

*Ingwer Borg, Patrick J.F. Groenen, Patrick Mair:*

## **Multidimensionale Skalierung**

*Sozialwissenschaftliche Forschungsmethoden,*

hrsg. von Martin Spieß, Wenzel Matiaske u.a., Band 1

ISBN 978-3-86618-438-1, ISBN 978-3-86618-538-8 (e-book pdf),

Rainer Hampp Verlag, München u. Mering 2010, 102 S., € 19.80

Die Multidimensionale Skalierung (MDS) ist eine Familie von Verfahren, die Objekte des Forschungsinteresses durch Punkte eines mehrdimensionalen (meist: 2-dimensionalen) Raums so darstellen, dass die Distanz zwischen je zwei Punkten in diesem Raum einem gegebenen Nähe-, Abstands-, Ähnlichkeits- oder Unähnlichkeitswert dieser Objekte optimal entspricht. Der Zweck der MDS liegt meist in der Visualisierung der wesentlichen Struktur der Daten. Diese soll dem Auge für eine explorative oder Theorie testende Analyse zugänglich gemacht werden. Als Daten lassen sich in der MDS außerordentlich viele Messgrößen verwenden wie etwa Korrelationen der Objekte über ihre Ausprägungen auf verschiedenen Variablen; direkt erhobene globale Ähnlichkeitsratings für Paare von Objekten; oder Co-Occurrence-Koeffizienten, die erfassen, wie oft ein Ereignis zusammen mit einem anderen auftritt.

Das Buch ist eine anwenderorientierte Einführung in die MDS. Es beschreibt die für die Praxis wichtigsten MDS-Modelle (inkl. der konfirmatorischen MDS) konzeptionell, mit wenigen Formeln, und unter Verwendung typischer Beispiele. Es bietet zudem eine Beschreibung von zwei umfassenden Computerprogrammen für die MDS (PROXSCAL in SPSS, SMACOF in R). Ausführlich diskutiert werden zudem typische Anwenderfehler in der MDS.

**Schlüsselwörter:** Multivariate Statistik, Datenanalyse, Visualisierungsverfahren, Urteilsbildung, Psychologische Methodik, Sozialwissenschaftliche Methodik

*Ingwer Borg* ist wissenschaftlicher Leiter der Abteilung Survey Design & Methodology (SDM) der GESIS (Mannheim) und Professor am Fachbereich Psychologie der Uni Gießen. Seine Forschung liegt überwiegend im Bereich von Mitarbeiterbefragungen, Werten und Einstellungen sowie Skalierungsverfahren.

*Patrick J.F. Groenen* ist Professor für Statistik am Ökonometrischen Institut der Erasmus School of Economics der Erasmus Universität in Rotterdam, Niederlande. Seine Forschung fokussiert auf Explorative Multivariate Datenanalyse, Multidimensionale Skalierung und numerische Algorithmen.

*Patrick Mair* ist Universitätsassistent am Institut für Statistik und Mathematik an der Wirtschaftsuniversität Wien. Seine Forschungsschwerpunkte sind psychometrische Methoden, kategoriale Datenanalyse, Modelle für latente Variablen sowie computationale Statistik allgemein.

# SOZIALWISSENSCHAFTLICHE FORSCHUNGSMETHODEN

herausgegeben von

Martin Spieß, Wenzel Matiaske,  
Ingwer Borg, Claudia Fantapié-Altobelli, Holger Hinz,  
Uwe Jirjahn, Bernhard Kittel, Manfred Kraft,  
Stefan Liebig, Rainer Oesterreich, Jost Reinecke,  
Kai-Uwe Schnapp, Rainer Schnell, Peter Sedlmeier,  
Winfried Seidel, Gerhard Tutz, Joachim Wagner

Band 1

Ingwer Borg, Patrick J.F. Groenen, Patrick Mair

# Multidimensionale Skalierung

Rainer Hampp Verlag

München und Mering 2010

### **Bibliografische Information der Deutschen Nationalbibliothek**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

ISBN 978-3-86618-438-1 (print)

ISBN 978-3-86618-538-8 (e-book)

SOZIALWISSENSCHAFTLICHE FORSCHUNGSMETHODEN: ISSN 1616-6876

DOI 10.1688/9783866185388

1. Auflage, 2010

© 2010 Rainer Hampp Verlag München und Mering  
Marktplatz 5 D – 86415 Mering  
[www.Hampp-Verlag.de](http://www.Hampp-Verlag.de)

Alle Rechte vorbehalten. Dieses Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne schriftliche Zustimmung des Verlags unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Mikroverfilmungen, Übersetzungen und die Einspeicherung in elektronische Systeme.

