
Auswertungsmethoden bei Experimenten
mit Nachwirkungen und
korrelierten Fehlern

Dissertation

zur Erlangung des Grades
eines Doktors der Naturwissenschaften
der Technischen Universität Dortmund

Der Fakultät Statistik
der Technischen Universität Dortmund
vorgelegt von

Oliver Sailer

Dortmund, im Juni 2009

1. Gutachter: Prof. Dr. Joachim Kunert

2. Gutachter: PD Dr. Guido Knapp

Tag der mündlichen Prüfung: 30. September 2009

*What we obtain too cheap, we esteem too lightly:
it is dearness only that gives every thing its value.*
(Thomas Paine)

Für meine Eltern

Inhaltsverzeichnis

1	Einleitung	1
2	Modelle und Versuchspläne zur Auswertung von Crossover-Studien	5
2.1	Lineare Modelle	5
2.1.1	Parameterschätzer in linearen Modellen	5
2.1.2	Tests in linearen Modellen	8
2.2	Ein lineares Modell für Crossover-Studien	10
2.2.1	Ein Gedankenexperiment	10
2.2.2	Carryover-Effekte und korrelierte Fehler	11
2.3	Versuchspläne für Crossover-Studien	13
3	Auswertung mit Hilfe der Randomisationstheorie	17
3.1	Rechtfertigung eines Modells	17
3.1.1	Randomisationsmodelle	17
3.1.2	Randomisationsverfahren	19
3.2	Auswertung im Zeilen-Spalten-Modell	20
3.2.1	Modelleigenschaften bei Abwesenheit von Nachwirkungen	20
3.2.2	Einfluss von Carryover-Effekten	26
3.3	Simulationsstudie	30
3.3.1	Aufbau	30
3.3.2	Ergebnisse	35
3.4	Bemerkungen	37
4	Das Verfahren von Kunert und Utzig	39
4.1	Abschätzung eines Worst-Case-Szenarios	40
4.1.1	Die obere Schranke von Kunert und Utzig	40
4.1.2	Korrigierte Tests	46
4.2	Die Maximierung des Varianzquotienten	47
4.2.1	Darstellung als Eigenwertproblem	47
4.2.2	Lösung des Maximierungsproblems	51
4.2.3	Verbesserte korrigierte Tests	62
5	Die modifizierte F-Test-Approximation	65
5.1	Das Näherungsverfahren nach Box	65

5.2	Anwendung auf das Modell mit Carryover-Effekten	67
5.2.1	Das MFA-Verfahren von Bellavance et al.	67
5.2.2	Vergleich mit konkurrierenden Auswertungen	71
6	Vergleich von konservativer Varianzschätzung und MFA-Verfahren	75
6.1	Gleichheit der Korrekturfaktoren	75
6.2	Simulationsstudie	77
6.2.1	Aufbau	78
6.2.2	Ergebnisse	80
6.3	Bemerkungen	87
7	Zusammenfassung	89
A	Tabellen und Abbildungen	93
B	R-Code	97
B.1	R-Code zu Kapitel 4	97
B.2	R-Code zu Kapitel 6	99
	Symbolverzeichnis	109
	Literaturverzeichnis	113

1 Einleitung

Die Auswertung geplanter Experimente dient dazu, Erkenntnisse über die Wirkungsweise von Produkten, Medikamenten, Verfahren oder ganz allgemein Behandlungen zu gewinnen. Dazu werden die Behandlungen auf Versuchseinheiten angewandt, an denen man dann eine oder mehrere Zielgrößen beobachtet. Diese Beobachtungen wertet man anschließend aus, um Rückschlüsse auf die Wirkung der Behandlungen zu ziehen.

Insbesondere in der klinischen Forschung, aber auch in der Lebensmittelindustrie oder bei psychometrischen Studien sind die Versuchseinheiten oft menschliche Probanden oder Versuchstiere. Die Zahl der Versuchseinheiten ist in diesen Anwendungsgebieten meist sehr gering. Dies hat vor allem ethische Gründe, zum Beispiel wenn die Probanden unangenehmen Haupt- oder Nebenwirkungen ausgesetzt sind, und folgt andererseits aus Kostengründen. So setzt man beispielsweise in der Lebensmittelsensorik oft wenige, für bestimmte Lebensmittel speziell geschulte Prüfer ein. Ein zu geringer Stichprobenumfang bedeutet aber, dass Unterschiede zwischen den Behandlungen nicht mehr zuverlässig festgestellt werden können. Daher benutzt man für derartige Untersuchungen oft *Crossover-Pläne*, bei denen jeder Proband nacheinander mehreren oder allen zu vergleichenden Behandlungen ausgesetzt wird. Dies reduziert die Zahl der für die Studie zu rekrutierenden Probanden, jede Versuchseinheit agiert gewissermaßen als ihre eigene Kontrolle. Crossover-Pläne werden insbesondere dann eingesetzt, wenn die Behandlungen lediglich von vorübergehender Wirkung sind. Dadurch kann der Einfluss mehrerer nacheinander verabreichter Behandlungen auf den Ausgangszustand der Versuchseinheiten sinnvoll untersucht werden.

Ein klassisches Beispiel für die Anwendung von Crossover-Plänen findet sich bei Cochran et al. (1941), welche die Milchleistung von Kühen in Abhängigkeit des Futters untersuchen. Crossover-Pläne sind auf diesem Forschungsgebiet auch heute die bevorzugten Versuchspläne (Kuoppala et al., 2008). Dies gilt für viele Tierversuche, bei denen die Tiere nicht durch die Behandlung geschädigt werden, zum Beispiel bei der Entwicklung von Hundefutter (Kempe et al., 2004). Crossover-Pläne eignen sich

ebenfalls für Studien in der Lebensmittelsensorik. Beispielsweise untersuchen Bakke und Vickers (2008) die Wahrnehmung von Bitterstoffen, Harker et al. (2008) den Geschmack von Kiwifrüchten und Laureati et al. (2008) die Wirkung der Anreicherung von Lebensmitteln mit Aromastoffen mit Hilfe solcher Versuchspläne. In psychometrischen Studien interessiert man sich oft dafür, wie Probanden bestimmte Aufgaben unter verschiedenen Bedingungen ausführen. Jones und Kenward (2003)[S. 243–246] beschreiben eine Anwendung von Crossover-Plänen in einer Studie von McNulty (1986) zur Wahrnehmung von bewegten Objekten auf Bildschirmen. Senn (2002)[Kapitel 5] erwähnt verschiedene Situationen, in denen Crossover-Pläne auch in klinischen Studien benutzt werden. Dazu gehören zum Beispiel Dosis-Findungs-Studien, Kombinationstherapien, Studien zur Bioäquivalenz und Vergleiche mit dem Gold Standard bei chronischen Krankheiten.

Werden mehrere Behandlungen im Zeitablauf angewendet, müssen bei der statistischen Modellierung *Nachwirkungen* von Behandlungen, sogenannte *Carryover-Effekte*, und *korrelierte Fehler* berücksichtigt werden.

Diese Arbeit beschäftigt sich mit der Frage, wie man Daten aus Experimenten mit Nachwirkungen und korrelierten Fehlern in linearen Modellen mit festen Effekten auswerten kann. Dabei beschränken wir uns auf Versuche mit nur einer Zielgröße und mindestens drei Behandlungen. Zur Auswertung der insbesondere in klinischen Studien häufig angewandten Pläne zum Vergleich von zwei Behandlungen existiert bereits eine Vielzahl von Ergebnissen. Zusammenfassungen dieser Ergebnisse finden sich zum Beispiel in Jones und Kenward (2003) oder Senn (2002). Dort werden auch Auswertungsverfahren für lineare Modelle mit zufälligen Effekten vorgestellt.

In Kapitel 2 werden zunächst grundlegende Begriffe und Ergebnisse der statistischen Inferenz für lineare Modelle wiederholt. Anschließend führen wir ein geeignetes Modell zur Auswertung von Crossover-Studien ein und befassen uns mit der Frage nach geeigneten Versuchsplänen für derartige Studien.

Wenn keine Carryover-Effekte auftreten, lassen sich die Beobachtungen aus Crossover-Studien mit Hilfe einfacher linearer Modelle analysieren. Azaïs und Druilhet (1997) und Bailey (1985) zeigen, wie man mit Hilfe geeigneter randomisierter nachbarbalancierter Versuchspläne eine Auswertung im einfachen Blockmodell beziehungsweise im Zeilen-Spalten-Modell rechtfertigen kann. In Kapitel 3 wird gezeigt, dass man die Klasse der Versuchspläne, für die das Zeilen-Spalten-Modell gerechtfertigt ist, über diese Ergebnisse hinaus erweitern kann. Damit lässt sich das Verfahren auch

auf unvollständige Blockpläne anwenden, das heißt auf Situationen, bei denen nicht alle Behandlungen bei jeder Versuchseinheit vorkommen. Wir werden ebenfalls untersuchen, inwiefern sich die Auswertung noch rechtfertigen lässt, wenn Carryover-Effekte vorliegen, und wie robust die Auswertung ist, wenn Versuchspläne benutzt werden, die den in Kapitel 3 betrachteten Voraussetzungen nur näherungsweise genügen.

Eine andere Vorgehensweise zur Auswertung von Crossover-Plänen besteht darin, mögliche Carryover-Effekte im Modell zu berücksichtigen. Die herkömmliche Auswertung mit Hilfe von Kleinste-Quadrate-Schätzern (KQ-Schätzern) im linearen Modell führt bei gleichzeitigem Vorliegen korrelierter Fehler zu verzerrten Varianzschätzern. Das hat zur Folge, dass F- und t-Tests auf der Grundlage dieser Schätzer das nominelle Testniveau nicht einhalten und unter Umständen stark konservativ sind. Von besonderer Bedeutung ist hierbei die Tatsache, dass die Korrelationsstruktur in der Regel nicht bekannt ist und bei geringen Stichprobenumfängen nur schlecht geschätzt werden kann. In der Literatur gibt es hierzu verschiedene Lösungsvorschläge. So kann man statt KQ-Schätzern verallgemeinerte KQ-Schätzer verwenden. Ein Spezialfall hierbei sind Repeated-Measurements-Modelle mit zufälligen Einheiteneffekten (Kenward und Roger, 1997).

Kunert und Utzig (1993) schlagen ein Verfahren auf Grundlage einer konservativen Varianzabschätzung vor. Sie gehen zunächst von einem Modell mit unkorrelierten Fehlern aus und untersuchen dann, wie sich die Varianzschätzer verhalten, falls doch korrelierte Beobachtungen vorliegen. Dazu betrachten sie den Quotienten aus der Varianz bei korrelierten Fehlern und dem Erwartungswert der Varianz bei unkorrelierten Fehlern. Das Worst-Case-Szenario bezüglich der Kovarianzmatrix der Fehler liefert dann einen Korrekturfaktor, mit dessen Hilfe man die Varianz konservativ abschätzen kann. Der ungünstigste Fall wurde bisher nur für Versuchspläne mit drei Versuchsperioden exakt berechnet, bei mehr als drei Perioden war lediglich eine grobe obere Schranke bekannt. In Kapitel 4 untersuchen wir das Verfahren von Kunert und Utzig (1993) näher und bestimmen dabei das Worst-Case-Szenario für eine beliebige Zahl von Behandlungen pro Versuchseinheit. Dies ermöglicht eine weniger konservative Varianzschätzung. Mit Hilfe unserer verbesserten Schranke konstruieren wir dann weniger konservative Tests auf Behandlungsunterschiede.

Bellavance et al. (1996) vergleichen verschiedene Testverfahren für das Modell mit Carryover-Effekten und korrelierten Fehlern. Dabei schneidet das sogenannte MFA-

Verfahren, das auf einer Approximation der F-Statistik nach Box (1954) beruht, sehr gut ab. Wir stellen dieses Verfahren in Kapitel 5 kurz vor.

In Kapitel 6 vergleichen wir das MFA- Verfahren mit unserer Verbesserung der Methode von Kunert und Utzig (1993). Hierbei untersuchen wir unter anderem das MFA-Verfahren für die ungünstigste Kovarianzmatrix. Theoretische Argumente sowie eine Simulationsstudie führen uns zum Schluss, dass das vermeintlich konservative Testverfahren von Kunert und Utzig (1993) mit unserer verbesserten Schranke nicht notwendig konservativ ist. Darüberhinaus ist die Power des MFA-Verfahrens oft größer als die unseres Konkurrenzverfahrens.

Wir fassen die wichtigsten Ergebnisse dieser Arbeit in Kapitel 7 zusammen und geben schließlich einen kurzen Ausblick auf mögliche Ergänzungen.

2 Modelle und Versuchspläne zur Auswertung von Crossover-Studien

In diesem Kapitel skizzieren wir zunächst einige bekannte Begriffe, Ergebnisse und Schreibweisen aus der Theorie der linearen Modelle, die in Lehrbüchern wie Rao et al. (2008) benutzt werden. Danach führen wir mit Hilfe eines Gedankenexperiments das grundlegende lineare Modell zur Auswertung von Crossover-Plänen ein. Anschließend stellen wir aus der Literatur bekannte geeignete Versuchspläne zur Auswertung von Crossover-Studien vor.

Grundsätzlich werden Vektoren und Matrizen in dieser Arbeit fett gesetzt, wobei Vektoren mit Hilfe von Kleinbuchstaben dargestellt und Matrizen großgeschrieben werden. Zufallsvariablen, Zufallsvektoren und ihre Realisationen werden stets kleingeschrieben. Eine Übersicht über die wichtigsten Symbole findet sich im Symbolverzeichnis (S. 109 ff.).

2.1 Lineare Modelle

2.1.1 Parameterschätzer in linearen Modellen

Wir betrachten lineare Modelle der Form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \mathbb{E}(\boldsymbol{\epsilon}) = \mathbf{0}, \quad \text{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}, \quad (2.1)$$

wobei $\mathbf{y} \in \mathbb{R}^N$ der Beobachtungsvektor, $\mathbf{X} \in \mathbb{R}^{N \times q}$ die nicht-stochastische Designmatrix, $\boldsymbol{\beta} \in \mathbb{R}^q$ der Parametervektor und $\boldsymbol{\epsilon} \in \mathbb{R}^N$ der Fehlervektor ist. $\boldsymbol{\Sigma}$ ist die

nicht-negativ definite (*n.n.d.*) Kovarianzmatrix von $\boldsymbol{\epsilon}$. E bezeichne Erwartungswerte und Cov Kovarianzmatrizen.

Falls \mathbf{X} keinen vollen Spaltenrang besitzt, gibt es keinen linearen erwartungstreuen Schätzer für $\boldsymbol{\beta}$. Für bestimmte Linearkombinationen des Parametervektors kann es aber erwartungstreue Schätzer geben.

Definition 2.1. Die parametrische Funktion $\tilde{\boldsymbol{\ell}}^T \boldsymbol{\beta}$ heißt schätzbar, falls es eine lineare Funktion $\mathbf{k}^T \mathbf{y}$ der Beobachtungen gibt, so dass $E(\mathbf{k}^T \mathbf{y}) = \tilde{\boldsymbol{\ell}}^T \boldsymbol{\beta}$. Ein Schätzer $\mathbf{k}^{*T} \mathbf{y}$ heißt bester linearer erwartungstreuer Schätzer oder BLUE für $\tilde{\boldsymbol{\ell}}^T \boldsymbol{\beta}$, falls er unter allen linearen erwartungstreuen Schätzern für $\tilde{\boldsymbol{\ell}}^T \boldsymbol{\beta}$ in Modell (2.1) die kleinste Varianz besitzt.

Oft nimmt man an, dass die Fehler in Modell (2.1) unabhängig identisch verteilt (u.i.v.) sind, also dass $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_N$. Wir schreiben \mathbf{A}^- für eine beliebige verallgemeinerte Inverse der Matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$. Eine mögliche Wahl für \mathbf{A}^- ist die Moore-Penrose-Inverse von \mathbf{A} , \mathbf{A}^+ , vergleiche Seber (2008)[Kapitel 7.1]. Weiterhin bezeichne $\text{Rg}(\mathbf{A})$ den Rang von \mathbf{A} und var die Varianz einer Zufallsvariablen. Aus Rao et al. (2008)[Theorem 3.6] folgt

Lemma 2.1. Der Kleinste-Quadrate-Schätzer (KQ-Schätzer) $\tilde{\boldsymbol{\ell}}^T \hat{\boldsymbol{\beta}}$ mit

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^- \mathbf{X}^T \mathbf{y} \quad (2.2)$$

ist in Modell (2.1) mit $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_N$ der BLUE für $\tilde{\boldsymbol{\ell}}^T \boldsymbol{\beta}$.

In diesem Fall ist

$$\text{var}(\tilde{\boldsymbol{\ell}}^T \hat{\boldsymbol{\beta}}) = \sigma^2 \tilde{\boldsymbol{\ell}}^T (\mathbf{X}^T \mathbf{X})^- \tilde{\boldsymbol{\ell}}.$$

Ein erwartungstreuer Schätzer für σ^2 ist gegeben durch

$$\hat{\sigma}^2 = \frac{1}{f_E} \hat{\boldsymbol{\epsilon}}^T \hat{\boldsymbol{\epsilon}},$$

wobei $f_E = N - \text{Rg}(\mathbf{X})$ und $\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}$.

Falls $\boldsymbol{\Sigma}$ positiv definit (*p.d.*) ist, so ist der verallgemeinerte KQ-Schätzer (Aitken-Schätzer) $\tilde{\boldsymbol{\ell}}^T \hat{\boldsymbol{\beta}}_{GLS}$ BLUE für $\tilde{\boldsymbol{\ell}}^T \boldsymbol{\beta}$ (Rao et al., 2008)[S. 217]. Dabei ist

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^- \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{y}.$$

Die Anwendung des Aitken-Schätzers setzt voraus, dass Σ bekannt ist. In praktischen Anwendungen ist diese Annahme oft unrealistisch. In diesem Fall kann man die Kovarianzmatrix der Fehler aus den Daten schätzen und in der Formel zur Berechnung des Aitken-Schätzers Σ durch einen Schätzer $\hat{\Sigma}$ ersetzen. Der resultierende Schätzer

$$\hat{\beta}_{EGLS} = \left(\mathbf{X}^T \hat{\Sigma}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \hat{\Sigma}^{-1} \mathbf{y}$$

heißt Empirical-Generalized-Least-Squares-Schätzer (*EGLS-Schätzer*). Er ist im Allgemeinen nicht linear in \mathbf{y} . Wir sehen in Kapitel 5.2.2, dass die EGLS-Schätzung für Crossover-Pläne mit geringem Stichprobenumfang nicht unbedingt sinnvoll ist. In dieser Arbeit beschränken wir uns daher auf den KQ-Schätzer (2.2).

Wenn das Modell (2.1) einen *Behandlungseffekt* $\tau \in \mathbb{R}^t$ enthält, lässt sich β in zwei Komponenten $\beta^{(1)} = \tau$ und $\beta^{(2)} \in \mathbb{R}^{q-t}$ so zerlegen, dass $\beta = [\tau^T, \beta^{(2)T}]^T$. Entsprechend partitioniert man die Designmatrix zu $\mathbf{X} = [\mathbf{T}, \mathbf{X}_2]$ und erhält

$$\mathbf{y} = [\mathbf{T}, \mathbf{X}_2] \begin{pmatrix} \tau \\ \beta^{(2)} \end{pmatrix} + \epsilon, \quad \mathbb{E}(\epsilon) = \mathbf{0}, \quad \text{Cov}(\epsilon) = \Sigma. \quad (2.3)$$

Wir interessieren uns für Aussagen über die Behandlungseffekte τ . Die weiteren Parameter $\beta^{(2)}$ sind dann Störgrößen, die lediglich berücksichtigt werden, um zu präziseren Aussagen über τ zu kommen. Von besonderem Interesse sind dabei *Behandlungskontraste*.

Definition 2.2. Sei $\ell \in \mathbb{R}^t$ und $\sum_{i=1}^t \ell_i = 0$. Dann heißt die parametrische Funktion $\psi = \ell^T \tau$ *Behandlungskontrast*. Falls zusätzlich $\sum_{i=1}^t \ell_i^2 = 1$, so heißt ψ *standardisierter Behandlungskontrast*.

In den folgenden Kapiteln greifen wir häufig auf *orthogonale Projektoren* zurück. Zu $\mathbf{A} \in \mathbb{R}^{m \times n}$ ist

$$\omega_{\mathbf{A}} = \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$$

der orthogonale Projektor auf den Spaltenraum von \mathbf{A} . Der orthogonale Projektor $\mathbf{I}_m - \omega_{\mathbf{A}}$ wird mit $\omega_{\mathbf{A}}^\perp$ bezeichnet. Beispielsweise können wir $\hat{\epsilon}^T \hat{\epsilon}$ in Lemma 2.1 durch $\mathbf{y}^T \omega_{\mathbf{X}}^\perp \mathbf{y}$ ausdrücken.

Lemma 2.2. Sei $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2]$ eine partitionierte Matrix, $\mathbf{C} = \mathbf{A}_1^T \boldsymbol{\omega}_{\mathbf{A}_2}^\perp \mathbf{A}_1$ und $\mathbf{D} = (\mathbf{A}_2^T \mathbf{A}_2)^-$. Dann ist

$$(\mathbf{A}^T \mathbf{A})^- = \begin{bmatrix} \mathbf{C}^- & -\mathbf{C}^- \mathbf{A}_1^T \mathbf{A}_2 \mathbf{D} \\ -\mathbf{D} \mathbf{A}_2^T \mathbf{A}_1 \mathbf{C}^- & \mathbf{D} + \mathbf{D} \mathbf{A}_2^T \mathbf{A}_1 \mathbf{C}^- \mathbf{A}_1^T \mathbf{A}_2 \mathbf{D} \end{bmatrix}$$

eine verallgemeinerte Inverse von $\mathbf{A}^T \mathbf{A}$.

Dieses Ergebnis ist ein Spezialfall von 14.41(b) in Seber (2008). Schreiben wir $\mathbf{0}_n$ für den Nullvektor des \mathbb{R}^n , so können wir $\psi = \boldsymbol{\ell}^T \boldsymbol{\tau}$ in Modell (2.1) darstellen als $\tilde{\boldsymbol{\ell}}^T \boldsymbol{\beta}$ mit $\tilde{\boldsymbol{\ell}} = (\boldsymbol{\ell}^T, \mathbf{0}_{q-t}^T)^T$. Mit Lemma 2.2 angewandt auf \mathbf{X} erhalten wir dann aus Lemma 2.1 den folgenden Satz, siehe auch Hedayat und Afsarinejad (1978).

Satz 2.1. Falls der Kontrast $\psi = \boldsymbol{\ell}^T \boldsymbol{\tau}$ in Modell (2.3) schätzbar ist, so gilt für den KQ-Schätzer von ψ

$$\hat{\psi} = \boldsymbol{\ell}^T \mathbf{C}^- \mathbf{T}^T \boldsymbol{\omega}_{\mathbf{X}_2}^\perp \mathbf{y},$$

wobei \mathbf{C}^- eine beliebige verallgemeinerte Inverse von $\mathbf{C} = \mathbf{T}^T \boldsymbol{\omega}_{\mathbf{X}_2}^\perp \mathbf{T}$ ist. Falls $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_N$, so ist $\hat{\psi}$ der BLUE für ψ und die Varianz des Kontrasts erfüllt

$$\text{var}(\hat{\psi}) = \sigma^2 \boldsymbol{\ell}^T \mathbf{C}^- \boldsymbol{\ell}.$$

Falls in einem Modell ein Schätzer $\widehat{\text{var}}(\hat{\psi})$ für die Varianz eines Behandlungskontrasts die tatsächliche Varianz im Mittel nicht unterschätzt, das heißt, falls

$$\text{E}(\widehat{\text{var}}(\hat{\psi})) \geq \text{var}(\hat{\psi}),$$

so nennt man $\widehat{\text{var}}(\hat{\psi})$ konservativ für $\text{var}(\hat{\psi})$. Andernfalls nennen wir den Varianzschätzer anti-konservativ.

2.1.2 Tests in linearen Modellen

In der Situation von Modell (2.3) kann man bestimmte Hypothesen bezüglich der Behandlungseffekte testen. Zum einen interessieren wir uns für den Test der Globalhypothese $H_0^g : \tau_1 = \dots = \tau_t$ gegen die Alternative $H_1^g : \exists i \neq j : \tau_i \neq \tau_j$. Zum anderen betrachten wir den Test auf den Behandlungskontrast ψ mit Nullhypothese $H_0^k : \psi = \psi_0$ gegen die zweiseitige Alternative $H_1^k : \psi \neq \psi_0$ mit festem $\psi_0 \in \mathbb{R}$.

Wir testen die obigen Hypothesen mit Hilfe des F -Tests und des t -Tests aus folgender Definition, vergleiche Witting (1985)[Kapitel 3.4 und 4.2].

Definition 2.3. Wir bezeichnen das ν -Quantil der F -Verteilung mit f_1 und f_2 Freiheitsgraden durch $F_{f_1, f_2; \nu}$. Der F -Test lehnt die Hypothese $H_0^g : \tau_1 = \dots = \tau_t$ zugunsten der Alternativen $H_1^g : \exists i \neq j : \tau_i \neq \tau_j$ ab, falls

$$F = \frac{MST}{MSE} > F_{t-1, f_E; 1-\alpha}.$$

Dabei ist $MST = \mathbf{y}^T(\boldsymbol{\omega}_X - \boldsymbol{\omega}_T)\mathbf{y}/(t-1)$, $MSE = \mathbf{y}^T \boldsymbol{\omega}_X^\perp \mathbf{y}/f_E = \hat{\sigma}^2$.

Wir bezeichnen das ν -Quantil der t -Verteilung mit f Freiheitsgraden durch $t_{f; \nu}$. Der zweiseitige t -Test lehnt die Hypothese $H_0^k : \psi = \psi_0$ zugunsten der Alternativen $H_1^k : \psi \neq \psi_0$ ab, falls

$$t = \frac{|\hat{\psi} - \psi_0|}{\sqrt{\widehat{\text{var}}(\hat{\psi})}} > t_{f_E; 1-\alpha/2}.$$

Dabei ist $\widehat{\text{var}}(\hat{\psi}) = \hat{\sigma}^2 \boldsymbol{\ell}^T \mathbf{C}^{-1} \boldsymbol{\ell}$ mit $\mathbf{C} = \mathbf{T}^T \boldsymbol{\omega}_{X_2}^\perp \mathbf{T}$.

Neben den genannten Nullhypothesen können auch noch weitere, möglicherweise zusammengesetzte Hypothesen über die Modellparameter getestet werden. Das Niveau eines Tests ist allgemein das Supremum der Ablehnwahrscheinlichkeiten unter der Nullhypothese und die *Power* oder *Macht* die Ablehnwahrscheinlichkeit für eine Parameterkombination aus der Alternativen. Wir nennen einen Test *konservativ zum nominellen Niveau* α , falls das Niveau des Tests kleiner oder gleich α ist. Ein Test heißt *anti-konservativ zum nominellen Niveau* α , falls er das Niveau α nicht einhält, das heißt, falls er ein Test zum Niveau $\alpha_1 > \alpha$ ist. Der größte Wert, bei dem F- beziehungsweise t-Test gerade nicht ablehnen, ist der *kritische Wert* des jeweiligen Tests.

Oft nimmt man an, dass die Modellfehler unabhängig identisch normalverteilt sind, so dass

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_N) \quad (2.4)$$

Es ist wohlbekannt, dass F- und t-Test im linearen Modell dann gewisse Optimalitätseigenschaften besitzen, siehe zum Beispiel Witting (1985)[Kapitel 3.4.3 und 4.2.1]. Insbesondere gilt

Lemma 2.3. *Unter (2.4) ist der F-Test aus Definition 2.3 in Modell (2.3) ein Test zum Niveau α für H_0^g gegen H_1^g und der t-Test aus Definition 2.3 ein Test zum Niveau α für H_0^k gegen H_1^k .*

Im nächsten Abschnitt motivieren wir anhand eines Gedankenexperiments ein konkretes Modell für Crossover-Studien.

2.2 Ein lineares Modell für Crossover-Studien

2.2.1 Ein Gedankenexperiment

Nehmen wir an, dass wir vier verschiedene farblose Limonaden mit Hilfe mehrerer geschulter Prüfer hinsichtlich der Intensität ihres Zitrusgeschmacks vergleichen wollen. Wir gehen davon aus, dass die Bewertungen der Versuchseinheiten, hier die Prüfer, wie im schematischen Versuchsplan in Abbildung 2.1 dargestellt, von verschiedenen Komponenten beeinflusst werden.

Jede der vier Behandlungen hat eine von der Versuchsperson und der Versuchsanordnung unabhängige Wirkung, den Behandlungseffekt. Jeder Prüfer kostet die vier Limonaden, in der Abbildung mit A , B , C und D gekennzeichnet, in einer durch den Versuchsplan vorgegebenen Reihenfolge. Es ist denkbar, dass einer der Tester Zitrusgeschmack stets stärker wahrnimmt als ein anderer der Prüfer, das heißt wir modellieren einen Einfluss der Versuchseinheiten, den *Einheiteneffekt*. Außerdem könnte es sein, dass das Zitrusaroma durch einen Gewöhnungseffekt nach einiger Zeit von allen Prüfern weniger stark wahrgenommen wird als zu Beginn der Testreihe, so dass wir einen *Periodeneffekt* mit in unser Modell aufnehmen. Neben dem direkten Einfluss der Behandlungen kann es auch noch zu einer Nachwirkung von Behandlungen auf folgende Behandlungen kommen. Nehmen wir an, dass die mit A gekennzeichnete Limonade so anhaltend süß schmeckt, dass jede direkt auf A folgende Limonade ebenfalls als süß wahrgenommen wird. Dieser von Limonade A ausgehende Carryover-Effekt hängt also nicht von der voran gehenden Versuchsperiode, sondern nur von der unmittelbar voran gehenden Behandlung ab. In dieser Arbeit verstehen wir unter dem Behandlungseffekt lediglich den direkten, unmittelbaren Einfluss einer Behandlung in der Periode, in der die Behandlung angewandt wird, und nicht den Carryover-Effekt der Behandlung.

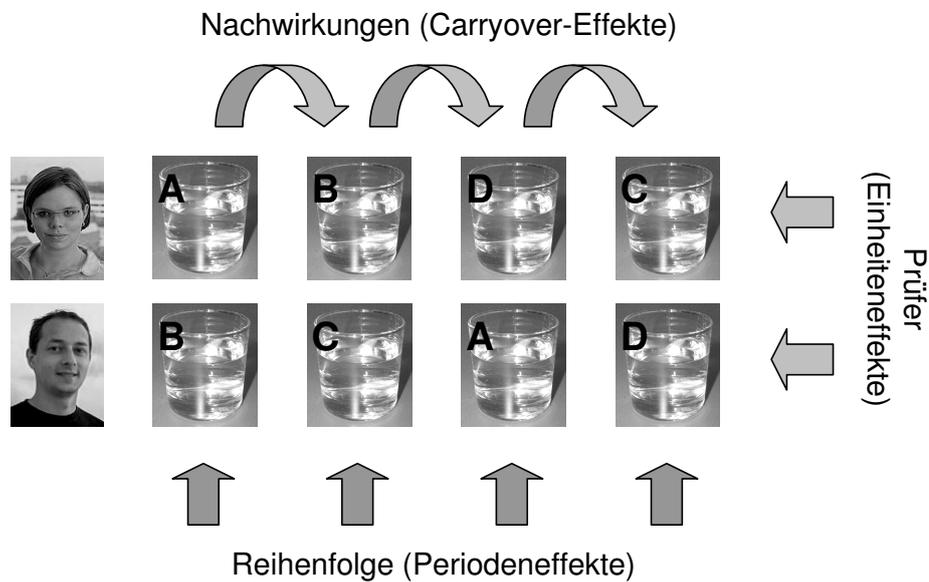


Abbildung 2.1: Schematischer Versuchsaufbau eines sensorischen Experiments zum Vergleich von vier Produkten. Jeder Prüfer verkostet die Proben in einer vorgegebenen Reihenfolge. Bei der Modellierung des Experiments werden Einheiten-, Perioden- und Carryover-Effekte berücksichtigt.

Alle oben beschriebenen Effekte werden als feste Effekte der konkreten Einheiten, Perioden und Behandlungen angesehen. Weitere, in obiger Darstellung nicht betrachtete Einflüsse auf die Bewertung der Limonaden modellieren wir als zufällige Restfehler. Wir gehen davon aus, dass Beobachtungen an verschiedenen Versuchspersonen unabhängig voneinander sind. Es ist aber durchaus denkbar, dass im Zeitablauf an einer Versuchseinheit gemessene Beobachtungen sich ähnlicher sind als Beobachtungen zwischen verschiedenen Einheiten. Wir lassen daher zu, dass die Fehler an einer Einheit korreliert sind. Das oben beschriebene Modell formulieren wir nun formal als lineares Modell.

2.2.2 Carryover-Effekte und korrelierte Fehler

Wir wollen $t \geq 3$ Behandlungen vergleichen. An jeder von n Versuchseinheiten werden p verschiedene Beobachtungen gemacht, so dass insgesamt $N = np$ Beobachtungen vorliegen. Dabei sei $p \leq t$. Wir schreiben y_{ij} für die j -te Beobachtung an Versuchs-

einheit i , $i = 1, \dots, n$, $j = 1, \dots, p$. Der Versuchsplan (Design) d ordnet (i, j) die Behandlung $d(i, j) \in \{1, \dots, t\}$ zu und für die Beobachtungen gelte das folgende lineare Modell (Stufken, 1996)

Definition 2.4. *Das Modell mit Carryover-Effekten und korrelierten Fehlern ist das lineare Modell*

$$y_{ij} = \mu + \alpha_i + \pi_j + \tau_{d(i,j)} + \rho_{d(i,j-1)} + \epsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, p. \quad (2.5)$$

Hier ist μ das allgemeine Mittel, α_i der Einheiten effekt von Einheit i , π_j der Periodeneffekt der Periode j , und $\tau_{d(i,j)}$ der Effekt der zugeordneten Behandlung. Weiterhin ist $\rho_{d(i,j-1)}$ der Carryover-Effekt der Behandlung $d(i, j-1)$, der auf die (i, j) -te Beobachtung nachwirkt. Dabei setzen wir $\rho_{d(i,0)} = 0$. Für den Fehlerterm ϵ_{ij} nehmen wir an, dass die zweiten Momente existieren, wobei $E(\epsilon_{ij}) = 0$ und

$$\text{Cov}(\epsilon_{ij}, \epsilon_{i'j'}) = \begin{cases} \sigma_{jj'}, & \text{falls } i = i' \\ 0, & \text{sonst} \end{cases}.$$

Beobachtungen zwischen verschiedenen Einheiten sind unkorreliert. Die Beobachtungen an einer Einheit sind möglicherweise korreliert, wobei diese Korrelationen für jede Einheit gleich sind. Diese Annahme kann durch die Randomisierung der Versuchseinheiten gerechtfertigt werden, vergleiche hierzu auch Kunert und Utzig (1993). Die ϵ_{ij} und damit auch die y_{ij} sind Zufallsvariablen.

Wir schreiben $\mathbf{1}_r$ für den Einsenvektor des \mathbb{R}^r , \mathbf{I}_r für die Einheitsmatrix des $\mathbb{R}^{r \times r}$ und $\mathbf{A}_1 \otimes \mathbf{A}_2$ für das Kroneckerprodukt zweier Matrizen \mathbf{A}_1 und \mathbf{A}_2 . In Vektorschreibweise erhalten wir dann für das Modell mit Carryover-Effekten

$$\mathbf{y} = \mathbf{1}_{np}\mu + \mathbf{U}\boldsymbol{\alpha} + \mathbf{P}\boldsymbol{\pi} + \mathbf{T}\boldsymbol{\tau} + \mathbf{F}\boldsymbol{\rho} + \boldsymbol{\epsilon}, \quad E(\boldsymbol{\epsilon}) = \mathbf{0}, \quad \text{Cov}(\boldsymbol{\epsilon}) = \mathbf{I}_n \otimes \mathbf{S}. \quad (2.6)$$

Dabei ist $\mathbf{y} = (y_{11}, \dots, y_{np})^T$ der Vektor der Beobachtungen, $\mathbf{1}_{np}$ der Einsenvektor der Länge np , $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)^T$ der Vektor der Einheiten effekte und $\boldsymbol{\pi} = (\pi_1, \dots, \pi_p)^T$ der Vektor der Periodeneffekte. $\mathbf{U} = \mathbf{I}_n \otimes \mathbf{1}_p$ und $\mathbf{P} = \mathbf{1}_n \otimes \mathbf{I}_p$ sind die entsprechenden Designmatrizen für Einheiten- und Periodeneffekte. Der Vektor der Behandlungseffekte ist gegeben durch $\boldsymbol{\tau} = (\tau_1, \dots, \tau_t)^T$ und der Vektor der

Nachwirkungseffekte durch $\boldsymbol{\rho} = (\rho_1, \dots, \rho_t)^T$. \mathbf{T} und \mathbf{F} sind die zum Versuchsplan d gehörigen Designmatrizen für direkte Effekte und Nachwirkungen. Mit der Notation

$$\mathbf{V} = \begin{bmatrix} \mathbf{0} & 0 \\ \mathbf{I}_{p-1} & \mathbf{0} \end{bmatrix}$$

ist $\mathbf{F} = (\mathbf{I}_n \otimes \mathbf{V})\mathbf{T}$. Für den Fehlervektor $\boldsymbol{\epsilon} = (\epsilon_{11}, \dots, \epsilon_{np})^T$ gilt $E(\boldsymbol{\epsilon}) = \mathbf{0}$ und $\text{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma} = \mathbf{I}_n \otimes \mathbf{S}$. Das Element an der Stelle (j, j') von $\mathbf{S} \in \mathbb{R}^{p \times p}$ ist gleich $\sigma_{jj'}$ und \mathbf{S} n.n.d.

Oft unterstellt man, dass die Beobachtungen y_{11}, \dots, y_{np} u.i.v. sind. In diesem Fall ist in Modell (2.6) $\mathbf{S} = \sigma^2 \mathbf{I}_p$ und damit $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_{np}$. Man erhält

$$\mathbf{y} = \mathbf{1}_{np}\mu + \mathbf{U}\boldsymbol{\alpha} + \mathbf{P}\boldsymbol{\pi} + \mathbf{T}\boldsymbol{\tau} + \mathbf{F}\boldsymbol{\rho} + \boldsymbol{\epsilon}, \quad E(\boldsymbol{\epsilon}) = \mathbf{0}, \quad \text{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}_{np}. \quad (2.7)$$

Insoweit wir Tests auf Behandlungseffekte untersuchen, nehmen wir an, dass die Fehler normalverteilt sind. Also gelte in Modell (2.7) Gleichung (2.4) und in Modell (2.6) sei $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{S})$. Kunert et al. (2002) zeigen, dass diese Annahme bei Crossover-Studien eine sinnvolle Auswertung selbst dann ermöglichen kann, wenn die Zielgröße \mathbf{y} nicht metrisch sondern lediglich ordinal skaliert ist.

Offensichtlich sind die Modelle (2.6) und (2.7) Spezialfälle von (2.3). Wenn man weiß, dass den Daten Modell (2.7) zugrunde liegt, so wendet man einfach die in Kapitel (2.1) beschriebenen KQ-Schätzer und Tests an, um Aussagen über die Behandlungseffekte zu gewinnen. Allerdings ist es bei Crossover-Studien oft nicht realistisch anzunehmen, dass (2.7) gilt. In Kapitel 4 wird untersucht, inwiefern eine modifizierte Auswertung unter (2.7) bei Gültigkeit von (2.6) möglich ist. In Kapitel 5 betrachten wir eine Auswertung unter (2.6). In Kapitel 3 wird mit der Randomisationstheorie eine alternative Herangehensweise an das Problem der Modellunsicherheit untersucht.

