

Eine inferentielle Clusteranalyse für Alternativdaten

VON HANS-PETER KRÜGER, GUSTAV A. LIENERT,
ALFRED GEBERT und ALEXANDER VON EYE, Nürnberg

Zusammenfassung, Summary, Résumé

Am Konzept der Axialsymmetrie multivariater Daten wird eine Methode entwickelt, mit der statistisch Cluster aus gleich interkorrelierenden Variablen gebildet werden können. Es wird eine heuristische Vorgehensweise empfohlen, die nicht wie die übliche deskriptive Clusteranalyse willkürliche Entscheidungen über Abbruchkriterien u.ä. verlangt. Die vorgeschlagene Methode ist insbesondere für mediandichotomierte kontinuierliche Variablen geeignet. Werden echte Alternativdaten verwendet, resultieren Cluster, die zusätzlich zur gleichen Interkorrelation Variablen mit gleicher Schwierigkeit aufweisen. Das Vorgehen wird an einem Beispiel illustriert.

An inferential cluster analysis for dichotomous data

By means of axial symmetry testing of multivariate data clusters may be defined statistically as equally intercorrelating variables (tests). A heuristic procedure is suggested which avoids arbitrary decisions as they are common in descriptive cluster analysis. The method is suitable for continuous variables dichotomized at their median. If applied to genuine binary variables the cluster analysis results in clusters of variables having equal Yes-No proportions as well as equal interassociations. In order to illustrate their approach the authors give a numerical example.

Une analyse de cluster interférentielle pour données alternatives

Sur le schéma de la symétrie axiale des données multivariées, est mise au point une méthode permettant de former par les voies statistiques des clusters résultant de variables à intercorrélation égale. On recommande un *modus procedendi* heuristique n'exigeant pas, comme l'analyse de cluster descriptive en usage, des décisions arbitraires par ex. sur les critères de rupture. La méthode proposée se prête en particulier au cas des variables continues à dichotomie médiane. Si les données sont véritablement alternatives, il en résulte des clusters présentant des variables non seulement d'intercorrélation égale, mais de difficulté égale. Un exemple sert d'illustration.

„ . . . clustering is vulnerable on two fronts; . . . ; the second, that the techniques themselves are not based on sound probability models and the results are poorly evaluated and unstable when evaluated.“ (HARTIGAN 1975,

p. 8). Auch in neueren Arbeiten werden Clusteranalysen meist als deskriptive Verfahren dargestellt und inferentielle Probleme bleiben weitgehend ausgeklammert; dies gilt sowohl für die Überblicksreferate (z.B. METZLER 1977) als auch für die Entwicklung neuer Verfahren (z.B. KLEITER und TIMMERMANN 1977, von EYE und WIRSING 1978). In den wenigen Fällen, in denen Wahrscheinlichkeitsmodelle zugrunde gelegt werden, handelt es sich meist um parametrische Restriktionen (zum Überblick siehe z.B. SINT 1978; BOCK 1974), die zudem nur selten zur Analyse und inferentiellen Qualifikation eingesetzt werden (wie bei EKLUND und WALLER 1976). Bei den wenigen kombinatorischen Ansätzen (z.B. FRANK 1978) ist noch mit numerischen Problemen zu rechnen.

Das Ziel der vorliegenden Arbeit ist es daher, ein einfaches Verfahren vorzustellen, das inferentiell qualifizierte Cluster liefert. Dazu wird das Konzept der Axialsymmetrie aufgegriffen, das von WALL (1976) in die Statistik eingeführt wurde.

In Abschnitt 1 dieser Arbeit soll daher das Konzept der Axialsymmetrie erläutert werden. Im Anschluß folgen die Beschreibung der Cluster-suchstrategie und ihre Anwendung auf ein empirisches Beispiel.

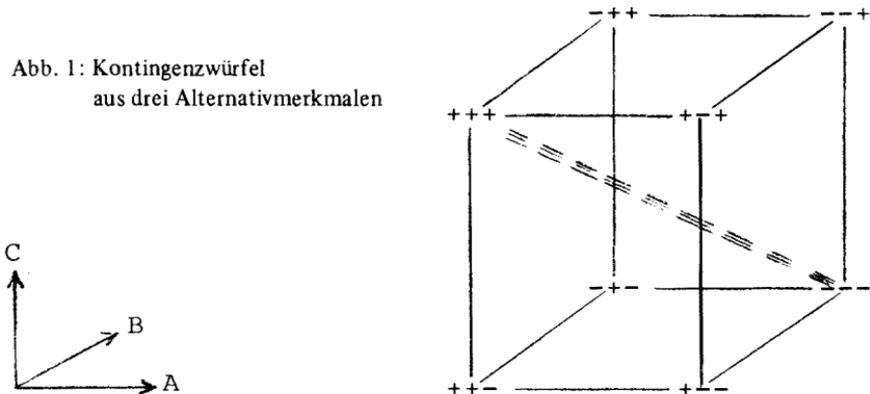
1. Das Konzept der Axialsymmetrie von t Variablen

Die Anwendung der hier vorgeschlagenen Clusteranalyse beruht auf zwei Voraussetzungen:

- a) Die t Variablen müssen identische univariate Randverteilungen mit gleichen Medianen haben (dies läßt sich durch Dichotomieren am Median erreichen);
- b) Die $t(t-1)/2$ bivariaten Randverteilungen müssen homogen und axial-symmetrisch sein.

