



**Varianzanalysen -
Prüfen der Voraussetzungen und
nichtparametrische Methoden
sowie
praktische Anwendungen mit R und SPSS**

Version 1.1
(2.3.2015)

Haiko Lüpsen

Regionales Rechenzentrum (RRZK)

Kontakt: Luepsen@Uni-Koeln.de

Universität zu Köln



Vorwort

Entstehung

In den letzten Jahren hatte ich mehrfach Kurse zum Thema „nichtparametrische Methoden mit SPSS“ bzw. Kurse zur Programmiersprache S und dem System R am RRZK gehalten. Dort hatte sich gezeigt, dass ein großes Interesse an nichtparametrischen statistischen Verfahren besteht, insbesondere im Bereich Varianzanalyse. Immerhin sind die dazu zählenden Verfahren, vom t-Test bis zur mehrfaktoriellen Analyse mit Messwiederholungen, die am meisten verwendeten. Umso erstaunlicher ist es, dass in den großen Statistiksystemen, insbesondere in SPSS, außer den alt bekannten 1-faktoriellen Klassikern Kruskal-Wallis- und Friedman-Tests keine nichtparametrischen Varianzanalysen angeboten werden. An Methoden mangelt es nicht, wie die nachfolgenden Kapitel sowie die angeführte Literatur zu diesem Thema zeigen.

Immerhin kann man mit mehr oder weniger Aufwand einige dieser Verfahren auch in SPSS durchführen, da sich manche auf die klassische Varianzanalyse zurückführen lassen. Solche Verfahren stehen daher im Vordergrund. Mit S bzw. R lassen sich naturgemäß alle Methoden programmieren. Auch da zeigen sich erstaunlicherweise große Lücken im Angebot. Daher sind im Anhang selbst erstellte R-Funktionen zu diesem Thema angeführt.

Da sich zwangsläufig vor Durchführung der Varianzanalyse die Frage stellt: In wieweit sind die Voraussetzungen für die parametrische Analyse erfüllt und wie robust sind die Verfahren, werden diese Fragen auch ausführlich behandelt. Manchmal reichen auch robuste Varianten der „klassischen“ Varianzanalyse, die hier natürlich auch vorgestellt werden.

Dieses waren die Themen meiner Kurse. In den entsprechenden Kursunterlagen waren die Antworten bzw. Lösungen zu den o.a. Fragen und Methoden nur skizziert. Da ich im WWW keine vergleichbare Zusammenstellung gefunden hatte, entschloss ich mich, die Kursunterlagen beider Kurse (SPSS und R) zu einem Skript „auszubauen“, das als Anleitung benutzt werden kann.

Umfang

Das Skript setzt voraus, dass der Leser zum einen mit Varianzanalysen (mehr oder weniger) vertraut ist und zum anderen mit R bzw. SPSS umgehen kann. So werden z.B. bei SPSS weitgehend die Angaben zu den Menüs, über die die einzelnen Funktionen erreichbar sind, zugunsten der SPSS-Syntax ausgespart.

Eine generelle Einführung in die Varianzanalyse, simple effects-Analysen, Ziehen der richtigen Schlüsse etc behandelt werden, ist geplant.

Ursprünglich war geplant, das Thema „multiple Mittelwertvergleiche und α -Adjustierungen“ ebenfalls in diesem Skript zu behandeln. Allerdings merkte ich schnell bei der Sichtung der Verfahren und der aktuellen Literatur, dass dies ein eigenes „Thema“ sein muss. Dementsprechend gibt es inzwischen dazu ein eigenes Skript, das an gleicher Stelle wie dieses abrufbar ist.

Spätere Versionen

Bei der Sichtung der Literatur war ich überrascht, wie viele nichtparametrische Verfahren zur Varianzanalyse seit Beginn der 90er Jahre entwickelt worden sind, obwohl ich mich seit jener Zeit etwas auf dem Laufenden gehalten habe. Denn ich hatte mich mit dem Thema selbst intensiv befasst und zu der Zeit das Programm Easystat entwickelt, mit dem auch mehrfaktorielle

nichtparametrische Analysen, auch mit Messwiederholungen, möglich waren.

Bei der Vielzahl von neueren Methoden handelt es sich allerdings auch vielfach um Modifikationen zu bekannten Verfahren. So gibt es inzwischen fast zahllose Varianten der ART (Aligned Rank Transform)-Verfahren.

Das vorliegende Skript wurde Ende 2014 komplett überarbeitet und sollte mittlerweile einigermaßen fehlerfrei sein. Es ist geplant, in späteren Versionen dieses Skripts weitere Verfahren zu berücksichtigen, insbesondere solche, die sich mit ein wenig Aufwand auch in SPSS realisieren lassen. Für die Zukunft sind auf ebensol Kovarianzanalysen geplant.

Lesehinweise

Zu jedem Versuchsplan, z.B. ohne bzw. mit Messwiederholungen, und zu jeder Methode gibt es nach einer kurzen Beschreibung des Verfahrens jeweils ein ausführliches Beispiel. Dieses wird dann einmal mit R sowie einmal mit SPSS durchgerechnet.

Die Ergebnistabellen aus R und SPSS sind zum Teil verkürzt wiedergegeben, d.h. Teile, die nicht zum Verständnis erforderlich sind, fehlen hier.

Historie

Version 1.0.1 (17.1.2015): Korrekturen an der Iman-Davenport-Korrektur.

Version 1.0.2 (10.2.2015): Korrekturen an den Formeln für die ART-Verfahren.

Version 1.1 (2.3.2015): Verwendung diverser neuer R-Funktionen sowie andere Korrekturen.

Inhaltsverzeichnis

1.	Allgemeines zur nichtparametrischen Statistik	1
1. 1	Wichtige Begriffe	1
1. 1. 1	Effizienz eines Tests	1
1. 1. 2	konservative und liberale Tests	2
1. 1. 3	starke und schwache Tests	2
1. 1. 4	robuste Tests	2
1. 2	Methoden für metrische Merkmale	2
1. 3	Methoden für ordinale Merkmale	3
1. 4	Methoden für dichotome Merkmale	3
1. 5	Methoden für nominale Merkmale	3
1. 6	Prüfung auf Normalverteilung	3
1. 7	Prüfung von Voraussetzungen	6
2.	Nichtparametrische Varianzanalysen - Übersicht der Methoden	7
2. 1	Kruskal-Wallis und Friedman	7
2. 2	Rank transform Tests (RT)	8
2. 3	Inverse normal transform (INT)	9
2. 4	Aligned rank transform (ART)	9
2. 5	Verallgemeinerte Kruskal-Wallis- und Friedman-Analysen (KWF, Puri & Sen)	10
2. 6	van der Waerden	11
2. 7	Bredenkamp Tests	12
2. 8	Akritas, Arnold & Brunner ATS Tests	12
2. 9	Wilcoxon analysis (WA)	13
2. 10	Weitere Varianzanalysen	13
2. 10. 1	Brunner, Dette und Munk	13
2. 10. 2	Quade	13
2. 10. 3	Skillings & Mack	14
2. 10. 4	Beasley & Zumbo	14
2. 11	Logistische Regression	14
2. 12	Voraussetzungen	14
2. 13	Vergleiche	15
3.	Funktionen zur Varianzanalyse in R und SPSS	16
3. 1	Funktionen in R	16
3. 2	Funktionen in SPSS	18
3. 3	Fehler bei der Rangberechnung	18
3. 4	Fehlende Werte	19
4.	Unabhängige Stichproben	20
4. 1	Voraussetzungen der parametrischen Varianzanalyse	21
4. 2	Die 1-faktorielle Varianzanalyse	23
4. 2. 1	Kruskal-Wallis-Test	23
4. 2. 2	Varianzanalysen für inhomogene Varianzen	24
4. 2. 3	Verfahren für nichtnormalverteilte Variablen	26
4. 2. 4	Weitere Verfahren	26

4.3	Die 2-faktorielle Varianzanalyse	26
4.3.1	Anmerkungen zur 2-faktoriellen Varianzanalyse	26
4.3.1.1	Balancierte und nichtbalancierte Versuchspläne	26
4.3.1.2	Die Interaktion	27
4.3.1.3	Reduzierung des statistischen Fehlers	28
4.3.2	Das parametrische Verfahren und Prüfung der Voraussetzungen	29
4.3.3	Varianzanalysen für inhomogene Varianzen	32
4.3.3.1	Verfahren von Box sowie von Brown & Forsythe	34
4.3.3.2	BDM-Test	34
4.3.3.3	Variablentransformationen	35
4.3.4	Rank transform-Tests (RT)	35
4.3.5	Verallgemeinerte Kruskal-Wallis- und Friedman-Analysen (KWF)	37
4.3.6	Aligned rank transform (ART)	39
4.3.7	normal scores- (INT-) und van der Waerden-Tests	43
4.3.8	ATS-Tests von Akritas, Arnold & Brunner	46
4.3.9	Bredenkamp Tests	46
4.4	Nichtparametrische Verfahren zur mehrfaktoriellen Varianzanalyse	47
4.5	Fazit	47
5.	Abhängige Stichproben - Messwiederholungen	50
5.1	Datenstruktur - Besonderheiten bei R und SPSS	51
5.2	Voraussetzungen der parametrischen Varianzanalyse	53
5.3	Die 1-faktorielle Varianzanalyse	56
5.3.1	Parametrischer Test und Prüfung der Voraussetzung	56
5.3.2	Der Friedman-Test	61
5.3.3	Verallgemeinerte Kruskal-Wallis- und Friedman-Tests (KWF)	62
5.3.4	rank transform (RT) und normal scores (INT)	64
5.3.5	van der Waerden	66
5.3.6	ATS-Tests von Akritas, Arnold & Brunner	68
5.3.7	Quade-Test	69
5.3.8	Skillings-Mack-Test	70
5.4	Die 2-faktorielle Varianzanalyse	70
5.4.1	Das parametrische Verfahren und Prüfung der Voraussetzungen	70
5.4.2	Rank transform-Tests (RT) und normal scores -Tests (INT)	75
5.4.3	Verallgemeinerte Kruskal-Wallis- und Friedman-Tests (KWF)	78
5.4.4	Aligned rank transform (ART)	84
5.4.5	ATS-Tests von Akritas, Arnold & Brunner	88
5.4.6	Bredenkamp Tests	90
5.5	Fazit	90
6.	Gemischte Versuchspläne	91
6.1	Voraussetzungen der parametrischen Varianzanalyse	91
6.2	Parametrische Varianzanalyse und Prüfung der Voraussetzungen	92
6.3	Rank transform-Tests (RT)	98
6.4	Verallgemeinerte Kruskal-Wallis- und Friedman-Tests (KWF)	100
6.5	Aligned rank transform (ART)	105
6.5.1	Ein Gruppierungs- und ein Messwiederholungsfaktor	105
6.5.2	Ein Gruppierungs- und zwei Messwiederholungsfaktoren	108
6.5.3	Zwei Gruppierungs- und ein Messwiederholungsfaktor	110

6. 6	normal scores-Tests (INT)	115
6. 7	van der Waerden-Tests	117
6. 8	ATS-Tests von Akritas, Arnold & Brunner	121
6. 9	Bredenkamp Tests	122
6. 9. 1	Ein Gruppierungs- und ein Messwiederholungsfaktor	122
6. 9. 2	Zwei Gruppierungs- und ein Messwiederholungsfaktor	124
6. 10	Fazit	126
7.	Analysen für dichotome Merkmale	128
7. 1	Anwendung der Verfahren für metrische Merkmale	129
7. 1. 1	Unabhängige Stichproben	129
7. 1. 2	Gemischte Versuchspläne	131
7. 2	Anwendung der Verfahren für ordinale Merkmale	132
8.	Logistische Regression	133
8. 1	dichotome abhängige Variablen	133
8. 2	ordinale abhängige Variablen	136
9.	Mittelwertvergleiche, Kontraste und Kodierungen	142
9. 1	Grundlagen	142
9. 2	Standard-Kontraste	144
9. 3	Auswahl der Kontraste	146
9. 4	nichtparametrische Kontraste für die RT-, ART- und KWF-Verfahren	147
9. 5	universelles Verfahren für Kontraste	151
9. 6	Kontraste bei logistischen Regressionen	152
9. 7	Kontraste für Messwiederholungen und Interaktionen	152
Anhang		156
1.	Umstrukturieren von Messwiederholungen in SPSS	156
1. 1	Umstrukturieren von Messwiederholungen in Fälle	156
1. 1. 1	ein Faktor und eine Analyse-Variable	156
1. 1. 2	mehrere Faktoren und eine Analyse-Variablen	159
1. 1. 3	ein Faktor und mehrere Analyse-Variablen	162
1. 2	Umstrukturieren von Fälle in Messwiederholungen	166
2.	Spezielle robuste F-Tests und andere Statistiken	169
2. 1	Box-Korrektur für heterogene Varianzen	169
2. 2	Brown-Forsythe F-Test für inhomogene Varianzen	169
2. 3	Box-Andersen F-Test für nichtnormalverteilte Variablen	170
2. 4	Box-Cox-Transformationen	170
2. 5	Fishers combined probability test	170
3.	R-Funktionen	171
3. 1	box.f: Box-F-Test für inhomogene Varianzen	171
3. 2	bf.f: Brown & Forsythe-F-Test für inhomogene Varianzen	171
3. 3	box.andersen.f: F-Test für nichtnormalverteilte Variablen	171
3. 4	boxm.test: Test auf Homogenität von Kovarianzmatrizen	172

3. 5	ats.2 und ats.3: 2- bzw. 3-faktorielle Varianzanalyse	172
3. 6	np.anova: nichtparametrische Varianzanalyse mittels der Verfahren von Puri & Sen und van der Waerden	172
3. 7	art1.anova: nichtparametrische Varianzanalyse nach dem ART-Verfahren (nur Gruppierungsfaktoren)	173
3. 8	art2.anova: nichtparametrische Varianzanalyse nach dem ART-Verfahren (nur Messwiederholungsfaktoren)	173
3. 9	art3.anova: nichtparametrische Varianzanalyse nach dem ART-Verfahren (für gemischte Versuchspläne)	174
4.	Literaturhinweise	175
Datensätze		
	Beispieldaten 1 (mydata1):	20
	Beispieldaten 2 (mydata2):	20
	Beispieldaten 3 (mydata3):	20
	Beispieldaten 4 (winer518):	49
	Beispieldaten 5 (mydata5):	49
	Beispieldaten 6 (winer568):	50
	Beispieldaten 7 (irish):	126
	Beispieldaten 8 (koch):	126

Alle Datensätze können von folgender Webseite heruntergeladen werden:

<http://www.uni-koeln.de/~luepsen/daten/>

1. Allgemeines zur nichtparametrischen Statistik

Parametrischen statistischen Verfahren (http://de.wikipedia.org/wiki/Parametrische_Statistik) liegt in der Regel ein mathematisches Modell zugrunde, das auf einer Verteilungsannahme beruht, häufig der Normalverteilung. Dabei müssen nicht unbedingt die Merkmale selbst der Verteilung folgen, häufig sind es auch abgeleitete Größen wie z.B. die Residuen. Die im Modell angenommene Verteilung hat Parameter (z.B. Mittelwert μ und Standardabweichung σ bei der Normalverteilung), über die sich dann die Parameter des Modells bestimmen lassen. Bei den nichtparametrischen Verfahren, auch *verteilungsfreie* Verfahren genannt (http://de.wikipedia.org/wiki/Parameterfreie_Statistik), wird in der Regel keine solche Verteilung angenommen.

Parametrische Verfahren werden meistens angewandt, wenn die abhängige Variable metrisch ist und zusätzliche Verteilungsvoraussetzungen, wie Normalverteilung der Residuen, erfüllt sind. Häufig kommen zusätzliche Voraussetzungen hinzu, wie z.B. Homogenität der Varianzen oder Unabhängigkeit der Beobachtungen. So z.B. bei der Varianz- oder Regressionsanalyse. Ist eine der Voraussetzungen nicht erfüllt, versucht man, äquivalente nichtparametrische Verfahren anzuwenden, sofern vorhanden. Letztere haben gegenüber den parametrischen eine geringere (asymptotische) Effizienz - mehr dazu im nächsten Kapitel, in der Regel zwischen 63.7% ($2/\pi$), z.B. beim Vorzeichen- und Mediantest, und $95,5\%$ ($3/\pi$), so beim Mann-Whitney U- und Kruskal-Wallis H-Test, falls alle Voraussetzungen erfüllt sind. Die Effizienz nichtparametrischer Tests kann allerdings auch umgekehrt über 100% , sogar beliebig hoch, liegen, wenn die Verteilungsvoraussetzungen nicht erfüllt sind. D.h. je weniger die Voraussetzungen eines parametrischen Tests erfüllt sind, desto eher ist zu einem nichtparametrischen Test zu raten.

Vielfach werden Vorbehalte gegen nichtparametrische Verfahren geltend gemacht, weil bei diesen nicht alle Informationen der Daten ausgeschöpft würden. Dieses mag zwar gelegentlich der Fall sein, z.B. beim Median-Test als nichtparametrische Varianzanalyse, gilt aber nicht allgemein und insbesondere nicht für die hier besprochenen Methoden. So hat z.B. Sawilowsky [16] in seiner Zusammenstellung auch diesen allgemeinen Punkt betrachtet. Demnach schneiden die (hier aufgeführten) nichtparametrischen Verfahren fast genau so gut ab, wie die parametrische Varianzanalyse. Und in den Fällen, wo die Voraussetzungen für die parametrischen Verfahren nicht mehr erfüllt sind, sind die nichtparametrischen ohnehin überlegen.

In Abhängigkeit vom Skalenniveau der abhängigen Variablen unterscheidet man die Verfahren.

1.1 Wichtige Begriffe

1.1.1 Effizienz eines Tests

Die (asymptotische) *relative Effizienz* (ARE) eines nichtparametrischen Tests A in Bezug auf einen parametrischen Test B (zur Prüfung derselben Hypothese) ist definiert als (das Grenzwertverhältnis für große n) n_B/n_A , den Quotienten der erforderlichen Stichprobenumfänge (n_A für Test A und n_B für Test B) zur Erlangung desselben Wertes für β , bei einem beliebigen (aber festen) α und unter der Annahme, dass die Voraussetzungen des parametrischen Tests erfüllt sind. (Dieser Grenzwert ist allerdings unabhängig von a .) D.h. eine Effizienz eines nichtparametrischen Tests A von 95% oder 67% gegenüber einem parametrischen Test B bedeutet, dass z.B. bei gleichen Mittelwertunterschieden der nichtparametrische Test eine ca. 5% ($(100-95)/95$) bzw. 50% ($(100-67)/67$) größere Stichprobe erfordert, um dieselbe Signifikanz zu erreichen. Dies schließt nicht aus, dass ein nichtparametrischer Test eine höhere Effizienz als der entsprechende parametrische haben kann, wenn die Voraussetzungen für den parametrischen nicht

erfüllt sind. So hat z.B. der Test von van der Waerden (vgl. Kapitel 2.6) für nichtnormalverteilte Variablen eine Effizienz größer als 1. Eine höhere Effizienz bedeutet immer auch eine größere Teststärke $1-\beta$.

Die Idee der asymptotischen relativen Effizienz ist folgende: Mit größer werdendem n wird auch der kleinste (Mittelwert-) Unterschied bei jedem Test einmal signifikant. Ein Test, der bis zu diesem Punkt ein kleineres n benötigt als ein anderer, kann als effizienter angesehen werden, da er mit einer kleineren Stichprobe auskommt.

1. 1. 2 konservative und liberale Tests

Ein Test reagiert *konservativ*, wenn die tatsächliche Wahrscheinlichkeit für einen Fehler 1. Art kleiner als das vorgegebene α ist. D.h. wenn z.B. bei einem $\alpha=0.05$ die Anzahl der irrtümlich abgelehnten Nullhypothesen unter 5% liegt. Entsprechend reagiert ein Test *liberal*, wenn die tatsächliche Wahrscheinlichkeit für einen Fehler 1. Art das vorgegebene α überschreiten kann. D.h. wenn z.B. bei einem $\alpha=0.05$ die Anzahl der irrtümlich abgelehnten Nullhypothesen nicht konsequent unter 5% liegt.

Ein Test A ist *konservativer (liberaler)* als ein Test B, wenn die tatsächliche Wahrscheinlichkeit für einen Fehler 1. Art für A kleiner (größer) als für B ist.

So ist z.B. bei den multiplen Mittelwertvergleichen der Newman-Keuls-Test ein liberaler Test, und der Newman-Keuls-Test ist liberaler als der Tukey-Test. Umgekehrt ist der Tukey-Test konservativer als der Newman-Keuls-Test. Konservative Tests sind in der Regel schwächer als liberale Tests.

1. 1. 3 starke und schwache Tests

Ein Test A ist *stärker (schwächer)* als ein Test B, wenn bei gleichem α und n die Wahrscheinlichkeit β für einen Fehler 2. Art bei Test A größer (kleiner) ist als bei Test B. D.h. bei Test A ist es leichter (schwieriger), einen Unterschied nachzuweisen als bei Test B.

1. 1. 4 robuste Tests

Ein Test wird als *robust* bezeichnet, wenn auch bei (moderaten) Verletzungen der Voraussetzungen die Ergebnisse noch korrekt sind. Das beinhaltet zweierlei: Zum einen wird die Rate für den Fehler 1. Art α eingehalten, d.h. bei z.B. $\alpha=0.05$ sind auch nur 5 von 100 Ergebnissen zufällig signifikant. Zum anderen verändert sich die Wahrscheinlichkeit für einen Fehler 2. Art β nicht drastisch, d.h. auch bei verletzten Voraussetzungen kann man noch signifikante Resultate erhalten.

1. 2 Methoden für metrische Merkmale

Bei diesen werden die Werte der Variablen in Ränge umgerechnet (vgl. [http://de.wikipedia.org/wiki/Rang_\(Statistik\)](http://de.wikipedia.org/wiki/Rang_(Statistik))). Auf diese werden dann die klassischen parametrischen Verfahren angewandt. So ist z.B. der Spearman-Rangkorrelationskoeffizient nichts anderes als der Pearson-Produkt-Moment-Korrelationskoeffizient der Ränge. Lediglich die Signifikanztests sind dann nicht mehr korrekt. Die korrekten Signifikanztests errechnen sich mit Mitteln der Kombinatorik, allerdings nur für kleine n (etwa <20) oder es werden asymptotische Signifikanztests angeboten, die nur für große n ($n>20$) gültig sind. In SPSS wird beides angeboten. Es konnte allerdings gezeigt werden, dass die Anwendung der klassischen parametrischen Verfahren auf die rangtransformierten Daten (ohne Anpassung der Signifikanztests) zu i.a. gültigen Ergebnis-

sen führt. Und dies sogar bei Verfahren, die sonst als sehr sensitiv bzgl. der Verletzungen von Voraussetzungen gelten, so z.B. multiple Mittelwertvergleiche und Diskriminanzanalyse, klassischen parametrischen Verfahren. (Vgl. dazu Conover & Iman [6].)

1.3 Methoden für ordinale Merkmale

Die oben erwähnten Verfahren für metrische Verfahren setzen voraus, dass eine Variable keine gleichen Werte hat. Durch sog. Bindungskorrekturen werden diese Verfahren allerdings auch anwendbar für ordinale Variablen, bei denen typischerweise Werte mehrfach vorkommen und dieser Tatsache bei der Rangberechnung durch die sog. Bindungen Rechnung getragen wird. Inzwischen sind in allen diesen Verfahren Bindungskorrekturen eingebaut.

In den letzten Jahren sind auch zunehmend Modelle für ordinale Merkmale entwickelt worden, denen die *relativen Effekte* zugrunde liegen, u.a. von Akritas, Arnold und Brunner [26]. Die daraus resultierenden Verfahren haben eine vergleichsweise hohe Effizienz, z.B. im Gegensatz zum Median-Test, der auch ohne Rangtransformationen metrischer Variablen auskommt. Mehr dazu in Kapitel 2.8.

1.4 Methoden für dichotome Merkmale

Dichotome Variablen könnte man einfach unter die nominalen Variablen subsummieren. Sie spielen aber eine Sonderrolle: Zum einen gestalten sich viele Formeln und mathematische Verfahren einfacher, wenn ein Merkmal nur zwei Ausprägungen hat. Zum anderen haben viele Simulationen gezeigt, dass man dichotome Variablen bei größeren Fallzahlen vielfach genauso handhaben kann wie metrische Variablen. So z.B. bei der Varianzanalyse. Hinzu kommt, dass man dichotome Variablen als Extremfall einer ordinalen Variablen betrachten kann und somit die dafür konzipierten Verfahren anwenden kann. Tatsächlich sind Verfahren für dichotome Variablen häufig identisch mit den äquivalenten für ordinale Variablen, z.B. der Phi-Koeffizient (Abhängigkeitsmaß) als Spezialfall des Spearman-Korrelationskoeffizienten oder Cochran's Q-Test als Spezialfall von Friedmans Varianzanalyse. (Vgl. dazu Cochran, W.G. [7] und Lunney, G.H., [8].)

1.5 Methoden für nominale Merkmale

Hier sind die polychotomen Merkmale angesprochen, also solche mit drei oder mehr Ausprägungen. Für solche Variablen gibt es vergleichsweise wenig statistische Methoden. Hinzu kommt, dass diese nicht immer trivial anzuwenden und die Ergebnisse nicht immer leicht verständlich sind. Entsprechende Methoden werden hier nicht vorgestellt.

1.6 Prüfung auf Normalverteilung

Die Normalverteilung spielt eine bedeutende Rolle bei der Entscheidung für oder gegen parametrische Verfahren. Insbesondere bei metrischen abhängigen Variablen wird i.a. eine Prüfung auf Normalverteilung vorgenommen, und zwar der Residuen e , die Bestandteil jedes varianzanalytischen Modells sind, z.B.

$$x_{ijm} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijm} \quad (i=1,\dots,I, j=1,\dots,J \text{ und } m=1,\dots,n_{ij})$$

Im einfachen Fall der Analyse ohne Messwiederholungen ist die Normalverteilung der Residuen äquivalent mit der Normalverteilung der abhängigen Variablen in jeder Zelle, allerdings auf keinen Fall mit der Normalverteilung der abhängigen Variablen insgesamt. (Letzteres würde ja

selten der Fall sein, da das untersuchte Merkmal für die einzelnen Zellen unterschiedliche Mittelwerte haben wird, die zu mehreren unterschiedlichen Gipfeln in der Gesamtverteilung führen würden.) Wollte man die abhängige Variable zellenweise auf Normalverteilung prüfen - wie es z.B. beim t-Test häufig gemacht wird - so müsste man eine Reihe von Prüfungen vornehmen, wo für jede von diesen nur ein geringes n zur Verfügung stünde, manchmal vielleicht weniger als 5. Damit lässt sich eine Normalverteilung weder beweisen noch widerlegen, egal mit welchem Verfahren. Das gleiche gilt natürlich auch, wenn man zellenweise die Residuen auf Normalverteilung überprüfen wollte.

Daher ist es erforderlich, alle Residuen e_{ijm} zusammen auf Normalverteilung zu überprüfen, denn dadurch kumulieren sich die n_{ij} zu einem brauchbaren n . Als Methoden gibt es sowohl Tests, u.a. der Shapiro-Wilk- oder der klassische Kolmogorov-Smirnov-Test, als auch Grafiken, u.a. Histogramme oder *normal probability Plots*.

Bei den Tests steckt man in einem Dilemma: Zum einen ist die Normalverteilungsvoraussetzung eher für kleinere Stichproben relevant als für größere, da bei großem n nach dem *zentralen Grenzwertsatz* ohnehin die Test-Statistiken die erforderlichen Verteilungsvoraussetzungen erfüllen. Zum anderen sprechen statistische Tests bei kleinem n nicht an, d.h. die Nullhypothese muss angenommen und eine Abweichung von der Normalverteilung kann nicht nachgewiesen werden.

Daher empfiehlt es sich, die Normalverteilung visuell über Grafiken zu überprüfen. Normal probability Plots sind insbesondere für Unerfahrene schwerer interpretierbar (siehe unten), so dass letztlich Histogramme das Verfahren der Wahl sind. Um nicht zu irreführenden Ergebnissen zu kommen, muss allerdings die Intervallzahl auf die Anzahl Beobachtungen n abgestimmt sein. Eine einfache aber dennoch sehr gute Faustregel ist

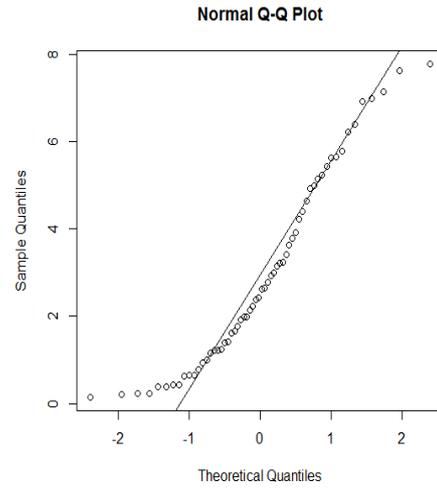
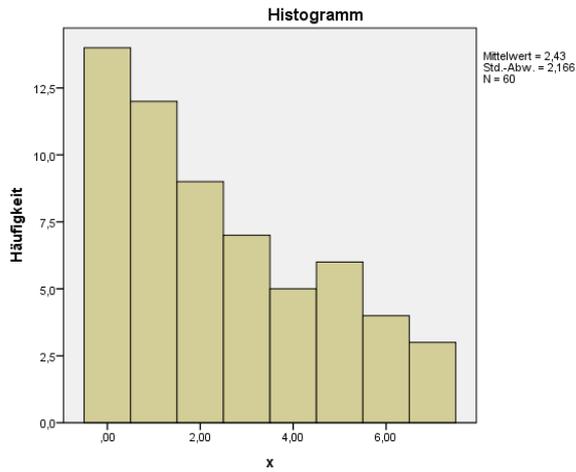
$$\text{Anzahl Intervalle} \sim \sqrt{n}$$

Aber auch dabei ist Vorsicht geboten, insbesondere wenn wie in SPSS gnadenlos die gewünschte Intervallzahl produziert wird: Bei diskreten (also nicht-stetigen) Merkmalen sollten alle Intervalle dieselbe Anzahl von Merkmalsausprägungen, also dieselbe Intervallbreite haben.

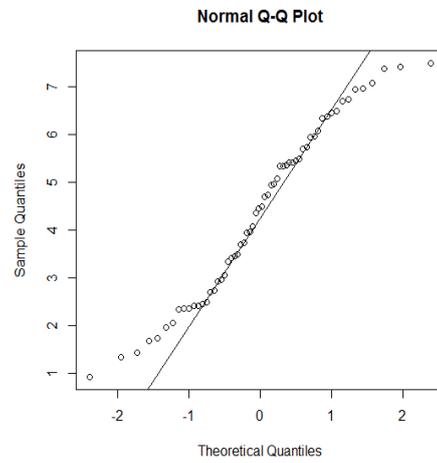
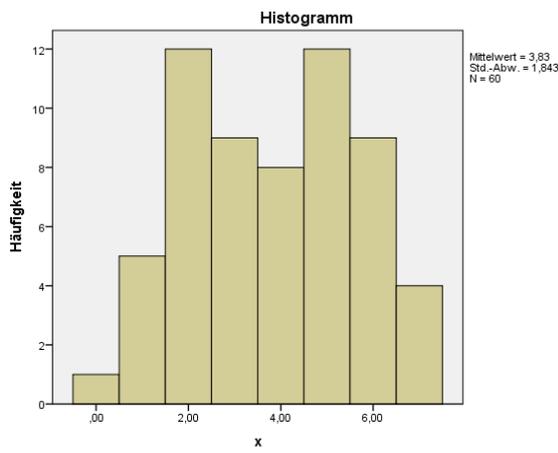
Andernfalls zeigt das Histogramm ein verzerrtes Verteilungsbild. In R wird bei `hist(x, breaks=k, ...)` diese Regel automatisch beachtet. In SPSS sollte die Intervallzahl anstatt über „Anzahl der Intervalle“ besser über die „Intervallbreite“ gesteuert werden.

Bei den *normal probability Plots*, oder allgemein bei den *Quantile-Quantile-Plots*, kurz *Q-Q-Plots* genannt (vgl. http://en.wikipedia.org/wiki/Normal_probability_plot), wird die empirische (kumulative) Verteilung mit der theoretischen, hier der Normalverteilung, verglichen. Üblicherweise ist die empirische Stichprobenverteilung y und die theoretische x . Leider ist das bei SPSS genau umgekehrt. Dabei wird zu jedem beobachteten Wert das Quantil y ermittelt und mit dem Quantil x der Vergleichsverteilung als Punkt eingezeichnet. Im Idealfall liegen also die Punkte auf einer Geraden. Im Gegensatz zu den Histogrammen sind diese Grafiken unabhängig von Intervalleinteilungen, die möglicherweise ein Bild „verzerrten“ können.

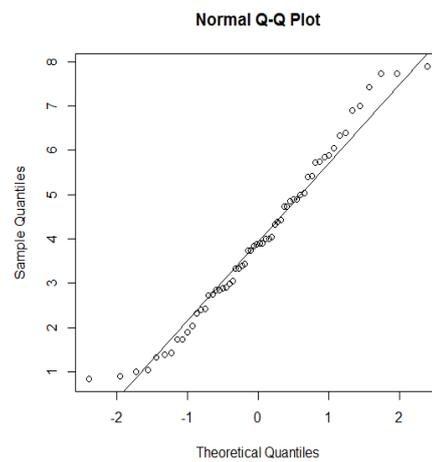
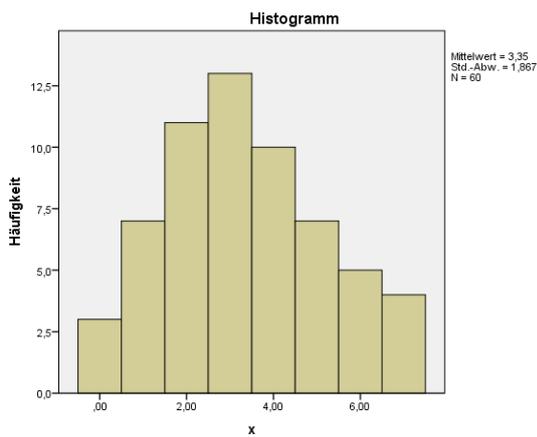
Aber sowohl die Interpretation von Histogrammen auch der Q-Q-Plots bedarf ein wenig Erfahrung. Die wichtigsten Kennzeichen einer Normalverteilung sind Symmetrie und Eingipfligkeit. Nachfolgend werden einige typische Verteilungsformen aufgezeigt, die zum Teil nicht mehr als normal eingestuft werden können. Das Ergebnis des Shapiro-Wilk-Tests, alle basierend auf einem $n=60$, wird zur Verdeutlichung ebenfalls angegeben:



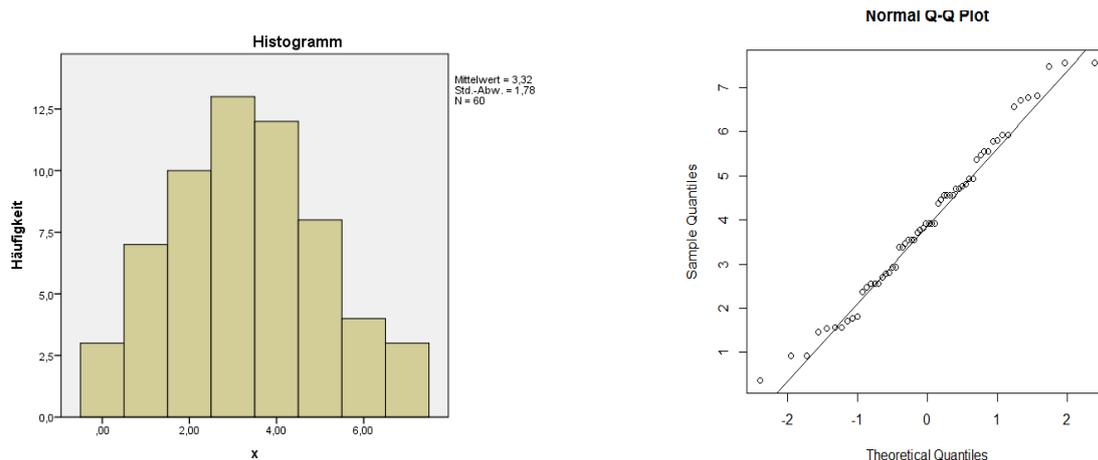
stark linksschiefe Verteilung (W=0.894 - p=0.001)



zweigipflige Verteilung (W=0.944 - p=0.008)



leicht linksschiefe Verteilung (W=0.955 - p=0.027)



(fast) normale Verteilung ($W=0.962 - p=0.056$)

Während die beiden ersten Beispiele eher krasse Fälle von nichtnormalverteilten Werten darstellen, wird manch einem kaum ein Unterschied zwischen den letzten beiden Histogrammen auffallen, die immerhin unterschiedliche Resultate aufweisen. Das rechte ist deutlich symmetrischer und daher eher als normalverteilt zu akzeptieren.

1.7 Prüfung von Voraussetzungen

Eine Warnung soll dieses Kapitel beenden. Am Thema „Prüfung von Voraussetzungen“ scheiden sich nämlich die Gemüter. Es wird nicht uneingeschränkt empfohlen, generell alle Voraussetzungen der parametrischen Anova zu prüfen. Der Grund: Zum einen sind die Prüfverfahren selbst unzuverlässig, d.h. sie können sowohl eine Abweichung von einer Voraussetzung anzeigen, obwohl diese gar nicht gegeben ist, als auch umgekehrt. Zum anderen haben diese Prüfverfahren wiederum Voraussetzungen, die nicht selten schärfer sind als die des eigentlichen Verfahrens, also hier der Varianzanalyse. Dagegen kann man sich, zumindest in beschränktem Maße, auf die Robustheit der Varianzanalyse verlassen. Vor diesem Hintergrund hatte Box [76] den inzwischen vielfach zitierten Satz geschrieben:

To make a preliminary test on variances is rather like putting to sea in a row boat to find out whether conditions are sufficiently calm for an ocean liner to leave port!

Diese Problematik wird in [5] behandelt, wo auch einige Beispiele dazu zu finden sind.

2. Nichtparametrische Varianzanalysen - Übersicht der Methoden

Nichtparametrische Varianzanalysen werden angewendet, wenn die Voraussetzungen für die parametrische Analyse nicht gegeben sind, d.h. wenn die abhängige Variable entweder metrisch ist und die Voraussetzungen „Normalverteilung der Residuen“ sowie „Varianzhomogenität“ nicht ausreichend erfüllt sind - als robustes Verfahren kann die Varianzanalyse i.a. einige Abweichungen von den idealen Voraussetzungen „vertragen“ - oder aber wenn die abhängige Variable ordinales oder dichotomes Skalenniveau hat. Darüber hinaus gibt es auch *semiparametrische* Verfahren, eine Mischform aus parametrischem und nichtparametrischem Modell, z.B. wenn an die Verteilung der abhängigen Variablen keine Bedingungen gestellt werden, aber eine Form der Varianzhomogenität vorausgesetzt wird. Während beim parametrischen Modell die abhängige Variable genau ein Verteilungsmodell annimmt, können beim nichtparametrischen Ansatz quasi beliebige Verteilungsformen auftreten. Und so ist es nicht verwunderlich, dass man praktisch für jedes Verfahren eine Verteilungsform für die abhängige Variable finden kann, so dass die Ergebnisse unbefriedigend sind: von der Verletzung des α -Risikos bis zu übermäßig konservativen Tests. Dies haben zahlreiche Simulationen gezeigt. Sogar Wikipedia widmet diesem Thema einen eigenen Artikel (vgl. [1]). Insofern ist es in der Praxis wenig hilfreich, die Voraussetzungen für die nichtparametrischen Verfahren allzupenibel zu überprüfen.

Neben den hier im Vordergrund stehenden „echten“ nichtparametrischen Verfahren darf nicht vergessen werden, dass es auch eine Reihe von robusten Tests für den Fall inhomogener Varianzen gibt, die vorzugsweise dann angewandt werden können und sollten, wenn die abhängige Variable metrisch ist, aber keine Varianzhomogenität vorliegt. Die Methoden werden in späteren Kapiteln vorgestellt.

Darüber hinaus gehören auch in diesen Kontext varianzanalytische Methoden für dichotome Merkmale, worauf später in Kapitel 7 kurz eingegangen wird.

Die wichtigsten Methoden werden im Folgenden kurz vorgestellt. Salazar-Alvarez et. al. [39] geben einen guten Überblick der nichtparametrischen Methoden zur mehrfaktoriellen Varianzanalyse. Eine leicht verständliche Einführung in diese Methoden bieten Erceg-Hurn und Mirosevich [5].

Welche Ansätze (Methoden) gibt es überhaupt? Dabei geht es im Wesentlichen um solche, die asymptotische Tests verwenden, also etwa für $n > 20$ (mit n Anzahl der Beobachtungen), wobei die Fallzahl bei abhängigen Stichproben durchaus geringer sein kann. In den letzten 15 Jahren sind eine Vielzahl von neuen Methoden zur nichtparametrischen Datenanalyse entwickelt worden, von denen nur die „wichtigsten“ hier erwähnt werden können. Dabei stehen solche im Vordergrund, die sich leicht mit Standardsoftware wie SPSS durchführen lassen. Trivialerweise lassen sich alle Verfahren in R (und natürlich S-Plus) realisieren.

2.1 Kruskal-Wallis und Friedman

Die klassischen nichtparametrischen Varianzanalysen sind die 1-faktoriellen Analysen mit den Tests von Kruskal & Wallis im Fall von unabhängigen Stichproben sowie dem von Friedman im Fall von abhängigen Stichproben (Messwiederholungen). Diese sind in (fast) allen gängigen Lehrbüchern ausführlich beschrieben.

Die asymptotische Effizienz des Kruskal-Wallis-Tests liegt bei 0.955, die des Friedman-Tests bei $0.955 * K / (K + 1)$, also z.B. 0.64 (für $K=2$) und 0.87 (für $K=10$), wobei K die Anzahl der

Gruppen (Versuchsbedingungen) ist. D.h. für große Stichproben ist der K-W-Test kaum schlechter als die parametrische Varianzanalyse.

Vielfach ist zu lesen, dass der Kruskal-Wallis-Test nicht nur auf Mittelwertunterschiede der zu vergleichenden Stichproben an, sondern verschiedentlich auch auf Unterschiede der Streuung und Schiefe anspricht (vgl. Wilcox in [45]). Andere Autoren teilen dagegen nicht diese Bedenken (vgl. Marascuilo & McSweeney [50]). Vargha & Delaney haben dieses Problem ausführlich untersucht und kommen zu dem Schluss, dass ein geringes Risiko besteht, dass der Test im Falle inhomogener Varianzen das α -Risiko leicht verletzt, also auch darauf anspricht. Daher wurden auch diverse „robuste“ Varianten dieses Tests entwickelt, u.a. der von Brunner, Dette und Munk (vgl. Kapitel 2.10).

Der Friedman-Test hat dazu im Vergleich eine geringe Effizienz. Iman und Davenport [36] haben den χ^2 -Wert des Friedman-Tests in einen F-Wert transformiert:

$$F = \frac{(n-1)\chi^2}{n(k-1) - \chi^2} \quad (2-1)$$

wobei n die Anzahl der Merkmalsträger ist. Dieser F-Wert mit $k-1$ Zähler-FG und $(k-1)(n-1)$ Nenner-FG hat deutlich bessere Eigenschaften und verleiht dem Friedman-Test eine höhere Effizienz. Für die Tests von reinen Messwiederholungseffekten bei mehrfaktoriellen Analysen, d.h. von Haupteffekten oder Interaktionen von Messwiederholungsfaktoren, ist oben $(k-1)$ durch die Zählerfreiheitsgrade des Tests zu ersetzen.

Die Anwendung dieser Korrektur erübrigt sich selbstverständlich, wenn der χ^2 -Wert bereits als signifikant ausgewiesen worden ist. Wie auch die Beispiele in den Kapiteln 5 und 6 zeigen, sollte man von dieser Korrektur nicht zu viel erwarten.

Es sei noch erwähnt, dass es eine analoge Umrechnung des χ^2 -Werts des Kruskal-Wallis-Tests in einen F-Wert von Iman und Davenport gibt (vgl. Conover & Iman [6]). Die ist dann allerdings mit dem F-Test des RT-Verfahrens (siehe nächstes Kapitel) identisch.

Eine Erweiterung dieser Verfahren auf mehrfaktorielle Versuchspläne erfolgt in Kapitel 2.5. Dort wird auch kurz gezeigt, dass sich die Verfahren von Kruskal & Wallis sowie von Friedman auf die „klassische“ Varianzanalyse zurückführen lassen.

2.2 Rank transform Tests (RT)

Dies sind klassische Anova-F-Tests angewandt auf Rangdaten. D.h. alle Werte der abhängigen Variablen, über Gruppen und Messwiederholungen hinweg, werden zunächst in Ränge umgerechnet, bevor dann eine klassische Varianzanalyse mit F-Tests durchgeführt wird. Das Verfahren wurde 1981 von Iman und Conover [6] vorgeschlagen und galt lange als eine brauchbare Lösung, bis in den 90er Jahren Simulationen einige Schwächen aufzeigten. So wird u.a. eine Verletzung des α -Risikos für den Test der Interaktion berichtet, wenn zugleich signifikante Haupteffekte bestehen (vgl. u.a. [17]). Der Grund dafür: die Additivität der Haupt- und Interaktionseffekte, d.h. die Unabhängigkeit der Tests, bleibt bei der Rangtransformation nicht erhalten (vgl. Beasley & Zumbo, [53]). Auf der anderen Seite konnten Hora und Iman [37] sowohl theoretisch als auch durch Simulationen zeigen, dass zum einen die Tests der Haupteffekte in jedem Fall asymptotisch, d.h. für größere n , valide sind, d.h. dass das Risiko für den Fehler 1. Art konsequent eingehalten wird, und zum anderen diese Tests stärker sind als die klassischen Tests von Kruskal-Wallis und Friedman oder auch als der von Quade.

Der Reiz dieser Methode liegt in der Einfachheit. Sie ist auch empfehlenswert, solange nicht eine Interaktion als signifikant ausgewiesen wird und zugleich mindestens ein Haupteffekt signifikant ist. In einem solchen Fall sollte dann für diese Variable das ART-Verfahren (siehe unten) angewandt werden.

2.3 Inverse normal transform (INT)

Eine Verbesserung der o.a. RT-Methode bringt die *inverse Normalverteilungs-Transformation* (*inverse normal transform*, INT). Bei dieser werden die oben erzeugten gleichverteilten RT-Werte R_i in (standard-) normalverteilte Scores umgerechnet:

$$\Phi^{-1}(R_i/(n+1)) \quad (2-2)$$

wobei Φ die Standardnormalverteilung und n die Anzahl aller Werte insgesamt ist. (Diese Division durch $n-1$ ist erforderlich, um den Wertebereich $1 \dots n$ in das Intervall $0 \dots 1$ zu transformieren.) Wie bei der o.a. RT-Methode werden dann für die transformierten Werte (normal scores) die klassischen F-Tests durchgeführt. Von dieser Transformation gibt es mehrere Varianten, die sich im Wesentlichen auf eine Formel zurückführen lassen:

$$\Phi^{-1}((R_i - c)/(n + 1 - 2c)) \quad (2-3)$$

Die o.a. zuerst aufgeführte, vielfach als *normal score test* bezeichnete Variante, erhält man z.B. über $c=0$. Huang [41] hat mittels Simulationen gezeigt, dass bei Verwendung dieser Methode (im Gegensatz zur RT-Methode) das α -Risiko auch für die Interaktionen nicht verletzt wird. Zu einem ähnlichen Ergebnis kommen Mansouri und Chang [44]. Eine ausführliche Darstellung dieser Methoden ist bei Beasley, Erickson & Allison [40] zu finden. Allerdings zeigen Letztere Beispiele auf, bei denen dennoch das α -Risiko leicht verletzt wird.

Das INT-Verfahren geht u.a. auf van der Waerden in den 50er-Jahren zurück (vgl. Kapitel 2.7). Es ist zuletzt durch die Analyse von Gendaten wieder aktuell und beliebt geworden, da es auf der einen Seite ähnlich leicht wie das RT-Verfahren zu rechnen ist und auf der anderen Seite die falsch signifikanten Testergebnisse weitgehend vermeidet und zudem noch eine hohe Effizienz hat.

2.4 Aligned rank transform (ART)

Eine andere Methode, die bei der o.a. RT-Methode möglichen fälschlich signifikanten Interaktionen zu vermeiden, wenn zugleich signifikante Haupteffekte vorliegen, bieten die *aligned rank transforms* oder auch *aligned rank tests* (ART). Hierbei werden zunächst die Daten bzgl. der Haupteffekte bereinigt. Hierzu gibt es u.a. zwei Methoden, eine einfache und eine etwas aufwändigere, die bei balancierten Versuchsplänen zu demselben Ergebnis führen, im allgemeinen Fall jedoch *nicht*.

- Der *naive approach* (ART1): Zunächst werden von der Kriteriumsvariablen die Haupteffekte der Faktoren subtrahiert, die an der zu untersuchenden Interaktion beteiligt sind. Die Ergebnisvariable wird anschließend in Ränge umgerechnet und dann mit dem RT-Verfahren weiter analysiert, um die Interaktion zu testen. Für den Test der Interaktion wird also anstatt x die Variable $x_{ijm} - \alpha_i - \beta_j$ untersucht, oder mit den Werten der Stichprobe:

$$x'_{ijm} = x_{ijm} - \bar{x}_i - \bar{x}_j + 2\bar{x} \quad (2-4)$$

wobei \bar{x}_i , \bar{x}_j , \bar{x} die Gruppenmittelwerte bzgl. der Faktoren A und B bzw. der Gesamtmitwert sind.

- Der *standard approach* (ART2): Zunächst wird eine komplette Varianzanalyse der Kriteriumsvariablen (mit allen Effekten) durchgeführt. Zu den daraus resultierenden Residuen wird der untersuchte Interaktionseffekt, abzüglich der daran beteiligten Haupteffekte, addiert. Die Ergebnisvariable wird anschließend in Ränge umgerechnet und dann mit dem RT-Verfahren weiter analysiert, um die Interaktion zu testen. Für den Test der Interaktion wird also anstatt x die Variable

$$x'_{ijm} = e_m + (\alpha\beta_{ij} - \alpha_i - \beta_j + 2\mu) \quad (2 - 5)$$

untersucht, wobei e_m die Residuen des kompletten varianzanalytischen Modells, α_i , β_j , $\alpha\beta_{ij}$, μ die Effekte der Faktoren A und B bzw. der Gesamtmittelwert sind.

Dieses Verfahren wird daher auch mit RAA (*ranking after alignment*) bezeichnet. Neben den beiden o.a. Methoden gibt es inzwischen noch eine Vielzahl weiterer Varianten von ART. So wurden u.a. von Peterson [54] Alignments (Korrekturen) mittels robuster Mittelwerte wie Median oder getrimmter Mittelwerte anstatt des arithmetischen Mittels vorgeschlagen. Für die Datentransformation wird ein spezielles Programm (*ARTool*) angeboten (vgl. [20]), das Microsoft .NET 2.0 Framework voraussetzt. Die transformierten Daten können dann mit einem Standardprogramm wie SPSS analysiert werden. Dieses Verfahren schneidet bei Vergleichen relativ gut ab. Die Methode ist beschrieben in [19], [20] und [21].

Das ART-Verfahren kann aber auch mit ein wenig Aufwand ohne Zusatzsoftware in R - für R gibt es das Paket `ARTOOLS`, allerdings nicht für Designs mit Messwiederholungen- oder SPSS angewandt werden, wie die Beispiele in den nachfolgenden Kapiteln demonstrieren. Im Wesentlichen müssen Aggregatdaten wie Mittelwerte ermittelt werden, die in die Berechnungen einfließen. Es sei ausdrücklich darauf hingewiesen, dass der Aufwand des ART- gegenüber dem RT-Verfahren nicht generell erforderlich ist, um falsch signifikante Ergebnisse zu vermeiden. Lediglich in dem Fall, dass eine Interaktion als signifikant ausgewiesen wird und zugleich mindestens ein Haupteffekt signifikant ist, sollte für die untersuchte Variable das ART-Verfahren angewandt werden. Dennoch werden bei den Beispielen in diesem Skript meistens auch Alignments für die Haupteffekte durchgeführt, allerdings nur zu Demonstrationszwecken.

2.5 Verallgemeinerte Kruskal-Wallis- und Friedman-Analysen (KWF, Puri & Sen)

Mit diesem Namen wird im Folgenden ein Verfahren bezeichnet, das eine Verallgemeinerung der o.a. Tests von Kruskal & Wallis (bei unabhängigen Stichproben) bzw. von Friedman (bei abhängigen Stichproben) darstellt. Beide Tests lassen sich auf die „klassische“ 1-faktorielle Varianzanalyse zurückführen. Die verallgemeinerten Kruskal-Wallis- und Friedman-Analysen erweitern dies auf beliebige Versuchspläne.

Auch hier werden alle Werte zunächst in Ränge umgerechnet, bevor dann eine klassische Varianzanalyse durchgeführt wird. Anstatt des F-Tests wird allerdings ein χ^2 -Test durchgeführt. Die Testgröße errechnet sich im Fall von Versuchsplänen ohne Messwiederholungen als

$$\chi^2 = \frac{SS_{Effekt}}{MS_{total}} \quad (2 - 6a)$$

bzw. für Gruppierungsfaktoren im Fall von Versuchsplänen mit Messwiederholungen als

$$\chi^2 = \frac{SS_{Effekt}}{MS_{zwischen}} \quad (2 - 6b)$$

bzw. im Fall von Messwiederlungsfaktoren als

$$\chi^2 = \frac{SS_{\text{Effekt}}}{(SS_{\text{Effekt}} + SS_{\text{Fehler}}) / (df_{\text{Effekt}} + df_{\text{Fehler}})} \quad (2 - 7)$$

wobei SS_{Effekt} die Streuungsquadratsumme (Sum of Squares) des zu testenden Effektes, MS_{total} die Gesamtvarianz (Mean Square), MS_{zwischen} die Varianz (Mean Square) zwischen den Versuchspersonen und SS_{Fehler} die Streuungsquadratsumme des zum getesteten Effekt gehörenden Fehlers ist, gegebenenfalls ergänzt um Streuungsquadratsummen von Interaktionen, die alle aus der Anova-Tabelle abgelesen werden können. Falls nur ein Messwiederholungsfaktor vorliegt, entspricht der Nenner in 2-7 der Varianz $MS_{\text{innerhalb}}$. Als Freiheitsgrade für den χ^2 -Test nimmt man die Zählerfreiheitsgrade aus der Varianzanalyse. Für die Haupteffekte ergibt dies die Tests der bekannten nichtparametrischen Anova von Kruskal-Wallis bzw. Friedman. Diese Methode gilt als relativ konservativ, insbesondere für den Test der Interaktion. Sie kann aber bedenkenlos auf ordinale Merkmale angewandt werden.

Der Ansatz geht in die 60er Jahre zurück auf Bennett [29], wurde von Shirley [30] sowie Scheirer, Ray & Hare [31] erweitert und schließlich von Puri & Sen [12] systematisch dargestellt. Das Verfahren ist in der Literatur auch als *L-Statistik* von Puri & Sen bekannt. Die 1-faktorielle Analyse ist auch bei Winer [13] nachzulesen. Diese Tests werden im Folgenden auch mit Puri & Sen- oder KWF-Tests bezeichnet.

Die Umsetzung in R bzw. SPSS ist natürlich nicht ganz so simpel wie bei den RT- und INT-Verfahren. In der Regel genügt die Erzeugung einer neuen rangtransformierten Variablen. Allerdings müssen bei Messwiederholungen die Ränge nach dem Friedman-Verfahren, also fallweise, ermittelt werden, was insbesondere bei SPSS zusätzlichen Aufwand erfordert, nämlich die Umstrukturierung der Datenmatrix. Bei zwei oder mehr Messwiederholungsfaktoren wird der Aufwand allerdings größer. Hinzu kommt die Durchführung der χ^2 -Tests, die insbesondere in SPSS mit dem Taschenrechner erfolgen muss.

2.6 van der Waerden

Die Methode von van der Waerden (vgl. Wikipedia [42]) vereinigt gewissermaßen die INT-Methode sowie die Verfahren von Kruskal & Wallis und Friedman. Van der Waerden hat das Verfahren als 1-faktorielle Varianzanalyse für unabhängige Stichproben entwickelt: Zunächst werden wie bei der INT-Methode die normal scores errechnet. Auf diese wird dann der χ^2 -Test wie beim Kruskal-Wallis-Test angewandt, z.B. in der Berechnung wie bei (2-4). Mansouri und Chang [44] haben das Verfahren auf mehrfaktorielle Versuchspläne verallgemeinert. Dieses funktioniert dann so, dass die verallgemeinerten KWF-Tests auf die normal scores angewandt werden. Marascuilo und McSweeney [50] haben analog einen Test für abhängige Stichproben konstruiert, allerdings nur für einen Messwiederholungsfaktor. Für die Interaktion mehrerer Messwiederholungsfaktoren fehlt noch ein entsprechendes Verfahren.

Der Test hat bei 1-faktoriellen Varianzanalysen für unabhängige Stichproben eine asymptotische Effizienz von 1, ist also der parametrischen Varianzanalyse ebenbürtig, und ist im Fall, dass die Voraussetzungen der klassischen Varianzanalyse nicht erfüllt sind, ihr deutlich überlegen (vgl. Sheskin [38], der sich auf Conover sowie Marascuilo & McSweeney bezieht).

Der Aufwand ist exakt derselbe wie bei den im vorigen Kapitel skizzierten verallgemeinerten Kruskal-Wallis- und Friedman-Analysen.

2.7 Bredenkamp Tests

Schon 1974 entwickelte Bredenkamp [15] eine Verallgemeinerung der Tests von Kruskal-Wallis und Friedman auf zwei faktorielle Analysen. Und zwar werden unter Ausnutzung der Additivität der χ^2 -Werte die Interaktion auf mehrere 1-faktorielle Analysen zurückgeführt. Dazu wird ein einfaktorieller Test über alle Zellen hinweg durchgeführt und anschließend von dem resultierenden χ^2 -Wert die χ^2 -Werte der Haupteffekte subtrahiert. Die Methode lässt sich auch auf mehr als zwei Faktoren erweitern. Das Verfahren lässt sich ohne Aufwand mit Standardsoftware durchführen. Diese Methode ist mit dem Verfahren 2 identisch, ist allerdings zum einen nur für balancierte Versuchspläne gültig. Zum anderen bringt sie bei mehrfacher Messwiederholung falsche Ergebnisse. Die Tests sind auch ausführlich beschrieben bei Lienert (vgl. [1], S. 1024 ff).

2.8 Akritas, Arnold & Brunner ATS Tests

Akritas, Arnold und Brunner stellen ein anderes Modell mit beliebigen Verteilungen vor, das nicht einfach auf einer Umrechnung der Werte in Ränge basiert (vgl. [22], gut verständlich dargestellt in [26]).

Ein Begriff, der bei diesem Verfahren eine wichtige Rolle spielt, ist der *relative Effekt*. Er dient zur Unterscheidung zwischen zwei Verteilungen, etwa der Zufallsvariablen X_1 und X_2 . Der relative Effekt von X_2 zu X_1 ist definiert als $p^+ = P(X_1 \leq X_2)$, d.h. durch die Wahrscheinlichkeit, dass X_1 kleinere Werte annimmt als X_2 . Dabei hat X_1 eine stochastische Tendenz zu größeren Werten als X_2 , falls $p^+ < 1/2$ und eine stochastische Tendenz zu kleineren Werten, falls $p^+ > 1/2$ ist. Ausführliche Ausführungen hierzu sind bei E. Brunner & U. Munzel [3] zu finden.

Trotz des anderen Ansatzes mit beliebigen Verteilungen resultieren dann dennoch im Wesentlichen ähnliche F-Quotienten wie bei Rank transform Tests. Allerdings werden sehr viel differenziertere Freiheitsgrade verwendet. Wegen der Ähnlichkeit zu den F-Tests der Anova werden sie ATS (*Anova type statistic*) genannt. Parallel zu den ATS bieten die Autoren auch eine weitere χ^2 -verteilte Statistik WTS (*Wald type statistic*) an, die aber hier nicht berücksichtigt wird, da die ATS bessere Eigenschaften aufweist. Letztlich werden dabei, anstatt Mittelwerte, die relativen (Behandlungs) Effekte p_i verglichen, mit

$$p_i = (\bar{R}_i - 0,5)/n \quad (\text{mit } \bar{R}_i = \text{mittlerer Rang und } n = \sum n_i)$$

Dieser Ansatz wird in [23] auf multivariate Analysen, in [24] auf Analysen mit Messwiederholungen sowie in [35] auf Kovarianzanalysen erweitert. Bei letzteren sind sogar fehlende Werte erlaubt und es gibt Lösungen sowohl für den Fall homogener Varianzen-Kovarianzen (*compound symmetry*) als auch für den allgemeinen Fall. Diese Tests haben eine vergleichsweise hohe Effizienz und sind ausdrücklich auch für ordinale und dichtotome abhängige Variablen anwendbar. Sie zeichnen sich insbesondere dadurch aus, dass das α -Niveau immer exakt eingehalten wird. Negativ wird vermerkt, dass die Tests nicht nur auf Mittelwertunterschiede, sondern auch auf andere Verteilungsunterschiede ansprechen.

Für Varianzen mit Messwiederholungen gibt es inzwischen auch das R-Paket nparLD. In SPSS sind diese Tests wegen der erforderlichen umfangreichen Matrizenrechnungen noch nicht durchführbar.

2.9 Wilcoxon analysis (WA)

Hettmansperger and McKean [32] haben eine nichtparametrische Regression, *Wilcoxon Analysis* (WA), entwickelt, bei die Ränge der Residuen die zentrale Rolle spielen und somit der Einfluss von Ausreißern reduziert wird. Trivialerweise lässt sich der Ansatz auf die Varianzanalyse anwenden. Eine Erweiterung dieser Methode ist die *weighted Wilcoxon technique* (WW), bei der auch die x-Variablen in Ränge transformiert werden. Dieses Verfahren zählt zu den semiparametrischen, da es auf den Parametern der linearen Regression basiert. Diese Methode wird derzeit hier nicht behandelt. Es gibt aber das R-Paket `RFit` zur Anwendung dieser Methode in R [33].

2.10 Weitere Varianzanalysen

2.10.1 Brunner, Dette und Munk

Im Zusammenhang mit der Analyse von Kruskal und Wallis wurde oben der Test von Brunner, Dette und Munk (BDM-Test) erwähnt. Er bietet sich an, wenn die Streuungen der Gruppen als unterschiedlich anzusehen sind, da letztlich alle o.a. Methoden auf inhomogene Varianzen reagieren können. Das Verfahren ähnelt dem o.a. von Akritas, Arnold und Brunner, was nicht verwunderlich ist, da zum Teil dieselben Autoren beteiligt sind. Die Durchführung des Tests ist relativ komplex, da sie wie die ATS auf fortgeschrittener Matrix-Algebra basiert. Das Verfahren ist von Wilcox in [45] und [46] beschrieben. Das Verfahren gibt es als 1- und 2-faktorielle Varianzanalyse (für unabhängige Stichproben).

Ein anderer Test von *Rust & Fligner* ist ebenfalls in den o.a. Büchern von Wilcox beschrieben. Dieser wird allerdings gegenüber dem oben erwähnten BDM-Test als weniger empfehlenswert angesehen, insbesondere da er keine Bindungen erlaubt.

2.10.2 Quade

Der Test von *Quade* (vgl. Wilcox et al [47]) ist ein globaler Test auf Gleichheit der Mittelwerte bei Messwiederholungen, ähnlich dem Friedman-Test. Er liegt bislang nur als 1-faktorielle Analyse vor.

Die Idee ist folgende: Bei der Rangbildung R_{ji} für die Friedman-Analyse, bei der pro Fall/Merkmalsträger j ($j=1, \dots, N$) die Werte $i=1, \dots, K$ vergeben werden, ist nur eine geringe Differenzierung zwischen den k Gruppen möglich. Daher wird eine Fallgewichtung Q_j eingeführt, die Fälle mit einem größeren Wertespektrum bevorzugt. Q_j errechnet sich aus der Spannweite D_j der Werte eines Falls (Differenz von Maximum und Minimum der x_{ji}), die dann in Ränge umgerechnet wird. Aus beiden Rängen R_{ji} und Q_j zusammen wird dann das Produkt $W_{ji} = Q_j * R_{ji}$ errechnet. Zum Vergleich zweier Gruppen werden schließlich die Rangsummen von W_{ji} verwendet:

$$T_i = \left(\sum_{j=1}^n W_{ji} \right) / (N(N+1)/2)$$

die dann in einen t- oder z-Test umgerechnet werden.

Der Quade-Test hat für $K < 6$ eine größere Teststärke als der Friedman-Test und ist daher diesem überlegen (vgl. u.a. Wikipedia [43]). Auf der anderen Seite wird er nicht für ordinal-skalierte Variablen empfohlen.

2. 10. 3 Skillings & Mack

Der Test von Skillings & Mack ist ebenfalls eine Alternative zum Friedman-Test, also für abhängige Stichproben (Messwiederholungen), allerdings für den Fall von fehlenden Werten. Er ist anschaulich beschrieben von Chatfield und Mander [48]. Auch dieses Verfahren liegt bislang nur als 1-faktorielle Analyse vor.

Liegen weder fehlende Werte noch Bindungen vor, so liefern die Tests von Skillings & Mack und von Friedman dieselben Resultate. Im Fall von vielen Bindungen und/oder kleinen Fallzahlen ist dieser Test dem von Friedman leicht überlegen.

An dieser Stelle sei darauf hingewiesen, dass das in Kapitel 2.8 erwähnte Verfahren von Akritas, Arnold und Brunner in der Version des R-Pakets `npard` auch fehlende Werte zulässt.

2. 10. 4 Beasley & Zumbo

Beasley & Zumbo haben eine Reihe von Tests für die Interaktion bei gemischten Versuchsplänen, d.h. solchen mit sowohl Gruppierungs- als auch Messwiederholungsfaktoren, zusammengestellt (vgl. [53]). Neben einigen Verfahren, die relativ aufwändig zu programmieren sind, sind auch die Interaktion aus dem KWF- sowie aus dem ART-Verfahren angeführt. Deren Fazit: I.a. ist die ART-Prozedur den anderen vorzuziehen.

