

Zur Stabilität von Ergebnissen bei der Korrespondenzanalyse

Blasius, Jörg

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. (1988). Zur Stabilität von Ergebnissen bei der Korrespondenzanalyse. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 23, 47-62. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-204922>

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.

Zur Stabilität von Ergebnissen bei der Korrespondenzanalyse

von Jörg Blasius

1. Einleitung

Die Korrespondenzanalyse wurde Anfang der 60er Jahre in Frankreich entwickelt (BENZÉCRI 1964, ESCOFIER-CODIER 1965). Sie kann auf HIRSCHFELD (1935) zurückgeführt werden, der als erster eine algebraische Formulierung der Korrelation von Zeilen und Spalten einer Kontingenztafel gab. HIRSCHFELDs Gedanken wurden von GUTTMAN (1941, 1950) mit der 'principal components analysis' und von HAYASHI (1950, 1952, 1954) mit der 'quantification of qualitative data' aufgenommen. Die Korrespondenzanalyse kann als Weiterentwicklung dieser Verfahren angesehen werden. Ihr wichtigstes Merkmal ist, daß in der Regel qualitative Daten als Eingangsinformationen verwendet werden.

Bei der Korrespondenzanalyse handelt es sich primär um ein Verfahren zur simultanen grafischen Darstellung von Zeilen und Spalten einer (mehrerer) Kontingenztafel(n). Ähnlich wie bei der Hauptkomponentenanalyse, der multidimensionalen Skalierung oder der Clusteranalyse handelt es sich bei der Korrespondenzanalyse um ein Datenreduktionsverfahren - vorhandene Informationen sollen mit möglichst wenigen Faktoren (Dimensionen) abgebildet werden. Diese sind orthogonal zueinander. Insgesamt gesehen ist die Korrespondenzanalyse ein mächtiges Instrument zur explorativen Analyse kategorialer (oder höher skaliertes) Daten.

Während das Verfahren in Frankreich als sozialwissenschaftliche Standardmethode angesehen werden kann und es nicht nur in der von dem "Vater" der Korrespondenzanalyse BÉNÉCRI herausgegebenen Zeitschrift "Cahiers de l'Analyse des Données" eine Fülle von Anwendungsbeispielen gibt, wurde das Verfahren im deutsch- und englischsprachigen Raum bisher nur selten verwendet. Die wohl bekannteste Anwendung stammt aber auch hier aus Frankreich: "Die feinen Unterschiede" von Pierre BOURDIEU (1979, dt. 1982). In seinem empirischen Teil beschreibt BOURDIEU mit Hilfe der Korrespondenzanalyse "feine Unterschiede" innerhalb dreier von ihm differenzierter Klassen (Oberschicht, Mittelschicht, Arbeiterklasse) der Gesellschaft.

DANGSCHAT und BLASIUS (1987) verwendeten das Verfahren zur Beschreibung von räumlichen und sozialstrukturellen Ungleichheiten in Warschau. Mit Hilfe des Censur-Datensatzes aus dem Jahre 1978 wurden die Stadtteile der polnischen Hauptstadt

anhand von Variablen zum Gebäudealter, dem Bauträger, der Schulbildung, der Haushaltsgröße und dem Alter (kategorisiert) der Bewohner beschrieben. Es konnten Stadtteile benannt und mit Hilfe sozialstruktureller Merkmale beschrieben werden, in denen Personen mit höherer Schulbildung und in denen Personen mit einfacher Schulbildung überdurchschnittlich oft lebten.

BLASIUS (1987) verwendete das Verfahren zur Beschreibung von vier Subpopulationen von Stadtbewohnern (Personen, die ihre Freizeitaktivitäten überdurchschnittlich oft im eigenen Stadtteil ausüben, Personen die ihre Freizeitaktivitäten überdurchschnittlich oft in der Innenstadt ausüben, Personen, die in beiden Gebieten eine überdurchschnittliche Anzahl von Freizeitaktivitäten haben, und Personen, die in ihrer Freizeit relativ selten aktiv sind) anhand soziodemografischer Merkmale und Einstellungen zur Innenstadt bzw. dem eigenen Stadtteil. THIESSEN und ROHLINGER (1988) beschrieben mit Hilfe der Korrespondenzanalyse die Veränderung der haushaltlichen Aufgabenteilung (z.B. beim Kochen, Autofahren, Saubermachen, Urlaub/Ausflüge organisieren - insgesamt dreizehn Aktivitäten) zwischen Mann und Frau in den ersten vier Ehejahren (Datengrundlage ist ein Panel mit insgesamt drei Erhebungszeitpunkten). Hierbei unterschieden die Autoren nach "gemeinsamem Erledigen", "abwechselndem Erledigen" sowie "überwiegend Aufgabe der Frau" und "überwiegend Aufgabe des Mannes" - jeweils aus der Sicht beider Personen. Weitere Anwendungsbeispiele sowie ausführliche Darstellungen des Verfahrens sind den Lehrbüchern von GREENACRE (1984) und LEBART et al. (1984) zu entnehmen.

2. Daten

Ziel dieser Arbeit ist zu zeigen, daß die Korrespondenzanalyse entgegen anderen multivariaten Verfahren sehr robust gegenüber der Aufnahme weiterer Variablen ist. Als mehr oder weniger humorvoll gemeintes Beispiel wurden Daten zur kulturellen Kompetenz verwendet, die in der allgemeinen Bevölkerungsumfrage (ALLBUS) (1) im Jahre 1986 erhoben wurden. Als zu beschreibende Variable wurde eine Kombination der Merkmale Geschlecht und Alter mit insgesamt sechs Ausprägungen (Männer bis 39 Jahre, Männer von 40 bis 59 Jahren, Männer ab 60 Jahre sowie Frauen derselben Altersgruppen) gebildet. In einem ersten Schritt wurden die "vorhandenen", "teilweise vorhandenen" und "nicht vorhandenen" Fähigkeiten bezüglich Walzer tanzen, Schach spielen, Hosen umnähen und tanzen zu Popmusik mit der vorab beschriebenen Variablen kreuztabelliert.

