

Simultane Konfidenzintervalle für nichtparametrische Effekte in faktoriellen Modellen

Diplomarbeit

vorgelegt von
Sebastian Pfeiffer
aus Rostock

angefertigt am
Institut für Mathematische Stochastik
der Georg-August-Universität Göttingen

2010

Danksagung

Mein Dank gilt Herrn Prof. Dr. Edgar Brunner für die Vergabe dieser interessanten Arbeit und der Bereitstellung der hervorragenden Arbeitsmöglichkeiten in der Abteilung für Medizinische Statistik. Bei Frau Prof. Dr. Tatyana Krivobokova möchte ich mich für die Übernahme des Koreferats bedanken. Ein besonderer Dank ist auch an Herrn Dr. Frank Konietschke zu richten, für die geduldige und stetige Hilfe während der Betreuung der Diplomarbeit. Bei der Beseitigung grammatikalischer Fehler bedanke ich mich bei Stefan Birkholz.

Bei meinen Eltern bedanke ich mich für die moralische und finanzielle Unterstützung, so dass für mich ein Traum in Erfüllung gehen konnte.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Aufbau der Arbeit	2
2	Motivierende Beispiele	5
2.1	Anzahl der Corpora Lutea	5
2.2	Nierengewichte	6
2.3	Leukozyten-Migration ins Peritoneum	6
3	Grundlagen	9
3.1	Parametrische Motivation	9
3.2	Varianzanalyse	10
3.3	Multiples Testen und Multiple Kontraste	10
3.4	Relativer Effekt und Nichtparametrisches Behrens-Fisher Problem für zwei Stichproben	11
4	Nichtparametrische Modelle, Effekte und Hypothesen	15
4.1	Nichtparametrische Modelle	15
4.2	Nichtparametrische Effekte	18
4.2.1	Relative Effekte im Mehrstichprobenfall	18
4.2.2	Transitive relative Effekte	19
4.3	Kontrastmatrizen und Hypothesen	21
4.3.1	Formulierung über Erwartungswert und Verteilungsfunktion	21
4.3.2	Hypothesen in nichtparametrischen Effekten	22
4.3.3	Hypothesenkombinationen	27
4.3.4	Sequentielles Testproblem	28
5	Theorie	33
5.1	Punktschätzer und Rangdarstellungen	33
5.2	Multiple Kontrasttests	35
5.2.1	Asymptotische Verteilung	36
5.2.2	Singularität	38
5.2.3	Varianzschätzer	39
5.2.4	Simultane Testverfahren	43

5.2.5	Approximation mit der Normalverteilung	47
5.2.6	Approximation mit der t-Verteilung	48
5.2.7	Bereichserhaltende Transformation	50
5.3	ANOVA in relativen Effekten	52
6	Simulationen	55
7	Beispielauswertung	65
7.1	Anzahl der Corpora Lutea	65
7.2	Nierengewichte	67
7.3	Leukozyten-Migration ins Peritoneum	72
8	Zusammenfassung und Ausblick	75
A	Definitionen, Notationen und Sätze	77
A.1	Matrizenrechnung	77
A.2	Wahrscheinlichkeitstheorie	79

Tabellenverzeichnis

2.1	Anzahl der Corpora Lutea	5
2.2	Relative Nierengewichte	6
2.3	Anzahl der Leukozyten	7
4.1	Einfaktorielle Datenstruktur	16
4.2	Zweifaktorielle Datenstruktur	16
4.3	In b aufgetrennte Datensätze	28
4.4	In a aufgetrennte Datensätze	29
5.1	Struktur der Kovarianzmatrix \mathbf{Y}_N	40
5.2	Konsistente Kovarianzschätzer	42
6.1	Simulation: Niveau; Haupteffekt A in CRF-42; Ungleiche Varianzen und Stichprobenumfänge.	62
7.1	Simultane 95%-Konfidenzintervalle zu Corpora Lutea: Dunnett Effekt.	66
7.2	Simultane 95%-Konfidenzintervalle für die Wechselwirkungseffekte mit $\mathbf{C}_{AB} = \mathbf{C}_{Dun}^2 \otimes \mathbf{C}_{Dun}^5$ aus (7.1) unter Verwendung der t-Approximation.	69
7.3	Simultane 95%-Konfidenzintervalle zu den relativen Nierengewichten mit Dunnett-Effekt und t-Approximation.	70
7.4	Ergebnisse zu den relativen Nierengewichten, Haupteffekte und Wechselwirkung unter Verwendung der ATS.	71
7.5	Ergebnisse zu den relativen Nierengewichten, paarweise Vergleiche - Placebo vs. Dosisstufen, mit ATS.	71
7.6	Ausgesuchte simultane 95%-Konfidenzintervalle unter Normal-Approximation.	74

Abbildungsverzeichnis

4.1	Beispiel für den Verlauf der Werte von q , getrennt nach den Stufen von Faktor A, im MODELL CRF-23, mit Wechselwirkung (links) und ohne (rechts).	26
4.2	Ablauf einer Sequentiellen Testprozedur im MODELL 4.2 auf Haupteffekt A.	30
5.1	Equikoordinate bivariate Quantile von $f(N_1, N_2) \sim t_{0,E}$: zweiseitiges $1 - \alpha$ (links), obere einseitige für α und $1 - \alpha$ (rechts).	46
5.2	Zweiseitige Equikoordinate 95%Quantile von bivariat $N(\mathbf{0}, \mathbf{I}_2 + \rho(\mathbf{J}_2 - \mathbf{I}_2))$ -verteilten Zufallsvariablen für $\rho = \{0, 0.9\}$.	46
6.1	Simulation: Niveau; CRF- a , $a = \{2, 4, 8\}$, $\mathbf{C} = \mathbf{P}_a$ (links) und CRF- ab , $a = 2, b = \{2, 4, 8\}$, $\mathbf{C} = \frac{1}{a}\mathbf{1}'_a \otimes \mathbf{P}_b$ (rechts).	56
6.2	Simulation: Niveau; Haupteffekt A in verschiedenen Designs, Normal-Approximation (links), t-Approximation (rechts).	57
6.3	Simulation: Niveau; Sequentieller Test (mit multiplen Kontrasten) auf Haupteffekt A.	58
6.4	Simulation: Niveau; nach getesteter Wechselwirkung, Test auf Haupteffekt A - getrennt nach Stufen von Faktor B, mit Bonferroni-Adjustierung.	58
6.5	Simulation: Niveau; links: sequentieller Test auf Haupteffekt A mit Variante 1 (weiterhin Auftrennung der Daten; Test auf Wechselwirkung und einzelnen Haupteffekt A mit Matrix (6.1)) und rechts: Test unter Verwendung von $\mathbf{C}_{Seq,A}$ (6.2) - Variante 2 (keine Auftrennung der Daten)	59
6.6	Simulation: Niveau; Test unter Verwendung von \mathbf{C}_{Seq} (4.19) auf beliebigen Haupteffekt (Ablehnung, falls mindestens ein Haupteffekt signifikant getestet wird), CRF-22 (links oben), CRF-23 (rechts oben), CRF-25 (links unten) und CRF-222 (rechts unten).	60
6.7	Simulation: Niveau; Haupteffekt B in CRF-23, Normalverteilung (links oben), Diskrete Gleichverteilung (rechts oben), Exponentialverteilung (links unten) und Log-Normalverteilung (rechts unten).	62
6.8	Simulation: Power; Haupteffekt B in CRF-23, Normalverteilung (links oben), Diskrete Gleichverteilung (rechts oben), Exponentialverteilung (links unten) und Log-Normalverteilung (rechts unten).	63
7.1	Boxplots der Anzahl der Corpora Lutea.	65
7.2	Simultane 95%-Konfidenzintervalle: c_1 Placebo gegen Dosisstufe 1, ..., c_4 Placebo gegen Dosisstufe 4.	67

Abbildungsverzeichnis

7.3	Boxplots der relativen Nierengewichte.	68
7.4	Boxplots der Anzahl der Leukozyten.	72

1 Einleitung

1.1 Motivation

Die vorliegende Arbeit beschäftigt sich mit der Auswertung von Datensätzen, in denen die Messwerte nicht nur von einer festen Einflussgröße (Faktor), sondern von mehr als einem Faktor beeinflusst werden. Als einfaches zweifaktorielles Beispiel sei hierfür die möglicherweise unterschiedliche Wirkung einer Substanz auf weibliche und männliche Probanden genannt, so dass das Geschlecht neben der Substanz als Einflussgröße betrachtet wird. Auch drei- oder mehrfaktorielle Designs sind in der Praxis nicht unüblich. Fragestellungen bezüglich dieser Art von Studien betreffen zunächst die allgemeine Frage nach der Wirkung eines Faktors. Nach positivem Test auf Signifikanz tritt dann die Frage in den Vordergrund, durch welche Stufen dieses Ergebnis hervorgerufen wird. Weiterhin ist für die Analyse die Berechnung von Konfidenzintervallen unabdingbar. So wird die Variabilität der Daten in die Analyse miteinbezogen und die Ergebnisse visualisiert.

Unterschieden wird dabei zwischen parametrischen und nichtparametrischen Verfahren. Im Bereich der Parametrik ist eine in der Praxis häufig angewandte Methode, zur Auswertung von Daten mit zugrundeliegender faktorieller Struktur, die Varianzanalyse. Diese beantwortet aber zum einen die Fragestellungen nur global, außerdem sind hergeleitete Konfidenzintervalle teilweise nicht kompatibel zu den Testentscheidungen. Dies bedeutet, dass eine Hypothese abgelehnt wird, obwohl das nachfolgende berechnete Konfidenzintervall die Null enthält. Weiterhin müssen nach signifikant getesteter Wechselwirkung die Datensätze entsprechend aufgetrennt werden, um dann auf Effekte zu testen, welche durch die einzelnen Faktoren verursacht werden. Dies geschieht, da sich Wirkungen einzelner Faktorstufen aufheben können, falls diese zusammen betrachtet werden. Außerdem würden sich die Schätzer für die Haupteffekte, bei vorhandener Wechselwirkung, verzerren. In Bretz, Genz und Hothorn (2001) werden parametrische multiple Kontrasttests für Linearkombinationen von Erwartungswerten (Kontraste) vorgestellt. Mit diesen können simultane Konfidenzintervalle für diese Kontraste hergeleitet werden, welche kompatibel zu den Testentscheidungen aus den Statistiken sind und das multiple Niveau stark kontrollieren. Außerdem können die einzelnen Kontraste einzeln untersucht werden. Somit vereint diese Testprozedur die drei Schritte der Beantwortung der Global- und Teilhypothesen und der Konfidenzintervallberechnung. Eine Erweiterung auf heteroskedastische Daten findet sich in Hasler und Hothorn (2008).

In der nichtparametrischen Statistik gibt es als Pendant zur Varianzanalyse für einfaktorielle Modelle den Kruskal-Wallis-Test (Kruskal und Wallis, 1952 und Brunner und Munzel, 2002, S.108) und für mehrfaktorielle Modelle die ANOVA-Typ- und Wald-Typ-Statistik (Akritas, Arnold und Brunner, 1997). Diesen Rangverfahren liegen nichtparametrische Effekte zugrunde und die Hypothesen werden unter Annahme der Globalhypothese, dass bestimmte Linearfaktoren von Verteilungsfunktionen gleich sind ($CF = 0$), hergeleitet. Dieses Vorgehen wird mit einer einfachen Struktur der Kovarianzmatrix begründet. Für diese Tests lassen sich im Einfaktoriellen wie im Mehrfaktoriellen für die nichtparametrischen Effekte nur nichtsimultane Konfidenzintervalle konstruieren (Kulle, 1999; Domhof 2001). Damit parametrischen multiplen Kontrasttests simultane Konfidenzintervalle bestimmt werden können, wurden diese von Munzel und Hothorn (2001) auf die nichtparametrischen Effekte erweitert. In dieser Arbeit wurde nur der Tukey-Vergleich betrachtet, so dass in der Dissertation von Konietschke (2009) die Theorie auf beliebige Kontraste im einfaktoriellen Modell erweitert wurde. Als Grundlage dient die Lösung zum nichtparametrischen Behrens-Fisher Problem für zwei Stichproben (Brunner und Munzel, 2000).

Das Ziel dieser Arbeit ist, die Theorie im nächsten Schritt auf mehrfaktorielle Modelle (mit unverbundenen Stichproben) zu verallgemeinern. Aufgegriffen werden die (transitiven) ungewichteten relativen Effekte aus Kulle (1999 S.38ff.), Gao et al. (2008) und Gao und Alvo (2008). Diese Effekte lassen sich als Linearkombinationen paarweiser relativer Effekte darstellen. Die Variante der Darstellung mit Linearkombinationen ist Konietschke (2009) entnommen. Über diese transitiven Effekte lassen sich dann Kontraste bilden. Durch die „beliebige“ Konstruktion der Kontraste, beschränken sich diese nicht nur auf die aus der Varianzanalyse bekannten Kontraste, für den globalen Test auf Haupteffekt und Wechselwirkung. So ist in der Praxis nicht allein die Antwort auf diese globale Fragestellung interessant, sondern vielmehr, durch welche Stufe die Signifikanz auftritt, d.h. der Anwender erhält Informationen für jede Stufe des Kontrastes. Dies beschränkt sich nicht nur auf Kontraste für einfaktorielle Modelle, sondern kann beispielsweise auch auf die Kontraste für Haupteffekte in faktoriellen Modellen angewandt werden (Gao und Alvo, 2008). Eine mögliche Anwendung dazu wird am Beispiel der sequentiellen Vorgehensweise in der ANOVA gezeigt.

1.2 Aufbau der Arbeit

In Kapitel 2 wird jeweils ein Beispiel mit einem, zwei und drei Faktoren gezeigt. Im darauf folgenden Grundlagenkapitel werden die Beispiele allgemein parametrisch dargestellt. Außerdem wird für zwei unverbundene Stichproben die Lösung zum Behrens-Fisher Problem einleitend vorgestellt. Kapitel 4 beschäftigt sich mit den nichtparametrischen Modellen, den dafür benötigten nichtparametrischen Effekten und den Hypothesen bzw. Kontrasten mit denen Informationen aus den Effekten gewonnen werden können. Außerdem wird das se-

quentielle Testproblem erläutert und Lösungsvorschläge kurz diskutiert. Danach wird in Kapitel 5 die Theorie hergeleitet, um simultane Konfidenzintervalle zu berechnen. Kapitel 6 beinhaltet die Simulationsstudie und in Kapitel 7 werden die Beispiele ausgewertet. Eine Diskussion der Ergebnisse dieser Arbeit und ein Ausblick befinden sich in Kapitel 8.

2 Motivierende Beispiele

Im Folgenden werden drei Beispiele vorgestellt, welche aus Brunner und Munzel (2002) entnommen sind. Dabei hat das erste einen, das zweite zwei und das dritte Beispiel drei Faktoren.

2.1 Anzahl der Corpora Lutea

Untersucht wurden 92 weibliche Wistar-Ratten auf unerwünschte Wirkungen einer Substanz (Verum) auf die Fertilität. Das Verum wurde in vier Dosisstufen gegeben und mit einem Placebo verglichen. Nach der Sektion der Tiere wurde unter anderem die Anzahl der Corpora Lutea bestimmt. Die Stichprobenumfänge betragen für die Placebogruppe $n_1 = 22$ und für die vier weiteren Verumgruppen $n_2 = 17$, $n_3 = 20$, $n_4 = 16$, $n_5 = 17$.

TABELLE 2.1: Anzahl der Corpora Lutea

Substanz	Anzahl der Corpora Lutea
Placebo	9, 11, 11, 11, 12, 12, 12, 12, 12, 12, 13, 13, 13, 13, 13, 14, 14, 14, 14, 15, 16
Dosis 1	9, 10, 11, 11, 11, 11, 11, 12, 12, 12, 13, 13, 14, 14, 14, 15, 15
Dosis 2	9, 11, 12, 12, 13, 13, 13, 13, 13, 14, 14, 14, 14, 15, 15, 15, 15, 17, 17
Dosis 3	6, 10, 11, 12, 12, 12, 13, 13, 13, 13, 14, 14, 14, 15, 15, 16
Dosis 4	9, 10, 11, 11, 11, 13, 13, 13, 13, 13, 14, 14, 14, 14, 15, 15

2.2 Nierengewichte

In einer Toxizitätsstudie sollten unerwünschte toxische Wirkungen einer Substanz, gegenüber Placebo, untersucht werden, welche in vier steigenden Dosisstufen den Tieren verabreicht wurde. Dabei wurden an 45 weiblichen und 41 männlichen Wistar-Ratten die relativen Nierengewichte, bezüglich des Körpergewichts, bestimmt.

TABELLE 2.2: Relative Nierengewichte

Relative Nierengewichte in [‰]					
Geschlecht	Placebo	Verum			
		Dosis 1	Dosis 2	Dosis 3	Dosis 4
weiblich	7.11	6.23	7.40	6.65	9.26
	7.08	7.93	6.51	8.11	8.62
	5.95	7.59	6.85	7.37	7.72
	7.36	7.14	7.17	8.43	8.54
	7.58	8.03	6.76	8.21	7.88
	7.39	7.31	7.69	7.14	8.44
	8.25	6.91	8.18	8.25	8.02
	6.95	7.52	7.05		7.72
		7.32	8.75		8.27
			7.53		7.91
männlich	6.62	6.25	7.11	6.93	7.26
	6.65	6.95	5.68	7.17	6.45
	5.78	5.61	6.23	7.12	6.37
	5.63	5.40	7.11	6.43	6.54
	6.05	6.89	5.55	6.96	6.93
	6.48	6.24	5.90	7.08	6.40
	5.50	5.85	5.98	7.93	7.01
	5.37		7.14		7.74
					7.63
					7.62
				7.38	

2.3 Leukozyten-Migration ins Peritoneum

In dieser Studie wurde an insgesamt 160 Mäusen die Wirkung eines Verums auf das Immunsystem unter einer Stresssituation untersucht. In der Kontrollgruppe wurde anstelle des

TABELLE 2.3: Anzahl der Leukozyten

Anzahl der Leukozyten in [$10^6/ml$]							
Normalfutter				Mangelfutter			
Stimulation				Stimulation			
nur Glycogen		Glycogen + Staphylokokken		Glycogen		Glycogen + Staphylokokken	
Placebo	Verum	Placebo	Verum	Placebo	Verum	Placebo	Verum
3.3	12.6	11.1	32.4	2.7	7.5	25.2	12.6
5.7	7.2	11.4	18.6	4.8	4.2	10.5	28.8
4.2	37.8	16.5	28.8	2.1	3.3	15.6	17.4
5.1	11.1	6.0	13.8	6.0	3.3	9.0	20.7
49.2	8.7	9.3	15.9	2.7	3.0	9.9	15.3
7.5	18.3	15.5	17.4	2.7	9.2	10.8	4.2
11.7	11.4	6.0	18.3	3.6	4.5	6.3	19.5
4.5	29.1	8.4	13.7	2.7	11.7	15.0	15.1
18.3	48.0	8.7	25.2	4.2	9.3	6.7	22.7
18.3	14.1	6.0	13.7	7.5	3.6	16.8	14.1
7.5	15.9	10.5	10.8	5.7	5.7	10.5	15.9
8.1	12.0	4.4	14.4	3.3	8.1	17.4	10.2
5.4	12.3	6.3	15.0	3.9	6.0	7.0	15.7
6.0	44.4	6.8	17.7	3.9	6.0	12.9	11.3
16.2	13.5	8.1	19.5	6.6	11.4	9.3	15.9
7.8	19.8	4.3	26.4	6.3	5.1	4.8	21.3
8.1	15.3	12.9	18.9	3.3	11.1	8.7	27.9
5.7	32.7	9.9	11.3	4.5	12.9	5.3	13.8
6.9	18.0	6.9	12.7	4.2	5.4	3.9	9.3
5.1	15.0		6.6		8.4	4.9	

Verums ein Placebo verabreicht. Eine wesentliche Zielgröße war die Leukozyten-Migration ins Peritoneum. Als Vorbehandlung erhielt die eine Hälfte der Mäuse eine eiweißarme Ernährung (Mangelfutter), die andere Hälfte normales Futter. Einen Tag vor Eröffnung des Peritoneums erhielten in jeder Gruppe 40 Mäuse eine Injektion mit der Testsubstanz (Verum), während die anderen 40 Mäuse die gleiche Menge an Placebo erhielten. Acht Stunden später wurde in jeder dieser vier Gruppen die Migration der Leukozyten durch eine Injektion von Glycogen stimuliert, wobei außerdem jeweils bei der Hälfte der Mäuse 10^8 inaktivierte Staphylokokken zugesetzt wurden. Bei den insgesamt acht Gruppen wurde unter anderem die Anzahl der Leukozyten bestimmt. Aus technischen Gründen waren drei Tiere verstorben.

2 Motivierende Beispiele

3 Grundlagen

In diesem Abschnitt wird kurz erläutert, wie für die gezeigten Beispiele im Bereich der parametrischen Statistik ein Lösungsweg aussehen könnte. Weiterhin wird allgemein auf multiple Kontrasttests eingegangen und zum Schluss wird der relative Effekt eingeführt, da die in dieser Arbeit untersuchten ungewichteten relativen Effekte auf dem relativen Effekt aufbauen. Außerdem wird die Lösung des nichtparametrischen Behrens-Fisher Problems von Brunner und Munzel (2002) im unverbundenen Zweistichprobenfall zusammengefasst. Mit Hilfe dieser Lösung werden später simultane Konfidenzintervalle konstruiert.

3.1 Parametrische Motivation

Parametrische Verfahren setzen eine Klasse von Verteilungsfunktionen voraus, die mit Hilfe von Parametern beschrieben werden können. Die Hypothesen werden dann über die Parameter dieser Verteilungsfunktionen beschrieben. Eine Möglichkeit, Datensätze mit faktoriellen Strukturen zu untersuchen, bieten die Linearen Modelle (Ravishanker und Dey, 2002). Im nichtsingulären Modell lassen sich die Messwerte der Beispiele aus Kapitel 2 durch

$$\begin{aligned} X_{ik} &= \mu_i + \epsilon_{ik}, \\ X_{ijk} &= \mu_{ij} + \epsilon_{ijk} \quad \text{und} \\ X_{ijkl} &= \mu_{ijl} + \epsilon_{ijkl} \end{aligned}$$

modellieren. Dabei sind die Werte μ Lageparameter, welche die Verschiebungseffekte, die aus den jeweiligen Faktorstufenkombinationen resultieren, beschreiben. Die Variation der Messwerte X wird durch den Fehlerterm ϵ ausgedrückt. Eine andere Schreibweise ist es, die Effekte, welche durch die entsprechenden Faktoren bzw. Wechselwirkungen verursacht werden, einzeln zu beschreiben. In diesem singulären Modell werden die Parameter μ_i, μ_{ij} und μ_{ijl} aufgespalten und die Zufallsvariablen werden durch

$$\begin{aligned} X_{ik} &= \mu + \alpha_i + \epsilon_{ik} \\ X_{ijk} &= \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \quad \text{und} \\ X_{ijkl} &= \mu + \alpha_i + \beta_j + \gamma_l + (\alpha\beta)_{ij} + (\alpha\gamma)_{il} + (\beta\gamma)_{jl} + (\alpha\beta\gamma)_{ijl} + \epsilon_{ijkl} \end{aligned}$$

erklärt. Dabei ist dann μ der Gesamteffekt, α_i, β_j und γ_l die Effekte, welche allein jeweils durch die Faktoren A, B und C beschrieben werden. Gleiches gilt für die Wechselwirkungen, so wird mit $(\alpha\beta)_{ij}$ der Effekt, der durch die Wechselwirkung zwischen den Faktoren

A und B auftritt beschrieben. An diese Effekte werden die Reparametrisierungsbedingungen gestellt, beispielsweise $\sum_i \alpha_i = 0$ oder $\sum_{i,j} \alpha_{ij} = 0$, dass diese jeweils in der Summe verschwinden. Diese Art von Modellen werden in der Varianzanalyse behandelt.

3.2 Varianzanalyse

Mit der Varianzanalyse können in faktoriellen Modellen globale Hypothesen getestet werden. Im zweifaktoriellen Modell CRF- ab beispielsweise

$$\begin{aligned} H_0^{(A)} &: \alpha_1 = \alpha_2 = \dots = \alpha_a, \\ H_0^{(B)} &: \beta_1 = \beta_2 = \dots = \beta_b \quad \text{und} \\ H_0^{(AB)} &: (\alpha\beta)_{11} = (\alpha\beta)_{12} = \dots = (\alpha\beta)_{ab}. \end{aligned}$$

Dies bedeutet, dass sich die Varianzanalyse ausschließlich für die Beantwortung globaler Fragestellungen eignet und sich keine Informationen für die einzelnen Stufen $\alpha_1, \dots, \alpha_a$; β_1, \dots, β_b bzw. $(\alpha\beta)_{11}, \dots, (\alpha\beta)_{ab}$ für den Anwender erschließen. Diese Fragen sind nur mit multiplen Vergleichen und multiplen Kontrasttests genau zu beantworten.

Die Tests in der Varianzanalyse bauen auf der Verteilung von quadratischen Formen auf. Aus den Kontrastmatrizen lassen sich idempotente und symmetrische Matrizen (Projektoren) konstruieren, woraus sich mit SATZ A.17 die Verteilungen der quadratischen Formen ergeben, deren Quotienten dann F-Verteilungen folgen. Mehr zur Varianzanalyse in relativen Effekten ist in Abschnitt 5.3 zu finden.

3.3 Multiples Testen und Multiple Kontraste

Multiple Kontrasttests beruhen auf der Idee für q verschiedene Vergleiche q univariate Prüfgrößen zu konstruieren. Da q Tests simultan durchgeführt werden, besteht die Gefahr, dass das multiple Niveau nicht eingehalten wird. Eine weitere Bezeichnung für das multiple Niveau ist die *familienbezogene Fehlerrate* - family wise error rate (kurz: FWER). Wird für die Kontrolle der FWER die Gültigkeit der Globalhypothese vorausgesetzt, so spricht man von der *schwachen* Kontrolle der FWER, ansonsten von der *starken* Kontrolle. Bei beispielsweise zwei unabhängigen Tests, die jeweils zum Niveau $\alpha = 5\%$ testen, welche zusammen zur Beantwortung einer übergeordneten Hypothese herangezogen werden, ist die Wahrscheinlichkeit, dass mindestens ein Test fälschlicherweise eine wahre Hypothese ablehnt, ohne Korrektur nicht mehr 0.05 sondern $1 - (1 - 0.05)^2 = 0.0975$. So ist der multiple Fehler anstelle der 5% schon 9.75% . Entsprechend steigt die Wahrscheinlichkeit, einen Fehler 1. Art zu begehen, mit wachsender Anzahl der Tests. Eine allgemeine Lösung bietet die p-Wert Adjustierung mit bekannten Verfahren nach Bonferroni, Scheffé (1953) oder Tukey.

3.4 Relativer Effekt und Nichtparametrisches Behrens-Fisher Problem für zwei Stichproben

Bei multiplen Kontrasttests hingegen wird aus der (asymptotisch) multivariaten Verteilung der Prüfgrößen, welche die Korrelation untereinander berücksichtigt, der Schwellwert bestimmt, mit dem das multiple Niveau eingehalten wird (Bretz et al., 2001; Hochberg und Tamhane, 1987, S.29).

Der Grundgedanke multipler Kontrasttests liegt darin, Aussagen zu Hypothesen nicht nur global zu treffen, sondern auch genauere Informationen über die einzelnen Kontraste zu erhalten. Dabei wird nach dem Union-Intersection-Prinzip von Roy (siehe Hochberg und Tamhane, 1987, S.28) jede Globalhypothese als ein Schnitt einzelner Teilhypothesen angesehen, so dass durch die Beantwortung aller Teilhypothesen auch gleichzeitig die Globalhypothese getestet wird. Die Kontrastmatrix $C = (c'_1, \dots, c'_q)'$ besteht aus einzelnen Kontrasten c'_m , die zusammengefasst die Globalhypothese beschreiben. Jeder Kontrast erhält genau eine Prüfgröße, so dass jeder Kontrast für sich untersucht werden kann. Die Globalhypothese wird dann abgelehnt, wenn mindestens eine der q Teilhypothesen abgelehnt wird.

3.4 Relativer Effekt und Nichtparametrisches Behrens-Fisher Problem für zwei Stichproben

Da in der Nichtparametrik keine Voraussetzung an die Verteilungsfamilie gestellt wird, sind keine Parameter gegeben durch die eine Verteilungsfunktion näher charakterisiert wird. Daher kann man sich überlegen, wie zumindest zwei Verteilungsfunktionen zueinander in Verbindung stehen. Auf diesen Gedankengang baut der relative Effekt auf. Dazu seien zwei unverbundene Stichproben $X_{1k} \sim F_1$ und $X_{2k} \sim F_2$ mit $k = 1, \dots, n_i$, und $i = 1, 2$ gegeben. Damit lässt sich ein nichtparametrischer Effekt über die zwei Verteilungsfunktionen definieren.

DEFINITION 3.1. (*Relativer Effekt*)

Sei $X_{ik} \stackrel{iid}{\sim} F_i, i = 1, 2, k = 1, \dots, n_i$, dann wird

$$r = r_{12} = \int F_1 dF_2 = P(X_{11} < X_{21}) + \frac{1}{2}P(X_{11} = X_{21}) \quad (3.1)$$

als *relativer Effekt* bezeichnet.

Dieser wird auch *relativer Behandlungseffekt* genannt und geht auf Mann und Whitney (1947) zurück. Aus der Definition mit Wahrscheinlichkeiten lässt sich dieser Effekt als *Wahrscheinlichkeit* interpretieren:

Die Zufallsvariable X_{1k} tendiert im Vergleich zu X_{2k} stochastisch zu größeren Werten, falls $r < \frac{1}{2}$ gilt und analog zu kleineren Werten, falls $r > \frac{1}{2}$ gilt. Die Zufallsvariablen heißen weiterhin *stochastisch tendenziell gleich*, wenn $r = \frac{1}{2}$ gilt.

Eine weitere Eigenschaft ist die Invarianz unter ordnungserhaltenden Transformationen. Für unabhängige normalverteilte Zufallsvariablen $X_1 \sim N(\mu_1, \sigma_1^2)$ und $X_2 \sim N(\mu_2, \sigma_2^2)$ mit

möglicherweise ungleichen Varianzen gilt $r = \phi \left(\frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_2^2}} \right)$ (siehe Domhof, 2001, S.23) mit der Verteilungsfunktion der Standardnormalverteilung ϕ . Der relative Effekt ist dabei als standardisierter Effekt in Standardabweichungseinheiten zu verstehen. Somit sind die Hypothesen $H_0^\mu: \mu_1 = \mu_2$ und $H_0^r: r = \frac{1}{2}$, unter Normalverteilung, äquivalent. Aufgrund dessen wird H_0^r als nichtparametrisches Behrens-Fisher Problem bezeichnet. Die Theorie hierzu wird kurz im nächsten Abschnitt 3.4.1 behandelt. Für zwei unabhängige normalverteilte Stichproben mit unbekanntem und möglicherweise unterschiedlichen Varianzen σ_1^2 und σ_2^2 wird der Test auf gleiche Erwartungswerte $H_0^\mu: \mu_1 = \mu_2$ als parametrisches Behrens-Fisher Problem bezeichnet.