Man kann sich das Prinzip der Axialsymmetrie an einem dreidimensionalen Kontingenzwürfel verdeutlichen (Abbildung 1),

Abb. 1: Kontingenzwürfel
aus drei Alternativmerkmalen



der bei drei Alternativvariablen A, B und C entsteht. Läßt man einen solchen Würfel um seine Raumdiagonale zwischen den Zellen (+ + +) und (---) rotieren, fallen die in Tabelle 1 aufgeführten Zellen zusammen. Es zeigt sich, daß diese zusammenfallenden Zellen jeweils die gleiche Zahl von Plus- (bzw. Minus-) Zeichen gemeinsam haben (z.B. + + -, und + - + und - + + mit jeweils $p = 2$ Plus). Die Zahl dieser axialsymmetrischen Zellen ergibt sich aus dem Binomialkoeffizienten mit $n = t$ und $i = p$ als Zahl der Plus. Für $t = 3$ gibt es

$$\binom{3}{3} = 1 \text{ Zelle mit } p = 3 \text{ Plus } (+ + +)$$

$$\binom{3}{2} = 3 \text{ Zellen mit } p = 2 \text{ Plus } (+ + -, + - +, - + +)$$

$$\binom{3}{1} = 3 \text{ Zellen mit } p = 1 \text{ Plus } (+ --, - + -, ---)$$

$$\binom{3}{0} = 1 \text{ Zelle mit } p = 0 \text{ Plus } (---)$$

Von Axialsymmetrie wird dann gesprochen, wenn die Felder mit gleichem p auch gleich besetzt sind. Unter der Nullhypothese axialsymmetrischer Felder läßt sich für jedes $p = p^0$ ($p^0 = 1, 2, \dots, t$) der Erwartungswert berechnen:

$$(1) \quad e_{p^0} = n_{p^0} / \binom{t}{p^0}$$

Für $t = 3$ und $p^0 = 2$ ergibt sich so, wie aus Tabelle 1 ersichtlich

$$e_{p^0=2} = (n_{++-} + n_{+-+} + n_{-++}) / 3$$

Damit kann auch eine X^2 -Komponente berechnet werden nach

$$(2) \quad X^2 = (n_{p^0} - e_{p^0})^2 / e_{p^0}$$

Wegen gleicher Erwartungswerte ergibt sich die X^2 -Komponente rechnerisch einfacher, wenn pro p mit den $m = \binom{t}{p}$ Zellfrequenzen n_i gerechnet wird

$$(3) \quad X^2 = \frac{m \sum n_i^2}{\sum n_i} - \sum n_i$$

Die Summe dieser X^2 über alle Zellen ist asymptotisch χ^2 -verteilt mit

$$df = \sum_{p=0}^t \binom{t}{p} - 1 = 2^t - (t + 1) \text{ Freiheitsgraden.}$$

Die Ablehnung von H_0 kann zwei Gründe haben:

1. entweder sind mindestens zwei Variablen unterschiedlich in ihrer Schwierigkeit,
2. oder mindestens zwei Paare von Variablen haben unterschiedliche Korrelationen.

Die Anwendung der Formeln auf den Fall von $t = 2$ Variablen zeigt,

Tabelle 1

Test auf Axialsymmetrie bei $t = 3$ Alternativmerkmalen

p = Zahl der Plus	bei der Rotation zusammenfallende Zellen	beobachtete Zellenfrequenzen	erwartete Zellenfrequenzen	χ^2
3	+++	n_{+++}	-	-
2	++- +-+ --+	n_{++-} n_{+-+} n_{--+} } $\Sigma = n_2$	$n_2 / 3$ $n_2 / 3$ $n_2 / 3$	χ^2_{++-} χ^2_{+-+} χ^2_{--+} } $\Sigma = \chi^2_{p=2}$
1	+-- -+- ---+	n_{+--} n_{-+-} n_{---+} } $\Sigma = n_1$	$n_1 / 3$ $n_1 / 3$ $n_1 / 3$	χ^2_{+--} χ^2_{-+-} χ^2_{---+} } $\Sigma = \chi^2_{p=1}$
0	---	n_{---}	-	-

$$\Sigma \chi^2 = \Sigma \chi^2_{p=2} + \Sigma \chi^2_{p=1}$$

daß die Axialsymmetrie lediglich eine Verallgemeinerung des McNEMAR-Tests (McNEMAR 1947) darstellt, wobei der multivariate Test sensu WALL auch auf Unterschiede in den Interkorrelationen der Variablen reagiert. Muß H_0 abgelehnt werden, kann die Asymmetrie so lokalisiert werden, daß die X^2 der einzelnen p 's nach $\binom{t}{p} - 1$ Freiheitsgraden bewertet werden (z.B. $X^2_{p=2}$ in Tabelle 2 nach 2 Freiheitsgraden).

