

Das Behrens-Fisher-Problem für hochdimensionale Split-Plot-Designs

Markus Harden
 Abteilung Medizinische Statistik
 Universitätsmedizin Göttingen
 Humboldtallee 32
 Göttingen
 Markus.Harden@med.uni-goettingen.de

Zusammenfassung

HINTERGRUND: In vielen Anwendungsbeispielen aus Biologie und Medizin werden statistische Analysemethoden benötigt, um hochdimensionale Datensätze auswerten zu können. Solche Versuchsdesigns beinhalten weniger Probanden als Messungen, die an den Probanden erhoben wurden. Unter den Annahmen normalverteilter Daten und einer endlichen Anzahl von Dimensionen existieren verschiedene Ansätze für Approximationen, die dieses Problem lösen. Standardverfahren wie Hotellings T^2 sind aufgrund einer singulären Kovarianzmatrix nicht anwendbar. Da keine exakte Lösung dieses Problems vorliegt, basieren alle Lösungsansätze auf Approximationen, die von Geisser-Greenhouse (GG) oder Huynh-Feldt (HF) beschrieben wurden. Diese Approximationen können jedoch zu falschen Entscheidungen führen, falls ungleiche Kovarianzmatrizen vorliegen. Zusätzlich wird die Qualität der Approximationen durch ungleiche Stichprobenumfänge mitunter massiv verschlechtert. Die Abweichung von der zuvor festgelegten Fehlerrate 1. Art nimmt zu, wenn die Anzahl der Dimensionen steigt.

METHODEN: Basierend auf einer Momenten-Approximation von Box und Patnaik, wurde die Anova-Typ-Statistik (Brunner, Dette, & Munk, Box-type approximations in nonparametric factorial designs, 1997) auf hochdimensionale Split-Plot-Designs mit $a \geq 2$ Behandlungsgruppen angewendet (Box, 1954) (Patnaik, 1949). Die Schätzer für das Box'sche ϵ für den Ein- und Zwei-Stichprobenfall aus (Brunner, Repeated Measures under Non-Sphericity, 2009) werden für Split-Plot-Designs mit Repeated Measures verallgemeinert. Diese Schätzer sind dimensionsstabil, was bedeutet, dass die Qualität der Approximation nicht von der Anzahl der Dimensionen abhängt. Für Compound Symmetry (CS) und autoregressive (AR) Kovarianzstrukturen zeigen die Simulationen bessere Ergebnisse des beschriebenen Verfahrens, als die Approximation von GG und HF.

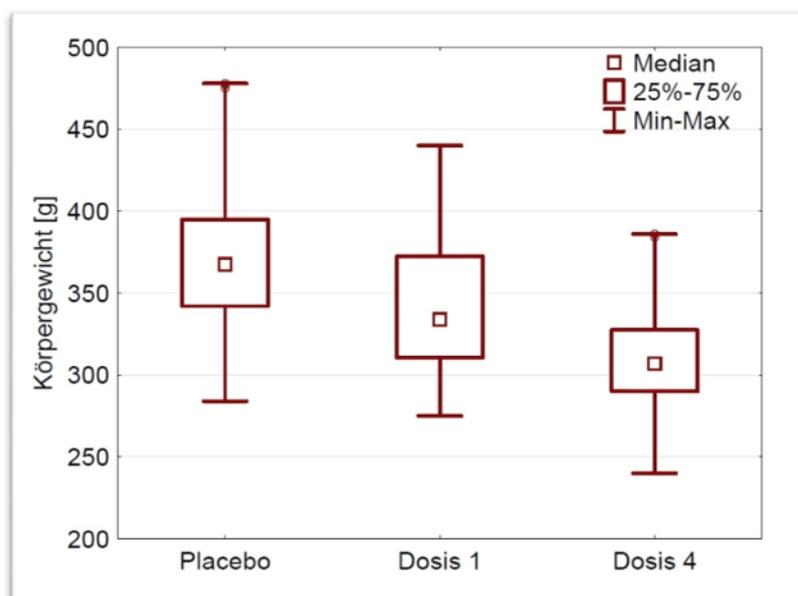
ERGEBNISSE: Die in diesem Vortrag vorgestellte Approximation führt zu besseren Ergebnissen als die in PROC GLM (SAS Version 9.3) enthaltenen Lösungsansätze, wie die von GG und HF. Das SAS-Makro F1_HF_F1 stellt daher eine Alternative zu den implementierten Approximationen dar, um Split-Plot-Designs mit hochdimensionalen Repeated Measures auszuwerten.

