

Schätzung von linearen Kontrasten in der Varianzanalyse mit SAS

Benjamin Mayer
Institut für Epidemiologie und
Medizinische Biometrie
Universität Ulm
Schwabstr. 13
89075 Ulm
benjamin.mayer@uni-ulm.de

Mareike Pfau
Institut für Epidemiologie und
Medizinische Biometrie
Universität Ulm
Schwabstr. 13
89075 Ulm
mareike.pfau@uni-ulm.de

Zusammenfassung

In der medizinischen Forschung müssen häufig Fragestellungen beantwortet werden, die den Vergleich verschiedener Gruppen beinhalten (z.B. Patientengruppen mit verschiedenen Expositionen, verschiedenen Behandlungsmethoden, etc.). Sollen dabei mehr als zwei Behandlungsgruppen gleichzeitig miteinander verglichen werden, so stellt unter der Voraussetzung bestimmter Kriterien an die zu analysierenden Daten die so genannte Varianzanalyse (ANOVA) die Auswertungsmethode der Wahl dar. Durch die Varianzanalyse kann der Einfluss einer oder auch mehrerer Faktoren auf eine stetige Zielgröße überprüft werden. In SAS kann die Varianzanalyse als lineares Modell betrachtet und umgesetzt werden. Die Einflussgrößen werden dann wie bei einer Regression modelliert und der Einfluss durch Schätzung der Parameter und anschließendem F-Test untersucht. Bei der klassischen ANOVA sagt das Ergebnis jedoch nur aus, ob es einen Unterschied zwischen den Gruppen gibt, man erfährt jedoch nicht, zwischen welchen Gruppen der Unterschied vorliegt. Hierfür bietet sich dann eine so genannte Kontrastberechnung an, mit deren Hilfe präzisere Alternativhypothesen getestet werden können. Dieser Beitrag soll die konkrete Verwendung von Kontrasten mit den in SAS hierfür vorgesehen Optionen vorstellen. Hierzu werden anhand eines Beispieldatensatzes aus dem Bereich Notfallmedizin die Möglichkeiten einer Kontrastberechnung in den SAS-Prozeduren MIXED, GLM und GLIMMIX vorgestellt. Grundlage für die Analysen bilden dabei reale Fragestellungen. Zudem sollen dabei die Kombination der entsprechenden CONTRAST-Statements mit ggf. notwendigen, weiteren Statements (z.B. RANDOM oder REPEATED) aufgezeigt werden. Im Zentrum steht dabei die Frage, welche Codierung konkret für die Definition bestimmter Kontraste verwendet werden muss, um die interessierenden Fragestellungen entsprechen beantworten zu können. In der praktischen Umsetzung zeigt sich, dass die gestellten Anforderungen an einen Kontrast nicht sehr hoch und auch leicht umsetzbar sind. Im Wesentlichen müssen die gewählten Kontrastkoeffizienten in Summe sich zu Null addieren. Insbesondere ist dabei festzuhalten, dass die Spezifizierung der Kontraste in allen verwendeten SAS-Prozeduren in der gleichen Weise vorzunehmen war. Eventuelle zusätzliche methodische Maßnahmen, die sich aufgrund der Kontrastberechnung in der Praxis ggf. ergeben, wie z.B. der Umgang mit multiplen Hypothesentests, müssen jedoch gesondert im Rahmen der Analyse berücksichtigt werden.

