ALMO
Statistiksystem

Korrespondenzanalyse
Nominale Faktorenanalyse

P30.8

Kurt Holm

Almo Statistik-System
http://www.almo-statistik.de/
kurt.holm@jku.at

2014

Autor: em. Prof. Dr. Kurt Holm, Universität Linz, Österreich

Im Text wird häufig auf das Dokument **P0** Bezug genommen. Dabei handelt es sich um das Almo-Dokument "Arbeiten mit Almo.PDF" (Dokument 0).

Weitere Almo-Dokumente

Die folgenden Dokumente können alle kostenlos von der Handbuchseite in

http://www.almo-statistik.de

heruntergeladen werden

0. Arbeiten_mit_Almo.PDF (1 MB)
1a. Eindimensionale Tabellierung.PDF (1,8 MB)
1b. Zwei- und drei-dimensionale Tabellierung.PDF (1.1 MB)
2. Beliebig-dimensionale Tabellierung.PDF (1.7 MB)
3. Nicht-parametrische Verfahren.PDF (0.9 MB)
4. Kanonische Analysen.PDF (1.8 MB)
Diskriminanzanalyse.PDF (1.8 MB)
enthält: Kanonische Korrelation, Diskriminanzanalyse, bivariate Korrespondenzanalyse, optimale Skalierung
5. Korrelation.PDF (1.4 MB)
6. Allgemeine multiple Korrespondenzanalyse.PDF (1.5 MB)
7. Allgemeines ordinales Rasch-Modell.PDF (0.6 MB)
7a. Wie man mit Almo ein Rasch-Modell rechnet.PDF (0.2 MB)
8. Tests auf Mittelwertsdifferenz, t-Test.PDF (1,6 MB)
9. Logitanalyse.pdf (1,2MB) enthält Logit- und Probitanalyse
9b. Bootstrap bei Logit- und Probitanalyse.pdf
10. Koeffizienten der Logitanalyse.PDF (0,06 MB)
11. Daten-Fusion.PDF (1,1 MB)
12. Daten-Imputation.PDF (1,3 MB)
13. ALM Allgemeines Lineares Modell.PDF (2.3 MB)
13a. ALM Allgemeines Lineares Modell II.PDF (2.7 MB)
13b. Bootstrap bei Allgemeinem Linearen Modell III.PDF
14. Ereignisanalyse: Sterbetafel-Methode,Kaplan-Meier,Cox-Regression(1,5MB)
15. Faktorenanalyse.PDF (1,6 MB)
15a. Bootstrap bei Faktorenanalyse.PDF
16. Konfirmatorische Faktorenanalyse.PDF (0,3 MB)
17. Clusteranalyse.PDF (3 MB)
18. Pisa 2012 Almo-Daten und Analyse-Programme.PDF (17 KB)
19. Guttman- und Mokken-Skalierung.PFD (0.8 MB)
20. Latent Structure Analysis.PDF (1 MB)
21. Statistische Algorithmen in C (80 KB)
22. Conjoint-Analyse (PDF 0,8 MB)
23. Ausreisser entdecken (PDF 170 KB)
24. Statistische Datenanalyse Teil I, Data Mining I
25. Statistische Datenanalyse Teil II, Data Mining II
26. Statistische Datenanalyse Teil III, Arbeiten mit Almo-Datenanalyse
27. Mehrfachantworten. Tabellierung von Fragen mit Mehrfachantworten
28. Metrische multidimensionale Skalierung (MDS) (0,4 MB)
29. Metrisches multidimensionales Unfolding (MDU) (0,6 MB)
30. Nicht-metrische multidimensionale Skalierung (MDS) (0,4 MB)
31. Pfadanalyse.PDF (0,7 MB)
32. Datei-Operationen mit Almo (1,1 MB)
33. Wählerstromanalyse und Wahlhochrechnung (1,6 MB)
34. Soziometrie. Auswertung soziometrischer Daten (0,5 MB)
35. Konfidenzintervall und p-Wert beim Bootstrap-Verfahren (200 KB)

Inhaltsverzeichnis

P30.8 Nominale Faktorenanalyse

Dichotome nominale Variable stellen für die Faktorenanalyse kein unüberwindbares Problem dar. Sie werden wie quantitative Variable behandelt. Arminger, 1979, S.157ff. stellt den Kalkül dar und diskutiert auch ausführlich die Probleme.

Für die Behandlung polytomer (und auch dichotomer), nominaler Variabler werden in Almo folgende Verfahren zur Verfügung gestellt:

1. Die "einfache" Faktorisierung der Dummies.
2. Das "Blockdiagonal-Verfahren" nach McDonald
3. Die multiple Korrespondenzanalyse

P30.8.0 Ein Beispiel

Mit dem Maskenprogramm Prog30m5 können alle drei Verfahren der nominalen Faktorenanalyse gerechnet werden.

Im Maskenprogramm wird folgendes Beispiel gerechnet:
Die nominalen Variablen, die faktorisiert werden sollen, sind:

```
Autokauf: (1)Porschen     (2)Mercedes      (3)VW
Beruf:    (1)Selbständig  (2)Arbeitnehmer  (3)Führungsposition
Fahrstil: (1)Aggressiv    (2)normal        (3)zurückhaltend
```

Die Datenmatrix ist folgende:

D.satz Nr.	V1 Auto	V2 Beruf	V3 Fahrstil
1	1	1	1
2	1	1	1
3	1	2	2
4	1	3	1
.	.	.	.
.	.	.	.
.	.	.	.
32	3	2	3
33	3	3	1
34	3	3	3

Eine nominale Faktorenanalyse nach dem Verfahren der multiplen Korrespondenzanalyse über diese 3 nominalen Variablen könnte zu einem 2- oder 3-dimensionalen Raum führen, in dem z.B. folgende Punkte dicht beieinander sind:

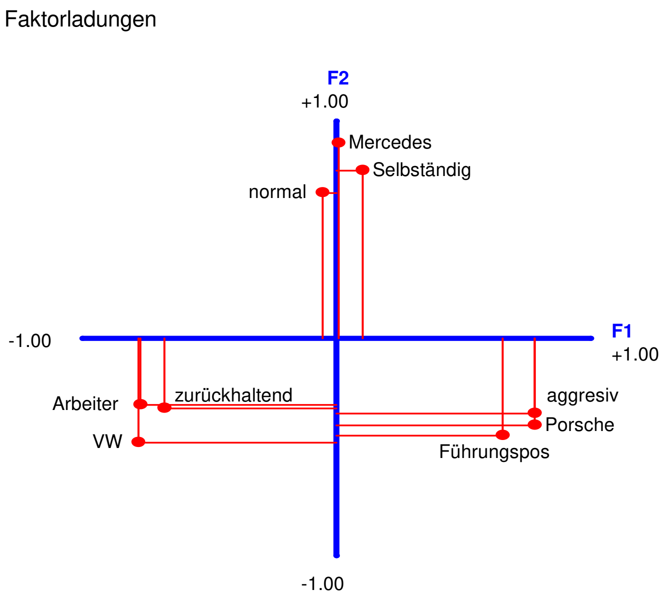
```
Mercedes, Selbständiger,     normaler Fahrstil
Porsche,  Führungsposition, aggresiver Fahrstil
VW,       Arbeitnehmer,      zurückhaltend
```

Bei 2 (vorgegebenen) Faktoren liefert Almo folgende Ladungsmatrix

		Faktor 1	Faktor 2
Auto	Porsche	0.7785	-0.4059
	Mercedes	0.0123	0.8955
	VW	-0.7742	-0.4792
Beruf	Selbstän	0.1056	0.7743
	Arbeiter	-0.7708	-0.3108
	Führungs	0.6512	-0.4537
Fahrstil	aggresiv	0.7778	-0.3468

normal	-0.0579	0.6719
zurückha	-0.6735	-0.3217

Grafisch dargestellt:



Damit wären 3 Typen von Autofahrern identifiziert. Dies ist auch wesentlich die Aufgabe der nominalen Faktorenanalyse:

Das Identifizieren von Typen und deren grafische Darstellung

Das wird im empirischen 3-Variablen-Beispiel von Carrol in Abschnitt P30.8.3.2 - wie wir meinen - eindrucksvoll illustriert. Das obige Autofahrer-Beispiel beruht auf mühsam konstruierten Daten.

Die nominale Faktorenanalyse ist ein Konkurrent zur Clusteranalyse. Das nachfolgende Almo-Programm bietet deswegen auch eine Option an, mit der eine Clusteranalyse für dieselben Daten "mit einem Mausklick" gerechnet werden kann. Siehe nachfolgend Abschnitt P30.8.3.7.

Optional können Faktorwerte berechnet werden und an die Datensätze als zusätzliche Variable angefügt und in eine neue Datei geschrieben werden. Im Rahmen der Korrespondenzanalyse werden die Faktorwerte "object scores" genannt.

Programm-Maske Prog30m5
Obige Ergebnisse entstanden aus Prog30m5. Diese Programm-Maske wird gefunden nach Klick auf den Knopf *Verfahren* in der Knopfleiste am oberen Rand des Almo-Fensters. Dann Klick auf den Eintrag *Korrespondenzanalyse.*

Prog30m5.Msk
Faktorenanalyse mit nominalen Variablen
mit Optionen

Die nominale Faktorenanalyse kann mit folgenden Verfahren gerechnet werden:

1. als multiple Korrespondenzanalyse

2. mit dem Blockdiagonal-Matrix-Verfahren nach McDonald

3. als gewöhnliche Faktorenanalyse der in Dummies aufgelösten nominalen Variablen

Beispiel: Die nominalen Variablen, die faktorisiert werden sollen, sind:

Autokauf: Porschen, Mercedes, VW
Beruf: Selbständig, Arbeitnehmer, Führungsposition,
Fahrstil: Aggresiv, normal, zurückhaltend

Eine nominale Faktorenanalyse über diese 3 nominalen Var. könnte zu einem 2- oder 3-dimensionalen Raum führen, in dem z.B. folgende Punkte dicht beieinander sind:

Mercedes,	Selbständiger,	normaler Fahrstil
Porsche,	Führungsposition,	aggresiver Fahrstil
VW,	Arbeitnehmer,	zurückhaltend

Optional können Faktorwerte berechnet werden und an die Datensätze als zusätzliche Variable angefügt und in eine neue Datei geschrieben werden. Im Rahmen der Korrespondenzanalyse werden die Faktorwerte "object scores" genannt

---

Prog30m5 kann auch für die Eingabe bereits fertiger Tabellen verwendet werden. Siehe Handbuch, P30.8.3.1.1
Hilfe

Was ist ein Kurzprogramm ? --> Hilfe
Bedienung --> Hilfe

1 Speicher fuer x Variable | Hilfe
Vereinbare Variable= 20 ;

2 Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

3 Datei der Variablennamen | Hilfe
"C:\Almo7\TESTDAT\Auto.nam"
zeige — zeige = Namensdatei in Output zeigen
leer = nicht

4 Freie Namensfelder | Hilfe
Name1=Auto:Porsche,Mercedes,VW;
Name2=Beruf:Selbständig,Arbeiter,Führungspos;
Name3=Fahrstil:aggresiv,normal,zurückhaltend;
erzeuge zusätzliche Namensfelder

Datei aus der gelesen wird

Hilfe bei Datei-Problemen

"C:\Almo7\Testdat\Auto.fre"

5 frei — Format der Daten — Hilfe

V1:3 — der Datensatz enthält diese Variablen
Bei Format DIREKT schreiben Sie: alle_V

6 Wenn Dateiformat FIX oder Nicht-Standard-FREI — Hilfe

zu faktorisierende nominale Variable — Hilfe

7 Auto,Beruf,Fahrstil

Verfahren der nominalen Faktorenanalyse

multiple_Korrespondenz

8 0= gewöhnliche Faktorenanalyse der in Dummies aufgelösten nominalen Variablen

Blockdiagonale = nominale Faktorennalyse mit Blockdiagonal-Matrix nach McDonald

multiple_Korrespondenz = multiple Korrespondenz-Analyse

9 Option: Ein- und Ausschliessen von Untersuchungseinheiten

10 Option: Umkodierungen und Kein-Wert-Angaben

11 Option: Spezielle Kein-Wert-Behandlung

12 Option: Untersuchungseinheiten gewichten

13 Option: Faktoren

14 Option: Rotation

15 Option: verschiedene Programm-Optionen

16 Option: Faktorwerte ermitteln und speichern

17 Option: "Aussehen" der auszugebenden Tabelle bzw. Matrix

18 Grafik-Optionen

19 Programmende

Erläuterungen zu den Eingabeboxen:
Das Programm entspricht weitgehend dem Prog30m2 für die quantitative Faktorenanalyse. Siehe hierzu das Almo-Dokument Nr. 15 "Faktorenanalyse".

Eingabebox 1 und 2: Vereinbarungen
Siehe Almo-Dokument Nr.0 „Arbeiten mit Almo“, Abschnitt P0.1 und P0.2.

Eingabebox 3 und 4: Variablennamen
Siehe Almo-Dokument Nr.0 „Arbeiten mit Almo“, Abschnitt P0.3.

Eingabebox 5 und 6: Datei
Siehe Almo-Dokument Nr.0 „Arbeiten mit Almo“, Abschnitt P0.4.

Eingabebox 7: Zu faktorisierende nominale Variable

Die Variable dürfen polytom und auch dichotom sein. Sie müssen nicht unbedingt ganzzahlig sein. Siehe hierzu den Hilfeknopf in der Box. Siehe auch Almo-Dokument Nr.0 „Arbeiten mit Almo“, Abschnitt P0.11.

Eingabebox 8: Verfahren der nominalen Faktorenanalyse

Hier wird das Verfahren selektiert, mit dem die nominalen Variablen faktorisiert werden sollen. Möglich sind:

```
        Eingabe
-----------------------
        0              = gewöhnliche Faktorenanalyse der in
                         Dummies aufgelösten nominalen Variablen

Blockdiagonale         = nominale Faktorennalyse mit
                         Blockdiagonal-Matrix nach McDonald

multiple_Korrespondenz = multiple Korrespondenz-Analyse
```

Eingabebox 9: Ein- und Ausschliessen von Untersuchungseinheiten
Siehe Almo-Dokument Nr.0 „Arbeiten mit Almo“, Abschnitt P0.7.

Eingabebox 10: Umkodierungen und Kein-Wert-Angaben
Siehe Almo-Dokument Nr.0 „Arbeiten mit Almo“, Abschnitt P0.5.

Eingabebox 11: Spezielle Kein-Wert-Behandlung
Almo bietet 7 Verfahren an, mit fehlenden Werten umzugehen. Zu berücksichtigen ist dabei, dass alle Analysevariable nominal sind. Siehe dazu Almo-Dokument Nr. 15 "Faktorenanalyse", Abschnitt P30.3.1, Eingabebox 10. Dort werden diese Verfahren ausführlich dargestellt.

0 *Analysevariable enthalten keine Kein-Wert-Fälle*

1 *Paarweises Ausscheiden (Voreinstellung)*

2 *Vollständiges Ausscheuden des gesamten Datensatzes, wenn nur eine der nominale Variablen einen Kein_Wert-Fall besitzt*

3 *wie 2*

4 *die zum Erwartungswert nächste empirisch vorkommende Codeziffer wird eingesetzt*

Die Berechnung des Erwartungswerts soll an einem Beispiel gezeigt werden. Die nominale Variable sei der Beruf mit den 3 Ausprägungen Arbeiter, Angestellte, Sonstige. Dabei wurden folgende Häufigkeiten ermittelt.

	Code	Häufigkeit	Anteil	Code*Anteil
Arbeiter	1	250	0.25	0.25
Angestellte	2	400	0.40	0.80
Sonstige	3	350	0.35	1.05

Summe				2.10

Der Erwartungswert ist 2.1. Die nächste empirisch vorkommende Codeziffer ist 2. Der KW-Einsetzungswert ist also 2

Ist die nominale Variable dichotom, dann ist der Kein-Wert-Einsetzungswert gleich der Codeziffer der häufigsten Ausprägung.

5 *wie 4*

6 *Der wahrscheinlichste Ausprägungswert wird eingesetzt.*

Die Vorgehensweise soll an einem Beispiel gezeigt werden. Die nominale Variable sei der Beruf mit den 3 Ausprägungen. Arbeiter, Angestellte, Sonstige. Dabei wurden folgende Häufigkeiten ermittelt.

	Code	Häufigkeit	in %	in % kummuliert
Arbeiter	1	250	25	25
Angestellte	2	400	40	65
Sonstige	3	350	35	100

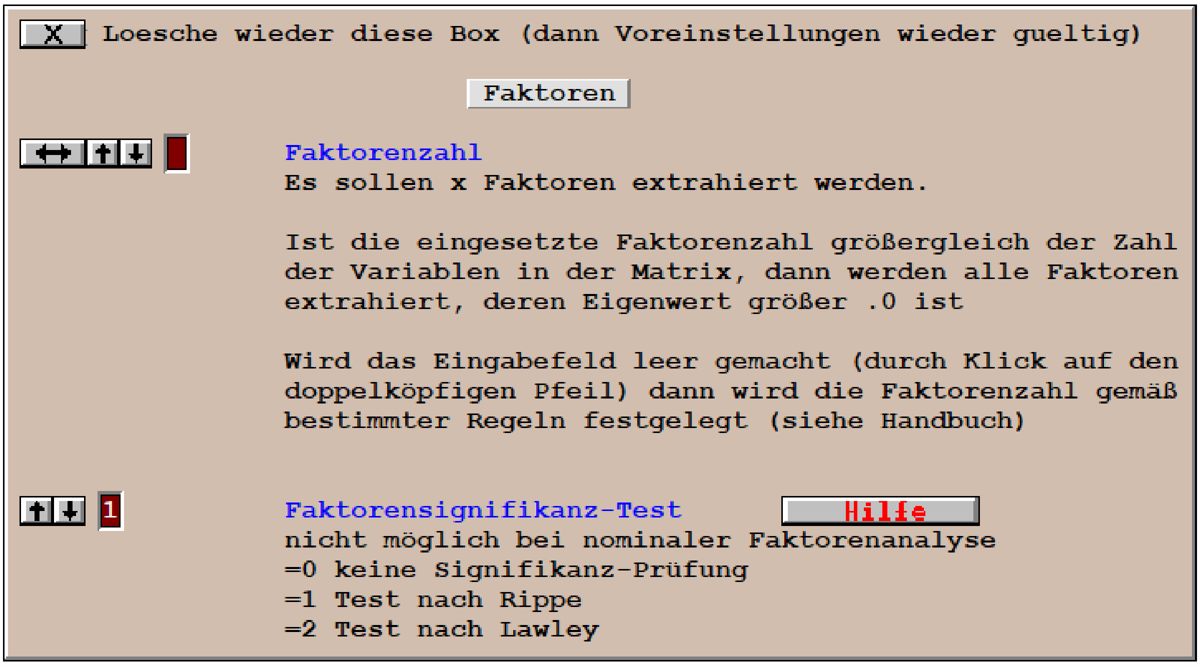
Dann wird eine gleichverteilte Zufallszahl zwischen 0 und 100 erzeugt.
Liegt sie zwischen

0	und	25, dann wird für den fehlenden Wert	1 eingesetzt
25		65	2
65		100	3

7 *wie 6*

Eingabebox 12: Untersuchungseinheiten gewichten
Siehe Almo-Dokument Nr.0 „Arbeiten mit Almo“, Abschnitt P0.8.

Eingabebox 13: Faktoren
Bleibt die Optionsbox geschlossen oder wird sie geöffnet und das Eingabefeld bleibt leer ….



… dann verfährt Almo so:

-- bei Korrepodenzanalyse und Blockdiagonal-Verfahren werden so viele Faktoren extrahiert wie nominale Variable vorhanden sind – unabhängig davon wie viele Ausprägungen vorhanden sind. Häufig werden dabei zu viele Faktoren extrahiert, so dass der Benutzer eine kleinere Faktorenzahl vorgegeben muss
-- bei „gewöhnlicher Faktorenanalyse mit Dummies (Eingabe „0“) werden so viele Faktoren extrahiert wie Eigenwerte größer 1.0 ermittelt werden können.

Wird eine Faktorenzahl vom Benutzer eingegeben, dann wird diese Faktorenzahl auch extrahiert. Empfehlenswert ist es dabei mehrere Faktorenzahlen auszuprobieren und die Lösung zu akzeptieren, die inhaltlich überzeugend ist.

Siehe dazu Almo-Dokument 15 "Faktorenanalyse", Abschnitt P30.3.7 in den Erläuterungen zu Prog30m2:

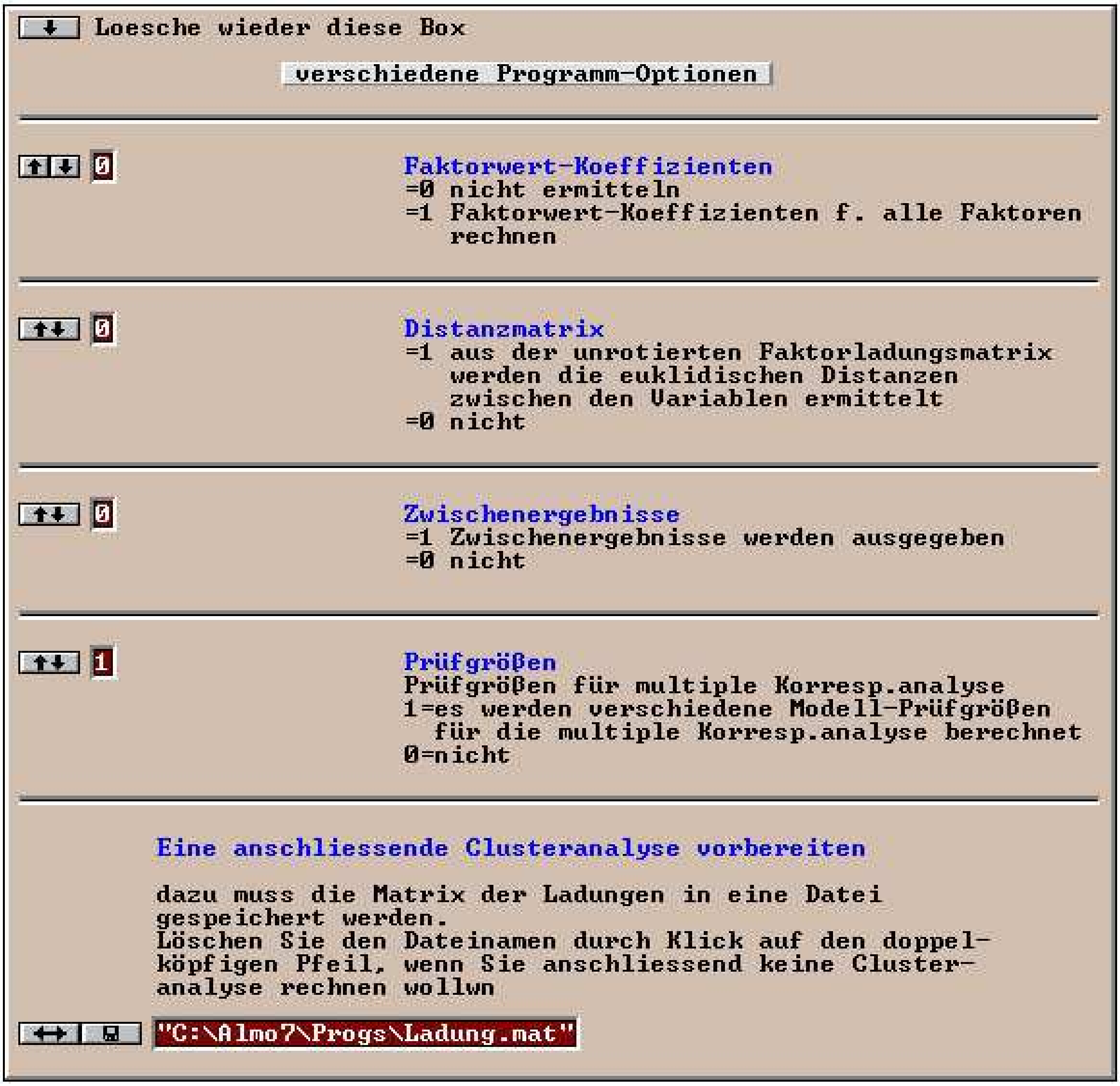
Eingabebox 14: Rotation

Eine recht- oder schiefwinklige Rotation oder eine Benutzer-definierte Rotation der Korrespondenzanalyse ist zwar möglich. Sie liefert allerdings selten Ergebnisse, die inhaltlich besser zu interpretieren sind als die der orthogonalen Analyse. Es macht aber Sinn, eine Rotationslösung zu probieren.