∞ *Dieses Buch ist auf säurefreiem und chlorfrei gebleichtem Papier gedruckt.*

*Liebe Leserinnen und Leser!*

*Wir wollen Ihnen ein gutes Buch liefern. Wenn Sie aus irgendwelchen Gründen nicht zufrieden sind, wenden Sie sich bitte an uns.*

---

# Inhaltsverzeichnis

<b>1</b>	<b>Erste Schritte</b> .....	1
<b>2</b>	<b>Vom Zweck der MDS</b> .....	7
2.1	MDS zur Aufdeckung latenter Urteilsdimensionen .....	7
2.2	Distanzformeln als Modelle der Urteilsbildung .....	9
2.3	MDS zur Visualisierung von Ähnlichkeitsdaten .....	14
2.4	MDS zur Überprüfung von Strukturhypothesen .....	16
<b>3</b>	<b>Der Stress einer MDS-Darstellung</b> .....	19
3.1	Überlegungen zur Güte einer MDS-Lösung .....	19
3.2	Bewertung des Stress .....	21
3.3	Andere Stress-Maße .....	23
<b>4</b>	<b>Proximitäten</b> .....	25
4.1	Direkt erhobene Proximitäten .....	25
4.2	Aus Datenvektoren abgeleitete Proximitäten .....	27
4.3	Proximitäten durch Konversion anderer Indices .....	28
4.4	Co-Occurrence Daten .....	28
<b>5</b>	<b>Einige MDS-Modelle</b> .....	31
5.1	Ordinale und metrische MDS .....	31
5.2	Euklidische und andere Distanzen .....	33
5.3	Replizierte Proximitäten in $N$ Datensätzen .....	33
5.4	MDS asymmetrischer Proximitäten .....	34
5.5	MDS-Modellierung interindividueller Unterschiede .....	36
5.6	Unfolding .....	38

<b>6</b>	<b>Konfirmatorische MDS</b> .....	41
6.1	Schwach-konfirmatorische MDS-Ansätze .....	42
6.2	Externe Nebenbedingungen auf den Punktkoordinaten .....	42
6.3	Regionale axiale Restriktionen .....	45
6.4	Herausforderungen der konfirmatorischen MDS .....	47
<b>7</b>	<b>Häufige Anwendungsfehler in der MDS</b> .....	49
7.1	Allzu salopper Umgang mit dem Begriff Distanz .....	49
7.2	Falsche Polung der Proximitäten .....	50
7.3	Verfrühter Abbruch der Iterationen .....	50
7.4	Falsche Startkonfiguration .....	50
7.5	Verwendung suboptimaler lokaler Minima .....	51
7.6	Degenerierte Lösungen in der ordinalen MDS .....	52
7.7	Reflexartige Interpretation „der“ Dimensionen .....	55
7.8	Behandlung störender Punkte .....	59
7.9	Unzulässige Vergleiche verschiedener MDS-Lösungen .....	60
7.10	Mechanisch-formale Bewertung des Stress .....	62
7.11	Überinterpretation von Dimensionsgewichten .....	63
7.12	Inhaltsleere MDS-Strukturen fast gleicher Proximitäten .....	64
<b>8</b>	<b>MDS-Algorithmen</b> .....	67
8.1	Klassische MDS .....	67
8.2	Iterative MDS-Algorithmen .....	69
<b>9</b>	<b>Computerprogramme für MDS</b> .....	73
9.1	PROXSCAL .....	73
9.1.1	PROXSCAL über Menüs .....	74
9.1.2	PROXSCAL über Kommandos .....	77
9.2	SMACOF, ein R-Package .....	81
9.2.1	Smacof allgemein .....	81
9.2.2	SmacofSym .....	83
9.2.3	SmacofIndDiff .....	87
9.2.4	SmacofRect .....	88
9.2.5	SmacofConstraint .....	90
9.2.6	Smacof Schlussanmerkungen .....	91
9.3	Andere Programme für die MDS .....	91
	<b>Literaturverzeichnis</b> .....	93
	<b>Autorenregister</b> .....	97
	<b>Index</b> .....	99



