

Nichtparametrische simultane Inferenz für faktorielle Repeated Measures Designs



Master-Thesis an der Fakultät für Mathematik und Informatik
Georg-August-Universität Göttingen

vorgelegt von:

Marius Placzek

Erstgutachter:

Prof. Dr. Edgar Brunner

Zweitgutachterin:

Prof. Dr. Tatyana Krivobokova

EIDESSTATTLICHE ERKLÄRUNG

Hiermit erkläre ich, dass ich die vorliegende Masterarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

.....
Ort, Datum

.....
Unterschrift

VORWORT UND DANKSAGUNG

Diese Masterarbeit wurde während des Sommersemesters 2013 am Institut für Mathematische Stochastik und Informatik im Institut für Medizinische Statistik der Universitätsmedizin Göttingen angefertigt.

Für die Vergabe des Themas und die hervorragende Betreuung möchte ich mich zuallererst bei Herrn Prof. Dr. Edgar Brunner bedanken, dessen Erfahrung und Ratschläge sowie ständige Ansprechbarkeit von großer Hilfe waren.

Ebenso danke ich Dr. Frank Konietzke, der jederzeit für Fragen und Diskussionen hilfsbereit zur Verfügung stand und im Team mit Prof. Brunner die Betreuung der Arbeit übernommen hat.

Des Weiteren bedanke ich mich auch bei allen Mitarbeitern der Abteilung für die gute Arbeitsatmosphäre im Institut für Medizinische Statistik und die Hilfsbereitschaft bei Fragen jeglicher Art.

Ein spezieller Dank geht an alle meine Master-Mitstreiter, allen voran mein Büronachbar Thomas Asendorf, für sowohl fachliche Diskussionen als auch ergiebige Mittagspausen.

Für die Übernahme der Zweitkorrektur danke ich Frau Prof. Dr. Tatyana Krivobokova.

Schließlich möchte ich mich noch bei allen Freunden und natürlich meiner Familie für die Unterstützung während des Studiums bedanken.

Göttingen, im September 2013

Inhaltsverzeichnis

| | |
|---|-----------|
| 1. Einleitung | 1 |
| 1.1. Notation | 4 |
| 1.2. Motivierendes Beispiel | 5 |
| 2. Nichtparametrisches Modell und Hypothesen | 6 |
| 2.1. Der relative Effekt | 7 |
| 2.2. Hypothesen | 9 |
| 2.3. Kontrastmatrizen | 11 |
| 3. Asymptotische Normalität der linearen Rangstatistiken | 14 |
| 4. Schätzen der Kovarianzmatrix | 24 |
| 5. Prozeduren für multiple Kontrasttests | 27 |
| 5.1. Bereichserhaltende Konfidenzintervalle | 29 |
| 5.2. Approximation für kleine Stichprobenumfänge | 31 |
| 6. Teststatistiken für Globaltests | 34 |
| 6.1. Wald-Typ Statistik | 34 |
| 6.2. ANOVA-Typ Statistik | 34 |
| 6.3. Asymptotische, parametrische ANOVA | 34 |
| 6.4. Hotellings T^2 Test | 36 |
| 7. Simulationen | 38 |
| 7.1. Erzeugung von Zufallsdaten | 38 |
| 7.2. Niveau | 39 |
| 7.3. Power | 50 |
| 8. Auswertung des Beispiels | 55 |
| 9. Zusammenfassung und Diskussion | 60 |
| A. Anhang | 61 |
| A.1. Wichtige Definitionen und Sätze | 61 |
| A.2. Tabellen und Daten | 64 |
| A.3. Auswertungsprogramm | 67 |

1. Einleitung

Faktorielle Repeated Measures Designs treten häufig in praktischen Anwendungen, z.B. in klinischen Studien in der Medizin, auf. Repeated Measures meint dabei, dass jedes Individuum oder jede Versuchseinheit mehr als einmal beobachtet wird. Dabei kann es sich nicht nur um Beobachtungen zu verschiedenen Zeitpunkten handeln, sondern möglicherweise auch um mehrere, unterschiedlich behandelte, Teilproben einer Probe. Faktoriell bedeutet dabei, dass neben dem Repeated Measures Faktor der Einfluss weiterer Faktoren auf die Zielgröße untersucht wird. So kann es beispielsweise mehrere Behandlungsgruppen geben (Placebo, Medikation 1, Medikation 2). In der Praxis stellt sich bei der Auswertung eines solchen Designs nicht nur die Frage, ob es einen Behandlungs- oder Zeiteffekt gibt, sondern, falls ein solcher aufgedeckt wird, zwischen welchen Behandlungsgruppen oder Zeitpunkten Unterschiede bestehen. Das heißt, statistisch ist nicht nur die Globalhypothese interessant, sondern insbesondere multiple Vergleiche zwischen den Faktorstufen. Die klassische Herangehensweise gliedert sich folglich in drei Schritte: Zunächst werden die globalen Nullhypothesen getestet. Falls verworfen wird, werden multiple Vergleiche durchgeführt. Im letzten Schritt werden zusätzlich Konfidenzintervalle für die Effekte berechnet. Letzterer Schritt wird explizit von den Zulassungsbehörden gefordert: *Estimates of treatment effects should be accompanied by confidence intervals, whenever possible...* (ICH E9 Guideline 1998, Kap.5.5, siehe [Lewis \(1999\)](#)). Häufig verwendet werden dazu schrittweise Verfahren. Diese können jedoch dazu führen, dass der Globaltest und die multiplen Tests nicht konsonant und kohärent sind ([Gabriel, 1969](#)), d.h., es kann vorkommen, dass die Globalhypothese verworfen wird, jedoch keine der Teilhypothesen, und umgekehrt. Zudem ist es möglich, dass die Testentscheidungen der multiplen Vergleiche und die entsprechenden Konfidenzintervalle der zugehörigen Effekte inkompatibel sind ([Bretz, Genz, und Hothorn, 2001](#)), d.h., ein Konfidenzintervall enthält die Null, obwohl die entsprechende Teilhypothese verworfen wird, und umgekehrt. Eine Möglichkeit, multiple Vergleiche durchzuführen und kompatible Konfidenzintervalle zu berechnen, bietet z.B. die klassische Bonferroni-Adjustierung. Diese tendiert jedoch dazu, konservativ zu werden, und hat somit eine geringe Power. In der vorliegenden Arbeit wird ein Verfahren vorgestellt, das für faktorielle Repeated Measures Designs die oben genannten drei Schritte in einem einzigen Schritt zusammenfasst und dabei konsonante Testentscheidungen und kompatible Konfidenzintervalle bereitstellt.

Parametrische Verfahren, die solche multiplen Kontrasttests mit kompatiblen simultanen Konfidenzintervallen (SCIs) realisieren, wurden im unverbundenen Fall (keine Repeated Measures) unter der Annahme von homoskedastischen normalverteilten Stichproben von [Mukerjee, Robertson, und Wright \(1987\)](#) und [Bretz et al. \(2001\)](#) vorgestellt. In letzterer Arbeit wird die gemeinsame multivariate t-Verteilung mit Korrelationsmatrix \mathbf{R} der verschiedenen Teststatistiken verwendet, um die Teilhypothesen der multiplen Vergleiche zu testen und somit die *familywise error rate* (FWER) im starken Sinn zu kontrollieren ([Hochberg und Tamhane, 1987](#)). Diese Ergebnisse wurden erweitert auf verallgemeinerte lineare Modelle ([Bretz, Hothorn, und Westfall, 2008](#)) und heteroskedastische Modelle ([Hasler und Hothorn \(2008\)](#) und [Herberich, Sikorski, und Hothorn \(2010\)](#)), weiterhin mit

1. Einleitung

der Einschränkung unabhängiger Daten. Die Erweiterung auf mehrfaktorielle Repeated Measures Designs folgte von [Miller \(2011\)](#). Parametrische Verfahren haben jedoch den Nachteil, dass sie restriktive Annahmen an das Modell machen, und bei Verletzung dieser Annahmen, z.B. Nichtnormalität, erhebliche Powerverluste auftreten. In der Praxis ist es aber nicht ungewöhnlich, dass ordinale Daten vorliegen oder solche, die von einer schiefen oder nichtstetigen Verteilung stammen. Aus diesem Grund sind nichtparametrische Verfahren vorzuziehen.

In der Literatur gibt es zwei Herangehensweisen, nichtparametrische Verfahren zu entwerfen: Zum einen lassen sich Hypothesen über die Verteilungsfunktionen selbst formulieren ([Akritas und Arnold \(1994\)](#); [Akritas und Brunner \(1997\)](#)). Zum anderen führt man den so genannten relativen Effekt ein ([Brunner und Puri \(2001\)](#)), im Fall von ordinalen Daten auch als ordinales Effektmaß bezeichnet ([Ryu und Agresti \(2008\)](#)). Im Fall von unverbundenen a Stichproben kann man ihn paarweise definieren $p_{ij} = \int F_i dF_j$ und nichtparametrische SCIs herleiten ([Munzel und Hothorn \(2001\)](#); [Ryu \(2009\)](#); [Konietschke \(2009\)](#)). In der Arbeit von [Pfeiffer \(2010\)](#) wird die von [Konietschke \(2009\)](#) entwickelte Methode auf faktorielle Modelle erweitert. Die paarweisen relativen Effekte sind jedoch nicht transitiv ([Brown und Hettmansperger, 2002](#)) und ihre Verwendung bei multiplen Vergleichen kann somit zu paradoxen Ergebnissen führen (vgl. Efrons Würfel, [Thangavelu und Brunner \(2006\)](#)) und ist nicht zu empfehlen. Besser geeignet ist der transitive Effekt $p_i = \int H dF_i$, der sich aus dem Vergleich mit einer mittleren Verteilungsfunktion $H = \sum_{i=1}^a w_i F_i$ ergibt (gewichtet: $w_i = n_i/N$; ungewichtet: $w_i = 1/a$). Im Fall von unabhängigen Stichproben verwenden [Brunner und Puri \(2001\)](#) sowie [Gao und Alvo \(2008\)](#) und [Gao, Alvo, Chen, und Li \(2008\)](#) diesen Effekt, um H_0^F zu testen, jedoch ohne simultane Konfidenzintervalle bereitzustellen. Auch der bekannte Kruskal-Wallis-Test ([Kruskal und Wallis, 1952](#)) testet H_0^F und verwendet dabei die über die gewichtete mittlere Verteilungsfunktion definierten Effekte. Für ein solches One-Way-Layout mit a unverbundenen Stichproben entwickelten [Konietschke, Hothorn, und Brunner \(2012\)](#) eine multiple Kontrastprozedur mit zugehörigen simultanen, kompatiblen Konfidenzintervallen für die relativen Behandlungseffekte. Für ein einfaches Repeated Measures Design (nur eine Stichprobe mit pro Individuum a Repeated Measurements) stellten [Konietschke, Bathke, Hothorn, und Brunner \(2010\)](#) ein Verfahren vor, das ebenso multiple Vergleiche sowie simultane Konfidenzintervalle bereitstellt. Beide Verfahren sind im frei verfügbaren R-Paket `nparcomp` implementiert. Schätzer für die relativen Effekte lassen sich leicht mit Hilfe von globalen, internen und paarweisen Rängen konstruieren, weshalb derartige nichtparametrische Verfahren auch als Rangverfahren bezeichnet werden.

In der vorliegenden Arbeit wird eine rangbasierte *Multiple Comparison Test Procedure* (MCTP) mit kompatiblen, simultanen Konfidenzintervallen in unbalancierten faktoriellen Repeated Measures Designs mit unabhängigen Beobachtungsvektoren und fester Dimension hergeleitet. Dabei sollen beliebige Kontraste, z.B. Tukey all-pairs oder Dunnett many-to-one, für die multiplen Vergleiche zur Verfügung stehen. Die Aufgabe wird also sein, die asymptotische, gemeinsame Verteilung der entsprechenden Teststatistiken unter beliebigen Alternativen zu bestimmen. In Zuge dessen muss die asymptotische Kovarianzmatrix der Rangstatistiken berechnet und geschätzt werden, wobei dies mit

1. Einleitung

geschickter Matrizenmultiplikation vereinfacht werden wird. Das Resultat soll eine Verallgemeinerung der in *nparcomp* implementierten Verfahren `mctp` und `mcpt.rm` darstellen und diese natürlich als Spezialfall enthalten.

Der Aufbau der Arbeit stellt sich wie folgt dar: Im zweiten Kapitel wird zunächst das zu betrachtende nichtparametrische Modell vorgestellt, insbesondere wird noch einmal der relative Effekt präsentiert. Darauf folgt die Diskussion der in diesem Modell verwendeten Hypothesen. Zudem wird ein kleiner Überblick über Kontraste bzw. Kontrastmatrizen gegeben, die zum Durchführen der multiplen Vergleiche zur Verfügung stehen. Im Anschluss daran werden in Kapitel 3 Schätzer für die relativen Effekte angegeben und deren Eigenschaften erläutert. Mit dem Vektor der relativen Effekte bzw. dessen Schätzer wird eine lineare Rangstatistik gebildet und deren asymptotische Normalität mit Hilfe des asymptotischen Äquivalenzsatzes gezeigt. Dabei wird die asymptotische Kovarianzmatrix hergeleitet, deren Schätzer in Kapitel 4 angegeben wird. In Kapitel 5 wird dann die MCTP konstruiert, nebst simultaner, bereichserhaltender Konfidenzintervalle sowie einer Approximation für kleine Stichprobenumfänge. Bevor in Kapitel 7 Simulationsergebnisse vorgestellt werden, werden in Kapitel 6 einige Teststatistiken für Globaltests aufgeführt, mit denen bei der Simulationsstudie verglichen wird. Anschließend wird das im nächsten Abschnitt präsentierte, motivierende Beispiel mit einer eigens implementierten R-Funktion, die die MCTP umsetzt, ausgewertet (Kapitel 8). Zum Abschluss bietet Kapitel 9 eine kleine Zusammenfassung und Diskussion.

1.1. Notation

Der Übersichtlichkeit halber wird in dieser Arbeit eine konsequente Notation eingehalten. Matrizen werden mit Großbuchstaben belegt und fett geschrieben. Vektoren werden ebenfalls fett gedruckt. Aus dem Kontext und der Dimensionierung sollte es klar sein, wann es sich um eine Matrix und wann um einen Vektor handelt. Skalare und Konstanten werden in normaler Schriftart mit Kleinbuchstaben belegt. Zufallsvariablen werden mit Großbuchstaben in normaler Schriftart gekennzeichnet. Schätzer werden stets mit einem Dach oder einer Tilde versehen. Beispielsweise könnte eine Kovarianzmatrix mit \mathbf{V} , ein zugehöriger Schätzer mit $\widehat{\mathbf{V}}$ oder $\widetilde{\mathbf{V}}$ bezeichnet werden. Dagegen könnte ein Freiheitsgrad als f definiert und mittels \widehat{f} oder \widetilde{f} geschätzt werden.

Eine der wichtigen Funktionen für eine quadratische Matrix \mathbf{M} ist zum einen die Spur, die auf Grund ihrer englischen Übersetzung *trace* mit $\text{tr}(\mathbf{M})$ abgekürzt wird. Zum anderen gibt es den Rang (engl. *rank*) einer Matrix. Dieser wird mit $\text{rk}(\mathbf{M})$ auf eine Matrix angewendet.

Häufig verwendete Matrixoperatoren sind vor allem die direkte Summe $\mathbf{M} \oplus \mathbf{N}$ und das Kronecker-Produkt $\mathbf{M} \otimes \mathbf{N}$.

Die stochastischen Operatoren wie der Erwartungswert, die Varianz und Kovarianz werden mit $\mathbb{E}(\cdot)$, $\mathbb{V}\text{ar}(\cdot)$ sowie $\mathbb{C}\text{ov}(\cdot)$ geschrieben.

Die Schreibweise \sim bedeutet *verteilt nach*, dagegen meint $\overset{\sim}{\sim}$ *asymptotisch verteilt wie*. Die Notation $\overset{\doteq}{\doteq}$ meint *asymptotisch äquivalent zu*. Eine Normalverteilung mit Erwartungswert $\boldsymbol{\mu}$ und Kovarianzmatrix \mathbf{V} wird mit $\mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$ beschrieben, eine Chi-Quadrat-Verteilung mit Freiheitsgrad f über die Darstellung χ_f^2 . Die F-Verteilung mit ihren beiden Freiheitsgraden f, f_0 wird mit $F(f, f_0)$ gekennzeichnet und die multivariate t-Verteilung mit ν Freiheitsgraden, Erwartungswert $\mathbf{0}$ und Korrelationsmatrix \mathbf{R} mit $T(\nu, \mathbf{0}, \mathbf{R})$.

Fast sichere Konvergenz wird mit $\xrightarrow{a.s.}$ beschrieben, Konvergenz in Wahrscheinlichkeit mit \xrightarrow{p} und Konvergenz in Verteilung entsprechend mit \xrightarrow{d} . Definitionen zu den verschiedenen Konvergenzarten sowie den Notationen zur Asymptotik finden sich im Anhang.

Die wichtigsten Matrizen sind:

- (i) Die $d \times d$ Einheitsmatrix \mathbf{I}_d .
- (ii) Der d -dimensionale Einservektor $\mathbf{1}_d = (1, \dots, 1)'$.
- (iii) Die $d \times d$ Einsermatrix $\mathbf{J}_d = \mathbf{1} \cdot \mathbf{1}'$.
- (iv) Die zentrierende Matrix $\mathbf{P}_d = \mathbf{I}_d - \frac{1}{d}\mathbf{J}_d$. Diese ist sowohl symmetrisch ($\mathbf{P}_d' = \mathbf{P}_d$) als auch idempotent ($\mathbf{P}_d \cdot \mathbf{P}_d = \mathbf{P}_d$). Ihr Rang ist $d - 1$ und somit ist auch $\text{tr}(\mathbf{P}_d) = d - 1$.

1.2. Motivierendes Beispiel

Als einführendes praktisches Beispiel soll an dieser Stelle eine Studie zur PCT-Konzentration vorgestellt werden (vgl. Brunner und Langer (2002), Abschnitt 1.3.12). Dabei handelt es sich um die Untersuchung des Verlaufes der Serum-Konzentration des Peptids Procalcitonin (PCT). Gemessen wurde dieser Verlauf prä- und post-operativ bei 71 Patienten bezüglich verschiedener Arten von Operationen in der Herz-, Thorax- und Gefäßchirurgie. Ausgewählt wurden nur Patienten, bei denen postoperativ keine Infektion festgestellt wurde. Unterteilt werden konnte in drei Gruppen, die drei unterschiedlichen chirurgischen Eingriffen entsprechen: Gruppe A fasst die 38 Patienten zusammen, bei denen während der Herzoperation eine Herz-Lungen-Maschine eingesetzt wurde. Bei den 17 Patienten aus Gruppe B musste eine Lungenoperation durchgeführt werden und Gruppe C besteht aus 16 Patienten mit einer Gefäß-Operation. Die PCT-Konzentration wurde zunächst prä-operativ (Zeitpunkt 1) festgestellt und dann direkt nach der Operation (Zeitpunkt 2), einen Tag (Zeitpunkt 3) und zehn Tage (Zeitpunkt 4) nach der Operation gemessen. Es handelt sich hier also um ein zweifaktorielles Design mit dem whole-Plot-Faktor *Behandlung* (3 Stufen) und dem sub-Plot-Faktor *Zeit* (4 Stufen). Zu beachten ist, dass die PCT-Konzentration nur dann gemessen werden kann, wenn sie eine gewisse untere Schranke (0.2) übertrifft, was bei jedem Patienten der Fall war. Die Original-Daten befinden sich im Anhang (Tabelle 12). Abbildung 1 gibt einen Eindruck der Profilverläufe in den drei Gruppen. Für den Anwender interessant war nun die Frage, ob sich der Verlauf der Serum-Konzentration in den Gruppen signifikant unterscheidet, d.h., kann eine Wechselwirkung von Behandlung und Zeit nachgewiesen werden? Von besonderem Interesse ist der Vergleich der ersten Gruppe mit den anderen beiden Gruppen. Das Beispiel wird in Abschnitt 8 ausgewertet.

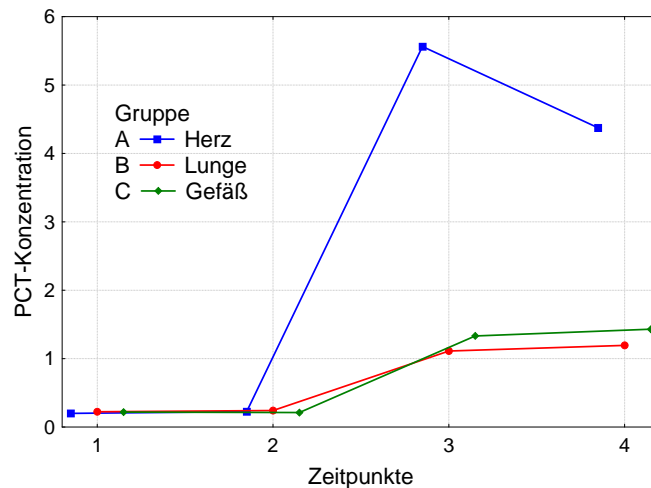


Abbildung 1: Mittelwertsverläufe der PCT-Konzentrationen.

2. Nichtparametrisches Modell und Hypothesen

Betrachtet wird ein Design mit $N = \sum_{i=1}^a n_i$ unabhängigen Versuchseinheiten, aufgeteilt auf a Gruppen. Bei jeder Versuchseinheit werden d Repeated Measures vorgenommen. Gemessen werden also Zufallsvektoren

$$\mathbf{X}_{ik} = (X_{i1k}, X_{i2k}, \dots, X_{idk}), \quad i = 1, \dots, a; \quad k = 1, \dots, n_i.$$

Es wird nun angenommen, dass das Modell marginal beschrieben werden kann, d.h. $X_{ijk} \sim F_{ij}$, mit den Verteilungsfunktionen

$$F_{ij}(x) = \mathbb{P}(X_{ijk} < x) + \frac{1}{2}\mathbb{P}(X_{ijk} = x).$$

Dabei handelt es sich bei $F_{ij}(x)$ um die normalisierte Version der Verteilungsfunktion, d.h. den Mittelwert aus linksstetiger ($F^-(x) = \mathbb{P}(X < x)$) und rechtsstetiger Version ($F^+(x) = \mathbb{P}(X \leq x)$). Bezüglich nichtparametrischer Modelle wurde diese normalisierte Version der Verteilungsfunktion erstmals von [Lévy \(1925\)](#) erwähnt. In der Folge verwendeten sie unter anderem [Ruymgaart \(1980\)](#), [Akritas und Brunner \(1997\)](#), [Munzel \(1999\)](#) und [Gao et al. \(2008\)](#), um asymptotische Ergebnisse für Rangstatistiken herzuleiten. Diese Resultate sind sowohl für Daten mit als auch ohne Bindungen gültig. Man beachte, dass es sich bei den F_{ij} bis auf den trivialen Fall einer Einpunktverteilung um beliebige Verteilungsfunktionen handeln darf. Zur Übersicht zeigt folgendes Schaubild die Anordnung der Daten:

| Faktor A | VE | Faktor D | | | | | Beobachtungsvektor |
|----------|-------|-------------|-----|-------------|-----|-------------|---------------------|
| | | 1 | ... | s | ... | d | |
| 1 | 1 | X_{111} | ... | X_{1s1} | ... | X_{1d1} | \mathbf{X}_{11} |
| | ⋮ | ⋮ | | ⋮ | | ⋮ | ⋮ |
| | k | X_{11k} | ... | X_{1sk} | ... | X_{1dk} | \mathbf{X}_{1k} |
| | ⋮ | ⋮ | | ⋮ | | ⋮ | ⋮ |
| | n_1 | X_{11n_1} | ... | X_{1sn_1} | ... | X_{1dn_1} | \mathbf{X}_{1n_1} |
| ⋮ | ⋮ | ⋮ | | ⋮ | | ⋮ | |
| a | 1 | X_{a11} | ... | X_{as1} | ... | X_{ad1} | \mathbf{X}_{a1} |
| | ⋮ | ⋮ | | ⋮ | | ⋮ | ⋮ |
| | k | X_{a1k} | ... | X_{ask} | ... | X_{adk} | \mathbf{X}_{ak} |
| | ⋮ | ⋮ | | ⋮ | | ⋮ | ⋮ |
| | n_a | X_{a1n_a} | ... | X_{asn_a} | ... | X_{adn_a} | \mathbf{X}_{an_a} |

Das allgemeine Modell fordert an dieser Stelle lediglich, dass

$$X_{ijk} \sim F_{ij}, \quad k = 1, \dots, n_i, \tag{2.1}$$

2. Nichtparametrisches Modell und Hypothesen

nicht aber, dass die Verteilungen in irgendeiner parametrischen Relation stehen. So wird zum Beispiel nicht einmal Homoskedastizität verlangt (Akritas und Brunner, 1997). Die Verteilungsfunktionen werden zusammengefasst im Vektor $\mathbf{F} = (F_{11}, F_{12}, \dots, F_{ad})'$.

2.1. Der relative Effekt

Eine Prozedur für multiple Vergleiche mit simultanen Konfidenzintervallen eignet sich beispielsweise zur Analyse von unterschiedlichen Behandlungsverfahren. Dazu wird aber erst einmal eine Größe benötigt, mit der man Unterschiede zwischen den Faktorstufen beschreiben kann. Insbesondere möchte man in der Lage sein, einen Behandlungseffekt zu beschreiben und für diesen Konfidenzintervalle anzugeben. Im parametrischen Modell, wenn sich die Verteilungsfunktionen klassifizieren lassen, kann man solche Unterschiede meistens durch Differenzen oder Quotienten von Parametern, die diese Funktionen charakterisieren, beschreiben. Der typische und schon häufig untersuchte Fall ist die Annahme der Normalverteilung, im Zuge derer Differenzen von Erwartungswerten bzw. Quotienten der Varianzen diesem Zweck dienen (vgl. auch ANOVA). Das allgemeine Modell (2.1) enthält keinerlei Parameter, mithilfe derer ein Unterschied zwischen den Verteilungen beschrieben werden könnte. In diesem nichtparametrischen Setting werden daher die Verteilungsfunktionen $F_{ij}(x)$ verwendet, um Behandlungseffekte zu definieren. Sei dazu $H = \frac{1}{ad} \sum_{i=1}^a \sum_{j=1}^d F_{ij}$ eine mittlere Verteilung in ungewichteter Form (vergleiche Brunner und Puri (2001); Gao et al. (2008)). Die so genannten relativen Effekte werden dann definiert als

$$p_{ij} = \int H dF_{ij}, \quad i = 1, \dots, a; \quad j = 1, \dots, d. \quad (2.2)$$

Man beachte, dass auf Grund dieser Definition von H die relativen Effekte nicht von den Stichprobenumfängen abhängen und daher besser zu interpretieren sind als solche, die mit Hilfe einer gewichteten mittleren Verteilung definiert werden. Zur Interpretation ist folgendes Lemma nützlich:

Lemma 2.1. *Seien X_{ij1} und X_{rs2} unabhängige Zufallsvariablen und $X_{ij1} \sim F_{ij}$ sowie $X_{rs2} \sim F_{rs}$. Dann ist*

$$\int F_{ij} dF_{rs} = \mathbb{P}(X_{ij1} < X_{rs2}) + \frac{1}{2} \mathbb{P}(X_{ij1} = X_{rs2}). \quad (2.3)$$

Beweis. Vgl. Brunner und Munzel (2002). Da der Beweis sehr kurz ist, wird er hier noch einmal angegeben:

$$\begin{aligned} \mathbb{P}(X_{ij1} < X_{rs2}) + \frac{1}{2} \mathbb{P}(X_{ij1} = X_{rs2}) &= \int [\mathbb{P}(X_{ij1} < x) + \frac{1}{2} \mathbb{P}(X_{ij1} = x)] dF_{rs}(x) \\ &= \int F_{ij} dF_{rs}. \end{aligned}$$

□

2. Nichtparametrisches Modell und Hypothesen

Mit Hilfe dieses Lemmas hat man nun eine einfache Anschauung: Ist X_{ijk} eine Beobachtung im beschriebenen Modell und Z eine Zufallsvariable unabhängig von X_{ijk} und $Z \sim H$. Dann gilt

$$p_{ij} = \mathbb{P}(Z < X_{ijk}) + \frac{1}{2}\mathbb{P}(Z = X_{ijk}). \quad (2.4)$$

Das neue Maß lässt sich also wie folgt interpretieren: Zunächst einmal kann man beim Vergleich mit der mittleren Verteilung H feststellen, dass für $p_{ij} < 1/2$ die Verteilung F_{ij} tendenziell weiter links als H liegt, im Fall $p_{ij} > 1/2$ weiter rechts als H . Für $p = 1/2$ ist keine Tendenz festzustellen. Aber auch zwei relative Effekte lassen sich in diesem Modell gut interpretierbar vergleichen: Ist $p_{ij} < p_{rs}$, dann tendieren Messungen, die von Verteilung F_{ij} stammen, zu kleineren Werten als solche von Verteilung F_{rs} . Im Fall $p_{ij} = p_{rs}$ gibt es zwischen F_{ij} und F_{rs} keine Tendenz zu größeren oder kleineren Werten. Die Definition der relativen Effekte (2.2) in der ungewichteten Form hat zum einen den Vorteil, dass sie nicht von den Stichprobenumfängen abhängen, zum anderen können keine paradoxen Aussagen (vgl. Efrons Würfel) entstehen, da jede Verteilung stets mit der festen Referenzverteilung H verglichen wird (Thangavelu und Brunner, 2006).

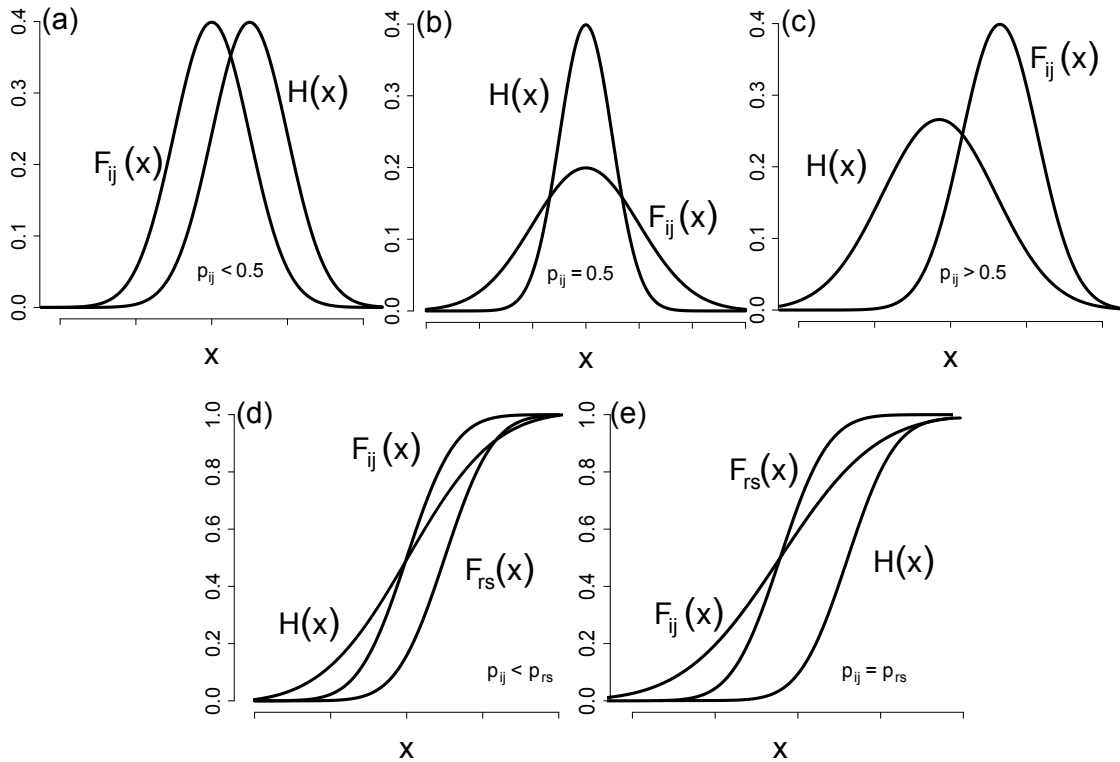


Abbildung 2: Visualisierung der p_{ij} : Dichtefunktionen (oben), Verteilungsfunktionen (unten). Graphiken aus Konietzsche und Hothorn (2012).

2. Nichtparametrisches Modell und Hypothesen

Die Verwendung des relativen Effektes bietet einen weiteren Vorteil.

Lemma 2.2. *Die relativen Effekte p_{ij} sind invariant unter einer beliebigen ordnungserhaltenden, messbaren Transformation $m(\cdot)$.*

Beweis. Siehe [Brunner und Munzel \(2002\)](#), Proposition 1.6. Es gilt

$$\begin{aligned} p_{ij} &= \mathbb{P}(Z < X_{ijk}) + \frac{1}{2}\mathbb{P}(Z = X_{ijk}) \\ &= \mathbb{P}(m(Z) < m(X_{ijk})) + \frac{1}{2}\mathbb{P}(m(Z) = m(X_{ijk})), \end{aligned}$$

da $m(\cdot)$ nach Voraussetzung eine messbare und ordnungserhaltende Transformation ist.

□ Diese Invarianz macht

den relativen Effekt zu einem Maß, das sich hervorragend eignet, um sowohl ordinale Daten, bei denen lediglich Informationen über die Ordnungsstruktur bekannt sind, als auch metrische Daten zu analysieren. Gerade metrische Daten, deren Mess-Skala nicht von vornherein bestimmt ist, stellen dadurch auch kein Problem dar. Zudem beruht es nicht auf Differenzen und ist somit ein robustes Unterscheidungsmaß.

Bezeichnet man nun die Größe aus (2.3) in Lemma 2.1 mit $w_{rsij} = \int F_{rs} dF_{ij}$, dann lässt sich der relative Effekt p_{ij} schreiben als

$$\begin{aligned} p_{ij} &= \int H dF_{ij} = \frac{1}{ad} \sum_{r=1}^a \sum_{s=1}^d \int F_{rs} dF_{ij} \\ &= \frac{1}{ad} \sum_{r=1}^a \sum_{s=1}^d w_{rsij}, \quad i = 1, \dots, a; \quad j = 1, \dots, d. \end{aligned} \tag{2.5}$$

Setze $\mathbf{p} = (p_{11}, p_{12}, \dots, p_{ad})' = \int H d\mathbf{F}$ als den Vektor der relativen Behandlungseffekte. Die Darstellung von p_{ij} in (2.5) wird es im Folgenden ermöglichen, eine einfache Darstellung der Kovarianzmatrix der linearen Rangstatistiken unter beliebigen Alternativen herzuleiten.

2.2. Hypothesen

Wollte man in diesem nichtparametrischen Modell lediglich globale Hypothesen testen, so würde man vorgehen wie im parametrischen Modell, jedoch ersetzen die Verteilungsfunktionen F_{ij} bzw. die relativen Effekte p_{ij} die Erwartungswerte μ_{ij} . Die Formulierungen der Hypothesen *kein Haupteffekt des Whole-Plot Faktors A*, *kein Haupteffekt des Repeated Measures Faktors D* sowie *kein Wechselwirkungseffekt AD* lauten somit

$$\begin{aligned} H_0^F(A) : (\mathbf{P}_a \otimes \frac{1}{d}\mathbf{1}'_d) \cdot \mathbf{F} = \mathbf{0} & \quad H_0^P(A) : (\mathbf{P}_a \otimes \frac{1}{d}\mathbf{1}'_d) \cdot \mathbf{p} = \mathbf{0} \\ H_0^F(D) : (\frac{1}{a}\mathbf{1}'_a \otimes \mathbf{P}_d) \cdot \mathbf{F} = \mathbf{0} & \quad \text{bzw.} \quad H_0^P(D) : (\frac{1}{a}\mathbf{1}'_a \otimes \mathbf{P}_d) \cdot \mathbf{p} = \mathbf{0} \\ H_0^F(AD) : (\mathbf{P}_a \otimes \mathbf{P}_d) \cdot \mathbf{F} = \mathbf{0} & \quad H_0^P(AD) : (\mathbf{P}_a \otimes \mathbf{P}_d) \cdot \mathbf{p} = \mathbf{0} \end{aligned} \tag{2.6}$$

Zum globalen Testen von Hypothesen der Form $H_0^F : \mathbf{C}\mathbf{F} = \mathbf{0}$ in einem Design mit beliebigen Faktoren und Kontrastmatrix \mathbf{C} sei auf die Arbeiten von [Akritas, Arnold](#),

2. Nichtparametrisches Modell und Hypothesen

und Brunner (1997) sowie Brunner und Puri (2001) verwiesen. In einem einfaktoriellen Modell ohne Repeated Measures (Faktor A mit a Levels) betrachten Gao et al. (2008) folgende Familie von Hypothesen

$$\Omega^F = \{H_0^F : \mathbf{c}'_\ell \mathbf{F} = 0, \ell = 1, \dots, q\}, \quad (2.7)$$

um rangbasierte, schrittweise Testprozeduren für many-to-one

$$H_0^F : \begin{cases} F_1 = F_2 \\ F_1 = F_3 \\ \vdots \\ F_1 = F_d \end{cases} \iff H_0^F : \mathbf{C}\mathbf{F} = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ -1 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & 0 & 0 & \dots & \dots & 1 \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_d \end{pmatrix} = \mathbf{0},$$

und all-pairs Vergleiche

$$H_0^F : \begin{cases} F_1 = F_2 \\ F_1 = F_3 \\ \vdots \\ F_1 = F_a \\ F_2 = F_3 \\ \vdots \\ F_{a-1} = F_a \end{cases} \iff H_0^F : \mathbf{C}\mathbf{F} = \begin{pmatrix} -1 & 1 & 0 & \dots & \dots & 0 & 0 \\ -1 & 0 & 1 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & 0 & 0 & 0 & \dots & \dots & 1 \\ 0 & -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \dots & \dots & -1 & 1 \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_a \\ \vdots \\ F_a \end{pmatrix} = \mathbf{0},$$

herzuleiten. Ausgehend von der Tatsache, dass Testprozeduren für H_0^F sich jedoch nur zum Testen eignen und nicht verwendet werden können, um Konfidenzintervalle für die zugrunde liegenden Behandlungseffekte $\delta_\ell = \mathbf{c}'_\ell \mathbf{p}$ zu konstruieren (ausgenommen bei Shift-Effekten), betrachten Konietzsche et al. (2012) die entsprechende Familie von Hypothesen mit relativen Effekten:

$$\Omega^p = \{H_0^p : \mathbf{c}'_\ell \mathbf{p} = 0, \ell = 1, \dots, q\}. \quad (2.8)$$

In ihrer Arbeit leiten sie eine Testprozedur für multiple Vergleiche für Ω^p und dazu kompatible, simultane Konfidenzintervalle für die Effekte $\delta_\ell = \mathbf{c}'_\ell \mathbf{p}$ her. Kompatibel meint in diesem Zusammenhang, dass, falls eine Teilhypothese verworfen wird, das entsprechende Konfidenzintervall nicht die Null enthält und umgekehrt. Man beachte, dass die Hypothese $H_0^F : \mathbf{C}\mathbf{F} = \mathbf{0}$ die Hypothese $H_0^p : \mathbf{C}\mathbf{p} = \mathbf{0}$ impliziert:

$$H_0^F : \mathbf{C}\mathbf{F} = \mathbf{0} \Rightarrow H_0^p : \mathbf{C}\mathbf{p} = \mathbf{C} \int H d\mathbf{F} = \int H d(\mathbf{C}\mathbf{F}) = \mathbf{0}.$$

Genauere Betrachtungen zu den obigen Hypothesen finden sich in Akritas et al. (1997) und Brunner und Munzel (2000). Zur Formulierung der Nullhypothesen bei multiplen Testprozeduren bedarf es also geeigneter Kontrastmatrizen, um die Fragen der Anwender korrekt auf das statistische Modell zu übertragen. Es lohnt sich somit, einen kleinen Blick auf deren Vielfalt zu werfen.