Siehe die ausführliche Darstellung der Rotationsverfahren im Almo-Dokument 15 "Faktorenanalyse", Abschnitt P30.1.4

Eingabebox 15: Option: verschiedene Programm-Optionen

Optionsbox geöffnet:



Eingabefeld 1, 2 und *3* ist in gleicher Weise beim Programm Prog30m2 in der Eingabebox "Optionen: Weitere Optionen" enthalten. Siehe dazu Almo-Dokument 15 "Faktorenanalyse", Abschnitt P30.3.7 in den Erläuterungen zu Prog30m2

Eingabefeld 4: Prüfgrößen
Prüfgrößen werden nur ermittelt, wenn als Verfahren in der Eingabebox 8 die multiple Korrespondenzanalyse eingesetzt wurde. Diese Prüfgrößen werden in Abschnitt P30.8.3.4 ausführlich erläutert.
Wird "0" eingegeben, dann werden keine Prüfgrößen ermittelt. Wird "1" eingegeben, dann werden verschiedene Modell-Prüfgrößen für die multiple Korresp.analyse berechnet

Eingabefeld 5: Eine anschliessende Clusteranalyse vorbereiten Wir haben oben ausgeführt, dass es die Aufgabe der nominalen Faktorenanalyse ist Typen (in unserem Beispiel: von Autofahrern) zu identifizieren. Dies ist auch die Aufgabe der Clusteranalyse. In Abschnitt P30.8.3.7 geht Johann Bacher ausführlich auf das Verhältnis von nominaler Faktorenanalyse und Clusteranalyse ein. Prog30m5 ermöglicht es nun, die Matrix der Faktorladungen so zu speichern, dass sie als

Eingabe für das Clusteranalyse-Programm verwendet werden kann. Schreiben Sie zu diesem Zwecke einen Dateinamen in das Eingabefeld 5.

Eingabebox 16: Faktorwerte ermitteln und speichern
Siehe dazu P30.3.8 in den Erläuterungen zu Prog30m2. Im späteren Abschnitt P30.8.3.2.1 geben wir ein Beispiel einer Faktorwert-Berechnung.

Die Faktorwertberechnung ist eigentlich nur dann sinnvoll, wenn die extrahierten (und eventuell rotierten) Faktoren inhaltlich interpretierbar sind. In unserem Beispiel ist dies kaum der Fall. Die Faktorwertberechnung ist aber dann sinnvoll, wenn es darum geht, die Individuen im Faktorenraum abzubilden, um ihre räumliche Nähe zu den identifizierten Typen abzuschätzen (siehe dazu Abschnitt P30.8.3.6).

Eingabebox 17: "Aussehen" der auszugebenden Tabelle bzw. Matrix
Siehe dazu Almo-Dokument Nr.0 „Arbeiten mit Almo", Abschnitt P0.9.

Eingabebox 18: Grafik-Optionen
Siehe dazu Almo-Dokument Nr.0 „Arbeiten mit Almo", Abschnitt.10.

P30.8.1 Die einfache Faktorisierung der Dummies

Die polytomen Variablen werden in 0-1 kodierte Dummies aufgelöst. Für diese wird die durchschnittliche Kreuzprodukte-, oder die Kovarianz-, oder die Korrelationsmatrix gebildet. Diese wird faktorisiert.

P30.8.1.1 Eingabe mit Maskenprogramm

Wir verwenden folgendes Beispiel:

Die nominalen Variablen, die faktorisiert werden sollen, sind:

```
    Autokauf: Porsche, Mercedes, VW,
    Beruf:    Selbständig, Arbeitnehmer, Führungsposition,
    Fahrstil: Aggressiv, normal, zurückhaltend
```

Wir rechnen zunächst das Verfahren der "einfachen" Faktorisierung der Dummies. Die drei nominalen Variablen werden entsprechend der Zahl ihrer Ausprägungen in je drei 0-1 kodierte Dummies aufgelöst. Im Prog30m5 muss in der Eingabebox „Verfahren der nominalen Faktorenanalyse" „0" eingetragen werden.

Eine nominale Faktorenanalyse über diese 3 nominalen Variablen könnte zu einem 2- oder 3-dimensionalen Raum führen, in dem z.B. folgende Punkte dicht beieinander sind:

```
Punktewolke 1:   Mercedes, Selbständiger,    normaler Fahrstil
            2:   Porsche,  Führungsposition, aggresiver Fahrstil
            3:   VW,       Arbeitnehmer,     zurückhaltend
```

Ausgabe

ALMO liefert aus dem Maskenprogramm Prog30m5 folgendes Ergebnis, das wir hier verkürzt wiedergeben.

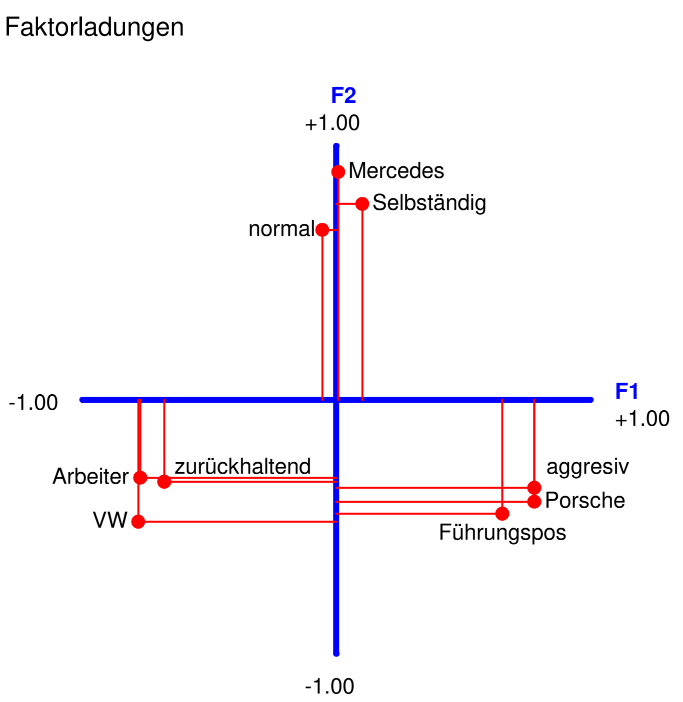
```
Ergebnisse aus Faktorenanalyse
------------------------------
```

```
3 Eigenwerte der Korrelationsmatrix sind groesser 1.0
Entsprechend versucht Almo 3 Faktoren fuer die nachfolgende
Faktorenanalyse zu extrahieren
-----------------------------------------------------------------

Eigenwerte je Faktor
    3.2968      2.7735      1.2570

Matrix der Faktorladungen
```

			Faktor 1	Faktor 2	Faktor 3
Auto	Porsche	V1-1	0.7784	-0.4059	-0.1519
Auto	Mercedes	V1-2	0.0123	0.8954	0.1561
Auto	VW	V1-3	-0.7741	-0.4792	-0.0041
Beruf	Selbstän	V2-1	0.1055	0.7742	0.4566
Beruf	Arbeiter	V2-2	-0.7708	-0.3108	-0.4087
Beruf	Führungs	V2-3	0.6512	-0.4537	-0.0469
Fahrst	aggresiv	V3-1	0.7777	-0.3467	0.1425
Fahrst	normal	V3-2	-0.0578	0.6718	-0.7120
Fahrst	zurückha	V3-3	-0.6735	-0.3216	0.5517

Wird die Faktorladungsmatrix grafisch dargestellt, so erhalten wir für die 2- und 3-dimensionale Lösung folgendes Bild.

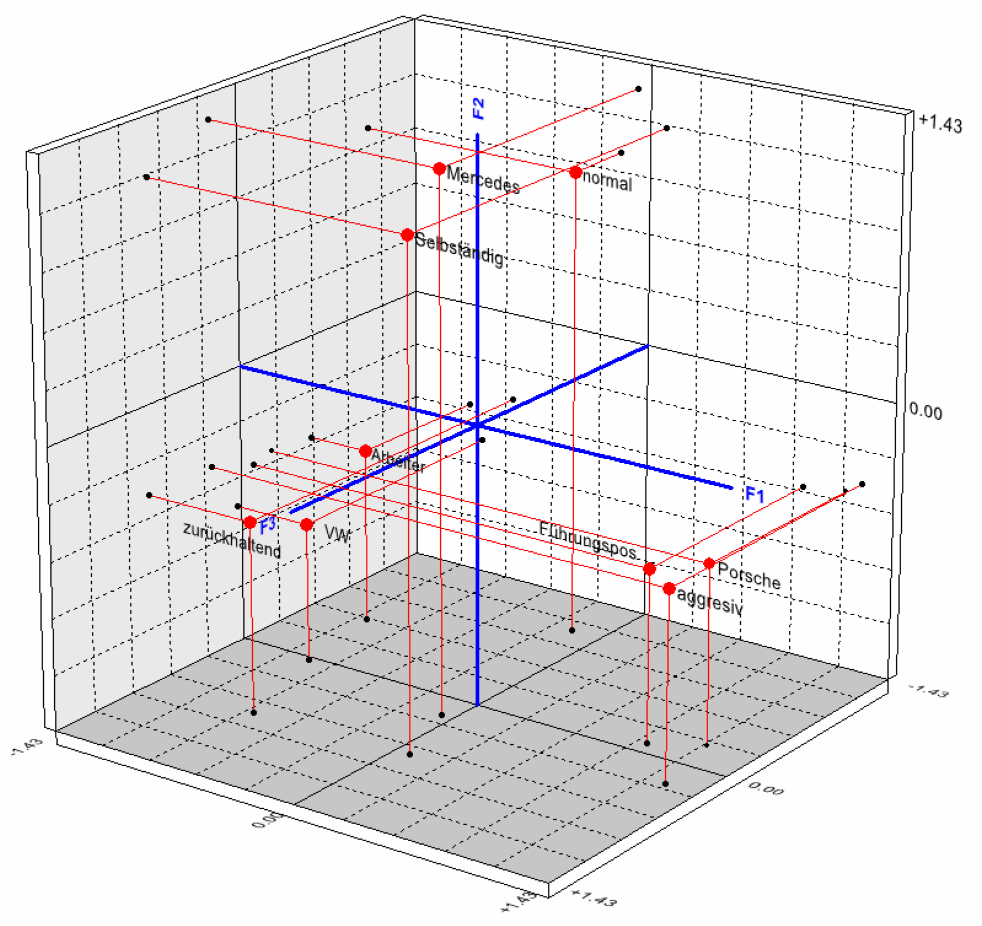
2-dimensionale Darstellung



Erläuterung: Es sind deutlich 3 Typen von Autofahrern zu erkennen:

```
Typ 1:   Mercedes, Selbständiger,     normaler Fahrstil
    2:   Porsche,  Führungsposition, aggresiver Fahrstil
    3:   VW,       Arbeitnehmer,     zurückhaltend
```

3-dimensionale Darstellung



Erläuterung: Es sind die oben beschriebenen 3 Autofahrer-Typen deutlich erkennbar. Die 3. Dimension bringt bei diesem Beispiel keine zusätzliche Erkenntnis.

Das Problem dieser Art nominaler Faktorenanalyse ist es, dass die Abhängigkeit, die zwischen den Dummies einer nominalen Variablen "naturgegeben" bestehen nicht berücksichtigt werden - wie dies jedoch explizit beim nachfolgend beschriebenen Blockdiagonal-Verfahren geschieht.

P30.8.2 Das "Blockdiagonal-Verfahren" nach McDonald

Wir folgen hier der kurzen und übersichtlichen Darstellung bei Arminger (1979, S.162ff).

Beim Blockdiagonalverfahren wird (im Unterschied zur „einfachen Faktorisierung der Dummies) berücksichtigt, dass die Dummies einer nominalen Variablen korrelieren. Die normale faktorenanalytische Zerlegung ist folgende (siehe auch Gleichung 3a in P30.1.2)

$$S^* = L \bullet L' + D$$

wobei **S*** = reproduzierte Korrelations- oder Kovarianzmatrix
L = Faktorladungsmatrix
D = Diagonalmatrix. In ihrer Diagonale steht $1.0 - h_i^2$ wobei h_i^2 die (geschätzte) Kommunalität der Variablen i ist.

Beim Blockdiagonalverfahren ist D nun folgende Blockdiagonalmatrix (für den Fall, dass 3 nominale Variable faktorisiert werden)

$\mathbf{D}_{11}$	0	0
0	$\mathbf{D}_{22}$	0
0	0	$\mathbf{D}_{33}$

$\mathbf{D}_{11}$ = Matrix der Dummies der nominalen Variablen 1
$\mathbf{D}_{22}$ = Matrix der Dummies der nominalen Variablen 2
$\mathbf{D}_{33}$ = Matrix der Dummies der nominalen Variablen 3

Im nachfolgend beschriebenen Rechenschritt 6 wird gezeigt, wie die Matrix **D** gewonnen wird. Vereinfacht kann gesagt werden: Beim Blockdiagonal-Verfahren nach McDonald wird die Inter-Korrelation der Dummies einer nominalen Variablen explizit berücksichtigt.

Wir wollen ein Beispiel aufgreifen, an dem McDonald sein Verfahren vorführte. Das Almo-Syntxprogramm dazu ist unter dem Namen "Nom_Fak.Alm" zu finden durch Klick auf den Knopf *alle Progs* in der Knopfleiste am Oberrand des Almo-Fensters. An 100 Männern (Rekruten der Schweizer Armee) wurden die 4 nominalen Variablen Haarfarbe, Augenfarbe, Schädelform, Körpergröße gemessen. Das Beispiel wird bei Arminger (1979, S. 168) detailliert ausgeführt.

Folgende Kovarianzmatrix der Variablenausprägungen wurde ermittelt.

Kovarianz-Matrix

	Haarfarb blond	Haarfarb rot	Haarfarb dunkel	Augenfar hell	Augenfar gemischt	Augenfar braun
Haarfarb blond	0.171	-0.033	-0.138	0.067	-0.019	-0.048
Haarfarb rot	-0.033	0.127	-0.094	0.030	-0.004	-0.026
Haarfarb dunkel	-0.138	-0.094	0.233	-0.097	0.023	0.074
Augenfar hell	0.067	0.030	-0.097	0.221	-0.118	-0.102
Augenfar gemischt	-0.019	-0.004	0.023	-0.118	0.230	-0.111
Augenfar braun	-0.048	-0.026	0.074	-0.102	-0.111	0.213
Schaedel schmal	-0.011	0.006	0.005	0.042	-0.048	0.006
Schaedel weit	0.011	-0.006	-0.005	-0.042	0.048	-0.006
Koerperg gross	0.035	0.035	-0.070	0.148	-0.054	-0.093
Koerperg klein	-0.035	-0.035	0.070	-0.148	0.054	0.093

	Schaedel schmal	Schaedel weit	Koerperg gross	Koerperg klein
Haarfarb blond	-0.011	0.011	0.035	-0.035
Haarfarb rot	0.006	-0.006	0.035	-0.035
Haarfarb dunkel	0.005	-0.005	-0.070	0.070
Augenfar hell	0.042	-0.042	0.148	-0.148
Augenfar gemischt	-0.048	0.048	-0.054	0.054
Augenfar braun	0.006	-0.006	-0.093	0.093
Schaedel schmal	0.213	-0.213	0.003	-0.003

Schaedel weit	-0.213	0.213	-0.003	0.003
Koerperg gross	0.003	-0.003	0.245	-0.245
Koerperg klein	-0.003	0.003	-0.245	0.245

Die nominalen Variablen wurden in Dummies aufgelöst. Für die Dummies wurde die Kovarianzmatrix S ermittelt. Der Algorithmus der Faktorenanalyse ist nun folgender: (Wir übernehmen die Darstellung bei Arminger, 1979, S.166/167; siehe dazu auch Bacher, 1994, Seite 116ff):

- Schritt 1: Berechne alle Eigenwerte und Eigenvektoren der Kovarianzmatrix **S** der Dummy Variablen
  S = C G C'
  mit **C** = $(c_1, \ldots c_p)$, **G** = diag $(g_1, \ldots g_p)$
  C ist die Matrix der Eigenvektoren
  G ist die Diagonalmatrix der Eigenwerte.
- Schritt 2: Bestimme die Anzahl der Faktoren, die extrahiert werden sollen. Da **S** eine Varianz-Kovarianzmatrix ist, ist es zur Verwendung des Kaiser Kriteriums (Eigenwerte > 1) notwendig, die Korrelationsmatrix
  $\mathbf{R} = \mathbf{V}^{-1}\, \mathbf{S}\, \mathbf{V}^{-1}$ **mit** $\mathbf{V}^2 = \mathbf{diag} \{ \mathbf{S} \}$
  zu faktorisieren.
  V ist die Diagonalmatrix der Standardabweichungen.
- Schritt 3: Setze den Iterationszähler q = 0
- Schritt 4: Berechne aus den ersten k Eigenvektoren
  $\mathbf{C_k} = (\mathbf{c_1}, \ldots \mathbf{c_k})$ und Eigenwerten $\mathbf{G_k} = (\mathbf{g_1}, \ldots \mathbf{g_k})$
  die orthogonalen Faktorladungen
  $\mathbf{L}^* = \mathbf{C_k}\, \mathbf{G_k}^{1/2}$
- Schritt 5: Normiere die Faktorladungen, so dass $\mathbf{L_i}'\, 1 = 0$,
  indem von l^*_{iaj} der Spaltenmittelwert über alle a abgezogen wird.
  $$l_{iaj} = l^*_{iaj} - \bar{l}^*_{ij} \quad \text{mit} \quad \bar{l}^*_{ij} = \frac{1}{r_i} \sum_{a=1}^{r_i} l^*_{iaj}$$
  wobei $a = 1, \ldots r_i$; $i = 1, \ldots p_i$; $j = 1, \ldots k$
- Schritt 6: Berechne **Dq = (S-LL')[*]M**
  M ist eine t*t Matrix mit 1 in den Blockdiagonalmatrizen $(\mathbf{M_{ii}})$ i = 1,...p
  $(\mathbf{M_{ii}}$ ist eine $r_i * r_i$ Matrix) und 0 sonst.
  Mit "[*]" symbolisieren wir die elementweise Verknüpfung von 2 Matrizen
  Beispiel: **C = A[*]B** <=> $c_{ij} = a_{ij}.b_{ij}$.
  Durch diese elementeweise Verknüpfung wird die geforderte Blockdiagonalform von $\mathbf{D_q}$ erreicht.
- Schritt 7: Berechne die Eigenwerte und Eigenvektoren von $\mathbf{S} - \mathbf{D_q}$.
  McDonald (1969a) schlägt als Abbruchkriterium vor, den Wert
  $$c = \frac{1}{p-k} \sum_{r=k+1}^{p} g_r^2$$ zu berechnen.
  Ist $|c_{q-1} - c_q| < \varepsilon$ wird das Verfahren abgebrochen, sonst erhöhe q um 1 und gehe zu Schritt 4.

Herman Denz hat diesen Algorithmus für Almo programmiert.

Almo liefert für die Daten aus McDonald folgende Ergebnis (gekürzt):

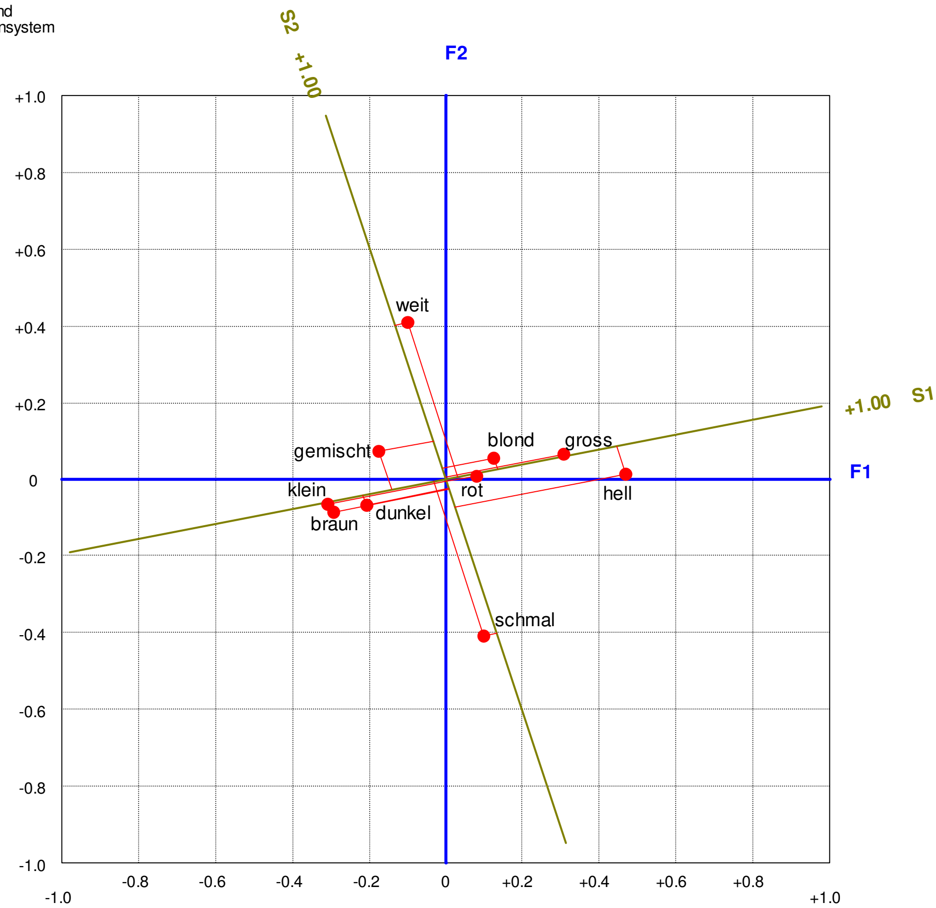
```
Matrix der Faktorladungen
```

		Faktor 1	Faktor 2
Haarfarb blond	V1-1	0.126	0.056
Haarfarb rot	V1-2	0.082	0.011
Haarfarb dunkel	V1-3	-0.209	-0.067
Augenfar hell	V2-1	0.471	0.013
Augenfar gemischt	V2-2	-0.177	0.074
Augenfar braun	V2-3	-0.294	-0.087
Schaedel schmal	V3-1	0.100	-0.408
Schaedel weit	V3-2	-0.100	0.408
Koerperg gross	V4-1	0.309	0.066
Koerperg klein	V4-2	-0.309	-0.066

Aus Quartimin-Rotation:
Matrix der auf die schiefwinkligen Achsen achsparallel projizierten Faktorladungen (Ladungsmatrix)

		Faktor 1	Faktor 2
Haarfarb blond	A1	0.138	0.031
Haarfarb rot	A2	0.082	-0.005
Haarfarb dunkel	A3	-0.221	-0.026
Augenfar hell	B1	0.455	-0.077
Augenfar gemischt	B2	-0.146	0.107
Augenfar braun	B3	-0.309	-0.030
Schaedel schmal	C1	-0.033	-0.424
Schaedel weit	C2	0.033	0.424
Koerperg gross	V4-1	0.316	0.005
Koerperg klein	V4-2	-0.316	-0.005

Faktorladungen im recht- und schiefwinkligen Koordinatensystem



Das schiefwinkelige Koordinatensystem ist nur geringfügig gegenüber dem rechtwinkeligen gedreht. Das Ergebnis kann in folgender Weise interpretiert werden:
Der 1. Faktor ist bipolar mit dem 1. Pol: helle Augenfarbe, groß, blond oder rot und dem 2. Pol: braune oder gemischte Augenfarbe, klein, dunkel. Simplifizierend könnte man sagen, der 1. Faktor ist Nord- versus Südländer. Der 2. Faktor wird allein durch die Schädelform gebildet.