Nichtparametrisches Behrens-Fisher Problem für zwei Stichproben

Für den Zweistichprobenfall werden für die Lösung des nichtparametrischen Behrens-Fisher-Problems die Ergebnisse von Brunner und Munzel (2000) kurz zusammengefasst. Schon Fligner und Pollicello (1981) beschäftigten sich mit diesem Thema, beschränkten sich dabei aber auf stetige Verteilungsfunktionen. Eine Verallgemeinerung auf den multivariaten Fall wurde in Brunner, Munzel und Puri (2002) abgehandelt. Diese Ergebnisse sind in der Theorie, Kapitel 5, nochmals genauer dargestellt.

Es sind zwei unverbundene Stichproben X_{1k} und X_{2k} gegeben. Dann seien $N = n_1 + n_2$ und r der relative Effekt aus Definition (3.1). Außerdem gelten die Voraussetzungen $N/n_i \leq N_0 < \infty$ und $Var(F_i(X_{jk})) = \sigma_{ij}^2 \geq \sigma_0^2 > 0$, für $i, j = 1, 2$ und $i \neq j$.

Beim Ersetzen der Verteilungsfunktionen $F_i, i = 1, 2$ durch die empirischen Counterparts $\hat{F}_i, i = 1, 2$ ergibt sich, unter Verwendung des Mittelranges $\bar{R}_2 = \frac{1}{n_2} \sum_{k=1}^{n_2} R_{2k}^{(12)}$, wobei $R_{2k}^{(12)}$ der Rang von X_{2k} in der gepoolten Stichprobe ist, der Plug-in Schätzer:

$$\hat{r} = \int \hat{F}_1 d\hat{F}_2 = \frac{1}{n_1} \left(\bar{R}_2 - \frac{n_2 + 1}{2} \right).$$

Hergeleitet werden soll die asymptotische Verteilung der Rangstatistik $\sqrt{N}(\hat{r} - r)$. Da aber die Ränge unabhängiger Zufallsvariablen nicht mehr unabhängig sind, kann der Zentrale Grenzwertsatz nicht direkt angewandt werden. Es lässt sich aber zeigen, dass

$$\sqrt{N}(\hat{r} - r) = \sqrt{N}(B_{12} + C_{12}),$$

mit $B_{12} = \left(\int F_1 d\hat{F}_2 - \int F_2 d\hat{F}_1 + 1 - 2r \right)$ und $C_{12} = \int (\hat{F}_1 - F_1) d(\hat{F}_2 - F_2)$ gilt. Aus dem Asymptotischen Äquivalenzsatz (Brunner und Munzel, 2002, S.207f.) folgt, dass $\sqrt{N}C_{12}$ in Wahrscheinlichkeit gegen Null konvergiert. Die Zufallsvariable B_{12} hingegen ist eine Linearkombination unabhängig verteilter Zufallsvariablen $F_1(X_{2k})$ und $F_2(X_{1k})$. Außerdem gilt $0 \leq F_1(X_{2k}) \leq 1$ und $0 \leq F_2(X_{1k}) \leq 1$, so dass diese Zufallsvariablen gleichmäßig beschränkt sind. Dadurch und mit den oben an die Varianzen und Stichprobenumfängen gestellten Voraussetzungen ist der Satz von Lindeberg-Feller anwendbar, welcher

3.4 Relativer Effekt und Nichtparametrisches Behrens-Fisher Problem für zwei Stichproben

die asymptotische Normalität von B_{12} und so auch der Rangstatistik sicherstellt:

$$\sqrt{N}(\hat{r} - r) \doteq \sqrt{N}B_{12} \dot{\sim} N(0, \sigma_N^2).$$

Dabei setzt sich σ_N^2 wie folgt aus $\sigma_{21}^2 = \text{Var}(F_2(X_{11}))$ und $\sigma_{12}^2 = \text{Var}(F_1(X_{21}))$ zusammen:

$$\sigma_N^2 = \text{Var}(\sqrt{N}B_{12}) = N \left(\frac{\sigma_{21}^2}{n_1} + \frac{\sigma_{12}^2}{n_2} \right). \quad (3.2)$$

Ein konsistenter Schätzer $\hat{\sigma}_N^2$ für σ_N^2 ist zu erreichen, indem die beiden Varianzen σ_{21}^2 und σ_{12}^2 durch

$$\hat{\sigma}_{21}^2 = \frac{1}{n_2^2(n_1 - 1)} \sum_{k=1}^{n_1} \left(R_{1k}^{(12)} - R_{1k}^{(1)} - \bar{R}_{1\cdot}^{(12)} + \frac{n_1 + 1}{2} \right)^2 \quad \text{und}$$

$$\hat{\sigma}_{12}^2 = \frac{1}{n_1^2(n_2 - 1)} \sum_{k=1}^{n_2} \left(R_{2k}^{(12)} - R_{2k}^{(2)} - \bar{R}_{2\cdot}^{(12)} + \frac{n_2 + 1}{2} \right)^2$$

in (3.2) ersetzt werden. Hierbei seien $R_{1k}^{(1)}$ und $R_{2k}^{(2)}$ stichprobeninterne Ränge. Aus der asymptotischen Normalverteilung von $\sqrt{N}(\hat{r} - r)$ folgt, dass die Zufallsvariable $T^r = \sqrt{N}(\hat{r} - r) / \hat{\sigma}_N$ asymptotisch standardnormalverteilt ist. Die Hypothese $H_0^r: r = \frac{1}{2}$ wird abgelehnt, wenn $|T^{1/2}| \geq z_{1-\alpha/2}$, wobei $z_{1-\alpha/2}$ das $(1 - \alpha/2)$ -Quantil der Standardnormalverteilung bezeichnet. Aus der Teststatistik T^r ergibt sich für r mit der Pivot-Methode das asymptotische zweiseitige $(1 - \alpha)$ -Konfidenzintervall

$$\left[\hat{r} - z_{1-\alpha/2} \frac{\hat{\sigma}_N}{\sqrt{N}}, \hat{r} + z_{1-\alpha/2} \frac{\hat{\sigma}_N}{\sqrt{N}} \right].$$

Für eine bessere Approximation bei kleinen Stichprobenumfängen kann die Verteilung von T^r durch eine t-Verteilung mit

$$\hat{f} = \frac{(n_1 \hat{\sigma}_{12}^2 + n_2 \hat{\sigma}_{21}^2)^2}{(n_1 \hat{\sigma}_{12}^2)^2 / (n_2 - 1) + (n_2 \hat{\sigma}_{21}^2)^2 / (n_1 - 1)} \quad (3.3)$$

Freiheitsgraden approximiert werden (Brunner und Munzel, 2000).

4 Nichtparametrische Modelle, Effekte und Hypothesen

4.1 Nichtparametrische Modelle

Behandelt werden in dieser Arbeit gekreuzte, vollständig randomisierte Versuchsmodelle mit unverbundenen Stichproben, die als CRF, Completely Randomized Factorial Design, bezeichnet werden. Unverbunden bedeutet, dass die Werte unabhängig sind. Somit sind zum Beispiel Versuchsanlagen mit abhängigen Messungen, wie der Verlauf von Werten zu unterschiedlichen Zeitpunkten, nicht beinhaltet. Gekreuzte Faktoren zeichnen sich dadurch aus, dass ihre Stufen ein kartesisches Produkt bilden.

Im Folgenden wird kurz $X_{ik} \stackrel{iid}{\sim} F_i$ (*independent and identically distributed*) geschrieben, falls unabhängige und identisch verteilte Zufallsvariablen gegeben sind, wobei $F_i(x) = P(X_{ik} < x) + \frac{1}{2}P(X_{ik} = x)$, $i = 1, \dots, a, k = 1, \dots, n_i$.

Zunächst wird ein Überblick zu den Modellen gegeben, welche in dieser Arbeit eine zentrale Rolle spielen.

MODELL 4.1. (CRF-a)

Für das einfaktorielle Modell mit a Stichproben seien

$$X_{ik} \stackrel{iid}{\sim} F_i,$$

mit $i = 1, \dots, a$ und n_i die Anzahl der Individuen in Stichprobe i , an denen eine Messung unternommen wurde.

Dann sei $\mathbf{X} = (X_{11}, \dots, X_{1n_1}, \dots, X_{a1}, \dots, X_{an_a})'$ der Vektor aller Zufallsvariablen, dessen Struktur in TABELLE 4.1 zu sehen ist.

Für zweifaktorielle Designs, wie die Nierengewichte aus Abschnitt 2.2, hat das Modell die folgende Gestalt.

MODELL 4.2. (CRF-ab)

Im Modell für zwei Faktoren gelte

$$X_{ijk} \stackrel{iid}{\sim} F_{ij},$$

wobei $i = 1, \dots, a, j = 1, \dots, b$ und $k = 1, \dots, n_{ij}$, mit a Faktorstufen des Faktors A , b Stufen des Faktors B und n_{ij} Individuen in der Faktorstufenkombination der i -ten Stufe von Faktor A und j -ten Stufe des Faktors B .

TABELLE 4.1: Einfaktorielle Datenstruktur

$i = 1$	X_{11}
	\vdots
	X_{1n_1}
\vdots	\vdots
$i = a$	X_{a1}
	\vdots
	X_{an_a}

Die Zufallsvariablen sind in dem Vektor

$\mathbf{X} = (X_{111}, \dots, X_{11n_1}, \dots, X_{1b1}, \dots, X_{1bn_1b}, \dots, X_{a11}, \dots, X_{a1n_a}, \dots, X_{ab1}, \dots, X_{abn_{ab}})'$ zusammengefasst und die gekreuzte Struktur ist in der TABELLE 4.2 dargestellt. Für drei

TABELLE 4.2: Zweifaktorielle Datenstruktur

Faktor A	Faktor B			
	$j = 1$	$j = 2$	\dots	$j = b$
$i = 1$	X_{111}	X_{121}	\dots	X_{1b1}
	\vdots	\vdots	\dots	\vdots
	X_{11n_1}	X_{12n_2}	\dots	$X_{1bn_{1b}}$
\vdots	\vdots	\vdots	\vdots	\vdots
$i = a$	X_{a11}	X_{a21}	\dots	X_{ab1}
	\vdots	\vdots	\dots	\vdots
	$X_{a1n_{a1}}$	$X_{a2n_{a2}}$	\dots	$X_{abn_{ab}}$

Faktoren ergibt sich auf gleiche Weise, durch Erweiterung mit einem zusätzlichen Index l , das folgende Modell.

MODELL 4.3. (CRF-abc)

Für ein Design mit drei Faktoren seien

$$X_{ijkl} \stackrel{iid}{\sim} F_{ijl},$$

mit $k = 1, \dots, n_{ijl}$ und zusätzlich $l = 1, \dots, c$ Stufen des Faktors C.

Für mehr als drei Faktoren definieren sich die Modelle analog, indem ein neuer Index für jeden Faktor hinzugefügt wird. Zur besseren Handhabung wird das Produkt der Faktorstufen mit d abgekürzt, beispielsweise $d: = ab$ im MODELL 4.2, und der Gesamtstichprobenumfang über alle Kombinationen der Stufen sei $N = \sum_{i=1}^a \sum_{j=1}^b n_{ij} = \sum_{t=1}^d n_t$. Die Verwendung des Indexes t wird im nächsten Abschnitt kurz erläutert.

Um die Theorie übersichtlicher darzustellen, besonders bei höherfaktoriellen Designs, wird eine Umindizierung vorgenommen und die faktorielle Struktur vernachlässigt, so dass auch der programmiertechnische Aufwand für die spätere Umsetzung durch diesen Schritt verringert wird. Die Indizes werden nach den Faktoren sortiert. Dieser Sortieralgorithmus kann formal im MODELL 4.2 mit der Funktion

$$f_{(2)} : \mathbb{N}^2 \rightarrow \mathbb{N}, \quad (i, j) \mapsto (i - 1)b + j$$

und im MODELL 4.3 mit

$$f_{(3)} : \mathbb{N}^3 \rightarrow \mathbb{N}, \quad (i, j, l) \mapsto (i - 1)bc + (j - 1)c + l$$

beschrieben werden. Die Menge der ab -Indexpaare bzw. der abc -Indextriplets wird dabei auf den Zahlenstrahl der natürlichen Zahlen, von 1 bis d , abgebildet. In dem einfaktoriellen Modell muss natürlich keine Umindizierung mehr unternommen werden. Der Index t läuft nun also von 1 bis d . Am Beispiel eines 2×3 Designs, MODELL 4.2 mit $a=2$ und $b=3$, wird die Anordnung der Paare und deren Projektion kurz deutlich gemacht:

$$\begin{array}{lll} (1, 1) \mapsto 1, & (1, 2) \mapsto 2, & (1, 3) \mapsto 3 \\ (2, 1) \mapsto 4, & (2, 2) \mapsto 5, & (2, 3) \mapsto 6. \end{array}$$

Diese Struktur überträgt sich auch auf den Vektor der Verteilungsfunktionen

$\mathbf{F} = (F_{11}, \dots, F_{1b}, \dots, F_{a1}, \dots, F_{ab})' = (F_1, \dots, F_d)'$ und den Zufallsvariablen

$\mathbf{X} = (\mathbf{X}'_1, \dots, \mathbf{X}'_d)'$ mit $\mathbf{X}_t = (X_{t1}, \dots, X_{tn_t})'$. Diese Ordnung wird sich auch auf weitere Vektoren übertragen, die im Laufe dieser Arbeit noch eingeführt werden.

Für das Herleiten asymptotischer Resultate müssen die Stichprobenumfänge und die Varianzen die folgenden Voraussetzungen erfüllen.

VORAUSSETZUNGEN 4.1.

$$(V1) \quad \min(n_1, \dots, n_d) \rightarrow \infty, \text{ s.d. } \frac{N}{n_t} \leq N_0 < \infty, \forall t = 1, \dots, d$$

$$(V2) \quad \text{Var}(F_t(X_{uk})) = \sigma_{tu}^2 \geq \sigma_0^2 > 0, \forall t, u = 1, \dots, d \text{ und } t \neq u.$$

Die erste Voraussetzung stellt sicher, dass die inversen relativen Stichprobenumfänge auch asymptotisch gleichmäßig beschränkt sind. Dass die Varianzen positiv und von Null weg beschränkt sind, besagt die zweite Voraussetzung. Durch die Verwendung der normalisierten Versionen sind nicht nur stetige Funktionen zulässig, sondern auch unstetige und diskrete wie speziell dichotome.

4.2 Nichtparametrische Effekte

In diesem Kapitel wird ein kurzer Überblick zu den nichtparametrischen Effekten gegeben. Beginnend mit dem paarweisen relativen Effekt, definiert über zwei Verteilungsfunktionen, wird der gewünschte nichtparametrische Effekt formuliert, mit dem in Abschnitt 5 weiter gearbeitet wird. In Domhof (2001) werden die Eigenschaften nichtparametrischer Effekte ausführlich behandelt.

4.2.1 Relative Effekte im Mehrstichprobenfall

Sollen mehr als zwei Verteilungen miteinander verglichen werden, kann für den paarweisen Vergleich der relative Effekt (3.1) auf zwei beliebige Verteilungsfunktion ausgedehnt werden und man erhält den paarweisen relativen Effekt r_{ij} .

DEFINITION 4.2. (*Paarweiser Relativer Effekt*)

Sei $X_{tk} \stackrel{iid}{\sim} F_t, t = 1, \dots, d, k = 1, \dots, n_t$, dann sind die paarweisen relativen Effekte definiert durch

$$r_{tu} = \int F_t dF_u = P(X_{t1} < X_{u1}) + \frac{1}{2}P(X_{t1} = X_{u1}). \quad (4.1)$$

Die Effekte (4.1) werden zusammenfassend im Vektor $\mathbf{r} = (r_{11}, \dots, r_{1d}, \dots, r_{d1}, \dots, r_{dd})'$ geschrieben. Die paarweisen Effekte sind nicht transitiv, so dass diese für den Mehrstichprobenfall nicht geeignet sind. Mit *Efron's paradoxen Würfeln* wurde in Gardner (1970), Thangavelu und Brunner (2007) und Brown und Hettmansperger (2002) die Nichttransitivität der paarweisen relativen Effekte erörtert und die paradoxen Aussagen diskutiert. Einen Ausweg bieten transitive Effekte, die über gemittelte Verteilungsfunktionen (siehe DEFINITION 4.3) definiert sind. Zunächst wird aber mit den *Efron'schen Würfeln* die Nichterfüllung der transitiven Eigenschaft gezeigt.

Efron'sche Würfel

In diesem Paradoxon sind vier Würfel gegeben, die mit den natürlichen Zahlen

$$W_1 = \{0, 0, 4, 4, 4, 4\}, W_2 = \{3, 3, 3, 3, 3, 3\}, W_3 = \{2, 2, 2, 2, 6, 6\}, W_4 = \{1, 1, 1, 5, 5, 5\}$$

belegt sind. Dann seien $p_{ij}, i, j = 1, \dots, 4$, die entsprechenden paarweisen relativen Effekte und dann gilt $p_{12} = p_{23} = p_{34} = 1/3$. Dies kann anschaulich so interpretiert werden, dass W_1 besser ist als W_2 , W_2 besser als W_3 und außerdem W_3 besser als W_4 . Dem natürlichen Verständnis und auch der Transitivität nach, sollte W_1 besser als W_4 sein. Es gilt aber $p_{41} = 1/3$, wonach die Eigenschaft „besser“ nicht transitiv ist. Um die Würfel besser einschätzen zu können, müssten diese jeweils mit einem „Würfel“ verglichen werden, der die Werte im gleichen Verhältnis die der vier Würfel beinhaltet. Dieser wird durch eine gemittelte Verteilungsfunktion in der Theorie repräsentiert.

4.2.2 Transitive relative Effekte

Um transitive Effekte zu erhalten, bieten sich zwei Vergleichsfunktionen an. Diese sind in der folgenden Definition kurz zusammengefasst.

DEFINITION 4.3. (*Gemittelte Verteilungsfunktionen*)

Mit

$$H(x) = \frac{1}{N} \sum_{t=1}^d n_t F_t(x) \quad (4.2)$$

bezeichnet man das gewichtete und mit

$$G(x) = \frac{1}{d} \sum_{t=1}^d F_t(x) \quad (4.3)$$

das ungewichtete Mittel der Verteilungsfunktionen.

In der gemittelten Version werden die Verteilungsfunktionen mit den entsprechenden Stichprobenumfängen gewichtet, so dass die Verteilungsfunktionen der Versuchsgruppen mit mehr Messungen auch einen höheren Einfluss haben. Im Gegensatz dazu sind in der ungewichteten Version (4.3) die Gewichte gleich. Aus diesen beiden Definitionen ergeben sich die entsprechenden transitiven Effekte.

DEFINITION 4.4. (*Transitive Effekt*)

Seien H und G das gewichtete bzw. ungewichtete Mittel der Verteilungsfunktionen aus DEFINITION 4.3, dann ist

$$p_t = \int H dF_t = \frac{1}{N} \sum_{u=1}^d n_u \int F_u dF_t = \frac{1}{N} \sum_{u=1}^d n_u r_{ut} \quad (4.4)$$

der gewichtete relative Effekt von F_t zu F_1, \dots, F_d und

$$q_t = \int G dF_t = \frac{1}{d} \sum_{u=1}^d \int F_u dF_t = \frac{1}{d} \sum_{u=1}^d r_{ut} = \bar{r}_{\cdot t} \quad (4.5)$$

der ungewichtete relative Effekt von F_t zu F_1, \dots, F_d .

Der Vektor aller Effekte in (4.5) lässt sich als $\mathbf{q} = (q_1, \dots, q_d)'$ = $\int G d\mathbf{F}$ schreiben und analog gilt für die Effekte (4.4) $\mathbf{p} = \int H d\mathbf{F}$.

Die Verwendung der gewichteten Version ist aber nicht vorteilhaft, denn durch die Abhängigkeit von den Stichprobenumfängen ist dieser Effekt keine konstante Größe mehr und verliert so seinen Charakter als feste Modellgröße. Somit wird die Auswertung späterer Testergebnisse erschwert, da der Gesichtspunkt der Variabilität über die Stichprobenumfänge nicht aus dem Auge gelassen werden kann. Außerdem ist das Testen der Hypothese H_0^p in unbalancierten Versuchsplänen nicht geeignet, da hier sogar die Hypothese selbst abhängig

von den Stichprobenumfängen ist. Eigenschaften und Interpretation der relativen Effekte wurden schon von Kulle (1999) behandelt und auch die weitere Theorie ist in dieser Arbeit abgehandelt worden. Die Nachteile der Effekte (4.4) sind dort bereits erkannt worden. Die Effekte (4.5) wurden auch in Gao et al. (2008) und Gao und Alvo (2008) verwendet.

Im Effekt (4.5) ist der Wert $r_{uu} = \frac{1}{2}$ enthalten und deshalb gilt $q_t \in [\frac{1}{2d}, 1 - \frac{1}{2d}]$. Somit kann dieser Effekt nicht als Wahrscheinlichkeit interpretiert werden. Wird der paarweise Effekt r_{uu} weggelassen, in welchem die Verteilungsfunktion $F_t(x)$ mit sich selbst verglichen wird, erhält man den Effekt

$$q_t^{(-t)} = \int G_t dF_t = \frac{1}{d-1} \sum_{u \neq t} \int F_u dF_t = \frac{1}{d-1} \sum_{u \neq t} r_{ut} \quad (4.6)$$

aus Domhof (2001).

In diesem Fall ist die ungewichtete Verteilungsfunktion

$G_t(x) = \frac{1}{d-1} \sum_{u \neq t} F_u(x)$ und es gilt $q_t^{(-t)} \in [0, 1]$, so dass dieser als Wahrscheinlichkeit betrachtet werden kann. Weitere Betrachtungen dazu sind in Domhof (2001) zu finden. Allerdings ist der Effekt (4.6) nicht transitiv, siehe Domhof (2001) Beispiel 3.5, so dass in dieser Arbeit mit Effekt (4.5) gearbeitet wird. Für eine leichtere programmiertechnische Umsetzung der Theorie bietet sich die folgende Darstellung der Effekte (4.5) an.

PROPOSITION 4.5. *(Darstellung der ungewichteten relativen Effekte)*

Seien $\mathbf{q} = (q_1, \dots, q_d)'$ und $\mathbf{r} = (r_{11}, \dots, r_{dd})'$ die Effekte aus (4.1) und (4.5). Weiterhin sei $\mathbf{W} = \frac{1}{d} \mathbf{1}'_d \otimes \mathbf{I}_d$. Dann gilt:

$$\mathbf{q} = \mathbf{W} \mathbf{r}.$$

Beweis. Es genügt die Behauptung für eine t-te Koordinate zu zeigen. Leicht einzusehen ist, dass die t-te Zeile von \mathbf{W} eine Aneinanderreihung von insgesamt d t-ten kanonischen Einheitsvektoren, geteilt durch d , ist: $\mathbf{w}_t = \frac{1}{d} (\mathbf{e}_d^{(t)'}, \dots, \mathbf{e}_d^{(t)'})'$. Weiterhin sei $\mathbf{r}_t = (r_{t1}, \dots, r_{td})'$, so dass dann gilt:

$$\begin{aligned} \mathbf{w}'_t \mathbf{r} &= \frac{1}{d} \langle (\mathbf{e}_d^{(t)'}, \dots, \mathbf{e}_d^{(t)'})', (r_{11}, \dots, r_{1d}, \dots, r_{d1}, \dots, r_{dd})' \rangle \\ &= \frac{1}{d} \left(\langle \mathbf{e}_d^{(t)'}, \mathbf{r}'_1 \rangle + \dots + \langle \mathbf{e}_d^{(t)'}, \mathbf{r}'_d \rangle \right) \\ &= \frac{1}{d} (r_{1t} + \dots + r_{dt}) = \frac{1}{d} \sum_{u=1}^d r_{ut} \\ &= \bar{r}_{\cdot t} \\ &= q_t. \end{aligned}$$

□

Zur Veranschaulichung der Matrix \mathbf{W} sei diese hier dargestellt:

$$\mathbf{W} = \left(\begin{array}{cccc|cccc} \frac{1}{d} & 0 & \dots & 0 & \frac{1}{d} & 0 & \dots & 0 \\ 0 & \frac{1}{d} & & \vdots & 0 & \frac{1}{d} & & \vdots \\ \vdots & & \ddots & 0 & \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{d} & 0 & \dots & 0 & \frac{1}{d} \end{array} \right)_{d \times dd}. \quad (4.7)$$

Mit diesem hergeleiteten transitiven Effekt \mathbf{q} können die vier Würfel besser miteinander, nicht nur paarweise, verglichen werden. So ergeben sich die Werte $q_1 = 35/72, q_2 = 36/72 = 1/2, q_3 = 37/72$ und $q_4 = 1/2$. So neigt W_3 zu etwas besseren, W_1 zu etwas schlechteren Werten, währenddessen W_2 und W_4 stochastisch tendenziell gleich sind zum ungewichteten Mittel G .

ANMERKUNG 4.1. Die transitive Eigenschaft gilt nicht nur, wenn gegen eine gemittelte Verteilungsfunktion verglichen wird, sondern auch beim Vergleich gegen eine feste Gruppe. So lässt sich beispielsweise mit dem Dunnett-Kontrast C_{Dun} (Test gegen eine Kontrollgruppe, siehe Abschnitt 4.3) $H_0: C_{Dun}\mathbf{r} = \mathbf{0}$ transitiv testen (Konietschke, 2009).

4.3 Kontrastmatrizen und Hypothesen

Ein wichtiger Abschnitt in dieser Arbeit betrifft die Bildung von Kontrastmatrizen beziehungsweise das Design von Hypothesenmatrizen. So wird zunächst die parametrische Statistik mit der Varianzanalyse betrachtet, gefolgt von der Nichtparametrik. Dabei können Hypothesen sowohl über die Erwartungswerte als auch über Verteilungsfunktionen formuliert werden (Brunner und Munzel, 2002, S.138ff.). Eine weitere Möglichkeit bieten die eingeführten nichtparametrischen Effekte. Zur besseren Veranschaulichung wird die ursprüngliche Indizierung aus den MODELLEN 4.1 bis 4.3 beibehalten.

4.3.1 Formulierung über Erwartungswert und Verteilungsfunktion

Bevor Hypothesen über Effekte aufgestellt werden wird in diesem Unterkapitel kurz die Möglichkeit diskutiert, diese auch über Erwartungswerte und Verteilungsfunktionen zu formulieren. Dafür ist es ausreichend, dies im MODELL 4.2 zu zeigen. Weitere Hypothesen in anderen faktoriellen Strukturen werden in Abschnitt 4.3.2 gezeigt.

Im Gebiet der Parametrik werden Bedingungen hinsichtlich ihrer Verteilungen an die Zufallsvariablen gestellt. So wird in der Varianzanalyse das Lineare Modell

$$X_{ijk} = \mu_{ij} + \epsilon_{ijk},$$

zugrunde gelegt, in dem die Zufallsvariablen X_{ijk} normalverteilt sind. Dabei sind μ_{ij} ein Lageparameter und ϵ_{ijk} ein normalverteilter Fehlerterm mit $E(\epsilon_{ijk}) = 0$ und $Var(\epsilon_{ijk}) = \sigma^2$. Die Theorie der Linearen Modelle wird im Buch von Ravishanker und Dey (2002) behandelt. Die Darstellung der Zufallsvariablen X_{ijk} entspricht einem homoskedastischen Verschiebungsmodell: $X_{ijk} \sim F_{ij}(x) = F(x - \mu_{ij})$. So sind die Verteilungen bis auf einen Verschiebungseffekt gleich, das heißt zwischen zwei verschiedenen Stichproben X_{ijk} und X_{rsk} , mit $(i, j) \neq (r, s)$, sind Unterschiede hinsichtlich ihrer Verteilungen durch die Lageparameter beschrieben. Es ist dann sinnvoll, über diese Parameter Hypothesen zu bilden. Beispielsweise schreibt sich die Nullhypothese, dass kein Haupteffekt A vorhanden ist, dann als $H_0^\mu(A): \bar{\mu}_{.1} = \dots = \bar{\mu}_{.a}$. Oder analog für keinen Haupteffekt B als $H_0^\mu(B): \bar{\mu}_{.j} = \dots = \bar{\mu}_{.b}$.

Das Lineare Modell ist mit der Annahme der Normalverteilung sehr restriktiv, da in der Praxis diese Voraussetzung oft nicht erfüllt ist.

Hingegen sind Einschränkungen in Bezug auf die Verteilungen, bis auf die VORAUSSETZUNGEN 4.1, in der nichtparametrischen Statistik nicht gegeben. Im Mittelpunkt stehen hier die transitiven Effekte. Aus diesen wird für Effekt (4.4) die Rangstatistik $\sqrt{N}(\hat{p} - p)$ konstruiert. Die unter der Hypothese $H_0^F: CF = 0$ asymptotisch multivariat normalverteilt ist (Akritas, Arnold und Brunner, 1997). C ist dabei eine Kontrastmatrix (siehe Abschnitt 4.3.2). Als Motivation für die Herleitung der asymptotischen Verteilung unter H_0^F , wird eine einfachere Struktur der Kovarianzmatrix angegeben (Brunner und Munzel, 2002, S.209). Die Hypothesen ergeben sich ähnlich wie im parametrischen Fall, nur das nun die Lokationsparameter durch die Verteilungsfunktionen ersetzt werden. Hypothesen über Verteilungsfunktionen sind aber nicht einfach zu interpretieren. So kann eine Ungleichheit zwischen zwei Verteilungen verschiedene Ursachen haben. Bei einer Normalverteilung können sowohl die Erwartungswerte, als auch die Varianzen oder beide Parameter gleichzeitig unterschiedlich sein.

Die Rangstatistik $\sqrt{N}(\hat{q} - q)$ beinhaltet die transitiven und ungewichteten Effekte q . Daher ist es natürlicher die Hypothesen über diese Effekte zu formulieren. Die Herleitung der asymptotischen Verteilung der Rangstatistik, unter der Globalhypothese $H_0^{(q)}: Cq = 0$ und die Struktur der Kovarianzmatrix wird in Kapitel 5 näher erläutert.

4.3.2 Hypothesen in nichtparametrischen Effekten

Die Theorie in dieser Arbeit wird für den ungewichteten relativen Effekt $q = (q_1, \dots, q_d)'$ (4.5) erarbeitet. So werden die Hypothesen über diesen Effekt formuliert. Die Kontrastmatrix $C \in \mathbb{R}^{q \times d}$ sei eine Matrix, welche insgesamt q Kontraste (Zeilenvektoren) beinhaltet. Nach der Definition einer Kontrastmatrix gilt $c_m' \mathbf{1} = 0$ für jeden Kontrast m . Für spätere Zwecke sei $\mathcal{I} = \{1, \dots, d\}$ und weiter $\mathcal{I}_m^+ = \{t \in \mathcal{I}, c_{mt} > 0\}$ und $\mathcal{I}_m^- = \{t \in \mathcal{I}, c_{mt} < 0\}$, die positiven bzw. negativen Indextmengen des m -ten Kontrastes. Außerdem sei $K = \max_{m=1, \dots, q} \sum_{u \in \mathcal{I}_m^+} c_{mu} = \max_{m=1, \dots, q} \sum_{u \in \mathcal{I}_m^-} |c_{mu}|$ eine Konstante, welche die maximale Sum-

me der positiven bzw. negativen Einträge in einem Kontrast, über alle m Kontraste, ist. Normierte Matrizen zeichnen sich dadurch aus, dass $\sum_{u \in \mathcal{I}_m^+} c_{mu} = 1, \forall m$ und so insbesondere auch $K = 1$ ist.