## Erste Schritte

Die Grundideen der MDS lassen sich recht einfach mit einem kleinen Beispiel demonstrieren. Tabelle 1.1 zeigt Korrelationen für die Häufigkeit verschiedener Verbrechen in 50 Bundesstaaten der USA. Man sieht hier z.B., dass dann, wenn in einem Bundesstaat relativ viele Fälle von Körperverletzung registriert wurden, auch relativ viele Morde vorkommen ( $r = .81$ ). Dagegen sagt eine hohe Mordrate so gut wie nichts über die Häufigkeit von Diebstahlsdelikten voraus ( $r = .06$ ).

Wir wollen nun diese Korrelationen mit der MDS skalieren. Das bedeutet, dass wir versuchen, die sieben Verbrechen durch sieben Punkte in einem Raum so darzustellen, dass je zwei Punkte umso *näher* zusammenliegen, je *größer* die Korrelation der Verbrechen ist, die diese Punkte repräsentieren. Dazu gehen wir wie folgt vor.

Wir nehmen sieben Karten und schreiben darauf jeweils ein Verbrechen, von Mord bis Autodiebstahl. Die Karten legen wir dann in beliebiger Anordnung auf einen Tisch, etwa so wie in Abb. 1.1 gezeigt. Dann messen wir die Distanzen aller Karten untereinander (Abb. 1.2) und vergleichen diese Messwerte mit den Korrelationen. Dieser Vergleich zeigt, dass die Punktekonfiguration in Abb. 1.1 die Daten nicht im gewünschten Sinn repräsentiert. So sollten z.B. die Punkte für Mord und Körperverletzung relativ nahe zusammenliegen, da diese Verbrechen mit 0.81 korrelieren; dagegen sollten die Punkte für Mord und Diebstahl weiter voneinander entfernt liegen, da diese Verbrechen nur mit 0.06 korrelieren. Wir versuchen daher, die Punkte nach und nach in kleinen Schritten („iterativ“) so zu verschieben, dass die Distanzen den Datenwerten genauer entsprechen. In Abb. 1.3 haben wir angedeutet, in welche Richtungen die Punkte (in diesem ersten Schritt) verschoben werden sollten, um die Übereinstimmung von Daten und Distanzen zu verbessern.

Nach einigen Iterationen sollte man auf diese Weise eine Konfiguration finden, die der in Abb. 1.4 nahe kommt. Da iterative Punktverschiebungen per Hand ziemlich mühsam sein können und auch nicht garantieren, dass man so eine *optimale*

Tabelle 1.1: Die Häufigkeiten verschiedener Verbrechen interkorreliert über 50 Bundesstaaten der USA.

Verbrechen	Mord	Verg.	Raub	Körperv.	Einbr.	Diebst.	Autod.
Mord	1.00	0.52	0.34	0.81	0.28	0.06	0.11
Vergewaltigung	0.52	1.00	0.55	0.70	0.68	0.60	0.44
Raub	0.34	0.55	1.00	0.56	0.62	0.44	0.62
Körperverletzung	0.81	0.70	0.56	1.00	0.52	0.32	0.33
Einbruch	0.28	0.68	0.62	0.52	1.00	0.80	0.70
Diebstahl	0.06	0.60	0.44	0.32	0.80	1.00	0.55
Autodiebstahl	0.11	0.44	0.62	0.33	0.70	0.55	1.00

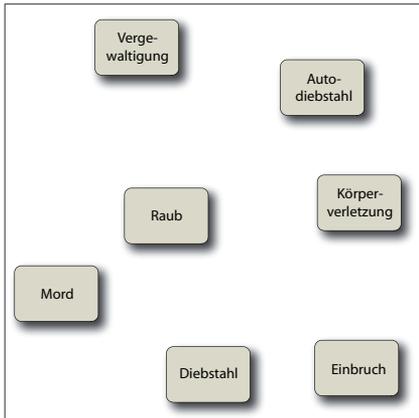


Abb. 1.1: Startkonfiguration für eine MDS der Daten aus Tabelle 1.1.