Die rechnerische Durchführung der Analysen erfolgte mit KORRES, einem von BLASIUS und ROHLINGER (1988) in der matrixorientierten Sprache SAS - PROC MATRIX geschriebenen Programmes. Die Eingabe ist bei diesem Programm denkbar einfach: In einem ersten Schritt wurde die aus Geschlecht und Alter kombinierte Variable mit den vier Merkmalen zur kulturellen Kompetenz kreuztabelliert, anschließend wurden die absoluten Häufigkeiten der vier Tabellen untereinander geschrieben. Diese so entstandene Matrix war das Eingangsmaterial für die erste Korrespondenzanalyse (siehe Tabelle 1), deren Ergebnisse sowohl in grafischer (Abbildung 1) als auch in numerischer Form erhältlich sind.

Tabelle 1: Dateneingabe bei KORRES

```
DATA N;  
INPUT YOUNGH MIDAGEM OLDM YOUNGW HIDAGEW OLDW ROWS $ ;CARDS;  
208 298 208 426 425 323 WALTZYES  
178 103 64 158 69 52 WALTZJAB  
235 84 52 109 27 54 WALTZNO  
287 146 114 87 47 23 CHESSYES  
143 68 44 118 54 27 CHESSJAB  
195 269 167 485 419 375 CHESSNO  
167 109 91 637 500 406 TROUSYES  
132 76 62 32 12 12 TROUSJAB  
308 291 165 23 10 11 TROUSNO  
386 145 33 535 214 67 POPDAYES  
138 105 33 92 88 32 POPDAJAB  
95 234 249 58 208 312 POPDANO  
  
PROC MATRIX ;
```

3. Ergebnisse

Die erste Achse (dargestellt durch die Abzisse) erklärt 63,9 Prozent der Varianz des gesamten Modelies. Hier lassen sich auf der rechten Seite die drei Subpopulationen (2) der Männer und auf der linken Seite die drei Subpopulationen der Frauen projizieren. Wir bezeichnen diese Achse als "Geschlechtsdimension". Bei einer Zuordnung der kulturellen Kompetenzen lassen sich den drei Gruppen der Männer insbesondere die Fähigkeit des Schachspielens und die Unfähigkeit bzw. die nur teilweise vorhandene Fähigkeit des Hosenumnäehens zuordnen; bei den drei Gruppen der Frauen ist es die vorhandene Fähigkeit des Hosenumnäehens und die nicht vorhandene Fähigkeit des Schachspielens. Somit gibt es zwei geschlechtsspezifische Variablen (Hosen umnähen, Schach spielen), wobei aber zu beachten ist, daß die genannten Zuordnungen auf einem relativen Vergleich der (sechs) Subpopulationen beruhen.

Mit der zweiten Achse (dargestellt durch die Ordinate), werden zusätzlich 32,2 Prozent der Gesamtvarianz des Modells erklärt, somit werden durch die ersten beiden Achsen 96,1 Prozent der Gesamtvarianz des Modells determiniert. Im oberen Abschnitt der zweiten Achse lassen sich die älteren Männer und die älteren Frauen und im unteren die jüngeren Männer und die jüngeren Frauen projizieren. Diese Achse wird als "Altersdimension" bezeichnet, wobei als diskriminierende Variable insbesondere die Fähigkeit (von jüngeren Männern und jüngeren Frauen) bzw. die Unfähigkeit (von älteren Männern und älteren Frauen) des Tanzens zu Popmusik nennenswert ist.

Werden lediglich die vier genannten kulturellen Fähigkeiten betrachtet, so lassen sich deren unterschiedliche Vorkommenshäufigkeiten in den Subpopulationen zu über 96 Prozent durch die Dimensionen "Geschlecht" und "Alter" interpretieren. Während sich mittels der grafischen Darstellung die Variablenausprägungen der kulturellen Fähigkeiten und der Subpopulationen durch einfache Projektion den beiden dargestellten Achsen zuordnen lassen, kann mit Hilfe der numerischen Ergebnisse festgestellt werden, mit welchen Achsen welche Variablenausprägungen/Subpopulationen wie stark korrelieren. Diese Korrelationen, die von uns in quadrierter Form angegeben werden, sowie einige andere Statistiken sind in Tabelle 2 (GENERAL STATISTIC) dargestellt.

