Probleme varianzanalytischer Auswertungen und deren praktische Lösung mit SPSS

André Hahn (1995)[1]

1 Einleitung

Dieses Papier beschäftigt sich mit technischen Problemen varianzanalytischer Auswertungsdesigns, die in der Praxis leider nur allzu häufig anzutreffen sind und dabei heftigst Kopfzerbrechen verursachen. So ist die Berechnung einer mehrfaktoriellen und multivariaten Varianzanalyse, die mehrere Gruppierungs- als auch Meßwiederholungsfaktoren enthalten mag, nur der Anfang und nicht das Ende aller Bemühungen. (Fast hätte ich „...Ende allen Übels“ geschrieben). Meist steht man nach einer mehr oder minder komplexen Gesamtanalyse (eines sogenannten „full models“) vor einer Reihe von Problemen. Etwa, wie gehe ich mit ungleichen Zellfrequenzen (unterschiedlich viele Personen in einer Zelle des Versuchsplans, z.B. wegen unerwarteter Missings) um? Oder, wie prüfe ich „einfache Haupteffekte“ für den schrecklichen Fall, daß ein Interaktionsterm der Gesamtanalyse statistisch bedeutsam ist? Anders: Was kann ich überhaupt mit einer Haupteffektanalyse an zusätzlichen Informationen gewinnen und worauf ist bei der Durchführung unbedingt zu achten? Um ein letztes und gleichzeitig wohl auch sehr häufiges Problem anzusprechen: Wie kann ich die Mittelwerte meines x-beliebigen Designs auf Verschiedenheit testen („multiple Mittelwertvergleiche“).
Jedes der angesprochenen Probleme ist für sich genommen (im Rahmen einfacher Designs) recht leicht zu lösen. So behilft man sich vielleicht im Falle einer signifikanten einfaktoriellen Varianzanalyse mit 5 Faktorstufen mit mehreren (un)abhängigen t-Tests, um zu ermitteln, welche der Faktorstufen sich signifikant voneinander unterscheiden. Oder man bestimmt einfache Haupteffekte, indem man einen bestimmten Teil der Untersuchungsstichprobe bei einer weiteren Analyse ausschließt (per „SELECT IF (gruppe=1)“-Befehl).
Tatsächlich findet man im Ergebnisteil sehr vieler empirischer Arbeiten eine noch elegantere Lösung: Das Problem wird einfach ignoriert. Entsprechende Statistiken werden nicht berichtet, „Subanalysen“ werden erst gar nicht durchgeführt. Urteile über Unterschiede zwischen Mittelwerten werden aufgrund subjektiver Eindrücke über das numerische Ausmaß eines Unterschieds getroffen (etwa per graphischer Inspektion). Das dies so ist, ist mit Sicherheit nicht auf die Böswilligkeit, Faulheit oder gar Unkenntnis der jeweiligen Autoren zurückzuführen. Trotzdem ist diese Situation natürlich sehr unbefriedigend, zumal sich für all diese Probleme schon in den Statistiklehrbüchern Lösungen finden lassen. Problemlösungen wie beispielsweise, daß man doch in diesem Falle am besten einen Scheffé-Test rechnet, sind meist hinlänglich bekannt.
Es ist möglich, daß fast immer ein rein technisches Problem, die Berichtlegung angemessener Statistiken verhindert. So unterstützt das Statistikprogramm XY vielleicht das gewünschte Verfahren nicht oder aber die entsprechende Kommandosyntax des Verfahrens ist ungenau, unvollständig oder gar unverständlich dokumentiert. Letzteres ist wohl der häufigste Fall und auf die ungeheure Mächtigkeit der rechnergestützten Analysesysteme zurückzuführen. Allein die Standardhandbücher zum SPSS-System in der Version 6.0 umfassen über 2.000 Seiten. Diese Handbücher sind von hoher Qualität (die Verfahren werden auch inhaltlich vorgestellt,

[1] Die vorliegenden Seiten verstehen sich als Entwurf und sind NICHT zitierfähig. Zudem übernehme ich keine Garantie für die Korrektheit der getroffenen Aussagen, da ich selbst auf die Annahme angeweisen bin, daß die in der einschlägigen Literatur getroffenen Aussagen korrekt sind. Die Urheber der jeweiligen Behauptungen - soweit sie nicht vom Autor stammen - werden durch entsprechende Referenzen „gebranntmarkt“.

allerdings unter (offensichtlich angenehmen) Verzicht auf mathematische Grundlagen) und viele Sozialwissenschaftler orientieren sich deshalb in ihrer praktischen Arbeit an diesen Büchern. Damit tragen die Autoren dieser Bücher eine Verantwortung, die ihnen höchst wahrscheinlich gar nicht bewußt ist. So etwa bei der Behandlung statistischer Detailprobleme, die dort bisweilen behandelt werden (z.B. Voraussetzungsprüfungen). Die Verantwortung liegt hier bei der Auswahl dargestellter Auswertungsprobleme. Problem A wird behandelt, Problem B eben nicht. Offensichtlich hat man Wert darauf gelegt, solche Problemfälle darzustellen, die bei annähernd jeder Varianzanalyse auftreten. Das eigentliche Defizit der Handbücher (wenn sie denn als Lehrbuchersatz herhalten) ist ihre Systematik. Selten finden sich Darstellungen, die der immanenten Logik des Analyseverfahrens folgen. Problemlöseschritte wie sie häufig bei sozialwissenschaftlichen Fragestellungen abgearbeitet werden, wären dabei wohl die idealste Form der Darstellung. Einige wenige Statistikhandbücher gehen auch diesen Weg und sind wohl auch deshalb entsprechend beliebt (Backhaus, et al., 1994; Tabachnik & Fidell, 1993). Im Falle der Varianzanalyse wäre eine mögliche sinnvolle Systematik: (a) Gesamtanalyse, (b) hierarchisch aufgebaute Haupteffektanalysen, (c) Mittelwertvergleiche und schließlich (d) Voraussetzungsanalysen [deren Befunde dann die Arbeitsschritte (a) bis (c) zunichte machen].

Um es vorweg zu nehmen: Der vorliegende Text folgt auch *nicht* dieser „idealen“ Systematik. Diese Systematik beschreibt ja auch lediglich die Reihenfolge der Auswertungsschritte, wie sie sich bei der konkreten Analyse an einem Datensatz in der Regel darstellt. Diese Systematik hätte den Nachteil, daß vielleicht hunderte verschiedene varianzanalytische Designs dargestellt werden müßten, um auch jeden möglichen Aufwertungsfall zu berücksichtigen. Das wiederum ist weder möglich, noch erscheint dieses Vorgehen sonderlich sinnvoll zu sein.

Die folgende Darstellung ist vielmehr eine Art „Baukasten“. Die Teile dieses Baukastens lassen sich dann miteinander kombinieren, sodaß man letztlich selbst in die Lage versetzt wird, die einzelnen Arbeitsschritte einer Varianzanalyse (unabhängig vom konkreten Design) zu planen, durchzuführen und die jeweiligen „Outputs“ zu interpretieren. Die Darstellung der Bausteine beginnt mit „Kontrasten“. Dieser erste Baustein behandelt die Frage, auf welche Weise sich die Mittelwerte zweier oder mehrerer Faktorenstufen eines Faktors auf Verschiedenheit prüfen lassen. Dieser Baustein kann dann später auch innerhalb der folgenden Bausteine verwendet werden. So etwa innerhalb des nächsten Bausteins „einfache Haupteffekte“. Zur Illustration ein Beispiel: eine zweifaktorielle Varianzanalyse erbrachte eine signifikante Interaktion zwischen Faktor A (drei Altersgruppen) und Faktor B (zwei Geschlechtsgruppen). Wir analysieren daraufhin die „einfachen Haupteffekte“ von Faktor A innerhalb der Stufen von Faktor B (also den Effekt des Alters innerhalb der Männer und der Frauen). Dabei können wir gleichzeitig zwei Kontraste des Altersfaktors prüfen. Nämlich etwa: Unterscheidet sich der Mittelwert der „jungen Männer“ von den „mittelalten Männern“ und unterscheidet sich der Mittelwert der „mittelalten Männern“ von den „alten Männern“ (die gleichen Vergleiche führen wir simultan auch für die Frauen durch). Die Teile des Baukastens abstrahieren dabei vom konkreten Design. Das heißt, die Teile sind sowohl für den univariaten, wie für den multivariaten Fall anwendbar. Sie beziehen sich sowohl auf Gruppierungsfaktoren, wie auf Meßwiederholungsfaktoren und auch auf deren Kombination im Fall sogenannter „mixed models“.

Bevor im folgenden Abschnitt der Baustein „Kontraste“ dargestellt wird, noch eine Bemerkung zu den im Text enthaltenen Formeln. Statistische Formeln schrecken häufig ab und führen dazu, daß ein Text erst gar nicht gelesen wird. Andererseits sind manche Formeln sehr einfach zu „lesen“ und erlauben zudem eine genauere Erläuterung eines Problems. In diesem Dilemma gefangen, habe hat sich der Autor entschlossen, Formeln in den Text aufzunehmen, soweit sie (und nur soweit sie) dem tieferen Verständnis eines Sachverhalts dienlich sind. Sie helfen sogar - unerwarteterweise - das Thema transparenter zu machen und erlauben letztlich auch Transferschlüsse. Die dargestellte Mathematik ist sicherlich mit eher mässigem Oberstufenwissen zu meistern. Der Autor ist sich hier so sicher, weil er selbst (bedauerlicherweise) gerade eben an dieses Niveau heranreicht.

2 Kontrastanalyse - Vergleich von Mittelwerten

Zu Beginn ein paar kurze einführende Worte zur Begrifflichkeit und zum logischen Aufbau dieses Abschnitts. Bei der Kontrastanalyse werden zwei Varianzverhältnisse miteinander verglichen, um (im einfachsten Fall) zwei Mittelwerte zweier Faktorstufen eines Faktors auf statistische Unterschiedlichkeit zu prüfen. Mittelwertvergleiche bilden einen Spezialfall eines Kontrasts (vgl. Beispiele unter 2.1.1). Grundsätzlich lassen sich statistische Tests für Kontraste danach unterscheiden, ob sie **a priori** (auch: **geplant**) oder aber **a posteriori (**auch: **post hoc** oder **nicht-geplant**) formuliert wurden. Es macht einen Unterschied, ob ich bereits vor der Durchführung eines Tests bestimmte Mittelwertunterschiede vermute oder aber im nachhinein beobachtete numerische Differenzen zwischen Mittelwerten vergleichen möchte.

