Robuste Mittelwertvergleiche mit gartenbaulichen Anwendungen

Vom Fachbereich Gartenbau
der Universität Hannover
zur Erlangung
des Grades eines

Doktors der Gartenbauwissenschaften
-Dr.rer.hort.-

genehmigte

Dissertation

von

Dipl.-Math.oec. **Michael Weichert**
geboren am 31.05.1970 in Bonn

2000

Referent:	Prof. Dr. L.A. Hothorn (Hannover)
Korreferent:	Prof. Dr. G. Hommel (Mainz)
Tag der Promotion:	21.1.2000

This one goes out to the one I love
This one goes out to the one I've left behind
A simple prop to occupy my time
This one goes out to the one I love

Fire (she's comin' down on her own, now)

This one goes out to the one I love
This one goes out to the one I've left behind
Another prop has occupied my time
This one goes out to the one I love

Songtext von R.E.M.

Abstract

Experimentelle Daten erfüllen häufig nicht die strengen Modellannahmen der Varianzanalyse. So treten zum einen (scheinbar) zu große oder zu kleine Werte auf, oder die Verteilung der Daten ist schief. In solchen Situationen sind Standardtests, die auf Mittelwerten und Standardabweichung beruhen, stark verzerrt. Unter einer kontaminierten Normalverteilung beispielsweise, also einer Verteilung die hauptsächlich eine Normalverteilung ist und vereinzelt „sehr“ große bzw. kleine Werte zeigt, ist der t-Test bei einem Unterschied von 2,1 Standardabweichungen in nur 30% aller Fälle in der Lage, überhaupt einen Unterschied festzustellen. Trotz dieser Problematik wird häufig auf die Standardtests und -Verfahren zurückgegriffen, da diese softwaretechnisch verfügbar sind und auch die Schätzung der absoluten Effekte (Mittelwerte) ermöglichen.

Das Ziel dieser Dissertation besteht darin, robuste Testverfahren für multiple Kontrasthypothesen zu entwickeln, die zudem eine Schätzung der interessierenden Parameter ermöglichen. Es wird insbesondere auf das Testen einseitiger paarweiser Hypothesen eingegangen.
Basierend auf drei robusten Schätzern, nämlich getrimmtes Mittel, Huber-m-Schätzer und Tiku MML-Schätzer, werden robuste Zwei-Stichproben-Tests zuerst diskutiert und dann auf den k-Stichproben-Fall verallgemeinert. Für die ersten beiden wird dabei gezeigt, daß sie k-variat normalverteilt sind. Im Rahmen einer Simulationsstudie wird das Ausmaß geprüft, in dem sich bei kleinen Fallzahlen multiple Kontraste dieser beiden Schätzer durch eine multivariate t-Verteilung approximieren lassen. Weiterhin werden diese multiplen Kontrasttests mit Rangverfahren und einem Bootstrap-Verfahren bezüglich Niveau und Güte unter verschiedenen Verteilungen verglichen.
Im Fall von paarweisen Kontrasten läßt sich auch die Wahrung des multiplen Niveaus über die Konstruktion eines Abschlußtests (AT) und dem sich daraus ergebenen schrittweisen Testverfahren gewährleisten. Hierbei stellt die Bestimmung der für den AT relevanten Mengen (streng erschöpfende Indexmengen) das Kernproblem dar. Zur Bestimmung dieser Mengen wird das Hypothesensystem der paarweisen Kontraste auf einen Graphen abgebildet. Anhand dieses Graphen läßt sich entscheiden, ob eine Menge für den AT relevant ist. Die graphentheoretische Herangehensweise bietet zudem die Möglichkeit, auch verschobene Hypothesen, insbesondere Äquivalenzhypothesen, mit in das Hypothesensystem aufzunehmen.

Die vorgestellten Verfahren werden jeweils an praktischen Beispielen demonstriert. Schwerpunkt liegt dabei auf der Bestimmung des höchsten effektiven Dosisschritts in einem Dosis-Wirkungs-Versuch als eine Anwendung einseitiger Hypothesensysteme. Zu den Verfahren sind SAS/IML MACROS erstellt worden, die in einem gesonderten Kapitel dokumentiert sind.

Schlagworte: *multiple Kontraste, robust, einseitige Hypothesen*

Abstract

Experimental data often do not fulfill the strict assumptions of variance analysis. Within a dataset some values are obviously to high or to small or the distribution of the data is skewed. In situations like these the use of standard-tests, which are based on means and standard deviation, is not appropriate. Under a contaminated normal distribution, for example, which is a distribution mostly normal and sporadically showing very high or small values, the t-test will state a difference in means in only 30% of all cases even when the actual difference is 2.1 sigma. Despite of this standard-tests and methods are used very often because of their availability in software-packages and the possibility of estimating effects (means).

The aim of this thesis is to develop robust tests for multiple contrast hypotheses. These tests in addition should allow an estimation of the interesting parameters. Special focus is put on testing one-sided pairwise hypotheses. Starting out from robust estimates like the trimmed mean, the Huber-m-estimator, and Tiku's MML-estimator robust two-sample-tests are discussed. These are then generalized to the k-sample-situation. For the trimmed mean and the Huber-m-estimator a multivariate limit theorem is proved. Via a simulation study a multi-t-approximation for small sample sizes is studied. Moreover, these robust multiple contrast-tests are compared to rank-tests and a bootstrap-test under a variety of distributions.
Looking at pairwise contrasts the multiple level can also be controlled by constructing a closure test and carrying out the consequent stepwise test procedure. The problem in constructing such a closure test system is to determine the relevant indexsets (strong exhaustive indexsets). To determine these indexsets the system of hypotheses is mapped into a graph. This graph will then determine whether an indexset is strong exhaustive or not. As an additional feature the graph-representation allows to consider one sided shifted hypotheses and particularly equivalence hypotheses, too.

The presented methods are demonstrated by practical examples. In particular the determination of the highest effective dose-step in a dose-response-analysis is presented as an application of one-sided hypotheses. The SAS/IML MACROS developed for the described methods and procedures are presented in separate chapter.

Keywords: *multiple contrasts, robustness, one-sided hypotheses*

Inhaltsverzeichnis

Abkürzungen und Symbole

Abkürzungen	CAU	Cauchyverteilung
	KNV	kontaminierte Normalverteilung
	LogNV	logarithmische Normalverteilung
	oBdA	ohne Beschränkung der Allgemeinheit
	W!	Widerspruch
	MED	minimal effektive Dosis
	HEDS	höchste effektive Dosis
Graphen	G	Graph
	$\vec{G}$	gerichteter Graph, Digraph
	$R(\cdot), \dot{R}(\cdot), \dot{R}_g(\cdot)$	Erreichbarkeitsmatrizen
Hypothesen	$H_.$	Nullhypothese
	$H_{i)}$	i-te nach p-Werten geordnete Nullhypothese
	$H_{\langle i,j \rangle}$	Nullhypothese für $\mu_i \geq \mu_j$
	$H_{[i,j]}$	Nullhypothese für $\mu_i \geq \mu_j + \rho\delta$, mit $\rho \in \{0,1\}$
	$A_{\langle i,j \rangle}$	Alternative $\mu_i > \mu_j$
	$H_{[i,j]}$	Alternative für $\mu_j < \mu_i + \rho\delta$, mit $\rho \in \{0,1\}$
Indexmengen	$\wp_l$	Potenzmenge von $\{1, \ldots, l\}$
	$\wp_l^0$	$\wp_l \setminus \emptyset$
	$\wp_{EI}$	Menge der erschöpfenden Schichten
	$\wp_{EI}^s$	Menge der streng erschöpfenden Schichten
	i^*	Maximales Element von I
	i^{**}	zweitgrößtes Element von I
	$SE(I)$	minimale I umfassende erschöpfende Indexmenge
	$\lvert \cdot \rvert$	Betrag; bei Mengen die Anzahl der Elemente
Operatoren	$\oplus$	(erweiterte) Boolsche Addition
	$\otimes$	(erweiterte) Boolsche Multiplikation
	$\times$	elementweise Boolsche Multiplikation von Matrizen
	$\oslash$	elementweise Boolsche Addition von Matrizen
	$\odot$	„Multiplikationsoperator“ für Matrizen
	$\bigcup$	Vereinigung
	$\dot{\bigcup}$	disjunkte Vereinigung
Parameter	μ, σ	Erwartungswert, Standardabweichung
	$\mu_t, \hat{\mu}_t$	getrimmter Erwartungswert, Schätzer für den getrimmten Erwartungswert
	σ_w^2, s_w^2	winsorisierte Varianz, Schätzer für die winsorisierte Varianz
	$\mu_\psi, \hat{\mu}_\psi$	m-Lokation, m-Lokationsschätzer
	σ_ψ^2, s_ψ^2	m-Skala, m-Skalenschätzer
	$\hat{\mu}_{MML}, \sigma_{MML}^2$	Lokations- und Skalenschätzer nach Tiku
Symbole	1_n	Vektor aus n Einsen
	∞	Unendlich
	α	Niveau des Tests
	γ	Trimmanteil
	C^3	Klasse der dreifach stetig differenzierbaren Funktionen

	E_k	k-dimensionale Einheitsmatrix
	O	Nullmatrix
	$\mathbb{N}$	Menge der natürlichen Zahlen
	$\mathbb{N}_l$	$\{1, \ldots, l\}$
	$\mathbb{R}$	Menge der reellen Zahlen
	$\mathbb{R}^k$	k-dimensionaler Vektorraum über $\mathbb{R}$
	$\mathbf{E}[]$	Erwartungswert
	S^{k-1}	Einheitskugel im R^k
	$\|\| \cdot \|\|$	Norm
	p_i	p-Wert zur Hypothese H_i
	ap_i	adjustierter p-Wert zur Hypothese H_i
	$p_{(i)}$	i-ter geordneter p-Wert
	$p_t, p_{t10}, p_{t20}, p_c$	p-Werte des t-, 10% getrimmten-, 20% getrimmten-Tests und des Huber-m-Tests
	t_{min_t}	$\min\{p_t, p_{t10}, p_{t20}\}$, bzw. $\min\{p_t, p_{t20}\}$
	t_{min_c}	$\min\{p_t, p_{c_{1,8}}\}$
	$t_{min_{tc}}$	$\min\{p_t, p_{t20}, p_{c_{1,8}}\}$
	$\xrightarrow{\mathcal{D}}$	konvergiert in Verteilung
	$.'$	Transposition
Verteilungen	$\mathcal{N}(\mu, \sigma^2)$	Normalverteilung mit Erwartungswert μ und Varianz σ^2
	$\mathcal{X}_2^2$	Chiquadratverteilung mit Freiheitsgrad 2
	$\Phi(\cdot)$	Verteilungsfunktion der Standardnormalverteilung
	$\Phi^{-1}(\cdot)$	Quantilfunktion der Standardnormalverteilung
	ν	Freiheitsgrade
	R	Korrelationsmatrix
	$t_{\nu,1-\alpha}$	$1-\alpha$ Quantil der t-Verteilung mit ν Freiheitsgraden
	$T_l(\infty, 1_l \cdot t, \nu, R)$	Verteilungsfunktion der l-variaten t-Verteilung mit Korellationsmatrix R und Freiheitsgrad ν an der Stelle t

Kapitel 1

Problemdarstellung

1.1 Einleitung

In der gartenbaulichen, pharmazeutischen und industriellen Praxis ist es Ziel, durch Versuche Einblick in die Effekte einer Behandlung oder Methode zu gewinnen. Neben der Bestimmung der Effekte und ihrer Größe ist es meist auch von Interesse, die verschiedenen Behandlungen oder Methoden gegeneinander abzugrenzen. Um solch eine Differenzierung zu systematisieren, wird für jeden interessierenden Vergleich eine statistische Hypothese aufgestellt. Werden z.B. zwei verschiedene Dünger hinsichtlich ihres Einflusses auf den Ertrag von Weizen untersucht, so interessiert, ob einer der beiden einen höheren Ertrag als der andere liefert. Um diese Entscheidung gesichert treffen zu können, wird die Hypothese aufgestellt, daß beide Dünger den gleichen Effekt haben. Wird diese Hypothese anhand des gewählten Tests verworfen, so ist ein signifikanter Unterschied zwischen den Düngern festgestellt worden[1].
Bei solch einer Vorgehensweise kann es vorkommen, daß man einen Unterschied entdeckt, obwohl in „Wahrheit“ keiner vorliegt. Ziel des Testverfahrens ist es, solche Fehler zu kontrollieren. In den meisten Versuchsanlagen werden nicht nur zwei verschiedene Behandlungen oder Sorten untersucht, sondern $k > 2$ verschiedene, wobei eine oder mehrere Kontrollen in der Anlage vorkommen können. Würde jeder einzelne der l interessierenden Vergleiche zu einem lokalen Niveau von $\alpha_{lokal} = 5\%$ gemacht, so läge die Wahrscheinlichkeit, falls keine Effekte auftreten, mindestens einen nicht existenten Unterschied zu finden, bei $1 - (0,95)^l\%$ (für $l = 5$ Vergleiche bei ca. 22%), gegeben daß die durchgeführten Vergleiche unabhängig sind . Eine multiple Vergleichsprozedur zeichnet sich dadurch aus, daß sie u.a. diesen globalen Fehler α_{global} kontrolliert. Eine einfache Lösung dieses Problems bietet die Bonferroni-Adjustierung, für die bei l Vergleichen das lokale α zu α_{global}/l gesetzt wird.

[1] Popper'sches Falsifikationsprinzip

Jedoch wird durch die Absenkung des Fehlers erster Art gleichzeitig die Wahrscheinlichkeit größer, einen wirklich vorliegenden Effekt nicht zu entdecken. Daher gilt es bei der Wahl des lokalen α einen möglichst hohen Wert zu wählen, so daß das globale Niveau den gewünschten Wert von α_{global} nicht übersteigt[2]. Die Lösungsansätze hierzu lassen sich grob in zwei Klassen von Verfahren einteilen, die Einschritt- und die Mehrschritt-Verfahren.
In einem Einschritt-Verfahren werden alle Tests simultan durchgeführt, und alle Testentscheidungen gleichzeitig getroffen. Demgegenüber werden bei einem Mehrschritt-Verfahren die Tests in einer bestimmten Reihenfolge nacheinander durchgeführt, wobei jede einzelne Testentscheidung von den vorhergegangenen abhängen kann. Dies kommt daher, daß bei den Mehrschritt-Verfahren die logischen Abhängigkeiten der Hypothesen untereinander genutzt werden, um das globale Niveau besser ausschöpfen zu können. Klassische Einschritt-Verfahren sind der Tukey-Test und der Dunnett-Test unter den Annahmen des ANOVA Modells oder auch die Bonferronisierung. Beispiele für Mehrschritt-Verfahren sind die α-Adjustierung nach Holm oder der Abschlußtest.
Bei gartenbaulichen oder pharmazeutischen Versuchen ist es häufig möglich, die Testalternative einzuschränken. In einem Dosisversuch für einen Dünger oder ein Herbizid mit einer Kontrolle interessiert den Versuchsansteller zum Beispiel nur, ob der Dünger ertragssteigernd ist, bzw. das Herbizid die Anzahl der Unkräuter reduziert. Die Alternative vereinfacht sich so zu einer einseitigen Fragestellung, d.h. es gilt nur zu entdecken, ob eine Effektsteigerung (Düngung) oder eine Effektsenkung (Unkraut) eintritt. Um einen multiplen Vergleich in einem Mehrschritt-Verfahren durchführen zu können, gilt es, die logischen Abhängigkeiten zu bestimmen. Für zweiseitige Hypothesen ist dieses Problem von Bernhard (1991) und Westfall (1997) gelöst worden. In dieser Arbeit wird eine Lösung für die in der Praxis wichtigeren einseitigen Hypothesen hergeleitet.

Die Standardverfahren zur Auswertung von multiplen Vergleichen basieren auf den Annahmen des klassischen ANOVA Modells. In diesem Modell wird gefordert, daß die Daten i) normal verteilt, ii) unkorreliert und iii) varianzhomogen sind und auf einer randomisierten Versuchsanlage basieren. In der gartenbaulichen und industriellen Realität sind hingegen die Bedingungen i) und iii) selten erfüllt. So sind industrielle Meßdaten nur zu ca. 2% normalverteilt[3]. Häufig sind in den Datensätzen abgesehen davon, daß keine Normalverteilung vorliegt, auch scheinbar zu große oder zu kleine Werte zu finden. Solche Ausreißer sind in wissenschaftlichen Routinedaten zu 1% bis 10%, in medizinischen sogar bis zu 25% enthalten (Hampel 1995). Ausreißer und Verteilungsab-

[2]Eine exakte Definition der verschieden Niveaus und Fehler ist in Abschnitt 1.4 gegeben.
[3]Ergebnis einer Studie bei Daimler-Chrysler, vorgetragen auf der MTQ-Messe Dortmund 1999.

weichungen treten im Gartenbau selbst bei einfachen Dosis- und Behandlungsversuchen auf. In Abbildung 1.1 ist der Effekt von 3 verschiedenen Dosen und einer Kontrolle auf den Durchmesser von Radieschen in Form von Boxplots gezeigt.

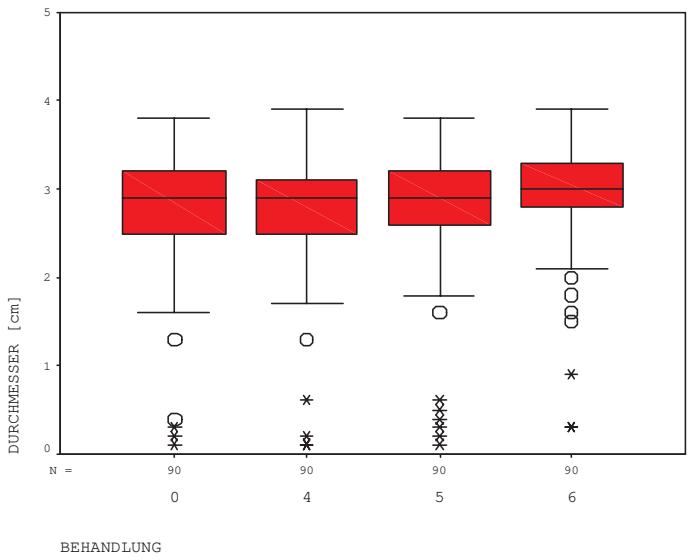


Abbildung 1.1: Durchmesser von Radieschen bei 3 verschiedenen Behandlungen und einer Kontrolle; Studentenversuch 4. Semester, FB Gartenbau

In einem Experiment zum Vergleich der Samenproduktion von Kohl mit sechs verschiedenen Behandlungsmethoden und vier Kontrollvarianten ergaben sich die durch den Boxplot in Abbildung 1.2 dargestellten Samenmengen in Tonnen pro $\frac{1}{4}$ha. (Daten aus Mead et al 1993). Neben den auftretenden Ausreißern kann auch nicht von gleichen Varianzen unter den verschiedenen Behandlungen ausgegangen werden.

Will man durch einen Vortest erst überprüfen, ob die Normalverteilungsannahme zutrifft, so stößt man auf zwei Probleme: Zum einen ist bei den üblichen Fallzahlen von bis zu 20 Beobachtungen je Gruppe die Wahrscheinlichkeit einer Fehlentscheidung zu groß, um überhaupt gesichert auf Normalverteilung schließen zu können. Zum anderen handelt es sich im eigentlichen Sinne um einen Äquivalenztest, da gezeigt werden soll, daß Normalverteilung vorliegt, was seinerseits eine noch höhere Fallzahl erfordert. Daher ist eine solche Vorgehensweise nicht realistisch.
Unter solchen Gegebenheiten weisen die auf den ANOVA-Annahmen basierenden Tests gravierende Probleme auf. So über- oder unterschreiten sie teilweise das vorgegebene Niveau, wobei im vorhinein nicht zu sagen ist welcher Fall auftreten wird. Außerdem fällt die Güte teilweise in drastischem Umfang ab. Einen Ausweg bietet die Verwendung robuster Verfahren. Diese zeichnen sich dadurch aus, daß sie zum einen unter Normal-

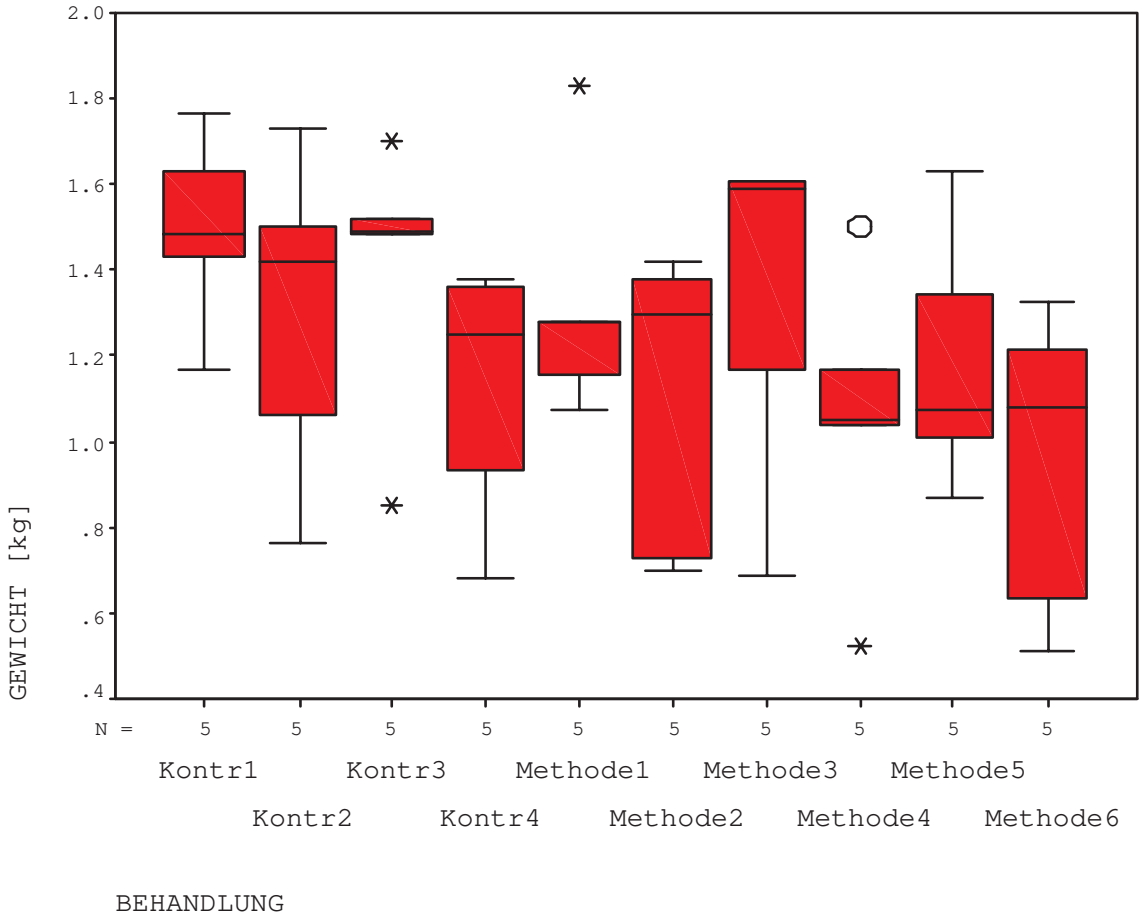


Abbildung 1.2: Samenproduktion von Kohl mit vier verschiedenen Kontrollvarianten, Quelle: Institut für Gemüsebau, persönliche Mitteilung

verteilung mit den Standardtests bezüglich der Einhaltung des Niveaus und der Güte konkurrieren können. Zum anderen sind diese Tests sowohl gegenüber Abweichungen von der Normalverteilung als auch gegenüber Ausreißern unempfindlich. D.h. sie halten weiterhin das Niveau ein und weisen eine weiterhin hohe Güte auf

Zur Lösung der oben aufgezeigten Probleme gliedert sich die Arbeit folgendermaßen: Diese Kapitel wird zunächst mögliche Ansätze zur Robustifizierung von multiplen Vergleichen mit ihren wesentlichen Charakteristika vorstellen. Danach werden die Ausreißermodelle vorgestellt, die in den Simulationen zur Untersuchung des Verhaltens der einzelnen Verfahren verwendet wurden. Die Einleitung schließt ab mit einer kurzen Einführung in die Theorie des multiplen Testens insbesondere unter Verwendung multipler Kontraste.
In Kapitel 2 sind anhand von drei gartenbaulichen Beispielen die auftretenden Probleme bei der Datenkondition ausführlich aufgezeigt. Kapitel 3 stellt vor, wie sich robuste multiple Mittelwertvergleiche durchführen lassen.Eine Möglichkeit, multiple (robuste)

Vergleiche durchzuführen besteht insbesondere darin, nur paarweise Vergleiche zu betrachten. Um nun die Einhaltung des multiplen Niveaus zu gewährleisten, müssen die sich aus den Vergleichen ergebenden p-Werte adjustiert werden. Eine optimale Adjustierung läßt sich durch die Verwendung des Abschlußtests erreichen. Für den Fall, daß alle Hypothesen einseitig formuliert sind, wird in Kapitel 4 gezeigt, wie sich der Abschlußtest durchführen und algorithmisieren läßt. Anschließend werden in Kapitel 5 die vorgestellten Verfahren an den Einführungsbeispielen demonstriert. Eine abschließende Betrachtung der Ziele dieser Arbeit und deren Lösungswege sowie Ausblicke auf weitere Fragestellungen gibt Kapitel 6. Die entwickelten SAS/IML Programme sind in Kapitel 7 dokumentiert.

1.2 Robustifizierungsansätze

Will man ein robustes Verfahren zur Auswertung der Daten verwenden, so bieten sich drei Hauptrichtungen an. Dies sind zum ersten parametrische Verfahren, die anstelle von Mittelwert und Standardabweichung robuste Schätzer für Lage und Streuung (location and scale) verwenden. Die davon abgeleiteten Testverfahren und Hypothesen unterscheiden sich von denen unter den Annahmen des ANOVA Modells nur durch die Verwendung dieser Schätzer. Zum zweiten gibt es die Rangverfahren, welche die Daten in Ränge transformieren und mit diesen Werten arbeiten. Die dritte und neueste Hauptrichtung sind Resampling-Verfahren, bei denen über eine Simulation der Verteilung der Teststatistik die Entscheidung über die Ablehnung der betrachteten Hypothese getroffen wird.
Diese drei Ansätze unterscheiden sich theoretisch in den von ihnen getesteten Hypothesen und den zugrundeliegenden Modellannahmen. In den parametrischen Verfahren werden Hypothesen über die Erwartungswerte (oder andere Parameter) der den Daten zugrundeliegenden Verteilung getestet. So lautet im Zweistichproben-Fall eine Hypothese z.B.: $\mu_1 = \mu_2$, wobei μ_1 und μ_2 die Erwartungswerte der Verteilungen von Gruppe 1 bzw. Gruppe 2 sind. Insbesondere werden bei robusten parametrischen Verfahren die Lage und die Streuung geschätzt, womit direkt auch Schätzer für die beobachteten Effekte vorliegen, wohingegen bei den Rangverfahren die Hypothesen bezüglich der Verteilungsfunktionen formuliert sind. In obigem Beispiel also $F_1 = F_2$, wobei F_1 und F_2 die Verteilungsfunktionen von Gruppe 1 bzw. Gruppe 2 sind. Nur für den Spezialfall von zwei Gruppen, in dem den Daten das Lokations-Skalen Modell unterstellt wird, d.h. den Daten, die gleiche Verteilungsfunktion mit unterschiedlicher Lage und/oder Streuung haben, sind die Hypothesen aus den parametrischen und den Rangverfahren äquivalent. Bei den Resampling-Testverfahren werden im allgemeinen wie bei den

Rangverfahren Hypothesen bezüglich der Verteilungsfunktionen getestet. Eine Bestimmung der Effekte aus diesen Testverfahren ist meist nicht direkt möglich, sondern erfordert ein gesondertes (Resampling-) Verfahren.
An die robusten Verfahren wird natürlich auch der Anspruch gestellt, im Falle von normalverteilten Daten nicht wesentlich schlechter zu sein als die hier optimalen Standardverfahren. Vergleicht man im Zwei-Stichproben-Fall den t-Test mit dem U-Test, so ergibt sich die relative Effizienz des U-Tests zu $3/\pi = 0,955$. Dies stellt für die robusten parametrischen Verfahren eine asymptotisch untere Grenze für ihre Effizienz dar. Im Endlichen ergibt sich simuliert eine Effizienz von $\approx 0,934$, bei einer Fallzahl von 10 Beobachtungen pro Gruppe und einem Lageunterschied von 1.18 Einheiten.
Unter der Zielvorgabe, multiple Mittelwertvergleiche durchzuführen und dabei auch Schätzer für die einzelnen Effekte zu erhalten, bieten sich somit die Verfahren an, die auf robusten Schätzern basieren. Daher werden in Kapitel 3 zuerst verschiedene robuste Schätzer eingeführt. Im zweiten Abschnitt werden daraus konstruierbare Zwei-Stichproben-Tests diskutiert, die sich z.B. über eine p-Wert Adjustierung zu einem multiplen Test kombinieren lassen. Der dritte Abschnitt behandelt dann robuste multiple Vergleiche, basierend auf multiplen Kontrasten, adjustierten Zwei-Stichproben p-Werten, Rangverfahren und multiple Bootstrap-Verfahren. Hierbei liegt das Hauptgewicht der zu den Abschnitten zwei und drei gemachten Untersuchungen auf den robusten parametrischen Verfahren, die abschließend mit den anderen robusten Verfahren in Simulationsstudien verglichen werden.

1.3 Ausreißermodelle

Obwohl es eine Vielzahl von Arbeiten zum Thema Ausreißer gibt, ist der Begriff des Ausreißers in einem Datensatz bis heute nicht klar und einheitlich definiert (Schultze 1997, Pigeot-Kübler 1993). Die intuitive Vorstellung von Ausreißern basiert darauf, daß Werte, die „verdächtig“ weit von der Masse der Daten entfernt liegen oder sich in „überraschender“ Weise vom Rest der Daten abheben, als Ausreißer zu erachten sind (Hawkins 1980, Barnett, Lewis 1984). Solch eine Definition ist kritisch zu betrachten, da sie stark subjektiv ist. Eine so subjektive Verfahrensweise findet sich auch in dem Übersichtsartikel von Beckman und Cook (1983), in dem Beobachtungen, welche nach Ansicht des Wissenschaftlers außerhalb der Hauptmasse der Daten liegen, als Ausreißer angesehen werden. Hierbei stützt sich die Beurteilung des einzelnen Forschers auf einen häufig unbewußt unterstellten, die Daten generierenden Zufallsmechanismus. Tritt bei einem Experiment ein Wert größer als vier auf, so würde er als Realisierung einer Standardnormalverteilung eher überraschen, wohingegen solch ein Wert bei einer

Cauchy-Verteilung nicht ungewöhnlich ist.
In der Literatur werden hauptsächlich zwei verschiedene Modelle der Ausreißerentstehung angenommen. Das eine Modell geht davon aus, daß die erhobenen Daten normalverteilt sind, jedoch werden einzelne Werte von einer abweichenden Verteilung generiert. In der Praxis entspricht das einem Meßverfahren und Experiment, in dem das beobachtete Merkmal normalverteilt ist, jedoch einige unbemerkte Meß- oder Ablesefehler auftreten. Das andere Modell ist eine Mißspezifikation der Verteilung, was bedeutet, daß von einer Normalverteilung ausgegangen wird, jedoch in Wirklichkeit eine andere Verteilung den Daten zugrundeliegt. Die Möglichkeiten, Daten entsprechend dieser beiden Modelle zu erzeugen, sind beliebig. So finden sich in verschiedensten Arbeiten immer wieder verschiedene Verteilungen, mit denen Daten für Simulationsstudien generiert werden.
Aus dieser Vielzahl sind für diese Arbeit einige häufig verwendete herausgegriffen, anhand derer die einzelnen robusten Verfahren in Simulationsstudien verglichen werden. Die Auswahl ist so getroffen, daß ein möglichst weites Spektrum an Nicht-Normalitäten abgedeckt wird. Ausgewählt sind neben der Normalverteilung die kontaminierte Normalverteilung, die Lognormal-, die Cauchy- und die $\mathcal{X}_2^2$-Verteilung, die u.a. von Yuen et al (1985) und Aboukalam (1992) verwendet wurden. Außerdem sind noch drei Vertreter der g-h-Verteilungsfamilie (Hoaglin 1985), die u.a. von Wilcox (1994) verwandt wurden, aufgenommen. Die einzelnen Verteilungen werden im folgenden anhand von Histogramm und Boxplot kurz vorgestellt.
Diese Arbeit betrachtet nur Auswertungsmöglichkeiten für Experimente, die Haupteffekte wie beispielsweise die Erhöhung des Ertrags untersuchen. Hingegen stellen bei Sicherheitsstudien, z.B. Giftigkeit einer gentechnisch veränderten Sorte, „Ausreißer“ gerade die interessierenden Werte dar. Eine Eliminierung der Werte oder Robustifizierung der Auswertemethode ist in solchen Fällen ein kontraindiziertes Konzept; statt dessen könnten Mischverteilungen aus „Respondern“ und „non-Respondern“ betrachtet werden (vgl. Hothorn 1994).

Normalverteilung

Die Standardnormalverteilung $\mathcal{N}(0,1)$ hat die Dichtefunktion: $\phi(x) := \frac{1}{\sqrt{(2\pi)}} \mathrm{e}^{-\frac{1}{2}x^2}$. Bei

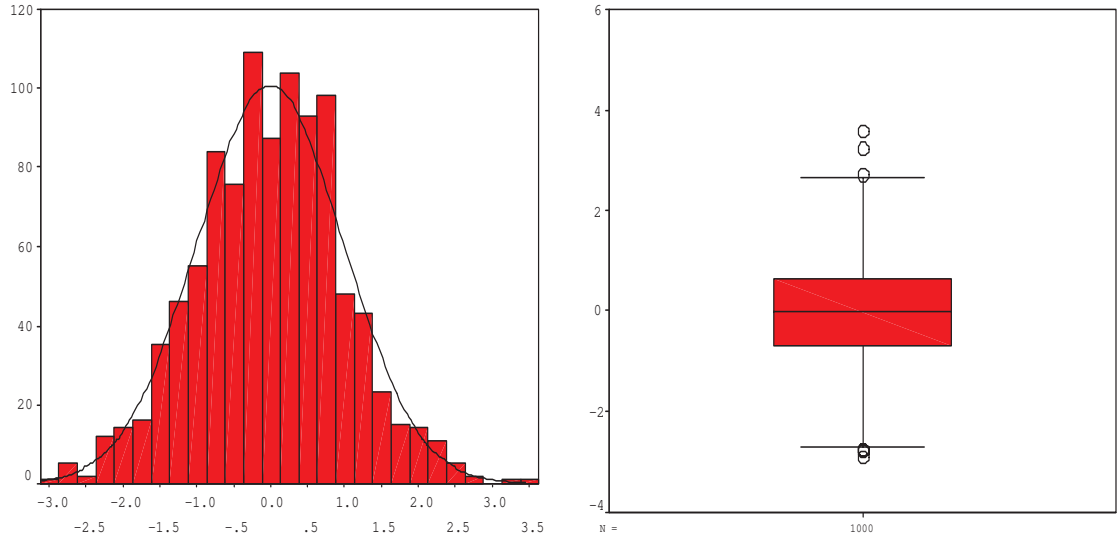


Abbildung 1.3: Histogramm und Boxplot von 1000 normalverteilten Datenpunkten

dem Boxplot fällt auf, daß einige Daten als Ausreißer gekennzeichnet sind (Kreise, Sterne). Aufgrund der Definition des Boxplots ist die Wahrscheinlichkeit für das Auftreten eines solchen „Ausreißers“ 0,7%. Daher ist mit 1000 Datenpunkten die Wahrscheinlichkeit, keinen „Ausreißer“ zu haben, fast 0.

Lognormalverteilung

Eine Zufallsvariable X heißt lognormalverteilt, wenn $\log(X)$ normalverteilt ist.

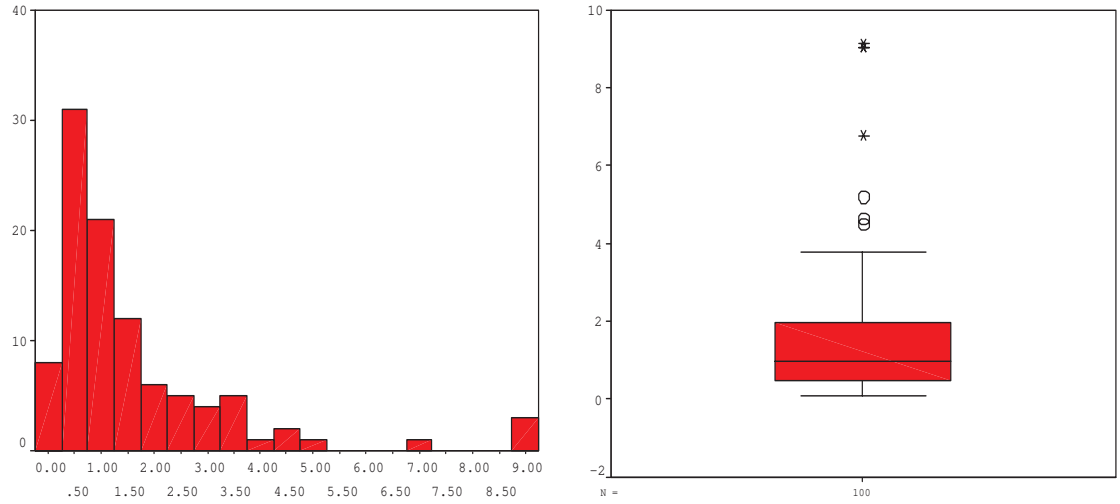


Abbildung 1.4: Histogramm und Boxplot von 100 lognormalverteilten Datenpunkten

Kontaminierte Normalverteilung

Eine kontaminierte Normalverteilung ergibt sich aus der Kombination zweier Normalverteilungen mit verschiedenen Varianzen. Hierbei werden o.B.d.A. $x\%$ der Daten von einer Standardnormalverteilung generiert und $1 - x\%$ von einer Normalverteilung mit einer Varianz größer eins. In den hier durchgeführten Simulationen ist $x = 80\%$ und die Standardabweichung der zweiten Normalverteilung 10.

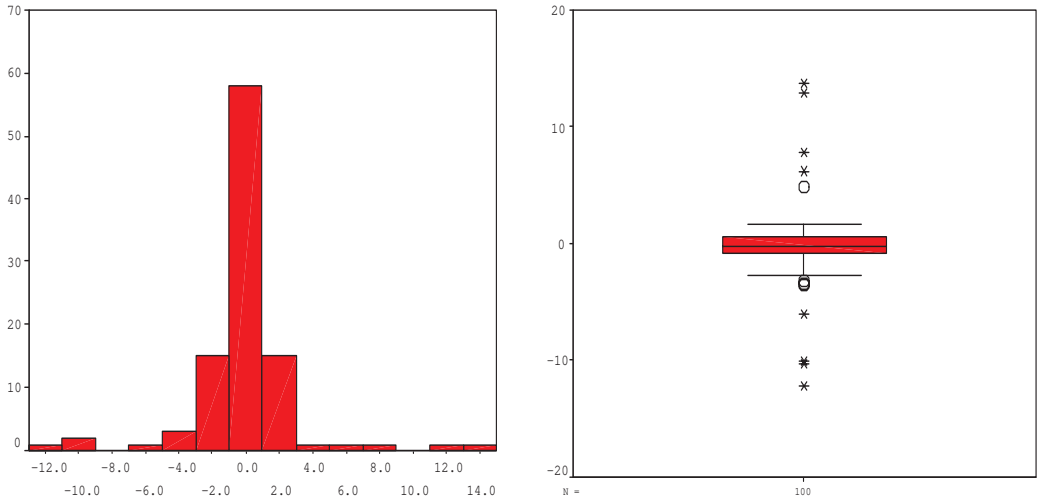


Abbildung 1.5: Histogramm und Boxplot von 100 kontaminiert normalverteilten Datenpunkten

Cauchy-Verteilung

Die Dichte der Cauchy-Verteilung ist $f(x) := \frac{1}{\pi(1+x^2)}$.

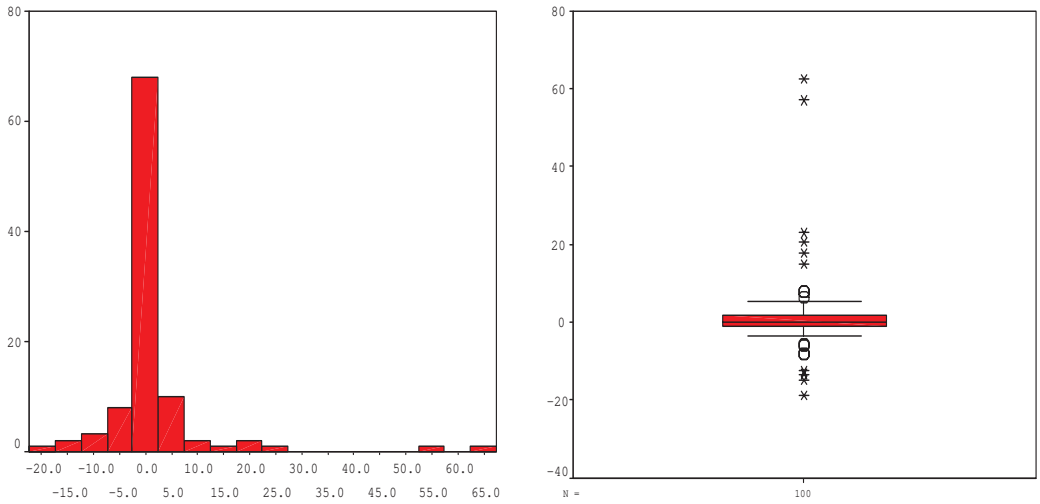


Abbildung 1.6: Histogramm und Boxplot von 100 cauchy-verteilten Datenpunkten

$\mathcal{X}_2^2$-Verteilung

Eine $\mathcal{X}_2^2$-verteilte Zufallsvariable erhält man zum Beispiel, indem man zwei unabhängig und identisch standardnormalverteilte Zufallsvariablen addiert.

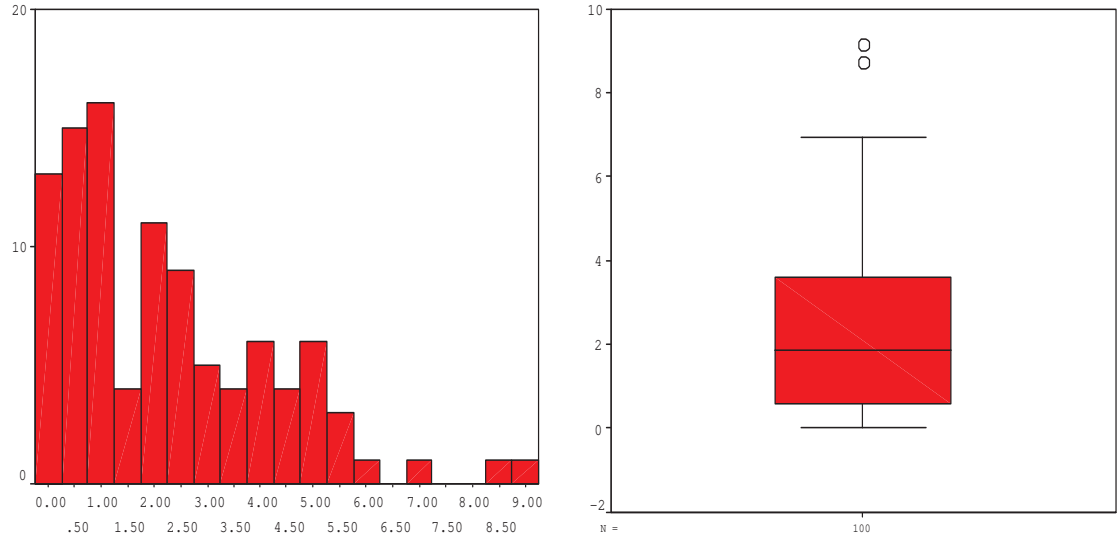


Abbildung 1.7: Histogramm und Boxplot von 100 $\mathcal{X}_2^2$-verteilten Datenpunkten

g-h-Verteilungsfamilie

Die g-h-Verteilungsfamilie ergibt sich durch Transformation der Standardnormalverteilung. Wenn X standardnormalverteilt ist, so ist die Variable

$$Y := \frac{e^{gX} - 1}{g} e^{\frac{hX^2}{2}} \tag{1.1}$$

g-h-verteilt, wobei g und h die Parameter der Verteilung sind. Ist $g = 0$, so wird der Grenzwert $Y = xe^{\frac{hX^2}{2}}$ von (1.1) genommen:

$\underline{(g;h) = (0;0,2)}$

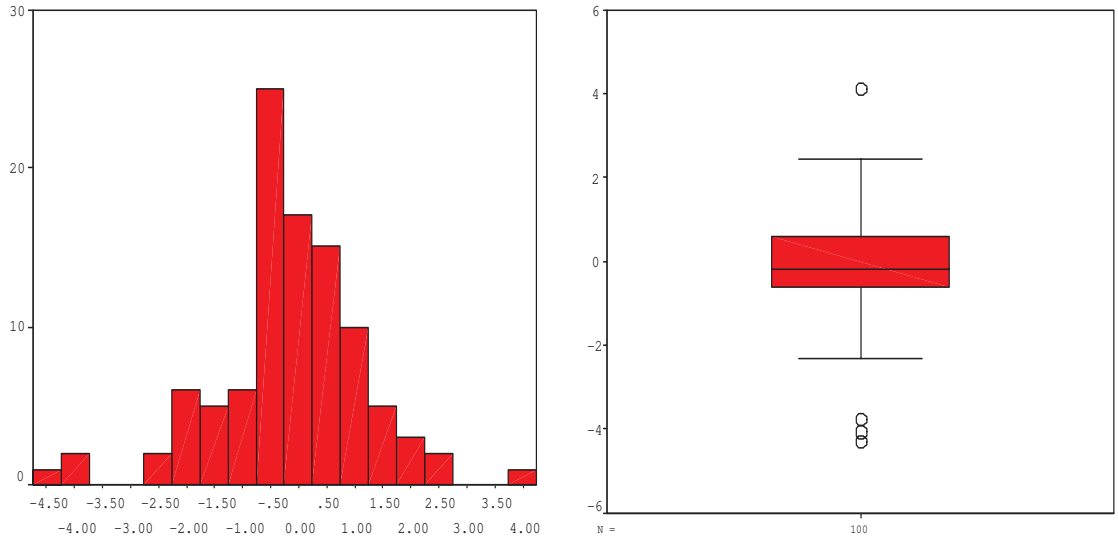


Abbildung 1.8: Histogramm und Boxplot von 100 $(g;h) = (0;0,2)$ verteilten Datenpunkten

$\underline{(g;h) = (0,5;0)}$

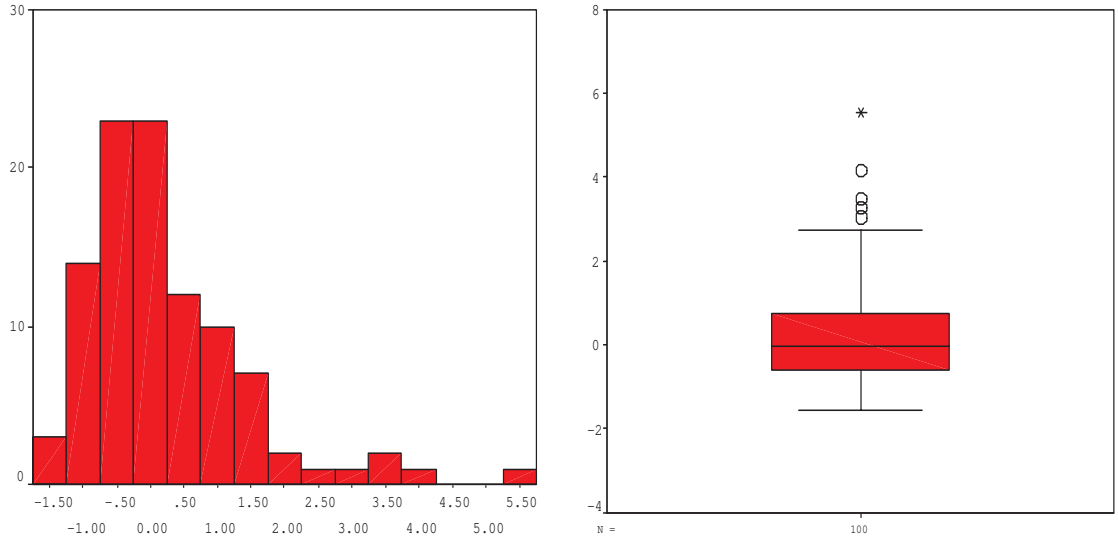


Abbildung 1.9: Histogramm und Boxplot von 100 $(g;h) = (0,5;0)$ verteilten Datenpunkten

$\underline{(g;h) = (0,5;0,2)}$

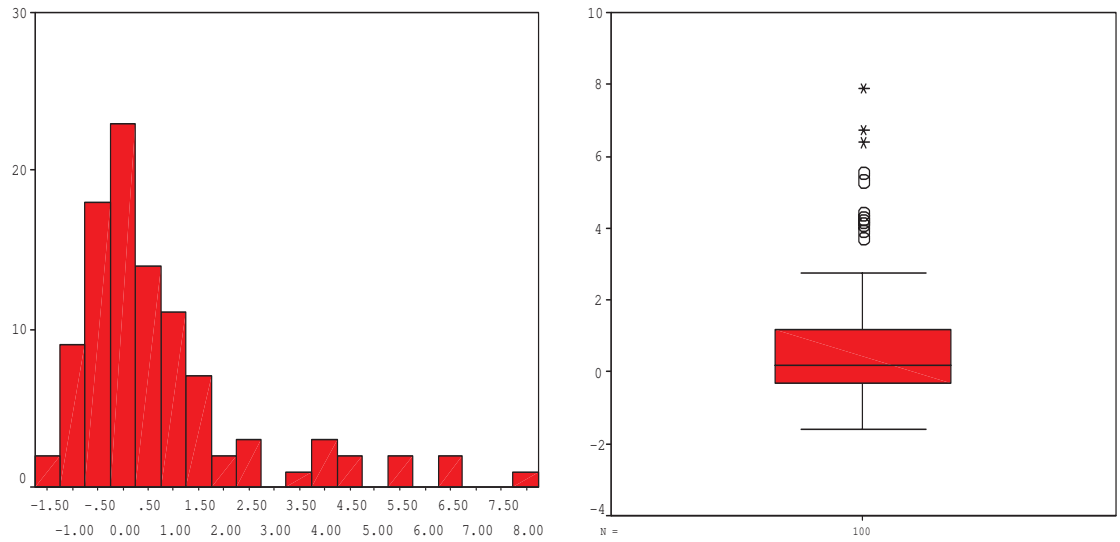


Abbildung 1.10: Histogramm und Boxplot von 100 $(g;h) = (0,5;0,2)$ verteilten Datenpunkten

1.4 Multiples Testen mittels multipler Kontraste

Wie in der Einleitung erwähnt, interessiert man sich im Gartenbau häufig für den Vergleich von zwei oder mehr Behandlungen. Der Effekt der Behandlung i wird hierbei durch den erwarteten Wert μ_i der Zielgröße unter der Behandlung charakterisiert. Stellt man eine Relation zwischen den verschiedenen erwarteten Werten auf, so bezeichnet man dies als Hypothese[4]. Ein Beispiel dafür wäre $\mu_1 = \mu_2$ oder $\mu_1 > \mu_2$, will man also bestimmen, ob der Effekt der Behandlung 1 gleich oder größer dem von Behandlung 2 ist. Im Bezug auf einen Test werden die Hypothesen unterschieden in Nullhypothesen und Alternativen. Im Beispiel der zwei Behandlungen interessiere, ob die Behandlung 1 einen größeren Effekt habe als die Behandlung 2. Als Hypothesen erhalten wir dann $\mu_1 \leq \mu_2$ und $\mu_1 > \mu_2$, wobei erstere die Nullhypothese H_0 und zweitere die Alternative H_A ist. Bei einem Test mit Nullhypothese und Alternative kann es passieren, daß der Test sich trotz Gültigkeit der Nullhypothese für die Alternative entscheidet. Solch einen Fehler nennt man *Fehler 1. Art*. Um sich gegen solche Fehler zu schützen, beschränkt man diese Irrtumswahrscheinlichkeit durch eine obere Schranke α. Diese Schranke α wird als Signifikanzniveau bezeichnet.

Liegt nun ein multiples Testproblem, d.h. ein Mehrhypothesenproblem vor, können von den einzelnen Nullhypothesen (Elementarhypothesen) mehrere wahr sein. Dabei kann jede einzelne der wahren Elementarhypothesen mit einer bestimmten Wahrscheinlichkeit irrtümlich abgelehnt werden. Der Begriff des Signifikanzniveaus läßt sich so in verschiedener Weise fassen:
Ist die Wahrscheinlichkeit, eine beliebig herausgegriffene Elementarhypothese fälschlicherweise abzulehnen, nicht größer als α, so sagt man, daß der entsprechende Test das *lokale Signifikanzniveau* einhält. Kontrolliert man nur das lokale Niveau, so hat man im allgemeinen keine Kontrolle über die Anzahl der gesamten Fehlentscheidungen (siehe Einleitung). Fordert man unter der Annahme, daß alle Nullhypothesen wahr sind, daß die Wahrscheinlichkeit, mindestens eine Nullhypothese abzulehnen, kleiner oder gleich α ist, so spricht man vom *globalen Niveau*. Läßt man noch die Bedingung fallen, daß alle Nullhypothesen wahr sein müssen, so gelangt man zum Begriff des *multiplen Niveaus*. Ein Testverfahren, bei dem die Wahrscheinlichkeit, mindestens eine wahre Nullhypothese abzulehnen durch α beschränkt ist, und zwar unabhängig davon, welche Nullhypothesen wahr sind, kontrolliert das *multiple Niveau*. Für die drei Niveaudefinitionen gilt, daß ein Verfahren, das das multiple Niveau einhält, auch das globale Niveau einhält. Weiter kontrolliert jedes Verfahren, welches das globale Niveau

[4]Die gesamte Arbeit bewegt sich im Kontext der Neyman Pearson Theorie.

hält, auch das lokale Niveau (Horn, Vollandt 1995). Die Umkehrungen dieser implizierten Niveaukontrollen gelten nicht. Neben der Fehlentscheidung, einen nicht existenten

Lokales Niveau α	Jede einzelne Elementarhypothese kann irrtümlich mit Wahrscheinlichkeit α abgelehnt werden
Globales Niveau α	Die Wahrscheinlichkeit, mindestens eine Elementarhypothese irrtümlich abzulehnen, ist höchstens α, vorausgesetzt, alle Elementarhypothesen sind wahr.
Multiples Niveau α	Die Wahrscheinlichkeit, mindestens eine Elementarhypothese irrtümlich abzulehnen, ist höchstens α, unabhängig von der Anzahl wahrer Elementarhypothesen.

Tabelle 1.1: Gegenüberstellung der drei Signifikanzniveau-Typen aus Horn, Vollandt (1995)

Effekt zu postulieren, interessiert auch, wahre vorliegende Effekte durch einen Test aufzudecken. Befindet man sich also unter der Alternative, d.h. es gibt einen Unterschied, so heißt der Fehler, sich jetzt für die Nullhypothese zu entscheiden, *Fehler 2. Art* und seine Wahrscheinlichkeit wird mit β bezeichnet.
Im allgemeinen ist jedoch nicht die Wahrscheinlichkeit des Fehlers 2. Art, sondern die *Güte* $= 1 - \beta$ eines Tests von Interesse. Dabei ist die Güte oder auch *Power* die Wahrscheinlichkeit, einen vorliegenden Effekt zu entdecken. Während bei einzelnen Tests die Power eindeutig ist, gibt es für multiple Testprozeduren verschiedene Powerkonzepte, vgl. z.B. Hochberg, Tamhane (1987): Für eine multiple Testprozedur, welche das multiple Niveau einhält, versteht man unter Power jeweils folgendes:

Lokale Power	die Wahrscheinlichkeit für die Ablehnung einer speziellen ausgewählten falschen Hypothese
Multiple Power	die Wahrscheinlichkeit für die Ablehnung mindestens einer der falschen Hypothesen
Globale Power (Totale Power)	die Wahrscheinlichkeit für die Ablehnung aller falschen Nullhypothesen
Erwartete durchschnittliche Power	die erwartete Anzahl von richtig abgelehnten falschen Nullhypothesen

Tabelle 1.2: Erweiterte Aufstellung der Powerarten aus Bernhard (1991)

Bei einem zweiseitigen Test, bei dem eine Nullhypothese der Form $\mu_1 = \mu_2$ getestet wird, kann noch ein weiterer Fehler auftreten. Sei wirklich $\mu_1 > \mu_2$ und der Test entscheidet, daß $\mu_1 < \mu_2$ ist, so entscheidet man sich richtig gegen die Nullhypothese, jedoch für die falsche Ordnung der Effekte. Solch einen Fehler nennt man *Fehler 3. Art*

(Bauer et al 1986).
Abhängig von der Fragestellung gibt es ein breites Sortiment an Verfahren, um multiple Tests, d.h. Test, die das multiple Niveau halten, durchzuführen. Will man alle Gruppen simultan miteinander vergleichen, so ist der Tukey-Test (1953) angebracht. Dies hat den Vorteil, daß mit „einem Test“ alle paarweisen Entscheidungen getroffen werden. Ist man von vornherein nur an einseitigen Vergleichen der einzelnen Behandlungsgruppen, wie zum Beispiel in einem Dosis-Wirkungs-Versuch, interessiert, so ist der One-Sided Studentized Range Test nach Hayter (1990) der Test der Wahl. Für einen Vergleich gegen eine Kontrolle eignet sich der Dunnett-Test (1955) in ein- oder zweiseitiger Form. Weiterhin gibt es Verfahren zur Bestimmung der besten Behandlung. Eine ausführliche Beschreibung dieser und weiterer Verfahren sind in Hochberg, Tamhane (1987) und Hsu (1996) beschrieben.
Wichtig ist, daß alle diese Verfahren von der Normalverteilung der Daten ausgehen. Wie unter den Einschritt-Verfahren gibt es auch unter den Mehrschritt-Verfahren eine große Anzahl verschiedener Methoden. Hier gilt jedoch, daß ein Teil der Verfahren „Abkürzungen“ des Abschlußtests sind. Daher und weil er die Grundlage des in Kapitel 4 verwendeten Verfahrens bildet, wird er im folgenden eingeführt.
Das Prinzip des Abschlußtests geht auf die Arbeit von Marcus et al (1976) zurück. Ausgehend von l zu testenden (Elementar-)Hypothesen $H_1, \ldots, H_l$ werden alle Schnitte dieser Elementarhypothesen gebildet. Die Entscheidung nach dem Abschluß-Prinzip verlangt nun, alle Elementarhypothesen sowie alle Schnitthypothesen zu testen. Hierbei ist jede der Hypothesen mit einem beliebigen Test zum Niveau α zu prüfen. Dies setzt voraus, daß insbesondere für die Schnitthypothesen geeignete Tests zur Verfügung stehen. Eine Elementarhypothese ist genau dann abzulehnen, wenn die Prüfung ihrer selbst, als auch jeder sie enthaltenden Schnitthypothese eine Ablehnung zum lokalen Niveau α ergab. Seien zum Beispiel vier verschiedene Sorten $(1, 2, 3, 4)$ hinsichtlich ihres Ertrags untereinander zu untersuchen. Die sechs zugehörigen Elementarhypothesen lauten dazu $H_{\{1,2\}} : \mu_1 = \mu_2$, $H_{\{1,3\}} : \mu_1 = \mu_3$, $H_{\{1,4\}} : \mu_1 = \mu_4$, $H_{\{2,3\}} : \mu_2 = \mu_3$, $H_{\{2,4\}} : \mu_2 = \mu_4$, $H_{\{3,4\}} : \mu_3 = \mu_4$. Als Schnitthypothesen ergeben sich: $\mu_1 = \mu_2 = \mu_3$, $\mu_1 = \mu_2 = \mu_4$, $\mu_1 = \mu_3 = \mu_4$, $\mu_2 = \mu_3 = \mu_4$, $\mu_1 = \mu_2\hat{\mu}_3 = \mu_4$, $\mu_1 = \mu_3\hat{\mu}_2 = \mu_4$, $\mu_1 = \mu_4\hat{\mu}_2 = \mu_3$, und $\mu_1 = \mu_2 = \mu_3 = \mu_4$. In diesem Fall könnte man für die „zusammenhängenden“ Schnitthypothesen jeweils F-Tests wählen und die Elementarhypothesen mit t-Tests prüfen. Eine formale Darstellung ist in Abschnitt 4.1.1 gegeben. Anstelle der F-Tests für die Schnitthypothesen ließen sich u.a. auch multiple Kontrasttests verwenden.

