The reason for doing so is that from our point of view most contrast definitions are more or less empirical. The proposals may be classified according to their development reasons:

- comparison of certain pre-determined classes or factors (Saville and Woods, 1991);
- representation of established multiple comparison procedures as MCTs (such as the approaches of Williams, 1971; Marcus, 1976; Hirotsu, 1979);
- ad hoc definitions of MCTs (Hothorn et al., 1997; Westfall, 1997);
- contrast tests based on certain optimisation criterions (Abelson and Tukey, 1963; Schaafsma and Smid, 1966).

For MCTs under total order restriction nothing has been published what would belong to the last category. In the following, an attempt is made to fill this gap. The main idea can be structured in the four steps:

1) decomposition of H_A into appropriate sub-hypotheses;
2) search for the maximin contrast for each sub-hypotheses;
3) adjustment for imbalances;
4) building the maximum over the resulting $2^k - 1$ single contrasts.

We shall illustrate the derivation for $k = 3$. We first decompose $H_A: \mu_0 \le \mu_1 \le \mu_2 \le \mu_3$ in all possible scenarios as

$$H_A = \bigcup_{i=1}^{7} H_{A(i)},$$

with

$$\begin{aligned}
&H_{A(1)}: \mu_0 = \mu_1 = \mu_2 < \mu_3,\\
&H_{A(2)}: \mu_0 < \mu_1 = \mu_2 = \mu_3,\\
&H_{A(3)}: \mu_0 = \mu_1 < \mu_2 = \mu_3,\\
&H_{A(4)}: \mu_0 < \mu_1 = \mu_2 < \mu_3,\\
&H_{A(5)}: \mu_0 < \mu_1 < \mu_2 < \mu_3,\\
&H_{A(6)}: \mu_0 = \mu_1 < \mu_2 < \mu_3,\\
&H_{A(7)}: \mu_0 < \mu_1 < \mu_2 = \mu_3.
\end{aligned} \tag{3.5}$$

Consequently, H_A has been decomposed into the 'smallest' possible sub-hypotheses. With its total number given by $2^k - 1$ (= 7 for k = 3) this approach will finally result in a $\left(2^k - 1\right)$–dimensional contrast. Note that no $H_{A(i)}$ can be decomposed further. Every true (but unknown) relation among the μ_i's will fall in exactly one of these sub-alternatives (as far as the general constraint H_A holds).

After the successful partition we are left with the question: which are suitable choices of contrasts for each sub-alternative? Clearly, a linear contrast would be a good choice for $H_{A(5)}$, but a bad one for $H_{A(1)}$ or $H_{A(2)}$ (convex and concave dose-response shapes, respectively). We will apply the ideas of Abelson and Tukey (1963) on our special case here. They derived a general methodology for finding the 'best' contrast, given any particular set of inequalities on the μ_i's. We have seven of such sets of inequalities and we determine with the methods of Abelson and Tukey for each set that contrast which maximises the minimum correlation between μ and $\boldsymbol{c}$ under the corresponding constraint, or, equivalently, which maximises the minimum power. Note that the maximin contrast given in (3.2) is not suitable here, as it was developed for the global alternative H_A. In contrast, we are seeking for $2^k - 1$ contrasts, which are locally optimal for each subspace of H_A. For a better understanding and more details of the techniques used below the reader is referred to the article of Abelson and Tukey and the terminology therein.

To illustrate the determination of suitable set of contrasts we explain the approach in detail for $H_{A(4)}$. Let the constraint given be

$$H_{A(4)}: \mu_0 < \mu_1 = \mu_2 < \mu_3. \qquad (3.6)$$

The boundaries of its subspace are given by matching inequalities, corner patterns and so-called corner vectors as follows:

Inequality	Corner pattern	Corner vector	SSD
$\mu_0 \le \mu_1$	$\mu_0 < \mu_1 = \mu_2 = \mu_3$	$(0, 1, 1, 1)$	$3/4$
$\mu_2 \le \mu_3$	$\mu_0 = \mu_1 = \mu_2 < \mu_3$	$(0, 0, 0, 1)$	$3/4$

The sum of squares of deviations (SSD) $\sum_i (\mu_i - \overline{\mu})^2$ gives the squared length of the $(\mu - \overline{\mu})-$ vectors, where $\overline{\mu} = \sum_i \mu_i / (k+1)$. This helps us determining the maximin contrast by solving

$$\begin{aligned} (s) \quad & c_0 + c_1 + c_2 + c_3 = 0 \\ (a) \quad & c_1 + c_2 + c_3 = \sqrt{\tfrac{3}{4}} \\ (b) \quad & c_3 = \sqrt{\tfrac{3}{4}}, \end{aligned}$$

where (*s*) is the contrast ensuring equation. Because of (3.6) and the additional condition $c_1 = c_2$ we get by successive substitution the solution $(-0.866,\ 0,\ 0,\ 0.866)$. As this contrast clearly satisfies both inequalities, it is the desired maximin contrast under the constraint (3.6).

When repeating above steps for each sub-hypothesis given under (3.5), the following set of contrasts is yielded:

$$\begin{aligned} \boldsymbol{c}_1 &= (-0.2887,\ -0.2887,\ -0.2887,\ 0.866), \\ \boldsymbol{c}_2 &= (-0.866,\ 0.2887,\ 0.2887,\ 0.2887), \\ \boldsymbol{c}_3 &= (-0.5,\ -0.5,\ 0.5,\ 0.5), \\ \boldsymbol{c}_4 &= (-0.866,\ 0,\ 0,\ 0.866), \\ \boldsymbol{c}_5 &= (-0.866,\ -0.134,\ 0.134,\ 0.866), \\ \boldsymbol{c}_6 &= (-0.5,\ -0.5,\ 0.134,\ 0.866), \\ \boldsymbol{c}_7 &= (-0.866,\ -0.134,\ 0.5,\ 0.5). \end{aligned} \qquad (3.7)$$

Each of them maximises the minimum correlation between μ and $\boldsymbol{c}$ under the corresponding restriction and is optimal in this sense. The next step involves the adjustment for possible imbalances. This is done by multiplying the contrasts of (3.7) with the associated sample size and subsequent centring so that the componentwise sum of the contrast remains 0, i.e. we set

$$c_{ij} = n_j c_{ij} - \sum_{l=0}^{k} n_l c_{il} \Big/ (k+1), \qquad j = 0, \ldots, k,\ i = 1, \ldots, 2^k - 1.$$

Finally the new proposed MCT is obtained by taking the maximum over the arising number of single contrasts. Note that for $n_j = n,\ j = 0, \ldots, k,\ \ c_{ij} = nc_{ij} - n\sum_l c_{il} \big/ (k+1) = nc_{ij}$ and therefore according to Lemma 1.5. the test statistic does not change. Other criteria of including unbalancedness could be applied and we refer to the example given in Section 3.5. for a more detailed discussion.

The whole procedure sketched above can be generalised straight forward to an arbitrary number of treatments *k*. Unfortunately, a closed form expression is not available but the contrast definitions are given up to *k* = 6 for the balanced case in the Appendix A.3. A SAS/IML program, which outputs these matrices for arbitrary *k* in the general unbalanced set-up is given in Appendix B.3.3. But before we proceed to the examples in the next section, we try to establish a link between the proposed MCT and the amalgamation procedure with

Remark 3.3.: Recall the amalgamation process discussed in Section 1.2. The max-min formula given there pools the means of two adjacent groups if and only if the pre-determined ordering is violated. One can then show easily that for *k* + 1 treatment groups to be compared under total order restriction exactly $2^k - 1$ different outcomes are possible in the alternative (not counting for the $2^k - \text{th}$ possibility, which occurs only when all means are amalgamated to a common average value). These $2^k - 1$ cases match uniquely to our decomposition (3.5) of the alternative space. Remember, that Williams' and Marcus' approaches consider only *k* respectively $\binom{k+1}{2}$ such situations. Therefore, one other reason for introducing Definition 3.3. is the search for a MCT that takes all $2^k - 1$ possible situations into account, which arise from the amalgamation process (see also Seidel (1999) for a brief discussion on this subject).

With this relation in mind we therefore are now able to state in accordance with Seidel (1999)

Definition 3.3.: Applying the procedure illustrated above we define the associated MCT as an *isotonic contrast*.

3.5. Example

We now come back to the example already analysed in Subsection 1.3.5. for the LRT and the original approaches of Williams and Marcus. The three new proposals of MCTs are compared to the three established tests, illustrating their use and evaluation. The p-values stated below were obtained by applying program B.2.9 in the Appendix with the error bound $\varepsilon = 0.0001$.

Example 3.1.: Let us return to the example from the Introduction of comparing the E.C.I. values of five different larva development stages and an adult form under the order restriction (1.2). The associated contrast matrices for the present imbalance $\boldsymbol{n} = (21,\ 10,\ 15,\ 17,\ 21,\ 4)$ and $k = 5$ are approximately

Williams contrast

$$\begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 & .84 & .16 \\ -1 & 0 & 0 & .4 & .5 & .1 \\ -1 & 0 & .26 & .3 & .37 & .07 \\ -1 & .15 & .22 & .25 & .31 & .06 \end{pmatrix}$$

Marcus contrast

$$\begin{pmatrix} -1 & .15 & .22 & .25 & .31 & .06 \\ -1 & 0 & .26 & .3 & .37 & .07 \\ -.68 & -.32 & .26 & .3 & .37 & .07 \\ -1 & 0 & 0 & .4 & .5 & .1 \\ -.68 & -.32 & 0 & .4 & .5 & .1 \\ -.46 & -.22 & -.33 & .4 & .5 & .1 \\ -1 & 0 & 0 & 0 & .84 & .16 \\ -.68 & -.32 & 0 & 0 & .84 & .16 \\ -.46 & -.22 & -.33 & 0 & .84 & .16 \\ -.33 & -.16 & -.24 & -.27 & .84 & .16 \\ -1 & 0 & 0 & 0 & 0 & 1 \\ -.68 & -.32 & 0 & 0 & 0 & 1 \\ -.46 & -.22 & -.33 & 0 & 0 & 1 \\ -.33 & -.16 & -.24 & -.27 & 0 & 1 \\ -.25 & -.12 & -.18 & -.2 & -.25 & 1 \end{pmatrix}$$

Isotonic contrast

$$\begin{pmatrix} -1.89 & .12 & -.79 & -1.16 & -1.89 & 5.6 \\ -5.44 & -2.26 & -3.7 & -4.28 & 12.75 & 2.93 \\ -4.49 & -1.31 & -2.75 & -3.33 & 6.65 & 5.23 \\ -8.3 & -3.81 & -5.85 & 7.21 & 8.85 & 1.91 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -11.6 & -5.22 & -.5 & 1.74 & 12.68 & 2.86 \\ -10.6 & -4.27 & .45 & 12.7 & 6.58 & 5.16 \\ -18 & 2.98 & 3.89 & 4.26 & 4.99 & 1.89 \\ -16.6 & 2.59 & 2.59 & 2.59 & 2.59 & 6.24 \\ -17.8 & .55 & .14 & -.02 & 13.48 & 3.66 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -18.3 & -1.5 & -.14 & 7.86 & 9.49 & 2.55 \\ -17 & -.24 & 1.13 & 4.83 & 5.45 & 5.83 \\ -18 & -1.25 & .12 & 2.36 & 13.29 & 3.48 \\ -17.1 & -.3 & 1.07 & 3.31 & 7.2 & 5.77 \end{pmatrix}$$

Note that all three of them are already adjusted for the unequal sample size allocation. Conducting the Williams contrast we get a p-value of 0.0028. Comparing it with the value of the original $\bar{t}$ (p = 0.0373) we have a significant improvement. Remember that the extended contrast method proposed is not identical to the original approach because of the use of different variance estimators. But we observe that the new version takes the sample size allocation better into account, as its variance estimator is not restricted to n_0 and n_k

only. If we switch the last two group sample sizes, no stronger effect can be observed (p = 0.0019), in contrast to $\bar{t}$ (p = 0.0017; remember the discussion in Example 1.4.). Performing the extension of Marcus' method we obtain a p-value of 0.0042 and a similar argumentation as before holds. Focus now the attention on the isotonic contrast. If we compute the p-value for the sample size corrected contrast given above one obtains 0.0036, which is similar to the p-values of the other trend tests. But performing the same test without this adjustment p is only 0.0132. Unfortunately, this kind of sample size adjustment is rarely mentioned in the literature, but this example shows the importance of applying the contrast principle correctly.

Other possibilities of adjusting the isotonic MCT for imbalances may be considered. For example, one could imagine each SCT as a pooled two-sample test, where each sample is divided into several (at least 1 but at most k) treatment groups. Williams' and Marcus' MCT are two examples for this approach. In the discussion of Williams' MCT following Definition 3.2. we already observed that the high doses are pooled subsequently and compared to $C-$. This approach results in pooled two-sample tests $C-$ versus high doses. The weights for combining the higher doses are determined according to the underlying sample size allocation. In the balanced case, for example, each dose would be assigned the same weight, as done in the discussion on p. 86. Marcus' MCT can be interpreted similarly. Applying this methodology to the proposed isotonic contrast one obtains p = 0.0032. This value lies close to the 0.0036 yielded before. We therefore conclude that both adjustments for imbalances achieved the goal, but they do not differ very much when compared to each other. Hence, for the remaining parts of the thesis we do not investigate the differences between the individual sample size adjustments. Instead, for the sake of convenience, we focus solely on the first proposed method.

	Williams' $\bar{t}$	Marcus' $\bar{t}^{\,\mathrm{mod}}$	LRT	Williams contrast	Marcus contrast	Isotonic contrast	Dunnett	$F-$test
p-value	0.0373	0.0481	0.0039	0.0028	0.0042	0.0036	0.0052	0.0535

The table above summarises the p-values of all tests performed for the present example. The following conclusions may be drawn:

- Williams' and Marcus' original tests are not suitable in unbalanced settings because of their chosen variance estimator. Moreover, their distribution functions under H_0 are still not feasible in such situations, simulation-based p-values are the state of art.
- The LRT behaves well but is not a clear best as one could expect before. One main problem is its unknown behaviour under H_A for general set-ups.
- All three proposed contrast tests behave similarly to each other and to the LRT. However, the power study in Chapter 4 will yield the result that the Williams contrast seems to have an inferior 'average' power.
- Dunnett's test yielded a surprisingly good p-value. In most of the cases, its power will lie behind that of the MCTs and the LRT, but it is a fairly simple and easy to evaluate candidate.
- The *F*–test, which takes no order restriction at all or even the many-to-one design into account, behaves poor and should not be used.

3.6. Reduction of the dimensionality of multiple contrast tests

In the preceding sections several MCTs have been proposed, some of them requiring the ability of evaluating high dimensional integrals. This is not always possible, primarily because of time limitation for an appropriate accuracy. Therefore we require a method which reduces the problem of the dimensionality. One technique widely used is the reduction onto a two dimensional integral if the correlation matrix $\boldsymbol{R}$ possesses a certain structure, for example the product correlation structure of Remark 1.1. Hayter (1989) provides a method for the special case of Tukey's all pair contrasts, reducing the original $\binom{k+1}{2}$–dimensional integral onto one over k dimensions. Somerville (1997, 1999) overcomes the whole problem by choosing a different approach at all. He finds for each random direction the minimum distance from the origin to a boundary. As the boundaries are defined on the $k + 1$ means, the dimension of his methodology is automatically restricted by k.

In the sequel we will proof a more general result, which shows that each MCT consisting of $q \geq k$ single contrasts can be reduced to the computation of an at most *k*-variate *t*-distribution. An example illustrates the theoretical result and the gain of time required for the computation.

To simplify the representation, we omit the index '0' in the next two results. First we establish a lemma which will be used for our main result in this section.

Lemma 3.3.: Let $\boldsymbol{C} = \left(c_{ij}\right)_{ij} = \left(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_q\right)^t$ be a given contrast matrix with rank $k-1$, where it is assumed that $k-1 \leq q$, $i = 1, \ldots, q,\ j = 1, \ldots, k$. Let further for $i = 1, \ldots, q,\ j = 1, \ldots, k$,

$$\boldsymbol{G} = \begin{pmatrix} * & \cdots & \cdots & * \\ 0 & \ddots & & \vdots \\ & \ddots & * & * \\ \vdots & & 0 & 0 \\ & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} k-1 \\ \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} q-k+1 \end{matrix}, \quad \text{i.e.} \quad \boldsymbol{G} = \begin{cases} g_{ij}, & i \leq j,\ i \neq k, \\ 0, & \text{else,} \end{cases} \quad \text{and} \quad \boldsymbol{T} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{B} & \boldsymbol{I}_{q-k+1} \end{pmatrix}_{q \times q}$$

so that

$$\boldsymbol{TC} = \boldsymbol{G}. \tag{3.8}$$

Here, $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ * & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ * & \ldots & * & 1 \end{pmatrix}$ is a lower triangular $(k-1)\times(k-1)$–matrix, $\boldsymbol{I}_{q-k+1}$ the $(q-k+1)$–dimensional unit matrix and $\boldsymbol{B}$ a $(q-k+1)\times(k-1)$–matrix. Then it follows that the matrices $\boldsymbol{T}$ and $\boldsymbol{G}$ exist and are uniquely determined.

Proof: Without loss of generalisation we assume $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{k-1}$ to be linearly independent (otherwise we exchange the corresponding rows).

Equation (3.8) is equivalent to

$$g_{ij} = \sum_{l=1}^{q} t_{il} c_{lj} .$$

According to our assumption of $\boldsymbol{T}$ and $\boldsymbol{G}$ we therefore have

$$g_{ij} = \begin{cases} \sum_{l=1}^{q} t_{il} c_{lj}, & i < j, \\ \sum_{l=1}^{q} t_{il} c_{li}, & i = j, \\ 0, & \text{else.} \end{cases} \qquad i = 1, \ldots, k,$$

Because of $t_{ii} = 1$, $\forall i = 1, \ldots, q$, we get for $i = 1$

$$g_{1j} = \sum_{l=1}^{1} t_{1l} c_{lj} = t_{11} c_{1j} = c_{1j}, \qquad j = 1, \ldots, k.$$

For $i = 2, \ldots, k-1$ we have

$$g_{ij} = \sum_{l=1}^{i} t_{il} c_{lj} = \sum_{l=1}^{i-1} t_{il} c_{lj} + c_{ij} = \begin{cases} 0, & j = 1, \ldots, i-1, \\ g_{ij}, & j \geq i. \end{cases} \tag{3.9}$$

According to the assumptions the resulting system of linear equations

$$\tilde{\boldsymbol{C}} \boldsymbol{x} = \boldsymbol{b}$$

has an unique solution $\boldsymbol{x} = \left(t_{i1}, \ldots, t_{i,i-1}\right)^t$, where $\tilde{\boldsymbol{C}} = \begin{pmatrix} c_{11} & \ldots & c_{1,i-1} \\ \vdots & & \vdots \\ c_{i-1,1} & \ldots & c_{i-1,i-1} \end{pmatrix}^t$ and $\boldsymbol{b} = -\left(c_{i1}, \ldots, c_{i,i-1}\right)^t$. Substituting $\boldsymbol{x}$ in (3.9) yields a unique solution for $g_{ii}, \ldots, g_{ik}$.

For $i = k, \ldots, q$ we get in accordance to our assumption of $\boldsymbol{T}$

$$g_{ij} = \sum_{l=1}^{i} t_{il} c_{lj} = \sum_{l=1}^{k} t_{il} c_{lj} + \underbrace{\sum_{l=k+1}^{i-1} t_{il} c_{lj}}_{=0} + t_{ii} c_{ij} = \sum_{l=1}^{k} t_{il} c_{lj} + c_{ij} = 0, \qquad j = 1, \ldots, k.$$

Again a unique solution of $\tilde{\boldsymbol{C}}'\boldsymbol{x}' = \boldsymbol{b}'$ is yielded for $\boldsymbol{x}' = (t_{i1}, \ldots, t_{ik})^t$, where $\tilde{\boldsymbol{C}}' = \begin{pmatrix} c_{11} & \ldots & c_{1k} \\ \vdots & & \vdots \\ c_{k1} & \ldots & c_{kk} \end{pmatrix}^t$ and $\boldsymbol{b}' = -(c_{i1}, \ldots, c_{ik})^t$. With this final step we determined the remaining entries of $\boldsymbol{T}$ and the lemma is proven.

The interpretation of the preceding lemma is based on the following fact. If $\boldsymbol{C}\overline{\boldsymbol{X}} \sim N_q(\boldsymbol{0}, \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^t)$, we get according to Lemma 2.1. $\boldsymbol{T}\boldsymbol{C}\overline{\boldsymbol{X}} = \boldsymbol{G}\overline{\boldsymbol{X}} \sim \boldsymbol{N}_{k-1}(\boldsymbol{0}, \boldsymbol{G}\boldsymbol{\Sigma}\boldsymbol{G}^t)$, where $\boldsymbol{G}\boldsymbol{\Sigma}\boldsymbol{G}^t$ is of the form $\begin{pmatrix} * & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$, i.e. only the upper $(k-1)\times(k-1)$ submatrix contains non-zero elements. With this result the first step has been taken to decrease the dimensionality problem. What has been left open is the question, how the integration region changes when applying the transformation $\boldsymbol{T}$. The answer is given in

Theorem 3.1.: Let $T^{MC} = \max\{T_1^{SC}, \ldots, T_q^{SC}\}$ be a MCT, where $T_l^{SC} = \frac{\boldsymbol{c}_l\overline{\boldsymbol{X}}}{s\sqrt{\boldsymbol{c}_l\boldsymbol{\Sigma}\boldsymbol{c}_l^t}}$ are the single contrast test statistics. Denote by $\boldsymbol{C} = (c_{ij})_{ij} = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_q)^t$ the $q \times k$ contrast matrix, $q \geq k-1$ and $\operatorname{rank}(\boldsymbol{C}) = k-1$. Then the corresponding distribution function can be expressed as a $(k-1)$–variate central *t*-distribution.

Proof: Without loss of generalisation we assume $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{k-1}$ to be linearly independent. If $q = k-1$, there is nothing to show. Therefore, assume $q > k-1$. With $\boldsymbol{T}$ from Lemma 3.3. set

$$\tilde{\boldsymbol{T}} = \boldsymbol{I}_q - \boldsymbol{T},$$

$\tilde{\boldsymbol{T}} = (\tilde{\boldsymbol{t}}_1, \ldots, \tilde{\boldsymbol{t}}_q)^t$. From the definition of the matrix $\boldsymbol{G}$ from Lemma 3.3. we get immediately

$$k = \min\{l = 1, \ldots, q: g_{lj} = 0 \ \forall j = 1, \ldots, k\}.$$

If we set further

$$m = m(l) = \max\left\{ j = 1, \ldots, l\colon \tilde{t}_{lj} \neq 0 \right\} < k,$$

then we have because of the construction of $\tilde{\boldsymbol{T}}$ for $l = k, \ldots, q$ and $d_l = \boldsymbol{c}_l \boldsymbol{\Sigma} \boldsymbol{c}_l^t$

$$T_l^{SC} = \frac{\tilde{\boldsymbol{t}}_l \boldsymbol{C} \overline{\boldsymbol{X}}}{s\sqrt{d_l}} = \frac{\sum_{i=1}^{k} \overline{X}_i \tilde{t}_{l.} c_{.i}}{s\sqrt{d_l}} = \frac{\sum_{i=1}^{k} \overline{X}_i \sum_{j=1}^{q} \tilde{t}_{lj} c_{ji}}{s\sqrt{d_l}} = \frac{\sum_{i=1}^{k} \overline{X}_i \sum_{j=1}^{m} \tilde{t}_{lj} c_{ji}}{s\sqrt{d_l}} =$$
$$= \frac{\sum_{j=1}^{m} \tilde{t}_{lj} \sum_{i=1}^{k} c_{ji} \overline{X}_i}{s\sqrt{d_l}} = \frac{\sum_{j=1}^{m} \tilde{t}_{lj} c_{j.} \overline{X}_{.}}{s\sqrt{d_l}} = \frac{\sum_{j=1}^{m} \tilde{t}_{lj} T_j^{SC} \sqrt{d_j}}{\sqrt{d_l}},$$

where the dot "." denotes the index, for which the matrix multiplication is defined. Therefore, we managed to represent each T_k^{SC}, T_{k+1}^{SC}, …, T_q^{SC} as a linear combination of at most $k-1 < q$ single contrast test statistics with the weighting coefficients $\tilde{\boldsymbol{t}}_l$. As we seek a simplified representation of $P\left(T_1^{SC} \leq c, \ldots, T_q^{SC} \leq c\right)$, we use the above facts and get the following equivalence:

$$T_l^{SC} \leq c \quad \Leftrightarrow \quad \frac{\sum_{j=1}^{m} \tilde{t}_{lj} T_j^{SC} \sqrt{d_j}}{\sqrt{d_l}} \leq c$$
$$\Leftrightarrow \quad T_m^{SC} \leq \frac{c\sqrt{d_l} - \sum_{j=1}^{m-1} \tilde{t}_{lj} T_j^{SC} \sqrt{d_j}}{\tilde{t}_{lm} \sqrt{d_m}}.$$

Each of the contrasts T_k^{SC}, T_{k+1}^{SC}, …, T_q^{SC} is fully described by the first *k* single contrasts. The sought probability term reduces therefore to an integral of dimension $k-1$ and the new integration bounds are the minimum over all conditioning equations on each random variable T_1^{SC}, T_2^{SC}, …, T_{k-1}^{SC}.

The following example will illustrate the techniques used so far.

Example 3.2.: Consider the Marcus contrast $(q = 6)$

$$\boldsymbol{C} = \begin{pmatrix} -1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & 0 & 1 \\ -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 1 \\ -1 & 0 & 0 & 1 \\ -1 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

in the balanced case $\boldsymbol{n} = (10, 10, 10, 10)$ with $\boldsymbol{\Sigma} = \mathrm{diag}(\frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10})$ and $k = 3$ (returning to the usual definition of *k*). According to Lemma 3.3. and the construction method contained in its proof we obtain

$$\boldsymbol{T} = \left(\begin{array}{ccc:ccc} 1 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ \hdashline 0 & \frac{2}{3} & -\frac{4}{3} & 1 & 0 & 0 \\ -\frac{3}{4} & \frac{1}{2} & -1 & 0 & 1 & 0 \\ -\frac{3}{4} & -\frac{1}{2} & 0 & 0 & 0 & 1 \end{array}\right) \text{ respectively } \boldsymbol{G} = \begin{pmatrix} -1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & -\frac{2}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

One checks equation (3.8) $\boldsymbol{TC} = \boldsymbol{G}$ and that $\boldsymbol{G\Sigma G}^t = \begin{pmatrix} \frac{4}{30} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{15} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{20} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$. The lower left block matrix of $\boldsymbol{T}$ therefore provides the weighting coefficients for the linear combinations obtained in the sequel. If we compute $(d_1, \ldots, d_q) = (\frac{2}{15}, \frac{1}{10}, \frac{3}{20}, \frac{2}{15}, \frac{2}{10}, \frac{3}{20})$ the following conditioning inequalities for the boundaries are yielded:

$$T_1^{SC} \le c,\ T_2^{SC} \le c,\ T_3^{SC} \le c,$$

$$T_4^{SC} \le c \quad \Leftrightarrow \quad T_3^{SC} \le \frac{c\sqrt{d_4} - \sum_{j=1}^{2} \tilde{t}_{4j} T_j^{SC} \sqrt{d_j}}{\tilde{t}_{43}\sqrt{d_3}} = \frac{c\sqrt{\frac{2}{15}} + \frac{2}{3}\sqrt{\frac{1}{10}} T_2^{SC}}{\frac{4}{3}\sqrt{\frac{3}{20}}}, \qquad (a)$$

$$T_5^{SC} \le c \quad \Leftrightarrow \quad T_3^{SC} \le \frac{c\sqrt{d_5} - \sum_{j=1}^{2} \tilde{t}_{5j} T_j^{SC} \sqrt{d_j}}{\tilde{t}_{53}\sqrt{d_3}} = \frac{c\sqrt{\frac{2}{10}} - \frac{3}{4}\sqrt{\frac{2}{15}} T_1^{SC} + \frac{1}{2}\sqrt{\frac{1}{10}} T_2^{SC}}{\sqrt{\frac{3}{20}}}, \qquad (b)$$

$$T_6^{SC} \le c \quad \Leftrightarrow \quad T_2^{SC} \le \frac{c\sqrt{d_6} - \sum_{j=1}^{1} \tilde{t}_{6j} T_j^{SC} \sqrt{d_j}}{\tilde{t}_{62}\sqrt{d_2}} = \frac{c\sqrt{\frac{3}{20}} + \frac{3}{4}\sqrt{\frac{2}{15}} T_1^{SC}}{\frac{1}{2}\sqrt{\frac{1}{10}}}. \qquad (c)$$

Finally, our original problem can be formulated as the three-dimensional integral

$$P\left(T_1^{SC} \le c, \ldots, T_q^{SC} \le c\right) = P\left(T_1^{SC} \le u_1, T_2^{SC} \le u_2, T_3^{SC} \le u_3\right),$$

where $u_1 = c$, $u_2 = \min\{c, \text{RHS of } (c)\}$ and $u_3 = \min\{c, \text{RHS of } (a), \text{RHS of } (b)\}$. Instead of evaluating the original six-variate *t*-distribution, the above derivation shows that it is sufficient to compute a three-dimensional one with upper integration bound (u_1, u_2, u_3). Evidently the gain of time gets rather large for increasing *k*. Computing above example for *c* = 1 on a Celeron 333 with a modified version of program B.2.9 one computes the original integral in about 30 seconds ($\varepsilon = 0.001$) respectively 30 minutes ($\varepsilon = 0.0001$). The trivariate integral however is done in 2 respectively 5 seconds. Therefore, high numbers *k* of treatment groups should be no obstacle to conduct MCTs, as the evaluation requires at most a *k*-variate integral, irrespective, which contrast sets are used (Marcus, isotonic, ...). And the numerical comparisons in Chapter 2 have shown that the multivariate *t*-distribution can be used up to *k* =10 without severe time limitation for an $\varepsilon = 0.001$ or even $\varepsilon = 0.0001$ in SAS/IML. Recently, Genz and Kwong (1999) have derived independently a similar method for computing singular multivariate normal probabilities without restrictions concerning ***R***, *k* and integration bounds.

4. Power comparison for normal data

In the following we compare in extensive the new derived MCTs of the preceding chapter with the original approaches of Williams, Marcus and the LRT of Bartolomew under a variety of conditions. Before this is done in Section 4.2., however, we first establish a closed form expression of the power function for arbitrary MCTs. This provides us a great advantage when working with MCTs, as no power formulas for the other tests considered here are available for general dimension *k*. We will have a look at the numerical aspects of evaluating the arising multivariate non-central *t*-distribution. Further, we shall discuss briefly the important practical issue of sample size determination for the design of experiments. In the final Section 4.3. we summarise the results obtained in 4.2. and give some conclusions to the practitioner.

4.1. Power expression of multiple contrast tests

Recall the notation given in Section 1.1. and in the introductory Subsection 1.3.4. of MCTs. We now examine the behaviour of MCTs in the alternative space. The following main result due to Genz and Bretz (1999) holds.

Theorem 4.1.: Let $T^{MC} = \max\left\{T_1^{SC}, \ldots, T_q^{SC}\right\}$ be a MCT, where $T_l^{SC} = \frac{\boldsymbol{c}_l \overline{\boldsymbol{X}}}{s\sqrt{\boldsymbol{c}_l \boldsymbol{\Sigma} \boldsymbol{c}_l^t}}$ are the single contrast test statistics. Denote by $\boldsymbol{C} = \left(c_{ij}\right)_{ij} = \left(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_q\right)^t$ the $q \times \left(k+1\right)$ contrast matrix. Under the alternative T^{MC} can be expressed according to a q–variate non-central t–distribution with the noncentrality parameter $\delta = \left(\frac{\sum_{i=0}^{k} c_{li}\mu_i}{\sigma\sqrt{\sum_{i=0}^{k} c_{li}^2/n_i}}\right)_{1 \le l \le q}$, ν degrees of freedom and the correlation matrix $\boldsymbol{R}$ given by Lemma 1.4.

Proof: We consider the behaviour under H_A by examining the power definition $P\left(T^{MC} \ge t \,\middle|\, H_A\right)$. Then the following steps are valid.

$$
\begin{aligned}
\text{Power} &= P\left(T^{MC} \geq t \,\middle|\, H_A\right) = \\
&= P\left(\max_{1\leq l\leq q} \left(\frac{\sum_{i=0}^{k} c_{li}\overline{X}_i}{s\sqrt{\sum_{i=0}^{k} c_{li}^2/n_i}} \right) \geq t \,\middle|\, H_A \right) = \\
&= 1 - P\left(\max_{1\leq l\leq q} \left(\frac{\sum_{i=0}^{k} c_{li}\overline{X}_i}{s\sqrt{\sum_{i=0}^{k} c_{li}^2/n_i}} \right) < t \,\middle|\, H_A \right) = \\
&= 1 - P\left(\frac{\sum_{i=0}^{k} c_{1i}\overline{X}_i}{s\sqrt{\sum_{i=0}^{k} c_{1i}^2/n_i}} < t \;\wedge\; \ldots \;\wedge\; \frac{\sum_{i=0}^{k} c_{qi}\overline{X}_i}{s\sqrt{\sum_{i=0}^{k} c_{qi}^2/n_i}} < t \,\middle|\, H_A \right) = \\
&= 1 - P\left(\frac{\frac{\sum_{i=0}^{k} c_{1i}\left(\overline{X}_i - \mu_i\right)}{\sigma\sqrt{\sum_{i=0}^{k} c_{1i}^2/n_i}} + \frac{\sum_{i=0}^{k} c_{1i}\mu_i}{\sigma\sqrt{\sum_{i=0}^{k} c_{1i}^2/n_i}}}{s/\sigma} < t \;\wedge\; \ldots \;\wedge\; \frac{\frac{\sum_{i=0}^{k} c_{qi}\left(\overline{X}_i - \mu_i\right)}{\sigma\sqrt{\sum_{i=0}^{k} c_{qi}^2/n_i}} + \frac{\sum_{i=0}^{k} c_{qi}\mu_i}{\sigma\sqrt{\sum_{i=0}^{k} c_{qi}^2/n_i}}}{s/\sigma} < t \right). \qquad (4.1)
\end{aligned}
$$

The assertion now follows from the representation (4.1) and one has only to show that the entries of $\boldsymbol{R}$ under H_A are indeed the same as under H_0. This is proven in the next Lemma.

Lemma 4.2.: For two SCTs T_1^{SC} and T_2^{SC} as defined in (1.11) the correlation $\rho = \mathrm{Corr}\left(T_1^{SC},\, T_2^{SC}\right)$ under H_A is given by

$$
\rho = \frac{\sum_{i=0}^{k} c_{1i}c_{2i}/n_i}{\sqrt{\left(\sum_{i=0}^{k} c_{1i}^2/n_i\right)\left(\sum_{i=0}^{k} c_{2i}^2/n_i\right)}}.
$$

Proof: By definition of the multivariate *t*–distribution it is sufficient to consider the correlation of $\left(X,\, Y\right) = \left(\sum c_{1i}\overline{X}_i,\, \sum c_{2i}\overline{X}_i\right) \sim N_2\left(\mu,\, \Sigma\right)$. Then the covariance is given by

$$
\begin{aligned}
\mathrm{Cov}\left(X,\, Y\right) &= E\left(X - EX\right)\left(Y - EY\right) = \\
&= E\left(\sum c_{1i}\overline{X}_i - E\left(\sum c_{1i}\overline{X}_i\right)\right)\left(\sum c_{2i}\overline{X}_i - E\left(\sum c_{2i}\overline{X}_i\right)\right) = \\
&= E\left(\sum c_{1i}\overline{X}_i - \sum c_{1i}\mu_i\right)\left(\sum c_{2i}\overline{X}_i - \sum c_{2i}\mu_i\right) = \\
&= \sum_i \sum_j c_{1i}c_{2j} E\left(\overline{X}_i - \mu_i\right)\left(\overline{X}_j - \mu_j\right) =
\end{aligned}
$$

$$= \sum_{i=0}^{k} c_{1i}c_{2i}E\left(\left(\overline{X}_i - \mu_i\right)^2\right) + \sum_{i,\,j=1,\,i\neq j}^{k} c_{1i}c_{2j} \underbrace{E\left(\left(\overline{X}_i - \mu_i\right)\left(\overline{X}_j - \mu_j\right)\right)}_{=\,0 \text{ under the independence of } \overline{X}_i \text{ and } \overline{X}_j} =$$

$$= \sum_{i=0}^{k} c_{1i}c_{2i}V\left(\overline{X}_i - \mu_i\right) =$$

$$= \sum_{i=0}^{k} c_{1i}c_{2i}V\left(\overline{X}_i\right) =$$

$$= \sum_{i=0}^{k} c_{1i}c_{2i}\,\frac{\sigma^2}{n_i}.$$

Because of $\rho = \mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$ above assertion follows directly when taking $\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)} = \sigma^2\sqrt{\left(\sum_{i=0}^{k}\frac{c_{1i}^2}{n_i}\right)\left(\sum_{i=0}^{k}\frac{c_{2i}^2}{n_i}\right)}$ into account.

With these results we are therefore able to calculate the power of an arbitrary MCT by a closed form expression. Examples of their evaluation are given in the power study of the next section. Instead, we focus now our attention on the numerical evaluation of (4.1), which involves the computation of multiple integrals belonging to the multivariate non-central *t*–distribution similar to the still unsolved problem (C) of Subsection 2.2.2.4. Let us denote the corresponding multivariate distribution function by $T_{\delta,\boldsymbol{R},\nu}(-\infty, \boldsymbol{t})$, where a_l = -∞ $\forall l$. In equation (4.1) $\boldsymbol{t} = (t, \ldots, t)$ stands for the $(1-\alpha)-$equipercentage point of the central *q*–variate *t*–distribution, so that $T(-\infty, \boldsymbol{t}) = 1-\alpha$, where α is the pre-specified type-I-error. We now make use of the relationship

$$P(\boldsymbol{t}'_\nu \leq \boldsymbol{b}) = \Pr\left(\boldsymbol{U} \leq \boldsymbol{b}\chi_\nu / \sqrt{\nu} - \delta\right).$$

Here $\boldsymbol{U}$ and χ_ν are independent random variables distributed as standardised *q*-variate normal and chi with ν degrees of freedom, $\boldsymbol{t}'_\nu = \boldsymbol{t}'_\nu(\delta) = (\boldsymbol{U} + \delta) \Big/ \left(\chi_\nu / \sqrt{\nu}\right)$. This leads to the explicit representation of the non-central *q*–variate *t*–distribution function $T_{\delta,\boldsymbol{R},\nu}(-\infty, \boldsymbol{b})$ by

$$P(\boldsymbol{t}'_\nu \le \boldsymbol{b}) = \frac{1}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} \int_0^\infty x^{\nu-1} e^{-\frac{x^2}{2}} \Phi_q\left(\boldsymbol{b}x/\sqrt{\nu} - \delta; \boldsymbol{0}, \boldsymbol{R}\right) dx, \tag{4.2}$$

where $\Phi_q(.)$ is the q–variate normal integral with expectation $\boldsymbol{0}$, correlation matrix $\boldsymbol{R}$ and upper integration bound $\boldsymbol{b}x/\sqrt{\nu} - \delta$. A correct evaluation of the probability term in equation (4.1) therefore consists of the following numerical problems. First we need the ability of computing the non-central q–variate t–distribution function $T_{\delta,\boldsymbol{R},\nu}(-\infty, \boldsymbol{b})$. This can be done by using the transformation indicated in equation (4.2). For the evaluation of the arising q-variate standard normal distribution $\Phi_q(.)$ several methods can be applied. We refer to the results obtained in Section 2.1. for this problem. The remaining single integral can be computed by using standard univariate integration techniques. Finally, we need to compute the critical value $\boldsymbol{t}$ in equation (4.1). The vector is the implicit solution of $h(\boldsymbol{t}) = 0$, where $h(\boldsymbol{t}) = T(-\infty, \boldsymbol{t}) - 1 + \alpha$. Therefore, the computation reduces to the problem discussed in Subsection 2.2.2.3. and the methods stated there can be applied. The evaluation of $h(\boldsymbol{t})$, on the other hand, requires the computation of $T(\boldsymbol{a}, \boldsymbol{b})$, what can be done by using the algorithms provided in Section 2.2.2.2.

One main objective in practical design studies is the determination of the required sample size for a specific experiment. For given values of α, β, effect vector $\mu = (\mu_0, \ldots, \mu_k)$ and a known or estimated (from preceding pilot studies, for example) variance σ^2 one needs to compute that sample size $\boldsymbol{n}$, which satisfies all these conditions. When using MCTs this problem can be solved by using above power expression iteratively. Because of the multivariate nature of the distribution (the inverse function of $T(\boldsymbol{a}, \boldsymbol{b})$ is not unique), closed form expressions for sample size determination do not exist. An iterative procedure, however, works always satisfactorily. In a sequential manner one calculates the power according to (4.1) for different sample size vectors $\boldsymbol{n}$ and increases or decreases the sample sizes in accordance with the resulting power values. If the power value lies below the pre-determined threshold, $1-\beta$ say, the sample size is enlarged by a certain amount. Otherwise, the power value lies above $1-\beta$ and one reduces $\boldsymbol{n}$ in order to find the minimum sample size constellation, which still yields a power above $1-\beta$.

4.2. Power study

It is impossible to compare the tests of interest from one point of view only. The six tests to be considered further include the approaches of Williams, Marcus and Bartholomew as well as the corresponding MCTs (Williams, Marcus and isotonic). We will investigate them by considering several aspects, for example the influence of the total sample size, the allocation of the sample sizes within the groups, degree of imbalance, the choice of a predefined α, the a-priori unknown dose-response relationship or the number of treatment groups among several other topics. This is done in Figures 4.1. through 4.4. In Figures 4.5. through 4.8. we will additionally investigate specific types of violations of the assumption, such as non-normal data, non-monotonic dose-response shapes and variance heterogeneity. Two further comparisons, Figure 4.9. and Table 4.1. give a deeper insight into the performance of the tests. This comparison study should be understood as an outlook, only. Because of the great number of influencing parameters no claim on completeness is made. Further on, one may argue that some of the situations analysed in the following are of minor practical importance, for example sample size allocations of the form $\boldsymbol{n} = (2, 18, 18, 2)$. But we want to give an overview as broad as possible within the restricted space of this thesis. And sometimes there are in fact real data sets following such 'pathological' patterns, as it is the example of the insect *Eupromus ruber* in the Introduction with the sample size varying between 4 and 21.