2. Die Clusterbildung über sukzessive Prüfung auf Axialsymmetrie

Zur Bildung von axialsymmetrischen Clustern wird folgender Algorithmus vorgeschlagen:

1. Berechne alle $\binom{t}{2}$ Vierfelder-Korrelationskoeffizienten (z.B. phi-Koeffizienten) und wähle die beiden Variablen A und B aus, die die höchste Interkorrelation zeigen. (Auf der zweidimensionalen Stufe entfällt eine Prüfung auf Axialsymmetrie, da mediandichotomierte Variablen per definitionem axialsymmetrisch sind).
2. Bilde alle dreidimensionalen Randtafeln, die das Variablenpaar AB enthalten (z.B. ABC, ABD, ABE usw.) und prüfe diese Tafeln auf Axialsymmetrie. Aus den nicht signifikanten χ^2 dieser Tafeln wähle jenes χ^2 , das die geringste Abweichung von der Axialsymmetrie andeutet (minimales χ^2). Daraus ergebe sich z.B. das Tripel ABC.
3. Bilde alle vierdimensionalen Randtafeln, die das Variablen-tripel ABC enthalten (z.B. ABCD, ABCE, ABCF usw.) und prüfe wie unter 2.
4. Diese Prozedur endet dann, wenn alle höherdimensionalen Randtafeln signifikante χ^2 -Werte annehmen. Damit ist ein Cluster abgeschlossen.
5. Aus den restlichen Variablen wird auf der Basis der Interkorrelationsmatrix (siehe Schritt 1) analog verfahren. Sind in der Korrelationsmatrix keine signifikanten Korrelationen mehr zu finden oder ist nur eine Variable übriggeblieben, werden diese Restvariablen nicht in die Clusterung einbezogen.

Ist man lediglich an einem heuristischen Vorgehen interessiert, kann das Signifikanzniveau α für alle Einzeltests beibehalten werden. Anders beim inferentiellen Testen, wo α auf die Zahl r der simultanen Tests adjustiert werden muß, d.h. man wird ein adjustiertes $\alpha = \alpha^0 r$ setzen. Die Einzeltests müssen demnach zu extremeren α^0 durchgeführt werden, um insgesamt ein vorgegebenes Signifikanzniveau α einhalten zu können. Für den χ^2 -Test gilt die übliche Restriktion, daß die Erwartungswerte nicht zu klein ausfallen sollten. Da sie in der Regel bei der Prüfung auf Axialsymmetrie alle etwa gleich sind, sollten sie den Wert von 3 nicht unterschreiten (WISE 1963).

3. Das Vorgehen an einem Beispiel

GEBERT untersuchte eine alters- und vorbildungsheterogene Gruppe von $N = 997$ Pbn mit dem IST von AMTHAUER (1973). Die Scores der Pbn wurden für jede der $t = 9$ Skalen möglichst nah am Median dichotomiert (die Skalen: 1 = SE = Satzergänzen, 2 = WA = Wortauswahl, 3 = AN = Analogien, 4 = GE = Gemeinsamkeiten, 5 = RA = Rechenaufgaben, 6 = ZR = Zahlenreihen, 7 = FA = Figurenauswahl, 8 = WÜ = Würfelaufgaben und 9 = ME = Merkaufgaben). Die Häufigkeitsmatrix ist dem Anhang zu entnehmen. Die Buchstaben A bis I bezeichnen dabei die Untertests des IST in der genannten Reihenfolge, jeweils hinter einer Konfiguration ist die Häufigkeit ihres Auftretens verzeichnet. Über Vierfelderkoeffizienten wurde die Korrelationsmatrix der Tabelle 2 gebildet, in deren oberer Dreiecksmatrix die φ -Koeffizienten, in deren unteren Dreiecksmatrix die χ^2 der Axialsymmetrie der Paare zu finden sind¹. Es zeigen sich durchweg signifikante Korrelationen, die zum einen auf einen Generalfaktor hinweisen, zum andern auf die Heterogenität der Stichprobe zurückzuführen sind. Sodann wurden die $\sum_{p=3}^9 \binom{t}{p} = 466$ 3er, 4er, . . . , p-er Konfigurationen gebildet und auf Axialsymmetrie geprüft. Da die Untertests des IST relativ hoch interkorrelieren, zum andern der Stichprobenumfang mit $N = 997$ groß ist, wurde ein $\alpha = 0.01$ vereinbart. Dieses α ist auf die Zahl der $\sum_{p=2}^9 \binom{p}{p} = 502$ (inclusive der Korrelationen mit $p = 2$) zu adjustieren, so daß $\alpha^0 = 0.01/502 = 0.00002$. Die χ^2 -Schranken zu diesem α^0 ergeben sich für die Freiheitsgrade

df = 4:	$\chi^2 = 27$	für Tripel
df = 11:	$\chi^2 = 41,5$	für Quadrupel
df = 26:	$\chi^2 = 66.5$	für Quintupel
df = 57:	$\chi^2 = 111$	für Sextupel
df = 120:	$\chi^2 = 194.5$	für Septupel

Für Freiheitsgrade $df > 30$ kann die Normalapproximation der χ^2 -Verteilung mit $Z = \sqrt{2 \chi^2 - df} - 1$ benutzt werden, um auch die Schranken für unkonventionelle α bestimmen zu können.