2. 11 Logistische Regression

Neben der bekannten logistischen Regression für dichotome Kriteriumsvariablen gibt es auch eine für ordinale Variablen. Unter dem Aspekt, dass die parametrische Varianzanalyse ein Spezialfall der linearen Regression ist, bei der die nominalen Prädiktoren passend kodiert werden, ist es einleuchtend, dass dasselbe Vorgehen auch bei der ordinalen logistischen Regression zu einer Varianzanalyse für ordinale Kriteriumsvariablen führt. Unter praktischen Aspekten müssen allerdings drei Einschränkungen gemacht werden:

- Erstens ist eine relativ hohe Fallzahl erforderlich,
- zweitens führt das Iterationsverfahren der Maximum-Likelihood-Schätzung nicht immer zum Erfolg, d.h. verschiedentlich gibt es kein Ergebnis, und
- drittens sollte die abhängige Variable nicht zu viele Ausprägungen haben (unter 10).

Im Gegensatz zu den zuvor aufgeführten Verfahren, die alle primär für metrische Kriteriumsvariablen konzipiert, allerdings auch für ordinale Variablen anwenbar sind, ist die ordinale logistische Regression eine Methode, die speziell auf ordinale Merkmale zugeschnitten ist. Die Anwendung ist allerdings nicht so ganz so einfach wie die der übrigen Verfahren.

2. 12 Voraussetzungen

Die meisten oben vorgestellten Verfahren basieren auf einer Rangtransformation und sind in erster Linie für metrische Variablen gedacht, die die Normalverteilungs-Voraussetzung nicht erfüllen, *nicht* jedoch für Variablen mit *beliebigen* Eigenschaften. D.h. hat die untransformierte Variable x ungleiche Varianzen, so kann das auch noch für die rangtransformierte Variable $R(x)$ gelten. Das gilt insbesondere für die RT-, ART-, und INT-Verfahren. U.a. haben Beasley & Zumbo [53] im Falle der ART-Prozedur darauf hingewiesen. Durch die Rangtransformation werden Verteilungsdeformationen bestenfalls abgemildert, nicht aber beseitigt. So ist es sinnvoll, gegebenenfalls auch $R(x)$ auf Varianzhomogenität zu überprüfen und gegebenenfalls einen der

in Kapitel 4.3.3 vorgestellten Tests zu benutzen. Verschiedentlich wird auch beim Kruskal-Wallis-Test darauf hingewiesen, dass dieser auch auf inhomogene Varianzen anspricht (vgl. Wilcox in [45]).

Im Fall von unabhängigen Stichproben empfiehlt sich zum Test der Varianzhomogenität von $R(x)$ der *Levene-Test*, da dieser (im Gegensatz zum klassischen F-Test, zum F_{\max} -Test oder zum Bartlett-Test) robust gegen Abweichungen von der Normalverteilung ist und auch für ordinale Variablen anwendbar ist. Allerdings gibt es kaum Alternativen für den Fall, dass sich die Varianzen auch nach der Transformation als inhomogen erweisen. Eine Möglichkeit besteht in der Box-Korrektur der Freiheitsgrade (vgl. Anhang 2.1), die im Fall der RT-, ART-, und INT-Verfahren angewendet werden kann. Leider ist die Box-Korrektur weder in R noch in SPSS standardmäßig verfügbar.

Im Fall von abhängigen Stichproben (Messwiederholungen) muss man notgedrungen den Mauchly-Test (vgl. Kapitel 5.2) benutzen, wenn dieser auch nicht annähernd die robuste Eigenschaften eines Levene-Tests hat. Es gibt zwar einen entsprechenden Test für Rangdaten von *Hallin und Paindaveine* [49], der aber noch nicht in den Softwaresystemen verfügbar ist. *Beasley und Zumbo* [53] propagieren, bei den F-Tests einfach eine der Korrekturen der Freiheitsgrade von Huynh-Feldt oder Greenhouse-Geisser vorzunehmen, ohne das Ergebnis des Mauchly-Tests zu berücksichtigen.

Auf der anderen Seite kann geschlossen werden: Erfüllen die nichttransformierten Daten die Voraussetzung der Varianzhomogenität, so gilt diese auch für die rangtransformierten Daten, so dass gegebenenfalls eine Überprüfung dafür entfallen kann.

Da bei dem KWF- und van der Waerden-Verfahren χ^2 - anstatt F-Tests durchgeführt werden, ist bei diesen die Voraussetzung der Varianzhomogenität, insbesondere der Sphärizität, von untergeordneter Bedeutung, so dass entsprechende Tests entfallen können. Dafür muss man allerdings konservativere Tests in Kauf nehmen.

2.13 Vergleiche

Die RT-, ART- und KFW-Methoden werden in [16] und [17] verglichen (durch eigene Simulationen und Verweise auf ähnliche Simulationsergebnisse anderer Autoren) hinsichtlich des Verhaltens von

- α , d.h. ob das vorgegebene α vom Test eingehalten wird, und
- β , d.h. wie konservativ der Test im Vergleich zum parametrischen Test reagiert.

Einen ähnlichen Vergleich hinsichtlich der nichtparametrischen Kovarianzanalyse gibt es in [18]. Einen umfangreichen Vergleich der Methoden, insbesondere zum Test der Interaktion bei der nichtparametrischen Varianzanalyse, ist bei Sawilowski [16] zu finden, der allerdings aus 1990 stammt und daher neuere Methoden nicht berücksichtigt. Dort werden aus zahllosen Artikeln die Pros und Contras der Verfahren zusammengestellt.

Mansouri und Chang [44] vergleichen die INT-Verfahren (normal scores und van der Waerden) u.a. mit dem RT-Verfahren. Ein Vergleich des ATS-Tests mit anderen Methoden wird in [34] geboten.

Eine Übersicht fast aller Verfahren mit einem Vergleich der Fehlerraten 1. Art und der Power auf Basis verschiedener Simulationen bietet Danaba [51], wenn auch diese Arbeit wegen typografischer Mängel nicht ganz einfach zu lesen ist.

3. Funktionen zur Varianzanalyse in R und SPSS

Auch für die nichtparametrischen Varianzanalysen greift man fast immer auf die klassischen parametrischen Methoden zurück, um anschließend die Ergebnisse weiterzuverarbeiten. Daher nachfolgend ein Überblick über die Möglichkeiten in R und SPSS.

3.1 Funktionen in R

Varianzanalysen sind in R nicht so problemlos durchzuführen, wie man erwarten sollte. Das hat im Wesentlichen zwei Gründe:

- Zum einen verwendet R für die in der Programmiersprache S vorgesehene Funktion `aov` die Berechnung der Streuungsquadrate vom Typ I, eine Methode, die zum einen problematisch ist und zum anderen von kaum einem anderen Programm benutzt wird (vgl. dazu das Kapitel 4.3.1.1). Weitere Hinweise hierzu bieten Scholer [111] und Meyer [112].
- Zum anderen müssen viele im Zusammenhang mit der Varianzanalyse erforderlichen Tests (z.B. Varianzhomogenitätstests oder multiple Mittelwertvergleiche) mühsam mit anderen Funktionen durchgeführt werden, was allerdings in R nicht unüblich ist.

Das hat dazu geführt, dass es inzwischen fast zahllose Funktionen zur Varianzanalyse in diversen hinzuzuladenden Paketen gibt. Von denen können hier nur wenige erwähnt werden.

Generell müssen die Faktoren, die unabhängigen Variablen, deren Einfluss getestet werden soll, vom Typ „factor“ sein, auch wenn sie nur zwei Stufen (Ausprägungen) haben. Darüber hinaus ist vielfach, insbesondere bei Messwiederholungen, eine numerische Fallkennung „subject“ erforderlich.

- `aov`
`aov (abh.Variable ~ Faktor1*Faktor2*..., Dataframe)`
für unabhängige Stichproben
`aov (abh.Variable ~ Faktor1*... + Error(subject/Faktor1*..., Dataframe)`
für abhängige Stichproben
`aov` berechnet Quadratsummen vom Typ I. Um solche vom Typ III zu erhalten, sind folgende Schritte erforderlich:
1. zu Beginn der Sitzung einmal:
`options (contrasts=c("contr.sum", "contr.poly"))` (vgl. Kapitel 9.4)
2. Wenn `model` das Ergebnis von `aov` enthält, dann werden die Quadratsummen vom Typ III mit Tests ausgegeben über
`drop1 (model, ~. , test="F")`
- `lm`
`anova (lm (abh.Variable ~ Faktor1*Faktor2*..., Dataframe)`
für unabhängige Stichproben
Um Quadratsummen vom Typ III zu erhalten, sind dieselben Schritte wie bei `aov` erforderlich.
Vorteil gegenüber `aov`: Die Ergebnisse, wie z.B. die Quadratsummen lassen sich weiterverarbeiten, was vielfach erforderlich ist.
- `lm` mit `Anova` (im Paket `car`)
`Anova (lm (abh.Variable ~ Faktor1*Faktor2*..., Dataframe), type="III")`
direkte Berechnung der Quadratsummen vom Typ III und mit weiterverarbeitbaren Ergebnissen.

- **ezANOVA (im Paket ez)**
`ezANOVA (Dataframe, .(abh.Variable), .(subject),
between=.(Faktoren), within=.(Faktoren))`
sowohl für Gruppierungsfaktoren (`between=.`) als auch für Messwiederholungsfaktoren (`within=.`)
bei Messwiederholungsfaktoren Ausgabe des Mauchly-Tests sowie der modifizierten Tests von Geisser & Greenhouse sowie von Huynh & Feldt, sonst Ausgabe des Levene-Tests. Berechnung der Quadratsummen vom Typ III möglich (`type=3`).
Diese Funktion ist zwar einfach zu benutzen, hat aber zwei Schwächen: zum einen muss immer eine numerische Fallkennung *subject* angegeben werden, zum anderen meldet sie häufig fälschlicherweise Eingabefehler oder ungültige Variablenangaben.
- **nparLD (im Paket nparLD)**
`nparLD (abh.Variable~Faktor1*Faktor2*..., Dataframe, subject)`
für nichtparametrische Analysen mit Messwiederholungen nach dem ATS-Verfahren von Akritas, Arnold & Brunner.
Es können auch Versuchspläne mit fehlenden Werten analysiert werden. Dafür stehen je nach Design die Funktionen `f1.ld.f1`, `f2.ld.f1`, `f1.ld.f2`, `ld.f1` und `ld.f2` zur Verfügung.
- **oneway.test**
`oneway.test (abh.Variable ~ Faktor, Dataframe)`
für unabhängige Stichproben
Robuste 1-faktorielle Varianzanalyse für inhomogene Varianzen nach dem Verfahren von Welch.
- **friedman.test**
`friedman.test (Datenmatrix)`
1-faktorielle nichtparametrische Varianzanalyse mit Messwiederholungen nach dem Verfahren von Friedman.
- **quade.test**
`quade.test (Datenmatrix)`
1-faktorielle nichtparametrische Varianzanalyse mit Messwiederholungen nach dem Verfahren von Quade.
- **waerden.test (im Paket agricolae)**
`waerden.test (abh.Variable, Faktor, group=F, console=T)`
1-faktorielle Varianzanalyse mit normal scores nach dem Verfahren von van der Waerden.
- **BDM (im Paket asbio)**
`BDM.test (abh.Variable, Faktor)`
`BDM.2way (abh.Variable, Faktor1, Faktor2)`
1- bzw. 2-faktorielle robuste Varianzanalyse nach dem Verfahren von Brunner, Dette, Munk.
- **SkiMack (im Paket Skillings.Mack)**
`SkiMack (as.matrix(Datenmatrix))`
1-faktorielle Varianzanalyse mit Messwiederholungen bei fehlenden Werten nach dem Verfahren von Skillings & Mack.

3.2 Funktionen in SPSS

Varianzanalysen sind mit SPSS vergleichsweise einfach durchzuführen. Zur Verfügung stehen:

- `Oneway`
`Oneway abh.Variable BY Faktor`
 (Menü: Mittelwerte vergleichen -> einfaktorische ANOVA)
 1-faktorielle Analyse für unabhängige Stichproben.
 Unter „Optionen“ kann der Levene-Test auf Gleichheit der Varianzen sowie die F-Tests von Welch und Brown & Forsythe im Falle von heterogenen Varianzen angefordert werden.
- `Unianova`
`Unianova abh.Variable BY Faktor1 Faktor2...`
 (Menü: Allgemeines lineares Modell -> Univariat)
 mehrfaktorielle Analyse für unabhängige Stichproben.
 Unter „Optionen“ kann der Levene-Test auf Gleichheit der Varianzen angefordert werden.
 Unter „Modell“ kann die Methode zur Berechnung der Streuungsquadrate gewählt werden (Typ I, II oder III).
- `GLM`
`GLM Messwiederholungsvariablen BY Faktor1 Faktor2...`
 `/WSFactor=... /WSDesign=... /Design=...`
 (Menü: Allgemeines lineares Modell -> Messwiederholung)
 mehrfaktorielle Analyse für unabhängige und abhängige Stichproben.
 Unter „Optionen“ kann der Levene-Test auf Gleichheit der Varianzen bzw. der Box-Test auf Gleichheit Kovarianzmatrizen angefordert werden.
 Unter „Modell“ kann die Methode zur Berechnung der Streuungsquadrate gewählt werden (Typ I, II oder III).
 Mauchly's Test auf Sphärität sowie der modifizierte Tests von Geisser & Greenhouse bzw. von Huynh & Feldt werden immer ausgegeben.
- `Nptests`
`Nptests`
 `/independent test (abh.Variable) group (Faktor) kruskal_wallis`
 `/related test (Messwiederholungsvariablen) friedman`
 (Menü: Nichtparametrische Verfahren -> k Stichproben ??)
 1-faktorielle nichtparametrische Analyse für unabhängige Stichproben (Kruskal-Wallis-Test) bzw.
 1-faktorielle nichtparametrische Analyse für abhängige Stichproben (Friedman-Test).

3.3 Fehler bei der Rangberechnung

Gelegentlich werden die Ränge mit der Funktion `rank` sowohl in R als auch in SPSS falsch berechnet. Das hört sich schlimm an, hat aber einen einfachen Grund: Rundungsfehler. Solche Fehler treten natürlich nicht auf, wenn die eingelesenen Variablen in Ränge umgerechnet werden, sondern nur dann, wenn abgeleitete statistische Variablen, wie z.B. Residuen, oder selbst neu errechnete Variablen, wie z.B. Variablensummen und -mittelwerte, in Ränge transformiert werden. Ein Beispiel soll das illustrieren:

Angenommen, es werden aus einer Reihe von Variablen mit den Werten -1, 0, 1 mehrere Mittelwerte gebildet, die dann zu einem Gesamtscore zusammengefasst werden. Dabei resultieren für zwei Probanden die folgenden Teilmittelwerte $1/3$ und $-1/3$ sowie $2/3$ bzw. $-2/3$, die natürlich nicht als Bruch sondern als Dezimalzahl gespeichert werden:

```
1: 0,6666667 - 0.3333333 - 0.3333333
2: - 0,6666667 + 0.3333333 + 0.3333333
```

Werden jetzt jeweils die Summen aus den drei Teilmittelwerten gebildet, erhält man:

```
1: 0.0000001
2: -0.0000001
```

Beide Summen müssten natürlich „theoretisch“ Null sein. Beim „normalen“ Rechnen macht diese Differenz von 0.0000001 , die durch Rundungsfehler entsteht, nichts aus, da sie verschwindend klein ist. Anders jedoch, wenn diese Summe in Ränge transformiert wird. Für die beiden o.a. Probanden sind die Summen nicht mehr gleich und erhalten dadurch verschiedene Ränge. Konkret wird dieses Problem häufiger bei dem *aligned rank transform*-Tests auftreten (vgl. Kapitel 4.3.2.3), da dort von Residuen Mittelwerte subtrahiert und das Ergebnis in Ränge umgerechnet werden.

In R lässt sich dieses Problem lösen: Dort gibt es die Funktion `round(x, digits=...)`, über die ein Vektor `x` auf die vorgegebene Anzahl von Dezimalstellen gerundet werden kann. In der Regel sollte ein Wert `digits=6` ausreichend sein. `round` muss dann vor der Rangberechnung auf die zu transformierende Variable angewandt werden. Würde man diese Funktion auf die Summe des o.a. Beispiels anwenden, so wären die Summen für beide Probanden Null.

3.4 Fehlende Werte

Fehlende Werte (*missing values*), insbesondere der abhängigen Variablen (*Kriterium*), sollten i.a. keine Probleme bereiten, sondern automatisch statistisch sinnvoll von den Programmen behandelt werden. Das funktioniert auch weitgehend so. Allerdings ist dabei zu bedenken, dass bei Messwiederholungen, zumindest bei den hier behandelten Standardmethoden, keine fehlenden Werte auftreten dürfen.

Bei der Benutzung von R empfiehlt es sich, im Fall von fehlenden Werten generell vor Durchführung der Varianzanalysen mit der Funktion `na.omit(...)` eine Teildatenmatrix der in der Analyse verwendeten Variablen (Faktoren und Kriterium) ohne fehlende Werte zu erzeugen. Dies ist ganz besonders in den folgenden Fällen ratsam:

- Die Funktion `ezANOVA` kann nicht mit fehlenden Werten umgehen, auch nicht bei Designs, die keine Messwiederholungen enthalten. Hier empfiehlt sich immer:

```
ezANOVA(na.omit(Dataframe), ...)
```

- Im Fall von fehlenden Werten bei Messwiederholungen müssen in jedem Fall (sowohl bei der Analyse mittels `aov` als auch mittels `ezANOVA`) vor der Umstrukturierung der Daten mittels `reshape` oder `make.rm` entsprechende Fälle (Versuchspersonen) komplett eliminiert werden.
- Bei den nichtparametrischen Analysen ist fast immer eine Rangtransformation erforderlich. Bei der Rangbildung mittels `rank(...)` erhalten standardmäßig (unsinnigerweise) auch fehlende Werte Ränge, nämlich die höchsten Ränge. Mittels des Parameters

```
rank(..., na.last="keep")
```

kann das vermieden werden.

4. Unabhängige Stichproben

Es wird im Folgenden angenommen, dass die Werte einer abhängigen Variablen x für K Gruppen mit Stichprobenumfängen n_i vorliegen. Üblicherweise werden die Gruppen, und damit die Stichproben, über eine Variable, die Gruppierungsvariable definiert. Diese wird i.a. *Gruppierungsfaktor* genannt, im Gegensatz zu den Messwiederholungsfaktoren. Bei mehrfaktoriellen Analysen entsprechend über mehrere Gruppierungsvariablen.

Beispieldaten 1 (mydata1):

Im Folgenden wird ein Datensatz verwendet, bei dem 2 Patientengruppen (Faktor A: Schizophrene und Depressive, je 9 Personen) jeweils in 3 Gruppen zu 3 Personen eingeteilt werden, die dann jeweils ein Medikament (Faktor B: drug 1, 2 oder 3) erhalten. Alle Zellen haben daher dieselbe Anzahl Versuchspersonen ($n=3$). Die abhängige Variable ist eine Beurteilung auf einer Skala von 0 bis 19, also quasi metrisch, wenn auch streng genommen als Beurteilung ordinal.

patients	drug 1	drug 2	drug 3
Schizophrene	8 4 0	10 8 6	8 6 4
Depressive	16 12 8	6 4 2	17 14 11

In R wie auch in SPSS werden hierfür die Variablennamen `patients`, `drug` und `x` verwendet. In R müssen `patients` und `drug` vom Typ „factor“ deklariert sein. In R hat der Dataframe den Namen `mydata1`.

Beispieldaten 2 (mydata2):

Im Weiteren wird ein Datensatz verwendet, bei dem 2 Patientengruppen (Faktor A: Kontrollgruppe und Behandlungsgruppe) jeweils in 4 Gruppen eingeteilt werden, die dann jeweils ein Medikament (Faktor B: drug 1, 2, 3 oder 4) erhalten. Die Zellenbestutzungszahlen sind in diesem Datensatz ungleich. Die abhängige Variable ist eine Beurteilung auf einer Skala von 1 bis 9, also ordinal.

group	drug 1	drug 2	drug 3	drug 4
Kontrolle	4 5 5 6	5 6 6 7 7	5 6 7 7	5 6 6 7 9
Behandlung	2 3 3	3 3 4 5	3 4 5 8	6 7 9 9

In R wie auch in SPSS werden hierfür die Variablennamen `group`, `drug` und `x` verwendet. In R müssen `group` und `drug` vom Typ „factor“ deklariert sein. In R hat der Dataframe den Namen `mydata2`.

Beispieldaten 3 (mydata3):

Darüber hinaus wird ein Datensatz verwendet, bei dem wieder 2 Patientengruppen (Faktor A: Kontrollgruppe und Behandlungsgruppe) jeweils in 4 Gruppen eingeteilt werden, die dann jeweils ein Medikament in 4 verschiedenen hohen Dosierungen (Faktor B: dosis 1, 2, 3 oder 4) erhalten. Die Zellenbestutzungszahlen sind in diesem Datensatz ungleich. Die abhängige Variable ist eine Beurteilung der Reaktion auf einer Skala von 1 bis 20. Durch Abbruch der Therapie kommt es hier zu unterschiedlichen n_i . Das Skalenniveau ist dasselbe wie im ersten Beispiel, also quasi metrisch, wenn auch streng genommen als Beurteilung ordinal.

gruppe	dosis 1	dosis 2	dosis 3	dosis 4
Kontrolle	4 5 7	5 6 7 6 7 8	4 6 8 9	5 6 7 9 10
Behandlung	4 5 6	6 6 7 7	5 7 11 12	5 9 11 14

In R wie auch in SPSS werden hierfür die Variablennamen `gruppe`, `dosis` und `x` verwendet. In R müssen `gruppe` und `dosis` vom Typ „factor“ deklariert sein. In R hat der Dataframe den Namen `mydata3`.

4.1 Voraussetzungen der parametrischen Varianzanalyse

Vom t-Test her kennt man zwei Voraussetzungen: Erstens müssen die Beobachtungen der abhängigen Variablen x in beiden Gruppen normalverteilt sein und zweitens müssen die Varianzen beider Gruppen homogen (statistisch gleich) sein. Dies lässt sich noch problemlos von zwei auf beliebig viele K Gruppen verallgemeinern. (Mit K wird im Folgenden die Anzahl von Stufen/Gruppen eines unspezifizierten Faktors bezeichnet.) Doch insbesondere die Normalverteilungsvoraussetzung kann auch anders formuliert werden: Die Residuen e_{ij} müssen normalverteilt sein, wobei sich die Residuen aus dem varianzanalytischen Modell ergeben, hier für den 1-faktoriellen Fall eines Faktors A mit I Stufen/Gruppen:

$$x_{im} = \mu + \alpha_i + e_{im} \quad (i=1, \dots, I \text{ und } m=1, \dots, n_i) \quad (4-1)$$

wobei $\alpha_i = \mu_i - \mu$ die Abweichungen des Gruppenmittelwertes vom Gesamtmittel sind, der *Effekt* von Faktor A mit I Stufen (Gruppen). Das Modell der 2- oder mehrfaktoriellen Analyse unterscheidet sich kaum von dem 1-faktoriellen, da diese auch nur eine einzige Residuenvariable e_{ijm} enthält. Dabei sei B der zweite Faktor, mit J Stufen (Gruppen) sowie den Effekten β_j :

$$x_{ijm} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijm} \quad (i=1, \dots, I, j=1, \dots, J \text{ und } m=1, \dots, n_{ij}) \quad (4-2)$$

(Auf die Interaktion $\alpha\beta_{ij}$ wird in Kapitel 4.3.1.2 kurz eingegangen.) Logisch sind zwar beide Bedingungen (Normalität innerhalb jeder Gruppe und Normalität der Residuen) identisch, doch in der Praxis ist es sinnvoll, die Gesamtheit der Residuen zu überprüfen. Weitere Erläuterungen zur Prüfung auf Normalverteilung sind in Kapitel 1.6 zu finden.

Die Varianzhomogenität ($\sigma_1^2 = \dots = \sigma_K^2$) wird zweckmäßigerweise mit dem *Levene-Test* überprüft, da dieser (im Gegensatz zum klassischen F-Test, zum F_{\max} -Test oder zum Bartlett-Test) robust gegen Abweichungen von der Normalverteilung ist. Einen kompletten Vergleich von 14 Tests zur Prüfung der Varianzhomogenität bietet Shuqiang Zhang [67].

Doch was, wenn eine der Voraussetzungen nicht erfüllt ist? Muss dann direkt zur nichtparametrischen Varianzanalyse gegriffen werden? Nein! Die Varianzanalyse ist ein sehr robustes statistisches Testverfahren (vgl. Kapitel 1.1). Hierzu gibt es zahlreiche Untersuchungen, insbesondere solche, die das Verhalten von β (Wahrscheinlichkeit für einen Fehler 2. Art) zum Inhalt haben. Brauchbare Übersichten findet man u.a. bei Field [63], Bortz [62] und Ito [61].

Zunächst einmal zwei generelle positive Aussagen:

- Je größer die Stichproben, desto weniger sind die Voraussetzungen noch relevant. Insbesondere ist nach dem zentralen Grenzwertsatz die Normalverteilungsvoraussetzung nur für kleinere Stichproben ($n_i < 50$) relevant.
- Bei annähernd gleichgroßen Stichprobenumfängen n_i wirken sich weder nichtnormalverteilte Residuen noch inhomogene Varianzen störend aus.

Zur Voraussetzung der Normalverteilung

(Details sind bei Wilcox [64], Osborne [70] sowie Lindman [69] nachzulesen):

- Moderate Abweichungen von der Normalverteilung, z.B. eine Schiefe, führen schlimmstenfalls zu einer leichten Vergrößerung von β . D.h. gegebenenfalls können Unterschiede nicht nachgewiesen werden. Oder positiv ausgedrückt: Signifikante Unterschiede können als gesichert gelten.
- Schmalgipflige, steile Verteilungen, d.h. mit negativem Exzess ([http://de.wikipedia.org/wiki/Wölbung_\(Statistik\)](http://de.wikipedia.org/wiki/Wölbung_(Statistik))), machen den F-Test konservativer. Breitgipflige Verteilungen machen dagegen den Test liberaler, können aber auch das α -Risiko vergrößern, allerdings nur in einem sehr geringen Maß (vgl. [61]).
- Drastische Abweichungen von der Normalverteilung können zu unbrauchbaren Ergebnissen führen, insbesondere wenn die Stichprobenumfänge n_i verschieden sind. (Der F-Test kann in solchen Fällen sowohl zu liberal als auch zu konservativ reagieren).
- Box & Andersen [74] haben einen F-Test entwickelt, der die Abweichung von der Normalverteilung durch eine Korrektur der Freiheitsgrade kompensiert (vgl. Anhang 2.3). Eine entsprechende R-Funktion ist im Anhang 3 zu finden.

Zur Voraussetzung der Varianzhomogenität:

- Moderate Abweichungen von der Varianzhomogenität führen ebenfalls schlimmstenfalls zu einer leichten Vergrößerung von β . Allerdings gilt auch hier, dass die Stichprobenumfänge n_i nicht zu stark divergieren dürfen.
- Bei ungleichen n_i gilt: Haben die großen Stichproben auch die größeren Varianzen, reagiert der F-Test konservativ. Haben dagegen die großen Stichproben die kleineren Varianzen, reagiert der F-Test liberal. Vgl. [75].
- Korrelieren im Falle inhomogener Varianzen die Zellenmittelwerte mit den -varianzen, nehmen also mit steigenden Zellenmittelwerten auch die Zellvarianzen zu, wird eine Datentransformation der Kriteriumsvariablen x empfohlen: gute Chancen bieten die einfachen Funktionen \sqrt{x} und $\log(x)$. Die Box-Cox-Transformationen [73] perfektionieren diese Idee.
- Es gibt Tests von Welch sowie von Brown & Forsythe (Details im Anhang 2.2), die keine Varianzhomogenität voraussetzen, allerdings auf Kosten der Teststärke. Von diesen ist der von Brown & Forsythe vorzuziehen (vgl. Jennifer J. Clinch et al. [68]). Diese sind für 1-faktorielle Varianzanalysen auch in R und SPSS enthalten. Sie lassen sich jedoch auch auf zwei Faktoren verallgemeinern.
- Darüber hinaus hat Box [72] eine Korrektur (genauer gesagt: Reduzierung) der Freiheitsgrade für den F-Test entwickelt, der die Heterogenität der Varianzen berücksichtigt. Diese erfordert zwar ein wenig Programmieraufwand, ist aber in R realisierbar. Näheres dazu in Winer [13], S. 109, sowie im Anhang 2.1.
- Für mehrfaktorielle Varianzanalysen wird manchmal empfohlen, den ART anstatt der „normalen“ Analyse zu verwenden (vgl. Mehmet Menduş et al. [66]). Allerdings gibt es dagegen auch Vorbehalte (vgl. Kapitel 2.12).

Details sind bei Glass et al. [65] sowie Osborne [70] nachzulesen. Eine gute Übersicht, insbesondere der robusten parametrischen Verfahren, ist bei Fan [75] zu finden. Dort wird auch ausdrücklich darauf hingewiesen, dass im Falle von Rangtransformationen vielfach Varianzhomogenitäten erhalten bleiben, was auch Beasley in seinen Untersuchungen [84] bestätigt.

Fazit: Ist die abhängige Variable metrisch, die Stichprobenumfänge n_i nicht zu unterschiedlich (etwa Faktor 3), die Abweichungen von der Normalverteilung der Residuen wie auch von der Varianzhomogenität moderat, so kann die parametrische Varianzanalyse durchgeführt und die Ergebnisse ohne Einschränkung interpretiert werden. Für alle anderen Fälle, insbesondere ordinale abhängige Variablen mit einem kleinen Wertebereich, ist eine nichtparametrische Analyse erste Wahl.

Neben den beiden o.a. Voraussetzungen gibt es allerdings noch eine dritte: die Unabhängigkeit der Beobachtungen. Diese lässt sich allerdings kaum „testen“, sondern setzt eher eine saubere Versuchsplanung voraus. Dies ist allerdings nicht Thema dieses Skripts.

Beispiele zur Prüfung der Voraussetzungen in R bzw. SPSS werden in den nachfolgenden Kapiteln, u.a. 4.3.2 vorgestellt.

4.2 Die 1-faktorielle Varianzanalyse

Getestet wird die Hypothese gleicher Gruppenmittelwerte:

$$\mu_1 = \mu_2 = \dots = \mu_K$$

was in der Terminologie des o.a. Modells 4-1 äquivalent ist zu:

$$\alpha_1 = \alpha_2 = \dots = \alpha_K = 0$$

4.2.1 Kruskal-Wallis-Test

Eine 1-faktorielle nichtparametrische Varianzanalyse erfolgt üblicherweise über den *Kruskal-Wallis-H-Test*, einer Verallgemeinerung des *Mann-Whitney-U-Tests* von zwei auf beliebig viele Gruppen. Die Logik sieht so aus, dass alle Werte in Ränge transformiert werden, so dass letztlich anstatt der Mittelwerte die mittlere Rangsummen verglichen werden. Für den Test wird ein Wert H errechnet, der χ^2 -verteilt ist mit $(K-1)$ Freiheitsgraden.

Derselbe Test lässt sich auch über eine 1-faktorielle klassische Varianzanalyse der Ränge der abhängigen Variablen durchführen. Dies wird in Abschnitt 4.3.5 ausführlich beschrieben.

mit R:

Sollen für den o.a. Datensatz 1 die Reaktionen bzgl. der 3 Medikamente (Faktor `drugs`) verglichen werden, lautet die Anweisung:

```
mydata1 <- within(mydata1, drug<-factor(drug))
kruskal.test(x, drug)
```

mit der Ausgabe

```
Kruskal-Wallis rank sum test

data:  x and drug
Kruskal-Wallis chi-squared = 2.023, df = 2, p-value = 0.3637
```

was zunächst einmal indiziert, dass die Reaktionen auf die 3 Medikamente sich nicht signifikant unterscheiden.

mit SPSS:

Hier muss beachtet werden, dass gegebenenfalls vorher das Skalenniveau der analysierten Variablen auf „Skala“ gesetzt wird.

```
Nptests
  /independent test (x) group (drug) kruskal_wallis (compare=pairwise).
```

mit folgender Ausgabe:

	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von ist über Kategorien von gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,364	Nullhypothese behalten.

Asymptotische Signifikanzniveaus werden angezeigt. Das Signifikanzniveau ist ,05.

Gesamtanzahl	18
Teststatistik	2,023
Freiheitsgrade	2
Asymptotische Sig. (zweiseitiger Test)	,364

4. 2. 2 Varianzanalysen für inhomogene Varianzen

Varianzhomogenität ist ja eine der Voraussetzungen für die „normale“ parametrische Varianzanalyse. Man muss aber im Falle von Inhomogenitäten nicht direkt zur nichtparametrischen Analyse greifen, denn es gibt bzgl. der Varianzhomogenität robuste Varianten der parametrischen Analyse. Zumal durch die meistens angewandten Rangtransformationen sich Streuungsunterschiede nicht notwendigerweise auflösen. Die bekanntesten sind die Tests von Welch bzw. von Brown & Forsythe, wovon letzterer der neuere und bessere ist. Allerdings sollte er nicht mit dem gleichnamigen Test zur Prüfung der Varianzhomogenität verwechselt werden. Trivialerweise dürfen diese Tests natürlich auch angewandt werden, wenn die Varianzen homogen sind. Im Falle von exakt gleichen Varianzen sind die F-Werte dieser Tests mit dem „normalen“ F-Test identisch, so dass es durchaus angebracht ist, diese Tests immer als 1-faktorielle Varianzanalyse zu benutzen. Beide Tests sind in SPSS verfügbar, während in R nur der Welch-Test standardmäßig aufrufbar ist. Anzumerken ist noch, dass die nichtganzzahligen Freiheitsgrade typisch für solche Tests sind, die keine Varianzhomogenität voraussetzen.

An dieser Stelle sollte auch der Test von Brunner, Dette und Munk, auch *BDM-Test* genannt, erwähnt werden. Eigentlich ist er ein nichtparametrischer Test und als Alternative zum Kruskal-Wallis-Test für den Fall stark inhomogener Varianzen gedacht. Aber er empfiehlt sich auch für den Fall normalverteilter Residuen. Ein Beispiel dazu folgt in Kapitel 4.3.3.

Für das nachfolgende Beispiel wird der Beispieldatensatz 3 benutzt und dort einfaktoriellement der Faktor `dosis` untersucht.

mit R:

Zunächst die Prüfung der Varianzhomogenität mittels des Levene-Tests:

```
leveneTest(x~dosis, center=mean, data=mydata3)
```

```
Levene's Test for Homogeneity of Variance (center = mean)
  Df F value Pr(>F)
group 3  4.9647 0.006675 **
  29
```

Infolge der stark signifikanten Inhomogenität ist anstatt des normalen F-Tests ein dafür geeigneter robuster F-Test zu wählen. Der Welch-Test ist durchführbar über die Funktion `oneway.test`. Für die Variable `x` aus dem Beispieldatensatz 3 mit dem Faktor `dosis` lautet die Anweisung:

```
oneway.test(x~doss, mydata3)
```

```
One-way analysis of means (not assuming equal variances)

data:  x and dosis
F = 3.8789, num df = 3.000, denom df = 13.308, p-value = 0.0433
```

Der p-Wert mit 0,043 belegt, dass die Dosis eine Wirkung zeigt.

Eine Funktion für den Brown-Forsythe-Test befindet sich im Anhang 2. Eine Anwendung wird in 4.3.3 gezeigt.

mit SPSS:

Beide Tests sind durchführbar über `Oneway` (Menü: Mittelwerte vergleichen -> Einfaktorielle Anova). Allerdings müssen diese robusten Tests über die „Optionen“ angefordert werden. Für die Variable `x` aus dem Beispieldatensatz 3 mit dem Faktor `dosis` lautet die Syntax:

```
Oneway x by dosis
  /statistics homogeneity brownforsythe welch.
```

In der Ausgabe erscheint nach dem Test auf Homogenität der Varianzen zunächst das Ergebnis für homogene Varianzen:

Test der Homogenität der Varianzen			
Levene-Statistik	df1	df2	Signifikanz
4,965	3	29	,007

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	45,672	3	15,224	3,130	,041
Innerhalb der Gruppen	141,056	29	4,864		
Gesamt	186,727	32			

danach die Tests für beliebige Varianzen, die hier sogar eine größere Signifikanz als der „normale“ F-Test zeigen, was häufig vorkommt, wenn Voraussetzungen des „normalen“ Tests nicht erfüllt sind.

Robuste Testverfahren zur Prüfung auf Gleichheit der Mittelwerte				
	Statistik ^a	df1	df2	Sig.
Welch-Test	3,879	3	13,308	,034
Brown-Forsythe	3,218	3	18,618	,047

4. 2. 3 Verfahren für nichtnormalverteilte Variablen

Wegen der großen Robustheit der Varianzanalyse hinsichtlich Abweichungen der Residuen von der Normalverteilung gibt es nur wenige Verfahren für metrische nichtnormalverteilte abhängige Variablen. Auf zwei soll hier kurz eingegangen werden.

Zum einen gibt es einen modifizierten F-Test von Box & Andersen [74] (vgl. auch Anhang 2.3), bei dem sich die Abweichung von der Normalverteilung in der Korrektur der Freiheitsgrade widerspiegelt, wie dies „üblicherweise“ auch bei den entsprechenden modifizierten F-Tests für heterogene Varianzen der Fall ist. Dieses Verfahren macht z.B. Sinn bei extrem schiefen Verteilungen. Eine entsprechende R-Funktion ist im Anhang 3 zu finden.

Erceg-Hurn & Mirosevich [5] erinnern an die Methode der *Winsorisierung* (oder auch *Trimmen*), die relativ selten angewandt wird, weil sie den Verdacht der Datenmanipulation aufkommen lässt, die aber statistisch durchaus sinnvoll ist. Hierbei werden ein fester Prozentsatz der größten und kleinsten Werte einer Variablen durch die nächstkleinere bzw. durch die nächstgrößere ersetzt. Häufig ersetzt man jeweils 5% der Werte, bei kleineren Stichproben auch jeweils 10% , am oberen Ende durch den nächstkleineren Wert sowie 5% bzw. 10% der Werte am unteren Ende. Dieses Verfahren ist sinnvoll insbesondere beim Vorliegen von Ausreißern.

4. 2. 4 Weitere Verfahren

Die nachfolgend für die 2-faktorielle Varianzanalyse beschriebenen Rank transform Tests (RT), normal scores-Test (INT) und von der Waerden-Tests sind ebenso als 1-faktorielle Analyse einsetzbar. Dagegen macht das ART-Verfahren nur im mehrfaktoriellen Design Sinn. Die ATS von Akritis & Co ist als 1-faktorielle Analyse nicht bekannt.

4. 3 Die 2-faktorielle Varianzanalyse

Bevor die einzelnen Methoden, von der parametrischen Analyse inklusive Prüfung der Voraussetzungen bis zu den verschiedenen nichtparametrischen Methoden, im Detail besprochen werden, sollen zunächst noch ein paar grundlegende Eigenschaften der mehrfaktoriellen Varianzanalyse erwähnt werden. Leser, die schon Erfahrungen auf dem Gebiet der Anova haben, werden damit schon vertraut sein.

4. 3. 1 Anmerkungen zur 2-faktoriellen Varianzanalyse

4. 3. 1. 1 Balancierte und nichtbalancierte Versuchspläne

Man unterscheidet zwischen *balancierten* (engl. *balanced*) und *nichtbalancierten* (engl. *unbalanced*) Versuchsplänen bzw. Zellenbesetzungszahlen. Bei balancierten Versuchsplänen sind die Zellenbestetzungszahlen zeilenweise oder spaltenweise proportional zueinander, z.B. bei einem Versuchsplan mit den Faktoren A (4 Stufen) und B (3 Stufen)

	B ₁	B ₂	B ₃
A ₁	10	12	16
A ₂	15	18	24
A ₃	20	24	32
A ₄	10	12	16

In diesem Beispiel sind die Zellenbesetzungszahlen der 2. bzw. 3. Spalte das 1,2-fache bzw 1,6-fache der 1. Spalte. Umgekehrt kann man auch erkennen, dass die Zellenbesetzungszahlen der 2. bzw. 3. Zeile das 1,5-fache bzw. das 2-fache der ersten Zeile sind.

Versuchspläne mit gleichen Zellenbesetzungszahlen sind natürlich immer balanciert. Solche, bei denen die o.a. Proportionalität nicht zutrifft, sind nichtbalanciert.

Diese Unterscheidung ist insofern relevant, als dass die Lösung für die 2- und mehrfaktorielle Varianzanalyse, d.h. die Berechnung der durch die einzelnen Faktoren bzw. Effekte erklärten Streuungen, bei nichtbalancierten Versuchsplänen nicht mehr eindeutig ist. Es gibt mehrere Schätzmethoden: Typ I, Typ II und Typ III, auf die hier nicht näher eingegangen werden soll. Von diesen ist die *Resgressionsmethode der kleinsten Quadrate* (LS), auch mit *Schätzungen vom Typ III* bezeichnet, die gebräuchlichste und unproblematischste.

4. 3. 1. 2 Die Interaktion

Soll der Einfluss zweier Einflussfaktoren A und B auf eine abhängige Variable x untersucht werden, so bringen zwei 1-faktorielle Varianzanalysen der Faktoren A und B nur die halbe Wahrheit hervor, mitunter sogar irreführende Ergebnisse. Neben den sog. *Haupteffekten* der Faktoren A und B, dem Einfluss von A bzw. B ohne Berücksichtigung des jeweils anderen Faktors, gibt es einen sog. *Interaktionseffekt* A*B, auch *Wechselwirkung* genannt. Dieser zeigt an, ob der Einfluss von A von B abhängig ist, und umgekehrt, ob der Einfluss von B von A abhängig ist. So kann es durchaus vorkommen, dass die Haupteffekte A und B nicht signifikant sind, dafür aber A*B. Dies besagt, dass ein Einfluss von A vorhanden ist, der je nach Gruppe (Stufe) des Faktors B unterschiedlich ausfällt, und umgekehrt, dass ein Einfluss von B vorhanden ist, der je nach Gruppe (Stufe) des Faktors A unterschiedlich ausfällt. In der Praxis heißt das, dass häufig der Einfluss eines Faktors erst dadurch zu Tage tritt, dass dieser in Zusammenhang mit einer anderen Einflussgröße analysiert wird.

Im mathematischen Modell für die 2-faktorielle Varianzanalyse

$$x_{ijm} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijm} \quad (4-3)$$

erscheint die Interaktion $\alpha\beta_{ij}$ als eine weitere erklärende Komponente von x, neben den Anteilen α_i , den durch Faktor A erklärten Abweichungen ($\mu - \mu_{Ai}$), sowie den β_j , den durch Faktor B erklärten Abweichungen ($\mu - \mu_{Bj}$). Während die Haupteffekte für A und B die Hypothesen

$$H_A: \alpha_i = 0 \text{ für } i=1,\dots,I \text{ (entspricht } \mu_{A1} = \mu_{A2} = \dots = \mu_{AI} \text{)}$$

$$H_B: \beta_j = 0 \text{ für } j=1,\dots,J \text{ (entspricht } \mu_{B1} = \mu_{B2} = \dots = \mu_{BJ} \text{)}$$

testen, wird über die Interaktion A*B die folgende Hypothese geprüft:

$$H_{AB}: \alpha\beta_{ij} = 0 \text{ für } i=1,\dots,I \text{ und } j=1,\dots,J$$

d.h. sowohl die durch A erklärten Abweichungen α_i sind für alle Stufen von B gleich groß als auch die durch B erklärten Abweichungen β_j sind für alle Stufen von A gleich groß.

Dies lässt sich grafisch durch einen sog. *Interaktionsplot* (in SPSS *Profilplot* genannt) veranschaulichen. Dort werden Mittelwertlinien des Faktors A getrennt für die Stufen des Faktors B gezeichnet. Ein nicht paralleler Verlauf der Kurven deutet auf eine signifikante Interaktion hin. Dies kann zum einen sein: Der Einfluss von A ist unterschiedlich stark für die Gruppen von B, oder der Einfluss von A ist für die Gruppen von B gegensätzlich. Bei der 2-faktoriellen Varianzanalyse lassen sich zwei solcher Plots erstellen: einmal erscheinen die Stufen von A auf der x-Achse und die Stufen von B als verschiedene Linien und einmal erscheinen die Stufen von B auf der x-Achse und die Stufen von A als Linien. Welches nun der aussagekräftigere Plot ist, muss individuell entschieden werden.

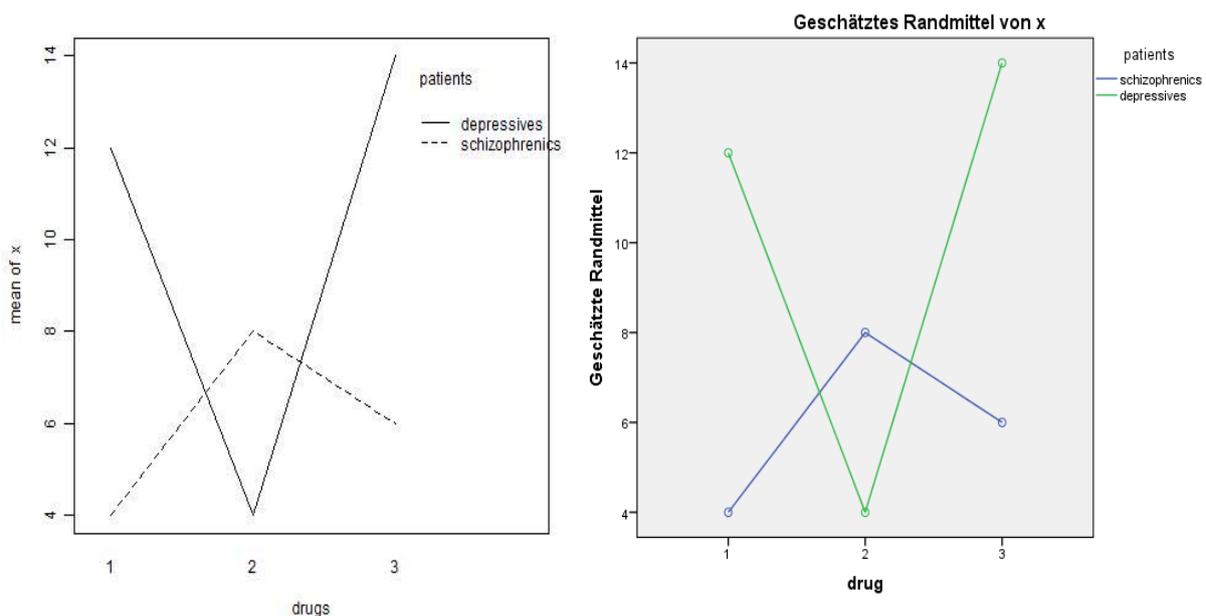
mit R

```
interaction.plot (factor1, factor2, x)
```

wobei die Variablen *factor1*, *factor2* vom Typ „factor“ sein müssen.

mit SPSS

In SPSS ist der Interaktionsplot erhältlich über die parametrische Varianzanalyse (Analysieren -> Allg. lineare Modell -> univariat -> Diagramme)



Interaktionsplot für den o.a. Datensatz: links mit R erstellt, rechts mit SPSS

4. 3. 1. 3 Reduzierung des statistischen Fehlers

Die folgenden Ausführungen gelten in erster Linie für die parametrische Varianzanalyse sowie für die „anderen“ Verfahren, bei denen die klassische Aufspaltung der Gesamtstreuung in Effekt- und Residuenstreuung vorgenommen wird. Das sind neben den robusten Verfahren für heterogene Varianzen in erster Linie die oben erwähnten Rank transform Tests (RT, INT und ART). Ferner gilt das Folgende ausschließlich für Versuchspläne mit Gruppierungsfaktoren und bei gemischten Versuchsplänen für die Messwiederholungsfaktoren.

Neben der Analyse der Wechselwirkung bringt die 2-faktorielle Analyse einen weiteren Gewinn gegenüber zwei 1-faktoriellen Analysen: Durch die Hinzunahme eines weiteren Einflussfaktors kann ein weiterer Anteil der Streuung von x erklärt werden. Die statistischen Tests der Faktoren erfolgen über F-Tests mit einem F-Wert, bei dem im Nenner die Residuenstreuung, die Reststreuung, erscheint. Wird letztere nun reduziert, vergrößert sich der F-Wert und damit verkleinert sich der daraus errechnete p-Wert, was eine höhere Signifikanz bedeutet.

Ausnahme: Falls ein hinzugenommener Faktors keinen Einfluss hat, auch nicht über die Interaktion, und keine zusätzliche Streuung erklärt, sollte dieser weggelassen werden. Denn der Haupteffekt sowie die Interaktion des hinzugenommenen Faktors beanspruchen Freiheitsgrade, die von denen der Residuenstreuung abgezogen werden. Und dadurch fallen die Tests für die anderen Effekte schlechter aus. Ob ein Faktor nun Teil eines Anova-Modells sein sollte oder nicht, muss der Untersuchende aufgrund der vorliegenden Hypothesen entscheiden.

Was hier für die Interaktion der 2-faktoriellen Varianzanalyse gesagt wurde, gilt analog für höhere Interaktionen bei der 3- und mehrfaktoriellen Analyse. Mit einem Unterschied: 3-fach und höhere Interaktionen sind zum einen sehr schwer zu interpretieren, sind aber (zum Glück) in der Praxis selten signifikant. Daher werden diese in der Regel nicht in die Modelle einbezogen.

4.3.2 Das parametrische Verfahren und Prüfung der Voraussetzungen

Zum Vergleich seien die Ergebnisse für die parametrische Analyse vorangestellt sowie die Tests auf Normalverteilung und Homogenität der Varianzen, und zwar zunächst für die Beispieldaten 1 mit einem balancierten Versuchsplan. Anschließend folgt die Analyse für die Beispieldaten 2 mit einem unbalancierten Design, allerdings ohne Wiedergabe der Tests der Voraussetzungen, die aber erfüllt sind:

mit R:

Da hier ein balancierter Versuchsplan ausgewertet wird, kann sowohl die `options-` als auch die `drop1-`Anweisung entfallen.

```
mydata1 <- within(mydata1, {drugs<-factor(drugs);
                        patients<-factor(patients)})
aov1 <- aov(x~patients*drugs, mydata)
summary(aov1)
```

erhält man als Ergebnis

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
patients	1	72	72.00	8.151	0.01449	*
drugs	2	48	24.00	2.717	0.10634	
patients:drugs	2	144	72.00	8.151	0.00581	**
Residuals	12	106	8.83			

Tabelle 4-1

Zur Prüfung der Normalverteilung der Residuen können diese aus dem Anova-Ergebnis über `aov1$residuals` gewonnen werden. Der Shapiro-Wilk-Test und der Levene-Test zur Prüfung der Homogenität der Varianzen können über folgende Anweisungen erfolgen:

```
library(car)
shapiro.test(aov1$residuals)
leveneTest(x~patients*drugs, data=mydata1)
```

mit folgender Ausgabe:

```
Shapiro-Wilk normality test

data:  aov1$residuals
W = 0.9372, p-value = 0.2592

Levene's Test for Homogeneity of Variance (center = mean)
  Df F value Pr(>F)
group 5  0.4377  0.814
  12
```

Hieraus ist zu schließen, dass sowohl die Voraussetzung der normalverteilten Residuen als auch der Homogenität der Varianzen erfüllt ist. Eine nichtparametrische Analyse ist daher nicht erforderlich.

Nachfolgend nun die Varianzanalyse für die Beispieldaten 2. Da es sich dabei nicht um einen balancierten Versuchsplan handelt, weichen die erforderlichen Kommandos von den oben aufgeführten etwas ab.

```
options (contrasts=c("contr.sum", "contr.poly"))
mydata2 <- within(mydata2, {drugs<-factor(drugs);
                        group<-factor(group) })
aov2 <- aov(x~group*drugs, mydata2)
drop1(aov2, ~. , test="F")
```

mit dem Ergebnis:

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)	
<none>			40.917	23.096			
group	1	12.024	52.940	29.598	7.3464	0.0119656	*
drugs	3	46.560	87.477	42.171	9.4827	0.0002319	***
group:drugs	3	17.932	58.848	29.089	3.6521	0.0260399	*

Tabelle 4-2

mit SPSS:

Die Prüfung der Voraussetzungen, d.h. die Analyse der Residuen sowie der Varianzhomogenität, sollte schon bei der Durchführung der Varianzanalyse berücksichtigt werden, indem sowohl unter „Speichern“ die Residuen (z.B. „standardisiert“) als zusätzliche Variable angefordert werden und unter „Optionen“ der Homogenitätstest angefordert wird. Die Syntax dafür:

```
Unianova  x by patients drug
  /save = zresid
  /print = homogeneity
  /design = patients drug patients*drug.
```

Die daraus erzeugte Varianzanalysetabelle:

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Korrigiertes Modell	264,000 ^a	5	52,800	5,977	,005
Konstanter Term	1152,000	1	1152,000	130,415	,000
patients	72,000	1	72,000	8,151	,014
drug	48,000	2	24,000	2,717	,106
patients * drug	144,000	2	72,000	8,151	,006
Fehler	106,000	12	8,833		
Gesamt	1522,000	18			
Korrigierte Gesamtvariation	370,000	17			

Tabelle 4-3

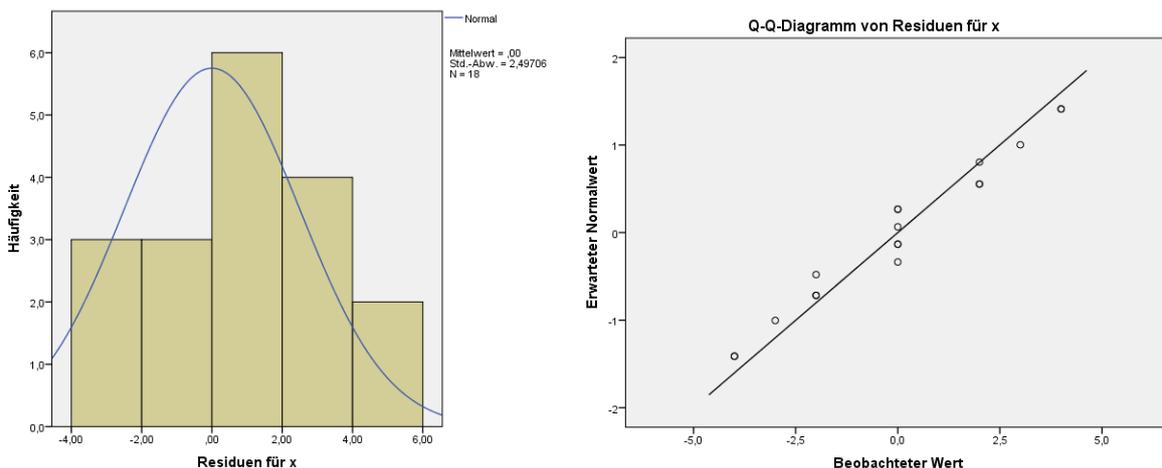
mit der Prüfung der Varianzhomogenität, die hier mit einem p-Wert von 0.814 gesichert ist:

Levene-Test auf Gleichheit der Fehlervarianzen ^a				
F	df1	df2	Sig.	
,438	5	12	,814	

Die Prüfung der Residuen auf Normalverteilung muss anschließend gesondert vorgenommen werden. Z.B. grafisch mittels eines Histogramms der in der Varianzanalyse erzeugten Residuenvariablen (RES_1) oder mittels des Shapiro-Wilks-Tests. Beides zusammen kann man über das Menü „Deskriptive Statistiken -> Explorative Datenanalyse“ erzeugen. Die SPSS-Syntax dazu:

```
Examine variables=RES_1
  /plot histogram npplot.
```

Zur besseren Interpretation des Histogramms sollte allerdings die Intervallzahl auf ca. \sqrt{n} geändert werde, d.h. in diesem Fall bei n=18 auf maximal 5 Intervalle. Der Zusatz npplot führt zu einem *normal probability plot* oder *Q-Q-Diagramm* (vgl. auch Kapitel 1.6) :



Beide zeigen keine deutlichen Abweichungen von der Normalverteilung.

Standardmäßig werden auch zwei Tests auf Normalverteilung ausgegeben: der klassische Kolmogorov-Smirnov- und der etwas modernere Shapiro-Wilk-Test, die hier ebenfalls keine Abweichungen von der Normalverteilung anzeigen:

Tests auf Normalverteilung						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Residuen für x	,167	18	,200*	,937	18	,259

Nachfolgend nun noch die Varianzanalyse für die Beispieldaten 2:

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Korrigiertes Modell	69,265 ^a	7	9,895	6,046	,000
group	12,024	1	12,024	7,346	,012
drugs	46,560	3	15,520	9,483	,000
group * drugs	17,932	3	5,977	3,652	,026
Fehler	40,917	25	1,637		

Tabelle 4-4

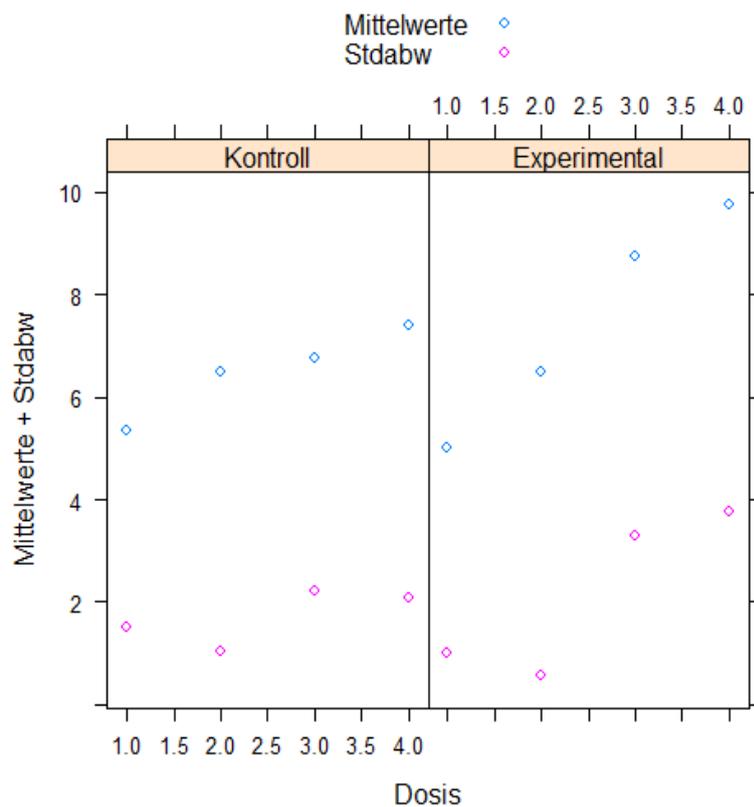
An dieser Stelle soll noch einmal auf die Ausführungen des Kapitels 4.3.1.3 zurückgekommen werden. Dort war darauf hingewiesen worden, dass durch die Hinzunahme eines Faktors häufig der statistische Fehler reduziert werden kann und Effekte erst bei mehrfaktoriellen Analysen als signifikant nachgewiesen werden können. Aus Tabelle 4-3 (Beispieldaten 1) konnten signifikante Effekte für den Faktor *patients* ($p=0,014$) sowie für die Interaktion ($p=0,006$) abgelesen werden. Würde man nur 1-faktorielle Analysen durchführen, so erhielte man keine Signifikanzen, abgesehen davon, dass Interaktionen ohnehin nur mehrfaktoriell erkennbar sind. Hier die Ergebnisse mit SPSS:

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
patients	72,000	1	72,000	3,866	,067
Fehler	298,000	16	18,625		

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
drug	48,000	2	24,000	1,118	,353
Fehler	322,000	15	21,467		

4.3.3 Varianzanalysen für inhomogene Varianzen

Für mehrfaktorielle Versuchspläne gibt es leider kaum robuste F-Tests speziell für heterogene Varianzen. Einzig den in Kapitel 4.1 erwähnten Test von Box (vgl. auch Anhang 2.1), der allerdings sehr konservativ reagiert. Darüber hinaus lässt sich der in Kapitel 4.2.2 vorgestellte Test von Brown & Forsythe (vgl. auch Anhang 2.2) auch auf Interaktionen und damit 2-faktorielle Designs erweitern. Für beide Verfahren ist im Anhang eine R-Funktion zu finden. An dieser Stelle sollte auch der Test von Brunner, Dette und Munk, auch *BDM-Test* genannt, erwähnt werden. Eigentlich ist er ein nichtparametrischer Test und als Alternative zum Kruskal-Wallis-Test für den Fall stark inhomogener Varianzen gedacht. Aber er empfiehlt sich auch für den Fall normalverteilter Residuen.



Alternativ wird verschiedentlich als Abhilfe empfohlen, die Kriteriumsvariable x zu transformieren. Genannt werden die Transformationen \sqrt{x} und $\log(x)$ (vgl. Kapitel 4.1). Allerdings bieten solche Transformationen keine Garantie, dass für die transformierte Variable Varianzhomogenität erreicht wird.

Für den Datensatz 3 zeigt die obige Grafik, dass bei diesem tatsächlich die Varianzen mit den Mittelwerten ansteigen. Der Levene-Test auf Varianzhomogenität zeigt übrigens mit einem $p=0,012$ einen relativ starken Unterschied der Zellvarianzen. Und da zugleich die Zellenbesetzungszahlen stark schwanken, von 3 bis 6, ist hier eine besondere Behandlung erforderlich.

Verschiedentlich werden für den Fall inhomogener Varianzen auch die Rangtransformation empfohlen, also Anwendung des RT-Verfahrens. Wie in Kapitel 2.12 dargelegt, kann diese Methode zum „Erfolg“ führen, muss es aber nicht. Auf ein Beispiel soll an dieser Stelle verzichtet werden, da dieses Verfahren ohnehin in den nachfolgenden Kapiteln ausführlich behandelt wird. Allerdings sei hier erwähnt, dass für den hier benutzten Datensatz 3 die Homogenität der Varianzen durch die Rangtransformation hergestellt werden kann. Nachfolgend die Ergebnisse (p-Werte) des Levene-Tests ohne und mit Rangtransformation:

Effekt	ohne Transformation	mit Rangtransformation	mit normal score-Transf.
gruppe	0.018854	0.39388	0.33077
dosis	0.0066747	0.054783	0.17853
gruppe*dosis	0.011643	0.36508	0.53687

Die beiden o.a. robusten F-Tests sowie der BDM-Test werden mit R gezeigt, während in SPSS Varianzanalysen mit transformierten Daten durchgeführt werden.

4.3.3.1 Verfahren von Box sowie von Brown & Forsythe

mit R

Zunächst einmal werden für den o.a. Datensatz 2-faktorielle Varianzanalysen gerechnet, und zwar mit den oben erwähnten F-Tests von Box sowie von Brown & Forsythe mit Hilfe der im Anhang aufgelisteten Funktion `box.f` bzw. `bf.f`:

```
box <- box.f(x~gruppe*dosis, mydata3)
box$anova
bf <- bf.f(x~gruppe*dosis, mydata3)
bf$anova
```

In der Anova-Tabelle des Box-Tests werden in den Spalten `Eps1` und `Eps2` die Korrekturfaktoren wiedergegeben, mit denen die Zähler- bzw. Nenner-Freiheitsgrade des F-Tests multipliziert werden und dann `Df1` bzw. `Df2` ergeben:

	Eps1	Eps2	Df1	Df2	Sum Sq	Mean Sq	F value	Pr(>F)
gruppe	1.000	0.794	1.00	19.85	9.12	9.116	1.8895	0.1846
dosis	0.708	0.618	2.12	15.45	45.92	15.307	3.1727	0.0677
gruppe:dosis	0.553	0.514	1.66	12.85	11.07	3.691	0.7650	0.4622
Residuals			25.00		120.62	4.825		

In der Anova-Tabelle der Tests von Brown & Forsythe wird neben den Zählerfreiheitsgraden des F-Tests (`Df`) noch die Nenner-Freiheitsgrade (`Df.err`) ausgewiesen:

	Df	Df.err	Sum Sq	Mean Sq	F value	Pr(>F)
gruppe	1	21.326	9.116	9.1162	1.4458	0.24239
dosis	3	18.618	45.922	15.3074	3.2354	0.04581 *
gruppe:dosis	3	12.344	11.072	3.6908	0.3227	0.80890
Residuals	25		120.617	4.8247		

Die Ergebnisse zeigen die Schwäche des Box-Tests gegenüber dem von Brown & Forsythe. Für die Ergebnisse der Varianzanalyse mit der transformierten Variable `x` sei auf den Abschnitt „SPSS“ verwiesen. Im Kapitel 4.3.9 werden alle Ergebnisse für diesen Datensatz, auch die von nichtparametrischen Verfahren, gegenübergestellt.

4.3.3.2 BDM-Test

mit R:

Der BDM-Test ist im Paket `asbio` u.a. als Funktion `BDM.2way` für eine 2-faktorielle Varianzanalyse enthalten. Nachfolgend ein Beispiel mit demselben oben benutzten Datensatz:

```
library(asbio)
with(mydata3, BDM.2way(x, gruppe, dosis))
```

Two way Brunner-Dette-Munk test				
	df1	df2	F*	P(F > F*)
X1	1.000000	14.05996	0.4143377	0.53013638
X2	2.786237	14.05996	2.9306761	0.07310691
X1:X2	2.786237	14.05996	0.3190448	0.79777127

In der Ausgabe werden mit `x1` und `x2` die beiden Faktoren bezeichnet, hier also Gruppe (`x1`)

und Dosis (x_2). Das Testergebnis zeigt, dass der BDM-Test noch konservativer reagiert als die beiden vorher durchgeführten Tests für heterone Varianzen.

4.3.3.3 Variablentransformationen

mit SPSS

Bei einer Transformation \sqrt{x} erhält man bei der Überprüfung der Varianzhomogenität immerhin noch einen p-Wert von 0,051, was allerdings akzeptabel wäre. Doch bei einer Transformation $\log(x)$ verbessert sich das Ergebnis auf $p=0,170$. Die entsprechende Varianzanalyse für die Variable $\ln x = \ln(x)$ ergibt:

Abhängige Variable: $\ln x$					
Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Korrigiertes Modell	1,072 ^a	7	,153	1,743	,144
Konstanter Term	112,524	1	112,524	1280,659	,000
Gruppe	,097	1	,097	1,104	,303
Dosis	,854	3	,285	3,241	,039
Gruppe*Dosis	,139	3	,046	,526	,669
Fehler	2,197	25	,088		
Gesamt	123,359	33			
Korrigierte Gesamtvariation	3,269	32			

so dass hier die log-Transformation wirklich zum Erfolg geführt hat, da zum einen die Varianzen „stabilisiert“ worden sind und zum anderen der Gruppen-Effekt signifikant ist.

4.3.4 Rank transform-Tests (RT)

Bei den einfachen Rank transform Tests (RT) wird lediglich vor der Durchführung der parametrischen Varianzanalyse die abhängige Variable in Ränge transformiert. Die statistischen Tests bleiben unverändert. Dieses Verfahren von Conover & Iman [6] ist in erster Linie für metrische Variablen gedacht, die die Normalverteilungs-Voraussetzung nicht erfüllen, nicht jedoch für Variablen mit beliebigen Eigenschaften. D.h. hat die untransformierte Variable x ungleiche Varianzen, so kann das auch noch für die rangtransformierte Variable R_x gelten. So kann es sinnvoll sein, gegebenenfalls auch R_x auf Varianzhomogenität zu überprüfen und gegebenenfalls einen der Tests in Kapitel 4.3.3 zu benutzen. Für die beiden nachfolgend benutzten Datensätze erübrigt sich dies allerdings, da in Kapitel 4.3.2 für diese keine Varianzinhomogenitäten nachgewiesen worden waren.

