



## **Nicht-metrische multidimensionale Skalierung Iterative MDS**

**P34**

**Johann Bacher / Kurt Holm / Heinrich Potuschak**

[www.almo-statistik.de](http://www.almo-statistik.de)  
[holm@almo-statistik.de](mailto:holm@almo-statistik.de)  
[johann.bacher@jku.at](mailto:johann.bacher@jku.at)  
[kurt.holm@jku.at](mailto:kurt.holm@jku.at)  
[heinrich.potuschak@linzag.net](mailto:heinrich.potuschak@linzag.net)

**2016**

## Weitere Almo-Dokumente

Die folgenden Dokumente können alle kostenlos von der Handbuchseite in

[www.almo-statistik.de](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
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)
14. Ereignisanalyse: Sterbetafel-Methode, Kaplan-Meier-Schätzer, Cox-Regression.PDF (1,5 MB)
15. Faktorenanalyse.PDF (1,6 MB)
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-System
27. Mehrfachantworten, Tabellierung von Fragen mit Mehrfachantworten (0.8 MB)
28. Metrische multidimensionale Skalierung (MDS) (0,4 MB)
29. Metrisches multidimensionales Unfolding (MDU) (0,6 MB)
30. Nicht-metrische multidimensionale Skalierung (MDS) (0,5 MB)
31. Pfadanalyse.PDF (0,7 MB)

# INHALTSVERZEICHNIS

<b>Nicht-metrische multidimensionale Skalierung (MDS)</b> .....	<b>4</b>
P34.0 Einführung .....	4
P34.0.1 Ein Beispiel .....	4
P34.0.2 Ein 2. Beispiel .....	5
P34.0.3 Terminologie .....	7
P34.0.4 Ordinale und quantitative Unähnlichkeiten.....	7
P34.0.5 Unvollständige Ähnlichkeits- bzw. Unähnlichkeitsmatrix .....	8
P34.1 Kalkül der nicht-metrischen MDS .....	10
P34.1.1 Der Stress-Koeffizient.....	10
P34.1.2 Die Berechnung der Disparitäten bei ordinalen Unähnlichkeiten.....	11
P34.1.3 Die Berechnung der Disparitäten bei quantitativen Unähnlichkeiten.....	12
P34.1.4 Das Gradienten-Verfahren .....	13
P34.1.5 Konkurrierende iterative MDS-Verfahren .....	15
P34.2 Dateneingabe.....	15
P34.2.1 Erstellen einer Un- oder Ähnlichkeitsmatrix.....	15
P34.2.2 Eingabe in Maskenprogramm Prog34ma .....	17
P34.2.3 Eingabe in Maskenprogramm Prog34mb.....	19
P34.2.4 Erläuterung zu den Eingabeboxen beider Programme .....	21
P34.2.5 Erläuterung zu den Optionsboxen von Prog34mb .....	26
P34.2.5.1 Optionsbox: Idealpunkte .....	27
P34.2.5.2 Optionsbox: Startwerte .....	28
P34.2.5.3 Optionsbox: Weitere Optionen .....	30
P34.2.5.4 Optionsbox: Rotation .....	33
P34.2.5.5 Optionsbox: "Aussehen" der auszugebenden Tabelle, Grafik-Optionen.....	34
P34.3 Ausgabe: Ergebnisse aus Prog34ma und Prog34mb.....	34
P34.3.1 Analyse mit 1 Dimension mit Prog34ma .....	34
P34.3.2 Analyse mit 2 Dimensionen mit Prog34mb .....	35
P34.3.3 Analyse mit 3 Dimensionen mit Prog34mb .....	39
P34.3.4 Welche Dimensionszahl ist die richtige ? .....	41
P34.3.5 Was tun wenn die vorgegebene Stress-Schwelle nicht erreicht wird.....	42
P34.3.6 Erweiterte Ausgabe bei Prog34mb.....	43
P34.3.6.1 Berechnungsschema, reproduzierte Distanzmatrix und Disparitätenmatrix. ....	44
P34.3.6.2 Streudiagramme der Unähnlichkeiten, Disparitäten und Distanzen .....	46
P34.3.6.3 Idealpunkte.....	49
P34.3.6.4 Verhältnis zum multidimensionalen Unfolding (MDU).....	52
P34.3.6.5 Vektorrepräsentation der Idealpunkte.....	52
P34.3.6.6 Projektion externer Information als Interpretationshilfe.....	53
P34.4 Vergleich Almo mit SPSS-Alscal und Proxscal .....	57
Exkurs: Vergleich metrische MDS und iterative MDS.....	59
Literatur.....	72

# Nicht-metrische multidimensionale Skalierung (MDS)

## P34.0 Einführung

Das Programm wurde von Heinrich Potuschak ursprünglich in Fortran geschrieben (siehe Potuschak in Sixtl, 1982, S.465 ff). Johann Bacher hat es an das ALMO-System adaptiert, durch wesentliche Programmteile erweitert und verschiedene Textteile dieses Almo-Dokuments verfasst. Kurt Holm hat das Programm auf quantitative (intervallskalierte) Daten erweitert, einige Programmteile ergänzt, die Grafikprogrammierung und die Programm-Masken hinzugefügt und mehrere Textabschnitte verfasst.

Siehe auch die beiden Almo-Dokumente:

- 28. Metrische multidimensionale Skalierung (MDS)
- 29. Metrisches multidimensionales Unfolding (MDU)

In Almo sind 4 Programm-Masken zur nicht-metrischen MDS enthalten

1. Prog34ma. Das Programm erfordert nur eine minimale Eingabe und liefert kurze Ergebnisse.
2. Prog34mb. Das Programm enthält viele Optionen, so die Möglichkeit *Startwerte* einzugeben und *Idealpunkte* einzufügen.
3. Prog34mc. Das Programm liest eine Datei von ordinalen Paarvergleichen ein, formt aus ihnen eine Ähnlichkeitsmatrix und rechnet dann, wie Prog34mb eine MDS (mit Optionen).
4. Prog34md. Das Programm rechnet zuerst eine hierarchische Clusteranalyse. Erzeugt dabei eine Distanzmatrix, die an die nicht-metrische MDS übergeben und gerechnet wird.

### P34.0.1 Ein Beispiel

Die Ähnlichkeiten bzw. Unähnlichkeiten zwischen sieben politischen Parteien sollen untersucht werden. Lassen sich diese durch eine Dimension (z.B. Links-Rechts-Dimension) oder durch mehrere Dimensionen erklären. Die Unähnlichkeitsmatrix ist folgende:

	Komm	Sozia	Arbei	Liber	Zentr	Chris	Konse
Kommunist. Par	0						
Sozialist. Par	8.7	0					
Arbeiterpartei	25.3	14.8	0				
Liberale	33.7	19.0	10.0	0			
Zentrumspartei	37.9	33.2	17.8	10.5	0		
Christliche P.	49.3	50.5	21.3	18.9	7.6	0	
Konservative	50.2	40.0	24.3	12.9	8.1	7.3	0

Beispiel aus: Fahrmeir/Hammerle, 1984: Multivariate statistische Verfahren. Berlin-New York, S. 678

**Wichtige Anmerkung:** Dieses Beispiel mit den 7 politischen Parteien wird im nachfolgenden Text immer wieder aufgegriffen. Im Almo-Programm wurde es dabei mit Zufallswerten als Startwerten und 2 Faktoren gerechnet. In Almo wird jedoch standardmäßig die Ladungsmatrix aus der metrischen MDS als Startwert verwendet. Dies ist die Voreinstellung. Dabei entsteht eine Lösung mit nur einem Faktor (Dimension), was vermutlich ohnehin die optimale Lösung ist. Siehe dazu auch Abschnitt P34.2.2, "Welche Dimensionszahl ist die richtige?". In den Programm-Masken Prog34mb,c,d kann in der Optionsbox "Startwerte" auf *Zufallswerte* umgestellt

werden, so dass das Parteien-Beispiel (so wie es hier vorgetragen wird) nachgerechnet werden kann.

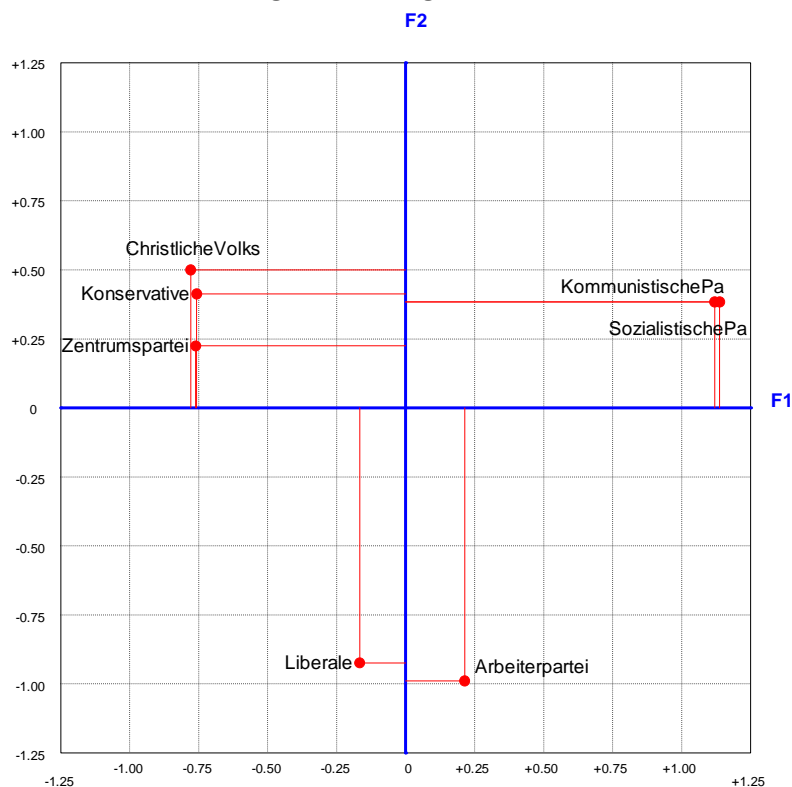
Die Eingabedaten für eine nicht-metrische MDS bestehen aus einer Unähnlichkeits-Matrix. Je kleiner ein Zahlenwert in obiger Matrix ist, umso näher sind sich die beiden Objekte. Ist die vorliegende Matrix eine Ähnlichkeitsmatrix, dann wird sie (Almo-intern) in eine Unähnlichkeitsmatrix gewandelt. Die nicht-metrische MDS zerlegt nun die Unähnlichkeitsmatrix in eine Ladungsmatrix. Folgendes Ergebnis entsteht:

Ladungsmatrix der Objekte

	Faktor 1	Faktor 2
Kommunistische Part.	1.1194	0.3844
Sozialistische Part.	1.1379	0.3844
Arbeiterpartei	0.2160	-0.9899
Liberale	-0.1689	-0.9243
Zentrumspartei	-0.7640	0.2273
Christliche Partei	-0.7813	0.5029
Konservative	-0.7591	0.4153

Stress 0.037

Diese Ladungsmatrix wird dann noch grafisch als 2-dimensionales Koordinatensystem in Form eines Punktediagramms abgebildet



Die Distanzen zwischen den Objekten werden somit grafisch sichtbar gemacht. Bei der nichtmetrischen MDS wird also versucht, die Objekte einer Unähnlichkeitsmatrix in einem r-dimensionalen Raum darzustellen, wobei man bestrebt ist, r klein zu halten (z.B. r=2 oder 3). Die Güte der Darstellung wird durch den *Stress-Koeffizienten* gemessen.

### P34.0.2 Ein 2. Beispiel

Die Distanzen in der Wahrnehmung verschiedener Automarken sind folgende

	Opel	Volkwag	Suzuki	Toyota	Mercede	BMW	Ferrari	Porsche	Lamborg	RollsRo
Opel	0	3.1	5.0	3.8	5.9	5.5	8.4	8.4	8.5	8.5
Volkwage	3.1	0	4.4	3.3	5.8	5.5	8.1	8.1	8.2	8.6
Suzuki	5.0	4.4	0	3.7	7.0	7.0	8.3	8.4	8.8	8.9
Toyota	3.8	3.3	3.7	0	5.3	4.2	8.3	8.3	8.7	8.2
Mercedes	5.9	5.8	7.0	5.3	0	2.7	6.9	6.4	6.6	5.8
BMW	5.5	5.5	7.0	4.2	2.7	0	6.8	6.4	6.4	7.0
Ferrari	8.4	8.1	8.3	8.3	6.9	6.8	0	3.0	0	6.6
Porsche	8.4	8.1	8.4	8.3	6.4	6.4	3.0	0	3.4	6.8
Lamborgh	8.5	8.2	8.8	8.7	6.6	6.4	0	3.4	0	6.3
RollsRoy	8.5	8.6	8.9	8.2	5.8	7.0	6.6	6.8	6.3	0

Daten aus Internetpaper "Multidimensionale Skalierung", BUGH Wuppertal, Lehrstuhl für empirische Wirtschafts- und Sozial-forschung, Fachbereich Wirtschaftswissenschaft, [http://www2.informatik.uni-osnabrueck.de/marc/lectures/zra\\_ss03/prgdat/mds.pdf](http://www2.informatik.uni-osnabrueck.de/marc/lectures/zra_ss03/prgdat/mds.pdf)

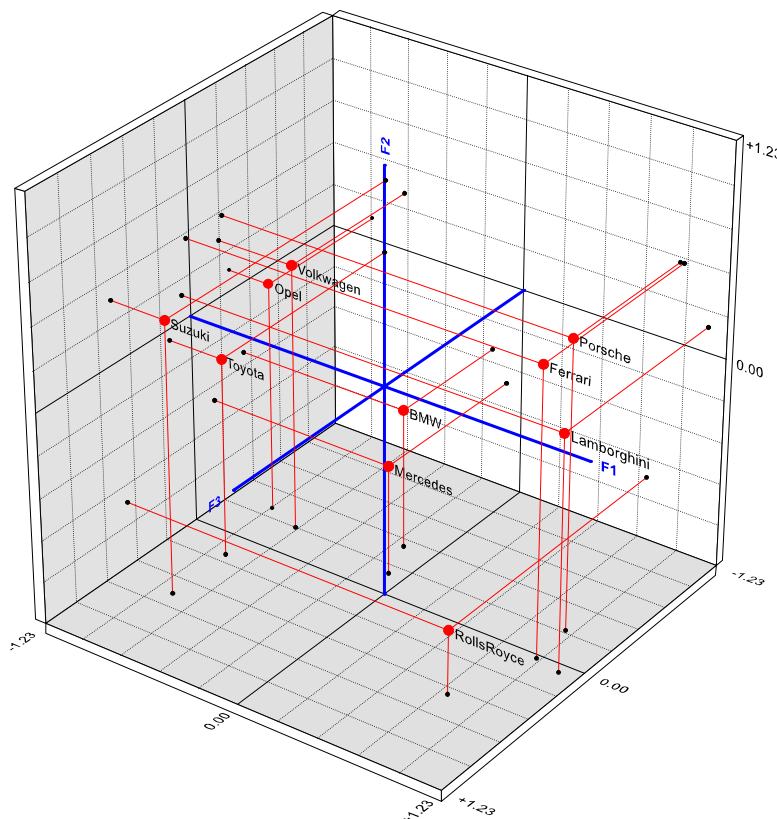
Das Programm wird unter dem Namen "Automarken2.Almo" gefunden durch Klick auf den Knopf "alle Progs" in der Knopfleiste am Oberrand des Almo-Fensters.

Wir fordern dieses mal 3 Dimensionen an. Almo liefert folgende Matrix der Ladungen auf die 3 Faktoren

	Faktor 1	Faktor 2	Faktor 3
Opel	-0.9740	0.1227	-0.3262
Volkwa	-0.7617	0.3356	-0.2445
Suzuki	-0.8855	0.3738	0.6303
Toyota	-0.8976	-0.0607	0.1752
Merced	-0.0972	-0.5884	-0.1769
BMW	-0.1893	-0.4102	-0.4326
Ferrar	0.9617	0.4643	0.0275
Porsch	0.9383	0.4648	-0.2703
Lambor	1.1190	0.1507	0.0676
RollsR	0.7864	-0.8525	0.5499

Stress-1 0.0229

Die Matrix wird von Almo dann noch als 3-dimensionales Punktediagramm dargestellt



Es sind 3 Cluster erkennbar:

1. Die "Sportwagen" Porsche, Ferrari, Lamborghini
2. Die "Alltagsautos" Volkswagen, Opel, Toyota, Suzuki (wobei Suzuki etwas auf Distanz zu den anderen 3 geht)
3. Die "Edelkarossen" Mercedes und BMW
4. Der "Einzelgänger" Rolls Royce

Im Almo-Grafik-Editor kann diese Grafik in vielfältiger Weise bearbeitet werden. So werden wir später diese Grafik in der Achse F1 spiegeln.

### P34.0.3 Terminologie

In der Terminologie der nicht-metrischen MDS wird zwischen "Unähnlichkeit", "Distanz" und "Disparität" unterschieden. Wir definieren vorläufig:

#### *Unähnlichkeit / Ähnlichkeit*

Das ist das was empirisch erhoben wurde. Entsprechend ist dann die Unähnlichkeitsmatrix die empirisch erhobene Datenmatrix. Ist die vorhandene Datenmatrix eine *Ähnlichkeitsmatrix* so stellt das kein Problem dar. Im Almo-Programm wird sie automatisch durch Multiplikation mit -1 (und Addition des absolut gesetzten kleinsten Wertes) zur Unähnlichkeitsmatrix gewandelt. Die empirischen Unähnlichkeiten werden im folgenden mit "u" symbolisiert

#### *Distanz*

Das ist die Entfernung zwischen zwei Punkten (Objekten) im von der MDS konstruierten Raum. Entsprechend ist die Distanzmatrix die Matrix der Entfernungen zwischen den Punkten im "MDS-Raum". Die *Distanz* wird aus den Koordinatenwerten der (anfänglich zufällig gewählten) "Startwerte"-Matrix errechnet bzw. im Verlauf des iterativen Algorithmus aus den jeweils entstandenen Koordinatenwerten. Die Distanzen werden durch "d" symbolisiert.

#### *Disparität*

Sie ist gleich der Distanz, die jedoch so verändert ist, dass ihre Rangordnung mit derjenigen der empirischen Unähnlichkeit übereinstimmt. Die Disparität wird im folgenden deswegen auch "rangkonforme Distanz" genannt. Dieser Begriff wird durch das nachfolgende Rechenbeispiel in den Abschnitten P34.0.2.1 und P34.0.2.2 deutlich werden. Die Disparitäten werden durch " $\hat{d}$ " symbolisiert. Da ihre Rangordnung derjenigen der empirischen Unähnlichkeiten entspricht, werden sie auch als "monotone Transformationen der empirischen Unähnlichkeiten" bezeichnet.

In der englischsprachigen Literatur wird von "dissimilarity", "distance" und "disparity" gesprochen.

### P34.0.4 Ordinale und quantitative Unähnlichkeiten.

Die nichtmetrische MDS wurde ursprünglich dafür entwickelt ordinal gemessene Unähnlichkeiten zwischen Objekten dimensional darzustellen. Sehr schnell wurde jedoch erkannt, dass der spezielle iterative Kalkül der MDS auch auf quantitativ gemessene Unähnlichkeiten anwendbar ist. Lediglich die Disparitäten werden anders berechnet. Sie sind "lineare Transformationen der empirischen Unähnlichkeiten" und nicht, wie bei ordinal gemessenen Unähnlichkeiten, "monotone Transformationen". Im folgenden wird das noch ausgeführt werden.

Die Bezeichnung *nicht-metrische* MDS für das in Almo enthaltene Programm ist also nicht korrekt. Sinnvoller wäre es, von *iterativer* MDS oder *Stress-optimierender* MDS zu sprechen. Damit würde dann auch der entscheidende Unterschied zur *metrischen* "faktoranalytischen"

MDS betont (die im Maskenprogramm Prog30ma realisiert und im Almo-Dokument Nr. 28 beschrieben ist).

Wir werden trotzdem den Begriff "nicht-metrische MDS" beibehalten. Aus zwei Gründen: Weil wir in früheren Almo-Versionen und Dokumenten diesen Begriff verwendet haben und - das ist der Hauptgrund - weil wir raten, bei quantitativen Unähnlichkeiten besser mit der metrischen MDS zu arbeiten und Prog30ma zu verwenden. Im Almo-Dokument Nr. 28 wird das ausführlich begründet.

Wir rechnen das obige 2. Beispiel "Automarken2.Alm" mit 3 Dimensionen, wobei wir in der Eingabebox für die einzulesende Matrix als Messniveau "quantitativ (intervall)" angeben. Es entsteht folgendes Ergebnis:

Ladungsmatrix der Objekte

		Faktor 1	Faktor 2	Faktor 3
Opel	V1	-1.0026	0.1402	-0.1998
Volkwa	V2	-0.7926	0.4207	-0.1517
Suzuki	V3	-0.8122	0.4451	0.6498
Toyota	V4	-0.8512	-0.2021	0.1053
Merced	V5	-0.0235	-0.6599	-0.2793
BMW	V6	-0.1496	-0.4021	-0.5387
Ferrar	V7	1.0235	0.3310	0.0651
Porsch	V8	0.9136	0.3944	-0.4093
Lambor	V9	1.0175	0.3377	0.0633
RollsR	V10	0.6770	-0.8050	0.6953

Stress-1 0.070211

Der Stress-1-Koeffizient ist nicht so gut wie der aus der ordinalen Analyse. Wenn wir mit Prog30ma eine metrische MDS rechnen (mit dem Beispielprogramm "Automarken2.Alm") dann erhalten wir einen noch schlechteren Stress-Koeffizienten. Allgemein gilt: Je höher die Anforderungen an das Messniveau umso schlechter der Stress-Koeffizient.

Die grafische Darstellung der Ladungsmatrix ist sehr ähnlich zum oben bereits abgebildeten Diagramm, so dass

### P34.0.5 Unvollständige Ähnlichkeits- bzw. Unähnlichkeitsmatrix

Der im Kruskalverfahren verwendete Algorithmus zur Bestimmung der Punktekonfiguration kann auch für unvollständige Unähnlichkeitsmatrizen verwendet werden. Das heißt, dass für einige Zellen der Unähnlichkeitsmatrix fehlende Werte (=KEIN\_WERT) vorliegen können. Dies ist jedoch in Prog34mb nicht zulässig, wenn als Startwerte die Ladungsmatrix aus der metrischen MDS eingesetzt wird.

Eine Konstellation mit KW-Fällen kann beispielsweise entstehen, wenn zur Gewinnung der Ähnlichkeitsmatrix die Methode des Paarvergleichs eingesetzt wird. Den Befragten werden nicht alle möglichen Objektpaare zur Ähnlichkeitsbeurteilung vorgelegt, sondern nur eine Auswahl.

Wir wollen die Analyse unvollständiger Unähnlichkeitsmatrizen für das obige Beispiel 1 der Unähnlichkeit von Parteien darstellen. Dazu werden in der Unähnlichkeitsmatrix einige Zellen auf "KEIN\_WERT" gesetzt. Die Matrix ist im Almo-Unterordner Testdat enthalten.

```
0.0
 8.7  0.0
25.3 14.8  0.0
33.7 19.0 10.0  0.0
  kw 33.2 17.8 10.5  0.0
49.3 50.5 21.3  kw  7.6  0.0
50.2 40.0 24.3 12.9  8.1  kw  0.0
```



Die in Almo verwendete Kein-Wert-Kodierung ist

kw oder KW

D.h. diese Bezeichnung in den Daten wird von Almo als fehlender Wert verstanden und entsprechend behandelt

Bei Zufallswerten als Startwerten und einem Schwellenwert von 0.05 für den Stress wird für die euklidischen Distanzen im zweidimensionalen Raum eine konvergente Lösung gefunden:

		Faktor 1	Faktor 2
Kommun	V1	1.4561	0.1144
Sozial	V2	0.9619	0.7880
Arbeit	V3	0.1691	0.5214
Libera	V4	-0.3218	0.0153
Zentru	V5	-0.2441	-0.7549
Christ	V6	-1.0044	-0.4130
Konser	V7	-1.0168	-0.2712

Stress-1    0.047436

Diese Lösung stimmt hinsichtlich der ersten Dimension relativ gut mit der Lösung für die vollständige Unähnlichkeitsmatrix überein.

Gesuchte Punktekonfiguration (vollständige Unähnlichkeitsmatrix)

		Faktor 1	Faktor 2
Kommunis	V1	1.1194	0.3843
Sozialis	V2	1.1379	0.3843
Arbeiter	V3	0.2159	-0.9898
Liberale	V4	-0.1689	-0.9243
Zentrums	V5	-0.7640	0.2272
Christli	V6	-0.7812	0.5029
Konserva	V7	-0.7590	0.4152
Stress		0.037	
Iterationen		13	
Startwert fuer ZZ.Generator=		123123	

Für die zweite Dimension ergeben sich dagegen deutliche Unterschiede. Dies ist u.a. darauf zurückzuführen, dass die Ergebnisse für zwei Dimensionen instabil sein können, da hier ein geringes Verhältnis zwischen der Objektzahl und Dimensionszahl vorliegt. Allgemein sollte folgendes Verhältnis zwischen der Zahl der Objekte und der maximalen Dimensionszahl bestehen (Sixtl 1982):

vollständige Unähnlichkeitsmatrix:    5 zu 1.

unvollständige Unähnlichkeitsmatrix:    7 zu 1.

Bei vollständigen und auch bei unvollständigen Unähnlichkeitsmatrizen sollten also bei 7 Objekten nur 1 Dimension gerechnet werden.

Wird nur 1 Dimension angefordert dann sind die Ergebnisse ausserordentlich ähnlich

		vollständige Daten	unvollständige Daten
Kommun	V1	-1.5746	-1.6052
Sozial	V2	-1.2380	-1.1612

Arbeit	V3	-0.2605	-0.3041
Libera	V4	0.1935	0.2009
Zentru	V5	0.6810	0.5960
Christ	V6	1.1350	1.1720
Konser	V7	1.0635	1.1016
Stress-1		0.0433	0.0959

### P34.1 Kalkül der nicht-metrischen MDS

Vorbemerkung: Eine einfache, "intuitive" Darstellung des Algorithmus der nicht-metrischen MDS bei ordinalen Unähnlichkeiten ist zu finden bei Jacoby; eine ausführliche und detaillierte Darstellung bei Bacher, Pöge, Welzig (2010), sowie bei Borg, Groenen (2005, S. 206ff) und besonders übersichtlich im Internetpaper der **BUGH** Wuppertal, Abschnitt 3.1 (siehe Literaturliste).

Das von KRUSKAL (1964) entwickelte Analyseverfahren setzt nur *ordinale* Information bezüglich der Ähnlichkeit bzw. Unähnlichkeit von Objekten voraus (=nichtmetrische MDS). Dabei wird nach einer Darstellung der Objekte in einem r-dimensionalen Raum gesucht, so dass zwischen den geschätzten Distanzen der Objekte im r-dimensionalen Raum und ihren empirischen Unähnlichkeiten eine monotone Beziehung besteht. Sind z.B.  $d_p(A,B)$  und  $d_p(C,D)$  die Distanzen zwischen den Objekten A und B bzw. zwischen C und D im r-dimensionalen Raum, dann soll für diese folgende Beziehung gelten:

$$u(A,B) \leq u(C,D) \Leftrightarrow d_p(A,B) \leq d_p(C,D)$$

$u(A,B)$  und  $u(C,D)$  sind die erhobenen Unähnlichkeiten zwischen den Objekten.

$d_p(A,B)$  ist die Distanz zwischen A und B, die in der allgemeinen Potenz- oder Minkowskimetrik mit Metrikparameter p gemessen wird,

$$(1) \quad d_p(A,B) = \left[ \sum_{k=1}^r |x_{Ak} - x_{Bk}|^p \right]^{\frac{1}{p}},$$

wobei  $x_{Ak}$  der Skalen- bzw. Koordinatenwert des Objekts A auf der Dimension k ist.  $x_{Bk}$  ist entsprechend der Skalenwert des Objekts B. Der gewohnte euklidische Raum besitzt den Metrikparameter  $p=2$ . Andere Metrikparameter, mit Ausnahme des Metrikparameters  $p=1$ , der sogenannten City-Block-Metrik, sind eher ungewöhnlich.