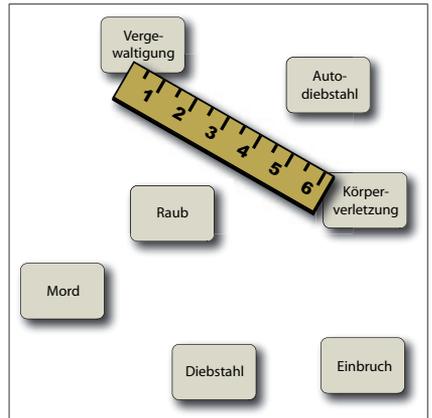


Abb. 1.2: Ausmessen der Distanzen mit dem Lineal.

Punktekonfiguration findet, wurde die hier gezeigte finale Lösung mit einem MDS-Computerprogramm erstellt.

Ein solches Programm ist z.B. PROXSCAL in SPSS. Für eine PROXSCAL-Anwendung legen wir zunächst die obige Korrelationsmatrix im File „CorrCrimes.sav“ ab. Dann brauchen wir nur die folgenden Kommandos:

```
GET FILE='CorrCrimes.sav'.
PROXSCAL VARIABLES=Mord to Autodiebstahl
      /PROXIMITIES=SIMILARITIES .
```

Wir teilen PROXSCAL hier also mit, dass es die Daten (genannt: „PROXIMITIES“) als Ähnlichkeiten („SIMILARITIES“) deuten soll (d.h. je größer der Datenwert, desto

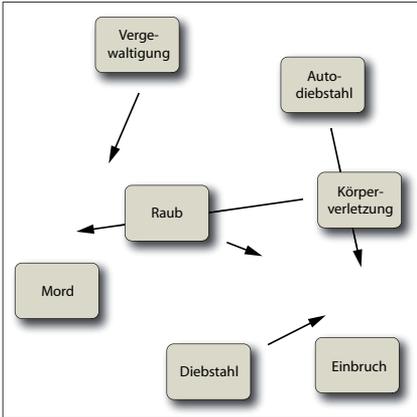


Abb. 1.3: Richtungen für Punktverschiebungen, die die MDS-Konfiguration verbessern sollten.

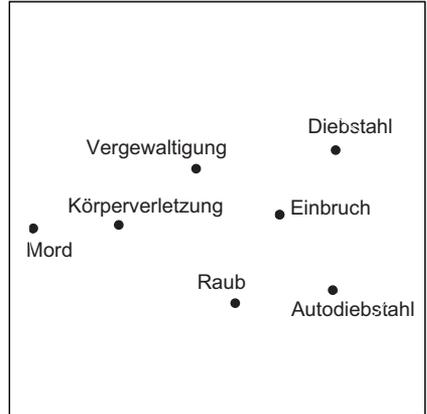


Abb. 1.4: Nach einigen Iterationen gefundene MDS-Repräsentation der Korrelationen aus Tabelle 1.1.

näher sollten die Punkte zusammenliegen). Weitere Angaben sind nicht nötig. Wir belassen es hier bei den Voreinstellungen des Programms.

Viele andere Programme existieren für die MDS. Eines mit schönen Grafiken ist das MDS-Modul im Statistikpaket SYSTAT. Man kann es, ebenso wie PROXSCAL, entweder mit Kommandos oder mit dem Anklicken verschiedener Optionen in einer grafischen Benutzeroberfläche bedienen. Dazu laden wir zunächst den Datenfile mit den Korrelationen und rufen dann die MDS-Prozedur auf. In der Oberfläche in Abb. 1.5 belassen wir es bei allen Voreinstellungen, außer dort, wo der Pfeil zu sehen ist: Dort fordern wir, dass die Beziehung („Regression“) der Daten zu den Distanzen „linear“ sein soll.