Das Verfahren wird sowohl am ersten Datensatz (`mydata1`) als auch am zweiten (`mydata2`) demonstriert.

mit R:

Für das o.a. erste Beispiel (Daten `mydata1`) sind die Anweisung wie folgt zu modifizieren:

```
mydata1 <- within(mydata1, {drugs<-factor(drugs);
                          patients<-factor(patients); rx<-rank(x)})
aov1r <- aov(rx~patients*drugs, mydata1)
summary(aov1r)
```

mit dem Ergebnis:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
patients	1	72.00	72.00	6.680	0.02389	*
drugs	2	56.58	28.29	2.625	0.11333	
patients:drugs	2	217.58	108.79	10.094	0.00268	**
Residuals	12	129.33	10.78			

Tabelle 4-5

Für das o.a. zweite Beispiel lauten die Anweisungen:

```
mydata2 <- within(mydata2, {drugs<-factor(drugs);
                           group<-factor(group); rx<-rank(x)})
aov2r <- aov(rx~group*drugs,mydata2)
drop1 (aov2r, ~., test="F")
```

mit der Ausgabe:

rx ~ group * drugs	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)	
<none>			1083.8	131.23			
group	1	364.17	1448.0	138.79	8.4003	0.0076982	**
drugs	3	1157.72	2241.5	149.21	8.9018	0.0003464	***
group:drugs	3	464.61	1548.4	137.00	3.5724	0.0281287	*

Tabelle 4-6

mit SPSS:

Zunächst muss über das Menü „Transformieren -> Rangfolge bilden“ bzw. über die Syntax

```
Rank variables=x (A) /rank into Rx.
```

x in Ränge transformiert werden, woraus die neue Variable Rx resultiert. Die Varianzanalyse von Rx:

Abhängige Variable: Rank of x					
Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Korrigiertes Modell	346,167 ^a	5	69,233	6,424	,004
patients	72,000	1	72,000	6,680	,024
drug	56,583	2	28,292	2,625	,113
patients * drug	217,583	2	108,792	10,094	,003
Fehler	129,333	12	10,778		

Tabelle 4-7

Für das o.a. zweite Beispiel:

Abhängige Variable: Rank of x					
Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Korrigiertes Modell	1820,713 ^a	7	260,102	6,000	,000
group	364,168	1	364,168	8,400	,008
drugs	1157,722	3	385,907	8,902	,000
group * drugs	464,611	3	154,870	3,572	,028
Fehler	1083,787	25	43,352		

Tabelle 4-8

Wie ein Vergleich mit den Ergebnissen der parametrischen Varianzanalyse (vgl. Kapitel 4.3.2) zeigt, weichen die Ergebnisse des Rank transform Tests nur geringfügig ab.

4.3.5 Verallgemeinerte Kruskal-Wallis- und Friedman-Analysen (KWF)

Diese Verfahren gehen gegenüber dem RT-Verfahren einen Schritt weiter: Es werden nicht die F-Tests verwendet, sondern aus den Streuungsquadratsummen (SS, Sum of Sq) werden χ^2 -Tests konstruiert. Diese können als Verallgemeinerung des Kruskal-Wallis-H-Test angesehen werden, da diese im 1-faktoriellen Fall mit letzterem identisch sind. Die χ^2 -Werte haben den Aufbau (vgl. Formel 2-6):

$$\chi^2 = \frac{SS_{\text{Effekt}}}{MS_{\text{total}}}$$

wobei SS_{Effekt} die Streuungsquadratsumme (SS, Sum of Squares) des zu testenden Effektes (A, B oder A*B) ist und MS_{total} die Gesamtvarianz (MS, Mean Square). Sie haben die gleichen Freiheitsgrade wie der Zähler des entsprechenden F-Tests.

Da bei der Errechnung der Testgröße nicht die Reduzierung des Fehlers durch andere im Versuchsplan berücksichtigte Faktoren eingeht, hat er zwangsläufig eine geringere Effizienz wie z.B. der o.a. Rank transform Test, der die in Kapitel 4.3.1.3 erwähnte Fehlerreduzierung durch mehrfaktorielle Designs ausnutzt, oder der unten aufgeführte ART.

Natürlich könnte man die o.a. χ^2 -Werte mit dem Taschenrechner ausrechnen und mit den kritischen Werten in den klassischen Tafelwerken vergleichen. Z.B. für den Test von Faktor `patients` (aus dem ersten Datensatz `mydata1`) errechnet man zunächst $MS_{\text{total}} = 27,94$. In SPSS ist dieser Wert aus der Zeile `Korrigierte Gesamtvariation` zu entnehmen (vgl. Tabelle 4-7: $475,500/17$), während in R die SS und df aufzusummieren sind (vgl. Tabelle 4-5: $(72,0 + 56,58 + 217,58 + 129,33) / (1 + 2 + 2 + 12)$). Anschließend die Testgröße:

$$\chi^2_{\text{patients}} = \frac{72,0}{27,94} = 2,58$$

Da der kritische Wert bei 1 Fg bei einem $\alpha=0.05$ 3,84 beträgt, bestätigt der errechnete χ^2 -Wert einen signifikanten Einfluss der Patientengruppen.

Mit SPSS ist man auch darauf beschränkt. Mit R lassen sich allerdings diese Schritte auch „programmieren“.

Nachfolgend wird das Verfahren mit R an den Beispieldaten 1 (`mydata1`) und 2 (`mydata2`) demonstriert, mit SPSS nur am ersten Datensatz.

mit R:

An dieser Stelle sollen die Berechnungen mit der Funktion `aov` durchgeführt werden. Die alternative Verwendung von `ezANOVA` wird in Kapitel 5 gezeigt.

Die o.a. Anova-Tabelle 4-5 `aov1r` für das erste Beispiel wird nun weiterverarbeitet.

- Als erstes ist das Objekt `aov1r` mithilfe der Funktion `anova` zu wandeln, damit die Werte in einer Matrix einzeln ansprechbar sind.

- Zunächst muss MS_{total} als Summe der Sum Sq-Spalte (2. Spalte) dividiert durch die Summe der df-Spalte (1. Spalte) berechnet werden.
- Anschließend wird die 2. Spalte durch die MS_{total} dividiert.
- Errechnen der p-Werte mithilfe der Freiheitsgrade der F-Werte in der 1. Spalte.
- Zum Schluss wird aus den Berechnungen ein Dataframe erstellt, für den die Effektnamen (Zeilenamen) von `aov1x` übernommen werden.

D.h. die oben in Kapitel 4.3.4 angeführten R-Kommandos sind zu ergänzen um:

```
aov1x <- anova(aov1r)
mstotal <- sum(aov1x[,2])/sum(aov1x[,1])
chisq <- aov1x[,2]/mstotal
df <- aov1x[,1]
pvalues <- 1-pchisq(chisq,df)
aov1y <- data.frame(chisq,df,pvalues)
row.names(aov1y) <- row.names(aov1x)
aov1y[1:3,]
```

Die daraus resultierende Ausgabe:

	chisq	df	pvalues
patients	2.574132	1	0.10862364
drugs	2.022958	2	0.36368065
patients:drugs	7.779005	2	0.02045552

Tabelle 4-9

Ein Vergleich mit den Tabellen 4-1 und 4-5 zeigt, dass in diesem Fall nicht alle Signifikanzen der parametrischen bzw. der Rank transform Tests mit den verallgemeinerten Kruskal-Wallis-Friedman-Tests reproduziert werden können. Anzumerken ist noch, dass das Testergebnis für den Faktor `drugs` (wie vorher bereits darauf hingewiesen) identisch ist mit dem Kruskal-Wallis H-Test, der 1-faktoriellen Analyse (vgl. Kapitel 4.2.1).

Für das o.a. zweite Beispiel sind auch hier wegen des unbalancierten Versuchsplans ein paar zusätzliche Schritte erforderlich. Insbesondere werden mit `drop1` die Streuungsquadrate vom Typ III ermittelt. Die Berechnung von MS_{total} erfolgt wie oben aus der ursprünglichen Varianzanalyse `aov2r` durch Summation der Streuungsquadratsummen `aov2x[,2]` und Residuen `aov2x[,1]`. `aov2r` muss wie im vorigen Beispiel mit `anova` in ein verarbeitbares Format gebracht werden. Zu beachten ist, dass die Ausgabe von `drop1`, auf `aov2s` gespeichert, eine redundante 1. Zeile enthält (vgl. Tabelle 4-6).

```
mydata2 <- within(mydata2, {drugs<-factor(drugs);
                           group<-factor(group); rx<-rank(x)})
aov2r <- aov(rx~group*drugs,mydata2)
aov2s <- drop1(aov2r, ~. , test="F")
aov2x <- anova(aov2r)
mstotal <- sum(aov2x[,2])/sum(aov2x[,1])
chisq <- aov2s[,2]/mstotal
df <- aov2s[,1]
pvalues <- 1-pchisq(chisq,df)
aov2y <- data.frame(chisq,df,pvalues)
row.names(aov2y) <- row.names(aov2s)
aov2y[2:4,]
```

mit der Ausgabe:

	chisq	df	pvalues
group	4.012175	1	0.045172850
drugs	12.755071	3	0.005197361
group:drugs	5.118797	3	0.163302051

Tabelle 4-10

Ein Vergleich mit den Tabellen 4-2 und 4-6 zeigt, dass auch in diesem Fall nicht alle Signifikanzen der parametrischen bzw. der Rank transform Tests mit den verallgemeinerten Kruskal-Wallis-Friedman-Tests reproduziert werden können.

mit SPSS:

Ausgangsbasis ist die Anova-Tabelle 4-7. Zunächst muss MS_{total} die Gesamtvarianz, in SPSS *korrigierte Gesamtvariation* bezeichnet, berechnet werden, da nur die Quadratsumme und Freiheitsgrade ausgegeben werden, nicht aber das Mittel der Quadrate (Mean Square):

$$MS_{total} = \frac{475}{17} = 27,94$$

Anschließend werden für jeden Effekt die χ^2 -Werte errechnet:

$$\chi^2_{patients} = \frac{72}{27,94} = 2,58 \quad df_{patients} = 1$$

$$\chi^2_{drugs} = \frac{56,68}{27,94} = 2,03 \quad df_{drugs} = 2$$

$$\chi^2_{Interaktion} = \frac{217,53}{27,94} = 7,78 \quad df_{Interaktion} = 2$$

Die 5%-Schranken für die χ^2 -Verteilung liegen bei 3,8 für $df=1$ bzw. 6,0 für $df=2$. Somit liegt nur ein signifikanter Interaktionseffekt vor. Ein Vergleich mit den Tabellen 4-3 und 4-7 zeigt, dass in diesem Fall nicht alle Signifikanzen der parametrischen bzw. der Rank transform Tests mit den verallgemeinerten Kruskal-Wallis-Friedman-Tests reproduziert werden können.

Auf die Berechnung für das zweite Beispiel kann hier verzichtet werden, da in SPSS nicht zwischen balancierten und unbalancierten Versuchsplänen unterschieden werden muss.

4.3.6 Aligned rank transform (ART)

Verschiedene Studien, u.a. von Sawilowsky, S., Blair, R. C., & Higgins, J. J. (vgl. [14]), haben gezeigt, dass für den Test der Interaktion, insbesondere nach dem o.a. Rank transform-Verfahren, der Fehler 1. Art nicht immer korrekt eingehalten wird, d.h. dass mehr Interaktionen zufällig signifikant sind, als es das vorgegebene α zulässt. Als Ursache wird angesehen, dass der Test der Interaktion nicht von den Tests der beiden Haupteffekte unabhängig ist. Als Lösung wird propagiert, zunächst ein komplettes Modell zu analysieren, anschließend für dessen Residuen die beiden Haupteffekte herauszupartialisieren, dann diese bereinigten Residuen in Ränge umzurechnen, um schließlich wiederum ein normales Modell mit Interaktion zu rechnen. Die Streuungsquadrate für die Haupteffekte sollten dann bei diesem Modell bei Null liegen. Die Haupteffekte sind dann aus der Analyse des ersten Modells zu entnehmen. Beim zweiten Mo-

dell interessiert dann lediglich der Test für die Interaktion. Im Folgenden werden auch ART-Tests der Haupteffekte durchgeführt, wenn das auch nicht erforderlich ist.

Die Schritte im Einzelnen:

- Durchführung einer (normalen) Anova mit Haupt- und Interaktionseffekten.
- Speichern der Residuen (e_m),
- Eliminieren des zu untersuchenden Effekts aus den Residuen:
 Interaktionseffekt: $e_m + (\alpha\beta_{ij} - \alpha_i - \beta_j + 2\mu)$
 Haupteffekte: $e_m + (\alpha_i + \beta_j - \mu)$
 bzw. wenn beide Haupteffekte separat getestet werden sollen:
 Haupteffekt A: $e_m + \alpha_i$
 Haupteffekt B: $e_m + \beta_j$
 bzw. im Fall einer 3-faktoriellen Varianzanalyse für die 3-fach-Interaktion:
 Interaktionseffekt: $e_m + (\alpha\beta\gamma_{ijl} - \alpha\beta_{ij} - \alpha\gamma_{il} - \beta\gamma_{jl} + \alpha_i + \beta_j + \gamma_l)$
- Umrechnung der bereinigten Residuen in Ränge.
- Durchführung einer normalen Anova mit Haupt- und Interaktionseffekten mit den Rängen, aus der dann der untersuchte Effekt abgelesen werden kann.

Es soll nun im Folgenden für den Beispieldatensatz 2 überprüft werden, ob die oben ausgewiesene Signifikanz der Interaktion garantiert ist.

mit R:

Zunächst die Durchführung des Verfahrens „per Hand“, d.h. das Alignment, also die Umrechnung der Werte wird elementar vorgenommen.

Dazu wird als erstes mit `x(mydata2[, 3])` die klassische Anova errechnet (`aov3`) und daraus die Residuen ermittelt. Zu den Residuen werden dann einmal zur Ermittlung der Interaktion dieser Effekt addiert (`rab`) sowie einmal zur Ermittlung des Haupteffekte die entsprechende Effekt addiert (`ra` und `rb`). Anschließend werden die bereinigten Residuen in Ränge transformiert (`rabr` bzw. `rar`). Zur Überprüfung der Interaktion bzw. der Haupteffekte wird jeweils ein komplettes Modell mit diesen Residuenrängen analysiert. Gemäß den Anmerkungen in Kapitel 3.3 zu Fehlern bei der Rangberechnung empfiehlt es sich, vorher die bereinigten Residuen mittels `round` auf 7 Dezimalstellen zu runden.

```
mydata2 <- within(mydata2, {drugs<-factor(drugs); group<-factor(group)})
aov3 <- aov(x~group*drugs, mydata2)
rab <- aov3$residuals
ra <- rab
# Zellenmittelwerte
mij <- ave(mydata2[, 3], mydata2[, 1], mydata2[, 2], FUN=mean)
ai <- ave(mydata2[, 3], mydata2[, 1], FUN=mean) # Effekte Faktor A
bj <- ave(mydata2[, 3], mydata2[, 2], FUN=mean) # Effekte Faktor B
mm <- mean(mydata2[, 3]) # Gesamtmittel
# Bereinigung der Residuen
rab <- rab + (mij - ai - bj + 2*mm) # Interaktion
ra <- ra + (ai + bj - mm) # Haupteffekte
rabr <- rank(round(rab, digits=7)) # Runden und
rar <- rank(round(ra, digits=7)) # Umrechnung in Ränge
aov3ab <- aov(rabr~group*drugs, mydata2) # Anova Interaktion
drop1(aov3ab, ~. , test="F") # Ergebnis Interaktionseffekt
```

```
aov3a <- aov(rar~group*drugs,mydata2)      # Anova Haupteffekte
drop1(aov3a, ~. , test="F")                # Ergebnis Haupteffekte
```

mit den Ergebnissen für den Interaktionseffekt:

rabr ~ group * drugs	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
<none>			2094.9	152.97		
group	1	15.16	2110.1	151.21	0.1809	0.67423
drugs	3	2.48	2097.4	147.01	0.0099	0.99862
group:drugs	3	876.49	2971.4	158.51	3.4866	0.03058 *

sowie für die Haupteffekte:

rar ~ group * drugs	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
<none>			1223.2	135.22		
group	1	363.94	1587.1	141.81	7.4385	0.0115045 *
drugs	3	1407.31	2630.5	154.49	9.5879	0.0002159 ***
group:drugs	3	4.14	1227.3	129.33	0.0282	0.9934370

Tabelle 4-11

Vergleicht man diese Ergebnisse mit dem Ergebnis der Rank transform Tests von x (vgl. Tabelle 4-6), können sowohl die Interaktion als auch die Haupteffekte als gesichert angesehen werden. Es sei noch angemerkt, dass die beiden o.a. Ergebnisse für die Interaktion sowie die Haupteffekte ohne die Rundung mittels `round` leicht von den obigen abweichen.

Seit Anfang 2015 wird das Paket `ARTool` für R zur Verfügung gestellt, mit dessen Hilfe die Umrechnung der Werte bequem vorgenommen werden kann. Dazu dient die Funktion `art`, die u.a. unter `$aligned.ranks` die Ränge der umgerechneten Werte für alle Effekte als Dataframe enthält. Die beiden Argumente der Funktion sind mit denen von `aov` identisch. Doch Vorsicht: die Namen der Variablen sind die Namen der Effekte, in diesem Beispiel also `group`, `drugs` und `group:drugs`, also in der Regel mit den Faktornamen identisch und sollten daher umbenannt werden.

Nachfolgend die Durchführung des ART-Verfahrens zur Ermittlung des bereinigten Tests für die Interaktion. `mydata2.a` ist das Ergebnis von `art`, das mit dem Ausgangsdatensatz mittels `cbind` zusammengeführt wird. Dabei erhalten die Variablennamen durch die Angabe `aligned=` das Präfix `aligned`.

```
library(ARTool)
mydata2.a <- art(x~group*drugs,mydata2)$aligned.ranks
mydata2x <- cbind(mydata2,aligned=mydata2.a)
names(mydata2x)[4:6] <- c("a.group","a.drugs","a.gd")
drop1(aov(a.gd~group*drugs,mydata2x), ~. , test="F")
```

a.gd ~ group * drugs	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
<none>			2094.9	152.97		
group	1	15.16	2110.1	151.21	0.1809	0.67423
drugs	3	2.48	2097.4	147.01	0.0099	0.99862
group:drugs	3	876.49	2971.4	158.51	3.4866	0.03058 *

mit SPSS:

- Zunächst wird für x die klassische Anova (Unianova) errechnet und dabei die Residuen gespeichert.
- Dann müssen mittels `Aggregate` die Effekte als Mittelwerte für die Gruppen ermittelt werden: m_{ij} für die Interaktion, a_i für Faktor `group` und b_j für Faktor `drugs`. Diese werden in die Arbeitsdatei eingefügt.
- Zu den Residuen werden dann einmal zur Ermittlung der Interaktion dieser Effekt addiert (r_{ab}) sowie einmal zur Ermittlung der Haupteffekte deren Effekte addiert (r_a und r_b).
- Anschließend werden die bereinigten Residuen in Ränge transformiert (r_{abr} bzw. r_{ar}).
- Zur Überprüfung der Interaktion bzw. der Haupteffekte wird jeweils ein komplettes Modell mit diesen Residuenrängen analysiert.

```

Unianova  x by group drugs
  /save=resid (rab)
  /design=group drugs group*drugs.

Compute ra=rab.

Aggregate
  /outfile=* mode=addvariables
  /break=group drugs   /mij=mean(x) .
Aggregate
  /outfile=* mode=addvariables
  /break=group         /ai=mean(x) .
Aggregate
  /outfile=* mode=addvariables
  /break=drugs         /bj=mean(x) .
Aggregate
  /outfile=* mode=addvariables
  /break=              /mm=mean(x) .

Compute rab=rab + (mij - ai - bj + 2*mm) .
Compute ra =ra  + (ai + bj - mm) .

Rank variables=ra rab (A)
  /rank into rar rabr.

Unianova  rabr by group drugs
  /design=group drugs group*drugs.
Unianova  rar  by group drugs
  /design=group drugs group*drugs.

```

mit den Ergebnissen für den Interaktionseffekt:

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
group	10,592	1	10,592	,131	,721
drugs	40,762	3	13,587	,167	,917
group * drugs	938,767	3	312,922	3,856	,021
Fehler	2028,817	25	81,153		

sowie für die Haupteffekte:

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
group	319,765	1	319,765	5,638	,026
drugs	1267,690	3	422,563	7,450	,001
group * drugs	8,802	3	2,934	,052	,984
Fehler	1418,017	25	56,721		

Tabelle 4-12

Vergleicht man diese Ergebnisse mit dem Ergebnis der Rank transform Tests von x (vgl. Tabelle 4-8), können sowohl die Interaktion als auch die Haupteffekte als gesichert angesehen werden.

4.3.7 normal scores- (INT-) und van der Waerden-Tests

Bei der einfachen *inverse normal transformation* (INT) wird lediglich vor der Durchführung der parametrischen Varianzanalyse zunächst die abhängige Variable x in Ränge $R(x)$ transformiert und anschließend über die inverse Normalverteilung in normal scores umgerechnet:

$$nscore_i = \Phi^{-1}(R(x_i)/(N+1))$$

wobei N die Anzahl der Werte ist und Φ^{-1} die Umkehrfunktion der Normalverteilung. Die statistischen Tests bleiben unverändert. Dieses Verfahren ist wie beim o.a. RT-Verfahren in erster Linie für metrische Variablen gedacht, die die Normalverteilungs-Voraussetzung nicht erfüllen, nicht jedoch für Variablen mit beliebigen Eigenschaften. D.h. hat die untransformierte Variable x ungleiche Varianzen, so kann das auch noch für die transformierte Variable $norm(R(x))$ gelten. So kann es sinnvoll sein, gegebenenfalls auch $norm(R(x))$ auf Varianzhomogenität zu überprüfen und gegebenenfalls einen der Tests in Kapitel 4.3.3 zu benutzen. In den nachfolgenden Beispielen wird darauf verzichtet, da bereits die nichttransformierten Daten homogen sind.

Bei dem Verfahren von *van der Waerden* werden anstatt der „klassischen“ F-Tests die χ^2 -Tests des Kruskal-Wallis-Tests bzw. wie bei der o.a. KWF-Methode gerechnet. Die χ^2 -Werte haben den Aufbau (vgl. Formel 2-6a):

$$\chi^2 = \frac{SS_{Effekt}}{MS_{total}}$$

wobei SS_{Effekt} die Streuungsquadratsumme (SS, Sum of Squares) des zu testenden Effektes (A, B oder A*B) ist und MS_{total} die Gesamtvarianz (MS, Mean Square). Sie haben die gleichen Freiheitsgrade wie der Zähler des entsprechenden F-Tests. (Vgl. auch Kapitel 4.3.5.)

Im folgenden Beispiel wird der zuletzt benutzte Datensatz `mydata2` verwendet.

mit R:

Wegen des nichtbalancierten Versuchsplans müssen zunächst mittels `option` die Standard-Kontraste zugewiesen werden sowie nach der Anova mit `aov` mittels `drop1` Quadratsummen vom Typ III errechnet werden. `nc` enthält die Anzahl der Merkmalsträger, die bei der Umrechnung in normal scores einfließt.

```
options (contrasts=c("contr.sum", "contr.poly"))
nc      <- dim(mydata2) [1]
Rx      <- rank(x)
nsx     <- qnorm(Rx/(nc+1))
aov2ns  <- aov(nsx~group*drugs, mydata2)
aov2ns1 <- drop1(aov2ns, ~. , test="F")
```

Diese Anweisungen dienen zunächst für die Analyse der normal scores (INT-Verfahren) mit folgendem Ergebnis:

Model:						
nsx ~ group * drugs						
	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
<none>			9.3768	-25.5229		
group	1	2.9521	12.3289	-18.4905	7.8708	0.0095852 **
drugs	3	10.6917	20.0684	-6.4128	9.5019	0.0002289 ***
group:drugs	3	4.1290	13.5058	-19.4817	3.6696	0.0256032 *

Tabelle 4-13

Für die Durchführung der van der Waerden-Tests sind noch zusätzlich die folgenden Anweisungen erforderlich, um die χ^2 -Tests durchzuführen (vgl. auch das Beispiel in Kapitel 4.3.5):

```
aov2ns  <- anova(aov2ns)
mstotal <- sum(aov2ns[,2])/sum(aov2ns[,1])
chisq   <- aov2ns[,2]/mstotal
df      <- aov2ns[,1]
pvalues <- 1-pchisq(chisq,df)
aov2vdw <- data.frame(chisq,df,pvalues=round(pvalues,digits=5))
row.names(aov2vdw) <- row.names(aov2ns1)
aov2vdw[2:4,]
```

	chisq	df	pvalues
group	3.710002	1	0.05409
drugs	13.436428	3	0.00378
group:drugs	5.189060	3	0.15847

Ein Vergleich mit den Tabellen 4-6 und 4-13 zeigt, dass in diesem Fall nicht alle Signifikanzen der Rank transform Tests bzw. des einfachen normal scores-Tests mit den van der Waerden-Tests reproduziert werden können.

Alternativ kann das van der Waerden-Verfahren auch mit der Funktion `np.anova` (vgl. Anhang 3.6) durchgeführt werden. Der Aufruf ist praktisch identisch mit dem der Standardfunktion `aov`. Über den Zusatz `method=1` wird anstatt des KWF-Verfahrens das van der Waerden-Verfahren durchgeführt. Nachfolgend die Ein- und Ausgabe:

```
np.anova(x~group*drugs, mydata2, method=1)
```

generalized van der Waerden tests				
	Df	Sum Sq	Chi Sq	Pr(>Chi)
group	1	2.9521	3.7100	0.054087
drugs	3	10.6917	13.4364	0.003782
group:drugs	3	4.1290	5.1891	0.158465
Residuals	25	9.3768		

mit SPSS:

Die Rang-Transformation sowie die Umrechnung in normal scores werden zweckmäßigerweise über das Syntax-Fenster vorgenommen. Das für die Umrechnung erforderliche n (Anzahl der Fälle, Variable `nc`) wird über `Aggregate` ermittelt. Die Ergebnisvariable wird `nsx` genannt:

```
Aggregate
  /outfile=* mode=addvariables
  /break= /nc=NU(x).
Rank Variables=x / rank into Rx.
compute nsx=Idf.normal(Rx/(nc+1),0,1).
execute.
```

Abhängige Variable: nsx					
Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Korrigiertes Modell	16,086 ^a	7	2,298	6,127	,000
Konstanter Term	,163	1	,163	,435	,516
group	2,952	1	2,952	7,871	,010
drugs	10,692	3	3,564	9,502	,000
group * drugs	4,129	3	1,376	3,670	,026
Fehler	9,377	25	,375		
Gesamt	25,463	33			
Korrigierte Gesamtvariation	25,463	32			

Tabelle 4-14

Für den van der Waerden-Test müssen wie beim KWF-Verfahren (Kapitel 4.3.5) χ^2 -Werte errechnet werden. Zunächst muss MS_{total} die Gesamtvarianz, in SPSS *korrigierte Gesamtvariation* bezeichnet, berechnet werden, da nur die Quadratsumme und Freiheitsgrade ausgegeben werden, nicht aber das Mittel der Quadrate (Mean Square):

$$MS_{total} = \frac{25,46}{32} = 0,796$$

Anschließend werden für jeden Effekt die χ^2 -Werte errechnet:

$$\chi^2_{group} = \frac{2,95}{0,796} = 3,71 \quad df_{patients} = 1$$

$$\chi^2_{drugs} = \frac{10,69}{0,796} = 13,43 \quad df_{drugs} = 3$$

$$\chi^2_{Interaktion} = \frac{4,13}{0,796} = 5,19 \quad df_{Interaktion} = 3$$

Die 5%-Schranken für die χ^2 -Verteilung liegen bei 3,8 für $df=1$ bzw. 7,8 für $df=3$. Somit liegt nur ein signifikanter Haupteffekt (Faktor `drugs`) vor. Ein Vergleich mit den Tabellen 4-8 und 4-14 zeigt, dass in diesem Fall nicht alle Signifikanzen der Rank transform Tests bzw. des einfachen normal scores-Tests mit den van der Waerden-Tests reproduziert werden können.

4.3.8 ATS-Tests von Akritas, Arnold & Brunner

Während auf der einen Seite die Rank transform Tests inklusive dem ART heuristische Verfahren sind, die sich in der Praxis mehr oder weniger als brauchbar erwiesen haben, und auf der anderen Seite die verallgemeinerten Kruskal-Wallis-Friedman-Tests keine hohe Effizienz haben, sind die Tests von Akritas, Arnold & Brunner aufwändig hergeleitet worden. Die Berechnung ist vergleichsweise kompliziert, so dass sie in SPSS nicht möglich ist und in R einen erheblichen Programmieraufwand erfordert. Sie ist allerdings sehr übersichtlich dokumentiert in dem Buch von Edgar Brunner und Ullrich Munzel (vgl. [26]). Eine entsprechende R-Funktion ist im Anhang 3 zu finden.

mit R:

Zur Verfügung stehen zwei Funktionen: `ats.2` und `ats.3` für 2- bzw. 3-faktorielle Analysen. Das Verfahren soll am 2. Datensatz demonstriert werden. Nachfolgend die Anweisungen sowie die Ausgabe:

```
ats.2(x~group*drugs, mydata2)
```

Response: x			
	Df	F value	Pr(>F)
group	1.0000	9.0685	0.009704
drugs	2.2550	9.7698	0.001914
group:drugs	2.6995	3.5608	0.046886
Residuals	13.4386		

die zeigt, dass mit diesem Verfahren alle drei Signifikanzen der Rank transform Tests (vgl. Tabelle 4-6) und des ART (vgl. Tabelle 4-11) reproduziert werden können.

4.3.9 Bredekamp Tests

Der Test von Bredekamp ist ein Spezialfall der Tests von Puri & Sen (KWF), und zwar ausschließlich für balancierte Versuchspläne. Er bedient sich des H-Tests von Kruskal-Wallis, also nur der 1-faktoriellen Varianzanalyse, und ist sehr einfach durchzuführen. Dabei ist zu bedenken, dass dieses Verfahren noch aus einer Zeit stammt, bevor die (inzwischen vielen) neueren Verfahren zur nichtparametrischen Varianzanalyse publiziert waren.

Das Verfahren beruht auf der Additivität des χ^2 -Tests. Zunächst wird ein H-Test über alle Zellen hinweg gerechnet. Von diesem χ^2 -Wert werden die Werte aus den H-Tests für Faktor A und Faktor B subtrahiert. Das Ergebnis ist der χ^2 -Wert für die Interaktion. Analog werden die Freiheitsgrade ermittelt.

H-Testwerte (χ^2 -Werte)	Freiheitsgrade
χ^2_{AB}	$IJ-1$
- χ^2_A	$I-1$
- χ^2_B	$J-1$
$\chi^2_{AB} - \chi^2_A - \chi^2_B$	$(I-1)(J-1)$

Da das Prozedere mit R und SPSS gleichermaßen abläuft, soll hier nur mit SPSS ein Beispiel durchgerechnet werden.

mit SPSS:

Es wird hier eine 2-faktorielle Varianzanalyse für das erste Datenbeispiel durchgeführt, das einen balancierten Versuchsplan beinhaltet. Zunächst müssen die Zellen für den ersten H-Test einmal durchnummeriert werden:

Durchnummerierung der Zellen:= (Patients - 1)*#Drugs + Drug

Anschließend werden H-Tests für die Zellen, für Faktor `patients` und Faktor `drug` gerechnet. Die SPSS-Syntax hierfür:

```
Compute zelle=(patients-1)*3 + drug.
Nptests /independent test (x) group (zelle) kruskal_wallis.
Nptests /independent test (x) group (patients) kruskal_wallis.
Nptests /independent test (x) group (drug) kruskal_wallis.
```

Die SPSS-Ergebnisse sind in folgender Tabelle zusammengefasst:

Effekt	H-Testwerte (χ^2 -Werte)	Freiheitsgrade	Signifikanz
Zellen	12,376	5	
patients	2,574	1	n.s. (< 3,8)
drug	2,023	2	n.s. (< 3,8)
patients*drug	7,779	2	s. (> 6,0)

Tabelle 4-15

Diese Ergebnisse decken sich mit denen aus den Tests von Puri & Sen (vgl. Tabelle 4-9).

4.4 Nichtparametrische Verfahren zur mehrfaktoriellen Varianzanalyse

Die in 4.3. vorgestellten Verfahren lassen sich alle ohne Weiteres auf drei und mehr Faktoren erweitern.

4.5 Fazit

Ist das zu analysierende Merkmal metrisch, sollte man zunächst die Voraussetzungen prüfen und danach entscheiden, ob überhaupt in Anbetracht der Robustheit der Varianzanalyse ein nichtparametrisches Verfahren erforderlich ist. Die einfachsten Wege der nichtparametrischen Varianzanalyse sind natürlich der simple Rank transform Test (RT) und die normal scores-Tests (INT), zumal diese eine relativ hohe Effizienz haben. Da es allerdings gegen das RT-Verfahren einige Bedenken gibt, kann man im Zweifelsfall bei einer signifikanten Interaktion die etwas aufwändigere ART-Variante wählen, wenn auch diese nicht unangefochten ist. Mit den verallgemeinerten Kruskal-Wallis-Friedman-Tests (Puri & Sen-Tests) ist man zwar auf der sicheren Seite, verliert aber an Effizienz, wie aus den Ergebnissen des Kapitels 4.3.5 zu ersehen ist. Dies ist dadurch zu erklären, dass bei mehrfaktoriellen nicht wie bei den anderen Verfahren die Residuenstreuung durch die Hinzunahme anderer Faktoren reduziert und dadurch der Test verbessert wird (vgl. Kapitel 4.3.1.3). Die Tests von Akritis & Co stellen vermutlich eine der zuverlässigsten Lösungen dar, sind allerdings mit SPSS nicht realisierbar.

Hat die abhängige Variable dagegen ordinales Skalenniveau mit einer geringen Anzahl von Ausprägungen, sind eher die verallgemeinerten Kruskal-Wallis-Friedman-Tests bzw. die Tests

von Akritis & Co vorzuziehen.

Abschließend werden für die drei benutzten Datensätze die Ergebnisse aller Verfahren, und zwar die p-Werte für alle drei Effekte, in einer Tabelle gegenüber gestellt. Dabei sind nicht alle hier aufgeführten Ergebnisse in den vorangegangenen Kapiteln wiedergegeben worden. Allerdings sollte man nicht die hier erzielten Ergebnisse verallgemeinern.

.Ergebnisse für den Datensatz 1 (`mydata1`):

Verfahren	patients	drugs	Interaktion
parametrisch	0.014	0.106	0.006
Rank transform Test (RT)	0.024	0.113	0.003
Aligned Rank Transform (ART)	0.019	0.116	0.005
Verall. Kruskal-Wallis-Friedman-Tests (KWF) Bredenkamp Tests	0.109	0.364	0.020
normal scores (INT)	0.027	0.126	0.005
van der Waerden	0.102	0.354	0.027
Akritis, Arnold & Brunner ATS	0.029	0.129	0.005

.Ergebnisse für den Datensatz 2 (`mydata2`):

Verfahren	group	drugs	Interaktion
parametrisch	0.012	0.001	0.026
Rank transform Test (RT)	0.008	0.001	0.028
Aligned Rank Transform(ART)	0.023	0.001	0.005
Verall. Kruskal-Wallis-Friedman-Tests (KWF) Bredenkamp Tests	0.109	0.364	0.016
normal scores (INT)	0.010	0.001	0.026
van der Waerden	0.054	0.004	0.158
Akritis, Arnold & Brunner ATS	0.010	0.002	0.047

Bei beiden Datensätzen zeigt sich, dass die verallgemeinerten Kruskal-Wallis-Friedman- sowie die van der Waerden-Tests vielfach die Signifikanzen der anderen Methoden nicht reproduzieren können.

Der Datensatz 3 (`mydata3`) zeichnete sich durch stark inhomogene Varianzen aus. D.h. hier ist davon auszugehen, dass die Signifikanzen des parametrischen Tests nicht gesichert sind und daher die anderen Verfahren vorzuziehen sind. Allerdings sollte man nicht die hier erzielten Ergebnisse verallgemeinern und etwa schließen, dass der Brown-Forsythe-Test oder eine Variablentransformation wie z.B. $\log(x)$ den anderen Methoden überlegen sind.

Verfahren	gruppe	dosis	Interaktion
parametrisch	0.211	0.035	0.524
Box-Test für heterogene Varianzen	0.185	0.068	0.462
Brown-Forsythe F-Test für heterogene Var.	0.242	0.046	0.809
log(x)-Transformation	0.303	0.039	0.669
BDM-Test	0.530	0.073	0.798
Rank transform Test	0.508	0.064	0.809
Aligned Rank Transform	0.217	0.077	0.522
normal scores (INT)	0.366	0.048	0.725
van der Waerden	0.371	0.048	0.759
Verall. Kruskal-Wallis-Friedman-Tests (KWF) Bredenkamp Tests	0.498	0.057	0.827
Akritas, Arnold & Brunner ATS	0.513	0.058	0.783

5. Abhängige Stichproben - Messwiederholungen

Es wird im Folgenden davon ausgegangen, dass ein Merkmal x K -mal (unter verschiedenen Bedingungen) erhoben wurde, so dass Variablen x_1, \dots, x_K vorliegen, deren Mittelwerte verglichen werden sollen. Z.B. können von dem Merkmal Herzfrequenz HF mehrere Messungen vorliegen, z.B. der Ruhewert, der Wert direkt nach Beendigung des Joggens sowie Werte 10 und 20 Minuten nach Beendigung, also insgesamt 4 Werte. Die Struktur kann aber auch hier mehrfaktoriell sein, wenn z.B. o.a. HF-Messungen einmal ohne Einnahme eines Medikaments und einmal mit Einnahme vorgenommen worden sind.

Beispieldaten 4 (winer518):

Der folgende Datensatz ist dem Buch von B.J.Winer [13], S. 518, entnommen. Die Einstellung zu einem Thema wurde von Männern und Frauen dreimal im Abstand von mehreren Monaten auf einer ordinalen Skala von 1 - 9 (negativ - positiv) erfasst:

Geschlecht	Versuchsperson	t1	t2	t3
Männer	1	4	7	2
	2	3	5	1
	3	7	9	6
	4	6	6	2
	5	5	5	1
Frauen	6	8	2	5
	7	4	1	1
	8	6	3	4
	9	9	5	2
	10	7	1	1

In R muss `Geschlecht` vom Typ „factor“ deklariert sein, ebenso die für die Umstrukturierung zu ergänzende Fallkennzeichnung, etwa `vpn`. In R hat der Dataframe den Namen `winer518`.

Beispieldaten 5 (mydata5):

Im folgenden Datensatz geht es um die sfähigkeit in Abhängigkeit von der Einnahme von 2 verschiedenen Medikamenten. 8 Personen, 4 Männer und 4 Frauen, müssen 3 verschiedene Aufgaben (1, 2, 3) lösen, einmal ohne Einnahme eines Präparats (Kontrollmessung K) sowie je einmal nach Einnahme von Medikament A bzw. B (A, B). Das Kriterium ist die Fehlerzahl, mit der eine Aufgabe bearbeitet wurde. Dieses ist zwar eigentlich metrisch, wegen des kleinen Wertebereichs aber eher ordinal zu handhaben.

Geschlecht	Versuchsperson	Kontrolle K			Medikament A			Medikament B		
		Aufgabe 1			Aufgabe 2			Aufgabe 3		
		1	2	3	1	2	3	1	2	3
Männer	1	3	3	1	4	4	2	5	4	3
	2	2	0	0	3	2	2	4	3	3
	3	5	4	3	5	3	3	6	3	4
	4	3	5	2	4	4	3	4	4	4
Frauen	5	2	2	1	2	2	2	5	2	3
	6	4	1	0	3	2	1	5	2	2
	7	3	2	1	3	2	1	4	3	2
	8	1	3	0	5	2	1	6	3	3

In R muss `Geschlecht` vom Typ „factor“ deklariert sein, ebenso die für die Umstrukturierung zu ergänzende Fallkennzeichnung, etwa `vpn`. In R hat der Dataframe den Namen `mydata5`, in dem die 9 Messwiederholungsvariablen die Namen `v1`, . . . , `v9` haben.

Beispieldaten 6 (winer568):

Der folgende Datensatz ist dem Buch von B.J.Winer [13], S. 568, entnommen. Hierbei handelt es sich um ein Lernexperiment, bei dem in 4 aufeinanderfolgenden Versuchen (Faktor Zeit) jeweils ein Score von 0 bis 20 erzielt werden konnte. Die 12 Versuchspersonen sind bzgl. 2 Kriterien A bzw. B (Faktoren A und B) in jeweils 2 Gruppen eingeteilt worden.:

A	B	Versuchsperson	V1	V2	V3	V4
A1	B1	1	1	6	5	7
		2	0	6	7	9
		3	3	8	8	9
	B2	4	2	7	12	15
		5	1	6	8	9
		6	3	7	10	11
A2	B1	7	1	2	7	12
		8	1	1	4	10
		9	1	1	4	8
	B2	10	2	2	8	12
		11	3	2	10	15
		12	2	2	7	13

In R hat der Dataframe den Namen `winer568`.

5.1 Datenstruktur - Besonderheiten bei R und SPSS

In der Regel liegen die Daten in Form einer Datenmatrix vor, bei der die Zeilen den Erhebungseinheiten (Fällen) entsprechen, also typischerweise Versuchspersonen, und die Spalten den erhobenen Merkmalen (Variablen). Liegen z.B. von der Variablen Herzfrequenz HF die oben aufgeführten 4 Werte vor, so sind diese normalerweise als 4 Variablen (z.B. `HF_Ruhe`, `HF_0`, `HF_10` und `HF_20`), also 4 Spalten, in der Datenmatrix zu finden. Bei den meisten Statistikprogrammen, so auch bei SPSS, werden dann zum Vergleich der Messwiederholungen diese Variablen angegeben.

Nicht so bei R. Hier werden die Messwiederholungen von Variablen nicht als Spalten, sondern als Zeilen in der Datenmatrix wiederholt. Dies erfordert zwei zusätzliche Kennungen:

- eine Kennzeichnung der Erhebungseinheit, üblicherweise Fall- oder Versuchspersonennummer, sowie
- eine Kennung der Messwiederholung, ähnlich einem Gruppierungsfaktor.

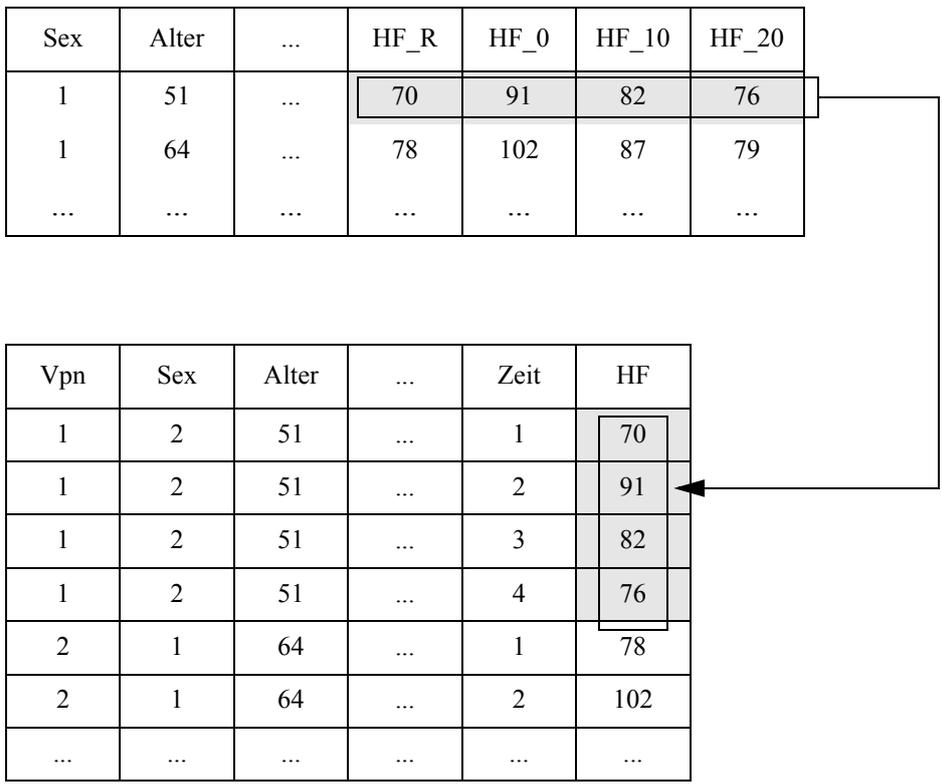
Für die statistischen Funktionen ist es ganz wichtig, dass beide Variablen vom Typ „factor“ sind, insbesondere da die Funktionen auch fehlerfrei durchlaufen, wenn diese Deklaration vergessen wurde. Nur: Die Ergebnisse sind dann falsch. Variablen, die nicht mehrfach gemessen wurden, wie z.B. `Geschlecht`, bleiben dann in den Wiederholungszeilen für die Messwiederholungen konstant.

Zum Wandeln der Datenstruktur, um Versuchspläne mit Messwiederholungen in R analysieren zu können, genügt in der Regel der Aufruf einer entsprechenden Funktion. Seit den Anfängen

von R ist im WWW die Funktion `make.rm` zu finden, die bequem einen Dataframe mit einem Messwiederholungsfaktor umstrukturiert. Inzwischen bietet R standardmäßig die Funktion `reshape`, mit der sowohl Messwiederholungen in Fälle (Parameter `direction=long`), mit ein wenig Aufwand auch für mehrfaktorielle Designs, gewandelt werden können, als auch umgekehrt Fälle in Messwiederholungen (Parameter `direction=wide`).

Allerdings ist eine solche Umstrukturierung verschiedentlich auch bei SPSS erforderlich, und zwar zur Berechnung der Ränge. SPSS bietet nur eine Funktion zur Berechnung von Rängen, und zwar für eine Variable über alle Fälle, also spaltenweise. Bei Messwiederholungen ist allerdings auch die zeilenweise Rangberechnung erforderlich. Daher müssen die Messwiederholungen wie oben skizziert in mehrere Zeilen umgewandelt werden. SPSS bietet dazu Verfahren an. Diese sind ausführlich im Anhang 1 beschrieben.

Der erforderliche Umwandlungsprozess soll an zwei Beispielen veranschaulicht werden. Zunächst einmal an dem einfachen Fall eines Mermals HF, das zu 4 Zeitpunkten beobachtet worden ist (siehe oben): zuerst die Ausgangsbasis, darunter die erforderliche Struktur mit den zusätzlichen Variablen `vpn` (Fallkennzeichnung) und `zeit` (Kennzeichnung der Messwiederholung):



Nachfolgend der etwas kompliziertere Fall von zwei Merkmalen, systolischer und diastolischer Blutdruck (Sys.. bzw. Dia..), die zum einen zu 3 Zeitpunkten (..1, ..2, ..3) und zum anderen ohne und mit einer Medikamentendosierung (..o, ..m) gemessen worden sind. Auch hier sind 3 neue Variablen erforderlich: `vpn` (Fallkennzeichnung), `Dosis` (Messwiederholung Dosierung) und `zeit` (Messwiederholung Zeit). Zunächst die Ausgangsstruktur:

Sex	Alter	Sys1o	Dia1o	Sys2o	Dia2o	Sys3o	Dia3o	Sys1m	Dia1m	Sys2m	Dia2m	Sys3m	Dia3m
2	51	100	71	112	76	121	85	102	69	114	72	118	80
1	64	105	82	116	88	125	93	109	85	114	88	120	93
...

und hier die Daten nach der Umstrukturierung:

Vpn	Sex	Alter	Dosis	Zeit	Sys	Dia
1	2	51	1	1	100	71
1	2	51	1	2	112	76
1	2	51	1	3	121	85
1	2	51	2	1	102	69
1	2	51	2	2	114	72
1	2	51	2	3	118	80
...	

5.2 Voraussetzungen der parametrischen Varianzanalyse

Hier geht es zunächst einmal um Versuchspläne, die ausschließlich abhängige Stichproben beinhalten, also ohne Gruppierungsfaktoren. Für die 1-faktorielle Varianzanalyse lautet das Modell dann für einen Faktor C mit I Messwiederholungen/Stufen - nachfolgend wird gelegentlich auch wieder die Anzahl mit K bezeichnet:

$$x_{im} = \mu + \gamma_i + \pi_m + e_{im} \quad (i=1, \dots, I \text{ und } m=1, \dots, N) \quad (5-1)$$

wobei N die Anzahl der Merkmalsträger/Versuchspersonen ist. Gegenüber dem entsprechenden Modell ohne Messwiederholungen (vgl. Kapitel 4.1) gibt es einen personenspezifischen Effekt: π_m . Die Voraussetzungen betreffen wiederum die Normalverteilung der Residuen und die Varianzhomogenität. Schaut man in die Lehrbücher, so wird dort kaum das Thema Normalverteilung behandelt, sondern im Wesentlichen die Varianzhomogenität, da die, im Gegensatz zur Analyse ohne Messwiederholungen, eine sehr viel größere Bedeutung hat.

Doch zunächst zur Normalverteilung der Residuen. Bei Varianzanalysen mit Messwiederholungen gibt es mehrere Residuen, denn jeder Test eines Wiederholungsfaktors hat seine eigene Fehlervarianz (und die ihr zugrunde liegenden Residuen), über die die Effektvarianz beurteilt wird. Diese Streuungen müssen alle aus normalverteilten Grundgesamtheiten kommen. Dazu sind zwei Tests erforderlich:

Zum einen müssen die Residuen e_{im} geprüft werden. Hier genügt es nicht, die Abweichungen vom Zellenmittelwert \bar{x}_i zu betrachten, vielmehr müssen die personenspezifischen Abweichungen ebenfalls berücksichtigt werden. Dazu muss von den Abweichungen $x_{im} - \bar{x}_i$ noch π_m subtrahiert werden. (Werden diese nicht subtrahiert, können extreme Werte der π_m zu Abweichungen der Residuen von der Normalverteilung führen.) Auch hier entspricht es weder

der Forderung noch ist es praktikabel, die Voraussetzung für jede Messwiederholung bzw. Zelle separat zu überprüfen. Vielmehr sollte man alle Residuen zu einer Variablen zusammenfassen und analysieren. Auf die Ermittlung der Residuen wird in Kapitel 5.3.1 näher eingegangen. Mehr zur Überprüfung auf Normalverteilung im Kapitel 1.6.

Zum anderen müssen die personenspezifischen Abweichungen π_m (*Personeneffekt*) auf Normalverteilung überprüft werden. Diese errechnen sich als Mittelwerte aller Messwiederholungen einer Versuchsperson, wovon noch der Mittelwert $\bar{\pi}$ abzuziehen ist. Für die Überprüfung kann allerdings der letzte Schritt entfallen, da er für die Verteilungsform nicht relevant ist.

Ergeben beide Tests keine Abweichungen, so können alle Residuen als normalverteilt angenommen werden, da sich diese aus den beiden o.a. Residuen zusammensetzen. Zur Prüfung kann wieder zum einen der Shapiro-Wilks-Test, zum anderen grafische Verfahren herangezogen werden. Durch die bei R erforderliche Umstrukturierung der Daten, ist es dort bequem, eine globale Residuen-Variable zu bestimmen und zu untersuchen. Bei SPSS bedarf es dazu etwas mehr Aufwand. Mehr dazu im Kapitel 5.3.1.

Dazu kommt wieder die Voraussetzung der Varianzhomogenität. (Allerdings nur für den Fall $K > 2$. Denn im Fall $K = 2$ kann zum Vergleich der beiden Variablen einfach deren Differenz verwendet werden.) Diese umfasst allerdings hier mehr als die Gleichheit der Varianzen der K zu vergleichenden Variablen: $\sigma_1^2 = \dots = \sigma_K^2$. Die Voraussetzung heißt *Sphärität* der aus den K Variablen gebildeten Kovarianzmatrix. Formal lautet die Bedingung:

$$\sigma_{x_1-x_2}^2 = \sigma_{x_1-x_3}^2 = \sigma_{x_2-x_3}^2 = \dots$$

d.h. die Varianzen von allen Differenzen je zweier Variablen sind gleich. Diese Bedingung ist nicht leicht nachzuvollziehen. Es gibt aber noch eine andere „verständlichere“ Bedingung, die *Compound Symmetry*. Bei dieser wird gefordert, dass zum einen alle K Varianzen gleich sind, und zum anderen die Korrelationen - eigentlich die Kovarianzen, was aber äquivalent ist - je zweier (verschiedener) Variablen gleich sind. Diese Bedingung impliziert die Sphärität.

Diese Bedingung der Sphärität wird für jeden der Tests der Messwiederholungsfaktoren gefordert. Liegt also z.B. ein Design mit zwei Messwiederholungsfaktoren C und D vor, so ist ein entsprechender Test für die Effekte von C, D und C*D durchzuführen.

Zur Prüfung der Sphärität wird allgemein der *Mauchly-Test* verwendet, so auch in R und SPSS. Dieser Test hat allerdings im Vergleich zu einigen anderen Tests Nachteile: Zum einen reagiert er empfindlich auf Abweichungen von der multivariaten (!) Normalverteilung der K zu vergleichenden Variablen, und zum anderen gibt es bessere, effizientere Tests (vgl. dazu Moulton [81]). Es sei noch darauf hingewiesen, dass für diese Tests die Anzahl der Beobachtungen N größer als die Anzahl der Messwiederholungen sein muss. Andernfalls kann der Test nicht durchgeführt werden und alle Werte werden mit 0 ausgegeben.

Die Prüfung beider Voraussetzungen in R bzw. SPSS wird in Kapitel 5.3.1 beschrieben.

Auch hier stellt sich die Frage: Was ist zu tun, wenn eine der Voraussetzungen nicht erfüllt ist? Die in Kapitel 4.1 angeführte Robustheit der Verfahren hinsichtlich Abweichungen von der Normalverteilung gilt hier ganz besonders, da keine unterschiedlichen n_i vorliegen. Abweichungen von der Varianzhomogenität, hier von der Sphärität, sind dagegen gravierender, können aber statistisch aufgefangen werden. Sowohl *Geisser & Greenhouse* als auch *Huynh &*

Feldt haben modifizierte F-Tests entwickelt, die auch bei Abweichungen von der Sphärität angewandt werden können. Hierbei werden (wie häufig in der Statistik, z.B. bei der Welch-Approximation für den klassischen t-Test) die Zähler- und Nenner-Freiheitsgrade des F-Tests entsprechend der Abweichung von Sphärität verkleinert. Hierfür wird ein Korrekturfaktor ε errechnet. Der F-Wert selbst bleibt davon unberührt. Als Konsequenz daraus reagiert der F-Test konservativer, je stärker die Abweichung ist. Von diesen beiden alternativen Tests ist der von Geisser & Greenhouse der konservativere. In SPSS (GLM Messwiederholungen) werden sowohl der Mauchly-Test als auch beide modifizierten F-Tests automatisch immer ausgegeben. In R gibt es Funktionen, die den Mauchly-Test wie auch die F-Tests von Geisser & Greenhouse sowie von Huynh & Feldt ausgeben, u.a. `ezANOVA` in dem Paket `ez`.

Beasley hat in einer umfangreichen Studie [84] gezeigt, dass zum einen das Aligned Rank Transform (ART) Verfahren auch bei Daten, die weder normalverteilt sind noch die Sphärität erfüllen, sowohl der Fehler 1. Art α eingehalten wird, als auch der Fehler 2. Art unter Kontrolle bleibt. Darüber hinaus wird darauf hingewiesen, dass bei einer „einfachen“ Rangtransformation Verteilungseigenschaften meist erhalten bleiben, wenn auch in abgeschwächter Form. (Hierauf wird auch von Fan [75] aufmerksam gemacht.) D.h. dass z.B. bei Anwendung des Rank transform Tests (RT, ART und INT) bei Varianzanalysen mit Messwiederholungen eine Korrektur der Freiheitsgrade nach Huynh-Feldt oder Greenhouse-Geisser angebracht ist, wie dies von Beasley und Zumbo [53] propagiert wird. Das Ergebnis des Mauchly-Tests auf Sphärität interessiert in dem Zusammenhang nicht, da dessen Voraussetzungen ohnehin kaum erfüllt sein werden. Das Verhalten der Kovarianzmatrizen im Zusammenhang mit Rangtransformationen ist von Bryan [85] ausführlich untersucht worden, ist aber zu speziell, um hier wiedergegeben zu werden.

Verschiedentlich wird auch vorgeschlagen, zum Test eines Messwiederholungsfaktors anstatt der klassischen univariaten Tests einen multivariaten Test, z.B. *Hotellings Spur*, zu verwenden. Hierbei werden zunächst für die K Messwiederholungen x_1, \dots, x_K einer Variablen x $K-1$ Differenzen $d_1 = x_2 - x_1, d_2 = x_3 - x_2, \dots$ errechnet. Der Ausgangshypothese entspricht dann, dass alle diese d_i gleich 0 sind. Dies wird über eine multivariate Varianzanalyse geprüft. Der Vorteil: Diese verlangt nicht die Voraussetzung der Varianzhomogenität (Sphärität). Der Nachteil: Es wird eine multivariate Normalverteilung gefordert, wobei anzumerken ist, dass diese Voraussetzung sehr essentiell ist. (Dieses Prozedere ist auch ausführlich bei Beasley & Zumbo [53] beschrieben.) SPSS gibt übrigens bei Analysen mit Messwiederholungen immer zuerst die Ergebnisse der multivariaten Varianzanalyse aus.

Dann bleibt die nichtparametrische Varianzanalyse für den Fall ordinaler abhängiger Variablen vorbehalten.

Der Vollständigkeit wegen sei noch erwähnt, dass es auch Modelle für Anovas mit Messwiederholungen gibt, die andere Strukturen der Varianz-Kovarianzmatrix als die o.a. Sphärität voraussetzen, so z.B. autoregressive und unstrukturierte. R bietet dafür auch mit der Funktion `gls` im Paket `nlme` Lösungen. Eine leicht verständliche Übersicht bietet [86].

Gute Erläuterungen der Voraussetzungen zu Varianzanalysen bieten der Klassiker B.J. Winer [13] und R.N. Cardinal [83]. Beide gehen jedoch nicht auf Details zur Überprüfung der Normalverteilung ein.

5.3 Die 1-faktorielle Varianzanalyse

Angenommen es liegt ein Faktor A mit K Messwiederholungen vor. Getestet wird die Hypothese gleicher Gruppenmittelwerte bzw. gleicher Abweichungen vom Gesamtmittelwert:

$$\mu_1 = \mu_2 = \dots = \mu_K \quad \text{bzw.} \quad \gamma_1 = \gamma_2 = \dots = \gamma_K = 0$$

5.3.1 Parametrischer Test und Prüfung der Voraussetzung

An den Beispieldaten 4, allerdings hier ohne Berücksichtigung der Gruppenstruktur, soll zum Vergleich die parametrische Varianzanalyse durchgeführt und die Prüfung der Voraussetzungen, Varianzhomogenität und Normalverteilung der Residuen, demonstriert werden.

Zur Berechnung der Residuen gibt es folgende Möglichkeit: Der oder die Messwiederholungsfaktoren C, D,.. werden als Gruppierungsfaktoren gehandhabt. Dazu muss der Datensatz umstrukturiert werden, indem die Messwiederholungen in Fälle gewandelt werden. (Dies ist in R ohnehin für Analysen mit Messwiederholungen erforderlich.) Dann wird folgendes Modell (*ohne* Messwiederholungen) analysiert:

$$C + Vpn \quad \text{bzw.} \\ C*D + Vpn$$

wobei Vpn die Fallkennung, z.B. Versuchspersonennummer, ist. Die Residuen dieses Modells sind die Residuen des Modells mit Messwiederholungen auf C (und D).

Dies ist zwar prinzipiell auch bei SPSS möglich, verursacht aber wegen der erforderlichen Umstrukturierung etwas Aufwand. SPSS gibt allerdings für jede Messwiederholungsvariable x_i andere Residuen aus: $e'_{im} = x_{im} - \gamma_i$. Aus dem Modell 5-1 ergibt sich für diese $e'_{im} = \pi_m + e_{im}$, d.h. um die Residuen e_{im} zu erhalten, müssen von den e'_{im} die π_m subtrahiert werden. Die erforderlichen Schritte sind dann:

- Speichern der Residuen: e'_{im} ,
- Ermitteln des Personeneffekts π_m aus $p_m = \left(\sum_i^I x_{im} \right) / I$ und $\pi_m = (p_m - \bar{p})$,
- und schließlich $e_{im} = e'_{im} - \pi_m$.

(Die Subtraktion von \bar{p} von p_m zur Ermittlung von π_m kann entfallen, da sie für die Beurteilung der Residuen e_{im} ohne Bedeutung ist.)

Für größere n ($n > 20$) können diese e_i für $i=1, \dots, I$ auf Normalverteilung überprüft werden. Die I Testergebnisse, etwa die p-Werte p_1, \dots, p_I können z.B. über *Fishers combined probability test* (vgl. Anhang 2.5) zu einem Testergebnis zusammengefasst werden.

Für kleinere n müssten die I Variablen zu einer mit $N*I$ Werten zusammengefasst werden, entweder per copy & paste oder wieder mittels der aufwändigen Umstrukturierung. Dann sollte aber besser der erste oben beschriebene Weg gewählt werden.

mit R:

- Zunächst erhält der Dataframe `winer518` eine Fallkennzeichnung, hier `vpn` genannt.
- `Geschlecht` und `vpn` müssen als „factor“ deklariert werden.
- Mittels einer der Funktionen `reshape` oder `make.rm` bekommt der Dataframe die für Messwiederholungen erforderliche Struktur, wobei die abhängige Variable den Namen

score und der Faktor den Namen Zeit erhalten.

- Das Ergebnis wird `winer518t` benannt.
- Zeit muss als „factor“ deklariert werden.
- Schließlich wird die Anova mit der „Standardfunktion“ `aov` durchgeführt, wobei durch den Modellterm `Error (Vpn/Zeit)` die Messwiederholungen auf dem Faktor `Zeit` gekennzeichnet werden:

```
Vpn      <- 1:10
winer518 <- cbind(Vpn,winer518)
winer518 <- within(winer518,
  {Geschlecht<-factor(Geschlecht); Vpn<-factor(Vpn)})
winer518t<- reshape(win518,direction="long",timevar="Zeit",
  v.names="score", varying=c("t1","t2","t3"),idvar="Vpn")
winer518t<- within(winer518t, Zeit<-factor(Zeit))
aov1     <- aov(score~Zeit+Error(Vpn/Zeit),winer518t)
summary (aov1)
```

mit dem Ergebnis:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	9	59.87	6.652		
Error: Vpn:Zeit					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Zeit	2	58.07	29.033	7.926	0.0034 **
Residuals	18	65.93	3.663		

Tabelle 5-1

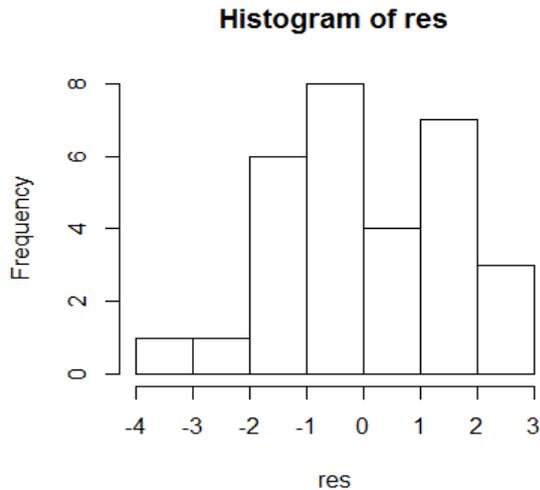
Bei der analogen Benutzung von `make.rm` für die Umstrukturierung werden die Variablen 1 und 2 als „konstante“ Variablen und 3 bis 5 als Messwiederholungsvariablen festgelegt. Die abhängige Variable (default: `repdat`) und Faktor (default: `constrasts`) werden umbenannt in `score` bzw. `Zeit`.

```
Vpn      <- 1:10
winer518 <- cbind(Vpn,winer518)
winer518 <- within(winer518,
  {Geschlecht<-factor(Geschlecht); Vpn<-factor(Vpn)})
winer518t <- make.rm (winer518, 1:2, 3:5)
names(winer518t) [3:4] <- c("score","Zeit")
aov1 <- aov(score~Zeit+Error(Vpn/Zeit),winer518t)
summary (aov1)
```

Für die Prüfung der Voraussetzungen bietet das Ergebnisobjekt `aov1` keine Möglichkeiten. Zunächst einmal zu den Residuen e_{im} . Diese lassen sich, wie oben erläutert, bequem als Residuen eines Anova-Modells ohne Messwiederholungen ermitteln:

```
aov2<-aov(score~Zeit+Vpn, winer518t)
res<-aov2$residuals
hist(res)
shapiro.test(res)
```

mit folgenden Ergebnissen für die Tests auf Normalverteilung:



Shapiro-Wilk normality test

data: res

W = 0.9695, p-value = 0.5255

Das Histogramm zeigt mit einer leichten Rechtsschiefe eine geringe Abweichung von der Normalverteilung, die allerdings nicht als bösartig angesehen werden muss. Diese resultiert zum Teil auch aus der zu großen Intervallzahl. Dahingegen weist der Shapiro-Test keine Abweichung aus.

Zur Überprüfung der Normalverteilung der versuchspersonenspezifischen Abweichungen π_m müssen diese ebenfalls erst ermittelt werden. Dazu muss man auf den ursprünglichen Dataframe `winer518` zurückgreifen und die Summen oder Mittelwerte der Variablen `t1`, `t2` und `t3` berechnen. Diese können dann wie üblich überprüft werden. Die Ergebnisse werden hier wegen der zu geringen Fallzahl ($n=10$) nicht wiedergegeben.

```
pm <- (winer518$t1 + winer518$t2 + winer518$t3)/3
hist(pm)
shapiro.test(pm)
```

Zur Überprüfung der Varianzhomogenität, in diesem Fall also der Sphärizität, findet man die Funktion `mauchly.test`. Ein Versuch, diese auf einen der bislang erzeugten Dataframes oder ein Anova-Ergebnisobjekt anzuwenden, scheitert. Denn diese Funktion erwartet ein `mlm`- oder ein `SSD`-Objekt. Beide sind nur mit erheblichem Aufwand und einigen „linear model“-Kenntnissen zu bekommen.