Schlüsselwörter: PROC ANOVA, linearer Kontrast, contrast, lineares Modell

1 Einleitung

Im Rahmen der Auswertung medizinische Forschungsdaten sollen üblicherweise verschiedene Subpopulationen des betreffenden Datensatzes miteinander verglichen werden. Dies können beispielsweise Subgruppen mit verschiedenen Expositionen sein (Raucher vs. Nichtraucher, schadstoffbelastet vs. gesund, usw.), die im Hinblick auf das Auftreten einer Krankheit untersucht werden, oder aber verschiedene Behandlungsmethoden sollen bezüglich ihrer Wirksamkeit verglichen werden. Von Interesse kann auch die Entwicklung eines medizinisch relevanten Parameters über die Zeit sein (z.B. Blutdruck). Sollen nur zwei Gruppen miteinander verglichen werden, so kann unter der Voraussetzung normalverteilter Daten beispielsweise der t-Test für einen solchen Vergleich genutzt werden. In der Praxis gibt es oftmals jedoch mehr als nur zwei Gruppen, z.B. beim Vergleich von vier Dosenstufen eines Medikaments oder von drei verschiedene Medikamenten. In diesen Fällen kommen varianzanalytische Ansätze wie die ANOVA (*engl. analysis of variance*) zum Einsatz. Unter der Verwendung einer ANOVA kann der Einfluss eines oder mehrerer Faktoren auf eine Zielgröße überprüft werden, und ebenso wenn man an der Wirksamkeit einer Behandlung interessiert ist und zudem erfahren möchte, ob das Geschlecht einen Effekt hat auf die zuvor definierte Zielgröße. Die ANOVA kann also sowohl in einem univariaten, als auch in einem multiplen Setting genutzt werden. Im letzteren Fall kann man zusätzlich auch Wechselwirkungen untersuchen, die zwischen den Faktoren auftreten können. Die Ausprägungen, die ein Faktor annehmen kann, werden dabei als Stufen bezeichnet.

Das Ziel der Varianzanalyse ist es herauszufinden, ob die Varianz der Zielgröße auf die verschiedenen Ausprägungen der Einflussgrößen zurückzuführen ist oder durch zufällige Schwankungen zu erklären ist. Dazu vergleicht die Varianzanalyse die Varianz zwischen den Gruppen mit der Varianz in den Gruppen. Es wird die Nullhypothese H_0 : „Es existiert kein Unterschied zwischen den Gruppen“ gegen die Alternativhypothese H_1 : „Es existieren mindestens zwei Gruppen, die sich unterscheiden“ getestet. Die typische in der Literatur verwendete Umsetzung der ANOVA erfolgt durch sogenannte mittlere Quadrate (mean squares). Dazu wird die gesamte Variabilität, also die totale Quadratsumme, in zwei Quadratsummen aufgeteilt, auf deren Basis dann die F-verteilte Teststatistik konstruiert wird.

Die Hypothesenprüfung der ANOVA sagt „nur“ aus, ob es einen Unterschied zwischen den Gruppen gibt, man erfährt jedoch nicht, auf welche Gruppen dieser Unterschied zurückzuführen ist. Man bekommt zudem keine Auskunft über die Richtung des Unterschiedes. Mit Hilfe der Berechnung so genannter linearer Kontraste besteht jedoch die Möglichkeit präzisere Hypothesen testen zu können.

In diesem Beitrag soll eine Übersicht gegeben werden, welche Möglichkeiten die gängigen Prozeduren in SAS zur Umsetzung einer Varianzanalyse inklusive der Berechnung linearer Kontraste verfügbar sind. Die Grundidee der Varianzanalyse soll dabei kurz für den ein- und zweifaktoriellen Fall beschrieben werden. Zudem wird auf den Spezialfall eines Messwiederholungsdesigns eingegangen. Die Berechnung von linearen Kontrasten in SAS wird anhand eines realen Beispieldatensatzes vorgestellt.

2 Material und Methoden

2.1 Methodik der Varianzanalyse

2.1.1 Notation

In diesem Kapitel wird nachfolgende Notation verwendet:

Y_{ij}	Zielvariable
μ	durchschnittlicher Effekt
α_i	Veränderung bzgl. Ausprägung i der Einflussgröße A
ε_{ij}	Fehlerterm
n_i	Anzahl Beobachtungen mit Ausprägung i
n	Anzahl Beobachtungen