#### P34.1.1 Der Stress-Koeffizient

Ein Maß für die Übereinstimmung zwischen empirischer und durch die r-dimensionale Darstellung reproduzierter Ordnungsrelation der Unähnlichkeit zwischen Objektpaaren ist der sogenannte Stress, der zwischen null (perfekte Modellanpassung) und eins (überhaupt keine Modellanpassung) variieren kann.

Der Stress ist definiert als (Sixtl 1982: 349)

$$(2) \quad S = \left[ \frac{\sum_{A < B} (d_p(A,B) - \hat{d}_p(A,B))^2}{\sum d_p(A,B)^2} \right]^{\frac{1}{2}} = \left[ \frac{R}{T} \right]^{\frac{1}{2}},$$

wobei  $d_p(A,B)$  die (mit Metrikparameter p) berechneten *Distanzen* sind.  $\hat{d}_p(A,B)$  sind die *Disparitäten*. Das sind die transformierte Distanzen, deren Rangordnung mit den empirischen

Unähnlichkeiten übereinstimmen. Wir nennen sie deswegen auch modell- oder rangkonforme Distanzen. Da es verschiedene Stress-Koeffizienten gibt (siehe etwa Busing 2010) wird obiges Stress-Maß in der Literatur "Stress-1" genannt.

Die Berechnung des Stress soll am obigen Beispiel aufgezeigt werden. Das nicht-metrische MDS nach dem Kalkül von Kruskal ist ein iteratives Verfahren. Nach mehreren Iterationen wird schließlich eine Lösung gefunden, die einen vom Benutzer vorgegebenen Stress-1-Wert erreicht bzw. unterschreitet.

Zuerst wird eine Ladungsmatrix (z.B. aus Zufallswerten) festgelegt. Aus ihr werden gemäß obiger Gleichung 1 die ersten Distanzen  $d_p(A,B)$  ermittelt. Im ersten Iterationsschritt entstehen dann folgende Ergebnisse

A	B	$u(A,B)$	$d_p(A,B)$	$\hat{d}_p(A,B)$	$(d_p(A,B) - \hat{d}_p(A,B))^2$	$d_p(A,B)^2$
2	6	50,5	1,309	1,49	0,032	1,714
1	7	50,2	1,146	1,49	0,117	1,314
1	6	49,3	2,011	1,49	0,272	4,043
2	7	40,0	0,373	1,48	1,232	0,139
1	5	37,9	1,493	1,48	0,000	2,230
1	4	33,7	1,710	1,48	0,051	2,923
2	5	33,2	0,784	1,48	0,488	0,615
1	3	25,3	0,205	1,48	1,633	0,042
3	7	24,3	1,351	1,48	0,017	1,826
3	6	21,3	2,145	1,48	0,439	4,603
2	4	19,0	2,235	1,48	0,565	4,994
4	6	18,9	1,635	1,48	0,023	2,675
3	5	17,8	1,651	1,48	0,028	2,726
2	3	14,8	1,591	1,48	0,012	2,531
4	7	12,9	2,286	1,48	0,645	5,225
4	5	10,5	1,622	1,48	0,019	2,630
3	4	10,0	1,678	1,48	0,038	2,814
1	2	8,7	1,389	1,39	0,000	1,930
5	7	8,1	1,044	1,08	0,001	1,090
5	6	7,6	0,585	1,08	0,245	0,342
6	7	7,3	1,611	1,08	0,282	2,596

A = Zeile in empirischer Unähnlichkeitsmatrix (oberes Dreieck)

B = Spalte

$u(A,B)$  = Unähnlichkeitswert in Zeile A Spalte B

$d_p(A,B)$  = Distanz, berechnet aus Koordinatenwerten der Startwertematrix bzw. der iterativ gewonnenen Ladungsmatrizen

$\hat{d}_p(A,B)$  = Disparitäten (oder rangkonforme Distanzen)

$(d_p(A,B) - \hat{d}_p(A,B))^2$  = Zähler der Stress-Formel

$d_p(A,B)^2$  = Nenner der Stress-Formel

### P34.1.2 Die Berechnung der Disparitäten bei ordinalen Unähnlichkeiten

Die empirischen Unähnlichkeiten sind  $u(A,B)$ . Die aus der Startwerte-Matrix bzw. aus den iterativ gewonnenen Koordinatenwerten berechneten Distanzen sind  $d_p(A,B)$ , die Disparitäten (die rangkonformen Distanzen) sind  $\hat{d}_p(A,B)$ . Die Objektpaare sind

entsprechend den Werten ihrer Unähnlichkeiten der Größe nach absteigend geordnet. Es wäre problemlos auch möglich die Objektpaare aufsteigend zu ordnen. So geschieht es im Almo-Programm. Ein Vergleich der Größenordnung der berechneten Distanzen mit den empirischen Unähnlichkeiten zeigt, dass für die ersten drei Objektpaare die Ordnungsrelationen nicht übereinstimmen. Eine Übereinstimmung wird erreicht, wenn die Werte der ersten drei Objektpaare gemittelt werden. Die Disparitäten (die rangkonformen Distanzen) sind daher gleich

$$\hat{d}_p(A,B) = \frac{1.309 + 1.146 + 2.011}{3} = 1.49$$

Diese nennen wir in der Ergebnisliste von Almo "Disparitäten".

Auch bei den nächsten 14 Objektpaaren - vom Paar (2, 7) bis zum Paar (3, 4) – stimmt die Anordnung der berechneten Distanzen mit den empirischen Unähnlichkeiten nicht überein. Die Disparitäten sind daher

$$\hat{d}_p(A,B) = \frac{0,373 + \dots + 1,678}{14} = 1.48$$

und ergeben sich als Durchschnitt der 14 berechneten Distanzen. Schließlich müssen noch die letzten drei Objektpaare gemittelt werden, um rangkonforme Distanzen zu erreichen. Für die Größenanordnung der rangkonformen Distanzen gilt nun die geforderte Übereinstimmung mit der Anordnung der empirischen Unähnlichkeiten:

$$u_p(A,B) \leq u_p(C,D) \Leftrightarrow \hat{d}_p(A,B) \leq \hat{d}_p(C,D).$$

In der Spalte  $(d_p(A,B) - \hat{d}_p(A,B))^2$  wird die Übereinstimmung zwischen berechneten und rangkonformen Distanzen ermittelt, deren Summe in den Stress als Zähler einfließt. Für die Spalten  $(d_p(A,B) - \hat{d}_p(A,B))^2$  und  $d_p(A,B)^2$  ergeben sich nachfolgende Summen:

$$R = \sum (d_p(A,B) - \hat{d}_p(A,B))^2 = 6,142$$

$$T = \sum d_p(A,B)^2$$

Der Stress nimmt daher ein Wert von  $STRESS = \sqrt{\frac{R}{T}} = 0,354$

Die *Disparitäten* sind nicht nur rangkonform transformierte Distanzen. Da ihre Rangordnung derjenigen der empirischen Unähnlichkeiten entspricht, sind sie auch monoton transformierte Unähnlichkeiten. Deswegen gehen auch sie in die Stress-Berechnung ein. Der Stress-Koeffizient drückt die Übereinstimmung von empirischen Daten mit den vom Modell reproduzierten Daten aus. Im ordinalen Modell der nichtmetrischen MDS sind die Disparitäten legitime monotone Transformationen der empirischen Unähnlichkeiten. Sie sind sozusagen deren legitime "Stellvertreter".

### **P34.1.3 Die Berechnung der Disparitäten bei quantitativen Unähnlichkeiten**

Der MDS-Kalkül nach Kruskal ist ursprünglich auf den Fall beschränkt gewesen, dass die Unähnlichkeiten zwischen den Objekten ordinal gemessen wurden. Er kann aber relativ einfach auch auf den Fall ausgedehnt werden, dass die Unähnlichkeiten quantitativ gemessen wurden. Lediglich die Disparitäten sind dann anders zu berechnen.

Bei jedem Iterationsschritt wird eine Regressionsanalyse gerechnet. Die Unähnlichkeiten bilden dabei die unabhängige Variable und die errechneten Distanzen  $d$  die abhängige. Mit dem daraus gefundenen Regressionskoeffizienten und der Konstanten wird dann für jede Unähnlichkeit ein Wert prognostiziert. Dies ist dann die Disparität  $\hat{d}$ .

Die Vorgehensweise ist folgende:

Aus der Ladungsmatrix der Startwerte werden die Distanzen  $d$  errechnet. Dann wird eine Regressionsanalyse mit den empirischen Unähnlichkeiten  $u$  als unabhängige Variable und den Distanzen  $d$  als abhängige Variable gerechnet. Dabei entsteht der Regressionskoeffizient  $b_1$  und die Konstante  $k_1$ . Die Ziffer 1 soll ausdrücken, dass diese Koeffizienten aus dem Iterationsschritt 1 hervorgegangen sind. Die Disparitäten  $\hat{d}_1$  entstehen als "Prognosewerte" gemäß  $\hat{d}_1 = b_1 * u + k_1$ . Damit ist  $\hat{d}_1$  eine lineare Transformation der Unähnlichkeiten  $u$ .

Dann werden wie beim ordinalen Rechengang der Stress berechnet und nach dem (im folgenden beschriebenen) Gradientenverfahren die Ladungsmatrix entsprechend (geringfügig) verändert. Danach beginnt der 2. Iterationsschritt: Aus der neuen Ladungsmatrix werden die Distanzen  $d_2$  berechnet, dann eine Regressionsanalyse gerechnet, die  $b_2$  und  $k_2$  liefert und die neuen Disparitäten aus  $\hat{d}_2 = b_2 * u + k_2$  ermittelt usw. Dieses iterative Verfahren wird solange fortgesetzt bis die vorgegebene Stress-Schwelle unterschritten wird bzw. eine Abbruchbedingung erfüllt ist.

### P34.1.4 Das Gradienten-Verfahren

Wenn der Stress-Wert unbefriedigend hoch ist – als „sehr gut“ gelten Stresswerte kleiner 0,05 (siehe P34.2) – werden neue Koordinaten (Skalenwerte) der Objekte berechnet. Dafür wird das sogenannte Gradientenverfahren eingesetzt (Sixtl 1982: 341-347). In einem Iterationsschritt werden die Koordinaten der Objekte so verschoben, dass der Stress minimiert wird.

Technisch ausgedrückt wird für jeden Koordinatenwert  $x_{ik}$  des Objekts  $i$  in der Dimension  $k$ , die erste partielle Ableitung des Stress in Abhängigkeit vom Koordinatenwert  $x_{ik}$  berechnet:

$$\frac{\delta STRESS(iter)}{\delta x_{ik}^{iter}},$$

Die partielle Ableitung wird als Gradient bezeichnet. Die Koordinaten werden anschließend verschoben mit

$$x_{ik}^{iter+1} = x_{ik}^{iter} - \alpha(iter) \cdot \frac{\delta STRESS(iter)}{\delta x_{ik}^{iter}}.$$

$\alpha$  ist die sogenannte Schrittlänge. Sie gibt bildlich gesprochen an, wie „weit“ man in Richtung des Minimums geht.

Die Schrittlänge wird in Almo in Abhängigkeit von folgenden Faktoren berechnet:

- der vorausgehenden Schrittlänge  $\alpha(iter - 1)$ .
- der relativen Verbesserung  $\frac{STRESS(iter)}{STRESS(iter - 1)}$  des Stresses aktuellen Iteration gegenüber dem Stress der vorausgehenden Lösung
- der relativen Verbesserung  $\frac{STRESS(iter)}{STRESS(iter - 5)}$  des Stresses gegenüber dem Stress der

fünf Iterationen zurückliegenden Lösung

- der Kreuzproduktsumme  $\frac{\sum \frac{\delta STRESS(iter-1)}{\delta x_{ik}^{iter-1}} \cdot \frac{\delta STRESS(iter)}{\delta x_{ik}^{iter}}}{g(iter) \cdot g(iter-1)}$  der partiellen Ableitungen der aktuellen und der vorausgehenden Iterationen
- der mittleren Quadratsumme  $g = \sqrt{\frac{\sum \left( \frac{\delta STRESS(iter)}{\delta x_{ik}^{iter}} \right)^2}{m}}$  der partiellen Ableitungen des aktuellen Iterationsschrittes

Durch die Berechnung der Schrittlänge soll erreicht werden, dass „größere“ Schritte in Richtung Minimum gemacht werden, wenn deutliche Verbesserungen erzielt werden können. Die Festlegung der Werte basiert auf Erfahrungen. Als kleinster Wert für die Schrittlänge ist 0,001 definiert. Wird dieser Wert unterschritten, bricht Almo die Iteration ab. Anstelle einer optimalen Berechnung könnte daher mit einer Schrittlänge von 0,0011 gerechnet werden. Dadurch würde sich aber die Rechenzeit verlängern.

Die partielle Ableitung des Stress nach dem Koordinatenwert  $x_{ik}$  ergibt sich als (Sixtl 1982: 347)

$$\frac{\delta STRESS(iter)}{\delta x_{ik}^{iter}} = STRESS(iter) \sum_{j=1}^m \left[ \frac{d_{ij} - \hat{d}_{ij}}{R} - \frac{d_{ij}}{T} \right] \cdot \frac{|x_{ik} - x_{jk}|^{p-1}}{d_{ij}^{p-1}} \cdot \sigma_{ij,k} \cdot$$

$d_{ij}$  sind die (mit Metrikparameter  $p$ ) berechneten Distanzen zwischen den Objekten  $i$  und  $j$ .

$\hat{d}_{ij}$  sind die Disparitäten (deren Rangordnung mit den empirischen Unähnlichkeiten übereinstimmen). Anmerkung: Aus Gründen der Einfachheit schreiben wir  $d_{ij}$  bzw.  $\hat{d}_{ij}$  statt

$d_p(A, B)$  bzw.  $\hat{d}_p(A, B)$ . Der Index  $i$  steht also für das Objekt A, der Index  $j$  für das Objekt B.

Auf die Spezifikation des Metrikparameters  $p$  wird verzichtet.

Der Parameter  $\sigma_{ij,k}$  ist definiert als

$$\sigma_{ij,k} = \begin{cases} -1 & x_{ik} < x_{jk} \\ 0 & x_{ik} = x_{jk} \\ +1 & x_{ik} > x_{jk} \end{cases}$$

und misst die Richtung der Veränderung, also die Richtung des Schritts in Richtung Minimum.

$R$  ist der Zähler und  $T$  der Nenner der Stressformel (Formel geschrieben für euklidische Metrik  $p=2$ ). Siehe auch oben.

$$S = \left[ \frac{\sum_{A < B} (d_p(A, B) - \hat{d}_p(A, B))^2}{\sum d_p(A, B)^2} \right]^{\frac{1}{2}} = \left[ \frac{R}{T} \right]^{\frac{1}{2}}$$

Die Iteration wird abgebrochen, wenn

- ein definiertes Minimum unterschritten ist, das durch den Benutzer in der Eingabe definiert werden kann (siehe Abschnitt P34.3),
- eine maximale Zahl an Iterationen überschritten ist, die ebenfalls vom Benutzer definiert werden kann (siehe Abschnitt P34.3), oder
- keine wesentliche Verbesserung erzielt oder ein lokales Minimum vermutet wird. In diesem Fall beginnt Almo mit einer neuen Startkonfiguration (sofern Zufallszahlen als Startwerte eingesetzt wurden). Die Zahl der Versuche kann ebenfalls definiert werden (siehe Abschnitt P34.4).

Die Berechnung der Modellparameter (Stress, Gradienten, Schrittweite usw.) kann mit Almo dokumentiert werden, wenn in den Programm-Masken Prog34mb,c,d in der Optionsbox "weitere Optionen" Zwischenergebnisse angefordert werden (siehe Abschnitt P34.4).

### **P34.1.5 Konkurrierende iterative MDS-Verfahren**

In der Literatur ist eine Fülle von Varianten des oben beschriebenen Verfahrens zu finden. Neben dem oben beschriebenen *Stress-1* nach Kruskal sind noch bedeutsam: *Roh-Stress* und *S-Stress*. Im Programm ALSCAL wird im iterativen Prozess der S-Stress minimiert, im SPSS-PROXSCAL der Rohstress. Auch das oben beschriebene Gradienten-Verfahren wurde vielfach modifiziert. Borg/Groenen (2005) beschreiben sehr ausführlich den Kalkül des Programms SMACOF, für das auch ein R-Programm existiert (de Leeuw/ Mair, 2009). Diese miteinander konkurrierenden Verfahren und Programme erbringen unterschiedliche Ergebnisse - bei eindeutiger Konfiguration mit nur minimalen Unterschieden. Die Verfahren und Programme haben ihre jeweiligen Vorteile und Nachteile. Das *beste* oder sogar das *richtige* Verfahren/Programm gibt es nicht. Auch Aussagen, die man gelegentlich lesen kann, wie "das moderne Programm X ist besser als das veraltete Programm Y" sollten mit Skepsis aufgenommen werden.

## **P34.2 Dateneingabe**

### **P34.2.1 Erstellen einer Un- oder Ähnlichkeitsmatrix**

Die Almo-Programme zur nichtmetrische MDS setzen voraus, dass eine Un- oder Ähnlichkeitsmatrix für bestimmte Objekte vorliegt. Diese kann mit folgenden anderen Almo-Programmen hergestellt werden.

1. Durch die Almo-Programme zur Ähnlichkeits-Skalierung, z.B. Prog33m2.

Das Programm wird gefunden durch Klick in der Knopfleiste auf den Knopf "Verfahren/Skalierungsverfahren". Prog33m2 erzeugt aus dem Paarvergleich von mehreren Objekten eine Ähnlichkeitsmatrix. Im bereits oben erwähnten MDS-Programm Prog34mc wird aus Paarvergleichen zuerst eine Ähnlichkeitsmatrix erzeugt und dann eine MDS gerechnet. Es ist also nicht notwendig das Almo-Programm Prog33m2 zur Ähnlichkeits-Skalierung zu verwenden.

2. Durch die Almo-Programme zur hierarchischen Clusteranalyse, z.B. Prog36md  
Das Programm wird gefunden durch Klick auf den Knopf "Verfahren/Clusteranalyse". Es rechnet eine hierarchische Clusteranalyse und formt eine Distanzmatrix der geclusterten Objekte. Im bereits oben erwähnten MDS-Programm Prog34md wird in einem Rechengang eine Clusteranalyse und dann eine MDS gerechnet. Es ist also nicht notwendig das Almo-Programm zu verwenden.

3. Durch die Almo-Programme zur Korrelation, z.B. Prog19bm

Das Programm wird gefunden durch Klick auf den Knopf "Verfahren/ Korrelation" In diesem Programm muss die Eingabe-Box "Option: Schreibe errechnete Matrix in Datei" geöffnet werden. Eine Korrelationsmatrix von Variablen kann zwar als Ähnlichkeitsmatrix verstanden werden, es muss aber kritisch überprüft werden, ob die miteinander korrelierten Variablen als Objekte verstanden werden dürfen, die Distanzen zueinander besitzen.

Zuerst soll die einfache Programm-Maske Pro34ma dargestellt werden. Sie benötigt nur eine minimale Eingabe des Benutzers und liefert auch nur knappe Ergebnisse. Danach wird die Programm-Maske Prog34mb dargestellt. Sie bietet viele Optionen an und liefert ein detailliertes Ergebnis und zusätzliche Grafiken. Anschliessend werden dann die einzelnen Eingabe-Boxen beider Programme beschrieben.

Beide Programme werden gefunden durch Klick auf den Knopf "Verfahren / MDS/MDU" am Oberrand des Almo-Fensters.



## P34.2.2 Eingabe in Maskenprogramm Prog34ma

Prog34ma.Msk

Nichtmetrische multidimensionale Skalierung (MDS) nach Kruskal

Bei der nichtmetrischen MDS wird versucht, Vergleichs-Objekte in einem p-dimensionalen Raum darzustellen, wobei p klein sein soll (z.B. 2 oder 3).

Beispiel: Die Distanzen zwischen 7 politischen Parteien sollen untersucht werden. Lassen sich diese durch eine Dimension (z.B. Links-Rechts-Dimension) oder durch mehrere Dimensionen erklären. Die Unähnlichkeitsmatrix ist folgende:

	Komm	Sozia	Arbei	Liber	Zentr	Chris	Konse
Kommunist. P.	0						
Sozialist. P.	8.7	0					
Arbeiterpart.	25.3	14.8	0				
Liberales	33.7	19.0	10.0	0			
Zentrumspartei	37.9	33.2	17.8	10.5	0		
Christliche P.	49.3	50.5	21.3	18.9	7.6	0	
Konservative	50.2	40.0	24.3	12.9	8.1	7.3	0


Beispiel aus: Fahrmeir/Hammerle, 1984: Multivariate statistische Verfahren. Berlin-New York, S. 678

Siehe Almo-Dokument 30 "Nicht-metrische multidimensionale Skalierung"



Programm-Bedienung ---> Hilfe



Speicher fuer x Variable Hilfe

Vereinbare Variable= 20

 Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert


Datei der Variablennamen Hilfe








 


 

zeige = Namensdatei in Output zeigen  
leer = nicht


**Freie Namensfelder** **Hilfe**

 Leere alle Eingabefelder dieser Sub-Box


 Name1=KommunistischePartei;
 Name2=SozialistischePartei;
 Name3=Arbeiterpartei;
 Name4=Liberale;
 Name5=Zentrumspartei;
 Name6=ChristlicheVolkspartei;
 Name7=Konservative;

 [erzeuge zusätzliche Namensfelder](#)


**Datei aus der die Matrix gelesen wird** **Hilfe**

 ".\TESTDAT\Kruskal.Mat"


---

 **7** Zahl der Objekte (Zeilen bzw. Spalten der Matrix)


---

 **0** Messniveau der Objekte **Hilfe**  
 0 = ordinal  
 1 = quantitativ (intervall-skaliert)


---


 **0** Matrix besteht **Hilfe**  
 0= aus unterem Dreieck mit Diagonale u. Zusatz-Informationen  
 1= aus unterem Dreieck mit Diagonale  
 2= aus oberem Dreieck mit Diagonale  
 3= Matrix ist quadratisch und symmetrisch


---


 **0** 0= die eingelesene Matrix ist eine Unähnlichkeitsmatrix  
 1= ist eine Ähnlichkeitsmatrix

**Programmparameter** **Hilfe**

 **1:3** Zahl der Dimensionen des gesuchten Raums  
 Im Beispiel: 1 Dimension, dann 2, dann 3 **Hilfe**

 **2** Metriken des Raums **Hilfe**  
 1=City-Block-Metrik  
 2=euklidische Metrik (empfohlen)

 **0.05** Stress. Abbruchkriterium **Hilfe**  
**Hilfe**

 **Grafik-Optionen**

[Programmende](#)

### P34.2.3 Eingabe in Maskenprogramm Prog34mb

Prog34mb.Msk  
 Nichtmetrische multidimensionale Skalierung (MDS) nach Kruskal  
 mit Optionen  
 mit vorgegebenen Startwerten  
 und mit Idealpunkten  
 für ordinale und intervall-skalierte Variable

Bei der nichtmetrischen MDS wird versucht, Vergleichs-Objekte in einem p-dimensionalen Raum darzustellen, wobei p klein sein soll (z.B. 2 oder 3).

Beispiel: Die Distanzen zwischen 7 politischen Parteien sollen untersucht werden. Lassen sich diese durch eine Dimension (z.B. Links-Rechts-Dimension) oder durch mehrere Dimensionen erklären. Die Unähnlichkeitsmatrix ist folgende:

	Komm	Sozia	Arbei	Liber	Zentr	Chris	Konse
Kommunist. P.	0						
Sozialist. P.	8.7	0					
Arbeiterpart.	25.3	14.8	0				
Liberale	33.7	19.0	10.0	0			
Zentrumspartei	37.9	33.2	17.8	10.5	0		
Christliche P.	49.3	50.5	21.3	18.9	7.6	0	
Konservative	50.2	40.0	24.3	12.9	8.1	7.3	0

Beispiel aus: Fahrmeir/Hammerle, 1984: Multivariate statistische Verfahren. Berlin-New York, S. 678

Doppelklick  
|

weitere Beispiel-Programme:

Ähnlichkeitsmatrix ---> [".\Almo\_Bsp\Aehnlich2.Alm" ]  
 Ähnlichkeit von Automarken ---> [".\Almo\_Bsp\Automarken2.Alm" ]  
 Ähnlichkeit von Getraenken ---> [".\Almo\_Bsp\Getraenke.Alm" ]  
 USA-Präs.schafts-Kandidaten ---> [".\Almo\_Bsp\Kandidaten.Alm" ]  
 Siehe auch die metrische MDS  
 zum Parteien-Beispiel ---> [".\Almo\_Bsp\PolitPartei.Alm" ]

Siehe Almo-Dokument 30 "Nicht-metrische multidimensionale Skalierung"

Programm-Bedienung --->

Speicher fuer x Variable

Vereinbare Variable=

Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

Datei der Variablennamen

zeige = Namensdatei in Output zeigen  
leer = nicht

Freie Namensfelder

Namen für die Objekte

<input type="checkbox"/>	Name1=KommunistischePartei
<input type="checkbox"/>	Name2=SozialistischePartei
<input type="checkbox"/>	Name3=Arbeiterpartei
<input type="checkbox"/>	Name4=Liberale
<input type="checkbox"/>	Name5=Zentrumspartei
<input type="checkbox"/>	Name6=ChristlicheVolkspartei
<input type="checkbox"/>	Name7=Konservative

erzeuge zusätzliche Namensfelder

Freie Namensfelder

Namen für die Idealpunkte  
nur wenn Idealpunkte eingesetzt werden

Leere alle Eingabefelder dieser Sub-Box

<input type="checkbox"/>	Name11=Person1
<input type="checkbox"/>	Name12=Person2
<input type="checkbox"/>	Name13=Person3

erzeuge zusätzliche Namensfelder

Datei aus der die Matrix gelesen wird












---

Matrix besteht

0= aus unterem Dreieck mit Diagonale u. Zusatz-Informationen  
1= aus unterem Dreieck mit Diagonale  
2= aus oberem Dreieck mit Diagonale  
3= Matrix ist quadratisch und symmetrisch

---

0= die eingelesene Matrix ist eine Unähnlichkeitsmatrix  
1= ist eine Ähnlichkeitsmatrix

Analyse-Variable (Objekte)	
 <input type="text" value="v1:7"/>	
 <input type="text" value="0"/>	Messniveau 0 = ordinal 1 = quantitativ (intervall-skaliert)
	<a href="#">Hilfe</a>
Programmparameter	
 <input type="text" value="2"/>	Zahl der Dimensionen des gesuchten Raums
 <input type="text" value="2"/>	Metriken des Raums 1=City-Block-Metrik 2=euklidische Metrik (empfohlen)
 <input type="text" value="0.05"/>	Stress. Abbruchkriterium
	<a href="#">Hilfe</a> <a href="#">Hilfe</a>
	Option: Startwerte
	<a href="#">Hilfe</a>
	Option: Idealpunkte
	<a href="#">Hilfe</a>
	Option: weitere Optionen
	Option: Rotation
	<a href="#">Hilfe</a>
	Option: "Aussehen" der auszugebenden Tabelle bzw. Matrix
	Grafik-Optionen
<a href="#">Programmende</a>	

### P34.2.4 Erläuterung zu den Eingabeboxen beider Programme

Die ersten grauen Eingabeboxen der beiden Programme sind gleich. Danach folgen die sandfarbenen Optionsboxen von Prog34mb.

#### **Eingabe-Box: Speicher für x Variable.**

Siehe dazu Almo-Dokument 0, Arbeiten mit Almo, Abschnitt P0.1

Eigentlich würde es ausreichen so viele Variable zu vereinbaren wie sie Matrix umfasst.

#### **Eingabe-Box: Option "Weitere Vereinbarungen".**

Siehe dazu Almo-Dokument 0, Arbeiten mit Almo, Abschnitt P0.2

Die Optionsbox braucht für unser Beispiel nicht geöffnet zu werden

#### **Eingabe-Box: Datei der Variablennamen**

Siehe dazu Almo-Dokument 0, Arbeiten mit Almo, Abschnitt P0.3

Für unser Beispiel mit den Ähnlichkeiten zwischen den 7 Parteien haben wir keine Datei der Variablennamen angelegt. Die Eingabefelder bleiben leer.