Wie der Name „a priori“ schon vermuten läßt, erfordert der a priori Test eines Kontrast die vorherige Festlegung von Unterschiedshypothesen und sollte daher NIE als Folge der Ergebnisbetrachtung angewendet werden (also post hoc). Das Prinzip der Hypothesenüberprüfung würde verletzt werden, Verallgemeinerungen sind nicht zulässig (Eimer, 1978, S. 58). Es kann also Sinn machen, die eigene mentale Trägheit zu überwinden und sich bereits vor einer Analyse Gedanken zu machen, welche der Gruppen sich eigentlich unterscheiden sollten. Bestehen keine a priori Hypothesen über zu erwartende Unterschiede zwischen Faktorstufen, so sollten **a posteriori** Vergleiche durchgeführt werden. Diese Prüfverfahren werden im Abschnitt 2.2 vorgestellt. Die besondere Stärke der a priori Testung von Vergleichen liegt darin, daß ein gezielter Kontrast signifikant sein kann, obwohl die varianzanalytische Haupteffektprüfung *keinen* signifikanten F-Wert erbracht hat (Kirk, 1968, S. 110). Im Berechnungsbeispiel 2 (vgl. Tabelle 3) werden wir einen solchen Fall darstellen.

Außer der Unterscheidung nach a priori vs. a posteriori lassen sich Kontraste noch danach unterscheiden, ob sie **orthogonal** oder aber **non-orthogonal** sind. Hierbei geht es um das Problem statistischer Unabhängigkeit der angeforderten oder gewünschten Kontraste. Aus dieser Aufteilung ergeben sich also vier „Kontrastarten“: (a) orthogonale a priori Kontraste, (b) non-orthogonale a priori Kontraste, (c) orthogonale a posteriori Kontraste und (d) non-orthogonale a posteriori Kontraste. In SPSS sind die Kontrastarten (a), (b) und (d) verfügbar. Um die „Lücke“ für orthogonale a posteriori Kontraste (Fall c) zu schließen, soll auch ein Verfahren von Tukey (1949) vorgestellt werden. Dieser Test kann mittels SPSS-Outputs recht leicht „zu Fuß“ berechnet werden. Die weitere Darstellung folgt der Logik dieser Aufteilung. Wir beginnen also mit a priori Kontrasten und stellen die Frage „nummero uno“: Was ist eigentlich ein (orthogonaler) Kontrast?

2.1 A priori Kontraste

2.1.1 Orthogonale a priori Kontraste

Was sind orthogonale Kontraste?

Gehen wir von einem einfachen Fall einer einfaktoriellen Varianzanalyse (Gruppierungsfaktor A) mit vier festen Stufen aus (A_1, A_2, A_3 und A_4). **Die statistische Anzahl „orthogonaler“ (also unabhängiger) Vergleiche zwischen den vier Mittelwerten entspricht den Freiheitsgraden eines Faktors (p-1)**. Im vorliegenden Fall sind also nur 3 Vergleiche (=Kontraste) im statistischen Sinne unabhängig voneinander. Das Problem besteht nun darin, daß man sehr viel mehr als nur 3 Unterschiedsfragen stellen kann (hier sind 25 verschiedene denkbar):

Beispiele:

- A_1 vs A_2 / A_1 vs A_3 / A_1 vs A_4 / A_2 vs A_3 / A_2 vs A_4 / A_3 vs A_4
- A_1+A_2 vs A_3 / A_1+A_2 vs A_4 / A_1+A_3 vs A_2 / A_1+A_3 vs A_4 / usw.
- A_1+A_2 vs A_3+A_4 / A_1+A_3 vs A_2+A_4 / A_1+A_4 vs A_2+A_3
- $A_1+A_2+A_3$ vs A_4 / $A_1+A_2+A_4$ vs A_3/ usw.

Die Beispiele machen deutlich, daß man auch einzelne Stufen zusammenziehen kann, um sie mit einer anderen Stufe oder einer anderen „durchschnittlichen“ Stufe zu vergleichen. Etwa: Unterscheidet sich Behandlungsform A_4 vom Schnitt der Behandlungsformen A_1 bis A_3?
Wie bereits oben erläutert: Bei einem Faktor mit 4 Stufen sind statistisch nur 3 Vergleiche orthogonal zueinander, wobei allerdings nicht einmal jede beliebige Kombination von 3 Vergleichen möglich ist. Zur Verdeutlichung des Problems der Orthogonalität ein kurzes „inhaltliches“ Beispiel:

> Drei Personen A, B und C besitzen einen bestimmten Geldbetrag. Wir wissen: (1) A hat 10 DM mehr als B und (2) C hat genausoviel wie A und B im Mittel besitzen. Natürlich wäre es jetzt interessant zu wissen (3), wieviel mehr Geld A als C hat. Allerdings ist diese dritte Frage nicht unabhängig von Frage (1) und Frage (2). Wir können sogar aufgrund der Antworten auf die Fragen (1) und (2) die Antwort auf die dritte errechnen. - Na?

Wir stehen also bei Berechnung spezifischer Vergleiche innerhalb einer Varianzanalyse vor der schwierigen Aufgabe **vorher** feststellen zu müssen, welche unserer Fragen voneinander unabhängig sind. Für diesen Problem haben sich kluge Statistiker eine Technik überlegt, die es einem leicht macht „orthogonale“ Fragen zu formulieren. Natürlich kann man die Technik auch nutzen, um zu überprüfen, ob die gestellten Fragen eigentlich orthogonal sind. Dazu verwendet man sog. „orthogonale Koeffizienten“ als Komponenten eines Zahlensatzes (Zahlensatz = Kontrast) (vgl. Tabelle 1).

Tabelle 1. Orthogonale Koeffizienten

Kontrast	a_1	a_2	a_3	a_4	a_i
1	0	0	-1	1	0
2	-1	1	0	0	0
3	1	1	-1	-1	0
	0	2	-2	0	0

Regeln für Tabelle 1:

- Es gibt soviele Zahlensätze (Kontraste) wie Freiheitsgrade des Faktors (p-1)
- Bedeutung der Koeffizienten:
    1. alle Koeffizienten mit gleichem Wert sind gleichgeschaltet
    2. alle Koeffizienten mit ungleichem Wert werden miteinander verglichen
    3. alle Koeffizienten mit dem Wert „0“ werden bei dem Vergleich nicht berücksichtigt.
- Drei Regeln für die Vergabe der Koeffizienten:
    1. die Summe der Koeffizienten eines Kontrastes ist „0“: $\sum_i c_i = 0$ (Formel 1)
    2. die Summe der Produkte aus den Koeffizienten zweier Kontraste ist „0“ (Beispiel: Summe der Kontrasteprodukte (1) und (2) = (0)(-1)+(0)(1)+(-1)(0)+(1)(0) = 0). Formal: $\sum_i c_{ij} * c_{ik} = 0$ (Formel 2)
    3. die Summe der Spaltensumme der Koeffizienten ist „0“ (Hilfsregel: gilt nur bei einer Konstrastanzahl >2)

Die hier beschriebene Technik ist einfach und effizient. Sie besitzt zudem den Vorteil einer Implementation in analoger Form in der SPSS-Procedure MANOVA. Bevor die entsprechenden Anweisungen vorgestellt werden, sei noch daran erinnert, daß bei jeder Varianzanalyse nur jeweils *eine* Gruppe von orthogonalen Kontrasten durchgeführt werden können. Wird mehr als nur eine Konfiguration von Kontrasten analysiert, so sind die einzelnen ermittelten t-Werte (oder F-Werte) nicht mehr unabhängig voneinander. Dieses Problem entspricht demjenigen der multiplen Testung verschiedener statistischer Hypothesen an ein und demselben Datensatz. In diesem Fall wäre es nötig das α-Niveau (Fehler I Art) zu adjustieren: α‘=1(1-α)^p (wobei p der Anzahl nicht orthogonaler Vergleiche entspricht; vgl. ausführlichere Diskussion zum Wesen der α-Adjustierung im Abschnitt 2.1.4).

Die `CONTRAST`-Anweisung in SPSS.

Kommen wir zur praktischen Vorgehensweise mittels SPSS und der CONTRAST-Anweisung in der (M)ANOVA-Procedure. Ausgangspunkt bildet folgende Datentabelle mit einem vierstufigen Gruppierungsfaktor `GRUPPE` („between-group factor“).

Tabelle 2. Berechnungsbeispiel: Model I - Datentabelle (n=5; p=4).

	A_1	A_2	A_3	A_4
	5	7	9	9
	4	8	7	10
	3	5	8	7
	2	6	5	8
	1	4	6	6
M=	3,00	6,00	7,00	8,00

Anmerkung. Daten wurden dem Beispiel von Eimer (1978, S. 35) entnommen.

Die SPSS-Anweisung für die in Tabelle 1 enthaltenen Kontraste lautet:

```
MANOVA av BY gruppe(1 4)
  /CONTRAST (gruppe)=SPECIAL
                      (1  1  1  1,
                       0  0 -1  1,
                      -1  1  0  0,
                       1  1 -1 -1)
  /PRINT PARAM(ESTIM) SIGNIF(EFSIZE) CELLINFO(MEANS)
  /CINTERVAL INDIVIDUAL(.95) UNIVARIATE
  /METHOD=UNIQUE
  /ERROR WITHIN+RESIDUAL
  /DESIGN.
```

Erläuterungen. Mit „av“ wird die abhängige Variable bezeichnet. Mit „gruppe“ der Gruppierungsfaktor A mit seinen vier Stufen. Unter der Anweisung CONTRAST definieren wir den gewünschten orthogonalen Kontrast selbst. (SPSS bietet zudem die Möglichkeit aus einer Anzahl vorgefertigter Kontraste - teils orthogonaler, teils nicht-orthogonaler Natur - zu wählen, z.B. DIFFERENCE, HELMERT, usw. Die verfügbaren Kontraste in SPSS werden weiter unten erläutert.) Die CONTRAST-Anweisung ist derart formatiert, daß sie einfach zu lesen ist und im Aufbau Tabelle 1 entspricht! Im Unterschied zu Tabelle 1 besteht die „erste“ Zeile des SPECIAL-CONTRASTs aus lauter „einsen“. Diese Zeile ist in der Regel für die Berechnung irrelevant, aber natürlich syntaktisch absolut notwendig (SPSS wird sonst eine seiner berüchtigten Fehlermeldungen ausgeben). Diese Zeile kann aber beispielsweise sinnvoll genutzt werden, wenn die Zellen des Gruppierungsfaktors „gruppe“ nicht gleich besetzt sind (unterschiedliche Stichprobengrößen in jeder Faktorstufe). Der besseren Übersicht halber trägt man in die erste Zeile einfach die jeweilige Stichprobengröße ein. Die ungleichen Zellfrequenzen werden jedoch in jedem Falle bei der Kontrastberechnung berücksichtigt! Die Signifkanztestungen werden mit der Anweisung PARAMETER(ESTIMATES) angefordert und finden sich auch im Output unter dem so bezeichneten Abschnitt. Die Anweisung „/CINTERVAL INDIVIDUAL(.95) UNIVARIATE“ fordert ein 95%iges Konfidenzintervall getrennt für jeden der berechneten Kontraste an. Man kann natürlich beliebige andere Vertrauensgrenzen wählen (95% entspricht der Voreinstellung). Alle Kommandos ab „/METHOD“ entsprechen der Voreinstellung und sind nur der Vollständigkeit halber aufgeführt.