Wesentlich einfacher ist die Benutzung der Funktion `ezANOVA` aus dem Paket `ez`, bei der Mauchly Test im Fall von Messwiederholungen automatisch ausgegeben wird:

```
library(ez)
ezANOVA (winer518t, score, Vpn, within=Zeit)
```

```
$ANOVA
  Effect DFn DFd      F      p p<.05      ges
2   Zeit    2   18 7.926188 0.003397427 * 0.3158086

$`Mauchly's Test for Sphericity`
  Effect      W      p p<.05
2   Zeit 0.6441534 0.1721699

$`Sphericity Corrections`
  Effect      GGe      p[GG] p[GG]<.05      HFe      p[HF] p[HF]<.05
2   Zeit 0.7375466 0.00859794 * 0.8472485 0.005821856 *
```

Tabelle 5-2

Die ersten Zeilen enthalten die schon oben erzielte Varianzanalyse (vgl. Tabelle 5-1). Anzumerken ist, dass darin „ges“ die *generalized effect size* (Effektgröße η^2) ist. Darunter das Ergebnis des Tests von Mauchly, das keine Signifikanz, und somit Varianzhomogenität zeigt. Die letzten Zeilen bieten für den Fall heterogener Varianzen die beiden alternativen Signifikanzberechnungen für die Varianzanalyse von Geisser & Greenhouse (GG) sowie Huynh & Feldt (HF), jeweils mit dem Zusatz „e“ für den Korrekturfaktor der Freiheitsgrade ϵ bzw. dem Zusatz „p“ für die Irrtumswahrscheinlichkeit.

mit SPSS:

Varianzanalysen mit Messwiederlungen erhält man in SPSS über das Menü „Allgemeines lineares Modell -> Messwiederholung“.

Die Anweisungen für den Beispieldatensatz 4 mit Speicherung der Residuen lauten:

```
GLM t1 t2 t3
  /wsfactor=Zeit 3 polynomial
  /save=resid
  /wsdesign=Zeit.
```

Die Ausgabe umfasst u.a. die zunächst interessierende Varianzanalyse in folgender Tabelle:

Tests der Innersubjekteffekte						
Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Zeit	Sphärizität angenommen	58,067	2	29,033	7,926	,003
	Greenhouse-Geisser	58,067	1,475	39,365	7,926	,009
	Huynh-Feldt	58,067	1,694	34,268	7,926	,006
	Untergrenze	58,067	1,000	58,067	7,926	,020
Fehler(Zeit)	Sphärizität angenommen	65,933	18	3,663		
	Greenhouse-Geisser	65,933	13,276	4,966		
	Huynh-Feldt	65,933	15,250	4,323		
	Untergrenze	65,933	9,000	7,326		

Tabelle 5-3

Die „normale“ Signifikanzüberprüfung für den Faktor *zeit* ist in der Zeile „Sphärität angenommen“ abzulesen. Die beiden Zeilen „Greenhouse-Geisser“ und „Huynh-Feldt“ bieten alternative Tests für den Fall, dass die Voraussetzung der Sphärität, also der Varianzhomogenität, nicht erfüllt ist. Der Mauchly-Test zur Überprüfung dieser Voraussetzung enthält die folgende Tabelle:

Mauchly-Test auf Sphärizität							
Innersubjekteffekt	Mauchly- W	Approximiertes Chi-Quadrat	df	Sig.	Epsilon		
					Greenhouse -Geisser	Huynh -Feldt	Untergrenze
Zeit	,644	3,519	2	,172	,738	,847	,500

aus der hervorgeht, dass die Varianzhomogenität erfüllt ist. Die rechten Spalten „Epsilon“ enthalten den Korrekturfaktor der Freiheitsgrade ϵ für den entsprechenden Test, der in der o.a. Varianzanalysetabelle zur Berechnung der Signifikanzen verwendet wird.

Die Überprüfung der Residuen auf Normalverteilung bei Messwiederholungen ist in SPSS mit etwas Aufwand verbunden. Zum einen gibt es die am Anfang dieses Kapitels beschriebene Möglichkeit über ein varianzanalytisches Modell ohne Messwiederholungen, was aber eine Umstrukturierung des Datensatzes erfordert. Ein Beispiel dazu folgt in Kapitel 6.2. Zum anderen kann man auf den Residuen e'_{im} aufbauen, die SPSS bei Messwiederholungsmodellen ausgibt. Dies soll hier kurz gezeigt werden.

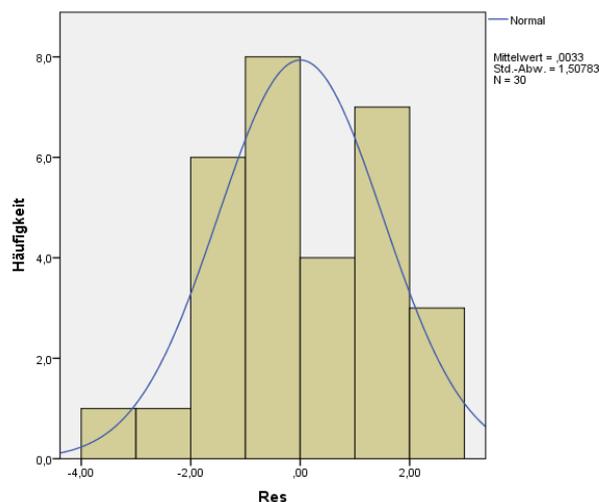
Es wird für jede Messwiederholungsvariable (t_1, t_2, t_3) eine Residuenvariable erzeugt: Res_1, Res_2, Res_3. Von diesen muss nun zunächst der Personeneffekt π_m abgezogen werden, der allerdings vorher noch berechnet werden muss. Nachfolgend die Kommandos hierfür, wobei im zweiten Schritt der Mittelwert von π errechnet wird - hier einfach über Descriptive und Einsetzen des Wertes 4.27, alternativ über Aggregate. Allerdings ist, wie oben bemerkt, die Subtraktion des Mittelwert von π nicht erforderlich.

```
Compute Pi=Mean(t1,t2,t3).
Descriptive Variables=Pi.
Compute R1 = Res_1 - (Pi-4.27).
Compute R2 = Res_2 - (Pi-4.27).
Compute R3 = Res_3 - (Pi-4.27).
```

Bei größeren Stichproben könnte jede dieser Variablen separat auf Normalverteilung überprüft werden, nicht aber bei kleineren wie hier $n=10$. Weder ein Histogramm noch ein Test können hier ein klares Bild geben. Zwei der Möglichkeiten, die Residuenvariablen zu einer einzigen zusammenzufassen, sollen hier kurz skizziert werden.

Zum einen können im Dateneditor über copy & paste sämtliche Residuenvariablen (hier: R1, R2 und R3) zu einer zusammengefügt werden. Dies dürfte, insbesondere bei nicht zu großen Datensätzen, der einfachste Weg sein.

Alternativ wird der Datensatz umstrukturiert, so dass die Messwiederholungen zu Fällen werden, hier also die Variablen R1, R2 und R3 zu einer Variablen Res, deren Werte sich jeweils auf 3 Fälle verteilen. Die Vorgehensweise ist ausführlich im Anhang 1 beschrieben. Die Variable Res kann nun über ein Histogramm oder über den Shapiro-Wilk-Test (erhältlich über das Menü „Deskriptive Statistiken -> Explorative Datenanalyse“ und dort bei „Diagramme“ „Normalverteilungsdiagramm mit Tests“ aktivieren) auf Normalverteilung überprüft werden.



Tests auf Normalverteilung						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Res	,120	30	,200*	,969	30	,526

Das Histogramm zeigt mit einer leichten Rechtsschiefe eine geringe Abweichung von der Normalverteilung, die allerdings nicht als bösartig angesehen werden muss. Diese resultiert zum Teil auch aus der zu großen Intervallzahl. Dahingegen weist der Shapiro-Test keine Abweichung aus.

5.3.2 Der Friedman-Test

Der Friedman-Test ist das nichtparametrische Pendant zur 1-faktoriellen Varianzanalyse mit Messwiederholungen. (Er wird gelegentlich auch irreführend als 2-faktorielle Varianzanalyse bezeichnet, da rein formal neben dem betrachteten Messwiederholungsfaktor noch der „Faktor“ Vpn in die Rechnung einfließt.) Der Algorithmus sieht so aus, dass zunächst innerhalb jeder Vpn die Werte in Ränge 1,...,I (mit I Anzahl der Stufen) transformiert werden, danach mit diesen wie gewohnt weitergerechnet wird, aber zum Schluss anstatt eines F-Tests ein χ^2 -Test durchgeführt wird (vgl. auch Kapitel 2.5 und 5.3.3). An den Beispieldaten 4, allerdings hier ohne Berücksichtigung der Gruppenstruktur, soll die Berechnung gezeigt werden.

mit R:

Die Funktion `friedman.test` kann auf zwei verschiedene Arten benutzt werden:

- zum einen mittels Eingabe der zu analysierenden Datenmatrix (Dataframe `winer518`), allerdings nicht vom Typ „data.frame“, sondern vom Typ „matrix“ (Umwandlung z.B. über `as.matrix`), wobei die Daten die ursprüngliche Struktur haben müssen.
- zum anderen mittels Angabe eines Modells wie in `aov`, wobei die Daten wie für `aov` umstrukturiert sein müssen (Dataframe `winer518t`),

Variante 1:

```
friedman.test (as.matrix(winer518[,3:5]))
```

Variante 2:

```
friedman.test (score~Zeit | Vpn, data=winer518t)
```

Die Ausgabe ist bei beiden natürlich identisch:

```
Friedman rank sum test
Friedman chi-squared = 9.5556, df = 2, p-value = 0.008415
```

mit SPSS:

Hier muss beachtet werden, dass gegebenenfalls vorher das Skalenniveau der analysierten Variablen auf „Skala“ gesetzt wird. Die Syntax für den Friedman-Test:

```
Nptests /related test(t1 t2 t3) friedman(compare=pairwise).
```

mit der Ausgabe:

Übersicht über Hypothesentest

		Nullhypothese	Test	Sig.	Entscheidung
Gesamtanzahl	10	1 Die Verteilungen von , and sind gleich.	Friedmans Zweifach-Rangvarianzanalyse verbundener Stichproben	,008	Nullhypothese ablehnen.
Teststatistik	9,556				
Freiheitsgrade	2				
Asymptotische Sig. (zweiseitiger Test)	,008				

Asymptotische Signifikanzen werden angezeigt. Das Signifikanzniveau ist ,05.

Das Ergebnis ist zwar signifikant. Dennoch soll hier kurz noch die Iman & Davenport-Korrektur gezeigt werden (vgl. Formel 2-1):

$$F = \frac{(10 - 1) \cdot 9,5556}{10 \cdot (3 - 1) - 9,5556} = 8,308$$

Dieser F-Wert hat 2 Zähler-FG und 20 Nenner-FG. Der entsprechende p-Wert: 0.00236, der tatsächlich etwas kleiner ausfällt als der p-Wert des Friedman-Tests.

5.3.3 Verallgemeinerte Kruskal-Wallis- und Friedman-Tests (KWF)

Obwohl das Ergebnis dieses Verfahrens mit dem o.a. Friedman-Test identisch ist und daher, wie oben beschrieben, leicht mit den in R bzw. SPSS verfügbaren Funktionen durchführbar ist, soll dieses Verfahren von Puri & Sen kurz beschrieben werden, da sich damit zeigen lässt, dass es letztlich eine Verallgemeinerung der Tests von Kruskal-Wallis bzw. von Friedman ist.

Folgende Schritte sind durchzuführen:

- Für jede Erhebungseinheit (Versuchsperson) werden die Werte in Ränge (1,...,I) transformiert.
- Mit den Rängen wird eine parametrische Varianzanalyse mit Messwiederholungen durchgeführt.
- Auf Basis der Anova-Tabelle wird folgender χ^2 -Test aufgestellt (vgl. Formel 2-7):

$$\chi^2 = \frac{SS_{Effekt}}{(SS_{Effekt} + SS_{Fehler}) / (df_{Effekt} + df_{Fehler})}$$

wobei SS_{Effekt} die Streuungsquadratsumme (Sum of Squares) des zu testenden Effektes (A), SS_{Fehler} die Streuungsquadratsumme des Fehlers ist sowie df die entsprechenden Freiheitsgrade.

- Der χ^2 -Wert ist dann in den Tafeln für den χ^2 -Test auf Signifikanz zu überprüfen, wobei die Freiheitsgrade die Zählerfreiheitsgrade (df_{Effekt}) des entsprechenden F-Tests sind.
- Schließlich kann noch die Iman & Davenport-Korrektur (Formel 2-1a) angewandt werden, falls der χ^2 -Test nicht signifikant war.

Die Schritte sollen am Datensatz des Beispiels 4 demonstriert werden.

mit R:

Basis ist der oben in Kapitel 5.3.1 erstellte Dataframe `winer518t`. Zuerst wird mittels der Funktion `ave` die Variable `score` für jeden Wert von `vpn` in Ränge `rscore` transformiert.

Der Dataframe wird um diese Variable ergänzt. Für `rscore` wird dann eine Varianzanalyse durchgeführt:

```
rscore <- ave(winer518t$score, winer518t$Vpn, FUN=rank)
winer518t <- cbind(winer518t,rscore)
aovr <- aov(rscore~Zeit+Error(Vpn/Zeit),winer518t)
summary(aovr)
```

Error: Vpn:Zeit						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Zeit	2	8.6	4.300	8.234	0.00289	**
Residuals	18	9.4	0.522			

Hieraus (Spalten „Sum Sq“ bzw. „Df“) wird der χ^2 -Wert errechnet:

$$\chi^2 = 8.6 / ((8.6 + 9.4)/(2+18)) = 8.6/0.9 = 9.555556$$

der mit den o.a. Werten (vgl. Kapitel 5.3.2) übereinstimmt. Es sei darauf hingewiesen, dass in Kapitel 5.4.3 ein R-Funktion für diese Methode vorgestellt wird.

mit SPSS:

- Zunächst müssen für den Datensatz über das Menü „Daten -> Umstrukturieren“ die Messwiederholungen in Fälle transformiert werden (siehe dazu im Anhang 1.1.1).
- Die Variable `score` wird dann über das Menü „Transformieren -> Rangfolge bilden“ in Ränge umgerechnet, wobei im Feld „Sortieren nach“ die Variable `Vpn` eingetragen werden muss, damit die Rangbildung pro Vpn vorgenommen wird. Ergibt Variable `Rscore`.
- Danach muss der Datensatz wieder zurück in das „normale“ Format mit Messwiederholungen transformiert werden (vgl. 1.2). Dabei werden aus `Rscore` wieder 3 Variablen `Rscore.1`, `Rscore.2`, `Rscore.3`.
- Abschließend wird dann eine Varianzanalyse mit Messwiederholungen (Menü: „Allgemeines lineares Modell -> Messwiederholung“) für `Rscore.1`, ... gerechnet.

Die Syntax für den ersten Schritt der Umstrukturierung, der Rangbildung bzw. den zweiten Schritt der Umstrukturierung in der SPSS-Syntax:

```
Varstocases
  /Id=Vpn
  /Make score from t1 t2 t3
  /index=Zeit(3)
  /keep=patients
  /null=keep.

Rank variables=score (A) by Vpn
  /rank.

Sort cases by Vpn Zeit.
casestovars
  /Id=Vpn
  /index=Zeit
  /groupby=variable.

GLM Rscore.1 Rscore.2 Rscore.3
  /WSfactor =Zeit 3 Polynomial
  /WSdesign Zeit.
```

Tests der Innersubjekteffekte						
Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Zeit	Sphärizität angenommen	8,600	2	4,300	8,234	,003
	Greenhouse-Geisser	8,600	1,122	7,666	8,234	,015
	Huynh-Feldt	8,600	1,170	7,349	8,234	,013
	Untergrenze	8,600	1,000	8,600	8,234	,018
Fehler(Zeit)	Sphärizität angenommen	9,400	18	,522		
	Greenhouse-Geisser	9,400	10,097	,931		
	Huynh-Feldt	9,400	10,532	,893		
	Untergrenze	9,400	9,000	1,044		

Hieraus (Spalten „Quadratsumme“ bzw. „df“) wird der χ^2 -Wert errechnet:

$$\chi^2 = 8.6 / ((8.6 + 9.4)/(2+18)) = 8.6/0.9 = 9.555556$$

der mit den o.a. Werten übereinstimmt.

5.3.4 rank transform (RT) und normal scores (INT)

Bei der einfachen *rank transform* (RT) wird lediglich vor der Durchführung der parametrischen Varianzanalyse zunächst die abhängige Variable x über alle Messwiederholungen hinweg in Ränge $R(x)$ transformiert. Beim einfachen *inverse normal transformation* (INT) werden anschließend zusätzlich die Ränge $R(x_i)$ über die inverse Normalverteilung in normal scores umgerechnet:

$$nscore_i = \Phi^{-1}(R(x_i)/(M+1))$$

wobei M die Anzahl aller Werte ist, also $N \cdot I$ (mit N Anzahl der Merkmalsträger und I Anzahl der Messwiederholungen), und Φ^{-1} die Umkehrfunktion der Normalverteilung. Die statistischen Tests bleiben unverändert. Beide Verfahren sind in erster Linie für metrische Variablen gedacht, die die Normalverteilungs-Voraussetzung nicht erfüllen, nicht jedoch für Variablen mit beliebigen Eigenschaften. D.h. hat die untransformierte Variable x ungleiche Varianzen, so kann das auch noch für die transformierte Variable $\text{norm}(R(x))$ gelten. So kann es sinnvoll sein, gegebenenfalls auch $\text{norm}(R(x))$ auf Sphärizität zu überprüfen. Hierfür steht allerdings nur der Mauchly-Test zur Verfügung, der selbst u.a. Normalverteilung voraussetzt, so dass dessen Ergebnisse unter Vorbehalt zu interpretieren sind. Beasley und Zumbo [53] propagieren daher, bei den F-Tests einfach eine der Korrekturen der Freiheitsgrade von Huynh-Feldt oder Greenhouse-Geisser vorzunehmen, ohne das Ergebnis des Mauchly-Tests zu berücksichtigen.

Das INT-Verfahren soll am Datensatz des Beispiels 4 für den Faktor `zeit` demonstriert werden.

mit R:

Ausgangsbasis ist wieder der in Kapitel 5.3.1 erstellte Dataframe `winer518t`. Zunächst wird die Kriteriumsvariable `score` in Ränge (`rscore`) transformiert, anschließend diese in normal scores umgerechnet, wobei die Anzahl der Fälle `nc` einfließt. Die Varianzanalyse wird mit `ezANOVA` (Paket `ez`) durchgeführt, um mittels des Tests von Mauchly die Varianzhomogenität (Sphärizität) die adjustierten Signifikanztests von Geisser-Greenhouse und Huynh-Feldt zu erhalten:

```
library(ez)
nc      <- dim(winer518t)[1]
winer518t <- within(winer518t, rscore<-rank(score))
winer518t <- within(winer518t, nscore<-qnorm(rscore/(nc+1)))
ezANOVA (winer518t, nscore, Vpn, within=Zeit)
```

Nachfolgend die Ergebnisse für das normal score (INT)-Verfahren:

```
$ANOVA
  Effect DFn DFd      F      p p<.05      ges
2   Zeit   2  18 8.570491 0.002427323 * 0.309934

$`Mauchly's Test for Sphericity`
  Effect      W      p p<.05
2   Zeit 0.5617469 0.09957784

$`Sphericity Corrections`
  Effect      GGe      p [GG]      HFe      p [HF]
2   Zeit 0.6952879 0.007838653 0.7823034 0.00559601
```

Die Varianzhomogenität ist zwar erfüllt ($p=0.100$). Dennoch liest man zweckmäßigerweise das Ergebnis für den Zeit-Effekt nicht im oberen ANOVA-Teil ($p=0.002$), sondern im unteren Sphericity Corrections ($p_{[HF]}$) ab ($p=0.0056$) ab.

mit SPSS:

Wie in Kapitel 5.3.3 sind die folgenden Schritte erforderlich, um die Werte über die Messwiederholungen hinweg in Ränge transformieren zu können:

- Zunächst müssen für den Datensatz über das Menü „Daten -> Umstrukturieren“ die Messwiederholungen in Fälle transformiert werden (siehe dazu im Anhang 1.1.1).
- Die Variable `score` wird dann über das Menü „Transformieren -> Rangfolge bilden“ in Ränge umgerechnet. Ergibt Variable `Rscore`.
- Diese Variable `Rscore` wird nun in normal scores umgerechnet. Dazu muss noch vorab über `Aggregate` die Anzahl der Werte `nc` ermittelt werden, da die Ränge durch $(n+1)$ dividiert werden. Die Ergebnisvariable wird `nscore` genannt.
- Danach muss der Datensatz wieder zurück in das „normale“ Format mit Messwiederholungen transformiert werden (vgl. 1.2). Dabei werden aus `nscore` wieder 3 Variablen `nscore.1`, `nscore.2`, `nscore.3`.
- Abschließend wird dann eine Varianzanalyse mit Messwiederholungen (Menü: „Allgemeines lineares Modell -> Messwiederholung“) für `nscore.1, ..` gerechnet.

Nachfolgend die Syntax für diese Schritte:

```
Varstocases
  /Id=Vpn
  /Make score from t1 t2 t3
  /index=Zeit(3)
  /keep=Geschlecht
  /null=keep.

Aggregate
  /outfile=* mode=addvariables
  /break= /nc=NU(score).
```

```

Rank Variables=score / rank into Rscore.
compute nscore=Idf.normal(Rscore/(nc+1),0,1).

Sort cases by Vpn Zeit.
casestovars
  /Id=Vpn
  /index=Zeit
  /groupby=variable.

GLM nscore.1 nscore.2 nscore.3
  /WSfactor = Zeit 3 Polynomial
  /WSdesign Zeit

```

Nachfolgend zunächst der Test auf Varianzhomogenität, der zwar mit $p=0,100$ gerade noch akzeptabel ist, aber ohnehin keine Rolle spielen sollte. Denn zweckmäßigerweise sollten die Ergebnisse für die Varianzanalyse (in der zweiten Tabelle) ohnehin einer der Zeilen mit den adjustierten Testergebnissen, z.B. Huynh-Feldt, entnommen werden.

Mauchly-Test auf Sphärizität ^a						
Innersubjekt- effekt	Mauchly-W	Approximiertes Chi-Quadrat	df	Sig.	Epsilon ^b	
					Greenhouse- Geisser	Huynh-Feldt
Zeit	,562	4,614	2	,100	,695	,782

Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Zeit	Sphärizität angenommen	6,909	2	3,454	8,570	,002
	Greenhouse-Geisser	6,909	1,391	4,968	8,570	,008
	Huynh-Feldt	6,909	1,565	4,416	8,570	,006
	Untergrenze	6,909	1,000	6,909	8,570	,017
Fehler(Zeit)	Sphärizität angenommen	7,255	18	,403		
	Greenhouse-Geisser	7,255	12,515	,580		
	Huynh-Feldt	7,255	14,081	,515		
	Untergrenze	7,255	9,000	,806		

Danach ist der Zeit-Effekt mit $p=0,006$ signifikant.

5.3.5 van der Waerden

Bei dem Verfahren von *van der Waerden* werden anstatt der „klassischen“ F-Tests die χ^2 -Tests des Kruskal-Wallis-Tests bzw. wie bei der o.a. KWF-Methode gerechnet. Allerdings wird eine andere Transformation in Ränge vorgenommen als beim o.a. INT-Verfahren: Wie beim Friedman- bzw. allgemein beim KWF-Verfahren werden die Ränge $1, \dots, I$ fallweise vergeben.

Folgende Schritte sind durchzuführen:

- Für jede Erhebungseinheit (Versuchsperson) werden die Werte in Ränge $(1, \dots, I)$ transformiert.
- Die Ränge werden in normal scores umgerechnet (vgl. Formel 2-2):

$$nscore_i = \Phi^{-1}(R(x_i)/(I+1))$$
- Mit diesen wird eine parametrische Varianzanalyse mit Messwiederholungen durchgeführt.

- Auf Basis der Anova-Tabelle wird folgender χ^2 -Test aufgestellt (vgl. Formel 2-7):

$$\chi^2 = \frac{SS_{\text{Effekt}}}{(SS_{\text{Effekt}} + SS_{\text{Fehler}}) / (df_{\text{Effekt}} + df_{\text{Fehler}})}$$

wobei SS_{Effekt} die Streuungsquadratsumme (Sum of Squares) des zu testenden Effektes, SS_{Fehler} die Streuungsquadratsumme des Fehlers ist sowie df die entsprechenden Freiheitsgrade.

- Der χ^2 -Wert ist dann in den Tafeln für den χ^2 -Test auf Signifikanz zu überprüfen, wobei die Freiheitsgrade die Zählerfreiheitsgrade (df_{Effekt}) des entsprechenden F-Tests sind.

Die Schritte sollen am Datensatz des Beispiels 4 für den Faktor `zeit` demonstriert werden. Da Überprüfung der Spherizität kann entfallen, da hier χ^2 - anstatt F-Tests durchgeführt werden.

mit R:

Ausgangsbasis ist wieder der in Kapitel 5.3.1 erstellte Dataframe `winer518t`. Zuerst wird mittels der Funktion `ave` die Variable `score` für jeden Wert von `vpn` in Ränge `rscore` transformiert, diese anschließend in normal scores `nscore` umgerechnet. Der Dataframe wird um diese Variablen ergänzt. Für `nscore` wird dann eine Varianzanalyse durchgeführt:

```
rscore <- ave(winer518t$score, winer518t$vpn, FUN=rank)
nscore <- qnorm(rscore/4)
winer518t <- cbind(winer518t, rscore, nscore)
summary(aov(nscore~Zeit+Error(vpn/Zeit), winer518t))
```

Error: v _{pn} :Zeit					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Zeit	2	3.847	1.9237	8.163	0.003 **
Residuals	18	4.242	0.2357		

Der χ^2 -Wert des van der Waerden-Tests, der 2 FG hat, errechnet sich nun leicht per Hand:

$$\chi^2 = 3.847 / ((3.847 + 4.242)/(2+18)) = 3.85 / 0.4045 = 9.52$$

Dieser ist auf dem 1%-Niveau signifikant (kritischer Wert: 9.2).

mit SPSS:

Wie im Kapitel 5.3.3 sind die folgenden Schritte erforderlich, um fallweise die Werte in Ränge transformieren zu können:

- Zunächst müssen für den Datensatz über das Menü „Daten -> Umstrukturieren“ die Messwiederholungen in Fälle transformiert werden (siehe dazu im Anhang 1.1.1).
- Die Variable `score` wird dann über das Menü „Transformieren -> Rangfolge bilden“ in Ränge umgerechnet, wobei im Feld „Sortieren nach“ die Variable `vpn` eingetragen werden muss, damit die Rangbildung pro `vpn` vorgenommen wird. Ergibt Variable `rscore`.
- Diese Variable `rscore` wird nun in normal scores umgerechnet. Dabei werden die Ränge durch $(I+1)$, hier also 4, dividiert. Die Ergebnisvariable wird `nscore` genannt.
- Danach muss der Datensatz wieder zurück in das „normale“ Format mit Messwiederholungen transformiert werden (vgl. 1.2). Dabei werden aus `nscore` wieder 3 Variablen `nscore.1`, `nscore.2`, `nscore.3`.

- Abschließend wird dann eine Varianzanalyse mit Messwiederholungen (Menü: „Allgemeines lineares Modell -> Messwiederholung“) für `nscore.1, . .` gerechnet.

Nachfolgend die Syntax für diese Schritte sowie die Anova-Tabelle:

```
Varstocases
  /Id=Vpn
  /Make score from t1 t2 t3
  /index=Zeit(3)
  /keep=Geschlecht
  /null=keep.
```

```
Rank Variables=score by Vpn / rank into Rscore.
compute nscore=Idf.normal(Rscore/4,0,1).
```

```
Sort cases by Vpn Zeit.
casestovars
  /Id=Vpn
  /index=Zeit
  /groupby=variable.
```

```
GLM nscore.1 nscore.2 nscore.3
  /WSfactor =Zeit 3 Polynomial
  /WSdesign Zeit
```

Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Zeit	Sphärizität angenommen	3,847	2	1,924	8,163	,003
	Greenhouse-Geisser	3,847	1,118	3,441	8,163	,015
	Huynh-Feldt	3,847	1,165	3,303	8,163	,014
	Untergrenze	3,847	1,000	3,847	8,163	,019
Fehler(Zeit)	Sphärizität angenommen	4,242	18	,236		
	Greenhouse-Geisser	4,242	10,062	,422		
	Huynh-Feldt	4,242	10,482	,405		
	Untergrenze	4,242	9,000	,471		

Hieraus (Spalten „Quadratsumme“ bzw. „df“) wird der χ^2 -Wert des van der Waerden-Tests errechnet, der 2 FG hat:

$$\chi^2 = 3.847 / ((3.847 + 4.242)/(2+18)) = 3.85 / 0.4045 = 9.52$$

Dieser ist auf dem 1%-Niveau signifikant (kritischer Wert: 9.2).

5.3.6 ATS-Tests von Akritas, Arnold & Brunner

Den von Akritas, Arnold und Brunner entwickelten ATS-Test gibt es auch für Varianzanalysen mit Messwiederholungen. Während in R dazu das Paket `nparLD` zur Verfügung steht, gibt es in SPSS derzeit keine Möglichkeit zur Anwendung dieses Verfahrens.

mit R:

Die 1-faktorielle Analyse mittels `nparLD` soll am Datensatz des Beispiels 4 gezeigt werden. Ausgangsbasis ist wieder der oben erstellte Dataframe `winer518t`. Die Analyse kann mittels zwei Funktionen erfolgen:

- `nparLD` ist eine universelle Funktion für alle verarbeitbaren Designs.
- `ld.f1` erlaubt fehlende Werte bei den Messwiederholungen, gibt einen Mittelwertplot aus sowie eine Reihe weiterer, hier allerdings nicht interessierender Statistiken aus.

Beide geben sowohl die WTS als auch die interessantere ATS aus. Die Ausgabe unterscheidet sich nicht hinsichtlich der Ausgabe dieser Statistiken. Nachfolgend zunächst die Eingabe für beide Varianten, wobei zu beachten ist, dass bei `nparLD` trotz Angabe des Dataframes die Variablennamen nicht automatisch gefunden werden. Daher muss bei beiden Funktionen entweder jeder Variablenname zusammen mit dem Dataframe-Namen in der üblichen Form, z.B. `winer518t$score` angegeben werden oder ein `attach(Dataframe)` ausgeführt werden:

```
library(nparLD)
attach(winer518t)
nparLD(score~Zeit, winer518t, Vpn)
ld.f1(score, Zeit, Vpn, time.name="Zeit")
```

Bei `ld.f1` muss die Variable zweimal angegeben werden: zum einen zur Identifikation des Faktors, zum anderen in “...” als Name des Faktors für die Ausgabe.

Nachfolgend die Ausgabe von `nparLD`, die die Signifikanz des Friedman-Tests bestätigt:

```
Call:
score ~ Zeit

Wald-Type Statistic (WTS):
      Statistic df      p-value
Zeit  43.42399  2 3.720494e-10

ANOVA-Type Statistic (ATS):
      Statistic      df      p-value
Zeit   8.369437 1.433543 0.001127567
```

5.3.7 Quade-Test

Das Verfahren von *Quade* war in Kapitel 2.10.2 skizziert worden. An den Beispieldaten 4, allerdings hier ohne Berücksichtigung der Gruppenstruktur, soll die Berechnung gezeigt werden. R bietet dazu die Funktion `quade.test`.

mit R:

Nachfolgend die Ein- und Ausgabe. Eine Umstrukturierung ist wie bei der Friedman-Analyse nicht erforderlich:

```
quade.test(as.matrix(winer518[,3:5]))
```

```
Quade test

data:  as.matrix(winer518[, 3:5])
Quade F = 6.2019, num df = 2, denom df = 18, p-value = 0.008935
```

Das Ergebnis bestätigt allerdings nicht, dass der Quade-Test bei kleinerer Anzahl von Messwiederholungen stärker ist als der Friedman-Test ($p=0,0084$).

5.3.8 Skillings-Mack-Test

Das Verfahren von *Skillings & Mack* war in Kapitel 2.10.3 erwähnt worden. An den Beispieldaten 4, allerdings hier ohne Berücksichtigung der Gruppenstruktur, soll die Berechnung gezeigt werden. R bietet dazu die Funktion `SkiMack` im Paket `Skillings.Mack`.

mit R:

Nachfolgend die Ein- und Ausgabe (auszugsweise). Eine Umstrukturierung ist wie bei der Friedman-Analyse nicht erforderlich:

```
library(Skillings.Mack)
SkiMack(as.matrix(winer518[,3:5]))
```

```
Skillings-Mack Statistic = 13.545455 , p-value = 0.139438
Note: the p-value is based on the chi-squared distribution with df=9
```

Dass dieser Test hier schlechter als der Friedman-Test abschneidet, ist höchstwahrscheinlich den Bindungen zuzuschreiben.

5.4 Die 2-faktorielle Varianzanalyse

Mit der 2-faktoriellen Varianzanalyse mit Messwiederholungen ist hier ein Design ohne Gruppierungsfaktoren, ausschließlich mit zwei Messwiederholungsfaktoren gemeint, hier mit C und D bezeichnet, jeweils mit *I* bzw. *J* Stufen. Sie unterscheidet sich zunächst einmal, anders als bei Analysen ohne Messwiederholungen, nicht entscheidend von der 1-faktoriellen. Auch hier nimmt man in der Regel den Umweg über die parametrische Analyse, auch wenn der Friedman-Test häufig irreführend als 2-faktorielle Analyse bezeichnet wird.

Während für die Analysen mit R ohnehin die Datenmatrix umstrukturiert werden muss und für die nichtparametrischen Tests kein gesonderter Aufwand entsteht, muss zur Rangberechnung an dieser Stelle auch in SPSS eine solche Umstrukturierung vorgenommen werden.

5.4.1 Das parametrische Verfahren und Prüfung der Voraussetzungen

Auch hier soll zunächst einmal zum Vergleich die parametrische Varianzanalyse durchgeführt werden, und zwar anhand der Beispieldaten 5, jedoch ohne Berücksichtigung der Gruppeneinteilung in Männer und Frauen.

Im Gegensatz zum Datensatz 4 (`winer518`) aus dem letzten Kapitel zeigt hier Mauchlys Test signifikante Abweichungen von der Sphärität. Für jeden der drei Tests C, D und C*D (im Beispiel: `Medikament`, `Aufgabe` und Wechselwirkung) wird die dafür relevante Sphärität überprüft. Da sowohl für `Medikament` als auch für die Wechselwirkung Mauchlys Test signifikant ist, sollten anstatt des „normalen“ F-Tests die Approximationen von Geisser & Greenhouse oder von Huynh & Feldt verwendet werden. Entscheidet man sich für letztere, so erhält man aus den Tabellen 5-5 (R) bzw. 5-6 (SPSS) für den Medikamenten-Effekt einen p-Wert, der nur geringfügig über dem „normalen“ liegt. Für den Interaktionseffekt bedeutet dies jedoch den Verlust der Signifikanz, da der p-Wert des „normalen“ Tests 0,023 beträgt gegenüber einem $p=0,058$ für die Huynh & Feldt-Approximation.

mit R:

Zunächst einmal muss der Dataframe `mydata5` mit der Funktion `reshape` umstrukturiert

werden (Ziel: `mydata5t`). Während `mydata5` zwei Messwiederholungsfaktoren beinhaltet, kann `reshape` nur einen verarbeiten. Die Funktion muss daher zweimal aufgerufen werden:

- Beim ersten `reshape`-Aufruf werden die Stufen des Faktors `Medikament` in Zeilen gewandelt, während die Stufen des Faktors `Aufgaben` als Variablen behandelt werden. Die umzustrukturierenden Variablen `v1, ., v9` können einfach durch die lfd Nummer 1:9 angegeben werden. Die neuen abhängigen Variablen werden `a1, a2, a3` genannt.
- Beim zweiten `reshape`-Aufruf wird dann der Faktor `Aufgaben` umstrukturiert. Allerdings darf dann `vpn` nicht mehr als ID-Variable spezifiziert werden, da die `Vpn`-Werte nach dem ersten Aufruf von `reshape` mehrfach vorkommen und daher nicht zur Identifikation herangezogen werden können. (Dadurch wird eine neue ID-Variable `id` angefügt, die aber ignoriert werden kann.) Die neue abhängige Variable wird `Fehler` genannt. Über den Parameter `times=1:3` werden die Werte des Faktors (`Medikament` bzw. `Aufgabe`) festgelegt. Abschließend müssen noch die beiden Variablen `Medikament` und `Aufgabe` vom Typ „factor“ deklariert werden.

```
Vpn      <- 1:8
mydata5  <- cbind(Vpn, mydata5)
names(mydata5)[2] <- "Geschlecht"
mydata5  <- within(mydata5,
  {Vpn<-factor(Vpn); Geschlecht<-factor(Geschlecht)})
mydata5a <- reshape((mydata5, direction="long", varying=3:11, idvar="Vpn",
  timevar="Medikament", times=1:3, v.names=c("a1", "a2", "a3")))
mydata5b <- reshape((mydata5a, direction="long",
  varying=c("a1", "a2", "a3"), idvar="Vpn",
  timevar="Aufgabe", times=1:3, v.names="Fehler")
mydata5t <- within(mydata5b, {Medikament<-factor(Medikament);
  Aufgabe<-factor(Aufgabe)})
```

Nach dem ersten Aufruf von `reshape` hat der Dataframe folgende Struktur:

	Vpn	Geschlecht	Medikament	a1	a2	a3
1.1	1	1	1	3	3	1
2.1	2	1	1	2	0	0
3.1	3	1	1	5	4	3
4.1	4	1	1	3	5	2
5.1	5	2	1	2	2	1
6.1	6	2	1	4	1	0
7.1	7	2	1	3	2	1
8.1	8	2	1	1	3	0
1.2	1	1	2	4	4	2
2.2	2	1	2	3	2	2
...

und nach dem zweiten Aufruf von `reshape` :

	Vpn	Geschlecht	Medikament	Aufgabe	Fehler	id
1.1	1	1	1	1	3	1
2.1	2	1	1	1	2	2
3.1	3	1	1	1	5	3
4.1	4	1	1	1	3	4
5.1	5	2	1	1	2	5
6.1	6	2	1	1	4	6
7.1	7	2	1	1	3	7
8.1	8	2	1	1	1	8
9.1	1	1	2	1	4	9
10.1	2	1	2	1	3	10

Dieser Dataframe ist nun für Varianzanalysen mit Messwiederholungen geeignet. Diese wird nun zunächst wieder mit `aov` durchgeführt, wobei jetzt zwei Messwiederholungsfaktoren zu berücksichtigen sind. Beide sind für den Error-Term als eingebettet in `Vpn` zu deklarieren, wobei die Klammern dringend erforderlich sind:

```
aov1 <- aov (Fehler ~ Medikament*Aufgabe
            + Error(Vpn/(Medikament*Aufgabe)), mydata5t)
```

Die Ausgabe wirkt auf den ersten Blick etwas unübersichtlich, da jeder Effekt einen eigenen Fehlerterm (Residuals) besitzt.

```
Error: Vpn
      Df Sum Sq Mean Sq F value Pr(>F)
Residuals  7  32.65   4.665

Error: Vpn:Medikament
      Df Sum Sq Mean Sq F value Pr(>F)
Medikament  2 27.444  13.722  20.83 6.37e-05 ***
Residuals 14   9.222   0.659

Error: Vpn:Aufgabe
      Df Sum Sq Mean Sq F value Pr(>F)
Aufgabe   2  40.78  20.389  20.55 6.83e-05 ***
Residuals 14  13.89   0.992

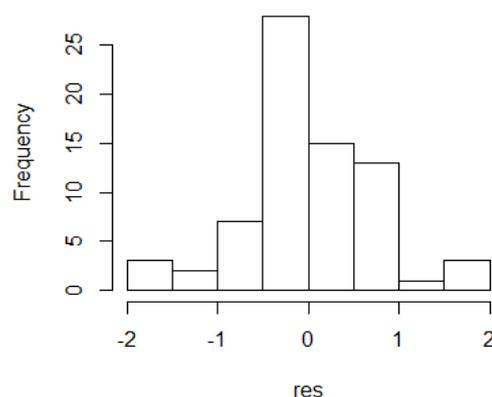
Error: Vpn:Medikament:Aufgabe
      Df Sum Sq Mean Sq F value Pr(>F)
Medikament:Aufgabe  4   6.056   1.5139   3.361 0.0229 *
Residuals          28 12.611   0.4504
```

Tabelle 5-4

Das Ergebnis: Sowohl zwischen den beiden Medikamenten bzw. der Kontrollmessung als auch zwischen den drei Aufgaben bestehen hinsichtlich der Bearbeitung der Aufgaben (Fehlerzahl) signifikante Unterschiede. Hinzu kommt eine signifikante Wechselwirkung beider Faktoren. Auf Details der Interpretation soll hier nicht eingegangen werden.

Die Prüfung der Voraussetzungen erfolgt wie bei der 1-faktoriellen Analyse (vgl. Kapitel 5.3.1). Die Residuen erhält man über folgendes Anova-Modell, das auch auf dem zuletzt erstellten Dataframe `mydata5b` aufsetzt. Diese können dann wie üblich betrachtet werden:

```
aov2 <- aov (Fehler ~ Medikament*Aufgabe + Vpn, mydata5t)
res <- aov2$residuals
hist(res)
```



Die versuchspersonenspezifische Abweichungen π_m basieren auf dem ursprünglichen Dataframe `mydata5`. Für den Test auf Normalverteilung genügt es, die Personenmittelwerte der 9 abhängigen Variablen zu betrachten, die bequem mittels `rowMeans` errechnet werden können. Auf die Ausgabe wird hier verzichtet:

```
hist(rowMeans(mydata5[,3:11]))
```

Die Varianzhomogenität bzw. Sphärität wird wieder mit der Funktion `ezANOVA` des Pakets `ez` geprüft. Die Spezifikation des Modells ist damit deutlich einfacher:

```
library(ez)
ezANOVA(mydata5t, Fehler, Vpn, within=. (Medikament, Aufgabe))
```

Das Ergebnis, das hinsichtlich der Tests auf Sphärität bereits oben interpretiert wurde:

	Effect	DFn	DFd	F	p	p < .05	ges
2	Medikament	2	14	20.831325	6.367208e-05	*	0.28641832
3	Aufgabe	2	14	20.552000	6.833046e-05	*	0.37358443
4	Medikament:Aufgabe	4	28	3.361233	2.286928e-02	*	0.08135846

\$`Mauchly's Test for Sphericity`							
	Effect	W	p	p < .05			
2	Medikament	0.35012339	0.04292036	*			
3	Aufgabe	0.86860800	0.65534724				
4	Medikament:Aufgabe	0.02042957	0.01630533	*			

\$`Sphericity Corrections`							
	Effect	GGe	p [GG]	HFe	p [HF]		
p [HF] < .05							
2	Medikament	0.6061059	0.0011688272	0.6649945	7.533244e-04		
3	Aufgabe	0.8838670	0.0001589182	1.1602880	6.833046e-05		
4	Medikament:Aufgabe	0.4258173	0.0752372276	0.5487419	5.794030e-02		

Tabelle 5-5

mit SPSS:

Die Spezifikation für die Syntax (mit Speicherung der 9 Residuenvariablen) ist relativ einfach:

```
GLM v1 v2 v3 v4 v5 v6 v7 v8 v9
  /wsfactor=Medikament 3 polynomial Aufgabe 3 polynomial
  /save=resid
  /wsdesign=Medikament Aufgabe Medikament*Aufgabe.
```

Mit folgenden relevanten Tabellen: des Mauchly-Tests und der Anova-Tabelle:

Mauchly-Test auf Sphärität							
Innersubjekteffekt	Mauchly-W	Approximiertes Chi-Quadrat	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Untergrenze
Medikament	,350	6,297	2	,043	,606	,665	,500
Aufgabe	,869	,845	2	,655	,884	1,000	,500
Medikament * Aufgabe	,020	21,075	9	,016	,426	,549	,250

Tabelle 5-7

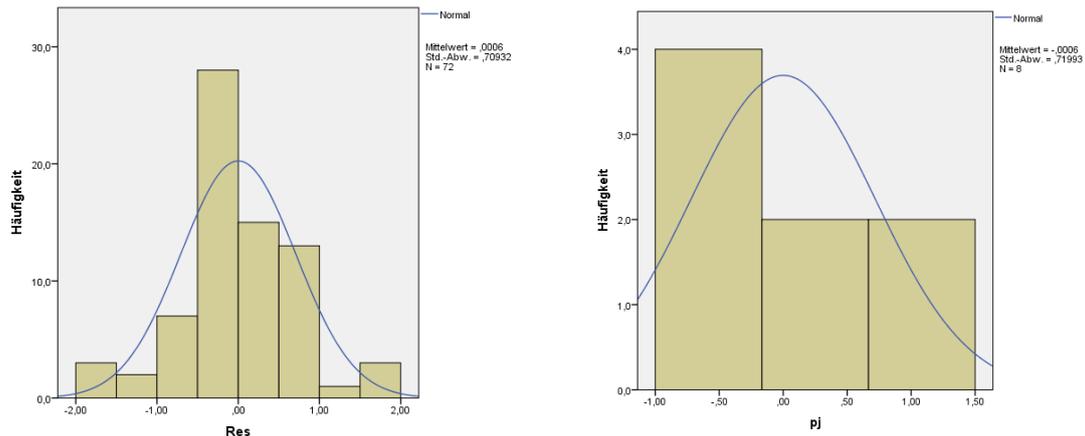
Tests der Innersubjekteffekte						
Quelle		Quadratsumme	df	Mittel der Quadrate	F	Sig.
Medikament	Sphärizität angenommen	27,444	2	13,722	20,831	,000
	Greenhouse-Geisser	27,444	1,212	22,640	20,831	,001
	Huynh-Feldt	27,444	1,330	20,635	20,831	,001
	Untergrenze	27,444	1,000	27,444	20,831	,003
Fehler (Medikament)	Sphärizität angenommen	9,222	14	,659		
	Greenhouse-Geisser	9,222	8,485	1,087		
	Huynh-Feldt	9,222	9,310	,991		
	Untergrenze	9,222	7,000	1,317		
Aufgabe	Sphärizität angenommen	40,778	2	20,389	20,552	,000
	Greenhouse-Geisser	40,778	1,768	23,068	20,552	,000
	Huynh-Feldt	40,778	2,000	20,389	20,552	,000
	Untergrenze	40,778	1,000	40,778	20,552	,003
Fehler (Aufgabe)	Sphärizität angenommen	13,889	14	,992		
	Greenhouse-Geisser	13,889	12,374	1,122		
	Huynh-Feldt	13,889	14,000	,992		
	Untergrenze	13,889	7,000	1,984		
Medikament * Aufgabe	Sphärizität angenommen	6,056	4	1,514	3,361	,023
	Greenhouse-Geisser	6,056	1,703	3,555	3,361	,075
	Huynh-Feldt	6,056	2,195	2,759	3,361	,058
	Untergrenze	6,056	1,000	6,056	3,361	,109
Fehler (Medikmt* Aufgabe)	Sphärizität angenommen	12,611	28	,450		
	Greenhouse-Geisser	12,611	11,923	1,058		
	Huynh-Feldt	12,611	15,365	,821		
	Untergrenze	12,611	7,000	1,802		

Tabelle 5-6

Das Ergebnis des Mauchly-Tests und dessen Konsequenzen wurden bereits am Anfang dieses Kapitels erörtert. Werden die 9 Residuenvariablen zu einer zusammengefasst, erhält man für die Überprüfung auf Normalverteilung ein Ergebnis, das keine bedeutsamen Abweichungen erkennen lässt:

Tests auf Normalverteilung						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Residuen	,130	72	,004	,968	72	,063

Besser ist es aber, wie in Kapitel 5.3.1. demonstriert, vorher von den Residuen den Versuchspersoneneffekt abzuziehen. Der Shapiro-Wilk-Test ergibt dann ein $p=0,173$. Unten links das dazugehörige Histogramm, unten rechts das Histogramm für die π_m , das allerdings bei $n=8$ kaum Aussagefähigkeit hat und daher i.a. entfallen kann:



5.4.2 Rank transform-Tests (RT) und normal scores -Tests (INT)

Bei den einfachen Rank transform Tests wird lediglich vor der Durchführung der parametrischen Varianzanalyse die abhängige Variable über alle Werte (Fälle und Messwiederholungen) hinweg in Ränge transformiert. Die statistischen Tests bleiben unverändert. Dieses Verfahren von Conover & Iman [6] ist in erster Linie für metrische Variablen gedacht, die die Normalverteilungs-Voraussetzung nicht erfüllen, nicht jedoch für Variablen mit beliebigen Eigenschaften. Insofern sollte die Möglichkeit genutzt werden, für die rangtransformierten Daten anstatt des normalen F-Tests die Variante von Huynh & Feldt anzuwenden.

Das INT-Verfahren unterscheidet sich von dem RT-Verfahren nur marginal: Die Ränge $R(x_i)$ werden noch in normal scores (INT) umgerechnet:

$$nscore_i = \Phi^{-1}(R(x_i)/(M+1))$$

wobei M die Anzahl aller Werte ist, also $N*I*J$ (mit N Anzahl der Merkmalsträger und I und J Anzahl der Messwiederholungen der Faktoren C und D), sowie Φ^{-1} die Umkehrfunktion der Normalverteilung.

Bei den Ergebnissen zeigt sich, dass durch die INT-Transformation - im Gegensatz zur RT-Transformation - die Varianzhomogenität nicht beseitigt werden konnte. Aber auf der anderen Seite sind die Ergebnisse qualitativ dieselben, wenn man bei der normal scores-Analyse die Huynh-Feldt-adjustierten F-Tests wählt.

Da die Ausgabe bei beiden Verfahren dieselbe ist, werden die Ergebnistabellen (etwas verkürzt) lediglich einmal in der (leichter lesbaren) Version von SPSS wiedergegeben.

mit R:

Um die Sphärität prüfen zu können bzw. die adjustierten F-Tests zu erhalten, wird die Varianzanalyse mit `ezANOVA` (Paket `ez`) durchgeführt. Ausgehend vom oben (Kapitel 5.4.1) erstellten Dataframe `mydata5t` sind folgende Anweisungen erforderlich:

```
library(ez)
RFehler <- rank(mydata5t$Fehler)
mydata5t <- cbind(mydata5t, RFehler)
ezANOVA(mydata5t, RFehler, Vpn, within=.(Medikament, Aufgabe))
```

Da alle drei Mauchly-Tests nicht signifikant sind, kann die Anova-Tabelle („ANOVA“) herangezogen werden, deren Ergebnisse zum Teil (Medikament und Interaktion) sogar bes-

ser sind, als bei der „rein parametrischen“ unter Verwendung der Huynh & Feldt-Approximationen (vgl. Tabelle 5-5).

Für die Berechnung der normal scores sowie deren Varianzanalyse sind die o.a. Anweisungen zu ergänzen:

```
nc          <- dim(mydata5t) [1]
mydata5t <- within(mydata5t, nsFehler<-qnorm(RFehler/(nc+1)))
ezANOVA (mydata5t, nsFehler, Vpn, within=.(Medikament,Aufgabe))
```

mit SPSS:

- Zunächst müssen für den Datensatz über das Menü „Daten -> Umstrukturieren“ die Messwiederholungen in Fälle transformiert werden (siehe dazu Anhang 1.1.2).
- Die Variable `Fehler` wird dann über das Menü „Transformieren -> Rangfolge bilden“ in Ränge umgerechnet.
- Danach muss der Datensatz wieder zurück in das „normale“ Format mit Messwiederholungen transformiert werden (vgl. Anhang 1.2).
- Abschließend wird dann eine Varianzanalyse mit Messwiederholungen (Menü: „Allgemeines lineares Modell -> Messwiederholung“) für die Variablen `RFehler.1.1`, `RFehler.1.2`, ..., `RFehler.3.3` gerechnet, die bei der Umstrukturierung gebildet werden:

Die Syntax für den ersten Schritt der Umstrukturierung, der Rangbildung bzw. des zweiten Schritts der Umstrukturierung in der SPSS-Syntax:

```
Varstocases
  /Id=Vpn
  /make Fehler from v1 v2 v3 v4 v5 v6 v7 v8 v9
  /index=Medikament(3) Aufgabe(3)
  /keep=Geschlecht
  /null=keep.

Rank variables=Fehler (A)
  /rank into RFehler.

Sort cases by Vpn Medikament Aufgabe.
Casestovars
  /Id=Vpn
  /index=Medikament Aufgabe
  /groupby=variable.

GLM RFehler.1.1 RFehler.1.2 RFehler.1.3 RFehler.2.1 RFehler.2.2
    RFehler.2.3 RFehler.3.1 RFehler.3.2 RFehler.3.3
  /WSfactor=Medikament 3 Polynomial Aufgabe 3 Polynomial
  /WSdesign=Medikament Aufgabe Medikament*Aufgabe.
```

Für die Errechnung der normal scores muss die Rank-Anweisung durch die folgenden ersetzt werden:

```
Aggregate
  /outfile=* mode=addvariables
  /break= /nc=NU(Fehler).
Rank Variables=Fehler / rank into RFehler.
compute nFehler=Idf.normal(RFehler/nc,0,1).
```

Und in den GLM-Anweisungen ist entsprechend `RFehler...` durch `nFehler` zu ersetzen.

Hier nun die Ergebnisse in der Version von SPSS:
Zunächst für das RT-Verfahren, und zwar der Mauchly-Test:

Mauchly-Test auf Sphärizität							
Innersubjekteffekt	Mauchly W	Approx. Chi-Quadrat	df	Sig.	Epsilon		
					Greenhouse -Geisser	Huynh Feldt	Unter grenze
Medikament	,470	4,524	2	,104	,654	,743	,500
Aufgabe	,922	,485	2	,785	,928	1,000	,500
Medikament * Aufgabe	,070	14,377	9	,125	,490	,679	,250

sowie das Ergebnis für die Varianzanalyse auf Basis der Rangtransformation, bei dem wegen der für alle drei Tests gegebenen Sphärizität die jeweils erste Zeile genommen werden kann. Die Ergebnisse sind zum Teil (Medikament und Interaktionen) sogar besser sind, als bei der „rein parametrischen“ unter Verwendung der Huynh-Feldt-Approximationen (vgl. Tabelle 5-6):

Tests der Innersubjekteffekte						
Quelle		Quadrat summe vom Typ III	df	Mittel der Quadrate	F	Sig.
Medikament	Sphärizität angen.	5419,083	2	2709,542	21,880	,000
	Greenhouse-Geisser	5419,083	1,308	4144,310	21,880	,001
	Huynh-Feldt	5419,083	1,486	3645,959	21,880	,000
	Untergrenze	5419,083	1,000	5419,083	21,880	,002
Fehler (Medikament)	Sphärizität angen.	1733,750	14	123,839		
	Greenhouse-Geisser	1733,750	9,153	189,415		
	Huynh-Feldt	1733,750	10,404	166,638		
	Untergrenze	1733,750	7,000	247,679		
Aufgabe	Sphärizität angen.	8037,750	2	4018,875	18,529	,000
	Greenhouse-Geisser	8037,750	1,856	4330,863	18,529	,000
	Huynh-Feldt	8037,750	2,000	4018,875	18,529	,000
	Untergrenze	8037,750	1,000	8037,750	18,529	,004
Fehler (Aufgabe)	Sphärizität angen.	3036,583	14	216,899		
	Greenhouse-Geisser	3036,583	12,991	233,737		
	Huynh-Feldt	3036,583	14,000	216,899		
	Untergrenze	3036,583	7,000	433,798		
Medikament * Aufgabe	Sphärizität angen.	1099,667	4	274,917	2,774	,046
	Greenhouse-Geisser	1099,667	1,962	560,571	2,774	,098
	Huynh-Feldt	1099,667	2,718	404,605	2,774	,074
	Untergrenze	1099,667	1,000	1099,667	2,774	,140
Fehler (Medikament* Aufgabe)	Sphärizität angen.	2774,500	28	99,089		
	Greenhouse-Geisser	2774,500	13,732	202,049		
	Huynh-Feldt	2774,500	19,025	145,833		
	Untergrenze	2774,500	7,000	396,357		

Tabelle 5-9

Nun das Ergebnis für das normal score (INT)-Verfahren, zunächst der Mauchly-Test:

Mauchly-Test auf Sphärizität ^a						
Innersubjekteffekt	Mauchly-W	Approx.. Chi- Quadrat	df	Sig.	Epsilon ^b	
					Greenhouse -Geisser	Huynh-Feldt
Medikament	,350	6,297	2	,043	,606	,665
Aufgabe	,869	,845	2	,655	,884	1,000
Medikament * Aufgabe	,020	21,075	9	,016	,426	,549

der zeigt, dass lediglich für den Effekt `Aufgabe` durch die Transformation die Varianzheterogenität beseitigt werden konnte. Abgesehen davon empfehlen Beasley & Zumbo [53] ohnehin, in jedem Fall die adjustierten F-Tests, z.B. den von Huynh-Feldt, zu verwenden. Nachfolgend die (um die Fehlerterme) verkürzte Anova-Tabelle:

Tests der Innersubjekteffekte						
Quelle		Quadrats. vom Typ III	df	Mittel der Quadrate	F	Sig.
Medikament	Sphärizität angen.	27,444	2	13,722	20,831	,000
	Greenhouse-Geisser	27,444	1,212	22,640	20,831	,001
	Huynh-Feldt	27,444	1,330	20,635	20,831	,001
	Untergrenze	27,444	1,000	27,444	20,831	,003
Aufgabe	Sphärizität angen.	40,778	2	20,389	20,552	,000
	Greenhouse-Geisser	40,778	1,768	23,068	20,552	,000
	Huynh-Feldt	40,778	2,000	20,389	20,552	,000
	Untergrenze	40,778	1,000	40,778	20,552	,003
Medikament * Aufgabe	Sphärizität angen.	6,056	4	1,514	3,361	,023
	Greenhouse-Geisser	6,056	1,703	3,555	3,361	,075
	Huynh-Feldt	6,056	2,195	2,759	3,361	,058
	Untergrenze	6,056	1,000	6,056	3,361	,109

5. 4. 3 Verallgemeinerte Kruskal-Wallis- und Friedman-Tests (KWF)

Dieses Verfahren wurde für den Fall von Messwiederholungen auf einem Faktor bereits in Kapitel 5.3.3 vorgestellt. Es werden fallweise (wie beim Friedman-Test) die Ränge 1,...,K (hier also die Ränge 1,...,I*J) vergeben, eine parametrische Varianzanalyse durchgeführt und aus den Streuungsquadratsummen (SS, Sum of Sq) werden χ^2 -Tests konstruiert. Dieses kann als Verallgemeinerung des Friedman-Test angesehen werden, da es im 1-faktoriellen Fall mit letzterem identisch sind, wie oben gezeigt wurde (vgl. Kapitel 5.3.2). Hier im Fall von Messwiederholungen auf zwei (oder mehr) Faktoren (C und D) gibt es allerdings ein kleines Problem: Das in Kapitel 5.3.3 vorgestellte Verfahren, verallgemeinert auf zwei Faktoren, führt für die Tests der Haupteffekte zu χ^2 -Werten, die etwas zu klein sind, also zu einem leicht konservativen Testverhalten neigen. Man hat somit für den Test der Haupteffekte drei Möglichkeiten:

- 1) Berechnung nur einer Rangtransformation, dafür Inkaufnahme konservativer χ^2 -Tests für die beiden Haupteffekte,
- 2) Testen der Haupteffekte über separate Friedman-Analysen, die allerdings zuvor die Berechnung diverser Summen voraussetzen,
- 3) Berechnung einer zusätzlichen Rangtransformation speziell für die Tests der beiden Haupteffekte.

Sinnvoll erscheint die erste Variante. Liegen die Testergebnisse für die Haupteffekte knapp unter der Signifikanzgrenze, können nachträglich die exakten Friedman-Analysen hierfür durchgeführt werden.

Folgende Schritte sind für den Test der Interaktion bzw. für die o.a. Varianten 1) und 2) durchzuführen:

- Für jede Erhebungseinheit (Versuchsperson) werden die Werte in Ränge (1,...,I*J) transformiert (I*J=Anzahl der gesamten Messwiederholungen).
- Mit den Rängen wird eine parametrische Varianzanalyse mit Messwiederholungen durchgeführt.
- Auf Basis der Anova-Tabelle wird folgender χ^2 -Test aufgestellt: Die χ^2 -Tests für die Haupteffekte und die Interaktion werden wie folgt errechnet (vgl. Formel 2-6):

$$\chi^2 = \frac{SS_{\text{Effekt}}}{(SS_{\text{Effekt}} + SS_{\text{Fehler}}) / (df_{\text{Effekt}} + df_{\text{Fehler}})}$$

wobei SS_{Effekt} die Streuungsquadratsumme (Sum of Squares) des zu testenden Effektes (C, D oder C*D) ist, SS_{Fehler} die Streuungsquadratsumme des zum Effekt gehörenden Fehlers ist sowie df die entsprechenden Freiheitsgrade.

- Die χ^2 -Werte sind dann in den Tafeln für den χ^2 -Test auf Signifikanz zu überprüfen, wobei die Freiheitsgrade die Zählerfreiheitsgrade (df_{Effekt}) des entsprechenden F-Tests sind.
- Die χ^2 -Werte sollten alternativ gemäß Iman & Davenport (vgl. Formel 2-1b) in F-Werte umgerechnet werden, in diesem Fall entspricht dies:

$$F = \frac{(N-1)\chi^2}{df_1 + df_2 - \chi^2}$$

wobei df_1 die Zähler- und df_2 die Nennerfreiheitsgrade des entsprechenden F-Tests sind.

Entscheidet man sich für die Variante 2), so müssen zunächst für jede Stufe des einen Messwiederholungsfaktors C die Summe über die Stufen des anderen Messwiederholungsfaktors D berechnet werden und umgekehrt. Für jede der beiden Summenvektoren werden dann Friedman-Analysen zum Test des jeweiligen Haupteffekts durchgeführt.

Im Fall von Variante 3) sind folgende Schritte erforderlich:

- Der Datensatz muss zunächst umstrukturiert werden, so dass die Messwiederholungen zu Fällen werden.
- Berechnung für jede Stufe des einen Messwiederholungsfaktors C und für jede Versuchsperson die Summe über die Stufen des anderen Messwiederholungsfaktors D über Aggregatdaten-Berechnung: ergibt eine Variable s_C mit I Werten.
- Berechnung für jede Stufe des einen Messwiederholungsfaktors D und für jede Versuchsperson die Summe über die Stufen des anderen Messwiederholungsfaktors C über Aggregatdaten-Berechnung: ergibt eine Variable s_D mit J Werten.
- Transformation von s_C in Ränge innerhalb jeder Versuchsperson: ergibt Variable r_{SC} .
- Transformation von s_D in Ränge innerhalb jeder Versuchsperson: ergibt Variable r_{SD} .
- Berechnung von $r_X = (r_{SC}-1) * I + r_{SD}$ (mit I Anzahl der Stufen von C).
- Der Datensatz muss wieder umstrukturiert werden, so dass die Fälle zu Messwiederholungen werden.

- Mit den dabei entstandenen $I*J$ Variablen R_X wird eine 2-faktorielle parametrische Varianzanalyse durchgeführt.
- Die Haupteffekte von C und D werden nun wie oben bei der ersten Variante mittels eines χ^2 -Tests überprüft (vgl. Formel 2-6).

Die Schritte der ersten Variante sollen am Datensatz des Beispiels 5 demonstriert werden.

mit R:

Zunächst wird die elementare Berechnung, anschließend eine R-Funktion hierfür vorgestellt. Diese Berechnung wird wieder mit der Funktion `ezANOVA` (Paket `ez`) durchgeführt. Dieses Mal aus folgendem Grund: Bei Analysen mit Messwiederholungen ist das Ergebnisobjekt von `aov` vom Typ „aovlist“ (anstatt vom Typ „aov“). Diese sind aber äußerst kompliziert aufgebaut, so dass eine Weiterverarbeitung von Ergebnissen wie die „Sum of Sq“ und „Df“ einen erheblichen Programmieraufwand erfordert, wohingegen die Anova-Tabelle von `ezANOVA` ein simpler Dataframe ist.