2.3 Versuchspläne für Crossover-Studien

Es bleibt noch zu klären, welche Art von Versuchsplänen für die Auswertung der hier beschriebenen Crossover-Studien geeignet ist. Wir können jede Versuchseinheit als Block der Länge p auffassen. Wir bezeichnen mit $\Lambda_{t,n,p}$ die Menge der Versuchspläne zum Vergleich von t Behandlungen mit n Blöcken der Länge p . Einen Versuchsplan

$d \in \Lambda_{t,n,p}$ bezeichnen wir im Folgenden auch als *Crossover-Plan*. Wir stellen die Einheiten der in den Kapiteln 3 und 6 abgebildeten Versuchspläne als Spalten von d dar und die Perioden als Zeilen. Wir können $d \in \Lambda_{t,n,p}$ auch als Blockplan in den Perioden auffassen. Damit haben wir in den Zeilen p Blöcke der Länge n . Kiefer (1975) benutzt folgende

Definition 2.5. Sei d ein Blockplan mit n_B Blöcken der Länge l_B und r_{di} gebe an, wie oft Behandlung i in d vorkommt. Weiterhin gebe n_{dij} an, wie oft Behandlung i in Block j vorkommt und wir definieren $h_{dih} = \sum_j n_{dij} n_{dhj}$. Ein Blockplan, der die folgenden Eigenschaften BBD1-BBD3 erfüllt, heißt *balancierter Blockplan*.

BBD1. $r_{di} = n_B l_B / t$ für alle i .

BBD2. $h_{dih} = h_{d12}$ für alle $i < h$.

BBD3. $|n_{dij} - l_B / t| < 1$ für alle i, j .

Falls $d \in \Lambda_{t,n,p}$ ein balancierter Blockplan in Zeilen und Spalten ist, so heißt *verallgemeinertes Youden-Design oder GYD*. Falls für ein GYD d ein $m \in \mathbb{N}$ existiert, so dass $n = mt$ oder falls die Zahl der Perioden ein ganzzahliges Vielfaches von t ist, so heißt d *reguläres GYD*.

Die Definition verallgemeinerter Youden-Designs berücksichtigt auch Pläne, bei denen an einer Einheit nicht alle Behandlungen auftreten. Solche Pläne wird man insbesondere bei großem t anwenden.

GYD besitzen die Eigenschaft, dass die Behandlungen möglichst gleichmäßig über die Zeilen und Spalten verteilt sind. Bei Crossover-Plänen scheint es intuitiv sinnvoll, zu verhindern, dass eine Behandlung stets den gleichen Nachfolger hat. Stattdessen bietet es sich an, die Reihenfolge der Behandlungen so zu wählen, dass eine Behandlung gleich oft auf jede andere Behandlung folgt. Wenn dann eine Behandlung eine besonders starke Nachwirkung hat, so wirkt sich dies in gleicher Weise auf alle Behandlungen aus. Formal haben wir die

Definition 2.6. Gegeben sei ein Crossover-Plan $d \in \Lambda_{t,n,p}$. Gegeben sei weiterhin ein Modell, in dem in der ersten Periode keine Nachwirkungen auftreten. Dann heißt die Behandlung $d(i, j)$ *Vorgänger* von $d(i, j + 1)$, falls $j \in \{1, \dots, p - 1\}$. Die Behandlung

$d(i, j)$ heißt Nachfolger von $d(i, j - 1)$, falls $j \in \{2, \dots, p\}$. Der Plan d heißt nachbarbalanciert, falls jede Behandlung in d genau $n(p - 1)/(t(t - 1))$ Mal auf jede andere Behandlung folgt und nie auf sich selbst.

In einem Modell, in dem in der ersten Periode die Nachwirkungen der Behandlungen der letzten Periode auftreten, spricht man von zyklischen Nachwirkungen. In diesem Fall ist die Behandlung $d(i, j)$ stets ein Vorgänger, und zwar von $d(i, j + 1)$, falls $j \in \{1, \dots, p - 1\}$ und von $d(i, 1)$, falls $j = p$. Jede Behandlung $d(i, j)$ ist Nachfolger, und zwar von $d(i, j - 1)$, falls $j \in \{2, \dots, p\}$ und von $d(i, p)$, falls $j = 1$. Der Plan d heißt dann zyklisch nachbarbalanciert, falls jede Behandlung in d genau $np/(t(t - 1))$ Mal auf jede andere Behandlung folgt und nie auf sich selbst.

Kunert und Utzig (1993) betrachten eine Untermenge der nachbarbalancierten verallgemeinerten Youden-Designs.

Definition 2.7. Sei $d \in \Lambda_{t,n,p}$. Falls es eine natürliche Zahl m gibt, so dass $n = mt$, dann heißt d uniform auf den Perioden. Sei d ein nachbarbalanciertes GYD, das uniform auf den Perioden ist. Sei weiter $p \leq t$ und d erfülle zusätzlich die folgenden Eigenschaften

VB1. Die Anzahl der Einheiten, bei denen die Behandlungen i und j gemeinsam in den ersten $p - 1$ Perioden vorkommen, ist für alle $i \neq j$ gleich.

VB2. Die Anzahl der Einheiten, bei denen Behandlung i in den ersten $p - 1$ Perioden vorkommt und Behandlung j in der letzten Periode vorkommt, ist für alle $i \neq j$ gleich.

Dann heißt d vollständig balanciert. Die Menge aller vollständig balancierter Crossover-Pläne bezeichnen wir als $\Lambda_{t,n,p}^*$.

Es folgt aus Kushner (1997), dass vollständig balancierte Pläne in Modell (2.7) universell optimal sind im Sinne von Kiefer (1975) unter allen Plänen, bei denen jede Behandlung bei jeder Einheit höchstens einmal vorkommt.

Man beachte, dass nicht für jedes (t, n, p) ein GYD oder gar ein vollständig balancierter Plan existiert. Beispiele für vollständig balancierte Versuchspläne sind nachbarbalancierte lateinische Quadrate, Williams-Designs (Williams, 1949) oder die von Patterson (1952) vorgeschlagenen Versuchspläne. Wakeling und MacFie (1995) erläutern

die Konstruktion solcher Pläne. Eine Übersicht über die Konstruktion und Optimalitätseigenschaften verschiedener Crossover-Pläne findet sich beispielsweise in Bate und Jones (2006, 2008). Mit Hilfe des Zusatzpakets *crossdes* (Sailer, 2005) lassen sich nachbarbalancierte Versuchspläne im Softwarepaket R (R Development Core Team, 2009) konstruieren.

3 Auswertung mit Hilfe der Randomisationstheorie

Das in Kapitel 2 vorgestellte Modell mit Carryover-Effekten und korrelierten Fehlern (2.5) erscheint plausibel für das dort vorgestellte Experiment. Es stellt sich aber die Frage, ob nicht noch weitere, möglicherweise unbekannte, systematische Effekte in einem konkreten Experiment auftreten, die durch den zufälligen Fehler ϵ nicht adäquat wiedergegeben werden, oder ob die im Modell berücksichtigten Effekte nicht anders wirken. Die Stärke des Carryover-Effekts einer Behandlung könnte zum Beispiel davon abhängen, auf welche Behandlung er trifft. Senn (2002)[Kapitel 10] stellt das Modell (2.5) zumindest für klinische Studien grundsätzlich in Frage, da es gängigen pharmakokinetischen Theorien widerspreche. Er schlägt vor, Crossover-Experimente von vorne herein so zu planen, dass keine Nachwirkungen auftreten. Dies erreicht man durch ausreichend lange *Wash-out* Perioden zwischen den Behandlungsperioden.

Selbst wenn wir den Carryover-Effekt bei der Auswertung nicht berücksichtigen, bleibt die Frage, ob ein plausibel erscheinendes Modell die Daten gut erklärt. Es wäre wünschenswert, ein Verfahren anzuwenden, welches eine Auswertung in einem ganz bestimmten Modell rechtfertigen kann. Eine Möglichkeit hierzu bietet die Randomisationstheorie. Die hier vorgestellten Ergebnisse beruhen im Wesentlichen auf den Arbeiten von Kunert und Sailer (2006, 2007).

3.1 Rechtfertigung eines Modells

3.1.1 Randomisationsmodelle

Bailey (1981) und Bailey und Rowley (1987) beschreiben die Vorgehensweise bei der Auswertung eines geplanten Experiments im Sinne der Randomisationstheorie. Die Randomisationstheorie versucht möglichst wenig Annahmen für die Auswertung eines

Versuchsplans vorzugeben. Das hier benutzte Randomisationsmodell unterstellt, dass die Beobachtung y_{ij} nur von zwei Komponenten abhängt. Einerseits dem unabhängig von den Behandlungen an Einheit i in Periode j erzielten Wert η_{ij} und dem *additiven* Behandlungseffekt τ_r , falls Einheit i in Periode j die Behandlung r erhält. Dieses Randomisationsmodell

$$y_{ij} = \eta_{ij} + \tau_r \quad (3.1)$$

schließt Carryover-Effekte aus.

Ein Spezialfall von (3.1) ist das Zeilen-Spalten-Modell mit

$$\eta_{ij} = \alpha_i + \pi_j + \epsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, p.$$

Hier sind die α_i die Einheitenwirkungen, die π_j die Periodeneffekte und die ϵ_{ij} u.i.v. Zufallsvariablen mit Erwartungswert 0 und Varianz σ^2 . In Vektorschreibweise erhält man in der Notation von (2.6) für das Zeilen-Spalten-Modell

$$\begin{aligned} \mathbf{y} &= \mathbf{T}_\delta \boldsymbol{\tau} + \boldsymbol{\eta} \\ &= \mathbf{T}_\delta \boldsymbol{\tau} + \mathbf{U} \boldsymbol{\alpha} + \mathbf{P} \boldsymbol{\pi} + \boldsymbol{\epsilon}. \end{aligned} \quad (3.2)$$

Die Behandlungsdesignmatrix \mathbf{T}_δ hängt vom Versuchsplan δ ab, wobei δ mit Hilfe einer Randomisierung zufällig aus der Menge aller zur Auswahl stehender Versuchspläne ausgewählt wird. Ein weiterer Spezialfall von (3.1) ist das einfache Blockmodell mit $\boldsymbol{\eta} = \mathbf{U} \boldsymbol{\alpha} + \boldsymbol{\epsilon}$.

Nachdem das Randomisationsmodell, die Menge der für die Anwendung in Frage kommenden Versuchspläne und das Randomisationsverfahren gewählt wurden, betrachtet man nun die Auswertung des Experiments. Hierbei interessieren wir uns für Behandlungskontraste $\psi = \boldsymbol{\ell}^T \boldsymbol{\tau}$. Dazu berechnen wir einen Schätzer $\hat{\psi}$ für ψ in dem Modell, das wir durch die Randomisierung rechtfertigen möchten. Weiterhin sei $\widehat{\text{var}}(\hat{\psi})$ ein Schätzer für die Varianz des Kontrastschätzers $\text{var}(\hat{\psi})$.

Definition 3.1. *Ein Modell heißt gerechtfertigt, falls für jedes feste $\boldsymbol{\eta} = (\eta_{11}, \dots, \eta_{np})^T$ in (3.1) gilt, dass für alle Behandlungskontraste $\psi = \boldsymbol{\ell}^T \boldsymbol{\tau}$*

$$E_{\delta}(\hat{\psi}) = \psi$$

und

$$E_{\delta}(\widehat{\text{var}}(\hat{\psi})) = \text{var}_{\delta}(\hat{\psi}).$$

Der Index δ gibt an, dass die Erwartungswerte bezüglich der Verteilung der zufälligen Auswahl des Versuchsplans δ gebildet werden, während $\boldsymbol{\eta}$ festgehalten wird. Man betrachtet also die bedingten Erwartungswerte y_{ij} gegeben η_{ij} . Die einzige zufällige Komponente bei der Berechnung der Erwartungswerte ist die Randomisierung des Versuchsplans.

3.1.2 Randomisationsverfahren

Ein einfaches Randomisationsverfahren für Crossover-Pläne besteht darin, die Reihenfolge der Behandlungen für jede Einheit eines Ausgangsplans d_0 separat zu randomisieren. Das bedeutet, dass die Behandlungsreihenfolge für Einheit i im randomisierten Versuchsplan δ aus einer Gleichverteilung über die Menge der Permutationen der Behandlungsreihenfolge für diese Einheit in d_0 gezogen wird. Allerdings lässt sich dieser Ansatz nicht mit nachbarbalancierten Versuchsplänen umsetzen. Die uneingeschränkte Randomisierung der Behandlungsreihenfolgen zerstört die Balance der Nachwirkungen.

Azaïs (1987) untersucht ein Modell mit zyklischen Nachwirkungen. In diesem Modell nimmt man an, dass auch in der ersten Periode Nachwirkungen auftreten und zwar die der Behandlungen der letzten Periode. Bei der Randomisierung werden nun die Behandlungen nicht beliebig, sondern zyklisch vertauscht. Dies kann man sich bildlich vorstellen, indem man die Behandlungen einer Einheit in einem Kreis aufschreibt und diesen Kreis dann an einer zufällig ausgewählten Position aufschneidet und so eine neue Folge von Behandlungen erhält. Wird die Reihenfolge der Behandlungen einer Einheit in einem zyklisch nachbarbalancierten Plan zyklisch vertauscht, dann bleibt die Balance offensichtlich erhalten. Es ist klar, dass auch die Randomisierung von Behandlungsnamen, das heißt die zufällige Zuordnung aus einer Gleichverteilung über die Zahlen $\{1, \dots, t\}$, und die Randomisierung der Einheiten aus einer Gleichverteilung über $\{1, \dots, n\}$ die Balance bezüglich Carryover-Effekten nicht zerstören. Azaïs (1987) zeigt, dass die Randomisierung von Einheiten, Behandlungsnamen und die zyklische Randomisierung der Behandlungsreihenfolgen zusammen das einfache Blockmodell rechtfertigen, sofern keine Nachwirkungseffekte vorliegen. Falls Carryover-Effekte vor-

liegen, so zeigen Azaïs und Druilhet (1997), dass dieses Verfahren die Verzerrung durch die Nachwirkungseffekte bezüglich gewisser Kriterien minimiert.

Zyklische Nachwirkungen sind zum Beispiel dann zu erwarten, wenn vor der ersten Periode auf jede Einheit bereits die Behandlung angewandt wird, die diese Einheit in der letzten Periode erhält. Die Beobachtungen dieser Vorperiode werden für die Auswertung nicht genutzt (Azaïs, 1987). In klinischen Studien ist eine solche Vorperiode kaum praktikabel. Der zusätzliche organisatorische Aufwand und die zusätzliche Belastung für die Probanden würden in Kauf genommen, damit etwaige Nachwirkungen zyklisch sind, mit dem Ziel ein Modell zu rechtfertigen, welches nur dann gilt, falls eben diese Nachwirkungen nicht vorkommen. Tatsächlich wurde das Verfahren in Azaïs (1987) für agrarwissenschaftliche Feldversuche vorgeschlagen. Bei derartigen Studien werden die Ränder der Versuchsfelder oft bepflanzt, ohne dass diese Parzellen, die *border plots*, mit in die Auswertung eingehen. In der Lebensmittelsensorik bietet es sich durchaus an, die Prüfer vor Beginn der eigentlichen Verkostung mit einer Probe auf die verschiedenen Produkte einzustimmen. Allerdings wird man hier eher ein neutrales Produkt verwenden, so dass durch dieses Vorgehen keine zyklischen Nachwirkungen generiert werden.

Bailey (1985) betrachtet eine Randomisierung, die nicht von zyklischen Nachwirkungen ausgeht. Sie zeigt, dass die Randomisierung von Behandlungsnamen und Einheiten alleine bereits das Zeilen-Spalten-Modell (3.2) rechtfertigt, sofern der Ausgangsplan ein nachbarbalanciertes verallgemeinertes lateinisches Quadrat ist. Beispiele für solche Versuchspläne sind beispielsweise Williams-Designs (Williams, 1949) oder Pläne, bei denen jede Zeile eines Williams-Designs gleich häufig wiederholt wird. Grundsätzlich gilt für alle verallgemeinerten lateinischen Quadrate, dass $p \geq t$. Bei Crossover-Plänen mit vielen Behandlungen kann dies eine Einschränkung für den praktischen Nutzen des Ergebnisses darstellen.

3.2 Auswertung im Zeilen-Spalten-Modell

3.2.1 Modelleigenschaften bei Abwesenheit von Nachwirkungen

Kunert und Sailer (2007) erweitern das Ergebnis von Bailey (1985) auf eine größere Klasse von Versuchsplänen. Zu Details der in diesem Kapitel aufgeführten Beweise

sei auf Kunert und Sailer (2007) verwiesen. Zunächst leiten wir einige orthogonale Projektoren her, auf die wir im Folgenden zurückgreifen.

Lemma 3.1. *Sei $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2]$ eine partitionierte Matrix. Dann gilt*

$$\boldsymbol{\omega}_{\mathbf{A}} = \boldsymbol{\omega}_{\mathbf{A}_2} + \boldsymbol{\omega}_{\boldsymbol{\omega}_{\mathbf{A}_2}^\perp \mathbf{A}_1}$$

und

$$\boldsymbol{\omega}_{\mathbf{A}}^\perp = \boldsymbol{\omega}_{\mathbf{A}_2}^\perp - \boldsymbol{\omega}_{\boldsymbol{\omega}_{\mathbf{A}_2}^\perp \mathbf{A}_1}.$$

Beweis. Mit Hilfe von Lemma 2.2 rechnet man leicht nach, dass

$$\mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T = \boldsymbol{\omega}_{\mathbf{A}_2} + \boldsymbol{\omega}_{\boldsymbol{\omega}_{\mathbf{A}_2}^\perp \mathbf{A}_1},$$

vergleiche auch Seber (2008)[2.50(b)]. □

Lemma 3.2. *Sei $\boldsymbol{\omega}_{\mathbf{1}_r}^\perp = \mathbf{Q}_r$ die zentrierende Matrix und seien \mathbf{U} und \mathbf{P} wie in (2.6). Dann gilt*

$$\boldsymbol{\omega}_{\mathbf{1}_r}^\perp = \mathbf{Q}_r = \mathbf{I}_r - \frac{1}{r} \mathbf{1}_r \mathbf{1}_r^T,$$

$$\boldsymbol{\omega}_{\mathbf{U}}^\perp = \mathbf{I}_n \otimes \mathbf{Q}_p$$

und

$$\boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}]}^\perp = \mathbf{Q}_n \otimes \mathbf{Q}_p.$$

Beweis.

$$\begin{aligned} \boldsymbol{\omega}_{\mathbf{1}_r}^\perp &= \mathbf{I}_r - \mathbf{1}_r (\mathbf{1}_r^T \mathbf{1}_r)^{-1} \mathbf{1}_r^T \\ &= \mathbf{I}_r - \frac{1}{r} \mathbf{1}_r \mathbf{1}_r^T. \end{aligned}$$

Weiterhin ist mit $\mathbf{U} = \mathbf{I}_n \otimes \mathbf{1}_p$ und $\mathbf{P} = \mathbf{1}_n \otimes \mathbf{I}_p$

$$\begin{aligned} \boldsymbol{\omega}_{\mathbf{U}}^\perp &= \mathbf{I}_{np} - (\mathbf{I}_n \otimes \mathbf{1}_p) (\mathbf{I}_n \otimes p)^{-1} (\mathbf{I}_n \otimes \mathbf{1}_p^T) \\ &= \mathbf{I}_n \otimes \mathbf{Q}_p \end{aligned}$$

und daher $\boldsymbol{\omega}_{\mathbf{U}}^\perp \mathbf{P} = \mathbf{1}_n \otimes \mathbf{Q}_p$. Mit Hilfe von Lemma 3.1 folgt sofort, dass

$$\begin{aligned} \boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}]}^\perp &= \boldsymbol{\omega}_{\mathbf{U}}^\perp - \boldsymbol{\omega}_{\boldsymbol{\omega}_{\mathbf{U}}^\perp \mathbf{P}} \\ &= \mathbf{Q}_n \otimes \mathbf{Q}_p. \end{aligned} \quad \square$$

Sei \mathbf{C}_δ^- eine verallgemeinerte Inverse von $\mathbf{C}_\delta = \mathbf{T}_\delta^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{T}_\delta$,

$$\hat{\boldsymbol{\tau}} = \mathbf{C}_\delta^- \mathbf{T}_\delta^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{y}$$

und $\boldsymbol{\psi} = \boldsymbol{\ell}^T \boldsymbol{\tau}$ ein beliebiger Kontrast. Nach Satz 2.1 und Lemma 3.2 ist der KQ-Schätzer von $\boldsymbol{\psi}$ im Zeilen-Spalten-Modell (3.2) dann gegeben durch $\hat{\boldsymbol{\psi}} = \boldsymbol{\ell}^T \hat{\boldsymbol{\tau}}$.

Lemma 3.3. (Kunert und Sailer, 2007). Sei $p \leq t$. Der Ausgangsplan $d_0 \in \Lambda_{t,n,p}$ sei ein reguläres GYD. Man randomisiere Behandlungsnamen und Einheiten. Dann ist der Schätzer $\hat{\boldsymbol{\psi}} = \boldsymbol{\ell}^T \mathbf{C}_\delta^- \mathbf{T}_\delta^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{y}$ aus dem Zeilen-Spalten-Modell (3.2) im Randomisationsmodell (3.1) unverzerrt.

Beweis. Wenn der Ausgangsplan d_0 die Behandlungsdesignmatrix \mathbf{H}_d besitzt und Behandlungsnamen und Einheiten randomisiert werden, dann gibt es Permutationsmatrizen $\mathbf{\Pi} \in \mathbb{R}^{n \times n}$ und $\mathbf{\Gamma} \in \mathbb{R}^{t \times t}$, so dass $\mathbf{T}_\delta = (\mathbf{\Pi} \otimes \mathbf{I}_p) \mathbf{H}_d \mathbf{\Gamma}$ die Behandlungsdesignmatrix des randomisierten Plans ist. Mit der Notation $\mathbf{D}_d = \mathbf{H}_d^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{H}_d$ ist dann $\mathbf{C}_\delta = \mathbf{\Gamma}^T \mathbf{D}_d \mathbf{\Gamma}$. Für reguläre GYD existiert ein $c \in \mathbb{R}$, so dass $\mathbf{D}_d = c \mathbf{Q}_t$ und somit auch $\mathbf{C}_\delta = c \mathbf{Q}_t$. Daraus folgt

$$\begin{aligned} \hat{\boldsymbol{\tau}} &= \mathbf{C}_\delta^- \mathbf{T}_\delta^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) (\mathbf{T}_\delta \boldsymbol{\tau} + \boldsymbol{\eta}) \\ &= \mathbf{Q}_t \boldsymbol{\tau} + \frac{1}{c} \mathbf{Q}_t \mathbf{\Gamma}^T \mathbf{H}_d^T (\mathbf{Q}_n \mathbf{\Pi}^T \otimes \mathbf{Q}_p) \boldsymbol{\eta}. \end{aligned} \quad (3.3)$$

Der Erwartungswert des Kontrastschätzers berechnet sich als arithmetisches Mittel über alle möglichen Permutationsmatrizen $\mathbf{\Pi}$ und $\mathbf{\Gamma}$. Wegen $\boldsymbol{\ell}^T \mathbf{1}_t = 0$ ist $\boldsymbol{\ell}^T \mathbf{Q}_t = \boldsymbol{\ell}^T$. Weiterhin ist $\frac{1}{t!} \sum_{\mathbf{\Gamma}} \mathbf{\Gamma}^T = \frac{1}{t} \mathbf{1}_t \mathbf{1}_t^T$. Mit

$$\mathbb{E}_\delta (\boldsymbol{\ell}^T \hat{\boldsymbol{\tau}}) = \boldsymbol{\ell}^T \boldsymbol{\tau} + \frac{1}{t!n!c} \boldsymbol{\ell}^T \sum_{\mathbf{\Gamma}} \sum_{\mathbf{\Pi}} \mathbf{\Gamma}^T \mathbf{H}_d^T (\mathbf{\Pi}^T \mathbf{Q}_n \otimes \mathbf{Q}_p) \boldsymbol{\eta}$$

folgt die Behauptung, da der zweite Summand in obiger Gleichung verschwindet. \square

Lemma 3.4. (Kunert und Sailer, 2007). Sei $p \leq t$. Der Ausgangsplan $d_0 \in \Lambda_{t,n,p}$ sei ein reguläres GYD. Man randomisiere Behandlungsnamen und Einheiten. Dann gilt

$$\text{var}_\delta (\hat{\boldsymbol{\psi}}) = \frac{1}{c(n-1)(p-1)} \boldsymbol{\ell}^T \boldsymbol{\ell} \boldsymbol{\eta}^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \boldsymbol{\eta}.$$

Beweis. Es ist

$$\boldsymbol{\ell}^T \hat{\boldsymbol{\tau}} = \boldsymbol{\ell}^T \boldsymbol{\tau} + \frac{1}{c} \boldsymbol{\ell}^T \boldsymbol{\Gamma}^T \mathbf{H}_d^T (\mathbf{Q}_n \boldsymbol{\Pi}^T \otimes \mathbf{Q}_p) \boldsymbol{\eta},$$

vergleiche (3.3). Da hier der erste Summand konstant ist und der zweite Summand Erwartungswert 0 hat, ist

$$\begin{aligned} \text{var}_\delta \left(\hat{\boldsymbol{\psi}} \right) &= \mathbb{E}_\delta \left(\left[\frac{1}{c} \boldsymbol{\ell}^T \boldsymbol{\Gamma}^T \mathbf{H}_d^T (\mathbf{Q}_n \boldsymbol{\Pi}^T \otimes \mathbf{Q}_p) \boldsymbol{\eta} \right]^2 \right) \\ &= \frac{1}{c^2 t! n!} \sum_{\boldsymbol{\Gamma}} \sum_{\boldsymbol{\Pi}} \boldsymbol{\ell}^T \boldsymbol{\Gamma}^T \mathbf{H}_d^T (\mathbf{Q}_n \boldsymbol{\Pi}^T \otimes \mathbf{Q}_p) \boldsymbol{\eta} \boldsymbol{\eta}^T (\boldsymbol{\Pi} \mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{H}_d \boldsymbol{\Gamma} \boldsymbol{\ell} \\ &= \frac{1}{c^2 t!} \sum_{\boldsymbol{\Gamma}} \boldsymbol{\ell}^T \boldsymbol{\Gamma}^T \mathbf{H}_d^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{M}_1 (\mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{H}_d \boldsymbol{\Gamma} \boldsymbol{\ell}. \end{aligned}$$

Hier ist

$$\mathbf{M}_1 = \frac{1}{n!} \sum_{\boldsymbol{\Pi}} (\boldsymbol{\Pi}^T \otimes \mathbf{I}_p) \boldsymbol{\eta} \boldsymbol{\eta}^T (\boldsymbol{\Pi} \otimes \mathbf{I}_p).$$

Nun gibt es von $\boldsymbol{\eta}$ abhängige Matrizen \mathbf{M} und $\tilde{\mathbf{M}} \in \mathbb{R}^{p \times p}$, so dass

$$\mathbf{M}_1 = \mathbf{I}_n \otimes (\mathbf{M} - \tilde{\mathbf{M}}) + \mathbf{1}_n \mathbf{1}_n^T \otimes \tilde{\mathbf{M}}.$$

Schreibt man nun $\mathbf{Z}_\eta = \mathbf{Q}_n \otimes (\mathbf{Q}_p (\mathbf{M} - \tilde{\mathbf{M}}) \mathbf{Q}_p)$, so ergibt sich für die Varianz des geschätzten Behandlungskontrasts

$$\text{var}_\delta \left(\hat{\boldsymbol{\psi}} \right) = \frac{1}{c^2 t!} \sum_{\boldsymbol{\Gamma}} \boldsymbol{\ell}^T \boldsymbol{\Gamma}^T \mathbf{H}_d^T \mathbf{Z}_\eta \mathbf{H}_d \boldsymbol{\Gamma} \boldsymbol{\ell}.$$

Bezeichne $\text{tr}(\mathbf{A})$ die Spur einer quadratischen Matrix \mathbf{A} . Für eine beliebige quadratische Matrix \mathbf{A} mit Zeilen- und Spaltensummen 0 gilt

$$\frac{1}{t!} \sum_{\boldsymbol{\Gamma}} \boldsymbol{\Gamma}^T \mathbf{A} \boldsymbol{\Gamma} = \frac{\text{tr}(\mathbf{A})}{t-1} \mathbf{Q}_t.$$

Man rechnet leicht nach, dass dann

$$\text{var}_\delta \left(\hat{\boldsymbol{\psi}} \right) = \frac{1}{c^2} \frac{\text{tr}(\mathbf{H}_d^T \mathbf{Z}_\eta \mathbf{H}_d)}{t-1} \boldsymbol{\ell}^T \boldsymbol{\ell}. \quad (3.4)$$

Indem wir die Behandlungsdesignmatrix von d_0 durch $\mathbf{H}_d = [\mathbf{H}_{d_1}^T, \dots, \mathbf{H}_{d_n}^T]^T$ ausdrücken, können wir $\mathbf{H}_d^T \mathbf{Z}_\eta \mathbf{H}_d$ darstellen als

$$\sum_i \mathbf{H}_{d_i}^T \mathbf{Q}_p (M - \tilde{M}) \mathbf{Q}_p \mathbf{H}_{d_i} - \frac{1}{n} \left(\sum_i \mathbf{H}_{d_i}^T \right) \mathbf{Q}_p (M - \tilde{M}) \mathbf{Q}_p \left(\sum_i \mathbf{H}_{d_i} \right).$$

Da $p \leq t$, gibt es Permutationsmatrizen $\Delta_1, \dots, \Delta_n \in \mathbb{R}^{t \times t}$, so dass $\mathbf{H}_{d_i} = [\mathbf{I}_p, \mathbf{0}] \Delta_i$ und $\mathbf{H}_{d_i} \mathbf{H}_{d_i}^T = \mathbf{I}_p$. Aufgrund der Balance bezüglich der Perioden für reguläre GYD existieren stets $c_1, c_2 \in \mathbb{R}$, so dass

$$\left(\sum_i \mathbf{H}_{d_i} \right) \left(\sum_i \mathbf{H}_{d_i} \right)^T = c_1 \mathbf{I}_t + c_2 \mathbf{1}_t \mathbf{1}_t^T.$$

Damit können wir die Spur von $\mathbf{H}_d^T \mathbf{Z}_\eta \mathbf{H}_d$ berechnen. Wir nutzen dabei aus, dass die Spur eines Matrixproduktes bei zyklischer Vertauschung der Faktoren unverändert bleibt (Seber, 2008)[4.14]. Es gilt

$$\begin{aligned} \text{tr}(\mathbf{H}_d^T \mathbf{Z}_\eta \mathbf{H}_d) &= \sum_i \text{tr} \left(\mathbf{H}_{d_i} \mathbf{H}_{d_i}^T \mathbf{Q}_p (M - \tilde{M}) \mathbf{Q}_p \right) \\ &\quad - \frac{1}{n} \text{tr} \left(\left(\sum_i \mathbf{H}_{d_i} \right) \left(\sum_i \mathbf{H}_{d_i} \right)^T \mathbf{Q}_p (M - \tilde{M}) \mathbf{Q}_p \right) \\ &= \sum_i \text{tr} \left(\mathbf{Q}_p (M - \tilde{M}) \mathbf{Q}_p \right) - \frac{1}{n} \text{tr} \left(c_1 \mathbf{Q}_p (M - \tilde{M}) \mathbf{Q}_p \right) \\ &= \left(n - \frac{c_1}{n} \right) \frac{1}{n-1} \text{tr}(\mathbf{Z}_\eta). \end{aligned}$$

Man rechnet nach, dass

$$\begin{aligned} \text{tr}(\mathbf{Z}_\eta) &= \text{tr} \left(\frac{1}{n!} \sum_{\Pi} (\Pi^T \mathbf{Q}_n \otimes \mathbf{Q}_p) \eta \eta^T (\mathbf{Q}_n \Pi \otimes \mathbf{Q}_p) \right) \\ &= \eta^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \eta. \end{aligned}$$

Kunert und Sailer (2007) bemerken, dass für reguläre GYD

$$c = \left(n - \frac{c_1}{n} \right) \frac{p-1}{t-1}.$$

Durch einsetzen in (3.4) folgt die Behauptung. □

Lemma 3.5. (Kunert und Sailer, 2007). Sei $p \leq t$. Der Ausgangsplan $d_0 \in \Lambda_{t,n,p}$ sei ein reguläres GYD. Man randomisiere Behandlungsnamen und Einheiten. Dann gilt

$$\mathbb{E}_\delta \left(\widehat{\text{var}} \left(\hat{\psi} \right) \right) = \frac{1}{c(n-1)(p-1)} \boldsymbol{\ell}^T \boldsymbol{\ell} \boldsymbol{\eta}^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \boldsymbol{\eta}.$$

Beweis. Es ist $\text{var}(\hat{\psi}) = \sigma^2 \boldsymbol{\ell}^T \mathbf{C}_\delta^- \boldsymbol{\ell}$. Ein Schätzer für die Varianz ist gegeben durch

$$\widehat{\text{var}} \left(\hat{\psi} \right) = \hat{\sigma}^2 \boldsymbol{\ell}^T \mathbf{C}_\delta^- \boldsymbol{\ell}, \quad (3.5)$$

wobei

$$\hat{\sigma}^2 = \frac{1}{f_E} \left(\mathbf{y}^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{y} - \hat{\boldsymbol{\tau}}^T \mathbf{C}_\delta \hat{\boldsymbol{\tau}} \right),$$

vergleiche Lemma (2.1). Im Zeilen-Spalten-Modell haben wir $f_E = (n-1)(p-1) - t + 1$ Freiheitsgrade zum Schätzen der Varianz.

Man beachte, dass $\mathbf{C}_\delta = c \mathbf{Q}_t$ und $\mathbf{Q}_t \mathbf{T}_\delta^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) = \mathbf{T}_\delta^T (\mathbf{Q}_n \otimes \mathbf{Q}_p)$. Damit zeigt man, dass

$$\begin{aligned} \mathbf{y}^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{y} - \hat{\boldsymbol{\tau}}^T \mathbf{C}_\delta \hat{\boldsymbol{\tau}} &= \mathbf{y}^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{y} - \frac{1}{c} \mathbf{y}^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{T}_\delta \mathbf{T}_\delta^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{y} \\ &= \boldsymbol{\eta}^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \boldsymbol{\eta} - \frac{1}{c} \boldsymbol{\eta}^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{T}_\delta \mathbf{T}_\delta^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \boldsymbol{\eta} \\ &= \text{tr}(\mathbf{Z}_\eta) \\ &\quad - \frac{1}{c} \text{tr}(\mathbf{H}_d^T (\mathbf{Q}_n \boldsymbol{\Pi}^T \otimes \mathbf{Q}_p) \boldsymbol{\eta} \boldsymbol{\eta}^T (\boldsymbol{\Pi} \mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{H}_d). \end{aligned}$$

Nun ist aber $\mathbf{Z}_\eta = \mathbb{E}_\delta((\mathbf{Q}_n \boldsymbol{\Pi}^T \otimes \mathbf{Q}_p) \boldsymbol{\eta} \boldsymbol{\eta}^T (\boldsymbol{\Pi} \mathbf{Q}_n \otimes \mathbf{Q}_p))$. Durch einsetzen von $\text{tr}(\mathbf{Z}_\eta)$ und $\text{tr}(\mathbf{H}_d^T \mathbf{Z}_\eta \mathbf{H}_d)$ folgt dann

$$\begin{aligned} \mathbb{E}_\delta(\hat{\sigma}^2) &= \frac{1}{(n-1)(p-1) - t + 1} \left(\text{tr}(\mathbf{Z}_\eta) - \frac{1}{c} \text{tr}(\mathbf{H}_d^T \mathbf{Z}_\eta \mathbf{H}_d) \right) \\ &= \frac{1}{(n-1)(p-1)} \boldsymbol{\eta}^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \boldsymbol{\eta}. \end{aligned}$$

Mit (3.5) ist

$$\mathbb{E}_\delta \left(\widehat{\text{var}} \left(\hat{\psi} \right) \right) = \frac{1}{c} \mathbb{E}_\delta(\hat{\sigma}^2) \boldsymbol{\ell}^T \boldsymbol{\ell},$$

woraus sofort die Behauptung folgt. \square

Da in Lemma 3.3 - 3.5 jeweils ℓ beliebig war, haben wir bereits den folgenden Satz für reguläre GYD bewiesen.

Satz 3.1. (Kunert und Sailer, 2007). Sei $p \leq t$. Der Ausgangsplan $d_0 \in \Lambda_{t,n,p}$ sei ein reguläres GYD. Dann rechtfertigt die Randomisierung von Behandlungsnamen und Einheiten das Zeilen-Spalten-Modell.

Man beachte, dass diese Aussage nicht davon abhängt, ob d_0 nachbarbalanciert ist oder nicht.

3.2.2 Einfluss von Carryover-Effekten

Falls doch Carryover-Effekte vorliegen, so ergibt sich aus dem Randomisationsmodell (3.1) das Modell

$$y_{ij} = \eta_{ij} + \tau_{\delta(i,j)} + \rho_{\delta(i-1,j)}. \quad (3.6)$$

Für den Fall, dass für die η_{ij} in (3.6) das Zeilen-Spalten-Modell gilt, zeigt Kunert (1984), dass nachbarbalancierte Versuchspläne sehr gute Optimalitätseigenschaften aufweisen. Kunert und Utzig (1993) weisen allerdings darauf hin, dass das Modell (3.6) nicht durch Randomisierung gerechtfertigt werden kann.