Um die Berechnung der Axialsymmetrie zu verdeutlichen, wird diese in Tabelle 3 für die ersten 5 Subtests des IST durchgeführt.

¹ Bei exakt mediandichotomierten Daten müßten diese χ^2 alle Null sein. Da die Testskalen jedoch unimodale Häufigkeitsverteilungen über der Skalenmitte aufwiesen, war in einigen Fällen nur eine „möglichst gute“ Dichotomierung möglich.

Tabelle 2

Phi-Koeffizienten und Chi-Quadratwerte aus der Prüfung der bivariaten Axialsymmetrie für die 9 Untertests des IST 70

	SE	WA	AN	GE	RA	ZR	FA	WÜ	ME
1 SE	-	.417	.475	.459	.391	.350	.247	.171	.301
2 WA	2.505	-	.505	.440	.404	.401	.238	.138	.361
3 AN	1.237	0.238	-	.549	.497	.498	.292	.192	.386
4 GE	0.301	4.629	3.240	-	.440	.423	.265	.169	.407
5 RA	0.557	5.369	3.828	0.058	-	.492	.302	.193	.330
6 ZR	1.935	9.013	7.367	0.895	0.571	-	.286	.185	.443
7 FA	0.451	0.516	0.071	1.322	1.943	4.056	-	.262	.191
8 WÜ	0.293	0.595	0.122	0.966	1.433	3.192	0.011	-	.227
9 ME	0.000	2.285	1.059	0.275	0.508	2.256	0.419	0.314	-

In der oberen Dreiecksmatrix sind die phi-Koeffizienten, in der unteren Dreiecksmatrix die χ^2 des Tests auf Axialsymmetrie zu finden.

Tabelle 3

Häufigkeiten i und χ^2 -Komponenten des 5-Tupels 1 2 3 4 5

i	$\binom{5}{i}$	f_i	e_i	χ^2_i
5	1	212	-	-
4	5	42+28+14+28+34	29.2	14.41
3	10	11+12+6+9+12+9+17+18+6+13	11.3	13.11
2	10	21+14+13+11+15+10+18+15+10+12	13.9	8.12
1	5	36+41+18+23+34	30.4	12.01
0	1	235	-	-

$\chi^2 = 47.65$

Die erste Konfiguration ist + + + + + und enthält diejenigen Pbn, die in allen 5 Untertests supramediane Scores haben. Die Häufigkeit ergibt sich aus der Summation der ersten 16 Neuner-Konfigurationen des Anhangs. In all diesen 16 Konfigurationen tritt das 5-tupel + + + + + auf. Die Summe ergibt sich so zu $88 + 10 + 24 + 6 + 19 + 3 + 18 + 6 + 12 + 9 + 2 + 4 + 5 + 0 + 3 + 3 = 212$. Diese Zahl 212 ist in der ersten Zeile der Tabelle 3 unter $i = 5$ wiederzufinden. Analog mit der zweiten Konfiguration + + + + - verfahren ($i = 4$). Dazu werden die Konfigurationen 17 bis 32 des Anhangs zusammengezählt (Summe = 42). Ebenfalls ein $i = 4$ liefern aber die Konfigurationen + + + - + (Summe der Konfigurationen 33 bis 48 des Anhangs = 28), + + - + + (Konfigurationen 49–64 = 14), + - + + + (Konfigurationen 65 bis 80 = 28) und - + + + + (Konfigurationen 81 bis 96 = 34). Diese Häufigkeiten stehen unter $i = 4$ in der zweiten Zeile der Tabelle 3. Der Erwartungswert für die Häufigkeiten zu $i = 4$ ergibt sich als der Mittelwert der beteiligten Häufigkeiten, also:

$$e_{i=4} = (42 + 28 + 14 + 28 + 34)/5 = 29.2$$

Tabelle 4

Clustersuche bei den neun Subtests des IST 70

4a : Bildung von 3-Tupeln, df = 4				
Subtest-Kombinationen		X	$p(X_{\min}^2)^1$	
1	3 4	11.8	0.0189	
2	3 4	17.8		
	3 4 5	14.5		
	3 4 6	21.2		
	3 4 7	79.7 **		
	3 4 8	131.9 **		
	3 4 9	30.8 **		
4b : Bildung von 4-Tupeln, df = 11				
1	2 3 4	28.1	0.0031	
1	3 4 5	29.2		
1	3 4 6	40.6		
1	3 4 7	107.0 **		
1	3 4 8	191.6 **		
1	3 4 9	52.6 **		
4c : Bildung von 5-Tupeln, df = 26				
1	2 3 4 5	47.7	0.00585	
1	2 3 4 6	66.4		
1	2 3 4 7	149.0 **		
1	2 3 4 8	260.7 **		
1	2 3 4 9	74.0 **		
4 d : Bildung von 6-Tupeln, df = 57				
1	2 3 4 5 6	110.8		
1	2 3 4 5 7	186.6 **		
1	2 3 4 5 8	316.5 **		
1	2 3 4 5 9	110.6	0.000027	
4e : Bildung von 7-Tupeln, df = 120				
1	2 3 4 5 6 9	223,3 **		
1	2 3 4 5 7 9	254.9 **		
1	2 3 4 5 8 9	403.5 **		
4f : Zusammenfassung der restlichen Subtests				
	6 7 8	12.3	0.0154	