Tabelle 2: Numerische Ergebnisse der ersten Korrespondenzanalyse

GENSTAT	GENERAL STATISTIC								
	MASS	SQCOR	INR	LOCI	QCOR1	INR1	LOC2	QCOR2	INR2
YOUNGH	0.202	0.981	0.280	0.562	0.816	0.357	-0.252	0.165	0.143
MIDAGEH	0.158	0.893	0.119	0.375	0.665	0.124	0.220	0.228	0.084
OLDM	0.105	0.967	0.121	0.313	0.304	0.058	0.463	0.663	0.249
YOUNGW	0.226	0.988	0.183	-0.293	0.377	0.108	-0.373	0.611	0.347
MIDAGEW	0.170	0.956	0.124	-0.441	0.952	0.184	0.031	0.005	0.002
OLDW	0.139	0.950	0.173	-0.467	0.624	0.169	0.338	0.326	0.175
WALTZYES	0.155	0.902	0.037	-0.212	0.661	0.039	0.128	0.241	0.028
WALTZJAB	0.051	0.955	0.015	0.219	0.595	0.014	-0.170	0.360	0.016
WALTZNO	0.046	0.798	0.055	0.467	0.645	0.056	-0.227	0.153	0.026
CHESSYES	0.058	0.968	0.091	0.654	0.962	0.137	-0.051	0.006	0.002
CHESSJAB	0.037	0.977	0.017	0.253	0.496	0.013	-0.249	0.480	0.026
CHESSNO	0.156	0.975	0.054	-0.296	0.909	0.076	0.080	0.066	0.011
TROUSYES	0.156	0.992	0.168	-0.536	0.954	0.251	-0.106	0.037	0.020
TROUSJAB	0.027	0.979	0.053	0.738	0.975	0.081	0.047	0.004	0.001
TROUSNO	0.066	0.976	0.222	0.929	0.916	0.318	0.239	0.060	0.042
POPDAYES	0.113	0.990	0.115	-0.001	0.000	0.000	-0.531	0.990	0.353
POPDAJAB	0.040	0.580	0.016	0.225	0.456	0.011	-0.118	0.124	0.006
POPDAANO	0.095	0.983	0.156	-0.072	0.011	0.003	0.670	0.972	0.470



In den Spalten QCOR1 und QCOR2 sind die quadrierten Korrelationen (analog interpretierbar zu den quadrierten Ladungen bei der Hauptkomponentenanalyse) mit den ersten beiden Achsen dargestellt. Die Summe der quadrierten Korrelationen der berücksichtigten Faktoren ist in der Spalte SQCOR aufgeführt, dessen Komponenten sind den Kommunalitäten bei der Hauptkomponentenanalyse vergleichbar. In den Spalten LOCI und LOC2 ist dargestellt, auf welcher Seite und wie weit die einzelnen Variablenausprägungen und Subpopulationen vom Achsenkreuz (dem Nullpunkt) entfernt sind.

In den Spalten INR1 und INR2 (inertia, Trägheitsgewicht) ist dargestellt, inwieweit die einzelnen Achsen durch die Variablen(-ausprägungen) bzw. Subpopulationen determiniert werden, ihre Summen sind spalten- und zeilenweise getrennt auf eins normiert. Während mit den Koeffizienten QCOR beschrieben wird, welcher Anteil der Varianz der einzelnen Variablenausprägungen durch die einzelnen Achsen determiniert wird, ist es bei dem Koeffizienten INR1 (INR2) umgedreht: Hier wird die erklärte Varianz der Achsen durch die Variablenausprägungen angegeben. Die zeilenweisen Summen dieser Werte sind, gewichtet mit den Erklärungsanteilen der einzelnen Achsen (den Eigenwerten), in der Spalte INR aufgeführt. In dieser Spalte ist abzulesen, wie stark das gesamte Modell von den einzelnen Ausprägungen der unterschiedlichen kulturellen Kompetenzen auf der Ebene der Zeilenvariablen und von den einzelnen Subpopulationen auf der Ebene der Spaltenvariablen determiniert wird. In der Spalte MASS ist die relative Häufigkeit der Variablenausprägungen und der Subpopulationen aufgeführt, wiederum getrennt für Zeilen und Spalten auf eins normiert.

Wählen wir als Schwellenwert für die Zuordnung einer Subpopulation/Variablenausprägung zu einer Achse .60 (d.h. 60 Prozent der Gesamtvarianz werden durch die jeweilige Achse determiniert), so laden auf der ersten Achse die jüngeren Männer und die im mittleren Alter, negativ korreliert hinsichtlich der verwendeten kulturellen Merkmale mit den älteren Frauen und denen mittleren Alters. Den beiden Gruppen der Männer lassen sich die Unfähigkeit bzw. die nur teilweise vorhandene Fähigkeit des Walzer-Tanzens und des Umnähens von Hosen zuordnen, ebenso die Fähigkeit des Schachspielens. Demgegenüber sind die Mitglieder der beiden genannten Gruppen der Frauen überdurchschnittlich oft in der Lage, Hosen umzunähen bzw. Walzer zu tanzen, während sie relativ selten Schach spielen können. Insgesamt betrachtet sind die eigentlichen Gegensatzpaare auf der als "Geschlechtsdimension" bezeichneten Achse die jüngeren Männer und die mittleren Alters versus die älteren Frauen und die Frauen mittleren Alters.

Mit der zweiten, als "Dimension des Alters" bezeichneten Achse, korreliert das Gegensatzpaar "ältere Männer" versus "jüngere Frauen". Differenziert werden beide Gruppen durch die Fähigkeit auf der Seite der jüngeren Frauen bzw. die Unfähigkeit auf der Seite

der älteren Männer des Tanzens zu Popmusik. Die Variablenausprägungen "etwas fähig Schach zu spielen" und "etwas fähig zum Tanzen zu Popmusik" laden auf keiner der beiden Achsen, sie sind mit den sechs Subpopulationen nahezu unkorreliert.

Während auf der Ebene der Subpopulationen die jüngeren Männer die erste Achse zu 35,7 Prozent determinieren (siehe INR1), ist es auf der Ebene der Zeilenvariablen insbesondere die Variable "(Un)fähigkeit des Hosenumnärens", die mit ihren drei Ausprägungen insgesamt 65,0 Prozent (es sind die drei Werte in der Spalte INR1 zu addieren) der ersten Achse determiniert. Beim zweiten Faktor sind als herausragende Determinationsmerkmale auf der Ebene der Subpopulationen die älteren Männer und die jüngeren Frauen zu nennen, die zusammen 59,6 Prozent der Varianz dieser Achse erklären und auf der Ebene der Zeilenvariablen die 'Fähigkeit bzw. Unfähigkeit, zu Popmusik zu tanzen' (82,3 Prozent). Die insgesamt sechs Ausprägungen der Variablen "Walzer tanzen" und "Schach spielen" sowie die nur teilweise vorhandene Fähigkeit, zu Popmusik zu tanzen, haben dagegen nur eine geringe Bedeutung zur Determination der ersten beiden Achsen.