## Eingabe-Box: Freie Namensfelder

Freie Namensfelder Hilfe

Namen für die Objekte

- ↔ Name1=KommunistischePartei
- ↔ Name2=SozialistischePartei
- ↔ Name3=Arbeiterpartei
- ↔ Name4=Liberale
- ↔ Name5=Zentrumspartei
- ↔ Name6=ChristlicheVolkspartei
- ↔ Name7=Konservative

... erzeuge zusätzliche Namensfelder

Den Objekten (im Beispiel: den 7 Parteien), die die Ähnlichkeitsmatrix bilden, können Namen gegeben werden. Das ist nicht obligatorisch, jedoch empfehlenswert, da dann die Ergebnisse besser zu lesen sind. Die Namens-Nummerierung muss mit 1 beginnen und fortlaufend sein.

In Prog34mb (mit Optionen) ist noch folgende zusätzliche Eingabebox enthalten

Freie Namensfelder Hilfe

Namen für die Idealpunkte  
nur wenn Idealpunkte eingesetzt werden

↔ Leere alle Eingabefelder dieser Sub-Box

- ↔ Name11=Person1
- ↔ Name12=Person2
- ↔ Name13=Person3

... erzeuge zusätzliche Namensfelder

Sie wird nur ausgefüllt, wenn im Programm sogenannte "Idealpunkte" angefordert werden. Sonst sollten den Eingabefelder geleert werden. Die erste Namensnummer muss höher sein als die letzte für die Objekte und dann kontinuierlich forlaufen. Das wird später in Abschnitt P34.1.3.2 noch ausführlich dargestellt.

## Eingabe-Box: Datei aus der gelesen wird

Siehe dazu Almo-Dokument 0, Arbeiten mit Almo, Abschnitt P0

Dialogbox: Datei aus der die Matrix gelesen wird

Hilfe

".\TESTDAT\Kruskal.Mat"

Zahl der Objekte (Zeilen bzw. Spalten der Matrix) 7

Messniveau der Objekte 0 Hilfe

0 = ordinal  
1 = quantitativ (intervall-skaliert)

Matrix besteht 0 Hilfe

0= aus unterem Dreieck mit Diagonale u. Zusatz-Informationen  
1= aus unterem Dreieck mit Diagonale  
2= aus oberem Dreieck mit Diagonale  
3= Matrix ist quadratisch und symmetrisch

0= die eingelesene Matrix ist eine Unähnlichkeitsmatrix  
1= ist eine Ähnlichkeitsmatrix

### 1. Eingabefeld: Datei

Geben Sie in der Eingabebox „Datei aus der die Matrix gelesen wird“ zuerst den Namen der Datei an, in der sich die Matrix befindet. Sie können den vollen Pfad- und Dateinamen schreiben oder die in Windows zulässige Kurzversion mit dem Punkt. Der Punkt ersetzt in unserem Fall den Pfadteil "C:\Almo ". Einfacher ist es, wenn Sie auf den Knopf mit dem Öffne-Symbol klicken. Es erscheint dann eine Dialogbox, in der Sie gefragt werden, ob Sie die Datei anschauen wollen oder ob Sie in das Eingabefeld einen (neuen) Namen einsetzen wollen. Klicken Sie auf "Namen einsetzen". Es erscheint dann die gewohnte Datei-Auswahlbox von Windows, in der Sie die Datei suchen, in der sich Ihre Matrix befindet.

### 2. Eingabefeld: Zahl der Objekte

Die Zahl der Objekte, die die Matrix bilden, muß mitgeteilt werden, muß mitgeteilt werden

### 3. Eingabefeld: Messniveau

Der MDS-Kalkül nach Kruskal ist ursprünglich auf den Fall beschränkt gewesen, dass die Unähnlichkeiten zwischen den Objekten ordinal gemessen wurden. Er kann aber relativ einfach auch auf den Fall ausgedehnt werden, dass die Unähnlichkeiten quantitativ gemessen wurden. Lediglich die Disparitäten werden dann anders berechnet. Siehe dazu oben Abschnitt P34.0.2.3. "Quantitativ" heißt: Die Unähnlichkeiten sind mindestens auf dem Niveau der Interval-Skala gemessen.

### 4. Eingabefeld: Matrix besteht aus ...

Die einzulesende Matrix muss selbstverständlich symmetrisch sein. Der Benutzer muss angeben, welche Form die Matrix besitzt. Die Matrix besteht

- 0= aus dem unterem Dreieck mit Diagonale und 3 Sternen davor und 2 dahinter
- 1= aus dem unterem Dreieck mit Diagonale
- 2= aus dem oberem Dreieck mit Diagonale

3= Matrix ist quadratisch (und symmetrisch)

Format 0 ist ein spezielles Almo-Format, das auch hier im MDS-Programm in seiner einfachsten Variante (mit den Sternen als Ersatz für zusätzliche Matrix-Informationen) verwendet werden kann. Siehe Handbuch, Teil 2, Almo-Programmiersprache, Abschnitt 43.1.1.

*Beispiel für Format 0: Unteres Dreieck mit Diagonale, 3 Sterne davor, 2 danach*

```
*
*
*
3.6426
0.1564 4.8809
0.1158 0.8062 6.1752
0.7936 0.2069 0.6262 6.3897
0.8457 0.0718 0.2919 5.4792 6.9798
*
*
```

*Beispiel für Format 1: Unteres Dreieck mit Diagonale*

```
3.6426
0.1564 4.8809
0.1158 0.8062 6.1752
0.7936 0.2069 0.6262 6.3897
0.8457 0.0718 0.2919 5.4792 6.9798
```

*Beispiel für Format 2: Oberes Dreieck mit Diagonale*

```
3.6426 0.1564 0.1158 0.7936 0.8457
      4.8809 0.8062 0.2069 0.0718
            6.1752 0.6262 0.2919
                  6.3897 5.4792
                        6.9798
```

*Beispiel für Format 3: Ganze quadratisch-symmetrische Matrix*

```
3.6426 0.1564 0.1158 0.7936 0.8457
0.1564 4.8809 0.8062 0.2069 0.0718
0.1158 0.8062 6.1752 0.6262 0.2919
0.7936 0.2069 0.6262 6.3897 5.4792
0.8457 0.0718 0.2919 5.4792 6.9798
```

Beachte: Ist die einzulesende Matrix eine Unähnlichkeitsmatrix dann müssen die Diagonalglieder gleich 0 sein. Siehe nachfolgend 3.Eingabefeld.

*5. Eingabefeld: Ähnlichkeits- oder Unähnlichkeitsmatrix*

Ist die einzulesende Matrix bereits eine Unähnlichkeitsmatrix, dann überprüft Almo, ob negative Werte auftreten. Ist dies der Fall, dann wird der größte negative Wert zu allen Matrixwerten (absolut) hinzuaddiert. Alle Diagonalwerte werden dann noch zwangsweise auf 0 gesetzt.

Wird die Matrix vom Benutzer als Ähnlichkeitsmatrix gekennzeichnet, dann wird sie von Almo automatisch in eine Unähnlichkeitsmatrix transformiert. Die Ähnlichkeitskoeffizienten werden im Programm mit -1.0 multipliziert. Dann wird der größte negative Zahlenwert zu allen Matrixwerten (absolut) hinzuaddiert. Alle Diagonalwerte werden zwangsweise auf 0 gesetzt



## Eingabe-Box: Programmparameter

Programmparameter		
<input type="text" value="1:3"/>	Zahl der Dimensionen des gesuchten Raums Im Beispiel: 1 Dimension, dann 2, dann 3	<input type="button" value="Hilfe"/>
<input type="text" value="2"/>	Metriken des Raums 1=City-Block-Metrik 2=euklidische Metrik	<input type="button" value="Hilfe"/>
<input type="text" value="0.05"/>	Stress. Abbruchkriterium	<input type="button" value="Hilfe"/>

Die Anweisungen des Programmparameterblocks haben folgende Bedeutung:

### 1. Eingabefeld: Zahl der Dimensionen

Mit dieser Anweisung wird die Anzahl der Dimensionen für die gesuchte Punktekonfiguration festgelegt. Eingeben werden kann

nur 1 Zahlenwert. Beispiel: 2

Dann wird eine Analyse mit 2 Dimensionen gerechnet

mehrere durch Komma getrennte Zahlenwerte. Beispiel: 2,3

Dann wird entsprechend eine Analyse mit 2 Dimensionen, danach eine mit 3 Dimensionen gerechnet.

ein durch Doppelpunkt gekennzeichnetes Zahlenintervall. Beispiel: 1:4

Dann wird eine Analyse mit 1 Dimension,

danach eine mit 2 Dimensionen

danach eine mit 3 Dimensionen

danach eine mit 4 Dimensionen gerechnet

Wird das Eingabefeld leer gemacht, dann wird (für den Benutzer nicht sichtbar)  $1:n/5$  eingesetzt.  $n$  ist die Anzahl der Variablen (Objekte). Es können maximal 10 Dimensionen berechnet werden. Allgemein sollte folgendes Verhältnis zwischen der Zahl der Objekte und der maximalen Dimensionszahl bestehen (Sixtl 1982): 5 zu 1. Beispiel: Werden 20 Objekte analysiert, rechnet Almo mit maximal 4 Dimensionen. Ergibt sich bei Division mit 5 ein Rest, wird dieser abgeschnitten. Bei 23 Objekten rechnet Almo ebenfalls mit maximal 4 Dimensionen, außer der Anwender vereinbart eine andere Zahl.

### 2. Eingabefeld: Metrik des Raums

Mit diesen Anweisungen kann der Benutzer den Metrikparameter  $P$  der Minkowskimetrik variieren.

Möglich ist

1. nur eine Zahlenangabe, z.B. 2 (auch 2.0 mit Dezimalteil)

2. oder zwei durch Komma getrennte Zahlenangaben, z.B. 1, 3

Mit diesen Anweisungen kann der Benutzer den Metrikparameter  $P$  der Minkowskimetrik variieren.  $P$  nimmt in diesem Falle die Werte 1.0, 2.0, 3.0 an.

Für jede dieser Metriken rechnet Almo eine Analyse.

In der Regel ist nur die Eingabe der Metrik 2 sinnvoll. Dann wird die gewohnte euklidische Metrik verwendet. Gelegentlich wird auch noch der Metrik-Parameter 1 eingesetzt. Dies ist die "City-Block-Metrik". Andere Metriken sind inhaltlich kaum zu interpretieren.

### 3. Eingabefeld: Stress - Abbruchkriterium

Die Formel für den Stress-1-Koeffizienten wurde zu Beginn dieses Almo-Dokuments angegeben. Wird das Eingabefeld leer gemacht, dann gilt die Voreinstellung von 0.05.

Der Benutzer schreibt beispielsweise in das Eingabefeld die Zahl 0.1. Damit kann er die oberste Schranke für den Stress der gewünschten Punktekonfiguration festsetzen. Liegt der Stresswert für eine bestimmte Punktekonfiguration nach 30 Iterationen über diesem Schwellwert oder wird nach 5 Iterationen keine wesentliche Verbesserung mehr erzielt, dann bricht das Iterationsverfahren ab. Beim einfachen Prog34ma wird damit das Programm beendet. Bei Prog34mb mit Optionen werden dem Benutzer für diese Situation mehrere Optionen angeboten. Sie werden nachfolgend in den Abschnitten XXX vorgetragen.

Almo startet dann einen neuen Lösungsversuch mit anderen Startwerten. Weitere Abbruchkriterien sind: Nach fünf Iterationen wird keine wesentliche Verbesserung erzielt oder es wird ein lokales Minimum erreicht. Die Zahl der Lösungsversuche ist auf 5 beschränkt.

## P34.2.5 Erläuterung zu den Optionsboxen von Prog34mb

Prog34mb verfügt über Optionen, die die Ergebnisse verbessern und erweitern können. Besonders wertvoll sind die beiden folgenden:

1. Es können "gute" **Startwerte** eingegeben werden, von denen der Benutzer vermutet, dass sie gute Schätzwerte der endgültigen Ladungen sind. Oder es kann veranlasst werden, dass automatisch (für den Benutzer nicht sichtbar) eine metrische MDS vorausgerechnet wird und deren Ergebnis-Ladungen als Startwerte eingesetzt werden.

2. Es können **Idealpunkte** eingesetzt werden. D. h. es können Subjekte (Personen) eingeführt werden, die Präferenz- bzw. Distanz-Urteile zu den Objekten äußern. Diese Subjekte werden in der nichtmetrischen MDS nach Kruskal (etwas verwirrend) "Idealpunkte" genannt. Die MDS erzeugt dann als Ergebnis eine gemeinsame Ladungsmatrix der Subjekte und Objekte. Entsprechend werden dann die Idealpunkte auch als Punkte in das gemeinsame Punkte-Diagramm eingesetzt. Damit kann dann auch die Distanz der Subjekte zu den Objekten und der Subjekte zueinander festgestellt werden.

In Prog34mb ist deswegen noch noch folgende zusätzliche Eingabebox enthalten, auf die wir oben bereits hingewiesen haben

The screenshot shows a dialog box titled "Freie Namensfelder". Inside, there is a text label "Namen für die Idealpunkte" in blue, followed by a red button labeled "Hilfe". Below this is the instruction "nur wenn Idealpunkte eingesetzt werden". A list of options is shown, each with a double-headed arrow icon: "Leere alle Eingabefelder dieser Sub-Box", "Name11=Person1", "Name12=Person2", and "Name13=Person3". At the bottom, there is a button with three dots and the text "erzeuge zusätzliche Namensfelder".

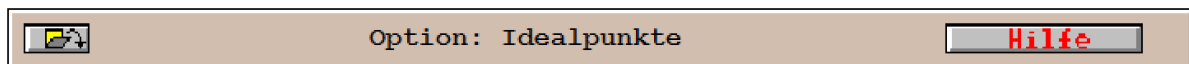
Sollen Subjekte (Personen) eingeführt werden, die Präferenz- bzw. Distanz-Urteile zu den Objekten äußern, dann werden diese in der MDS nach Kruskal als sogenannte "Idealpunkte" eingeführt.

Werden Idealpunkte eingesetzt dann müssen diesen in der Optionsbox "Idealpunkte" Variablennummern zugefügt werden. Siehe nachfolgend diese Box. Die Nummern dürfen beliebig sein. Sie dürfen jedoch nicht identisch mit den Variablennummern der Objekte sein. Und sie müssen kleiner sein als die (in der Vereinbare-Box) vereinbarte Variablenzahl. Wir haben in unserem Beispiel die Nummern 11,12,13 verwendet. Die Variablen in der Unähnlichkeitsmatrix haben die Nummern 1 bis 7. In der Vereinbare-Box wurden 20 Variable vereinbart.

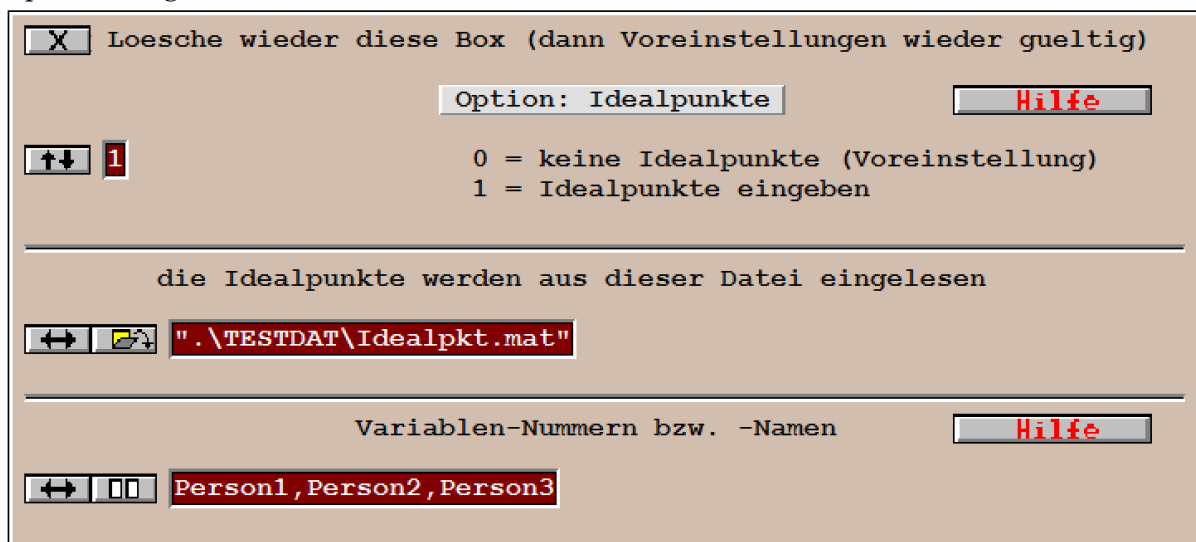
Den Idealpunkten können auch (aber müssen nicht) Namen gegeben werden. Das geschieht in der obigen Eingabebox. Erhält z.B. ein Idealpunkt die Variablennummer 11 dann kann z.B. folgender Name vergeben werden

Name 11 = Person1;

### P34.2.5.1 Optionsbox: Idealpunkte



Optionsbox geöffnet



Eine Person äußert gegenüber den Objekten bestimmte Präferenzen bzw. Distanzen. Diese Person soll nun als Punkt in den Raum gestellt werden, in dem sich bereits die Objekte befinden. Dieser Punkt wird "Idealpunkt" genannt. Dieser etwas seltsam anmutende Begriff ist so zu verstehen:

Befindet sich unter den Objekten eines, zu dem die Person eine Distanz von 0 hat, dann ist dieses für die Person ein "ideales" Objekt. Das ist dann auch der Punkt der Person im Raum. Ein solches Objekt muss jedoch nicht notwendigerweise real in den Daten existieren. Es gilt also

Gesucht wird der Punkt im Raum *wäre dort ein Objekt* zu dem die Person eine Distanz von 0 hat. Das wäre für die Person ein ideales Objekt. Das ist ihr Idealpunkt.

Wir wollen das Vorgehen für 3 Personen P1, P2, P3 zeigen. Die Zahl der Personen ist in Almo nicht beschränkt

Objekt	Personen		
	P1	P2	P3
-----	--	--	--
KommunistischeP	0	3	6
SozialistischeP	1	2	5
Arbeiterpartei	2	1	5
Liberale	3	4	4
Zentrumspartei	4	4	3
ChristlicheVolk	5	4	2
Konservative	6	5	1

Die Zahlenwerte bezeichnen Distanzen. Die Person P1 hat eine Distanz von 0 zur Kommunistischen Partei und eine maximale Distanz zu den Konservativen. P3 ist fast das exakte Gegenteil zu P1. Nur bei den Sozialisten und der Arbeiterpartei unterscheidet P3 nicht. Zu beiden beträgt die Distanz 5 Einheiten. Die Zahlenwerte sind ordinal, d.h. es handelt sich um Rangziffern, nicht um quantitative Ziffern. Ihre absolute Größe ist irrelevant.

Almo berechnet nun die Ladungen der 3 Personen für die Zahl der Dimensionen, die in der Eingabebox "Programmparameter" zuvor angegeben wurde.

Die Daten können empirisch gewonnen worden sein oder auch nur hypothetisch unterstellt worden sein.

#### 1. Eingabefeld:

Hier wird 1 eingegeben, wenn Idealpunkte berechnet werden sollen

#### 2. Eingabefeld:

Es muss die Datei angegeben werden, in der die Idealpunkte gespeichert sind. In unserem Beispiel enthält die Datei folgende Daten

```
0 3 6
1 2 5
2 1 5
3 4 4
4 4 3
5 4 2
6 5 1
```

#### 3. Eingabefeld:

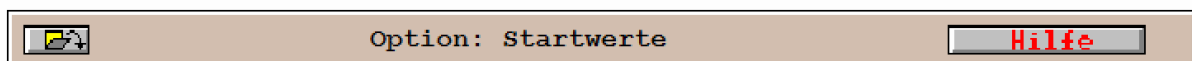
Die den Idealpunkten zugewiesenen Variablennummern oder die in obiger Eingabebox "Freie Namensfelder" definierten Variablennamen müssen einhetragen werden. In unserem Beispiel also entweder

V11, 12,13

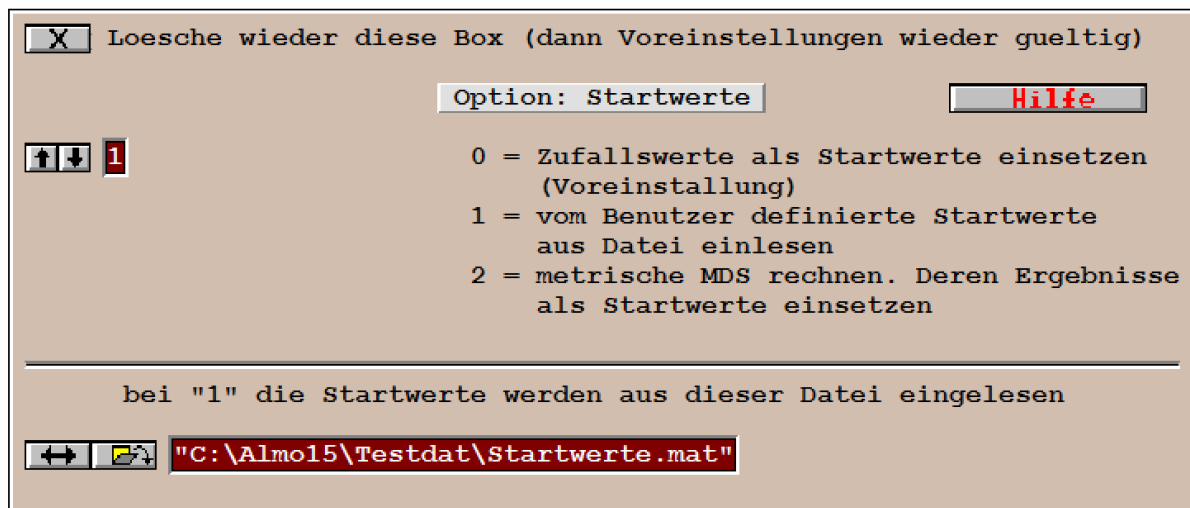
oder

Person1, Person2, Person3

### P34.2.5.2 Optionsbox: Startwerte



Wird die Optionsbox geöffnet, dann sieht man folgendes



Das Kruskalverfahren ist ein iteratives Verfahren, das ausgehend von einer vermutlichen Lösung (als Startwerten) zur endgültigen Lösung gelangt.

Im oben beschriebenen Prog34ma setzt Almo Zufallswerte als Startwerte ein, die dann in einem Iterations-Verfahren bis hin zur endgültigen Ladungsmatrix der Punkte verbessert werden. Die endgültige Ladungsmatrix wird erheblich von den Startwerten bestimmt. Andere Startwerte erzeugen andere Ladungsmatrizen. Allerdings sind sie im Wesentlichen nur dadurch verschieden, dass die Faktoren bzw. Achsen gedreht, gespiegelt, gestreckt und gekürzt sind. Dies wird in den Punkte-Diagrammen sichtbar. Diese sind in ihrer Struktur sehr ähnlich.

Werden "gute" Startwerte eingesetzt, dann hat das zur Folge, (1) dass weniger Iterationen gebraucht werden, um die vom Benutzer vorgegebene Stress-Schwelle zu erreichen bzw. zu unterschreiten und (2) dass die gefundene Ladungsmatrix der unbekanntenen wahren Ladungsmatrix (hoffentlich) sehr nahe kommt.

In der Programm-Maske Prog34mb besteht nun die Möglichkeit, Startwerte vorzugeben.

#### *Bei Eingabe von 0*

werden Zufallswerte eingesetzt. Dies ist die Voreinstellung. D.h. der Benutzer braucht die Optionsbox nicht zu öffnen

#### *Bei Eingabe von 1: Startwerte aus Datei*

Es werden Startwert aus einer Datei eingelesen. Im 2. Eingabefeld der Optionsbox muss der Name dieser Datei mitgeteilt werden, z.B. als vollständiger Pfad-und Dateiname

```
"C:\Almo15\Progs\Startwerte.mat"
```

Die Apostrophe dürfen nicht vergessen werden. Möglich ist auch die Windows-Kurzform

```
".\Progs\Startwerte.mat"
```

Da Almo aus dem Ordner "C:\Almo15" gestartet wird (dort befindet sich Almo.exe) und Progs ein Unterordner von Almo15 ist, kann dieser Pfadteil durch einen Punkt symbolisiert werden.

Selbstverständlich muss der Benutzer zuvor über das Menü "Datei/Neue Datei anlegen" eine Datei dieses Namens erzeugt haben und in diese die Startwerte geschrieben haben.

Am besten ist es als Ordner, in den die Datei gestellt wird, den Almo-Unterordner "Progs" zu

selektieren; keinesfalls den Ordner "Testdat" (wie in obigem Bild). Also erzeugt dann ein neues Fenster, in das die Daten geschrieben werden.

Der Benutzer gibt inhaltlich begründete Startwerte ein, die vom Almo-Programm aus einer Datei eingelesen werden - und zuvor vom Benutzer selbstverständlich in diese geschrieben werden mussten. Wenn er eine maximal 2-dimensionale Lösung erwartet, dann spekuliert er über die vermutliche Ladungsmatrix und gibt diese als Matrix der Startwerte mit 2 Spalten ein. Wenn der Benutzer mehrere Dimensionen anfordert, z.B. 1,2,3, dann muss die Spaltenzahl der Startwertematrix der maximalen Dimensionszahl entsprechen, bei 1,2,3 Dimensionen also 3 Spalten.

Im Beispiel in Prog34mb sollen die Ähnlichkeitsbeziehungen zwischen sieben politischen Parteien untersucht werden. Folgende theoretische Startkonfiguration wird vorgegeben:

Dimension 1: Diese soll die Rechts-Links-Dimension abbilden  
Ein negativer Zahlenwert bedeutet "links", ein positiver "rechts".  
Die kommunistische Partei erhält folglich einen Startwert von "-3" und die Konservativen einen Startwert von "+3".

Dimension 2: Diese soll den Grad des Extremismus abbilden.  
Den vermuteten "linken" und "rechten" Parteien wird ein positiver Startwert von "1" zugewiesen, den Parteien in der Mitte ein Startwert von "0".

Im Beispiel in Prog34mb wird so folgende Startwerte-Matrix eingegeben:

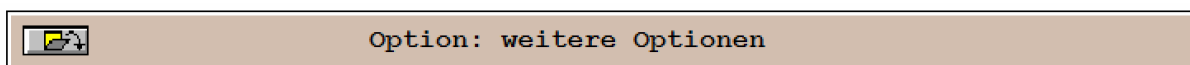
```
-3 1
-2 1
-1 0
 0 0
 1 0
 2 1
 3 1
```

### *Bei Eingabe von 2: Ladungsmatrix aus metrischer MDS als Startwerte*

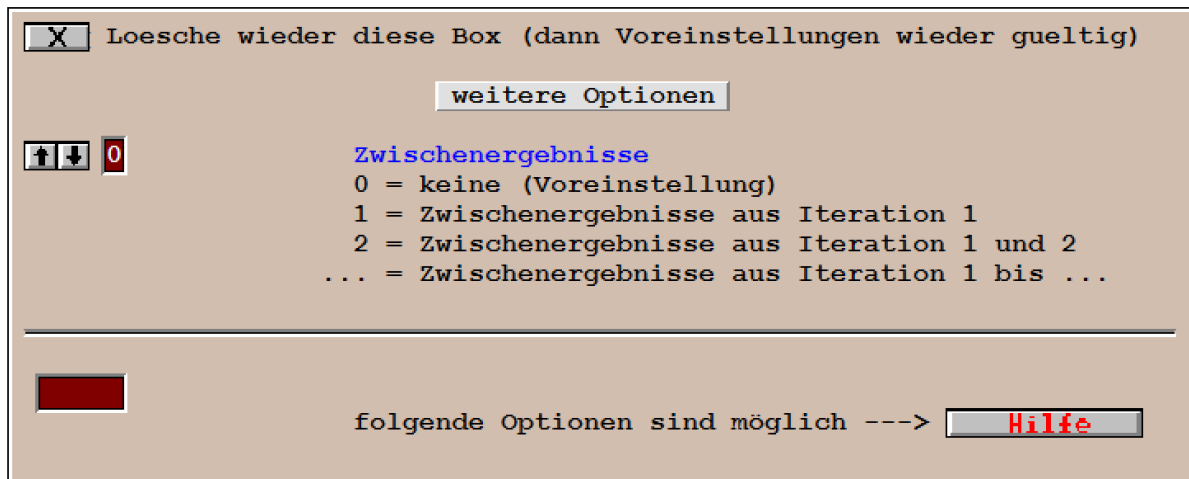
Zuerst eine metrische MDS rechnen und deren Ergebnisse als Startwerte einsetzen. Also rechnet automatisch, ohne dass das für den Benutzer sichtbar wird, eine metrische MDS und übergibt deren gefundene Ladungsmatrix als Startwerte an den Kalkül der nicht-metrischen MDS.