The following power study was conducted using the SAS-algorithms provided in the preceding chapters. The concrete constellation of the parameters is given in the legend of each figure. The power values of the MCTs are exact up to at least two significant digits (Solow procedure; Figs. 4.3. through 4.5. and 4.7. and 4.8.) or four digits (lattice rule implementation; Fig. 4.9. and Table 4.1.). They were achieved by virtue of Theorem 4.1. and the discussion afterwards. All other results (power and size) were obtained by simulation techniques with 10,000 runs each. The generation of the normal random samples was conducted with the SAS- function `RANNOR`. Because of the binomial nature of the outcomes the expected magnitude of the error due to simulation is estimated by the asymptotic two-sided 95%-confidence interval

$$\left[\hat{p} - 1.96\sqrt{\frac{p(1-p)}{10{,}000}};\ \hat{p} + 1.96\sqrt{\frac{p(1-p)}{10{,}000}}\right].$$

For example, for the size estimation p = 0.05 one expects the estimate to lie within $[0.0457;\ 0.0543]$. For the determination of quantiles exact values were available for Williams' $\bar{t}$ –test (balanced case), the LRT and all MCTs. Otherwise, 9,999 simulation runs were additionally conducted to obtain an estimate for them (Marcus' test, unbalanced $\bar{t}$ –test).

First we shall examine the performance of all six tests under the null distribution for normal distributed data. Figure 4.1. shows the size estimation for three levels of α (0.01, 0.05 and 0.1) in both balanced and unbalanced set-ups. On the abscise the sample size n_0 of the control group is given, whereas the boxes inside the Figure indicate the sample size allocation $\boldsymbol{n} = (n_0, n_1, n_2, n_3)$ belonging to it. All curves lie well below the corresponding upper confidence bounds for the simulation (bold straight lines). One concludes therefore that all tests in fact maintain the α–level. Williams' test tends even to be a little conservative for higher values of α.

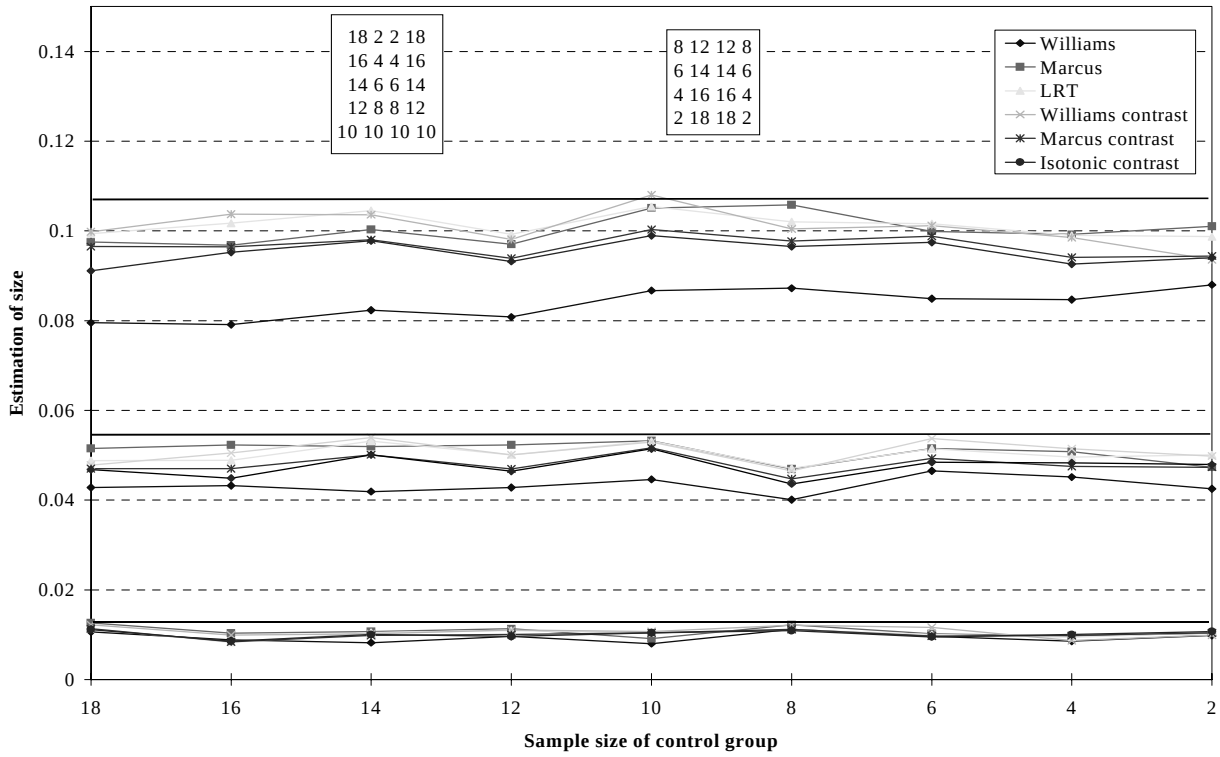


Figure 4.1. Estimation of size, unbalanced case with sample size allocation given in the Figure, k = 3, total sample size = 40 for α = 0.01, 0.05 and 0.1.

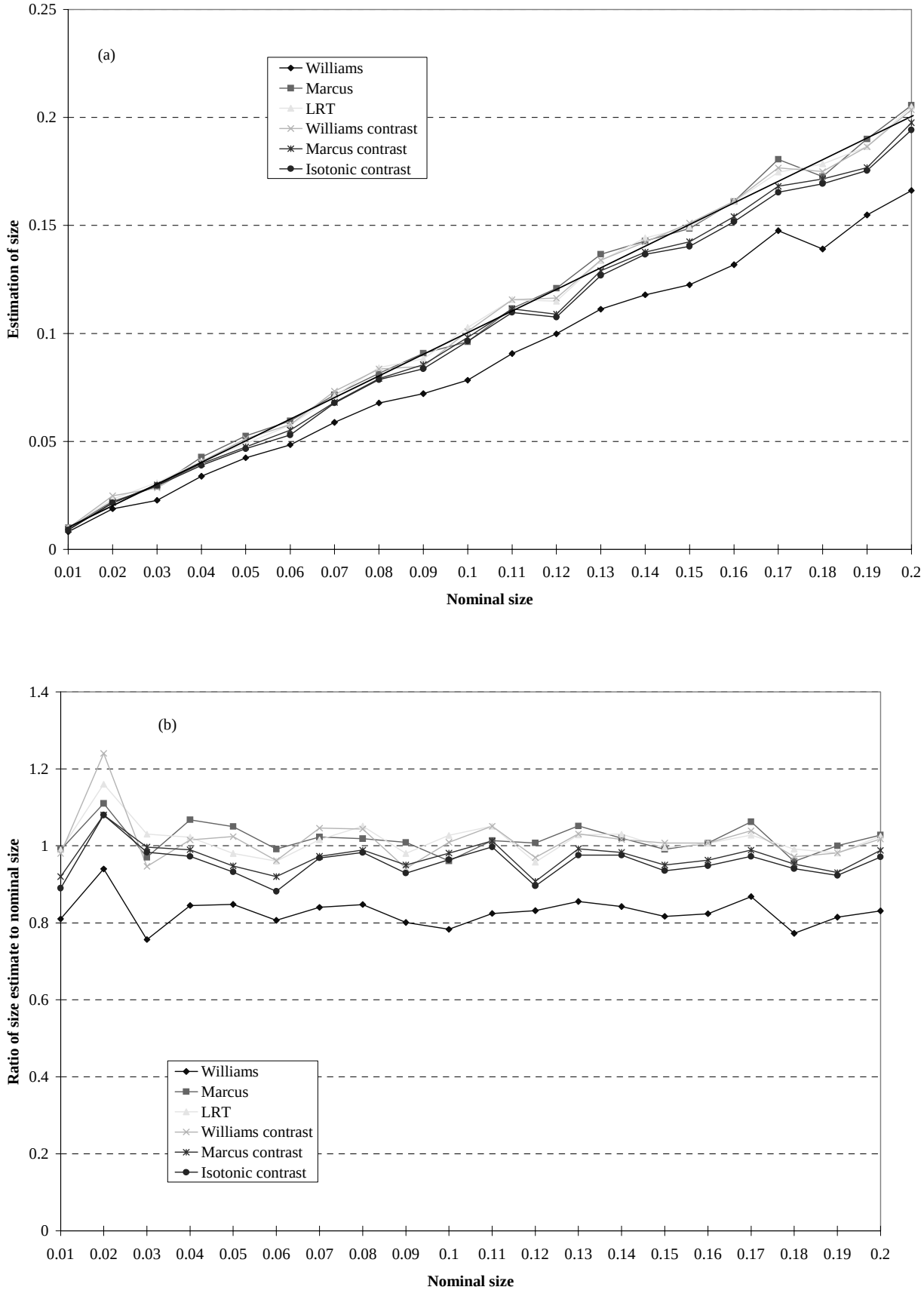


Figure 4.2. Estimation of size, unbalanced with sample size allocation $(14, 10, 10, 6)$, $k = 3$ for (a) absolute deviations, and (b) ratio to nominal size.

To further investigate empirically the conservative nature of Williams test, we estimated different sizes by simulation. Figure 4.2. shows clearly that the $\bar{t}$–test in fact maintains its size 20% below the nominal size over the whole range investigated $(0.01 \le \alpha \le 0.2)$. In contrast, all other tests keep the ratio $\hat{\alpha}/\alpha$ at the constant ratio 1 and hence maintain the nominal size fairly well. Figure 4.2.(a) shows the absolute deviations of $\hat{\alpha}$ from α and one notices the consequences for higher values of α.

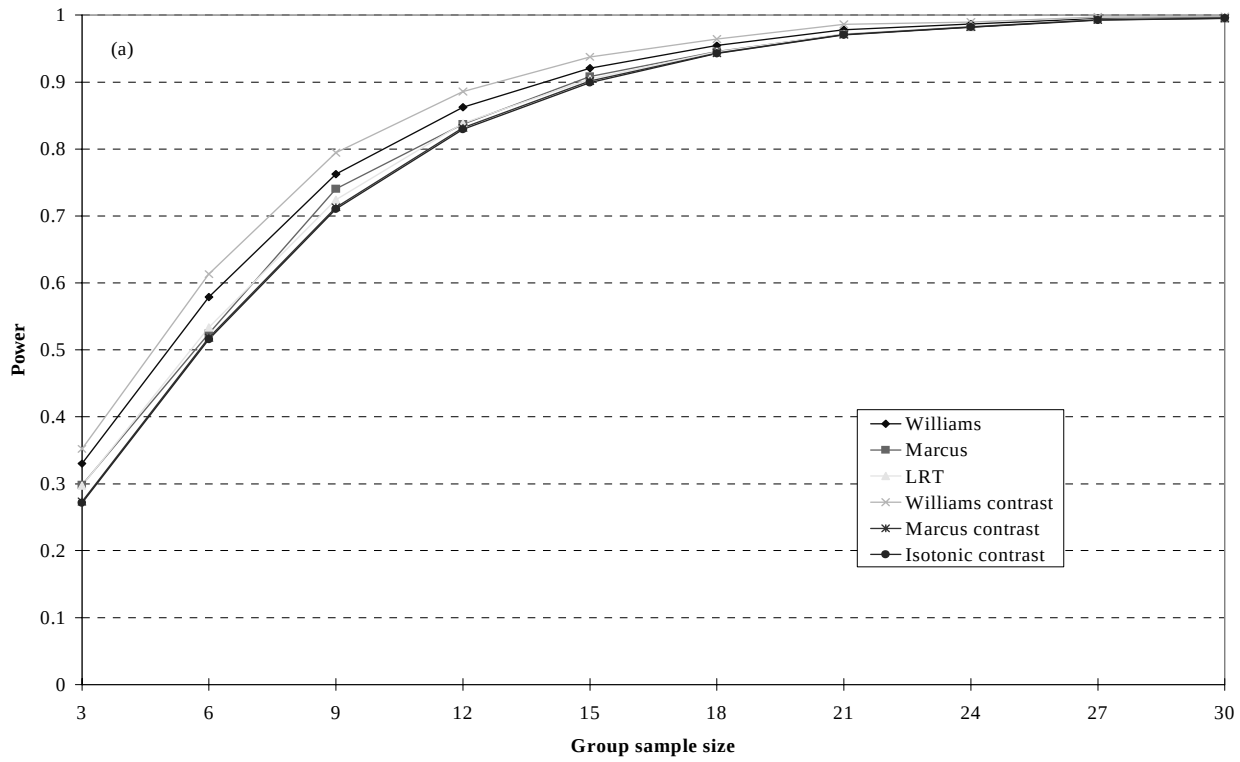
Figure 4.3. investigates the power of the tests for different dose-response shapes in the balanced situation. In the concave case (a) (the first efficient dose step occurs at low doses) both Williams' test and Williams MCT are uniformly better than the other tests. But with increasing departures from this shape the power gains vanish and become negligible around a linear shape (b). In the other extreme, convex profile of the μ_i's (the first efficient dose step occurs at high doses), both tests are inferior to the other ones. This behaviour is indeed consistent with the discussion following Definition 3.2. (p. 86) and the case of contrasts specifically designed for concave profiles. Moreover, one can demonstrate the power differences between Williams and the corresponding MCT analytically. Imagine the situation for $k = 3$. Then the isotonic estimate $\hat{\mu}_k$ just builds the maximum over

$$\left\{ \frac{n_1\overline{X}_1 + n_2\overline{X}_2 + n_3\overline{X}_3}{n_1 + n_2 + n_3},\ \frac{n_2\overline{X}_2 + n_3\overline{X}_3}{n_2 + n_3},\ \overline{X}_3 \right\}.$$

Calculating the variances of each term gives

$$\left\{ \frac{\sigma^2}{n_1 + n_2 + n_3},\ \frac{\sigma^2}{n_2 + n_3},\ \frac{\sigma^2}{n_3} \right\}.$$

Williams did not take this fact into consideration for his statistic $\bar{t}$. He just included the worst case, which is here σ^2/n_3 (yielding the maximum value of the three terms, resulting in the smallest value of the whole statistics). This approach has the drawback of an increased conservativeness, which could be demonstrated in the simulations. In the convex case this approach does not change the overall result, because the (fixed) choice of the variance is then correct. But for concave profiles the adequate choice results in higher test statistics, thus leading to more power. The simplification of the test statistics is possible only in cost of power.



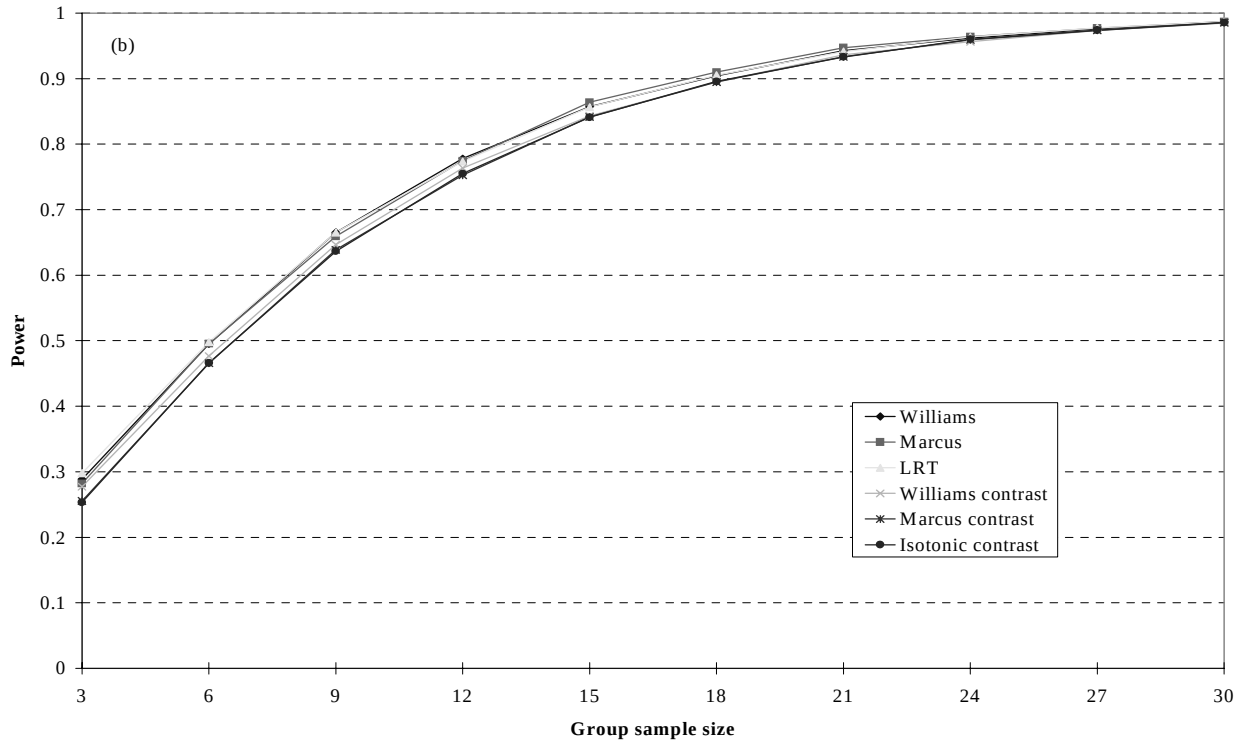


Figure 4.3. Power comparison, balanced case, $\alpha = 0.05$, $k = 3$ for (a) concave profile $\mu = (0, 1, 1, 1)$, (b) linear profile $\mu = (0, \frac{1}{3}, \frac{2}{3}, 1)$ and (c) convex profile $\mu = (0, 0, 0, 1)$.

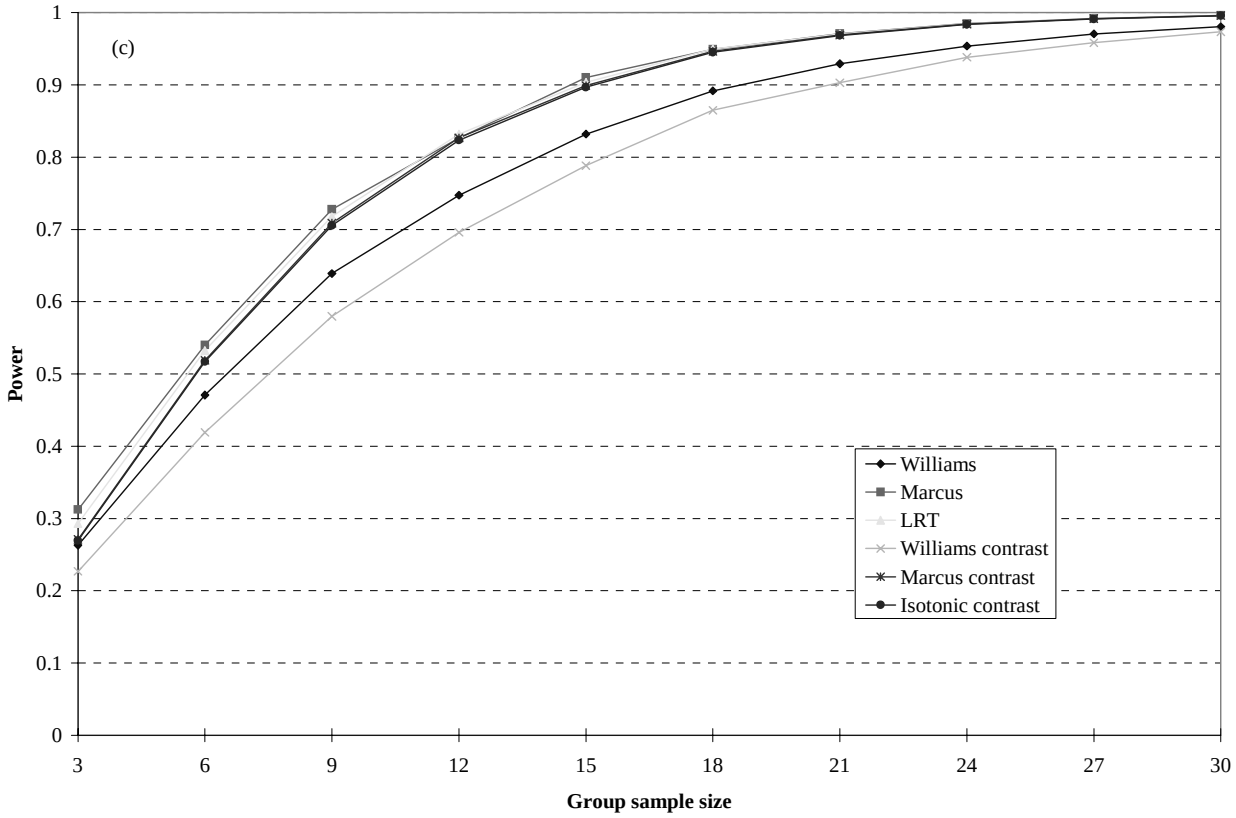


Figure 4.3. *(continued)*

Next we consider in Figure 4.4. the unbalanced situation. The results are complex and have to be differentiated with respect to dose-response shapes and whether the control group has more sampling units $(n_0 > n_i)$ or less $(n_i > n_0)$ than the remaining dose groups. Williams' MCT is best for all sample size allocations investigated in the concave case (a). The power advantages diminish for non-concave shapes (b) and (c). Marcus original test behaves poor in (a) for $n_0 > n_i$ (up to 40% power loss), but is reasonable for $n_0 < n_i$. For the step profile (b) it is uniformly dominated by the LRT and both Marcus' and isotonic MCT. For convex profiles (d), however, Marcus' $\bar{t}^{\,\mathrm{mod}}$ is much better than Williams' version for all sample size situations. In general, the LRT and both Marcus' and the isotonic MCT seem to be least dependent on shape and sample size allocation. All three tests perform similar, if not identical, in terms of power. For strong concave profiles they are beaten by Williams' MCT, but they behave best for the step (b) and linear (c) shapes. Finally, for convex curves and $n_0 < n_i$ they yield a power gain of over 40% with respect to the other procedures.

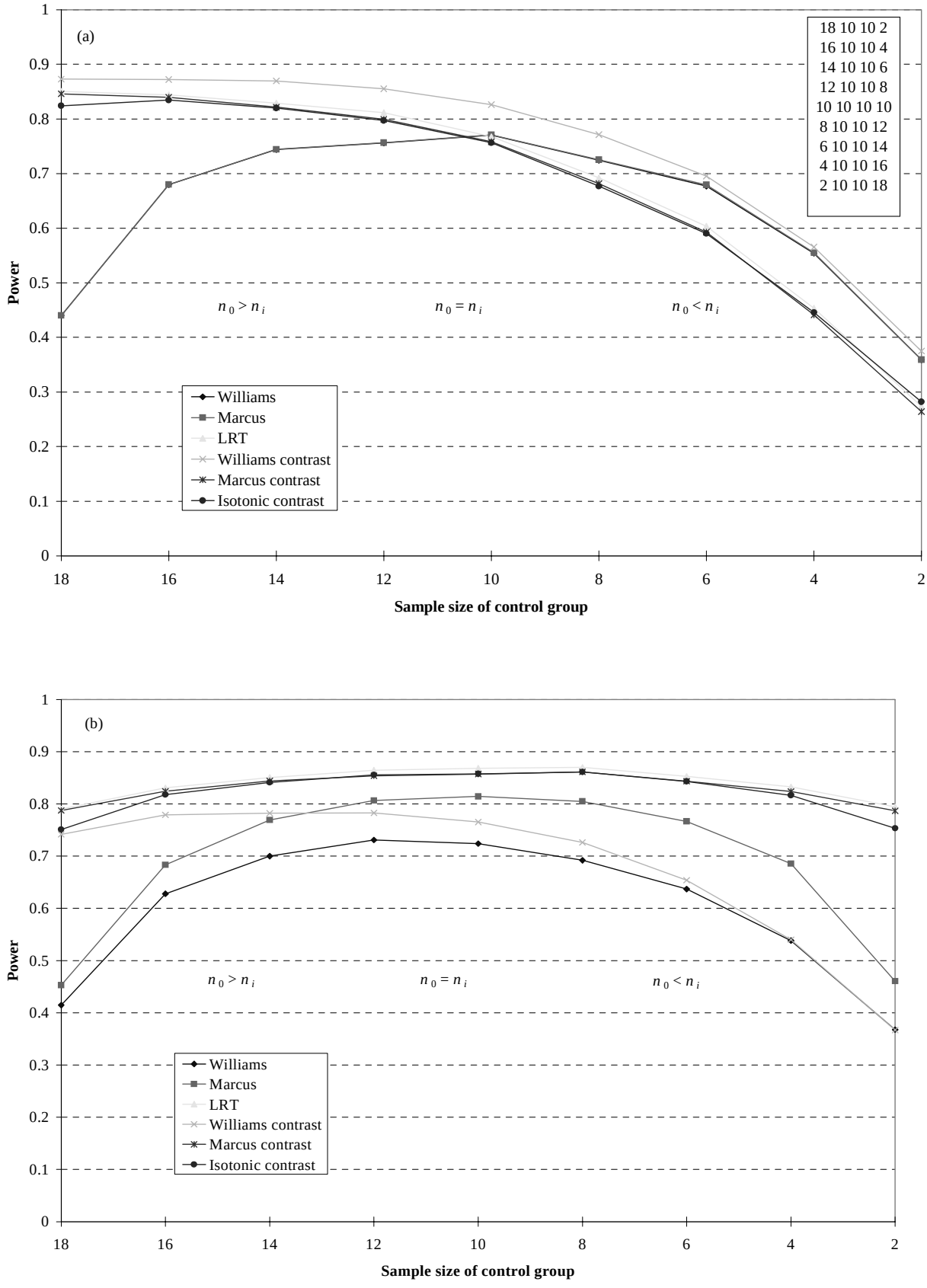
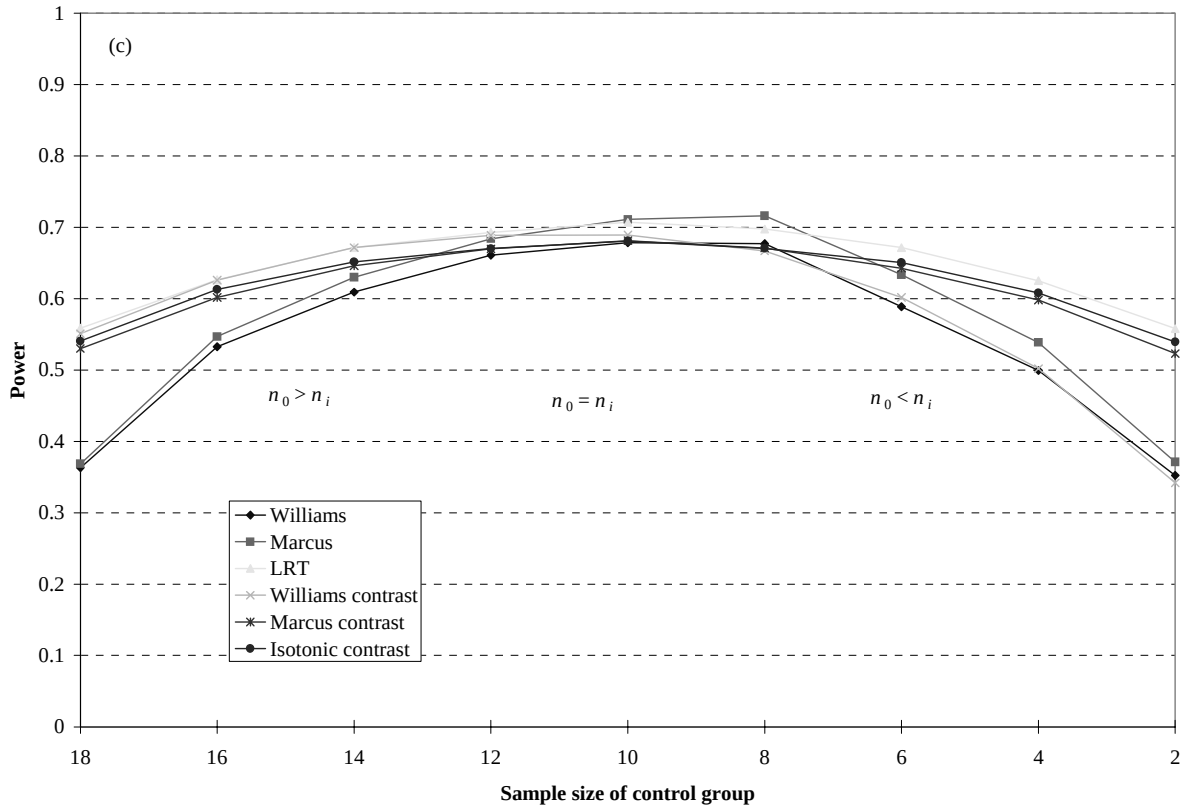


Figure 4.4. Power comparison, unbalanced case with sample size allocation given in (a), total sample size = 40, α = 0.05, k = 3 for (a) concave profile, $\mu = (0, 1, 1, 1)$, (b) step profile, $\mu = (0, 0, 1, 1)$, (c) linear profile, $\mu = (0, \frac{1}{3}, \frac{2}{3}, 1)$ and (d) convex profile, $\mu = (0, 0, 0, 1)$.



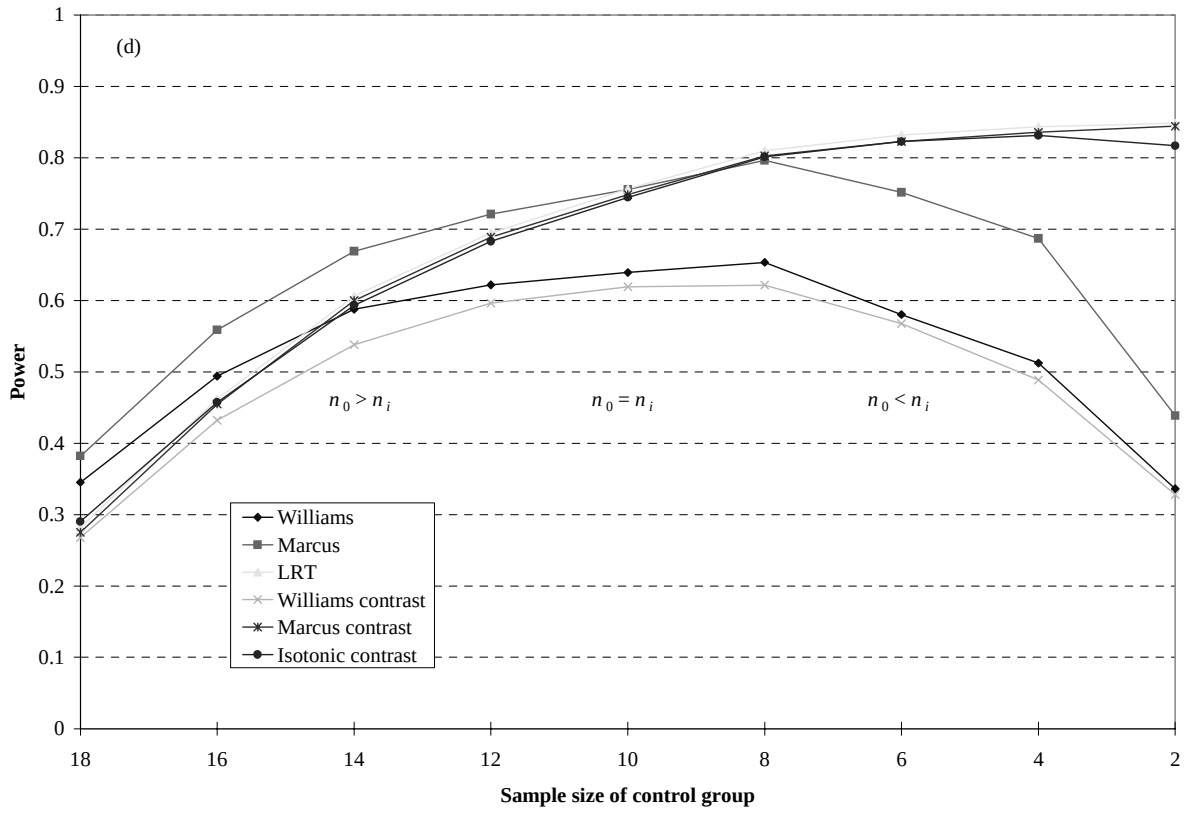


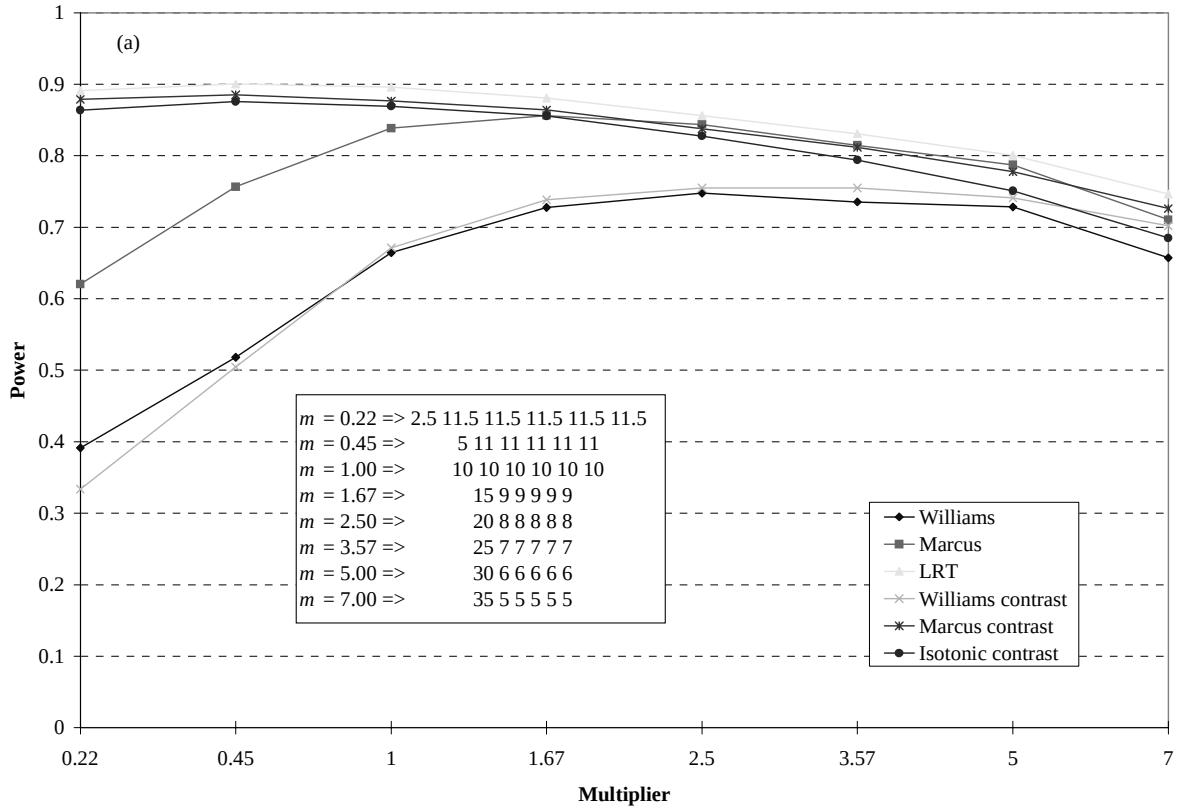
Figure 4.4. *(continued)*

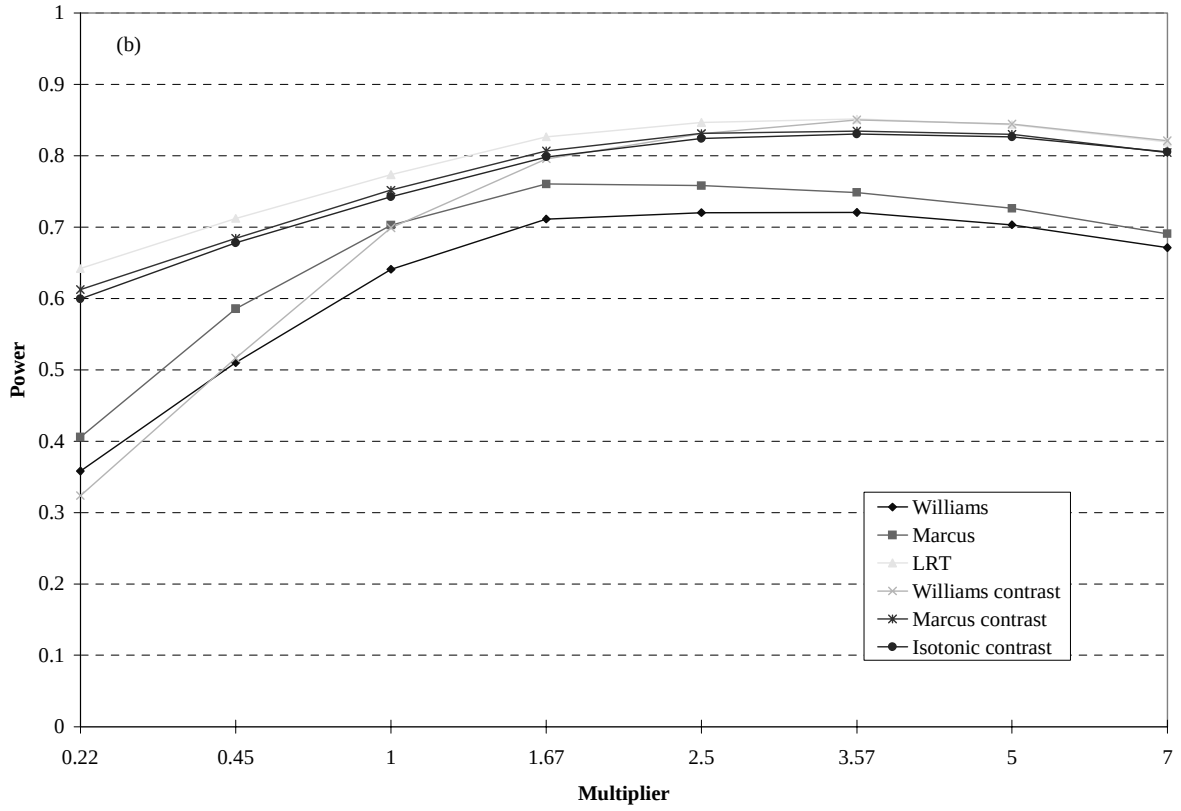
Figure 4.5. investigates a series of situations and shapes different from the usually analysed ones in the literature. Typical violations of the monotonicity assumption (1.2) are sketched in

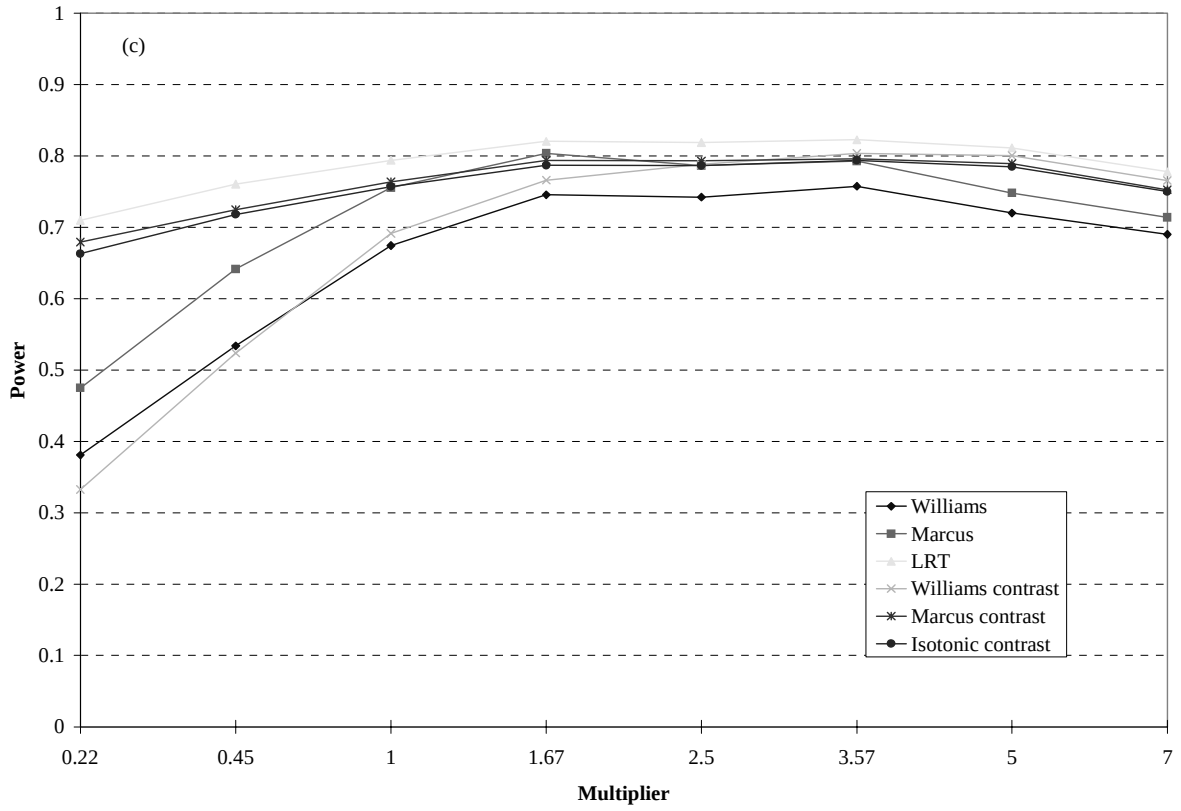
(a) U-shape at low doses, and
(b) down-turn at high doses. Case
(c) reflects a typical S-shaped response, a pattern different from strict linear and convex,
(d) investigates a saturation-type pattern.

The situations were investigated for partial balanced cases $n_0 = mn_i$, where $n_i = n \quad \forall i = 1, \ldots, k$. The multiplier *m* denotes the ratio of sample size between control and dose group. The motivation for these special sample size allocations arises from the $\sqrt{k}-$rule (e.g. Hochberg and Tamhane, 1987, p. 168), first established by Dunnett (1955) in the many-to-one context. No such rules were derived in the literature for the present trend situations. However, Williams (1972, p. 523) concluded for his test that "... *the optimum value ... varies for different cases but generally lies between* $1.1\sqrt{\mathrm{k}}$ and $1.4\sqrt{\mathrm{k}}$." In fact we can conclude the validity of this conjecture for the other trend tests as well from Figure 4.5.(c) (the only figure, which assures the monotonicity assumption). The optimum value indeed seems to lie within the range given by Williams (1972). On the other side, the curves are very flat and the power differences are marginal over a broad range of values for *m*. For the non-monotonous situations considered in a), b) and d) such general assessments regarding optimum sample size allocations does not hold any more.

All in all the tests crystallised above (LRT, Marcus and isotonic MCT) perform again similar to each other and are better than the other tests. They are less shape dependent and are therefore more robust against such specific dose-response patterns. Only in the last situation Williams' MCT is competitive, yielding power values which lie sometimes more than 5% above the values of the three distinguished tests. This performance is well explained by the approximate underlying concave dose response shape.