Beide Computerprogramme (PROXSCAL und SYSTAT) errechnen praktisch die gleichen MDS-Darstellungen für die Korrelationen in Tabelle 1.1. Sie sind nicht nur optimal, sondern auch gut, wie Abb. 1.6 zeigt: Der Zusammenhang der Daten und der Distanzen ist fast perfekt linear ( $r = -0.99$ ). Man macht also keine Fehler, wenn man die Abstände zwischen den Punkten der MDS-Konfiguration wie die empirisch ermittelten Korrelationen deutet.

Was hat man durch diese MDS gewonnen? Zunächst bekommt man statt der 21 verschiedenen Korrelationen eine *visuelle Darstellung der empirischen Zusammenhänge*. Das ermöglicht es, die Struktur dieser Zusammenhänge *mit dem Auge zu explorieren*. Wie in Abb. 1.7 gezeigt bilden die Verbrechen, bei denen Personen zu Schaden kommen bzw. bei denen es um Eigentumsdelikte geht, jeweils eine *Nachbarschaft*. Ist also z.B. die Mordrate in einem Bundesstaat hoch, dann gibt es in diesem Bundesstaat auch relativ viele Körperverletzungen und viele Vergewalti-

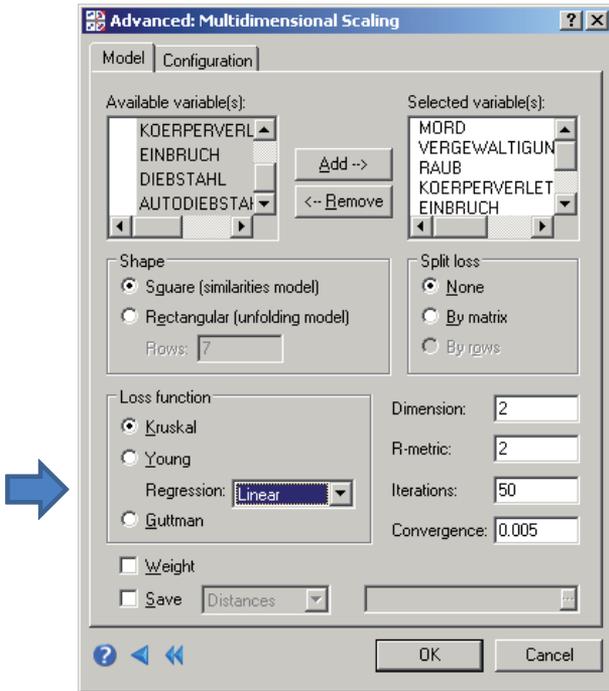


Abb. 1.5: Anweisungen für eine MDS in SYSTAT.

gungen. Ähnliches gilt für die Eigentumsdelikte. Raub liegt quasi zwischen diesen Nachbarschaften, evtl. deshalb, weil es beide Verbrechenarten verbindet.

Diese Information ist im Wesentlichen auf der ersten *Hauptachse* dargestellt, d.h. einer Geraden, die hier waagrecht durch die Punktwolke verläuft<sup>1</sup>. Die zweite Hauptachse, auf der Diebstahl und Autodiebstahl weit auseinander liegen, scheint dagegen nicht interpretierbar: In welcher Hinsicht sind Diebstahl (im Original: „larceny“) und Autodiebstahl („autotheft“) denn so verschieden? Auf diese Frage scheint es keine Antwort zu geben. Die zweite Dimension der MDS-Lösung stellt daher möglicherweise nur Fehler in den Daten dar. Es liegt daher nahe zu testen, ob vielleicht eine 1-dimensionale MDS-Darstellung für diese Daten ausreicht.

<sup>1</sup> Die erste Hauptachse ist eine Gerade, die so in der Punktwolke liegt, dass die Summe der (quadratischen) Abstände der Punkte von ihr minimal ist. Anders ausgedrückt: Diese Gerade verläuft so, dass die Varianz der Projektionen der Punkte auf sie maximal ist. Die zweite Hauptachse steht senkrecht auf der ersten und erklärt das Maximum der verbleibenden Varianz.