2.3. Kontrastmatrizen

Bei einer Kontrastmatrix \mathbf{C} handelt es sich um eine Matrix, bei der sich jede Zeile, d.h. jeder Kontrast, zu Null aufsummiert. Angenommen wir befinden uns in einem One-Way Layout mit a Gruppen und wollen q Vergleiche durchführen, dann hätte die Kontrastmatrix folgende Form:

$$\mathbf{C} = \begin{pmatrix} \mathbf{c}'_1 \\ \vdots \\ \mathbf{c}'_\ell \\ \vdots \\ \mathbf{c}'_q \end{pmatrix} = \begin{pmatrix} c_{11} & \dots & c_{1a} \\ \vdots & \dots & \vdots \\ c_{\ell 1} & \dots & c_{\ell a} \\ \vdots & \dots & \vdots \\ c_{q1} & \dots & c_{qa} \end{pmatrix} \in \mathbb{R}^{q \times a}, \quad \sum_{i=1}^a c_{\ell i} = 0 \quad \forall \ell = 1, \dots, q. \quad (2.9)$$

Jeder Kontrast $\mathbf{c}'_\ell = (c_{\ell 1}, \dots, c_{\ell a})$ multipliziert mit dem Vektor der relativen Effekte $\mathbf{p} = (p_1, \dots, p_a)'$ ergibt einen einzelnen Vergleich entsprechend der Fragestellung im Versuch. Im Folgenden sollen einige bereits vordefinierte, in der Praxis häufig verwendete Kontrastmatrizen vorgestellt werden:

1. *Tukey* (all-pairs Vergleiche):

Die Tukey-Kontrastmatrix (Tukey, 1953) ist eine der bekanntesten Kontrastmatrizen. Sie realisiert bei einem Faktor mit a Levels alle paarweisen Vergleiche der zugehörigen Erwartungswerte (im parametrischen Modell) bzw. relativen Effekte oder Verteilungsfunktionen (im nichtparametrischen Modell). Es ist also $q = \binom{a}{2}$ die Anzahl der Kontraste bzw. Vergleiche.

$$H_0^p : \begin{cases} p_1 = p_2 \\ p_1 = p_3 \\ \vdots \\ p_1 = p_a \\ p_2 = p_3 \\ \vdots \\ p_{a-1} = p_a \end{cases} \iff H_0^p : \mathbf{C}\mathbf{p} = \begin{pmatrix} -1 & 1 & 0 & \dots & \dots & 0 & 0 \\ -1 & 0 & 1 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & 0 & 0 & 0 & \dots & \dots & 1 \\ 0 & -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \dots & \dots & -1 & 1 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_a \end{pmatrix} = \mathbf{0}.$$

2. *Dunnnett* (many-to-one Vergleiche):

Nicht weniger bekannt ist die Dunnnett-Kontrastmatrix (Dunnnett, 1955), die alle paarweisen Vergleiche einer gewählten Faktorstufe mit allen anderen Faktorstufen beschreibt. So lassen sich z.B. alle Dosisstufen gegen eine Placebo-Gruppe testen. Häufig wird die Vergleichsgruppe als erste Faktorstufe gesetzt. Man spricht dabei von einer Kontrollgruppe und einem so genannten Baseline-Vergleich. Es werden also $q = a - 1$ Vergleiche durchgeführt.

$$H_0^p : \begin{cases} p_1 = p_2 \\ p_1 = p_3 \\ \vdots \\ p_1 = p_d \end{cases} \iff H_0^p : \mathbf{C}\mathbf{p} = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ -1 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & 0 & 0 & \dots & \dots & 1 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_d \end{pmatrix} = \mathbf{0}.$$

2. Nichtparametrisches Modell und Hypothesen

3. Average (Vergleiche mit dem Mittelwert):

Hier werden die einzelnen Faktorstufen jeweils mit dem Mittelwert über alle Faktorstufen verglichen. Dies ist ähnlich dem Vorgehen bei der Varianzanalyse (ANOVA). Die Anzahl der Kontraste ist gleich der Zahl der Faktorstufen, d.h. $q = a$. Ist $\bar{p} = \frac{1}{a} \sum_{i=1}^a p_i$, dann hat man als Kontrastmatrix offenbar $\mathbf{C} = \mathbf{P}_a$, die zentrierende Matrix:

$$H_0^p : \begin{cases} p_1 = \bar{p}. \\ p_2 = \bar{p}. \\ \vdots \\ p_a = \bar{p}. \end{cases} \iff H_0^p : \mathbf{Cp} = \begin{pmatrix} 1 - \frac{1}{a} & -\frac{1}{a} & \cdots & -\frac{1}{a} & -\frac{1}{a} \\ -\frac{1}{a} & 1 - \frac{1}{a} & -\frac{1}{a} & \cdots & -\frac{1}{a} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\frac{1}{a} & \cdots & \cdots & -\frac{1}{a} & 1 - \frac{1}{a} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_d \end{pmatrix} = \mathbf{0}.$$

4. Changepoint:

Mit diesen $q = a - 1$ Kontrasten wird versucht, unter den Faktorstufen einen Changepoint in den relativen Effekten zu finden. Dazu werden systematisch Linearkombinationen von relativen Effekten miteinander verglichen. Gewichtet wird entsprechend der zugehörigen Stichprobenumfänge, um die optimale Power zu erreichen.

$$H_0^p : \begin{aligned} p_1 &= \frac{\sum_{i=2}^a n_i p_i}{\sum_{i=2}^a n_i} \\ \frac{\sum_{i=1}^2 n_i p_i}{\sum_{i=1}^2 n_i} &= \frac{\sum_{i=3}^a n_i p_i}{\sum_{i=3}^a n_i} \\ &\vdots \\ \frac{\sum_{i=1}^{a-1} n_i p_i}{\sum_{i=1}^{a-1} n_i} &= p_a \end{aligned}$$

$$\iff H_0^p : \mathbf{Cp} = \begin{pmatrix} -1 & \frac{n_2}{\sum_{i=2}^a n_i} & \cdots & \frac{n_{a-1}}{\sum_{i=2}^a n_i} & \frac{n_a}{\sum_{i=2}^a n_i} \\ -\frac{n_1}{\sum_{i=1}^2 n_i} & -\frac{n_2}{\sum_{i=1}^2 n_i} & \frac{n_3}{\sum_{i=3}^a n_i} & \cdots & \frac{n_a}{\sum_{i=3}^a n_i} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\frac{n_1}{\sum_{i=1}^{a-1} n_i} & \cdots & \cdots & -\frac{n_{a-1}}{\sum_{i=1}^{a-1} n_i} & 1 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_d \end{pmatrix} = \mathbf{0}.$$

Dies ist nur ein kleiner Ausschnitt von möglichen Kontrastmatrizen. Die Entscheidung, welche Kontrastmatrix zur Analyse eines Versuchs gewählt wird, hängt einzig und allein von den Fragestellungen der Studie ab. Auch benutzerdefinierte Kontrastmatrizen sind möglich, solange sie die Eigenschaft einer Kontrastmatrix besitzen und auf den Versuch abgestimmt sind.

Da sich die vorliegende Arbeit mit einem mehrfaktoriellen Design beschäftigt, wird an dieser Stelle kurz veranschaulicht, wie in diesem Fall multiple Vergleiche für die Haupteffekte und Wechselwirkung mit den soeben angesprochenen Kontrastmatrizen formuliert werden können.

2. Nichtparametrisches Modell und Hypothesen

- *Haupteffekt A:*
Um Vergleiche bezüglich des Haupteffektes des whole-plot Faktors A durchzuführen, muss zunächst zentriert und über die Messpunkte des Repeated Measures Faktors D gemittelt werden. Die Kontrastmatrix $\mathbf{C}_A \in \mathbb{R}^{q \times ad}$ ergibt sich also als $\mathbf{C}_A = \mathbf{C}_a \cdot (\mathbf{P}_a \otimes \frac{1}{d} \mathbf{1}'_d) = \mathbf{C}_a \otimes \frac{1}{d} \mathbf{1}'_d$. Dabei ist $\mathbf{C}_a \in \mathbb{R}^{q \times a}$ eine beliebige Kontrastmatrix wie sie z.B. im vorigen Abschnitt 2.3 vorgestellt wurden. Bezeichnen nun wieder $\mathbf{c}'_\ell = (c_{\ell 11}, c_{\ell 12}, \dots, c_{\ell ad})$, $\ell = 1, \dots, q$ die einzelnen Kontraste von \mathbf{C}_A , dann hat man, wie schon gesehen, $\Omega^p = \{H_0^p : \mathbf{c}'_\ell \mathbf{p} = 0, \ell = 1, \dots, q\}$ als Familie von Hypothesen.
- *Haupteffekt D:*
Analog zur gerade beschriebenen Vorgehensweise beim Haupteffekt des whole-plot Faktors wird, um multiple Vergleiche zwischen den Stufen des Repeated Measures Faktors D durchzuführen, die Kontrastmatrix $\mathbf{C}_D \in \mathbb{R}^{q \times ad}$ konstruiert als Produkt aus einer Kontrastmatrix für einen einzelnen Faktor \mathbf{C}_d und der entsprechenden Zentrierung $\frac{1}{a} \mathbf{1}'_a \otimes \mathbf{P}_d$. Auf Grund der Beziehung $\mathbf{C}_d \cdot \mathbf{P}_d = \mathbf{0}$ vereinfacht sich dieses Produkt wiederum zu $\mathbf{C}_A = \frac{1}{a} \mathbf{1}'_a \otimes \mathbf{C}_d$.
- *Wechselwirkung AD*
Bezüglich der Wechselwirkung machen multiple Vergleiche interpretatorisch nur wenig Sinn. Daher verwendet man die für die Wechselwirkung bekannte Zentrierung als Kontrastmatrix $\mathbf{C}_{AD} = \mathbf{P}_a \otimes \mathbf{P}_d$. Man hat dann als Familie von Nullhypothesen $\Omega^p = \{H_0^p : p_{ij} - \bar{p}_{i.} - \bar{p}_{.j} + \bar{p}.. = 0, \text{ für alle } i = 1, \dots, a; j = 1, \dots, d\}$. Alternativ kann man auch einen Average-Kontrast anwenden, um die einzelnen Zellen mit dem Mittelwert zu vergleichen.

3. Asymptotische Normalität der linearen Rangstatistiken

Rangschätzer der in (2.5) definierten Größen p_{ij} lassen sich gewinnen, indem die unbekanntenen Verteilungsfunktionen $F_{ij}(x)$ durch die entsprechenden empirischen Gegenstücke $\widehat{F}_{ij} = n_i^{-1} \sum_{k=1}^{n_i} c(x - X_{ijk})$, $i = 1, \dots, a$; $j = 1, \dots, d$ ersetzt werden. Dabei ist

$$c(x) = \begin{cases} 0, & x < 0, \\ 1/2, & x = 0, \\ 1, & x > 0. \end{cases} \quad (3.1)$$

Damit lassen sich die so genannten Platzierungen (Orban und Wolfe, 1982) mithilfe von Rängen darstellen:

$$\begin{aligned} n_i \widehat{F}_{ij}(X_{ijk}) &= \sum_{k'=1}^{n_i} c(X_{ijk} - X_{ijk'}) \\ &= R_{ijk}^{(ij)} - \frac{1}{2}, \end{aligned} \quad (3.2)$$

$$\begin{aligned} n_r \widehat{F}_{rs}(X_{ijk}) &= \sum_{k'=1}^{n_r} c(X_{ijk} - X_{rsk'}) \\ &= \sum_{r'=1}^a \sum_{s'=1}^d \sum_{k'=1}^{n_r} c(X_{ijk} - X_{r'sk'}) - \sum_{\substack{r'=1 \\ (r',s') \neq (r,s)}}^a \sum_{s'=1}^d \sum_{k'=1}^{n_{r'}} c(X_{ijk} - X_{r's'k'}) \\ &= R_{ijk} - R_{ijk}^{(-rs)} \\ &= R_{ijk}^{(ij+rs)} - R_{ijk}^{(ij)}, \quad (i, j) \neq (r, s). \end{aligned} \quad (3.3)$$

Hier bezeichnet $R_{ijk} = 1/2 + \sum_{i'=1}^a \sum_{j'=1}^d \sum_{k'=1}^{n_{i'j'}} c(X_{ijk} - X_{i'j'k'})$ im Fall von Bindungen den Mittelrang von X_{ijk} unter allen Beobachtungen. Liegen keine Bindungen vor, handelt es sich um den normalen Rang. Zur Vereinfachung wird in der Folge nur noch von Rängen gesprochen. Bei $R_{ijk}^{(ij)}$ handelt es sich um den internen Rang der Beobachtung X_{ijk} unter allen n_i Beobachtungen der Kombination Gruppe i und Repeated Measurement j . Dabei ist $R_{ijk}^{(-rs)}$ der Rang von X_{ijk} unter allen Beobachtungen ausgenommen der n_r Messwerte der Kombination Gruppe r und Repeated Measurement s . Die Notation $R_{ijk}^{(ij+rs)}$ meint den Rang von X_{ijk} unter den $n_i + n_r$ Beobachtungen der Kombinationen (i, j) und (r, s) . Nun lassen sich die relativen Effekte p_{ij} aus (2.5) mittels der Platzierungen aus (3.3) schätzen:

$$\widehat{p}_{ij} = \int \widehat{H} d\widehat{F}_{ij} = \frac{1}{ad} \sum_{r=1}^a \sum_{s=1}^d \int \widehat{F}_{rs} d\widehat{F}_{ij}. \quad (3.4)$$

3. Asymptotische Normalität der linearen Rangstatistiken

Man benötigt also

$$\begin{aligned}\widehat{w}_{rsij} &= \int \widehat{F}_{rs} d\widehat{F}_{ij} = \frac{1}{n_i} \sum_{k=1}^{n_i} \widehat{F}_{rs}(X_{ijk}) \\ &\stackrel{(3.3)}{=} \frac{1}{n_i n_r} \sum_{k=1}^{n_i} R_{ijk}^{(ij+rs)} - R_{ijk}^{(ij)} = \frac{1}{n_r} \left(\overline{R}_{ij\cdot}^{(ij+rs)} - \frac{n_i + 1}{2} \right).\end{aligned}\tag{3.5}$$

$$\tag{3.6}$$

Dabei ist $\overline{R}_{ij\cdot}^{(ij+rs)} = n_i^{-1} \sum_{k=1}^{n_i} R_{ijk}^{(ij+rs)}$ der Mittelwert der Ränge in Kombination (i, j) und $n_i^{-1} \sum_{k=1}^{n_i} R_{ijk}^{(ij)} = (n_i + 1)/2$. Damit ergibt sich über (3.4)

$$\widehat{p}_{ij} = \frac{1}{ad} \sum_{r=1}^a \sum_{s=1}^d \widehat{w}_{rsij} = \frac{1}{ad} \sum_{r=1}^a \sum_{s=1}^d \frac{1}{n_r} \left(\overline{R}_{ij\cdot}^{(ij+rs)} - \frac{n_i + 1}{2} \right).\tag{3.7}$$

Sei $\widehat{\mathbf{p}} = (\widehat{p}_{11}, \widehat{p}_{12}, \dots, \widehat{p}_{ad})'$ der Vektor dieser Schätzer. Beachte, dass die \widehat{p}_{ij} asymptotisch unverzerrte und konsistente Schätzer der p_{ij} sind, wie das folgende Lemma zeigt.

Lemma 3.1. *Sei $\widehat{\mathbf{p}} = (\widehat{p}_{11}, \widehat{p}_{12}, \dots, \widehat{p}_{ad})'$ und \widehat{p}_{ij} wie in (3.4). Sei zudem der kleinste Stichprobenumfang bezeichnet mit $n_0 = \min_{i=1, \dots, a} \{n_i\}$. Dann gilt*

1. $\mathbb{E}(\widehat{p}_{ij}) = p_{ij} + O(\frac{1}{n_i}), i = 1, \dots, a,$
2. $\mathbb{E}(\widehat{p}_{ij} - p_{ij})^2 = O(\frac{1}{n_0}), i = 1, \dots, a; j = 1, \dots, d.$

Der Schätzer \widehat{p}_{ij} ist also konsistent und asymptotisch unverzerrt, wenn die kleinste Stichprobe groß genug ist, d.h. $\min_{i=1, \dots, a} \{n_i\} \rightarrow \infty$. Somit ist auch $\widehat{\mathbf{p}}$ asymptotisch unverzerrt und konsistent für \mathbf{p} .

Beweis. Vgl. Brunner und Langer (2002).

Um zunächst die erste der beiden Aussagen zu zeigen, definiere $H^{(i)} = \frac{1}{ad} \sum_{j'=1}^d F_{ij'}$ und damit $p_j^{(i)} = \int H^{(i)} dF_{ij} = \frac{1}{ad} \sum_{j'=1}^d \int F_{ij'} dF_{ij}$. Sei zudem $\Delta_j^{(i)} = \frac{1}{ad} \sum_{j'=1}^d \Delta_{j'j}^{(i)}$, wobei $\Delta_{j'j}^{(i)} = \mathbb{P}(X_{ij'k} < X_{ijk}) + \frac{1}{2} \mathbb{P}(X_{ij'k} = X_{ijk}) = \mathbb{E}[c(X_{ijk} - X_{ij'k})]$. Dann ergibt sich für den Erwartungswert des Schätzers \widehat{p}_{ij}

$$\begin{aligned}\mathbb{E}(\widehat{p}_{ij}) &= \mathbb{E}\left(\int \widehat{H} d\widehat{F}_{ij}\right) \\ &= \mathbb{E}\left(\frac{1}{n_i} \sum_{k=1}^{n_i} \widehat{H}(X_{ijk})\right) \\ &= \mathbb{E}\left(\frac{1}{n_i} \sum_{k=1}^{n_i} \frac{1}{ad} \sum_{i'=1}^a \sum_{j'=1}^d \frac{1}{n_{i'}} \sum_{k'=1}^{n_{i'}} c(X_{ijk} - X_{i'j'k'})\right)\end{aligned}$$

3. Asymptotische Normalität der linearen Rangstatistiken

$$\begin{aligned}
&= \frac{1}{ad} \frac{1}{n_i} \sum_{k=1}^{n_i} \sum_{\substack{a \\ i'=1 \\ j'=1 \\ i' \neq i}} \sum_{\substack{d \\ k'=1 \\ j'=1}} \frac{1}{n_{i'}} \sum_{k'=1}^{n_{i'}} \mathbb{E}[c(X_{ijk} - X_{i'j'k'})] && (i, k) \neq (i', k'), i \neq i' \\
&\quad \underbrace{= p_{ij} - \frac{1}{ad} \frac{1}{n_i} \sum_{k=1}^{n_i} \frac{1}{n_i} \sum_{k=1}^{n_i} \sum_{j'=1}^d \int F_{ij'} dF_{ij}} \\
&+ \frac{1}{ad} \frac{1}{n_i} \sum_{k=1}^{n_i} \frac{1}{n_i} \sum_{\substack{k'=1 \\ k' \neq k}}^{n_i} \sum_{j'=1}^d \mathbb{E}[c(X_{ijk} - X_{i'j'k'})] && (i, k) \neq (i', k'), i = i' \\
&\quad \underbrace{ad \cdot p_j^{(i)}} \\
&+ \frac{1}{ad} \frac{1}{n_i} \sum_{k=1}^{n_i} \frac{1}{n_i} \sum_{\substack{k'=1 \\ k'=k}}^{n_i} \sum_{j'=1}^d \mathbb{E}[c(X_{ijk} - X_{i'j'k'})] && (i, k) = (i', k') \\
&\quad \underbrace{ad \cdot \Delta_j^{(i)}} \\
&= p_{ij} - p_j^{(i)} + \frac{n_i - 1}{n_i} p_j^{(i)} + \frac{1}{n_i} \Delta_j^{(i)} \\
&= p_{ij} + \frac{1}{n_i} \left(\Delta_j^{(i)} - p_j^{(i)} \right). \tag{3.8}
\end{aligned}$$

Da aber $|\Delta_j^{(i)}| \leq 1$ und auch $|p_j^{(i)}| \leq 1$, konvergiert die Verzerrung $\frac{1}{n_i} (\Delta_j^{(i)} - p_j^{(i)})$ für $n_i \rightarrow \infty$ gegen 0 und \hat{p}_{ij} ist asymptotisch unverzerrt. Somit ist $\hat{\boldsymbol{p}}$ asymptotisch unverzerrt, falls $n_0 = \min_{i=1, \dots, a} \{n_i\} \rightarrow \infty$.

Der zweite Teil des Lemmas benötigt eine kleine Vorbereitung: Es ist

$$\mathbb{E}[\widehat{F}_{ij}(X_{i'j'k}) - F_{ij}(X_{i'j'k})]^2 \leq \frac{1}{n_i}. \tag{3.9}$$

Um dies zu zeigen (vgl. auch [Brunner und Munzel \(2002\)](#)), sei $G(x, y, z)$ die gemeinsame Verteilung von $(X_{i'j'k}, X_{ijl}, X_{ijl'})'$ sowie $G_1(x, y)$ die gemeinsame Verteilung von $(X_{i'j'k}, X_{ijl})'$ und $G_2(x, z)$ die gemeinsame Verteilung von $(X_{i'j'k}, X_{ijl'})'$. Für $l \neq l'$ folgt, dass entweder $X_{ijl'}$ von $X_{i'j'k}$ und X_{ijl} unabhängig ist oder dass X_{ijl} von $X_{i'j'k}$ und $X_{ijl'}$ unabhängig ist. Damit gilt für $l \neq l'$ entweder $G(x, y, z) = G_1(x, y) \cdot F_{ij}(z)$ oder $G(x, y, z) = G_2(x, z) \cdot F_{ij}(y)$. Falls X_{ijl} von $X_{i'j'k}$ und $X_{ijl'}$ unabhängig ist, erhält man für den Erwartungswert unter Anwendung des Satzes von Fubini

$$\begin{aligned}
&\mathbb{E}([c(X_{i'j'k} - X_{ijl}) - F_{ij}(X_{i'j'k})][c(X_{i'j'k} - X_{ijl'}) - F_{ij}(X_{i'j'k})]) \tag{3.10} \\
&= \int \int \int [c(x - y) - F_{ij}(x)][c(x - z) - F_{ij}(x)] dG(x, y, z) \\
&= \int \int [c(x - z) - F_{ij}(x)] \int [c(x - y) - F_{ij}(x)] dF_{ij}(y) dG_2(x, z) = 0,
\end{aligned}$$

da $\int [c(x - y) - F_{ij}(x)] dF_{ij}(y) = F_{ij}(x) - F_{ij}(x) = 0$ ist. Analoges lässt sich zeigen, falls

3. Asymptotische Normalität der linearen Rangstatistiken

$X_{ijl'}$ von $X_{i'j'k}$ und X_{ijl} unabhängig ist. Damit folgt

$$\begin{aligned}
& \mathbb{E} \left[\widehat{F}_{ij}(X_{i'j'k}) - F_{ij}(X_{i'j'k}) \right]^2 \\
&= \frac{1}{n_i^2} \sum_{l=1}^{n_i} \sum_{l'=1}^{n_i} \mathbb{E} \left([c(X_{i'j'k} - X_{ijl}) - F_{ij}(X_{i'j'k})] [c(X_{i'j'k} - X_{ijl'}) - F_{ij}(X_{i'j'k})] \right) \\
&= \frac{1}{n_i} \sum_{l=1}^{n_i} \mathbb{E} [c(X_{i'j'k} - X_{ijl}) - F_{ij}(X_{i'j'k})]^2 \leq \frac{1}{n_i}. \tag{3.11}
\end{aligned}$$

Aus diesem Resultat kann man nun folgende Abschätzung gewinnen:

$$\begin{aligned}
\mathbb{E} \left(\widehat{H}(X_{i'j'k}) - H(X_{i'j'k}) \right)^2 &= \mathbb{E} \left(\frac{1}{ad} \sum_{i=1}^a \sum_{j=1}^d [\widehat{F}_{ij}(X_{i'j'k}) - F_{ij}(X_{i'j'k})] \right)^2 \\
&\stackrel{\text{Jensen}}{\leq} \frac{1}{ad} \sum_{i=1}^a \sum_{j=1}^d \mathbb{E} \left[\widehat{F}_{ij}(X_{i'j'k}) - F_{ij}(X_{i'j'k}) \right]^2 \\
&\stackrel{(3.11)}{\leq} \frac{1}{a} \sum_{i=1}^a \frac{1}{n_i} \leq \frac{1}{a} \sum_{i=1}^a \frac{1}{n_0} = \frac{1}{n_0}. \tag{3.12}
\end{aligned}$$

Nun lässt sich die eigentliche Aussage zeigen. Betrachte

$$\begin{aligned}
(\widehat{p}_{ij} - p_{ij})^2 &= \left(\int \widehat{H} d\widehat{F}_{ij} - \int H dF_{ij} \right)^2 = \left(\int [\widehat{H} - H] d\widehat{F}_{ij} + \int H d[\widehat{F}_{ij} - F_{ij}] \right)^2 \\
&= \left(\int [\widehat{H} - H] d\widehat{F}_{ij} + \frac{1}{n_i} \sum_{k=1}^{n_i} [H(X_{ijk}) - \int H dF_{ij}] \right)^2.
\end{aligned}$$

Bei der Berechnung des Erwartungswertes dieses Ausdrucks lässt sich nun zunächst die c_r -Ungleichung (mit $r = 2$) anwenden und weiter mit der Jensen-Ungleichung abschätzen:

$$\begin{aligned}
\mathbb{E}(\widehat{p}_{ij} - p_{ij})^2 &\leq 2 \cdot \mathbb{E} \left(\frac{1}{n_i} \sum_{k=1}^{n_i} [\widehat{H}(X_{ijk}) - H(X_{ijk})] \right)^2 + 2 \cdot \mathbb{E} \left(\frac{1}{n_i} \sum_{k=1}^{n_i} [H(X_{ijk}) - \int H dF_{ij}] \right)^2 \\
&\leq \frac{2}{n_i} \sum_{k=1}^{n_i} \overbrace{\mathbb{E}[\widehat{H}(X_{ijk}) - H(X_{ijk})]^2}^{\leq \frac{1}{n_0} \text{ vgl. (3.12)}} \\
&\quad + \frac{2}{n_i^2} \sum_{k=1}^{n_i} \sum_{k'=1}^{n_i} \underbrace{\mathbb{E}[(H(X_{ijk}) - \int H dF_{ij})(H(X_{ijk'}) - \int H dF_{ij})]}_{\text{unabhängig für } k \neq k'} \\
&= \frac{2}{n_0} + \frac{2}{n_i^2} \sum_{k=1}^{n_i} \underbrace{\mathbb{E}[H(X_{ijk}) - \int H dF_{ij}]^2}_{\leq 1} \\
&\leq \frac{2}{n_0} + \frac{2}{n_i} \leq \frac{4}{n_0}. \tag{3.13}
\end{aligned}$$

3. Asymptotische Normalität der linearen Rangstatistiken

Dies zeigt also die Konsistenz von \hat{p}_{ij} bzgl. p_{ij} für $n_0 = \min_{i=1,\dots,a}\{n_i\} \rightarrow \infty$ und überträgt sich wiederum auf $\hat{\mathbf{p}} = (\hat{p}_{11}, \hat{p}_{12}, \dots, \hat{p}_{ad})'$ für $\mathbf{p} = (p_{11}, p_{12}, \dots, p_{ad})'$. \square

Der folgende Satz beinhaltet die grundlegende asymptotische Äquivalenz, die es dann ermöglichen wird, die asymptotische Kovarianzmatrix in einer einfachen Form darzustellen:

Satz 3.2. (Asymptotischer Äquivalenzsatz) Sei $\mathbf{F} = (F_{11}, F_{12}, \dots, F_{ad})'$ der Vektor der Verteilungsfunktionen, $\hat{\mathbf{F}} = (\hat{F}_{11}, \hat{F}_{12}, \dots, \hat{F}_{ad})'$ der Vektor der empirischen Verteilungsfunktionen, außerdem $H = \frac{1}{ad} \sum_{r=1}^a \sum_{s=1}^d F_{rs}$ eine mittlere Verteilungsfunktion und \hat{H} das entsprechende empirische Gegenstück. Unter der Annahme $\sum_{i=1}^a n_i = N \rightarrow \infty$, so dass $N/n_i \leq N_0 < \infty$, für alle $i = 1, \dots, a$, gilt

$$\sqrt{N} \int \hat{H} d(\hat{\mathbf{F}} - \mathbf{F}) \doteq \sqrt{N} \int H d(\hat{\mathbf{F}} - \mathbf{F}).$$

Dabei bedeutet das Zeichen \doteq , dass die beiden Folgen von Zufallsvariablen auf der linken und rechten Seite asymptotisch äquivalent sind.

Beweis. Es ist ausreichend, den Beweis für die ij -te Komponente durchzuführen. Schreibe dazu

$$\sqrt{N} \int \hat{H} d(\hat{F}_{ij} - F_{ij}) = \sqrt{N} \int H d(\hat{F}_{ij} - F_{ij}) + \sqrt{N} \int (\hat{H} - H) d(\hat{F}_{ij} - F_{ij}). \quad (3.14)$$

Es muss also nur gezeigt werden, dass gilt:

$$\mathbb{E} \left(\sqrt{N} \int (\hat{H} - H) d(\hat{F}_{ij} - F_{ij}) \right)^2 \rightarrow 0. \quad (3.15)$$

Betrachte hierfür

$$\begin{aligned} B_{ij} &:= \int (\hat{H} - H) d(\hat{F}_{ij} - F_{ij}) \\ &= \frac{1}{n_i} \sum_{k=1}^{n_i} \left[(\hat{H}(X_{ijk}) - H(X_{ijk})) - \int (\hat{H}(x) - H(x)) dF_{ij} \right]. \end{aligned}$$

Für den ersten Term der Summe ergibt sich

$$\begin{aligned} (\hat{H}(X_{ijk}) - H(X_{ijk})) &= \frac{1}{ad} \sum_{r=1}^a \sum_{s=1}^d (\hat{F}_{rs}(X_{ijk}) - F_{rs}(X_{ijk})) \\ &= \frac{1}{ad} \sum_{r=1}^a \sum_{s=1}^d \frac{1}{n_r} \sum_{l=1}^{n_r} c(X_{ijk} - X_{rsl}) - F_{rs}(X_{ijk}). \end{aligned}$$

Der zweite Term der Summe liefert

$$\begin{aligned} \int (\hat{H}(x) - H(x)) dF_{ij} &= \int \left[\frac{1}{ad} \sum_{r=1}^a \sum_{s=1}^d \left(\frac{1}{n_r} \sum_{l=1}^{n_r} c(x - X_{rsl}) - F_{rs}(x) \right) \right] dF_{ij} \\ &= \frac{1}{ad} \sum_{r=1}^a \sum_{s=1}^d \frac{1}{n_r} \sum_{l=1}^{n_r} \int [c(x - X_{rsl}) - F_{rs}(x)] dF_{ij}. \end{aligned}$$

3. Asymptotische Normalität der linearen Rangstatistiken

Definiere nun, um die Notation abzukürzen,

$$\varphi(X_{ijk}, X_{rsl}) := c(X_{ijk} - X_{rsl}) - F_{rs}(X_{ijk}) - \int [c(x - X_{rsl}) - F_{rs}(x)] dF_{ij}.$$

Damit lässt sich zusammenfassen:

$$B_{ij} = \frac{1}{ad} \frac{1}{n_i} \sum_{k=1}^{n_i} \sum_{r=1}^a \sum_{s=1}^d \frac{1}{n_r} \sum_{l=1}^{n_r} \varphi(X_{ijk}, X_{rsl}).$$

Zu untersuchen ist somit

$$\begin{aligned} & \mathbb{E} \left(\sqrt{N} \int (\widehat{H} - H) d(\widehat{F}_{ij} - F_{ij}) \right)^2 = \mathbb{E}(\sqrt{N} B_{ij})^2 \\ &= \frac{N}{n_i^2 (ad)^2} \sum_{k=1}^{n_i} \sum_{k'=1}^{n_i} \sum_{r=1}^a \sum_{r'=1}^a \sum_{s=1}^d \sum_{s'=1}^d \frac{1}{n_r n_{r'}} \sum_{l=1}^{n_r} \sum_{l'=1}^{n_{r'}} \mathbb{E} \left[\varphi(X_{ijk}, X_{rsl}) \varphi(X_{ijk'}, X_{r's'l'}) \right] \end{aligned}$$

Unterscheidet sich nun einer der drei Indizes i, r, r' von den anderen beiden, dann folgt auf Grund der Unabhängigkeit mit den gleichen Argumenten wie im Beweis von Lemma 3.1, Gleichung (3.10), dass $\mathbb{E}[\varphi(X_{ijk}, X_{rsl}) \varphi(X_{ijk'}, X_{r's'l'})] = 0$. Damit hat man als erste Abschätzung

$$\mathbb{E}(\sqrt{N} B_{ij})^2 \leq \frac{N}{n_i^4 (ad)^2} \sum_{k=1}^{n_i} \sum_{k'=1}^{n_i} \sum_{s=1}^d \sum_{s'=1}^d \sum_{l=1}^{n_i} \sum_{l'=1}^{n_i} \mathbb{E} \left[\varphi(X_{ijk}, X_{isl}) \varphi(X_{ijk'}, X_{is'l'}) \right].$$

Genauso lässt sich weiter abschätzen, indem man feststellt, dass, falls sich einer der Indizes k, l, k', l' von allen anderen drei Indizes unterscheidet, ebenso wegen der Unabhängigkeit $\mathbb{E}[\varphi(X_{ijk}, X_{isl}) \varphi(X_{ijk'}, X_{is'l'})] = 0$ gilt. Man hat also

$$\mathbb{E}(\sqrt{N} B_{ij})^2 \leq \frac{N}{n_i^4 (ad)^2} \sum_{k=1}^{n_i} \sum_{s=1}^d \sum_{s'=1}^d \sum_{l=1}^{n_i} \mathbb{E} \left[\varphi(X_{ijk}, X_{isl}) \varphi(X_{ijk}, X_{is'l}) \right].$$

Vergleiche zu diesen Abschätzungen auch [Akritas und Brunner \(1997\)](#), Beweis von Theorem 2.2. Dort wird der Beweis zwar mit einer gewichteten mittleren Verteilungsfunktion geführt, der Term $\mathbb{E}[\varphi(X_{ijk}, X_{rsl}) \varphi(X_{ijk'}, X_{r's'l'})]$ ist allerdings unabhängig von der Wahl der mittleren Verteilungsfunktion.

Schätzt man nun noch $\mathbb{E}[\varphi(X_{ijk}, X_{isl}) \varphi(X_{ijk}, X_{is'l})] \leq 1$ mit Eins ab, erhält man

$$\mathbb{E}(\sqrt{N} B_{ij})^2 \leq \frac{N}{n_i^4 (ad)^2} n_i d^2 n_i = \frac{N}{n_i^2 a^2} \stackrel{\text{Vor.}}{\leq} \frac{N_0}{a^2 n_i} \xrightarrow{n_i \rightarrow \infty} 0.$$

Es sind also $\sqrt{N} \int \widehat{H} d(\widehat{\mathbf{F}} - \mathbf{F})$ und $\sqrt{N} \int H d(\widehat{\mathbf{F}} - \mathbf{F})$ asymptotisch äquivalent. \square

An dieser Stelle sollte man bemerken, dass Satz 3.2 unter $H_0^F : \mathbf{C}\mathbf{F} = \mathbf{0}$ zu einer einfacheren asymptotischen Äquivalenz für $\sqrt{N}\mathbf{C}(\widehat{\mathbf{p}} - \mathbf{p})$ führt

$$\sqrt{N}\mathbf{C}(\widehat{\mathbf{p}} - \mathbf{p}) \stackrel{H_0^F}{\doteq} \sqrt{N}\mathbf{C} \int H d\widehat{\mathbf{F}},$$

3. Asymptotische Normalität der linearen Rangstatistiken

da in diesem Fall $\mathbf{C} \int \mathbf{F} d\widehat{\mathbf{H}} = \int \mathbf{C} \mathbf{F} d\widehat{\mathbf{H}} = \mathbf{0}$. Damit vereinfacht sich die asymptotische Kovarianzmatrix (vgl. Akritas et al. (1997), Gao et al. (2008) und Brunner und Puri (2001) für entsprechende Schätzer und verschiedene Globaltests z.B. im One-Way Layout). In dieser Arbeit soll jedoch, wie in Abschnitt 2.2 diskutiert, eine multiple Testprozedur unter H_0^p hergeleitet werden. Man fährt wie folgt fort, um die asymptotische Kovarianzmatrix von $\sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p})$ auszurechnen:

Eine leichte Umformung in Satz 3.2 liefert

$$\sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p}) \doteq \int H d\widehat{\mathbf{F}} + \int \widehat{\mathbf{H}} d\mathbf{F} - 2 \int H d\mathbf{F} \quad (3.16)$$

Somit erhält man für die ij -te Komponente

$$\sqrt{N}(\widehat{p}_{ij} - p_{ij}) \doteq \int H d\widehat{F}_{ij} + \int \widehat{H} dF_{ij} - 2 \int H dF_{ij} \quad (3.17)$$

$$\frac{\sqrt{N}}{ad} \sum_{r=1}^a \sum_{s=1}^d (\widehat{w}_{rsij} - w_{rsij}) \doteq \frac{\sqrt{N}}{ad} \sum_{r=1}^a \sum_{s=1}^d \left(\int F_{rs} d\widehat{F}_{ij} + \int \widehat{F}_{rs} dF_{ij} - 2 \int F_{rs} dF_{ij} \right).$$

Also betrachte

$$\begin{aligned} \sqrt{N}(\widehat{w}_{rsij} - w_{rsij}) &\doteq \sqrt{N} \left(\int F_{rs} d\widehat{F}_{ij} + \int \widehat{F}_{rs} dF_{ij} - 2 \int F_{rs} dF_{ij} \right) \\ &= \sqrt{N} \left(\frac{1}{n_i} \sum_{k=1}^{n_i} [F_{rs}(X_{ijk}) - w_{rsij}] - \frac{1}{n_r} \sum_{k=1}^{n_r} [F_{ij}(X_{rsk}) - w_{ijrs}] \right) \\ &= \sqrt{N} \left(\frac{1}{n_i} \sum_{k=1}^{n_i} Y_{rsijk} - \frac{1}{n_r} \sum_{k=1}^{n_r} Y_{ijrsk} + 1 - 2w_{rsij} \right) \\ &= \sqrt{N} Z_{rsij}. \end{aligned} \quad (3.18)$$

Dabei sind also Z_{rsij} Summen von unabhängigen Zufallsvariablen $Y_{rsijk} = F_{rs}(X_{ijk})$ und $Y_{ijrsk} = F_{ij}(X_{rsk})$. Die nicht beobachtbaren Zufallsvariablen Z_{rsij} werden zusammengefasst in den Vektoren $\mathbf{Z}_{ij} = (Z_{11ij}, Z_{12ij}, \dots, Z_{adij})'$ und diese wiederum in $\mathbf{Z} = (\mathbf{Z}'_{11}, \mathbf{Z}'_{12}, \dots, \mathbf{Z}'_{ad})' \in \mathbb{R}^{(ad)^2}$.