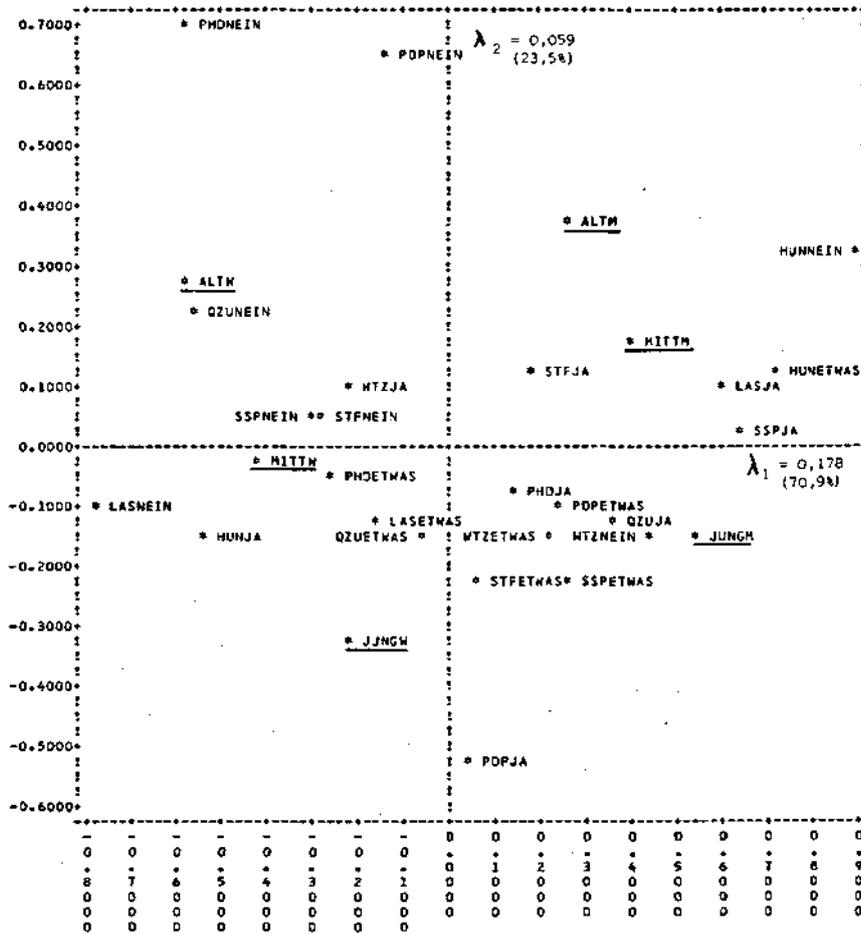
4. Aufnahme zusätzlicher Variablen

Um die Robustheit der Korrespondenzanalyse zu testen, wurden für die nachfolgende Analyse vier weitere kulturelle Fähigkeiten aufgenommen: "Lampe anschließen", "Photografieren", "Steuerformular auszufüllen" und "Einstellen einer Quarzuhr", jeweils wieder mit den drei Ausprägungen (5) "Fähigkeit vorhanden", "etwas vorhanden", "nicht vorhanden". Die grafische Darstellung der Ergebnisse ist in Abbildung 2 wiedergegeben.

Bei Betrachtung von Abbildung 2 fällt auf, daß ebenso wie in der ersten Analyse die erste Achse die "Geschlechtsdimension" ist. Dem positiven Bereich dieser Achse lassen sich die drei Subpopulationen der Männer, dem negativen Bereich die drei Subpopulationen der Frauen zuordnen. Ähnliches gilt auch für den Vergleich der zweiten Achsen ("Dimensionen des Alters"), auch hier polarisieren die Extremgruppen der Altersgruppierungen. Bei den Zuordnungen der kulturellen Kompetenzen zu den Subpopulationen sind auf der "Geschlechtsdimension" bei den Frauen insbesondere die überdurchschnittlich oft vorhandene Unfähigkeit, eine Lampe anzuschließen bzw. eine Quarzuhr einzustellen und bei den Männern die relativ oft vorhandene Fähigkeit, eine Lampe anzuschließen, hinzugekommen. Dem zweiten Faktor kann zusätzlich die Unfähigkeit des Fotografierens als beschreibendes Merkmal hinzugefügt werden.



Abbildung 2: Grafische Darstellung der zweiten Korrespondenzanalyse



- JUNGM Männer bis 39 Jahre
- MITTM Männer von 40 bis 59 Jahren
- ALTM Männer ab 60 Jahre
- JUNGW Frauen bis 39 Jahre
- MITTW Frauen von 40 bis 59 Jahren
- ALTW Frauen ab 60 Jahre
- WTZJA kann Walzer tanzen
- WTZETWAS kann etwas Walzer tanzen
- WTZNEIN kann nicht Walzer tanzen
- SSPJA kann Schach spielen
- SSPETWAS kann etwas Schach spielen
- SSPNEIN kann nicht Schach spielen
- HUNJA kann Hosen umnähen
- HUNETWAS kann etwas Hosen umnähen
- HUNNEIN kann nicht Hosen umnähen
- POPJA kann zu Popmusik tanzen
- POPETWAS kann etwas zu Popmusik tanzen
- POPNEIN kann nicht zu Popmusik tanzen