1 $\alpha^* = .01/502 = 0.0000199$

** = signifikante Abweichung von der Axialsymmetrie

Für jede der beteiligten Häufigkeiten wird nach $(b - e)^2/e$ eine χ^2 -Komponente berechnet, deren Summe 14.41 ebenfalls in Tabelle 3 zu finden ist. Nachdem für alle $i = 1, \dots, 5$ dieses χ^2 berechnet wurde, ergibt sich aus der Summe dieser χ^2 der Gesamtwert für die Axialsymmetrie.

Im folgenden werden alle 9 Subtests nach diesem Verfahren analysiert. Ausgangspunkt ist die höchste Interkorrelation in Tabelle 2, die zwischen den Skalen AN und GE besteht. In Tabelle 4 a stehen alle Tripel, die das Paar AN-GE enthalten.

χ^2 -Werte mit einem $p < \alpha^0$ sind gekennzeichnet. Am geringsten wird die Abweichung von der Axialsymmetrie bei Skala SE, so daß das Tripel SE-AN-GE den Ausgangspunkt für Tabelle 4b bildet. Dort kommt Skala WA hinzu. Schließlich werden in Tabelle 4 c die Skala RA, in Tabelle 4 d die Skala ME hinzugenommen. Tabelle 4 e zeigt, daß das Cluster damit abgeschlossen ist. Von den Restvariablen ZR, FA und WÜ haben ZR und FA die höchste Korrelation ($\phi = .286$ in Tabelle 2). Dieses Tripel ist zugleich axialsymmetrisch, wie das $\chi^2 = 12.3$ in Tabelle 4 f. zeigt.

Es entstehen so zwei Cluster:

- a) Cluster 1 aus den Skalen mit verbalem Material (die Rechenaufgaben sind als Textaufgaben formuliert) und
- b) Cluster 2 aus den Skalen ZR, FA und WÜ, die konstruktives, analytisches und theoretisches Denken mit Betonung der space-Komponente vereinigen.

Ein Blick in Tabelle 4 zeigt, daß nicht notwendig alle niederen Konfigurationen eines axialsymmetrischen Clusters ebenfalls axialsymmetrisch sein müssen. Die im Beispiel auftretenden Asymmetrien sind auf die Untertests 3 (= AN) und 9 (= ME) zurückzuführen: immer, wenn diese beiden in einem Cluster auftreten, ist die Axialsymmetrie zurückzuweisen. Erst im Sextupel des Clusters 1 werden beide Tests kompatibel. Ein solches Ergebnis ist ein aus Cluster- und Faktorenanalyse vertrautes Bild. Dabei liegt hier kein Artefakt vor, das etwa aus der „Verdünnung“ der Besetzungszahlen bei höherdimensionalen Tafeln resultierte (die Erwartungswerte liegen bei Cluster 1 einheitlich weit über 5). Eher ist an eine inhaltlich zu interpretierende Überlagerung von axialen Asymmetrien bei niedriger dimensionalen Tafeln zu denken.

Liegen wie im Beispiel alle $p = 2, 3, \dots, t$ -dimensionalen Konfigurationen vor, kann die Clusterung auch anders vorgenommen werden. Statt von den Paaren ($p = 2$) auszugehen, sucht man die Konfiguration höchster Dimension, die mit dem Konzept der Axialsymmetrie vereinbar ist (deszendierendes Verfahren). Tabelle 4 d zeigt, daß dies bei der 6er-Konfiguration

SE – WA – AN – GE – RA – ME

Tabelle 5

Subtesttupel für $n < 6$ aus dem ersten Subtestcluster

Subtest-Kombination	χ^2
1 2 3 4 5 .	47.7
1 2 3 4 . 9	74.0 *
1 2 3 . 5 9	62.5
1 2 . 4 5 9	49.1
1 . 3 4 5 9	73.9 *
. 2 3 4 5 9	69.3 *
1 2 3 4 . .	28.1
1 2 3 . 5 .	27.1
1 2 3 . . 9	42.9 *
1 2 . 4 5 .	15.0
1 2 . 4 . 9	32.5
1 2 . . 5 9	20.5
1 . 3 4 5 .	29.2
1 . 3 4 . 9	52.6 *
1 . 3 . 5 9	46.2 *
1 . . 4 5 9	29.6
. 2 3 4 5 .	31.9
. 2 3 4 . 9	46.3 *
. 2 3 . 5 9	44.6 *
. 2 . 4 5 9	21.7
. . 3 4 5 9	50.6 *
1 2 3 . . .	10.3
1 2 . 4 . .	6.3
1 2 . . 5 .	6.1
1 2 . . . 9	13.4
1 . 3 4 . .	11.8
1 . 3 . 5 .	15.3
1 . 3 . . 9	26.1
1 . . 4 5 .	5.0
1 . . 4 . 9	21.6
1 . . . 5 9	7.1
. 2 3 4 . .	17.8
. 2 3 . 5 .	19.0
. 2 3 . . 9	22.7
. 2 . 4 5 .	8.5
. 2 . 4 . 9	10.0
. 2 . . 5 9	9.5
. . 3 4 5 .	15.5
. . 3 4 . 9	30.8 *
. . 3 . 5 9	27.1 *
. . . 4 5 9	11.2