Definiere

$$\mathbf{w}_{rs} = (w_{rs11}, w_{rs12}, \dots, w_{rsad})' = \int F_{rs} d\mathbf{F} \quad (3.19)$$

$$\mathbf{W} = (\mathbf{w}_{11} : \mathbf{w}_{12} : \dots : \mathbf{w}_{ad}) \in \mathbb{R}^{ad \times ad} \quad (3.20)$$

und entsprechend für die Schätzer

$$\widehat{\mathbf{w}}_{rs} = (\widehat{w}_{rs11}, \widehat{w}_{rs12}, \dots, \widehat{w}_{rsad})' = \int \widehat{F}_{rs} d\widehat{\mathbf{F}} \quad (3.21)$$

$$\widehat{\mathbf{W}} = (\widehat{\mathbf{w}}_{11} : \widehat{\mathbf{w}}_{12} : \dots : \widehat{\mathbf{w}}_{ad}) \in \mathbb{R}^{ad \times ad}. \quad (3.22)$$

3. Asymptotische Normalität der linearen Rangstatistiken

Mit der Matrix $\mathbf{E}_{ad} = \frac{1}{ad} \mathbf{1}'_{ad} \otimes \mathbf{I}_{ad}$ lassen sich der Vektor der relativen Effekte sowie sein Schätzer mittels \mathbf{W} bzw. $\widehat{\mathbf{W}}$ gemäß ihrer Darstellung in (2.5) bzw. (3.7) schreiben als

$$\mathbf{p} = \mathbf{E}_{ad} \cdot \text{vec}(\mathbf{W}) \quad (3.23)$$

$$\widehat{\mathbf{p}} = \mathbf{E}_{ad} \cdot \text{vec}(\widehat{\mathbf{W}}) \quad (3.24)$$

Dabei bezeichnet $\text{vec}(\cdot)$ den Vektor-Operator einer Matrix, der aus den Spalten einer Matrix einen einzigen Vektor erstellt ($\text{vec}(\mathbf{W}) = (\mathbf{w}'_{11}, \mathbf{w}'_{12}, \dots, \mathbf{w}'_{ad})'$). Es ist also

$$\sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p}) = \sqrt{N} \mathbf{E}_{ad} \cdot \text{vec}(\widehat{\mathbf{W}} - \mathbf{W}) \quad (3.25)$$

und die in (3.18) gegebene asymptotische Äquivalenz führt zu

$$\sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p}) \doteq \sqrt{N} \mathbf{E}_{ad} \mathbf{Z}, \quad (3.26)$$

sodass die asymptotische Kovarianzmatrix von $\sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p})$ gegeben ist durch

$$\mathbf{V}_N = \text{Cov}(\sqrt{N} \mathbf{E}_{ad} \mathbf{Z}) = \mathbf{E}_{ad} \boldsymbol{\Sigma} \mathbf{E}'_{ad}, \quad (3.27)$$

wobei

$$\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_{rs,ij})_{r,i=1,\dots,a}^{s,j=1,\dots,d} = \text{Cov}(\sqrt{N} \mathbf{Z}) \in \mathbb{R}^{(ad)^2 \times (ad)^2}. \quad (3.28)$$

Hier ist

$$\boldsymbol{\Sigma}_{rs,rs} = \text{Cov}(\sqrt{N} \mathbf{Z}_{rs}) = [\sigma_{rs}(p, q, p', q')]_{p,p'=1,\dots,a}^{q,q'=1,\dots,d} \in \mathbb{R}^{ad \times ad} \quad (3.29)$$

$$\boldsymbol{\Sigma}_{rs,ij} = \text{Cov}(\sqrt{N} \mathbf{Z}_{rs}, \sqrt{N} \mathbf{Z}_{ij}) = [\sigma_{rs,ij}(p, q, p', q')]_{p,p'=1,\dots,a}^{q,q'=1,\dots,d} \in \mathbb{R}^{ad \times ad} \quad (3.30)$$

und

$$\sigma_{rs}(p, q, p', q') = N \text{Cov}(Z_{pqrs}, Z_{p'q'rs}) \quad (3.31)$$

$$\sigma_{rs,ij}(p, q, p', q') = N \text{Cov}(Z_{pqrs}, Z_{p'q'ij}), \quad (r, s) \neq (i, j). \quad (3.32)$$

Man beachte, dass $\mathbb{E}(\mathbf{Z}) = \mathbf{0}$, $Z_{rsrs} = 0$ und $Z_{rsij} = -Z_{ijrs}$. Unter Berücksichtigung, dass X_{ijk} und $X_{i'j'k'}$ unabhängig sind für $i \neq i'$ oder $k \neq k'$, erhält man

$$\frac{\sigma_{rs}(p, q, p', q')}{N} = \begin{cases} \tau_r^{(s,s)}(p, q, p', q'), & r \neq p, p' \wedge p \neq p', \\ \tau_r^{(s,s)}(p, q, p, q') + \tau_r^{(q,q')}(r, s, r, s), & r \neq p, p' \wedge p = p', \\ \tau_r^{(s,s)}(r, q, p', q') - \tau_r^{(q,s)}(r, s, p', q'), & r = p \wedge p \neq p' \wedge q \neq s, \\ \tau_r^{(s,s)}(p, q, r, q') - \tau_r^{(s,q')}(p, q, r, s), & r = p' \wedge p \neq p' \wedge q' \neq s, \\ \tau_r^{(s,s)}(p, q, r, q') - \tau_r^{(s,q')}(r, q, r, s) \\ - \tau_r^{(q,s)}(r, s, r, q') + \tau_r^{(q,q')}(r, s, r, s), & r = p = p' \wedge q \neq s \wedge q' \neq s, \\ 0, & (r = p \wedge q = s) \vee (r = p' \wedge q' = s), \end{cases} \quad (3.33)$$

3. Asymptotische Normalität der linearen Rangstatistiken

sowie für $(r, s) \neq (i, j)$

$$\frac{\sigma_{rs,ij}(p, q, p', q')}{N} = \begin{cases} \tau_r^{(s,j)}(p, q, p', q'), & r = i \wedge p \neq i, p' \wedge r \neq p', \\ -\tau_r^{(s,q')}(p, q, i, j), & r = p' \wedge p \neq i, p' \wedge r \neq i, \\ -\tau_p^{(q,j)}(r, s, p', q'), & p = i \wedge r \neq i, p' \wedge p \neq p', \\ \tau_p^{(q,q')}(r, s, i, j), & p = p' \wedge r \neq i, p' \wedge p \neq i, \\ \tau_r^{(s,j)}(p, q, r, q') - \tau_r^{(s,q')}(p, q, r, j), & r = i = p' \wedge p \neq i, p' \wedge q' \neq j, \\ -\tau_p^{(q,j)}(r, s, p, q') + \tau_p^{(q,q')}(r, s, i, j), & p = i = p' \wedge r \neq i, p' \wedge q' \neq j, \\ \tau_r^{(s,j)}(r, q, p', q') - \tau_r^{(q,j)}(r, s, p', q'), & r = i = p \wedge p' \neq i, p \wedge q \neq s, \\ -\tau_r^{(s,q')}(r, q, i, j) + \tau_r^{(q,q')}(r, s, i, j), & p = r = p' \wedge i \neq r, p \wedge q \neq s, \\ \tau_r^{(s,j)}(p, q, p, q') + \tau_p^{(q,q')}(r, s, r, j), & r = i \wedge p = p' \wedge r \neq p' \wedge p \neq i, \\ -\tau_r^{(s,q')}(p, q, p, j) - \tau_p^{(q,j)}(r, s, r, q'), & r = p' \wedge p = i \wedge r \neq i \wedge p \neq p', \\ \tau_r^{(s,j)}(r, q, r, q') - \tau_r^{(s,q')}(r, q, r, j) & r = p = p' = i \wedge q \neq s \wedge q' \neq j \\ -\tau_r^{(q,j)}(r, s, r, q') + \tau_r^{(q,q')}(r, s, r, j) & \wedge q \neq j \wedge s \neq q' \wedge q \neq q' \end{cases}, \\ 0, & (p, q) = (r, s) \vee (p', q') = (i, j) \\ & \vee (p, r \neq p' \wedge p, r \neq i) \end{cases} \quad (3.34)$$

Dabei sind

$$\tau_r^{(s,j)}(p, q, p', q') = \frac{1}{n_r} \mathbb{E}[(F_{pq}(X_{rs1}) - w_{pqrs})(F_{p'q'}(X_{rj1}) - w_{p'q'rj})] \quad (3.35)$$

unbekannt und müssen geschätzt werden. Es sei an dieser Stelle angemerkt, dass lediglich die Darstellung der p_{ij} als Linearkombination der w_{rsij} die einfache Darstellung der Struktur von \mathbf{V}_N wie in (3.27) gesehen ermöglicht. Die asymptotische Normalität der linearen Rangstatistik $\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p})$ wird im folgenden Satz angegeben.

Satz 3.3. (Asymptotische Normalität) *Sei \mathbf{V}_N wie in (3.27). Gilt für $N \rightarrow \infty$, dass $N/n_i \leq N_0 < \infty$, $\forall i = 1, \dots, a$, und gilt für $\mathbf{V}_N \rightarrow \mathbf{V}$, dass $\text{rk}(\mathbf{V}_N) = \text{rk}(\mathbf{V})$, für alle $N \geq M_0 < \infty$, dann besitzt die lineare Rangstatistik $\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p})$, asymptotisch für $N \rightarrow \infty$, eine multivariate Normalverteilung mit Erwartungswert $\mathbf{0}$ und Kovarianzmatrix \mathbf{V}_N .*

Beweis. Der Beweis folgt im Wesentlichen jenem für Theorem 2 in [Konietschke et al. \(2012\)](#). Auch hier wird aufgeteilt in einen regulären und einen singulären Teil der Verteilung. Ersterer wird mit der Cramer-Wold Technik bewiesen, letzterer als multivariate Ein-Punkt-Normalverteilung aufgefasst. Seien also $\lambda_{i,N}, i = 1, \dots, ad$, die Eigenwerte von \mathbf{V}_N und sei

$$\lambda_N^* = \min_{i=1, \dots, ad} \{\lambda_{i,N} | \lambda_i > 0\}$$

der kleinste Eigenwert größer als Null. Da laut Voraussetzung $\mathbf{V}_N \rightarrow \mathbf{V}$ für $N \rightarrow \infty$ so, dass $\text{rk}(\mathbf{V}_N) = \text{rk}(\mathbf{V})$, gibt es eine Konstante $c_0 > 0$ und $M_0 < \infty$, sodass $\lambda_N^* \geq c_0 \forall N \geq M_0$. Ohne Beschränkung der Allgemeinheit gelte $\lambda_{1,N}, \dots, \lambda_{j,N} \rightarrow 0$, wohingegen

3. Asymptotische Normalität der linearen Rangstatistiken

$\lambda_{j+1,N}, \dots, \lambda_{ad,N} \geq c_0 \forall N \geq M_0$ seien. Nach dem Spektralsatz zur Diagonalisierbarkeit von Matrizen gibt es nun eine invertierbare Matrix $\mathbf{\Lambda}$, sodass

$$\mathbf{\Lambda} \mathbf{V}_N \mathbf{\Lambda}' = \mathbf{D} = \bigoplus_{i=1}^2 \mathbf{D}_i,$$

wobei $\mathbf{D}_1 = \text{diag}\{\lambda_{1,N}, \dots, \lambda_{j,N}\}$ und $\mathbf{D}_2 = \text{diag}\{\lambda_{j+1,N}, \dots, \lambda_{ad,N}\}$. Es ist also $\mathbf{D}_1 \rightarrow \mathbf{0}$, wie oben o.B.d.A. angenommen. Ist $j = ad$, dann ist $\mathbf{V}_N = \mathbf{0}$ und die Verteilung kann als multivariate Ein-Punkt-Normalverteilung bezeichnet werden. Die asymptotische multivariate Normalität der Summen von unabhängigen Zufallsvariablen $\sqrt{N} \mathbf{E}_{ad} \mathbf{Z}$ lässt sich nun mit der Cramer-Wold Technik beweisen. Sei dazu $\mathbf{k} = (k_1, \dots, k_{ad})'$ ein beliebiger Vektor von Konstanten. Da $\mathbf{\Lambda}$ invertierbar ist und somit eine bijektive Abbildung beschreibt, existiert für jedes solche \mathbf{k} ein Vektor $\tilde{\mathbf{k}}$ mit $\mathbf{k}' = \tilde{\mathbf{k}}' \mathbf{\Lambda}$. Man verifiziert nun die Lindeberg-Bedingung für $N \mathbf{k}' \mathbf{E}_{ad} \mathbf{Z}$, um den zentralen Grenzwertsatz von Lindeberg-Feller anwenden zu können, via

$$L_N = \frac{\sqrt{N} \mathbf{k}' \mathbf{E}_{ad} \mathbf{Z}}{\sqrt{\text{Var}(\sqrt{N} \mathbf{k}' \mathbf{E}_{ad} \mathbf{Z})}} = \frac{N \mathbf{k}' \mathbf{E}_{ad} \mathbf{Z}}{\sqrt{\text{Var}(N \mathbf{k}' \mathbf{E}_{ad} \mathbf{Z})}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (3.36)$$

Dabei bedient man sich eines Korollars bezüglich der Lindeberg-Bedingung, wonach es für unabhängige und gleichmäßig beschränkte Zufallsvariablen ausreicht, dass die Summe der Varianzen divergiert, um die Bedingung zu erfüllen. Betrachte dazu

$$\begin{aligned} \text{Var}(N \mathbf{k}' \mathbf{E}_{ad} \mathbf{Z}) &= \text{Var}(\sqrt{N} \mathbf{k}' (\sqrt{N} \mathbf{E}_{ad} \mathbf{Z})) = N \mathbf{k}' \mathbf{V}_N \mathbf{k} \\ &= N \tilde{\mathbf{k}}' \mathbf{\Lambda} \mathbf{V}_N \mathbf{\Lambda}' \tilde{\mathbf{k}} = N \tilde{\mathbf{k}}' \mathbf{D} \tilde{\mathbf{k}} \geq \min_{s=j+1, \dots, ad} \{\tilde{k}_s^2\} N c_0 \rightarrow \infty. \end{aligned}$$

Dies zeigt also, dass die Summen der Varianzen von $N \mathbf{k}' \mathbf{E}_{ad} \mathbf{Z}$ für jedes beliebige $\mathbf{k} \neq \mathbf{0}$ für $N \rightarrow \infty$ divergieren. Da die Zufallsvariablen $N/n_i Y_{ijrsk}$ bzw. $N/n_r Y_{rsijk}$ gleichmäßig beschränkt sind durch die Voraussetzung $N/n_i \leq N_0 < \infty$, ist die Lindeberg-Bedingung erfüllt und (3.36) gilt. Die Cramer-Wold Technik liefert dann die multivariate Normalität von $\sqrt{N} \mathbf{E}_{ad} \mathbf{Z}$ und also $\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p})$. □

Ein Schätzer für \mathbf{V}_N wird im nächsten Abschnitt vorgestellt.

4. Schätzen der Kovarianzmatrix

Es ist ausreichend, die unbekanntenen Größen $\tau_r^{(s,j)}(p, q, p', q') = 1/n_r \text{Cov}(Y_{pqrs1}, Y_{p'q'rj})$ in (3.35) konsistent zu schätzen. Wären die Zufallsvariablen Y_{ijrsk} beobachtbar, dann ließe sich $\theta_r^{(s,j)}(p, q, p', q') = \text{Cov}(Y_{pqrs1}, Y_{p'q'rj1})$ schätzen mittels des empirischen Kovarianzschätzers

$$\tilde{\theta}_r^{(s,j)}(p, q, p', q') = \frac{1}{n_r - 1} \sum_{k=1}^{n_r} (Y_{pqrsk} - \bar{Y}_{pqrs})(Y_{p'q'rjk} - \bar{Y}_{p'q'rj}).$$

Sie sind jedoch nicht beobachtbar, sodass sie durch beobachtbare Zufallsvariablen ersetzt werden müssen, die ihnen möglichst nahe kommen. Dazu verwendet man die in (3.3) angegebenen Platzierungen sowie die Schätzer aus (3.5), um die zentrierten Platzierungen zu gewinnen. Sei also

$$\begin{aligned} D_{rsk}(p, q) &= \hat{F}_{pq}(X_{rsk}) - \hat{w}_{pqrs} = \hat{Y}_{pqrsk} - \frac{1}{n_r} \sum_{k=1}^{n_r} \hat{Y}_{pqrsk} \\ &= \frac{1}{n_p} \left((R_{rsk}^{(rs+pq)} - R_{rsk}^{(rs)}) - (\bar{R}_{rs}^{(rs+pq)} - \frac{n_r + 1}{2}) \right). \end{aligned} \quad (4.1)$$

Dann lässt sich folgender Kovarianzschätzer konstruieren

$$\hat{\theta}_r^{(s,j)}(p, q, p', q') = \frac{1}{(n_r - 1)} \sum_{k=1}^{n_r} D_{rsk}(p, q) \cdot D_{rjk}(p', q').$$

Damit ergibt sich als Rangschätzer für $\tau_r^{(s,j)}(p, q, p', q')$ in (3.35)

$$\hat{\tau}_r^{(s,j)}(p, q, p', q') = \frac{1}{n_r} \hat{\theta}_r^{(s,j)}(p, q, p', q'). \quad (4.2)$$

Ersetzt man nun die Größen $\tau_r^{(s,j)}(p, q, p', q')$ in (3.33) und (3.34) durch $\hat{\tau}_r^{(s,j)}(p, q, p', q')$ erhält man die Schätzer $\hat{\sigma}_{rs}(p, q, p', q')$ und $\hat{\sigma}_{rs,ij}(p, q, p', q')$. Der entstehende Schätzer der asymptotischen Kovarianzmatrix Σ von $\sqrt{N}\mathbf{Z}$ wird dann mit

$$\hat{\Sigma}_N = (\hat{\Sigma}_{rs,ij})_{r,i=1,\dots,a}^{s,j=1,\dots,d}$$

bezeichnet. Damit und mit der Beziehung (3.27) lässt sich dann ein Schätzer $\hat{\mathbf{V}}_N$ der asymptotischen Kovarianzmatrix von $\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p})$ konstruieren

$$\hat{\mathbf{V}}_N = \mathbf{E}_{ad} \hat{\Sigma}_N \mathbf{E}'_{ad}. \quad (4.3)$$

Der folgende Satz stellt sicher, dass es sich bei $\hat{\mathbf{V}}_N$ um einen konsistenten Schätzer für \mathbf{V}_N handelt.

Satz 4.1. (Konsistenz von $\hat{\mathbf{V}}_N$) *Sei $\hat{\mathbf{V}}_N = \mathbf{E}_{ad} \hat{\Sigma}_N \mathbf{E}'_{ad}$ der Schätzer der Kovarianzmatrix wie oben hergeleitet. Gilt $\min\{n_1, \dots, n_a\} \rightarrow \infty$, so dass $\frac{N}{n_i} \leq N_0 < \infty$, dann ist $\hat{\mathbf{V}}_N$ konsistent für \mathbf{V}_N , d.h. $\hat{\mathbf{V}}_N - \mathbf{V}_N \xrightarrow{a.s.} \mathbf{0}$.*

4. Schätzen der Kovarianzmatrix

Beweis. Vgl. auch [Konietschke et al. \(2012\)](#). Bevor zum eigentlichen Beweis übergegangen werden kann, muss zunächst eine kleine Vorbereitung getroffen werden. Sei $\|\cdot\|_\infty$ die Supremumsnorm und seien die D_{rsk} wie in (4.1) definiert. Seien weiterhin

$$\tilde{D}_{rsk}(p, q) = Y_{pqrs} - \frac{1}{n_r} \sum_{k=1}^{n_r} Y_{pqrsk}.$$

Dann gilt für $n_p \rightarrow \infty$, dass $D_{rsk}(p, q) - \tilde{D}_{rsk}(p, q) \xrightarrow{a.s.} 0$, denn

$$\begin{aligned} |D_{rsk}(p, q) - \tilde{D}_{rsk}(p, q)| &= |\hat{Y}_{pqrs} - Y_{pqrs} + \frac{1}{n_r} \sum_{k=1}^{n_r} (\hat{Y}_{pqrsk} - Y_{pqrsk})| \\ &\leq |\hat{F}_{pq}(X_{rsk}) - F_{pq}(X_{rsk})| + \frac{1}{n_r} \sum_{k=1}^{n_r} |\hat{F}_{pq}(X_{rsk}) - F_{pq}(X_{rsk})| \\ &\leq \left\| \hat{F}_{pq} - F_{pq} \right\|_\infty + \max_{k=1, \dots, n_r} |\hat{F}_{pq}(X_{rsk}) - F_{pq}(X_{rsk})| \\ &\stackrel{GC}{\leq} 2 \left\| \hat{F}_{pq} - F_{pq} \right\|_\infty \xrightarrow{a.s.} 0, \quad n_p \rightarrow \infty. \end{aligned} \quad (4.4)$$

Hier wurde im letzten Schritt der Satz von Glivenko-Cantelli verwendet.

Nun zum eigentlichen Teil, der Konsistenz von $\hat{\mathbf{V}}_N$. Da sowohl a als auch d endlich sind, genügt es die Konsistenz elementweise zu zeigen. Dazu reicht es per Konstruktion aus, nachzuweisen, dass $|\hat{\theta}_r^{(s,j)}(p, q, p', q') - \theta_r^{(s,j)}(p, q, p', q')| \xrightarrow{a.s.} 0$, falls $n_p, n_{p'} \rightarrow \infty$. Das starke Gesetz der großen Zahlen vereinfacht hier den Beweis, da für $n_p, n_{p'} \rightarrow \infty$ $|\hat{\theta}_r^{(s,j)}(p, q, p', q') - \theta_r^{(s,j)}(p, q, p', q')| \xrightarrow{a.s.} 0$ gilt. Demnach muss nur noch gezeigt werden, dass $|\hat{\theta}_r^{(s,j)}(p, q, p', q') - \tilde{\theta}_r^{(s,j)}(p, q, p', q')| \xrightarrow{a.s.} 0$, falls $n_p, n_{p'} \rightarrow \infty$.

$$\begin{aligned} &|\hat{\theta}_r^{(s,j)}(p, q, p', q') - \tilde{\theta}_r^{(s,j)}(p, q, p', q')| \\ &= \left| \frac{1}{n_r - 1} \sum_{k=1}^{n_r} [D_{rsk}(p, q) D_{rjk}(p', q') - \tilde{D}_{rsk}(p, q) \tilde{D}_{rjk}(p', q')] \right| \\ &\leq \frac{1}{n_r - 1} \sum_{k=1}^{n_r} \left| D_{rsk}(p, q) [D_{rjk}(p', q') - \tilde{D}_{rjk}(p', q')] - \tilde{D}_{rjk}(p', q') [\tilde{D}_{rsk}(p, q) - D_{rsk}(p, q)] \right| \\ &\leq \frac{1}{n_r - 1} \sum_{k=1}^{n_r} \left| D_{rjk}(p', q') - \tilde{D}_{rjk}(p', q') \right| + \frac{1}{n_r - 1} \sum_{k=1}^{n_r} \left| \tilde{D}_{rsk}(p, q) - D_{rsk}(p, q) \right| \\ &\leq \frac{n_r}{n_r - 1} \left(\max_{k=1, \dots, n_r} |D_{rjk}(p', q') - \tilde{D}_{rjk}(p', q')| + \max_{k=1, \dots, n_r} |\tilde{D}_{rsk}(p, q) - D_{rsk}(p, q)| \right) \\ &\stackrel{(4.4)}{\leq} \underbrace{\frac{2n_r}{n_r - 1}}_{\rightarrow 2 < \infty} \left(\left\| \hat{F}_{p'q'} - F_{p'q'} \right\|_\infty + \left\| \hat{F}_{pq} - F_{pq} \right\|_\infty \right) \xrightarrow{a.s.} 0, \quad n_p, n_{p'} \rightarrow \infty. \end{aligned}$$

Damit sind $\hat{\theta}_r^{(s,j)}(p, q, p', q')$ und also $\hat{\tau}_r^{(s,j)}(p, q, p', q')$ konsistent. Damit folgt auch die Konsistenz der Linearkombinationen $\hat{\sigma}_{rs}(p, q, p', q')$ und $\hat{\sigma}_{rs,ij}(p, q, p', q')$, den Elementen von $\hat{\mathbf{V}}_N$. Dies beendet den Beweis.

4. Schätzen der Kovarianzmatrix

□

Die asymptotische Verteilung von $\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p})$ und der Schätzer $\hat{\mathbf{V}}_N$ können nun verwendet werden, um Testprozeduren für multiple Vergleiche und kompatible simultane Konfidenzintervalle für $\delta_\ell = \mathbf{c}'_\ell \mathbf{p}$ herzuleiten.

5. Prozeduren für multiple Kontrasttests

Die folgenden Schritte zur Herleitung einer solchen Prozedur entsprechen jener Vorgehensweise, wie sie schon von [Konietschke et al. \(2012\)](#) für ein One-Way Layout angewandt wurde. Um eine Prozedur für multiple Kontrasttests (MCTP, *multiple contrast test procedure*) bezüglich der in (2.8) definierten Familie von Hypothesen zu konstruieren, müssen zunächst Teststatistiken für jede individuelle Hypothese $H_0^p : \mathbf{c}'_\ell \mathbf{p} = 0$ angegeben werden. Definiere dazu $\hat{v}_{\ell\ell} = \mathbf{c}'_\ell \hat{\mathbf{V}}_N \mathbf{c}_\ell$. Nun sei

$$T_\ell^p = \sqrt{N} \frac{\mathbf{c}'_\ell (\hat{\mathbf{p}} - \mathbf{p})}{\sqrt{\hat{v}_{\ell\ell}}}, \quad l = 1, \dots, q. \quad (5.1)$$

Auf Grund der asymptotischen Normalität von $\sqrt{N} \mathbf{c}'_\ell (\hat{\mathbf{p}} - \mathbf{p})$ und des Satzes von Slutsky folgt, dass $T_\ell^p \xrightarrow{d} \mathcal{N}(0, 1)$. Die Teststatistiken T_ℓ^p werden zusammengefasst im Vektor

$$\mathbf{T} = (T_1^p, \dots, T_q^p)'. \quad (5.2)$$

Man beachte, dass die MCTP und kompatible simultane Konfidenzintervalle (SCI, *simultaneous confidence intervals*) mit Hilfe der asymptotischen multivariaten Verteilung der in (5.2) definierten Teststatistik \mathbf{T} hergeleitet werden. Die asymptotische Verteilung von \mathbf{T} wird zudem für beliebige Alternativen bestimmt. Daher ist sie auch vollständig spezifiziert unter jeder beliebigen Variation der Nullhypothese.

Korollar 5.1. *Sei $v_{\ell\ell} = \mathbf{c}'_\ell \mathbf{V}_N \mathbf{c}_\ell$ und $v_{\ell m} = \mathbf{c}'_\ell \mathbf{V}_N \mathbf{c}_m$. Gelten die Annahmen aus Satz 3.3, dann besitzt die Statistik \mathbf{T} asymptotisch eine multivariate Normalverteilung mit Erwartungswert $\mathbf{0}$ und Korrelationsmatrix $\mathbf{R} = (r_{\ell m})_{\ell, m=1, \dots, q}$, wobei die Einträge durch $r_{\ell m} = v_{\ell m} / \sqrt{v_{\ell\ell} v_{m m}}$ gegeben sind.*

Beweis. Sei $\mathbf{C} = (\mathbf{c}'_\ell)_{\ell=1, \dots, q}$ die Kontrastmatrix bestehend aus den q einzelnen Kontrasten \mathbf{c}'_ℓ . Der Beweis folgt sofort aus der asymptotischen multivariaten Normalität von $\sqrt{N} \mathbf{C} (\hat{\mathbf{p}} - \mathbf{p})$ und dem Satz von Slutsky. □

Lemma 5.2. *Die Familie von Hypothesen Ω^p und die entsprechende Teststatistik \mathbf{T} bilden asymptotisch eine gemeinsame Testfamilie.*

Beweis. Unter den Annahmen von Satz 3.3 besitzt \mathbf{T} asymptotisch eine multivariate Normalverteilung mit Erwartungswert $\mathbf{0}$ und Korrelationsmatrix \mathbf{R} . Also ist die asymptotische gemeinsame Verteilung von \mathbf{T} vollständig spezifiziert unter $H_0^p : \bigcap_{l=1}^q \{\mathbf{c}'_l \mathbf{p} = 0\}$. Jede einzelne Teststatistik T_ℓ^p konvergiert für $N \rightarrow \infty$ zu einer Standardnormalverteilung. Insbesondere ist also die asymptotische Verteilung von T_ℓ^p unabhängig von der Verteilung von T_m^p für $\ell \neq m$. Das bedeutet, dass für eine beliebige Menge von Indizes $J \subseteq \{1, \dots, q\}$ die asymptotische gemeinsame Verteilung von $\mathbf{T}^J = \{T_j^p, j \in J\}$ vollständig spezifiziert ist unter jedem beliebigen Durchschnitt von Hypothesen $H_0^p : \bigcap_{j \in J} \{\mathbf{c}'_j \mathbf{p}\}$. □

5. Prozeduren für multiple Kontrasttests

Wie schon von [Konietschke et al. \(2012\)](#) gezeigt, wird nun aus dieser gemeinsamen Testfamilie $\{\Omega^p, \mathbf{T}\}$ eine simultane Testprozedur (STP, *simultaneous test procedure*) gewonnen. Sei dazu $z_{1-\alpha, \mathbf{2}, \mathbf{R}}$ das zweiseitige äquikoordinate $(1 - \alpha)$ -Quantil von $\mathcal{N}(\mathbf{0}, \mathbf{R})$ und entsprechend $z_{1-\alpha, \mathbf{1}, \mathbf{R}}$ das einseitige äquikoordinate $(1 - \alpha)$ -Quantil. Das heißt also: Sind $\mathbf{X} = (X_1, \dots, X_q) \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$, dann ist ([Bretz et al., 2001](#))

$$\mathbb{P}\left(\bigcap_{\ell=1}^q \{-z_{1-\alpha, \mathbf{2}, \mathbf{R}} \leq X_\ell \leq z_{1-\alpha, \mathbf{2}, \mathbf{R}}\}\right) = 1 - \alpha.$$

Abbildung 3 veranschaulicht die äquikoordinate Quantile.

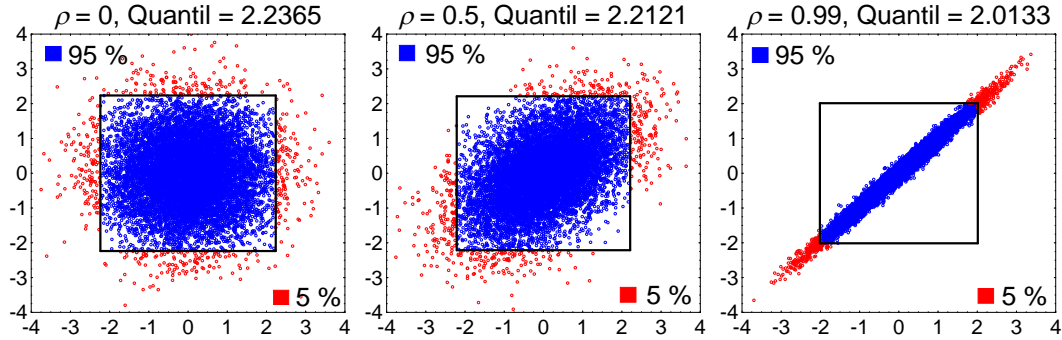


Abbildung 3: Zweiseitige äquikoordinate Quantile der $\mathcal{N}(\mathbf{0}, \mathbf{I}_2 + \rho(\mathbf{J}_2 - \mathbf{I}_2))$ -Verteilung mit unterschiedlich starken Korrelationen $\rho = 0, 0.5, 0.99$. Graphiken aus [Konietschke et al. \(2012\)](#).

Ausführungen zur numerischen Berechnung der äquikoordinate Quantile finden sich in [Bretz et al. \(2001\)](#) und [Genz und Bretz \(2009\)](#). Um $z_{1-\alpha, \mathbf{2}, \mathbf{R}}$ allerdings in einer Testprozedur zu verwenden, muss zunächst die unbekannte asymptotische Korrelationsmatrix \mathbf{R} geschätzt werden. Seien dazu $\hat{v}_{\ell\ell}$ und $\hat{v}_{\ell m}$ die konsistenten Schätzer von $v_{\ell\ell}$ und $v_{\ell m}$ aus Korollar 5.1. Diese erhält man, indem \mathbf{V}_N durch $\hat{\mathbf{V}}_N$ aus Satz 4.1 ersetzt wird. Dementsprechend ergibt sich ein konsistenter Schätzer der Korrelationsmatrix \mathbf{R} durch einfaches Einsetzen, d.h. $\hat{\mathbf{R}} = (\hat{r}_{\ell m})_{\ell, m=1, \dots, q}$, wobei $\hat{r}_{\ell m} = \hat{v}_{\ell m} / \sqrt{\hat{v}_{\ell\ell} \hat{v}_{m m}}$. Damit handelt es sich nach [Gabriel \(1969\)](#) bei der Menge $\{\Omega^p, \mathbf{T}, z_{1-\alpha, \mathbf{2}, \hat{\mathbf{R}}}\}$, bestehend aus einer Familie von Hypothesen Ω^p , einer zugehörigen Teststatistik \mathbf{T} und einem einzigen kritischen Wert $z_{1-\alpha, \mathbf{2}, \hat{\mathbf{R}}}$ für alle einzelnen Hypothesen, um eine asymptotische simultane Testprozedur (STP). Die strenge Fehlerkontrolle dieser Methode zeigt der nächste Satz.

Satz 5.3. *Für $N \rightarrow \infty$ ist für die STP $\{\Omega^p, \mathbf{T}, z_{1-\alpha, \mathbf{2}, \hat{\mathbf{R}}}\}$ eine strenge Kontrolle der familienbezogenen Fehlerrate (FWER, familywise error rate) gegeben.*

Beweis. Wie in Lemma 5.2 gesehen handelt es sich bei $\{\Omega^p, \mathbf{T}\}$ um eine asymptotische gemeinsame Testfamilie. Außerdem ist die STP $\{\Omega^p, \mathbf{T}, z_{1-\alpha, \mathbf{2}, \mathbf{R}}\}$ kohärent per Konstruktion, d.h. es kann nicht sein, dass eine Hypothese angenommen wird, während eine durch sie implizierte Hypothese verworfen wird. Somit sind die Annahmen von [Gabriel \(1969\)](#), Theorem 2, erfüllt und dieses liefert die strenge Kontrolle der FWER für diese

5. Prozeduren für multiple Kontrasttests

STP. Nun muss die unbekannte Korrelationsmatrix \mathbf{R} durch den konsistenten Schätzer $\widehat{\mathbf{R}}$ ersetzt werden. Betrachte jetzt die Abbildung $\mathbf{f}(\mathbf{R}) = z_{1-\alpha, \mathbf{2}, \mathbf{R}}$. Da \mathbf{f} stetig ist, folgt

$$z_{1-\alpha, \mathbf{2}, \widehat{\mathbf{R}}} - z_{1-\alpha, \mathbf{2}, \mathbf{R}} = \mathbf{f}(\widehat{\mathbf{R}}) - \mathbf{f}(\mathbf{R}) \xrightarrow{p} 0, \text{ für } N \rightarrow \infty.$$

Das bedeutet, $z_{1-\alpha, \mathbf{2}, \widehat{\mathbf{R}}}$ ist ein konsistenter Schätzer für $z_{1-\alpha, \mathbf{2}, \mathbf{R}}$. Somit kontrolliert auch die STP $\{\Omega^p, \mathbf{T}, z_{1-\alpha, \mathbf{2}, \widehat{\mathbf{R}}}\}$ für $N \rightarrow \infty$ die FWER streng. □

Für ausreichend große Stichprobenumfänge kann also wie folgt getestet werden: Jede Teilhypothese $H_0^p : \mathbf{c}'_\ell \mathbf{p} = 0$ wird zum zweiseitigen multiplen Level α verworfen, falls $|T_\ell^p| \geq z_{1-\alpha, \mathbf{2}, \widehat{\mathbf{R}}}$. Asymptotische simultane $(1 - \alpha)$ -Konfidenzintervalle für die Behandlungseffekte $\delta_\ell = \mathbf{c}'_\ell \mathbf{p}$ erhält man mittels

$$\left[\mathbf{c}'_\ell \widehat{\mathbf{p}} - z_{1-\alpha, \mathbf{2}, \widehat{\mathbf{R}}} \sqrt{\widehat{v}_{\ell\ell}/N} ; \mathbf{c}'_\ell \widehat{\mathbf{p}} + z_{1-\alpha, \mathbf{2}, \widehat{\mathbf{R}}} \sqrt{\widehat{v}_{\ell\ell}/N} \right]. \quad (5.3)$$

Dabei bringt diese Konstruktion einen wichtigen Vorteil der vorgestellten Methode: Die Testentscheidung zu $H_0^p : \mathbf{c}'_\ell \mathbf{p} = 0$ und das zugehörige simultane Konfidenzintervall (SCI) sind kompatibel. Was die globale Nullhypothese $H_0^p : \mathbf{C}\mathbf{p} = \mathbf{0}$ betrifft, so wird diese zum zweiseitigen multiplen Niveau α verworfen, sobald

$$\max\{|T_1^p|, \dots, |T_q^p|\} \geq z_{1-\alpha, \mathbf{2}, \widehat{\mathbf{R}}}.$$

Einseitige Konfidenzintervalle können auf die gleiche Art und Weise konstruiert werden,

$$\left[-1 ; \mathbf{c}'_\ell \widehat{\mathbf{p}} + z_{1-\alpha, \mathbf{1}, \widehat{\mathbf{R}}} \sqrt{\widehat{v}_{\ell\ell}/N} \right] \text{ oder } \left[\mathbf{c}'_\ell \widehat{\mathbf{p}} - z_{1-\alpha, \mathbf{1}, \widehat{\mathbf{R}}} \sqrt{\widehat{v}_{\ell\ell}/N} ; 1 \right]. \quad (5.4)$$

5.1. Bereichserhaltende Konfidenzintervalle

Die simultanen Konfidenzintervalle in (5.3) sind nicht notwendigerweise bereichserhaltend, d.h., die untere Grenze kann kleiner als -1 sein bzw. die obere Grenze größer als 1 . Die Behandlungseffekte $\delta_\ell = \mathbf{c}'_\ell \mathbf{p}$ sind jedoch für $\sum_{i=1}^a \sum_{j=1}^d |c_{\ell ij}| \leq 2$ durch diese Grenzen nach unten und oben beschränkt. Man kann jedoch bereichserhaltende SCIs für die Behandlungseffekte $\delta_\ell = \mathbf{c}'_\ell \mathbf{p}$ konstruieren, indem man Cramérs multivariaten Delta-Satz anwendet. Dazu seien im Folgenden die linearen Kontraste $\mathbf{c}'_\ell = (c_{\ell 11}, c_{\ell 12}, \dots, c_{\ell ad})$ normiert, d.h. $|c_{\ell ij}| \leq 1$, für alle $l = 1, \dots, q; i = 1, \dots, a; j = 1, \dots, d$. Bezeichne weiter

$$g(x_\ell) = \frac{1}{2} \log \left(\frac{1 + x_\ell}{1 - x_\ell} \right) \text{ und } g^{-1}(y_\ell) = \frac{\exp(2y_\ell) - 1}{\exp(2y_\ell) + 1}, \quad \ell = 1, \dots, q, \quad (5.5)$$

die Fisher-Transformation von x_ℓ und die entsprechende inverse Funktion. Schreibe außerdem $\mathbf{g}(\mathbf{x}) = (g(x_1), \dots, g(x_q))'$. Der Vektor der Behandlungseffekte $\delta_\ell = \mathbf{c}'_\ell \mathbf{p}$ wird mit $\mathbf{d} = (\delta_1, \dots, \delta_q)'$ bezeichnet, entsprechend der Vektor der konsistenten Schätzer $\widehat{\delta}_\ell = \mathbf{c}'_\ell \widehat{\mathbf{p}}$ mit $\widehat{\mathbf{d}} = (\widehat{\delta}_1, \dots, \widehat{\delta}_q)'$. Schlussendlich sei $\Psi = \text{diag}(1 - \delta_1^2, \dots, 1 - \delta_q^2)^{-1}$ die Jakobi-Matrix von $\mathbf{g}(\mathbf{d})$. Dann lässt sich folgendes Korollar formulieren.

5. Prozeduren für multiple Kontrasttests

Korollar 5.4. *Gelten die Annahmen von Satz 3.3, dann ist der Vektor $\sqrt{N}(\mathbf{g}(\hat{\mathbf{d}}) - \mathbf{g}(\mathbf{d}))$ multivariat $\mathcal{N}(\mathbf{0}, \mathbf{\Gamma})$ verteilt, wobei $\mathbf{\Gamma} = \mathbf{\Psi} \mathbf{V}_N \mathbf{\Psi}'$.*

Beweis. Das Korollar folgt sofort aus der Kombination von Satz 3.3 und Cramérs multivariatem Delta-Satz. \square

Definiert man $\hat{\mathbf{\Psi}} = \text{diag}(1 - \hat{\delta}_1^2, \dots, 1 - \hat{\delta}_q^2)^{-1}$, dann lässt sich wiederum für $\mathbf{\Gamma}$ aus Korollar 5.4 ein konsistenter Schätzer $\hat{\mathbf{\Gamma}} = \hat{\mathbf{\Psi}} \hat{\mathbf{V}}_N \hat{\mathbf{\Psi}}'$ angeben. Zum Testen der Teilhypothese $H_0^p : \mathbf{c}'_\ell \mathbf{p} = 0$ wird nun die Teststatistik

$$\tilde{T}_\ell^p = \sqrt{N} \frac{g(\hat{\delta}_\ell) - g(\delta_\ell)}{\sqrt{\hat{\gamma}_{\ell\ell}}}$$

gebildet, wobei $\hat{\gamma}_{\ell\ell}$ das ℓ -te Diagonalelement von $\hat{\mathbf{\Gamma}}$ darstellt. Die q Teststatistiken werden wieder in einem Vektor gesammelt

$$\tilde{\mathbf{T}} = (\tilde{T}_1^p, \dots, \tilde{T}_q^p)'. \quad (5.6)$$

Unter den Annahmen von Korollar 5.1 besitzt $\tilde{\mathbf{T}}$, asymptotisch für $N \rightarrow \infty$, eine multivariate Normalverteilung mit Erwartungswert $\mathbf{0}$ und Korrelationsmatrix \mathbf{R} . Bestimmt man hier also die obere und untere Grenze des Konfidenzintervalls für $g(\delta_\ell)$, nämlich

$$\begin{aligned} p_{\ell,L}^g &= g(\hat{\delta}_\ell) - z_{1-\alpha, 2, \hat{\mathbf{R}}} \sqrt{\hat{\gamma}_{\ell\ell}/N}, \\ p_{\ell,U}^g &= g(\hat{\delta}_\ell) + z_{1-\alpha, 2, \hat{\mathbf{R}}} \sqrt{\hat{\gamma}_{\ell\ell}/N}, \end{aligned} \quad (5.7)$$

dann lassen sich bereichserhaltende simultane $(1 - \alpha)$ -Konfidenzintervalle für die Behandlungseffekte $\delta_\ell = \mathbf{c}'_\ell \mathbf{p}$ gewinnen, indem rücktransformiert wird:

$$\left[g^{-1}(p_{\ell,L}^g) ; g^{-1}(p_{\ell,U}^g) \right]. \quad (5.8)$$

Da $g(x_\ell)$ und $g^{-1}(y_\ell)$ jeweils strikt monotone Transformationen sind, sind die bereichserhaltenden SCIs per Konstruktion kompatibel zu ihren entsprechenden Teilhypthesen.