Schlüsselwörter: Ungleiche Kovarianzmatrizen, Box'sches Epsilon, GLM-Prozedur, Makro F1_HF_F1

Tabelle 2: Deskriptive Kennzahlen der gemessenen Körpergewichte in [g]

Gruppe	Anzahl	Mittelwert	Median	Std	Min	Max
Placebo	286	369,38	367,50	39,42	284	478
Dosis 1	264	342,10	334,00	40,73	275	440
Dosis 4	528	309,33	307,00	29,00	240	386

In diesem Beispiel werden 3 Dosisgruppen der männlichen Parental-Generation aus dem Jahr 1981 betrachtet. Die Gewichtsentwicklung der $N = 49$ Tiere wurde über $d = 22$ Zeitpunkte betrachtet. Die ungleiche Anzahl der Tiere je Dosisgruppe ($n_0 = 13$, $n_1 = 12$ und $n_4 = 24$) ist darauf zurückzuführen, dass die Dosisstufe 4 erst im 2. Jahr der Studie betrachtet wurde und aus diesem Grund die doppelte Anzahl der Tiere herangezogen wurde, um über die Jahre hinweg ein annähernd balanciertes Versuchsdesign auswerten zu können. Einige deskriptive Kennzahlen sind für die drei Dosisstufen in Tabelle 2 aufgeführt. Man erkennt, dass die Mittelwerte über die $d = 22$ Zeitpunkte mit steigender Dosis abnehmen. Die Varianzen in den Gruppen Placebo und Dosis 1 sind ähnlich groß, während die empirische Varianz in der Gruppe Dosis 4 deutlich kleiner ist. Der Zeitverlauf des mittleren Körpergewichtes ist in der Abbildung 1 dargestellt. Der Eindruck abnehmender mittlerer Körpergewichte mit steigender Dosis lässt sich durch diese Mittelwertplots bestätigen. Die Verteilung der Daten ist in Abbildung 2 dargestellt und lässt sich gut durch eine Normalverteilung approximieren. Dafür spricht, dass die Daten auf einer stetigen Skala gemessen wurden und die Verteilung der Messwerte in annähernd symmetrisch ist.

**Abbildung 2:** Boxplot des gemessenen Körpergewichtes, unterteilt nach Behandlung.

1 Einleitung

1.1 Motivation

Bei einem Split-Plot-Design mit Repeated Measures handelt es sich um einen Versuchsaufbau, bei dem an unabhängigen, auf mehrere Behandlungsgruppen aufgeteilten Versuchsobjekten wiederholt dieselben Messungen vorgenommen werden. Zwischen diesen Messungen besteht, sofern sie zu demselben Versuchsobjekt gehören, eine Abhängigkeit, die bei der späteren statistischen Modellierung und der Wahl des Auswertungsverfahrens berücksichtigt werden muss. Die Bezeichnung Split-Plot kommt ursprünglich aus der Agrarwissenschaft, in der mehrere Anbaufelder (sog. Wholeplot-Faktor) in Parzellen (sog. Subplot-Faktor) unterteilt und diese beispielsweise mit unterschiedlichem Saatgut bepflanzt wurden. Ein solches Split-Plot-Design ist exemplarisch in Tabelle 1 dargestellt. Unter der Annahme, dass es sich bei den gemessenen Daten um Realisationen normalverteilter Zufallsvariablen handelt, existieren zahlreiche sowohl uni- als auch multivariate Verfahren, die sich für die Auswertung eines Split-Plot-Designs anbieten. Zu jedem dieser Verfahren gibt es bestimmte Voraussetzungen, die für die Anwendung erfüllt sein müssen, um zu verlässlichen Ergebnissen zu gelangen.

Tabelle 1: Versuchsaufbau eines zweifaktoriellen Split-Plot-Designs für a Behandlungsgruppen mit je n_i unabhängigen Versuchsobjekten, an denen d abhängige Messungen durchgeführt werden.

Gruppe	Individuum	Repeated Measures			
		1	...	d	
1	1	X_{111}	...	X_{11d}	
	\vdots	\vdots	...	\vdots	
	n_1	X_{1n_11}	...	X_{1n_1d}	
\vdots	\vdots	\vdots	...	\vdots	
a	1	X_{a11}	...	X_{a1d}	
	\vdots	\vdots	...	\vdots	
	n_a	X_{an_a1}	...	X_{an_ad}	

Für die klassische Varianzanalyse ist die Annahme eines homoskedastischen Modells gefordert. Die multivariate Teststatistik von Hotelling verlangt, dass die Anzahl der Versuchsobjekte größer als die Anzahl der Versuchswiederholungen ist ($n_i > d$, $i = 1, \dots, a$), da die inverse empirische Kovarianzmatrix berechnet werden muss, die im Falle $n_i < d$ singulär ist (Hotelling, 1931).

In Statistikprogrammen wie SAS wird für die Auswertung eines solchen Versuchsplans eine Adjustierung der Freiheitsgrade aus der Varianzanalyse mit einem Korrekturterm vorgeschlagen (SAS Institute Inc., 2012): In PROC GLM der SAS Version 9.3 sind die Approximationen von Geisser-Greenhouse, Huynh-Feldt und Lecoutre implementiert, wobei die Arbeit von Lecoutre einen Fehler aus der Publikation von Huynh-Feldt berichtigt (vgl. (Geisser & Greenhouse, 1958), (Huynh & Feldt, 1976), (Lecoutre, 1991)).