Ausgehend vom in 5.4.1 erstellten Dataframe `mydata5t` werden zunächst mittels der Funktionen `ave` und `rank` pro Vpn die Fehlerwerte in Ränge umgerechnet und an den Dataframe angehängt. Beim Aufruf von `ezANOVA` werden mittels des Parameters `detailed` die „Sum of Sq“ sowie die „Df“ ausgegeben, die für die weiteren Berechnungen benötigt werden. Vom Ergebnis interessiert nur die Komponente `ANOVA` mit der entsprechenden Tabelle, wobei die letzten Spalten, u.a. mit den p-Werten, hier nicht wiedergegeben werden:

```
mydata5t <- within(mydata5t, RFehler<- ave(Fehler,Vpn,FUN=rank)
aov2r      <- ezANOVA(mydata5t,RFehler,Vpn,within=(Medikament,Aufgabe),
                    detailed=T)
aov2ra     <- aov2r$ANOVA
aov2ra
```

	Effect	DFn	DFd	SSn	SSd	F
1	(Intercept)	1	7	1800.00000	-1.529641e-15	-8.237229e+18
2	Medikament	2	14	111.06250	2.827083e+01	2.749963e+01
3	Aufgabe	2	14	150.58333	6.475000e+01	1.627928e+01
4	Medikament:Aufgabe	4	28	26.41667	4.691667e+01	3.941385e+00

Tabelle 5-10

Die Spalten `ssn` und `ssd` (4. und 5. Spalte) enthalten die SS_{Effekt} bzw. den dazugehörigen Fehlerterm SS_{Fehler} , die Spalten `DFn` und `DFd` die entsprechenden Freiheitsgrade (2. und 3. Spalte). Mit folgenden Anweisungen lassen sich die χ^2 -Werte berechnen und auf Signifikanz überprüfen, wobei bewusst die erste Zeile „Intercept“ unterdrückt wird:

```
denom <- (aov2ra[,4]+aov2ra[,5]) / (aov2ra[,2]+aov2ra[,3])
chisq <- aov2ra[,4] / denom
df     <- aov2ra[,2]
pvalue <- 1-pchisq(chisq,df)
data.frame(Effekt=aov2ra[,1],Chisq=chisq,Df=df,
           Pvalue=round(pfvalue,digits=7))[2:4]
```

	Effekt	Chisq	DF	Pvalue
2	Medikament	12.75359	2	0.001700566
3	Aufgabe	11.18885	2	0.003718529
4	Medikament:Aufgabe	11.52727	4	0.021235612

Hierauf aufbauend kann die Iman-Davenport-Korrektur durchgeführt werden. Dazu wird zunächst das n (Anzahl Versuchspersonen n_{subj}) ermittelt. Die Spalten 2 und 3 enthalten die erforderlichen Freiheitsgrade Df_n und Df_d :

```
nsubj <- dim(mydata5) [1]
dfn <- aov2ra[,2]
dfd <- aov2ra[,3]
fvalue <- ((n-1)*chisq) / (dfn+dfd-chisq)
pfvalue <- 1-pf(fvalue,dfn,dfd)
data.frame(Effekt=aov2ra[,1],Fvalue=fvalue,Dfn=dfn,Dfd=dfd,
           Pvalue=round(pfvalue,digits=7))
```

	Effekt	Fvalue	Df1	Df2	Pvalue
1	(Intercept)	Inf	1	7	0.0000000
2	Medikament	27.499632	2	14	0.0000142
3	Aufgabe	16.279279	2	14	0.0002223
4	Medikament:Aufgabe	3.941385	4	28	0.0116312

Die p-Werte zeigen alle Effekte als signifikant an, so dass für die Haupteffekte `Medikament` und `Aufgabe` ein zusätzlicher Friedman-Test entfällt. Andernfalls wäre zum Beispiel für den Test von `Medikament` die folgende Eingabe auf Basis des Dataframes `mydata5t` erforderlich gewesen:

```
sm <- with(mydata5t, aggregate(Fehler, list(Vpn,Medikament), FUN=sum))
with(sm, friedman.test(x,Group.2,Group.1))
```

Hierdurch wird zunächst ein Dataframe `sm` mit den Variablen `x` (Fehlersumme), `Group.1` (`Vpn`) und `Group.2` (`Aufgabe`) erzeugt, wobei `Group.1` und `Group.2` vom Typ „factor“ sind. Das Ergebnis des Friedman-Tests für den Faktor `Group.2` (`Aufgabe`):

```
data: x, Group.2 and Group.1
Friedman chi-squared = 14, df = 2, p-value = 0.0009119
```

Alternativ kann auch die Funktion `np.anova` (vgl. Anhang 3.6) angewandt werden. Der Aufruf ist praktisch identisch mit dem der Standardfunktion `aov`. Basis ist auch hierfür der umstrukturierte Datensatz (`mydata5t`). Eingabe und Ausgabe:

```
np.anova(Fehler~Medikament*Aufgabe+Error(Vpn/(Medikament*Aufgabe)),
         mydata5t)
```

generalized Kruskal-Wallis/Friedman (Puri & Sen) tests including Iman & Davenport F-tests							
	Df	Sum Sq	Chisq	Pr(>Chi)	F value	Pr(>F)	
Medikament	2	111.062	12.7536	0.0017006	27.4996	1.416e-05	***
Residuals	14	28.271					
Aufgabe	2	150.583	11.1889	0.0037185	16.2793	0.0002223	***
Residuals	14	64.750					
Medikament:Aufgabe	4	26.417	11.5273	0.0212356	3.9414	0.0116312	*
Residuals	28	46.917					

mit SPSS:

- Zunächst müssen für den Datensatz über das Menü „Daten -> Umstrukturieren“ die Messwiederholungen in Fälle transformiert werden (siehe dazu Anhang 1.1.2).
- Die Variable `Fehler` wird dann über das Menü „Transformieren -> Rangfolge bilden“ in

Ränge umgerechnet, wobei (im Gegensatz zum Beispiel für den o.a. Rank transform Test) im Feld „Sortieren nach“ die Variable V_{pn} eingetragen werden muss, damit die Rangbildung für jede V_{pn} einzeln vorgenommen wird.

- Danach muss der Datensatz wieder zurück in das „normale“ Format mit Messwiederholungen transformiert werden (vgl. Anhang 1.2).
- Abschließend wird dann eine Varianzanalyse mit Messwiederholungen (Menü: „Allgemeines lineares Modell -> Messwiederholung“) für Fehler gerechnet.

Die Syntax für den ersten Schritt der Umstrukturierung, der Rangbildung innerhalb von V_{pn} bzw. des zweiten Schritts der Umstrukturierung in der SPSS-Syntax:

```
Varstocases
  /Id=Vpn
  /make Fehler from v1 v2 v3 v4 v5 v6 v7 v8 v9
  /index=Medikament(3) Aufgabe(3)
  /keep=Geschlecht
  /null=keep.

Rank variables=Fehler (A) by Vpn
  /rank.

Sort cases by Vpn Medikament Aufgabe.
Casestovars
  /Id=Vpn
  /index=Medikament Aufgabe
  /groupby=variable.
```

Nach der Umstrukturierung haben die rangtransformierten Variablen v_1, v_2, \dots die Namen $R_{Fehler.1.1}, R_{Fehler.1.2}, \dots, R_{Fehler.3.3}$. Die Syntax für die Varianzanalyse lautet:

```
GLM R_{Fehler.1.1} R_{Fehler.1.2} R_{Fehler.1.3} R_{Fehler.2.1} R_{Fehler.2.2}
  R_{Fehler.2.3} R_{Fehler.3.1} R_{Fehler.3.2} R_{Fehler.3.3}
  /WSfactor=Medikament 3 Polynomial Aufgabe 3 Polynomial
  /WSdesign=Medikament Aufgabe Medikament*Aufgabe.
```

Bei dem Ergebnis der Varianzanalyse kann der Mauchly-Test hier außer Acht bleiben. Somit ist in der Anova-Tabelle nur die Zeile „Sphärizität angen.“ relevant:

Tests der Innersubjekteffekte						
Quelle		Quadrat summe	df	Mittel der Quadrate	F	Sig.
Medikament	Sphärizität angen.	111,063	2	55,531	27,500	,000
Fehler (Medikament)	Sphärizität angen.	28,271	14	2,019		
Aufgabe	Sphärizität angen.	150,583	2	75,292	16,279	,000
Fehler (Aufgabe)	Sphärizität angen.	64,750	14	4,625		
Medikament * Aufgabe	Sphärizität angen.	26,417	4	6,604	3,941	,012
Fehler (Medikament*Aufgabe)	Sphärizität angen.	46,917	28	1,676		

Tabelle 5-11

Die χ^2 -Werte müssen nun „mit der Hand“ aus den Werten der o.a. Tabelle (Spalten „Quadratsumme“ und „df“) berechnet werden:

$$\chi^2_{\text{Medikament}} = \frac{111,06}{(111,06 + 28,27)/(2 + 14)} = 12,75 \quad df_{\text{Medikament}} = 2$$

$$\chi^2_{\text{Aufgabe}} = \frac{150,58}{(150,58 + 64,75)/(2 + 14)} = 11,19 \quad df_{\text{Aufgabe}} = 2$$

$$\chi^2_{\text{Interaktion}} = \frac{26,42}{(26,42 + 46,92)/(4 + 28)} = 11,53 \quad df_{\text{Interaktion}} = 4$$

Die 5%- bzw. 1%-Schranken der χ^2 -Verteilung liegen bei 6,0 bzw. 9,2 (df=2) und bei 9,5 bzw. 13,3 (df=4). Somit sind alle Effekte signifikant.

Schließlich noch die Umrechnung der χ^2 -Werte in F-Werte gemäß Iman & Davenport. Theoretisch könnte dies an dieser Stelle entfallen, da alle Tests als signifikant ausgewiesen wurden und somit keine Verbesserung durch die Transformation sinnvoll erscheint. Dennoch soll die Berechnung hier demonstriert werden:

$$F_{\text{Medikament}} = \frac{(8-1) \cdot 12,75}{2 + 14 - 12,75} = 27,46 \quad df_{\text{Zaehler}} = 2 \quad df_{\text{Nenner}} = 14$$

$$F_{\text{Aufgabe}} = \frac{(8-1) \cdot 11,19}{2 + 14 - 11,19} = 16,32 \quad df_{\text{Zaehler}} = 2 \quad df_{\text{Nenner}} = 14$$

$$F_{\text{Interaktion}} = \frac{(8-1) \cdot 11,53}{4 + 28 - 11,53} = 3,94 \quad df_{\text{Zaehler}} = 4 \quad df_{\text{Nenner}} = 28$$

Die 5%- bzw. 1%-Schranken der F-Verteilung für df=2 und df=4 liegen bei 3,74 bzw. 6,51 sowie für df=4 und df=28 bei 2,71 bzw. 4,07.

Die χ^2 -Tests der beiden Haupteffekte `Medikament` und `Aufgabe` erweisen sich als signifikant, so dass für diese ein zusätzlicher Friedman-Test entfällt. Andernfalls wäre zum Beispiel für den Test von `Medikament` die folgende Eingabe auf Basis des Datensatzes `mydata5` erforderlich gewesen:

```
compute m1=v1+v2+v3.
compute m2=v4+v5+v6.
compute m3=v7+v8+v9.
Nptests /related test(m1 m2 m3) Friedman.
```

womit die Summen für die 3 Medikamentenstufen über die 3 Aufgaben berechnet werden und anschließend die Friedman-Varianzanalyse durchgeführt wird:

Gesamtanzahl	8
Teststatistik	14,000
Freiheitsgrade	2
Asymptotische Sig. (zweiseitiger Test)	,001

5.4.4 Aligned rank transform (ART)

Das Prinzip des Aligned rank transform-Tests wurde oben bereits erläutert (vgl. Kapitel 4.3.6).

Die Schritte noch einmal im Einzelnen:

- Durchführung einer (normalen) Anova mit Haupt- und Interaktionseffekten.
- Speichern der Residuen (e_m),
- Eliminieren des zu untersuchenden Effekts aus den Residuen:
 Interaktionseffekt: $e_m + (\alpha\beta_{ij} - \alpha_i - \beta_j + 2\mu)$
 Haupteffekte: $e_m + (\alpha_i + \beta_j - \mu)$
- Umrechnung der bereinigten Residuen in Ränge.
- Durchführung einer normalen Anova mit Haupt- und Interaktionseffekten mit den Rängen, aus der dann der untersuchte Effekt abgelesen werden kann.

Es sei noch einmal darauf aufmerksam gemacht, dass die ART-Tests für die beiden Haupteffekte statistisch nicht erforderlich sind.

Dieses Verfahren stellt in erster Linie eine Verbesserung des o.a. Rank transform Tests da, um die Haupt- und Interaktionseffekte sauber zu trennen (vgl. Kapitel 4.3.6). Es ist also in erster Linie für metrische Variablen gedacht, die die Normalverteilungs-Voraussetzung nicht erfüllen, nicht jedoch für Variablen mit beliebigen Eigenschaften. Insofern sollte die Möglichkeit genutzt werden, die rangtransformierten Daten mittels des Mauchly-Tests auf Varianzhomogenität bzw. Spherizität zu überprüfen, um dann gegebenenfalls anstatt des normalen F-Tests die Variante von Huynh & Feldt anzuwenden. Oder alternativ ohne Beachtung des Mauchly-Tests die adjustierten F-Tests z.B. von Huynh & Feldt verwenden. Nach Beasley [84] spielt bei dieser ART-Methode die Spherizität keine Rolle, so dass ein Blick auf den Mauchly-Test entfallen kann und in der Anova-Tabelle ausschließlich der „normale“ F-Test von Bedeutung ist.

Es soll nun im Folgenden für den Beispieldatensatz 5 überprüft werden, ob die oben ausgewiesene Signifikanz der Interaktion garantiert ist.

mit R:

Zunächst wird die elementare Berechnung, anschließend eine R-Funktion hierfür vorgestellt. Ausgehend vom oben (Kapitel 5.4.1) erstellten Dataframe `mydata5t` werden zunächst

- die Residuen der Varianzanalyse mit den Faktoren `Medikament` und `Aufgabe` ermittelt (vgl. dazu 5.3.1),
- die Effekte `ma` des Faktors `Medikamente` bzw. `mb` des Faktors `Aufgaben` berechnet,
- die Zellenmittelwerte `mab` sowie den Gesamtmittelwert `mm`,
- in der Variablen `rabr` die Residuen um die Haupteffekte bereinigt und in Ränge transformiert,
- in der Variablen `rar` die Residuen um den Interaktionseffekt bereinigt und in Ränge transformiert.
- Anschließend werden Varianzanalysen für `rabr` zum Test des Interaktionseffekts

durchgeführt:

```
aov3r <- aov(Fehler~Medikament*Aufgabe + Vpn, mydata5t)
mydata5s <- cbind(mydata5t, resid=aov3r$residuals)
mydata5s <- within(mydata5s,
  { ma <- ave(Fehler,Medikament,FUN=mean);
    mb <- ave(Fehler,Aufgabe,FUN=mean);
    mab<- ave(Fehler,Medikament,Aufgabe, FUN=mean);
    mm <- mean(Fehler)})
mydata5s <- within(mydata5s,
  { rabr<- rank(round(resid-mab+ma+mb-mm,digits=7));
    rar <- rank(round(resid-ma-mb+2*mm,digits=7))})
aov3rab <- aov(rabr~Medikament*Aufgabe+Error(Vpn/(Medikament*Aufgabe)),
  mydata5s)
summary(aov3rab)
aov3ra <- aov(rar~Medikament*Aufgabe+Error(Vpn/(Medikament*Aufgabe)),
  mydata5s)
summary(aov3ra)
```

Nachfolgend zunächst die Ergebnisse der Anova zum Test des Interaktionseffekts, dessen Signifikanz ($p=0.017$) danach bestätigt ist:

Error: Vpn						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Residuals	7	518.6	74.08			
Error: Vpn:Medikament						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Medikament	2	10	5.0	0.011	0.989	
Residuals	14	6215	443.9			
Error: Vpn:Aufgabe						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Aufgabe	2	32	16.0	0.019	0.981	
Residuals	14	11491	820.8			
Error: Vpn:Medikament:Aufgabe						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Medikament:Aufgabe	4	4363	1090.8	3.617	0.0169 *	
Residuals	28	8443	301.5			

Tabelle 5-12

sowie der Ergebnisse für rar zum Test der Haupteffekte, die ebenfalls beide signifikant

sind:

Error: Vpn						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Residuals	7	52	7.429			
Error: Vpn:Medikament						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Medikament	2	8042	4021	25.11	2.34e-05	***
Residuals	14	2242	160			
Error: Vpn:Aufgabe						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Aufgabe	2	12830	6415	23.58	3.29e-05	***
Residuals	14	3808	272			
Error: Vpn:Medikament:Aufgabe						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Medikament:Aufgabe	4	185	46.32	0.33	0.855	
Residuals	28	3931	140.40			

Tabelle 5-13

Schließlich noch die Alternative mit der R-Funktion `art2.anova` (vgl. Anhang 3.8). Der Aufruf ist praktisch identisch mit dem der Standardfunktion `aov`. Basis ist auch hierfür der umstrukturierte Datensatz (`mydata5t`). Eingabe und Ausgabe:

```
art2.anova(Fehler~Medikament*Aufgabe+Error(Vpn/(Medikament*Aufgabe)),
mydata5t)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Medikament	2	27.4	13.72	20.8313	6.367e-05	***
Residuals	14	9.2	0.66			
Aufgabe	2	40.8	20.39	20.5520	6.833e-05	***
Residuals	14	13.9	0.99			
Medikament:Aufgabe	4	4363.0	1090.76	3.6173	0.01692	*
Residuals	28	8443.0	301.54			

mit SPSS:

Wie beim Rank Transform-Test (vgl. Kapitel 5.4.2) muss zunächst der Datensatz umstrukturiert werden, wobei die Messwiederholungen in Fälle gewandelt werden.

```
Varstocases
  /Id=Vpn
  /make Fehler from v1 v2 v3 v4 v5 v6 v7 v8 v9
  /index=Medikament(3) Aufgabe(3)
  /keep=Geschlecht
  /null=keep.
```

Mit diesem Datensatz wird zur Ermittlung der Residuen des Modells mit den Faktoren `Medikament` und `Aufgaben` eine Varianzanalyse (ohne Messwiederholungen, dafür mit dem Faktor `Vpn` der Versuchspersonenkennung) gerechnet (im Menü „Modell“ „Anpassen“ wählen, die Interaktion von `Medikament` und `Aufgaben` für die rechte Seite auswählen sowie den Haupteffekt `Vpn`):

```
Unianova Fehler by Medikament Aufgabe Vpn
  /save=resid
  /design=Aufgabe*Medikament Vpn.
```

Über `Aggregate` werden nun die Mittelwerte für Medikament (a_i), Aufgaben (b_j), Zellen (m_{ij}) und gesamt (mm) berechnet, um die Effekte von den Residuen abzuziehen und das Ergebnis in Ränge umzurechnen:

- rab bzw. die Ränge $rabr$ zum Test der Interaktion
- ra bzw. rar zum Test der Haupteffekte

```
Aggregate
  /outfile=* mode=addvariables
  /break=Medikament Aufgabe /mij=mean(Fehler).
Aggregate
  /outfile=* mode=addvariables
  /break=Medikament /ai=mean(Fehler).
Aggregate
  /outfile=* mode=addvariables
  /break=Aufgabe /bj=mean(Fehler).
Aggregate
  /outfile=* mode=addvariables
  /break= /mm=mean(Fehler).
Compute rab = res_1 + (mij - ai - bj + 2*mm).
Compute ra = res_1 + (ai + bj - mm).
Rank variables=ra rab (A)
  /rank into rar rabr.
execute.
```

Anschließend wird der Datensatz wieder in die ursprüngliche Form transformiert:

```
Sort cases by Vpn Medikament Aufgabe.
Casestovars /Id=Vpn
  /index=Medikament Aufgabe
  /groupby=variable.
```

Schließlich wird dann für $rabr$, die im umstrukturierten Datensatz die Namen $rabr.1.1$, $rabr.1.2, \dots$ hat, bzw. rar , eine Varianzanalyse mit Messwiederholungen mit den Faktoren **Medikament** und **Aufgaben** gerechnet:

```
GLM rabr.1.1 rabr.1.2 rabr.1.3 rabr.2.1 rabr.2.2 rabr.2.3
  rabr.3.1 rabr.3.2 rabr.3.3
  /wsfactor=Medikament 3 Polynomial Aufgabe 3 Polynomial
  /wsdesign=Medikament Aufgabe Medikament*Aufgabe.
```

Nachfolgend die Ergebnisse für den Test der Interaktion. Nach Beasley [84] spielt bei dieser ART-Methode die Sphärität keine Rolle, so dass ein Blick auf den Mauchly-Test entfallen kann und in der Anova-Tabelle ausschließlich die Zeile „Sphärität angenommen“ von Bedeutung ist:

Tests der Innersubjekteffekte						
Quelle		Quadrat- summe vom Typ III	df	Mittel der Quadrate	F	Sig.
Medikament	Sphärität angen.	9,146	2	4,573	,010	,990
Fehler(Medikament)	Sphärität angen.	6282,410	14	448,744		

Aufgabe	Sphärizität angen.	30,896	2	15,448	,019	,981
Fehler(Aufgabe)	Sphärizität angen.	11459,493	14	818,535		
Medikament * Aufgabe	Sphärizität angen.	4313,458	4	1078,365	3,573	,018
Fehler (Medikament*Aufgabe)	Sphärizität angen.	8449,486	28	301,767		

Tabelle 5-14

bzw. die Anova-Tabelle für den Test der Haupteffekte:

Tests der Innersubjekteffekte						
Quelle		Quadrat- summe vom Typ III	df	Mittel der Quadrate	F	Sig.
Medikament	Sphärizität angen.	8042,312	2	4021,156	25,113	,000
Fehler(Medikament)	Sphärizität angen.	2241,687	14	160,121		
Aufgabe	Sphärizität angen.	12830,333	2	6415,167	23,584	,000
Fehler(Aufgabe)	Sphärizität angen.	3808,167	14	272,012		
Medikament * Aufgabe	Sphärizität angen.	185,292	4	46,323	,330	,855
Fehler (Medikament*Aufgabe)	Sphärizität angen.	3931,208	28	140,400		

Tabelle 5-15

5.4.5 ATS-Tests von Akritas, Arnold & Brunner

Den von Akritas, Arnold und Brunner entwickelten ATS-Test gibt es auch für mehrfaktorielle Varianzanalysen mit Messwiederholungen. Während in R dazu das Paket `npard` zur Verfügung steht, gibt es in SPSS derzeit keine Möglichkeit zur Anwendung dieses Verfahrens.

mit R:

Die 2-faktorielle Analyse mittels `npard` soll am Datensatz des Beispiels 5 gezeigt werden. Ausgangsbasis ist wieder der oben (vgl. Kapitel 5.4.1) erstellte Dataframe `mydata5t`. Die Analyse kann mittels zwei Funktionen erfolgen:

- `npard` ist eine universelle Funktion für alle verarbeitbaren Designs.
- `ld.f2` erlaubt fehlende Werte bei den Messwiederholungen, gibt einen Mittelwertplot aus sowie eine Reihe weiterer, hier allerdings nicht interessierende Statistiken aus.

Beide geben sowohl die WTS als auch die ATS aus. Die Ausgabe unterscheidet sich nicht hinsichtlich der Wiedergabe dieser Statistiken. Nachfolgend zunächst die Eingabe für beide Varianten, wobei zu beachten ist, dass bei der Funktion `npard` trotz Angabe des Dataframes die Variablennamen nicht automatisch gefunden werden. Daher muss entweder jeder Variablenname zusammen mit dem Dataframe-Namen in der üblichen Form, z.B. `mydata5t$Fehler` angegeben werden oder ein `attach(Dataframe)` ausgeführt werden.

```
attach(mydata5t)
nparLD(Fehler~Medikament*Aufgabe,mydata5t,mydata5t$Vpn)
ld.f2(score,Medikament,Aufgabe,Vpn,
      time1.name="Medikament",time2.name="Aufgabe") -> ano
round(ano$ANOVA.test,4)
```

Bei `ld.f2` müssen die Faktoren zweimal angegeben werden: zum einen zur Identifikation des Faktors, zum anderen in “...” als Name des Faktors für die Ausgabe.

Nachfolgend die Ausgabe von `nparLD`:

```
Call:
Fehler ~ Medikament * Aufgabe

Wald-Type Statistic (WTS):
```

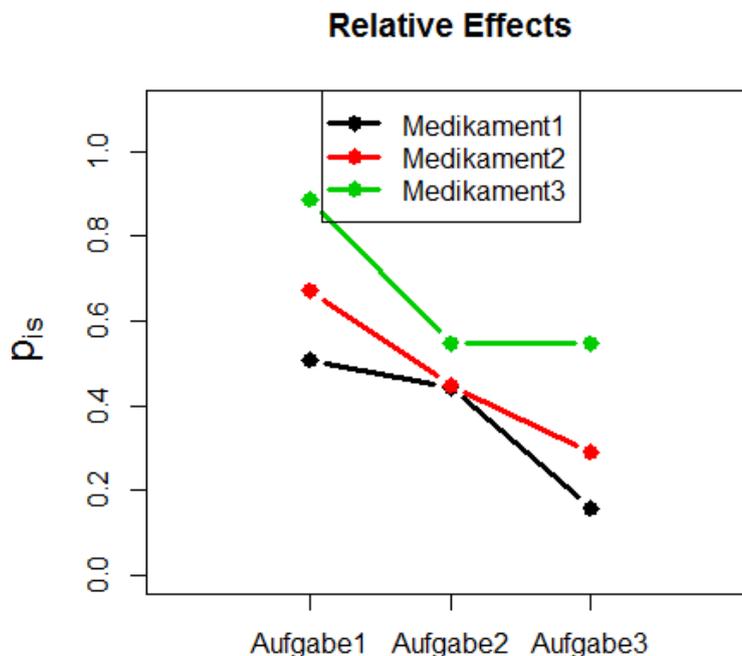
	Statistic	df	p-value
Medikament	44.43367	2	2.245694e-10
Aufgabe	43.50097	2	3.580012e-10
Medikament:Aufgabe	12.38836	4	1.468530e-02

```
ANOVA-Type Statistic (ATS):
```

	Statistic	df	p-value
Medikament	21.8795	1.3076	0.0000
Aufgabe	18.5288	1.8559	0.0000
Medikament:Aufgabe	2.7744	1.9617	0.0635

Tabelle 5-16

`ld.f2` gibt noch zusätzlich einen Interaktionsplot aus (siehe nächste Seite), allerdings der relativen Effekte (vgl. Kapitel 2.8) anstatt der Mittelwerte, da sich ja die Hypothesen auf erstere beziehen:



Bei der Ausgabe von `ld.f2` gibt es die Möglichkeit, einzelne Teile auszugeben, etwa die ATS- (Anova-) Tabelle (`..$ANOVA.test`) oder die WTS- (Wald-Test-) Tabelle (`..$wald.test`). Dies hat den Vorteil, dass man über die Funktion `round` die Zahlendarstellung der Art `xxxe-nn` ändern kann.

	Statistic	df	p-value
Medikament	21.8795	1.3076	0.0000
Aufgabe	18.5288	1.8559	0.0000
Medikament:Aufgabe	2.7744	1.9617	0.0635

5.4.6 Bredenkamp Tests

Bredenkamp (vgl. [1], S. 1024 ff und [15]) hat für den Versuchsplan mit Messwiederholungen auf zwei Faktoren auch eine Lösung vorgeschlagen, die sich aber nicht mit dem verallgemeinerten Friedman-Test (vgl. Kapitel 5.4.3) deckt. Das Verfahren erfolgt analog zu dem für unabhängige Stichproben (vgl. Kapitel 4.3.8):

- Für den Test von Faktor A wird für jede Stufe von A die Summe der abhängigen Variablen über die Stufen von Faktor B ermittelt. Anschließend wird ein Friedman-Test über diese Summen durchgeführt.
- Der Test von Faktor B erfolgt analog zu dem o.a. Test für Faktor A.
- Für den Test der Interaktion wird zunächst ein Friedman-Test über alle Messwiederholungen durchgeführt. Von dem resultierenden χ^2 -Wert werden die beiden χ^2 -Werte von den Tests von A und B subtrahiert. Analog werden die Freiheitsgrade subtrahiert. Das Ergebnis wird schließlich in der Tabelle der χ^2 -Verteilung überprüft.

Leider gibt es zu diesem Verfahren keine vergleichenden Tests.

5.5 Fazit

Im Prinzip können hier nur die Ausführungen von Kapitel 4.5 wiederholt werden. Allerdings hat man hier mit den modifizierten F-Tests von Greenhouse & Geisser als auch Huynh & Feldt bessere Möglichkeiten, inhomogenen Varianzen zu begegnen.

Für die SPSS-Benutzer kommt hier erschwerend hinzu, dass wegen der Rangbildung der Messwiederholungen bei vielen nichtparametrischen Verfahren eine zweimalige Umstrukturierung des Datensatzes erforderlich wird. Dies lässt vielleicht den Untersucher auf die Robustheit der Varianzanalyse bauen.

Verfahren	Medikament	Aufgabe	Interaktion
parametrisch	< 0.001	< 0.001	0.023
parametrisch - Greenhouse & Geisser	< 0.001	< 0.001	0.075
parametrisch - Huynh & Feldt	0.001	< 0.001	0.068
Rank transform Test	< 0.001	< 0.001	0.046
normal scores (INT)	< 0.001	< 0.001	0.061
Aligned Rank Transform	< 0.001	< 0.001	0.017
Verall. Kruskal-Wallis-Friedman-Tests (KWF)	0.002	0.004	0.021
Verall. KWF-Tests mit Iman-Davenport-Korr.	< 0.001	< 0.001	0.012
Akritis, Arnold & Brunner ATS	< 0.001	< 0.001	0.063

Tabelle 5-20

Abschließend werden in o.a. Tabelle 5-20 für den oben benutzten Datensatz (`mydata5`) die Ergebnisse aller Verfahren, und zwar die p-Werte für alle drei Effekte, in einer Tabelle gegenübergestellt. Zu beachten ist, dass signifikante Abweichungen von der Varianzhomogenität (hier: Sphärität) vorliegen, so dass das parametrische Verfahren ohne Korrekturen zu möglicherweise falschen Signifikanzen führen kann. Schließlich sollte man - wie schon oben gesagt - die hier erzielten Ergebnisse nicht verallgemeinern.

6. Gemischte Versuchspläne

Unter *gemischten Versuchsplänen*, auch *Split Plot Designs* genannt, versteht man in der Regel solche, in denen sowohl Messwiederholungsfaktoren als auch Gruppierungsfaktoren (ohne Messwiederholungen) enthalten sind. So wird im Folgenden davon ausgegangen, dass ein Merkmal x K -mal (unter verschiedenen Bedingungen) erhoben wurde, so dass Variablen x_1, \dots, x_K vorliegen, deren Mittelwerte verglichen werden sollen. Die Struktur kann aber auch hier mehrfaktoriell sein. Die Ausgangssituation ist also ähnlich wie in Kapitel 5. Hier kommt allerdings hinzu, dass die Beobachtungseinheiten, z.B. Versuchspersonen, Gruppen zugeordnet sind, deren Einfluss ebenfalls getestet werden soll.

Der einfachste Fall der in diesem Abschnitt betrachteten Versuchspläne beinhaltet jeweils einen Gruppierungsfaktor sowie einen Messwiederholungsfaktor. Allerdings unterscheiden sich mehrfaktorielle Designs, etwa mit zwei oder mehr Gruppierungsfaktoren oder mehreren Messwiederholungsfaktoren, nicht grundsätzlich von dem hier behandelten einfachen Fall. Verschiedentlich wird auf die Ausdehnung auf mehr als zwei Faktoren kurz eingegangen. Für den Fall zweier Messwiederholungsfaktoren sind zum Teil die Ergebnisse des letzten Kapitels 5 hier anzuwenden. Beispiele für 3-faktorielle Versuchspläne bieten die Datensätze 5, mit zwei Messwiederholungsfaktoren und einem Gruppierungsfaktor, sowie 6, mit einem Messwiederholungsfaktor und zwei Gruppierungsfaktoren, die zu Beginn des Kapitels 5 vorgestellt wurden.

An die Datenstruktur werden dieselben Anforderungen gestellt wie in Kapitel 5.1 beschrieben.

Im Folgenden wird weitgehend der einfache 2-faktorielle Fall behandelt. Ein entsprechender Datensatz bieten die Beispieldaten 4.

6.1 Voraussetzungen der parametrischen Varianzanalyse

Hier geht es um Versuchspläne, die sowohl abhängige als auch unabhängige Stichproben beinhalten. Für den einfachsten Fall einer 2-faktoriellen Varianzanalyse mit einem Gruppierungsfaktor A (mit I Gruppen) und einem Messwiederholungsfaktor C (mit J Wiederholungen) lautet das Modell dann:

$$x_{ijm} = \mu + \alpha_i + \gamma_j + \alpha\gamma_{ij} + \pi_m + \varepsilon_{ijm} \quad (i=1, \dots, I, j=1, \dots, J \text{ und } m=1, \dots, n) \quad (6-1)$$

Auch hier gibt es einen personenspezifischen Effekt: π_m . Die Voraussetzungen betreffen wiederum die Normalverteilung der Residuen und die Varianzhomogenität. Und hier kumulieren sich jetzt die Voraussetzungen der Analysen ohne Messwiederholungen (siehe Kapitel 4.1) sowie der Analysen mit Messwiederholungen (siehe Kapitel 5.2), die hier allerdings zum Teil etwas abgewandelt werden. Dazu kommen allerdings noch weitere, auf die nachfolgend näher eingegangen wird.

Doch zunächst wieder zur Normalverteilung der Residuen sowie der Personeneffekte π_m : Hier sind dieselben Schritte erforderlich wie in Kapitel 5.2 beschrieben.

Zur Varianzhomogenität hinsichtlich der Messwiederholungen: Auch hier ist wie in 5.2 beschrieben der Mauchly-Test auf Sphärizität (für alle Messwiederholungsfaktoren und deren Interaktionen) durchzuführen. Und im Fall von Inhomogenitäten wird wieder die Approximation von Huynh & Feldt (alternativ von Geisser & Greenhouse) empfohlen.

Doch was ist mit den Varianzhomogenitätstests aus Kapitel 4.1? Die sehen in diesem Versuchsplan anders aus: Statt der Homogenität der Zellvarianzen wird hier die Homogenität der Fehler

(oder Residuen-) varianzen gefordert. Man kann sich das folgendermaßen vorstellen: Es wird für jede Gruppe des/der Gruppierungsfaktoren eine Varianzanalyse für den/die Messwiederholungsfaktoren gerechnet. Dann hat jeder Test eines Effektes (der Messwiederholungen) einen „eigenen“ Error-Term. Jeder dieser Fehlerterme muss nun über die Gruppen hinweg homogen sein. Das zu überprüfen ist ein mühseliges Unterfangen, zumal diese Fehlerterme von den Programmen nicht gruppenweise ausgewiesen werden. Es sei denn, man rechnet wirklich für jede Gruppe eine Varianzanalyse und vergleicht die Ergebnisse. Ähnlich wie bei der Analyse der Residuen kann man sich damit behelfen, dass für jede Messwiederholungsvariable ein Test auf Homogenität der Varianzen durchgeführt wird, z.B. mit dem schon mehrfach erwähnten Levene-Test. So macht es auch SPSS.

Modifizierte F-Tests zur Kompensierung von Varianzheterogenitäten, wie etwa in Kapitel 4.2.2 oder 4.3.3 vorgestellt, sind für dieses Design nicht verbreitet. Lediglich Huynh [87] hat für diesen Fall eine *general approximate procedure* (GA) sowie eine *improved general approximate procedure* (IGA) entwickelt, auf die aber hier nicht eingegangen wird. Daher ist im Fall von stark inhomogenen Varianzen auf nichtparametrische Verfahren auszuweichen.

Doch damit nicht genug. Der o.a. Mauchly-Test wird für alle Gruppen zusammen durchgeführt, d.h. es wird *eine* Kovarianzmatrix errechnet und geprüft. Das setzt aber voraus, dass die Kovarianzmatrizen für alle Gruppen (statistisch) gleich sind, um sie zu einer zusammenfassen zu können. Diese Homogenität der Kovarianzmatrizen muss wiederum geprüft werden, und zwar mit dem *Box-M-Test*. Anzumerken ist, dass streng genommen, statistisch gleiche Kovarianzmatrizen die o.a. Varianzhomogenität implizieren.

Schließlich ist darauf hinzuweisen, dass dieser Box-Test, ähnlich wie der Mauchly-Test, multivariate Normalverteilung der Messwiederholungsvariablen voraussetzt. Das ist wesentlich mehr, als für die eigentliche Varianzanalyse gefordert wird. An dieser Stelle sollte man sich an die Bemerkungen in Kapitel 1.7 erinnern: Die Voraussetzungen zur Prüfung der Voraussetzungen sind restriktiver als die Voraussetzungen der eigentlichen Analyse und sind selten erfüllt. D.h. die Ergebnisse dieser Voraussetzungstests sind mit besonderer Vorsicht zu betrachten. Was liegt hier näher, als direkt ein nichtparametrisches Verfahren anzuwenden?

6.2 Parametrische Varianzanalyse und Prüfung der Voraussetzungen

Auch hier soll zunächst einmal zum Vergleich die parametrische Varianzanalyse durchgeführt und die Prüfung der Voraussetzungen gezeigt werden. Das Prozedere wie auch die Ergebnisse sind zum Teil zwangsläufig mit denen aus Kapitel 5.3.1 identisch. Dieses wird noch einmal für den Fall gemischter Versuchspläne erläutert.

Zur Berechnung der Residuen gibt es folgende Möglichkeit: Der oder die Messwiederholungsfaktoren C, D,.. werden als Gruppierungsfaktoren gehandhabt. Dazu muss der Datensatz umstrukturiert werden, indem die Messwiederholungen in Fälle gewandelt werden. (Dies ist in R ohnehin für Analysen mit Messwiederholungen erforderlich.) Dann wird folgendes Modell *ohne* Messwiederholungen analysiert:

$$A * C * D + V_{pn} \quad (6-2)$$

wobei V_{pn} die Fallkennung, z.B. Versuchspersonennummer, ist. Die Residuen dieses Modells sind die Residuen des Modells mit dem Gruppierungsfaktor A sowie mit Messwiederholungen auf C (und D). Dies gilt auch analog für mehrere Gruppierungsfaktoren A, B,...

Dies ist zwar prinzipiell auch bei SPSS möglich, verursacht aber wegen der erforderlichen Um-

strukturierung etwas Aufwand. SPSS gibt allerdings für jede Messwiederholungsvariable x_j andere Residuen aus: $e'_{ijm} = x_{ijm} - \alpha\gamma_{ij} - \alpha_i - \gamma_j$. Aus dem Modell 6-1 ergibt sich für diese $e'_{ijm} = \pi_m + e_{ijm}$, d.h. um die Residuen e_{ijm} zu erhalten, müssen von den e'_{ijm} die π_m subtrahiert werden. Die erforderlichen Schritte sind dann:

- Speichern der Residuen: e'_{ijm} ,
- Ermitteln des Personeneffekts π_m aus

$$p_m = \left(\sum_j x_{jm} \right) / J \text{ und}$$

$a_i =$ Mittelwert der p_m für Gruppe i :

$$\pi_m = (p_m - \bar{p} - a_i),$$

- und schließlich $e_{ijm} = e'_{ijm} - \pi_m$.

(Die Subtraktion von \bar{p} von p_m zur Ermittlung von π_m kann entfallen, da sie für die Beurteilung der Residuen e_{im} ohne Bedeutung ist.)

Wie bei dieser Art der Residuen-Ermittlung diese gehandhabt und beurteilt werden können, wurde bereits in Kapitel 5.3.1 erläutert. Wie man sieht, ist dieses Verfahren relativ aufwändig, insbesondere wenn das Design mehrere Gruppierungsfaktoren enthält. Insofern empfiehlt es sich, das oben skizzierte Verfahren 6-2 anzuwenden.

Wenn man in den nachfolgenden Beispielrechnungen das Ergebnis des Mauchly-Tests hier mit dem aus 5.3.1 vergleicht, mögen die unterschiedliche Ergebnisse irritieren, da ja eigentlich die Gruppenstruktur nicht in den Test einfließen sollte. Tut sie aber doch. Denn hier werden im Gegensatz zum Modell ohne Gruppierungsfaktoren *gepoolte* Kovarianzmatrizen errechnet. D.h. die Berechnung erfolgt quasi gruppenweise, bevor die Matrizen zusammengefasst werden. Der Unterschied kann u.a. durch die verschiedenen Gruppenmittelwerte verursacht werden. Hierher rührt auch die in 6.1 erwähnte Voraussetzung der Homogenität der Kovarianzmatrizen.

mit R:

Ausgangsbasis ist wieder der in Kapitel 5.3.1 erstellte Dataframe `winer518t`. Zunächst wird die Anova zum Vergleich mit der „Standardfunktion“ `aov` durchgeführt, wenn das auch i.a. nicht sinnvoll ist, weil die Funktion `ezANOVA` zugleich den Mauchly-Test durchführt (siehe unten). Dabei werden durch den Modellterm `Error(Vpn/Zeit)` die Messwiederholungen auf dem Faktor `Zeit` gekennzeichnet:

```
data(winer518t)
aov1 <- aov(score~Geschlecht*Zeit+Error(Vpn/Zeit),winer518t)
summary(aov1)
```

Error: Vpn						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Geschlecht	1	3.33	3.333	0.472	0.512	
Residuals	8	56.53	7.067			
Error: Vpn:Zeit						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Zeit	2	58.07	29.033	22.05	2.52e-05	***
Geschlecht:Zeit	2	44.87	22.433	17.04	0.000109	***
Residuals	16	21.07	1.317			

Tabelle 6-1

Zunächst einmal zur Prüfung der Residuen ε_{ijm} auf Normalverteilung. Diese lassen sich, wie oben erläutert, bequem als Residuen eines Anova-Modells ohne Messwiederholungen ermitteln:

```
aov2<-aov(score~Geschlecht*Zeit+Vpn, winer518t)
res<-aov2$residuals
hist(res)
shapiro.test(res)
```

Da die Ergebnisse denen aus Kapitel 5.3.1 weitgehend ähnlich sind, wird auf die Wiedergabe hier verzichtet. Die Überprüfung der Normalverteilung der versuchspersonenspezifischen Abweichungen π_m ist dieselbe wie in Kapitel 5.3.1.

Die Überprüfung der Spharizität mittels des Mauchly-Tests wird mit der Funktion `ezANOVA` des Pakets `ez` vorgenommen:

```
library(ez)
ezANOVA(winer518t, score, Vpn, within=Zeit, between=Geschlecht)
```

\$ANOVA							
	Effect	DFn	DFd	F	p	p<.05	ges
2	Geschlecht	1	8	0.4716981	5.116202e-01		0.04118616
3	Zeit	2	16	22.0506329	2.522847e-05	*	0.42800983
4	Geschlecht:Zeit	2	16	17.0379747	1.086241e-04	*	0.36635819
\$`Mauchly's Test for Sphericity`							
	Effect	W	p	p<.05			
3	Zeit	0.9306201	0.7775055				
4	Geschlecht:Zeit	0.9306201	0.7775055				
\$`Sphericity Corrections`							
	Effect	GGe	p [GG]	HFe	p [HF]		
3	Zeit	0.9351214	4.280809e-05	1.209851	2.522847e-05		
4	Geschlecht:Zeit	0.9351214	1.683544e-04	1.209851	1.086241e-04		

Tabelle 6-2

Der Aufbau der Tabelle 6-2 wurde bereits kurz in 5.3.1 erläutert. Die Anova-Tabelle ist natürlich mit der in Tabelle 6-1 identisch. Da der Mauchly-Test keine Signifikanz zeigt, werden die Ergebnisse aus der ersten Tabelle (ANOVA) verwendet.

Für den Box-M-Test auf Homogenität gibt es zwar eine Funktion `boxM` im Paket `biotools`, diese brachte aber bei einem Vergleich falsche Resultate. Deswegen sei auf die entsprechende Funktion `boxm.test` im Anhang 3 verwiesen. Diese verlangt als Eingabe einen Dataframe mit den Messwiederholungsvariablen (also *nicht* den umstrukturierten Datensatz), hier also `winer518`, sowie den Gruppierungsfaktor, der vom Typ „factor“ sein muss:

```
boxm.test(winer518[,3:5], winer518$Geschlecht)
```

Die Ausgabe enthält alle Statistiken, es können aber auch Teilresultate abgefragt werden. Relevant ist lediglich der p-Wert, hier 0,622, wonach die Homogenität gewährleistet ist:

MBox	F	df1	df2	P
-----	7.5870	0.7344	6	463
-----				0.6221
Covariance matrices are not significantly different.				

Zur Überprüfung der Gleichheit der Fehlervarianzen werden hier der Einfachheit halber für die drei Messwiederholungsvariablen (Variablenindizes 3,4,5) jeweils die Gruppenvarianzen mit dem Levene-Test überprüft. Auch hier wird der ursprüngliche Dataframe `winer518` benutzt. In diesem Fall liegt nur ein Gruppierungsfaktor vor. Somit lassen sich alle Variablen mittels `apply` in einem Funktionsaufruf überprüfen:

```
library(car)
apply(winer518[,3:5], 2, leveneTest, win_518$Geschlecht)
```

```
$t1
Levene's Test for Homogeneity of Variance (center = median)
  Df F value Pr(>F)
group 1      0.1 0.7599
      8

$t2
Levene's Test for Homogeneity of Variance (center = median)
  Df F value Pr(>F)
group 1      0      1
      8

$t3
Levene's Test for Homogeneity of Variance (center = median)
  Df F value Pr(>F)
group 1     0.05 0.8287
      8
```

Da keines der Ergebnisse signifikant ist, kann die Varianzhomogenität angenommen werden.

mit SPSS:

Varianzanalysen mit Messwiederholungen erhält man in SPSS über das Menü „Allgemeines lineares Modell -> Messwiederholung“. Die Syntax für den Beispieldatensatz 4 (`winer518`) mit Ausgabe der Homogenitätstests lautet:

```
GLM t1 t2 t3
  /wsfactor=Zeit 3 polynomial
  /print homogeneity
  /wsdesign=Zeit
  /design=Geschlecht.
```

mit folgender Ausgabe des Mauchly-Tests, der Anova-Tabelle für die Messwiederholungseffekte (Innersubjekteffekte) und der Anova-Tabelle für den Gruppierungsfaktor (Zwischensubjekteffekte), wobei der Mauchly-Test keine Inhomogenitäten zeigt, so dass die Ergebnisse der Zeile „Sphärität angenommen“ verwendet werden können:

Mauchly-Test auf Sphärität ^a							
Innersubjekt- effekt	Mauchly-W	Approximiertes Chi-Quadrat	df	Sig.	Epsilon ^b		
					Greenhouse- Geisser	Huynh-Feldt	
Zeit	,931	,503	2	,778	,935	1,000	

Tests der Innersubjekteffekte						
Quelle		Quadrat- summe vom Typ III	df	Mittel der Quadrate	F	Sig.
Zeit	Sphärizität angen.	58,067	2	29,033	22,051	,000
	Greenhouse-Geisser	58,067	1,870	31,048	22,051	,000
	Huynh-Feldt	58,067	2,000	29,033	22,051	,000
Zeit * Geschlecht	Sphärizität angen.	44,867	2	22,433	17,038	,000
	Greenhouse-Geisser	44,867	1,870	23,990	17,038	,000
	Huynh-Feldt	44,867	2,000	22,433	17,038	,000
Fehler(Zeit)	Sphärizität angen.	21,067	16	1,317		
	Greenhouse-Geisser	21,067	14,962	1,408		
	Huynh-Feldt	21,067	16,000	1,317		

Tests der Zwischensubjekteffekte					
Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Konstanter Term	546,133	1	546,133	77,283	,000
Geschlecht	3,333	1	3,333	,472	,512
Fehler	56,533	8	7,067		

Tabelle 6-3

Darüberhinaus werden über den Zusatz `/print homogeneity` der Box-M-Test zur Überprüfung der Gleichheit der Kovarianzmatrizen für die beiden Gruppen sowie für alle 3 Variablen ein Levene-Test auf Gleichheit der Zellenvarianzen ausgegeben:

Box-Test auf Gleichheit der Kovarianzmatrizen	
Box-M-Test	7,587
F	,734
df1	6
df2	463,698
Sig.	,622

Der Box-Test zeigt keine Ungleichheit der Varianzen, so dass eine Voraussetzung für die Durchführung des Mauchly-Tests gegeben ist.

Levene-Test auf Gleichheit der Fehlervarianzen ^a				
	F	df1	df2	Sig.
t1	,159	1	8	,700
t2	,000	1	8	1,000
t3	,015	1	8	,905

Da alle drei Tests nicht signifikant sind, kann auch die Homogenität der Fehlervarianzen angenommen werden.

Bleibt noch die Überprüfung der Residuen auf Normalverteilung. Dazu wird das am Eingang dieses Kapitels genannte Modell ohne Messwiederholungen 6-2 gerechnet. Zunächst muss der Datensatz umstrukturiert werden, so dass aus den 3 Messwiederho-

lungen jeweils 3 Fälle erzeugt werden. Das ist im Anhang 1.1.1 ausführlich beschrieben. Die Syntax hierfür lautet:

```
Varstocases
  /id=Vpn
  /make score from t1 t2 t3
  /index=Zeit(3)
  /keep=Geschlecht
  /null=keep.
```

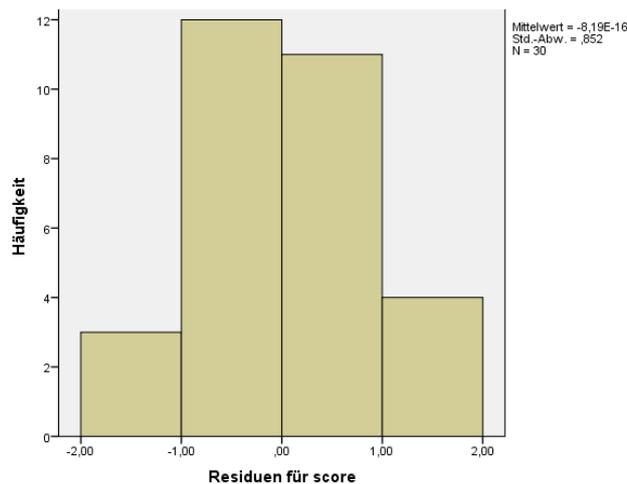
Die ersten Fälle des umstrukturierten Datensatzes sehen etwa folgendermaßen aus:

	Vpn	Geschlecht	Zeit	score
1	1	1	1	4
2	1	1	2	7
3	1	1	3	2
4	2	1	1	3
5	2	1	2	5
6	2	1	3	1
7	3	1	1	7
8	3	1	2	9

Für diesen Datensatz wird nun eine Varianzanalyse mit den Faktoren `Vpn`, `Geschlecht` und `Zeit` gerechnet, wobei das Modell angepasst werden muss: Anstatt des gesättigten Modells sind neben den Haupteffekten die Interaktion `Geschlecht*Zeit` auszuwählen. Ferner müssen die Residuen gespeichert werden, die anschließend den Namen `RES_1` haben. Schließlich werden diese dann in `Examine` (Explorative Datenanalyse) mittels Shapiro-Test und Histogramm auf Normalverteilung überprüft. Die Anweisungen hierfür:

```
Unianova score BY Geschlecht Zeit Vpn
  /save=resid
  /design=Geschlecht Zeit Geschlecht*Zeit Vpn.

Examine variables=RES_1
  /plot histogram.
```



Das automatisch erzeugte Histogramm basiert zunächst auf 11 Intervallen, was bei einem n von 30 keinen Sinn macht. Möglich wären hier 4, 5 oder 6 Intervalle (vgl. Kapitel 1.6), so dass eine Nachbereitung mit dem Grafikeditor erforderlich ist und o.a. Abbildung erzeugt.

Tests auf Normalverteilung						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
RES_1	,126	30	,200*	,959	30	,288

Auf Basis dieser Ergebnisse kann man die Residuen als normalverteilt annehmen.

6.3 Rank transform-Tests (RT)

Bei dem Rank transform Test werden lediglich die Werte der abhängigen Variablen über alle Messwiederholungen und Gruppen hinweg in Ränge gewandelt, um mit diesen dann eine „normale“ parametrische Varianzanalyse zu rechnen. Auch hier sollte man den Mauchly-Test durchführen, um die korrigierten F-Tests von Huynh & Feldt zu benutzen, falls die Sphärität nicht gegeben ist. Dieses Verfahren soll wieder am Beispieldatensatz 4 demonstriert werden.

mit R:

Ausgangsbasis ist wieder der in Kapitel 5.3.1 erstellte Dataframe `winer518t`. Zunächst wird die Variable `score` in Ränge umgerechnet, anschließend die Anova mit der `ezANOVA` durchgeführt, um die Ergebnisse des Mauchly-Tests zu erhalten:

```
data(winer518t)
within(winer518t, Rscore<-rank(score))
ezANOVA(winer518t, Rscore, Vpn, within=Zeit, between=Geschlecht)
```

\$ANOVA							
	Effect	DFn	DFd	F	p	p<.05	ges
2	Geschlecht	1	8	0.6079316	4.580116e-01		0.05306095
3	Zeit	2	16	22.3581811	2.325487e-05	*	0.42329132
4	Geschlecht:Zeit	2	16	16.0426724	1.502651e-04	*	0.34497155

\$`Mauchly's Test for Sphericity`				
	Effect	W	p	p<.05
3	Zeit	0.9861432	0.9523355	
4	Geschlecht:Zeit	0.9861432	0.9523355	

Tabelle 6-3

Da der Mauchly-Test keine Signifikanzen zeigt, können die unkorrigierten F-Tests benutzt werden. Dessen Ergebnisse weichen nicht nennenswert von denen der o.a. parametrischen Analyse (Tabelle 6-2) ab. Die Voraussetzung der Normalverteilung braucht hier nicht geprüft werden.

mit SPSS:

Ausgangspunkt ist hier der im Kapitel 5.3.3 umstrukturierte Datensatz. Für diesen wird zunächst die Variable `score` in Ränge gewandelt und erhält den Namen `Rscore`, bevor der Datensatz dann wieder in die Ausgangsform zurücktransformiert wird (vgl. Anhang 1.2). Dabei wird `Rscore` für die 3 Zeitstufen zu `Rscore.1`, `Rscore.2`, `Rscore.3`, Schließlich wird dann für diese Variablen wie im vorigen Kapitel die parametrische Varianzanalyse durchgeführt.

```
Rank variables=score (A)
  /Rank into Rscore.

Sort cases by Vpn Zeit.

Casestovars
  /Id=Vpn
  /Index=Zeit
  /Groupby=variable.
```

Hier der Datensatz nach der erneuten Umstrukturierung:

	Vpn	Geschlecht	score.1	score.2	score.3	Rscore.1	Rscore.2	Rscore.3
1	1	1	4	7	2	14,000	26,000	8,500
2	2	1	3	5	1	11,500	18,000	3,500
3	3	1	7	9	6	26,000	29,500	22,500
4	4	1	6	6	2	22,500	22,500	8,500
5	5	1	5	5	1	18,000	18,000	3,500
6	6	2	8	2	5	28,000	8,500	18,000
7	7	2	4	1	1	14,000	3,500	3,500
8	8	2	6	3	4	22,500	11,500	14,000
9	9	2	9	5	2	29,500	18,000	8,500
10	10	2	7	1	1	26,000	3,500	3,500

```
GLM Rscore.1 Rscore.2 Rscore.3
  /wsfactor=Zeit 3 polynomial
  /wsdesign=Zeit
  /design=Geschlecht.
```

Nachfolgend zunächst der Test auf Sphärität (Varianzhomogenität), danach die Ergebnisse der Varianzanalyse für den Effekt des Gruppierungsfaktors und zuletzt die Effekte des Messwiederholungsfaktors (Innersubjekteffekte). Bei diesen interessieren nur die Zeilen „Sphärität angenommen“, da der entsprechende Mauchly-Test keine Signifikanzen aufweist.

Mauchly-Test auf Sphärität ^a						
Innersubjekteffekt	Mauchly W	Approxim. Chi-Quadrat	df	Sig.	Epsilon ^b	
					Greenhouse- Geisser	Huynh-Feldt
Zeit	,986	,098	2	,952	,986	1,000

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Konstanter Term	7207,500	1	7207,500	82,156	,000
Geschlecht	53,333	1	53,333	,608	,458
Fehler	701,833	8	87,729		

Quelle		Quadrat- summe	df	Mittel der Quadrate	F	Sig.
Zeit	Sphärizität angen.	698,600	2	349,300	22,358	,000
	Greenhouse-Geisser	698,600	1,973	354,140	22,358	,000
	Huynh-Feldt	698,600	2,000	349,300	22,358	,000
	Untergrenze	698,600	1,000	698,600	22,358	,001
Zeit * Geschlecht	Sphärizität angen.	501,267	2	250,633	16,043	,000
	Greenhouse-Geisser	501,267	1,973	254,106	16,043	,000
	Huynh-Feldt	501,267	2,000	250,633	16,043	,000
	Untergrenze	501,267	1,000	501,267	16,043	,004
Fehler(Zeit)	Sphärizität angen.	249,967	16	15,623		
	Greenhouse-Geisser	249,967	15,781	15,839		
	Huynh-Feldt	249,967	16,000	15,623		
	Untergrenze	249,967	8,000	31,246		

Tabelle 6-4

Die Ergebnisse weichen nicht nennenswert von denen der o.a. parametrischen Analyse ab. Weitere Voraussetzungen brauchen hier nicht geprüft werden.

6.4 Verallgemeinerte Kruskal-Wallis- und Friedman-Tests (KWF)

Bei diesem Verfahren wird die Rangtransformation gegenüber dem o.a. RT-Verfahren anders vorgenommen: Zum einen erhält jede Erhebungseinheit (Vpn) einen Rang, zum anderen werden wie beim Friedman-Test pro Vpn Ränge für die einzelnen Messwiederholungen vergeben. Beide Ränge werden dann zu einem zusammengefasst. Darüber hinaus werden nicht die F-Tests verwendet, sondern aus den Streuungsquadratsummen (SS, Sum of Sq) werden χ^2 -Tests konstruiert. Die Tests der Haupteffekte (Geschlecht und Zeit) sind mit denen von Kruskal-Wallis bzw. von Friedman identisch.

Folgende Schritte sind für eine Analysevariable x durchzuführen:

- Im Folgenden ist K =Anzahl der gesamten Messwiederholungen.
Die Anzahl der Analysevariablen ist im Beispiel 4 genau eine.
- Für die Analyse-Variable x (Variablen x_1, \dots, x_k) die Summe aller Messwiederholungen (Sum) errechnen
- Diese Summe Sum in Ränge (R_{Sum}) umrechnen.
- Für jede Erhebungseinheit (Versuchsperson) werden die Werte x_1, \dots, x_k in Ränge $(1, \dots, K)$ transformiert und ergeben R_{x1}, \dots, R_{xk} .
- Für jede Messwiederholung $i=1, \dots, K$
 $(R_{Sum} - 1) * k + R_{xi}$
berechnen
- Mit diesen Rängen wird eine parametrische Varianzanalyse mit Messwiederholungen durchgeführt.
- Auf Basis der Anova-Tabelle werden folgende χ^2 -Tests aufgestellt:
Für die Effekte ohne Messwiederholungsfaktoren, z.B. A, B, A*B (vgl. Formel 2-6b):

$$\chi^2 = \frac{SS_{Effekt}}{MS_{zwischen}}$$

und für die Effekte (Haupteffekte und Interaktionen) mit Messwiederholungsfaktoren z.B. C, D, A*C, A*D, B*C, ...A*B*C,... (vgl. Formel 2-7):

$$\chi^2 = \frac{SS_{\text{Effekt}}}{(SS_X + SS_{\text{Fehler}})/(df_X + df_{\text{Fehler}})}$$

wobei

- SS_{Effekt} die Streuungsquadratsumme (Sum of Squares) des zu testenden Effektes,
- MS_{zwischen} die Varianz der gesamten Zwischensubjektstreuung (MS, Mean Square), die die Streuung aller Gruppierungsfaktoren und deren Interaktionen sowie der damit verbundenen Fehlerstreuung beinhaltet,
- SS_{Fehler} die Streuungsquadratsumme des zum Effekt gehörenden Fehlers ist sowie
- SS_X die Streuungsquadratsummen aller Effekte, die SS_{Fehler} als Fehlerterm haben, also insbesondere der zu testende Effekt SS_{Effekt} sowie Interaktionen mit allen Gruppierungsfaktoren,
- df die entsprechenden Freiheitsgrade.

(Der Nenner der χ^2 -Tests für die Messwiederholungseffekte entspricht genau $MS_{\text{innerhalb}}$, also der Varianz innerhalb der Versuchspersonen.)

- Die χ^2 -Werte sind dann in den Tabellen für den χ^2 -Test auf Signifikanz zu überprüfen, wobei die Freiheitsgrade die Zählerfreiheitsgrade (df_{Effekt}) des entsprechenden F-Tests sind.
- Die χ^2 -Werte für die Tests, die ausschließlich Messwiederholungsfaktoren beinhalten, also C, D, C*D, sollten alternativ gemäß Iman & Davenport (vgl. Formel 2-1b) in F-Werte umgerechnet werden. In diesem Fall entspricht dies:

$$F = \frac{(N-1)\chi^2}{df_X + df_{\text{Fehler}} - \chi^2}$$

wobei df_X und df_{Fehler} die o.a. Freiheitsgrade sind.

Die Schritte sollen wiederum für ein 2-faktorielles Design am Datensatz des Beispiels 4 (`winer518`) demonstriert werden. Das oben beschriebene Verfahren ist auch auf 3- und mehrfaktorielle Versuchspläne anwendbar. Bei mehreren Messwiederholungsfaktoren sind die Ausführungen von Kapitel 5.4.3 zu beachten.

mit R:

Auch hier wird zunächst die elementare Berechnung vorgestellt. Die ersten Schritte sind weitgehend dieselben wie in Kapitel 5.3.1.

- Die Summe der Variablen `t1`, `t2`, `t3` errechnen und diese in Ränge (`Rsum`) wandeln.
- Eine Fallkennzeichnung, hier `vpn` genannt, erzeugen.
- Die beiden neuen Variablen in den Dataframe einfügen.
- `Geschlecht` und `vpn` als „factor“ deklarieren.
- Dataframe `winer518` mittels der Funktion `reshape` die für Messwiederholungen erforderliche Struktur geben, mit den neuen Variablen `zeit` als Faktor und abhängige Variable `score`.

- Die neue Variable `zeit` als Faktor deklarieren.
- Die Messwiederholungsvariablen pro `Vpn` in Ränge `Rscore` umrechnen.
- Aus `Rsum` und `Rscore` die zu analysierende Variable `Ry` bilden.
- Schließlich wird die Anova mit `aov` oder `ezANOVA` durchgeführt.
(Falls die χ^2 -Werte „mit der Hand“ ausgerechnet werden, empfiehlt sich die Verwendung von `aov`. Sollen dagegen die Berechnung in R programmiert werden, ist `ezANOVA` vorzuziehen.)

```
Rsum      <- rank(rowSums(winer518[,3:5]))
Vpn       <- 1:10
winer518  <- cbind(Vpn, Rsum, winer518)
winer518  <- within(winer518,
                    {Geschlecht<-factor(Geschlecht); Vpn<-factor(Vpn)})
winer518t<- reshape(winer518,direction="long",timevar="Zeit",
                    v.names="score", varying=c("t1","t2","t3"),idvar="Vpn")
winer518t<- within(winer518t, Zeit<-factor(Zeit))
Rscore    <- ave(winer518t$score,winer518t$Vpn,FUN=rank)
Ry        <- (Rsum-1)*3 + Rscore
aov3      <- aov(Ry~Geschlecht*Zeit+Error(Vpn/Zeit),winer518t)
summary(aov3)
```

bzw. mit `ezANOVA`, wobei zu beachten ist, dass alle verwendeten Variablen Teil des angegebenen Dataframes sein müssen (während `aov` da weniger penibel ist und auch andere Variablen akzeptiert, sofern sie die passende Länge haben) und dass mit `detailed=T` die Streuungsquadratsummen ausgegeben werden:

```
winer518t <- cbind(winer518t,Rscore,Ry)
ezANOVA(winer518t,Ry,Vpn,within=Zeit,between=Geschlecht,detailed=T)
```

Die Ausgabe von `aov`:

Error: Vpn						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Geschlecht	1	24.3	24.3	0.089	0.773	
Residuals	8	2176.2	272.0			
Error: Vpn:Zeit						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Zeit	2	8.6	4.300	34.4	1.61e-06	***
Geschlecht:Zeit	2	7.4	3.700	29.6	4.20e-06	***
Residuals	16	2.0	0.125			

Tabelle 6-5

In diesem Fall ist es am einfachsten, die χ^2 -Werte aus den Spalten „Sum Sq“ und „Df“ „mit der Hand“ auszurechnen:

$$MS_{\text{zwischen}} = \frac{24,3 + 2176,2}{1 + 8} = 244,5$$

$$\chi_{\text{Geschlecht}}^2 = \frac{24,3}{244,5} = 0,1$$

$$\chi_{\text{Zeit}}^2 = \frac{8,6}{(8,6 + 7,4 + 2,0)/(2 + 2 + 16)} = 9,56$$

$$\chi_{\text{Interaktion}}^2 = \frac{7,4}{(8,6 + 7,4 + 2,0)/(2 + 2 + 16)} = 8,22$$

Die 5%- bzw. 1%-Schranken der χ^2 -Verteilung liegen bei 3,8 bzw. 6,6 (df=1) sowie bei 6,0 bzw. 9,2 (df=2). Somit sind der Effekt *zeit* sowie die Interaktion stark signifikant.

Die Ausgabe der Anova-Tabelle von `ezANOVA` (zum Vergleich):

\$ANOVA							
	Effect	DFn	DFd	SSn	SSd	F	p
1	(Intercept)	1	8	7207.5	2176.2	26.49572650	8.771197e-04
2	Geschlecht	1	8	24.3	2176.2	0.08933002	7.726474e-01
3	Zeit	2	16	8.6	2.0	34.40000000	1.606176e-06
4	Geschlecht:Zeit	2	16	7.4	2.0	29.60000000	4.199689e-06

Hier bezeichnen *ssn* die Sum of Squares des jeweiligen Effekts und *ssa* die Streuung des dazugehörigen Fehler- (Residuen) terms. Bei diesem vergleichsweise einfachen Design sind die χ^2 -Werte für die Effekte der Gruppierungs- wie auch der Messwiederholungsfaktoren gleich aufgebaut. Liegen allerdings mehrere Gruppierungsfaktoren vor, ist das Prozedere etwas schwieriger, da bei MS_{zwischen} mehr als Effekt- und Residuenstreuung zu berücksichtigen sind. Auf entsprechende R-Anweisungen soll hier verzichtet werden.

Die Umrechnung der χ^2 -Werte in F-Werte gemäß Iman & Davenport erübrigt sich hier, da diese nur für den Effekt *zeit* vorgenommen werden kann, was bereits früher gezeigt wurde.

Alternativ kann auch die Funktion `np.anova` (vgl. Anhang 3.6) angewandt werden. Der Aufruf ist praktisch identisch mit dem der Standardfunktion `aov`. Basis ist auch hierfür der umstrukturierte Datensatz (`winer518t`). Eingabe und Ausgabe:

```
np.anova(score~Geschlecht*Zeit+Error(Vpn/Zeit), winer518t)
```

generalized Kruskal-Wallis/Friedman (Puri & Sen) tests including Iman & Davenport F-tests							
	Df	Sum Sq	Chisq	Pr(>Chi)	F value	Pr(>F)	
Geschlecht	1	24.3	0.1064	0.74424			
Residuals Btw. Vpn	8	2030.4					
Zeit	2	8.6	9.5556	0.00841	8.2340	0.003478	**
Geschlecht:Zeit	2	7.4	8.2222	0.01639	6.2830	0.009686	**
Residuals	16	2.0					

mit SPSS:

Ausgangspunkt ist der Beispieldatensatz 4. Folgende Schritte sind zu machen:

- Errechnen der Summe der Messwiederholungsvariablen (*sum*) Transformation in Ränge (*rsum*).
- Umstrukturieren des Datensatzes, so dass aus den 3 Messwiederholungen jeweils 3 Fälle erzeugt werden. Das ist im Anhang 1.1.1 ausführlich beschrieben.
- Pro *Vpn* aus den Werten von *score* die Ränge *r_{score}* errechnen.

- Aus R_{Sum} und R_{score} die zu analysierende Variable R_y errechnen.
- Zurücktransformieren des Datensatzes wie in Kapitel 6.2.2., wobei aus R_y für die 3 Zeitpunkte die Variablen $R_{y.1}$, $R_{y.2}$, $R_{y.3}$ entstehen.
- Durchführen der Varianzanalyse

Die hierfür erforderlichen SPSS-Anweisungen:

```
compute sum=t1+t2+t3.

rank variables=Sum (A)
  /rank into RSum.

Varstocases
  /Id=Vpn
  /make score from t1 t2 t3
  /index=Zeit(3)
  /keep=Geschlecht Sum RSum
  /null=keep.

rank variables=score(A) by Vpn
  /rank into RScore.

compute Ry=(RSum-1)*3 + RScore.