Man beachte an dieser Stelle, dass der Wertebereich der relativen Effekte nicht $[0, 1]$ ist, sondern eingeschränkt werden kann zu

$$\frac{1}{2ad} \leq p_{ij} = \frac{1}{ad} \sum_{i=1}^a \sum_{j=1}^d [\mathbb{P}(X_{ij} < X_{rs}) + \frac{1}{2} \mathbb{P}(X_{ij} = X_{rs})] \leq 1 - \frac{1}{2ad},$$

da $0 \leq \mathbb{P}(X_{ij} < X_{rs}) + \frac{1}{2} \mathbb{P}(X_{ij} = X_{rs})$ für $(i, j) \neq (r, s)$, während für $(i, j) = (r, s)$, folgt, dass $\mathbb{P}(X_{ij} < X_{ij}) + \frac{1}{2} \mathbb{P}(X_{ij} = X_{ij}) = \frac{1}{2}$. Da der Fall $(i, j) = (r, s)$ zur Summe gehört, können die Grenzen 0 und 1 nicht mehr erreicht werden. Das bedeutet aber, dass der Bereich, den die Behandlungseffekte $\delta_\ell = \mathbf{c}'_\ell \mathbf{p}$ annehmen können, auch eingeschränkt

5. Prozeduren für multiple Kontrasttests

werden kann zu $[-1 + \frac{1}{2ad}, 1 - \frac{1}{2ad}]$. Demnach müsste also vor der Fisher-Transformation dieses Intervall auf $[-1, 1]$ transformiert werden und anschließend auch rücktransformiert werden, um tatsächlich exakt bereichserhaltende Konfidenzintervalle zu erhalten. Allerdings würde man dann das Risiko eingehen, dass, falls die untere oder obere Grenze von $[-1 + \frac{1}{2ad}, 1 - \frac{1}{2ad}]$ tatsächlich einmal angenommen wird - z.B. im Zusammenhang mit einer überaus wirksamen Therapie, die dafür sorgt, dass alle Patienten ab einem Zeitpunkt den höchsten Score erzielen - nach $-\infty$ oder $+\infty$ transformiert wird und die Auswertung stoppt. Man müsste dann z.B. diese *besten* Daten aus der Auswertung nehmen. Daher wird in dieser Arbeit darauf verzichtet, das Intervall $[-1 + \frac{1}{2ad}, 1 - \frac{1}{2ad}]$ zunächst auf $[0, 1]$ zu strecken, und in Kauf genommen, dass die Konfidenzintervalle nur in extremsten Fällen minimal nicht bereichserhaltend sind.

5.2. Approximation für kleine Stichprobenumfänge

Da die in den vorangegangenen Abschnitten vorgestellten Aussagen, z.B. Satz 3.3, asymptotischer ($N \rightarrow \infty$) Natur sind, stellt sich die Frage, ob die Qualität der Approximation durch die multivariate Normalverteilung im Falle von nicht allzu großen Stichprobenumfängen ausreichend gut ist. Die Erfahrung zeigt, dass die Konvergenz der Teststatistiken zu ihren asymptotischen multivariaten Normalverteilungen vergleichsweise langsam ist. Zu solch einem Ergebnis kommen auch [Konietschke et al. \(2012\)](#) im Falle eines One-Way-Layouts und adaptieren folglich eine von [Brunner, Dette, und Munk \(1997\)](#) sowie [Gao et al. \(2008\)](#) vorgestellte Box-Approximation ([Box, 1954](#)), um die Verteilung der Statistik \mathbf{T} mittels einer multivariaten t -Verteilung $T(\nu, \mathbf{0}, \hat{\mathbf{R}})$ mit ν Freiheitsgraden, Erwartungswert $\mathbf{0}$ und Korrelationsmatrix $\hat{\mathbf{R}}$ zu approximieren. Daher ist es interessant zu sehen, inwiefern es möglich ist, eine derartige Approximation im hier betrachteten Design anzugeben, die ebenso Rücksicht auf kleinere Stichprobenumfänge nimmt. Für jeden linearen Kontrast $\mathbf{c}'_l = (c_{l11}, c_{l12}, \dots, c_{lad})$, $l = 1, \dots, q$, definiert man dazu

$$A_{lijk} = c_{lij} \left(H(X_{ijk}) - \frac{1}{ad} F_{ij}(X_{ijk}) \right) - \frac{1}{ad} \sum_{r=1}^a \sum_{\substack{s=1 \\ (r,s) \neq (i,j)}}^d c_{lrs} F_{rs}(X_{ijk}).$$

Erinnert man sich nun an die asymptotische Äquivalenz in (3.16) und ordnet wie folgt um

$$\begin{aligned} & \sqrt{N}(\hat{p}_{ij} - p_{ij}) \\ \doteq & \sqrt{N} \left(\frac{1}{ad} \frac{1}{n_i} \sum_{k=1}^{n_i} \sum_{r=1}^a \sum_{s=1}^d F_{rs}(X_{ijk}) - \frac{1}{ad} \sum_{r=1}^a \sum_{s=1}^d \frac{1}{n_r} \sum_{k=1}^{n_r} F_{ij}(X_{rsk}) + 1 - 2p_{ij} \right) \\ = & \sqrt{N} \left(\frac{1}{n_i} \sum_{k=1}^{n_i} H(X_{ijk}) - \frac{1}{ad} \sum_{r=1}^a \sum_{s=1}^d \frac{1}{n_r} \sum_{k=1}^{n_r} F_{ij}(X_{rsk}) + 1 - 2p_{ij} \right) \\ = & \sqrt{N} \left(\frac{1}{n_i} \sum_{k=1}^{n_i} (H(X_{ijk}) - F_{ij}(X_{ijk})) - \frac{1}{ad} \sum_{r=1}^a \sum_{\substack{s=1 \\ (r,s) \neq (i,j)}}^d \frac{1}{n_r} \sum_{k=1}^{n_r} F_{ij}(X_{rsk}) + 1 - 2p_{ij} \right), \end{aligned}$$

5. Prozeduren für multiple Kontrasttests

dann lässt sich $\sqrt{N}\mathbf{c}'_\ell(\hat{\mathbf{p}} - \mathbf{p})$ in asymptotischer Äquivalenz als Linearkombination der $A_{lij k}$ schreiben

$$\begin{aligned}
\sqrt{N}\mathbf{c}'_\ell(\hat{\mathbf{p}} - \mathbf{p}) &\doteq \sqrt{N} \sum_{i=1}^a \sum_{j=1}^d \left(\frac{1}{n_i} \sum_{k=1}^{n_i} c_{lij} (H(X_{ijk}) - F_{ij}(X_{ijk})) \right. \\
&\quad \left. - \frac{1}{ad} \sum_{\substack{r=1 \\ (r,s) \neq (i,j)}}^a \sum_{s=1}^d \frac{1}{n_r} \sum_{k=1}^{n_r} c_{lij} F_{ij}(X_{rsk}) + c_{lij}(1 - 2p_{ij}) \right) \\
&= \sqrt{N} \sum_{i=1}^a \sum_{j=1}^d \left(\frac{1}{n_i} \sum_{k=1}^{n_i} c_{lij} (H(X_{ijk}) - F_{ij}(X_{ijk})) \right. \\
&\quad \left. - \frac{1}{n_i} \sum_{k=1}^{n_i} \frac{1}{ad} \sum_{\substack{r=1 \\ (r,s) \neq (i,j)}}^a \sum_{s=1}^d c_{lrs} F_{rs}(X_{ijk}) + c_{lij}(1 - 2p_{ij}) \right) \\
&= \sqrt{N} \sum_{i=1}^a \sum_{j=1}^d \frac{1}{n_i} \sum_{k=1}^{n_i} A_{lij k} + \sqrt{N} \sum_{i=1}^a \sum_{j=1}^d c_{lij}(1 - 2p_{ij}). \tag{5.9}
\end{aligned}$$

Der zweite Term ist für festes \mathbf{c}_ℓ nicht stochastisch. Daher lässt sich die Varianz ausrechnen via

$$\begin{aligned}
&\text{Var} \left(\sqrt{N} \sum_{i=1}^a \sum_{j=1}^d \frac{1}{n_i} \sum_{k=1}^{n_i} A_{lij k} + \sqrt{N} \sum_{i=1}^a \sum_{j=1}^d c_{lij}(1 - 2p_{ij}) \right) \\
&= \text{Var} \left(\sqrt{N} \sum_{i=1}^a \sum_{j=1}^d \frac{1}{n_i} \sum_{k=1}^{n_i} A_{lij k} \right) \\
&= N \sum_{i=1}^a n_i^{-1} \text{Var} \left(\sum_{j=1}^d A_{lij 1} \right) = N \sum_{i=1}^a \omega_{\ell i}^2 / n_i. \tag{5.10}
\end{aligned}$$

Dabei wurde bezüglich der vorletzten Gleichheit die Unabhängigkeit von $A_{lij k}$ und $A_{\ell i' j k}$ für $i \neq i'$ verwendet. Zudem wurde ausgenutzt, dass $A_{lij k}$ und $A_{lij k'}$ unabhängig und identisch verteilt sind für $k \neq k'$. Im letzten Schritt ist $\omega_{\ell i}^2 = \text{Var} \left(\sum_{j=1}^d A_{lij 1} \right)$. Diese Varianzen sind unbekannt, können aber mit der gleichen Argumentation wie im Beweis von Satz 4.1 konsistent geschätzt werden. Schreibe dazu

$$\omega_{\ell i} = \text{Var} \left(\sum_{j=1}^d A_{lij 1} \right) = \sum_{j=1}^d \text{Var}(A_{lij 1}) + \sum_{\substack{j=1 \\ j \neq j'}}^d \sum_{j'=1}^d \text{Cov}(A_{lij 1}, A_{lij 1'})$$

und schätze wie folgt

$$\hat{\omega}_{\ell i}^2 = \sum_{j=1}^d \sum_{j'=1}^d \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (B_{lij k} - \bar{B}_{lij \cdot}) \cdot (B_{lij' k} - \bar{B}_{lij' \cdot}), \tag{5.11}$$

5. Prozeduren für multiple Kontrasttests

wobei

$$B_{lijk} = c_{lij} \left(\widehat{H}(X_{ijk}) - \frac{1}{ad} \widehat{F}_{ij}(X_{ijk}) \right) - \frac{1}{ad} \sum_{\substack{r=1 \\ (r,s) \neq (i,j)}}^a \sum_{s=1}^d c_{lrs} \widehat{F}_{rs}(X_{ijk})$$

und

$$\bar{B}_{lij\cdot} = \frac{1}{n_i} \sum_{k=1}^{n_i} B_{lijk}$$

sind.

Analog zu [Gao und Alvo \(2008\)](#) kann die Verteilung von \mathbf{T} nun approximiert werden durch eine multivariate t-Verteilung $T(\nu, \mathbf{0}, \widehat{\mathbf{R}})$ mit $\nu = \max\{1, \min_{\ell=1, \dots, q} \{\nu_1, \dots, \nu_q\}\}$ Freiheitsgraden ([Konietschke et al., 2012](#)). Dabei ist

$$\nu_\ell = \left(\sum_{i=1}^a \widehat{\omega}_{\ell i}^2 / n_i \right)^2 / \sum_{i=1}^a \widehat{\omega}_{\ell i}^4 / (n_i^2 (n_i - 1)), \quad \ell = 1, \dots, q.$$

6. Teststatistiken für Globaltests

Um später die Güte der vorgestellten multiplen Testprozedur einordnen zu können, sollen an dieser Stelle zwei nichtparametrische Teststatistiken für Globaltests sowie als parametrische Gegenparts eine asymptotische ANOVA sowie Hotellings T^2 -Test angegeben werden.

6.1. Wald-Typ Statistik

Eine Möglichkeit, um die Globalhypothesen H_0^p aus (2.6) zu testen, bietet die Wald-Typ Statistik (WTS). Dabei handelt es sich um die quadratische Form

$$Q_N = N\hat{\mathbf{p}}'\mathbf{C}'[\mathbf{C}\hat{\mathbf{V}}_N\mathbf{C}']^+\mathbf{C}\hat{\mathbf{p}}, \quad (6.1)$$

die unter der Nullhypothese $H_0^p : \mathbf{C}\mathbf{p} = \mathbf{0}$ und unter den Annahmen von Satz 3.3 asymptotisch eine zentrale χ_f^2 -Verteilung besitzt. Dabei ist $f = \text{rk}(\mathbf{C})$ die Zahl der Freiheitsgrade und $[\cdot]^+$ bezeichnet die Moore-Penrose Inverse einer Matrix.

6.2. ANOVA-Typ Statistik

Es ist bekannt, dass die Wald-Typ Statistik den Nachteil besitzt, erst bei sehr großen Stichprobenumfängen eine ausreichende Approximation durch die χ_f^2 -Verteilung zu bieten. Daher konstruierten Brunner et al. (1997), Brunner und Puri (1999) für nichtparametrische multivariate und Repeated Measures Designs die so genannte ANOVA-Typ Statistik (ATS). Die Idee basiert auf einer Box-Approximation (Box, 1954). Dazu betrachtet man $\mathbf{M} = \mathbf{C}'[\mathbf{C}\mathbf{C}']^-\mathbf{C}$. Hier bezeichnet $[\cdot]^-$ die verallgemeinerte Inverse einer Matrix. Nun lässt sich zeigen, dass die Hypothesen $H_0^p(\mathbf{C}) : \mathbf{C}\mathbf{p} = \mathbf{0}$ und $H_0^p(\mathbf{M}) : \mathbf{M}\mathbf{p} = \mathbf{0}$ äquivalent sind, d.h. $\mathbf{C}\mathbf{p} = \mathbf{0} \Leftrightarrow \mathbf{M}\mathbf{p} = \mathbf{0}$. Die ANOVA-Typ Statistik lautet dann

$$A_N = \frac{N \cdot \text{tr}(\mathbf{M}\hat{\mathbf{V}}_N)}{\text{tr}(\mathbf{M}\hat{\mathbf{V}}_N\mathbf{M}\hat{\mathbf{V}}_N)}\hat{\mathbf{p}}'\mathbf{M}\hat{\mathbf{p}} \quad (6.2)$$

und hat unter der Nullhypothese $H_0^p : \mathbf{M}\mathbf{p} = \mathbf{0}$ approximativ eine zentrale χ_f^2 -Verteilung, wobei f geschätzt wird via

$$\hat{f} = \frac{[\text{tr}(\mathbf{M}\hat{\mathbf{V}}_N)]^2}{\text{tr}(\mathbf{M}\hat{\mathbf{V}}_N\mathbf{M}\hat{\mathbf{V}}_N)}.$$

Hier bezeichnet $\text{tr}(\cdot)$ die Spur einer quadratischen Matrix.

Ist $\text{rk}(\mathbf{M}) = 1$, dann gilt $Q_N = A_N$ und $f = 1$ (Brunner und Munzel, 2002).

6.3. Asymptotische, parametrische ANOVA

Dieser Ansatz startet mit dem allgemeinen parametrischen Modell unter der Annahme, dass

$$\mathbb{E}(\mathbf{X}_{ik}) = \boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{id})' \quad (6.3)$$

$$\text{Cov}(\mathbf{X}_{ik}) = \mathbf{V}_i \in \mathbb{R}^{d \times d}. \quad (6.4)$$

6. Teststatistiken für Globaltests

Dementsprechend formuliert man das allgemeine Modell, nachdem man die Beobachtungsvektoren zusammengefasst hat zu $\mathbf{X} = (\mathbf{X}'_{11}, \dots, \mathbf{X}'_{1n_1}, \mathbf{X}'_{21}, \dots, \mathbf{X}'_{ad})'$:

$$\mathbf{X} = \left(\bigoplus_{i=1}^a \mathbf{1}_{n_i} \otimes \mathbf{I}_d \right) \boldsymbol{\mu} + \boldsymbol{\varepsilon} = \mathbf{D}\boldsymbol{\mu} + \boldsymbol{\varepsilon} \quad (6.5)$$

mit

$$\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}'_1, \dots, \boldsymbol{\varepsilon}'_a)', \quad \boldsymbol{\varepsilon}_i = (\boldsymbol{\varepsilon}'_{i1}, \dots, \boldsymbol{\varepsilon}'_{in_i})', \quad i = 1, \dots, a, \quad (6.6)$$

$$\boldsymbol{\varepsilon}_{ik} = (\varepsilon_{i1k}, \dots, \varepsilon_{idk})'. \quad (6.7)$$

Dabei sind die $\boldsymbol{\varepsilon}_{ik}$ unabhängige, identisch verteilte Zufallsvektoren mit

$$\begin{aligned} \mathbb{E}(\boldsymbol{\varepsilon}_{ik}) &= \mathbf{0} \\ \text{Cov}(\boldsymbol{\varepsilon}_{ik}) &= \mathbf{V}_i \in \mathbb{R}^{d \times d}. \end{aligned}$$

Also ist

$$\text{Cov}(\boldsymbol{\varepsilon}) = \mathbf{S} = \bigoplus_{i=1}^a (\mathbf{I}_{n_i} \otimes \mathbf{V}_i).$$

Die Kovarianzmatrix \mathbf{S} ist jedoch unbekannt, sodass der *best linear unbiased estimator* $\boldsymbol{\mu}_{BLUE} = (\mathbf{D}'\mathbf{S}^{-1}\mathbf{D})^{-1}\mathbf{D}'\mathbf{S}^{-1}\mathbf{X}$ nicht ausgerechnet werden kann. Man verwendet daher den Kleinste-Quadrate-Schätzer $\boldsymbol{\mu}_{OLSE} = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{X} = \bar{\mathbf{X}}$. (*ordinary least squares estimator*). Er ist zwar erwartungstreu für $\boldsymbol{\mu}$ aber nicht BLUE. Zu testen sind die Hypothesen

$$\begin{aligned} H_0^\mu(A) : \mathbf{H}_A \boldsymbol{\mu} &= (\mathbf{P}_a \otimes \frac{1}{d} \mathbf{1}'_d) \cdot \boldsymbol{\mu} = \mathbf{0} \\ H_0^\mu(D) : \mathbf{H}_D \boldsymbol{\mu} &= (\frac{1}{a} \mathbf{1}'_a \otimes \mathbf{P}_d) \cdot \boldsymbol{\mu} = \mathbf{0} \\ H_0^\mu(AD) : \mathbf{H}_{AD} \boldsymbol{\mu} &= (\mathbf{P}_a \otimes \mathbf{P}_d) \cdot \boldsymbol{\mu} = \mathbf{0}. \end{aligned}$$

Allgemein soll also die Hypothese $H_0^\mu : \mathbf{H}\boldsymbol{\mu} = \mathbf{0}$ getestet werden. Betrachte dazu

$$\text{Cov}(\sqrt{N}\mathbf{H}\hat{\boldsymbol{\mu}}) = \mathbf{H} \cdot \bigoplus_{i=1}^a \frac{N}{n_i} \mathbf{V}_i \cdot \mathbf{H}' = \mathbf{H}\boldsymbol{\Sigma}\mathbf{H}'.$$

Es lässt sich nun zeigen, unter den Annahmen, dass $N/n_i \leq N_0 < \infty$ und der kleinste Eigenwert von $\mathbf{V}_i, i = 1, \dots, a$ größer Null ist,

$$\begin{aligned} \sqrt{N}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) &\overset{\cdot}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \quad \text{und} \\ \sqrt{N}\mathbf{H}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) &\overset{\cdot}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{H}\boldsymbol{\Sigma}\mathbf{H}'). \end{aligned}$$

Die Kovarianzmatrix $\boldsymbol{\Sigma}$ kann erwartungstreu und konsistent geschätzt werden durch $\hat{\boldsymbol{\Sigma}} = \bigoplus_{i=1}^a \frac{N}{n_i} \hat{\mathbf{V}}_i$. Man bildet nun eine Statistik vom ANOVA-Typ

$$F_H = N \cdot \bar{\mathbf{X}}' \mathbf{H}' [\mathbf{H}\mathbf{H}']^{-1} \mathbf{H} \bar{\mathbf{X}} \overset{\cdot}{\sim} \sum_{i=1}^a \sum_{j=1}^d \lambda_{ij} C_{ij},$$

6. Teststatistiken für Globaltests

wobei λ_{ij} die Eigenwerte von $\mathbf{T}\boldsymbol{\Sigma}$ und C_{ij} unabhängig und identisch χ_1^2 -verteilte Zufallsvariablen sind. Dabei ist $\mathbf{T} = \mathbf{H}'[\mathbf{H}\mathbf{H}']^{-1}\mathbf{H}$.

Man approximiert die Verteilung von $\sum_{i=1}^a \sum_{j=1}^d \lambda_{ij} C_{ij}$ wiederum durch eine gestreckte χ_f^2 -Verteilung, so dass die ersten beiden Momente übereinstimmen, und erhält

$$F_N = \frac{N}{\text{tr}(\widehat{\mathbf{T}\boldsymbol{\Sigma}})} \overline{\mathbf{X}}' \mathbf{T} \overline{\mathbf{X}} \cdot \underset{\sim}{\sim} F(\widehat{f}, \infty) \quad \text{mit} \quad \frac{[\text{tr}(\mathbf{T}\boldsymbol{\Sigma})]^2}{\text{tr}(\mathbf{T}\boldsymbol{\Sigma}\mathbf{T}\boldsymbol{\Sigma})}.$$

Eine ANOVA-Tabelle veranschaulicht das Hergeleitete

| Faktor | Matrix $\mathbf{T} = \mathbf{H}'[\mathbf{H}\mathbf{H}']^{-1}\mathbf{H}$ | Quadratform F_N (ANOVA-Typ) | Approximative Verteilung |
|--|--|---|-------------------------------|
| A | $\mathbf{T}_A = \mathbf{P}_a \otimes \frac{1}{d} \mathbf{J}_d$ | $\frac{N}{\text{tr}(\mathbf{T}_A \widehat{\boldsymbol{\Sigma}})} \sum_{i=1}^a (\tilde{X}_{i\cdot} - \tilde{X}_{\dots})^2$ | $F(\widehat{f}_A, \infty)$ |
| D | $\mathbf{T}_D = \frac{1}{a} \mathbf{J}_a \otimes \mathbf{P}_d$ | $\frac{N}{\text{tr}(\mathbf{T}_D \widehat{\boldsymbol{\Sigma}})} \sum_{j=1}^d (\tilde{X}_{\cdot j} - \tilde{X}_{\dots})^2$ | $F(\widehat{f}_D, \infty)$ |
| AD | $\mathbf{T}_{AD} = \mathbf{P}_a \otimes \mathbf{P}_d$ | $\frac{N}{\text{tr}(\mathbf{T}_{AD} \widehat{\boldsymbol{\Sigma}})} \sum_{i=1}^a \sum_{j=1}^d (\tilde{X}_{ij\cdot} - \tilde{X}_{i\cdot} - \tilde{X}_{\cdot j} - \tilde{X}_{\dots})^2$ | $F(\widehat{f}_{AD}, \infty)$ |
| Mittelwerte: $\tilde{X}_{i\cdot} = \frac{1}{d} \sum_{j=1}^d \bar{X}_{ij\cdot}, \quad \tilde{X}_{\cdot j} = \frac{1}{a} \sum_{i=1}^a \bar{X}_{ij\cdot}, \quad \bar{X}_{ij\cdot} = \frac{1}{n_i} \sum_{k=1}^{n_i} X_{ijk}, \quad \tilde{X}_{\dots} = \frac{1}{a} \sum_{i=1}^a \sum_{j=1}^d \tilde{X}_{i\cdot}$ | | | |
| Schätzer: $\mathbf{X}_{ik} = (X_{i1k}, \dots, X_{idk})', \quad \bar{\mathbf{X}}_i = (\bar{X}_{i1\cdot}, \dots, \bar{X}_{id\cdot})'$ $\widehat{\mathbf{V}}_i = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (\mathbf{X}_{ik} - \bar{\mathbf{X}}_i)(\mathbf{X}_{ik} - \bar{\mathbf{X}}_i)', \quad i = 1, \dots, a$ $\widehat{\boldsymbol{\Sigma}} = \bigoplus_{i=1}^a \frac{N}{n_i} \widehat{\mathbf{V}}_i, \quad \widehat{f} = \frac{[\text{tr}(\mathbf{T}\widehat{\boldsymbol{\Sigma}})]^2}{\text{tr}(\mathbf{T}\widehat{\boldsymbol{\Sigma}}\mathbf{T}\widehat{\boldsymbol{\Sigma}})}$ | | | |

6.4. Hotellings T^2 Test

Im Fall von zwei Stichproben mit d Repeated Measures bietet Hotellings T^2 Test einen in der Praxis häufig verwendeten exakten Test. Dabei wird angenommen, dass die Beobachtungsvektoren $\mathbf{X}_{ik}, i = 1, 2; k = 1, \dots, n_i$ multivariat normalverteilt sind, genauer

$$\mathbf{X}_{ik} \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}).$$

Das heißt, die Kovarianzmatrizen in den beiden Gruppen werden zusätzlich als gleich vorausgesetzt ($\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$). Man beachte, dass dies eine enorm einschränkende Annahme ist. Möchte man nun die Hypothese $H_0^{\mu}: \mathbf{C}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \mathbf{0}$ testen, wobei \mathbf{C} eine $c \times d$ -Matrix mit Rang c ($c \leq d$) ist, so beobachtet man zunächst, dass

$$\mathbf{C}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2) \sim \mathcal{N}(\mathbf{C}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2), \frac{n_1 + n_2}{n_1 n_2} \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}').$$

Sind $\mathbf{S}_1, \mathbf{S}_2$ die empirischen Kovarianzmatrizen und $\mathbf{S}_{zi} = \mathbf{C}\mathbf{S}_i\mathbf{C}', i = 1, 2$, dann lassen sich gepoolte Schätzer von $\boldsymbol{\Sigma}$ und $\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}'$ angeben:

$$\mathbf{S} = \frac{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2}{n_1 + n_2 - 2} \quad \mathbf{S}_z = \frac{(n_1 - 1)\mathbf{S}_{z1} + (n_2 - 1)\mathbf{S}_{z2}}{n_1 + n_2 - 2}.$$

6. Teststatistiken für Globaltests

Nun ist

$$(n_1 + n_2 - 2)\mathbf{S}_z \sim W_c(n_1 + n_2 - 2, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}') \text{ (Wishart Verteilung),}$$

und somit hat die Statistik

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{\mathbf{X}}_{1.} - \bar{\mathbf{X}}_{2.})' \mathbf{C}' (\mathbf{C}\mathbf{S}\mathbf{C}')^{-1} \mathbf{C} (\bar{\mathbf{X}}_{1.} - \bar{\mathbf{X}}_{2.})$$

eine $T_{c, n_1+n_2-2, \delta}^2$ -Verteilung mit Nichtzentralitätsparameter

$$\delta = \frac{n_1 n_2}{n_1 + n_2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \mathbf{C}' (\mathbf{C}\mathbf{S}\mathbf{C}')^{-1} \mathbf{C} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2).$$

Die Statistik

$$F = \frac{(n_1 + n_2 - 2) - c + 1}{(n_1 + n_2 - 2)c} T^2 = \frac{n_1 + n_2 - c - 1}{(n_1 + n_2 - 2)c} T^2$$

hat dann eine $F_{c, n_1+n_2-c-1, \delta}$ -Verteilung. Gilt $H_0^\mu : \mathbf{C}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \mathbf{0}$, dann hat F eine zentrale F_{c, n_1+n_2-c-1} -Verteilung. Für weitere Details und eine ausführliche Herleitung siehe [Davis \(2002\)](#), Abschnitt 3.4.

7. Simulationen

In diesem Kapitel soll nun die hergeleitete Methode der multiplen Kontrasttests und simultanen Konfidenzintervalle auf ihre Güte untersucht werden. Dazu wird sie in den folgenden Simulationsstudien mit den Globaltests mittels ANOVA-Typ Statistik (ATS, Abschnitt 6.2, Ränge) und parametrischer, asymptotischer ANOVA (Abschnitt 6.3, keine Ränge) verglichen. Für Designs mit lediglich zwei Gruppen wird zusätzlich der parametrische Hotellings T^2 -Test (Abschnitt 6.4) betrachtet. Simuliert wird sowohl das Niveau als auch die Power der Tests. Um Simulationen durchzuführen, müssen natürlich zunächst Daten derart generiert werden, sodass deren Verteilung bzw. deren Parameter bekannt sind. Nur so können anhand dieser Parameter Settings generiert werden, in denen feststeht, ob ein Test die Nullhypothese richtigerweise oder fälschlicherweise verwirft. In der vorliegenden Arbeit wurde die frei verfügbare Statistik-Software R (URL <http://CRAN.R-project.org/>) verwendet.

7.1. Erzeugung von Zufallsdaten

Wie auch im Technical Report von Brunner und Placzek (2011) werden die Daten generiert, indem die unabhängigen Beobachtungsvektoren \mathbf{X}_{ik} , $i = 1, \dots, a$; $k = 1, \dots, n_i$ durch die folgende lineare Transformation erhalten werden:

$$\mathbf{X}_{ik} = \sigma_i \mathbf{V}^{1/2} \mathbf{Z}_{ik} + c_i B_{ik} \mathbf{1}_d. \quad (7.1)$$

Dabei repräsentiert die Zufallsvariable B_{ik} den Blockeffekt, auch Individualeffekt genannt, des k -ten Individuums in Gruppe i und c_i den zugehörigen Skalierungsfaktor. Sie ist unabhängig von \mathbf{Z}_{ik} und ermöglicht es, in allen Gruppen unterschiedliche oder auch nicht vorhandene Blockeffekte zu simulieren. Der Zufallsvektor \mathbf{Z}_{ik} erzeugt den Versuchsfehler, die Matrix $\mathbf{V} = (v_{kl})_{k,l=1,\dots,d}$ die Kovarianzstruktur der Repeated Measures. Dabei ist zu beachten, dass es sich tatsächlich nur um die Kovarianzstruktur der Repeated Measures handelt. Wird ein Blockeffekt additiv hinzugefügt, verändert dies die Kovarianzstruktur. Zur Skalierung der Kovarianzmatrizen stehen σ_i zur Verfügung. Die Generierung der einzelnen Größen erhält man folgendermaßen:

$$\begin{aligned} \mathbf{Z}_{ik} &= (Z_{ik1}, \dots, Z_{ikd})' \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d), \quad i = 1, \dots, a, \quad k = 1, \dots, n_i, \\ \mathbf{B}_i &= (B_{i1}, \dots, B_{in_i})' \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n_i}), \quad i = 1, \dots, a, \\ \mathbf{V} &= \begin{cases} \mathbf{I}_d & , \text{ für eine Compound Symmetry Struktur,} \\ v_{kl} = \rho^{|k-l|} & , \quad 0 < \rho < 1, \text{ für eine autoregressive Struktur,} \\ v_{kl} = d - |k - l| & , \text{ für eine (linear abfallende) Toeplitz Struktur.} \end{cases} \end{aligned} \quad (7.2)$$

Um in dieser Arbeit nicht nur Daten zu untersuchen, die mit Hilfe einer multivariaten Normalverteilung generiert wurden, werden bei den Simulationen zusätzlich multivariat lognormalverteilte Daten erzeugt. Diese erhält man, indem die wie oben erzeugten Beobachtungsvektoren exponiert werden $\widetilde{\mathbf{X}}_{ik} = \exp(\mathbf{X}_{ik})$. Es gilt nämlich, dass wenn \mathbf{X} eine multivariate Normalverteilung $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ besitzt, dann hat $\mathbf{Y} = \exp(\mathbf{X})$ eine multivariate

7. Simulationen

Lognormalverteilung mit (Johnson, Balakrishnan, und Kotz, 1980)

$$\begin{aligned}\mathbb{E}[\mathbf{Y}]_i &= \exp(\mu_i + \frac{1}{2}\Sigma_{ii}) \\ \text{Cov}[\mathbf{Y}]_{ij} &= \exp[\mu_i + \mu_j + \frac{1}{2}(\Sigma_{ii} + \Sigma_{jj})](\exp[\Sigma_{ij}] - 1).\end{aligned}$$

Dabei ist anzumerken, dass man an dieser Stelle offenbar die Kontrolle über die Kovarianzmatrizen ein Stück weit aufgeben muss. Man kann lediglich die Kovarianzstrukturen der zu Grunde liegenden Normalverteilungen kontrollieren. Zudem ist der Erwartungswert abhängig von der Kovarianzmatrix, was vor allem im Fall von ungleichen Kovarianzmatrizen beachtet werden muss.

7.2. Niveau

Will man das Niveau eines Test simulieren, so muss eine Situation geschaffen werden, in der die Nullhypothese gilt, und dann empirisch festgestellt werden, wie oft diese wahre Nullhypothese fälschlicherweise verworfen wird. Im zu untersuchenden Design sind drei Effekte möglich: der Gruppeneffekt A , der Zeiteffekt D und der Wechselwirkungseffekt AD . In der Simulationsstudie werden unterschiedliche Settings betrachtet. Zunächst kann man natürlich Daten derart erzeugen, dass keiner dieser Effekte vorhanden ist, somit alle drei Nullhypothesen $H_0(A)$ (kein Gruppeneffekt), $H_0(D)$ (kein Zeiteffekt) und $H_0(AD)$ (keine Wechselwirkung) zutreffen. Dann führt man die jeweiligen Globaltests n_{sim} -mal zu einem Signifikanzniveau α durch und bestimmt jeweils das empirische Niveau, d.h. den Anteil verworfener Resultate. Je näher das empirische Niveau dem vorgegebenen, so genannten nominalen Niveau α ist, desto besser ist die Performance des Tests. Interessant ist jedoch auch, wie sich das Niveau bzw. der Test gezielt auf einen Effekt verhält, wenn nur diese Nullhypothese wahr ist. Zum Beispiel könnte man untersuchen, wie sich das Niveau bei einem Test auf $H_0(AD)$ verhält, wenn sowohl ein Gruppen- als auch ein Zeiteffekt vorhanden ist. Analoges ist möglich beim Testen von $H_0(A)$ bzw. $H_0(D)$ mit jeweils vorhandenen anderen beiden Effekten. An dieser Stelle muss man sich jedoch Gedanken machen, ob dann noch alle Tests unter der Nullhypothese durchgeführt werden. Das heißt, sobald additiv - wie im linearen Modell - Effekte hinzugefügt werden, muss man sich die Frage stellen, ob dann die nichtparametrischen Tests noch unter H_0^F bzw. H_0^p testen oder ob man sich schon unter der Alternative befindet. Die parametrischen Verfahren testen dagegen unter der so konstruiert wahren Nullhypothese H_0^μ . Dies wird genauer ausgeführt.

Simuliert wird immer mit $n_{sim} = 10000$ Simulationsläufen. Die Stichprobenumfänge liegen dabei pro Gruppe zwischen 10 und 30 Versuchseinheiten. Als Kovarianzstrukturen stehen wie schon oben erwähnt CS, $\text{AR}(\rho)$ und TPL zur Verfügung, die anhand der σ_i skaliert werden können. Kombiniert mit den Stichprobenumfängen ergeben sich die bekannten Situationen: balancierte und unbalancierte Stichprobenumfänge, das sogenannte positive pairing sowie negative pairing. Beim positive pairing sind die kleinen Stichprobenumfänge gepaart mit kleinen Varianzen, also kleinem σ_i , und die großen Stichprobenumfänge mit den großen Varianzen. Entsprechend umgekehrt ist die Situation beim negative pairing.

7. Simulationen

In den Tabellen und Graphiken werden die unterschiedlichen Tests wie folgt abgekürzt: Für den multiplen Kontrasttest wird - dabei handelt es sich immer um die Approximation mit der multivariaten t-Verteilung (Abschnitt 5.2) kombiniert mit der Fisher-Transformation (Abschnitt 5.1) - die Bezeichnung MCTP verwendet. Die nichtparametrische ANOVA-Typ Statistik erhält die Abkürzung ATS und die asymptotische parametrische Varianzanalyse (dabei handelt es sich auch um eine ANOVA-Typ Statistik) wird einfach mit ANOVA bezeichnet. Hotellings T^2 -Test wird mit T^2 abgekürzt. Man beachte, dass die vier Methoden nicht die gleichen Hypothesen testen: MCTP und die nichtparametrische ATS verwenden die relativen Effekte und testen H_0^p . Die parametrische ANOVA-Typ Statistik und Hotellings T^2 hingegen verwenden die Erwartungswerte und testen H_0^μ .

Den Anfang macht in dieser Simulationsstudie ein balanciertes 3×3 Design. Hier meint $a \times d$, dass der Gruppenfaktor A gerade a Stufen und der Zeitfaktor D entsprechend d Stufen aufweist. Simuliert werden, wie oben besprochen, die Fehlerraten des Fehlers erster Art - fälschliches Verwerfen einer wahren Hypothese - zu einem nominalen Niveau von $\alpha = 5\%$ in $n_{sim} = 10000$ Simulationsläufen für alle drei Effekte: Gruppeneffekt, Zeiteffekt und Wechselwirkungseffekt. Die Daten werden in jedem Simulationslauf mit Hilfe der multivariaten Normalverteilung so erzeugt, dass keiner dieser drei Effekte vorhanden ist, d.h. alle drei Nullhypothesen $H_0(A)$, $H_0(D)$ und $H_0(AD)$ treffen zu. Exemplarisch

Tabelle 1: Simulierte Fehlerraten des Fehlers erster Art zum nominalen Niveau $\alpha = 0.05$ in 10000 Simulationsläufen bei gleichen Stichprobenumfängen und gleichen Skalierungsfaktoren der Kovarianzmatrizen. Erzeugung der Daten erfolgt mittels multivariater Normalverteilung und derart, dass kein Effekt vorhanden ist.

| Balanced 3×3 Design, Kovarianzstruktur: CS | | | | | | | | | |
|---|------------------------|--------|--------|----------------|--------|--------|---------------------|--------|--------|
| | Gruppeneffekt A | | | Zeiteffekt D | | | Wechselwirkung AD | | |
| Kontrast | MCTP | ATS | ANOVA | MCTP | ATS | ANOVA | MCTP | ATS | ANOVA |
| | $n_1 = n_2 = n_3 = 10$ | | | | | | | | |
| T | 0.0484 | 0.0585 | 0.0521 | 0.0426 | 0.0456 | 0.0524 | 0.0384 | 0.0431 | 0.0433 |
| A | 0.053 | 0.0606 | 0.054 | 0.0401 | 0.0473 | 0.0474 | 0.0344 | 0.0412 | 0.0434 |
| C | 0.0515 | 0.0578 | 0.0532 | 0.0417 | 0.0429 | 0.0474 | 0.0253 | 0.0435 | 0.0427 |
| D | 0.0505 | 0.0597 | 0.0547 | 0.0415 | 0.0454 | 0.0456 | 0.0369 | 0.0427 | 0.0451 |
| | $n_1 = n_2 = n_3 = 20$ | | | | | | | | |
| T | 0.0521 | 0.0532 | 0.0542 | 0.0486 | 0.0476 | 0.048 | 0.0405 | 0.0424 | 0.0452 |
| A | 0.0507 | 0.0584 | 0.051 | 0.0418 | 0.0439 | 0.047 | 0.0436 | 0.0485 | 0.0513 |
| C | 0.0462 | 0.0522 | 0.0471 | 0.0459 | 0.0439 | 0.048 | 0.0318 | 0.0448 | 0.0444 |
| D | 0.046 | 0.0519 | 0.0495 | 0.0415 | 0.0466 | 0.0474 | 0.0404 | 0.0446 | 0.0461 |
| | $n_1 = n_2 = n_3 = 30$ | | | | | | | | |
| T | 0.0508 | 0.0545 | 0.0531 | 0.0461 | 0.0491 | 0.0495 | 0.043 | 0.0456 | 0.0465 |
| A | 0.0523 | 0.0563 | 0.0522 | 0.0535 | 0.0503 | 0.0519 | 0.0428 | 0.0463 | 0.0472 |
| C | 0.0498 | 0.0562 | 0.0542 | 0.0474 | 0.0478 | 0.0516 | 0.0335 | 0.0457 | 0.049 |
| D | 0.0495 | 0.0541 | 0.0537 | 0.0462 | 0.0488 | 0.0494 | 0.0398 | 0.0444 | 0.0437 |

7. Simulationen

wurde als Kovarianzstruktur in den Gruppen eine Compound Symmetry (CS) gewählt. Die Skalierungsfaktoren sind identisch $\sigma_1 = \sigma_2 = \sigma_3 = 1$. Als Kontraste wurden die in Abschnitt 2.3 vorgestellten Kontrastmatrizen für Tukey (T), Average (A), Changepoint (C) und Dunnett (D) Kontraste verwendet. Dabei ist anzumerken, dass nur die Average-Kontraste einen fairen Vergleich zum parametrischen ANOVA-Verfahren bieten, da die Varianzanalyse auch auf solch einem Kontrast aufbaut.