- LASJA kann Lampe anschließen
- LASETWAS kann etwas Lampe anschließen
- LASNEIN kann nicht Lampe anschließen
- PHOJA kann fotografieren
- PHOETWAS kann etwas fotografieren
- PHONEIN kann nicht fotografieren
- STFJA kann Steuerformular ausfüllen
- STFETWAS kann Steuerformular etwas ausfüllen
- STFNEIN kann Steuerformular nicht ausfüllen
- QZUJA kann Quarzuhr einstellen
- QZUETWAS kann Quarzuhr etwas einstellen
- QZUNEIN kann Quarzuhr nicht einstellen



Gegenüber dem ersten Ergebnis haben sich die Anteile der erklärten Varianzen der beiden ersten Achsen geändert. War im ersten Beispiel deren Verhältnis 63,9 zu 32,2 Prozent, so beträgt es im zweiten 70,9 zu 23,5 Prozent. Inhaltlich bedeutet dies, daß bei den hinzugefügten Variablen die geschlechtsspezifische Dimension verhältnismäßig stärker war als bei den ersten vier Variablen. Trotz dieser Verschiebung der Dimensionsanteile sind aber bei der grafischen Darstellung die Positionen der zuerst aufgenommen Variablenausprägungen zueinander und zu den Subpopulationen nahezu identisch geblieben. Dies gilt auch für das Ausmaß der quadrierten Korrelationen, was mittels der numerischen Darstellung gezeigt werden kann (Tabelle 3).

Tabelle 3: Numerische Ergebnisse der zweiten Korrespondenzanalyse

GENSTAT	GENERAL STATISTIC								
	MASS	SQCOR	INR	LOCI	QCOR1	INR1	LOC2	QCOR2	INR2
JUNGM	0.204	0.949	0.259	0.532	0.885	0.324	-0.144	0.065	0.071
MITTM	0.159	0.911	0.132	0.403	0.777	0.145	0.167	0.134	0.075
ALTH	0.106	0.953	0.094	0.261	0.303	0.040	0.382	0.649	0.261
JUNGW	0.225	0.976	0.142	-0.217	0.297	0.059	-0.328	0.679	0.409
MITTW	0.169	0.920	0.126	-0.415	0.919	0.163	-0.015	0.001	0.001
ALTW	0.138	0.946	0.246	-0.589	0.772	0.268	0.280	0.174	0.183
WTZJA	0.078	0.808	0.022	-0.218	0.679	0.021	0.095	0.130	0.012
WTZETWAS	0.026	0.924	0.008	0.226	0.651	0.007	-0.146	0.274	0.009
WTZNEIN	0.023	0.673	0.030	0.447	0.602	0.026	-0.153	0.071	0.009
SSPJA	0.029	0.932	0.050	0.637	0.931	0.066	0.023	0.001	0.000
SSPETWAS	0.019	0.944	0.009	0.267	0.561	0.007	-0.220	0.383	0.015
SSPNEIN	0.079	0.939	0.031	-0.302	0.920	0.040	0.044	0.019	0.003
HUNJA	0.079	0.999	0.096	-0.531	0.917	0.124	-0.159	0.083	0.034
HUNETWAS	0.013	0.954	0.029	0.713	0.924	0.038	0.128	0.030	0.004
HUNNEIN	0.033	0.991	0.123	0.905	0.880	0.153	0.320	0.110	0.058
POPJA	0.057	0.996	0.065	0.048	0.008	0.001	-0.532	0.987	0.272
POPETWAS	0.020	0.618	0.009	0.238	0.522	0.006	-0.103	0.097	0.004
POPNEIN	0.048	0.996	0.087	-0.148	0.048	0.006	0.661	0.948	0.351
LASJA	0.068	0.997	0.101	0.604	0.968	0.138	0.105	0.029	0.013
LASETWAS	0.009	0.502	0.003	-0.158	0.299	0.001	-0.130	0.203	0.002
LASNEIN	0.049	0.998	0.123	-0.784	0.981	0.170	-0.105	0.018	0.009
PHOJA	0.092	0.823	0.011	0.136	0.628	0.010	-0.076	0.196	0.009
PHOETWAS	0.023	0.887	0.007	-0.259	0.860	0.009	-0.046	0.027	0.001
PHONEIN	0.011	0.764	0.049	-0.580	0.309	0.021	0.705	0.456	0.094
STFJA	0.053	0.677	0.015	0.187	0.490	0.010	0.116	0.187	0.012
STFETWAS	0.030	0.939	0.007	0.057	0.058	0.001	-0.224	0.881	0.025
STFNEIN	0.040	0.699	0.017	-0.272	0.685	0.017	0.039	0.014	0.001
QZUJA	0.070	0.993	0.042	0.366	0.885	0.052	-0.127	0.107	0.019
QZUETWAS	0.012	0.470	0.003	-0.054	0.053	0.000	-0.153	0.418	0.005
QZUNEIN	0.041	0.988	0.063	-0.569	0.842	0.075	0.237	0.146	0.039