Die durch die nicht-metrische MDS gefundene Ladungsmatrix für die Objekte wird erheblich durch die Startwerte bestimmt. Der Gedanke ist nun naheliegend, die bei einer metrischen MDS gefundene Ladungsmatrix als Startwerte-Matrix für Prog34mb zu verwenden. Man wird natürlich die Frage stellen: Warum wird nicht gleich die Ladungsmatrix aus der metrischen MDS als endgültiges Ergebnis betrachtet? Wenn (nach Meinung des Forschers !!) die Ähnlichkeits-Daten ordinal sind, dann darf nur die nicht-metrische MDS gerechnet werden. Die Ladungsmatrix aus der metrischen MDS kann dann aber, wohl als die bestmögliche Schätzung, als Startwerte-Matrix verwendet werden.

### **P34.2.5.3 Optionsbox: Weitere Optionen**



Wird die Optionbox geöffnet dann sieht man folgendes



### 1. Eingabefeld: Zwischenergebnisse

0 = es werden keine Zwischenergebnisse ausgegeben (das ist die Voreinstellung)

1 = die Berechnungen im 1. Iterationsschritt des Kruskal-Kalküls werden ausgegeben.

2 = die Berechnungen im 1. und 2. Iterationsschritt des Kruskal-Kalküls werden ausgegeben.

x = die Berechnungen vom 1. bis zum x-ten Iterationsschritt des Kruskal-Kalküls werden ausgegeben. Für x kann zwar eine beliebig große Zahl eingesetzt werden. Da aber sehr viel Ausgabertext erzeugt wird, sollte der Benutzer nicht übertreiben.

Es werden nur die Zwischenergebnisse der Iterationen aus dem 1. Lösungsversuch ausgegeben. D.h. für einen eventuell notwendigen 2. Versuch werden keine Zwischenergebnisse ausgegeben.

### 2. Eingabefeld: Weitere Optionen

Hier können Optionen eingesetzt werden, die nicht über eine der anderen Optionsboxen aktivierbar sind. Es können beliebig viele Angaben hintereinander in das Eingabefeld geschrieben werden, beispielsweise so:

```
Option 5=0.5; Option 6=40; Option 7=10;
```

Folgende Optionen können eingesetzt werden:

**Option 5=...;** (Schrittweite des Metrikparameters)

Der Benutzer hat beispielsweise in der Box "Programmparameter" für die Metrik des Raums angegeben: 1, 3

Almo unterstellt dann eine Schrittweite von 1.0, rechnet also 3 Analysen für die Metriken 1, 2 und 3

Wird `Option 5 = 0.5;` gesetzt, dann beträgt die Schrittweite des Metrikparameters 0.5.

Es werden also folgende Metrikparameter eingesetzt: 1 1.5 2 2.5 3

Die Voreinstellung ist `Option 5=1.0;` Siehe auch oben in Abschnitt P34.1.2 die Erläuterungen zur Eingabebox "Programmparameter"

**Option 6=...;** (Zahl der Iterationen)

Beispiel: `Option6=40;` Mit dieser Anweisung kann der Benutzer die maximale Anzahl von Iterationen für jeden Lösungsversuch festlegen. Die Voreinstellung ist "`Option 6=30;`".

**Option 7=...;** (Zahl der Lösungsversuche)

Werden vom Benutzer Startwerte über eine Datei eingegeben oder wird das Ergebnis aus einer metrischen MDS als Startkonfiguration eingegeben, dann gibt es nur einen Lösungsversuch. Die Zahl der Iterationen innerhalb des einen Lösungsversuchs kann durch Option 6 beeinflusst werden. Werden hingegen Zufallswerte als Startwerte eingegeben, dann kann mit Option 7 die Zahl der Lösungsversuche erhöht werden. Bei z.B. "Option 7=10;" werden für jede Dimensionszahl und jeden Metrikparameter maximal 10 Lösungsversuche gerechnet, um eine konvergente Lösung zu finden. "Option 7=5;" ist die Voreinstellung. Bei jedem Lösungsversuch erfolgen dann maximal so viele Iterationen wie voreingestellt sind (=30) oder in Option 6 festgelegt wurden

**Option 8=...;** (Zufalls-Startwert)

Sind Zufallswerte als Startwerte eingestellt, dann kann der Zufallszahlengenerator beeinflusst werden. Die Startzahl kann verändert werden. Die Voreinstellung ist 123123. Ein Durchrechnen des Kruskalverfahrens mit unterschiedlichen Ausgangslösungen kann aus mehreren Gründen vorteilhaft sein:

1. Zur Bestimmung der Dimensionszahl und des Metrikparameters.
2. Zum Schutz vor lokalen Minima
3. Wegen der Abhängigkeit der gefundenen Punktekonfiguration von der Ausgangslösung.

**Option 34=...;** (erforderliche Stress-Verbesserung)

Wenn eine Iteration gegenüber der vorhergehenden eine Stress-Verminderung von unter x % erbringt, dann wird die Iteration noch einmal mit einer anderen Schrittweite wiederholt. Voreingestellt sind 5 %. Dieser Wert kann vermindert werden (z.B. auf 2.5 %) oder auch erhöht werden.

BEACHTTE: Das Prozentzeichen wird nicht geschrieben - nur der Zahlenwert.

BEACHTTE: Bei quantitativen Unähnlichkeiten wird der voreingestellte oder veränderte Wert von Almo automatisch halbiert.

**Option 35=...;** (erforderliche Stress-Verbesserung zwischen 5 Iterationen)

ab Iteration 10 wird geprüft, ob sich bei Iteration i der Stress im Vergleich zur aktuellen Iteration i um x% verbessert (vermindert) hat. Voreingestellt ist 5. Dieser Wert kann vermindert werden (z.B. auf 2.5 %) oder auch erhöht werden.

BEACHTTE: Das Prozentzeichen wird nicht geschrieben - nur der Zahlenwert.

**Option 36=...;** (Ausgabe der Lösungsversuche)

Werden Zufallszahlen als Startwerten eingesetzt, dann rechnet Almo in aller Regel mehrere Lösungsversuche mit verschiedenen Zufalls-Startwerten und gibt nur die Ergebnisse aus dem bestem Lösungsversuch aus. Dies ist die Voreinstellung. Wird Option 36=1 gesetzt, dann werden alle Lösungsversuche ausgegeben und zusätzlich die beste nochmals zum Schluss

**Option 37=...;** (Zahl der Iterationen bei Idealpunkt-Berechnung)

Die Idealpunkte werden (vergleichbar der MDS) in einem iterativen Kalkül ermittelt. Die Zahl der Iterationen kann vom Benutzer eingestellt werden, z.B. auf Option37=20. Die Voreinstellung ist 30

**Option 38=...;** (Wert der geforderten Stress-Schwelle bei Idealpunkt-Berechnung)

Für die Idealpunkte wird (vergleichbar der MDS) ein Mindest-Stress-1-Wert vorgegeben, der im iterativen Kalkül erreicht werden soll. Die Höhe dieses Stress-Wertes kann vom Benutzer eingestellt werden, z.B. Option38=0.01. Die Voreinstellung ist 0.05.

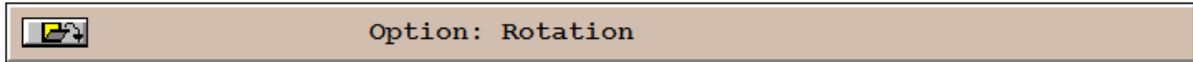


**Option 39=...;** (Punkte-Wanderung)

Die Position eines Objekts im MDS-Raum wird von einer Iteration zur nächsten hin- und hergeschoben. Für ein Objekt z.B. für das 3. Objekt kann die Punkte-Wanderung zahlenmäßig und grafisch dargestellt werden durch Option39=3;

In Abschnitt P34.3.6.7 werden wir das an einem Beispiel detailliert darstellen.

#### P34.2.5.4 Optionsbox: Rotation



Almo liefert eine 2- oder 3-dimensionale Grafik. Die Lage der Koordinatenachsen kann im Grafik-Editor beliebig gedreht werden. Die Punkte können auch um die Achsen gespiegelt werden. Rotation bedeutet, dass die Koordinatenachse gedreht werden, wobei die Punkte aber stehen bleiben und nicht mitgedreht werden. Die Koordinatenachsen bleiben dabei rechtwinklig aufeinander stehen - oder der Winkel zwischen ihnen wird verändert. Sinn der Rotation ist es eine Position zu finden, bei der die Achsen inhaltlich sinnvoll interpretierbar sind.

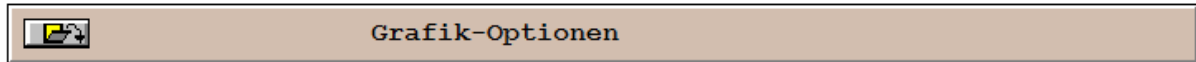
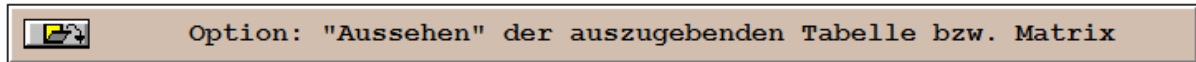
Eine Rotation ist in der Regel nur sinnvoll, wenn "Punktwolken" vorhanden sind, die deutlich voneinander getrennt sind und in sich gut geschlossen sind. Die vorhandenen rechtwinkligen oder schiefwinkligen Rotationskalküle verwenden bestimmte Prinzipien, die eher bei der klassischen Faktorenanalyse als bei der MDS angebracht sind. Lesen Sie die ausführliche Darstellung im Almo-Dokument 15 "Faktorenanalyse", Abschnitt P30.1.9.

Wenn Sie aber doch rotieren wollen, dann öffnen Sie die Optionsbox. Sie sehen dann folgendes:



Im Almo-Dokument Nr. 15 "Faktorenanalyse" wird diese Eingabebox und der Vorgang der Rotation ausführlich dargestellt. Siehe dort die Abschnitte P30.1.4, P30.2.3 Seite 29 und

### P34.2.5.5 Optionsbox: "Aussehen" der auszugebenden Tabelle, Grafik-Optionen



Siehe zu diesen beiden Optionsboxen das Almo-Dokument 0 "Arbeiten mit Almo", Abschnitt P0.7.3 und P0.7.4.

### P34.3 Ausgabe: Ergebnisse aus Prog34ma und Prog34mb

Zuerst sollen die Ergebnisse gezeigt werden, die beide Programme, das einfache Prog34ma und das mit Optionen ausgestattete Prog34mb liefern. Danach werden die Ergebnisse vorgestellt, die Prog34mb zusätzlich ausgibt.

#### P34.3.1 Analyse mit 1 Dimension mit Prog34ma

Unaehnlichkeitsmatrix

		Kommunis	Sozialis	Arbeiter	Liberale	Zentrums	Christli	Konserva
		V1	V2	V3	V4	V5	V6	V7
Kommunis	V1	0	8.7000	25.3000	33.7000	37.9000	49.3000	50.2000
Sozialis	V2	8.7000	0	14.8000	19.0000	33.2000	50.5000	40.0000
Arbeiter	V3	25.3000	14.8000	0	10.0000	17.8000	21.3000	24.3000
Liberale	V4	33.7000	19.0000	10.0000	0	10.5000	18.9000	12.9000
Zentrums	V5	37.9000	33.2000	17.8000	10.5000	0	7.6000	8.1000
Christli	V6	49.3000	50.5000	21.3000	18.9000	7.6000	0	7.3000
Konserva	V7	50.2000	40.0000	24.3000	12.9000	8.1000	7.3000	0

Als Startkonfiguration ist in Prog34ma die Ladungsmatrix aus der metrischen MDS voreingestellt. Das kann in Prog34mb verändert werden, jedoch nicht im einfachen Prog34ma. Entsprechend der Benutzeranweisung wird zunächst eine 1-dimensionale Lösung versucht

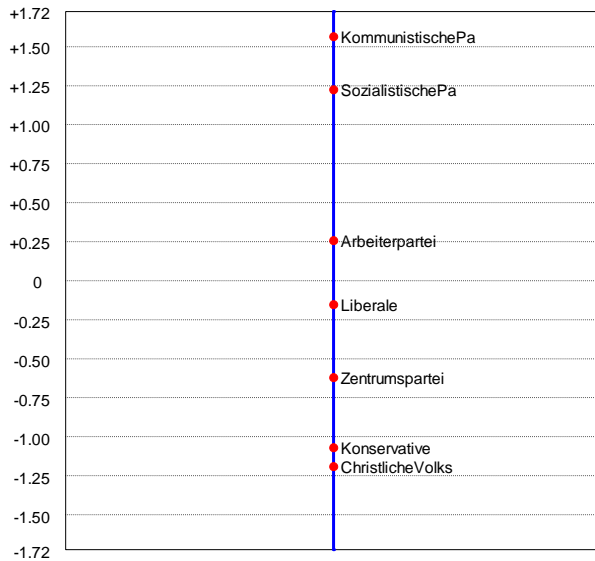
Ladungsmatrix der Objekte

		Faktor 1
Kommunis	V1	1.5646
Sozialis	V2	1.2227
Arbeiter	V3	0.2621
Liberale	V4	-0.1598
Zentrums	V5	-0.6248
Christli	V6	-1.1940
Konserva	V7	-1.0708

Stress-1 0.0475  
Anpassungsindex 0.9915  
Iterationen 3

Startwerte sind Ergebnismatrix aus metrischer MDS

Almo zeichnet folgendes 1-dimensionale Punktediagramm



Gemäß der Benutzereingabe müsste nun eine Lösung mit 2 und danach mit 3 Dimensionen gerechnet werden. Also meldet allerdings dass aus der metrischen MDS nur 1 Faktor entsteht. Die Analyse wird damit beendet.

### P34.3.2 Analyse mit 2 Dimensionen mit Prog34mb

Wir rechnen nun Prog34mb mit den 7-Parteien-Daten - mit folgenden Einstellungen:

Dimensionen: 2  
 Startwerte: Zufallswerte  
 Stress-Schwelle: 0.05

Wir rechnen also mit folgenden Einstellungen

**Programmparameter**

↔	<b>2</b>	Zahl der Dimensionen des gesuchten Raums	Hilfe
↔	<b>2</b>	Metriken des Raums 1=City-Block-Metrik 2=euklidische Metrik (empfohlen)	Hilfe
↔	<b>0.05</b>	Stress. Abbruchkriterium	Hilfe Hilfe

X Loesche wieder diese Box (dann Voreinstellungen wieder gueltig)

**Option: Startwerte** Hilfe

↑↓	<b>0</b>	0 = Zufallswerte als Startwerte einsetzen 1 = vom Benutzer definierte Startwerte aus Datei einlesen 2 = metrische MDS rechnen. Deren Ergebnisse als Startwerte einsetzen (Voreinstellung)	
----	----------	---	--

---

bei "1" die Startwerte werden aus dieser Datei eingelesen

↔ [File Icon] [Color Box]

Die Optionsbox "Startwerte" wurde geöffnet. Im 1. Eingabefeld wurde 0 (=Zufallswerte)

eingesetzt. Das 2. Eingabefeld wurde leer gemacht.

Almo liefert folgende Ergebnisse:

Dimensionen 2 Metrik 2.0 Versuch 1

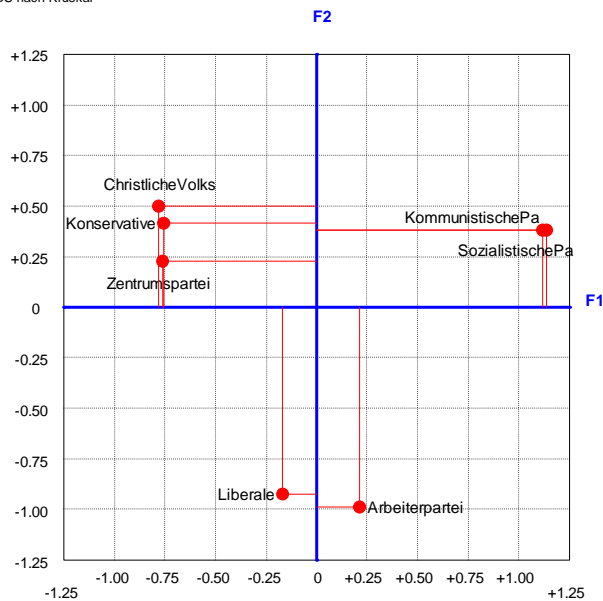
Gesuchte Punktekonfiguration  
Ladungsmatrix der Objekte

		Faktor 1	Faktor 2
Kommunis	V1	1.1194	0.3844
Sozialis	V2	1.1379	0.3844
Arbeiter	V3	0.2160	-0.9899
Liberale	V4	-0.1689	-0.9243
Zentrums	V5	-0.7640	0.2273
Christli	V6	-0.7813	0.5029
Konserva	V7	-0.7591	0.4153

Stress 0.0373  
Anpassungsindex 0.9930  
Iterationen 13  
Startwert f. Zufallsgenerator 123123

Gleich im 1. Versuch wird nach 13 Iterationen die vorgegebene Stress-Schwelle von 0.05 unterschritten. Almo zeichnet folgendes 2-dimensionale Punktediagramm

MDS nach Kruskal



Kommunistische und sozialistische Partei sind so nahe beieinander, dass in der Grafik die beiden Punkte sich beinahe überdecken.

Nun rechnen wir dieselbe Analyse mit einer vorgegebenen Stress-Schwelle von 0. Ein Stress von 0 kann bei empirischen (fehlerbehafteten) Daten nicht auftreten. Der iterative Kalkül wird also irgendwie beendet werden müssen. Almo liefert folgendes Ergebnis:

Einstellungen

```
-----
Matrix: Unaehnlichkeitsmatrix
Messniveau: ordinal
Dimensionen: 2
Metrik: 2
Startkonfiguration: Zufallswerte
```

Zahl der Lösungs-Versuche: 5

**Stress-Schwelle: 0**

geforderte Stress-Verbesserung von einer Iteration zu naechsten: 1 %

geforderte Stress-Verbesserung zwischen 5 Iterationen: 5 %

**maximale Zahl der Iterationen: 30**

Erläuterung: Bedeutsam für uns sind die beiden Einstellungen "Stress-Schwelle 0" und "maximale Iterationszahl 30". Letzere bedeutet dass der iterative Kalkül nach spätestens 30 Iterationen abgebrochen wird, wenn ein Stress von 0 nicht erreicht wird.

Es wurden 5 Loesungsversuche mit verschiedenen Zufalls-Startzahlen gerechnet

Versuch	Zufallszahl	Stress-1
1	123123	0.000012
2	3469186011	0.013692
3	720542595	0.026036
4	4221235691	0.287572
5	288604435	0.024224

Den besten Stress-1- Wert erbrachte: Versuch 1

Dieser Loesungsversuch wird nun ausgegeben

Erläuterung: Da Zufallswerten als Startwerte eingesetzt wurden, versucht Almo in maximal 5 Lösungsversuchen den geforderten Stress von 0 zu erreichen. Dabei werden jedes mal 30 Iterationen probiert. Diese Zahl könnte in der Optionsbox "Zwischenergebnisse und weitere Optionen" mit Option 6 verändert (z.B.auf 100 gesetzt) werden. Der Lösungsversuch, der den besten Stress-Koeffizienten erbringt wird ausgegeben. Das ist in unserem Beispiel gleich der 1. Versuch.

Ladungsmatrix der Objekte

		Faktor 1	Faktor 2
Kommun	V1	1.1194	0.4715
Sozial	V2	1.1115	0.2805
Arbeit	V3	0.1807	-0.9849
Libera	V4	-0.0943	-0.9573
Zentru	V5	-0.7459	0.2837
Christ	V6	-0.7857	0.4533
Konser	V7	-0.7857	0.4533

Stress-1 0.000012  
Anpassungsindex 1.000000  
Versuch Nr. 1  
Iterationen 31  
Startwert f. Zufallsgenerator 123123

Stress-1 nach Kruskal je Iteration  
Stress-Minderung

Iteration	Stress-1	in %
1	0.354052	-
2	0.312395	11.77
3	0.303071	2.98
4	0.292731	3.41
5	0.288192	1.55
6	0.276584	4.03
7	0.265676	3.94
8	0.244357	8.02
9	0.197681	19.10
10	0.101847	48.48
11	0.063361	37.79
12	0.057285	9.59
13	0.037302	34.88
14	0.019683	47.23
15	0.008427	57.19

16	0.005213	38.14
17	0.003432	34.17
18	0.001960	42.88
19	0.001135	42.11
20	0.000586	48.37
21	0.000443	24.40
22	0.000340	23.31
23	0.000213	37.38
24	0.000149	29.78
25	0.000116	22.57
26	0.000074	35.78
27	0.000048	34.73
28	0.000035	28.62
29	0.000026	23.64
30	0.000019	29.74
31	0.000012	33.01

Gesamtzahl der notwendigen Wiederholungen aus  
 allen Iterationen wegen nicht ausreichender  
 Stress-Verbesserung in Versuch 1: 1

Achtung: Anzahl zulaessiger Iterationen wurde ueberschritten  
 ----- vorgegebene Stress-Schwelle nicht erreicht  
 iterativer Kalkuel wird beendet

\*\*\*\*\* WARNUNG

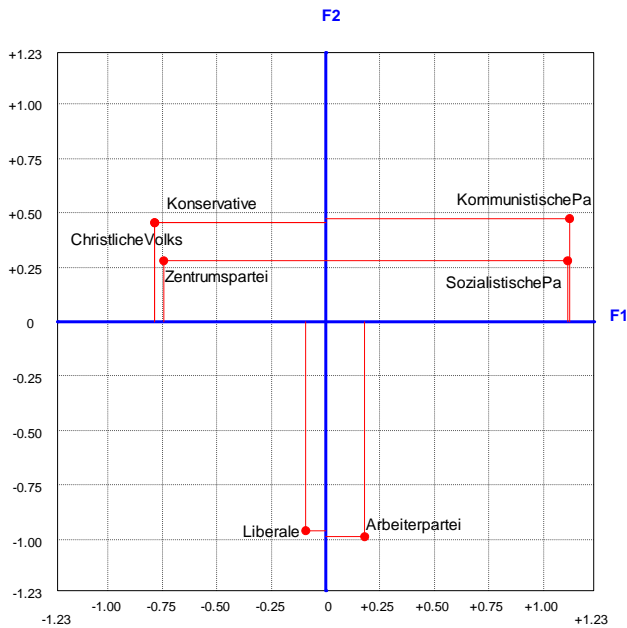
Die vorgegebene Stress-Schwelle wurde nicht erreicht  
 Abhilfe ----->

Erläuterung: Ab ca. der 13. Iteration ist zwar die prozentuale Stress-Minderung groß, die absolute Minderung jedoch sehr gering. Almo teilt noch mit, dass 1 Wiederholung notwendig war. Wenn der Benutzer auf den Hilfeknopf klickt (in der Almo-Ausgabe, nicht hier) dann wird ihm erklärt, was damit gemeint ist.

*Wiederholung:* Ist die Stress-Verbesserung innerhalb eines Iterations-Schrittes kleiner als 1 % (oder sogar negativ, was vorkommen kann), dann wird die Iteration etwas modifiziert wiederholt. Insgesamt darf das in Almo 5 Mal geschehen. Nach 5 Wiederholungen, die nicht zu einer "gelungenen Iteration" geführt habe, greift das Programm zurück auf eine frühere gelungene Iteration als neue Ausgangs-Konfiguration. Wiederholungen sind also gewissermaßen Iterationen 2. Ordnung innerhalb eines Iterationsschrittes. Meldet Almo sehr viele Wiederholungen, dann ist das ein Zeichen dafür, dass der Iterationsprozess sich erschöpft hat, d.h. eine weitere Stress-Verbesserungen nicht mehr zu erwarten ist.

Almo bringt dann noch eine Warnung, dass die vorgegebene Stress-Schwelle nicht erreicht wurde. Klickt der Benutzer auf den Hilfeknopf (in der Almo-Ausgabe, nicht hier), dann wird ihm, kurz zusammengefasst, empfohlen, das erzielte Ergebnis als das best mögliche zu akzeptieren. In folgendem Abschnitt wird das ausführlich behandelt.

Die gefundene Ladungsmatrix wird nun noch grafisch dargestellt:



Gegenüber der Grafik aus der Analyse, die nur einen Stress-Koeffizienten von 0.0373 erzielt hat, sind nun doch einige Unterschiede sichtbar geworden. Die 3 Cluster

Konservative	Kommunisten	Arbeiterpartei
Christliche	Sozialisten	Liberales
Zentrum		

sind unverändert klar voneinander getrennt. Konservative und Christliche sind zu einem Punkt verschmolzen. Dagegen sind jetzt Kommunisten und Sozialisten stärker auseinander gerückt.

Prog34mb gibt noch eine Vielzahl weiterer Ergebnisse aus. Sie werden in Abschnitt P34.2.5 im Detail noch gezeigt.

### P34.3.3 Analyse mit 3 Dimensionen mit Prog34mb

Abschliessend wird noch eine Analyse mit 3 Dimensionen gerechnet und wieder mit Zufallswerten als Startwerte, einer Stress-Schwelle von 0 und maximal 30 Iterationen Folgendes Ergebnis (gekürzt) entsteht:

Es wurden 5 Lösungsversuche mit verschiedenen Zufalls-Startzahlen gerechnet Den besten Stress-1- Wert erbrachte: Versuch 5

Ladungsmatrix der Objekte

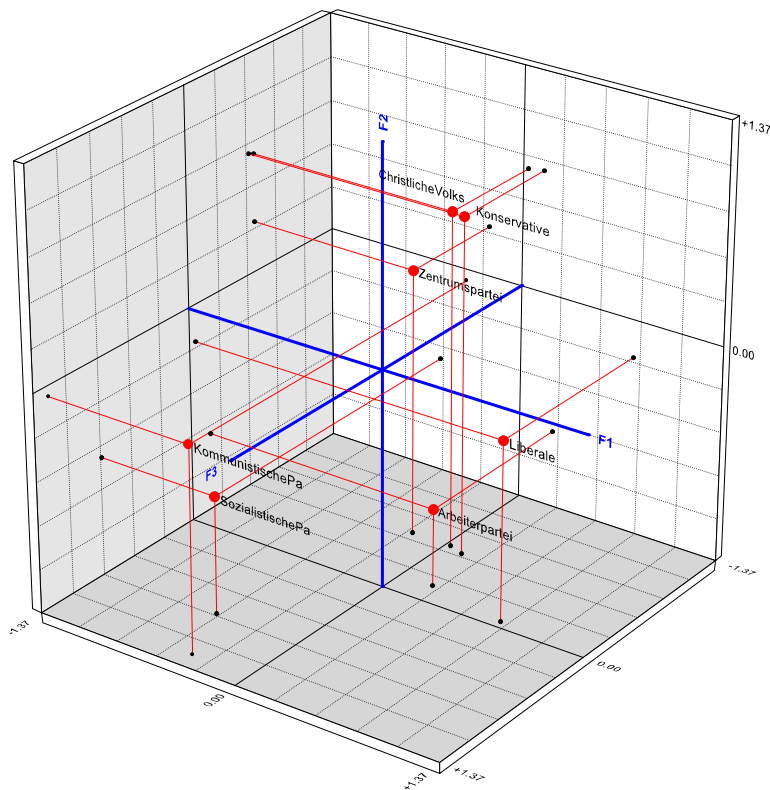
		Faktor 1	Faktor 2	Faktor 3
Kommun	V1	-0.3892	-0.0702	1.2453
Sozial	V2	-0.5654	-0.6281	0.8084
Arbeit	V3	0.2316	-0.8775	-0.1712
Libera	V4	0.7779	-0.2361	-0.0578
Zentru	V5	-0.2312	0.3135	-0.6216
Christ	V6	0.0318	0.7409	-0.6259
Konser	V7	0.1446	0.7575	-0.5773

Stress-1 0.001880  
 Anpassungsindex 0.999974  
 Versuch Nr. 1  
 Iterationen 31  
 Startwert f. Zufallsgenerator 1693344163

Stress-1 nach Kruskal je Iteration		
Iteration	Stress-1	Stress-Minderung in %
1	0.274272	-
2	0.200147	27.03
3	0.152488	23.81
4	0.116379	23.68
5	0.105071	9.72
6	0.074796	28.81
7	0.032930	55.97
8	0.017822	45.88
.	.	.
.	.	.
27	0.001979	1.97
28	0.001951	1.42
29	0.001924	1.35
30	0.001902	1.13
31	0.001880	1.19

Die Zahl zulaessiger Iterationen (30) wurde ueberschritten  
Die vorgegebene Stress-Schwelle wurde nicht erreicht

Almo zeichnet folgendes 3-dimensionale Punktediagramm



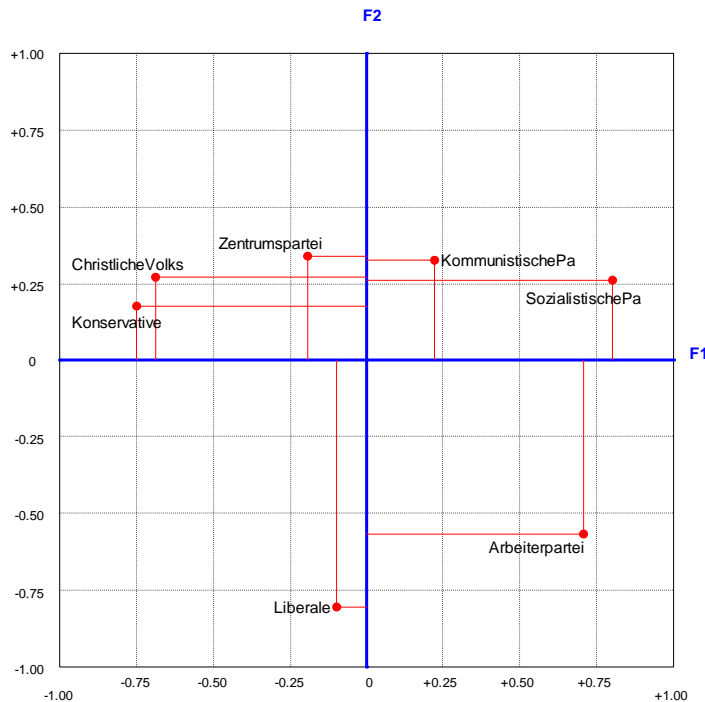
Prog34mb gibt noch eine Vielzahl weiterer Ergebnisse aus. Sie werden in Abschnitt P34.2.5 im Detail noch gezeigt.