Die Ergebnisse sind in Tabelle 1 aufgeführt. Betrachtet man das Resultat, kann man feststellen, dass alle drei Verfahren bezüglich Gruppen- und Zeiteffekt das Niveau recht gut einhalten. Die Testprozedur basierend auf multiplen Kontrasttests ist zwar beim Zeiteffekt und den kleinsten Stichprobenumfängen minimal konservativ, dies verschwindet aber mit steigender Stichprobengröße, so, wie es bei einem asymptotischen Verfahren sein sollte. Ähnliches gilt für die ANOVA. Bezüglich des Gruppeneffektes hält MCTP das Niveau akkurat ein, ATS und ANOVA sind minimal liberal. Ein deutlicher Unterschied zwischen den gewählten Kontrasten lässt sich nicht feststellen. Anders sieht es beim Wechselwirkungseffekt aus: Hier sind alle drei Verfahren leicht konservativ, am stärksten bei den kleinsten Stichprobenumfängen. MCTP schneidet hier am schlechtesten ab. Interessant zu beobachten ist auch, dass das schlechteste Ergebnis auf den Test mittels Changepoint-Kontrast zurückgeht. Dies liegt jedoch an den Korrelationen, die der Kontrast erzeugt. Der nächste Schritt ist das Zulassen von ungleichen Stich-

Tabelle 2: Simulierte Fehlerraten des Fehlers erster Art zum nominalen Niveau $\alpha = 0.05$ in 10000 Simulationsläufen bei ungleichen Stichprobenumfängen und ungleichen Kovarianzmatrizen (**positive pairing**). Erzeugung der Daten erfolgt mittels multivariater Normalverteilung und derart, dass kein Effekt vorhanden ist.

| Unbalanced 3 × 3 Design, $n_1 = 14, n_2 = 20, n_3 = 25, \sigma_1 = 1, \sigma_2 = 2, \sigma_3 = 3$ | | | | | | | | | |
|---|------------------------|--------|--------|---------------------|--------|--------|--------------------------|--------|--------|
| | Gruppeneffekt <i>A</i> | | | Zeiteffekt <i>D</i> | | | Wechselwirkung <i>AD</i> | | |
| Kontrast | MCTP | ATS | ANOVA | MCTP | ATS | ANOVA | MCTP | ATS | ANOVA |
| | CS | | | | | | | | |
| T | 0.049 | 0.0539 | 0.0512 | 0.0491 | 0.0531 | 0.0489 | 0.0457 | 0.0464 | 0.0443 |
| A | 0.0514 | 0.0555 | 0.052 | 0.0461 | 0.0487 | 0.0478 | 0.0484 | 0.0466 | 0.0424 |
| C | 0.051 | 0.0533 | 0.0509 | 0.0466 | 0.0472 | 0.0495 | 0.0322 | 0.0424 | 0.0401 |
| D | 0.0495 | 0.0558 | 0.0479 | 0.0461 | 0.0471 | 0.0496 | 0.0416 | 0.0456 | 0.0414 |
| | AR(0.6) | | | | | | | | |
| T | 0.0465 | 0.0545 | 0.0483 | 0.0448 | 0.0462 | 0.0461 | 0.0394 | 0.0442 | 0.0449 |
| A | 0.0495 | 0.0543 | 0.0451 | 0.0422 | 0.042 | 0.0454 | 0.0427 | 0.043 | 0.0375 |
| C | 0.0479 | 0.0552 | 0.0482 | 0.0442 | 0.0426 | 0.0459 | 0.0276 | 0.0402 | 0.0401 |
| D | 0.0493 | 0.0528 | 0.0444 | 0.0451 | 0.0455 | 0.0455 | 0.0373 | 0.0413 | 0.0394 |
| | TPL | | | | | | | | |
| T | 0.0503 | 0.0561 | 0.0466 | 0.0426 | 0.0391 | 0.0414 | 0.0389 | 0.0402 | 0.0414 |
| A | 0.0472 | 0.0538 | 0.0448 | 0.0415 | 0.0377 | 0.038 | 0.0417 | 0.0415 | 0.036 |
| C | 0.0476 | 0.0527 | 0.0436 | 0.0468 | 0.0395 | 0.0374 | 0.0288 | 0.041 | 0.0362 |
| D | 0.0501 | 0.0536 | 0.0475 | 0.0429 | 0.0368 | 0.0412 | 0.0391 | 0.0414 | 0.0423 |

7. Simulationen

probenumfängen und ungleichen Kovarianzmatrizen. Tabelle 2 zeigt die Ergebnisse der Simulationen, bei denen als Stichprobenumfänge $n_1 = 14$, $n_2 = 20$, $n_3 = 25$ und als Skalierungsfaktoren der Kovarianzmatrizen $\sigma_1 = 1$, $\sigma_2 = 2$, $\sigma_3 = 3$ gewählt wurden. Damit ist man im Setting des *positive pairing*. Als Kovarianzstrukturen wurden sowohl die Compound Symmetry (CS), als auch eine autoregressive Struktur mit Parameter $\rho = 0.6$ (AR(0.6)) und eine linear abfallende Toeplitz Struktur (TPL) verwendet. Die Fehlerraten unterscheiden sich für die CS Struktur nicht großartig von denen im vorigen Setting. Bezüglich der Wechselwirkung ist die Performance von MCTP bis auf den Changepoint-Kontrast sogar besser. Bei den anderen beiden Kovarianzstrukturen lässt sich ein leicht konservativeres Verhalten feststellen, vor allem beim Zeit- und Wechselwirkungseffekt bei der Toeplitz-Struktur. Der Fall des *negative pairing* wird in Tabelle 3 behandelt. Hier ist $n_1 = 14$, $n_2 = 20$, $n_3 = 25$ und $\sigma_1 = 3$, $\sigma_2 = 2$, $\sigma_3 = 1$. Sonst ist das Setting genauso wie im Fall des soeben analysierten positive pairing.

Auch die Ergebnisse unterscheiden sich nicht großartig: Für die Compound Symmetry Struktur bieten alle drei Verfahren gute Fehlerraten für den Gruppen- sowie den Zeiteffekt. Bei der Wechselwirkung geht die Tendenz wieder in Richtung konservatives Testen. Bei der autoregressiven und der Toeplitz Struktur liefert die ANOVA das beste empirische Niveau, während die anderen beiden Verfahren entweder leicht konservativ (MCTP) oder leicht liberal (ATS) werden. Bei der Wechselwirkung sind wieder alle drei Tests relativ konservativ.

Tabelle 3: Simulierte Fehlerraten des Fehlers erster Art bei ungleichen Stichprobenumfängen und Kovarianzmatrizen (**negative pairing**). Erzeugung der Daten erfolgt mittels multivariater Normalverteilung und derart, dass kein Effekt vorhanden ist.

| Unbalanced 3×3 Design, $n_1 = 14, n_2 = 20, n_3 = 25, \sigma_1 = 3, \sigma_2 = 2, \sigma_3 = 1$ | | | | | | | | | |
|--|-----------------|--------|--------|--------------|--------|--------|-------------------|--------|--------|
| | Gruppeneffekt A | | | Zeiteffekt D | | | Wechselwirkung AD | | |
| Kontrast | MCTP | ATS | ANOVA | MCTP | ATS | ANOVA | MCTP | ATS | ANOVA |
| | CS | | | | | | | | |
| T | 0.0479 | 0.0548 | 0.0488 | 0.0479 | 0.0486 | 0.0494 | 0.038 | 0.0432 | 0.0405 |
| A | 0.0414 | 0.0571 | 0.049 | 0.0479 | 0.0504 | 0.0473 | 0.0408 | 0.0418 | 0.0403 |
| C | 0.0434 | 0.0545 | 0.0488 | 0.0523 | 0.053 | 0.0479 | 0.0292 | 0.0452 | 0.0401 |
| D | 0.0486 | 0.0542 | 0.0478 | 0.0512 | 0.0521 | 0.0484 | 0.0395 | 0.047 | 0.0407 |
| | AR(0.6) | | | | | | | | |
| T | 0.0423 | 0.0549 | 0.0443 | 0.0447 | 0.0464 | 0.0511 | 0.0316 | 0.0404 | 0.0403 |
| A | 0.0402 | 0.0586 | 0.0478 | 0.0466 | 0.0432 | 0.0466 | 0.0314 | 0.04 | 0.0412 |
| C | 0.043 | 0.0551 | 0.0485 | 0.0447 | 0.0457 | 0.0481 | 0.0213 | 0.0401 | 0.0404 |
| D | 0.0457 | 0.0547 | 0.0456 | 0.0425 | 0.0414 | 0.0456 | 0.0379 | 0.0422 | 0.0417 |
| | TPL | | | | | | | | |
| T | 0.046 | 0.0595 | 0.0486 | 0.0415 | 0.0411 | 0.0439 | 0.0317 | 0.0415 | 0.0414 |
| A | 0.0458 | 0.0592 | 0.0523 | 0.0412 | 0.0409 | 0.046 | 0.0293 | 0.0368 | 0.0385 |
| C | 0.0438 | 0.0578 | 0.0492 | 0.0427 | 0.0385 | 0.0442 | 0.0199 | 0.0413 | 0.0423 |
| D | 0.0456 | 0.0597 | 0.0521 | 0.0447 | 0.0418 | 0.0419 | 0.0353 | 0.042 | 0.0442 |

7. Simulationen

Nachdem nun das Niveau für alle drei Nullhypothesen gleichzeitig untersucht wurde, werden jetzt die einzelnen Nullhypothesen gezielt einzeln betrachtet. Dabei werden die Daten immer so generiert, dass gerade die anderen beiden Effekte vorhanden sind. Beispielsweise erzeugt man Daten mit Gruppen- und Zeiteffekt, aber keinem Wechselwirkungseffekt, und untersucht dann das empirische Niveau für $H_0(AD)$. Hier muss man jedoch, wie zu Beginn dieses Kapitels erwähnt, aufpassen, welche Hypothesen gelten. In Abschnitt 2.2 wurde schon gezeigt, dass aus $H_0^F : \mathbf{CF} = \mathbf{0}$ bezüglich relativer Effekte $H_0^p : \mathbf{Cp} = \mathbf{0}$ folgt. Selbstverständlich folgt $H_0^\mu : \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ ebenso aus H_0^F . Die Umkehrungen gelten im Allgemeinen jedoch nicht (siehe Abbildung rechts). Wenn nun die Effekte additiv hinzugefügt werden, wie aus den linearen Modellen bekannt, dann stellt sich natürlich die Frage, ob unter der Nullhypothese H_0^μ auch die Hypothese H_0^p ihre Gültigkeit behält. Ansonsten würde man die nichtparametrischen Tests über die relativen Effekte - MCTP und ATS - nur unter der Alternative durchführen. Um dies einmal zu veranschaulichen, kann man die relativen Effekte unter Annahme der Normalverteilung einfach ausrechnen und sich entsprechende Situationen anschauen. Nimmt man im 2×2 Design an, dass $\mathbf{X}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \mathbf{V}_1)$ und $\mathbf{X}_2 \sim \mathcal{N}(\boldsymbol{\mu}_2, \mathbf{V}_2)$, dann ergeben sich als relative Effekte



Abbildung 4: Implikationen für die Hypothesen via Verteilungsfunktionen, relativer Effekte und Erwartungswerte.

$$\begin{aligned}
 p_{11} &= \frac{1}{4} \left[\frac{1}{2} + \Phi\left(\frac{\mu_{11} - \mu_{12}}{\sqrt{\mathbf{V}_1^{(1,1)} + \mathbf{V}_1^{(2,2)}}}\right) + \Phi\left(\frac{\mu_{11} - \mu_{21}}{\sqrt{\mathbf{V}_1^{(1,1)} + \mathbf{V}_2^{(1,1)}}}\right) + \Phi\left(\frac{\mu_{11} - \mu_{22}}{\sqrt{\mathbf{V}_1^{(1,1)} + \mathbf{V}_2^{(2,2)}}}\right) \right] \\
 p_{12} &= \frac{1}{4} \left[\frac{1}{2} + \Phi\left(\frac{\mu_{12} - \mu_{11}}{\sqrt{\mathbf{V}_1^{(2,2)} + \mathbf{V}_1^{(1,1)}}}\right) + \Phi\left(\frac{\mu_{12} - \mu_{21}}{\sqrt{\mathbf{V}_1^{(2,2)} + \mathbf{V}_2^{(1,1)}}}\right) + \Phi\left(\frac{\mu_{12} - \mu_{22}}{\sqrt{\mathbf{V}_1^{(2,2)} + \mathbf{V}_2^{(2,2)}}}\right) \right] \\
 p_{21} &= \frac{1}{4} \left[\frac{1}{2} + \Phi\left(\frac{\mu_{21} - \mu_{22}}{\sqrt{\mathbf{V}_2^{(1,1)} + \mathbf{V}_2^{(2,2)}}}\right) + \Phi\left(\frac{\mu_{21} - \mu_{11}}{\sqrt{\mathbf{V}_2^{(1,1)} + \mathbf{V}_1^{(1,1)}}}\right) + \Phi\left(\frac{\mu_{21} - \mu_{12}}{\sqrt{\mathbf{V}_2^{(1,1)} + \mathbf{V}_1^{(2,2)}}}\right) \right] \\
 p_{22} &= \frac{1}{4} \left[\frac{1}{2} + \Phi\left(\frac{\mu_{22} - \mu_{21}}{\sqrt{\mathbf{V}_2^{(2,2)} + \mathbf{V}_2^{(1,1)}}}\right) + \Phi\left(\frac{\mu_{22} - \mu_{11}}{\sqrt{\mathbf{V}_2^{(2,2)} + \mathbf{V}_1^{(1,1)}}}\right) + \Phi\left(\frac{\mu_{22} - \mu_{12}}{\sqrt{\mathbf{V}_2^{(2,2)} + \mathbf{V}_1^{(2,2)}}}\right) \right]
 \end{aligned}$$

Damit kann man nun konkret bestimmen, was mit den nichtparametrischen Hypothesen passiert, wenn Haupteffekte oder ein Wechselwirkungseffekt addiert werden. Tabelle 4 zeigt, welche Hypothesen gelten, wenn solche Effekte mittels des Erwartungswertvektors $\boldsymbol{\mu} = (\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22})'$ erzeugt werden. In der oberen Hälfte sind die Kovarianzmatrizen gleich, d.h. $\mathbf{V}_1 = \mathbf{V}_2$, in der unteren Hälfte ungleich $\mathbf{V}_1 = \frac{1}{2}\mathbf{V}_2$. Sind die Kovarianzmatrizen gleich, kann man aus der obigen Darstellung der relativen Effekte nachrechnen, dass, falls H_0^μ gilt, auch H_0^p zutrifft. Dies spiegelt auch die Tabelle wider. Es gelten jeweils alle Hypothesen, bis auf die, die auf den Effekt testen, der hinzugefügt wurde. Bei ungleichen Kovarianzmatrizen ist dies nur der Fall beim Hinzufügen eines Gruppeneffektes A . Sowohl $H_0^p(D)$ als auch $H_0^p(AD)$ bleiben gültig. Fügt man jedoch lediglich einen Zeiteffekt D hinzu, gilt in den relativen Effekten $H_0^p(AD)$ nicht mehr. Entsprechendes lässt sich bei der Wechselwirkung feststellen: Wird nur eine Wechselwirkung addiert, ist zusätzlich

7. Simulationen

Tabelle 4: Parametrische und nichtparametrische Hypothesen beim additiven Hinzufügen von Haupt- und Wechselwirkungseffekten

| $V_1 = V_2$ | | | |
|------------------------|---|--|---|
| Effekt | A | D | AD |
| μ | $(0, 0, 1, 1)'$ | $(0, 1, 0, 1)'$ | $(0, 1, 1, 0)'$ |
| p | $(0.37, 0.37, 0.63, 0.63)'$ | $(0.37, 0.63, 0.37, 0.63)'$ | $(0.37, 0.63, 0.63, 0.37)'$ |
| $H_0^\mu(A)$ | $0 = \mu_1. \neq \mu_2. = 1$ | $0.5 = \mu_1. = \mu_2. = 0.5$ | $0.5 = \mu_1. = \mu_2. = 0.5$ |
| $H_0^p(A)$ | $0.37 = p_1. \neq p_2. = 0.63$ | $0.5 = p_1. = p_2. = 0.5$ | $0.5 = p_1. = p_2. = 0.5$ |
| $H_0^\mu(D)$ | $0.5 = \mu_1 = \mu_2 = 0.5$ | $0 = \mu_1 \neq \mu_2 = 1$ | $0.5 = \mu_1 = \mu_2 = 0.5$ |
| $H_0^p(D)$ | $0.5 = p_1 = p_2 = 0.5$ | $0.37 = p_1 \neq p_2 = 0.63$ | $0.5 = p_1 = p_2 = 0.5$ |
| $H_0^\mu(AD)$ | $0 = \mu_{11} - \mu_{12} = \mu_{21} - \mu_{22} = 0$ | $-1 = \mu_{11} - \mu_{12} = \mu_{21} - \mu_{22} = -1$ | $-1 = \mu_{11} - \mu_{12} \neq \mu_{21} - \mu_{22} = 1$ |
| $H_0^p(AD)$ | $0 = p_{11} - p_{12} = p_{21} - p_{22} = 0$ | $-0.26 = p_{11} - p_{12} = p_{21} - p_{22} = -0.26$ | $-0.26 = p_{11} - p_{12} \neq p_{21} - p_{22} = 0.26$ |
| $V_1 = \frac{1}{2}V_2$ | | | |
| Effekt | A | D | AD |
| μ | $(0, 0, 1, 1)'$ | $(0, 1, 0, 1)'$ | $(0, 1, 1, 0)'$ |
| p | $(0.391, 0.391, 0.609, 0.609)'$ | $(0.38, 0.62, 0.4, 0.6)'$ | $(0.38, 0.62, 0.6, 0.4)'$ |
| $H_0^\mu(A)$ | $0 = \mu_1. \neq \mu_2. = 1$ | $0.5 = \mu_1. = \mu_2. = 0.5$ | $0.5 = \mu_1. = \mu_2. = 0.5$ |
| $H_0^p(A)$ | $0.391 = p_1. \neq p_2. = 0.609$ | $0.5 = p_1. = p_2. = 0.5$ | $0.5 = p_1. = p_2. = 0.5$ |
| $H_0^\mu(D)$ | $0.5 = \mu_1 = \mu_2 = 0.5$ | $0 = \mu_1 \neq \mu_2 = 1$ | $0.5 = \mu_1 = \mu_2 = 0.5$ |
| $H_0^p(D)$ | $0.5 = p_1 = p_2 = 0.5$ | $0.38 = p_1 \neq p_2 = 0.62$ | $0.4914 = p_1 \neq p_2 = 0.5086$ |
| $H_0^\mu(AD)$ | $0 = \mu_{11} - \mu_{12} = \mu_{21} - \mu_{22} = 0$ | $-1 = \mu_{11} - \mu_{12} = \mu_{21} - \mu_{22} = -1$ | $-1 = \mu_{11} - \mu_{12} \neq \mu_{21} - \mu_{22} = 1$ |
| $H_0^p(AD)$ | $0 = p_{11} - p_{12} = p_{21} - p_{22} = 0$ | $-0.2391 = p_{11} - p_{12} \neq p_{21} - p_{22} = -0.2048$ | $-0.2391 = p_{11} - p_{12} \neq p_{21} - p_{22} = 0.2048$ |

$H_0^p(D)$ nicht mehr korrekt. Die Alternative, in der man sich dann befindet, ist zwar minimal, aber lässt sich im Fall der Normalverteilungsannahme exakt angeben. Umgekehrt kann man natürlich auch Fälle konstruieren, in denen die nichtparametrischen Hypothesen weiter gelten und die parametrischen falsch werden. Z.B. gilt für $\mu = (0.6591, 0, 1, 0)'$ und $V_1 = \frac{1}{5}V_2$, dass $p = (0.5575333, 0.4155335, 0.5844706, 0.4424626)$ und

$$\begin{aligned}
 0.6591 &= \mu_1. \neq \mu_2. = 1 & 0.4865334 &= p_1. \neq p_2. = 0.5134666 \\
 0.82955 &= \mu_1 \neq \mu_2 = 0 & 0.571002 &= p_1 \neq p_2 = 0.428998 \\
 0.6591 &= \mu_{11} - \mu_{12} \neq \mu_{21} - \mu_{22} = 1 & 0.142 &= p_{11} - p_{12} = p_{21} - p_{22} = 0.142,
 \end{aligned}$$

also ist lediglich die nichtparametrische Nullhypothese $H_0^p(AD)$ wahr, während die entsprechende parametrische Hypothese $H_0^\mu(AD)$ nicht richtig ist.

Ein interessantes praktisches Beispiel, das an dieser Stelle angeführt werden soll und eben eine solche Situation beinhaltet, stellt die Herzklappenstudie zur Untersuchung der Verkalkung von Ersatz-Herzklappen dar (Mirzaie, Brunner, Mahbub-ul Latif, Schmitto, Fatehpur, und Schöndube, 2007). Gemessen wurde der Kalkgehalt in mg in der Herzklappe jeweils 12 Wochen, nachdem sie der Ratte eingesetzt worden war (Tabelle 13). Untersucht wurden zwei Typen von Herzklappen - Mosaic und Intact - (Faktor A), die sich in der Herstellungsweise unterscheiden. Konserviert wurden die Herzklappen jeweils in zwei verschiedenen Lösungen, GA und MMS (Faktor D). Jeder Gruppe wurden 10 Individuen zugeteilt. Es handelt sich hier

Tabelle 5: Mittelwerte und relative Effekte der Herzklappenstudie

| | | GA | MMS |
|-------------|--------|----------|----------|
| $\hat{\mu}$ | Mosaic | 3,60 mg | 1,80 mg |
| | Intact | 66,00 mg | 34,00 mg |
| \hat{p} | Mosaic | 0.3625 | 0.1375 |
| | Intact | 0.8750 | 0.6250 |

7. Simulationen

also auch um ein 2×2 Design. In der Analyse zeigt sich nun, dass Tests, die die parametrischen Hypothesen testen, sowohl die beiden Haupteffekte also auch einen Wechselwirkungseffekt entdecken. Nichtparametrische Tests, die über den relativen Effekt und die entsprechenden Hypothesen testen, decken ebenso die Haupteffekte auf, jedoch keinen Wechselwirkungseffekt. Ein Blick auf Tabelle 5 liefert eine Erklärung: Während in der Intact-Gruppe im Mittel ein Unterschied der Verkalkung zwischen GA und MMS von 32 mg zu sehen ist, fällt dieser in der Mosaic-Gruppe mit lediglich 1.8 mg sehr gering aus. Das liegt jedoch daran, dass die Werte in der Intact-Gruppe von vornherein relativ hoch sind. So ist z.B. in der Mosaic-Gruppe gar kein derartig hoher Abfall möglich, da der Mittelwert für GA nur bei 3.6 mg liegt und der Wert nicht unter Null fallen kann. Eine genauere Betrachtung verrät jedoch, dass es sich beim MMS-Wert jeweils um ungefähr die Hälfte des GA-Wertes handelt. Es liegt also ein multiplikativer Effekt vor. Logarithmiert man die Daten vor der Analyse, so beträgt die Differenz in der Mosaic-Gruppe 0.69 und in der Intact-Gruppe 0.66 und auch der parametrische Test findet keine Wechselwirkung, sondern nur noch die Haupteffekte. Die nichtparametrischen Tests liefern in beiden Fällen das gleiche Ergebnis, da der relative Effekt invariant unter ordnungserhaltenden Transformationen ist (vgl. Abbildung 10). Dies ist also ein praktisches Beispiel für die oben diskutierten Situationen, bei denen die parametrischen und nichtparametrischen Hypothesen nicht übereinstimmen. Bei den folgenden Simulationen sollte man also immer daran denken, dass die nichtparametrischen Tests beim parametrischen Hinzufügen von Zeiteffekt und/oder Wechselwirkungseffekt teilweise unter einer minimalen Alternative durchgeführt werden. Dies muss hier jedoch in Kauf genommen werden.

Die Betrachtungen dazu werden im 2×2 Design stattfinden, da sie sehr zeitaufwendig sind und der Zeitaufwand mit der Dimension steigt. Zudem lässt sich das additive Hinzufügen von Effekten im 2×2 Design noch besser kontrollieren und koordinieren. Hier wird nun zusätzlich Hotellings T^2 -Test betrachtet, ein in der Praxis häufig verwendetes parametrisches Verfahren. Begonnen wird mit einem balancierten 2×2 Design, bei dem die Daten mit der multivariaten Normalverteilung generiert werden. Getestet werden Stichprobenumfänge von $n_1 = n_2 = 10$ sowie $n_1 = n_2 = 20$. Tabelle 6 listet die simulierten Fehlerraten für die einzelnen Effekte, wobei jeder Abschnitt der Tabelle ein eigenes Setting darstellt, wie oben angesprochen. Beispielsweise wurden die Daten im Abschnitt *Gruppeneffekt A* mit Zeiteffekt und Wechselwirkungseffekt erzeugt, aber ohne Gruppeneffekt. Dann wurde das Niveau der Tests auf *kein Gruppeneffekt A* empirisch bestimmt.

Wertet man diese Resultate aus, so sieht man, dass MCTP bezüglich des Gruppeneffektes das Niveau recht gut einhält, wohingegen ATS und ANOVA leicht liberal sind. Wie erwartet verbessern die größeren Stichprobenumfänge dies. Beim Zeiteffekt ist MCTP leicht konservativ, ANOVA bleibt minimal liberal, genauso wie ATS. Nahezu identisch ist die Situation bei der Wechselwirkung. Das beste empirische Niveau zeigt Hotellings T^2 bei allen drei Effekten, wobei dies nicht überraschend ist, wenn man bedenkt, dass es sich hier um einen exakten Test handelt und alle Annahmen, die er erfordert, erfüllt sind (Normalität, gleiche Kovarianzmatrizen). Versucht man zwischen den Kovarianzstrukturen CS, AR(0.6) und TPL einen Unterschied festzustellen, so lässt sich lediglich

7. Simulationen

festhalten, dass hier die autoregressive Struktur die schlechtesten Fehlerraten verschuldet. Hier ist allerdings zu bemerken, dass es sich im 2×2 Design bei allen drei Strukturen um Compound Symmetry Strukturen handelt, lediglich mit unterschiedlich starken Korrelationen.

An dieser Stelle wird das soeben besprochene Setting erneut simuliert mit dem einzigen Unterschied, dass die Daten nun von einer multivariaten Lognormalverteilung stammen (Tabelle 11, Anhang).

Vergleicht man diese Ergebnisse mit den vorherigen Resultaten, deren Daten mit einer multivariaten Normalverteilung erzeugt wurden, stellt man fest, dass die Lognormalverteilung bei den nichtparametrischen Tests (MCTP, ATS) keinen großen Einfluss hat. Die Fehlerraten unterscheiden sich nahezu nicht. Bei den parametrischen Tests (ANOVA, T^2) zeigt der Wechsel der Verteilungen jedoch seine Wirkung. Die Annahme der Normalverteilung ist nicht mehr erfüllt und die Tests werden leicht konservativ, am besten bei Hotellings T^2 zu beobachten.

Nachdem nun das empirische Niveau für die einzelnen globalen Tests im balancierten Fall so untersucht wurde, dass jeweils die anderen beiden Effekte vorhanden waren, soll dies nun auch im unbalancierten Fall geschehen.

Tabelle 6: Simulierte Fehlerraten des Fehlers erster Art zum nominalen Niveau $\alpha = 0.05$ in 10000 Simulationsläufen bei gleichen Stichprobenumfängen und gleichen Skalierungsfaktoren der Kovarianzmatrizen. Erzeugung der Daten erfolgt mittels **multivariater Normalverteilung** und derart, dass jeweils nur die zwei anderen Effekte vorhanden sind, wenn auf den dritten Effekt getestet wird.

| Balanced 2×2 Design | | | | | | | | |
|--|---------------------|--------|--------|--------|------------------|--------|--------|--------|
| Kovarianzstruktur | MCTP | ATS | ANOVA | T^2 | MCTP | ATS | ANOVA | T^2 |
| | $n_1 = n_2 = 10$ | | | | $n_1 = n_2 = 20$ | | | |
| | Gruppeneffekt A | | | | | | | |
| CS | 0.0494 | 0.0665 | 0.0595 | 0.0554 | 0.0515 | 0.0599 | 0.0593 | 0.0494 |
| AR(0.6) | 0.0534 | 0.0694 | 0.0618 | 0.0521 | 0.0541 | 0.0598 | 0.0585 | 0.0495 |
| TPL | 0.055 | 0.0718 | 0.0604 | 0.0529 | 0.0515 | 0.0583 | 0.0551 | 0.0485 |
| | Zeiteffekt D | | | | | | | |
| CS | 0.0446 | 0.0542 | 0.0603 | 0.0526 | 0.0483 | 0.0575 | 0.0541 | 0.0467 |
| AR(0.6) | 0.04 | 0.0509 | 0.0605 | 0.0529 | 0.0458 | 0.0536 | 0.0588 | 0.0483 |
| TPL | 0.0447 | 0.0514 | 0.067 | 0.0528 | 0.0468 | 0.051 | 0.0579 | 0.0494 |
| | Wechselwirkung AD | | | | | | | |
| CS | 0.041 | 0.0539 | 0.0625 | 0.0454 | 0.0452 | 0.0546 | 0.0569 | 0.0513 |
| AR(0.6) | 0.0378 | 0.0515 | 0.0637 | 0.0485 | 0.0469 | 0.051 | 0.058 | 0.0493 |
| TPL | 0.0398 | 0.0504 | 0.0662 | 0.0476 | 0.0457 | 0.0541 | 0.056 | 0.0504 |

7. Simulationen

Begonnen wird wieder mit Daten, die von einer multivariaten Normalverteilung stammen. Die Stichprobengrößen sind $n_1 = 10$ und $n_2 = 20$ und mit Hilfe der Skalierungsfaktoren für die Kovarianzmatrizen ergeben sich drei Spalten: Mit $\sigma_1 = \sigma_2 = 1$ hat man den einfachen unbalancierten Fall mit gleichen Kovarianzmatrizen, $\sigma_1 = 1, \sigma_2 = 2$ sorgt für den Fall des *positive pairing* und $\sigma_1 = 2, \sigma_2 = 1$ entsprechend für *negative pairing*. Die simulierten Fehlerraten befinden sich in Tabelle 7. Betrachtet man den ersten der drei Blöcke, also den einfachen unbalancierten Fall, so lässt sich feststellen, dass MCTP bei allen drei Effekten ein recht gutes empirisches Niveau zeigt. Sehr gute Ergebnisse liefert T^2 , da hier wieder alle Annahmen erfüllt sind, wohingegen ATS und auch ANOVA ein wenig liberal werden. Im zweiten Block - positive pairing - zeigen sich bei MCTP und ATS bezüglich des Zeit- und Wechselwirkungseffektes ziemlich liberale Resultate. Dies muss daran liegen, dass dort nicht mehr H_0^p gilt, sondern eine Alternative, die sich hier bemerkbar macht (vgl. Diskussion oben). D.h., das parametrische Hinzufügen eines Zeiteffektes resultiert in einem minimalen nichtparametrischen Wechselwirkungseffektes bzw. das parametrische Hinzufügen eines Wechselwirkungseffektes sorgt für einen leichten nichtparametrischen Zeiteffekt. Hier zeigt ANOVA das beste, wenn auch leicht liberale Verhalten. Hotellings T^2 ist konservativ. Im dritten Block, dem negative pairing, zeigt sich bei den nichtparametrischen Verfahren ein ähnliches Verhalten. MCTP und ATS sind relativ liberal. Die parametrischen Methoden sind ebenso liberal. Die schlechtesten Resultate liefert T^2 . Was die Kovarianzstrukturen betrifft lässt sich einzig herausheben, dass die autoregressive Struktur für die extremsten Ergebnisse verantwortlich ist.

Tabelle 7: Simulierte Fehlerraten des Fehlers erster Art zum nominalen Niveau $\alpha = 0.05$ in 10000 Simulationsläufen bei ungleichen Stichprobenumfängen. Erzeugung der Daten erfolgt mittels **multivariater Normalverteilung** und derart, dass jeweils nur die zwei anderen Effekte vorhanden sind, wenn auf den dritten Effekt getestet wird.

| Unbalanced 2×2 Design, $n_1 = 10, n_2 = 20$ | | | | | | | | | | | | |
|--|------------------------------|--------|--------|--------|------------------------------|--------|--------|--------|------------------------------|--------|--------|--------|
| Kovarianzstruktur | MCTP | ATS | ANOVA | T^2 | MCTP | ATS | ANOVA | T^2 | MCTP | ATS | ANOVA | T^2 |
| | $\sigma_1 = 1, \sigma_2 = 1$ | | | | $\sigma_1 = 1, \sigma_2 = 2$ | | | | $\sigma_1 = 2, \sigma_2 = 1$ | | | |
| | Gruppeneffekt A | | | | | | | | | | | |
| CS | 0.0538 | 0.0643 | 0.0616 | 0.0479 | 0.0539 | 0.0593 | 0.0617 | 0.0381 | 0.0472 | 0.0634 | 0.0644 | 0.0807 |
| AR(0.6) | 0.0525 | 0.0649 | 0.0603 | 0.0471 | 0.0513 | 0.0608 | 0.0539 | 0.0393 | 0.0594 | 0.0783 | 0.0691 | 0.0921 |
| TPL | 0.0581 | 0.0709 | 0.0642 | 0.047 | 0.0528 | 0.0623 | 0.0593 | 0.039 | 0.0599 | 0.0759 | 0.0696 | 0.0924 |
| | Zeiteffekt D | | | | | | | | | | | |
| CS | 0.0481 | 0.0607 | 0.0633 | 0.0478 | 0.0723 | 0.0754 | 0.0553 | 0.0381 | 0.0621 | 0.0734 | 0.0724 | 0.0895 |
| AR(0.6) | 0.0468 | 0.061 | 0.0622 | 0.047 | 0.1038 | 0.0984 | 0.0576 | 0.0405 | 0.0692 | 0.0814 | 0.0672 | 0.09 |
| TPL | 0.0472 | 0.0605 | 0.0634 | 0.0497 | 0.0739 | 0.0763 | 0.0593 | 0.0389 | 0.0593 | 0.073 | 0.0675 | 0.0889 |
| | Wechselwirkung AD | | | | | | | | | | | |
| CS | 0.0462 | 0.0603 | 0.0618 | 0.0473 | 0.07 | 0.0717 | 0.0626 | 0.0402 | 0.0574 | 0.076 | 0.0756 | 0.0921 |
| AR(0.6) | 0.0441 | 0.0565 | 0.0574 | 0.05 | 0.0998 | 0.0895 | 0.0574 | 0.0377 | 0.0709 | 0.0815 | 0.0734 | 0.0858 |
| TPL | 0.0488 | 0.0587 | 0.0612 | 0.0506 | 0.0683 | 0.0724 | 0.06 | 0.0372 | 0.0571 | 0.0743 | 0.0708 | 0.0888 |

7. Simulationen

Die gleiche Analyse wurde - wie auch im balancierten Fall schon gesehen - mit multivariat lognormalverteilten, simulierten Daten wiederholt. Die simulierten Fehlerraten (Tabelle 8) unterscheiden sich ein wenig von denen, die mit der multivariaten Normalverteilung erhalten wurden. Im einfachen unbalancierten Fall zeigen MCTP und ANOVA bezüglich des Gruppeneffektes leicht liberale und bezüglich Zeit- und Wechselwirkungseffektes leicht konservative Ergebnisse. ATS ist durchweg leicht liberal und T^2 konservativ. Im Fall des positive pairing ist MCTP bei allen drei Effekten leicht liberal. Das andere nichtparametrische Verfahren, ATS, zeigt auch ein liberales Verhalten, genauso wie T^2 . Einzig ANOVA liefert recht gute empirische Niveaus, wenn auch leicht konservativ bei der Wechselwirkung. Beim negative pairing sind die Ergebnisse ähnlich, einzig T^2 ist nun recht konservativ. Auffallend ist, dass alle vier Verfahren beim Simulieren mit lognormalverteilten Daten bezüglich des Gruppeneffektes leichte Probleme bekommen, d.h. vergleichsweise relativ liberal werden. Dieses Verhalten ist so offensichtlich, dass man davon ausgehen muss, dass hier bei Generieren der Daten eine leichte Alternative entstanden ist.

Tabelle 8: Simulierte Fehlerraten des Fehlers erster Art zum nominalen Niveau $\alpha = 0.05$ in 10000 Simulationsläufen bei ungleichen Stichprobenumfängen. Erzeugung der Daten erfolgt mittels **multivariater Lognormalverteilung** und derart, dass jeweils nur die zwei anderen Effekte vorhanden sind, wenn auf den dritten Effekt getestet wird.

| Unbalanced 2×2 Design, $n_1 = 10, n_2 = 20$ | | | | | | | | | | | | |
|--|------------------------------|--------|--------|--------|------------------------------|--------|--------|--------|------------------------------|--------|--------|--------|
| Kovarianzstruktur | MCTP | ATS | ANOVA | T^2 | MCTP | ATS | ANOVA | T^2 | MCTP | ATS | ANOVA | T^2 |
| | $\sigma_1 = 1, \sigma_2 = 1$ | | | | $\sigma_1 = 1, \sigma_2 = 2$ | | | | $\sigma_1 = 2, \sigma_2 = 1$ | | | |
| | Gruppeneffekt A | | | | | | | | | | | |
| CS | 0.0567 | 0.0695 | 0.0583 | 0.0432 | 0.055 | 0.068 | 0.0577 | 0.0879 | 0.0742 | 0.0744 | 0.0749 | 0.0306 |
| AR(0.6) | 0.0639 | 0.0745 | 0.0578 | 0.0402 | 0.1005 | 0.0968 | 0.0493 | 0.0554 | 0.0984 | 0.0954 | 0.0739 | 0.0239 |
| TPL | 0.0622 | 0.0742 | 0.0464 | 0.0349 | 0.0636 | 0.0804 | 0.0498 | 0.0925 | 0.0863 | 0.0848 | 0.0738 | 0.0183 |
| | Zeiteffekt D | | | | | | | | | | | |
| CS | 0.0475 | 0.0562 | 0.0481 | 0.0448 | 0.0528 | 0.0615 | 0.0501 | 0.0478 | 0.0564 | 0.0646 | 0.0522 | 0.041 |
| AR(0.6) | 0.0448 | 0.0583 | 0.0429 | 0.0385 | 0.0575 | 0.0669 | 0.0549 | 0.0879 | 0.0616 | 0.0664 | 0.0509 | 0.0399 |
| TPL | 0.0449 | 0.0562 | 0.035 | 0.0336 | 0.073 | 0.0714 | 0.0416 | 0.0497 | 0.0551 | 0.065 | 0.0404 | 0.0342 |
| | Wechselwirkung AD | | | | | | | | | | | |
| CS | 0.0439 | 0.0607 | 0.0467 | 0.0419 | 0.0451 | 0.0584 | 0.0496 | 0.0528 | 0.0523 | 0.063 | 0.0506 | 0.0383 |
| AR(0.6) | 0.0408 | 0.0563 | 0.0416 | 0.0378 | 0.0527 | 0.0651 | 0.0433 | 0.0482 | 0.0591 | 0.0682 | 0.0497 | 0.0371 |
| TPL | 0.0431 | 0.0568 | 0.0358 | 0.0348 | 0.0485 | 0.0606 | 0.0377 | 0.0485 | 0.0485 | 0.0589 | 0.0422 | 0.0349 |

7. Simulationen

Zum Abschluss der Analyse des Niveaus sollen an dieser Stelle noch zwei höherdimensionale Designs, ein 3×3 und ein 2×5 Design, auf die Art und Weise analysiert werden, wie sie eben schon im 2×2 Design angewandt wurde. Exemplarisch wird dabei der Test auf $H_0(AD)$ (*keine Wechselwirkung*) untersucht, während die Daten mittels multivariater Normalverteilung so simuliert werden, dass sowohl ein Gruppen- als auch ein Zeiteffekt vorhanden ist. Es werden fünf Situationen betrachtet: balancierte Stichprobenumfänge mit $n_i = 10$, balancierte Stichprobenumfänge mit $n_i = 20$, unbalancierte Stichprobenumfänge mit $n_1 = 10, n_2 = 15, n_3 = 20$ bzw. $n_1 = 10, n_2 = 20$ und dazu positive pairing mit $\sigma_1 = 1, \sigma_2 = 1.5, \sigma_3 = 2$ bzw. $\sigma_1 = 1, \sigma_2 = 2$ und negative pairing mit umgekehrten Skalierungsfaktoren. Als Kovarianzstruktur wurde die autoregressive Struktur AR(0.6) gewählt und zur besten Vergleichbarkeit mit der ANOVA wurde als Kontrastmatrix der Average-Kontrast verwendet.