Eine wesentliche Einschränkung dieser Verfahren liegt in der restriktiven Annahme gleicher Kovarianzmatrizen in den Behandlungsgruppen (Referenzen ebenda). In einfachen Simulationsstudien kann gezeigt werden, dass das Fehlerniveau dieser Approximationen nicht mehr eingehalten wird, falls diese Annahme verletzt ist. Dieser Effekt wird noch verstärkt, wenn ungleiche Stichprobenumfänge vorliegen. Aus dieser Problematik resultierte die Motivation ein approximatives Testverfahren zu entwickeln, mit dem hochdimensionale Split-Plot-Designs für beliebig viele Behandlungsgruppen bei einer festen Anzahl an Dimensionen d ausgewertet werden können, ohne dabei im Vorfeld wesentliche Annahmen über die Kovarianzmatrizen zu treffen. Als Ausgangspunkt dient die Arbeit von (Brunner, Repeated Measures under Non-Sphericity, 2009), der ein solches Verfahren bereits für den Ein- bzw. Zwei-Stichprobenfall präsentiert hat. Dieser Ansatz wurde mit dieser Arbeit auf $a \geq 2$ Stichproben verallgemeinert.

1.2 Beispiel: Gewichtsentwicklung bei Wistar-Ratten

In einer breit angelegten Teratogenitäts- und Fertilitätsstudie an Wistar-Ratten in den Jahren 1980 und 1981 wurde unter anderem bei Tieren der Parental-Generation untersucht, ob die tägliche orale Behandlung mit einer Testsubstanz in vier Dosisstufen zu signifikanten Unterschieden in der Körpergewichtsentwicklung führt. Dazu wurde das Körpergewicht der Versuchstiere zu $d = 22$ Zeitpunkten bestimmt. Die Testsubstanz wurde den Tieren täglich mit starrer Schlucksonde in den Dosen 0,1 mg/ 100 g Körpergewicht, 0,8 mg/ 100 g Körpergewicht, 6,4 mg/ 100 g Körpergewicht sowie 12,8 mg/ 100 g Körpergewicht verabreicht. Hierbei entsprechen 0,1 mg/ 100 g Körpergewicht der humantherapeutischen Tagesdosis des Wirkstoffes. Zusätzlich wurde eine Gruppe betrachtet, der anstelle des Verum ein Placebo, in diesem Fall 1 ml Leitungswasser je 100 g Körpergewicht, verabreicht wurde.

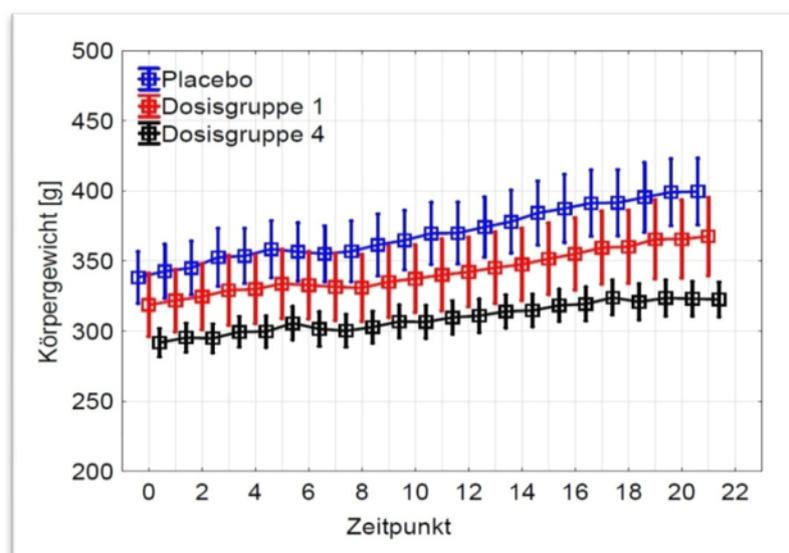


Abbildung 1: Entwicklung des mittleren Körpergewichtes in [g] über die Zeit (Obere Kurve Placebo, untere Kurve Dosis 4)

2 Modelle und Hypothesen

In diesem Abschnitt werden das statistische Modell beschrieben und die Hypothesen formuliert, die aus einem allgemeinen Split-Plot-Design mit den zwei festen Faktoren Behandlung und Zeit resultieren. Es werden N Versuchsobjekte betrachtet, von denen $n_i > 0$ Individuen der i -ten von insgesamt a Behandlungsgruppen zugeordnet werden. An jedem Versuchsobjekt erfolgen wiederholt zu d Zeitpunkten Messungen. Die Anzahl an Behandlungen a und Messwiederholungen d wird als fest und damit reproduzierbar angesehen. Es bezeichne $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{id})' \in \mathbb{R}^d$ den Erwartungsvektor und $\boldsymbol{\Sigma}_i \in \mathbb{R}^{d \times d}$ die positiv definite Kovarianzmatrix der Versuchsobjekte aus der i -ten Behandlungsgruppe. Unter der Annahme multivariat normalverteilter Daten lässt sich jeder Realisierungsvektor der $N = \sum_{i=1}^a n_i$ Versuchsobjekte schreiben als

$$\mathbf{X}_{ij} \sim N_d(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i); i = 1, \dots, a; j = 1, \dots, n_i$$