Die 3 Cluster, die oben bei der 2-dimensionalen Analyse identifiziert wurden, können auch hier ungefähr wieder erkannt werden.

Interessant ist es nun, die Projektion der Punkte auf die hintere Wand zu betrachten und zu vergleichen mit dem oben abgebildeten 2-dimensionalen Punktediagramm in Abschnitt P34.2.2. Auf der hinteren Wand ist der 2-dimensionale "Sub-Raum" F1-F2 des



3-dimensionalen Raums F1-F2-F3 abgebildet.



Im Almo-Grafik-Editor wurde durch Klick auf "Diverse Position" oder (mit gleichem Effekt) auf "Anderer Grafiktyp / 2-dim-Koordin.system" die Projektion der Punkte auf die hintere Wand als eigene Grafik erzeugt. Dann muss noch F2 gespiegelt werden und der Schieber "Rotieren" bewegt werden. Danach entsteht obige 2-dimensionale Grafik. Sie ist gegenüber dem Punkte-Abbild auf die hintere Wand des 3-dimensionalen Würfels um F2 gespiegelt und im Uhrzeigersinne gedreht (rotiert). Die Punktekonfiguration bleibt dadurch unverändert. Gegenüber der Grafik aus der 2-Dimensionen-Analyse aus Abschnitt P34.2.2 ist obige Grafik schwach ähnlich. Das ist ein bedeutsamer Unterschied zur metrischen MDS. Dort ist die Punktekonfiguration aus einer Analyse mit 2 Dimensionen identisch mit der 2-dimensionalen "Sub-Konfiguration" aus einer Analyse mit 3 Dimensionen. Bei der iterativen MDS ist das nicht der Fall. Wir erkennen hier ein Problem, auf das wir im Exkurs ... zurückkommen werden: Die iterative MDS minimiert den Stress, es gelingt ihr aber nicht die räumliche Struktur der Vergleichs-Objekte korrekt zu rekonstruieren. Wir werden diese Kritik noch einengen auf den Fall, dass die Objekte quantitativ (intervall-skaliert) sind und die Metrik die euklidische ist.

### P34.3.4 Welche Dimensionszahl ist die richtige ?

Die Stress-Koeffizienten der 3 Lösungen sind folgende:

```
0.043276 bei 1 Faktor
0.000012 bei 2 Faktoren
0.001880 bei 3 Faktoren
```

Alle 3 Stresswerte wurden mit Prog34mb gerechnet, wobei eine Stress-Schwelle von 0 vorgegeben wurde.

Der beste Stress-Wert entsteht bei 2 bzw. 3 Faktoren. Der Unterschied zur 1-Faktor-Lösung ist deutlich. Die Frage ist, ob er signifikant ist. Ein Signifikanztest (etwa entsprechend dem Rippe-Test oder anderen bei der Faktorenanalyse) ist nicht bekannt. Verdächtig ist die Verschlechterung des Stress-Koeffizienten von der 2- zur 3-Faktor-

Lösung. Zu erwarten ist eher, dass durch einen weiteren Faktor der Erklärungswert der Lösung zunimmt. Wenn das Gegenteil eintritt, muss angenommen werden, dass der weitere Faktor artifizuell ist. Auch stellt sich die Frage, ob die gewonnenen Faktoren inhaltlich interpretierbar sind. Der 1. Faktor ist sicherlich als "Links-Rechts-Dimension" zu interpretieren. Beim 2. und dem 3. Faktor ist eine Bezeichnung jedoch unklar. Auch wurde darauf hingewiesen, dass zwischen der Zahl der Objekte und der maximalen Dimensionszahl ein Verhältnis von maximal 5 zu 1 bestehen soll. Demzufolge wäre die 1-Faktor-Lösung die richtige. Werden, wie das oben gezeigt wurde, die Ladungen aus der *metrischen* MDS als Startwerte eingesetzt (das ist ja bei Prog34ma so voreingestellt und kann dort auch nicht verändert werden), dann wird vom Kalkül nur 1 Faktor errechnet und dann abgebrochen. Auch dies weist darauf hin, dass die 1-Faktor-Lösung die korrekte ist.

Ein formales Entscheidungskriterium existiert nicht. Letztendlich entscheidet der Forscher. Es muss auch bedacht werden, dass das Programm abbricht sobald eine vom Programmierer in den Kalkül hineinprogrammierte Abbruch-Bedingung greift

### **P34.3.5 Was tun wenn die vorgegebene Stress-Schwelle nicht erreicht wird.**

Dann gibt Almo die Ergebnisse nach dem letzten Iterationsschritt aus. Der Benutzer kann dann folgendermaßen reagieren:

1. Er akzeptiert den vom Programm erzielten Stress-1-Wert als den besten erreichbaren. Das ausgegebene Ergebnis wird akzeptiert.

Normalerweise möchte man einen möglichst niedrigen Stress erreichen. Man gibt dann eine sehr niedrige Stress-1-Schwelle vor, z.B. 0.01 oder sogar nur 0 und lässt dann das Programm solange iterieren bis es abbrechen muss, weil keine nennenswerte Stress-Verringerung mehr erzielt werden kann. Diesen letzten erzielten Stress-Wert wird man als besten erreichbaren akzeptieren. Die so gefundene Lösung ist korrekt. Die Meldung, dass möglicherweise ein lokales Minimum vorliegt, kann man in der Regel ignorieren.

2. In der Eingabebox "Programmparameter" wird die Stress-Schwelle auf einen höheren Wert hinaufgesetzt. Die Warnung erscheint dann nicht. Der Nachteil dieser Vorgehensweise ist, dass wahrscheinlich ein niedriger Stress-1-Wert erzielbar gewesen wäre - auf den man nunmehr verzichtet hat.

In der Literatur wird die Güte einer MDS-Lösung folgendermaßen beurteilt:

Stress-1	Goodness of fit
-----	-----
größer 0.2	nicht ausreichend
0.2 bis 0.1	angemessen
0.1 bis 0.05	gut
kleiner 0.05	sehr gut

Wir würden etwas strenger urteilen und bereits Stress-Werte über 0.1 skeptisch betrachten.

3. Die Zahl der Dimensionen wird um 1 erhöht, z.B. von 2 auf 3.
4. Die Optionsbox "Startwerte" wird geöffnet und ein anderes

Startwerte-Verfahren gewählt. Das beste Ergebnis erzielt man erfahrungsgemäß mit "Ergebnisse aus metrischer MDS als Startwerte".

Dies ist allerdings ohnehin voreingestellt.

- Die Optionsbox "weitere Optionen" wird geöffnet und dort eine oder mehrere der Optionen 6, 7, 8, 34 eingesetzt.

Bei Programm-Maske Prog34ma ist dies nicht möglich.

Wir empfehlen folgende Vorgehensweise:

Man gibt eine sehr niedrige Stress-1-Schwelle vor, z.B. 0.01 oder sogar nur 0 und lässt dann das Programm solange iterieren bis es abbrechen muss, weil keine nennenswerte Stress-Verringerung mehr erzielt werden kann. Diesen letzten erzielten Stress-Wert wird man als besten erreichbaren akzeptieren.

### P34.3.6 Erweiterte Ausgabe bei Prog34mb

Bei Prog34mb (mit Optionen) werden nach der oben gezeigten grafischen Darstellung der Ladungsmatrix weitere Ergebnisse ausgegeben.

Wir rechnen zeigen diese Ergebnisse aus einer Analyse mit folgenden Einstellungen

Programmparameter

<input type="text" value="2"/>	Zahl der Dimensionen des gesuchten Raums	Hilfe
<input type="text" value="2"/>	Metriken des Raums 1=City-Block-Metrik 2=euklidische Metrik (empfohlen)	Hilfe
<input type="text" value="0.05"/>	Stress. Abbruchkriterium	Hilfe Hilfe

Loesche wieder diese Box (dann Voreinstellungen wieder gueltig)

Option: Startwerte

Hilfe

0 = Zufallswerte als Startwerte einsetzen  
1 = vom Benutzer definierte Startwerte aus Datei einlesen  
2 = metrische MDS rechnen. Deren Ergebnisse als Startwerte einsetzen (Voreinstellung)

bei "1" die Startwerte werden aus dieser Datei eingelesen

#### Einstellungen

Matrix: Unaehnlichkeitsmatrix

Messniveau: ordinal

Dimensionen: 2

Metrik: 2

Startkonfiguration: Zufallswerte

Zahl der Lösungs-Versuche: 5

Stress-Schwelle: 0.05

geforderte Stress-Verbesserung von einer Iteration zu naechsten: 1 %

geforderte Stress-Verbesserung zwischen 5 Iterationen: 5 %

maximale Zahl der Iterationen: 30

### P34.3.6.1 Berechnungsschema, reproduzierte Distanzmatrix und Disparitätenmatrix

Wir wollen die erweiterte Ausgabe für eine Analyse für Prog34mb zeigen, bei der 2 Dimensionen und die euklidische Metrik vorgegeben wurden. Gerechnet wurden mit Zufallswerten als Startwerten.

Letztes abschliessendes Berechnungsschema  
 fuer Distanzen, Disparitaeten und Stress-1

-----  
 P = Nummer des Punktes in den nachfolgenden Streudiagrammen  
 R = Rang des Objektpaares ij  
 i = Objekt i (=Zeile der Matrix)  
 j = Objekt j (=Spalte der Matrix)  
 u = empirische Unaehnlichkeiten  
 d = "Distanzen" (aus Ladungsmatrix der Objekte berechnet)  
 dd= "Disparitaeten" (monoton transformierte Unaehnlichkeiten)  
 \*\*= Symbol fuer Potenzieren

P	R	i	j	u	d	dd	(d-dd)**2	d**2
21	1	7	6	7.3000	0.0904	0.0904	0.0000	0.0082
15	2	6	5	7.6000	0.2762	0.1610	0.0133	0.0763
20	3	7	5	8.1000	0.1881	0.1610	0.0007	0.0354
1	4	2	1	8.7000	0.0186	0.1610	0.0203	0.0003
6	5	4	3	10.0000	0.3904	0.3904	0.0000	0.1524
10	6	5	4	10.5000	1.2962	1.2962	0.0000	1.6802
19	7	7	4	12.9000	1.4638	1.4638	0.0000	2.1427
3	8	3	2	14.8000	1.6548	1.5902	0.0042	2.7385
9	9	5	3	17.8000	1.5626	1.5902	0.0008	2.4417
14	10	6	4	18.9000	1.5530	1.5902	0.0014	2.4120
5	11	4	2	19.0000	1.8495	1.7499	0.0099	3.4206
13	12	6	3	21.3000	1.7952	1.7499	0.0021	3.2229
18	13	7	3	24.3000	1.7103	1.7499	0.0016	2.9250
2	14	3	1	25.3000	1.6446	1.7499	0.0111	2.7047
8	15	5	2	33.2000	1.9084	1.8724	0.0013	3.6421
4	16	4	1	33.7000	1.8364	1.8724	0.0013	3.3724
7	17	5	1	37.9000	1.8899	1.8899	0.0000	3.5717
17	18	7	2	40.0000	1.8973	1.8934	0.0000	3.5996
11	19	6	1	49.3000	1.9043	1.8934	0.0001	3.6264
16	20	7	1	50.2000	1.8787	1.8934	0.0002	3.5294
12	21	6	2	50.5000	1.9229	1.9229	0.0000	3.6974
							0.0682	49.0000
							=SS	=TT

Summe[(d-dd)\*\*2] = SS= 0.0682  
 Summe[ d\*\*2 ] = TT= 49.0000  
 Stress =Wurzel(SS/TT)= 0.0373

Betrachten wir die erste Zeile des Berechnungsschemas:

P	R	i	j	u	d	dd	(d-dd)**2	d**2
21	1	7	6	7.3000	0.0904	0.0904	0.0000	0.0082

Das Objektpaar i=6 mit j=7 (also *Christliche Partei* mit *Zentrum*) besitzt mit 7.3000 die kleinste Unähnlichkeit (Rangplatz R=1).

Dieses Objektpaar nimmt in der Unähnlichkeitsmatrix die Zelle 21 ein. In den folgenden Streudiagrammen wird es entsprechend als Punkt P=21 eingezeichnet.

Die Matrixzellen werden in folgender Weise durchnummeriert: Es wird nur das untere Dreieck ohne Diagonale verwendet und in folgender Weise von oben nach unten nummeriert

Objekt	A	B	C	D	E	....
A	-					

B	1	-				
C	2	3	-			
D	4	5	6	-		
E	7	8	9	-		
F	10	11	12	13	-	
.	.	.	.	.	.	.
.	.	.	.	.	.	.

Im Streudiagramm entspricht also z.B. der Punkt 2 der Distanz zwischen Objekt C und A, der Punkt 13 der Distanz zwischen F und D

Aus der gefundenen Ladungsmatrix wird die Distanz  $d=0.0904$  errechnet. Siehe Zelle 21 (6,7) in nachfolgender Distanzmatrix. Die Disparität ist für dieses Objektpaar mit  $dd=0.0904$  gleich groß.

Distanzen  
(vom Modell der MDS aus der Ladungsmatrix reproduziert)

		Kommuni	Soziali	Arbeite	Liberal	Zentrum	Christl	Konserv
		V1	V2	V3	V4	V5	V6	V7
Kommun	V1	0	0.0186	1.6446	1.8364	1.8899	1.9043	1.8787
Sozial	V2	0.0186	0	1.6548	1.8495	1.9084	1.9229	1.8973
Arbeit	V3	1.6446	1.6548	0	0.3904	1.5626	1.7952	1.7103
Libera	V4	1.8364	1.8495	0.3904	0	1.2962	1.5530	1.4638
Zentru	V5	1.8899	1.9084	1.5626	1.2962	0	0.2762	0.1881
Christ	V6	1.9043	1.9229	1.7952	1.5530	0.2762	0	0.0904
Konser	V7	1.8787	1.8973	1.7103	1.4638	0.1881	0.0904	0

Disparitäten

		Kommuni	Soziali	Arbeite	Liberal	Zentrum	Christl	Konserv
		V1	V2	V3	V4	V5	V6	V7
Kommun	V1	0	0.1610	1.7499	1.8724	1.8899	1.8934	1.8934
Sozial	V2	0.1610	0	1.5902	1.7499	1.8724	1.9229	1.8934
Arbeit	V3	1.7499	1.5902	0	0.3904	1.5902	1.7499	1.7499
Libera	V4	1.8724	1.7499	0.3904	0	1.2962	1.5902	1.4638
Zentru	V5	1.8899	1.8724	1.5902	1.2962	0	0.1610	0.1610
Christ	V6	1.8934	1.9229	1.7499	1.5902	0.1610	0	0.0904
Konser	V7	1.8934	1.8934	1.7499	1.4638	0.1610	0.0904	0

Almo gibt dann noch den Stress-2 (nach Kruskal) aus

Stress-2 nach Kruskal

Stress je Zeile

V1	Kommunistische	=	0.083023
V2	Sozialistische	=	0.085582
V3	Arbeiterpartei	=	0.078039
V4	Liberales	=	0.063467
V5	Zentrumspartei	=	0.061847
V6	ChristlicheVol	=	0.058515
V7	Konservative	=	0.023048

Stress-2 pauschal = 0.067734

Dieser Stresskoeffizient ermöglicht es festzustellen, ob und welches Objekt für einen schlechten Stress-1-Koeffizienten verantwortlich ist. Der pauschale Stress-2-Koeffizient ist der Mittelwert aus den einzelnen Werten je Objekt. Er ist ungefähr doppelt so groß wie der Stress-1-Koeffizient.

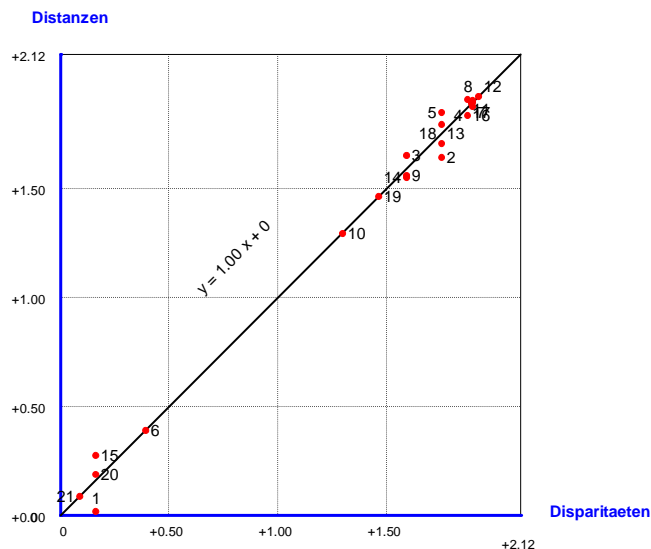
Unter die reproduzierte Distanzmatrix haben wir die Disparitätenmatrix geschrieben. Die beiden differieren nur wenig. Die Güte der MDS-Schätzung wird offenkundig, wenn nun die Distanzen und die Disparitäten miteinander korreliert werden und in einem Streudiagramm gezeichnet werden.

### P34.3.6.2 Streudiagramme der Unähnlichkeiten, Disparitäten und Distanzen

Es werden 3 Streudiagramme ausgegeben.

1. Streudiagramm der "Disparitäten" gegen "Distanzen"
2. Streudiagramm der "Unähnlichkeiten" gegen "Disparitäten"
3. Streudiagramm der "Unähnlichkeiten" gegen "Distanzen"

#### 1. Streudiagramm der Disparitäten und reproduzierten Distanzen



Regressionsgleichung:  $y = 1.000 * x - 0.000$   
 Korrelation zwischen Disparitäten und Distanzen  $r = 0.9965$   
 $r$ -Quadrat (Anpassungsindex)  $= 0.9930$

Was ein Punkt repräsentiert, z.B. der Punkt 15, kann aus dem oben abgebildeten "letzten abschliessenden Berechnungsschema" entnommen werden.

P	R	i	j	u	d	dd	(d-dd)**2	d**2
15	2	6	5	7.6000	0.2762	0.1610	0.0133	0.0763

Der Punkt P=15 entspricht dem Objektpaar i=6 (Christliche Partei) versus j=5 (Zentrum). Die vom MDS-Modell reproduzierte Distanz zwischen den beiden beträgt d=0.2762 (das ist der Ordinatenwert des Punktes) und die Disparität dd=0.1610 (das ist der Abszissenwert des Punktes).

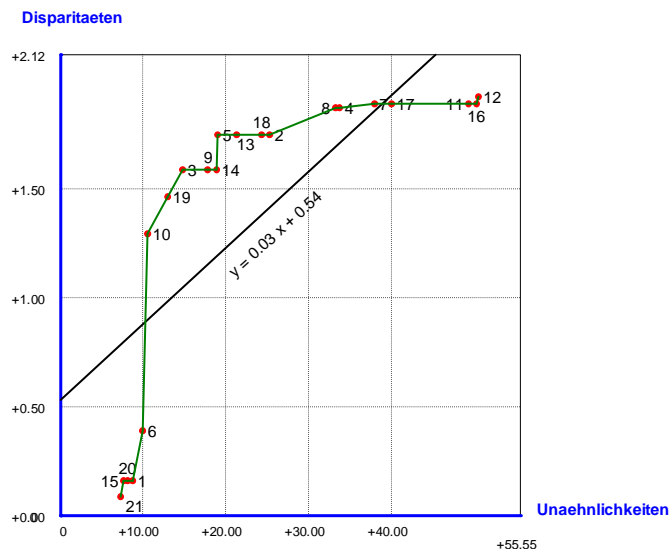
Die beiden Zahlenreihen d und dd korrelieren sehr hoch mit 0.9965. Der quadrierte Korrelationskoeffizient wird auch "Anpassungsindex" genannt. In das Diagramm wurde eine Regressionsgerade eingezeichnet, für die folgende lineare Gleichung gefunden wurde.

$$\text{Distanz} = 1 * \text{Disparität} + 0$$

Bei der nicht-metrischen MDS liefert dieses Streudiagramm eine zusätzliche optische Information. Je dichter die Punkte bei der Regressionsgeraden liegen, umso besser ist es

gelingen, die monoton transformierten Unähnlichkeiten (=die Disparitäten) und die vom Modell reproduzierten Distanzen aneinander anzupassen. Die Reproduktion ist perfekt, wenn alle Punkte auf der Regressionsgeraden liegen. Der quadrierte Korrelationskoeffizient wird auch "Anpassungsindex" genannt. Er ist gewissermaßen das Gegenstück zum Stress-Koeffizienten. Er ist 1.0 wenn der Stress-Koeffizient 0.0 ist.

## 2. Streudiagramm der "Unähnlichkeiten" gegen "Disparitäten"



Regressionsgleichung:  $y = 0,035 * x + 0,536$   
 Korrelation zwischen Unähnlichkeiten und Disparitäten  $r = 0,7430$   
 $r$ -Quadrat =  $0,5521$

### Ein Punkt im Diagramm besitzt als Koordinatenwerte

auf der x-Achse: die empirische Unähnlichkeit zwischen Objekt i und j  
 auf der y-Achse: die Disparität, d.h. die monoton transformierte Unähnlichkeit zwischen Objekt i und j

Welches Objektpaar welcher Punktenummer entspricht, kann im Almo-Ergebnis aus dem "letzten abschliessenden Berechnungsschema" wie bereits oben beim 1. Streudiagramm gezeigt, abgelesen werden.

Im Diagramm werden die Punkte in aufsteigender Rangordnung durch Linien verbunden. Dadurch entsteht die "Shepard-Kurve" (nach R. Shepard, einem der Pioniere der MDS). Im "Shepard-Diagramm", das als 3. Streudiagramm in der Ausgabe gezeichnet wird, wird die Streuung der Distanzen um diese "Shepard-Kurve" gezeigt.

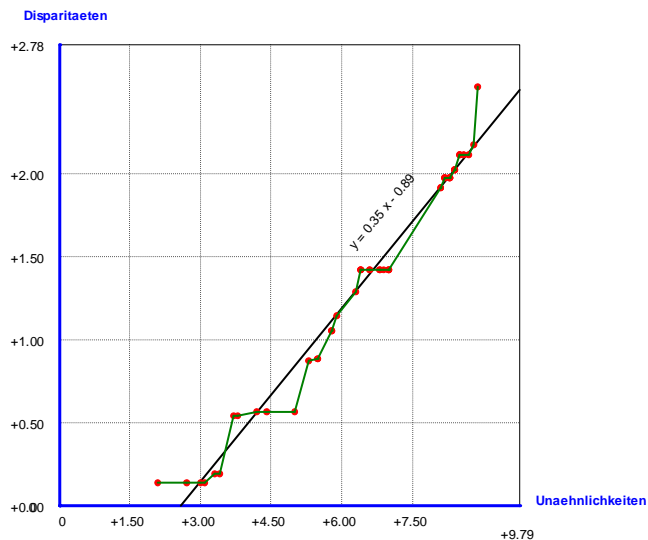
Der Zweck dieses Streudiagramms ist ein doppelter:

a. Wenn die monotone Transformation gelungen ist, dann muss für alle Punkte gelten, dass bei zunehmender Unähnlichkeit auch die Disparitäten zunehmen oder zumindest gleich bleiben. Bei gleicher Unähnlichkeit dürfen die Disparitäten gleich aber auch größer oder kleiner sein. Das wird "schwache Monotonie" genannt. Die Disparitäten sind also "schwach monotone" Transformationen der empirischen Unähnlichkeiten.

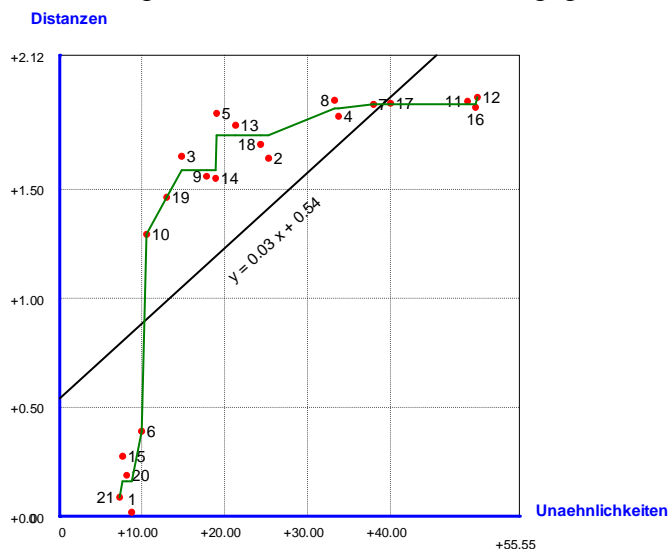
b. Die "Shepard-Kurve" kann sich mehr oder weniger eng um die Regressionsgerade "herumschlängeln". Daraus kann ersichtlich werden, ob sich der monotone Zusammenhang

zwischen Unabhängigkeiten und Disparitäten einem learen annähert (womit eventuell eine metrische MDS gerechtfertigt wäre).

In unserem Beispiel ist dies nicht der Fall. Regressionsgerade und "Shepard-Kurve" verlaufen getrennt. Anders ist dies beim Beispiel mit den 10 Automarken. Siehe dazu das Maskenprogramm ".\Almo\_Bsp\Automarken2.Alm". Es wird gefunden durch Klick auf das Menü "Almo/Liste aller Almo-Programme", dann "Automarken2.Alm". Bei diesem Beispiel entsteht folgendes Diagramm, bei dem die Shepard-Kurve sich eng um die Regressionsgerade "schlängelt".



### 3. Streudiagramm der "Unähnlichkeiten" gegen "Distanzen" (Shepard-Diagramm)



Regressionsgleichung:  $y = 0.035 * x + 0.543$   
 Korrelation zwischen Unähnlichkeiten und Distanzen  $r = 0.7345$   
 $r$ -Quadrat =  $0.5395$

Ein Punkt im Diagramm besitzt als Koordinatenwerte

auf der x-Achse: die empirische Unähnlichkeit zwischen Objekt  $i$  und  $j$



auf der y-Achse: die vom MDS\_Modell reproduzierte (aus der endgültigen Ladungsmatrix errechnete) Distanz zwischen Objekt i und j

Welches Objektpaar welcher Punktnummer entspricht, kann im Almo-Ergebnis aus dem "letzten abschliessenden Berechnungsschema" wie bereits oben beim 1. Streudiagramm gezeigt, abgelesen werden.