Die durch eine Kontrastmatrix definierte Globalhypothese kann als

$$H_0^{(q)}: \mathbf{C}\mathbf{q} = \mathbf{0} \quad (4.8)$$

geschrieben werden. Anstelle der zweiseitigen können auch einseitige Hypothesen definiert werden, so dass für die obere

$$H_{0,o}^{(q)}: \mathbf{C}\mathbf{q} \leq \mathbf{0}$$

und für die untere

$$H_{0,u}^{(q)}: \mathbf{C}\mathbf{q} \geq \mathbf{0}$$

geschrieben wird. Im Folgenden wird sich auf die zweiseitige Hypothesen (4.8) beschränkt, die Interpretationen ändern sich entsprechend, falls die einseitigen betrachtet werden. Da \mathbf{C} aus q Kontrasten zusammengesetzt ist, ist $H_0^{(q)}$ (4.8) der Durchschnitt der Teilhypothesen

$$H_m^{(q)}: \bigcap \{ \mathbf{c}'_m \mathbf{q} = 0 \},$$

mit $m = 1, \dots, q$ und $\mathbf{c}_m = (c_{m1}, \dots, c_{md})'$.

Ein wichtiger Punkt in dieser Arbeit ist, dass die Kontrastmatrizen spezifisch auf die Fragestellungen der Anwender angepasst werden können. Da die Menge dieser Kontraste sehr vielfältig ist, werden im Folgenden nur einige praxisrelevante Kontraste behandelt. Diese werden in unterschiedlichen faktoriellen Designs erklärt.

CRF-a

Bei nur einem Faktor stehen Kontrastmatrizen im Vordergrund, welche die Beziehungen zwischen den Stufen eines Faktors untersuchen. In Dosis-Wirkungsstudien, wie im Beispiel 2.1, wird die Wirkung einer Substanz untersucht und mit den Werten einer Placebogruppe verglichen. Diese Placebogruppe dient dann als Kontrolle, so bietet sich hier der Dunnett-Kontrast (Dunnett, 1955)

$$\mathbf{C}_{Dun}^a = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{pmatrix}_{(a-1) \times a}, \quad (4.9)$$

mit den Teilhypothesen

$$\begin{aligned} H_1^{(q)} &: q_1 - q_2 = 0 \\ &\vdots \\ H_a^{(q)} &: q_1 - q_a = 0 \end{aligned}$$

an.

Ist keine Faktorstufe als Kontrollgruppe geeignet, kann auch gegen das Mittel verglichen werden. Diese Vergleiche liegen Varianzanalysen zugrunde und werden hier als Average-Kontraste bezeichnet. Die Kontrastmatrix ist ein Projektor und hat die Form

$$\mathbf{P}_a = \mathbf{I}_a - \frac{1}{a} \mathbf{J}_a = \begin{pmatrix} 1 - \frac{1}{a} & -\frac{1}{a} & \dots & -\frac{1}{a} \\ -\frac{1}{a} & 1 - \frac{1}{a} & \dots & -\frac{1}{a} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{a} & -\frac{1}{a} & \dots & 1 - \frac{1}{a} \end{pmatrix}_{a \times a},$$

mit den Teilhypothesen

$$\begin{aligned} H_1^{(q)} &: q_1 - \bar{q} = 0 \\ &\vdots \\ H_a^{(q)} &: q_a - \bar{q} = 0, \end{aligned}$$

wobei $\bar{q} = \frac{1}{a} \sum_{i=1}^a q_i$.

Wenn es möglich ist die Faktorstufen zu ordnen, kann auch ein Unterschied zwischen zwei folgenden Dosisstufen von Interesse sein. Der sequentielle Kontrast

$$\mathbf{C}_{seq}^a = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -1 \end{pmatrix}_{(a-1) \times a}, \quad (4.10)$$

mit

$$\begin{aligned} H_1^{(q)} &: q_1 - q_2 = 0 \\ H_2^{(q)} &: q_2 - q_3 = 0 \\ &\vdots \\ H_a^{(q)} &: q_{a-1} - q_a = 0. \end{aligned}$$

ist dabei eine Variante. Bei Verwendung der relativen Kontrasteffekte (Konietschke, 2009) können die sequentiellen Kontrasteffekte, wie in Abschnitt 4.2.1 gezeigt, zu scheinbar paradoxen Aussagen führen. Mit den transitiven Effekten besteht diese Gefahr nicht.

CRF-ab

Ist der Versuchsplan in zwei Faktoren A und B aufgeteilt, stellt sich die Frage, ob die nicht-parametrischen Effekte in den Stufen von Faktor A oder Faktor B unterschiedlich sind. Außerdem ist die Frage nach einer Wechselwirkung zwischen beiden Faktoren zu beantworten. Im Folgenden wird der in der ANOVA typischerweise verwendete Test gegen das Mittel P_a genutzt. Als erstes werden die Tests auf Haupteffekte betrachtet. Die Globalhypothese für den Test auf Haupteffekt A wird mit

$$\begin{aligned} H_1^{(q)}(A) &: \bar{q}_{1\cdot} - \bar{q}_{\cdot\cdot} = 0 \\ &\vdots \\ H_a^{(q)}(A) &: \bar{q}_{a\cdot} - \bar{q}_{\cdot\cdot} = 0 \end{aligned}$$

beschrieben. In diesem Fall ist die Kontrastmatrix dann $C = C_A = P_a \otimes \frac{1}{b} \mathbf{1}'_b$. Auf ähnliche Weise formuliert man die Globalhypothese für den Test auf Haupteffekt B durch

$$\begin{aligned} H_1^{(q)}(B) &: \bar{q}_{\cdot 1} - \bar{q}_{\cdot\cdot} = 0 \\ &\vdots \\ H_b^{(q)}(B) &: \bar{q}_{\cdot b} - \bar{q}_{\cdot\cdot} = 0, \end{aligned}$$

mit $C = C_B = \frac{1}{a} \mathbf{1}'_a \otimes P_b$. Dabei sind $\bar{q}_{i\cdot} = \frac{1}{a} \sum_{i=1}^a q_{ij}$, $\bar{q}_{\cdot j} = \frac{1}{b} \sum_{j=1}^b q_{ij}$ und $\bar{q}_{\cdot\cdot} = \frac{1}{ab} \sum_{i,j} q_{ij}$ für $i = 1, \dots, a$, und $j = 1, \dots, b$. Die jeweilige Globalhypothese setzt sich aus Teilhypothesen zusammen, in denen die Mittelwerte über die einzelnen Stufen der Faktoren A und B mit der Konstante der Gesamtmittelwerte verglichen werden. Zur besseren Veranschaulichung kann diese Konstante auch vernachlässigt werden, so dass die Globalhypothesen für die Tests auf Haupteffekte auch durch

$$H_0^{(q)}(A): \bar{q}_{1\cdot} = \bar{q}_{2\cdot} = \dots = \bar{q}_{a\cdot} \quad (4.11)$$

und

$$H_0^{(q)}(B): \bar{q}_{\cdot 1} = \bar{q}_{\cdot 2} = \dots = \bar{q}_{\cdot b} \quad (4.12)$$

ausgedrückt werden können. Die (Zweifach-) Wechselwirkung zwischen zwei Faktoren A und B wird mit $C = C_{AB} = P_a \otimes P_b$ getestet, und die Teilhypothesen haben dann die

folgende Gestalt:

$$\begin{aligned}
 H_1^{(q)}(AB) &: q_{11} - \bar{q}_{1.} - \bar{q}_{.1} + \bar{q}_{..} = 0 \\
 &\vdots \\
 H_b^{(q)}(AB) &: q_{1b} - \bar{q}_{1.} - \bar{q}_{.b} + \bar{q}_{..} = 0 \\
 &\vdots \\
 H_{ab}^{(q)}(AB) &: q_{ab} - \bar{q}_{a.} - \bar{q}_{.b} + \bar{q}_{..} = 0.
 \end{aligned}$$

Eine Wechselwirkung besagt, dass sich zwei Faktoren gegenseitig beeinflussen. So ist dann das Profil der Werte für eine Stufe von Faktor A, über die Stufen des Faktors B, verschieden zum Profil für eine andere Stufe des Faktors A (siehe ABBILDUNG 4.1). So können die Wirkungen von Medikamenten für weibliche und männliche Probanden unterschiedlich sein, so dass diese beiden Gruppen getrennt voneinander untersucht werden müssen.

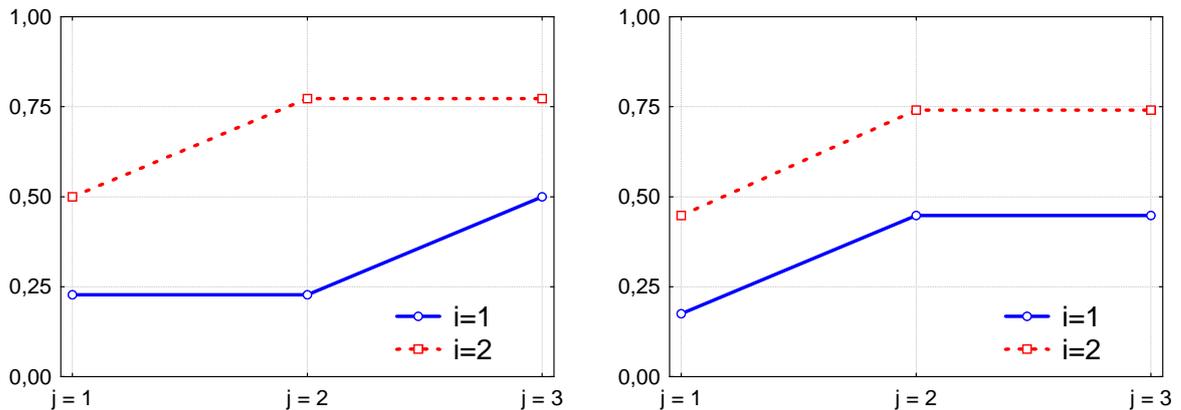


ABBILDUNG 4.1: Beispiel für den Verlauf der Werte von q , getrennt nach den Stufen von Faktor A, im MODELL CRF-23, mit Wechselwirkung (links) und ohne (rechts).

CRF-abc

Die Erweiterungen der Kontrastmatrizen für Haupteffekt A und B auf das MODELL 4.3 sind $C_A = P_a \otimes \frac{1}{b} \mathbf{1}'_b \otimes \frac{1}{c} \mathbf{1}'_c$ und $C_B = \frac{1}{a} \mathbf{1}'_a \otimes P_b \otimes \frac{1}{c} \mathbf{1}'_c$. Für Haupteffekt C lässt ich analog die

Kontrastmatrix $C_C = \frac{1}{a}\mathbf{1}'_a \otimes \frac{1}{b}\mathbf{1}'_b \otimes P_c$ bilden. Die Teilhypothesen schreiben sich dann als

$$H_i^{(q)}(A) : \bar{q}_{i..} - \bar{q}_{...} = 0,$$

$$H_j^{(q)}(B) : \bar{q}_{.j.} - \bar{q}_{...} = 0 \quad \text{und}$$

$$H_l^{(q)}(C) : \bar{q}_{.l} - \bar{q}_{...} = 0$$

für $i = 1, \dots, a, j = 1, \dots, b$ und $l = 1, \dots, c$. Für die Wechselwirkungen ist $C_{AB} = P_a \otimes P_b \otimes \frac{1}{c}\mathbf{1}'_c$, $C_{AC} = P_a \otimes \frac{1}{b}\mathbf{1}'_b \otimes P_c$ und $C_{BC} = \frac{1}{a}\mathbf{1}'_a \otimes P_b \otimes P_c$. So haben die Teilhypothesen die Gestalt

$$H_{ij}^{(q)}(AB) : \bar{q}_{ij.} - \bar{q}_{i..} - \bar{q}_{.j.} + \bar{q}_{...} = 0,$$

$$H_{il}^{(q)}(AC) : \bar{q}_{i.l} - \bar{q}_{i..} - \bar{q}_{.l} + \bar{q}_{...} = 0 \quad \text{und}$$

$$H_{jl}^{(q)}(BC) : \bar{q}_{.jl} - \bar{q}_{.j.} - \bar{q}_{.l} + \bar{q}_{...} = 0.$$

Die Haupteffekte und die (Zweifach-) Wechselwirkungen lassen sich wie im Modell CRF- ab interpretieren. Eine Dreifach-Wechselwirkung ABC wird mit der Kontrastmatrix $C_{ABC} = P_a \otimes P_b \otimes P_c$ getestet und die Teilhypothesen sind

$$H_{ijl}^{(q)}(ABC) : q_{ijk} - \bar{q}_{.jl} - \bar{q}_{i.l} - \bar{q}_{ij.} + \bar{q}_{i..} + \bar{q}_{.j.} + \bar{q}_{.l} - \bar{q}_{...} = 0.$$

Je mehr Faktoren sich gegenseitig beeinflussen, umso schwieriger wird die Interpretation. Daher ist es wichtig, nicht zu viele Faktoren in einen Versuchsplan einzubauen.

4.3.3 Hypothesenkombinationen

Im Folgenden wird auf die Möglichkeit eingegangen, die Kontrastmatrizen C beliebig aus einzelnen Kontrasten c'_m zusammensetzen. Am Beispiel der Globalhypothese (4.8) im MODELL 4.2 soll ein Vorteil der multiplen Kontrasttestprozedur (MCTP) gezeigt werden. Bekannt ist (siehe z.B. Brunner und Munzel, 2000), dass in der ANOVA die Hypothese $H_0^{(q)}(A)$ mit der Kontrastmatrix $C_A = P_a \otimes \frac{1}{b}\mathbf{1}'_b$ getestet werden kann. Die Globalhypothese kann aber ebenfalls durch Teilhypothesen beschrieben werden, welche andere Informationen enthalten als die Abweichungen vom Mittelwert $\bar{q}_i - \bar{q}_{...}$. Neben dem Schnitt der Average-Hypothesen erhält man die globale Hypothese (4.11) auch aus dem Schnitt der Hypothesen von Dunnett (4.9) und der sequentiellen Hypothesen (4.10), angewandt auf die Stufen des Faktors A. So gilt:

$$H_0^{(q)}(A): \bar{q}_1 = \bar{q}_2 = \dots = \bar{q}_a \Leftrightarrow \bigcap_{i=1}^a \{\bar{q}_i = \bar{q}_{..}\} \Leftrightarrow \bigcap_{i=2}^a \{\bar{q}_i = \bar{q}_1\} \Leftrightarrow \bigcap_{i=1}^{a-1} \{\bar{q}_i = \bar{q}_{(i+1)}\}. \quad (4.13)$$

Diese vier Globalhypothesen sind äquivalent und das Ereignis „kein Haupteffekt A“ lässt sich demnach unter anderem durch diese verschiedenen Arten definieren. So wird gegen den Mittelwert verglichen, gegen eine Kontrollgruppe und gegen die nachfolgende Stufe. Je nach den Bedürfnissen des Anwenders kann der eine oder andere Weg im Mittelpunkt des Interesses stehen. Für rein globale Tests, wie die mit der ANOVA, gehen die Informationen aus den Teilhypothesen verloren, so dass für diese eine Variation der Teilhypothesen uninteressant ist.

4.3.4 Sequentielles Testproblem

Eine Art der Kombination ist auch die Vereinigung mehrerer Haupteffekte und Wechselwirkungen in einer Hypothese. Dies lässt sich auf das sequentielle Testproblem anwenden.

Im Blickpunkt steht hier der korrekte Umgang mit dem Datensatz \mathbf{X} . Beispielhaft sei das MODELL 4.2 gegeben. Bei nicht vorhandener Wechselwirkung kann der Datensatz zusammenbleiben, wie in TABELLE 4.2, da die Verläufe parallel sind (ABBILDUNG 4.1 rechts). Die Globalhypothesen Haupteffekt A und B definieren sich wie in (4.11) bzw. (4.12) über die gemittelten Effekte. Ist hingegen eine Wechselwirkung AB vorhanden, so sind die entsprechenden Verläufe unterschiedlich (ABBILDUNG 4.1 links) und die Schätzer verzerrt. In diesem Fall muss für jeden Verlauf einzeln überprüft werden, ob ein Effekt existiert.

Dafür werden die Datensätze aufgetrennt. Soll auf Haupteffekt A getestet werden, dann wird der Datensatz zwischen den Stufen des Faktors B aufgetrennt und man erhält b einfaktorielle Datensätze mit jeweils a Stufen (siehe TABELLE 4.3).

TABELLE 4.3: In b aufgetrennte Datensätze

<table border="1" style="border-collapse: collapse; width: 60px; height: 60px;"> <tr><td style="padding: 2px 10px;">A_1</td><td style="padding: 2px 10px;">X_{111} \vdots $X_{11n_{11}}$</td></tr> <tr><td style="padding: 2px 10px;">\vdots</td><td style="padding: 2px 10px;">\vdots</td></tr> <tr><td style="padding: 2px 10px;">A_a</td><td style="padding: 2px 10px;">X_{a11} \vdots $X_{a1n_{a1}}$</td></tr> </table>	A_1	X_{111} \vdots $X_{11n_{11}}$	\vdots	\vdots	A_a	X_{a11} \vdots $X_{a1n_{a1}}$...	<table border="1" style="border-collapse: collapse; width: 60px; height: 60px;"> <tr><td style="padding: 2px 10px;">A_1</td><td style="padding: 2px 10px;">X_{1b1} \vdots $X_{1bn_{1b}}$</td></tr> <tr><td style="padding: 2px 10px;">\vdots</td><td style="padding: 2px 10px;">\vdots</td></tr> <tr><td style="padding: 2px 10px;">A_a</td><td style="padding: 2px 10px;">X_{ab1} \vdots $X_{abn_{ab}}$</td></tr> </table>	A_1	X_{1b1} \vdots $X_{1bn_{1b}}$	\vdots	\vdots	A_a	X_{ab1} \vdots $X_{abn_{ab}}$
A_1	X_{111} \vdots $X_{11n_{11}}$													
\vdots	\vdots													
A_a	X_{a11} \vdots $X_{a1n_{a1}}$													
A_1	X_{1b1} \vdots $X_{1bn_{1b}}$													
\vdots	\vdots													
A_a	X_{ab1} \vdots $X_{abn_{ab}}$													

Die entsprechenden Datensätze seien mit $\mathbf{X}_a^{(1)}, \dots, \mathbf{X}_a^{(b)}$ bezeichnet, wobei

$$\mathbf{X}_a^{(j)} = (X_{1j1}, \dots, X_{1jn_{1j}}, \dots, X_{aj1}, \dots, X_{ajn_{aj}}) \text{ für } j = 1, \dots, b.$$

Für Haupteffekt B, wird \mathbf{X} zwischen den a Stufen des Faktors A getrennt (siehe TABELLE 4.4).

Analog ergeben sich die Datensätze $\mathbf{X}_b^{(1)}, \dots, \mathbf{X}_b^{(a)}$, mit

$$\mathbf{X}_b^{(i)} = (X_{i11}, \dots, X_{i1n_{i1}}, \dots, X_{ib1}, \dots, X_{ibn_{ib}}) \text{ für } i = 1, \dots, a.$$

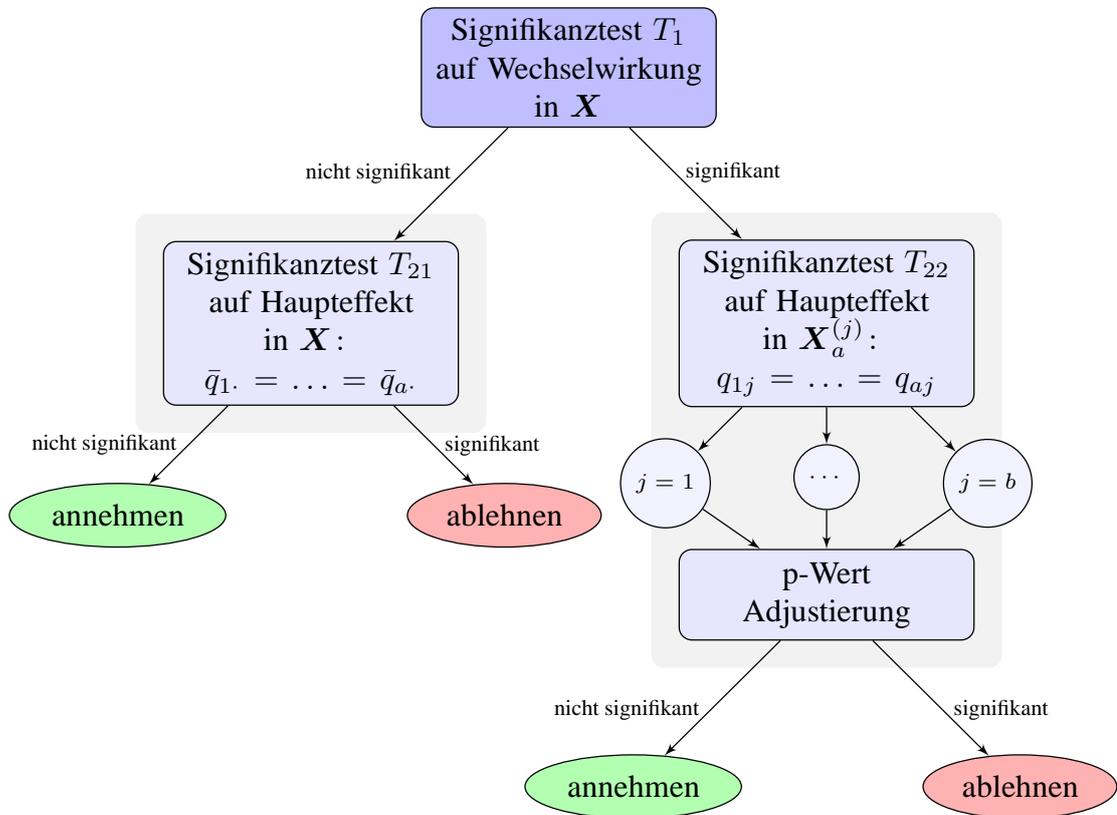


ABBILDUNG 4.2: Ablauf einer Sequentiellen Testprozedur im MODELL 4.2 auf Haupteffekt A.

Durch das sequentielle Vorgehen mit dem Vortest auf Wechselwirkung und dem folgenden Test auf Haupteffekt, jeweils zum Niveau α , hängen die Testentscheidungen voneinander ab, so dass der Test das Niveau nicht einhält. Dies wird durch Simulationen (siehe AB-BILDUNG 6.3 in Abschnitt 6) bestätigt. Eine weiterer Grund liegt darin, dass bei fälschlicher Weise vorher getesteter Wechselwirkung und danach folgender Auftrennung der Daten das Niveau, beim Test auf Haupteffekt, massiv überschritten wird (siehe AB-BILDUNG 6.4). Eine Möglichkeit um das Niveau einzuhalten bietet die in Abschnitt 4.3.2 angesprochene eigenständige Kombinierbarkeit von Kontrasten. Die Variation innerhalb dieser ist recht vielfältig. Zum einen könnten die Kontraste der Haupteffekte und Wechselwirkung in einen größeren Kontrast vereinigt werden. Die Kontraste von Wechselwirkung und Haupteffekt werden also gleichzeitig (simultan) getestet, und das globale Niveau wird eingehalten. Dies sei hier mit *Variante 1* bezeichnet. Angewandt auf das Beispiel von AB-BILDUNG 4.2 würde

sich die Kontrastmatrix

$$\mathbf{C} = (\mathbf{C}'_{AB} : \mathbf{C}'_A)' \quad (4.16)$$

ergeben. Der rechte Arm (die Auftrennung der Daten nach Wechselwirkung) der ABILDUNG 4.2 würde beibehalten werden. Die Tests auf Wechselwirkung T_1 und Haupteffekt T_{21} hingegen würden mit den entsprechenden Kontrasten aus \mathbf{C} untersucht werden. Das Auftrennen der Daten ist aber stilistisch nicht erstrebenswert, so dass Kontraste gesucht sind welche diesen Schritt ersetzen. Die Hypothesen welche diese Informationen ausdrücken sind (für Haupteffekt A) $q_{1j} = \dots = q_{qj}$ für $j = 1, \dots, b$. Die daraus resultierende Kontrastmatrix ist

$$\mathbf{C}_A^{(B)} = (\mathbf{P}_a \otimes \mathbf{e}_b^{(1)} : \dots : \mathbf{P}_a \otimes \mathbf{e}_b^{(b)})', \quad (4.17)$$

so dass die Kontrastmatrix

$$\mathbf{C}_{Seq,A} = (\mathbf{C}'_{AB} : \mathbf{C}'_A : \mathbf{C}_A^{(B)'})' \quad (4.18)$$

alle wichtigen Informationen beinhaltet, um das sequentielle Testproblem mit einer Matrix zu beschreiben und gleichzeitig das multiple Niveau nicht zu überschreiten (hier: *Variante 2*). Bei gleichzeitigem Interesse für den Faktor B könnte die Matrix auch die Gestalt

$$\mathbf{C}_{Seq} = (\mathbf{C}'_{AB} : \mathbf{C}'_A : \mathbf{C}'_B : \mathbf{C}_A^{(B)' : \mathbf{C}_B^{(A)'})', \quad (4.19)$$

mit $\mathbf{C}_B^{(A)' = \text{diag}(a) \otimes \mathbf{P}_b$ haben. Es ist aber zu beachten, dass die Matrix nicht zu viele Kontraste enthält, da die Tests ansonsten sehr konservativ werden. So sollte vorher überlegt werden, welche Informationen aus den Daten gewonnen werden sollen und welche Kontraste „wirklich wichtig“ sind. Der Test, welcher die sequentielle Testprozedur umgeht und dafür die Kontrastmatrix verwendet, wird in dieser Arbeit als *Test unter Verwendung von \mathbf{C}_{Seq} (bzw. $\mathbf{C}_{Seq,A}$ oder $\mathbf{C}_{Seq,B}$)* bezeichnet. Eine Anwendung dazu ist in Abschnitt 7 den Beispielauswertungen, insbesondere Beispiel 7.2 zu finden, sowie Simulationen im Abschnitt 6.

5 Theorie

Im ersten Teil dieses Abschnitts werden Schätzer für die Effekte gegeben. Danach wird die Rangstatistik vorgestellt. Mit dessen asymptotischen Verteilung werden anschließend simultane Tests entwickelt. Damit lassen sich dann unter anderem simultane Konfidenzintervalle bestimmen. Zum Schluss wird eine bessere Approximation für kleine Stichprobenumfänge und außerdem eine bereichserhaltende Transformation gezeigt.

5.1 Punktschätzer und Rangdarstellungen

In diesem Abschnitt soll ein Überblick der Schätzer von den Verteilungsfunktionen und den benötigten relativen Effekten gegeben werden. Außerdem wird die jeweilige Erwartungstreue und Konsistenz gezeigt. Die grundlegendsten Bausteine sind die Verteilungsfunktionen, welche nicht bekannt sind und daher geschätzt werden. Um auch nichtstetige Verteilungsfunktionen zuzulassen, wird die normalisierte empirische Version verwendet, die z.B. auch in Brunner und Munzel (2002, S.31) definiert ist.

DEFINITION 5.1. (*Normalisierte Empirische Verteilungsfunktion*)

Für eine Stichprobe X_1, \dots, X_n mit $X_k \sim F(x)$ und $k = 1, \dots, n$ Beobachtungen, heißt die Funktion

$$\hat{F}(x) = \frac{1}{n} \sum_{k=1}^n c(x - X_k)$$

normalisierte empirische Verteilungsfunktion von X_1, \dots, X_n , wobei $c(x)$ die normalisierte Zählfunktion aus DEFINITION A.14 ist.

Im Folgenden wird diese kurz als *empirische Verteilungsfunktion* bezeichnet. In den Rangdarstellungen für die Schätzer der Verteilungsfunktionen werden die Mittelränge (A.15) verwendet, da diese ebenfalls über die normalisierte Zählfunktion definiert sind. Die Rangvergabe ist in statistischen Softwarepaketen implementiert und daher in der Praxis gut anwendbar. Ist eine Menge von d Stichproben gegeben, so seien die empirischen Verteilungsfunktionen im Vektor $\hat{\mathbf{F}} = (\hat{F}_1, \dots, \hat{F}_d)'$ angeordnet. Soll eine empirische Verteilungsfunktion an einer festen Stelle ausgewertet werden, erhält man die Rangdarstellung

$$\hat{F}_t(X_{uk}) = \frac{1}{n_t} \left(R_{uk}^{(tu)} - R_{uk}^{(u)} \right) \quad (5.1)$$

für $t \neq u; t, u = 1, \dots, d$ und $k = 1, \dots, n_u$. Dabei ist $R_{uk}^{(u)}$ der interne Rang der Zufallsvariable X_{uk} in der Stichprobe u und mit $R_{uk}^{(tu)}$ wird der Rang der Zufallsvariable R_{uk} innerhalb der Stichprobe bezeichnet, in welcher die t -te und u -te Stichprobe zusammengefasst sind. Der nächste Schritt führt direkt zu den Punktschätzern für die paarweisen relativen Effekte r_{tu} aus (4.1):

$$\hat{r}_{tu} = \int \hat{F}_t d\hat{F}_u = \frac{1}{n_t} \left(R_{u \cdot}^{(tu)} - \frac{n_u + 1}{2} \right) = \frac{1}{n_u} \sum_{k=1}^{n_u} \hat{F}_t(X_{uk}). \quad (5.2)$$

Diese Darstellung erhält man ebenfalls über die Zählfunktion und da

$$R_{uk}^{(tu)} = \frac{1}{2} + \sum_{k'=1}^{n_t} c(X_{uk} - X_{tk'}) + \sum_{k'=1}^{n_u} c(X_{uk} - X_{uk'}) \text{ gilt. Außerdem ist}$$

$R_{u \cdot}^{(tu)} = \frac{1}{n_u} \sum_{k=1}^{n_u} R_{uk}$. Die Schätzer der gemittelten Verteilungsfunktionen (4.2) und (4.3) könnten nun mit Rängen geschrieben werden (vgl. Kulle, 1999, S.13 und S.43), dies wird aber nicht benötigt, da die Darstellung aus PROPOSITION 4.5 gilt.

PROPOSITION 5.2. (*Darstellung des Punktschätzers für die ungewichteten relativen Effekte*)
Seien $\hat{\mathbf{q}} = (\hat{q}_1, \dots, \hat{q}_d)'$ und $\hat{\mathbf{r}} = (\hat{r}_{11}, \dots, \hat{r}_{dd})'$ die Punktschätzer der Effekte (4.1) und (4.5). Weiterhin sei \mathbf{W} die Matrix (4.7). Dann gilt:

$$\hat{\mathbf{q}} = \mathbf{W} \hat{\mathbf{r}}.$$

Beweis. Der Beweis geht analog zur PROPOSITION 4.5, nur dass die Effekte durch ihre Schätzer ersetzt werden. \square

Nebenbei erwähnt sei, dass $\hat{\mathbf{q}} = \int G d\hat{\mathbf{F}}$ gilt. Wichtig sind noch die Eigenschaften der empirischen Verteilungsfunktion und der Punktschätzer für die Effekte (4.1) und (4.5). Diese stehen für die Qualität der Schätzer. Ein erwartungstreuer Schätzer besagt, dass der Erwartungswert des Schätzers gleich dem zu schätzenden Wert ist. Während die Konsistenz bedeutet, dass der Schätzer in Wahrscheinlichkeit gegen den wahren Wert konvergiert. Für die empirischen Verteilungsfunktionen gelten die Eigenschaften aus dem folgendem Lemma.