Lemma 3.6. (Kunert und Sailer, 2007). Es sei $n = mt$ für ein $m \in \mathbb{N}$ und $p \leq t$. Der Ausgangsplan $d_0 \in \Lambda_{t,n,p}$ sei ein verallgemeinertes Youden-Design, das uniform auf den Perioden ist. Es gelte Modell (3.6) und man randomisiere Behandlungsnamen und Einheiten. Dann ist der Kontrastschätzer $\hat{\psi}$ aus Kapitel 3.2.1 im Allgemeinen verzerrt und die Verzerrung beträgt $-\ell^T \boldsymbol{\rho}/p$. Für die Varianz des Kontrastschätzers gilt

$$\text{var}_{\delta}(\hat{\psi}) \geq \frac{1}{c(n-1)(p-1)} \ell^T \boldsymbol{\ell} \boldsymbol{\eta}^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \boldsymbol{\eta}.$$

Beweis. Modell (3.6) lautet in Matrixschreibweise

$$\mathbf{y} = \mathbf{T}_{\delta} \boldsymbol{\tau} + \mathbf{F}_{\delta} \boldsymbol{\rho} + \boldsymbol{\eta},$$

wobei $\mathbf{F}_{\delta} = (\mathbf{I}_n \otimes \mathbf{V}) \mathbf{T}_{\delta}$ mit

$$\mathbf{V} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{I}_{p-1} & \mathbf{0} \end{bmatrix}.$$

Wir definieren

$$\mathbf{u}_\delta = \frac{1}{c} (\mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{T}_\delta \boldsymbol{\ell}$$

und erhalten unter (3.6)

$$\begin{aligned} \hat{\psi} &= \frac{1}{c} \boldsymbol{\ell}^T \mathbf{T}_\delta^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) (\mathbf{T}_\delta \boldsymbol{\tau} + \mathbf{F}_\delta \boldsymbol{\rho} + \boldsymbol{\eta}) \\ &= \mathbf{u}_\delta^T (\mathbf{T}_\delta \boldsymbol{\tau} + \boldsymbol{\eta}) + \mathbf{u}_\delta^T \mathbf{F}_\delta \boldsymbol{\rho}. \end{aligned}$$

Mit Lemma 3.3 und Lemma 3.4 folgt sofort, dass $\mathbf{u}_\delta^T (\mathbf{T}_\delta \boldsymbol{\tau} + \boldsymbol{\eta}) = \boldsymbol{\ell}^T \boldsymbol{\tau} + \mathbf{u}_\delta^T \boldsymbol{\rho}$, $E_\delta(\mathbf{u}_\delta^T \boldsymbol{\rho}) = 0$ und $\text{var}_\delta(\mathbf{T}_\delta \boldsymbol{\tau} + \boldsymbol{\eta}) = v_m$, wobei

$$v_m = \frac{1}{c(n-1)(p-1)} \boldsymbol{\ell}^T \boldsymbol{\ell} \boldsymbol{\eta}^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \boldsymbol{\eta}.$$

Weiterhin ist

$$\mathbf{u}_\delta^T \mathbf{F}_\delta \boldsymbol{\rho} = \frac{1}{c} \boldsymbol{\ell}^T \boldsymbol{\Gamma}^T \mathbf{H}_d^T (\mathbf{Q}_n \otimes \mathbf{Q}_p \mathbf{V}) \mathbf{H}_d \boldsymbol{\Gamma} \boldsymbol{\rho}.$$

Mit der gleichen Vorgehensweise wie in Lemma 3.4 zeigt man, dass

$$E_\delta (\boldsymbol{\Gamma}^T \mathbf{H}_d^T (\mathbf{Q}_n \otimes \mathbf{Q}_p \mathbf{V}) \mathbf{H}_d \boldsymbol{\Gamma}) = \frac{\text{tr} (\mathbf{H}_d^T (\mathbf{Q}_n \otimes \mathbf{Q}_p \mathbf{V}) \mathbf{H}_d)}{t-1} \mathbf{Q}_t,$$

wobei $\text{tr} (\mathbf{H}_d^T (\mathbf{Q}_n \otimes \mathbf{Q}_p \mathbf{V}) \mathbf{H}_d)$ sich darstellen lässt als

$$\sum_i \text{tr} (\mathbf{H}_{di} \mathbf{H}_{di}^T \mathbf{Q}_p \mathbf{V}) - \frac{1}{n} \text{tr} \left(\left(\sum_i \mathbf{H}_{di} \right) \left(\sum_i \mathbf{H}_{di} \right)^T \mathbf{Q}_p \mathbf{V} \right).$$

Schreibt man wieder $\mathbf{H}_{di} = [\mathbf{I}_p, \mathbf{0}] \boldsymbol{\Delta}_i$, so sieht man, dass

$$\begin{aligned} \text{tr} (\mathbf{H}_{di} \mathbf{H}_{di}^T \mathbf{Q}_p \mathbf{V}) &= \text{tr} (\mathbf{Q}_p \mathbf{V}) \\ &= -\frac{p-1}{p}. \end{aligned}$$

Da d_0 uniform auf den Perioden ist, ist außerdem $\sum_i \mathbf{H}_{di} = n \mathbf{1}_p \mathbf{1}_p^T / t$. Dies impliziert, dass

$$E_\delta (\mathbf{u}_\delta^T \mathbf{F}_\delta \boldsymbol{\rho}) = -\frac{1}{c} \frac{n}{t-1} \frac{p-1}{p} \boldsymbol{\ell}^T \boldsymbol{\rho}.$$

Schließlich ist für d_0 $c = n(p-1)/(t-1)$. Daher gilt für den Erwartungswert des Kontrastschätzers

$$\mathbb{E}_\delta(\hat{\psi}) = \boldsymbol{\ell}^T \boldsymbol{\tau} - \frac{1}{p} \boldsymbol{\ell}^T \boldsymbol{\rho}.$$

Man rechnet nach, dass

$$\begin{aligned} \mathbb{E}_\delta(\mathbf{u}_\delta^T \mathbf{F}_\delta \boldsymbol{\rho} \mathbf{u}_\delta^T (\mathbf{T}_\delta \boldsymbol{\tau} + \boldsymbol{\eta})) &= \frac{1}{n!t!} \sum_{\mathbf{\Pi}} \sum_{\mathbf{\Gamma}} \left(\left(\frac{1}{c} \boldsymbol{\ell}^T \boldsymbol{\Gamma}^T H_d^T (\mathbf{Q}_n \otimes \mathbf{Q}_p \mathbf{V}) H_d \boldsymbol{\Gamma} \boldsymbol{\rho} \right) \right. \\ &\quad \left. \cdot \left(\boldsymbol{\ell}^T \boldsymbol{\tau} + \frac{1}{c} \boldsymbol{\ell}^T \boldsymbol{\Gamma}^T H_d^T (\mathbf{Q}_n \boldsymbol{\Pi}^T \otimes \mathbf{Q}_p) \boldsymbol{\eta} \right) \right) \\ &= \frac{1}{t!} \sum_{\mathbf{\Gamma}} \left(\frac{1}{c} \boldsymbol{\ell}^T \boldsymbol{\Gamma}^T H_d^T (\mathbf{Q}_n \otimes \mathbf{Q}_p \mathbf{V}) H_d \boldsymbol{\Gamma} \boldsymbol{\rho} \boldsymbol{\ell}^T \boldsymbol{\tau} \right) \\ &= \mathbb{E}_\delta(\mathbf{u}_\delta^T \mathbf{F}_\delta \boldsymbol{\rho}) \boldsymbol{\ell}^T \boldsymbol{\tau} \\ &= \mathbb{E}_\delta(\mathbf{u}_\delta^T \mathbf{F}_\delta \boldsymbol{\rho}) \mathbb{E}_\delta(\mathbf{u}_\delta^T (\mathbf{T}_\delta \boldsymbol{\tau} + \boldsymbol{\eta})). \end{aligned}$$

Demnach sind $\mathbf{u}_\delta^T \mathbf{F}_\delta \boldsymbol{\rho}$ und $\mathbf{u}_\delta^T (\mathbf{T}_\delta \boldsymbol{\tau} + \boldsymbol{\eta})$ unkorreliert und für die Varianz von $\hat{\psi}$ ergibt sich

$$\begin{aligned} \text{var}_\delta(\mathbf{u}_\delta^T (\mathbf{T}_\delta \boldsymbol{\tau} + \boldsymbol{\eta} + \mathbf{F}_\delta \boldsymbol{\rho})) &= v_m + \text{var}_\delta(\mathbf{u}_\delta^T \mathbf{F}_\delta \boldsymbol{\rho}) \\ &\geq v_m. \end{aligned} \tag{3.7} \quad \square$$

Der folgende Satz zeigt, dass es von Vorteil sein kann, nachbarbalancierte Versuchspläne einzusetzen, wenn mit Carryover-Effekten gerechnet wird.

Satz 3.2. (Kunert und Sailer, 2007). Sei d_0 unter den Bedingungen von Lemma 3.6 vollständig balanciert. Dann hängt die Varianz von $\hat{\psi}$ nicht von den Carryover-Effekten ab. Der Varianzschätzer $\widehat{\text{var}}(\hat{\psi})$ ist ein konservativer Schätzer für die Varianz von $\hat{\psi}$.

Beweis. Der erste Teil der Behauptung ist gezeigt, falls in (3.7) $\text{var}_\delta(\mathbf{u}_\delta^T \mathbf{F}_\delta \boldsymbol{\rho}) = 0$ für alle $\boldsymbol{\rho} \in \mathbb{R}^t$. Für vollständig balancierte Pläne ist

$$\mathbf{H}_d^T (\mathbf{Q}_n \otimes \mathbf{Q}_p \mathbf{V}) \mathbf{H}_d = -\frac{n(p-1)}{p(t-1)} \mathbf{Q}_t,$$

vergleiche Kunert und Utzig (1993). Damit ist $\mathbf{\Gamma}^T \mathbf{H}_d^T (\mathbf{Q}_n \mathbf{\Pi}^T \otimes \mathbf{Q}_p) (\mathbf{\Pi} \otimes \mathbf{V}) \mathbf{H}_d \mathbf{\Gamma} \boldsymbol{\rho}$ konstant in $\mathbf{\Gamma}$ und $\mathbf{\Pi}$ und daher

$$\text{var}_\delta (\mathbf{u}_\delta^T \mathbf{F}_\delta \boldsymbol{\rho}) = \text{var}_\delta \left(\frac{1}{c} \boldsymbol{\ell}^T \mathbf{\Gamma}^T \mathbf{H}_d^T (\mathbf{Q}_n \otimes \mathbf{Q}_p \mathbf{V}) \mathbf{H}_d \mathbf{\Gamma} \boldsymbol{\rho} \right) = 0.$$

Um den zweiten Teil der Behauptung zu zeigen, betrachten wir

$$\hat{\boldsymbol{\tau}} = \mathbf{C}_\delta^- \mathbf{T}_\delta^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) (\mathbf{T}_\delta \boldsymbol{\tau} + \boldsymbol{\eta}) - \frac{1}{k} \mathbf{Q}_t \boldsymbol{\rho}.$$

Nun hängt

$$f_E \hat{\sigma}^2 = \mathbf{y}^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{y} - \hat{\boldsymbol{\tau}}^T \mathbf{C}_\delta \hat{\boldsymbol{\tau}}$$

wie in Lemma 3.5 nicht von $\boldsymbol{\tau}$ ab. Damit können wir hier $\boldsymbol{\tau} = \mathbf{0}$ setzen. Mit der Notation $\tilde{\boldsymbol{\tau}} = \mathbf{C}_\delta^- \mathbf{T}_\delta^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \boldsymbol{\eta}$ folgt

$$f_E \hat{\sigma}^2 = (\boldsymbol{\eta} + \mathbf{F}_\delta \boldsymbol{\rho})^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) (\boldsymbol{\eta} + \mathbf{F}_\delta \boldsymbol{\rho}) - \left(\tilde{\boldsymbol{\tau}} - \frac{1}{p} \mathbf{Q}_t \boldsymbol{\rho} \right)^T \mathbf{D}_d \left(\tilde{\boldsymbol{\tau}} - \frac{1}{p} \mathbf{Q}_t \boldsymbol{\rho} \right).$$

Nutzt man aus, dass $E_\delta(\tilde{\boldsymbol{\tau}}) = \mathbf{0}$ und dass für vollständig balanciertes d_0

$$\mathbf{F}_\delta^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{F}_\delta = \frac{n(p-1)(pt-t-1)}{pt(t-1)} \mathbf{Q}_t,$$

vergleiche Kunert und Utzig (1993), so ergibt sich schließlich

$$\begin{aligned} f_E E_\delta (\hat{\sigma}^2) &= \boldsymbol{\eta}^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \boldsymbol{\eta} - E_\delta (\tilde{\boldsymbol{\tau}}^T \mathbf{C}_\delta \tilde{\boldsymbol{\tau}}) + \frac{n(p-1)(pt-t-1)}{pt(t-1)} \boldsymbol{\rho}^T \mathbf{Q}_t \boldsymbol{\rho} \\ &\quad - \frac{1}{k^2} \frac{n(p-1)}{t-1} \boldsymbol{\rho}^T \mathbf{Q}_t \boldsymbol{\rho}. \end{aligned}$$

Im Modell ohne Carryover-Effekte ist hingegen

$$f_E E_\delta (\hat{\sigma}^2) = \boldsymbol{\eta}^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \boldsymbol{\eta} - E_\delta (\tilde{\boldsymbol{\tau}}^T \mathbf{C}_\delta \tilde{\boldsymbol{\tau}}).$$

Diesen Ausdruck kürzen wir mit σ_0 ab. Falls Carryover-Effekte vorliegen, ist dann

$$f_E E_\delta (\hat{\sigma}^2) - \sigma_0 = \frac{n(p-1)}{p(t-1)} \left(p-1 - \frac{1}{p} - \frac{1}{t} \right) \boldsymbol{\rho}^T \mathbf{Q}_t \boldsymbol{\rho} \geq 0,$$

da $2 \leq p \leq t$. Da nach Satz 3.1 im Modell ohne Carryover-Effekte der Varianzschätzer erwartungstreu ist, folgt für vollständig balanciertes d_0 bei Vorliegen von Carryover-Effekten

$$E_\delta \left(\widehat{\text{var}} \left(\hat{\psi} \right) \right) \geq \text{var}_\delta \left(\hat{\psi} \right).$$

□

3.3 Simulationsstudie

Wir haben gezeigt, dass die Randomisation von Behandlungsnamen und Versuchseinheiten das Zeilen-Spalten Modell (3.2) für geeignete Versuchspläne im Sinne von Definition 3.1 rechtfertigt, falls keine Nachwirkungen vorliegen. Kunert und Sailer (2006) untersuchen mit Hilfe einer Simulationsstudie, ob sich die Auswertung im Zeilen-Spalten-Modell auch für Versuchspläne eignet, die den Bedingungen von Satz 3.1 nicht genügen. Weiterhin wird untersucht, wie verschiedene Versuchspläne abschneiden, wenn doch Carryover-Effekte vorliegen. Dieser Abschnitt beschreibt den Aufbau und die Ergebnisse dieser Simulationsstudie. Zunächst stellen wir die fünf Versuchspläne in der Studie vor.

3.3.1 Aufbau

Daten

Die Daten simulieren ein sensorisches Experiment, bei dem bis zu 30 Prüfer jeweils fünf Produkte nacheinander verkosten und auf einer Linienskala von 0 bis 10 bewerten. Die Beobachtungen werden für jede Versuchseinheit unabhängig aus der Lognormalverteilung generiert. Indem die Daten für jede Einheit in aufsteigender Reihenfolge sortiert werden, erzeugt man künstliche, nicht-additive Periodeneffekte. Die Beobachtungen werden schließlich zensiert, indem Werte größer als 10 auf 10 gesetzt werden. Der so erzeugte Ausgangsdatensatz ist in Tabelle A.1 in Anhang A abgebildet.

Man beachte, dass die Daten nicht dem Zeilen-Spalten-Modell genügen, da die Periodeneffekte nicht additiv sind. Genüßten die Daten dem Zeilen-Spalten-Modell, so bräuhete man das Modell nicht durch eine Randomisierung zu rechtfertigen. Auf diese Daten addiert man nach der Randomisierung Behandlungseffekte und Nachwirkungen entsprechend der unten aufgeführten Designs und Modellszenarien.

Periode	Einheit									
	1	2	3	4	5	6	7	8	9	10
1	1	2	3	4	5	3	4	5	1	2
2	5	1	2	3	4	4	5	1	2	3
3	2	3	4	5	1	2	3	4	5	1
4	4	5	1	2	3	5	1	2	3	4
5	3	4	5	1	2	1	2	3	4	5

Tabelle 3.1: Der Crossover-Plan $d_1 \in \Lambda_{5,10,5}$ mit Einheiten als Spalten und Perioden als Zeilen.

Versuchspläne

Der Plan d_1 aus Tabelle 3.1 ist ein vollständig balancierter Plan mit $t = 5$ Behandlungen für $n = 10$ Einheiten und $p = 5$ Perioden.

Plan d_2 entsteht durch Streichung der letzten Einheit in d_1 . Dieser Plan ist ein reguläres GYD, aber kein vollständig balancierter Plan.

Wie bereits erwähnt, existiert nicht für jede Wahl von (t, n, p) ein nachbarbalanciertes verallgemeinertes Youden-Design. Ball (1997) und Périnel und Pagès (2004) schlagen vor, in solchen Fällen näherungsweise balancierte Versuchspläne zu benutzen. Dabei wird zu einer vorgegebenen Zahl von Behandlungen, Einheiten und Versuchsperioden mit Hilfe eines numerischen Verfahrens ein Versuchsplan bestimmt, bei dem die Bedingungen BBD1 - BBD3 aus Definition 2.5 sowie die Balance der Nachwirkungen möglichst wenig verletzt sind (Périnel und Pagès, 2004). Diese Versuchspläne genügen im Allgemeinen nicht den Voraussetzungen von Satz 3.1.

Der Plan d_3 aus Tabelle 3.2 ist ein näherungsweise balancierter Plan aus Périnel und Pagès (2004) mit $t = 10$ Behandlungen für $n = 30$ Einheiten und $p = 5$ Perioden. Dieser Plan ist weder ein balancierter Blockplan noch ist er nachbarbalanciert.

Der Plan d_4 aus Tabelle 3.3 besteht aus nur zwei verschiedenen Behandlungssequenzen. Er ist in extremer Weise unbalanciert bezüglich der Perioden und in keiner Weise nachbarbalanciert.

Der Plan d_5 aus Tabelle 3.4 ist ein reguläres und nachbarbalanciertes, aber nicht vollständig balanciertes GYD. Es gibt drei Einheiten, die in den ersten beiden Perioden die Behandlungen 1 und 2 erhalten aber nur eine Einheit, die die Behandlungen 1 und 4 in den ersten beiden Perioden erhält. Das steht im Widerspruch zu Eigenschaft VB1 aus Definition 2.7.

Periode	Einheit														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	2	10	7	3	8	4	9	1	2	2	4	1	8	6	9
2	3	2	4	9	4	7	10	9	6	10	1	7	2	8	1
3	5	9	8	4	3	2	7	5	4	1	10	9	5	3	6
4	7	7	10	5	6	8	1	6	10	8	9	3	4	2	2
5	6	3	5	10	9	1	4	7	8	5	6	8	9	1	7

Periode	Einheit														
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	7	6	5	5	10	3	7	3	6	1	4	10	9	1	8
2	8	5	3	6	3	10	5	1	5	9	7	4	8	2	6
3	6	8	1	10	7	2	1	8	9	3	10	6	2	4	7
4	3	9	4	1	5	9	4	7	2	8	6	1	5	3	10
5	4	10	2	3	2	8	6	10	4	5	2	3	7	9	5

Tabelle 3.2: Der Crossover-Plan $d_3 \in \Lambda_{10,30,5}$ mit Einheiten als Spalten und Perioden als Zeilen.

Periode	Einheit																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3
3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5
5	5	5	5	5	5	5	5	5	5	5	1	1	1	1	1	1	1	1	1	1

Tabelle 3.3: Der Crossover-Plan $d_4 \in \Lambda_{5,20,5}$ mit Einheiten als Spalten und Perioden als Zeilen.

Periode	Einheit											
	1	2	3	4	5	6	7	8	9	10	11	12
1	1	2	3	4	1	2	3	4	1	2	3	4
2	2	3	4	1	3	4	1	2	2	1	4	3
3	4	1	2	3	2	3	4	1	3	4	1	2

Tabelle 3.4: Der Crossover-Plan $d_5 \in \Lambda_{4,12,3}$ mit Einheiten als Spalten und Perioden als Zeilen.

Randomisierung

In jeder der $n_{iter} = 10000$ Iterationen der Simulation wird aus den Ausgangsplänen durch Randomisierung von Behandlungsnamen und Einheiten ein neuer Versuchsplan erzeugt. Tritt an der Stelle (i, j) im randomisierten Plan die Behandlung $\delta_{(i,j)} = r$ auf, so addieren wir in dieser Iteration den Behandlungseffekt τ_r zur entsprechenden Beobachtung im Ausgangsdatensatz. Ist die Behandlung $\delta_{(i,j)}$ ein Nachfolger der Behandlung s , so addieren wir an der Stelle (i, j) des Ausgangsdatensatzes den Nachwirkungseffekt ρ_s auf die Beobachtung.

Modellparameter

Wir betrachten sechs Kombinationen von direkten Behandlungseffekten und Nachwirkungen und nehmen O.E. an, dass man sich für den Elementarkontrast $\psi = \tau_1 - \tau_2$ interessiert.

- Fall 1.** Alle Behandlungseffekte sind gleich und es gibt keine Nachwirkungen.
- Fall 2.** Die Behandlungen haben unterschiedliche Behandlungseffekte aber keine Nachwirkungen.
- Fall 3.** Die Behandlungen 1 und 2 besitzen den gleichen Behandlungseffekt und haben keine Nachwirkung. Behandlung 3 habe aber eine Nachwirkung.
- Fall 4.** Wie Fall 3, aber mit unterschiedlichen Behandlungseffekten bei Behandlung 1 und 2.
- Fall 5.** Die Behandlungen 1 und 2 haben die gleichen Behandlungseffekte aber unterschiedliche Carryover-Effekte.
- Fall 6.** Wie Fall 5, aber mit unterschiedlichen Behandlungseffekten bei Behandlung 1 und 2.

Kunert und Sailer (2006) wählen dazu τ_1 , ρ_1 und ρ_3 wie in Tabelle 3.5. Alle anderen Behandlungs- und Carryover-Effekte werden auf Null gesetzt. Damit ergeben sich die in der Tabelle aufgeführten Werte von ψ und $\ell^T \rho$.

Fall	τ_1	ρ_1	ρ_3	ψ	$\ell^T \boldsymbol{\rho}$
1	0	0	0	0	0
2	-1	0	0	-1	0
3	0	0	10	0	0
4	-1	0	10	-1	0
5	0	10	0	0	10
6	-1	10	0	-1	10

Tabelle 3.5: Die Wahl der Modellparameter in den sechs Modellszenarien.

Kriterien

In jeder Iteration i berechnet man für den randomisierten Versuchsplan δ_i den geschätzten Kontrast $\hat{\psi}_{(i)}$ und den Schätzer für die Varianz $\widehat{\text{var}}(\hat{\psi})_{(i)}$. Um die 1. Bedingung von Definition 3.1 zu überprüfen berechnen wir

$$Z_1 = \sqrt{10000} \frac{\bar{\hat{\psi}} - \psi}{\sqrt{s_{\hat{\psi}}^2}},$$

wobei $\bar{\hat{\psi}} = \frac{1}{10000} \sum_i (\hat{\psi}_{(i)})$ das arithmetische Mittel der Kontrastschätzer und

$$s_{\hat{\psi}}^2 = \frac{1}{10000 - 1} \sum_i (\hat{\psi}_{(i)} - \bar{\hat{\psi}})^2$$

die empirische Varianz. Aus dem dem zentralen Grenzwertsatz folgt, dass Z_1 approximativ standardnormalverteilt ist. Wir schließen, dass der Kontrastschätzer nicht erwartungstreu ist, falls in der Simulation $|Z_1| > u_{0.975}$, wobei $u_{0.975} = 1.96$ das 97,5%-Quantil der Standardnormalverteilung ist. Wir teilen die Simulation in 100 Teile zu je 100 Iterationen auf. Innerhalb jeder Gruppe $g = 1, \dots, 100$ berechnen wir die empirische Varianz s_g^2 der Kontrastschätzer und das arithmetische Mittel $\widehat{\text{var}}_g$ der Varianzschätzer $\widehat{\text{var}}(\hat{\psi})_{(i)}$. Dann ist

$$Z_2 = \sqrt{100} \frac{\frac{1}{100} \sum_g D_g}{\sqrt{\frac{1}{99} \sum_g (D_g - \frac{1}{100} \sum_h D_h)^2}}$$

mit

$$D_g = s_g^2 - \widehat{\text{var}}_g, \quad g = 1, \dots, 100,$$

approximativ standardnormalverteilt. Wir schließen, dass die Varianzschätzer der Kontrastschätzer nicht erwartungstreu sind, sofern $|Z_2| > 1.96$.

Für die Beurteilung der Auswertung im Zeilen-Spalten-Modell ist es auch interessant zu wissen, wie sich Tests auf Kontraste in den Fällen 1 bis 6 verhalten. Dazu berechnen wir t-Statistik

$$t_{(i)} = \frac{\hat{\psi}_{(i)}}{\sqrt{\widehat{\text{var}}(\hat{\psi})_{(i)}}}, \quad i = 1 \dots, 10000. \quad (3.8)$$

Mit Hilfe eines Permutationstests wird überprüft, ob die t-Statistiken aus der t-Verteilung mit $f_E = (n - 1)(p - 1) - t - 1$ Freiheitsgraden stammen. Sei Z_3 die Anzahl der der Permutationsstatistiken, die kleiner sind als das 5%-Quantil $t_{f_E;0.05}$ der t-Verteilung. Kunert und Sailer (2006) lehnen die Hypothese, dass die t-Statistiken tatsächlich t-verteilt sind ab, falls $Z_3 \notin [458, 542]$. Diese kritischen Werte werden aus einer Binomialverteilung mit Parametern 10000 und 0.05 berechnet.

3.3.2 Ergebnisse

Für die Fälle 1 und 2 zeigt Satz 3.1, dass sich das Zeilen-Spalten-Modell für die regulären GYD d_1 , d_2 und d_5 rechtfertigen lässt. Wie in Tabelle 3.6 zu sehen ist, bestätigt sich dies in der Simulationsstudie. Für die genannten Versuchspläne liegen weder Z_1 noch Z_2 im Ablehnbereich der Tests aus 3.3.1. Für den näherungsweise balancierten Versuchsplan d_3 , der den Voraussetzungen von Satz 3.1 nicht genügt, wird die Erwartungstreue von $\hat{\psi}$ nur knapp abgelehnt, während Z_2 und Z_3 nicht im Ablehnbereich liegen. Die t-Statistiken (3.8) sind für die genannten Pläne näherungsweise t-verteilt, obwohl die Ausgangsdaten nicht normalverteilt sind. Lediglich für den Versuchsplan d_4 wird die Varianz des Kontrastschätzers $\hat{\psi}$ unterschätzt ($Z_2 > 1.96$). Dies hat zur Folge, dass der t-Test signifikant anti-konservativ wird ($Z_3 > 542$). Die Randomisierung von Behandlungsnamen und Einheiten kann die Auswertung im Zeilen-Spalten-Modell für diesen in den Perioden nicht balancierten Plan nicht rechtfertigen.

Für die Fälle 3 und 4 folgt aus Satz 3.2, dass die mit Hilfe des Crossover-Plans d_1 geschätzte Varianz von $\hat{\psi}$ auch durch die extrem hohen Carryover-Effekte nicht erhöht wird. Das bestätigt sich in der Simulation, vergleiche Tabelle 3.7. Die Varianz wird wie im theoretischen Ergebnis vorhergesagt in der Simulation konservativ geschätzt.

Design	Fall 1			Fall 2		
	Z_1	Z_2	Z_3	Z_1	Z_2	Z_3
d_1	-0.36	-1.28	476	-0.36	-1.28	6154
d_2	-1.35	-0.20	471	-1.35	-0.20	5343
d_3	2.22	-1.18	466	2.22	-1.18	7728
d_4	-0.74	6.61	563	-0.74	6.61	6344
d_5	0.07	-0.03	500	0.07	-0.03	3331

Tabelle 3.6: Simulationsergebnisse bei den Fällen 1 und 2. Angegeben sind die Kriterien Z_1 , Z_2 und Z_3 für die Versuchspläne d_1 bis d_5 .

Design	Fall 3			Fall 4		
	Z_1	Z_2	Z_3	Z_1	Z_2	Z_3
d_1	-0.36	-715.70	0	-0.36	-715.70	0
d_2	-1.67	-248.82	0	-1.67	-248.82	17
d_3	0.87	-108.90	28	0.87	-108.90	925
d_4	0.26	78.69	3748	0.26	78.69	5452
d_5	0.81	-268.78	0	0.81	-268.78	27

Tabelle 3.7: Simulationsergebnisse bei den Fällen 3 und 4. Angegeben sind die Kriterien Z_1 , Z_2 und Z_3 für die Versuchspläne d_1 bis d_5 .

Design	Fall 5			Fall 6		
	Z_1	Z_2	Z_3	Z_1	Z_2	Z_3
d_1	-404.78	-765.76	48	-404.78	-765.76	3005
d_2	-272.56	-387.71	199	-272.56	-387.71	2555
d_3	-353.40	-230.47	5096	-353.40	-230.47	9707
d_4	-43.18	78.64	4153	-43.18	78.64	6706
d_5	-408.89	-289.44	3896	-408.89	-289.44	7749

Tabelle 3.8: Simulationsergebnisse bei den Fällen 5 und 6. Angegeben sind die Kriterien Z_1 , Z_2 und Z_3 für die Versuchspläne d_1 bis d_5 .

Allerdings ist die Varianzschätzung so extrem konservativ, dass in Fall 4 die Nullhypothese mit Hilfe des t-Tests nicht abgelehnt werden kann. Die Versuchspläne d_2 , d_3 und d_5 erfüllen die Bedingungen von Satz 3.2 nicht. Die Varianz des Kontrastschätzers ist hier in den Fällen 3 und 4 durchweg höher als in den Fällen 1 und 2. Allerdings ist die Varianzschätzung noch immer konservativ, da die Carryover-Effekte zu einer sehr stark erhöhten geschätzten Varianz führen. Dementsprechend ergibt sich ein konservativer t-Test. Bei Plan d_4 erhöht sich die Varianz des Kontrastschätzers durch die Carryover-Effekte in der Simulation um mehr als den Faktor 100, während die geschätzte Varianz unverändert bleibt. Das hat einen sehr stark anti-konservativen t-Test zur Folge.

Wie man anhand von Tabelle 3.8 erkennen kann, ist der Kontrastschätzer $\hat{\psi}$ in den Fällen 5 und 6 bei allen Versuchsplänen verzerrt. Immerhin wird die Varianz $\text{var}_\delta(\hat{\psi})$ bei den Plänen d_1 , d_2 , d_3 und d_5 noch immer konservativ geschätzt. Die Carryover-Effekte führen aber zu einer so starken Verzerrung von $\hat{\psi}$, dass die t-Tests entweder in Fall 5 anti-konservativ sind oder in Fall 6 eine geringe Power aufweisen.

3.4 Bemerkungen zur Modellrechtfertigung durch Randomisierung

Wenn durch die Planung einer Crossover-Studie sichergestellt werden kann, dass keine Carryover-Effekte auftreten, dann kann eine Auswertung im Zeilen-Spalten-Modell durchaus gerechtfertigt werden. Die Randomisierung von Behandlungsnamen und Einheiten kann die Auswertung im Zeilen-Spalten-Modell auch dann rechtfertigen, wenn dieses Modell den Daten gar nicht zugrunde liegt. Die theoretischen Ergebnisse hierzu gelten für vollständig balancierte Pläne. Aber auch für näherungsweise vollständig balancierte Pläne wie die von Périnel und Pagès (2004) vorgestellten Designs ist das Zeilen-Spalten-Modell in der Praxis anwendbar, da Definition 3.1 hier bestenfalls minimal verletzt wird. Die Tatsache, dass nicht für jede Wahl von (t, n, p) ein vollständig balancierter Plan existiert, ist also keine große Einschränkung.

Darüberhinaus kann die Randomisierung auch bei nicht normalverteilten Daten dazu führen, dass die herkömmlichen t- und F-Tests unverzerrt sind. Obwohl die Daten in der Simulationsstudie nicht normalverteilt sind, sind die t-Statistiken der Permutationstests annähernd t-verteilt. Dies veranschaulicht, dass die Normalvertei-

lungsannahme für die Auswertung geplanter randomisierter Experimente nicht unbedingt notwendig ist. Darauf weisen auch Kunert et al. (2002) hin. Dennoch sollte man beachten, dass für manche nicht normalverteilten Datensätze die asymptotische Verteilung der t-Statistiken im Permutationstest nicht gleich der t-Verteilung ist. Ein t-Test kann dann zu falschen Testentscheidungen führen.

Dies gilt beispielsweise dann, wenn es in den Daten einen einzigen großen Ausreißer gibt. Nur die Aussagen zum Mittelwert und der Varianz der Kontraste gelten unabhängig von der Verteilung der Daten.

Die Randomisierung nach Azaïs (1987) rechtfertigt, anders als die Randomisierung nach Bailey (1985), sogar eine Auswertung im einfachen Blockmodell. Es sei aber darauf hingewiesen, dass der Begriff der gerechtfertigten Auswertung in Definition 3.1 nicht bedeutet, dass eine solche Auswertung auch notwendigerweise Tests hoher Power zulässt. Wenn in einem Versuch tatsächlich Periodeneffekte auftreten, sollten diese auch modelliert werden, da ansonsten die Periodeneffekte vom Fehlerterm absorbiert werden und die Power sinkt, siehe hierzu auch Kunert (1998b). In jedem Fall ist es wichtig, dass das zu rechtfertigende Modell zu der vorgenommenen Randomisierung passt.

Wenn nicht ausgeschlossen werden kann, dass Carryover-Effekte auftreten, dann bieten nachbarbalancierte Pläne einen gewissen Schutz vor Carryover-Effekten. Sobald aber starke Nachwirkungen von Behandlungen auftreten, sollten diese bei der Auswertung explizit berücksichtigt werden. Auch die Randomisierung nachbarbalancierter Pläne kann dann nicht verhindern, dass die Kontrastschätzer stark verzerrt und Tests auf Behandlungskontraste eventuell anti-konservativ sind.

In den folgenden Kapiteln dieser Arbeit betrachten wir daher das Modell (2.6), welches Carryover-Effekte explizit berücksichtigt.

4 Das Verfahren von Kunert und Utzig

In Kapitel 3 haben wir gesehen, dass die Schätzer für Behandlungskontraste im linearen Modell ohne Nachwirkungseffekte trotz Randomisierung verzerrt sind, falls Nachwirkungseffekte vorliegen. In diesem Kapitel untersuchen wir eine Auswertung für das Modell mit Carryover-Effekten und korrelierten Fehlern (2.6). Wie Kunert und Utzig (1993) bemerken, kann dieses Modell nicht durch Randomisierung gerechtfertigt werden. Das bedeutet nicht, dass bei einer Auswertung in Modellen mit Carryover-Effekten nicht randomisiert werden darf. Wir gehen für das hier vorgestellte Verfahren davon aus, dass Versuchseinheiten und Behandlungsnamen randomisiert werden. Dies ist gerade die Randomisierung, die wir in Kapitel 3 näher untersucht haben.

Modell (2.6) unterstellt, dass den Beobachtungen an jeder Versuchseinheit die gleiche unbekannte Kovarianzmatrix \mathbf{S} zugrunde liegt. Diese Annahme lässt sich durch die Randomisierung der Versuchseinheiten rechtfertigen (Kunert und Utzig, 1993). Wäre \mathbf{S} bekannt, so würden wir Aitken-Schätzer für die Behandlungseffekte berechnen. In Kapitel 5 betrachten wir ein Verfahren, bei dem \mathbf{S} aus den Daten geschätzt wird. Kunert und Utzig (1993) umgehen die Schätzung der unbekanntes Kovarianzmatrix, da eine solche Schätzung bei kleinem Stichprobenumfang unpräzise ist. Stattdessen werten sie Crossover-Studien unter Modell (2.7) aus und korrigieren die zugehörigen KQ-Schätzer so, dass sie in (2.6) konservative Varianzschätzer erhalten. In diesem Kapitel beschreiben wir zunächst das Verfahren von Kunert und Utzig (1993) kurz und leiten dann ausführlich eine verbesserte Variante des Verfahrens her.

Im Folgenden kennzeichnen wir Erwartungswerte von Zufallsvariablen unter Modell (2.7) mit $E_{\mathbf{I}}$, (Schätzer für) Varianzen mit $\text{var}_{\mathbf{I}}$ ($\widehat{\text{var}}_{\mathbf{I}}$) und Kovarianzmatrizen mit $\text{Cov}_{\mathbf{I}}$. Unter Modell (2.6) verwenden wir entsprechend die Notation $E_{\mathbf{S}}$, $\text{var}_{\mathbf{S}}$, $\widehat{\text{var}}_{\mathbf{S}}$ und $\text{Cov}_{\mathbf{S}}$. Zur Auswertung benutzen wir die vollständig balancierten Versuchspläne aus

Definition 2.7, da sie in Modell (2.7), wie bereits in Kapitel 2.3 erwähnt, hervorragende Optimalitätseigenschaften besitzen.

4.1 Abschätzung eines Worst-Case-Szenarios

4.1.1 Die obere Schranke von Kunert und Utzig

Da die Behandlungseffekte in (2.7) nicht schätzbar sind, betrachten wir zum Vergleich verschiedener Behandlungen Behandlungskontraste $\psi = \boldsymbol{\ell}^T \boldsymbol{\tau}$. O.E. seien die Kontraste standardisiert, das heißt $\sum_i \ell_i^2 = 1$. Diese Kontraste schätzen wir mit Hilfe der KQ-Methode in Modell (2.7). Kunert und Utzig (1993) untersuchen die Eigenschaften der Kontrastschätzer $\hat{\psi}$, wenn in Wirklichkeit Modell (2.6) vorliegt. Wir wiederholen nun kurz die Ergebnisse dieser Autoren.

Lemma 4.1. *Sei $d \in \Lambda_{t,n,p}^*$. Definiere die Informationsmatrizen $\mathbf{C}_d = \mathbf{T}^T \boldsymbol{\omega}_{[U,P,F]}^\perp \mathbf{T}$, $\mathbf{C}_{d11} = \mathbf{T}^T \boldsymbol{\omega}_{[U,P]}^\perp \mathbf{T}$, $\mathbf{C}_{d12} = \mathbf{T}^T \boldsymbol{\omega}_{[U,P]}^\perp \mathbf{F}$ und $\mathbf{C}_{d22} = \mathbf{F}^T \boldsymbol{\omega}_{[U,P]}^\perp \mathbf{F}$. Dann gilt*

$$\begin{aligned} \mathbf{C}_d &= c_d \mathbf{Q}_t, \\ \mathbf{C}_{d11} &= c_{d11} \mathbf{Q}_t, \\ \mathbf{C}_{d12} &= c_{d12} \mathbf{Q}_t \\ \text{und} \quad \mathbf{C}_{d22} &= c_{d22} \mathbf{Q}_t, \end{aligned}$$

wobei

$$\begin{aligned} c_d &= \frac{n(p-1)}{t-1} \left(1 - \frac{t}{p(pt-t-1)} \right), \\ c_{d11} &= \frac{n(p-1)}{t-1}, \\ c_{d12} &= \frac{-n(p-1)}{p(t-1)} \\ \text{und} \quad c_{d22} &= \frac{n(p-1)(pt-t-1)}{p(t-1)t}, \end{aligned}$$

vergleiche Kunert und Utzig (1993).

Nach Satz 2.1 ist der BLUE für ψ in Modell (2.7) gegeben durch

$$\hat{\psi} = \boldsymbol{\ell}^T \mathbf{C}_d^{-1} \mathbf{T}^T \boldsymbol{\omega}_{[U,P,F]}^\perp \mathbf{y}.$$

In Modell (2.7) ist $\hat{\psi}$ unverzerrt für ψ , das heißt $E_{\mathbf{I}}(\hat{\psi}) = \psi$. Aber auch in (2.6) ist $\hat{\psi}$ erwartungstreu, denn

$$\begin{aligned} E_{\mathbf{S}}(\hat{\psi}) &= E_{\mathbf{S}}\left(\frac{1}{c_d} \boldsymbol{\ell}^T \mathbf{T}^T \boldsymbol{\omega}_{[U, P, F]}^\perp \left([U, P, F] (\boldsymbol{\alpha}^T, \boldsymbol{\pi}^T, \boldsymbol{\rho}^T)^T + T\boldsymbol{\tau} + \boldsymbol{\epsilon}\right)\right) \\ &= \frac{1}{c_d} \boldsymbol{\ell}^T \mathbf{C}_d \boldsymbol{\tau} + \frac{1}{c_d} \boldsymbol{\ell}^T \mathbf{T}^T \boldsymbol{\omega}_{[U, P, F]}^\perp E_{\mathbf{S}}(\boldsymbol{\epsilon}) \\ &= \frac{c_d}{c_d} \boldsymbol{\ell}^T \boldsymbol{\tau} + 0 \\ &= \psi. \end{aligned}$$

In Modell (2.7) ist der Schätzer für σ^2 gegeben durch $\hat{\sigma}^2 = \frac{1}{f_E} \mathbf{y}^T \boldsymbol{\omega}_{[U, P, F, T]}^\perp \mathbf{y}$, wobei

$$\begin{aligned} f_E &= np - 1 - (n - 1) - (p - 1) - (t - 1) - (t - 1) \\ &= np - 2t - n - p + 3 \end{aligned}$$

die Anzahl der Freiheitsgrade zum Schätzen von σ^2 angibt. Für einen standardisierten Kontrast ψ ist dann

$$\widehat{\text{var}}_{\mathbf{I}}(\hat{\psi}) = \hat{\sigma}^2 \boldsymbol{\ell}^T \mathbf{C}_d^{-1} \boldsymbol{\ell} = \frac{\hat{\sigma}^2}{c_d}$$

ein unverzerrter Schätzer für $\text{var}_{\mathbf{I}}(\hat{\psi})$.