Ein weiteres Beispiel

Wir wollen das "Autofahrer"-Beispiel aus Abschnitt P30.8.1 auch mit dem Blockdiagonalverfahren rechnen. Zu diesem Zwecke muss in der Eingabebox „Verfahren der nominalen Faktorenanalyse" das Wort „Blockdiagonale" eingegeben werden.

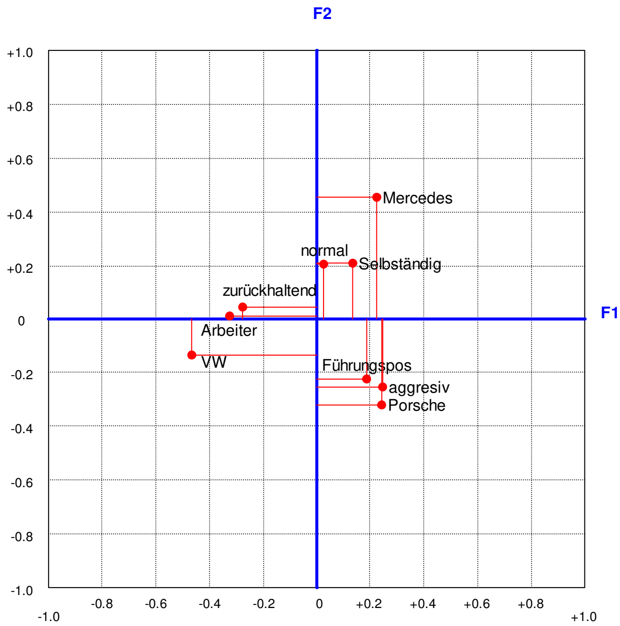
Wir erhalten folgendes (gekürztes) Ergebnis:

```
Ergebnisse aus Faktorenanalyse
Nominale Faktorenanalyse mit Blockdiagonalmatrix
------------------------------------------------
```

```
Schwellenwert fuer iterierte nominale Faktorenanalyse= 0.001
max. Zahl der Iterationen f. nominale Faktorenanalyse= 20
Zahl der tatsaechlich durchgefuehrten Iterationen     = 5
------------------------------------------------------------------------

Eigenwerte
    0.6326      0.5186      0.2481      0.0168

Matrix der Faktorladungen
```

			Faktor 1	Faktor 2	Faktor 3
Auto	Porsche	V1-1	0.2902	-0.2746	0.1008
Auto	Mercedes	V1-2	0.0885	0.3649	-0.1196
Auto	VW	V1-3	-0.3788	-0.0902	0.0188
Beruf	Selbstän	V2-1	0.1622	0.2546	-0.2016
Beruf	Arbeiter	V2-2	-0.3733	-0.0472	0.1978
Beruf	Führungs	V2-3	0.2111	-0.2073	0.0038
Fahrst	aggresiv	V3-1	0.2611	-0.2431	-0.0382
Fahrst	normal	V3-2	0.0748	0.3485	0.2848
Fahrst	zurückha	V3-3	-0.3360	-0.1053	-0.2466

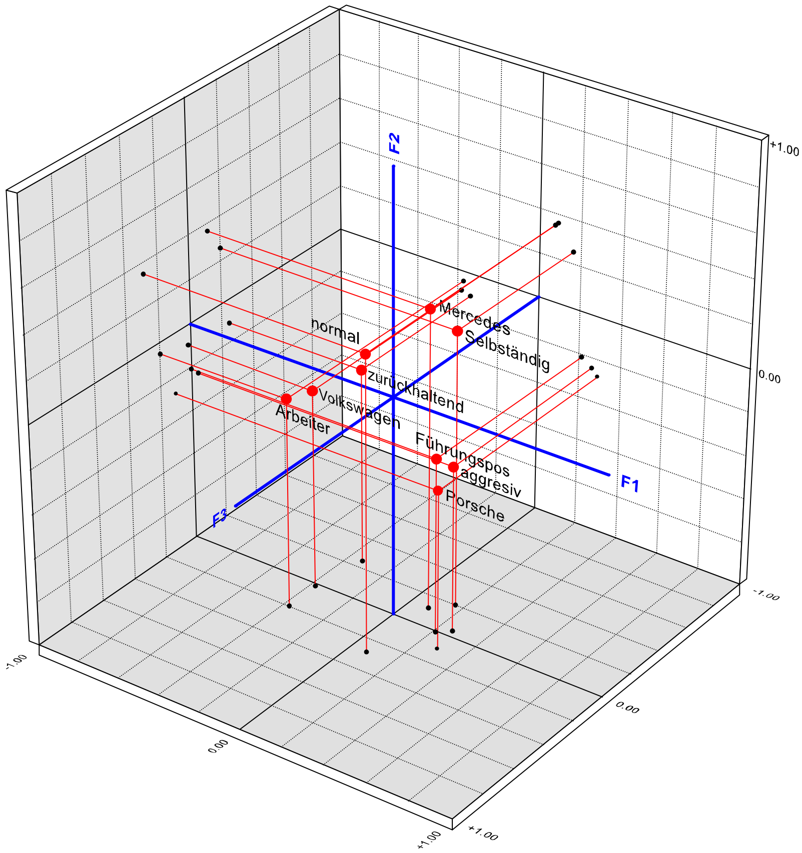
Grafisch dargestellt erhalten wir folgendes Ergebnis für die ersten 2 Faktoren:



Erläuterung: Es sind deutlich 3 Typen von Autofahrern zu erkennnen.

```
Typ 1:   Mercedes,  Selbständiger,      normaler Fahrstil
    2:   Porsche,   Führungsposition,   aggresiver Fahrstil
    3:   VW,        Arbeitnehmer,       zurückhaltend
```

3-dimensionale Darstellung



Erläuterung: Es sind die oben beschriebenen Autofahrer-Typen deutlich erkennnbar. Die 3. Dimension bringt bei diesem Beispiel keine zusätzliche Erkenntnis.

P30.8.3 Die multiple Korrespondenzanalyse (MCA)

Wir sprechen von "bivariater" Korrespondenzanalyse, wenn der Zusammenhang zwischen zwei nominalen Variablen untersucht wird. Die "bivariate" Korrespondenzanalyse haben wir im Almo-Dokument Nr. 4 „Kanonische Analysen", Abschnitt P29.4 dargestellt. Wir sprechen von "multipler" Korrespondenzanalyse, wenn 3 oder mehr Variable analysiert werden. Wir werden jedoch noch zeigen, dass die multiple Korrespondenzanalyse auch problemlos auf 2 nominale Variable angewendet werden kann. Sie erbringt dieselben Ergebnisse.

Was wir hier als „multiple Korrespondenzanalyse" (MCA) bezeichnen, wird bei SPSS "homogenity analysis" genannt. Die Ergebnisse, die unser MCA liefert, sind dieselben, die die SPSS-Prozedur "homals" erzeugt, obwohl die Kalküle von MCA und "homals" völlig verschieden sind.

Wir übernehmen im folgenden den MCA-Kalkül von J.C. Carrol / P.E. Green (1988). Dieser ist einfacher und eleganter (!) als der Kalkül der älteren Korrespondenzanalyse.

Betrachten wir folgendes Beispiel:

Die nominalen Variablen, die faktorisiert werden sollen, sind:

```
    Autokauf: Porsche, Mercedes, VW,
    Beruf:    Selbständig, Arbeitnehmer, Führungsposition,
    Fahrstil: Aggresiv, normal, zurückhaltend
```

Eine Korrespondenzanalyse über diese 3 nominalen Variablen könnte zu einem 2- oder 3-dimensionalen Raum führen, in dem z.B. folgende Punkte dicht beieinander liegen:

```
Punktewolke 1:   Mercedes, Selbständiger,     normaler Fahrstil
            2:   Porsche,  Führungsposition, aggresiver Fahrstil
            3:   VW,       Arbeitnehmer,      zurückhaltend
```

Wir haben also 3 Typen identifiziert.

Kalkül der MCA

Diese 3 nominal-polytomen Variablen werden in Dummies aufgelöst. Für die Dummies wird die durchschnittliche Kreuzprodukte-Matrix **G** gebildet ("durchschnittlich" heißt, dass die Kreuzproduktematrix elementeweise noch durch n, die Zahl der Untersuchungseinheiten, dividiert wird).

Die Diagonalmatrix $\mathbf{H}^{-1/2}$ wird gebildet. Sie besteht aus dem Kehrwert der Wurzel der Diagonalelementen von **G**. Das einzelne Element ist also $1/\sqrt{g_{ii}}$ (dabei ist g_{ii} der Anteilswert der Untersuchungseinheiten in einer Ausprägung.)

Die Matrix G wird in folgender Weise *skaliert*:

$$G^* = H^{-1/2} \cdot G \cdot H^{-1/2}$$

G* ist eine Art Kovarianzmatrix.

Der 1. Faktor der Matrix **G*** besteht aus Konstanten. Es wird die Residualmatrix **G**** gebildet, die von diesem 1. Faktor "befreit" ist.

Wir bezeichnen **G**** als *"MCA-Matrix"*. Wird in nachfolgend abgebildeter Programm-Maske Prog30m5 die Optionsbox "Verschiedene Programm-Optionen" geöffnet und Zwischenergebnisse angefordert dann gibt Almo aus

Skalierungsvektor fuer MCA-Matrix
1.758 1.758 1.683 1.758 1.758 1.683 1.844 1.758 1.617

MCA-Matrix: Ausgangsmatrix fuer multiple Korrespondenz-Analyse

		Auto Porsche V1-1	Auto Merced V1-2	Auto VW V1-3	Beruf Selbst V2-1	Beruf Arbeiter V2-2	Beruf Führung V2-3	Fahrst aggress V3-1	Fahrst normal V3-2	Fahrst zurückh V3-3
Auto	Porsche	0.676	-0.324	-0.338	-0.142	-0.233	0.358	0.454	-0.142	-0.268
Auto	Mercedes	-0.324	0.676	-0.338	0.404	-0.233	-0.164	-0.213	0.313	-0.101
Auto	VW	-0.338	-0.338	0.647	-0.251	0.445	-0.186	-0.231	-0.164	0.353
Beruf	Selbstän	-0.142	0.404	-0.251	0.676	-0.324	-0.338	-0.022	0.131	-0.101
Beruf	Arbeiter	-0.233	-0.233	0.445	-0.324	0.676	-0.338	-0.308	0.040	0.234
Beruf	Führungs	0.358	-0.164	-0.186	-0.338	-0.338	0.647	0.317	-0.164	-0.127
Fahrstil	aggressi	0.454	-0.213	-0.231	-0.022	-0.308	0.317	0.706	-0.308	-0.335
Fahrstil	normal	-0.142	0.313	-0.164	0.131	0.040	-0.164	-0.308	0.676	-0.352
Fahrstil	zurückha	-0.268	-0.101	0.353	-0.101	0.234	-0.127	-0.335	-0.352	0.618

G** wird nun in der üblichen Weise faktorisiert. Die dabei entstehende Faktorladungsmatrix **L** muss wieder reskaliert werden:

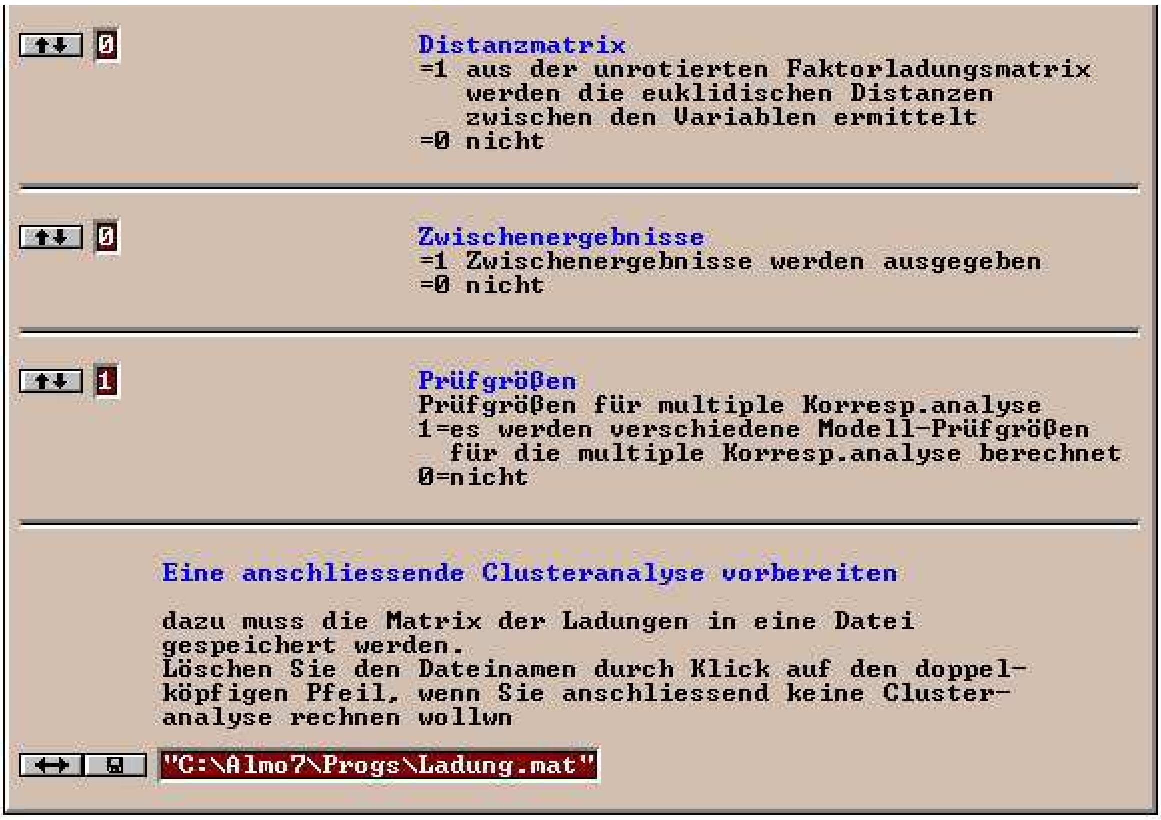
$$L^* = H^{-1/2} \cdot L$$

L* ist die gesuchte orthogonale Faktorladungsmatrix. Sie kann in gewohnter Weise schief - oder rechtwinkelig rotiert werden. Auch die Berechnung von Faktorwerten (in der Sprache der MCA: "object scores") ist selbstverständlich möglich.

P30.8.3.1 Eingabe mit Maskenprogramm Prog30m5

Lediglich in der Eingabebox "Verfahren" muss mit Mausklick auf multiple Korrespondenzanalyse umgeschaltet werden.

Wir wollen nun auch die Distanzmatrix und die „Prüfgrößen für die multiple Korrespondenzanalyse“ ermitteln und wir wollen eine spätere Clusteranalyse vorbereiten. Zu diesem Zweck wird die Optionsbox „verschiedene Programm-Optionen“ geöffnet.



Wird eine Distanzmatrix angefordert, dann ermittelt Almo aus der orthogonalen Faktorladungsmatrix die räumlichen Distanzen zwischen den Dummies der nominalen Variablen. In unserem Beispiel entsteht folgende Distanzmatrix (gekürzt):

Matrix der euklidischen Distanzen zwischen den Variablen

			Auto Porsche V1-1	Auto Mercedes V1-2	...
Auto	Porsche	V1-1	0	2.2332	...
Auto	Mercedes	V1-2	2.2332	0	...
Auto	VW	V1-3	2.1873	2.2284	...
Beruf	Selbstän	V2-1	2.1576	0.5188	...
Beruf	Arbeiter	V2-2	2.2721	2.2306	...
Beruf	Führungs	V2-3	0.2843	2.1104	...
Fahrst	aggresiv	V3-1	0.4811	2.1963	...
Fahrst	normal	V3-2	2.1373	1.2925	...
Fahrst	zurückha	V3-3	2.1896	1.9701	...

Werden die Prüfgrößen für die MCA angefordert, dann werden eine Reihe von Maßzahlen, die im Rahmen der klassischen Korrespondenzanalyse üblicherweise ermittelt werden, ausgegeben. In Abschnitt P30.8.3.4 werden sie ausführlich dargestellt.

Mit der orthogonalen Faktorladungsmatrix als Eingabedaten ist es möglich, eine Clusteranalyse zu rechnen, die für unser Beispiel dieselben Autofahrertypen nachweisen müsste. Wir werden darauf in Abschnitt P30.8.3.7 ausführlich eingehen. Im letzten Eingabefeld der Optionsbox wird Almo aufgefordert, eine anschließende Clusteranalyse vorzubereiten.

Ergebnis

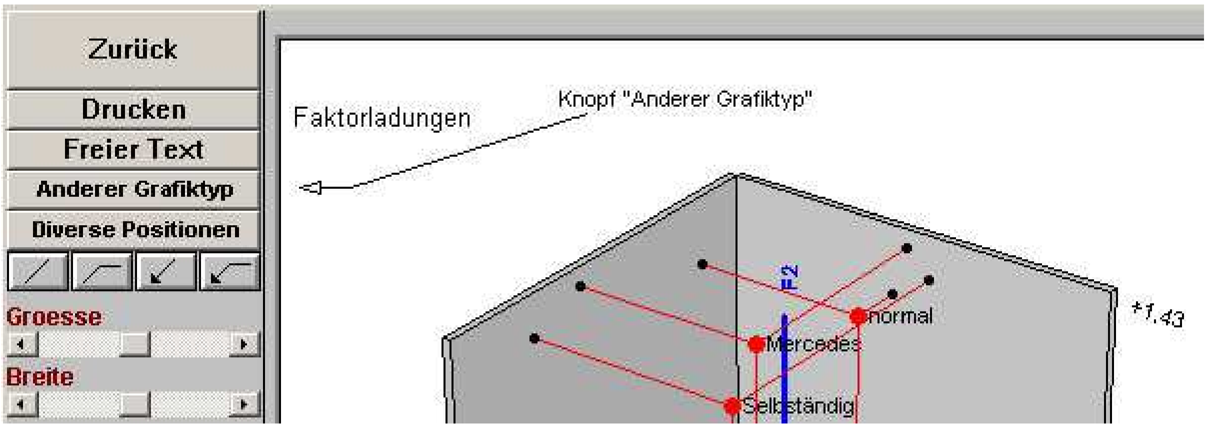
ALMO liefert folgendes Ergebnis, das wir hier verkürzt wiedergeben.

```
Matrix der Faktorladungen
(=Matrix der "category quantifications")
```

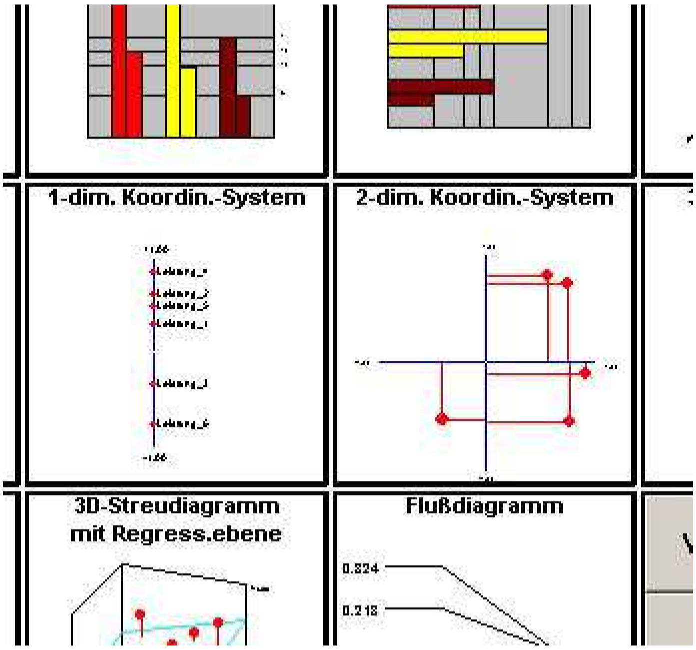
			Faktor 1	Faktor 2	Faktor 3
Auto	Porsche	V1-1	1.1605	-0.5125	-0.2392
Auto	Mercedes	V1-2	-0.0698	1.2980	0.2026
Auto	VW	V1-3	-0.9997	-0.7200	0.0335
Beruf	Selbstän	V2-1	0.0919	1.1290	0.6656
Beruf	Arbeiter	V2-2	-1.0893	-0.5353	-0.5557
Beruf	Führungs	V2-3	0.9142	-0.5441	-0.1007
Fahrst	aggresiv	V3-1	1.2477	-0.4588	0.2309
Fahrst	normal	V3-2	-0.1703	0.9557	-1.0397
Fahrst	zurückha	V3-3	-0.8156	-0.4557	0.7020

Almo stellt diese Matrix grafisch dar. Betrachten wir zuerst die 2-dimensionale Darstellung.

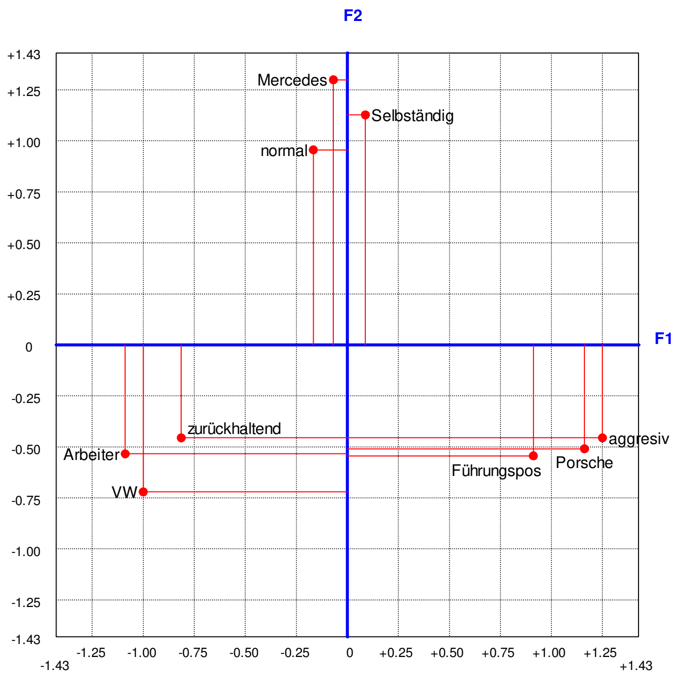
Diese erhält man, wenn man im Grafik-Fenster auf der linken Bildschirmseite auf den Knopf "Anderer Grafiktyp" klickt. Wir zeigen hier einen Ausschnitt aus dem Grafikfenster.