SPSS-Output 1. Berechnungsresultat der Kontraste aus Tabelle 1.

```
-> MANOVA av BY gruppe(1 4)
->   /CONTRAST (gruppe)=SPECIAL
->                      (1  1  1  1,
->                       0  0 -1  1,
->                      -1  1  0  0,
->                       1  1 -1 -1)
->   /PRINT PARAM(ESTIM) SIGNIF(EFSIZE) CELLINFO(MEANS)
->   /CINTERVAL INDIVIDUAL(.95)
->   /METHOD=UNIQUE
->   /ERROR WITHIN+RESIDUAL
->   /DESIGN.
 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 Cell Means and Standard Deviations
 Variable .. AV
      FACTOR           CODE                  Mean  Std. Dev.          N

  GRUPPE                1                   3,000      1,581          5
  GRUPPE                2                   6,000      1,581          5
  GRUPPE                3                   7,000      1,581          5
  GRUPPE                4                   8,000      1,581          5
 For entire sample                          6,000      2,406         20
 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
* * * * * * A n a l y s i s   o f   V a r i a n c e -- design   1 * * * * * *

 Tests of Significance for AV using UNIQUE sums of squares
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL           40,00      16      2,50
 GRUPPE                    70,00       3     23,33      9,33      ,001
 (Model)                   70,00       3     23,33      9,33      ,001
 (Total)                  110,00      19      5,79

 R-Squared =           ,636
 Adjusted R-Squared =  ,568
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 Estimates for AV
 --- Individual univariate ,9500 confidence intervals
 GRUPPE

  Parameter     Coeff.  Std. Err.    t-Value     Sig. t Lower -95%  CL- Upper

        2   1,00000000    1,00000    1,00000     ,33219   -1,11991    3,11991
        3   3,00000000    1,00000    3,00000     ,00848     ,88009    5,11991
        4   -6,0000000    1,41421   -4,24264     ,00062   -8,99800   -3,00200
```

Jede Zeile der `ESTIMATES for AV` steht dabei für genau einen berechneten Kontrast (Nummerierung in der Reihenfolge der Spezifikation der `SPECIAL`-Anweisung, d.h. der Parameter mit der Nummer „2“ steht für die zweite Zeile der `SPECIAL`-Anweisung - also für den ersten Kontrast). Wir könnten also für die Berichtlegung unserer Untersuchung beispielsweise formulieren: „Gruppe 1 und Gruppe 2 unterscheiden sich signifikant voneinander $t(16)=3.0$, $p<.01$, allerdings erwies sich der erwartete Unterschied zwischen Gruppe 3 und 4 als nicht bedeutend $t(16)=1.0$, $p=.33$“ (die Freiheitsgrade des *t*-Wertes entsprechen den Freiheitsgraden des Fehlerterms ("`WITHIN+RISIDUAL`") aus der varianzanalytischen Ergebnistabelle).

Der Ergebnis-Output von SPSS enthält im Abschnitt `ESTIMATES for AV` ebenfalls die angeforderten 95%igen Konfidenzintervalle. Schließt das Konfidenzintervall die „0“ ein, so ist der Kontrast nicht signifikant. Beispiel: Der „tatsächliche“ Unterschied zwischen Gruppe 1 und Gruppe 2 bewegt sich mit 95%iger Wahrscheinlichkeit in der Population zwischen 0,88 und 5,12. Die Bestimmung von Konfidenzintervallen ist die einzige Möglichkeit mit SPSS andere als orthogonale a priori Kontraste auf Signifikanz zu prüfen (vgl. Abschnitt 2.2). Das Konfidenzintervall wird nach folgender Formel berechnet:

$$\left|\overline{A}_j - \overline{A}_{j'}\right| \pm t_{\alpha\,/\,2,\,\upsilon}\sqrt{MS_{error}\left(\frac{(C_j)^2}{n_j}+\frac{(C_{j'})^2}{n_{j'}}\right)}; df = N - p$$

Wobei $\overline{A}_j$ dem Mittelwert des ersten Kontrastteils und $\overline{A}_{j'}$ dem Mittelwert des zweiten Kontrastteils entspricht; C_j bezeichnen die orthogonalen Koeffizienten aus Tabelle 1; MS_{error} ist die unerklärte Fehlervarianz und kann der varianzanalytischen Tabelle (`WITHIN+RESIDUAL`) entnommen werden; n_j und $n_{j'}$ entspricht den Zellfrequenzen (Anzahl von VPn innnerhalb einer Faktorstufe) der miteinander zu vergleichenden Kontraste; p entspricht der Anzahl der Faktorstufen von Faktor A und N steht für den Gesamtstichprobenumfang. Der t-Wert kann einer „handelsüblichen" Tabelle entnommen werden (bei υ-Freiheitsgraden des Fehlerterms [`WITHIN+RISIDUAL;` kann der varianzanalytischen Ergebnistabelle entnommen werden] und definiertem α-Niveau von hier 5%).

Beispiel: $t_{\alpha=.05\,/\,2,16} = 2.120$ (Tabellenwert für υ=16 und α=.05); (Gruppe 1 - Gruppe 2) = 3.0; MS_{error} = 2.50; n_j und $n_{j'}$ = 5

$$|3-6| \pm 2.12\sqrt{2.50\left(\frac{(-1)^2}{5}+\frac{(1)^2}{5}\right)} = 3 \pm 2.12\sqrt{1} = 3 \pm 2.12 = 0.88 / 5.12$$

Dieses Ergebnis entspricht auch dem SPSS-Output (Abschnitt `ESTIMATES for AV` für den diesem Kontrast zugeordneten Parameter mit der Nummer „3").

***t*- oder *F*-Werte**. Jeder t-Wert des Outputs kann per Quadrierung in einen F-Wert umgerechnet werden. Dies ist für den Fall interessant, daß man lieber eine F-Statistik als eine t-Statistik als Beleg berichten möchte. Bei Verwendung des F-Werts wäre der Zählerfreiheitsgrad immer 1 und der Nennerfreiheitsgrad immer gleich dem Fehlerterm ("`WITHIN+RISIDUAL`"). Im vorliegenden Beispiel wäre der Nennerfreiheitsgrad also „16". Wir könnten also beispielsweise formulieren: „Der Kontrast zwischen X und Y ist mit F(1,16)=Quadrat(t-Wert), $p < .01$ signifikant".

Wer überprüfen möchte, ob die berechneten orthogonalen Kontraste auch wirklich orthogonal sind, der kann dies aufgrund des Outputs nachrechnen. Dazu quadriere man zunächst die t-Werte des Abschnitts `Parameter Estimates` und bilde dann den Mittelwert dieser Quadrate. Dieser Wert muß dann dem F-Wert für Faktor A (hier also `gruppe`) entsprechen. Die Erklärung für diesen Überprüfungsweg ist leicht nachvollziehbar: Der F-Wert des Haupteffekts von Faktor „`gruppe`" aus der varianzanalytischen Tabelle (hier: $F_{(3,16)}$=9.33) gibt ein *durchschnittliches* Verhältnis von Varianzen wieder, wenn mehr als ein Freiheitsgrad vorhanden ist. Der F-Wert gibt also in diesem Fall nur an, daß mindestens zwei der vier Gruppen voneinander verschieden sind, aber eben nicht welche. Bei der Berechnung der orthogonalen Kontraste nutzen wir jetzt jeden Freiheitsgrad des Faktors. Wenn man soll will, zerlegen wir also die erklärende Varianz des Faktors A in (hier:) 3 additive Komponenten, sprich Kontraste. Dieses Prüfverfahren kann leider *nicht* angewendet werden, wenn sich die Zellfrequenzen der Faktorstufen unterscheiden (was wohl eher die Regel als die Ausnahme ist).

Die hier vorgestellte Form der orthogonalen Kontrastbildung und -analyse funktioniert auch mit Meßwiederholungsfaktoren, sog. „within-subject"-Faktoren (Beispiel: bei Peter wurde *vor* und *nach* dem Treatment nach Variable A gefragt). Die Spezifikation sowie natürlich auch die Interpretation erfolgt analog. Allerdings können aufgrund einer technischen Besonderheit von

SPSS nicht einfach „nicht-orthogonale“ Kontraste (vgl. Abschnitt 2.1.3) spezifiziert werden. Im Meßwiederholungsfalle orthogonalisiert SPSS die Kontraste automatisch (sie werden übrigens auch noch normalisiert; vgl. die „orthonormalized transformation matrix“ in SPSS, Norusis, 1993, S. 119). SPSS macht dies für die Berechnung des multivariaten *F*-Werts (vgl. Perfomance Plus, 1988, S. 4-7).

Formelanmerkung für Spezialisten und solche, die es werden wollen.

Die folgenden Formeln habe ich in erste Linie deshalb in diesen Text aufgenommen, weil sie es ermöglichen die Logik der Kontrastanalyse zu verfolgen und Zusammenhänge zu vorhandenem Grundlagenwissen aufzuzeigen. Außerdem hat man eine einfache Möglichkeit den Output von SPSS nachzurechnen. Dies mag etwa ein Bedürfnis sein, wenn unklar bleibt, ob SPSS eigentlich ungleiche Beobachtungszahlen in den Faktorstufen berücksichtigt. Die Formel für **orthogonale a priori *t*-Tests** lautet (Kirk, 1968, S. 74):

$$t = \frac{C_j(\overline{A}_i) + C_{j'}(\overline{A}_{i'})}{\sqrt{2MS_{error} / n}}; df = p(n-1) \text{ [Formel 3]}$$

Wobei $\overline{A}_i$ dem Mittelwert des ersten Kontrastteils und $\overline{A}_{i'}$ dem Mittelwert des zweiten Kontrastteils entspricht; C_j bezeichnen die orthogonalen Koeffizienten aus Tabelle 1; MS_{error} ist die unerklärte Fehlervarianz und kann der varianzanalytischen Tabelle (`WITHIN+RESIDUAL`) entnommen werden; *n* entspricht der Zellfrequenz (Anzahl von VPn innnerhalb einer Faktorstufe); und *p* schließlich die Anzahl der Faktorstufen von Faktor A.

Und wie lautet die Formel bei **ungleichen Zellfrequenzen**? Die Formel für ungleiche Zellfrequenzen unterscheidet sich nur im Nenner und ist eine einfache Erweiterung der oberen Formel (Kirk, 1968, S.74):

$$t = \frac{C_j(\overline{A}_i) + C_{j'}(\overline{A}_{i'})}{\sqrt{MS_{error}\left(\frac{(C_j)^2}{n_j} + \frac{(C_{j'})^2}{n_{j'}}\right)}}; df = N - p \text{ [Formel 4]}$$

Hier stehen die n_j und $n_{j'}$ für die Zellfrequenzen der miteinander zu vergleichenden Kontraste und N steht für den Gesamtstichprobenumfang.