Die Einflussgrößen seien immer qualitativ, während die Zielvariable quantitativ und stetig sei. Die Einflussfaktoren können generell fest oder zufällig sein, auch eine Kombination aus festen und zufälligen Faktoren ist möglich. Feste Faktoren sind solche, die $a \geq 2$ fest vorgegebene Stufen haben und bei Wiederholung einer Studie wieder mit denselben Stufen des Faktors auftreten würden. Beispiele sind Geschlecht oder Behandlung (Medikament). Zufällige Faktoren besitzen hingegen nicht systematisch festgelegte oder vorgegebene Stufen [1]. Es werden nur einige Stufen aus einer Grundgesamtheit von möglichen Stufen verwendet, um Rückschlüsse für alle Stufen zu treffen. Typische Beispiele sind Versuchsperson oder Versuchstier. Die Varianzanalyse setzt voraus, dass die abhängige Variable für jede Gruppe normalverteilt ist. Eine weitere wichtige Voraussetzung ist die Homoskedastizität der Varianzen der abhängigen Variablen, d.h. die Varianz in den Gruppen sollte etwa gleich groß sein. Außerdem müssen die Messungen unabhängig sein. Für große Fallzahlen ist die Varianzanalyse robust gegenüber Verletzungen der Normalverteilungsannahme. Auch leichte Verletzungen der Varianzhomogenität sind bei großer Fallzahl und gleich großen Gruppen unproblematisch [2].

2.1.2 Einfaktorielle Varianzanalyse

Bei der einfaktoriellen Varianzanalyse wird der Einfluss eines Faktors A auf die Zielvariable untersucht. Die verschiedenen Ausprägungen a können z.B. verschiedene Dosisstufen eines Medikaments oder verschiedene Behandlungsmethoden sein. Der einfaktoriellen Varianzanalyse wird das folgende statistische Modell zugrunde gelegt:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \text{ mit } 1 \leq i \leq a, 1 \leq j \leq n_i, n = \sum_{i=1}^a n_i$$

mit $\varepsilon_{ij} \sim N(0, \sigma^2)$ unabhängig für alle i, j . Hierbei wird die Omnibushypothese

$$H_0: \alpha_1 = \dots = \alpha_a = 0$$

getestet und für die Berechnung der zugehörigen Teststatistik die folgenden Quadratsummen verwendet:

$$QST = \sum_{i=1}^a \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2 \quad (\text{Totale Quadratsumme})$$

$$QSA = \sum_{i=1}^a n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 \quad (\text{Quadratsumme bzgl. Effektgröße})$$

$$QSE = \sum_{i=1}^a \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2 \quad (\text{Fehler-Quadratsumme})$$

Es gilt der Zusammenhang $QST = QSA + QSE$. Wird die Varianzanalyse jedoch als lineare Modell aufgefasst, ist die Berechnung der Quadratsummen nicht zwingend erforderlich [3].

2.1.3 Zwei-/Mehrfaktorielle Varianzanalyse

Soll der Einfluss mehrerer Variablen und deren Ausprägungen auf die Zielvariable gleichzeitig getestet werden, so wird eine mehrfaktorielle Varianzanalyse verwendet. Dabei können zusätzlich zu den Haupteffekten der Faktoren auch die Wechselwirkungen zwischen den Einflussgrößen mit in das Modell aufgenommen werden. Der zweifaktoriellen Varianzanalyse wird folgendes Modell zugrunde gelegt:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

mit

$$1 \leq i \leq a, 1 \leq j \leq b, 1 \leq k \leq n_0, n = a \cdot b \cdot n_0$$

und $\varepsilon_{ij} \sim N(0, \sigma^2)$ unabhängig für alle i, j, k . Hierbei charakterisiert der Modellparameter γ_{ij} die Wechselwirkung zwischen den Stufen i und j der Einflussfaktoren A und B . Analog zur einfaktoriellen Varianzanalyse können die Quadratsummen berechnet werden, die sich wie folgt ergeben:

	Quadratsumme
A	$b \cdot n_0 \sum_{i=1}^a (\bar{Y}_{i..} - \bar{Y}_{...})^2$
B	$a \cdot n_0 \sum_{j=1}^b (\bar{Y}_{.j.} - \bar{Y}_{...})^2$
A*B	$n_0 \sum_{i=1}^a \sum_{j=1}^b (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$
Fehler	$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_0} (Y_{ijk} - \bar{Y}_{ij.})^2$
Total	$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_0} (Y_{ijk} - \bar{Y}_{...})^2$

Bei der zweifaktoriellen Varianzanalyse werden nunmehr unter der Annahme des obigen Modells drei Omnibushypothesen getestet.