Im Diagramm werden neben den Punkten noch die Regressionsgerade und die "Shepard-Kurve" gezeigt.

Ein Punkt xy auf der Shepard-Kurve ist so zu interpretieren:

Der Abszissenwert x ist die empirische Unähnlichkeit zwischen den Objekten i und j. Der Ordinatenwert y ist die Disparität, d.h. die zulässige monoton transformierte Unähnlichkeit zwischen i und j

Ein Punkt des Streudiagramms (beispielsweise der Punkt 1) mit dem Abszissenwert x und dem Ordinatenwert y zeigt welche Distanz y das MDS-Modell für das Objektpaar ij reproduziert, wenn die tatsächliche empirische Unähnlichkeit den Wert x besitzt. (Da Distanzen und Disparitäten gleich skaliert sind, können sie gemeinsam im Diagramm abgebildet werden).

Liegt ein Punkt des Streudiagramms direkt auf der Shepard-Kurve, dann ist es gelungen, diesen Punkt exakt aus dem Modell der MDS zu reproduzieren. Liegt er oberhalb oder unterhalb der Shepard-Kurve dann ist die Reproduktion fehlerhaft. Wird der Punkt des Streudiagramms vertikal auf die Shepard-Kurve projiziert, dann entspricht die Länge dieser Projektionsstrecke dem Reproduktionsfehler. Der Fehler wurde in der Almo-Ausgabe des "letzten abschliessenden Berechnungsschemas" bezeichnet mit

d-dd

Aus dieser Differenz wird letztendlich der Stress-Koeffizient berechnet.

Das Ausmaß der Streuung der Punkte um die Shepard-Kurve ist somit ein grafischer Hinweis darauf, wie gut oder schlecht es gelungen ist durch das nicht-metrische MDS-Modell die Dimensionalität der Objekt-Unähnlichkeiten zu entdecken.

Die in das Diagramm noch eingezeichnete Regressionsgerade ist nur von geringer Relevanz. Man kann folgendes erkennen: Die "Shepard-Kurve" kann sich mehr oder weniger eng um die Regressionsgerade "herumschlängeln". Daraus kann ersichtlich werden, ob sich der monotone Zusammenhang zwischen Unabhängigkeiten und Disparitäten einem leeren annähert (womit eventuell eine metrische MDS gerechtfertigt wäre). Die Regressionsgerade kann im Grafik-Editor gelöscht werden. Klicken Sie auf den großen Grafik-Knopf. Damit gelangen Sie in den Grafik-Editor. Auf der rechten Werkzeugleiste oben muss der Eintrag "Regressionsgerade" geköscht werden. Danach in der linken Werkzeug-Leiste oben Klick auf den Knopf "Einsetzen".

### **P34.3.6.3 Idealpunkte**

In Abschnitt P34.1.5.1 wurde die Eingabe in die Optionsbox "Idealpunkte" beschrieben. Hier sollen nun die Ergebnisse erläutert werden, die aus Prog34mb ausgegeben werden.

Wir wollen nochmals kurz definieren: Idealpunkt ist der Punkt, den die Person in dem durch die Objekte vorgegebenen Raum einnimmt.

Eine Person äußert gegenüber den Objekten bestimmte Präferenzen bzw. Distanzen. Diese Person soll nun als Punkt in den Raum gestellt werden, in dem sich bereits die Objekte befinden. Dieser Punkt wird "Idealpunkt" genannt. Dieser etwas seltsam anmutende Begriff ist so zu verstehen:

Befindet sich unter den Objekten eines, zu dem die Person eine Distanz von 0 hat, dann ist dieses für die Person ein ideales Objekt. Das ist dann auch der Punkt der Person im Raum. Es ist für die Berechnung des Idealpunktes jedoch nicht notwendig, dass ein solches Objekt existiert - obwohl dieser Fall durchaus auftreten kann. Allgemein gilt.

Gesucht wird der Punkt im Raum *wäre dort ein Objekt* zu dem die Person eine Distanz von 0 hat. Das wäre für die Person ein ideales Objekt. Das ist ihr Idealpunkt.

In die vom Programm berechnete Punktekonfiguration der Objekte werden nun in einem 2. Schritt die Personen abgebildet, von denen wir ihre Präferenzen bzw. Distanzen gegenüber den 7 Parteien kennen. Wir wollen das Vorgehen für 3 Personen P1, P2, P3 zeigen. Die Zahl der Personen ist in Almo nicht beschränkt

Objekt	Personen		
	P1	P2	P3
KommunistischeP	0	3	6
SozialistischeP	1	2	5
Arbeiterpartei	2	1	5
Liberale	3	4	4
Zentrumspartei	4	4	3
ChristlicheVolk	5	4	2
Konservative	6	5	1

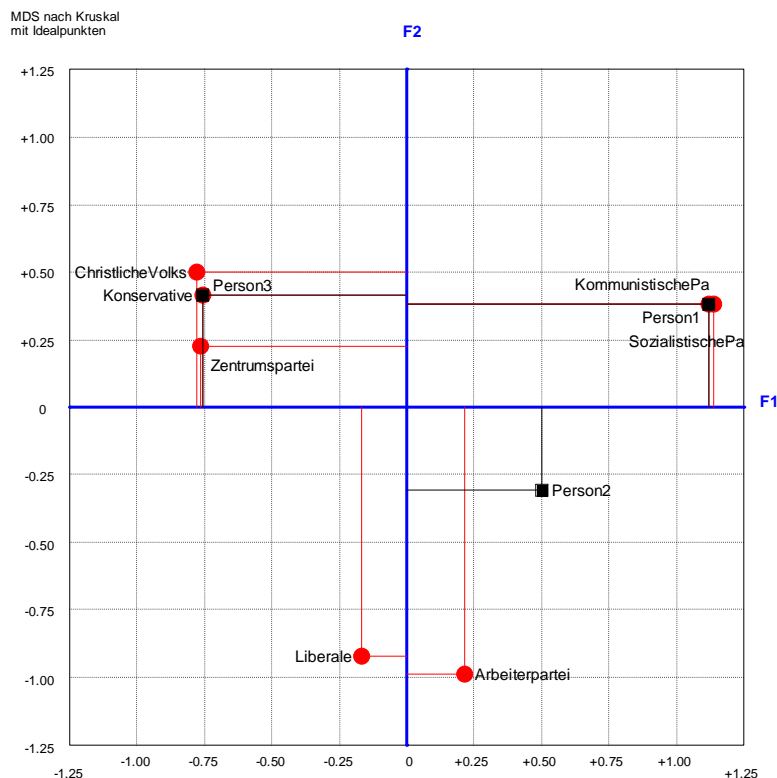
Die Zahlenwerte bezeichnen Distanzen. Die Person P1 hat eine Distanz von 0 zur Kommunistischen Partei und eine maximale Distanz zu den Konservativen. P3 ist fast das exakte Gegenteil zu P1. Nur bei den Sozialisten und der Arbeiterpartei unterscheidet P3 nicht. Zu beiden beträgt die Distanz 5 Einheiten. Die Zahlenwerte sind ordinal, d.h. es handelt sich um Rangziffern, nicht um quantitative Ziffern. Ihre absolute Größe ist irrelevant.

Almo berechnet nun die Ladungen der 3 Personen für die 2 Dimensionen. Wir hängen die Ladungen an die Ladungsmatrix der Parteien an

Gesuchte Punktekonfiguration

		Faktor 1	Faktor 2
Kommunis	V1	1.1194	0.3844
Sozialis	V2	1.1379	0.3844
Arbeiter	V3	0.2160	-0.9899
Liberale	V4	-0.1689	-0.9243
Zentrums	V5	-0.7640	0.2273
Christli	V6	-0.7813	0.5029
Konserva	V7	-0.7591	0.4153
Person1		1.1194	0.3844
Person2		0.5025	-0.3080
Person3		-0.7591	0.4153

Die Ladungsmatrix wird von Almo als 2-dimensionales Punktediagramm dargestellt.



Da die Person1 zur Kommunistischen Partei eine Distanz von 0 hat fallen Personenpunkt und Objektpunkt zusammen. Kommunistische und sozialistische Partei sind so nahe beieinander, dass in der Grafik die beiden Punkte sich beinahe überdecken.

Die 3 Personen sind neben der Objektgruppe (die 7 Parteien) eine zweite andersartige Gruppe, die in dem Raum abgebildet werden, den die Objektgruppe aufspannt und besetzt.

Der Unterschied zwischen den beiden Gruppen ist dieser: Die Objekte werden auf ihre Ähnlichkeit bzw. Unähnlichkeit *untereinander* untersucht. Die Personen werden auf ihre Ähnlichkeit bzw. Unähnlichkeit *gegenüber* den Objekten untersucht. Zusätzlich werden dabei auch noch die Distanzen zwischen den Personen sichtbar. Dagegen muss allerdings folgendes eingewendet werden: Der Idealpunkte-Kalkül berechnet die Position einer Person  $i$  im MDS-Raum ohne die anderen Personen und ihre Präferenzen bzw. Distanzen gegenüber den Objekten zu berücksichtigen. Der Kalkül verwendet nur die Distanzen zwischen der Person  $i$  und den Objekten.

Eine Frage, die vor allem bei sozialwissenschaftlichen Untersuchung oft vergessen wird zu stellen, lautet:

Ähnlichkeit bzw. Unähnlichkeit worin ?

Betrachten wir ein Beispiel. Personen werden zuerst befragt, wie sie die *Ähnlichkeit* von Parteien in ihrer Wirtschaftspolitik sehen (etwa durch einen Paarvergleich) und danach welche *Präferenz* sie gegenüber den einzelnen Parteien empfinden. Offensichtlich sind das verschiedene Messdimensionen. Einmal geht es um *Wahrnehmung* und einmal um *Bewertung*. Für die MDS spielt diese Frage jedoch keine Rolle. Ihr ist es egal, aus was inhaltlich die metrische Distanz zwischen zwei Punkten besteht.

Wir würden urteilen, dass es zulässig ist, Daten dieser Art mit der nicht-metrischen

MDS zu analysieren. Das ist sogar die Stärke dieses Verfahrens. Jedoch sollte sich der Forscher darüber im Klaren sein, was die jeweilige Distanz inhaltlich bedeutet. So ist es z.B. notwendig, sich folgende Frage zu stellen: Die Distanzen zwischen den Personen sind nun zusätzlich auch bekannt. Was bedeuten sie inhaltlich? Drücken sie die Distanzen zwischen den Personen in ihren Sympathien gegenüber den Parteien aus oder ihre Unterschiede in der Wahrnehmung der Ähnlichkeiten der Parteien? Diese Frage stellt sich nicht beim Verfahren des multidimensionalen Unfolding MDU.

#### **P34.3.6.4 Verhältnis zum multidimensionalen Unfolding (MDU)**

Beim Verfahren des "multidimensionalen Unfolding" (MDU) entfällt die erste Befragung. Die Ähnlichkeiten der Objekte (im Beispiel: der Parteien) wird nicht erfragt. Es wird nur die 2. Befragung durchgeführt. D.h. die Personen werden zu ihren Präferenzen gegenüber den Parteien befragt. Aus dieser Information werden die Distanzen zwischen den Parteien aus dem MDU-Kalkül *gefolgert*. Sie müssen nicht empirisch erhoben werden (was auch kritisch zu betrachten ist). Im Unterschied zur nicht-metrischen MDS mit Idealpunkten, ist jedoch klar, dass sich alle errechneten Distanzen (zwischen den Objekten, zwischen den Personen und zwischen Personen und Objekten) inhaltlich auf dieselbe Variable beziehen, auf die Präferenz.

Siehe dazu das Almo-Programm zur metrischen MDU und das Almo-Handbuch 29 "Metrisches multidimensionales Unfolding".

#### **P34.3.6.5 Vektorrepräsentation der Idealpunkte**

Bei der Idealpunkt-Repräsentation wird die Person (allgemein: die externe Variable) als Punkt in den MDS-Raum gestellt. Bei der Vektorrepräsentation wird die Person als Vektor in dem berechneten Raum hineinprojiziert. In nachfolgender Grafik wird verständlich, was damit gemeint ist. Der Richtungsvektor wird dabei so bestimmt, dass die Projektionen der Objekte auf den Richtungsvektor mit den Präferenzen der Personen gegenüber den Objekten maximal korreliert. Im Programm wird die Produkt-Moment-Korrelation maximiert.

Die Vektorrepräsentation ist eine alternative Möglichkeit der Abbildung der Präferenzen einer Person. Es wird angenommen, dass der Idealpunkt einer Person außerhalb des MDS-Raums liegt und durch einen Richtungsvektor dargestellt werden kann. Die Präferenzen einer Person ergeben sich durch rechtwinkelige Projektion der Objekte auf diesem Vektor.

Für die Person 3 ermittelt ALMO folgende Ergebnisse:

Projektionen der Objekte bezueglich Kriteriumsvariable 3 (Person 3)

Objekt	emp.Wert	Projekt.	Distanz
1 KommunistischeP	6.000	0.978	0.667
2 SozialistischeP	5.000	0.996	0.672
3 Arbeiterpartei	5.000	0.470	0.897
4 Liberale	4.000	0.082	0.936
5 Zentrumspartei	3.000	-0.797	0.017
6 ChristlicheVolk	2.000	-0.887	0.278
7 Konservative	1.000	-0.842	0.199

Korrelation= 0.931

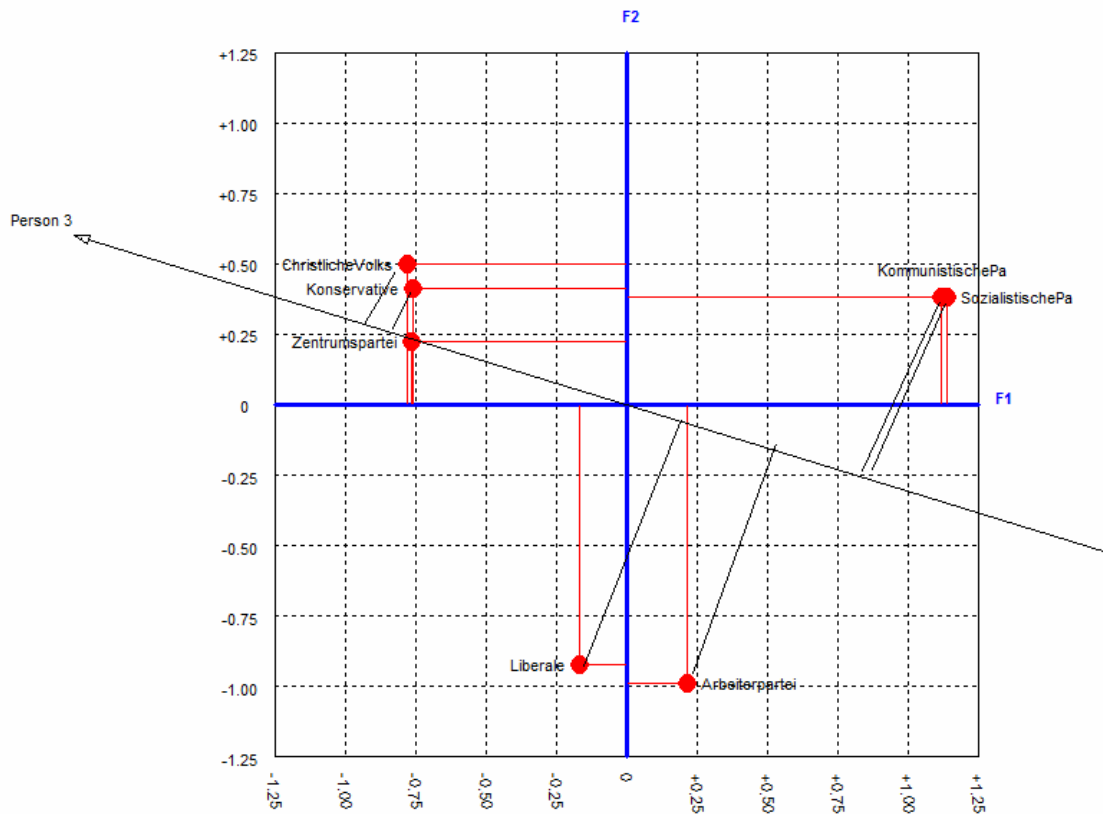
Zahl der Iterationen=5

Koordinaten des Vektors 0.96 -0.26

In nachfolgender Grafik läuft der Vektor der Person 3 durch den Ursprung 0/0 zum Punkt F1=0.96 / F2=-0.26

Durch die Vektordarstellung wird die Präferenz der Person relativ gut abgebildet. Im

Unterschied zur Idealpunktdarstellung wird angenommen, dass ein größerer Wert eine stärkere Präferenz ausdrückt. Sind die Präferenzen invers kodiert, so dass eine größere Zahl eine schwächere Präferenz ausdrückt, so sind die Vorzeichen des Richtungsvektors umzudrehen, um die Präferenzen abzubilden. Dies ist im vorliegenden Beispiel der Fall. Der Richtungsvektor lautet daher:  $(-0.96 \ 0.26)$ . Der Verlauf des Vektors in der Grafik ändert sich dadurch nicht.



Die in obiger Tabelle mit "Distanz" bezeichnete Spalte enthält die Entfernung des jeweiligen Objekt-Punktes vom Vektor der Person 3. So beträgt z.B. die Entfernung vom Punkt der liberalen Partei vom Vektor 0.936. Das ist das Maß an Präferenz, das die Person 3 gegenüber dieser Partei empfindet. Der Vektor und die Projektionslinien auf ihn werden von Almo nicht automatisch gezeichnet. Der Benutzer muss sie selbst einzeichnen. Siehe dazu Almo-Handbuch Teil 1 "Bedienungsanleitung", Abschnitt 10.1.6.

### P34.3.6.6 Projektion externer Information als Interpretationshilfe

Die Datei der Personen, die in den Raum der Objekte mit ihren Idealpunkten eingefügt werden, enthält in der Regel empirisch gewonnene Daten. Das muss jedoch nicht so sein. Dies können auch artifizielle Personen sein, die vom Forscher mit wohl überlegten Daten eingegeben werden. Sie können dann als Interpretationshilfe der MDS-Ergebnisse dienen.

Betrachten wir ein Beispiel: Es sollen drei Personen (wir nennen sie auch externe Variable oder Kriteriumsvariable) in den Raum der Objekte hineinprojiziert werden. Wir stellen die Ausprägungen der Variablen gleich in Tabellenform dar.

-3	3	1
-2	2	2
-1	1	2

```

0  4  3
1  4  4
2  4  5
3  5  6

```

Diese Tabelle muss in eine Datei eingeschrieben werden und in der Optionsbox "Idealpunkte" als Datei eingetragen werden.

Die erste Spalte ist eine artifizielle Person, die die Links-Rechts-Anordnung repräsentieren soll. Für sie soll eine Vektordarstellung berechnet werden. Die zweite und dritte Spalte sind Parteipräferenzen von zwei Personen. Für die Parteipräferenzen sollen Idealpunkte gesucht werden.

Es werden folgende Ergebnisse erzielt:

Vektorrepräsentation der Kriteriumsvariablen

-----  
Projektionen der Objekte bezüglich Kriteriumsvariable 1  
V11 Person1 (Links-Rechts-Einstufung)

	Objekt	emp.Wert	Projekt	Distanz
1	KommunistischeP	-3.000	-1.321	0.490
2	SozialistischeP	-2.000	-1.244	0.296
3	Arbeiterpartei	-1.000	-0.372	0.462
4	Liberale	0.000	0.293	0.180
5	Zentrumspartei	1.000	0.505	0.590
6	ChristlicheVolk	2.000	1.104	0.021
7	Konservative	3.000	1.036	0.120

Korrelation= 0.970  
Zahl der Iterationen= 5  
Koordinaten des Vektors: -0.91 -0.41

Für die Links-Rechts-Dimension wird eine relativ hohe Korrelation erzielt. Es ergibt sich folgende Anordnung der Objekte auf der Variablen:

links: Objekt 1 (=Kommunistische Partei) -1.321  
Rechts: Objekt 6 (=Christliche Volkspartei) 1.104

Der Richtungsvektor geht durch den Nullpunkt und hat auf der 1. Dimension einen Wert von -0.91 und auf der zweiten von -0.41.

Für die Parteipräferenzen der 2. Person (2. Spalte) ergibt sich folgendes Bild:

Idealpunktrepräsentation der Kriteriumsvariablen

-----  
Distanzen der Objekte zum Idealpunkt 2 V12 Praeferenz Person2

	Objekt	emp. Wert	Distanz	erw.Dis tanz
3	Arbeiterpartei	1.000	0.252	0.252
2	SozialistischeP	2.000	0.637	0.637
1	KommunistischeP	3.000	1.153	1.050
4	Liberale	4.000	0.946	1.050
5	Zentrumspartei	4.000	1.516	1.516
6	ChristlicheVolk	4.000	1.771	1.728
7	Konservative	5.000	1.685	1.728

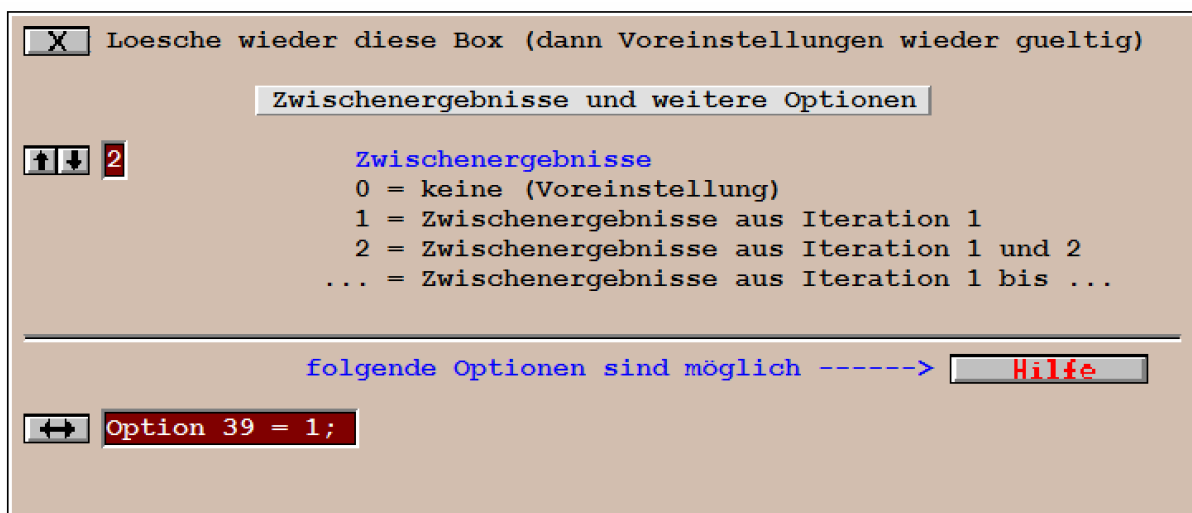
```

Stress =                0.048
Zahl der Iterationen=    2
Koordinaten des Idealpunktes:  0.39  0.64

```

### P34.3.6.7 "Punkte-Wanderung"

Wir rechnen das Beispielprogramm "Automarken2.Alm", das in Abschnitt P34.0.2 bereits vorgestellt wurde. Es wird in Almo gefunden durch Klick auf den Knopf *alle Progs* in der Knopfleiste am Oberrand des Almo-Fensters. Im Programm wird die Optionsbox "weitere Optionen" geöffnet. Wir fordern Zwischenergebnisse z.B. für die ersten 2 Iterationen an und schreiben im 2. Eingabefeld "Option 39 = 1;". Damit veranlassen wir Almo zu zeigen, wie der Punkt für Objekt 1, das ist das Auto *Opel*, im MDS-Raum durch den Kruskal-Kalkül von einer Iteration zur nächsten hin und her geschoben wird. Mit Option 39 kann für eines der Objekte dessen "Punkte-Wanderung" zahlenmäßig und grafisch gezeigt werden.



Da nach der Iteration 1 (das ist die Startkonfiguration) der vorgegebene Stress-Wert noch nicht erreicht wurde, läuft der Kruskal-Kalkül. Durch Hin- und Herschieben der Punkte wird versucht den Stress weiter zu verringern. Werden in der Programm-Maske Zwischenergebnisse angefordert, dann gibt Almo in unserem Beispiel für den Punkt 1, Opel, die Verschiebung aus. Von Iteration 1 zu 2 ist dies

neuer Koordinatenwert = alter Koordinatenwert - Schrittlänge\*Gradient

neuer Koordinatenwert für Objekt 1 Opel auf der Dimension 1  
 $-0.9343 - 3.764 * 0.011 = -0.9764$

neuer Koordinatenwert für Objekt 1 Opel auf der Dimension 2  
 $0.1033 - 3.764 * -0.018 = 0.1728$

	Faktor 1	Faktor 2
	-----	-----
aus den alten Ladungen	-0.9343	-0.9764
wird jetzt	0.1033	0.1728

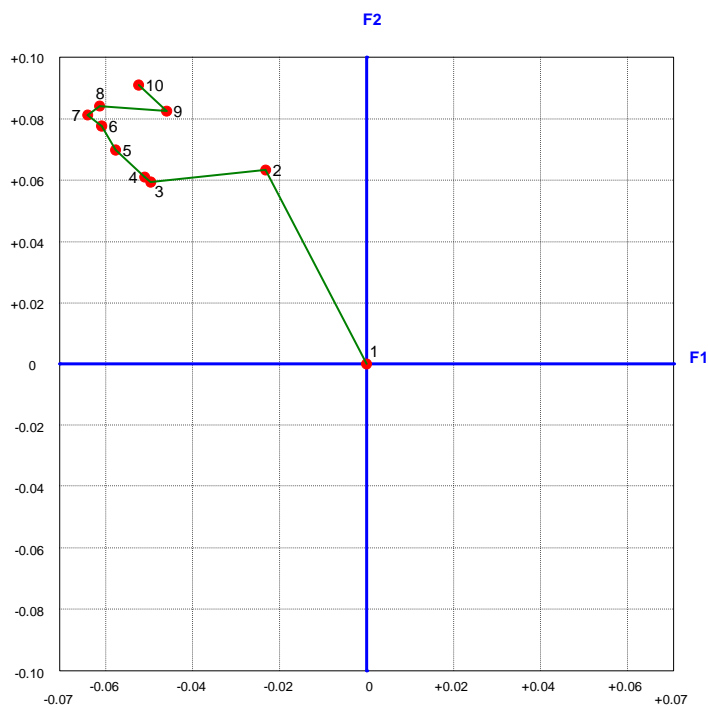
Wir wollen für Punkt 1 Opel die Verschiebungen über die Iterationen 1 bis 10 verfolgen. Dazu wird noch der jeweils erzielte Stress-1-Koeffizient und die prozentuale Stress-Minderung angegeben.

Iteration	Koordinatenwerte		Stress-1	Stress-Minderung in %
	Faktor 1	Faktor 2		
1	-0.9343	0.1033	0.094877	-
2	-0.9575	0.1695	0.073630	22.39
3	-0.9839	0.1654	0.051002	30.73
4	-0.9853	0.1672	0.050355	1.27
5	-0.9920	0.1764	0.048568	3.55
6	-0.9950	0.1845	0.047670	1.85
7	-0.9982	0.1885	0.047336	0.70
8	-0.9954	0.1915	0.046860	1.00
9	-0.9802	0.1897	0.046481	0.81
10	-0.9865	0.1986	0.046164	0.68

Für nachfolgende Grafik werden die Koordinaten der Wanderungs-Punkte so verschoben dass der Punkt für Iteration 1 im Ursprung liegt. Von allen Wanderungspunkten wird der Koordinatenwert aus der 1. Iteration subtrahiert. Damit wird die Grafik etwas gespreizt.