Casestovars
  /Id=Vpn
  /Index=Zeit
  /Groupby=variable.

GLM Ry.1 Ry.2 Ry.3
  /wsfactor=Zeit 3 polynomial
  /wsdesign=Zeit
  /design=Geschlecht.
```

Nachfolgend die Ergebnisse der Varianzanalyse, zunächst die Effekte des Messwiederholungsfaktors (Innersubjekteffekte), danach der Effekt des Gruppierungsfaktors (Zwischen-subjekt-effekte). Eine Prüfung von Voraussetzungen, etwa der Sphärizität, entfällt hier, so dass in der Anova-Tabelle nur die Zeilen mit den unkorrigierten F-Tests interessieren.

Quelle		Quadratsumme	df	Mittel der Quadrate	F	Sig.
Zeit	Sphärizität angen.	8,600	2	4,300	34,400	,000
Zeit * Geschlecht	Sphärizität angen.	7,400	2	3,700	29,600	,000
Fehler(Zeit)	Sphärizität angen.	2,000	16	,125		

Tabelle 6-6a

Quelle	Quadratsumme	df	Mittel der Quadrate	F	Sig.
Konstanter Term	7207,500	1	7207,500	26,496	,001
Geschlecht	24,300	1	24,300	,089	,773
Fehler	2176,200	8	272,025		

Tabelle 6-6b

Aus den Spalten „Quadratsumme“ und „df“ werden nun die χ^2 -Werte berechnet, zunächst aus Tabelle 6-6b für den Gruppierungsfaktor *Geschlecht*, danach aus Tabelle 6-6a für die Messwiederholungseffekte:

$$MS_{zwischen} = \frac{24,3 + 2176,2}{1 + 8} = 244,5$$

$$\chi_{Geschlecht}^2 = \frac{24,3}{244,5} = 0,1$$

$$\chi_{Zeit}^2 = \frac{8,6}{(8,6 + 7,4 + 2,0)/(2 + 2 + 16)} = 9,56$$

$$\chi_{Interaktion}^2 = \frac{7,4}{(8,6 + 7,4 + 2,0)/(2 + 2 + 16)} = 8,22$$

Die 5%- bzw. 1%-Schranken der χ^2 -Verteilung liegen bei 3,8 bzw. 6,6 (df=1) sowie bei 6,0 bzw. 9,2 (df=2). Somit sind der Effekt „Zeit“ sowie die Interaktion stark signifikant.

Die Umrechnung der χ^2 -Werte in F-Werte gemäß Iman & Davenport erübrigt sich hier, da diese nur für den Effekt *Zeit* vorgenommen werden kann, was bereits früher gezeigt wurde.

6.5 Aligned rank transform (ART)

Das Prinzip des Aligned rank transform-Tests wurde oben bereits erläutert (vgl. Kapitel 4.3.6 und 5.4.4). Würde man jedoch dasselbe Verfahren auf ein gemischtes Design anwenden, so erhielte man „merkwürdige“ Signifikanzen. Der Grund: der Effekt des Gruppierungsfaktors α_i lässt sich nicht vom Personeneffekt π_m trennen. Daher muss hier ein anderer Weg eingeschlagen werden (vgl. dazu Beasley [84]). Da es letztlich nur um einen „sauberen“ Test für die Interaktion geht, genügt es, nur für diesen das ART-Verfahren anzuwenden. Die Haupteffekte werden über die o.a. Rank transform Tests (Kapitel 6.3) ermittelt. Aber der Aufwand zur Überprüfung der Interaktion lohnt auch nur dann, wenn der RT hierfür eine Signifikanz ergab, da letztlich mit dem ART nur der liberalere RT abgesichert wird.

Auf Folgendes sei noch aufmerksam gemacht: Beasley [84] hat zwar auf die Vorzüge des ART im Fall von gemischten Modellen auch bei nichtsphärischen Kovarianzmatrizen und nichtnormalen Daten hingewiesen, dennoch haben Kowalchuk et al. [88] gezeigt, dass dies nicht mehr gilt, wenn die Kovarianzmatrizen nicht mehr gleich (homogen) sind. Allerdings empfiehlt sich nicht, hier den Box-Test durchzuführen, um diese Voraussetzung zu überprüfen, da der Box-Test selbst sehr viel mehr voraussetzt, so u.a. multivariate Normalverteilung, so dass der Test in diesem Zusammenhang letztlich unbrauchbar wird.

Hier ist es erforderlich, den einfachen Fall der 2-faktoriellen Analyse und die beiden Fälle der 3-faktoriellen Analyse getrennt zu behandeln. Hieraus lassen sich dann auch Lösungen für höher-faktorielle Versuchspläne ableiten.

6.5.1 Ein Gruppierungs- und ein Messwiederholungsfaktor

Die Schritte im Einzelnen:

- Durchführung einer (normalen) Anova mit Haupt- und Interaktionseffekten für die Ränge R_x der Kriteriumsvariablen x . Hieraus werden nur die Haupteffekte verwendet.
- per *naive approach* (vgl. Formel 2-4): Eliminieren des Hauptffekts γ_j der Messwiederho-

lungen sowie des Personeneffekts π_m aus der Kriteriumsvariablen x :

$$e_{jm} = x_{jm} - (\pi_m + \gamma_j - \mu)$$

alternativ per *standard approach* (vgl. Formel 2-5): Berechnung der Residuen e_{jm} wie in Kapitel 6.2, anschließend Addition des „reinen“ Interaktionseffekts:

$$e_{jm} = e_{jm} + \alpha\gamma_{ij} - (\pi_m + \gamma_j - \mu)$$

- Umrechnung der so errechneten Residuen e_{jm} in Ränge.
- Durchführung einer Anova mit Haupt- und Interaktionseffekten mit den Rängen, aus der dann der Interaktionseffekt abgelesen werden kann.

Als Beispiel soll nachfolgend wieder der bereits verwendete Datensatz 4 dienen.

mit R:

Ausgangsbasis ist wieder der in Kapitel 5.3.1 erstellte Dataframe `winer518t`. In Kapitel 6.3 wurde der Rank transform Test durchgeführt, aus dem die Haupteffekte abzulesen sind (Tabelle 6-3).

Für die Berechnung der Residuen e_{jm} wird hier der o.a. „naive approach“ gewählt. Dazu müssen zunächst die Effekte γ_j (mb) und π_m (mp) sowie der Gesamtmittelwert (mm) berechnet werden, um sie von der Kriteriumsvariablen `score` abzuziehen. Diese werden dann nach Rundung auf 6 Stellen in Ränge transformiert, um darauf die Varianzanalyse anzuwenden.

```
attach(winer518t)
mb <- tapply(score, Zeit, mean)
mp <- tapply(score, Vpn, mean)
mm <- mean(score)
ek <- score
n <- dim(winer518t)[1]
for (k in 1:n) {j=Zeit[k]; i=Vpn[k]
  ek[k] <- ek[k] - mb[j] - mp[i] + mm }
ek <- rank(round(ek, digits=6))
summary(aov(ek~Geschlecht*Zeit+Error(Vpn/Zeit), winer518t))
```

Die Anova-Tabelle zeigt einen signifikanten Interaktionseffekt, während die anderen beiden Haupteffekte keine Bedeutung haben:

Error: Vpn						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Geschlecht	1	2.133	2.1333	2.265	0.171	
Residuals	8	7.533	0.9417			
Error: Vpn:Zeit						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Zeit	2	0.2	0.1	0.002	0.998	
Geschlecht:Zeit	2	1550.9	775.4	18.132	7.72e-05	***
Residuals	16	684.3	42.8			

Alternativ kann auch die Funktion `art3.anova` (vgl. Anhang 3.9) angewandt werden. Der Aufruf ist praktisch identisch mit dem der Standardfunktion `aov`. Basis ist auch hierfür der umstrukturierte Datensatz (`winer518t`). Eingabe und Ausgabe:

```
art3.anova(score~Geschlecht*Zeit+Error(Vpn/Zeit), winer518t)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Geschlecht	1	53.33	53.33	0.6079	0.4580116
Zeit	2	698.60	349.30	22.3582	2.325e-05 ***
Geschlecht:Zeit	2	820.87	410.43	16.9440	0.0001119 ***

Der Unterschied für das Ergebnis der Interaktion im Vergleich zur vorigen Tabelle liegt in den unterschiedlichen ART-Verfahren begründet, oben der „naive approach“, bei dem einfach von den beobachteten Werten der Zeit- und der Vpn-Effekt abgezogen werden, und in der Funktion `art3.anova` der „standard approach“, der auf den Residuen basiert. Über den zusätzlichen Parameter `method=1` kann aber auch die Berechnung nach dem „naive approach“ angefordert werden.

mit SPSS:

Wie bei der Durchführung der Rank transform-Tests muss zunächst der Datensatz umstrukturiert werden, wobei die Messwiederholungen in Fälle gewandelt werden. Dies wurde bereits in Kapitel 6.3 einmal durchgeführt. Für die Berechnung der Residuen e_{jm} wird hier der o.a. „naive approach“ gewählt.

Über `Aggregate` werden nun die Mittelwerte für Personen (`mp`), Zeit (`mb`) und gesamt (`mm`) berechnet und in der Arbeitsdatei ergänzt, um die Effekte von den Werten der Kriteriumsvariablen `score` abzuziehen und das Ergebnis in Ränge umzurechnen:

```
Aggregate
  /outfile=* mode=addvariables
  /break=id /mp=mean(score).
Aggregate
  /outfile=* mode=addvariables
  /break=Zeit /mb=mean(score).
Aggregate
  /outfile=* mode=addvariables
  /break= /mm=mean(score).

Compute ek = score - (mp + mb - mm).

Rank variables=ek (A) /rank into rek.
execute.
```

Anschließend wird der Datensatz wieder in die ursprüngliche Form transformiert:

```
Sort cases by Vpn Zeit.
Casestovars
  /Id=Vpn
  /index=Zeit
  /groupby=variable.
```

Schließlich wird dann für `rek`, die im umstrukturierten Datensatz die Namen `rek.1`, `rek.2`, ... hat, eine Varianzanalyse mit Messwiederholungen mit den Faktoren `Geschlecht` und `Zeit` gerechnet:

```
GLM rek.1 rek.2 rek.3 by Geschlecht
  /wsfactor=Zeit 3 Polynomial
  /wsdesign=Zeit
  /design=Geschlecht.
```

Nachfolgend die Anova-Tabelle der Variablen `rek.1` . . . für den bereinigten Test der Interaktion, wobei nur die Zeilen „Sphärizität angenommen“ relevant sind. Demnach ist die Signifikanz der Interaktion gesichert.

Tests der Innersubjekteffekte						
Quelle		Quadrat summe	df	Mittel der Quadrate	F	Sig.
Zeit	Sphärizität angen.	,200	2	,100	,002	,998
Zeit * Geschlecht	Sphärizität angen.	1550,867	2	775,433	18,132	,000
Fehler(Zeit)	Sphärizität angen.	684,267	16	42,767		

6.5.2 Ein Gruppierungs- und zwei Messwiederholungsfaktoren

Der Gruppierungsfaktor wird mit A, die beiden Messwiederholungsfaktoren mit C und D bezeichnet. Die Schritte im Einzelnen:

- Durchführung einer (normalen) Anova mit Haupt- und Interaktionseffekten für die Ränge R_x der Kriteriumsvariablen x (vgl. Kapitel 5.4.2). Hieraus werden nur die Haupteffekte verwendet. Für die Haupteffekte der Messwiederholungsfaktoren C und D können allerdings auch die bereinigten Tests wie in Kapitel 5.4.4 errechnet werden.
- Die Interaktion C*D, ein reiner Messwiederholungseffekt, wird mit der ART wie in Kapitel 5.4.4 ermittelt, wobei Faktor A außer Acht gelassen wird.
- Für die Interaktion A*C, ein gemischter Interaktionseffekt, werden die Werte der Kriteriumsvariablen x über die Stufen von Faktor D gemittelt (oder summiert), um mit diesen Werten die ART wie im vorhergehenden Kapitel 6.5.1 durchzuführen.
- Für die Interaktion A*D ist das Verfahren analog der Interaktion A*C durchzuführen.

Ein bereinigter Test für die 3er Interaktion A*C*D ist kein entsprechendes Verfahren bekannt.

Als Beispiel soll nachfolgend der bereits verwendete Datensatz 5 dienen:

- Die Haupteffekte `Medikament` und `Aufgabe` wurden bereits mit dem Rank transform-Test in Kapitel 5.4.2 ermittelt. Dort ist es kein Problem, auch den Faktor `Geschlecht` miteinzubeziehen.
- Der Interaktionseffekt `Medikament*Aufgabe` wurde in Kapitel 5.4.4 ermittelt.
- Bleiben noch die Interaktionen `Geschlecht*Medikament` und `Geschlecht*Aufgabe`, von denen nur die erste hier behandelt wird, da das Verfahren für beide identisch ist.

mit R:

Ausgangsbasis ist der in Kapitel 5.4.1 erstellte und in 5.4.4 verwendete Dataframe `mydata5t`. Zunächst werden mittels `aggregate` die Summen von `Fehler` über die 3 Aufgabenstufen berechnet. Dabei entsteht ein neuer Dataframe (`mydata5s`) mit den Mittelwerten als Variable x .

```

      Vpn Geschlecht Medikament      x
1      1          1           1 2.3333333
2      2          1           1 0.6666667
3      3          1           1 4.0000000
4      4          1           1 3.3333333
5      5          2           1 1.6666667
6      6          2           1 1.6666667
7      7          2           1 2.0000000
8      8          2           1 1.3333333
9      1          1           2 3.3333333
10     2          1           2 2.3333333
..     ..          ..          ..

```

Für die Berechnung der Residuen e_{jm} (vgl. vorigen Abschnitt) müssen zunächst die Effekte μ_j (mb) und π_m (mp) sowie der Gesamtmittelwert (mm) berechnet werden, um diese von der Kriteriumsvariablen x abzuziehen. Diese werden dann nach Rundung auf 6 Stellen in Ränge transformiert, um darauf die Varianzanalyse anzuwenden. Hierfür wird diesmal wieder `ezANOVA` verwendet, wobei zu beachten ist, dass alle verwendeten Variablen Teil des angegebenen Dataframes sein müssen. D.h. in diesem Fall muss die neu erzeugte Variable `ez` mit `cbind` angehängt werden.

```

data(mydata5t)
library(ez)
mydata5s <- aggregate(Fehler,
                      mydata5t[,c("Vpn", "Geschlecht", "Medikament")], mean)
attach(mydata5s)
mb <- tapply(x, Medikament, mean)
mp <- tapply(x, Vpn, mean)
mm <- mean(x)
ek <- x
n <- dim(mydata5s)[1]
for (k in 1:n) {j=Medikament[k]; i=Vpn[k]
  ek[k] <- ek[k] - mb[j] - mp[i] + mm }
ek <- rank(round(ek, digits=6))
ezANOVA(cbind(mydata5s, ek), ek, Vpn,
        within=. (Medikament), between=. (Geschlecht)) $ANOVA

```

Das Ergebnis für die Interaktion ist nicht signifikant. Hätte man sich diese Interaktion beim Rank transform-Test (RT) angeschaut, hätte man sich die Durchführung des ART hierfür sparen können.

	Effect	DFn	DFd	F	p	p < .05
2	Geschlecht	1	6	0.14555256	0.7159674	
3	Medikament	2	12	0.04571522	0.9554795	
4	Geschlecht:Medikament	2	12	0.62084221	0.5538958	

mit SPSS:

Wie bei der Durchführung der Rank transform-Tests muss zunächst der Datensatz umstrukturiert werden, wobei die Messwiederholungen in Fälle gewandelt werden. Dies wurde bereits in Kapitel 5.4.2 einmal durchgeführt und in 5.4.4 wieder verwendet. Zunächst werden mittels `aggregate` die Mittelwerte von `Fehler` über die 3 Aufgabenstufen berechnet. Die Syntax dafür sowie ein Ausschnitt der Ergebnismatrix (`mydata5s`):

```

Dataset Declare mydata5s.
Aggregate
  /outfile='mydata5s'
  /break=id Geschlecht Medikament
  /Mfehler=mean(Fehler) .

```

	id	Geschlecht	Medikament	Mfehler
1	1	1	1	2,33
2	1	1	2	3,33
3	1	1	3	4,00
4	2	1	1	,67
5	2	1	2	2,33
6	2	1	3	3,33
7	3	1	1	4,00
8	3	1	2	3,67
9	3	1	3	4,33
10	4	1	1	3,33

Über `Aggregate` werden nun die Mittelwerte für Personen (μ_i), Zeit (μ_j) und gesamt (μ) berechnet, um die Effekte von den Werten der Kriteriumsvariablen `Mfehler` abzuziehen und das Ergebnis in Ränge umzurechnen. Die Anweisungen hierfür sind weitgehend identisch mit denen des vorigen Abschnitts. Lediglich `score` ist durch `Mfehler` zu ersetzen.

Anschließend wird der Datensatz mit den Anweisungen wie im vorigen Abschnitt wieder in die ursprüngliche Form transformiert.

Schließlich wird dann für `rek`, die im umstrukturierten Datensatz die Namen `rek.1`, `rek.2`, ... hat, eine Varianzanalyse mit Messwiederholungen mit den Faktoren `Geschlecht` und `Medikament` gerechnet (Anweisungen siehe voriger Abschnitt). Nachfolgend die Anova-Tabelle für den bereinigten Test der Interaktion, wobei nur die Zeilen „Sphärität angenommen“ relevant sind. Demnach liegt für die Interaktion keine Signifikanz vor.

Tests der Innersubjekteffekte						
Quelle		Quadrat summe	df	Mittel der Quadrate	F	Sig.
Medikament	Sphärität angen.	7,750	2	3,875	,046	,955
Medikament * Geschlecht	Sphärität angen.	110,583	2	55,292	,656	,537
Fehler(Medikament)	Sphärität angen.	1012,167	12	84,347		

6.5.3 Zwei Gruppierungs- und ein Messwiederholungsfaktor

Die Gruppierungsfaktoren werden mit A und B, der Messwiederholungsfaktor mit C bezeichnet, die Effekte mit α_i , β_j bzw. γ_l . Die Schritte im Einzelnen:

- Durchführung einer (normalen) Anova mit Haupt- und Interaktionseffekten für die Ränge R_x der Kriteriumsvariablen x (vgl. Kapitel 5.4.2). Hieraus werden nur die Haupteffekte verwendet.
- Für die Interaktion $A*B$, ein Effekt ohne Messwiederholungen, werden die Werte der Kriteriumsvariablen x über die Stufen von Faktor C gemittelt (oder summiert), um mit diesen Werten die ART wie im Kapitel 4.3.6 durchzuführen.
- Für die Interaktionen $A*C$ und $B*C$ sind zunächst die Residuen e_k des kompletten Modells zu berechnen (vgl. Kapitel 6.2).

- Für die Interaktion A*C sind zu den Residuen zunächst der Interaktionseffekt zu addieren und danach der Messwiederholungseffekt γ_l sowie der Personeneffekt π_m zu subtrahieren:

$$e_k(a) = e_k + \alpha\gamma_{il} - (\pi_m + \gamma_l - \mu)$$

- Für die Interaktion B*C wird analog A*C vorgegangen.

$$e_k(b) = e_k + \beta\gamma_{jl} - (\pi_m + \gamma_l - \mu)$$

- Umrechnung der so errechneten Residuen $e_k(a)$ sowie $e_k(b)$ in Ränge.
- Durchführung einer Anova mit Haupt- und Interaktionseffekten jeweils mit den Rängen $R(e_k(a))$ bzw. $R(e_k(b))$, aus der dann der jeweilige Interaktionseffekt abgelesen werden kann.

Ein bereinigter Test für die 3er Interaktion A*B*C ist kein entsprechendes Verfahren bekannt.

Das Verfahren soll am Datensatz 6 demonstriert werden. Die Anova-Tabelle der 3-faktoriellen Varianzanalyse für Rx, aus der die Haupteffekte A, B, und Zeit abzulesen sind:

	Effect	DFn	DFd		F	p	p < .05	ges
2	A	1	8	3.3160388	0.1060896			0.22755888
3	B	1	8	8.1885856	0.0211004	*		0.42112020
5	Zeit	3	24	235.4228709	0.0000000	*		0.89487936
4	A:B	1	8	0.1732461	0.6881851			0.01515789
6	A:Zeit	3	24	25.8348420	0.0000001	*		0.48298681
7	B:Zeit	3	24	4.8246813	0.0090990	*		0.14854504
8	A:B:Zeit	3	24	0.9709958	0.4226642			0.03392018

Tabelle 6-7

D.h. die Haupteffekte B und Zeit sind signifikant, insbesondere aber auch die Interaktionen A*Zeit sowie B*Zeit, die nun mittels dem ART gesondert berechnet werden. Zur Demonstration soll allerdings auch die Interaktion A*B untersucht werden, wenn dies auch nicht erforderlich ist.

mit R:

Zunächst muss wieder der Dataframe `winer568` umstrukturiert und die Faktoren als solche deklariert werden. Das Ergebnis wird `winer568t` benannt. Anschließend werden für die Analyse der Interaktionen A*C und B*C die Residuen des Modells $A*B*C+V_{pn}$ ermittelt:

```
winer568t <- reshape(winer568, direction="long", timevar="Zeit",
  v.names="x", varying=c("V1", "V2", "V3", "V4"), idvar="Vpn")
winer568t <- within(winer568t, {A<-factor(A); B<-factor(B);
  Zeit<-factor(Zeit); Vpn<-factor(Vpn) })
ek <- aov(x~A*B*Zeit+Vpn, winer568t)$residuals
```

Anschließend werden die Effekte für die beiden untersuchten Interaktionen (`mac` bzw. `mbc`), die Zeit (`mc`) sowie den Personeneffekt `mv` ausgerechnet und gemäß o.a. Formel mit den Residuen `ek` verrechnet, um schließlich für `eka` und `ekb` eine Varianzanalyse durchzuführen:

```
attach(winer568t)
mc <- tapply(x, Zeit, mean)
mv <- tapply(x, Vpn, mean)
mac <- tapply(x, winer568t[, c("A", "Zeit")], mean)
mbc <- tapply(x, winer568t[, c("B", "Zeit")], mean)
mm <- mean(x)
```

```

n <- dim(winer568t) [1]
eka <- ek
ekb <- ek
for (k in 1:n) {ia=A[k]; ib=B[k]; ic=Zeit[k]; m=Vpn[k]
  eka[k] <- eka[k] + mac[ia,ic] - mc[ic] -mv[m] + mm
  ekb[k] <- ekb[k] + mbc[ib,ic] - mc[ic] -mv[m] + mm }
reka<-rank(round(eka,digits=7))
rekb<-rank(round(ekb,digits=7))
library(ez)
ezANOVA(cbind(winer568t,reka),reka,Vpn,
         between=.(A,B),within=.(Zeit))$ANOVA
ezANOVA(cbind(winer568t,rekb),rekb,Vpn,
         between=.(A,B),within=.(Zeit))$ANOVA

```

Bei der Varianzanalyse für `reka` zeigt der Mauchly-Test auf Varianzhomogenität mit $p=0,029$ eine signifikante Abweichung an. Aber in jedem Fall ist vorsichtshalber in der Anova-Ausgabe die Signifikanz im Teil `Sphericity Corrections` und dort unter „p[HF]“ (Huynh-Feldt-korrigiert) abzulesen, allerdings *ausschließlich* für die Interaktion A*Zeit (auf die Tabelle für `ekb` wird hier verzichtet). Der p-Wert (0,00006) bestätigt den oben mit dem RT-Test errechneten Einfluss von A*Zeit:

\$`Sphericity Corrections`					
	Effect	GGe	p [GG]	HFe	p [HF]
5	Zeit	0.4925664	0.9606032485	0.5774698	0.9751581
6	A:Zeit	0.4925664	0.0001875066	0.5774698	0.0000645
7	B:Zeit	0.4925664	0.7383084419	0.5774698	0.7730265
8	A:B:Zeit	0.4925664	0.8874259252	0.5774698	0.9150948

Nun zur Interaktion A*B.

- Zunächst werden mittels `aggregate` die Summen von v_1, \dots, v_4 über die 4 Zeitstufen berechnet. Dabei entsteht ein neuer Dataframe (`winer568s`) mit den Mittelwerten als Variable `x`.
- Wie in Kapitel 4.3.6 werden die Effekte `mab` (Interaktion), `ma` (Faktor A) sowie `mb` (Faktor B) errechnet.
- Ermittlung der Residuen `ek` der Varianzanalyse des Modells A*B,
- Addition bzw. Subtraktion der vorher errechneten Effekte von `ek`,
- Durchführung der Varianzanalyse für `ek` zur Kontrolle des Effekts A*B:

```

winer568s <- aggregate(winer568t$x, winer568t[,c("Vpn","A","B")], mean)
attach(winer568s)
ma <- tapply(x,A,mean)
mb <- tapply(x,B,mean)
mab <- tapply(x,list(A,B),mean)
mm <- mean(x)
ek <- aov(x~A*B,winer568s)$residuals
n <- dim(winer568s) [1]
for (k in 1:n) {ia=A[k]; ib=B[k]
  ek[k] <- ek[k] + mab[ia,ib] - ma[ia] - mb[ib] + mm }
rek <- rank(ek)
summary(aov(ek~A*B,winer568s))

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	2.08	2.083	0.122	0.736
B	1	0.33	0.333	0.020	0.892
A:B	1	0.75	0.750	0.044	0.839
Residuals	8	136.33	17.042		

Auch hier ist - wie bereits in Kapitel 6.5.1 - das ART-Verfahren mit der Funktion `art3.anova` (vgl. Anhang 3.9) bequem durchführbar. Basis ist auch hier der umstrukturierte Datensatz `winer568t`. Nachfolgend Eingabe und Ausgabe:

```
art3.anova(x~A*B*Zeit+Error(Vpn/Zeit),winer568t)
```

	Df	Sum of Sq	F value	Pr(>F)
A	1	18.8	3.2609	0.108588
B	1	75.0	13.0435	0.006866 **
A:B	1	0.7	0.0440	0.839079
Zeit	3	6637.2	235.4229	< 2.2e-16 ***
A:Zeit	3	3528.8	22.7165	3.421e-07 ***
B:Zeit	3	1764.9	6.8443	0.001714 **

mit SPSS:

Zunächst muss wieder der Datensatz aus Beispiel 6 wie in Kapitel 6.2 umstrukturiert werden, wobei `vpn` die Vpn-Kennzeichnung ist. Anschließend werden für die Analyse der Interaktionen A*C und B*C die Residuen (Variable `res_1`) des Modells (ohne Messwiederholungen) `A*B*C+Vpn` ermittelt:

```
Unianova x by Vpn A B Zeit
  /Save=resid
  /design=A*B*Zeit Vpn.
```

Anschließend werden die Effekte für die beiden untersuchten Interaktionen (`mac` bzw. `mbc`), die Zeit (`mc`) sowie den Personeneffekt `mv` ausgerechnet, der Arbeitsdatei angehängt und gemäß o.a. Formel mit den Residuen `ek` verrechnet, um schließlich für `eka` und `ekb` eine Varianzanalyse durchzuführen:

```
Aggregate
  /outfile=* mode=addvariables
  /break=id /mp=mean(x).
Aggregate
  /outfile=* mode=addvariables
  /break=Zeit /mc=mean(x).
Aggregate
  /outfile=* mode=addvariables
  /break=A Zeit /mac=mean(x).
Aggregate
  /outfile=* mode=addvariables
  /break=B Zeit /mbc=mean(x).
Aggregate
  /outfile=* mode=addvariables
  /break= /mm=mean(x).
Compute eka = res_1 + mac - (mp + mc - mm).
Compute ekb = res_1 + mbc - (mp + mc - mm).
Rank variables=eka (A) /rank into reka.
Rank variables=ekb (A) /rank into rekb.
execute.
```

Nun wird wie Kapitel 6.3 der Datensatz in die ursprüngliche Form zurücktransformiert. Dabei werden aus den zu analysierenden Rängen von eka und ekb die Messwiederholungsvariablen $reka.1, \dots, reka.4$ bzw. $rekb.1, \dots, rekb.4$. Bei der Varianzanalyse mit Messwiederholungen für $reka$ zeigt der Mauchly-Test mit $p=0,027$ eine signifikante Abweichung von der Varianzhomogenität. Vorsichtshalber sollte in jedem Fall die Signifikanz des Effekts in der Zeile „Huynh-Feldt“ abgelesen werden. Allerdings kann aus der Tabelle *ausschließlich* der Effekt $A \cdot \text{Zeit}$ entnommen werden. Der p -Wert ($< 0,001$) bestätigt den oben mit dem RT-Test errechneten Einfluss von $A \cdot \text{Zeit}$. (Auf die Ausgabe für $rekb$ wird hier verzichtet):

Tests der Innersubjekteffekte						
Quelle		Quadratsumme	df	Mittel der Quadrate	F	Sig.
Zeit	Sphärizität angen.	2,250	3	,750	,015	,998
	Greenhouse-Geisser	2,250	1,478	1,523	,015	,962
	Huynh-Feldt	2,250	2,412	,933	,015	,993
Zeit * A	Sphärizität angen.	3552,083	3	1184,028	23,039	,000
	Greenhouse-Geisser	3552,083	1,478	2403,814	23,039	,000
	Huynh-Feldt	3552,083	2,412	1472,632	23,039	,000
Zeit * B	Sphärizität angen.	38,167	3	12,722	,248	,862
	Greenhouse-Geisser	38,167	1,478	25,829	,248	,719
	Huynh-Feldt	38,167	2,412	15,823	,248	,821
Zeit * A * B	Sphärizität angen.	12,083	3	4,028	,078	,971
	Greenhouse-Geisser	12,083	1,478	8,177	,078	,873
	Huynh-Feldt	12,083	2,412	5,010	,078	,950

Nun zur Interaktion $A \cdot B$. Ausgangsbasis ist die oben erzeugte umstrukturierte Arbeitsdatei.

Zunächst werden mittels `aggregate` die Summen von v_1, \dots, v_4 über die 4 Zeitstufen berechnet. Dabei muss eine neue Datei mit den Mittelwerten als Variable mx angelegt werden.

```
Dataset Declare winer568s.
Aggregate      /outfile='winer568s'
               /break=id A B /mx=MEAN(x).
```

Ermittlung der Residuen (Variable Res_1) der Varianzanalyse des Modells $A \cdot B$.

Wie in Kapitel 4.3.6 werden die Effekte mab (Interaktion), ma (Faktor A) sowie mb (Faktor B) errechnet. Anschließend Addition bzw. Subtraktion der vorher errechneten Effekte von Res_1 :

```
Aggregate      /outfile=* mode=addvariables
               /break=A B /mab=mean(mx).
Aggregate      /outfile=* mode=addvariables
               /break=A /ma=mean(mx).
Aggregate      /outfile=* mode=addvariables
               /break=B /mb=mean(mx).
Aggregate      /outfile=* mode=addvariables
               /break= /mm=mean(mx).

Compute ek = res_1 + mab - (ma + mb - mm).

Rank variables=ek (A) /rank into rek.
execute.
```

Durchführung der Varianzanalyse für ϵ_k zur Kontrolle des Effekts A*B, wonach die Interaktion A*B nicht signifikant ist.

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
A	2,083	1	2,083	,122	,736
B	,333	1	,333	,020	,892
A * B	,750	1	,750	,044	,839
Fehler	136,333	8	17,042		

6.6 normal scores-Tests (INT)

Bei dem *normal score-* bzw. *inverse normal transform-Verfahren* (INT) werden lediglich die Werte der abhängigen Variablen x über alle Messwiederholungen und Gruppen hinweg zunächst in Ränge $R(x)$ gewandelt und anschließend in normal scores umgerechnet:

$$nscore_i = \Phi^{-1}(R(x_i)/(KN + 1))$$

wobei K die Anzahl der Messwiederholungen und N die Anzahl der Merkmalsträger ist, also KN die Anzahl der Beobachtungen. Mit diesen scores wird dann eine „normale“ parametrische Varianzanalyse gerechnet. Auch hier sollte man den Mauchly-Test durchführen, um die korrigierten F-Tests von Huynh & Feldt zu benutzen, falls die Spherizität nicht gegeben ist. Dieses Verfahren soll wieder am Beispieldatensatz 4 demonstriert werden. Die Ergebnisse zeigen, dass dieses Verfahren „besser“ abschneidet als das einfachere Rank transform (RT).

mit R:

Ausgangsbasis ist wieder der in Kapitel 5.3.1 erstellte Dataframe `winer518t`. Die Berechnung der normal scores erfolgt genauso wie Kapitel 5.3.4. Die Varianzanalyse enthält allerdings hier zusätzlich den Test des Faktor `Geschlecht`. Die Analyse wird mit `ezANOVA` (Paket `ez`) durchgeführt. Die Varianzhomogenität (Spherizität) war schon in Kapitel 5.3.4 bestätigt worden:

```
library(ez)
ezANOVA(winer518t, nscore, Vpn, within=Zeit, between=Geschlecht)
```

	Effect	DFn	DFd	F	p	p < .05	ges
2	Geschlecht	1	8	0.4589120	5.172406e-01		0.04306605538
3	Zeit	2	16	26.1823940	9.001193e-06	*	0.41354670288
4	Geschlecht:Zeit	2	16	19.4945215	5.137485e-05	*	0.34428051545

mit SPSS:

Die Schritte im Einzelnen:

- Zunächst muss der Datensatz umstrukturiert werden, so dass aus den 3 Messwiederholungen jeweils 3 Fälle erzeugt werden. Das ist im Anhang 1.1.1 ausführlich beschrieben. Dabei wird die abhängige Variable `score` gebildet.
- Über Aggregate wird die Anzahl der Werte `nc` ermittelt.
- Die Werte werden in Ränge umgerechnet.

- Über die inverse Normalverteilung (`Idf.normal`) werden die Ränge in normal scores umgerechnet.
- Der Datensatz wird zurück in die ursprüngliche Form transformiert. Daraus resultieren aus `nscore` die Variablen `nscore.1,...`
- Schließlich kann die parametrische Varianzanalyse auf die Variablen `nscore.1,...` angewandt werden.

Die Syntax hierfür sowie nachfolgend die Ausgabe der Anova-Tabellen. Die Varianzhomogenität (Sphärität) war schon in Kapitel 5.3.4 bestätigt worden, so dass für die Messwiederholungseffekte nur die Zeilen „Sphärität angenommen“ relevant sind.

```

Varstocases
  /Id=Vpn
  /Make score from t1 t2 t3
  /index=Zeit(3)
  /keep=Geschlecht
  /null=keep.

Aggregate
  /outfile=* mode=addvariables
  /break= /nc=NU(score).
Rank Variables=score / rank into Rscore.
compute nscore=Idf.normal(Rscore/(nc+1),0,1).
Sort cases by Vpn Zeit.

Casestovars
  /Id=Vpn
  /index=Zeit
  /groupby=variable.

GLM nscore.1 nscore.2 nscore.3
  /wsfactor=Zeit 3 polynomial
  /wsdesign=Zeit
  /design=Geschlecht.

```

Tests der Zwischensubjekteffekte					
Quelle	Quadratsumme	df	Mittel der Quadrate	F	Sig.
Konstanter Term	,003	1	,003	,003	,955
Geschlecht	,441	1	,441	,459	,517
Fehler	7,686	8	,961		

Tests der Innersubjekteffekte						
Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Zeit	Sphärität angen.	6,909	2	3,454	26,182	,000
Zeit * Geschlecht	Sphärität angen.	5,144	2	2,572	19,495	,000
Fehler(Zeit)	Sphärität angen.	2,111	16	,132		

6.7 van der Waerden-Tests

Das Verfahren von van der Waerden verläuft zunächst ähnlich mit dem KWF-Verfahren (vgl. Kapitel 6.4). D.h. zum einen erhalten die einzelnen Fälle Ränge (R_{Sum}) entsprechend der Summe der Messwiederholungen, und zum anderen werden die Werte der Messwiederholungen pro Fall analog dem Friedman-Test in Ränge R_{xi} transformiert. Die Ränge werden jeweils in normal scores umgerechnet. Beide scores werden addiert. Schließlich werden die χ^2 -Tests wie beim KWF-Verfahren durchgeführt. Wie schon in Kapitel 2.6 erwähnt gibt es für den Fall mehrerer Messwiederholungsfaktoren kein entsprechendes Verfahren.

Folgende Schritte sind für eine Analysevariable x durchzuführen:

- Im Folgenden ist k =Anzahl der gesamten Messwiederholungen. (Die Anzahl der Analysevariablen ist im Beispiel 4 genau eine.)
- Für die Analyse-Variable x (Variablen x_1, \dots, x_k) pro Fall die Summe aller Messwiederholungen (S_{um}) errechnen.
- Diese Summe S_{um} in Ränge (R_{Sum}) umrechnen.
- Umrechnung von R_{Sum} in normal scores: $n_{sum} = \Phi^{-1}(R_{sum}/(N+1))$, wobei N die Anzahl der Fälle ist.
- Für jede Erhebungseinheit (Versuchsperson) werden die Werte x_1, \dots, x_k in Ränge $(1, \dots, K)$ transformiert und ergeben R_{x1}, \dots, R_{xk} .
- Umrechnung von R_{xi} in normal scores: $n_{score} = \Phi^{-1}(R_{xi}/(K+1))$.
- Für jede Messwiederholung $i=1, \dots, K$
 $n_{sx} = n_{sum} + n_{score}$
 berechnen
- Mit diesen normal scores wird eine parametrische Varianzanalyse mit Messwiederholungen durchgeführt.
- Auf Basis der Anova-Tabelle werden folgende χ^2 -Tests aufgestellt:
 Für die Effekte ohne Messwiederholungsfaktoren, z.B. A, B, A*B (vgl. Formel 2-6b):

$$\chi^2 = \frac{SS_{Effekt}}{MS_{zwischen}}$$

und für die Effekte (Haupteffekte und Interaktionen) mit Messwiederholungsfaktoren z.B. C, D, A*C, B*C, A*D, ..., A*B*C, ... (vgl. Formel 2-7):

$$\chi^2 = \frac{SS_{Effekt}}{(SS_X + SS_{Fehler}) / (df_X + df_{Fehler})}$$

wobei

- SS_{Effekt} die Streuungsquadratsumme (Sum of Squares) des zu testenden Effektes ,
- $MS_{zwischen}$ die Varianz der gesamten Zwischensubjektstreuung (MS, Mean Square), die die Streuung aller Gruppierungsfaktoren und deren Interaktionen sowie der damit verbundenen Fehlerstreuung beinhaltet,
- SS_{Fehler} die Streuungsquadratsumme des zum Effekt gehörenden Fehlers ist sowie
- SS_X die Streuungsquadratsummen aller Effekte, die SS_{Fehler} als Fehlerterm haben, also insbesondere der zu testende Effekt SS_{Effekt} sowie Interaktionen mit allen Gruppierungsfaktoren,

- df die entsprechenden Freiheitsgrade.
- Die χ^2 -Werte sind dann in den Tabellen für den χ^2 -Test auf Signifikanz zu überprüfen, wobei die Freiheitsgrade die Zählerfreiheitsgrade (df_{Effekt}) des entsprechenden F-Tests sind.

Die Schritte sollen wiederum am Datensatz des Beispiels 4 demonstriert werden. Da Überprüfung der Sphärizität kann entfallen, da hier χ^2 - anstatt F-Tests durchgeführt werden

mit R:

Auch hier wieder zunächst die elementare Berechnung, anschließend unter Verwendung einer R-Funktion für dieses Verfahren. Ausgangsbasis ist der Dataframe `winer518`. Die Schritte zur Erlangung der Anova-Tabelle, mit deren Hilfe die χ^2 -Tests errechnet werden können, sind weitgehend identisch mit denen aus Kapitel 6.4. Zusätzlich wird zunächst die Anzahl der Merkmalsträger `nc` ermittelt, mit deren Hilfe die normal scores `nsum` für die Merkmalsträger berechnet werden. Ebenso werden die normal scores `nscore` für die 3 Messwiederholungen berechnet. Die Summe aus beiden zusammen bilden die normal scores `nsx`, auf deren Basis die Varianzanalyse durchgeführt wird:

```
Rsum      <- rank(rowSums(winer518[,3:5]))
nc        <- dim(winer518)[1]
nsum      <- qnorm(Rsum/(nc+1))
Vpn       <- 1:10
winer518  <- cbind(winer518, Vpn, Rsum, nsum)
winer518  <- within(winer518,
  {Geschlecht<-factor(Geschlecht); Vpn<-factor(Vpn)})
winer518t<- reshape(winer518, direction="long", timevar="Zeit",
  v.names="score", varying=c("t1","t2","t3"), idvar="Vpn")
winer518t<- within(winer518t, Zeit<-factor(Zeit))
Rscore    <- ave(winer518t$score, winer518t$Vpn, FUN=rank)
nscore    <- qnorm(Rscore/4)
nsx       <- nsum + nscore
aov3      <- aov(nsx~Geschlecht*Zeit+Error(Vpn/Zeit), winer518t)
summary(aov3)
```

Zunächst die Ausgabe der (parametrischen) Anova:

```
Error: Vpn
      Df Sum Sq Mean Sq F value Pr(>F)
Geschlecht  1  0.368  0.3681    0.165  0.695
Residuals   8 17.833  2.2291

Error: Vpn:Zeit
      Df Sum Sq Mean Sq F value Pr(>F)
Zeit      2  3.847  1.9237   33.81 1.8e-06 ***
Geschlecht:Zeit  2  3.331  1.6657   29.27 4.5e-06 ***
Residuals   16  0.910  0.0569
```

Nun zur Berechnung der χ^2 -Werte:

Aus dem oberen Teil der Anova-Tabelle ist zu entnehmen:

$$MS_{zwischen} = \frac{0,368 + 17,833}{1 + 8} = 2,022$$

$$\chi^2_{Geschlecht} = \frac{0,368}{2,022} = 0,182$$

Aus dem unteren Teil der Anova-Tabelle ist zu entnehmen:

$$\chi_{Zeit}^2 = \frac{3,847}{(3,847 + 3,331 + 0,910)/(2 + 2 + 16)} = \frac{3,847}{0,4044} = 9,513$$

$$\chi_{Interaktion}^2 = \frac{3,331}{(3,847 + 3,331 + 0,910)/(2 + 2 + 16)} = \frac{3,331}{0,4044} = 8,24$$

Die 5%- bzw. 1%-Schranken der χ^2 -Verteilung liegen bei 3,8 bzw. 6,6 (df=1) sowie bei 6,0 bzw. 9,2 (df=2). Somit sind der Effekt „Zeit“ sowie die Interaktion stark signifikant.

Alternativ kann auch die Funktion `np.anova` (vgl. Anhang 3.6) angewandt werden. Der Aufruf ist praktisch identisch mit dem der Standardfunktion `aov`. Über den Parameter `method=1` wird das van der Aarden-Verfahren ausgewählt. Basis ist auch hierfür der umstrukturierte Datensatz (`winer518t`). Eingabe und Ausgabe:

```
np.anova(score~Geschlecht*Zeit+Error(Vpn/Zeit), winer518t, method=1)
```

generalized van der Waerden tests				
	Df	Sum Sq	Chisq	Pr(>Chi)
Geschlecht	1	0.3681	0.1944	0.65929
Residuals Btw.Vpn	8	16.6742		
Zeit	2	3.8475	9.5124	0.00860
Geschlecht:Zeit	2	3.3315	8.2367	0.01627
Residuals	16	0.9104		

mit SPSS:

Ausgangspunkt ist der Beispieldatensatz 4. Folgende Schritte sind erforderlich:

- Errechnen der Summe der Messwiederholungsvariablen (`sum`)
- Transformation der Summe in Ränge (`Rsum`).
- Ermitteln der Anzahl der Fälle (`nc`) mittels `Aggregate`.
- Umwandeln von `Rsum` in normal scores (Variable `nsum`) mittels `Idf.normal`.
- Umstrukturieren des Datensatzes, so dass aus den 3 Messwiederholungen jeweils 3 Fälle erzeugt werden. Das ist im Anhang 1.1.1 ausführlich beschrieben. Daraus resultiert die abhängige Variable `score`.
- Pro `Vpn` aus den Werten von `score` die Ränge `Rscore` errechnen.
- Umrechnen in normal score `nscore` mittels `Idf.normal`.
- Aus `nsum` und `nscore` die zu analysierende Variable `nsx` als deren Summe errechnen.
- Zurücktransformieren des Datensatzes wie in Kapitel 6.2.2., wobei aus `nsx` für die 3 Zeitpunkte die Variablen `nsx.1`, `nsx.2`, `nsx.3` entstehen.
- Durchführen der Varianzanalyse für die Variablen `nsx.1`, `nsx.2`, `nsx.3`.
- Berechnung der χ^2 -Werte gemäß Formeln 2-6 bzw. 2-7.

Die hierfür erforderlichen SPSS-Anweisungen:

```

compute sum=t1+t2+t3.
rank variables=Sum (A)
  /rank into RSum.
Aggregate
  /outfile=* mode=addvariables
  /break= /nc=NU(RSum) .
compute nsum=Idf.normal (RSum/(nc+1) , 0, 1) .

Varstocases
  /Id=Vpn
  /Make score from t1 t2 t3
  /index=Zeit(3)
  /keep=Geschlecht Sum RSum nsum nc
  /null=keep.

Rank Variables=score by Vpn / rank into Rscore.
compute nscore=Idf.normal (Rscore/4, 0, 1) .
compute nsx=nsum+nscore.
Sort cases by Vpn Zeit.

Casestovars
  /Id=Vpn
  /index=Zeit
  /groupby=variable.

GLM nsx.1 nsx.2 nsx.3
  /wsfactor=Zeit 3 polynomial
  /wsdesign=Zeit
  /design=Geschlecht.
    
```

Zunächst die Ausgabe der (parametrischen) Anova:

Tests der Zwischensubjekteffekte					
Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Konstanter Term	9,006E-005	1	9,006E-005	,000	,995
Geschlecht	,368	1	,368	,165	,695
Fehler	17,833	8	2,229		

Tabelle 6-9a

Tests der Innersubjekteffekte						
Quelle		Quadratsumme	df	Mittel der Quadrate	F	Sig.
Zeit	Sphärizität angen.	3,847	2	1,924	33,808	,000
Zeit * Geschlecht	Sphärizität angen.	3,331	2	1,666	29,274	,000
Fehler(Zeit)	Sphärizität angen.	,910	16	,057		

Tabelle 6-9b

Aus Tabelle 6-9a ist zu entnehmen:

$$MS_{zwischen} = \frac{0,368 + 17,833}{1 + 8} = 2,022$$

$$\chi^2_{Geschlecht} = \frac{0,368}{2,022} = 0,182$$

Aus Tabelle 6-9b ist zu entnehmen:

$$\chi_{Zeit}^2 = \frac{3,847}{(3,847 + 3,331 + 0,910)/(2 + 2 + 16)} = \frac{3,847}{0,4044} = 9,513$$

$$\chi_{Interaktion}^2 = \frac{3,331}{(3,331 + 0,910)/(2 + 16)} = \frac{3,331}{0,2356} = 14,14$$

Die 5%- bzw. 1%-Schranken der χ^2 -Verteilung liegen bei 3,8 bzw. 6,6 (df=1) sowie bei 6,0 bzw. 9,2 (df=2). Somit sind der Effekt „Zeit“ sowie die Interaktion stark signifikant.

6.8 ATS-Tests von Akritas, Arnold & Brunner

Die von Akritas, Arnold und Brunner entwickelten ATS-Test gibt es auch für mehrfaktorielle Varianzanalysen mit gemischten Designs. Während in R dazu das Paket `nparLD` zur Verfügung steht, gibt es in SPSS derzeit keine Möglichkeit zur Anwendung dieses Verfahrens.

mit R:

Die 2-faktorielle Analyse mittels `nparLD` soll ebenfalls am Datensatz des Beispiels 4 gezeigt werden. Ausgangsbasis ist wieder der oben (vgl. 5.4.1) erstellte umstrukturierte Dataframe `winer518t`. Die Analyse kann mittels zwei Funktionen erfolgen:

- `nparLD` ist eine universelle Funktion für alle verarbeitbaren Designs.
- `f1.ld.f1` erlaubt fehlende Werte bei den Messwiederholungen, gibt einen Mittelwertplot aus sowie eine Reihe weiterer, hier allerdings nicht interessierender Statistiken. (Darüber hinaus gibt es entsprechende Funktionen für 3-faktorielle Designs: `f2.ld.f1` für zwei Gruppierungs- und einen Messwiederholungsfaktor sowie `f1.ld.f2` für einen Gruppierungs- und zwei Messwiederholungsfaktoren.)

Beide geben sowohl die WTS als auch die ATS aus. Die Ausgabe unterscheidet sich nicht hinsichtlich der Wiedergabe dieser Statistiken. Nachfolgend zunächst die Eingabe für beide Varianten, wobei zu beachten ist, dass bei `nparLD` trotz Angabe des Dataframes die Variablennamen nicht automatisch gefunden werden. Daher muss bei beiden Funktionen entweder jeder Variablenname zusammen mit dem Dataframennamen in der üblichen Form, z.B. `winer518t$score` angegeben werden oder ein `attach` ausgeführt werden:

```
library(nparLD)
attach(winer518t)
nparLD(score~Geschlecht*Zeit, winer518t, Vpn)
f1.ld.f1(score, Zeit, Geschlecht, Vpn,
         time.name="Zeit", group.name="Geschlecht") -> ano
round(ano$ANOVA.test, 3)
```

Bei `f1.ld.f1` müssen die Faktoren zweimal angegeben werden: zum einen zur Identifikation des Faktors, zum anderen in „...“ als Name des Faktors für die Ausgabe. Diese Funktion gibt noch zusätzlich einen Interaktionsplot aus, allerdings der relativen Effekte (vgl. Kapitel 2.5) anstatt der Mittelwerte, da sich ja die Hypothesen auf erstere beziehen:

Die Ergebnisse von `nparLD`:

Wald-Type Statistic (WTS):			
	Statistic	df	p-value
Geschlecht	0.6079316	1	4.355677e-01
Zeit	40.2018842	2	1.863253e-09
Geschlecht:Zeit	36.3186594	2	1.298683e-08
ANOVA-Type Statistic (ATS):			
	Statistic	df	p-value
Geschlecht	0.6079316	1.000000	4.355677e-01
Zeit	22.3581811	1.972665	2.515147e-10
Geschlecht:Zeit	16.0426724	1.972665	1.281568e-07

Bei der Ausgabe von `f1.l.d.f1` gibt es die Möglichkeit, einzelne Teile auszugeben, etwa die ATS- (Anova-) Tabelle (`.$ANOVA.test`) oder die WTS- (Wald-Test-) Tabelle (`..wald.test`). Dies hat den Vorteil, dass man über die Funktion `round` die Zahlendarstellung der Art `xxe-nn` ändern kann.

	Statistic	df	p-value
Geschlecht	0.6079	1.0000	0.4356
Zeit	22.3582	1.9727	0.0000
Geschlecht:Zeit	16.0427	1.9727	0.0000

6.9 Bredenkamp Tests

Zunächst sei noch einmal darauf hingewiesen, dass die Tests von Bredenkamp (vgl. [1], S. 1024 ff und [15]) letztlich mit den verallgemeinerten Kruskal-Wallis- und Friedman-Tests (vgl. u.a. Kapitel 6.4) identisch sind. Lediglich die Berechnung erfolgt auf einem anderen Weg. Insbesondere für SPSS-Benutzer können die Tests von Bredenkamp bei gemischten Versuchsplänen von Nutzen sein, da zum einen die ATS aus dem vorigen Kapitel nicht zur Verfügung stehen und zum anderen bei diesen Tests keine Umstrukturierungen der Daten erforderlich sind. Daher werden nachfolgend Beispiele nur mit SPSS gerechnet. Hinzu kommt, dass es in R erhebliche Schwierigkeiten bereitet, Friedman-Tests für Teildatensätze durchzuführen, was bei den Bredenkamp Tests erforderlich ist.

Für das Prozedere werden die einzelnen Versuchspläne unterschieden. Im ersten Fall der 2-faktoriellen Analyse wird das Grundprinzip gezeigt und in den 3-faktoriellen Analysen dann erweitert.

6.9.1 Ein Gruppierungs- und ein Messwiederholungsfaktor

Die drei Effekte (Gruppierungsfaktor A, Messwiederholungsfaktor B sowie die Interaktion) werden wie folgt überprüft:

- Haupteffekt A: pro Erhebungseinheit (z.B. Versuchsperson) wird die Summe aller Messwiederholungen errechnet. Hierauf wird dann der Kruskal-Wallis-H-Test angewandt.
- Haupteffekt B: ein Friedman-Test wird über die Messwiederholungen durchgeführt, wobei die Gruppeneinteilung durch den Faktor A ignoriert wird.
- Interaktion: Unter Ausnutzung der Additivität der χ^2 -Werte wird für jede Stufe des Faktors A ein Friedman-Test für B errechnet, die resultierenden χ^2 -Werte aufsummiert, davon der χ^2 -Wert des Friedman-Tests des Haupteffekts B abgezogen und schließlich der Restwert

anhand der Tabelle der χ^2 -Verteilung auf Signifikanz überprüft:.

χ^2 -Testwerte	Freiheitsgrade
$\chi^2_B(A_1)$	$l-1$
$+ \chi^2_B(A_2)$	$l-1$
$+ \dots$	\dots
$+ \chi^2_B(A_k)$	$l-1$
$- \chi^2_B$	$l-1$
$Summe(\chi^2_B(A_i)) - \chi^2_B$	$(k-1)(l-1)$

mit SPSS:

Es wird wieder der Datensatz 4 (winer518) benutzt. Zunächst muss das Skalenniveau der Variablen t1, t2, t3 auf „Skala“ gesetzt werden, anschließend deren Summe tsum errechnet, damit der Kruskal-Wallis-Test zum Test des Geschlechtseffekts sowie der Friedman-Test für t1, t2, t3 zum Test des Zeiteffekts durchgeführt werden können. Danach wird wiederum der Friedman-Test durchgeführt, allerdings dann mittels Split File für die beiden Geschlechtsgruppen getrennt.

```
compute tsum=t1+t2+t3.
Nptests
  /Independent test (tsum) group (Geschlecht) kruskal_wallis.
Nptests
  /Related test(t1 t2 t3) friedman.

Sort cases by Geschlecht.
Split File separate by Geschlecht.
Nptests
  /Related test(t1 t2 t3) friedman.
```

Die Ausgabe zeigt zunächst links den K-W-Test (für „Geschlecht“), rechts den Friedman-Test (für „Zeit“):

Gesamtanzahl	10	Gesamtanzahl	10
Teststatistik	,099	Teststatistik	9,556
Freiheitsgrade	1	Freiheitsgrade	2
Asymptotische Sig. (zweiseitiger Test)	,753	Asymptotische Sig. (zweiseitiger Test)	,008

Die Teststatistiken (χ^2 -Werte) für die beiden Friedman-Test zur Ermittlung der Interaktion:

$$\chi^2_{Zeit}(\text{Männer}) = 9,333 \quad (2 \text{ Fg})$$

$$\chi^2_{Zeit}(\text{Frauen}) = 8,444 \quad (2 \text{ Fg})$$

Zieht man von der Summe 17,777 (4 Fg) den o.a. $\chi^2_{Zeit}(\text{gesamt})$ mit dem Wert 9,556 (2 Fg) ab, so erhält man $\chi^2_{Interaktion} = 8,222$ mit 2 Fg. Die 5%-Schranke für die χ^2 -Verteilung liegt

bei 6,0 für $df=2$, so dass die Interaktion als signifikant angesehen werden kann. Die Ergebnisse decken sich mit denen aus Kapitel 6.4.

6.9.2 Zwei Gruppierungs- und ein Messwiederholungsfaktor

Das Prinzip aus dem vorigen Abschnitt wird nun auf drei Faktoren erweitert. Allerdings können die Bredenkamp Tests nur für balancierte Versuchspläne (vgl. Kapitel 4.3.1.1) angewandt werden. Im Folgenden werden die beiden Gruppierungsfaktoren mit A und B (mit Gruppenzahl I bzw. J), der Messwiederholungsfaktor mit C (mit Gruppenzahl K) bezeichnet. Die Effekte werden wie folgt überprüft:

- Haupteffekte A und B:
pro Erhebungseinheit (z.B. Versuchsperson) wird die Summe aller Messwiederholungen errechnet. Hierauf wird dann jeweils für A und B der Kruskal-Wallis-H-Test angewandt.
- Interaktion A*B: Zunächst wird ein H-Test über alle Zellen hinweg gerechnet. Von diesem χ^2 -Wert werden die Werte aus den H-Tests für Faktor A und Faktor B subtrahiert. Das Ergebnis ist der χ^2 -Wert für die Interaktion A*B. Analog werden die Freiheitsgrade ermittelt.
- Haupteffekt C: ein Friedman-Test wird über die Messwiederholungen durchgeführt, wobei die Gruppeneinteilung durch die Faktoren A und B ignoriert wird.
- Interaktion A*C: Unter Ausnutzung der Additivität der χ^2 -Werte wird für jede Stufe des Faktors A ein Friedman-Test für C errechnet, die resultierenden χ^2 -Werte aufsummiert, davon der χ^2 -Wert des Friedman-Tests des Haupteffekts C abgezogen und schließlich der Restwert anhand der Tabelle der χ^2 -Verteilung auf Signifikanz überprüft.
- Interaktion B*C: analog Interaktion A*C.
- Interaktion A*B*C: für jede der $I*J$ Zellen von A*B wird ein Friedman-Test für C errechnet, die resultierenden χ^2 -Werte aufsummiert, davon die χ^2 -Werte des Haupteffekts C, der Interaktion A*C sowie der Interaktion B*C abgezogen und schließlich der Restwert anhand der Tabelle der χ^2 -Verteilung auf Signifikanz überprüft. Die dafür erforderlichen Freiheitsgrade errechnen sich analog zu den χ^2 -Werten.

mit SPSS:

Die Berechnungen sollen am Datensatz 6 (`winer568`) demonstriert werden. Zunächst muss das Skalenniveau der Variablen `v1`, ..., `v4` auf „Skala“ gesetzt werden, anschließend deren Summe `vsum` errechnet, womit zwei H-Tests zur Prüfung der Effekte A und B durchgeführt werden. Für `v1`, ..., `v4` wird eine Friedman-Analyse zum Test des Zeiteffekts gerechnet. Hier empfiehlt es sich, die „alten“ Anweisungen für die nichtparametrischen Tests (`Npar Tests`) zu verwenden, da bei diesen die Ergebnisse „direkt“ im Ausgabefenster angezeigt werden und nicht erst über ein Doppelklick in einem separaten Fenster erscheinen.

```
compute vsum=v1+v2+v3+v4.
Npar tests
  /K-W = vsum by A.
Npar tests
  /K-W = vsum by B.
Npar tests
  /Friedman = v1 v2 v3 v4.
```

Nachfolgend werden nur die χ^2 -Werte protokolliert:

$$\chi^2_A = 1,468 \quad (1 \text{ Fg})$$

$$\chi^2_B = 5,872 \quad (1 \text{ Fg})$$

$$\chi^2_{\text{Zeit}} = 32,635 \quad (3 \text{ Fg})$$

Für die Interaktion A*B wird zunächst eine Zellennummer Z_{elle} errechnet, für die Gruppierung dann ein H-Test bzgl. v_{sum} errechnet. Anschließend von der resultierenden Teststatistik die beiden o.a. Statistiken für A und B subtrahiert:

```
compute Zelle=(a-1)*2+b.
Npar tests
  /K-W = vsum by Zelle.
```

$$\chi^2_{\text{Zellen}} = 7,399 \quad (3 \text{ Fg})$$

$$\chi^2_{A*B} = \chi^2_{\text{Zellen}} - \chi^2_A - \chi^2_B = 7,399 - 1,468 - 5,872 = 0,059 \quad (1 \text{ Fg})$$

Für die Interaktion A*Zeit werden jeweils Friedman-Tests für die zwei Stufen von A errechnet, die resultierenden χ^2 -Werte addiert und davon der oben errechnete Wert χ^2_{Zeit} subtrahiert:

```
Sort cases by A.
Split File separate by A.
Npar tests
  /Friedman = v1 v2 v3 v4.
```

$$\chi^2_{\text{Zeit}}(A_1) = 16,932 \quad (3 \text{ Fg})$$

$$\chi^2_{\text{Zeit}}(A_2) = 17,357 \quad (3 \text{ Fg})$$

$$\chi^2_{A*\text{Zeit}} = \chi^2_{\text{Zeit}}(A_1) + \chi^2_{\text{Zeit}}(A_2) - \chi^2_{\text{Zeit}} = 16,932 + 17,357 - 32,635 = 1,654 \quad (3 \text{ Fg})$$

Analog erhält man für die Interaktion B*Zeit:

$$\chi^2_{\text{Zeit}}(B_1) = 16,158$$

$$\chi^2_{\text{Zeit}}(B_2) = 16,966$$

$$\chi^2_{B*\text{Zeit}} = \chi^2_{\text{Zeit}}(B_1) + \chi^2_{\text{Zeit}}(B_2) - \chi^2_{\text{Zeit}} = 16,158 + 16,966 - 32,635 = 0,489 \quad (3 \text{ Fg})$$

Für die Interaktion A*B*Zeit werden zuerst für alle 4 Zellen von A*B jeweils ein Friedman-Test für den Faktor Zeit gerechnet und die resultierenden Teststatistiken (χ^2 -Werte) addiert. Davon werden dann der oben errechnete Wert χ^2_{Zeit} sowie die χ^2 -Werte der Interaktionen A*Zeit und B*Zeit subtrahiert:

```
Sort cases by Zelle.
Split File separate by Zelle.
Npar tests
  /Friedman = v1 v2 v3 v4.
```

$$\chi^2_{\text{Zeit}}(A_1B_1) = 8,379 \quad (3 \text{ Fg})$$

$$\chi^2_{\text{Zeit}}(A_1B_2) = 9,000 \quad (3 \text{ Fg})$$

$$\chi^2_{\text{Zeit}}(A_2B_1) = 8,786 \quad (3 \text{ Fg})$$

$$\chi^2_{\text{Zeit}}(A_2B_2) = 8,786 \quad (3 \text{ Fg})$$

$$\begin{aligned}
\chi^2_{A*B*Zeit} &= \chi^2_{Zeit}(A_1B_1) + \chi^2_{Zeit}(A_1B_2) + \chi^2_{Zeit}(A_2B_1) + \chi^2_{Zeit}(A_2B_2) \\
&\quad - \chi^2_{Zeit} - \chi^2_{A*Zeit} - \chi^2_{B*Zeit} \\
&= 8,379 + 9,000 + 8,786 + 8,786 - 1,654 - 0,489 - 32,635 \\
&= 0,172 \quad (3 \text{ Fg})
\end{aligned}$$

Die p-Werte für die drei Haupteffekte werden in SPSS ja ausgegeben. Die χ^2 -Werte für die Interaktionen müssen mit den tabellierten kritischen Werten verglichen werden. Die 5%-Schranke für die χ^2 -Verteilung liegt bei 3,8 bzw. 9,0 für $df=1$ bzw. $df=3$, so dass keine Interaktion als signifikant angesehen werden kann. Vergleicht man diese Ergebnisse mit denen der ART-Methode (Kapitel 6.5.3), so zeigt sich deutlich, dass bei diesem Verfahren die Tests der Interaktionen relativ konservativ ausfallen.

6.10 Fazit

Auch hier gelten zunächst einmal die Ausführungen der Kapitel 4.5 und 5.5. Allerdings sind bei den gemischten Versuchsplänen noch Voraussetzungen hinzugekommen. Insbesondere setzen die Tests von Mauchly sowie von Box, die ja nur zur Prüfung von Voraussetzungen dienen, eigentlich multivariate Normalverteilungen der Messwiederholungsvariablen bzw. der Residuen voraus. An dieser Stelle möge man sich an die Bemerkungen des Kapitels 1.7. erinnern. R-Benutzer können diesem einfach mit den ATS von Akritas, Arnold und Brunner begegnen. Dieses Verfahren erfordert keine Voraussetzungen, hat eine gute Effizienz und steht als Package `nparLD` zur Verfügung. Der SPSS-Benutzer wird dagegen vielfach mit „Augen zu und durch“ handeln müssen.

Abschließend werden für die oben benutzten Datensätze die Ergebnisse aller Verfahren, und zwar die p-Werte für alle Effekte, in einer Tabelle gegenüber gestellt. Schließlich sollte man - wie schon oben gesagt - die hier erzielten Ergebnisse nicht verallgemeinern.

.Datensatz 4 (`winer518`) :

Verfahren	Geschlecht	Zeit	Interaktion
parametrisch	0.511	< 0.001	0.001
parametrisch - Greenhouse & Geisser		< 0.001	0.001
parametrisch - Huynh & Feldt		< 0.001	0.001
Rank transform Test (RT)	0.458	< 0.001	< 0.001
normal score (INT)	0.517	< 0.001	< 0.001
Aligned Rank Transform (ART)	0.171	< 0.001	< 0.001
Verall. Kruskal-Wallis-Friedman-Tests (KWF) Bredenkamp Tests	0.752	0.0084	0.0164
Verall. KWF-Tests mit Iman-Davenport-Korr.		0.0024	0.0048
van der Waerden	0.67	0.0086	0.0008
Akritas, Arnold & Brunner ATS	0.37	< 0.001	< 0.001

Datensatz 6 (winer568):

(nicht alle Ergebnisse wurden in den vorangegangenen Kapiteln protokolliert.).

Verfahren	A	B	Zeit	A*B	A*Zeit	B*Zeit
parametrisch	0.100	0.018	< 0.001	0.810	< 0.001	0.002
Rank transform Test (RT)	0.106	0.021	< 0.001	0.688	< 0.001	0.009
normal scores (INT)	0.251	0.015	< 0.001	0.718	< 0.001	0.104
Aligned Rank Transform(ART)	0.106	0.021	< 0.001	0.894	< 0.001	0.002
Verall. Kruskal-Wallis-Friedman-Tests (KWF) Bredenkamp Tests	0.227	0.015	< 0.001	0.809	0.650	0.921
Verall. KWF-Tests mit Iman-Davenport-Korr.			< 0.001		0.6982	0.935
van der Waerden	0.195	0.017	< 0.001	0.842	0.6405	0.926
Akritis, Arnold & Brunner ATS	0.069	0.004	< 0.001	0.677	< 0.001	0.008

Der eklatante Unterschied der verallgemeinerten Kruskal-Wallis-Friedman- und der van der Waerden-Tests gegenüber den anderen Verfahren hinsichtlich der Interaktionen A*Zeit und B*Zeit ist auf die geringe Residuenstreuung der Messwiederholungseffekte zurückzuführen. Diese geht bei der dort vorgenommenen Rangbildung zum Teil verloren.

7. Analysen für dichotome Merkmale

Für dichotome abhängige Variablen gibt es grundsätzlich zwei Möglichkeiten zur Durchführung einer Varianzanalyse: die oben beschriebenen Verfahren oder die weiter unten angeführte logistische Regression (siehe Kapitel 8.1).

Beispieldatensatz 7 (irish):

Hier wurden 1107 irische Schulkinder zu ihrer Einstellung und Gebrauch der irischen Sprache befragt. Erhoben wurden u.a.:

Variablenname	Bedeutung	Ausprägungen
(school) type	Schultyp	1=secondary (Gymnasium) 2=community (Mischung aus Gymn. und Berufsschule) 3=vocational (Berufsschule)
(school) location	Lage	1=urban (städtisch) 2=rural (ländlich)
sex	Geschlecht	1=male 2=female
income	Einkommen	1=high 2=medium 3=low
vocabula	Vokabular	1=bad 2=poor 3=good 4=excellent
usage	Nutzung	1=never 2=little 3=regular
attitude	Einstellung	1=negative 2=neutral 3=positive

Diesen Daten liegt kein Versuchsplan zugrunde, wie sonst vielfach bei Varianzanalysen. D.h. die Daten wurden erhoben, ohne dass darauf geachtet wurde, dass die Gruppierungsvariablen (Schultyp, Schullage und Geschlecht) orthogonal zueinander oder zumindest unabhängig voneinander sind. Dies erschwert Varianzanalysen insofern, als dass zum einen die Effekte nicht unabhängig voneinander sind und zum anderen die Hinzunahme z.B. von Interaktionseffekten die Tests der anderen Effekte deutlich beeinflusst und somit keine klare Interpretation der Effekte möglich ist. Konkret: Geschlecht und Schultyp sowie Schultyp und Einkommen sind voneinander abhängig. Da nicht orthogonale Faktoren aber bei Untersuchungen häufig der Fall ist, wurde dieser Datensatz bewusst als Gegenstück zu den bislang vorgestellten ausgewählt, die allesamt Versuchspläne beinhalten.

Die Daten wurden früher als Beispieldatensatz mit SPSS ausgeliefert. Die primäre Herkunft der Daten lässt sich nicht mehr klären.

Beispieldaten 8 (koch):

Bei diesem Datensatz handelt es sich um klinische Daten von 340 Patienten, die in ein Krankenhaus eingeliefert worden waren. Zu Beginn wurden die Leiden der Patienten in leicht (0) und schwer (1) klassifiziert (Faktor *severity*). Ein Teil der Patienten wurde daraufhin behandelt (Faktor *treat*). Anschließend wurden alle im Abstand von mehreren Tagen dreimal untersucht (Faktor *time*). Dabei wurde eine Person entweder als krank (0) oder normal (1) eingestuft (Variable *outcome*). Der Datensatz stammt von Koch [95] und umfasst eigentlich noch weitere Informationen, wie z.B. Behandlungen zwischen den Untersuchungsterminen. Deren Analyse würde jedoch eine Kovarianzanalyse erfordern. Daher werden diese hier nicht berücksichtigt. Im „Original“ liegt der Datensatz „umstrukturiert“ vor, d.h. die Werte der 3 Zeitpunkte als jeweils 3 Fälle. Nachfolgend ein Auszug:

	case_id	severity	treat	outcome	time013	t013trea	time012	t012trea
1	1	0	0	1	0	0	0	0
2	1	0	0	1	1	0	1	0
3	1	0	0	1	3	0	2	0
4	2	0	0	1	0	0	0	0
5	2	0	0	1	1	0	1	0
6	2	0	0	1	3	0	2	0
7	3	0	0	1	0	0	0	0
8	3	0	0	1	1	0	1	0
9	3	0	0	1	3	0	2	0
10	4	0	0	1	0	0	0	0

In der Standardform für Messwiederholungen sehen die ersten Fälle folgendermaßen aus:

	case_id	severity	treat	outcome.0	outcome.1	outcome.2
1	1	0	0	1	1	1
2	2	0	0	1	1	1
3	3	0	0	1	1	1
4	4	0	0	1	1	1
5	5	0	0	1	1	1

7.1 Anwendung der Verfahren für metrische Merkmale

Dichotome Merkmale verhalten sich vielfach wie metrische Merkmale. Simulationen haben gezeigt, dass man dichotome Variablen bei größeren Fallzahlen vielfach genauso handhaben kann wie metrische Variablen. So auch bei der Varianzanalyse (vgl. dazu Cochran, W.G. [7] und Lunney, G.H., [8].) Danach werden sowohl α -Level wie auch β eingehalten. Für das erforderliche n gilt: Liegen die relativen Häufigkeiten der beiden Ereignisse über 0,2, so genügen 20 Freiheitsgrade für den Fehlerterm, andernfalls sind mindestens 40 Freiheitsgrade erforderlich. Die Untersuchungen betrafen allerdings nur Versuchspläne mit gleichen Zellenbesetzungszahlen. Es ist allerdings zu vermuten, dass die Ergebnisse auch für ungleiche n_{ij} gültig sind. D'Agostino [91] sowie Cleary & Angel [92] haben die Untersuchungen von Lunney zwar bestätigt, allerdings etwas abgeschwächt mit der Bedingung, dass die relativen Häufigkeiten zwischen 0,25 und 0,75 liegen sollten, da andernfalls die Varianzen zu unterschiedlich werden können. Bogard [93] hat die wichtigste Literatur zu diesem Thema mit Zitaten zusammengestellt. Erstaunlicherweise gibt es hierzu kaum neuere Ergebnisse bzw. Veröffentlichungen. Im Gegensatz zur u.a. Logistischen Regression kann diese Vorgehensweise auch bei Messwiederholungen angewandt werden.

Tests auf Homogenität der Varianzen bzw. auf Sphärizität im Fall von Messwiederholungen entfallen hier, da die Varianzen $p(1-p)$ sich aus den Mittelwerten p errechnen lassen.

7.1.1 Unabhängige Stichproben

An dieser Stelle soll ein Beispiel gerechnet werden, und zwar für den Datensatz 7. Als Kriteriumsvariable wird `vocabulary` gewählt, allerdings dichotomisiert: 0=(1/bad, 2/poor) und 1=(3/good, 4/excellent). Als Faktoren: Geschlecht, Schultyp und Einkommen. Wegen der Pro-

blematik der Abhängigkeit der Faktoren, auf die bei der Beschreibung des Datensatzes kurz aufmerksam gemacht wurde, wird zum einen eine 2-faktorielle Varianzanalyse mit den Faktoren `sex` und `income` durchgeführt, da diese voneinander unabhängig sind. Der Einfluss von `type` wird wegen der Abhängigkeit von `sex` und `income` separat untersucht, wenn auch der Effekt des Schultyps vom Geschlecht und Einkommen ein wenig mitbeeinflusst wird. Die Interaktionen `sex*type` und `income*type` machen wegen der Abhängigkeit keinen Sinn. Die relativen Häufigkeiten des Kriteriums liegen mit 0,21 bzw. 0,68 im geforderten Bereich.

mit R:

Zunächst muss die 4-stufige abhängige Variable `vocabula` dichotomisiert werden (Variable `dvocabul`), bevor „wie gewohnt“ mit `aov` und `drop1` die parametrische Varianzanalyse darauf angewandt wird:

```
irish <- within(irish,dvocabul<-as.integer(vocabula)>2)
options (contrasts=c("contr.sum","contr.poly"))
drop1(aov(dvocabul~sex*income,irish), ~. ,test="F")
drop1(aov(dvocabul~type,irish), ~. ,test="F")
```

mit folgendem Ergebnis für die Analyse der Effekte von `sex` und `income`:

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
<none>			262.58	-1580.8		
sex	1	0.6298	263.21	-1580.2	2.6408	0.1044
income	2	12.5531	275.13	-1533.1	26.3175	6.843e-12 ***
sex:income	2	0.4187	263.00	-1583.0	0.8777	0.4160

sowie für die Analyse des Effekts von `type` :

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
<none>			261.52	-1591.3		
type	2	15.009	276.53	-1533.5	31.68	4.186e-14 ***

mit SPSS

Zunächst muss die 4-stufige abhängige Variable `vocabula` dichotomisiert werden (Variable `dvocabul`), bevor „wie gewohnt“ mit `Unianova` die parametrische Varianzanalyse darauf angewandt wird.