LEMMA 5.3. (*Empirische Verteilungsfunktion*)

Es seien $\hat{F}_t(x), t = 1, \dots, d$ empirische Verteilungsfunktionen. Dann gilt unter Voraussetzung (VI) $\forall t$:

- (1) $E(\hat{F}_t(x)) = F_t(x),$
- (2) $\|\hat{F}_t - F_t\|_\infty \xrightarrow{f.s.} 0, n_t \rightarrow \infty.$

Beweis. Siehe z.B. Domhof (2001, Satz 4.2). \square

Damit sind die Erwartungstreue und starke Konsistenz empirischer Verteilungsfunktionen gezeigt. Daraus folgen für die empirische ungewichtete Verteilungsfunktion

$$\hat{G}(x) = \frac{1}{d} \sum_{t=1}^d \hat{F}_t(x) \quad (5.3)$$

die gleichen Eigenschaften. Nach Brunner und Munzel (2002, Korollar 4.6) gilt auch die L_2 -Konsistenz von \widehat{F}_t zu F_t .

LEMMA 5.4. (*Paarweiser relativer Effekt*)

Es seien $\widehat{r}_{tu} = \int \widehat{F}_t d\widehat{F}_u$ die Punktschätzer für $r_{tu} = \int F_t dF_u$ und $N_{tu} = n_t + n_u$, dann gilt unter Voraussetzung (VI) $\forall t, u = 1, \dots, d$:

$$(1) \quad E(\widehat{r}_{tu}) = r_{tu},$$

$$(2) \quad \|\widehat{\mathbf{r}} - \mathbf{r}\|_2^2 = \sum_{t=1}^d \sum_{u=1}^d E(\widehat{r}_{tu} - r_{tu})^2 \xrightarrow{N_{ij} \rightarrow \infty} 0.$$

Beweis. Siehe z.B. Brunner, Munzel und Puri (2002, Proposition 3.2). □

LEMMA 5.4 (2) besagt die (komponentenweise) L_2 -Konsistenz des Punktschätzers \mathbf{r} . Weiterhin folgt aus der starken Konsistenz der Verteilungsfunktionen die starke Konsistenz der nichtparametrischen Effekte (4.5).

LEMMA 5.5. Es sei $\widehat{\mathbf{q}} = \mathbf{W}\widehat{\mathbf{r}}$ der Punktschätzer für $\mathbf{q} = \mathbf{W}\mathbf{r}$, dann gilt unter Voraussetzung (VI):

$$(1) \quad E(\widehat{\mathbf{q}}) = \mathbf{q},$$

$$(2) \quad \widehat{\mathbf{q}} - \mathbf{q} \xrightarrow{f.s.} 0.$$

Beweis. Die erste Behauptung folgt mit der Linearität des Erwartungswertes und mit LEMMA 5.4 (1), die zweite Behauptung lässt sich mit Koniettschke (2009, A.2) zeigen. □

ANMERKUNG 5.1. Mit Koniettschke (2009, A.2) gilt auch die starke Konsistenz von \mathbf{r} . Außerdem gilt mit LEMMA 5.4 (2) die (komponentenweise) L_2 -Konsistenz des Punktschätzers $\widehat{\mathbf{q}}$.

Für die Kontraste $\mathbf{c}'_m \mathbf{q}$ können die gleichen Eigenschaften gezeigt werden.

LEMMA 5.6. Es sei $\mathbf{c}'_m \widehat{\mathbf{q}}$ der Punktschätzer für $\mathbf{c}'_m \mathbf{q}$, dann gilt unter Voraussetzung (VI):

$$(1) \quad E(\mathbf{c}'_m \widehat{\mathbf{q}}) = \mathbf{c}'_m \mathbf{q},$$

$$(2) \quad \mathbf{c}'_m (\widehat{\mathbf{q}} - \mathbf{q}) \xrightarrow{f.s.} 0.$$

Beweis. Der Vektor \mathbf{c}'_m besteht aus Konstanten. So folgen die beiden Behauptungen mit der Linearität des Erwartungswertes und der Stetigkeit der Funktion $h(\mathbf{q}) = \mathbf{c}'_m \mathbf{q}$. □

5.2 Multiple Kontrasttests

Zum Testen von Kontrasten bzw. Hypothesen wird zuerst die asymptotische Verteilung der Rangstatistik

$$\sqrt{N}(\widehat{\mathbf{q}} - \mathbf{q}) \tag{5.4}$$

benötigt. Diese soll in diesem Kapitel hergeleitet werden. Danach wird die unbekannte Kovarianzmatrix geschätzt, woraus sich dann die Teststatistiken und Konfidenzintervalle ergeben. Behandelt werden außerdem eine bessere Approximation der Verteilung der Prüfgrößen und bereichserhaltende Transformationen für die Konfidenzintervalle.

5.2.1 Asymptotische Verteilung

Für asymptotische Resultate der Rangstatistik (5.4) können, wie in 3.4.1 bereits erwähnt, zentrale Grenzwertsätze nicht direkt angewendet werden, da die Summen von Rängen unabhängiger Zufallsvariablen nicht unabhängig sind. Es besteht aber die Möglichkeit einen Vektor $\sqrt{N}\mathbf{W}\mathbf{B}_N$ von Zufallsvariablen zu finden, welcher asymptotisch die gleiche Verteilung wie die Rangstatistik (5.4) besitzt und dessen Komponenten Summen von unabhängigen Zufallsvariablen sind. Asymptotische Äquivalenzsätze bilden für dieses Thema die Grundlage, mit denen eine beliebig genaue Annäherung von (5.4) an $\sqrt{N}\mathbf{W}\mathbf{B}_N$, für N groß genug, gezeigt werden kann. In der Arbeit von Kulle (1999) wird für die ungewichteten relativen Effekte die Schreibweise $\mathbf{q} = \int Gd\mathbf{F}$ genommen und der Asymptotische Äquivalenzsatz hat dann die folgende Gestalt.

SATZ 5.7. (Asymptotischer Äquivalenzsatz I)

Seien $\hat{G}(x)$ und $\hat{F}_t(x)$ die in (5.3) und (5.1) definierten Schätzer der Verteilungsfunktionen $G(x)$ und $F_t(x)$ und es gelte die VORAUSSETZUNG (V1), dann gilt

$$\sqrt{N} \left(\int \hat{G}d(\hat{\mathbf{F}} - \mathbf{F}) \right) \doteq \sqrt{N} \left(\int Gd(\hat{\mathbf{F}} - \mathbf{F}) \right).$$

Beweis. Siehe Kulle (1999, Satz 4.8). □

Aus diesem Satz folgt für die Rangstatistik (5.4) die Darstellung

$$\sqrt{N}(\hat{q}_t - q_t) \doteq \sqrt{N} \left(\int Gd\hat{F}_t - \int F_t d\hat{G} + 1 - 2q_t \right) \quad (5.5)$$

für die t-te Komponente oder allgemein zusammengefasst im Vektor

$$\sqrt{N}(\hat{\mathbf{q}} - \mathbf{q}) \doteq \sqrt{N} \left(\int Gd\hat{\mathbf{F}} - \int \mathbf{F}d\hat{G} + \mathbf{1}_d - 2\mathbf{q} \right). \quad (5.6)$$

Eine weitere Möglichkeit besteht darin, die Relationen aus den PROPOSITIONEN 4.5 und 5.1 auszunutzen, so dass (5.4) als $\sqrt{N}(\hat{\mathbf{q}} - \mathbf{q}) = \sqrt{N}(\mathbf{W}\hat{\mathbf{r}} - \mathbf{W}\mathbf{r}) = \sqrt{N}\mathbf{W}(\hat{\mathbf{r}} - \mathbf{r}) = \mathbf{W}\sqrt{N}(\hat{\mathbf{r}} - \mathbf{r})$ geschrieben werden kann. Für die Rangstatistik der paarweisen relativen Effekte $\sqrt{N}(\hat{\mathbf{r}} - \mathbf{r})$ kann die Verallgemeinerung des Nichtparametrischen Behrens-Fisher Problems von Brunner, Munzel und Puri (2002) genommen werden. Dafür werden zunächst die Zufallsvariablen

$$B_{tu} = \left(\int F_t d\hat{F}_u - \int F_u d\hat{F}_t + 1 - 2r \right) = \bar{Y}_{tu} - \bar{Y}_{ut} + 1 - 2r_{tu}, \quad (5.7)$$

definiert, mit $Y_{tuk} = F_t(X_{uk})$, $Y_{utk} = F_u(X_{tk})$ und entsprechend $\bar{Y}_{tu.} = \frac{1}{n_u} \sum_{k=1}^{n_u} Y_{tuk}$, für $u, t = 1, \dots, d$. Leicht einzusehen ist, dass $B_{tt} = 0$ für $t = 1, \dots, d$ und $B_{ut} = -B_{tu}$. Die B_{tu} sind dann im Vektor $\mathbf{B}_N = (B_{11}, \dots, B_{dd})$ zusammengefasst. Im folgenden Satz wird die asymptotische Äquivalenz für die Rangstatistik $\sqrt{N}(\hat{\mathbf{r}} - \mathbf{r})$ geregelt.

SATZ 5.8. (*Asymptotischer Äquivalenzsatz II*)

Für den Vektor der paarweisen relativen Effekte \mathbf{r} und deren Schätzer $\hat{\mathbf{r}}$ (5.2), gilt unter den VORAUSSETZUNGEN 4.1, dass

$$\|\sqrt{N}(\hat{\mathbf{r}} - \mathbf{r}) - \sqrt{N}\mathbf{B}_N\|_2^2 \rightarrow 0.$$

Beweis. Siehe Brunner, Munzel und Puri (2002, Theorem 3.3). □

Aus SATZ 5.8 folgt nun direkt, dass

$$\sqrt{N}(\hat{q}_t - q_t) = \sqrt{N}\mathbf{w}'_t(\hat{\mathbf{r}} - \mathbf{r}) \doteq \sqrt{N}\frac{1}{d} \sum_{u=1}^d B_{ut} \quad (5.8)$$

und als Vektor

$$\sqrt{N}(\hat{\mathbf{q}} - \mathbf{q}) = \sqrt{N}\mathbf{W}(\hat{\mathbf{r}} - \mathbf{r}) \doteq \sqrt{N}\mathbf{W}\mathbf{B}_N. \quad (5.9)$$

Durch einfaches nachrechnen kann gezeigt werden, dass die Darstellungen (5.5) und (5.8) identisch sind.

KOROLLAR 5.9. Für die Rangstatistik (5.4) seien die beiden asymptotischen Äquivalenzen (5.5) und (5.8) gegeben, dann gilt

$$\sqrt{N} \left(\int Gd\hat{F}_t - \int F_t d\hat{G} + 1 - 2q_t \right) = \sqrt{N}\frac{1}{d} \sum_{u=1}^d B_{ut} = \sqrt{N}\bar{B}_{.t}.$$

Beweis. Für die Darstellung (5.8) gilt mit der Definition der Zufallsvariablen aus (5.7)

$$\begin{aligned} \sqrt{N}\frac{1}{d} \sum_{u=1}^d B_{ut} &= \sqrt{N}\frac{1}{d} \sum_{u=1}^d (\bar{Y}_{ut.} - \bar{Y}_{tu.} + 1 - 2r_{ut}) \\ &= \sqrt{N}(\bar{Y}_{.t.} - \bar{Y}_{t..} + 1 - 2\bar{r}_{.t}). \end{aligned}$$

Das gleiche Resultat erhält man für (5.5), mit $\int F_u d\hat{F}_t = \bar{Y}_{ut.}$ und $\int F_t d\hat{F}_u = \bar{Y}_{tu.}$:

$$\begin{aligned} &\sqrt{N} \left(\int Gd\hat{F}_t - \int F_t d\hat{G} + 1 - 2 \int GdF_t \right) \\ &= \sqrt{N} \left(\int \frac{1}{d} \sum_{u=1}^d F_u d\hat{F}_t - \int \frac{1}{d} \sum_{u=1}^d F_t d\hat{F}_u + 1 - 2 \int \frac{1}{d} \sum_{u=1}^d F_u dF_t \right) \\ &= \sqrt{N} \left(\frac{1}{d} \sum_{u=1}^d (\bar{Y}_{ut.} - \bar{Y}_{tu.} + 1 - 2\bar{r}_{.t}) \right) \\ &= \sqrt{N}(\bar{Y}_{.t.} - \bar{Y}_{t..} + 1 - 2\bar{r}_{.t}). \end{aligned}$$

Außerdem gilt $\sqrt{N}(\bar{Y}_{.t.} - \bar{Y}_{t..} + 1 - 2\bar{r}_{.t}) = \sqrt{N}\bar{B}_{.t}$. □

Natürlicherweise gilt diese Identität auch für (5.6) und (5.9). Für den Beweis der asymptotischen Normalität von $\sqrt{N}(\hat{\mathbf{q}} - \mathbf{q})$ wird eine weitere Voraussetzung benötigt, da die Kovarianzmatrix $\mathbf{V}_N = \text{Cov}(\sqrt{N}\mathbf{W}\mathbf{B}_N)$ von den Stichprobenumfängen abhängt.

VORAUSSETZUNG 5.10.

(V3) Die Kovarianzmatrix \mathbf{V}_N konvergiere gegen \mathbf{V} , so dass $\text{rang}(\mathbf{V}_N) = \text{rang}(\mathbf{V})$.

Für die Darstellung (5.6) wird in Kulle (1999) die asymptotische Verteilung hergeleitet, so dass sich mit KOROLLAR 5.9 die gleichen Ergebnisse für (5.9) ergeben.

SATZ 5.11. (Asymptotische Normalität)

Sind die VORAUSSETZUNGEN (V1) und (V3) erfüllt, dann folgt $\sqrt{N}(\hat{\mathbf{q}} - \mathbf{q})$ asymptotisch einer multivariaten Normalverteilung mit Erwartungswert $\mathbf{0}$ und Kovarianzmatrix $\mathbf{V}_N = \text{Cov}(\sqrt{N}\mathbf{W}\mathbf{B}_N)$.

Beweis. Folgt aus Kulle (1999, Satz 4.15) unter Verwendung des KOROLLARS 5.9. □

Wie in Brunner, Munzel und Puri (2002, Theorem 3.5) zu lesen ist, kann die asymptotische Normalität von Rangstatistiken nichtparametrischer Effekte auch auf einem anderen Weg bewiesen werden. Für den Beweis wird aber vorausgesetzt, dass die Eigenwerte $\lambda_1, \dots, \lambda_d$ von \mathbf{V}_N größer als Null sind. Die Kovarianzmatrix \mathbf{V}_N erfüllt diese Eigenschaft nicht, da diese nicht regulär ist. Das Thema der Singularität wird im nächsten Abschnitt kurz behandelt.

5.2.2 Singularität

Bereits in der Arbeit von Boysen (2002), wurde die Singularität der Kovarianzmatrix im Bereich der nichtparametrischen Effekte untersucht. Zunächst kann man feststellen, dass die Summe aller ungewichteten relativen Effekte eine feste Konstante ist.

KOROLLAR 5.12.

Für die Effekte (4.5), gilt

$$\sum_{t=1}^d q_t = \frac{d}{2}.$$

Beweis. Mit $r_{tu} + r_{ut} = 1$, folgt

$$\sum_{t=1}^d q_t = \frac{1}{d} \sum_{u=1}^d \sum_{t=1}^d r_{tu} = \frac{1}{d} \left(\underbrace{d \frac{1}{2}}_{t=u} + \sum_{u=1}^d \sum_{\substack{t=1 \\ t \neq u}}^d r_{tu} \right) = \frac{1}{2} + \frac{1}{d} \frac{d(d-1)}{2} = \frac{d}{2}.$$

□

Im Mittel hat ein Effekt q_t also den Wert $1/2$. Da auch $\hat{r}_{tu} + \hat{r}_{ut} = 1$ gilt überträgt sich die Eigenschaft auf die Schätzer \hat{q}_t . Sind alle Werte von \mathbf{q} bis auf einen bekannt, lässt sich der letzte unbekannte Wert, o.B.d.A. sei dies q_d , bestimmen: $q_d = d/2 - \sum_{t=1}^{d-1} q_t$. Für den Vektor

$\hat{\mathbf{q}}$ erhält man einen linearen Zusammenhang, dafür sei $\mathbf{A} = (\mathbf{I}_{d-1} \vdots -\mathbf{1}_{d-1})'$, $\mathbf{b} = (\mathbf{0}'_d, d/2)'$ und $\mathbf{q}^{(-d)} = (q_1, \dots, q_{d-1})'$ der zu \mathbf{q} identische Vektor ohne Eintrag den q_d . Dann gilt

$$\mathbf{q} = \mathbf{A}\mathbf{q}^{(-d)} + \mathbf{b}.$$

Damit lässt sich zeigen, dass die Kovarianzmatrix \mathbf{V}_N singulär ist. Denn es gilt

$$\mathbf{V}_N \doteq Cov(\sqrt{N}(\hat{\mathbf{q}} - \mathbf{q})) = \mathbf{A}Cov(\sqrt{N}(\hat{\mathbf{q}}^{(-d)} - \mathbf{q}^{(-d)}))\mathbf{A}',$$

womit

$$r(Cov(\sqrt{N}(\hat{\mathbf{q}} - \mathbf{q}))) = r(\mathbf{A}Cov(\sqrt{N}(\hat{\mathbf{q}}^{(-d)} - \mathbf{q}^{(-d)}))\mathbf{A}') \leq r(\mathbf{A}) = d - 1 < d$$

folgt. Die Kovarianzmatrix der Rangstatistik (5.4) ist nicht von vollem Rang und so ist mit

$$\det(Cov(\sqrt{N}(\hat{\mathbf{q}} - \mathbf{q}))) = 0$$

auch $\det(\mathbf{V}_N) = 0$.

Wie auch in Boysen (2002) im Punkt 4.0.1 nachzulesen ist, enthält \mathbf{q} also redundante Informationen, so dass ein Wert theoretisch weggelassen werden könnte. Andererseits müssten die bekannten Kontrastmatrizen umgeschrieben werden. Der programmiertechnische Aufwand zur Realisation dieser Theorie würde sich nicht signifikant verringern und für die Interpretation sind, besonders graphisch, alle Werte interessant. Daher ist es nicht notwendig bzw. sinnvoll die weitere Theorie für $\mathbf{q}^{(-d)}$ zu betrachten.

Diese Singularität überträgt sich auch auf den konsistenten Schätzer der Kovarianzmatrix $\hat{\mathbf{V}}_N$, welcher als nächstes hergeleitet wird.

5.2.3 Varianzschätzer

Für die in Abschnitt 5.2.1 eingeführte Kovarianzmatrix

$$\mathbf{V}_N = Cov(\sqrt{N}\mathbf{W}\mathbf{B}_N) = \mathbf{W}Cov(\sqrt{N}\mathbf{B}_N)\mathbf{W}', \quad (5.10)$$

wird nun ein konsistenter Schätzer hergeleitet. Dazu wird zunächst die Kovarianzmatrix $\mathbf{Y}_N := Cov(\sqrt{N}\mathbf{B}_N)$ betrachtet und ihre Struktur näher erläutert, woraus dann nach (5.10) \mathbf{V}_N berechnet wird. Danach werden die Einträge der Kovarianzmatrix \mathbf{Y}_N konsistent geschätzt, woraus sich wiederum ein konsistenter Schätzer für \mathbf{V}_N ergibt. Zuvor soll wie erwähnt die Struktur im Mittelpunkt stehen. Dabei habe die Kovarianzmatrix der Zufallsvariablen \mathbf{B}_{tu} die folgende Gestalt

$$\mathbf{Y}_N = \begin{pmatrix} y_{11,11} & \cdots & y_{11,dd} \\ \vdots & \ddots & \vdots \\ y_{dd,11} & \cdots & y_{dd,dd} \end{pmatrix}_{d^2 \times d^2},$$

mit den Einträgen $y_{tu,rs} = (\mathbf{Y}_N)_{tu,rs} = Cov(\sqrt{N}B_{tu,rs})$. Bekanntermaßen stehen auf der Diagonalen die Varianzen $y_{tu,tu} = Cov(\sqrt{N}B_{tu,tu}, \sqrt{N}B_{tu,tu}) = Var(\sqrt{N}B_{tu})$, für die gezeigt werden kann, dass

$$\begin{aligned} Var(\sqrt{N}B_{tu}) &= NVar\left(\int F_t d\widehat{F}_u - \int F_u d\widehat{F}_t + 1 - 2r_{tu}\right) \\ &= N\left(\frac{1}{n_u^2} \sum_{k=1}^{n_u} Var(F_t(X_{uk})) + \frac{1}{n_t^2} \sum_{k=1}^{n_t} Var(F_u(X_{tk}))\right) \\ &= N\left(\frac{1}{n_u} \sigma_{tu}^2 + \frac{1}{n_t} \sigma_{ut}^2\right) \end{aligned}$$

gilt. Dabei ist zu beachten, dass die Stichproben unverbunden und so die Zufallsvariablen X_{tk} und X_{uk} , für $t \neq u$ unabhängig sind. Dieses Resultat erhält man ebenfalls, wenn die Lösung zum Behrens-Fisher Problem (Abschnitt 3.4.1) auf zwei beliebige Stichproben t und u erweitert wird (Brunner et al., 2002b). Neben den Varianzen müssen weitere Fallunterscheidungen vorgenommen werden. Allgemein gilt

$$\begin{aligned} Cov(\sqrt{N}B_{tu}, \sqrt{N}B_{rs}) &= NCov(\bar{Y}_{tu.} - \bar{Y}_{ut.}, \bar{Y}_{rs.} - \bar{Y}_{sr.}) \\ &= N(Cov(\bar{Y}_{tu.}, \bar{Y}_{rs.}) - Cov(\bar{Y}_{tu.}, \bar{Y}_{sr.}) - Cov(\bar{Y}_{ut.}, \bar{Y}_{rs.}) + Cov(\bar{Y}_{ut.}, \bar{Y}_{sr.})). \end{aligned}$$

Durch die Unverbundenheit der Stichproben ergibt sich die weitere Kovarianzstruktur, welche in der TABELLE 5.1 gelistet ist. Dabei verschwindet jeweils, für jede gezeigte Kombination, genau ein Summand nicht. Zu beachten ist, dass $NCov(\bar{Y}_{ut.}, \bar{Y}_{rs.}) = -NCov(\bar{Y}_{tu.}, \bar{Y}_{rs.})$ gilt, wie auch $NCov(\bar{Y}_{tu.}, \bar{Y}_{sr.}) = -NCov(\bar{Y}_{tu.}, \bar{Y}_{rs.})$. Für $(t \wedge u) \neq (r \vee s)$ hingegen gilt $Cov(B_{tu}, B_{rs}) = 0$, wie auch für $(t = u) \vee (r = s)$, da dann $\bar{Y}_{tt} = 0$ bzw. $\bar{Y}_{rr} = 0$ konstant sind.

TABELLE 5.1: Struktur der Kovarianzmatrix \mathbf{Y}_N

Index	nicht verschwindender Summand	$NCov(B_{tu}, B_{rs})$
$t = r \wedge u \neq s$	$NCov(\bar{Y}_{ut.}, \bar{Y}_{st.})$	$y_{tu,ts} = \frac{N}{n_t} Cov(F_u(X_{t1}), F_s(X_{t1}))$
$s = u \wedge t \neq r$	$NCov(\bar{Y}_{tu.}, \bar{Y}_{ru.})$	$y_{tu,ru} = \frac{N}{n_u} Cov(F_t(X_{u1}), F_r(X_{u1}))$
$t = s \wedge u \neq r$	$NCov(\bar{Y}_{ut.}, \bar{Y}_{rt.})$	$y_{tu,rt} = -\frac{N}{n_t} Cov(F_u(X_{t1}), F_r(X_{t1}))$
$u = r \wedge t \neq s$	$NCov(\bar{Y}_{tu.}, \bar{Y}_{su.})$	$y_{tu,ts} = -\frac{N}{n_u} Cov(F_t(X_{u1}), F_s(X_{u1}))$

Die Schätzung der Einträge erfolgt in zwei Schritten. Unter der Annahme, dass die Zufallsvariablen Y_{tuk} beobachtbar sind, können die Varianzen und Kovarianzen als erstes erwar-

tungstreu aus den Werten \mathbf{X} (siehe Abschnitt 4.1) geschätzt werden. Für die „Varianzschätzer“ gilt dann

$$\tilde{\sigma}_{tu}^2 = \frac{1}{n_u - 1} \sum_{k=1}^{n_u} (Y_{tuk} - \bar{Y}_{tu.})^2 \quad \text{und} \quad \tilde{\sigma}_{ut}^2 = \frac{1}{n_t - 1} \sum_{k=1}^{n_t} (Y_{utk} - \bar{Y}_{ut.})^2. \quad (5.11)$$

Ähnlich sieht dies mit der Stichprobenkovarianz aus, so ist

$$\tilde{y}_{tu,ts} = \widetilde{Cov}(\sqrt{N}B_{tu}, \sqrt{N}B_{ts}) = \frac{N}{n_t(n_t - 1)} \sum_{k=1}^{n_t} (Y_{utk} - \bar{Y}_{ut.})(Y_{stk} - \bar{Y}_{st.}), \quad (5.12)$$

am Beispiel von $t = r \wedge u \neq s$. Allerdings sind die Verteilungsfunktionen F_t nicht bekannt, und dadurch sind die Y_{tuk} nicht beobachtbar. Daher ist die Bezeichnung von (5.11) und (5.12) als „Schätzer“ formell nicht ganz richtig, da diese von unbekanntem Werten abhängen. Als zweites werden daher die Verteilungsfunktionen durch die empirischen Verteilungsfunktionen \hat{F}_t (5.1) ersetzt, da diese wegen der L_2 -Konsistenz „nah genug“ an den nicht-beobachtbaren Zufallsvariablen $Y_{tuk} = F_t(X_{uk})$ liegen. So erhält man die beobachtbaren Zufallsvariablen

$$\hat{Y}_{tuk} = \hat{F}_t(X_{uk}) \quad \text{und} \quad \hat{Y}_{tu.} = \frac{1}{n_u} \sum_{k=1}^{n_u} \hat{F}_t(X_{uk}). \quad (5.13)$$

Diese Schätzer können nach Orban und Wolfe (1980, 1982) mit Rängen dargestellt werden, so dass

$$\hat{Y}_{tuk} = \frac{1}{n_t} (R_{uk}^{(tu)} - R_{uk}^{(u)}) \quad \text{und} \quad \hat{Y}_{tu.} = \frac{1}{n_t} (\bar{R}_{u.}^{(tu)} - \frac{n_u + 1}{2}),$$

wobei mit $R_{uk}^{(tu)}$ der Rang der Zufallsvariable X_{uk} in der Stichprobe $\mathbf{X}_{(tu)} = (\mathbf{X}'_t, \mathbf{X}'_u)'$ bezeichnet wird. Wie in (5.11) und (5.12) zu erkennen ist, werden die Differenzen der Schätzer (5.13) benötigt. So wird für die Übersichtlichkeit diese Darstellung abgekürzt

$$D_{tuk} = \hat{Y}_{tuk} - \hat{Y}_{tu.} = \frac{1}{n_t} \left(R_{uk}^{(tu)} - R_{uk}^{(u)} - \bar{R}_{u.}^{(tu)} + \frac{n_u + 1}{2} \right).$$

Damit gilt für die Schätzer der Varianzen für die Zufallsvariablen Y_{tuk}

$$\hat{\sigma}_{tu}^2 = \frac{1}{n_u(n_u - 1)} \sum_{k=1}^{n_u} D_{tuk}^2, \quad \text{und} \quad \hat{\sigma}_{ut}^2 = \frac{1}{n_t(n_t - 1)} \sum_{k=1}^{n_t} D_{utk}^2,$$

so dass dann

$$\hat{y}_{tu,tu} = N \left(\frac{\hat{\sigma}_{tu}^2}{n_u} + \frac{\hat{\sigma}_{ut}^2}{n_t} \right) \quad (5.14)$$

gilt. Die Schätzer der Kovarianzen aus TABELLE 5.1 sind TABELLE 5.2 zu entnehmen. Auch für die Schätzer gilt dann $\hat{y}_{tt,rs} = \hat{y}_{tu,rr} = 0$ und $\hat{y}_{ut,rs} = -\hat{y}_{tu,rs}$ bzw. $\hat{y}_{tu,sr} = -\hat{y}_{tu,rs}$.

ANMERKUNG 5.2. Der so konstruierte Varianzschätzer (5.14) wird trotz der VORAUSSETZUNG (V1) gleich Null, wenn die Stichproben X_t und X_u voneinander getrennt sind. In diesem Fall muss für die praktische Anwendung der Schätzer ersetzt werden und es bietet sich die kleinste von Null weg verschiedene Varianz an.

TABELLE 5.2: Konsistente Kovarianzschätzer

Index	$N\widehat{Cov}(B_{tu}, B_{rs})$
$t = r \wedge u \neq s$	$\widehat{y}_{tu,ts} = \frac{N}{n_t(n_t-1)} \sum_{k=1}^{n_t} D_{utk} D_{stk}$
$s = u \wedge t \neq r$	$\widehat{y}_{tu,ru} = \frac{N}{n_u(n_u-1)} \sum_{k=1}^{n_u} D_{tuk} D_{ruk}$
$t = s \wedge u \neq r$	$\widehat{y}_{tu,rt} = -\frac{N}{n_t(n_t-1)} \sum_{k=1}^{n_t} D_{utk} D_{rtk}$
$u = r \wedge t \neq s$	$\widehat{y}_{tu,us} = -\frac{N}{n_u(n_u-1)} \sum_{k=1}^{n_u} D_{tuk} D_{suk}$

Wie für die nichtparametrischen Effekte kann für den Kovarianzschätzer die starke Konsistenz gezeigt werden.

SATZ 5.13. (Stark Konsistenter Kovarianzschätzer $\widehat{\mathbf{Y}}_N$)

Mit der VORAUSSETZUNG (V1) sind die Schätzer aus (5.14) und TABELLE 5.2 stark konsistent.

Beweis. Siehe Konietschke (2009, Satz 6.2.1). □

Wie in (5.10) gilt für die Schätzer die Beziehung $\widehat{\mathbf{V}}_N = \mathbf{W}\widehat{\mathbf{Y}}_N\mathbf{W}'$, mit $\widehat{\mathbf{Y}}_N = (\widehat{y}_{tu,rs})_{t,u,r,s=1,\dots,d}$ und so folgt aus SATZ 5.13 die starke Konsistenz von $\widehat{\mathbf{V}}_N$. Aus der nun bekannten asymptotischen Verteilung der Rangstatistik (5.4) und unter Verwendung des Kovarianzschätzers könnte ein Test für die Hypothese $H_0 : \mathbf{q} = 1/2 \cdot \mathbf{1}_d$ hergeleitet werden. Im einfaktoriellen Modell könnte dies genügen, im mehrfaktoriellen weniger. Ein wichtiger Punkt ist aber die Möglichkeit, auf die Anwenderfragen einzugehen. Dafür wurden in Abschnitt 4.3 mögliche interessante Kontrastmatrizen gezeigt. Um solche Kontraste zu testen, muss die asymptotische Verteilung von $\sqrt{N}(\mathbf{C}\widehat{\mathbf{q}} - \mathbf{C}\mathbf{q})$ bestimmt werden. Aus $\sqrt{N}(\mathbf{C}\widehat{\mathbf{q}} - \mathbf{C}\mathbf{q}) = \mathbf{C}\sqrt{N}(\widehat{\mathbf{q}} - \mathbf{q})$ folgt mit SATZ 5.11 und den Eigenschaften von Erwartungswert und Varianz bei linearen Transformationen

$$\sqrt{N}(\mathbf{C}\widehat{\mathbf{q}} - \mathbf{C}\mathbf{q}) \overset{d}{\rightsquigarrow} N(\mathbf{0}, \mathbf{V}_N^{(C)}), \quad (5.15)$$

mit $\mathbf{V}_N^{(C)} = \mathbf{C}\mathbf{V}_N\mathbf{C}'$. Auch hier folgt, dass $\widehat{\mathbf{V}}_N^{(C)} = \mathbf{C}\widehat{\mathbf{V}}_N\mathbf{C}'$ stark konsistent ist.