Kunert und Utzig (1993) setzen jetzt voraus, dass die Behandlungsnamen im Versuchsplan randomisiert sind. Das bedeutet, dass d aus einem Versuchsplan $d_0 \in \Lambda_{t, n, p}^*$ bestimmt wird, indem die Behandlungen einer Permutation der Zahlen $1, \dots, t$ zugeordnet werden. Offensichtlich ist dann auch $d \in \Lambda_{t, n, p}^*$. Die Randomisierung führt dazu, dass $\boldsymbol{\ell}$ zufällig wird. Dies hat auf die bisherigen Ergebnisse keinen Einfluss, da sie unabhängig von der Wahl von $\boldsymbol{\ell}$ waren. Man sieht dann leicht ein, dass alle standardisierten Behandlungskontraste gleiche Varianz besitzen. Schreibe nun $\tilde{\mathbf{S}} = \mathbf{Q}_p \mathbf{S} \mathbf{Q}_p$ für die *zentrierte Kovarianzmatrix* und s_{11} für das erste Diagonalelement von $\tilde{\mathbf{S}}$. Dann ist nach Kunert und Utzig (1993)

$$\begin{aligned} E_{\mathbf{S}}\left(\widehat{\text{var}}_{\mathbf{I}}(\hat{\psi})\right) &= \frac{1}{c_d f_E} \left((n - 1) \text{tr}(\tilde{\mathbf{S}}) - \frac{n}{c_{d22}} \left(\text{tr}(\tilde{\mathbf{S}}) - \frac{(t - 1)s_{11}}{t} \right) \right) \\ &\quad - \frac{t - 1}{f_E} \text{var}_{\mathbf{S}}(\hat{\psi}). \end{aligned}$$

Hierbei ist

$$\text{var}_{\mathcal{S}}(\hat{\psi}) = \frac{n}{(t-1)c_d^2} \text{tr} \left(\tilde{\mathbf{S}} - 2\gamma \mathbf{V}^T \tilde{\mathbf{S}} + \gamma^2 \mathbf{V}^T \tilde{\mathbf{S}} \mathbf{V} - \frac{1}{t} \gamma^2 \mathbf{1}_p^T \mathbf{V}^T \tilde{\mathbf{S}} \mathbf{V} \mathbf{1}_p \right) \quad (4.1)$$

mit

$$\gamma = \frac{c_{d12}}{c_{d22}} = \frac{-t}{pt - t - 1} \quad (4.2)$$

und

$$\mathbf{V} = \begin{bmatrix} \mathbf{0} & 0 \\ \mathbf{I}_{p-1} & \mathbf{0} \end{bmatrix}.$$

Im Allgemeinen ist $E_{\mathcal{S}}(\widehat{\text{var}}_{\mathbf{I}}(\hat{\psi})) \neq \text{var}_{\mathcal{S}}(\hat{\psi})$. Wir betrachten nun den Quotienten dieser beiden Ausdrücke, das Verhältnis zwischen wahrer Varianz des Kontrastschätzers $\hat{\psi}$ und der mittleren geschätzten Varianz von $\hat{\psi}$ in Modell (2.6).

Definition 4.1. Für festes $(t, n, p) \in \mathbb{N}^3$ sei $d \in \Lambda_{t,n,p}^*$ und die Behandlungsnamen von d seien randomisiert. Sei $\mathbf{S} \in \mathbb{R}^{p \times p}$ n.n.d. und \mathcal{S} die Menge aller n.n.d. Matrizen des $\mathbb{R}^{p \times p}$.

Dann ist der Varianzquotient $k_{t,n,p}$ definiert durch

$$k_{t,n,p}(\mathbf{S}) = \frac{\text{var}_{\mathcal{S}}(\hat{\psi})}{E_{\mathcal{S}}(\widehat{\text{var}}_{\mathbf{I}}(\hat{\psi}))}. \quad (4.3)$$

Das Maximum $k_{t,n,p}^*$ des Varianzquotienten bezüglich $\mathbf{S} \in \mathcal{S}$ nennen wir Konstante von Kunert und Utzig, das heißt

$$k_{t,n,p}^* = \max_{\mathbf{S} \in \mathcal{S}} k_{t,n,p}(\mathbf{S}). \quad (4.4)$$

Eine Matrix, für die das Maximum angenommen wird, nennen wir ungünstigste Kovarianzmatrix.

Analog kann man auch eine günstigste Kovarianzmatrix \mathbf{S} definieren, die den Varianzquotienten minimiert. Wir sprechen von günstigen Kovarianzmatrizen, wenn sie zu einem kleinen Varianzquotienten führen und von ungünstigen Kovarianzmatrizen, wenn sie mit einem großen Varianzquotienten einhergehen.

Der Varianzquotient $k_{t,n,p}(\mathbf{S})$ hängt nach Konstruktion für festes (t, n, p) nicht von der konkreten Wahl des Versuchsplans ab, er ist unabhängig von der Wahl von ℓ und

nimmt für $\mathbf{S} = \sigma^2 \mathbf{I}_p$ den Wert 1 an. Falls der Quotient für ein konkretes \mathbf{S} kleiner als 1 ist, so ist der Varianzschätzer in Modell (2.6) konservativ. Wesentlich schlimmer ist der Fall $k_{t,n,p}(\mathbf{S}) > 1$, dann wird die wahre Varianz unterschätzt, das heißt der Varianzschätzer ist anti-konservativ. Mit Hilfe der nächsten Definition kommen wir zu einer ersten Aussage über die Konstante von Kunert und Utzig.

Definition 4.2. Für festes $(t, n, p) \in \mathbb{N}^3$ sei

$$k_{t,n,p}^v = \begin{cases} \left(\frac{(t-1)c_d(u(z^2+z+1)+v)}{f_E n(w(z^2+z+1)-z^2\gamma-r)} - \frac{t-1}{f_E} \right)^{-1}, & \text{falls } p = 3 \\ \left(\frac{(t-1)c_d u}{f_E n(w-\gamma)} - \frac{t-1}{f_E} \right)^{-1}, & \text{falls } p = 4 \\ \left(\frac{(t-1)c_d u}{f_E n(1-\gamma)^2} - \frac{t-1}{f_E} \right)^{-1}, & \text{falls } p \geq 4 \end{cases}.$$

Wir definieren die obere Schranke von Kunert und Utzig $k_{t,n,p}^{\text{oben}}$ durch

$$k_{t,n,p}^{\text{oben}} = \begin{cases} k_{t,n,p}^v, & \text{falls } k_{t,n,p}^v > 0 \\ \infty & \text{sonst} \end{cases}. \quad (4.5)$$

Dabei ist γ wie in (4.2), $u = n - 1 - n/c_{d22}$, $v = n(t+1)/(2tc_{d22})$, $w = 1 + \gamma^2$, $r = (t+1)\gamma^2/(2t)$ und

$$z = \frac{1}{\gamma u} \left(-(u+v)\gamma - ru - vw - \sqrt{(-(u+v)\gamma - ru - vw)^2 - u(ru + vw)\gamma} \right).$$

Satz 4.1. (Kunert und Utzig, 1993). Die obere Schranke von Kunert und Utzig $k_{t,n,p}^{\text{oben}}$ erfüllt

$$k_{t,n,p}^{\text{oben}} \geq k_{t,n,p}^*.$$

Die Aussage wird in Kunert und Utzig (1993) hergeleitet. Details der Herleitung entnehme man Utzig (1990).

Man beachte, dass die obere Schranke von Kunert und Utzig von den Designparametern (t, n, p) abhängt und auch mit Hilfe eines Taschenrechners leicht bestimmt werden kann. Für die praktische Anwendung ist es erforderlich, dass die Schranke darüberhinaus möglichst scharf ist. Wenn nämlich die Beobachtungen innerhalb eines Blockes doch unkorreliert sein sollten, so führt die Anwendung der Konstanten wie wir in der Simulationsstudie in Kapitel 6.2 sehen werden bereits zu konservativen Tests

auf Behandlungskontraste. Wenn dann die obere Schranke für die Konstante $k_{t,n,p}^*$ nicht scharf ist, wird ein Test umso konservativer ausfallen.

Die obere Schranke ist im Fall $p = 3$ scharf. Man beachte, dass eine Matrix genau dann Kovarianzmatrix ist, wenn sie n.n.d. ist. Die Herleitung der Schranken für $p \geq 4$ basiert auf der Eigenschaft einer n.n.d. Matrix \mathbf{S} , dass $\mathbf{x}^T \mathbf{S} \mathbf{x} \geq 0$ für alle \mathbf{x} . Kunert und Utzig (1993) wählen dazu ein \mathbf{x} geschickt. Wie wir in Kapitel 4.2 sehen werden, ist die so hergeleitete Schranke leider nicht scharf. Für den Fall $p = 3$ nutzen die Autoren die Eigenschaft, dass die Minoren von \mathbf{S} nicht negativ sind. Diese Eigenschaft ließe sich mit einigem Aufwand dazu nutzen, auch im Fall $p = 4$ die Konstante $k_{t,n,4}^*$ zu bestimmen, für allgemeines p ist diese Vorgehensweise jedoch nicht praktikabel.

Einen Überblick über die obere Schranke aus Definition 4.2 gibt Tabelle 4.1, siehe auch Kunert und Utzig (1993)[S. 926]. Das R-Programm `k.upper` in Anhang B.1 berechnet zu gegebenem t, n, p die obere Schranke von Kunert und Utzig.

$k_{t,n,p}^{oben}$ hängt von n ab, konvergiert aber sehr schnell. Bei Versuchsplänen mit geringer Periodenanzahl wird die Varianz zum Teil stark korrigiert, zum Beispiel um den Faktor 2.11 bei einem Plan zum Vergleich von vier Behandlungen mit acht Versuchspersonen, die jeweils alle Behandlungen zugeteilt bekommen. Die oberen Schranken für $p = 4$ und $p = 5$ sind bei gleichem t und p zum Teil größer als bei $p = 3$, da die entsprechenden Schranken weniger scharf sind. Mit wachsendem p nähert sich die Konstante schließlich der Zahl 1, siehe zum Beispiel den Fall $p = 100$. Allerdings kommen derartige Versuchspläne in der Praxis selten oder gar nicht vor.

Konservative Varianzschätzer lassen sich auch für andere Modelle als das hier betrachtete berechnen. Kunert (1987) untersucht ein Modell, bei dem die Einheiten in Gruppen unterteilt werden und bei dem in unterschiedlichen Gruppen unterschiedliche Periodeneffekte auftreten. Dort wird auch eine obere Schranke für die Varianz von Schätzern des Carryover-Effekts hergeleitet. Wenn man zeigen möchte, dass Carryover-Effekte bei einer Studie vernachlässigbar klein sind, dann könnte wie Kunert (1987) bemerkt eine untere Schranke für die Varianz von Carryover-Effekten eventuell von Interesse sein. Kunert (1998a) berechnet eine konservative Varianzabschätzung für zirkulär balancierte Versuchspläne in Verbindung mit der von Azais (1987) vorgeschlagenen Randomisierung, vergleiche Kapitel 3.1.2.

$p = 3$		t						
m	3	4	5	6	7	10	100	
1								
2	1.80	1.77	1.75	1.74	1.73	1.71	1.68	
4	1.51	1.49	1.47	1.46	1.46	1.45	1.43	
6	1.48	1.45	1.44	1.43	1.43	1.42	1.40	
1000000	1.44	1.41	1.40	1.39	1.39	1.38	1.36	
$p = 4$		t						
m	3	4	5	6	7	10	100	
1		∞	∞	∞	∞	84.90	18.80	
2		2.11	2.10	2.09	2.08	2.07	2.04	
4		1.80	1.78	1.78	1.77	1.76	1.74	
6		1.74	1.72	1.71	1.71	1.70	1.67	
1000000		1.65	1.63	1.62	1.61	1.60	1.58	
$p = 5$		t						
m	3	4	5	6	7	10	100	
1			3.37	3.29	3.24	3.16	2.99	
2			1.96	1.96	1.96	1.95	1.93	
4			1.79	1.78	1.78	1.77	1.76	
6			1.75	1.74	1.74	1.73	1.72	
1000000			1.68	1.68	1.67	1.66	1.65	
$p = 6$		t						
m	3	4	5	6	7	10	100	
1				2.01	2.00	1.98	1.94	
2				1.65	1.64	1.64	1.63	
4				1.56	1.56	1.56	1.55	
6				1.54	1.54	1.54	1.53	
1000000				1.51	1.51	1.50	1.49	
$p = 7$		t						
m	3	4	5	6	7	10	100	
1					1.65	1.64	1.62	
2					1.49	1.48	1.48	
4					1.44	1.44	1.43	
6					1.43	1.42	1.42	
1000000					1.40	1.40	1.40	
$p = 10$		t						
m	3	4	5	6	7	10	100	
1						1.32	1.32	
2						1.28	1.28	
4						1.26	1.26	
6						1.26	1.26	
1000000						1.25	1.25	
$p = 100$		t						
m	3	4	5	6	7	10	100	
1							1.02	
2							1.02	
4							1.02	
6							1.02	
1000000							1.02	

Tabelle 4.1: Die obere Schranke $k_{t,n,p}^{oben}$ (4.5) von Kunert und Utzig für Pläne zum Vergleich von t Behandlungen in $n = mt$ Blöcken der Länge p .

4.1.2 Korrigierte Tests

Wenn wir nun die Varianz eines Kontrastschätzers unter Modell (2.6) schätzen wollen, so können wir den Varianzschätzer $\widehat{\text{var}}_{\mathbf{I}}(\hat{\psi})$ mit $k_{t,n,p}^{\text{oben}}$ multiplizieren. Damit stellen wir sicher, dass wir die wahre Varianz von $\hat{\psi}$ im Mittel nicht unterschätzen, wenn wir fälschlicherweise Modell (2.7) zugrundelegen, denn für jedes beliebige \mathbf{S} ist

$$\begin{aligned} k_{t,n,p}^{\text{oben}} \mathbb{E}_{\mathbf{S}} \left(\widehat{\text{var}}_{\mathbf{I}} \left(\hat{\psi} \right) \right) &\geq k_{t,n,p}^* \mathbb{E}_{\mathbf{S}} \left(\widehat{\text{var}}_{\mathbf{I}} \left(\hat{\psi} \right) \right) \\ &\geq k_{t,n,p}(\mathbf{S}) \mathbb{E}_{\mathbf{S}} \left(\widehat{\text{var}}_{\mathbf{I}} \left(\hat{\psi} \right) \right) \\ &= \text{var}_{\mathbf{S}} \left(\hat{\psi} \right), \end{aligned}$$

vergleiche Definition 4.1 und Satz 4.1.

Damit kann man die Schranke $k_{t,n,p}^{\text{oben}}$ nutzen, um die Tests auf Behandlungseffekte aus Definition 2.3 zu korrigieren. Kunert und Utzig (1993)[S. 926 f.] schlagen folgenden t-Test der Nullhypothese $H_0^k : \psi = \psi_0$ gegen die Alternative $H_1^k : \psi \neq \psi_0$ vor, und zeigen, dass dieser Test asymptotisch konservativ ist.

Lehne H_0^k zum Niveau α ab, falls

$$\frac{|\hat{\psi} - \psi_0|}{\sqrt{k_{t,n,p}^{\text{oben}} \widehat{\text{var}}_{\mathbf{I}} \left(\hat{\psi} \right)}} > t_{f_E; 1-\alpha/2}.$$

Dabei sind $\hat{\psi} = \boldsymbol{\ell}^T \mathbf{C}_d^- \mathbf{T}^T \boldsymbol{\omega}_{[U,P,F]}^\perp \mathbf{y}$, $\widehat{\text{var}}_{\mathbf{I}}(\hat{\psi}) = \hat{\sigma}^2 \boldsymbol{\ell}^T \mathbf{C}_d^- \boldsymbol{\ell}$, und $\hat{\sigma}^2 = \mathbf{y}^T \boldsymbol{\omega}_{[U,P,F,T]}^\perp \mathbf{y} / f_E$.

Statt der Teststatistik kann man äquivalent auch den kritischen Wert korrigieren, so dass man H_0^k ablehnt, falls

$$\frac{|\hat{\psi} - \psi_0|}{\sqrt{\widehat{\text{var}}_{\mathbf{I}} \left(\hat{\psi} \right)}} > \sqrt{k_{t,n,p}^{\text{oben}}} t_{f_E; 1-\alpha/2}. \quad (4.6)$$

Der korrigierte globale F-Test auf Behandlungsunterschiede lehnt entsprechend die Nullhypothese $H_0^g : \tau_1 = \dots = \tau_t$ ab, falls

$$\frac{MST}{MSE} > k_{t,n,p}^{\text{oben}} F_{t-1, f_E; 1-\alpha},$$

wobei $MST = \mathbf{y}^T (\boldsymbol{\omega}_{[U,P,F,T]} - \boldsymbol{\omega}_{[U,P,F]}) \mathbf{y} / (t-1)$ und $MSE = \hat{\sigma}^2$.

4.2 Die Maximierung des Varianzquotienten

4.2.1 Darstellung als Eigenwertproblem

In diesem Abschnitt wird die Konstante von Kunert und Utzig aus Definition 4.1 exakt bestimmt. Wir leiten zunächst noch einmal den Zähler und Nenner des Varianzquotienten (4.3) her und kommen so zu einer neuen Darstellung des Maximierungsproblems (4.4). Weitere Hilfssätze ermöglichen schließlich die Bestimmung des Maximums. Mit der Notation $\mathbf{Q}_{p,t} = \mathbf{I}_p - \frac{1}{t} \mathbf{1}_p \mathbf{1}_p^T$ erhalten wir unser erstes Ergebnis.

Lemma 4.2.

$$\text{var}_{\mathbf{S}}(\hat{\psi}) = \frac{n}{(t-1)c_d^2} \text{tr}(\mathbf{A}_d \tilde{\mathbf{S}}).$$

Dabei ist

$$\mathbf{A}_d = \mathbf{I}_p - \gamma(\mathbf{V} + \mathbf{V}^T) + \gamma^2 \mathbf{V} \mathbf{Q}_{p,t} \mathbf{V}^T,$$

γ wie in (4.2) und $\tilde{\mathbf{S}}$ die zentrierte Kovarianzmatrix.

Beweis. Da die Spur eines Matrixproduktes sich bei zyklischer Vertauschung der Faktoren nicht ändert ist

$$\begin{aligned} \text{tr}\left(\mathbf{V}^T \tilde{\mathbf{S}} \mathbf{V} - \frac{1}{t} \mathbf{1}_p^T \mathbf{V}^T \tilde{\mathbf{S}} \mathbf{V} \mathbf{1}_p\right) &= \text{tr}\left(\mathbf{V}^T \mathbf{Q}_p \mathbf{S} \mathbf{Q}_p \mathbf{V} - \frac{1}{t} \mathbf{1}_p \mathbf{1}_p^T \mathbf{V}^T \mathbf{Q}_p \mathbf{S} \mathbf{Q}_p \mathbf{V}\right) \\ &= \text{tr}(\mathbf{Q}_p \mathbf{V} \mathbf{Q}_{p,t} \mathbf{V}^T \mathbf{Q}_p \mathbf{S}). \end{aligned}$$

Weiterhin gilt $\text{tr}(\mathbf{V}^T \tilde{\mathbf{S}}) = \text{tr}(\tilde{\mathbf{S}} \mathbf{V})$, so dass mit (4.1)

$$\begin{aligned} \text{var}_{\mathbf{S}}(\hat{\psi}) &= \frac{n}{(t-1)c_d^2} \text{tr}(\mathbf{Q}_p \mathbf{Q}_p \mathbf{S} - \gamma \mathbf{Q}_p (\mathbf{V} + \mathbf{V}^T) \mathbf{Q}_p \mathbf{S} + \gamma^2 \mathbf{Q}_p \mathbf{V} \mathbf{Q}_{p,t} \mathbf{V}^T \mathbf{Q}_p \mathbf{S}) \\ &= \frac{n}{(t-1)c_d^2} \text{tr}(\mathbf{Q}_p (\mathbf{I} - \gamma(\mathbf{V} + \mathbf{V}^T) + \gamma^2 \mathbf{V} \mathbf{Q}_{p,t} \mathbf{V}^T) \mathbf{Q}_p \mathbf{S}) \\ &= \frac{n}{(t-1)c_d^2} \text{tr}(\mathbf{Q}_p \mathbf{A}_d \mathbf{Q}_p \mathbf{S}) \\ &= \frac{n}{(t-1)c_d^2} \text{tr}(\mathbf{A}_d \tilde{\mathbf{S}}), \end{aligned}$$

wobei $\mathbf{A}_d = \mathbf{I}_p - \gamma(\mathbf{V} + \mathbf{V}^T) + \gamma^2 \mathbf{V} \mathbf{Q}_{p,t} \mathbf{V}^T$. □

Lemma 4.3.

$$\mathbb{E}_{\mathcal{S}} \left(\widehat{\text{var}}_{\mathbf{I}} \left(\hat{\psi} \right) \right) = \frac{n}{c_d^2 f_E} \text{tr} \left(\mathbf{B}_d \tilde{\mathbf{S}} \right).$$

Dabei ist

$$\mathbf{B}_d = \left(\frac{c_d(n-1)}{n} - 1 \right) \mathbf{I}_p + \gamma (\mathbf{V} + \mathbf{V}^T) + p\gamma \mathbf{V} \mathbf{Q}_{p,t} \mathbf{V}^T,$$

γ wie in (4.2) und $\tilde{\mathbf{S}}$ die zentrierte Kovarianzmatrix.

Beweis. Es ist

$$\hat{\sigma}^2 = \frac{1}{f_E} \mathbf{y}^T \boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}, \mathbf{F}, \mathbf{T}]}^{\perp} \mathbf{y},$$

$$\mathbb{E}_{\mathcal{S}}(\mathbf{y}) = \mathbf{1}_{np} \mu + \mathbf{U} \boldsymbol{\alpha} + \mathbf{P} \boldsymbol{\pi} + \mathbf{T} \boldsymbol{\tau} + \mathbf{F} \boldsymbol{\rho}$$

und $\text{Cov}_{\mathcal{S}}(\mathbf{y}) = \boldsymbol{\Sigma}$, vergleiche Kapitel 2.1.1. Weiterhin ist $\mathbf{1}_{np}$ im Spaltenraum von $[\mathbf{U}, \mathbf{P}, \mathbf{T}, \mathbf{F}]$, so dass

$$\mathbb{E}_{\mathcal{S}} \left(\widehat{\text{var}}_{\mathbf{I}} \left(\hat{\psi} \right) \right) = \frac{1}{c_d} \mathbb{E}_{\mathcal{S}}(\hat{\sigma}^2)$$

wobei

$$\begin{aligned} \mathbb{E}_{\mathcal{S}}(\hat{\sigma}^2) &= \frac{1}{f_E} \left(\text{tr} \left(\boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}, \mathbf{F}, \mathbf{T}]}^{\perp} \boldsymbol{\Sigma} \right) + \mathbb{E}_{\mathcal{S}}(\mathbf{y})^T \boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}, \mathbf{F}, \mathbf{T}]}^{\perp} \mathbb{E}_{\mathcal{S}}(\mathbf{y}) \right) \\ &= \frac{1}{f_E} \text{tr} \left(\boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}, \mathbf{F}, \mathbf{T}]}^{\perp} \boldsymbol{\Sigma} \right). \end{aligned} \quad (4.7)$$

Wir wollen nun $\text{tr} \left(\boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}, \mathbf{F}, \mathbf{T}]}^{\perp} \boldsymbol{\Sigma} \right)$ herleiten. Dazu wenden wir die Projektorzerlegungsformel aus Lemma 3.1 an. Lemma 3.2 impliziert, dass

$$\begin{aligned} \boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}, \mathbf{F}]}^{\perp} &= \boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}]}^{\perp} - \boldsymbol{\omega}_{\boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}]}^{\perp} \mathbf{F}} \\ &= \mathbf{Q}_n \otimes \mathbf{Q}_p - (\mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{F} (\mathbf{F}^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{F})^{-1} \mathbf{F}^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \\ &= \mathbf{Q}_n \otimes \mathbf{Q}_p - \frac{1}{c_{d22}} (\mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{F} \mathbf{F}^T (\mathbf{Q}_n \otimes \mathbf{Q}_p). \end{aligned} \quad (4.8)$$

Man beachte, dass $\mathbf{1}_t^T \mathbf{T} = \mathbf{1}_{np}^T$ und dass $\boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}, \mathbf{F}]}^{\perp} \mathbf{1}_{np} = 0$. Daraus folgt

$$\begin{aligned} \boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}, \mathbf{F}, \mathbf{T}]}^{\perp} &= \boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}, \mathbf{F}]}^{\perp} - \boldsymbol{\omega}_{\boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}, \mathbf{F}]}^{\perp} \mathbf{T}} \\ &= \boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}, \mathbf{F}]}^{\perp} - \boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}, \mathbf{F}]}^{\perp} \mathbf{T} (\mathbf{T}^T \boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}, \mathbf{F}]}^{\perp} \mathbf{T})^{-1} \mathbf{T}^T \boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}, \mathbf{F}]}^{\perp} \\ &= \boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}, \mathbf{F}]}^{\perp} - \frac{1}{c_d} \boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}, \mathbf{F}]}^{\perp} \mathbf{T} \mathbf{T}^T \boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}, \mathbf{F}]}^{\perp}. \end{aligned}$$

Nun ist

$$\begin{aligned}
\text{tr}(\boldsymbol{\omega}_{[U,P,F]}^\perp \boldsymbol{\Sigma}) &= \text{tr}((\mathbf{Q}_n \otimes \mathbf{Q}_p) \boldsymbol{\Sigma}) - \frac{1}{c_{d22}} \text{tr}((\mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{F} \mathbf{F}^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) \boldsymbol{\Sigma}) \\
&= \text{tr}((\mathbf{Q}_n \otimes \mathbf{Q}_p) (\mathbf{I}_n \otimes \mathbf{S}) (\mathbf{Q}_n \otimes \mathbf{Q}_p)) \\
&\quad - \frac{1}{c_{d22}} \text{tr}(\mathbf{F}^T (\mathbf{Q}_n \otimes \mathbf{Q}_p) (\mathbf{I}_n \otimes \mathbf{S}) (\mathbf{Q}_n \otimes \mathbf{Q}_p) \mathbf{F}) \\
&= \text{tr}(\mathbf{Q}_n \otimes \tilde{\mathbf{S}}) - \frac{1}{c_{d22}} \text{tr}(\mathbf{F}^T (\mathbf{Q}_n \otimes \tilde{\mathbf{S}}) \mathbf{F}).
\end{aligned}$$

Für vollständig balancierte Pläne ist außerdem $\sum_{i=1}^n \mathbf{T}_i = \frac{n}{t} \mathbf{1}_p \mathbf{1}_t^T$. Damit ergibt sich

$$\begin{aligned}
\mathbf{F}^T (\mathbf{Q}_n \otimes \tilde{\mathbf{S}}) \mathbf{F} &= [\mathbf{T}_1^T \mathbf{V}^T, \dots, \mathbf{T}_n^T \mathbf{V}^T] (\mathbf{Q}_n \otimes \tilde{\mathbf{S}}) [\mathbf{T}_1^T \mathbf{V}^T, \dots, \mathbf{T}_n^T \mathbf{V}^T]^T \\
&= [\mathbf{T}_1^T \mathbf{V}^T \tilde{\mathbf{S}}, \dots, \mathbf{T}_n^T \mathbf{V}^T \tilde{\mathbf{S}}] [\mathbf{T}_1^T \mathbf{V}^T, \dots, \mathbf{T}_n^T \mathbf{V}^T]^T \\
&\quad - \frac{1}{n} \mathbf{1}_n^T \otimes \sum_{i=1}^n \mathbf{T}_i^T \mathbf{V}^T \tilde{\mathbf{S}} [\mathbf{T}_1^T \mathbf{V}^T, \dots, \mathbf{T}_n^T \mathbf{V}^T]^T \\
&= \sum_{i=1}^n (\mathbf{T}_i^T \mathbf{V}^T \tilde{\mathbf{S}} \mathbf{V} \mathbf{T}_i) - \frac{1}{n} \left(\frac{n}{t} \mathbf{1}_t \mathbf{1}_p^T \mathbf{V}^T \tilde{\mathbf{S}} \mathbf{V} \right) \frac{n}{t} \mathbf{1}_p \mathbf{1}_t^T
\end{aligned}$$

und wegen $\mathbf{T}_i \mathbf{T}_i^T = \mathbf{I}_p$

$$\begin{aligned}
\text{tr}(\mathbf{F}^T (\mathbf{Q}_n \otimes \tilde{\mathbf{S}}) \mathbf{F}) &= \text{tr}(n \mathbf{V}^T \tilde{\mathbf{S}} \mathbf{V} - \frac{n}{t} \mathbf{1}_p \mathbf{1}_p^T \mathbf{V}^T \tilde{\mathbf{S}} \mathbf{V}) \\
&= n \text{tr}(\mathbf{V} \mathbf{Q}_{p,t} \mathbf{V}^T \tilde{\mathbf{S}}).
\end{aligned}$$

Folglich erhalten wir

$$\text{tr}(\boldsymbol{\omega}_{[U,P,F]}^\perp \boldsymbol{\Sigma}) = (n-1) \text{tr}(\tilde{\mathbf{S}}) - \frac{n}{c_{d22}} \text{tr}(\mathbf{V} \mathbf{Q}_{p,t} \mathbf{V}^T \tilde{\mathbf{S}}). \quad (4.9)$$

Analog rechnet man mit Hilfe von (4.8) leicht nach, dass

$$\begin{aligned}
\text{tr}(\boldsymbol{\omega}_{[U,P,F]}^\perp \mathbf{T} \mathbf{T}^T \boldsymbol{\omega}_{[U,P,F]}^\perp \boldsymbol{\Sigma}) &= \text{tr}(\mathbf{T}^T \boldsymbol{\omega}_{[U,P,F]}^\perp \boldsymbol{\Sigma} \boldsymbol{\omega}_{[U,P,F]}^\perp \mathbf{T}) \\
&= \text{tr}((\mathbf{T}^T - \gamma \mathbf{Q}_t \mathbf{F}^T) \\
&\quad \cdot (\mathbf{Q}_n \otimes \mathbf{Q}_p) \boldsymbol{\Sigma} (\mathbf{Q}_n \otimes \mathbf{Q}_p) (\mathbf{T} - \gamma \mathbf{F} \mathbf{Q}_t))
\end{aligned}$$

$$\begin{aligned}
&= n \operatorname{tr}(\tilde{\mathbf{S}}) \\
&\quad - 2n\gamma \operatorname{tr}(\mathbf{Q}_{p,t} \mathbf{V}^T \tilde{\mathbf{S}}) + n\gamma^2 \operatorname{tr}(\mathbf{Q}_{p,t} \mathbf{V}^T \tilde{\mathbf{S}} \mathbf{V}).
\end{aligned}$$

Zusammenfassend sieht man, dass

$$\begin{aligned}
\operatorname{tr}(\boldsymbol{\omega}_{[\mathbf{U}, \mathbf{P}, \mathbf{F}, \mathbf{T}]}^\perp \boldsymbol{\Sigma}) &= \operatorname{tr} \left(\left((n-1) \mathbf{I}_p - \frac{n}{c_{d22}} (\mathbf{V} \mathbf{Q}_{p,t} \mathbf{V}^T) \right) \tilde{\mathbf{S}} \right) \\
&\quad - \frac{n}{c_d} \operatorname{tr} \left((\mathbf{I}_p - 2\gamma \mathbf{Q}_{p,t} \mathbf{V}^T + \gamma^2 \mathbf{V} \mathbf{Q}_{p,t} \mathbf{V}^T) \tilde{\mathbf{S}} \right) \\
&= \frac{n}{c_d} \operatorname{tr}(\mathbf{B}_d^* \tilde{\mathbf{S}}), \tag{4.10}
\end{aligned}$$

wobei

$$\mathbf{B}_d^* = \left(\frac{c_d(n-1)}{n} - 1 \right) \mathbf{I}_p + \gamma (\mathbf{V} + \mathbf{V}^T) - \left(\frac{c_d}{c_{d22}} + \gamma^2 \right) \mathbf{V} \mathbf{Q}_{p,t} \mathbf{V}^T.$$

Durch einsetzen von c_d , c_{d12} und c_{d22} aus Lemma 4.1 sieht man, dass $p\gamma = -c_d/c_{d22} - \gamma^2$ und damit $\mathbf{B}_d^* = \mathbf{B}_d$. Mit (4.7) folgt schließlich die Behauptung. \square

Mit Lemma 4.2 und 4.3 ergibt sich eine neue Darstellung für den Varianzquotienten.

Satz 4.2. *Der Varianzquotient (4.3) ist gegeben durch*

$$k_{t,n,p}(\mathbf{S}) = \frac{f_E}{t-1} \frac{\operatorname{tr}(\mathbf{Q}_p \mathbf{A}_d \mathbf{Q}_p \mathbf{S})}{\operatorname{tr}(\mathbf{Q}_p \mathbf{B}_d \mathbf{Q}_p \mathbf{S})}, \tag{4.11}$$

wobei

$$\mathbf{A}_d = \mathbf{I}_p - \gamma (\mathbf{V} + \mathbf{V}^T) + \gamma^2 \mathbf{V} \mathbf{Q}_{p,t} \mathbf{V}^T$$

und

$$\mathbf{B}_d = \left(\frac{c_d(n-1)}{n} - 1 \right) \mathbf{I}_p + \gamma (\mathbf{V} + \mathbf{V}^T) + p\gamma \mathbf{V} \mathbf{Q}_{p,t} \mathbf{V}^T.$$

Diesen Ausdruck können wir mit Hilfe der Spektralzerlegung von \mathbf{S} geschickt umformen. Seien λ_i , $i = 1, \dots, p$, die Eigenwerte von \mathbf{S} und \mathbf{s}_i die zugehörigen Eigenvektoren, dann gilt $\mathbf{S} = \sum_{i=1}^p \lambda_i \mathbf{s}_i \mathbf{s}_i^T$ (Seber, 2008)[16.44(a)]. Damit erhalten wir

$$\begin{aligned}
k_{t,n,p}(\mathbf{S}) &= \frac{f_E}{t-1} \frac{\operatorname{tr}(\mathbf{Q}_p \mathbf{A}_d \mathbf{Q}_p \sum_{i=1}^p \lambda_i \mathbf{s}_i \mathbf{s}_i^T)}{\operatorname{tr}(\mathbf{Q}_p \mathbf{B}_d \mathbf{Q}_p \sum_{i=1}^p \lambda_i \mathbf{s}_i \mathbf{s}_i^T)} \\
&= \frac{f_E}{t-1} \frac{\sum_{i=1}^p \lambda_i \mathbf{s}_i^T \mathbf{Q}_p \mathbf{A}_d \mathbf{Q}_p \mathbf{s}_i}{\sum_{i=1}^p \lambda_i \mathbf{s}_i^T \mathbf{Q}_p \mathbf{B}_d \mathbf{Q}_p \mathbf{s}_i}.
\end{aligned}$$

Wir maximieren $k_{t,n,p}(\mathbf{S})$, indem wir eine ungünstigste Kovarianzmatrix \mathbf{S} wählen, die diesen Ausdruck maximiert. Das Problem ließe sich mit Hilfe des Satzes von Rayleigh-Ritz, vergleiche Horn und Johnson (1985)[S. 176], unter zwei Bedingungen lösen. Zum einen dürfte hier statt eines Quotienten von Summen nur ein Summand im Zähler und Nenner stehen. Zum anderen müsste $\mathbf{Q}_p \mathbf{B}_d \mathbf{Q}_p$ p.d. sein, was offensichtlich nicht der Fall ist. Das nächste Lemma bietet eine Möglichkeit, das Problem zu lösen.

4.2.2 Lösung des Maximierungsproblems

Lemma 4.4. *Seien $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$ und \mathbf{S} n.n.d., wobei $\tilde{\mathbf{A}}\mathbf{1}_p = \tilde{\mathbf{B}}\mathbf{1}_p = \mathbf{0}$ und $\text{Rg}(\tilde{\mathbf{B}}) = p - 1$. Dann gilt*

$$\frac{\text{tr}(\tilde{\mathbf{A}}\mathbf{S})}{\text{tr}(\tilde{\mathbf{B}}\mathbf{S})} \leq \lambda^*.$$

Dabei ist λ^ der größte Eigenwert von $\tilde{\mathbf{B}}^+ \tilde{\mathbf{A}}$ und $\tilde{\mathbf{B}}^+$ die Moore-Penrose-Inverse von $\tilde{\mathbf{B}}$. Sei \mathbf{v} ein Eigenvektor von $\tilde{\mathbf{B}}^+ \tilde{\mathbf{A}}$ zum Eigenwert λ^* . Dann gilt für $\mathbf{S} = \mathbf{v}\mathbf{v}^T$ Gleichheit.*

Beweis. Sei $\tilde{\mathbf{U}} = [\mathbf{u}_1, \dots, \mathbf{u}_p]$ mit $\mathbf{u}_p = \mathbf{1}_p/\sqrt{p}$. Dann haben wir die Spektralzerlegung $\tilde{\mathbf{U}}^T \tilde{\mathbf{B}} \tilde{\mathbf{U}} = \mathbf{D}$, wobei \mathbf{D} eine Diagonalmatrix mit Diagonalelementen $\gamma_1, \dots, \gamma_p$, wobei $\gamma_1 \geq \dots \geq \gamma_{p-1} > 0$ und $\gamma_p = 0$ die Eigenwerte von $\tilde{\mathbf{B}}$ sind. Dabei ist

$$\tilde{\mathbf{U}}^T \tilde{\mathbf{U}} = \tilde{\mathbf{U}} \tilde{\mathbf{U}}^T = \mathbf{I}_p.$$

Definiert man

$$\mathbf{W} = \begin{bmatrix} \frac{1}{\sqrt{\gamma_1}} & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \frac{1}{\sqrt{\gamma_{p-1}}} & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix} \tilde{\mathbf{U}}^T,$$

dann ist \mathbf{W} invertierbar, die letzte Zeile von \mathbf{W} ist gleich $\mathbf{1}_p^T/\sqrt{p}$ und es gilt

$$\mathbf{W} \tilde{\mathbf{B}} \mathbf{W}^T = \begin{bmatrix} \mathbf{I}_{p-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}.$$

Wir suchen

$$\lambda^* = \max_{\mathbf{S} \in \mathcal{S}} \frac{\operatorname{tr}(\tilde{\mathbf{A}}\mathbf{S})}{\operatorname{tr}(\tilde{\mathbf{B}}\mathbf{S})},$$

wobei \mathcal{S} die Menge aller n.n.d. Matrizen ist. Wegen der Invertierbarkeit von \mathbf{W} ist $\mathbf{S}_{\mathbf{W}} = \mathbf{W}^{T^{-1}}\mathbf{S}\mathbf{W}^{-1}$ ebenfalls n.n.d. und zu jeder n.n.d. Matrix $\mathbf{S}_{\mathbf{W}}$ existiert ein \mathbf{S} , so dass $\mathbf{S}_{\mathbf{W}} = \mathbf{W}^{T^{-1}}\mathbf{S}\mathbf{W}^{-1}$, das heißt

$$\mathcal{S} = \left\{ \mathbf{S}_{\mathbf{W}} = \mathbf{W}^{T^{-1}}\mathbf{S}\mathbf{W}^{-1} : \mathbf{S} \in \mathcal{S} \right\}.$$

Damit können wir schreiben

$$\begin{aligned} \operatorname{tr}(\tilde{\mathbf{A}}\mathbf{S}) &= \operatorname{tr}(\tilde{\mathbf{A}}\mathbf{W}^T\mathbf{W}^{T^{-1}}\mathbf{S}\mathbf{W}^{-1}\mathbf{W}) \\ &= \operatorname{tr}(\tilde{\mathbf{A}}\mathbf{W}^T\mathbf{S}_{\mathbf{W}}\mathbf{W}) \end{aligned}$$

und es folgt

$$\lambda^* = \max_{\mathbf{S}_{\mathbf{W}} \in \mathcal{S}} \frac{\operatorname{tr}(\tilde{\mathbf{A}}\mathbf{W}^T\mathbf{S}_{\mathbf{W}}\mathbf{W})}{\operatorname{tr}(\tilde{\mathbf{B}}\mathbf{W}^T\mathbf{S}_{\mathbf{W}}\mathbf{W})}.$$

Nun ist aber

$$\begin{aligned} \operatorname{tr}(\tilde{\mathbf{B}}\mathbf{W}^T\mathbf{S}_{\mathbf{W}}\mathbf{W}) &= \operatorname{tr}(\mathbf{W}\tilde{\mathbf{B}}\mathbf{W}^T\mathbf{S}_{\mathbf{W}}) \\ &= \operatorname{tr}\left(\begin{bmatrix} \mathbf{I}_{p-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} \mathbf{S}_{\mathbf{W}}\right). \end{aligned}$$

Weiterhin ist wegen $\tilde{\mathbf{A}}\mathbf{1}_p = \mathbf{0}$

$$\begin{aligned} \mathbf{W}\tilde{\mathbf{A}}\mathbf{W}^T &= \begin{bmatrix} \mathbf{W}_1\tilde{\mathbf{A}}\mathbf{W}_1^T & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I}_{p-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} \mathbf{W}\tilde{\mathbf{A}}\mathbf{W}^T \begin{bmatrix} \mathbf{I}_{p-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}, \end{aligned}$$

wobei $\mathbf{W}^T = [\mathbf{W}_1^T, \mathbf{1}_p/\sqrt{p}]$. Folglich können wir uns bei der Suche nach dem Maximum auf eine Teilmenge der n.n.d. Matrizen beschränken. Es gilt

$$\lambda^* = \max_{\mathbf{S} \in \mathcal{S}^*} \frac{\text{tr}(\mathbf{W} \tilde{\mathbf{A}} \mathbf{W}^T \mathbf{S})}{\text{tr}(\mathbf{S})},$$

wobei

$$\mathcal{S}^* = \left\{ \mathbf{S} \in \mathbb{R}^{p \times p} : \mathbf{S} = \begin{bmatrix} \mathbf{S}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \text{ und } \mathbf{S}^* \in \mathbb{R}^{(p-1) \times (p-1)} \text{ n.n.d.} \right\}$$

Wir können nun jedes $\mathbf{S} \in \mathcal{S}^*$ über seine Spektralzerlegung darstellen als $\sum_{i=1}^{p-1} \lambda_i \mathbf{x}_i \mathbf{x}_i^T$, wobei $\lambda_i \geq 0$, $\mathbf{x}_i^T \mathbf{x}_i = 1$, $\mathbf{x}_i^T \mathbf{x}_j = 0 \quad \forall i \neq j$. Dabei ist $\mathbf{x}_i = [x_{i1}, \dots, x_{i,p-1}, 0]^T$. Dann haben wir

$$\begin{aligned} \lambda^* &= \max \frac{\text{tr} \left(\sum_{i=1}^p \lambda_i \mathbf{x}_i \mathbf{x}_i^T \mathbf{W} \tilde{\mathbf{A}} \mathbf{W}^T \right)}{\sum_{i=1}^p \lambda_i} \\ &= \max \frac{\lambda_i \mathbf{x}_i^T \mathbf{W} \tilde{\mathbf{A}} \mathbf{W}^T \mathbf{x}_i}{\sum_{i=1}^p \lambda_i}. \end{aligned}$$

Sei nun \mathbf{z}^* ein Eigenvektor zum größten Eigenwert $\tilde{\lambda}$ von $\mathbf{W} \tilde{\mathbf{A}} \mathbf{W}^T$. Dann gilt mit dem Satz von Rayleigh-Ritz, vergleiche Horn und Johnson (1985)[S. 176], dass

$$\forall \mathbf{x}_i : \mathbf{x}_i^T \mathbf{W} \tilde{\mathbf{A}} \mathbf{W}^T \mathbf{x}_i \leq \tilde{\lambda}$$

und daher $\lambda^* \leq \tilde{\lambda}$. Es ist sogar Gleichheit möglich, denn wegen

$$\mathbf{W} \tilde{\mathbf{A}} \mathbf{W}^T = \begin{bmatrix} \mathbf{W}_1 \tilde{\mathbf{A}} \mathbf{W}_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

ist der Eigenvektor \mathbf{z}^* zum größten Eigenwert von $\mathbf{W} \tilde{\mathbf{A}} \mathbf{W}^T$ von der Form $\mathbf{z}^* = [z_1, \dots, z_{p-1}, 0]^T$. Man beachte, dass $\mathbf{W} \tilde{\mathbf{A}} \mathbf{W}^T$ und $\mathbf{W}^T \tilde{\mathbf{W}} \tilde{\mathbf{A}}$ die gleichen Eigenwerte haben. Daher ist $\tilde{\lambda}$ auch der größte Eigenwert von

$$\mathbf{W}^T \mathbf{W} \tilde{\mathbf{A}} = \tilde{\mathbf{U}} \begin{bmatrix} \frac{1}{\gamma_1} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \frac{1}{\gamma_{p-1}} & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} \tilde{\mathbf{U}}^T \tilde{\mathbf{A}}$$

$$\begin{aligned}
&= \tilde{\mathbf{U}} \begin{bmatrix} \frac{1}{\gamma_1} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \frac{1}{\gamma_{p-1}} & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix} \tilde{\mathbf{U}}^T \tilde{\mathbf{A}} \\
&= \tilde{\mathbf{B}}^+ \tilde{\mathbf{A}},
\end{aligned}$$

da $\tilde{\mathbf{U}}^T = [\mathbf{u}_1, \dots, \mathbf{u}_p]^T$ mit $\mathbf{u}_p = \mathbf{1}_p/\sqrt{p}$, also $\mathbf{u}_p^T \tilde{\mathbf{A}} = \mathbf{0}$ und $\tilde{\mathbf{B}}^+ = \tilde{\mathbf{U}}^T \mathbf{D}^+ \tilde{\mathbf{U}}$. Man sieht leicht, dass nun $\mathbf{v} = \mathbf{W}^T \mathbf{z}^*$ ein Eigenvektor von $\tilde{\mathbf{B}}^+ \tilde{\mathbf{A}}$ zum Eigenwert λ^* ist. \square

Wir wollen Lemma 4.4 anwenden, wobei $\tilde{\mathbf{A}}_d = \mathbf{Q}_p \mathbf{A}_d \mathbf{Q}_p$ die Rolle von $\tilde{\mathbf{A}}$ übernimmt und $\tilde{\mathbf{B}}_d = \mathbf{Q}_p \mathbf{B}_d \mathbf{Q}_p$ die Rolle von $\tilde{\mathbf{B}}$. Dazu zeigen wir, dass diese Matrizen die Voraussetzungen des Lemmas erfüllen.