Beim Vergleich der quadrierten Korrelationen (QCOR) aus beiden Analysen sind keine nennenswerten Abweichungen festzustellen; die Zuordnung von Variablenausprägungen zu den einzelnen Achsenabschnitten (Schwellenwert: .60) ist geblieben. Geändert haben sich lediglich die Werte der Trägheitsgewichte (INR) auf der Ebene der Zeilenvariablen, welche sich in etwa halbiert haben. Inhaltlich bedeutet dies, daß die neu aufgenommenen Variablen in etwa die gleiche Determinationskraft für den Darstellungsraum haben wie die ersten vier. War in der ersten Analyse die erste Achse auf der Ebene der Zeilenvariablen insbesondere durch die "Fähigkeit/Unfähigkeit eine Hose umzunähen" determiniert (65,0%), so hat in der zweiten die Variable "Fähigkeit/Unfähigkeit, eine Lampe anzuschließen" etwa die gleiche Bedeutung (31,5% bzw. 30,9%, zusammen 62,4%). Auf der zweiten Achse ist keine zusätzliche Variable gefunden worden, die eine ähnlich starke Bedeutung für die Beschreibung der Dimension "Alter" hat wie die "Fähigkeit/Unfähigkeit, zu Popmusik zu tanzen". Werden die durch die Verdopplung der Variablen entstandenen Veränderungen im Datensatz betrachtet, so ist die sehr hohe Reproduktion der Koeffizientenwerte der vier Ausgangsvariablen im zweiten Analysegang auffällig.

Beiden Beispielen gemeinsam war, daß sie lediglich Variablen mit zwei Dimensionen (Geschlecht und Alter) enthielten. Um eine dritte Dimension zu erhalten, wurde eine Dummyvariable mit drei Ausprägungen konstruiert (Tabelle 4), in der den jüngeren Männern die gleiche Eigenschaft zugesprochen wurde wie den älteren Frauen (DUMMY1) und den älteren Männern die gleiche wie den jüngeren Frauen (DUMMY3). Der mittleren Variablenausprägung (DUMMY2) wurden keine besonderen Verteilungsmerkmale hinsichtlich der Subpopulationen gegeben. Diese künstlich konstruierte Variable bildete zusammen mit den vier Variablen der ersten Analyse das Eingangsmaterial der dritten Korrespondenzanalyse, deren grafische Ergebnisse in Abbildung 3 dargestellt sind.

Wie im Vergleich mit Abbildung 1 zu erkennen ist, hat sich an der relativen Lage der Variablenausprägungen und Subpopulationen zueinander nahezu nichts geändert. Werden hingegen die relativen Anteile der Achsen betrachtet, so fällt ein dritter Faktor auf, der 16,6 Prozent der Gesamtvarianz erklärt. Trotz dieser neuen Dimension ist das Verhältnis der ersten beiden Achsen zueinander nahezu konstant geblieben, nur auf niedrigerem Niveau (von 63,9 zu 32,2 Prozent nach 52,5 zu 27,4 Prozent). Daß zwei der neu aufgenommenen Dummyvariablenausprägungen auf der dritten Achse laden, läßt sich u.a. mit Hilfe der numerischen Darstellung zeigen (Tabelle 5).



Tabelle 4: Dummyvariable

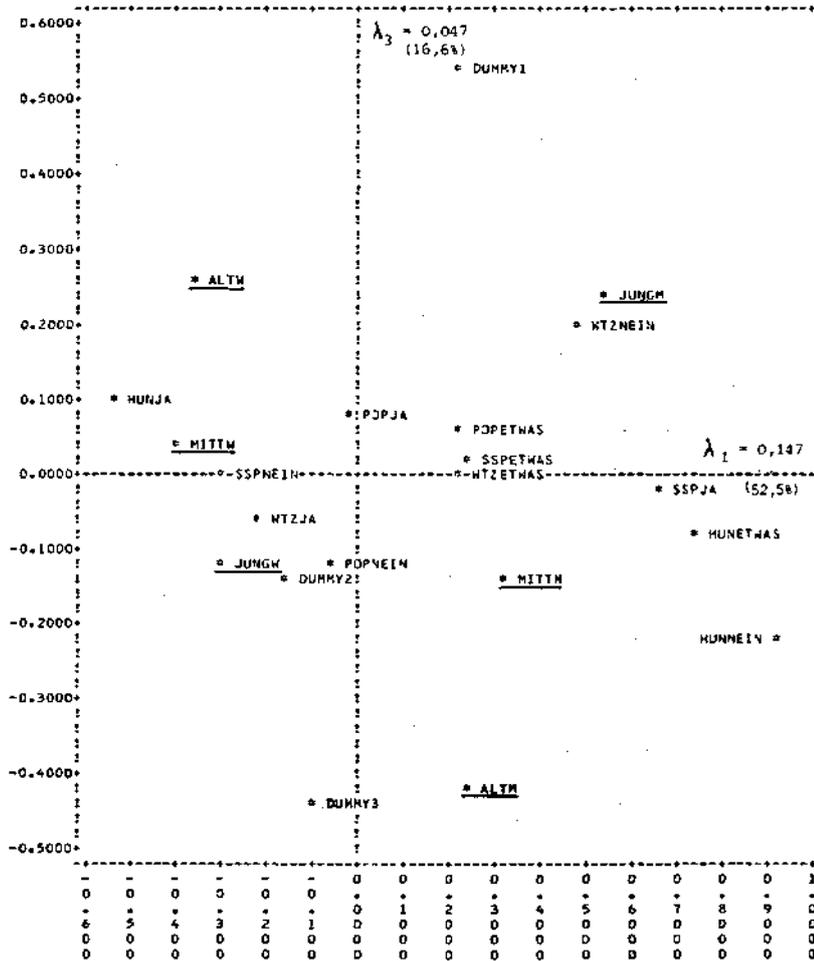
A	DUMMY-CONSTRUCTION							SUM
	YOUNGM	MIDAGEM	OLDM	YOUNGM	MIDAGEM	OLDM	SUM	
DUMMY1	418 (2/3)	162 (1/3)	36 (1/9)	77 (1/9)	174 (1/3)	285 (2/3)	1152	
DUMMY2	70 (2/9)	162 (1/3)	73 (2/9)	154 (2/9)	174 (1/3)	95 (2/9)	728	
DUMMY3	139 (1/9)	162 (1/3)	218 (2/3)	463 (2/3)	174 (1/3)	47 (1/9)	1203	
SUM	627	486	327	694	522	427	3083	