Tabelle 9 zeigt die Simulationsergebnisse. Im 3×3 Design zeigen alle drei Verfahren in den fünf Situationen ein minimal konservatives Verhalten, am stärksten jedoch MCTP, vor allem bei kleinen Stichprobenumfängen und unbalancierten Daten. Die Performance von ATS und ANOVA ist daher leicht besser. Im 2×5 Design zeigt MCTP dann durchweg bessere Ergebnisse als ATS und ANOVA, wenn auch leicht konservativ, bis auf den Fall des positive pairing, dort leicht liberal. Die ATS ist recht konservativ, mit Ausnahme des positive pairing, und die ANOVA ist durchgehend konservativ. Die besten Ergebnisse in den Fällen mit gleichen Kovarianzmatrizen liefert Hotellings T^2 , aber auch das schlechteste beim negative pairing, also ungleichen Kovarianzmatrizen.

Tabelle 9: Simulierte Fehlerraten des Fehlers erster Art zum nominalen Niveau $\alpha = 0.05$ in 10000 Simulationsläufen im 2×2 und 3×3 Design. Untersucht wird der Test auf *keine Wechselwirkung AD* bei vorhandenem Gruppen- und Zeiteffekt. Erzeugung der Daten erfolgt mittels multivariater Normalverteilung.

| Verschiedene Settings, Kovarianzstruktur: AR(0.6), Kontrast: Average | | | | | | | |
|---|---------------------|--------|--------|---------------------|--------|--------|--------|
| Setting | 3×3 Design | | | 2×5 Design | | | |
| | MCTP | ATS | ANOVA | MCTP | ATS | ANOVA | T^2 |
| balanced $n_i = 10$ | 0.037 | 0.0439 | 0.0464 | 0.0417 | 0.0366 | 0.0416 | 0.0495 |
| balanced $n_i = 20$ | 0.0445 | 0.0422 | 0.0424 | 0.0493 | 0.0363 | 0.0379 | 0.0483 |
| unbalanced | 0.0322 | 0.0428 | 0.0416 | 0.0388 | 0.0344 | 0.0365 | 0.0506 |
| positive paired | 0.0452 | 0.0502 | 0.0428 | 0.057 | 0.0502 | 0.0395 | 0.0557 |
| negative paired | 0.0355 | 0.0445 | 0.0429 | 0.0474 | 0.043 | 0.0377 | 0.1379 |

7.3. Power

Die Power eines Tests beschreibt die Wahrscheinlichkeit, einen tatsächlich vorliegenden Unterschied bzw. einen tatsächlich vorhandenen Effekt aufzudecken. In den Powersimulationen muss also immer eine Situation geschaffen werden, in der man beim Generieren der Daten die Kontrolle über die Größe eines gezielt betrachteten Effekts besitzt. Dann kann man diesen Unterschied z.B. in den Erwartungswerten vergrößern und beobachten, wie gut es dem Test gelingt den Effekt zu erkennen. Dazu führt man wieder $n_{sim} = 10000$ Simulationsläufe durch und bestimmt den Anteil der richtigerweise verworfenen Hypothesen. Beispielsweise lässt sich ein Gruppeneffekt A durch unterschiedliche Erwartungswerte beim Erzeugen der Daten generieren. Dann testet man $H_0(A)$ (*kein Gruppeneffekt*) n_{sim} -mal und bestimmt die empirische Wahrscheinlichkeit, dass der Test verworfen wird. Variiert man dann noch die Größe des Unterschieds in den Erwartungswerten und trägt die Ergebnisse in einer Graphik auf, erhält man die typischen Powerkurven.

Die vorliegende Arbeit hält bekanntlich drei mögliche Effekte bereit, bezüglich derer die drei Verfahren aus den Niveausimulationen - MCTP, ATS, ANOVA, T^2 - auf die Power ihrer Globaltests $H_0(A)$ (*kein Gruppeneffekt*), $H_0(D)$ (*kein Zeiteffekt*) und $H_0(AD)$ (*keine Wechselwirkung*) untersucht werden können. Dies geschieht wiederum im 2×2 Design in den bekannten Settings. Als Kovarianzstruktur dient erneut die autoregressive AR(0.6) Struktur. Verglichen wird zudem zwischen Simulationen mittels multivariater Normalverteilung und multivariater Lognormalverteilung.

Den Anfang machen die Powersimulationen für den Haupteffekt A . Der Effekt wird additiv hinzugefügt als $\delta = 0, 0.1, 0.2, \dots, 1.9, 2$ und stellt einen Unterschied in der Erwartungswerten der beiden Gruppen dar. Die Powerkurven sind in Abbildung 5 zu sehen. Die Power aller vier Testverfahren ist vergleichbar gut, wenn die Daten aus einer multivariaten Normalverteilung simuliert werden (linke Spalte). MCTP ist dabei immer minimal besser. Beim Übergang von der ersten zur zweiten Zeile der Abbildung kann man sehr gut sehen, wie sich die Power mit steigender Stichprobengröße verbessert. Betrachtet man die Zeilen drei und vier, also die unbalancierten Settings, lässt sich ein gewisser Powerabfall beobachten, was jedoch zu erwarten war, da auch in den Niveausimulationen die positive und negative pairings die schwierigeren Situationen darstellten. Auffällig ist Hotellings T^2 mit einer viel schlechteren Power im Fall des positive pairing und einer leicht nach oben verschobenen Powerkurve beim negative pairing. Dies lässt sich wieder mit den ungleichen Kovarianzmatrizen erklären. Vergleicht man nun die beiden Spalten, d.h. die Powerkurven für die Daten aus multivariater Normalverteilung und multivariater Lognormalverteilung, kann man für die nichtparametrischen Verfahren (MCTP, ATS) nahezu keine Veränderung beobachten. Anders sieht es bei den parametrischen Verfahren aus: Hier lässt sich eine deutlich schlechtere Power im Vergleich zu den beiden parametrischen Methoden beobachten.

Die gleichen Powersimulationen werden an dieser Stelle noch für den Haupteffekt D angegeben. Der Zeiteffekt wird wie der Gruppeneffekt additiv hinzugefügt, indem jeweils auf den zweiten Zeitpunkt aller Daten $\delta = 0, 0.05, 0.1, \dots, 0.95, 1$ addiert wird. Abbildung 6 zeigt die entsprechenden Powerkurven. Das Resultat stellt sich so wie im Fall des

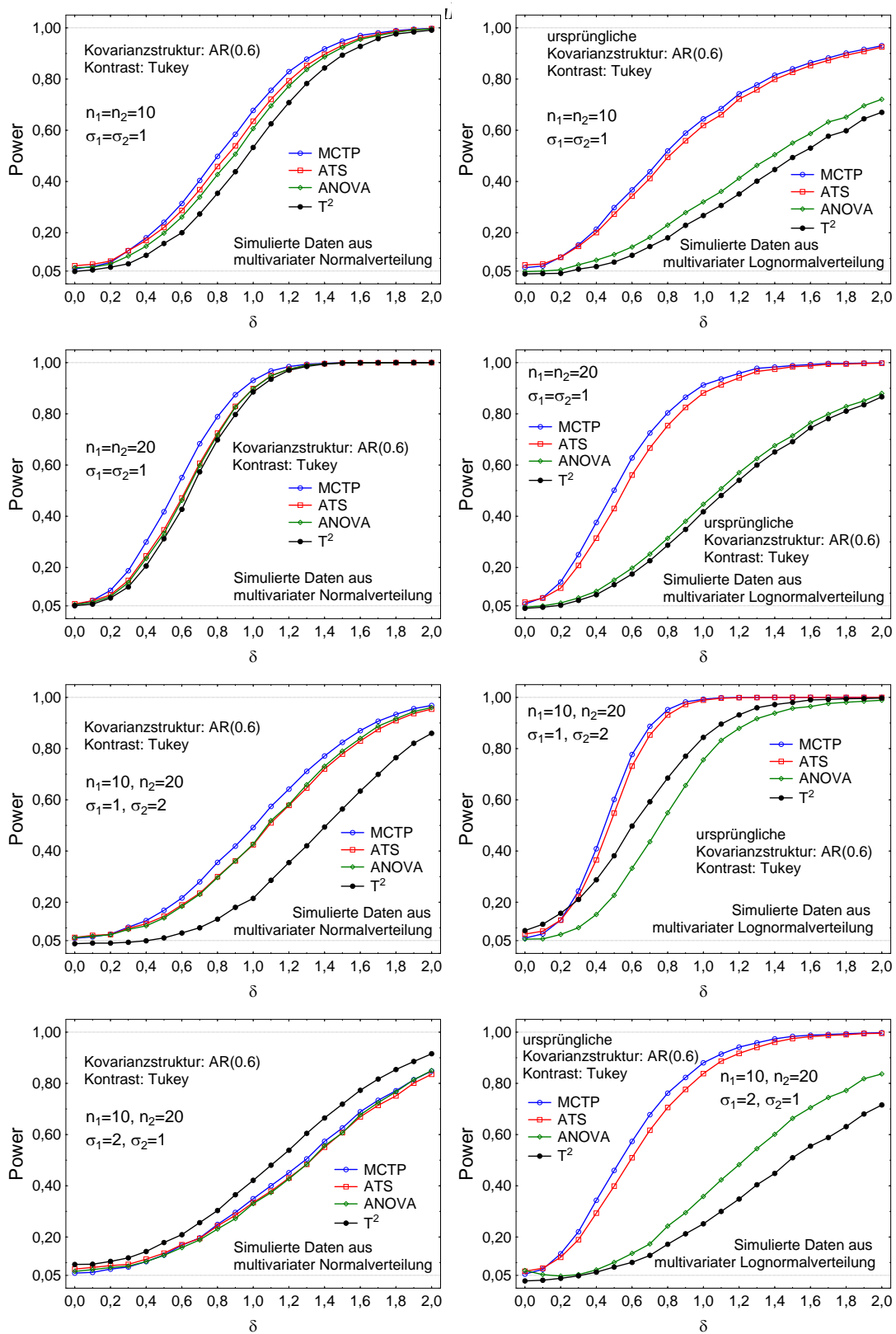


Abbildung 5: Power des Tests auf *kein Haupteffekt A* in unterschiedlichen 2×2 Settings: balanced, positive und negative pairing. Größeres δ bedeutet größerer Gruppeneffekt *A*.

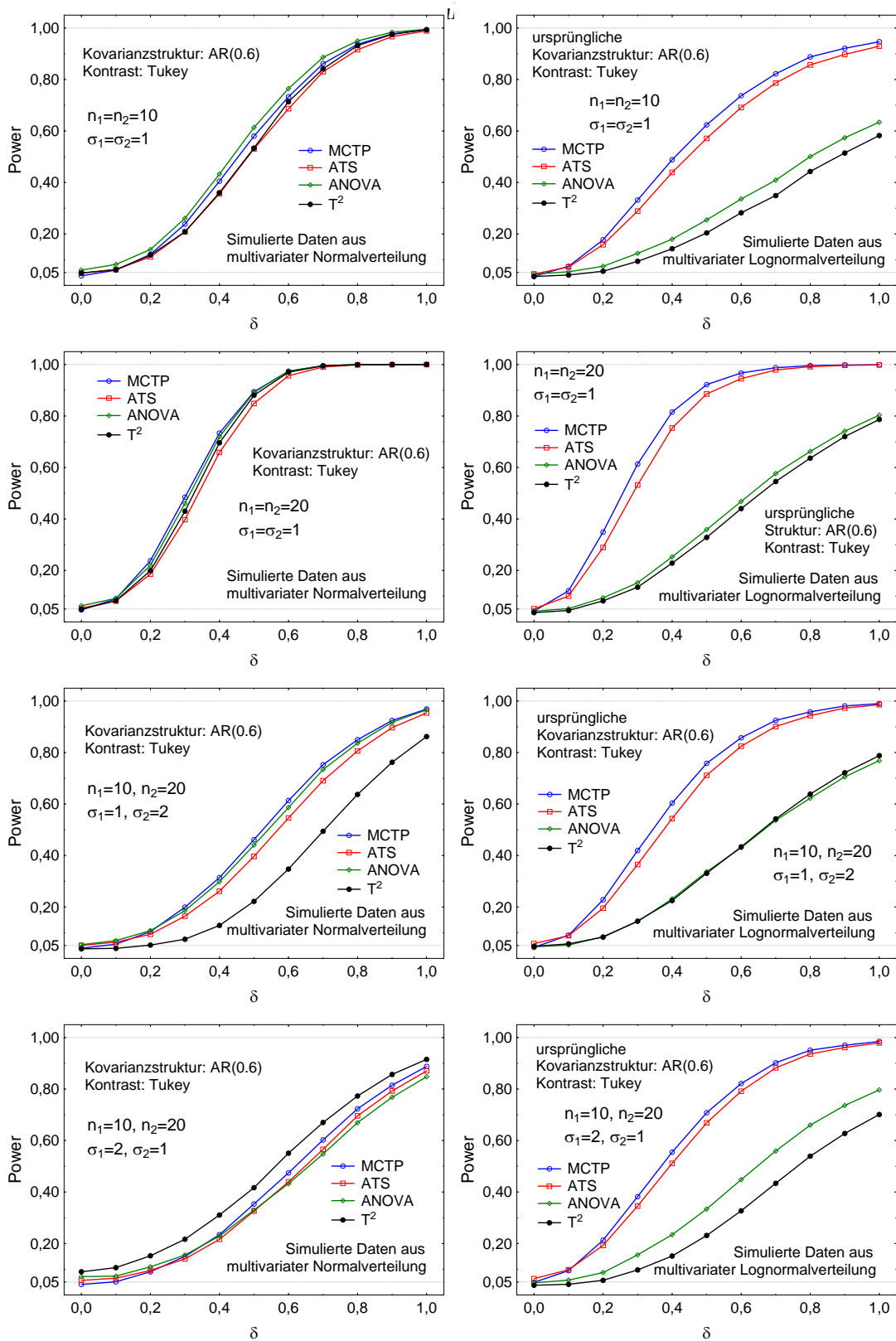


Abbildung 6: Power des Tests auf kein Haupteffekt D in unterschiedlichen 2×2 Settings: balanced, positive und negative pairing. Größeres δ bedeutet größerer Zeiteffekt D .

7. Simulationen

Haupteffektes A dar. Die nichtparametrischen Verfahren übertreffen die parametrischen, falls die Annahme der Normalverteilung nicht mehr gilt (lognormalverteilte Daten). Bei Daten aus der multivariaten Normalverteilung fällt lediglich Hotellings T^2 negativ auf; ansonsten ist kein Verfahren dem anderen überlegen.

Für den Wechselwirkungseffekt werden nur zwei Settings betrachtet (siehe Abbildung 7): Die Graphik links zeigt die Powerkurve für das balancierte Design mit Stichprobenumfängen von $n_i = 20$ sowie die Kurve für das unbalancierte *positive pairing* bei Daten aus einer Normalverteilung. Rechts sind wieder entsprechende Situationen mit Daten aus der Lognormalverteilung dargestellt. Die Schaubilder zeigen das in dieser Simulationsstudie gewohnte Bild. Im Fall mit der Normalverteilung haben die vier Verfahren eine vergleichbar gute Power. Bei lognormalverteilten Daten haben MCTP und ATS als nichtparametrische Tests eine deutlich höhere Power als ANOVA und T^2 (parametrisch).

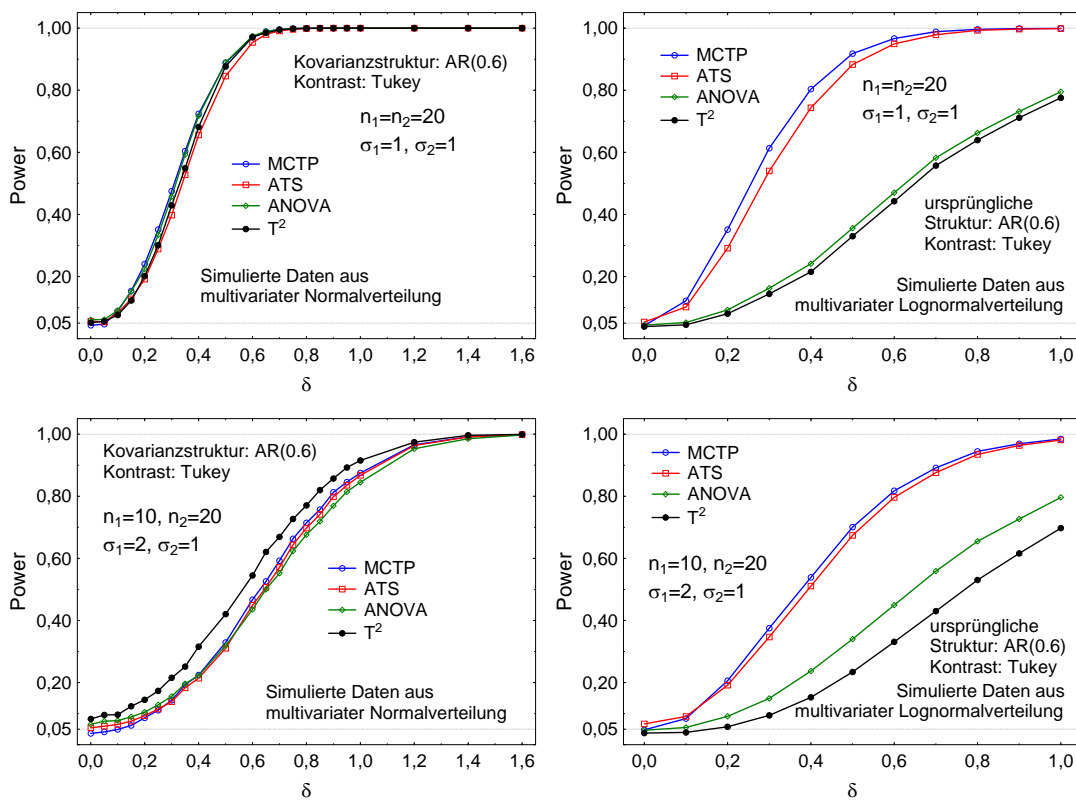


Abbildung 7: Power des Tests auf keine Wechselwirkung AD in unterschiedlichen 2×2 Settings: balanced und negative pairing. Größeres δ bedeutet größerer Wechselwirkungseffekt AD .

7. Simulationen

Insgesamt kann man die Resultate der vorgestellten Simulationsstudie wie folgt zusammenfassen: Sowohl in den Niveausimulationen als auch in den Powersimulationen konnte man feststellen, dass in den zahlreichen Settings unter der Normalverteilungsannahme kein Verfahren einen bedeutenden Vorteil gegenüber den anderen Verfahren bietet. Sowohl das Testverfahren mit multiplen Kontrasttests, als auch die ANOVA-Typ Statistik, parametrische ANOVA und Hotellings T^2 haben in gewissen Settings leichte Probleme, alles in allem halten sie jedoch das Niveau - mal ein wenig konservativ oder liberal - ein. Auch die Simulationen zur Power sind zufriedenstellend. Lässt man die für die parametrischen Verfahren elementare Annahme der Normalverteilung weg, übertreffen die nichtparametrischen Verfahren (MCTP und ATS) die parametrischen Methoden vor allem bei der Power. Die Performance von MCTP und ATS ist auf ähnlich gutem Niveau. Wenn man aber bedenkt, dass für MCTP die Globaltests nur Nebenprodukte aus den multiplen Vergleichen sind, wohingegen ATS, ANOVA und Hotellings T^2 lediglich Globaltests bieten, ist der klare Vorteil von MCTP und dessen zusätzlicher Informationsgewinn offensichtlich.

8. Auswertung des Beispiels

Zur Auswertung von realen Datensätzen mit der in dieser Arbeit präsentierten Methode wurde in der Statistik-Software R eine Funktion `f1_mctp_f1` implementiert, die sich für Versuchspläne mit einem whole-Plot Faktor (Behandlung, Gruppe) gekreuzt mit einem sub-Plot Faktor (Repeated Measures, Zeit) eignet. Sie eignet sich somit für die Analyse des in Abschnitt 1.2 vorgestellten Beispieldatensatzes. Dort war die Behandlung (Herz-, Lungen- und Gefäßoperation) gekreuzt mit dem Faktor Zeit (4 Zeitpunkte). Von Interesse waren die Profilverläufe der PCT-Konzentration in den drei Behandlungsgruppen, insbesondere der Vergleich von herz- mit lungen- und gefäßoperierten Patienten. Die Wechselwirkung ist also genauer zu untersuchen. Die Daten müssen derartig eingelesen werden, dass eine Spalte die Zielvariable, in diesem Fall die PCT-Konzentrationen, enthält, eine Spalte einen Vektor, der die Gruppenzugehörigkeit angibt, und eine Spalte, die die Repeated Measures indiziert. Eingelesen werden kann beispielsweise über eine Textdatei, die sich im aktuellen *working directory* befinden muss:

```
R> pctstudy <- read.table("pctstudy.txt", header=TRUE)
R> pctstudy
  group time  pct
1     1   1  0.2
2     1   2  0.2
3     1   3  3.6
4     1   4  2.6
5     1   1  0.2
6     1   2  0.2
7     1   3  0.4
...
151    1   3  2.6
152    1   4  2.1
153    2   1  0.3
154    2   2  0.2
...
283    3   3  0.6
284    3   4  0.4
```

Dann kann auch schon direkt die Funktion eingeladen (*working directory*) und mit der richtigen Formel aufgerufen werden:

```
R> source("f1_mctp_f1.r")
R> f1_mctp_f1( formula = pct~group*time, data = pctstudy ,
              type = "Tukey", alternative = "two.sided",
              conf.level = 0.95, asy.method = "fisher", plot.simci = TRUE )
```

Hier wurde für die multiplen Vergleiche jeweils der all-pairs Tukey-Kontrast gewählt. Zur Verfügung stehen natürlich auch jegliche in dieser Arbeit angeführten Kontraste (Dunnett, Average, Changepoint) sowie einige mehr. Die Tests werden zweiseitig durchgeführt. Möglich wären hier auch einseitige ("**greater**", "**less**"). Als Konfidenzlevel ist 95% gewählt und als asymptotische Methode "**fisher**". Dabei handelt es sich um die Approximation mit der multivariaten t-Verteilung, kombiniert mit der Fisher-Transformation für bereichserhaltende Konfidenzintervalle. Die Option `plot.simci` fragt ab, ob ein Plot der Ergebnisse gewünscht ist. Die Ausgabe lautet wie folgt:

8. Auswertung des Beispiels

\$Data.Info

| | Sample Size | Effect.1 | Effect.2 | Effect.3 | Effect.4 | |
|---|-------------|----------|-----------|-----------|-----------|-----------|
| 1 | 1 | 38 | 0.2548294 | 0.3038660 | 0.8523119 | 0.8072751 |
| 2 | 2 | 17 | 0.3134035 | 0.3045912 | 0.6525892 | 0.7062528 |
| 3 | 3 | 16 | 0.2819429 | 0.2787396 | 0.6103375 | 0.6338611 |

\$Analysis.Factor.A\$Multiple.Comparisons

| | Estimator | Lower | Upper | Statistic | p.Value |
|-------|-----------|--------|--------|-----------|-------------|
| 2 - 1 | -0.060 | -0.131 | 0.011 | -2.090 | 0.109499241 |
| 3 - 1 | -0.103 | -0.180 | -0.025 | -3.260 | 0.008958075 |
| 3 - 2 | -0.043 | -0.143 | 0.058 | -1.055 | 0.540975885 |

\$Analysis.Factor.A\$Overall

| | Quantile | p.Value | p.Value.WTS | p.Value.ATS | p.parANOVA |
|---|----------|-------------|-------------|-------------|-------------|
| 1 | 2.482463 | 0.008958075 | 0.001786744 | 0.0278743 | 0.004719797 |

\$Analysis.Factor.D\$Multiple.Comparisons

| | Estimator | Lower | Upper | Statistic | p.Value |
|-------|-----------|--------|-------|-----------|--------------|
| 2 - 1 | 0.012 | -0.030 | 0.055 | 0.791 | 0.8523284 |
| 3 - 1 | 0.422 | 0.339 | 0.455 | 14.678 | 0.000000e+00 |
| 4 - 1 | 0.432 | 0.365 | 0.448 | 20.863 | 0.000000e+00 |
| 3 - 2 | 0.409 | 0.324 | 0.449 | 13.429 | 3.996803e-15 |
| 4 - 2 | 0.420 | 0.350 | 0.442 | 18.429 | 0.000000e+00 |
| 4 - 3 | 0.011 | -0.052 | 0.073 | 0.465 | 0.9645010 |

\$Analysis.Factor.D\$Overall

| | Quantile | p.Value | p.Value.WTS | p.Value.ATS | p.parANOVA |
|---|----------|---------|-------------|-------------|--------------|
| 1 | 2.728174 | 0 | 0 | 0 | 8.721574e-07 |

\$Analysis.Interaction.AD\$Multiple.Comparisons

| | Estimator | Lower | Upper | Statistic | p.Value |
|----|-----------|--------|--------|-----------|--------------|
| 1 | -0.083 | -0.134 | -0.032 | -4.916 | 0.0009519551 |
| 2 | -0.046 | -0.099 | 0.006 | -2.686 | 0.1026193397 |
| 3 | 0.093 | 0.037 | 0.147 | 5.105 | 0.0006655486 |
| 4 | 0.037 | -0.019 | 0.092 | 2.019 | 0.3222310998 |
| 5 | 0.036 | -0.035 | 0.106 | 1.553 | 0.5882085863 |
| 6 | 0.015 | -0.053 | 0.082 | 0.665 | 0.9815470783 |
| 7 | -0.047 | -0.124 | 0.032 | -1.817 | 0.4280168854 |
| 8 | -0.004 | -0.064 | 0.057 | -0.190 | 0.9999799134 |
| 9 | 0.047 | -0.022 | 0.116 | 2.082 | 0.2930854424 |
| 10 | 0.032 | -0.034 | 0.097 | 1.487 | 0.6292075977 |
| 11 | -0.046 | -0.140 | 0.048 | -1.485 | 0.6303837415 |
| 12 | -0.033 | -0.107 | 0.041 | -1.362 | 0.7067000841 |

\$Analysis.Interaction.AD\$Overall

| | Quantile | p.Value | p.Value.WTS | p.Value.ATS | p.parANOVA |
|---|----------|--------------|-------------|-------------|------------|
| 1 | 3.056486 | 0.0006655486 | 1.84478e-06 | 0.002165007 | 0.00338428 |

8. Auswertung des Beispiels

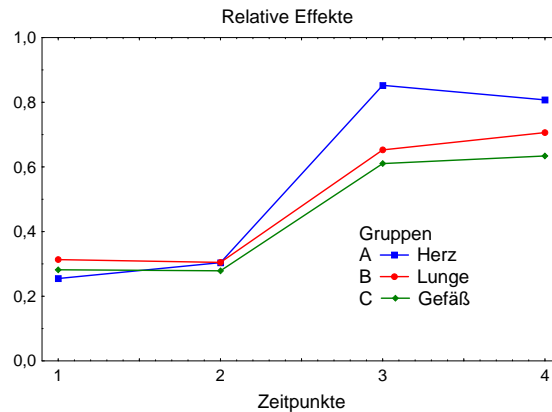


Abbildung 8: Relative Effekte in der PCT-Studie

Die Ausgabe beginnt mit einer Tabelle mit den geschätzten relativen Effekten. Zu jeder Gruppe ist sowohl der Stichprobenumfang angegeben als auch jeweils die relativen Effekte an jedem der $d = 4$ Zeitpunkte. Abbildung 8 veranschaulicht die Profile der relativen Effekte.

Darauf folgen die Analysen der Haupteffekte A (Behandlung) und D (Zeit) sowie des Wechselwirkungseffektes. Jeder Abschnitt besteht dabei aus einer Tabelle zu den multiplen Vergleichen, gefolgt von den p-Werten für den entsprechenden Globaltest. In der Tabelle befindet sich eine Spalte mit den Schätzern für die Vergleiche (z.B. bei Haupteffekt A in Zeile 1 der Wert von $\hat{p}_2 - \hat{p}_1$.) mit den zugehörigen unteren und oberen Grenzen der Konfidenzintervalle. Dazu eine Spalte mit den Teststatistiken und eine mit den zugehörigen p-Werten. Bei den Globaltests sind neben dem p-Wert, der sich als Nebenprodukt aus den multiplen Vergleichen ergibt, zusätzlich die p-Werte der Tests mit nichtparametrischer ANOVA-Typ Statistik (ATS), nichtparametrischer Wald-Typ Statistik (WTS) und parametrischer ANOVA-Typ Statistik (ANOVA) aufgeführt.

Der Globaltest auf *kein Haupteffekt A* wird verworfen mit $p = 0.009$, da der Vergleich von Gruppe 1 (Herz) zu Gruppe 3 (Gefäß) einen signifikanten Unterschied aufdeckt. Die Hypothesen bezüglich der anderen beiden Vergleiche können nicht verworfen werden.

Der Globaltest bezüglich der Hypothese, dass kein Zeiteffekt D vorliegt wird ebenso verworfen ($p < 0.0001$). Hier zeigen nur die Vergleiche von Zeitpunkt 2 mit Zeitpunkt 1 und von Zeitpunkt 4 mit Zeitpunkt 3 keinen signifikanten Unterschied. Hier wird direkt der Vorteil der multiplen Kontrasttests deutlich. Die Globalhypothesen werden verworfen und der Anwender weiß sofort in welchen Faktorstufen die Unterschiede liegen, die zum Verwerfen der Nullhypothese geführt haben.

Bei der Wechselwirkung werden die relativen Effekte, zentriert um Gruppen- und Zeiteffekt, $p_{ij} - \bar{p}_i - \bar{p}_j + \bar{p}.$ gegen Null getestet. Auch hier wird die Globalhypothese verworfen, da die Nullhypothesen bezüglich p_{11} ($H_0^p : p_{11} - \bar{p}_1 - \bar{p}_{.1} + \bar{p} = 0$) und p_{13} verworfen werden.

Abbildung 9 zeigt die zusätzlich ausgegebene Graphik, die für jeden Effekt übersichtlich

8. Auswertung des Beispiels

die Schätzer und deren Konfidenzintervalle darstellt.

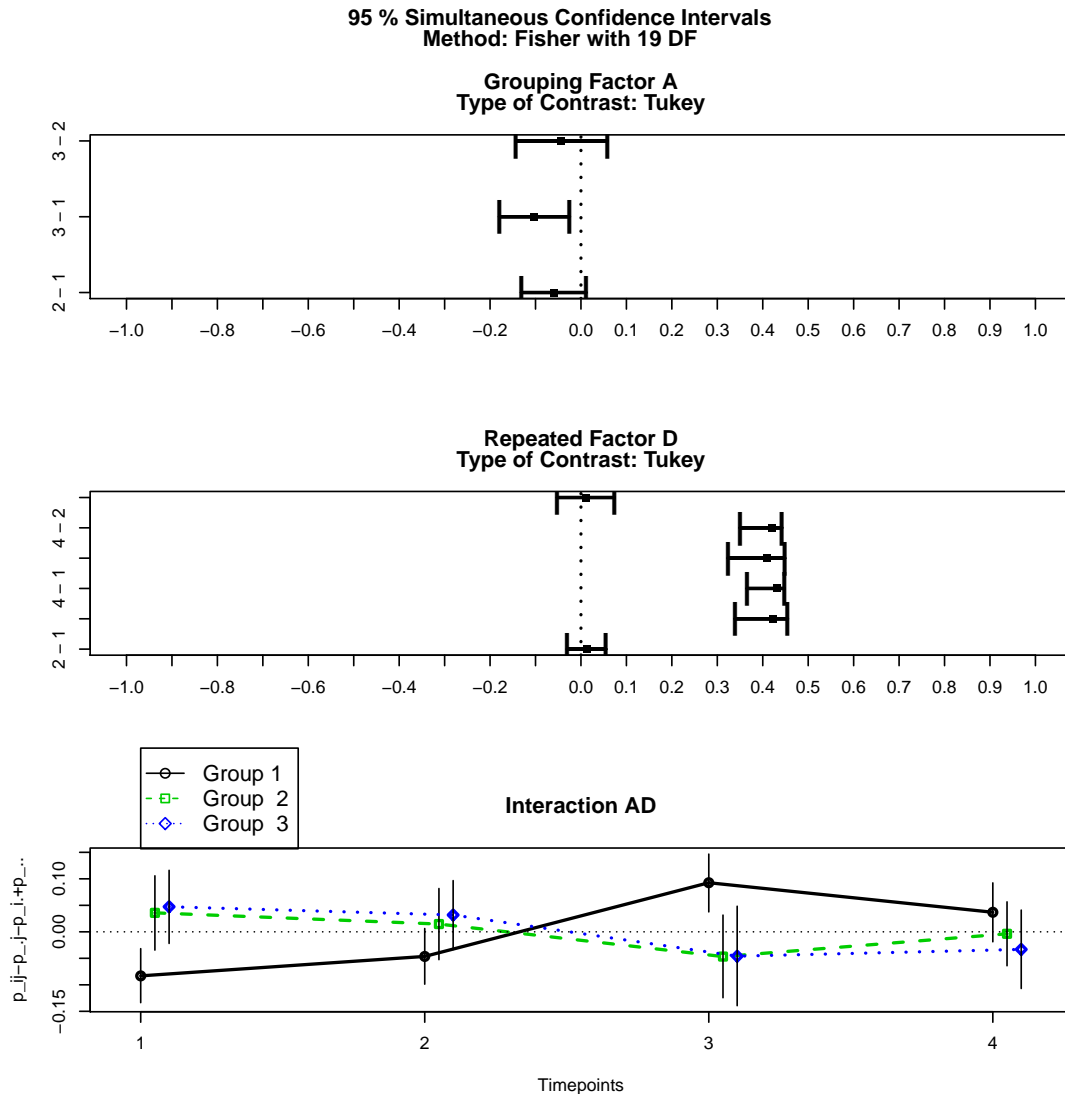


Abbildung 9: Plot der simultanen Konfidenzintervalle in der Auswertung der PCT-Studie.

Hier lässt sich wiederum eine praktische Eigenschaft beobachten. Da die Konfidenzintervalle kompatibel mit den Testentscheidungen sind, gilt: Sobald eine Hypothese verworfen wird, enthält das entsprechende Konfidenzintervall nicht die Null. Man kann also an der Graphik direkt ablesen, welche Hypothesen verworfen werden und somit z.B. sofort feststellen, zwischen welchen Faktorstufen signifikante Unterschiede aufgedeckt wurden. Nun wurde ein Wechselwirkungseffekt aufgedeckt. Den Anwender interessierte jedoch auch insbesondere: Die Profilverläufe welcher Behandlungsgruppen unterscheiden sich?

8. Auswertung des Beispiels

Abbildung 8 legt schon die Vermutung nahe, dass die Profile von Lungen- und Gefäßoperation ähnlich sind und sich von dem der Herzoperation unterscheiden. Dies muss jedoch auch statistisch bestätigt werden. Daher sind bezüglich der Wechselwirkung zusätzlich paarweise Vergleiche zwischen den Gruppen notwendig. Tabelle 10 zeigt die entsprechenden p-Werte. Hier bestätigt sich die Annahme: Nur der Vergleich zwischen Lungen- und Gefäßoperation zeigt keine signifikante Wechselwirkung. Das heißt also, dass sich der Profilverlauf der PCT-Konzentration in der Gruppe der Patienten mit einer Herzoperation von den Profilverläufen in den beiden anderen Gruppen unterscheidet. Man beachte, dass hier auf Grund des Abschlusstest-Prinzips keine Adjustierung des Niveaus vorgenommen werden muss, da die Schnitthyothese (kein Wechselwirkungseffekt zwischen den drei Gruppen) verworfen wurde. An dieser Stelle kann man auch einen Blick auf die Haupteffekte werfen. Der Gruppeneffekt interessiert in diesem Versuch nicht so sehr wie der Zeiteffekt. Wird ein Wechselwirkungseffekt festgestellt, ist die Interpretation der Haupteffekte jedoch schwierig. Beim Vergleich von Lungen- und Gefäßoperation liegt keine Wechselwirkung vor, sodass hier der Zeiteffekt einwandfrei aufzudecken ist ($p_{overall} < 0.0001$). Zu den Zeitpunkten 1 und 2 sind die PCT-Konzentrationen signifikant niedriger als zu den Zeitpunkten 3 und 4. Es ist also ein Anstieg des PCT-Levels zu beobachten.

Tabelle 10: Paarvergleiche für den Wechselwirkungseffekt

| Vergleich | p-Wert |
|---------------|---------|
| Herz ↔ Lunge | 0.00065 |
| Herz ↔ Gefäß | 0.00304 |
| Lunge ↔ Gefäß | 0.83182 |

Um die Auswertung zusammenzufassen, lässt sich somit feststellen: Es kann statistisch gezeigt werden, dass der Verlauf der Serum-Konzentration PCT in den Gruppen signifikant unterschiedlich ist. Dabei liegen Unterschiede zwischen Herz- und Lungenoperation sowie zwischen Herz- und Gefäßoperation. Der entscheidende Ausschlag kann durch die ausgegebenen Graphiken lokalisiert werden zwischen Zeitpunkt 2 und 3. Die signifikanten Haupteffekte der Gruppen und der Zeit sind auf Grund der vorhandenen Wechselwirkung schwer zu interpretieren. Die paarweisen Vergleiche der Gruppen liefern jedoch zumindest beim Vergleich von Lungen- und Gefäßoperation einen signifikanten Anstieg der PCT-Konzentration über die Zeit.

9. Zusammenfassung und Diskussion

In dieser Arbeit wurde eine rangbasierte nichtparametrische Multiple Comparison Test Procedure für faktorielle Repeated Measures Designs vorgestellt. Dazu wurden simultane, kompatible und bereichserhaltende Konfidenzintervalle hergeleitet, wie in der Praxis von Regulationsbehörden gefordert. Das Verfahren stellt eine Verallgemeinerung der Arbeiten von [Konietschke et al. \(2012\)](#) für ein One-Way-Layout mit unverbundenen Stichproben und [Konietschke et al. \(2010\)](#) für ein einfaches Repeated Measures Design mit nur einer Behandlungsgruppe. Der Ausgangspunkt ist eine lineare Rangstatistik basierend auf den transitiven relativen Effekten, deren asymptotische Verteilung eine multivariate Normalverteilung darstellt. Die gemeinsame Verteilung der Teststatistiken für die multiplen Vergleiche wird dementsprechend mit einer multivariaten Normalverteilung approximiert oder, bei kleinen Stichprobenumfängen, mit einer multivariaten t-Verteilung. Konsistente Schätzer für die asymptotische Kovarianz- bzw. Korrelationsmatrix wurden angegeben. Die Konstruktion der Kontrasttests stellt keinerlei Annahmen an die Kovarianzstruktur oder Homoskedastizität in den Gruppen. Zudem ist jede beliebige Kontrastmatrix zur Formulierung von multiplen Vergleichen zugelassen, je nachdem, welche Fragestellung der Anwender zu beantworten versucht.