```
dvocabula <- vocabula gt 2.
Unianova dvocabula by Sex Income
  /Design = Sex Income Sex*Income.
Unianova dvocabula by Type
  /Design = Type.
```

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
sex	,630	1	,630	2,641	,104
income	12,553	2	6,277	26,317	,000
sex * income	,419	2	,209	,878	,416
Fehler	262,581	1101	,238		

type	15,009	2	7,505	31,680	,000
Fehler	261,524	1104	,237		

7. 1. 2 Gemischte Versuchspläne

Als Beispiel wird hier der Datensatz 8 von Koch verwendet, der zum einen eine dichotome abhängige Variable (*outcome*) und zum anderen 2 Gruppierungsfaktoren (*severity* und *treat*) sowie einen Messwiederholungsfaktor (*time*) beinhaltet. *outcome* hat mit 48 bzw. 52 Prozent ideale relative Häufigkeiten. Der Mauchly-Test auf Varianzhomogenität (genauer Sphärizität) entfällt hier wie oben bereits erläutert. Damit erübrigen sich auch die in Kapitel 5.1 erwähnten robusten Tests von Huynh & Feldt bzw. Greenhouse & Geisser.

mit R:

Der Datensatz muss zwar nicht umstrukturiert werden, jedoch die Variablen *severity*, *treat*, *time012* sowie *case_id* als Faktoren deklariert werden. Darüber hinaus muss gegebenenfalls *outcome* über *as.numeric* numerische Werte erhalten. Die Varianzanalyse erfolgt hier einmal über *ezANOVA*, wobei zu beachten ist, dass wegen ungleicher Zellenbesetzungszahlen über *type=3* die Quadratsummen vom Typ III angefordert werden müssen:

```
ezANOVA (koch, outcome, case_id, between=.(severity,treat),
         within=time012, type=3)
```

	Effect	DFn	DFd	F	p
2	severity	1	336	90.89621790	3.166354e-19
3	treat	1	336	40.81026220	5.591147e-10
5	time012	2	672	60.68707191	5.908176e-25
4	severity:treat	1	336	0.09022516	7.640769e-01
6	severity:time012	2	672	2.68142786	6.919789e-02
7	treat:time012	2	672	12.79599590	3.515413e-06
8	severity:treat:time012	2	672	0.41843893	6.582447e-01

mit SPSS:

Für eine Varianzanalyse mit Messwiederholungen muss der Datensatz in die entsprechende Form umstrukturiert werden (vgl. Anhang 1.2), wobei die Messwiederholungsvariablen *outcome.0*, *outcome.1*, *outcome.2* entstehen. Die Syntax für die Anova lautet dann:

```
GLM outcome.0 outcome.1 outcome.2 BY severity treat
  /WSfactor=Zeit 3 Polynomial
  /WSdesign=Zeit
  /Design=severity treat severity*treat.
```

Nachfolgend zunächst die Tabelle der Effekte mit dem Faktor *zeit*, danach die Tabelle für die Effekte der Gruppierungsfaktoren *severity* und *treat*:

Quelle		Quadrat summe	df	Mittel der Quadrate	F	Sig.
Zeit	Sphärizität angen.	23,844	2	11,922	60,687	,000
Zeit * severity	Sphärizität angen.	1,054	2	,527	2,681	,069
Zeit * treat	Sphärizität angen.	5,028	2	2,514	12,796	,000
Zeit * severity * treat	Sphärizität angen.	,164	2	,082	,418	,658
Fehler(Zeit)	Sphärizität angen.	132,017	672	,196		

Tests der Zwischensubjekteffekte					
Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Konstanter Term	296,013	1	296,013	1510,983	,000
severity	17,807	1	17,807	90,896	,000
treat	7,995	1	7,995	40,810	,000
severity * treat	,018	1	,018	,090	,764
Fehler	65,825	336	,196		

7.2 Anwendung der Verfahren für ordinale Merkmale

Zur 1-faktoriellen Varianzanalyse eines dichotomen Merkmals verwendet man üblicherweise den χ^2 -Test im Fall eines Gruppierungsfaktors bzw. Cochrans Q-Test im Fall eines Messwiederholungsfaktors. Diese Tests sind aber nichts anderes als der Kruskal-Wallis H-Test bzw. die Friedman-Varianzanalyse, wenn man bei diesen die ordinale Variable nur zwei Werte annehmen lässt und die Bindungskorrekturen verwendet. Somit lassen sich trivialerweise die in den Kapiteln 4.3.5, 5.3.3 und 6.4 beschriebenen verallgemeinerte Kruskal-Wallis-Friedman-Tests auf dichotome Merkmale anwenden.

Ferner weisen Akritas, Arnold und Brunner (vgl. [22] und [24]) ausdrücklich darauf hin, dass ihre ATS (Anova type statistic) nicht nur für ordinale, sondern auch dichotome Merkmale anwendbar sind.

Im Gegensatz zur u.a. Logistischen Regression können diese Methoden auch bei Messwiederholungen angewandt werden.

Auf Beispiele soll hier verzichtet werden, da die Anwendung dieser Verfahren in den vorangegangenen Kapiteln ausführlich beschrieben wurde.

8. Logistische Regression

8.1 dichotome abhängige Variablen

Die bekannteste logistische Regression ist die *binär-logistische Regression*, bei der ein Modell mit einer dichotomen (d.h. binären) abhängigen Variablen y (mit Werten 0 und 1) und v Prädiktoren x_1, x_2, \dots, x_v aufgestellt wird. Typischerweise ist dabei die abhängige Variable nicht y selbst, sondern $P(y=1)$, d.h. die Wahrscheinlichkeit, dass y den Wert 1 annimmt. Dadurch ist der Wertebereich der Funktion das komplette Intervall $[0,1]$:

$$P(y = 1) = \frac{e^{b_0 + b_1 x_1 + \dots + b_v x_v}}{1 + e^{b_0 + b_1 x_1 + \dots + b_v x_v}}$$

Für die unabhängigen Variablen (Prädiktoren) gelten die üblichen Bedingungen, d.h. für nominale Prädiktoren müssen Kontrastvariablen gebildet werden.

Zum weiteren Verständnis im Kontext der Varianzanalyse ist es an dieser Stelle nicht erforderlich, auf dieses Modell näher einzugehen. Die logistische Regression ist inzwischen soweit etabliert, dass sie in vielen einführenden Statistik-Lehrbüchern beschrieben wird. Eine Einführung bieten z.B. Diaz-Bone & Künemund [96] oder auch Wikipedia [97].

Allerdings ist an dieser Stelle noch nicht die Beziehung zur Varianzanalyse direkt erkennbar. Dazu sei angemerkt, dass die (parametrische) Varianzanalyse nichts anderes als eine lineare Regression mit nominalen Prädiktoren ist, nämlich den Faktoren, die wie oben angedeutet in Kontrastvariable transformiert werden. Und wenn genau diese Transformation bei der binären oder ordinalen logistischen Regression angewandt wird, erhält man ein Modell für eine dichotome oder ordinale Varianzanalyse. Hierbei gibt es jedoch einen Stolperstein: Für die Transformation der nominalen Faktoren in Kontraste gibt es zahlreiche Lösungen (vgl. Kapitel 9.1.2), die allerdings hinsichtlich der Tests der einzelnen Kontraste nicht immer zu demselben Ergebnis führen. Hinzu kommt, dass zunächst einmal, wie bei der Regression üblich, der Effekt jeder einzelnen Kontrastvariablen separat getestet wird. Einige Programme, insbesondere der binär-logistischen Regression, fassen allerdings die Tests für die Kontrastvariablen eines Faktors zu einem Gesamtergebnis zusammen, woraus der Effekt dieses Faktors zu entnehmen ist. Dieser globale Effekttest ist allerdings von dem gewählten Kontrasttyp unabhängig.

Ein Nachteil gegenüber den o.a. varianzanalytischen Verfahren liegt in der bislang noch fehlenden Möglichkeit zur Behandlung von Messwiederholungen. Auf der anderen Seite gibt es die Möglichkeit zur Verarbeitung von Versuchsplänen mit leeren Zellen. Wie auch insgesamt die Logistische Regression relativ liberal hinsichtlich der Voraussetzungen ist. Schaut man in die Literatur, so sucht man vergebens nach „handfesten“ Voraussetzungen, obwohl die zur Lösung eingesetzte Maximum-Likelihood-Methode sehr sensibel ist. (So kann es durchaus vorkommen, dass keine Lösung gefunden werden kann, weil die mathematische Schätzmethode nicht konvergiert. Das liegt an der mathematischen „Kondition“. Denn im Gegensatz zur Varianzanalyse wird die Lösung der Logistischen Regression nicht „direkt“ errechnet, sondern über ein Iterationsverfahren näherungsweise gefunden. Oder aber auch nicht.) Um Probleme zu vermeiden, sind nur zwei Dinge zu beachten:

- ein hinreichend großer Stichprobenumfang n , mindestens 10 pro Prädiktor bzw. geschätztem Parameter (wobei die Empfehlungen, sofern erwähnt, zum Teil stark divergieren). Da bei der Varianzanalyse ein Faktor als nominal skalierte Variable mit K Merkmalsausprägungen in $(K-1)$ Kontrastvariable transformiert und für die Interaktionen auch deren Produkte als Prädiktoren verwendet werden, bedeutet das für das n : ca. $10 \cdot (\text{Anzahl der Zellen})$.

- ein „vernünftiges“ Modell, d.h. u.a. ohne überflüssige (nicht erklärende) und ohne kollineare Variablen. Diese Forderung erübrigt sich allerdings beim Einsatz als Varianzanalyse.

Mit der logistischen Regression sind i.a. drei Signifikanztests verbunden:

- Ein Test des gesamten Modells, d.h. aller Effekte zusammen, über einen χ^2 -Test des log likelihood-Wertes. Sind Effekte der Faktoren vorhanden, so sollte dieser Test signifikant sein.
- Ein „klassischer“ χ^2 -Anpassungstest des Modells, der also prüft, in wieweit die Daten mit dem Modell vereinbar sind. Dieser sollte nicht signifikant sein.
- Die Signifikanzüberprüfung eines Regressionskoeffizienten (auf Verschiedenheit von 0) oder eines Effekts über die Wald-Statistik mittels des χ^2 -Tests.

Bei der binär-logistischen Regression wird ähnlich der Anova-Tabelle für jeden Effekt der Wald-Test automatisch ausgegeben, so dass die beiden Modell-Tests keine zusätzlichen Informationen geben. Bei der ordinalen Regression müssen die Wald-Tests recht aufwändig angefordert werden. Da kann es nützlich sein, über die Modell-Tests vorab zu erfahren, ob dieser Aufwand überhaupt erforderlich ist.

Als Beispiel wird hier wie in Kapitel 7.1.1 der Datensatz 7 mit `dvocabul`, der dichotomisierten Variable `vocubula` (Wortschatz), als abhängige Variable verwendet. Mit Hilfe der Logistischen Regression können allerdings alle drei Einflussfaktoren simultan untersucht werden, was die Interpretation der Effekte nicht gerade vereinfacht. Allerdings werden die Interaktionen `sex*type` und `income*type` auch hier weggelassen, die die beteiligten Faktoren nicht unabhängig voneinander sind. Für die oben angesprochene Transformation der Faktoren in Kontrastvariablen wird hier, wie in der Varianzanalyse üblich, die Effekt-Kodierung („Deviation“) vorgenommen. Mit dem Test eines Kontrasts wird dann die Abweichung der entsprechenden Ausprägung vom Mittelwert getestet. Alternativ könnten auch die einfache Kodierung gewählt werden, bei der Unterschiede einer Ausprägung zur letzten Ausprägung getestet werden. Die Anzahl von Zellen beträgt 36, so dass ein n von ca. 360 wünschenswert ist, was mit 1107 mehr als erfüllt ist.

mit R:

Zur Logistischen Regression bietet R u.a. die Funktion `glm` an. Hierbei ist die Angabe der Verteilungsfamilie `binomial` als Fehlerverteilung erforderlich, um das logistische Regressionsmodell zugrunde zu legen. Die oben angesprochene Effekt-Kodierung der Faktoren wird hier über den Parameter `contr.sum` der `options`-Anweisung vorgenommen. Die `anova`-Funktion erlaubt hier die Ausgabe einer Anova-Tabelle:

```
options(contrasts=c("contr.sum", "contr.poly"))
irish.glm <- glm(dvocabul~sex+income+type+sex:income,
               family=binomial, irish)
anova(irish.glm, test="Chisq")
```

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			1106	1533.8	
sex	1	3.985	1105	1529.8	0.04592 *
income	2	51.091	1103	1478.7	8.050e-12 ***
type	2	42.807	1101	1435.9	5.065e-10 ***
sex:income	2	1.876	1099	1434.0	0.39140

Fordert man über `summary(..)` eine Zusammenfassung der Ergebnisse, erhält man eine Tabelle der Einzelvergleiche, bei denen jeweils eine Stufe eines Faktors gegen den Mittelwert verglichen wird:

(Intercept)	-0.24638	0.09190	-2.681	0.00734	**
sex1	0.14066	0.07250	1.940	0.05237	.
income1	0.50340	0.09633	5.226	1.74e-07	***
income2	-0.09991	0.08818	-1.133	0.25723	
type1	0.27648	0.10134	2.728	0.00636	**
type2	0.65142	0.11727	5.555	2.78e-08	***
sex1:income1	-0.06826	0.09427	-0.724	0.46902	
sex1:income2	0.11108	0.08813	1.260	0.20752	

mit SPSS:

Die Logistische Regression ist in SPSS über das Menü „Regression -> binär logistisch...“ erreichbar. Nominale Prädiktoren, also Faktoren, müssen in der Menü-Oberfläche als „kategorial“ vereinbart werden. Hierbei bietet SPSS die Möglichkeit, für jeden Faktor die Kontraste individuell zu wählen. Meistens empfiehlt es sich, den Typ „Deviation“ zu wählen, da dann die Tests, die für alle Stufen (bis auf die letzte) ausgegeben werden, die Abweichungen der jeweiligen Kategorie vom Mittelwert überprüfen. Interaktionen müssen explizit angefordert werden. Die Syntax für die Analyse lautet:

```
Logistic regression variables=dvocabul
  /method = enter Sex Income Type Sex*Income
  /contrast(Sex) =Deviation
  /contrast(Income) =Deviation
  /contrast(Type) =Deviation
  /contrast(dvocabul) =Indicator.
```

mit der Tabelle für alle Effekte, in der sowohl die globalen Tests als auch die der einzelnen Kontraste (Variablenamen mit (..)), den Einzelvergleichen der (*K-I*) ersten Stufen eines Faktors gegen den Mittelwert (aller Stufen), enthalten sind:

	Regressions- koeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)
sex(1)	,141	,073	3,764	1	,052	1,151
income			27,460	2	,000	
income(1)	,503	,096	27,308	1	,000	1,654
income(2)	-,100	,088	1,284	1	,257	,905
type			38,325	2	,000	
type(1)	,276	,101	7,444	1	,006	1,318
type(2)	,651	,117	30,854	1	,000	1,918
income * sex			1,875	2	,392	
income(1) by sex(1)	-,068	,094	,524	1	,469	,934
income(2) by sex(1)	,111	,088	1,589	1	,208	1,117
Konstante	-,246	,092	7,187	1	,007	,782

8.2 ordinale abhängige Variablen

Das Modell der binär-logistischen Regression lässt sich in ein Modell für eine ordinale abhängige Variable y verallgemeinern, indem nicht mehr $P(y=1)$, sondern $P(y \leq j)$ als die abhängige Variable verwendet wird, mit $j=1, \dots, m$, wenn m die Anzahl der Merkmalsausprägungen von y ist:

$$P(y \leq j) = \frac{e^{b_{0j} + b_{1j}x_1 + \dots + b_{vj}x_v}}{1 + e^{b_{0j} + b_{1j}x_1 + \dots + b_{vj}x_v}}$$

(v ist wieder die Anzahl der Prädiktoren.) Während bei der binär-logistischen Regression nur eine Modellgleichung aufgestellt wird, sind es bei der ordinalen $m-1$ Modellgleichungen. D.h. es müssten $(m-1)*v$ Parameter geschätzt werden. Dieses Modell wird üblicherweise vereinfacht, indem für jeden Prädiktor i ($i=1, \dots, v$) die Koeffizienten der jeweiligen Merkmalsausprägungen als gleich angenommen werden: $b_{i1}=b_{i2}=\dots=b_{i(m-1)}$. Dies Modell heißt dann *proportional odds model*.

Zu den Voraussetzungen der dichotomen logistischen Regression kommt im Falle ordinaler Kriteriumsvariablen allerdings erschwerend die Anzahl der Ausprägungen von y hinzu, weil sich dadurch die Anzahl der Zellen vervielfacht. Daher ist dieses Verfahren i.a. nur für abhängige Variablen y mit 3 bis 5 Ausprägungen empfehlenswert.

Wie kann man sich die Bedingung gleicher Regressionskoeffizienten vorstellen? Dazu ein Beispiel: Eine Aufgabe wird mit Schulnoten 1 bis 6 beurteilt, und es soll der Einfluss von Geschlecht und Alter untersucht werden. Hinsichtlich des Geschlechts besagt die Bedingung: Wenn sich Mädchen und Jungen bei guten Noten (1 und 2) unterscheiden, dann unterscheiden sie sich auch bei guten bis mittleren Noten (1 bis 3) sowie bei guten bis schwachen (1 bis 4). Oder umgekehrt: wenn sie sich in einer Gruppe nicht unterscheiden, dann auch in keiner anderen. Die Gleichheit der Koeffizienten geht sogar noch soweit, dass die Mädchen-Jungen-Unterschiede in allen Notengruppen gleich groß sind. Ähnlich verhält es sich mit dem Alter. Wenn mit zunehmendem Alter die Wahrscheinlichkeit für eine gute Note steigt, dann gilt das ebenso für die Wahrscheinlichkeit einer guten bis mittleren Note oder einer nicht schlechten Note (1 bis 4).

Für die Anwendung des *proportion odds model* muss allerdings die Gleichheit der Koeffizienten mit den Daten vereinbar sein. Das wird mit dem „Parallelitätstest für Linien“ (*parallel lines test*) überprüft. Bei diesem werden die Abweichungen (ähnlich den Residuen) beider Modelle (einmal mit gleichen und einmal mit individuellen Koeffizienten) verglichen. Fällt dieser signifikant aus, bedeutet dies zunächst, dass die individuellen Koeffizienten eine signifikante Verbesserung der Anpassung erbringen. Das heißt aber, dass das vereinfachte Regressionsmodell nicht angewandt werden kann. Um diesen Test durchzuführen, müssen allerdings alle $(m-1)*v$ Parameter geschätzt werden, was ein hinreichend großes n erfordert. R bietet allerdings mit der Funktion `vglm` im Paket `VGAM` auch eine Lösung des o.a. Modells, bei dem die Gleichheit der Koeffizienten nicht gefordert wird.

Wenn für den Test ohnehin schon das Modell mit den individuellen Koeffizienten geschätzt werden muss, dann könnte man ja einfach damit anstatt mit dem vereinfachten Modell arbeiten. Nur: man hat dann eine riesige Anzahl von Koeffizienten, die einzeln kaum interpretierbar sind. Für einen Faktor mit K Gruppen (Stufen) resultieren alleine $(K-1)(m-1)$ Koeffizienten. Daher ist man bestrebt, das Modell mit gleichen Koeffizienten zu wählen.

Aber damit sind noch nicht alle Probleme aus dem Weg geräumt. Sollte man „zufällig“ ein Mo-

dell zum einen mit R und zum anderen mit SPSS rechnen, so wird man direkt irritiert sein, dass die Ergebnisse überhaupt nicht in Einklang zu bringen sind. Die Ursache: Das Modell ist ja zunächst einmal ein Regressionsmodell. Bei diesem werden in beiden Fällen automatisch Faktoren, d.h. nominale Prädiktoren, in Kontraste transformiert (vgl. Kapitel 9.1). Doch die Wahl des Kontrastes fällt bei beiden Programmen verschieden aus: R nimmt standardmäßig „einfache“ Kontraste mit der ersten Gruppe als Referenzgruppe, SPSS zwar auch „einfache“ Kontraste, aber mit der letzten Gruppe als Referenzgruppe. Dadurch fallen die Tests der Kontraste verschieden aus.

Erschwerend kommt hinzu, dass beide Programme apriori neben den Einzeltests der Kontraste keinen globalen, zusammenfassenden Test ausgeben, aus dem der Effekt eines Faktors abzulesen wäre. Lediglich bei SPSS kann solch ein Test angefordert werden. Bei R empfiehlt es sich, für den Faktor orthogonale Kontraste festzulegen, so dass sich die Einzeltests zu einem globalen Test aufsummieren.

Als Beispiel wird hier der Datensatz 7 (*irish*) benutzt, und zwar soll der Einfluss von Geschlecht (*sex*) und Schultyp (*type*) auf den Wortschatz (*vocabula*) untersucht werden.

mit R:

In R stehen eine Reihe von Funktionen zur ordinalen logistischen Regression zur Verfügung, u.a.:

- `polr (Modell, data=Dataframe)` aus dem Paket `MASS`
- `clm (Modell, data=Dataframe)` aus dem Paket `ordinal`
- `vglm (Modell, family=cumulative(parallel=T/F))` aus dem Paket `VGAM`, die sowohl das vereinfachte Modell (`parallel=T`) als auch das Modell mit individuellen Regressionskoeffizienten (`parallel=F`) handhaben kann.
- `npmlt (Modell, link="clogit")` aus dem Paket `mixcat`

Wie oben angedeutet bietet R keinen globalen Test für den Effekt eines Faktors. Dieser kann nur näherungsweise aus den Tests für die einzelnen Kontraste des Faktors ermittelt werden. Eine Voraussetzung dafür: die Kontraste müssen orthogonal sein. Dies sind z.B. die `contr.sum` (vgl. Kapitel 9.2). Die R-Funktionen geben für jeden Kontrast immer eine Testgröße aus, nämlich den Quotienten aus Parameterschätzung und Schätzfehler. Dieser ist normalerweise ein z-Wert, der für größere n immer normalverteilt ist, gelegentlich auch einen t-Wert, der allerdings wie ein z-Wert behandelt werden kann. Die Vorgehensweise ist dann:

- Durch Quadrieren jedes z-Wertes erhält man jeweils einen χ^2 -Wert, was der Prüfstatistik des *Wald-Tests* entspricht,
- Aufsummieren der zu einem Effekt gehörenden χ^2 -Werte, was wiederum einen χ^2 -Wert ergibt.
- Testen der Summe auf Signifikanz anhand der χ^2 -Verteilung, wobei die Anzahl der Freiheitsgrade der Anzahl Summanden entspricht.

Nachfolgend die Anweisungen für die ordinale Regression, hier mit `clm`, wobei zu beachten ist, dass nicht nur die Faktoren (hier `sex` und `type`) vom Typ „factor“ sein müssen, sondern auch die abhängige Variable vom Typ „ordered factor“. Die `options`-Anweisung

bewirkt, dass bei der Transformation der Faktoren das Effekt-Kodieren (`contr.sum`) angewandt wird.

```
irish <- within(irish, vocabula<-ordered(vocabula);
               sex<-factor(sex); type<-factor(type))
options(contrasts=c("contr.sum", "contr.poly"))
summary(clm(vocabula~sex*type, data=irish))
```

mit folgender Ausgabe für die Koeffizienten:

Coefficients:				
	Estimate	Std. Error	z value	Pr(> z)
sex1	-0.20542	0.14399	-1.427	0.1537
type1	-0.07987	0.12311	-0.649	0.5165
type2	0.93545	0.15661	5.973	2.33e-09 ***
sex1:type1	0.73844	0.16396	4.504	6.67e-06 ***
sex1:type2	-0.48368	0.20039	-2.414	0.0158 *

Zu den Tests der Effekte:

- Der Effekt von `sex` ist direkt aus der Tabelle mit $p_{\text{sex}} = 0.1537$ ablesbar, da der Faktor nur zwei Stufen hat.

- Der Effekt von `type`:

$$\chi^2_{\text{type}} = 0.649^2 + 5.973^2 = 36.1$$

und die Signifikanzprüfung ergibt $p_{\text{type}} < 0.001$.

- Der Effekt von `sex*type`:

$$\chi^2_{\text{sex*type}} = 4.504^2 + 2.414^2 = 26.1$$

und die Signifikanzprüfung ergibt $p_{\text{sex*type}} < 0.001$.

Bleibt noch zu prüfen, ob das vereinfachte *proportional odds model* überhaupt angewandt werden darf. Dazu wird mit der Funktion `vglm` einmal das einfache Modell (`fit1`) gerechnet und einmal das Modell mit den individuellen Koeffizienten (`fit2`). Der Unterschied der Anpassungsgüte wird mittels der Maßzahl „deviance“ auf Signifikanz überprüft:

```
fit1 <- vglm(vocabula~sex*type, data=irish, family=cumulative(parallel=T))
fit2 <- vglm(vocabula~sex*type, data=irish, family=cumulative(parallel=F))
pchisq(deviance(fit1)-deviance(fit2),
       df=df.residual(fit1)-df.residual(fit2), lower.tail=F)
```

Der p-Wert von 0.196 indiziert die Verträglichkeit des vereinfachten Modells mit den Daten.

Möchte man oben die Quadrierung der z-Werte direkt aus dem Ergebnisobjekt der Funktion `clm` (oder einer der anderen Funktionen) vornehmen, so steht man vor einem kleinen Problem. Die Ausgabe von `summary` erlaubt leider nicht die Adressierung der einzelnen Spalten. Und die Ergebnisobjekte selbst sind äußerst komplex aufgebaut. Einzig `npmlt` bietet die Koeffizienten und Standardfehler als Ergebnisobjekte an:

```
erg <- npmlt(irish$vocabula~irish$sex*irish$type)
zvalues <- erg$coefficients/erg$SE.coefficients
chisq <- zvalues^2
```

mit SPSS:

In SPSS steht für die ordinale logistische Regression der Modul `PLUM` (*polytomous universal model*) zur Verfügung, im Menü über Regression -> Ordinal. Faktoren, d.h. nominal skalierte Prädiktoren mit K Ausprägungen, werden automatisch in $K-1$ Kontraste transformiert (vgl. Kapitel 9.1), derart dass für diese $b_i=0$ getestet wird. Die oben erwähnten globalen Tests der Effekte sind nur über die Syntax anforderbar. Für den Faktor `sex` erübrigt sich solch ein Test, da für eine 2-stufige Variable dieser mit dem Test des Koeffizienten identisch ist.

```
PLUM vocabula BY sex type
  /link = logit
  /location = sex type sex*type
  /print = fit parameter summary tparallel
  /test (0,0) = type 1 0 0;
                    type 0 1 0
  /test (0,0) = sex*type 1 0 0 0 0 0 ;
                    sex*type 0 1 0 0 0 0 .
```

Erläuterungen hierzu:

Über `location` werden die zu testenden Effekte angegeben.

Über `test` wird jeweils ein globaler Effektttest angefordert, wobei auf der rechten Seite so viele Kontraste aufgeführt werden müssen, wie Parameter geschätzt werden, also (K_A-1) (mit K_A als Anzahl Stufen/Gruppen von Faktor A). Für jeden Kontrast wird hinter `test` ein Hypothesenwert in `(..)` angegeben, also i.a. 0. Bei Interaktionen beträgt die Anzahl der Kontraste $(K_A-1)(K_B-1)$ mit jeweils $K_A K_B$ Kontrastkoeffizienten. Als Koeffizienten werden zweckmäßigerweise nur 0 und 1 gewählt, wodurch die Hypothese lautet: alle Koeffizienten sind gleich 0.

Das wesentliche Ergebnis steckt in der Tabelle der Regressionsparameter, oben unter „Schwelle“ die Parameter b_{0j} ; sowie unter „Lage“ die Parameter b_i , die nach Annahme nicht von der Merkmalsausprägung j abhängen. Durch die nominalen Prädiktoren und deren Transformation in $(K-1)$ Kontraste und damit $(K-1)$ Parameter sind davon einige redundant, die dann mit 0 ausgegeben werden:

Parameterschätzer								
		Schätzer	Standard fehler	Wald	Fg	Sig.	Konfidenz intervall 95%	
							Unterg.	Oberg.
Schwelle	[vocabula = 1]	-,293	,275	1,139	1	,286	-,831	,245
	[vocabula = 2]	,891	,276	10,414	1	,001	,350	1,432
	[vocabula = 3]	1,935	,280	47,694	1	,000	1,386	2,484

Lage	[sex=1]	-,460	,341	1,823	1	,177	-1,128	,208
	[sex=2]	0 ^a	.	.	0	.	.	.
	[type=1]	,776	,288	7,246	1	,007	,211	1,341
	[type=2]	1,791	,333	28,937	1	,000	1,138	2,444
	[type=3]	0 ^a	.	.	0	.	.	.
	[sex=1] * [type=1]	,993	,367	7,337	1	,007	,275	1,712
	[sex=1] * [type=2]	-,229	,415	,304	1	,581	-1,043	,585
	[sex=1] * [type=3]	0 ^a	.	.	0	.	.	.
	[sex=2] * [type=1]	0 ^a	.	.	0	.	.	.
	[sex=2] * [type=2]	0 ^a	.	.	0	.	.	.
	[sex=2] * [type=3]	0 ^a	.	.	0	.	.	.

Bei den „globalen“ Effekttests werden zunächst die Kontraste noch einmal einzeln getestet, deren Ergebnis mit den o.a. identisch ist. Anschließend folgen die gewünschten Gesamttests. Auf welchen Faktor sich diese beziehen, ist nur über die davor angezeigten Kontrastkoeffizienten erkennbar. Also nachfolgend zunächst der Test für `type`, danach für `sex*type`:

Testergebnisse		
Wald	Freiheitsgrade	Sig.
35,100	2	,000

Testergebnisse		
Wald	Freiheitsgrade	Sig.
23,614	2	,000

Von besonderem Interesse ist noch der Parallelitätstest. Da dieser nicht signifikant ist, darf das vereinfachte *proportional odds model* angewandt werden.

Parallelitätstest für Linien ^a				
Modell	-2 Log-Likelihood	Chi-Quadrat	Freiheitsgrade	Sig.
Nullhypothese	99,933			
Allgemein	86,421	13,511	10	,196

Die Nullhypothese gibt an, daß die Lageparameter (Steigungskoeffizienten) über die Antwortkategorien übereinstimmen.

Was passiert, wenn das n bezogen auf die Anzahl der Zellen nicht ausreichend ist? Wollte man z.B. eine ordinale Regression mit den Daten des Beispiels 2 (`mydata2`) rechnen, dann stößt man auf dieses Problem: Die Kriteriumsvariable hat 8 Ausprägungen und das Design hat 8 Zellen, also gibt es insgesamt 64 Zellen. Aber auf der anderen Seite nur 33 Beobachtungen. Man könnte zunächst das Problem abmildern, indem Merkmalsausprägungen der abhängigen Variablen zusammengefasst werden, z.B. von 8 auf 4 reduzieren. Das kann gelegentlich gut gehen, in diesem Fall aber nicht. Es kann nämlich keine „gesicherte“ Lösung gefunden werden. Sowohl R als auch SPSS geben in solchen Fällen Warnungen aus, etwa in R:

```
Warning message:  
(1) Hessian is numerically singular: parameters are not uniquely  
determined  
In addition: Absolute convergence criterion was met, but relative  
criterion was not met
```

oder in SPSS:

Warnungen
Es gibt 15 (46,9%) Zellen (also Niveaus der abhängigen Variablen über Kombinationen von Werten der Einflußvariablen) mit Null-Häufigkeiten.
Es wurden unerwartete Singularitäten in der Fisher-Informationsmatrix gefunden. Möglicherweise liegt eine quasi-vollständige Trennung der Daten vor. Einige Parameter werden sich Unendlich nähern.
Die PLUM-Prozedur wird trotz der obigen Warnung(en) fortgesetzt. Die anschließend angezeigten Ergebnisse basieren auf der letzten Iteration. Die Zulässigkeit der Anpassungsgüte des Modells ist unsicher.

Zwar kann sowohl in R als auch in SPSS die Anzahl der Iterationen zur Berechnung der Lösung vergrößert werden, was aber selten hilft. In solchen Fällen kann nur davon abgeraten werden, die Ergebnisse zu verwenden.

9. Mittelwertvergleiche, Kontraste und Kodierungen

In der Regel ist es erforderlich, im Anschluss an eine Varianzanalyse Mittelwertvergleiche durchzuführen. Denn signifikante Effekte besagen nur, dass zwischen irgendwelchen Gruppen Mittelwertunterschiede bestehen, geben aber keinen weiteren Aufschluss darüber, welche Gruppen oder Stufen dies nun sind. Für diese Fragestellung unterscheidet man grundsätzlich:

- *geplante* Vergleiche, *apriori-Vergleiche* oder *Kontraste*, die als Hypothesen bereits *vor* der Untersuchung, d.h. vor Erhebung des Datenmaterials, vorliegen, und
- *multiple Mittelwertvergleiche* oder *posthoc-Tests*, für die keine speziellen Hypothesen vorliegen und die üblicherweise durchgeführt werden, wenn die Varianzanalyse einen signifikanten Effekt aufzeigt, der dann näher analysiert werden soll. Das allgemeinste, aber auch schwächste Verfahren in dieser Kategorie sind die *paarweisen Vergleiche mit α -Adjustierungen*.

Alpha-Adjustierungen und multiplen Vergleichen ist ein separates Skript gewidmet (vgl. [9]). Dieses Skript beschränkt sich auf allgemeine Grundlagen zu Kontrasten, da diese zum Verständnis in den Kapiteln 7 und 8 erforderlich sind. Ausführliche Darstellungen sind auch im Internet zu finden, so z.B. bei Gonzalez [103].

9.1 Grundlagen

Vielfach existieren bei der Varianzanalyse eines Merkmals zusätzlich zur globalen Hypothese gleicher Mittelwerte noch spezielle Hypothesen. Liegen z.B. 3 Gruppen vor, etwa eine Kontrollgruppe K sowie 2 Experimentalgruppen A und B, so könnten diese lauten: Vergleich der Mittelwerte von K gegen A sowie K gegen B. Solche Hypothesen müssen allerdings bereits *vor* der Untersuchung festliegen. Solche speziellen Vergleiche heißen *apriori-Vergleiche* oder *Kontraste*. Hierbei können nicht nur jeweils die Mittelwerte von zwei Gruppen verglichen werden, sondern allgemein eine Linearkombination der Mittelwerte auf den Wert 0. Bei o.a. Beispiel etwa den Mittelwert von K gegen den Durchschnitt der Mittelwerte von A und B, d.h. die beiden Experimentalgruppen unterscheiden sich „im Schnitt“ von der Kontrollgruppe hinsichtlich der Mittelwerte. Die Linearkombination ist dann $1 \cdot \mu_K - 0.5 \cdot (\mu_A + \mu_B)$. Theoretisch können sogar bei der Zusammenfassung von Gruppen gewichtete Mittel gebildet werden, etwa $(0.333 \cdot \mu_A + 0.667 \cdot \mu_B)$, wenn etwa die B-Gruppe doppelt so stark berücksichtigt werden soll wie die A-Gruppe.

Hat ein Faktor k Gruppen (Schichten), so ist ein Kontrast C über k Koeffizienten c_j definiert:

$$C = c_1 \mu_1 + c_2 \mu_2 + \dots + c_k \mu_k$$

wobei die Nebenbedingung $c_1 + c_2 + \dots + c_k = 0$ eingehalten werden muss. Diese Summe wird dann auf den Wert 0 getestet. Im parametrischen Fall errechnet sich die Testgröße dann als

$$SS_C = \frac{(c_1 \bar{x}_1 + c_2 \bar{x}_2 + \dots + c_k \bar{x}_k)^2}{\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2} + \dots + \frac{c_k^2}{n_k}}$$

und entspricht dem Anteil der Streuung SS_{Effekt} , der durch diesen Kontrast erklärt wird. Somit lässt sich diese Streuung SS_C analog mit dem F-Test auf Signifikanz überprüfen :

$$F = \frac{SS_C}{MS_{\text{Fehler}}}$$

wobei dieser F-Wert 1 Zähler-Fg hat und Nenner-Fg dem Test von SS_{Effekt} zu entnehmen sind.

Es gibt aber noch eine andere, in R bevorzugte, Darstellung dieses Tests, und zwar mittels eines t-Tests, wobei in Erinnerung gerufen wird, dass allgemein $t_n = \sqrt{F_{1,n}}$ gilt:

$$t = \frac{C}{s_e} = \sqrt{F}$$

wobei C der o.a. Kontrastschätzer und s_e der Standardfehler (des Kontrastschätzers) ist.

Es sei noch erwähnt, dass die Skalierung der c_j ohne Bedeutung ist, d.h. Kontraste $c_j' = a \cdot c_j$ ergeben dasselbe Resultat wie die Kontraste c_j .

In der Regel hat der Untersucher mehrere Hypothesen, aus denen dann mehrere Kontraste resultieren. Hierfür gelten dann folgende Regeln bzw. Eigenschaften:

- Es dürfen nur $(k-1)$ Kontraste getestet werden.
- Zwei Kontraste C_1 mit Koeffizienten $c_{11}+c_{12}+\dots+c_{1k}$ und C_2 mit Koeffizienten $c_{21}+c_{22}+\dots+c_{2k}$ heißen *orthogonal*, d.h. sind unabhängig voneinander, wenn die folgende Bedingung erfüllt ist:

$$\frac{c_{11}c_{21}}{n_1} + \frac{c_{12}c_{22}}{n_2} + \dots + \frac{c_{1k}c_{2k}}{n_k} = 0$$

- Eine Menge von Kontrasten heißt *orthogonal*, wenn alle Paare orthogonal sind.
- Werden $(k-1)$ orthogonale Kontraste C_1, C_2, \dots, C_k mit Streuungen $SS_{C_1}, SS_{C_2}, \dots, SS_{C_{(k-1)}}$ getestet, dann gilt $SS_{C_1} + SS_{C_2} + \dots + SS_{C_{(k-1)}} = SS_{\text{Effekt}}$, d.h. die gesamte durch den Faktor erklärte Streuung lässt sich in $(k-1)$ einzeln erklärbare Streuungen unterteilen.

Sind die zu untersuchenden Kontraste nicht orthogonal oder sollen mehr als $(k-1)$ Kontraste geprüft werden, so sind die einzelnen Testergebnisse nicht mehr unabhängig voneinander. In solchen Fällen ist eine α -Korrektur (siehe dazu [9]) vorzunehmen. Speziell hierfür ist u.a. das Verfahren von *Dunn & Bonferroni* konzipiert.

Beispiel:

Für die o.a. Situation eines Faktors mit den Gruppen K, A und B werden 2 Kontraste definiert: K-A sowie K-B. Daraus resultieren folgende Koeffizienten c_j :

Gruppe	Kontraste	
	C_1	C_2
K	1	1
A	-1	0
B	0	-1

Diese beiden Kontraste sind nicht orthogonal, denn $1 \cdot 1 + (-1) \cdot 0 + 0 \cdot (-1) = 1$.

Wird dagegen zum einen die Kontrollgruppe K gegen das Mittel von A und B verglichen und zum anderen die beiden Experimentalgruppen A und B gegeneinander, dann resultieren

daraus die Koeffizienten c_j :

	Kontraste	
Gruppe	C ₁	C ₂
K	2	0
A	- 1	1
B	- 1	- 1

Diese beiden Kontraste sind orthogonal, denn $2 \cdot 0 + (-1) \cdot 1 + (-1) \cdot (-1) = 0$.

Die Kontraste oder Kodierungen haben auch eine andere Funktion: Bei der Regression müssen Prädiktoren mit nominalem Skalenniveau dichotomisiert werden. Die „naive“ Art, ein nominales Merkmal f mit k Ausprägungen in mehrere dichotome d_1, \dots, d_k zu transformieren, ist normalerweise so, dass d_j genau dann den Wert 1 hat, wenn f den Wert j hat, und sonst 0. Da von diesen k Variablen zwangsläufig eine redundant ist - jede beliebige von diesen lässt sich aus den übrigen errechnen, z.B. $d_k = 1 - d_1 - d_2 - \dots - d_{k-1}$, muss eine weggelassen werden. Diese Kodierung, das *dummy coding*, ist nicht die einzige Möglichkeit, ein nominales Merkmal zu transformieren. Nachfolgend werden die Standardmethoden für die Kodierung und Kontrastbildung vorgestellt.

9.2 Standard-Kontraste

Prinzipiell kann der Benutzer natürlich individuelle Kontraste festlegen, was sowohl in R als auch in SPSS mit ein wenig Aufwand verbunden ist. Es gibt aber eine Reihe von „Standard“-Kontrasten, die für einen Faktor vereinbart werden können. Allerdings ist die Namensgebung nicht einheitlich. Hierbei sind Kontraste und Kodierungen (nominaler Variablen) zu unterscheiden. Bei Kontrasten muss die Nebenbedingung $c_1 + c_2 + \dots + c_k = 0$ eingehalten werden, bei Kodierungen nicht.

Dummy Coding / Indikator / Einfach bzw. Simple (SPSS)/ `contr.treatment (R)`

Statistisch werden alle Gruppen gegen eine vorgegebene, üblicherweise die erste oder letzte, paarweise verglichen, nämlich die, die bei den oben erwähnten d_j nicht repräsentiert ist. Die „Referenzgruppe“ kann sowohl bei R als auch bei SPSS festgelegt werden. Dies wird angewandt, wenn eine Gruppe die Vergleichsgruppe ist, meist die sog. Kontrollgruppe. Anzumerken ist, dass bei SPSS die Koeffizienten dieselben sind, wie beim Effekt-Kodierung bei R, aber die Ergebnisse denen eines Vergleichs mit einer vorgegebenen Gruppe entsprechen:

Gruppe	Kontraste R				Kontraste SPSS			
	1	2	...	(k-1)	1	2	...	(k-1)
1	1	0		0	1	0		0
2	0	1		0	0	1		0
...	0	0						
k-1	0	0		1	0	0		1
k	0	0		0	- 1	- 1		- 1

Effekt-Kodierung / Abweichung bzw. Deviation (SPSS) / contr.sum (R)

Dies sind orthogonale Kontraste, die letztlich der Varianzanalyse zugrunde liegen. Durch diese werden nämlich die Abweichungen vom Gesamtmittelwert getestet. Da nur $(k-1)$ Vergleiche erlaubt sind, muss der Test für eine Gruppe entfallen. Dies ist üblicherweise (in R und SPSS) die letzte Gruppe. Die Koeffizienten:

Gruppe	Kontraste R				Kontraste SPSS			
	1	2	...	(k-1)	1	2	...	(k-1)
1	1	0		0	$(k-1)/k$	$-1/k$		$-1/k$
2	0	1		0	$-1/k$	$(k-1)/k$		$-1/k$
...	0	0						
k-1	0	0		1	$-1/k$	$-1/k$		$(k-1)/k$
k	-1	-1		-1	$-1/k$	$-1/k$		$-1/k$

Helmert-Kodierung / Differenz bzw. Difference (SPSS) / contr.helmert (R)

Bei dieser Bildung von orthogonalen Kontrasten werden sukzessive folgende Gruppen miteinander verglichen: 1-2, (1,2)-3, (1,2,3)-4 usw. wobei mit (.,.) der Mittelwert der entsprechenden Gruppen bezeichnet wird.

Gruppe	Kontraste R und SPSS			
	1	2	...	(k-1)
1	-1	$-1/2$		$-1/(k-1)$
2	1	$-1/2$		$-1/(k-1)$
...	0	1		
k-1	0	0		$-1/(k-1)$
k	0	0		1

umgekehrte Helmert-Kodierung / Helmert (SPSS)

Bei dieser Bildung von orthogonalen Kontrasten werden sukzessive die erste gegen alle folgenden Gruppen miteinander verglichen, die zweite gegen alle folgenden usw. (Diese Kontraste sind in R nicht verfügbar.)

Gruppe	Kontraste SPSS			
	1	2	...	(k-1)
1	1	0		0
2	$-1/(k-1)$	1		0
...	$-1/(k-1)$	$-1/(k-2)$		

	Kontraste SPSS			
Gruppe	1	2	...	(k-1)
k-1	- 1/(k-1)	- 1/(k-2)		1
k	- 1/(k-1)	- 1/(k-2)		- 1

Wiederholt bzw. Repeated (SPSS)

Bei dieser Kodierung werden sukzessive zwei aufeinander folgende Gruppen miteinander verglichen: 1-2, 2-3, 3-4 usw. Diese werden sinnvollerweise bei Messwiederholungsfaktoren eingesetzt. (Diese Kontraste sind in R nicht verfügbar.)

	Kontraste SPSS			
Gruppe	1	2	...	(k-1)
1	1	0		0
2	- 1	1		0
...	0	- 1		
k-1	0	0		1
k	0	0		- 1

Polynomial

Diese Kontraste dienen der Trendanalyse und setzen ordinales Skalenniveau des Faktors voraus. Die Kontrastkoeffizienten errechnen sich aus den sog. orthogonalen Polynomen. In dieser Version des Skripts wird nicht näher darauf eingegangen.

Ausführliche Erläuterungen der Standard-Kontraste sind unter [101] für R bzw. unter [102] für SPSS zu finden.

9.3 Auswahl der Kontraste

R bietet die o.a. Standard-Kontraste über die folgenden Funktionen:

```
contr.treatment(k, base=j) (j=Nummer der Vergleichsgruppe)
contr.sum(k)
contr.helmert(k)
contr.poly(k)
```

wobei k die Anzahl der Gruppen ist. Die Auswahl erfolgt über das Kommando

```
contrasts(Faktorname) <- contr.name
```

Es gibt auch eine Voreinstellung für Objekte vom Typ „factor“:

```
contr.treatment(k, base=k) für „normale“ Faktoren
contr.poly(k) für „ordered factors“
```

die dann z.B. bei der Verwendung von „factor“-Variablen bei der Regression verwendet werden. Die Voreinstellung kann über

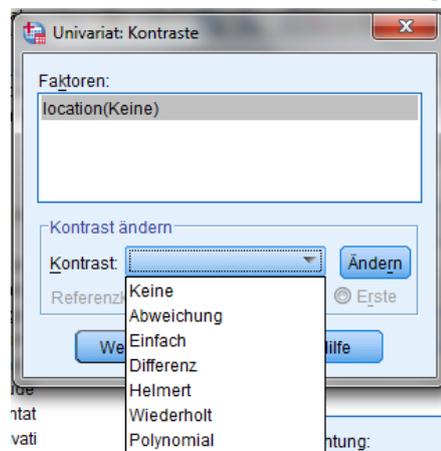
```
options(contrasts=c("contr.name1", "contr.name2"))
```

geändert werden und über `getOption("contrasts")` abgefragt werden. Hierbei wird `contr.name1` für „normale“ Faktoren und `contr.name2` für für „ordered factors“ übernommen. (Vgl. auch Anmerkungen zur Funktion `aov` in Kapitel 3.1.)

Bei SPSS gibt es in den Routinen zur Varianzanalyse sowie zur binär logistischen Regression zum einen das Unterkommando

```
/Contrast(Faktorname)=name
```

wobei `name` einer der oben für SPSS angeführten *englischen* Kontrastnamen ist, zum anderen in den Eingabemasken den Button „Kontraste“, der zu der folgenden Auswahl führt:



Dabei darf allerdings nicht der „Ändern“-Button vergessen werden.

9.4 nichtparametrische Kontraste für die RT-, ART- und KWF-Verfahren

Einige der im Kapitel 2 vorgestellten nichtparametrischen Varianzanalysen lassen sich ja auf die parametrischen Standardverfahren zurückführen, so insbesondere die RT-, die ART-, die INT- sowie die verallgemeinerte Kruskal-Wallis- und Friedman-Analysen (KWF). Die Analyse von Kontrasten ist darin problemlos möglich.

Als erstes sollen Kontrast-Vergleiche in Verbindung mit dem RT-Verfahren, und zwar am Beispiel des Datensatzes 2 (`mydata2`) mit dem Faktor `drug` demonstriert werden. Zunächst einmal wird angenommen, dass die erste Gruppe eine Vergleichsgruppe ist, gegen die die anderen drei Gruppen getestet werden sollen.

mit R:

Die Tabelle 4.6 in Kapitel 4.3.4 zeigt für den Faktor `drug` einen signifikanten Effekt an, der nun weiter untersucht werden soll. Dabei besteht die Hypothese, dass der Mittelwert der ersten Gruppe sich von allen anderen unterscheidet. Diese kann mit den „einfach“-Kontrasten (`contr.treatment`) geprüft werden. Dazu ist `gls` aus dem Paket `nlme` als Varianzanalysefunktion zu verwenden, die zwar keine Anova-Tabelle ausgibt, dafür aber die Kontraste:

```
library(nlme)
```

```
contrasts(mydata2$drug) <- contr.treatment(4, base=1)
aovc <- gls(rx~group*drugs, mydata2)
summary(aovc)
```

Neben ein paar weiter nicht interessierenden Ergebnissen wird eine Tabelle aller Kontraste mit Tests ausgegeben. Hierbei ist anzumerken, dass bedingt durch die 2-faktorielle Analyse auch Kontraste für den anderen Faktor (`group`) sowie für die Interaktion ausgegeben werden. Die Zeilen `drugs2,...,drugs4` enthalten die Vergleiche mit `drug1`:

	Value	Std. Error	t-value	p-value
(Intercept)	8.2500	2.514377	3.2811303	3.043817e-03
group1	5.2500	2.514377	2.0879920	4.714492e-02
drugs2	5.9750	3.346511	1.7854415	8.632831e-02
drugs3	9.3750	3.426519	2.7360130	1.127545e-02
drugs4	16.7125	3.346511	4.9940068	3.785352e-05
group1:drugs2	1.7250	3.346511	0.5154622	6.107586e-01
group1:drugs3	-1.3750	3.426519	-0.4012819	6.916220e-01
group1:drugs4	-7.9125	3.346511	-2.3644026	2.613481e-02

Tabelle 9-1

mit SPSS:

Die Tabelle 4.8 in Kapitel 4.3.4 zeigt für den Faktor `drug` einen signifikanten Effekt an, der nun weiter untersucht werden soll. Dabei besteht die Hypothese, dass der Mittelwert der ersten Gruppe sich von allen anderen unterscheidet. Diese kann mit den „simple“-Kontrasten geprüft werden. Dazu ist bei den Anweisungen für die oben erwähnte Analyse die Zeile

```
/Contrast(drugs)=Simple(1)
```

einzufragen, wobei das „(1)“ die Nummer der Vergleichsgruppe angibt, also hier die erste:

```
Unianova x by patients drug
/Contrast(drugs)=Simple(1)
/save = zresid
/print = homogeneity
/design = patients drug patients*drug.
```

Die Ausgabe dazu sollte selbsterklärend sein:

Kontrastergebnisse (K-Matrix)			
Einfacher Kontrast ^a	Abhängige Variable		
	Rx		
Niveau 2 vs. Niveau 1	Kontrastschätzer	5,975	
	Hypothesenwert	0	
	Differenz (Schätzung - Hypothesen)	5,975	
	Standardfehler	3,347	
	Sig.	,086	
	95% Konfidenzintervall für die Differenz	Untergrenze	-,917
	Obergrenze	12,867	

Niveau 3 vs. Niveau 1	Kontrastschätzer		9,375
	Hypothesenwert		0
	Differenz (Schätzung - Hypothesen)		9,375
	Standardfehler		3,427
	Sig.		,011
	95% Konfidenzintervall für die Differenz	Untergrenze	2,318
		Obergrenze	16,432
Niveau 4 vs. Niveau 1	Kontrastschätzer		16,713
	Hypothesenwert		0
	Differenz (Schätzung - Hypothesen)		16,713
	Standardfehler		3,347
	Sig.		,000
	95% Konfidenzintervall für die Differenz	Untergrenze	9,820
		Obergrenze	23,605
a. Referenzkategorie = 1			

Tabelle 9-2

Das Vorgehen ist im Zusammenhang mit dem ART-Verfahren (vgl. Kapitel 4.3.6) völlig identisch.

Ein wenig anders ist es bei Verwendung des KWF-Verfahrens (vgl. Kapitel 4.3.5). Hier müssen die χ^2 -Werte für jeden Vergleich „mit der Hand“ ausgerechnet werden, was ein wenig mühselig ist, zumal SPSS nicht die Testgröße ausgibt:

$$\chi^2 = t^2 \cdot \frac{MS_{Fehler}}{MS_{total}} \quad t = \frac{C}{s_e}$$

wobei

- t die t-verteilte Teststatistik ist, die bei SPSS erst errechnet werden muss aus
- C der Kontrastwert (in SPSS: Kontrastschätzer) und
- s_e der Standardfehler (des Kontrastschätzers),
- MS_{Fehler} die Fehlervarianz (aus der Anova-Tabelle zu entnehmen)
- MS_{total} die Gesamtvarianz, die bereits für die Anova-Tests ermittelt worden war (vgl. Kapitel 4.3.5).

Die χ^2 -Werte haben jeweils 1 Fg und müssen anhand der Tabellen der χ^2 -Verteilung auf Signifikanz überprüft werden. Aus Tabelle 4-8 in Kapitel 4.3.5 lässt sich $MS_{Fehler} = 43,35$ sowie $MS_{total} = 2904,5/32 = 90,77$ errechnen.

mit R:

In der Anova-Tabelle für diese Daten (Tabelle 4-6) fehlt ein Wert für MS_{Fehler} . Dieser muss gegebenenfalls mit `aov` neu errechnet werden und ergibt `msfehler` mit dem Wert 43,35. Zur Berechnung der χ^2 -Werte müssen die t-Werte aus der Tabelle 9-1 quadriert, mit MS_{Fehler} sowie durch MS_{total} dividiert werden. Das kann in R programmiert werden. (Die Berechnung „per Hand“ kann dem Abschnitt „SPSS“ entnommen werden.) Wenn `aovc` das oben ermittelte Ergebnisobjekt von `gls` ist, dann lässt sich mit folgenden Anweisungen daraus zunächst die Kontrasttabelle `ctabelle`, die t-Werte `twerte` und schließlich die χ^2 -Werte `chisq`:

```

ctabelle<- as.data.frame(summary(aovc)$tTable)
twerte  <- ctabelle$"t-value"
names(twerte)<- row.names(ctabelle)
aov2r   <- anova(aov(rx~group*drugs,mydata2))
mstotal <- sum(aov2r[,2])/sum(aov2r[,1])
msfehler<- aov2r[4,3]
chisq   <- twerte^2*msfehler/mstotal
pvalues <- 1-pchisq(chisq,1)
data.frame(chisq,pvalues)

```

mit der nachfolgenden Ausgabe, worin die Zeilen drugs2,...,drugs4 die gewünschten Testergebnisse enthalten:

	chisq	pvalues
(Intercept)	5.14197182	0.0233541081
group1	2.08228611	0.1490168492
drugs2	1.52255843	0.2172327363
drugs3	3.57535389	0.0586429521
drugs4	11.91189867	0.0005577652
group1:drugs2	0.12690430	0.7216636075
group1:drugs3	0.07690983	0.7815296246
group1:drugs4	2.67008813	0.1022503615

Tabelle 9-3

mit SPSS:

Die Berechnung soll nur für den ersten Vergleich (drug1 - drug2) gezeigt werden:

$$\chi^2 = \left(\frac{5,975}{3,347} \right)^2 \cdot \frac{43,35}{90,77} = 1,52$$

Der kritische χ^2 -Wert bei 1 Fg beträgt 3,84, so dass kein Unterschied zwischen drug1 und drug2 nachgewiesen werden kann.

Das vorige Beispiel wird dahingehend modifiziert, dass drug1 und drug2 als etablierte Präparate angenommen werden, während drug3 und drug4 als neu angesehen werden. Daher sollen zum einen die beiden alten Präparate (1-2) sowie die beiden neuen Präparate (3-4) verglichen werden, zum anderen die alten zusammen gegen die neuen zusammen ((1,2)-(3,4)). Daraus resultiert folgende Kontrastmatrix:

Gruppe	Kontraste		
	1	2	3
drug1	1	0	1
drug2	- 1	0	1
drug3	0	1	- 1
drug4	0	- 1	- 1

Tabelle 9-4

Nachfolgend werden nur die Anweisungen für die Benutzer-spezifischen Kontraste aufgeführt. Die Ausgabe ist praktisch identisch mit der der Standard-Kontraste im vorigen Beispiel.

mit R:

Auch hier dient natürlich wieder die Funktion `gls` aus dem Paket `nlme` zur Analyse der Kontraste. Lediglich die Spezifikation der Koeffizienten differiert. Die Werte müssen spaltenweise eingegeben, als Matrix mit 3 Spalten definiert und dann übergeben werden:

```
library(nlme)
cont <- matrix( c(1,-1,0,0, 0,0,1,-1, 1,1,-1,-1), ncol=3)
contrasts(mydata2$drugs) <- cont
aovc <- gls(rx~group*drugs,mydata2)
summary(aovc)
```

mit SPSS:

Auch hier ist nur eine kleine Modifikation der Anweisungen des letzten Beispiels erforderlich. Die Kontrast-Anweisung lautet:

```
/Contrast(drugs) = Special(1 -1 0 0 0 0 1 -1 1 1 -1 -1)
```

Die Ausführungen dieses Abschnitts gelten gleichermaßen für Analysen mit Messwiederholungen.

9.5 universelles Verfahren für Kontraste

Wenn die nichtparametrische Varianzanalyse nicht auf die parametrische zurückgeführt werden kann, steht damit auch nicht mehr die Kontrastfunktionalität der Standardroutinen von R und SPSS zur Verfügung. D.h. man verfügt nur über die Funktion zur Durchführung einer Varianzanalyse. Damit lassen sich aber immerhin durch passendes Umkodieren der Gruppen/Faktorvariablen sowohl zwei Gruppen vergleichen als auch Gruppen von Gruppen vergleichen. Das soll wieder am oben verwendeten Datensatz 2 (`mydata2`) erläutert werden.

Es sollen die Kontraste aus Tabelle 9-4 getestet werden. Vor jedem der drei Vergleiche muss die Gruppenvariable `drugs` so umkodiert werden, dass jeweils nicht verwendete Werte auf Missing gesetzt werden. Dies erfolgt mit einer Hilfsvariablen `d`.

mit R:

Die Kontraste sollen im Anschluss an eine Kruskal-Wallis-Varianzanalyse durchgeführt werden. Es wird darauf aufmerksam gemacht, dass die `levels`-Angaben aus der `factor`-Definition der Gruppierungsvariablen (hier `drugs`) auf `d` übertragen werden, aber anschließend nicht mehr stimmen, da die Anzahl der Stufen von `d` auf zwei reduziert wurde. Das kann bei verschiedenen Funktionen zu Problemen führen. Gegebenenfalls muss dies in einer `factor`-Anweisung korrigiert werden.