Iteration	Koordinatenwerte	
	Faktor 1	Faktor 2
1	0.0000	0.0000
2	-0.0232	0.0662
3	-0.0497	0.0621
4	-0.0511	0.0639
5	-0.0577	0.0731
6	-0.0607	0.0813
7	-0.0640	0.0852
8	-0.0612	0.0882
9	-0.0460	0.0864
10	-0.0523	0.0953

Punkte-Wanderung



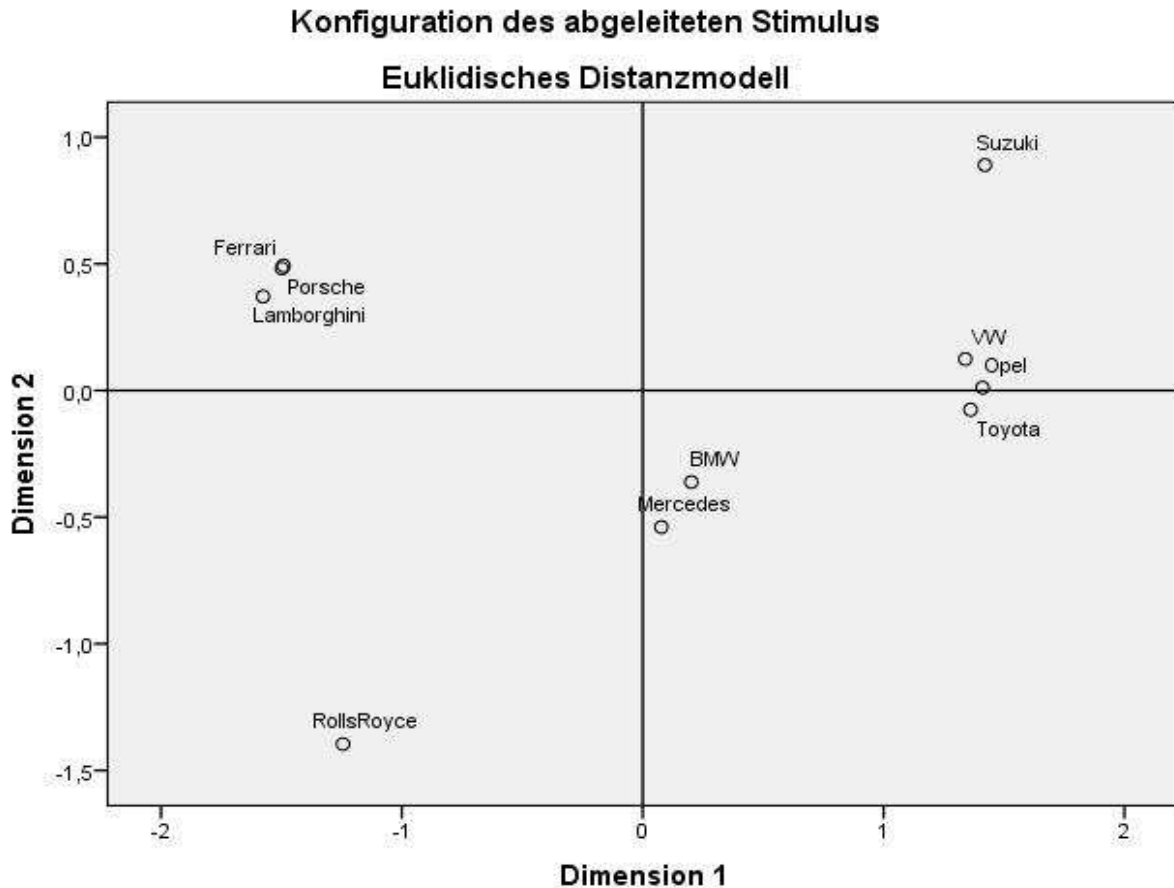
Es ist zu erkennen, dass der Punkt Opel in den ersten 7 Iterationen bei F1 nach links und bei F2 nach oben wandert. Nach Iteration 5 verbessert sich dann der Stress nur noch vernachlässigbar.



### P34.4 Vergleich Almo mit SPSS-Alscal und Proxscal

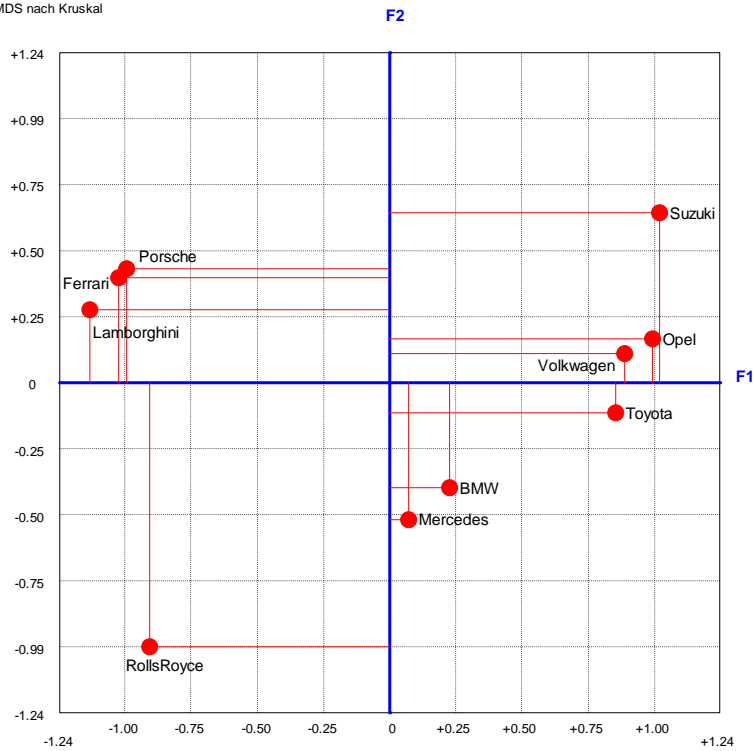
Mit Almo wurde das Programm "Automarken2.Alm" gerechnet. Man findet es durch Klick auf das Menü "Almo/Liste aller Almo-Programme". Mit SPSS wurde die Datei "Automarken\_mat.sav" gerechnet. Sie ist enthalten im Almo-Ordner TESTDAT. Der Benutzer kann auch den File "Automarken.sps" als Syntaxdatei in SPSS öffnen und ausführen. Er befindet sich ebenfalls im Ordner TESTDAT.

#### Punktediagramm aus SPSS-Alscal

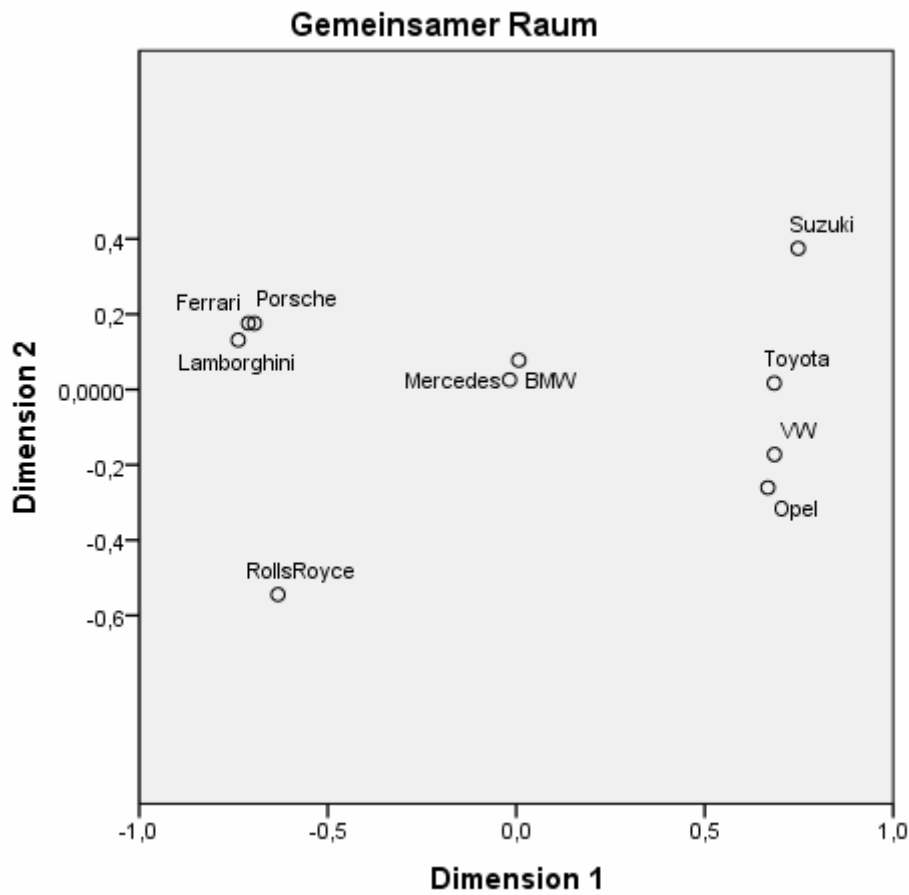


#### Punktediagramm aus Almo

MDS nach Kruskal



### Punktendiagramm aus SPSS-Proxscal



Um die Diagramme besser vergleichen zu können, wurde bei Almo im Grafik-Editor der

Faktor F1 gespiegelt (von rechts nach links gedreht). D.h. das Vorzeichen des 1. Faktors wurde umgedreht (nur für die Grafik, nicht in nachfolgender Almo-Ladungsmatrix). Das Almo-Programm Prog34mb wurde mit der Ladungsmatrix aus der metrischen MDS als Startwerte gerechnet.

Ladungsmatrix der Objekte

	ALMO		ALSCAL		PROXSCAL	
	Faktor 1	Faktor 2	Faktor 1	Faktor 2	Faktor 1	Faktor 2
Opel	-0.9898	0.1685	1,4106	,0116	.667	-.260
Volkwa	-0.8866	0.1114	1,3394	,1243	.685	-.173
Suzuki	-1.0170	0.6395	1,4208	,8904	.748	.375
Toyota	-0.8537	-0.1159	1,3607	-,0759	.685	.017
Merced	-0.0721	-0.5162	,0772	-,5391	-.018	.026
BMW	-0.2278	-0.3966	,2020	-,3611	.007	.077
Ferrar	0.9900	0.4297	-1,4915	,4927	-.711	.176
Porsch	1.0205	0.3958	-1,4980	,4821	-.694	.176
Lambor	1.1297	0.2780	-1,5766	,3709	-.737	.131
RollsR	0.9068	-0.9942	-1,2445	-1,3958	-.632	-.545
	Stress	0.049	Stress	0,0551	Stress	0.0372

Die Ladungsmatrizen sind auf den ersten Blick verschieden. Die Grafiken stimmen jedoch weitgehend überein. D.h. die Struktur der Punktekonfiguration ist weitgehend dieselbe. Das gilt generell beim Vergleich der Ergebnisse aus verschiedenen iterativen nicht-metrischen MDS-Verfahren. Wird mit dem Almo-Programm "Automarken.Alm" eine metrische MDS gerechnet, dann entsteht ebenfalls eine gut übereinstimmende Punktekonfiguration. Das Programm wird gefunden durch Klick auf das Menü "Almo/Liste aller Almo-Programme".

Kurt Holm

### **Exkurs: Vergleich metrische MDS und iterative MDS**

Wir werden im folgenden eine Simulation durchführen. Durch diese wird die Arbeitsweise der iterativen "stress-minimierenden" Verfahren sichtbar und der grundlegende Unterschied zum Verfahren der metrischen ("faktorenanalytischen") MDS offenbar.

Zur Terminologie: Wir werden im folgenden vereinfacht von

*metrischer MDS* und *iterativer MDS*

sprechen.

Die Merkmale unserer Simulation sind folgende:

1. Wir haben die Unähnlichkeitsmatrix aus dem Automarken-Beispiel aus P34.0.2 geringfügig so verändert, dass sie perfekt 3-dimensional ist. Gerechnet haben wir dabei mit 12 Kommastellen. Die doppelt zentrierte Unähnlichkeitsmatrix besitzt 3 positive Eigenwerte. Der 4. Eigenwert ist exakt 0. Die Matrix ist im Ordner TESTDAT unter dem Namen "Auto\_3dim.mat" enthalten und die SPSS-Version unter dem Namen "Auto\_3dim.sav".

2. Im Vergleich werden gerechnet:

- (1) die metrische MDS mit Almo
- (2) die iterative MDS mit Almo
- (3) die iterative MDS mit SPSS-Proxscal

(4) die iterative MDS mit SPSS-Alscal

3. Die 4 Programme werden zuerst mit 3 Faktoren (Dimensionen) und dann mit 2 Faktoren gerechnet
4. Die Simulation wird für quantitative Unähnlichkeiten bzw. Unähnlichkeiten auf dem Intervall-Niveau durchgeführt. Das ist eine bedeutsame Einschränkung. Für ordinale Variable sind nur die iterativen Verfahren verfügbar. Diese wurden ursprünglich auch nur für dieses Messniveau entwickelt.
5. Die Simulation wird nur für die euklidische Metrik durchgeführt. Sehr selten wird in empirischen Forschungen die "city-block"-Metrik eingesetzt. Alle anderen Metriken spielen in der empirischen Forschung kaum eine Rolle.

Zwischenbemerkung:

Wichtig ist in diesem Zusammenhang folgendes: Die MDS wird eingesetzt um Objekt-Unähnlichkeiten räumlich darzustellen - präziser: um sie im *euklidischen* Raum korrekt darzustellen (zahlenmäßig als Ladungsmatrix und grafisch als Punktediagramm). Die iterativen Verfahren streben danach den Stress zu minimieren. Das muss aber nicht unbedingt dazu führen, dass auch die euklidische Punktekonfiguration gefunden wird, die bei der metrischen MDS entsteht. Siehe dazu nachfolgend das mit dem iterativen Kruskal-Verfahren mit zufälligen Startwerten gerechnete Beispiel in Bild 5.

Werden ordinale Unähnlichkeiten im euklidischen Raum abgebildet, dann wird dabei unterstellt, dass die Unähnlichkeiten tatsächlich quantitativ sind, jedoch mit den gegebenen Instrumenten nur ordinal gemessen werden können; dass es jedoch durch das gewählte Analyse-Verfahren gelungen ist, die zugrunde liegende quantitative Struktur der Unähnlichkeiten aufzudecken. Für ordinal gemessene Unähnlichkeiten existieren nur die iterativen, "stress-minimierenden" Verfahren.

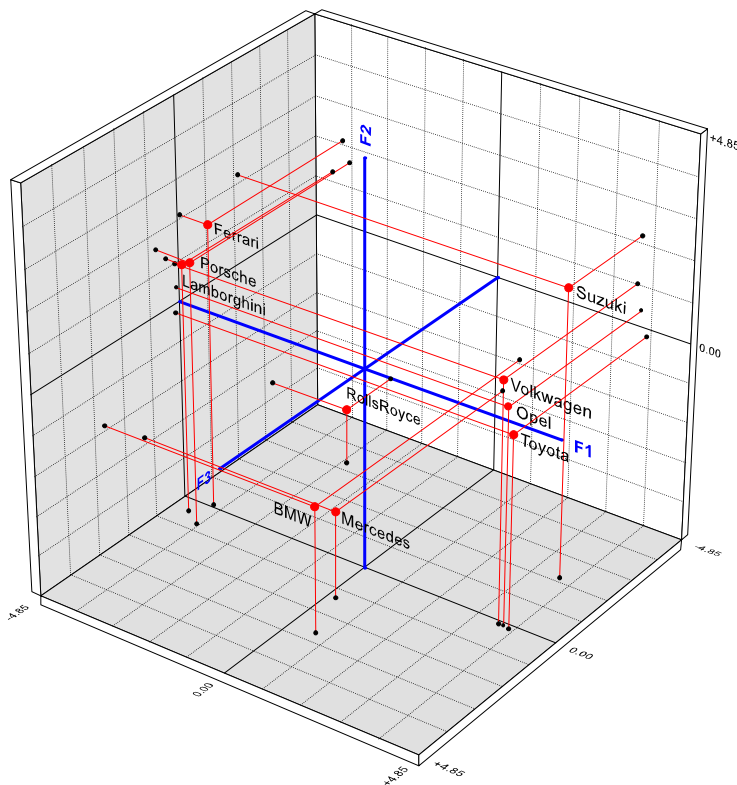
Zuerst wird eine metrische (faktorenanalytische) MDS gerechnet. Das Programm ist (durch Klick auf den Knopf *alle Progs* in der Knopfleiste am Oberrand des Almo-Fensters) unter dem Namen "metrMDS\_3dim.Alm" zu finden. Dabei entsteht folgende Ladungs-matrix für 3 Faktoren

3-dimensionale Matrix der Faktorladungen aus metrischer MDS  
(Ergebnis aus Prog "metrMDS\_3dim.Alm")

		Faktor 1	Faktor 2	Faktor 3
Opel	V1	3.6170	0.3755	0.0848
Volkwage	V2	3.5083	0.9681	0.1096
Suzuki	V3	3.5673	2.1300	-2.0885
Toyota	V4	3.7712	-0.2035	0.1224
Mercedes	V5	0.1811	-2.7555	1.2839
BMW	V6	0.6122	-1.8120	2.5856
Ferrari	V7	-4.1109	2.0314	-0.0821
Porsche	V8	-3.9372	1.5550	0.7375
Lamborgh	V9	-4.4048	1.2016	0.4196
RollsRoy	V10	-2.8042	-3.4905	-3.1727

Das Modell der metrischen MDS reproduziert exakt die empirische Distanzmatrix. Der 4. Eigenwert ist exakt 0, der Stress-Koeffizient ist ebenfalls exakt 0.00000. Das grafische Abbild der Ladungsmatrix ist folgendes:

Bild 1: 3-dimensionale Grafik aus metrischer MDS (Prog "metrMDS\_3dim.Alm")



Es ist klar ersichtlich, dass die 4 Autos Suzuki, BMW, Mercedes, Rolls-Royce in die 3. Dimension abweichen. Wird in der Programm-Maske die Faktorenzahl auf 2 reduziert, dann entstehen die in obiger 3-Faktoren-Ladungsmatrix enthaltenen Faktoren 1 und 2. Das ist bei den iterativen Verfahren nicht so. Die ersten 2 Faktoren aus einer 3-dimensionalen iterativen Analyse sind anders als die beiden Faktoren aus einer 2-dimensionalen iterativen Analyse. Sie sind auch nicht linear proportional zueinander.

Welche Ergebnisse liefern nun die iterativen MDS-Programme Proxscal,Alscal und aus Almo

### Proxscal

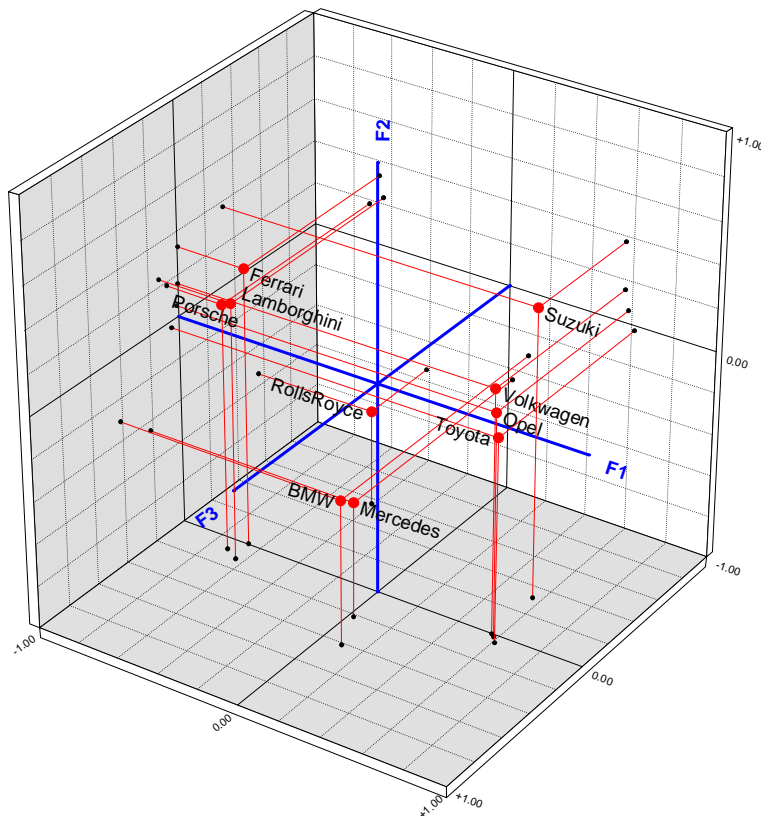
Wir rechnen zunächst mit SPSS-Proxscal die Unähnlichkeitsmatrix "Auto\_3dim.sav" (aus dem Almo-Ordner TESTDAT) mit folgenden Einstellungen:

```

/INITIAL=SIMPLEX           (in Proxscal voreingestellt)
/TRANSFORMATION=INTERVAL  (intervall-skalierte Unähnlichkeiten)
    
```

Dabei entsteht ein Stress-1 nach Kruskal von 0.012, dabei wurde der iterative Prozess nach 8 Iterationen abgebrochen, weil die Verbesserung kleiner wurde als das Konvergenzkriterium. Es entstand folgende Grafik (zum besseren Vergleich mit dem Almo-Grafik-Editor gezeichnet)

Bild 2: 3-dimensionale Grafik aus Proxscal (mit INITIAL=SIMPLEX)



Die Punktekonfiguration ist fast identisch mit der aus der metrischen MDS in Bild 1. Die 3-dimensionale Struktur der Autodaten wurde zwar nicht exakt, aber hinreichend gut durch das iterative Verfahren entdeckt.

Wird mit

```
/ INITIAL=TORGERSON
/ TRANSFORMATION=INTERVAL
```

gerechnet, dann wird als Startkonfiguration die Ladungsmatrix aus der metrischen MDS eingesetzt. Die 3-dimensionale Struktur der Auto-Daten wird sofort erkannt und ein Stress-1-Koeffizient von 0.0 ausgegeben. Das Punktediagramm ist nach Spiegelung um F2 identisch mit dem aus der metrischen MDS (Bild 1). Die Maßzahlen an den 3 Achsen sind modellbedingt anders.

Wird mit

```
/ INITIAL=RANDOM(2)
/ TRANSFORMATION=INTERVAL
```

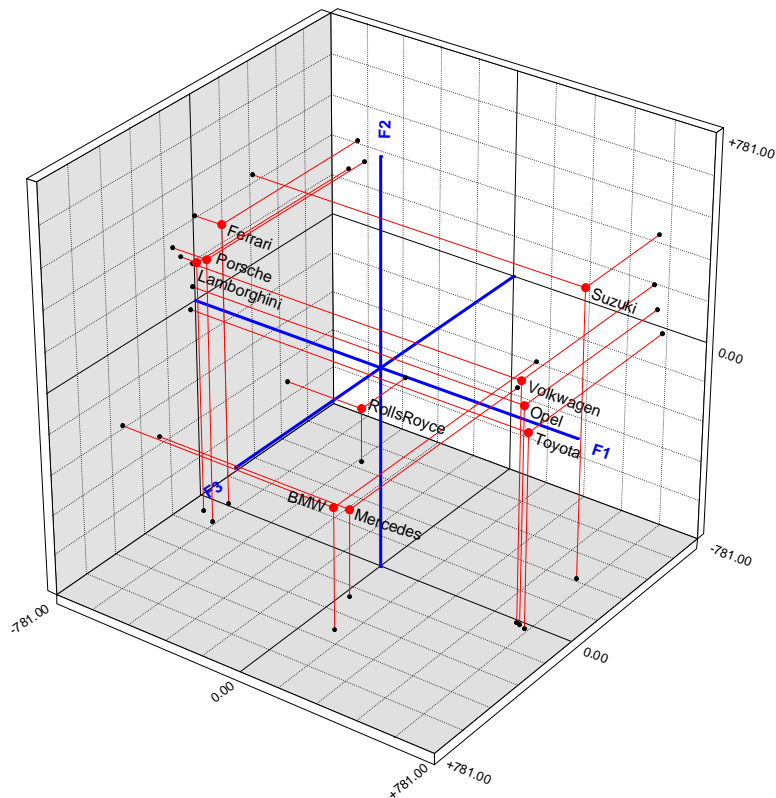
gerechnet, dann werden Zufallswerte aus 2 Startkonfigurationen eingesetzt und ein Stress-1 von 0.230 ausgegeben. Wird das gleiche Programm nochmals gestartet, dann verwendet Proxscal andere Zufallswerte und liefert einen Stress-1 von 0.0406. Dies weist auch darauf hin, dass für die iterativen MDS-Verfahren die Startwerte entscheidend sein können.

Bei

```
/ INITIAL=RANDOM(10)
/ TRANSFORMATION=INTERVAL
/ CRITERIA=DIMENSIONS(3,3) MAXITER(200) DIFFSTRESS(.00001) MINSTRESS(.00001)
```

(also maximal 10 zufälligen Startkonfigurationen) entsteht ein Stress-1 von 0.0046. Das Punktediagramm (mit Almo-Grafik-Editor gezeichnet) ist

Bild 3: 3-dimensionale Grafik aus Proxscal (mit zufälligen Startwerten)



Wird um F1, F2 und F3 gespiegelt, dann ist die grafische Punktekonfiguration fast gleich zur der der metrischen MDS in Bild 1.

### Alscal

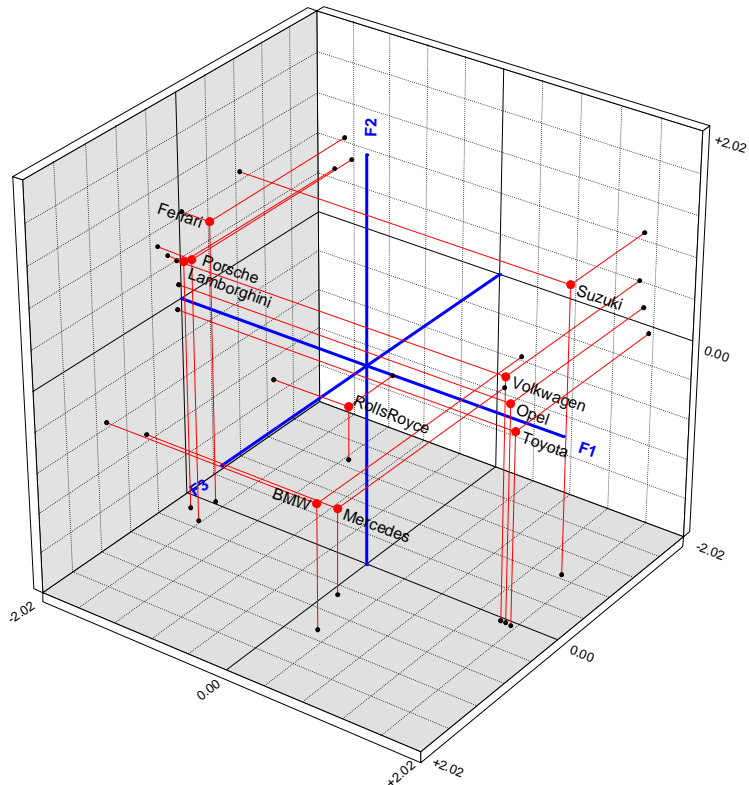
Bei SPSS-Alscal kann der Benutzer bei der dialoggesteuerten Eingabe die Startwerte nicht beeinflussen. Es ist jedoch zu vermuten, dass die Ergebnismatrix aus einer metrischen MDS verwendet wird. Folgende Parameter wurden eingegeben:

```
/LEVEL=INTERVAL
/CRITERIA=CONVERGE(0.0001) STRESSMIN(0.0005) ITER(50) CUTOFF(0) DIMENS(3,3)
```

Es entstand ein Stress-1 nach Kruskal von 0,00008.

Die grafische Punktekonfiguration ist folgende

Bild 4: 3-dimensionale Grafik aus Alscal



Nach Spiegelung um F3 ist die Punktekonfiguration fast identisch mit der aus der metrischen MDS in Bild 1.

### Almo

Wir rechnen zuerst mit dem Almo-Programm "iterMDS\_3dim.Alm". Als Startkonfiguration wird die Voreinstellung übernommen, d.h. es wird die Ladungsmatrix aus der metrischen MDS eingesetzt. Das Programm behält diese Lösung bei, transformiert sie linear (wie das der Kruskal-Kalkül erfordert) und gibt einen Stress-1-Wert von 0.0 aus.

Das Punktediagramm ist anders als oben in Bild 1. Wird um F1 und F3 gespiegelt (Vorzeichenumkehr) dann sind die beiden gleich.