Mit dem Wissen über die Asymptotische Verteilung und den Schätzer der Kovarianzmatrix können jetzt Tests hergeleitet werden.

5.2.4 Simultane Testverfahren

Dieser Abschnitt beschäftigt sich mit der Theorie von multiplen Testverfahren, insofern sie für diese Arbeit wichtig ist; die bisherige Notation wird beibehalten. Diese Vorarbeit ist wichtig, um die starke Kontrolle der FWER (*familywise error rate*) sicher zu stellen. Für mehr Informationen über multiple Testverfahren sei auf Hochberg und Tamhane (1987), verwiesen. Zunächst sei die Hypothesenfamilie durch

$$\Omega = \{H_m^{(q)} : \mathbf{c}'_m \mathbf{q} = 0, m = 1, \dots, q\} \quad (5.16)$$

beschrieben. Zu jeder Teilhypothese $H_m^{(q)}$ lässt sich weiterhin eine studentisierte Teststatistik

$T_m = \sqrt{N} \frac{\mathbf{c}'_m (\hat{\mathbf{q}} - \mathbf{q})}{\sqrt{\hat{v}_m^{(C)}}}$, mit $\hat{v}_m^{(C)} = (\hat{\mathbf{V}}_N^{(C)})_{m,m}$, angeben. Die T_m sind asymptotisch standardnormalverteilt und die Statistiken werden im Vektor

$$\mathbf{T} = (T_1, \dots, T_q)' = \left(\sqrt{N} \frac{\mathbf{c}'_1 (\hat{\mathbf{q}} - \mathbf{q})}{\sqrt{\hat{v}_1^{(C)}}}, \dots, \sqrt{N} \frac{\mathbf{c}'_q (\hat{\mathbf{q}} - \mathbf{q})}{\sqrt{\hat{v}_q^{(C)}}} \right)' \quad (5.17)$$

zusammengefasst. Die multivariate Normalverteilung von \mathbf{T} lässt sich wie folgt herleiten.

Die Rangstatistik $\sqrt{N}(\mathbf{C}\hat{\mathbf{q}} - \mathbf{C}\mathbf{q})$ ist nach (5.15) asymptotisch multivariat normalverteilt, woraus sich in bekannter Weise standardisierte Teststatistiken ergeben

$$\mathbf{T}^* = (T_1^*, \dots, T_q^*)' = \left(\sqrt{N} \frac{\mathbf{c}'_1 (\hat{\mathbf{q}} - \mathbf{q})}{\sqrt{v_1^{(C)}}}, \dots, \sqrt{N} \frac{\mathbf{c}'_q (\hat{\mathbf{q}} - \mathbf{q})}{\sqrt{v_q^{(C)}}} \right)',$$

mit $v_m^{(C)} = (\mathbf{V}_N^{(C)})_{m,m}$ dem m-ten Diagonaleintrag der Kovarianzmatrix $\mathbf{V}_N^{(C)}$. Für die einzelnen T_m^* , $m = 1, \dots, q$ gilt dann, dass sie asymptotisch standardnormalverteilt sind und weiterhin für den Vektor der Statistiken \mathbf{T}^* , dass $E(\mathbf{T}^*) = \mathbf{0}$. Die einzelnen Einträge aus \mathbf{T}^* sind nicht unabhängig voneinander und so ist $Cov(\mathbf{T}^*) = \mathbf{E}_N^{(C)}$ die asymptotische Kovarianzmatrix, welche gleichzeitig die Korrelationsmatrix von $\mathbf{V}_N^{(C)}$ ist. Als Einträge erhält man auf der Diagonalen

$$\rho_{m,m} = \frac{1}{v_m^{(C)}} Var \left(\sqrt{N} \mathbf{c}'_m \mathbf{W} \mathbf{B}_N \right) = 1$$

und auf gleiche Weise für die Nichtdiagonaleinträge

$$\rho_{m,n} = \frac{Cov(\mathbf{c}'_m \mathbf{W} \mathbf{B}_N, \mathbf{c}'_n \mathbf{W} \mathbf{B}_N)}{\sqrt{Var(\mathbf{c}'_m \mathbf{W} \mathbf{B}_N) Var(\mathbf{c}'_n \mathbf{W} \mathbf{B}_N)}}.$$

Die Varianzen $v_m^{(C)}$ und die Korrelationsmatrix $\mathbf{E}_N^{(C)}$ sind nicht bekannt und müssen noch durch entsprechende Schätzer ersetzt werden. Für diesen Schritt kann die Vorarbeit aus dem Abschnitt 5.2.3 entnommen werden, so dass man stark konsistente Schätzer durch $\hat{v}_m^{(C)} = (\hat{\mathbf{V}}_N^{(C)})_{m,m}$ und $\hat{\mathbf{E}}_N^{(C)}$, mit den Einträgen $\hat{\rho}_{m,n} = \frac{\hat{v}_{m,n}^{(C)}}{\sqrt{\hat{v}_m^{(C)} \hat{v}_n^{(C)}}}$, erhält. $\hat{\mathbf{E}}_N^{(C)}$ ist so mit die geschätzte Korrelationsmatrix von $\sqrt{N}\mathbf{C}(\hat{\mathbf{q}} - \mathbf{q})$, die mit den Einträgen von $\hat{\mathbf{V}}_N^{(C)}$ konstruiert wird, woraus deren starke Konsistenz sofort folgt. Damit erhält man unter VORAUSSETZUNG (V1) für den Vektor \mathbf{T} (5.17) die asymptotische multivariate Normalität

$$\mathbf{T} = (T_1, \dots, T_q)' = \left(\sqrt{N} \frac{\mathbf{c}'_1(\hat{\mathbf{q}} - \mathbf{q})}{\sqrt{\hat{v}_1^{(C)}}}, \dots, \sqrt{N} \frac{\mathbf{c}'_q(\hat{\mathbf{q}} - \mathbf{q})}{\sqrt{\hat{v}_q^{(C)}}} \right)' \rightsquigarrow N(\mathbf{0}, \hat{\mathbf{E}}_N^{(C)}). \quad (5.18)$$

Das Tupel $\{\Omega, \mathbf{T}\}$ ist dann nach Gabriel (1969) eine asymptotische Testfamilie, da jeder Teilhypothese $H_m^{(q)}$ eine Prüfgröße T_m zugeordnet werden kann, deren Verteilung unter $H_m^{(q)}$ vollständig bestimmt ist. Auch die asymptotische Verteilung des Vektors der Prüfgrößen \mathbf{T} ist unter der Globalhypothese

$$H_0^{(q)} : \mathbf{C}\mathbf{q} = \bigcap_{m=1}^q \{\mathbf{c}'_m \mathbf{q} = 0\} = \mathbf{0} \quad (5.19)$$

mit der multivariaten Normalverteilung vollständig bestimmt. Speziell gilt diese Eigenschaft auch für $\mathbf{T}^J = (T_j, j \in J)$ mit der Teilmenge $J \subseteq I = \{1, \dots, q\}$ unter der Hypothese $\tilde{H}_0^{(q)} : \bigcap_{j \in J} H_j^{(q)} = 0$. Nach Gabriel (1969), werden diese asymptotische Testfamilien *gemeinsam* genannt.

SATZ 5.14. (*Asymptotisch gemeinsame Testfamilie*)

Die Testfamilie $\{\Omega, \mathbf{T}\}$ ist eine asymptotisch gemeinsame Testfamilie nach Gabriel (1969).

Beweis. Zu zeigen ist, dass für jede Teilmenge $J \subseteq I$ die asymptotisch gemeinsame Verteilung der Prüfgrößen \mathbf{T}^J vollständig unter der Hypothese $\tilde{H}_0^{(q)}$ bestimmt ist. Die Gedanken, welche zu diesem Schluss führen, sind in Koniettschke (2009, Satz 6.3.1) nachzulesen. \square

Die konsistenten Varianzschätzer $\hat{v}_m^{(C)}$ wurden im Abschnitt 5.2.3 unabhängig von der Gültigkeit der Teil- und Globalhypothese hergeleitet. Im Gegensatz dazu fordern andere nicht-parametrische multiple Kontrasttests, z.B. Munzel und Hothorn (2001), die Gültigkeit der Globalhypothese, so dass diese Testfamilien die Eigenschaften für das Attribut *gemeinsam* nicht erfüllen. Um Aussagen bezüglich der Hypothesen zu treffen, wird in Gabriel (1969) der Begriff der simultanen Testfamilie eingeführt. Dazu wird das Tupel (Ω, \mathbf{T}) um eine Konstante $\xi \in \mathbb{R}$ erweitert. Für einen speziellen Schwellenwert ξ , wird $(\Omega, \mathbf{T}, \xi)$ auch als simultane Testprozedur (STP) bezeichnet. Theoretisch könnte zur Bewertung jeder einzelnen Teilhypothese $H_m^{(q)}$ durch die Prüfgröße T_m jeweils ein eigener Schwellenwert ξ_m genommen werden. In Hochberg und Tamhane (1987, S.29f.) werden Gründe aufgeführt,

warum es genügt und vorteilhafter ist einen gemeinsamen Wert ξ zu verwenden. Durch die asymptotische multivariate Normalität der Prüfgrößen T (5.18) und der in Abschnitt 5.2.6 beschriebenen Approximation mit der t-Verteilung sind gerade die Quantile der beiden genannten Verteilungen gefragt. Ein nicht trivialer Algorithmus zur numerischen Berechnung wird in Bretz, Genz und Hothorn (2001) vorgestellt. Diese Quantile, welche die Korrelation berücksichtigen, werden als *equikoordinate* Quantile bezeichnet, da in jeder Koordinate der gleiche Wert steht. Zum besseren Verständnis von equikoordinaten Quantilen sollen diese an einem Beispiel veranschaulicht werden. Die ABBILDUNGEN 5.1¹ und 5.2 sind aus Konietschke (2009, S.53-56) entnommen. Dazu ist ein bivariatverteilter Vektor \mathbf{X} gegeben, mit Erwartungswert $\mathbf{0}$ und Korrelationsmatrix \mathbf{E} , so dass $\mathbf{X} = (X_1, \dots, X_q) \sim N(\mathbf{0}, \mathbf{E})$. Das equikoordinate multivariate zweiseitige $(1 - \alpha)$ -Quantil wird mit $z_{1-\alpha,2,\mathbf{E}}$ bezeichnet und ist definiert durch

$$P\left(\bigcap_{m=1}^q \{-z_{1-\alpha,2,\mathbf{E}} \leq X_m \leq z_{1-\alpha,2,\mathbf{E}}\}\right) = 1 - \alpha.$$

Entsprechend ist das obere equikoordinate multivariate einseitige $(1 - \alpha)$ -Quantil $z_{1-\alpha,1,\mathbf{E}}^o$ als

$$P\left(\bigcap_{m=1}^q \{z_{1-\alpha,1,\mathbf{E}}^o \leq X_m\}\right) = 1 - \alpha$$

und das untere $z_{1-\alpha,1,\mathbf{E}}^u$ als

$$P\left(\bigcap_{m=1}^q \{X_m \leq z_{1-\alpha,1,\mathbf{E}}^u\}\right) = 1 - \alpha$$

definiert. Aus der Symmetrie der Normalverteilung bezüglich des Erwartungswertes $\mathbf{0}$ folgt, dass $z_{1-\alpha,1,\mathbf{E}}^u = -z_{1-\alpha,1,\mathbf{E}}^o$. Zweiseitige und obere einseitige Quantile sind in der ABBILDUNG 5.1 beispielhaft gezeigt.

Eine Eigenschaft solcher Quantile ist die Abhängigkeit von der Korrelation. Dies wird in der ABBILDUNG 5.2 verdeutlicht. Dazu wurden 10.000 bivariate $N(\mathbf{0}, \mathbf{I}_2 + \rho(\mathbf{J}_2 - \mathbf{I}_2))$ -verteilte Zufallsvariablen mit den Korrelationskoeffizienten $\rho = \{0, 0.9\}$ erzeugt und dann die entsprechenden equikoordinaten bivariaten zweiseitigen 95%-Quantile berechnet. So werden die Werte der equikoordinaten Quantile mit wachsender Korrelation kleiner und passen sich der ellipsoiden Form der Verteilung an.

Für die Teilhypothesen können mittels der Prüfgrößen T_m und dem Schwellenwert ξ Aussagen getroffen werden. Zur Herleitung des Schwellenwertes ξ kann die multivariate Normalverteilung von T (5.18) herangezogen werden. Die simultane Testprozedur kontrolliert die FWER im starken Sinne zum Niveau α nach Definition, wenn ξ so gewählt wird, dass

$$P(-\xi \leq T_1 \leq \xi, \dots, -\xi \leq T_q \leq \xi) = 1 - \alpha$$

¹In Konietschke (2009) werden die Abbildungen der ein- und zweiseitigen Quantile für eine bivariate t-Verteilung angelegt. Bei einer bivariaten Normalverteilung sind die Niveaulinien ebenfalls Ellipsen, bzw. Kreise bei Unkorreliertheit, so dass diese auch in diesem Fall ihre Daseinsberechtigung haben.

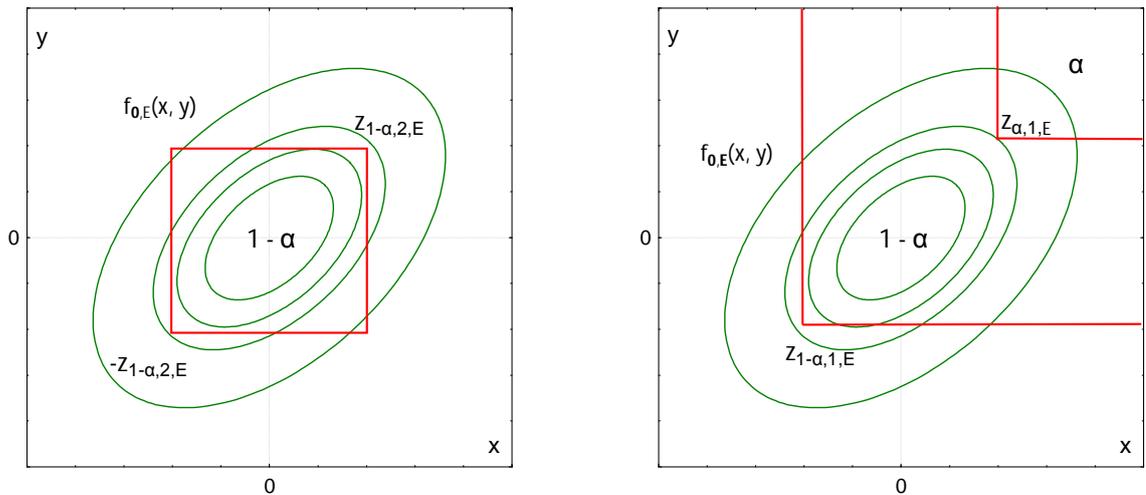


ABBILDUNG 5.1: Equikoordinate bivariate Quantile von $f(N_1, N_2) \sim t_{0,E}$: zweiseitiges $1 - \alpha$ (links), obere einseitige für α und $1 - \alpha$ (rechts).

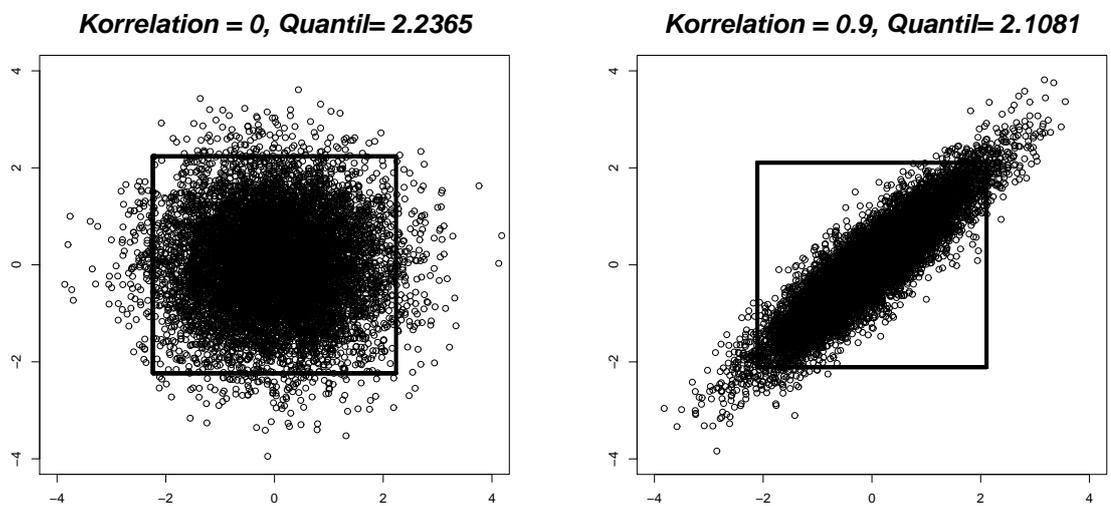


ABBILDUNG 5.2: Zweiseitige Equikoordinate 95%Quantile von bivariat $N(\mathbf{0}, \mathbf{I}_2 + \rho(\mathbf{J}_2 - \mathbf{I}_2))$ -verteilten Zufallsvariablen für $\rho = \{0, 0.9\}$.

gilt. Die theoretischen Erkenntnisse aus diesem Abschnitt gelten auch für die einseitigen Hypothesenmengen

$$\Omega_o = \{H_m^{(q)} : \mathbf{c}'_m \mathbf{q} \leq 0, m = 1, \dots, q\}$$

und

$$\Omega_u = \{H_m^{(q)} : \mathbf{c}'_m \mathbf{q} \geq 0, m = 1, \dots, q\}. \quad (5.20)$$

In den nächsten beiden Abschnitten werden als erstes für die Normalverteilung Teststatistiken, Schwellenwerte und Konfidenzintervalle bestimmt und danach für kleine Stichprobenumfänge eine Approximation mit der t-Verteilung behandelt.

5.2.5 Approximation mit der Normalverteilung

Aus der multivariaten Normalität von \mathbf{T} (5.18) können jetzt Schwellenwerte ξ gewonnen werden, mit denen sich die Statistiken aus \mathbf{T} auswerten und simultane Konfidenzintervalle konstruieren lassen. Zum Schluss kann außerdem die starke Kontrolle der FWER (nach Gabriel, 1969) gezeigt werden.

Für die zweiseitige Alternative erhält man die simultane Testprozedur $(\Omega, \mathbf{T}, z_{1-\alpha,2,\hat{\mathbf{E}}_N^{(C)}})$, mit dem equikoordinaten zweiseitigen $(1 - \alpha)$ -Quantil der $N(\mathbf{0}, \hat{\mathbf{E}}_N^{(C)})$ -Verteilung. Analog sind für die beiden einseitigen Alternativen die simultanen Testprozeduren mit $(\Omega_o, \mathbf{T}, z_{1-\alpha,1,\hat{\mathbf{E}}_N^{(C)}})$ und $(\Omega_u, \mathbf{T}, z_{1-\alpha,1,\hat{\mathbf{E}}_N^{(C)}})$ mit dem entsprechendem einseitigen Quantil $z_{1-\alpha,1,\hat{\mathbf{E}}_N^{(C)}}$ definiert. Da die Verteilung von \mathbf{T} nur asymptotisch gilt, werden die multiplen Niveaus im Folgenden auch nur asymptotisch, also für genügend große Stichprobenumfänge, eingehalten. Da \mathbf{q} unbekannt ist, erfolgt die Auswertung der Hypothese unter der Annahme von $\mathbf{q} = \frac{1}{2}\mathbf{1}_d$, so dass unter Annahme der Hypothese $\mathbf{c}'_m \mathbf{q} = 0$ für $m = 1, \dots, q$ gilt. Die Gestalt von \mathbf{T} ändert sich und erhält daher eine andere Bezeichnung

$$\mathbf{T}_N = (T_{N,1}, \dots, T_{N,q})' = \left(\sqrt{N} \frac{\mathbf{c}'_1 \hat{\mathbf{q}}}{\sqrt{\hat{v}_1^{(C)}}}, \dots, \sqrt{N} \frac{\mathbf{c}'_q \hat{\mathbf{q}}}{\sqrt{\hat{v}_q^{(C)}}} \right)'. \quad (5.21)$$

Für die zweiseitige Alternative wird die Globalhypothese (4.8) zum multiplen Niveau α abgelehnt, wenn

$$\max_{m=1,\dots,q} |T_{N,m}| \geq z_{1-\alpha,2,\hat{\mathbf{E}}_N^{(C)}}. \quad (5.22)$$

Mit der Pivot-Methode erhält man aus den T_m in (5.17) ein simultanes Konfidenzintervall

$$P \left(\bigcap_{m=1}^q \left\{ \mathbf{c}'_m \mathbf{q} \in \left[\mathbf{c}'_m \hat{\mathbf{q}} - z_{1-\alpha,2,\hat{\mathbf{E}}_N^{(C)}} \sqrt{\frac{\hat{v}_m^{(C)}}{N}}, \mathbf{c}'_m \hat{\mathbf{q}} + z_{1-\alpha,2,\hat{\mathbf{E}}_N^{(C)}} \sqrt{\frac{\hat{v}_m^{(C)}}{N}} \right] \right\} \right) \xrightarrow{N \rightarrow \infty} 1 - \alpha,$$

welches kompatibel zur Testentscheidung (5.22) ist. Für die einzelnen Kontraste wird die Teststatistik mit dem Wert aus dem equikoordinaten Quantil verglichen und die einzelnen

Konfidenzintervalle sind in den eckigen Klammern gegeben. Des Weiteren werden die einseitigen Hypothesen zum multiplen Niveau α abgelehnt, wenn für die obere Alternative

$$\max_{m=1,\dots,q} T_{N,m} \geq z_{1-\alpha,1,\widehat{\mathbf{E}}_N^{(C)}}$$

bzw. für die untere

$$\min_{m=1,\dots,q} T_{N,m} = \max_{m=1,\dots,q} -T_{N,m} \geq z_{1-\alpha,1,\widehat{\mathbf{E}}_N^{(C)}}$$

gilt. Die entsprechenden simultanen Konfidenzintervalle sind

$$P \left(\bigcap_{m=1}^q \left\{ \mathbf{c}'_m \mathbf{q} \in \left[\mathbf{c}'_m \widehat{\mathbf{q}} - z_{1-\alpha,1,\widehat{\mathbf{E}}_N^{(C)}} \sqrt{\frac{\widehat{v}_m^{(C)}}{N}}, K \right] \right\} \right) \xrightarrow{N \rightarrow \infty} 1 - \alpha$$

und

$$P \left(\bigcap_{m=1}^q \left\{ \mathbf{c}'_m \mathbf{q} \in \left[-K, \mathbf{c}'_m \widehat{\mathbf{q}} + z_{1-\alpha,1,\widehat{\mathbf{E}}_N^{(C)}} \sqrt{\frac{\widehat{v}_m^{(C)}}{N}} \right] \right\} \right) \xrightarrow{N \rightarrow \infty} 1 - \alpha,$$

mit K aus Abschnitt 4.3.2 Seite 21. Dass diese Vorgehensweise die familienbezogene Fehlerrate einhält, besagt der folgende Satz mit der darauf folgenden Anmerkung.

SATZ 5.15. (*Starke Kontrolle der FWER*)

Es seien die Hypothesenmengen wie in (5.16) und (5.20) definiert. Dann kontrollieren die simultanen Testprozeduren $(\Omega, \mathbf{T}, z_{1-\alpha,2,\widehat{\mathbf{E}}_N^{(C)}})$, $(\Omega_o, \mathbf{T}, z_{1-\alpha,1,\widehat{\mathbf{E}}_N^{(C)}})$ und $(\Omega_u, \mathbf{T}, z_{1-\alpha,1,\widehat{\mathbf{E}}_N^{(C)}})$ die FWER asymptotisch im starken Sinne.

Beweis. (Ω, \mathbf{T}) , (Ω_o, \mathbf{T}) und (Ω_u, \mathbf{T}) , sind asymptotisch gemeinsame Testfamilien und die genannten simultanen Testprozeduren sind nach Hochberg und Tamhane (1987, Theorem 3.2) kohärent. Mit der Kohärenz und da die einzelnen Globalhypothesen abgelehnt werden, wenn mindestens eine Teilhypothese verworfen wird, folgt die Behauptung mit SATZ A.16 (vgl. Konietzschke 2009, Satz 6.4.1). \square

ANMERKUNG 5.3. *Da $\widehat{\mathbf{E}}_N^{(C)}$ ein stark konsistenter Schätzer ist, gelten die Aussagen von SATZ 5.16 auch für die simultanen Testprozeduren $(\Omega, \mathbf{T}, z_{1-\alpha,2,\widehat{\mathbf{E}}_N^{(C)}})$, $(\Omega_o, \mathbf{T}, z_{1-\alpha,1,\widehat{\mathbf{E}}_N^{(C)}})$ und $(\Omega_u, \mathbf{T}, z_{1-\alpha,1,\widehat{\mathbf{E}}_N^{(C)}})$.*

5.2.6 Approximation mit der t-Verteilung

Simulationsstudien haben gezeigt, dass bei der Approximation mit der Normalverteilung das Niveau bzw. multiple Niveau nicht eingehalten wird und die Tests liberal sind, besonders für kleine Stichprobenumfänge. Die finiten univariaten (und multivariaten) $(1 - \alpha)$ -Quantile der t-Verteilung mit ν Freiheitsgraden $t_{1-\alpha,\nu}$ sind stets größer ($\forall \nu$) als die der

Normalverteilung $z_{1-\alpha}$. So wurde für das Behrens-Fisher-Problem (in Abschnitt 3.4.1) von Brunner und Munzel auch ein Freiheitsgrad für den Zweistichprobenfall hergeleitet und anschließend geschätzt. Die Verbindung der ungewichteten relativen Effekte zu den paarweisen relativen Effekte $\mathbf{q} = \mathbf{W}\mathbf{r}$ wurde in PROPOSITION 4.5 gezeigt. Dies ist ähnlich zu den in Koniettschke (2009) konstruierten relativen Kontrasteffekten, welche ebenfalls aus den paarweise gebildeten relativen Effekten gewonnen werden. So wird in dieser Arbeit der Freiheitsgrad der t-Verteilung ν , mit der die zentrierte Rangstatistik verglichen wird, auf ähnliche Weise konstruiert. Dazu wird der Freiheitsgrad (3.3) auf zwei beliebige Verteilungsfunktionen verallgemeinert

$$\hat{f}_{tu} = \frac{(n_t \hat{\sigma}_{tu}^2 + n_u \hat{\sigma}_{ut}^2)^2}{(n_t \hat{\sigma}_{tu}^2)^2 / (n_u - 1) + (n_u \hat{\sigma}_{ut}^2)^2 / (n_t - 1)}$$

gilt. Für den Fall $t = u$ wird $\hat{f}_{tt} = 1$ gesetzt. Letztlich ist der Schätzer des Freiheitsgrades für die t-Verteilung dann

$$\nu = \max \left(1, \min_{r=1, \dots, d} (\mathbf{W}\hat{\mathbf{f}}) \right)$$

wobei $\hat{\mathbf{f}} = (\hat{f}_{11}, \dots, \hat{f}_{dd})'$. Unter der VORAUSSETZUNG (V1) divergieren die Freiheitsgrade f_{tu} , folglich auch ν , und da die t-Verteilung, für $N \rightarrow \infty$, gegen die Normalverteilung konvergiert, sind die asymptotischen Resultate äquivalent. Die Resultate für die Approximation mit der t-Verteilung erhält man ähnlich wie bei der Normal-Approximation, nur dass die Normalverteilungsquantile durch die entsprechenden equikoordinaten $(1 - \alpha)$ -Quantile der $t(\mathbf{0}, \nu, \hat{\mathbf{E}}_N^{(C)})$ -Verteilung $t_{1-\alpha, \nu, \hat{\mathbf{E}}_N^{(C)}}$ ersetzt werden. So wird die Globalhypothese $H_0^{(q)}: \bigcap_{m=1}^q \{\mathbf{c}'_m \mathbf{q} = 0\}$ zum multiplen Niveau α zweiseitig abgelehnt, wenn

$$\max_{m=1, \dots, q} |T_{N,m}| \geq t_{1-\alpha, 2, \nu, \hat{\mathbf{E}}_N^{(C)}}.$$

Die kompatiblen simultanen Konfidenzintervalle sind durch

$$P \left(\bigcap_{m=1}^q \left\{ \mathbf{c}'_m \hat{\mathbf{q}} \in \left[\mathbf{c}'_m \hat{\mathbf{q}} - t_{1-\alpha, 2, \nu, \hat{\mathbf{E}}_N^{(C)}} \sqrt{\frac{\hat{v}_m^{(C)}}{N}}, \mathbf{c}'_m \hat{\mathbf{q}} + t_{1-\alpha, 2, \nu, \hat{\mathbf{E}}_N^{(C)}} \sqrt{\frac{\hat{v}_m^{(C)}}{N}} \right] \right\} \right) \xrightarrow{N \rightarrow \infty} 1 - \alpha$$

gegeben. Für die einseitigen Alternativen werden die Hypothesen $H_{0,o}^{(q)}: \bigcap_{m=1}^q \{\mathbf{c}'_m \mathbf{q} \leq 0\}$ bzw. $H_{0,u}^{(q)}: \bigcap_{m=1}^q \{\mathbf{c}'_m \mathbf{q} \geq 0\}$ zum multiplen Niveau α einseitig abgelehnt, wenn

$$\max_{m=1, \dots, q} T_{N,m} \geq t_{1-\alpha, 1, \nu, \hat{\mathbf{E}}_N^{(C)}}$$

bzw.

$$\max_{m=1, \dots, q} -T_{N,m} \geq t_{1-\alpha, 1, \nu, \hat{\mathbf{E}}_N^{(C)}}.$$

Wiederum dazu kompatible simultane Konfidenzintervalle sind

$$P \left(\bigcap_{m=1}^q \left\{ \mathbf{c}'_m \mathbf{q} \in \left[\mathbf{c}'_m \hat{\mathbf{q}} - t_{1-\alpha, 1, \nu, \hat{\mathbf{E}}_N^{(C)}} \sqrt{\frac{\hat{v}_m^{(C)}}{N}}, K \right] \right\} \right) \xrightarrow{N \rightarrow \infty} 1 - \alpha$$

bzw.

$$P \left(\bigcap_{m=1}^q \left\{ \mathbf{c}'_m \mathbf{q} \in \left[-K, \mathbf{c}'_m \hat{\mathbf{q}} + t_{1-\alpha, 1, \nu, \hat{\mathbf{E}}_N^{(C)}} \sqrt{\frac{\hat{v}_m^{(C)}}{N}} \right] \right\} \right) \xrightarrow{N \rightarrow \infty} 1 - \alpha.$$

Die asymptotisch starke Kontrolle der FWER für die simultanen Testprozeduren $(\Omega, \mathbf{T}, t_{1-\alpha, 2, \hat{\mathbf{E}}_N^{(C)}})$, $(\Omega_o, \mathbf{T}, t_{1-\alpha, 1, \hat{\mathbf{E}}_N^{(C)}})$ und $(\Omega_u, \mathbf{T}, t_{1-\alpha, 1, \hat{\mathbf{E}}_N^{(C)}})$ gilt mit der asymptotischen Gleichheit von Normal- und t-Verteilung und SATZ 5.15.

Bekanntermaßen kann auch der adjustierte p-Wert zur Beurteilung herangezogen werden. Diesen erhält man, in dem die jeweilige Teststatistik $T_{N,m}$ equikoordinat, also als Vektor $T_{N,m} \mathbf{1}_q$, an die multivariate Normal- bzw. t-Verteilung übergeben wird.