Kontraste

Ein Großteil multipler Tests läßt sich über sogenannte Kontraste bzw. aus ihnen konstruierte Tests beschreiben (Bretz 1999). Dafür gehe man von dem einfaktoriellen ANOVA-Modell mit festen Effekten aus:

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

Die X_{ij} seien unabhängig und identisch normalverteilte Beobachtungen mit unbekannten Erwartungswerten $\mu_1, \ldots, \mu_k$ und gemeinsamer Varianz σ^2. Hierbei ist n_i die Anzahl der Beobachtungen in Gruppe i. Gilt es, nun die Hypothese $H_1 : \mu_1 = \ldots = \mu_k$ zu testen, so läßt sich dieses anhand eines (single) Kontrasts tun, der definiert ist durch:

$$t^{SC} := \frac{\sum_{i=1}^{k} c_i \bar{X}_i}{s\sqrt{\sum_{i=1}^{k} \frac{c_i}{n_i}}} \sim t_{\nu,1-\alpha}. \qquad \text{mit } \sum_i c_i = 0 \tag{1.2}$$

Die Kontrastkoeffizienten können dabei so gewählt werden, daß bestimmte interessierende Alternativen gezielt aufgedeckt werden können. Solche Hypothesen nennt man auch Kontrasthypothesen. Da t^{SC} ein Quotient aus einer standardnormalverteilten und einer unabhängigen chiquadratverteilten Variable mit ν Freiheitsgraden ist, folgt, daß ein single Kontrast univariat zentral t-verteilt mit ν Freiheitsgraden ist. Zur einfachen Darstellung faßt man die Kontrastkoeffizienten in einem Vektor zusammen: $c = (c_1, c_2, \ldots, c_k)$.
Verwendet man nicht nur einen Kontrast sondern mehrere zum Testen der Hypothese H_1, so ist nach Bretz (1999) der Vektor der Teststatistiken multivariat t-verteilt, mit $\nu = \sum n_i - k$ Freiheitsgraden und Korrelationsmatrix R. Die Einträge der Korrelationsmatrix $R = (\rho_{ab})$ für l single Kontraste $(c_{1i}, \ldots, c_{ki})$, $i = 1, \ldots, l$ ergeben:

$$\rho_{ab} := \frac{\sum_{j=1}^{k} \frac{c_{ja} c_{jb}}{n_j}}{\sqrt{\left(\sum_{j=1}^{k} \frac{c_{ja}^2}{n_j}\right)\left(\sum_{j=1}^{k} \frac{c_{jb}^2}{n_j}\right)}}.$$

Verwendet man nun als Teststatistik das Maximum der Einzelteststatistiken, so spricht man von einem multiplen Kontrasttest.

$$t^{MC} := \max\{t_1^{SC}, \ldots, t_l^{SC}\}$$

Die Verteilung eines solchen multiplen Kontrasttests ergibt sich aus der Beziehung:

$$\begin{aligned} \mathrm{P}\left(t^{MC} \le t\right) &= \mathrm{P}\left(\max\{t_1^{SC}, \ldots, t_l^{SC}\} \le t\right) \\ &= \mathrm{P}\left(t_1^{SC} \le t, \ldots, t_l^{SC} \le t\right) \\ &= T_l(-\infty, 1_l \cdot t, \nu, R). \end{aligned}$$

Ein multipler Kontrasttest ist somit multivariat t-verteilt. Entsprechend ist das zu t^{MC} gehörige Isokoordinaten- Quantil der multivariaten t-Verteilung $1 - T_l(-\infty, 1_q \cdot t^{MC}, \nu, R)$.
So lassen sich z.B. die beiden im Gartenbau weit verbreiteten Standardtests von Tukey und Dunnett als multiple Kontrasttests schreiben. Den Tukey-Test erhält man durch die Verwendung der Kontraste:

$$\begin{aligned} c_{12} &:= (-1, 1, 0, \ldots, 0) \\ c_{21} &:= (1, -1, 0, \ldots, 0) \\ c_{13} &:= (-1, 0, 1, 0, \ldots, 0) \\ c_{31} &:= (1, 0, -1, 0, \ldots, 0) \\ &\vdots \\ c_{kk-1} &:= (0, \ldots, -1, 1) \\ c_{k-1k} &:= (0, \ldots, 1, -1). \end{aligned}$$

Die Darstellung des Tukey-Tests durch multiple Kontraste hat sogar den Vorteil, daß im Fall von unbalancierten Daten ($n_i \neq n_j$) nicht die konservative Tukey-Kramer-Approximation verwendet werden muß.
Der Vergleich mit der Kontrolle (Test von Dunnett) ergibt sich aus den Kontrasten:

$$\begin{aligned} c_{12} &:= (-1, 1, 0, \ldots, 0) \\ c_{21} &:= (1, -1, 0, \ldots, 0) \\ c_{13} &:= (-1, 0, 1, 0, \ldots, 0) \\ c_{31} &:= (1, 0, -1, 0, \ldots, 0) \\ &\vdots \\ c_{1k} &:= (-1, \ldots, 0, 1) \\ c_{k1} &:= (1, \ldots, 0, -1). \end{aligned}$$

Will man nur einseitig auf ansteigende Effekte gegenüber der Kontrolle testen, so reicht es aus, die Kontraste c_{1i}, $i = 2, \ldots, k$ zu verwenden.
Geht man davon aus, daß der gemessene Effekt mit steigender Gruppennummer zunimmt, wie z.B. in einem Dosis-Wirkungs-Versuch, so hat Bretz (1999) einen neuen Kontrasttest vorgestellt, der für diese Situation geeignet ist. Die Konstruktion der zu verwendenden Kontraste ist in Bretz (1999) beschrieben.
Häufig wird die Entscheidung, ob ein Test ablehnt, anhand der berechneten Testgröße und dem sich ergebenden kritischen Wert (Quantil der Verteilung der Testgröße) getroffen, der sich aus dem Stichprobenumfang und dem geforderten Niveau ergibt. Anstelle

dieser Ja/Nein-Entscheidung kann man auch den zu der Testgröße gehörigen p-Wert berechnen, der die Wahrscheinlichkeit ist, daß die Testgröße diesen oder einen größeren Wert annimmt. Der p-Wert ist somit das größtmögliche Niveau, zu dem der Test bei diesen Daten noch ablehnen würde. Man erhält also bei vorgegebenem Niveau zusätzlich noch die „Stärke“ der Ablehnung des Tests.

Kapitel 2

Beispiele

Die in Kapitel 1.1 erwähnten Probleme bezüglich der Datenkondition werden in den folgenden drei Beispielen veranschaulicht. In Kapitel 5 werden die in dieser Arbeit vorgestellten Verfahren an diesen Daten demonstriert und mit denen der nicht-robusten Verfahren verglichen.

2.1 Länge von Kuckuckseiern

Kuckucke sind bekannt dafür, daß sie ihre Eier nicht selber ausbrüten, sondern sie in bestehende Gelege anderer Vögel legen. In den Nestern von sechs verschiedenen Vogelarten, die der Kuckuck als Wirtseltern für seine Jungen wählt, wurde die Länge der Kuckuckseier gemessen. Es ist bekannt, daß die Eier des Kuckucks farblich an die des Geleges angepaßt sind. Darüber hinaus stellt sich die Frage, ob die Längen der Eier und damit die Größe ebenfalls in Abhängigkeit von der als Zieheltern ausgewählten Vogelart variieren. Der Datensatz stammt aus dem DASL-Projekt[1] und ist dort unter dem Namen „Cockoo eggs in nests other birds“ verzeichnet. Die Vogelarten sind Wiesenpieper (Wiese), Baumpieper (Baum), Spatz, Rotkehlchen (Rot), Bachstelze (Bach) und Zaunkönig (Zaun). Aus dem Boxplot (Abbildung 2.1)ersieht man, daß:

1. die Varianzhomogenität nur annähernd gegeben ist,
2. bei Spatz und Wiesenpieper Ausreißer in den Daten vorkommen,
3. die Verteilungen bei Baumpieper und Zaunkönig schief sind.

Damit sind die Voraussetzungen des ANOVA Modells verletzt.

[1] (http://lib.stat.cmu.edu/DASL/DataArchive.html)

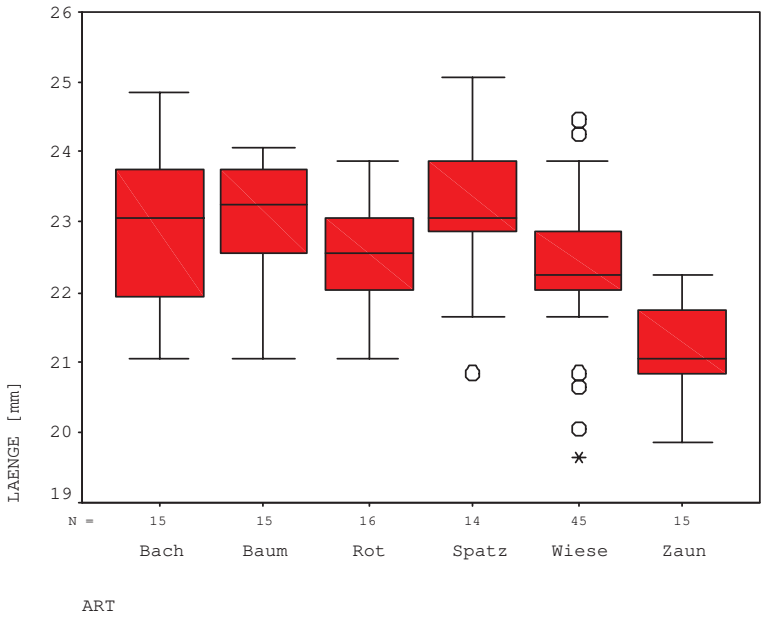


Abbildung 2.1: Boxplot der Längen der Kuckuckseier

2.2 Einfluß verschiedener Belichtungsvarianten auf die Kulturdauer bei Calibrachoa-Hybriden

An der LVA Hannover Ahlem wurde im Frühjahr 1988 ein Versuch zur Beurteilung des Einflusses verschiedener Belichtungsvarianten auf das Wachstum und Blühverhalten von Calibrachoa-Hybriden gemacht.
Ziel der Untersuchung war es herauszufinden, durch welche Belichtungsmaßnahme sich die Pflanzen frühzeitig in guter Verlaufsqualität zur Blüte bringen lassen. In dem Versuch wurden drei Sorten mit sechs verschiedenen Belichtungsvarianten behandelt. Die Behandlungen waren im einzelnen:

A) 8h natürliches Licht mit Verdunkelung von 16.00 bis 8.00 Uhr,

B) 8h natürliches Licht plus 8h photoperiodisches Licht unter der Verdunkelung,

C) 12h photoperiodische Tagesverlängerung,

D) 12h Assimilationslicht,

E) 16h photoperiodische Tagesverlängerung,

F) 16h Assimilationslicht.

Im vorliegenden Datensatz fehlen die Daten der ersten Variante.
An jedem Tag wurde für jede Pflanze die Anzahl der offenen Blüten bestimmt. Im folgenden wird eine Pflanze als blühend angesehen, wenn die erste Blüte geöffnet ist. Die

Zeit vom Topfen bis zur Blüte wird im folgenden Blühdauer genannt. Da interessiert, welche Belichtungsvariante für welche Sorte die kürzeste Blühdauer zur Folge hat, und nicht bekannt ist, wie die verschiedenen Behandlungen wirken, ist ein Vergleich aller Behandlungen gegeneinander pro Sorte das durchzuführende Testverfahren. Die

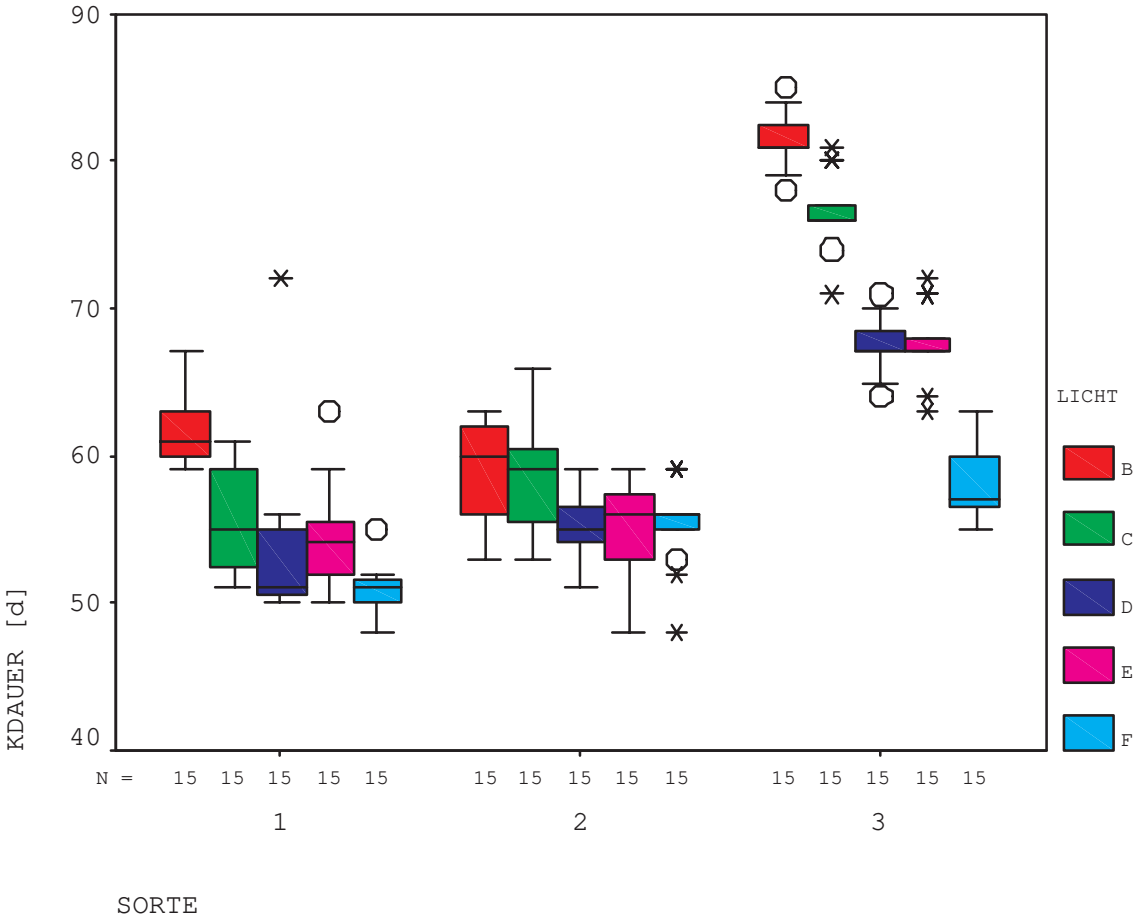


Abbildung 2.2: Boxplot der Kulturdauern (KDAUER) je Sorte und Behandlung, Quelle: Dr. Ludolph, persönliche Mitteilung

Boxplots (Abbildung 2.2) der Daten pro Sorte zeigen

1. keine Varianzhomogenität,
2. teilweise sehr schiefe Verteilungen, wie bei Sorte 1 Behandlung B, Sorte 2 Behandlung B und E,
3. moderate bis extreme Ausreißer, wie bei Sorte 3 in den Behandlungen C und E oder Sorte 1 in den Behandlungen D und E.

2.3 Durchlässigkeit einer Membran

In einem umfangreichen Projekt wurde unter drei verschiedenen Bedingungen der Stoffdurchgang (Permeation) durch eine Membran untersucht. Hierbei wurde alle 12h die noch nicht permeierte Menge des Stoffes bestimmt. Unter den verschiedenen Bedingungen wurden zusätzlich noch verschiedene Behandlungen angewandt, wobei immer eine Kontrollgruppe pro Bedingung mitgeführt wurde. Die Behandlungen bestehen zum Teil in Applikation verschiedener Dosierungen einer Substanz. Exemplarisch für die Ergebnisse des Versuchs sind hier die nicht permeierten Stoffmengen zum Endzeitpunkt des Experiments (nach 96h) für die 3. Bedingung dargestellt (Abbildung 2.3). Hierbei ist die Behandlung D in drei Konzentrationsstufen (1-3) angewandt worden, und außerdem wurden sechs weitere mit den Codierungen E und G-K vorgenommen. Es gilt festzustellen, welche der Behandlungen die geringste Rückstandsmenge hat, wobei zusätzlich bei Behandlung D von einer monotonen Dosis-Wirkungsbeziehung ausgegangen werden kann. Die zu testenden Hypothesen lauten deshalb für die Behandlungen D untereinander: $\mu_{D1} \geq \mu_{D2}$, $\mu_{D1} \geq \mu_{D3}$, $\mu_{D2} \geq \mu_{D3}$. Weiterhin werden alle Behandlungen einseitig mit der Kontrolle verglichen, und die restlichen paarweisen Hypothesen sind zweiseitig. Die Daten (Konzentrationen) wurden vor der Analyse logarithmustransformiert, da der zugrundeliegenden Mechanismus (Modell) dies nahelegt. Zum einen läßt sich die obige Fragestellung mit den Standardmethoden des ANOVA Modells, unter der Maßgabe das multiple Niveau zu kontrollieren, nur unzufriedenstellend beantworten. Da ein Teil der Tests zweiseitig ist und alle Behandlungen miteinander zu vergleichen sind, bleibt nur der Tukey-Test. Dadurch, daß der Tukey-Test ein zweiseitiger all-Paar Vergleich ist, werden elf einseitige Vergleiche zu viel gemacht. Damit wird es erschwert, an den interessierenden Stellen Unterschiede zu zeigen. Zum anderen sind die Voraussetzungen der Varianzhomogenität definitiv verletzt. Weiterhin treten insbesondere in der Kontrolle Ausreißer auf. Das Problem, nur die interessierenden Tests simultan durchzuführen und hierbei teilweise nur einseitig zu testen läßt sich durch die Verwendung multipler Kontraste lösen. Eine andere Möglichkeit besteht darin, nur Zwei-Stichprobe-Tests durchzuführen und deren p-Werte zu adjustieren. Um nicht durch den Einfluß der Ausreißer die Tests zu verzerren, verwende man in den beiden Methoden robuste Schätzer.

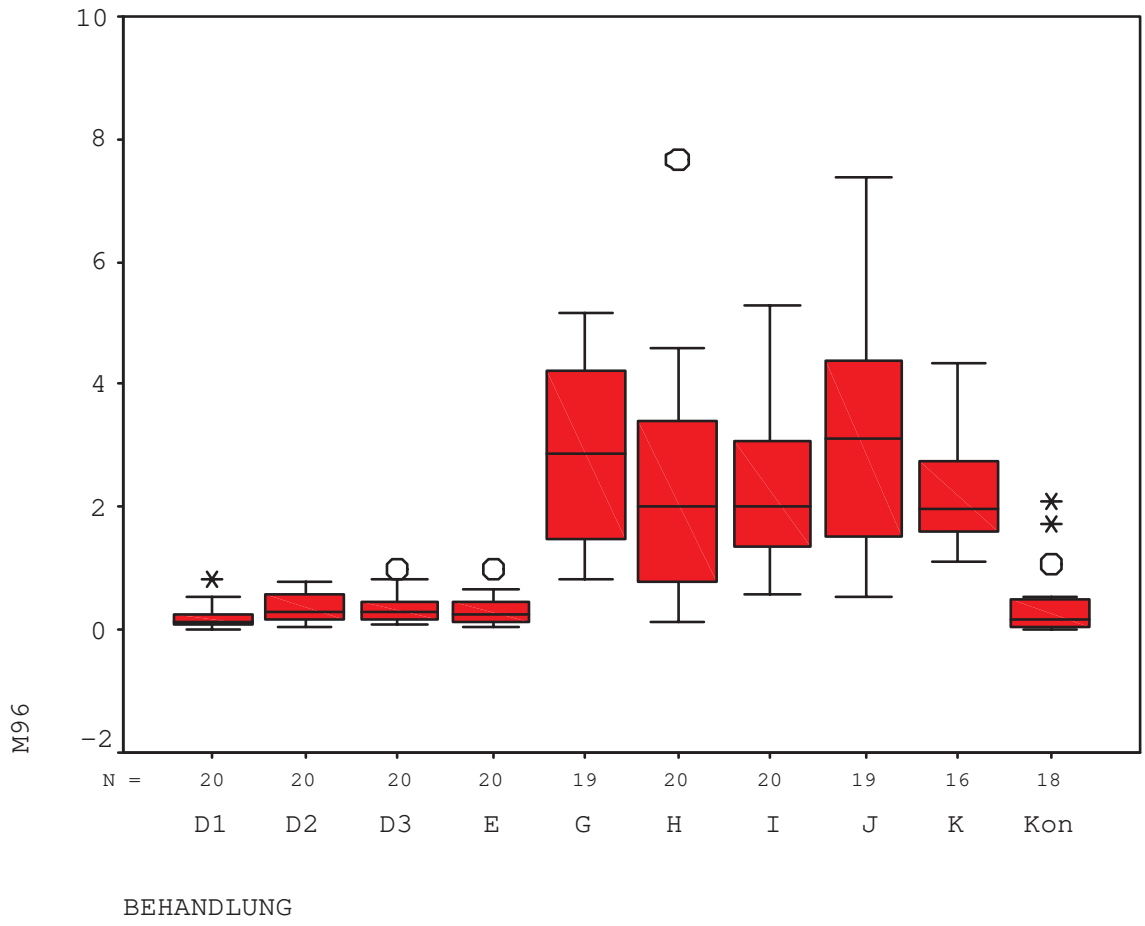


Abbildung 2.3: Boxplot der nicht permeierten Mengen nach 96h, Quelle: Diplomarbeit Martin Krämer, Lehrgebiet Bioinformatik

Kapitel 3

Robuste Mittelwertvergleiche

3.1 Robuste Schätzer

Die klassische Methode, zwei Mittelwerte zu vergleichen, ist der t-Test. Dieser Test ist unter der Voraussetzung, daß die Daten beider Gruppen normalverteilt sind, der gleichförmig beste Test (*uniform most powerful*, UMP). Ist mindestens eine der beiden Gruppen nicht normalverteilt, so ist von vornherein nicht klar, ob der Test das Niveau einhält. Auch das Verhalten der Power des Tests, also die Wahrscheinlichkeit, einen vorliegenden Unterschied zu erkennen, verändert sich situationsspezifisch.
Das Prinzip dennoch weiterverwenden zu können, werden im folgenden Ansätze vorgestellt, die eine robuste Schätzung der Lage bzw. der Streuung ermöglichen. Die hier vorgestellten Methoden sind eine Auswahl, die zum einen durch die Ergebnisse der Princton Studie (Andrews et al 1972) motiviert ist und zum anderen neuere erfolgversprechende Verfahren beinhaltet. Als erstes Verfahren ist das Trimmen und Winsorisieren ausgewählt, da es dem intuitiven Umgang mit Ausreißern am nächsten kommt. Das intuitive Weglassen von „zu großen“ und „zu kleinen“ Beobachtungen wird hierbei ersetzt durch einen systematischen Umgang mit großen und kleinen Werten. Weiter wird die m-Schätzung nach Huber (1981) vorgestellt. Bei diesem Verfahren werden die Parameter so geschätzt, daß die Summe der gewichteten Residuen, d.h. die Abstände der Werte vom „Mittel“ , Null ist. Als Gewichtsfunktionen sind dabei solche Funktionen gewählt, die besonders große Residuen relativ oder absolut geringer gewichten als kleinere. Als drittes Verfahren wird noch die modifizierte maximum likelihood Schätzung (Tiku et al 1986) vorgestellt, die als eine Abwandlung des Trimmens verstanden werden kann.

3.1.1 Trimmen und Winsorisieren

Erscheint bei der Durchführung des Versuchs der eine oder andere Meßwert suspekt, weil er zu groß oder zu klein ist, mag man versucht sein, diese Werte zu ignorieren. Zu solch einer subjektiven Auswahl ist zu bemerken, daß sie der Manipulation „Tür und Tor öffnet“ . Daher ist in einer entsprechenden Situation ein Verfahren angebracht, das von sich aus scheinbar zu stark abweichende Daten speziell behandelt. Ein solcher Ansatz zum Umgang mit Ausreißern besteht darin, prinzipiell die kleinsten und größten Beobachtungen bei der Schätzung des Lageparameters nicht zu berücksichtigen. Hierzu eignet sich beispielsweise das getrimmte Mittel (Dixon, Tukey 1968).

Definition 3.1 (Getrimmtes Mittel)
Das getrimmte Mittel einer Verteilung F ist definiert als:

$$\begin{aligned} \mu_t := \mu_t(\gamma, F) &:= \frac{1}{1-2\gamma} \int_{F^{-1}(\gamma)}^{F^{-1}(1-\gamma)} x \, dF(x) \\ &= \frac{1}{1-2\gamma} \int_{\gamma}^{1-\gamma} F^{-1}(x) \, dx \end{aligned} \tag{3.1}$$

Hierbei ist γ der Anteil der Verteilung, der zensiert wird.

Ist F eine symmetrische Verteilung, so ist dieser Lageparameter identisch mit dem Erwartungswert, falls dieser existiert.
Einen Schätzer für μ_t erhält man, indem anstelle der wahren Verteilungsfunktion F die empirische Verteilungsfunktion der Daten in die Gleichung (3.1) eingesetzt wird.

$$\begin{aligned} \hat{\mu}_t &:= \frac{1}{1-2\gamma} \int_{F_n^{-1}(\gamma)}^{F_n^{-1}(1-\gamma)} x \, dF_n(x) \\ &= \frac{1}{n(1-2\gamma)} \left[\sum_{i=\lceil \gamma n \rceil + 1}^{n - \lceil \gamma n \rceil} X_{(i)} + p \cdot (X_{(\lceil \gamma n \rceil + 1)} + X_{(n - \lceil \gamma n \rceil)}) \right] \end{aligned} \tag{3.2}$$

mit $p = \lceil \gamma n \rceil - \gamma n$.
Bei der Schätzung der Lage anhand des getrimmten Mittels werden also die $\lceil \gamma n \rceil$ kleinsten und größten Werte nicht berücksichtigt. Für den Fall, daß nur zu große oder nur zu kleine Werte nicht in die Schätzung mit eingehen sollen, weil z.B. bei dem gemessenen Merkmal keine zu kleinen Werte vorkommen können, läßt sich auch ein einseitig getrimmtes Mittel definieren als:

$$\begin{aligned} \hat{\mu}_{to} &:= \frac{1}{1-\gamma} \int_{0}^{F_n^{-1}(1-\gamma)} x \, dF_n(x) \\ \hat{\mu}_{tu} &:= \frac{1}{1-\gamma} \int_{F_n^{-1}(\gamma)}^{1} x \, dF_n(x). \end{aligned}$$

Bemerkung:
Die Trimmschätzer gehören zu der Klasse der L-Schätzer[1].
Zum Beispiel ergibt sich $\hat{\mu}_t$ durch die Wahl der Gewichte:

$$a_i = \begin{cases} 0 & \text{, falls } i \leq \lfloor \gamma n \rfloor \text{ oder } i > n - \lfloor \gamma n \rfloor \\ \frac{1}{n(1-2\gamma)} & \text{, falls } \lfloor \gamma n \rfloor < i \leq \lfloor n - \gamma n \rfloor. \end{cases}$$

Stigler (1973) hat die asymptotische Verteilung dieser Schätzer hergeleitet.

Satz 3.1 (Staute, Sheater 1990, S.106)
Unter der Voraussetzung, daß F an den Trimmstellen streng monoton wachsend ist, gilt

$$\sqrt{n}(\hat{\mu}_t - \mu_t) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(\gamma, F))$$

mit asymptotischer Varianz $\sigma^2(\gamma, F)$, wie zum Beispiel in Staute, Sheater (1990) beschrieben:

$$\sigma^2(\gamma, F) = \mathbf{E}\left[\mathrm{IF}^2_{\gamma,F}(x)\right] =: b^2_F(\gamma)\sigma^2.$$

Dabei ist $\mathrm{IF}_{\gamma,F}$ die Influenzfunktion, wie sie ebenda definiert ist. Hier hat sie die Form

$$(1-2\gamma)\mathrm{IF}_{\gamma,F}(x) = \begin{cases} F^{-1}(\gamma) - \mu_w, & x < F^{-1}(\gamma) \\ x - \mu_w, & F^{-1}(\gamma) \leq x \leq F^{-1}(1-\gamma) \\ F^{-1}(1-\gamma) - \mu_w, & F^{-1}(1-\gamma), \end{cases}$$

mit μ_w als dem winsorisierten Mittelwert der Verteilung F.

Definition 3.2 (Winsorisierter Erwartungswert)
Der winsorisierte Erwartungswert einer Zufallsvariablen ist definiert als:

$$\mathbf{E}_w[X] := \int_{F^{-1}(\gamma)}^{F^{-1}(1-\gamma)} x \, dF(x) + \gamma F^{-1}(\gamma) + \gamma F^{-1}(1-\gamma). \tag{3.3}$$

Für den winsorisierten Mittelwert $\mathbf{E}_w[X]$ schreibe man kurz μ_w, für alle höheren um μ_w zentrierten winsorisierten Momente schreibe man

$$\mu_{kw} := \mathbf{E}_w[X - \mu_w]^k,$$

beziehungsweise für $k = 2$ auch $\sigma^2_w = \mu_{2w}$.
Für eine Stichprobe $X_1, \ldots, X_n$ und die zugehörige geordnete Stichprobe $X_{(1)}, \ldots, X_{(n)}$ definiere die winsorisierte Stichprobe als

$$Y_{(i)} = \begin{cases} X_{(\lfloor \gamma n \rfloor + 1)}, & i \leq \lfloor \gamma n \rfloor \\ X_{(i)}, & \lfloor \gamma n \rfloor < i \leq n - \lfloor \gamma n \rfloor \\ X_{(n - \lfloor \gamma n \rfloor)}, & n - \lfloor \gamma n \rfloor < i \end{cases}$$

[1]Zur Klassifizierung von Schätzern siehe z.B. Huber (1981).

und die winsorisierte Varianz der Stichprobe als

$$s_w^2 := \frac{1}{n-1} SSQ_w := \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2.$$

Einen Schätzer für die Varianz $\sigma^2(\gamma, F)$ erhält man, indem die Verteilungsfunktion durch die empirische Verteilungsfunktion ersetzt wird, die damit die Form

$$s_W^2 = \frac{s_w^2}{(1-2\gamma)^2 n}$$

erhält. Dieser Varianzschätzer ist konsistent für $\sigma^2(\gamma, F)$ (Shorack 1997).
Im folgenden werden noch leichte Modifikationen dieses Schätzers vorgestellt, die auf den Vorschlag von Tukey und McLaughlin (1963) beziehungsweise auf Mudholkar et al (1991) zurückgehen:

$$s_Y^2 = \frac{(n-1)s_w^2}{h-1},$$

mit $h = (1-2\gamma)n$ und

$$s_M^2 = (n-1)s_W^2.$$

Zusammengefaßt unterscheiden sich die drei Varianzschätzer einzig in ihrem Faktor vor der winsorisierten Summe der Quadrate SSQ_w. Außer den in der Literatur vorge-

Schätzer	s_w^2	s_W^2	s_Y^2	s_M^2
Faktor	$\frac{1}{n-1}$	$\frac{1}{(1-2\gamma)^2 n(n-1)}$	$\frac{1}{h-1}$	$\frac{1}{(1-2\gamma)^2 n}$
Author	÷	Wilcox (1994)	Yuen (1974)	Mudholkar et al (1991)

Tabelle 3.1: Varianzfaktoren

stellten Varianzschätzern beziehungsweise genauer ihren zugehörigen Faktoren sind im Rahmen der Simulationsstudie von Kapitel 3.2 noch weitere betrachtet worden. Diese waren jedoch den hier vorgestellten unterlegen.
Ein generelles Problem bei Trimmschätzern ist die Wahl des Trimmanteils. In der Literatur bewegt er sich meist zwischen 10 und 20% (z.B. Andrews et al 1972, Mudholkar et al 1991, Yuen, Dixon 1973). Die Idee, den Anteil zu trimmender Daten dynamisch, d.h. von den Daten selbst abhängend, zu bestimmen, wurde von Leger und Romano (1990) umfassend theoretisch aufgegriffen. Als zu wählender Trimmanteil wird der Anteil gewählt, der ein passend gewähltes Funktional, das vom zugehörigen Trimmschätzer abhängt, minimiert. Dieses wird in ein Bootstrap-Verfahren eingebettet. Leger und Romano zeigen von dem so berechneten getrimmten Mittel, daß es asymptotisch normalverteilt ist und die kleinste Varianz unter den getrimmten Mitteln besitzt. Ein Spezialfall ihres theoretischen Konzepts stellt hierbei das von Jaeckel

(1971) eingeführte „optimal getrimmte Mittel“ dar. Weitere umfassende theoretische Resultate inklusive eines Zentralen Grenzwertsatzes bei zufälligem Trimmanteil hat Shorack (1997) veröffentlicht.

3.1.2 m-Schätzer

Bei den m-Schätzern handelt es sich um eine Klasse von implizit definierten Schätzern.

Definition 3.3 (m-Lokation)
Sei F eine beliebige Verteilungsfunktion und ψ eine ungerade, monoton wachsende Funktion ($\not\equiv 0$). So definiere man $\mu_\psi := \mu_\psi(F)$ implizit als die Lösung der Gleichung:

$$\int \psi(x - \mu_\psi)\, dF(x) = 0. \tag{3.4}$$

Der so definierte Schätzer muß nicht eindeutig sein; die Gleichung (3.4) kann keine, eine oder mehrere Lösungen haben. Für den Fall, daß mehr als eine Lösung existiert, liegen alle Lösungen in einem endlichen Intervall; als Lokation läßt sich dann der Mittelpunkt dieses Intervalls nehmen. Um μ_ψ zu schätzen, ersetze in man (3.4) F durch F_n.

Definition 3.4 (m-Schätzer)
Der m-Schätzer $\hat{\mu}_\psi$ ist die implizite Lösung der Gleichung

$$\int \psi(x - \hat{\mu}_\psi)\, dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} \psi(X_i - \hat{\mu}_\psi) = 0.$$

Unter gewissen Voraussetzungen gilt für den m-Schätzer auch, daß er asymptotisch normalverteilt ist.

Satz 3.2 (Theorem 4.2 Staudte, Sheater 1990)
Sei F symmetrisch um $\mu_0 = \mu_\psi(F)$, ψ ungerade und monoton wachsend und $h(\mu) := \mathbf{E}\left[\psi(X - \mu)\right]$ existent. Existieren weiterhin $h'(\mu) = -\mathbf{E}\left[\psi'(X - \mu)\right] < 0$ und $h''(\mu)$, wobei $h''(\mu)$ in einer Umgebung von μ_0 beschränkt sei, so gilt, daß (3.4) eine eindeutige Lösung hat und

$$\sqrt{n}(\hat{\mu}_\psi - \mu_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(\psi, F)).$$

Hierbei ist $\sigma^2(\psi, F) = \mathbf{E}\left[\mathrm{IF}^2_{\psi,F}(X)\right]$.

Mit $\mathrm{IF}_{\psi,F}$ ist wieder die Influenzfunktion bezeichnet, die hier die Form

$$\mathrm{IF}_{\psi,F}(x) = \frac{\psi(x - \mu_\psi)}{\mathbf{E}\left[\psi'(X - \mu_\psi)\right]} \tag{3.5}$$

hat. Von einem Schätzer für $\sigma^2(\psi, F)$ ist zu erwarten, daß er konsistent ist. In natürlicher Weise bietet sich hier wieder die Ersetzung von F durch F_n in (3.5) an:

$$s^2_\psi := \frac{1}{n-1} \frac{\frac{1}{n}\sum_{i=1}^{n} \psi^2(X_i - \hat{\mu}_\psi)}{\left(\frac{1}{n}\sum_{i=1}^{n} \psi'(X_i - \hat{\mu}_\psi)\right)^2}.$$

Der Vorfaktor $(n-1)$ anstelle von n, der sich beim Ersetzen von F durch F_n in (3.5) ergeben würde, ist gewählt in Anlehnung an den klassischen Varianzschätzer (Huber 1981).
Ein so definierter Lageparameter μ_ψ hat den Nachteil, daß er nicht skaleninvariant ist. Die Skaleninvarianz läßt sich erhalten, indem in der Definition des Lageparameters der Term $\psi(x-\mu_\psi)$ durch $\psi\left(\frac{x-\mu_\psi}{\sigma(F)}\right)$ ersetzt wird, wobei $\sigma(F)$ der Skalenparameter ist. Da im allgemeinen die Skala nicht bekannt ist, muß diese ebenfalls geschätzt werden. Auch dies ist mit einer m-Schätzung möglich.

Definition 3.5 (m-Skala)
Sei F eine beliebige Verteilungsfunktion und χ eine gerade Funktion ($\not\equiv 0$). So definiere man $\sigma_\chi := \sigma_\chi(F)$ implizit als die Lösung der Gleichung

$$\int \chi\left(\frac{x}{\sigma_\chi}\right)\, dF(x) = 0. \tag{3.6}$$

Um nun zu einer gemeinsamen Schätzung von Lokation und Skala zu kommen, sind die sich aus (3.4) und (3.6) ergebenden Schätzgleichungen simultan zu lösen. Huber (1981) schlägt als simultanen Ansatz folgende Gleichungen vor (Huber's Proposal 2):

$$\sum_{i=1}^{n} \psi\left(\frac{x_i-\hat{\mu}_\psi}{s_\psi}\right) = 0,$$
$$\sum_{i=1}^{n} \psi^2\left(\frac{x_i-\hat{\mu}_\psi}{s_\psi}\right) = (n-1)\beta,$$

mit

$$\beta = \mathbf{E}\left[\psi^2\right]. \tag{3.7}$$

Die Funktion ψ sei hier so gewählt, daß sie ungerade und monoton wachsend ist, mit $0 \le \psi' \le 1$. Die beiden so definierten Schätzer sind nach Hampel et al (1986) asymptotisch unabhängig, und für symmetrisches F ist s_ψ konsistent.
Da dieses Gleichungssystem im allgemeinen nicht explizit lösbar ist, müssen die Lösungen für $\hat{\mu}_\psi$ und s_ψ iterativ bestimmt werden. In der Literatur (Huber 1981, Hampel et al 1986) wird zur Bestimmung das Newton-Verfahren vorgeschlagen, das schon nach dem ersten Schritt abgebrochen wird. Als Startwerte sollten möglichst einfach zu bestimmende und „robuste“ Ausgangswerte für $\hat{\mu}_0$ und s_0 gewählt werden. Übliche Vorschläge für $\hat{\mu}_0$ sind der Median und für s_0 das 1.483-fache der *Median Absolut Deviation* (MAD), wobei sich für s_0 auch die *length of the shortest half* (Rousseeuw, Leroy 1988) als Startwert eignet. Diese durch das einmalige Iterieren gewonnenen Schätzer werden *Einschritt m-Schätzer* genannt. Wie in Huber (1981) gezeigt, haben diese Schätzer die

gleichen asymptotischen Eigenschaften wie die exakten Nullstellen, jedenfalls bei ungeradem ψ und symmetrischem F.
Sind mit $\hat{\mu}_0$ und s_0 die Startwerte bezeichnet, so haben die Einschritt-Schätzer die Gestalt

$$\hat{\mu}_\psi := \hat{\mu}_0 + \frac{\sum_{i=1}^n \psi\left(\frac{x_i-\hat{\mu}_0}{s_0}\right) s_0}{\sum_{i=1}^n \psi'\left(\frac{x_i-\hat{\mu}_0}{s_0}\right)} \tag{3.8}$$

$$s_\psi^2 := \frac{1}{(n-1)\beta} \sum_{i=1}^n \psi^2\left(\frac{x_i-\hat{\mu}_0}{s_0}\right) s_0^2. \tag{3.9}$$

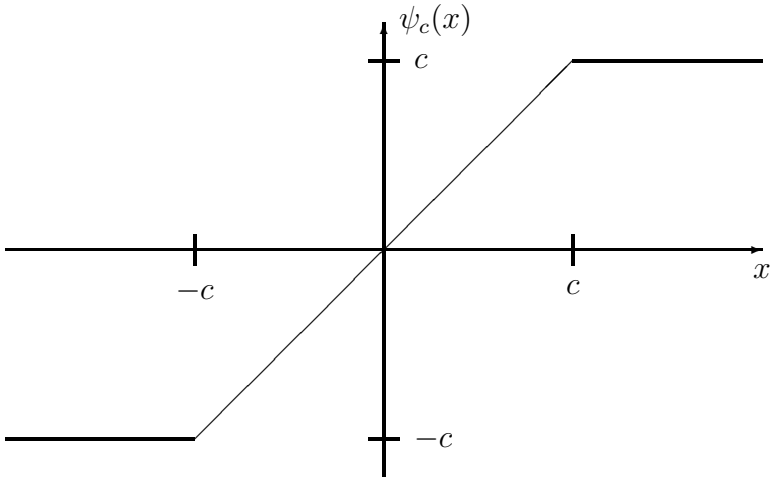
Die Forderung der Monotonie von ψ hat zur Folge, daß selbst sehr große Werte, die den Anschein haben, Ausreißer oder Meßfehler zu sein, noch in die Schätzung der Parameter eingehen. Will man dies vermeiden, also soll $\psi(x) = 0$ für $|x| > c$ gelten, so muß man die Monotonieforderung aufgegeben, wodurch zwar die Eindeutigkeit der Lösung von (3.4) verloren geht, dafür aber große Ausreißer vollständig verworfen werden. Die Verwendung der Einschrittschätzer hat zudem den Vorteil, daß sie immer ein Ergebnis liefert, denn Wellmann (1994) hat gezeigt, daß das Iterationsverfahren für beschränkte ψ nicht immer konvergiert. In dieser Arbeit wird der folgende m-Schätzer untersucht:

Huber-m-Schätzer

Der Huber-m-Schätzer ergibt sich durch die Wahl der Gewichtsfunktion

$$\psi_c(x) := \begin{cases} -c & x < -c \\ x & -c \le x \le c \\ c & c < x \end{cases} .$$

Das c ist hierbei ein noch frei zu wählender Parameter, der bestimmt, ab welchem Wert die Beobachtungen nicht mehr in ihrer vollen Größe in die Parameterschätzung eingehen.



Graph der Gewichtsfunktion

Das β aus Gleichung (3.7) lautet in diesem Fall:

$$\begin{aligned}
\beta &= \int_{-\infty}^{\infty} \psi_c^2(x)\phi(x)\,dx \\
&= \int_{-\infty}^{-c} c^2\phi(x)\,dx + \int_{-c}^{c} x^2\phi(x)\,dx + \int_{c}^{\infty} c^2\phi(x)\,dx \\
&= c^2\Phi(-c) + \int_{-c}^{c} x^2 \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}x^2}\,dx + c^2(1-\Phi(c)) \\
&= 2c^2(1-\Phi(c)) + \frac{1}{\sqrt{2\pi}}\left(\left[-x\mathrm{e}^{-\frac{1}{2}x^2}\right]_{-c}^{c} + \int_{-c}^{c} \mathrm{e}^{-\frac{1}{2}x^2}\,dx\right) \\
&= 2c^2(1-\Phi(c)) + 2\Phi(c) - 1 - \sqrt{\frac{2}{\pi}} c\mathrm{e}^{-\frac{1}{2}c^2}.
\end{aligned}$$

Speziell für das Zwei-Stichproben-Problem hat Akritas (1991) einen weiteren komplexeren Vorschlag zur alleinigen m-Schätzung der Lokationsdifferenz gemacht. Weitere Ansätze zur gemeinsamen Schätzung von Lage und Streuung im Ein-Stichproben-Fall sind in Jurećková und Sen (1982) und in Öztürk (1998) zu finden. Jedoch handeln beide Arbeiten nur die simultane Schätzung der Parameter ab und gehen nicht auf eine mögliche Verwendung der Schätzer in Zwei- oder k-Stichproben-Tests ein.

3.1.3 Tikus modified maximum likelihood-Schätzer

Ausgehend davon, daß die Dichte der den Daten zugrundeliegenden Verteilung von der Form $\sigma^{-1}f((x-\mu)\sigma^{-1})$ ist, wobei μ und σ der Lokations- bzw. Skalenparameter sind, läßt sich die likelihood-Funktion für Typ II (beidseitig) zensierte Daten bestimmen. Mit $z = (x-\mu)\sigma^{-1}$, r_1 als der unteren Anzahl, r_2 als der oberen Anzahl an zensierten Daten und $F(x) = int_{-\infty}^{x} f(y)\,dy$ lautet die likelihood-Funktion:

$$L = \frac{n!}{r_1!r_2!}\sigma^{-(n-r_1-r_2)}\left[\prod_{i=r_1+1}^{n-r_2} f(z_{(i)})\right][F(z_{(r_1+1)})]^{r_1}[1-F(z_{(n-r_2)})]^{r_2},$$

mit $z_{(i)} = (x_{(i)}-\mu)\sigma^{-1}$.
Die maximum likelihood-Schätzer sind die Lösung der Gleichungen

$$\frac{\partial \log L}{\partial \mu} = 0 \quad \text{und} \quad \frac{\partial \log L}{\partial \sigma} = 0. \tag{3.10}$$

Diese sind jedoch für die meisten Verteilungen nicht explizit lösbar. Im Fall der Normalverteilung erhalten die Gleichungen (3.10) die Gestalt:

$$\frac{\partial \log L}{\partial \mu} = \frac{n}{\sigma}\left[\frac{1}{n}\sum_{i=r_1+1}^{n-r_2} z_{(i)} - q_1 g_1(z_{(r_1+1)}) + q_2 g_2(z_{(n-r_2)})\right] = 0 \tag{3.11}$$

$$\begin{aligned}
\frac{\partial \log L}{\partial \sigma} &= \frac{n}{\sigma}\Big[-(1-q_1-q_2) + \frac{1}{n}\sum_{i=r_1+1}^{n-r_2} z_{(i)}^2 - q_1 z_{(r_1+1)} g_1(z_{(r_1+1)}) \\
&\qquad + q_2 z_{(n-r_2)} g_2(z_{(n-r_2)})\Big] = 0.
\end{aligned} \tag{3.12}$$

Hierbei sind $q_1 = r_1/n,\ q_2 = r_2/n,\ g_1(z) = \varphi(z)/\Phi(z)$ und $g_2(z) = \varphi(z)/(1-\Phi(z))$. Anstelle einer iterativen Lösung der Gleichungen (3.11) und (3.12), schlägt Tiku (1986) eine leichte Abänderung der Gleichungen vor, die eine explizite Lösung zuläßt. Die Idee besteht darin, die beiden Funktionswerte $g_1(z)$ und $g_2(z)$ an den betrachteten Stellen linear zu approximieren. Die zugehörige likelihood-Funktion sei mit L^* bezeichnet. Setzt man diese lineare Approximation in (3.11) und (3.12) ein, so lassen sich die Nullstellen wieder explizit bestimmen. Die Schätzer ergeben sich dann zu (Tiku 1986, Seiten 34-37):

$$\begin{aligned} \hat{\mu}_{MML} &= K + D\hat{\sigma}_{MML}, \\ s_{MML} &= \frac{B + \sqrt{B^2 + 4AC}}{2\sqrt{A(A-1)}} \end{aligned}$$

mit

$$\begin{aligned} q_i &= \frac{r_i}{n}, \\ h_1 &= \Phi^{-1}\left(q_1 - \sqrt{\frac{q_1(1-q_1)}{n}}\right), \\ k_1 &= \Phi^{-1}\left(q_1 + \sqrt{\frac{q_1(1-q_1)}{n}}\right), \\ h_2 &= \Phi^{-1}\left(1 - q_2 - \sqrt{\frac{q_2(1-q_2)}{n}}\right), \\ k_2 &= \Phi^{-1}\left(1 - q_2 + \sqrt{\frac{q_2(1-q_2)}{n}}\right), \\ \beta_1 &= -\frac{g_1(k_1) - g_1(h_1)}{k_1 - h_1}, \\ \alpha_1 &= g_1(h_1) - h_1\beta_1, \\ \beta_2 &= \frac{g_2(k_2) - g_2(h_2)}{k_2 - h_2}, \\ \alpha_2 &= g_2(h_2) - h_2\beta_2, \\ m &= n - r_1 - r_2 + r_1\beta_1 + r_2\beta_2, \\ A &= n - r_1 - r_2, \\ D &= \frac{r_2\alpha_2 - r_1\alpha_1}{m} \\ K &= \frac{1}{m}\left[\sum_{i=r_1+1}^{n-r_2} x_{(i)} + r_1\beta_1 x_{(r_1+1)} + r_2\beta_2 x_{(n-r_2)}\right], \\ B &= r_2\alpha_2(x_{(n-r_2)} - K) - r_1\alpha_1(x_{(r_1+1)} - K) \text{ und} \\ C &= \sum_{i=r_1+1}^{n-r_2} x_{(i)}^2 + r_1\beta_1 x_{(r_1+1)}^2 + r_2\beta_2 x_{(n-r_2)}^2 - mK^2. \end{aligned}$$

Für diese Schätzer existiert auch ein Resultat über die asymptotische Verteilung:

Satz 3.3 (Lemma 2.7.1 Tiku 1986)
Asymptotisch ist der Vektor $(\hat{\mu}_{MML}, s_{MML})$ bivariat normalverteilt mit Erwartungswerten $\mathbf{E}[\hat{\mu}_{MML}] = \mu$ und $\mathbf{E}[s_{MML}] = \sigma$ und Kovarianzmatrix:

$$\Sigma = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{21} \end{pmatrix}^{-1} \qquad J_{12} = J_{21},$$

mit

$$J_{11} = -\mathbf{E}\left[\frac{\partial^2 \log L^*}{\partial \mu^2}\right], \quad J_{12} = -\mathbf{E}\left[\frac{\partial^2 \log L^*}{\partial \mu \partial \sigma}\right] \text{ und } J_{22} = -\mathbf{E}\left[\frac{\partial^2 \log L^*}{\partial \sigma^2}\right].$$

Falls $r = r_1 = r_2$ ist, so ist $J_{12} = 0$, das heißt, die Schätzer sind asymptotisch unabhängig.

3.2 Robuste 2-Stichprobenverfahren

Um die Lage zweier Stichproben zu vergleichen, lassen sich verschiedene Wege einschlagen. Hier sind einerseits die parametrischen Verfahren, die auf den Vergleich über (robust) geschätzte Parameter - hier Lage und Streuung - abzielen, und auf der anderen Seite die nichtparametrischen Verfahren (Rangtransformationen) zu nennen. Im folgenden werden beide Ansätze betrachtet, wobei die parametrischen Verfahren auf den im vorigen Kapitel vorgestellten robusten Schätzern aufbauen. Die gewählten Notationen stimmen mit denen aus Kapitel 3.1 überein.

3.2.1 Verwendung von robusten Schätzern

Aus den in Kapitel 3.1 vorgestellten robusten Schätzern werden nun 2-Stichprobentests konstruiert.

Getrimmter t-Test

Die im folgenden vorgestellten verschiedenen Teststatistiken basieren auf den in 3.1.1 vorgestellten Schätzern für Lage und Streuung. Als erste stellten Yuen und Dixon (1973) einen robusten 2-Stichprobentest vor, den Yuen in einer Arbeit von 1974 verbessert hat. Die verwendete Teststatistik lautet

$$T_Y = \frac{\hat{\mu}_{t1} - \hat{\mu}_{t2}}{\sqrt{\frac{s^2_{Y1}}{h_1} + \frac{s^2_{Y2}}{h_2}}}. \tag{3.13}$$

Hierbei steht der zusätzliche Index 1 bzw. 2 für die jeweilige Gruppe. Yuen schlägt vor, die Verteilung der Teststatistik (3.13) durch eine t-Verteilung mit ν_Y Freiheitsgraden, die denen des Welch-Tests nachgebildet sind, zu approximieren:

$$\frac{1}{\nu_Y} = \frac{\hat{c}^2}{h_1 - 1} + \frac{(1-\hat{c})^2}{h_2 - 1}, \qquad \hat{c} = \frac{s^2_{Y1}/h_1}{s^2_{Y1}/h_1 + s^2_{Y2}/h_2}. \tag{3.14}$$

In ihren Untersuchungen beschränkte sich Yuen auf das Verhalten dieser Teststatistik bei symmetrischen Verteilungen. Für den Fall, daß die zugrundeliegende Verteilung nicht symmetrisch ist, schlägt Wilcox (1994, 1990) eine auf einer Cornish-Fisher-Expansion basierende Modifikation der Teststatistik (3.13) vor. Durch diese Modifikation werden Schätzungen für Schiefe und Exzeß mit in der Teststatistik berücksichtigt. So sollen möglichen Verzerrungen bei Nicht-Normalverteilungen korrigiert werden. Zusätzlich verwendet Wilcox einen leicht modifizierten Varianzschätzer.

$$T_W = \frac{(\hat{\mu}_{t1} - \hat{\mu}_{t2}) + \Gamma + \Xi(\hat{\mu}_{t1} - \hat{\mu}_{t2})^2}{s_{Wp}} \tag{3.15}$$

Die gepoolte Varianz $\check{s}^2_{pW}$ und die Korrektoren für Schiefe und Exzeß (Γ, Ξ) werden bestimmt zu:

$$\begin{aligned} s^2_{Wp} &= s^2_{W1} + s^2_{W2} \\ \Gamma &= \frac{\hat{\mu}_{31w}/n_1 - \hat{\mu}_{32w}/n_2}{6 s^2_{Wp}} \\ \Xi &= \frac{\hat{\mu}_{31w}/n_1 - \hat{\mu}_{32w}/n_2}{3 s^4_{Wp}}. \end{aligned}$$

Die Freiheitsgrade der approximativen t-Verteilung bleiben gleich denen der Yuen-Statistik. Da sich der von Wilcox vorgeschlagene Varianzschätzer von dem Yuen'schen unterscheidet, wird auch die Teststatistik

$$T_w = \frac{\hat{\mu}_{t1} - \hat{\mu}_{t2}}{s_{Wp}}$$

in den Simulationen betrachtet. Mudholkar et al (1991) schlagen eine anders motivierte Schätzung der gepoolten Varianz und einen anders geschätzten Freiheitsgrad vor. Als Teststatistik geben sie an

$$T_M = \frac{\hat{\mu}_{t1} - \hat{\mu}_{t2}}{s_{Mp}\sqrt{\frac{b_1^2}{n_1} + \frac{b_2^2}{n_2}}}. \tag{3.16}$$

Der Varianzschätzer s^2_{Mp} und die b_i's aus (3.16) sind hier definiert als

$$\begin{aligned} b_i^2 &:= b_i^2(\gamma_i) := 1 + 0,48\gamma_i + 1,21\gamma_i^2 \\ w_i &:= w_i(\gamma_i) := (n_i - 1) \cdot (0,5 - 1,62\gamma_i + 1,91\gamma_i^2 - 1,85\gamma_i^3) \\ s^2_{Mp} &:= \left(w_1 \frac{s^2_{M1}}{b_1^2} + w_2 \frac{s^2_{M2}}{b_2^2} \right) (w_1 + w_2)^{-1}, \end{aligned}$$

wobei γ_i der Trimmanteil in der $i-$ten Gruppe ist. Als Freiheitsgrad für eine approximative t-Verteilung wird

$$\nu_M = 2(w_1 + w_2) \tag{3.17}$$

genommen. Für kleine Fallzahlen schlagen Modholkar et al (1991) noch einen Korrekturfaktor A vor, mit dem die Teststatistik modifiziert werden sollte. Dieser ergibt sich aus der Anpassung der empirischen Varianz von T_M an die der $t_{\nu_M}-$Verteilung

$$\frac{1}{A} := 1 - 1,3\frac{\gamma}{\nu_M} + 7,5\frac{\gamma^2}{\nu_M} + 16\frac{\gamma}{\nu_M^2} - 150\frac{\gamma^3}{\nu_M^2},$$

mit $\gamma = \frac{1}{2}(\gamma_1 + \gamma_2)$ und ν_M dem approximativen Freiheitsgrad aus (3.17). Eine Übertragung der Ergebnisse der Herleitung von Mudholkar et al (1991) auf die Teststatistik (3.13) von Yuen führt zu der Teststatistik

$$T_{YM} = \frac{\hat{\mu}_{t1} - \hat{\mu}_{t2}}{\sqrt{\frac{s^2_{M1}}{h_1} + \frac{s^2_{M2}}{h_2}}} \tag{3.18}$$

mit Freiheitsgraden

$$\frac{2}{\nu_{YM}} = \frac{\tilde{c}^2}{w_1} + \frac{(1-\tilde{c})^2}{w_2}$$

und

$$\tilde{c} = \frac{b_1^2 s_{M1}^2/n_1}{b_1^2 s_{M1}^2/n_1 + b_2^2 s_{M2}^2/n_2}.$$

Um die verschiedenen hier vorgestellten Teststatistiken auf multiple Kontraste verallgemeinern zu können, werden die Teststatistiken (3.13) und (3.18) noch modifiziert zu

$$\tilde{T}_Y = \frac{\hat{\mu}_{t1} - \hat{\mu}_{t2}}{s_{Yp}\sqrt{\frac{1}{h_1} + \frac{1}{h_2}}}, \tag{3.19}$$

$$\tilde{T}_{YM} = \frac{\hat{\mu}_{t1} - \hat{\mu}_{t2}}{s_{Mp}\sqrt{\frac{1}{h_1} + \frac{1}{h_2}}}, \tag{3.20}$$

mit $s_{Yp}^2 := \frac{(h_1-1)s_{Y1}^2 + (h_2-1)s_{Y2}^2}{h_1+h_2-2}$.
Sind die Fallzahlen und die Trimmung in beiden Gruppen gleich, so ist $\tilde{T}_Y = T_Y$ und $\tilde{T}_{YM} = T_{YM}$.
Im Hinblick auf multiple Tests wird der Ansatz von Mudholkar nicht weiterverfolgt, da die empirische Anpassung der Verteilung keine Verallgemeinerung zuläßt. Außerdem ist im Multivariaten nicht mehr die Frage nach den „richtigen“ Freiheitsgraden von Interesse, sondern die Korrelationsmatrix der einzelnen Vergleiche. Ebenso ist für den Vorschlag von Wilcox keine direkte Verallgemeinerung möglich, so daß nur der zudem leicht zu handhabende Ansatz von Yuen weiterverfolgt wird.

M-Schätzer t-Test

Wie in Gleichungen (3.8, 3.9) definiert, lassen sich Lage und Streuung simultan aus einer Stichprobe schätzen. Im Zwei-Stichproben-Fall kann man zum einen in jeder Stichprobe getrennt die beiden Parameter schätzen und aus den vier Werten einen Test konstruieren oder, falls man von homogenen Varianzen ausgeht, simultan die beiden Lageparameter und die gemeinsame Streuung schätzen. Zum ersten Ansatz veröffentlichten Yuen, Lee und Tajuddin (1985) eine Methode, die jedoch sehr speziell ist. Da sich die Teststatistik ihren Untersuchungen nach nicht direkt durch eine t-Verteilung approximieren läßt, modifizieren sie sowohl die Teststatistik als auch die Freiheitsgrade durch je eine empirisch bestimmte „Tunigkonstante“, die von der gewählten Gewichtsfunktion ψ abhängt:

$$T_M = \frac{\hat{\mu}_{\psi 1} - \hat{\mu}_{\psi 2}}{\sqrt{\frac{(n_1-1)s_{\psi 1} + (n_2-1)s_{\psi 2}}{n_1+n_2-2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}.$$

Für diese Testgröße gilt, daß sie asymptotisch normalverteilt ist. Im Endlichen sind die zugehörigen kritischen Werte durch eine t-Verteilung mit $n_1 + n_2 - 2$ Freiheitsgraden approximierbar (Aboukalam 1992). Den Freiheitsgrad analog zum Welch-Test aus den Daten zu berechnen, ist nach den theoretischen Resultaten von Lucas (1997) nicht sinnvoll, da die Robustheit des Tests nur für festen Feiheitsgrad gewährleistet ist.
Unter der Annahme von homogenen Varianzen läßt sich aus der gepoolten Stichprobe die gemeinsame Varianz schätzen, was zu den drei folgenden impliziten Gleichungen führt (Weichert, Hothorn 2000):

$$\sum_{i=1}^{n_1} \psi\left(\frac{x_{i1} - \hat{\mu}_{\psi 1}}{s_\psi}\right) = 0, \tag{3.21}$$

$$\sum_{i=1}^{n_2} \psi\left(\frac{x_{i2} - \hat{\mu}_{\psi 2}}{s_\psi}\right) = 0, \tag{3.22}$$

$$\sum_{k=1}^{2}\sum_{i=1}^{n_k} \psi^2\left(\frac{x_{ik} - \hat{\mu}_{\psi k}}{s_\psi}\right) = (n_1 + n_2 - 1)\beta. \tag{3.23}$$

Dabei ergibt sich

$$T_W = \frac{\hat{\mu}_{\psi 1} - \hat{\mu}_{\psi 2}}{s_\psi \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

zu der Testgröße, deren kritische Werte mit der t-Verteilung mit $n_1 + n_2 - 2$ Freiheitsgraden angenähert werden.

Tikus MML-Schätzer t-Test

Zum robusten Vergleich zweier Lageparameter anhand der Tiku-Schätzer werden in Tiku (1986) die Schätzer für Lage und Streuung im t-Test durch die jeweiligen Pendants der MML-Schätzung ersetzt. Ein Tiku-t-Test hat somit die Form

$$T_{MML} = \frac{\hat{\mu}_{MML1} - \hat{\mu}_{MML2}}{s_{MMLp}\sqrt{\frac{1}{m_1} + \frac{1}{m_2}}}.$$

Dabei wird die gepoolte Varianz s^2_{MMLp} geschätzt durch

$$s^2_{MMLp} = \frac{(A_1 - 1)s^2_{MML1} + (A_2 - 1)s^2_{MML2}}{A_1 + A_2 - 2}.$$

Für den Fall, daß beide Verteilungen Normalverteilungen mit gleicher Varianz σ^2 sind, gilt:

Satz 3.4 (Theorem 4.3.1 Tiku 1986)
Unter Normalverteilung ist die Verteilung von T_{MML} unter der Nullhypothese für große Fallzahlen eine zentrale t-Verteilung mit $A_1 - A_2 - 2$ Freiheitsgraden, während unter der Alternative T_{MML} nicht-zentral t-verteilt ist mit $A_1 - A_2 - 2$ Freiheitsgraden und Nichtzentralitätsparameter $\Delta = (\mu_1 - \mu_2)^2/(\sigma^2(1/m_1 + 1/m_2))$.

3.2.2 Minimum-Statistik

Wie schon bei Dunnett (1982), Carroll (1979) oder Wilcox (1997) aus den Simulationsergebnissen zu ersehen ist, gibt es keinen in allen Situationen besten Test, sondern je nach wahrer Verteilung der Daten weist ein anderer Test die höchste Güte auf. So ist z.B. bei normalverteilten Daten der t-Test UMP, während er bei kontaminiert normalverteilten Daten einen deutlichen Powerverlust gegenüber den Yuen'schen Tests aufweist. Eine Möglichkeit, um aus dem Dilemma herauszukommen, möglicherweise den schlechtesten Test für die jeweils vorliegenden Daten zu wählen, besteht darin, das Maximum von verschiedenen Teststatistiken als neue Teststatistik zu nehmen oder analog den minimalen p-Wert. Efron und Tibshirani (1993) schlagen in einem Beispiel vor, den minimalen p-Wert des t-Tests, des 10% getrimmten Tests, des 25% getrimmten Tests und des Median-Tests zu nehmen und die Verteilung dieses Tests zu bootstrapen. Diese Idee wird nur als Beispiel für die Verwendung des Bootstraps bei heuristischen Statistiken angeführt, deren Verteilung nur extrem schwer zu bestimmen ist. Der auf dieser neuen Teststatistik basierende Test wird nie eine optimale Güte im Vergleich zum besten Test haben, jedoch ist nach den Ergebnissen von Neuhäuser(1996) zu vermuten, daß solch ein Test nie der schlechteste sein wird.
Da im allgemeinen die Verteilung dieses Maximums nicht einfach zu bestimmen ist, wird hier dem Vorschlag von Efron und Tibshirani gefolgt, und die Methode des Bootstrapens verwendet, um die Verteilung der Teststatistik zu bestimmen. Das Bootstrapen ist ein vielversprechendes Verfahren, da es im Ein-Stichproben-Fall sogar eine bessere Approximation der Verteilung für trim- und m-Schätzer liefert als die Verwendung der asymptotischen Verteilung (Yang 1985, Hall, Padmanabhan 1992). Da die hier gemeinsam betrachteten Tests teilweise verschiedene Freiheitsgrade haben, wird den Vorschlägen von Efron und Tibshirani (1993) und Westfall und Young (1993), gefolgt und es werden die p-Werte anstelle der Teststatistiken verwendet. Im folgenden setze man den p-Wert des t-Tests als p_t, den des 10% und 20% getrimmten Tests als p_{t10} und p_{t20} und den des Huber-m-Testes mit $c = 1,8$ als $p_{c_{1,8}}$. Aufgrund der Ergebnisse für die simulierte Power der einzelnen Zwei-Stichproben-Tests unter den verschiedenen Verteilungen werden die folgenden Minima untersucht:

1. Das Minimum der p-Werte aus t-Test und Yuen's Test mit 10% getrimmtem und 20% getrimmtem Mittel

$$t_{min_t} := \min\{p_t, p_{t10}, p_{t20}\}.$$

2. Das Minimum der p-Werte aus t-Test und dem Huber-m-Test mit $c = 1,8$

$$t_{min_c} := \min\{p_t, p_{c_{1,8}}\}.$$

3. Das Minimum der p-Werte aus t-Test, 20% getrimmtem Test und dem Huber-m-Test mit $c = 1,8$

$$t_{min_{tc}} := \min\{p_t, p_{t20}, p_{c_{1,8}}\}.$$

3.2.3 Rangverfahren

Der am weitesten verbreitete Rangtest ist der Wilcoxon- oder Mann-Whitney-Test. Dieser wurde von Brunner und Puri (1996) für kleine Fallzahlen noch verbessert. Ausgehend von den Beobachtungen $X_{11}, \ldots, X_{1n_1}$ der ersten und denen der zweiten Gruppe $X_{21}, \ldots, X_{2n_2}$ werden Ränge über die gesamte Stichprobe gebildet. Hierbei werden die Ränge der Stichprobenwerte der Gruppe i mit $R_{i1}, \ldots, R_{in_i}$ bezeichnet. Die aus den Rängen berechnete gepoolte Varianz

$$s_R^2 := \frac{1}{n_1 + n_2 - 2} \sum_{i=1}^{2} \sum_{j=1}^{n_i} (R_{ij} - \bar{R}_{i\cdot})^2$$

geht in die Teststatistik

$$t_R := \frac{\bar{R}_{1\cdot} - \bar{R}_{2\cdot}}{s_R \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

der Ränge ein. Diese ist approximativ t-verteilt mit $n_1 + n_2 - 3$ Freiheitsgraden. Solch eine vorgehensweise wird auch asymptotische Rangtransformation (ART) genannt.