* bedeutet signifikante Abweichung von der Axialsymmetrie

gegeben ist, die wir oben als Cluster 1 definiert hatten. In Tabelle 5 sind alle Tafeln niedrigerer Dimension aufgeführt, die sich aus diesem Cluster ergeben. Die signifikant von der Axialsymmetrie abweichenden Cluster sind mit einem Stern versehen. Wieder zeigt sich, daß diese Abweichung immer dann auftritt, wenn die Skalen AN und ME zusammentreffen.

4. Diskussion

Die nach dem vorgeschlagenen Verfahren gebildeten Cluster erfüllen auf multivariater Ebene die Voraussetzungen der klassischen Clusteranalyse: die Variablen des ersten Clusters sind untereinander höher interkorreliert als zu den Variablen des zweiten Clusters usf. (TRYON 1939; BOLSHEV 1970; KENDALL 1970).

Verwendet man, wie im vorangegangenen Abschnitt beschrieben wurde, das deszendierende Verfahren, so ergibt sich der Vorteil, daß auf das Konzept der höchsten Korrelation verzichtet werden kann. Dennoch ist der Rechenaufwand erheblich höher, so daß die Wahl des Verfahrens von der Art und Menge des zu bearbeitenden Datenmaterials geleitet werden muß.

Ein weiterer Vorteil des deszendierenden Verfahrens kann darin gesehen werden, daß die Wahrscheinlichkeit geringer ist, Entscheidungen darüber treffen zu müssen, welche Elemente fusioniert werden sollen. Dieses Problem tritt immer dann auf, wenn die niedrigsten χ^2 -Werte im Betrag gleich sind (In Tabelle 4 sind die Sechstupel 1 - 5 + 6 und 1 - 5 + 3 nahezu gleich). In diesem Fall kann z.B. jede der entstandenen Möglichkeiten durchgerechnet werden, wobei die Entscheidung dann auf die Lösung fällt, die in der Folge zu den kleineren χ^2 -Werten führt. Ähnliche Aufgaben werden auch bei anderen agglomerativen Verfahren angewendet, in denen es zu analogen Situationen kommen kann.

In der praktischen Anwendung auf echte (nicht dichotomierte) Alternativvariablen (Ja-Nein-Items) sind die eindimensionalen Randsummen in der Regel nicht homogen. Wird das Verfahren auf solche Datensätze angewendet, resultieren Cluster, die nicht nur gleich interkorreliert sind, sondern auch gleiche Itemschwierigkeiten haben. Liegen mehrstufige Variablen vor (wie etwa psychiatrische Ratings), sollten diese möglichst nah am Median dichotomiert werden. Prinzipiell ist eine Erweiterung dieses Ansatzes auf multinäre Items möglich. In praxi bedeutet das jedoch ein sehr schnelles Anwachsen der Möglichkeiten und damit, unter den Restriktionen des χ^2 -Tests, die Notwendigkeit sehr großer Stichproben.

Ähnlich wie die Faktorenanalyse kann die Clusteranalyse auch auf die gestürzte Datenmatrix angewendet werden. Dann werden die N Pbn

so gruppiert, daß Cluster von Pbn entstehen, die eine ähnliche Merkmalstruktur haben. In diesem Fall kann allerdings nicht mehr von der Unabhängigkeit der Merkmale ausgegangen werden (anders als bei den N unabhängigen Pbn der nicht-invertierten Clusteranalyse). Damit entfällt auch eine inferentielle Prüfung. Heuristisch ist ein solches Verfahren aber immer dann anzuwenden, wenn die Zahl der Pbn relativ gering gegenüber der Zahl der Merkmale ist.

Anhang

Häufigkeiten der nach Mediandichotomierung entstandenen Konfigurationen der 9 Subtests des IST (A = Satzergänzungen, B = Wortauswahl, C = Analogien, D = Gemeinsamkeiten, E = Rechenaufgaben, F = Zahlenreihen, G = Figurenauswahl, H = Würfelaufgaben, I = Merkaufgaben; + = supramediante, - = submediante Scores).

Siehe nebenstehende Falttafel.