Übrigens läßt sich auch die **Quadratsumme** (bzw. mittlere Quadratsumme) **eines Kontrasts** sehr leicht per Hand berechnen (vgl. Hand & Taylor, 1986, S. 25f). (*Anmerkung*. Bei Kontrasten ist die mittlere Quadratsumme immer gleich der Quadratsumme, da ja immer nur durch „einen“ Freiheitsgrad geteilt wird):

$$QS_{(Kontrast)} = \frac{n\left[C_1\overline{A}_1 + C_2\overline{A}_2 + \ldots + C_j\overline{A}_j\right]^2}{C_1^2 + C_2^2 + \ldots + C_j^2} \text{ [Formel 5]}$$

und damit ergibt sich natürlich auch der *F*-Wert als:

$$F = \frac{QS_{(Kontrast)} / 1}{MS_{(Error)}} \text{ ; mit } df_{(Zähler)} = 1 \text{ und } df_{(Nenner)} = N-p \text{ [Formel 6]}$$

Den so berechneten *F*-Wert kann man natürlich auch wieder in den *t*-Wert, der unter dem Abschnitt „`Parameter Estimates`" für diesen Kontrast von SPSS ausgegeben wird, umrechnen: $t = \sqrt{F}$.

Für den Fall **ungleicher Zellfrequenzen** lautet die Formel zur Berechnung der Quadratsumme (vgl. Bortz, 1985, S. 328) sehr ähnlich. Allerdings ist sie für praktische Berechnungen kaum zu verwenden, da die Kontrastkoeffizienten C_1, C_2...C_j ebenfalls neu berechnet werden müssen (in diesem Fall verkomplizieren sich nämlich die beiden oben erläuterten Regeln für Kontrastkoeffizienten erheblich: (1) $\sum_i n_i * C_i = 0$ [Formel 1a] und (2) $\sum_i n_i * C_{ij} * C_{ik} = 0$ [Formel 2a]):

$$QS_{(Kontrast)} = \frac{\left[n_1 C_1 \overline{A}_1 + n_2 C_2 \overline{A}_2 + \ldots + n_j C_j \overline{A}_j\right]^2}{n_1 C_1^2 + n_2 C_2^2 + \ldots + n_j C_j^2} \quad \text{[Formel 7]}$$

Wer diese Formel verwenden will, der kann hier nicht einfach die orthogonalen Koeffizienten aus Tabelle 1 einsetzen. Wie man an die notwendigen Kontrastkoeffizienten berechnet, beschreibt Bortz (1985, S. 326f).
Der Autor hat mit der obigen Formel für den a priori *t*-Test überprüft, ob die von SPSS berechneten *t*-Werte der Kontraste auch für den Fall ungleicher Fallzahlen korrekt sind. Ergebnis: Sie sind es natürlich. Hierzu im folgenden ein ausführliches Beispiel.

2.1.2 Orthogonale a priori Kontraste im Falle ungleicher Zellfrequenzen

Tabelle 3 enthält die Rohdaten eines kleinen Experiments mit drei Faktorstufen. Als wir die Treatmentstufe 1 (10 „sinnvolle“ Silben lernen) und 2 (10 „sinnlose“ Silben lernen) durchführten, meldeten sich jeweils 3 Studenten. Erst nachdem sich herumgesprochen hatte, daß unsere Untersuchung überaus spannend sei, kamen mehr Studenten. Zu diesem Zeitpunkt führten wir gerade Treatmentstufe 3 durch (10 „selbstproduzierte“ Silben lernen). Insgesamt nahmen 7 Studenten an dieser letzten Lernbedingung teil.

Tabelle 3. Berechnungsbeispiel: Model Ia - Datentabelle (n_1=3; n_2=3; n_3=7; p=3).

	A_1 (sinnvolle Silben)	A_2 (sinnlose Silben)	A_3 (selbstproduzierte Silben)
	7	3	7
	6	2	6
	5	1	5
			4
			3
			2
			1
M=	6,00	2,00	4,00

Anmerkung. Daten wurden dem Beispiel von Eimer (1978, S. 37) entnommen.

Wir haben in unserem Experiment nach einer Wartezeit von 5 Minuten die Teilnehmer gebeten, soviele Silben zu erinnern und niederzuschreiben, wie sie in der Lage sind im Gedächtnis wieder zu finden. Abhängige Variable war also einfach die Reproduktionsleistung (im Sinne eines „incidental free recall“). Zu Auswertungszwecken gaben wir schnell die Daten ein, um folgende (M)ANOVA zu berechnen. Die SPSS-Anweisung lautet:

```
MANOVA av1 BY gruppe(1 3)
  /CONTRAST (gruppe)=SPECIAL
                     (3  3  7,
                      1  0 -1,
                     -1  2 -1,
  /PRINT PARAM(ESTIM) SIGNIF(EFSIZE) CELLINFO(MEANS)
  /CINTERVAL INDIVIDUAL(.95) UNIVARIATE
  /METHOD=UNIQUE
  /ERROR WITHIN+RESIDUAL
  /DESIGN.
```

An der Definition der `CONTRAST`-Anweisung ist abzulesen, daß wir a priori angenommen haben, daß das Lernen „sinnvoller“ Silben, dem Lernen „selbstproduzierter“ Silben überlegen ist. Außerdem prüfen wir, ob - wie vermutet - das Lernen „sinnloser“ Silben zu den nachweisbar schlechtesten Lernleistungen führt. Die beiden spezifizierten Kontraste sind orthogonal, da die Summe der Kontrastprodukte (1)(-1)+(0)(2)+(-1)(-1) null ergibt.

SPSS-Output 2. Berechnungsresultat der Daten aus Tabelle 3.

```
-> MANOVA av1 BY gruppe(1 3)
->   /CONTRAST (gruppe)=SPECIAL
->                    (3  3  7,
->                     1  0 -1,
->                    -1  2 -1)
->   /PRINT PARAM(ESTIM) SIGNIF(EFSIZE) CELLINFO(MEANS)
->   /CINTERVAL INDIVIDUAL(.95)
->   /METHOD=UNIQUE
->   /ERROR WITHIN+RESIDUAL
->   /DESIGN.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 Cell Means and Standard Deviations
 Variable .. AV1
      FACTOR           CODE                  Mean  Std. Dev.          N

  GRUPPE                1                   6,000      1,000          3
  GRUPPE                2                   2,000      1,000          3
  GRUPPE                3                   4,000      2,160          7
 For entire sample                          4,000      2,160         13

 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

* * * * * * A n a l y s i s   o f   V a r i a n c e -- design   1 * * * * * *

 Tests of Significance for AV1 using UNIQUE sums of squares
 Source of Variation          SS      DF        MS         F  Sig of F

 WITHIN+RESIDUAL           32,00      10      3,20
 GRUPPE                    24,00       2     12,00      3,75      ,061

 (Model)                   24,00       2     12,00      3,75      ,061
 (Total)                   56,00      12      4,67

 R-Squared =           ,429
 Adjusted R-Squared =  ,314

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 Estimates for AV1
 --- Individual univariate ,9500 confidence intervals

 GRUPPE

  Parameter     Coeff.  Std. Err.    t-Value     Sig. t Lower -95%  CL- Upper

        2   2,00000000    1,23443    1,62019     ,13626    -,75047    4,75047
        3   -6,0000000    2,40634   -2,49341     ,03180  -11,36166    -,63834
```

Der Output zeigt zweierlei. Zunächst stellt sich der Effekt des Faktors „Lernmaterial“ als im statistischen Sinne nicht bedeutsam für die beobachtete Lernleistung heraus. Insbesondere ist unsere Hypothese zu verwerfen, daß das Lernen „sinnvoller“ Silben, dem Lernen „selbstprodzierter“ Silben überlegen ist, $t_{(10)}$=1,62; p<.14. Zum anderen bestätigt sich aber unsere zweite Hypothese, nämlich, daß das Lernen „sinnloser“ Silben im Vergleich zum Durchschnitt der anderen beiden Materialbedingungen zu den nachweisbar schlechtesten Lernleistungen führt ($t_{(10)}$=-2,49; p<.04).

Dieses Beispiel zeigt darüber hinaus, daß ein spezifischer a priori Kontrast statistisch bedeutend sein kann, obwohl sich bei der Haupteffektprüfung ein negativer Befund (vgl. SPSS-Output 2) ergab. Prüfen wir mit der oben beschriebenen Formel [4] für ungleiche Zellfrequenzen (Kirk, 1968, S.74), ob SPSS den richtigen *t*-Wert für die beiden Kontraste errechnet hat:

$$t = \frac{C_j(\bar{A}_i) + C_{j'}(\bar{A}_{i'})}{\sqrt{MS_{error}\left(\frac{(C_j)^2}{n_j} + \frac{(C_{j'})^2}{n_{j'}}\right)}} ; df = N - p \text{ [Formel 4]}$$

Kontrast 1:

$$t = \frac{(1)(6.0) + (-1)(4.0)}{\sqrt{3{,}20 \frac{1^2}{3} + \frac{-1^2}{7}}} = \frac{2}{\sqrt{3{,}20(0{,}33 + 0{,}14)}} = \frac{2}{\sqrt{1.504}} = \frac{2}{1.23} = 1.63$$

Kontrast 2:

$$t = \frac{(-1)(6.0) + (2)(2.0) + (-1)(4.0)}{\sqrt{3{,}20 \frac{-1^2}{3} + \frac{2^2}{3} + \frac{-1^2}{7}}} = \frac{-6}{\sqrt{3{,}20(0{,}33 + 1.33 + 0{,}14)}} = \frac{-6}{\sqrt{5.76}} = \frac{-6}{2.4} = -2.50$$

Die beiden errechneten *t*-Werte zeigen, daß die von SPSS unter den "`Estimates for AV1`" errechneten *t*-Werte tatsächlich die ungleichen Fallzahlen berücksichtigen. Die kleinen Abweichungen sind auf Rundungsfehler zurückzuführen. Da sich die Berechnung der Quadratsummen für die beiden Kontraste im Falle ungleicher Zellfrequenzen verhältnismäßig aufwendig darstellt, sollte darauf im allgemeinen verzichtet werden. Dadurch können wir leider nicht auf einfache Art und Weise mittels des SPSS-Outputs *F*-Werte für unsere Kontraste berechnen (im Fall ungleicher Zellfrequenzen ist $F \neq t^2$ - dies gilt selbstverständlich nicht allgemein, sondern nur für die „`t-Values`“ des SPSS-Outputs). Aus dem gleichen Grund ist die Prüfung der Orthogonalität der Kontraste ebenfalls nicht auf einfache Art möglich. Prinzipiell geht es natürlich, wie die folgenden Berechnungen zeigen. Bestimmen wir also für die beiden Kontraste zunächst neue Kontrastkoeffizienten, die die Regeln (1) und (2) erfüllen:

