

Die komplementäre Anwendung zweier Verfahren: Korrespondenzanalyse und Konfigurationsfrequenzanalyse

Blasius, Jörg; Lautsch, Erwin

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Blasius, J., & Lautsch, E. (1990). Die komplementäre Anwendung zweier Verfahren: Korrespondenzanalyse und Konfigurationsfrequenzanalyse. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 27, 110-133. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-202527>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Die komplementäre Anwendung zweier Verfahren: Korrespondenzanalyse und Konfigurationsfrequenzanalyse

von Jörg Blasius und Erwin Lautsch*

1. Einleitung

Nicht nur in den Sozialwissenschaften besteht ein Streit zwischen explorativem und konfirmatorischem Vorgehen bei der quantitativen Datenanalyse. Während dem einen das explorative Vorgehen als wahllos erscheint, da Zusammenhänge lediglich beschrieben, nicht aber erklärt werden können, hält der andere die Genauigkeit der mittels konfirmatorischer Verfahren gewonnenen Ergebnisse für wenig realistisch. Während der explorativ auswertende Forscher nicht selten dem Verdacht ausgesetzt ist, "keine Theorie zu haben", wird dem konfirmatorisch auswertenden häufig ein "fitting the data" vorgeworfen.

Ein weiteres Problem in den Sozialwissenschaften ist, daß die vorhandenen Daten relativ selten metrisches Meßniveau haben. Dieses wird aber für die Anwendung bei der Mehrzahl der multivariaten Analyseverfahren vorausgesetzt. Als Ausweg für diese Diskrepanz zwischen dem Meßniveau der vorhandenen Daten und den Anforderungen der Verfahren werden häufig Metrisierungen vorgenommen bzw. nicht-metrische Daten per Definition als metrisch erklärt. Im ersten Fall werden oft Skalen konstruiert, bei denen die Ursprungsdaten lediglich als Durchschnittswerte vorkommen und daher nur noch schwer interpretierbar sind, im zweiten Fall sind dagegen Fehler aufgrund der ungenauen Analyseform vorprogrammiert.

In dem vorliegenden Artikel werden zwei multivariate Verfahren vorgestellt, bei denen es keine Voraussetzungen bezüglich des Skalenniveaus gibt. Bei den Verfahren handelt es sich um die Ende der 60er Jahre von *Lienert* entwickelte Konfigurationsfrequenzanalyse (KEA) (*Lienert* 1969, *Krauth und Lienert* 1973, *Lienert* 1988, *Lautsch und von Weber* 1990, *von Eye* 1990) und um die erst in jüngster Zeit viel diskutierte Korrespondenzanalyse (*Greenacre* 1984, *Lebart et al.* 1984), die in den Sozialwissenschaften mit Ausnahme von französischsprachigen Publikationen bislang selten verwendet wurde (s. *z.B. Bourdieu* 1982, 1988, *Blasius* 1987, *Blasius und Winkler* 1989, *Dangschat und Blasius* 1987, *Thiessen und Rohlinger* 1988).

Während die Konfigurationsfrequenzanalyse in die Gruppe der konfirmatorischen Verfahren einzuordnen ist, zählt die Korrespondenzanalyse zu den explorativen Techniken. In diesem Beitrag sollen beide Verfahren kurz vorgestellt und Schnittpunkte der komplementären Verwendung aufgezeigt werden.

*Dr. Erwin Lautsch ist Priv. Doz. am Institut für Rechtswissenschaften, Otto-Nuschke-Str. 10/11, 1086 Berlin.

2. Daten

Datengrundlage ist eine schriftliche Befragung von 567 Alkoholikern des Hamburger Universitätskrankenhauses Eppendorf (UKE), die sich zwischen 1980 und 1985 beraten ließen und sich über eine Therapie informierten. Mittels dieser Befragung sollte herausgefunden werden, ob es eine bestimmte Alkoholikerpopulation gibt, deren Mitglieder überdurchschnittlich oft an einer Therapie teilnahmen und durch welche Indikatoren diese charakterisiert werden kann. Ferner sollten jene Personen beschrieben werden, die übermäßig oft nicht zur Therapie erschienen bzw. die lediglich an einer Vorbereitungsgruppe teilnahmen.

Als beschreibende Merkmale der drei genannten Personengruppen wurden verwendet: das Eingeständnis von Alkoholproblemen ("Werde damit nicht alleine fertig": stimmt; stimmt etwas; stimmt nicht), das Drängen des Partners/der Familie und der Rat des Arztes zur Therapieaufnahme (beide Variablen ebenfalls in trichotomisierter Form), der geplante Alkoholgenuß nach einer Therapie (nie wieder Alkohol; Jahre kein Alkohol; moderates Trinken), Angaben über Magenerkrankungen (nein, keine Magenerkrankung; ja, ohne Krankenhausaufenthalt; ja, mit Krankenhausaufenthalt), Kontakt zu Selbsthilfegruppen (nein; gelegentlich; regelmäßig), der Erwerbsstatus (berufstätig; arbeitslos; Hausfrau/Rentner) und der Familienstand (ledig; verheiratet; getrennt-lebend; geschieden; verwitwet).

In dem genannten Zeitraum nahmen 135 der 567 Personen des Erstgesprächs (23.8%) an einer Therapie teil. Von den verbleibenden nahmen 187 (33.0%) lediglich an der Vorbereitungsgruppe teil, die anderen 245 Personen (43.2%) kamen nicht wieder.

3. Korrespondenzanalyse

Die Korrespondenzanalyse wurde unter verschiedenen Bezeichnungen in der statistischen Literatur eingeführt. In der bekanntesten und hier vorzustellenden Form wurde sie in Frankreich entwickelt (*Benzécri et al.* 1973), sie gilt dort als sozialwissenschaftliche Standardmethode. Seit 1976 gibt es mit "Les Cahiers de l'Analyse des Données" eine Zeitschrift, die nahezu ausschließlich Artikel über die Korrespondenzanalyse (incl. zahlreicher Anwendungsbeispiele) enthält. Trotz relativ früher Schriften von *Benzécri* und anderen französischen Statistikern sowie zahlreicher sozialwissenschaftlicher Anwendungen, insbesondere von französischen Sozialwissenschaftlern, wurde die Korrespondenzanalyse außerhalb von Frankreich erst wahrgenommen, als englischsprachige Literatur (insbesondere die Lehrbücher von *Greenacre* (1984) und *Lebart et al.* (1984)) zur Verfügung stand. In den nächsten Jahren dürfte die Korrespondenzanalyse aufgrund der visuellen Anschaulichkeit ihrer grafischen Ausgabe und nicht zuletzt auch durch die Implementation entsprechender Prozeduren

in die neuesten Versionen der großen Statistikpakete (SAS, SPSS, BMDP) auch außerhalb von Frankreich zunehmend Anwendung finden.

Bei der Korrespondenzanalyse wird in "multiple" und "einfache" unterschieden, obwohl beide Verfahren letztlich auf dem gleichen Algorithmus basieren (vgl. *Greenacre* 1984, S. 126 ff.). Bei der einfachen Korrespondenzanalyse werden - mit einer für alle Tabellen identischen zu beschreibenden Variable - die einzelnen Kreuztabellen untereinander geschrieben. Durch diese Art der Dateneingabe gibt es keine Probleme mit der Häufigkeit der Zellenbesetzung, wie dies z.B. bei der log-linearen Analyse oder der KFA der Fall ist. Die Eingabe bei der multiplen Korrespondenzanalyse kann sowohl in Form einer Burt-Matrix erfolgen, also in Form einer symmetrischen Matrix, bei der jede Variable mit jeder anderen kreuztabelliert und die Tabellen spalten- und zeilenweise miteinander verknüpft werden, als auch in Form einer Indikatormatrix. Letztere Eingabeform wird insbesondere dann verwendet, wenn einzelne Analyseeinheiten, in der Regel Individuen, in einem mehrdimensionalen Raum lokalisiert und beschrieben werden sollen, so z.B. von *Bourdieu* (1988) zur Beschreibung von Hochschullehrern im "Raum der Fakultäten". Die Verwendung von Burt-Matrizes ist relativ selten; aufgrund der Duplizität von Zeilen- und Spaltenausprägungen gibt es weder eine (mehrere) zu beschreibende Variable(n) noch Individuen (oder andere Analyseeinheiten), die anhand von Merkmalen beschrieben werden sollen.