Literatur

- AMTHAUER, R.: Intelligenz-Struktur-Test. 4. Auflage. Göttingen: Hogrefe, 1973.
- BOCK, H. H.: Automatische Klassifikation. Göttingen: Vandenhoeck & Ruprecht, 1974.
- BOLSHEV, L. N.: Cluster analysis. London: Proc. Int. Inst. Stat., 1970.
- EKLUND, D. L. and R. A. WALLER: Inference for cluster sampling. *Biometrical Journal* 18 (1976) 193–204.
- FRANK, O.: Inferences concerning cluster structure. In: CORSTEN L. C. A. (Ed.): *Compstat 1978: Proceedings in computational statistics*. Wien: Physica Verlag, 1978, p. 259–265.
- HARTIGAN, J. A.: *Clustering Algorithms*. New York: John Wiley, 1975.
- KENDALL, M. G.: Cluster Analysis. In: WATANABE, S. (Ed.): *Frontiers of pattern analysis*. New York: Academic Press, 1970, p. 291–307.
- KLEITER, E. F. und G. TIMMERMANN: Über einen Algorithmus zur Hierarchischen Voraussetzungs-Struktur-Analyse. *Psychologische Beiträge* 19 (1977) 355–390.
- McNEMAR, Q.: Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 12 (1947) 153–157.
- METZLER, P.: Clusteranalyse – Einführung, Methoden, Anwendungsbeispiele. *Probleme und Ergebnisse der Psychologie* 63 (1977) 19–36.

- SINT, P. P.: Cluster Analysis, an Introduction. In: SKARABIS, H. and SINT, P. P. (Ed.). Compstat lectures 1. Würzburg: Physica Verlag, 1978, p. 46–56.
- TYRON, R. C.: Cluster analysis. Ann Arbor: Edwards Brothers, 1939.
- von EYE, A. and M. WIRSING: An attempt for a mathematical foundation and evaluation of Macs, a method for multidimensional automatical cluster detection. Biometrical Journal 20 (1978) 671–682.
- WALL, K. D.: Ein Test auf Symmetrie in einer j-dimensionalen Kontingenztafel. EDV in Medizin und Biologie 7 (1976) 57–64.
- WISE, U. E.: Multinomial probabilities and the χ^2 and the X^2 distributions. Biometrika 50 (1963) 145.

Dr. H.-P. Krüger
Prof. Dr. Dr. G. A. Lienert
Dr. A. Gebert
Dr. A. von Eye
Universität Erlangen-Nürnberg
FB 11 – Psychologie
Regensburger Str. 160
8500 Nürnberg