3.2.4 Simulationsstudien

In den Simulationsstudien wird immer zusätzlich zu den verschiedenen robusten Verfahren der t-Test betrachtet. Im Fall von normalverteilten Daten muß dieser der Test mit der höchsten Power sein, wobei die Powerkurven der robusten Verfahren möglichst nahe bei der des t-Tests liegen sollten, wohingegen bei nicht normalverteilten Daten die robusten Tests eine höhere Power haben sollten.
In Abhängigkeit von der zugrundeliegenden Verteilung ist der Lageunterschied beider Gruppen so gewählt, daß der t-Test eine Power von 70% bei einem Niveau von $\alpha = 5\%$ hat. Hat der t-Test eine sehr niedrige Power im Vergleich zu den anderen Tests, so ist der Lageunterschied so gewählt, daß einer der robusten Einzeltests eine Güte von ca. 70% hat. Die von der Gesamtfallzahl abhängigen Lageunterschiede sind in Tabelle 3.2 zusammengestellt. Die Anzahl der Wiederholungen betrug bei den Bootstrap-Verfahren 1000 Bootstrap-Wiederholungen innerhalb von 5000 Gesamtwiederholungen. Bei den anderen Verfahren wurde 10.000 mal wiederholt. Eine oberere 99% Konfidenzschranke für die simulierten Niveaus $\hat{\alpha}$ ergibt sich approximativ zu: $\hat{\alpha} + 2,326 \cdot \sqrt{\alpha(1-\alpha)n_{sim}^{-1}}$ (Efron, Tibshirani 1993). D.h. für ein nominales Niveau von $\alpha = 5\%$ liegen bei 10.000

N_{tot}	NV	LogNV	CAU	χ^2_2	KNV	(0,0.2)	(0.5,0)	(0.5,0.2)
20	1.18	2	2.4	2.23	2.1	1.6	1.4	1.9
40	0.81	1.5	1.5	1.56	1.56	1.2	1	1.4
60	0.655	1.245	1.2	1.28	1	1	0.8	1.2

Tabelle 3.2: Lageunterschiede in Abhängigkeit von der zugrundeliegenden Verteilung und Gesamtfallzahl

Simulationen alle geschätzten Niveaus unter 5,5% im Konfidenzintervall. Bei 5000 Wiederholungen sind simulierte Niveaus kleiner als 5,71% noch im Konfidenzintervall zu $\alpha = 5\%$. In den Tabellen der simulierten Niveaus sind die Werte außerhalb des Konfidenzbereichs fett gedruckt. Bei den Powertabellen sind die Werte der liberalen Tests kursiv hervorgehoben. Alle Simulationen wurden in SAS/IML programmiert.

Vergleich unter Normalverteilung

Das erste Bewertungskriterium für die Tests ist ihr Verhalten unter der Nullhypothese. Zum Vergleich sind in Tabelle 3.3 die simulierten Niveaus zu 1%, 5% und 10% im balancierten Fall aufgeführt. Dabei ist mit huber_1 das Schätzverfahren, das direkt eine gepoolte Varianz schätzt, und mit huber_2 das Verfahren von Yuen ohne die empirischen Anpassungsfaktoren bezeichnet. Es ist zu ersehen, daß alle Tests das Niveau

	$n_1 = n_2 = 10$			$n_1 = n_2 = 20$		
	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$
t-Test	1,0	5,0	9,7	1,0	5,0	9,7
10% trim	1,1	5,1	10,0	1,1	5,0	9,8
20% trim	1,3	**5,6**	10,9	1,1	5,4	10,4
huber_1 c=1,8	1,1	5,2	10,3	1,2	5,0	10,2
huber_1 c=2,4	1,0	5,2	10,1	1,0	5,1	10,0
huber_2 c=1,8	0,8	4,4	9,9	0,9	4,5	10,2
tiku 10%	1,1	5,4	10,5	1,1	5,0	10,1
tiku 20%	**1,4**	**6,0**	**11,2**	1,2	5,4	10,5
$\min_t$	1,0	5,0	10,3	0,9	5,1	9,8
$\min_c$	0,9	4,6	9,6	1,0	5,0	9,8
$\min_{tc}$	0,9	4,9	10,1	0,9	5,2	9,9
Brunner	1,2	5,3	10,7	1,0	4,9	10,4

Tabelle 3.3: Simulierte Niveaus unter Normalverteilung

unter Normalverteilung einhalten, wobei der 20% getrimmte Tiku-Test leicht liberal ist. Bezüglich der Power bei den Lageunterschieden aus Tabelle 3.4 ergibt sich folgendes Bild: Der t-Test weist, wie zu erwarten, die höchste Power auf, wobei die Huber-m-Schätzer-Tests nur eine geringfügig niedrigere Power zeigen. Insbesondere hat der

$n_1 = n_2$	t-Test	10% trim	20% trim	c=1,8	c=2,4	tiku 10%	tiku 20%
10	70,4	65,7	*59,9*	68,9	70,0	67,5	*63,0*
20	70,7	68,0	63,7	70,1	70,7	68,6	65,8
30	70,0	68,0	64,1	69,8	70,2	68,6	65,8

$n_1 = n_2$	$\min_t$	$\min_c$	$\min_{tc}$	Brunner
10	66,2	68,2	68,1	68,8
20	68,6	70,1	68,8	68,2
30	68,4	70,2	68,7	68,0

Tabelle 3.4: Simulierte Power unter Normalverteilung bei $\alpha = 5\%$

ART-Test eine geringere Power als der m-Schätzer-Test. Die geringste Power haben die trimmenden Tests. Hierbei ist der Tiku-Test immer dem einfachen Trimmen überlegen. Insgesamt liegen die Powerwerte aller Verfahren in einem Band von ca. 6%-Punkten bei kleinen Fallzahlen (10 Beobachtungen pro Gruppe), das mit wachsender Fallzahl enger wird.
Die Power der Minimum-Tests liegt in der Größenordnung des ART-Tests, wobei der Minimum-Test aus t-, trim- und m-Test die gleiche Power wie der ART-Test zeigt. Besonders fällt auf, daß der 20% getrimmte Test einen Powerverlust zwischen 6% und 10% im Vergleich zum t-Test hat, was sich jedoch auf die Power des $\min_t$-Test nicht negativ auswirkt, da dieser annähernd die Power des t-Tests hat. Allgemein ist für die Minimum-Tests zu sehen, daß ihre Power zwischen der des t-Tests und der des jeweiligen robusten Tests liegt, sie unter Normalverteilung demnach eine Verbesserung der robusten Tests darstellen. Diese an einem Punkt beschriebenen Unterschiede zeigen sich auch über das ganze Spektrum an Lokationsunterschieden wie aus den Abbildungen 3.1 und 3.2 zu erkennen ist. Im Fall von unbalancierten Fallzahlen zeigen sich fast alle

n_1/n_2	6/34	8/32	10/30	12/28	14/26	16/24	18/22	20/20
t-Test	5,4	4,9	5,1	4,7	5,4	5,1	5,2	5,0
10% trim	**6,6**	**5,8**	5,0	4,7	**5,6**	5,3	4,9	5,0
20% trim	**5,6**	**5,6**	4,9	5,2	**6,1**	5,5	5,0	5,4
huber_1 c=1,8	5,5	4,9	5,1	5,0	5,5	5,3	5,1	5,0
huber_1 c=2,4	5,5	5,0	5,1	4,9	**5,6**	5,1	5,1	5,1
huber_2 c=1,8	5,0	5,0	5,5	4,7	5,1	4,7	5,3	4,5
tiku 10%	5,1	5,0	4,1	5,0	5,1	5,0	5,0	5,0
tiku 20%	5,5	5,7	5,1	5,3	5,2	5,4	5,4	5,4
$\min_t$	5,1	4,5	4,9	4,6	5,3	5,1	4,9	5,1
$\min_c$	5,4	4,8	5,1	4,7	5,4	5,0	5,7	5,0
$\min_{tc}$	5,0	4,5	4,9	4,7	5,3	5,0	4,9	5,2
Brunner	4,7	4,9	4,8	4,5	5,4	4,9	5,0	4,9

Tabelle 3.5: Simulierte Niveaus zu $\alpha = 5\%$ unter Normalverteilung bei Unbalanciertheit

Tests als niveautreu (siehe Tabelle 3.5). Nur bei extremen Fallzahlunterschieden (z.B. $n_1 = 8, n_2 = 32$) werden die 20% getrimmten Tests liberal. Dies ist auf den Effekt der Freiheitsgradreduktion von 8 auf 4 und des daher stark veränderten Quantil zurückzuführen. Bezüglich der Güte bei Unbalanciertheit zeigen alle Tests einen Abfall bei steigender Unbalanciertheit. Zwischen den beiden verschiedenen c's beim huber_1-Test liegt kein Güteunterschied vor. Ansonsten verhalten sich die Tests wie bei balancierten Fallzahlen.

n_1/n_2	6/34	8/32	10/30	12/28	14/26	16/24	18/22	20/20
t-Test	42,8	52,5	57,8	62,5	66,6	68,7	69,7	70,7
10% trim	*45,6*	*54,8*	55,5	59,9	*64,8*	66,5	67,7	68,0
20% trim	*38,9*	*50,7*	51,4	56,7	*62,2*	61,4	64,3	63,7
huber_1 c=1,8	42,1	51,4	57,2	61,7	65,8	67,7	68,7	70,1
huber_1 c=2,4	42,7	52,0	57,6	62,0	*66,4*	68,4	69,3	70,7
huber_2 c =1,8	42,7	51,9	58,0	63,1	65,1	68,2	70,0	70,3
tiku 10%	41,9	50,9	55,6	62,4	65,0	67,7	69,3	68,6
tiku 20%	39,4	48,9	52,4	59,5	62,9	64,1	66,2	65,8
$\min_t$	40,5	49,7	55,3	59,9	64,5	66,4	67,7	68,6
$\min_c$	42,1	51,5	57,2	61,7	65,8	68,1	69,0	70,1
$\min_{tc}$	40,4	49,5	55,3	59,9	64,4	66,6	67,7	68,8
Brunner	40,7	48,7	56,0	60,0	65,2	67,1	68,3	68,2

Tabelle 3.6: Simulierte Power zu $\alpha = 5\%$ unter Normalverteilung bei Unbalanciertheit

Vergleich unter Ausreißerverteilungen

Nachdem sich alle betrachteten robusten Tests unter Normalverteilung bei üblichen Fallzahlen als niveautreu und gütestabil erweisen, interessiert ihr Verhalten in nichtnormalen Situationen. Zuerst ist auch hier die Forderung aufzustellen, daß die Tests das vorgegebene Niveau nicht überschreiten dürfen. In Tabelle 3.7 fallen dazu drei Tests auf. Zum einen erweist sich das Tiku-Verfahren mit 10% oder 20% Trimmung als liberal in einigen Situationen. Zum anderen hält der Huber-m-Test mit den für jede Gruppe einzeln geschätzten Varianzen als Einschritt-Verfahren nicht das Niveau. Deswegen werden diese Ansätze nicht weiter untersucht. Alle anderen Tests halten das Niveau, wobei der t-Test teilweise extrem konservativ (z.B. Cauchy-Verteilung) wird[2]. Auch der ART-Test zeigt in einigen Fällen solch ein konservatives Verhalten.
Bei gleichen Fallzahlen lassen sich folgende Schlüsse aus den Simulationsergebnissen (siehe Tabelle 3.8) ziehen: Zuallererst sehen wir, daß der t-Test in allen Fällen einen

[2]Die Tabellen zeigen die Ergebnisse bei $\alpha = 5\%$; diese unterscheiden sich nicht von denjenigen bei $\alpha = 1\%$ oder $\alpha = 10\%$.

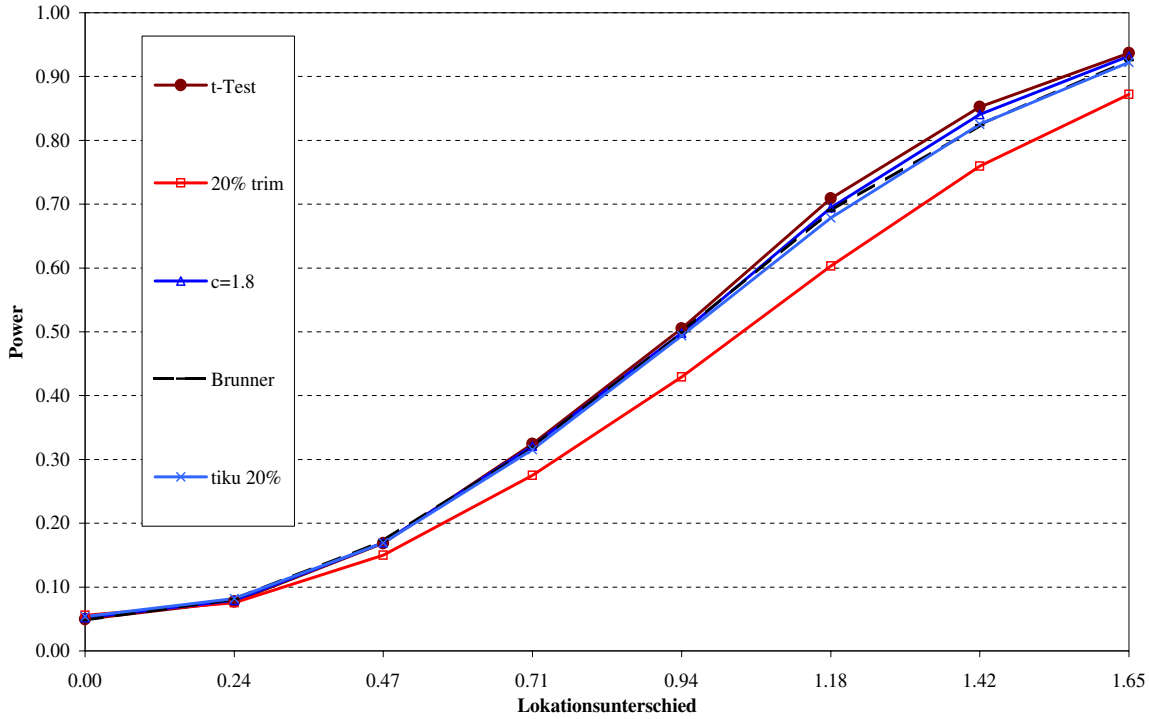


Abbildung 3.1: Power der robusten Tests im Vergleich zu t-Test und Rangtransformation unter Normalverteilung, $n_1 = n_2 = 10$

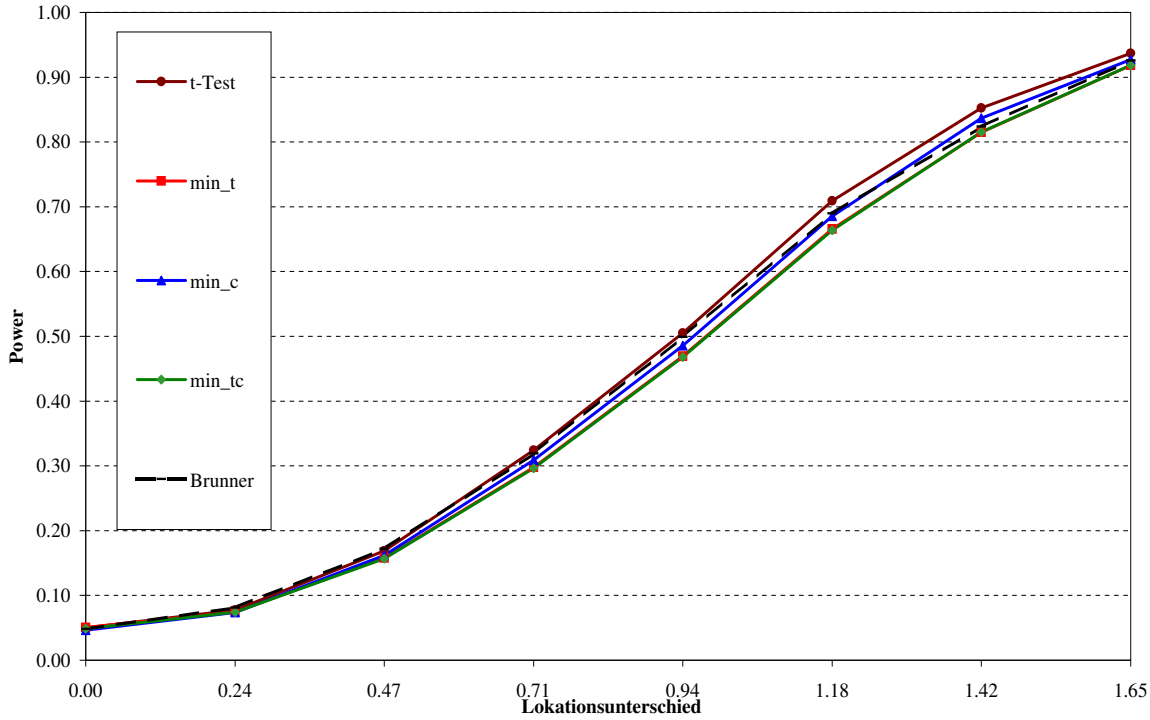


Abbildung 3.2: Power der Minimum-Tests im Vergleich zu t-Test und Rangtransformation unter Normalverteilung, $n_1 = n_2 = 10$

deutlichen Powerverlust gegenüber den robusten Tests hat. Trotz seiner Konservativität weist der ART-Test eine hohe Güte auf. Für die Minimum-Tests gilt auch hier

wieder, daß sich ihre Güte zwischen der des t-Tests und der des robusten Tests befindet. Insgesamt erweisen sich die Minimum-Tests als gütestabil über alle Situationen hinweg. Ist die zugrundeliegende Verteilung schief, so hat der $\min_c$-Test meist eine höhere Power als der $\min_t$-Test. Hat dagegen die Verteilung relativ viel Masse weit außerhalb, so verhält es sich gerade anders herum. Diesen Effekt mittelt der $\min_{tc}$-Test durch das Verwenden von beiden robusten Schätzern aus, so daß sich dieser Test als robust und gütestabil über alle Situationen erweist. Für den Tiku-Test fällt noch auf, daß er trotz seiner leichten Liberalität bei den betrachteten Fallzahlen nicht zu den besonders powervollen Tests zählt.

An dem Beispielen der Chiquadrat-Verteilung mit zwei Freiheitsgraden und der kontaminierten Normalverteilung ist aus den Abbildungen 3.3 bis 3.6 zu ersehen, daß diese punktuellen Beobachtungen sich über das ganze Spektrum der möglichen Lageunterschiede allgemein gelten. Insbesondere fällt bei der kontaminierten Normalverteilung auf, daß bei einem Lageunterschied größer 2 alle robusten Tests fast Power 1 haben, der t-Test jedoch lediglich über eine Power von 40 % verfügt.

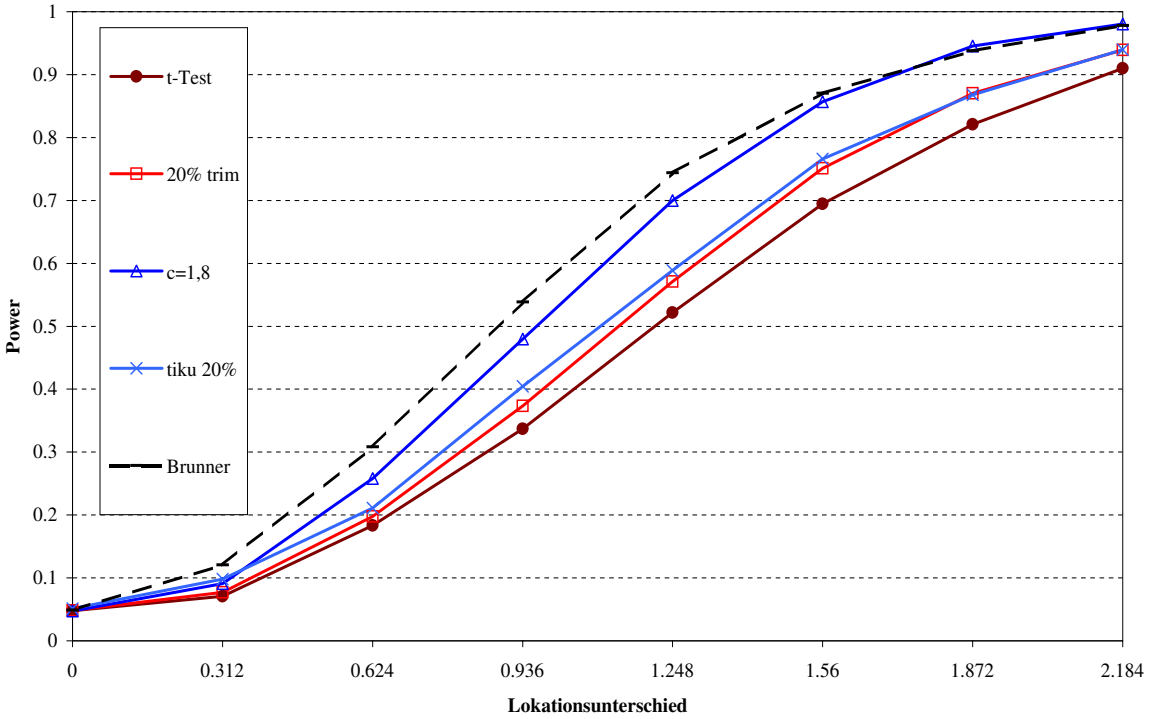


Abbildung 3.3: Power der robusten Tests im Vergleich zu t-Test und Rangtransformation unter einer $\mathcal{X}^2_2$-Verteilung, $n_1 = n_2 = 20$

n=10	LogNV	CAU	$\mathcal{X}_2^2$	KNV	(0;0,2)	(0,5;0)	(0,5;0,2)
t-Test	3,2	2,1	4,2	2,7	4,5	3,8	3,5
10% trim	3,2	2,7	4,1	2,5	4,5	4,1	4,2
20% trim	4,1	3,3	4,6	3,8	5,0	4,5	4,6
huber_1 c=1,8	3,4	4,8	4,4	3,6	4,4	4,1	4,1
huber_1 c=2,4	2,9	3,8	4,1	2,9	4,1	3,8	3,8
huber_2 c=1,8	**18,5**	**18,9**	**7,2**	**6,8**	**9,1**	**14,3**	**13,9**
tiku 10%	4,4	4,1	4,5	3,7	**5,8**	4,8	4,7
tiku 20%	4,5	4,3	5,0	5,0	**5,9**	5,1	5,0
$\min_t$	4,3	4,1	4,9	4,5	4,9	4,8	4,6
$\min_c$	3,9	4,6	4,6	4,4	4,4	4,4	4,2
$\min_{tc}$	4,2	4,7	4,7	4,7	5,0	4,8	4,6
Brunner	3,5	1,9	4,4	2,5	4,5	4,3	3,8

n=20	LogNV	CAU	$\mathcal{X}_2^2$	KNV	(0;0,2)	(0,5;0)	(0,5;0,2)
t-Test	3,9	2,1	4,6	4,1	4,7	4,2	4,0
10% trim	3,9	3,1	4,3	2,5	4,9	4,7	4,2
20% trim	3,9	3,5	4,5	4,0	5,1	4,6	4,6
huber_1 c=1,8	3,8	5,0	4,8	4,6	4,7	4,6	4,6
huber_1 c=2,4	3,2	4,2	4,7	3,7	4,4	4,4	4,1
huber_2 c=1,8	**18,9**	**20,2**	**7,2**	**7,9**	**10,2**	**12,7**	**15,7**
tiku 10%	**5,8**	**7,3**	**5,7**	**6,8**	**6,2**	**5,6**	**5,9**
tiku 20%	5,2	**7,0**	**5,7**	**5,6**	**6,1**	**5,7**	**5,9**
$\min_t$	4,8	4,0	4,9	5,0	4,8	4,8	4,7
$\min_c$	4,8	4,6	5,0	5,0	5,0	5,0	4,8
$\min_{tc}$	4,8	4,3	5,0	5,1	4,9	5,1	4,8
Brunner	4,1	2,0	4,7	3,6	4,4	4,8	4,3

n=30	LogNV	CAU	$\mathcal{X}_2^2$	KNV	(0;0,2)	(0,5;0)	(0,5;0,2)
t-Test	4,1	1,9	5,1	4,6	4,8	4,2	4,2
10% trim	4,1	3,6	5,0	3,0	4,7	4,7	4,3
20% trim	4,2	4,4	5,6	5,0	4,6	4,8	4,5
huber_1 c=1,8	3,8	5,2	5,2	5,1	4,8	4,8	4,5
huber_1 c=2,4	3,6	4,3	5,1	4,2	4,5	4,4	4,1
huber_2 c=1,8	**19,2**	**22,3**	**7,3**	**8,7**	**9,9**	**13,5**	**16,0**
tiku 10%	**6,6**	**9,4**	**5,8**	**9,6**	**6,3**	**5,6**	**6,9**
tiku 20%	**5,9**	**7,9**	5,2	**6,3**	**5,9**	5,5	**6,7**
$\min_t$	4,8	4,7	5,3	5,4	4,8	4,8	4,9
$\min_c$	4,9	4,7	5,3	5,3	5,1	5,1	4,8
$\min_{tc}$	4,8	4,8	5,3	5,3	5,0	4,9	4,9
Brunner	4,0	2,0	4,9	4,7	4,7	4,3	4,5

Tabelle 3.7: Simulierte Niveaus unter Nicht-Normalverteilung für $\alpha = 5\%$, $n_1 = n_2 = n = 10, 20, 30$

n=10	LogNV	CAU	$\mathcal{X}_2^2$	KNV	(0;0,2)	(0,5;0)	(0,5;0,2)
t-Test	69,4	24,8	69,6	29,3	71,0	71,6	71,3
10% trim	81,4	54,1	72,1	58,7	79,4	74,5	83,2
20% trim	84,4	68,3	70,9	76,0	78,9	71,7	85,6
huber$_1$ c=1,8	92,7	67,0	83,7	74,9	79,9	80,5	88,4
huber$_1$ c=2,4	90,1	60,3	81,2	68,0	77,9	78,3	86,4
tiku 10%	81,1	56,7	74,2	63,4	*78,8*	76,0	83,8
tiku 20%	85,7	71,5	73,8	79,1	*79,5*	74,2	86,6
min$_t$	85,5	67,9	75,0	76,1	79,7	75,7	86,2
min$_c$	91,6	64,6	81,9	72,9	78,6	79,1	87,3
min$_{tc}$	90,5	68,9	79,8	77,3	79,4	77,1	87,5
Brunner	88,4	59,9	80,2	69,5	77,4	80,2	85,7

n=20	LogNV	CAU	$\mathcal{X}_2^2$	KNV	(0;0,2)	(0,5;0)	(0,5;0,2)
t-Test	70,4	13,2	69,5	25,2	74,5	73,7	71,8
10% trim	88,3	49,5	75,3	66,7	86,3	78,2	89,2
20% trim	92,1	69,2	75,1	88,3	87,4	77,6	91,8
huber$_1$ c=1,8	96,8	61,3	85,7	83,3	85,8	84,2	93,0
huber$_1$ c=2,4	95,4	54,8	83,2	76,7	84,1	82,0	90,9
tiku 10%	*87,1*	*52,7*	*75,4*	*68,0*	*85,4*	*79,5*	*87,9*
tiku 20%	91,2	*69,5*	*76,4*	*88,4*	*88,1*	*78,8*	*91,4*
min$_t$	91,4	66,4	76,8	86,5	86,5	79,2	91,1
min$_c$	96,4	57,7	84,2	80,5	84,7	83,2	91,9
min$_{tc}$	95,8	65,7	82,4	86,5	86,4	81,2	92,1
Brunner	96,9	62,3	86,6	82,3	86,1	85,6	93,1

n=30	LogNV	CAU	$\mathcal{X}_2^2$	KNV	(0;0,2)	(0,5;0)	(0,5;0,2)
t-Test	69,2	9,8	70,0	16,2	76,5	71,7	73,0
10% trim	90,0	51,3	75,5	54,3	88,6	77,7	92,3
20% trim	94,0	71,2	*76,4*	75,2	89,5	76,3	95,0
huber$_1$ c=1,8	98,1	61,9	87,1	66,4	87,6	83,1	95,4
huber$_1$ c=2,4	96,6	54,7	84,1	58,6	86,1	81,0	93,9
tiku 10%	*88,2*	*52,8*	*76,5*	*59,3*	*88,0*	*78,0*	*91,1*
tiku 20%	*93,1*	*71,0*	77,9	*76,7*	*90,1*	78,6	*97,3*
min$_t$	92,9	67,7	77,0	69,9	88,5	78,1	94,0
min$_c$	97,6	57,3	85,4	61,2	86,3	81,8	94,4
min$_{tc}$	97,0	66,4	83,7	69,5	88,6	80,4	94,8
Brunner	98,7	63,7	90,6	68,8	89,0	86,4	96,4

Tabelle 3.8: Simulierte Power unter Nicht-Normalverteilung bei $\alpha = 5\%$, $n_1 = n_2 = n = 10, 20, 30$

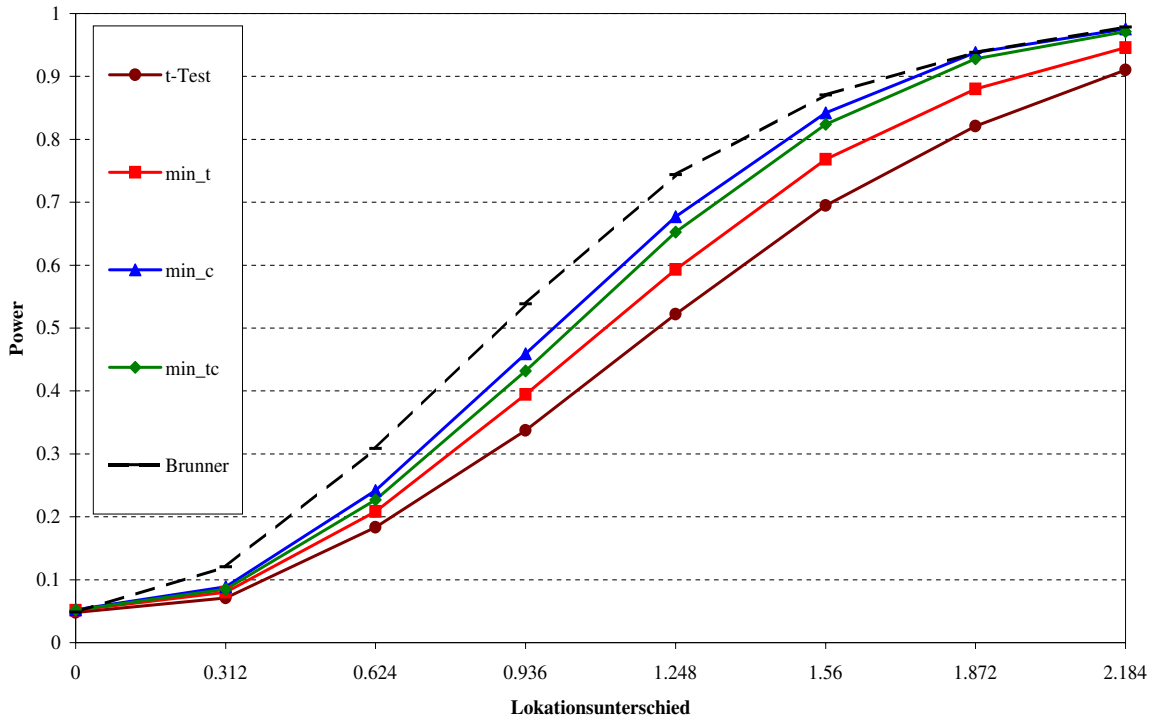


Abbildung 3.4: Power der Minimum-Tests im Vergleich zu t-Test und Rangtransformation unter einer $\mathcal{X}_2^2$-Verteilung, $n_1 = n_2 = 20$

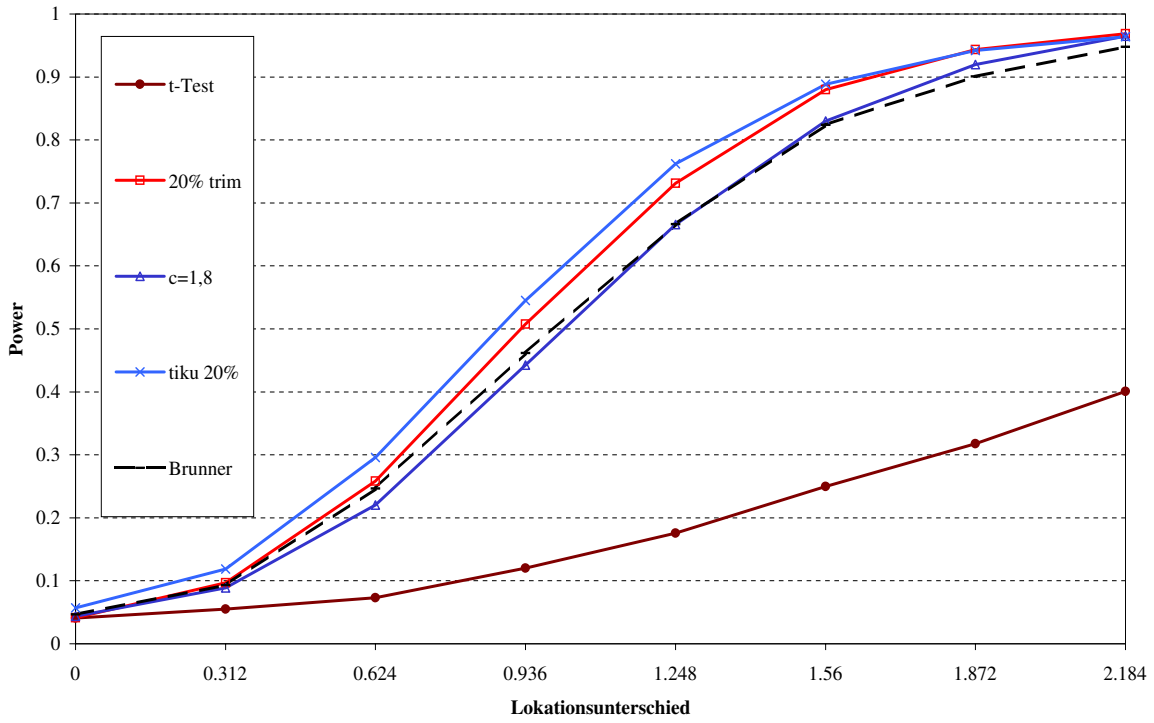


Abbildung 3.5: Power der robusten Tests im Vergleich zu t-Test und Rangtransformation unter einer kontaminierten Normalverteilung, $n_1 = n_2 = 20$

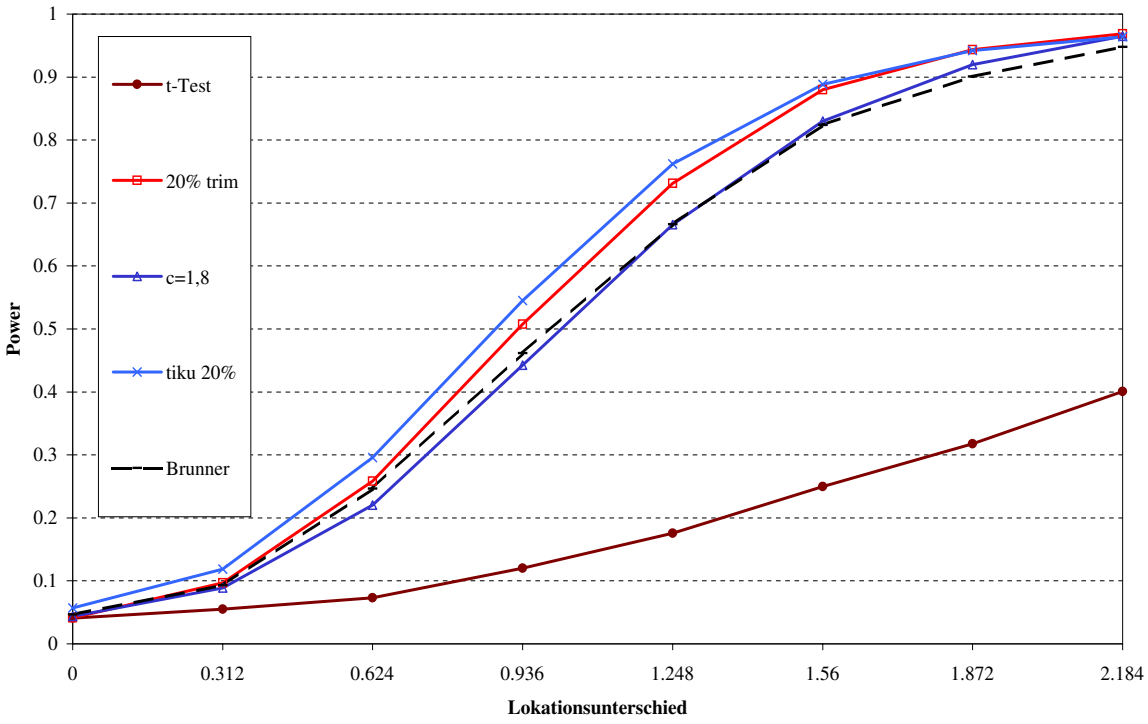


Abbildung 3.6: Power der Minimum-Tests im Vergleich zu t-Test und Rangtransformation unter einer kontaminierten Normalverteilung, $n_1 = n_2 = 20$

3.3 Robuste Multiple Vergleiche

Um zu robusten multiplen Verfahren zu kommen, lassen sich verschiedene Wege einschlagen. Zum einen werden im folgenden die robusten 2-Stichprobentests auf den k-Stichproben-Fall verallgemeinert. Andererseits lassen sich die robusten 2-SP-Tests auch in die verschiedenen α-Adjustierungsverfahren wie zum Beispiel in die Verfahren von Holm (1979), Hochberg (1988) oder Hommel (1988) einsetzen. Dieses Verfahren hat den Vorteil, daß die multivariate Verteilung der Lageschätzer nicht bestimmt werden muß.
Da nur die asymptotischen Verteilungen bekannt sind, wird untersucht, inwieweit sich für kleine Fallzahlen die Verteilungen der Testgrößen durch die multivariate t-Verteilung approximieren lassen. Ein anderer Ausweg läge darin, ein Bootstrap-Verfahren zu verwenden. Ein wichtiges Beurteilungskriterium bei der Untersuchung der verschiedenen Verfahren ist die Einhaltung der globalen Niveaus. Zudem werden die Güten der einzelnen Verfahren in einer Simulationsstudie mit den vorgestellten Verteilungen aus Kapitel 1 verglichen. Insbesondere werden die robusten Analoga zum Dunnett- und Tukey-Test behandelt.

3.3.1 Simultane Vergleiche auf Basis der multivariaten t-Verteilung

Die in den vorhergehenden Kapiteln vorgestellten parametrischen Schätzer und Zwei-Stichproben-Tests werden in diesem Abschnitt zu k-Stichproben-Tests verallgemeinert. Ausgehend von der asymptotischen multivariaten Normalverteilung des Vektors der Lageschätzer, die in den einzelnen Abschnitten gezeigt wird, ist es in parametrischen Fall naheliegend, für kleine Fallzahlen die Verteilung des Schätzvektors durch die multivariate t-Verteilung zu approximieren. Dafür werden folgende Notationen verwendet:

$$\begin{array}{ll} X_{11}, X_{12}, \ldots, X_{1n_1} & \text{Beobachtungen in Gruppe 1} \\ X_{21}, X_{22}, \ldots, X_{2n_2} & \text{Beobachtungen in Gruppe 2} \\ \vdots & \\ X_{k1}, X_{k2}, \ldots, X_{kn_k} & \text{Beobachtungen in Gruppe } k \end{array}$$

Hierbei ist $n_i\,(i = 1, \ldots, k)$ die Anzahl der Beobachtungen in der i-ten Gruppe. Wenn mit $\hat{\mu}_1, \ldots, \hat{\mu}_k$ die Lageschätzer für die einzelnen Gruppen bezeichnet werden, so setze man $\hat{\mu} := (\hat{\mu}_1, \ldots, \hat{\mu}_k)'$. Ein Kontrast $C := (c_1, \ldots, c_k)$ mit $\sum c_i = 0$ von $\hat{\mu}$ läßt sich somit schreiben als $C\hat{\mu}$. Sind für einen multiplen Kontrast die einzelnen Kontraste mit $C_1, \ldots, C_g$ bezeichnet, so hat der multiple Kontrast die Form $t^{MC} = \max\{t_1^{SC}, \ldots, t_g^{SC}\}$

mit

$$\begin{pmatrix} t_1^{SC} \\ \vdots \\ t_g^{SC} \end{pmatrix} := \begin{pmatrix} C_1 \\ \vdots \\ C_g \end{pmatrix} \cdot \begin{pmatrix} \hat{\mu}_1 \\ \vdots \\ \hat{\mu}_k \end{pmatrix}. \tag{3.24}$$

Die Darstellung in (3.24) zeigt, daß mit Kenntnis der (asymptotischen) Verteilung von $\hat{\mu}$ auch die Verteilung von t^{MC} über den Abbildungssatz bekannt ist.

Trimmen und Winsorisieren

Um die Verteilung eines multiplen Tests basierend auf Trimmschätzern zu bestimmen, ist zuerst die multivariate Verteilung des Vektors, bestehend aus den Schätzern je Gruppe, zu ermitteln. Sei zuerst angenommen, daß die Varianz bekannt und in allen Gruppen gleich ist. Unter diesen Bedingungen und der Voraussetzung von Satz 3.1 ist der Vektor der Trimmschätzer asymptotisch multivariat normalverteilt.

Satz 3.5
Seien $X_{11}, \ldots, X_{1n_1}, X_{21}, \ldots, X_{kn_k}$ unabhängig und identisch nach einer Verteilungsfunktion F verteilt. Weiterhin sei F monoton wachsend an den Trimmstellen und es gelte $\frac{n_j}{\min n_i} \to 1$ für $\min n_i \to \infty$, so gilt:

$$\sqrt{\min n_i} \left(\begin{pmatrix} \hat{\mu}_{t1} \\ \vdots \\ \hat{\mu}_{tk} \end{pmatrix} - \begin{pmatrix} \mu_{t1} \\ \vdots \\ \mu_{tk} \end{pmatrix} \right) \xrightarrow{\mathcal{D}} \mathcal{N}_k(O, \sigma^2(\gamma, F) \cdot E_k). \tag{3.25}$$

Beweis: Unter den Voraussetzungen gilt für jede der Komponenten mit Satz 3.1

$$\sqrt{\frac{\min n_i}{n_j}} \sqrt{n_j}(\hat{\mu}_{tj} - \mu_{tj}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(\gamma, F)), \qquad j = 1, \ldots, k.$$

Damit folgt für die charakteristischen Funktionen von $\hat{\mu}_{tj} - \mu_{tj}$ mit dem Stetigkeitssatz für charakteristische Funktionen (Billingsley 1995, Seite 349), daß $\phi_{jn}(s) \to \phi_j(s) = \mathrm{e}^{-\frac{1}{2}t^2\sigma^2(\gamma,F)}$ für alle $t \in \mathbb{R}$. Um nun (3.25) zu zeigen, wird:

$$\begin{aligned} &\sqrt{\min n_i}\,(a_1, \ldots, a_k) \cdot (\hat{\mu}_{t1} - \mu_{t1}, \ldots, \hat{\mu}_{tk} - \mu_{tk})' \\ &=: (a_1, \ldots, a_k) \cdot (\tilde{\mu}_{t1}, \ldots, \tilde{\mu}_{tk})' \\ &= \sum_{j=1}^{k} a_j \tilde{\mu}_{tj} = a' \tilde{\mu}_t \end{aligned} \tag{3.26}$$

mit $\|a\|_2 = 1$ betrachtet. Für diese Summe ist zu zeigen, daß sie für alle $a \in S^{k-1}$ gegen eine Normalverteilung konvergiert. Die charakteristische Funktion von (3.26) läßt sich

aufgrund der Unabhängigkeit der $\tilde{\mu}_{ti}$ schreiben als:

$$\begin{aligned}
\phi_{a'\tilde{\mu}_t}(s) &= \prod_{j=1}^{k} \phi_{a_j\tilde{\mu}_{tj}}(s) \\
&= \prod_{j=1}^{k} \phi_{\tilde{\mu}_{tj}}(a_j s) \\
&\longrightarrow \prod_{j=1}^{k} \mathrm{e}^{-\frac{1}{2}(a_j s)^2\sigma^2(\gamma,F)} \qquad n \to \infty \\
&= \mathrm{e}^{-\frac{1}{2}\sum_{i=j}^{k}(a_j s)^2\sigma^2(\gamma,F)} \\
&= \mathrm{e}^{-\frac{1}{2}s^2\sigma^2(\gamma,F)\sum_{j=1}^{k}a_j^2} \\
&= \mathrm{e}^{-\frac{1}{2}s^2\sigma^2(\gamma,F)}.
\end{aligned}$$

Wiederum mit dem Stetigkeitssatz ergibt sich, daß jede Projektion von $\tilde{\mu}_t$ eindimensional normalverteilt ist.

$\triangledown$

Nun ist im allgemeinen die Varianz $\sigma^2(\gamma, F)$ nicht bekannt. Wird die unbekannte Varianz jedoch konsistent geschätzt, so gilt:

Lemma 3.6
Mit den Voraussetzungen von Satz 3.5 und einem konsistenten Schätzer $\hat{\sigma}^2(\gamma)$ für $\sigma^2(\gamma, F)$ gilt

$$\frac{\sqrt{\min n_i}}{\hat{\sigma}(\gamma)} \left(\begin{pmatrix} \hat{\mu}_{t1} \\ \vdots \\ \hat{\mu}_{tk} \end{pmatrix} - \begin{pmatrix} \mu_{t1} \\ \vdots \\ \mu_{tk} \end{pmatrix} \right) \xrightarrow{\mathcal{D}} \mathcal{N}_k(O, E_k). \tag{3.27}$$

Damit ergibt sich aus der Kombination von (3.24) und (3.27), daß ein multipler Kontrast von Trimmschätzern asymptotisch multivariat normalverteilt ist. Es bleibt die Frage, wie die Verteilung im Endlichen lautet. In Analogie zum klassischen multiplen Kontrasttest, basierend auf Mittelwert und Standardabweichung, läßt sich vermuten, daß zumindest eine Approximation durch die multivariate t-Verteilung für kleine Fallzahlen machbar ist. Im Gegensatz zum Zwei-Stichproben-Fall, in dem es galt, den „richtigen“ Freiheitsgrad zu bestimmen, gilt es hier, die „richtige“ Korrelationsmatrix zu finden. Für zwei Kontraste $(c_{a1}, \ldots, c_{ak}), (c_{b1}, \ldots, c_{bk})$ lautet der Eintrag in der normalen Korrelationsmatrix $R = (\rho_{ab})_{a,b=1,\ldots,g}$ (Bechhofer, Dunnett 1982):

$$\rho_{ab} := \frac{\sum_{i=1}^{k} \frac{c_{ai}c_{bi}}{n_i}}{\sqrt{\left(\sum_{i=1}^{k} \frac{c_{ai}^2}{n_i}\right)\left(\sum_{i=1}^{k} \frac{c_{bi}^2}{n_i}\right)}}. \tag{3.28}$$

Da durch das Trimmen nicht alle n_i Beobachtungen der $i-$ten Gruppe verwendet werden, ersetze man n_i durch h_i

$$\rho_{ab} := \frac{\sum_{i=1}^{k} \frac{c_{ia}c_{ib}}{h_i}}{\sqrt{\left(\sum_{i=1}^{k} \frac{c_{ia}^2}{h_i}\right)\left(\sum_{i=1}^{k} \frac{c_{ib}^2}{h_i}\right)}}.$$

Die gepoolte Varianz schätze man konsistent in direkter Verallgemeinerung des Zwei-Stichproben-Falls durch

$$\hat{\sigma}^2 := \frac{1}{\sum h_i - k} \sum_{i=1}^{k} (h_i - 1) s_{Yi}^2$$

bzw. die Freiheitsgrade durch $\nu = \sum_{i=1}^{k}(h_i - 1)$. Das approximative Konfidenzintervall für jeden Kontrast ergibt sich so zu:

$$\left(-\infty, \sum_{i=1}^{k} c_i \hat{\mu}_{ti} + T_g(-\infty, 1_g t^{MC}, \nu, R)\hat{\sigma}\sqrt{\sum_{i=1}^{k} \frac{c_i^2}{h_i - 1}} \right].$$

M-Schätzer

Wie im vorherigen Abschnitt wird zuerst gezeigt, daß der Vektor bestehend aus den m-Schätzern für die einzelnen Gruppen asymptotisch normalverteilt ist. Um die asymptotische Normalität des Vektors aus den m-Schätzern zu zeigen, wird die Gültigkeit der Voraussetzungen der Proposition 3.1 aus Huber 1973 gezeigt. Dieser Satz impliziert dann direkt die gewünschte Normalität, wie in dem Artikel gezeigt wird.
Die Beobachtungen seien wie in Abschnitt 3.3.1 bezeichnet, und es gelte $\min n_i \to \infty$. Weiterhin seien die X_{ij} unabhängig nach einer Verteilungsfunktion F verteilt, die nicht notwendigerweise symmetrisch sein muß, so daß nach folgender Setzung

$$Y_i := X_{jl} := \theta_j + U_i, \qquad i = \sum_{m=0}^{j-1} n_m + l \tag{3.29}$$

die U_i unabhängig und identisch verteilt sind. In der Form von Gleichung (1.1) in Huber 1973 lautet (3.29):

$$Y_i = \sum_{j=1}^{k} c_{ij}\theta_j + U_i$$

beziehungsweise in Matrixschreibweise

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_{n_1} \\ Y_{n_1+1} \\ \vdots \\ Y_{n_1+n_2} \\ \vdots \\ Y_{\sum n_i - n_k + 1} \\ \vdots \\ Y_{\sum n_i} \end{pmatrix} = \begin{pmatrix} 1 & 0 & & \\ \vdots & \vdots & & \\ 1 & 0 & & \\ 0 & 1 & & O \\ \vdots & \vdots & & \\ 0 & 1 & & \\ & & \ddots & \\ & O & & 1 \\ & & & \vdots \\ & & & 1 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_k \end{pmatrix} + \begin{pmatrix} U_1 \\ \vdots \\ U_{\sum n_i} \end{pmatrix}$$

Im folgenden schreibe man kurz:

$$Y = C\theta + U.$$

Mit dieser Definition gilt nun:

$$CC^T = \begin{pmatrix} n_1 & & & O \\ & n_2 & & \\ & & \ddots & \\ O & & & n_k \end{pmatrix}$$

und

$$\Gamma := C(C^TC)^{-1}C^T = \begin{pmatrix} \frac{1}{n_1} & \cdots & \frac{1}{n_1} & & & \\ \vdots & \ddots & \vdots & & O & \\ \frac{1}{n_1} & \cdots & \frac{1}{n_1} & & & \\ & & & \ddots & & \\ & & & & \frac{1}{n_k} \cdots \frac{1}{n_k} & \\ & O & & & \vdots \ddots \vdots & \\ & & & & \frac{1}{n_k} \cdots \frac{1}{n_k} & \end{pmatrix}.$$

Satz 3.7 (Proposition 3.1 Huber 1973)
Unter den folgenden Bedingungen ist der Vektor aus den Lage-m-Schätzern asymptotisch normalverteilt:

1. $k \cdot \max \gamma_{ii} \to 0$
2. ψ ist konvex und $\psi \in C^3$
3. $\mathbf{E}\left[\psi(U_i)\right] = 0$, U_i unabhängig und identisch verteilt.

Für Bedingung 1 gilt $k \cdot \max \gamma_{ii} = \frac{k}{\min n_i} \to 0$, da $\min n_i \to \infty$. Die weiteren Bedingungen sind für die jeweilige Wahl der Gewichtsfunktion im einzelnen zu prüfen. Für den Huber-m-Schätzer ist Bedingung 3 aufgrund der Symmetrie erfüllt, auch ist die Existenz der dritten Ableitung gegeben. Die gewählte Funktion ψ ist jedoch nicht konvex. Die Konvexitätsforderung wird jedoch nur für die Existenz der Schätzer benötigt, die auf Grund der Wahl der Einschritt-Schätzer gewährleistet ist.
Zur Durchführung eines multiplen Tests schätze man anhand der Gleichungen (3.8)(3.9) $\hat{\mu}_{\psi i}$ und $s^2_{\psi i}$ für jede Gruppe einzeln. Da die beiden Schätzer asymptotisch unabhängig sind, läßt sich die gepoolte Varianz schätzen durch

$$s^2_\psi := \frac{1}{n_1 + \cdots + n_k - k} \sum_{i=1}^{k} (n_i - 1) s^2_{\psi i}. \tag{3.30}$$

Mit dem Lemma von Sluzky (z.B. Henze 1995) erhält man somit die asymptotische Normalität.

Satz 3.8

Unter den oben aufgelisteten Bedingungen gilt für den Vektor $\hat{\mu}$ der Huber-m-Schätzer mit dem gepoolten Varianzschätzer aus (3.30)

$$\frac{1}{s_\psi} \left(\begin{pmatrix} \hat{\mu}_{\psi 1} \\ \vdots \\ \hat{\mu}_{\psi k} \end{pmatrix} - \begin{pmatrix} \mu_{\psi 1} \\ \vdots \\ \mu_{\psi k} \end{pmatrix} \right) \xrightarrow{\mathcal{D}} \mathcal{N}_k(O, E_k).$$

Da sich in den Simulationen für den Zwei-Stichproben-Fall gezeigt hat, daß die Schätzung der gepoolten Varianz durch (3.30) für Fallzahlen unter 30 und nicht normalverteilte Daten zu einem zu liberalen Test führt, wird der Ansatz aus den Gleichungen (3.21) bis (3.23) noch weiter verfolgt. Für k Gruppen erhält man die impliziten Gleichungen:

$$\sum_{i=1}^{n_j} \psi \left(\frac{x_{ij} - \hat{\mu}_{\psi j}}{s_\psi} \right) = 0, \quad j = 1, \ldots, k \text{ und} \tag{3.31}$$

$$\sum_{j=1}^{k} \sum_{i=1}^{n_j} \psi^2 \left(\frac{x_{ij} - \hat{\mu}_{\psi j}}{s_\psi} \right) = (\sum n_j - 1)\beta. \tag{3.32}$$

Nach Bachmaier und Precht (1995) bzw. Hampel et al (1986) sind die Schätzer aus (3.31) und (3.32) asymptotisch unabhängig. Anstelle der exakten Nullstellen werden wieder die Einschritt-Schätzer für $\hat{\mu}_{\psi 1}, \ldots, \hat{\mu}_{\psi k}$ und s_ψ analog zu (3.8) und (3.9) verwendet.

$$\hat{\mu}_{\psi j} := \hat{\mu}_{0j} + \frac{\sum_{i=1}^{n_j} \psi \left(\frac{x_{ij} - \hat{\mu}_{0j}}{s_0} \right) s_0}{\sum_{i=1}^{n_j} \psi' \left(\frac{x_{ij} - \hat{\mu}_{0j}}{s_0} \right)} \tag{3.33}$$

$$\hat{s}^2_\psi := \frac{1}{(\sum n_j - 1)\beta} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \psi^2 \left(\frac{x_{ij} - \hat{\mu}_{0j}}{s_0} \right) s^2_0, \tag{3.34}$$

mit den Gruppenmedianen als Startwerten μ_{0j} und der 1,483-fachen MAD über die gepoolte Stichprobe als Startwert s_0 für σ_ψ. Der sich so ergebende Varianzschätzer s_ψ wird noch, wie in Bachmaier und Precht (1997) vorgeschlagen, mit Hubers Korrekturfaktor κ^2 (Huber 1973) modifiziert:

$$\kappa := 1 + \frac{k}{n} \cdot \frac{\text{var}(\psi')}{\text{ave}(\psi')^2}$$

$$\text{mit ave}(\psi') := \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \psi' \left(\frac{x_{ij} - \hat{\mu}_{\psi j}}{s_\psi} \right)$$

$$\text{und var}(\psi') := \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \psi' (\left(\frac{x_{ij} - \hat{\mu}_{\psi j}}{s_\psi} \right) - \text{ave}(\psi'))^2$$

Die Quantile eines mit diesem Schätzer durchgeführten multiplen Kontrasttests werden durch eine multivariate t-Verteilung mit der sich aus den Kontrastkoeffizienten und den Fallzahlen n_i nach (3.28) berechneten Korrelationsmatrix approximiert.
Das approximative Konfidenzintervall für jeden Kontrast ergibt sich so zu:

$$\left(-\infty, \sum_{i=1}^{k} c_i \hat{\mu}_{\psi i} + T_g(-\infty, 1_g t^{MC}, N-k, R) \hat{s}_\psi \sqrt{\sum_{i=1}^{k} \frac{c_i^2}{n_i}} \right].$$

3.3.2 Rangverfahren

Zu Verfahren, die auf Rängen basieren, werden im folgenden zwei Ansätze vorgestellt. Dies sind zum einen eine direkte Verallgemeinerung des Zwei-Stichproben-Rangverfahrens aus Abschnitt 3.2.3 und zum anderen der von Hettmansperger und McKean (1998) verfolgte Ansatz. Ein umfassendere Darstellung der Theorie der Rangtests ist z.B. in Bühning und Trenkler (1992) oder auch Hajek et al (1999) gegeben.
In der Verallgemeinerung des Verfahrens aus Abschnitt 3.2.3 werden die Ränge $R_{11}, \ldots, R_{kn_k}$ über alle Beobachtungen $X_{11}, \ldots, X_{kn_k}$ gebildet. Nach dieser Rangtransformation wird die gepoolte Varianz geschätzt durch:

$$s_R^2 := \frac{1}{\sum_{i=1}^{k} n_i - k} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (R_{ij} - \bar{R}_{i.})^2.$$

Als Teststatistik für einen einzelnen Kontrast $c = (c_1, \ldots, c_k)$ ergibt sich damit:

$$T^{SCR} := \frac{\sum_{i=1}^{k} c_i \bar{R}_{i.}}{s_R \sqrt{\sum_{i=1}^{k} \frac{c_i^2}{n_i}}}.$$

Die exakte Verteilung eines multiplen Kontrasttests bestehend aus diesen Testgrößen läßt sich über den Abbildungssatz aus der exakten Verteilung der Ränge bestimmen.

Diese läßt sich unter Verwendung kombinatorischer Argumenteherleiten (siehe z. B. Hajek, Sidak 1967). Da die Bestimmung der exakten Verteilung rechnerisch aufwendig ist, denn sie unterscheidet sich für jede Kombination von Fallzahlen, wird im Folgenden eine Approximation durch eine multivariate t-Verteilung mit der sich aus den Fallzahlen und den Kontrasten nach (3.28) ergebenden Korrelationsmatrix weiterverfolgt. Im Gegensatz zur Zwei-Stichproben Situation sind hier die Hypothesen der Parametrischen Tests mit denen dieses Rangtests nicht äquivalent. Währen die parametrischen Tests weiterhin die Gleichheit der Lageparameter testen, testet der Rangtest auf Gleichheit der Verteilungsfunktionen.
Diese einfache Rangtransformation hat den Vorteil, daß sie einfach durchzuführen ist und bei einfaktoriellen Designs in Fällen von nicht normalen Daten eine höhere Power aufweist als der analoge „Standard-Kontrasttest“ (Conover, Iman 1981), während bei höherfaktoriellen oder komplexeren Designs dieses Verfahren nicht tauglich ist (Hettmansperger, McKean 1998). Ist man hingegen an Hypothesen bezüglich der Verteilungsfunktionen der einzelnen Gruppen interessiert, so haben Akritas et al (1997) gezeigt, daß dieses Verfahren angemessen ist.
Neben dem „total rankin“ , d.h. der Rangbildung über die gesamte Stichprobe könnte man auch ein paarweises Rangbilden in Betracht ziehen. Bei dieser Form der Rangbildung werden die gemeinsamen Ränge über jeweils nur zwei Gruppen gebildet. Bei dieser Vorgehensweise sind die Rangmittel der Gruppen nicht eindeutig, so daß sich eine Teststatistik für einen Kontrast nicht eindeutig definieren läßt.
Anstelle die Ränge über die gesamte Stichprobe zu bilden (total-ranking) Eine andere Form über Ränge einen multiplen Test zu definieren wählen Hettmansperger und McKean (1998). Dafür schreiben sie das Ein-Faktor-Modell der Varianzanlyse wie in (3.29) als:

$$Y = C\theta + U. \tag{3.35}$$

Setzt man $a = \mathbf{E}[U_i]$ und $1_N := (1, \ldots, 1)^T$, dem Vektor aus $N = \sum n_i$ Einsen, so wird (3.35) zu:

$$Y = 1_N a + C\theta + (U - 1_N a) =: 1_N a + C\theta + U^*$$

mit $\mathbf{E}[U^*] = 0_N$. Setzen wir nun noch Ω als den von den Spalten von C aufgespannten Raum, so läßt sich das Modell (3.35) in der Form

$$Y = 1_N a + \eta + U^*, \qquad \eta \in \Omega \tag{3.36}$$

schreiben. Bezüglich des Modells (3.36) wird η durch den Punkt geschätzt, der einen geeigneten Abstand zwischen Y und Ω minimiert. Wählt man den euklidischen Abstand, so erhält man die normalen Kleinste-Quadrate- Schätzer $\theta_1 = \bar{X}_{i.}$. Die Idee

von Hettmansperger und McKean ist es nun, einen Abstand zu wählen, der geringer sensitiv auf Ausreißer reagiert. Ein solcher Abstand ist der, der durch die Pseudonorm

$$||v||_W := \sum_{i=1}^{N} \sqrt{12} \left(\frac{R(v_i)}{N+1} - \frac{1}{2} \right) v_i$$

definiert wird (vgl. Hettmansperger, McKean 1998). Hierbei steht $R(v_i)$ für den Rang, den der Wert v_i in dem Vektor v hat. Mit der so definierten Rang-Pseudonorm definieren Hettmansperger und McKean den R-Schätzer für η als die Lösung von

$$D_W(Y, \Omega) = ||Y - \hat{\eta}||_W = \min_{\eta \in \Omega} ||Y - \eta||.$$

Aus dem so geschätzten η erhält man den Schätzer für θ als $\hat{\theta} = (C^T C)^{-1} C^T \hat{\eta}$. Schätzt man nun noch a durch $\hat{a} = \text{median}\{Y_i - (C\hat{\theta})_i\}$, so haben Hettmansperger und McKean gezeigt, daß $\hat{a}$ und $\hat{\theta}$ asymptotisch gemeinsam multivariat normalverteilt, unkorreliert und konsistente Schätzer für a und θ sind. Hier gilt es auch wieder im Rahmen der Simulationen zu untersuchen, inwieweit sich im Endlichen multiple Kontraste mit diesen Schätzern durch eine multivariate t-Verteilung approximieren lassen. Da alle Daten mit in die Berechnungen eingehen, wird wie bei der m-Schätzung die aus den Fallzahlen n_i nach 3.28 berechnete Korrelationsmatrix verwendet.