Lemma 4.5. *Die Matrix $\tilde{\mathbf{A}}_d = \mathbf{Q}_p \mathbf{A}_d \mathbf{Q}_p$ aus Lemma 4.2 ist n.n.d. und $\tilde{\mathbf{A}}_d \mathbf{1}_p = \mathbf{0}$.*

Beweis. Falls es eine Matrix \mathbf{L} gibt, so dass $\mathbf{Q}_p \mathbf{A}_d \mathbf{Q}_p = \mathbf{L} \mathbf{L}^T$, dann ist $\mathbf{Q}_p \mathbf{A}_d \mathbf{Q}_p$ n.n.d.. Sei

$$\mathbf{L} = \mathbf{Q}_p (\mathbf{I} - \gamma \mathbf{V}) \left(\mathbf{I} - \left(\frac{1 + \sqrt{1 - p/t}}{p} \right) \mathbf{1}_p \mathbf{1}_p^T \right).$$

Damit folgt sofort, dass $\mathbf{L} \mathbf{L}^T = \mathbf{Q}_p \mathbf{A}_d \mathbf{Q}_p$. Wegen $\mathbf{Q}_p = \omega_{\mathbf{1}_p}^\perp$ folgt $\tilde{\mathbf{A}}_d \mathbf{1}_p = \mathbf{0}$. \square

Man kann zeigen, dass $\tilde{\mathbf{B}}_d = [\mathbf{I}_p, \mathbf{0}] \omega_{[\mathbf{U}, \mathbf{P}, \mathbf{F}, \mathbf{T}]}^\perp [\mathbf{I}_p, \mathbf{0}]^T$. Daraus folgt sofort, dass $\tilde{\mathbf{B}}_d$ n.n.d. Allerdings benötigen wir für Lemma 4.4 zusätzlich, dass $\text{Rg}(\tilde{\mathbf{B}}_d) = p - 1$. Wir zeigen dies über einen Umweg und führen zunächst den Begriff der *diagonal dominanten* Matrix ein.

Definition 4.3. *Eine Matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ heißt stark diagonal dominant, falls*

$$a_{ii} > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \quad \forall i = 1, \dots, n.$$

Lemma 4.6. *Sei $\mathbf{A} \in \mathbb{R}^{n \times n}$ symmetrisch und stark diagonal dominant. Dann ist \mathbf{A} p.d.*

Zum Beweis siehe zum Beispiel Horn und Johnson (1985)[S. 349].

Lemma 4.7. *Die Matrix \mathbf{B}_d aus Lemma 4.3 ist für alle $d \in \Lambda_{t,n,p}^*$ p.d.*

Beweis. Wegen

$$\mathbf{V}\mathbf{Q}_{p,t}\mathbf{V}^T = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & \frac{t-1}{t} & \frac{-1}{t} & \cdots & \frac{-1}{t} \\ 0 & \frac{-1}{t} & \ddots & \ddots & \vdots \\ 0 & \vdots & \ddots & \ddots & \frac{-1}{t} \\ 0 & \frac{-1}{t} & \cdots & \frac{-1}{t} & \frac{t-1}{t} \end{bmatrix}$$

ist

$$\mathbf{B}_d = \begin{bmatrix} b_{11} & \gamma & 0 & 0 & \cdots & 0 \\ \gamma & b_{22} & (1 - \frac{p}{t})\gamma & \frac{-p\gamma}{t} & \cdots & \frac{-p\gamma}{t} \\ 0 & (1 - \frac{p}{t})\gamma & \ddots & \ddots & \ddots & \vdots \\ 0 & \frac{-p\gamma}{t} & \ddots & \ddots & \ddots & \frac{-p\gamma}{t} \\ \vdots & \vdots & \ddots & \ddots & \ddots & (1 - \frac{p}{t})\gamma \\ 0 & \frac{-p\gamma}{t} & \cdots & \frac{-p\gamma}{t} & (1 - \frac{p}{t})\gamma & b_{22} \end{bmatrix}.$$

Hierbei sind

$$b_{11} = \frac{c_d(n-1)}{n} - 1$$

und
$$b_{22} = \frac{c_d(n-1)}{n} - 1 + p\gamma \frac{t-1}{t},$$

wobei $\gamma = -t/(pt - t - 1)$, vergleiche (4.2).

Dies führt zu den folgenden Ungleichungen. \mathbf{B}_d ist stark diagonal dominant, falls

- (a). $b_{11} > |\gamma|$, vergleiche die erste Zeile von \mathbf{B}_d .
- (b). $b_{22} > |\gamma| + |(1 - \frac{p}{t})\gamma| + (p-3)|\frac{-p\gamma}{t}|$, vergleiche die zweite Zeile von \mathbf{B}_d .
- (c). $b_{22} > 2|(1 - \frac{p}{t})\gamma| + (p-4)|\frac{-p\gamma}{t}|$, vergleiche die dritte bis $(p-1)$ -te Zeile von \mathbf{B}_d (falls $p > 3$).
- (d). $b_{22} > |(1 - \frac{p}{t})\gamma| + (p-3)|\frac{-p\gamma}{t}|$, vergleiche die letzte Zeile von \mathbf{B}_d .

Da $|\gamma| > 0$ ist (d) erfüllt, falls (b) gilt. Wegen

$$\begin{aligned} |\gamma| + \left| \frac{-p\gamma}{t} \right| &= -(1 + \frac{p}{t})\gamma \\ &< -(1 - \frac{p}{t})\gamma \\ &= |(1 - \frac{p}{t})\gamma| \end{aligned}$$

ist (c) erfüllt falls (b) gilt. Schließlich ist wegen $b_{22} < b_{11}$ mit (b) insbesondere auch (a) erfüllt. Es genügt also für jedes $p > 3$ die obige Bedingung (b) zu zeigen.

Nun gilt

$$\begin{aligned} |\gamma| + \left| \left(1 - \frac{p}{t}\right)\gamma \right| + (p-3) \left| \frac{-p\gamma}{t} \right| &= -\gamma - \left(1 - \frac{p}{t}\right)\gamma - (p-3) \frac{p\gamma}{t} \\ &= - \left(2 + (p-4) \frac{p}{t}\right) \gamma. \end{aligned}$$

Für alle vollständig balancierten Blockpläne gibt es ein $m \in \mathbb{N}$, so dass $n = mt$ und da $t \geq p$ können wir t durch $p + r$ ersetzen, wobei $r \in \mathbb{N} \cup \{0\}$. Setzen wir jetzt noch γ und c_d ein, so können wir (b) durch die Designparameter m , p und r ausdrücken. Durch ausmultiplizieren und zusammenfassen der Potenzen in p sieht man leicht, dass (b) wahr ist, falls $f : \mathbb{Z}^3 \rightarrow \mathbb{R}$ mit

$$\begin{aligned} f(m, p, r) &= mp^5 + (2mr - 2m - 4)p^4 + (mr^2 - 4mr - 6r - m + 9)p^3 \\ &\quad + (-2mr^2 - 2r^2 - mr + 7r + 2m - 2)p^2 + (-r^2 + 3mr + 2r - 3)p \\ &\quad + mr^2 - r \end{aligned}$$

nur positive Werte annimmt. Betrachte nun die partielle Ableitung von f nach m ,

$$\begin{aligned} \frac{\partial f}{\partial m}(m, p, r) &= p^5 + (2r - 2)p^4 + (r^2 - 4r - 1)p^3 + (-2r^2 - r + 2)p^2 + 3rp + r^2 \\ &= p^2 \left((p-2)r^2 + (2p^2 - 4p - 1)r + (p-2)(p^2 - 1) \right) + 3rp + r^2. \end{aligned}$$

Da $p \geq 3$ gilt

$$\begin{aligned} \frac{\partial f}{\partial m}(m, p, r) &\geq p^2 \left(r^2 + (2p-1)r + (p^2 - 1) \right) + 3rp + r^2 \\ &> 0, \end{aligned}$$

Daher ist mit $f(m_0, p, r) > 0$ auch $f(m_1, p, r) > 0$ für alle $m_1 > m_0$ gezeigt.

Wir führen jetzt eine Fallunterscheidung durch. Falls möglich wenden wir Lemma 4.6 an, um zu zeigen, dass \mathbf{B}_d p.d. ist. Falls dies nicht möglich ist, zeigen wir entweder anderweitig, dass \mathbf{B}_d p.d. ist oder wir zeigen, dass es im betrachteten Falls kein $d \in \Lambda_{t,n,p}^*$ gibt, so dass der Fall für die Aussage von Lemma 4.7 irrelevant ist.

- (i). $m = 1, p = 3$ und $r \geq 0$.
- (ii). $m \geq 2, p = 3$ und $r \geq 0$.
- (iii). $m = 1, p = 4$ und $r \geq 0$.
- (iv). $m \geq 2, p = 4$ und $r \geq 0$.
- (v). $m \geq 1, p \geq 5$ und $r \geq 0$.

(i). Eine notwendige Bedingung für die Existenz eines nachbarbalancierten Plans ist, dass $(n-1)p/(t(t-1)) = 3/t$ eine natürliche Zahl ist. Dies gilt hier nur für $t = 3$, aber es gibt kein vollständig balanciertes Design für $n = t = p = 3$. Der Fall ist daher irrelevant für die Behauptung.

(ii). Hier ist beispielsweise $f(2, 3, r) = -r^2 + 14r + 36 < 0$ für $r = 17$, so dass wir Lemma 4.6 nicht direkt anwenden können. Wenn aber

$$\mathbf{R}\mathbf{B}_d\mathbf{R} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix} \mathbf{B}_d \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

p.d. ist, so auch \mathbf{B}_d selbst, da die obige Diagonalmatrix \mathbf{R} Vollrang hat, vergleiche hierzu Graybill (1969)[S. 334]. Man rechnet leicht nach, dass $\mathbf{R}\mathbf{B}_d\mathbf{R}$ stark diagonal dominant ist und wendet Lemma 4.6 auf $\mathbf{R}\mathbf{B}_d\mathbf{R}$ an.

(iii). Es ist $f(1, 4, r) = -3r^2 - 13r - 12 < 0$ für alle $r \geq 0$. Eine notwendige Bedingung für die Existenz eines nachbarbalancierten Designs ist, dass $(n-1)p/(t(t-1)) = 4/t$ eine natürliche Zahl ist. Dies gilt hier nur für $t = 4$, das heißt für $r = 0$. Man berechnet die Eigenwerte von \mathbf{B}_d für $n = t = p = 4$ numerisch und stellt fest, dass diese alle positiv sind und daher \mathbf{B}_d p.d. Jede andere Wahl von r ist irrelevant für die Behauptung, da dann kein vollständig balanciertes Design existiert.

(iv). Hier gilt $f(2, 4, r) = 30r^2 + 239r + 468 > 0$. Daraus folgt sofort $f(m, 4, r) \geq f(2, 4, r) > 0$.

(v). Man betrachte die Unterfälle $m \geq 1, p = 5$ und $r = 0$, $m \geq 1, p \geq 6$ und $r = 0$, $m \geq 1, p \geq 5$ und $r = 1$, $m \geq 1, p \geq 5$ und $r = 2$ sowie $m \geq 1, p \geq 5$ und $r \geq 3$. Mit Hilfe einfacher Algebra zeigt man jeweils, dass $f > 0$ ist und wendet Lemma 4.6 an.

□

Lemma 4.8. *Sei $d \in \Lambda_{t,n,p}^*$. Dann ist die Matrix $\tilde{\mathbf{B}}_d = \mathbf{Q}_p \mathbf{B}_d \mathbf{Q}_p$ n.n.d. mit Rang $p - 1$ und es gilt $\tilde{\mathbf{B}}_d \mathbf{1}_p = \mathbf{0}$.*

Beweis. $\tilde{\mathbf{B}}_d$ ist n.n.d. da \mathbf{Q}_p n.n.d. ist und \mathbf{B}_d ist p.d. nach Lemma 4.7. Es gilt also $\text{Rg}(\mathbf{B}_d) = p$. Weiterhin ist $\text{Rg}(\mathbf{Q}_p) = \text{tr}(\mathbf{Q}_p) = p - 1$. Bekanntlich gilt für zwei Matrizen $\mathbf{A}_1, \mathbf{A}_2$, dass $\text{Rg}(\mathbf{A}_1 \mathbf{A}_2) \leq \text{Rg}(\mathbf{A}_2)$ (Seber, 2008)[3.12]. Da \mathbf{B}_d^{-1} existiert, ist dann

$$\text{Rg}(\mathbf{Q}_p) = \text{Rg}(\mathbf{B}_d^{-1} \mathbf{B}_d \mathbf{Q}_p) \leq \text{Rg}(\mathbf{B}_d \mathbf{Q}_p) \leq \text{Rg}(\mathbf{Q}_p),$$

und damit $\text{Rg}(\mathbf{B}_d \mathbf{Q}_p) = \text{Rg}(\mathbf{Q}_p) = p - 1$. Andererseits existiert ein \mathbf{L} , so dass $\mathbf{B}_d = \mathbf{L} \mathbf{L}^T$ da \mathbf{B}_d p.d. ist. Man beachte, dass $\text{Rg}(\mathbf{A}^T \mathbf{A}) = \text{Rg}(\mathbf{A})$ (Seber, 2008)[3.3(b)]. Dies impliziert

$$\text{Rg}(\mathbf{B}_d \mathbf{Q}_p) \leq \text{Rg}(\mathbf{L}^T \mathbf{Q}_p) = \text{Rg}\left((\mathbf{L}^T \mathbf{Q}_p)^T \mathbf{L}^T \mathbf{Q}_p\right) = \text{Rg}(\mathbf{Q}_p \mathbf{B}_d \mathbf{Q}_p) \leq \text{Rg}(\mathbf{B}_d \mathbf{Q}_p)$$

und somit $\text{Rg}(\tilde{\mathbf{B}}_d) = \text{Rg}(\mathbf{Q}_p \mathbf{B}_d \mathbf{Q}_p) = p - 1$. Wegen $\tilde{\mathbf{B}}_d \mathbf{1}_p = \mathbf{Q}_p \mathbf{B}_d \mathbf{Q}_p \mathbf{1}_p = \mathbf{0}$ ist die Behauptung bewiesen. □

Damit können wir nun die Hauptaussage dieses Kapitels formulieren.

Satz 4.3. *Für $(t, n, p) \in \mathbb{N}^3$ seien*

$$\tilde{\mathbf{A}}_d = \mathbf{Q}_p (\mathbf{I} - \gamma (\mathbf{V} + \mathbf{V}^T) + \gamma^2 \mathbf{V} \mathbf{Q}_{p,t} \mathbf{V}^T) \mathbf{Q}_p,$$

$$\tilde{\mathbf{B}}_d = \mathbf{Q}_p \left(\left(\frac{c_d(n-1)}{n} - 1 \right) \mathbf{I}_p + \gamma (\mathbf{V} + \mathbf{V}^T) + p\gamma \mathbf{V} \mathbf{Q}_{p,t} \mathbf{V}^T \right) \mathbf{Q}_p$$

und \mathbf{v} ein Eigenvektor zum größten Eigenwert λ^* von $\tilde{\mathbf{B}}_d^+ \tilde{\mathbf{A}}_d$. Sei \mathbf{S} eine beliebige Kovarianzmatrix in Modell (2.6). Falls dann zu (t, n, p) ein $d \in \Lambda_{t,n,p}^*$ existiert, so gilt

$$k_{t,n,p}(\mathbf{S}) \leq \frac{f_E}{t-1} \lambda^* \quad \text{für alle } \mathbf{S}$$

und

$$k_{t,n,p}(\mathbf{S}) = \frac{f_E}{t-1} \lambda^* \quad \text{für } \mathbf{S} = \mathbf{v} \mathbf{v}^T.$$

In diesem Fall ist die Konstante von Kunert und Utzig gegeben durch

$$k_{t,n,p}^* = \frac{f_E}{t-1} \lambda^*.$$

Beweis. Lemma 4.5 und 4.8 zeigen, dass die Voraussetzungen von Lemma 4.4 erfüllt sind. Die Aussage ist bewiesen, wenn wir Lemma 4.4 auf die Darstellung des Varianzquotienten (4.11) in Satz 4.2 anwenden. \square

Mit den Methoden aus Kapitel 4.2.2 kann man auch eine exakte untere Schranke für den Varianzquotienten $k_{t,n,p}$ und eine dazugehörige günstigste Kovarianzmatrix bestimmen. Sei $\mathbf{v}^{(p-1)}$ ein Eigenvektor zum zweitkleinsten Eigenwert $\lambda^{(p-1)}$ von $\tilde{\mathbf{B}}_d^+ \tilde{\mathbf{A}}_d$. Man sieht leicht, dass dann

$$k_{t,n,p}(\mathbf{S}) \geq \frac{f_E}{t-1} \lambda^{(p-1)} \quad \text{für alle } \mathbf{S}$$

und

$$k_{t,n,p}(\mathbf{S}) = \frac{f_E}{t-1} \lambda^{(p-1)} \quad \text{für } \mathbf{S} = \mathbf{v}^{(p-1)} \mathbf{v}^{(p-1)T}.$$

Wir nutzen dieses Ergebnis in der Simulationsstudie in Kapitel 6.2. Für die praktische Anwendung ist die Minimierung des Varianzquotienten jedoch von geringerem Interesse als die Maximierung, da wir bei unbekannter Kovarianzmatrix konservative Tests konstruieren möchten und keine anti-konservativen. Von größerem Interesse für die Praxis ist stattdessen die Berechnung der Konstanten von Kunert und Utzig für große Stichproben.

Korollar 4.1. *Sei unter den Voraussetzungen von Satz 4.3 $\lambda_{\tilde{\mathbf{A}}_d}^*$ der größte Eigenwert von $\tilde{\mathbf{A}}_d$, $\mathbf{S} = \mathbf{v}\mathbf{v}^T$ ungünstigste Kovarianzmatrix und $k_{t,\infty,p}(\mathbf{S}) = \lim_{n \rightarrow \infty} k_{t,n,p}(\mathbf{S})$. Dann gilt*

$$k_{t,\infty,p}(\mathbf{S}) = \frac{\lambda_{\tilde{\mathbf{A}}_d}^*}{1 + \frac{\gamma}{p}}.$$

Beweis. Wenn λ^* der größte Eigenwert von $\tilde{\mathbf{B}}_d^+ \tilde{\mathbf{A}}_d$ ist, dann ist $k_{t,n,p} = f_E \lambda^* / (t-1)$ der größte Eigenwert von $((t-1)\tilde{\mathbf{B}}_d/f_E)^+ \tilde{\mathbf{A}}_d$. Nun hängt $\tilde{\mathbf{A}}_d$ nicht von n ab. Wegen $f_E = n(p-1) - 2t - p + 3$ und $\gamma = -t/(pt - t - 1)$ folgt

$$\begin{aligned} (t-1) \lim_{n \rightarrow \infty} \frac{c_d}{f_E} &= (t-1) \lim_{n \rightarrow \infty} \frac{n(p-1)}{n(p-1) - 2t - p + 3} \frac{1 - \frac{t}{p(pt-t-1)}}{t-1} \\ &= 1 + \frac{\gamma}{p}. \end{aligned}$$

Weiterhin lässt sich $(t-1)\tilde{\mathbf{B}}_d/f_E$ schreiben als

$$\mathbf{Q}_p \left(\left((t-1) \frac{c_d}{f_E} \frac{n-1}{n} - \frac{t-1}{f_E} \right) \mathbf{I}_p + \frac{(t-1)\gamma}{f_E} (\mathbf{V} + \mathbf{V}^T + p\mathbf{V}\mathbf{Q}_{p,t}\mathbf{V}^T) \right) \mathbf{Q}_p$$

Man rechnet leicht nach, dass dann $(t-1)\tilde{\mathbf{B}}_d/f_E$ für $n \rightarrow \infty$ gegen $(1+\gamma/p)\mathbf{Q}_p$ konvergiert. Da $\mathbf{Q}_p\tilde{\mathbf{A}}_d = \tilde{\mathbf{A}}_d$ erhält man somit

$$\lim_{n \rightarrow \infty} \left(\frac{t-1}{f_E} \tilde{\mathbf{B}}_d \right)^+ \tilde{\mathbf{A}}_d = \frac{\tilde{\mathbf{A}}_d}{\left(1 + \frac{\gamma}{p}\right)}. \quad \square$$

Wenn zusätzlich zu $n \rightarrow \infty$ auch $t \rightarrow \infty$, so sieht man leicht, dass $\gamma \rightarrow -1/(p-1)$ und $\mathbf{Q}_{p,t} \rightarrow \mathbf{I}_p$. Das hat zur Folge, dass

$$\lim_{t \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{\tilde{\mathbf{A}}_d}{1 + \frac{\gamma}{p}} = p_1 \mathbf{Q}_p \left(\mathbf{I}_p + \frac{1}{p-1} (\mathbf{V} + \mathbf{V}^T) + \frac{1}{(p-1)^2} \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-1} \end{bmatrix} \right) \mathbf{Q}_p,$$

wobei $p_1 = p(p-1)/(p(p-1)+1)$. Damit ist $\lim_{t \rightarrow \infty} \lim_{n \rightarrow \infty} \tilde{\mathbf{A}}_d/(1+\gamma/p)$ für große p näherungsweise gleich \mathbf{Q}_p und folglich

$$\lim_{p \rightarrow \infty} \lim_{t \rightarrow \infty} \lim_{n \rightarrow \infty} k_{t,n,p}^* = 1.$$

In Tabelle 4.2 sind einige Werte von $k_{t,n,p}^*$ vertafelt. Man beachte, dass nicht für alle Werte von (t, n, p) in der Tabelle ein vollständig balancierter Plan existiert. In Anhang B.1 findet man das R-Programm `k.star`, welches zu gegebenem t, n, p die Konstante von Kunert und Utzig berechnet.

Man erkennt, dass $k_{t,n,p}^*$ in allen drei Parametern fällt. Die Konstante konvergiert mit wachsendem n schnell. Bei sehr großer Periodenanzahl (etwa $p = 100$) liegt die Konstante nahe bei 1. Dann kommt es selbst im ungünstigsten Fall nur zu einer minimalen mittleren Varianzunterschätzung. Allerdings dürften Versuchspläne mit derart großer Periodenanzahl in der Praxis kaum vorkommen. Für Anwendungen in der pharmazeutischen Forschung sind mehr als vier Behandlungen in vier Perioden selten. Dort liegt der Korrekturfaktor in der Größenordnung von etwa 1.4 - 1.6. Studien in der Lebensmitteltechnologie können durchaus zehn Behandlungen in zehn Perioden umfassen. Hier liegt der Korrekturfaktor bei ungünstiger Kovarianzstruktur immerhin noch bei etwa 1.25. Psychometrische Anwendungen haben zum Teil 20-50 Behandlungen mit großer Periodenanzahl. In diesem Fall nimmt $k_{t,n,p}^*$ Werte um 1.1 an.

Ein Vergleich der Werte in Tabelle 4.2 mit denen zur oberen Schranke in Tabelle 4.1 zeigt, dass insbesondere bei Plänen mit vier bis sechs Perioden $k_{t,n,p}^*$ deutlich kleiner

$p = 3$		t						
m	3	4	5	6	7	10	100	
1								
2	1.80	1.77	1.75	1.74	1.73	1.71	1.68	
4	1.51	1.49	1.47	1.46	1.46	1.45	1.43	
6	1.48	1.45	1.44	1.43	1.43	1.42	1.40	
∞	1.44	1.41	1.40	1.39	1.39	1.38	1.36	
$p = 4$		t						
m	3	4	5	6	7	10	100	
1		5.09	4.51	4.20	4.01	3.71	3.23	
2		1.61	1.61	1.60	1.60	1.59	1.58	
4		1.49	1.48	1.48	1.47	1.47	1.46	
6		1.47	1.46	1.45	1.45	1.44	1.43	
∞		1.44	1.43	1.42	1.42	1.41	1.40	
$p = 5$		t						
m	3	4	5	6	7	10	100	
1			1.96	1.93	1.92	1.89	1.83	
2			1.49	1.49	1.49	1.49	1.48	
4			1.43	1.42	1.42	1.42	1.41	
6			1.41	1.41	1.40	1.40	1.39	
∞			1.39	1.38	1.38	1.38	1.37	
$p = 6$		t						
m	3	4	5	6	7	10	100	
1				1.61	1.60	1.59	1.56	
2				1.41	1.41	1.41	1.40	
4				1.37	1.37	1.36	1.36	
6				1.36	1.35	1.35	1.35	
∞				1.34	1.34	1.33	1.33	
$p = 7$		t						
m	3	4	5	6	7	10	100	
1					1.46	1.45	1.44	
2					1.35	1.35	1.34	
4					1.32	1.32	1.31	
6					1.31	1.31	1.31	
∞					1.30	1.30	1.29	
$p = 10$		t						
m	3	4	5	6	7	10	100	
1						1.27	1.27	
2						1.24	1.23	
4						1.22	1.22	
6						1.22	1.22	
∞						1.21	1.21	
$p = 100$		t						
m	3	4	5	6	7	10	100	
1							1.02	
2							1.02	
4							1.02	
6							1.02	
∞							1.02	

Tabelle 4.2: $f_E \lambda^* / (t - 1)$ aus Satz 4.3 für verschiedene Werte von $(t, n = mt, p)$. Für einen vollständig balancierten Plan zum Vergleich von t Behandlungen in $n = mt$ Blöcken der Länge p ist dies gleich der Konstanten $k_{t,n,p}^*$ von Kunert und Utzig.

ist als $k_{t,n,p}^{oben}$. Für ein nachbarbalanciertes lateinisches Quadrat mit $t = n = p = 4$ ist die obere Schranke von Kunert und Utzig sogar unendlich, während die mittlere Varianzüberschätzung bei ungünstiger Kovarianzmatrix hier durchaus beschränkt ist, da $k_{t,n,p}^* = 5.09$. Dieser extreme Fall ist wegen des geringen Stichprobenumfangs nicht sehr relevant. Bei Versuchsplänen mit $p = 4$ oder $p = 5$ und $m > 1$ würden wir auf Grundlage der oberen Schranke einen um etwa 20-30% höheren Korrekturfaktor auf unsere Varianzschätzung anwenden als nötig. Bei Vorliegen von günstigen Kovarianzstrukturen führt dies zu besonders konservativen Varianzschätzungen. Wie man in Abbildung 4.1 erkennen kann, wirkt sich die Verbesserung der Schranke von Kunert und Utzig bei größerer Periodenanzahl weniger deutlich aus. Hier ist für vollständig balancierte verallgemeinerte lateinische Quadrate mit $m = 2$ beziehungsweise $m \rightarrow \infty$ die relative Erhöhung des Korrekturfaktors bei Verwendung von $k_{t,n,p}^{oben}$ statt $k_{t,n,p}^*$ abgebildet.

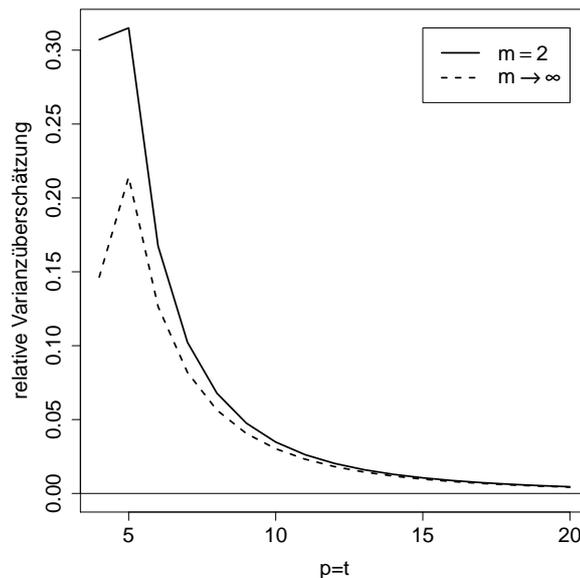


Abbildung 4.1: Relative Varianzüberschätzung $\frac{k_{t,n,p}^{oben}}{k_{t,n,p}^*} - 1$ bei Verwendung von $k_{t,n,p}^{oben}$ statt $k_{t,n,p}^*$. Betrachtet werden Pläne $d \in \Lambda_{t,mt,t}^*$ mit $m = 2$ und $m \rightarrow \infty$.

4.2.3 Verbesserte korrigierte Tests

Mit Hilfe der exakten konservativen Konstanten können die Tests aus Kapitel 4.1.2 für den Fall $p > 3$ verbessert werden.

Der F-Test auf Behandlungsunterschiede lehnt die Globalhypothese H_0^g ab, falls

$$\frac{MST}{MSE} > k_{t,n,p}^* F_{t-1, f_E; 1-\alpha}.$$

Der zweiseitige t-Test lehnt H_0^k zum Niveau α ab, falls

$$\frac{|\hat{\psi} - \psi_0|}{\sqrt{\widehat{\text{var}}_I(\hat{\psi})}} > \sqrt{k_{t,n,p}^* t_{f_E; 1-\alpha/2}}. \quad (4.12)$$

Die Korrektur mit Hilfe der Konstanten von Kunert und Utzig ist optimal in dem Sinne, dass jeder größere Korrekturfaktor zu Tests geringerer Power führt und jeder kleinere Korrekturfaktor die Varianz der Behandlungskontraste im Worst-Case-Szenario im Mittel unterschätzt.

Im nächsten Kapitel wird ein alternatives Verfahren zur Auswertung von Crossover-Plänen in Modell (2.6) vorgestellt, bei dem die Kovarianzmatrix der Fehler geschätzt wird.

5 Die modifizierte F-Test-Approximation

5.1 Das Näherungsverfahren nach Box

Box (1954) weist darauf hin, dass die Verteilung quadratischer Formen N -variater normalverteilter Zufallsvektoren eine Mischung χ^2 -verteilter Zufallsvariablen ist. Er gibt eine Reihenentwicklung für die Verteilung einer solchen Mischverteilung $\sum_{j=1}^r \lambda_j \chi_{\nu_j}^2$ an, wobei die Faktoren λ_j und die Freiheitsgrade ν_j von der Kovarianzmatrix Σ und der konkret gewählten quadratischen Form abhängen. Darauf aufbauend wird eine Approximation an die Verteilung der quadratischen Form bestimmt. Sie ist so gewählt, dass die Verteilung der Approximation die gleichen ersten und zweiten Momente besitzt wie die quadratische Form.

Lemma 5.1. (Box, 1954). Sei $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \Sigma)$, $\Sigma \in \mathbb{R}^{N \times N}$ n.n.d. und $\mathbf{M} \in \mathbb{R}^{N \times N}$, so dass $q_{\mathbf{M}} = \mathbf{x}^T \mathbf{M} \mathbf{x} \sim \sum_{j=1}^r \lambda_j \chi_{\nu_j}^2$. Sei weiter $z \sim b \chi_h^2$, wobei

$$b = \frac{\text{var}(q_{\mathbf{M}})}{\text{E}(q_{\mathbf{M}})}$$
$$h = \frac{2(\text{E}(q_{\mathbf{M}}))^2}{\text{var}(q_{\mathbf{M}})}.$$

und

Dann besitzen $q_{\mathbf{M}}$ und z die gleichen ersten und zweiten Momente.

Box (1954) schlägt nun vor, Zähler und Nenner der F-Statistik getrennt mit Hilfe obiger Methode anzunähern. Die Unabhängigkeit von Zähler und Nenner wird dabei nicht explizit verlangt. Für die praktische Anwendung beachte man, dass $\text{E}(q_{\mathbf{M}}) = \text{tr}(\mathbf{M}\Sigma)$ und $\text{var}(q_{\mathbf{M}}) = 2 \text{tr}((\mathbf{M}\Sigma)^2)$, siehe Box (1954)[S. 290 f.]. Dies führt uns zu folgender Definition.

Definition 5.1. Sei $q_{\mathbf{M}} = \mathbf{x}^T \mathbf{M} \mathbf{x}$ wie in Lemma 5.1 und $q_{\tilde{\mathbf{M}}} = \mathbf{x}^T \tilde{\mathbf{M}} \mathbf{x}$ eine weitere quadratische Form. Definiere

$$F = \frac{q_{\mathbf{M}} \operatorname{Rg}(\tilde{\mathbf{M}})}{q_{\tilde{\mathbf{M}}} \operatorname{Rg}(\mathbf{M})}.$$

Dann approximiert die Box-Korrektur die Verteilung von F durch bF_{h_1, h_2} , wobei

$$b = \frac{\operatorname{Rg}(\tilde{\mathbf{M}}) \operatorname{tr}(\mathbf{M} \boldsymbol{\Sigma})}{\operatorname{Rg}(\mathbf{M}) \operatorname{tr}(\tilde{\mathbf{M}} \boldsymbol{\Sigma})},$$

$$h_1 = \frac{(\operatorname{tr}(\mathbf{M} \boldsymbol{\Sigma}))^2}{\operatorname{tr}((\mathbf{M} \boldsymbol{\Sigma})^2)}$$

und

$$h_2 = \frac{(\operatorname{tr}(\tilde{\mathbf{M}} \boldsymbol{\Sigma}))^2}{\operatorname{tr}((\tilde{\mathbf{M}} \boldsymbol{\Sigma})^2)}.$$

Im Allgemeinen ist F weder nach bF_{h_1, h_2} verteilt noch stimmen Erwartungswert und Varianz von F und bF_{h_1, h_2} überein. Falls aber $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_p$, so ist $b = 1$, $h_1 = \operatorname{Rg}(\mathbf{M})$ sowie $h_2 = \operatorname{Rg}(\tilde{\mathbf{M}})$, da orthogonale Projektoren idempotent sind und Spur und Rang den gleichen Wert annehmen. Wenn dann Zähler und Nenner stochastisch unabhängig sind, so gilt tatsächlich $F \sim bF_{h_1, h_2}$.

Obiges Verfahren lässt sich auf eine Vielzahl linearer Modelle anwenden, bei denen die Fehler nicht u.i.v. sind. So basiert schon die Approximation der Freiheitsgrade für den Zweistichproben-t-Test bei ungleichen Varianzen nach Welch (1937) auf der gleichen Idee. Eine weitere Anwendung findet sich in Geisser und Greenhouse (1958). Die Autoren betrachten die Zweifachklassifikation mit einem festen und einem zufälligen Effekt. Sie schätzen den Korrekturfaktor b und die Freiheitsgrade aus Definition 5.1 für beliebige Kovarianzmatrizen so ab, dass ein konservativer Test entsteht.

Huynh und Feldt (1970) untersuchen die Box-Korrektur für Repeated-Measurements-Designs mit zufälligen Blockeffekten. Sie gehen von einem Modell ohne Nachwirkungseffekte aus und stellen fest, dass die F-Statistik unter der Nullhypothese genau dann tatsächlich F-verteilt ist, wenn der Korrekturfaktor nach Box (1954) gleich 1 ist. Sie zeigen, dass dies gleichbedeutend ist mit der Eigenschaft, dass die Kovarianzmatrix sphärisch ist.

Definition 5.2. Sei $\mathbf{x} = [x_1, \dots, x_p]^T$ Zufallsvektor mit Kovarianzmatrix \mathbf{S} . \mathbf{S} heißt sphärisch, falls ein $c \in \mathbb{R}$ existiert, so dass für alle $i \neq j$

$$\text{var}(x_i - x_j) = c.$$

Spezialfälle sphärischer Kovarianzstrukturen sind vollständig symmetrische Kovarianzmatrizen. Ein trivialer Spezialfall ist der Fall $\mathbf{S} = \sigma^2 \mathbf{I}_p$.