Die mehrfaktorielle Varianzanalyse stellt eine analoge Erweiterung der zweifaktoriellen Varianzanalyse dar. Der Unterschied besteht darin, dass zu den Haupteffekten jetzt nicht nur Wechselwirkungen zwischen zwei Einflussgrößen mit in das Modell aufgenommen werden (können), sondern von mehreren. Die inhaltliche Interpretation komplexerer Wechselwirkungen unter Beteiligung von mehr als 2 Faktoren stellt jedoch ein schwieriges Unterfangen dar.

2.1.4 Varianzanalyse mit Messwiederholungen

Sollte ein Analysedatensatz Messwiederholungen enthalten, so sind die vorliegenden Daten nicht mehr voneinander unabhängig und dadurch die Voraussetzungen der vorgestellten Modelle nicht mehr erfüllt. In diesen Fällen muss eine zusätzliche Voraussetzung erfüllt sein, die Sphärizität genannt wird und immer dann überprüft werden muss, wenn mehr als zwei Stufen vorliegen. Ist die Sphärizitätsbedingung erfüllt, so sind die Varianzen der Differenzen aller Stufen der unabhängigen Variablen gleich [4].

2.2 Lineare Kontraste

Die Varianzanalyse gibt zunächst nur Auskunft darüber, ob es einen Unterschied zwischen den einzelnen Faktorstufen gibt oder nicht. Oft ist man jedoch daran interessiert, welche Gruppen sich im Speziellen unterscheiden. Eine Möglichkeit zur Testung zielgerichteter Hypothesen ist die Berechnung linearer Kontraste. Ein Kontrast dient als Gewichtung der Mittelwerte. Es können dabei verschiedene Stufen oder Kombinationen von Stufen miteinander verglichen werden. Mit $\mu_i = \mu + a_i$ als Erwartungswert von Stufe i und $\lambda = (\lambda_1, \dots, \lambda_a)$ als ein Vektor bekannter Konstanten heißt die Linearkombination

$$c = \sum_{i=1}^a \lambda_i \mu_i$$

ein linearer Kontrast, falls die Summe der λ_i 's gleich Null ist. Wichtig für die Praxis ist, dass die absolute Größe der λ_i für das Testen eines Kontrastes nicht von Bedeutung ist, sondern allein das Verhältnis zueinander [5].

2.3 Umsetzung in SAS

Im Folgenden werden verschiedene SAS-Prozeduren vorgestellt und erklärt, wie mit diesen Varianzanalysemodellen bzw. lineare Modelle in SAS angepasst werden können. Dabei wird auch insbesondere erklärt, wie lineare Kontraste in diesen Prozeduren verwendet werden können.

2.3.1 PROC MIXED

Mit der Prozedur PROC MIXED können gemischt lineare Modelle angepasst werden, die sowohl feste als auch zufällige Einflussgrößen beinhalten. Die festen Variablen und entsprechende Wechselwirkungsterme werden im MODEL-Statement übergeben, während die zufälligen Variablen ggf. im RANDOM-Statement gelistet sind.