3.3.3 Multiple Bootstrap-Verfahren

Eine ausführliche Darstellung der auf Bootstrap basierenden multiplen Testmethoden ist in Westfall und Young (1993) zu finden. Ein Großteil der dort beschriebenen Verfahren ist in der SAS Prozedur PROC MULTTEST implementiert. Daher wird hier zum Vergleich der unterschiedlichen robusten Methoden die PROC MULTTEST in den Simulationen verwendet. Die Vorgehensweise beim in der PROC MULTTEST verwendeten Bootstrap ist deshalb in folgenden beschrieben.
Das implementierte Verfahren berechnet aus den Rohdaten in einem ersten Schritt die Mittelwerte jeder Gruppe und die gepoolte Varianz. Für jeden interessierenden single-Kontrast wird dann die zugehörige t-Statistik nach (1.2) und deren p-Wert berechnet. Da das Bootstrappen unter der globalen Nullhypothese erfolgen sollte (Effron, Tibshirani 1993, Westfall, Young 1993), werden die Daten jeder Gruppe mittelwertzentriert. Würde dies nicht getan, so würden im Fall der Alternative die Bootstrapdaten aus einer multimodalen Verteilung gezogen. Von der Verteilung dieser Daten ist nun nicht anzunehmen, daß sie eine Normalverteilung ist oder die Situation unter der Nullhypothesen widerspiegelt (Westfall 1999, persönliche Mitteilung). Aus diesen Daten werden dann die Bootstrap-Stichproben mit Zurücklegen gezogen. Für jede Boostrap-Stichprobe werden die Gruppenmittelwerte, die gepoolte Varianz, die t-Statistiken und

die p-Werte für die einzelnen Kontraste berechnet. Als Testgröße wird nun der kleinste dieser Bootstrap p-Werte verwendet.
In einem Zähler für jeden single-Kontrast wird gezählt, wie oft der minimale Bootstrap p-Wert kleiner oder gleich dem p-Wert aus den Originaldaten ist. Der adjustierte p-Wert für jeden single-Kontrast ergibt sich dann als der Quotient aus dem Zähler für diesen Kontrast und der Anzahl der Bootstrap-Wiederholungen. Eine globale Ablehnung liegt genau dann vor, wenn der kleinste der adjustieren p-Werte kleiner als das vorgegebene Niveau ist.

Die Verfahrensweise des Bootstrappens läßt sich auch wieder verwenden, um k-Stichproben-Minimum-Tests analog denen aus Abschnitt 3.2.2 zu konstruieren. Wie im Zwei-Stichproben-Fall werden wieder verschiedene robuste Tests und der auf Mittelwerten und gepoolter Standardabweichung basierende Test zu einem Minimum-Test kombiniert. Dafür setze man im folgenden den p-Wert des multiplen Kontrasttests mit Mittelwert und Standardabweichung als p_t, den aus 20% getrimmtem Mittel und winsorisierter Varianz als p_{t20} und den aus dem m-Schätzer mit $c = 1,8$ als $p_{c_{1,8}}$. Aufgrund der Ergebnisse im Zwei-Stichproben-Fall werden die folgenden Minima betrachtet:

1. Das Minimum der p-Werte aus dem Test mit 20% getrimmten Mitteln und dem „normalen“ Test
$$t_{\min_t} := \min\{p_t, p_{t20}\}.$$
2. Das Minimum der p-Werte aus dem Test mit m-Schätzern ($c = 1,8$) und dem „normalen“ Test
$$t_{\min_c} := \min\{p_t, p_{c_{1,8}}\}.$$
3. Das Minimum der p-Werte aus dem Test mit 20% getrimmten Mitteln, m-Schätzern ($c = 1,8$) und dem „normalen“ Test
$$t_{\min_{tc}} := \min\{p_t, p_{t20}, p_{c_{1,8}}\}.$$

3.3.4 Adjustierte Zwei-Stichproben p-Werte

Neben den komplexen Verfahren aus den vorherigen Abschnitten, in denen die α-Adjustierung über die Verwendung der multivariaten t-Verteilung erfolgte, lassen sich auch mehre Zwei-Stichproben-Tests über eine „direkte“ α-Adjustierung zu einem multiplen Test verknüpfen. Die einfachste Methode ist die Adjustierung nach Bonferroni. Seien l Tests durchgeführt worden und die zugehörigen p-Werten mit $p_1, \ldots, p_l$ bezeichnet. Die Bonferroni-adjustierten p-Werte sind dann
$$p_i^{bon} := \min\{1, l \cdot p_i\}, \quad i = 1, \ldots, l.$$

Diese Vorgehensweise hat eine sehr geringe Güte. Diese läßt sich ohne Einschränkungen verbessern, indem man die p-Werte der Größe nach ordnet $p_{(1)} \leq \ldots \leq p_{(l)}$. Nach Holm (1979) ergeben sich die adjustierten p-Werte zu

$$\begin{aligned} p_{(1)}^{holm} &:= l \cdot p_{(1)} \\ p_{(i)}^{holm} &:= \min\{1, \max\{p_{(i-1)}^{holm}, (l-i+1) \cdot p_{(i)}\}\}, \quad i = 2, \ldots, l. \end{aligned}$$

Eine weitere Verbesserung der Güte läßt sich durch die Nutzung logischer Abhängigkeiten der Hypothesen untereinander erzielen. Auf diese Form der p-Wert-Adjustierung wird ausführlich in Kapitel 4 eingegangen.

3.3.5 Simulationsstudien

Am Beispiel des All-paar-Vergleichs (Tukey-Test) und des einseitigen Vergleichs mit einer Kontrolle (Dunnett-Test) wurden im Rahmen der durchgeführten Simulationsstudie Niveau- und Güte-Verhalten der robusten multiplen Tests untersucht. Verglichen werden der „standard“ Tukey-Test (Dunnett-Test) mit den entsprechenden multiplen Kontrasttests aus Trimmschätzern und Huber-m-Schätzern. Weiterhin sind die beiden Bootstap-Verfahren aus Abschnitt 3.3.3 und die einfache Rangtransformation in den Vergleich mit einbezogen. Für die Untersuchung des Güteverhaltens wurden drei verschiedene Erwartungswertprofile betrachtet, ein konvexes $0 = \mu_1 = \mu_2 < \mu_3$, ein lineares $0 = \mu_1 < \mu_2 = \frac{1}{2}\mu_3 < \mu_3$ und ein konkaves $0 = \mu_1 < \mu_2 = \mu_3$. Der dritte und höchste Erwartungswert wurde in Abhängigkeit von der zugrundeliegenden Verteilung und Fallzahl entsprechend Tabelle 3.2 gewählt. Die Anzahl der Wiederholungen betrug bei den Bootstrap-Verfahren 1000 Bootstrap-Wiederholungen innerhalb von 1000 Gesamtwiederholungen, bei den anderen Verfahren wurde 10000 mal wiederholt. In den folgenden Tabellen ist jeweils die multiple Power angegeben, da sich in den durchgeführten Simulationen die Resultate für die anderen Powerdefinitionen bezüglich des Vergleichs der Verfahren nicht unterscheiden. Liegt das simulierte Niveau oberhalb des 99%-igen Konfidenzintervalls, so ist der Wert fett geduckt. Die zu liberalen Tests bzw. Situationen gehörenden Powerwerte sind kursiv hervorgehoben.
Der Großteil der Simulationen wurden in SAS unter Verwendung des Programmiermoduls IML durchgeführt. Hierbei wurden die verwendeten Zufallszahlen mit den Standard-SAS-Fuktionen (RANNOR, RANCAU, RANBIN) erzeugt. Da für das Rangverfahren nach Hettmansperger ein Programm zur Verfügung stand, wurden die Simulationen hierfür in C implementiert. Hier wurde der Zufallszahlengenerator `ran2` aus Press et al (1995) verwendet.
Im Bezug auf die folgenden Untersuchungsergebnisse ist generell zu bedenken, daß es

keinen gleichmäßig besten Test über alle denkbaren Ausreißersituationen geben wird. Daher wird zum einen untersucht in wie weit die robusten Test den Standardtests unter Normalverteilung unterlegen sind. Zum anderen werden die Test unter den verschiedenen „Ausreißerverteilungen“ betrachtet, um zu ermitteln in welchen Situationen welche Test vorteilhafter sind.

Vergleich unter Normalverteilung

Aus den Tabellen 3.9 und 3.10 ist zu ersehen, daß alle Verfahren im Rahmen der Simulationsgenauigkeit das Niveau einhalten. Demnach erfüllen alle Verfahren die Grundanforderung der Niveautreue. Ein einziger liberaler Wert tritt hier beim $\min_t$-Test auf. Das Auftreten eines solchen Wertes liegt im Rahmen der zu erwartenden Niveauverletzungen bei der Anzahl der hier simulierten Niveaus. Beim Vergleich der Güten (Tabellen 3.11 und 3.12) zeigt sich eine erste Differenzierung der Verfahren. Die beiden Standardtests weisen, wie zu erwarten ist, die höchste Power auf. Jedoch sind sowohl die Tests basierend auf dem m-Schätzer als auch der Bootstrap-Test nach Westfall nur marginal schlechter. Dahingegen fallen die auf den Trimmschätzern basierenden Tests mit zwischen 8 und 10 Prozentpunkten deutlich zurück. Die drei verschiedenen Minimum-Tests weisen alle eine geringfügig (1-3%) niedrigere Power als der Standardtest auf. Beim Rangtest nach Hettmansperger fällt auf, daß er bei den hier verwendeten Fallzahlen sehr konservativ ist, was sich in der Güte mit einem Verlust von 4 bis 10 Prozentpunkten gegenüber dem Standardtest niederschlägt. Insbesondere zeigt er eine geringere Güte als der Rangtest nach Brunner, obwohl der Brunner-Test eine allgemeinere Hypothese testet (siehe auch Abschnitt 3.3.2). Daher wäre zu erwarten, daß der Brunner-Test eine geringere Güte aufweist.

All-paar	$n_1 = n_2 = n_3 = 10$			$n_1 = n_2 = n_3 = 20$		
	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$
tukey	1,0	4,9	9,5	1,1	5,0	10,0
20% trim	1,2	4,8	10,3	1,2	4,9	9,8
huber c=1,8	1,1	5,4	10,3	0,9	5,0	10,5
$\min_t$	1,0	5,0	10,4	0,8	4,7	9,8
$\min_c$	0,9	5,2	10,3	0,9	4,6	10,0
$\min_{tc}$	1,0	5,0	10,4	0,8	4,7	9,8
westfall	0,9	4,8	9,7	0,8	4,9	10,1
Brunner	1,2	5,2	10,2	1,1	5,2	10,0
Hettmansperger	0,9	3,8	7,7	1,0	4,4	8,7

Tabelle 3.9: Simulierte Niveaus unter Normalverteilung, All-paar

Vergl. mit Kontrolle	$n_1 = n_2 = n_3 = 10$			$n_1 = n_2 = n_3 = 20$		
	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$
dunnett	0,9	4,8	9,9	1,0	4,7	9,6
20% trim	1,1	4,9	9,8	1,0	5,0	9,7
huber c=1,8	1,1	4,9	10,6	1,0	5,0	10,6
$\min_t$	1,0	**5,9**	11,0	1,7	5,3	10,5
$\min_c$	1,0	5,5	10,5	1,4	5,6	9,7
$\min_{tc}$	1,1	5,6	11,2	1,7	5,2	10,3
westfall	1,0	4,9	10,0	0,5	4,9	9,9
Brunner	1,2	5,2	10,2	1,1	5,2	9,8
Hettmansperger	0,8	3,8	6,9	0,7	3,8	7,3

Tabelle 3.10: Simulierte Niveaus unter Normalverteilung, Vergleich mit Kontrolle

All-paar	$n_1 = n_2 = n_3 = 10$			$n_1 = n_2 = n_3 = 20$		
$\alpha = 5\%$	konvex	linear	konkav	konvex	linear	konkav
tukey	72,1	59,8	72,6	72,3	60,1	72,2
20% trim	59,3	48,2	59,3	64,7	51,8	64,2
huber c=1,8	71,3	58,7	71,3	72,6	58,7	72,3
$\min_t$	67,7	55,3	67,3	69,8	57,4	69,4
$\min_c$	71,4	57,0	74,1	71,7	59,4	70,8
$\min_{tc}$	67,8	55,0	67,4	69,4	57,4	69,4
westfall	72,1	60,6	72,3	72,3	60,1	72,8
Brunner	70,6	58,5	70,6	70,8	57,5	70,6
Hettmansperger	63,9	53,1	63,8	66,9	56,4	68,0

Tabelle 3.11: Simulierte Power unter Normalverteilung, All-paar

Verg. mit Kontrolle	$n_1 = n_2 = n_3 = 10$			$n_1 = n_2 = n_3 = 20$		
$\alpha = 5\%$	konvex	linear	konkav	konvex	linear	konkav
dunnett	73,7	75,7	86,6	73,1	74,9	85,9
20% trim	64,4	66,4	78,4	66,9	68,1	80,9
huber c=1,8	73,1	74,4	85,9	72,2	73,4	85,9
$\min_t$	*70,6*	*71,4*	*83,0*	70,4	71,5	85,0
$\min_c$	71,7	72,7	83,9	71,2	73,2	86,2
$\min_{tc}$	70,2	71,8	82,7	70,6	71,1	84,6
westfall	73,8	75,2	86,5	74,3	75,7	85,7
Brunner	71,7	74,0	84,5	70,4	73,7	84,7
Hettmansperger	61,5	62,5	75,1	63,9	65,9	77,3

Tabelle 3.12: Simulierte Power unter Normalverteilung, Vergleich mit Kontrolle

Vergleich unter Ausreißerverteilungen

Nach den Betrachtungen und Ergebnissen im Zwei-Stichproben-Fall werden hier nur noch die logarithmische Normalverteilung, die Cauchy-Verteilung, die $\mathcal{X}_2^2$-Verteilung und die kontaminierte Normalverteilung als nichtnormale Verteilungen herangezogen.

Aus den Tabellen 3.13 und 3.14 ist zu ersehen, daß alle Verfahren das 5%-Niveau einhalten. Die beiden Niveauüberschreitungen bei den Minimum-Tests sind nicht als systematische Fehler zu sehen. Zum einen ist das Bootstrappen unter einer $\mathcal{X}_2^2$ Verteilung der Konstruktion nach ein niveautreues Verfahren. Zum anderen existiert bei der Cauchyverteilung kein einziges Moment (nicht einmal der Erwartungswert), wodurch die klassische asymptotische Theorie keine Aussagen bezüglich des Verhaltens von Teststatistiken basierend auf solchen Daten ermöglicht. Es fällt auf, daß die Standardtests unter Cauchy-verteilten Daten beide extrem konservativ sind. Das gleiche Verhalten zeigen auch der Rangtest nach Hettmansperger und das Bootstrap-Verfahren. Demgegenüber weist das zweite Rangverfahren (Brunner) eine gleichmäßige Ausschöpfung des Niveaus auf. Dieses Ergebnis verwundert nicht, da der Brunner-Test als Hypothese die Gleichheit der verschiedenen Verteilungen testet, was in allen Fällen der Fall ist. Wohingegen der Rangtest nach Hettmansperger, wie auch die restlichen Test die Hypothese der Gleichheit der Lageparameter testet. Generell fällt auf, daß die Standardverfahren als auch das Bottstrap-Verfahen bei Ausreißern in beide Richtungen (zu große und zu keine Werte), insbesondere bei kleinen Fallzahlen, dazu tendieren konservativ zu sein, was ein Folge des stark vergrößerten Varianzschätzers ist.

Bei der Betrachtung der Power ist zuerst zu bemerken, daß das Erwartungswertprofil

All-paar	$n_1 = n_2 = n_3 = 10$				$n_1 = n_2 = n_3 = 20$			
$\alpha = 5\%$	LogNV	CAU	$\mathcal{X}_2^2$	KNV	LogNV	CAU	$\mathcal{X}_2^2$	KNV
tukey	3,5	1,7	4,3	2,4	3,5	1,7	4,5	4,0
20% trim	3,2	2,6	3,9	3,1	3,8	3,6	4,4	3,9
huber c=1,8	3,9	4,6	4,9	3,8	3,4	5,5	4,5	4,4
$\min_t$	5,2	4,6	4,2	4,9	4,7	4,0	5,7	5,6
$\min_c$	5,2	5,4	4,1	4,4	4,6	3,8	**5,9**	5,2
$\min_{tc}$	5,2	5,0	4,1	4,9	4,7	4,6	5,6	5,5
westfall	3,7	2,1	4,3	3,1	4,3	2,5	4,8	4,4
Brunner	5,2	5,2	5,4	4,8	4,8	5,2	4,9	5,2
Hettmansperger	4,2	2,2	4,6	2,9	3,9	2,5	4,7	2,2

Tabelle 3.13: Simulierte Niveaus unter Nicht-Normalverteilung, All-paar

keinen Einfluß auf die relative Ordnung der verschiedenen Tests zueinander hat. Auch zeigen sich zwischen einseitigem (Dunnett) und zweiseitigem Testen die Ergebnisse bezüglich der Über- oder Unterlegenheit konsistent.

Wie schon im Zwei-Stichproben-Fall zeigen die auf Mittelwert und Standardabweichung basierenden Tests einen Powerverlust gegenüber den robusten Tests. Auffallend ist, daß die Bootstrap-Methode, die in der Prozedur PROC MULTTEST implementiert ist, eine nur geringfügig höhere Power zeigt als der jeweilige Standardtest. Dies liegt in der

Verg. mit Kontrolle	$n_1 = n_2 = n_3 = 10$				$n_1 = n_2 = n_3 = 20$			
$\alpha = 5\%$	LogNV	CAU	$\mathcal{X}_2^2$	KNV	LogNV	CAU	$\mathcal{X}_2^2$	KNV
dunnett	4,9	2,9	5,3	4,1	4,8	2,9	5,2	5,0
20% trim	4,2	3,4	4,7	3,6	4,6	4,2	4,9	4,4
huber c=1,8	4,8	5,2	5,2	4,4	4,3	5,0	5,1	4,4
$\min_t$	4,6	3,3	5,2	4,6	4,3	**5,9**	3,6	5,4
$\min_c$	4,8	3,9	5,1	4,5	4,2	5,7	3,7	5,2
$\min_{tc}$	4,4	3,9	5,1	4,6	4,6	5,7	3,8	5,1
westfall	4,1	3,1	4,9	5,0	4,0	2,6	4,5	5,6
Brunner	4,8	5,1	5,1	5,0	5,1	5,1	5,1	4,9
Hettmansperger	4,4	2,3	3,7	2,4	3,3	2,2	3,0	2,2

Tabelle 3.14: Simulierte Niveaus unter Nicht-Normalverteilung, Vergleich mit Kontrolle

Transformation der Teststatistik in den p-Wert über eine t-Verteilung begründet und ist von daher zu erwarten (Westfall 1999, persönliche Mitteilung).
Im Gegensatz zur Zwei-Stichproben-Situation ist bei kleinen Fallzahlen ($n_i = 10$) der Huber-m-Test immer besser als der trimm-Test. Außerdem ist er in allen Situationen unter den Test mit der höchste Güte. Erst bei größeren Fallzahlen hängt es von der Schiefe ab, ob der Trimm- oder der Huber-m-Schätzer-Test die höhere Güte aufweist. Beim Rang-Test nach Hettmansperger fällt auf, daß seine Güte bei kleinen Fallzahlen noch stark von der Schiefe der zugrundeliegenden Verteilung abhängt. So ist er bei symmetrischen Verteilungen z.B. der kontaminierten Normalverteilung mit einer Power von 49,7% (konvex) der schlechteste der robusten Tests, hingegen bei schiefen Verteilungen unter den Besten Tests z.B. $\mathcal{X}_2^2$-Verteilung mit einer Power von 89,3% (konvex). Mit wachsender Fallzahl wird diese Unterlegenheit jedoch geringer (siehe KNV, CAU). Der Brunner-Rang-Test zeigt ein gleichmäßig stabiles Güteverhalten, hat jedoch den Nachteil, keine Parameterschätzer oder Konfidenzintervalle bereitzustellen. Weiterhin zeigt sich das der parametrische Huber-m-Test gleichmäßig besser als der Brunner-Test ist, was wieder in der Verschiedenheit der getesteten Hypothesen begründet ist. Außerdem ist er dem jeweils besten Test um ≈ 4 Prozentpunkte unterlegen. Bei dem Vergleich der Power unter Cauchy-Verteilung für die beiden Fallzahlen ($n_1 = 10, n_i = 20$) fällt auf, das die Power für die Standardtest, den Huber-m-Test, den daraus gebildeten Minimum und dem Bootstrap-Test bei der höheren Fallzahl geringer sind. Diese Tests zeichnen sich gegenüber den restlichen dadurch aus, daß sie über eine „empfindlichere“ Schätzung der Streuung verfügen. Aufgrund der Tatschache, daß die Cauchy-Verteilung sehr flach ist, steigt mit der Anzahl der Werte die Anzahl der sehr weit außen liegenden Werte stärker als bei den restlichen Verteilungen. Die führt zu einem größeren Schätzer für die Streuung und dadurch zu einer selteneren Ablehnung der Nullhypothese.
Die Minimum-Tests erweisen sich, wie schon im Zwei-Stichproben-Fall, als „Mittler“

zwischen ihren Einzeltests, sind jedoch nur in der Lage, global zu entscheiden, ob überhaupt irgendein Unterschied besteht.

3.3.6 Zusammenfassende Anwendungsempfehlung

Generell ist aus den Simulationsergebnissen zu sehen, daß Tests basierend auf Mittelwert und Standardabweichung sehr empfindlich auf Abweichungen von der Normalverteilung reagieren. Daher lassen sich folgend Empfehlungen zur Robustifizierung von Parameterschätzungen und Testentscheidungen geben:

Auswertung eines Zwei-Stichproben-Experiments Bei Fallzahlen bis zu 20 Beobachtungen pro Gruppe sollten die Tests von Tiku und der Huber-m-Test mit einer aus den Einzelvarianzen gepoolten Varianz nicht verwendet werden, da beide Tests im allgemeinen sehr liberal werden können. Ist man an der Schätzung der Parameter und der Bestimmung von Konfidenzintervallen interessiert, so ist die Verwendung des Huber-m-Tests anzuraten. Insbesondere weil der Huber-m-Test unter Normalverteilung nur marginal schlechter als der t-Test ist, ist er immer eine gute Alternative zum t-Test. Ist die Bestimmung der Parameter und der p-Wert für den Test auf Unterschied das gewünschte Ergebnis, so bietet sich der Minimum-Test aus t-, 20% trim- und Huber-m-Statistik an. Im Kapitel Software ist unter Abschnitt 7.1 das Listing eines SAS-Macros angegeben, das den getrimmten-t-Test, den huber-m-Test und den minimum-Test basieren auf t-Test, trim-Test und m-Test durchführt.

Multiple Kontraste Im Fall von k-Stichproben empfiehlt es sich bei kleinen Fallzahlen, multiple Kontraste mit den Huber-m-Schätzern zu testen, wohingegen der Rang-Test nach Hettmansperger nicht bei kleinen Fallzahlen verwendet werden sollte. Bei größeren Fallzahlen hängt es wiederum von der Schiefe der zugrundeliegenden Verteilung ab, ob die Verwendung von trim-Schätzern oder Huber-m-Schätzern vorteilhafter ist. Weiterhin gewinnt hierbei der Rang-Test nach Hettmansperger, der auch eine Parameterschätzung und die Bestimmung von Konfidenzintervallen ermöglicht, an Power, so daß er bei Fallzahlen von mehr als 25 Beobachtungen pro Gruppe dem Brunner-Rang-Verfahren vorzuziehen ist. Obwohl die Minimum-Tests wieder gut zwischen t-Test und den robusten Tests vermitteln, haben sie den Nachteil, daß sie nur eine Globalentscheidung treffen und eine Parameterschätzung passend zum p-Wert nicht ermöglichen. Bei mittleren Fallzahlen (ca. $15 < n_i < 25$) sollte man sich je nach erwarteter Schiefe zwischen dem Trimmen und dem Huber-Verfahen entscheiden. Ist zu erwarten, daß die Verteilung symmetrisch ist so sollte getrimmt werden, ansonsten empfiehlt sich das

huber-Verfahren. Sind die Fallzahlen größer($n_i > 25$), so ist der Test nach Hettmansperger eine sehr robuste und schiefeunabhängige Alternative zu diesen beiden Verfahren. Problematisch können extremen Situationen der Art, daß viele Gruppen ($k \geq 10$) mit wenig Fallzahl ($n_i = 5, \ldots, 10$) untersucht werden sollen sein. Im Rahmen von Simulationsstudien hat Ringland (1983) gezeigt, daß das Huber-Verfahen nicht niveautreu ist. Bei den Untersuchungen wurden jedoch als Startwerte für die iterative Lösung der Gleichungen (3.31) und (3.32) Mittelwert und Varianz verwendet, welche nicht der Forderung nach „robusten“ Startwerten entsprechen.
Der Code eines SAS-Macros zur Berechnung der p-Werte eines multiplen Kontrasts basierend auf trim- oder huber-m-Schätzern ist im Kapitel Software im Abschnitt 7.2 angegeben.

LogNV	All-paar			Vergl. mit Kontrolle		
	konvex	linear	konkav	konvex	linear	konkav
tukey/dunnett	67,7	59,5	68,3	69,3	72,2	79,7
20% trim	84,3	77,5	84,1	86,4	87,5	92,2
huber c=1,8	96,0	90,3	94,0	94,8	95,3	97,3
$\min_t$	85,8	77,7	85,0	86,1	87,6	92,2
$\min_c$	94,3	88,3	93,0	93,4	94,2	97,0
$\min_{tc}$	92,8	85,9	91,5	93,0	93,3	96,3
westfall	72,0	62,9	72,7	68,2	75,0	80,1
Brunner	93,9	89,3	86,9	92,9	94,5	93,6
Hettmansperger	97,9	94,4	96,3	96,3	96,7	98,2
CAU	All-paar			Vergl. mit Kontrolle		
	konvex	linear	konkav	konvex	linear	konkav
tukey/dunnett	17,0	14,4	17,8	20,8	23,5	29,5
20% trim	66,2	56,9	66,2	69,7	71,7	80,3
huber c=1,8	69,5	58,3	70,9	71,0	72,8	83,3
$\min_t$	69,6	61,0	68,2	68,0	70,6	81,8
$\min_c$	67,0	56,8	63,1	65,8	68,9	80,9
$\min_{tc}$	71,2	61,9	70,6	40,0	73,7	83,8
westfall	19,8	19,0	22,6	22,1	23,7	30,6
Brunner	60,9	56,0	61,3	63,5	70,5	76,9
Hettmansperger	49,6	39,2	48,8	59,3	59,4	70,5
$\mathcal{X}_2^2$	All-paar			Vergl. mit Kontrolle		
	konvex	linear	konkav	konvex	linear	konkav
tukey/dunnett	70,7	60,1	70,5	70,9	74,4	83,2
20% trim	69,8	58,2	69,1	73,9	75,1	83,8
huber c=1,8	87,1	76,5	85,5	86,4	86,8	92,7
$\min_t$	75,0	68,2	73,3	75,3	76,4	86,8
$\min_c$	83,8	76,4	82,6	82,9	84,1	92,1
$\min_{tc}$	81,2	73,7	79,2	81,2	81,3	90,4
westfall	72,5	61,3	71,5	69,7	73,6	82,6
Brunner	85,9	76,3	79,8	84,8	87,1	89,6
Hettmansperger	89,3	81,0	87,2	86,4	87,7	92,6
KNV	All-paar			Vergl. mit Kontrolle		
	konvex	linear	konkav	konvex	linear	konkav
tukey/dunnett	23,1	18,9	26,8	26,0	30,1	38,6
20% trim	73,2	63,2	73,3	76,9	78,0	84,6
huber c=1,8	77,6	68,3	77,3	78,2	80,6	88,0
$\min_t$	75,1	66,0	74,5	74,8	77,3	83,9
$\min_c$	74,6	65,7	74,0	74,2	75,4	84,1
$\min_{tc}$	78,9	69,8	78,2	78,9	80,1	87,1
westfall	26,5	20,7	27,9	27,2	30,6	36,3
Brunner	71,1	64,7	70,4	72,9	77,9	84,2
Hettmansperger	49,7	42,9	51,0	52,3	53,3	60,9

Tabelle 3.15: Simulierte multiple Power unter Nicht-Normalverteilung, $k = 3$, $n_i = 10$, $\alpha = 5\%$

LogNV	All-paar			Vergl. mit Kontrolle		
	konvex	linear	konkav	konvex	linear	konkav
tukey/dunnett	70,4	59,9	69,9	70,6	72,3	81,0
20% trim	94,9	87,9	92,8	93,9	93,5	96,9
huber c=1,8	98,8	95,8	97,8	97,9	98,3	99,2
$\min_t$	93,4	87,3	91,8	93,0	92,8	95,1
$\min_c$	97,9	94,3	96,5	97,8	96,9	99,1
$\min_{tc}$	97,7	93,1	96,2	97,2	95,9	98,7
westfall	72,7	63,1	73,4	70,0	72,1	82,0
Brunner	99,1	97,4	96,6	98,6	98,6	98,8
Hettmansperger	99,8	99,0	99,3	99,4	99,3	99,7
CAU	All-paar			Vergl. mit Kontrolle		
	konvex	linear	konkav	konvex	linear	konkav
tukey/dunnett	8,8	6,6	8,1	11,7	13,2	16,9
20% trim	68,9	58,3	69,5	70,4	72,9	82,9
huber c=1,8	63,5	51,9	63,1	64,2	66,7	77,9
$\min_t$	70,8	55,3	67,9	66,5	68,2	78,4
$\min_c$	62,4	46,0	58,7	58,1	61,3	68,7
$\min_{tc}$	68,9	54,4	66,7	65,0	68,2	77,4
westfall	11,4	9,7	11,6	12,9	15,0	19,8
Brunner	64,3	54,1	64,8	66,0	69,5	79,1
Hettmansperger	66,0	54,8	66,3	58,8	60,4	72,5
$\mathcal{X}_2^2$	All-paar			Vergl. mit Kontrolle		
	konvex	linear	konkav	konvex	linear	konkav
tukey/dunnett	70,6	58,8	69,9	70,9	73,0	83,7
20% trim	76,1	64,1	75,3	76,6	77,9	87,0
huber c=1,8	89,7	79,6	87,9	87,9	88,8	94,4
$\min_t$	77,2	67,1	77,4	76,2	77,5	86,4
$\min_c$	87,4	74,2	85,8	86,3	85,1	92,6
$\min_{tc}$	85,5	73,1	84,3	83,5	83,9	91,4
westfall	72,1	60,1	70,6	71,5	72,0	83,2
Brunner	92,6	84,4	87,1	91,0	91,7	93,8
Hettmansperger	95,3	88,6	92,8	91,3	92,3	96,0
KNV	All-paar			Vergl. mit Kontrolle		
	konvex	linear	konkav	konvex	linear	konkav
tukey/dunnett	21,4	17,0	22,4	24,4	27,5	36,6
20% trim	88,3	79,9	88,5	88,8	89,0	94,7
huber c=1,8	85,7	76,0	85,7	85,3	86,2	93,3
$\min_t$	86,7	77,9	85,2	86,3	84,9	92,2
$\min_c$	81,9	71,4	79,6	80,0	82,1	89,8
$\min_{tc}$	87,6	79,7	86,2	87,0	85,5	94,0
westfall	23,5	19,2	22,5	25,6	29,2	36,2
Brunner	83,9	76,8	84,7	84,7	87,3	93,4
Hettmansperger	74,6	64,7	73,7	73,4	75,5	83,1

Tabelle 3.16: Simulierte multiple Power unter Nicht-Normalverteilung, $k = 3$, $n_i = 20$, $\alpha = 5\%$

Kapitel 4

Bestimmung logischer Abhängigkeiten von paarweisen Hypothesen

Der folgende Abschnitt behandelt schwerpunktmäßig das multiple Testen paarweiser, einseitiger Nullhypothesen, d.h. es werden in jedem Einzeltest nur zwei Gruppen (z.B. Dosisstufen, Behandlungen) miteinander verglichen. Solch eine Situation liegt zum Beispiel bei einem Vergleich von k Behandlungen mit einer Kontrolle vor (μ_0 vs. μ_i, $i = 1, \ldots, k$), oder wenn in einem Dosis-Wirkungs-Versuch die einzelnen Dosisschritte interessieren (μ_i vs. μ_{i+1}). Wie bei zweiseitigen Fragestellungen führen auch bei einseitigen Fragestellungen logische Abhängigkeiten der Hypothesen zu einer Vereinfachung des aus ihnen gebildeten Abschlußtestsystems. Folgendes Beispiel soll die Bedeutung logischer Abhängigkeiten veranschaulichen.
In einem Dosis-Wirkungs-Versuch seien neben einer Kontrolle zwei verschieden hohe Dosen eines Mittels verwendet worden. In solch einem Design ist es für den Versuchsansteller häufig von Interessse, ob zum einen die Dosierungen des Mittels einen Effekt in eine vorgegebene Richtung haben (z.B. Ertragssteigerung), und ob zum anderen mit steigender Dosis der Effekt zunimmt. In Nullhypothesen ausgedrückt lautet die erste Frage $\mu_K \geq \mu_{D1}$ und $\mu_K \geq \mu_{D2}$ und die zweite $\mu_{D1} \geq \mu_{D2}$. In diesem kleinen Beispiel ist nun die Nullhypothese, ob die zweite Dosis besser ist als die Kontrolle, logisch abhängig von den beiden anderen. Sind nämlich die beiden anderen falsch, gilt also: $\mu_K < \mu_{D1}$ und $\mu_{D1} < \mu_{D2}$, so folgt auch, daß $\mu_K < \mu_{D2}$ ist. Dies bedeutet, daß die dritte Hypothese nicht mehr wahr sein kann.
Sollen in einem mehrere Gruppen umfassenden Versuch alle oder einige paarweise Vergleiche durchgeführt werden, wobei das multiple Niveau eingehalten werden soll, so lassen sich diese logischen Abhängigkeiten ausnutzen. Führt man den Abschlußtest mit Bonferroni Globaltests für jede einzelne auftretende Schnitthypothese durch, so

erhält man bessere (höhere) lokale α-Schranken als dies beim Holm-Verfahren der Fall wäre.
Zur Bestimmung der logischen Abhängigkeiten wird das Hypothesensystem in die Graphentheorie abgebildet. Anhand des sich ergebenden Graphen lassen sich die logischen Abhängigkeiten leicht und anschaulich bestimmen. Zudem werden Algorithmen angegeben, die die Bestimmung aller logischen Abhängigkeiten ermöglichen. Dafür werden im folgenden Abschnitt zuerst einige Begriffe aus dem Bereich des Hypothesentestens und aus der Graphentheorie eingeführt, sowie hilfreiche Sätze angegeben. In den darauffolgenden Abschnitten wird zuerst für Hypothesensysteme von zweiseitigen Tests eine Möglichkeit vorgestellt, das Shaffer-Verfahren (1986) durchzuführen, und danach beschrieben, wie sich die Hypothesen des Abschlußtests (AT) für einseitige All-paar- und Beliebige-Paar- Vergleiche konstruieren lassen. Unter Verwendung dieser Ergebnisse werden die Algorithmen von Bernhard (1991) erweitert, um den Abschlußtest computergestützt durchführen zu können. In einer Anwendung werden verschiedene Methoden, zur Bestimmung des höchsten effektiven Dosisschritts der des AT über Einzelvergleichen gegenübergestellt. Abschließend wird gezeigt wie sich auch Hypothesensysteme mit verschobenen einseitigen Nullhypothesen behandeln lassen.

4.1 Definitionen

In den folgenden zwei Abschnitten werden die im späteren verwendeten Begriffe aus der Theorie des Hypothesentestens und der Graphentheorie vorgestellt und erläutert. Die vorgestellten Begriffe und Sätze bilden eine Auswahl, welche für die speziellen, hier behandelten Fragestellungen benötigt wird.

4.1.1 Grundbegriffe des Hypothesentestens

Notation und Begriffsbildung sind an Bernhard (1991) angelehnt, insbesondere, da die hier erzielten Ergebnisse Erweiterungen und Anwendungen der Resultate aus dieser Arbeit sind.
Es seien $l > 1$ zu prüfende Elementarhypothesen (Nullhypothesen) H_i, $i = 1, \ldots, l$ gegeben. Einen beliebigen Schnitt dieser Elementarhypothesen nennt man *Schnitthypothese*. Zur Charakterisierung aller Schnitthypothesen verwendet man die Potenzmenge $\wp_l$ der Hypothesenindizes $\{1, \ldots, l\}$ beziehungsweise $\wp_l^0 := \wp_l \setminus \{\emptyset\}$. Die Schnitthypothese H_I ist dann erklärt durch

$$H_I := \bigcap_{i \in I} H_i.$$

Die Menge I wird hierbei als Indexmenge bezeichnet. Weiterhin wird die Menge $H_0 := H_{\{1,\ldots,l\}}$ Globalhypothese genannt. Das System aller Schnitthypothesen bezeichne man mit

$$\hbar := \{H_I : \ I \in \wp_l^0\}.$$

Ausgehend von einem System von Elementarhypothesen gelangt man zu einem durchschnittsabgeschlossenen Hypothesensytem $\hbar$, in dem noch alle Schnitthypothesen mit in das Hypothesensystem aufgenommen werden. Dabei kommt es im allgemeinen vor, daß verschieden indizierte Schnitthypothesen die gleiche Hypothese repräsentieren, d.h. $H_I = H_J$ für $I \neq J$. Solche Schnitthypothesen heißen *redundant*. Desweiteren ist gefordert, daß $H_0 \neq \emptyset$ ist. Da die in dieser Arbeit behandelten Problemstellungen nur paarweise Vergleiche als Elementarhypothesen beinhalten, ist diese Bedingung im folgenden immer erfüllt. Gilt für eine Schnitthypothese, daß sich die in Relation zueinander befindlichen Erwartungswerte in mindestens zwei disjunkte Gruppen trennen lassen, so wird eine solche Hypothese auch *Partitionshypothese* genannt. Sind zum Beispiel $H_1 : \ \mu_1 = \mu_2$, $H_2 : \ \mu_2 = \mu_3$ und $H_3 : \ \mu_3 = \mu_4$, so ist die Schnitthypothese $H_{\{1,3\}}$ eine Partitionshypothese.

Beispiel (All-paar-Vergleich) Seien $k = 3$ Gruppen/Dosen mit den $l = 3$ Elementarhypothesen $H_1 : \ \mu_1 = \mu_2$, $H_2 : \ \mu_2 = \mu_3$ und $H_3 : \mu_1 = \mu_3$ von Interesse. Für die Schnitthypothesen gilt

$$H_{\{1,2\}} = H_{\{1,3\}} = H_{\{2,3\}} = H_{\{1,2,3\}} : \ \mu_1 = \mu_2 = \mu_3.$$

Damit enthält $\hbar$ redundante Schnitthypothesen. Um das Hypothesensystem $\hbar$ vollständig zu beschreiben, würde es genügen, die Indexmengen $I = \{\{1\}, \{2\}, \{3\}, \{1, 2\}\}$ anzugeben.
Zur eindeutigen Charakterisierung eines Hypothesensystems durch eine Teilmenge von $\wp_l^0$ definiere man für Indexmengen die Eigenschaft, erschöpfend zu sein.

Definition 4.1 (Definition 1.3, Bernhard 1991)
Sei $I \in \wp_l^0$. Eine Menge I ist *erschöpfend*, falls $H_I \neq \emptyset$ und für alle $J \in \wp_l^0$ mit $H_I = H_J$ folgt $J \subseteq I$.
Weiterhin ist $\wp_{EI} := \{I \in \wp_k^0 : \ I \text{ erschöpfend}\}$ die Menge aller *erschöpfenden Indexmengen* und $EL := \{|I| : \ I \in \wp_{EI}\}$ die Menge der *erschöpfenden Schichten*.

In Bergmann (1987) wird gezeigt, daß es zu jeder Indexmenge I mit $H_I \neq \emptyset$ genau eine kleinste erschöpfende Indexmenge gibt, die I umfaßt und zur gleichen Schnitthypothese führt. Diese Indexmenge sei mit $SE(I)$ bezeichnet.

Satz 4.1 (Bergmann 1987)
Zu jeder Indexmenge $I \in \wp_l^0$ ist die kleinste, umfassende, erschöpfende Indexmenge $SE(I)$ eindeutig bestimmt.

Diese Definition von erschöpfend läßt sich noch modifizieren, wie in Hommel und Bernhard (1999) angegeben:

Definition 4.2 (Hommel, Bernhard 1999)
Sei $I \in \wp_l^0$. Eine Menge I ist *streng erschöpfend*, falls genau alle Hypothesen H_i mit $i \in I$ wahr sein können.
Weiterhin ist $\wp_{EI}^s := \{I \in \wp_l^0 : \ I \text{ streng erschöpfend}\}$ die Menge aller *im strengen Sinn erschöpfenden Indexmengen* und $EL^s := \{|I| : \ I \in \wp_{EI}^s\}$ die Menge der *streng erschöpfenden Schichten*.

Satz 4.2
Sei $I \in \wp_l^0$, so folgt für I aus der Eigenschaft, streng erschöpfend zu sein, daß auch I erschöpfend ist.

Beweis:
Sei $I \in \wp_l^0$ streng erschöpfend. Außerdem sei mit A_i die Alternative von H_i bezeichnet, bzw. $A_I = \bigcap_{i \in I} A_i$. Da I streng erschöpfend ist, gilt $H_I \neq \emptyset$ und $H_I \cap A_{\mathbb{N}_l \setminus I} \neq \emptyset$. Wäre I nicht erschöpfend, so gäbe es ein $i \in \mathbb{N}_l \setminus I$ mit $H_I = H_{I \cup i} = H_I \cap H_i$. Dann würde gelten

$$\emptyset \neq H_I \cap A_{\mathbb{N}_l \setminus I} = H_I \cap H_i \cap A_{\mathbb{N}_l \setminus \{I \cup i\}} \cap A_i = \emptyset. \quad \text{(W!)}$$

Also ist I erschöpfend.

$\bigtriangledown$

Daß die Eigenschaft, im strengen Sinn erschöpfend zu sein, eine schärfere Forderung ist, d.h. $\wp_{EI}^s \subseteq \wp_{EI}$, und die kleinste, umfassende, streng erschöpfende Indexmenge nicht eindeutig ist, zeigt folgendes Beispiel:

Beispiel (Dosisschritte und Vergleich zur Kontrolle) Seien $k = 3$ Gruppen/Dosen mit den $l = 3$ Elementarhypothesen $H_1 : \ \mu_1 \geq \mu_2$, $H_2 : \ \mu_2 \geq \mu_3$ und $H_3 : \mu_1 \geq \mu_3$ von Interesse. (Wie z.B. in der Einleitung formuliert.) Als Schnitthypothesen ergeben sich dann:

$$\begin{aligned} H_{\{1,2\}} = H_{\{1,2,3\}} : &\quad \mu_1 \geq \mu_2 \geq \mu_3 \\ H_{\{1,3\}} : &\quad \mu_1 \geq \mu_2 \ \wedge \ \mu_1 \geq \mu_3 \\ H_{\{2,3\}} : &\quad \mu_2 \geq \mu_3 \ \wedge \ \mu_1 \geq \mu_3. \end{aligned}$$

Die Menge der kleinsten, erschöpfenden Indexmengen ist hier

$$I = \{\{1\}, \{2\}, \{3\}, \{1,3\}, \{2,3\}, \{1,2,3\}\},$$

während mit der strengeren Forderung die Elementarhypothese H_3 nicht erschöpfend ist, da sie nicht wahr sein kann, wenn H_1 und H_2 falsch sind. Somit ist die Menge der streng erschöpfenden Indexmengen

$$I^s = \{\{1\}, \{2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}.$$

D.h., zur Indexmenge $\{3\}$ gibt es zwei kleinste, umfassende, streng erschöpfende Indexmengen.
Am obigen Beispiel zeigte sich, daß für Hypothesensysteme mit einseitigen Nullhypothesen die erschöpfenden und die streng erschöpfenden Indexmengen nicht zusammenfallen. Sind jedoch alle Hypothesen zweiseitige lineare Kontrasthypothesen, so ist jede erschöpfende Indexmenge auch streng erschöpfend.

Satz 4.3 (Hommel 1999)
Sind alle Elementarhypothesen H_i, $i = 1, \ldots, l$ zweiseitige lineare Kontrasthypothesen, so gilt:

1. $H_I \neq \emptyset$ für alle $I \subset \mathbb{N}_l$,

2. Jede erschöpfende Indexmenge ist auch streng erschöpfend.

Bezüglich der durch die Indexmengen $\wp_{EI}$ und $\wp^s_{EI}$ generierten Hypothesensysteme gilt, daß sie sich maximal durch die leere Menge unterscheiden.

Satz 4.4
Seien H_i, $i = 1, \ldots, l$ Elementarhypothesen gegeben, dann unterscheiden sich $\{H_I : I \in \wp_{EI}\}$ und $\{H_I : I \in \wp^s_{EI}\}$ maximal durch die leere Menge.

Beweis:
Nach Satz 4.2 ist $\wp^s_{EI} \subseteq \wp_{EI}$. Also ist auch $\hbar^s := \{H_I : I \in \wp^s_{EI}\} \subseteq \{H_I : I \in \wp_{EI}\} =: \hbar$. Sei nun $H_J \in \hbar \setminus \hbar^s$. D.h. $J \notin \wp^s_{EI}$ und alle in H_J enthaltenen Elementarhypothesen sind wahr. Könnten die restlichen $\mathbb{N}_l \setminus J$ Hypothesen alle falsch sein, so wäre $J \in \wp^s_{EI}$ (W!) oder $H_J = \emptyset$. Demnach müssen zusätzlich noch K Hypothesen mit $1 \leq |K| \leq l - |J|$ wahr sein. Damit ist $H_J = H_{J \cup K}$ und $J \cup K \in \wp^s_{EI}$.

∇

Ausgehend von einer Menge von Hypothesen $H_1, \ldots, H_l$ läßt sich nun anhand der streng erschöpfenden Indexmengen eine Testprozedur konstruieren, die dem Abschlußtest entspricht und somit das multiple Niveau einhält. Hierfür werden alle Elementarhypothesen H_j mit $j \notin A$ verworfen, wobei die Akzeptanzmenge

$$A := \bigcup \{I : I \text{ streng erschöpfend, } \min\{p_i : i \in I\} > \frac{\alpha}{|I|}\}$$

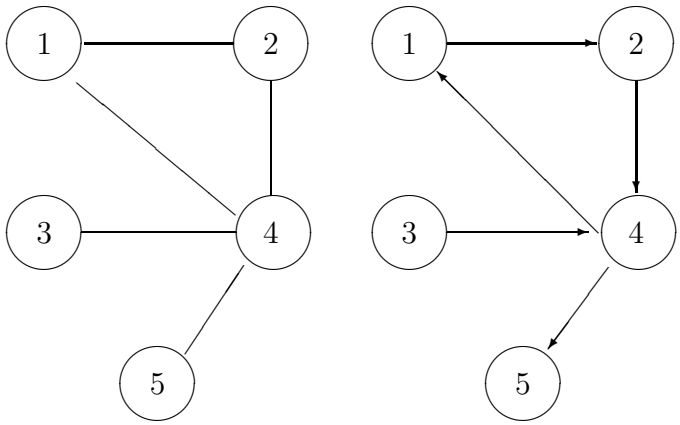
die Indexmenge der beizubehaltenden Nullhypothesen ist. Diese Prozedur ist identisch mit dem Abschlußtest, der die Schnitthypothesen mit Bonferroni-Globaltests testet (Bergmann, Hommel 1988).

4.1.2 Graphentheoretischer Ansatz

Definition von Graphen und Grundbegriffen

In diesem Abschnitt werden die für die Abbildung von Hypothesensystemen benötigten Begriffe aus der Graphentheorie eingeführt. Die Darstellung orientiert sich an „Kapitel 2.1 Grundbegriffe der Graphentheorie“ aus Neumann, Morlock (1993).
Ein *Graph* besteht aus einer nicht leeren Menge V von Knoten, einer Menge E mit $E \cap V = \emptyset$ und einer Inzidenzabbildung, die jedem $e \in E$ genau zwei Knoten $i, j \in V$ zuordnet.[1] Ist das e zugeordnete Paar i, j ungeordnet, so heißt e *Kante* mit den Endknoten i, j. Ein Graph, der nur aus Kanten besteht, wird *ungerichteter Graph* oder *Graph* G genannt. Ist das e zugeordnete Paar von Knoten i, j geordnet, wobei i der erste und j der zweite Knoten der Paares (i, j) ist, dann heißt e *Pfeil* mit dem Anfangsknoten i und dem Endknoten j. Ein solcher Graph heißt *gerichteter Graph* oder *Digraph* $\vec{G}$. Weiterhin nehme man an, daß in einem Graphen zwischen zwei Knoten maximal eine Kante existiert, und in einem Digraphen zwei Knoten maximal über zwei entgegengesetzte Pfeile verbunden sind. Somit ist jede Kante eindeutig über ihre Endknoten i, j festgelegt; sie wird im folgenden mit dem Symbol $[i, j]$ bezeichnet. Genauso ist jeder Pfeil durch Anfangsknoten i und Endknoten j eindeutig bestimmt und wird mit $\langle i, j \rangle$ bezeichnet. Ein Graph mit der Knotenmenge V und Kantenmenge E wird durch das Symbol $[V, E]$, und ein Digraph mit der Knotenmenge V und der Pfeilmenge E wird durch das Symbol $\langle V, E \rangle$ dargestellt.
Beide Arten von Graphen lassen sich einfach durch Matrizen beschreiben. Sei $V = \{1, \ldots, k\}$ die Knotenmenge des zugrundeliegenden Graphen beziehungsweise gerichteten Graphen. Dann heißt die dem Graphen $G = [V, E]$ zugeordnete symmetrische

[1] Die Symbole V für die Knotenmenge und E für die Kantenmenge sowie e für eine Kante kommen von den englischen Bezeichnungen „vertex“ für Ecke und „edge“ für Kante.



Graph 1 Digraph 1

Abbildung 4.1: Graphen

$k \times k-$ Matrix $U(G)$ mit den Elementen

$$u_{ij} := \begin{cases} 1, & \text{falls } [i,j] \in E \\ 0, & \text{sonst} \end{cases} \quad (i,j = 1,\ldots,k)$$

Adjazenzmatrix von G. Entsprechend wird die dem Digraphen $\vec{G} = \langle V, E\rangle$ zugeordnete $k \times k-$ Matrix $U(\vec{G})$ mit den Elementen

$$u_{ij} := \begin{cases} 1, & \text{falls } \langle i,j\rangle \in E \\ 0, & \text{sonst} \end{cases} \quad (i,j = 1,\ldots,k)$$

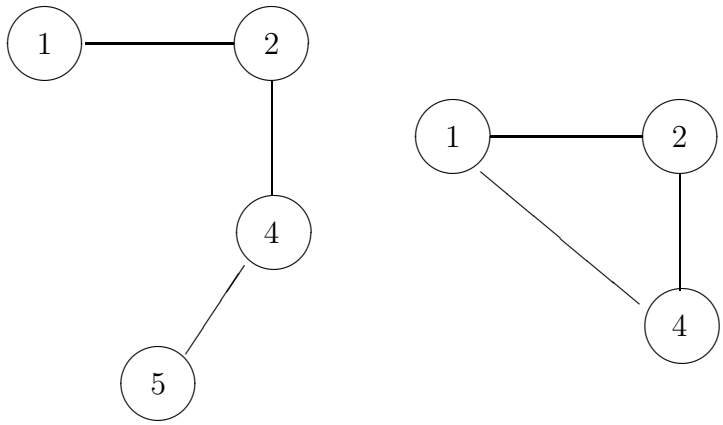
Adjazenzmatrix von $\vec{G}$ genannt. Die Adjazenzmatrizen zu den Graphen aus Abbildung 4.1 sind somit

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Graph 1 Digraph 1

Ein Graph G wird *vollständig* genannt, wenn jeder Knoten direkt mit allen anderen Knoten des Graphen verbunden ist. Analog wird ein Digraph $\vec{G}$ vollständig genannt, wenn zu je zwei Knoten i, j auch die Pfeile $\langle i,j\rangle$ und $\langle j,i\rangle$ existieren. Ein vollständiger Graph mit k Knoten hat $k(k-1)/2$ Kanten, während ein vollständiger Digraph mit k Knoten $k(k-1)$ Pfeile besitzt.

Ein *Teilgraph* $G' = [V', E']$ eines Graphen $G = [V, E]$ ist durch die Eigenschaften $V' \subseteq V$ und $E' \subseteq E$ definiert. Weiterhin heißt die Knotenmenge eines vollständigen Teilgraphen von G *Clique*. Umfaßt die Clique maximal viele Knoten aus V, daß heißt, es



Teilgraph von Graph 1 maximale Clique in Graph 1

Abbildung 4.2: Teilgraph, Clique

gibt keine andere Clique in G, die mehr Knoten umfaßt, so ist sie eine *maximale Clique* (Beispiel siehe Abbildung 4.2). Analog ist ein Teilgraph eines Digraphen definiert.
In einem Digraphen wird eine Folge $\langle i_0, i_1\rangle, \langle i_1, i_2\rangle, \ldots, \langle i_{r-1}, i_r\rangle$ von Pfeilen *Pfeilfolge* mit dem Anfangsknoten i_0 und dem Endknoten i_r oder kurz Pfeilfolge von i_0 nach i_r genannt und mit dem Symbol $\langle i_0, i_1, \ldots, i_r\rangle$ bezeichnet. Ist $i_0 \neq i_r$, so wird die Pfeilfolge offen, anderenfalls geschlossen genannt. Eine Pfeilfolge mit lauter verschiedenen Knoten heißt *Weg*, und eine geschlossene Pfeilfolge mit lauter verschiedenen „Zwischenknoten“ wird *Zyklus* genannt. Zum Beispiel ist im Digraph 1 ein Weg: $\langle 3, 4\rangle, \langle 4, 1\rangle, \langle 1, 2\rangle$ oder ein Zyklus $\langle 4, 1\rangle, \langle 1, 2\rangle, \langle 2, 4\rangle$. Ein Digraph, der keinen Zyklus enthält, heißt *zyklenfrei*. Ein Knoten j in $\vec{G}$ heißt von einem Knoten i aus *erreichbar*, wenn es einen Weg von i nach j gibt. Jeder Knoten hat hierbei die Eigenschaft, von sich selbst aus erreichbar zu sein. Die Menge der von einem Knoten erreichbaren Knoten wird mit $R(i)$ bezeichnet, die Menge derjenigen Knoten, von denen aus i erreichbar ist, mit $\bar{R}(i)$. Zudem sei $\dot{R}(i) = R(i) \setminus \{i\}$ die Menge der von i erreichbaren, aber verschiedenen Knoten. So ist im Digraph 1 $R(1) = \{1, 2, 4, 5\}$, $\bar{R}(1) = \{2, 3, 4\}$ und $\dot{R}(1) = \{2, 4, 5\}$.

Eigenschaften und grundlegende Algorithmen

Im folgenden wird von Interesse sein, welche Knoten in einem gerichteten Graphen von einem anderen Knoten aus erreichbar sind. Diese Fragestellung läßt sich mit Hilfe der Adjazenzmatrix unter Verwendung von Boolscher Algebra elegant beantworten. Als Boolsche Addition und Multiplikation sind die folgenden Abbildungen zu verstehen:

$$\begin{array}{rccl} & \{0,1\} \times \{0,1\} & \longrightarrow & \{0,1\} \\ \oplus : & & & \begin{cases} 0, & \text{falls } a = b = 0 \\ 1, & \text{sonst} \end{cases} \\ & a \oplus b & \longmapsto & \end{array}$$

$$\otimes : \begin{array}{ccc} \{0,1\} \times \{0,1\} & \longrightarrow & \{0,1\} \\ a \otimes b & \longmapsto & \begin{cases} 1, & \text{falls } a = b = 1 \\ 0, & \text{sonst} \end{cases} \end{array}$$

Damit lassen sich für die Adjazenzmatrizen A, B, die zum einen quadratisch sind und zum anderen nur aus Nullen und Einsen bestehen, ebenfalls Boolsche Verknüpfungen definieren:

$$A \oplus B := (a_{ij} \oplus b_{ij})_{i,j=1,\ldots,k}$$
$$A \otimes B := (c_{ij})_{i,j=1,\ldots,k} \text{ mit } c_{ij} := (a_{i1} \otimes b_{1j}) \oplus \ldots \oplus (a_{ik} \otimes b_{kj}).$$

Zusätzlich definiere man noch eine zweite Multiplikation zweier Matrizen A, B aus $\{0,1\}^{k \times k}$:

$$A \times B := (c_{ij})_{i,j=1,\ldots,k} \text{ mit } c_{ij} := a_{ij} \otimes b_{ij}.$$

Über die Bestimmung der Erreichbarkeitsmatrix geben nun die folgenden drei Sätze Auskunft:

Satz 4.5 (Satz 3.6 Noltemeier 1976)
Sei A eine Adjazenzmatrix eines Digraphen $\vec{G}$ und $A^p = (a_{ij}^{(p)})$. Es ist $a_{ij}^{(p)} = 1$ genau dann, wenn es eine Pfeilfolge der Länge p vom Knoten i zum Knoten j gibt. Insbesondere zeigt $a_{ii}^{(p)}$ an, daß es einen Zyklus der Länge p gibt, der durch den Knoten i läuft.

Lemma 4.6 (Korollar 3.7 Noltemeier 1976)
Die Erreichbarkeitsmatrix R ist die Matrix, in der zu jedem Knoten die von ihm erreichbaren Knoten aufgeführt sind. Diese Matrix eines Digraphen mit k Knoten und Adjazenzmatrix A erhält man durch

$$R(A) = E_k \oplus A \oplus A^2 \oplus \ldots \oplus A^{k-1}.$$

Für die Anwendung interessiert auch eine Abwandlung der Erreichbarkeitsmatrix. Bei dieser Matrix $\dot{R}$ soll ein Diagonalelement r_{ii} nur dann gleich 1 sein, wenn es mindestens einen Zyklus durch i gibt.

Lemma 4.7
Die Erreichbarkeitsmatrix $\dot{R}$, eines Digraphen mit k Knoten und Adjazenzmatrix A, ist die Matrix, in der zu jedem Knoten die von ihm erreichbaren Knoten aufgeführt sind und die Diagonalelemente r_{ii} nur dann gleich 1 sind, wenn es mindestens einen Zyklus durch i gibt. Diese Matrix erhält man durch

$$\dot{R}(A) = A \oplus A^2 \oplus \ldots \oplus A^k.$$

Lemma 4.8
Ist $\text{Spur}(\dot{\text{R}}(\text{A})) = 0$, so ist der Digraph zyklenfrei.

Lemma 4.8 liefert uns insbesondere ein einfaches Entscheidungskriterium für die Zyklenfreiheit.

Beispiel Betrachte die Adjazenzmatrix des Digraphen 1 von Seite 74.

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$ Ausgangsmatrix

$$A^2 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$ Es gibt einen Weg über zwei Pfeile von 1 nach 4, von 2 nach 1, von 2 nach 5, von 3 nach 1, von 3 nach 5 und von 4 nach 2.

$$A^3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$ Es gibt drei Zyklen der Länge 3, sowie einen Weg über drei Pfeile von 1 nach 5 und von 3 nach 2.

$$A^4 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$ in diesem Schritt sind keine neuen Wege gefunden worden.

Die Erreichbarkeitsmatrizen ergeben sich hier zu:

$$R(A) = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \text{ und } \dot{R}(A) = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

4.2 Zweiseitige Hypothesen

4.2.1 Darstellung des All-paar-Vergleichs durch Graphen

Die zu vergleichenden Erwartungswerte seien $\mu_1, \ldots, \mu_k$. Eine Nullhypothese, die für den Test auf Ungleichheit von μ_i und μ_j steht, sei mit $H_{(i,j)} : \ \mu_i = \mu_j$ bezeichnet.
Ein System von paarweisen Hypothesen $H = \{H_{(i,j)} : \ (i,j) \in I\}$ wird über zwei Ab-

bildungen mit einem Graphen identifiziert:

$$\begin{array}{lccl} 1. & \{\mu_1, \ldots, \mu_k\} & \longrightarrow & V = 1, \ldots, k \\ & \mu_i & \longmapsto & i \\ \\ 2. & H & \longrightarrow & E \subseteq \{1, \ldots, k\} \times \{1, \ldots, k\} \\ & H_{(i,j)} & \longmapsto & (i, j). \end{array}$$

Auf diese Weise wird je ein Erwartungswert mit einem Knoten und jede Hypothese mit der Kante, die die zu vergleichenden Knoten (Erwartungswerte) verbindet, identifiziert. Zum Beispiel ergibt das Hypothesensytem $\mu_1 = \mu_2$, $\mu_2 = \mu_4$, $\mu_3 = \mu_4$, $\mu_4 = \mu_1$, $\mu_4 = \mu_5$ als Graph genau den Graph 1 von Seite 74.
Im folgenden wird das Hypothesensystem betrachtet, das zum All-paar-Vergleich von k Gruppen gehört. Zwischen den erschöpfenden Indexmengen des All-paar-Vergleichs und den Cliquen eines Graphen besteht eine eineindeutige Beziehung.

Satz 4.9
Sei H das Hypothesensystem für den All-paar-Vergleich von k Gruppen und G der zugehörige Graph. Jede erschöpfende Indexmenge entspricht genau einer Zerlegung des Graphen in Cliquen und Einzelknoten.

Beweis: $\Longrightarrow$:
Sei I eine erschöpfende Indexmenge. Ist H_I eine Partitionshypothese, so sei $I = \dot{\bigcup}_{i=1}^{j} I_i$ die Zerlegung in Schnitthypothesen. Für jede Schnitthypothese gilt, daß sie der Form $\mu_{i_1} = \ldots = \mu_{i_m}$ ist, da sie zu einer erschöpfenden Indexmenge gehört. Damit entspricht ihr genau die Clique $\{i_1, \ldots, i_m\}$. Da die Zerlegung disjunkt ist, sind auch die Cliquen disjunkt. Die Erwartungswerte, die in keiner der Hypothesen vorkommen, bilden die Einzelknoten.
$\Longleftarrow$: Seien $C_1, \ldots, C_l$ die Cliquen einer Zerlegung des Graphen und $v_1, \ldots, v_m$ die Einzelknoten. In jeder Clique C_i ist jeder Knoten mit allen anderen verbunden, d.h. alle möglichen paarweisen Vergleiche zweier Mittelwerte in der Clique sind durch eine Kante repräsentiert. Damit ist die Schnitthypothese, die C_i entspricht, maximal. Da alle den C_i entsprechenden Schnitthypothesen maximal sind, ist die disjunkte Vereinigung auch maximal, denn bei Hinzunahme einer weiteren Elementarhypothese (Kante) würde die Hypothese verändert. Die Einzelknoten entsprechen den Erwartungswerten, die in keiner Hypothese vorkommen.