Die Korrespondenzanalyse ist ein exploratives Analyseverfahren, statistische Signifikanztests wie bei der log-linearen Analyse oder der KFA gibt es nicht. Die Interpretation der Ergebnisse erfolgt deskriptiv - von daher sind den Einsatzmöglichkeiten der Korrespondenzanalyse Grenzen gesetzt. Sinnvoll ist ihr Einsatz aber immer dann, wenn Strukturen in mehrdimensionalen Daten gefunden werden sollen, also als hypothesengenerierendes Verfahren. Ferner kann die Korrespondenzanalyse zur vertiefenden Beschreibung von inferenzstatistisch gewonnenen Erkenntnissen verwendet werden (vgl. z.B. *van der Heijden et al.* 1989).

Die Korrespondenzanalyse ist primär eine Technik zur grafischen Darstellung von Spalten und Zeilen einer (mehrerer) Kontingenztabelle(n). Hierbei gibt es zwar eine eindeutige Distanzfunktion zwischen den Zeilen und zwischen den Spalten, nicht aber zwischen Spalten **und** Zeilen (vgl. *Greenacre und Hastie* 1987, *Greenacre* 1989). Wie bei der Hauptkomponentenanalyse gibt es einen Satz von orthogonalen Vektoren, die einen niederdimensionalen Raum (maximal $(r-1)$ mit $r \leq c$, wobei r =Anzahl der Spalten und c =Anzahl der Zeilen) aufspannen. Neben der grafischen Darstellung der Ergebnisse gibt es eine numerische Ausgabe, mittels derer eine weitergehende Interpretation der Daten möglich ist. Mit Hilfe der dazugehörigen Koeffizienten ist es möglich, Zeilen und Spalten der zu analysierenden Kontingenztabelle(n) zu vergleichen.



4. Konfigurationsfrequenzanalyse (KFA)

Die KFA wurde von *Lienert* (1969) ursprünglich als Klassifikationsmethode entwickelt. Sie ist eine unter mehreren Zugängen, die zu einer bestimmten Typologie von Individuen führt, wobei die Besonderheit der KFA in der inferenzstatistischen Anwendungsmöglichkeit besteht.

Im klassischen Modell der KFA sollen Individuen mit identischen Merkmalsausprägungen (dargestellt durch die einzelnen Zellen einer mehrdimensionalen Kontingenztabelle) als "typische Klasse" mittels einer statistischen Testentscheidung identifiziert werden. So kann z.B. geprüft werden, ob Personen mit dem Merkmalsprofil "geplanter zukünftiger Alkoholgenuß: Abstinenz", "der Partner/die Familie drängt stark zur Therapie" und "Magenleiden ja, aber ohne Krankenhausaufenthalt" eine statistisch bedeutsame Klasse manifestieren. In dem genannten Beispiel wäre eine dreidimensionale Kontingenztabelle zu konstruieren, deren Zellen aus allen möglichen Merkmalskombinationen der drei Variablen (also 27) bestehen.

Entsprechend dem Modell der KFA ist eine Personengruppe mit identischen Merkmalsausprägungen "statistisch bedeutsam", wenn unter der Annahme der statistischen Unabhängigkeit der Variablen die beobachtete Zellenhäufigkeit statistisch signifikant von der erwarteten abweicht. Für die konkrete Testentscheidung gibt es mehrere Testprozeduren, von denen in den meisten Fällen die nach Lehmann (mit Stetigkeitskorrektur und Alpha-Adjustierung) zu empfehlen ist (*Lautsch und von Weber* 1990). Merkmalskombinationen, die mit dem Modell der statistischen Unabhängigkeit nicht vereinbar sind, werden als "Typen" bzw. als "Antitypen" bezeichnet. Wir sprechen von einem "Typ", wenn die beobachtete Zellenhäufigkeit größer als die der erwarteten ist, und von einem "Antityp", wenn die beobachtete Häufigkeit kleiner als die erwartete ist.

Der Unterschied zur log-linearen Analyse besteht darin, daß die KFA ausschließlich zellenorientiert ist, d.h. es werden Merkmalskombinationen gesucht, die unter den o.g. Annahmen als Ausreißer interpretiert werden können. Eine Zerlegung der Varianz in Haupteffekte und Interaktionseffekte unterschiedlicher Ordnung sowie die Prüfung bestimmter Modellannahmen sind bei der standardmäßigen Anwendung der KFA nicht vorgesehen. Mittels Weiterentwicklungen der KFA ist es inzwischen möglich, auch spezifische Klassifikationsprobleme zu lösen (z.B. durch die hierarchische KFA, die Mehrstichproben-KFA, die Prädiktions-KFA). In empirischen Anwendungen werden diese Verfahren in der Regel als nicht-parametrische Pendant klassischer multivariater Verfahren verwendet (vgl. *Lienert* 1988, *Lautsch und von Weber* 1990, *von Eye* 1990).

Um zu erklären, welche der drei Therapiekategorien mit welchen beschreibenden Merkmalskombinationen einhergehen, verwenden wir ein modifiziertes Modell - die (multiple) Prädiktions-KFA (vgl. *Lienert* 1988). Die hierfür zu konstruierende Kontingenztafel hat zwei Dimensionen: Die erste ist durch die drei Merkmalsausprägungen der abhängigen Größe (Therapiepartizipation) gegeben, die zweite durch die Merkmalsausprägungskombinationen der unabhängigen (der beschreibenden) Variablen. Ein "Prädiktionstyp" bzw. ein "Prädiktionsantityp" liegt vor, wenn eine Kombination von Merkmalsausprägungen der unabhängigen Variablen mit einer Ausprägung der abhängigen Variable signifikant häufiger bzw. signifikant seltener als erwartet (unter der Annahme der statistischen Unabhängigkeit) vorkommt. Während beim KFA-Basismodell alle untersuchten Merkmale gleichberechtigt in die Analyse eingehen (symmetrisches Modell), wird bei der Prädiktions-KFA nach zu erklärenden (abhängigen) und erklärenden (unabhängigen) Merkmalen (asymmetrisches Modell) unterschieden. Zellenweise getestet wird bei der Prädiktions-KFA die Nullhypothese, daß zwischen dem zu erklärenden Merkmal "Therapiepartizipation" und den erklärenden Merkmalen (den Merkmalskombinationen derselben) kein Zusammenhang besteht.

Mit steigender Anzahl der beschreibenden Variablen steigt die Anzahl der Kategorien der zu bildenden Merkmalskombinationen (die resultierende Variable wird auch als "synthetische Variable" bezeichnet) schnell an. So gibt es bei 5 Merkmalen mit je 3 Ausprägungen 243 mögliche Merkmalskombinationen - eine (Prädiktions-)KFA wäre damit nur noch bei einer sehr hohen Fallzahl durchführbar. Aus diesem Grund muß die Anzahl der beschreibenden Merkmale auf jene Variablen reduziert werden, die wesentlich zur Erklärung des abhängigen Merkmals beitragen. Ein mögliches Hilfsmittel bei der Suche der "wesentlichen Variablen" ist die Korrespondenzanalyse.

5. Verbindung von KFA und Korrespondenzanalyse

Soll einerseits nicht auf die Anschaulichkeit der Korrespondenzanalyse verzichtet werden, und soll andererseits eine Interpretation im inferenzstatistischen Sinn möglich sein, so können beide Verfahren ergänzend verwendet werden. Ein analoges Vorgehen wählen u.a. *van der Heijden und de Leeuw* (1985), *Goodman* (1986) und *van der Heijden et al.* (1989) bei ihrer komplementären Verwendung von log-linearen Modellen und der Korrespondenzanalyse.

Bei einer kombinierten Anwendung eines explorativen und eines konfirmatorischen Verfahrens kann - sofern die (wenigen) relevanten Variablen bekannt sind - im ersten Schritt das inferenzstatistische Verfahren (entsprechend der Fragestellung z.B. die log-lineare Analyse) zur Verifizierung von Interaktionen zwischen den Variablen verwendet werden. Im zweiten

Schritt kann dann die Korrespondenzanalyse auf zweifache Art eingesetzt werden: Zum einen zur explorativen Beschreibung von Variablenzusammenhängen, die zwar theoretisch relevant sind, aber im Test nicht passen, und zum anderen zur Interpretation von Interaktionsparametern, die nur schwach signifikant sind.

Eine weitere Möglichkeit der gemeinsamen Verwendung der Verfahren besteht darin, im ersten Schritt mittels der Korrespondenzanalyse nicht-triviale Strukturen in mehrdimensionalen Daten zu explorieren. Im zweiten Schritt können die gefundenen Zusammenhänge (anhand eines weiteren Datensatzes) auf Signifikanz getestet werden.