In einer Simulationsstudie wurde dann das Verfahren mit bekannten und in der Praxis verwendeten Globaltests verglichen, wobei festgestellt werden konnte, dass die neue MCTP den bekannten Verfahren bezüglich der Einhaltung des Fehlerniveaus weder deutlich unter- noch überlegen ist. Vergleicht man die Power, so bietet die nichtparametrische MCTP vor allem unter Nichtnormalität einen deutlichen Vorteil gegenüber parametrischen Verfahren. Ansonsten ist die Güte der Tests auf einem ähnlichen Level. Allerdings darf man nicht vergessen, dass die Globaltests der multiplen Testprozedur lediglich Nebenprodukte aus den multiplen Vergleichen sind. Das heißt also, der große Vorteil der multiplen Testprozedur ist, dass nicht nur eine Signifikanzaussage bezüglich der Globalhypothese getroffen wird. Mittels der multiplen Vergleiche und simultanen Konfidenzintervalle besteht ein beachtlicher zusätzlicher Informationsgewinn über die Behandlungseffekte. Gerade die Auswertung des motivierenden Beispiels im letzten Abschnitt machte dies deutlich und demonstrierte, wie vor allem die kompatiblen Konfidenzintervalle dazu beitragen können, einem Anwender die statistische Analyse verständlich zu vermitteln. Hier ist vor allem die in R implementierte Funktion und deren Option zur Ausgabe der Plots der Behandlungseffekte und deren Konfidenzintervalle zu erwähnen.

Was die Implementation betrifft, so könnte der nächste Schritt darin bestehen, die Funktion `f1_mctp_f1` für faktorielle Repeated Measures Designs in das R-Paket *nparcomp* zu integrieren. Eine weitere Aufgabe für die Zukunft könnte des Weiteren lauten, den Fall zu untersuchen, in dem die Dimension der Repeated Measures gegen Unendlich geht. In der vorliegenden Arbeit wurde diese immer als fest angenommen. Aber gerade in der Medizin ermöglichen es neue Messtechniken, immer mehr wiederholte Messungen durchzuführen. In dieser Richtung besteht somit Bedarf.

A. Anhang

A.1. Wichtige Definitionen und Sätze

Definition A.1. (Konvergenzarten) Sei X_n eine Folge von Zufallsvariablen, $n \geq 1$. Es gibt verschiedene Arten der Konvergenz:

1. Konvergenz in Wahrscheinlichkeit:

Man sagt, X_n konvergiert in Wahrscheinlichkeit gegen die Zufallsvariable X , falls $\forall \varepsilon > 0$ gilt:

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \rightarrow 0.$$

In diesem Fall schreibt man $X_n \xrightarrow{p} X$.

2. Fast sichere Konvergenz:

Die Folge X_n konvergiert fast sicher gegen X , falls $\mathbb{P}(\lim_{n \rightarrow \infty} X_n = X) = 1$, oder äquivalent dazu, falls $\forall \varepsilon > 0$ gilt

$$\mathbb{P}\left(\bigcup_{k \geq n} \{|X_k - X| \geq \varepsilon\}\right) \rightarrow 0 \text{ für } n \rightarrow \infty.$$

Man schreibt dann $X_n \xrightarrow{\text{a.s.}} X$.

3. Konvergenz im p -ten Mittel:

Gilt für $p > 0$, dass $\mathbb{E}|X_n - X|^p \rightarrow 0$, dann konvergiert X_n im p -ten Mittel gegen X . Dies wird abgekürzt mit $X_n \xrightarrow{L_p} X$.

4. Konvergenz in Verteilung:

Sind $X_n \sim F_n(x)$ und $X \sim F(x)$, dann konvergiert X_n in Verteilung gegen X , falls $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ für jeden Stetigkeitspunkt x gilt. Hier schreibt man $X_n \xrightarrow{d} X$.

Bemerkung. Da aus $X_n \xrightarrow{L_p} X$ folgt, dass $X_n \xrightarrow{p} X$, und ersteres in der Regel praktischer nachzuweisen ist, zeigt man häufig das stärkere Resultat der Konvergenz im p -ten Mittel, um Konvergenz in Wahrscheinlichkeit zu zeigen.

Definition A.2. (Asymptotische Äquivalenz) Seien X_n und Y_n Folgen von Zufallsvariablen. Konvergiert die Differenz der Folgen in Wahrscheinlichkeit gegen 0, also $|X_n - Y_n| \xrightarrow{p} 0$ für $n \rightarrow \infty$, dann heißen X_n and Y_n asymptotisch äquivalent und man schreibt:

$$X_n \doteq Y_n.$$

Definition A.3. (Asymptotische Verteilungsgleichheit) Sei X_n eine Folge von Zufallsvariablen. Sei weiter F eine Verteilungsfunktion und $U \sim F$ eine Zufallsvariable mit dieser Verteilung. Wenn $X_n \doteq U$, dann ist X_n asymptotisch verteilt wie U bzw. folgt asymptotisch der Verteilung F und man schreibt:

$$X_n \dot{\sim} F.$$

A. Anhang

Satz A.4. (c_r -Ungleichung) Für beliebige (abhängige oder unabhängige) Zufallsvariablen X und Y gilt

$$\mathbb{E}|X + Y|^r \leq c_r \cdot [\mathbb{E}(|X|^r) + \mathbb{E}(|Y|^r)],$$

mit

$$c_r = \begin{cases} 1 & , \text{ für } 0 < r \leq 1, \\ 2^{r-1} & , \text{ für } r > 1. \end{cases}$$

Beweis. Siehe **Loève (1977)**, S.157. □

Satz A.5. (Jensensche Ungleichung) Sei X eine Zufallsvariable mit $\mathbb{E}(X) < \infty$ und sei $g(\cdot)$ eine konvexe Funktion. Dann gilt

$$g[\mathbb{E}(X)] \leq \mathbb{E}[g(X)].$$

Falls $X \sim F(x)$ ist, gilt insbesondere für die konvexe Funktion $g(x) = x^2$

$$\left(\int x dF(x) \right)^2 \leq \int x^2 dF(x).$$

Speziell folgt weiter

$$\left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2 \leq \frac{1}{n} \sum_{i=1}^n x_i^2.$$

Beweis. Siehe **Loève (1977)**, S.161. □

Satz A.6. (Lindeberg-Feller) Die Zufallsvariablen $X_i \sim F_i(x)$ seien unabhängig und es sei $\mathbb{E}(X_i) = \mu_i$ und $\text{Var}(X_i) = \sigma_i^2 > 0$, $i = 1, \dots, n$. Es bezeichne $C_n^2 = \sum_{i=1}^n \sigma_i^2$ die Summe der Varianzen. Dann gilt für $n \rightarrow \infty$

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \frac{\sigma_i}{C_n} = 0 \quad \text{und} \quad \frac{1}{C_n} \sum_{i=1}^n (X_i - \mu_i) \xrightarrow{d} U \sim \mathcal{N}(0, 1)$$

genau dann, wenn $\forall \varepsilon > 0$ die Lindeberg-Bedingung

$$\lim_{n \rightarrow \infty} \frac{1}{C_n^2} \sum_{i=1}^n \int_{|x - \mu_i| > \varepsilon C_n} (x - \mu_i) dF_i(x) = 0$$

erfüllt ist. Dabei meint \xrightarrow{d} Konvergenz in Verteilung.

Korollar A.7. (Lindeberg-Bedingung) Die Zufallsvariablen X_i , $i = 1, \dots, n$, seien unabhängig und gleichmäßig beschränkt und es sei $\text{Var}(X_i) = \sigma_i^2 > 0$, $i = 1, \dots, n$. Dann folgt, dass die Lindeberg-Bedingung genau dann erfüllt ist, wenn $\sum_{i=1}^n \sigma_i^2 \rightarrow \infty$ für $n \rightarrow \infty$ gilt.

Beweis. Siehe **Gnedenko (1962)**. □

A. Anhang

Satz A.8. (Cramér-Wold Technik) *Seien $\mathbf{X}_n = (X_{n1}, \dots, X_{nd})'$ sowie $\mathbf{X} = (X_1, \dots, X_d)'$ zwei d -dimensionale Zufallsvektoren. Dann konvergiert \mathbf{X}_n in Verteilung gegen \mathbf{X} genau dann, wenn jede feste Linearkombination von \mathbf{X}_n in Verteilung gegen die entsprechende Linearkombination von \mathbf{X} konvergiert*

$$\mathbf{X}_n \xrightarrow{d} \mathbf{X} \Leftrightarrow \mathbf{t}'\mathbf{X}_n \xrightarrow{d} \mathbf{t}'\mathbf{X} \quad \forall \mathbf{t} \in \mathbb{R}^d.$$

Beweis. Cramér und Wold (1936). □

Satz A.9. (Cramérs multivariater Delta-Satz) *Sei $\phi : \mathbb{D}_\phi \subset \mathbb{R}^k \mapsto \mathbb{R}^m$ eine Abbildung definiert auf einer Teilmenge des \mathbb{R}^k und differenzierbar in $\boldsymbol{\theta}$. Seien \mathbf{T}_n Zufallsvektoren mit Werten im Wertebereich von ϕ . Falls $r_n(\mathbf{T}_n - \boldsymbol{\theta}) \xrightarrow{d} \mathbf{T}$ für die Zahlen $r_n \rightarrow \infty$, dann gilt auch $r_n(\phi(\mathbf{T}_n) - \phi(\boldsymbol{\theta})) \xrightarrow{d} \phi'_\theta(\mathbf{T})$. Zudem konvergiert die Differenz zwischen $r_n(\phi(\mathbf{T}_n) - \phi(\boldsymbol{\theta}))$ und $\phi'_\theta(r_n(\mathbf{T}_n - \boldsymbol{\theta}))$ in Wahrscheinlichkeit gegen Null. Dabei meint*

$$\phi'_\theta = \begin{pmatrix} \frac{\partial \phi_1}{\partial x_1}(\boldsymbol{\theta}) & \dots & \frac{\partial \phi_1}{\partial x_k}(\boldsymbol{\theta}) \\ \vdots & & \vdots \\ \frac{\partial \phi_m}{\partial x_1}(\boldsymbol{\theta}) & \dots & \frac{\partial \phi_m}{\partial x_k}(\boldsymbol{\theta}) \end{pmatrix}$$

die Jakobimatrix von ϕ im Punkt $\boldsymbol{\theta}$.

Beweis. van der Vaart (1998), S.25ff. □

Bemerkung. Häufig tritt die Situation auf, dass $\sqrt{n}(\mathbf{T}_n - \boldsymbol{\theta})$ zu einer multivariaten Normalverteilung $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ konvergiert. Der Satz besagt dann, dass $\sqrt{n}(\phi(\mathbf{T}_n) - \phi(\boldsymbol{\theta}))$ gegen eine $\mathcal{N}(\phi'_\theta \boldsymbol{\mu}, \phi'_\theta \boldsymbol{\Sigma} (\phi'_\theta)^T)$ -Verteilung konvergiert.

Satz A.10. (Kontrolle der FWER) *Sei $(\boldsymbol{\Omega}, \mathbf{T}, \xi)$ eine kohärente simultane Testprozedur. Die Wahrscheinlichkeit, dass $(\boldsymbol{\Omega}, \mathbf{T}, \xi)$ eine wahre Nullhypothese zum Niveau α ablehnt, ist α , sofern die Globalhypothese stimmt. Sie ist höchstens α , unabhängig von der Gültigkeit der Globalhypothese, wenn $(\boldsymbol{\Omega}, \mathbf{T})$ eine geschlossene oder gemeinsame Testfamilie ist.*

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

A.2. Tabellen und Daten

Tabelle 11: Simulierte Fehlerraten des Fehlers erster Art zum nominalen Niveau $\alpha = 0.05$ in 10000 Simulationsläufen bei gleichen Stichprobenumfängen und gleichen Skalierungsfaktoren der Kovarianzmatrizen. Erzeugung der Daten erfolgt mittels **multivariater Lognormalverteilung** und derart, dass jeweils nur die zwei anderen Effekte vorhanden sind, wenn auf den dritten Effekt getestet wird.

| Balanced 2×2 Design | | | | | | | | |
|--|---------------------|--------|--------|--------|------------------|--------|--------|--------|
| Kovarianzstruktur | MCTP | ATS | ANOVA | T^2 | MCTP | ATS | ANOVA | T^2 |
| | $n_1 = n_2 = 10$ | | | | $n_1 = n_2 = 20$ | | | |
| | Gruppeneffekt A | | | | | | | |
| CS | 0.051 | 0.0682 | 0.0533 | 0.0413 | 0.0509 | 0.0603 | 0.0504 | 0.0433 |
| AR(0.6) | 0.0494 | 0.0703 | 0.0466 | 0.0365 | 0.0516 | 0.058 | 0.043 | 0.0381 |
| TPL | 0.0483 | 0.0659 | 0.0378 | 0.0304 | 0.0531 | 0.0577 | 0.0358 | 0.0319 |
| | Zeiteffekt D | | | | | | | |
| CS | 0.0501 | 0.061 | 0.0464 | 0.0371 | 0.0507 | 0.0567 | 0.0462 | 0.0409 |
| AR(0.6) | 0.0513 | 0.0589 | 0.0499 | 0.0401 | 0.0484 | 0.0582 | 0.0458 | 0.0416 |
| TPL | 0.0494 | 0.0612 | 0.038 | 0.0299 | 0.0458 | 0.0548 | 0.035 | 0.0312 |
| | Wechselwirkung AD | | | | | | | |
| CS | 0.0429 | 0.0594 | 0.0463 | 0.0339 | 0.0511 | 0.0585 | 0.0453 | 0.0406 |
| AR(0.6) | 0.0466 | 0.0557 | 0.0454 | 0.0352 | 0.0474 | 0.0594 | 0.0414 | 0.0361 |
| TPL | 0.0477 | 0.0589 | 0.0372 | 0.0296 | 0.0467 | 0.0572 | 0.0347 | 0.0297 |

A. Anhang

Tabelle 12: Original-Daten der Studie bzgl. der PCT-Konzentrations-Verläufe

| PCT-Konzentrations-Verläufe | | | | | | | | | | | | |
|-----------------------------|--------|-----------|-----|------|------|------|--------|-----------|-----|-----|-----|------|
| Pat. | Gruppe | Zeitpunkt | | | | Pat. | Gruppe | Zeitpunkt | | | | |
| | | 1 | 2 | 3 | 4 | | | 1 | 2 | 3 | 4 | |
| 1 | A | 0,2 | 0,2 | 3,6 | 2,6 | 14 | B | 0,3 | 0,2 | 0,5 | 0,3 | |
| 3 | | 0,2 | 0,2 | 0,4 | 0,4 | 28 | | 0,2 | 0,2 | 3,4 | 2,8 | |
| 6 | | 0,2 | 0,2 | 1 | 1 | 29 | | 0,2 | 0,2 | 0,6 | 0,4 | |
| 7 | | 0,2 | 0,2 | 0,7 | 0,6 | 40 | | 0,2 | 0,2 | 1,3 | 0,8 | |
| 8 | | 0,2 | 0,2 | 0,3 | 0,2 | 41 | | 0,2 | 0,2 | 0,5 | 0,7 | |
| 9 | | 0,2 | 0,2 | 1,6 | 1,4 | 61 | | 0,2 | 0,2 | 0,5 | 0,8 | |
| 15 | | 0,2 | 0,2 | 2 | 1,1 | 62 | | 0,2 | 0,2 | 0,7 | 0,5 | |
| 16 | | 0,2 | 0,2 | 28,7 | 21,5 | 63 | | 0,2 | 0,2 | 0,2 | 0,3 | |
| 18 | | 0,2 | 0,2 | 0,7 | 0,4 | 64 | | 0,2 | 0,2 | 1,5 | 1,5 | |
| 23 | | 0,2 | 0,2 | 7,6 | 7,6 | 75 | | 0,2 | 0,2 | 0,5 | 3,9 | |
| 24 | | 0,2 | 0,2 | 18,4 | 17,5 | 79 | 0,2 | 0,8 | 0,8 | 0,6 | | |
| 25 | | 0,2 | 0,3 | 12,2 | 8,7 | 80 | 0,2 | 0,2 | 0,3 | 0,3 | | |
| 30 | | 0,2 | 0,2 | 4,1 | 3,6 | 86 | 0,3 | 0,2 | 0,3 | 0,3 | | |
| 32 | | 0,2 | 0,2 | 1,9 | 1,3 | 88 | 0,2 | 0,2 | 0,2 | 0,4 | | |
| 33 | | 0,2 | 0,2 | 4,1 | 3,4 | 94 | 0,2 | 0,2 | 0,4 | 0,3 | | |
| 36 | | 0,2 | 0,2 | 0,8 | 0,7 | 97 | 0,2 | 0,2 | 0,2 | 0,3 | | |
| 37 | | 0,2 | 0,2 | 1,3 | 0,7 | 103 | 0,4 | 0,3 | 7 | 6,1 | | |
| 38 | | A | 0,2 | 0,2 | 0,7 | 0,7 | 12 | C | 0,2 | 0,2 | 1,2 | 0,9 |
| 42 | | | 0,2 | 0,2 | 9,3 | 9,1 | 17 | | 0,2 | 0,2 | 0,4 | 0,3 |
| 44 | | | 0,2 | 0,2 | 0,4 | 0,3 | 20 | | 0,2 | 0,2 | 0,7 | 0,6 |
| 45 | | | 0,2 | 0,5 | 6,6 | 5,1 | 26 | | 0,2 | 0,2 | 0,7 | 0,5 |
| 46 | | | 0,2 | 0,2 | 5,6 | 4,8 | 27 | | 0,2 | 0,2 | 4,3 | 1,8 |
| 47 | | | 0,2 | 0,2 | 0,4 | 0,3 | 31 | | 0,2 | 0,2 | 0,3 | 0,2 |
| 49 | | | 0,2 | 0,2 | 4 | 3,4 | 35 | | 0,2 | 0,2 | 8,4 | 14,3 |
| 51 | | | 0,2 | 0,2 | 0,9 | 0,8 | 43 | | 0,2 | 0,2 | 0,2 | 0,3 |
| 52 | | | 0,2 | 0,2 | 7,3 | 4,9 | 50 | | 0,2 | 0,2 | 0,5 | 0,3 |
| 53 | | | 0,2 | 0,2 | 1,3 | 0,7 | 72 | | 0,2 | 0,2 | 0,7 | 0,3 |
| 55 | | | 0,2 | 0,2 | 5,8 | 3,9 | 78 | | 0,2 | 0,2 | 0,2 | 0,2 |
| 56 | | | 0,2 | 0,2 | 0,5 | 0,3 | 81 | | 0,2 | 0,2 | 2,5 | 1,7 |
| 57 | | | 0,2 | 0,2 | 0,5 | 0,2 | 93 | | 0,2 | 0,2 | 0,2 | 0,3 |
| 66 | | | 0,2 | 0,2 | 27,9 | 18,4 | 90 | | 0,2 | 0,2 | 0,2 | 0,5 |
| 68 | | | 0,2 | 0,2 | 11,3 | 9 | 92 | | 0,2 | 0,2 | 0,2 | 0,3 |
| 69 | | | 0,2 | 0,3 | 29,9 | 24,4 | 104 | | 0,5 | 0,4 | 0,6 | 0,4 |
| 70 | | | 0,2 | 0,2 | 2,5 | 1,7 | | | | | | |
| 82 | | | 0,2 | 0,2 | 0,7 | 0,8 | | | | | | |
| 83 | | | 0,2 | 0,2 | 2,3 | 1,6 | | | | | | |
| 89 | | | 0,2 | 0,4 | 1,5 | 1 | | | | | | |
| 101 | 0,2 | | 0,5 | 2,6 | 2,1 | | | | | | | |

A. Anhang

Tabelle 13: Original-Daten der Herzklappenstudie

| Herzklappentyp | | | | | |
|----------------|---------------|-----|--------|---------------|------|
| Mosaic | | | Intact | | |
| Klappe | Konservierung | | Klappe | Konservierung | |
| | GA | MMS | | GA | MMS |
| 1 | 3,2 | 0,7 | 11 | 67,7 | 33,1 |
| 2 | 3,5 | 3 | 12 | 63,9 | 37,3 |
| 3 | 5,5 | 0,6 | 13 | 66 | 26,7 |
| 4 | 3,7 | 1,6 | 14 | 63,4 | 32,7 |
| 5 | 3,7 | 1,8 | 15 | 64 | 37 |
| 6 | 3,2 | 2,3 | 16 | 64 | 35,2 |
| 7 | 3,5 | 1,7 | 17 | 69,4 | 33 |
| 8 | 3,6 | 2,4 | 18 | 63,1 | 32,7 |
| 9 | 3,3 | 3,2 | 19 | 69,2 | 37,1 |
| 10 | 3,2 | 3,3 | 20 | 69 | 38,4 |

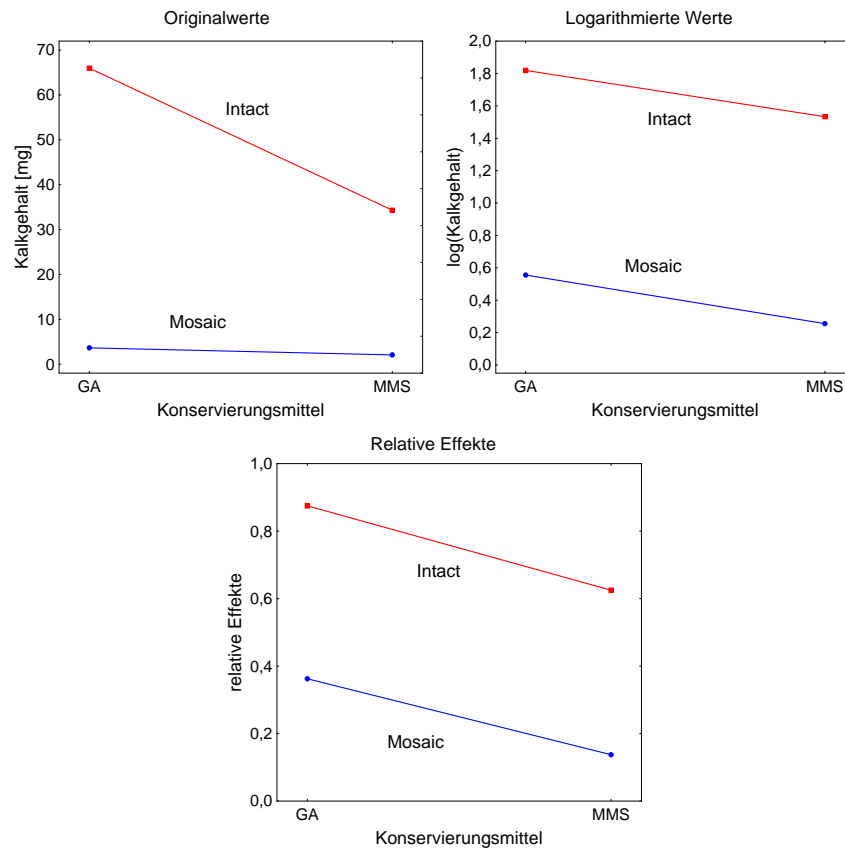


Abbildung 10: Visualisierung der Eckdaten der Herzklappenstudie.

A.3. Auswertungsprogramm

```

#-----#
f1_mctp_f1<-function(formula, data , type = c("Tukey","Dunnett","Sequen","Williams",
      "Changepoint", "AVE", "McDermott", "Marcus", "UmbrellaWilliams",
      "UserDefined"), conf.level=0.95,
      alternative = c("two.sided", "less", "greater"),
      asy.method = c("fisher", "mult.t", "normal","resample"),
      plot.simci=FALSE, control.group=NULL, control.rep=NULL, rounds=3,
      contrast.matrix.group=NULL,contrast.matrix.rep=NULL,correlation=FALSE,
      analyze.group=TRUE,analyze.rep=TRUE,analyze.inter=TRUE){

'%+%'<-function(A,B){
a.1<-nrow(A)
a.2<-ncol(A)
b.1<-nrow(B)
b.2<-ncol(B)
cbind(rbind(A,matrix(0,ncol=a.2,nrow=b.1)),rbind(matrix(0,ncol=b.2,nrow=a.1),B))
}
#-----Necessary R Packages-----#

mvtnorm <- require(mvtnorm, quietly = TRUE)
multcomp <- require(multcomp, quietly = TRUE)
MASS<- require(MASS, quietly = TRUE)

#-----Quality Checks-----#
if (conf.level >= 1 || conf.level <= 0) {
  stop("The confidence level must be between 0 and 1!")
  if (is.null(alternative)) {
    stop("Please declare the alternative! (two.sided, lower, greater)")
  }
}

#-----Containment of given parameters-----#
type <- match.arg(type)
alternative <- match.arg(alternative)
asy.method <- match.arg(asy.method)

#-----Arrange the data-----#
dat <- model.frame(formula, data)
if (ncol(dat) != 3) {
  stop("Specify one response and two class variables in the formula (first:
      grouping variable, second: repeated measures factor)")
}
if (is.numeric(dat[, 1]) == FALSE) {
  stop("Response variable must be numeric")}

response <- dat[, 1]
factora <- as.factor(dat[, 2])
factorb <- as.factor(dat[, 3])
factors<-list(factora,factorb)
samples<-split(response,factors)
fla <- levels(factora)
flb <- levels(factorb)
a <- nlevels(factora)
d <- nlevels(factorb)

```

A. Anhang

```

n<-sapply(samples,length)
N<-sum(n)/d

#-----Compute the point estimators-----#

intRanks<-lapply(samples,rank)
tmp1<-sort(rep(1:(a*d),a*d))
tmp2<-rep(1:(a*d),a*d)
pairRanks<-lapply(1:((a*d)^2),function(arg)
rank(c(samples[[tmp1[arg]]],samples[[tmp2[arg]]])))
tmp1<-sort(rep(1:(a*d),a*d))

w.rs.pq<-function(r,s,p,q,pR,iR,n,a,d){
prank<-Rrs.in.rs.plus.pq(p,q,r,s,pR,n,a,d)
nr<-length(Rrs.in.rs(r,s,iR,a,d))
irank<-Rrs.in.rs(p,q,iR,a,d)
np<-length(irank)
out<-1/nr*((1/np*sum(prank-irank)))
return(out)}

Rrs.in.rs.plus.pq<-function(r,s,p,q,pR,n,a,d){
out<-pR[[(a*(s-1)+(r-1))*a*d+(q-1)*a+p]][1:n[[tmp1[(a*(s-1)+(r-1))*a*d+(q-1)*a+p]]]]
return(out)}

Rrs.in.rs<-function(r,s,iR,a,d){
out<-iR[[(s-1)*a+r]]
return(out)}

Drs.pq<-function(r,s,p,q,pR,iR,n,a,d){
prank<-Rrs.in.rs.plus.pq(r,s,p,q,pR,n,a,d)
np<-length(Rrs.in.rs(p,q,iR,a,d))
irank<-Rrs.in.rs(r,s,iR,a,d)
nr<-length(irank)
out<-1/np*((prank-irank)-(1/nr*sum(prank-irank)))
return(out)}

tau.hat<-function(r,s,l,p,q,t,u,pR,iR,n,a,d){
nr<-length(Rrs.in.rs(r,s,iR,a,d))
out<-1/nr/(nr-1)*sum(Drs.pq(r,s,p,q,pR,iR,n,a,d)*Drs.pq(r,l,t,u,pR,iR,n,a,d))
return(out)}

index<-c()
w<-c()
help<-expand.grid(1:d,1:a,1:d,1:a)
for (j in 1:(a*d)^2){
r<-help[j,4]
s<-help[j,3]
p<-help[j,2]
q<-help[j,1]
ind<-(r-1)*d*a*d+(s-1)*a*d+(p-1)*d+q
w[ind]<-w.rs.pq(r,s,p,q,pairRanks,intRanks,n,a,d)
}
E<-1/(a*d)*matrix(rep(diag(a*d),a*d),nrow=a*d)

```

A. Anhang

```

p.hat<-E%*%w

S<-matrix(rep(0,(a*d)^4),ncol=(a*d)^2,nrow=(a*d)^2)
H<-matrix(rep(0,(a*d)^4),ncol=(a*d)^2,nrow=(a*d)^2)
#-----#

cov.pqrs.pqil<-function(p,q,r,s,i,l,pR,iR,n,a,d){
out<-0
if (p!=r && p!=i && r!=i){
out<-tau.hat(p,q,q,r,s,i,l,pR,iR,n,a,d)}

if (p!=r && p!=i && r==i){
out<-tau.hat(p,q,q,r,s,r,l,pR,iR,n,a,d)+tau.hat(r,s,l,p,q,p,q,pR,iR,n,a,d)}

if (p==r && r!=i && s!=q){
out<-tau.hat(p,q,q,p,s,i,l,pR,iR,n,a,d)-tau.hat(p,s,q,p,q,i,l,pR,iR,n,a,d)}

if (p==i && r!=i && l!=q){
out<-tau.hat(p,q,q,r,s,p,l,pR,iR,n,a,d)-tau.hat(p,q,l,r,s,p,q,pR,iR,n,a,d)}

if (p==r && r==i && p==i && s!=q && l!=q){
out<-tau.hat(p,q,q,r,s,p,l,pR,iR,n,a,d)-tau.hat(p,q,l,p,s,p,q,pR,iR,n,a,d)
-tau.hat(p,s,q,p,q,p,l,pR,iR,n,a,d)+tau.hat(p,s,l,p,q,p,q,pR,iR,n,a,d)}
return(out)
}

pR<-pairRanks
iR<-intRanks
#-----#

for (p in 1:a){
  for (q in 1:d){
    for (r in 1:a){
      for (s in 1:d){
        qh<-q
        rh<-r
        sh<-s
        for (t in p:a){
          for (u in qh:d){
            for (i in rh:a){
              for (l in sh:d){
ind1<-(p-1)*d*a*d+(q-1)*a*d+(r-1)*d+s
ind2<-(t-1)*d*a*d+(u-1)*a*d+(i-1)*d+l
indrspq<-(r-1)*d*a*d+(s-1)*a*d+(p-1)*d+q
indiltu<-(i-1)*d*a*d+(l-1)*a*d+(t-1)*d+u

while(H[ind1,ind2]==0){

if(H[indrspq,ind2]!=0 && H[ind1,ind2]==0){
S[ind1,ind2]<- -S[indrspq,ind2]
H[ind1,ind2]<-H[ind1,ind2]+1}

if(H[ind1,indiltu]!=0&& H[ind1,ind2]==0){
S[ind1,ind2]<- -S[ind1,indiltu]

```

A. Anhang

```

H[ind1,ind2]<-H[ind1,ind2]+1}

if(H[indrspq,indiltu]!=0 && H[ind1,ind2]==0){
S[ind1,ind2]<-S[indrspq,indiltu]
H[ind1,ind2]<-H[ind1,ind2]+1}

if (p==t && q==u && H[ind1,ind2]==0){
S[ind1,ind2]<-cov.pqrs.pqil(p,q,r,s,i,l,pR,iR,n,a,d)
H[ind1,ind2]<-H[ind1,ind2]+1
}

else{
if (p==i && q==l && H[ind1,ind2]==0){
S[ind1,ind2]<- -cov.pqrs.pqil(p,q,r,s,t,u,pR,iR,n,a,d)
H[ind1,ind2]<-H[ind1,ind2]+1
}

else{
if (r==t && s==u && H[ind1,ind2]==0){
S[ind1,ind2]<- -cov.pqrs.pqil(r,s,p,q,i,l,pR,iR,n,a,d)
H[ind1,ind2]<-H[ind1,ind2]+1
}

else{
if (r==i && s==l && H[ind1,ind2]==0){
S[ind1,ind2]<- cov.pqrs.pqil(r,s,p,q,t,u,pR,iR,n,a,d)
H[ind1,ind2]<-H[ind1,ind2]+1
}

else{
if(r==i && p!=i && p!=t && r!=t && H[ind1,ind2]==0){
S[ind1,ind2]<-tau.hat(r,s,l,p,q,t,u,pR,iR,n,a,d)
H[ind1,ind2]<-H[ind1,ind2]+1}

if(r==t && p!=i && p!=t && r!=i && H[ind1,ind2]==0){
S[ind1,ind2]<- -tau.hat(r,s,u,p,q,i,l,pR,iR,n,a,d)
H[ind1,ind2]<-H[ind1,ind2]+1}

if(p==i && r!=i && r!=t && p!=t && H[ind1,ind2]==0){
S[ind1,ind2]<- -tau.hat(p,q,l,r,s,t,u,pR,iR,n,a,d)
H[ind1,ind2]<-H[ind1,ind2]+1}

if(p==t && r!=i && r!=t && p!=i && H[ind1,ind2]==0){
S[ind1,ind2]<-tau.hat(p,q,u,r,s,i,l,pR,iR,n,a,d)
H[ind1,ind2]<-H[ind1,ind2]+1}

if(r==i && i==t && r==t && p!=i && p!=t && u!=1 && H[ind1,ind2]==0){
S[ind1,ind2]<-tau.hat(r,s,l,p,q,r,u,pR,iR,n,a,d)-tau.hat(r,s,u,p,q,r,l,pR,iR,n,a,d)
H[ind1,ind2]<-H[ind1,ind2]+1}

if(p==i && i==t && p==t && r!=i && r!=t && u!=1 && H[ind1,ind2]==0){
S[ind1,ind2]<- -tau.hat(p,q,l,r,s,p,u,pR,iR,n,a,d)+tau.hat(p,q,u,r,s,i,l,pR,iR,n,a,d)
H[ind1,ind2]<-H[ind1,ind2]+1}

```

A. Anhang

```

if(r==i && i==p && r==p && t!=r && t!=p && q!=s && H[ind1,ind2]==0){
S[ind1,ind2]<-tau.hat(r,s,l,r,q,t,u,pR,iR,n,a,d)-tau.hat(r,q,l,r,s,t,u,pR,iR,n,a,d)
H[ind1,ind2]<-H[ind1,ind2]+1}

if(t==r && t==p && r==p && i!=r && i!=p && q!=s && H[ind1,ind2]==0){
S[ind1,ind2]<-tau.hat(r,s,u,r,q,i,l,pR,iR,n,a,d)+tau.hat(r,q,u,r,s,i,l,pR,iR,n,a,d)
H[ind1,ind2]<-H[ind1,ind2]+1}

if(r==i && p==t && r!=t && p!=i && H[ind1,ind2]==0){
S[ind1,ind2]<-tau.hat(r,s,l,p,q,p,u,pR,iR,n,a,d)+tau.hat(p,q,u,r,s,r,l,pR,iR,n,a,d)
H[ind1,ind2]<-H[ind1,ind2]+1}

if(r==t && p==i && r!=i && p!=t && H[ind1,ind2]==0){
S[ind1,ind2]<-tau.hat(r,s,u,p,q,p,l,pR,iR,n,a,d)-tau.hat(p,q,l,r,s,r,u,pR,iR,n,a,d)
H[ind1,ind2]<-H[ind1,ind2]+1}

if(r==p && p==i && i==t && q!=s && u!=l && q!=l && s!=u && q!=u && H[ind1,ind2]==0){
S[ind1,ind2]<-tau.hat(r,s,l,r,q,r,u,pR,iR,n,a,d)-tau.hat(r,s,u,r,q,r,l,pR,iR,n,a,d)
-tau.hat(r,q,l,r,s,r,u,pR,iR,n,a,d)+tau.hat(r,q,u,r,s,r,l,pR,iR,n,a,d)
H[ind1,ind2]<-H[ind1,ind2]+1}
}
}
}
}
H[ind1,ind2]<-H[ind1,ind2]+(H[ind1,ind2]==0)*1
}

        sh<-1}
        rh<-1}
        qh<-1}
    }
}
}
}
S<-N*(S+t(S)-diag(diag(S)))

V<-E*%S*%t(E)

#-----Compute the MCTP and SCI-----#

#####dfT.mult.T#####
#-----#

Alijk<-function(Con,l,i,j,pR,iR,n,a,d){
res<-rep(0,n[[i]])
for (r in 1:a){
for (s in 1:d){
res<-res+((Con[l,(i-1)*d+j]-Con[l,(r-1)*d+s])/n[[r]]*(Rrs.in.rs.plus.pq(i,j,r,s,pR,n,a,d)
-Rrs.in.rs(i,j,iR,a,d) )/a/d
}
}
return(res)
}

```

A. Anhang

```

wlis<-function(Con,l,pR,iR,n,a,d){
#N<-sum(n)/d
out<-c()
for(i in 1:a){
res<-0
for(j in 1:d){
for(k in 1:d){
res<-res+( 1/(n[[i]]-1)* sum((Alijk(Con,l,i,j,pR,iR,n,a,d)-
mean(Alijk(Con,l,i,j,pR,iR,n,a,d)))*(Alijk(Con,l,i,k,pR,iR,n,a,d)
-mean(Alijk(Con,l,i,k,pR,iR,n,a,d)))) )
}}
out[i]<-res
}
out<-out+(out==0)*0.001
return(out)
}

dfTmultT<-function(Con,pR,iR,n,a,d){
sss<-c()
for(j in 1:a){sss[j]<-n[[j]]}
res<-c()
for(l in 1:nrow(Con)){
res[l]<-sum(wlis(Con,l,pR,iR,n,a,d)/sss)^2/sum(wlis(Con,l,pR,iR,n,a,d)^2/sss^2/(sss-1))
}
out<-max(1,min(res))
return(out)
}

#----parametric ANOVA-----#
X<- rep(1,n[[1]])%x%diag(d)
for (tt in 2:a){
X<-X%+(rep(1,n[[tt]])%x%diag(d))
}
Y.q<-solve(t(X)%*%X)%*%t(X)%*%response
YMats<-list()
for(kk in 1:a){
YMats[[kk]]<-matrix(response[(factora==factora[kk])],ncol=d,byrow=TRUE)
}

Vdach<-matrix(rep(0,d^2),nrow=d,ncol=d)
ColMeans1<- colMeans(YMats[[1]])
for(i in 1:n[[1]]){
Vdach<-Vdach+(YMats[[1]][i,]-ColMeans1)%*%t(YMats[[1]][i,]-ColMeans1)
}
Vdach<-N/n[[1]]/(n[[1]]-1)*Vdach

for(kk in 2:a){
Vdachkk<-matrix(rep(0,d^2),nrow=d,ncol=d)
ColMeanskk<- colMeans(YMats[[kk]])
for(i in 1:n[[kk]]){
Vdachkk<-Vdachkk+(YMats[[kk]][i,]-ColMeanskk)%*%t(YMats[[kk]][i,]-ColMeanskk)
}
Vdachkk<-N/n[[kk]]/(n[[kk]]-1)*Vdachkk
Vdach<-Vdach%+%Vdachkk

```

A. Anhang

```

}
#-----#
#####

if(a>1){
#-----Factor A - Group-----#
if (type=="UserDefined"){
if(is.null(contrast.matrix.group)){stop("Please enter a contrast matrix!")}
CA<-contrast.matrix.group
rownames(CA)<-paste("C",1:nrow(CA))
colnames(CA)<-fla}

if (type !="UserDefined"){
if (is.null(control.group)){iconA<-1}
if (!is.null(control.group)){iconA<-which(fla==control.group)}
CA<-contrMat(rep(1,a),type,base=iconA) }

nca<-nrow(CA)
connames.A<-rownames(CA)
ConA<-matrix(CA, ncol=a)
rownames(ConA)<-connames.A
colnames(ConA)<-colnames(CA)
M<-(diag(a)-1/a*matrix(rep(1,a^2),nrow=a,ncol=a))%x%t(rep(1/d,d))
ConA<-ConA%*%M
ConAp<-ConA%*%p.hat
CVA<-ConA%*%V%*%t(ConA)
if(nca==1){if(CVA==0){CVA[[1]]<-0.001}}
if(nca>1){CVA<-CVA+diag((diag(CVA)==0)*0.001)}
dCVA<-sqrt(diag(CVA))
rhobfA<-cov2cor(CVA)

#-----WTS-----
Q<-N*t(ConAp)%*%ginv(CVA)%*%ConAp
f<-qr(ConA)$rank

#-----ATS
M<-t(ConA)%*%ginv(ConA)%*%t(ConA)%*%ConA
An<-N*sum(diag(M%*%V))/sum(diag(M%*%V%*%M%*%V))*t(p.hat)%*%M%*%p.hat
fn<-sum(diag(M%*%V))^2/sum(diag(M%*%V%*%M%*%V))

#-----parametric ANOVA
TAM<-(diag(a)-1/a*matrix(rep(1,a^2),nrow=a,ncol=a))%x%(1/d*matrix(rep(1,d^2),nrow=d,ncol=d))
StatA<-N/sum(diag(TAM%*%Vdach))*t(Y.q)%*%TAM%*%Y.q
fA<-sum(diag(TAM%*%Vdach))^2/sum(diag(TAM%*%Vdach%*%TAM%*%Vdach))

switch(alternative,
#-----Two-sided alternative-----#
two.sided={
WTS.A<-(2*min(pchisq(Q,f),1-pchisq(Q,f)))
ATS.A<-(2*min(pchisq(An,fn),1-pchisq(An,fn)))
para.pA<-((2*min(pf(StatA,fA,Inf),1-pf(StatA,fA,Inf))))
},
#-----Alternative= LOWER-----#
less={