Zusätzlich können über CONTRAST-Statements die entsprechenden linearen Kontraste angefordert werden, die für die Analyse von Interesse sind. Das Statement besteht aus einem Label für den Kontrast, der Angabe der Einflussgröße, deren Mittelwerte gewichtet werden sollen, und den Kontrastkoeffizienten. Feste Variablen werden vor dem „|“ angegeben, zufällige dahinter. Es können bis zu $a-1$ linear unabhängige Kontraste simultan getestet werden. Diese müssen dann durch „;“ voneinander getrennt werden. Mit der Option „/E“ kann in der Ausgabe die Matrix aller Kontrastkoeffizienten ausgegeben werden. Der konkrete Schätzwert eines Kontrastes lässt sich über das ESTIMATE-Statement ausgeben. Hierbei wird dieselbe Syntax zugrunde gelegt wie im CONTRAST-Statement.

```
PROC MIXED DATA=daten;  
  CLASS fix_var1 zuf_var1;  
  MODEL abh_var= fix_var1 ;  
  RANDOM zuf_var1 fix_var1*zuf_var1;  
  CONTRAST 'Name des Kontrasts'  
    fix_var werte | zuf_var werte /E;  
  CONTRAST 'Simultane Kontraste (2-dim)'  
    fix_var werte | zuf_var werte,  
    fix_var werte | zuf_var werte;  
RUN;
```

Abbildung 1: Programmcode für Basismodell in PROC MIXED

2.3.2 PROC GLM

Eine Alternative zu PROC MIXED ist die Verwendung der Prozedur PROC GLM, deren Syntax etwas anders ist. Ein wesentlicher Unterschied besteht im Umgang mit zufälligen Einflussfaktoren, diese müssen wie die festen Einflussgrößen im MODEL-Statement aufgeführt und zusätzlich im RANDOM-Statement aufgelistet werden. Mit den Optionen „/ss1“, „/ss2“, „/ss3“ und „/ss4“ kann im MODEL-Statement ausgewählt werden welche mittleren Quadratsummen ausgegeben werden sollen. Hier wird also gewählt, ob die Quadratsummen vom Typ I, II, III oder IV berechnet werden sollen. Es können auch alle vier ausgegeben werden. Im Fall balancierter Daten sind diese alle gleich.

Für die Berechnung linearer Kontraste benötigt man auch bei PROC GLM das CONTRAST-Statement, das in gleicher Weise wie bei PROC MIXED zu verwenden ist.

```

PROC GLM DATA=daten;
    CLASS fix_var1 zuf_var1;
    MODEL abh_var= fix_var1 zuf_var1 fix_var1*zuf_var1;
    CONTRAST 'Name des Kontrasts' fix_var1 werte;
    RANDOM zuf_var1 fix_var1*zuf_var1 /TEST;
RUN;

```

Abbildung 2: Programmcode für Basismodell in PROC GLM

2.3.3 PROC GLIMMIX

Ist die Zielvariable nicht stetig, sondern binär, so kann ein Modell mit PROC GLIMMIX angepasst werden. Die Standardaufrufe wie CLASS, RANDOM funktionieren analog zu PROC MIXED. Beim MODEL-Statement kann man zwischen zwei Varianten wählen. Nach dem „=“ stehen wie bei PROC MIXED die fixen Effektgrößen. Vor dem „=“ kann entweder die response-Variable angegeben werden oder zwei Variablen events/trials. Dabei wird letzterer Syntax nur für binomialverteilte Daten verwendet. Auch bei PROC GLIMMIX können durch das CONTRAST-Statement spezifische Kontraste getestet werden. Der Aufruf erfolgt analog zu PROC GLM und PROC MIXED.

```

PROC GLIMMIX DATA=daten;
    CLASS fix_var1 zuf_var1;
    MODEL abh_var_en= fix_var1;
    CONTRAST 'Name des Kontrasts' fix_var1 werte;
    RANDOM zuf_var1 fix_var1*zuf_var1 /TEST;
RUN;

```

Abbildung 3: Programmcode für Basismodell in PROC GLIMMIX

3 Ergebnisse eines Anwendungsbeispiels

Anhand eines Anwendungsbeispiels soll nun der Einsatz der vorgestellten Ansätze in SAS zur Berechnung interessierender Kontraste veranschaulicht werden. Die Auswertungen erfolgten mit SAS 9.4.