$\triangledown$

Lemma 4.10
Sei H das Hypothesensystem für einen Beliebige-Paare-Vergleich mit k Gruppen, und

sei G der zugehörige Graph. Weiterhin sei mit $\bar{G}$ der vollständige Graph für die k Gruppen bezeichnet. Dann gilt:
Jede erschöpfende Indexmenge entspricht genau einem Schnitt von einer Zerlegung des Graphen $\bar{G}$ in Cliquen und Einzelknoten mit dem Graphen G.

4.2.2 Durchführung der dynamischen Shaffer-Prozedur

Zur Durchführung von All-paar-Vergleichen unter Einhaltung des multiplen Niveaus stehen mehrere Methoden zur Verfügung. Die einfachste ist die Bonferronisierung, bei der jeder Test zum Niveau α/l bei $l = \frac{1}{2}k(k-1)$ Vergleichen durchgeführt wird. Eine einfache Verbesserung der Trennschärfe läßt sich durch das Verwenden der Holm-Prozedur erreichen (Holm 1979). Hierfür werden die p-Werte der Größe nach geordnet und beginnend mit dem kleinsten solange mit $\alpha/(l-i)$ verglichen, bis eine Hypothese nicht mehr abgelehnt werden kann; i ist hierbei die Anzahl der schon abgelehnten Hypothesen. Eine weitere Verbesserung hat Shaffer (1986) vorgestellt. Auch hier werden die geordneten p-Werte mit adjustierten α Schranken verglichen, die größer oder gleich denen aus der Holm-Prozedur sind. Dafür werden die p-Werte aufsteigend jeweils mit α/i verglichen, wobei i die Anzahl der maximal noch gültigen Nullhypothesen einschließlich der zu testenden ist. Die Anzahl der noch maximal gültigen Nullhypothesen in jedem Schritt hängt jedoch von den schon abgelehnten Hypothesen ab. Die ungünstigsten Teilerverläufe für 3-10 Gruppen sind bei Shaffer (1986) oder Holland und Copenhaver (1987) tabelliert. So lauten die Teiler bei vier Gruppen $(6, 3, 3, 3, 2, 1)$, d.h. der kleinste p-Wert wird mit $\alpha/6$ verglichen, die drei nächst größeren mit $\alpha/3$ und die beiden letzten mit $\alpha/2$ bzw. α. Arbeitet man nur mit diesen statischen Teilern, so spricht man von der 1. Shaffer-Prozedur. Soll jedoch dynamisch die α-Schranke in jedem Schritt bestimmt werden (2. Shaffer-Prozedur oder auch dynamische Shaffer-Prozedur), so ist dies anhand von Satz 4.9 unter Zuhilfenahme eines Graphen einfach durchzuführen.

Shaffer-Prozedur Es seien k Gruppen paarweise zu vergleichen. Die Anzahl der zweiseitigen Vergleiche ist damit $l = \frac{k(k-1)}{2}$, und die geordneten p-Werte seien $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(l)}$. Weiterhin sei G der nach obiger Definition dem Problem zugeordnete Graph.

i) $i = 1, j = l$.

ii) Ist $p_{(i)} \leq \alpha/j$, dann lehne $H_{(i)}$ ab und fahre fort, sonst terminiere.

iii) Entferne die Kante, die $H_{(i)}$ entspricht, aus dem Graph.

iv) $i = i + 1$; ist $i = l + 1$, so terminiere.

v) Bestimme die Cliquenzerlegung von G, die zum einen die Kante zur Hypothese $H_{(i)}$ enthält und zum anderen maximal viele Kanten umfaßt. Die Anzahl der Kanten in dieser Zerlegung sei j. Gehe zu ii).

Ebenso läßt sich die Shaffer-Prozedur für Beliebige-Paare-Vergleiche unter Zuhilfenahme eines Graphen nach Lemma 4.10 formulieren. Jedoch ist die Durchführung aufgrund der Bestimmung der Cliquenzerlegung aus Punkt v) unübersichtlich. Zudem sind schon andere Lösungen für die zweiseitige Fragestellung bekannt, die nicht nur paarweise Kontraste sondern beliebige Kontraste umfassen. In der Dissertation von Bernhard (1991) sind Algorithmen angegeben, die ein Hypothesensystem von beliebigen Kontrasten durch einen Abschlußtest testen, welcher noch trennschärfer als die Shaffer-Prozedur ist. Der Algorithmus zum Testverfahren mit der höchsten Trennschärfe ist in Abschnitt 4.3.2 beschrieben. Vom Prinzip her wird die Potenzmenge der Indizes durchlaufen und für jede Indexmenge geprüft, ob sie erschöpfend ist. Diese Überprüfung erfolgt anhand der Ränge der den Indexmengen entsprechenden Hypothesenmatrizen. Hierbei gilt nämlich, daß eine Indexmenge genau dann erschöpfend ist, wenn sich zu ihr keine weitere Elementarhypothese hinzunehmen läßt, ohne daß der Rang der entsprechenden Hypothesenmatrix zunimmt (Bernhard 1991, Lemma 1.8).
Ähnlich gibt Westfall (1997) an, wie sich für beliebige Kontraste mit Methoden der linearen Algebra logische Abhängigkeiten unter den Hypothesen fassen lassen. Die Vorgehensweise hierbei zielt auf eine direkte Bestimmung von adjustierten p-Werten ab. Dafür werden die p-Werte der einzelnen Nullhypothesen geordnet $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(l)}$. Entsprechend werden die Nullhypothesen mit $H_{(1)}$ bis $H_{(l)}$ bezeichnet. Die Nullhypothesen werden sequentiell verworfen anhand der Größe der original p-Werte. Hierbei gilt, daß $H_{(j)}$ nur dann verworfen werden kann, wenn zuvor auch $H_{(1)}$ bis $H_{(j-1)}$ verworfen wurden. Dafür wird die Menge $\bar{S}_j$ definiert, die alle möglichen wahren Nullhypothesen einschießlich $H_{(j)}$ enthält, gegeben, daß die Nullhypothesen $H_{(i)}$, $i = 1, \ldots, j-1$ falsch sind. Mit $S = \mathbb{N}_l$ setze man

$$\begin{aligned}
\bar{S}_1 &= \{S\} \\
\bar{S}_2 &= \left\{K \subset S | 2 \in K \ \wedge \bigcap_{m \in K} H_{(m)} \cap A_{(1)} \neq \emptyset\right\} \\
\bar{S}_3 &= \left\{K \subset S | 3 \in K \ \wedge \bigcap_{m \in K} H_{(m)} \cap A_{(1)} \cap A_{(2)} \neq \emptyset\right\} \\
&\vdots
\end{aligned}$$

$$\begin{aligned}
\bar{S}_j &= \left\{ K \subset S | j \in K \ \wedge \bigcap_{m \in K} H_{(m)} \cap \bigcap_{m=1}^{j-1} A_{(m)} \neq \emptyset \right\} \\
&\vdots \\
\bar{S}_l &= \{\{l\}\}
\end{aligned}$$

Diese Mengen werden noch vereinfacht, indem alle Teilmengen eliminiert werden. Mit

$$M_j := \max_{K \in S_j} |K|$$

ergeben sich die adjustierten p-Werte dann durch

$$\begin{aligned}
ap_{(1)} &= M_1 p_{(1)} = l p_{(1)} \\
ap_{(2)} &= \max\{ap_{(1)}, M_2 p_{(2)}\} \\
&\vdots \\
ap_{(j)} &= \max\{ap_{(j-1)}, M_j p_{(j)}\} \\
&\vdots \\
ap_{(l)} &= \max\{ap_{(l-1)}, p_{(l)}\}
\end{aligned}$$

Bei dieser Vorgehensweise liegt das Problem in der Bestimmung der Mengen $K \in \bar{S}_j$. Sei mit c_i der zur Hypothese H_i gehörende Kontrast bezeichnet. Zum Beispiel $H_i : \mu_1 - \mu_2 = 0 \Rightarrow c_i = (1, -1, 0, \ldots, 0)$ oder $H_i : \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 - \mu_3 = 0 \Rightarrow c_i = (\frac{1}{2}, \frac{1}{2}, -1, 0, \ldots, 0)$. Weiterhin sei mit C_K die Matrix bezeichnet, deren Spalten aus den Kontrastvektoren c_i, $i \in K$ besteht, und sei $\mathcal{C}(A)$ der reelle Vektorraum, der durch die Spalten von A aufgespannt wird. Nach Westfall (1997) ist für $K \in \bar{S}_j$ notwendig und hinreichend, daß $j \in K$ und $c_{(i)} \notin \mathcal{C}(C_K)$ für alle $i = 1, \ldots, j-1$ ist. Um dies zu entscheiden, konstruiere man mit $C_{(j)}$, der Matrix, deren Spalten aus den Kontrastvektoren $c_{(1)}, \ldots, c_{(j-1)}$ bestehen, die Matrix $C_{(j)}^{\perp K} = (E_k - C_K(C_K' C_K)^{-} C_K')C_{(j)}$. Ist $C_{(j)}^{\perp K} = O$, dann ist $K \in \bar{S}_j$. Beide Verfahrensweisen lassen sich nicht auf den Fall von einseitigen Hypothesen verallgemeinern, wie folgendes Gegenbeispiel zeigt:

Gegenbeispiel

Sei die Situation des Beispiels für Dosisschritte von Seite 71 gegeben mit den Hypothesen $H_1 : \mu_1 \geq \mu_2$, $H_2 : \mu_2 \geq \mu_3$, $H_3 : \mu_1 \geq \mu_3$ mit den entsprechenden Kontrasten $c_1 = (1, -1, 0)$, $c_2 = (0, 1, -1)$, $c_3 = (1, 0, -1)$. Für die einseitigen Tests seien die p-Werte: $p_1 = 0,01$, $p_2 = 0,03$, $p_3 = 0,031$. Nach der Vorgehensweise von Bernhard (1991) sind die folgenden Matrizen und ihre Ränge zu betrachten:

$$Rang(-1\,1\,0) = 1 \qquad Rang\begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} = 2$$

$$Rang(0\,-1\,1) = 1 \quad Rang\begin{pmatrix} 0 & -1 & 1 \\ -1 & 0 & 1 \end{pmatrix} = 2 \quad Rang\begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} = 2$$

$$Rang(-1\,0\,1) = 1 \quad Rang\begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} = 2$$

Demnach würden die Indexmengen $\{1\}, \{2\}, \{3\}, \{1,2,3\}$ als erschöpfend erkannt. Wie in dem Beispiel gezeigt, fehlen hier noch die beiden Indexmengen $\{1,3\}$ und $\{2,3\}$. Wendet man das Verfahren aus Westfall (1997) an, so erhält man für $\bar{S}_1 = \{1,2,3\}$. Für $\bar{S}_2$ ist zu überprüfen, ob $c_1 \notin C_K$ mit $K = \{2\}, \{2,3\}$ ist. Entsprechend der oben angegebenen Bedingung berechne man

$$\begin{aligned} C_{(1)}^{\perp\{2\}} &= \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}\left((0\,1\,-1)\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}\right)^{-}(0\,1\,-1)\right]\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \\ &= \frac{1}{2}\begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} \\ C_{(1)}^{\perp\{2,3\}} &= \left[E_3 - \begin{pmatrix} 0 & -1 \\ 1 & 0 \\ -1 & 1 \end{pmatrix}\left(\left(\begin{pmatrix} 0 & -1 \\ 1 & 0 \\ -1 & 1 \end{pmatrix}'\begin{pmatrix} 0 & -1 \\ 1 & 0 \\ -1 & 1 \end{pmatrix}\right)^{-}\begin{pmatrix} 0 & -1 \\ 1 & 0 \\ -1 & 1 \end{pmatrix}'\right]\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \end{aligned}$$

Damit sind $\bar{S}_2 = \{\{2\}\}$ und $\bar{S}_3 = \{\{3\}\}$. Als adjustierte p-Werte erhält man dann $ap_1 = 0,03$, $ap_2 = 0,03$ und $ap_3 = 0,031$. Die kleinstmöglichen p-Werte erhält man, bei dieser Art Vorgehensweise, aus dem Abschlußtest. Diese lauten $ap_1 = 0,03$, $ap_2 = 0,06$ und $ap_3 = 0,06$. Folglich garantiert die Verwendung des Verfahrens von Westfall nicht mit der Argumentation über den Abschlußtest die Einhaltung des multiplen Niveaus.

4.2.3 Beispiel

Zur Schädlingsüberwachung von Anbauflächen lassen sich mit Klebstoff bestrichene Tafeln verwenden, die im Feld aufgestellt werden. Anhand der gefangenen Arten und der Fanghäufigkeit lassen sich so die Insektenpopulationen kontrollieren. In einem Feldversuch untersuchten Wilson und Shade (1967) den Effekt, den die Farbe der verwendeten Tafeln auf die Fanghäufigkeit des Getreideblattkäfers hat. Ein Auszug aus den Daten ist in Hsu (1996) ab Seite 140 gegeben. Es wurden sechs Tafeln in fünf verschiedenen Farben aufgestellt, wobei die Daten als Einweganlage betrachtet werden. Es wurden die mittleren Fangzahlen aus Tabelle 4.1 beobachtet. Der Schätzer für die Standard-

Farbe	Abkürzung	Fallzahl	Mittelwert
gelb	g	6	49,2
orange	o	6	34,2
rot	r	6	26,4
blau	b	6	17,2
weiß	w	6	15,0

Tabelle 4.1: Mittlere Fangzahlen je Farbe

abweichung ist $\hat{\sigma} = 7,956$ mit $5(6-1) = 25$ Freiheitsgraden. Als paarweise zweiseitige p-Werte aus dem multiplen t-Test erhalten wir damit die Werte aus Tabelle 4.2. Die

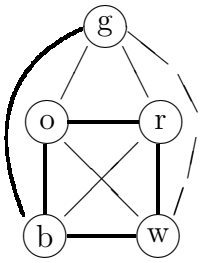
	orange	rot	blau	weiß
gelb	$0,0046$	$4,1 \cdot 10^{-5}$	$2,7 \cdot 10^{-7}$	$8,5 \cdot 10^{-8}$
orange		$0,0761$	$0,0007$	$0,0002$
rot			$0,056$	$0,020$
blau				$0,67$

Tabelle 4.2: p-Werte der zweiseitigen Vergleiche

Ordnung der p-Werte lautet somit:

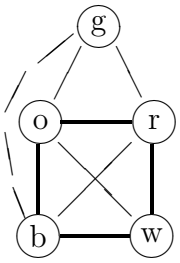
$$p_{gw} < p_{gb} < p_{gr} < p_{ow} < p_{ob} < p_{og} < p_{rw} < p_{rb} < p_{or} < p_{bw}.$$

Die Prozedur liefert die Graphen, Cliquenzerlegungen und α-Teiler aus Abbildung 4.3. Als globales Niveau ist $\alpha = 5\%$ gewählt. In den Graphen ist jeweils die Kante, die der Hypothese des aktuell kleinsten p-Werts entspricht, unterbrochen markiert. In Schritt ii) des Algorithmus aus 4.2.2 ist $p_{(1)} = p_{gw} = 8,5 \cdot 10^{-8} < 0,05/10$. Mit der Shaffer-Prozedur wird gelb als attraktivste Farbe identifiziert, und orange attraktiver als blau und weiß. Das in Hsu (1996) verwendete Tukey-Verfahren kommt zu dem gleichen Ergebnis. Im allgemeinen ist es von den Daten abhängig, ob dieses Verfahren nur gleichwertig oder besser (trennschärfer) ist.



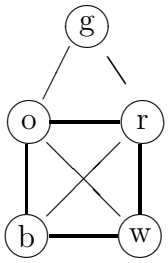
1.

Cliquenzerlegung:
{g,o,r,b,w}
Kanten: $j = 10$
$p_{gw} = 8,5 \cdot 10^{-8}$
$< 0,005$



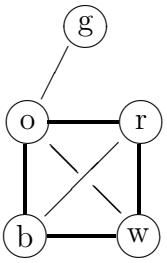
2.

{g,o,r,b}, $j = 6$
$p_{gb} = 2,7 \cdot 10^{-7}$
$< 0,0083$



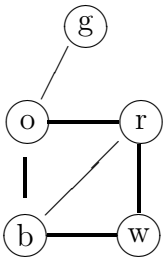
3.

{g,o,r}{bw} oder
{o,b,w}{g,r}, $j = 4$
$p_{gr} = 0,0004 < 0,0125$



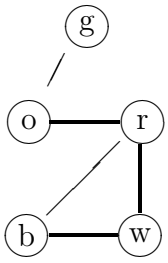
4.

{o,r,b,w}, $j = 6$
$p_{ow} = 2,1 \cdot 10^{-4}$
$< 0,0083$



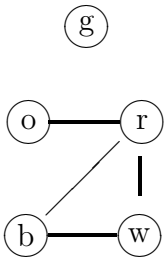
5.

{g,o,r,b}, $j = 3$
$p_{ob} = 7,2 \cdot 10^{-4}$
$< 0,016$



6.

{g,o} {r,b,w}, $j = 4$
$p_{go} = 0,0046$
$< 0,0125$



7.

{r,b,w}, $j = 3$
$p_{rw} = 0,02 > 0,016$
Abbruch

Abbildung 4.3: Graphen des Farbvergleichs

4.3 Einseitige Hypothesen

In der Praxis sind häufig die gestellten Fragen und Hypothesen einseitig formuliert, so zum Beispiel ob Mittel B dem Mittel A überlegen ist. Weitere Beispiele sind multiple Vergleiche gegen eine Kontrolle (Welche der Behandlungen sind besser als die Kontrolle?), die Bestimmung der minimalen effektiven Dosis (Welche Dosis ist die niedrigste, die besser als die Kontrolle ist?) oder die Bestimmung des höchsten effektiven Dosisschritts. Wie in Abschnitt 4.2.2 gezeigt lassen sich die Vorgehensweisen von Westfall (1997) und Bernhard (1991) nicht auf einseitige Hypothesensysteme übertragen. Dies liegt darin begründet, daß einseitige Hypothesen nicht lineare Teilräume des $\mathbb{R}^k$ sind, sondern Halbräume. Für diese gilt, daß der Schnitt zweier Halbräume i.a. kein Halbraum mehr ist, und Halbräume auch keine Vektorräume sind. Damit ist die theoretische Grundlage für die Rechenoperationen aus beiden Verfahren nicht gegeben.
Im folgenden sei ein Hypothesensystem von nur einseitigen Nullhypothesen betrachtet, d.h. die Elementarhypothesen haben die Form $H_{\langle i,j \rangle} : \mu_i \geq \mu_j$. Besteht ein Hypothesensystem aus ein- und zweiseitigen Hypothesen, so läßt es sich in ein rein einseitiges transformieren. Dafür werden die zweiseitigen Hypothesen $\mu_i = \mu_j$ in zwei einseitige,

$\mu_i \geq \mu_j$ und $\mu_i \leq \mu_j$, zerlegt, was auch generell für zweiseitige Hypothesensysteme gilt. Durch diese Form der Transformation erhält man zusätzlich die Kontrolle über den Fehler 3. Art (Holm 1979, Bauer et al 1986). Bei der Bezeichnung der Hypothesen werden die folgenden Notationen verwendet:
Die Elementarhypothesen werden mit $H_i,\ i = 1, \ldots l$ bzw. $H_{\langle i,j \rangle} :\ \mu_i \geq \mu_j$ bezeichnet, wobei die Reihenfolge der Indizes die Ordnung der Mittelwerte bezeichnet. Die Schnitthypothese zu der Indexmenge $I = \{i_1, \ldots, i_m\}$ ist mit $H_{\{i_1,\ldots,i_m\}}$ bezeichnet, und die Schnitthypothesen, die Ungleichungsketten entsprechen, d.h. $\mu_{i_1} \geq \mu_{i_2} \geq \ldots \geq \mu_{i_m}$, sind mit $H_{\langle i_1,i_2,\ldots,i_m \rangle}$ bezeichnet.
Um das multiple Niveau über alle Entscheidungen zu kontrollieren, sind auch hier wieder die Adjustierungen nach Bonferroni, Holm oder Shaffer möglich. Weitere Verbesserungen wären die Verwendung des Hochberg-Verfahrens (Hochberg 1988) oder das step-up-Verfahren nach Hommel (1988), welche jedoch nicht für alle Abhängigkeitsstrukturen anwendbar sind, da sie auf dem Test von Simes(1986) beruhen. Für die Trennschärfe der einzelnen Verfahren gilt ($\leq$ bedeutet trennschärfer):

$$\text{Bonferroni} \leq \text{Holm} \leq \text{Shaffer} \leq \text{Abschlußtest mit Bonferroni Globaltests}$$

Zur Durchführung des Abschlußtests mit Bonferroni-Globaltests haben Bergmann und Hommel (1988) eine Prozedur angegeben, im folgenden Prozedur 4 (P4B) genannt, die eine computergestützte Durchführung ermöglicht. Unter anderem hat Bernhard (1991) einen effizienten Algorithmus konstruiert, um diese Prozedur durchzuführen. Für die Anwendung des Algorithmus ist notwendig, daß zu einer beliebigen Kombination von Elementarhypothesen die kleinste sie umfassende, erschöpfende Hypothese bestimmt werden kann. Für den Fall von nur zweiseitigen paarweisen oder Kontrasthypothesen ist dieses Problem in Berhard's Arbeit gelöst. Bei einseitigen paarweisen Hypothesen läßt sich dieses Problem durch Betrachtung von gerichteten Graphen lösen. Für beide Definitionen von erschöpfenden Indexmengen werden in den nachfolgenden Abschnitten Verfahren eingeführt, mit denen die kleinste umfassende erschöpfende Indexmenge bestimmt werden kann; ihre Implementierung wird beschrieben. Hierbei ist zu beachten, daß die kleinste umfassende erschöpfende Indexmenge einer Indexmenge eindeutig ist, die kleinste umfassende streng erschöpfende jedoch nicht. Durch die Verschärfung der Eigenschaft, erschöpfend zu sein, verbessert sich die Trennschärfe des Verfahrens noch einmal, da weniger Schnitthypothesen getestet werden. Nachfolgend wird zuerst die Prozedur P4B vorgestellt, wie sie in Bernhard (1991) angegeben ist. Danach ist beschrieben, wie sich die kleinste umfassende erschöpfende Indexmenge (Hypothese) zu einer beliebigen Indexmenge (Schnitt-/Partitionshypothese) bestimmen läßt. Verwen-

det man die schärfere Definition von erschöpfend, so muß die Prozedur abgewandelt werden, was im 4. Abschnitt beschrieben ist. Zuletzt wird als eine Anwendung des Verfahrens die Bestimmung des höchsten effektiven Dosisschritts (HEDS) mit der Prozedur 4 beschrieben und bezüglich der Power mit anderen Verfahren zur Bestimmung des HEDS verglichen.

4.3.1 Darstellung durch gerichtete Graphen

Die Erwartungswerte, die zu vergleichen sind, seien mit $\mu_1, \ldots, \mu_k$ bezeichnet. Eine einseitige Hypothese der Art $\mu_i \geq \mu_j$ sei mit $H_{\langle i,j \rangle}$ bezeichnet. Ein System von paarweisen einseitigen Hypothesen $H = \{H_{\langle i,j \rangle} : (i,j) \in I\}$ identifiziere man über zwei Abbildungen mit einem Digraphen

$$\begin{array}{lccl} 1. & \{\mu_1, \ldots, \mu_k\} & \longrightarrow & V = 1, \ldots, k \\ & \mu_i & \longmapsto & i \\ & & & \\ 2. & H & \longrightarrow & E \subseteq \{1, \ldots, k\} \times \{1, \ldots, k\} \\ & H_{\langle i,j \rangle} & \longmapsto & \langle i, j \rangle \end{array}$$

Durch die erste Abbildung wird je ein Mittelwert mit einem Knoten identifiziert, während die zweite jede Hypothese $\mu_i \geq \mu_j$ mit dem Pfeil von i nach j identifiziert. So ergibt das Hypothesensystem $\mu_1 \geq \mu_2$, $\mu_2 \geq \mu_4$, $\mu_3 \geq \mu_4$, $\mu_4 \geq \mu_1$, $\mu_4 \geq \mu_5$ genau den Digraph 1 von Seite 74. Betrachten wir nun den Schnitt von zwei Hypothesen $H_{\langle g,h \rangle}$ und $H_{\langle i,j \rangle}$, so kann der Fall eintreten, daß $h = i$ ist, d.h. die Hypothesen lauten $\mu_g \geq \mu_h$ und $\mu_h \geq \mu_j$. Damit folgt auch, daß $\mu_h \geq \mu_j$ ist, also der Schnitt der beiden Hypothesen die Hypothese $H_{\langle h,j \rangle}$ impliziert. Ist $h \neq i$ bzw. $g \neq j$, wird durch den Schnitt keine weitere mögliche Elementarhypothese impliziert. Wenn man diese Situation mit Graphen darstellt, können die Graphen aus Abbildung 4.4 auftreten: Die im fünften Graphen implizierte Hypothese $\mu_g \geq \mu_j$ entspricht genau dem Pfeil, zu dem es einen Weg mit einer Länge größer eins gibt. Allgemein ist zu sehen, daß jede Hypothesenungleichungskette $\mu_{i_1} \geq \ldots \geq \mu_{i_m}$ genau dem Weg $\langle i_1, \ldots, i_m \rangle$ entspricht.

4.3.2 Beschreibung der Prozedur P4B aus Bernhard (1991)

Um den Abschlußtest für l Elementarhypothesen computergestützt durchführen zu können, wird ein Verfahren benötigt, das die Potenzmenge $\wp_l^0$ systematisch generiert. In dem verwendeten Verfahren geschieht dies rekursiv. Die spezielle Ordnung, die bei dieser Erzeugung zugrunde liegt, wird zudem dazu genutzt, teilweise Indexmengen zu überspringen, die zu schon generierten, erschöpfenden Indexmengen führen. Zur Beschreibung der Rekursionsvorschrift werden die Symbole i^* für das größte und i^{**} für

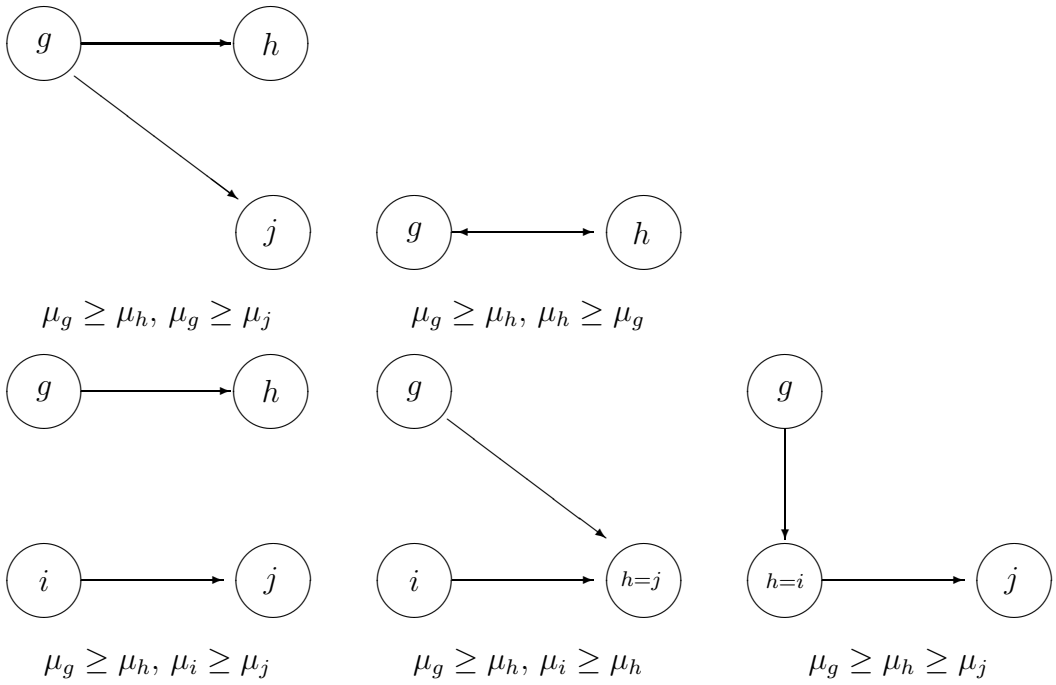


Abbildung 4.4: Zwei einseitige Hypothesen als Graphen

das zweitgrößte Element einer Indexmenge I verwendet, jeweils unter der Voraussetzung, daß $|I| \geq 1$ bzw. $|I| \geq 2$ ist.

Rekursionsvorschrift zur Erzeugung von $\wp_l$

i) Beginne mit $k = 1$ und $I_1 = \{1\}$[2]

ii) Für alle $k = 1, \ldots, 2^l - 1$ ist die nachfolgende Indexmenge:

$$I_{k+1} := \begin{cases} I_k \cup \{i^* + 1\} & \text{, falls } i^* < l \quad \textbf{(R1)} \\ (I_k \setminus \{i^*, i^{**}\}) \cup \{i^{**} + 1\} & \text{, falls } i^* = n,\ |I_k| > 1 \quad \textbf{(R2)} \\ \emptyset & \text{, falls } i^* = n,\ |I_k| = 1 \quad \textbf{(R2)} \end{cases}$$

Nachdem die Menge der erschöpfenden Indexmengen durch $\wp_{EI} = \{I \in \wp_l^0 : I \text{ erschöpfend}\}$ definiert ist, liegt es nahe, die erschöpfenden Indexmengen selber durch sukzessives Durchlaufen der Potenzmenge und jeweiliges Prüfen auf die Eigenschaft, erschöpfend zu sein, zu generieren. Läßt sich zu jeder Indexmenge die kleinste sie umfassende erschöpfende bestimmen und nutzt man die durch die Rekursion benutzte Ordnung der Potenzmenge aus, so lassen sich zwei Kriterien formulieren. Diese prüfen, ob auf die Bestimmung von $SE(I)$ verzichtet werden kann, bzw. ob die berechnete Indexmenge $SE(I)$ schon generiert wurde. Das erste Kriterium kommt vor der Berechnung von

[2]Der für die Indexmengen verwendete Index k steht in keinem Zusammenhang mit der Anzahl der Erwartungswerte

$SE(I)$ zum Tragen und prüft, ob $SE(I)$ überhaupt zu berechnen ist, während das zweite nach der Bestimmung von $SE(I)$ prüft, ob diese Menge schon generiert wurde. Zur Vereinfachung setze man noch $C(I) := SE(I) \setminus I$ als die Menge der in I_k implizierten Hypothesen.

Kriterium 1 (K1) Gilt für die im $k-$ten Schritt generierte Indexmenge I_k, daß $i^* \in C(I_k \setminus \{i^*\})$ ist, so kann auf die Bestimmung von $SE(I_k)$ verzichtet werden, und die nächste zu betrachtende Indexmenge wird erzeugt durch

$$I_{k+1} := \begin{cases} (I_k \setminus \{i^*\}) \cup \{\min M\} & \text{, falls } M \neq \emptyset \\ \text{erzeugt mit } \mathbf{(R2)} & \text{, falls } M = \emptyset, \end{cases}$$

wobei $M := \{i^* + 1, \ldots, l\} \setminus C(I_k \setminus \{i^*\})$ ist.

Kriterium 2 (K2) Seien im $k-$ten Rekursionsschritt I_k und $SE(I_k) = I_k \cup C(I_k)$ bestimmt worden. Ist $D := (C(I_k) \setminus C(I_k \setminus \{i^*\}) \neq \emptyset$ und $\min D < \max I_k$, so wurde $SE(I_k)$ bereits erzeugt, und die nächste zu betrachtende Indexmenge ist

$$I_{k+1} := \begin{cases} (I_k \setminus \{i^*\}) \cup \{\min M\} & \text{, falls } M \neq \emptyset \\ \text{erzeugt mit } \mathbf{(R2)} & \text{, falls } M = \emptyset, \end{cases}$$

wobei $M := \{i^* + 1, \ldots, l\} \setminus C(I_k \setminus \{i^*\})$ ist.

Durch die Anwendung der beiden Kriterien läßt sich der Rechenaufwand teilweise erheblich verringern. Aufgrund der der Potenzmenge zugrundeliegenden Ordnung und deren Eigenschaften (siehe Bernhard 1991, S.37ff) muß nur ein Teil der Mengen $C(I_k \setminus \{i_k^*\})$ gespeichert werden. Durch systematisches Anwenden der Kriterien 1 und 2 läßt sich somit folgender, in Abbildung 4.5 dargestellter Algorithmus zur Erzeugung aller erschöpfenden Indexmengen konstruieren. Hierbei sind die Mengen $C(I_k)$ unter J_j mit $j = |I_k|$ abgelegt. Zu Beginn werden alle Mengen initialisiert.

i) Zuerst wird Kriterium 1 geprüft, d.h. ob $i_k^* \in J_{j-1} = C(I_k \setminus \{i_k^*\})$ mit $j = |I_k|$ ist.

ii) Die Bestimmung der kleinsten I_k umfassenden erschöpfenden Indexmenge wird wie in Abschnitt 4.3.3 beschrieben durchgeführt.

iii) Prüfung des Kriteriums 2. Ist dies nicht erfüllt, so ist $SE(I_k)$ eine „neue“ erschöpfende Indexmenge.

iv) Ist I_{k+1} leer, so sind alle Mengen der Potenzmenge betrachtet worden, und der Algorithmus stoppt.

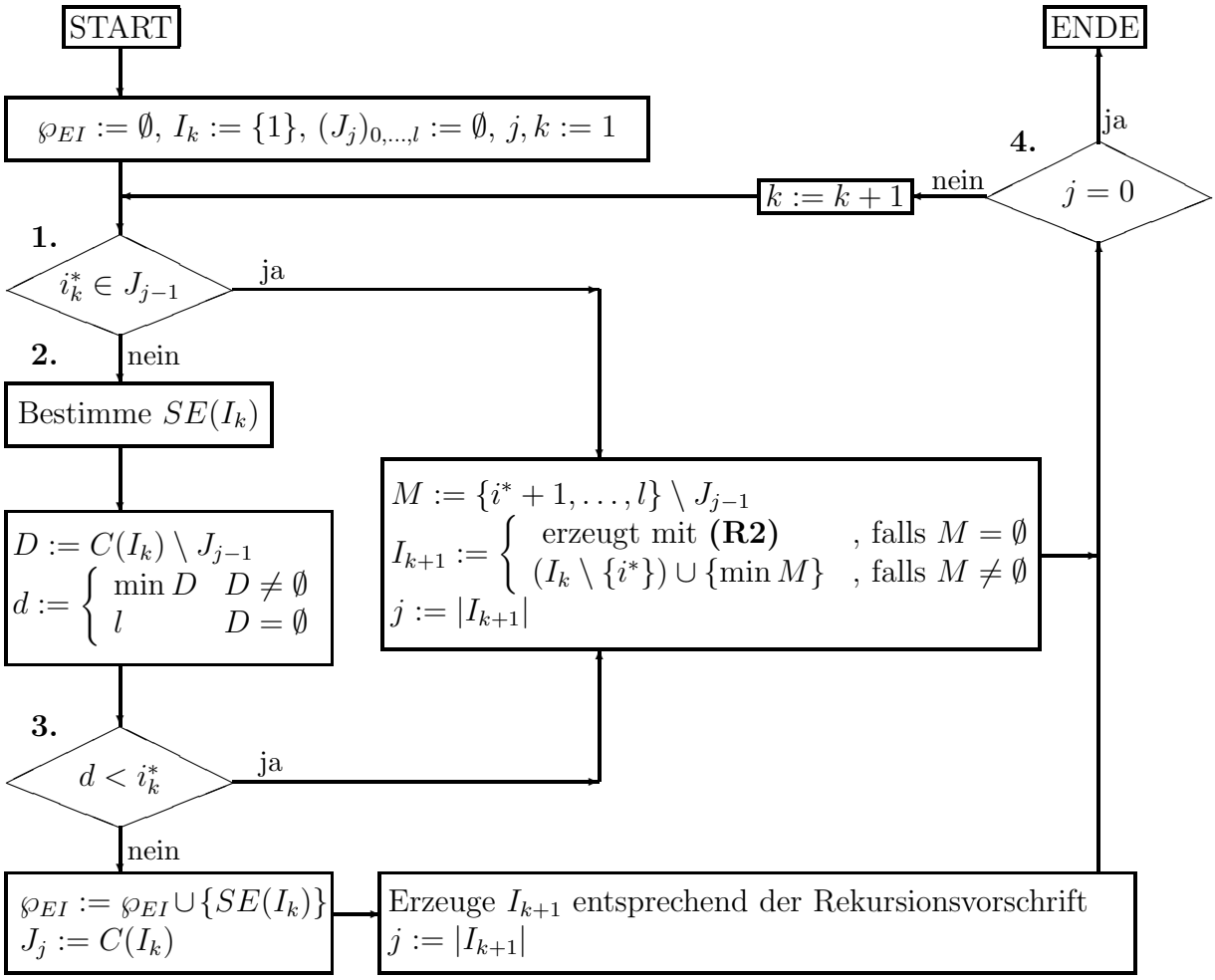


Abbildung 4.5: Flußdiagramm zur Bestimmung von $\wp_{EI}$

Damit steht ein Algorithmus zur Verfügung, mit dem die Menge der erschöpfenden Indexmengen systematisch erzeugt wird. Anstatt erst alle erschöpfenden Indexmengen zu bestimmen und danach die Testentscheidungen zu treffen, ist es u.a. aus Speicherplatzgründen einfacher und sinnvoll, direkt im Anschluß an die Bestimmung einer erschöpfenden Indexmenge die Entscheidung über die betroffenen Hypothesen zu treffen. Daher ist die Entscheidungsprozedur in den Algorithmus eingebunden.
Eine weitere Verkürzung liefert das Vorschalten der Holm-Prozedur, da Hypothesen, die durch die Holm-Prozedur verworfen werden, auch durch den AT verworfen werden; sie sind aher bei den Berechnungen nicht mehr zu betrachten. Die so modifizierte Prozedur heißt P4Bk; die Prozedur P4B erhält man durch Auslassen der vorgeschalteten Holm-Prozedur.

Prozedurbeschreibung P4Bk (vgl. Abb. 4.16) Gegeben seien die geordneten p-Werte $p_{(1)} \leq \ldots \leq p_{(l)}$ mit den zugehörigen Elementarhypothesen $H_{(1)}, \ldots, H_{(l)}$. Beginnend mit der Prozedur von Holm, werden die geordneten p-Werte $p_{(i)}$ mit den Schranken $s = \alpha/(l - i + 1)$ verglichen. Im Fall $p_{(i)} > s$ oder $i = l + 1$ stoppt die Holm-Prozedur. Ist $i = 1$ oder $i = l + 1$, so stoppt die gesamte Prozedur, und die Ablehnungsmenge R_{P4B} ist $\emptyset$ bzw. $\mathbb{N}_l$.
Ansonsten beginne man die eigentliche Prozedur mit $I_k = \{i\}$ und $A = \emptyset$, d.h. der leeren Akzeptanzmenge.
Nacheinander werden alle noch interessanten erschöpfenden Indexmengen $SE(I_k)$ erzeugt und $m = |SE(I_k)|$ sowie $i = \min I_k$ bestimmt. Ist $p_i > \alpha/m$, wird die Indexmenge zur Akzeptanzmenge hinzugefügt, und die nächste Indexmenge erzeugt. Ansonsten wird direkt die nächste Indexmenge generiert.
Sobald $|I_{k+1}| = 0$ ist, stoppt die Prozedur, und R_{P4B} ergibt sich zu $\mathbb{N}_l \setminus A$.
In dem Flußdiagramm wird bis zum Punkt **1.** das Holm-Verfahren durchgeführt. Der Teil bis **2.** generiert die nächste erschöpfende Indexmenge entsprechend dem Algorithmus aus Abbildung 4.5. Im letzten Teil wird der Bonferroni-Test für die gerade bestimmte erschöpfende Indexmenge durchgeführt.

4.3.3 Prozedur P4B (einfach)

Sei für k Gruppen mit l einseitigen (Elementar)Hypothesen $H_i = H_{\langle i_m, j_m \rangle}$, $m = 1, \ldots, l$, $i_m, j_m \in \mathbb{N}_k$ das (Abschlußtest)System der Hypothesen $\hbar = \{H_I : I \in \wp_l^0\}$ betrachtet. Um nun die Prozedur 4 durchführen zu können, ist ein Verfahren erforderlich, das zu einer Indexmenge die kleinste sie umfassende und erschöpfende Indexmenge bestimmt. Nach Definition 4.1 ist diese Menge dadurch gekennzeichnet, daß sie die maximal mögliche Anzahl an Elementarhypothesen umfaßt. D.h. für eine Indexmenge I sind die Elementarhypothesen zu bestimmen, die - außer den schon in ihr liegenden - in ihr enthalten sind. Dabei kann eine Hypothese nur dann zusätzlich enthalten sein, wenn sie durch die schon in I liegenden Hypothesen impliziert wird.
Eine Hypothese $H_{\langle j_1, j_2 \rangle}$ wird genau dann von einer Hypothesenmenge H_I impliziert, wenn sich aus ihr eine Hypothesenungleichung der Art $\mu_{j_1} = \mu_{i_1} \geq \mu_{i_2} \geq \ldots \geq \mu_{i_n} = \mu_{j_2}$ konstruieren läßt. Das bedeutet für die Bestimmung der Menge $SE(I)$, daß alle sich durch Ungleichungsketten von Hypothesen aus I konstruierbaren paarweisen Hypothesen zu bestimmen sind. Von den so konstruierten Hypothesen sind alle die zu I hinzuzunehmen, die ihrerseits auch Elementarhypothesen sind. Zur Bestimmung von $SE(I)$ haben wir somit den Algorithmus:

i) Setze: $I_l = \mathbb{N}_l$ Indexmenge aller Elementarhypothesen (Indexmenge der Globalhypothese).

ii) Sei $I \subset \mathbb{N}_l$ die Indexmenge, für die $SE(I)$ zu bestimmen ist. Nehme die zugehörigen Paare von Indizes $\langle i_m, j_m \rangle$ mit $m \in I$. $SE(I) = I$.

iii) Bilde alle möglichen Ungleichungsketten $\mu_{k_0} \geq \ldots \geq \mu_{k_p}$ mit $k_{r-1} = i_m, k_r = j_m$ wobei $m \in I$ ist (r=1, ... p). Dies ergibt die implizierten Hypothesen $H_{\langle k_o, k_p \rangle}$: $\mu_{k_o} \geq \mu_{k_p}$.

iv) Nimm alle $H_{\langle i_o, i_p \rangle}$ zu $SE(I)$ hinzu, die auch in I_l enthalten sind.

Zur Bestimmung aller möglichen implizierten Hypothesen aus iii) betrachte man den zu I gehörigen gerichteten Graphen. Nach Lemma 4.6 lassen sich aus der Erreichbarkeitsmatrix dieses Graphen alle existierenden Wege in dem Graphen ablesen und damit auch alle implizierten Hypothesen. Zur Implementierung der Bestimmung der Menge $SE(I)$ ergibt sich damit der Algorithmus:

i) Sei $I_l = \mathbb{N}_l$ die Indexmenge aller Elementarhypothesen und I die Indexmenge, zu der SE(I) zu bestimmen ist.

ii) Setze $U(I_l)$ als die Adjazenzmatrix zu I_l und $U(I)$ als die Adjazenzmatrix zu I.

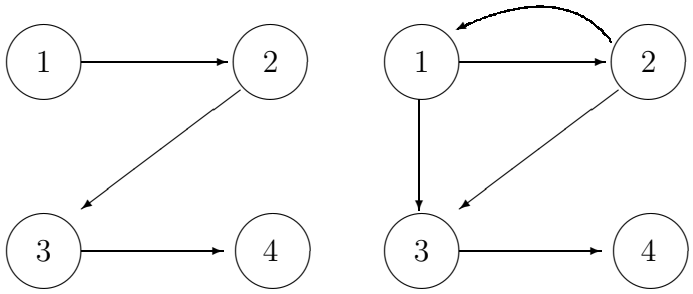
iii) Berechne $\dot{R}(U(I))$ und damit $U(SE(I)) = \dot{R}(U(I)) \times U(I_l)$

iv) $SE(I)$ besteht aus allen Hypothesen, für die eine 1 in $U(SE(I))$ vorkommt.

Beispiel Die Elementarhypothesen seien $H_1 = H_{\langle 1,3 \rangle} : \mu_1 \geq \mu_3$, $H_2 = H_{\langle 2,3 \rangle} : \mu_2 \geq \mu_3$, $H_3 = H_{\langle 3,4 \rangle} : \mu_3 \geq \mu_4$, $H_4 = H_{(1,2)} : \mu_1 = \mu_2$. Teile zuerst die zweiseitige Hypothese H_4 in zwei einseitige $H_5 = H_{\langle 1,2 \rangle} : \mu_1 \geq \mu_2$ und $H_6 = H_{\langle 2,1 \rangle} : \mu_2 \geq \mu_1$ auf. Also ist $k = 4$ und $l = 6$.
Gesucht sei $SE(I)$ zu $I = \{2, 3, 5\}$. Die Graphen und Adjazenzmatrizen haben so die Gestalt

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Adjazenzmatrix zu I Adjazenzmatrix zu H_0



Graph zu I Graph zu H_0

Damit ergibt sich $U(SE(I))$ zu:

$$U(SE(I)) = \Big(\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \Big) \times \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Durch die Summe werden anschaulich alle möglichen implizierten Hypothesen bestimmt. Durch die Multiplikation mit der Adjazenzmatrix der Globalhypothese werden dann die Hypothesen selektiert, die als Elementarhypothesen existieren. In diesem Fall impliziert I die Hypothesen: $\mu_1 \geq \mu_3$, $\mu_1 \geq \mu_4$ und $\mu_2 \geq \mu_4$, von denen nur $\mu_1 \geq \mu_3$ als einzige noch eine weitere Elementarhypothese ist.
Die Implementierung dieses Verfahrens in SAS/IML ist in Anhang 7 beschrieben.

4.3.4 Prozedur P4B (streng)

Nach Definition 4.2 ist eine Indexmenge $I = \{i_1, \ldots, i_m\}$ streng erschöpfend, wenn genau alle in ihr enthaltenen Elementarhypothesen wahr sein können. Was auch bedeutet, daß die restlichen Hypothesen $\mathbb{N}_l \setminus I$ falsch sind. Eine Elementarhypothese H_j ihrerseits kann nur dann nicht wahr sein, wenn sie im Widerspruch zu einem Schnitt H der restlichen $\mathbb{N}_l \setminus \{j\}$ Hypothesen und Alternativen steht. Dies ist genau dann der Fall, wenn solch ein Schnitt die Alternative von H_j impliziert. Um also zu prüfen, ob eine Indexmenge im strengen Sinn erschöpfend ist, muß für jede enthaltene Elementarhypothese geprüft werden, ob nicht ihre Alternative durch die restlichen Hypothesen impliziert

wird. Ein erster Algorithmus zur Bestimmung aller im strengen Sinne erschöpfenden Indexmengen ist folgender:

i) Erzeuge alle Mengen von $\wp_l^0$.

ii) Prüfe für jede Menge einzeln, ob sie widerspruchsfrei ist.

Dieses erste Verfahren ist extrem aufwendig, da in Schritt ii) für jede einzelne „wahre“ Hypothese die Widerspruchsfreiheit zu prüfen ist. Weil jede im strengen Sinn erschöpfende Indexmenge auch „normal“ erschöpfend ist, läßt sich das Verfahren dahingehend verkürzen, daß nur für erschöpfende Indexmengen die Widerspruchsfreiheit geprüft wird. D.h., der Algorithmus von Prozedur 4 muß im 2. Schritt (Abb. 4.16) dahingehend erweitert werden, daß für die Menge $SE(I_k)$ die Widerspruchsfreiheit zusätzlich geprüft wird. Eine starke Vereinfachung ist dann möglich, wenn der Graph der Globalhypothese zyklenfrei ist. Der Schlüssel zur Identifizierung von Widersprüchen ist hier das Vorhandensein von Zyklen in dem zu den Indexmengen gehörenden Graphen. Um dies zu zeigen, wird in diesem Abschnitt der zu der Indexmenge gehörende Graph erweitert.
Für den *erweiterten Graph* werden zur Pfeilmenge noch die Pfeile hinzugefügt, die den falschen Hypothesen $\mathbb{N}_l \setminus I$ entsprechen, wobei für die Hypothese $H_{\langle i,j\rangle}$ der der Alternative $A_{\langle j,i\rangle}$ entsprechende Pfeil von j nach i eingesetzt wird.
Man betrachte einen Zyklus $\langle j_1, j_2, \ldots, j_l, j_1\rangle$ in einem erweiterten Graphen. Entspricht in diesem Zyklus mindestens ein Pfeil einer Alternative, so bedeutet dies für die entsprechende Schnitthypothese, daß $\mu_{j_1} > \mu_{j_1}$ ist. D.h., diese Indexmenge ist nicht im strengen Sinn erschöpfend. Andererseits sei I nicht im strengen Sinn erschöpfend, demnach gibt es eine Elementarhypothese $H_i = H_{\langle i,j\rangle}$, die im Widerspruch zu dem Schnitt der restlichen Hypothesen $I \setminus \{i\}$ bzw. Alternativen $\mathbb{N}_l \setminus I$ steht. Da der Schnitt in disjunkte Hypothesenketten bezüglich der betroffenen Elementarhypothesen und Alternativen zerfällt, gibt es einen Schnitt der Form $H_{j,j_1}/A_{j_1,j} \cap H_{j_1,j_2}/A_{j_2,j_1} \cap \ldots \cap H_{j_l,i}/A_{i,j_l}$, wobei mindestens eine Alternative in diesem Schnitt vorkommt. Dem Schnitt entspricht der Weg $\langle j, j_1, \ldots, j_l, i\rangle$. Verlängert man diesen um die Hypothese/Pfeil $H_{\langle i,j\rangle}/\langle i, j\rangle$, so erhält man einen Zyklus, der mindestens einen Alternativenpfeil enthält. Damit ist der folgende Satz gezeigt:

Satz 4.11
Seien l einseitige, paarweise Elementarhypothesen gegeben mit der Indexmenge $\mathbb{N}_l$ und $I \subset \mathbb{N}_l$ eine beliebige Indexmenge. Sei $\tilde{G}$ der Graph, dessen Knoten den in $\mathbb{N}_l$ vorkommenden Erwartungswerten entsprechen, und dessen Kanten zum einen aus den in I

enthaltenen Hypothesen und zum anderen aus den $\mathbb{N}_l \setminus I$ entsprechenden Alternativen bestehen. Dann sind äquivalent:

1. I ist streng erschöpfend.

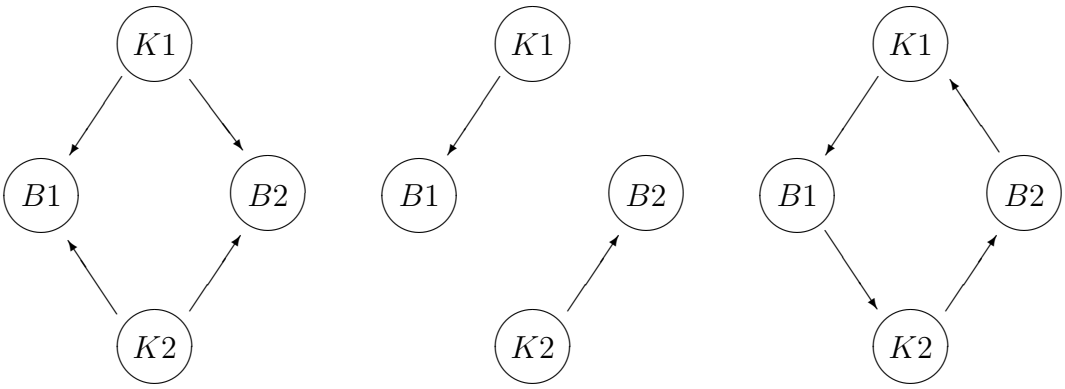
2. Es gibt einen Zyklus in $\tilde{G}$, in dem mindestens ein Alternativenpfeil vorkommt.

Beispiel Bei einem Vergleich zweier neuer Behandlungen $(B1, B2)$ gegen zwei bekannte $(K1, K2)$ sei von Interesse, ob die neuen Mittel besser sind als die herkömmlichen. In Hypothesen ausgedrückt heißt das

$$H_1 : \mu_{K1} \geq \mu_{B1},\ H_2 : \mu_{K1} \geq \mu_{B2},\ H_3 : \mu_{K2} \geq \mu_{B1},\ H_4 : \mu_{K2} \geq \mu_{B2}.$$

Man betrachte die Indexmenge $I = \{1, 4\}$. Diese ist erschöpfend, jedoch nicht streng erschöpfend, denn sind die Hypothesen 1 und 4 wahr und 2 und 3 falsch, so würde gelten:

$$\mu_{K1} \overset{H_1}{\geq} \mu_{B1} \overset{A_3}{>} \mu_{K2} \overset{H_4}{\geq} \mu_{B2} \overset{A_2}{>} \mu_{K1} \quad \text{(W!)}$$



Graph zu H_0 | Graph zu I | erweiterter Graph $\tilde{G}$ von I

Abbildung 4.6: Graphen zum Beispiel

Im erweiterten Graphen ist dies genau der Zyklus, der in Abbildung 4.6 im dritten Graphen zu sehen ist.

Ist nun die Globalhypothese zyklenfrei, so hat dies zur Folge, daß diese immer einen Alternativenpfeil beinhalten, falls in kleineren Indexmengen Zyklen auftreten. Denn gäbe es eine Indexmenge, die einen Zyklus nur aus Hypothesenpfeilen beinhaltet, so

würde dieser auch in der Globalhypothese enthalten sein. In diesem Fall ist die Bestimmung der streng erschöpfenden Indexmengen leicht unter Verwendung von Lemma 4.7 zu bewerkstelligen:

Algorithmus zur Bestimmung der streng erschöpfenden Indexmengen bei zyklenfreien Globalhypothesen

i) Sei $I_l = \mathbb{N}_l$ die Indexmenge aller Elementarhypothesen und I die Indexmenge, von der $SE(I)$ zu bestimmen ist.

ii) Setze $U(I_l)$ als die Adjazenzmatrix zu I_l und $U(I)$ als die Adjazenzmatrix zu I.

iii) Berechne $\dot{R}(U(I))$ und damit $U(SE(I)) = \dot{R}(U(I)) \times U(I_l)$.

iv) Bestimme $\dot{R}([U(I_l) - U(SE(I)]^T \oplus U(SE(I)))$; ist ein Diagonalelement = 1, so ist SE(I) nicht streng erschöpfend.

v) Ist $SE(I)$ streng erschöpfend, so besteht sie aus allen Hypothesen, für die eine 1 in $U(SE(I))$ vorkommt.

Enthält die Globalhypothese Zyklen, so ist bei beliebigen Schnitthypothesen nicht von vornherein klar, ob die auftretenden Zyklen auch Alternativenpfeile enthalten. Bei der Bestimmung der Zyklen ist es daher nötig zu wissen, ob ein Pfeil einer Hypothese oder einer Alternative entspricht. Dafür führe man als weitere Eintragung in die Adjazenzmatrix zu den Werten 0 für die Abwesenheit eines Pfeils, 1 für die Nullhypothese, den Wert a für die Alternative ein. Für diese drei Werte erweitere man dann die Abbildungen $\oplus$ und $\otimes$ zu

$\oplus$	0	1	a
0	0	1	a
1	1	1	a
a	a	a	a

$\otimes$	0	1	a
0	0	0	0
1	0	1	a
a	0	a	a

Damit erhalten wir analog zu Satz 4.5 für erweiterte Graphen:

Satz 4.12

Sei A eine Adjazenzmatrix eines erweiterten Graphen $\tilde{G}$ und $A^p = (a_{ij}^{(p)})$. Es ist $a_{ij}^{(p)} > 0$ genau dann, wenn es einen Weg der Länge p vom Knoten i zum Knoten j gibt. Ist $a_{ij}^{(p)} = 1$, so besteht der Weg nur aus Hypothesenpfeilen; ist $a_{ij}^{(p)} = a$, so ist mindestens ein Alternativenpfeil im Weg enthalten.

Beweis: Vollständige Induktion

Es gilt $a_{ij}^p = (a_{i1}^{(p-1)} \otimes a_{1j}) \oplus \ldots \oplus (a_{in}^{(p-1)} \otimes a_{nj})$.

$p = 2$: Für jeden Summanden $a_{ik} \otimes a_{kj}$ gilt:

$= 0 \iff$ mindestens einer der beiden Multiplikatoren ist $= 0$, d.h. es existiert kein direkter Weg von i nach j über k.

$= 1 \iff$ beide Multiplikatoren $= 1$, d.h. es existiert der Weg von i nach j über k, der aus zwei Hypothesenpfeilen besteht.

$= a \iff$ sonst, d.h. es gibt einen Weg von i nach j über k, der mindestens einen Alternativenpfeil enthält.

Damit ist

$a_{ij}^2 = 0 \iff$ Es existiert kein Weg der Länge 2 von i nach j.
$a_{ij}^2 = 1 \iff$ Es existiert mindestens ein Weg von i nach j der Länge 2; keiner der Wege beinhaltet einen Alternativenpfeil.
$a_{ij}^2 = a \iff$ Es existiert mindestens ein Weg der Länge 2 von i nach j, der mindestens einen Alternativenpfeil enthält.

$\Longrightarrow A^2$ enthält alle Wege der Länge 2.

$p \to p+1 : A^p$ enthält alle Wege der Länge p, mit der Argumentation von $p = 2$ folgt die Behauptung.

$\triangledown$

Lemma 4.13
Die Erreichbarkeitsmatrix $\dot{R}(\tilde{A})$ einer Adjazenzmatrix $\tilde{A}$ eines erweiterten Graphen mit k Knoten erhalten wir durch

$$\dot{R}(\tilde{A}) = \tilde{A} \oplus \tilde{A}^2 \oplus \ldots \oplus \tilde{A}^k.$$

Ist in $\dot{R}(\tilde{A})$ ein Diagonalelement $= a$, so enthält der erweiterte Graph mindestens einen Zyklus, der einen Alternativenpfeil beinhaltet.

Das liefert uns den

Algorithmus zur Bestimmung der streng erschöpfenden Indexmengen

i) Sei $I_l = \mathbb{N}_l$ die Indexmenge aller Elementarhypothesen und I die Indexmenge, von der $SE(I)$ zu bestimmen ist.

ii) Setze $U(I_l)$ als die Adjazenzmatrix zu I_l und $U(I)$ als die Adjazenzmatrix zu I.

iii) Berechne $\dot{R}(U(I))$ und damit $U(SE(I)) = \dot{R}(U(I)) \times U(I_l)$.

iv) Bestimme $\dot{R}(a \cdot [U(I_l) - U(SE(I)]^T \oplus U(SE(I)))$; ist ein Diagonalelement $= a$, so ist SE(I) nicht streng erschöpfend.

v) Ist $SE(I)$ streng erschöpfend, so besteht sie aus allen Hypothesen, für die eine 1 in $U(SE(I))$ vorkommt.

Beispiel (Fortsetzung) Neben den schon formulierten Hypothesen soll zudem noch gesichert werden, daß die zweite Dosis einen höheren Effekt als die erste hat, zusätzlich also noch die Hypothese $H_5 : \mu_{K1} \geq \mu_{K2}$ getestet wird. Wieder ist $I = \{1, 4\}$

erschöpfend.
Im Schritt ii) des Algorithmus ergeben sich die Matrizen:

$$U(I_l) = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad U(I) = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

In Schritt iii) erhalten wir $U(SE(I)) = U(I)$. Als nächstes wird:

$$\begin{aligned} & a \cdot (U(I_l) - U(SE(I))^T \oplus U(SE(I)) \\ = \; & a \left(\begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \right)^T \oplus \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ = \; & \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ a & 0 & a & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ = \; & \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & a & 0 & 0 \\ a & 0 & a & 0 \end{pmatrix} \end{aligned}$$

berechnet. Für $\dot{R}(\cdot)$ sind der zweite und dritte Summand

$$\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & a & 0 & 0 \\ a & 0 & a & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & a & 0 & 0 \\ a & 0 & a & 0 \end{pmatrix} = \begin{pmatrix} 0 & a & 0 & 0 \\ a & 0 & a & 0 \\ 0 & 0 & 0 & a \\ 0 & a & a & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & a & 0 & 0 \\ a & 0 & a & 0 \\ 0 & 0 & 0 & a \\ 0 & a & a & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & a & 0 & 0 \\ a & 0 & a & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & a \\ 0 & a & a & 0 \\ a & 0 & a & 0 \\ 0 & a & 0 & a \end{pmatrix}.$$

Damit sind in $\dot{R}(\cdot)$ Diagonalelemente $= a$. Also ist $SE(I)$ nicht streng erschöpfend.

4.3.5 Adjustierte p-Werte

Anstatt sich ein globales Niveau vorzugeben und dann „nur“ eine Entscheidung über die Ablehnung einer Hypothese zu erhalten, lassen sich auch adjustierte p-Werte mit diesem Verfahren bestimmen. Diese geben zu jeder Hypothese das kleinste multiple Niveau an, zu dem sich die Hypothese noch ablehnen läßt. Wright hat 1992 das generelle Konzept zur Konstruktion von adjustierten p-Werten für beliebige auf dem Abschlußtest beruhende Verfahren beschreiben.
Insbesondere lassen adjustierte p-Werte sich als deskriptives Maß für die „Plausibilität“ einzelner Hypothesen betrachten (vgl. Hommel, Bernhard 1991, Bernhard 1991). Als weitere Vorteile von adjustierten p-Werten schreiben Westfall und Young (1993) auf Seite 11:

> The reason for using adjusted p-values are the same as those for using ordinary p-values: (1) the statistican does not force a particular α upon the consumer of the report, and (2) cumbersome table look-ups are avoided. Table look-ups are usually difficult (if not impossible) in many Simultaneous Test Procedure applications, because of limited tables.

Definition 4.3 (Bernhard 1991)
Gegeben seien die p-Werte der Elementarhypothesen $H_1, \ldots, H_l$ und eine multiple Testprozedur.
Für $i = 1, \ldots, l$ heißt ap_i der zu der Hypothese H_i *adjustierte p-Wert*, falls ap_i das kleinste multiple Signifikanzniveau ist, für das die multiple Testprozedur die Hypothese H_i ablehnen kann.

Entsprechend den Ergebnissen aus Bernhard (1991) lassen sich zu Prozedur P4B adjustierte p-Werte wie folgt bestimmen:

Ermittlung adjustierter p-Werte zu P4B Man ordne die p-Werte der Größe nach $p_{(1)} \leq \ldots \leq p_{(l)}$ und setze $ap_{(i)} := 0$ für $i = 1, \ldots, l$.
Mit $i = 1$ beginnend erzeuge nacheinander alle streng erschöpfenden Indexmengen I mit $(i) \in I$ und $\{(1), \ldots, (i-1)\} \cap I = \emptyset$. Für jedes I setze man

$$ap_{(j)} := \max\{ap_{(j)}, |I| \cdot p_{(i)}\}, \quad \forall (j) \in I.$$

Nachdem für $i = l$ das Verfahren durchlaufen wurde, setze man alle $ap_j > 1$ zu 1. Dann enthält $(ap_{(i)})_{i=1,\ldots,l}$ alle adjustierten p-Werte der Hypothesen $H_{(i)}$ mit $i = 1, \ldots, l$.