Im folgenden wird zunächst mittels der Korrespondenzanalyse die zu beschreibende Variable "Therapieerfolg" mit den drei Ausprägungen "nur am Erstgespräch teilgenommen", "nur an der Vorbereitungsgruppe teilgenommen" und "Therapieerfolgreicher" mit Hilfe der im Abschnitt "Daten" genannten Variablen beschrieben. Daran anschließend werden die explorativ gefundenen bedeutsamen unabhängigen Variablen mit der Variable "Therapieerfolg" kreuztabelliert und geprüft, ob es "Prädiktionstypen" oder "Prädiktionsantitypen" im Sinne der Prädiktions-KFA gibt. Im letzten Schritt werden die nachgewiesenen "Prädiktionstypen" und "Prädiktionsantitypen" zur besseren Anschaulichkeit, und zur Darstellung der (Un-)Ähnlichkeiten zwischen ihnen, mittels der Korrespondenzanalyse in einem zweidimensionalen Unterraum abgebildet.

6. Erster Schritt: Ergebnisse der ersten Korrespondenzanalyse

Eingangsmaterial der Korrespondenzanalyse sind die zeilenweise untereinander geschriebenen absoluten Häufigkeiten der Kreuztabellen der Variablen "Therapieerfolg" mit den im Abschnitt 2 aufgeführten Variablen.

**Tabelle 1: Eingangsdaten der ersten Korrespondenzanalyse**

	nur Erst- gespräch	nur Vb.- gruppe	Therapie- teilnahme
Berufstätig	108	84	58
Arbeitslos	110	75	61
Hausfrau/Rentner	26	26	15
Keine Alkoholprobleme	37	19	3
Etwas Alkoholprobleme	53	36	15
Starke Alkoholprobleme	130	122	116
Familie drängt nicht	97	57	55
Familie drängt etwas	49	48	29
Familie drängt stark	70	66	42
Arzt drängt nicht	61	38	24
Arzt drängt etwas	46	30	12
Arzt drängt stark	117	106	95
Abstinenz	158	132	109
Jahre ohne Alkohol	25	19	14
Moderates Trinken	55	31	11
Kein Magenleiden	152	114	67
Magenleiden ohne K.haus	27	26	41
Magenleiden mit K.haus	26	24	11
Selbsthilfegruppe nie	128	83	56
Selbsthilfegruppe geleg.	78	68	44
Selbsthilfegruppe regel.	36	25	32
Ledig	73	50	34
Verheiratet	88	63	46
Getrennt-lebend	22	25	11
Geschieden	56	47	40
Verwitwet	6	2	4

Für eine erste Bewertung der in den Daten vorhandenen Varianz kann die Chi-Quadrat-Statistik herangezogen werden. In Tabelle 2 sind die einzelnen quadrierten Abweichungen der empirischen von den erwarteten Werten, dividiert durch die jeweiligen Erwartungswerte, aufgeführt. Die Summe der in Tabelle 2 aufgeführten Werte ist der Gesamt-Chi-Quadrat-Wert



($\chi^2=110.5$ mit $df=50$). Dieser Wert wird durch das Gesamt-N (das N von Tabelle 1) dividiert (der resultierende Wert wird als "total inertia" bezeichnet; total inertia=0.0257), der resultierende Wert entspricht der Spur der zu zerlegenden Residualmatrix und damit der Summe der Eigenwerte, die zu den den Projektionsraum aufspannenden Eigenvektoren gehört (vgl. *Blasius und Rohlinger* 1989, S. 100ff.).

Tabelle 2: Chi-Quadrat-Statistik

	nur Erst- gespräch	nur Vb.- gruppe	Therapie- teilnahme
Berufstätig	0.01	0.03	0.13
Arbeitslos	0.23	0.46	0.02
Hausfrau/Rentner	0.24	0.69	0.10
Keine Alkoholprobleme	5.53	0.01	8.98
Etwas Alkoholprobleme	1.66	0.09	4.20
Starke Alkoholprobleme	4.69	0.00	7.82
Familie drängt nicht	0.67	2.06	0.34
Familie drängt etwas	0.43	1.00	0.09
Familie drängt stark	0.47	0.91	0.04
Arzt drängt nicht	1.37	0.16	1.17
Arzt drängt etwas	1.89	0.03	4.14
Arzt drängt stark	2.60	0.01	4.02
Abstinenz	0.90	0.00	1.46
Jahre ohne Alkohol	0.00	0.00	0.00
Moderates Trinken	4.45	0.03	6.73
Kein Magenleiden	0.68	0.16	2.43
Magenleiden ohne K.haus	4.30	0.80	14.37
Magenleiden mit K.haus	0.00	0.75	0.99
Selbsthilfegruppe nie	1.72	0.29	1.24
Selbsthilfegruppe geleg.	0.12	0.46	0.11
Selbsthilfegruppe regel.	0.35	1.05	3.88
Ledig	0.53	0.06	0.46
Verheiratet	0.18	0.06	0.08
Getrennt-lebend	0.31	1.81	0.69
Geschieden	0.42	0.00	0.78
Verwitwet	0.15	0.97	0.40

Anhand von Tabelle 2 wird ersichtlich, daß die größten Abweichungen von empirischen und erwarteten Werten in der Spalte "Therapie Teilnehmer" mit den Variablenausprägungen "Magenleiden, aber ohne Krankenhausaufenthalt", "keine Alkoholprobleme" und "starke Alkoholprobleme" sind. Die Zellen dieser drei Variablenkombinationen weisen somit die größte Varianz auf. Dementgegen hat die Ausprägung "Jahre ohne Alkohol" keine Varianz in der zu erklärenden Variable - die unter der Annahme der statistischen Unabhängigkeit berechneten Erwartungswerte sind gleich denen der beobachteten. Schon an dieser Stelle wird deutlich, welche Merkmalskombinationen miteinander korrelieren (positiv, wenn der Erwartungswert kleiner dem empirischen ist oder negativ, wenn der Erwartungswert größer als der empirische ist).

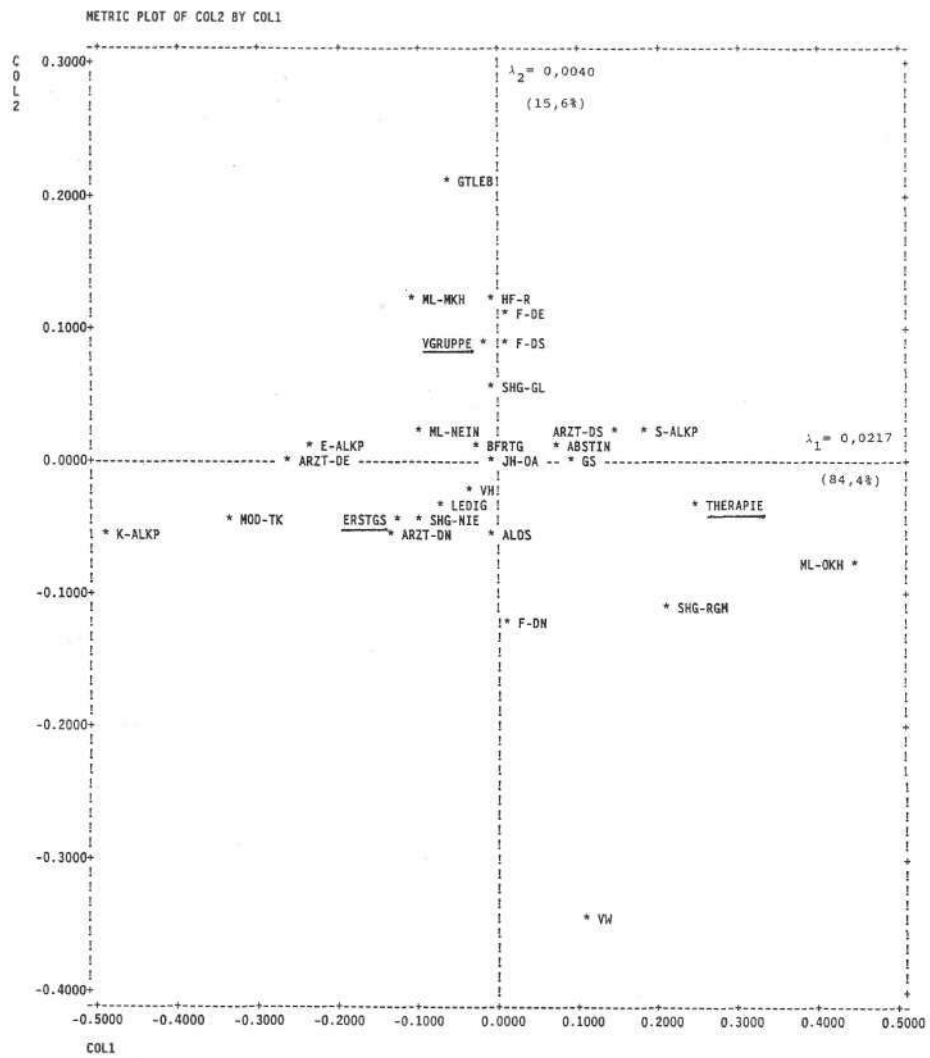
Neben den bereits erwähnten Korrelationen des Merkmales "Therapie Teilnehmer" dürfte es Zusammenhänge zwischen dem zu erklärenden Merkmal "nur an der Vorbereitungsgruppe teilgenommen" und den beschreibenden Merkmalen "Familie drängt nicht", "Familie drängt etwas", "Familie drängt stark", "getrennt-lebend", "gelegentliche Teilnahme an einer Selbsthilfegruppe" und "verwitwet" geben. Im Vergleich zu den Abweichungen in der Spalte "Therapie Teilnehmer" sind diese jedoch wesentlich niedriger, d.h. der Anteil der zu erklärenden Gesamtvarianz ist wesentlich geringer. In der noch durchzuführenden Korrespondenzanalyse wird sich dies dahingehend widerspiegeln, daß die Ausprägung "nur an der Vorbereitungsgruppe teilgenommen" gegenüber der Ausprägung "Therapie Teilnehmer" genau den Anteil zur Determination des Darstellungsraumes beiträgt, der dem Verhältnis der beiden spaltenweise auf summierten Abweichungen entspricht (64.67 zu 11.89).