Tabelle 5: Numerische Darstellung der Korrespondenzanalyse mit einer Dummyvariablen

GENSTAT	GENERAL STATISTIC													
	MASS	SQCOR	INR	LOC1	QCOR1	INR1	LOC2	QCOR2	INR2	LOC3	QCOR3	INR3		
JUNGH	0.203	0.997	0.277	0.544	0.771	0.407	-0.174	0.079	0.080	0.238	0.147	0.246		
MITTM	0.158	0.868	0.101	0.320	0.571	0.110	0.179	0.178	0.066	-0.146	0.119	0.072		
ALTM	0.105	0.832	0.126	0.234	0.164	0.039	0.283	0.239	0.110	-0.421	0.529	0.401		
JUNGW	0.226	0.995	0.208	-0.304	0.357	0.182	-0.391	0.591	0.449	-0.110	0.047	0.059		
MITTW	0.170	0.932	0.106	-0.397	0.904	0.182	0.059	0.020	0.008	0.040	0.009	0.006		
ALTW	0.139	0.980	0.181	-0.356	0.346	0.120	0.399	0.436	0.288	0.269	0.198	0.216		
WTZJA	0.123	0.953	0.030	-0.215	0.681	0.039	0.122	0.218	0.024	-0.061	0.054	0.010		
WTZETWAS	0.041	0.972	0.012	0.212	0.564	0.013	-0.181	0.408	0.017	-0.003	0.000	0.000		
WTZWEIN	0.037	0.934	0.044	0.482	0.690	0.058	-0.196	0.114	0.018	0.209	0.130	0.034		
SSPJA	0.046	0.973	0.073	0.653	0.961	0.133	-0.069	0.011	0.003	-0.017	0.001	0.000		
SSPETWAS	0.030	0.988	0.014	0.247	0.472	0.012	0.257	0.513	0.026	0.020	0.003	0.000		
SSPWEIN	0.125	0.979	0.043	-0.295	0.899	0.074	0.088	0.080	0.013	0.002	0.000	0.000		
HUNJA	0.125	0.995	0.135	-0.532	0.937	0.240	-0.077	0.020	0.010	0.108	0.039	0.031		
HUNNETWAS	0.021	0.981	0.042	0.735	0.969	0.078	0.014	0.000	0.000	-0.080	0.012	0.003		
HUNNEIN	0.053	0.982	0.178	0.918	0.896	0.303	0.184	0.036	0.023	-0.218	0.050	0.054		
POPJA	0.090	0.995	0.092	-0.014	0.001	0.000	-0.527	0.971	0.326	0.082	0.023	0.013		
POPETWAS	0.032	0.573	0.013	0.225	0.455	0.011	-0.102	0.094	0.004	0.053	0.025	0.002		
POPWEIN	0.076	0.973	0.124	-0.060	0.008	0.002	0.656	0.933	0.423	-0.123	0.033	0.024		
DUMMY1	0.075	0.998	0.102	0.212	0.118	0.023	0.190	0.095	0.035	0.546	0.785	0.482		
DUMMY2	0.048	0.529	0.019	-0.153	0.208	0.008	0.127	0.143	0.010	-0.142	0.179	0.021		
DUMMY3	0.079	0.950	0.079	-0.104	0.038	0.006	-0.256	0.231	0.067	-0.439	0.680	0.325		



Abbildung 4: Grafische Darstellung der Korrespondenzanalyse mit einer Dummyvariablen, erste gegen dritte Achse



JUNGM	Männer bis 39 Jahre	SSPJA	kann Schach spielen
MITTM	Männer von 40 bis 59 Jahren	SSPETWAS	kann etwas Schach spielen
ALTM	Männer ab 60 Jahre	SSPNEIN	kann nicht Schach spielen
JUNGW	Frauen bis 39 Jahre	HUNJA	kann Hosen umnähen
MITTW	Frauen von 40 bis 59 Jahren	HUNETWAS	kann etwas Hosen umnähen
ALTW	Frauen ab 60 Jahre	HUNNEIN	kann nicht Hosen umnähen
WZJJA	kann Walzer tanzen	POPJA	kann zu Popmusik tanzen
WZETWAS	kann etwas Walzer tanzen	POPETWAS	kann etwas zu Popmusik tanzen
WZNEIN	kann nicht Walzer tanzen	POPNEIN	kann nicht zu Popmusik tanzen



Werden die quadrierten Korrelationen (QCOR) der beschreibenden Merkmale (der kulturellen Fähigkeiten) der ersten Analyse mit der zuletzt durchgeführten verglichen, so fällt auf, daß diese sich nur unbedeutend unterscheiden. Demgegenüber laden bei den Subpopulationen die "älteren Männer" nun auf der dritten Achse, positiv korreliert mit der dritten und negativ korreliert mit der ersten Ausprägung der Dummyvariablen. Inhaltlich bedeutet dies, daß diese Subpopulation am besten durch die (künstlichen) Variablenausprägungen beschrieben werden kann.

Die Wichtigkeit der Dummyvariablen zur Beschreibung der dritten Achse wird auch aus der Spalte INR3 ersichtlich, wo zu erkennen ist, daß diese Variable 82,2 Prozent (es sind die drei Werte aufzuaddieren) der dritten Achse determiniert. Zur grafischen Illustration ist die dritte Achse gegen die erste Achse geplottet worden (Abbildung 4).