Da hier $I = 1, \ldots, l$ streng erschöpfend ist, kann obiges Verfahren mit $i = 2$ und $(a_{(j)} = l \cdot p_{(1)}$ für $i = 1, \ldots, l$ beginnen. Daß dieses Verfahren korrekt die adjustierten p-Werte nach P4B bestimmt, ist in Satz 1.27 in Bernhard (1991) gezeigt. Durch die Form der Implementierung des Durchlaufs durch die Potenzmenge in P4B ist nur eine kleine Abwandlung des Flußdiagramms aus Abbildung 4.16 erforderlich. In dieser Implementierung werden nämlich sukzessive zuerst alle Indexmengen generiert, die die Hypothese $H_{(1)}$ enthalten, dann dienigen, die die Hypothese $H_{(2)}$ enthalten usw.. Zum einen entfällt die vorgeschaltete Holm-Prozedur, an deren Stelle die Initialisierung der adjustierten p-Werte tritt, und zum anderen tritt an die Stelle des Vergleichs $p_i > \alpha/m$ und der beiden Entscheidungen die Manipulation der adjustierten p-Werte. Eine Implementierung des Verfahrens in SAS/IML ist im Anhang 7 angegeben.

4.3.6 Beispiel

Anstelle des Vergleichs von Mittelwerten lassen sich unter anderem auch einseitige Vergleiche von Binomialwahrscheinlichkeiten mit diesem Verfahren betrachten. So ist zum Beispiel für eine Chemikalie von Interesse, in welchem Umfang sie genotoxischen Schaden in der Form, daß ganze Chromosomen verloren gehen oder hinzukommen, verursacht. Das Auftreten solch eines Chromosomen-Satzes wird als aneuploidie bezeichnet, wobei Zellen mit einem Verlust von Chromosomen hypodiploid und Zellen mit einer Zunahme an Chromosomen hyperdiploid heißen. In einem Dosis-Wirkungs-Versuch mit chinesischen Hamstern untersuchten Dulout und Natarajan (1987) die Wirkung von Diethylstilbestrol (DES) auf die Bildung von aneuploiden Zellen. Das Design bestand aus zwei verschiedenen Negativkontrollen und drei Dosisgruppen. Zum einen sollten die beiden Kontrollen auf Gleichwertigkeit getestet werden (Äquivalenznachweis). Zum andern interessierte die Versuchsansteller, welche der Dosierungen eine Erhöhung der Wahrscheinlichkeit des Auftretens von aneuploiden Zellen gegenüber einer ausgewälten Kontrolle bewirkt, und ob die Dosisinkremente eine Effektsteigerung bewirken. Aus dem Gesamtversuch, in dem zu mehreren Zeiten nach der Behandlung gemessen wurde, wird hier die Bildung hypodiploider Zellen zu dem empirisch am besten diskriminierenden Zeitpunkt herausgenommen. Die Versuchsergebnisse sind in Tabelle 4.3 zusammengefaßt.

Im folgenden wird nur die zweite Fragestellung, d.h. der Vergleich der verschiedenen Dosierungen untereinander und mit der Vehikelsubstanz als Kontrolle (nur TS) weiterverfolgt. Diese Fragestellung läßt sich durch folgende Elementarhypothesen abbilden, wobei der Index ts für nur Trägersubstanz und 5, 10, 15 für die jeweilige Dosisstufe

	Kontrollen		DES Dosen $[\mu g/ml]$		
	unbeh.	nur TS	5	10	25
aneuploide Zellen	13	6	24	19	26
Zellen gesamt	300	150	200	150	175
Mutationswahrscheinlichkeit q_i	0,043	0,040	0,120	0,126	0,148

Tabelle 4.3: Hypodiploider Effekt in Zellen von Chinesischen Hamstern, TS = Trägersubstanz

stehen:

$$\begin{aligned}
\text{Vergleich: Kontrolle vs. Dosis } 5\mu g/ml: &\quad H_1 = H_{\langle ts,5\rangle} : q_{ts} \geq q_5 \\
\text{Vergleich: Kontrolle vs. Dosis } 10\mu g/ml: &\quad H_2 = H_{\langle ts,10\rangle} : q_{ts} \geq q_{10} \\
\text{Vergleich: Kontrolle vs. Dosis } 15\mu g/ml: &\quad H_3 = H_{\langle ts,15\rangle} : q_{ts} \geq q_{15} \\
\text{Vergleich: Dosis } 5\mu g/ml \text{ vs. Dosis } 10\mu g/ml: &\quad H_4 = H_{\langle 5,10\rangle} : q_5 \geq q_{10} \\
\text{Vergleich: Dosis } 10\mu g/ml \text{ vs. Dosis } 15\mu g/ml: &\quad H_5 = H_{\langle 10,15\rangle} : q_{10} \geq q_{15}
\end{aligned}$$

Die betrachteten Hypothesen entsprechen genau denen, die sich bei einer simultanen Bestimmung von minimal effektiver Dosis und HEDS zu testen sind. In Tabelle 4.4 sind die p-Werte der einzelnen Vergleiche für Fischers exakten Test zusammengestellt. Auf eine Auswertung nach Hommel und Krummenauer (1998) wurde verzichtet, da die Erläuterung der Verfahren hier im Vordergrund steht. Für die Elementarhypothesen

	5	10	15
nur TS	0,0057	0,00543	$7,09 \cdot 10^{-4}$
5	-	0,488	-
10	-	-	0,342

Tabelle 4.4: p-Werte der einseitigen Vergleiche

erhalten wir mit der Prozedur P4B zu $\alpha = 5\%$ die Signifikanzen bzw. adjustierten p-Werte aus Tabelle 4.5. Die Testresultate lassen den Schluß zu, daß DES genotoxisch ist.

	5	10	15
nur TS	s; 0,01629	s; 0.01629	s; 0.00355
5	-	ns; 0,684	-
10	-	-	ns; 0,684

Tabelle 4.5: Signifikante Unterschiede (s, ns) und adjustierte p-Werte der Elementarhypothesen

Weiterhin legen die p-Werte der Vergleiche mit der Kontrolle nahe, die höchste Dosis zu verwenden, falls Mutationen erwünscht sind. Jedoch sind die Dosisschritte selbst nicht mehr signifikant, also kann eine Steigerung der Wirkung mit Eskalation der Dosis relativ zur nächst niedrigen Dosis nicht erreicht werden.
Um den Ablauf des Algorithmus zu veranschaulichen und die Unterschiede zwischen dem Hypothesensystem der Potenzmenge, dem aus erschöpfenden Indexmengen und dem auf streng erschöpfenden Indexmengen beispielhaft aufzuzeigen, ist in Abbildung 4.7 die Potenzmenge $\wp_5^0$ für die fünf Hypothesen dargestellt. Die streng erschöpfenden Indexmengen sind fett gedruckt und die Indexmengen, die zusätzlich nur erschöpfend sind, sind kursiv markiert. Anstelle der 31 Hypothesen der Potenzmenge reduziert sich

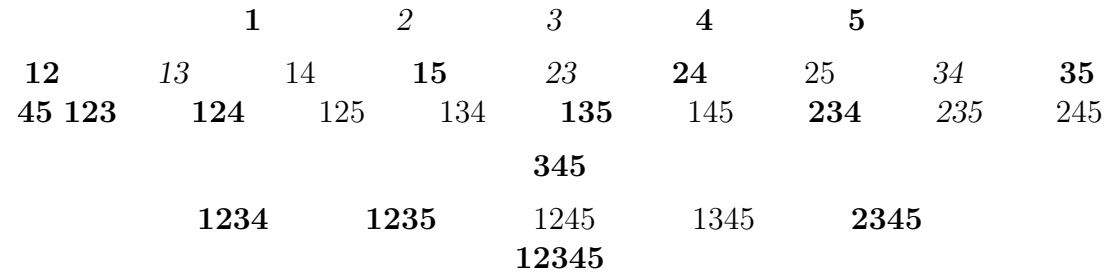


Abbildung 4.7: Potenzmenge, erschöpfende (kursiv) und streng erschöpfende Indexmengen (fett) bei 5 Gruppen

die Anzahl der zu testenden Hypothesen auf 17. Der Ablauf des Algorithmus aus Abbildung 4.16 ohne das vorgeschaltete Holm-Verfahren berechnet die streng erschöpfenden Indexmengen in folgender Folge:

k	j	I_k	K1	$SE(I_k)$	D	d	K2	J_j	streng e.	$\wp_{EI}$
1	1	{1}	-	{1}	$\emptyset$	5	-	$\emptyset$	ja	{1}
2	2	{1,2}	-	{1,2}	$\emptyset$	5	-	$\emptyset$	ja	{1,2}
3	3	{1,2,3}	-	{1,2,3}	$\emptyset$	5	-	$\emptyset$	ja	{1,2,3}
4	4	{1-4}	-	{1-4}	$\emptyset$	5	-	$\emptyset$	ja	{1-4}
5	5	{1-5}	-	{1-5}	$\emptyset$	5	-	$\emptyset$	ja	{1-5}
6	4	{1,2,3,5}	-	{1,2,3,5}	$\emptyset$	5	-	$\emptyset$	ja	{1,2,3,5}
7	3	{1,2,4}	-	{1,2,4}	$\emptyset$	5	-	$\emptyset$	ja	{1,2,4}
8	4	{1,2,4,5}	-	{1-5}	{3}	3	+			
9	3	{1,2,5}	-	{1,2,3,5}	{3}	3	+			
10	2	{1,3}	-	{1,3}	$\emptyset$	5	-	$\emptyset$	nein	
11	3	{1,3,4}	-	{1-4}	{2}	2	+			
⋮		⋮								
28	2	{4,5}	-	{4,5}	$\emptyset$	5	-	$\emptyset$	ja	{4,5}
29	1	{5}	-	{5}	$\emptyset$	5	-	$\emptyset$	ja	{5}

4.3.7 Anwendungen zur Bestimmung des HEDS

Um den höchsten effektiven Dosisschritt in einem Dosis-Wirkungs Versuch zu bestimmen, gibt es in der Literatur eine Vielzahl von Ansätzen. Diese werden im folgenden vorgestellt und teilweise erweitert. Neben diesen Verfahren läßt sich auch mit einer auf einseitigen paarweisen Vergleichen basierenden multiplen Testprozedur die HEDS bestimmen.
Die Verfahren zur HEDS-Bestimmung lassen sich grob unterteilen in Verfahren, die unter Ordnungsrestriktion gelten und Verfahren ohne Bedingungen an die Alternative. Die Beschränkung der Alternative rührt daher, daß bei Dosis-Wirkungs-Versuchen i.a. davon ausgegangen wird, daß ein Effekt mit steigender Dosierung zunimmt oder mindestens gleich bleibt. Eine weitere Unterteilung der Verfahren besteht darin, ob sie auf Zwei-Stichproben-Tests oder allgemeinen Kontrast-Tests basieren. Im Vergleich der Kontrast-Tests mit den Zwei-Stichproben-Tests haben die Kontrast-Tests den Vorteil eines gepoolten Varianzschätzers mit größerem Freiheitsgrad, was zu kleineren kritischen Werten führt. Sind jedoch die Varianzen der einzelnen Dosen inhomogen, so sind die Ergebnisse der Kontrast-Tests verfälscht, inbesondere wenn häufig mit steigender Dosis auch eine Varianz-Zunahme verbunden ist (Brandt 1996). Außerdem ermöglicht die Wahl von Zwei-Stichproben-Tests die Kontrolle des direktionalen Fehlers (Bauer 1994).
Bei einem Dosis-Wirkungs-Versuch ist teilweise nicht nur der HEDS von Interesse, sondern auch die minimal effektive Dosis (MED). Würde man beide getrennt aus den Daten eines Versuchs bestimmen, so müßte zur Wahrung des multiplen Niveaus eine α-Adjustierung der Form gemacht werden, daß beide Bestimmungen zum Niveau $\alpha/2$ erfolgen. Eine simultane Bestimmung von HEDS und MED ist zudem sinnvoll, weil es fragwürdig ist, einen HEDS zu bestimmen, wenn nicht zumindest eine MED existiert. Insbesondere ist es in Designs mit einer Kontrolle meist schwerer, die Signifikanz der einzelnen Dosisschritte zu zeigen, als einige der Unterschiede von höheren Dosen zur Kontrolle.
Im folgenden seien k verschiedene ansteigende Dosen mit den zugehörigen interessierenden Parametern $\mu_1, \ldots, \mu_k$ betrachtet. Typischerweise handelt es sich hierbei um die Erwartungswerte des gemessenen Effekts. Zur Bestimmung der MED interessieren die Vergleiche jeder Dosis mit der Kontrolle, hier Dosis 1. Die naheliegenden Hypothesen haben somit die Gestalt

$$\text{(MED)} \qquad H_{Mi}: \quad \mu_1 \geq \mu_i, \quad i = 2, \ldots, k. \tag{4.1}$$

Die p-Werte des paarweisen einseitigen Tests für die Hypothese H_{Mi} sei mit p_{Mi} bezeichnet. Für die Ermittlung des HEDS sind die Hypothesen

$$\text{(HEDS)} \qquad H_{Hi}: \quad \mu_i \geq \mu_{i+1}, \quad i = 1, \ldots, k-1 \tag{4.2}$$

von Interesse, wobei die p-Werte mit p_{Hi} bezeichnet seien. Die beiden Hypothesensysteme sind bis auf die „erste“ identische Hypothese H_{M2} beziehungsweise H_{H1} disjunkt.

4.3.7.1 Verfahren, die nur den HEDS bestimmen

Parallel zu jedem der beschriebenen Verfahren wird an einem Beispiel das jeweilige Verfahren illustriert. Der Beispielversuch besteht aus einer Kontrolle (1) und drei Dosisgruppen ($k = 4$). Die p-Werte aus einseitigen t-Tests für die HEDS-Hypothesen seien $p_{H1} = 0,04$, $p_{H2} = 0,015$ und $p_{H3} = 0,02$. Getestet wird zu einem multiplen Niveau von $\alpha = 5\%$.

4.3.7.1.1 Verfahren ohne Ordnungsrestriktion

Adjustierte p-Werte Die einfachste Möglichkeit, das multiple Niveau zu kontrollieren, ist, wie bei Lee und Spurrier (1995) zu finden, die p-Werte p_{Hi} mit der Bonferronischranke $\alpha/(k-1)$ zu vergleichen. Der höchste signifikante Dosisschritt ist dann der HEDS. Eine Verbesserung bringt das Holm-Verfahren, das die p-Werte der Größe nach ordnet. Die so geordneten p-Werte $p_{M(i)}$ werden dann vom kleinsten aufwärts mit wachsenden α-Schranken $\alpha/(k-i)$ verglichen, bis der erste Vergleich nicht mehr signifikant ist. Der höchste der signifikanten Dosisschritte ist dann der HEDS. Eine Adjustierung anhand der multivariaten t-Verteilung ist theoretisch trennschärfer. Der Powergewinn ist aufgrund der Struktur der Korrelationsmatrix jedoch nur marginal (vgl. Hommel 1989, Lee, Spurrier 1995).
Im Beispiel bedeutet das für die α-Adjustierung nach Bonferroni oder Holm, die p-Werte der Hypothesen aus (4.1) $p_{H1} = 0,04, p_{H2} = 0,015$ und $p_{H3} = 0,02$ zu verwenden. Bei der Bonferroni-Adjustierung werden alle p-Werte mit $\alpha/3 = 0,016$ verglichen, demnach wird nur die Hypothese H_{H2} abgelehnt, und der Dosisschritt von 2 nach 3 wird als der HEDS erkannt. Beim Holm-Verfahren wird der kleinste p-Wert $p_{H2} = 0,015$ mit $\alpha/3 = 0,016$ verglichen, dieser ist signifikant. Damit wird der zweitkleinste p-Wert $p_{H3} = 0,02$ mit $\alpha/2 = 0,025$ verglichen, welcher ebenfalls signifikant ist. Jetzt wäre noch p_{H1} mit α zu vergleichen; da der höchste Dosisschritt schon signifikant ist, interessiert das Ergebnis an dieser Stelle nicht mehr. Der HEDS ist hierbei der Schritt von Dosis 3 nach 4.

Diese Form der „Abkürzung“ wird eine weitere Verbesserung der α-Schranken ermöglichen, was durch eine andere Modellierung der Fragestellung mit speziellen Hypothesen erreicht wird.

HEDS mit Bauer/Budde Hypothesen Anstelle der Hypothesen in (4.2) werden hierbei alternativ die folgenden formuliert

$$\tilde{H}_{Hi}: \quad \mu_i \geq \ldots \geq \mu_k, \quad i = 1, \ldots, k-1 \tag{4.3}$$

mit den zugehörigen Alternativen

$$\tilde{A}_{Hi}: \exists\, i \leq j \leq k-1: \ \mu_j < \mu_{j+1}, \quad i = 1, \ldots, k-1.$$

Die Idee der folgenden Prozedur geht auf die modifizierte Shaffer-Prozedur (Shaffer 1986) zurück.
Die Prozedur verwendet die p-Werte der Paarvergleiche zu den Hypothesen aus (4.2), da sich nach dem union-intersection-Prinzip jedes der $\tilde{H}_{Hi}$ dadurch testen läßt, daß die Hypothesen $H_{Hj},\ j \geq i$ betrachtet werden. Im ersten Schritt wird das Minimum aller p-Werte $p_{Hj} = \min_1^{k-1} p_{Hi}$ bestimmt. Ist $p_{Hj} \leq \alpha/(k-1)$, so werden $\tilde{H}_{Hj}$ und alle $\tilde{H}_{Hi}$ mit $i < j$ verworfen. Andernfalls wird keine Nullhypothese verworfen. In jedem weiteren Schritt verbleiben noch die Hypothesen $\tilde{H}_{Hj+1}, \ldots, \tilde{H}_{Hk-1}$ zu testen. Ist nun $p_{Hl} = \min_{j+1}^{k-1} p_{Hi} \leq \alpha/(k-j-1)$, so verwerfe man die Hypothesen $\tilde{H}_{Hi}$ mit $j < i \leq l$. Dies führe man solange fort, bis eine Hypothese nicht mehr verworfen werden kann. Ist die Hypothese $\tilde{H}_l$ die letzte noch verworfene, so ist der höchste effektive Dosisschritt der Schritt von Dosis l zu Dosis $l+1$. Daß dieses Verfahren das multiple Niveau bezüglich der Hypothesen (4.3) kontrolliert, ist in Bauer und Budde (1994) gezeigt. Der Beweis erfolgt über die Konstruktion eines Abschlußtests.
Für das Beispiel lauten die Hypothesen

$$\tilde{H}_{M1}: \mu_1 \geq \mu_2 \geq \mu_3 \geq \mu_4, \quad \tilde{H}_{M2}: \mu_2 \geq \mu_3 \geq \mu_4, \quad \tilde{H}_{M3}: \mu_3 \geq \mu_4.$$

Im ersten Schritt ist das Minimum aller p-Werte $p_{H2} = 0,015 < 0,016$. Somit werden H_{H1} und H_{H2} verworfen, und im nächsten Schritt verbleibt nur noch p_{H3}, welches mit α verglichen wird. Bei diesem Verfahren wird der letzte Dosisschritt als HEDS erkannt. Analog schlagen die Autoren ein Verfahren zur Bestimmung der MED vor, das hier auch beschrieben wird, da es in eines der neuen Verfahren Eingang findet:
Anstelle der Hypothesen in (4.1) werden die folgenden formuliert:

$$\tilde{H}_{Mi}: \quad \mu_1 \geq \ldots \geq \mu_i, \quad i = 2, \ldots, k \tag{4.4}$$

mit den Alternativen

$$\tilde{A}_{Hi} : \exists\, 1 < j \leq i :\ \mu_1 < \mu_j, \quad i = 2, \ldots, k.$$

Auch diese Prozedur verwendet wieder p-Werte aus Paarvergleichen, hier die zu den Hypothesen aus (4.1). Ebenfalls nach dem union-intersection- Prinzip läßt sich jede der $\tilde{H}_{Mi}$ dadurch testen, daß die Hypothesen $H_{Mj},\ j \geq i$ betrachtet werden. Im ersten Schritt wird das Minimum aller p-Werte $p_{Mj} = \min_2^k p_{Mi}$ bestimmt. Ist $p_{Mj} \leq \alpha/(k-1)$, so werden $\tilde{H}_{Mj}$ und alle $\tilde{H}_{Mi}$ mit $i > j$ verworfen. Andernfalls wird keine Nullhypothese verworfen. In jedem weiteren Schritt verbleiben noch die Hypothesen $\tilde{H}_{M2}, \ldots, \tilde{H}_{Hj-1}$ zu testen. Ist nun $p_{Ml} = \min_2^{j-1} p_{Mi} \leq \alpha/(k-j-1)$, so verwerfe man die Hypothesen $\tilde{H}_{Mi}$ mit $l \leq i < j$. Dies führe man solange fort, bis eine Hypothese nicht mehr verworfen werden kann. Auch hier ließe sich anstelle der Bonferronisierung eine multivariate t-Verteilung verwenden, was mit dem gleichen Argument wie oben keinen wesentlichen Powergewinn bringt.
In ihrer Arbeit fordern Bauer und Budde die Monotonie in der Alternative der einzelnen Hypothesen, was für den Fall, daß nur paarweise Tests verwendet werden, nicht gefordert werden muß. Auf den Originalvorschlag zum Testen der Hypothesen aus (4.4) wird als nächstes eingegangen, da dabei eine monotone Ordnung der Erwartungswerte in der Alternative unterstellt wird.

4.3.7.2 Verfahren mit Ordnungsrestriktion

Der Begriff „Ordnungsrestriktion“ wird hier so aufgefaßt, daß den Erwartungswerten der einzelnen Behandlungen eine gewisse Ordnung unterstellt wird. Es wird gefordert, daß die Effekte mit steigender Dosis ansteigen, mindestens jedoch gleich bleiben. Für eine Kontrolle (1) und drei aufsteigende Dosen (2,3,4) bedeutet das

$$\mu_1 \leq \mu_2 \leq \mu_3 \leq \mu_4.$$

HEDS nach Bauer/Budde Anstelle der p-Werte der paarweisen Hypothesen (4.1) verwenden Bauer und Budde (1994) p-Werte von Reverse-Helmert- Kontrasten. Dabei wird wie folgt vorgegangen: Zum Testen jeder der Hypothesen

$$\tilde{H}_{Hi} :\ \mu_i \geq \ldots \geq \mu_k$$

wird der p-Wert $\tilde{p}_{Hi}$ des Kontrasts

$$\tilde{c}_{Hi} = (\underbrace{0, \ldots, 0}_{i-1}, -(k-i), \underbrace{1, \ldots, 1}_{k-1})$$

verwendet. Mit diesen p-Werten wird nun das Verfahren aus dem Abschnitt *HEDS mit Bauer/Budde Hypothesen* durchgeführt.
Für das Beispiel mit vier Gruppen lauten die Reverse-Helmert-Kontraste demnach

$$\begin{aligned} \tilde{c}_{H1} &= (-3,1,1,1), \\ \tilde{c}_{H2} &= (0,-2,1,1), \\ \tilde{c}_{H3} &= (0,0,-1,1). \end{aligned}$$

Anstatt in diesem Verfahren in jedem Schritt die Bonferronischranke zu benutzen, ließe sich auch die multivariate t-Verteilung verwenden, was jedoch bei diesen Kontrasten aufgrund ihrer Unkorreliertheit keinen Vorteil bringen würde.

Die sechs folgenden Verfahren testen das durch die Hypothesen (4.4) aufgestellte Abschlußtest-System und sind daher Niveau-α-Verfahren, d.h. auf jeder Stufe kann zum Niveau α getestet werden.

Likelihood-Ratio-Test Zum Testen der Hypothesen $\tilde{H}_{Hi}$ läßt sich auch der Likelihood-Ratio-Test verwenden, der von Bartholomew (1959) zum Testen von Trendhypothesen eingeführt wurde. Eine ausführliche Beschreibung des Tests und seiner Verteilung findet sich in Bretz (1999) oder Robertson et al (1988).

Test nach Rom et al (1994) Sie betrachten nur Hypothesen der Form $H_{i,j}: \mu_i = \mu_{i+1} = \ldots = \mu_j$. Sequentiell wird zuerst die Homogenitätshypothese für alle k Mittelwerte getestet, dann die zwei Hypothesen für $k-1$ aufeinanderfolgende Mittelwerte, dann die drei Hypothesen für $k-2$ aufeinanderfolgende Mittelwerte usw. , wobei jede Hypothese mit einem linearen Kontrast getestet wird. Eine Hypothese wird zum Niveau α verworfen unter den Bedingungen, daß

i) jede Hypothese, die nicht zum Niveau α verworfen wurde, beibehalten wird, und

ii) jede Hypothese, die von beibehaltenen Hypothesen impliziert wird, ebenfalls beibehalten wird, und

iii) jede der Hypothesen H_{i_1,j_1} beibehalten wird, für die gilt, daß H_{i_2,j_2} beibehalten wurde und $i_1 < j_1 < i_2 < j_2$ ist.

Anstatt die einzelnen Trendhypothesen mit linearen Kontrasten zu testen, lassen sich diese auch mit multiplen Kontrasten, wie zum Beispiel den isotonischen Kontrasten oder dem Likelihood-Ratio-Test testen, da der lineare Kontrast quasi-lineare Dosis-Wirkungs-Verläufe annimmt.

4.3.7.3 Verfahren, die zusätzlich MED-Hypothesen betrachten

Wie anfangs erwähnt, kann es von Interesse sein, gleichzeitig MED und HEDS im selben Versuch zu bestimmen. Auch wenn nur der HEDS gesucht wird, kann es trotzdem vorteilhaft sein, die MED Hypothesen zusätzlich zu betrachten, da diese insbesondere bei einem monotonen Dosis-Wirkungseffekt „leichter“ zu zeigen sind. In den Verfahren betrachtet man dann die dementsprechenden Elementarhypothesen nicht mehr. Testet man den HEDS mit dem einseitigen Studentized-Range-Test nach Hayter (1990), so werden die MED Hypothesen automatisch mitgetestet.

4.3.7.3.1 Verfahren ohne Ordnungsrestriktion

One-sided Studentized-Range-Test Dieser Test ist ein simultaner, einseitiger Allpaar-Vergleich in der Form, daß die als Gruppe 1 ausgezeichnete Dosis mit allen weiteren Dosen einseitig auf Anstieg getestet wird. Die Gruppe 2 wird mit allen höheren Gruppen verglichen, usw. Bei der Durchführung des Tests ist daher darauf zu achten, wie die Gruppen/Dosen zu ordnen sind, damit auch die interessierenden Hypothesen getestet werden.

MED und HEDS im Abschlußtest Betrachten wir zu den HEDS Hypothesen die MED Hypothesen mit, so läßt sich ein Abschlußtest mit den Hypothesen (4.1) und (4.2) als Elementarhypothesen konstruieren. Werden die redundanten Hypothesen nicht ausgeschlossen, so ist schon für $k = 4$ das Hypothesensystem nicht mehr übersichtlich darstellbar, während das „verschärfte“ Hypothesensystem weniger Schnitt- und Partitionshypothesen insgesamt sieben Hypothesen (davon zwei Elementarhypothesen) weniger enthält. Mit den Bezeichnungen: a $\sim \ \mu_1 \geq \mu_2$, b $\sim \ \mu_1 \geq \mu_3$, c $\sim \ \mu_1 \geq \mu_4$, d $\sim \ \mu_2 \geq \mu_3$, e $\sim \ \mu_3 \geq \mu_4$ ergeben sich die Hypothesen aus Tabelle 4.6: In der gleichen Weise läßt sich aus den Bauer/Budde-Hypothesen (4.4) und (4.3) ein AT-System konstruieren.

MED und HEDS mit Bauer/Budde-Hypothesen Bei der Bildung des AT-Systems treten zusätzlich Partitionshypothesen auf. Diese setzen sich immer aus einer MED und einer HEDS Hypothese dergestalt zusammen:

$$\mu_1 \geq \ldots \geq \mu_i \ \wedge \ \mu_j \geq \ldots \geq \mu_k, \quad i < j.$$

Testet man nun wieder mit den paarweisen Tests, so hat für $k = 4$ das AT-System die Form aus Abbildung 4.9. An den Hypothesen steht zusätzlich, mit welchem Niveau der

a	ab	abc	abcd	abcde
b	**ac**	abd	abce	
c	ae	ace	bcde	
d	**bc**	bcd		
e	bd	**bce**		
	be	cde		
	cd			
	ce			
	de			

Tabelle 4.6: Abschlußtest für MED und HEDS simultan, $k = 4$.
Fett gedruckt sind die redundanten Hypothesen.

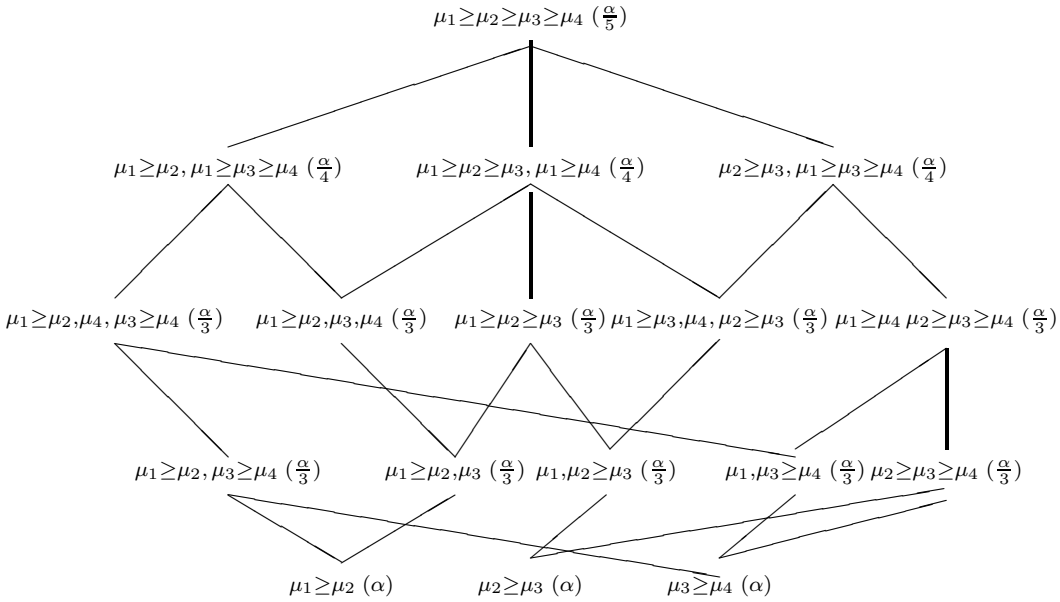


Abbildung 4.8: Abschlußtest, $k = 4$, in Klammern steht das lokale Niveau, zu dem der minimale p-Wert getestet wird.

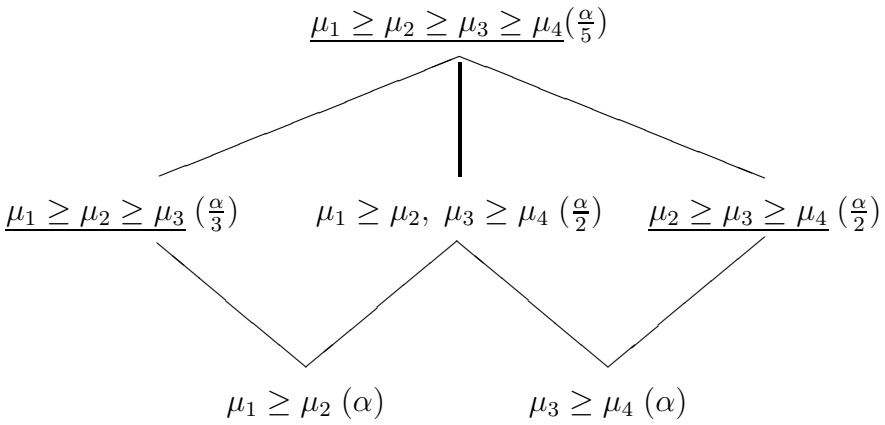


Abbildung 4.9: Abschlußtest mit den Bauer/Budde Hypothesen, $k = 4$, in Klammern steht das lokale Niveau, zu dem der minimale p-Wert getestet wird.

minimale p-Wert der enthaltenen Elementarhypothesen verglichen werden muß. Dabei liegt der HEDS genau dann bei dem Schritt $2 \rightarrow 3$, wenn die unterstrichenen Hypothesen ablehnt werden, und der Test für die Hypothese $\mu_3 \geq \mu_4$ nicht ablehnt. Für $k = 5$ ergibt sich das Hypothesensystem aus Abbildung 4.10. Hier müssen die unterstrichenen

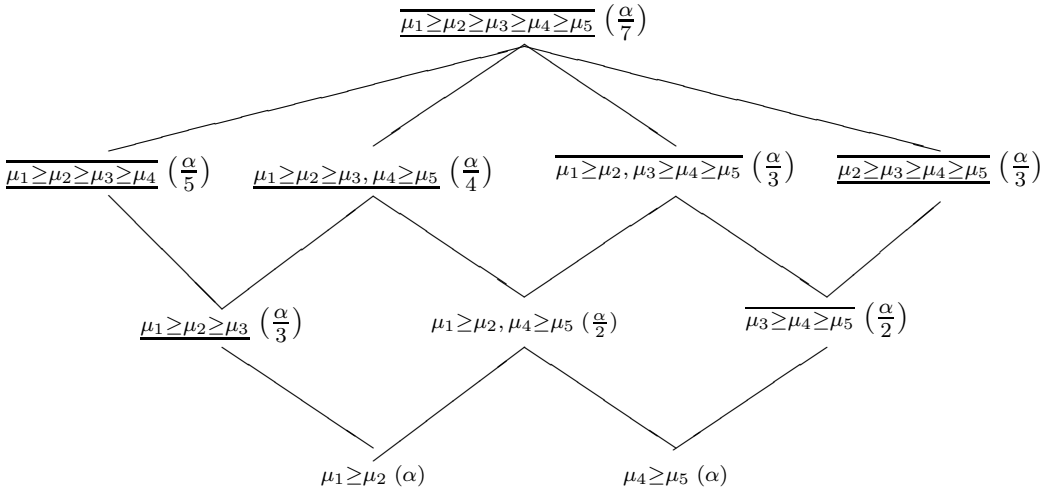


Abbildung 4.10: Abschlußtest mit den Bauer/Budde-Hypothesen, $k = 5$, in Klammern steht das lokale Niveau, zu dem der minimale p-Wert getestet wird.

Hypothesen abgelehnt werden, damit der HEDS bei $2 \rightarrow 3$ liegt, und die Hypothesen $\mu_3 \geq \mu_4 \geq \mu_5$ und $\mu_4 \geq \mu_5$ dürfen nicht abgelehnt werden. Für den HEDS bei $3 \rightarrow 4$ müssen die überstrichenen Hypothesen abgelehnt werden, und die Hypothese $\mu_4 \geq \mu_5$ darf nicht verworfen werden.
Anstatt dieses komplexe System zu verwenden, was vor allem ohne Rechnerunterstützung in der Praxis zu undurchsichtig ist, wandele man das Bauer/Budde-Verfahren zur HEDS-Bestimmung nur auf der ersten Stufe ab. Anstatt beim Globaltest nur das Minimum der p-Werte der HEDS Hypothesen (4.2) zu bilden, nehme man auch noch die p-Werte der MED-Hypothesen (4.1) hinzu. Dieses Minimum wird dann zur Bonferronischranke $\alpha/(2k-3)$ getestet. Liegt hier Signifikanz vor, so verfahre man weiter wie beim Ausgangsverfahren von Seite 105.
Die drei in diesem Abschnitt beschriebenen Abschußtestverfahren vereinfachen sich, wenn der Alternative eine Ordnung unterstellt wird.

4.3.7.3.2 Verfahren unter Ordnungsrestriktion Unter Ordnungsrestriktion vereinfacht sich einerseits das AT-System, und zum anderen lassen sich einige Hypothesen zu höheren α-Schranken testen. Im Fall von $k = 4$ vereinfacht sich das System aus Abbildung 4.8 zu dem in Abbildung 4.11 dargestellten. Es wird insbesondere

die Globalhypothese zum vollen α -Niveau getestet. Sind die beiden größten Sprünge

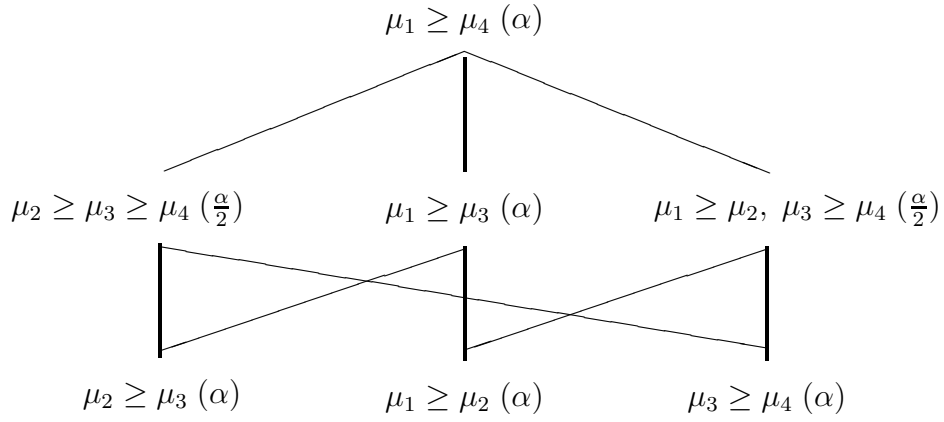


Abbildung 4.11: Abschlußtestsystem bei geordneten Alternativen, $k = 4$, in Klammern steht das lokale Niveau, zu dem der minimale p-Wert getestet wird.

$(\mu_1 \leftrightarrow \mu_4,\ \mu_1 \leftrightarrow \mu_3)$ überhaupt signifikant, so werden in diesem System die drei HEDS-Hypothesen zu maximal $\alpha/2$ anstelle zu $\alpha/3$ getestet, wie es bei Bonferronisierung oder beim Holm-Verfahren der Fall wäre. Dies ist zum Beispiel der Fall, wenn bei gleichen Fallzahlen und gepoolter Varianz der Effekt mit der Dosis ansteigt. Hat zudem die Hypothese H_{H3} den kleinsten p-Wert unter den HEDS-Hypothesen, so würden die anderen beiden Dosisschritte je zu α getestet werden, selbst wenn dies in der Praxis nicht unbedingt von Interesse ist, da die HEDS schon gefunden ist.

Im Fall von 5 Dosisstufen erhält man das AT-System aus Abbildung 4.12. Auch hier zeigt sich, daß die erste HEDS zu $\alpha/3$ und nicht zu $\alpha/4$ getestet wird, wenn der Sprung $(\mu_1 \leftrightarrow \mu_5)$ zum Niveau α signifikant ist.
Beim abgewandelten Bauer/Budde-Verfahren ändert sich nur, daß die Globalhypothese der „letzten“ MED-Hypothese H_{Mk} zu α getestet wird.

4.3.7.4 Vergleich der verschiedenen Verfahren

Die Power der richtigen HEDS-Entscheidung wird für vier Gruppen durch eine Simulationsstudie für alle beschriebenen Verfahren bestimmt. Alle möglichen Erwartungswertprofile, die zu den drei HEDS-Entscheidungen führen können, sind in Abbildung 4.13 aufgelistet, wobei in den Profilen g,h,j,k,l,n,o,p,q noch die Größe der Sprünge variieren kann, so daß die p-Werte verschieden geordnet sein können. Die unterschiedlichen p-Wert-Ordnungen sind in Tabelle 4.7 angegeben. Bei den Simulationen ist der zu entdeckende HEDS bei der Fallzahl 10 zu 1.25 gewählt, was einer Power von 73% bei einem

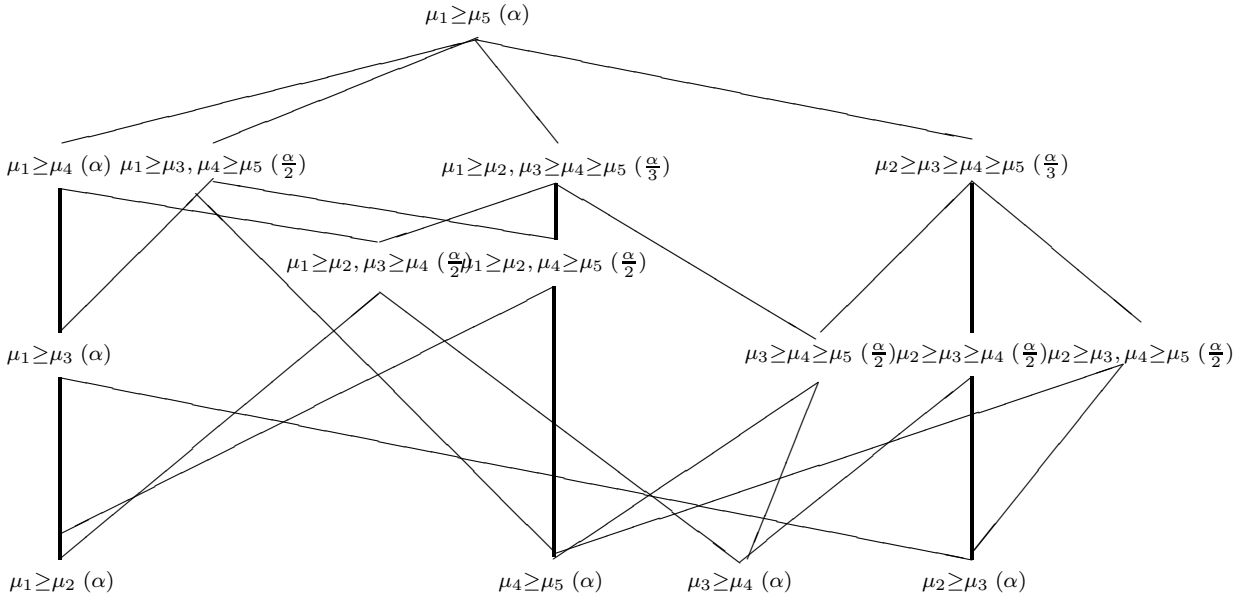


Abbildung 4.12: Abschlußtestsystem bei geordneten Alternativen, $k = 5$, in Klammern steht das lokale Niveau, zu dem der minimale p-Wert getestet wird.

Profil	
g)	$p_{12} < p_{23}$, $p_{23} < p_{12}$
h)	$p_{12} < p_{23} < p_{14}$, $p_{23} < p_{12} < p_{14}$, $p_{12} < p_{14} < p_{23}$, $p_{23} < p_{14} < p_{12}$ $p_{14} < p_{12} < p_{23}$, $p_{14} < p_{23} < p_{12}$
j,l,o)	$p_{12} < p_{34}$, $p_{34} < p_{12}$
p)	$p_{12} < p_{23} < p_{34}$, $p_{23} < p_{12} < p_{34}$, $p_{12} < p_{34} < p_{23}$, $p_{23} < p_{34} < p_{12}$ $p_{34} < p_{12} < p_{23}$, $p_{34} < p_{23} < p_{12}$
k,n,q)	$p_{23} < p_{34}$, $p_{34} < p_{23}$

Tabelle 4.7: Mögliche Anordnungen für die p-Werte der paarweisen Vergleiche für die Profile aus Abb. 4.13, wobei mit p_{ij} der p-Wert eines beliebigen Tests für den Vergleich der beiden Gruppen i und j bezeichnet ist

einzelnen Test entspricht. Die in den Simulationen gewählten Mittelwertkonstellationen sind im Anhang A.1 aufgelistet, ebenso die vollständigen simulierten Wahrscheinlichkeiten für die richtige Schätzung, wie auch die Über- oder Unterschätzung des HEDS. Ein Großteil der hier beschriebenen Verfahren trifft seine Entscheidungen basierend auf den p-Werten von Paarvergleichen. Hierbei ist zu beachten, daß die Ordnung der p-Werte nur dann Rückschlüsse auf die Ordnung der Erwartungswerte zuläßt, wenn die Fallzahl in allen Gruppen gleich ist und ein gemeinsamer Varianzschätzer verwendet wird. Unter Ordnungsrestriktion sieht man, daß bei allen Profilen der Test von Bauer und Budde die höchste Power besitzt. Dagegen sind die beiden Abschlußtests mit Powerverlusten zwischen 5 und 20% deutlich schlechter. Diese Unterlegenheit ist auf die Empfindlichkeit des Globaltests zurückzuführen, der nur die Hypothese $\mu_1 \geq \mu_4$ testet,

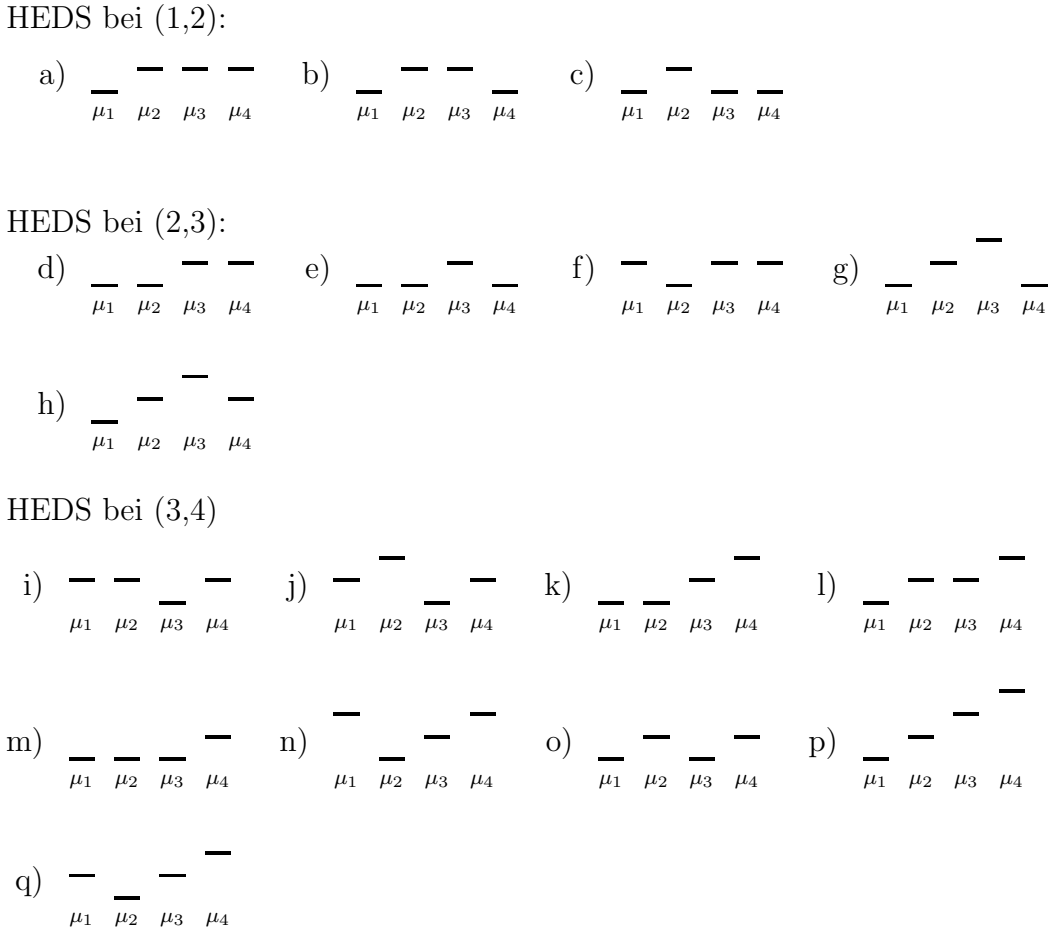


Abbildung 4.13: Erwartungswertprofile, $k = 4$

was sich darin zeigt,daß die beiden AT-Varianten ohne Ordnungsrestriktion in fast allen Fällen eine höhere Power als die mit Ordungsrestriktion haben. Insbesondere fällt auf, daß der von Rom et al (1994) vorgeschlagene Test bei dem Stepprofil $\mu_1 < \mu_2 = \mu_3 = \mu_4$ total zusammenbricht, was aus dem Test für die Hypothese $\mu_2 = \mu_3 = \mu_4$ zu sehen ist. Da hier der lineare Kontrasttest nicht ablehnt, kann aufgrund der angegebenen Bedingungen kein effektiver Dosisschritt mehr gefunden werden. Für alle Verfahren ist die Wahrscheinlichkeit, den HEDS zu unterschätzen, falls dies aufgrund des Erwartungswertprofils möglich ist zwischen 10% und 20% teilweise sogar noch höher. Bei Profilen, in denen die Mittelwerte nicht monoton in der Dosis sind, haben sowohl die beiden AT-Varianten als auch das Verfahren von Rom einen totalen Powerzusammenbruch. Bei den Profilen k und l, die durchaus in der Praxis realistisch sind, haben alle Verfahren eine über 10% liegende Chance, den HEDS zu unterschätzen.
Für die Tests ohne Ordnungsrestriktion fällt auf, daß der Hayter-Test die geringste

Power hat, wohingegen das abgewandelte Verfahren von Bauer und Budde die höchste Power aufweist. Zusammenfassend läßt sich sagen, daß unter Ordnungsrestriktion der Test von Bauer und Budde zu empfehlen ist und bei beliebiger Alternative das abgewandelte Bauer/Budde-Verfahren, wobei es bei allen Verfahren Situationen gibt, in denen die Chance der Überschätzung des wahren HEDS über 20% liegt.

4.3.8 Einseitige verschobene Hypothesen

Bis jetzt wurden nur einseitige Nullhypothesen der Form $H_{\langle i,j\rangle} : \mu_i \geq \mu_j$ betrachtet. Im folgenden wird die entwickelte Methodik auf Nullhpothesen der Form $H_{[i,j]} : \mu_i \geq \mu_j + \delta$ erweitert. D.h., es werden Nullhypothesen der Form $H_{\langle i,j\rangle} : \mu_i \geq \mu_j$ und $H_{[i,j]} : \mu_i \geq \mu_j + \delta$ zugelassen. Hierbei ist gefordert, daß der Verschiebeparameter δ für alle Hypothesen gleich und, oBdA > 1 ist. Somit lassen sich insbesondere Systeme paarweiser Hypothesen, die Äquivalenzhypothesen beinhalten, abbilden. Entsprechend der Vorgehensweise in den Abschnitten 4.3.3 und 4.3.4 wird zuerst gezeigt, wie für eine beliebige Indexmenge ihre kleinste sie umfassende erschöpfende Indexmenge bestimmt werden kann. In einem zweiten Schritt wird dann gezeigt, wie sich für eine erschöpfende Indexmenge überprüfen läßt, ob sie streng erschöpfend ist. Eine weitere technische Forderung ist, daß für den Vergleich zweier Mittelwerte μ_i und μ_j entweder eine Hypothese der Form $H_{\langle i,j\rangle} : \mu_i \geq \mu_j$ getestet werden soll oder eine beliebige Auswahl der verschobenen Hypothesen $H_{[i,j]} : \mu_i \geq \mu_j + \delta$ und $H_{[j,i]} : \mu_j \geq \mu_i + \delta$. (Beide verschobenen Hypothesen würden für einen Äquivalenznachweis benötigt werden.)
Läßt man verschobene Nullhypothesen zu, so tritt zusätzlich der Effekt auf, daß Schnitte von Elementarhypothesen leer sein können. Zum Beispiel tritt dies beim Äquivalenztest für zwei Gruppen auf. $\mu_1 \geq \mu_2 + \delta$, $\mu_2 \geq \mu_1 + \delta \Rightarrow \mu_1 \geq \mu_1 + 2\delta$ (W!), d.h., der Schnitt der beiden einseitigen Nullhypothesen ist leer.

4.3.8.1 Bestimmung der kleinsten umfassenden erschöpfenden Indexmenge

Für eine beliebige Indexmenge I seien die in den Nullhypothesen vorkommenden Mittelwerte mit $\mu_1, \ldots, \mu_k$ bezeichnet. Die in I enthaltenen Nullhypothesen lassen sich schreiben als $H_{[i,j]} : \mu_i \geq \mu_j + \rho_{ij} \cdot \delta$, mit $\rho_{ij} \in \{0, 1\}$ und $i, j \in \{1, \ldots, k\}$.
Für zwei Nullhypothesen der Art $\mu_1 \geq \mu_2 + \rho_{12} \cdot \delta$ und $\mu_2 \geq \mu_3 + \rho_{23} \cdot \delta$ ergibt sich dann folgender Schluß: $\mu_1 \geq \mu_2 + \rho_{12} \cdot \delta \geq \mu_3 + (\rho_{12} + \rho_{23})\delta$. Allgemein haben so alle in I implizierten Hypothesenungleichungen, die keine Zyklen enthaltenden, die Form

$$\mu_{i_1} \geq \mu_{i_o} + \delta \sum_{j=1}^{o-1} \rho_{i_j i_{j+1}} := \mu_{i_o} + \delta \cdot a \quad \text{für } i_j \neq i_m \, (j, m < o),\ i_o \neq i_j \, (j > 1). \tag{4.5}$$

(Die sich ergebende Hypothesenungleichung kann ein Zyklus sein!) Man betrachte zuerst den Fall $i_o \neq i_1$: Existiert in I_l keine Elementarhypothese der Art $\mu_{i_1} \geq \mu_{i_o} + \rho_{i_1 i_o} \cdot \delta$, so stellt die Ungleichung (4.5) einen nichtleeren Schnitt von Elementarhypothesen dar und hat keine weitere Auswirkung auf die Indexmenge I. Ist in I_l eine Elementarhypothese der Art $H_{[i_1,i_o]} : \mu_{i_1} \geq \mu_{i_o} + \rho_{i_1 i_o} \cdot \delta$ vorhanden, so müssen zwei Fälle unterschieden werden. Ist $a \geq \rho_{i_1 i_o}$, so gilt $\mu_{i_1} \geq \mu_{i_o} + a\delta \Rightarrow \mu_{i_1} \geq \mu_{i_o} + \rho_{i_1 i_o} \cdot \delta$. D.h., die Elementarhypothese $H_{[i_1,i_o]}$ wird von den Elementarhypothesen aus I impliziert. Ist hingegen $a < \rho_{i_1 i_o}$, so wird $H_{[i_1,i_o]}$ nicht von der Ungleichung (4.5) impliziert.
Ist $i_o = i_1$, liegt also ein Zyklus vor, so liegt für $a > 0$ ein Widerspruch vor. Die Indexmenge repräsentiert demnach die leere Menge und ist im Abschlußtestsystem nicht zu betrachten. Andernfalls ist (4.5) eine immer wahre Ungleichung. Zur Bestimmung von $SE(I)$ haben wir somit den Algorithmus

i) Setze: $I_l = \mathbb{N}_l$ Indexmenge aller Elementarhypothesen (Indexmenge der Globalhypothese).

ii) Sei $I \subset \mathbb{N}_l$ die Indexmenge, für die $SE(I)$ zu bestimmen ist. Nehme die zugehörigen Paare von Indizes $\langle i_m, j_m \rangle$ mit $m \in I$. $SE(I) = I$.

iii) Bilde alle möglichen Ungleichungsketten $\mu_{k_0} \geq \ldots \geq \mu_{k_p} + \delta \sum \rho_m$ mit $k_{r-1} = i_m, k_r = j_m$ wobei $m \in I$ ist (r=1, ... p).

iv) Für jede Ungleichungskette $H_{[i_o,\ldots,i_p]} : \mu_{i_o} \geq \ldots \geq \mu_{i_p} + \delta \cdot a$ prüfe, ob $i_p = i_o$ ist.

- Ja: Ist $a > 0$, dann liegt ein Widerspruch vor und die Indexmenge ist nicht zulässig; andernfalls fahre mit der nächsten Ungleichungskette fort.
- Nein: Ist in I_l keine Elementarhypothese der Form $\mu_{i_o} \geq \mu_{i_p} + \rho_{i_o i_p} \cdot \delta$ vorhanden, so fahre mit der nächsten Ungleichungskette fort. Andernfalls:
  - $a \geq \rho_{i_o i_p}$: Füge die Elementarhypothese $H_{[i_o,i_p]}$ bzw. $H_{\langle i_o,i_p \rangle}$ zu $SE(I)$ hinzu.
  - $a < \rho_{i_o i_p}$: Fahre mit der nächsten Ungleichungskette fort.

Schritt iv) des Algorithmus läßt sich noch dahingehend vereinfachen, daß für jede Ungleichungskette von μ_{i_o} nach μ_{i_p} nur die mit maximalem a zu überprüfen ist.
Zur Abbildung des Problems auf einen Digraphen identifiziere man wie in Abschnitt 4.3.1 die Erwartungswerte mit den Knoten und die Hypothesen $H_{\langle i,j \rangle}$ bzw. $H_{[i,j]}$ mit den Pfeilen von i nach j. Zusätzlich wird jeder Pfeil, der eine Hypothese $H_{\langle i,j \rangle}$ repräsentiert, mit dem Gewicht 1, jeder der eine Hypothese $H_{[i,j]}$ repräsentiert mit dem Gewicht δ versehen. So entspricht dem Hypothesensystem $H_1 : \mu_1 \geq \mu_2 + \delta$, $H_2 : \mu_2 \geq \mu_1 + \delta$, $H_3 :$

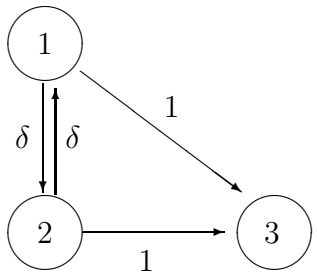


Abbildung 4.14: Graph mit gewichteten Pfeilen

$\mu_1 \geq \mu_3$, $H_4 : \mu_2 \geq \mu_3$ der *gewichtete Digraph* aus Abbildung 4.14. Die Adjazenzmatrix U für solch einen Digraphen setze man hier so, daß für jeden Pfeil sein Gewicht in die Adjazenzmatrix eingetragen wird. Die Adjazenzmatrix für den Beispieldigraphen lautet demnach:

$$\begin{pmatrix} 0 & \delta & 1 \\ \delta & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

In Abwandlung des Algorithmus zur Bestimmung der kritischen Wege (CPM-Algorithmus Neumann, Morlock 1993), definiere man die folgenden Operationen für die quadratischen Matrizen A, B gewichteter Digraphen.

$$\begin{aligned} A \oslash B &:= (\max\{a_{ij}, b_{ij}\})_{i,j=1,\ldots,k} \\ A \odot B &:= (c_{ij})_{i,j=1,\ldots,k} \text{ mit } c_{ij} = \max_{l=1}^{k}\{a_{il} \circ b_{lj}\} \end{aligned} \tag{4.6}$$

Wobei $a \circ b$ definiert ist durch

$\circ$	0	1	δ_2
0	0	0	0
1	0	1	δ_2
δ_1	0	δ_1	$\delta_1 + \delta_2$

Lemma 4.14
Sei $A = U(I)$ die Adjazenzmatrix eines gewichteten Digraphen mit k Knoten zur Indexmenge $I \subset I_l$ und $U = U(I_l)$ die Adjazenzmatrix der Globalhypothese. Aus der Matrix

$$\dot{R}_g(A) := A \oslash (A \odot A) \oslash (A \odot A \odot A) \oslash \ldots \oslash (\underbrace{A \odot \cdots \odot A}_{k\text{mal}})$$

lassen sich dann folgende Informationen ablesen:

- Aus den Diagonalelementen r_{ii} von $\dot{R}_g(A)$:
  - Ist mindestens ein $r_{ii} > 1$, so sind die Nullhypothesen der Indexmenge, die zu dem Graphen gehört, widersprüchlich.

- Aus den restlichen Elementen $r_{ij}, i \neq j$ von $\dot{R}_g(A)$:
  - Ist $0 < u_{ij} \leq r_{ij}$, so wird die dem Pfeil von i nach j entsprechende Elementarhypothese von den in I enthaltenen Hypothesen impliziert.

Mit Lemma 4.14 ergibt sich folgender, zur Implementierung geeigneter Algorithmus zur Bestimmung der Menge $SE(I)$:

i) Sei $I_l = \mathbb{N}_l$ die Indexmenge aller Elementarhypothesen und I die Indexmenge, zu der SE(I) zu bestimmen ist.

ii) Setze $U(I_l)$ als die Adjazenzmatrix zu I_l und $U(I)$ als die Adjazenzmatrix zu I.

iii) Berechne $\dot{R}_g(U(I))$ und überprüfe Bedingungen aus Lemma 4.14. Ist mindestens ein Diagonalelement von $\dot{R}_g(U(I)) > 1$, so ist I nicht weiter zu betrachten.

iv) $SE(I)$ besteht zusätzlich aus allen Hypothesen, die nach Schritt iii) impliziert wurden.

Für das obige Beispiel ergeben sich so $\dot{R}_g(U(I_l))$ und $\dot{R}_g(U(\{1,3\}))$ zu:

$$
\begin{aligned}
U(I_l) &= \begin{pmatrix} 0 & \delta & 1 \\ \delta & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \qquad U(\{1,3\}) = \begin{pmatrix} 0 & \delta & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \\
\dot{R}_g(U(I_l)) &= \begin{pmatrix} 0 & \delta & 1 \\ \delta & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \oslash \begin{pmatrix} 2\delta & 0 & \delta \\ 0 & 2\delta & \delta \\ 0 & 0 & 0 \end{pmatrix} \oslash \begin{pmatrix} 0 & 3\delta & 2\delta \\ 3\delta & 0 & 2\delta \\ 0 & 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} 2\delta & 3\delta & 2\delta \\ 3\delta & 2\delta & 2\delta \\ 0 & 0 & 0 \end{pmatrix} \\
\dot{R}_g(U(\{1,3\})) &= \begin{pmatrix} 0 & \delta & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \oslash \begin{pmatrix} 0 & 0 & \delta \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \oslash \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} 0 & \delta & \delta \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}
\end{aligned}
$$

Da in $\dot{R}_g(U(I_l))$ zwei Diagonalelemente größer als 1 sind, ist die Globalhypothese die leere Menge. Weiterhin ist die Indexmenge $\{1,4\}$ nicht erschöpfend, da sie die Elementarhypothese H_3 impliziert; die kleinste $\{1,4\}$ umfassende erschöpfende Indexmenge ist demnach $\{1,3,4\}$.