Abbildung 1: Legende

ERSTGS	nur Erstgespräch	ARZT-DS	Arzt drängt stark
VGRUPPE	nur Vorbereitungsgruppe	ABSTIN	Abstinenz
THERAPIE	Therapie Teilnehmer	JH-OA	Jahre ohne Alkohol
		MOD-TK	Moderates Trinken
BFRTG	Berufstätig	ML-NEIN	Kein Magenleiden
ALOS	Arbeitslos	ML-OKH	Magenleiden ohne Knaus
HF-R	Hausfrau/Rentner	ML-MKH	Magenleiden mit Knaus
K-ALKP	Keine Alkoholprobleme	SHG-NJE	Selbsthilfegruppe nie
E-ALKP	Etwas Alkoholprobleme	SHG-GL	Selbsthilfegruppe geleg.
S-ALKP	Starke Alkoholprobleme	SHG-RGM	Selbsthilfegruppe regel.
F-DN	Familie drängt nicht	LEDIG	Ledig
F-DE	Familie drängt etwas	VH	Verheiratet
F-DS	Familie drängt stark	GTLEB	Getrennt-lebend
ARZT-DN	Arzt drängt nicht	GS	Geschieden
ARZT-DE	Arzt drängt etwas	VW	Verwitwet



Abbildung 1: Grafische Darstellung der ersten Korrespondenzanalyse*



*Zur besseren Übersicht wurden die Ausprägungen der zu erklärenden Variablen "Therapie" unterstrichen.

Wie aus Abbildung 1 ersichtlich wird, bilden auf der Ebene der Variablen "Therapie Teilnehmer" die Teilnehmer (THERAPIE) und die Nichtteilnehmer (ERSTGS) die gegensätzlichen Positionen auf der ersten Achse, diese erklärt 84.4% der Gesamtvariation des Modells. Die zweite Achse erklärt die verbleibenden 15.6% der Gesamtvariation des Modells. Sie ist auf der Ebene der Spaltenvariablen durch die Ausprägung "nur zur Vorbereitungsgruppe erschienen" gekennzeichnet.

Dem positiven Abschnitt der ersten Achse, also dem Abschnitt, dem auch die Ausprägung "Therapie Teilnehmer" zuzuordnen ist, lassen sich die beschreibenden Merkmale "Magenleiden ja, aber ohne Krankenhausaufenthalt" (ML-OKH), "geschieden" (GS) und "zukünftiges Trinkverhalten: Abstinenz" (ABSTIN) zuordnen. Im gleichen Achsenabschnitt lassen sich dem zu beschreibenden Merkmal "nur am Erstgespräch teilgenommen" folgende beschreibende Merkmale zurechnen: Arzt drängt etwas (ARZT-DE), etwas Alkoholprobleme (E-ALKP), keine Teilnahme an Selbsthilfegruppen (SHG-NIE), die Familienstände "ledig" (LEDIG) und "verheiratet" (VH), zukünftig moderates Trinkverhalten (MOD-TK) und keine Alkoholprobleme (K-ALKP). Dem positiven Abschnitt der zweiten Achse, also dem gleichen Abschnitt wie dem Merkmal "nur an der Vorbereitungsgruppe teilgenommen", lassen sich die Variablenausprägungen "Hausfrau/Rentner" (HF-R), "gelegentliche Teilnahme an Selbsthilfegruppen" (SHG-GL) sowie die beiden Merkmale "Familie/Partner drängt etwas" (F-DE) und "Familie/Partner drängt stark" (F-DS) zuordnen.

Wie bereits erwähnt, wurde bislang noch keine eindeutige Distanzfunktion gefunden, mittels derer es möglich ist, Zeilen- und Spaltenausprägungen direkt miteinander zu vergleichen. Somit sollte die grafische Darstellung vorsichtig interpretiert werden. Um (genauere) Aussagen über die (Un-)Ähnlichkeit von Spalten- und Zeilenausprägungen machen zu können, kann ein Vergleich über die Ähnlichkeit der Winkel zum Schwerpunkt (dargestellt durch das Achsenkreuz) erfolgen. Diese Winkel zum Achsenkreuz, oder genauer deren quadrierte Kosinus, sind neben anderen Koeffizienten in der numerischen Ausgabe der Korrespondenzanalyse enthalten (s. Tabelle 3). In den Spalten 5 und 8 (QCOR1, QCOR2) stehen die quadrierten Korrelationen (die quadrierten Kosinus der Winkel zwischen den Eigenvektoren und den vom Schwerpunkt ausgehenden Vektoren der Merkmale) der Variablenausprägungen von Zeilen und Spalten mit den beiden Achsen. Die Zeilensummen der quadrierten Korrelationen der berücksichtigten Achsen sind in der zweiten Spalte (SQCOR) aufgeführt. Die Komponenten von SQCOR sind analog den Kommunalitäten der Hauptkomponentenanalyse interpretierbar, die von QCOR1 und QCOR2 denen der dazugehörigen (quadrierten) Faktorladungen. Da in unserem Beispiel nur zwei Dimensionen existieren, ergeben die in SQCOR aufgeführten Zeilensummen immer Eins. Die Spalten QCOR1 und QCOR2 sind die wichtigsten Interpretationshilfen, da hier abzulesen ist, auf welcher Achse welche Variablenausprägungen positiv oder negativ "laden" (das Vorzeichen ist den Spalten LOCI und LOC2 zu entnehmen).