(a)
Power
Multiplier
0.22
0.45
1
1.67
2.5
3.57
5
7
m = 0.22 => 2.5 11.5 11.5 11.5 11.5 11.5
m = 0.45 => 5 11 11 11 11 11
m = 1.00 => 10 10 10 10 10 10
m = 1.67 => 15 9 9 9 9 9
m = 2.50 => 20 8 8 8 8 8
m = 3.57 => 25 7 7 7 7 7
m = 5.00 => 30 6 6 6 6 6
m = 7.00 => 35 5 5 5 5 5
Williams
Marcus
LRT
Williams contrast
Marcus contrast
Isotonic contrast


(b)
Power
Multiplier
0.22
0.45
1
1.67
2.5
3.57
5
7
Williams
Marcus
LRT
Williams contrast
Marcus contrast
Isotonic contrast



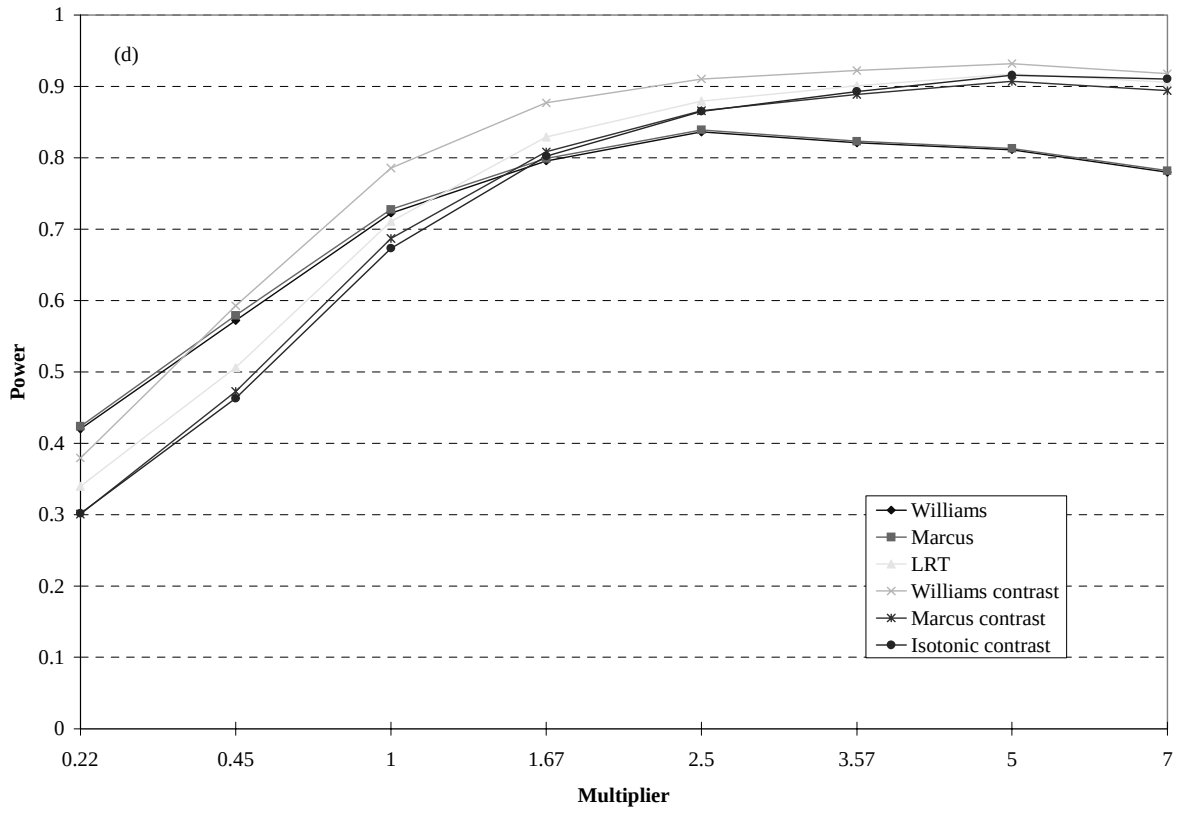


Figure 4.5. Power comparison, partially balanced case with sample size allocation given in (a), total sample size = 60, α = 0.05, k = 5 for (a) $\mu = \left(0, -\frac{1}{5}, 0, \frac{1}{3}, \frac{2}{3}, 1\right)$, (b) $\mu = (0, 0.2, 0.4, 0.8, 1, 0.8)$, (c) $\mu = (0, 0.1, 0.3, 0.6, 0.7, 1)$ and (d) $\mu = (0, 0.7, 0.9, 0.8, 0.9, 1)$.

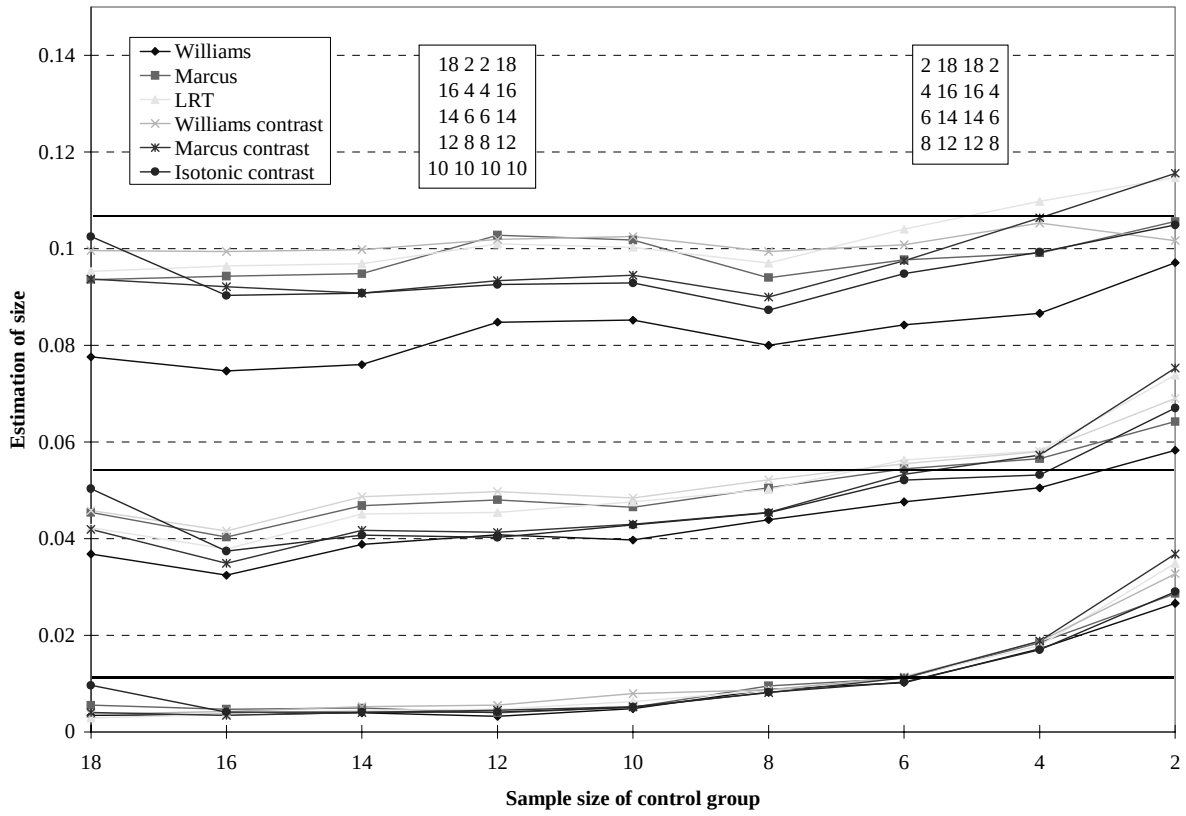
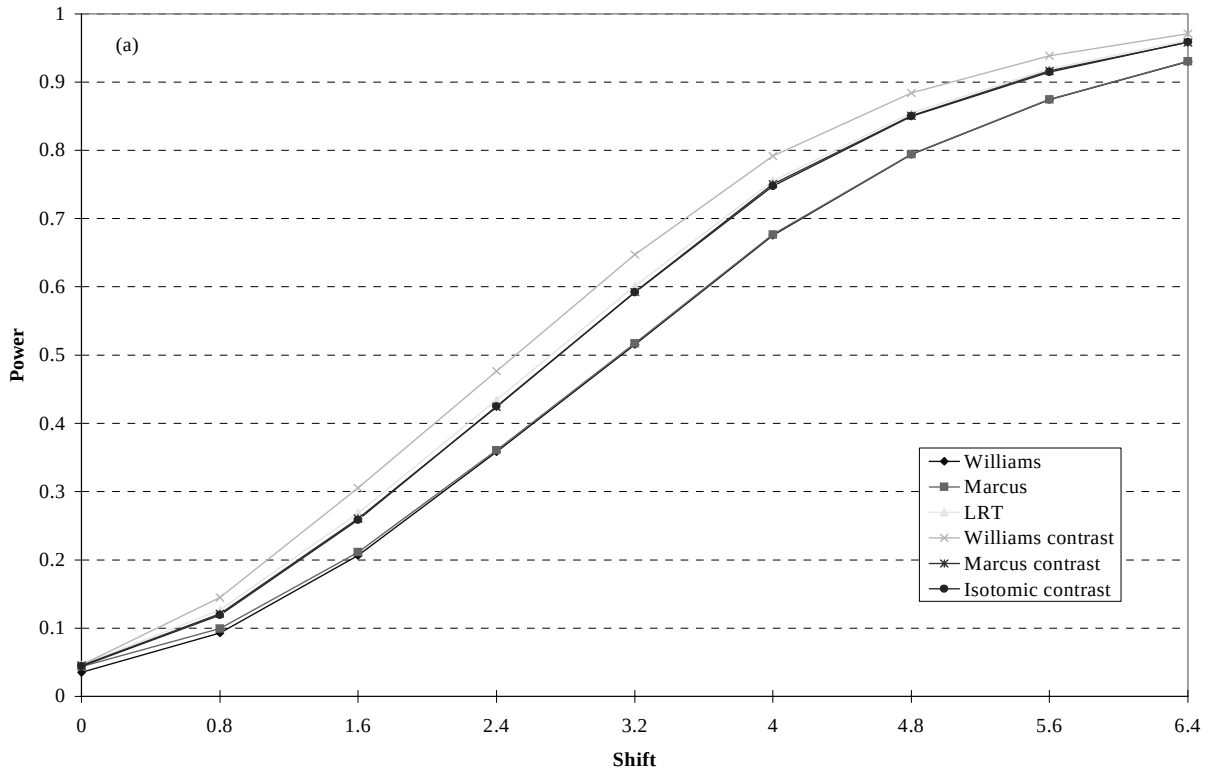


Figure 4.6. Estimation of size, contaminated normal data, unbalanced case with sample size allocation given in the Figure, k = 3, total sample size = 40 for α = 0.01, 0.05 and 0.1.

Next we look at the behaviour of the tests for non-normal data, though possessing equal variances. The contaminated normal distribution $pN(0, 1)+(1-p)N(0, \sigma^2)$ is examined in Figures 4.6 and 4.7. for p = 0.8 and σ^2 = 10. The size estimates for different imbalances of the six tests are given in Figure 4.6. in the same sample size set-up of Figure 4.1. All tests maintain the size surprisingly well. The LRT is somewhat more liberal than Williams' test, as already noticed in the Introduction. While the size estimates of the LRT lie eight times above the respective upper confidence intervals for the nominal sizes (bold lines), this is only in three times the case for Williams' $\bar{t}$. Analogously, Williams' MCT is less liberal than the other two MCTs, in particular Marcus' contrast. Similar results were also obtained for other values of σ^2 (20 and 30) and other types of underlying non-normal distributuins (Cauchy, chi, lognormal, ...).



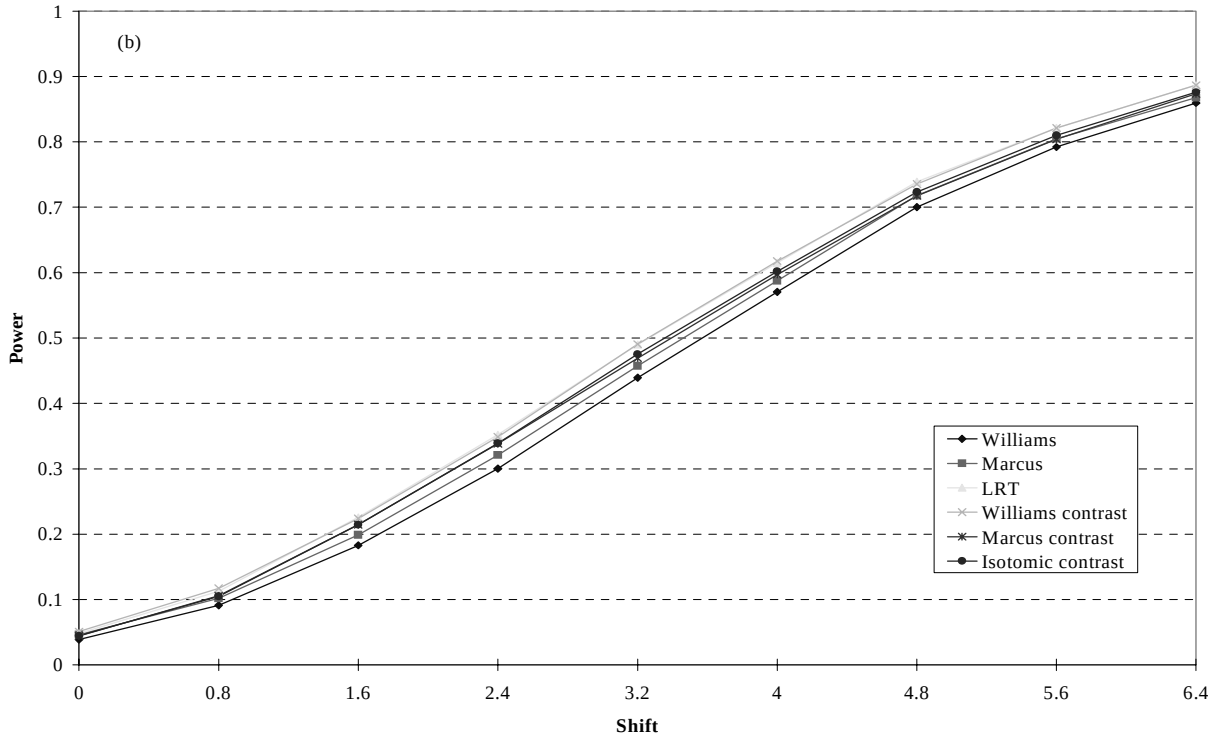


Figure 4.7. Power comparison, contaminated normal data, unbalanced case with the sample size allocation $(14, 10, 10, 6)$, $\alpha = 0.05$, $k = 3$ for (a) concave profile, $\mu = (0, \theta, \theta, \theta)$, (b) linear profile, $\mu = \left(0, \frac{\theta}{3}, \frac{2\theta}{3}, \theta\right)$ and (c) convex profile, $\mu = (0, 0, 0, \theta)$.

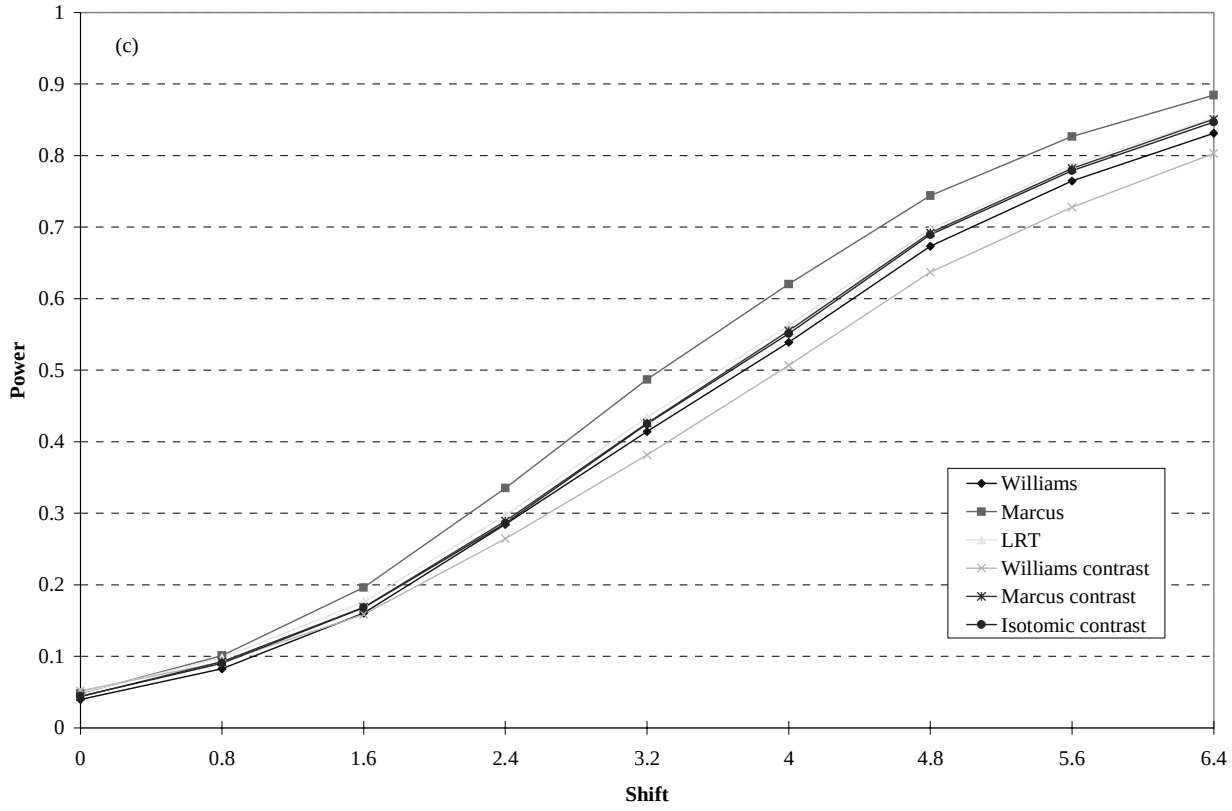
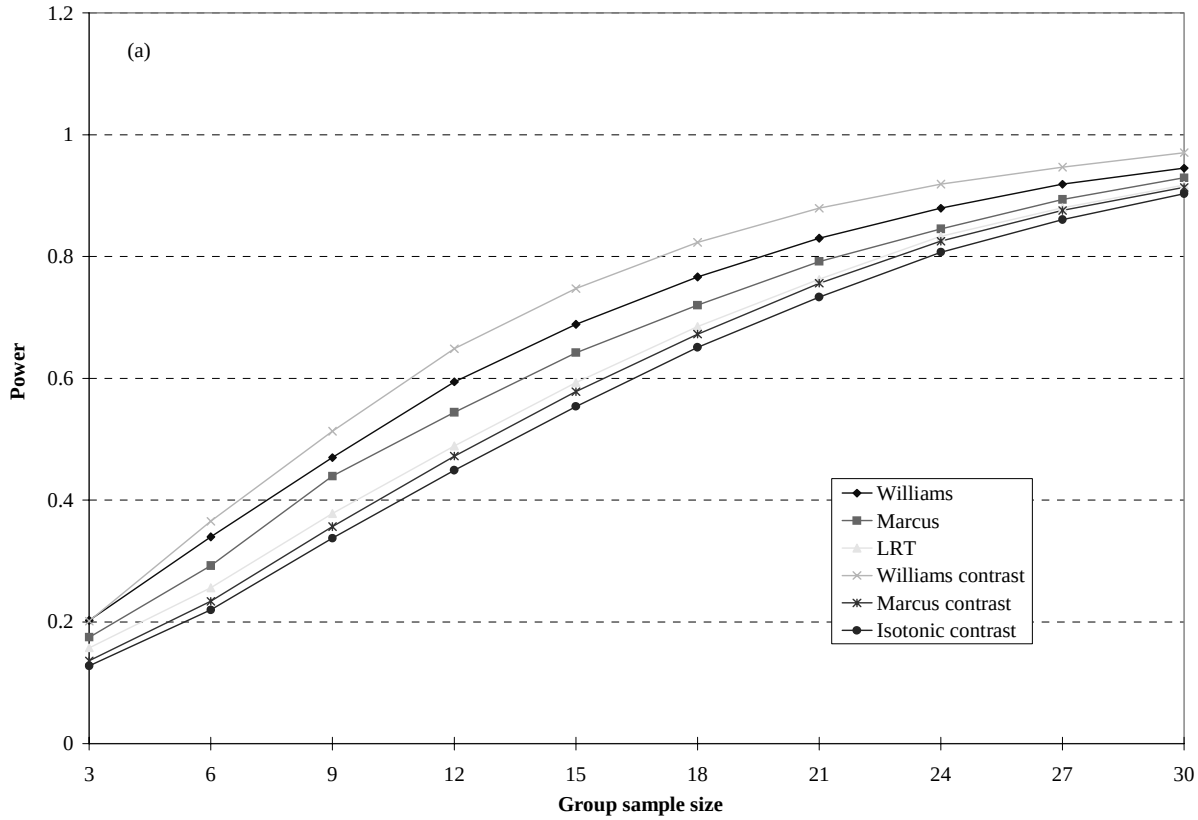


Figure 4.7. *(continued)*

For the sample size allocation $(14, 10, 10, 6)$ Figure 4.7. plots the power values in dependence of the shift parameter $\theta = \mu_k - \mu_0$, assuming the underlying contaminated normal distribution described above. Similar results as obtained before in the normal case hold again. Williams' MCT performs best for concave (a) through linear (b) profiles. In the convex case Marcus' $\bar{t}^{\,\mathrm{mod}}$ behaves well. Other types of non-normal distributions have been investigated, too. But the performance of the power curves in relation to each other did not change dramatically and further results are not reported here. As seen before, Williams' $\bar{t}$ behaves best when considering the size performances, but on the other side its power values lie well below the other curves. In the convex case (c) only Williams' MCT performs worse.



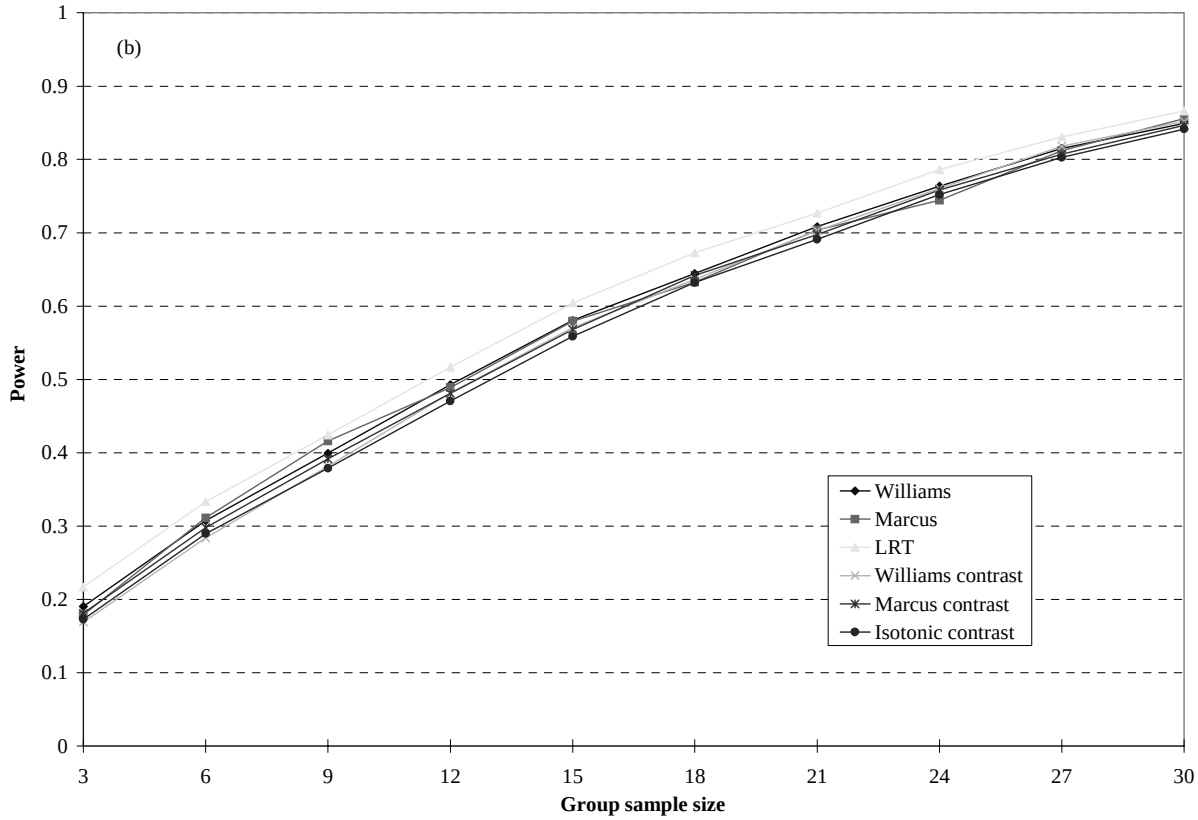


Figure 4.8. Power comparison, assumed variances $(1, 1.2, 1.4, 1.6, 1.8, 2)$, balanced case, $\alpha = 0.05$, $k = 5$ for (a) concave profile $\mu = (0, 1, 1, 1, 1, 1)$, (b) linear profile $\mu = (0, 0.2, 0.4, 0.6, 0.8, 1)$ and (c) convex profile $\mu = (0, 0, 0, 0, 0, 1)$.

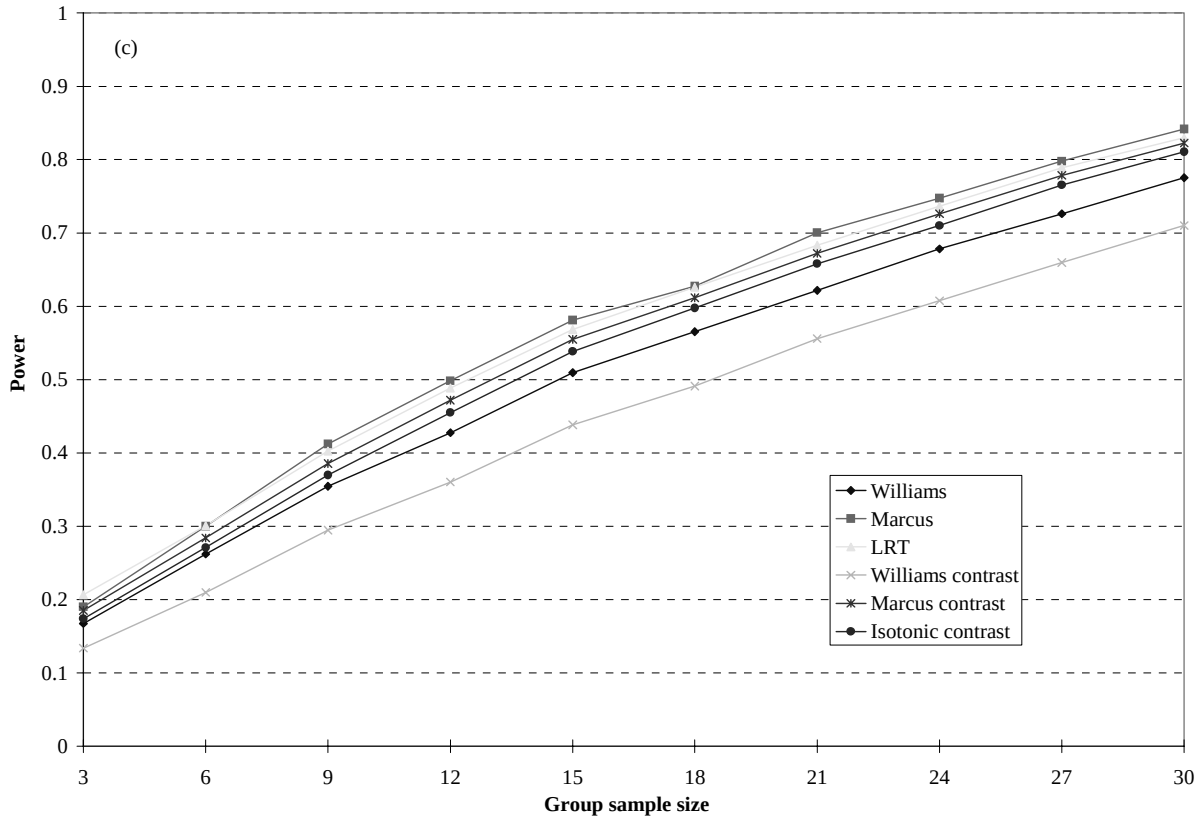


Figure 4.8. *(continued)*

Variance heterogeneity is common in practice. We therefore investigated this special violation of the assumption for the tests considered so far in Figure 4.8. We analysed the frequent case of increasing variance with respect to increasing doses or treatments. For $n_i = 12$ or 15 the approach of Williams beats the LRT by approximately 10% in the concave case. In the convex case, however, Williams $\bar{t}$ loses only up to 6 or 7%. This is an evidence for the statements given in the Introduction, which say that Williams' test is considered to be more robust against variance heterogeneity than the LRT. Unfortunately, Williams' MCT does not share this virtue. In the concave case it outperforms by far the LRT (over 15% power differences), but lies significantly below the LRT in the convex case.

Figure 4.9. represents a different view of comparing the power of the six tests. It demonstrates more clearly for k = 2 the behaviour of the tests in the balanced case for different underlying dose-response shapes. The power values are given with respect to β, which is the angle μ makes with the ray on which $x_0 = x_1 < x_2$. The restriction $\mu \in H_A$ is equivalent to $0° \le \beta \le 60°$ (e.g. Robertson et al., 1988, p. 91). For example, $\beta = 0°$ represents the case $\mu_0 = \mu_1 < \mu_2$, $\beta = 30°$ denotes $\mu_1 - \mu_0 = \mu_2 - \mu_1 > 0$ and $\beta = 60°$ stands for the concave profile $\mu_0 < \mu_1 = \mu_2$. For Figure 4.9. $\Delta = \sqrt{\sum_{i=0}^{k} n_i (\mu_i - \bar{\mu})^2} = 2$ has been chosen, which is a measure for the distance from μ to H_0. The pair (β, Δ) can be thought of more formally as the equivalent polar coordinates of a point in H_A. Marcus' original test behaves very good, independently from the investigated sample size in either (a) or (b). Williams' $\bar{t}$ -test, and even more distinct the MCT version, perform good for concave profiles ($\beta = 40°$ through 60°), but behave worse for underlying convex shapes (i.e. $\beta = 0°$ through 20°). The LRT and Marcus' MCT are again very similar to each other. Their main characteristics is their relative shape independence, as their power curves are more flat than the other tests, while maintaining a high average power. Note that for k = 2 the definitions of the isotonic and Marcus' MCT are identical in the balanced case. Finally, one could assess empirically a symmetric power performance for both distinguished MCTs similar to the LRT. This behaviour leads to a meaningful comparison of maximum and minimum power values in Table 4.1.

As a final comparison Table 4.1. presents power values for selected points in the alternative space. According to Robertson et al. (1988, p. 94), the minimum power of the LRT over H_A for a fixed $\Delta > 0$ is believed to occur at $\mu_0 < \mu_1 = \ldots = \mu_k$ and $\mu_0 = \mu_1 = \ldots = \mu_{k-1} < \mu_k$. (note the symmetry of the LRT around $\beta = 30°$ in Figure 4.9.). On the other side, the maximum power is believed to occur at $\mu_1 - \mu_0 = \mu_2 - \mu_1 = \ldots = \mu_k - \mu_{k-1} > 0$. These conjectures has been proven for k = 2, but no general methodology seems to be available for arbitrary k. Table 4.1. compares the power at these selected points for different values of Δ, whereas for the minimum case only the convex shape is considered. The power of the MCTs were calculated with an accuracy of $\varepsilon = 1 \times 10^{-4}$. The values of the other tests, however, were obtained by simulation (10,000 runs), resulting in an accuracy of about 1×10^{-2}. Both the isotonic and the Marcus MCT are very similar, differing at most from the third digit on. This is indeed consistent with our conclusions from the previous Figures. Further, the LRT is better than both MCTs for all points considered. But the differences get never larger than 2%.

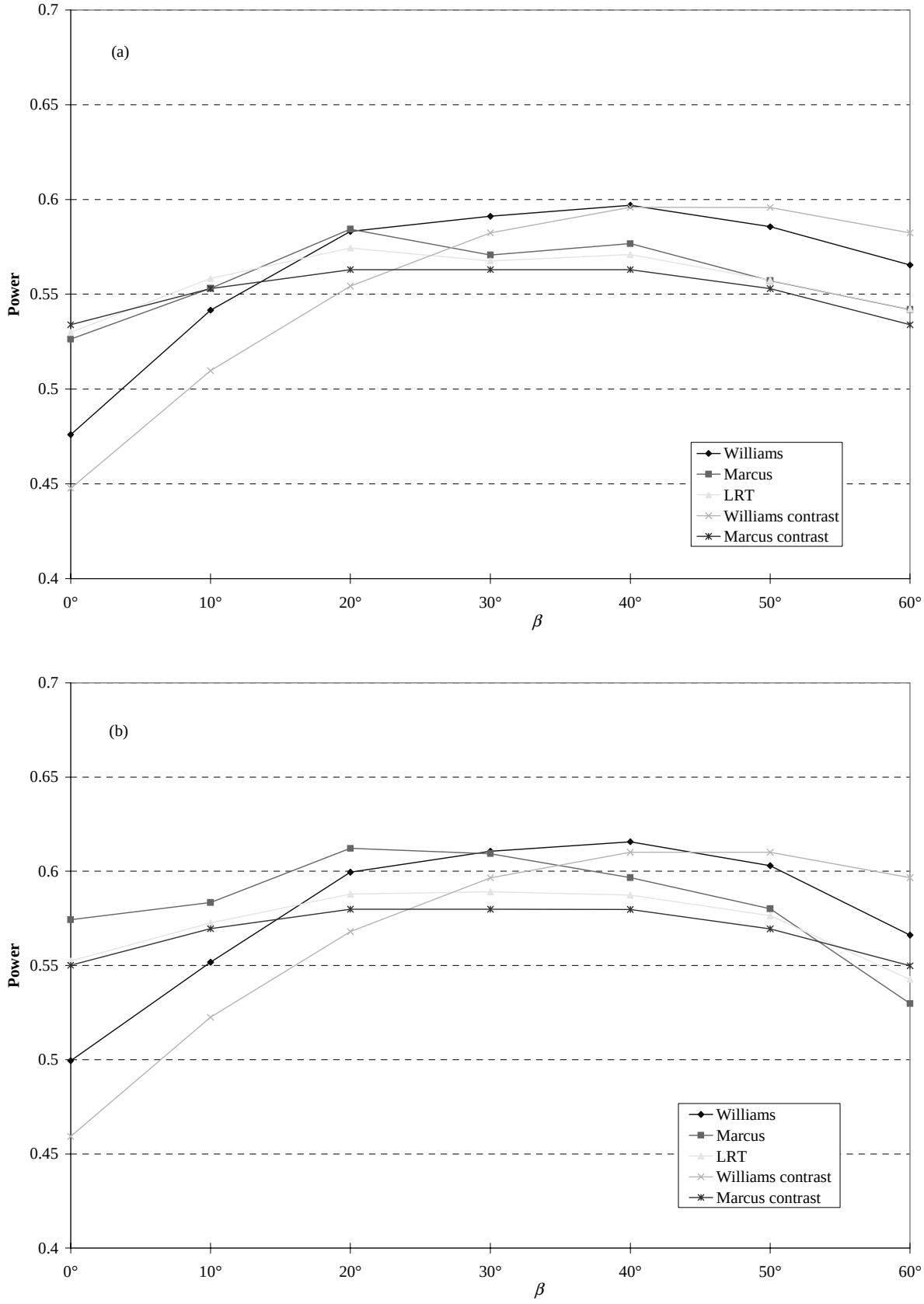


Figure 4.9. Power comparison, balanced case, $\alpha = 0.05$, $k = 2$, $\Delta = 2$ for (a) $n_i = 8$ and (b) $n_i = 14$.

Furthermore, the differences between the maximum and minimum values are always smaller than 8% for a given Δ. These 8% are therefore an upper bound of the maximum power loss due to unfavourable shapes and strengthens our conclusion that the power of the LRT (and hence both the isotonic and Marcus' MCT) is fairly stable over H_A, at least in the balanced case.

	$\nu = 20$ ($n_i = 6$)		$\nu = 40$ ($n_i = 11$)	
	(minimum)	*(maximum)*	*(minimum)*	*(maximum)*
		$\Delta = 1$		
Williams	0.1774	0.2308	0.1839	0.237
Marcus	0.1994	0.23	0.2156	0.2474
LRT	0.1921	0.2287	0.197	0.2343
Williams contrast	0.1535	0.2203	0.1571	0.2266
Marcus contrast	0.188	0.2191	0.1947	0.2274
Isotonic contrast	0.1881	0.2197	0.195	0.2281
		$\Delta = 2$		
Williams	0.4366	0.5647	0.454	0.5875
Marcus	0.4951	0.5573	0.5142	0.5763
LRT	0.4912	0.5651	0.5214	0.5852
Williams contrast	0.3826	0.547	0.395	0.5641
Marcus contrast	0.4888	0.5432	0.5098	0.5661
Isotonic contrast	0.4886	0.5456	0.5095	0.5685
		$\Delta = 3$		
Williams	0.7381	0.8647	0.7507	0.8726
Marcus	0.8153	0.8591	0.8306	0.8732
LRT	0.8157	0.8658	0.8272	0.8725
Williams contrast	0.6785	0.8451	0.6965	0.8605
Marcus contrast	0.8077	0.8418	0.8291	0.8624
Isotonic contrast	0.8071	0.8442	0.8288	0.8645

Table 4.1. Comparison of power values at selected points in H_A (balanced case, $k = 3$).

4.3. Conclusions

From the computations done we emphasise again the well known fact that there is no uniformly most powerful test among the class of trend tests. No test could be recommended in the sense of achieving the best power over the whole alternative. If prior information regarding the dose-response shape is available, powerful single contrast tests can be defined and conducted (see also Section 3.1., pp. 81, for a short discussion). In the general case, however, the various trend tests have to be compared with regard to a certain meaningful power measure. One of those few adequate methodologies would be to compute both the maximum and minimum over the alternative space. Unfortunately, corresponding expressions are not available yet for MCTs and the LRT for $k > 2$ – not to mention Williams' and Marcus' approaches. Table 4.1. was a first attempt in this direction.

In order to give the practitioner in spite of that some recommendations and conclusions from the conducted power study we refer to the more empirical 'average power'. As no rigid comparison methods are available, we try to find those tests, which are least dependent on shape and sample size allocation, hence maintaining a stable power at an acceptable level throughout H_A. As already noted elsewhere in the literature, too, the LRT of Bartolomew fulfils these conditions better than the competing methods. But, as we could see from the preceding Figures, both Marcus' and the isotonic MCT perform very similar when compared to the LRT. In many situations they even lead to practically identical power values. They approximate the LRT so well, that the resulting losses do not exceed 5% throughout the whole alternative space. Williams' original $\bar{t}$–test and the corresponding MCT perform especially well for concave profiles. But both get worse for convex profiles. Marcus' $\bar{t}^{\,\mathrm{mod}}$–test seems to compete well with the LRT in balanced cases (especially for $k = 2$, as seen from Figure 4.9.), but in unbalanced situations it may perform poor with up to 40% power difference in some cases. Similar to Williams' $\bar{t}$, it does not incorporate the sample size allocation appropriately and can therefore not be recommended for general imbalances (see also the discussion in Example 1.4.).

For non-normal data and variance heterogeneity the LRT is somewhat more liberal than the approaches of Marcus and Williams. The size behaviour of MCTs seems to depend also on the present contrast definition. We could conclude in terms of power that all tests are more or less robust against violations of the assumptions (including non-monotonicity of the means). However, further comparisons are required to establish the robustness limits for the particular tests.

5. Power comparison for binomial data

Variables with only two outcomes, i.e. „finding/no finding“ or „effect/no effect“ occur frequently in both biological and non-biological studies. As an example, we will discuss later in more detail the influence of germinal temperature on the occurrence of anomalies in young kohlrabi plants (Habegger and Wiebe, 1985). In this experiment the number of responders, i.e. those plants showing an abnormality, is determined for each temperature treatment. The experimental question is whether temperature has an significant influence, provided that temperature decrease leads, if at all, to an increase of anomaly among kohlrabi plants (see Subsection 5.3.4. for the data).

Mead et al. (1993, p. 288) report an experiment in order to determine the presence or absence of *Agrostis tenuis* upon the effect of burned heathland. Denoting treatments A through D in accordance with the time past since the last burning (4, 3, 2 and 1 year(s), respectively), the following numbers of quadrants (from 100 samples each) containing *Agrostis tenuis* were obtained: 44, 36, 40 and 28 for A through D. Does the frequency of occurrence of *Agrostis tenuis* vary monotonously from area to area?

Furthermore, continuos endpoints are sometimes dichotomised at an a-priori determined cut-point into responder and non-responder. This approach overcomes difficulties with violation of assumptions for the distribution or variance. Moreover, healing rates are easier to understand by physicians than p-values. Suissa and Blais (1995, p. 247), for example, conclude that "... *frequently focuses on the proportion of subjects who fall below or above a clinically relevant cut-off value ... The customary approach to analyse such data is to dichotomise the continuous outcome measure and use statistical techniques based on binary data* ...". See Sankey and Weissfeld (1998) for a recent discussion on this subject.

In Section 5.1. we introduce some new notations and concepts for this chapter. Moreover, a short survey of the most frequently used binomial trend tests is given. Multiple contrast tests are introduced in the present framework. Next we derive in Section 5.2. asymptotic power formulas in closed form for contrast tests. We discuss further the important aspect of sample size determination and provide a closed form expression in the single contrast case. In Section

5.3. a series of generalisations is analysed regarding the performance under the alternative (continuity correction, pooled versus unpooled versions, exact conditional and unconditional tests). Brief power comparisons are conducted in the corresponding subsections. Finally, the above mentioned example of anomalies among kohlrabi is analysed in Subsection 5.3.4., with special focus on the effects of the suggestions given in the subsections before.

5.1. Notations

When comparing several independent binomial properties, the results are usually summarised as shown in Table 5.1. Hence, r_i is the observed value of the binomial random variable $R_i \sim \text{Bin}(n_i, \pi_i)$, $i = 0, 1, \ldots, k$, and π_i is the probability of response at dose level d_i.

Dose level	d_0	d_1	...	d_k	Σ
Number of responders	r_0	r_1	...	r_k	R
Number at risk	n_0	n_1	...	n_k	N

Table 5.1. Summarising results of an experiment involving the comparison of proportions.