Almo präsentiert dann eine Übersicht über die verschiedenen "Anderen Grafiktypen", in die die Grafik transformiert werden kann. Wir zeigen wieder nur einen Ausschnitt aus dieser Übersicht:



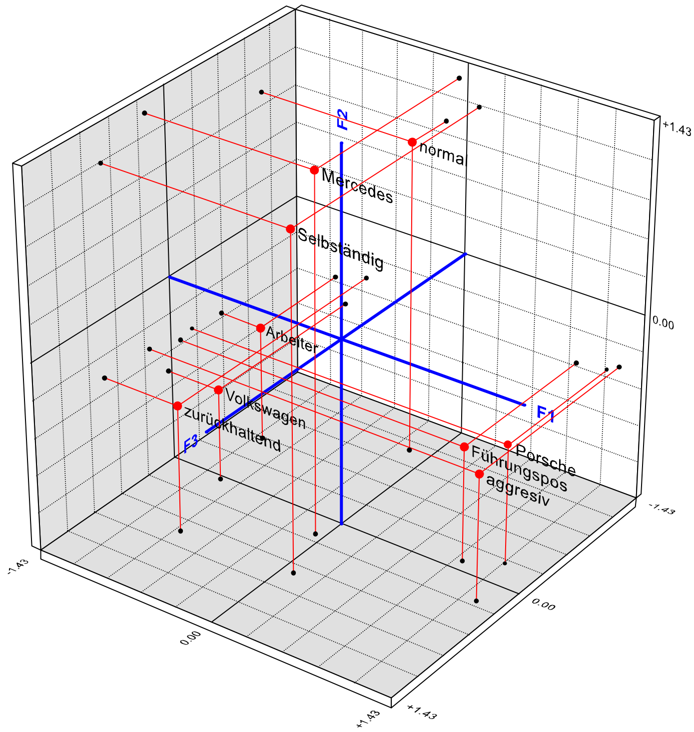
Klicken Sie in das kleine Fenster "2-dim Koordin.-System". Es werden dann nur die ersten beiden Achsen gezeigt. Sie sehen folgendes:



Erläuterung: Es sind deutlich 3 Typen von Autofahrern zu erkennen.

Typ 1:	Mercedes,	Selbständiger,	normaler Fahrstil
2:	Porsche,	Führungsposition,	aggresiver Fahrstil
3:	VW,	Arbeitnehmer,	zurückhaltend

Die 3-dimensionale Darstellung:



********** ***Erläuterung***:

In beiden Darstellungen wird der Zusammenhang der 3 nominalen Variable ersichtlich. Wir haben 3 Typen von Autofahrern identifiziert

```
Typ 1:  Mercedes, Selbständiger,    normaler Fahrstil
Typ 2:  Porsche,  Führungsposition, aggresiver Fahrstil
Typ 3:  VW,       Arbeitnehmer,     zurückhaltend
```

```
                                            ******************** MITTEILUNG
              Die Faktorladungsmatrix wurde in folgende Datei gespeichert
                                             "C:\Almo7\Progs\Ladung.mat"
Mit Doppelklick auf den Dateinamen wird die Datei in ein Fenster geladen
```

*************** ***Erläuterung:***

Almo teilt mit, dass es - für eine spätere Clusteranalyse – die Faktorladungsmatrix in eine Datei gespeichert hat. Siehe dazu auch Johann Bacher in Abschnitt P30.8.3.7. Danach wird dem Benutzer ein Clusteranalyse-Programm als Syntaxprogramm angeboten.

```
ALMO-Programm fuer eine anschliessende
Clusteranalyse:
#####################################
#AlmPrg#

VEREINBARE
  Variable=100;

ANFANG
Name1=:
  Porsche
 ,Mercedes
 ,VW
 ,Selbstän
```

```
 ,Arbeiter
 ,Führungs
 ,aggresiv
 ,normal
 ,zurückha;
Name2=XFaktor1;
Name3=XFaktor2;
Name4=XFaktor3;
ENDE

ANFANG
   Programm = 36;              # hierarchische Clusteranalyse #
   A_Quantitative_V = V2:4;
   Modell = Ward_Linkage;
   Distanzmass = quad_euklid;
   Objekte = 9;
   Min_Clusterzahl = 2;
   Max_Clusterzahl = 3;
Ende_Programmparameter

Lese V2:4 aus Datei 7 "C:\Almo7\Progs\Ladung.mat" Format frei
   leerzu Ende;
Gehe_in_Programm
Gehe_zu Lese
ENDE
                                                ******************** MITTEILUNG
                  Das ALMO-Clusteranalyse-Programm wurde unter dem Namen
                                                    "C:\Almo7\Progs\Ladung.alm"
                                                                  gespeichert
         Wenn Sie auf den Dateinamen doppelklicken, dann wird es geladen
                                             Sie koennen es dann gleich rechnen
#####################################
```

***************** Erläuterung:**
Das Clusteranalyse-Programm wird auch in eine Datei gespeichert. Dem Benutzer wird mitgeteilt, dass er nur auf den Dateinamen doppelklicken muss. Dann wird das Programm geladen. Durch Klick auf den Knopf "Rechne" wird es dann gerechnet.

Matrix der euklidischen Distanzen zwischen den Variablen

			Auto Porsche V1-1	Auto Mercedes V1-2	Auto VW V1-3
Auto	Porsche	V1-1	0	2.2332	2.1873
Auto	Mercedes	V1-2	2.2332	0	2.2284
Auto	VW	V1-3	2.1873	2.2284	0
Beruf	Selbstän	V2-1	2.1576	0.5188	2.2383
Beruf	Arbeiter	V2-2	2.2721	2.2306	0.6240
Beruf	Führungs	V2-3	0.2843	2.1104	1.9267
Fahrstil	aggresiv	V3-1	0.4812	2.1963	2.2712
Fahrstil	normal	V3-2	2.1373	1.2925	2.1559
Fahrstil	zurückha	V3-3	2.1896	1.9701	0.7421

***************** Erläuterung:**
Die Distanzmatrix wird ausgegeben. Wir geben hier nur einen Teil wieder. Betrachten wir in der 1. Spalte die Porsche-Fahrer und die 3 letzten Werte in dieser Spalte.

0.4812 dies ist die Distanz des Porsche-Fahrers zum "aggressiven Fahrstil"
2.1373 dies ist die Distanz des Porsche-Fahrers zum "normalen Fahrstil"
2.1896 dies ist die Distanz des Porsche-Fahrers zum "zurückhaltenden Fahrstil"

Die geringste Distanz hat der Porsche-Fahrer also zum aggressiven Fahrstil.

Die nachfolgenden "Modellprüfgrößen für die Korrespodenzanalyse" werden im Abschnitt P30.8.3.4 ausführlich erläutert.

```
mittlere Residuenabweichungen:

Faktoren      RMR     GFIR
--------------------------------
    0        0.253    0.000
    1        0.162    0.591
    2        0.089    0.875
    3        0.060    0.943

RMR  = Wurzel aus den mittleren Residuenquadraten nur fuer Dummies
       von unterschiedlichen nominalen Variablen
GFIR = Anpassungsindex zu RMR-Koeffizient (=erklaerter Chi-Quadrat-Wert)

kophenetische Korrelationskoeffizienten:

Dimensionen   KOPH
--------------------------
    1        0.788
    2        0.976
    3        0.992

KOPH = kophenetische Korrelation nur fuer Dummies  von unterschiedlichen
nominalen Variablen

Chi-Quadrat-Pruefgroessen:
                               Signifikanz
Dimensionen      Chi      df    100*(1-p)
------------------------------------------
    0          58.55      20      100
    1          32.54      10      100
    2          36.07       6      100
    3           6.92       2       97

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

Masszahlen fuer Distanzinterpretation:


mittlere Distanzabweichungen:

Dimensionen   DIS     GFID
----------------------------------
    0        4.322    0.000
    1        2.848    0.566
    2        1.434    0.890
    3        0.864    0.960

DIS  = Wurzel aus den mittleren Distanzdifferenzen nur fuer Dummies
       von unterschiedlichen nominalen Variablen
GFID = Anpassungsindex zu DIS


Stress-Koeffizienten:

Dimensionen   Stress    GFIS
----------------------------
    1          0.372    0.756
```

2	0.024	0.998
3	0.058	0.986

Gamma-Koeffizienten:

Dimensionen	Gamma
1	0.704
2	0.843
3	0.861

Matrix der Varianzbeitraege der Dummies zu den Faktoren

			Faktor 1	Faktor 2	Faktor 3
Auto	Porsche	V1-1	0.4357	0.0849	0.0185
Auto	Mercedes	V1-2	0.0015	0.5451	0.0132
Auto	VW	V1-3	0.3527	0.1829	0.0004
Beruf	Selbstän	V2-1	0.0027	0.4123	0.1433
Beruf	Arbeiter	V2-2	0.3838	0.0927	0.0999
Beruf	Führungs	V2-3	0.2949	0.1044	0.0035
Fahrst	aggresiv	V3-1	0.4579	0.0619	0.0156
Fahrst	normal	V3-2	0.0093	0.2955	0.3497
Fahrst	zurückha	V3-3	0.2543	0.0794	0.1884

Varianzbeitraege der Dummies
zu allen extrahierten Faktoren zusammen

Auto	Porsche	V1-1	0.5392
Auto	Mercedes	V1-2	0.5599
Auto	VW	V1-3	0.5361
Beruf	Selbstän	V2-1	0.5584
Beruf	Arbeiter	V2-2	0.5765
Beruf	Führungs	V2-3	0.4030
Fahrst	aggresiv	V3-1	0.5355
Fahrst	normal	V3-2	0.6546
Fahrst	zurückha	V3-3	0.5222

Durch die nominalen Variablen erklaerte Varianz
(=discrimination measures)

hinsichtlich Faktor 1

		absolut	in %
V1	Auto	0.7901	36.0225
V2	Beruf	0.6816	31.0747
V3	Fahrstil	0.7217	32.9027
Eigenwert		2.1933	100%

hinsichtlich Faktor 2

		absolut	in %
V1	Auto	0.8131	43.7247
V2	Beruf	0.6096	32.7822
V3	Fahrstil	0.4369	23.4931
Eigenwert		1.8595	100%

hinsichtlich Faktor 3

		absolut	in %
V1	Auto	0.0322	3.8646
V2	Beruf	0.2469	29.6396
V3	Fahrstil	0.5539	66.4958
Eigenwert		0.8329	100%

Korrelationen der Dummies mit den Faktoren

			Faktor 1	Faktor 2	Faktor 3
Auto	Porsche	V1-1	0.8025	-0.3544	-0.1654
Auto	Mercedes	V1-2	-0.0483	0.8976	0.1401
Auto	VW	V1-3	-0.7383	-0.5317	0.0247
Beruf	Selbstän	V2-1	0.0635	0.7807	0.4603
Beruf	Arbeiter	V2-2	-0.7533	-0.3702	-0.3843
Beruf	Führungs	V2-3	0.6751	-0.4018	-0.0744
Fahrst	aggresiv	V3-1	0.8054	-0.2962	0.1490
Fahrst	normal	V3-2	-0.1178	0.6609	-0.7190
Fahrst	zurückha	V3-3	-0.6417	-0.3585	0.5524

Schiefwinklige Rotation mit 3 Faktoren

Quartimin-Kriterium 1.3530
letzte Iterationsdifferenz bei Quartimin-Rotation 0.0001

Aus Quartimin-Rotation:
Matrix der Korrelationen zwischen
den schiefwinkligen Achsen

	Faktor 1	Faktor 2	Faktor 3
Faktor 1	1.0000	0.2016	-0.5696
Faktor 2	0.2016	1.0000	0.1807
Faktor 3	-0.5696	0.1807	1.0000

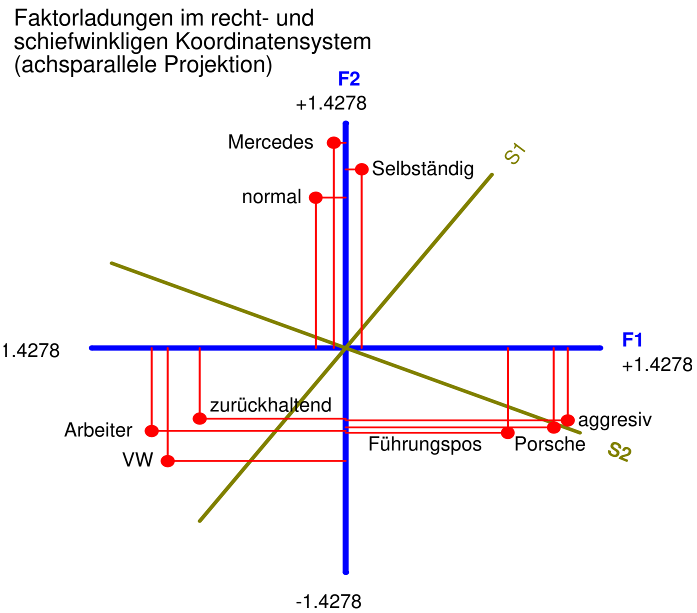
Matrix der Winkel zwischen den schiefwinkligen Achsen

	Faktor 1	Faktor 2	Faktor 3
Faktor 1	0	78.3660	-55.2742
Faktor 2	78.3660	0	79.5882
Faktor 3	-55.2742	79.5882	0

Aus Quartimin-Rotation:
Matrix der auf die schiefwinkligen Achsen
achsparallel projizierten Faktorladungen
(Ladungsmatrix)

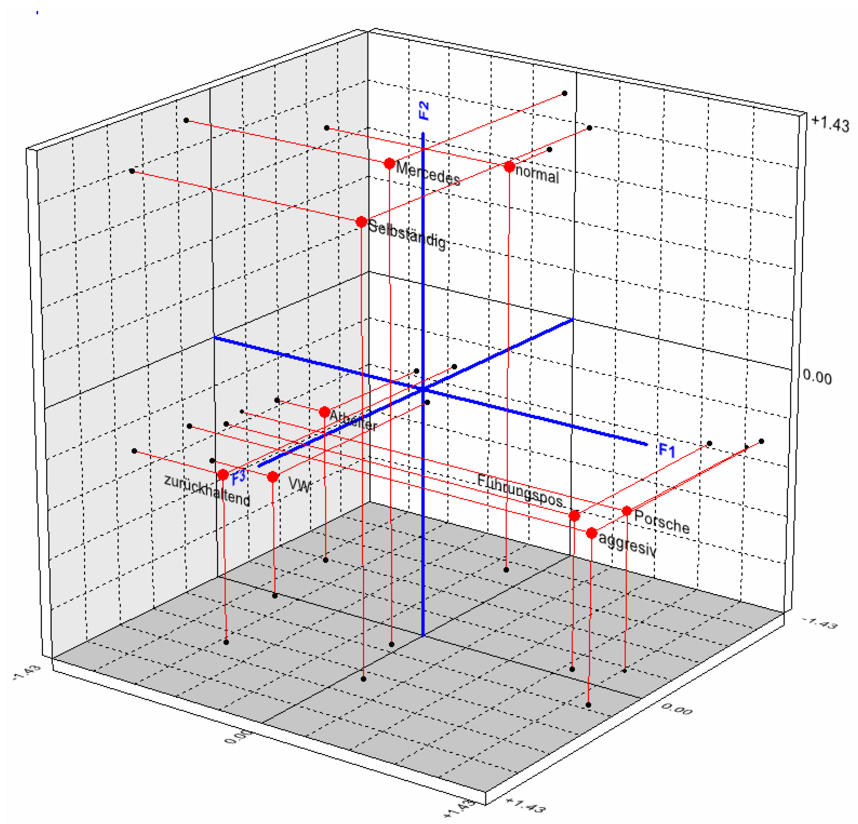
	Faktor 1	Faktor 2	Faktor 3
Auto Porsche A1	-0.1311	1.3335	-0.1120
Auto Mercedes A2	1.0034	-0.7293	-0.2875
Auto VW A3	-0.7995	-0.5538	0.3663
Beruf Selbstän B1	1.4649	-0.7758	0.3529
Beruf Arbeiter B2	-1.3609	-0.3831	-0.4389
Beruf Führungs B3	-0.0953	1.0624	0.0788
Fahrst aggresiv V3-1	0.4481	1.1243	0.4463
Fahrst normal V3-2	-0.6021	0.0195	-1.6827
Fahrst zurückha V3-3	0.1648	-0.8814	1.0805

Almo stellt diese Matrix grafisch dar. Betrachten wir zuerst die 2-dimensionale Darstellung



********** ***Erläuterung***:
F1, F2 sind die ursprünglichen orthogonalen Achsen. Die schiefwinkligen Achsen S1, S2 haben kaum einen Bezug zu den 3 Autofahrertypen. Die schiefwinklige Rotation ist in unserem Beispiel sinnlos. Auch eine orthogonale Varimax-Rotation, die wir hier nicht zeigen, bringt kein interpretierbares Ergebnis.

Die 3-dimensionale Darstellung:



********** ***Erläuterung:***
F1, F2 und F3 sind die ursprünglichen orthogonalen Achsen. S1, S2 und S2 sind die schiefwinkligen Achsen, die durch unser Rotationsprogramm erzeugt wurden.
Die Grafik überfordert das räumliche 3D-Vorstellungsvermögen des Benutzer. Trotzdem ist zu erkennen, dass die schiefwinkligen Achsen nicht befriedigend durch die 3 Punktewolken verlaufen.

Das ist eine Erfahrung, die man sehr oft macht. Der Sinn der Rotation ist es ja, Achsen zu finden, die durch die Punktewolken laufen (oder mindestens ihnen sehr nahe kommen) und die dann inhaltlich als "Hintergrund-Dimensionen" interpretierbar sind. Eine Rotation, ob recht- oder schiefwinklig, bringt sehr selten ein interpretierbares Ergebnis.

P30.8.3.1.1 Maskenprogramm Prog30m5: Eingabe mit fertiger Tabelle

Wenn die Daten nicht als Individualdaten vorliegen, sondern bereits als Tabelle, dann kann das Maskenprogramm Prog30m5 ebenfalls verwendet werden.
Beachte: Durch entsprechende Selektion in der Eingabebox „Verfahren der nominalen Faktorenanalyse“ kann das Maskenprogramm nicht nur für die MCA, sondern auch für das Blockdiagonalverfahren und die Faktorisierung der Dummies verwendet werden.

Betrachten wir das Autofahrer-Beispiel.

Die Tabelle muss in der Form einer multidimensionalen Kontingenztabelle bzw. als "OLAP-Würfel" geschrieben werden. Siehe dazu auch Almo-Dokument Nr. 2 "Beliebig-dimensionale Tabellierung, Abschnitt P11.6.1 und das Maskenprogramm Prog11m2 (das Sie durch Klick auf "Verfahren/Tabellierung" erreichen).

Für das "Autofahrer-Beispiel wäre die Tabelle in folgender Weise zu schreiben:

```
                                              in dieser Weise ist
    diese Tabelle soll eingegeben werden      sie zu schreiben
    ------------------------------------      -------------------

   Auto      Beruf       Fahrstil   Häufigkeit
   ----      -----       --------   ----------
 Porsche  Selbstständig aggressiv     2              1  1  1  2
 Porsche  Selbstständig normal        0              1  1  2  0
 Porsche  Selbstständig zurückhalt    0              1  1  3  0
 Porsche  Arbeitnehmer  aggressiv     0              1  2  1  0
 Porsche  Arbeitnehmer  normal        1              1  2  2  1
 Porsche  Arbeitnehmer  zurückhalt    0              1  2  3  0
    .          .            .         .
    .          .            .         .
    .          .            .         .
 VW       Führungspos   aggressiv     1              3  3  1  1
 VW       Führungspos   normal        0              3  3  2  0
 VW       Führungspos   zurückhalt    1              3  3  3  1
```

In der 1. Zeile steht die Gruppe der Porschefahrer, die selbständig sind und aggressiv fahren. Diese Gruppe umfasst 2 Personen. Die 2. Gruppe ist leer, etc. ...

Leere Gruppen können - müssen aber nicht geschrieben werden. Auch die Reihenfolge, in der die Datensätze hinter einander stehen, ist beliebig.

Gehen Sie folgendermaßen vor:

1. Erzeugen Sie ein neues Fenster durch Klick auf das Menü "Datei/Neue Datei anlegen" oder einfach durch Klick auf den ersten Knopf ganz links in der Knopfleiste. Almo präsentiert die Datei-Auswahl-Box. Entscheiden Sie sich für ein bestimmtes Verzeichnis und einen bestimmten Dateinamen, z.B.

   `"C:\Almo\Progs\Tabelle.tab"`

2. In dieses Fenster schreiben Sie die Tabelle in der Form wie oben beschrieben. Leere Datendätze können weg bleiben. Die Aufeinderfolge der Datensätze ist beliebig.
   Damit das Maskenprogramm Prog30m5 auch für die Eingabe derartiger fertiger Tabellen verwendet werden kann, müssen nur in 3 Boxen folgende Eingaben gemacht werden:

   1. In der Eingabebox "Freie Namensfelder"

   Die Häufigkeitsvariable erhält einen Namen. Beachte, dass bei Variablennamen deutsche Umlaute verboten sind.

2. In der Eingabebox "Datei aus der gelesen wird"

Geben Sie den Namen der Datei an, in der sich die fertige Tabelle befindet.

3. Die Eingabebox "Option: Untersuchungseinheiten gewichten" muss geöffnet werden.

In die geöffnete Optionsbox wird dann eingetragen

Die Häufigkeitsvariable wird als Gewichtungsfaktor verwendet.

In den anderen Eingabeboxen des Programms müssen Sie dann die Einträge Ihren Daten und Analyse-Absichten entsprechend anpassen.

P30.8.3.2 Beispiel einer multiplen Korrespondenzanalyse (MCA)

Wegen ihrer Bedeutung in der empirischen Forschung und Datenanalyse wollen wir ein weiteres Beispiel für die MCA im Detail betrachten.

Wir wollen ein Beispiel von Carroll und Green verwenden.

(Siehe D.Carroll/P.F.Green: An INDSCAL-based approach to multiple correspondence analysis, Journal of Marketing Research, Vol XXV, May 1988, S.193-203).

25 Studenten wurden folgende Fragen gestellt:

1. Was trinken Sie am liebsten?
   (1) Coke
   (2) 7-Up
   (3) Dr. Pepper
   (4) Shasta grape
2. Wieviel Geld geben Sie durchschnittlich jede Woche für (nichtalkoholische) Getränke aus?
   (1) weniger als $2.00
   (2) $2.00 - 3.99
   (3) $4.00 und mehr
3. Welchen Imbiß nehmen Sie am liebsten zu einem (nichtalkoholischen) Getränk?
   (1) pretzels
   (2) peanuts
   (3) M&M's

(4) Fritos
(5) dried fruits

Die Frage ist, ob die befragten Studenten in Typen differenziert werden können.

P30.8.3.2.1 Analyse mit Maskenprogramm Prog30m5

Das Prog30m5 haben wir bereits dargestellt und seine Eingabeboxen ausführlich erläutert. Für unser Beispiel entsprechend ausgefüllt, erhalten wir folgendes Programm, das Sie auch im Menü „Almo/Beispiel-Programm laden“ unter dem Namen „Carrol2.Alm“ finden.