Tabelle 3: Numerische Ausgabe der ersten Korrespondenzanalyse

GENSTAT	MASS	SQCOR	INR	LOCI	QCOR1	INR1	LOC2	QCOR2	INR2
ERSTGS	0.427	1.000	0.307	-0.127	0.870	0.316	-0.049	0.130	0.257
VGRUPPE	0.330	1.000	0.108	-0.017	0.036	0.005	0.090	0.964	0.666
THERAPIE	0.243	1.000	0.585	0.246	0.979	0.679	-0.036	0.021	0.078
BFRTG	0.058	1.000	0.002	-0.024	0.813	0.002	0.011	0.187	0.002
ALOS	0.057	1.000	0.006	-0.007	0.015	0.000	-0.054	0.985	0.041
HF-R	0.016	1.000	0.009	-0.006	0.002	0.000	0.124	0.998	0.060
K-ALKP	0.014	1.000	0.131	-0.493	0.987	0.154	-0.057	0.013	0.011
E-ALKP	0.024	1.000	0.054	-0.239	0.996	0.063	0.015	0.004	0.001
S-ALKP	0.086	1.000	0.113	0.183	0.989	0.133	0.019	0.011	0.008
F-DN	0.049	1.000	0.028	0.008	0.004	0.000	-0.121	0.996	0.177
F-DE	0.029	1.000	0.014	0.005	0.002	0.000	0.110	0.998	0.088
F-DS	0.041	1.000	0.013	0.012	0.018	0.000	0.089	0.982	0.081
ARZT-DN	0.029	1.000	0.024	-0.137	0.859	0.025	-0.056	0.141	0.022
ARZT-DE	0.020	1.000	0.055	-0.262	1.000	0.065	0.002	0.000	0.000
ARZT-DS	0.074	1.000	0.060	0.143	0.981	0.070	0.020	0.019	0.007
ABSTIN	0.093	1.000	0.021	0.076	0.987	0.025	0.009	0.013	0.002
JH-OA	0.014	1.000	0.000	-0.006	0.622	0.000	-0.005	0.378	0.000
MOD-TK	0.023	1.000	0.101	-0.336	0.979	0.118	-0.050	0.021	0.014
ML-NEIN	0.078	1.000	0.030	-0.097	0.964	0.034	0.019	0.036	0.007
ML-OKH	0.022	1.000	0.176	0.449	0.972	0.203	-0.076	0.028	0.031
ML-MKH	0.014	1.000	0.016	-0.112	0.439	0.008	0.127	0.561	0.057
SHG-NIE	0.062	1.000	0.029	-0.099	0.807	0.028	-0.048	0.193	0.036
SHG-GL	0.044	1.000	0.006	-0.009	0.021	0.000	0.059	0.979	0.039
SHG-RGM	0.022	1.000	0.048	0.210	0.777	0.044	-0.112	0.223	0.068
LEDIG	0.037	1.000	0.010	-0.076	0.863	0.010	-0.030	0.137	0.008
VH	0.046	1.000	0.003	-0.032	0.646	0.002	-0.024	0.354	0.006
GTLEB	0.014	1.000	0.025	-0.061	0.076	0.002	0.211	0.924	0.150
GS	0.033	1.000	0.011	0.091	0.996	0.013	0.005	0.004	0.000
VW	0.003	1.000	0.014	0.107	0.090	0.001	-0.339	0.910	0.080



In der ersten Spalte (MASS) ist die relative Masse der einzelnen Variablenausprägungen angegeben, deren Werte für Zeilen und Spalten getrennt auf "Eins" normiert wurden. Hier ist z.B. abzulesen, daß die "Therapie Teilnehmer" die Masse .243 haben, d.h. ihr Anteil an allen Nennungen der berücksichtigten Variablen beträgt 24.3%. Zusätzlich zu den analogen Koeffizienten der Hauptkomponentenanalyse gibt es bei der Korrespondenzanalyse das "Trägheitsgewicht" (INR1, INR2). Dieses wird berechnet aus dem Quadrat der Entfernung der Projektion auf der Achse zum Achsenkreuz, multipliziert mit der Masse. Anhand der Spalten INR1 und INR2 wird ersichtlich, wie stark die einzelnen Achsen von den einzelnen Variablenausprägungen, getrennt für Zeilen und Spalten auf Eins normiert, determiniert werden. Die gleiche Interpretation, bezogen auf das Gesamtmodell, gilt auch für das Gesamtträgheitsgewicht (INR). Die Komponenten dieser Spalte ergeben sich aus den mit den erklärten Varianzen der Eigenwerte multiplizierten "Trägheitsgewichten" der einzelnen Achsen. Die Werte dieser Spalte entsprechen auch dem (normierten) Verhältnis der zeilen- bzw. spaltenweise aufsummierten Werte der Chi-Quadrat-Statistik (vgl. Tabelle 2).

Aus Tabelle 3 wird ersichtlich, daß - wie schon anhand der grafischen Darstellung zu erkennen war - mit der ersten Achse die Variablenausprägungen "Therapie Teilnehmer" (QCOR1=.979) und "nur am Erstgespräch teilgenommen" (QCOR1=.870) mit unterschiedlichen Vorzeichen (vgl. Spalte LOCI) korrelieren; d.h. beide Merkmale korrelieren hinsichtlich der verwendeten beschreibenden Variablen negativ. Ferner sind sie, entsprechend dem Modell der Korrespondenzanalyse, mit jenen Merkmalen unkorreliert, die auf einer anderen Achse laden.

Aufgespannt wird der Raum auf der Ebene der zu beschreibenden Variable insbesondere durch die Ausprägung "Therapie Teilnehmer" auf der ersten Achse (INR1=.679) - diese Variablenausprägung erklärt auf der Ebene der Spalten 67.9% der Gesamtvariation der ersten Achse sowie durch die Variablenausprägung "nur an der Vorbereitungsgruppe teilgenommen" auf der zweiten Achse (INR2=.666). Auf der Ebene der beschreibenden Variablen wird die geometrische Ausrichtung der ersten Achse durch die Ausprägungen "habe keine Alkoholprobleme" (INR 1= .154), "habe starke Alkoholprobleme" (INR 1= .133), "zukünftiges Trinkverhalten: moderat" (INR 1= .118) und "Magenleiden ja, aber ohne Krankenhausaufenthalt" (INR 1 =.203) determiniert. Die zweite Achse wird auf der Ebene der beschreibenden Variablen insbesondere durch die drei Ausprägungen der Variablen "Familie/ Partner drängt zur Therapie" (diese drei Ausprägungen erklären zusammen 34.6% der geometrischen Ausrichtung der zweiten Achse) und durch das Merkmal "getrennt- lebend" (INR2=.150) erklärt.

Mit dem positiven Abschnitt der ersten Achse - also mit dem Abschnitt, mit dem die Gruppe der Therapie Teilnehmer beschrieben werden kann - gibt es quadrierte Korrelationen von über .90 mit den Merkmalen "habe starke Alkoholprobleme", "Arzt drängt stark", "zukünftiges

Trinkverhalten: Abstinenz", "Magenleiden ja, aber ohne Krankenhausaufenthalt" und "geschieden". Mit dem negativen Abschnitt der ersten Achse, also dem Abschnitt, auf dem auch die Variablenausprägung "nur am Erstgespräch teilgenommen" läßt, gibt es quadrierte Korrelationen von über .90 mit den Variablenausprägungen "habe keine Alkoholprobleme", "habe etwas Alkoholprobleme", "Arzt drängt nicht", "Arzt drängt etwas", "zukünftiges Trinkverhalten: moderat" und "kein Magenleiden". Diese zuletzt genannten Merkmale sind mit dem Merkmal "Therapie Teilnehmer" negativ korreliert, d.h. diese Merkmale treffen für diese Gruppe von Personen überdurchschnittlich selten zu.

Entsprechend dem Modell der Korrespondenzanalyse sind die Achsen unkorreliert. Zusammen mit dem beschreibenden Merkmal "nur an der Vorbereitungsgruppe teilgenommen" korrelieren folgende Variablenausprägungen mit dem positiven Abschnitt der zweiten Achse: "Hausfrau/Rentner", "Familie/Partner drängt etwas", "Familie/Partner drängt stark", "gelegentliche Teilnahme an Treffen von Selbsthilfegruppen" und "getrennt-lebend". Negativ mit den genannten Merkmalen, also mit dem negativen Abschnitt der zweiten Achse, korrelieren die Variablenausprägungen "arbeitslos", "die Familie/der Partner drängt nicht" und "verwitwet".

Diesen Ergebnissen zufolge sind akute Magenleiden, die zum Befragungszeitpunkt noch nicht im Krankenhaus behandelt wurden, der wichtigste Indikator zur Therapie Teilnahme. Ebenfalls einen hohen positiven Einfluß auf die Therapie Teilnahme hat die Einsicht, daß "starke Alkoholprobleme" bestehen. Wird von dem Patienten jedoch angegeben, daß "keine Alkoholprobleme bestehen" bzw. daß sein zukünftiges Trinkziel "moderates Trinken" ist, so ist die Wahrscheinlichkeit hoch, daß er nach dem Erstgespräch nicht wiederkommt. Ferner hat das "Drängen der Familie/des Partners" keinen Einfluß auf die Entscheidung für oder gegen eine Therapie Teilnahme, es führt lediglich dazu, daß der Patient überdurchschnittlich häufig ausschließlich an der Vorbereitungsgruppe teilnimmt.

Die Variable, deren Ausprägungen alle mit der ersten Achse hoch korrelieren, ist "Alkoholprobleme". Zudem hat sie mit insgesamt 35.3% den von allen Variablen größten Anteil an der geometrischen Ausrichtung der ersten Achse. Analoges gilt in bezug auf die zweite Achse für die Variable "die Familie/der Partner drängt". Somit scheinen beide Variablen gut für die Beschreibung der Variable "Therapie Teilnahme" geeignet zu sein.