Für den Fall, dass die F-Verteilung nicht gilt, schlagen Huynh und Feldt (1976) ein Verfahren vor, das auf der Schätzung der Kovarianz beruht und weniger konservativ ist als die Methode nach Geisser und Greenhouse (1958).

5.2 Anwendung auf das Modell mit Carryover-Effekten

Bellavance et al. (1996) wenden die Box-Korrektur auf Modelle mit Nachwirkungseffekten an. Dieser Abschnitt beschreibt das Vorgehen dieser Autoren im Detail. In der Literatur finden sich mehrere Vergleiche dieses Verfahrens mit konkurrierenden Auswertungen. Darauf gehen wir am Ende dieses Abschnittes ein.

5.2.1 Das MFA-Verfahren von Bellavance et al.

Der F-Test aus Definition 2.3 lehnt in Modell (2.7) die Nullhypothese H_0^g ab, falls $MST/MSE > F_{t-1, f_E; 1-\alpha}$. Hier sind MST und MSE wie in Kapitel 4.1.2.

In Modell (2.6) verwirft man mit der Box-Korrektur H_0^g zugunsten von H_1^g , falls

$$\frac{MST}{MSE} > b^M F_{h_1^M, h_2^M; 1-\alpha}.$$

Dabei sind

$$\begin{aligned} b^M &= \frac{f_E}{t-1} \frac{\text{tr}((\boldsymbol{\omega}_{[U, P, F, T]} - \boldsymbol{\omega}_{[U, P, F]}) \boldsymbol{\Sigma})}{\text{tr}(\boldsymbol{\omega}_{[U, P, F, T]}^\perp \boldsymbol{\Sigma})}, \\ h_1^M &= \frac{(\text{tr}((\boldsymbol{\omega}_{[U, P, F, T]} - \boldsymbol{\omega}_{[U, P, F]}) \boldsymbol{\Sigma}))^2}{\text{tr}((\boldsymbol{\omega}_{[U, P, F, T]} - \boldsymbol{\omega}_{[U, P, F]}) \boldsymbol{\Sigma})^2} \end{aligned} \quad (5.1)$$

und

$$h_2^M = \frac{\left(\text{tr} \left(\boldsymbol{\omega}_{[U,P,F,T]}^\perp \boldsymbol{\Sigma} \right) \right)^2}{\text{tr} \left(\left(\boldsymbol{\omega}_{[U,P,F,T]}^\perp \boldsymbol{\Sigma} \right)^2 \right)},$$

da $\text{Rg}(\boldsymbol{\omega}_{[U,P,F,T]}^\perp) = f_E$ und $\text{rg}(\boldsymbol{\omega}_{[U,P,F,T]} - \boldsymbol{\omega}_{[U,P,F]}) = t - 1$. Man beachte, dass b^M , h_1^M und h_2^M jeweils von den Designparametern t, n, p und über $\boldsymbol{\Sigma} = \mathbf{I}_n \otimes \mathbf{S}$ von der Kovarianzmatrix \mathbf{S} abhängen. Wenn die Abhängigkeit von einer konkreten Matrix \mathbf{S} gekennzeichnet werden soll, schreiben wir im Folgenden auch $b^M(\mathbf{S})$ für b^M .

In der Praxis ist \mathbf{S} unbekannt und muss aus den Daten geschätzt werden. Einen einfachen Schätzer erhält man, falls die Behandlungssequenzen für die Einheiten im Versuchsplan wiederholt werden. Man betrachte dazu folgende Notation. Der Plan d bestehe aus n_{seq} verschiedenen Behandlungssequenzen, wobei die i -te Behandlungssequenz n_i Mal in d vorkomme. Für den Rest dieses Kapitels schreiben wir $\mathbf{y}_{ij} = (y_{ij1}, \dots, y_{ijp})^T$ für den Vektor der Beobachtungen an der j -ten Einheit der i -ten Behandlungssequenz und $\bar{\mathbf{y}}_i = \sum_{j=1}^{n_i} \mathbf{y}_{ij} / n_i$ für den Vektor der durchschnittlichen Beobachtungen in der i -ten Sequenz. Wir definieren den Schätzer $\hat{\boldsymbol{\Sigma}}$ für $\boldsymbol{\Sigma}$ wie in Jones und Kenward (1989)[S. 284] durch $\hat{\boldsymbol{\Sigma}} = \mathbf{I}_n \otimes \hat{\mathbf{S}}$ mit

$$\hat{\mathbf{S}} = \frac{1}{n - n_{seq}} \sum_{i=1}^{n_{seq}} \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_i) (\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)^T. \quad (5.2)$$

Jones und Kenward (1989) verwenden ein Modell mit zufälligen Effekten und bemerken, dass obiger Schätzer in diesem Modell unverzerrt ist. Bellavance et al. (1996) verwenden den Kovarianzschätzer $\hat{\boldsymbol{\Sigma}}$ für $\boldsymbol{\Sigma} = \mathbf{I}_n \otimes \mathbf{S}$ in Modell (2.6) und behaupten, dass $\hat{\boldsymbol{\Sigma}}$ unverzerrt für $\boldsymbol{\Sigma}$ sei. Man sieht jedoch leicht, dass dies im Allgemeinen nicht der Fall ist, solange nicht alle Einheiten einer Sequenz den gleichen festen Einheiten-effekt besitzen. Um die Notation einfach zu halten, beschränken wir uns im folgenden Lemma auf den Fall, dass jede Behandlungssequenz im Versuchsplan gleich häufig vorkommt, also dass $n_1 = \dots = n_{n_{seq}} = n_w$, wobei $n_w = n/n_{seq}$.

Lemma 5.2. *Sei $d \in \Lambda_{t,n,p}^*$ und $\hat{\mathbf{S}}$ wie in (5.2) und $n_1 = \dots = n_{n_{seq}} = n_w$. Sei α_{ij} der feste Einheiten-effekt der j -ten Einheit der i -ten Sequenz in d und $\bar{\alpha}_i = \frac{1}{n_w} \sum_j \alpha_{ij}$. Dann gilt in Modell (2.6)*

$$E_{\mathbf{S}}(\hat{\mathbf{S}}) = \mathbf{S} + c_{\alpha} \mathbf{1}_p \mathbf{1}_p^T,$$

wobei

$$c_{\alpha} = \frac{1}{n - n_{seq}} \sum_{i=1}^{n_{seq}} \sum_{j=1}^{n_w} (\alpha_{ij} - \bar{\alpha}_i)^2.$$

Beweis. In (2.6) ist $y_{ijk} = \mu + \alpha_{ij} + \pi_k + \tau_{d(ij,k)} + \rho_{d(ij,k-1)} + \epsilon_{ijk}$, wobei ϵ_{ijk} der Fehler in der k -ten Periode bei der j -ten Einheit in Behandlungssequenz i ist. Definiere $\bar{\epsilon}_{i \cdot k} = \frac{1}{n_w} \sum_j \epsilon_{ijk}$ und bezeichne mit $\bar{y}_{i \cdot k}$ das k -te Element von $\bar{\mathbf{y}}_i$. Schreiben wir den (k_1, k_2) -ten Eintrag von $\hat{\mathbf{S}}$ als \hat{s}_{k_1, k_2} , so erhalten wir

$$\begin{aligned} \hat{s}_{k_1, k_2} &= \frac{1}{n - n_{seq}} \sum_{i=1}^{n_{seq}} \sum_{j=1}^{n_w} (y_{ijk_1} - \bar{y}_{i \cdot k_1}) (y_{ijk_2} - \bar{y}_{i \cdot k_2}) \\ &= \frac{1}{n - n_{seq}} \sum_{i=1}^{n_{seq}} \sum_{j=1}^{n_w} (\alpha_{ij} - \bar{\alpha}_i + \epsilon_{ijk_1} - \bar{\epsilon}_{i \cdot k_1}) (\alpha_{ij} - \bar{\alpha}_i + \epsilon_{ijk_2} - \bar{\epsilon}_{i \cdot k_2}) \\ &= \frac{1}{n - n_{seq}} \sum_{i=1}^{n_{seq}} \sum_{j=1}^{n_w} (\alpha_{ij} - \bar{\alpha}_i)^2 \\ &\quad + \frac{1}{n - n_{seq}} \sum_{i=1}^{n_{seq}} \sum_{j=1}^{n_w} (\alpha_{ij} - \bar{\alpha}_i) (\epsilon_{ijk_1} + \epsilon_{ijk_2} - \bar{\epsilon}_{i \cdot k_1} - \bar{\epsilon}_{i \cdot k_2}) \\ &\quad + \frac{1}{n - n_{seq}} \sum_{i=1}^{n_{seq}} \sum_{j=1}^{n_w} (\epsilon_{ijk_1} - \bar{\epsilon}_{i \cdot k_1}) (\epsilon_{ijk_2} - \bar{\epsilon}_{i \cdot k_2}). \end{aligned}$$

Es ist $E_{\mathbf{S}}(\epsilon_{ijk}) = 0$ und $E_{\mathbf{S}}(\epsilon_{ijk_1} \epsilon_{ijk_2}) = s_{k_1, k_2}$, dem (k_1, k_2) -ten Eintrag in \mathbf{S} , sowie $E_{\mathbf{S}}(\epsilon_{ij_1 k_1} \epsilon_{ij_2 k_2}) = 0$ für $j_1 \neq j_2$. Man rechnet nach, dass dann

$$\begin{aligned} E_{\mathbf{S}}(\hat{s}_{k_1, k_2}) &= c_{\alpha} + E_{\mathbf{S}} \left(\frac{1}{n - n_{seq}} \sum_{i=1}^{n_{seq}} \sum_{j=1}^{n_w} (\epsilon_{ijk_1} - \bar{\epsilon}_{i \cdot k_1}) (\epsilon_{ijk_2} - \bar{\epsilon}_{i \cdot k_2}) \right) \\ &= c_{\alpha} + s_{k_1, k_2}. \end{aligned}$$

Daraus folgt sofort die Behauptung. □

Wir können den Schätzer $\hat{\mathbf{S}}$ dennoch für den korrigierten F-Test benutzen. Denn $\mathbf{Q}_p \hat{\mathbf{S}} \mathbf{Q}_p$ ist in (2.6) ein erwartungstreuer Schätzer für die zentrierte Kovarianzmatrix $\tilde{\mathbf{S}} = \mathbf{Q}_p \mathbf{S} \mathbf{Q}_p$. Da $k_{t, n, p}(\mathbf{S})$ nur über $\tilde{\mathbf{S}}$ von \mathbf{S} abhängt, stört uns der Term c_{α} bei der Box-Korrektur nicht.

Für den Fall, dass die Behandlungssequenzen in d nicht wiederholt werden, kann dieses Verfahren nicht benutzt werden. Jones und Kenward (1989)[S. 287–289] beschreiben, wie man die Varianz in nicht wiederholten Plänen schätzen kann. Dies ist vor allem dann relevant, wenn $p < t$ und vollständig balancierte Pläne benutzt werden. Solche unvollständigen Blockpläne benötigen oft schon dann einen großen Stichprobenumfang, wenn jede Behandlungssequenz nur einmal im Plan vorkommt.

Bellavance et al. (1996) nennen die oben beschriebene Anwendung der Box- Korrektur *modifizierte F-Test-Approximation* oder kurz *MFA-Verfahren*. Zusammenfassend hat man folgende

Definition 5.3. *Bei der modifizierten F-Test-Approximation verwirft man $H_0^g : \tau_1 = \dots = \tau_t$ zugunsten von $H_1^g : \exists i \neq j : \tau_i \neq \tau_j$, falls*

$$\frac{MST}{MSE} > \hat{b}^M F_{\hat{h}_1^M, \hat{h}_2^M; 1-\alpha}.$$

Hier sind

$$\hat{b}^M = \frac{f_E}{t-1} \frac{\text{tr} \left((\boldsymbol{\omega}_{[U, P, F, T]} - \boldsymbol{\omega}_{[U, P, F]}) \hat{\boldsymbol{\Sigma}} \right)}{\text{tr} \left(\boldsymbol{\omega}_{[U, P, F, T]}^\perp \hat{\boldsymbol{\Sigma}} \right)},$$

$$\hat{h}_1^M = \frac{\left(\text{tr} \left((\boldsymbol{\omega}_{[U, P, F, T]} - \boldsymbol{\omega}_{[U, P, F]}) \hat{\boldsymbol{\Sigma}} \right) \right)^2}{\text{tr} \left(\left((\boldsymbol{\omega}_{[U, P, F, T]} - \boldsymbol{\omega}_{[U, P, F]}) \hat{\boldsymbol{\Sigma}} \right)^2 \right)}$$

und

$$\hat{h}_2^M = \frac{\left(\text{tr} \left(\boldsymbol{\omega}_{[U, P, F, T]}^\perp \hat{\boldsymbol{\Sigma}} \right) \right)^2}{\text{tr} \left(\left(\boldsymbol{\omega}_{[U, P, F, T]}^\perp \hat{\boldsymbol{\Sigma}} \right)^2 \right)}.$$

Dabei sind MST und MSE wie in Kapitel 4.1.2 und $\hat{\boldsymbol{\Sigma}} = \mathbf{I}_n \otimes \hat{\mathbf{S}}$ mit $\hat{\mathbf{S}}$ wie in (5.2).

Für \hat{b}^M wird im Folgenden auch die Schreibweise $\hat{b}^M(\mathbf{S})$ benutzt, wenn die Abhängigkeit von \mathbf{S} betont werden soll.

Analog zu obigem F-Test kann man auch einen t-Test für H_0^k gegen H_1^k auf Grundlage des MFA-Verfahrens konstruieren. Man lehnt dann H_0^k ab, falls

$$\frac{|\hat{\psi} - \psi_0|}{\sqrt{\widehat{\text{var}}_{\mathbf{I}}(\hat{\psi})}} > \sqrt{\widehat{b}^M} t_{\widehat{h}_2^M; 1-\alpha/2},$$

vergleiche (4.6).

5.2.2 Vergleich mit konkurrierenden Auswertungen

Bellavance et al. (1996) vergleichen vier Auswertungen in Modell (2.6). Neben der gewöhnlichen KQ-Schätzung (*OLS*) und dem MFA-Verfahren sind dies eine Anwendung von Pearson-Kurven und eine empirische Variante des Aitken-Schätzers (EGLS, vergleiche 2.1.1). In ihrer Simulationsstudie überprüfen sie den Fehler 1. Art beim Test auf Behandlungseffekte. Die Autoren wählen sechs verschiedene Kovarianzmatrizen für zwei vollständig balancierte Pläne zum Vergleich von drei Behandlungen in drei Perioden ($n=18$ und $n=72$). Das OLS-Verfahren schneidet bei den fünf nicht sphärischen Kovarianzmatrizen in der Studie erwartungsgemäß schlecht ab. Das wahre Testniveau liegt hier bei Tests zum nominellen Niveau 5% bei bis über 12%. Bei kleinem Stichprobenumfang ($n=18$) führt die EGLS-Methode auch bei sphärischer Kovarianzmatrix zu deutlich anti-konservativen Tests. Der Fehler 1. Art liegt hier bei etwa 10%. Mit Hilfe von Pearson-Kurven lassen sich Verteilungen schätzen. Dabei werden 4 Momente einer Verteilung berücksichtigt. Dies kann man nutzen, um die Verteilung von Teststatistiken zu approximieren, vergleiche Solomon und Stephens (1978). In der Simulationsstudie von Bellavance et al. (1996) ergaben sich mit dieser Methode bei kleinem Stichprobenumfang Fehler 1. Art um 5% - 7%. Diese Werte liegen nahe bei den durch das MFA-Verfahren erzielten Werten. Das MFA-Verfahren lieferte aber insgesamt das beste Ergebnis. Hier lag das empirische Niveau unabhängig von der Kovarianzmatrix sehr nahe bei 5%.

Correa und Bellavance (2001) untersuchen mit Hilfe einer Simulationsstudie die Power der MFA-Methode und vergleichen das Verfahren darüberhinaus mit zwei nicht-parametrischen Tests. Einer dieser Tests wurde von Bellavance und Tardif (1995) speziell für Crossover-Pläne mit drei Behandlungen und drei Perioden entwickelt. Der zweite nichtparametrische Test wurde von Öhrvik (1998) vorgeschlagen und kann auch für Pläne mit mehr als drei Behandlungen benutzt werden. Der Fehler 1. Art wird sowohl vom MFA-Verfahren als auch von den nichtparametrischen Tests sehr gut kontrolliert. Die Power des MFA-Verfahrens ist in dieser Studie auch bei sphärischer

Kovarianzstruktur nicht schlechter als die des OLS-Verfahrens und größer als die des Tests von Bellavance und Tardif (1995). Der Test auf Behandlungseffekte von Öhrvik (1998) wird durch Carryover-Effekte verzerrt. Die Power dieses Verfahrens hängt von der Richtung der Nachwirkungen ab. Angesichts dieser Verzerrung ist es fraglich, ob das Verfahren angewandt werden sollte, wenn tatsächlich mit Nachwirkungen gerechnet wird.

Correa und Bellavance (2001) untersuchen darüberhinaus die Verletzung der Normalverteilungsannahme. Neben der multivariaten Normalverteilung betrachten sie Fehler aus der Gammaverteilung. Bei nicht normalverteilten Fehlern halten die nichtparametrischen Tests das Niveau im Wesentlichen ein, während das MFA-Verfahren hier bei einigen Simulationseinstellungen empirische Niveaus von bis zu 8% liefert. Die Aussagen über die Power der Tests bei Normalverteilung bestätigen sich auch bei der Gammaverteilung. Ein weiterer nichtparametrischer Test für Crossover-Studien mit verallgemeinerten lateinischen Quadraten wird von Tardif et al. (2005) vorgeschlagen.

Die von Bellavance et al. (1996) betrachtete EGLS-Methode führt zu stark anti-konservativen Tests. Kenward und Roger (1997) schlagen ein Verfahren vor, das sich ebenfalls als eine EGLS-Variante darstellen lässt. Chen und Wei (2003) vergleichen die MFA-Methode mit Hilfe einer Simulationsstudie mit der Auswertung nach Kenward und Roger (1997). Anders als die MFA-Methode hält das Verfahren von Kenward und Roger (1997) das Testniveau bei kleinem Stichprobenumfang nicht ein. Die Power der MFA-Methode ist deutlich niedriger als die der Methode von Kenward und Roger (1997), falls keine vollständig balancierte Kovarianzmatrix vorliegt. Chen und Wei (2003), die Versuchspläne mit drei und vier Behandlungen betrachten, kommen zu dem Schluß, dass die MFA-Methode bei Studien mit kleinem Stichprobenumfang $n < 24$ dem Verfahren von Kenward und Roger (1997) vorzuziehen ist. Sie weisen außerdem darauf hin, dass die größere Power sich vor allem dadurch erklärt, dass die Einheitenefekte bei Kenward und Roger (1997) als zufällig modelliert werden. Sie folgern, dass dieser Vorteil verloren ginge, wenn man in EGLS-Verfahren die Einheiten als feste Effekte modelliert. Jones und Kenward (2003)[S. 262 f.] empfehlen das MFA-Verfahren als einfach anzuwendendes Auswertungsverfahren, geeignet für kleine Stichprobenumfänge.

Zusammenfassend kann man feststellen, dass die in der Literatur betrachteten Simulationsstudien das MFA-Verfahren zur Auswertung von Crossover-Studien empfehlen.

Das Verfahren zeichnet sich gegenüber seinen Konkurrenten insbesondere dadurch aus, dass es auch bei geringen Stichprobenumfängen das Testniveau gut einhält.

Im nächsten Kapitel vergleichen wir das MFA-Verfahren mit unserer Verbesserung des Verfahrens von Kunert und Utzig (1993). Dabei wird auch das Worst-Case-Verhalten des MFA-Verfahrens untersucht.

6 Vergleich von konservativer Varianzschätzung und MFA-Verfahren

In diesem Kapitel vergleichen wir das Verfahren von Kunert und Utzig (1993) mit dem MFA-Verfahren von Bellavance et al. (1996). Im nächsten Abschnitt wird zunächst eine interessante Verbindung zwischen beiden Verfahren aufgezeigt. Bei bekannter Kovarianzmatrix ist der Varianzquotient $k_{t,n,p}(\mathbf{S})$ aus Definition 4.1 gleich dem Korrekturfaktor $b^M(\mathbf{S})$ in (5.1). Diese Tatsache erleichtert den Vergleich beider Verfahren.

6.1 Gleichheit der Korrekturfaktoren

Man betrachte den Beweis von Lemma 4.3. Gleichung (4.9) besagt, dass

$$\operatorname{tr}(\boldsymbol{\omega}_{[U,P,F]}^\perp \boldsymbol{\Sigma}) = (n-1) \operatorname{tr}(\tilde{\mathbf{S}}) - \frac{n}{c_{d22}} \operatorname{tr}(\mathbf{V} \mathbf{Q}_{p,t} \mathbf{V}^T \tilde{\mathbf{S}}).$$

Weiterhin gilt mit (4.10)

$$\begin{aligned} \operatorname{tr}(\boldsymbol{\omega}_{[U,P,F,T]}^\perp \boldsymbol{\Sigma}) &= \frac{n}{c_d} \operatorname{tr}(\mathbf{B}_d^* \tilde{\mathbf{S}}) \\ &= \frac{n}{c_d} \operatorname{tr}(\mathbf{Q}_p \mathbf{B}_d \mathbf{Q}_p \mathbf{S}). \end{aligned}$$

Man rechnet leicht nach, dass dann

$$\begin{aligned} \operatorname{tr}((\boldsymbol{\omega}_{[U,P,F,T]} - \boldsymbol{\omega}_{[U,P,F]}) \boldsymbol{\Sigma}) &= \operatorname{tr}((\boldsymbol{\omega}_{[U,P,F]}^\perp - \boldsymbol{\omega}_{[U,P,F,T]}^\perp) \boldsymbol{\Sigma}) \\ &= \frac{n}{c_d} \operatorname{tr}((\mathbf{I}_p - \gamma(\mathbf{V} + \mathbf{V}^T) + \gamma^2 \mathbf{V} \mathbf{Q}_{p,t} \mathbf{V}^T) \boldsymbol{\Sigma}) \\ &= \frac{n}{c_d} \mathbf{Q}_p \mathbf{A}_d \mathbf{Q}_p \mathbf{S}. \end{aligned}$$

Mit (5.1) und Satz 4.2 impliziert dies, dass

$$\begin{aligned} b^M(\mathbf{S}) &= \frac{f_E}{t-1} \frac{\text{tr} \left((\boldsymbol{\omega}_{[U,P,F,T]} - \boldsymbol{\omega}_{[U,P,F]}) \boldsymbol{\Sigma} \right)}{\text{tr} \left(\boldsymbol{\omega}_{[U,P,F,T]}^\perp \boldsymbol{\Sigma} \right)} \\ &= \frac{f_E}{t-1} \frac{\frac{n}{c_d} \text{tr} \left(\mathbf{Q}_p \mathbf{A}_d \mathbf{Q}_p \mathbf{S} \right)}{\frac{n}{c_d} \text{tr} \left(\mathbf{Q}_p \mathbf{B}_d \mathbf{Q}_p \mathbf{S} \right)} \\ &= k_{t,n,p}(\mathbf{S}). \end{aligned}$$

Damit haben wir den folgenden Satz bewiesen.

Satz 6.1. *Sei $d \in \Lambda_{t,n,p}^*$, $k_{t,n,p}(\mathbf{S})$ der Varianzquotient (4.3) und b^M der Korrekturfaktor aus (5.1). Dann gilt $b^M(\mathbf{S}) = k_{t,n,p}(\mathbf{S})$.*

Falls nun \mathbf{S} eine ungünstigste Kovarianzmatrix ist, so gilt $b^M(\mathbf{S}) = k_{t,n,p}^*$, das heißt die Konstante von Kunert und Utzig ist gleich dem Korrekturfaktor der MFA-Methode, der sich für die ungünstigste Kovarianzmatrix ergibt. Wenn die Kovarianzmatrix geschätzt wird, so ist im Allgemeinen $\hat{b}^M(\mathbf{S}) \neq k_{t,n,p}(\mathbf{S})$. Es gilt aber folgende Aussage.

Korollar 6.1. *Sei $d \in \Lambda_{t,n,p}^*$ ein Crossover-Plan, bei dem jede Behandlungssequenz n_w Mal wiederholt wird. Sei \mathbf{S} ungünstigste Kovarianzmatrix und $\hat{b}^M(\mathbf{S})$ der geschätzte Korrekturfaktor aus Definition 5.3. Es gelte Modell (2.6) mit der Normalverteilungsannahme $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{S})$. Dann gilt mit Wahrscheinlichkeit 1, dass $\hat{b}^M(\mathbf{S}) = k_{t,n,p}^*$.*

Beweis. Für diesen Beweis benutzen wir die Notation aus Kapitel 5.2.1 für Pläne, bei denen die Behandlungssequenzen wiederholt werden, siehe S.68–69. Nach Satz 4.3 ist $\mathbf{S} = \mathbf{v}\mathbf{v}^T$ für ein geeignetes $\mathbf{v} = (v_1, \dots, v_p)^T$, das heißt \mathbf{S} hat höchstens Rang 1 und alle Beobachtungen einer Einheit sind perfekt linear korreliert. Dann gilt mit Wahrscheinlichkeit 1, dass $\mathbf{y}_{ij} = y_{ij1}\mathbf{v}/v_1$ und $\bar{\mathbf{y}}_i = \bar{y}_{i,1}\mathbf{v}/v_1$. Mit (5.2) folgt dann mit Wahrscheinlichkeit 1

$$\begin{aligned} \hat{\mathbf{S}} &= \frac{1}{n - n_{seq}} \sum_{i=1}^{n_{seq}} \sum_{j=1}^{n_w} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)^T \\ &= \frac{1}{(n - n_{seq})v_1^2} \sum_{i=1}^{n_{seq}} \sum_{j=1}^{n_w} (y_{ij1} - \bar{y}_{i,1})\mathbf{v}\mathbf{v}^T(y_{ij1} - \bar{y}_{i,1}). \end{aligned}$$

Mit der Notation

$$c_{\mathbf{S}} = \frac{1}{(n - n_{seq})v_1^2} \sum_{i=1}^{n_{seq}} \sum_{j=1}^{n_w} (y_{ij1} - \bar{y}_{i.1})^2$$

erhält man $\hat{\Sigma} = \mathbf{I}_n \otimes c_{\mathbf{S}} \mathbf{S} = c_{\mathbf{S}} \Sigma$. Die Konstante $c_{\mathbf{S}}$ ist mit Wahrscheinlichkeit 1 nicht gleich 0, so dass in Definition 5.3 mit Wahrscheinlichkeit 1

$$\begin{aligned} \hat{b}^M(\mathbf{S}) &= \frac{f_E}{t-1} \frac{\text{tr} \left((\omega_{[U,P,F,T]} - \omega_{[U,P,F]}) \hat{\Sigma} \right)}{\text{tr} \left(\omega_{[U,P,F,T]}^\perp \hat{\Sigma} \right)} \\ &= \frac{f_E}{t-1} \frac{c_{\mathbf{S}} \text{tr} \left((\omega_{[U,P,F,T]} - \omega_{[U,P,F]}) \Sigma \right)}{c_{\mathbf{S}} \text{tr} \left(\omega_{[U,P,F,T]}^\perp \Sigma \right)} \\ &= b^M(\mathbf{S}). \end{aligned} \quad \square$$

Im Worst-Case-Fall wird die Kovarianzmatrix \mathbf{S} also insofern perfekt geschätzt, als das MFA-Verfahren den gleichen Korrekturfaktor realisiert wie das Verfahren, das die Konstante von Kunert und Utzig benutzt. Die obigen Resultate lassen sich dazu nutzen, das MFA-Verfahren besser zu studieren. Die in Kapitel 5.2.2 zitierten Simulationsstudien betrachten stets nur wenige willkürlich ausgewählte Kovarianzmatrizen. Nun kann das MFA-Verfahren beispielsweise auch für ungünstigste Kovarianzmatrizen untersucht werden.

6.2 Simulationsstudie

Bei den Tests auf der Grundlage des MFA-Verfahrens (Kapitel 5.2.1) werden anders als bei den Tests nach dem Verfahren von Kunert und Utzig (Kapitel 4.2.3) auch die Freiheitsgrade bei der Berechnung des kritischen Werts korrigiert. Die Frage, welches der beiden Verfahren bei Vorliegen ungünstigster Kovarianzmatrizen besser ist, reduziert sich mit Korollar 6.1 auf die Frage, ob die Tests mit Freiheitsgradapproximation nach Box (1954) besser sind als die Tests ohne Freiheitsgradkorrektur. Befindet man sich nicht im Worst-Case-Szenario, so spielt ein weiterer Faktor eine Rolle. Es muss dann zusätzlich geklärt werden, ob der Vorteil des Verfahrens von Kunert und Utzig, keine Schätzung der Kovarianzmatrix vornehmen zu müssen, stärker wiegt als der Vorteil des MFA-Verfahrens, nicht für ungünstigste Kovarianzmatrizen zu korrigieren, wenn diese gar nicht vorliegen.

Weiterhin wollen wir klären, ob die Korrektur der Varianzschätzung mit Hilfe von $k_{t,n,p}^*$ tatsächlich zu besseren Tests führt als die Korrektur mit $k_{t,n,p}^{oben}$. Die im Folgenden beschriebene Simulationsstudie hilft, diese Fragen zu beantworten. Die Berechnungen wurden mit Hilfe des Softwarepakets R (R Development Core Team, 2009) durchgeführt. Der zugehörige R-Code ist in Anhang B.2 aufgeführt.

6.2.1 Aufbau

Daten

Wir gehen davon aus, dass die Einheiten randomisiert sind und daher jede Einheit die gleiche Kovarianzmatrix aufweist. Um Daten zu erzeugen, die Modell (2.6) genügen, wählt man zunächst μ , α und π . Dann legt man die Behandlungseffekte und Carryover-Effekte fest, siehe hierzu den Abschnitt *Modellparameter* dieses Kapitels. In jeder Iteration der Simulation werden die Einheiten und Behandlungsnamen randomisiert. Schließlich werden Pseudozufallszahlen aus der multivariaten Normalverteilung mit vorgegebener Kovarianzmatrix erzeugt.

Versuchspläne

Da die obere Schranke $k_{t,n,p}^{oben}$ für $p = 3$ scharf ist, wählen wir $p > 3$, um das Verfahren von Kunert und Utzig mit unserer Verbesserung dieses Verfahrens vergleichen zu können. Um das MFA-Verfahren mit der Kovarianzschätzung \hat{S} aus (5.2) anwenden zu können, wählen wir Versuchspläne, bei denen jede Behandlungssequenz wiederholt wird. Wir betrachten zwei verschiedene Stichprobenumfänge, um den Einfluss des Stichprobenumfangs auf die Schätzung von S zu erfassen.

Eine mögliche Wahl ist der Crossover-Plan d_6 aus Tabelle 6.1 zum Vergleich von $t = 4$ Behandlungen in $n = 12$ Blöcken der Länge $p = 4$. Das Design ist ein vollständig balanciertes verallgemeinertes lateinisches Quadrat. Jede der vier Behandlungssequenzen in d_6 wird dreimal wiederholt.

Mit Hilfe des Plans d_6 kann ein Crossover-Plan für $n = 48$ Einheiten zum Vergleich von vier Behandlungen in vier Perioden konstruiert werden. Dazu wiederholt man jede der vier Behandlungssequenzen in d_6 zwölfmal statt lediglich dreimal. Das so definierte vollständig balancierte verallgemeinerte lateinische Quadrat nennen wir d_7 .

Periode	Einheit											
	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	1	2	2	2	3	3	3	4	4	4
2	2	2	2	3	3	3	4	4	4	1	1	1
3	4	4	4	1	1	1	2	2	2	3	3	3
4	3	3	3	4	4	4	1	1	1	2	2	2

Tabelle 6.1: Der vollständig balancierte Crossover-Plan $d_6 \in \Lambda_{4,12,4}^*$ mit Einheiten als Spalten und Perioden als Zeilen.

Modellparameter

Für jeden der beiden Versuchspläne werden verschiedene Kovarianzmatrizen \mathbf{S} , von der günstigsten bis zur ungünstigsten, untersucht. Hierzu betrachten wir für d_6 eine günstigste Kovarianzmatrix \mathbf{S}_g , die Matrix $0.5\mathbf{I}_4$ mit $b^M = 1$ und eine ungünstigste Kovarianzmatrix \mathbf{S}_u . Durch Konvexkombination werden weitere günstige Kovarianzmatrizen der Form $c_i\mathbf{S}_g + (1 - c_i)0.5\mathbf{I}_4$ und ungünstige Kovarianzmatrizen der Form $c_i0.5\mathbf{I}_4 + (1 - c_i)\mathbf{S}_u$ zwischen den Extremen erzeugt. Die Kovarianzmatrizen werden wie bei Chen und Wei (2003)[S. 2827] so skaliert, dass die durchschnittliche Varianz von Beobachtungen aus verschiedenen Perioden bei allen Kovarianzmatrizen identisch ist. Dadurch kann man die Power bei verschiedenen Kovarianzmatrizen besser vergleichen. Für d_7 geht man analog vor.

Eine weitere einfache Kovarianzstruktur sind autoregressive Prozesse 1. Ordnung (*AR(1)-Prozesse*). Das (i, j) -te Element der zugehörigen Kovarianzmatrix ist von der Form $\sigma^2\rho_{AR}^{|i-j|}$, vergleiche Brockwell und Davis (2002)[S. 53]. Der Bereich, den der Varianzquotient bei AR(1)-Prozessen mit positivem Korrelationsparameter ρ_{AR} annimmt, lässt sich für gegebene Designparameter (t, n, p) numerisch leicht bestimmen. Er ist für d_6 gegeben durch $[1, 1.2273]$. Bei d_7 decken die AR(1)-Prozesse das Intervall $[1, 1.2229]$ ab. Die entsprechenden Intervalle sind in den Abbildungen in Kapitel 6.2.2 zur Orientierung als vertikale schwarze Linien eingezeichnet. In der Simulationsstudie betrachten wir verschiedene AR(1)-Prozesse, wobei wir den Korrelationsparameter ρ_{AR} zwischen 0 und 1 variieren. Diese Kovarianzmatrizen werden bei der statistischen Modellierung realer Experimente eher zur Anwendung kommen als die mit Hilfe von \mathbf{S}_g beziehungsweise \mathbf{S}_u erzeugten Kovarianzstrukturen.

Zur Untersuchung der Tests unter der Nullhypothese wählen wir $\boldsymbol{\tau} = \mathbf{0}$ und setzen darüberhinaus auch alle Carryover-Effekte auf Null. Zur Untersuchung der Power wählen wir für d_6 $\tau_1 = 0.8$ und $\rho_1 = 0.4$, bei d_7 setzen wir $\tau_1 = 0.4$ und $\rho_1 = 0.2$. Damit erreichen wir, dass die Power nicht zu nahe an 0 und 1 liegt, was den Vergleich der Testverfahren erleichtert. Man beachte, dass die Auswertung in Modell (2.6) für etwaige Nachwirkungseffekte $\boldsymbol{\rho}$ korrigiert und die Varianzschätzer bei vollständig balancierten Designs nicht von $\boldsymbol{\rho}$ abhängen. Das bedeutet, dass man in der Simulationsstudie für d_6 und d_7 unabhängig von der Wahl von $\boldsymbol{\rho}$ stets die gleichen Ergebnisse erhält.

Kriterien

Für jede untersuchte Kovarianzmatrix werden für die zu vergleichenden Verfahren jeweils der F-Test und der t-Test in Modell (2.6) zum nominellen Niveau $\alpha = 5\%$ untersucht.

Zunächst betrachten wir die kritischen Werte der Tests ohne Korrektur der Varianzschätzung, mit der Korrektur auf Grundlage von $k_{t,n,p}^{oben}$ sowie der Korrektur auf Basis von $k_{t,n,p}^*$. Diese kritischen Werte vergleichen wir mit denen des MFA-Verfahrens für den idealisierten Fall, dass die Kovarianzmatrix \mathbf{S} bekannt ist beziehungsweise perfekt aus den Daten geschätzt wird. Dies ermöglicht eine Bewertung des Potenzials des MFA-Verfahrens bei einer verbesserten Schätzung von \mathbf{S} .

Danach betrachten wir die Quantile der Verteilung der kritischen Werte des MFA-Verfahrens für den realistischen Fall, dass \mathbf{S} mit Hilfe von $\hat{\mathbf{S}}$ aus (5.2) aus den Daten geschätzt wird. Damit erhalten wir einen Eindruck von der Variabilität der Ergebnisse bei der MFA-Methode.

Schließlich vergleichen wir das Niveau und die Power der Tests bei den Verfahren für verschiedene Kovarianzmatrizen. Ein gutes Verfahren sollte nicht anti-konservativ sein, also das Niveau für alle Kovarianzmatrizen einhalten. Darüberhinaus sollte es auf der Alternativen eine möglichst große Power aufweisen.

6.2.2 Ergebnisse

Der Test der Globalhypothese bei d_6

Bei Versuchsplan d_6 erhält man für den gewöhnlichen F-Test der Globalhypothese $H_0^g : \tau_1 = \dots = \tau_t$ in Modell (2.7) einen kritischen Wert von $F_{3,27;0.95} = 2.96$. Da

$k_{4,12,4}^{oben} = 1.8746$ und $k_{4,12,4}^* = 1.5195$, lauten die kritischen Werte für das ursprüngliche und das verbesserte Verfahren von Kunert und Utzig $k_{4,12,4}^{oben} F_{3,27;0.95} = 5.55$ und $k_{4,12,4}^* F_{3,27;0.95} = 4.50$, vergleiche Kapitel 4.1.2 und 4.2.3. Ist die Kovarianzmatrix \mathbf{S} bekannt, so kann man mit $k_{4,12,4}(\mathbf{S})$ korrigieren und erhält abhängig von \mathbf{S} kritische Werte zwischen 1.61 und 4.50. Abbildung 6.1 (linke Seite) zeigt diese kritischen Werte in Abhängigkeit vom Varianzquotienten $k_{4,12,4}(\mathbf{S})$ als blaue (gewöhnlicher F-Test), rote (obere Schranke von Kunert und Utzig), grüne (Konstante von Kunert und Utzig) und violette (Korrektur mit $k_{4,12,4}(\mathbf{S})$) Kurve. Per Konstruktion schneiden sich die blaue und violette Kurve bei $k_{4,12,4}(\mathbf{S}) = 1$ und die grüne und violette Kurve bei $k_{4,12,4}^*$.

Das MFA-Verfahren korrigiert zusätzlich die Freiheitsgrade. Bei bekanntem \mathbf{S} führt dies in Abhängigkeit von \mathbf{S} zu den in Abbildung 6.1 (linke Seite) schwarz eingezeichneten kritischen Werten $b^M F_{h_1^M, h_2^M; 0.95}$, vergleiche Definition 5.3. Dabei steht die durchgezogene schwarze Kurve für die kritischen Werte bei den Kovarianzmatrizen aus der Konvexkombination von ungünstigster und günstigster Kovarianzmatrix. Die gestrichelte Kurve zeigt die kritischen Werte für die AR(1)-Prozesse. Die kritischen Werte sind für gegebenes $k_{t,n,p}$ nahezu identisch.

Man beachte, dass die Freiheitsgradapproximation die kritischen Werte erhöht, falls $k_{t,n,p}(\mathbf{S}) \neq 1$. Die schwarze Kurve liegt bei der ungünstigsten Kovarianzstruktur oberhalb der grünen und sogar oberhalb der roten Kurve. Bei bekannter Kovarianzmatrix würde das MFA-Verfahren erwartungsgemäß bei vielen Kovarianzstrukturen häufiger ablehnen als das Verfahren von Kunert und Utzig. Bei großem Varianzquotienten hingegen lehnt das Verfahren von Kunert und Utzig für diesen Versuchsplan häufiger ab, und zwar selbst dann, wenn man zur Korrektur der kritischen Werte die unscharfe Schranke $k_{4,12,4}^{oben}$ verwendet.