Caroll2.ALM
Beispiel nach Caroll/Green: An INDSCAL-based approach to MCA

verwendet wird

Prog30m5.Msk
Faktorenanalyse mit nominalen Variablen
mit Optionen

Was ist ein Kurzprogramm ? --> Hilfe
Bedienung --> Hilfe

1 Speicher fuer x Variable | Hilfe
Vereinbare Variable= 20 ;

2 Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

3 Datei der Variablennamen | Hilfe
zeige = Namensdatei in Output zeigen
leer = nicht

4 Freie Namensfelder | Hilfe
Name1=Drink:Coke, 7-Up, Dr.Pepper, Grape;
Name2=Money:bis $2, $2-4, $4 und mehr;
Name3=Snack:Pretzels, Peanuts, M&M, Fritos, dried Fruits;
erzeuge zusätzliche Namensfelder

5 Datei aus der gelesen wird | Hilfe bei Datei-Problemen
"C:\Almo7\Testdat\Caroll.fre"
frei | Format der Daten | Hilfe
V1:3 | der Datensatz enthält diese Variablen
Bei Format DIREKT schreiben Sie: alle_V

6 Wenn Dateiformat FIX oder Nicht-Standard-FREI | Hilfe

7 zu faktorisierende nominale Variable | Hilfe
Drink,Money,Snack

8

Verfahren der nominalen Faktorenanalyse

multiple_Korrespondenz

0= gewöhnliche Faktorenanalyse der in Dummies aufgelösten nominalen Variablen

Blockdiagonale = nominale Faktorennalyse mit Blockdiagonal-Matrix nach McDonald

multiple_Korrespondenz = multiple Korrespondenz-Analyse

9 Option: Ein- und Ausschliessen von Untersuchungseinheiten

10 Option: Umkodierungen und Kein-Wert-Angaben

11 Option: Spezielle Kein-Wert-Behandlung

12 Option: Untersuchungseinheiten gewichten

13

Loesche wieder diese Box

Faktoren

3

Faktorenzahl
Es sollen x Faktoren extrahiert werden.

Erfolgt keine Angabe, d.h. wird das Eingabefeld leer gemacht, dann wird die Faktorenzahl gemäß bestimmter Regeln festgelegt (siehe Handbuch). Erfolgt jedoch eine Angabe, dann wird ihr auch entsprochen.

0

Faktorensignifikanz-Test
(nicht möglich bei nominaler Faktorenanalyse)
=1 zuerst wird ermittelt,ob die zu faktorisierende Matrix überhaupt signifikant ist. Dann wird die Signifikanz von jedem extrahierten Faktor bestimmt
=0 keine Signifikanz-Prüfung

14

Loesche wieder diese Box

Rotation

0

Zahl der zu rotierenden Faktoren
keine Angabe= die extrahierten Faktoren werden rotiert
=0 nicht rotieren

Beispiele fuer weitere mögliche Angaben:
=2 schiefwinklige Rotation für 2 Faktoren
=2:4 schiefwinklige Rotation für die ersten 2 Faktoren, dann für die ersten 3 Faktoren etc. bis zu den ersten 4 Faktoren.

Rotationstyp
= schiefwinklig; Quartimin-Rotation mit Verbesserung durch die Gruppen-Rotation (Voreinstellung)
= rechtwinklig; Varimax-Rotation

1

Rotationslösung
=1 die Ergebnisse der durch die "Gruppen-Rotation" verbesserten Quartimin-Rotation werden ausgegeben (Voreinstellung)
=2 die reine Quartimin-Lösung wird zusätzlich ausgegeben
=3 nur Quartimin-Lösung rechnen und ausgeben

Loesche wieder diese Box

verschiedene Programm-Optionen

0

Faktorwert-Koeffizienten
=0 nicht ermitteln
=1 Faktorwert-Koeffizienten f. alle Faktoren rechnen

1

Distanzmatrix
=1 aus der unrotierten Faktorladungsmatrix werden die euklidischen Distanzen zwischen den Variablen ermittelt
15 =0 nicht

1

Zwischenergebnisse
=1 Zwischenergebnisse werden ausgegeben
=0 nicht

1

Prüfgrößen
Prüfgrößen für multiple Korresp.analyse
1=es werden verschiedene Modell-Prüfgrößen für die multiple Korresp.analyse berechnet
0=nicht

Eine anschliessende Clusteranalyse vorbereiten

dazu muss die Matrix der Ladungen in eine Datei gespeichert werden.
Löschen Sie den Dateinamen durch Klick auf den doppelköpfigen Pfeil, wenn Sie anschliessend keine Clusteranalyse rechnen wollwn

```
[↓] Loesche wieder diese Box

        Option: Faktorwerte ermitteln und speichern

[↔] 3                      Zahl der Faktorwertvariablen,
                           die gebildet werden sollen

[↔] Runde 0.01
                           die Faktorwert-Variable transformieren, z.B.
                           +4 addieren, dann mit 10 multiplizieren
                           dann auf Ganzzahl runden
                           - ist nicht obligatorisch -

[↔] 20                     Maximal x % der faktorisierten Variablen,
                           dürfen Kein_Wert besitzen. Für sie wird
                           der Mittelwert eingesetzt.
                           Sonst werden die Faktorwert-Variable
                           auf Kein_Wert gesetzt
```

16

```
[↔][■] "C:\Almo7\PROGS\Fakwert"
                                                   [Hilfe]
                        Geben Sie einen neuen Dateinamen
                        ohne Erweiterung an
                        Almo erzeugt 3 Dateien:

                        1. eine nicht lesbare Almo-Arbeitsdatei
                           mit der Erweiterung ___.dir

                        2. eine anschaubare Datei im freien Format
                           mit der Erweiterung ___.fre

                        3. eine Datei der Variablennamen
                           mit der Erweiterung ___.nam

 In den unter 1. und 2. angegebenen neuen Dateien sind nun enthalten:
   - die Variablen aus der alten Datei
   - die Faktorwertvariable, die als letzte Variablen
     hinter die Variablen der alten Datei gestellt wurden

 In der unter 3. angegebenen neuen Datei der Variablennamen sind nun
 enthalten:
   - die Variablennamen aus der alten Datei einschliesslich der
     in der Box "Freie Namensfelder" angegebenen Namen
   - die Namen "Fakwert" für die neuen angehängten Faktorwertvariablen
     Sie müssen diesen noch die Nummern 1, 2, ... anhängen

                                     bitte lesen -----> [Hilfe]
```

17 [↓] Option: "Aussehen" der auszugebenden Tabelle bzw. Matrix

18 [↓] Grafik-Optionen

19 Programmende

Almo liefert aus dem Maskenprogramm folgende Ergebnisse (gekürzt)

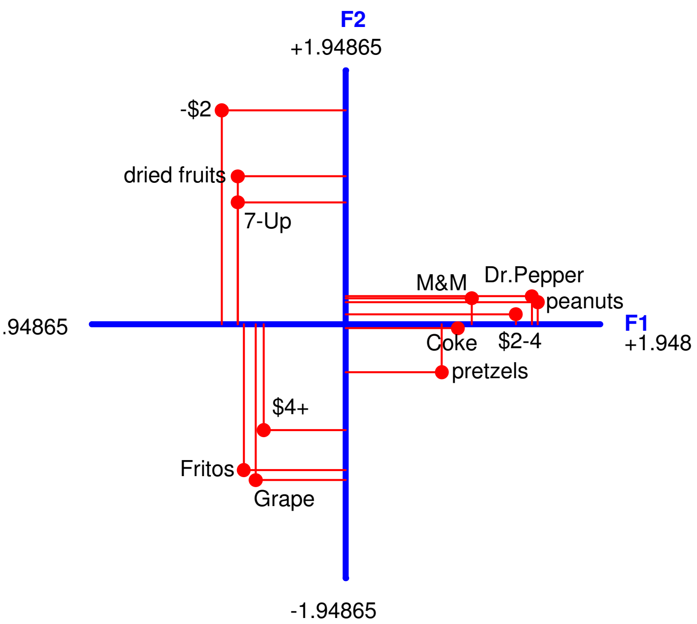
```
Ergebnisse aus Faktorenanalyse
Multiple Korrespondenz-Analyse

Koeffizienten fuer Faktoren

Eigenwerte
    2.7272      2.2593      1.6842

Prozent der Varianz
   30.3020     25.1037     18.7138
Zu erklaerende Gesamtvarianz=        9.0000
```

```
Durch 3 Faktoren erklaerte Varianz=  6.6708
Prozentsatz der erklaerten Varianz= 74.1195

Durch die nominalen Variablen erklaerte Varianz je Faktor (in %)
   90.9061     75.3111     56.1413

Matrix der Faktorladungen
(=Matrix der "category quantifications")
```

			Faktor 1	Faktor 2	Faktor 3
Drink	Coke	V1-1	0.8605	-0.0439	-1.5962
Drink	7-Up	V1-2	-0.8228	0.9313	0.1473
Drink	Dr.Peppe	V1-3	1.4275	0.2254	1.2378
Drink	Grape	V1-4	-0.6941	-1.1940	0.0875
Money	-$2	V2-1	-0.9537	1.6383	-0.0946
Money	$2-4	V2-2	1.2952	0.0875	-0.0873
Money	$4+	V2-3	-0.6262	-0.8162	0.1145
Snack	pretzels	V3-1	0.7450	-0.3731	-1.7715
Snack	peanuts	V3-2	1.4670	0.1772	0.3007
Snack	M&M	V3-3	0.9725	0.2092	1.3567
Snack	Fritos	V3-4	-0.7857	-1.1141	0.3203
Snack	dried fr	V3-5	-0.8245	1.1318	-0.2121

Grafik für eine 2-dimensionale Lösung

Faktorladungen



2 Typen sind eindeutig identifizierbar:
1. -$2, dried fruits, 7-Up
2. $4+, Fritos , grape

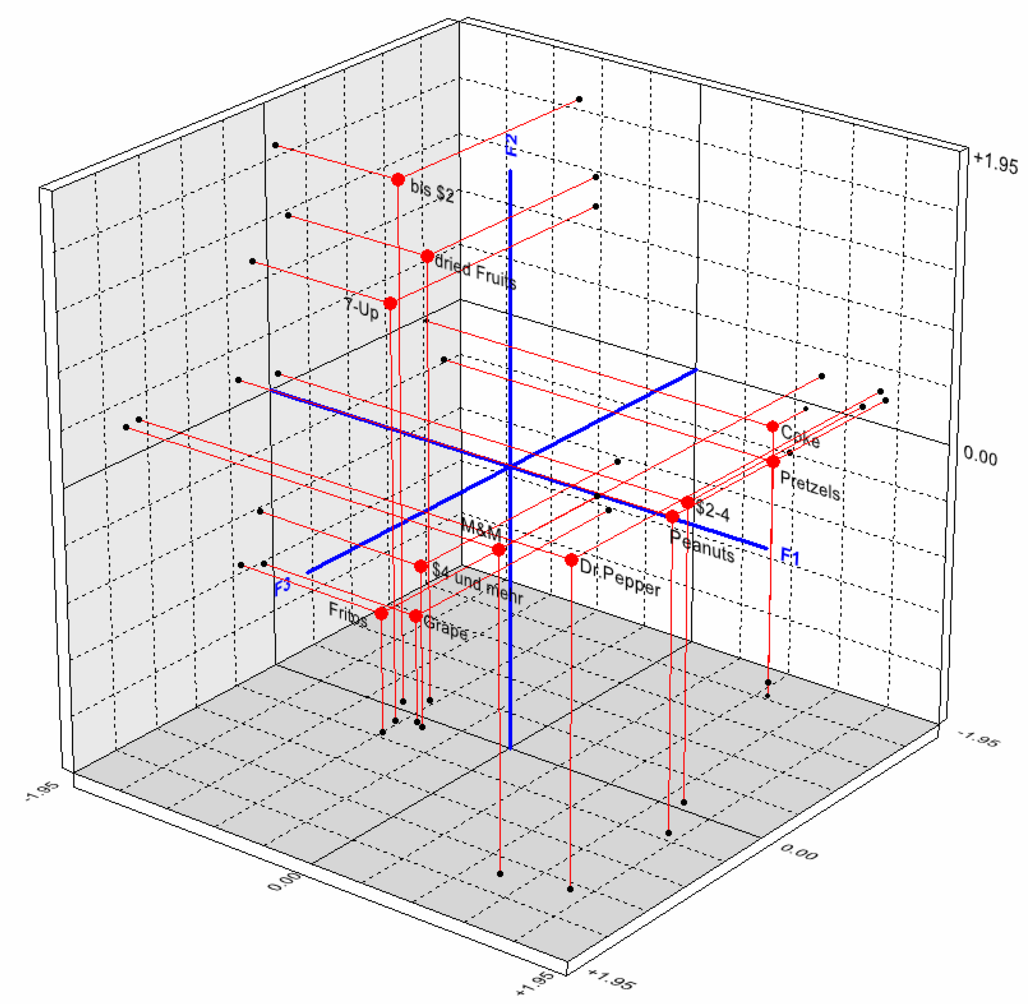
In einen Typ fallen zusammen:

Money: $ 2-4
Snack: pretzels, M&M und peanuts

Drink: Coke und Dr. Pepper

Die Frage ist nun, ob sich bei 3-dimensionaler Darstellung dieser gemeinsame Typ auflösen lässt.

Grafik für die 3-dimensionale Lösung



Der gemeinsame Typ läßt sich ungefähr so auflösen

1. Money: $2-4
   Snack: pretzels (und peanuts)
   Drink: Coke
2. Money: $ 2-4
   Snack: M&M (und peanuts)
   Drink: Dr. Pepper

Der 3. Faktor ist also noch signifikant

Matrix der euklidischen Distanzen zwischen den Variablen
(Wir geben sie hier nur teilweise aus)

			Drink Coke V1-1	Drink 7-Up V1-2	Drink Dr.Pepper V1-3	
Drink	Coke	V1-1	0	2.6125	2.9027	.....
Drink	7-Up	V1-2	2.6125	0	2.5983	.....

Drink	Dr.Peppe	V1-3	2.9027	2.5983	0	.....
Drink	Grape	V1-4	2.5642	2.1301	2.7999	.....
Money	-$2	V2-1	2.8942	0.7586	3.0727	.....
Money	$2-4	V2-2	1.5757	2.2919	1.3389	.....
Money	$4+	V2-3	2.3945	1.7589	2.5622	.....
Snack	pretzels	V3-1	0.3904	2.8003	3.1433	.....
Snack	peanuts	V3-2	2.0038	2.4156	0.9391	.....
Snack	M&M	V3-3	2.9659	2.2819	0.4705	.....
Snack	Fritos	V3-4	2.7439	2.0531	2.7449	.....
Snack	dried fr	V3-5	2.4774	0.4116	2.8276	.....

Die nachfolgenden Modellprüfgrößen für die Korrespondenzanalyse werden im nächsten Abschnitt P30.8.3.4 ausführlich erläutert.

```
mittlere Residuenabweichungen:

Faktoren      RMR     GFIR
--------------------------------
    0        0.290    0.000
    1        0.191    0.565
    2        0.110    0.857
    3        0.049    0.971

RMR  = Wurzel aus den mittleren Residuenquadraten nur fuer Dummies
       von unterschiedlichen nominalen Variablen
GFIR = Anpassungsindex zu RMR-Koeffizient (=erklaerter Chi-Quadrat-Wert)


kophenetische Korrelationskoeffizienten:

Dimensionen   KOPH
--------------------------
    1        0.753
    2        0.931
    3        0.993

KOPH = kophenetische Korrelation nur fuer Dummies
       von unterschiedlichen nominalen Variablen


Chi-Quadrat-Pruefgroessen:
                                   Signifikanz
Dimensionen      Chi      df    100*(1-p)
------------------------------------------
    0          98.79      37      100
    1          30.08      18       96
    2          11.59       7       88
    3           0.00      -3        0

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

Masszahlen fuer Distanzinterpretation:


mittlere Distanzabweichungen:

Dimensionen   DIS     GFID
----------------------------------
    0        7.631    0.000
    1        5.569    0.467
    2        4.341    0.676
    3        2.637    0.881

DIS  = Wurzel aus den mittleren Distanzdifferenzen nur fuer Dummies
       von unterschiedlichen nominalen Variablen
GFID = Anpassungsindex zu DIS
```

Stress-Koeffizienten:

Dimensionen	Stress	GFIS
1	0.311	0.811
2	0.287	0.788
3	0.185	0.876

Gamma-Koeffizienten:

Dimensionen	Gamma
1	0.469
2	0.473
3	0.706

Matrix der Varianzbeitraege der Dummies zu den Faktoren

			Faktor 1	Faktor 2	Faktor 3
Drink	Coke	V1-1	0.1481	0.0003	0.5095
Drink	7-Up	V1-2	0.2166	0.2775	0.0069
Drink	Dr.Peppe	V1-3	0.4075	0.0101	0.3064
Drink	Grape	V1-4	0.1349	0.3992	0.0021
Money	-$2	V2-1	0.1819	0.5367	0.0017
Money	$2-4	V2-2	0.6039	0.0027	0.0027
Money	$4+	V2-3	0.1725	0.2931	0.0057
Snack	pretzels	V3-1	0.0888	0.0222	0.5021
Snack	peanuts	V3-2	0.2582	0.0037	0.0108
Snack	M&M	V3-3	0.1513	0.0070	0.2945
Snack	Fritos	V3-4	0.1728	0.3475	0.0287
Snack	dried fr	V3-5	0.1903	0.3586	0.0126

Varianzbeitraege der Dummies
zu allen extrahierten Faktoren zusammen

Drink	Coke	V1-1	0.6580
Drink	7-Up	V1-2	0.5011
Drink	Dr.Peppe	V1-3	0.7241
Drink	Grape	V1-4	0.5362
Money	-$2	V2-1	0.7205
Money	$2-4	V2-2	0.6094
Money	$4+	V2-3	0.4715
Snack	pretzels	V3-1	0.6132
Snack	peanuts	V3-2	0.2728
Snack	M&M	V3-3	0.4528
Snack	Fritos	V3-4	0.5491
Snack	dried fr	V3-5	0.5616

Durch die nominalen Variablen erklaerte Varianz
(=discrimination measures)

hinsichtlich Faktor 1

		absolut	in %
V1	Drink	0.9072	33.2648
V2	Money	0.9584	35.1424
V3	Snack	0.8616	31.5928
Eigenwert		2.7272	100%

```
hinsichtlich Faktor 2
                                  absolut      in %
V1     Drink                       0.6873    30.4210
V2     Money                       0.8327    36.8568
V3     Snack                       0.7393    32.7222
---------------------------------------------------
Eigenwert                          2.2593     100%

hinsichtlich Faktor 3
                                  absolut      in %
V1     Drink                       0.8251    48.9900
V2     Money                       0.0103     0.6121
V3     Snack                       0.8488    50.3979
---------------------------------------------------
Eigenwert                          1.6842     100%
```

Korrelationen der Dummies mit den Faktoren

			Faktor 1	Faktor 2	Faktor 3
Drink	Coke	V1-1	0.4302	-0.0219	-0.7981
Drink	7-Up	V1-2	-0.5644	0.6388	0.1011
Drink	Dr.Peppe	V1-3	0.7137	0.1127	0.6189
Drink	Grape	V1-4	-0.4328	-0.7446	0.0546
Money	-$2	V2-1	-0.4768	0.8191	-0.0473
Money	$2-4	V2-2	0.9713	0.0656	-0.0655
Money	$4+	V2-3	-0.5551	-0.7235	0.1015
Snack	pretzels	V3-1	0.3251	-0.1628	-0.7731
Snack	peanuts	V3-2	0.5417	0.0654	0.1110
Snack	M&M	V3-3	0.4244	0.0913	0.5921
Snack	Fritos	V3-4	-0.4900	-0.6947	0.1997
Snack	dried fr	V3-5	-0.5141	0.7058	-0.1323

P30.8.3.2.3 Faktorwert-Berechnung

Die nachfolgenden Ergebnisse werden von Almo nur ausgegeben, wenn eine Faktorwertberechnung angefordert wurde.

Die Faktorwertberechnung ist eigentlich nur dann sinnvoll, wenn die extrahierten (und eventuell rotierten) Faktoren inhaltlich interpretierbar sind. In unserem Beispiel ist dies kaum der Fall. Die Faktorwertberechnung ist aber dann sinnvoll, wenn es darum geht, die Individuen im Faktorenraum abzubilden, um ihre räumliche Nähe zu den identifizierten Typen abzuschätzen (siehe dazu Abschnitt P30.8.3.6 *Korrespondenanalyse mit Positionierung der Individuen im (mehrdimensionalen) Faktorenraum*).

Unrotierte Faktor-Betaladungen (Faktorwert-Koeffizienten)
(Hauptkomponenten-Loesung)

			Faktor 1	Faktor 2	Faktor 3
Drink	Coke	V1-1	0.3155	-0.0194	-0.9477
Drink	7-Up	V1-2	-0.3017	0.4122	0.0875
Drink	Dr.Peppe	V1-3	0.5234	0.0998	0.7349
Drink	Grape	V1-4	-0.2545	-0.5285	0.0519
Money	-$2	V2-1	-0.3497	0.7251	-0.0562
Money	$2-4	V2-2	0.4749	0.0387	-0.0518
Money	$4+	V2-3	-0.2296	-0.3612	0.0679
Snack	pretzels	V3-1	0.2731	-0.1651	-1.0518
Snack	peanuts	V3-2	0.5379	0.0784	0.1785
Snack	M&M	V3-3	0.3566	0.0926	0.8055
Snack	Fritos	V3-4	-0.2881	-0.4931	0.1901

Snack dried fr V3-5	-0.3023	0.5009	-0.1259

Als Beispiel wird die Faktorwert-Berechnung fuer den 1.Datensatz gezeigt

Jede einzelne Variable wird standardisiert (sofern die Korrelations-Matrix faktorisiert wurde) und mit dem Faktorwert-Koeffizienten multipliziert. Die Formel ist folgende:

```
    (Variablenwert - Mittelwert) * FakwKoeff / Standabwg
```

Wurde die Kovarianzmatrix faktorisiert, dann ist Standabwg = 1
Wurde die (durchschnittliche) Kreuzproduktematrix faktorisiert, dann ist Mittelwert = 0 und Standabwg = 1

```
V1.01                           (0 - 0) * 0.315558 / 1
V1.02                         + (1 - 0) * -0.301706 / 1
V1.03                         + (0 - 0) * 0.52342 / 1
V1.04                         + (0 - 0) * -0.254514 / 1
V2.01                         + (1 - 0) * -0.349725 / 1
V2.02                         + (0 - 0) * 0.474916 / 1
V2.03                         + (0 - 0) * -0.229633 / 1
V3.01                         + (0 - 0) * 0.273175 / 1
V3.02                         + (0 - 0) * 0.5379 / 1
V3.03                         + (0 - 0) * 0.356604 / 1
V3.04                         + (0 - 0) * -0.288114 / 1
V3.05                         + (1 - 0) * -0.302338 / 1
Faktorwert-Variable V4       = -0.953769


V1.01                           (0 - 0) * -0.0194343 / 1
V1.02                         + (1 - 0) * 0.412204 / 1
V1.03                         + (0 - 0) * 0.0998026 / 1
V1.04                         + (0 - 0) * -0.528496 / 1
V2.01                         + (1 - 0) * 0.725113 / 1
V2.02                         + (0 - 0) * 0.0387363 / 1
V2.03                         + (0 - 0) * -0.36129 / 1
V3.01                         + (0 - 0) * -0.165143 / 1
V3.02                         + (0 - 0) * 0.0784489 / 1
V3.03                         + (0 - 0) * 0.0926147 / 1
V3.04                         + (0 - 0) * -0.49313 / 1
V3.05                         + (1 - 0) * 0.500954 / 1
Faktorwert-Variable V5       = 1.63827


V1.01                           (0 - 0) * -0.94773 / 1
V1.02                         + (1 - 0) * 0.0875001 / 1
V1.03                         + (0 - 0) * 0.734941 / 1
V1.04                         + (0 - 0) * 0.0519923 / 1
V2.01                         + (1 - 0) * -0.0562138 / 1
V2.02                         + (0 - 0) * -0.051869 / 1
V2.03                         + (0 - 0) * 0.06799 / 1
V3.01                         + (0 - 0) * -1.05183 / 1
V3.02                         + (0 - 0) * 0.178565 / 1
V3.03                         + (0 - 0) * 0.805531 / 1
V3.04                         + (0 - 0) * 0.190178 / 1
V3.05                         + (1 - 0) * -0.125964 / 1
Faktorwert-Variable V6       = -0.0946777
```

Faktorwert-Variable V4	V5	V6
-0.95376	1.638271	-0.094677
-0.77226	-1.382916	0.310160
1.06364	-0.145841	-2.051429
1.32837	0.097750	-0.821034
1.35494	0.231153	1.488603
-0.95376	1.638271	-0.094677
-0.77226	-1.382916	0.310160

1.06364	-0.145841	-2.051429
-0.77226	-1.382916	0.310160
1.53623	0.216987	0.861637
-0.95376	1.638271	-0.094677
1.35494	0.231153	1.488603
-0.95376	1.638271	-0.094677
-0.77226	-1.382916	0.310160
-0.77226	-1.382916	0.310160
1.53623	0.216987	0.861637
-0.78648	-0.388832	-0.005981
-0.95376	1.638271	-0.094677
1.35494	0.231153	1.488603
1.06364	-0.145841	-2.051429
-0.81945	-0.442216	0.345668
-0.21641	0.120229	-1.005704
-0.17473	0.143528	0.961021
-0.81945	-0.442216	0.345668
-0.21097	-1.054929	-0.931847

Almo ermöglicht es auch, was in den üblichen Programmen zur MCA nicht möglich ist, die Faktorladungsmatrix schief- oder rechtwinklig zu rotieren und für die rotierten Faktoren Faktorwerte zu ermitteln.