The hypotheses of interest are formulated equivalently to (1.1) and (1.2). The goal is to test

$$H_0: \pi_0 = \pi_1 = \ldots = \pi_k \tag{5.1}$$

against the one-sided alternative

$$H_A: \pi_0 \leq \pi_1 \leq \ldots \leq \pi_k, \pi_0 < \pi_k. \tag{5.2}$$

The trend test most frequently used in the literature is that according to Cochran (1954) and Armitage (1955), termed CA hereafter. Furthermore, it is also recommended by several national and international guidelines for detecting a (linear) trend in proportions (e.g. National

Archives and Record Service, 1985). Assuming the widely employed linear logistic model $\log\left(\frac{\pi_i}{1-\pi_i}\right) = a + bd_i$, Tarone and Gart (1980) succeeded in demonstrating the equivalence of the CA to the score test for testing H_0': $b = 0$ against H_A': $b > 0$ derived from the partial derivatives of the log-likelihood function (so-called $C(\alpha)$–test). Furthermore, Tarone and Gart (1980) have shown that the CA is uniformly optimal under the logistic law and it is asymptotically efficient for all monotone alternatives. The test statistic is defined as

$$T^{CA} = \sum_{i=0}^{k} r_i d_i \,. \tag{5.3}$$

Under H_0 one obtains the conditional mean of T^{CA} for given R as $E\left(T^{CA} \middle| R\right) = \pi \sum_i n_i d_i$, where the common success rate π is estimated by $p = R/N$. Setting $T^{CA'} = T^{CA} - E\left(T^{CA} \middle| R\right) = \sum_i r_i (d_i - d)$, $d = \sum_i n_i d_i / N$, an approximate test based on the asymptotic standard normal distribution is

$$T_a^{CA} = \frac{T^{CA'}}{\sqrt{V\left(T^{CA'} \middle| R\right)}},$$

where $V\left(T^{CA'} \middle| R\right) = \pi(1-\pi)\left[\sum n_i (d_i - d)^2\right]$. This form is due to Nam (1987). The advantage is its easy and intuitive derivation. However, in practice (see, for example, Portier and Hoel, 1984) usually the form given next is employed. In the following we will work mainly with this representation.

Lemma 5.1.: With the notation given above T_a^{CA} can also be written as

$$T_a^{CA} = \frac{\sum_i \left(r_i - \frac{n_i}{N} R\right) d_i}{\sqrt{\sum_i \frac{n_i d_i^2}{N} - \left(\sum_i \frac{n_i d_i}{N}\right)^2}} \sqrt{\frac{N}{R(N-R)}} \,. \tag{5.4}$$

Proof: The following algebraic transformations yield immediately the assertion:

$$T_a^{CA} = \frac{\sum_i r_i \left(d_i - \frac{\sum_j n_j d_j}{N} \right)}{\sqrt{\frac{R(N-R)}{N^2} \sum_i n_i \left(d_i - \frac{\sum_j n_j d_j}{N} \right)^2}} = \frac{\sum_i r_i d_i - R \sum_j \frac{n_j d_j}{N}}{\sqrt{\sum_i n_i \left(d_i - \frac{\sum_j n_j d_j}{N} \right)^2}} \sqrt{\frac{N^2}{R(N-R)}} =$$

$$= \frac{\sum_i r_i d_i - \sum_i R \frac{n_i d_i}{N}}{\sqrt{\frac{N}{N} \left\{ \sum_i n_i d_i^2 + \underbrace{\sum_i n_i}_{=N} \left(\sum_j \frac{n_j d_j}{N} \right)^2 - 2 \sum_i n_i d_i \left(\sum_j \frac{n_j d_j}{N} \right) \right\}}} \sqrt{\frac{N^2}{R(N-R)}} =$$

$$= \frac{\left(\sum_i r_i d_i - \sum_i R \frac{n_i d_i}{N} \right) \sqrt{\frac{N}{R(N-R)}}}{\sqrt{\sum_i \frac{n_i d_i^2}{N} + \left(\sum_j \frac{n_j d_j}{N} \right)^2 - 2 \sum_i \frac{n_i d_i}{N} \left(\sum_j \frac{n_j d_j}{N} \right)}} = \frac{\left(\sum_i r_i d_i - \sum_i R \frac{n_i d_i}{N} \right) \sqrt{\frac{N}{R(N-R)}}}{\sqrt{\sum_i \frac{n_i d_i^2}{N} - \left(\sum_j \frac{n_j d_j}{N} \right)^2}}.$$

The CA, however, bears several disadvantages. The dose scores d_i $(d_0 \leq \ldots \leq d_k)$ are fixed in advance. One critical point is their choice, because they influence the power markedly. Armitage (1955) and Graubard and Korn (1987) recommended equally spaced scores if no information on the shape of the dose-response is known a-priori. Williams (1988, p. 424) proposed a modification using the maximum of extreme score functions, which is relatively robust against varying dose-response shapes. Moreover, the CA-test is a test on slope in a linearised regression model. The linearity assumption, however, is highly unrealistic in practice. Hoel and Portier (1994), for example, analysed the data of 315 chemicals obtained from the National Toxicology Program and stated that "*tumour site data were more often consistent with a quadratic than a linear response* ...". In Chapter 4 we have already observed the impact of varying dose response shapes on the power of a specific test. Shapes may differ from convex through concave, and even down-turns at high doses occur sometimes. The behaviour of the CA-test, when slightly deviating from its linearity assumption, is sketched in the following artificial example. Let k = 3, n_i = n = 150 and the incidences be (0.01, **0.02**, 0.03, 0.04). Conducting the CA-test (5.4) with equidistant scores leads to a p-value of 0.0393. A small change in this perfectly linear relationship by increasing the second rate to 0.03, i.e. considering the non-linear shape (0.01, **0.03**, 0.03, 0.04), results in p = 0.0753. Therefore, the null hypothesis could not be rejected any more. Increasing the effect resulted in even worse p-values.

In order to overcome the problems of shape dependence at least partially, different approaches have been proposed. Based on Bartholomew (1959), a likelihood ratio test has been introduced by Barlow et al. (1972) and Oluyede (1994). Collings et al. (1981, p. 785-791) extensively compared the CA to the isotonic test for balanced set-ups and concluded that "*the isotonic test holds little advantage over the regression test.*" Agresti and Coull (1998) compared in this context the power differences between general order restricted inference and the use of linear logit models for binary responses. They got mixed results, but all in all their data seem to suggest an improved performance of the ordinary trend tests. Recently, Hirji and Tang (1998) screened several trend tests for binary data. They compared the size, power and the shape of the power function of nine two-sided exact, mid-p and asymptotic trend tests. We will refer several times to this study in the course of this chapter. For an overview of robust trend tests for binomial data, as they are required to incorporate historical control data, we refer to Smythe et al. (1987) and the references therein.

In the framework of this thesis we apply the concept of MCTs developed in the preceding chapters on the present situation. Dichotomous contrasts have already been investigated since Knoke (1976). But most of these articles deal only with the Dunnett contrast, see Koch (1996) for a recent overview of dichotomous unrestricted many-to-one comparisons. Robertson et al. (1988) mentioned general contrast tests for dichotomous data, but they referred only to the results for normal variates. Neuhäuser and Hothorn (1997, 1998) published several possibilities of defining such contrast tests. The aim of this chapter is therefore to investigate further this class of tests under a variety of aspects.

With the notation of Table 5.1. we denote the individual outcome of such an experiment by x_{ij}, $i = 0, 1, \ldots, k$, $j = 1, \ldots, n_i$. The $x_{ij} \sim \text{Bin}(1, \pi_i)$ are Bernoulli distributed with expectation $E(x_{ij}) = \pi_i$ and variance $V(x_{ij}) = \pi_i(1-\pi_i)$. Therefore, $r_i = \sum_{j=1}^{n_i} x_{ij}$. Define further the empirical group estimates $p_i = \sum_j x_{ij} / n_i = r_i / n_i$ of π_i. To test now the null hypothesis (5.1) against the alternative (5.2) consider the (single) contrast test

$$T^{SC} = \sum_{i=0}^{k} c_i p_i = \sum_{i=0}^{k} \frac{c_i}{n_i} r_i \qquad (5.5)$$

with the contrast ensuring equation $\sum_i c_i = 0$. Consequently, the expectation of (5.5) under H_0 is $E\left(T^{SC}\right) = \sum_i \frac{c_i}{n_i} E\left(r_i\right) = \pi \sum_i c_i = 0$. The variance under H_0 is $V\left(T^{SC}\right) = \pi(1-\pi)\sum_i c_i^2 / n_i$. According to Neuhäuser (1996) the asymptotic test statistics

$$T_a^{SC} = \frac{T^{SC} - E\left(T^{SC}\right)}{\sqrt{V\left(T^{SC}\right)}} = \frac{\sum_i \frac{c_i}{n_i} r_i}{\sqrt{p(1-p)\sum_i c_i^2 / n_i}} \tag{5.6}$$

is standard normal distributed for large N under H_0. Hence, in the same way as in the case of normal variates, we define the MCTs

$$T^{MC} = \max\left\{T_1^{SC}, \ldots, T_q^{SC}\right\} \tag{5.7}$$

respectively

$$T_a^{MC} = \max\left\{T_{1,a}^{SC}, \ldots, T_{q,a}^{SC}\right\}. \tag{5.8}$$

Again, the MCTs are supposed to lead to test statistics which depend less on the underlying dose-response shapes. As a heuristic reason one can imagine that several kinds of dose-response pattern are mapped onto appropriate single contrast vectors. The maximum over them takes therefore the best single test statistic. The following theorem establishes the asymptotic behaviour of T_a^{MC} under the null hypothesis. It says that the joint distribution of $\left\{T_{1,a}^{SC}, \ldots, T_{q,a}^{SC}\right\}$ follows a q–variate normal distribution with expectation vector $\boldsymbol{0}$ and a correlation matrix $\boldsymbol{R}$ to be established in a closed form expression in the lemma following afterwards.

Theorem 5.1.: Let $\left\{T_{1,a}^{SC}, \ldots, T_{q,a}^{SC}\right\}$ be given and let $\boldsymbol{C}$ denote the corresponding contrast matrix. Then the joint distribution of the single contrasts can be expressed asymptotically through the q–variate normal distribution $N_q\left(\boldsymbol{C}\pi, \boldsymbol{C}\Sigma\boldsymbol{C}^t\right)$, where $\Sigma = \mathrm{diag}\left(\frac{\pi_i(1-\pi_i)}{n_i}\right)$.

Proof: According to Baringhaus (1998) we assume without loss of generalisation $n_i = n$. We have first to show that the vector

$$\boldsymbol{p} = \begin{pmatrix} p_0 \\ \vdots \\ p_k \end{pmatrix} = \frac{1}{n} \begin{pmatrix} \sum_j x_{0j} \\ \vdots \\ \sum_j x_{kj} \end{pmatrix} = \frac{1}{n} \sum_j \begin{pmatrix} x_{0j} \\ \vdots \\ x_{kj} \end{pmatrix}$$

is asymptotically mvn distributed. With respect to Lemma 2.5. it is sufficient to show that $\boldsymbol{t}^t \boldsymbol{p} = \sum_{i=0}^{k} t_i p_i$ is asymptotically univariate normal distributed $\forall\, \boldsymbol{t} = (t_0, \ldots, t_k)^t \in \mathbb{R}^{k+1}$. One calculates

$$\boldsymbol{t}^t \boldsymbol{p} = \sum_{i=0}^{k} t_i p_i = \sum_{i=1}^{k} t_i \sum_{j=1}^{n} x_{ij} \Big/ n = \sum_{j=1}^{n} \sum_{i=1}^{k} t_i x_{ij} \Big/ n = \sum_{j=1}^{n} y_j ,$$

where $y_j = \sum_{i=1}^{k} t_i x_{ij} \Big/ n$. Consequently, the expectation and the variance of y_j are given by

$$E[y_j] = E\left[\sum_{i=1}^{k} t_i x_{ij} \Big/ n\right] = \sum_i t_i E[x_{ij}] \Big/ n = \sum_i t_i \pi_i \Big/ n =: a_j / n =: a/n \quad \forall j \text{, and}$$

$$V[y_j] = V\left[\sum_{i=1}^{k} t_i x_{ij} \Big/ n\right] = \sum_i t_i^2 V[x_{ij}] \Big/ n^2 = \sum_i t_i^2 \pi_i (1-\pi_i) \Big/ n^2 =: \sigma_j^2 =: \sigma^2 < +\infty \quad \forall j .$$

Now, the y_j are i.i.d. with finite variance and the central limit theorem yields

$$\frac{1}{\sigma\sqrt{n}} \left(\sum_{j=1}^{n} y_j - na/n \right) \xrightarrow{d} N(0, 1).$$

Therefore, $\boldsymbol{t}^t \boldsymbol{p}$ is asymptotically univariate $N(a, n\sigma^2) = N\left(\boldsymbol{t}^t \pi, \sum_i \pi_i (1-\pi_i) t_i^2 \Big/ n\right)$ distributed and we get with the above mentioned Lemma 2.5. the intermediate result

$$\boldsymbol{p} \sim N_{k+1}\left(\pi,\ \frac{1}{n}\begin{pmatrix}\pi_0(1-\pi_0) & & 0\\ & \ddots & \\ 0 & & \pi_k(1-\pi_k)\end{pmatrix}\right) =: N_{k+1}(\pi,\ \Sigma).$$

The assertion of the Theorem is proven in a final step based on Lemma 2.1.:

$$\boldsymbol{Cp} \sim N_q\left(\boldsymbol{C}\pi,\ \frac{\boldsymbol{C}}{n}\begin{pmatrix}\pi_0(1-\pi_0) & & 0\\ & \ddots & \\ 0 & & \pi_k(1-\pi_k)\end{pmatrix}\boldsymbol{C}^t\right) = N_q\left(\boldsymbol{C}\pi,\ \boldsymbol{C}\Sigma\boldsymbol{C}^t\right).$$

Lemma 5.2.: The correlation under H_0 between two single contrasts $T_{i,a}^{SC}$ and $T_{j,a}^{SC}$ from Theorem 5.1. with the contrast vectors $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ $(1 \le i,\ j \le q)$ is given by

$$\rho_{ij} = \frac{\sum_l c_{il}c_{jl}/n_l}{\sqrt{\left(\sum_l c_{il}^2/n_l\right)\left(\sum_l c_{jl}^2/n_l\right)}}. \qquad (5.9)$$

Proof: The assertion follows directly from the last line of the proof to Theorem 5.1. and remembering that under H_0 $\pi_0 = \pi_1 = \ldots = \pi_k$. As we use the pooled estimator $p = R/N$ the assertion follows.

Remark 5.1.: Note, that at present we are working with the pooled estimate *p* only. The asymptotic test statistics defined in (5.6) and (5.8) were introduced in the way they usually appear in the literature. Hence, we call those asymptotic contrast tests, which bear the pooled estimate *p* in the variance term, *pooled contrast tests*. In Subsection 5.3.2. we shall introduce a new class of dichotomous MCTs by considering the p_i's instead, consequently named *unpooled contrast tests*. In such cases one should use another representation of the correlation matrix $\boldsymbol{R}$ instead of (5.9). We refer to the corresponding subsection for further details.

5.2. Power expression of multiple contrast tests

Similarly to Subsection 4.1. we derive a power expression in closed form for asymptotic MCTs. We divide the representation in single and multiple contrasts because of an easier introduction of the results. We first establish the power formula in the case of SCTs for given type–I–error α, sample size and dose-response shape. A similar expression for the CA has been derived by Nam (1987).

Theorem 5.2.: Let T_a^{SC} be a binomial asymptotic SCT. Denote by $\boldsymbol{c}_1$ the $1\times(k+1)$ contrast vector. For $\pi = \sum_i n_i \pi_i / N$ the asymptotic power of T_a^{SC} is then given by

$$1-\beta = 1-\Phi\left(\frac{z_{1-\alpha}\sqrt{\pi(1-\pi)\sum_i c_i^2/n_i} - \sum_i c_i\pi_i}{\sqrt{\sum_i \pi_i(1-\pi_i)c_i^2/n_i}}\right), \tag{5.10}$$

where $z_{1-\alpha}$ is the $(1-\alpha)$–quantile of the standard normal distribution Φ.

Proof: Considering the behaviour of T_a^{SC} under H_A,

$$T_a^{SC} \sim N\left(E_A\left(T_a^{SC}\right), V_A\left(T_a^{SC}\right)\right) = N\left(\frac{E_A\left(T^{SC}\right)}{\sqrt{V_0\left(T^{SC}\right)}}, \frac{V_A\left(T^{SC}\right)}{V_0\left(T^{SC}\right)}\right)$$

follows, where the indices denote whether the expectation or variance is obtained under H_A or H_0. The expectation and variance of T^{SC} under the alternative are given by $E_A\left(T^{SC}\right) = \sum_i c_i\pi_i$ and $V_A\left(T^{SC}\right) = \sum_i c_i^2 \frac{\pi_i(1-\pi_i)}{n_i}$. The power is defined as the probability of correctly rejecting the null hypothesis. Then the following transformations are valid:

$$\text{power} = 1-\beta = P\left(T_a^{SC} \geq z_{1-\alpha} \middle| H_A\right)$$

$$\Leftrightarrow \qquad 1-\beta = P\left(\frac{T_a^{SC} - E_A\left(T_a^{SC}\right)}{\sqrt{V_A\left(T_a^{SC}\right)}} \geq \frac{z_{1-\alpha} - E_A\left(T_a^{SC}\right)}{\sqrt{V_A\left(T_a^{SC}\right)}}\right)$$

$$\Leftrightarrow \qquad 1-\beta = 1-\Phi\left(\frac{z_{1-\alpha}-E_A\left(T_a^{SC}\right)}{\sqrt{V_A\left(T_a^{SC}\right)}}\right)$$

$$\Leftrightarrow \qquad 1-\beta = 1-\Phi\left(\frac{z_{1-\alpha}\sqrt{\pi(1-\pi)\sum_i c_i^2/n_i}-\sum_i c_i\pi_i}{\sqrt{\sum_i \pi_i(1-\pi_i)c_i^2/n_i}}\right).$$

Lemma 5.3.: Denote by $\lfloor x \rfloor$ the greatest integer smaller than x. For $n_i \in \mathbb{N}$ and with the same notation of Theorem 5.2. the following two assertions hold in order to determine the required sample size when conducting a single contrast test.

a) For $t_i = n_0/n_i$, $i = 0, \ldots, k$, fixed a-priori,

$$n_0 = \left\lfloor \frac{\left(z_{1-\alpha}\sqrt{\pi(1-\pi)\sum_i c_i^2 t_i} + z_{1-\beta}\sqrt{\sum_i \pi_i(1-\pi_i)c_i^2 t_i}\right)^2}{\left(\sum_i c_i\pi_i\right)^2} \right\rfloor + 1. \qquad (5.11)$$

b) In the balanced case $n_i = n \quad \forall i$,

$$n_0 = \left\lfloor \frac{\left(z_{1-\alpha}\sqrt{\pi(1-\pi)\sum_i c_i^2} + z_{1-\beta}\sqrt{\sum_i \pi_i(1-\pi_i)c_i^2}\right)^2}{\left(\sum_i c_i\pi_i\right)^2} \right\rfloor + 1. \qquad (5.12)$$

Proof: Applying the inverse function Φ^{-1} on equation (5.10) and subsequent rearranging yield $\sum_i c_i\pi_i = z_{1-\alpha}\sqrt{\pi(1-\pi)\sum_i c_i^2/n_i} + z_{1-\beta}\sqrt{\sum_i \pi_i(1-\pi_i)c_i^2/n_i}$. Substituting $t_i = n_0/n_i$ assertion a) follows. Part b) follows from a) for $t_i = 1 \quad \forall i$.

The importance of the last lemma lies in the fact that starting from (5.10) we were able to derive a closed form expression for the determination of a reasonable asymptotic sample size when conducting SCTs. Equation (5.11) yields the minimum sample size of the control group required to detect a significant relevant monotonous dose-response relationship for given α, β and relationships t_i. An approximate sample size of the i^{th} group is immediately found by $n_i = n_0/t_i$, where the relationship t_i is determined in advance. This constellation might be relevant, for example, when one selects according to the well-known recommendation

$n_0 = n_i\sqrt{k}$. In this case, the n_i are chosen to be equal for i = 1, ..., k, and n_0 to be higher, corresponding proportionally to the number of treatment levels. This optimality rule for sample size allocation in the many-to-one design was published in the normal set-up by Dunnett (1955). No such investigation for dichotomous data, however, was found. In the balanced case the sample size determination (5.12) follows from the preceding expression by setting $t_i = 1$.

Note that in the univariate case, i.e. considering SCTs or the CA, the resulting power and sample size formulas are easy to evaluate as only the univariate standard normal distribution is required beside some simple arithmetic operations. Such calculations can therefore be conducted by use of any statistical software package. Things change, when focusing now on MCTs. This is due to their multivariate nature. By virtue of Theorem 5.1. we trace the calculations back to the mvn distribution discussed in extensive in Chapter 2. First we provide an analogous result to Theorem 5.2. for MCTs.

Theorem 5.3.: Let $T_a^{MC} = \max\left\{T_{1,a}^{SC}, \ldots, T_{q,a}^{SC}\right\}$ be a binomial asymptotic MCT. Denote by $\boldsymbol{C} = \left(c_{ij}\right)_{ij} = \left(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_q\right)^t$ the $q \times (k+1)$ contrast matrix. Let further $\boldsymbol{e} = \left(E_A\left(T_{1,a}^{SC}\right), \ldots, E_A\left(T_{q,a}^{SC}\right)\right)$ and $\boldsymbol{v} = \left(v_1, \ldots, v_q\right) = \left(\sqrt{V_A\left(T_{1,a}^{SC}\right)}, \ldots, \sqrt{V_A\left(T_{q,a}^{SC}\right)}\right)$ be the vectorially summarised expectations and variances of the SCTs as given above. Then the power of T_a^{MC} is given by

$$1-\beta = 1-\boldsymbol{\Phi}_q\left(\left(\boldsymbol{z}_{q,1-\alpha} - \boldsymbol{e}\right)\mathrm{diag}\left(\tfrac{1}{v_1}, \ldots, \tfrac{1}{v_q}\right); \boldsymbol{0}, \boldsymbol{R}\right),$$

where $\boldsymbol{z}_{q,1-\alpha} = \left(z_{q,1-\alpha}, \ldots, z_{q,1-\alpha}\right)$ stands for the q-variate normal $(1-\alpha)-$equipercentage point and the elements of $\boldsymbol{R}$ are given by

$$\rho_{lm} = \frac{\sum_i c_{li} c_{mi} \frac{\pi_i(1-\pi_i)}{n_i}}{\sqrt{\left(\sum_i c_{li}^2 \frac{\pi_i(1-\pi_i)}{n_i}\right)\left(\sum_i c_{mi}^2 \frac{\pi_i(1-\pi_i)}{n_i}\right)}}, \qquad 1 \le l, m \le q. \tag{5.13}$$

Proof: From Theorem 5.1. we know that the joint distribution of $\left\{T_{1,a}^{SC}, \ldots, T_{q,a}^{SC}\right\}$ can be expressed through $N_q\left(\boldsymbol{C\pi}, \boldsymbol{C\Sigma C}^t\right)$ for large sample sizes. The representation of $\boldsymbol{R}$ follows from equation (5.9). With the notations given so far the power of a MCT can be expressed according to

$$
\begin{aligned}
1-\beta &= P\left(T_a^{MC} \geq z_{q,1-\alpha} \middle| H_A\right) \\
&= P\left(\max_{1\leq l\leq q}\left\{T_{l,a}^{SC}\right\} \geq z_{q,1-\alpha} \middle| H_A\right) \\
&= 1-P\left(\max_{1\leq l\leq q}\left\{T_{l,a}^{SC}\right\} < z_{q,1-\alpha} \middle| H_A\right) \\
&= 1-P\left(T_{1,a}^{SC} < z_{q,1-\alpha} \wedge \ldots \wedge T_{q,a}^{SC} < z_{q,1-\alpha} \middle| H_A\right) \\
&= 1-P\left(\frac{T_{1,a}^{SC}-E_A\left(T_{1,a}^{SC}\right)}{\sqrt{V_A\left(T_{1,a}^{SC}\right)}} < \frac{z_{1-\alpha}-E_A\left(T_{1,a}^{SC}\right)}{\sqrt{V_A\left(T_{1,a}^{SC}\right)}} \wedge \ldots \wedge \frac{T_{q,a}^{SC}-E_A\left(T_{1,a}^{SC}\right)}{\sqrt{V_A\left(T_{q,a}^{SC}\right)}} < \frac{z_{1-\alpha}-E_A\left(T_{q,a}^{SC}\right)}{\sqrt{V_A\left(T_{q,a}^{SC}\right)}}\right) \\
&= 1-\boldsymbol{\Phi}_q\left(\left(\boldsymbol{z}_{q,1-\alpha}-\boldsymbol{e}\right)\mathrm{diag}\left(\tfrac{1}{v_1}, \ldots, \tfrac{1}{v_q}\right); \boldsymbol{0}, \boldsymbol{R}\right).
\end{aligned}
$$

For the sample size determination an analogous way of deriving an approximate formula as in the univariate case fails to succeed. This lies in the multivariate nature of the problem and the non-uniqueness of Φ_q^{-1} for $q>1$. However, if a closed formula can not be presented here, iterative methods similar to the ones presented on page 104 will work good as well. The idea behind this approach is that the power expression (5.13) increases monotonously with *N*. Therefore, with a given starting sample size allocation $\boldsymbol{n}=\left(n_0, \ldots, n_k\right)$ one calculates the corresponding power and decides upon the result, whether a higher sample size is required or not. This procedure can be repeated subsequently, until a fixed threshold value $1-\beta$ is reached.

We summarise briefly the results obtained so far. We successfully extended the technique of building the maximum over several contrasts to dichotomous data. Corresponding test statistics have been introduced and power formulas established in closed form. The natural progress would be to compare the behaviour of MCTs to other trend tests, for example the CA or a binary LRT. However, we do not carry out such comparisons for two reasons. First, the

isotonic trend test mentioned earlier is much less applied on binary data than in the case of normal data. The CA is definitive the method of choice. But its test statistic is more or less equivalent to single contrasts as it can be deducted by comparing (5.4) and (5.6). In balanced cases, for example, the choices of $\boldsymbol{d} = (0, 1, 2, 3)$ and $\boldsymbol{c} = (-3, -1, 1, 3)$ lead to exactly the same test statistics. This means that the CA is in fact a single contrast, in the balanced case a linear one. Therefore, there is no need to compare CA to MCTs on its own. Furthermore, power comparison similar to Chapter 4 leads to very similar curves and conclusion. An extensive power study has been conducted but the results from the case of normal variates just repeat. We therefore conclude similarly to Robertson et al. (1988, p. 168) that "... *the results ... will apply approximately if the normal variates are replaced by proportions.*" Referring to the conclusions given in Section 4.3. we advance in the next sections by trying to illustrate a series of important issues regarding dichotomous testing by focusing only on Marcus' contrast test. A SAS/IML-implementation of the power formulas in the univariate case (5.10) and in the multivariate case (5.13) is contained in the Appendix.

5.3. Introduction of new contrast tests

In a series of brief notes we introduce in the following some generalisations to the common form (5.5) and (5.6) of dichotomous contrast tests. Short comparisons among the several proposed methods are included in each subsection.

5.3.1. Continuity correction

In Section 5.1. we approximated the unknown exact distributions of the CA (5.3) and the contrast test (5.5) using the standard normal distribution. To account for the fact that a discrete distribution is approximated by a continuos function, it is traditional to include a continuity correction (c.c. in the following). The introduction of such c.c.'s, however, is not unique. The exact value of the c.c. should depend on the discreteness of the approximated distribution. If, for example, the discrete outcomes are 1, 2, 3, ... one would expect a c.c. of 0.5 to behave adequately. Otherwise, for the discrete values 4, 8, 12, ... a c.c. of 2 would perhaps be more appropriate.

For a discussion in depth within the common $2\times 2-$design we refer to Andrés and Mato (1996) and the references therein. They examined a total of 20 different methods for incorporating a c.c. and compared their performances – but without obtaining a reasonable solution. In multi-sample designs, such as investigated in the present thesis, the choice of a 'best' correction is even more difficult. The present section is therefore not thought as an detailed analysis on this subject. In contrast, it should be made clear that sometimes a c.c. might be required in real data situations. The practitioner should always be aware, whether to include or not to include a c.c. A working solution is applied to MCTs and some numerical results will show the reliability of the method within the situations investigated.

Consider the modified single contrast

$$T_a^{SC} = \frac{\sum_i \frac{c_i}{n_i} r_i - \frac{\Delta}{2}}{\sqrt{p(1-p)\sum_i c_i^2 / n_i}},$$

where $\Delta/2$ denotes the c.c. The aim is to find an appropriate value $\Delta > 0$ for the possible outcomes of the original test statistic $\sum_i \frac{c_i}{n_i} r_i$. Unfortunately, when the dose scores of the CA are not equally spaced, an adequate c.c. for all outcomes is missing. Consequently, Thomas et al. (1977) proposed to consider for the CA both the minimum and maximum interval between adjacent doses and therefore to provide lower and upper bounds. Applying their methods on contrast tests, we get the following possible c.c.s:

(cc1) $\Delta_1 = \max\limits_{1\le i\le k}\left\{\frac{c_i}{n_i} - \frac{c_{i-1}}{n_{i-1}}\right\};$

(cc2) $\Delta_2 = \min\limits_{1\le i\le k}\left\{\frac{c_i}{n_i} - \frac{c_{i-1}}{n_{i-1}}\right\}.$

Furthermore, we additionally investigate

(cc3) $\Delta_3 = \frac{\sum_i \left|\frac{c_i}{n_i} - \frac{c_{i-1}}{n_{i-1}}\right|}{k},$

which computes the arithmetic average over all adjacent intervals. The idea is to compensate possible exaggerated effects due to considering only the maximum or minimum in (cc1) and (cc2).

Figure 5.1. provides the size behaviour of Marcus' MCT including above three c.c.'s for two different risk levels. The values were obtained by performing a simulation study with 9,999 runs. At first it should be mentioned that the uncorrected MCT and the version cc2 including the minimum correction lead to the same test statistics because the maximum contrast always contains a pair of identical neighbouring contrast coefficients (recall the definition of Marcus' MCT in Section 3.3.). Therefore, both curves lie above each other. For the cases where $n_0 > n_i$ the MCT does not maintain the α–level of 5%. The actual size increases up to 10%. In the second part (b) rule cc1 performs clearly best, but gets too conservative in (a), especially if $n_0 < n_i$. Here, cc3 seems to behave better. It maintains the size and is still a little less conservative than the cc1. In situation (b), however, it also has its problems. Concluding, one can say that the size behaviour depends in a complicated fashion on several factors, not only on the total sample size, but also on its allocation, the risk π, the α–level and so on. From these two single figures one can already imagine, how difficult it is to obtain an adequate c.c. for all scenarios.

The power performance is illustrated in Figure 5.2. The power values were calculated analogously to Theorem 5.3. (the formulas generalise in a straight forward manner) and the SAS program provided in Appendix B.5.1 can be used for the calculations. Clearly, the uncorrected version yields the highest power together with rule cc2 (which in these cases again lead to identical test statistics). Rules cc3 and cc2 have a uniform lower power what is consistent with their improved size behaviour.

Other c.c.'s than the ones introduced above exist. Westfall and Lin (1988), for example, proposed to minimise a weighted function of the deviations of the approximating p-values from the exact permutation p-values for small totals $\sum_i r_i$. Next, one should employ the obtained estimate of the c.c. for the present data set with larger totals. However, this approach also represents a rather empirical method for estimating a reliable c.c. The search gets even more complicated by taking two further parameters into account (the weight function to be used and the total $\sum_i r_i$).

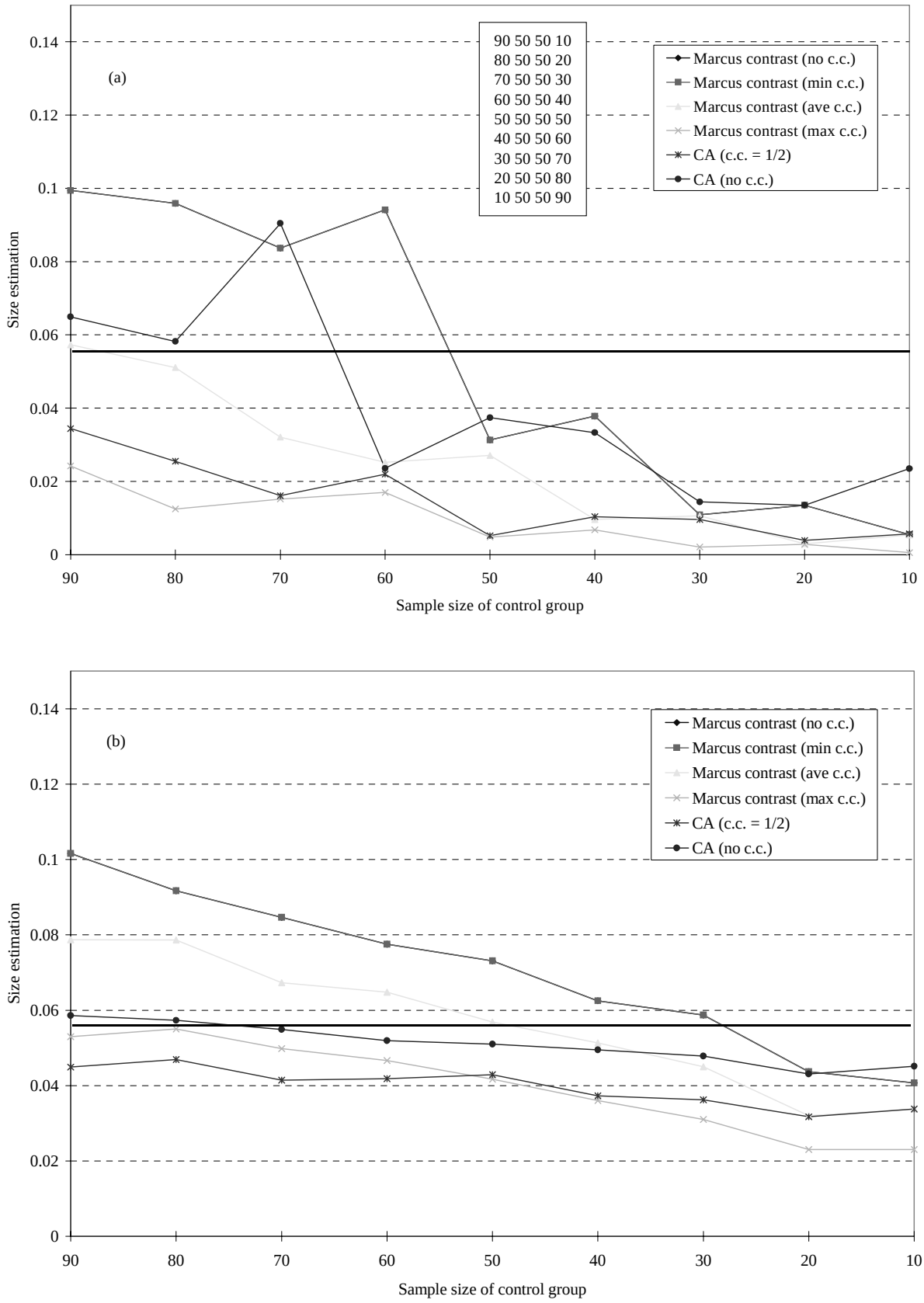


Figure 5.1. Estimation of size of Marcus' contrast test for several continuity corrections, unbalanced case with sample size allocation given in (a), $k = 3$, total sample size = 200, $\alpha = 0.05$ for (a) $\pi = 0.01$ and (b) $\pi = 0.1$.

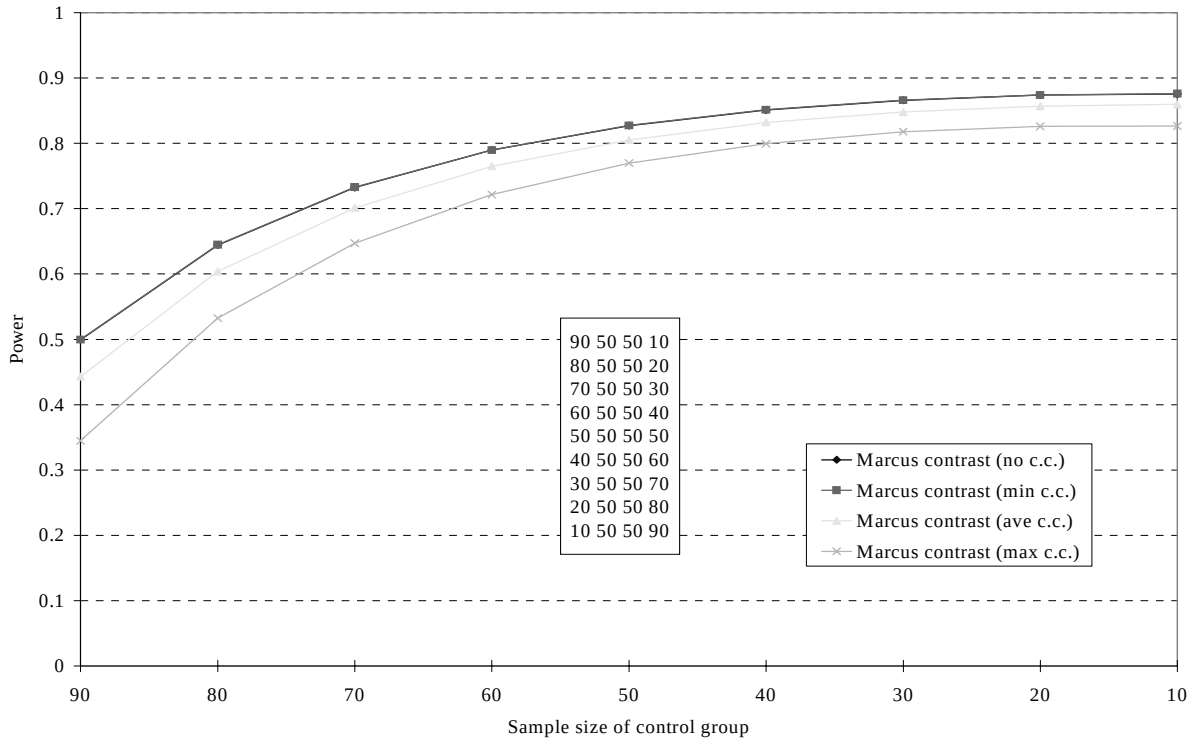


Figure 5.2. Power comparison of Marcus' contrast test for several continuity corrections, unbalanced case with sample size allocation given in the Figure, k = 3, total sample size = 200, α = 0.05 for the convex profile $\pi = (0.05, 0.05, 0.05, 0.2)$.

From above discussion the following conclusions are drawn. At first, one should always have in ones mind that no optimal c.c. in the multi-sample design exists. The test statistics might get too conservative for a series of situations or they do not maintain the size in other cases. According to one's expectation, the power decreases with respect to increasing c.c. But the power differences are sometimes rather high. In Figure 5.2. we could observe differences up to 10% in the usual balanced case. Rules cc1 and cc3 seem to be more adequate than cc2 to control the size α. Combining these results with the power values, rule cc3 might be the method of choice. It is not as conservative as cc1, but maintains the size much better than cc2 or even the uncorrected test. But the practitioner should always be aware whether a continuity correction is really needed. Especially in balanced or nearly balanced situations such adjustments might not be required and they decrease the power significantly. Continuity corrections should therefore always '*... be invoked with caution*' (Hirji and Tang, 1998, p. 958). One important result, which Andrés and Mato (1996) obtained from their study and which continues to hold for the present multi-sample situation, is that the performance of c.c. does not depend on the total sample size but rather on its allocation and the relationship among the risks.

5.3.2. Unpooled version

Similarly to the case of normal variates the introduction of an adequate variance estimator is not unique either when conducting asymptotic dichotomous tests. In equation (5.6) we proposed the use of a pooled variance estimator, i.e. we considered $p = R/N$. Other possibilities, however, exist. Koch (1996) reviewed several binomial many-to-one procedures and compared their performances (in particular the size behaviour) extensively. Bristol (1993) introduced power expressions in closed form in the same set-up for several Dunnett-versions: pooling over all groups, no pooling at all, pairwise pooling and an arcsine transformation. A subset pooling makes no sense in the framework of trend tests and the arcsine transformation is known to perform well for high sample sizes only (Koch, 1996). Therefore, we restrict our attention to the unpooled version by extending the results of Bristol (1993) to general MCTs.

In analogy to (5.6) we define the asymptotic unpooled single contrast

$$T_a^{uSC} = \frac{\sum_i \frac{c_i}{n_i} r_i}{\sqrt{\sum_i p_i (1-p_i) c_i^2 / n_i}}.$$

From now on T_a^{pSC} denotes the pooled contrast and former T_a^{SC}. Consequently, T_a^{uMC} denotes the asymptotic MCT, which takes the maximum over q unpooled SCTs. Then we can derive a closed form expression for the power exactly in the same way as done in Theorem 5.2., i.e. we determine the probability of correctly rejecting the null hypothesis.

Theorem 5.4.: Let $T_a^{uMC} = \max\left\{T_{1,a}^{uSC}, \ldots, T_{q,a}^{uSC}\right\}$ be a binomial asymptotic unpooled MCT. Denote by $\boldsymbol{C} = \left(c_{ij}\right)_{ij} = \left(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_q\right)^t$ the $q \times (k+1)$ contrast matrix and $\delta = \left(\frac{\sum c_{li} \pi_i}{\sqrt{\sum \pi_i (1-\pi_i) c_{li}^2 / n_i}} \right)_{1 \le l \le q}$. Then the asymptotic power of T_a^{uMC} is given by

$$1 - \beta = 1 - \boldsymbol{\Phi}_q\left(\boldsymbol{z}_{q,1-\alpha} - \delta; \boldsymbol{0}, \boldsymbol{R}\right),$$

where $\boldsymbol{z}_{q,1-\alpha} = \left(z_{q,1-\alpha}, \ldots, z_{q,1-\alpha}\right)$ stands for the q-variate normal $(1-\alpha)-$equipercentage point and the elements of $\boldsymbol{R}$ are given by equation (5.13).