5.2.7 Bereichserhaltende Transformation

Die Werte der paarweisen und ungewichteten relativen Effekte \mathbf{r} und \mathbf{q} liegen aus der Definition heraus im Intervall $[0, 1]^q$. Daraus folgt, dass $\mathbf{C}\mathbf{q} \in [-K, K]$ gilt, mit K aus Abschnitt 4.3.2, Seite 21. Diese Eigenschaften gelten auch für die Schätzer. Liegt der wahre Wert $\mathbf{C}\mathbf{q}$ nahe den Grenzen $-K$ bzw. K oder auch bei kleinen Stichprobenumfängen, kann es passieren, dass die Grenzen des Konfidenzintervalls außerhalb des Definitionsbereiches $[-K, K]$ liegen. Diese Intervalle sind in der Praxis dann nicht interpretierbar. Im Folgenden wird daher eine Möglichkeit aufgezeigt, welche auch als Transformationsmethode bezeichnet wird, um die Konfidenzintervalle wieder bereichserhaltend zu transformieren. Dafür sei $\mathbf{g} = (g_1(\mathbf{c}'_1 \mathbf{q}), \dots, g_q(\mathbf{c}'_q \mathbf{q})) : (-K, K)^q \rightarrow \mathbb{R}^q$ eine Funktion welche die Kontraste $\mathbf{c}'_m \mathbf{q}$ auf die reelle Achse \mathbb{R} transformiert und $\mathbf{g}^{-1} : \mathbb{R}^q \rightarrow (-K, K)^q$ die Umkehrfunktion. Dabei sollte \mathbf{g}

VORAUSSETZUNGEN 5.16.

(V4) stetig, bijektiv und streng monoton steigend

sein. Zusätzlich wird gefordert, dass $g_m(\mathbf{c}'_m \mathbf{q})$

(V5) Funktionen sind, mit stetigen ersten Ableitungen und $g'_m(\mathbf{c}'_m \mathbf{q}) \neq 0$.

Die asymptotische Verteilung der Rangstatistik der transformierten Kontraste wird durch den multivariaten δ -Satz von Cramer bestimmt. So ist die Rangstatistik mit den durch \mathbf{g} transformierten Kontrasten asymptotisch multivariat normalverteilt, wenn die VORAUSSETZUNGEN (V1), (V2), (V4) und (V5) erfüllt sind. Dann ist

$$\sqrt{N} (\mathbf{g}(\mathbf{C}\hat{\mathbf{q}}) - \mathbf{g}(\mathbf{C}\mathbf{q})) \overset{\cdot}{\sim} N \left(\mathbf{0}, \mathbf{J}\mathbf{V}_N^{(C)} \mathbf{J}' \right), \quad (5.23)$$

mit der Jacobi-Matrix \mathbf{J} von g

$$\mathbf{J} = \frac{\partial g_m}{\partial \mathbf{c}'_m \mathbf{q}}(\mathbf{C}\mathbf{q}) = g'_m(\mathbf{c}'_m \mathbf{q}), \quad \text{für } m = n$$

und 0 sonst. Daraus folgt, dass $\mathbf{J} = \text{diag}(g'_1(\mathbf{c}'_1 \mathbf{q}), \dots, g'_q(\mathbf{c}'_q \mathbf{q}))$ eine Diagonalmatrix ist. Aus der asymptotischen Verteilung werden, wie zuvor mit der Pivot-Methode, die simultanen Konfidenzintervalle bestimmt und danach die Grenzen mit der Umkehrfunktion zurücktransformiert, so dass diese im gewünschten Intervall liegen und dadurch bereichserhaltend sind. Eigenschaften der Jacobi-Matrix \mathbf{J} und der asymptotischen Kovarianzmatrix $\mathbf{V}_N^{(C)}(\mathbf{g}) = \mathbf{J}\mathbf{V}_N^{(C)}\mathbf{J}'$ sind im nächsten Satz gezeigt.

SATZ 5.17. (Starke Konsistenz)

Es sei $\mathbf{J} = \text{diag}(g'_1(\mathbf{c}'_1 \mathbf{q}), \dots, g'_q(\mathbf{c}'_q \mathbf{q}))$ die Diagonalmatrix von $g(\mathbf{C}\mathbf{q})$. Dann gilt unter der VORAUSSETZUNG (VI), dass $\widehat{\mathbf{J}} = \text{diag}(g'_1(\mathbf{c}'_1 \widehat{\mathbf{q}}), \dots, g'_q(\mathbf{c}'_q \widehat{\mathbf{q}}))$ ein stark konsistenter Schätzer für \mathbf{J} ist und so auch $\widehat{\mathbf{J}}\widehat{\mathbf{V}}_N^{(C)}\widehat{\mathbf{J}}'$. Außerdem sind die Korrelationsmatrizen der Rangstatistiken $\sqrt{N}(\mathbf{C}\widehat{\mathbf{q}} - \mathbf{C}\mathbf{q})$ und $\sqrt{N}(g(\mathbf{C}\widehat{\mathbf{q}}) - g(\mathbf{C}\mathbf{q}))$ asymptotisch identisch.

Beweis. Siehe Koniettschke (2009, Satz 6.6.1). □

ANMERKUNG 5.4. Die Korrelationsmatrizen von $\widehat{\mathbf{V}}_N^{(C)}$ und $\widehat{\mathbf{V}}_N^{(C)}(\mathbf{g})$ sind identisch, da \mathbf{J} eine Diagonalmatrix ist. So ändern sich die equikoordinaten Quantile $z_{1-\alpha, \widehat{\mathbf{E}}_N^{(C)}}$ bzw. $t_{1-\alpha, \widehat{\mathbf{E}}_N^{(C)}}$ durch die Transformation der Effekte nicht.

Eine geeignete Funktion, welche die Eigenschaften erfüllt, hat die Einträge $g_m(\mathbf{c}'_m \mathbf{q}) = \frac{1}{2} \log \left(\frac{K + \mathbf{c}'_m \mathbf{q}}{K - \mathbf{c}'_m \mathbf{q}} \right)$ mit der Umkehrfunktion $g_m^{-1}(\mathbf{c}'_m \mathbf{q}) = \frac{\exp(2\mathbf{c}'_m \mathbf{q}) - 1}{\exp(2\mathbf{c}'_m \mathbf{q}) + 1}$ und den Einträgen $\mathbf{J}_{m,m} = \frac{1}{K^2 - (\mathbf{c}'_m \mathbf{q})^2}$ in \mathbf{J} . Für $K > 1$ sind die transformierten Grenzen in $(-1, 1) \subset (-K, K)$, da $g^{-1}: \mathbb{R}^q \rightarrow (-1, 1)^q$. Um eine Einheitlichkeit in die Transformationen zu bringen, sollte man sich auf $K=1$, also normierte Kontrastmatrizen, beschränken.

Dann folgt weiter $(\mathbf{V}_N^{(C)}(\mathbf{g}))_{m,m} = v_m^{(C)}(\mathbf{g}) = v_m^{(C)} / (K^2 - (\mathbf{c}'_m \mathbf{q})^2)^2$. Analog gilt dies, wenn $\mathbf{c}'_m \widehat{\mathbf{q}}$ transformiert wird. Die Funktion g wird auch als Fisher-Transformation bezeichnet und ist aus Koniettschke (2006, S.67f.) entnommen. Der Vektor der Rangstatistiken ist dann

$$\mathbf{T}_N^g = (T_{N,1}^g, \dots, T_{N,q}^g)' = \left(\frac{\sqrt{N} g_1(\mathbf{c}'_1 \widehat{\mathbf{q}}) \cdot (K^2 - (\mathbf{c}'_1 \widehat{\mathbf{q}})^2)}{\sqrt{\widehat{v}_1^{(C)}}}, \dots, \frac{\sqrt{N} g_q(\mathbf{c}'_q \widehat{\mathbf{q}}) \cdot (K^2 - (\mathbf{c}'_q \widehat{\mathbf{q}})^2)}{\sqrt{\widehat{v}_q^{(C)}}} \right)'$$

und die zweiseitige Alternative der transformierten Globalhypothese wird zum multiplen Niveau α abgelehnt, wenn

$$\max_{m=1, \dots, q} T_{N,m}^g \geq z_{1-\alpha, 2, \widehat{\mathbf{E}}_N^{(C)}}.$$

Das simultane zweiseitige bereichserhaltende Konfidenzintervall ist, unter VORAUSSETZUNG (V1) durch

$$P \left(\bigcap_{m=1}^q \left\{ g_m(\mathbf{c}'_m \mathbf{q}) \in \left[g_m^{-1}(c_{m,u}^g), g_m^{-1}(c_{m,o}^g) \right] \right\} \right) \xrightarrow{N \rightarrow \infty} 1 - \alpha$$

gegeben mit $c_{m,u}^g = g_m(\mathbf{c}'_m \hat{\mathbf{q}}) - z_{1-\alpha, 2, \hat{\mathbf{E}}_N^{(C)}} \sqrt{\frac{\hat{v}_m^{(C)}(\mathbf{g})}{N}}$ und

$c_{m,o}^g = g_m(\mathbf{c}'_m \hat{\mathbf{q}}) + z_{1-\alpha, 2, \hat{\mathbf{E}}_N^{(C)}} \sqrt{\frac{\hat{v}_m^{(C)}(\mathbf{g})}{N}}$, wobei $\hat{v}_m^{(C)}(\mathbf{g}) = \hat{v}_m^{(C)} / (K^2 - (\mathbf{c}'_m \hat{\mathbf{q}})^2)^2$. Einseitige bereichserhaltende simultane Konfidenzintervalle sind durch

$$P \left(\bigcap_{m=1}^q \left\{ g_m(\mathbf{c}'_m \mathbf{q}) \in \left[g_m^{-1} \left(g_m(\mathbf{c}'_m \hat{\mathbf{q}}) - z_{1-\alpha, 1, \hat{\mathbf{E}}_N^{(C)}} \sqrt{\frac{\hat{v}_m^{(C)}(\mathbf{g})}{N}} \right), K \right] \right\} \right) \xrightarrow{N \rightarrow \infty} 1 - \alpha$$

und

$$P \left(\bigcap_{m=1}^q \left\{ g_m(\mathbf{c}'_m \mathbf{q}) \in \left[-K, g_m^{-1} \left(g_m(\mathbf{c}'_m \hat{\mathbf{q}}) + z_{1-\alpha, 1, \hat{\mathbf{E}}_N^{(C)}} \sqrt{\frac{\hat{v}_m^{(C)}(\mathbf{g})}{N}} \right) \right] \right\} \right) \xrightarrow{N \rightarrow \infty} 1 - \alpha$$

gegeben.

5.3 ANOVA in relativen Effekten

Zum Testen globaler Hypothesen können die quadratischen Formen der Wald-Typ Statistik (WTS) und der Anova-Typ Statistik (ATS) genommen werden. Simulationen (Brunner et al., 1997) haben in der Vergangenheit gezeigt, dass die WTS gerade bei kleinen Stichprobenumfängen das Niveau nicht einhält. Daher wird nur die ATS betrachtet. Später in der Simulationsstudie (Kapitel 6) werden globale Ergebnisse der multiplen Kontrasttests und der ATS miteinander verglichen. Die ATS mitsamt ihrer approximativ asymptotischen Verteilung wird im Folgenden übersichtshalber zusammenfassend dargestellt.

Nach (5.15) gilt unter der Hypothese $H_0^{(q)}$ (5.18)

$$\sqrt{N} \mathbf{C} \hat{\mathbf{q}} \overset{\cdot}{\sim} N(\mathbf{0}, \mathbf{C} \mathbf{V}_N \mathbf{C}').$$

Außerdem gilt für $\mathbf{T} = \mathbf{C}'[\mathbf{C}\mathbf{C}']^{-1}\mathbf{C}$

$$\mathbf{C} \mathbf{q} = \mathbf{0} \Leftrightarrow \mathbf{T} \mathbf{q} = \mathbf{0}.$$

Die Äquivalenz der Hypothesen folgt, da $C'[CC']^{-}$ eine verallgemeinerte Inverse zu C ist. Weiterhin ist T ein Projektor und damit symmetrisch und idempotent. Damit folgt aus SATZ A.17

$$N\hat{\mathbf{q}}\mathbf{T}\mathbf{q} \dot{\sim} \sum_{m=1}^q \lambda_m Z_m,$$

mit den unabhängigen Zufallsvariablen $Z_m \sim \chi_1^2$ und den Eigenwerten $\lambda_m, m = 1, \dots, q$, von $\mathbf{T}\mathbf{V}_N$. Die Eigenwerte sind unbekannt und so auch die gewichtete Summe. Diese kann aber nach Box (1954) mit einer gestreckten χ^2 -Verteilung approximiert werden. Mit

VORAUSSETZUNG 5.18.

(V5) $Sp(\mathbf{T}\mathbf{V}_N) \geq k > 0$, k konstant,

wird ausgeschlossen, dass $\sqrt{N}\mathbf{C}\hat{\mathbf{q}}$ asymptotisch eine Ein-Punkt-Verteilung ist. Nach Kulle (1999, S.32f.) gilt, dass

$$F_N(\mathbf{T}) = \frac{N\hat{\mathbf{q}}'\mathbf{T}\hat{\mathbf{q}}}{Sp(\mathbf{T}\mathbf{V}_N)} \dot{\sim} \frac{\chi_f^2}{f}, \quad \text{mit} \quad f = \frac{(Sp(\mathbf{T}\mathbf{V}_N))^2}{Sp(\mathbf{T}\mathbf{V}_N\mathbf{T}\mathbf{V}_N)}.$$

Die Kovarianzmatrix \mathbf{V}_N ist unbekannt und wird durch den Schätzer $\hat{\mathbf{V}}_N$ ersetzt. Die resultierende Verteilung ist im folgenden Satz angegeben.

SATZ 5.19. (Approximative Verteilung der ATS)

Unter den VORAUSSETZUNGEN (VI) und (V5) gilt

$$F_N(\mathbf{T}) = \frac{N\hat{\mathbf{q}}'\mathbf{T}\hat{\mathbf{q}}}{Sp(\mathbf{T}\hat{\mathbf{V}}_N)} \dot{\sim} \frac{\chi_{\hat{f}}^2}{\hat{f}}, \quad \text{mit} \quad \hat{f} = \frac{(Sp(\mathbf{T}\hat{\mathbf{V}}_N))^2}{Sp(\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}\hat{\mathbf{V}}_N)}.$$

Beweis. Siehe Kulle (1999, Satz 3.24 und Satz 4.21). □

6 Simulationen

Zur besseren Beurteilung der entwickelten Tests werden Simulationen durchgeführt. Im Folgenden werden in diesen Bereich fallende Begriffe kurz erklärt. Im Mittelpunkt von Simulationen stehen zum einen die Einhaltung des Niveaus α und zum anderen die Power. Mit dem Niveau eines Tests wird der Fehler 1. Art bezeichnet, welcher die Wahrscheinlichkeit ist, die Nullhypothese fälschlicherweise abzulehnen, d.h. die Hypothese wird verworfen, obwohl diese wahr ist. Wird häufiger als das vorgegebene α abgelehnt, so wird dies als liberal bezeichnet, im Gegensatz dazu hat ein Test konservativen Charakter, wenn seltener als α die Nullhypothese verworfen wird. Mit der Power wird die Wahrscheinlichkeit bezeichnet, mit der ein vorhandener Unterschied aufgedeckt wird. Hält der Test das Niveau schnell ein, also auch für kleine Stichprobenumfänge, und werden möglichst kleine tatsächliche Unterschiede sicher aufgedeckt, dann wird dies für den Test als positiv bewertet.

Die Simulationen wurden zum Niveau $\alpha = 5\%$ mit jeweils $n_{sim} = 10.000$ Wiederholungen durchgeführt. Für die bessere Beurteilung des Niveauperhaltens ist in einzelnen Niveau-graphiken ein 99%-Zufallsstreifen mit eingezeichnet, der die zufälligen Schwankungen der Simulationsergebnisse mit einbezieht. Für $\alpha = 5\%$ und $n_{sim} = 10.000$ erhält man das Intervall $[0.0444, 0.0556]$, d.h. bei 10.000 Simulationen ist das Simulationsergebnis bis auf ein halbes Prozent genau. In den Niveausimulationen wurde weiterhin für unterschiedliche Stichprobenumfänge $n = \{5, 10, 15, 20, \dots, 500\}$ simuliert. Für große Stichprobenumfänge $n > 50$ wurde simuliert, da u.U. das Niveau erst spät eingehalten wird.

Im Folgenden seien, bis nichts anderes erwähnt wird, die Zufallsvariablen, mit denen simuliert wurde, aus der Standardnormalverteilung erzeugt und außerdem die Stichprobenumfänge gleich.

Zum besseren Verständnis für die Interpretation der Ergebnisse muss erwähnt werden, dass sich die Werte der equikoordinaten Quantile vergrößern, wenn zu einer Kontrastmatrix weitere Kontraste hinzugefügt werden, die keine Vielfachen von genau einem bereits vorhandenen Kontrast sind.

Normal- vs. t-Approximation

In einer ersten Simulationsstudie soll das finite Verhalten multipler Testverfahren (siehe Abschnitt 5.2.4) unter Verwendung der Normal-Approximation (Abschnitt 5.2.5) und der t-Approximation (Abschnitt 5.2.6) untersucht werden. Hierzu wurden standardnormalverteilte CRF- a Designs mit $a = \{2, 4, 8\}$ Stufen und CRF- $2b$ Designs mit $b = \{2, 4, 8\}$ Stufen erzeugt. Als Kontrastmatrizen wurden im CRF- a Design einheitlich die zentrierende Matrix $P_a = I_a - \frac{1}{a}J_a$ und im CRF- $2b$ Design die Kontrastmatrix für „kein Haupteffekt B“

$C = \frac{1}{a} \mathbf{1}'_a \otimes P_b$ verwendet, die bereits in Kapitel 4.3 vorgestellt wurden. Die Ergebnisse sind in ABBILDUNG 6.1 dargestellt.

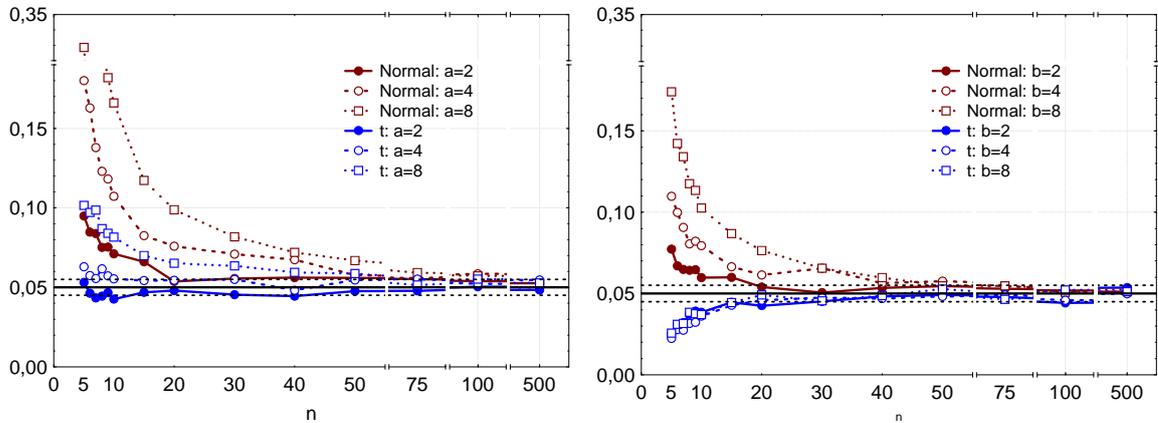


ABBILDUNG 6.1: Simulation: Niveau; CRF- a , $a = \{2, 4, 8\}$, $C = P_a$ (links) und CRF- ab , $a = 2, b = \{2, 4, 8\}$, $C = \frac{1}{a} \mathbf{1}'_a \otimes P_b$ (rechts).

Aus den Simulationskurven ist zu erkennen, dass im CRF- a Design, mit wachsender Anzahl der Stufen der Test liberaler wird, sowohl bei der Normal- als auch der t-Approximation. Wobei besonders die Normal-Approximation sehr liberal ist und erst ab $n > 50$ die Simulationskurven im Zufallsstreifen liegen. Im MODELL CRF- $2b$ wird die Liberalität der Normal-Approximation geringer, was mit den größeren Stichprobenumfängen begründet werden kann, die durch den zusätzlichen Faktor dazukommen. Im zweifaktoriellen Modell scheint die t-Approximation stabil zu sein und behält ihre konservative Eigenschaft, außerdem liegt das simulierte Niveau ab circa $n > 15$ im Zufallsstreifen.

Im nächsten Schritt wird kurz untersucht, wie sich die Liberalität der Normal-Approximation und die Konservativität der t-Approximation verhalten. Hierzu wurden im MODELL 4.2 für eine unterschiedliche Anzahl von Stufen, $a, b \in \{2, 3\}$, simuliert. Dazu sei auf die ABBILDUNG 6.2 verwiesen. In Bezug darauf kann gesagt werden, dass die Normal-Approximation besser das Niveau einhält, je weniger Stufen der Faktor hat auf dessen Einfluss hin getestet wird und je mehr Stufen bzw. je mehr Stichprobenumfänge durch die weiteren Faktoren hinzukommen. Dies ist verständlich, da die Normalverteilung der Rangstatistik asymptotisch gilt. Die konservative Eigenschaft der t-Approximation wird ebenfalls verstärkt, wenn sich die Anzahl der zu testenden Stufen erhöht.

Kontrastmatrizen zum sequentiellen Testproblem

In den mehrfaktoriellen Modellen wurde das sequentielle Testproblem angesprochen und Lösungen dazu vorgeschlagen (Abschnitt 4.3.4). Als erstes wird in diesem Abschnitt ge-

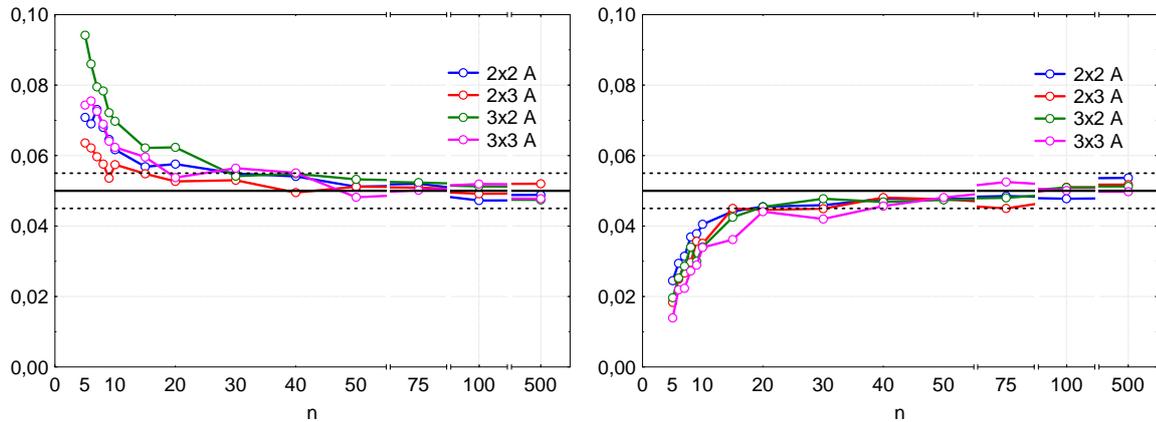


ABBILDUNG 6.2: Simulation: Niveau; Haupteffekt A in verschiedenen Designs, Normal-Approximation (links), t-Approximation (rechts).

zeigt, dass die sequentielle Testprozedur nach ABBILDUNG 4.2 das Niveau nicht einhält. Dazu wurden im CRF- ab Design die Faktorstufen $a_1 = b_1 = 2; a_2 = b_2 = 3$ und $a_3 = 4, b_3 = 3$ ausgewählt und zunächst mittels $C_{AB} = P_a \otimes P_b$ auf Wechselwirkung getestet. Fiel dieser Test negativ aus, wurde mit $C_A = P_a \otimes \frac{1}{b} \mathbf{1}_b$ auf Haupteffekt A getestet. Wobei hingegen bei getesteter Wechselwirkung die aufgetrennten Datensätze $X_a^{(j)}$ (siehe Abschnitt 4.3.4) jeweils mit P_a untersucht wurden, mit anschließender p-Wert Adjustierung nach Bonferroni. So ist der Fehler 1. Art, wenn die Tests auf Haupteffekt A mit C_A bzw. P_A ablehnen, obwohl die Hypothesen wahr sind. Zum besseren Verständnis wurde diese Vorgehensweise in ABBILDUNG 4.2 graphisch anschaulich dargestellt. Die Simulationskurven sind in ABBILDUNG 6.3 zu sehen.

Das Niveau der sequentiellen Testprozedur liegt bei circa 7% und hält damit das Niveau von 5% nicht ein. Eine Ursache dafür ist die deutliche Überschreitung des Niveaus, wenn auf Haupteffekt getestet wird und vorher nach signifikant getesteter Wechselwirkung fälschlicherweise die Daten aufgetrennt wurden. Zum Nachweis dieser Behauptung wurde ebenfalls für die CRF-22, CRF-33 und CRF-43 Designs auf Haupteffekt A getestet. Allerdings wurden solange neue Zufallswerte erzeugt, bis eine Wechselwirkung getestet wurde. Danach erfolgte, getrennt nach den Stufen des Faktors B, der Test auf Haupteffekt A. Damit das multiple Niveau eingehalten wird, wurden die p-Werte mit der Methode von Bonferroni adjustiert. Aus ABBILDUNG 6.4 ist zu erkennen, dass unter den genannten Gegebenheiten das Niveau massiv überschritten wird.

Eine vorgeschlagene Lösung war, die Matrizen für die Wechselwirkung und Haupteffekt A in einer Matrix (4.16)

$$C = (C'_{AB}; C'_A)' \quad (6.1)$$

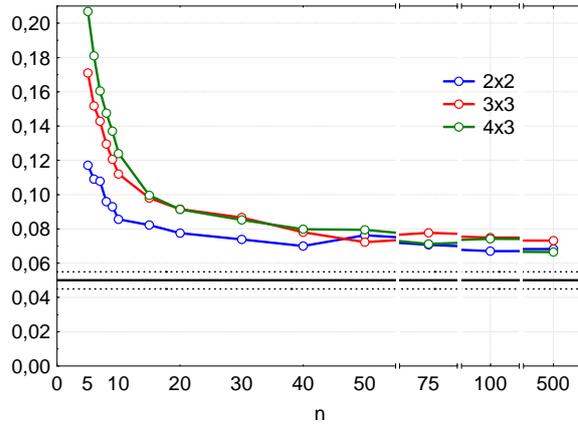


ABBILDUNG 6.3: Simulation: Niveau; Sequentieller Test (mit multiplen Kontrasten) auf Haupteffekt A.

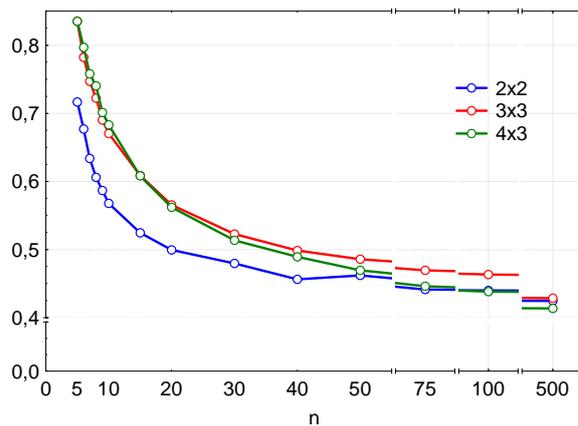


ABBILDUNG 6.4: Simulation: Niveau; nach getesteter Wechselwirkung, Test auf Haupteffekt A - getrennt nach Stufen von Faktor B, mit Bonferroni-Adjustierung.

zu vereinigen und die Tests auf Wechselwirkung und Haupteffekt A mit der Matrix (6.1) durchzuführen, aber die Auftrennung der Daten weiter beizubehalten, wenn vorher eine Wechselwirkung getestet wurde. Dies erhielt die Bezeichnung *Variante 1*. Soll auch die Auftrennung der Daten durch Kontraste beschrieben werden, so hat die Kontrastmatrix die Gestalt

$$C_{Seq,A} = (C'_{AB} : C'_A : C^{(B)'}_A)', \quad (6.2)$$

wobei $C^{(B)}_A = (P_a \otimes e^{(1)}_b : \dots : P_a \otimes e^{(b)}_b)'$. Mit der Matrix $C^{(B)}_A$ wird in jeder Stufe von

Faktor B auf Haupteffekt A getestet. Diese Kontraste werden untersucht, wenn eine Wechselwirkung AB getestet wurde, da dann unterschiedliche Verläufe zu erwarten sind. Der Test unter Verwendung der Kontrastmatrix (6.2) ist *Variante 2* und erhielt in Abschnitt 4.3.4 die Bezeichnung *Test unter Verwendung von $C_{Seq,A}$* . Zum Vergleich wurde für beide Varianten, wie in der Simulation zum sequentiellen Test (ABBILDUNG 6.3), in den zweifaktoriellen Designs CRF-22, CRF-33 und CRF-43, jeweils auf Haupteffekt A getestet. Die Simulationskurven dazu sind in ABBILDUNG 6.5 zu finden.

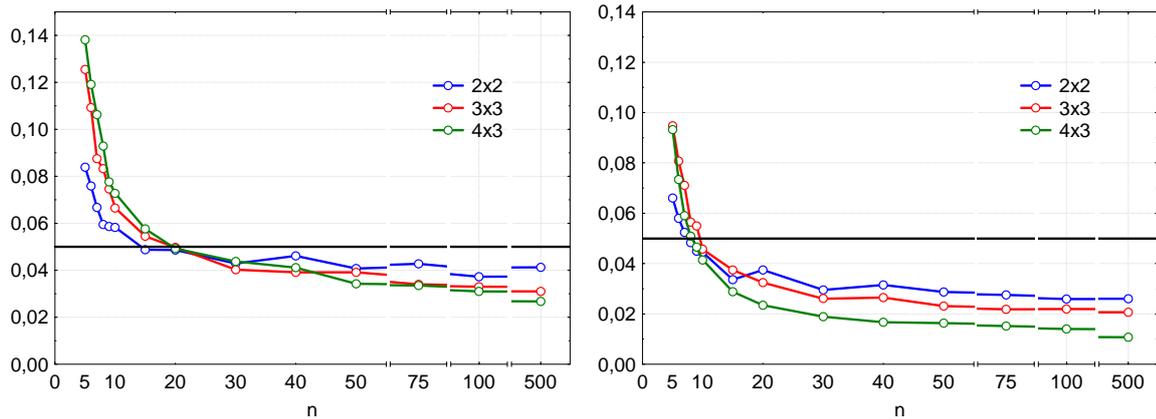


ABBILDUNG 6.5: Simulation: Niveau; links: sequentieller Test auf Haupteffekt A mit Variante 1 (weiterhin Auftrennung der Daten; Test auf Wechselwirkung und einzelnen Haupteffekt A mit Matrix (6.1)) und rechts: Test unter Verwendung von $C_{Seq,A}$ (6.2) - Variante 2 (keine Auftrennung der Daten)

Beide Varianten überschreiten das Niveau für $n > 15$ nicht. Da in *Variante 2* (rechts) die Kontrastmatrix (6.2) mehr Kontraste enthält, ist diese konservativer. Soll das Auftrennen der Daten also mit Kontrasten ausgedrückt werden, muss man akzeptieren, dass der Test konservativer wird.