P30.8.3.3 MCA und bivariate Korrespondenzanalyse

Die im Almo-Dokument Nr.4 "Kanonische Analysen", Abschnitt P29.4 dargestellte bivariate Korrespondenzanalyse verwendet den Kalkül der kanonischen Korrelation. Wir wollen nun das dort vorgestellte Beispiel von Greenacre aus P29.4.2 als MCA rechnen. Wir erhalten folgende Ergebnisse

Multiple Korrespondenzanalyse (MCA)

Matrix der Faktorladungen
(=Matrix der "category quantifications")

	Faktor 1	Faktor 2	Faktor 3
V1-1	0.1919	1.4356	-2.4930
V1-2	-0.7557	1.8029	1.1838
V1-3	1.1107	0.0789	0.1810
V1-4	-0.6798	-0.4278	-0.1160
V1-5	0.5868	-0.5847	0.2838
V2-1	1.1478	0.2259	0.0312
V2-2	-0.2902	-1.0453	-0.7726
V2-3	-0.5729	-0.0545	0.9012
V2-4	-0.8573	1.4655	-0.9205

Bivariate Korrespondenzanalyse

Gemeinsame Matrix aller Variablen
der unstandardisierten, nicht-normalisierten Gewichte

			Faktor 1	Faktor 2	Faktor 3
Beruf	sen.Mana	V1-1	0.2405	1.9357	-3.4903
Beruf	jun.Mana	V1-2	-0.9471	2.4310	1.6574
Beruf	sen.Empl	V1-3	1.3920	0.1065	0.2535
Beruf	jun.Empl	V1-4	-0.8519	-0.5769	-0.1625
Beruf	Secretar	V1-5	0.7354	-0.7884	0.3973
Raucher	Nicht	V2-1	1.4385	0.3046	0.0437
Raucher	Leicht	V2-2	-0.3637	-1.4094	-1.0817
Raucher	Mittel	V2-3	-0.7180	-0.0735	1.2617
Raucher	Schwer	V2-4	-1.0744	1.9760	-1.2889

Wir erkennen, dass die Ergebnisse um eine Proportionalitätskonstante je Faktor verschieden sind. Um die Ergebnisse der bivariaten Analyse zu erhalten, müssen die Ergebnisse aus der MCA multipliziert werden mit:

1.2532 für den 1. Faktor
1.3484 für den 2. Faktor
1.4000 für den 3. Faktor

Die Formel zur Umformung der Ergebnisse lautet:

$$u_{ij} = f_{ij}\sqrt{\tfrac{2}{e_j}}$$

u_{ij} = unstandardisierte kanonische Gewichtszahl der Dummy i auf dem kanonischen Faktor j
f_{ij} = Faktorladung der Dummy i auf den Faktor j aus MCA
e_j = Eigenwert des Faktors j aus MCA

Die Proportionalitätskonstante p für Faktor j ergibt sich also sehr einfach aus

$$p_j = \sqrt{\tfrac{2}{e_j}}$$

Für den 1.Faktor ist sie also : $p_j = \sqrt{\tfrac{2}{1.2734}} = 1.2532$

Die kanonisch normalisierten Gewichtszahlen aus der bivariaten Korrespondenzanalyse erhalten wir dann auch sehr einfach:

$$r_j = e_j - 1$$
$$k_{ij} = u_{ij}\sqrt{r_j}$$

r_j = j-te kanonische Korrelation
k_{ij} = kanonisch normalisierte Gewichtszahl der Dummy i auf dem kanonischen Faktor j. Durch Einsetzen der Gleichung ineinander erhalten wir

$$\boldsymbol{k}_{ij} = f_{ij}\sqrt{\frac{\boldsymbol{e}_j - \mathbf{1}}{\boldsymbol{e}_j}} * \sqrt{2}$$

Die Konstante $\sqrt{2}$ könnte problemlos aus der Formel gestrichen werden. Auch bei der Ermittlung der Faktorwerte ("object scores") können bivariate und multiple Korrespondenanalysen auseinander abgeleitet werden. Werden die Faktorwertkoeffizienten der MCA multipliziert mit $\sqrt{2e_j}$ dann entstehen die unstandardisierten kanonischen Gewichtszahlen

$$u_{ij} = w_{ij}\sqrt{2e_j}$$

w_{ij} = Faktorwertkoeffizient der Dummy i für den Faktor j

Und umgekehrt können aus den Ergebnissen der bivariaten Korrespondenzanalyse die Faktorwertkoeffizienten aus der MCA ermittelt werden gemäß:

$$\boldsymbol{w}_{ij} = \frac{\boldsymbol{u}_{ij}}{\sqrt{2} \cdot \sqrt{\boldsymbol{r}_j + 1}}$$

Die Konstante $\sqrt{2}$ könnte wieder problemlos aus der Formel gestrichen werden.

P30.8.3.4 Johann Bacher: Modellprüfgrößen für die Korrespondenzanalyse

Die bivariate und multiple Korrespondenzanalyse können als explorative Verfahren zur räumlichen Darstellung einer zwei- bzw. mehrdimensionalen Tabelle betrachtet werden. Dies war der Ausgangspunkt der Entwicklung der Korrespondenzanalyse im französischsprachigen Raum (siehe dazu auch Abschnitt P29.4).

Für die bivariate Korrespondenzanalyse bestand das Ziel darin, eine räumliche Darstellung zu finden, bei der die 1. Dimension ein Maximum des Chi-Quadratwertes erklärt, die 2. Dimension ein Maximum des verbleibenden Chi-Quadratwertes usw. Zwischen dem Chi-Quadratwert und den Eigenwerten der bivariaten Korrespondenzanalyse besteht folgender Zusammenhang:

CHI/N = e1 + e2 + ...

CHI ist der Chi-Quadratwert der untersuchten zweidimensionalen Tabelle, N ist die in die Tabellierung eingehende Fallzahl und e1, e2, ... sind die berechneten Eigenwerte. Der mittlere Chi-Quadratwert CHI/N wird in der Literatur zur Korrespondenzanalyse als **Trägheit (="inertia")** bezeichnet.

Diese mathematische Beziehung zwischen Chi-Quadratwert bzw. Trägheit und den Eigenwerten läßt sich für das in P29.4 und P30.8.2.1 untersuchte Beispiel von Greenacre leicht nachvollziehen.

1. Wir berechnen mit dem Almo-Tabellierungsprogramm P10 den Chi-Quadratwert für die beiden Variablen "berufliche Stellung" (=V1; 5 Ausprägungen) und "Raucherverhalten" (=V2; 4 Ausprägungen). Für die Tabelle ergibt sich ein Chi-Quadratwert von 16.4416. Da 193 Fälle in die Berechnung der Tabelle eingingen, ergibt sich eine Trägheit von 0.08519 (=16.4416/193).
2. Wir rechnen mit Programm Prog29m2 eine bivariate Korrespondenzananalyse. Die maximale Faktorenzahl ist gleich dem Minimum aus der Zeilen- und Spaltenzahl minus 1. Im Beispiel ist die maximale Faktorenzahl somit gleich 3. Die Eigenwerte dieser drei Faktoren sind:

Faktor	Kanonische Korrelation	Eigenwert	erklärter Chi-Quadratwert in %	kumuliert
1	0.27342	0.07476	87.8 %	87.8 %
2	0.10009	0.01002	11.8%	99.6%
3	0.02034	0.00041	0.4%	100.0%
Gesamt		0.08519	= Trägheit 100.0 %	

Werden die Eigenwerte aufaddiert ergibt sich die Trägheit bzw. der mittlere Chi-Quadratwert (=CHI/N) von 0.08519. Setzen wir die Eigenwerte der einzelnen Faktoren zu diesem Gesamtwert in Beziehung, ergibt sich folgendes Bild: Der 1. Faktor erklärt 87.8 % des Chi-Quadratwertes, der 2. Faktor 11.8 % und der letzte Faktor 0.4 %. Die beiden ersten Faktoren zusammen erklären 99.6 %. Die Interpretation "erklärter Chi-Quadratwert" ist zulässig, da bei Relativaussagen dem Skalierungsfaktor 1/N keine Bedeutung zukommt.

Das Konzept der Trägheit kann auch auf die multiple Korrespondenzanalyse übertragen werden. Die Trägheit ist in diesem Fall definiert als die Summe der bivariaten Chi-Quadratwerte dividiert durch die Fallzahl.

Betrachten wir dazu das in Abschnitt P30.8.2 untersuchte Beispiel von Carroll und Green. Hier wird eine multiple Korrespondenzanalyse für folgende drei Variablen gerechnet: "Lieblingsgetränk" ("drink", V1, 4 Ausprägungen), "Ausgaben für Lieblingsgetränke" ("money", V2, 3 Ausprägungen) und "Lieblingsimbiß" ("snack", V3, 5 Ausprägungen).

Für diese drei Variablen ergeben sich folgende bivariate Chi-Quadratwerte:

Tabelle	Chi-Quadratwerte aus Programm P10		Zahl der Zellen
V1 mit V2	32.3232	(Signifikanz: 99.993)	4*3 = 12
V1 mit V3	34.4335	(Signifikanz: 99.917)	4*5 = 20
V2 mit V3	32.0346	(Signifikanz: 99.978)	3*5 = 15
Gesamt	98.7913		47

Die Summe der bivariaten Chi-Quadratwerte ist gleich 98.7913. Wir wollen diese Größe im folgenden als Chi-Quadratwert für die paarweise Unabhängigkeit der Variablen bezeichnen. Dieser Chi-Quadratwert für paarweise Unabhängigkeit darf nicht mit dem im Programm 11 berechneten Chi-Quadratwert auf allseitige Unabhängigkeit verwechselt werden. Nur im Fall einer statistischen Unabhängigkeit sind beide Chi-Quadratwerte gleich 0, andernfalls ist der Chi-Quadratwert auf allseitige Unabhängigkeit i.d.R. größer.

Da in die Tabellierung 25 Personen eingingen, beträgt die Trägheit 3.951652.

Auf der Grundlage des Chi-Quadratwertes für paarweise Unabhängigkeit bzw. der Trägheit können nun für die multiple Korrespondenzanalyse folgende Modellprüfgrößen berechnet werden:

1. den durch die ersten h Faktoren erklärten Chi-Quadratwert (Chi-Quadratbeitrag der ersten h Faktoren)
2. die Signifikanz des Faktors h+1.

Zur Berechnung dieser beiden Modellprüfgrößen wird die bei h Faktoren (z.B. h=1, h=2, ..) verbleibende Residualmatrix R(h) benötigt, da zwischen dem Chi-Quadratwert für paarweise Unabhängigkeit bzw. der Trägheit und den Eigenwerten keine mathematisch nachweisbare Beziehung besteht, wenn mehr als zwei Variablen untersucht werden. Auf technische Details der Berechnung soll hier nicht eingegangen werden. Wir geben nur die entsprechenden von unserem Programm berechneten Modellprüfgrößen wieder. Diese Modellprüfgrößen können im Maskenprogramm Prog30m5 berechnet werden.

Es muss die Optionsbox „verschiedene Programm-Optionen" geöffnet werden und das 4. Eingabefeld auf „1" gesetzt werden. Siehe P30.8.0, Erläuterungen zu Eingabebox 15.

ALMO liefert folgende Ausgabe:

```
Maßzahlen fuer faktorenanalytische Interpretation:

mittlere Residuenabweichungen:
```

Faktoren	RMR	GFIR
0	0.290	0.000
1	0.191	0.565
2	0.110	0.857
3	0.049	0.971
4	0.048	0.973

```
RMR  = Wurzel aus den mittleren Reisudenquadraten nur fuer Dummies von
       unterschiedlichen nominalen Variablen
GFIR = Anpassungsindex zu RMR-Koeffizient (=erklärter Chi-Quadrat-Wert)
```

Der GFIR-Index ("goodness-of-fit index for residuals") entspricht dem bei der bivariaten Korrespondenzanalyse angeführten kumulierten erklärten Chi-Quadratwerten. Der 1. Faktor erklärt also 56.5 %, die beiden ersten Faktoren 85.7 % usw. des Chi-Quadratwertes auf paarweise Unabhängigkeit. Der GFIR-Index wird aus den RMR-Werten wie folgt berechnet:

```
GFIR(h)=1-RMR(h)*RMR(h)/RMR(0)*RMR(0)
```

Für den ersten Faktor ergibt sich ein Wert von:

`GFIR(1)=1-0.191*0.191/0.290*0.290=0.566`

(Die Abweichung der 3. Kommastelle entsteht durch Rundungsfehler).

Vergleicht man die GFIR-Werte mit den erklärten Varianzen, wie sie bei der Faktorenanalyse berechnet werden, so zeigt sich, dass die erklärten Varianzen die Modellanpassung unterschätzen. Sie vermitteln ein zu pessimistisches Bild über die Modellanpassung.

Faktoren	Eigenwert	erkl.Varianz	GFIR in %
1	2.7272	30.3 %	56.5 %
2	2.2593	55.4 %	85.7 %
3	1.6842	74.1 %	97.1 %
4	1.0081	85.3 %	97.3 %

Es läßt sich zudem zeigen, dass die Gesamtvarianz bei der MCA-Analyse eine rein rechentechnisch bedingte Größe ist. Wir empfehlen daher die Verwendung der GFIR-Werte.

Das Quadrat des RMR-Wertes für 0 Dimensionen ist gleich der Trägheit dividiert durch die Zahl der Zellen, die in die Berechnung des RMR-Wertes eingingen. In unserem Beispiel sind dies 47 Zellen. Der quadrierte RMR-Wert ist daher

RMR(0)*RMR(0)=(CHI/N)/47=Trägheit/47= 3.951652/12=0.0841=0.290*0.290

Für die Signifikanz der Faktoren ergibt sich folgendes Bild:
Chi-Quadrat-Pruefgroessen:

Dimensionen	Chi	df	100*(1-p)
0	98.79	37	100
1	30.08	18	96
2	11.59	7	88
3	0.00	-3	0
4	0.00	-6	0

Bei 0 Faktoren (Dimensionen) beträgt der Chi-Quadratwert - wie wir bereits wissen - 98.79. Er ist mit 37 Freiheitsgraden von 0 verschieden. Dies bedeutet, dass eine MCA-Analyse sinnvoll ist, da ein signifikanter Zusammenhang vorliegt, der durch eine räumliche Darstellung erklärt werden kann. Nach dem 1. Faktor ergibt sich ein Chi-Quadratwert von 30.08. Auch dieser ist signifikant von Null verschieden. Der Zusammenhang kann somit nicht ausschließlich durch einen Faktor erklärt werden. Auch der 2. Faktor leistet noch eine signifkante Erklärung, der 3. Faktor dagegen nicht mehr, da der Chi-Quadratwert nach 2 Faktoren (=11.59) nicht mehr signifikant ist. Der Chi-Quadratwert nach 3 und 4 Dimensionen kann nicht berechnet werden. Es ergibt sich eine negative Zahl von Freiheitsgraden. Diese entsteht dadurch, dass erwartete Häufigkeiten berechnet werden, die kleiner/gleich 0 sind.

Weitere Modellprüfgrößen:

Die bisher behandelten Modellprüfgrößen der MCA sind in Analogie zur bivariaten Korrespondenzanalyse definiert. Sie geben Anhaltspunkte über die Brauchbarkeit einer faktorenanalytischen Interpretation der MCA. Mit faktorenanalytischer Interpretation ist gemeint, dass bei der Interpretation der Ergebnisse nach zugrunde liegenden gemeinsamen Dimensionen gesucht wird. Von dieser faktorenanalytischen Interpretation ist eine clusteranalytische Interpretation zu unterscheiden, bei der versucht wird, die Ausprägungen zu Gruppen zusammenzufassen. Für die Beurteilung der clusteranalytischen und der faktorenanalytischen Interpretation sind unterschiedliche Modellprüfgrößen erforderlich. Allgemein wird für eine clusteranalytische Interpretation eine größere Faktorenzahl benötigt als für eine faktorenanalytische Interpretation.

Unser Programm berechnet insgesamt folgende Prüfgrößen:

Maßzahlen für die faktorananalytische Interpretation:

1. GFIR-Index (siehe oben)
2. Chi-Quadrattest für paarweise Unabhängigkeit (siehe oben)
3. kophenetische Korrelation: Dies gibt an, wie gut die reproduzierten mit den empirischen MCA-Werten korrelieren. Eine Korrelation nahe bei 1.0 bedeutet eine gute Modellanpassung.

Maßzahlen für die clusteranalytische Interpretation
(für die Interpretation der Distanzen):

4. mittlere Distanzabweichung
5. Anpassungsindex zu den Distanzabweichungen. Ein Wert nahe bei 1.0 bedeutet eine gute Modellanpassung.
6. Streß-Koeffizient nach Kruskal. Eine Wert nahe bei Null bedeutet eine gute Modellanpassung.
7. Anpassungsindex für den Streßkoeffizienten: Ein Wert nahe bei 1.0 bedeutet eine gute Modellanpassung.
8. Gamma-Korrelationskoeffizient. Ein Wert nahe bei 1.0 bedeutet eine gute Modellanpassung.

Für das Beispiel von Carroll und Green ergeben sich folgende Maßzahlen für die clusteranalytische Interpretation:

```
mittlere Distanzabweichungen:
```

Dimensionen	DIS	GFID
0	7.631	0.000
1	5.569	0.467
2	4.341	0.676
3	2.637	0.881
4	1.053	0.981

```
DIS  = Wurzel aus den mittleren Distanzdifferenzen nur für Dummies von
       unterschiedlichen nominalen Variablen
GFID = Anpassungsindex zu DIS
```

```
Stress-Koeffizienten:
```

Dimensionen	Streß	GFID
1	0.311	0.811
2	0.287	0.788
3	0.185	0.876
4	0.048	0.991

```
Gamma-Koeffizienten:
```

Dimensionen	Gamma
1	0.469
2	0.473
3	0.706
4	0.945

Den Maßzahlen ist zu entnehmen, dass für eine Interpretation der Distanzen mindestens 3 Dimensionen benötigt werden. Erst bei 3 Dimensionen ergeben sich Anpassungsindizes größer 0.85 bzw. ein Gamma größer 0.7.

P30.8.3.5 Korrespondenzanalyse mit Positionierung von Gruppen von Untersuchungseinheiten im (mehrdimensionalen) Faktorenraum ("supplementary variables")

Unsere folgenden Ausführungen gelten nicht nur für den multiplen, sondern auch für den bivariaten Fall. Der Benutzer kann wieder zwischen 3 Verfahren der nominalen Faktorenanalyse wählen:

1. Der multiplen Korrespondenzanalyse (MCA)
2. dem Blockdiagonal-Verfahren
3. der Faktorisierung von Dummies

Betrachten wir das Autofahrer-Beispiel, das wir nun etwas abwandeln:

Die nominalen Variablen, die faktorisiert werden, sind:

```
Autokauf: Porsche, Mercedes, VW
Fahrstil: Aggressiv, normal, zurückhaltend
```

Eine nominale Faktorenanalyse über diese 2 nominalen Variablen führt zu einem 2-dimensionalen Faktorenraum, in dem folgende Punkte dicht beieinander liegen:

```
Mercedes und normaler Fahrstil
Porsche  und aggressiver Fahrstil
VW und zurückhaltend
```

Die Untersuchungspersonen werden nun gruppiert entsprechend ihrem Beruf in

```
Selbständige, Arbeitnehmer, Führungsposition,
```

und entsprechend ihrem Alter in

```
jung, mittel, alt
```

Es wird unterschieden zwischen:

1. den zu faktorisierenden Variablen (Autokauf, Fahrstil)
2. den Gruppierungsvariablen (="supplementary Variables") im Beispiel: der Beruf, das Alter

Die Vorgehensweise ist folgende: Zuerst wird eine nominale Faktorenanalyse für die zu faktorisierenden Variablen (Autokauf,Fahrstil) gerechnet. Dann werden für alle Untersuchungseinheiten die Faktorwerte für die 2 extrahierten Faktoren berechnet. Dann wird für jede Berufsgruppe der durchschnittliche Faktorwert in den beiden Faktoren gebildet. Diese durchschnittlichen Faktorwerte werden an die Faktorladungsmatrix angehängt. Damit werden die Berufsgruppen im 2-dimensionalen Faktorraum "positioniert". Genau so wird für die 3 Altersgruppen verfahren.

Wurden die Untersuchungseinheiten gewichtet, z.B. durch die Anweisung

```
„Wenn Geschlecht=1 dann Gewicht1=1.5; EndeWenn"
```

dann wird auch der durchschnittliche Faktorwert durch entsprechende Gewichtung errechnet – im Beispiel würden also Männer (Geschlecht = 1) mit einem Gewichtsfaktor von 1.5 in diesen durchschnittlichen Faktorwert eingehen.