2.1 Statistisches Modell

Der Faktor Behandlung ist mit dem Faktor Zeit gekreuzt, die Versuchsobjekte sind unter der Behandlung verschachtelt und mit der Zeit gekreuzt. Es handelt sich um ein Split-Plot-Design mit dem Wholeplot-Faktor Behandlung und dem Subplot-Faktor Zeit. Die in diesem Modell auftretenden Effekte sind die der Faktoren Behandlung, Zeit und die der Wechselwirkung aus Behandlung und Zeit. In der Sprache der linearen Modelle lässt sich jede Beobachtung in die im Modell modellierten Effekte zerlegen. Das lineare Modell hat die Gestalt

$$X_{ijk} = \mu + \alpha_i + \beta_k + (\alpha\beta)_{ik} + \epsilon_{ijk}$$

mit $\epsilon_{ij} \sim N_d(\mathbf{0}, \boldsymbol{\Sigma}_i), i = 1, \dots, a; j = 1, \dots, n_i; k = 1, \dots, d$. An dieser Stelle bezeichnet μ den Gesamteffekt, α_i den i -ten Behandlungseffekt, β_k den k -ten Zeiteffekt und $(\alpha\beta)_{ik}$ den Einfluss der Wechselwirkung der (i,k) -ten Faktorkombination. Diese Effekte lassen sich ebenso durch die Erwartungsvektoren aus (2.1) darstellen. Für die Summen der jeweiligen Effekte gilt: $\sum_{i=1}^a \alpha_i = \sum_{k=1}^d \beta_k = \sum_{i,k} (\alpha\beta)_{ik} = 0$.

2.2 Hypothesen

Nachdem die Effekte des Modells bestimmt und abgegrenzt sind, lassen sich nun die zu testenden Hypothesen formulieren. Da sich alle festen Effekte durch die Erwartungswerte darstellen lassen, erscheint es sinnvoll, auch die Hypothesen mit dem Erwartungsvektor $\boldsymbol{\mu} = (\boldsymbol{\mu}'_1, \dots, \boldsymbol{\mu}'_a)'$ zu bilden. Die Nullhypothesen für kein Behandlungseffekt, kein Zeiteffekt und keine Wechselwirkung lauten:

$$\begin{aligned} H_0(\alpha): \alpha_1 = \dots = \alpha_a &\Leftrightarrow \bar{\mu}_{\cdot 1} = \dots = \bar{\mu}_{\cdot a} \\ H_0(\beta): \beta_1 = \dots = \beta_d &\Leftrightarrow \bar{\mu}_{\cdot 1} = \dots = \bar{\mu}_{\cdot d} \\ H_0(\alpha\beta): (\alpha\beta)_{ik} = 0 &\Leftrightarrow \mu_{ik} - \bar{\mu}_{i \cdot} - \bar{\mu}_{\cdot k} + \mu = 0 \quad \forall i, k. \end{aligned}$$

Jede der soeben formulierten Hypothesen lässt sich in einer äquivalenten Form $H_0: \mathbf{T}\boldsymbol{\mu} = \mathbf{0}$ mit passender Projektionsmatrix \mathbf{T} darstellen, die sich durch Konstruktion aus einer beliebigen Hypothesenmatrix \mathbf{H} mit der Formel

$$\mathbf{T} = \mathbf{H}'(\mathbf{H}\mathbf{H}')^{-1}\mathbf{H}$$

ergibt und eindeutig ist. Gemeinsam mit den Projektionsmatrizen haben die Nullhypothesen die Gestalt

$$H_0(\alpha): (\mathbf{P}_a \otimes \frac{1}{d}\mathbf{J}_d)\boldsymbol{\mu} = \mathbf{0}$$

$$H_0(\beta): (\frac{1}{a}\mathbf{J}_a \otimes \mathbf{P}_d)\boldsymbol{\mu} = \mathbf{0}$$

$$H_0(\alpha\beta): (\mathbf{P}_a \otimes \mathbf{P}_d)\boldsymbol{\mu} = \mathbf{0}.$$

An dieser Stelle bezeichnet $\mathbf{P}_a = \mathbf{I}_a - \frac{1}{a}\mathbf{J}_a \in \mathbb{R}^{a \times a}$ die zentrierende Matrix aus der Einheitsmatrix \mathbf{I}_a und der nur aus Einsen bestehenden Matrix \mathbf{J}_a . Im Allgemeinen ist die Projektionsmatrix das Kroneckerprodukt aus einem Wholeplot- und einem Subplotprojektor

$$H_0: (\mathbf{T}_W \otimes \mathbf{T}_S)\boldsymbol{\mu} = \mathbf{0}$$

mit geeigneten Projektoren $\mathbf{T}_W \in \mathbb{R}^{a \times a}$ und $\mathbf{T}_S \in \mathbb{R}^{d \times d}$.