5. Fazit

Werden die dargestellten Ergebnisse insgesamt betrachtet, so ist die hohe Robustheit des Verfahrens gegenüber zusätzlichen Variablen beachtlich. Einmal gefundene Zusammenhänge zwischen Variablenausprägungen bleiben auch nach Aufnahme zusätzlicher Variablen erhalten. Dies gilt sogar, wenn eine neue Variable aufgenommen wird, die eine zusätzliche Dimension definiert. Im letzten Fall kann sich aber die Zuordnung von Spalten- zu Zeilenvariablen(ausprägungen) ändern, wenn, wie im vorgestellten Beispiel, eine Subpopulation durch eine der zusätzlich aufgenommenen Variablenausprägungen besser beschrieben werden kann als durch die zuvor berücksichtigten.

Die vorgestellten Ergebnisse sollten auch eine Aufforderung sein, weiter über die Korrespondenzanalyse nachzudenken und ihre Anwendungsmöglichkeiten - insbesondere im Bereich der Sozialwissenschaften - auszunutzen. Mit der Korrespondenzanalyse ist ein robustes Instrument zur explorativen Datenanalyse vorhanden. Die Daten müssen hierbei nicht, wie z.T. bei anderen multivariaten Verfahren, auf ein höheres Skalenniveau gebracht werden, um den Anwendungsvoraussetzungen der Methode zu genügen.

Anmerkungen

- 1) Der Datensatz ist im Zentralarchiv für empirische Sozialforschung unter der Studiennummer 1500 archiviert.
- 2) Zur optischen Unterscheidung wurden die sechs Subpopulationen in der grafischen Darstellung nachträglich unterstrichen.



- 3) Um zu übersichtlichen grafischen Darstellungen zu gelangen, wurden die Ordinaten in dieser und in den folgenden Abbildungen gestreckt. Die in den jeweils nachfolgenden Tabellen angegebenen quadrierten Korrelationen der Variablenausprägungen mit den einzelnen Achsen sind daher nicht direkt aus den Grafiken ablesbar.
- 4) Auf die durchaus mögliche Aufnahme weiterer Achsen in die Tabelle 'GENERAL STATISTIC' wurde verzichtet, da mittels der ersten beiden bereits über 96 Prozent der Gesamtvarianz des Modells erklärt werden.
- 5) Die Beschränkung auf drei Ausprägungen pro Variable basiert einzig auf inhaltlichen Überlegungen. Es ist durchaus möglich, Variablen mit einer unterschiedlichen Anzahl von Ausprägungen als Eingabematerial zu verwenden (z.B. GREENACRE 1984, BLASIUS 1987).
- 6) Die nachträglich aufgenommenen Variablen haben einen Anteil von 44,1 Prozent an der Determination des gesamten Raumes, dieser Wert ergibt sich durch Addition der letzten zwölf Werte in der Spalte INR.

Literatur

BENZÉCRI, J.-P. 1964. Course de Linguistique Mathématique. Paris: Publications de l'Institut de Statistique de l'Université de Paris.

BLASIUS, J. 1987. Einstellungen zur Hamburger Innenstadt. Eine Auswertung mit Hilfe der Korrespondenzanalyse.
ZA-Information, 21,29-51.

BLASIUS, J. & ROHLINGER, H. 1988. KORRES - Ein multivariates Programm zur Auswertung von zweidimensionalen Kontingenztabellen.
In: FAULBAUM, F. & UEHLINGER, H. (Hrsg.):
Fortschritte der Statistik - Software 1. Stuttgart: Gustav Fischer Verlag, 387-397.

BOURDIEU, P. 1982. Die feinen Unterschiede. Frankfurt: Suhrkamp.

DANGSCHAT, J. & BLASIUS, J. 1987. Social and Spatial Disparities in Warsaw in 1978: An Application of Correspondence Analysis to a 'Socialist City'. Urban Studies, 24, 173-191.

ESCOFIER-CORDIER, B. 1965. L'Analyse des Correspondances. Thèse de Doctorat. Rennes: Université de Rennes.



GREENACRE, M. 1984. Theory and Applications of Correspondence Analysis. London: Akademie Press.

GUTTMAN, L. 1941. The Quantification of a Class of Attributes: A Theory and Method of Scale Construction.

In: HORST, P. (Eds.): The Prediction of Personal Adjustment. New York: Social Science Research Council, 319-348.

GUTTMAN, L. 1950. The Principal Components of Scale Analysis.

In: STOUFFER, S.S. et al. (Eds.): Measurement and Prediction, Princeton, New Jersey: Princeton University Press, 312-361.

HAYASHI, C. 1950. On the Quantification of Qualitative Data from the Mathematico - Statistical Point of View. *Annals of the Institute of Statistical Mathematics*, 2, 35-47.

HAYASHI, C. 1952. On the Prediction of Phenomena from Qualitative Data and the Quantification of Qualitative Data from the Mathematico-Statistical Point of View. *Annals of the Institute of Statistical Mathematics*, 3, 69-98.

HAYASHI, C. 1954. Multidimensional Quantification - with the Applications to Analysis of Social Phenomena. *Annals of the Institute of Statistical Mathematics*, 5, 231-245.

HIRSCHFELD, H. O. 1935. A Connection between Correlation and Contingency. *Proceedings of the Cambridge Philosophical Society*, 31, 520-524.

LUDOVIC, L., MORINEAU, A & KENNETH, M. W. 1984. Multivariate Descriptive Statistical Analysis. Correspondence Analysis and Related Techniques for large Matrices. New York: John Wiley.

THIESSEN, V. & ROHLINGER, H. 1988. Die Verteilung von Aufgaben und Pflichten im ehelichen Haushalt. Erscheint in: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 40, Heft 4.