Werden in "iterMDS\_3dim.Alm" folgende Einstellungen gewählt:

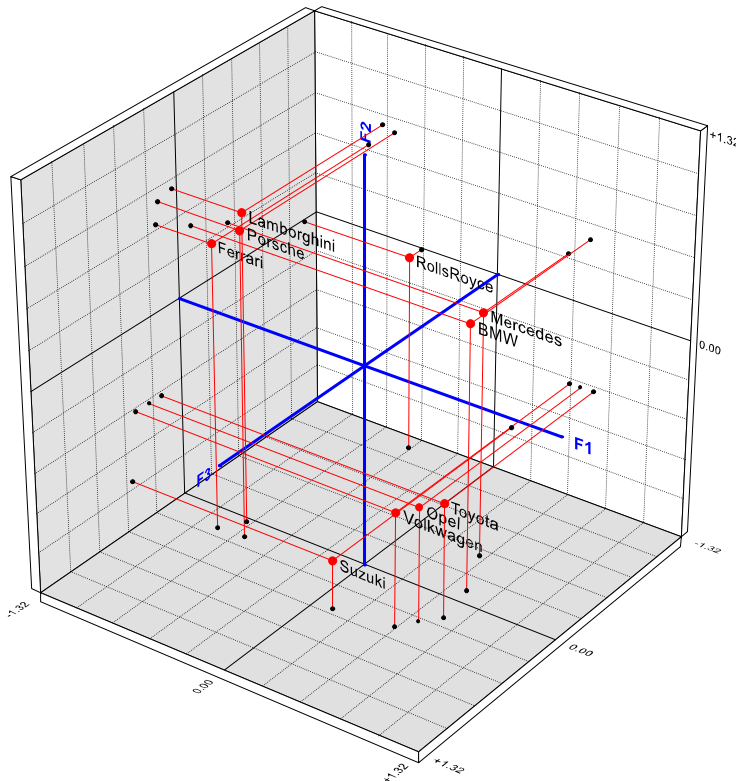
- als Startwerte Zufallszahlen
- 0 als vorgegebene Stress-Schwelle
- Option 6 = 100; also maximal 100 Iterationen

dann entsteht ein Stress-1 von 0.000000. (Almo gibt 6 Dezimalstellen aus). Vereinfacht gesagt: Der iterative Kalkül strebt korrekt gegen einen Stress von 0.0.

Die Grafik ist folgende

Bild 5: 3-dimensionale Grafik aus Almo (iterativ mit zufälligen Startwerten)





Die Grafik ist seltsam anders. Die 3 Gruppen

Gruppe 1	Gruppe 2	Gruppe 3
-----	-----	-----
Lamborghini	Toyota	Mercedes
Porsche	Opel	BMW
Ferrari	Volkswagen	

sind wie in Bild 1 deutlich voneinander getrennt. Auch die Sonderstellung von Suzuki und Rolls Royce ist erkennbar. Jedoch ist bei RollsRoyce und Suzuki gegenüber Bild 1 im Faktor F2 eine Vorzeichenumkehr erkennbar, bei Mercedes und BMW bei Faktor F3.

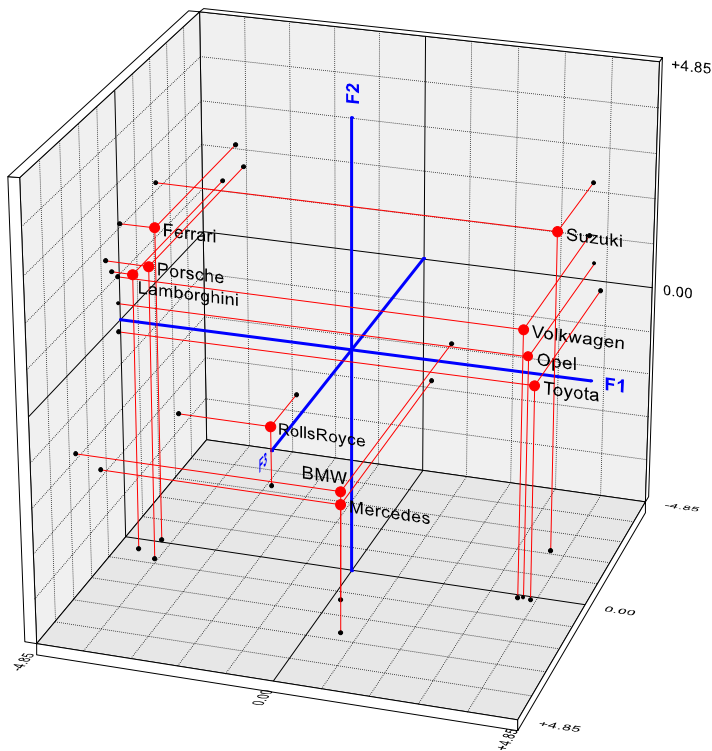
Hier wird nun ersichtlich, dass ein sehr guter Stress-1-Wert auch bei verschiedener Punktekonfiguration auftreten kann. Wie wir bereits oben ausgeführt haben, streben die iterativen Verfahren danach, den Stress zu minimieren. Wie hier ersichtlich wird muss das aber nicht unbedingt dazu führen, dass auch die euklidische Punktekonfiguration gefunden wird, die bei der metrischen MDS entsteht.

### 2-dimensionale Lösung.

Wir rechnen nun mit Almo, Proxscal und Alscal eine 2-dimensionale Lösung. Zunächst betrachten wir aber nochmals die 3-dimensionale Lösung aus der metrischen MDS (aus dem Programm "metrMDS\_3dim.Alm").

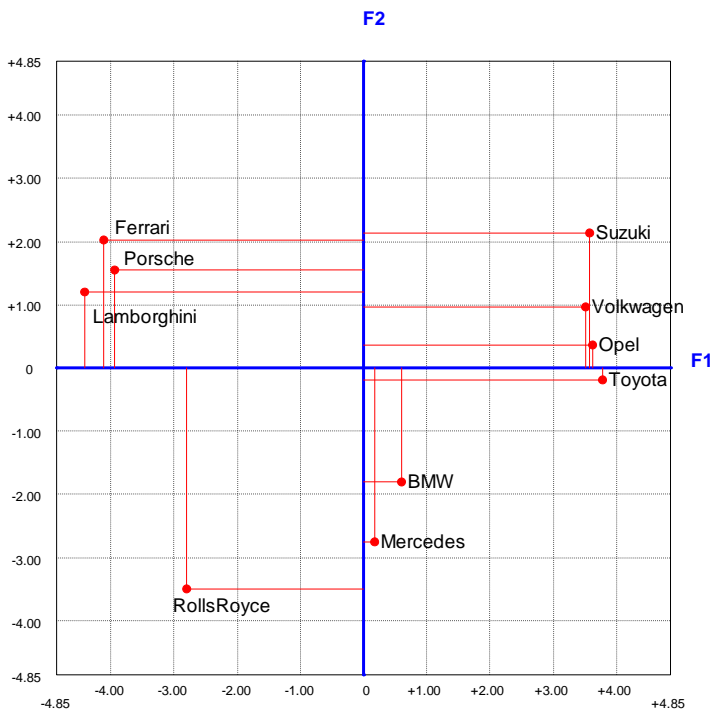
Wird die Grafik in Bild 1 im Grafik-Editor etwas horizontal gedreht, dann erkennt man das Abbild der 2-dimensionalen Lösung auf der Rückwand des aufgeschnittenen Würfels.

Bild 6: 3-dimensionale Grafik aus metrischer MDS (Prog "metrMDS\_3dim.Alm")  
(Projektion der Punkte auf Rückwand)



Wird im Grafik-Editor in der linken Werkzeugleiste auf den Knopf "Anderer Grafiktyp" geklickt und das 2-dimensionale Koordinatensystem selektiert, dann erhält man folgende Grafik der Rückwand des Würfels

Bild 7: 2-dimensionale Grafik aus metrischer MDS



Wird nun in der Programm-Maske für die metrische MDS die Faktorenzahl auf 2 gesetzt und neu gerechnet, dann liefert Almo exakt dasselbe Ergebnis und dieselbe Grafik. Der Stresskoeffizient dieser 2-dimensionalen Lösung ist jetzt mit 0.13497 deutlich schlechter. Das ist verständlich, da die Abweichung der oben genannten 4 Autos in die 3. Dimension

nicht berücksichtigt wurde und somit die Distanzmatrix nicht vollständig reproduziert werden konnte. Aber festzuhalten ist:

*Die Darstellung der Punktekongfiguration im reduzierten 2-dimensionalen Raum ist korrekt.*

Und:

*Wüßten wir nicht, dass die Distanzmatrix tatsächlich 3-dimensional ist, dann erhielten wir so den korrekten 2-dimensionalen Sub-Raum der eigentlich 3-dimensionalen Punktekongfiguration.*

Welche Ergebnisse liefern nun die iterativen Almo-MDS-Programme?

### Almo

Gerechnet wird mit dem iterativen MDS-Programm "iterMDS\_2dim.Alm". Es wird gefunden nach Klick auf den Knopf "alle Progs" in der Knopfleiste.

Wir rechnen eine Analyse mit 2 Dimensionen. Als Messniveau wird "quantitativ (intervall)" angegeben, als Startkonfiguration wird die Voreinstellung belassen, d.h. es werden die beiden ersten Faktoren aus der metrischen MDS eingesetzt. Als Stress-Schwelle wird 0.05 vorgegeben.

Almo liefert im 1. Iterationsschritt das Ergebnis aus der vorgegebenen Startkonfiguration. Dies ist die Ladungsmatrix aus der metrischen MDS, die noch normiert wurde (was belanglos ist)

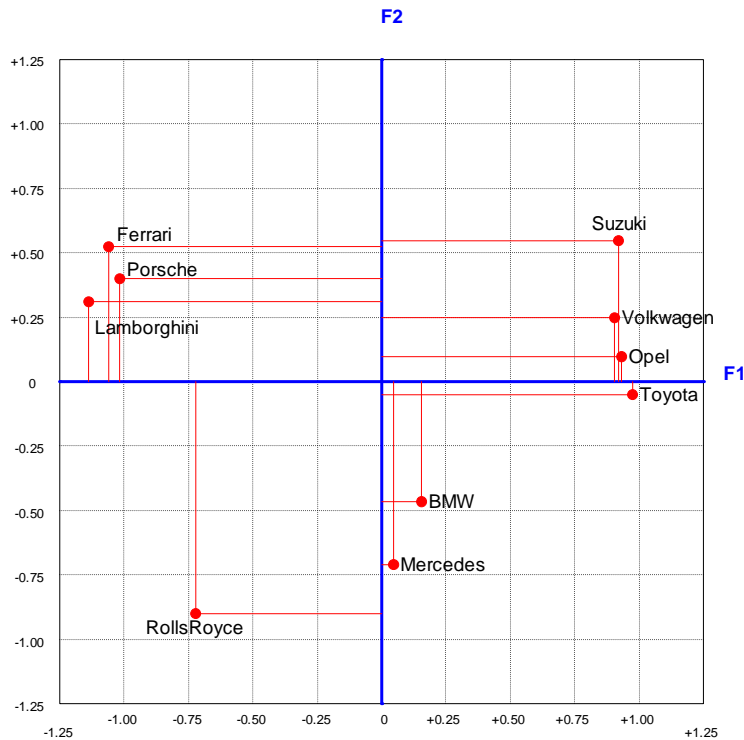
Ladungsmatrix der Startkonfiguration aus Iteration 1 der iterativen MDS (identisch mit Ergebnismatrix aus metrischer MDS)

		Faktor 1	Faktor 2
Opel	V1	-0.9367	0.0972
Volkwa	V2	-0.9085	0.2507
Suzuki	V3	-0.9238	0.5516
Toyota	V4	-0.9766	-0.0527
Merced	V5	-0.0469	-0.7136
BMW	V6	-0.1585	-0.4692
Ferrari	V7	1.0646	0.5260
Porsche	V8	1.0196	0.4027
Lamborghini	V9	1.1407	0.3112
RollsRoyce	V10	0.7262	-0.9039

Um die grafische Darstellung der Ladungsmatrix aus der 1. Iteration zu bekommen, muss ein Trick angewendet werden. Man rechnet das Programm ein 2. Mal, wobei man in der Optionsbox "weitere Optionen" schreibt: Option 6 = 1; Dadurch wird das Programm nach der 1. Iteration beendet und die bis dahin erzielten Ergebnisse mit Grafik ausgegeben.

Die grafische Darstellung des Ergebnisses ist folgende

Bild 8: 2-dimensionale Grafik aus nicht-metrischer MDS bei 1. Iteration mit Ladungsmatrix aus metrischer MDS als Startwerte-Matrix



Die Achse F1 in der Grafik wurde gespiegelt (Vorzeichenumkehr)  
 Werden die beiden Grafiken in Bild 7 und 8 miteinander verglichen, dann ist zu erkennen dass sie, abgesehen von der Bemäßung, gleich sind.

Als Stresskoeffizient wird 0.1134 ausgegeben und nicht wie oben bei der metrischen MDS 0.13497. Das ist zunächst irritierend, erklärt sich aber dadurch, dass die eingegebene Startwerte-Matrix normiert (bzw. standardisiert) wird. Bei der Normierung wird von jeder Ladung ihr Spaltenmittelwert subtrahiert und dann mit der Standardabweichung aus allen Ladungen aus allen Faktoren dividiert.

*Dies gilt immer: Bei identischer Punktekonfiguration ist der Stress-Koeffizient aus der metrischen MDS immer schlechter als der aus der nicht-metrischen MDS. Das heißt aber, dass man den Stress-Koeffizienten aus metrischer und iterativer MDS nicht dazu verwenden sollte zu entscheiden, welches Verfahren die bessere Lösung liefert.*

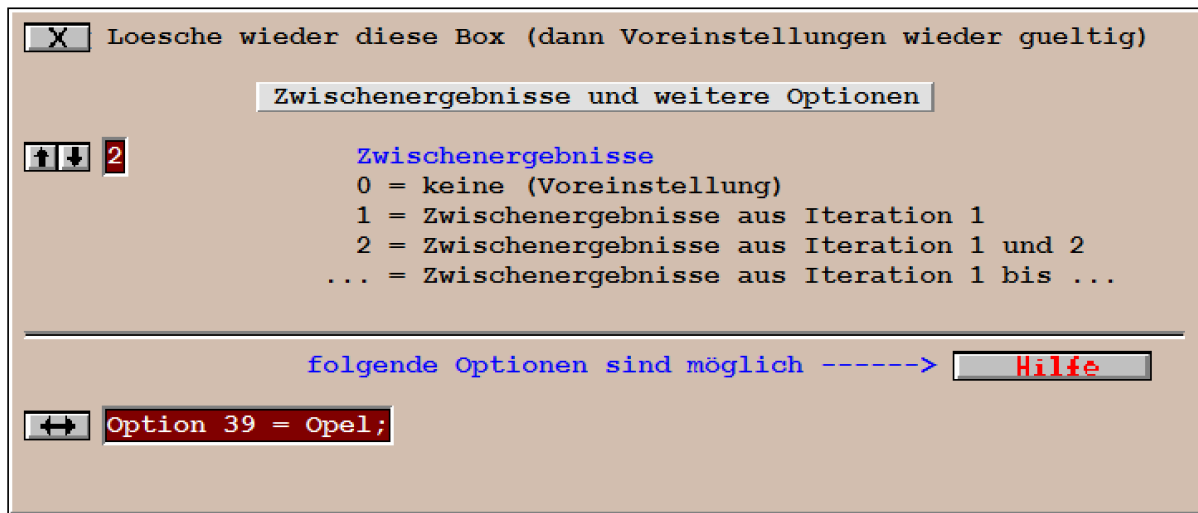
Da der vorgegebene Stress-Wert noch nicht erreicht wurde, läuft der Kruskal-Kalkül weiter - obwohl die korrekte 2-dimensionale Konfiguration bereits gefunden wurde. Durch Hin- und Herschieben der Punkte wird versucht den Stress weiter zu verringern. Wird in der Programm-Maske die Optionsbox "weitere Optionen" geöffnet und Zwischenergebnisse für 2 Iterationen angefordert, dann gibt Almo beispielhaft für den Punkt 1 (das ist die Automarke Opel) die Verschiebung aus. Von Iteration 1 zu 2 ist dies

neuer Koordinatenwert fuer Objekt 1 auf der Dimension 1  
 $-0.9367 - 3.654 \cdot -0.022 = -0.8576$

neuer Koordinatenwert fuer Objekt 1 auf der Dimension 2  
 $0.0972 - 3.654 \cdot 0.006 = 0.0764$

	Faktor 1	Faktor 2
	-----	-----
aus den alten Ladungen	-0.9367	0.0972
wird jetzt	-0.8576	0.0764

Wir wollen für Punkt 1 (Opel) die Verschiebungen über die Iterationen 1 bis 10 (nach der Almo abbricht) verfolgen. Das erreichen wir dadurch das wir in der Optionsbox "weitere Optionen" schreiben "Option 39 = 1;" (oder: Option 39 = Opel"). Die Optionsbox sieht also folgendermaßen aus:



Almo liefert folgende Ausgabe:

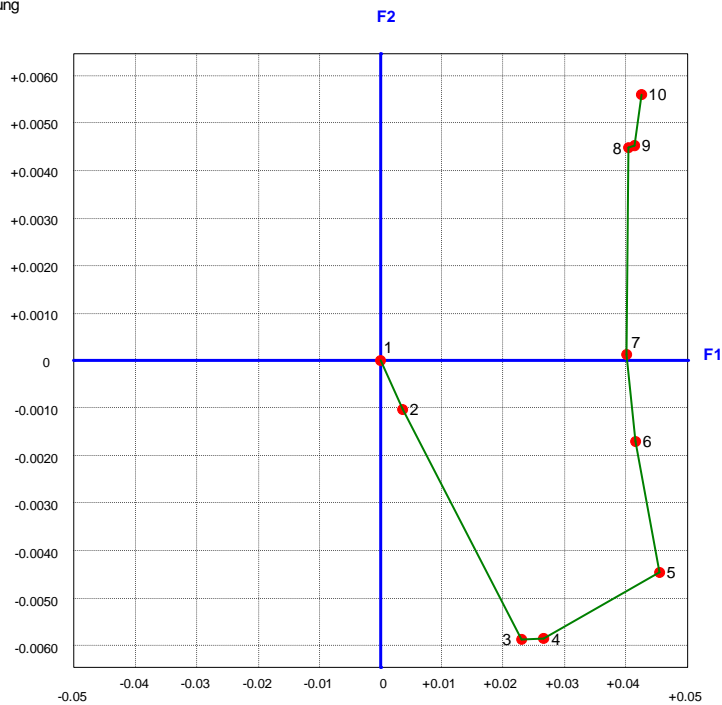
Wanderung des Objektpunktes 1 Opel ueber 10 Iterationen

Iteration	Faktor 1	Faktor 2	Stress-Minderung	
			Stress-1	in %
1	-0.9367	0.0972	0.113351	-
2	-0.9327	0.0962	0.108002	4.72
3	-0.9115	0.0914	0.086797	19.63
4	-0.9075	0.0914	0.084055	3.16
5	-0.8870	0.0928	0.079277	5.68
6	-0.8913	0.0955	0.078447	1.05
7	-0.8929	0.0974	0.078162	0.36
8	-0.8925	0.1017	0.077812	0.45
9	-0.8914	0.1018	0.077768	0.06
10	-0.8903	0.1028	0.077738	0.04

Um die Punktwanderung in einer Grafik besser sichtbar zu machen, wird die Grafik "gespreizt". Das geschieht durch folgende Transformation: Die Koordinaten der Wanderungs-Punkte werden so verschoben dass der Punkt für Iteration 1 im Ursprung liegt. Die Koordinatenwerte der 1. Iteration werden sehr einfach von allen Iterationen subtrahiert.

Iteration	Faktor 1	Faktor 2
1	0.0000	0.0000
2	0.0040	-0.0010
3	0.0251	-0.0059
4	0.0291	-0.0058
5	0.0496	-0.0045
6	0.0454	-0.0017
7	0.0438	0.0001
8	0.0442	0.0045
9	0.0453	0.0045
10	0.0464	0.0056

Bild 9: Punkte-Wanderung von Opel in Iteration 1 bis 10



In obigem Bild 8 haben wir für F1 das Vorzeichen umgedreht. Das müsste auch hier geschehen. Wir tun dies nicht. Der Benutzer muss sich die Grafik "auf den Kopf gestellt" vorstellen.

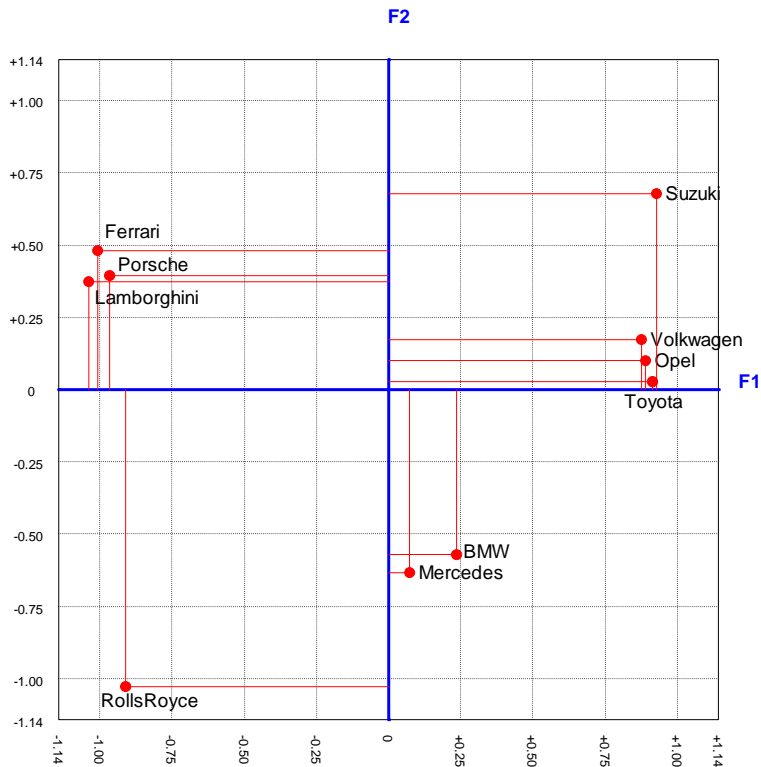
Es ist zu erkennen, dass der Punkt "Opel" zuerst nach unten und dann nach oben wandert. In obiger Tabelle ist der Stress-1 in der 3. Spalte angegeben. Nach Iteration 5 verbessert sich dann der Stress nur noch vernachlässigbar.

Nach 5 Iterationen wird ein Stress von 0.0793 und nach 10 von 0.0777 erreicht. Gegenüber der Startkonfiguration aus der metrischen MDS mit 0.1134 wurde eine deutliche Verbesserung erzielt.

*Die Punkt-Verschiebungen der eigentlich 3-dimensionalen Konfiguration im reduzierten 2-dimensionalen Raum hat zu einer Verbesserung des Stress-1 geführt. Diese wird damit bezahlt, dass die korrekte 2-dimensionale Abbildung der Punktekonfiguration, d.h. die Projektion der Punkte auf die Rückwand des Würfels verändert wird.*

Grafisch dargestellt ist die endgültige Punktekonfiguration nach 10 Iterationen dann folgende

Bild 10: 2-dimensionale Grafik aus nicht-metrischer MDS nach 10. Iteration mit Ladungsmatrix aus metrischer MDS als Startwerte-Matrix



Beim Vergleich mit Bild 8 (aus der 2-dimensionalen metrischen MDS) erkennt man:

Gleich geblieben sind:

Die Struktur der Punktekonfiguration ist dieselbe. Deutlich voneinander getrennt sind die 3 Gruppen (a) die Sportwagen, (b) die Alltagsautos, (c) die Komfortautos und die zwei Einzelgänger RollsRoyce und Suzuki.

Verändert sind:

1. Die 3 Sportwagen Ferrari, Lamborghini, Porsche sind enger zusammengedrückt, genau so die "Alltags-Autos" VW, Opel, Toyota und auch die beiden Komfort-Autos BMW und Mercedes
2. Die Automarke Rolls Royce ist deutlich nach unten verschoben und Suzuki nach oben

### Folgerungen

Aus unserem Simulationsexperiment leiten wir folgende Überlegungen und Empfehlungen an den Benutzer ab:

1. Sind die Unähnlichkeiten quantitativ (intervall-skaliert) dann liefert die metrische MDS immer ein korrektes Ergebnis. Empirische Daten sind im Unterschied zu unseren simulierten Daten fehlerbehaftet. Es kann also nicht gelingen, die Dimensionszahl des Raums der Objekt-Unähnlichkeiten exakt zu bestimmen. Die metrische MDS liefert aber immer den korrekten Sub-Raum des unbekannt-dimensionalen euklidischen Raums. Die iterativen Verfahren verzerren die Punktekonfiguration im Sub-Raum. Unseres Erachtens gibt es keinen Grund dafür, ein iteratives Verfahren zu wählen, wenn die Unähnlichkeiten quantitativ sind und die Metrik euklidisch. Bei ordinalen Unähnlichkeiten hingegen sind nur die iterativen Verfahren verfügbar. Seltsamerweise ist im weit verbreiteten SPSS keine Prozedur für die metrische MDS vorhanden.

2. Will der Benutzer jedoch, aus welchen Gründen auch immer, ein iteratives Verfahren bei quantitativen Unähnlichkeiten einsetzen, dann sollte er bestrebt sein, die "richtige" Dimensionszahl zu finden. Nur bei dieser tritt die in unserer Simulation vorgeführte Verzerrung der Punktekonfiguration im Sub-Raum nicht auf.

Der Benutzer sollte nicht eine 2-dimensionale Lösung anstreben, wie dies so oft bei empirischen Forschungsarbeiten beobachtbar ist - nur weil sich 2 Dimensionen grafisch anschaulich darstellen lassen.

## **Literatur**

Eine kurze und sehr übersichtliche Darstellung (auch zur metrischen MDS) ist im Internet zu finden bei

**Jacoby, W. G.** : Multidimensional Scaling, an Introduction, 2012, im Internet herunter ladbar <http://polisci.msu.edu/jacoby/iu/mds2012/outline/2012%20IU%20MDS%20Outline,%2012-2-12.pdf>

Eine ausführliche und detaillierte Darstellung der nicht-metrischen MDS nach Kruskal ist enthalten in

**Bacher, J. / Böge, A. / Wenzig K.** : Clusteranalyse, 3. Auflage, Kapitel 4 Nichtmetrische multidimensionale Skalierung, 2010, Oldenbourg Verlag, München

---

**Borg, I., Groenen, P.**: Modern multidimensional scaling: theory and applications, 2005, New York, Springer, 2. Auflage

**Borg, I.**: Multidimensionale Skalierung, in Wolf/Best(Hrsg.): Handbuch der sozialwissenschaftlichen Datenanalyse, 2010, VS Verlag Wiesbaden

**BUGH Wuppertal**: "Multidimensionale Skalierung", Lehrstuhl für empirische Wirtschafts- und Sozialforschung, Fachbereich Wirtschaftswissenschaft, 2001, im Internet herunter ladbar bei

[http://www2.informatik.uni-osnabrueck.de/marc/lectures/zra\\_ss03/prgdat/mds.pdf](http://www2.informatik.uni-osnabrueck.de/marc/lectures/zra_ss03/prgdat/mds.pdf) )

**Gäbler von, E.E.**: Mehrdimensionale Skalierung aus inferenzstatistischer und anwendungsorientierter Sicht, Dissertation Tübingen, 1982.

**Hammerle, A. / Pape, H.**: Grundlagen der mehrdimensionalen Skalierung, in: Fahrmeir, L./Hamerle, A. (Hg.): Multivariate statistische Verfahren. Berlin und New York, 1984, S. 133 - 688.

**Kruskal J. B.**: Nonmetric multidimensional scaling, Psychometrika, 1964, 29, S. 115-131

**de Leeuw, J. / Mair P.** :Multidimensional Scaling Using Majorization: SMACOF in R, Journal of statistical software, August 2009, Volume 31, Issue 3. (Auch im Internet zu finden)

**Sixtl, F.**: Meßmethoden der Psychologie, 2. Aufl., Weinheim und Basel, 1982.

**Takane, Y., Young / F. W., de Leeuw, J.** (1977). Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features.



Psychometrika, 42, 7-67.

**Warren S. Torgerson:** Theory and Method of Scaling, New York, Wiley, 1958