Neben dem Test auf einen Haupteffekt, kann auch gleichzeitig auf mehrere Haupteffekte getestet werden. So enthält die Matrix (4.19)

$$C_{Seq} = (C'_{AB} : C'_A : C'_B : C^{(B)'}_A : C^{(A)'}_B)'$$

alle Kontraste, um nicht-sequentiell und simultan auf Haupteffekt A und Haupteffekt B zu testen. So wurden in dieser Matrix die Kontraste für die Wechselwirkung AB und für die Haupteffekte A und B vereint. Hinzu kommen die Kontraste auf Haupteffekt A und B, getrennt nach den Stufen des jeweils anderen Faktors, für den Fall einer vorhandenen Wechselwirkung.

So wird als nächstes untersucht wie diese Tests (mit C_{Seq}) das Niveau einhalten und wie konservativ diese sind. Dabei soll allgemein auf Haupteffekt getestet werden, d.h. der Test lehnt ab, wenn mindestens ein Haupteffekt als signifikant getestet wird.

Im Design mit drei Faktoren erweitert sich die Kontrastmatrix entsprechend und ist in Abschnitt 7.3 zu finden (7.2). Analog wird abgelehnt, wenn mindestens einer der drei Haupteffekte als signifikant getestet wird. Die ABBILDUNG 6.6 zeigt die Niveausimulationen für die zweifaktoriellen Modelle CRF-2b mit $b = \{2, 3, 5\}$ und das dreifaktorielle Modell CRF-222.

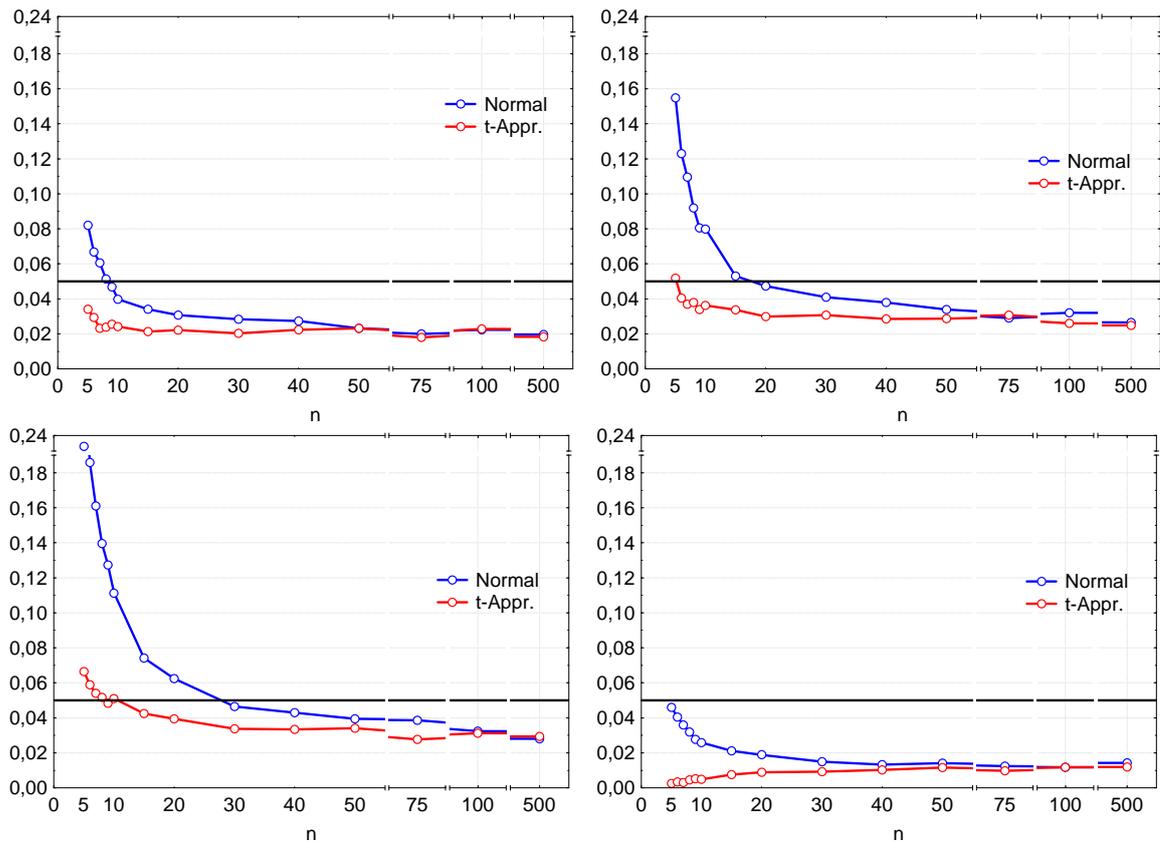


ABBILDUNG 6.6: Simulation: Niveau; Test unter Verwendung von C_{Seq} (4.19) auf beliebigen Haupteffekt (Ablehnung, falls mindestens ein Haupteffekt signifikant getestet wird), CRF-22 (links oben), CRF-23 (rechts oben), CRF-25 (links unten) und CRF-222 (rechts unten).

Die Simulationen zeigen, dass für genügend große Stichprobenumfänge das Niveau eingehalten wird. Wie auch aus ABBILDUNG 6.5 zu erschließen ist, sind die Tests unter Verwendung von C_{Seq} konservativ. Aus den Simulationen ist außerdem zu erkennen, dass mit

wachsender Zahl von Faktorstufen die Tests etwas liberaler werden. So scheint das „wahre“ Niveau im Modell CRF-22 ungefähr bei 0.02 zu liegen, während dies im Modell CRF-25 bei schätzungsweise 0.03 liegt. Wird allein das Niveau für einen Faktor betrachtet, also ob beispielsweise Haupteffekt A vorliegt, so wird der Test noch konservativer, da nur die Kontraste für den Faktor A betrachtet werden. Daher sollten die Matrizen nur aus den Kontrasten bestehen, welche für den Test tatsächlich genutzt werden.

Für die Auswertung der Beispiele kann aus den Simulationen festgehalten werden, dass für die Nierengewichte (CRF-25 Design) die t-Approximation genutzt wird, da die Normal-Approximation für die Stichprobenumfänge 7 bis 11 sehr liberal ist. Für das dreifaktorielle Beispiel der Leukozyten (CRF-222 Design) ist die t-Approximation hingegen zu konservativ ist. Daher wird dort auf die Normal-Approximation zurückgegriffen.

Niveausimulationen

In diesem Abschnitt wird das Niveau der Normal- und der t-Approximation, der Fisher-Transformation (Abschnitt 5.2.7) und der ATS (Abschnitt 5.3) untersucht. Als erstes für unterschiedliche Verteilungsfunktionen und als zweites für ungleiche Varianzen und Stichprobenumfänge. Die sequentielle Testprozedur wird hier nicht betrachtet.

Bei den unterschiedlichen Verteilungsfunktionen wurde im zweifaktoriellen CRF-23 Design, mit gleichen Stichprobenumfängen und gleichen Varianzen, simuliert und auf Haupteffekt B mit $C_B = \frac{1}{a} \mathbf{1}'_a \otimes P_b$ getestet. Dazu wurden den Zufallsvariablen zufällige Werte aus der Standardnormalverteilung, Diskreten Gleichverteilung (1 bis 10), Exponentialverteilung ($\lambda = 2$) und der Log-Normalverteilung zugewiesen. Die Simulationskurven dazu sind in ABBILDUNG 6.7 zu finden.

Da die Verläufe über die Verteilungen ähnlich sind, scheint die Art der Verteilung keinen auffälligen Einfluss zu besitzen. Betrachtet man die unterschiedlichen Approximationen und Statistiken, so konvergiert die ATS als erstes in den Zufallsstreifen, während die Normal- und t-Approximation und die Fisher-Transformation ab $n = 30$ adäquate Resultate liefern. Im zweiten Teil wird der Fall der Heterogenität und Unbalanciertheit bei normalverteilten Zufallsvariablen betrachtet. Dazu wurde im CRF-42 Design auf Haupteffekt A mit $C_A = P_a \otimes \frac{1}{b} \mathbf{1}'_b$ getestet. Unterschieden wird dabei zwischen steigenden Varianzen und fallenden Varianzen, bei gleichbleibend steigenden Stichprobenumfängen. Die Ergebnisse sind in TABELLE 6.1 aufgelistet.

Zu erkennen ist, dass bei fallenden Varianzen die Tests bei kleineren Stichproben liberaler sind als bei steigenden Varianzen. Die ATS scheint dabei im späteren Verlauf sogar konservativ zu werden.

Powersimulationen

Wie in der Niveausimulation wurde auch hier für vier verschiedene Verteilungen die Power simuliert (Normal-, diskrete Gleich-, Exponential- und Log-Normalverteilung), mit einem

6 Simulationen

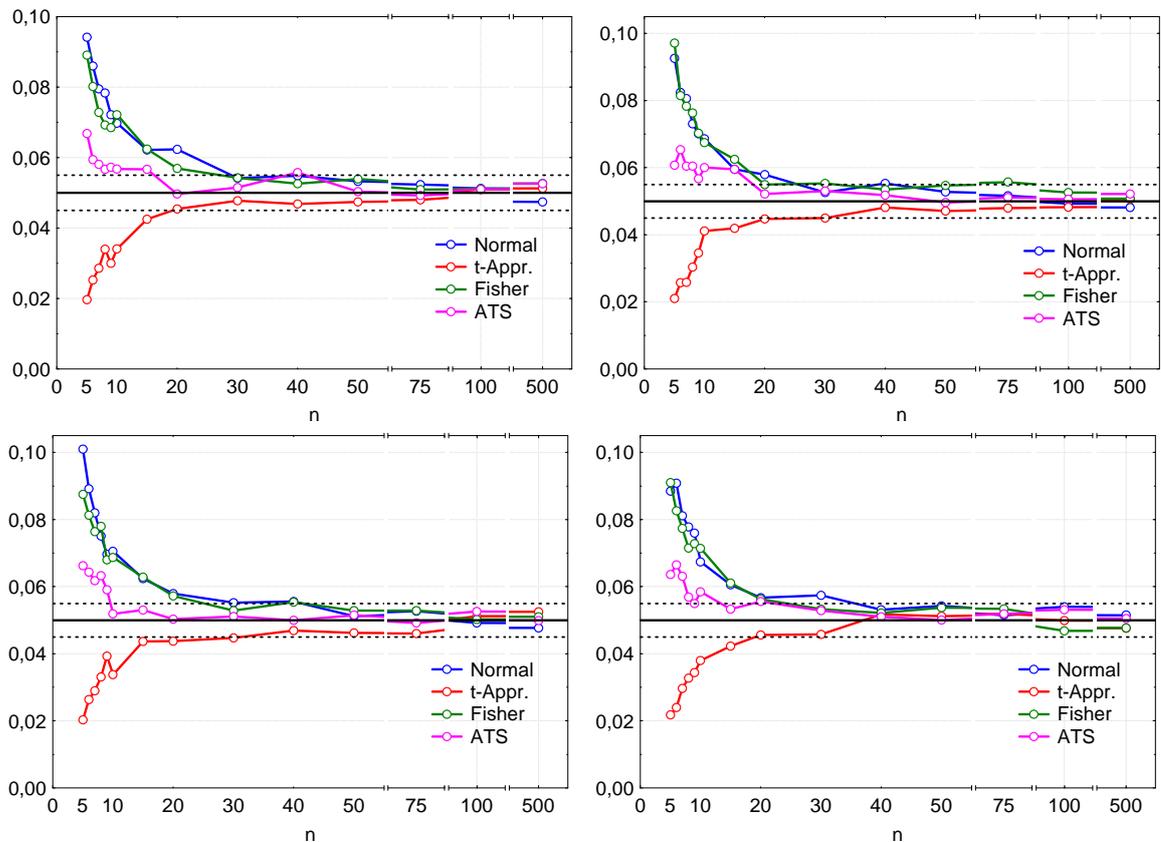


ABBILDUNG 6.7: Simulation: Niveau; Haupteffekt B in CRF-23, Normalverteilung (links oben), Diskrete Gleichverteilung (rechts oben), Exponentialverteilung (links unten) und Log-Normalverteilung (rechts unten).

TABELLE 6.1: Simulation: Niveau; Haupteffekt A in CRF-42; Ungleiche Varianzen und Stichprobenumfänge.

Stichprobenumfänge	Var: 1,2,3,4,6,8,11,15			Var: 15,11,8,6,4,3,2,1		
	Normal	t-Appr.	ATS	Normal	t-Appr.	ATS
3,4,5,6,7,8,9,10	0.0948	0.0188	0.0544	0.1234	0.0254	0.0827
8,10,12,14,16,18,20,22	0.0641	0.0387	0.0516	0.0735	0.0237	0.0541
15,18,21,24,27,30,33,36	0.0588	0.0413	0.0478	0.0634	0.0336	0.0452
20,25,30,35,40,45,50,55	0.0530	0.0480	0.0494	0.0584	0.0381	0.0476
30,40,50,60,70,80,90,100	0.0535	0.0466	0.0496	0.0547	0.0343	0.0420
45,60,75,90,105,120,135,150	0.0491	0.0492	0.0490	0.0572	0.0448	0.0428

Stichprobenumfang von $n = 30$ in jeder Stufe. Getestet wurde wiederum im CRF-23 Design auf Haupteffekt B mit $C_B = \frac{1}{a} \mathbf{1}'_a \otimes P_b$. Die Verläufe sind in ABBILDUNG 6.8 zu finden.

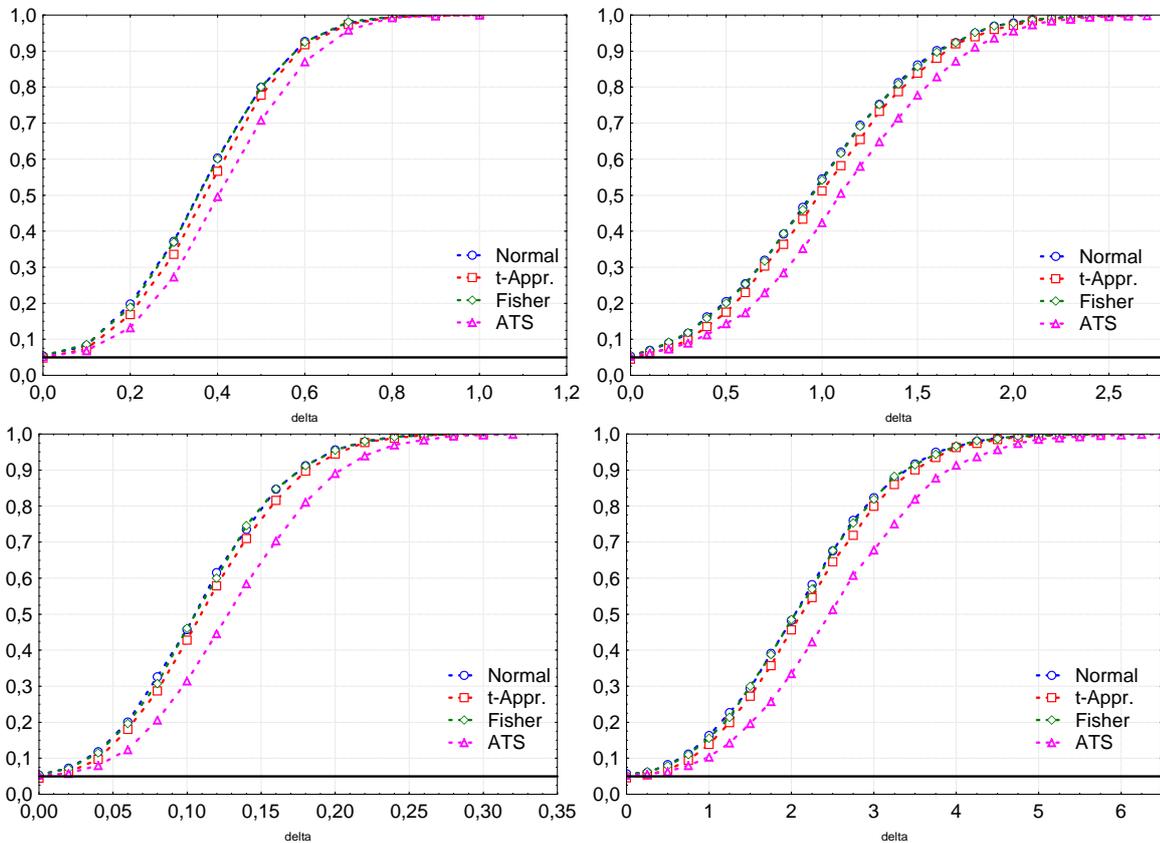


ABBILDUNG 6.8: Simulation: Power; Haupteffekt B in CRF-23, Normalverteilung (links oben), Diskrete Gleichverteilung (rechts oben), Exponentialverteilung (links unten) und Log-Normalverteilung (rechts unten).

Ähnliche Powerkurven ergeben sich auch bei anderen Konstellationen wie unterschiedlichen Varianzen oder Stichprobenumfängen, zum Beispiel bei normalverteilten Zufallsvariablen. Die Power der Normal-Approximation und Fisher-Transformation sind nahezu identisch, da die Fisher-Transformation nur die Grenzen der simultanen Konfidenzintervalle verändert. Die t-Approximation fällt um bis zu 5% Punkte ab, da diese allgemein konservativer ist und die globale Anova-Typ-Statistik um bis zu 16% Punkte bei der Exponentialverteilung.

Aus der Simulationsstudie ergeben sich folgende Erkenntnisse: Die ATS hält zwar generell das Niveau am besten ein, allerdings hat diese eine relativ schlechte Power (zudem

kann die ATS nur global testen). Als zweites kann geschlussfolgert werden, dass die Fisher-Transformation keinen Einfluss auf die Eigenschaften bezüglich des Niveaus und der Power hat. Weiterhin ist bei kleinen Stichprobenumfängen die t- der Normal-Approximation vorzuziehen, da ansonsten das Niveau erheblich überschritten wird. Sind die Stichprobenumfänge hingegen groß genug, so dass die Normal-Approximation das Niveau einhält, kann diese genutzt werden, da deren Power etwas höher ist.

Im Bezug auf die sequentielle Testprozedur konnte gezeigt werden, dass diese das vorgegebene Niveau nicht einhält. Der Test unter Verwendung von C_{Seq} hingegen hält das Niveau ein, wird allerdings konservativ.

7 Beispielauswertung

Mit der hergeleiteten Theorie können die in Abschnitt 2 vorgestellten Beispiele mit simultanen Konfidenzintervallen ausgewertet werden.

7.1 Anzahl der Corpora Lutea

In dieser Studie wurde die Wirkung einer Substanz auf die Fertilität bei weiblichen Ratten untersucht. Die Substanz wurde in vier unterschiedlichen Dosisstufen verabreicht, während in der Kontrollgruppe ein Placebo eingesetzt wurde. Die Stichprobenumfänge der einzelnen Stufen sind zwischen 16 und 20. Die Boxplots zu den Daten sind in ABBILDUNG 7.1 dargestellt.

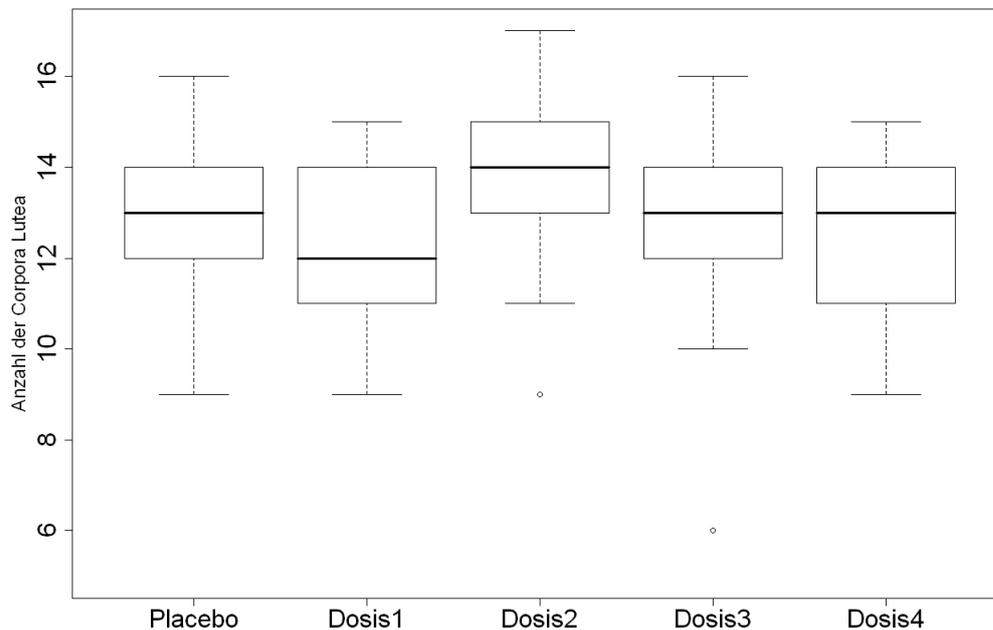


ABBILDUNG 7.1: Boxplots der Anzahl der Corpora Lutea.

Bei solchen Versuchen bietet sich der Dunnett-Kontrast (siehe (4.9)) an, so dass die Gruppen der vier Dosisstufen gegen die Placebogruppe verglichen werden. Da im CRF-51 De-

sign und für die in diesem Beispiel gegebenen Stichprobenumfänge die t-Approximation besser geeignet ist, sollten die Daten mit dieser ausgewertet werden. Zum Vergleich werden dennoch die Ergebnisse von der Normal-Approximation mit angegeben. Die 95%-Konfidenzintervalle, unter Anwendung der t- und Normal-Approximation, und die Schätzer $c'_m \hat{q}$ der Kontraste $c'_m q$ sind in TABELLE 7.1 zu finden.

TABELLE 7.1: Simultane 95%-Konfidenzintervalle zu Corpora Lutea: Dunnett Effekt.

Vergleich	Schätzer $c'_m \hat{q}$	Normal		t-Approx.	
		SCI	p-Wert	SCI	p-Wert
Placebo vs. Dosis 1	0.060	[-0.162, 0.282]	0.923	[-0.178, 0.298]	0.920
Placebo vs. Dosis 2	-0.165	[-0.365, 0.036]	0.146	[-0.380, 0.051]	0.175
Placebo vs. Dosis 3	-0.046	[-0.274, 0.182]	0.972	[-0.291, 0.199]	0.971
Placebo vs. Dosis 4	-0.040	[-0.260, 0.179]	0.980	[-0.276, 0.195]	0.979

Ein signifikanter Unterschied zwischen der Placebogruppe und einer der Substanzgruppen liegt nicht vor. Die Grenzen der Konfidenzintervalle sind bereichserhaltend, so dass auf eine zusätzliche Angabe der Fisher-Konfidenzintervalle in TABELLE 7.1 verzichtet wird. Zur Veranschaulichung sind die simultanen Konfidenzintervalle unter Verwendung der t-Approximation in der ABBILDUNG 7.2 graphisch dargestellt.

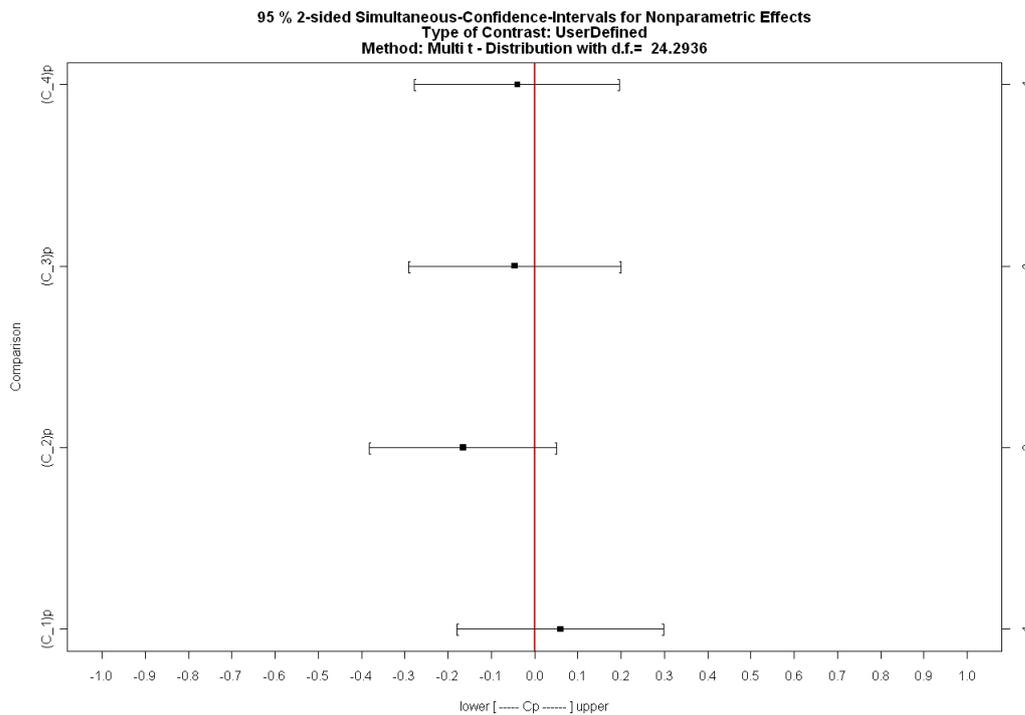


ABBILDUNG 7.2: Simultane 95%-Konfidenzintervalle:
c_1 Placebo gegen Dosisstufe 1 , . . . , c_4 Placebo gegen Dosisstufe 4.

7.2 Nierengewichte

In dieser Studie wurden die relativen Nierengewichte von männlichen und weiblichen Wistar Ratten gemessen, um die Wirkung einer Substanz zu untersuchen. Faktor A sei das Geschlecht mit zwei Stufen und Faktor B die Substanz, wobei entweder ein Placebo oder eine Substanz in einer von vier Dosisstufen verabreicht wurde. So wird das zweifaktorielle MODELL CRF-25 betrachtet. An diesem ersten (mehrfaktoriellen) Beispiel soll das sequentielle Testproblem aus Abschnitt 4.3.4 und die dort vorgestellte Herangehensweise veranschaulicht werden.

Da die Versuchstiere dem Geschlecht nach getrennt protokolliert wurden ist die erste Frage, ob eine Wechselwirkung vorliegt. So könnten Ratten verschiedenen Geschlechts unterschiedlich auf einzelne Dosisstufen reagieren. Die zweite Frage beschäftigt sich damit, ob eine Signifikanz zwischen den Dosisstufen existiert. Dabei ist ein Vergleich zwischen der Placebo-Gruppe mit denen der Dosisstufen-Gruppen von besonderem Interesse. Für letzteres bietet sich daher der Dunnett-Effekt (4.9) zur Auswertung an. Ein Test des Haupteffekts Geschlecht hingegen ist nebensächlich, da männliche Ratten größer als weibliche sind.

Zunächst seien in der ABBILDUNG 7.3 die Boxplots dargestellt, getrennt nach männlichen und weiblichen Tieren.

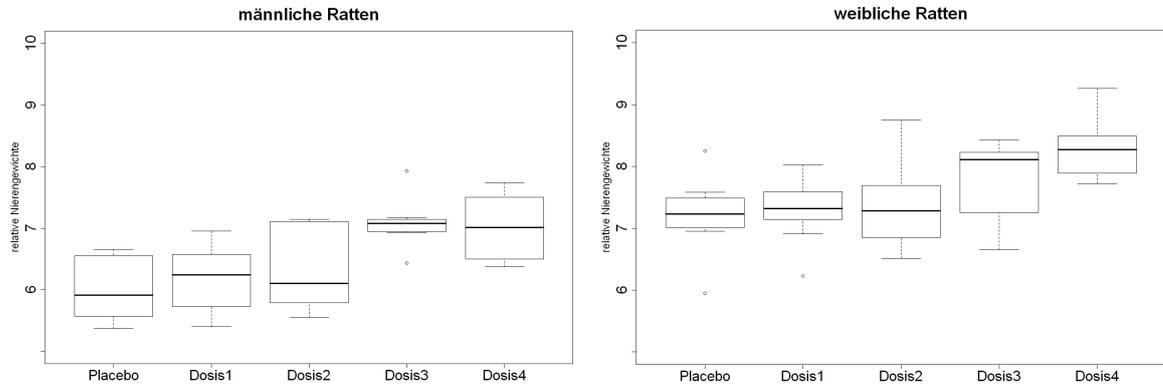


ABBILDUNG 7.3: Boxplots der relativen Nierengewichte.

Um die beiden unterschiedlichen Vorgehensweisen nochmals deutlich zu machen, wird zunächst sequentiell auf Wechselwirkung und Haupteffekt Substanz getestet. Danach wird mit dem hergeleiteten Testverfahren unter Verwendung von C_{Seq} simultan auf das Vorliegen einer Wechselwirkung und Unterschiede zur Kontrollgruppe getestet. Aus selbigem Grund steht in dieser Auswertung die Untersuchung zwischen den Dosisstufen im Mittelpunkt und nicht die der Wechselwirkung. Abschließend werden die Ergebnisse aus dem multiplen Kontrasttest gegen die der ANOVA verglichen.

Sequentieller Test

In der sequentiellen Herangehensweise wird als erstes die Wechselwirkung betrachtet. Die dazu bekannte Kontrastmatrix aus der Varianzanalyse enthält die Vergleiche gegen das Mittel $P_n, n \in \mathbb{N}^+$, so dass $C_{AB} = P_a \otimes P_b$. Da in diesem Fall aber die Dunnett-Kontraste C_{Dun} genutzt werden, hat die Kontrastmatrix zum Test auf Wechselwirkung die Gestalt

$$C_{AB} = C_{Dun}^2 \otimes C_{Dun}^5 = \frac{1}{2} \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (7.1)$$

Für den i-ten Kontrast ist $c_i \mathbf{q} = ((q_{11} - q_{1i}) - (q_{21} - q_{2i}))/2$, so dass mit den Teilhypothesen $c_i \mathbf{q} = 0$ getestet wird, ob die Differenzen der Effekte zwischen der Placebogruppe und der i-ten Dosisstufe bei den männlichen und weiblichen Ratten gleich sind. TABELLE 7.2

zeigt die Kontrastschätzer $c'_m \hat{q}$ der Wechselwirkung und die entsprechenden p-Werte und simultanen Konfidenzintervalle.

TABELLE 7.2: Simultane 95%-Konfidenzintervalle für die Wechselwirkungseffekte mit $C_{AB} = C_{Dun}^2 \otimes C_{Dun}^5$ aus (7.1) unter Verwendung der t-Approximation.

	Kontrast			
	c_1	c_2	c_3	c_4
$c'_m \hat{q}$	0.004	-0.035	-0.084	-0.003
p-Wert	>0.999	0.966	0.613	>0.999
SCI	[-0.191, 0.198]	[-0.248, 0.179]	[-0.295, 0.127]	[-0.156, 0.150]

Eine signifikante Wechselwirkung liegt für keinen Kontrast vor, so dass auf eine geschlechts-spezifische Auswertung verzichtet werden kann. Die Kontrastmatrix zum Testen des Haupteffektes Substanz ist gegeben durch

$$C_B = C_{Dun}^{(2,5)} = \frac{1}{2} \mathbf{1}'_2 \otimes C_{Dun}^5 = \frac{1}{2} \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

Die ersten bzw. letzten fünf Spalten beschreiben dabei jeweils einen Dunnett-Kontrast. So testet der i -te Kontrast $c_i q = ((q_{11} - q_{1i}) + (q_{21} - q_{2i}))/2$, ob die Effekte q zwischen der Placebo- und der i -ten Dosisstufengruppe über die Geschlechter gleich sind. Die Ergebnisse dazu stehen in TABELLE 7.3 und werden mit denjenigen aus dem Test unter Verwendung von C_{Seq} verglichen.