Im folgenden Abschnitt soll die Frage beantwortet werden, ob den Therapie Teilnehmern, denjenigen, die nur die Vorbereitungsgruppe besuchten und denjenigen, die nur zum Erstgespräch kamen, statistisch signifikante unterschiedliche Merkmalskombinationen der beiden ausgewählten Variablen zugewiesen werden können, die diese im Sinne der Prädiktions-KFA als "Prädiktionstypen" oder als "Prädiktionsantitypen" kennzeichnen.

7. Zweiter Schritt: Konfigurationsfrequenzanalyse

Zur Durchführung der Prädiktions-KFA wird eine zweidimensionale Kreuztabelle mit der abhängigen Merkmalsgröße "Therapiepartizipation" einerseits und der aus den Merkmalen "Eingeständnis von Alkoholproblemen" und "Familie/Partner drängt zur Therapiepartizipation" kombinierten Variablen andererseits gebildet¹.

Tabelle 4: Eingabedaten der Prädiktions-KFA

Alkoholprobleme ¹	Familie drängt ²	nur Erstgespräch	nur Vb.-gruppe	Therapiepartizipation
1	1	21	5	3
1	2	9	4	0
1	3	7	10	0
2	1	18	10	5
2	2	12	9	1
2	3	17	16	9
3	1	57	41	47
3	2	27	34	28
3	3	42	40	33

¹ (1) keine Alkoholprobleme, (2) etwas Alkoholprobleme, (3) starke Alkoholprobleme

² (1) drängt nicht, (2) drängt etwas, (3) drängt stark

Die in Tabelle 4 aufgeführten absoluten Häufigkeiten sind das Eingangsmaterial für die Prädiktions-KFA. Für alle Zellen der Kontingenztabelle wird als Testgröße Lehmachers asymptotischer hypergeometrischer Komponententest mit der Kontinuitätskorrektur nach Küchenhoff verwendet (s. *Lautsch und von Weber* 1990, S. 60ff.). Die Testentscheidungen werden zum einen explorativ (zweiseitige Testentscheidung ohne Alpha-Adjustierung) und zum anderen inferenzstatistisch (einseitige Testentscheidung mit Alpha-Adjustierung nach Holms) durchgeführt² (s. *Lautsch und von Weber* 1990, S. 69f.). Der kritische Wert liegt beim zweiseitigen Test bei $u=1.96$ (5%-Niveau), beim einseitigen Test mit Alpha-Adjustierung nach Holms gibt es keinen konstanten Wert.

- 1 Damit die Unabhängigkeit der Ergebnisse, die mit den unterschiedlichen Verfahren erzielt wurden, gewährleistet ist, hätte vor Beginn der Datenauswertung die Stichprobe gesplittet werden müssen. Da wir aber lediglich eine Auswertungsstrategie angeben und zudem das Leistungsvermögen beider Verfahren diskutieren wollen, wurde auf ein Splitten der Stichprobe verzichtet
- 2 Als Programm zur KFA wurde SICFA verwendet (*Lautsch und von Weber* 1990).



Tabelle 5: Ergebnisse der Prädiktions-KFA

Alkohol- probleme ¹	Familie drängt ²	Therapie- teilnahme ³	emp. Wert	erw. Wert	Lehmacher-Test ⁴	
					ohne α -Adj.	mit α -Adj.
1	1	1	21	12.1	2.80 (1)	3.27 (1)
1	1	2	5	9.7	1.99(-1)	1.70
1	1	3	3	7.2	1.97(-1)	1.65
1	2	1	9	5.4	1.94	1.76
1	2	2	4	4.4	1.28	.00
1	2	3	0	3.2	2.04(-1)	1.78
1	3	1	7	7.1	1.42	.00
1	3	2	10	5.7	2.11 (1)	1.99
1	3	3	0	4.2	2.44(-1)	2.13
2	1	1	18	13.7	1.93	1.38
2	1	2	10	11.0	1.54	.21
2	1	3	5	8.2	1.70	1.14
2	2	1	12	9.1	1.75	1.04
2	2	2	9	7.4	1.57	.53
2	2	3	1	5.5	2.23(-1)	2.01
2	3	1	17	17.5	1.64	.00
2	3	2	16	14.1	1.68	.49
2	3	3	9	10.5	1.54	.36
3	1	1	57	60.3	1.93	.56
3	1	2	41	48.5	2.11(-1)	1.46
3	1	3	47	36.2	2.49 (1)	2.34
3	2	1	27	37.0	2.40(-1)	2.25
3	2	2	34	29.8	1.92	.92
3	2	3	28	22.2	2.06 (1)	1.43
3	3	1	42	47.8	2.00(-1)	1.14
3	3	2	40	38.5	1.83	.23
3	3	3	33	28.7	1.92	.93

¹ (1) keine Alkoholprobleme, (2) etwas Alkoholprobleme, (3) starke Alkoholprobleme

² (1) drängt nicht, (2) drängt etwas, (3) drängt stark

³ (1) nur Erstgespräch, (2) nur an Vorbereitungsgruppe teilgenommen, (3) Therapieteilnahme

⁴ (1) "Prädiktionstyp", (-1) "Prädiktionsantityp"

Gesamt- $\chi^2=43.5$ mit $df=16$

Aus Tabelle 5 wird ersichtlich, daß es bei inferenzstatistischen Testentscheidungen (mit Alpha-Adjustierung) nur einen "Prädiktionstyp" gibt. Dieser ist dadurch charakterisiert, daß die zusammengesetzte Variablenausprägung "keine Alkoholprobleme, die Familie/der Partner drängt nicht zur Therapieteilnahme" einen signifikanten Einfluß auf die Variablenausprägung "nur am Erstgespräch teilgenommen" hat. Weitere "Prädiktionstypen" sowie auch "Prädikationsantypen" sind lediglich tendenziell, d.h. wenn auf die Alpha-Adjustierung verzichtet wird, vorhanden. Demnach käme es relativ oft nur zur ausschließlichen Teilnahme am Erstgespräch, wenn angegeben wurde, daß keine Alkoholprobleme bestehen und daß die Familie/der Partner nur etwas zur Therapie drängte. Relativ häufig zur Teilnahme an der Vorbereitungsgruppe, aber nicht an der Therapie, käme es, wenn die Familie/der Partner zwar stark zur Therapieteilnahme drängte, aber gleichzeitig "keine Alkoholprobleme" angegeben wurde. Tendenzuelle Charakteristika für eine Therapieteilnahme sind die Angaben "habe starke Alkoholprobleme, die Familie/der Partner drängt nicht" und "habe starke Alkoholprobleme, die Familie/der Partner drängt etwas".

Mit der Prädiktions-KFA konnten zwar "Prädiktionstypen" und "Prädikationsantypen" ermittelt werden, jedoch kann nicht, wie mittels der loglinearen Analyse möglich ist, sichtbar gemacht werden, durch welche Haupteffekte bzw. durch welche Interaktionseffekte unterschiedlicher Ordnung diese determiniert werden. Damit sind auch die Vergleichsmöglichkeiten der (Un-)Ähnlichkeiten der einzelnen "Prädiktionstypen" und "Prädikationsantypen" sehr begrenzt, was insbesondere für die lediglich tendenziell nachgewiesenen zutrifft. Als exploratives Hilfsmittel kann für diesen Zweck wiederum die Korrespondenzanalyse verwendet werden, somit ist eine weitergehende Beschreibung der Ergebnisse der KFA möglich.

8. Dritter Schritt: Grafische Darstellung der Ergebnisse der KFA

In diesem Abschnitt werden die im Abschnitt 7 nachgewiesenen, als auch die als nicht signifikant ermittelten, "Prädiktionstypen" und "Prädikationsantypen" graphisch abgebildet, so daß (Un-)Ähnlichkeiten zwischen diesen sichtbar werden. Eingabeinformation für die zweite Korrespondenzanalyse sind die absoluten Häufigkeiten, die aus der Kreuztabellierung mit der in der KFA verwendeten zusammengesetzten Variable (neun Merkmale) und der Variable "Therapieteilnahme", hervorgegangen sind. Die Tabelle ist identisch mit Tabelle 4.