In der Praxis ist \mathbf{S} unbekannt. Die kritischen Werte des F-Tests bei der MFA-Methode in diesem Fall sind in Abbildung 6.1 (rechte Seite) dargestellt. Die durchgezogenen grauen Linien zeigen sieben empirische Quantile des kritischen Werts $\hat{b}^M F_{\hat{h}_1^M, \hat{h}_2^M; 0.95}$ bei den durch Konvexkombination erzeugten Kovarianzmatrizen. Eingezeichnet sind von unten nach oben das 0.1%-Quantil, das 2.5%-Quantil, das 5%-Quantil, der Median, das 95%-Quantil, das 97.5%-Quantil und schließlich das 99.9%-Quantil. Die durchgezogene orange Linie gibt das arithmetische Mittel der kritischen Werte bei diesen Kovarianzmatrizen an. Die gestrichelte orange und schwarze Kurve sowie die gestrichelten grauen Kurven geben die Ergebnisse für die AR(1)-Prozesse

wieder. Die Unterschiede zwischen den Kovarianzmatrizen sind bei gleichem $k_{t,n,p}$ nur gering. Bei den AR(1)-Prozessen streuen die kritischen Werte bei großem Korrelationsparameter etwas weniger stark als bei den Matrizen aus der Konvexkombination.

Man erkennt, dass das arithmetische Mittel und der Median sehr nahe am schwarz eingezeichneten kritischen Wert liegen, der sich bei bekannter Kovarianzmatrix ergeben würde. Allerdings streuen die kritischen Werte stark. Dies liegt an der geringen Zahl der Wiederholungen von Behandlungssequenzen im Versuchsplan. Bei sphärischer Kovarianz können sich aber nur in sehr seltenen Fällen konservativere Tests ergeben als mit dem Verfahren von Kunert und Utzig, siehe dazu die Kurven für die 97.5%- und 99.9%-Quantile an der Stelle $k_{4,12,4}(\mathbf{S}) = 1$. Für Varianzquotienten in unmittelbarer Nähe von $k_{4,12,4}^*$ streuen die geschätzten kritischen Werte nur wenig. Im Worst-Case-Fall ist die Streuung im Einklang mit Korollar 6.1 gleich 0. Also lehnt das MFA-Verfahren bei ungünstigster Kovarianzmatrix stets seltener ab als das Verfahren von Kunert und Utzig.

Ein Vergleich der empirischen Niveaus der Testverfahren zeigt, dass das Verfahren von Kunert und Utzig nicht notwendig zu konservativen Tests führt. Abbildung 6.2 (linke Seite) zeigt das empirische Niveau der Testverfahren in der Simulationsstudie. Wenn ein Test das Niveau von 5% einhält, so wird er in der Simulation mit 10000 Iterationen in 95% aller Fälle ein empirisches Niveau zwischen 0.0458 und 0.0542 liefern. Dieses aus der Binomialverteilung berechnete Konfidenzintervall ist in der Abbildung grau hinterlegt. Der gewöhnliche F-Test (blau) verletzt das Niveau bekanntlich stark, wenn der wahre Varianzquotient groß ist. Obwohl das Verfahren von Kunert und Utzig (rot, grün) eine konservative Varianzschätzung beinhaltet, ist das Verfahren bei großem Varianzquotienten mangels Freiheitsgradkorrektur anti-konservativ. Dies gilt selbst dann, wenn die Schranke $k_{4,12,4}^{oben}$ verwendet wird. Wenn hingegen der Varianzquotient im mittleren Bereich liegt, dann führt die Korrektur mit $k_{4,12,4}^*$ zu konservativen Tests. Bei unkorrelierten Fehlern liegt das empirische Niveau hier nur bei 1.12%, mit Korrektur mit $k_{4,12,4}^{oben}$ sogar nur bei 0.44%. Die zum MFA-Verfahren gehörenden schwarzen Kurven in Abbildung 6.2 (linke Seite) liegen teilweise leicht oberhalb, teilweise leicht unterhalb des grau markierten Konfidenzintervalls. Das MFA-Verfahren hält das nominelle Niveau also nicht perfekt ein, das empirische Niveau liegt zwischen 3.05% und 5.72%. Damit ist das MFA-Verfahren aber selbst bei Schätzung von \mathbf{S} aus nur drei Sequenzwiederholungen weder besonders konservativ noch stark anti-konservativ. Dies gilt insbesondere in dem Bereich, der durch die AR(1)-Prozesse mit

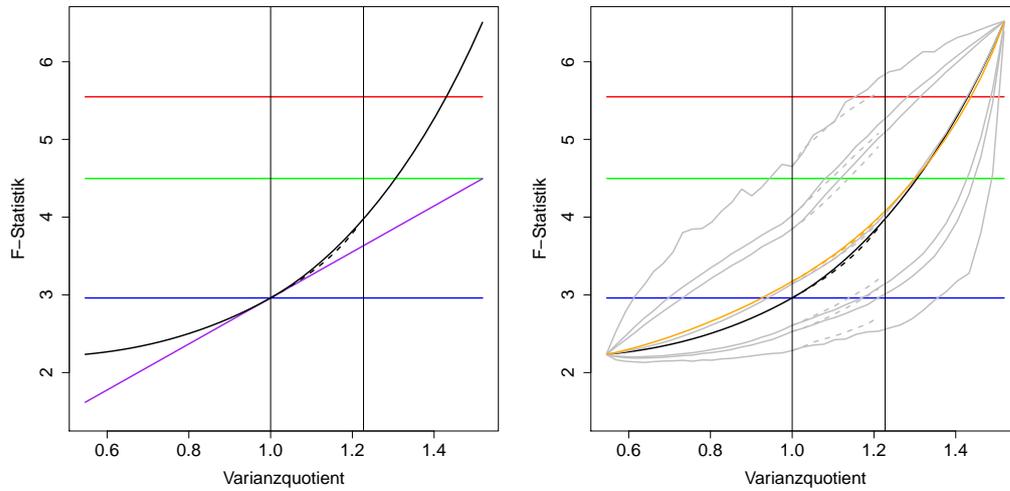


Abbildung 6.1: Kritische Werte des F-Tests bei Plan d_6 in Abhängigkeit vom Varianzquotienten $k_{4,12,4}(\mathbf{S})$ bei bekannter Kovarianzmatrix (links) und bei unbekannter Kovarianzmatrix (rechts). Dargestellt werden der gewöhnliche F-Test (blau), die Korrektur mit $k_{4,12,4}^{oben}$ (rot), die Korrektur mit $k_{4,12,4}^*$ (grün), die Korrektur mit $k_{4,12,4}(\mathbf{S})$ (violett) und das MFA-Verfahren bei bekannter Kovarianzmatrix (schwarz). Die grauen Kurven zeigen das 0.1%- , 2.5%- und 5%-Quantil, den Median, das 95%- , 97.5%- und 99.9%-Quantil der empirischen Verteilung der kritischen Werte beim MFA-Verfahren, die orange Linie das arithmetische Mittel. Die gestrichelten Kurven stellen die Ergebnisse für Kovarianzmatrizen bei AR(1)-Prozessen dar.

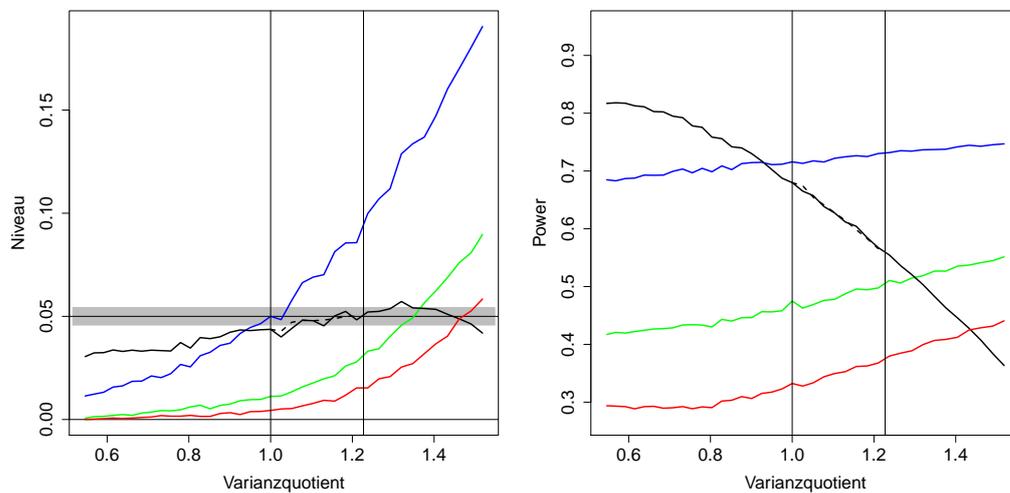


Abbildung 6.2: Niveau (links) und Power (rechts) des F-Tests bei Plan d_6 in Abhängigkeit vom Varianzquotienten $k_{4,12,4}(\mathbf{S})$ für den gewöhnlichen F-Test (blau), die Korrektur mit $k_{4,12,4}^{oben}$ (rot), die Korrektur mit $k_{4,12,4}^*$ (grün) und das MFA-Verfahren (schwarz – durchgezogen für Kovarianzmatrizen aus Konvexkombination, gestrichelt für AR(1)-Prozesse). Der grau hinterlegte Bereich gibt ein 95%-Konfidenzintervall an.

positivem Korrelationsparameter abgedeckt wird, und zwar sowohl für die Kovarianzmatrizen aus der Konvexkombination als auch für die AR(1)-Prozesse. Im Worst-Case-Szenario ist der F-Test hier sogar leicht anti-konservativ.

Da das Verfahren von Kunert und Utzig zum Teil sehr konservativ ist, ergibt sich eine entsprechend niedrige Power. Die Power ist bei Korrektur mit der oberen Schranke $k_{4,12,4}^{oben}$ niedriger als bei der Korrektur mit unserer exakten Schranke $k_{4,12,4}^*$. In Abbildung 6.2 (rechte Seite) sieht man, dass die Power des F-Tests hier nur bei sehr ungünstigen Kovarianzmatrizen größer ist als bei der MFA-Methode, deren Power fallend in $k_{t,n,p}(\mathbf{S})$ ist. In diesem Bereich ist das Verfahren von Kunert und Utzig jedoch anti-konservativ, so dass die höhere Power kein Argument für die Güte des Verfahrens ist. In dem Bereich, in dem beide Verfahren das Niveau einhalten, weist das MFA-Verfahren fast immer die größere Power auf. Lediglich in einem kleinen Bereich um $k_{t,n,p}(\mathbf{S}) = 1.32$ ergab sich in der Simulation für den Test nach Korrektur mit $k_{4,12,4}^*$ eine höhere Power bei einem Testniveau von höchstens 5%. Man beachte, dass die Power des MFA-Verfahrens in $k_{t,n,p}(\mathbf{S})$ fällt. Bei gegebenem $k_{t,n,p}$ hängt sie aber fast gar nicht davon ab, ob eine Kovarianzmatrix aus der Konvexkombination oder aus dem AR(1)-Prozess genutzt wird.

Zusammenfassend zeigt sich, dass das MFA-Verfahren für den F-Test der Globalhypothese bei Plan d_6 geeignet ist. Das Verfahren von Kunert und Utzig ist bei unbekannter Kovarianzmatrix \mathbf{S} selbst dann nicht sinnvoll, wenn die exakte obere Schranke genutzt wird. Es besteht die Gefahr, dass das Verfahren entweder extrem konservativ ist oder aber, falls eine ungünstigste Kovarianzmatrix vorliegt, sogar zu anti-konservativen Tests führt.

Die Ergebnisse unterscheiden sich bei Verwendung der mit Hilfe von \mathbf{S}_g beziehungsweise \mathbf{S}_u erzeugten Kovarianzstrukturen nur sehr geringfügig von den Ergebnissen auf Grundlage der AR(1)-Prozesse. Dies gilt auch für die weiteren Abschnitte dieser Auswertung, so dass wir im Folgenden nicht weiter auf die zwei Klassen von Kovarianzmatrizen eingehen.

Der Test auf Kontraste bei d_6

Für den t-Test der Hypothese $\tau_1 - \tau_2 = 0$ ergibt sich ein ähnliches Bild wie für den F-Test. Allerdings spielen die Freiheitsgrade beim t-Test eine geringere Rolle als beim F-Test. In Abbildung 6.3 (linke Seite) zeigt sich, dass bei der MFA-Methode bei

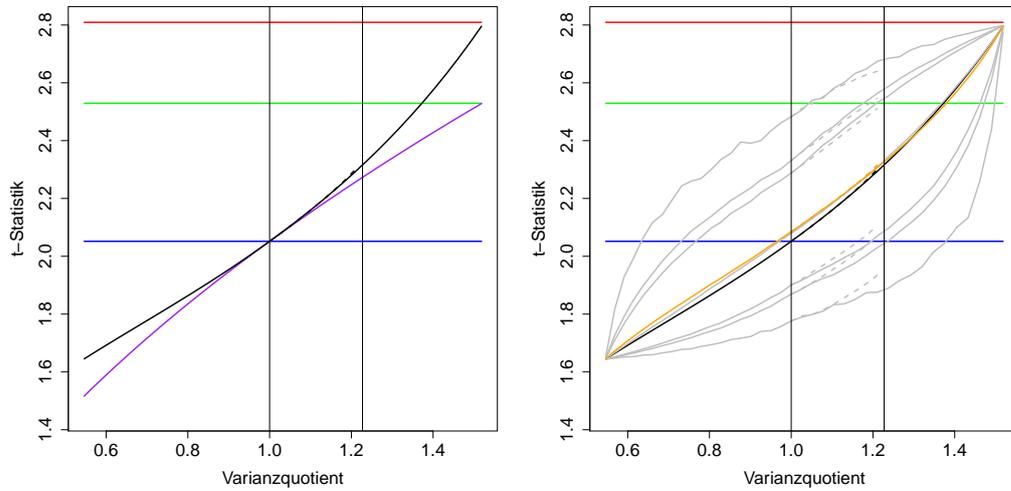


Abbildung 6.3: Kritische Werte des t -Tests bei Plan d_6 in Abhängigkeit vom Varianzquotienten $k_{4,12,4}(\mathbf{S})$ bei bekannter Kovarianzmatrix (links) und bei unbekannter Kovarianzmatrix (rechts). Dargestellt werden der gewöhnliche t -Test (blau), die Korrektur mit $k_{4,12,4}^{oben}$ (rot), die Korrektur mit $k_{4,12,4}^*$ (grün), die Korrektur mit $k_{4,12,4}(\mathbf{S})$ (violett) und das MFA-Verfahren bei bekannter Kovarianzmatrix (schwarz). Die grauen Kurven zeigen das 0.1%-, 2.5%- und 5%-Quantil, den Median, das 95%-, 97.5%- und 99.9%-Quantil der empirischen Verteilung der kritischen Werte beim MFA-Verfahren, die orange Linie das arithmetische Mittel. Die gestrichelten Kurven stellen die Ergebnisse für Kovarianzmatrizen bei AR(1)-Prozessen dar.

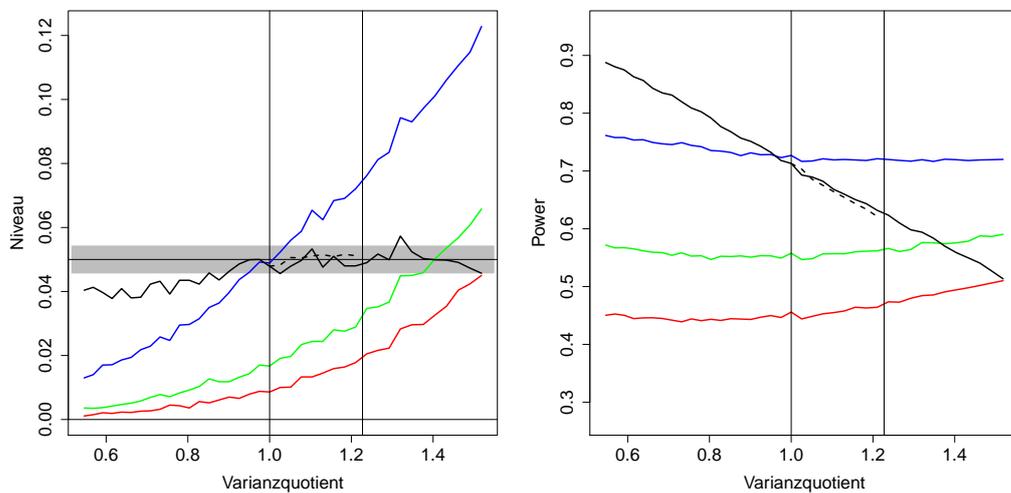


Abbildung 6.4: Niveau (links) und Power (rechts) des t -Tests bei Plan d_6 in Abhängigkeit vom Varianzquotienten $k_{4,12,4}(\mathbf{S})$ für den gewöhnlichen t -Test (blau), die Korrektur mit $k_{4,12,4}^{oben}$ (rot), die Korrektur mit $k_{4,12,4}^*$ (grün) und das MFA-Verfahren (schwarz – durchgezogen für Kovarianzmatrizen aus Konvexkombination, gestrichelt für AR(1)-Prozesse). Der grau hinterlegte Bereich gibt ein 95%-Konfidenzintervall an.

bekanntem \mathbf{S} die kritischen Werte im Bereich $k_{t,n,p} \in (0.8, 1.2)$ kaum größer sind als bei der Variante ohne Freiheitsgradkorrektur. Bei ungünstigsten Kovarianzmatrizen hingegen ist der kritische Wert mit Korrektur deutlich größer als im Fall ohne Korrektur, reicht hier aber nicht über den kritischen Wert des ursprünglichen Verfahrens nach Kunert und Utzig. Anhand der Abbildung könnte man den Eindruck gewinnen, dass beim t-Test der kritische Wert des MFA-Verfahrens bei bekannter Kovarianzmatrix \mathbf{S} im Worst-Case-Szenario gleich dem kritischen Wert bei Korrektur mit $k_{t,n,p}^{oben}$ sei. Wie Abbildung A.3 in Anhang A für den Crossover-Plan d_7 zeigt, ist dies im Allgemeinen nicht der Fall. Dort liegt der kritische Wert des MFA-Verfahrens auch bei ungünstigster Kovarianzmatrix stets deutlich niedriger.

Anhand von Abbildung 6.3 (rechte Seite) erkennt man, dass die kritischen Werte des MFA-Verfahrens auch beim Test des Elementarkontrasts erheblich variieren. Dies hat aber keinen negativen Einfluss auf das Testniveau. Wie man in Abbildung 6.4 (linke Seite) sieht, hält die MFA-Methode das 5% Niveau für alle \mathbf{S} hinreichend gut ein. Verwendet man die Konstante von Kunert und Utzig zur Korrektur der kritischen Werte des t-Tests, so ist der Test im Worst-Case Szenario mit einem empirischen Niveau von 6.58% noch leicht anti-konservativ. Verwendet man die obere Schranke $k_{4,12,4}^{oben}$ zur Korrektur des Tests, so hält man hier zwar das Niveau ein, die Power ist aber sehr niedrig, siehe Abbildung 6.4 (rechte Seite). Wir stellen fest, dass das MFA-Verfahren auch beim t-Test dem Verfahren von Kunert und Utzig überlegen ist.

F-Test und t-Test bei d_7

Die Ergebnisse für den Plan d_6 mit kleinem Stichprobenumfang gelten grundsätzlich auch für Pläne mit größerem Stichprobenumfang. Bei Plan d_7 mit 12 Sequenzwiederholungen ist das Verfahren von Kunert und Utzig wieder stark konservativ, falls günstige Kovarianzstrukturen vorliegen. Im Worst-Case-Szenario hingegen sind sowohl der F-Test als auch der t-Test aus Kapitel 4.2.3 noch immer leicht anti-konservativ (empirisches Niveau von 7.87% beziehungsweise 5.91%). Beim MFA-Verfahren kann \mathbf{S} dank des größeren Stichprobenumfangs wesentlich präziser geschätzt werden. Daher schwanken die kritischen Werte des Verfahrens weniger und die Anpassung an die F-Verteilung gelingt besser. Das Testniveau des F- und t-Tests liegt für alle Kovarianzmatrizen sehr nahe am nominellen Niveau von 5%. Insgesamt lässt sich feststellen, dass das MFA-Verfahren sich sehr gut zur Auswertung von d_7 eignet, während das

Verfahren von Kunert und Utzig nicht zu empfehlen ist. Weitere Details der Ergebnisse der Simulationsstudie zu d_7 entnehme man den Abbildungen A.1 bis A.4 in Anhang A.

6.3 Bemerkungen zu den Verfahren

Das MFA-Verfahren ist dem Verfahren von Kunert und Utzig sowohl im Worst-Case-Szenario als auch bei günstigen Kovarianzmatrizen überlegen. Im Worst-Case-Szenario schätzt das MFA-Verfahren die korrekte Kovarianzmatrix und ist aufgrund der Freiheitsgradapproximation weniger anti-konservativ. Bei kleinem Varianzquotienten ist die Power des Verfahrens mit konservativer Varianzschätzung sehr niedrig. Im Gegensatz dazu ermöglicht das MFA-Verfahren in diesem Fall selbst bei kleinem Stichprobenumfang trotz der unpräzisen Schätzung von \mathbf{S} eine zufriedenstellende Approximation an die F-Verteilung und besitzt eine wesentlich größere Power. Die Korrektur des F-Tests mit $k_{t,n,p}^*$ kann sogar zu schlechteren Ergebnissen führen als die Korrektur mit $k_{t,n,p}^{oben}$, da das verbesserte Verfahren für eine größere Menge von Kovarianzmatrizen zu anti-konservativen Tests führt als das ursprüngliche Verfahren von Kunert und Utzig.

Kunert und Utzig (1993) haben die konservative Varianzschätzung unter anderem auch vorgeschlagen, um konservative Tests auf Behandlungskontraste zu erhalten. Sie zeigen, dass der t-Test (4.6) asymptotisch konservativ ist. Ihre Argumentation lässt sich auch auf den verbesserten t-Test (4.12) übertragen. Die Simulationsstudie deutet darauf hin, dass diese Asymptotik auch bei d_7 mit immerhin $n = 48$ Einheiten bei der ungünstigsten Kovarianzmatrix noch nicht vollständig greift. Das empirische Niveau liegt hier bei 6.58% statt 5%.

Um F- und t-Tests zu erhalten, die bereits bei kleinem Stichprobenumfang und ungünstigster Kovarianzmatrix konservativ sind, könnte man auch die Freiheitsgrade konservativ abschätzen. Ein vielversprechender Ausgangspunkt hierzu wäre die Kombination des Korrekturfaktors $k_{t,n,p}^*$ mit den Freiheitsgraden des MFA-Verfahrens bei ungünstigster Kovarianzmatrix. Damit erhält man dann kritische Werte, die noch größer sind als die Werte auf Grundlage von $k_{t,n,p}^*$ und Tests, die stets höchstens so häufig ablehnen wie das MFA-Verfahren. Dieses Vorgehen hat aber den entscheidenden Nachteil, dass es bei günstigen Kovarianzmatrizen zu Tests mit noch niedrigerer

Power führt als das Verfahren von Kunert und Utzig. Eine konservative Freiheitsgradkorrektur ist daher nicht zu empfehlen.

Wir haben in dieser Arbeit den Fall $p = 2$ nicht betrachtet. Jede Kovarianzmatrix \mathbf{S} ist in diesem Fall vollständig symmetrisch. Der Korrekturfaktor $b^M(\mathbf{S})$ und die Konstante $k_{t,n,2}^*$ nehmen dann stets den Wert 1 an. Die Varianz des Kontrastschätzers wird daher im Mittel nicht unterschätzt. Folglich kann auf die Korrektur der Teststatistiken des F- und t-Tests im Fall $p = 2$ verzichtet werden.

7 Zusammenfassung

In dieser Arbeit haben wir uns mit der Auswertung von Crossover-Studien befasst. Solche Studien sind oft durch kleine Stichprobenumfänge, Nachwirkungen von Behandlungen und korrelierte Beobachtungen gekennzeichnet. Bei unseren Untersuchungen haben wir uns auf Crossover-Pläne zum Vergleich von mindestens drei Behandlungen beschränkt. Weiterhin haben wir Experimente betrachtet, bei denen sich die vorliegenden Daten durch lineare Modelle mit festen Effekten hinreichend gut beschreiben lassen und für die statistische Tests auf Grundlage der Normalverteilung angemessen sind. Einige grundlegende Begriffe hierzu wurden in Kapitel 2 eingeführt.

In Kapitel 3 wurde die Auswertung im Zeilen-Spalten-Modell ohne Berücksichtigung etwaiger Nachwirkungseffekte untersucht. Wir haben gezeigt, dass sich dieses Modell durch die Randomisierung von Einheiten und Behandlungsnamen rechtfertigen lässt, sofern geeignete Versuchspläne verwendet werden und keine Nachwirkungen vorliegen. Dabei konnte die Klasse von Versuchsplänen, für die dieses Resultat gilt, gegenüber Bailey (1985) auf alle regulären verallgemeinerten Youden-Designs erweitert werden (Kunert und Sailer, 2007). Darüberhinaus legt eine Simulationsstudie nahe, dass die Randomisierung von Einheiten und Behandlungsnamen das Zeilen-Spalten-Modell auch dann zumindest approximativ rechtfertigt, wenn näherungsweise balancierte Versuchspläne verwendet werden (Kunert und Sailer, 2006).

Falls Carryover-Effekte vorliegen, so lässt sich die Auswertung unter Vernachlässigung von Carryover-Effekten nicht mehr rechtfertigen. Schätzer von Behandlungs-kontrasten sind dann verzerrt, was sowohl zu stark konservativen als auch zu anti-konservativen Tests führen kann. Aus diesem Grund sollten Modelle benutzt werden, die etwaige Carryover-Effekte explizit berücksichtigen.

Ein intuitiv plausibles Modell in diesem Zusammenhang ist das Modell (2.6), das Einheiteneffekte, Periodeneffekte, Behandlungseffekte, Nachwirkungseffekte und einen Fehlerterm enthält. Oft nimmt man dabei an, dass die Fehler unkorreliert sind (Modell 2.7). Dies ist jedoch für Beobachtungen an einer Einheit nicht immer gegeben. Die Aus-

wertung in Modell (2.7) kann dann zu stark anti-konservativen Tests führen. Kunert und Utzig (1993) haben eine Korrektur für Tests auf Behandlungseffekte vorgeschlagen, die auf einer oberen Schranke für die Varianz der Behandlungskontraste beruht, der Konstanten von Kunert und Utzig. Diese Konstante wird von Kunert und Utzig (1993) für Pläne mit $p = 3$ Perioden exakt bestimmt und für $p > 3$ nach oben abgeschätzt. Ein Problem bei diesem Verfahren besteht darin, dass es bei günstigen Kovarianzmatrizen der Modellfehler zu stark konservativen Tests führt.

Wir haben dieses Verfahren in Kapitel 4 verbessert. Dazu wurde die exakte Konstante von Kunert und Utzig für beliebige vollständig balancierte Pläne hergeleitet. Es zeigt sich, dass die obere Schranke teilweise um mehr als 30% größer ist als die exakte Schranke. Benutzt man unsere exakte Schranke zur Korrektur von Tests auf Behandlungseffekte, so besitzen diese Tests eine deutlich größere Power als Tests auf Grundlage der oberen Schranke von Kunert und Utzig. Die Korrektur mit Hilfe der Konstanten von Kunert und Utzig ist optimal in dem Sinne, dass jeder größere Korrekturfaktor zu Tests geringerer Power führt und jeder kleinere Korrekturfaktor die Varianz der Behandlungskontraste im Worst-Case-Szenario im Mittel unterschätzt.

In Kapitel 5 wurde das MFA-Verfahren von Bellavance et al. (1996) vorgestellt. Dieses Verfahren berücksichtigt mögliche Korrelationen zwischen den Beobachtungen an einer Einheit und wird zum Beispiel von Chen und Wei (2003), Correa und Bellavance (2001) und Jones und Kenward (2003)[S. 262 f.] für die Auswertung von Daten im Modell mit Carryover-Effekten und korrelierten Fehlern vorgeschlagen. Da diese Arbeiten das Verfahren allerdings weder für die zur Schätzung der Varianz des Kontrastschätzers ungünstigste Kovarianzmatrix untersuchen noch mit dem Verfahren von Kunert und Utzig vergleichen, war zu klären, ob unsere Verbesserung des Verfahrens der MFA-Methode vorzuziehen ist.

Wir haben in Kapitel 6 gezeigt, dass der Korrekturfaktor des MFA-Verfahrens bei bekannter Kovarianzmatrix gleich der Konstanten von Kunert und Utzig ist und dass dieser Korrekturfaktor ausgerechnet im Worst-Case-Szenario perfekt geschätzt wird. Damit war zu erwarten, dass das MFA-Verfahren in diesem Fall das Testniveau gut einhält.

Mit Hilfe einer Simulationsstudie haben wir gezeigt, dass das Verfahren von Kunert und Utzig bei ungünstigster Kovarianzmatrix insbesondere bei kleinem, aber auch bei mittelgroßem Stichprobenumfang sogar anti-konservativ ist. Beim MFA-Verfahren hingegen werden auch die Freiheitsgrade korrigiert, so dass dieses Verfahren das Test-

niveau zwar nicht exakt, aber doch hinreichend gut einhält. Letzteres gilt nicht nur im Worst-Case-Szenario sondern im gesamten Spektrum möglicher Kovarianzmatrizen. Dort wo beide Verfahren das Niveau einhalten, hat das MFA-Verfahren eine größere Power als unsere Verbesserung des Verfahrens von Kunert und Utzig. Das MFA-Verfahren ist also das bessere der beiden Verfahren.

Unsere Variante des Verfahrens von Kunert und Utzig ließe sich durch eine konservative Freiheitsgradkorrektur so modifizieren, dass sie tatsächlich auch im Worst-Case-Szenario das Niveau einhielte. Allerdings hätte ein solches Verfahren bei unkorrelierten Beobachtungen eine noch niedrigere Power und wäre somit nicht zu empfehlen.

Zusammenfassend stellen wir fest, dass die Kovarianzmatrix bei der Auswertung im Modell mit Carryover-Effekten und korrelierten Fehlern (2.6) explizit berücksichtigt werden sollte, sofern nicht ausgeschlossen werden kann, dass die Kovarianzmatrix sphärisch ist. Das MFA-Verfahren ist hierfür bereits bei kleinen Stichprobenumfängen gut geeignet. Als Einschränkung ist lediglich zu erwähnen, dass die Schätzung der Kovarianzmatrix im Originalverfahren von Bellavance et al. (1996) Versuchspläne voraussetzt, bei denen jede Behandlungssequenz wiederholt wird. Diese Einschränkung könnte durch Verwendung anderer Kovarianzschätzer aufgehoben werden.

Dem MFA-Verfahren liegt eine sehr flexible Approximation zugrunde, die auf Box (1954) zurückgeht. Das Verfahren ließe sich leicht auf weitere lineare Modelle übertragen. Auch das Verfahren zur Bestimmung der exakten konservativen Varianzabschätzung kann man prinzipiell auf andere Modelle übertragen. So könnte man zum Beispiel die Schranke für zirkulär balancierte Versuchspläne in Kunert (1998a) mit der Methode aus Kapitel 4 verbessern und dann mit einer Adaption des MFA-Verfahrens an dieses Modell vergleichen.

A Tabellen und Abbildungen

Periode	Einheit									
	1	2	3	4	5	6	7	8	9	10
1	0.18	0.35	0.32	0.06	0.49	0.55	0.21	0.48	0.21	0.67
2	0.49	0.71	0.69	0.43	0.72	0.86	0.52	0.67	0.57	1.04
3	1.24	1.23	0.77	0.84	1.56	1.06	1.07	0.73	1.91	1.65
4	2.42	2.53	1.99	3.01	1.91	1.36	1.67	1.07	1.96	2.06
5	3.68	4.43	3.16	10.00	4.88	1.42	3.70	1.50	6.08	3.82

Periode	Einheit									
	11	12	13	14	15	16	17	18	19	20
1	0.88	0.90	0.38	0.23	0.50	0.27	0.37	0.24	0.71	0.45
2	1.16	1.02	0.67	0.47	0.52	0.28	0.87	0.47	1.35	0.80
3	1.23	1.05	1.41	0.48	2.60	0.98	1.25	0.92	3.63	1.20
4	2.18	1.16	1.66	1.46	5.29	2.62	3.10	1.03	5.23	1.78
5	10.00	2.34	3.40	3.38	6.75	2.64	3.53	1.23	5.82	5.84

Periode	Einheit									
	21	22	23	24	25	26	27	28	29	30
1	0.20	1.18	0.37	0.44	0.30	0.71	0.49	0.51	0.33	0.53
2	0.34	1.93	0.41	0.53	0.52	0.83	0.87	0.84	0.52	1.23
3	0.61	2.61	1.70	1.95	1.12	1.50	1.32	0.94	1.12	1.25
4	2.04	6.37	3.49	2.50	2.73	1.81	1.37	2.39	1.75	1.28
5	3.14	8.49	4.70	4.05	2.91	2.29	1.73	3.14	1.87	4.00

Tabelle A.1: *Der Ausgangsdatensatz der Simulationsstudie von Kunert und Sailer (2006) ohne Behandlungseffekte oder Carryover-Effekte.*

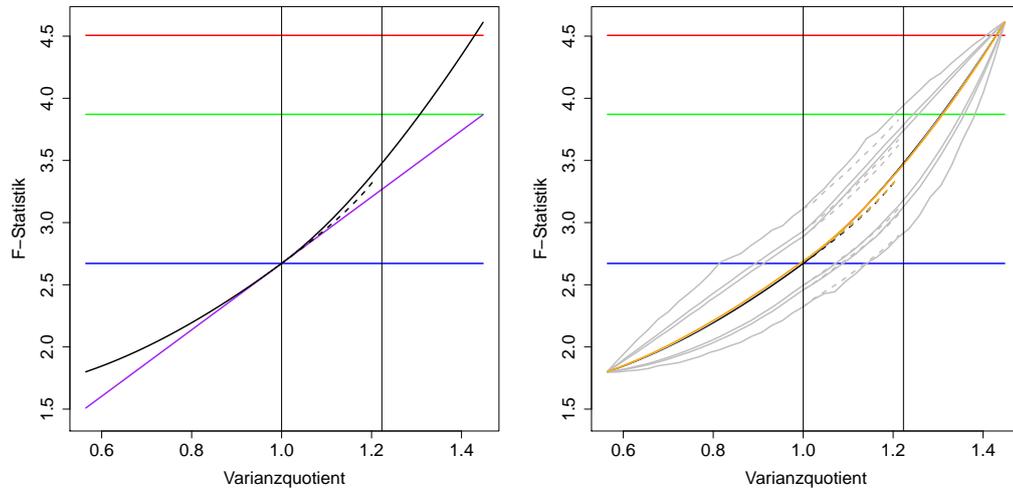


Abbildung A.1: Kritische Werte des F-Tests bei Plan d_7 in Abhängigkeit vom Varianzquotienten $k_{4,48,4}(S)$ bei bekannter Kovarianzmatrix (links) und bei unbekannter Kovarianzmatrix (rechts). Dargestellt werden der gewöhnliche F-Test (blau), die Korrektur mit $k_{4,48,4}^{oben}$ (rot), die Korrektur mit $k_{4,48,4}^*$ (grün), die Korrektur mit $k_{4,48,4}(S)$ (violett) und das MFA-Verfahren bei bekannter Kovarianzmatrix (schwarz). Die grauen Kurven zeigen das 0.1%-, 2.5%- und 5%-Quantil, den Median, das 95%-, 97.5%- und 99.9%-Quantil der empirischen Verteilung der kritischen Werte beim MFA-Verfahren, die orange Linie das arithmetische Mittel. Die gestrichelten Kurven stellen die Ergebnisse für Kovarianzmatrizen bei AR(1)-Prozessen dar.

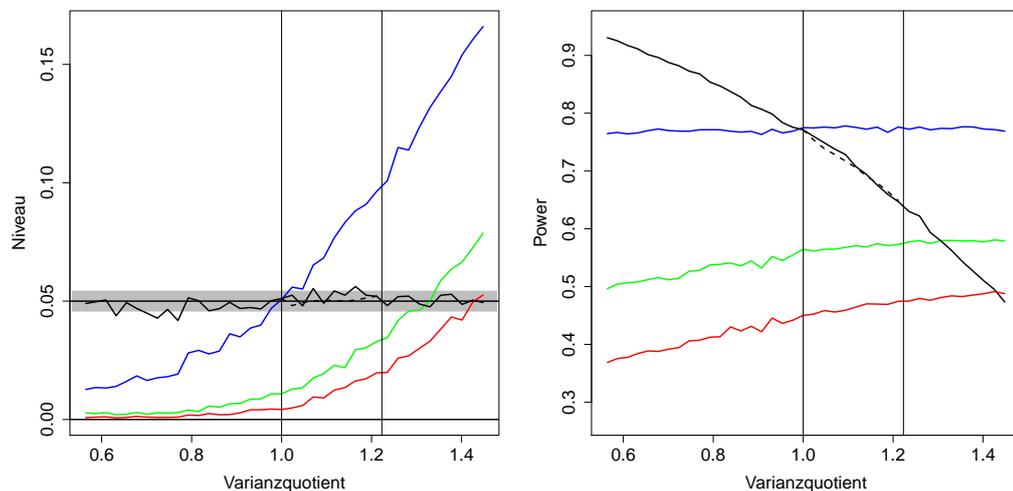


Abbildung A.2: Niveau (links) und Power (rechts) des F-Tests bei Plan d_7 in Abhängigkeit vom Varianzquotienten $k_{4,48,4}(S)$ für den gewöhnlichen F-Test (blau), die Korrektur mit $k_{4,48,4}^{oben}$ (rot), die Korrektur mit $k_{4,48,4}^*$ (grün) und das MFA-Verfahren (schwarz – durchgezogen für Kovarianzmatrizen aus Konvexkombination, gestrichelt für AR(1)-Prozesse). Der grau hinterlegte Bereich gibt ein 95%-Konfidenzintervall an.

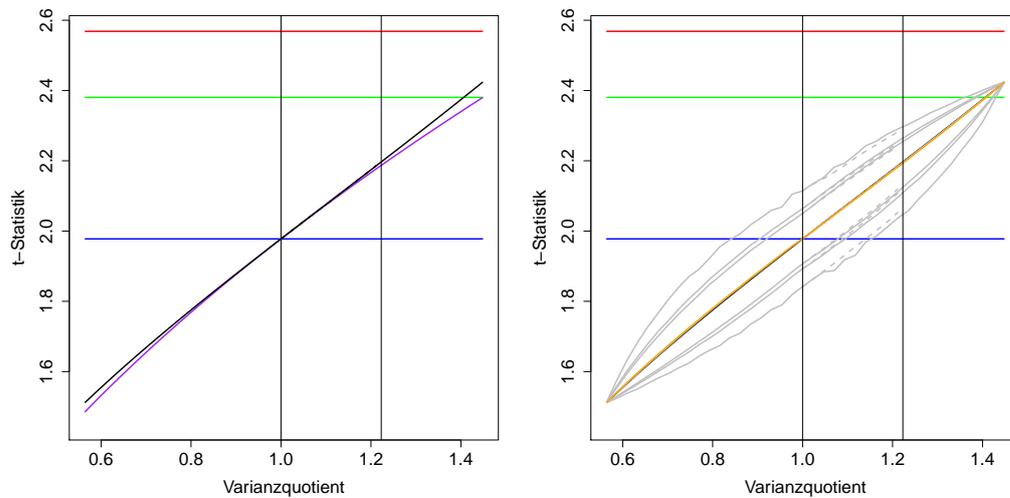


Abbildung A.3: Kritische Werte des t -Tests bei Plan d_7 in Abhängigkeit vom Varianzquotienten $k_{4,48,4}(S)$ bei bekannter Kovarianzmatrix (links) und bei unbekannter Kovarianzmatrix (rechts). Dargestellt werden der gewöhnliche t -Test (blau), die Korrektur mit $k_{4,48,4}^{oben}$ (rot), die Korrektur mit $k_{4,48,4}^*$ (grün), die Korrektur mit $k_{4,48,4}(S)$ (violett) und das MFA-Verfahren bei bekannter Kovarianzmatrix (schwarz). Die grauen Kurven zeigen das 0.1%, 2.5%- und 5%-Quantil, den Median, das 95%, 97.5% und 99.9%-Quantil der empirischen Verteilung der kritischen Werte beim MFA-Verfahren, die orange Linie das arithmetische Mittel. Die gestrichelten Kurven stellen die Ergebnisse für Kovarianzmatrizen bei $AR(1)$ -Prozessen dar.