Damit die Regel (1) $\sum_i n_i * C_i = 0$ für den Fall ungleicher Zellfrequenzen erfüllt wird, suchen wir für den ersten Kontrast „1 0 -1“ (n_1=3; n_2=3; n_3=7) neue Kontrastkoeffizienten, sodaß gilt:

Kontrast 1 = 3 * C_{11} + 3 * C_{21} + 7 * C_{31} = 0 - Regel (1)

Da wir für C_{21}=0 setzen ist (die Treatmentstufe ist nicht Teil des Vergleichs), ist die Lösung einfach: C_{11}=7, C_{21}=0 und C_{31}=-3. Für den zweiten Kontrast „-1 2 -1“ (n_1=3; n_2=3; n_3=7) suchen wir auf die gleiche Art Kontrastkoeffizienten, für die gilt:

Kontrast 2 = 3 * C_{12} + 3 * C_{22} + 7 * C_{32} = 0 - Regel (1)

Da der zweite Kontrast nur dann orthogonal zum ersten Kontrast ist, wenn auch die zweite Regel erfüllt ist ($\sum_i n_i * C_{ij} * C_{ik} = 0$), muß der zweite Kontrast auch die folgende zusätzliche Bedingung erfüllt sein:

Kontrast 2 = 3 * C_{12} * C_{11} + 3 * C_{22} * C_{21} + 7 * C_{32} * C_{31} = 0 - Regel (2)

Da C_{11}=7, C_{21}=0 und C_{31}=-3 ist, errechnen wir durch einsetzen, daß $21C_{12}$=$21C_{32}$, daß heißt also einfach, daß C_{12} = C_{32} ist. Dies vereinfacht die Regel (1) für Kontrast 2, sodaß gilt:

Kontrast 2 = 3 * C_{22} + 10 * C_{32} = 0   - Regel (1)

Als mögliche Werte kämen dann für C_{22} = -10 und C_{32}= 3 (=C_{12}) in Frage. Die Koeffizienten für Kontrast 1 lauten 7, 0 und -3 und für Kontrast 2 3, -10 und 3. Wir prüfen noch einmal:

```
Kontrast 1 = [3 * 7] + [3 * 0]      + [7 * (-3)] = 0 - Regel (1)
Kontrast 2 = [3 * 3] + [3* (-10)]   + [7 * 3]    = 0 - Regel (1)
-----------------------------------------------------------------
             [3*7*3] + [3*0*(-10)] + [7*(-3)*3] = 0 - Regel (2)
```

Die so müselig errechneten neuen Kontrastkoeffizienten setzen wir nur in die Formel [7] für ungleiche Zellfrequenzen ein (vgl. Bortz, 1985, S. 328):

$$QS_{(Kontrast)} = \frac{\left[n_1C_1\overline{A}_1 + n_2C_2\overline{A}_2 + \ldots + n_jC_j\overline{A}_j\right]^2}{n_1C_1^2 + n_2C_2^2 + \ldots + n_jC_j^2} \quad \text{[Formel 7]}$$

Kontrast 1: C_{11}=7, C_{21}=0 und C_{31} =-3 und n_1=3, n_2=3 und n_3=7

$$QS(Kontrast-1) = \frac{\left[(3)(7)(6.0) \;+\; (7)(\text{-}3)(4.0)\right]^2}{(3)(7)^2 + (7)(\text{-}3)^2} = \frac{[126-84]^2}{147+63} = \frac{42^2}{210} = \frac{1764}{210} = 8{,}4$$

Kontrast 2: C_{12}=3, C_{22}=-10 und C_{32} =3 und n_1=3, n_2=3 und n_3=7

$$QS_{(Kontrast\,-\,2)} = \frac{\left[(3)(3)(6.0) + (3)(\text{-}10)(2.0) + (7)(3)(4.0)\right]^2}{(3)(3)^2 + (3)(\text{-}10)^2 + (7)(3)^2} = \frac{[54-60+84]^2}{27+300+63} = \frac{78^2}{390} = \frac{6084}{390} = 15{,}6$$

Prüfen wir zunächst die Orthogonalität der beiden Kontraste mit

$QS_{\text{Faktor A}} = QS_{\text{Kontrast-1}} + QS_{\text{Kontrast -2}}$ = 24,0. [Formel 8]

Dies entspricht dem Ergebnis der varianzanalytischen Tabelle im SPSS-Output 2 für den Gruppierungsfaktor „`gruppe`“. Durch Teilung der errechneten Quadratsummen von Kontrast 1 bzw. 2 durch die mittlere Fehlerquadratsumme MS_{Error} = 3,20 erhalten wir als *F*-Werte $F_{(1,10)}$= 2.63 für Kontrast 1 und $F_{(1,10)}$= 4.88 für Kontrast 2. **Fazit**: Da (a) ungleiche Zellfrequenzen eher die Regel, denn die Ausnahme bilden und (b) die Berechnung der Quadratsummen für die gewünschten Kontraste im Falle ungleicher Zellfrequenzen grusselige Ausmaße annimmt, wird man sich in der Praxis mit der Bestimmung von *t*-Werten begnügen.

2.1.3 Die `CONTRAST`-Optionen in SPSS

Außer den Möglichkeiten, die die `SPECIAL`-Option des `Contrast`-Kommandos bietet, stellt SPSS eine Reihe „vorgefertigter“ Kontraste zur Verfügung. Diese seien hier kurz genannt (vgl. Norusis, 1993, S. 51f).

- **`Deviation`**. Jede Stufe eines Faktors (mit Ausnahme der Referenzkategorie=Faktorstufe) wird mit dem Gesamtmittelwert verglichen. Dies ist die Voreinstellung.

- **`Simple`**. Jede Stufe eines Faktors (mit Ausnahme der Referenzkategorie) wird mit der Referenzkategorie verglichen.
- **`Difference`**. Auch als inverser Helmert-Kontrast bekannt. Jede Stufe eines Faktors (mit Ausnahme der ersten Stufe) wird mit dem Durchschnitt der vorgeordneten Stufen verglichen.
- **`Helmert`**. Jede Stufe eines Faktors (mit Ausnahme der letzten Stufe) wird mit dem Durchschnitt der nachfolgenden Stufen verglichen.
- **`Repeated`**. Jede Stufe eines Faktors (mit Ausnahme der ersten Stufe) wird mit der vorgeordneten Stufe verglichen (dieser Kontrast ist nicht orthogonal!).
- **`Polynomial`**. Der Faktor wird in lineare, quadratische, kubische und so weiter (je nach Anzahl der Faktorstufen) Kontraste aufgeteilt. Es wird angenommen, daß die Faktorstufen äquidistant sind (gleicher Abstand der Faktorstufen im Sinne einer Intervalskala).

Für den `Deviation` und den `Simple`-Kontrast kann man eine Referenzkategorie (Faktorstufe) angeben, die dann vom Vergleich ausgeschlossen wird. `LAST` wählt dabei die numerisch größte Faktorstufe (Voreinstellung) und `FIRST` wählt die numerisch kleinste Faktorstufe.

2.1.4 Non-orthogonale a priori Kontraste

Häufig steht man vor dem Problem, viel mehr Vergleiche anstellen zu wollen, als Freiheitsgrade zur Verfügung stehen. Die interessierenden Vergleiche sind also non-orthogonal. Dieses Problem stellt sich übrigens auch häufig im Falle sorgfältigster experimenteller Versuchsplanung. In diesem Abschnitt geht es um ein Verfahren, das es erlaubt sich über die Begrenzung orthogonaler a priori Kontraste hinwegzusetzen und weitere eben „nicht-orthogonale“ Kontraste zu berechnen.

Der Ausdruck „nicht-orthogonal“ bedeutet soviel wie „nicht unabhängig“ voneinander. Man könnte diesen Ausdruck nun für einen typischen Fall akademischer Sprachakrobatik halten und demzufolge die doppelte Verneinung durch ein Positiv vereinfachen wollen. Dies wäre aber nicht richtig, da „nicht-orthogonal“ eben nicht soviel wie „abhängig“ bedeutet. Beispiel: Die folgenden zwei Aussagen sind - statistisch gesehen - nicht orthogonal zueinander: (1) Elephanten sind größer als Tiger. (2) Wale sind größer als Elephanten. Schreiben wir diese beiden Kontrastaussagen zur Prüfung in die in Abschnitt 2.1.1 eingeführte Tabellenform.

Tabelle 4. Ein Beispiel nicht-orthogonale Koeffizienten (vgl. Text).

Kontrast	Tiger	Elephant	Wal	♠a_i
1	1	-1	0	0
2	0	1	-1	0
Produkte	(1)(0) = 0	(-1)(1) = -1	(0)(-1) = 0	-1

Aus der letzten Zeile von Tabelle 4 ist ersichtlich, daß die Summe der Produkte dieser Koeffizienten nicht 0 ist. Obwohl also im statistischen Sinne nicht gesagt werden kann, daß die beiden Aussagen voneinander unabhängig sind, so kann man auch nicht sagen, daß die beiden Aussagen voneinander "abhängig" wären. Genau genommen überschneiden sich die beiden Aussagen lediglich, ohne dabei jedoch ihren eigenen Aussagewert zu verlieren (vgl. Eimer, 1978, S. 71). Genau solche Überschneidungen treten aber sehr häufig auf. Nehmen wir als Beispiel den typischen Fall alle paarweise möglichen Vergleiche zwischen den Mittelwerten der Faktorstufen berechnen zu wollen (im Falle eines Faktors mit 4 Faktorstufen ergäben sich also

$k(k\text{-}1)/2 = 6$ Vergleiche). Natürlich ergeben sich bei 6 Vergleichen nicht nur Vergleiche mit Überschneidung, sondern auch einige inhaltlich völlig redundante Vergleiche (für das Beispiel in Tabelle 4 wäre etwa der Vergleich Tiger vs Wal redundant). Allerdings mag es dem Wissenschaftler ja auch darauf ankommen, eine solche Information - obwohl redundant - zum Zwecke der Betonung zu erwähnen. In der natürlich-sprachlichen Fassung psychologischer Theorien - nicht allerdings in deren strukturalistischer Rekonstruktion - dürfte sich auch die ein oder andere redundante Aussage befinden. In diesem Sinne kann ich nichts verwerfliches an einer redundanten statistischen Aussage finden - im Gegenteil mögen dadurch die häufig komplizierten Forschungsbefunde verständlicher werden.