Test unter Verwendung von C_{Seq}

Die Konstruktion der Kontrastmatrix C_{Seq} zur Umgehung der sequentiellen Testprozedur wurde bereits in Abschnitt 4.3.4 erklärt. Dort wurde auch erwähnt, dass sich der Anwender vor dem Test überlegen sollte, welche Kontraste für die Auswertung wichtig sind, damit der Test nicht zu konservativ wird. Es muss aber beachtet werden, dass die Kontrastmatrix vor der Auswertung feststeht. Da bei nachträglicher Änderung durch Hinzu- oder Wegnahme von Kontrasten der Test verfälscht wird.

In diesem Beispiel kann berücksichtigt werden, dass ein biologisch bedingter Unterschied im Faktor A (Geschlecht) existiert. Dadurch können die Kontraste für den Haupteffekt Geschlecht, C_A und $C_A^{(B)}$, weggelassen werden. Dann erhält man die Kontrastmatrix

$$\mathbf{C}_{Seq,B} = \mathbf{C}_{Seq} = \left(\mathbf{C}'_{AB} : \mathbf{C}_B : \mathbf{C}_B^{(A)} \right)' \text{ mit}$$

$$\mathbf{C}_B^{(A)} = \mathbf{I}_2 \otimes \mathbf{C}_{Dun}^5 = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

Mit $\mathbf{I}_2 \otimes \mathbf{C}_{Dun}^5$ wird der Haupteffekt Substanz, getrennt nach Geschlechtern, getestet. Auch mit dieser Vorgehensweise liegt keine signifikante Wechselwirkung vor. So werden für den Test des Haupteffektes Substanz die Werte entnommen, die zur Matrix \mathbf{C}_B gehören. Die Kontraste aus $\mathbf{C}_B^{(A)} = \mathbf{I}_2 \otimes \mathbf{C}_{Dun}^5$ werden nicht weiter betrachtet. Die Ergebnisse für die Kontraste aus \mathbf{C}_B sind ebenfalls in TABELLE 7.3 dargestellt.

TABELLE 7.3: Simultane 95%-Konfidenzintervalle zu den relativen Nierengewichten mit Dunnett-Effekt und t-Approximation.

Vergleich	Schätzer $\mathbf{c}'_m \hat{\mathbf{q}}$	$\mathbf{C}_B = \mathbf{C}_{Dun}^{(2,5)}$		\mathbf{C}_{Seq}	
		SCI	p-Wert	SCI	p-Wert
Placebo vs. Dosis 1	-0.043	[-0.235, 0.149]	0.905	[-0.273, 0.187]	0.991
Placebo vs. Dosis 2	-0.078	[-0.288, 0.132]	0.663	[-0.330, 0.174]	0.894
Placebo vs. Dosis 3	-0.258	[-0.464, -0.052]	0.016	[-0.505, -0.011]	0.040
Placebo vs. Dosis 4	-0.330	[-0.505, -0.155]	0.001	[-0.539, -0.120]	0.003

Der Test unter Verwendung von \mathbf{C}_{Seq} ist konservativer als der sequentielle Test (\mathbf{C}_B). Aber auch in diesem Fall ist ein signifikanter Unterschied zwischen dem Placebo und den Dosisstufen drei und vier feststellbar. Die Konfidenzintervalle sind auch hier bereichserhaltend, so dass auf eine zusätzliche Angabe der Fisher-Konfidenzintervalle verzichtet wird.

ANOVA vs. multipler Kontrasttest

In diesem Abschnitt wird das Beispiel klassisch mit der ATS (Abschnitt 5.3) global ausgewertet und mit den Werten aus dem simultanen Testverfahren verglichen. Im Letzteren wurde mittels multiplen Kontrasttests der sequentielle Test durchgeführt. Der sequentielle Test wird hier als Vergleichstest verwendet, da in der Varianzanalyse die Datensätze ebenfalls aufgetrennt werden. Nach nicht signifikant getesteter Wechselwirkung wurde unter Verwendung der Dunnett-Kontrastmatrix $\mathbf{C}_B = \mathbf{C}_{Dun}^{(2,5)}$ der Test im Faktor Substanz durchgeführt

(siehe TABELLE 7.3). In der zweifaktoriellen Varianzanalyse wird mit $C_A = P_a \otimes \frac{1}{b} \mathbf{1}'_b$, $C_B = \frac{1}{a} \mathbf{1}'_a \otimes P_b$ und $C_{AB} = P_a \otimes P_b$ auf Haupteffekt Geschlecht (A) und Substanz (B) und Wechselwirkung getestet. Die Ergebnisse hierzu sind in TABELLE 7.4 zusammengefasst.

TABELLE 7.4: Ergebnisse zu den relativen Nierengewichten, Haupteffekte und Wechselwirkung unter Verwendung der ATS.

Faktor	ATS	\hat{f}	p-Wert
HE Geschlecht	96.17	1	< 0.001
HE Substanz	9.15	3.79	< 0.001
WW	0.61	3.58	0.723

Wie mit dem multiplen Kontrasttest wird auch in der Varianzanalyse die Wechselwirkung nicht signifikant getestet, ebenso ist der biologisch bedingte Unterschied im Faktor Geschlecht signifikant. Auch der Test auf Haupteffekt Substanz ist signifikant. Die gleichen Werte zum Haupteffekt Substanz ergeben sich unter Verwendung von $C_B = C_{Dun}^{(2,5)}$.

Der multiple Kontrasttest hat nun zum einen Vorteil, dass zu jedem einzelnen Kontrast entscheidungsfindende Werte mit angegeben werden. Wie aus TABELLE 7.3 zu entnehmen ist, kann direkt gefolgert werden, dass es zwischen Placebo und den Dosisstufen drei und vier einen signifikanten Unterschied gibt. In der Varianzanalyse müssen dafür extra Paarvergleiche durchgeführt werden. Die Resultate dazu sind in TABELLE 7.5 dargestellt.

TABELLE 7.5: Ergebnisse zu den relativen Nierengewichten, paarweise Vergleiche - Placebo vs. Dosisstufen, mit ATS.

Vergleich	ATS	\hat{f}	p-Wert
Placebo vs. Dosis 1	0.53	1	0.932
Placebo vs. Dosis 2	1.25	1	0.527
Placebo vs. Dosis 3	8.25	1	0.008
Placebo vs. Dosis 4	19.2	1	< 0.001

Bevor die p-Werte ausgewertet werden können müssen diese adjustiert werden, da das multiple Niveau andernfalls nicht eingehalten wird. Dafür wird die Bonferroni-Adjustierung mit dem Faktor vier verwendet, da die vier Vergleiche gegen das Placebo zusammen den Dunnett-Effekt beschreiben, für den das Niveau eingehalten werden soll. Dann ist einer der paarweisen Tests als signifikant zu bewerten, wenn der p-Wert kleiner als $\alpha/4 = 0.05/4 = 0.0125$ ist. Dies trifft sowohl für die Dosisstufe drei als auch für Dosisstufe vier zu. Diese Ergebnisse decken sich mit denen der multiplen Kontrastprozedur, als auch mit denen des sequentiellen Tests. Neben den p-Werten werden auch die simultanen Konfidenzintervalle

zur Beurteilung herangezogen (siehe Abschnitt 5.2.5 und 5.2.6). Analog zu den p-Werten, lassen sich im multiplen Kontrasttest die Konfidenzintervalle gleichzeitig für jeden Kontrast angeben (siehe TABELLE 7.3). In der ANOVA werden wiederum für jeden Kontrast einzelne Konfidenzintervalle hergeleitet. So existieren verschiedene Verfahren um Konfidenzintervalle zu konstruieren, welche das multiple Niveau einhalten. Beispielsweise gibt es ähnlich zur Bonferroni-Adjustierung der p-Werte die Konstruktionsmethode von Bonferroni, in der gewöhnlicherweise jedes Konfidenzintervall zum Niveau von, angewandt auf dieses Beispiel, $\alpha/4$ konstruiert wird. Auf die Angabe dieser Konfidenzintervalle soll hier verzichtet werden.

7.3 Leukozyten-Migration ins Peritoneum

Ziel dieser Studie war es unter anderem die Wirkung einer Substanz unter einer Stresssituation zu untersuchen. In der Kontrollgruppe bzgl. der Substanzgruppen wurde den Mäusen anstelle des Verums ein Placebo verabreicht. Die Stresssituation wurde durch die Einnahme von Mangelfutter erreicht, wobei deren Kontrollgruppe Normalfutter zu sich nahm. Ein weiterer Faktor ist die Stimulation der Migration der Leukozyten durch Glycogen zum einen und zum anderen mit zusätzlich inaktivierten Staphylokokken. Faktor A sei die Art des Futters, Faktor B die Stimulation und Faktor C die Substanz. In ABBILDUNG 7.4 sind die Box-Plots dargestellt, getrennt nach den zwei Stufen des Futters, Normalfutter ($i=2$) und Mangelfutter ($i=1$). Dabei steht G für die Stimulation mit Glycogen ($j=1$) und G+S für Glycogen und Staphylokokken ($j=2$).

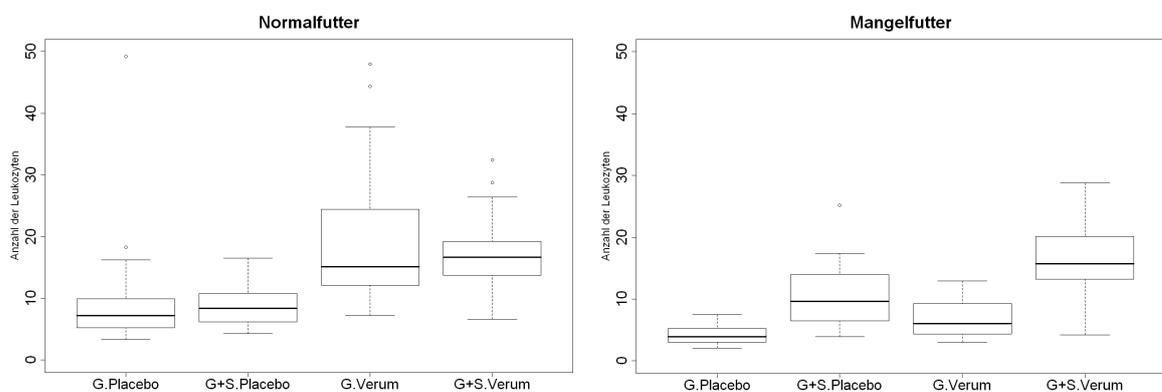


ABBILDUNG 7.4: Boxplots der Anzahl der Leukozyten.

Jede der acht Stufenkombinationen hat 19 bzw. 20 Stichprobenumfänge, daher werden die Daten mit der Normal-Approximation (Abschnitt 5.2.5) untersucht.

Dieser dreifaktorielle Versuch soll ebenfalls mit dem multiplen Kontrasttest unter Verwen-

ung von C_{Seq} ausgewertet werden. Da jeder Faktor jeweils nur zwei Stufen hat (CRF-222), bleibt die Anzahl der Kontraste überschaubar. Auf eine Auswertung mittels der sequentiellen Vorgehensweise und der ANOVA wird an dieser Stelle verzichtet.

Für den Vergleich von zwei Stichproben bietet sich nur der Kontrast $c = (-1, 1)'$ an. Dann ist für $d = 2$, $r(P_d) = 1$, so dass für alle Wechselwirkungen und Haupteffekte die Kontrastmatrix jeweils nur aus einem Kontrast besteht. Zum Beispiel gilt für die Kontrastmatrix der Dreifach-Wechselwirkung $C_{ABC} = P_a \otimes P_a \otimes P_a \in \mathbb{R}^{8 \times 8}$, dass deren Kontrastzeilen, bis auf die Multiplikation mit -1, alle identisch sind. Die insgesamt sieben Kontraste für die Wechselwirkungen und die Haupteffekte sind in der Matrix

$$C_{WW,HE} = \frac{1}{4} \begin{pmatrix} 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{pmatrix}$$

enthalten. Die oberste Zeile beschreibt den Test auf Dreifach-Wechselwirkung, gefolgt von den Wechselwirkungen AB, AC und BC und zum Schluss die Haupteffekte A, B und C. Treten Wechselwirkungen auf, so müssen die Haupteffekte dementsprechend getrennt untersucht werden. Beispielhaft für die Zweifach-Wechselwirkung AB beschreibt die Matrix

$$C_{A,B} = \left((C_A^{(B)'} : C_B^{(A)'}) \right)' = \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & -1 & -1 \\ 1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & -1 & 1 \end{pmatrix}$$

diesen Schritt. Die ersten beiden Zeilen testen auf Haupteffekt A (Futter), getrennt nach den Stufen der Stimulierung ($j=1, j=2$) und die beiden folgenden auf Haupteffekt B (Stimulation), getrennt nach den Stufen des Futters ($i=1, i=2$). Ähnliche Matrizen lassen sich für die Fälle konstruieren, dass die Wechselwirkungen AC ($C_{A,C}$) und BC ($C_{B,C}$) auftreten. Tritt sogar eine Wechselwirkung ABC auf, so wird nach allen Stufenkombinationen komplett getrennt untersucht. Dies drückt die Matrix

$$C_{A,B,C} = \left(((1, -1) \otimes I_4)' : (I_2 \otimes (1, -1) \otimes I_2)' : (I_4 \otimes (1, -1))' \right)' \text{ aus.}$$

Die Matrix

$$C_{Seq} = \left(C'_{WW,HE} : C'_{A,B} : C'_{A,C} : C'_{B,C} : C'_{A,B,C} \right)' \quad (7.2)$$

enthält nun alle nötigen Kontraste, um dem sequentiellen Testproblem zu entgehen. Die Ergebnisse sind in der TABELLE 7.6 aufgelistet, wobei nur die Kontraste angegeben sind die betrachtet werden müssen.

TABELLE 7.6: Ausgesuchte simultane 95%-Konfidenzintervalle unter Normal-Approximation.

Kontrast	Schätzer $c'_m \hat{q}$	SCI	p-Wert
WW ABC	0.017	[-0.077, 0.111]	0.999
WW AB	-0.181	[-0.272, -0.091]	<0.001
WW AC	0.063	[-0.029, 0.155]	0.420
WW BC	0.012	[-0.082, 0.106]	>0.999
HE Futter (G)	-0.351	[-0.473, -0.230]	<0.001
HE Futter (G+S)	0.012	[-0.125, 0.148]	>0.999
HE Stimulation (Mangelfutter)	-0.384	[-0.503, -0.266]	<0.001
HE Stimulation (Normalfutter)	-0.022	[-0.163, 0.120]	>0.999
HE Substanz	-0.270	[-0.363, -0.177]	<0.001

Eine Dreifach-Wechselwirkung besteht nicht, dafür eine Zweifach-Wechselwirkung zwischen den Faktoren Futter und Stimulation. Daher müssen die Haupteffekte für Futter und Stimulation jeweils getrennt nach den Stufen des anderen Faktors untersucht werden. So ist die Anzahl der Leukozyten in den Gruppen Mangelfutter und Normalfutter unterschiedlich, wenn diese nur mit Glycogen stimuliert wurden, wobei hingegen unter Stimulation von Glycogen und Staphylokokken kein signifikanter Unterschied getestet wurde. Für die Stimulation gilt, dass unter Mangelfutter die Stimulation eine Rolle spielt, unter Normalfutter hingegen nicht. Der Faktor Substanz steht in keiner Wechselwirkung zu den anderen Faktoren. So wird im Test auf Haupteffekt Substanz nicht zwischen den Stufen der zwei weiteren Faktoren unterschieden und der Test ergibt einen stark signifikanten Unterschied zwischen der Placebo- und Verumgruppe.

8 Zusammenfassung und Ausblick

In dieser Arbeit wurden simultane Konfidenzintervalle und multiple Kontrasttests für nicht-parametrische relative Effekte in faktoriellen Designs entwickelt. Im Vergleich zur Varianzanalyse liefern diese Informationen für die einzelnen Faktorstufen mit Hilfe von adjustierten p -Werten und simultanen Konfidenzintervallen. Des Weiteren wurde das sequentielle Vorgehen in faktoriellen Designs mit dem Vortest auf Wechselwirkung und folgendem Test auf Haupteffekt, mit möglicher Auftrennung der Daten, kritisch mit Hilfe von Simulationen untersucht.

Das sequentielle Verfahren kann durch spezielle Kontrastmatrizen, die die Kontrastmatrizen für Haupteffekte und Wechselwirkungen vereinen, mit den neu entwickelten Verfahren umgangen werden. So erhält man ohne die Auftrennung der Daten eine kompaktere Testtheorie.

Für kleine Stichprobenumfänge wurde außerdem eine Approximation hergeleitet, deren praktische Einsetzbarkeit durch Simulationen überprüft wurde.

In dieser Arbeit wurde an die Stichproben die Bedingung gestellt, dass diese unverbunden sind. So bietet sich eine Erweiterung auf verbundene Daten, wie faktoriell longitudinale Daten, an. Simulationen haben gezeigt, dass das neue Verfahren zur Umgehung der sequentiellen Testprozedur konservativ ist. Mit der Anwendung von Score-Funktionen könnte untersucht werden die Tests effizienter zu gestalten.

Durch die vielfältige Kombinierbarkeit der Kontrastmatrizen können noch Vorschriften oder Anleitungen zum richtigen Umgang ausführlicher ausgearbeitet werden.

A Definitionen, Notationen und Sätze

A.1 Matrizenrechnung

Matrizen und Vektoren werden stets **fett** geschrieben. Eine Matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ hat dabei die Gestalt:

$$\mathbf{A} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}.$$

Die zu \mathbf{A} transponierte Matrix wird mit \mathbf{A}' bezeichnet.

DEFINITION A.1. (*Spezielle Matrizen*)

1. *Einheitsmatrix*

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{pmatrix}_{n \times n}$$

2. *Einser-Vektor*

$$\mathbf{1}_n = (1, \dots, 1)'_{1 \times n},$$

Nullvektor

$$\mathbf{0}_n = (0, \dots, 0)'_{1 \times n},$$

kanonischer i-te Einheitsvektor

$$\mathbf{e}_n^{(i)} = (0, \dots, 0, \underset{\uparrow}{1}, 0, \dots, 0)'_{1 \times n}$$

3. *Einser-Matrix*

$$\mathbf{J}_n = \mathbf{1}_n \mathbf{1}_n' = \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix}_{n \times n}$$

4. zentrierende Matrix

$$\mathbf{P}_n = \mathbf{I}_n - \frac{1}{n} \mathbf{J}_n = \begin{pmatrix} 1 - \frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1 - \frac{1}{n} & & \vdots \\ \vdots & & \ddots & -\frac{1}{n} \\ -\frac{1}{n} & \cdots & -\frac{1}{n} & 1 - \frac{1}{n} \end{pmatrix}_{n \times n}$$

DEFINITION A.2. (Kronecker-Produkt)

Für beliebige Matrizen $\mathbf{A} \in \mathbb{R}^{m \times n}$ und $\mathbf{B} \in \mathbb{R}^{p \times q}$ heißt

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11} \mathbf{B} & \cdots & a_{1n} \mathbf{B} \\ \vdots & & \vdots \\ a_{m1} \mathbf{B} & \cdots & a_{mn} \mathbf{B} \end{pmatrix}_{mp \times nq}$$

Kronecker-Produkt von \mathbf{A} und \mathbf{B} .

DEFINITION A.3. (Jacobi-Matrix)

Die Funktion $\mathbf{f} : \mathbb{R}^d \rightarrow \mathbb{R}^k$, $(x_1, \dots, x_d)' = \mathbf{x} \mapsto \mathbf{f}(\mathbf{x})$ sei differenzierbar, mit $\mathbf{f} = (f_1, \dots, f_k)'$. Dann heißt die Matrix

$$\mathbf{J}_f := \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial x_1} & \cdots & \frac{\partial f_k}{\partial x_d} \end{pmatrix}_{k \times d}$$

Jacobi-Matrix von \mathbf{f} .

DEFINITION A.4. (Skalarprodukt) Für zwei Vektoren $\mathbf{v} = (v_1, \dots, v_n)$, $\mathbf{w} = (w_1, \dots, w_n) \in \mathbb{R}^n$ heißt

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}' \mathbf{w} = \sum_{i=1}^n v_i \cdot w_i$$

Skalarprodukt der Vektoren \mathbf{v} und \mathbf{w} .

DEFINITION A.5. (Spur) Für eine quadratische Matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ ist

$$Sp(\mathbf{S}) = \sum_{i=1}^n s_{ii}$$

die Spur der Matrix \mathbf{S} .

DEFINITION A.6. (Verallgemeinerte Inverse) Für eine Matrix \mathbf{A} ist heißt \mathbf{A}^- verallgemeinerte Matrix, wenn $\mathbf{A} \mathbf{A}^- \mathbf{A} = \mathbf{A}$.

A.2 Wahrscheinlichkeitstheorie

DEFINITION A.7. (Konvergenz in Wahrscheinlichkeit)

Eine Folge von Zufallsvariablen X_n konvergiert stochastisch gegen X , kurz: $X_n \xrightarrow{p} X$, wenn

$$P(|X_n - X| \geq \epsilon) \rightarrow 0 \quad \forall \epsilon > 0$$

gilt.

DEFINITION A.8. (Fast sichere Konvergenz)

Eine Folge von Zufallsvariablen X_n konvergiert fast sicher gegen X , kurz: $X_n \xrightarrow{f.s.} X$, wenn

$$P\left(\limsup_{n \rightarrow \infty} \{|X_n - X| > \epsilon\}\right) = 0 \quad \forall \epsilon > 0$$

gilt.

DEFINITION A.9. (Konvergenz im p -ten Mittel)

Eine Folge von Zufallsvariablen X_n konvergiert im p -ten Mittel gegen X , kurz: $X_n \xrightarrow{L_p} X$, falls für $p > 0$

$$E |X_n - X|^p \rightarrow 0$$

gilt.

DEFINITION A.10. (Konvergenz in Verteilung)

Eine Folge von Zufallsvariablen $X_n \sim F_n$ konvergiert in Verteilung gegen $X \sim F(x)$, kurz: $X_n \xrightarrow{L} X$, falls für alle Stetigkeitspunkte x

$$\lim_{n \rightarrow \infty} P(X_n \leq x) = F(x)$$

gilt.

DEFINITION A.11. (Asymptotische Äquivalenz)

Zwei Folgen von Zufallsvariablen X_n und Y_n heißen asymptotisch äquivalent, kurz: $X_n \doteq Y_n$, falls $X_n - Y_n \xrightarrow{p} 0$ für $n \rightarrow \infty$ gilt.

SATZ A.12. (Cramer)

Die Abbildung $g : \mathbb{R}^d \rightarrow \mathbb{R}^k$ sei auf einer Umgebung von $\mu \in \mathbb{R}^d$ stetig differenzierbar mit Jacobi-Matrix $J_g(\mu)$. Es sei \mathbf{X}_n eine Folge d -dimensionaler Zufallsvektoren mit $\sqrt{n}(\mathbf{X}_n - \mu) \xrightarrow{L} \mathbf{X}$. Dann gilt $\sqrt{n}[g(\mathbf{X}_n) - g(\mu)] \xrightarrow{L} J_g(\mu)\mathbf{X}$. Insbesondere gilt, falls $\sqrt{n}(\mathbf{X}_n - \mu) \xrightarrow{L} N(\mathbf{0}, \Sigma)$,

$$\sqrt{n}[g(\mathbf{X}_n) - g(\mu)] \xrightarrow{L} N(\mathbf{0}, J_g(\mu)\Sigma J_g'(\mu)).$$

Beweis. Siehe Ferguson (1996, Kapitel 5, Theorem 7.6). □

DEFINITION A.13. (Verteilungsfunktion)

Für eine Zufallsvariable X heißt

$$\begin{aligned} F^-(x) &= P(X < x) \quad \text{links-stetige,} \\ F^+(x) &= P(X \leq x) \quad \text{rechts-stetige,} \\ F(x) &= \frac{1}{2}[F^+(x) + F^-(x)] \quad \text{normalisierte} \end{aligned}$$

Version der Verteilungsfunktion X .

DEFINITION A.14. (Zählfunktion)

Die Funktion heißt

$$\begin{aligned} c^-(x) &= \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases} \quad \text{links-stetige,} \\ c^+(x) &= \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \quad \text{rechts-stetige,} \\ c(x) &= \frac{1}{2}[c^+(x) + c^-(x)] \quad \text{normalisierte} \end{aligned}$$

Version der Zählfunktion.

DEFINITION A.15. (Mittelränge von Konstanten)

Seien X_1, \dots, X_n beliebige reelle Zahlen, dann heißt

$$R_i = \frac{1}{2} + \sum_{j=1}^n c(X_i - X_j)$$

der Mittelrang von X_i unter allen Zahlen von X_1, \dots, X_n .

SATZ A.16. (Kontrolle der FWER)

Es sei $(\Omega, \mathbf{T}, \xi)$ eine kohärente simultane Testprozedur. Die Wahrscheinlichkeit, dass $(\Omega, \mathbf{T}, \xi)$ eine wahre Nullhypothese zum Niveau α ablehnt, ist α , sofern die Globalhypothese stimmt. Sie ist höchstens α , unabhängig von der Gültigkeit der Globalhypothese, wenn (Ω, \mathbf{T}) , eine geschlossene oder gemeinsame Testfamilie ist.

Beweis. Siehe Gabriel (1969, Theorem 2). □

SATZ A.17. (Verteilung quadratischer Formen)

Sei $\mathbf{Y} \sim N(\mathbf{0}, \mathbf{V})$ mit $\mathbf{V} \in \mathbb{R}^{q \times q}$, $r(\mathbf{V}) \leq q$ und \mathbf{S} eine symmetrische Matrix. Dann gilt

$$\mathbf{Y}'\mathbf{S}\mathbf{Y} \sim \sum_{i=1}^q \lambda_i Z_i,$$

mit $Z_i \sim \chi_1^2$, $i = 1, \dots, q$ und λ_i die Eigenwerte von $\mathbf{S}\mathbf{V}$.

Beweis. Siehe Mathai und Provost (1992, Kapitel 3.1, Representation 3.1a.1). □

Literaturverzeichnis

- [1] Akritas M.G., Arnold S.F., Brunner E., 1997. Nonparametric hypotheses and rank statistics for unbalanced factorial designs. *Journal of the American Statistical Association* **92**, 258-265.
- [2] Box G.E.P., 2008. Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems I. Effect of the Inequality of Variance in the One-Way Classification. *The Annals of Mathematical Statistics* **25**, 290-302.
- [3] Boysen L., 2002. Analyse von intra-individuellen Effekten bei longitudinalen Daten. *Diplomarbeit am Institut für Mathematische Stochastik der Universität Göttingen*.
- [4] Bretz F., Genz A., Hothorn L.A., 2001. On the Numerical Availability of Multiple Comparison Procedures. *Biometrical Journal* **43**, 645-656.
- [5] Brown B.M., Hettmansperger T.P., 2002. Kruskal-Wallis, multiple comparisons and Efron dice. *Australia and New Zealand Journal of Statistics* **44**, 427-438.
- [6] Brunner E., Dette H., Munk A., 1997. Box-Type Approximations in Nonparametric Factorial Designs. *Journal of the American Statistical Association* **92**, 1494-1502.
- [7] Brunner E., Munzel U., 2002a. Nichtparametrische Datenanalyse. *Springer, Berlin*.
- [8] Brunner E., Munzel U., 2000. The Nonparametric Behrens-Fisher Problem: Asymptotic Theory and a Small-Sample Approximation. *Biometrical Journal* **42**, 17-25.
- [9] Brunner E., Munzel U., Puri M.L., 2002b. The multivariate nonparametric Behrens-Fisher problem. *Journal of Statistical Planning and Inference* **108**, 37-53.
- [10] Domhof S., 2001. Nichtparametrische relative Effekte. *Dissertation an der Universität Göttingen*.
- [11] Dunnett C.W., 1955. A multiple comparison procedure for comparing several treatments with a control. *Journal of American Statistical Association* **50**, 1096-1121.
- [12] Feller W., 1968. An Introduction to Probability Theory and Its Applications. *Wiley, New York*.
- [13] Ferguson T.S., 1996. A Course in Large Sample Theory. *Chapman & Hall, London*.

- [14] Fligner M., Policello G., 1981. Robust rank procedures for the Behrens-Fisher problem. *Journal of the American Statistical Association* **76**, 162-168.
- [15] Gabriel K.R., 1969. Simultaneous Test Procedures - Some Theory of Multiple Comparisons. *The Annals of Mathematical Statistics* **40**, 224-250.
- [16] Gao X., Alvo M., 2008. Nonparametric multiple comparison procedures for unbalanced Two-way layouts. *Journal of Statistical Planning and Inference* **138**, 3674-3686.
- [17] Gao X., Alvo M., Chen J., Li G., 2008. Nonparametric multiple comparison procedures for unbalanced One-way layouts. *Journal of Statistical Planning and Inference* **138**, 2574-2591.
- [18] Gardner M., 1970. The paradox of the nontransitive dice and the elusive principle of indifference. *Scientific American* **223**, 110-124.
- [19] Hasler M., Hothorn L.A., 2008. Multiple Contrast Tests in the Presence of Heteroscedasticity. *Biometrical Journal* **50**, 793-800.
- [20] Hochberg Y., Tamhane A.C., 1987. Multiple Comparison Procedures. *Wiley*, New York.
- [21] Konietschke F., 2009. Simultane Konfidenzintervalle für nichtparametrische relative Kontrasteffekte. *Dissertation an der Universität Göttingen*.
- [22] Konietschke F., 2006. Konstruktion einfacher Konfidenzintervalle für Linearkombinationen von Erfolgswahrscheinlichkeiten. *Diplomarbeit am Institut für Mathematische Stochastik der Universität Göttingen*.
- [23] Konietschke F., Brunner E., Hothorn L.A., 2008. Nonparametric Relative Contrast Effects: Asymptotic Theory and Small Sample Approximations *Research Report at the University of Hannover*.
- [24] Kruskal W.H., Wallis W.H., 1952. Use of ranks in one-criterion variance. *Journal of the American Statistical Association* **47**, 583-621.
- [25] Kulle B., 1999. Nichtparametrisches Behrens-Fisher-Problem im Mehr-Stichprobenfall. *Diplomarbeit am Institut für Mathematische Stochastik der Universität Göttingen*.
- [26] Mann H.B., Whitney D.R., 1947. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* **18**, 50-60.
- [27] Mathai A.M., Provost S.B., 1992. Quadratic Forms in Random Variables. Theory and Applications. *Marcel Dekker*, New York.
- [28] Munzel U., Hothorn L.A., 2001. A Unified Approach to Simultaneous Rank Test Procedures in the Unbalanced One-way Layout. *Biometrical Journal* **43**, 553-569.

- [29] Orban J., Wolfe D.A., 1980. Distribution-free partialy sequentiell placement procedures. *Communications in Statistics* **9**, 883-904.
- [30] Orban J., Wolfe D.A., 1982. A Class of Distribution-Free Two Sample Tests Based on Placements. *Journal of the American Statistical Association* **77**, 666-672.
- [31] Ravishanker N., Dey D.K., 2002. A first course in linear model theory. *Chapman & Hall*, London.
- [32] Scheffé H., 1953. A method for judging all contrasts in the analysis of variance. *Biometrika* **40**, 87-104.
- [33] Thangavelu K., Brunner E. 2007. Wilcoxon-Mann-Whitney Test for Stratified Samples and Efron's Paradox Dice. *Journal of Statistical Planning and Inference* **137**, 720-737.