```
kruskal.test(mydata2$x,drugs) # gloabler Vergleich

d <- mydata2$drugs # Vergleich 1-2
d[d==3|d==4] <- NA
d<-factor(d,levels=c(1,2))
kruskal.test(mydata2$x,d)

d <- mydata2$drugs # Vergleich 3-4
d[d==1|d==2] <- NA
d<-factor(d,levels=c(3,4))
kruskal.test(mydata2$x,d)
```

```
d <- mydata2$drugs # Vergleich (1,2) - (3,4)
d[d==1|d==2] <- 1
d[d==3|d==4] <- 4
d<-factor(d, levels=c(1,4))
kruskal.test(mydata2$x, d)
```

Der globale χ^2 -Wert beträgt 11,2 . Die χ^2 -Werte der drei Kontraste: 1,97 (1-2), 2,61 (3-4) und 7,32 ((1,2)-(3,4)) mit der Summe von 11,9, die ungefähr dem globalen Wert entspricht, da die Kontraste orthogonal sind.

mit SPSS:

Die Kontraste sollen im Anschluss an eine Kruskal-Wallis-Varianzanalyse durchgeführt werden.

```
NPtests /independent test (x) group (drugs) Kruskal_Wallis.

* Vergleich 1-2 .
Recode drugs (1=1) (2=2) (3,4=sysmis) into d.
NPtests /independent test (x) group (d) Kruskal_Wallis.

* Vergleich 3-4 .
Recode drugs (3=3) (4=4) (1,2=sysmis) into d.
NPtests /independent test (x) group (d) Kruskal_Wallis.

* Vergleich (1,2)-(3,4) .
Recode drugs (1,2=1) (3,4=4) into d.
NPtests /independent test (x) group (d) Kruskal_Wallis.
```

Der globale χ^2 -Wert beträgt 11,2 . Die χ^2 -Werte der drei Kontraste: 1,97 (1-2), 2,61 (3-4) und 7,32 ((1,2)-(3,4)) mit der Summe von 11,9, die ungefähr dem globalen Wert entspricht, da die Kontraste orthogonal sind.

Aus diesem Beispiel geht das generelle Prozedere hervor. So lassen sich auch die im vorigen Abschnitt vorgenommenen Vergleiche der `drug2`, . . . , `drug4` gegen `drug1` durchführen.

9.6 Kontraste bei logistischen Regressionen

Bei der logistischen Regression gibt es für nominale Prädiktoren Standard-Kontraste. Wenn in R ein Prädiktor als „factor“ deklariert ist, wird für diesen automatisch die Kodierung gewählt, die in der `options(contrasts...)`-Anweisung festgelegt wurde (vgl. Kapitel 9.3). In SPSS kann bei der binär-logistischen Regression wie oben in 9.3 dargestellt die Kodierung gewählt werden. Speziellere Kontraste müssen wie oben in 9.5 skizziert über Umkodierungen analysiert werden. Beispiele sind in Kapitel 8 zu finden.

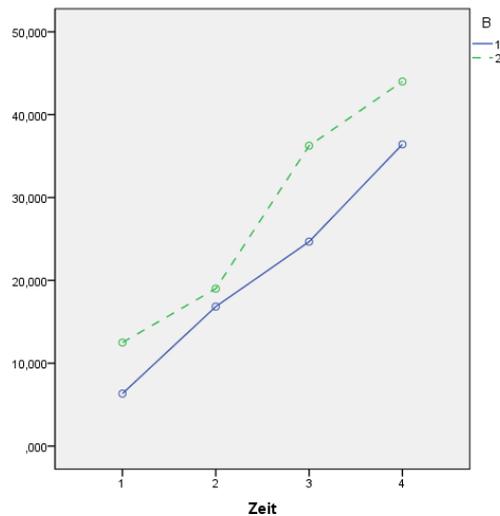
9.7 Kontraste für Messwiederholungen und Interaktionen

Aus dem eingangs (Kapitel 9.1) angeführten Signifikanztest für einen Kontrast kann abgelesen werden, dass dafür lediglich die Varianz MS_{Error} erforderlich ist, die praktisch den Nenner des entsprechenden F-Tests für den untersuchten Effekt darstellt. Somit sind zumindest im Fall der RT-, ART-, INT- und KWF-Analysen Kontrastanalysen gleichermaßen für Versuchspläne mit Messwiederholungen durchführbar.

Sind für zwei Faktoren A und B Kontraste festgelegt worden, $I-1$ Kontraste für A sowie $J-1$ Kontraste für B, so resultieren aus den Produkten der jeweiligen Kontraste $(I-1)(J-1)$ Kontraste

für die Interaktion A*B. Damit lassen sich auch Interaktionen im Detail untersuchen. Sind in R bzw. SPSS für zwei Faktoren A und B Kontraste definiert worden, so werden automatisch auch diese Kontraste für die Interaktion A*B ausgegeben.

Dies soll am Datensatz 6 (`winer568`) demonstriert werden. Dieser umfasst die Gruppierungsfaktoren A und B sowie den Messwiederholungsfaktor `zeit`. Tabelle 6-7 in Kapitel 6.5.3 enthielt die Anova-Tabelle für das RT-Verfahren. Die Signifikanzen waren dort mittels des ART-Verfahrens verifiziert worden, so dass problemlos die einfach rangtransformierten Daten verwendet werden können. Hier soll jetzt die Interaktion B*Zeit näher betrachtet werden. Hierbei besteht die Vermutung, dass zwischen je zwei aufeinanderfolgenden Zeitpunkten der Anstieg der Werte für die Gruppen von B unterschiedlich stark verläuft.



*Interaktionsplot B*Zeit*

Hierzu werden für den Faktor `zeit` die Standard-Kontraste „wiederholt“ festgelegt, bei denen die Zeitpunkte 1-2, 2-3 und 3-4 verglichen werden, sowie für Faktor B die Effekt-Kodierung

mit SPSS:

Hierzu werden zunächst analog den Berechnungen in Kapitel 6.3 die Daten umstrukturiert, so dass aus den Variablen v_1, \dots, v_4 eine Variable v entsteht. Anschließend wird diese Kriteriumsvariable v über alle Faktoren A, B und Zeit hinweg in Ränge transformiert (Variable RV) und schließlich die Daten wieder in die ursprüngliche Form zurücktransformiert, woraus u.a. die Messwiederholungsvariablen $RV.1, \dots, RV.4$ gebildet werden. Mit diesen Daten kann nun die Varianzanalyse durchgeführt werden. Im Unterkommando `wsfactor` werden mit `Repeated` die gewünschten Kontraste für `zeit` festgelegt, im Unterkommando `contrast` für die Gruppierungsfaktoren A und B.

```
GLM RV.1 RV.2 RV.3 RV.4 by A B
  /wsfactor=Zeit 4 Repeated
  /contrast(A)=Deviation
  /contrast(B)=Deviation
  /plot=profile(Zeit*B)
  /wsdesign=Zeit
  /design=A B A*B.
```

Die Ergebnisse der Varianzanalyse sind in Tabelle 6-7 (Kapitel 6.5.3) zusammengefasst (dort allerdings in der Ausgabe von R). Nachfolgend nun die Ausgabe der Kontraste für den Faktor `zeit`.

Hier interessieren die Ergebnisse des letzten Blocks $\text{Zeit} \times \text{B}$. Daraus geht hervor, dass (vermutlich wegen der geringen Fallzahl) nur zwischen den Zeitpunkten 2 und 3 („Niveau 2 vs. Niveau 3“) ein unterschiedlich starker Anstieg der Werte nachgewiesen werden kann.

Tests der Innersubjektkontraste						
Quelle	Zeit	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Zeit	Niveau 1 vs. Niveau 2	867,000	1	867,000	71,383	,000
	Niveau 2 vs. Niveau 3	1887,521	1	1887,521	122,932	,000
	Niveau 3 vs. Niveau 4	1140,750	1	1140,750	86,777	,000
Zeit * A	Niveau 1 vs. Niveau 2	800,333	1	800,333	65,894	,000
	Niveau 2 vs. Niveau 3	379,688	1	379,688	24,729	,001
	Niveau 3 vs. Niveau 4	280,333	1	280,333	21,325	,002
Zeit * B	Niveau 1 vs. Niveau 2	48,000	1	48,000	3,952	,082
	Niveau 2 vs. Niveau 3	266,021	1	266,021	17,326	,003
	Niveau 3 vs. Niveau 4	48,000	1	48,000	3,651	,092

mit R:

Ausgangsbasis ist der in Kapitel 6.5.3 erstellte Datensatz `winer568t`.

- Zunächst müssen für die Faktoren die Kontraste festgelegt werden. Da die Standard-Kontraste „wiederholt“ in R nicht verfügbar sind, müssen diese als Koeffizienten-Matrix vorgegeben werden.
- Für A und B bietet `contr.sum` die Effekt-Kodierung.
- Die Kontraste werden wie oben über die Funktion `gls` des Pakets `nlme` getestet. Allerdings muss in diesem Fall der Faktor `zeit` als Messwiederholungsfaktor deklariert werden. Dies erfolgt in `gls` über die Spezifikation der Fallkennung (`Vpn`) sowie der Struktur für die Kovarianzen der Messwiederholungsvariablen, die hier mit „*compound symmetry*“ festgelegt wird, was der sonst üblichen Sphärizität entspricht (vgl. Kapitel 5.2):

```
corr = corCompSymm(, form= ~ 1 | Vpn)
```

Die Kommandos lauten dann:

```
library(nlme)
cont4 <- matrix( c(1,-1,0,0, 0,1,-1,0, 0,0,1,-1), ncol=3)
contrasts(winer568t$Zeit) <- cont4
contrasts(winer568t$A) <- contr.sum
contrasts(winer568t$B) <- contr.sum
aovgls <- gls(Rx~A*B*Zeit, data=winer568t,
             corr = corCompSymm(, form= ~ 1 | Vpn))
summary(aovgls)
```

Zunächst vorab die oben erzeugte Kontrastmatrix `cont4`:

```
> cont4
      [,1] [,2] [,3]
[1,]    1    0    0
[2,]   -1    1    0
[3,]    0   -1    1
[4,]    0    0   -1
```

Hier der Teil der Ausgabe, der die Kontrast-Tests enthält:

Coefficients:				
	Value	Std.Error	t-value	p-value
(Intercept)	24.500000	1.2012621	20.395216	0.0000
A1	2.187500	1.2012621	1.821001	0.0780
B1	-3.437500	1.2012621	-2.861574	0.0074
Zeit1	-15.083333	0.7663867	-19.681101	0.0000
Zeit2	-21.666667	0.8849471	-24.483573	0.0000
Zeit3	-15.708333	0.7663867	-20.496616	0.0000
A1:B1	0.500000	1.2012621	0.416229	0.6800
A1:Zeit1	-2.104167	0.7663867	-2.745568	0.0098
A1:Zeit2	3.958333	0.8849471	4.472960	0.0001
A1:Zeit3	4.395833	0.7663867	5.735790	0.0000
B1:Zeit1	0.354167	0.7663867	0.462125	0.6471
B1:Zeit2	2.708333	0.8849471	3.060447	0.0044
B1:Zeit3	0.354167	0.7663867	0.462125	0.6471
A1:B1:Zeit1	0.750000	0.7663867	0.978618	0.3351
A1:B1:Zeit2	1.500000	0.8849471	1.695017	0.0998
A1:B1:Zeit3	0.875000	0.7663867	1.141721	0.2620

Hier interessieren die Ergebnisse der Zeilen B1:Zeit. Daraus geht hervor, dass (vermutlich wegen der geringen Fallzahl) nur zwischen den Zeitpunkten 2 und 3 (B1:Zeit2) ein unterschiedlich starker Anstieg der Werte nachgewiesen werden kann.

Anzumerken ist noch, dass über `anova(aovgls)` auch eine Anova-Tabelle erzeugt werden kann:

Denom. DF: 32				
	numDF	F-value	p-value	
(Intercept)	1	415.9648513	<.0001	
A	1	3.3160463	0.0780	
B	1	8.1886042	0.0074	
Zeit	3	235.4226927	<.0001	
A:B	1	0.1732465	0.6800	
A:Zeit	3	25.8348225	<.0001	
B:Zeit	3	4.8246777	0.0070	
A:B:Zeit	3	0.9709950	0.4185	

A. Anhang

1. Umstrukturieren von Messwiederholungen in SPSS

Dieses ist z.B. erforderlich zur Rangbildung von Messwiederholungen.

1.1 Umstrukturieren von Messwiederholungen in Fälle

Vorzunehmen im Menü: „Daten -> Umstrukturieren“

1.1.1 ein Faktor und eine Analyse-Variable

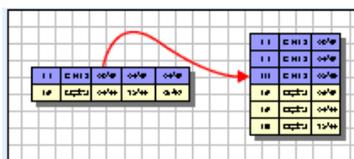
Als Beispiel dient der Datensatz 4 (vgl. Kapitel 5)

	Geschlecht	t1	t2	t3
1	1	4	7	2
2	1	3	5	1
3	1	7	9	6
4	1	6	6	2
5	1	5	5	1
6	2	8	2	5
7	2	4	1	1
8	2	6	3	4
9	2	9	5	2
10	2	7	1	1

• Datenumstrukturierung

1. Option:

Umstrukturieren ausgewählter Variablen in Fälle



Folgende Möglichkeiten stehen Ihnen zur Verfügung:

- Umstrukturieren ausgewählter Variablen in Fälle

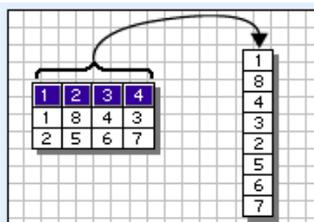
Verwenden Sie diese Option, wenn jeder Fall in den aktuellen Daten Variablen enthält, die im neuen Datenblatt in Gruppen verwandter Fälle angeordnet werden sollen.

-> Weiter

• Anzahl der Variablengruppen

1. Option:

Eine (Variablengruppe)



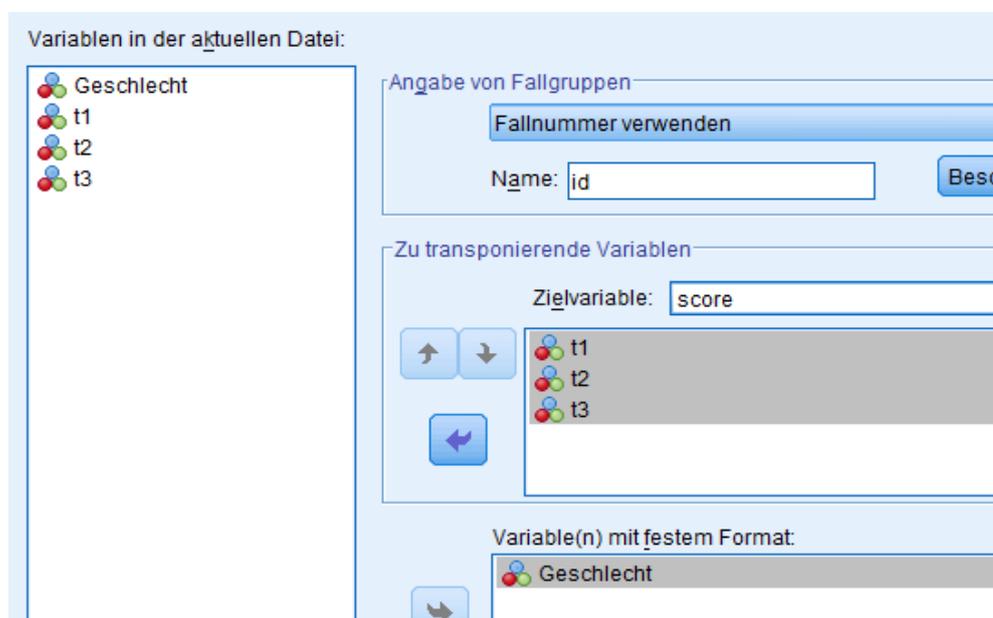
Wieviele Variablengruppen möchten Sie umstrukturieren?

- Eine (beispielsweise w1, w2 und w3)

-> Weiter

- **Auswählen von Variablen**

- Fallnummer verwenden,
 - kann eine vorhandene Fallkennung sein, z.B. Vpn
 - ist aber frei wählbar
 - erhält standardmäßig den Namen id
- zu transponierende Variablen:
 - hier die Messwiederholungsvariablen eintragen
 - und einen gemeinsamen Namen geben, hier: „score“
- Variablen mit festem Format:
 - hier die "konstanten" Variablen (ohne Messwiederholung) eintragen
 - (z.B. Alter, Geschlecht etc)

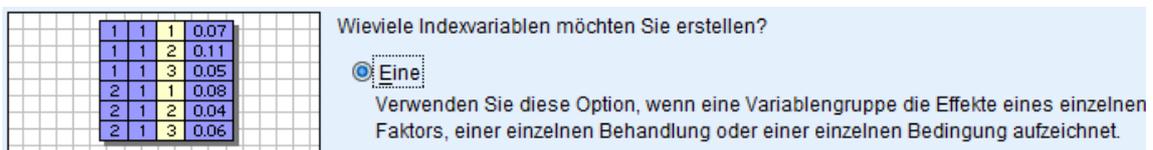


-> Weiter

- **Wieviele Indexvariablen möchten Sie erstellen?**

1. Option:

Eine (Indexvariablen)



-> Weiter

- **Erstellen einer Indexvariablen**

(Diese kann numerisch oder alphanumerisch sein.)

- Art des Indexwertes:
 - fortlaufende Zahlen

b. Name und Label der Indexvariablen:

kann frei gewählt werden (standardmäßig: Index1), hier: „Zeit“

	Name	Variablenlabel	Stufen	Indexwerte
1	Zeit		3	1, 2, 3

-> Weiter (es folgen dann noch Optionen) oder Fertigstellen

- **Optionen**

a. Verarbeitung nicht ausgewählter Variablen (die oben weder als zu transponierende noch als "konstante" deklariert worden waren):

(normalerweise) beibehalten und als Variablen mit festem Format behandeln

b. System Missing: Einen Fall in der neuen Datei erstellen

-> Weiter

Die hier aufgeführten Schritte können auch über die SPSS-Syntax realisiert werden:

```
Varstocases
  /Id=id
  /Make score from t1 t2 t3
  /index=Zeit(3)
  /keep=patients
  /null=keep.
```

Das Ergebnis der Umstrukturierung:

	id	Geschlecht	Zeit	score
1	1	1	1	4
2	1	1	2	7
3	1	1	3	2
4	2	1	1	3
5	2	1	2	5
6	2	1	3	1
7	3	1	1	7
8	3	1	2	9
9	3	1	3	6
10	4	1	1	6
11	4	1	2	6
12	4	1	3	2

1. 1. 2 mehrere Faktoren und eine Analyse-Variablen

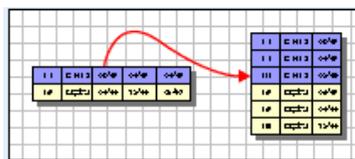
Als Beispiel dient der Datensatz 5 (vgl. Kapitel 5)

	Geschlecht	v1	v2	v3	v4	v5	v6	v7	v8	v9
1	1	3	3	1	4	4	2	5	4	3
2	1	2	0	0	3	2	2	4	3	3
3	1	5	4	3	5	3	3	6	3	4
4	1	3	5	2	4	4	3	4	4	4
5	2	2	2	1	2	2	2	5	2	3
6	2	4	1	0	3	2	1	5	2	2
7	2	3	2	1	3	2	1	4	3	2
8	2	1	3	0	5	2	1	6	3	3

• Datenumstrukturierung

1. Option:

Umstrukturieren ausgewählter Variablen in Fälle



Folgende Möglichkeiten stehen Ihnen zur Verfügung:

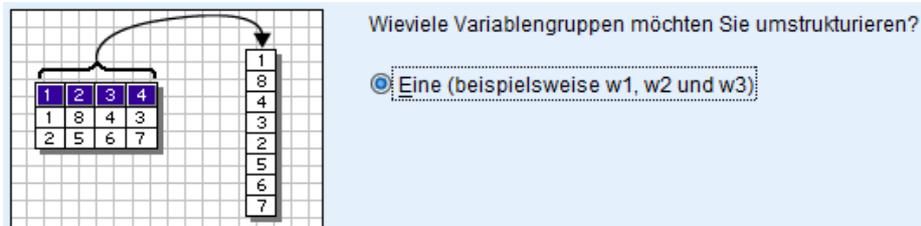
- Umstrukturieren ausgewählter Variablen in Fälle

Verwenden Sie diese Option, wenn jeder Fall in den aktuellen Daten Variablen enthält, die im neuen Datenblatt in Gruppen verwandter Fälle angeordnet werden sollen.

-> Weiter

- **Anzahl der Variablengruppen**

1. Option: Eine (Variablengruppe)



-> Weiter

- **Auswählen von Variablen**

a. Fallnummer verwenden,

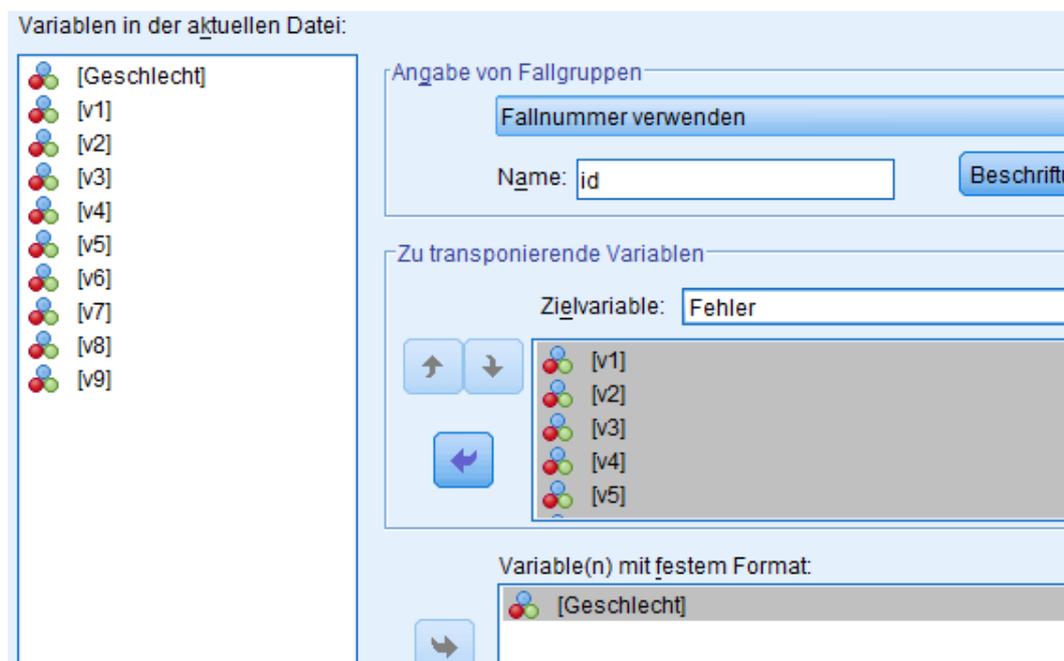
- kann eine vorhandene Fallkennung sein, z.B. Vpn
- ist aber frei wählbar
- erhält standardmäßig den Namen id

b. zu transponierende Variablen:

hier die Messwiederholungsvariablen eintragen und einen gemeinsamen Namen geben, hier: „Fehler“

c. Variablen mit festem Format:

hier die "konstanten" Variablen (ohne Messwiederholung) eintragen (z.B. Alter, Geschlecht etc)



-> Weiter

- **Wieviel Indexvariablen möchten Sie erstellen?**

2. Option:

Mehrere (Indexvariablen) und Anzahl der Messwiederholungsfaktoren festlegen

1	1	1	1	0.07
1	1	1	2	0.11
1	1	1	3	0.05
1	1	2	1	0.08
1	1	2	2	0.04
1	1	2	3	0.06

 Mehrere
Wie viele?

Verwenden Sie diese Option, wenn eine Variablen­gruppe die Effekte mehrerer Faktoren, Behandlungen oder Bedingungen aufzeichnet.

-> Weiter

- **Erstellen mehrerer Indexvariablen)**

In der folgenden Tabelle müssen für jeden Messwiederholungsfaktor Name und wahlweise Label frei gewählt (standardmäßig: Index1, Index2) sowie für jede die Anzahl der Stufen festgelegt werden, hier „Medikament“ und „Aufgabe“. Hierbei ist die Reihenfolge zu beachten: in der Variablenreihenfolge variiert der erste Faktor am langsamsten, der letzte am schnellsten. Und das Produkt der Stufen muss die Anzahl der Messwiederholungsvariablen ergeben:

Namen, Label und Anzahl der Ebenen für Indexvariablen:

	Name	Variablenlabel	Stufen	Indexwerte
1	Medikament		3	1, 2, 3
2	Aufgabe		3	1, 2, 3

Gesamtzahl kombinierter Ebenen (Produkt): 9

-> Weiter

- Optionen

a. Verarbeitung nicht ausgewählter Variablen (die oben weder als zu transponierende noch als "konstante" deklariert worden waren): (normalerweise) beibehalten und als Variablen mit festem Format behandeln

b. System Missing: Einen Fall in der neuen Datei erstellen

Verarbeitung nicht ausgewählter Variablen

Variable(n) aus neuer Datendatei entfernen

Beibehalten und als Variable(n) mit festem Format behandeln

System Missing (fehlender Wert) oder leere Werte in allen transponierten Variablen

Einen Fall in der neuen Datei erstellen

Daten verwerfen

Variable zum Zählen von Fällen

Anzahl neuer Fälle zählen, die vom Fall in den aktuellen Daten erstellt wurden

Name:

Beschriftung:

-> Weiter

-> Fertigstellen

Wenn keine Namen festgelegt worden waren, hat die Analyse-Variablen anschließend die Namen `trans1` und `Index1`, `Index2`,... sind standardmäßig die Kennzeichnungen der Messwiederholung für die jeweiligen Faktoren.

Die hier aufgeführten Schritte können auch über die SPSS-Syntax realisiert werden:

```
Varstocases
  /Id=id
  /make Fehler from v1 v2 v3 v4 v5 v6 v7 v8 v9
  /index=Medikament(3) Aufgabe(3)
  /keep=Geschlecht
  /null=keep.
```

Das Ergebnis der Umstrukturierung:

	id	Geschlecht	Medikament	Aufgabe	Fehler
1	1	1	1	1	3
2	1	1	1	2	3
3	1	1	1	3	1
4	1	1	2	1	4
5	1	1	2	2	4
6	1	1	2	3	2
7	1	1	3	1	5
8	1	1	3	2	4
9	1	1	3	3	3
10	2	1	1	1	2
11	2	1	1	2	0
12	2	1	1	3	0

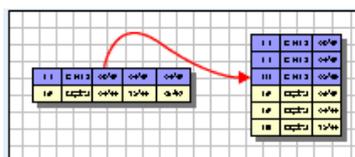
1. 1. 3 ein Faktor und mehrere Analyse-Variablen

Als Beispiel dient der Datensatz 4 (vgl. Kapitel 5), wobei die 3 Aufgaben nicht als Faktor, sondern als 3 Variablen interpretiert werden und lediglich ein Faktor Medikament vorhanden ist.

• Datenumstrukturierung

1. Option: Umstrukturieren ausgewählter Variablen in Fälle

-> Weiter



Folgende Möglichkeiten stehen Ihnen zur Verfügung:

- Umstrukturieren ausgewählter Variablen in Fälle

Verwenden Sie diese Option, wenn jeder Fall in den aktuellen Daten Variablen enthält, die im neuen Datenblatt in Gruppen verwandter Fälle angeordnet werden sollen.

• Anzahl der Variablengruppen

2. Option:

Mehrere (Variablengruppen) sowie Anzahl der Analyse-Variablen festlegen (hier 3)

1	2	3	4	5	6
1	8	4	0.3	0.9	0.4
2	5	6	0.7	0.1	0.7

1	0.3
8	0.9
4	0.4
2	0.7
5	0.1
6	0.7

Mehrere (beispielsweise w1, w2, w3 und h1, h2, h3, usw.)

Anzahl

-> Weiter

- **Auswählen von Variablen**

a. Fallnummer verwenden,

- kann eine vorhandene Fallkennung sein, z.B. Vpn
- ist aber frei wählbar
- erhält standardmäßig den Namen id

b. zu transponierende Variablen:

hier die Messwiederholungsvariablen für die 1. abhängige Variable eintragen und bei „Zielvariable“ einen gemeinsamen Namen geben, hier: „Aufgabe1“ diesen Schritt dann für die anderen abhängigen Variablen wiederholen, indem im Pulldown-Menü rechts neben der Zielvariablen nacheinander die nächsten Variablen ausgewählt werden, deren Voreinstellung `trans1`, `trans2`, ... ist.

c. Variablen mit festem Format:

hier die "konstanten" Variablen (ohne Messwiederholung) eintragen (z.B. Alter, Geschlecht etc)

Variablen in der aktuellen Datei:

- [Geschlecht]
- [v1]
- [v2]
- [v3]
- [v4]
- [v5]
- [v6]
- [v7]
- [v8]
- [v9]

Angabe von Fallgruppen

Fallnummer verwenden

Name:

Zu transponierende Variablen

Zielvariable:

- [v1]
- [v4]
- [v7]

Variable(n) mit festem Format:

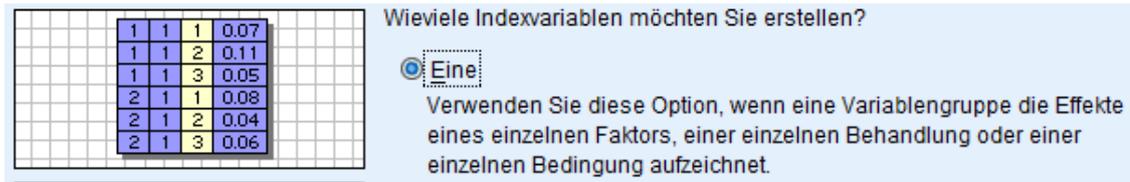
- [Geschlecht]

-> Weiter

- **Erstellen von Indexvariablen)**

1. Option:

Eine (Indexvariablen)



Wieviele Indexvariablen möchten Sie erstellen?

Eine

Verwenden Sie diese Option, wenn eine Variablengruppe die Effekte eines einzelnen Faktors, einer einzelnen Behandlung oder einer einzelnen Bedingung aufzeichnet.

1	1	1	0.07
1	1	2	0.11
1	1	3	0.05
2	1	1	0.08
2	1	2	0.04
2	1	3	0.06

Liegt ein mehrfaktorielles Design vor, wie etwa im vorigen Abschnitt, so können bei der 2. Option die Anzahl der Messwiederholungsfaktoren festgelegt werden.

-> Weiter

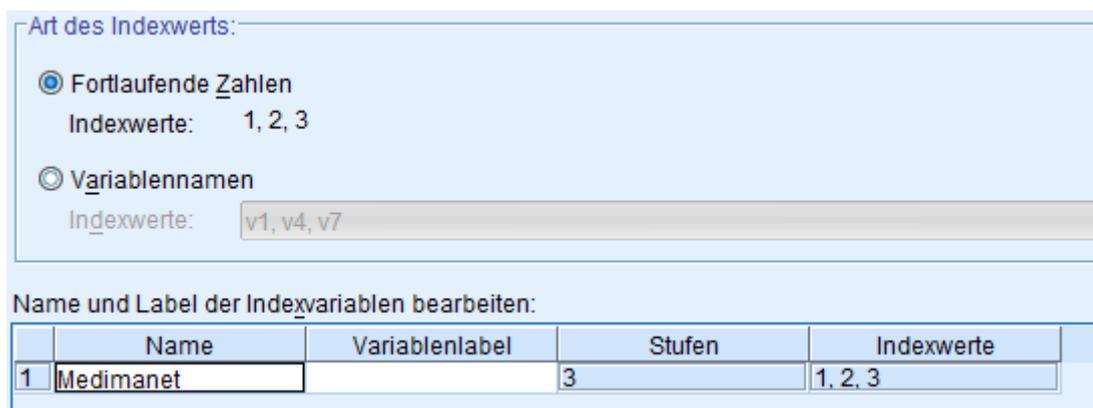
- **Erstellen einer Indexvariablen**

(Diese kann numerisch oder alphanumerisch sein.)

a. Art des Indexwertes:
fortlaufende Zahlen

b. Name und Label der Indexvariablen:

kann frei gewählt werden (standardmäßig: Index1), hier: „Medikament“. Die Stufenzahl ergibt sich aus den anderen Angaben.



Art des Indexwerts:

Fortlaufende Zahlen
Indexwerte: 1, 2, 3

Variablennamen
Indexwerte: v1, v4, v7

Name und Label der Indexvariablen bearbeiten:

	Name	Variablenlabel	Stufen	Indexwerte
1	Medimanet		3	1, 2, 3

-> Weiter

- **Optionen**

a. Verarbeitung nicht ausgewählter Variablen (die oben weder als zu transponierende noch als "konstante" deklariert worden waren):

(normalerweise) beibehalten und als Variablen mit festem Format behandeln

b. System Missing: Einen Fall in der neuen Datei erstellen

Verarbeitung nicht ausgewählter Variablen

Variable(n) aus neuer Datendatei entfernen

Beibehalten und als Variable(n) mit festem Format behandeln

System Missing (fehlender Wert) oder leere Werte in allen transponierten Variablen

Einen Fall in der neuen Datei erstellen

Daten verwerfen

Variable zum Zählen von Fällen

Anzahl neuer Fälle zählen, die vom Fall in den aktuellen Daten erstellt wurden

Name:

Beschriftung:

Falls keine Namen vereinbart worden waren, haben die Analyse-Variablen anschließend die Namen `trans1`, `trans2`, ... und `Index1` ist standardmäßig der Kennzeichnung der Messwiederholung.

Die hier aufgeführten Schritte können auch über die SPSS-Syntax realisiert werden:

```
Varstocases
  /Id=id
  /make Aufgabe1 from v1 v4 v7
  /make Aufgabe2 from v2 v5 v8
  /make Aufgabe3 from v3 v6 v9
  /Index=Medikament(3)
  /Keep=Geschlecht
  /Null=keep.
```

Das Ergebnis der Umstrukturierung:

	id	Geschlecht	Medikament	Aufgabe1	Aufgabe2	Aufgabe3
1	1	1	1	3	3	1
2	1	1	2	4	4	2
3	1	1	3	5	4	3
4	2	1	1	2	0	0
5	2	1	2	3	2	2
6	2	1	3	4	3	3
7	3	1	1	5	4	3
8	3	1	2	5	3	3
9	3	1	3	6	3	4
10	4	1	1	3	5	2
11	4	1	2	4	4	3
12	4	1	3	4	4	4

1.2 Umstrukturieren von Fälle in Messwiederholungen

Vorzunehmen im Menü: „Daten -> Umstrukturieren“

- **Datenumstrukturierung**

2. Option: Umstrukturieren ausgewählter Variablen in Fälle

11	EM13	0010	0010
11	EM14	0010	0010
11	EM13	0010	0010
12	EM13	0010	0010
12	EM14	0010	0010
12	EM13	0010	0010

- **Umstrukturieren ausgewählter Fälle in Variablen**

Verwenden Sie diese Option, wenn Gruppen verwandter Fälle vorliegen, die neu angeordnet werden sollen, sodass die Daten aus den einzelnen Gruppen im neuen Datenblatt als einzelner Fall dargestellt werden.

-> Weiter

- **Auswählen von Variablen**

a. Bezeichnervariablen:

Fallkennzeichnung/nummer (z.B. id oder Vpn)

b. Indexvariable:

Kennzeichnungen der Messwiederholung, hier „Zeit“

(z.B. 1-faktoriell: Index1 bzw. mehrfaktoriell Index1, Index2,...)

Variablen in der aktuellen Datei:

- [patients]
- [score]
- Rank of score by id [Rscore]

Bezeichnervariable(n):

- id

Indexvariable(n):

- Zeit

Alle übrigen Variablen werden automatisch „sinnvoll“ als konstante oder Messwiederholungsvariable zugeordnet.

-> Weiter

- **Sortieren von Daten**

2. Option:

Nein - Daten wie gegenwärtig sortiert verwenden

(Bei 1. Option werden zuerst alle Wiederholungen einer Analyse-Variablen hintereinander ausgegeben, vor denen der nächsten Analysevariablen,

bei 2. Option werden zuerst die ersten Werte aller Analyse-Variablen hintereinander aus-

gegeben, vor allen Werten der zweiten Messwiederholung etc)

Sollen die aktuellen Daten sortiert werden?

Ja - Daten werden nach Bezeichner- und Indexvariablen sortiert.

The image shows two SPSS data grids. The left grid contains data with columns for group, condition, and time. The right grid shows the same data sorted by group and then by time. A dialog box asks if the current data should be sorted, with the 'Yes' option selected.

-> Weiter

- **Optionen**

Anordnung der neuen Variablengruppen

Nach ursprünglicher Variable sortieren (z. B.: w1 w2 w3, h1 h2 h3)

Nach Index sortieren (z. B.: w1 h1, w2 h2, w3 h3)

Variable zum Zählen von Fällen

Anzahl der Fälle in den aktuellen Daten zählen, mit denen ein neuer Fall erstellt wird

Name:

Beschriftung:

Indikatorvariablen

Indikatorvariablen erstellen

Stammname:

The image shows the 'Options' dialog box in SPSS. It has three sections: 'Anordnung der neuen Variablengruppen' with two radio buttons, 'Variable zum Zählen von Fällen' with a checkbox and two text fields, and 'Indikatorvariablen' with a checkbox and a text field.

Die Optionen sind i.a. nicht erforderlich.

-> Weiter

-> Fertigstellen

Die neuen Namen der Messwiederholungen der einzelnen Analyse-Variablen sind Name.1, Name2, .. (wenn ein Name vorgegeben wurde) andernfalls trans.1, trans.2, ... Bei mehrfaktoriellen Designs haben diese jeweils den Zusatz der Kennzeichnung der Messwiederholung z.B. .1.1, .1.2, ..., 2.1, 2.2, ...

Die hier aufgeführten Schritte können auch über die SPSS-Syntax realisiert werden:

```
Sort cases by id Zeit.
casestovars
  /Id=id
  /index=Zeit
  /groupby=variable.
```

Und das Ergebnis der Umstrukturierung:

	id	patients	score.1	score.2	score.3	Rscore.1	Rscore.2	Rscore.3
1	1	1	4	7	2	2,000	3,000	1,000
2	2	1	3	5	1	2,000	3,000	1,000
3	3	1	7	9	6	2,000	3,000	1,000
4	4	1	6	6	2	2,500	2,500	1,000
5	5	1	5	5	1	2,500	2,500	1,000
6	6	2	8	2	5	3,000	1,000	2,000
7	7	2	4	1	1	3,000	1,500	1,500
8	8	2	6	3	4	3,000	1,000	2,000
9	9	2	9	5	2	3,000	2,000	1,000
10	10	2	7	1	1	3,000	1,500	1,500

2. Spezielle robuste F-Tests und andere Statistiken

Im Folgenden werden drei robuste F-Tests vorgestellt, deren Formeln in der Literatur nicht weit verbreitet sind und daher hier zitiert werden.

2.1 Box-Korrektur für heterogene Varianzen

Es liegen k Gruppen (Haupteffekt oder Interaktion) mit Varianzen s_i^2 vor. Der F-Test

$$F = \frac{MS_{Effekt}}{MS_{Fehler}}$$

kann bzgl. der Heterogenität der Varianzen korrigiert werden, indem die Zähler- und Nennerfreiheitsgrade adjustiert (genauer: reduziert) werden. Die Zählerfreiheitsgrade $df1$ werden dabei mit ε_1 multipliziert, die Nennerfreiheitsgrade $df2$ mit ε_2 . Diese Korrekturfaktoren errechnen sich wie folgt:

$$\bar{s}^2 = (\sum s_i^2)/k$$

$$c^2 = \left(\sum (s_i^2 - \bar{s}^2)^2 \right) / (k \cdot \bar{s}^4)$$

$$\varepsilon_1 = \left(1 + \frac{k-2}{k-1} c^2 \right)^{-1} \quad \varepsilon_2 = (1 + c^2)^{-1}$$

Hierbei lassen sich \bar{s}^2 als durchschnittliche Varianz und c^2 als Streuung der Varianzen interpretieren. Es ist leicht zu erkennen, dass im Falle gleicher Varianzen $c^2=0$ wird und damit ε_1 und ε_2 den Wert 1 bekommen.

2.2 Brown-Forsythe F-Test für inhomogene Varianzen

Es liegen k Gruppen (Haupteffekt oder Interaktion) mit Varianzen s_i^2 , Zellenbesetzungen n_i vor. Brown & Forsythe bilden den folgenden Quotienten, der annähernd F-verteilt ist:

$$F = \frac{SS_{Effekt}}{SS_{Fehler}}$$

Hierbei errechnet sich SS_{Error} (mit $n = \sum n_i$)

$$SS_{Error} = \sum \left(1 - \frac{n_i}{n} \right) s_i^2$$

Die Nennerfreiheitsgrade des F-Tests berechnen sich

$$df = \left(\sum \frac{m_i^2}{n_i - 1} \right)^{-1} \quad m_i = \left(1 - \frac{n_i}{n} \right) s_i^2 / (SS_{Error})$$

2.3 Box-Andersen F-Test für nichtnormalverteilte Variablen

Bei diesem modifizierten F-Test werden dessen Zähler- und Nennerfreiheitsgrade mit dem Parameter d multipliziert. Dieser errechnet sich im Wesentlichen aus der Varianz und dem Exzess der Variablen x . Die folgende Berechnung des Korrekturparameters d ist gültig für annähernd gleiche n_i . Sei daher n die Anzahl der Beobachtungen pro Gruppe. Es sei erwähnt, dass es auch eine etwas kompliziertere Formel für stark differierende n_i gibt.

$$S_2 = \sum_i^k \sum_j^n (x_{ij} - \bar{x})^2 \quad S_4 = \sum_i^k \sum_j^n (x_{ij} - \bar{x})^4$$

Daraus werden zwei Zwischengrößen berechnet:

$$k_2 = S_2 / (n - 1)$$

$$k_4 = [n(n + 1)S_4 - 3(n - 1)S_2^2] / [(n - 1)(n - 2)(n - 3)]$$

Schließlich errechnet sich hieraus d als

$$d = 1 + \frac{1}{n} \frac{k_4}{k_2^2}$$

2.4 Box-Cox-Transformationen

Hier geht es darum, einen passenden Parameter a zu finden, so dass die Funktion, angewandt auf die abhängige Variable, varianzstabilisierend wirkt.

$$f(x) = \frac{x^a - 1}{a}$$

Für den Parameter a gilt:

- $0 < a < 1$ rechtsschiefe Verteilungen symmetrisch machen
- $1 < a$ linksschiefe Verteilungen symmetrisch machen

Schließlich gilt, dass $f(x) \rightarrow \log(x)$ für $a \rightarrow 0$.

Mehr dazu unter:

<http://de.wikipedia.org/wiki/Box-Cox-Transformation>

2.5 Fishers combined probability test

Mit *Fishers combined probability test* können mehrere unabhängig voneinander gewonnene Testergebnisse zur gleichen Hypothese H_0 über deren p-Werte zusammengefasst werden. Das Verfahren ist für beliebige Tests anwendbar, also z.B. auch für den W-Test von Shapiro und Wilk zur Überprüfung eines Merkmals auf Normalverteilung, etwa für k Variablen oder k Stichproben. Werden für k Tests die p-Werte P_1, \dots, P_k erzielt, dann wird mit der folgenden Testgröße X die Hypothese geprüft, dass für alle k Tests H_0 richtig ist:

$$X = -2[\ln(P_1) + \ln(P_2) + \dots + \ln(P_k)]$$

X ist χ^2 -verteilt mit $2k$ Freiheitsgraden.

Mehr dazu unter https://en.wikipedia.org/wiki/Fishers_method

3. R-Funktionen

Die folgenden Funktionen sind alle im Verzeichnis

<http://www.uni-koeln.de/~luepsen/R/>

zu finden und können von dort heruntergeladen werden.

3.1 **box.f: Box-F-Test für inhomogene Varianzen**

Durchführung einer 1- oder 2-faktoriellen Varianzanalyse (ohne Messwiederholungen) unter Verwendung der robusten F-Tests von Box (vgl. Anhang 2.1) zur Kompensierung von Varianzinhomogenitäten.

Aufruf: `box.f (Modell, Dataframe)`

Aufrufparameter:

Modell	varianzanalytisches Modell (vgl. Funktion aov) Beispiel: $x \sim A*B$
Dataframe	Datensatz, Objekt vom Type Dataframe

3.2 **bf.f: Brown & Forsythe-F-Test für inhomogene Varianzen**

Durchführung einer 1- oder 2-faktoriellen Varianzanalyse (ohne Messwiederholungen) unter Verwendung der robusten F-Tests von Brown & Forsythe (vgl. Anhang 2.2) zur Kompensierung von Varianzinhomogenitäten.

Aufruf: `bf.f (Modell, Dataframe)`

Aufrufparameter:

Modell	varianzanalytisches Modell (vgl. Funktion aov) Beispiel: $x \sim A*B$
Dataframe	Datensatz, Objekt vom Type Dataframe

3.3 **box.andersen.f: F-Test für nichtnormalverteilte Variablen**

Durchführung einer 1- oder 2-faktoriellen Varianzanalyse (ohne Messwiederholungen) unter Verwendung der robusten F-Tests von Box & Andersen (vgl. Anhang 2.3) zur Kompensierung von Abweichungen von der Normalverteilung.

Aufruf: `box.andersen.f (Modell, Dataframe)`

Aufrufparameter:

Modell	varianzanalytisches Modell (vgl. Funktion aov) Beispiel: $x \sim A*B$
Dataframe	Datensatz, Objekt vom Type Dataframe

Ergebnisobjekte:

anova	Anova-Tabelle
eps	Korrekturfaktor d zur Korrektur der Freiheitsgrade

3.4 **boxm.test: Test auf Homogenität von Kovarianzmatrizen**

Durchführung des Box M-Tests zur Prüfung der Homogenität von Kovarianzmatrizen. Dieser Code ist von Michail T. Tsagris (School of Mathematical Sciences, University of Nottingham).

Aufruf: `boxm.test (Messwiederholungsvariablen, Gruppierungsfaktor, a=0.05)`

Aufrufparameter:

Messwiederholungsvariablen	Variablen des Messwiederholungsfaktors als Dataframe
Gruppierungsfaktor	Vektor mit den Werten des Gruppierungsfaktors
a	Signifikanzniveau (default: 0.05)

Beispiel: `boxm.test (winer[, c("V3", "V4", "V5")], winer$V2)`

3.5 **ats.2 und ats.3: 2- bzw. 3-faktorielle Varianzanalyse**

ats.2 führt eine 2-faktorielle Varianzanalyse (ohne Messwiederholungen) nach dem Verfahren von Akritas, Arnold und Brunner [26] durch sowie ats.3 eine 3-faktorielle Analyse. Errechnet wird die F-verteilte ATS (anova type statistic). Leere Zellen sind nicht erlaubt.

Aufruf: `ats.2 (Modell, Dataframe)`
bzw. `ats.3 (Modell, Dataframe)`

Aufrufparameter:

Modell	varianzanalytisches Modell (vgl. Funktion aov) Beispiel: $x \sim A*B$
Dataframe	Datensatz, Objekt vom Type Dataframe

3.6 **np.anova: nichtparametrische Varianzanalyse mittels der Verfahren von Puri & Sen und van der Waerden**

np.anova führt eine mehrfaktorielle Varianzanalyse (mit und ohne Messwiederholungen) wahlweise nach den Verfahren von Puri & Sen (L-Statistik, verallgemeinerte Kruskal-Wallis- und Friedman-Analysen) oder van der Waerden durch. Im Fall von Messwiederholungen muss der Datensatz die gleiche Struktur haben, wie sie von aov oder ezANOVA gefordert wird. Bei dem Verfahren von van der Waerden ist nur maximal ein Messwiederholungsfaktor möglich.

Aufruf: `np.anova (Modell, Dataframe)` Methode von Puri & Sen
bzw. `np.anova (Modell, Dataframe, method=1)` Methode von van der Waerden

Aufrufparameter:

Modell	varianzanalytisches Modell (vgl. Funktion aov) Beispiele: $x \sim A*B$ oder $\text{score} \sim \text{gruppe}*\text{Zeit}+\text{Error}(\text{Vpn}/\text{Zeit})$
Dataframe	Datensatz, Objekt vom Type Dataframe
method	0 (Methode von Puri & Sen) oder 1 (Methode von van der Waerden)
compact	im Falle von Messwiederholungen: T: alle Tests in einer Dataframe-Tabelle (default) F: für jeden Fehlerterm eine getrennte Tabelle (wie bei <code>summary(aov)</code>)

3.7 **art1.anova: nichtparametrische Varianzanalyse nach dem ART-Verfahren (nur Gruppierungsfaktoren)**

`art1.anova` führt eine mehrfaktorielle Varianzanalyse ohne Messwiederholungen nach dem ART-Verfahren (Aligned Rank Transform) durch.

Aufruf: `art1.anova (Modell, Dataframe, method=..., main=...)`

Aufrufparameter:

Modell	varianzanalytisches Modell (vgl. Funktion aov) Beispiel: $x \sim A*B$
Dataframe	Datensatz, Objekt vom Type Dataframe
method	0: Berechnung der Residuen über eine Regression (default) 1: Berechnung der Residuen als Abweichungen vom Zellenmittelwert
main	F: für die Tests der Haupteffekte nur das RT-Verfahren (default) T: für die Tests der Haupteffekte ebenfalls das ART-Verfahren

3.8 **art2.anova: nichtparametrische Varianzanalyse nach dem ART-Verfahren (nur Messwiederholungsfaktoren)**

`art2.anova` führt eine mehrfaktorielle Varianzanalyse mit Messwiederholungen auf zwei Faktoren nach dem ART-Verfahren (Aligned Rank Transform) durch. Für die Haupteffekte werden die Tests nach dem einfachen RT-Verfahren ausgegeben.

Aufruf: `art2.anova (Modell, Dataframe, method=...)`

Aufrufparameter:

Modell	varianzanalytisches Modell (vgl. Funktion aov) Beispiel: $x \sim \text{Medi}* \text{Aufgabe} + \text{Error}(\text{Vpn}/(\text{Medi}* \text{Aufgabe}))$
Dataframe	Datensatz, Objekt vom Type Dataframe
method	0: Berechnung der Residuen über eine Regression (default) 1: Berechnung der Residuen als Abweichungen vom Zellenmittelwert

3.9 **art3.anova: nichtparametrische Varianzanalyse nach dem ART-Verfahren (für gemischte Versuchspläne)**

`art3.anova` führt eine mehrfaktorielle Varianzanalyse für Versuchspläne mit mindestens einem Gruppierungsfaktor und ein oder zwei Messwiederholungsfaktoren nach dem ART-Verfahren (Aligned Rank Transform) durch. Für die Haupteffekte werden die Tests nach dem einfachen RT-Verfahren ausgegeben.

Aufruf: `art3.anova (Modell, Dataframe, method=...)`

Aufrufparameter:

Modell	varianzanalytisches Modell (vgl. Funktion <code>aov</code>) Beispiel: <code>score ~ gruppe*Zeit+Error(Vpn/Zeit)</code>
Dataframe	Datensatz, Objekt vom Type <code>Dataframe</code>
method	0: Berechnung der Residuen über eine Regression (default) 1: Berechnung der Residuen als Abweichungen vom Zellenmittelwert

Literaturhinweise

Allgemeines zur nichtparametrischen Statistik

- [1] G.A.Lienert (1987): *Verteilungsfreie Methoden in der Biostatistik* - Band 1 und 2, (der „Klassiker“)
- [2] J. Bortz, G.A. Lienert, K. Boehnke (2008): *Verteilungsfreie Methoden in der Biostatistik*, Springer, (gekürzte Neuauflage des Klassikers)
- [3] E. Brunner & U. Munzel (2002): *Nichtparametrische Datenanalyse - unverbundene Stichproben*, Springer, ISBN 3-540-43375-9
- [4] W.J. Conover (1980): *Practical nonparametric Statistics*, Wiley, (Standardverfahren mit einigen Zusatzinformationen)
- [5] David M. Erceg-Hurn and Vikki M. Miroseovich (2008): *Modern robust statistical methods*, American Psychologist, Vol. 63, No. 7, 591–601
- [6] Conover, W. J. & Iman, R. L. (1981): *Rank transformations as a bridge between parametric and nonparametric statistics*. American Statistician 35 (3): 124–129.
- [7] Cochran, W.G. (1950): *The comparison of percentages in matched samples*. Biometrika 3
- [8] Lunney, G.H. (1970): *Using Analysis of Variance with a dichotomous dependent variable: an empirical study*. Journal of Educational Measurement Volume 7, Issue 4
- [9] Haiko Lüpsen (2014): *Multiple Mittelwertvergleiche - parametrisch und nichtparametrisch - sowie alpha-Adjustierungen mit praktischen Anwendungen mit R und SPSS*, Universität zu Köln, <http://www.uni-koeln.de/~luepsen/statistik/buch/mult-comp.pdf>

Nichtparametrische Varianzanalysen

- [11] http://en.wikipedia.org/wiki/ANOVA_on_ranks
- [12] Puri, M.L. & Sen, P.K. (1985): *Nonparametric Methods in General Linear Models*, Wiley, New York
- [13] B.J. Winer et.al. (1991): *Statistical Principles in Experimental Design*, S. 1028 ff bzw. S. 1024 ff)
- [14] Sawilowsky, S., Blair, R. C., & Higgins, J. J. (1989): *An investigation of the type I error and power properties of the rank transform procedure in factorial ANOVA*, Journal of Educational Statistics 14 (3): 255–267
- [15] J. Bredenkamp (1974): *Nonparametrische Prüfung von Wechselwirkungen*, Psychologische Beiträge 16, 398-416
- [16] Sawilowsky, S. (1990): *Nonparametric tests of interaction in experimental design*. Review of Educational Research 60: 91–126.
- [17] Larry E. Toothaker and De Newman (1994): *Nonparametric Competitors to the Two-Way ANOVA*, Journal of Educational and Behavioral Statistics, Vol. 19, No. 3, pp. 237-273

- [18] Stephen F. Olejnik and James Algina (1985): *A Review of Nonparametric Alternatives To Analysis of Covariance*, Evaluation Review 9: p 51-83
- [19] Wobbrock, J. O., Findlater, L., Gergle, D. & Higgins, J. (2011): *The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only ANOVA Procedures*, Computer Human Interaction - CHI , pp. 143-146
- [20] ARTool: <http://depts.washington.edu/aimgroup/proj/art/>
- [21] H. Mansouri, R. L. Paige & J. G. Surles (2004): *Aligned Rank Transform Techniques for Analysis of Variance and Multiple Comparisons*, Communications in Statistics - Theory and Methods, Volume 33, Issue 9
- [22] Michael G. Akritas, Steven F. Arnold & Edgar Brunner (1997): *Nonparametric Hypotheses and Rank Statistics for Unbalanced Factorial Designs*, Journal of the American Statistical Association, Volume 92, Issue 437 , pages 258-265
- [23] Ullrich Munzel, Edgar Brunner (2000): *Nonparametric methods in multivariate factorial designs*, Journal of Statistical Planning and Inference, Volume 88, Issue 1, Pages 117–132
- [24] Edgar Brunner, Ulrich Munzel and Madan L. Puri (1999): *Rank-Score Tests in Factorial Designs with Repeated Measures*, Journal of Multivariate Analysis 70, 286-317
- [25] Kimihiro Noguchi, Yulia R. Gel, Edgar Brunner, Frank Konietzschke (2012): *nparLD: An R Software Package for the Nonparametric Analysis of Longitudinal Data in Factorial Experiments*, Journal of Statistical Software , Volume 50, Issue 12.
- [26] Edgar Brunner, Ullrich Munzel (2013): *Nichtparametrische Datenanalyse, Unverbundene Stichproben*, Springer, 126 ff.
- [27] Cochran, W.G. (1950): *The comparison of percentages in matched samples*. Biometrika 3
- [28] Lunney, G.H. (1970): *Using Analysis of Variance with a dichotomous dependent variable: an empirical study*. Journal of Educational Measurement Volume 7, Issue 4
- [29] Bennett, B.M. (1968) *Rank-order tests of linear hypotheses*, J. of Stat . Society B 30: 483-489.
- [30] Shirley, E.A. (1981): *A distribution-free method for analysis of covariance based on ranked data*, Journal of Applied Statistics 30: 158-162.
- [31] James Scheirer, William S. Ray, Nathan Hare (1976): *The Analysis of Ranked Data Derived from Completely Randomized Factorial Designs*. Biometrics. 32(2). International Biometric Society, S. 429–434
- [32] Thomas P. Hettmansperger, Joseph W. McKean (2011): *Robust Nonparametric Statistical Methods*, CRC Press
- [33] John D. Kloke and Joseph W. McKean (2012): *Rfit : Rank-based estimation for linear models*,
http://journal.r-project.org/archive/2012-2/RJournal_2012-2_Kloke+McKean.pdf

- [34] Sonja Hahn, Frank Konietzschke and Luigi Salmaso (2013): *A comparison of efficient permutation tests for unbalanced ANOVA in two by two designs - and their behavior under heteroscedasticity*, arXiv.org Cornell University, <http://arxiv.org/pdf/1309.7781.pdf>
- [35] Michael Akritas and Edgar Brunner (2003): *Nonparametric Models for ANOVA and ANCOVA, a Review*. in “Recent Advances and Trends in Nonparametric Statistics” (Eds. M.G. Akritas and D.N. Politis), 79-91.
- [36] Iman, R.L. and Davenport, J.M. (1976): *New approximations to the exact distribution of the Kruskal-Wallis test statistic*, *Comm, Statist, A5*, pp 1335-1348
- [37] Stephen C. Hora and W. J. Conover (1984): *The F Statistic in the Two-Way Layout with Rank-Score Transformed Data*, *Journal of the American Statistical Association*, Vol. 79, No. 387, pp. 668-673
- [38] David J. Sheskin (2004): *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall
- [39] M.I. Salazar-Alvarez, V.G.Tercero-Gomez, M. Temblador-Pérez, A.E. Cordero-Franco, W.J. Conover (2014): *Nonparametric analysis of interactions: a review and gap analysis*, *Proceedings of the 2014 Industrial and Systems Engineering Research Conference*, Y. Guan and H. Liao (eds.)
- [40] T. Mark Beasley, Stephen Erickson, and David B. Allison (2009): *Rank-Based Inverse Normal Transformations are Increasingly Used, But are They Merited?* *Behavioural Genetics*, 39 (5), pp 380-395
- [41] M.L. Huang (2007): *A Quantile-Score Test for Experimental Design*, *Applied Mathematical Sciences*, Vol. 1, No 11, pp 507-516
- [42] http://en.wikipedia.org/wiki/Van_der_Waerden_test
- [43] [http://de.wikipedia.org/wiki/Friedman-Test_\(Statistik\)](http://de.wikipedia.org/wiki/Friedman-Test_(Statistik))
- [44] H. Mansouri and G. H. Chang (1995): *A Comparative Study of Some Rank Tests for Interaction*, *Computational Statistics and Data Analysis*, 19, 85-96
- [45] Rand R. Wilcox (2003): *Applying Contemporary Statistical Techniques*, Elsevier
- [46] Rand R. Wilcox (2012): *Introduction to Robust Estimation and Hypothesis Testing*, Elsevier
- [47] Rand R. Wilcox (2013): *New Statistical Procedures for the Social Sciences: Modern Solutions To Basic Problems*, Psychology Press, Lawrence Erlbaum Assoc
- [48] Mark Chatfield and Adrian Mander (2009): *The Skillings–Mack test*, *Stata Journal*, 9(2): pp 299–305.
- [49] Marc Hallin and Davy Paindaveine (2006): *Optimal Rank-Based Tests for Sphericity*, *The Annals of Statistics*, Vol. 34, No. 6, pp 2707–2756
- [50] Leonard A. Marascuilo & Maryellen McSweeney (1977): *Nonparametric and distribution-free methods for the social sciences*, Brooks/Cole Pub. Co.

- [51] Danbaba, Abubakar (2009): *A Study of Robustness of Validity and Efficiency of Rank Tests in AMMI and Two-Way ANOVA Tests*, Thesis, University of Ilorin (Nigeria)
- [52] András Vargha and Harold D. Delaney (1998): *The Kruskal-Wallis Test and Stochastic Homogeneity*, Journal of Education and Behavioral Statistics, vol. 23 no. 2, pp 170-192
- [53] T.Mark Beasley & Bruno D. Zumbo (2009): *Aligned Rank Tests for Interactions in Split-Plot Designs: Distributional Assumptions and Stochastic Heterogeneity*, Journal of Modern Applied Statistical Methods, Vol 8, NO. 1 , pp 16-50
- [54] Kathleen Peterson (2002): *Six Modifications Of The Aligned Rank Transform Test For Interaction*, Journal Of Modern Applied Statistical Methods Winter 2002, Vol. 1, No. 1, pp 100-109

Voraussetzungen zur parametrischen Varianzanalyse

- [61] P.K.Ito (1980): *Robustness of Anova and Manova Test Procedures* in Handbook of Statistics, Vol. 1, (P.R.Krishnaiah,ed.)
- [62] Jürgen Bortz (1984): *Statistik*, Springer Lehrbuch, Berlin
- [63] Andy Field (2009): *Discovering Statistics using SPSS*, Sage Publications, London
- [64] R.R.Wilcox (2005): *Introduction to robust estimation and hypothesis testing*, Burlington MA; Elsevier
- [65] G.V.Glass et al. (1972): *Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance*, Review of Educational Research, 42(3), pp 237-288
- [66] Mehmet Mendes, Soner Yiğit (2013): *Type I error and test power of different tests for testing interaction effects in factorial experiments*, Statistica Neerlandica, Vol 67 Issue 1, pp 1-26
- [67] Shuqiang Zhang (1998): *Fourteen Homogeneity of Variance Tests: When and how to use them*, Annual Meeting of the American Educational Research Association, San Diego
- [68] Jennifer J. Clinch and H. J. Keselman (1982): *Parametric Alternatives to the Analysis of Variance*, Journal of Educational Statistics, Vol. 7, No. 3, pp. 207-214
- [69] Lindman, H. R. (1974): *Analysis of variance in complex experimental designs*. San Francisco: W. H. Freeman & Co.
- [70] Jason W. Osborne (2008): *Best Practices in Quantitative Methods*, Sage Publications
- [71] <https://www.statsoft.com/Textbook/ANOVA-MANOVA>
- [72] 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, pp 290-302
- [73] <http://onlinestatbook.com/2/transformations/box-cox.html>

- [74] G.E.P. Box and S.L. Andersen (1955): *Permutation Theory in the Derivation of robust criteria and the study of departures from assumption*, Journal of the Royal Statistical Society, Series B, Vol XVII, No 1
- [75] Weihua Fan (2006): *Robust means modelling: An Alternative to Hypothesis Testing of Mean Equality in Between-subject Designs under Variance Heterogeneity and Nonnormality*, Dissertation, University of Maryland
<http://drum.lib.umd.edu/bitstream/1903/3786/1/umi-umd-3627.pdf>
- [76] G.E.P. Box (1953): *Non-normality and tests on variances*, Biometrika 40, pp. 318-335

Voraussetzungen zur parametrischen Varianzanalyse mit Messwiederholungen

- [81] Samuel Moulton: *Mauchly Test*, http://www.wjh.harvard.edu/~moulton/mauchly_test.pdf
- [82] Cornell, J. E., Young, D. M., Seaman, S. L., & Kirk, R. E. (1992). *Power comparisons of eight tests for sphericity in repeated measures designs*. Journal of Educational Statistics, 17, 233-249.
- [83] Rudolf N. Cardinal (2004): *ANOVA in practice, and complex ANOVA designs*, http://egret.psychol.cam.ac.uk/psychology/graduate/Guide_to_ANOVA.pdf
- [84] T.Mark Beasley (2002): *Multivariate Aligned Rank Test for Interactions in multiple Group repeated Measures Design*, Multivariate Behavioral Research, 37 (2), 197-226
- [85] Jennifer Joanne Bryan (2009): *Rank transforms and tests of interaction for repeated measures experiments with various covariance structures*, Oklahoma State University, Dissertation
- [86] http://www.ats.ucla.edu/stat/r/seminars/Repeated_Measures/repeated_measures.htm
- [87] https://en.wikipedia.org/wiki/Fishers_method

Versuchspläne

- [87] H. Huynh (1978): *Some approximate tests for repeated measurement designs*, Psychometrika 43, 161-175
- [88] Rhonda K. Kowalchuk, H. J. Keselman & James Algina (2003): *Repeated Measures Interaction Test with Aligned Ranks*, Multivariate Behavioral Research, Volume 38, Issue 4

Logistische Regression

- [91] Ralph B. D'Agostino (1971): *A Second Look at Analysis of Variance on Dichotomous Data*, Journal of Educational Measurement, Vol. 8, No. 4, pp. 327-333
- [92] Paul D. Cleary and Ronald Angel (1984): *The Analysis of Relationships Involving Dichotomous Dependent Variables*, Journal of Health and Social Behavior, 25, pp. 334-348.
- [93] Matt Bogard (2011): *Linear Regression and Analysis of Variance with a Binary Dependent Variable*, <http://econometricsense.blogspot.de/2011/08/linear-regression-and-analysis-of.html>

- [94] Dawson, Robert J. MacG. (1995): *The 'Unusual Episode' Data Revisited*, Journal of Statistics Education, 3
- [95] Koch, G.G., Landis, J.R., Freeman, J.L., Freeman, D.H., and Lehnen, R.G. (1977): *A general methodology for the analysis of experiments with repeated measurement of categorical data*. Biometrics, 33, 133-158.
- [96] Rainer Diaz-Bone & Harald Künemund (2003): *Einführung in die binäre logistische Regression*, Freie Universität Berlin, Mitteilungen aus dem Schwerpunktbereich Methodenlehre, Heft Nr. 56
<http://www.rainer-diaz-bone.de/Logreg.pdf>
- [97] http://en.wikipedia.org/wiki/Logistic_regression

Kontraste und Kodierungen

- [101] R Library: Contrast Coding Systems for categorical variables:
http://www.ats.ucla.edu/stat/r/library/contrast_coding.htm
- [102] *Regression with SPSS: Chapter 5: Additional coding systems for categorical variables in regression analysis* :
<http://www.ats.ucla.edu/stat/spss/webbooks/reg/chapter5/spsreg5.htm>
- [103] Richard Gonzalez: *Contrasts and Post Hoc tests (Lecture Notes)* (2009), University of Michigan, Ann Arbor, <http://www-personal.umich.edu/~gonzo/coursenotes/file3.pdf>

Besonderheiten bei Varianzanalysen mit R

- [111] Falk Scholer: *ANOVA (and R)*, <http://goanna.cs.rmit.edu.au/~fscholer/anova.php>
- [112] Bertolt Meyer: *Obtaining the same ANOVA results in R as in SPSS - the difficulties with Type II and Type III sums of squares* ,
<http://myowelt.blogspot.de/2008/05/obtaining-same-anova-results-in-r-as-in.html>