Im folgenden wird nun eine Methode vorgestellt, die es erlaubt solche überschneidenden oder aber redundanten Fragen statistisch zu beantworten. In SPSS ist Dunn's Verfahren multipler **a priori** Mittelwertvergleiche (Dunn, 1961) implementiert. Das von Dunn beschriebene Verfahren wird in SPSS unter dem Namen *Bonferroni t-Statistik* geführt. Diese Bezeichnung hat Miller (1966) für Dunn’s Verfahren eingeführt, da das Verfahren letztlich auf Bonferroni zurückgeht. Soviel zur Begriffsverwirrung.
Dunn’s Verfahren multipler Mittelwertvergleiche oder alternativ die Bonferroni *t*-Statistik kann man also nutzen, um *alle möglichen* und nicht nur alle orthogonalen „geplanten“ Vergleiche zwischen den Mittelwerten anzustellen. Das Verfahren beruht darauf, das „α-Niveau“ für den ganzen Satz geplanter Vergleiche aufzuteilen. Im Einzelnen:

Ziel der Bonferroni *t*-Statistik ist es, alle Einzelvergleiche eines Satzes statistischer Hypothesen auf einem gemeinsamen - für jeden der Einzelvergleiche gültigen - α-Niveau abzusichern. Es soll also sichergestellt werden, daß wir H_0 nicht fälschlicherweise verwerfen, nur weil wir mehr Einzelvergleiche prüfen als statistisch orthogonal zueinander sind. Was das Verfahren von Dunn und überhaupt jedes Verfahren für die Analyse nicht-orthogonaler Vergleiche also im Grunde tut, ist schlicht eine Adjustierung des Alpha-Niveaus vorzunehmen. Die zugrunde liegende Logik entspricht dabei dem Fall multipler, am **gleichen Datensatz getrennt** durchgeführter Hypothesenprüfungen. Wird über 2 oder mehr **unabhängige** Hypothesen getrennt entschieden (etwa über mehrere unabhängige *t*-Tests), beträgt die Wahrscheinlichkeit, daß mindestens eine H_0 fälschlicherweise verworfen wird

$$\alpha' = 1-(1-\alpha)^p, \text{ [Formel 9]}$$

wobei p der Anzahl der Hypothesen und α dem Signifikanzniveau für die einzelnen Tests entspricht. Bei einem nominellen α-Niveau von 5 % (nominell = für den ganzen Hypothesensatz festgelegtes α-Niveau) und nur 3 **unabhängigen** Vergleichen beträgt die tatsächliche Wahrscheinlichkeit mindestens eine Nullhypothese fälschlicherweise zu verwerfen 14 %. Bei 10 Vergleichen liegt die Wahrscheinlichkeit bereits bei 40 %. Entsprechende Wahrscheinlichkeiten lassen sich auch für abhängige Vergleiche berechnen. Leider ist dies ziemlich kompliziert (entfällt daher hier), führt aber in jedem Fall zu geringeren Wahrscheinlichkeiten. Die grundsätzliche Frage, die sich stellt, lautet also etwa: Soll die Wahrscheinlichkeit für einen Fehler vom Typ I für jeden einzelnen Vergleich festgelegt werden oder für eine größere konzeptuelle Einheit wie etwa für eine bestimmte Sammlung (Anzahl) von Vergleichen? In Abschnitt 2.1.1 und 2.1.2 haben wir bereits eine Antwort kennengelernt. In diesem Fall („geplanter orthogonaler Vergleiche“) wird das α-Niveau für jeden Vergleich einzeln festgelegt. Hier wird also gar nichts „adjustiert“. Für geplante (a priori) aber auch für nicht geplante (a posteriori, vgl. Abschnitt 2.2) **non-orthogonale** Vergleiche wird die

Fehlerwahrscheinlichkeit jedoch für eine definierte Ansammlung an Vergleichen bestimmt. So auch, wie bereits einleitend gesagt, das Vorgehen von Dunn (vgl. auch Kirk, 1968, S. 79-81).

Sie hat vorgeschlagen das nominelle α-Niveau ganz einfach für die *Anzahl geplanter Vergleiche* aufzuteilen. Nehmen wir an, wir wollten 10 Vergleiche berechnen und würden das α-Niveau für diesen Hypothesensatz auf .05 setzen, dann ergäbe sich ein α-Niveau für jeden der Vergleiche von

$$\alpha_i = \alpha / C = .05 / 10 = .005, \text{ [Formel 10]}$$

C entspricht der Anzahl der geplanten Vergleiche. Die Adjustierung des α-Niveau's richtet sich nach der Anzahl der geplanten Vergleiche. Die entsprechenden *a posteriori* Verfahren (vgl. Abschnitt 2.2) richten sich hingegen nach der Anzahl der Mittelwerte bzw. der Anzahl der Faktorstufen.

In SPSS wird nun leider nicht einfach ein Bonferroni *t*-Wert ausgegeben (plus *p*-Wert), um einen bestimmten Kontrast zu prüfen. Stattdessen wird ein Konfidenzintervall für jeden der spezifizierten Kontraste berechnet. Diese Konfidenzintervalle werden in SPSS als „simultaneous confidence intervals" oder auch als „joint confidence intervals" bezeichnet. Diese Bezeichnung macht, nach dem was gerade erläutert wurde, auch Sinn. Für jeden der Kontraste/Vergleiche legen wir „simultan" das gleiche adjustierte α-Niveau zugrunde und jeder Vergleich/Kontrast stammt aus dem „Verbund" des Hypothesensatzes für den ein nominelles α-Niveau festgelegt wurde (in der Regel 5%).

Zur Illustration betrachten wir noch einmal das zweite Berechnungsbeispiel (vgl. Tabelle 3; SPSS-Output 2). Wir können nun statt der orthogonalen Kontraste jetzt auch zwei beliebige andere Kontraste berechnen. Vergleichen wir beispielsweise nun die Lernleistung „sinnloser Silben" (A_2) jeweils mit den beiden anderen Treatmentstufen („sinnvolle Silben" A_1; selbstproduzierte Silben, A_3). Für diesen Fall lautet die SPSS-Anweisung:

```
MANOVA av1 BY gruppe(1 3)
  /CONTRAST (gruppe)=SPECIAL
                     (3  3  7,
                     -1  1  0,
                      0  1 -1)
  /PRINT PARAM(ESTIM) SIGNIF(EFSIZE) CELLINFO(MEANS)
  /CINTERVAL JOINT(.95) UNIVARIATE(BONFER)
  /METHOD=UNIQUE
  /ERROR WITHIN+RESIDUAL
  /DESIGN.
```

Betrachtet man diese Anweisung, so sieht man, daß es nur einen, aber entscheidenden Unterschied zur bereits bekannten Berechnung orthogonaler a priori Kontraste gibt: Das „`/CINTERVAL`"-Subkommando. Lautete die Anweisung zuvor „`/CINTERVAL INDIVIDUAL(.95) UNIVARIATE`", so lautet sie jetzt „`/CINTERVAL JOINT(.95) UNIVARIATE(BONFER)`". Für jeden der beiden gewünschten Kontraste wird also ein Konfidenzintervall auf der beschriebenen α-adjustierten Basis berechnet. Im Berechnungsbeispiel haben wir ein nominelles α-Niveau von 5% festgelegt (durch Festlegung des Konfidenzintervalls auf 95%). Da SPSS die Anzahl simultan berechenbarer Kontraste auf *k*-1 (Anzahl der Faktorstufen) begrenzt, beträgt das α-Niveau für jeden der zwei Kontraste im Berechnungsbeispiel 2,5%. Das Konfidenzintervall wird daher etwas größer sein, als im orthogonalen a priori Fall, bei welchem ja das α-Niveau nicht adjustiert wird (vgl. SPSS-Output 2). Das Ergebnis für die „Bonferroni-adjustierten“ Kontraste findet sich im SPSS-Output 3.