ABCDEFGHI	N	ABCDEFGHI	N	ABCDEFGHI	N	ABCDEFGHI	N	ABCDEFGHI	N	ABCDEFGHI	N	ABCDEFGHI	N	ABCDEFGHI	N
++++++	88	+++	1	+++++	7	+-+	0	-++++	10	-++++	2	-++++	4	-++++	0
++++++	10	+++++	1	+++++	3	+++++	2	+++++	3	+++++	1	+++++	0	+++++	2
++++++	24	+++	0	+++++	2	+-+	1	-++++	7	-++++	0	-++++	1	-++++	1
++++++	6	+++	1	+++++	3	+++++	1	+++++	2	+++++	0	+++++	1	+++++	1
++++++	19	+++++	4	+++++	3	+++++	0	+++++	0	+++++	0	+++++	0	+++++	0
++++++	3	+++++	0	+++++	0	+++++	2	+++++	0	+++++	0	+++++	0	+++++	0
++++++	18	+++++	2	+++++	1	+++++	0	+++++	4	+++++	0	+++++	2	+++++	1
++++++	6	+++++	1	+++++	1	+++++	0	+++++	2	+++++	0	+++++	0	+++++	0
++++++	12	+++++	0	+++++	5	+++++	0	+++++	1	+++++	0	+++++	1	+++++	1
++++++	9	+++++	1	+++++	0	+++++	1	+++++	3	+++++	1	+++++	1	+++++	0
++++++	2	+++++	1	+++++	1	+++++	0	+++++	0	+++++	0	+++++	0	+++++	0
++++++	4	+++++	1	+++++	0	+++++	0	+++++	0	+++++	1	+++++	1	+++++	0
++++++	5	+++++	0	+++++	0	+++++	0	+++++	1	+++++	0	+++++	0	+++++	3
++++++	0	+++++	0	+++++	0	+++++	1	+++++	0	+++++	0	+++++	0	+++++	1
++++++	3	+++++	0	+++++	0	+++++	0	+++++	0	+++++	1	+++++	1	+++++	2
++++++	3	+++++	1	+++++	2	+++++	1	+++++	1	+++++	0	+++++	1	+++++	0
++++++	6	+++++	0	+++++	0	+++++	0	+++++	1	+++++	2	+++++	3	+++++	0
++++++	1	+++++	1	+++++	1	+++++	0	+++++	1	+++++	0	+++++	0	+++++	1
++++++	5	+++++	2	+++++	1	+++++	1	+++++	3	+++++	0	+++++	0	+++++	0
++++++	2	+++++	0	+++++	0	+++++	0	+++++	0	+++++	1	+++++	1	+++++	0
++++++	5	+++++	0	+++++	0	+++++	1	+++++	2	+++++	1	+++++	2	+++++	2
++++++	0	+++++	0	+++++	1	+++++	0	+++++	0	+++++	0	+++++	0	+++++	0
++++++	5	+++++	1	+++++	1	+++++	0	+++++	3	+++++	1	+++++	1	+++++	0
++++++	1	+++++	0	+++++	1	+++++	0	+++++	1	+++++	0	+++++	1	+++++	1
++++++	2	+++++	2	+++++	2	+++++	0	+++++	0	+++++	0	+++++	1	+++++	0
++++++	2	+++++	2	+++++	1	+++++	2	+++++	1	+++++	0	+++++	0	+++++	0
++++++	1	+++++	2	+++++	0	+++++	3	+++++	1	+++++	0	+++++	0	+++++	2
++++++	4	+++++	0	+++++	0	+++++	1	+++++	0	+++++	0	+++++	0	+++++	3
++++++	3	+++++	0	+++++	0	+++++	2	+++++	1	+++++	2	+++++	0	+++++	1
++++++	0	+++++	1	+++++	0	+++++	1	+++++	0	+++++	0	+++++	0	+++++	3
++++++	1	+++++	0	+++++	1	+++++	2	+++++	0	+++++	2	+++++	2	+++++	2
++++++	4	+++++	1	+++++	0	+++++	0	+++++	3	+++++	1	+++++	4	+++++	8
++++++	8	+++++	1	+++++	3	+++++	0	+++++	6	+++++	2	+++++	1	+++++	2
++++++	1	+++++	0	+++++	2	+++++	1	+++++	3	+++++	0	+++++	0	+++++	0
++++++	2	+++++	0	+++++	0	+++++	0	+++++	2	+++++	2	+++++	1	+++++	4
++++++	0	+++++	0	+++++	0	+++++	3	+++++	0	+++++	2	+++++	2	+++++	1
++++++	2	+++++	0	+++++	0	+++++	1	+++++	0	+++++	3	+++++	0	+++++	2
++++++	2	+++++	0	+++++	1	+++++	1	+++++	0	+++++	0	+++++	1	+++++	0
++++++	3	+++++	0	+++++	1	+++++	0	+++++	2	+++++	1	+++++	0	+++++	0
++++++	2	+++++	0	+++++	1	+++++	1	+++++	1	+++++	2	+++++	1	+++++	2
++++++	2	+++++	2	+++++	0	+++++	1	+++++	0	+++++	1	+++++	0	+++++	0
++++++	1	+++++	2	+++++	0	+++++	0	+++++	0	+++++	0	+++++	2	+++++	3
++++++	1	+++++	0	+++++	0	+++++	0	+++++	0	+++++	0	+++++	0	+++++	1
++++++	0	+++++	0	+++++	0	+++++	0	+++++	0	+++++	2	+++++	1	+++++	2
++++++	1	+++++	0	+++++	0	+++++	0	+++++	1	+++++	0	+++++	0	+++++	2
++++++	1	+++++	1	+++++	3	+++++	3	+++++	2	+++++	1	+++++	1	+++++	12
++++++	1	+++++	0	+++++	2	+++++	1	+++++	2	+++++	2	+++++	0	+++++	3
++++++	0	+++++	0	+++++	1	+++++	0	+++++	2	+++++	0	+++++	0	+++++	4
++++++	0	+++++	0	+++++	0	+++++	1	+++++	0	+++++	0	+++++	1	+++++	0
++++++	0	+++++	0	+++++	0	+++++	1	+++++	1	+++++	1	+++++	1	+++++	3
++++++	1	+++++	0	+++++	0	+++++	2	+++++	0	+++++	1	+++++	0	+++++	2
++++++	0	+++++	0	+++++	0	+++++	1	+++++	0	+++++	1	+++++	0	+++++	2
++++++	0	+++++	0	+++++	0	+++++	2	+++++	0	+++++	1	+++++	0	+++++	1
++++++	2	+++++	3	+++++	0	+++++	1	+++++	0	+++++	0	+++++	1	+++++	1
++++++	3	+++++	1	+++++	0	+++++	1	+++++	0	+++++	0	+++++	3	+++++	2
++++++	0	+++++	0	+++++	0	+++++	0	+++++	0	+++++	1	+++++	1	+++++	3
++++++	0	+++++	0	+++++	0	+++++	2	+++++	0	+++++	1	+++++	0	+++++	2
++++++	0	+++++	0	+++++	0	+++++	1	+++++	0	+++++	2	+++++	0	+++++	1
++++++	2	+++++	3	+++++	0	+++++	1	+++++	0	+++++	0	+++++	1	+++++	1
++++++	3	+++++	1	+++++	0	+++++	1	+++++	0	+++++	0	+++++	3	+++++	2
++++++	0	+++++	0	+++++	0	+++++	0	+++++	0	+++++	1	+++++	1	+++++	3
++++++	0	+++++	0	+++++	0	+++++	0	+++++	0	+++++	1	+++++	0	+++++	2
++++++	0	+++++	2	+++++	2	+++++	4	+++++	4	+++++	4	+++++	4	+++++	43
++++++	0	+++++	3	+++++	1	+++++	3	+++++	0	+++++	5	+++++	1	+++++	11
++++++	1	+++++	3	+++++	0	+++++	11	+++++	5	+++++	14	+++++	1	+++++	99