3 Testverfahren

3.1 Mehrstichprobenfall

In diesem Kapitel werden $N = \sum_{i=1}^a n_i$, $n_i > 0$ unabhängige Versuchsobjekte

$$\mathbf{X}_{ij} \sim N_d(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i); i = 1, \dots, a; j = 1, \dots, n_i$$

mit Erwartungsvektoren $\boldsymbol{\mu}_i$ und positiv definiten Kovarianzmatrizen $\boldsymbol{\Sigma}_i$ betrachtet, von denen n_i Realisierungsvektoren der i -ten von $a \geq 2$ Behandlungsgruppen zugeordnet sind. Aus der Verteilungsannahme der Individuen ergibt sich für die Verteilung des i -ten Behandlungsmittelwertes, dass

$$\bar{\mathbf{X}}_i \sim N_d(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i/n_i) \text{ bzw. } \sqrt{N} \cdot \bar{\mathbf{X}}_i \sim N_d(\sqrt{N} \cdot \boldsymbol{\mu}_i; N/n_i \cdot \boldsymbol{\Sigma}_i).$$

Ordnet man diese unabhängigen Gruppenmittelwerte in einem Vektor untereinander an, folgt für die Verteilung des Vektors

$$\sqrt{N} \cdot \bar{\mathbf{X}} \sim N_{ad}(\sqrt{N} \cdot \boldsymbol{\mu}, \sqrt{N} \cdot \mathbf{V}),$$

mit $\boldsymbol{\mu} = (\boldsymbol{\mu}'_1, \dots, \boldsymbol{\mu}'_a)'$ und der Blockdiagonalmatrix $\mathbf{V} = \bigoplus_{i=1}^a \boldsymbol{\Sigma}_i/n_i$. Wie bereits in Kapitel 2.2 beschrieben, lassen sich alle Hypothesen des Modells über den Erwartungsvektor $\boldsymbol{\mu}$ formulieren, der erwartungstreu und konsistent durch den Vektor der Stichprobenmittelwerte $\bar{\mathbf{X}}$ geschätzt werden kann. Da die Kovarianzmatrix \mathbf{V} andernfalls degeneriert, wird an die Stichprobenumfänge die Bedingung

$$0 < \frac{N}{n_i} \leq n^* < \infty \text{ für } N \rightarrow \infty, i = 1, \dots, a \text{ und } n^* \in \mathbb{R}$$

gestellt. Diese Bedingung stellt sicher, dass die Stichproben „gleichmäßig“ wachsen und wird für den Konsistenznachweis der Schätzer benötigt. n^* ist dabei eine obere Grenze für den Quotienten aus Gesamtzahl der Individuen N und Anzahl der Individuen je Behandlungsgruppe n_i . Zum Testen der Nullhypothesen wird die quadratische Form

$$Q = N \cdot \bar{\mathbf{X}}' \mathbf{T} \bar{\mathbf{X}}$$

mit Projektionsmatrix $\mathbf{T} = (\mathbf{T}_W \otimes \mathbf{T}_S) \in \mathbb{R}^{ad \times ad}$ verwendet. Es lässt sich zeigen, dass Q unter der Nullhypothese $H_0: \mathbf{T}\boldsymbol{\mu} = \mathbf{0}$ als Summe χ^2 -verteilter Zufallsvariablen $Q = \sum_{s=1}^{ad} \lambda_s C_s$ dargestellt werden kann, wobei die C_s unabhängige χ_1^2 -verteilte Zufallsvariablen sind und λ_s die Eigenwerte der Matrix $\mathbf{T}\mathbf{V}\mathbf{T}$ für $s = 1, \dots, ad$ sind.

Für die approximative Verteilung der quadratischen Form Q wird die Idee einer Zwei-Momenten-Approximation von Box und Patnaik verwendet (Box, 1954), (Patnaik, 1949).