SPSS-Output 3. Bonferroni adjustierte a priori Kontraste (Daten aus Tabelle 3).

```
...

Estimates for AV1
 --- Joint univariate ,9500 BONFERRONI confidence intervals

 GRUPPE
  Parameter      Coeff.  Std. Err.    t-Value     Sig. t Lower -95%  CL- Upper

        2   -4,0000000    1,46059   -2,73861     ,02088   -7,84686    -,15314
        3   -2,0000000    1,23443   -1,62019     ,13626   -5,25119    1,25119
 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

Für die Interpretation des SPSS-Outputs können *nicht* die Werte der „`t-Value`“-Spalte oder der „`Sig. t`“-Spalte herangezogen werden. Diese Werte sind immer gleich und nur für den orthogonalen a priori Fall direkt verwendbar (vgl. SPSS-Output 2). Wie bereits unter Abschnitt 2.1.1 erläutert ist ein Kontrast genau dann *nicht* signifikant, wenn das Konfidenzintervall die „0“ einschließt. Der Vergleich der Lernleistung für „sinnlose Silben“ (A_2) mit derjenigen für „sinnvolle Silben“ (A_1) ist also signifikant, während der Vergleich für „sinnlose Silben“ (A_2) mit „selbstproduzierten Silben“ (A_3) nicht signifikant ist.
Rechentechnisch ergeben sich die Konfidenzintervalle für a priori Vergleiche nach dem Verfahren von Dunn genauso wie bereits im Abschnitt 2.1.1 erläutert. Allerdings wird nicht der kritische *t*-Wert verwendet, sondern eben Dunn’s *t*-Wert (oder auch Bonferroni *t*-Wert). Die Formel für das Dunn/Bonferroni-Konfidenzintervall lautet:

$$\left|\overline{A}_j - \overline{A}_{j'}\right| \pm t'D_{\alpha/2;C,\upsilon}\sqrt{MS_{error}\left(\frac{(C_j)^2}{n_j}+\frac{(C_{j'})^2}{n_{j'}}\right)} \text{ [Formel 11]}$$

Der kritische Dunn-Wert „$t'D_{\alpha/2;C,\upsilon}$“ muß einer Tabelle entnommen werden. Eine solche Tabelle findet sich im Anhang (Tabelle A.1). *C* entspricht der Anzahl der Vergleiche die zwischen den *k* Mittelwerten vorgenommen wurden und υ den Freiheitsgraden des Fehlerterms MS_{error}. Im vorliegenden Fall müßten wir einen $t'D$- Wert nachschlagen für C=2 Kontraste, υ=10 Freiheitsgrade und einem nominellen α-Niveau von .05 ($t'D$ = 2.64). Der nunmehr geübte Formelleser möge doch einmal das Konfidenzintervall für den ersten berechneten Kontrast (Parameter 2) im SPSS-Output mittels dieser Angaben nachrechnen. Der Autor hat als obere Grenze -.1456 und als untere Grenze des Konfidenzintervalls -7.85 errechnet. Dies entspricht dem SPSS-Output (bis auf die üblichen Rundungsfehler).

Dunn (1961) hat ihr Verfahren mit „a posteriori“ Verfahren nach Tukey und Scheffé (vgl. Abschnitt 2.2) verglichen. Wenn sehr viele Faktorstufen *k* (Mittelwerte) vorliegen, aber nur eine begrenzte Anzahl von Vergleichen *C* zwischen den Mittelwerten vorgenommen werden sollen, resultieren beim Dunn-Verfahren kleinere Konfidenzintervalle als bei jedem a posteriori Verfahren. Die Teststärke ist in diesem Fall also höher. Allerdings erweisen sich die a posteriori Verfahren als günstiger (bezogen auf die Teststärke), wenn *k* relativ klein ist und *C* relativ groß. Das liegt daran, daß sich bei Dunn die Adjustierung des α-Niveau’s nach der Anzahl der geplanten Vergleiche *C* richtet. Die entsprechenden *a posteriori* Verfahren richten sich hingegen nach der Anzahl der Mittelwerte *k*, also der Anzahl der Faktorstufen.

Abschließend noch zwei technische Hinweise. Dunn hat etwas später (Dunn & Massey, 1965) auch Verfahren für die Berechnung simultaner Konfidenzintervalle vorgeschlagen, die auf der *t*-Verteilung beruhen. Soweit der Autor das beurteilen kann, liegt dem von SPSS ausgegebenen „joint confidence interval", das *t*-verteilungsbasierte Verfahren zugrunde. Der zweite Hinweis bezieht sich auf das Problem, daß SPSS die Anzahl simultan berechenbarer Kontraste auf *k*-1-Faktorstufen begrenzt. Leider kann man innerhalb der „`CONTRAST`"-Anweisung nicht einfach alle Kontraste, die berechnet werden sollen, spezifizieren. Praktisch führt das dazu, daß einfach mehrere Analysen hintereinander gerechnet werden, bis eben Ergebnisse für alle gewünschten Kontraste vorliegen. Die jeweils ausgegebenen Konfidenzintervalle beruhen dann natürlich *nicht* auf der Gesamtzahl der durchführten Vergleiche C , wie es für eine korrekte Interpretation nötig wäre, sondern nur auf der reduzierten Kontrastanzahl von *k*-1. Will man mehr als eben *k*-1 Kontraste berechnen, so kommt man nicht umhin Dunn's-Tabelle zu verwenden. Die Adjustierungsbereitschaft von SPSS endet genau an diesem Punkt. Der nachgeschlagene $t'D$- Wert entspricht der „kritischen Differenz" für alle durchführten Kontraste. Der $t'D$- Wert muß dann nur mit der Kontrastdifferenz der „`Coeff.`"-Spalte (vgl. SPSS Output 3) verglichen werden. Ist die empirische Differenz größer als der „kritische" Differenzwert $t'D$, so ist der Kontrast signifikant. Im obigen Beispiel betrug $t'D$ = 2.64 und war somit kleiner als Kontrast 1 und größer als Kontrast 2. Kontrast 1 ist demnach signifkant ($p < .05$), Kontrast 2 hingegen nicht.

2.2 A posteriori Kontraste

Häufig hat man bei exploratischen Datenanalysen - wie sie typisch für Feldstudien sind - keine a priori Hypothesen aufgestellt. Dennoch möchte man gern bestimmte Kontraste berechnen (also post hoc Vergleiche durchführen). In SPSS ist nur ein Verfahren für multiple **a posteriori** Mittelwertvergleiche implementiert. Es handelt sich um den berühmten Scheffé-Test (Scheffé, 1953, 1959), der als ein sehr konservatives Verfahren gilt. Der Scheffé-Test ist ein Verfahren für non-orthogonale Kontraste. Er adjustiert das „α-Niveau" gar für alle überhaupt möglichen Kontraste eines jeweiligen Designs (etwa für alle 25 möglichen Kontraste des unter Abschnitt 2.1.1 dargestellten Beispiels).

Tukey's **a posteriori** Verfahren (Tukey, 1949) ist nicht in SPSS integriert. Dennoch wird das Verfahren von Tukey vorgestellt, da es weniger konservativ als der Scheffé-Test ist und außerdem verhältnismäßig leicht „zu Fuß" mit Hilfe des SPSS-Output berechnet werden kann. Tukey's Verfahren ist ebenfalls für den Test non-orthogonaler Kontraste entwickelt, schließt allerdings die Möglichkeit zur Prüfung orthogonaler Kontraste ein. Tukey's Test hat eine deutlich höhere Teststärke als das Verfahren von Scheffé. Andersherum könnte man grob formuliert sagen, daß positiven (signifikanten) Testergebnissen mittels Scheffé getrost „Glauben" geschenkt werden darf.

Es sei der Vollständigkeit halber erwähnt, daß eine Reihe weiterer Verfahren für die Berechnung nicht-orthogonaler **a posteriori** Kontraste zur Verfügung stehen (vgl. Bortz, 1985, Kap. 7.3; Eimer, 1978, Kap. 8; Kirk, 1968, Kap. 3; Timm, 1975; Winer, Brown & Michels, 1991).

2.2.1 Der Scheffé Test

Genau wie beim Verfahren von Dunn oder der Bonferroni *t*-Statistik ist das Ziel des Scheffé-Tests, alle Einzelvergleiche eines Satzes statistischer Hypothesen auf einem gemeinsamen (für jeden der Einzelvergleiche gültigen) α-Niveau abzusichern. Anders als beim a priori

Verfahren, daß die Adjustierung des α-Niveau's nach der tatsächlichen Anzahl der Vergleiche vornimmt, richten sich die *a posteriori* Verfahren (und damit auch der Scheffé-Test) nach der Anzahl der Mittelwerte bzw. der Anzahl der Faktorstufen. Das adjustierte α-Niveau nach dem Scheffé-Test berechnet sich einfach aus der Anzahl der Faktorstufen eines Faktors minus 1 (*k*-1) als

$$\alpha' = 1 - (1-\alpha)^{k-1}, \text{ [Formel 9a]}$$

wobei hier k der Anzahl der Fakorstufen und α dem gemeinsamen Signifikanzniveau der definierten Anzahl von Vergleichen entspricht (also in der Regel 5 %). Im Falle des Scheffé-Tests besteht die Anzahl definierter Vergleiche aus allen überhaupt möglichen Einzelvergleichen (also auch aus Vergleichen, die sich auf Kombinationen von Mittelwerten beziehen).

Beispiel: Für einen Faktor A mit drei Faktorstufen

- 6 Vergleiche sind möglich:
  - A_1 vs A_2
  - A_2 vs A_3
  - A_1 vs A_3
  - A_1 vs A_2+A_3
  - A_2 vs A_1+A_3
  - A_3 vs A_1+A_2
- Für das Set der 6 Vergleiche legen wir ein 5 %-Niveau fest.

- Daraus ergibt sich nach Umformung von Formel [9a] für jeden der 6 Einzelvergleiche ein Signifikanzniveau von 1,7%, da

$$\alpha = 1 - (1-\alpha_{(Set)})^{\frac{1}{k-1}} = 1-(1-0{,}05)^{\frac{1}{3-1}} = 0.017 \text{. [Formel 9b]}$$

In SPSS wird - genau wie beim a priori Verfahren nicht-orthogonaler Kontraste (vgl. Abschnitt 2.1.4) - ein Konfidenzintervall für jeden der spezifizierten Kontraste berechnet. Betrachten wir deshalb auch das gleiche Berechnungsbeispiel von oben erneut. Allerdings wollen wir nun Scheffé-adjustierte und nicht Bonferroni-adjustierte Kontraste berechnen lassen. Für diesen Fall ändert sich die SPSS-Anweisung, wie folgt:

```
MANOVA av1 BY gruppe(1 3)
  /CONTRAST (gruppe)=SPECIAL
                     (3  3  7,
                     -1  1  0,
                      0  1 -1)
  /PRINT PARAM(ESTIM) SIGNIF(EFSIZE) CELLINFO(MEANS)
  /CINTERVAL JOINT(.95) UNIVARIATE(SCHEFFE)
  /METHOD=UNIQUE
  /ERROR WITHIN+RESIDUAL
  /DESIGN.
```

Betrachtet man die Syntax, so sieht man, daß „`/CINTERVAL JOINT(.95) UNIVARIATE(BONFER)`“ einfach durch „`/CINTERVAL JOINT(.95) UNIVARIATE(SCHEFFE)`“ ersetzt wurde. Für jeden der zwei spezifizierten Kontraste wird ein Konfidenzintervall auf der beschriebenen α-adjustierten Basis berechnet. Im

Berechnungsbeispiel haben wir wieder ein nominelles α-Niveau von 5% festgelegt (durch Festlegung des Konfidenzintervalls auf 95%). Das α-Niveau für jeden der zwei Kontraste beträgt 1,7% (kann wie oben gezeigt durch Umformung von [Formel 9a] errechnet werden). Das Konfidenzintervall wird daher wiederum etwas größer sein, als im nicht-orthogonalen a priori Fall, bei welchem das α-Niveau „nur" auf 2,5% pro Kontrast adjustiert wurde (vgl. SPSS-Output 3). Das Ergebnis für die „Scheffé-adjustierten" Kontraste findet sich im SPSS-Output 4.

SPSS-Output 4. Scheffé-adjustierte a posteriori Kontraste (Daten aus Tabelle 3).