Proof: Similarly to Theorem 5.1. one obtains that the joint distribution of $\left\{T_{1,a}^{uSC}, \ldots, T_{q,a}^{uSC}\right\}$ can be expressed through $N_q\left(\boldsymbol{C}\boldsymbol{\pi}, \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^t\right)$ for large sample sizes. Furthermore, the representation of $\boldsymbol{R}$ follows from equation (5.9). Finally, the power can be expressed according to

$$
\begin{aligned}
1-\beta &= P\left(T_a^{uMC} \geq z_{q,1-\alpha} \middle| H_A\right) = \\
&= P\left(\max_{1\leq l\leq q}\left\{T_{l,a}^{uSC}\right\} \geq z_{q,1-\alpha} \middle| H_A\right) = \\
&= 1-P\left(\max_{1\leq l\leq q}\left\{T_{l,a}^{uSC}\right\} < z_{q,1-\alpha} \middle| H_A\right) = \\
&= 1-P\left(T_{j,a}^{uSC} < z_{q,1-\alpha}, \quad j=1,\ldots,q \middle| H_A\right) = \\
&\approx 1-P\left(\sum_i c_{ji}p_i < z_{q,1-\alpha}\sqrt{\textstyle\sum_i \pi_i(1-\pi_i)c_{ji}^2/n_i}, \quad j=1,\ldots,q \middle| H_A\right) = \\
&= 1-P\left(\frac{\sum_i c_{ji}p_i - \sum_i c_{ji}\pi_i}{\sqrt{\sum_i \pi_i(1-\pi_i)c_{ji}^2/n_i}} < z_{q,1-\alpha} - \frac{\sum_i c_{ji}\pi_i}{\sqrt{\sum_i \pi_i(1-\pi_i)c_{ji}^2/n_i}}, \quad j=1,\ldots,q\right) = \\
&= 1-\Phi_q\left(\boldsymbol{z}_{q,1-\alpha}-\delta;\, \boldsymbol{0},\, \boldsymbol{R}\right).
\end{aligned}
$$

We next investigate briefly the performance of the new class both under H_0 and H_A. For an easier comparison to previous results, similar constellations already analysed in the preceding subsection has been chosen again. Figure 5.3. illustrates for the same unbalanced situations and background risk π as in Figure 5.1. the size behaviour of Marcus' MCT. Again, continuity corrections are straight forward included in above definition and power expression of T_a^{uMC}. For the same c.c.'s proposed in Subsection 5.3.1. we get the following results. For small values of π both the maximum and the average rules are very conservative with an actual size well below 2%. On the other hand, the uncorrected version and the minimum correction lead again to identical results (see the discussion on page 140). Both do not control α for stronger imbalances when $n_0 < n_i$. Things change when π increases. All tests perform worse with increasing imbalances for situations of $n_0 < n_i$. The actual sizes can reach values up to 40% (!) for extreme imbalances. No control of the α-level is given any more. The maximum c.c. with

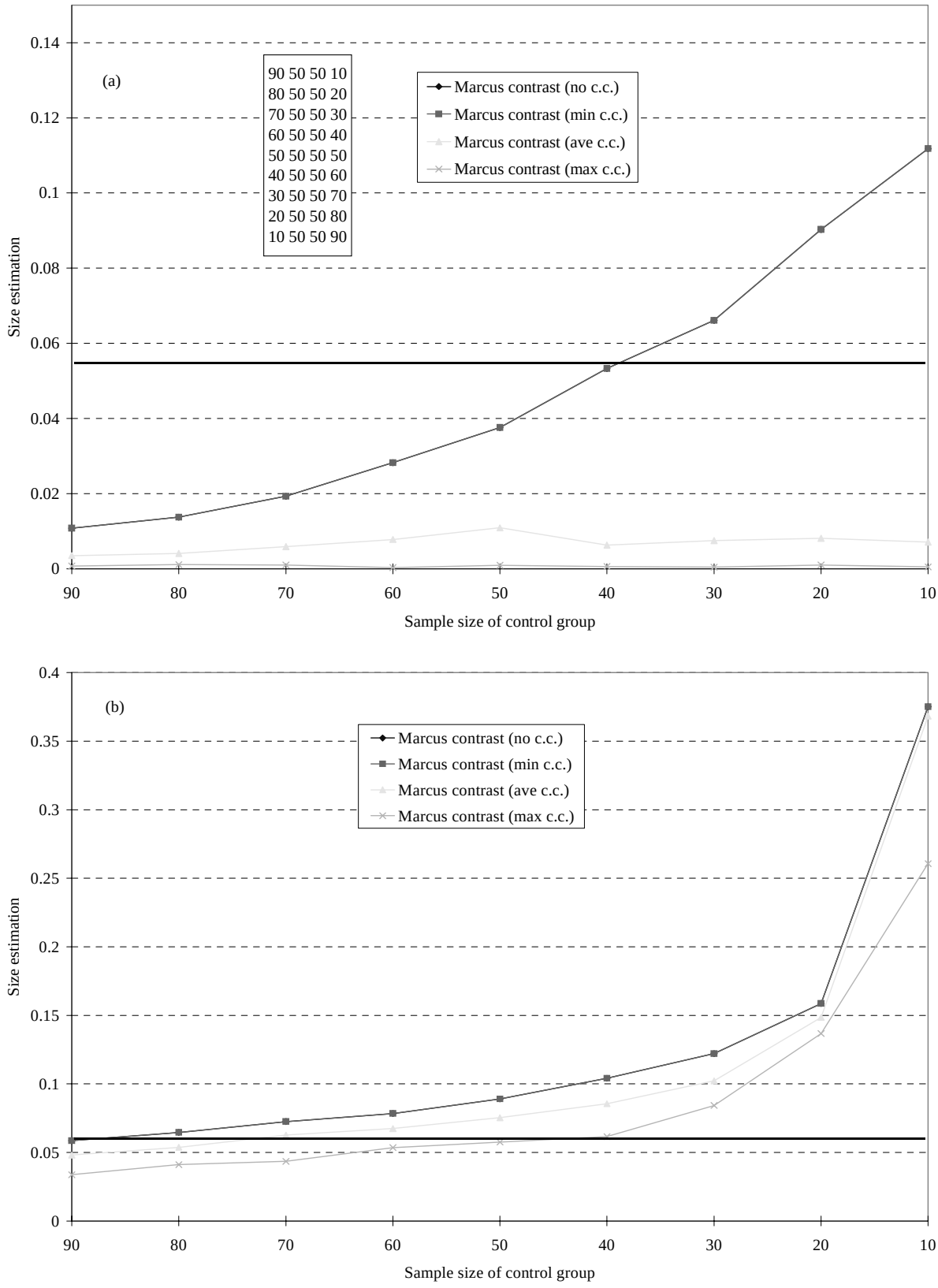


Figure 5.3. Estimation of size of Marcus' unpooled contrast test for several continuity corrections, unbalanced case with sample size allocation given in (a), k = 3, total sample size = 200, α = 0.05 for (a) $\pi = 0.01$ and (b) $\pi = 0.1$.

$\hat{\alpha} = 26\%$ even performs 'best'. This behaviour is in fact consistent with simulations in the literature for similar problems. Koch (1996), for example, obtained similar size excesses for an unpooled Dunnett version investigated in detail.

Even if we could conclude from Figure 5.3. that T_a^{uMC} is not an level-α test any more, we provide for the sake of completeness some power comparisons for the sample size allocations investigated so far (Figure 5.4.). The power advantages of the unpooled versions over the pooled tests are not as big as one could expect from the previous size behaviour. In fact, for situations with $n_0 > n_i$ T_a^{pMC} tends to be more powerful than T_a^{uMC} for all c.c.'s investigated. In balanced situations or if $n_0 < n_i$, T_a^{uMC} are clearly better, with power differences up to 15% for the respective c.c.'s.

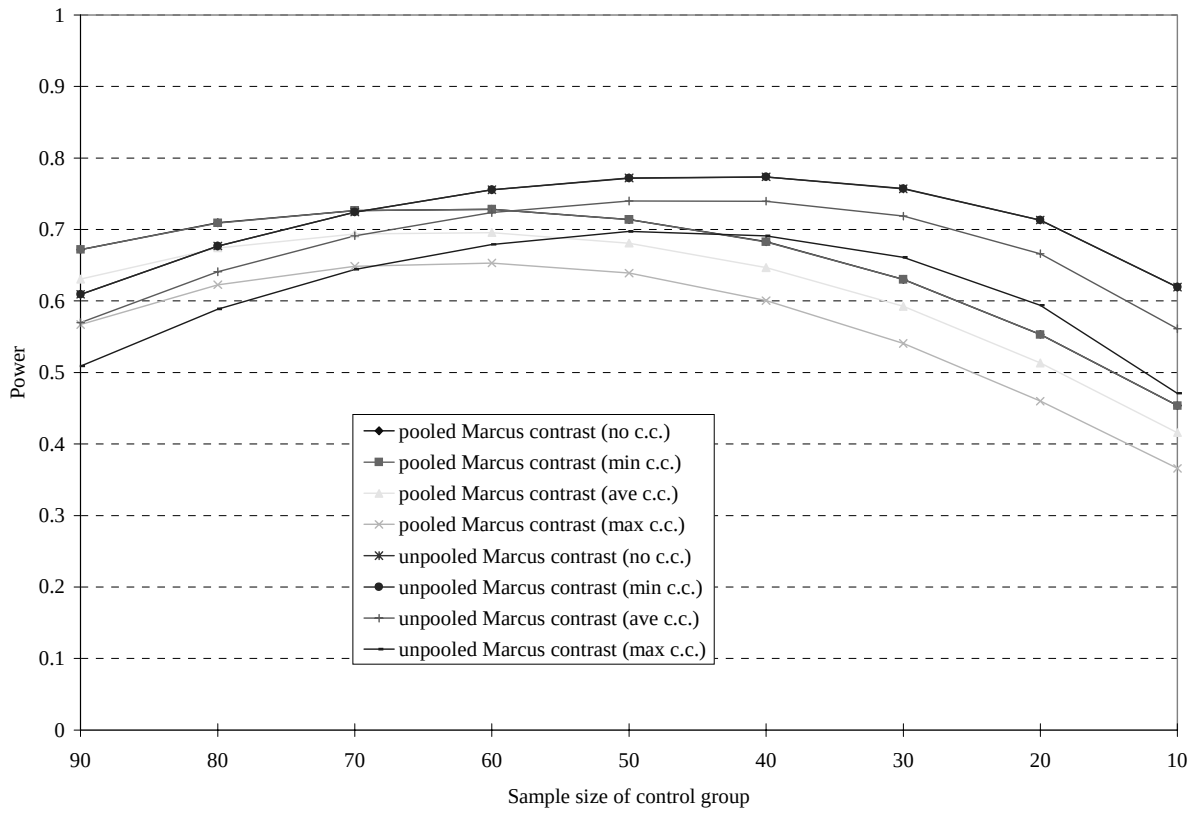


Figure 5.4. Power comparison of Marcus' unpooled contrast test for several continuity corrections, unbalanced case with sample size allocation given in the Figure 5.3., k = 3, total sample size = 200, α = 0.05 for the linear profile $\pi = (0.05, 0.1, 0.15, 0.2)$.

From above considerations we can only give the recommendation to avoid unpooled MCTs, if either strong imbalances occur (in particular if $n_0 << n_i$), or the background risk is very high. Moreover, the power advantages are not that high that one could ignore the poor size behaviour. Otherwise T_a^{uMC} might be conducted with caution, i.e. in balanced situation or $n_0 > n_i$, if π is not too large. Again, as in the case of T_a^{pMC}, one should always be aware, whether to use a c.c. or not.

5.3.3. Exact conditional and unconditional versions

From the preceding two subsections we could conclude that asymptotic tests do not control the pre-determined α-level for all set-ups. In particular for situations of small sample sizes (a problem of frequent occurrence in practice) and strong imbalances (again of practical importance due to possible mortality effect in high dose groups or due to allocation according to the $\sqrt{k}$–rule) the asymptotic versions, both pooled and unpooled, failed to succeed. Continuity corrections have seen to yield limited improvements only.

In this subsection we shall analyse exact MCTs, which make use of the underlying exact permutational distribution of the present data set. Such tests perform in particular well in small sample situations and fill therefore the gap left by the asymptotic counterparts. In the following we extend the results of Neuhäuser (1996) who in turn applied successfully the ideas of Williams (1988) and Storer and Kim (1990) on contrast tests and the underlying multi-sample situation.

We first consider the case of conditioning on the total sum of responses $R = \sum_i r_i$. Williams (1988) proposed an algorithm to calculate exact significance probabilities based on the multivariate hypergeometric distribution. The exact conditional p-value is then given by

$$P\left(T \geq t_0 \middle| \sum_i r_i = R,\ H_0\right) = \frac{\sum_{\{\boldsymbol{r} \in \Gamma_R,\, T(\boldsymbol{r}) \geq t_0\}} \prod_{i=0}^{k} \binom{n_i}{r_i}}{\binom{N}{R}},$$

where $\Gamma_R = \{\boldsymbol{r} = (r_0, \ldots, r_k)\colon r_i = 0, 1, \ldots, \min(n_i, R), \sum_i r_i = R\}$ denotes the total sample space and t_0 is the observed test statistic.

Another possibility of conducting exact tests is not to condition on a fixed total of responses. Instead, R is regarded as a random variable what leads to a larger sample space and less discrete underlying test statistics. Improved size and power performances therefore hold. However, the debate which methodology would be more appropriate (conditional or unconditional approach) does not fall in the scope of the thesis and we refer to the references given in Andrés and Mato (1996).

Koch (1996) successfully extended the procedure of Storer and Kim (1990) on situations of comparing several groups. Their original method determines in the classical 2×2-design the distribution of the test statistic using the MLE $p = \frac{r_0 + r_1}{n_0 + n_1}$ under H_0:

$$P(\boldsymbol{r} = (r_0, r_1) | H_0) = \binom{n_0}{r_0} \binom{n_1}{r_1} p^{r_0 + r_1} (1 - p)^{n_0 + n_1 - r_0 - r_1}.$$

Generalising the resulting approximate unconditional test to multiple sample situations, one obtains for the significance level

$$P(T \geq t_0 | H_0) = \sum_{\{\boldsymbol{r} \in \Gamma, T(\boldsymbol{r}) \geq t_0\}} \prod_{i=0}^{k} \binom{n_i}{r_i} p^{r_i} (1 - p)^{n_i - r_i}$$

where $\Gamma = \{\boldsymbol{r} = (r_0, \ldots, r_k)\colon r_i = 0, 1, \ldots, n_i\} = \bigcup_{R=0}^{N} \Gamma_R$.

To compare both exact procedures, we conducted a size and power simulation study. Several situations were investigated, from which we report the most important results. Exact power evaluations are only possible for the conditional test and are still numerical awkward (Tang et al., 1995; Mehta et al., 1998; Corcoran et al., 1998). Therefore we restricted ourselves to a simulation study with 1,000 simulation evaluations each. Table 5.2. presents the values calculated for the profiles $\pi = (0.05, 0.05, 0.05, 0.05)$ and $\pi = (0.05, 0.1, 0.15, 0.2)$. In accordance with above discussions only Marcus' MCT has been included in the results which

Sample size	Asymptotic	Conditional		Unconditional	
n_i	T_a^{pMC}	T_a^{pMC}	T^{MC}	T_a^{pMC}	T^{MC}
5	0.157 (0.014)	0.054 (0.001)	0.050 (0.001)	0.095 (0.001)	0.050 (0.012)
10	0.236 (0.037)	0.174 (0.006)	0.185 (0.007)	0.183 (0.007)	0.186 (0.020)
15	0.312 (0.044)	0.308 (0.017)	0.294 (0.015)	0.309 (0.013)	0.274 (0.035)
20	0.384 (0.053)	0.382 (0.024)	0.359 (0.022)	0.385 (0.020)	0.390 (0.035)
25	0.451 (0.059)	0.432 (0.030)	0.445 (0.029)	0.438 (0.029)	0.447 (0.039)

Table 5.2. Power (and size) of Marcus' MCT for k = 3, balanced case, $\alpha = 5\%$.

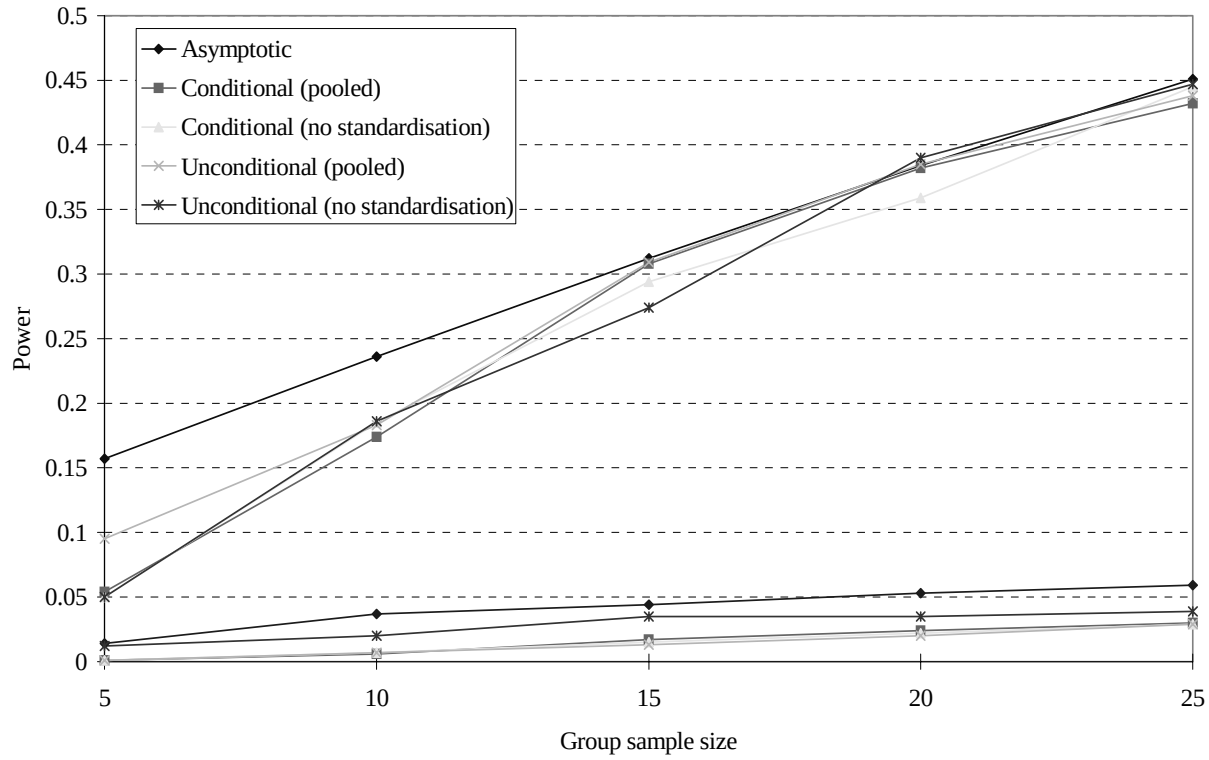


Figure 5.5. Graphical illustration of Table 5.2.

Sample size	Asymptotic	Conditional		Unconditional	
n_i	T_a^{pMC}	T_a^{pMC}	T^{MC}	T_a^{pMC}	T^{MC}
25 5 5 5	0.368	0.235	0.213	0.277	0.225
	(0.081)	(0.014)	(0.010)	(0.029)	(0.017)
30 10 10 10	0.426	0.390	0.339	0.401	0.347
	(0.095)	(0.037)	(0.027)	(0.047)	(0.034)
35 15 15 15	0.485	0.442	0.450	0.469	0.440
	(0.095)	(0.037)	(0.032)	(0.037)	(0.032)
40 20 20 20	0.541	0.537	0.493	0.550	0.509
	(0.081)	(0.029)	(0.021)	(0.035)	(0.028)

Table 5.3. Power (and size) of Marcus' MCT for $k = 3$, unbalanced case, $\alpha = 5\%$.

are graphically shown in Figure 5.5. We applied above formulas on both standardised (5.7) and non-standardised (5.8) pooled statistics for both conditional and unconditional approach, respectively. No continuity corrections at all were considered. Unpooled versions were originally included as well but performed less powerful in our studies. Their results are therefore omitted here.

It comes clearly out that the asymptotic uncorrected test performs best in terms of power. Especially for small sample sizes it yields the highest power values whereas all exact tests show a poor power behaviour for n_i = 5 or 10. With increasing sample sizes however, the differences diminish and the exact tests are as good as the asymptotic ones. The other side of the coin is to investigate the respective size performances. In fact the asymptotic test holds the size fairly well, in particular for smaller sample sizes. Only for n_i = 25 the test gets slightly liberal. However, these are very special results, which hold in the balanced case. As seen from Table 5.3. the size behaviour becomes critical if imbalances occur. The uncorrected asymptotic test clearly fails to maintain the nominal 5%-level. Nevertheless, these results demonstrate that uncorrected asymptotic tests might work well in the balanced case (even for n_i = 5). The exact versions are all very conservative with an actual size less than 4% in most of the cases. Moreover, they are up to now computationally not feasible, limiting their application in fact to small sample designs.

Other possible definitions instead of the conventional $p = \Pr(T \geq t_0)$ exist in the context of permutation tests. All of them are motivated by the conservativeness of the exact tests and try to alleviate this problem. On the other side, the α–level is possibly not guaranteed any more, although this property is one major reason for the application of exact tests. One well-known p-value adjustment is the mid–*p* concept of Lancaster (1961) given by $p = P(T > t_0) + \frac{1}{2} P(T = t_0)$. More recently, Chen et al. (1997) introduced another less conservatived method of calculating *p* by excluding some permutations whose probabilities of occurrence are greater than the probability of the observed outcome. For a general up to date review of randomisation procedures we refer to Seidel (1999).

5.3.4. Example

Recall the example (p. 126) of investigating the influence of germinal temperature on the occurrence of anomalies in young kohlrabi plants (Habegger and Wiebe, 1985). Table 5.4. summarises the data obtained from the randomised one-way layout for 6 temperature treatments, where 16° is regarded as the control group (standard treatment).

Temperature	16°	14°	11°	8°	5°	2°	Σ
Number of plants bearing anomalies	3	3	9	10	14	12	51
Number at risk	100	100	100	100	100	100	600
p_i	0.03	0.03	0.09	0.10	0.14	0.12	0.085

Table 5.4. Summarising results of occurrence of anomalies in young kohlrabi plants.

We want to analyse, whether a statistical significant trend with respect to decreasing temperature treatments exists. At first we conduct the pooled and unpooled asymptotic Marcus MCT, both with including different continuity corrections. The following table summarises the test statistics together with the respective p-value.

Test	Pooled asymptotic test				Unpooled asymptotic test			
C.c.	no c.c.	min c.c.	ave c.c.	max c.c.	no c.c.	min c.c.	ave c.c.	max c.c.
Statistic	3.5857	3.5857	3.5499	3.496	4.1553	4.1553	4.1175	3.9664
P-value	0.0011	0.0011	0.0013	0.0016	0.0001	0.0001	0.0002	0.0003

Clearly, the uncorrected tests yield the highest test statistics and hence the lowest p-values. The minimum c.c. does not change the statistic and lead to the same p-values. The unpooled version are slightly better, but all p-values are so small that the null hypothesis can be rejected without hesitation. Conducting conditional exact tests for the different versions introduced above results in similar p-values to those given for T_a^{uMC}. The conduction of unconditional tests were impossible because of the high total sample size of 600.

6. Estimation of the minimum effective dose

We now focus on the problem of estimating the smallest dose that produces a response significantly different from the control (minimum effective dose, abbreviated MED hereafter). This goal can be achieved within the closure test principle of Marcus et al. (1976). Proceeding from the two main assumptions, comparison to a control and monotone dose-response relationship, it can be shown (e.g. Hothorn et al., 1997) that the identification of the MED consists of a series of inferences in a specified order. Failure of achieving the desired inference at any step renders subsequent comparisons unnecessary. In dose-response studies, for example, it is desirable for a method to not declare a lower dose efficacious if this could be done for a higher dose. Following this goal, we achieve this procedure by answering the question "$\mu_i > \mu_0$" in a stepwise fashion, continuing only while the answer is affirmative. As the closure principle imposes no restriction on the choice of the tests used, all approaches considered above fit into this context and will be applied on an example in this chapter. For more details on this topic we refer to the recent articles of Rom et al. (1994), Tamhane et al. (1996), Bauer (1997), Amaratunga and Ge (1998), Sidik and Morris (1999), Dunnett and Tamhane (1998) and with particular attention on efficacy and safety of compounds Ruberg (1995).

The closure test principle according to Marcus et al. (1976) is a widely used method for combining several statistical statements and simultaneously controlling the arising global error. Starting from the finite set of hypotheses $\left\{H_i \middle| i \in I,\ I = \{1, \ldots, p\}\right\}$ to be investigated it forms the closure of this family by taking all possible non-empty intersections $H_S = \bigcap_{i \in S} H_i,\ \forall S \subseteq I$. If an α-level test for each hypothesis H_S is available, then the closed testing procedure rejects any hypothesis H_S if and only if every H_T is rejected by its associated α-level test for all $T \supseteq S$. This procedure is proven to hold the global error α (see Hochberg and Tamhane (1987, pp. 54) for the proof and a further comprehensive overview on this subject). Two main advantages characterise this procedure. On the one hand it leaves the choice of the individual tests entirely free and one can chose the appropriate test procedure for the analyse of a designated hypothesis. Even the change of individual test procedures within the closure test is allowed. Second, under all multiple test it is that one which yields the

highest power and therefore provides the highest statistical information (Horn and Vollandt, 1995, p. 19). However, an overblown set of hypotheses to be tested may reduce its efficacy.

Applying these theoretical results under the consideration of the two additional assumptions comparison to a control and monotone dose-response assumption we get the following systematic for the realisation (Hothorn et al., 1997). The set of null hypotheses to be tested is defined as

$$H_{0(i)}: \mu_0 = \mu_1 = \ldots = \mu_i \qquad (6.1)$$

versus

$$H_{A(i)}: \mu_0 \leq \mu_1 \leq \ldots \leq \mu_{i-1} \leq \mu_i,\ \mu_0 < \mu_i, \qquad k \geq i \geq 1. \qquad (6.2)$$

Then the sequence

$$H_{0(k)},\ H_{0(k-1)},\ \ldots,\ H_{0(2)},\ H_{0(1)}$$

will be tested hierarchically at the conditional level α = 5% until a null hypothesis fails to be rejected. In other words, the procedure begins with testing $H_{0(k)}$, the simultaneous comparison of all treatment groups with the negative control (the question whether a global trend exists or not). If there is no indication for a trend the null hypothesis will not be rejected and the procedures stops. Otherwise $H_{0(k-1)}$ is investigated. This process holds on until a null hypothesis is not rejected any more or the sequence above comes to the end, in which case all treatment groups are considered to be different from the negative control. The first to derive this sequential procedure for order restricted inferences was Williams himself (1971, 1972), who applied this methodology on his own new developed test. But he did not generalise his approach in the sense of the results established later by Marcus et al. (1976).

A general testing procedure to estimate a MED based on these considerations is given in the following algorithm. The T_i's stand for the adequate trend tests for $[C- = D_0, D_1, \ldots, D_i]$ at stage *i*.

1. **INPUT** α, k, T_i.
2. Initialise $i = k$, $crit = 0$.
3. **REPEAT**
   a) Evaluate $p = p(T_i) = \text{p - value of } T_i$.
   b) **IF** $p > \alpha$ **THEN** $crit = 1$.
   ELSE $i = i - 1$.
   UNTIL ($i < 1$ **OR** $crit = 1$).
4. Set $MED = i + 1$.
5. **OUTPUT** *MED*, *p*.

If the procedure fails to detect a significance at the first step, *MED* is virtually set as $k + 1$. This means that no true MED has been found, i.e. no monotonous dose-response relationship is considered to hold. Otherwise the algorithm outputs the estimate of the MED together with the corresponding p-value calculated.

As an illustration of the procedure we apply the six trend tests considered in Chapter 4. Consider the data in Table 6.1., taken from Ruberg (1995). The example is chosen from an experiment involving the comparison of nine active dose groups against a negative control (vehicle group) with six animals per group. Because the dose-response curve follows a sigmoidale course (see Figure 6.1.), several dose-response shapes will appear during the process of identifying the MED. This example is therefore very illustrative to investigate the behaviour of the several procedures in dependency of convex, concave and other profiles. In Table 6.1. the test statistics (upper entry) and the respective p-values (lower entry) are given for Williams' and Marcus' original tests, the LRT and all three multiple contrast tests considered so far.

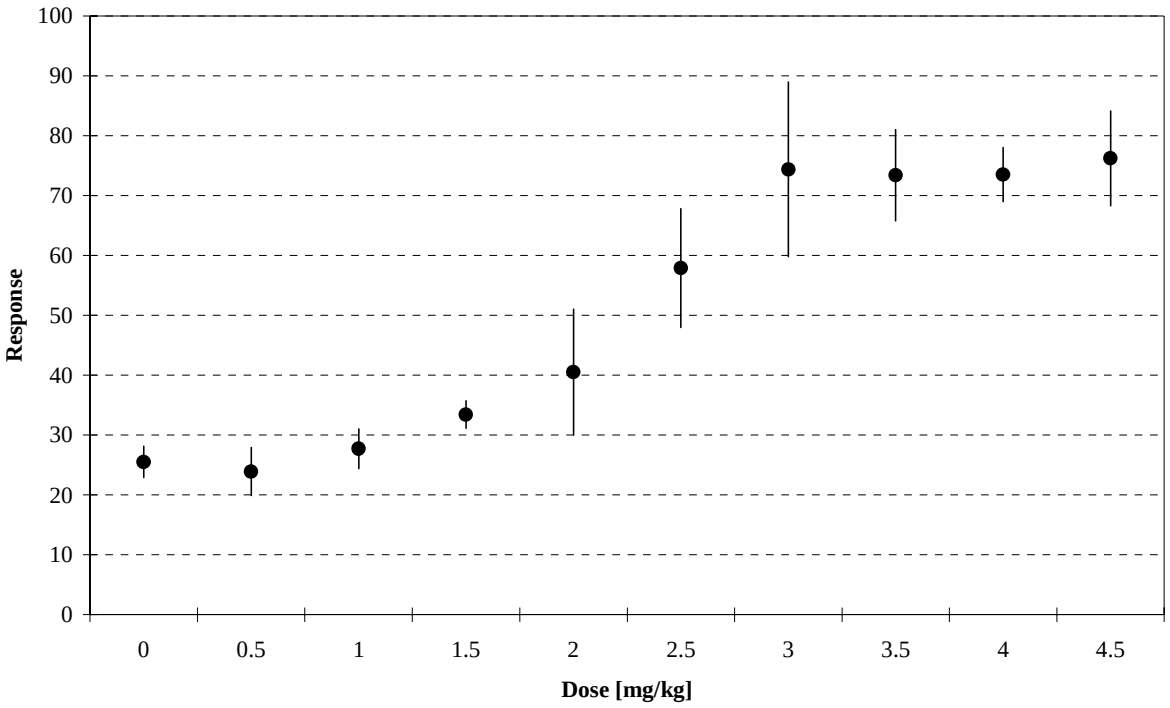


Figure 6.1. Example dose-response data represented through mean ± standard deviation (Ruberg, 1995).

The p-values were obtained as follows. Because of the present balanced design $(n_i = 6)$, the SAS-call (1.8) was used for Williams' $\bar{t}$. The conduction of Marcus' original test required 9,999 simulation runs. The p-values of the LRT were obtained making use of the SAS/IML algorithm B.2.4 in the Appendix (which is based on the method of Sun, see Subsection 2.1.2.3.). For the MCTs the lattice rule implementation of Genz and Bretz (1999) was used instead (see Subsection 2.2.2.2. and the corresponding implementation B.2.9; ε = 0.001). Note that computing the p-values for the isotonic contrast for high values of *k* can be very tedious without an dimensionality reduction as presented in Section 3.6., even if the Solow procedure is conducted.

As we assume a common variance, one calculates $s^2 = 60.078$. Furthermore, applying Lemma 1.2. one obtains for the restricted estimates

$$\hat{\mu} = (24.7,\ 24.7,\ 27.7,\ 33.4,\ 40.5,\ 57.9,\ 73.77,\ 73.77,\ 73.77,\ 76.2).$$

Dose (mg/kg)	Mean Std. dev.	Williams' $\bar{t}$	Marcus' $\bar{t}^{\,\mathrm{mod}}$	LRT	Williams contrast	Marcus contrast	Isotonic contrast
4.5	76.2	11.3295	11.5083	0.9012	13.815	20.8937	21.0361
	7.9	< 0.0001	0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
4	73.5	10.7857	10.9645	0.8950	13.2098	19.0921	19.2691
	4.5	< 0.0001	0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
3.5	73.4	10.7857	10.9645	0.8856	12.4887	16.9405	17.2627
	7.6	< 0.0001	0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
3	74.4	10.7857	10.9645	0.8603	10.9273	14.2842	14.0582
	14.6	< 0.0001	0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
2.5	57.9	7.2402	7.4189	0.7316	7.2402	8.9614	8.8126
	9.9	< 0.0001	0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
2	40.5	3.3519	3.5307	0.4212	3.3519	4.0769	4.1953
	10.5	0.0008	0.0006	0.0005	0.0017	0.0004	0.0002
1.5	33.4	1.7653	1.9441	0.2002	1.7653	2.2449	2.2273
	2.3	0.0511	0.0494	0.0455	0.069	0.0418	0.0435
1	27.7	0.4916	0.6704	0.0381	0.4916	0.7741	0.7741
	3.3	0.3752	0.5056	0.3434	0.3862	0.3442	0.3442
0.5	23.9	–0.3575	0	0	–0.3575	–0.3575	–0.3575
	4	0.6389	1.0000	1.0000	0.6389	0.6389	0.6389
0	25.5	%	%	%	%	%	%
	2.6						

Table 6.1. Estimation of the minimum effective dose: data and summary statistics for different multiple procedures; upper entry: test statistic, lower entry: p-value.

According to Williams (1971, 1972) one computes $\hat{\mu}$ only once and use these estimates for all subsequent steps of the sequential procedure. We applied this approach also for Marcus' $\bar{t}^{\,\mathrm{mod}}$. The analogous way of calculation for the MCTs is to maintain all dose groups in the contrast matrices. At lower steps *i*, where only the first $i + 1$ dose groups should actually be considered, the remaining last $k - i$ contrast coefficients are set as '0'. In contrast, when conducting the LRT the isotonic estimates were evaluated anew at each step.

Because the response in the high doses is very strong, no test has problems to declare them significant. But focusing the attention to the dose 1.5 mg/kg, we notice that by reason of the convex shape for the last 3 + 1 doses, Williams' test as well as the Williams-type multiple contrast do not reject the null hypothesis (p-values 0.0511 and 0.069, respectively). Therefore they declare the dose D = 2.0 mg/kg to be the MED. The other four tests, in contrast, perform better, achieving an improved MED-estimation. But caution has to be taken upon the result for Marcus' $\bar{t}^{\,\mathrm{mod}}$. The resulting p-value 0.0494 has to be interpreted with great care because of the simulation evaluation conducted. Using a different seed for the random number generator other p-values will be obtained. Certainly, a series of values will lie above the 0.05-limit (in which case the null hypothesis would not be rejected). Therefore, similar to Example 1.4. we are again confronted with the situation that our final decision upon rejecting or not the null hypothesis is directly determined through the arbitrary definition of the seed.

So far we considered only the minimum dose, which shows a significant response increase with respect to a control group under the total order assumption. However, other relevant doses or treatments could also be of interest for the practitioner. A vast literature exists for estimating such doses, for example the maximum dose, which is still equivalent to the control group (MEQD), the no-observed-adverse-effect-level (NOAEL), the highest efficient dose step (HEDS) or the maximum tolerated dose (MTD). All of these problems can be analysed by use of appropriate multiple contrast tests (not necessarily under the total order restriction, see also p. 28).

Instead we focus now briefly on a further dose, called the maximum effective dose. Remmenga et al. (1997) define it as "... *the dose above which no significant improvement in efficacy is obtained.*". In other words, one tries to estimate the smallest dose, which still shows a maximum effect. For our example in Figure 6.1. we would clearly point D = 3.0 mg/kg out. Using the example above we show how the methodology developed so far can be varied in order to apply it to different settings and problems of interest. For this we define the following procedure to obtain the estimate. Instead of comparing several treatments to C_-, we invert the problem and compare the treatments to the maximum dose. As the monotonicity assumption is still supposed to hold, the trend tests considered in this thesis can again be applied. The only difference is that one now tests on a decreasing response with respect to the maximum dose. Furthermore, we state analogously to (6.1) and (6.2) the hypotheses

$$H'_{0(i)}: \mu_k = \mu_{k-1} = \ldots = \mu_i \qquad (6.3)$$

versus

$$H'_{A(i)}: \mu_k \geq \mu_{k-1} \geq \ldots \geq \mu_{i+1} \geq \mu_i,\ \mu_k > \mu_i, \quad 0 \leq i \leq k-1. \qquad (6.4)$$

Afterwards a similar sequential procedure as conducted for the MED is applied. The sequence $H'_{0(0)}$, $H'_{0(1)}$, …, $H'_{0(k-2)}$, $H'_{0(k-1)}$ will be tested hierarchically until a null hypothesis fails to be rejected. First, $H'_{0(0)}$ is tested, the simultaneous comparison of all lower groups with the highest dose group (the question whether a global trend exists or not). If there is an indication for a decreasing trend, $H'_{0(1)}$ is investigated. Otherwise the procedure stops and no maximum effective dose exists (the procedure failed to show an overall trend). This process holds on until a null hypothesis is not rejected any more or the sequence above comes to the end. In the latter case, all treatment groups are considered to be different from the maximum dose D_k and the sought maximum efficient dose is set as D_k itself. Otherwise, the first not rejected dose is set as the desired maximum efficient dose. For an illustration of this procedure the associated test statistics and p-values of above example are summarised in Table 6.2. It transpires that all tests clearly distinguish D = 3 mg/kg to be the estimate of the desired maximum effective dose.

Before leaving this chapter it should be noticed that the trend test approach is not the only possible way of estimating a maximum effective dose. In fact it is often required to control the second type error in such situations (consumers risk). In these cases it would be adequate to follow a stepwise procedure (beginning at D_k), where at each stage an intersection union test is conducted. At each step both a significant difference to the control (either trend test under order restriction or pairwise testing without order restriction) and a two-sided equivalence to the higher doses have to be established by rejecting the corresponding null hypotheses. The smallest dose, for which this condition is satisfied as well as for all higher doses, is estimated to be the sought maximum effective dose. Because this methodology lies beyond the scope of this thesis we do not consider it further. Instead we refer to Bauer et al. (1998) and Hothorn and Hauschke (1999) for a deeper discussion on this subject (use of the intersection union test and the principle of equivalence testing in dose finding studies, respectively).

Dose (mg/kg)	Williams' $\bar{t}$	Marcus' $\bar{t}^{\,\mathrm{mod}}$	LRT	Williams contrast	Marcus contrast	Isotonic contrast
0	11.5083	11.5083	0.9012	13.8210	20.8937	21.0361
	< 0.0001	0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
0.5	11.5083	11.5083	0.8959	13.1003	19.2177	19.3073
	< 0.0001	0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
1	10.8379	10.8379	0.8785	11.7791	16.7612	16.7322
	< 0.0001	0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
1.5	9.5642	9.5642	0.8448	10.1277	13.6568	13.7009
	< 0.0001	0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
2	7.9776	7.9776	0.7688	7.9776	9.5751	9.9665
	< 0.0001	0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
2.5	4.0893	4.0893	0.4689	4.0893	4.6568	4.6568
	0.0001	0.0001	0.0002	0.0001	0.0001	0.0001
3	0.5438	0.5438	0.0216	0.6659	0.6659	0.6659
	0.3756	0.5878	0.4634	0.3511	0.4711	0.4719

Table 6.2. Estimation of the maximum efficient dose: data and summary statistics for different multiple procedures; upper entry: test statistic, lower entry: p-value.

7. Summary and complements

This thesis has been inspired by the particular problem of assessing monotone dose-response relationships. This problem is frequently encountered in practice in the context of actively proofing a significant monotonous dependence of the response on increasing doses or treatments.

One of the most popular statistical approaches in this context is the trend test of Williams (1971, 1972). However, several disadvantages are inherent to this procedure and the research for its improvement led to the present work. In the course of the thesis we could observe several main features, which severely restrict the use of Williams' $\bar{t}$.

- The null distribution is difficult to compute, especially in the general unbalanced set-up. Quantiles and p-values are hardly available apart from the balanced case.
- As seen from Example 1.4. (pp. 29) the test statistic $\bar{t}$ does not incorporate sufficiently possible unequal replications. Due to its variance estimator, $\bar{t}$ is very sensitive against different allocations of sample sizes.
- Moreover, $\bar{t}$ has been shown to behave sometimes poor with regard to both size and power. Figure 4.2. demonstrated the conservative nature of $\bar{t}$, while the power study revealed a stronger dependence of Williams' test on the underlying dose-response shape than it is the case of its competitors.

Therefore, the aim of the thesis was to improve and extend Williams' $\bar{t}$ with respect to these restrictions. The method of choice in the present work was the application of the principle of multiple contrast tests. In fact, much research work has been paid for their development. Chapter 2 was devoted to their numerical availability and the computation of multivariate normal and *t*–probabilities. Results obtained there are essential for both theory and application of MCTs. By virtue of Sections 2.1. and 2.2. we are now able to evaluate MCTs under both H_0 and H_A – a distinguishing feature when compared to competing trend tests, where exact power calculations are available at most for small values of *k*. Consequently, power expressions in closed form of MCTs were derived for normal and dichotomous data. Their derivation is essential when conducting a post-hoc analysis, or, even more important, for

design and statistical planning of experiments. Sample size determinations are implicitly solved by virtue of these power formulas.

In Chapter 3 a solution for extending Williams' $\bar{t}$ to general unbalanced settings was proposed. By incorporating the ideas of Williams into an adequate definition of a contrast matrix, we succeeded in improving the $\bar{t}$–test in the sense of above discussion. Due to a different variance estimator we managed to overcome the power and size deficits of the original approach. But beside from extending Williams' test, attention has been given towards the derivation of two further MCTs under total order restriction. First we extended in analogy to $\bar{t}$ the modified Williams test described by Marcus (1976). Next we provided an isotonic contrast, which owns two main characteristics. On the one hand, the methodology of Abelson and Tukey (1963) was applied for each of the arising 2^k-1 contrasts, resulting in 'locally' optimal single contrasts. On the other side, a link was established to the well-known maximum likelihood estimators under order restriction by decomposing the global null hypothesis into smaller subsets, which correspond uniquely to the possible outcomes of the max-min formula (1.5). Particular applications for a further enhanced use of these MCTs were introduced in Chapters 5 and 6.

In Section 4.2. an extensive power and size study was conducted to compare the different trend tests which emerged in the course of the thesis. The likelihood ratio test of Bartholomew (1959, 1961) is well-known to yield the highest 'average' power among the trend tests available at present. This was also one of the main conclusions from the power study (see Section 4.3. for a brief discussion of the results). But despite this main advantage the LRT seems nevertheless to be of restricted use in practice. Several reasons may help to explain this phenomena. Agresti and Coull (1998, p. 148), for example, state that "*the area would be well-served by an applied version of the fine theoretical text book of Robertson et al. (1988)*." Similarly, Tang and Lin (1997) observed that "*although the development of the LRT seems quite complete, many practitioners might find the amount of details excessive.*" Moreover, the null distribution of the LRT was regarded long time computationally infeasible, especially in unbalanced settings. Up to now, many articles are published under this assumption and provide alternative methods for analysing trend situations (see for example Tang and Lin, 1997, Hothorn et al., 1997, Bailey, 1998, and McDermott, 1999). However, one major conclusion of the thesis is to reinforce that numerical methods for evaluating orthant

probabilities (and hence level probabilities) indeed exist. Based on the formulas of Sun (1988a, b) slightly extended and improved SAS/IML programs were provided in Subsections 2.1.2.3. and 2.1.3. Together with the randomised lattice rules introduced in Section 2.1.2.2. the practitioner is able to conduct the LRT under total order restriction without any restriction of the number of groups.