4.3.8.2 Überprüfung der Forderung, streng erschöpfend zu sein

Entsprechend der Definition 4.2 ist eine Indexmenge $SE(I)$ streng erschöpfend, wenn genau alle in ihr enthaltenen Elementarhypothesen wahr sein können. Das bedeutet, daß alle anderen Elementarhypothesen falsch sein müssen. Im folgenden ist also der Fall betrachtet, daß alle Nullhypothesen aus $SE(I)$ wahr sind, und alle Nullhypothesen aus $I_l \setminus SE(I)$ falsch sind. Entsprechend können vier verschiedene Typen von Hypothesen auftreten:

i) $H_{\langle i,j\rangle} : \mu_i \geq \mu_j$

ii) $A_{\langle j,i\rangle} : \mu_i < \mu_j$

iii) $H_{[i,j]} : \mu_i \geq \mu_j + \delta$

iv) $A_{[j,i]} : \mu_i < \mu_j + \delta \Leftrightarrow \mu_i - \delta < \mu_j$

Zur Beantwortung der Frage, ob eine erschöpfende Indexmenge streng erschöpfend ist, betrachte man den gewichteten Digraphen, in dem jeder Knoten einem Mittelwert entspricht und die Pfeile die Bewertungen aus Tabelle 4.8 enthalten. Tritt hierbei der

Hypothese	Pfeil	Bewertung
$H_{\langle i,j\rangle} : \mu_i \geq \mu_j$	von i nach j	1
$A_{\langle j,i\rangle} : \mu_i < \mu_j$	von j nach i	a
$H_{[i,j]} : \mu_i \geq \mu_j + \delta$	von i nach j	δ
$A_{[j,i]} : \mu_j > \mu_i - \delta$	von j nach i	$-\delta$

Tabelle 4.8: Hypothesen mit zugehörigen Pfeilen und Bewertungen

Fall auf, daß ein Pfeil die Bewertungen δ und $-\delta$ bekommen würde, so wird er mit δ bewertet. Dieser Fall kann genau dann auftreten, wenn in I_l ein Nullhypothesenpaar der Art $H_{[i,j]} : \mu_i \geq \mu_j + \delta$, $H_{[j,i]} : \mu_j \geq \mu_i + \delta$ enthalten ist und nur eine der beiden Hypothesen in $SE(I)$ auftritt. Sei oBdA $H_{[i,j]} \in SE(I)$, dann lauten die beiden sich ergebenden Hypothesen $\mu_i \geq \mu_j + \delta$ und $\mu_i > \mu_j - \delta$. Der Schnitt beider Hypothesen ist somit $H_{[i,j]} : \mu_i \geq \mu_j + \delta$.
Analog zu den Überlegungen in Abschnitt 4.3.4 kann nur dann ein Widerspruch, und damit ein leerer Schnitt, auftreten, wenn in dem gewichteten Digraphen ein Zyklus enthalten ist, wobei jeder mögliche Zyklus ein potentieller Widerspruch ist. Dafür betrachte man einen Zyklus und dessen Pfeilbewertungen. Sind die Pfeile nur mit 1'en und a's bewertet, so liegen die Situationen aus Abschnitt 4.3.4 vor. D.h., sind alle Pfeile mit 1 bewertet, so liegt kein Widerspruch vor; ist mindestens einer mit einem a bewertet,

so liegt ein Widerspruch vor. Betrachte nun den Fall, daß außer 1'en und a's genau ein Pfeil mit δ bewertet sei, die zugehörige Hypothese sei $\mu_i \geq \mu_j + \delta$. Die sich aus den restlichen Pfeilen ergebende Hypothesenkette hat die Form $\mu_j \geq \ldots \geq \mu_i$, wobei manche der Ungleichungen echte sein können. Als Schnitt ergibt sich so: $\mu_j \geq \ldots \geq \mu_i \geq \mu_j + \delta$ (W!). Analog ergeben sich Widersprüche, wenn mehrere Kanten mit δ bewertet sind. Sei in dem Zyklus genau eine Kante mit $-\delta$ bewertet und alle restlichen mit 1'en und a's. Die entsprechende Hypothese sei $\mu_i > \mu_j - \delta$. Ist die Hypothesenkette der restlichen Pfeile so benannt wie oben, erhalten wir eine widerspruchsfreie Hypothesenkette $\mu_j \geq \ldots \geq \mu_i \geq \mu_j - \delta$. Ist zusätzlich ein Pfeil mit δ bewertet, so erhält man die widersprüchliche Kette $\mu_j \geq \ldots \geq \mu_i > \mu_j$ (W!). Allgemein sieht man, daß die Hypothesenkette dann widersprüchlich ist, wenn die Anzahl der mit δ bewerteten Pfeile höher ist als die der mit $-\delta$ bewerteten.
Zusammengefaßt ergibt sich, daß eine erschöpfende Indexmenge genau dann streng erschöpfend ist, wenn ihr zugehöriger bewerteter Digraph keine Zyklen enthält, in denen mindestens ein Pfeil mit einem a bewertet ist und in keinem Zyklus die Anzahl der mit δ bewerteten Pfeile höher ist als die der mit $-\delta$ bewerteten. Bei dem zu lösenden Problem handelt es wiederum um die Bestimmung kritischer Pfade, hier Zyklen. Für die Operation $\odot$ aus (4.6) wird die Verknüpfungsmatrix von $\circ$ erweitert zu:

$\circ$	0	1	a	$-b_2\delta$	$b_2\delta$
0	0	0	0	0	0
1	0	1	a	$-b_2\delta$	$b_2\delta$
a	0	a	a	$-b_2\delta$	$b_2\delta$
$-b_1\delta$	0	$-b_1\delta$	$-b_1\delta$	$-(b_1+b_2)\delta$	$(b_2-b_1)\delta^*$
$b_1\delta$	0	$b_1\delta$	$b_1\delta$	$(b_1-b_2)\delta^*$	$(b_1+b_2)\delta$

* ist $|b_1 - b_2| = 0$, so setze $(b_1 - b_2)\delta = a$ bzw. $(b_2 - b_1)\delta = a$

Zur Vervollständigung der Definition der „Addition“ $\oslash$ sind die Pfeilbewertungen wie folgt geordnet:

$$-b\delta < 0 < a < b\delta, \quad b \in \mathbb{N}$$

Diese Anordnung folgt aus den Implikationen

$$\mu_i \geq \mu_j + b\delta \Rightarrow \mu_i > \mu_j \Rightarrow \mu_i \geq \mu_j \Rightarrow \mu_i > \mu_j - \delta$$

Lemma 4.15
Sei $A = U(SE(I))$ die Adjazenzmatrix eines gewichteten Digraphen mit k Knoten zur erschöpfenden Indexmenge $SE(I) \subset I_l$ und $U = U(I_l)$ die Adjazenzmatrix der Globalhypothese. Ist mindestens ein Diagonalelement der Matrix

$$\dot{R}_g(A) := A \oslash (A \odot A) \oslash (A \odot A \odot A) \oslash \ldots \oslash (\underbrace{A \odot \cdots \odot A}_{k\text{mal}})$$

gleich a oder ein Vielfaches von δ, so ist $SE(I)$ nicht streng erschöpfend.

Der Algorithmus von Seite 117 wird so noch um die folgenden Schritte erweitert:

vi) Stelle die zu $SE(I)$ gehörige Adjazenzmatrix U mit den Pfeilbewertungen aus Tabelle 4.8 auf.

vii) Berechne $\dot{R}_g(U)$; ist mindestens ein Diagonalelement gleich a oder ein positiv Vielfaches von δ, so ist $SE(I)$ nicht streng erschöpfend.

4.3.9 Beispiel einer Studie mit je einer Negativ- und Positiv-Kontrolle

Im Rahmen einer Studie mit einem Kalziumkanal-Blocker für Patienten mit chronischer Agina pectoris wurde ein neues Medikament getestet[3]. Neben dem Medikament in drei Dosierungen 50, 100 und 150mg wurden eine Plazebogruppe (0) und eine aktive Kontrollgruppe (a), behandelt mit Amlodipine, in das Versuchsdesign aufgenommen. Zur Demonstration des Verfahrens aus dem vorhergegangenen Abschnitt seinen hier die folgenden Fragestellungen von Interesse: Zum einen wird gefragt, welche der Dosen besser ist als das Plazebo. Weiter interessiert, welche der Dosen äquivalent zur aktiven Kontrolle sind, hierbei sei die Äquivalenzgrenze $\delta = 30$ Einheiten der gemessenen Größe. (Diese Hypothesen wurden auch in Bauer et al (1998) aufgestellt). Zusätzlich wird hier noch gefragt, ob die aktive Kontrolle besser als die Plazebobehandlung ist. Mit den Bezeichungen $\mu_i\,(i = 50, 100, 150)$ für die Effekte der Dosisgruppen und μ_0 bzw. μ_a für die Effekte des Plazebos bzw. der aktiven Kontrolle sind die folgenden Hypothesen zu testen:

$$\begin{aligned}
H_{\langle 0,50\rangle} &: \mu_0 \geq \mu_{50}\ , \\
H_{\langle 0,100\rangle} &: \mu_0 \geq \mu_{100}, \\
H_{\langle 0,150\rangle} &: \mu_0 \geq \mu_{150}, \\
H_{\langle 0,a\rangle} &: \mu_0 \geq \mu_a, \\
H_{[a,50]} &: \mu_a \geq \mu_{50} + \delta, \\
H_{[a,100]} &: \mu_a \geq \mu_{100} + \delta, \\
H_{[a,150]} &: \mu_a \geq \mu_{150} + \delta.
\end{aligned}$$

Die Mittelwerte, Standardabweichungen und Fallzahlen der einzelnen Behandlungsgruppen sind:
Als gepoolte Standardabweichung erhält man aus den Daten $\hat{\sigma} = 82,9$.
Die p-Werte der einzelnen t-Tests ergeben sich somit zu:

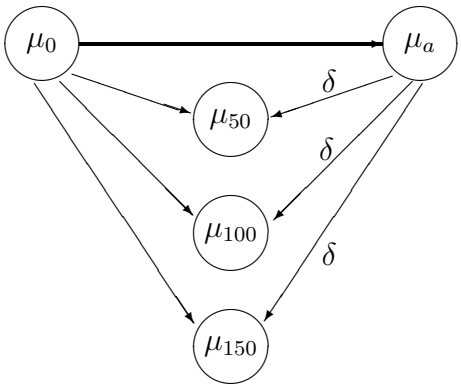
[3]Die Daten stammen aus der Arbeit von Bauer et al (1998).

	Behandlung				
	0	50	100	150	a
$\bar{x}$	57,5	76,8	109,5	105,3	67,3
$\hat{\sigma}$	75,0	75,5	87,1	85,7	90,1
n	62	60	60	60	59

Tabelle 4.9: Mittelwerte, Streuungen und Fallzahlen der einzelnen Behandlungen

$$
\begin{aligned}
p_{(6)} &:= p_{\langle 0,50\rangle} = 0,1,\\
p_{(3)} &:= p_{\langle 0,100\rangle} = 0,0001,\\
p_{(4)} &:= p_{\langle 0,150\rangle} = 0,0007,\\
p_{(7)} &:= p_{\langle 0,a\rangle} = 0,26,\\
p_{(5)} &:= p_{[a,50]} = 0,0049,\\
p_{(2)} &:= p_{[a,100]} = 0,0001,\\
p_{(1)} &:= p_{[a,150]} = 0,0001.
\end{aligned}
$$

Als multiples Niveau sei $\alpha = 5\%$ gewählt. Der zugehörige Graph zur Globalhypothese hat dann die Form:

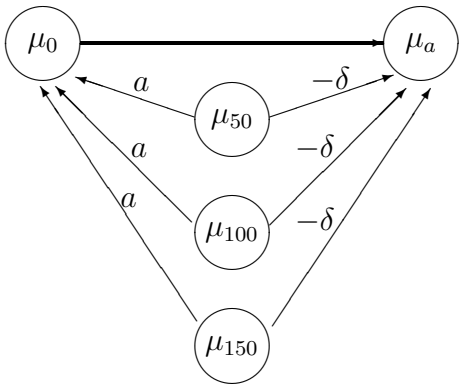


$$
\begin{array}{c|c}
Beh. & 0-150 \\
0 & \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & \delta & \delta & \delta \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \\
a & \\
50 & \\
100 & \\
150 &
\end{array}
$$

Abbildung 4.15: Gerichteter Graph und Adjazenzmatrix der globalen Nullhypothese

Da die vier kleinsten p-Werte zur Bonferronischranke von $\alpha/7$ abgelehnt werden und damit auch bei einem vorgeschalteten Holm-Verfahren, betrachte man das so verbleibende Hypothesensystem mit den Hypothesen $H_5 := H_{\langle 0,a\rangle}$, $H_6 := H_{[a,50]}$ und $H_7 := H_{\langle 0,50\rangle}$[4]. Die Information, daß die restlichen Hypothesen abgelehnt sind, wird natürlich weiter verwendet. Außerdem gilt hier, daß keine Elementarhypothese eine andere impliziert, alle Elementarhypothesen daher erschöpfend sind. Entsprechend der Vorgehensweise von Prozedur 4B wird mit der Indexmenge $\{5\}$ gestartet. Von dieser ist zu überprüfen, ob sie streng erschöpfend ist. Der sich ergebende Graph und die zugehörige Adjazenzmatrix haben die Gestalt:

[4]Die restlichen drei Hypothesen sind nicht nach ihren p-Werten geordnet.



$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ a & -\delta & 0 & 0 & 0 \\ a & -\delta & 0 & 0 & 0 \\ a & -\delta & 0 & 0 & 0 \end{pmatrix}$$

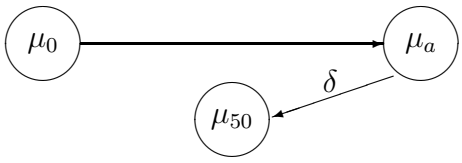
In dem Graph ist zu sehen, daß kein Weg mehr als drei Knoten umfassen kann, daher sind nur A^2 und A^3 zu berechnen.

$$A^2 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ a & -\delta & 0 & 0 & 0 \\ a & -\delta & 0 & 0 & 0 \\ a & -\delta & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ a & -\delta & 0 & 0 & 0 \\ a & -\delta & 0 & 0 & 0 \\ a & -\delta & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & a & 0 & 0 & 0 \\ 0 & a & 0 & 0 & 0 \\ 0 & a & 0 & 0 & 0 \end{pmatrix}$$

$$A^3 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ a & -\delta & 0 & 0 & 0 \\ a & -\delta & 0 & 0 & 0 \\ a & -\delta & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & a & 0 & 0 & 0 \\ 0 & a & 0 & 0 & 0 \\ 0 & a & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\dot{R}_g(A) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ a & a & 0 & 0 & 0 \\ a & a & 0 & 0 & 0 \\ a & a & 0 & 0 & 0 \end{pmatrix}$$

Da kein Diagonalelement von $\dot{R}_g(U)$ gleich a oder größer als Null ist, ist die Indexmenge $\{5\}$ streng erschöpfend. Als nächstes ist die Indexmenge $\{5, 6\}$ zu betrachten, von der zuerst zu prüfen ist, ob sie erschöpfend ist. Dafür betrachte man den Graph mit den beiden Hypothesen H_5 und H_6; da die Behandlungen 50mg und 100mg in den Hypothesen nicht vorkommen, können ihre Knoten im Graphen weggelassen werden.



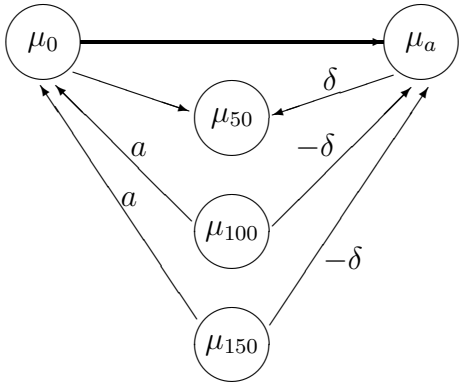
$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & \delta \\ 0 & 0 & 0 \end{pmatrix}$$

Die Erreichbarkeitsmatrix $\dot{R}_g(A)$ ergibt sich so durch

$$A^2 = \begin{pmatrix} 0 & 0 & \delta \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad A^3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\dot{R}_g(A) \;=\; \begin{pmatrix} 0 & 1 & \delta \\ 0 & 0 & \delta \\ 0 & 0 & 0 \end{pmatrix}.$$

Aus dem Element rechts oben in $\dot{R}_g(A)$ sieht man, daß die beiden Hypothesen die Hypothese $H_7 : \mu_0 \geq \mu_{50}$ implizieren. D.h., die Indexmenge $\{5, 6\}$ ist nicht erschöpfend, und die kleinste sie umfassende erschöpfende Indexmenge ist $\{5, 6, 7\}$. Zur Überprüfung, ob $\{5, 6, 7\}$ auch streng erschöpfend ist, sind der Graph und die zugehörige Adjazenzmatrix



$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & \delta & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ a & -\delta & 0 & 0 & 0 \\ a & -\delta & 0 & 0 & 0 \end{pmatrix}$$

zu betrachten. Wie bei der vorigen Indexmenge sind auch hier nur A^2 und A^3 zu berechen.

$$A^2 \;=\; \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & \delta & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ a & -\delta & 0 & 0 & 0 \\ a & -\delta & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & \delta & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ a & -\delta & 0 & 0 & 0 \\ a & -\delta & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & \delta & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & a & a & 0 & 0 \\ 0 & a & a & 0 & 0 \end{pmatrix}$$

$$A^3 \;=\; \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & \delta & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ a & -\delta & 0 & 0 & 0 \\ a & -\delta & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & \delta & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & a & a & 0 & 0 \\ 0 & a & a & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \delta & 0 & 0 \\ 0 & 0 & \delta & 0 & 0 \end{pmatrix}$$

$$\dot{R}_g(A) \;=\; \begin{pmatrix} 0 & 1 & \delta & 0 & 0 \\ 0 & 0 & \delta & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ a & a & \delta & 0 & 0 \\ a & a & \delta & 0 & 0 \end{pmatrix}$$

Da in $\dot{R}_g(A)$ alle Diagonalelemente gleich Null sind, ist die Indexmenge $\{5, 6, 7\}$ streng erschöpfend. Sukzessive werden noch die Indexmengen $\{5, 7\}, \{6\}, \{6, 7\}$ und $\{7\}$ geprüft, die alle streng erschöpfend sind. Somit sind die im Abschlußtest auftretenden Schnitt- und Elementarypothesen durch die Indexmengen

$$\{5\},\ \{6\},\ \{7\},\ \{5, 7\},\ \{6, 7\},\ \{5, 6, 7\}$$

repräsentiert. Demnach wird der kleinste der drei restlichen p-Werte $p_{(5)} = p_6 = 0,0049$ zu $\alpha/3 = 0,0166$ getestet, der zweitkleinste $p_{(6)} = p_5 = 0,1$ zu $\alpha/2 = 0,025$ und der größte $p_{(7)} = p_7 = 0,26$ zu $\alpha = 0,05$.
Als Ergebnis können alle Hypothesen bis auf $H_{\langle 0,50\rangle}$ und $H_{\langle 0,a\rangle}$ abgelehnt werden. Demnach werden alle drei Dosierungen als äquivalent zur aktiven Kontrolle erklärt, jedoch kann die aktive Kontrolle dem Plazebo gegenüber nicht als besser erkannt werden. Während die Dosierungen mit 100mg und 150mg als signifikant wirksamer als das Plazebeo erkannt werden, gilt dies nicht für die geringste Dosierung.

4.3.10 Diskussion

In diesem Kapitel wurde geklärt, wie sich logische Abhängigkeiten von paarweisen Hypothesen anhand von Graphen visualisieren und systematisch bestimmen lassen. Diese sind insbesondere dann von Interesse, wenn mehrere Elementarhypothesen unter Wahrung eines multiplen Niveaus getestet werden sollen. Verwendet man nämlich das Abschlußtestverfahren, so reduziert sich die Anzahl der zu testenden Hypothesen entsprechend dem „Grad“ der logischen Abhängigkeiten. Das Hauptaugenmerk lag hierbei auf der Bestimmung der logischen Abhängigkeiten einseitiger Hypothesen, da sich in der Literatur keine allgemeine Lösung dieses Problems findet. Ein früherer graphentheoretischer Ansatz stammt von Hochberg und Conforti (1987), dieser ist jedoch nur für spezielle p-Wert-Ordnungen verwendbar und wurde nicht weiterverfolgt (Hochberg, persönliche Mitteilung).
Als Besonderheit beim schrittweisen Testen von einseitigen paarweisen Hypothesen fällt auf, daß es vorkommen kann, daß einige der Elementarhypothesen nicht einzeln getestet werden (siehe Beispiel Dosisschritte, Seite 71). D.h., die Entscheidung über die Ablehnung dieser Elementarhypothesen wird durch die Entscheidungen bezüglich der restlichen Hypothesen impliziert. Weiterhin kann es im Abschlußtest bei Verwendung von Bonferroni-Globaltests für die Schnitthypothesen vorkommen, daß die Ablehnentscheidungen nicht monoton in den p-Werten sind. Verwendet man hingegen die in dieser Arbeit beschriebenen adjustierten p-Werte, so wird diese Monotonie erzwungen. Ein Programm zur Berechnung dieser adjustierten p-Werte für paarweise einseitige Hypothesen ist im Kapitel „7. Software“ im Abschnitt 7.3 angegeben.
Die Verfahren von Bernhard (1991) und Westfall (1997) lassen sich ebenfalls verwenden, um einseitige paarweise, sogar lineare Kontrast-Hypothesen zu testen. Um diese Verfahren verwenden zu können, formuliere man die Nullhypothesen als Punkthypothesen (zweiseitig). Anstatt jedoch die Hypothesen mit einem zweiseitigen Test zu testen, verwende man einen einseitigen bezüglich der interessierenden Richtung. Da nach einer

Ablehnung der zweiseitigen Hypothese die Richtungsentscheidung zulässig ist, erhält man so die gewünschten Entscheidungen bei Kontrolle des multiplen Niveaus. Bei solch einer Vorgehensweise bleibt jedoch die Kontrolle der Fehler 3. Art fraglich (Westfall, persönliche Mitteilung). Ist mindestens eine der Nullhypothesen verschoben, so sind die Verfahren von Westfall und Bernhard nicht mehr anwendbar, da diese Hypothesen verschobenen linearen Räumen entsprechen. Daher ist die Herangehensweise über lineare Verktorräume nicht mehr möglich.
Bei der Anwendung des Abschlußtests bei der simultanen Bestimmung des höchsten effektiven Dosisschritts (HEDS) und der minimalen effektiven Dosis (MED) zeigte das Verfahren eine hohe Power über alle betrachten Profile hinweg. Ist man nur an dem HEDS interessiert, so ist unter Ordnugsrestiktion das Verfahren von Bauer und Budde (1994) gleichmäßig besser. Auch bei nichtmonotonen Erwartungswertprofilen zeigt das Verfahren von Bauer und Budde, jedoch mit paarweisen Tests durchgeführt, die gleichmäßig höchste Güte. Je nach Profil weist hier das Abschlußtestverfahren eine bis zu neun Prozentpunkte geringere Güte auf.

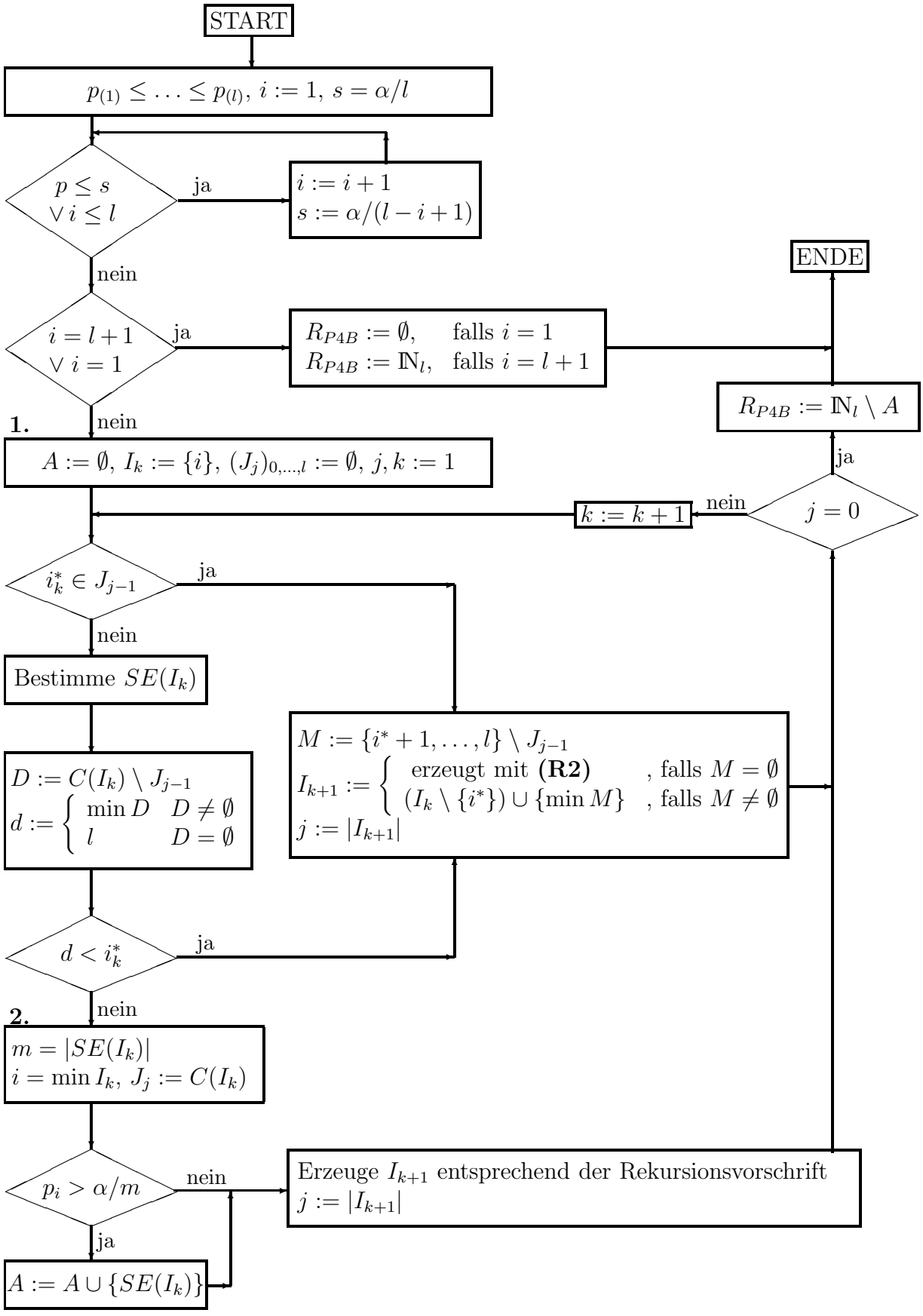


Abbildung 4.16: Flußdiagramm zur Durchführung von P4B

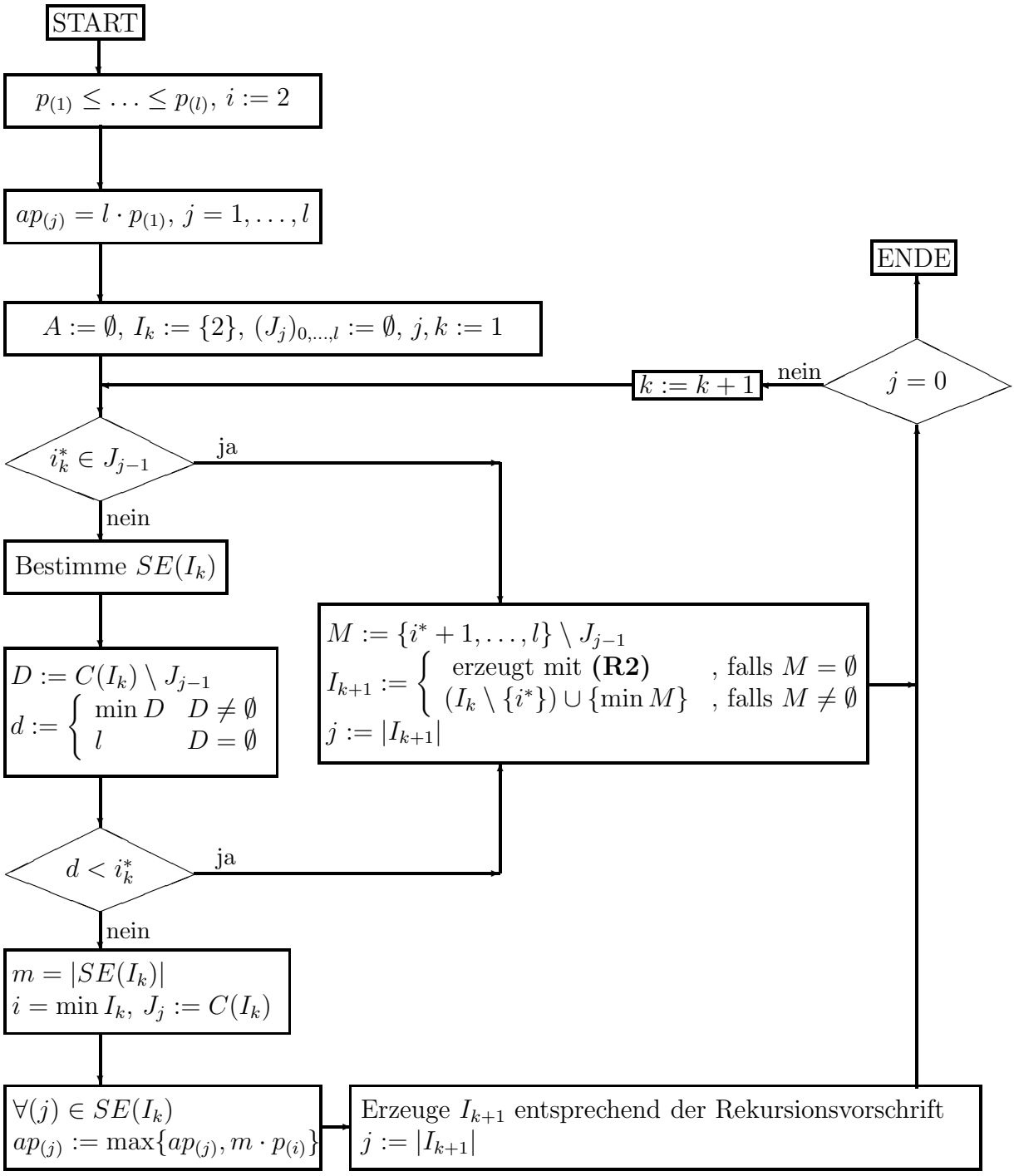


Abbildung 4.17: Flußdiagramm zur Bestimmung der adjustierten p-Werte nach P4B

Kapitel 5

Anwendungen

An den Beispielen aus Kapitel 2 werden die in dieser Arbeit vorgestellten Verfahren bezüglich ihrer Entscheidungen verglichen. Als multiples Niveau ist $\alpha = 5\%$ gewählt. Für die Schätzungen der Parameter nach dem Ansatz von Hettmansperger und McKean wurde das Program RGLM (2.0) von Kapenga, McKean und Vidmar (1995) verwendet.

5.1 Länge von Kuckuckseiern

In diesem Beispiel wird danach gefragt, ob die Größe der Eier des Kuckucks denen der ausgewählten Ziehart angepasst sind. Die Boxplots aus Abbildung 2.1 lassen vermuten, daß dies der Fall ist. Weiterhin interessiert, bezüglich welcher der Wirtsvogelarten sich die Eigrößen unterscheiden. Zur Beantwortung dieser Fragen wird deswegen ein simultaner All-paar-Vergleich verwendet. Die p-Werte, die sich für jeden zweiseitigen Einzelvergleich ergeben, sind in Anhang A.3 in Tabelle A.1 zusammengefaßt.
Alle Verfahren kommen zu dem Ergebnis, daß die Eigröße des Kuckucks sich bezüglich der jeweiligen Zieheltern unterscheidet. Die Eier, die bei Zaunkönigen im Nest gefunden wurden, sind signifikant kleiner als bei allen anderen Arten. Weiterhin sind die Eier in Nestern des Wiesenpiepers kleiner als die in denen des Baumpiepers. Nur bei der Unterscheidung zwischen Spatz und Wiesenpieper kann der Rangtest nach Brunner nicht ablehnen, während alle anderen Tests hier einen Unterschied feststellen. Aufgrund des größeren Varianzschätzers (siehe Tabelle 5.2) kann der Test anhand der Rangschätzer nach Hettmansperger und McKean nur die Eier in den Zaunköniggelegen als kleiner klassifizieren.
Ein anderer Ansatz ist, anstatt die Größen der Kuckuckseier miteinander zu vergleichen, zu überprüfen, ob die simultanen 95% Konfidenzintervalle der Kuckuckseilängen innerhalb der Größenspanne der jeweiligen Vogelart (Tabelle 5.1) liegen. Aus den einzelnen Verfahren zu Schätzung der Parameter erhält man für die „mittlere“ Eilänge und

Vogelart	Eilänge [mm]
Bachstelze	16,7 - 22,3
Baumpieper	18,0 - 23,5
Rotkehlchen	16,9 - 22,3
Spatz	19,1 - 25,4
Wiesenpieper	17,2 - 21,4
Zaunkönig	14,7 - 18,9

Tabelle 5.1: Eilängen verschiedener Vogelarten

deren Standardabweichung die Werte aus Tabelle 5.2. Mit diesen Parametern ergeben

	Bach	Baum	Rot	Spatz	Wiese	Zaun	$\hat{\sigma}$
μ, σ	22,90	23,09	22,58	23,12	22,30	21,13	0,9093
trimmen	22,87	23,25	22,61	23,21	22,27	21,16	0,7216
Huber-m	22,90	23,14	22,59	23,17	22,32	21,13	0,8980
Hettmansperger	22,85	23,04	22,46	23,09	22,25	21,05	1,2403

Tabelle 5.2: Geschätzte mittlere Eilängen und Standardabweichungen

sich die 95% Konfidenzintervalle für die Eilängen zu den in Abbildung 5.1 abgebildeten.

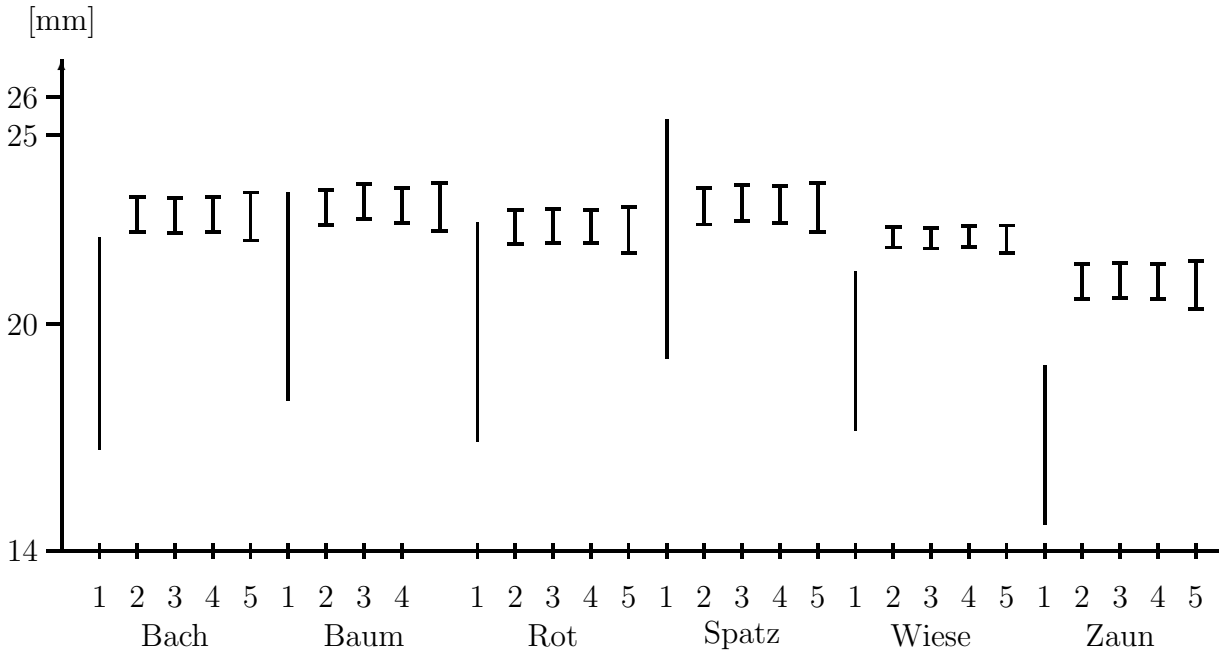


Abbildung 5.1: 95% Konfidenzintervalle für die Längen der Kuckuckseier
1: Längenbereich der Vogelart; 2: KI aus μ und σ; 3: KI der Trimmschätzer; 4: KI der Huber-m-Schätzer; 5: KI nach Hettmansperger

Die Konfidenzintervalle zeigen, daß der Kuckuck die Größe seiner Eier nicht exakt an die des Zielgeleges anpaßt. Jedoch variiert, wie sich auch bei dem multiplen Vergleich gezeigt hat, die Eigröße artspezifisch. D.h., das die Eier des Kuckucks bei kleineren Vogelarten tendenziell kleiner sind.

5.2 Einfluß verschiedener Belichtungsvarianten

Bei der Auswertung des Experiments, das als zweites Beispiel in Kapitel 2 vorgestellt wurde, fließt ein, daß die Sorten unterschiedlich auf die Umweltbedingungen reagieren. Deswegen soll pro Sorte der Effekt der Belichtung untersucht werden. Da von vornherein nicht bekannt ist, ob eine Verlängerung der Belichtungsdauer eine Verkürzung der Kultivierungsdauer möglich macht, wird ein All-paar-Vergleich durchgeführt.
Für Sorte 1 erhält man das Resultat, daß alle Behandlungen von mehr als 8h Dauer eine Verkürzung der Kultivierungsdauer möglich machen unabhängig von dem verwendeten Test (Tabelle 5.3). Bei den Vergleichen der Behandlungen 16h Assimilationslicht mit 12h und 16h photoperiodischer Tagesverlängerung zeigt sich der getrimmte Test weitaus konservativer als die restlichen Verfahren. Außerdem kommen hier beide Rangtests bei dem Vergleich von 16h photoperiodischer Tagesverlängerung mit 16h Assimilationslicht zu einer anderen Entscheidung als die restlichen Tests.
Bei der zweiten Sorte fällt auf, daß der Huber-m-Test beim Vergleich der Variante 8h photoperiodischer Tagesverlängerung mit den 3 letzten Behandlungen als einziger keinen Unterschied feststellt. Zudem zeigen das Trimmen und der Huber-m-Test beim Vergleich der Behandlung 12h photoperiodischer Tagesverlängerung mit den drei letzteren Behandlungen nicht signifikante p-Werte ($> 10\%$), wohingegen die restlichen Test weitaus kleinere p-Werte liefern.
Die Testergebnisse zu Sorte 3 spiegeln die aus dem Boxplot zu ersehenden Unterschiede signifikant ($p<0.0001$) wieder. D.h., für die beiden Belichtungsarten nimmt mit steigender Behandlungsdauer die Kultivierungsdauer ab. Weiterhin führt bei gleicher Behandlungsdauer die Verwendung des Assimilationslichts zu einer kürzeren Kultivierungsdauer.

tukey huber 20% trim westfall Brunner Hettmansperger	12h photo. Tagesverlängerung	12h Assimilationslicht	16h photo. Tagesverlängerung	16h Assimilationslicht
	0,001	0,001	0,001	0,001
	0,001	0,001	0,001	0,001
8h photo.	0,001	0,001	0,001	0,001
	0,001	0,001	0,001	0,001
	0,001	0,001	0,001	0,001
	0,001	0,001	0,001	0,001
		0,636	0,896	0,009
		0,174	0,821	0,002
12h photo.		0,174	0,990	0,012
		0,651	0,906	0,008
		0,120	0,895	0,001
		0,096	0,841	0,001
			0,987	0,258
			0,763	0,490
12h Assi.			0,790	0,790
			0,990	0,259
			0,550	0,222
			0,566	0,515
				0,093
				0,051
16h photo.				0,174
				0,091
				0,044
				0,024

Tabelle 5.3: p-Werte der paarweisen Vergleiche mit den verschiedenen Methoden für Sorte 1

tukey huber 20% trim westfall Brunner Hettmansperger	12h photo. Tagesverlängerung	12h Assimilationslicht	16h photo. Tagesverlängerung	16h Assimilationslicht
	0.998	0.012	0.008	0.012
	0.994	0.119	0.109	0.145
8h photo.	0.946	0.025	0.030	0.046
	0.998	0.013	0.008	0.013
	0.956	0.007	0.009	0.016
	0.991	0.001	0.001	0.001
		0.028	0.018	0.028
		0.263	0.245	0.306
12h photo.		0.140	0.164	0.222
		0.028	0.018	0.028
		0.052	0.062	0.097
		0.004	0.004	0.004
			1	1
			1	1
12h Assi.			1	1
			1	1
			1	1
			1	1
				1
				1
16h photo.				1
				1
				1
				1

Tabelle 5.4: p-Werte der paarweisen Vergleiche mit den verschiedenen Methoden für Sorte 2

5.3 Durchlässigkeit einer Membran

Die in Abschnitt 2.3 gestellten Fragen lassen sich auf die folgenden Kontraste abbilden.

Vergleich	zugehöriger Kontrast									
	Kon	D1	D2	D3	E	G	H	I	J	K
Kon vs. D1	-1	1	0	0	0	0	0	0	0	0
Kon vs. D2	-1	0	1	0	0	0	0	0	0	0
Kon vs. D3	-1	0	0	1	0	0	0	0	0	0
Kon vs. E	-1	1	0	0	1	0	0	0	0	0
⋮					⋮					
Kon vs. K	-1	0	0	0	0	0	0	0	0	1
D1 vs. D2	0	-1	1	0	0	0	0	0	0	0
D1 vs. D3	0	-1	0	1	0	0	0	0	0	0
D2 vs. D3	0	0	-1	1	0	0	0	0	0	0
D1 vs. E	0	-1	0	0	1	0	0	0	0	0
	0	1	0	0	-1	0	0	0	0	0
D1 vs. G	0	-1	0	0	0	1	0	0	0	0
	0	1	0	0	0	-1	0	0	0	0
⋮					⋮					
J vs. K	0	0	0	0	0	0	0	0	-1	1
	0	0	0	0	0	0	0	0	1	-1

Die Entscheidungen der verschiedenen Verfahren unterscheiden sich nicht bei denenigen Behandlungen mit kleiner Varianz (Kon, D1, D2, D3, E). So lassen sich die Behandlungen Kon, D1, D2, D3 und E untereinander bezüglich der gemessenen Zielvariablen nicht unterscheiden, wogegen sich bei den Behandlungen G, H, I, J und K eine Effektsteigerung gegenüber den restlichen Behandlungen nachweisen läßt. Zu unterschiedlichen Entscheidungen kommen die verschiedenen Verfahren bei den Vergleichen innerhalb der Gruppen G, H, I, J, K, die im Boxplot (Abbildung 2.3 auf Seite 22) die relativ großen Varianzen aufweisen. Hier sind nur der Test basierend auf den Huber-m-Schätzern und der Rang-Test nach Hettmansperger und McKean in der Lage, die Behandlungen zu diskriminieren. Während der Rang-Test die beiden Behandlungen mit den höchsten Medianen (G,J) von den anderen dreien (H,I,K) trennen kann, erkennt der m-Test den Unterschied zwischen den Behandlungen G und I nicht.

vs.	westfall	μ, σ	huber	20% trim	Brunner	Hettmansperger
Kon vs. D1	1	1	1	1	0,997	1
Kon vs. D2	1	1	1	1	0,897	1
Kon vs. D3	1	1	1	1	0,964	1
Kon vs. E	1	1	1	1	1	1
Kon vs. G	0,001	0,001	0,001	0,001	0,001	0,001
Kon vs. H	0,001	0,001	0,001	0,001	0,001	0,001
Kon vs. I	0,001	0,001	0,001	0,001	0,001	0,001
Kon vs. J	0,001	0,001	0,001	0,001	0,001	0,001
Kon vs. K	0,001	0,001	0,001	0,001	0,001	0,001
D1 vs. D2	1	1	1	1	0,319	0,998
D1 vs. D3	1	1	1	1	0,478	0,999
D2 vs. D3	1	1	1	1	0,997	1
D1 vs. E	1	1	1	1	0,870	1
D1 vs. G	0,001	0,001	0,001	0,001	0,001	0,001
D1 vs. H	0,001	0,001	0,001	0,001	0,001	0,001
D1 vs. I	0,001	0,001	0,001	0,001	0,001	0,001
D1 vs. J	0,001	0,001	0,001	0,001	0,001	0,001
D1 vs. K	0,001	0,001	0,001	0,001	0,001	0,001
D2 vs. E	1	1	1	1	1	1
D2 vs. G	0,001	0,001	0,001	0,001	0,001	0,001
D2 vs. H	0,001	0,001	0,001	0,001	0,001	0,001
D2 vs. I	0,001	0,001	0,001	0,001	0,001	0,001
D2 vs. J	0,001	0,001	0,001	0,001	0,001	0,001
D2 vs. K	0,001	0,001	0,001	0,001	0,001	0,001
D3 vs. E	1	1	1	1	1	1
D3 vs. G	0,001	0,001	0,001	0,001	0,001	0,001
D3 vs. H	0,001	0,001	0,001	0,001	0,001	0,001
D3 vs. I	0,001	0,001	0,001	0,001	0,001	0,001
D3 vs. J	0,001	0,001	0,001	0,001	0,001	0,001
D3 vs. K	0,001	0,001	0,001	0,001	0,001	0,001
E vs. G	0,001	0,001	0,001	0,001	0,001	0,001
E vs. H	0,001	0,001	0,001	0,001	0,001	0,001
E vs. I	0,001	0,001	0,001	0,001	0,001	0,001
E vs. J	0,001	0,001	0,001	0,001	0,001	0,001
E vs. K	0,001	0,001	0,001	0,001	0,001	0,001
G vs. **H**	0,808	0,832	0,045	0,314	0,566	0,002
G vs. I	0,884	0,902	0,148	0,488	0,991	0,006
G vs. J	0,973	0,980	0,977	0,999	1	0,882
G vs. **K**	0,879	0,898	0,165	0,456	1	0,001
H vs. I	1	1	1	1	0,989	1
H vs. **J**	0,14	0,151	0,003	0,112	0,398	0,001
H vs. K	1	1	1	1	0,971	1
I vs. **J**	0,201	0,215	0,015	0,207	0,960	0,001
I vs. K	1	1	1	1	1	1
J vs. **K**	0,224	0,240	0,0188	0,198	0,993	0,001

Tabelle 5.5: Multiple p-Werte der einzelnen Vergleiche

Kapitel 6

Zusammenfassung und Ausblick

Ausgangspunkt dieser Arbeit war das Problem, daß experimentelle Daten im allgemeinen nicht den strengen ANOVA- Modellannahmen entsprechen. Zum einen treten extreme Werte auf, bei denen nicht klar ist, ob sie Meßfehler sind. Zum anderen sind die Daten oft nicht symmetrisch verteilt. Daher ist die Verwendung von Mittelwert und Standardabweichung in Tests wie dem t-Test oder multiplen Kontrasttests, im folgenden Standardtests genannt, nicht als sinnvoll zu erachten. Einen Ausweg, der weiterhin die Schätzung von Effekten erlaubt, ist die Verwendung robuster Schätzer für „Mittelwert“ und „Standardabweichung“ . Im Rahmen dieser Arbeit sind dafür trim-, Huber-m- und Tiku MML-Schätzer vorgestellt worden. Aus diesen wurden in einem ersten Schritt Zwei-Stichproben-Tests konstruiert. Diese Tests sind bei normalverteilten Daten Niveau-α-Tests und bezüglich ihrer Güte nur geringfügig schlechter als der t-Test. Unter Ausreißerverteilungen zeigt sich, daß der Tiku-Test nicht niveautreu ist, und die robustifizierten Tests eine höhere Power als der t-Test haben. Jedoch ließ sich kein über alle Situationen poweroptimaler Test identifizieren. Die Optimalität, d.h. welcher Test die höchste Güte hat, hängt insbesondere von der Schiefe der zugrundeliegenden Verteilung ab. Um zu einem gleichmäßig „guten“ Test zu gelangen wurde die Idee von Neuhäuser (1996) aufgegriffen, das Maximum verschiedener Teststatistiken zu betrachten. Aufgrund der verschiedenen Freiheitsgrade der verwendeten Teststatistiken wurden anstelle des Maximums der Teststatistiken das Minimum der korespondierenden p-Werte verwendet. Da die Verteilungen der hier betrachteten Minima nicht explizit zu bestimmen waren, wurde ein Bootstrap-Verfahren zur Durchführung des Tests verwendet. Von den betrachteten Minimum-Tests zeigte der Minimum-Test aus t-, Huber-m- und trim-Schätzer eine gleichmäßig gute Güte über alle betrachteten Situationen. Daher ist im Zwei-Stichproben-Fall dieser Minimum-Test zu empfehlen. Im zweiten Schritt wurden die Ergebnisse für trim- und m-Schätzer aus dem Zwei-Stichproben-Fall auf k-Stichproben verallgemeinert. Hier wurde gezeigt, daß die Vektoren aus den k Lage

trim- bzw. m-Schätzern asymptotisch multivariat normal verteilt sind. Im Rahmen einer Simulationsstudie wurde dann überprüft, in wie weit sich im Endlichen multiple Kontraste dieser Schätzer durch die multi-t-Verteilung approximieren lassen. Unter Normalverteilung wie auch den Ausreißerverteilungen hielten alle multiplen Kontrasttests die Niveaus 1%, 5% und 10% ein. Daher kann man die Approximation durch die multi-t-Verteilung ab einer Fallzahl von 10 Beobachtungen pro Gruppe als geeignet erachten. Bezüglich der Güte unter Normalverteilung zeigen die Verfahren basierend auf dem Huber-m-Schätzer sowie die gebootstrapten Minimum-Tests und der Rangtest nach Brunner nur eine minimal geringere Power als der Standardtest. Hingegen zeigt das Rangverfahren nach Hettmansperger und McKean bei Fallzahlen bis 20 Beobachtungen eine geringere Güte. In nichtnormalen Situationen bei kleinen Fallzahlen zeigt der Huber-m-Schätzer basierte multiple Kontrasttest, im Gegensatz zum Zwei-Stichproben-Fall, eine gleichmäßig hohe Güte. Daher ist der Huber-m-Schätzer basierte multiple Kontrasttest zur Schätzung von Parametern und Konfidenzintervallen bei Fallzahlen bis 15 Beobachtungen pro Gruppe zu empfehlen. Hat man über 25 Beobachtungen pro Gruppe so ist das Rangverfahren von Hettmansperger und KcKean zu empfehlen. Außerdem haben die Simulationsergebnisse gezeigt, daß man prinzipiell bei nichtnormalen Daten von der Verwendung der Standardtests absehen sollte.
Als eine weitere Möglichkeit, robuste multiple Mittelwertvergleiche durchzuführen, wurde in einer weiteren Betrachtung ein schrittweises Verfahren basierend auf dem Abschlußtest betrachtet. Hierbei stand die Betrachtung paarweiser einseitiger Hypothesen im Mittelpunkt. Ausgehend von l Elementarhypothesen müßten im Abschlußtestsystem i.a. $2^l - 1$ Schnitthypothesen betrachtet werden, sind jedoch die einzelnen Hypothesen logisch abhängig, so reduziert sich die Anzahl der zu betrachtenden Schnitthypothesen. Im Fall von zweiseitigen single Kontrast Hypothesen ist das Problem der Bestimmung der relevanten Schnitthypothesen von Bernhard (1991) und Westfall (1997) gelöst worden.
In dieser Arbeit wurden Verfahren der Graphentheorie verwendet, um für einseitige paarweise Hypothesen die relevanten Schnitthypothesen zu bestimmen. Dieser Ansatz konnte dahingehend noch erweitert werden, daß auch für verschobene einseitige Hypothesen die relevanten Schnitthypothesen bestimmt werden können. Somit sind insbesondere auch Hypothesensysteme behandelbar, die Äquivalenzhypothesen beinhalten. Außerdem kann der graphentheoretische Ansatz auch verwendet werden, um die dynamische Shaffer-Prozedur (auch Shaffer 2 Prozedur genannt) anschaulich für zwei- oder einseitige Hypothesen durchzuführen. Für eine kleine Anzahl an Gruppen ($k < 6$) läßt sich so mit „Papier und Bleistift“ eine der Holm-Prozedur i.a. überlegene

α-Adjustierung duchführen.
Im Rahmen einer Anwendung - Bestimmung der höchsten effektiven Dosis - wurden verschiedene Verfahren inklusive Verfahren, die auf dem AT basieren, verglichen. Hierbei erwiesen sich die von Bauer und Budbe vorgeschlagenen Verfahren als die mit der höchsten Güte. Die Wahl des Verfahrens hängt einzig von der Annahme bezüglich der Alternative ab, ob sie geordnet ist oder nicht.

In dieser Arbeit wurden die trim- und die Huber-m-Schätzer mit multiplen Kontrasten verwendet. Weitergehende Untersuchungen könnten sich mit der Verwendung dieser Schätzer in anderen Testverfahren, wie zum Beispiel dem F-Test oder einer Erweiterung auf mehrfaktorielle Versuche beschäftigen.
Da hier nur der Fall von einseitigen Mittelwertvergleichen explizit behandelt wurde, sei noch erwähnt, daß die Vorgehensweise sich direkt auf alle Situationen übertragen läßt, in denen die Meßgröße für die zu vergleichenden Effekte stetig ist. So ist das Verfahren z.B. auch zum paarweisen Vergleich von mehreren Eintrittswahrscheinlichkeiten verwendbar. Eine weiterführende Frage bezüglich des Testens einseitiger Hypothesen ist, ob eines der beiden Verfahren - das nach Westfall bzw. Bernhard oder die Verwendung das Abschlußests mit einseitigen Hypothesen - das andere bezüglich der Power dominiert.

Im Rahmen dieser Arbeit sind zudem SAS Macros erstellt wurden. So steht ein Macro zur Verfügung, mit dem sich die robusten Zwei-Stichproben-Tests in gleicher Weise wie mit der PROC TTEST aus SAS durchführen lassen. Ein weiteres Macro ermöglicht die Berechnung der p-Werte der einzelnen Kontraste eines multiplen Kontrasttests. Weiterhin ist für Systeme von einseitigen paarweisen Hypothesen ein Macro erstellt worden, das den Abschlußtest mit Bonferroni Globaltests für diese Hypothesen durchführt und die adjustierten p-Werte der einzelnen Hypothesen liefert.

Kapitel 7

Software

In diesem Kapitel sind die SAS MACROS, mit denen sich die Tests aus den Abschnitten 3.2.1 - 3.2.2 und dem Abschnitt 3.3.1 durchführen lassen zusammengestellt. Weiterhin ist das SAS/IML Programm dokumentiert, mit dem sich die Prozedur P4B mit streng erschöpfenden Indexmengen durchführen läßt. Das Progamm liefert die adjustierten p-Werte für den Abschlußtest mit Bonferroni-Globaltests und kann beliebige Kombinationen von einseitigen Hypothesen verarbeiten. Alle Programme sind auch über die Hompage des LG Bioinformatik am Fachbereich Gartenbau der Universität Hannover verfügbar und auf der beiliegenden Diskette im Verzeichnis „Programme“ zu finden. (http://www.bioinf.uni-hannover.de/index.html)

7.1 Macro zur Berechnung des getrimmten- und huber-t-Tests inklusive der Minimum-Tests

Das Macro ROBTTEST hat die folgenden Übergabeparameter:

- var: Name der auszuwertenden Variable
- class: Name der Variablen, die die Gruppen bezeichnet
- data: Name des Datensatzes (default = letzter verwendeter Datensatz)
- N_BOOT: Anzahl der Bootstrapwiederholungen (default=10000)
- c: Tuningkonstante (default = 1,8)
- trim: Anteil der zu trimmenden Daten (default = 0,1)

Mit dem Beispieldatensatz:

```
DATA test;
INPUT gruppe wert @@;
CARDS;
1 1 1 2 1 3 1 4 1 5
2 1 2 2 2 3 2 4 2 5 2 6 2 7 2 8 2 9 2 10
;
```

liefert der Aufruf:

```
%ROBTTEST( data=test, var=wert, class=gruppe, N_BOOT=10000, trim=.2, c=2.0);
```

die folgende, der der PROC TTEST analoge Ausgabe:

ROBTTEST MACRO

Variable: wert

Test	gruppe	N	Mean	|T|	DF	Prob>|T|
20%-trim	1	5	3.000000	1.371758	7	.21249
	2	10	5.500000			
huber c=2.0	1	5	3.000000	1.469140	13	.14805
	2	10	5.500000			

Number of bootstrap replications: 10000

Test	p-value
min t,10%-,20%-trim	.33360
min t,c=1.8	.27200
min t,20%-trim,c=1.8	.33280

Abbildung 7.1: Ausgabe des Macros ROBTTEST

```
%MACRO ROBTTEST( var, class, data = _LAST_, N_BOOT = 10000, c=1.8, trim = .1);
  ****************************;
  * Sortieren des Datensatzes *;
  ****************************;
  PROC SORT DATA=&data;
    BY &class;
  RUN;
  PROC IML;
    ****************************************;
    * Einlesen des Datensatzes in eine Matrix *;
    ****************************************;
    boot = 1;
    USE  &data;
    READ ALL VAR {&var} INTO owerte;
    READ ALL VAR {&class} INTO class;
    anz_beob = NROW( owerte);
    df_f     = anz_beob - 2;
    count    = 1;
    DO UNTIL(( class[ count] ^= class[ count + 1]));
      count = count + 1;
    END;
    n_i = count // df_f+2-count;
    ma  = MAX( n_i);
```

```
werte = j( ma, 2 , .);
werte[ 1: n_i[ 1], 1] = owerte[ 1:n_i[ 1]];
werte[ 1: n_i[ 2], 2] = owerte[ n_i[ 1]+1:df_f+2];

unten = FLOOR( &trim * n_i);
oben  = n_i - unten;
df_t  = SUM( oben - unten) - 2;
t_m   = .;
t_m_b = .;
mittel_t = .;
mittel_m = .;
dummy = .;

******************;
* globale arrays *;
******************;
mediane   = j( 1, 2, .);
med_ab_b  = j( ma, 2, .);
sigma_b   = 0;

*********;
* TTEST *;
*********;

START t_test( owerte) GLOBAL( n_i, ma);

  mittel  = owerte[ +, ] # T( n_i ## (-1));
  z_werte = owerte - j( ma, 1, 1) * mittel;
  var     = ( SSQ( z_werte[ 1:n_i[ 1], 1]) + SSQ( z_werte[ 1:n_i[ 2], 2]))
            / ( SUM( n_i) - 2);
  var     = var * SUM( n_i ## (-1));

  zaehler = mittel[ 1] - mittel[ 2];

  td_e    = zaehler /SQRT( var);

  RETURN( td_e);

FINISH; * t_test *;

******************;
* YUEN TRIMM-TEST *;
******************;

START trimz( owerte, unten, oben, mittelw) GLOBAL( n_i, ma);

  t_werte = j( ma, 2, 0);
  raenge  = j( ma, 2, .);
  werte   = owerte;
  grenzen = j( 2, 2, .);
  var_gr  = j( 1, 2, .);
  trim    = T( unten);
  ********************************************;
  * Berechnung der Schaetzung von mu und sigma *;
  ********************************************;
  DO sp = 1 TO 2;
    raenge[ 1:n_i[ sp], sp] = RANK( werte[ 1:n_i[ sp], sp]);
             * Raenge in jeder Gruppe bestimmen;
  END;
  DO sp = 1 TO 2;
    DO ze = 1 TO n_i[ sp];
      tw = raenge[ ze, sp];
      IF( unten[ sp] < tw & tw <= oben[ sp]) THEN DO;
        IF( tw = unten[ sp]+1) THEN grenzen[ 1, sp] = werte[ ze, sp];
        IF( tw = oben[ sp]) THEN grenzen[ 2, sp] = werte[ ze, sp];
        t_werte[ ze, sp] = 1;
      END;
    END;
  END;
  werte    = werte # t_werte;
  mittelw  = werte[ +, ];
  w_mittel = mittelw + trim # grenzen[ +, ];
  divisor  = T( oben - unten);
  mittelw  = mittelw # ( divisor ## (-1));
  w_mittel = w_mittel # ( T( n_i)  ## (-1));
  ****************************************;
```

```
  * Berechnung der verschiedenen 2. Momente *;
  ******************************************;
  ssd_1    = ( werte - j( ma, 1, 1) * w_mittel) # t_werte;
  ssd_2    = ( grenzen - j( 2, 1, 1) * w_mittel);
  DO sp = 1 TO 2;
    var_gr[ sp]   = ( SSQ( ssd_1[ , sp]) + SSQ( ssd_2[ , sp]) * trim[ sp]);
  END;
  ****************************;
  * Berechnung der Testgroessen *;
  ****************************;
  zaehler  = mittelw[ 1] - mittelw[ 2];
  s_y      = SQRT( SUM( var_gr) / ( SUM(divisor) - 2) * SUM( divisor ## (-1)));
  t_y      = zaehler / s_y;
  RETURN( t_y);
FINISH; * trimz *;
*******************************************;
* Initialisierung f\"{u}r den Huber m-Schaetzer *;
*******************************************;
START init( owerte, sigma_b, med_ab_b, mediane) GLOBAL( n_i, ma);
  med_ge    = j( SUM( n_i), 1, .);
  ******************************************************;
  * Berechnung der ersten Schaetzung von mu und sigma *;
  ******************************************************;
  DO i = 1 TO 2;
    mediane[ ,i]  = MEDIAN( owerte[ 1:n_i[ i], i]);
  END;
  zentren  = j( ma, 1, 1) * mediane;
  med_ab   = owerte - zentren;
  med_ge[ 1:n_i[ 1]] = med_ab[ 1:n_i[ 1], 1];
  base = n_i[ 1];
  med_ge[ base+1:SUM( n_i)] = med_ab[ 1:n_i[ 2], 2];
  sigma_b  = 1.483 * MEDIAN( ABS( med_ge));
  IF( sigma_b = 0) THEN RETURN( -1);
  ELSE DO;
    med_ab_b = med_ab / sigma_b;
    RETURN( 0);
  END;
FINISH; * init *;
START huber( owerte, c, med_ab_b, sigma_b, mediane, tz, mittelw) GLOBAL( ma, n_i);
  design  = j( ma, 2, 1);
  DO i = 1 TO 2;
    IF( ma > n_i[ i]) THEN design[ n_i[ i]+1:ma, i] = .;
  END;
  beta      = 2 * c * c * ( 1 - PROBNORM( c)) + 2 * PROBNORM( c) - 1
              - SQRT( 2 / 3.1415 ) * c * exp( -.5 * c * c);
  anz_beob  = SUM( n_i);
  df_f      = anz_beob - 2;
  phi       = j( ma, 2, 1);
  phi_s     = j( ma, 2, 1);
  phi_s  = ( ABS( med_ab_b) <= j( ma, 2, c)) # design;
  phi    = phi_s # med_ab_b + SIGN( med_ab_b) # ( j( ma, 2, 1) - phi_s)
           # j( ma, 2, c) # design;
  summe   = phi[ +, ];
  summe_s = phi_s[ +, ];
  IF( summe_s[><] = 0) THEN DO;
    RETURN( -1);
  END;
  ELSE DO;
    summe_s = summe_s ## ( -1);
    phi_q   = phi ## 2;
    summe_q = SUM( phi_q);
    mittelw = mediane + summe # summe_s * sigma_b;
    sigma_q = sigma_b * summe_q / ( ( anz_beob - 1) * beta);
```

```
    sigma   = SQRT( sigma_q);
    ****************************************;
    * Bis hier Start m-Schaetzer Berechnung *;
    ****************************************;
    med_ab = ( owerte - j( ma, 1, 1) * mittelw) / sigma;
    phi_s  = ( ABS( med_ab) <= j( ma, 2, c)) # design;
    phi    = phi_s # med_ab + SIGN( med_ab) # ( j( ma, 2, 1) - phi_s)
             # j( ma, 2, c) # design;
    summe_q  = SUM( phi ## 2);
    summe_s  = SUM( phi_s);
    IF( summe_s = 0) THEN DO;
      RETURN( -1);
    END;
    ELSE DO;
      zwischen = SUM( phi_s - ( ( summe_s / anz_beob) * j( ma, 2, 1)) ## 2)
                 / summe_s;
      kappa    = 1 + ( zwischen * zwischen * 2 / anz_beob);
      f_dach   = anz_beob * anz_beob * summe_q * kappa * kappa
                 / ( df_f * summe_s * summe_s);
      ******************************;
      * Berechnung der Testgroesse *;
      ******************************;
      tgr   =  (( mittelw[ 1] - mittelw[ 2]) / ( sigma * SQRT( f_dach)))
               #( SQRT( SUM ( n_i ## (-1))) ## (-1));
      te    = tgr;
      tz    = ABS( tgr);
      RETURN( 0);
    END;
  END;
FINISH; * huber *;

t_z    = ABS( t_test( werte));
pt_z   = PROBT( t_z, df_f);

t_t    = ABS( trimz( werte, unten, oben, mittel_t));
pt_t   = PROBT( t_t, df_t);

IF( &trim = 0.1) THEN DO;
  unten1 = unten;
  oben1  = oben;
  df_t_1 = df_t;
  t_t_1  = t_t;
  pt_t_1 = pt_t;
END;
ELSE DO;
  unten1 = FLOOR( 0.1 * n_i);
  oben1  = n_i - unten1;
  df_t_1 = SUM( oben1 - unten1) - 2;
  t_t_1  = ABS( trimz( werte, unten1, oben1, dummy));
  pt_t_1 = PROBT( t_t_1, df_t_1);
END;
IF( &trim = 0.2) THEN DO;
  unten2 = unten;
  oben2  = oben;
  df_t_2 = df_t;
  t_t_2  = t_t;
  pt_t_2 = pt_t;
END;
ELSE DO;
  unten2  = FLOOR( 0.2 * n_i);
  oben2   = n_i - unten2;
  df_t_2  = SUM( oben2 - unten2) - 2;
  t_t_2   = ABS( trimz( werte, unten2, oben2, dummy));
  pt_t_2  = PROBT( t_t_2, df_t_2);
END;

error   = 0;
error_f = 0;
ok_b = init( werte, sigma_b, med_ab_b, mediane);
IF( ok_b = -1) THEN error = 1;
ELSE DO;
```

```
  ok_b = huber( werte, &c, med_ab_b, sigma_b, mediane, t_m_b, mittel_m);
  IF( ok_b = -1) THEN error_b = 1;
  ELSE pt_t_mb = PROBT( t_m_b, df_f);
  ok = huber( werte, 1.8, med_ab_b, sigma_b, mediane, t_m, dummy);
  IF( ok = -1) THEN error_f = 1;
  ELSE DO;
    pt_t_m  = PROBT( t_m, df_f);
    tmax_tc = MAX( t_z, t_m);
    pmax_3  = MAX( pt_z, pt_t_m, pt_t_2);
  END;
END;

p_ttest = 2 * ( 1 - pt_z);
p_trim  = 2 * ( 1 - pt_t);
p_trim1 = 2 * ( 1 - pt_t_1);
p_trim2 = 2 * ( 1 - pt_t_2);
p_m     = 2 * ( 1 - pt_t_m);

IF( boot = 1) THEN DO;

  pmax_ttt = MAX( pt_z, pt_t_1, pt_t_2);

  al_tttb = 0;
  al_tcb  = 0;
  al_3b   = 0;

  *************;
  * Bootstrap *;
  *************;

  DO wdh_b = 1 TO &N_BOOT;

    boot_pos = RANUNI( j( anz_beob, 1, -1));
    boot_hil = INT( anz_beob * boot_pos + j( anz_beob, 1, 1));
    werte_h  = owerte(|boot_hil,|);
    werte    = j( ma, 2, .);
    pos      = 0;
    DO i = 1 TO 2;
      werte[ 1:n_i[ i], i] = werte_h[ pos + 1:pos + n_i[ i]];
      pos = pos + n_i[ i];
    END;

    error_l = 0;
    IF(( error ^= 1 & error_f  ^= 1)) THEN DO;
      t18zb = 0;
      x = init( werte, sigma_b, med_ab_b, mediane);
      IF((x  ^= -1)) THEN DO;
        ok = huber( werte, 1.8, med_ab_b, sigma_b, mediane, t18zb, dummy);
        IF( ok = -1) THEN DO;
          error_l = 1;
          wdh_b   = wdh_b - 1;
        END;
      END;
      ELSE DO;
        error_l = 1;
        wdh_b   = wdh_b - 1;
      END;
    END;
    IF(( error_l  ^= 1)) THEN DO;

        t_zb   = ABS( t_test( werte));
        t_t1zb = ABS( trimz( werte, unten1, oben1, dummy));
        t_t2zb = ABS( trimz( werte, unten2, oben2, dummy));

        pbt_tz   = PROBT( t_zb, df_f);
        pbt_t18z = PROBT( t18zb, df_f);
        pbt_t1z  = PROBT( t_t1zb, df_t_1);
        pbt_t2z  = PROBT( t_t2zb, df_t_2);

        bmax_ttt = MAX( pbt_tz, pbt_t1z, pbt_t2z);
        bmax_tc  = MAX( t_zb, t18zb);
        bmax_3   = MAX( pbt_tz, pbt_t18z, pbt_t2z);

        IF( pmax_ttt > bmax_ttt) THEN al_tttb = al_tttb + 1;
        IF( tmax_tc  > bmax_tc ) THEN al_tcb  = al_tcb + 1;
        IF( pmax_3   > bmax_3  ) THEN al_3b   = al_3b + 1;
```

```
      END;
    END;
    al_tttb = al_tttb / &N_BOOT;
    al_tcb  = al_tcb / &N_BOOT;
    al_3b   = al_3b / &N_BOOT;
    p_tttb  = 2 * ( 1 - al_tttb);
    p_tcb   = 2 * ( 1 - al_tcb);
    p_3b    = 2 * ( 1 - al_3b);
  END;
  PRINT 'ROBTTEST MACRO';
  FILE PRINT;
  PUT 'Variable: ' " &var";
  PRINT,;
  PUT 'Test' @15 "&class" @28 'N' @40 'Mean' @55 '|T|' @65 'DF' @73 'Prob>|T|';
  DO i = 1 TO 80;
    PUT '-' @;
  END;
  PUT;
  t = 100 * &trim;
  PUT t 2.0 "%-trim" @15 (class[ 1]) 6.0 @23 (n_i[ 1]) 6.0 @32 (mittel_t[ 1])
      12.6 @48 t_t 10.6 @;
  PUT @61 df_t 6.0 @75 p_trim 6.5;
  PUT @15 (class[ anz_beob]) 6.0  @23 (n_i[ 2]) 6.0 @32 (mittel_t[ 2]) 12.6 ;
  PUT;
  PUT 'huber c=' "&c" @15 (class[ 1]) 6.0 @23 (n_i[ 1]) 6.0 @32 (mittel_m[ 1])
      12.6 @48 t_m_b 10.6 @;
  PUT @61 df_f 6.0 @75 p_m 6.5;
  PUT @15 (class[ anz_beob]) 6.0 @23 (n_i[ 2]) 6.0 @32 (mittel_m[ 2]) 12.6;
  PUT;
  PUT 'Number of bootstrap replications: ' "&N_BOOT";
  PUT;
  PUT 'Test' @30 'p-value';
  DO i = 1 TO 36;
    PUT '-' @;
  END;
  PUT;
  PUT 'min t,10%-,20%-trim' @31 p_tttb 6.5;
  PUT 'min t,c=1.8' @31 p_tcb 6.5;
  PUT 'min t,20%-trim,c=1.8' @31 p_3b 6.5;
  CLOSEFILE PRINT;
  QUIT;
%MEND ROBTTEST;
```