3.1 Beispieldatensatz „Notfallmedizin“

Die Grundlage für die verwendeten Daten war eine experimentelle Studie zur Durchführung einer Koniotomie. An der Studie haben 48 Ärzte mit teilgenommen, die aufgrund ihres Erfahrungsgrades zu gleichen Anteilen in zwei Gruppen eingeteilt wurden (erfahren vs. unerfahren). Im Rahmen der Studie wurden an Puppenmodellen zwei verschiedene Koniotomiesets (Set 1, Set 2) an vier verschiedenen Halskonfigurationen (kurz-dick (kd), kurz-schlank (ks), lang-dick (ld) und lang-schlank (ls)) untersucht. Jeder Arzt hat dabei jedes Set an jeder Halskonfiguration verwendet. Es liegen damit insgesamt

384 Messungen vor. Es wurde die für den Eingriff benötigte Zeit gemessen und u.a. erfasst, ob Komplikationen aufgetreten sind. Als Hauptfragestellungen der Studie sollte u.a. untersucht werden, ob die benötigte Zeit für den Eingriff und die Fehlerrate zwischen den Sets unterschiedlich ist innerhalb der Subgruppen der Halsanatomien bzw. Erfahrungsgrad der Ärzte.

Für die Analyse des stetigen Endpunktes der benötigten Zeit wurde das folgende Basismodell in PROC MIXED verwendet:

```
PROC MIXED DATA=osma;
  CLASS id set konfiguration erfahrung;
  MODEL time = set konfiguration erfahrung
    set*konfiguration set*erfahrung /SOLUTION;
  REPEATED /SUBJECT=id;
RUN;
```

Abbildung 4: Basismodell in PROC MIXED zur Kontrastberechnung

Des Weiteren sind in Abbildung 5 beispielhaft die entsprechenden CONTRAST-Statements zu sehen, wie sie für die Berechnung der Gruppenvergleiche „dicker vs. schlanker Hals“ und „Set 1 vs. Set 2 in erfahrenen und unerfahrenen Ärzten“ verwendet wurden:

```
CONTRAST 'Konfiguration dick vs schlank'
  konfiguration -1 1 -1 1;

CONTRAST 'Set1 vs Set2 x erfahrung'
  erfahrung*set -1 1 1 -1;
```

Abbildung 5: Spezifische CONTRAST-Statements in PROC MIXED

Insgesamt zeigten sich deutliche Unterschiede bezüglich dem verwendeten Set ($p < 0.0001$), der Halsanatomie ($p = 0.0005$), sowie der Erfahrung des Arztes ($p < 0.0001$). Insbesondere war auch die Wechselwirkung von Set und Erfahrung signifikant ($p = 0.0210$). Dies spiegelte sich u.a. dadurch wieder, dass der durchschnittliche Unterschied zwischen Set 1 und Set 2 bei unerfahrenen Ärzten doppelt so hoch war, als in erfahrenen Ärzten. Dieses Erkenntnis folgte unmittelbar aus den entsprechend berechneten Kontrasten.

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
Set1 vs Set2 erfahren	24.6656	7.1570	46	3.45	0.0012
Set1 vs Set2 unerfahren	48.8479	7.1570	46	6.83	<.0001

Abbildung 6: Spezifische Gruppenvergleiche mittels ESTIMATE-Statement

Das entsprechende Basismodell zur Erfassung von Effekten der o.g. Einflussgrößen auf die Wahrscheinlichkeit einer Komplikation wurde modelliert via PROC GLIMMIX:

```
PROC GLIMMIX DATA=osma;
  CLASS konfiguration erfahrung set;
  MODEL komplikation(EVENT='1')= konfiguration
    erfahrung set konfiguration*set erfahrung*set
    /DIST=BINARY;
RUN;
```

Abbildung 7: Basismodell in PROC GLIMMIX zur Kontrastberechnung

Wie den Ergebnissen zu entnehmen ist, gibt es einige signifikante Gruppenunterschiede, jedoch besteht kein konstanter Effekt zwischen den beiden Sets in allen betrachteten Subgruppen.