```
...
Estimates for AV1
 --- Joint univariate ,9500 SCHEFFE confidence intervals

 GRUPPE

  Parameter      Coeff.  Std. Err.     t-Value     Sig. t Lower -95%  CL- Upper

        2   -4,0000000    1,46059   -2,73861     ,02088   -8,18394     ,18394
        3   -2,0000000    1,23443   -1,62019     ,13626   -5,53608    1,53608
```

Für die Interpretation des SPSS-Outputs können *nicht* die Werte der „`t-Value`"-Spalte oder der „`Sig. t`"-Spalte herangezogen werden. Diese Werte sind für alle innerhalb von SPSS verfügbaren Kontrastarten gleich und nur für den orthogonalen a priori Fall direkt verwendbar (vgl. SPSS-Output 2). Wie bereits unter Abschnitt 2.1.1 erläutert ist ein Kontrast genau dann *nicht* signifikant, wenn das Konfidenzintervall die „0" einschließt. An diesem Beispiel zeigt sich nun auch die geringere Teststärke des Scheffé-Tests. So wurde im a priori non-orthogonalen Fall (vgl. Abschnitt 2.1.4) der Vergleich der Lernleistung für „sinnlose Silben" (A_2) mit derjenigen für „sinnvolle Silben" (A_1) noch signifikant, während derselbe Vergleich im vorliegenden Fall einer a posteriori Scheffé-Testung nicht signifikant ist.

Die statistische Kenngröße für den Scheffé-Test wird bei Kirk (1968, S. 113) als *„kritischer" F'*-Wert angegeben, der mit dem *„beobachteten" F*-Wert zu vergleichen ist. Der *kritische F'*-Wert wird definiert als

$$F' = \left[(k-1)F_{\alpha;\,k-1,\,N-k}\right], \text{ [Formel 12a]}$$

wobei k der Anzahl der Faktorstufen entspricht und $F_{\alpha;\,k-1,\,N-k}$ dem Tabellenwert der *F*-Verteilung. Die Berechnungsformel für die *F*-Statistik ist übrigens der Formel [4] sehr ähnlich (da $F = t^2$). Der *beobachtete F*-Wert ergibt sich als:

$$F = \frac{\left[C_j(\overline{A}_i) + C_{j'}(\overline{A}_{i'}) + \ldots + C_{j''}(\overline{A}_{i''})\right]^2}{MS_{error}\left(\frac{(C_j)^2}{n_j} + \frac{(C_{j'})^2}{n_{j'}} + \ldots + \frac{(C_{j''})^2}{n_{j''}}\right)} \text{ [Formel 12b]}$$

Leider gibt SPSS, wie erwähnt, diesen *F*-Wert für den Scheffé-Test nicht aus, sondern nur das entsprechende Konfidenzintervall. Rechentechnisch ergeben sich die Konfidenzintervalle für a posteriori Vergleiche nach dem Verfahren von Scheffé als:

$$\left|\overline{A}_j - \overline{A}_{j'}\right| \pm \sqrt{F'_{kritisch}}\sqrt{MS_{error}\left(\frac{(C_j)^2}{n_j} + \frac{(C_{j'})^2}{n_{j'}}\right)} \text{ [Formel 12c]}$$

Als Beispiel wird der erste Kontrast (Vergleich der Lernleistung für „sinnlose Silben“ (A_2) mit derjenigen für „sinnvolle Silben“ (A_1)) „nachgerechnet“:

$$F' = \left[(k-1)F_{\alpha;\,k-1,\,N-k}\right] = \left[(3-1)F_{.05;2,10}\right] = 2*4.1 = 8.2,$$

$$F = \frac{\left[C_j(\bar{A}_i) + C_{j'}(\bar{A}_{i'}) + \ldots + C_{j''}(\bar{A}_{i''})\right]^2}{MS_{error}\left(\frac{(C_j)^2}{n_j} + \frac{(C_{j'})^2}{n_{j'}} + \ldots + \frac{(C_{j''})^2}{n_{j''}}\right)} = \frac{\left[-1(6.0) + 1(2.0) + 0(4.0)\right]^2}{3.20\left(\frac{(-1)^2}{3} + \frac{(1)^2}{3} + \frac{(0)^2}{7}\right)} = \frac{16}{2.13} = 7.5$$

$$\left|\bar{A}_j - \bar{A}_{j'}\right| \pm \sqrt{F'_{kritisch}}\sqrt{MS_{error}\left(\frac{(C_j)^2}{n_j} + \frac{(C_{j'})^2}{n_{j'}}\right)} = -6.0 + 2.0 \pm \sqrt{8.2}\sqrt{3.20\left(\frac{(-1)^2}{3} + \frac{(1)^2}{3}\right)} = -4.0 \pm 4.18$$

Der kritische *F*-Wert ist mit *F'*=8.2 größer als der beobachtete *F*-Wert von *F*=7,5. Der erste Kontrast ist also nicht signifikant. Das 95%ige Konfidenzintervall für den Kontrast beinhaltet die „0“ (untere Grenze=0,18; obere Grenze=-8.18). Dieses Konfidenzintervall entspricht der SPSS-Ausgabe (vgl. SPSS-Output 4).

Abschließend sei nach angemerkt, daß aus der technischen SPSS-Begrenzung auf *k*-1-Faktorstufen simultan berechenbarer Kontraste, kein Problem entsteht, wie dies am Ende des Abschnitts 2.1.4 zur Berechnung von nicht-orthogonalen a priori Kontrasten geschildert wurde. Die jeweils im Output ausgegebenen Konfidenzintervalle sind im Falle von Scheffé immer für *alle überhaupt nur möglichen* Kontraste adjustiert.

2.2.2 Der Tukey Test

Es kann vorkommen, daß trotz einer Gesamtsignifikanz in der einfaktoriellen Varianzanalyse kein Paarvergleich nach dem Scheffé-Test signifikant wird. Der Grund hierfür ist darin zu sehen, daß das mathematische Rational, das dem Scheffé-Test zugrunde liegt, nicht nur von allen möglichen Paarvergleichen, sondern von *allen Einzelvergleichen überhaupt* (also auch von Vergleichen, die sich auf Kombinationen von Mittelwerten beziehen) ausgeht. Liegt eine Gesamtsignifikanz vor, muß allerdings mindestens einer der möglichen Einzelvergleiche, der jedoch kein Paarvergleich sein muß, signifikant sein (Bortz, 1985, S. 332).

Um diesen Nachteil des Scheffé-Tests zu umgehen, könnte man auf den Tukey-Test (Tukey, 1957), der auch unter dem Namen HSD-Test („honestly significant difference“) gekannt ist, zurückgreifen. Dieses a posteriori Verfahren adjustiert das α-Niveau nur für *alle paarweisen* Vergleiche zwischen den Mittelwerten und nicht für alle überhaupt nur denkbaren Einzelvergleiche. Die Anzahl aller paarweisen Vergleiche bestimmt sich nach der Anzahl der Faktorstufen *k* als *k*(*k*-1)/2. Bei einem Faktor mit 3 Faktorstufen ergeben sich beispielsweise drei paarweise Vergleiche.

Leider ist der Tukey-Test nicht in der MANOVA-Procedure von SPSS verfügbar (wohl aber als „post hoc“-Vergleich in der ONEWAY-Procedure). Dies ist liegt möglicherweise an der Testvoraussetzung, daß die Zellfrequenzen, *n*, der Faktorstufen gleich oder annähernd gleich sein sollten. Für den Fall, daß eine höhere Teststärke benötigt wird als sie der Scheffé-Test anbietet, sollte ein Tukey-Test gerechnet werden. Der Test kann sehr leicht mit den Angaben aus dem SPSS-Output „zu Fuß“ (also mit dem Taschenrechner) gerechnet werden. Aus dem SPSS-Output (vgl. SPSS-Output 1 oder 2) benötigt man die Mittelwerte der Faktorstufen, die

Schätzung des Fehlerterms MS_{error} und die zugehörigen υ-Freiheitsgrade des Fehlerterms MS_{error}. Nun benötigt man nur noch die Berechnungsformel von Tukey, in die man diese Werte einsetzt. Für den Tukey-Test berechnet man einen „*q*"-Wert (dies ist eine spezielle Verteilung, die sich aus der bekannten *t*-Verteilung ableitet). Der *q*-Wert berechnet sich als

$$q = \frac{C_j(\overline{A}_i) + C_{j'}(\overline{A}_{i'})}{\sqrt{MS_{error} / n}} \text{ [Formel 13]}$$

Diesen *beobachteten q*-Wert muß man nur noch mit dem *kritischen q'*-Wert der Tabelle (vgl. Anhang, Tabelle A.2) vergleichen (*r* in Tabelle A.2 entspricht der Anzahl der Faktorstufen). Ist $q > q'$, so ist der betrachtete paarweise Vergleich signifikant. Als Beispiel dient wieder der erste Kontrast (Vergleich der Lernleistung für „sinnlose Silben" (A_2) mit derjenigen für „sinnvolle Silben" (A_1)). Da das Berechnungsbeispiel ungleiche Zellfrequenzen enthält (vgl. Tabelle 3), muß zunächst das „harmonische" Mittel der Zellfrequenz bestimmt werden. Kirk (1968, S. 90) nennt dazu folgende einfache Formel:

$$\tilde{n} = \frac{k}{\left[(1/n_1) + (1/n_2) + \ldots + (1/n_j)\right]} \text{ [Formel 14]}$$

Da $\tilde{n} = \frac{3}{\left[(1/1) + (1/1) + (1/7)\right]} = 3{,}71$ ergibt sich der *q*-Wert als

$$q = \frac{C_j(\overline{A}_i) + C_{j'}(\overline{A}_{i'})}{\sqrt{MS_{error} / n}} = \frac{-1(6.0) + 1(2.0)}{\sqrt{3.20 / 3.71}} = |-4.31|$$ und der $q'_{\alpha,\upsilon}$-Wert als $q'_{.05,10}$=3.88.

Da $q > q'$ ist der erste Kontrast des Berechnungsbeispiels auf dem 5%-Niveau signifikant. Womit auch gleichzeitig gezeigt wird, das der Tukey-Test ein höhere Teststärke als der Scheffé-Test aufweist (der Scheffé-Test wurde für diesen Kontrast nicht signifikant, vgl. Abschnitt 2.2.1). Zum Abschluß der Ausführungen über den Tukey-Test sei noch kurz die Formel zur Berechnung der Konfidenzgrenzen genannt:

$$\left|\overline{A}_j - \overline{A}_{j'}\right| \pm q'_{\alpha,\upsilon}\sqrt{\frac{MS_{error}}{n}} \text{ [Formel 15].}$$

3 Haupteffektanalysen

...to be continued...

Literatur

Backhaus, K., Erichson, B., Plinke, W., Schuchard-Ficher, C. & Weiber, W. (1989). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Fünfte, revidierte Auflage. Berlin: Springer.

Bortz, (1985). *Lehrbuch der Statistik für Sozialwissenschaftler*. 2te Auflage. Berlin: Springer.

Dunn, O.J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, *56*, 52-64.

Dunn, O.J. & Massey, F.J., Jr. (1965). Estimation of multiple contrasts using *t*-distributions. *Journal of the American Statistical Association*, 60, 573-583.

Eimer, E. (1978). *Varianzanalyse*. Stuttgart: Kohlhammer.

Hand, D.J. & Taylor, C.C. (1986). *Multivariate analysis of variance and repeated measures*. London: Chapman and Hall.

Kirk, R.E. (1968). *Experimental design: Procedures for the behavioral sciences*. Belmont, CA: Brooks/Cole Publishing.

Miller, R.G. (1961). *Simultaneous statistical inference*. New York: McGraw-Hill.

Norusis, M.J. (1993). *SPSS for Windows: Advanced Statistics, Release 6.0*. Chicago, IL: SPSS Inc.

Perfomance Plus (1988). *Non-orthogonal contrasts in MANOVA repeated measures design* (S. 4-7).

Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika*, *40*, 87-104.

Scheffé, H. (1959). *The analysis of variance*. New York: Wiley & Sons.

Timm, N.H. (1975). *Multivariate analysis with applications in education and psychology*. Monterey, Calif.: Brooks/Cole.

Tukey, J.W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, *5*, 99-114.

Winer, B.J., Brown, D.R. & Michels, K.M. (1991). *Statistical principles in experimental design*. 3rd. ed. New York: McGraw-Hill.