7.2 Robuste multiple Kontrasttests

Zur Kalkulation der p-Werte werden in dem Macro die Quantile der multivariaten t-Verteilung mit der Prozedur *mvt_lr* aus Bretz (1999) berechnet.
Das Macro ROBKONTR hat die folgenden Übergabeparameter:

- var: Name der auszuwertenden Variable
- class: Name der Variablen, die die Gruppen bezeichnet
- data: Name des Datensatzes (default = letzter verwendeter Datensatz)
- contrast: Name des Datensatzes, der die zu betrachtenden Kontraste enthält
- c: Tuningkonstante (default = 1,8)
- trim: Anteil der zu trimmenden Daten (default = 0,2)

Mit dem Beispieldatensatz:

```
DATA test;
INPUT gruppe wert @@;
CARDS;
1 1 1 2 1 3 1 4 1 5
2 1 2 2 2 3 2 4 2 5 2 6 2 7 2 8 2 9 2 10
3 2 3 2 3 3 3 5 3 5
;
```

und den einseitigen Dunnett-Kontrasten bezüglich Guppe 1:

```
DATA con;
INPUT c d1 d2;
CARDS;
-1 1 0
-1 0 1
;
```

liefert der Aufruf:

```
%ROBKONTR( data=test, var=wert, class=gruppe, contrast=con, trim=.2, c=1.8);
```

die folgende Ausgabe:

```
                                                    ROBKONTR MACRO

     Variable: wert
     Huber-m-estimation c=1.8


     Contrast                  T          Prob>T
     ------------------------------------------
          -1.0  1.00  0.00    1.89000     0.07301
          -1.0  0.00  1.00    0.34000     0.50860


     0.2-trim-estimation


     Contrast                  T          Prob>T
     ------------------------------------------
          -1.0  1.00  0.00    1.49000     0.06759
          -1.0  0.00  1.00    0.28000     0.50513
```

Abbildung 7.2: Ausgabe des Macros ROBKONTR

```
%MACRO ROBKONTR( data=daten, class=gruppen, var=wert, contrast=contrast, c=1.8, trim=0.2);
PROC SORT DATA=&data;
  BY &class;
RUN;
PROC SORT OUT = gruppen NODUPKEY;
  BY &class;
RUN;
PROC UNIVARIATE NOPRINT;
  VAR &var;
  OUTPUT OUT=anz_g  N=n;
RUN;
DATA anz_g;
  SET anz_g;
  CALL SYMPUT('Anz_g',n);
RUN;
PROC IML;
  USE work.&data;
  READ ALL VAR{&class} INTO owerte_g;
  READ ALL VAR{&var} INTO owerte;
  USE work.&contrast;
  READ ALL INTO contrast;
  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 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;
  RETURN( intval);
FINISH;
START trim( werte, mittel_2, sigma_2, divisor2, anteil) GLOBAL( gruppen, n_i, max_n);
  unten_2 = INT( anteil * n_i);
  oben_2  = n_i - unten_2;
  unten_2 = unten_2 + 1;
  grenzen2 = j( 2, gruppen, 0);
  t_werte2 = j( max_n, gruppen, 0);
  raenge   = j( max_n, gruppen, .);
  sigmal_2 = j( 1, gruppen, 0);
  ********************************************;
  * Berechnung der Schaetzung von mu und sigma *;
  ********************************************;
  DO sp = 1 TO gruppen;
    raenge[ 1:n_i[ sp], sp] = RANK( werte[ 1:n_i[ sp], sp]);
                         * Raenge in jeder Gruppe bestimmen;
  END;
  DO sp = 1 TO gruppen;
    DO ze = 1 TO n_i[ sp];
      tw = raenge[ ze, sp];
      IF( unten_2[ sp] <= tw & tw <= oben_2[ sp]) THEN DO;
        IF( tw = unten_2[ sp]) THEN grenzen2[ 1, sp] = werte[ ze, sp];
        IF( tw = oben_2[ sp]) THEN grenzen2[ 2, sp] = werte[ ze, sp];
        t_werte2[ ze, sp] = 1;
      END;
    END;
  END;
  werte_2  = werte # t_werte2;
  mittel_2 = werte_2[ +, ];
  w_mitt_2 = mittel_2 + ( unten_2 - 1) # grenzen2[ +, ];
  divisor2 = oben_2 - unten_2;
  mittel_2 = mittel_2 / ( divisor2 + 1);
  w_mitt_2 = w_mitt_2 / n_i;
  ****************************************;
  * Berechnung der verschiedenen 2. Momente *;
  ****************************************;
  ssd_12   = ( werte - j( max_n, 1, 1) * w_mitt_2) # t_werte2;
  ssd_22   = ( grenzen2 - j( 2, 1, 1) * w_mitt_2);
  DO sp = 1 TO gruppen;
    sigmal_2[ sp]  =  SSQ( ssd_12[ , sp]) + SSQ( ssd_22[ , sp])
                      * ( unten_2[ sp] - 1);
  END;
  sigma_2 = SQRT( SUM( sigmal_2) / SUM( divisor2));
  divisor2 = divisor2 + 1;
FINISH;
START init( werte, sigma_b, med_ab_b, mediane) GLOBAL( gruppen, n_i, max_n, n_ges);
  med_ge    = j( n_ges, 1, .);
  ***************************************************;
  * Berechnung der ersten Schaetzung von mu und sigma *;
  ***************************************************;
  DO i = 1 TO gruppen;
    mediane[ ,i]  = MEDIAN( werte[ 1:n_i[ i], i]);
  END;
  zentren  = j( max_n, 1, 1) * mediane;
  med_ab   = werte - zentren;
  med_ge[ 1:n_i[ 1]] = med_ab[ 1:n_i[ 1], 1];
  base = n_i[ 1];
  DO i = 2 TO gruppen-1;
    med_ge[ base+1:base+n_i[ i]] = med_ab[ 1:n_i[ i], i];
    base = base + n_i[ i];
  END;
```

```
  med_ge[ base+1:n_ges] = med_ab[ 1:n_i[ gruppen], gruppen];
  sigma_b  = 1.483 * MEDIAN( ABS( med_ge));
  IF( sigma_b = 0) THEN RETURN( -1);
  ELSE DO;
    med_ab_b = med_ab / sigma_b;
    RETURN( 0);
  END;
FINISH;
START huber( werte, c, med_ab_b, sigma_b, mediane, mittel, sigma)
GLOBAL( design, max_n, n_i, gruppen, n_ges, beta, df_f);
  phi       = j( max_n, gruppen, 1);
  phi_s     = j( max_n, gruppen, 1);
  phi_s  = ( ABS( med_ab_b) <= j( max_n, gruppen, c)) # design;
  phi    = phi_s # med_ab_b + SIGN( med_ab_b) # ( j( max_n, gruppen, 1) - phi_s)
           # j( max_n, gruppen, c) # design;
  summe   = phi[ +, ];
  summe_s = phi_s[ +, ];
  IF( summe_s[><] = 0) THEN DO;
    RETURN( -1);
  END;
  ELSE DO;
    summe_s = summe_s ## ( -1);
    phi_q   = phi ## 2;
    summe_q = SUM( phi_q);
    mittelw = mediane + summe # summe_s * sigma_b;
    sigma_q = sigma_b * summe_q / ( ( n_ges - 1) * beta);
    sigma   = SQRT( sigma_q);
    ****************************************;
    * Bis hier Start m-Schaetzer Berechnung *;
    ****************************************;
    med_ab = ( werte - j( max_n, 1, 1) * mittelw) / sigma;
    phi_s  = ( ABS( med_ab) <= j( max_n, gruppen, c)) # design;
    phi    = phi_s # med_ab + SIGN( med_ab) # ( j( max_n, gruppen, 1) - phi_s)
             # j( max_n, gruppen, c) # design;
    summe_q  = SUM( phi ## 2);
    summe_s  = SUM( phi_s);
    IF( summe_s = 0) THEN DO;
      RETURN( -1);
    END;
    ELSE DO;
      zwischen = SUM( phi_s - ( ( summe_s / n_ges) * j( max_n, gruppen, 1)) ## 2)
                 / summe_s;
      kappa    = 1 + ( zwischen * zwischen * gruppen / n_ges);
      f_dach   = n_ges * n_ges * summe_q * kappa * kappa
                 / ( df_f * summe_s * summe_s);
      sigma    = sigma * SQRT( f_dach);
      mittel   = mittelw;
      RETURN( 0);
    END;
  END;
FINISH;
j = 1;
gruppen = &Anz_g;
DO i = 1 TO gruppen - 1;
g_val = owerte_g[ j];
n = 0;
  DO UNTIL(g_val ^= owerte_g[ j]);
    j = j + 1;
    n = n + 1;
  END;
n_i = n_i || n;
END;
n_ges    = NROW( owerte);
n_i = n_i || (n_ges - SUM(n_i));
```

```
max_n    = MAX( n_i);
df_f     = n_ges - gruppen;

c        = &c;
beta     = 2 * c * c * ( 1 - PROBNORM( c)) + 2 * PROBNORM( c) - 1
           - SQRT( 2 / 3.1415 ) * c * exp( -.5 * c * c);

design  = j( max_n, gruppen, 1);
werte   = j( max_n, gruppen, .);
base = 0;
DO i = 1 TO gruppen;
  IF( max_n > n_i[ i]) THEN design[ n_i[ i]+1:max_n, i] = .;
  werte[ 1:n_i[ i], i] = owerte[ base + 1: base + n_i[ i], 1];
  base = base + n_i[ i];
END;

mediane  = j( 1, gruppen, .);
med_ab_b = j( max_n, gruppen, .);
sigma_b  = 0;
sigma_h  = 0;
mittel_h = j( 1, gruppen, 0);
sigma_t  = 0;
mittel_t = j( 1, gruppen, 0);

ERROR = 0;
ok = init( werte, sigma_b, med_ab_b, mediane);
IF( ok = -1) THEN ERROR = 1;
ELSE DO;
  ok = huber( werte, c, med_ab_b, sigma_b, mediane, mittel_h, sigma_h);
  IF( ok = -1) THEN ERROR = 1;
  ELSE DO;

    CALL trim( werte, mittel_t, sigma_t, divisor2, &trim);

    tuk_zh  = contrast * T( mittel_h);
    tuk_zt  = contrast * T( mittel_t);
    tuk_n   = SQRT(( contrast ## 2) * T( n_i ## (-1)));
    tuk_n_t = SQRT(( contrast ## 2) * T( divisor2 ## (-1)));
    t_tuk_h = MAX( .1, ROUND( MAX(( tuk_zh / tuk_n) / sigma_h), 0.01));
    t_tuk_t = MAX( .1, ROUND( MAX(( tuk_zt / tuk_n_t) / sigma_t), 0.01));

    th      = ROUND( ( tuk_zh / tuk_n) / sigma_h, 0.01);
    tt      = ROUND( ( tuk_zt / tuk_n_t) / sigma_t, 0.01);
    df_t    = SUM( divisor2);

    RUN CORR( contrast, n_i);
    r1 = r;
    RUN CORR( contrast, divisor2);
    r2 = r;

    dim = NROW( contrast);
    p_werte = j( dim, 2, 0);
    DO i = 1 TO dim;
      th_l = j( 1, dim,th[ i]);
      p_werte[ i, 1] = 1 - mvt_lr( df, th_l,r1, 0.001);
      tt_l = j( 1, dim, th[ i]);
      p_werte[ i, 2] = 1 - mvt_lr( df_t, tt_l,r2, 0.001);
    END;

  END;
END;

PRINT 'ROBKONTR MACRO';

FILE PRINT;
PUT 'Variable: ' "&var";
PUT 'Huber-m-estimation c=' "&c";
PRINT,;
PUT 'Contrast' +(6*gruppen-2) 'T' +10 'Prob>T';
DO i = 1 TO (6*gruppen+24);
  PUT '-' @;
END;
PUT;
DO i = 1 TO dim;
  DO j = 1 TO gruppen;
    PUT @(j*6) (contrast[ i, j] ) 4.2@;
```

```
    END;
    PUT (th[ i]) 10.5 +4 (p_werte[ i, 1]) 7.5;
  END;
  PUT;
  PUT;
  PUT "&trim" '-trim-estimation';
  PRINT,;
  PUT 'Contrast' +(6*gruppen-2) 'T' +10 'Prob>T';
  DO i = 1 TO (6*gruppen+24);
    PUT '-' @;
  END;
  PUT;
  DO i = 1 TO dim;
    DO j = 1 TO gruppen;
      PUT @(j*6) (contrast[ i, j] ) 4.2@;
    END;
    PUT (tt[ i]) 10.5 +4 (p_werte[ i, 2]) 7.5;
  END;
QUIT;
%MEND;
```

7.3 Adjustierte p-Werte einseitiger Paarvergleiche

Damit das Macro korrekte Ergebnisse liefert, müssen die zu vergleichenden Mittelwerte von 1 bis k durchnummeriert sein.
Das Macro P4B_EINS hat die folgenden Übergabeparameter:

- p_wert: Variable für den p-Wert des einseitigen Tests
- gross: Variable, die die Indizes der in der Nullhypothese größereren Parameter enthält
- klein: Variable, die die Indizes der in der Nullhypothese kleineren Parameter enthält

Für das Hypothesensystem

$$H_{11} : \mu_1 \geq \mu_2, H_{12} : \mu_1 \leq \mu_2, H_2 : \mu_1 \geq \mu_3, H_3 : \mu_1 \geq \mu_4, H_4 : \mu_1 \geq \mu_5,$$
$$H_5 : \mu_2 \geq \mu_3, H_6 : \mu_2 \geq \mu_4, H_7 : \mu_2 \geq \mu_5$$

mit den p-Werten

$$p_{11} = 0,552, p_{12} = 0.448, p_2 = 0,00134, p_3 = 0,00155, p_4 = 0,0000781,$$
$$p_5 = 0,00572, p_6 = 0,00543, p_7 = 0,000709$$

hat der Eingabedatensatz die Form:

```
DATA test;
INPUT p von nach;
CARDS;
0.552 1 2
0.448 2 1
0.00134 1 3
0.00155 1 4
0.0000781 1 5
0.00572 2 3
0.00543 2 4
0.000709 2 5
;
```

Mit dem Aufruf:

```
%P4B_EINS( p_wert=p, gross=von, klein=nach )
```

erhält man die Ausgabe:

The SAS System

OBS	P_WERT	VON	NACH	ADJ_P
1	0.00008	1	5	0.00062
2	0.00071	2	5	0.00425
3	0.00134	1	3	0.00804
4	0.00155	1	4	0.00804
5	0.00543	2	4	0.01629
6	0.00572	2	3	0.01629
7	0.44800	2	1	0.89600
8	0.55200	1	2	0.89600

Abbildung 7.3: Ausgabe des Macros P4B_EINS

```
%MACRO P4B_EINS( p_wert=p, gross=von, klein=nach, data= _LAST_);
PROC SORT DATA = &data;
  BY &p_wert;
RUN;
PROC IML;
START NEXT_I( I_k, R2, anzI, i_stern, i_zstern, min_i) GLOBAL( k);
  IF(( R2 ^= 1) & (i_stern < k)) THEN DO;
      i_zstern = i_stern;
      i_stern  = i_stern + 1;
      I_k[ i_stern] = 1;
      anzI = anzI + 1;
    END;
    ELSE DO;
      IF( anzI ^= 1) THEN DO;
        anzI = anzI - 1;
        I_k[ i_stern] = 0;
        I_k[ i_zstern] = 0;
        i_stern = i_zstern + 1;
        I_k[ i_stern] = 1;
        IF( SUM( I_k) = 1) THEN DO;
          min_i = i_stern;
          i_zstern = -1;
        END;
        ELSE DO;
          min_i = I_k[<:>];
          i = i_zstern-1;
          DO WHILE( I_k[ i] ^= 1);
            i = i - 1;
          END;
          i_zstern = i;
        END;
      END;
      ELSE anzI = 0;
    END;
FINISH; * NEXT_I;
START SCHRITT2( I_k, anzI, i_stern, i_zstern, min_i, J) GLOBAL( k);
  leer = 0;
  IF( i_stern = k) THEN leer = 1;
  ELSE DO;
    M = j( i_stern, 1, 0) // j( k-i_stern, 1, 1);
    M = ( M - J[ anzI-1]) <> j( k, 1, 0);
    IF( SUM( M) = 0) THEN leer = 1;
    ELSE DO;
      i = i_stern;
      DO UNTIL( M[ i] = 1);
        i = i + 1;
      END;
      I_k[ i_stern] = 0;
      I_k[ i] = 1;
      i_stern = i;
    END;
  END;
```

```
  IF( leer = 1) THEN RUN NEXT_I( I_k, 1, anzI, i_stern, i_zstern, min_i);
FINISH; * SCHRITT2;
*****************************************;
* Hauptprogramm *************************;
*****************************************;
  * Matrix mit den p-Werten und den zugehoerigen Vergleichen;
  USE &data;
  READ ALL VAR{ &p_wert &gross &klein} INTO P;
  * Niveau;
  alpha = .05;
  * Anzahl der Vergleiche;
  k = NROW( P);
  * Anzahl der Knoten;
  n_knot = MAX(P[ , 2:3]);
  * Inzidenzmatrix der Globalhypothese;
  H0  = j( n_knot, n_knot, 0);
  * Adjustierte p-Werte;
  Ap  = P[ 1, 1] * k * j( k, 1, 1);
  * Indizes der aktuellen Schnitthypothese;
  I_k = j( k, 1, 0);
  * J wie in Gudruns Arbeit;
  J   = j( k, k+1, 0);
  * Vergleichsmatrizen fuer Wege;
  nicht = j( n_knot, n_knot, 1);
  weg   = j( n_knot, n_knot, n_knot);
  * Initialisierung;
  I_k[ 2] = 1;
  anzI = 1; * entspricht dem j in Gudruns Arbeit;
  i_stern = 2;
  i_zstern = -1;
  min_i = 2;
  DO i = 1 TO k;
    H0[ P[ i, 2], P[ i, 3]] = 1;
  END;
  * Progammschritt;
  DO UNTIL( anzI=0); * Brich ab, wenn I_k die leere Menge ist;
    * Ueberpruefung der Bedingung K1;
    IF ( J[ i_stern, anzI] = 1) THEN K1 = 1;
    ELSE DO;
      K1 = 0; * Bedingung K1 ist nicht erfuellt;
      * Initialisirung und Generierung der Inzidenzmatrix,;
      * die zu der aktuellen Schnitthypothese gehoert;
      H_k = j( n_knot, n_knot, 0);
      DO i = 1 TO k;
        IF( I_k[ i] = 1) THEN H_k[ P[ i, 2], P[ i, 3]] = 1;
      END;
      * Bestimmung der maximalen umfassenden erschoepfenden Indexmenge;
      H = H_k;
      SEH_k = H_k;
      SEH = H_k;
      i = 1;
      IF( anzI ^= 1) THEN
        DO UNTIL( (MIN( SEH - H) ^= -1) | i = anzI);
          SEH = SEH_k;
          i = i + 1;
          H = H * H_k;
          H = ( H >= j( n_knot, n_knot, 1));
          SEH_k = SEH_k <> H;
        END;
      * Check ob SEH_k Hypothesen enthaelt, die in H0 und nicht in H_k sind;
      SEH  = (SEH_k >< H0);
      CH_k = H_k - SEH;
      * Berechnung der Mengen C(I_k) und D;
```

```
      IF( SUM( CH_k) = 0) THEN DO;
        * I_k ist selbst erschoepfend;
        d     = k;
        SEI_k = I_k;
        CI_k  = j( k, 1, 0);
      END;
      ELSE DO;
        * I_k ist nicht erschoepfend;
        SEH_k = SEH;
        SEI_k = j( k, 1, 0);
        DO i = 1 TO k;
          IF( SEH_k[ P[ i, 2], P[ i, 3]] = 1) THEN SEI_k[ i] = 1;
        END;
        CI_k = SEI_k - I_k;
        D = ( CI_k - J[ , anzI]) <> j(k ,1, 0);
        IF( MAX( D) = 0) THEN d = k; * D ist die leere Menge;
        ELSE d = D[<:>]; * Finde den ersten Wert in D, der <> 1 ist;
      END;

      IF( d < i_stern) THEN K1=1; * Kriterium K2 ist erfuellt;
      ELSE DO;
        * Kriterium K2 ist nicht erfuellt;
        m = SUM( SEI_k);
        J[ , anzI+1] = CI_k;

        * Ueberpruefung, ob die Menge auch im strengen Sinn erschoepfend ist;
        GV = (( n_knot + 1) * T( H0 - SEH_k)) <> SEH_k;  * Aktueller Graph;
        GH = GV;
        G  = GV;
        i  = 1;
        zycel = 0;

        DO UNTIL( zycel > n_knot | i >= n_knot);
          GH     = GH * GV;
          i      = i + 1;
          noway  = GH < nicht;
          oneway = (GH <= weg) - noway;
          altway = GH > weg;
          GH     = oneway + ( altway * ( n_knot + 1));
          G      = G <> GH;
          zycel  = TRACE( G);
        END;

        IF( zycel <= n_knot ) THEN DO;
          Ap = (Ap <> ( SEI_k * m * P[ min_i, 1])) >< j( k, 1, 1);
        END;
      END;
    END;

    IF( K1) THEN RUN SCHRITT2( I_k, anzI, i_stern, i_zstern, min_i, J);
    ELSE RUN NEXT_I( I_k, 0, anzI, i_stern, i_zstern, min_i);

  END;

  CREATE out FROM Ap[ COLNAME='adj_p'];
  APPEND FROM Ap;
QUIT;

DATA testout;
  MERGE test out;
RUN;

PROC PRINT;
RUN;

%MEND;
```

Abbildungsverzeichnis

Literaturverzeichnis

Aboukalam, M.A.F. (1992), Some robust two-sample test statistics based on m-estimators of location, Communications in Statistics B **21(1)**, Seiten 133-148

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., Tukey, J.W. (1972), Robust estimators of location, Princeton University Press

Akritas M.G., Arnold, S.F., Brunner, E. (1997), Nonparametric hypotheses and rank statistics for unbalanced factorial designs, JASA **92**, Seiten 258ff

Bachmaier, M., Precht, M. (1995), Robust confidence intervals for contrasts based on a likelihood ratio test, Statistical Papers **36**, Seiten 215-236

Bachmaier, M., Precht, M. (1997), Robust multiple confidence intervals for contrasts, Computational Statistics & Data Analysis **25**, Seiten 25-42

Barnett V., Lewis, T. (1984), Outliers in statistical data, Wiley

Bartholomew, D.J. (1959), A test of homogeneity for ordered alternatives, Biometrika **46**, Seiten 36-48

Basu, A., Lindsay, B.G. (1994), Minimum disparity estimation for continuous models: Efficiency, distributions and robustness, Annals of the Institute of Statistical Mathematics **46**, Seiten 683-705

Basu, S., Sarkar, S., Basu, A. (1997), Robust tests for the equality of two population means under the normal model, Communications in Statistics B **26**, Seiten 333-353

Bauer, P., Hackel, P., Hommel, G., Sonnemann, E. (1986), Multiple testing of pairs of one-sided hypotheses, Metrika **33**, Seiten 121-127

Bauer, P., Budde, M. (1994), Multiple testing for detecting efficient dose steps, Biometric Journal **36(1)**, Seiten 3-15

Bauer, P., Röhmel, J., Maurer, W., Hothorn, L. (1998), Testing strategies in multi-dose experiments including active control, Staistics in Medicine **17**, Seiten 2133-2146

Bechhofer, R.E., Dunnett, C.W. (1982), Multiple comparisons for orthogonal contrasts: examples and tables, Technometrics **24.3**, Seiten 213-222

Beckman, R.J., Cook, R.D. (1983), Outlier....s (with discussion), Technometrics **25**, Seiten 119-163

Bergmann, B. (1987), Multiple Testprozeduren bei redundanten Schnitthypothesen - Modifikationen zur Erhöhung der Trennschärfe, Diplomarbeit, Mainz

Bergmann, B., Hommel, G. (1988), Improvements of general multiple test procedures for redundant systems of hypotheses, In: Bauer, P., Hommel, G., Sonnemann, E., Multiple Hypothesenprüfung - Multiple hypothesis testing, Springer, Seiten 100-115

Bernhard, G.(1991), Computergestützte Durchführung von multiplen Testprozeduren - Verbesserte Algorithmen und Powervergleich -, Dissertation, Universität Mainz

Billingsley, P. (1995), Probability and measure, Wiley

Brandt, A. (1996), Trendtests für location-scale Alternativen, Dissertation, Universität Hannover

Bretz, F. (1999), Powerful modifications of William's test on trend, Dissertation, Universität Hannover

Brunner, E., Puri, M.,L. (1996), Nonparametric methods in design and analysis of experiments, pre-print

Bühning, H., Trenkler, G. (1992), Nichtparametrische statistische Methoden, de Gryter

Carroll, R.J. (1979), On estimating variance of robust estimators when the errors are asymmetric, JASA **74**, Seiten 674-679

Conover, W.J., Iman, R.L. (1981), Rank transformations as a bridge between parametric and nonparametric statistics, The American Statistician **35(3)**, Seiten 124-132

Dixon, W.J., Tukey, J.W. (1968), Approximate behaviour of the distribution of Winsorized t (trimming/Wisorization 2), Technometrics **10**, Seiten 83-98

Dulout, F.N., Natarajan, A.T. (1987), A simple and reliable in vitro test system for the analysis of induced aneuploidy as well as other cytogenetic end-points using Chinese hamster cells Mutagenesis **2**, Seiten 121-126

Dunnett, C.W. (1955), A multiple comparison procedure for comparing several treatment with a control, Journal of the American Statistical Association **50**, Seiten 1096-1121

Dunnett, C.W. (1982), Robust multiple comparisons, Communications in Statistics A **11(22)**, Seiten 2611-2629

Efron, B., Tibshirani, R.,J. (1993), An introduction into the bootstrap, Chapmann & Hall

Hajek, J., Sidak, Z. (1967), Theory of rank tests, Academic Press

Hajek, J., Sidak, Z., Sen, P.K. (1999), Theory of rank tests, Academic Press

Hall, P., Padmanabhan, R.A. (1992), On the Bootstrap and the Trimmed Mean, Journal of Multivariate Analysis **41**, Seiten 132-153

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A. (1986), Robust statistics, Wiley

Hampel, F.R. (1995), Wozu brauchen wir Robuste Statistik?, Vortrag auf der ROeS-Tagung, Rapperswil

Hawkins, D.M. (1980), Identification of outliers, Chapmann & Hall

Hayter, A.J. (1990), A one-sided Studentized Range Test for testing against a simple ordered alternative, JASA **85.411**, Seiten 778-785

Henze, N. (1995), Stochastik II (Skriptum), Universität Karlsruhe

Hettmansperger, T.P. McKean, J.W. (1999), Robust nonparametric statistical methods, Arnold

Hirotsu, C., Kuriki, S., Hayter, A.J. (1992), Multiple comparison procedures based on the maximal component of the cumulative chi-squared statistic, Biometrika **79**, Seiten 381-392

Hoaglin, D.C. (1985), Summarising shape numerically: the g-and-h distribution, in Exploring data tables, trends, and shapes (eds. Hoaglin, Mosteller and Tukey), Wiley, Seiten 461-511

Hochberg, Y. (1988), A sharper Bonferroni procedure for multiple tests of significance, Biometrika **75**, Seiten, 800-802

Hochberg, Y., Conforti, M. (1987), Sequentially rejective pairwise testing procedures, Journal of Statistical Planning and Inference **17**, Seiten 193-208

Hochberg, Y., Tamhane, A. (1987), Multiple comparison procedures, Wiley

Holland, B.S., Copehaver, M. (1987), An improved sequentially rejective Bonferroni test procedure, Biometrics **43**, Seiten 417-423

Holm, S. (1979), A simple sequentially rejective multiple test procedure, Scandinavian Journal of Statistics **6**, Seiten 65-70

Hommel, G. (1988), A stagewise rejective multiple test procedure based on a modified Bonferroni test, Biometrika **73**, Seiten, 383-386

Hommel, G. (1989), A comparison of two modified Bonferroni procedures, Biometrika **76(3)**, Seiten 624-625

Hommel, G. (1999), Concepts for the description of logical relationships among hypotheses, Vortrag auf der AG Sitzung „Multiple Verfahren“ , Mainz 7.-8.10.1999

Hommel, G, Bernhard, G. (1991), Multiple hypothesis testing, International summer school on computational aspects of model choice, 1.-14. Juli in Prag

Hommel, G., Bernhard, G. (1999), Bonferroni procedures for logically related hypotheses, Journal of statistical planing and inference **82**, Seiten 119-128

Hommel, G., Krummenauer, F. (1998) Improvements and modifications of Tarone's multiple test procedure for discrete data, Biometrics **54**, Seiten 673-681

Horn, M., Vollandt, R. (1995), Multiple Tests und Auswahlverfahren, Gustav Fischer Verlag

Hothorn, L.A. (1994), Biostatistical analysis of the micronucleus mutagenicity assay based on the assumption of a mixing distribution, Environmental Health Perspectives **102** Supplement 1, Seiten 33-38

Hsu, J.C. (1996), Multiple comparisons, Theory and methods, Chapman and Hall

Huber, P. (1981), Robust statistics, Wiley

Huber, P. (1973), Robust regression: Asymptotics, conjectures and Monte Carlo, Annals of Statistics **1**, Seiten 799-821

Huber, P. (1984), Finite sample breakdown of m- and p-estimators, Annals of Statistics **12**, Seiten 119-126

Jaeckel, L.A. (1971), Some flexible estimates of location, The Annals of Statistics **42(5)**, Seiten 1540-1552

Jurećková, J., Sen, P.K. (1982), Simultaneous m-estimator of the common location and the scale-ratio in the two-sample problem, Math. Operationsforsch. Statist., Series Statistics **13(2)**, Seiten 163-169

Kapenga, J.A, McKean J.W., Vidmar, T.J. (1995), RGLM Robust general linear model package version 2.0 (Manual and Program)

Lee, R.E., Spurrier, J.D. (1995), Successive comparisons between ordered treatments, Journal of Statistical Planning and Inference **43**, Seiten 323-330

Léger, C., Romano, P. (1990), Bootstrap adaptive estimation: The trimmed-mean example, The Canadian Journal of Statistics **18(4)**, Seiten 297-314

Lindsay, B.G. (1994), Efficiency versus robustness: The case for minimum Hellinger distance and related methods, Annals of Statistics **22**, Seiten 1081-1114

Lucas, A. (1997), Robustness of the student t based m-estimator, Communications in Statistics A **26(5)**, Seiten 1165-1182

Marcus, R., Perlitz, E., Gabriel, K.R. (1976), On closed testing procedures with special reference to ordered analysis of variance, Biometrika **63**, Seiten 655-660

Markatou, M. (1996), Robust statistical inference: Weighted likelihood's or usual m-estimation, Communications in Statistics A **25**, Seiten 2597-2613

Mead, R., Curnow, R.N., Hasted, A.M. (1983), Statistical methods in agriculture and experimental biology, Chapman & Hall

Mudholkar, A., Mudholkar, G.S., Srivastava, D.K. (1991), A construction and appraisal of pooled trimmed t statistics, Communications in Statistics A **20(4)**, Seiten 1345-1359

Neuhäuser, M (1996),Trendtests bei a priori unbekanntem Erwartungswertprofil, Dissertation, Universität Dortmund

Neumann, K., Morlock, M. (1993), Operations Research, Hanser

Noltemeier, H. (1976), Graphentheorie mit Algorithmen und Anwendungen, de Gruyter

Öztürk, Ö. (1998), A robust and almost fully efficient m-estimator, Austral.&New Zeeland J. Statisics **40(4)**, Seiten 415-424

Pigeot-Kübler, I. (1993), Multiple Testtheorie in der Ausreißererkennung, Habilitationsschrift, Dortmund

Press, W.H. Teukolsky, S.A., Vetterling, W.T., Flannery, B.P (1995), Numerical recipes in C second edition, Cambridge University Press

Ringland, J.T. (1983), Robust multiple comparisons, JASA **78**, Seiten 145-151

Robertson, T., Wright, F.T., Dykstra, R.L. (1988), Order restricted statistical inference, Wiley

Rom, D.M., Costello, R.J., Connell, L.T. (1994), On closed test procedures for dose-response analysis, Statistics in Medicine **13**, Seiten 1583-1596

Rousseeuw, P.J., Leroy, A.M. (1988), A robust scale estimator based on the shortest half, Statistica Neerlandica 42, Seiten103-116

Schultze, V. (1997), Robuste Schätzung und Ausreißeridentifikation in exponentialverteilten Zufallsstichproben, Dissertation, Dortmund

Shaffer, J.P. (1986), Modified sequentially rejective multiple test procedures, JASA **81**, Seiten 826-831

Shorack, G.R. (1997), Uniform CLT, WLLN, LIL and bootstrapping in a data analytic approach to trimmed L-statistics, Journal of Statistical Planning and Inference **60**, Seiten 1-44

Simes, R.J. (1986), An improved Bonferroni procedure for multiple tests of significance, Biometrika **73**, Seiten 751-754

Staudte, R.G., Sheater, S.J. (1990), Robust estimation and testing, Wiley

Stigler, S.M. (1973), The asymptotic distribution of the trimmed mean, Annals of Statistics **1**, Seiten 472-477

Tiku, M.L., Tan, W.Y., Balakrishnan, N. (1986), Robust inference, Marcel Dekker

Tukey, J.W. (1953), The problem of multiple comparisons, Monographie

Tukey, J.W., MCLaughlin, D.H. (1963), Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/Winsorization 1, Sankhya Series A **25**, Seiten 331-352

Weichert, M., Hothorn, L.A. (2000), A minimum test - appropriate for normal distribution and robust against outliers, Communications in Statistics B, (eingereicht)

Welch, B.L. (1947), The generalisation of 'Student's' problem when several different population variances are involved, Biometrika **34**, Seiten 28-35

Wellmann, J. (1994), Robuste statistische Verfahren und Ausreißeridentifikation beim Modell der Einfachklassifikation mit zufälligen Effekten, Dissertation, Universität Dortmund

Westfall, P.H. (1997), Multiple testing of general contrasts using logical constraints and correlations, JASA **92.437**, Seiten 299-306

Westfall, P.H., Young S.S. (1993), Resampling-based multiple testing, Wiley

Wilcox, R.R. (1990), Comparing the means of two independent groups, Biomertic Journal **32(7)**, Seiten 771-780

Wilcox, R.R. (1994), Some results on the Tukey-McLaughlin and Yuen methods for trimmed means when distributions are skewed, Biometric Journal **36**, Seiten 259-273

Wilcox, R.R. (1997), Introduction to robust estimation and hypothesis testing, Academic Press

Wilson, M. C., Shade, R. E. (1967), Relative attractiveness of various luminescent colors to the cereal leaf beatle and the meadow spittlebug, Journal of Economic Entomology **60**, Seiten 578-580

Wright, S.P. (1992), Adjusted p-values for simultaneous inference, Biometrics **48**, Seiten 1005-1013

Yang, S.-S. (1985), On bootstrapping a class of differentiable statistical functionals with applications to l- and m- estimates, Statistica Neerlandica **39(4)**, Seiten 375-385

Yuen, K.K., Dixon, W.J. (1973), The approximate behaviour and performance of the two-sample trimmed t, Biometrika **60**, Seiten 369-374

Yuen, K.K. (1974), The two-sample trimmed t for unequal population variances, Biometrika **61**, Seiten 165-170

Yuen Fung, K., Lee, H., Tajuddin, I. (1985), Some robust test statistics for the two-sample location problem, The Statistician **35**, Seiten 175-182

Anhang A

Tabellen

A.1 Mittelwerte der einzelnen Profile aus 4.3.7.4

Profil	μ_1	μ_2	μ_3	μ_4	Profil	μ_1	μ_2	μ_3	μ_4
a	0	1,25	1,25	1,25	j2	1,25	2,75	0	1,25
b	0	1,25	1,25	0	k1	0	0	1,5	2,75
c	0	1,25	0	0	k2	0	0	1	2,25
d	0	0	1,25	1,25	l1	0	1,5	1,5	2,75
e	0	0	1,25	0	l2	0	1	1	2,25
f	1,25	0	1,25	1,25	m	0	0	0	1,25
g1	0	0,75	2	0	n1	2,75	0	1,5	2,75
g2	0	1,5	2,75	0	n2	2,25	0	1	2,25
h1	0	1,5	2,75	1	o1	0	1,5	0	1,25
h2	0	0,75	2	0,5	o2	0	1	0	1,25
h3	0	1,5	2,75	1,4	p1	0	1,5	2,85	4,1
h4	0	0,75	2	1	p2	0	1,35	2,85	4,1
h5	0	1,5	2,75	1,7	p3	0	1,35	2,35	3,6
h6	0	1	2,25	1,5	p4	0	1	2,35	3,6
h7	0	1	2,25	2,25	p5	0	1	1,75	3
h8	0	1,5	2,75	2,75	p6	0	0,75	1,75	3
i	1,25	1,25	0	1,25	q1	1,5	0	1,5	2,75
j1	1,25	2,25	0	1,25	q2	1	0	1	2,25

A.2 Simulationsergebnisse zu Abschnitt 4.3.7.4

In der Tabelle sind die Verfahren mit den Kurzformen

alle Profile

Verfahren	AT mit paarweisen Hyp.	Lee/Spurrier	Holm
Kurzform	AT	Lee	Holm
Verfahren	AT mit Bauer/Budde Hyp.	Bauer/Budde paarweise Hyp.	Hayter
Kurzform	AT/BB	BB	Hay

nur Ordnungsprofile

Verfahren	AT mit Ordnungsrestiktion	AT/BB mit Ordnungsrestriktion
Kurzform	AT_O	AT/BB_O
Verfahren	Rom et al (1994)	Bauer/Budde mit Reverse-Helmert-Kontrasten
Kurzform	Rom	BB_rh

bezeichnet. In der Spalte Power steht high für die Wahrscheinlichkeit, den HEDS zu überterschätzen, low für die Wahrscheinlichkeit, den HEDS zu unterschätzen und hit für die Power, den HEDS richtig zu erkennen. In den Ergebnissen der Tests unter Ordnungsrestriktion sind die Profile kursiv gedruckt, die nicht monotonen Profilen entsprechen.

		mit Ordnungsrestriktion			
Profil	Power	AT/BB_O	AT_O	BB_rh	Rom
a	high	0,049	0,049	0,049	0,026
	hit	0,648	0,672	0,833	0,019
	low	0,000	0,000	0,000	0,000
b	high	0,001	0,001	0,000	0,000
	hit	0,048	0,048	0,541	0,000
	low	0,000	0,000	0,000	0,000
c	high	0,008	0,008	0,019	0,000
	hit	0,012	0,041	0,149	0,000
	low	0,000	0,000	0,000	0,000
d	high	0,027	0,027	0,048	0,052
	hit	0,623	0,631	0,813	0,679
	low	0,016	0,016	0,059	0,000
e	high	0,000	0,000	0,000	0,000
	hit	0,040	0,040	0,301	0,036
	low	0,004	0,004	0,095	0,000
f	high	0,008	0,008	0,047	0,023
	hit	0,011	0,011	0,793	0,037
	low	0,000	0,000	0,000	0,000
g1	high	0,000	0,000	0,000	0,000
	hit	0,040	0,040	0,084	0,001
	low	0,010	0,010	0,563	0,000
g2	high	0,000	0,000	0,000	0,000
	hit	0,040	0,040	0,012	0,000
	low	0,010	0,010	0,934	0,000

Profil	Power	AT/BB_O	AT_O	BB_rh	Rom
h1	high	0,000	0,000	0,000	0,000
	hit	0,560	0,560	0,154	0,003
	low	0,156	0,156	0,840	0,000
h2	high	0,000	0,000	0,000	0,000
	hit	0,238	0,238	0,230	0,014
	low	0,055	0,055	0,579	0,000
h3	high	0,000	0,000	0,000	0,000
	hit	0,718	0,718	0,306	0,030
	low	0,200	0,200	0,692	0,000
h4	high	0,000	0,000	0,000	0,000
	hit	0,560	0,560	0,461	0,139
	low	0,114	0,114	0,462	0,000
h5	high	0,000	0,000	0,000	0,000
	hit	0,764	0,764	0,446	0,116
	low	0,213	0,213	0,554	0,000
h6	high	0,000	0,000	0,000	0,000
	hit	0,741	0,741	0,593	0,292
	low	0,172	0,172	0,395	0,000
h7	high	0,031	0,031	0,049	0,053
	hit	0,754	0,761	0,831	0,750
	low	0,166	0,166	0,119	0,000
h8	high	0,036	0,036	0,049	0,053
	hit	0,754	0,756	0,831	0,750
	low	0,203	0,203	0,120	0,000
i	high	0,000	0,000	0,000	0,000
	hit	0,049	0,049	0,721	0,003
	low	0,000	0,000	0,000	0,000
j1	high	0,000	0,000	0,000	0,000
	hit	0,049	0,049	0,722	0,000
	low	0,000	0,001	0,002	0,000
j2	high	0,000	0,000	0,000	0,000
	hit	0,049	0,049	0,723	0,000
	low	0,000	0,001	0,004	0,000
k1	high	0,000	0,000	0,000	0,000
	hit	0,782	0,782	0,867	0,867
	low	0,131	0,211	0,133	0,131
k2	high	0,000	0,000	0,000	0,000
	hit	0,781	0,781	0,864	0,867
	low	0,108	0,161	0,133	0,112
l1	high	0,000	0,000	0,000	0,000
	hit	0,783	0,783	0,807	0,780
	low	0,196	0,197	0,192	0,101
l2	high	0,000	0,000	0,000	0,000
	hit	0,781	0,781	0,804	0,780
	low	0,126	0,129	0,181	0,071

Profil	Power	AT/BB_O	AT_O	BB_rh	Rom
m	high	0,000	0,000	0,000	0,000
	hit	0,716	0,716	0,765	0,697
	low	0,007	0,010	0,057	0,008
n1	high	0,000	0,000	0,000	0,000
	hit	0,049	0,049	0,867	0,262
	low	0,000	0,000	0,133	0,000
n2	high	0,000	0,000	0,000	0,000
	hit	0,049	0,049	0,861	0,169
	low	0,000	0,000	0,129	0,000
o1	high	0,000	0,000	0,000	0,000
	hit	0,716	0,716	0,756	0,013
	low	0,028	0,137	0,129	0,030
o2	high	0,000	0,000	0,000	0,000
	hit	0,716	0,716	0,746	0,132
	low	0,025	0,090	0,092	0,023
p1	high	0,000	0,000	0,000	0,000
	hit	0,851	0,851	0,867	0,867
	low	0,140	0,149	0,133	0,131
p2	high	0,000	0,000	0,000	0,000
	hit	0,847	0,847	0,867	0,867
	low	0,137	0,153	0,133	0,132
p3	high	0,000	0,000	0,000	0,000
	hit	0,827	0,827	0,865	0,867
	low	0,156	0,170	0,135	0,121
p4	high	0,000	0,000	0,000	0,000
	hit	0,821	0,821	0,867	0,867
	low	0,139	0,177	0,133	0,131
p5	high	0,000	0,000	0,000	0,000
	hit	0,797	0,797	0,858	0,866
	low	0,152	0,178	0,141	0,103
p6	high	0,000	0,000	0,000	0,000
	hit	0,796	0,796	0,865	0,867
	low	0,138	0,183	0,135	0,121
q1	high	0,000	0,000	0,000	0,000
	hit	0,716	0,716	0,867	0,850
	low	0,020	0,028	0,133	0,020
q2	high	0,000	0,000	0,000	0,000
	hit	0,716	0,716	0,861	0,832
	low	0,019	0,027	0,129	0,020

		ohne Ordnungsrestriktion					
Profil	Power	AT/BB	AT	BB	Hay	Holm	Lee
a	high	0,051	0,049	0,044	0,021	0,044	0,034
	hit	0,706	0,699	0,699	0,641	0,699	0,709
	low	0,000	0,000	0,000	0,000	0,000	0,000
b	high	0,025	0,017	0,020	0,011	0,020	0,017
	hit	0,711	0,689	0,720	0,648	0,720	0,723
	low	0,000	0,000	0,000	0,000	0,000	0,000
c	high	0,021	0,013	0,024	0,010	0,024	0,017
	hit	0,632	0,640	0,707	0,643	0,707	0,714
	low	0,000	0,000	0,000	0,000	0,000	0,000
d	high	0,026	0,026	0,032	0,010	0,021	0,017
	hit	0,709	0,716	0,702	0,636	0,713	0,715
	low	0,016	0,016	0,011	0,009	0,011	0,013
e	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,710	0,684	0,722	0,639	0,722	0,721
	low	0,017	0,013	0,012	0,009	0,012	0,013
f	high	0,015	0,012	0,031	0,010	0,021	0,017
	hit	0,623	0,635	0,701	0,636	0,711	0,715
	low	0,000	0,000	0,000	0,000	0,000	0,000
g1	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,775	0,720	0,741	0,639	0,741	0,721
	low	0,133	0,136	0,116	0,135	0,116	0,137
g2	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,777	0,722	0,774	0,639	0,774	0,721
	low	0,218	0,267	0,215	0,332	0,215	0,268
h1	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,777	0,748	0,774	0,639	0,774	0,721
	low	0,218	0,241	0,215	0,332	0,215	0,268
h2	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,775	0,727	0,741	0,639	0,741	0,721
	low	0,133	0,132	0,116	0,135	0,116	0,137
h3	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,777	0,765	0,774	0,639	0,774	0,721
	low	0,218	0,226	0,215	0,332	0,215	0,268
h4	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,775	0,747	0,741	0,639	0,741	0,721
	low	0,133	0,130	0,116	0,135	0,116	0,137
h5	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,777	0,774	0,774	0,639	0,774	0,721
	low	0,218	0,219	0,215	0,332	0,215	0,268
h6	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,777	0,769	0,756	0,639	0,756	0,721
	low	0,177	0,176	0,165	0,216	0,165	0,201
h7	high	0,031	0,031	0,036	0,010	0,030	0,017
	hit	0,754	0,761	0,735	0,636	0,741	0,715
	low	0,166	0,166	0,155	0,211	0,155	0,193

Profil	Power	AT/BB	AT	BB	Hay	Holm	Lee
h8	high	0,036	0,036	0,038	0,010	0,036	0,017
	hit	0,754	0,756	0,751	0,636	0,753	0,715
	low	0,203	0,203	0,200	0,324	0,200	0,256
i	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,651	0,650	0,722	0,650	0,722	0,721
	low	0,002	0,003	0,004	0,004	0,004	0,005
j1	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,704	0,663	0,749	0,650	0,749	0,721
	low	0,087	0,127	0,106	0,140	0,106	0,134
j2	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,755	0,676	0,772	0,650	0,772	0,721
	low	0,175	0,254	0,187	0,281	0,187	0,238
k1	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,782	0,782	0,857	0,650	0,779	0,721
	low	0,132	0,212	0,131	0,321	0,210	0,267
k2	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,780	0,780	0,813	0,650	0,758	0,721
	low	0,112	0,171	0,105	0,208	0,161	0,197
l1	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,783	0,783	0,776	0,650	0,775	0,721
	low	0,196	0,197	0,189	0,285	0,190	0,243
l2	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,780	0,780	0,753	0,650	0,751	0,721
	low	0,127	0,129	0,112	0,148	0,114	0,144
m	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,719	0,719	0,726	0,650	0,724	0,721
	low	0,011	0,014	0,012	0,012	0,014	0,017
n1	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,764	0,676	0,857	0,650	0,776	0,721
	low	0,127	0,293	0,130	0,319	0,211	0,265
n2	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,715	0,664	0,813	0,650	0,756	0,721
	low	0,092	0,190	0,103	0,205	0,159	0,194
o1	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,759	0,730	0,772	0,650	0,772	0,721
	low	0,177	0,206	0,187	0,281	0,187	0,238
o2	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,731	0,723	0,749	0,650	0,749	0,721
	low	0,093	0,101	0,106	0,140	0,106	0,134
p1	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,851	0,851	0,860	0,650	0,850	0,721
	low	0,140	0,149	0,140	0,346	0,149	0,278
p2	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,847	0,847	0,863	0,650	0,847	0,721
	low	0,137	0,153	0,137	0,347	0,152	0,278

Profil	Power	AT/BB	AT	BB	Hay	Holm	Lee
p3	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,827	0,827	0,838	0,650	0,824	0,721
	low	0,156	0,170	0,154	0,325	0,168	0,271
p4	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,821	0,821	0,857	0,650	0,819	0,721
	low	0,139	0,177	0,137	0,330	0,175	0,273
p5	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,797	0,797	0,804	0,650	0,782	0,721
	low	0,152	0,178	0,140	0,235	0,162	0,223
p6	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,796	0,796	0,824	0,650	0,782	0,721
	low	0,138	0,183	0,130	0,255	0,172	0,233
q1	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,770	0,733	0,857	0,650	0,776	0,721
	low	0,127	0,243	0,130	0,319	0,211	0,265
q2	high	0,000	0,000	0,000	0,000	0,000	0,000
	hit	0,742	0,726	0,813	0,650	0,756	0,721
	low	0,094	0,162	0,103	0,205	0,159	0,194

A.3 Kuckuckseier p-Werte

Tukey Huber 20% trim Westfall Brunner Hettmansperger	Baum	Rot	Spatz	Wiese	Zaun
	0,993	0,915	0,986	0,228	0,001
	0,977	0,927	0,964	0,279	0,001
Bach	0,977	0,927	0,963	0,279	0,001
	0,993	0,914	0,987	0,232	0,001
	0,936	0,977	0,976	0,345	0,001
	0,994	0,970	0,995	0,585	0,002
		0,606	1	0,046	0,001
		0,530	1	0,030	0,001
Baum		0,530	1	0,037	0,001
		0,610	1	0,048	0,001
		0,540	1	0,024	0,001
		0,759	1	0,203	0,001
			0,567	0,900	0,001
			0,491	0,914	0,001
Rot			0,490	0,914	0,001
			0,568	0,900	0,001
			0,676	0,864	0,001
			0,785	0,983	0,018
				0,041	0,001
				0,035	0,001
Spatz				0,036	0,001
				0,042	0,001
				0,055	0,001
				0,240	0,001
					0,001
					0,001
Wiese					0,001
					0,001
					0,001
					0,019

Tabelle A.1: p-Werte der paarweisen Vergleiche mit den verschiedenen Methoden

Anhang B

Datensätze

Die in dieser Arbeit betrachteten Versuche bzw. ihre Daten sind auf der beiliegenden Diskette in Verzeichnis „Daten“ als Textfiles abgelegt. Die Datenstuktur der einzelnen Files ist im folgenden erläutert.

B.1 Radieschen

Die Daten des Radieschenversuchs sind unter dem Dateinamen „Radieschen.txt“ zu finden. In der ersten Spalte mit dem Titel „beh“ sind die vier verschiedenen Behandlungen 0,2,4 und 6 codiert. Die drei folgenden Spalten enthalten jeweils einen gemessenen Endpunkt pro Radieschen. Es wurden Gewicht (Gew), Blattfläche (Blatt) und Durchmesser (Durchm) der Radieschen gmessen.

B.2 Kohl

Die Daten des Kohlversuchs sind unter dem Dateinamen „Kohl.txt“ zu finden. Gemessen wurde das Gewicht der Kohlköpfe in fünf Wiederholungen.

Kontrolle				Behandlung					
1	2	3	4	1	2	3	4	5	6
1,63	1,73	1,49	1,25	1,07	0,73	0,69	0,52	1,63	1,08
1,48	1,42	1,70	1,36	1,28	1,42	1,59	1,50	1,34	1,33
1,43	1,50	1,52	0,93	1,28	1,30	1,61	1,17	1,07	0,63
1,76	1,06	1,48	1,38	1,83	1,38	1,61	1,05	1,01	1,21
1,17	0,76	0,85	0,68	1,16	0,70	1,17	1,04	0,87	0,51

B.3 Kuckuck

Die Kuckuckseilängen sind unter dem Dateinamen „Vogel.txt“ zu finden. In der Datei sind die Vogelnamen wie in Kapitel 2 abgekürzt.

Bachstelze	Baumpieper	Rotkehlchen	Spatz	Zaunkönig	Wiesenpieper		
21.05	21.05	21.05	20.85	19.85	19.65	22.05	22.45
21.85	21.85	21.85	21.65	20.05	20.05	22.05	22.65
21.85	22.05	22.05	22.05	20.25	20.65	22.05	22.65
21.85	22.45	22.05	22.85	20.85	20.85	22.05	22.85
22.05	22.65	22.05	23.05	20.85	21.65	22.05	22.85
22.45	23.25	22.25	23.05	20.85	21.65	22.25	22.85
22.65	23.25	22.45	23.05	21.05	21.65	22.25	22.85
23.05	23.25	22.45	23.05	21.05	21.85	22.25	23.05
23.05	23.45	22.65	23.45	21.05	21.85	22.25	23.25
23.25	23.45	23.05	23.85	21.25	21.85	22.25	23.25
23.45	23.65	23.05	23.85	21.45	22.05	22.25	23.45
24.05	23.85	23.05	23.85	22.05	22.05	22.25	23.65
24.05	24.05	23.05	24.05	22.05	22.05	22.25	23.85
24.05	24.05	23.05	25.05	22.05	22.05	22.45	24.25
24.85	24.05	23.25		22.25	22.05	22.45	24.45
		23.85					

B.4 Calibrachoa-Hybriden

In dem File „Cali.txt“ sind die Messwerte des Beleuchtungsversuchs an der LVA Ahlem abgelegt.
Spalte eins enthält die Sortencodierung (1,2,3), Spalte zwei die Buchstaben für die Beleuchtungsvariante und in der dritten Spalte ist die Kultivierungsdauer in Tagen angegeben.

B.5 Membran

Das File „Membran.txt“ beinhaltet die Daten des Permeabilitäts-Versuchs.
In der ersten Spalte ist die Behandlung notiert, wobei KON für die Kontrolle, D1 bis D3 für die drei aufsteigenden Dosierungen der Behandlung D und die restlichen Buchstaben ebenfalls für je eine Behandlung stehen.

Bevor der Buchdeckel sich schließt möchte ich mich bei all denen, die mich beim Entstehen dieser Arbeit unterstützt haben, bedanken. Insbesondere bei:

L.A. Hothorn, meinem Betreuer und Chef für die Unterstützung, Ermutigung und das Vertrauen in das Gelingen in dieser Arbeit.

G. Hommel, der trotz der Distanz zwischen Mainz und Hannover immer für Fragen ein offenes Ohr hatte.

Peter Westfall und Gudrun Bernhard, die mir den einen oder andern Denkanstoß gaben.

Tom Hettmansperger und Joe McKean, die mir unkompliziert und schnell in „letzter“ Minute mit einer Software zur Durchführung ihres Verfahrens aushalfen.

Frank Bretz und Dirk Seidel, meinen Mitdoktoranden für die kleinen und großen Diskussionen zum Thema oder auch anderen statistischen Problemen.

Hans Bernhard Wiezorke und Dirk Henkels, die mir mit ihren C-Kenntnissen aushalfen.

Andreas Fick, Christiane Stadler und Christoph Zinn, die die Arbeit als unbefangene lasen.

Clemens Buczilowski und Hanne Visser, die Computer- wie Verwaltungstechnisch mir den Rücken stärkten.

Und das alles hätte es nie gegeben, wenn nicht da noch ein Elternpaar den Grundstein dafür gelegt hätte.

Lebenslauf

Michael Weichert

geboren 31.05.1970 in Bonn
ledig

Berufstätigkeit

seit 10/1996	Lehrgebiet Bioinformatik, Universität Hannover: Wissenschaftlicher Assistent
1995-1999	Babtec Informationssysteme GmbH, Solingen: Freier Mitarbeiter
09-10/1997	Schering AG, Berlin: Praktikum mit Werkvertrag
10/1996	Prognos AG, Basel: Werkvertrag zusammen mit Herrn H.B. Wiezorke
1991-1996	Institut für Mathematische Stochastik, Universität Karlsruhe (TH): Wissenschaftliche Hilfskraft

Ausbildung

10/1989-07/1996	Studium der Wirtschaftsmathematik: Universität Karlsruhe (TH) Nebenfächer - Betriebswirtschaft mit den Schwerpunkten Finanzwirtschaft und Operations Research - Informatik
1986-1989	Staatliches Gymnasium Lauf an der Pegnitz
1983-1986	Humboldt-Gymnasium, Solingen
1980-1983	Clara-Schumann-Gymnasium, Bonn
1976-1980	Grundschule, Bonn