Das nachfolgende Almo-Prog erlaubt es, beliebig viele nominale Variable mit der maximalen Faktorenzahl zu faktorisieren und beliebig viele Gruppierungsvariable zu verwenden

Eingabe mit Maskenprogramm Prog31m1

Prog31m1.Msk
Bivariate und multiple Korrespondenzanalyse
(als Faktorenanalyse mit nominalen Variablen)
mit Positionierung von Gruppen von Untersuchungs-
einheiten im (mehrdimensionalen) Faktorenraum
("supplementary Variables")

Zum Verhältnis der bivariaten zur
multiplen Korrespondenzanalyse --> **Hilfe 79**

Beispiel:
Die nominalen Variablen, die faktorisiert werden, sind:
Autokauf: Porschen, Mercedes, VW
Fahrstil: Aggresiv, normal, zurückhaltend

Eine nominale Faktorenanalyse über diese 2 nominalen Var. führt zu einem 2-dimensionalen Faktorenraum, in dem folgende Punkte dicht beieinanderder liegen:
Mercedes, normaler Fahrstil
Porsche, aggresiver Fahrstil
VW, zurückhaltend

Die Untersuchungspersonen werden nun gruppiert entsprechend ihrem Beruf in
Selbständige, Arbeitnehmer, Führungsposition,

und entsprechend ihrem Alter in
jung, mittel, alt

Wir nennen diese Variable "Positionierungsvariable" oder "Gruppierungsvariable"

Es wird unterschieden zwischen:
1. den zu faktorisierenden Variablen (Autokauf, Fahrstil)
2. den Gruppierungsvariablen (="supplementary Variables")
im Beispiel: der Beruf, das Alter

Die Vorgehensweise ist folgende:
Zuerst wird eine nominale Faktorenanalyse für die zu faktorisierenden Variablen (Autokauf,Fahrstil) gerechnet. Dann werden für alle Untersuchungseinheiten die Faktorwerte für die 2 extrahierten Faktoren berechnet. Dann wird für jede Berufsgruppe der durchschnittliche Faktorwert in den beiden Faktoren gebildet. Diese durchschnittlichen Faktorwerte werden an die Faktorladungsmatrix angehängt. Damit werden die Berufsgruppen im 2-dimensionalen Faktorraum "positioniert". Genau so wird für die 3 Altersgruppen verfahren.

Das nachfolgende Almo-Prog erlaubt es, beliebig viele nominale Variable mit der maximalen Faktorenzahl zu faktorisieren und beliebig viele Gruppierungsvariable zu verwenden

Was ist ein Kurzprogramm ? --> Hilfe
Bedienung --> Hilfe

1 Speicher fuer x Variable Hilfe
Vereinbare Variable= 20 ;

2 Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

3 Datei der Variablennamen | Hilfe

"C:\Almo7\TESTDAT\Auto.nam"

zeige — zeige = Namensdatei in Output zeigen; leer = nicht

4 Freie Namensfelder | Hilfe

Name1=Auto:Porsche,Mercedes,VW;
Name2=Beruf:Selbst,Arbeitn,Führungspos;
Name3=Fahrstil:aggressiv,normal,zurückhaltend;
Name4=Altersgruppe:jung,mittelalter,alt;

erzeuge zusätzliche Namensfelder

5 Datei aus der gelesen wird | Hilfe bei Datei-Problemen

"C:\Almo7\Testdat\Auto2.fre"

frei — Format der Daten | Hilfe

V1:4 — der Datensatz enthält diese Variablen. Bei Format DIREKT schreiben Sie: alle_V

6 Wenn Dateiformat FIX oder Nicht-Standard-FREI | Hilfe

7 zu faktorisierende nominale Variable | Hilfe

Auto,Fahrstil

8 nominale Positionierungsvariable (Gruppierungsvariable) | Hilfe

Beruf,Altersgruppe

9 Verfahren der nominalen Faktorenanalyse

multiple_Korrespondenz

0= gewöhnliche Faktorenanalyse der in Dummies aufgelösten nominalen Variablen

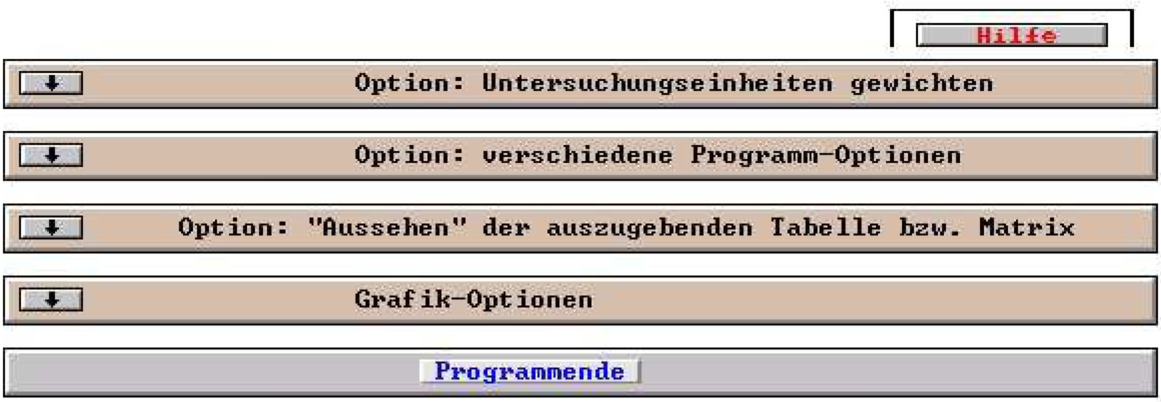
Blockdiagonale = nominale Faktorenanalyse mit Blockdiagonal-Matrix nach McDonald

multiple_Korrespondenz = multiple Korrespondenz-Analyse

10 Option: Faktoren

11 Option: Ein- und Ausschliessen von Untersuchungseinheiten

12 Option: Umkodierungen und Kein-Wert-Angaben



Erläuterungen zu den Eingabeboxen
Alle Eingabeboxen wurden bereits bei Prog30m5 bzw. Prog30m2 erläutert. Siehe Abschnitt P30.8.0 bzw. P30.3.1.

Ergebnisse:
Im 1. Teil der Ergebnisliste werden die Ergebnisse der MCA dargestellt. Wir geben hier nur die Matrix der Faktorenladung und der Faktorwertkoeffizienten wieder

```
Matrix der Faktorladungen
(=Matrix der "category quantifications")
```

			Faktor 1	Faktor 2
Auto	Porsche	V1-1	1.2951	-0.2469
Auto	Mercedes	V1-2	-0.3868	1.1788
Auto	VW	V1-3	-0.8325	-0.8542
Fahrstil	aggressi	V3-1	1.3614	-0.3638
Fahrstil	normal	V3-2	-0.2414	1.2121
Fahrstil	zurückha	V3-3	-0.8430	-0.7457

```
Unrotierte Faktor-Betaladungen (Faktorwert-Koeffizienten)
```

			Faktor 1	Faktor 2
Auto	Porsche	V1-1	0.7748	-0.1698
Auto	Mercedes	V1-2	-0.2314	0.8109
Auto	VW	V1-3	-0.4981	-0.5876
Fahrstil	aggressi	V3-1	0.8145	-0.2503
Fahrstil	normal	V3-2	-0.1444	0.8337
Fahrstil	zurückha	V3-3	-0.5043	-0.5129

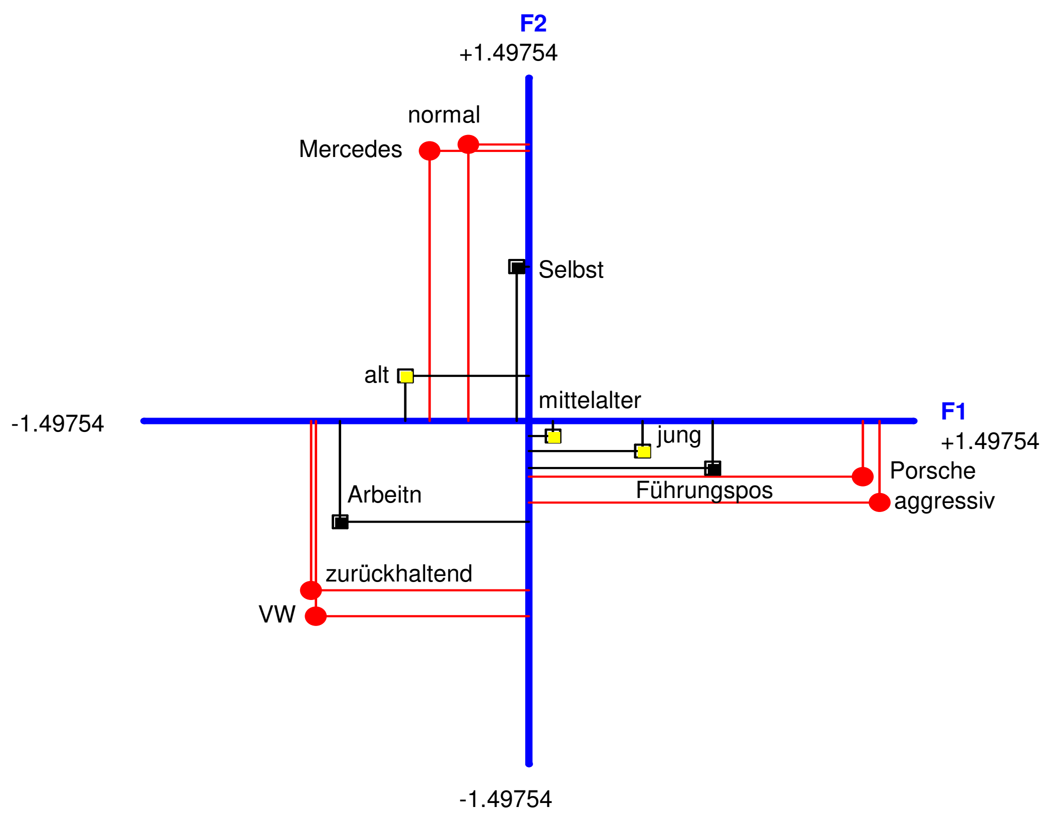
Im 2. Teil der Ergebnisliste wird dann die gemeinsame Matrix der Faktorenladung und der positionierten Gruppen gebracht.

```
Gemeinsame Matrix der Faktorladungen und der positionierten Gruppen

Die Gruppen wurden gebildet nach den Merkmalen: V2 Beruf
                                                V4 Altersgruppe
Sie wurden im Faktorenraum positioniert gemäß ihrem mittleren
Faktorwert in den Faktoren 1 bis 2
```

			Faktor 1	Faktor 2	
Auto	Porsche	V1-1	1.2951	-0.2469	
Auto	Mercedes	V1-2	-0.3868	1.1788	
Auto	VW	V1-3	-0.8326	-0.8542	faktorisierte Variable
Fahrstil	aggressi	V3-1	1.3614	-0.3638	
Fahrstil	normal	V3-2	-0.2414	1.2121	
Fahrstil	zurückha	V3-3	-0.8430	-0.7457	
Beruf	Selbst	V2-1	-0.0538	0.6762	
Beruf	Arbeitn	V2-2	-0.7316	-0.4457	
Beruf	Führungs	V2-3	0.7199	-0.2113	
Altersgr	jung	V4-1	0.4457	-0.1359	Positionierungsvariable
Altersgr	mittelal	V4-2	0.0955	-0.0695	(Gruppierungsvariable)
Altersgr	alt	V4-3	-0.4863	0.1997	

Die Faktorladungen der Positionierungsvariablen sind durchschnittliche Faktorwerte. Nach Klick auf den Grafik-Knopf präsentiert dann Almo folgendes Koordinatendiagramm.

Faktorladungen und
positionierte Gruppen



Die kreisrunden Punkte repräsentieren die Dummies der faktorisierten Variablen. Die schwarzen Vierecke stellen die Berufe und die hellen Vierecke das Lebensalter dar.

Wir erkennen, dass die Gruppe der Selbständigen sich nahe beim Typ Mercedes/normaler Fahrstil" befindet etc. Die 3 Altersgruppen sind nicht eindeutig den Typen zuzuordnen.

Hinweis: Wenn Sie als Positionierungsvariable dieselben Variablen, wie die zu faktorisierenden angeben, dann liegen die schwarzen und hellen Vierecke auf den entsprechenden runden Punkten. Dies ist der geometrische Ausdruck dafür, dass die Faktorisierung für eine Ausprägung einer nominalen Variablen gleich den Mittelwert der Faktorenwerte derjenigen Personen ist, die zu dieser Ausprägung gehören. Beispiel: Der mittlere Faktorenwert der Selbständigen ist gleich der Faktorenladung für die Dummy „Selbständig".

P30.8.3.6: Korrespondenanalyse mit Positionierung der Individuen im (mehrdimensionalen) Faktorenraum

Die folgenden Ausführungen gelten nicht nur für den multiplen, sondern auch für den bivariaten Fall.

Betrachten wir ein Beispiel:

Die nominalen Variablen, die faktorisiert werden, sind:

```
Autokauf: Porsche, Mercedes, VW
Fahrstil: Aggressiv, normal, zurückhaltend
```

Eine nominale Faktorenanalyse über diese 2 nominalen Variablen führt zu einem 2-dimensionalen Faktorenraum, in dem folgende Punkte dicht beieinander liegen:

```
Mercedes und normaler Fahrstil
Porsche  und aggressiver Fahrstil
VW       und zurückhaltend
```

Die einzelnen Untersuchungspersonen werden nun mit ihren Faktorwerten (=object scores) im Faktorenraum positioniert.

Untersuchungspersonen, die in den zu faktorisierenden Variablen (im Beispiel: Autokauf und Fahrstil) gleich sind, haben auch die gleichen Faktorwerte. Es müssen also nicht alle individuellen Untersuchungspersonen im Faktorraum positioniert werden. Es genügt, wenn sie nach den zu faktorisierenden Variablen gruppiert werden und diese Gruppen im Faktorenraum positioniert werden. In unserem Beispiel gibt es 9 Gruppen:

```
Mercedes/aggressiv,
Mercedes/normal,........,VW/zurückhaltend.
```

Die Vorgehensweise ist folgende:
Zuerst wird eine nominale Faktorenanalyse für die zu faktorisierenden Variablen (Autokauf, Fahrstil) gerechnet.
Dann werden die Untersuchungseinheiten gruppiert und für diese Gruppen die Faktorwerte für die extrahierten Faktoren berechnet. Diese Faktorwerte werden an die Faktorladungsmatrix angehängt. Damit werden die 9 Gruppen im Faktorraum "positioniert".

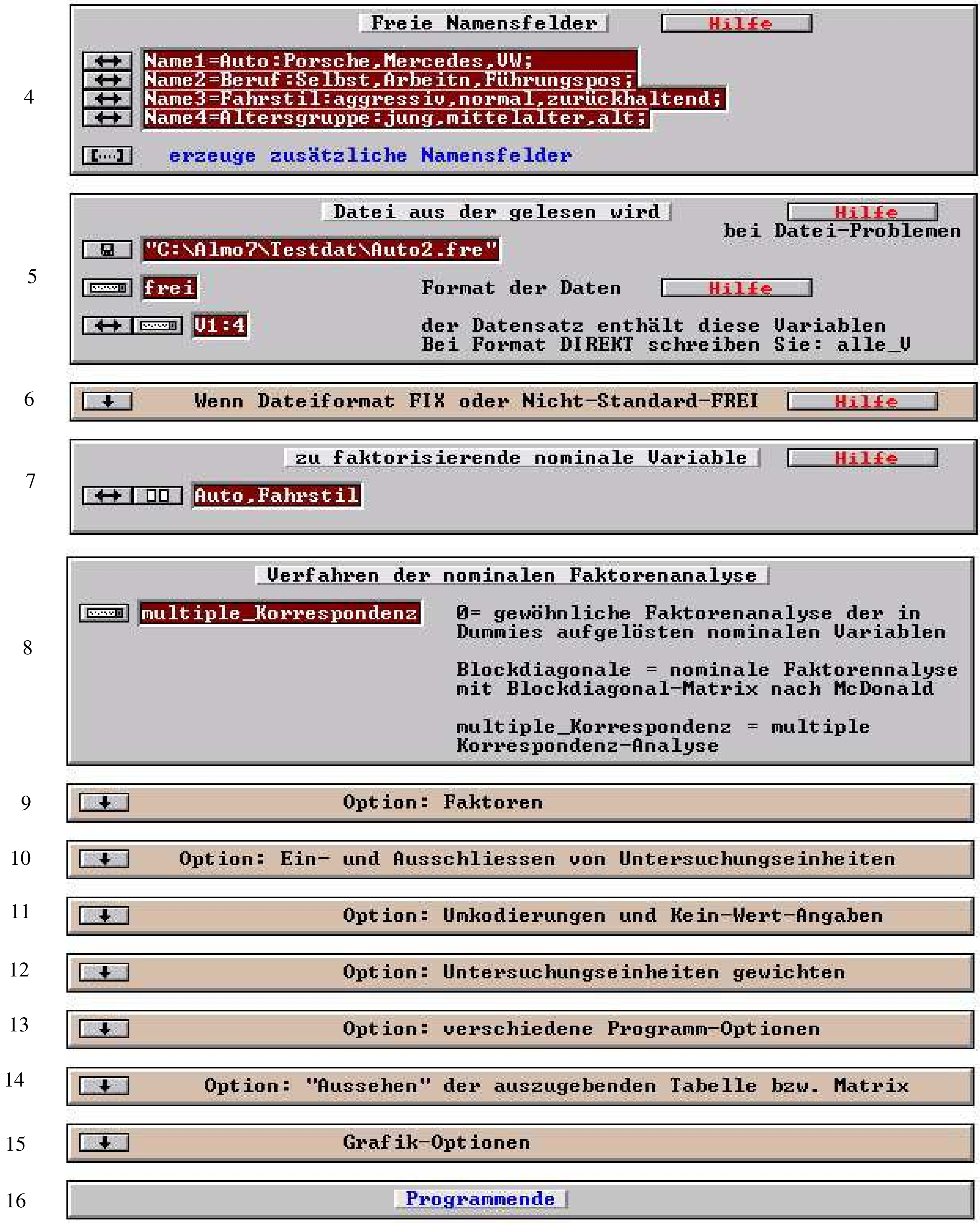
Eingabe mit Maskenprogramm

```
                  Prog31m2.Msk
        Bivariate und multiple Korrespondenzanalyse
        (als Faktorenanalyse mit nominalen Variablen)
    mit Positionierung der individuellen Untersuchungs-
        einheiten im (mehrdimensionalen) Faktorenraum

   Zum Verhältnis der bivariaten zur
   multiplen Korrespondenzanalyse    -->  [ Hilfe ]

Beispiel:
Die nominalen Variablen, die faktorisiert werden, sind:
         Autokauf: Porschen, Mercedes, VW
         Fahrstil: Aggresiv, normal, zurückhaltend

Eine nominale Faktorenanalyse über diese 2 nominalen Var.
führt zu einem 2-dimensionalen Faktorenraum, in dem
folgende Punkte dicht beieinanderder liegen:
           Mercedes und normaler Fahrstil
           Porsche  und aggresiver Fahrstil
           VW       und zurückhaltend

Die einzelnen Untersuchungspersonen werden nun mit ihren
Faktorwerten (=object scores) im Faktorenraum positioniert.

Untersuchungspersonen, die in den zu faktorisierenden
Variablen (im Beispiel: Autokauf und Fahrstil) gleich
sind, haben auch die gleichen Faktorwerte. Es müssen also
nicht alle individuellen Untersuchungspersonen im Faktor-
raum positioniert werden. Es genügt, wenn sie nach den zu
faktorisierenden Variablen gruppiert werden und diese
Gruppen im Faktorenraum positioniert werden. In unserem
Beispiel gibt es 9 Gruppen: Mercedes/aggresiv, Mercedes/
normal,.........,VW/zurückhaltend.

Die Vorgehensweise ist folgende:
Zuerst wird eine nominale Faktorenanalyse für die zu
faktorisierenden Variablen (Autokauf,Fahrstil) gerechnet.
Dann werden die Untersuchungseinheiten gruppiert und für
diese Gruppen die Faktorwerte für die extrahierten Faktoren
berechnet. Diese Faktorwerte werden an die Faktorladungs-
matrix angehängt. Damit werden die 9 Gruppen im Faktorraum
"positioniert".

Das nachfolgende Almo-Prog erlaubt es, beliebig viele
nominale Variable mit der maximalen Faktorenzahl zu
faktorisieren und die entsprechenden Gruppen gleicher
Untersuchungspersonen zu bilden
          Was ist ein Kurzprogramm ? -->  [ Hilfe ]
                          Bedienung  -->  [ Hilfe ]
```

1
```
          Speicher fuer x Variable  [ Hilfe ]
Vereinbare Variable= 20 ;
```

2
```
[↓]  Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert
```

3
```
          Datei der Variablennamen  [ Hilfe ]
[↔|💾] "C:\Almo7\TESTDAT\Auto.nam"
[↔|↓]  zeige                  zeige = Namensdatei in Output zeigen
                              leer  = nicht
```



Erläuterungen zu den Eingabeboxen
Alle Eingabeboxen wurden bereits bei Prog30m5 bzw. Prog30m2 erläutert. Siehe Abschnitt P30.8.0 bzw. P30.3.1.

Ergebnisse:
Gegen Ende der Ergebnisliste bringt Almo folgende Ausgabe:

Die Untersuchungseinheiten werden mit ihren Faktorwerten (=object scores) im Faktorenraum positioniert.