3.1.1 Box-Approximation

Die Verteilung der quadratischen Form Q wird mit einer gestreckten χ^2 -Verteilung approximiert, so dass die ersten zwei Momente übereinstimmen

$$Q \approx g \cdot \chi_f^2.$$

Durch Umstellung des Gleichungssystems erhält man für die unbekannt Parameter

$$g = N \cdot \frac{Sp[(\mathbf{T}\mathbf{V})^2]}{Sp[\mathbf{T}\mathbf{V}]} \quad f = \frac{Sp[\mathbf{T}\mathbf{V}]^2}{Sp[(\mathbf{T}\mathbf{V})^2]}$$

wobei $Sp[\cdot]$ die Spur einer quadratischen Matrix bezeichnet. Gemeinsam mit einer weiteren Momenten-Approximation der quadratischen Form

$$N \cdot Sp[\widehat{\mathbf{T}\mathbf{V}}] \approx g_0 \cdot \chi_{f_0}^2$$

gelangt man schließlich zu einer approximativ F-verteilter Teststatistik

$$T = \frac{Q}{N \cdot Sp[\widehat{\mathbf{T}\mathbf{V}}]} \approx F(f, f_0),$$

deren Freiheitsgrade f und f_0 aus den Daten geschätzt werden müssen. Die Nullhypothese $H_0: \mathbf{T}\boldsymbol{\mu} = \mathbf{0}$ wird verworfen, wenn die Teststatistik T das $(1 - \alpha)$ -Quantil der F-Verteilung überschreitet.

3.2 Schätzung der Freiheitsgrade

Eine wesentliche Herausforderung dieser Teststatistik liegt in der korrekten Schätzung der Freiheitsgrade. Da bisher keine exakte Lösung dieses Problems gefunden wurde, beschränken sich alle Ansätze auf Approximationen. Die in SAS implementierten Frei-

heitsgradschätzer beruhen auf der Korrektur der Freiheitsgrade einer Varianzanalyse. Simulationen wie in Tabelle 3 belegen, dass die „alten“ Schätzer verzerrt sind und diese Verzerrung von der Dimension d abhängt. Dies hat zur Folge, dass das Fehlerniveau 1. Art nicht eingehalten wird.

Tabelle 3: Vergleich von Quotienten aus Spur- und Plug-in-Schätzern und dem wahren Freiheitsgrad für eine Behandlungsgruppe mit $n = 10$ Individuen, $\Sigma = I_d + \mathbf{1}_{d \times d}$ und $n_{sim} = 1.000$ Simulationen.

d	\hat{f}/f	$Var(\hat{f})/f^2$	\tilde{f}/f	$Var(\tilde{f})/f^2$
5	1.0450280	0.0547921	1.0450280	0.0082674
10	1.0471479	0.0561270	1.0471479	0.0024427
20	1.0600445	0.0750266	0.3159279	0.0004857
30	1.0447874	0.0707148	0.2324651	0.0001429
40	1.0525466	0.0718624	0.1849800	0.0000617
50	1.0467603	0.0667174	0.1532802	0.0000277
75	1.0678702	0.0703435	0.1076151	< 0.00001
100	1.0690598	0.0688765	0.0827766	< 0.00001

Tabelle 4: Vergleich von Quotienten aus Spur- und Plug-in-Schätzern und dem wahren Freiheitsgrad für eine Behandlungsgruppe mit $n = 100$ Individuen, $\Sigma = I_d + \mathbf{1}_{d \times d}$ und $n_{sim} = 1.000$ Simulationen.

d	\hat{f}/f	$Var(\hat{f})/f^2$	\tilde{f}/f	$Var(\tilde{f})/f^2$
5	1.0003738	0.0004094	0.9571377	0.0003368
10	1.0006961	0.0004257	0.9110077	0.0002897
20	1.0002897	0.0004268	0.8328381	0.0002033
30	0.9997950	0.0004087	0.7678061	0.0001418
40	0.9999840	0.0003907	0.7125716	0.0001009
50	0.9998202	0.0004216	0.6645817	0.0000822
75	1.0010005	0.0004217	0.5693934	0.0000437
100	1.0011240	0.0004233	0.4977943	0.0000256

Die in dem Makro F1_HD_F1 verwendeten Spurschätzer sind unverzerrt und konvergieren, unabhängig von d , gegen die zu schätzenden Parameter f und f_0 . In Tabelle 3 und Tabelle 4 ist exemplarisch dargestellt, dass die Spurschätzer ein besseres Konvergenzverhalten besitzen, als die Plug-in-Schätzer, die mit steigender Dimension immer kleiner werden und deren Varianz scheinbar gegen 0 konvergiert. Es kann sogar theoretisch gezeigt werden, dass die Spurschätzer das von (Brunner, Repeated Measures under Non-Sphericity, 2009) geprägte Kriterium der Dimensionsstabilität erfüllen, was anschaulich bedeutet, dass die Konvergenz der Schätzer gleichmäßig beschränkt ist und diese Schranke nicht von der Anzahl der Messwiederholungen d abhängt.