But from all these investigations new problems and questions come up. With Marcus' and the isotonic MCT two further competing contrast definitions were found, which perform better in the power study than Williams' approach, i.e. they depend less on the underlying dose-response shape while achieving reasonable good power values. In fact they even behaved practically identical to the benchmark LRT in many situations. The question naturally arises for a deeper comparison between the LRT and MCTs. Which of these two approaches should be preferred for analysing a real data situation?

As mentioned before, the LRT has the strong advantage of good power properties. Moreover, due to the books of Barlow et al. (1972) and Robertson et al. (1988) and many further articles in the literature the LRT is thoroughly investigated from a theoretical point of view. But as seen from the study in Chapter 4 the power advantages of the LRT are not as big as one could expect. The power differences are rarely greater than 1 or 2%, when compared to Marcus' or the isotonic contrast. Furthermore, the LRT has seen to be slightly more liberal than its competitors, in particular Williams original test, in the presence of strong violations of the classical ANOVA-assumptions (non-normal data, variance heterogeneity). Common to both approaches (LRT and MCTs) is their simple generalisation to other experimental set-ups (non-parametric analysis, higher factorial layouts, ...; see also the discussion provided below). This distinguishes both approaches from many other trend tests, which are still only available for the analysis of special settings. Whereas the LRT has a theoretical and mathematical sounded background, the MCT, on the other side, has the claim of being a somewhat unified approach among multiple tests, not necessarily under total order restriction. We have seen that most of the trend tests published are actually multiple contrast tests, which include the approaches of Hirotsu, Hayter, McDermott, Williams and Marcus. Therefore, the results obtained for arbitrary MCTs in the thesis are valid for all of these tests. In particular, we are able to evaluate these tests under both H_0 and H_A. This is a further major advantage of MCTs when compared to the LRT. For the latter test, power expressions in closed form are only

available for $k = 2$ or $k = 3$. Therefore, sample size determination and design of experiments can not be carried out with the LRT. Next, the construction of simultaneous confidence intervals for the LRT is still a question of research. Robertson et al. (1988) devoted only two pages on this important subject. In the case of MCTs, however, one-sided intervals for linear combinations of normal means with coverage probability $1 - \alpha$ are easily available with

$$\sum_{i=0}^{k} c_i \mu_i \geq \sum_{i=0}^{k} c_i \overline{X}_i - t_{q,\nu,1-\alpha,\boldsymbol{R}} s \sqrt{\sum_{i=0}^{k} c_i^2 / n_i} \,,$$

where $t_{q,\nu,1-\alpha,\boldsymbol{R}}$ is the corresponding q–variate t–quantile with ν degrees of freedom and correlation matrix $\boldsymbol{R}$ (see Subsection 2.2.2.3.). One further main advantage of MCTs over the LRT involves their numerical availability. Unfortunately most of the tests introduced and developed in the course of the thesis are not implemented at present in the statistical software packages (with exception of the balanced Williams). Algorithmic implementations, such as presented here for both the LRT and the MCTs, are hardly available to the practitioner and moreover difficult to understand. Hence, for a numerical method to be accepted by the practitioner it has to be widely available, easy to use and has to provide an intuitive insight. The SAS-procedure `PROC MULTTEST` (SAS Institute Inc., 1997, p. 777) meets all these requirements. The contrast statements within the procedure call allows the practitioner an easy

```
DATA seeding;
  INPUT yield rate @@;
  CARDS;
  25.4    50  22.4    50  25.2    50  24.4    50  24.2    50  22      50
  26.2    75  26.2    75  25.2    75  26.4    75  25      75  27.8    75
  27.6   100  27.6   100  26     100  25.8   100  26.2   100  25.8   100
  27.6   125  28.2   125  26.8   125  26.6   125  28     125  27.8   125
  27.2   150  28.2   150  26.8   150  25.6   150  27.2   150  27.6   150
  ;

PROC MULTTEST BOOT N=10000;
  CLASS rate;
  TEST MEAN (yield/UPPERTAILED);
  CONTRAST  '1'     -1      0      0      0     1;
  CONTRAST  '2'     -1      0      0     .5    .5;
  CONTRAST  '3'     -1      0  .3333  .3333 .3333;
  CONTRAST  '4'     -1    .25    .25    .25   .25;
  CONTRAST  '5'    -.5    -.5      0      0     1;
  CONTRAST  '6' -.3333 -.3333 -.3333      0     1;
  CONTRAST  '7'   -.25   -.25   -.25   -.25     1;
  CONTRAST  '8'     -1     -1      0      1     1;
  CONTRAST  '9'    -.5    -.5  .3333  .3333 .3333;
  CONTRAST '10' -.3333 -.3333 -.3333     .5    .5;
RUN;
```

MULTTEST PROCEDURE

Test for continuous variables:	Mean t-test
Tails for continuous tests:	Upper-tailed
Strata adjustment?	No
P-value adjustments:	Bootstrap
Center continuous variables?	Yes
Number of resamples:	10000
Seed:	69986

MULTTEST COEFFICIENTS

	Class				
Contrast	50	75	100	125	150
1	-1.00000	0.00000	0.00000	0.00000	1.00000
2	-1.00000	0.00000	0.00000	0.50000	0.50000
3	-1.00000	0.00000	0.33330	0.33330	0.33330
4	-1.00000	0.25000	0.25000	0.25000	0.25000
5	-0.50000	-0.50000	0.00000	0.00000	1.00000
6	-0.33330	-0.33330	-0.33330	0.00000	1.00000
7	-0.25000	-0.25000	-0.25000	-0.25000	1.00000
8	-1.00000	-1.00000	0.00000	1.00000	1.00000
9	-0.50000	-0.50000	0.33330	0.33330	0.33330
10	-0.33330	-0.33330	-0.33330	0.50000	0.50000

MULTTEST TABLES

		Class				
Variable	Statistic	50	75	100	125	150
YIELD	Mean	23.9333	26.1333	26.5000	27.5000	27.1000
	Std Dev	1.4236	1.0013	0.8649	0.6542	0.8741
	N	6.0000	6.0000	6.0000	6.0000	6.0000

Contrast	Raw_p	Boot_p
1	0.0001	0.0001
2	0.0001	0.0001
3	0.0001	0.0001
4	0.0001	0.0001
5	0.0002	0.0009
6	0.0013	0.0055
7	0.0126	0.0428
8	0.0001	0.0001
9	0.0001	0.0001
10	0.0001	0.0003

and flexible use of multiple contrasts, which are than approximated by a bootstrap or permutation method. The algorithm above illustrates the convenient use by analysing the seeding rate example of the Introduction. An uppertailed bootstrap test is conducted with

10,000 bootstrap replications. The contrast statements involve the definition of Marcus' contrast test ($q = k(k+1)/2 = 10$ single contrasts). The results are calculated quickly and are summarised below the algorithm. The final p-value is then the minimum value (`0.0001`) among the q = 10 adjusted bootstrap p-values towards the end of the output (marked bold for purposes of illustration here).

It is always difficult to summarise such a broad choice of subjects as investigated in this thesis. It becomes even more complicated if one is requested to provide recommendations to the practitioner. Nevertheless, the attempt is made in the following, although we are conscious of the inherent superficiality and shortcoming of detailed results.

- One should conduct a trend test only if one is sure upon the monotonicity assumption. Otherwise the results are not reliable and difficult to interpret.
- In general, the one-sided Dunnett-test is a good alternative to trend tests. It does not loose too much power under H_A, but it makes no monotonicity assumption and is therefore robust against violations of the monotonicity.
- Among the present trend tests no uniformly most powerful test exists. None of the introduced tests above can therefore be recommended to achieve the best power over the whole alternative space.
- The LRT seems to be that one with the highest average power.
- However, MCTs (especially those according to Marcus and the isotonic MCT) approximate the LRT very well, resulting in a power loss not greater than 1 or 2% over most parts of the alternative space. In many cases, MCTs perform identically to the LRT.
- Williams original test behaves especially good for concave profiles. But, as seen from the data example, it takes the sample size allocation not correctly into account.
- For non-normal data and variance heterogeneity the LRT is somewhat liberal. Williams $\bar{t}$ behaves best in such situations, MCTs depend on the contrast choice.
- Our final recommendation is to use MCTs, because their power loss is negligible in comparison to the LRT and they are easier to handle under a variety of aspects (confidence intervals, numerical implementation, power expression, generalisation to other set-ups).

The question arises upon further generalisations of MCTs. A few lines above we have already introduced for another reason a bootstrap MCT. Certainly, a bootstrap MCT is not only interesting because of its convenient use when applying SAS. If the underlying distribution is unknown, bootstrap methods are one applicable approach for analysing the observed data. For further reading we refer to standard text books, e.g. Davison and Hinkley (1997). For a short investigation of the performances, selected power values from Table 4.1. are compared with the corresponding simulated power of bootstrap MCTs. A SAS/IML program of Seidel (1999) has been used for these purposes, where the data are bootstrapped from the pooled sample, while assuming variance homogeneity. The simulations were conducted with standardised normal data generated by the SAS call `RANNOR`. The data in Table 7.1. (columns 4 and 5) suggest that the empirical bootstrap distribution indeed approximates the real one fairly well and the power values differ only negligible on the second or third significant digit.

	MCTs with mean square error		bootstrap MCTs with mean square error		bootstrap MCTs with total sum of error	
	minimum	*maximum*	*minimum*	*maximum*	*minimum*	*maximum*
Williams	0.4366	0.5647	0.4282	0.5528	0.4168	0.5556
Marcus	0.4951	0.5573	0.498	0.5534	0.4996	0.5654
LRT	0.4912	0.5651	0.491	0.5634	0.491	0.5634
Williams contrast	0.3826	0.547	0.3774	0.541	0.3578	0.5438
Marcus contrast	0.4888	0.5432	0.486	0.5384	0.4916	0.5564
Isotonic contrast	0.4886	0.5456	0.486	0.54	0.4904	0.5576

Table 7.1. Comparison of power values for different trend tests, $\nu = 20$ ($n_i = 6$), $\Delta = 2$, bootstrap tests with $5{,}000 \times 5{,}000$ replications (simulation $\times$ bootstrap replications).

Attention has been drawn several times in the course of this thesis that Cohen and Sackrowitz (1992, 1993) succeeded in improving several multiple tests by taking the total sum of errors instead of the usual mean square error. In particular, they have shown that studentised tests are inadmissible under specific order restrictions, i.e. they can be improved uniformly throughout the alternative by an appropriate test. However, improved tests were provided for some restricted examples and values of *k* only. In the light of their considerations, we included

the modified tests statistics in Table 7.1. Their power values were yielded with the same SAS-program cited above. The tests statistics under investigation were obtained by substituting the usual s^2 through the total sum of errors $\hat{\sigma}^2 = \sum_{i=0}^{k} \sum_{j=1}^{n_i} \left(X_{ij} - \overline{X} \right)^2 \Big/ \left(\sum_{i=0}^{k} n_i - 1 \right)$. Only the LRT has been left unchanged. It transpires from Table 7.1. that the differences, if any, are marginal. Due to the uncertainty inherent to the simulation approach, no final conclusion can be made upon $5{,}000 \times 5{,}000$ replications.

Up to now we have focused ourselves on situations under the simple (i.e. total) order assumption

$$H_A\colon \mu_0 \le \mu_1 \le \ldots \le \mu_k\,,\ \mu_0 < \mu_k\,.$$

However, other order restrictions are possible and frequently investigated in the literature. We cite a few further orderings which are closely related to above H_A. Grömping (1996) investigated in detail the incremental ordering

$$H_A'\colon \mu_0 < \mu_1 < \ldots < \mu_k$$

and proposed a test based on the comparison of adjacent doses within the intersection union principle (the null hypothesis is rejected only if the maximum p-value is smaller than the level α). Obviously, an adequate MCT would be defined by taking the minimum of $q = k$ adjacent pairwise contrasts. The whole theory developed in this thesis could be applied, including power expression, confidence intervals, numerical availability, etc. Next, we encountered already twice the slippage alternative

$$H_A''\colon \mu_0 = \ldots = \mu_i < \mu_{i+1} = \ldots = \mu_k\,, \quad 0 \le i \le k-1,$$

of exactly one shift in the response function. As seen, the approaches of Hirotsu (1979) and Sugiura (1994) belong to the class of MCTs, too. One further well-known order restriction is the classical many-to-one approach

$$H_A'''\colon \exists i\colon \mu_0 < \mu_i\,.$$

The most frequently used test for this set-up seems to be that of Dunnett (1955, 1964), which again is a MCT with a special contrast matrix. In this way many other restrictions of the alternative in the general *k*–sample problem can be found in the literature and one can imagine that for each situation adequate MCTs might be defined. Thus, the results of the present thesis are not only valid for the investigated total order, but generalise to other situations (see Robertson et al., 1988, or Hettmansperger and Norton, 1987, for further examples of order restrictions).

Further applications of MCTs are given in the area of non-parametric analysis. Bootstrap methods have already been mentioned briefly. Moreover, Seidel (1999) provides a detailed discussion of rank transformed statistics, including MCTs. Due to the multivariate central limit theorem the multivariate normal distribution holds asymptotically and the results achieved in Section 2.1. are fundamental for an enhanced use of rank transformed MCTs. Future research will also involve the inclusion of variance heterogeneity in MCTs. Several ad-hoc methods have already been proposed (see Grimes and Federer, 1984, Meng et al., 1993, and Bailey, 1998). But a systematic approach is still missing. Further applications are also expected to occur for higher factorial layouts. The present thesis has been devoted exclusively to the one-way layout. Seidel (1999) investigates also randomised block designs and general two-way layouts.

As a final conclusion we reinforce that the first steps for an enhanced application of MCTs were done, but much future research work is waiting to be conducted.

References

[Pages in the text on which the reference is cited are given in brackets at the end of the respective entry.]

Abelson, R.P. and Tukey, J.W. (1963) Efficient utilisation of non-numerical information in quantitative analysis: General theory and the case of simple order. *The Annals of Mathematical Statistics*, **34**, 1347-1369. [33, 81, 88, 89, 162]

Agresti, A. and Coull, B.A. (1998) An empirical comparison of inference using order-restricted and linear logit models for a binary response. C*ommunications in Statistics – Computation and Simulation*, **27**, 147-166. [5, 130, 162]

Akkerboom, J.C. (1990) *Testing problems with linear or angular inequality constraints.* Springer-Verlag, New York. [22]

Amaratunga, D. and Ge, N. (1998) Step-down trend tests for identifying the minimum effective dose. *Journal of Biopharmaceutical Statistics*, **8**, 151-162. [153]

Andrés, A.M. and Mato, A.S. (1996) Optimal correction for continuity and conditions for validity in the unconditional chi-squared test. *Computational Statistics and Data Analysis*, **21**, 609-626. Correction: **26** (1997), 235. [139, 142, 148]

Antonello, J.M., Clark, R.L. and Heyse, J.F. (1993) Application of the Tukey trend test procedure to assess developmental and reproductive toxicity. I. Measurement data. *Fundamentals in Applied Toxicology,* **21,** 52-58. [33]

Armitage, P. (1955) Tests for linear trends in proportions and frequencies. *Biometrics*, **11**, 375-386. [127, 129]

Ayer, M., Brunk, H.D., Ewing, G.M., Reid, W.T. and Silverman, E. (1955) An empirical distribution function for sampling with incomplete information. *The Annals of Mathematical Statistics*, **26**, 641-647. [14]

Bailey, S. (1998) Subchronic toxicity studies. In: *Design and analysis of animal studies in pharmaceutical development*, Chow, S.C. and Liu, J.P. (eds.), Marcel Dekker, New York, pp. 135-196. [79, 80, 162, 169]

Banno, H. and Yamagami, A. (1989) Food consumption and conversion efficiency of the larvae of *Eupromus ruber* (Dalman) (Coleoptera: Cerambycidae). *Applied Entomology and Zoology*, **24**, 174-179. [2, 29]

Baringhaus, L. (1998) *Personal communication.* [132]

Barlow, R.E., Bartholomew, D.J, Bremner, J.M. and Brunk, H.D. (1972) *Statistical inference under order restrictions.* Wiley, London. [15, 16, 130, 163]

Bartholomew, D.J. (1959) A test of homogeneity for ordered alternatives. *Biometrika*, **46,** 36-48. [4, 20, 130, 162]

Bartholomew, D.J. (1961) Ordered tests in the analysis of variance. *Biometrika*, **48,** 325-332. [4, 9, 11, 20, 162]

Bauer, P. (1997) A note on multiple testing procedures in dose finding. *Biometrics*, **53**, 1125-1128. [11, 153]

Bauer, P., Röhmel, J., Maurer, W. and Hothorn, L.A. (1998) Testing strategies in multidose trials including active control. *Statistics in Medicine*, **17**, 2133-2146. [1, 159]

Beckers, M. and Haegemans, A. (1992) Comparison of numerical integration techniques for multivariate normal integrals. Computer Science Department, Catholic University of Leuven, Belgium (preprint). [46]

Berger, J. (1991) Introduction. *Contemporary Mathematics*, **115**, 1-7. [75]

Bretz, F. (1999) *Berechnung der multivariaten Normal- und* t-*Verteilung.* Biometrische Beiträge der Universität Hannover, Band I, Hannover. [43, 60]

Bretz, F. and Hothorn, L.A. (1999) A powerful alternative to Williams test with application to toxicological dose-response relationships of normally distributed data. *Environmental and Ecological Statistics* (to appear). [5, 30, 32]

Bristol, D.R. (1993) One-sided multiple comparisons of response rates with a control. In: *Multiple comparisons, selection, and applications in biometry: A festschrift in honour of Charles W. Dunnett*, Hoppe, F.M. (ed.), Marcel Dekker, New York. [143]

Brunden, M.N. (1995) A review of isotonic regression theory - means from normal distributions. *Technical Report*, No. 9164-95-001, The Upjohn Company, Kalamazzo, Michigan. [53]

Brunk, H.D. (1955) Maximum likelihood estimates of monotone parameters. *The Annals of Mathematical Statistics*, **26**, 607-616. [11]

Capizzi, T., Survill, T.T., Heyse, J.F. and Malani, H. (1992) An empirical and simulated comparison of some tests for detecting progressiveness of response with increasing doses of a compound. *Biometrical Journal*, **34,** 275-289. [33]

Chen, J.J., Kodell, R.L. and Pearce, B.A. (1997) Significance levels of randomization trend tests in the event or rare ocurrences. *Biometrical Journal*, **39,** 327-337. [151]

Cheng, P.H. (1995) A new proof of the fundamental theorem in isotonic regression. *ASA Proceedings of the Biopharmaceutical Section*, 281-282. [15]

Childs, D.R. (1967) Reduction of the multivariate normal integral to characteristic form. *Biometrika*, **54**, 293-299. [49]

Chuang-Stein, C. and Agresti, A. (1997) A review of tests for detecting a monotone dose-response relationship with ordinal response data. *Statistics in Medicine,* **16**, 2599-2618. [34]

Cochran, W.G. (1954) Some methods for strengthening the common χ^2-test. *Biometrics*, **10,** 417-451. [127]

Cohen, A. and Sackrowitz, H.B. (1992) Improved tests for comparing treatments against a control and other one-sided problems. *Journal of the American Statistical Association,* **87**, 1137-1144. [20, 34, 167]

Cohen, A. and Sackrowitz, H.B. (1993) Inadmissibility of studentised tests for normal order restricted models. *The Annals of Statistics*, **21**, 746-752. [34, 167]

Collings, B.J., Margolin, B.H. and Oehlert, G.W. (1981) Analyses for binomial data, with application to the fluctuation test for mutagenicity. *Biometrics*, **37**, 775-794. [130]

Conaway, M., Pillers, C., Robertson, T. and Sconing, J. (1991) A circular cone test for testing homogeneity against a simple tree order. *The Canadian Journal of Statistics*, **19**, 283-296. [22]

Corcoran, C., Mehta, C.R. and Senchaudhuri, P. (1998) Computing the exact unconditional power for the Cochran-Armitage test for trend: A recommendation for carcinogenicity studies (preprint). [148]

Cornish, E.A. (1954) The multivariate *t*-distribution associated with a set of normal sample deviates. *Australian Journal of Physics*, **7**, 531-542. [55]

David, F.N. (1953) A note on the evaluation of the multivariate normal integral. *Biometrika*, **40**, 458-459. [48]

Davison, A.C. and Hinkley, D.V. (1997) *Bootstrap methods and their application*. Cambridge University Press, Cambridge. [167]

Deak, I. (1990) *Random number generators and simulation.* Akademiai Kiado, Budapest. [46, 65, 67]

Dunn, O.J. and Massey, F.J. (1965) Estimation of multiple contrasts using t distributions. *Journal of the American Statistical Association,* **60,** 573-583. [22, 82]

Dunnett, C.W. (1955) A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association,* **50,** 1096-1121. [10, 26, 28, 59, 74, 113, 136, 169]

Dunnett, C.W. (1964) New tables for multiple comparison with a control. *Biometrics*, **20**, 482-491. [10, 28, 169]

Dunnett, C.W. and Sobel, M. (1954) A bivariate generalisation of Student's distribution with tables for certain special cases. *Biometrika*, **41**, 153-169. [55]

Dunnett, C.W. and Tamhane, A.C. (1998) Some new multiple test procedures for dose finding. *Journal of Biopharmaceutical Statistics*, **8**, 353-366. [153]

Dwass, M. (1960) Some *k*-sample rank-order tests. In: *Contributions to Probability and Statistics*, Olkin, I. (ed.), Stanford University Press, pp. 198-202. [22, 82]

Edwards, D. and Berry, J.J. (1987) The efficiency of simulation-based multiple comparisons. *Biometrics*, **43**, 913-928. [60, 70, 71]

Fang, K.T. and Zhang, Y.T. (1990) *Generalised multivariate analysis*. Science Press, Beijing and Springer-Verlag, Berlin. [35, 58]

Fligner, M.A. and Wolfe, D.A. (1982) Distribution-free tests for comparing several treatments with a control. *Statistica Neelandica*, **36**, 119-127. [81]

Genz, A. (1992) Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, **1**, 141-150. [45, 46]

Genz, A. (1993) Comparison of methods for the computation of multivariate normal probabilities. *Computing Science and Statistics*, **25**, 400-405. [46]

Genz, A. and Bretz, F. (1999) Numerical computation of multivariate *t*–probabilities with application to power calculation of multiple contrasts. *Journal of Statistical Computation and Simulation* (to appear). [62, 63, 65, 101]

Genz, A. and Kwong, K.S. (1999) Numerical evaluation of singular multivariate normal distributions (preprint). [100]

Graubard, B.I. and Korn, E.L. (1987) Choice of column scores for testing independence in ordered 2x*k* contingency tables. *Biometrics*, **43**, 471-476. [129]

Grimes, B.A. and Federer, W.T. (1984) Comparison of means from populations with unequal variances. In: *W.G. Cochran's impact on statistics*, Rao, P.S.R.S. and Sedransk, J. (eds.), Wiley, New York, pp. 353-374. [169]

Grömping, U. (1996) *Tests for a monotone dose-response relation in models with ordered categorical dose with emphasis on likelihood ration tests for linear inequalities on normal means.* Ph.D. thesis, University of Dortmund. [168]

Gupta, S.S. (1963a) Probability integrals of multivariate normal and multivariate *t*. *The Annals of Mathematical Statistics*, **34**, 792-828. [43, 60]

Gupta, S.S. (1963b) Bibliography on the multivariate normal integrals and related topics. *The Annals of Mathematical Statistics*, **34**, 829-838. [43, 60]

Habegger, R. and Wiebe, H.J. (1985) Einfluß der Temperatur auf das Auftreten von Anomalien bei Kohlrabijungpflanzen. *Gartenbauwissenschaft*, **50**, 25-28. [126, 151]

Hayter, A.J. (1989) Pairwise comparisons of generally correlated means. *Journal of the American Statistical Association*, **84**, 208-213. [94]

Hayter, A.J. (1990) A one-sided studentised range test for testing against a simple ordered alternative. *Journal of the American Statistical Association*, **85**, 778-785. [34]

Hayter, A.J. and Liu, W. (1996) A note on the calculation of $\Pr\{X_1 < X_2 < \ldots < X_k\}$. *The American Statistician*, **50**, 365. [53]

Hayter, A.J., Kuriki, S. and Shimodaira, H. (1999a) The isotonic range statistic for testing against an ordered alternative (preprint). [19, 87]

Hayter, A.J., Miwa, T. and Liu, W. (1999b) Combining the advantages of one-sided and two-sided test procedures for comparing several treatments with a control (preprint). [34]

Hettmansperger, T.P. and Norton, R.M. (1987) Tests for patterned alternatives in *k*-sample problems. *Journal of the American Statistical Association*, **82**, 292-299. [169]

Hirji, K.F. and Tang, M.L. (1998) A comparison of tests for trend. *Communications in Statistics – Theory and Methods*, **27**, 943-963. [130, 140]

Hirotsu, C. (1979) The cumulative chi-squares method and a studentised maximal contrast method for testing an ordered alternative in a on-way analysis of variance model. *Reports of Statistical Application Research*, **26**, 12-21. [32, 80, 88, 168]

Hirotsu, C. (1982) Use of cumulative efficient scores for testing ordered alternatives in discrete models. *Biometrika*, **69**, 567-577. [32]

Hirotsu, C. (1997) Isotonic inference with particular interest in application to clinical trials. In: *Industrial Statistics*, Kitsos, C.P. and Edler, L. (eds.), Physica-Verlag, Heidelberg. [32, 80]

Hirotsu, C., Kuriki, S. and Hayter, A.J. (1992) Multiple comparison procedure based on the maximal component of the cumulative chi-squared statistic. *Biometrika*, **79**, 381-392. [33]

Hochberg, Y. and Tamhane, A.C. (1987) *Multiple comparison procedures.* Wiley, New York. [60, 113, 153]

Hoel, D. and Portier, C. (1994) Nonlinearity of dose-response functions for carcinogenicity. *Environmental Health Perspectives*, **102**, Supplement 1, 109-113. [129]

Hogg, R.V. (1965) On models and hypotheses with restricted alternatives. *Journal of the American Statistical Association*, **60**, 1153-1162. [28]

Horn, M. and Vollandt, R. (1995) *Multiple Tests und Auswahlverfahren.* Fischer Verlag, Jena. [154]

Hothorn, L.A. (1989) Robustness study on Williams and Shirley procedure, with application in toxicology. *Biometrical Journal*, **31,** 891-903. [6, 19]

Hothorn, L.A. (1995) Biostatistical analysis of the experimental design 'control versus k treatments' including a positive control group. In: *Biometrie in der chemisch-pharmazeutischen Industrie*, Vol. 6, Vollmar, J. (ed.), Fischer Verlag, Stuttgart. [1]

Hothorn, L.A., Neuhäuser, M. and Koch, H.F. (1997) Analysis of randomised dose-finding studies: closure test modifications based on multiple contrast tests. *Biometrical Journal*, **39,** 467-479. [82, 88, 153, 154, 162]

Hothorn, L.A. and Hauschke, D. (1998) Principles in statistical testing in randomised toxicological studies. In: *Design and analysis of animal studies in pharmaceutical development*, Chow, S.C. and Liu, J.P. (eds.), Marcel Dekker, New York, pp. 79-134. [78]

Hothorn, L.A. and Hauschke, D. (1999) Identifying the maximum safe dose: A multiple testing approach (preprint). [159]

House, D.E. (1986) A non-parametric version of Williams' test for a randomised block design. *Biometrics*, **42,** 187-190. [18]

Hsu, J.C. (1984) Constrained simultaneous confidence intervals for multiple comparisons with the best. *Annals of Statistics*, **12**, 1136-1144. [28]

Hsu, J.C. (1992) The factor analytic approach to simultaneous inference in the general linear model. *Journal of Computational and Graphical Statistics*, **1**, 151-168. [60]

Hsu, J.C. (1996) *Multiple comparisons*. Chapman and Hall, London. [60]

Hsu, J.C. and Nelson, B.L. (1990) Control variates for quantile estimation. *Management Science*, **36**, 835-951. [72]

Hsu, J.C. and Nelson, B.L. (1998) Multiple comparisons in the general linear model. *Journal of Computational and Graphical Statistics*, **7**, 23-41. [60, 70, 72]

Hu, X. (1998) An exact algorithm for projection onto a polyhedral cone. *Australian & New Zealand Journal of Statistics*, **40**, 165-170. [22]

Johnson, N.L. and Kotz, S. (1972) *Distributions in statistics: continuous multivariate distributions*. Wiley, New York. [35, 54]

Joe, H. (1995) Approximations to multivariate normal rectangle probabilities based on conditional expectations. *Journal of the American Statistical Association*, **90**, 957-964. [50]

Joe, S. (1990) Randomization of lattice rules for numerical multiple integration. *Journal of Computational and Applied Mathematics*, **31**, 299-304. [69]

Knoke, J.D. (1976) Multiple comparisons with dichotomous data. *Journal of the American Statistical Association*, **71,** 849-853. [22, 130]

Koch, H.F. (1996) *Teststatistiken für die 'many-to-one' Versuchsanlage im Falle dichotomer Ereignisse*. Ph.D. thesis, University of Hannover. [130, 143, 146, 148]

Korobov, N.M. (1960) Properties and calculation of optimal coefficients. *Soviet Mathematics – Doklady*, **1**, 696-700. [69]

Lancaster, H.O. (1961) Significance tests in discrete distributions. *Journal of the American Statistical Association*, **56,** 223-234. [151]

Lee, C.C. (1996) On estimation of monotone dose-response curves. *Journal of the American Statistical Association*, **91,** 1110-1119. [34]

Lee, R.E. and Spurrier, J.D. (1995) Successive comparisons between ordered treatments. *Journal of Statistical Planning and Inference*, **43**, 323-330. [80]

Lewis, E.L. and Mouw, J.T. (1978) *The use of contrast coefficients*. Southern Illinois University Press, Carbondale, USA. [79]

Marcus, R. (1976) The power of some tests of the equality of normal means against an ordered alternative. *Biometrika*, **63**, 177-183. [7, 9, 11, 19, 22, 87, 88, 162]

Marcus, R. and Peritz, E. (1976) Some simultaneous confidence bounds in normal models with restricted alternatives. *Journal of the Royal Statistical Society, Series B*, **38**, 157-165. [27, 28]

Marcus, R., Peritz, E. and Gabriel, K.B. (1976) On closed testing procedures with special reference to ordered analysis of variance. *Biometrika,* **63,** 655-660. [8, 153, 154]

Marsaglia, G. and Olkin, I. (1984) Generating correlation matrices. *SIAM Journal of Scientific and Statistical Computing*, **5**, 470-475. [74]

Martynov, G.V. (1981) Evaluation of the normal distribution function. *Journal of Soviet Mathematics*, **17**, 1857-1875. [43, 60]

Mathematica (1996) Mathematica Version 3.0, Wolfram Research Inc., Champaign, USA. [39]

McDermott, M.P. (1998) Generalised orthogonal contrast tests for homogeneity of ordered means. Talk held on the 'Joint Statistical Meeting', Section 55, Dallas. [6]

McDermott, M.P. (1999) Generalised orthogonal contrast tests for homogeneity of ordered means. *The Canadian Journal of Statistics* (preprint). [34, 162]

McDermott, M.P. and Mudholkar, G.S. (1993) A simple approach to testing homogeneity of order-constrained means. *Journal of the American Statistical Association,* **88,** 1371-1379. [34, 80]

Mead, R., Curnow, R.N. and Hasted, A.M. (1993) *Statistical methods in agriculture and experimental biology.* Chapman and Hall, London. [126]

Mehta, H., Capizzi, T. and Oppenheimer, L. (1984) Use of SAS software for trend and sequential trend test analysis. *Proceedings of the 9th SAS Users' Group International Conference,* 794-799. [22, 33]

Mehta, C.R., Patel, N.R. and Senchaudhuri, P. (1998) Exact power and sample size computations for the Cochran-Armitage trend test. *Biometrics*, **54**, 1615-1621. [148]

Meng, C.Y.K., Davis, S.B. and Roth, A.J. (1993) Robust contrast based trend tests. *ASA Proceedings of the Biopharmaceutical Section*, 127-132. [169]

Miller, KS. (1968) Some multivariate *t*-distributions. *The Annals of Mathematical Statistics*, **39**, 1605-1609. [55]

Miwa, T. and Hayter, A.J. (1999) Combining the advantages of one-sided and two-sided test procedures for comparing several treatments effects. *Journal of the American Statistical Association,* **94,** 302-308. [34]

Miwa, T., Hayter, A.J. and Wei, L. (1999) Calculations of level probabilities for normal random variables with unequal variances with applications to Bartholomew's test in unbalanced one-way models (preprint). [27, 28, 53]

Mount, K. (1999) A comparison of four tests for ordered proportions (preprint). [19]

Mudholkar, G. S. and McDermott, M.P. (1989) A class of tests for equality of ordered means. *Biometrika*, **76**, 161-168. [33]

Mukerjee, H., Robertson, T. and Wright, F.T. (1986) Multiple contrast tests for testing against a simple tree ordering. In: *Advances in order restricted statistical inference*, Dykstra, R (ed.), Springer-Verlag, Berlin, pp. 203-230. [6, 22, 82]

Mukerjee, H., Roberston, T. and Wright, F.T. (1987) Comparison of several treatments with a control using multiple contrasts. *Journal of the American Statistical Association,* **82,** 902-910. [6, 9, 22, 23, 81, 82]

Nam, J. (1987) A simple approximation for calculating sample size for detecting linear trends in proportions. *Biometrics,* **43**, 701-705. [128, 134]

National Archives and Record Service (1985) *Federal Register No. 50*, Vol. 50, Washington. [128]

Neuhäuser, M. (1996) *Trendtests bei a-priori unbekanntem Erwartungswertprofil*. Ph.D. thesis, University of Dortmund. [4, 8, 81, 131, 147]

Neuhäuser, M. and Hothorn, L.A. (1997) Trend tests for dichotomous endpoints with application to carcinogenicity studies. *Drug Information Journal,* **31**, 463-469. [130]

Neuhäuser, M. and Hothorn, L.A. (1998) An analogue of Jonckheere's trend test for parametric and dichotomous data. *Biometrical Journal,* **40**, 11-19. [130]

Niederreiter, H. (1992) *Random number generation and Quasi-Monte Carlo methods.* Society for Industrial and Applied Mathematics, Philadelphia. [68]

Oluyede, B.O. (1994) Tests for equality of several binomial populations against an order restricted alternative and model selection for one-dimensional multinomials. *Biometrical Journal*, **36**, 17-32. [130]

Pan, G. and Wolfe, D.A. (1996) Comparing groups with umbrella orderings. *Journal of the American Statistical Association,* **91,** 311- 317. [11]

Petersen, R.G. (1985) *Design and analysis of experiments*. Marcel Dekker, New York. [3]

Pincus, R. (1975) Testing linear hypotheses under restricted alternatives. *Mathematische Operationsforschung und Statistik*, **6**, 733-751. [22]

Portier, C. and Hoel, D. (1984) Type I error of trend tests in proportions and the design of cancer screens. *Communications in Statistics – Theory and Methods*, **13**, 1-14. [128]

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1992) *Numerical recipes in FORTRAN*. Cambridge University Press, Cambridge. [77]

Qian, S. (1994) Efficient programs for simulating chi-bar square distributions. *Computing Science and Statistics*, **26**, 232-236. [53]

Ralston, A. and Rabinowitz, P. (1978) *A first course in numerical analysis*. McGraw-Hill, New York. [71]

Remmenga, M.D., Milliken, G.A., Kratzer, D., Schwenke, J.R. and Rolka, H.R. (1997) Estimating the maximum effective dose in a quantitative dose-response experiment. *Journal of Animal Sciences*, **75**, 2174-2183. [158]

Ridders, C.J.F. (1979) A new algorithm for computing a single root or a real continuos function. *IEEE Transactions on Circuits and Systems*, **26**, 979-980. [77]

Robertson, T., Wright, F.T. and Dykstra, R.L. (1988) *Order restricted statistical inference*. Wiley, New York. [11, 14, 16, 20, 21, 22, 28, 29, 49, 52, 53, 84, 121, 130, 138, 162, 163, 164, 169]

Rom, D.M., Costello, R.J. and Connell, L.T. (1994) On closed test procedures for dose-response analysis. *Statistics in Medicine*, **13**, 1583-1596. [153]

Ruberg, S.J. (1989) Contrasts for identifying the minimum effective dose. *Journal of the American Statistical Association,* **84,** 816-822. [81]

Ruberg, S.J. (1995) Dose response studies II. Analysis and interpretation. *Journal of Biopharmaceutical Statistics*, **5**, 15-42. [153, 155, 156]

Sankey, S.S. and Weissfeld, L.A. (1998) A study of the effect of dichotomised ordinal data upon modeling. *Communications in Statistics – Simulation and Computation*, **27**, 871-887. [126]

SAS Institute Inc. (1997) *SAS/STAT Software: Changes and Enhancements through Release 6.12*. SAS Institute Inc., Cary, NC. [18, 84, 164]

Saville, D.J. and Wood, G.R. (1991) *Statistical Methods: The geometric approach.* Springer-Verlag, New York. [3, 4, 79, 80, 81, 88]

Schaafsma, W. and Smid, L.J. (1966) Most stringent somewhere most powerful tests against alternatives restricted by a number of linear inequalities. *The Annals of Mathematical Statistics*, **37**, 1161-1172. [81, 88]

Seidel, D. (1999) *Gerichtete Tests für geordnet kategoriale Daten bei sehr kleinen Fallzahlen.* Ph.D. thesis, University of Hannover. [53, 91, 151, 167, 169]

Shi, Y. and Meng, C. (1991) Bootstrap trend test procedures based on isotonic regression. *ASA Proceedings of the Biopharmaceutical Section*, 256-262. [53]

Shirley, E.A. (1977) A nonparametric equivalent of Williams' test for contrasting increasing dose levels of a treatment. *Biometrics*, **33,** 386- 389. [18]

Shirley, E.A. (1996) A literature review of statistical methods for the analysis of general toxicology data. In: *Statistics in Toxicology*, Morgan, B.J.T. (ed.), Clarendon Press, Oxford. [5, 34]

Sidik, K. and Morris, R.W. (1999) Nonparametric step-down test procedures for finding minimum effective dose. *Journal of Biopharmaceutical Statistics*, **9**, 217-240. [153]

Silvapulle, M.J. and Silvapulle, P. (1995) A score test against one-sided alternatives. *Journal of the American Statistical Association,* **90,** 342- 349. [34]

Simpson, D.G. and Margolin, B.H. (1990) Nonparametric testing for dose-response curves subject to downturns: Asymptotic power considerations. *The Annals of Statistics*, **18**, 373-390. [11]

Sloan, I.H. and Joe, S. (1994) *Lattice methods for multiple integration.* Clarendon Press, Oxford. [69]

Smythe, R.T., Krewski, D. and Dewanji, A. (1987) Robust tests for trend in binomial proportions. *Technical Report,* No. 70, Laboratory for Research in Statistics and Probability, Ontario, Canada. [130]

Solow, A.R. (1990) A method for approximating multivariate normal orthant probabilities. *Journal of Statistical Computation and Simulation*, **37**, 225-229. [44, 50]

Somerville, P. (1997) Multiple testing and simultaneous confidence intervals: calculation of constants. *Computational Statistics & Data Analysis*, **25**, 217-233. [28, 61, 94]

Somerville, P. (1998) Numerical computation of multivariate normal and multivariate *t*-probabilities over convex regions. *Journal of Computational and Graphical Statistics*, **7**, 529-544. [61]

Somerville, P. (1999) Critical values for multiple testing and comparisons: one step and step down procedures. *Journal of Statistical Planning and Inference* (to appear). [28, 61, 94]

Somerville, P. and Wang, M. (1994) Numerical computation of multivariate normal probabilities over convex regions. *Computing Science and Statistics*, **26**, 15-18. [61]

Steel, R.G.D. (1959) A multiple comparison rank sum test: Treatments versus control. *Biometrics*, **15**, 560-572. [10]

Storer, B.E. and Kim, C. (1990) Exact Properties of Some Exact Test Statistics for Comparing Two Binomial Proportions. *Journal of the American Statistical Association*, **85**, 146-155. [147, 148]

Sugiura, N. (1994) Approximating Bayes critical region for testing simple and tree ordered normal means. *Journal of Statistical Research,* **28**, 1-20. [82, 168]

Suissa, S. and Blais, L. (1995) Binary regression with continuous outcomes. *Statistics in Medicine*, **14**, 247-255. [126]

Sun, H.J. (1988a) A general reduction method for *n*–variate normal orthant probability. *Communications in Statistics – Theory and Methods*, **17**, 3913-3921. [49, 50, 163]

Sun, H.J. (1988b) A FORTRAN subroutine for computing normal orthant probabilities of dimensions up to nine. *Communications in Statistics – Simulation and Computation*, **17**, 1097-1111. [50, 163]

Sun, H.J. and Asano, C. (1989) Computing normal orthant probabilities with a Jacobi correlation matrix. *Computational Statistics Quarterly*, **1**, 27-41. [50]

Tamhane, A.C., Dunnett, C.W. and Hochberg, Y. (1996) Multiple test procedures for dose finding. *Biometrics,* **52,** 21-37. [17, 78, 85. 153]

Tang, D.I. and Lin, S.P. (1997) An approximate likelihood ratio test for comparing several treatments to a control. *Journal of the American Statistical Association,* **92,** 1155-1162. [5, 22, 162]

Tang, M.L., Hirji, K.F. and Vollset, S.E. (1995) Exact power computation for dose-response studies. *Statistics in Medicine*, **14**, 2261-2272. [148]

Tarone, R.E. and Gart, J.J. (1980) On the robustness of combined tests for trends in proportions. *Journal of the American Statistical Association*, **75**, 110-116. [128]

Thomas, D.G., Breslow, N. and Gart, J.J. (1977) Trend and homogeneity of proportions and life table data. *Computers and Biomedical Research*, **10**, 373-381. [139]

Tong, Y.L. (1990) *The multivariate normal distribution*. Springer, New York. [35, 36, 37, 41, 54, 58]

Tsai, K.T. and Chen, R.D. (1995) Robust Williams trend test. *ASA Proceedings of the Biopharmaceutical Section*, 65-70. [19]

Tukey, J.W. (1953) The problem of multiple comparisons. Unpublished manuscript reprinted in: *The Collected Works of John W. Tukey, Volume 8 (1994)*, Braun, H.I. (ed.), Chapman and Hall, New York. [28, 34]

Tukey, J.W., Ciminera, J.L. and Heyse, J.F. (1985) Testing the statistical certainty of a response to increasing doses of a drug. *Biometrics*, **41,** 295-301. [33]

Turnbull, T.A., Malani, H. and Capizzi, T.P. (1987) The power of some tests for comparing several doses of a compound with a zero dose control. *ASA Proceedings of the Biopharmaceutical Section*, 76-80. [22]

Wang, O. and Kennedy, W.J. (1997) Application of numerical interval analysis to contain self-validating results for multivariate probabilities in a massively parallel environment. *Statistics and Computing*, **7**, 163-171. [60]

Westfall, P.H. (1997) Multiple testing of general contrasts using logical constraints and correlations. *Journal of the American Statistical Association,* **92,** 299-306. [82, 88]

Westfall, P.H. and Lin, Y. (1988) Estimating optimal continuity corrections in run time. *ASA Proceedings of the Statistical Computing Section*, 297-298. [140]

Williams, D.A. (1971) A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics*, **27,** 103-117. Correction: **31** (1975), 1019. [5, 9, 11, 17, 18, 19, 30, 84, 88, 154, 157, 161]

Williams, D.A. (1972) The comparison of several dose levels with a zero dose control. *Biometrics*, **28,** 519-531. [5, 17, 18, 84, 113, 154, 157, 161]

Williams, D.A. (1986) A note on Shirley's non-parametric test for comparing several dose levels with a zero-dose control. *Biometrics*, **42,** 183-186. [18]

Williams, D.A. (1988) Tests for differences between several small proportions. *Applied Statistics*, **37**, 421-434. [19, 129, 147]

Wright, F.T. (1988) The one-way analysis of variance with ordered alternatives: a modification of Bartholomew's $\overline{E}^2$ test. *The Canadian Journal of Statistics*, **16**, 75-85. [22, 28, 53]

Yoshimura, I., Wakana, A. and Hamada, C. (1997) A performance comparison of maximum contrast methods to detect dose dependency. *Drug Information Journal,* **30**, 423-432. [84, 87]

Appendix A

The following pages contain the contrast coefficients in the balanced case of the three multiple contrast tests defined in the course of Chapter 3 (Sections 2 through 4). In the case of equal replications the sample size has no influence on the contrast definitions. The matrices were generated with programs B.3.1. through B.3.3. from Appendix B.