**Tabelle 6: Eingabedaten der zweiten Korrespondenzanalyse**

	nur Erst- gespräch	nur Vb.- gruppe	Therapie- teilnahme
Keine A.probl., Fam. drängt nicht	21	5	3
Keine A.probl., Fam. drängt etwas	9	4	0
Keine A.probl., Fam. drängt stark	7	10	0
Etwas A.probl., Fam. drängt nicht	18	10	5
Etwas A.probl., Fam. drängt etwas	12	9	1
Etwas A.probl., Fam. drängt stark	17	6	9
Starke A.probl., Fam. drängt nicht	57	41	47
Starke A.probl., Fam. drängt etwas	27	34	28
Starke A.probl., Fam. drängt stark	42	40	33

Tabelle 7: Chi-Quadrat-Statistik der zweiten Korrespondenzanalyse

	nur Erst- gespräch	nur Vb.- gruppe	Therapie- teilnahme
Keine A.probl., Fam. drängt nicht	6.63	2.28	2.48
Keine A.probl., Fam. drängt etwas	2.39	0.03	3.24
Keine A.probl., Fam. drängt stark	0.00	3.27	4.24
Etwas A.probl., Fam. drängt nicht	1.33	0.10	1.27
Etwas A.probl., Fam. drängt etwas	0.89	0.36	3.67
Etwas A.probl., Fam. drängt stark	0.01	0.27	0.21
Starke A.probl., Fam. drängt nicht	0.18	1.17	3.24
Starke A.probl., Fam. drängt etwas	2.71	0.60	1.51
Starke A.probl., Fam. drängt stark	0.71	0.06	0.65

$\chi^2=43.5$ mit $df=16$; total inertia=0.0861

Bei der Prädiktions-KFA konnte unter inferenzstatistischen Testbedingungen als einziger "Prädiktionstyp" die Merkmalskombination "keine Alkoholprobleme, die Familie/der Partner

drängt nicht" und "nur am Erstgespräch teilgenommen" nachgewiesen werden. Diese Zelle trägt auch bei der Korrespondenzanalyse am meisten zur Gesamtvariation des Modells bei, d.h. hier sind die größten Abweichungen von erwarteten und beobachteten Werten. Am wenigsten zur Gesamtvarianz trägt die Merkmalskombination "etwas Alkoholprobleme, die Familie/der Partner drängt stark" in allen Ausprägungen der zu beschreibenden Variablen bei - mit der Prädiktions-KFA konnte für diese drei Zellen nicht einmal tendenziell ein Zusammenhang zwischen den beiden Variablen nachgewiesen werden.

Wie auch schon in der ersten Korrespondenzanalyse ist das Gegensatzpaar auf der ersten Achse (erklärte Varianz: 72.0%) "Therapie Teilnehmer" vs. "nur am Erstgespräch teilgenommen". Interpretieren wir die Variablenausprägungen anhand der (Un-)Ähnlichkeit ihrer Winkel zum Achsenkreuz, können wir den mit der Prädiktions-KFA ermittelten "Prädiktions-typ" auch in der Korrespondenzanalyse finden. Auch in der Korrespondenzanalyse kann das Merkmal "Therapie Teilnehmer" am besten durch die Merkmalskombination "keine Alkoholprobleme, Familie drängt nicht" beschrieben werden; der Winkel (bezogen zum Achsenkreuz) zwischen diesen beiden Vektoren ist am kleinsten.

Abbildung 2: Legende

ERSTGS	nur Erstgespräch
VGRUPPE	nur Vorbereitungsgruppe
THERAPIE	Therapie Teilnehmer
APNFDN	Keine A.probl., Fam. drängt nicht
APNFDE	Keine A.probl., Fam. drängt etwas
APNFDS	Keine A.probl., Fam. drängt stark
APEFDN	Etwas A.probl., Fam. drängt nicht
APEFDE	Etwas A.probl., Fam. drängt etwas
APEFDS	Etwas A.probl., Fam. drängt stark
APJFDN	Starke A.probl., Fam. drängt nicht
APJFDE	Starke A.probl., Fam. drängt etwas
APJFDS	Starke A.probl., Fam. drängt stark



Abbildung 2: Grafische Darstellung der zweiten Korrespondenzanalyse

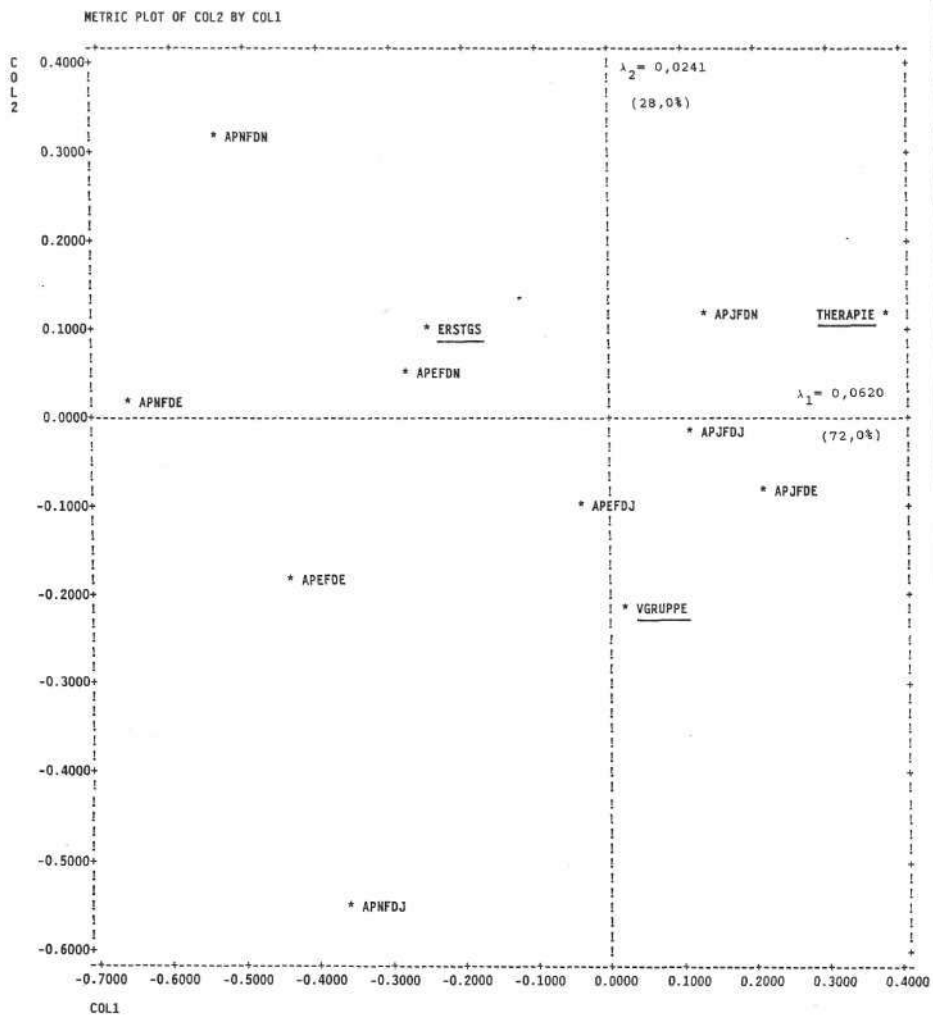


Tabelle 8: Numerische Darstellung der zweiten Korrespondenzanalyse

GENSTAT	MASS	SQCOR	INR	LOC1	QCOR1	INR1	LOC2	QCOR2	INR2
ERSTGS	0.416	1.000	0.341	-0.246	0.852	0.404	0.102	0.148	0.180
VGRUPPE	0.335	1.000	0.187	0.019	0.007	0.002	-0.219	0.993	0.663
THERAPIE	0.250	1.000	0.472	0.384	0.907	0.594	0.123	0.093	0.156
APNFDN	0.057	1.000	0.262	-0.541	0.746	0.271	0.316	0.254	0.237
APNFDE	0.026	1.000	0.130	-0.660	0.999	0.181	0.022	0.001	0.001
APNFDJ	0.034	1.000	0.173	-0.362	0.297	0.071	-0.557	0.703	0.434
APEFDN	0.065	1.000	0.062	-0.281	0.967	0.083	0.052	0.033	0.007
APEFDE	0.044	1.000	0.113	-0.437	0.853	0.134	-0.181	0.147	0.059
APEFDJ	0.083	1.000	0.011	-0.040	0.137	0.002	-0.100	0.863	0.035
APJFDN	0.287	1.000	0.105	0.134	0.565	0.083	0.117	0.435	0.164
APJFDE	0.176	1.000	0.111	0.215	0.853	0.131	-0.089	0.147	0.058
APJFDJ	0.228	1.000	0.033	0.109	0.960	0.043	-0.022	0.040	0.005

Neben dem einzigen (echten) "Prädiktionstyp" der Prädiktions-KFA können auch die anderen lediglich tendenziell (Lehmacher-Test ohne Alpha-Adjustierung) nachgewiesenen "Prädiktionstypen" mittels der Korrespondenzanalyse bestätigt werden - auch die Winkel zwischen diesen Vektoren sind relativ klein. Ferner können auch die "Prädiktionsantitypen", durch eine imaginäre Verbindung der jeweiligen Spalten- und Zeilenvariablenausprägungen, belegt werden - sie alle gehen relativ genau durch den Schwerpunkt oder anders ausgedrückt: Die Winkel zwischen diesen Vektoren sind sehr groß. Diese in beiden Dimensionen konträre Projektion auf den Achsen kann als negative Korrelation zwischen Zeilen- und Spaltenmerkmalen interpretiert werden.