4 Makro F1_HD_F1

Während die Zeitpunkte in PROC GLM in je einer Spalte ausgewiesen werden müssen, um SAS die Abhängigkeitsstruktur der Daten zu übermitteln, werden bei dem Makro F1_HD_F1 die Messwerte in einer Spalte untereinander angeordnet. Die Zuordnung von Messwiederholungen und Zeitpunkten erfolgt durch entsprechende Spalten. Eine Sortierung der Daten ist im Vorfeld nicht erforderlich und wird während der Auswertung durchgeführt. Der Aufruf des Makros und die Ausgabe haben die folgende Gestalt:

```
%F1_HD_F1(data=wistar,
          var=Messwert,
          time=Dim,
          subject=Ratte,
          group=Gruppe);
```

The SAS System				
a	n	d		
3	13	22		
	12			
	24			

Ausgabe				
	T	f	f0	p-Wert
Gruppe	10.589464	1	37	0.0024474
Dim	274.6913	1	127	0
Wechselwirkung	8.693539	1	127	0.0038029

Abbildung 3: Ausgabe des Makros F1_HD_F1 für den Beispieldatensatz.

Die Auswertung legt den Schluss eines signifikanten Gruppen- und Zeiteffektes und einer signifikanten Wechselwirkung nahe. Die Simulationen im folgenden Abschnitt sollen die Anfälligkeit der in SAS implementierten Approximationen verdeutlichen.

5 Simulationen

Anhand von Simulationen kann demonstriert werden, welchen Einfluss ungleiche Kovarianzmatrizen auf die Güte der besprochenen Testverfahren ausüben. Dazu werden die Approximationen von Geisser-Greenhouse, Huynh-Feldt und eine gewöhnliche Anova mit der neuen Approximation verglichen. In Anlehnung an den Datensatz werden Zufallszahlen für 3 Behandlungsgruppen mit einer Compound Symmetry-Struktur generiert, die sich nur in den Varianzen der Behandlungsgruppen und den Stichprobenumfängen unterscheiden. Es wird die Hypothese getestet, dass keine Wechselwirkung zwischen der Behandlung und Zeit vorliegt, da die erwarteten Freiheitsgrade am größten sind und Unterschiede zwischen den Approximationen am deutlichsten zutage treten.