A.1. Contrast coefficients of Williams' MCT according to Definition 3.1., balanced case

k = 1

CM
-1
1

k = 2

CM		
-1	0	1
-1	0.5	0.5

k = 3

CM			
-1	0	0	1
-1	0	0.5	0.5
-1	0.3333333	0.3333333	0.3333333

k = 4

CM				
-1	0	0	0	1
-1	0	0	0.5	0.5
-1	0	0.3333333	0.3333333	0.3333333
-1	0.25	0.25	0.25	0.25

k = 5

CM					
-1	0	0	0	0	1
-1	0	0	0	0.5	0.5
-1	0	0	0.3333333	0.3333333	0.3333333
-1	0	0.25	0.25	0.25	0.25
-1	0.2	0.2	0.2	0.2	0.2

k = 6

CM						
-1	0	0	0	0	0	1
-1	0	0	0	0	0.5	0.5
-1	0	0	0	0.3333333	0.3333333	0.3333333
-1	0	0	0.25	0.25	0.25	0.25
-1	0	0.2	0.2	0.2	0.2	0.2
-1	0.1666667	0.1666667	0.1666667	0.1666667	0.1666667	0.1666667

A.2. Contrast coefficients of Marcus' MCT according to Definition 3.2., balanced case

k = 1

CM
-1
1

k = 2

CM		
-1	0.5	0.5
-1	0	1
-0.5	-0.5	1

k = 3

CM			
-1	0.3333333	0.3333333	0.3333333
-1	0	0.5	0.5
-0.5	-0.5	0.5	0.5
-1	0	0	1
-0.5	-0.5	0	1
-0.333333	-0.333333	-0.333333	1

k = 4

CM				
-1	0.25	0.25	0.25	0.25
-1	0	0.3333333	0.3333333	0.3333333
-0.5	-0.5	0.3333333	0.3333333	0.3333333
-1	0	0	0.5	0.5
-0.5	-0.5	0	0.5	0.5
-0.333333	-0.333333	-0.333333	0.5	0.5
-1	0	0	0	1
-0.5	-0.5	0	0	1
-0.333333	-0.333333	-0.333333	0	1
-0.25	-0.25	-0.25	-0.25	1

k = 5

CM					
-1	0.2	0.2	0.2	0.2	0.2
-1	0	0.25	0.25	0.25	0.25
-0.5	-0.5	0.25	0.25	0.25	0.25
-1	0	0	0.3333333	0.3333333	0.3333333
-0.5	-0.5	0	0.3333333	0.3333333	0.3333333
-0.333333	-0.333333	-0.333333	0.3333333	0.3333333	0.3333333
-1	0	0	0	0.5	0.5
-0.5	-0.5	0	0	0.5	0.5
-0.333333	-0.333333	-0.333333	0	0.5	0.5
-0.25	-0.25	-0.25	-0.25	0.5	0.5
-1	0	0	0	0	1
-0.5	-0.5	0	0	0	1
-0.333333	-0.333333	-0.333333	0	0	1
-0.25	-0.25	-0.25	-0.25	0	1
-0.2	-0.2	-0.2	-0.2	-0.2	1

k = 6

CM						
-1	0.1666667	0.1666667	0.1666667	0.1666667	0.1666667	0.1666667
-1	0	0.2	0.2	0.2	0.2	0.2
-0.5	-0.5	0.2	0.2	0.2	0.2	0.2
-1	0	0	0.25	0.25	0.25	0.25
-0.5	-0.5	0	0.25	0.25	0.25	0.25
-0.333333	-0.333333	-0.333333	0.25	0.25	0.25	0.25
-1	0	0	0	0.3333333	0.3333333	0.3333333
-0.5	-0.5	0	0	0.3333333	0.3333333	0.3333333
-0.333333	-0.333333	-0.333333	0	0.3333333	0.3333333	0.3333333
-0.25	-0.25	-0.25	-0.25	0.3333333	0.3333333	0.3333333
-1	0	0	0	0	0.5	0.5
-0.5	-0.5	0	0	0	0.5	0.5
-0.333333	-0.333333	-0.333333	0	0	0.5	0.5
-0.25	-0.25	-0.25	-0.25	0	0.5	0.5
-0.2	-0.2	-0.2	-0.2	-0.2	0.5	0.5
-1	0	0	0	0	0	1
-0.5	-0.5	0	0	0	0	1
-0.333333	-0.333333	-0.333333	0	0	0	1
-0.25	-0.25	-0.25	-0.25	0	0	1
-0.2	-0.2	-0.2	-0.2	-0.2	0	1
-0.166667	-0.166667	-0.166667	-0.166667	-0.166667	-0.166667	1

A.3. Contrast coefficients of the isotonic MCT according to Definition 3.3., balanced case

k = 1

CM	
-7.071068	7.0710678

k = 2

CM		
-4.082483	-4.082483	8.1649658
-8.164966	4.0824829	4.0824829
-8.164966	0	8.1649658

k = 3

CM			
-2.886751	-2.886751	-2.886751	8.660254
-5	-5	5	5
-5	-5	1.339746	8.660254
-8.660254	2.8867513	2.8867513	2.8867513
-8.660254	0	0	8.660254
-8.660254	-1.339746	5	5
-8.660254	-1.339746	1.339746	8.660254

k = 4

CM

-2.236068	-2.236068	-2.236068	-2.236068	8.9442719
-3.651484	-3.651484	-3.651484	5.4772256	5.4772256
-3.651484	-3.651484	-3.651484	2.0101792	8.9442719
-5.477226	-5.477226	3.6514837	3.6514837	3.6514837
-5.477226	-5.477226	1.0050896	1.0050896	8.9442719
-5.477226	-5.477226	0	5.4772256	5.4772256
-5.477226	-5.477226	0	2.0101792	8.9442719
-8.944272	2.236068	2.236068	2.236068	2.236068
-8.944272	-2.22E-16	-2.22E-16	-2.22E-16	8.9442719
-8.944272	-1.00509	-1.00509	5.4772256	5.4772256
-8.944272	-1.00509	-1.00509	2.0101792	8.9442719
-8.944272	-2.010179	3.6514837	3.6514837	3.6514837
-8.944272	-2.010179	1.0050896	1.0050896	8.9442719
-8.944272	-2.010179	0	5.4772256	5.4772256
-8.944272	-2.010179	0	2.0101792	8.9442719

k = 5

CM

-1.825742	-1.825742	-1.825742	-1.825742	-1.825742	9.1287093
-2.886751	-2.886751	-2.886751	-2.886751	5.7735027	5.7735027
-2.886751	-2.886751	-2.886751	-2.886751	2.4182961	9.1287093
-4.082483	-4.082483	-4.082483	4.0824829	4.0824829	4.0824829
-4.082483	-4.082483	-4.082483	1.5593697	1.5593697	9.1287093
-4.082483	-4.082483	-4.082483	0.7004433	5.7735027	5.7735027
-4.082483	-4.082483	-4.082483	0.7004433	2.4182961	9.1287093
-5.773503	-5.773503	2.8867513	2.8867513	2.8867513	2.8867513
-5.773503	-5.773503	0.8060987	0.8060987	0.8060987	9.1287093
-5.773503	-5.773503	0	0	5.7735027	5.7735027
-5.773503	-5.773503	0	0	2.4182961	9.1287093
-5.773503	-5.773503	-0.700443	4.0824829	4.0824829	4.0824829
-5.773503	-5.773503	-0.700443	1.5593697	1.5593697	9.1287093
-5.773503	-5.773503	-0.700443	0.7004433	5.7735027	5.7735027
-5.773503	-5.773503	-0.700443	0.7004433	2.4182961	9.1287093
-9.128709	1.8257419	1.8257419	1.8257419	1.8257419	1.8257419
-9.128709	0	0	0	0	9.1287093
-9.128709	-0.806099	-0.806099	-0.806099	5.7735027	5.7735027
-9.128709	-0.806099	-0.806099	-0.806099	2.4182961	9.1287093
-9.128709	-1.55937	-1.55937	4.0824829	4.0824829	4.0824829
-9.128709	-1.55937	-1.55937	1.5593697	1.5593697	9.1287093
-9.128709	-1.55937	-1.55937	0.7004433	5.7735027	5.7735027
-9.128709	-1.55937	-1.55937	0.7004433	2.4182961	9.1287093
-9.128709	-2.418296	2.8867513	2.8867513	2.8867513	2.8867513
-9.128709	-2.418296	0.8060987	0.8060987	0.8060987	9.1287093
-9.128709	-2.418296	0	0	5.7735027	5.7735027
-9.128709	-2.418296	0	0	2.4182961	9.1287093
-9.128709	-2.418296	-0.700443	4.0824829	4.0824829	4.0824829
-9.128709	-2.418296	-0.700443	1.5593697	1.5593697	9.1287093
-9.128709	-2.418296	-0.700443	0.7004433	5.7735027	5.7735027
-9.128709	-2.418296	-0.700443	0.7004433	2.4182961	9.1287093

$\underline{k = 6}$

CM

-1.543033	-1.543033	-1.543033	-1.543033	-1.543033	-1.543033	9.258201
-2.390457	-2.390457	-2.390457	-2.390457	-2.390457	5.976143	5.976143
-2.390457	-2.390457	-2.390457	-2.390457	-2.390457	2.6940851	9.258201
-3.273268	-3.273268	-3.273268	-3.273268	4.3643578	4.3643578	4.3643578
-3.273268	-3.273268	-3.273268	-3.273268	1.9174362	1.9174362	9.258201
-3.273268	-3.273268	-3.273268	-3.273268	1.1407873	5.976143	5.976143
-3.273268	-3.273268	-3.273268	-3.273268	1.1407873	2.6940851	9.258201
-4.364358	-4.364358	-4.364358	3.2732684	3.2732684	3.2732684	3.2732684
-4.364358	-4.364358	-4.364358	1.2782908	1.2782908	1.2782908	9.258201
-4.364358	-4.364358	-4.364358	0.5703937	0.5703937	5.976143	5.976143
-4.364358	-4.364358	-4.364358	0.5703937	0.5703937	2.6940851	9.258201
-4.364358	-4.364358	-4.364358	2.379E-16	4.3643578	4.3643578	4.3643578
-4.364358	-4.364358	-4.364358	2.379E-16	1.9174362	1.9174362	9.258201
-4.364358	-4.364358	-4.364358	2.379E-16	1.1407873	5.976143	5.976143
-4.364358	-4.364358	-4.364358	2.379E-16	1.1407873	2.6940851	9.258201
-5.976143	-5.976143	2.3904572	2.3904572	2.3904572	2.3904572	2.3904572
-5.976143	-5.976143	0.6735213	0.6735213	0.6735213	0.6735213	9.258201
-5.976143	-5.976143	0	0	0	5.976143	5.976143
-5.976143	-5.976143	0	0	0	2.6940851	9.258201
-5.976143	-5.976143	-0.570394	-0.570394	4.3643578	4.3643578	4.3643578
-5.976143	-5.976143	-0.570394	-0.570394	1.9174362	1.9174362	9.258201
-5.976143	-5.976143	-0.570394	-0.570394	1.1407873	5.976143	5.976143
-5.976143	-5.976143	-0.570394	-0.570394	1.1407873	2.6940851	9.258201
-5.976143	-5.976143	-1.140787	3.2732684	3.2732684	3.2732684	3.2732684
-5.976143	-5.976143	-1.140787	1.2782908	1.2782908	1.2782908	9.258201
-5.976143	-5.976143	-1.140787	0.5703937	0.5703937	5.976143	5.976143
-5.976143	-5.976143	-1.140787	0.5703937	0.5703937	2.6940851	9.258201
-5.976143	-5.976143	-1.140787	3.172E-16	4.3643578	4.3643578	4.3643578
-5.976143	-5.976143	-1.140787	3.172E-16	1.9174362	1.9174362	9.258201
-5.976143	-5.976143	-1.140787	3.172E-16	1.1407873	5.976143	5.976143
-5.976143	-5.976143	-1.140787	3.172E-16	1.1407873	2.6940851	9.258201
-9.258201	1.5430335	1.5430335	1.5430335	1.5430335	1.5430335	1.5430335
-9.258201	-1.59E-16	-1.59E-16	-1.59E-16	-1.59E-16	-1.59E-16	9.258201
-9.258201	-0.673521	-0.673521	-0.673521	-0.673521	5.976143	5.976143
-9.258201	-0.673521	-0.673521	-0.673521	-0.673521	2.6940851	9.258201
-9.258201	-1.278291	-1.278291	-1.278291	4.3643578	4.3643578	4.3643578
-9.258201	-1.278291	-1.278291	-1.278291	1.9174362	1.9174362	9.258201
-9.258201	-1.278291	-1.278291	-1.278291	1.1407873	5.976143	5.976143
-9.258201	-1.278291	-1.278291	-1.278291	1.1407873	2.6940851	9.258201
-9.258201	-1.917436	-1.917436	3.2732684	3.2732684	3.2732684	3.2732684
-9.258201	-1.917436	-1.917436	1.2782908	1.2782908	1.2782908	9.258201
-9.258201	-1.917436	-1.917436	0.5703937	0.5703937	5.976143	5.976143
-9.258201	-1.917436	-1.917436	0.5703937	0.5703937	2.6940851	9.258201
-9.258201	-1.917436	-1.917436	3.172E-16	4.3643578	4.3643578	4.3643578
-9.258201	-1.917436	-1.917436	3.172E-16	1.9174362	1.9174362	9.258201
-9.258201	-1.917436	-1.917436	3.172E-16	1.1407873	5.976143	5.976143
-9.258201	-1.917436	-1.917436	3.172E-16	1.1407873	2.6940851	9.258201
-9.258201	-2.694085	2.3904572	2.3904572	2.3904572	2.3904572	2.3904572
-9.258201	-2.694085	0.6735213	0.6735213	0.6735213	0.6735213	9.258201
-9.258201	-2.694085	0	0	0	5.976143	5.976143
-9.258201	-2.694085	0	0	0	2.6940851	9.258201
-9.258201	-2.694085	-0.570394	-0.570394	4.3643578	4.3643578	4.3643578
-9.258201	-2.694085	-0.570394	-0.570394	1.9174362	1.9174362	9.258201
-9.258201	-2.694085	-0.570394	-0.570394	1.1407873	5.976143	5.976143
-9.258201	-2.694085	-0.570394	-0.570394	1.1407873	2.6940851	9.258201
-9.258201	-2.694085	-1.140787	3.2732684	3.2732684	3.2732684	3.2732684
-9.258201	-2.694085	-1.140787	1.2782908	1.2782908	1.2782908	9.258201
-9.258201	-2.694085	-1.140787	0.5703937	0.5703937	5.976143	5.976143
-9.258201	-2.694085	-1.140787	0.5703937	0.5703937	2.6940851	9.258201
-9.258201	-2.694085	-1.140787	3.172E-16	4.3643578	4.3643578	4.3643578
-9.258201	-2.694085	-1.140787	3.172E-16	1.9174362	1.9174362	9.258201
-9.258201	-2.694085	-1.140787	3.172E-16	1.1407873	5.976143	5.976143
-9.258201	-2.694085	-1.140787	3.172E-16	1.1407873	2.6940851	9.258201

Appendix B

This appendix contains the codes and numerical implementations of many procedures used throughout the thesis. Most of the programs are provided in SAS/IML. Doubtless more efficient programming techniques in terms of computer time could have been used, but only at the expense of effort and research. For the sake of convenience the representations on the subsequent pages are provided instead, which are easier to understand and to follow. The listings contain material from the whole thesis. In particular, the following programs were included.

B.1.1. SAS/IML module for calculating the MLE acc. to equation (1.5)

B.1.2. SAS/IML module for simulating p-values of Williams' trend test

B.1.3. SAS/IML module for simulating p-values of Marcus' trend test

B.2.1. SAS/IML module for calculating mvn probabilities acc. to Solow (1990)

B.2.2. SAS/IML module for calculating mvn probabilities acc. to Genz (1992, 1993)

B.2.3. SAS/IML module for calculating orthant probabilities acc. to Sun (1988a, b)

B.2.4. SAS/IML module for calculating p-values of the LRT under simple order

B.2.5. SAS/IML module for calculating generators of good lattice vectors acc. to (2.15)

B.2.6. FORTRAN program for calculating generators of good lattice vectors acc. (2.15)

B.2.7. SAS/IML module for calculating mv*t*–probabilities (acceptance-rejection)

B.2.8. SAS/IML module for calculating mv*t*–probabilities (Monte Carlo method)

B.2.9. SAS/IML module for calculating mv*t*–probabilities (randomised lattice rule)

B.2.10. SAS/IML module for calculating mv*t*–quantiles using the bisection method

B.2.11. SAS/IML module for calculating mv*t*–quantiles using the regula falsi

B.2.12. SAS/IML module for calculating mv*t*–quantiles using the seacnt method

B.2.13. SAS/IML module for calculating mv*t*–quantiles acc. to Ridders (1979)

B.3.1. SAS/IML module for computing the entries of Williams' MCT acc. to Def. 3.1.

B.3.2. SAS/IML module for computing the entries of Marcus' MCT acc. to Def. 3.2.

B.3.3. SAS/IML module for computing the entries of the isotonic MCT acc. to Def. 3.3.

B.4.1. SAS/IML module for computing the power of arbitrary MCTs for normal means

B.5.1. SAS/IML module for computing the power of arbitrary MCTs for binomial data

The representation of the algorithms is divided into several parts. First, a short list contains the input and output variables. Moreover, information is given on the required design of the input variables (whether they are input as scalars, vectors or matrices). Afterwards, the algorithm itself follows, sometimes accompanied by explaining comments. An example call illustrates the use of each program, with the output given below. To avoid repeated descriptions of certain program parts which are required at several places, a modular representation of the programs has been chosen. Therefore, the given programs are not always sufficient on its own. Instead, additional modules described before are required to be included.

The practitioner should be able to apply the modules on his own. In fact, in most cases it is sufficient to include the statements

```
PROC IML;
```

and

```
QUIT;
```

at the beginning and to the end of each program. Sometimes additional modules are required and the user is requested to include them by himself from the preceding algorithms.

The programs were written and tested under Windows NT 4.0 using Release 6.12 of the SAS System. Therefore, the modules refer especially to this context. However, users should be able to replicate these activities on other computer and operating systems as well. Nuances of the SAS System peculiar to a particular operating system are covered in the SAS Companion manuals. If online help is available, these nuances also appear in the Table of Contents section, also titled ‘SAS Companion for ...’.

B.1.1. SAS/IML module for calculating the maximum likelihood estimators according to equation (1.5)

```
/************************************************************************/
/* Module for calculation of MLE according to (1.5)                     */
/*                                                                      */
/* Input:  x = mean vector (row vector)                                 */
/*         n = sample size vector (row vector)                          */
/*                                                                      */
/* Output: m = MLE                                                      */
/*                                                                      */
/************************************************************************/

/* If trend is downward, just revert the sign of x  */



start MLE(x,n);
  k=ncol(x);
  m=j(1,k,.);
  a=j(1,k,.);
  c=j(1,k,.);

  do i=1 to k;
    do u=1 to i;
      do v=i to k;
        a[v]=sum(x[u:v]#n[u:v])/sum(n[u:v]);
      end;
      c[u]=min(a);
      a=j(1,k,.);
    end;
    m[i]=max(c);
    c=j(1,k,.);
  end;
  return(m);
finish;



/* Example call */

x={25 13  2 15 14 21  9 33 25 15 21 25};
n={ 1  1  1  1  1  1  1  1  1  1  1  1};

m=MLE(x,n);
print m [format=5.2];


/* Output */

    M
13.33 13.33 13.33 14.50 14.50 15.00 15.00 23.50 23.50 23.50 23.50 25.00
```

B.1.2. SAS/IML module for simulating p-values of Williams' trend test

```
/************************************************************************/
/* Module for calculation of p_values of Williams' trend test           */
/*                                                                      */
/* Input:  data    = data matrix (treatment * replications)             */
/*         n       = sample size vector (row vector)                    */
/*         simanz  = simulation number (scalar)                         */
/*                                                                      */
/* Output: p_value = p_value of test                                    */
/*                                                                      */
/* Required modules: MLE                                                */
/*                                                                      */
/************************************************************************/

/* Module 'will_unb' is the main module, module 'simul' conducts the simulation */



start will_unb(data,n);
  k=ncol(n);
  nu=sum(n)-k;
  x=j(1,k,0);
  do i=1 to k;
    x[i]=sum(data[i,1:n[i]])/n[i];
  end;
  m=MLE(x,n);

  s=0;
  do i=1 to k;
    s=s+sum((data[i,1:n[i]]-x[i])##2);
  end;
  s=s/nu;

  t_obs=(m[k]-x[1])/sqrt(s/n[1]+s/n[k]);
  p=simul(n,t_obs);
  return(p);
finish;

start simul(n,t_obs) global(simanz);
  k=ncol(n);
  nu=sum(n)-k;
  maxn=max(n);
  mu=j(1,k,0);
  count=0;

  do index1=1 to simanz;
    data=rannor(j(k,maxn,141071))+repeat(t(mu),1,maxn);

    x=j(1,k,0);
    do i=1 to k;
      x[i]=sum(data[i,1:n[i]])/n[i];
    end;
    m=MLE(x,n);

    s=0;
    do i=1 to k;
      s=s+sum((data[i,1:n[i]]-x[i])##2);
    end;
    s=s/nu;

    t=(m[k]-x[1])/sqrt(s/n[1]+s/n[k]);
```

```
    if t_obs>t then count=count+1;
  end;

  p=1-count/(simanz+1);
  return(p);
finish;



/* Example call */

data={25.4 22.4 25.2 24.4 24.2 22.0,
      26.2 26.2 25.2 26.4 25.0 27.8,
      27.6 27.6 26.0 25.8 26.2 25.8,
      27.6 28.2 26.8 26.6 28.0 27.8,
      27.2 28.2 26.8 25.6 27.2 27.6};
n={6 6 6 6 6};
simanz=9999;

p_value=will_unb(data,n);
print p_value [format=6.4];


/* Output */

P_VALUE
 0.0000
```

B.1.3. SAS/IML module for simulating p-values of Marcus' trend test

```
/************************************************************************/
/* Module for calculation of p_values of Marcus' trend test             */
/*                                                                      */
/* Input:  data    = data matrix (treatment * replications)             */
/*         n       = sample size vector (row vector)                    */
/*         simanz  = simulation number (scalar)                         */
/*                                                                      */
/* Output: p_value = p_value of test                                    */
/*                                                                      */
/* Required modules: MLE                                                */
/*                                                                      */
/************************************************************************/

/* Module 'marc_unb' is the main module, module 'simul' conducts the simulation */



start marc_unb(data,n);
  k=ncol(n);
  nu=sum(n)-k;
  x=j(1,k,0);
  do i=1 to k;
    x[i]=sum(data[i,1:n[i]])/n[i];
  end;
  m=MLE(x,n);

  s=0;
  do i=1 to k;
    s=s+sum((data[i,1:n[i]]-x[i])##2);
  end;
  s=s/nu;

  t_obs=(m[k]-m[1])/sqrt(s/n[1]+s/n[k]);
  p=simul(n,t_obs);
  return(p);
finish;

start simul(n,t_obs) global(simanz);
  k=ncol(n);
  nu=sum(n)-k;
  maxn=max(n);
  mu=j(1,k,0);
  count=0;

  do index1=1 to simanz;
    data=rannor(j(k,maxn,141071))+repeat(t(mu),1,maxn);

    x=j(1,k,0);
    do i=1 to k;
      x[i]=sum(data[i,1:n[i]])/n[i];
    end;
    m=MLE(x,n);

    s=0;
    do i=1 to k;
      s=s+sum((data[i,1:n[i]]-x[i])##2);
    end;
    s=s/nu;

    t=(m[k]-m[1])/sqrt(s/n[1]+s/n[k]);
    if t_obs>t then count=count+1;
```

```
  end;

  p=1-count/(simanz+1);
  return(p);
finish;



/* Example call */

data={25.4 22.4 25.2 24.4 24.2 22.0,
      26.2 26.2 25.2 26.4 25.0 27.8,
      27.6 27.6 26.0 25.8 26.2 25.8,
      27.6 28.2 26.8 26.6 28.0 27.8,
      27.2 28.2 26.8 25.6 27.2 27.6};
n={6 6 6 6 6};
simanz=9999;

p_value=marc_unb(data,n);
print p_value [format=6.4];


/* Output */

P_VALUE
 0.0000
```

B.2.1. SAS/IML module for calculating multivariate normal probabilities according to Solow (1990) for arbitrary dimension

```
/************************************************************************/
/* Module for calculation of mvn probabilities according to Solow       */
/*                                                                      */
/* Input:  z = upper integration bound (scalar)                         */
/*         R = correlation matrix                                       */
/*                                                                      */
/* Output: x = probability                                              */
/*                                                                      */
/************************************************************************/

/* The program here is restricted to one-sided upper equicoordinate mvn probabilities */



start solow(z,R);
  dim=ncol(R);
  D=probbnrm(z,z,round(R,1E-09))-probnorm(z)**2;
  sumi=1;
  do i=1 to dim-1;
    b=inv(D[i+1:dim,i+1:dim]+I(dim-i)*1E-09)*D[i,i+1:dim]`;
    sumj=sum(b[1:dim-i]#(1-probnorm(z)));
    sumi=sumi#(probnorm(z)+sumj);
  end;
  return(sumi#probnorm(z));
finish;



/* Example call */

R={        1 0.7071068         0         0,
   0.7071068         1       0.5         0,
           0       0.5         1 0.3333333,
           0         0 0.3333333         1};
z=1;

x=solow(z,r);
print x [format=6.4];


/* Output */

     X
0.5783
```

B.2.2. SAS/IML module for calculating multivariate normal probabilities according to Genz (1992, 1993) for arbitrary upper integration bounds and dimension $q < 33$

```
/*************************************************************************/
/* Module for calculation of mvn probabilities according to Genz         */
/*                                                                       */
/* Input:  z     = upper integration bound (row vector)                  */
/*         R     = correlation matrix                                    */
/*         eps   = accuracy                                              */
/*                                                                       */
/* Output: prob  = probability                                           */
/*         error = simulated standard error                              */
/*                                                                       */
/*************************************************************************/

/* The program here is restricted to one-sided upper mvn probabilities */


start mvn(R,b,eps);
  q=ncol(b);
  c=t(root(r))+1E-12;

  y=j(1,q,0);
  e=y;
  e[1]=probnorm(b[1]/c[1,1]);
  n=10;
  vec=0:q-2;

  p_vector={157 313 619 1249 2503 5003 10007 20011};

  mat={ 1   1   1   1   1    1    1     1,
       46 119 239 512 672 1850 3822  6103,
       46  93 178 136 652 1476 2325  2894,
       17  51  73 197 792  792 1206  8455,
       18  51 104 165 792  380 1927  3629,
       18  80 102 175 253  162 2286  1752,
       11  70 161 303 306  363  343  1920,
       11  70 161 155 153  137  378   652,
       11  93 106  18 288  186   81   146,
       36  62  57  27 288  186  182   156,
       36  15  57  27  29   33   76   136,
       36  15  36  24 128   36   21    44,
        4  19  22  24  64   38   21    31,
       30  15  22  24  16   36   21   161,
       31   9  22  24  16   48   20   161,
       31   9  22  14  16   48   21    11,
        6  20   6  14  16   12   21    11,
        6   9   6  14  64   12   11    13,
        3   9   6  14  16   12   11    13,
        3   9   6   8  16    6   11    13,
        3  16   6   8  16    6    7    13,
        3  16   6   8  16    6    7    22,
        3  16   6   8  16    6    7    13,
        3  16   6   8   8    6    7    13,
        3  16   6   8   8    6    4    13,
        3  16   6   8   8    5    4    16,
        3  16   4   3   8    5    4    16,
        3   4   4   3   8    4    4    13,
        3   4   4   3   8    4    4    13,
        3   4   4   3   8    4    4     8,
        3   4   4   3   8    4    4     8};
```

```
  do until(n>50 | error<eps);
    index=1;
    do until(index=9 | error<eps);
      p=p_vector[index];
      h=mat[q-1,index];
      z=mod(j(1,q-1,h)##vec,p);

      intval=0;
      varsum=0;

      do l=1 to n;
        latsum=0;
        rr=ranuni(j(1,q-1,141071));

        do j=1 to p;
          w=abs(2*mod(rr+j#z/p,1)-1);

          do i=2 to q;
            y[i-1]=probit(w[i-1]#e[i-1]+1E-12);
            e[i]=probnorm((b[i]-sum(c[i,1:i-1]*y[1:i-1]))/c[i,i]);
          end;

          f=e[#];
          latsum=latsum+(f-latsum)/j;
        end;

        varsum=varsum+(l-1)*(latsum-intval)**2/l;
        intval=intval+(latsum-intval)/l;
      end;

      error=3*sqrt(varsum/(n*(n-1)));
      index=index+1;
    end;

  n=n+2;
  end;
  prob=intval;
  print prob [format=6.4] error [format=6.4];
finish;



/* Example call */

R={        1 0.7071068         0         0,
   0.7071068         1       0.5         0,
           0       0.5         1 0.3333333,
           0         0 0.3333333         1};
z={1 1 1 1};
eps=0.0001;

run mvn(R,z,eps);


/* Output */

  PROB  ERROR
0.5831 0.0000
```

B.2.3. SAS/IML module for calculating orthant probabilities according to Sun (1988a, b) for dimension $q < 12$

```
/************************************************************************/
/* Module for calculation of orthant probabilities according to Sun     */
/*                                                                      */
/* Input:  corr = entries of diagonal line adjacent to the main diagonal */
/*                of the correlation matrix (row vector)                */
/*                                                                      */
/* Output: prob = probability                                           */
/*                                                                      */
/************************************************************************/

/* Calculation of orthant probabilities with Jacobi correlation matrix */



start orthant(corr);
  n=ncol(corr)+1;
  r=corr;

  coef=j(1,7,.);
  pai=3.141592653589793238;

  do i=1 to int(n/2)+1;
    coef[i]=1/(2**(n-i+1)*pai**(i-1));
  end;

  if n=1 then pr=coef[1];
  else if n=2  then pr=coef[1]+coef[2]*arsin(r[1]);
  else if n=3  then pr=coef[1]+coef[2]*(arsin(r[1])+arsin(r[2]));
  else if n=4  then pr=coef[1]+coef[2]*sasin(r,4) +coef[3]*fint(r,4);
  else if n=5  then pr=coef[1]+coef[2]*sasin(r,5) +coef[3]*si4(r,5);
  else if n=6  then pr=coef[1]+coef[2]*sasin(r,6) +coef[3]*si4(r,6)
                       +coef[4]*fint(r,6);
  else if n=7  then pr=coef[1]+coef[2]*sasin(r,7) +coef[3]*si4(r,7)
                       +coef[4]*si6(r,7);
  else if n=8  then pr=coef[1]+coef[2]*sasin(r,8) +coef[3]*si4(r,8)
                       +coef[4]*si6(r,8) +coef[5]*fint(r,8);
  else if n=9  then pr=coef[1]+coef[2]*sasin(r,9) +coef[3]*si4(r,9)
                       +coef[4]*si6(r,9) +coef[5]*si8(r,9);
  else if n=10 then pr=coef[1]+coef[2]*sasin(r,10)+coef[3]*si4(r,10)
                   +coef[4]*si6(r,10)+coef[5]*si8(r,10)+coef[6]*fint(r,10);
  else if n=11 then pr=coef[1]+coef[2]*sasin(r,11)+coef[3]*si4(r,11)
                   +coef[4]*si6(r,11)+coef[5]*si8(r,11)+coef[6]*si10(r,11);

  return(pr);
finish;



start sasin(r,n);
  sasin=0;
  do i=1 to n-1;
    sasin=sasin+arsin(r[i]);
  end;
  return(sasin);
finish;


start si4(r,n);
```

```
  si4=0;
  r1=dis(r,8,3,n-3);
  do i=1 to n-3;
    p=tran(r1,8,3,i);
    si4=si4+fint(p,4);
  end;
  do i=1 to n-4;
    sum=0;
    do j=i+3 to n-1;
      sum=sum+arsin(r[j]);
    end;
    si4=si4+arsin(r[i])*sum;
  end;
  return(si4);
finish;


start si6(r,n);
  si6=0;
  r1=dis(r,8,3,n-3);
  r2=dis(r,6,5,n-5);
  do i=1 to n-5;
    p=tran(r2,6,5,i);
    si6=si6+fint(p,6);
  end;
  do i=1 to n-6;
    sum=0;
    do j=i+5 to n-1;
      sum=sum+arsin(r[j]);
    end;
    p=tran(r1,8,3,i);
    si6=si6+fint(p,4)*sum;
  end;
  do i=4 to n-3;
    sum=0;
    do j=1 to i-3;
      sum=sum+arsin(r[j]);
    end;
    p=tran(r1,8,3,i);
    si6=si6+fint(p,4)*sum;
  end;
  do i=1 to n-7;
    sum=0;
    do j=i+3 to n-4;
      sum1=0;
      do k=j+3 to n-1;
        sum1=sum1+arsin(r[k]);
      end;
      sum=sum+arsin(r[j])*sum1;
    end;
    si6=si6+arsin(r[i])*sum;
  end;
  return(si6);
finish;


start si8(r,n);
  si8=0;
  r1=dis(r,8,3,n-3);
  r2=dis(r,6,5,n-5);
  r3=dis(r,4,7,n-7);

  do i=1 to n-7;
```

```
  p=tran(r3,4,7,i);
  si8=si8+fint(p,8);
end;
do i=1 to n-8;
  sum=0;
  do j=i+7 to n-1;
    sum=sum+arsin(r[j]);
  end;
  p=tran(r2,6,5,i);
  si8=si8+fint(p,6)*sum;
end;
do i=4 to n-5;
  sum=0;
  do j=1 to i-3;
    sum=sum+arsin(r[j]);
  end;
  p=tran(r2,6,5,i);
  si8=si8+fint(p,6)*sum;
end;
do i=1 to n-8;
  sum=0;
  do j=i+5 to n-3;
    p=tran(r1,8,3,j);
    sum=sum+fint(p,4);
  end;
  p=tran(r1,8,3,i);
  si8=si8+fint(p,4)*sum;
end;
do i=1 to n-9;
  sum=0;
  do j=i+5 to n-4;
    sum1=0;
    do k=j+3 to n-1;
      sum1=sum1+arsin(r[k]);
    end;
    sum=sum+arsin(r[j])*sum1;
  end;
  p=tran(r1,8,3,i);
  si8=si8+fint(p,4)*sum;
end;
do i=7 to n-3;
  sum=0;
  do j=1 to i-6;
    sum1=0;
    do k=j+3 to i-3;
      sum1=sum1+arsin(r[k]);
    end;
    sum=sum+arsin(r[j])*sum1;
  end;
  p=tran(r1,8,3,i);
  si8=si8+fint(p,4)*sum;
end;
sum=0;
do i=9 to n-1;
  sum=sum+arsin(r[i]);
end;
p=tran(r1,8,3,4);
si8=si8+arsin(r[1])*fint(p,4)*sum;
if n>10 then do;
  sum=0;
  do i=1 to n-9;
    sum=sum+arsin(r[i]);
  end;
```

```
    p=tran(r1,8,3,5);
    si8=si8+arsin(r[10])*fint(p,4)*sum;
    si8=si8+arsin(r[1])*arsin(r[4])*arsin(r[7])*arsin(r[10]);
  end;
  return(si8);
finish;


start si10(r,n);
  si10=0;
  r1=dis(r,8,3,n-3);
  r2=dis(r,6,5,n-5);
  r3=dis(r,4,7,n-7);
  r4=dis(r,2,9,n-9);

  do i=1 to n-9;
    p=tran(r4,2,9,i);
    si10=si10+fint(p,10);
  end;
  p=tran(r3,4,7,4);
  q=tran(r3,4,7,1);
  si10=si10+arsin(r[1])*fint(p,8)+arsin(r[10])*fint(q,8);
  p=tran(r1,8,3,1);
  q=tran(r2,6,5,6);
  si10=si10+fint(p,4)*fint(q,6);
  p=tran(r1,8,3,8);
  q=tran(r2,6,5,1);
  si10=si10+fint(p,4)*fint(q,6);
  return(si10);
finish;


start fint(r,m);
  a=j(1,19,.);
  b=j(1,11,.);
  c=j(1,7,.);
  p=j(1,20,.);
  gl=j(2,12,.);

  iint=12;
  g1={-0.981560634246719251 -0.904117256370474857 -0.769902674194304687
      -0.587317954286617447 -0.367831498998180194 -0.125233408511463915,
       0.047175336386511827 2 0.106939325995318431  0.160078328543346226
       0.203167426723065922  0.233492536538354809  0.249147045813402785};

  do i=1 to iint/2;
    gl[1,i]=0.5*(g1[1,i]+1);
    gl[1,iint+1-i]=0.5*(-g1[1,i]+1);
    gl[2,i]=0.5*g1[2,i];
    gl[2,iint+1-i]=gl[2,i];
  end;

  i=1;
  do j=1 to m+m-3 by 2;
    p[j]=1;
    p[j+1]=r[i];
    i=i+1;
  end;
  p[m+m-1]=1;

  sum1=0;
  k1=0;
  krit=0;
```

```
do until(k1=iint | krit=1);
  k1=k1+1;
  if m<=8 then do;
    do j=1 to 15;
      a[j]=p[j];
    end;
  end;
  else do;
    ta=1-(p[2]*gl[1,k1])**2;
    a[1]=ta-p[4]*p[4];
    do j=2 to 18 by 2;
      a[j]=p[j+2]*ta;
      a[j+1]=ta;
    end;
  end;

  sum2=0;
  k2=0;
  do until(k2=iint | krit=1);
    k2=k2+1;
    if m<=6 then do;
      do j=1 to 11;
        b[j]=p[j];
      end;
    end;
    else do;
      ub=a[1]*a[3]-(a[2]*gl[1,k2])**2;
      ub1=a[5]*ub;
      b[1]=ub1-a[1]*a[4]*a[4];
      do j=2 to 10 by 2;
        b[j]=a[j+2]*ub;
        b[j+1]=ub1;
      end;
    end;

    sum3=0;
    k3=0;
    do until(k3=iint | krit=1);
      k3=k3+1;
      if m=4 then do;
        do j=1 to 7;
          c[j]=p[j];
        end;
      end;
      else do;
        vc=b[1]*b[3]-(b[2]*gl[1,k3])**2;
        vc1=b[5]*vc;
        c[1]=vc1-b[1]*b[4]*b[4];
        do j=2 to 6 by 2;
          c[j]=b[j+2]*vc;
          c[j+1]=vc1;
        end;
      end;

      sum4=0;
      do k4=1 to iint;
        wd=c[1]*c[3]-(c[2]*gl[1,k4])**2;
        if wd<0 then wd=1E-09;
        wd1=c[5]*wd;
        d11=wd1-c[1]*c[4]*c[4];
        d12=c[6]*wd;
        d22=wd1;
        yyy=d11*d22;
```

```
          if yyy<0 then yyy=1E-09;
          xxx=d12/sqrt(yyy);
          if xxx<-1 then xxx=-1+1E-09;
          sum4=sum4+gl[2,k4]/sqrt(wd)*arsin(xxx);
        end;
        sum4=c[2]*sum4;
        if m^=4 then sum3=sum3+gl[2,k3]/sqrt(vc)*sum4;
        else krit=1;
      end;
      if m^=4 then do;
        sum3=b[2]*sum3;
        if m^=6 then sum2=sum2+gl[2,k2]/sqrt(ub)*sum3;
        else krit=1;
      end;
    end;
    if (m^=4 & m^=6) then do;
      sum2=a[2]*sum2;
      if m^=8 then sum1=sum1+gl[2,k1]/sqrt(ta)*sum2;
      else krit=1;
    end;
  end;

  if m=4 then return(sum4);
  else if m=6 then return(sum3);
  else if m=8 then return(sum2);
  else return(p[2]*sum1);
finish;


start dis(r,m,n,k);
  rr=j(k,n,.);
  do i=1 to k;
    do j=1 to n;
      rr[i,j]=r[i-1+j];
    end;
  end;
  return(rr);
finish;


start tran(rr,m,n,k);
  r=j(1,n,.);
  do i=1 to n;
    r[i]=rr[k,i];
  end;
  return(r);
finish;



/* Example call */

corr={0.7071068 0.5 0.3333333};

prob=orthant(corr);
print prob [format=6.4];


/* Output */

  PROB
0.1364
```