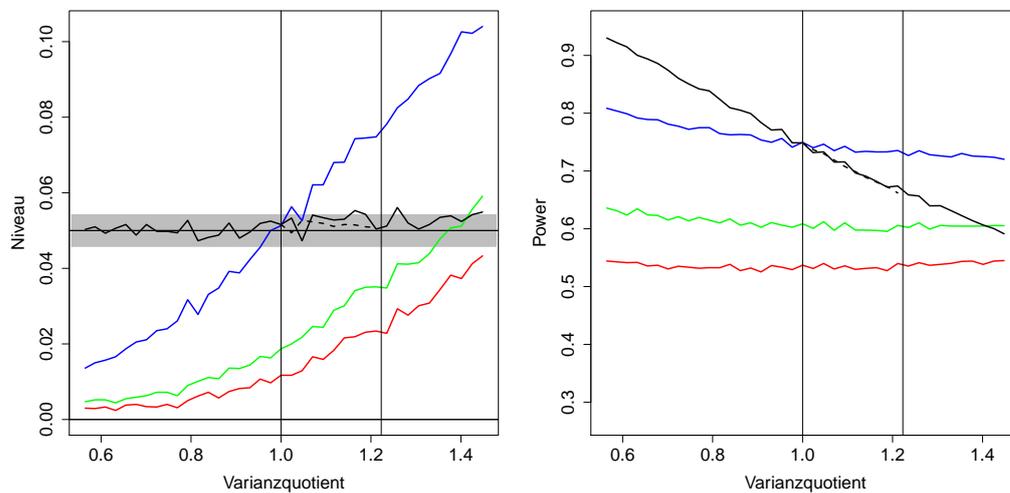


Abbildung A.4: Niveau (links) und Power (rechts) des t -Tests bei Plan d_7 in Abhängigkeit vom Varianzquotienten $k_{4,48,4}(S)$ für den gewöhnlichen t -Test (blau), die Korrektur mit $k_{4,48,4}^{oben}$ (rot), die Korrektur mit $k_{4,48,4}^*$ (grün) und das MFA-Verfahren (schwarz – durchgezogen für Kovarianzmatrizen aus Konvexkombination, gestrichelt für $AR(1)$ -Prozesse). Der grau hinterlegte Bereich gibt ein 95%-Konfidenzintervall an.

B R-Code

Zur Anwendung der im Folgenden aufgeführten R-Programme werden die Zusatzpakete MASS und crossdes benötigt.

B.1 R-Code zu Kapitel 4

Die Funktion `v.ratio` berechnet den Varianzquotienten $k_{t,n,p}^*$ (4.3) für gegebenes \mathbf{S} .

```
v.ratio <- function(trt,n,p,S){
  # trt: number of treatments
  # n: number of units
  # p: number of periods
  # S: covariance matrix
  # output variance ratio
  cf <- n*p-2*trt-n-p+3
  cd <- n*(p-1)*( 1-trt/(p*(p*trt-trt-1)) )/(trt-1)
  gamma0 <- -trt/(p*trt-trt-1)
  one <- rep(1,p)
  Qp <- crossdes:::Q.t(p)
  Qpt <- diag(one) - (1/trt)*one**t(one)
  V <- rbind(numeric(p),cbind(diag(rep(p-1)),0))
  Ainner <- diag(one)-gamma0*(V+t(V))+(gamma0^2)*V**Qpt**t(V)
  Binner <- (cd*(n-1)/n-1)*diag(one)+gamma0*(
    V+t(V)+p*V**Qpt**t(V))
  A <- Qp**Ainner**Qp
  B <- Qp**Binner**Qp
  if( (trt-p)>-0.001 ){
    cf*sum(diag(A**S))/((trt-1)*sum(diag(B**S)))
```

```

}
else{NA}
}

```

Die Funktion `k.star` berechnet die Konstante von Kunert und Utzig $k_{t,n,p}^*$ (4.4).

```

k.star <- function(trt,n,p){
  # trt: number of treatments
  # n: number of units
  # p: number of periods
  # output conservative constant of Kunert and Utzig
  cf <- n*p-2*trt-n-p+3
  cd <- n*(p-1)*( 1-trt/(p*(p*trt-trt-1)) )/(trt-1)
  gamma0 <- -trt/(p*trt-trt-1)
  one <- rep(1,p)
  Qp <- crossdes:::Q.t(p)
  Qpt <- diag(one) - (1/trt)*one%%t(one)
  V <- rbind(numeric(p),cbind(diag(rep(p-1)),0))
  Ainner <- diag(one)-gamma0*(V+t(V))+(gamma0^2)*V%%Qpt%%t(V)
  Binner <- (cd*(n-1)/n-1)*diag(one)+gamma0*(
    V+t(V)+p*V%%Qpt%%t(V))
  A <- Qp%%Ainner%%Qp
  B <- Qp%%Binner%%Qp
  if( ((trt-p)>-0.001)&(all(eigen(Binner)$val >10^(-12)))) ){
    cf*eigen(ginv(B))%%A$val[1]/(trt-1)
  }
  else{NA}
}

```

Die Funktion `k.upper` berechnet die obere Schranke von Kunert und Utzig $k_{t,n,p}^{oben}$ (4.5).

```

k.upper <- function(trt,n,p){
  # trt: number of treatments
  # n: number of units

```

```

# p: number of periods
# output upper bound of variance ratio
cd <- n*(p-1)*( 1-(trt/(p*(p*trt-trt-1))) )/(trt-1)
cd12 <- -n*(p-1)/(p*(trt-1))
cd22 <- n*(p-1)*(p*trt-trt-1)/( p*(trt-1)*trt )
f <- n*p -2*trt -n -p +3
u <- n - 1 -n/cd22
v <- n*(trt+1)/(2*trt*cd22)
w <- 1 + (cd12^2)/(cd22^2)
K <- -cd12/cd22
r <- (cd12^2)*(trt+1)/(2*trt*(cd22^2))
z <- -(K*u + K*v -r*u -v*w)/(K*u) + sqrt(
  (K*u + K*v -r*u -v*w)^2 + ((r*u + v*w)*K*u) )/(K*u)
if(p>4.5){
  a <- ( (trt-1)*cd*( n-1-n/cd22 ) )/( f*n*(
    1-cd12/cd22 )^2 ) - (trt-1)/f }
else{
  if(p>3.5){
    a <- ( (trt-1)*cd*( n-1-n/cd22 ) )/( f*n*(
      1-cd12/cd22 + (cd12^2)/(cd22^2) ) ) - (trt-1)/f
  }
  else{
    a <- ( (trt-1)*cd*( u*((z^2)+z+1)+v ) )/( f*n*(
      w*((z^2)+z+1) -K*z^2 -r ) ) - (trt-1)/f
  }
}
ifelse(1/a>0,1/a,"infty")
}

```

B.2 R-Code zu Kapitel 6

Die Funktion `extreme.matrix` berechnet eine ungünstigste und eine günstigste Kovarianzmatrix.

```

extreme.matrix <- function(trt,n,p){
  # trt: number of treatments
  # n: number of units
  # p: number of periods
  # output list with best and worst case covariance matrix
  cf <- n*p-2*trt-n-p+3
  cd <- n*(p-1)*( 1-trt/(p*(p*trt-trt-1)) )/(trt-1)
  gamma0 <- -trt/(p*trt-trt-1)
  one <- rep(1,p)
  Qp <- crossdes:::Q.t(p)
  Qpt <- diag(one) - (1/trt)*one%%t(one)
  V <- rbind(numeric(p),cbind(diag(rep(p-1)),0))
  Ainner <- diag(one) - gamma0*( V+t(V) ) + (
    gamma0^2)*V%%Qpt%%t(V)
  Binner <- ( cd*(n-1)/n -1 )*diag(one) + gamma0*(
    V + t(V) + p*V%%Qpt%%t(V) )
  A <- Qp%%Ainner%%Qp
  B <- Qp%%Binner%%Qp
  mat1 <- eigen(ginv(B)%%A)$vec[,1]%%t(eigen(ginv(B)%%A)$vec[,1])
  mat2 <- eigen(ginv(B)%%A)$vec[,p-1]%%t(
    eigen(ginv(B)%%A)$vec[,p-1])
  list(mat1,mat2)
}

```

Mit Hilfe dieser Funktion werden Kovarianzmatrizen für die Simulationsstudie erzeugt. Für d_6 wählen wir den folgenden Array von Matrizen. Für d_7 erzeugt man die Kovarianzmatrizen analog.

```

# generate scaled extreme matrices
N1 <- extreme.matrix(4,12,4)[[2]]
M1 <- N1*6/(3*sum(diag(N1))-2*sum(N1[upper.tri(N1)]))
N2 <- extreme.matrix(4,12,4)[[1]]
M2 <- N2*6/(3*sum(diag(N2))-2*sum(N2[upper.tri(N2)]))
Matrixarray1 <- array(0,dim=c(4,4,39))
Matrixarray1[, ,1] <- M1

```

```

Matrixarray1[, ,39] <- M2

# generate scaled matrix with variance ratio of 1
Matrixarray1[, ,20] <- diag(4)*6/(3*sum(diag(diag(4)))-2*sum(
  diag(4)[upper.tri(diag(4))]))

# convex combination of matrices
for(i in 2:19){
  N3 <- (20-i)*Matrixarray1[, ,1]+(i-1)*Matrixarray1[, ,20]
  Matrixarray1[, ,i] <- N3*6/(3*sum(diag(N3))-2*sum(N3[upper.tri(N3)]))
}
for(i in 21:38){
  N3 <- (39-i)*Matrixarray1[, ,20]+(i-20)*Matrixarray1[, ,39]
  Matrixarray1[, ,i] <- N3*6/(3*sum(diag(N3))-2*sum(N3[upper.tri(N3)]))
}

# for simulation use marginally p.d. matrices
Matrixarray1[, ,1] <- (Matrixarray1[, ,1]+.001*diag(4))
Matrixarray1[, ,1] <- Matrixarray1[, ,1]*6/(
  3*sum(diag(Matrixarray1[, ,1]))-2*sum(Matrixarray1[, ,1][upper.tri(
    Matrixarray1[, ,1])]))
Matrixarray1[, ,39] <- (Matrixarray1[, ,39]+.001*diag(4))
Matrixarray1[, ,39] <- Matrixarray1[, ,39]*6/(
  3*sum(diag(Matrixarray1[, ,39]))-2*sum(Matrixarray1[, ,39][upper.tri(
    Matrixarray1[, ,39])]))

```

Daneben betrachten wir Kovarianzmatrizen für AR(1)-Prozesse. Diese werden mit Hilfe der Funktion `AR1cov` für geeignete Werte des Korrelationsparameters erzeugt.

```

AR1cov <- function(p,rho){
  # p: number of periods
  # rho: correlation parameter
  # output covariance matrix
  AR1mat <- matrix(0,p,p)
  for(i in 1:p){

```

```

    for(j in 1:p){
      AR1mat[i,j] <- rho^(abs(i-j))
    }
  AR1mat
}

# generate scaled matrices according to AR(1)
Matrixarray2 <- array(0,dim=c(4,4,9))
Matrixarray2[,,1] <- AR1cov(4,0)
Matrixarray2[,,2] <- AR1cov(4,0.0816)
Matrixarray2[,,3] <- AR1cov(4,0.1688)
Matrixarray2[,,4] <- AR1cov(4,0.2626)
Matrixarray2[,,5] <- AR1cov(4,0.3645)
Matrixarray2[,,6] <- AR1cov(4,0.4762)
Matrixarray2[,,7] <- AR1cov(4,0.6001)
Matrixarray2[,,8] <- AR1cov(4,0.7395)
Matrixarray2[,,9] <- AR1cov(4,0.8994)
for(i in 1:9){
  Matrixarray2[,,i] <- Matrixarray2[,,i]*6/(
    3*sum(diag(Matrixarray2[,,i]))-2*sum(Matrixarray2[,,i][upper.tri(
      Matrixarray2[,,i])]))
}

```

Die Funktion `simulation` simuliert F- und t-Statistiken für den gewöhnlichen KQ-Ansatz und die MFA-Methode im Modell mit Carryover-Effekten und korrelierten Fehlern (2.6) bei Randomisierung von Einheiten und Behandlungsnamen. Die Funktion setzt voraus, dass der Versuchsplan `d` ein vollständig balancierter Plan mit Sequenzwiederholungen ist. Die Einheiten in `d` müssen nach Behandlungssequenzen sortiert sein.

```

simulation <- function(d,nw,mu=0,subj,period,tau,rho,sigma,
  alpha=0.05,niter=10000,dir,filename,seed=5416){
  # d: design matrix with units as rows
  # rows of d must be sorted by treatment sequence
  # nw: number of sequence replications in d

```

```

# mu: mean effect
# subj: subject effect
# period: period effect
# tau: treatment effect
# rho: carryover effect
# sigma: array of covariance matrices
# alpha: test size
# niter: number of iterations in simulation
# dir,filename: directory and filename to save output to
# seed: seed for random number generator
# output OLS F- and t-statistic, F- and t-quantiles for MFA, alpha,
# MFA parameters b, h_1 and h_2 as well as d and seed
trt <- max(d)
nsubj <- dim(d)[[1]]
nperiod <- dim(d)[[2]]
nobs <- nsubj*nperiod
nseq <- nsubj/nw
ncov <- dim(sigma)[[3]]

for(k in 1:ncov){
  eS <- eigen(sigma[, ,k], symmetric = TRUE, EISPACK = TRUE)
  ev <- eS$values
  tol <- 1e-06
  if (!all(ev >= -tol * abs(ev[1])))
    stop("Covariance matrix is not positive definite")
}

# compute design parameters
cd <- nsubj*(nperiod-1)*(1-trt/(nperiod*(
  nperiod*trt-trt-1)))/(trt-1)
cd12 <- -nsubj*(nperiod-1)/(nperiod*(trt-1))
cd22 <- nsubj*(nperiod-1)*(nperiod*trt-trt-1)/(nperiod*trt*(trt-1))
# error d.f.
dfe <- nobs-2*trt-nsubj-nperiod+3

```

```

# \omega^{\bot}_{UP}
woUP <- kronecker( crossdes:::Q.t(nsubj), crossdes:::Q.t(nperiod) )
V <- rbind(0, cbind(diag( nperiod-1 ), 0))
IV <- kronecker(diag(nsubj),V)
# for computation of shat below
matones <- matrix(1,nw,nw)
projshat <- ( diag(nseq*nw)-kronecker(diag(nseq),matones)/nw)/(
  nseq*(nw-1) )

subj <- rep(subj,each=nperiod) # U\alpha
period <- rep(period,nsubj) # P\pi
y0 <- mu + subj + period # y0=1\mu + U\alpha + P\pi

Fstat <- matrix( 0, nr=niter, nc=ncov)
tstat <- matrix( 0, nr=niter, nc=ncov)
b <- matrix(0,niter,ncov)
h1 <- matrix(0,niter,ncov)
h2 <- matrix(0,niter,ncov)
MFA.Fquant <- matrix( 0, nr=niter, nc=ncov)
MFA.tquant <- matrix( 0, nr=niter, nc=ncov)

set.seed(seed)
for(i in 1:niter){

  # randomization
  rand <- randomize(d)
  d1 <- rand[[1]]
  id <- rand[[2]]
  # treatment design matrix
  Td1 <- crossdes:::Td(d1)
  TT <- Td1%*%t(Td1)

  # projector matrices
  woUPF <- woUP - woUP%*%IV%*%TT%*%t(IV)%*%woUP/cd22

```

```

woUPFT <- woUPF - woUPF**TT**woUPF/cd
y1 <- y0 + Td1**tau + IV**Td1**rho
# generate random error
epsilon <- mmvrnorm(nsubj,sigma)
contrastcoeff <- woUPF**(Td1[,1]-Td1[,2])

for(k in 1:ncov){

  # full data vector
  y <- y1 + as.vector(t(epsilon[, ,k]))
  Y <- t(matrix(y,nperiod))
  SSE <- as.numeric(t(y)**woUPFT**y)
  # preliminary F-stat (SST/SSE)
  Fstat[i,k] <- (as.numeric(t(y)**woUPF**y)-SSE)/SSE
  contrast <- sum(contrastcoeff*y)
  # preliminary t-stat
  tstat[i,k] <- contrast / sqrt(SSE)
  # \hat{\Sigma} = I_n \otimes \hat{S}
  shat <- kronecker( diag(nsubj), t(
    Y[order(id),])**projshat**Y[order(id),] )
  # trace of projector matrix * covariance
  woUPFTshat <- woUPFT**shat
  woUPFshat <- woUPF**shat
  trP2S <- sum(diag( woUPFTshat ))
  trP1S <- sum(diag( woUPFshat )) - trP2S
  # preliminary Box-correction factor b
  b[i,k] <- trP1S / trP2S
  h1[i,k] <- trP1S^2 / sum(diag( ( woUPFshat - woUPFTshat )**((
    woUPFshat - woUPFTshat ) ) )
  h2[i,k] <- trP2S^2 / sum(diag( woUPFTshat**woUPFTshat ))
  MFA.Fquant[i,k] <- qf(1-alpha, h1[i,k], h2[i,k])
  MFA.tquant[i,k] <- qt(1-alpha/2, h2[i,k])
}
}

```

```

Fstat <- (dfe/(trt-1))*Fstat
# |t| for standardized elementary contrast btw. treatments 1 and 2
tstat <- abs( sqrt(dfe/(2*cd))*tstat )
b <- (dfe/(trt-1))*b

out <- list(Fstat,tstat,MFA.Fquant,MFA.tquant,alpha,b,h1,h2,d,seed)
save( out, file=file.path( dir, filename) )
}

```

Die Funktion `simulation` nutzt zwei Hilfsfunktionen. Dies ist zum einen der Zufallszahlengenerator `mmvrnorm` für die multivariate Normalverteilung, der auf der Funktion `mvrnorm` des R-Pakets `MASS` (Venables und Ripley, 2002) beruht. Zum anderen nutzen wir die Funktion `randomize`, die Einheiten und Behandlungsnamen permutiert.

```

mmvrnorm <- function(n,Sigma){
  # n: sample size
  # Sigma: array of covariance matrices
  # output array of data from multivariate normal distribution
  p <- dim(Sigma)[[1]]
  ncov <- dim(Sigma)[[3]]
  Z <- matrix(rnorm(p*n),n)
  X <- array(0,c(n,p,ncov))
  for(i in 1:ncov){
    eS <- eigen(Sigma[, ,i],symmetric=TRUE,EISPACK=TRUE)
    ev <- eS$values
    X[, ,i] <- t(eS$vectors%%diag(sqrt(pmax(ev,0)),p)%%t(Z))
  }
  X
}

randomize <- function(d){
  # d: design matrix with units as rows
  # output randomized design and unit id
  perm.n <- sample(nrow(d))
  perm.t <- sample(max(d))
}

```

```

d <- d[perm.n,]
list(matrix(perm.t[d],nc=ncol(d)),perm.n)
}

```

Mit Hilfe der Funktion `evaluate.simulation` wertet man schließlich die Ergebnisse der Simulation aus.

```

evaluate.simulation <- function(dir,filename,kupper,kstar){
  # dir,filename: directory and filename of simulation output
  # kupper: upper bound of variance ratio
  # kstar: conservative constant
  # output list of empirical power for F- and t-tests based on
  # OLS, k.upper, k.star and MFA as well as critical values
  # for the MFA F- and t-Tests

  # load simulation output from file
  load(paste(dir,filename,sep="",collapse=""))

  Fstat <- out[[1]]
  tstat <- out[[2]]
  MFA.Fquant <- out[[3]]
  MFA.tquant <- out[[4]]
  alpha <- out[[5]]
  b <- out[[6]]
  d <- out[[9]]
  trt <- max(d)
  nsubj <- dim(d)[[1]]
  nperiod <- dim(d)[[2]]
  niter <- dim(Fstat)[[1]]
  ncov <- dim(Fstat)[[2]]

  # OLS critical F- and t-values
  dfe <- nsubj*nperiod-2*trt-nsubj-nperiod+3
  Fquant <- qf(1-alpha,trt-1,dfe)
  tquant <- qt(1-alpha/2,dfe)

```

```

# test if null hypothesis is rejected
FtestOLS <- Fstat > Fquant
Ftestkupper <- Fstat > (kupper*Fquant)
Ftestkstar <- Fstat > (kstar*Fquant)
MFA.Fkrit <- b*MFA.Fquant
FtestMFA <- Fstat > MFA.Fkrit
ttestOLS <- tstat > tquant
ttestkupper <- tstat > (sqrt(kupper)*tquant)
ttestkstar <- tstat > (sqrt(kstar)*tquant)
MFA.tkrit <- sqrt(b)*MFA.tquant
ttestMFA <- tstat > MFA.tkrit

# empirical power (or test size under null)
pF1 <- apply( FtestOLS      , 2, sum)/niter
pF2 <- apply( Ftestkupper   , 2, sum)/niter
pF3 <- apply( Ftestkstar    , 2, sum)/niter
pF4 <- apply( FtestMFA      , 2, sum)/niter
pt1 <- apply( ttestOLS      , 2, sum)/niter
pt2 <- apply( ttestkupper   , 2, sum)/niter
pt3 <- apply( ttestkstar    , 2, sum)/niter
pt4 <- apply( ttestMFA      , 2, sum)/niter

out <- list(pF1,pF2,pF3,pF4, pt1,pt2,pt3,pt4, MFA.Fkrit, MFA.tkrit)
out
}

```

Eine Auswertung für den Plan d_6 unter der Alternativhypothese erhält man beispielsweise durch den folgenden R-Code. Alle weiteren Fälle in der Simulationsstudie wertet man analog aus.

```

d6 <- matrix(rep(williams(4),each=3),nc=4)
simulation(d6,3,0,seq(-1,1,12),seq(1,0,le=4)^2,c(0.8,0,0,0),
  c(0.4,0,0,0),Matrixarray1,0.05,10000,"D:\\", "d6-H1.R")
evaluate.simulation("D:\\", "d6-H1.R",k.upper(4,12,4),k.star(4,12,4))

```

Symbolverzeichnis

Seien $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{A}_1 \in \mathbb{R}^{m_1 \times n_1}$ und $\mathbf{A}_2 \in \mathbb{R}^{m_2 \times n_2}$ Matrizen. Dann bezeichne

\mathbf{A}^T	die Transponierte von \mathbf{A} ,
\mathbf{A}^-	eine verallgemeinerte Inverse von \mathbf{A} ,
\mathbf{A}^+	die Moore-Penrose-Inverse von \mathbf{A} ,
\mathbf{A}^{-1}	die Inverse von \mathbf{A} (falls $m = n$ und \mathbf{A} invertierbar),
$[\mathbf{A}_1, \mathbf{A}_2]$	eine partitionierte Matrix $\in \mathbb{R}^{m_1 \times (n_1+n_2)}$ (falls $m_1 = m_2$),
$\mathbf{A}_1 \otimes \mathbf{A}_2$	das Kronecker-Produkt von \mathbf{A}_1 und \mathbf{A}_2 ,
$\text{Rg}(\mathbf{A})$	den Rang von \mathbf{A} ,
$\text{tr}(\mathbf{A})$	die Spur von \mathbf{A} (falls $m = n$),
$\boldsymbol{\omega}_\mathbf{A}$	den orthogonalen Projektor $\mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \in \mathbb{R}^{m \times m}$ und
$\boldsymbol{\omega}_\mathbf{A}^\perp$	den orthogonalen Projektor $\mathbf{I}_m - \boldsymbol{\omega}_\mathbf{A}$.

Wir schreiben

$\mathbf{1}_n$	für den Einsenvektor der Länge n ,
\mathbf{I}_n	für die Einheitsmatrix des \mathbb{R}^n ,
\mathbf{Q}_n	für $\boldsymbol{\omega}_{\mathbf{1}_n}^\perp = \mathbf{I}_n - (1/n)\mathbf{1}_n\mathbf{1}_n^T$,
$\mathbf{Q}_{p,t}$	für $\mathbf{I}_p - (1/t)\mathbf{1}_p\mathbf{1}_p^T$ und
\mathbf{V}	für $\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{I}_{p-1} & \mathbf{0} \end{bmatrix}$.

Für Crossover-Pläne d sei

t	die Anzahl der Behandlungen,
n	die Anzahl der Einheiten,
p	die Anzahl der Perioden,
N	die Gesamtzahl np der Beobachtungen zu d ,
m	den Quotienten n/t (falls $n/t \in \mathbb{N}$),

$\Lambda_{t,n,p}$	die Menge der Crossover-Pläne zum Vergleich von t Behandlungen mit n Blöcken der Länge p und
$\Lambda_{t,n,p}^*$	die Menge der vollständig balancierten Crossover-Pläne zum Vergleich von t Behandlungen mit n Blöcken der Länge p .

In linearen Modellen bezeichne

\mathbf{y}	den Vektor der Beobachtungen $\in \mathbb{R}^N$,
μ	das allgemeine Mittel $\in \mathbb{R}$,
$\boldsymbol{\alpha}$	den Vektor der Einheiten effekte $\in \mathbb{R}^n$,
$\boldsymbol{\pi}$	den Vektor der Perioden effekte $\in \mathbb{R}^p$,
$\boldsymbol{\tau}$	den Vektor der Behandlungseffekte $\in \mathbb{R}^t$,
$\boldsymbol{\rho}$	den Vektor der Nachwirkungseffekte $\in \mathbb{R}^t$,
$\boldsymbol{\epsilon}$	den Fehlervektor $\in \mathbb{R}^N$,
\mathbf{U}	die Designmatrix der Einheiten effekte $\in \mathbb{R}^{N \times n}$,
\mathbf{P}	die Designmatrix der Perioden effekte $\in \mathbb{R}^{N \times p}$,
\mathbf{T}	die Behandlungsdesignmatrix $\in \mathbb{R}^{N \times t}$,
\mathbf{F}	die Designmatrix der Nachwirkungseffekte $\in \mathbb{R}^{N \times t}$,
$\boldsymbol{\psi} = \boldsymbol{\ell}^T \boldsymbol{\tau}$	einen Behandlungskontrast,
$\hat{\boldsymbol{\tau}}$	den KQ-Schätzer für die Behandlungseffekte $\boldsymbol{\tau}$,
$\hat{\boldsymbol{\psi}}$	den KQ-Schätzer für den Behandlungskontrast $\boldsymbol{\psi}$,
$\hat{\boldsymbol{\epsilon}}$	den Vektor der KQ-Residuen und
f_E	die Freiheitsgrade zum Schätzen des Fehlers.

Für die Zufallsvariable x_1 und den Zufallsvektor \boldsymbol{x} sei

$E(\boldsymbol{x})$	der Erwartungswert von \boldsymbol{x} ,
$\text{var}(x_1)$	die Varianz von x_1 und
$\text{Cov}(\boldsymbol{x})$	die Kovarianzmatrix von \boldsymbol{x} .

Speziell in Kapitel 3 schreiben wir

d_0	für einen Ausgangsplan vor der Randomisierung,
δ	für einen randomisierten Versuchsplan,
$\boldsymbol{\eta}$	für den von Behandlungen und Nachwirkungen unabhängigen Anteil der Beobachtungen \mathbf{y} ,

\mathbf{H}_d	für die Behandlungsdesignmatrix von d_0 ,
\mathbf{D}_d	für die Informationsmatrix $\mathbf{H}_d(\mathbf{Q}_n \otimes \mathbf{Q}_p)\mathbf{H}_d$ des Plans d_0 ,
$\mathbf{\Pi}$	für eine Permutationsmatrix $\in \mathbb{R}^{n \times n}$,
$\mathbf{\Gamma}$	für eine Permutationsmatrix $\in \mathbb{R}^{t \times t}$,
\mathbf{C}_δ	für die Informationsmatrix des randomisierten Plans δ ,
\mathbf{T}_δ	für die Behandlungsdesignmatrix von δ ,
\mathbf{F}_δ	für die Designmatrix der Nachwirkungseffekte von δ und
\mathbf{Z}_η	für die Matrix $\mathbf{Q}_n \otimes (\mathbf{Q}_p(\mathbf{M} - \tilde{\mathbf{M}})\mathbf{Q}_p)$.

Wir kennzeichnen in diesem Zusammenhang Erwartungswerte und Varianzen unter Randomisierung mit dem Index δ .

Es bezeichne

$\mathbf{\Sigma}$	die Kovarianzmatrix des Beobachtungsvektors \mathbf{y} ,
\mathbf{S}	die Kovarianzmatrix der Beobachtungen an einer Einheit,
$\tilde{\mathbf{S}}$	die zentrierte Kovarianzmatrix $\mathbf{Q}_p\mathbf{S}\mathbf{Q}_p$,
$\hat{\mathbf{S}}$	den Schätzer für die Kovarianzmatrix beim MFA-Verfahren und
\mathcal{S}	die Menge aller n.n.d. Matrizen des $\mathbb{R}^{p \times p}$.

In Kapitel 4 und 6 sei

γ	$= -t/(pt - t - 1)$,
\mathbf{A}_d	$= \mathbf{I}_p - \gamma(\mathbf{V} + \mathbf{V}^T) + \gamma^2\mathbf{V}\mathbf{Q}_{p,t}\mathbf{V}^T$,
$\tilde{\mathbf{A}}_d$	$= \mathbf{Q}_p\mathbf{A}_d\mathbf{Q}_p$,
\mathbf{B}_d	$= (c_d(n-1)/n-1)\mathbf{I}_p + \gamma(\mathbf{V} + \mathbf{V}^T) + p\gamma\mathbf{V}\mathbf{Q}_{p,t}\mathbf{V}^T$ und
$\tilde{\mathbf{B}}_d$	$= \mathbf{Q}_p\mathbf{B}_d\mathbf{Q}_p$.

Im Zusammenhang mit der Korrektur von F- und t-Test bezeichne

$k_{t,n,p}(\mathbf{S})$	den Varianzquotienten,
$k_{t,n,p}^{oben}$	die obere Schranke von Kunert und Utzig,
$k_{t,n,p}^*$	die Konstante von Kunert und Utzig,
$b^M(\mathbf{S})$	den Korrekturfaktor des MFA-Verfahrens und
h_1^M, h_2^M	die Freiheitsgradapproximation des MFA-Verfahrens.

Wir kennzeichnen Erwartungswerte, Varianzen, Varianzschätzer und Kovarianzmatrizen unter Modell (2.6) mit dem Index \mathbf{S} und unter Modell (2.7) mit dem Index \mathbf{I} .

Schließlich schreiben wir

H_0^g	für die Globalhypothese $\tau_1 = \dots = \tau_t$ des F-Tests,
H_1^g	für die Alternativhypothese $\exists i \neq j : \tau_i \neq \tau_j$,
H_0^k	für die Nullhypothese $\psi = \psi_0$ des t-Tests,
H_1^k	für die Alternativhypothese $\psi \neq \psi_0$,
$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	falls \mathbf{x} multivariat normalverteilt ist mit Parametern $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$,
$x_1 \sim \chi_\nu^2$	falls x_1 χ^2 -verteilt ist mit ν Freiheitsgraden,
$F_{f_1, f_2; \nu}$	für das ν -Quantil der F -Verteilung mit f_1 und f_2 Freiheitsgraden und
$t_{f; \nu}$	für das ν -Quantil der t -Verteilung mit f Freiheitsgraden.

Literaturverzeichnis

- AZAÏS, J. M. (1987): Design of experiments for studying intergenotypic competition. *Journal of the Royal Statistical Society B*, **49**: 334–345.
- AZAÏS, J. M. und DRUILHET, P. (1997): Optimality of neighbour balanced designs when neighbour effects are neglected. *Journal of Statistical Planning and Inference*, **64**: 353–367.
- BAILEY, R. A. (1981): A unified approach to design of experiments. *Journal of the Royal Statistical Society A*, **144**: 214–223.
- BAILEY, R. A. (1985): Restricted randomization for neighbour-balanced designs. *Statistics and Decisions Supplement*, **2**: 237–248.
- BAILEY, R. A. und ROWLEY, C. A. (1987): Valid randomization. *Proceedings of the Royal Society London A*, **410**: 105–124.
- BAKKE, A. und VICKERS, Z. (2008): Relationships between fungiform papillae density, prop sensitivity and bread roughness perception. *Journal of Texture Studies*, **39**: 569–581.
- BALL, R. D. (1997): Incomplete block designs for the minimisation of order and carry-over effects in sensory analysis. *Food Quality and Preference*, **8**: 111–118.
- BATE, S. T. und JONES, B. (2006): The construction of nearly balanced and nearly strongly balanced uniform cross-over designs. *Journal of Statistical Planning and Inference*, **136**: 3248–3267.
- BATE, S. T. und JONES, B. (2008): A review of uniform cross-over designs. *Journal of Statistical Planning and Inference*, **138**: 336–351.
- BELLAVANCE, F. und TARDIF, S. (1995): A nonparametric approach to the analysis of three-treatment three-period crossover designs. *Biometrika*, **82**: 865–875.
- BELLAVANCE, F., TARDIF, S. und STEPHENS, M. A. (1996): Tests for the analysis of variance of crossover designs with correlated errors. *Biometrics*, **52**: 607–612.
- BOX, G. E. P. (1954): 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.

- BROCKWELL, P. J. und DAVIS, R. A. (2002): *Introduction to time series and forecasting*. Springer, New York, 2. Auflage.
- CHEN, X. und WEI, L. (2003): A comparison of recent methods for the analysis of small-sample cross-over studies. *Statistics in Medicine*, **22**: 2821–2833.
- COCHRAN, W. G., AUTREY, K. M. und CANNON, C. Y. (1941): A double change-over design for dairy cattle feeding experiments. *Journal of Dairy Science*, **24**: 937–951.
- CORREA, J. A. und BELLAVANCE, F. (2001): Power comparison of robust approximate and non-parametric tests for the analysis of cross-over trials. *Statistics in Medicine*, **20**: 1185–1196.
- GEISSER, S. und GREENHOUSE, S. W. (1958): An extension of box's results on the use of the f distribution in multivariate analysis. *Annals of Mathematical Statistics*, **29**: 885–891.
- GRAYBILL, F. A. (1969): *Introduction to matrices with applications in statistics*. Wadsworth, Belmont, CA.
- HARKER, F. R., AMOS, R. L., WHITE, A., PETLEY, M. B. und WOHLERS, M. (2008): Flavor differences in heterogeneous foods can be detected using repeated measures of consumer preferences. *Journal of Sensory Studies*, **23**: 52–64.
- HEDAYAT, A. und AFSARINEJAD, K. (1978): Repeated measurements designs, II. *The Annals of Statistics*, **6**: 619–628.
- HORN, R. und JOHNSON, C. (1985): *Matrix analysis*. Cambridge University Press, Cambridge.
- HUYNH, H. und FELDT, L. S. (1970): Conditions under which mean square ratios in repeated measurement designs have exact f-distributions. *Journal of the American Statistical Association*, **65**: 1582–1589.
- HUYNH, H. und FELDT, L. S. (1976): Estimation of the box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics*, **1**: 69–82.
- JONES, B. und KENWARD, M. G. (1989): *Design and Analysis of Cross-Over Trials*, volume 34 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.
- JONES, B. und KENWARD, M. G. (2003): *Design and Analysis of Cross-over Trials*. Chapman & Hall, 2. Auflage.

- KEMPE, R., SAASTAMOINEN, M., HYYPPÄ, S. und SMEDS, K. (2004): Composition, digestibility and nutritive value of cereals for dogs. *Agricultural and Food Science*, **13**: 5–17.
- KENWARD, M. G. und ROGER, J. H. (1997): Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, **53**: 983–997.
- KIEFER, J. (1975): Construction and optimality of generalized youden designs. In: SRIVASTAVA, J. N. (Hrsg.), *A survey of statistical designs and linear models*, S. 333–353, North-Holland.
- KUNERT, J. (1984): Optimality of balanced uniform repeated measurements designs. *Annals of Statistics*, **12**: 1006–1017.
- KUNERT, J. (1987): On variance estimation in crossover designs. *Biometrics*, **43**: 833–845.
- KUNERT, J. (1998a): On the analysis of circular balanced crossover designs. *Journal of Statistical Planning and Inference*, **69**: 359–370.
- KUNERT, J. (1998b): Sensory experiments as crossover studies. *Food Quality and Preference*, **9**: 243–253.
- KUNERT, J., MEYNEERS, M. und ERDBRÜGGE, M. (2002): On the applicability of anova for the analysis of sensory data. In: DE STATISTIQUE, S. F. (Hrsg.), *7è Journées Européennes Agro-Industrie et Méthodes Statistiques*, S. 129–134.
- KUNERT, J. und SAILER, O. (2006): On nearly balanced designs for sensory trials. *Food Quality and Preference*, **17**: 219–227.
- KUNERT, J. und SAILER, O. (2007): Randomization of neighbour balanced generalized youden designs. *Journal of Statistical Planning and Inference*, **137**: 2045–2055.
- KUNERT, J. und UTZIG, B. P. (1993): Estimation of variance in cross-over designs. *Journal of the Royal Statistical Society B*, **55**: 919–927.
- KUOPPALA, K., RINNE, M., NOUSIAINEN, J. und HUHTANENA, P. (2008): The effect of cutting time of grass silage in primary growth and regrowth and the interactions between silage quality and concentrate level on milk production of dairy cows. *Livestock Science*, **116**: 171–182.
- KUSHNER, H. B. (1997): Optimal repeated measurements designs: The linear optimality equations. *Annals of Statistics*, **25**: 2328–2344.
- LAUREATI, M., PAGLIARINI, E. und CALCINONI, O. (2008): Does the enhancement of chemosensory stimuli improve the enjoyment of food in institutionalized elderly people? *Journal of Sensory Studies*, **23**: 234–250.

- MCNULTY, P. (1986): *Spatial Velocity Induction and Reference Mark Density*. Doktorarbeit, University of California, Santa Barbara.
- ÖHRVIK, J. (1998): Nonparametric methods in crossover trials. *Biometrical Journal*, **40**: 771–789.
- PATTERSON, H. D. (1952): The construction of balanced designs for experiments involving sequences of treatments. *Biometrika*, **39**: 32–48.
- PÉRINEL, E. und PAGÈS, J. (2004): Optimal nested cross-over designs in sensory analysis. *Food Quality and Preference*, **15**: 439–446.
- R DEVELOPMENT CORE TEAM (2009): *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Wien, URL <http://www.R-project.org>.
- RAO, C., TOUTENBURG, H., SHALABH und HEUMANN, C. (2008): *Linear Models and Generalizations: Least Squares and Alternatives*. Springer, Berlin, 3. Auflage.
- SAILER, O. (2005): crossdes: A package for design and randomization in crossover studies. *Rnews*, **5**: 24–27.
- SEBER, G. A. F. (2008): *A matrix handbook for statisticians*. Wiley.
- SENN, S. (2002): *Cross-over Trials in Clinical Research*. Wiley, 2. Auflage.
- SOLOMON, H. und STEPHENS, M. A. (1978): Approximations to density functions using pearson curves. *Journal of the American Statistical Association*, **73**: 153–160.
- STUFKEN, J. (1996): Optimal crossover designs. In: GHOSH, S. und RAO, C. R. (Hrsg.), *Handbook of Statistics, Vol. 13*, S. 63–90, Elsevier Science B.V.
- TARDIF, S., BELLAVANCE, F. und VAN EEDEN, C. (2005): A nonparametric procedure for the analysis of balanced crossover designs. *Canadian Journal of Statistics*, **33**: 471–488.
- UTZIG, B. P. (1990): *Analyse von Cross-over Modellen mit korrelierten Fehlern*. Diplomarbeit, Universität Trier, Trier.
- VENABLES, W. N. und RIPLEY, B. D. (2002): *Modern Applied Statistics with S*. Springer, New York, 4. Auflage.
- WAKELING, I. N. und MACFIE, H. J. H. (1995): Designing consumer trials for first and higher orders of carry-over effects when only a subset of k samples from p may be tested. *Food Quality and Preference*, **6**: 299–308.

WELCH, B. (1937): The significance of the difference between two means when the population variances are unequal. *Biometrika*, **29**: 350–362.

WILLIAMS, E. J. (1949): Experimental designs balanced for the estimation of residual effects. *Australian Journal of Scientific Research A*, **2**: 149–168.

WITTING, H. (1985): *Mathematische Statistik I*. Teubner, Stuttgart.