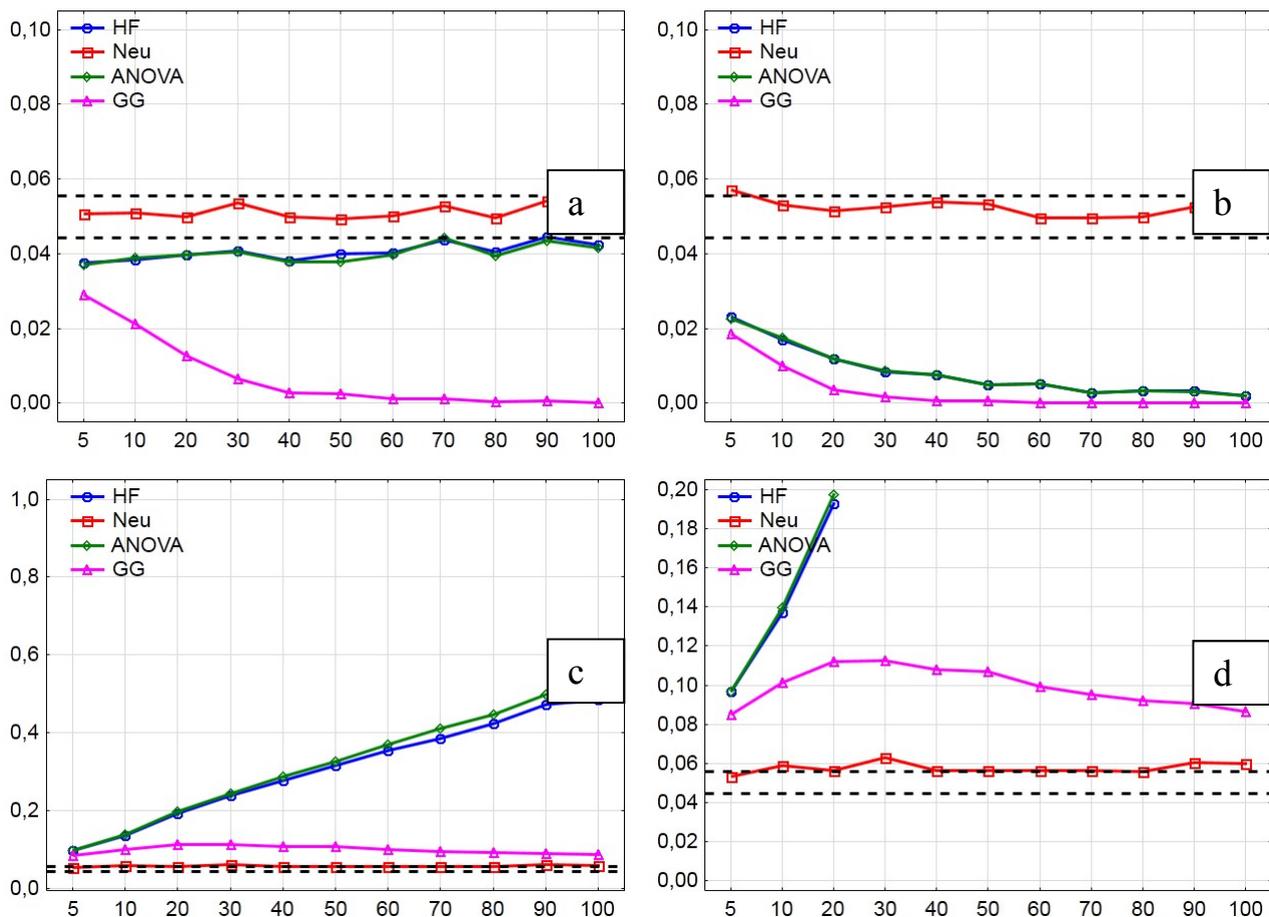


Abbildung 4:

- a) $n_1 = n_2 = n_3 = 10$ Probanden mit $\sigma_1^2 = 1522$, $\sigma_2^2 = 1682$, $\sigma_3^2 = 842$;
- b) $n_1 = 10, n_2 = 20, n_3 = 10$ Probanden mit $\sigma_1^2 = 1522$, $\sigma_2^2 = 1682$, $\sigma_3^2 = 842$;
- c) $n_1 = 10, n_2 = 10, n_3 = 20$ Probanden mit $\sigma_1^2 = 1522$, $\sigma_2^2 = 1682$, $\sigma_3^2 = 842$;
- d) Vergrößerung des Bereichs aus Setting c).

In der ersten Abbildung 4 a) werden je Gruppe 10 Beobachtungsvektoren erzeugt. Die Approximation von Geisser-Greenhouse wird mit steigender Dimension immer konservativer, die neue Approximation hält das Niveau für alle Dimensionen ein. Die Anova und die Approximation von Huynh-Feldt verhalten sich leicht konservativ. Die Abbildungen 4 b) und c) unterscheiden sich in der Datenerzeugung nur durch die Stichprobenumfänge. Die Approximationen Geisser-Greenhouse, Huynh-Feldt und Anova zeigen ein konservatives Verhalten, wenn weniger Probanden in der Gruppe mit kleiner Varianz beobachtet werden. Die Approximationen verhalten sich liberal, wenn viele Probanden in der Behandlungsgruppe mit kleiner Varianz beobachtet werden. Das Verhalten wird durch steigende Dimension verstärkt, wobei die Geisser-Greenhouse-Approximation mit steigender Dimension konservativer wird. Die neue Approximation hält das Niveau von $\alpha = 5\%$ in allen Fällen ein.

6 Diskussion

Die Simulationen legen den Schluss nahe, hochdimensionale Split-Plot-Designs mit Repeated Measures Daten mit dem Makro F1_HD_F1, anstelle der in SAS 9.3 implementierten Approximationen, auszuwerten. Die Folgen verletzter Annahmen wurden in Kapitel 5 demonstriert, besonders der Einfluss ungleicher Kovarianzmatrizen im Zusammenspiel mit ungleichen Stichprobenumfängen kann zu stark verfälschten Resultaten führen.

Das Makro F1_HD_F1 wird über die Homepage der Medizinischen Statistik der Universitätsmedizin Göttingen unter <http://www.ams.med.uni-goettingen.de/> zur Verfügung gestellt.

Literatur

- [1] Box, G. E. P., 1954. Some theorems on quadratic forms applied in the study of analysis of variance problems, ii. effects of inequality of variance and of correlation between errors in the two-way-classification. *The Annals of Mathematical Statistics*, pp. 484-498.
- [2] Brunner, E., 2009. Repeated Measures under Non-Sphericity. *Proceedings of the 6th St. Petersburg Workshop on Simulation*, pp. 605-609.
- [3] Brunner, E., Dette, H. & Munk, A., 1997. Box-type approximations in nonparametric factorial designs. *Journal of the American Statistical Association*.
- [4] Geisser, S. W. & Greenhouse, S., 1958. An extension of box's results on the use of the f distribution on multivariate analysis. *The Annals of Mathematical Statistics*, pp. 885-891.
- [5] Hotelling, H., 1931. A generalization of Student's ratio. *Annals of Mathematical Statistics*, pp. 360-378.
- [6] Huynh, H. & 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 and Behavioral Statistics*, pp. 69-82.
- [7] Lecoutre, B., 1991. A correction for the epsilon approximate test in repeated measures designs with two or more independent groups. *Journal of Educational and Behavioral Statistics*, pp. 371-372.
- [8] Patnaik, P. B., 1949. The non-central chi - and f-distribution and their applications. *Biometrika*, pp. 202-232.
- [9] SAS Institute Inc., 2012: The GLM Procedure, Syntax: REPEATED Statement. SAS/STAT® 12.1 User's Guide, August
http://support.sas.com/documentation/cdl/en/statug/65328/HTML/default/viewer.htm#statug_glm_syntax20.htm [27.03.2013]