Untersuchungseinheiten, die in den zu faktorisierenden Variablen gleich sind, haben auch die gleichen Faktorwerte. Sie bilden eine Gruppe mit gleichen Faktorwerten. Almo positioniert diese Gruppen und nicht die einzelnen Individuen

Die Merkmalskombinationen in nachfolgender Tabelle setzen sich aus folgenden Elementen zusammen

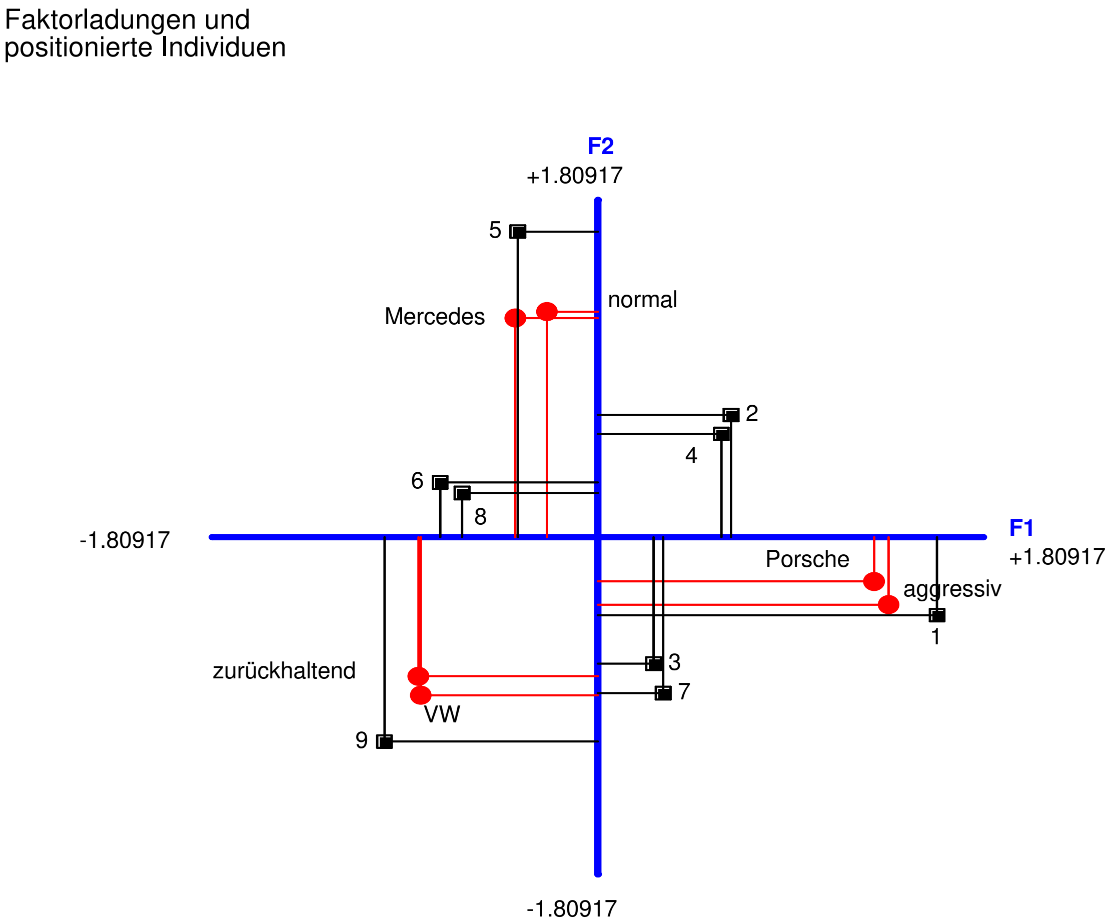
- 1 A Auto
  - A1 Porsche
  - A2 Mercedes
  - A3 VW
- 3 B Fahrstil
  - B1 aggressiv
  - B2 normal
  - B3 zurückhaltend

Gruppe	Variable	Merkmalskombination	Haeufigkeit
1	V20-1	A1 B1	8
2	V20-2	A1 B2	2
3	V20-3	A1 B3	1
4	V20-4	A2 B1	1
5	V20-5	A2 B2	7
6	V20-6	A2 B3	3
7	V20-7	A3 B1	1
8	V20-8	A3 B2	2
9	V20-9	A3 B3	9

Gemeinsame Matrix der Faktorladungen und der positionierten Gruppen
Die Gruppen wurden so gebildet wie oben angegeben

			Faktor 1	Faktor 2	
Auto	Porsche	V1-1	1.2951	-0.2469	
Auto	Mercedes	V1-2	-0.3868	1.1788	
Auto	VW	V1-3	-0.8326	-0.8542	faktorisierte Variable
Fahrstil	aggressi	V3-1	1.3614	-0.3638	
Fahrstil	normal	V3-2	-0.2414	1.2121	
Fahrstil	zurückha	V3-3	-0.8430	-0.7457	
Gruppe		V20-1	1.5893	-0.4201	
Gruppe		V20-2	0.6304	0.6639	
Gruppe		V20-3	0.2704	-0.6828	
Gruppe		V20-4	0.5830	0.5606	
Gruppe		V20-5	-0.3758	1.6447	
Gruppe		V20-6	-0.7358	0.2979	Gruppen von (gleichen)
Gruppe		V20-7	0.3163	-0.8379	Individuen
Gruppe		V20-8	-0.6425	0.2461	
Gruppe		V20-9	-1.0025	-1.1006	

Nach Klick auf den Grafik-Knopf wird folgende Grafik präsentiert.



Die runden Punkte stellen die Ausprägungen der faktorisierten Variablen dar. Die schwarzen Vierecke repräsentieren die Gruppen gleicher Individuen.

P30.8.3.7 Johann Bacher: Multiple Korrespondenzanalyse und Clusteranalyse

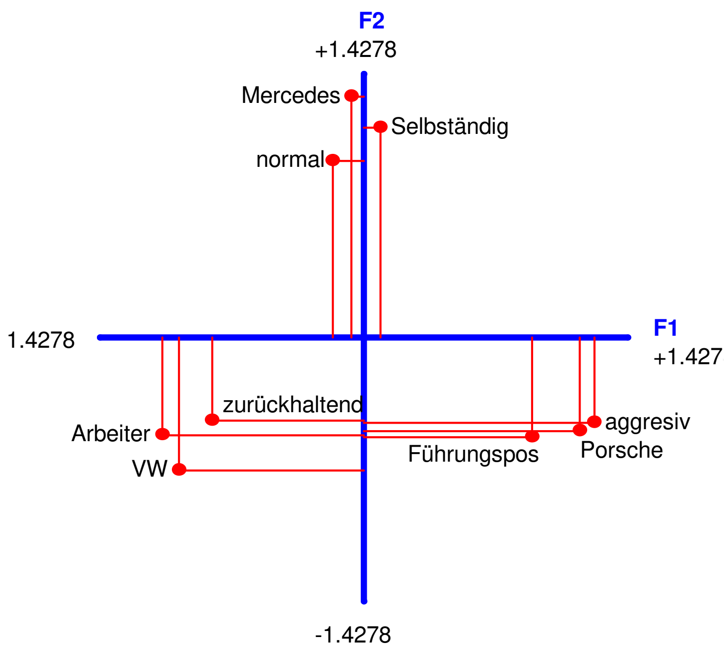
Die folgenden Ausführungen gelten in entsprechender Weise auch für die beiden anderen Methoden der nominalen Faktorenanalyse, also die "einfache Faktorisierung der Dummies" und das Blockdiagonal-Verfahren.

Die multiple Korrespondenzanalyse wird häufig zu einer graphischen Bestimmung von Typen/Clustern eingesetzt. Das Vorgehen besteht darin, dass in der zwei- oder seltener in der dreidimensionalen Darstellung die untersuchten Ausprägungen der in die Analyse einbezogenen nominalen Variablen, die eng beieinander liegen, zu Clustern oder Typen zusammengefaßt werden. Es wird also eine variablen- bzw. genauer ausprägungsorientierte Clusteranalyse durchgeführt: Ausprägungen unterschiedlicher nominaler Variablen werden zu Typen/Clustern zusammengefaßt.

Wir wollen das Vorgehen anhand eines fiktiven Beispiels darstellen: Mit Hilfe der multiplen Korrespondenzanalyse soll der Zusammenhang zwischen der beruflichen Stellung (selbständig, Führungsposition, Arbeiter), dem Fahrstil (aggressiv, normal, zurückhaltend) und der Automarke (Mercedes, Porsche, VW) untersucht werden. Das Beispiel ist in ALMO als Kurz- und Maskenprogramm "PROG30m5.alm" enthalten.

Die multiple Korrespondenzanalyse erbringt zwei bedeutsamen Faktoren mit Eigenwerten größer 1. In der graphischen Darstellung lassen sich drei Typen unterscheiden: Selbständige, die einen Mercedes fahren und einen normalen Fahrstil zeigen (Typ 1), Personen in Führungspositionen, die mit einem Porsche aggressiv fahren (Typ 2) und Arbeiter mit einem zurückhaltenden Fahrstil mit einem VW (Typ3).

Faktorladungen



In dem Beispiel bereitet die graphische Bestimmung der Typen keine Probleme. Probleme können allerdings dann auftreten, wenn viele Objekte (Ausprägungen) analysiert und/oder mehr als zwei oder drei Dimensionen bedeutsam sind. In diesem Fall kann wie folgt vorgegangen werden:

- Es wird eine multiple Korrespondenzanalyse gerechnet. Die Zahl der bedeutsamen Dimensionen wird bestimmt.
- Die Faktorladungsmatrix wird zwischengespeichert. Sie enthält als Zeilen die Objekte (Ausprägungen der untersuchten nominalen Variablen) und als Spalten die Koordinatenwerte (Faktorladungen) der Objekte in den berechneten Dimensionen.
- Die Faktorladungsmatrix wird in einer anschließenden hierarchischen Clusteranalyse als Datenmatrix eingelesen. In die Clusteranalyse gehen die Koordinatenwerte (Faktorladungen) als Klassifikationsvariable und die Objekte (Ausprägungen) als Klassifikationsobjekte ein. Als Distanzmaß kann die quadrierte euklidische Distanz verwendet werden, da diese für die multiple Korrespondenzanalyse definiert ist.

Variablenorientierte Clusteranalyse und Korrespondenzanalyse.

Das hier dargestellte Vorgehen wird von ALMO technisch unterstützt. Beim Maskenprogramm P30m5.msk in Abschnitt P30.8.0 wird durch entsprechenden Eintrag in die 15. Optionsbox „Option: verschiedene Programm-Optionen“, Eingabefeld 5, die Faktorladungsmatrix gespeichert und ein hierarchisches Clusteranalyse-Programm erzeugt.

In der Ergebnisausgabe protokolliert ALMO den Dateinamen, unter dem die Faktorladungsmatrix abgespeichert wurde, das erzeugte Programm zur Clusteranalyse und gibt dessen Namen an.

```
                                                    ******************** MITTEILUNG
       Die Faktorladungsmatrix wird in Datei "c:\almo6\progs\faklad.fre"
                                                                 gespeichert

ALMO-Programm fuer eine anschliessende
Clusteranalyse:

VEREINBARE
  Variable=100;

ANFANG
Name1=:
  Porsche
 ,Mercedes
```

```
 ,VW
 ,Selbstän
 ,Arbeiter
 ,Führungs
 ,aggresiv
 ,normal
 ,zurückha;
Name2=XFaktor1;
Name3=XFaktor2;
ENDE

ANFANG
   Programm = 36;              # hierarchische Clusteranalyse #
   A_Quantitative_V = V2:3;
   Modell = Ward_Linkage;
   Distanzmass = quad_euklid;
   Objekte = 9;
   Min_Clusterzahl = 2;
   Max_Clusterzahl = 3;
Ende_Programmparameter

Lese V2:3 aus Datei 2 'c:\almo6\progs\ladung.mat' Format frei
   leerzu Ende;
Gehe_in_Programm
Gehe_zu Lese
ENDE
                                                   ******************** MITTEILUNG
                  Das ALMO-Clusteranalyse-Programm wurde unter dem Namen
                                                   "c:\almo6\progs\faklad.alm"
                                                                   gespeichert
         Wenn Sie auf den Dateinamen doppelklicken, dann wird es geladen
                                                Sie koennen es dann gleich rechnen
#####################################
```

Der Dateiname für das Clusteranalyseprogramm wird wie folgt gebildet: Wird die Datei in der Optionsbox (in die die Ladungsmatrix gespeichert wird) "Ladung.mat" genannt, dann speichert Almo das Clusteranalyse-Programm mit der Endung ".Alm", im Beispiel also mit dem Nmen "Ladung.Alm".

Die Faktorladungsmatrix sollte daher in keine Datei mit der Namenserweiterung "ALM" geschrieben werden.

Wird das Clusteranalyseprogramm durch einen Doppelklick geladen und anschließend gerechnet, so deuten alle Teststatistiken zur Bestimmung der Clusterzahl auf drei Cluster hin. Diese lassen sich auch in dem Dendrogramm erkennen. Die drei Cluster entsprechen den aufgrund der graphischen Darstellung bei der multiplen Korrespondenzanalyse ermittelten Typen.

```
Minimum=   0.0, Maximum=  12.9
Dendrogramm:

Führungs   6
aggresiv   7
Porsche    1
zurückha   9
Arbeiter   5
VW         3
normal     8
Selbstän   4
Mercedes   2
```

Wie bereits erwähnt, eignet sich das hier dargestellte Vorgehen, wenn sehr viele Objekte (Ausprägungen) analysiert werden und/oder mehr als zwei oder drei bedeutsame Dimensionen vorliegen. In dem Beispiel ist ein derartiges Vorgehen nicht erforderlich.

Neben den beiden bisher behandelten Vorgehensweisen einer graphischen Typenbestimmung und einer Typenbestimmung durch eine vorausgehende Korrespondenzanalyse mit anschließender Clusteranalyse auf der Basis der berechneten Faktorenladungsmatrix bestehen noch folgende Möglichkeiten der Typenbildung:

- Typenbildung durch eine Rotation der berechneten Dimensionen
- Typenbildung durch eine variablenorientierte Clusteranalyse
- Typenbildung durch eine objektorientierte Clusteranalyse

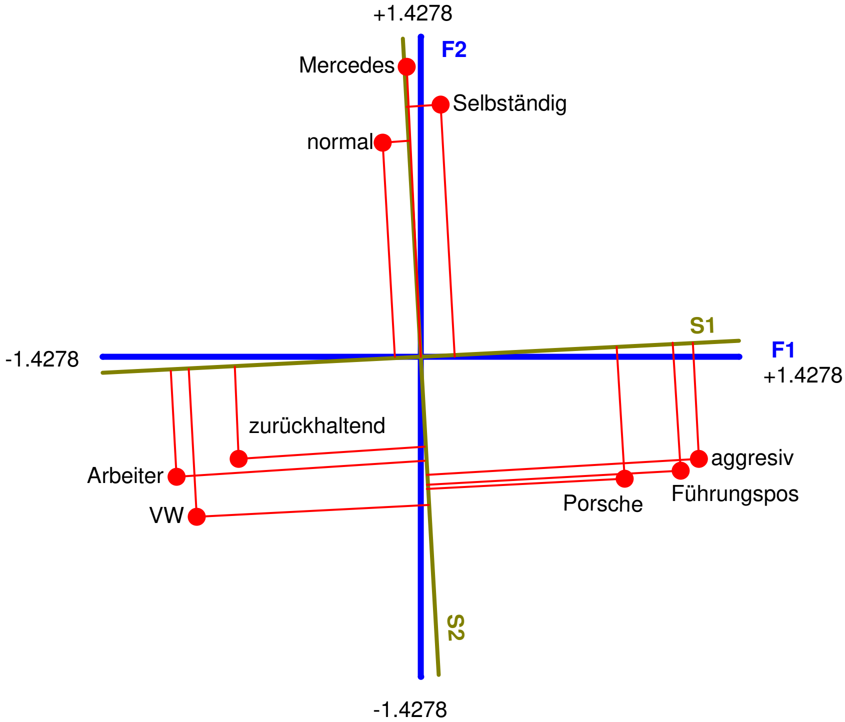
Der **Rotation der berechneten Dimensionen** liegt - wie bei der Faktorenanalyse allgemein - die Annahme einer Einfachstruktur zugrunde. Die Dimensionen (Faktoren) lassen sich so drehen, dass ein Objekt nur auf einer rotierten Achse/Dimension eine hohe Ladung besitzt und auf allen anderen eine geringe. Die Objekte, die eine Dimension bilden, können auf dieser eine positive oder negative Ladung aufweisen und entsprechend einen Typus und einen „Anti-Typus" bilden. Für unser Beispiel ergibt sich eine derartige Situation für die erste Dimension:

```
Aus Quartimin-Rotation:
Matrix der auf die schiefwinkligen Achsen
rechtwinklig projizierten Faktorladungen
(Strukturmatrix)
```

	Faktor 1	Faktor 2
Auto Porsche A1	1.1300	-0.5774
Auto Mercedes A2	0.0030	1.2999
Auto VW A3	-1.0386	-0.6622
Beruf Selbstän B1	0.1550	1.1219
Beruf Arbeiter B2	-1.1176	-0.4728
Beruf Führungs B3	0.8822	-0.5950
Fahrstil aggresiv V3-1	1.2201	-0.5288
Fahrstil normal V3-2	-0.1165	0.9638
Fahrstil zurückha V3-3	-0.8399	-0.4087

ALMO liefert folgende Grafik

Faktorladungen im recht- und
schiefwinkligen Koordinatensystem
(rechtwinklige Projektion)



Die Ausprägungen "Porsche", "Führungsposition" und "Porsche" besitzen auf der ersten rotierten Dimension positive Faktorladungen. Sie bilden einen Typus. Die Ausprägungen "VW", "Arbeiter" und "zurückhaltend" haben auf der ersten rotierten Dimension negative Faktorladungen. Sie bilden daher einen zweiten, einen Anti-Typus. Der Typus der Selbständigen mit einem normalen Fahrstil in einem Mercedes bilden den zweiten Faktor.

Eine Rotation der Dimensionen eignet sich somit ebenfalls zur Bestimmung von Typen, allerdings unter der Voraussetzung, dass eine Einfachstruktur vorliegt. Ist dies nicht der Fall, ist eine Rotation nur bedingt geeignet.

Bei einer variablenorientierten Clusteranalyse werden die Variablen geclustert. Das Vorgehen entspricht somit dem der multiplen Korrespondenzanalyse. Diese läßt sich als variablenorientiertes, unvollständiges Clusteranalyseverfahren betrachten, "unvollständig" deshalb, da nicht direkt Typen berechnet, sondern vom Anwender graphisch bestimmt werden (siehe dazu Bacher 1996). Eine direkte variablenorientierte Clusteranalyse für die Ausprägungen von nominalen Variablen ist nicht möglich, da empirisch die Ähnlichkeiten bzw. Unähnlichkeiten der Ausprägungen nicht bekannt sind (siehe dazu Bacher 1996). Nur eine variablenorientierte Clusteranalyse für die nominalen Variablen selbst ist möglich!

Objektorientierte Clusteranalyse.

Im Unterschied zu einer variablenorientierten Clusteranalyse werden bei einer objektorientierten Clusteranalyse die Untersuchungseinheiten, in unserem Beispiel die Personen, geclustert. Das Programm ist unter dem Namen "Prog36_Auto_Clust.Alm" in Almo enthalten. Es wird gefunden durch Klick auf das Menü "Almo / Liste aller Almo-Programme".

Für die Beispielsdaten ergeben sich drei Cluster, die im wesentlichen den drei (aus der Korrespondenzanalysedrei bereits bekannten) Typen entsprechen.

```
Masszahlen fuer Klassifikationsvariablen im Clustern 1:

Variable             n=      Min.      Max.        MA        SA    z-Wert
--------------------------------------------------------------------
  1 Auto
    1 Porsche        11      0.00      1.00      0.18      0.39     -1.16
    2 Mercedes       11      0.00      1.00      0.73      0.45      2.87
    3 VW             11      0.00      1.00      0.09      0.29     -2.88
  2 Beruf
    1 Selbstän       11      1.00      1.00      1.00      0.00     99.99
    2 Arbeiter       11      0.00      0.00      0.00      0.00      0.00
    3 Führungs       11      0.00      0.00      0.00      0.00      0.00
  3 Fahrstil
    1 aggresiv       11      0.00      1.00      0.27      0.45     -0.15
    2 normal         11      0.00      1.00      0.45      0.50      0.83
    3 zurückha       11      0.00      1.00      0.27      0.45     -0.78
------------------------------------------------------------------------

Masszahlen fuer Klassifikationsvariablen im Clustern 2:

Variable             n=      Min.      Max.        MA        SA    z-Wert
--------------------------------------------------------------------
  1 Auto
    1 Porsche        11      0.00      1.00      0.09      0.29     -2.56
    2 Mercedes       11      0.00      1.00      0.09      0.29     -2.56
    3 VW             11      0.00      1.00      0.82      0.39      3.81
  2 Beruf
    1 Selbstän       11      0.00      0.00      0.00      0.00      0.00
    2 Arbeiter       11      1.00      1.00      1.00      0.00     99.99
    3 Führungs       11      0.00      0.00      0.00      0.00      0.00
  3 Fahrstil
    1 aggresiv       11      0.00      0.00      0.00      0.00      0.00
    2 normal         11      0.00      1.00      0.36      0.48      0.26
    3 zurückha       11      0.00      1.00      0.64      0.48      1.67

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

Masszahlen fuer Klassifikationsvariablen im Clustern 3:

Variable             n=      Min.      Max.        MA        SA    z-Wert
--------------------------------------------------------------------
  1 Auto
    1 Porsche        12      0.00      1.00      0.67      0.47      2.41
    2 Mercedes       12      0.00      1.00      0.17      0.37     -1.40
    3 VW             12      0.00      1.00      0.17      0.37     -1.66
  2 Beruf
    1 Selbstän       12      0.00      0.00      0.00      0.00      0.00
    2 Arbeiter       12      0.00      0.00      0.00      0.00      0.00
    3 Führungs       12      1.00      1.00      1.00      0.00     99.99
  3 Fahrstil
    1 aggresiv       12      0.00      1.00      0.58      0.49      1.95
    2 normal         12      0.00      1.00      0.17      0.37     -1.40
    3 zurückha       12      0.00      1.00      0.25      0.43     -1.01

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
```

Die Typen sind allerdings nicht mehr so klar erkenntlich. Das Cluster 1 wird zwar ausschließlich von Selbständigen (MA=1.0) gebildet, es handelt sich dabei ferner überwiegend (zu 73%/MA=0.73) um Mercedes-Fahrer. Ein "normaler" Fahrstil tritt am häufigsten, allerdings nur mehr zu 45% (MA=0.45), auf. Dass die Typen nicht mehr so klar erkenntlich sind, liegt daran, dass es empirisch auch Mercedes-Fahrer mit einem aggressiven oder zurückhaltenden Fahrstil gibt, dass Selbständige eben auch einen Porsche oder einen VW fahren usw.

Fassen wir zusammen. Wir haben fünf Möglichkeiten der Typenbildung behandelt.

- Typenbildung auf der Basis der graphischen Darstellung der Ergebnisse der multiplen Korrespondenzanalyse. Dieses Vorgehen eignet sich nur bei einer geringen Objektzahl (Anzahl von Ausprägungen) und/oder bei zwei- oder drei bedeutsamen Dimensionen.
- Typenbildung durch eine vorausgehende multiple Korrespondenzanalyse mit einer anschließenden Clusteranalyse auf der Basis der Faktorladungsmatrix. Dieses Vorgehen eignet sich auch für eine größere Objektmenge und eine größere Zahl bedeutsamer Dimensionen.
- Typenbildung durch eine Rotation der berechneten Dimensionen. Dieses Vorgehen eignet sich ebenfalls für eine größere Objektmenge und eine größere Zahl bedeutsamer Dimensionen. Einfachstruktur der Daten wird allerdings vorausgesetzt.
- Typenbildung durch eine variablenorientierte Clusteranalyse der ursprünglichen Daten. Diese ist nicht möglich. Das hier dargestellte Vorgehen der Typenbildung durch eine vorausgehende multiple Korrespondenzanalyse mit einer anschließenden Clusteranalyse auf der Basis der Faktorladungsmatrix ist somit eine Möglichkeit, für die Ausprägungen von nominalen Variablen eine variablenorientierte Clusteranalyse durchzuführen.
- Typenbildung durch eine objektorientierte Clusteranalyse der ursprünglichen Daten. Dieses Vorgehen eignet sich für eine große Befragtenzahl und eine große Variablenzahl. Die Typen sind aber i.d.R. nicht mehr so deutlich erkennbar. Im Unterschied zu den vorausgehenden Verfahren ist allerdings nach der Analyse die Zuordnung der Personen zu den Typen bekannt. Bei den anderen Verfahren stellt sich dieses Problem erst. So z.B. ist zu entscheiden, welchen Typus ein Arbeiter als Mercedes-Fahrer mit einem aggressiven Fahrstil angehören soll. Anstelle der Ursprungsdaten können auch die "Faktorscores" der multiplen Korrespondenzanalyse verwendet werden.
-

Literatur

Arminger, Gerhard: Faktorenanalyse, Teubner, Stuttgart 1979

Mc Donald, R.P.: A generalized common factor analysis based on residual covariance matrices of prescribed structure, British Journal of mathematical and statistical Psychology, Vol 22, 1969a, S. 149-163

Mc Donald, R.P.: The common factor analysis of multicategory data, British Journal of mathematical and statistical Psychology, Vol 22, 1969b, S. 165-175

Bacher, J.: Clusteranalyse. 3. Auflage. Opladen, 2010, insbesonders Kap. 3

Blasius, J.:Korrespondenzanalyse, Oldenbourg Verlag, München, 2001

Carroll, D. **/Green**, P.F.: An Indscal-based approach to multiple correspondence analysis, Journal of Marketing Research, Vol XXV, May 1988, S.193-203

Greenacre M., **Blasius** J. Correspondence Analysis in the social sciences, Academic Press, Multivariate statistische Verfahren, de Gruyter, Berlin, New York 1984