Neben der Beschreibung der in der KFA gefundenen "Prädiktionstypen" und "Prädiktionsantitypen" können die Distanzen zwischen den Zeilen- sowie zwischen den Spaltenvariablen ebenfalls interpretiert werden; so z.B. die Ähnlichkeit der Merkmalskombinationen "keine Alkoholprobleme, die Familie/der Partner drängt etwas" und "etwas Alkoholprobleme, die Familie/ der Partner drängt nicht".

9. Fazit

Ziel dieses Beitrags war es, komplementäre Verwendungsmöglichkeiten der Korrespondenzanalyse und der KFA sowie Ähnlichkeiten dieser Verfahren aufzuzeigen. Hierbei konnte gezeigt werden, daß mit der (Prädiktions-)KFA statistisch signifikante "Typen" und "Antitypen" ermittelt werden können, wobei jedoch die Anzahl der zu verwendenden Variablen eng limitiert ist. Aus dieser Beschränkung folgt, daß die Auswahl der möglichen Variablen nicht mit dem Verfahren selbst erfolgen kann, sondern daß diese bereits bekannt sein müssen. Wenn dies nicht der Fall ist, können die relevanten Variablen vor der inferenzstatistischen Analyse mit einem explorativen Analyseverfahren ermittelt werden.

Bei der konfirmatorischen Datenanalyse entscheidet über "Signifikanz" und "Nicht-Signifikanz" oft nur die zweite Stelle hinter dem Komma, so daß die Testentscheidung häufig willkürlich erscheint. Um diese Differenzen nicht übermäßig zu bewerten, können mittels der Korrespondenzanalyse alle Abweichungen in den Daten grafisch (und numerisch) dargestellt und interpretiert werden. Ein derartiges Vorgehen bezüglich der komplementären Verwendung der Korrespondenzanalyse und log-linearer Modelle schlagen z.B. *Goodman* (1986) und *van der Heijden et al.* (1989) vor.

Die Korrespondenzanalyse eignet sich zum Aufspüren und Beschreiben nicht-trivialer Strukturen in mehrdimensionalen Daten, nicht aber zur Prüfung der Zusammenhänge auf Signifikanz. Sollen Aussagen nicht nur auf der beschreibenden Ebene, sondern auch auf der erklärenden gemacht werden, so muß ein konfirmatorisches Verfahren (z.B. die KFA) verwendet werden. Da sowohl die explorative Beschreibung von Zusammenhängen als auch das Testen auf Signifikanz wichtige Ziele der empirischen sozialwissenschaftlichen Forschung sind, bleibt die Frage: Warum nicht zwei Verfahren komplementär verwenden?

Mit diesem Beitrag wurde ein derartiger Vorschlag zur komplementären Verwendung von konfirmatorischen und explorativen Verfahren gemacht, hier dargestellt am Beispiel der Korrespondenzanalyse und der (Prädiktions-)KFA. Konfirmatorische und explorative Analyseverfahren sind u.E. auch keine Gegensätze, wie implizit und teilweise auch explizit immer wieder behauptet wird, sondern Ergänzungen. Nicht zu Unrecht fordert *Tukey* (1980) im Titel eines Beitrages: "We Need Both Exploratory and Confirmatory".



Literatur

J.-P. *Benzécri et collaborateurs* (1973):

L'Analyse des Données II. L'Analyse des Correspondances. Paris: Dunod.

Jörg *Blasius* (1987):

Einstellungen zur Hamburger Innenstadt. Eine Auswertung mit Hilfe der Korrespondenzanalyse. *ZA-Information*, 21, S. 29-51.

Jörg *Blasius* and Harald *Rohlinger* (1989):

KORRES - A Program for Multivariate Analysis of Categorical Data from Contingency Tables.

In: SAS Institute Inc. (eds.): SEUGI'89. Proceedings of the SAS Users Group International Conference. Cary, N.C.: SAS Institute Inc., S. 98-117.

Jörg *Blasius* und Joachim *Winkler* (1989):

Gibt es die feinen Unterschiede? Eine empirische Überprüfung der Bourdieuschen Theorie. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 41, S. 72-94.

Pierre *Bourdieu* (1982):

Die feinen Unterschiede. Kritik der gesellschaftlichen Urteilskraft. Frankfurt/M.: Suhrkamp.

Pierre *Bourdieu* (1988):

Homo Academicus. Frankfurt/M.: Suhrkamp.

Jens *Dangschat* and Jörg *Blasius* (1987):

Social and Spatial Disparities in Warsaw in 1978:

An Application of Correspondence Analysis to a 'Socialist' City. *Urban Studies*, 24, S. 173-191.

Alexander *von Eye* (1990):

Introduction to Configural Frequency Analysis. New York: Cambridge University Press.

Leo A. *Goodman* (1986):

Some Useful Extensions of the Usual Correspondence Analysis Approach and the Usual Log-Linear Models Approach in the Analysis of Contingency Tables. *International Statistical Review*, 54, S. 243-309.

Michael *Greenacre* (1984):

Theory and Applications of Correspondence Analysis. London: Academic.

Michael *Greenacre* (1989):

The Carroll-Green-Schaffer Scaling in Correspondence Analysis: A Theoretical and Empirical Appraisal. *Journal of Marketing Research*, 26, S. 358-365.

Michael *Greenacre* (1990):

SimCA 2.0. Pretoria.

Michael *Greenacre* and Trevor *Hastie* (1987):

The Geometric Interpretation of Correspondence Analysis.

Journal of the American Statistical Association, 82, S. 437-447.

Joachim **Krauth** (1983):

Typenanalyse. In: Jürgen **Bredenkamp** und Hubert **Feger** (Hrsg.): Forschungsmethoden der Psychologie, Bd. 4, Enzyklopädie der Psychologie. Göttingen: Hogrefe, S. 440-496.

Joachim **Krauth** und Gustav A. **Lienert** (1973):

KFA. Die Konfigurationsfrequenzanalyse und ihre Anwendung in Psychologie und Medizin. Freiburg: Alber.

Erwin **Lautsch** und Stefan **von Weber** (1990):

Die Konfigurationsfrequenzanalyse (KFA) - Methoden und Anwendungen. Berlin: Volk und Wissen.

Ludovic **Lebart**, Alain **Morineau** and Kenneth M. **Warwick** (1984):

Multivariate Statistical Analysis. Correspondence Analysis and Related Techniques for Large Matrices. New York: Wiley.

Gustav A. **Lienert** (1969):

Die "Konfigurationsfrequenzanalyse" als Klassifikationsmittel in der klinischen Psychologie. In: Martin **Irlé** (Hrsg.): Bericht 26. Kongreß der Deutschen Gesellschaft für Psychologie in Tübingen 1968. Göttingen: Hogrefe, S. 244-253.

Gustav A. **Lienert** (1988):

Angewandte Konfigurationsfrequenzanalyse. Frankfurt/M.: Athenäum.

Viktor **Thiessen** und Harald **Rohlinger** (1988):

Die Verteilung von Aufgaben und Pflichten im ehelichen Haushalt. Kölner Zeitschrift für Soziologie und Sozialpsychologie, 40, S. 640-658.

John **W. Tukey** (1980):

We Need Both Exploratory and Confirmatory. The American Statistician, 34, S. 23-25.

Peter G.M. **van der Heijden** and Jan **de Leeuw** (1985):

Correspondence Analysis Used Complementary to Loglinear Analysis. Psychometrika, 50, S. 429-447.

Peter G.M. **van der Heijden**, Antoine **de Falguerolles** and Jan **de Leeuw** (1989):

A Combined Approach to Contingency Table Analysis Using Correspondence Analysis and Log-Linear Analysis. Applied Statistics, 38, S. 249-292.