B.2.4. SAS/IML module for calculating p-values of the likelihood ratio test under simple order according to Bartholomew (1959, 1961) without restriction of the number *k* of treatment groups and sample size allocation

```
/************************************************************************/
/* Module for calculation of p_values of LRT under simple order         */
/*                                                                      */
/* Input:  xbar    = group means (row vector)                            */
/*         n       = sample size vector (row vector)                    */
/*         s       = variance (scalar)                                  */
/*                                                                      */
/* Output: p_value = p_value of test                                    */
/*                                                                      */
/* Required modules: MLE                                                */
/*                   Sun or Genz_mvn                                    */
/*                                                                      */
/************************************************************************/

/* Module 'plkwunba' calculates the level probabilities according to
                     Lemma 2.12.; code adapted from Seidel (1999);
   module 'LRT'      calculates p-values                              */



start plkwunba(gewicht,p_lkw);
  k=ncol(gewicht);
  p_1_k=j(k-1,k-1,1);
  p_1_k[2,]=j(1,k-1,0.5);
  p_lkw=j(k,1,0);

  do t=3 to k;
    do i=2 to t;
      mengen=0;
      anz=erzall(t,i, mengen);
      nrowmen=nrow(mengen);

      if t<k then index=j(nrowmen,t,1)||j(nrowmen,k-t,0);
      else index=j(nrowmen,k,1);

      if i>1 then do;
        do s=1 to nrowmen;
          gew=j(1,i,0); top=0;
          prod=1;
          do u=1 to i;
            gew[u]=sum(gewicht[top+1:top+mengen[s,u]]);
            if mengen[s,u]=2 then prod=prod*0.5;
            else if mengen[s,u]>2 then prod=prod*p_1_k[mengen[s,u],top+1];
            top=top+mengen[s,u];
          end;
          if i=2 then z=0.5;
          else z=orthant(i,gew);
          if s=1 then y=z;
          else y=y//z;

          if t<k then p_1_k[t,1]=p_1_k[t,1]-z*prod;
          else p_lkw[i]=p_lkw[i]+z*prod;
        end;
      end;

      gewi=gewicht[,1:k];
```

```
      do j=2 to k-t+1;
        index=j(nrowmen,1,0)||index[,1:k-1];
        gewi=gewi[,2:ncol(gewi)];
        mengen1=mengen||index;
        if i>1 then do;
          do s=1 to nrowmen;
            gew=j(1,i,0);
            top=0;
            prod=1;
            do u=1 to i;
              gew[u]=sum(gewi[top+1:top+mengen[s,u]]);
              if mengen[s,u]=2 then prod=prod*0.5;
              else if mengen[s,u]>2 then
                prod=prod*p_1_k[mengen[s,u],top+j-1+1];
              top=top+mengen[s,u];
            end;
            if i=2 then z=0.5;
            else z=orthant(i,gew);
            if s=1 then y=z;
            else y=y//z;
            p_1_k[t,j]=p_1_k[t,j]-z*prod;
          end;
        end;
      end;
    end;
  end;
finish;


start erzall(n,k,mengen);
  feld1=j(1,k,0);
  first=0;
  mtc1=0;
  t=0;
  h=0;
  do while(mtc1<2);
    i=nexcom2(n, k, feld1, mtc1,t,h);
    if all(feld1) then
      if first=0 then do;
        mengen=feld1;
        first=1;
      end;
    else mengen=mengen//feld1;
  end;
  return(nrow(mengen));
finish;


start nexcom2(n_ges, c, feld, mtc, t, h);
  if (mtc=0) then do;
    feld[1]=n_ges;
    if(c>1)then do i=2 to c;
      feld[i]=0;
    end;
    t=n_ges;
    h=0;
  end;
  else goto marke2;

  marke1: if (feld[c]=n_ges) then mtc=2;
    else mtc=1;
    return(1);
```

```
  marke2: if (t>1)then h=0;
    h=h+1;
    t=feld[h];
    feld[h]=0;
    feld[1]=t-1;
    feld[h+1]=feld[h+1]+1;
    goto marke1;
finish;



start LRT(x,s,n);
  k=ncol(n);
  nu=sum(n)-k;
  m=MLE(x,n);
  xquer=sum(x#n/n[+]);

  x1=sum(n#(m-xquer)##2);
  x2=sum(n#(m-x)##2);
  t=x1/(x1+x2+nu*s);
  run plkwunba(n,p_lkw);

  p=0;
  do i=2 to k;
    p=p+p_lkw[i]*(1-probf(t*(sum(n)-i)/((i-1)*(1-t)),i-1,sum(n)-i));
  end;
  return(p);
finish;



/* Example call */

n={6 6 6 6};
xbar={-76.2 -73.5 -73.4 -74.4};
s=60.078;

p_value=LRT(xbar,s,n);
print p_value [format=6.4];



/* Output */

P_VALUE
 0.4634
```

B.2.5. SAS/IML module for calculating generators of good lattice vectors according to the minimisation rule (2.15)

```
/************************************************************************/
/* Module for computing good lattice vectors of randomised lattice rules */
/*                                                                      */
/* Input:  p_vector = fineness of lattice (row vector)                  */
/*         dim      = dimension of distribution function (scalar)       */
/*                                                                      */
/* Output: p        = particular finess of lattice                      */
/*         q        = particular dimension                              */
/*         h        = generator of good lattice vector                  */
/*                                                                      */
/************************************************************************/

/* Calculation conducted according to equation (2.15) */

start lattvec(p_vector,dim);
  do q=2 to dim;
    do index=1 to ncol(p_vector);
      p=p_vector[index];
      numb=int(p/4)+1; numb1=int(p/2)+1;
      help=int(sqrt(p)); a=j(1,numb,.);
      a[1:help]=1:help; t=help;
      do j=help+1 to numb1;
        i=t+1;
        do until(krit=1 | krit=0);
          i=i-1;
          if i=1 then krit=0;
          else krit=mod(j#a[i],p);
          if krit=p-1 then krit=1;
        end;
        if krit=0 then do;
          t=t+1; a[t]=j;
        end;
      end;
      vec=0:q-2; krit=10;
      do i=1 to numb;
        sum=0;
        z=mod(j(1,q-1,a[i])##vec,p);
        do j=0 to p-1;
          zz=mod(z#j/p,1);
          f2=1+19.7392*(zz##2-zz+1/6);
          f2n=f2/2##2+0.75;
          sum=sum+f2n[#];
        end;
        sum=sum/p;
        if sum<krit then do;
          krit=sum; val=a[i];
        end;
      end;
      z=mod(j(1,q-1,val)##vec,p); h=z;
      print q p h;
    end;
  end;
finish;

/* Example call */

p_vector={157 313 619 1249 2503 5003 10007 20011 40009 80021}; dim=21;
run lattvec(p_vector,dim);
```

B.2.6. FORTRAN program for calculating generators of good lattice vectors according to the minimisation rule (2.15)

```
C       Module for computing good lattice vectors of randomised lattice rules.
C
C       Input:  p_vector = fineness of lattice
C               dim      = dimension of distribution function
C
C       Output: q        = particular dimension
C               val      = generator of good lattice vector
C               krit1    = minimum value according to (2.15)
C
C       The output is stored in 'qcalc.out'. To execute the program this file must exist prior
C       to running the program. Calculation conducted according to equation (2.15).


        real*8,dimension(:),allocatable::a,vec,z,zz,f2,f2n
        real*8 p_vector(9),fff,sum,val,krit1,xxx,pp,p,one
        integer numb,numb1,help,t,i,qq,krit
        character(len=8)date
        character(len=10)time

        data p_vector/157,313,619,1249,2503,5003,10007,20011,40009/
        dim=31
        Open(8,file='qcalc.out')

        call date_and_time(date,time)
        write(8,*)'date and time ',date,time

        do q=2,dim
          one=1.

          do index=1,3
            write (6,*) q,index
            p=p_vector(index)
            numb=int(p/4)+1
            numb1=int(p/2)+1
            help=int(sqrt(p))
            allocate(a(numb))
            do j=1,help
              a(j)=j
            enddo
            t=help
            do j=help+1,numb1
              i=t+1
              do
                i=i-1
                if (i.lt.2) then
                  krit=0
                else
c                   krit=int(j*a(i))-(int((j*a(i))/p)*p)
                   krit=mod(j*a(i),p)
                end if

                if (krit.eq.p-1) krit=1
                if ((krit.eq.0).or.(krit.eq.1)) exit
              enddo
              if (krit.eq.0) then
                t=t+1;
                a(t)=j;
              endif
```

```
    enddo

    pp=p
    qq=q-1
    allocate(vec(qq))
    do j=1,q-1
      vec(j)=j-1
    enddo
    krit1=10000000.
    do i=1,numb
      sum=0.
      allocate(z(qq))
      allocate(zz(qq))
      allocate(f2(qq))
      allocate(f2n(qq))
      do j=1,q-1
         z(j)=dmod(a(i)**vec(j),pp)
      enddo;
      do j=0,p-1
        fff=1.
        do jj=1,q-1
          zz(jj)=dmod(z(jj)*j/p,one)
          xxx=1+19.7392*((zz(jj)**2)-zz(jj)+(1./6.))
          f2(jj)=xxx
          f2n(jj)=f2(jj)/2.**2+0.75
          fff=fff*f2n(jj)
        enddo
        sum=sum+fff
      enddo

      sum=sum/p
      if (sum.lt.krit1) then
        krit1=sum
        val=a(i)
      endif

      deallocate(z,zz,f2,f2n)
    enddo
    allocate(z(qq))
    do j=1,q-1
       z(j)=dmod(val**vec(j),pp)
    enddo;
    write (8,*) val,krit1
    deallocate(a,vec,z)

  enddo

call date_and_time(date,time)
write(8,*)'date and time ',date,time,q

enddo
end
```

B.2.7. SAS/IML module for calculating multivariate *t*–probabilities according to Genz and Bretz (1999) for arbitrary upper integration bounds (acceptance-rejection)

```
/*************************************************************************/
/* Module for calculation of mvt probabilities according to Genz         */
/*                 (acceptance-rejection algorithm)                      */
/*                                                                       */
/* Input:  z     = upper integration bound (row vector)                  */
/*         R     = correlation matrix                                    */
/*         df    = degrees of freedom (scalar)                           */
/*         eps   = accuracy (scalar)                                     */
/*         n_max = maximum number of simulations (scalar)                */
/*                                                                       */
/* Output: prob  = probability                                           */
/*         error = simulated standard error                              */
/*                                                                       */
/*************************************************************************/

/* The program here is restricted to one-sided upper mvt probabilities */

start mvt_ar(df,b,r,eps,n_max);
  n=ncol(b);
  c=t(root(R));
  index=0; intval=0; varsum=0;
  y=j(1,n,0);
  do until(error<eps | index=n_max);
    w=ranuni(j(1,n,141071));
    f=1;
    y[1]=tinv(w[1],df);
    if c[1,1]*y[1]>b[1] then f=0;
    else do;
      do i=2 to n;
        y[i]=tinv(w[i],df+i-1)*sqrt((df+sum(y[1:i-1]##2))/(df+i-1));
        if sum(c[i,1:i]#t(y[1:i]))>b[i] then do;
          i=n; f=0;
        end;
      end;
    end;
    index=index+1;
    varsum=varsum+(index-1)*(f-intval)**2/index;
    intval=intval+(f-intval)/index;
    if index<100 then error=1;
    else error=3*sqrt(varsum/(index*(index-1)));
  end;
  prob=intval;
  print prob [format=6.4] error [format=6.4];
finish;

/* Example call */

R={  1   .5  .5  .9,
    .5    1 -.5  .5,
    .5  -.5   1  .5,
    .9   .5  .5   1};
z={1 1 1 1}; eps=0.001; n_max=10000; df=50;
run mvt_ar(df,z,r,eps,n_max);

/* Output */

  PROB  ERROR
0.6279 0.0145
```

B.2.8. SAS/IML module for calculating multivariate *t*–probabilities according to Genz and Bretz (1999) for arbitrary upper integration bounds (Monte Carlo method)

```
/************************************************************************/
/* Module for calculation of mvt probabilities according to Genz        */
/*                 (Monte Carlo algorithm)                              */
/*                                                                      */
/* Input:  z     = upper integration bound (row vector)                 */
/*         R     = correlation matrix                                   */
/*         df    = degrees of freedom (scalar)                          */
/*         eps   = accuracy (scalar)                                    */
/*         n_max = maximum number of simulations (scalar)               */
/*                                                                      */
/* Output: prob  = probability                                          */
/*         error = simulated standard error                             */
/*                                                                      */
/************************************************************************/

/* The program here is restricted to one-sided upper mvt probabilities */

start mvt_mc(df,b,r,eps,n_max);
  n=ncol(b);
  c=t(root(r))+1E-12;
  index=0;
  intval=0;
  varsum=0;
  f=j(1,n,0); y=f;
  f[1]=probt(b[1]/c[1,1],df);
  e=f;
  do until(error<eps | index=n_max);
    w=ranuni(j(1,n,141071));
    y[1]=tinv(w[1]*e[1],df);
    do i=2 to n;
      e[i]=probt((b[i]-sum(c[i,1:i-1]*y[1:i-1]))
               *sqrt((df+i-1)/(df+sum(y[1:i-1]##2)))/c[i,i],df+i-1);
      f[i]=e[i]*f[i-1];
      y[i]=tinv(w[i]*e[i]+1E-12,df+i-1)*sqrt((df+sum(y[1:i-1]##2))/(df+i-1));
    end;
    index=index+1;
    varsum=varsum+(index-1)*(f[n]-intval)**2/index;
    intval=intval+(f[n]-intval)/index;
    if index<100 then error=1;
    else error=3*sqrt(varsum/(index*(index-1)));
  end;
  prob=intval;
  print prob [format=6.4] error [format=6.4];
finish;

/* Example call */

R={  1   .5  .5  .9,
    .5    1 -.5  .5,
    .5  -.5   1  .5,
    .9   .5  .5   1};
z={1 1 1 1}; eps=0.001; n_max=10000; df=50;
run mvt_mc(df,z,r,eps,n_max);

/* Output */

  PROB  ERROR
0.6249 0.0080
```

B.2.9. SAS/IML module for calculating multivariate *t*–probabilities according to Genz and Bretz (1999) for arbitrary upper integration bounds and dimension $q < 33$ (randomised lattice rules)

```
/*************************************************************************/
/* Module for calculation of mvt probabilities according to Genz         */
/*                 (randomised lattice rule algorithm)                   */
/*                                                                       */
/* Input:  z     = upper integration bound (row vector)                  */
/*         R     = correlation matrix                                    */
/*         df    = degrees of freedom (scalar)                           */
/*         eps   = accuracy (scalar)                                     */
/*                                                                       */
/* Output: prob  = probability                                           */
/*         error = simulated standard error                              */
/*                                                                       */
/*************************************************************************/

/* The program here is restricted to one-sided upper mvt probabilities */

start mvt_lr(df,b,r,eps);
  q=ncol(b);
  c=t(root(r))+1E-12;
  y=j(1,q-1,0);
  e=j(1,q,0);
  e[1]=probt(b[1]/c[1,1],df);
  n=10;
  vec=0:q-2;
  p_vector={157 313 619 1249 2503 5003 10007 20011};
  mat={ 1   1   1   1   1    1    1     1,
       46 119 239 512 672 1850 3822  6103,
       46  93 178 136 652 1476 2325  2894,
       17  51  73 197 792  792 1206  8455,
       18  51 104 165 792  380 1927  3629,
       18  80 102 175 253  162 2286  1752,
       11  70 161 303 306  363  343  1920,
       11  70 161 155 153  137  378   652,
       11  93 106  18 288  186   81   146,
       36  62  57  27 288  186  182   156,
       36  15  57  27  29   33   76   136,
       36  15  36  24 128   36   21    44,
        4  19  22  24  64   38   21    31,
       30  15  22  24  16   36   21   161,
       31   9  22  24  16   48   20   161,
       31   9  22  14  16   48   21    11,
        6  20   6  14  16   12   21    11,
        6   9   6  14  64   12   11    13,
        3   9   6  14  16   12   11    13,
        3   9   6   8  16    6   11    13,
        3  16   6   8  16    6    7    13,
        3  16   6   8  16    6    7    22,
        3  16   6   8  16    6    7    13,
        3  16   6   8   8    6    7    13,
        3  16   6   8   8    6    4    13,
        3  16   6   8   8    5    4    16,
        3  16   4   3   8    5    4    16,
        3   4   4   3   8    4    4    13,
        3   4   4   3   8    4    4    13,
        3   4   4   3   8    4    4     8,
        3   4   4   3   8    4    4     8};
```

```
  do until(n>50 | error<eps);
    index=1;
    do until(index=9 | error<eps);
      p=p_vector[index];
      h=mat[q-1,index];
      z=mod(j(1,q-1,h)##vec,p);

      intval=0;
      varsum=0;

      do l=1 to n;
        latsum=0;
        rr=ranuni(j(1,q-1,141071));

        do j=1 to p;
          w=abs(2*mod(rr+j#z/p,1)-1);
          y[1]=tinv(w[1]*e[1],df);

          do i=2 to q-1;
            e[i]=probt((b[i]-sum(c[i,1:i-1]*y[1:i-1]))*sqrt((df+i-1)/
                   (df+sum(y[1:i-1]##2)))/c[i,i],df+i-1)+1E-12;
            y[i]=tinv(w[i]*e[i],df+i-1)
                   *sqrt((df+sum(y[1:i-1]##2))/(df+i-1));
          end;

          e[q]=probt((b[q]-sum(c[q,1:q-1]*y[1:q-1]))*sqrt((df+q-1)/
                   (df+sum(y[1:q-1]##2)))/c[q,q],df+q-1)+1E-12;
          f=e[#];
          latsum=latsum+(f-latsum)/j;
        end;

        varsum=varsum+(l-1)*(latsum-intval)**2/l;
        intval=intval+(latsum-intval)/l;
      end;

      error=3*sqrt(varsum/(n*(n-1)));
      index=index+1;
    end;
  n=n+2;
  end;
  prob=intval;
  print prob [format=6.4] error [format=6.4];
finish;


/* Example call */

df=50;
R={  1   .5  .5  .9,
    .5    1 -.5  .5,
    .5  -.5   1  .5,
    .9   .5  .5   1};
z={1 1 1 1};
eps=0.001;

run mvt_lr(df,z,r,eps);


/* Output */

  PROB  ERROR
0.6222 0.0008
```

B.2.10. SAS/IML module for calculating multivariate *t*–quantiles using the bisection method

```
/************************************************************************/
/* Module for calculation of mvt quantiles (bisection method)           */
/*                                                                      */
/* Input:  x1      = lower bound of bracket containing the root (scalar) */
/*         x2      = upper bound of bracket containing the root (scalar) */
/*         R        = correlation matrix                                */
/*         df       = degrees of freedom (scalar)                       */
/*         eps1     = accuracy of mvt-procedure (scalar)                */
/*         eps2     = accuracy of root finding procedure (scalar)       */
/*         quantile = pre-determined probability (scalar)               */
/*                                                                      */
/* Output: quan     = quantile                                          */
/*         prob     = probability                                       */
/*         diff     = difference of last two calculated quantiles       */
/*         index    = muber of iterations                               */
/*                                                                      */
/* Required modules: mvt_AR, mvt_MC or mvt_LR                           */
/*                                                                      */
/************************************************************************/

/* The program can also be used for mvn probabilities */

start bisec(x1,x2,eps1,eps2,R,df,alpha);
  f=-1;
  quan=1;
  index=0;

  do until(abs(diff)<eps2);
    if f>0 then do;
      diff=x2-quan;
      x2=quan;
    end;
    else do;
      diff=x1-quan;
      x1=quan;
    end;
    quan=(x1+x2)/2;
    prob=mvt(df,quan,r,eps1);
    f=prob-alpha;
    index=index+1;
  end;
  print quan [format=6.4] prob [format=6.4] diff [format=6.4] index;
finish;

/* Example call using mvt_LR */

R={         1 0.4403855 0.8257228,
    0.4403855         1 0.3636364,
    0.8257228 0.3636364         1};
df=34; quantile=0.95;
x1=0; x2=5;
eps1=0.0001; eps2=0.0001;
run bisec(x1,x2,eps1,eps2,R,df,quantile);

/* Output */

  QUAN   PROB   DIFF     INDEX
2.1021 0.9500 -.0001        17
```

B.2.11. SAS/IML module for calculating multivariate *t*–quantiles using the regula falsi

```
/************************************************************************/
/* Module for calculation of mvt quantiles (regula falsi)               */
/*                                                                      */
/* Input:  x1      = lower bound of bracket containing the root (scalar) */
/*         x2      = upper bound of bracket containing the root (scalar) */
/*         R        = correlation matrix                                */
/*         df       = degrees of freedom (scalar)                       */
/*         eps1     = accuracy of mvt-procedure (scalar)                */
/*         eps2     = accuracy of root finding procedure (scalar)       */
/*         quantile = pre-determined probability (scalar)               */
/*                                                                      */
/* Output: quan     = quantile                                          */
/*         prob     = probability                                       */
/*         diff     = difference of last two calculated quantiles       */
/*         index    = muber of iterations                               */
/*                                                                      */
/* Required modules: mvt_AR, mvt_MC or mvt_LR                           */
/*                                                                      */
/************************************************************************/

/* The program can also be used for mvn probabilities */

start regfalsi(x1,x2,eps1,eps2,R,df,alpha);
  f1=mvt(df,x1,r,eps1)-alpha;
  f2=mvt(df,x2,r,eps1)-alpha;
  index=0;
  do until(abs(diff)<eps2);
    dx=x2-x1;
    quan=x1+f1*dx/(f1-f2);
    f=mvt(df,quan,r,eps1)-alpha;
    if f<0 then do;
      diff=x1-quan;
      x1=quan;
      f1=f;
    end;
    else do;
      diff=x2-quan;
      x2=quan;
      f2=f;
    end;
    index=index+1;
  end;
  prob=f+alpha;
  print quan [format=6.4] prob [format=6.4] diff [format=6.4] index;
finish;

/* Example call using mvt_LR */

R={         1 0.4403855 0.8257228,
    0.4403855         1 0.3636364,
    0.8257228 0.3636364         1};
df=34; quantile=0.95;
x1=0; x2=5;
eps1=0.0001; eps2=0.0001;
run regfalsi(x1,x2,eps1,eps2,R,df,quantile);

/* Output */

  QUAN   PROB   DIFF     INDEX
2.1024 0.9500 0.0001        31
```

B.2.12. SAS/IML module for calculating multivariate *t*–quantiles using the seacnt method

```
/************************************************************************/
/* Module for calculation of mvt quantiles (secant method)              */
/*                                                                      */
/* Input:  x1      = lower bound of bracket containing the root (scalar) */
/*         x2      = upper bound of bracket containing the root (scalar) */
/*         R        = correlation matrix                                */
/*         df       = degrees of freedom (scalar)                       */
/*         eps1     = accuracy of mvt-procedure (scalar)                */
/*         eps2     = accuracy of root finding procedure (scalar)       */
/*         quantile = pre-determined probability (scalar)               */
/*                                                                      */
/* Output: quan     = quantile                                          */
/*         prob     = probability                                       */
/*         diff     = difference of last two calculated quantiles       */
/*         index    = muber of iterations                               */
/*                                                                      */
/* Required modules: mvt_AR, mvt_MC or mvt_LR                           */
/*                                                                      */
/************************************************************************/

/* The program can also be used for mvn probabilities */

start secant(x1,x2,eps1,eps2,R,df,alpha);
  quan=x2;
  f1=mvt(df,x1,r,eps1)-alpha;
  f=mvt(df,x2,r,eps1)-alpha;
  index=0;

  do until(abs(dx)<eps2);   print quan f f1;
    dx=(x1-quan)*f/(f-f1+1E-12);
    x1=quan;
    f1=f;
    quan=quan+dx;

    f=mvt(df,quan,r,eps1)-alpha;
    index=index+1;
  end;
  prob=f+alpha;
  print quan [format=6.4] prob [format=6.4] diff [format=6.4] index;
finish;

/* Example call using mvt_LR */

R={         1 0.4403855 0.8257228,
    0.4403855         1 0.3636364,
    0.8257228 0.3636364         1};
df=34;
x1=0;
x2=5;
eps1=0.0001;
eps2=0.0001;
quantile=0.95;

run secant(x1,x2,eps1,eps2,R,df,quantile);

/* Output */

Not convergent
```

B.2.13. SAS/IML module for calculating multivariate *t*–quantiles using the method according to Ridders (1979)

```
/************************************************************************/
/* Module for calculation of mvt quantiles (Ridders method)             */
/*                                                                      */
/* Input:  x1      = lower bound of bracket containing the root (scalar) */
/*         x2      = upper bound of bracket containing the root (scalar) */
/*         R        = correlation matrix                                */
/*         df       = degrees of freedom (scalar)                       */
/*         eps1     = accuracy of mvt-procedure (scalar)                */
/*         eps2     = accuracy of root finding procedure (scalar)       */
/*         quantile = pre-determined probability (scalar)               */
/*                                                                      */
/* Output: quan     = quantile                                          */
/*         prob     = probability                                       */
/*         diff     = difference of last two calculated quantiles       */
/*         index    = muber of iterations                               */
/*                                                                      */
/* Required modules: mvt_AR, mvt_MC or mvt_LR                           */
/*                                                                      */
/************************************************************************/

/* The program can also be used for mvn probabilities */



start ridders(x1,x2,eps1,eps2,R,df,alpha);
  f1=mvt(df,x1,r,eps1)-alpha;
  f2=mvt(df,x2,r,eps1)-alpha;

  quan_old=8;
  index=0;

  do until(abs(diff)<eps2);
    x=(x1+x2)/2;
    f=mvt(df,x,r,eps1)-alpha;

    wurz=sqrt(f*f-f1*f2);
    quan=x+(x-x1)*sign(f1-f2)*f/wurz;
    fquan=mvt(df,quan,r,eps1)-alpha;
    diff=quan-quan_old;
    quan_old=quan;

    if f*fquan<0 then do;
      x1=x;
      f1=f;
      x2=quan;
      f2=fquan;
    end;
    else do;
      if f1*fquan<0 then do;
        x2=quan;
        f2=fquan;
      end;
      else do;
        if f2*fquan<0 then do;
          x1=quan;
          f1=fquan;
        end;
      end;
```

```
    end;

    index=index+1;
  end;

  prob=fquan+alpha;
  print quan [format=6.4] prob [format=6.4] diff [format=6.4] index;
finish;


/* Example call using mvt_LR */

R={         1 0.4403855 0.8257228,
    0.4403855         1 0.3636364,
    0.8257228 0.3636364         1};
df=34;
x1=0;
x2=5;
eps1=0.0001;
eps2=0.0001;
quantile=0.95;

run ridders(x1,x2,eps1,eps2,R,df,quantile);



/* Output */

  QUAN   PROB   DIFF     INDEX
2.1022 0.9500 0.0000         4
```

B.3.1. SAS/IML module for computing the entries of Williams' MCT according to Definition 3.1.

```
/************************************************************************/
/* Module for computing the contrast matrix of Williams' MCT            */
/*                                                                      */
/* Input:  n  = sample size (row vector)                                */
/*                                                                      */
/* Output: cm = contrast matrix                                         */
/*                                                                      */
/************************************************************************/

/* The program calculates the coefficients adjusted for sample size
   allocation and for arbitrary number of treatment groups          */



start will_con(n);
  k1=ncol(n)-1;

  c=j(1,k1,.);
  cm=j(k1,k1,0);
  cm=j(k1,1,-1)||cm;

  do i=1 to k1;
    x=sum(n[k1-i+2:k1+1]);
    do j=1 to i;
      cm[i,k1-j+2]=n[k1-j+2]/x;
    end;
  end;

  return(cm);
finish;



/* Example call */

n={10 10 10 10};

cm=will_con(n);
print cm;



/* Output */

CM
-1         0         0         1
-1         0       0.5       0.5
-1 0.3333333 0.3333333 0.3333333
```

B.3.2. SAS/IML module for computing the entries of Marcus' MCT according to Definition 3.2.

```
/************************************************************************/
/* Module for computing the contrast matrix of Marcus' MCT              */
/*                                                                      */
/* Input:  n  = sample size (row vector)                                */
/*                                                                      */
/* Output: cm = contrast matrix                                         */
/*                                                                      */
/************************************************************************/

/* The program calculates the coefficients adjusted for sample size
   allocation and for arbitrary number of treatment groups          */



start marc_con(n);
  k=ncol(n);

  cm1=j(k-1,k,0);
  cm2=cm1;

  do i=1 to k-1;
    cm1[i,i+1:k]=t(n[i+1:k]/sum(n[i+1:k]));
  end;

  do i=1 to k-1;
    cm2[i,1:i]=t(n[1:i]/sum(n[1:i]));
  end;

  row=k*(k-1)/2;
  cm=j(row,k,0);
  index=1;
  do i=1 to k-1;
    do j=1 to i;
      cm[index,]=cm1[i,]-cm2[j,];
      index=index+1;
    end;
  end;

  return(cm);
finish;


/* Example call */

n={10 10 10 10};
cm=marc_con(n);
print cm;


/* Output */

       CM
       -1 0.3333333 0.3333333 0.3333333
       -1         0       0.5       0.5
     -0.5      -0.5       0.5       0.5
       -1         0         0         1
     -0.5      -0.5         0         1
-0.333333 -0.333333 -0.333333         1
```

B.3.3. SAS/IML module for computing the entries of the isotonic MCT according to Definition 3.3.

```
/************************************************************************/
/* Module for computing the contrast matrix of the isotonic MCT         */
/*                                                                      */
/* Input:  n  = sample size (row vector)                                */
/*                                                                      */
/* Output: cm = contrast matrix                                         */
/*                                                                      */
/************************************************************************/

/* The program calculates the coefficients adjusted for sample size
   allocation and for arbitrary number of treatment groups          */

start schnitt(vec,stelle);
  col=ncol(vec);
  if stelle=col then neu=vec[,1:col-1];
  else do;
    neu=vec[,1:col-1];
    neu[stelle:col-1]=vec[,stelle+1:col];
  end;
  return(neu);
finish;

start iso_con(n);
  k=ncol(n)-1;
  anz=2##k;
  cm=j(2##k-1,k+1,.);
  mat=j(anz,k,0);
  mat[2,k]=1;
  do i=2 to k;
    mat[2##(i-1)+1:2##i,k-i+1]=j(2##(i-1),1,1);
    mat[2##(i-1)+1:2##i,k-i+2:k]=mat[1:2##(i-1),k-i+2:k];
  end;
  mat=mat[2:2##k,];
  row=nrow(mat);
  do i=1 to row;
    y=mat[i,];
    x=j(1,k,1);
    krit=0;
    j=1;
    do until(krit=1);
      if y[j]=0 then do;
        if y[j]=y[j+1] then do;
          y=schnitt(y,j+1);
          x=schnitt(x,j+1);
          x[j]=x[j]+1;
        end;
        else j=j+1;
      end;
      else j=j+1;
      if j>=ncol(y) then krit=1;
    end;

    do j=1 to ncol(y);
      if y[j]=0 then x[j]=x[j]+1;
    end;
    j=1;
    do while(j<(ncol(y)-1));
      if y[j]=0 then do;
```

```
        l=j+1;
        do until(y[l]=0 | l>=ncol(y));
          l=l+1;
        end;
        if y[l]=0 then do;
          y=schnitt(y,j+1);
          x=schnitt(x,j+1);
        end;
      end;
      j=j+1;
    end;

    if i=row then do;
      y=y||{1};
      x=x||{1};
    end;
    s=mat[i,+];
    submat=I(s+1);
    do j=1 to s;
      do jj=j+1 to s+1;
        submat[j,jj]=1;
      end;
    end;
    ssd=j(1,s+1,0);
    do j=2 to s+1;
      ssd[j]=sqrt(sum(((submat[j,]-sum(submat[j,]#x)/x[+])##2)#x));
    end;
    c=j(1,s+1,ssd[s+1]);
    do j=s to 1 by -1;
      c[j]=ssd[j]-sum(c[j+1:s+1]);
    end;
    contrast=c;
    cm[i,1:x[1]]=contrast[1]/x[1];
    do j=2 to ncol(y);
      cm[i,sum(x[1:j-1])+1:sum(x[1:j])]=contrast[j]/x[j];
    end;
  end;

  do i=1 to row;
    do j=1 to k+1;
      if abs(cm[i,j])<1E-10 then cm[i,j]=0;
    end;
    cm[i,]=n#cm[i,]-cm[i,]*t(n)/(k+1);
  end;
  return(cm);
finish;

/* Example call */

n={10 10 10 10};
cm=iso_con(n);
print cm;

/* Output */

       CM
-2.886751 -2.886751 -2.886751  8.660254
       -5        -5         5         5
       -5        -5  1.339746  8.660254
-8.660254 2.8867513 2.8867513 2.8867513
-8.660254         0         0  8.660254
-8.660254 -1.339746         5         5
-8.660254 -1.339746  1.339746  8.660254
```

B.4.1. SAS/IML module for computing the power for arbitrary multiple contrasts for normal means according to equation (7.2) ($1 < q < 33$)

```
/************************************************************************/
/* Module for calculation of power of MCTs for normal means             */
/*                                                                      */
/* Input:  mu     = expected means (row vector)                         */
/*         cm     = contrast matrix                                     */
/*         s      = variance                                            */
/*         n      = sample size                                         */
/*         eps1   = accuracy of mvt-procedure (scalar)                  */
/*         eps2   = accuracy of root finding procedure (scalar)         */
/*         probab = 1 - α                                               */
/*                                                                      */
/* Output: power  = power value                                         */
/*                                                                      */
/* Required modules: mvt_AR, mvt_MC or mvt_LR                           */
/*                   Genz_mvn                                           */
/*                   bisec, regfalsi, secant or Ridders                 */
/*                                                                      */
/************************************************************************/

/* The program is designed for contrast matrices with less than 33 contrasts */


start corr(cm,sampsize) global(q,r,df,var);
  q=nrow(cm);
  df=sum(sampsize)-ncol(sampsize);
  rr=j(q,q,0);
  var=j(1,q,.);

  do i=1 to q-1;
    do j=i+1 to q;
      rr[i,j]=sum(cm[i,]#cm[j,]/sampsize)/
              sqrt(sum(cm[i,]##2/sampsize)*sum(cm[j,]##2/sampsize));
    end;
  end;

  r=rr+rr`+I(q);
  do i=1 to q;
    var[i]=sum(CM[i,]##2/sampsize);
  end;
finish;


start nu(x) global(eps3,quan,df,nc,r);
  g=(x#quan/sqrt(df)-nc);
  qwer=mvn(g,eps3,r);
  return(qwer#x##(df-1)#exp(-x#x/2));
finish;


start powermct(mu,cm,s,n,probab,eps1,eps2) global(q,r,df,var, eps3,quan,nc);
  eps3=eps1;
  aa=1E-07||.p;
  run corr(cm,n);
  quan=quantile(probab,eps3,eps2,r,df);
  nc=cm*mu`/sqrt(s*var`);
  call quad(power,"nu",aa) eps=1E-06;
  power=1-power#(1/2)##(df/2-1)/gamma(df/2);
  return(power);
```

```
finish;

/* Example call */

mu={0 .3333 .6667 1};
cm={-1     0     0     1,
    -1     0    .5    .5,
    -1 .3333 .3333 .3333};
s=1;
n={14 8 8 8};
probab=0.95;
eps1=0.001;
eps2=0.001;


powermct=powermct(mu,cm,s,n,probab,eps1,eps2);
print powermct [format=6.4];



/* Output */

 POWER
0.7156
```

B.5.1. SAS/IML module for computing the power for arbitrary multiple contrasts for binomial data ($1 < q < 33$)

```
/************************************************************************/
/* Module for calculation of power of MCTs for binomial data            */
/*                                                                      */
/* Input:  mu       = expected means (row vector)                       */
/*         cm       = contrast matrix                                   */
/*         n        = sample size                                       */
/*         alpha    = type-I-error                                      */
/*                                                                      */
/* Output: powlabel = row vector containing the labels                  */
/*         pow      = power value                                       */
/*                                                                      */
/* Required modules: Genz_mvn                                           */
/*                   Solow                                              */
/*                   bisec, regfalsi, secant or Ridders                 */
/*                                                                      */
/************************************************************************/

/* In contrast to program B.4.1 the Solow procedure was also included (for the
   calculation of quantiles) and therefore no error bounds 'eps' are passed;
   alternatively the Genz_mvn module can be used with some some smaller modifications.

   The program calculates the power for the following tests:

   MCT_PO: pooled multiple contrast, no continuity correction
   MCT_P1: pooled multiple contrast, minimum continuity correction (cc2)
   MCT_P2: pooled multiple contrast, average continuity correction (cc3)
   MCT_P3: pooled multiple contrast, maximum continuity correction (cc1)
   MCT_UO: unpooled multiple contrast, no continuity correction
   MCT_U1: unpooled multiple contrast, minimum continuity correction (cc2)
   MCT_U2: unpooled multiple contrast, average continuity correction (cc3)
   MCT_U3: unpooled multiple contrast, maximum continuity correction (cc1)          */



start powbino(n,profile,alpha,cm);

  kk=ncol(n)-1;
  quan=1-alpha;

  k=nrow(cm);
  pp=sum(n#profile)/sum(n);

  RR=j(k,k,0);
  do i=1 to k-1;
    do j=i+1 to k;
      RR[i,j]=sum(CM[i,]#CM[j,]/N)/sqrt(sum(CM[i,]##2/N)*sum(CM[j,]##2/N));
    end;
  end;
  R_O=RR+RR`+I(k);
  run  quantil(k,quan,r_O);

  aa=cm;
  zwres1=profile#(1-profile)/n;
  sigma=diag(zwres1);
  zwres2=aa*sigma*t(aa);
  zwres3=diag(sqrt(1/vecdiag(zwres2)));
  r=zwres3*zwres2*zwres3;
```

```
  pow=powcalc(r,k,kk,cm,n,pp,profile);
  powlabel={mct_p0 mct_p1 mct_p2 mct_p3 mct_u0 mct_u1 mct_u2 mct_u3};

  print powlabel pow [format=6.4];
finish;


start powcalc(r,k,kk,cm,n,pp,profile) global(power);
  help1=j(k,kk+1,.);
  do i=1 to k;
    help1[i,]=cm[i,]/n;
  end;
  help2=help1[,2:kk+1]||j(k,1,.);
  help3=abs(help1-help2);
  delta0=0;
  delta1=help3[,><]/2;
  delta2=help3[,+]/(2*kk);
  delta3=help3[,<>]/2;


  delta=delta0;
  x=sqrt(pp#(1-pp)#cm##2*t(1/n));
  xx1=(cm*t(profile)-delta)/x;
  xx2=sqrt(cm##2#profile#(1-profile)*t(1/n))/x;
  run  multinor(k,xx1,xx2,r);
  power1=power;

  delta=delta1;
  x=sqrt(pp#(1-pp)#cm##2*t(1/n));
  xx1=(cm*t(profile)-delta)/x;
  xx2=sqrt(cm##2#profile#(1-profile)*t(1/n))/x;
  run  multinor(k,xx1,xx2,r);
  power2=power;

  delta=delta2;
  x=sqrt(pp#(1-pp)#cm##2*t(1/n));
  xx1=(cm*t(profile)-delta)/x;
  xx2=sqrt(cm##2#profile#(1-profile)*t(1/n))/x;
  run  multinor(k,xx1,xx2,r);
  power3=power;

  delta=delta3;
  x=sqrt(pp#(1-pp)#cm##2*t(1/n));
  xx1=(cm*t(profile)-delta)/x;
  xx2=sqrt(cm##2#profile#(1-profile)*t(1/n))/x;
  run  multinor(k,xx1,xx2,r);
  power4=power;


  delta=delta0;
  x=sqrt(cm##2#profile#(1-profile)*t(1/n));
  xx1=(cm*t(profile)-delta)/x;
  xx2=1;
  run  multinor(k,xx1,xx2,r);
  power5=power;

  delta=delta1;
  x=sqrt(cm##2#profile#(1-profile)*t(1/n));
  xx1=(cm*t(profile)-delta)/x;
  xx2=1;
  run  multinor(k,xx1,xx2,r);
  power6=power;
```

```
  delta=delta2;
  x=sqrt(cm##2#profile#(1-profile)*t(1/n));
  xx1=(cm*t(profile)-delta)/x;
  xx2=1;
  run  multinor(k,xx1,xx2,r);
  power7=power;

  delta=delta3;
  x=sqrt(cm##2#profile#(1-profile)*t(1/n));
  xx1=(cm*t(profile)-delta)/x;
  xx2=1;
  run  multinor(k,xx1,xx2,r);
  power8=power;


  pow=j(1,8,.);
  pow[1]=power1;
  pow[2]=power2;
  pow[3]=power3;
  pow[4]=power4;
  pow[5]=power5;
  pow[6]=power6;
  pow[7]=power7;
  pow[8]=power8;

  return(pow);

finish;


/* Example call */

n={90 50 50 10};
mu={.05 .1 .15 .2};
alpha=0.05;
cm={-1     0     0     1,
    -1     0    .5    .5,
    -1 .3333 .3333 .3333};

run powbino(n,mu,alpha,cm);


/* Output */

POWLABEL
MCT_P0 MCT_P1 MCT_P2 MCT_P3 MCT_U0 MCT_U1 MCT_U2 MCT_U3


   POW
0.6292 0.6193 0.5740 0.5210 0.5319 0.5252 0.4752 0.4263
```

Acknowledgements

Many people influenced the writing of the thesis. I am indebted to my advisor L.A. Hothorn for his guidance, encouragement and support over the past years. This thesis is developed only based on a small portion of the joint work.

I thank L. Baringhaus for many helpful discussions and his willingness to supervise the thesis together with L.A. Hothorn. It was a joy to work with Alan Genz and I'm very grateful for his help upon the numerical evaluation of multivariate distributions. Chapter 2 would not exist in this form without his constant support. Moreover, I thank Paul Somerville, Tony Hayter, David Bristol, H. Stützel and W.D. Richter for continuous stimulating discussions, which improved many parts of the thesis.

I have enjoyed working with my colleagues from the research unit Bioinformatics. Special thanks go to Dirk Seidel, Michael Weichert, Markus Neuhäuser, Egbert Biesheuvel and Dieter Hauschke for the lot of patience to discuss my problems I had over the years and to read parts of the thesis. Hannelore Visser and Clemens Buczilowski provided the technical support.

And last but not least I also thank my parents, who gave me all available support and encouragement since my childhood.

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides Statt, daß ich die vorliegende Arbeit selbständig angefertigt habe und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe sowie diese Arbeit noch keiner anderen Prüfungskommission vorgelegt worden ist.

Hannover, den 31.3.99

(Frank Bretz)

Lebenslauf

Persönliche Angaben

Name:	Frank Bretz
Geburtstag:	14.10.71
Geburtsort:	Vlaardingen/NL
Familienstand:	Verheiratet seit 1993 mit Jiamei Zhou (Chinesin), ohne Kinder

Schulausbildung

10/74 bis 02/78	Kindergarten in Muroran/Japan
08/78 bis 08/80	Grundschule in Mülheim a.d. Ruhr
08/80 bis 12/80	Gemeinschaftsgrundschule in Mettmann
01/81 bis 12/88	Escola Corcovado in Rio de Janeiro/Brasilien
01/89 bis 07/91	Hans-Sachs-Gymnasium in Nürnberg;

Berufsausbildung

10/91 bis 05/96	Mathematikstudium mit Biologie als Nebenfach an der Universität Erlangen-Nürnberg;
10/96 bis 03/99	Promotion an der Universität Hannover, Fachbereich Gartenbau
seit 04/99	Informatiker in der klinischen Forschung bei Byk Gulden, Konstanz