```

A. Anhang

```

WTS.A<-(pchisq(Q,f))
ATS.A<-(pchisq(An,fn))
para.pA<-(pf(StatA,fA,Inf))
},
#-----Alternative= GREATER-----#
greater={
WTS.A<-(1-pchisq(Q,f))
ATS.A<-(1-pchisq(An,fn))
para.pA<-(1-pf(StatA,fA,Inf))
}
)
#-----#
}

if(d>1){
#-----Factor D - RepMeasure-----#
if (type=="UserDefined"){
if(is.null(contrast.matrix.rep)){stop("Please enter a contrast matrix!")}
CD<-contrast.matrix.rep
rownames(CD)<-paste("C",1:nrow(CD))
colnames(CD)<-flb}

if (type!="UserDefined"){
if (is.null(control.rep)){iconD<-1}
if (!is.null(control.rep)){iconD<-which(flb==control.rep)}
CD<-contrMat(rep(1,d),type,base=iconD) }

ncd<-nrow(CD)
connames.D<-rownames(CD)
ConD<-matrix(CD, ncol=d)
rownames(ConD)<-connames.D
colnames(ConD)<-colnames(CD)

M<-t(rep(1/a,a))%x%(diag(d)-1/d*matrix(rep(1,d^2),nrow=d,ncol=d))
ConD<-ConD%*%M
ConDp<-ConD%*%p.hat
CVD<-ConD%*%V%*%t(ConD)
if(ncd==1){if(CVD==0){CVD[[1]]<-0.001}}
if(ncd>1){CVD<-CVD+diag((diag(CVD)==0)*0.001)}
dCVD<-sqrt(diag(CVD))
rhobfD<-cov2cor(CVD)

#-----WTS----
Q<-N*t(ConDp)%*%ginv(CVD)%*%ConDp
f<-qr(ConD)$rank

#-----ATS
M<-t(ConD)%*%ginv(ConD%*%t(ConD))%*%ConD
An<-N*sum(diag(M%*%V))/sum(diag(M%*%V%*%M%*%V))*t(p.hat)%*%M%*%p.hat
fn<-sum(diag(M%*%V))^2/sum(diag(M%*%V%*%M%*%V))

#-----parametric ANOVA
TDM<- 1/a*matrix(rep(1,a^2),nrow=a,ncol=a)%x%(diag(d)-1/d*matrix(rep(1,d^2),nrow=d,ncol=d))
StatD<-N/sum(diag(TDM%*%Vdach))*t(Y.q)%*%TDM%*%Y.q

```


A. Anhang

```

fD<-sum(diag(TDM%*%Vdach))^2/sum(diag(TDM%*%Vdach%*%TDM%*%Vdach))

switch(alternative,
#-----Two-sided alternative-----#
two.sided={
WTS.D<-(2*min(pchisq(Q,f),1-pchisq(Q,f)))
ATS.D<-(2*min(pchisq(An,fn),1-pchisq(An,fn)))
para.pD<-((2*min(pf(StatD,fD,Inf),1-pf(StatD,fD,Inf))))
},
#-----Alternative= LOWER-----#
less={
WTS.D<-(pchisq(Q,f))
ATS.D<-(pchisq(An,fn))
para.pD<-(pf(StatD,fD,Inf))
},
#-----Alternative= GREATER-----#
greater={
WTS.D<-(1-pchisq(Q,f))
ATS.D<-(1-pchisq(An,fn))
para.pD<-(1-pf(StatD,fD,Inf))
}
)
#-----#
}

if(a>1 && d>1){
#-----Interaction-----#

CI<-contrMat(rep(1,a*d),type)
nci<-nrow(CI)
connames.I<-rownames(CI)
ConI<-matrix(CI, ncol=a*d)
cnames<-c()
for(i in 1:a){
for(j in 1:d){
cnames<-c(cnames,paste(" ",i,j," "))
}}

rownames(ConI)<-connames.I
colnames(ConI)<-cnames
M<- (diag(a)-1/a*matrix(rep(1,a^2),nrow=a,ncol=a))%x%(diag(d)
-1/d*matrix(rep(1,d^2),nrow=d,ncol=d))
ConI<-M#[((rowSums(CI%*%M==0)==a*d)==FALSE),]
nci<-nrow(ConI)
ConIp<-ConI%*%p.hat
CVI<-ConI%*%V%*%t(ConI)
dCVI<-sqrt(diag(CVI))
rhobfI<-cov2cor(CVI)

#-----WTS-----
Q<-N*t(ConIp)%*%ginv(CVI)%*%ConIp
f<-qr(ConI)$rank

```

A. Anhang

```

#----ATS
M<-t(ConI)%*%ginv(ConI)%*%t(ConI)%*%ConI
An<-N*sum(diag(M%*%V))/sum(diag(M%*%V%*%M%*%V))*t(p.hat)%*%M%*%p.hat
fn<-sum(diag(M%*%V))^2/sum(diag(M%*%V%*%M%*%V))
#----parametric ANOVA
TIM<- (diag(a)-1/a*matrix(rep(1,a^2),nrow=a,ncol=a))%x%(diag(d)
-1/d*matrix(rep(1,d^2),nrow=d,ncol=d))
StatI<-N/sum(diag(TIM%*%Vdach))*t(Y.q)%*%TIM%*%Y.q
fI<-sum(diag(TIM%*%Vdach))^2/sum(diag(TIM%*%Vdach%*%TIM%*%Vdach))

switch(alternative,
#-----Two-sided alternative-----#
two.sided={
WTS.I<-(2*min(pchisq(Q,f),1-pchisq(Q,f)))
ATS.I<-(2*min(pchisq(An,fn),1-pchisq(An,fn)))
para.pI<-((2*min(pf(StatI,fI,Inf),1-pf(StatI,fI,Inf))))
},
#-----Alternative= LOWER-----#
less={
WTS.I<-(pchisq(Q,f))
ATS.I<-(pchisq(An,fn))
para.pI<-(pf(StatI,fI,Inf))
},
#-----Alternative= GREATER-----#
greater={
WTS.I<-(1-pchisq(Q,f))
ATS.I<-(1-pchisq(An,fn))
para.pI<-(1-pf(StatI,fI,Inf))
}
)
#-----#
}
#-----Compute adjusted p-Values and SCI-----#
result.A<-paste("a<2 || analyze.group==FALSE")
result.D<-paste("d<2 || analyze.rep==FALSE")
result.I<-paste("a<2 || d<2 || analyse.inter==FALSE")

if(asy.method=="resample"){
pboots<-t(rmvnorm(nboot,sigma=V,method="chol"))/sqrt(N)
}

if (a>1 && analyze.group==TRUE){
#-----Grouping Factor A-----#
switch(
asy.method,
#-----Multi NORMAL-----#
normal={
p.adj.A<-c()
AsyMethod <- "Normal - Approximation"
TA<-sqrt(N)*(ConAp)/sqrt(c(diag(CVA)))
switch(alternative,
#-----Two-sided alternative-----#
two.sided={
for (pp in 1:nca) {

```

A. Anhang

```

p.adj.A[pp]<-1-pmvnorm(lower=-abs(TA[pp]*dCVA), upper= abs(TA[pp]*dCVA),
                      mean=rep(0,nca),sigma=CVA)[1]}
crit.A<- qmvnorm(conf.level, sigma=rhobfA, tail = "both")$quantile
Lower.A <- ConAp - crit.A/sqrt(N)*sqrt(c(diag(CVA)))
Upper.A <- ConAp + crit.A/sqrt(N)*sqrt(c(diag(CVA)))
},
#-----Alternative= LOWER-----#
less={
for (pp in 1:nca) {
p.adj.A[pp]<-pmvnorm(lower = -Inf*dCVA , upper = TA[pp]*dCVA,
                    mean = rep(0, nca), sigma = CVA)}
crit.A<- qmvnorm(conf.level, sigma = rhobfA, tail = "lower")$quantile
Lower.A <- rep(-1,nca)
Upper.A <- ConAp + crit.A/sqrt(N)*sqrt(c(diag(CVA)))
},
#-----Alternative= GREATER-----#
greater={
for (pp in 1:nca) {
p.adj.A[pp]<-1-pmvnorm(lower =-Inf*dCVA , upper =TA[pp]*dCVA,
                      mean = rep(0, nca), sigma = CVA)}
crit.A<- qmvnorm(conf.level, sigma = rhobfA, tail = "lower")$quantile
Lower.A <-ConAp - crit.A/sqrt(N)*sqrt(c(diag(CVA)))
Upper.A <- rep(1,nca)}
)
},
#-----Multi T-----#
mult.t={
dfT<-round(dfTmultT(ConA,pR,iR,n,a,d))
p.adj.A<-c()
  TA<-sqrt(N)*(ConAp)/sqrt(c(diag(CVA)))
AsyMethod <- paste("Multi - T with", round(dfT,rounds), "DF")
switch(alternative,
#-----Two-sided alternative-----#
two.sided={
for (pp in 1:nca) {
p.adj.A[pp]<-1-pmvt(lower=-abs(TA[pp]), upper= abs(TA[pp]), delta=rep(0,nca),
                  df=dfT,corr=rhobfA)[1]}

crit.A<- qmvt(conf.level, df=dfT, corr=rhobfA, tail = "both")$quantile
Lower.A <- ConAp - crit.A/sqrt(N)*sqrt(c(diag(CVA)))
Upper.A <- ConAp + crit.A/sqrt(N)*sqrt(c(diag(CVA)))
},
#-----Alternative= LOWER-----#
less={
for (pp in 1:nca) {
p.adj.A[pp]<-pmvt(lower = -Inf , upper = TA[pp],
                 delta = rep(0, nca), df=dfT, corr = rhobfA)}
crit.A<- qmvt(conf.level, df=dfT, corr = rhobfA, tail = "lower")$quantile
Lower.A <- rep(-1,nca)
Upper.A <- ConAp + crit.A/sqrt(N)*sqrt(c(diag(CVA)))
},
#-----Alternative= GREATER-----#
greater={
for (pp in 1:nca) {
p.adj.A[pp]<-1-pmvt(lower =-Inf, upper =TA[pp],

```

A. Anhang

```

        delta = rep(0, nca), df=dfT, corr = rhobfA)}
crit.A<- qmvt(conf.level, df=dfT, corr = rhobfA, tail = "lower")$quantile
Lower.A <-ConAp - crit.A/sqrt(N)*sqrt(c(diag(CVA)))
Upper.A <- rep(1,nca)}
)
},
#-----FISHER-TRANS-----#
fisher={
dfT<-round(dfTmultT(ConA,pR,iR,n,a,d))
p.adj.A<-c()
AsyMethod <- paste("Fisher with", round(dfT,rounds), "DF")
  Cfisher<-1/2*log((1+ConAp)/(1-ConAp))
Vfisherdev<-diag(c(1/(1-ConAp^2)))
Vfisher<-Vfisherdev%*%CVA%*%t(Vfisherdev)
TA<-sqrt(N)*Cfisher/sqrt(c(diag(Vfisher)))
switch( alternative,
#-----Two-sided alternative-----#
two.sided={
for (pp in 1:nca) {
p.adj.A[pp]<-1-pmvt(lower=-abs(TA[pp]), upper= abs(TA[pp]), delta=rep(0,nca),
                    df=dfT,corr=rhobfA)[1])
crit.A<- qmvt(conf.level, df=dfT, corr=rhobfA, tail = "both")$quantile
Lower1 <- ConAp - crit.A/sqrt(N)*sqrt(c(diag(CVA)))
Upper1 <- ConAp + crit.A/sqrt(N)*sqrt(c(diag(CVA)))
Lower.A <- (exp(2*Lower1)-1)/(exp(2*Lower1)+1)
Upper.A <- (exp(2*Upper1)-1)/(exp(2*Upper1)+1)
},
#-----Alternative= LOWER-----#
less={
for (pp in 1:nca) {
p.adj.A[pp]<-pmvt(lower = -Inf , upper = TA[pp],
                  delta = rep(0, nca), df=dfT, corr = rhobfA)}
crit.A<- qmvt(conf.level, df=dfT, corr = rhobfA, tail = "lower")$quantile
Lower.A <- rep(-1,nca)
Upper1 <- ConAp + crit.A/sqrt(N)*sqrt(c(diag(CVA)))
Upper.A <- (exp(2*Upper1)-1)/(exp(2*Upper1)+1)
},
#-----Alternative= GREATER-----#
greater={
for (pp in 1:nca) {
p.adj.A[pp]<-1-pmvt(lower =-Inf , upper =TA[pp],
                  delta = rep(0, nca), df=dfT, corr = rhobfA)}
crit.A<- qmvt(conf.level, df=dfT, corr = rhobfA, tail = "lower")$quantile
Lower1 <-ConAp - crit.A/sqrt(N)*sqrt(c(diag(CVA)))
Lower.A <- (exp(2*Lower1)-1)/(exp(2*Lower1)+1)
Upper.A <- rep(1,nca)}
)
},
resample={
p.adj.A<-c()
AsyMethod <- "Resampling"
TA<-sqrt(N)*(ConAp)/sqrt(c(diag(CVA)))
TAs<-matrix(ncol=nboot,nrow=nca)
for(i in 1:nboot){TAs[,i]<-sqrt(N)*(ConA%*%pboots[,i])/sqrt(c(diag(CVA)))}
}
}

```

A. Anhang

```

for(pp in 1:nca){
  p.adj.A[pp]<-2*min((mean(TA[pp]<=TAs[pp,])),1-(mean(TA[pp]<=TAs[pp,])))
}
crit.A<- sort(apply(TAs,2,max))[nboot*conf.level+1]
Lower.A <- ConAp - crit.A/sqrt(N)*sqrt(c(diag(CVA)))
Upper.A <- ConAp + crit.A/sqrt(N)*sqrt(c(diag(CVA)))
}
)
#-----#

Analysis.of.Relative.Effects.A <- data.frame(Estimator=round(ConAp,rounds),
      Lower=round(Lower.A,rounds), Upper=round(Upper.A,rounds),
      Statistic = round(TA,rounds), p.Value=p.adj.A)
Overall.A<-data.frame(Quantile=crit.A, p.Value=min(p.adj.A), p.Value.WTS=WTS.A,
      p.Value.ATS=ATS.A,p.parANOVA=para.pA)
result.A<-list(Contrast=ConA, Multiple.Comparisons=Analysis.of.Relative.Effects.A,
      Overall=Overall.A)
}
#-----#
#####
#-----#
if(d>1 && analyze.rep==TRUE){
#-----Repeated Measures Factor D-----#
switch(
asy.method,
#-----Multi NORMAL-----#
normal={
p.adj.D<-c()
AsyMethod <- "Normal - Approximation"
TD<-sqrt(N)*(ConDp)/sqrt(c(diag(CVD)))
switch(alternative,
#-----Two-sided alternative-----#
two.sided={
for (pp in 1:ncd) {
p.adj.D[pp]<-1-pmvnorm(lower=-abs(TD[pp]*dCVD), upper= abs(TD[pp]*dCVD),
      mean=rep(0,ncd),sigma=CVD)[1]}
crit.D<- qmvnorm(conf.level, sigma=rhobfD, tail = "both")$quantile
Lower.D <- ConDp - crit.D/sqrt(N)*sqrt(c(diag(CVD)))
Upper.D <- ConDp + crit.D/sqrt(N)*sqrt(c(diag(CVD)))
},
#-----Alternative= LOWER-----#
less={
for (pp in 1:ncd) {
p.adj.D[pp]<-pmvnorm(lower = -Inf*dCVD , upper = TD[pp]*dCVD,
      mean = rep(0, ncd), sigma = CVD)}
crit.D<- qmvnorm(conf.level, sigma = rhobfD, tail = "lower")$quantile
Lower.D <- rep(-1,ncd)
Upper.D <- ConDp + crit.D/sqrt(N)*sqrt(c(diag(CVD)))
},
#-----Alternative= GREATER-----#
greater={
for (pp in 1:ncd) {
p.adj.D[pp]<-1-pmvnorm(lower ==-Inf*dCVD , upper =TD[pp]*dCVD,
      mean = rep(0, ncd), sigma = CVD)}
}
}
}
}

```

A. Anhang

```

crit.D<- qmvnorm(conf.level, sigma = rhobfD, tail = "lower")$quantile
Lower.D <-ConDp - crit.D/sqrt(N)*sqrt(c(diag(CVD)))
Upper.D <- rep(1,ncd)}
)
},
#-----Multi T-----#
mult.t={
dfT<-round(dfTmultT(ConD,pR,iR,n,a,d))
p.adj.D<-c()
  TD<-sqrt(N)*(ConDp)/sqrt(c(diag(CVD)))
AsyMethod <- paste("Multi - T with", round(dfT,rounds), "DF")
switch(alternative,
#-----Two-sided alternative-----#
two.sided={
for (pp in 1:ncd) {
p.adj.D[pp]<-1-pmvt(lower=-abs(TD[pp]), upper= abs(TD[pp]), delta=rep(0,ncd),
                    df=dfT, corr=rhobfD)[1]}

crit.D<- qmvt(conf.level, df=dfT, corr=rhobfD, tail = "both")$quantile
Lower.D <- ConDp - crit.D/sqrt(N)*sqrt(c(diag(CVD)))
Upper.D <- ConDp + crit.D/sqrt(N)*sqrt(c(diag(CVD)))
},
#-----Alternative= LOWER-----#
less={
for (pp in 1:ncd) {
p.adj.D[pp]<-pmvt(lower = -Inf , upper = TD[pp],
                 delta = rep(0, ncd), df=dfT, corr = rhobfD)}
crit.D<- qmvt(conf.level, df=dfT, corr = rhobfD, tail = "lower")$quantile
Lower.D <- rep(-1,ncd)
Upper.D <- ConDp + crit.D/sqrt(N)*sqrt(c(diag(CVD)))
},
#-----Alternative= GREATER-----#
greater={
for (pp in 1:ncd) {
p.adj.D[pp]<-1-pmvt(lower =-Inf , upper =TD[pp],
                  delta = rep(0, ncd), df=dfT, corr = rhobfD)}
crit.D<- qmvt(conf.level, df=dfT, corr = rhobfD, tail = "lower")$quantile
Lower.D <-ConDp - crit.D/sqrt(N)*sqrt(c(diag(CVD)))
Upper.D <- rep(1,ncd)}
)
},
#-----FISHER-TRANS-----#
fisher={
dfT<-round(dfTmultT(ConD,pR,iR,n,a,d))
p.adj.D<-c()
AsyMethod <- paste("Fisher with", round(dfT,rounds), "DF")
  Cfisher<-1/2*log((1+ConDp)/(1-ConDp))
Vfisherdev<-diag(c(1/(1-ConDp^2)))
Vfisher<-Vfisherdev%*%CVD%*%t(Vfisherdev)
TD<-sqrt(N)*Cfisher/sqrt(c(diag(Vfisher)))
switch( alternative,
#-----Two-sided alternative-----#
two.sided={
for (pp in 1:ncd) {
p.adj.D[pp]<-1-pmvt(lower=-abs(TD[pp]), upper= abs(TD[pp]), delta=rep(0,ncd),

```

A. Anhang

```

df=dfT,corr=rhobfD)[1]}
crit.D<- qmvt(conf.level, df=dfT, corr=rhobfD, tail = "both")$quantile
Lower1 <- ConDp - crit.D/sqrt(N)*sqrt(c(diag(CVD)))
Upper1 <- ConDp + crit.D/sqrt(N)*sqrt(c(diag(CVD)))
Lower.D <- (exp(2*Lower1)-1)/(exp(2*Lower1)+1)
Upper.D <- (exp(2*Upper1)-1)/(exp(2*Upper1)+1)
},
#-----Alternative= LOWER-----#
less={
for (pp in 1:ncd) {
p.adj.D[pp]<-pmvt(lower = -Inf , upper = TD[pp],
delta = rep(0, ncd), df=dfT, corr = rhobfD)}
crit.D<- qmvt(conf.level, df=dfT, corr = rhobfD, tail = "lower")$quantile
Lower.D <- rep(-1,ncd)
Upper1 <- ConDp + crit.D/sqrt(N)*sqrt(c(diag(CVD)))
Upper.D <- (exp(2*Upper1)-1)/(exp(2*Upper1)+1)
},
#-----Alternative= GREATER-----#
greater={
for (pp in 1:ncd) {
p.adj.D[pp]<-1-pmvt(lower ==-Inf , upper =TD[pp],
delta = rep(0, ncd), df=dfT, corr = rhobfD)}
crit.D<- qmvt(conf.level, df=dfT, corr = rhobfD, tail = "lower")$quantile
Lower1 <-ConAp - crit.D/sqrt(N)*sqrt(c(diag(CVD)))
Lower.D <- (exp(2*Lower1)-1)/(exp(2*Lower1)+1)
Upper.D <- rep(1,ncd)}
)
}
)
#-----#

Analysis.of.Relative.Effects.D <- data.frame(Estimator=round(ConDp,rounds),
Lower=round(Lower.D,rounds), Upper=round(Upper.D,rounds),
Statistic = round(TD,rounds), p.Value=p.adj.D)
Overall.D<-data.frame(Quantile=crit.D, p.Value=min(p.adj.D), p.Value.WTS=WTS.D,
p.Value.ATS=ATS.D,p.parANOVA=para.pD)
result.D<-list(Contrast=ConD, Multiple.Comparisons=Analysis.of.Relative.Effects.D,
Overall=Overall.D)

#-----#
}
#-----#
#####
#-----#

if(a>1 && d>1 && analyze.inter==TRUE){
#-----Interaction AD-----#
switch(
asy.method,
#-----Multi NORMAL-----#
normal={
p.adj.I<-c()
AsyMethod <- "Normal - Approximation"
TI<-sqrt(N)*(ConIp)/sqrt(c(diag(CVI)))

```

A. Anhang

```

switch(alternative,
#-----Two-sided alternative-----#
two.sided={
for (pp in 1:nci) {
p.adj.I[pp]<-1-pmvnorm(lower=-abs(TI[pp]*dCVI), upper= abs(TI[pp]*dCVI),
                      mean=rep(0,nci),sigma=CVI)[1]}

crit.I<- qmvnorm(conf.level, sigma=rhobfI, tail = "both")$quantile
Lower.I <- ConIp - crit.I/sqrt(N)*sqrt(c(diag(CVI)))
Upper.I <- ConIp + crit.I/sqrt(N)*sqrt(c(diag(CVI)))
},
#-----Alternative= LOWER-----#
less={
for (pp in 1:nci) {
p.adj.I[pp]<-pmvnorm(lower = -Inf*dCVI , upper = TI[pp]*dCVI,
                    mean = rep(0, nci), sigma = CVI)}
crit.I<- qmvnorm(conf.level, sigma = rhobfI, tail = "lower")$quantile
Lower.I <- rep(-1,nci)
Upper.I <- ConIp + crit.I/sqrt(N)*sqrt(c(diag(CVI)))
},
#-----Alternative= GREATER-----#
greater={
for (pp in 1:nci) {
p.adj.I[pp]<-1-pmvnorm(lower = -Inf*dCVI , upper =TI[pp]*dCVI,
                      mean = rep(0, nci), sigma = CVI)}
crit.I<- qmvnorm(conf.level, sigma = rhobfI, tail = "lower")$quantile
Lower.I <-ConIp - crit.I/sqrt(N)*sqrt(c(diag(CVI)))
Upper.I <- rep(1,nci)}
)
},
#-----Multi T-----#
mult.t={
dfT<-round(dfTmultT(ConI,pR,iR,n,a,d))
p.adj.I<-c()
TI<-sqrt(N)*(ConIp)/sqrt(c(diag(CVI)))
AsyMethod <- paste("Multi - T with", round(dfT,rounds), "DF")
switch(alternative,
#-----Two-sided alternative-----#
two.sided={
for (pp in 1:nci) {
p.adj.I[pp]<-1-pmvt(lower=-abs(TI[pp]), upper= abs(TI[pp]), delta=rep(0,nci),
                   df=dfT, corr=rhobfI)[1]}

crit.I<- qmvt(conf.level, df=dfT, corr=rhobfI, tail = "both")$quantile
Lower.I <- ConIp - crit.I/sqrt(N)*sqrt(c(diag(CVI)))
Upper.I <- ConIp + crit.I/sqrt(N)*sqrt(c(diag(CVI)))
},
#-----Alternative= LOWER-----#
less={
for (pp in 1:nci) {
p.adj.I[pp]<-pmvt(lower = -Inf , upper = TI[pp],
                 delta = rep(0, nci), df=dfT, corr = rhobfI)}
crit.I<- qmvt(conf.level, df=dfT, corr = rhobfI, tail = "lower")$quantile
Lower.I <- rep(-1,nci)
Upper.I <- ConIp + crit.I/sqrt(N)*sqrt(c(diag(CVI)))
},

```


A. Anhang

```

#-----Alternative= GREATER-----#
greater={
for (pp in 1:nci) {
p.adj.I[pp]<-1-pmvt(lower =-Inf, upper =TI[pp],
                    delta = rep(0, nci), df=dfT, corr = rhobfI)}
crit.I<- qmvt(conf.level, df=dfT, corr = rhobfI, tail = "lower")$quantile
Lower.I <-ConIp - crit.I/sqrt(N)*sqrt(c(diag(CVI)))
Upper.I <- rep(1,nci)}
)
},
#-----FISHER-TRANS-----#
fisher={
dfT<-round(dfTmultT(ConI,pR,iR,n,a,d))
p.adj.I<-c()
AsyMethod <- paste("Fisher with", round(dfT,rounds), "DF")
Cfisher<-1/2*log((1+ConIp)/(1-ConIp))
Vfisherdev<-diag(c(1/(1-ConIp^2)))
Vfisher<-Vfisherdev%*%CVI%*%t(Vfisherdev)
TI<-sqrt(N)*Cfisher/sqrt(c(diag(Vfisher)))
switch( alternative,
#-----Two-sided alternative-----#
two.sided={
for (pp in 1:nci) {
p.adj.I[pp]<-1-pmvt(lower=-abs(TI[pp]), upper= abs(TI[pp]), delta=rep(0,nci),
                    df=dfT,corr=rhobfI)[1]}

crit.I<- qmvt(conf.level, df=dfT, corr=rhobfI, tail = "both")$quantile
Lower1 <- ConIp - crit.I/sqrt(N)*sqrt(c(diag(CVI)))
Upper1 <- ConIp + crit.I/sqrt(N)*sqrt(c(diag(CVI)))
Lower.I <- (exp(2*Lower1)-1)/(exp(2*Lower1)+1)
Upper.I <- (exp(2*Upper1)-1)/(exp(2*Upper1)+1)
},
#-----Alternative= LOWER-----#
less={
for (pp in 1:nci) {
p.adj.I[pp]<-pmvt(lower = -Inf, upper = TI[pp],
                  delta = rep(0, nci), df=dfT, corr = rhobfI)}
crit.I<- qmvt(conf.level, df=dfT, corr = rhobfI, tail = "lower")$quantile
Lower.I <- rep(-1,nci)
Upper1 <- ConIp + crit.I/sqrt(N)*sqrt(c(diag(CVI)))
Upper.I <- (exp(2*Upper1)-1)/(exp(2*Upper1)+1)
},
#-----Alternative= GREATER-----#
greater={
for (pp in 1:nci) {
p.adj.I[pp]<-1-pmvt(lower =-Inf, upper =TI[pp],
                    delta = rep(0, nci), df=dfT, corr = rhobfI)}
crit.I<- qmvt(conf.level, df=dfT, corr = rhobfI, tail = "lower")$quantile
Lower1 <-ConIp - crit.I/sqrt(N)*sqrt(c(diag(CVI)))
Lower.I <- (exp(2*Lower1)-1)/(exp(2*Lower1)+1)
Upper.I <- rep(1,nci)}
)
}
)
#-----#

```

A. Anhang

```

Analysis.of.Relative.Effects.I <- data.frame(Estimator=round(ConIp,rounds),
      Lower=round(Lower.I,rounds), Upper=round(Upper.I,rounds),
      Statistic = round(TI,rounds), p.Value=p.adj.I)
Overall.I<-data.frame(Quantile=crit.I, p.Value=min(p.adj.I), p.Value.WTS=WTS.I,
      p.Value.ATS=ATS.I,p.parANOVA=para.pI)
result.I<-list(Contrast=ConI, Multiple.Comparisons=Analysis.of.Relative.Effects.I,
      Overall=Overall.I)
}
#-----#
g<-paste("Group ",fla[1]," is denoted by '1', Group ", fla[2]," is denoted by '2'...")
effect.mat<-matrix(p.hat,ncol=d,nrow=a,byrow=TRUE)
n.info<-c()
for(i in 1:a){n.info[i]<-n[[i]]}
data.info2<-data.frame(Sample=1:a,Size=n.info,Effect=effect.mat)
result<-list(Group.Levels=g,Alternative=alternative,Asymptotic.Method=AsyMethod,
      Contrast=type,Data.Info=data.info2, Analysis.Factor.A=result.A,
      Analysis.Factor.D=result.D, Analysis.Interaction.AD=result.I)
#-----#
if (plot.simci == TRUE) {
par(mfrow=c(3,1),oma=c(0,0,2,0))

text.Ci<-paste(conf.level*100, "%", "Simultaneous Confidence Intervals")
Lowerp<-|"
  plot(ConAp,1:nca,xlim=c(-1,1), pch=15,axes=FALSE,xlab="",ylab="")
  points(Lower.A,1:nca, pch=Lowerp,font=2,cex=2)
  points(Upper.A,1:nca, pch=Lowerp,font=2,cex=2)
  abline(v=0, lty=3,lwd=2)
  for (ss in 1:nca){
  polygon(x=c(Lower.A[ss],Upper.A[ss]),y=c(ss,ss),lwd=2)}
  axis(1, at = seq(-1, 1, 0.1))
  axis(2,at=1:nca,labels=connames.A)
  box()
  title(main=c("Grouping Factor A",paste("Type of Contrast:",type)))
#-----#
  plot(ConDp,1:ncd,xlim=c(-1,1), pch=15,axes=FALSE,xlab="",ylab="")
  points(Lower.D,1:ncd, pch=Lowerp,font=2,cex=2)
  points(Upper.D,1:ncd, pch=Lowerp,font=2,cex=2)
  abline(v=0, lty=3,lwd=2)
  for (ss in 1:ncd){
  polygon(x=c(Lower.D[ss],Upper.D[ss]),y=c(ss,ss),lwd=2)}
  axis(1, at = seq(-1, 1, 0.1))
  axis(2,at=1:ncd,labels=connames.D)
  box()
  title(main=c("Repeated Factor D", paste("Type of Contrast:",type)))
#-----#
ConIp<-as.vector(ConIp)

plot(ConIp[1:d], type="o", col=1,
      axes=FALSE, ann=FALSE,ylim=c(min(Lower.I),
      max(Upper.I)),xlim=c(0.95,d+a*0.05),lwd=2)
axis(1, at=1:d, lab=1:d)
for (ss in 1:d){
  polygon(x=c(ss,ss),y=c(Lower.I[ss],Upper.I[ss]),lwd=1,col=1)}

```

A. Anhang

```
axis(2)
box()
group<-c("Group 1")
for(i in 1:(a-1)){
  lines(x=(1:d)+i*0.05,y=ConIp[(i*d+1):((i+1)*d)],type="o",
        pch=21+i,lty=1+i,col=2+i,lwd=2)
  for (ss in 1:d){
    polygon(x=c(ss+i*0.05,ss+i*0.05),y=c(Lower.I[i*d+ss],
      Upper.I[i*d+ss]),lwd=1,col=2+i)}
  group<-c(group, paste("Group ", i+1))
}
abline(h=0, lty=3,lwd=1)
par(xpd=NA)
legend(1,max(Upper.I)+0.2, group, cex=1.2, col=c(1,2:a+1),
      pch=21:(21+a-1), lty=1:a)
par(xpd=F)
title(xlab="Timepoints")
title(ylab="pij-p.j-pi.+p...")
title(main="Interaction AD")

  title(c(text.Ci,paste("Method:", AsyMethod )),outer=TRUE)
}
#-----#
if (correlation == TRUE){
result$Covariance.A <- CVA; result$Covariance.D<-CVD; result$Covariance.AD<-CVI
result$Correlation.A<-rhobfA;result$Correlation.D<-rhobfD
result$Correlation.AD<-rhobfI
}

return(result)
}
#-----#
```

Literatur

- M. G. Akritas und S. F. Arnold. Fully nonparametric hypotheses for factorial designs i: multivariate repeated measures designs. *Journal of the American Statistical Association*, 89:336–343, 1994.
- M. G. Akritas, S.F. Arnold, und E. Brunner. Nonparametric hypotheses and rank statistics for unbalanced factorial designs. *Journal of the American Statistical Association*, 92:258–265, 1997.
- M.G. Akritas und E. Brunner. A unified approach to rank tests for mixed models. *Journal of Statistical Planning and Inference*, 61:249–277, 1997.
- G. E. P. Box. Some theorems on quadratic forms applied in the study of analysis of variance problems, i. effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics*, 25:290–302, 1954.
- F. Bretz, A. Genz, und L. A. Hothorn. On the numerical availability of multiple comparison procedures. *Biometrical Journal*, 43(5):645–656, 2001.
- F. Bretz, T. Hothorn, und P. Westfall. Multiple comparison procedures in linear models. *Compstat 2008: Proceedings In Computational Statistics*, pages 423–431, 2008.
- B. M. Brown und T. P. Hettmansperger. Kruskal-wallis, multiple comparisons and efron dice. *Australian and New Zealand Journal of Statistics*, 44:427 – 438, 2002.
- Domhof S. Brunner, E. und F. Langer. *Nonparametric Analysis of Longitudinal Data*. John Wiley & Sons, New York, 2002.
- E. Brunner und U. Munzel. The nonparametric behrens-fisher problem: Asymptotic theory and a small-sample approximation. *Biometrical Journal*, 1(1):17–21, February 2000.
- E. Brunner und U. Munzel. *Nichtparametrische Datenanalyse*. Springer-Verlag, New York, 2002.
- E. Brunner und M. Placzek. A box-type approximation for general two-sample repeated measures: Technical report. 2011. URL <http://www.ams.med.uni-goettingen.de/download/Techrep-Brunner-Placzek.pdf>.
- E. Brunner und M. L. Puri. Nonparametric methods in factorial designs. *Statistical Papers*, 42(1):1–52, 2001.
- E. Brunner, H. Dette, und A. Munk. Box-type approximations in nonparametric factorial designs. *Journal of the American Statistical Association*, 92:1494–1502, 1997.
- Munzel U. Brunner, E. und M.L. Puri. Rank score tests in factorial designs with repeated measures. *Journal of Multivariate Analysis*, 70:286–317, 1999.

Literatur

- H. Cramér und H. Wold. Some theorems on distribution functions. *Journal of the London Mathematical Society*, 11:290–294, 1936.
- C. S. Davis. *Statistical Methods for the Analysis of Repeated Measurements*. Springer-Verlag, New York, Berlin, Heidelberg, 2002.
- C. W. Dunnett. A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, 50(272):1096–1121, 1955.
- K. R. Gabriel. Simultaneous test procedures - some theory of multiple comparisons. *Annals of Mathematical Statistics*, 40:224–250, 1969.
- X. Gao und M Alvo. Nonparametric multiple comparison procedures for unbalanced two-way factorial designs. *Journal of Statistical Planning and Inference*, 138:3674–3686, 2008.
- X. Gao, M. Alvo, J. Chen, und G. Li. Nonparametric multiple comparison procedures for unbalanced one-way factorial designs. *Journal of Statistical Planning and Inference*, 138(8):2574–2591, 2008.
- A. Genz und F. Bretz. *Computation of Multivariate Normal and t Probabilities*. Springer-Verlag, New York, 2009.
- B. V. Gnedenko. *The Theory of Probability*. Chelsea Publishing, New York, 2nd edition, 1962.
- M. Hasler und L. A. Hothorn. Multiple contrast tests in the presence of heteroscedasticity. *Biometrical Journal*, 50(5):793–800, October 2008.
- E. Herberich, J. Sikorski, und T. Hothorn. *Multiple Comparison Procedures*. Wiley, New York, 2010.
- Y. Hochberg und A. C. Tamhane. *Multiple Comparison Procedures*. Wiley, New York, 1987.
- N.L. Johnson, N. Balakrishnan, und S. Kotz. *Continuous Multivariate Distributions*. John Wiley & Sons, New York, 1980.
- F. Konietschke. *Simultane Konfidenzintervalle für Nichtparametrische Relative Kontrasteffekte*. PhD thesis, Georg-August Universität Göttingen, 2009.
- F. Konietschke und L. A. Hothorn. Evaluation of toxicological studies using a nonparametric shirley-type trend test for comparing several dose levels with a control group. 4:14–27, 2012.
- F. Konietschke, A.C. Bathke, L. A. Hothorn, und E. Brunner. Testing and estimation of purely nonparametric effects in repeated measures designs. *Computational Statistics and Data Analysis*, 54:1895–1905, 2010.

Literatur

- F. Konietzschke, L. A. Hothorn, und E. Brunner. Rank-based multiple test procedures and simultaneous confidence intervals. *Electronic Journal of Statistics*, 6:738–759, 2012.
- W.H. Kruskal und W.A. Wallis. Use of ranks in one-criterion variance analysis. *J. Amer. Stat. Ass.*, 47:583–621, 1952.
- J. A. Lewis. Statistical principles for clinical trials (ich e9) - an introductory note on an international guideline. *Statistics in Medicine*, 18:1903–1942, 1999.
- M. Loève. *Probability Theory I*. Springer Verlag, Berlin, 1977.
- P. Lévy. *Calcul des probabilités*. Gauthiers-Villars, Paris, 1925.
- S. Müller. *Simultane Konfidenzintervalle in Repeated Measures Designs*. Diplomarbeit, Georg-August-Universität Göttingen, 2011.
- M. Mirzaie, E. Brunner, A.H.M. Mahbub-ul Latif, J. Schmitto, S. Fatehpur, und F. Schöndube. A new storage solution for porcine aortic valves. *Annals of Thoracic and Cardiovascular Surgery*, 13, 2007.
- H. Mukerjee, T. Robertson, und F. T. Wright. Comparison of several treatments with a control using multiple contrasts. *Journal of the American Statistical Association*, 82: 902–910, 1987.
- U. Munzel. Linear rank score statistics when ties are present. *Statistics and Probability Letters*, 41:389–395, 1999.
- U. Munzel und L. A. Hothorn. A unified approach to simultaneous rank test procedures in the unbalanced one-way layout. *Biometrical Journal*, 43(5):553–569, 2001.
- J. Orban und D.A. Wolfe. A class of distribution-free twosample tests based on placements. *Journal of the American Statistical Association*, 77:666–672, 1982.
- S. Pfeiffer. *Simultane Konfidenzintervalle für nichtparametrische Effekte in faktoriellen Modellen*. Diplomarbeit, Georg-August-Universität Göttingen, 2010.
- F.H. Ruymgaart. A unified approach to the asymptotic distribution theory of certain midrank statistics. *Statistique non Paramétrique Asymptotique*, 118, J.P. Raoult. *Lecture Notes on Mathematics*, 821, 1980.
- E. Ryu. Simultaneous confidence intervals using ordinal effect measures for ordered categorical outcomes. *Statistics in Medicine*, 28(25):3179–3188, November 2009.
- E. Ryu und A. Agresti. Modeling and inference for an ordinal effect size measure. *Statistics in Medicine*, 27(25):1703 – 1717, October 2008.
- K. Thangavelu und E. Brunner. Wilcoxon mann-whitney test and efrons paradox dice. *Journal of Statistical Planning and Inference*, 137:720 – 737, 2006.

Literatur

- J. W. Tukey. The problem of multiple comparisons. *Unpublished manuscript reprinted in: The Collected Works of John W. Tukey*, 8:Chapman and Hall, June 1953.
- A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1998.