Estimates						
Label	Estimate	Standard Error	DF	t Value	Pr > t	Exponentiated Estimate
Set1 vs Set2	-0.4841	0.2098	374	-2.31	0.0216	0.6162
Set1 vs Set2 erfahren	-0.7160	0.3005	374	-2.38	0.0177	0.4887
Set1 vs Set2 unerfahren	-0.2523	0.2927	374	-0.86	0.3893	0.7770
Set1 vs Set2 kd	-0.9889	0.4319	374	-2.29	0.0226	0.3720
Set1 vs Set2 ks	0.08956	0.4129	374	0.22	0.8284	1.0937
Set1 vs Set2 ld	-0.6970	0.4203	374	-1.66	0.0981	0.4981
Set1 vs Set2 ls	-0.3402	0.4124	374	-0.83	0.4099	0.7116

Abbildung 8: Spezifische Gruppenvergleiche mit Angabe des Odds Ratios

4 Diskussion

In der medizinischen Forschung stellen varianzanalytische Konzepte ein gängiges Auswertungsverfahren dar. Die Berechnung von linearen Kontrasten ist hierbei für die Beantwortung inhaltlicher Fragestellungen enorm wichtig, da die Resultate einer ANOVA generell keine Aussage über spezifische Gruppenvergleiche machen kann. Sollen die Kontrastthesen konfirmatorisch interpretiert werden, so müssen die Kontrastgewichte, deren absoluter Wert keinen Einfluss auf die Berechnung hat, bereits vor der Datenanalyse festgelegt werden. Ein signifikanter Kontrasteffekt kann dann als Bestätigung der zugrunde liegenden Forschungshypothese angesehen werden, während a-posteriori-Kontraste nur einen Hypothesen-generierenden Charakter haben. Einzig wesentliche Voraussetzung an die Kontrastkoeffizienten ist, dass deren Summe gleich Null ist. Zudem muss genau darauf geachtet werden, dass die Koeffizienten in der richtigen Reihenfolge angegeben werden in den entsprechenden CONSTAST-Statements. Hier

bietet jedoch die Class Level Information-Tabelle im Standard-SAS-Output eine geeignete Hilfe, da man dieser Tabelle entnehmen kann, in welcher Reihenfolge die Stufen der Einflussfaktoren angeordnet sind. In SAS können keine Brüche als Kontrastkoeffizienten eingegeben werden. Im CONTRAST-Statement sollte man daher darauf achten, die Dezimalzahl mit genügend Nachkommastellen anzugeben. Ansonsten kann es passieren, dass SAS den Kontrast als „not estimable“ bezeichnet. Wie in diesem Beitrag gezeigt wurde, folgt die Angabe der CONTRAST-Statements über verschiedene SAS-Prozeduren hinweg demselben Muster.

Literatur

- [1] W. Gaus und R. Muche. Medizinische Statistik: Angewandte Biometrie für Ärzte und Gesundheitsberufe. Schattauer, 2014. – ISBN 9783794529315
- [2] W.J. Kennedy und M. G. Marasinghe. SAS for Data Analysis: Intermediate Statistical Methods. Springer, 2008. – ISBN 9780387773711
- [3] L. Fahrmeir, I. Pigeot, R. Künstler und G. Tutz. Statistik: Der Weg zur Datenanalyse. 7. Auflage. Springer, 2011. – ISBN 9783642019388
- [4] J. Bortz und C. Schuster. Statistik für Human- und Sozialwissenschaftler. 7., vollständig überarbeitete und erweiterte Auflage. Springer, 2010. – ISBN9783642127694
- [5] W. Hall und K.D. Bird. Simultaneous Multiple Comparison Procedures in Psychiatric Research. In: Australian and New Zealand Journal of Psychiatry 20 (1986), S. 46–54