

IBM SPSS Categories 24

IBM

Hinweis

Vor Verwendung dieser Informationen und des darin beschriebenen Produkts sollten die Informationen unter „Bemerkungen“ auf Seite 57 gelesen werden.

Produktinformation

Diese Ausgabe bezieht sich auf Version 24, Release 0, Modifikation 0 von IBM SPSS Statistics und alle nachfolgenden Releases und Modifikationen, bis dieser Hinweis in einer Neuausgabe geändert wird.

Diese Veröffentlichung ist eine Übersetzung des Handbuchs

IBM SPSS Categories 24,

herausgegeben von International Business Machines Corporation, USA

© Copyright International Business Machines Corporation 2016

Informationen, die nur für bestimmte Länder Gültigkeit haben und für Deutschland, Österreich und die Schweiz nicht zutreffen, wurden in dieser Veröffentlichung im Originaltext übernommen.

Möglicherweise sind nicht alle in dieser Übersetzung aufgeführten Produkte in Deutschland angekündigt und verfügbar; vor Entscheidungen empfiehlt sich der Kontakt mit der zuständigen IBM Geschäftsstelle.

Änderung des Textes bleibt vorbehalten.

Herausgegeben von:

TSC Germany

Kst. 2877

Januar 2016

Inhaltsverzeichnis

Kapitel 1. Einführung in die Prozeduren für die optimale Skalierung von kategorialen Daten 1

Was ist optimale Skalierung?	1
Warum sollte optimale Skalierung verwendet werden?	1
Messniveau der optimalen Skalierung	2
Auswahl des Messniveaus der optimalen Skalierung	2
Transformationsdiagramme	3
Kategoriecodes	3
Welche Prozedur ist für Ihre Anwendung am besten geeignet?	5
Kategoriale Regression	6
Kategoriale Hauptkomponentenanalyse	6
Nicht lineare kanonische Korrelationsanalyse	7
Korrespondenzanalyse	8
Mehrfachkorrespondenzanalyse	9
Multidimensionale Skalierung	10
Multidimensionale Entfaltung	10
Seitenverhältnis in Diagrammen mit optimaler Skalierung	11

Kapitel 2. Kategoriale Regression (CATREG) 13

Definieren der Skala in der kategorialen Regression	14
Kategoriale Regression: Diskretisierung	14
Kategoriale Regression: Fehlende Werte	15
Kategoriale Regression: Optionen	15
Kategoriale Regression: Regularisierung	16
Kategoriale Regression: Ausgabe	17
Kategoriale Regression: Speichern	17
Kategoriale Regression: Transformationsdiagramme	18
Zusätzliche Funktionen beim Befehl CATREG	18

Kapitel 3. Kategoriale Hauptkomponentenanalyse (CATPCA) 19

Definieren von Skala und Gewichtung in CATPCA	20
Kategoriale Hauptkomponentenanalyse: Diskretisierung	21
Kategoriale Hauptkomponentenanalyse: Fehlende Werte	21
Kategoriale Hauptkomponentenanalyse: Optionen	22
Kategoriale Hauptkomponentenanalyse: Ausgabe	23
Kategoriale Hauptkomponentenanalyse: Speichern	24
Kategoriale Hauptkomponentenanalyse: Objekt- und Variablendiagramme	24
Kategoriale Hauptkomponentenanalyse: Kategoriediagramme	25
Kategoriale Hauptkomponentenanalyse: Ladungsdiagramme	25
Kategoriale Hauptkomponentenanalyse: Bootstrap	26
Zusätzliche Funktionen beim Befehl CATPCA	26

Kapitel 4. Nicht lineare kanonische Korrelationsanalyse (OVERALS) 27

Definieren von Bereich und Skala	28
Definieren eines Bereichs	29
Nicht lineare kanonische Korrelationsanalyse: Optionen	29
Zusätzliche Funktionen beim Befehl OVERALS	30

Kapitel 5. Korrespondenzanalyse 31

Definieren des Zeilenbereichs in der Korrespondenzanalyse	32
Definieren des Spaltenbereichs in der Korrespondenzanalyse	32
Korrespondenzanalyse: Modell	33
Korrespondenzanalyse: Statistiken	34
Korrespondenzanalyse: Diagramme	34
Zusätzliche Funktionen beim Befehl CORRESPONDENCE	35

Kapitel 6. Mehrfachkorrespondenzanalyse 37

Definieren der Variablen Gewichtung in der Mehrfachkorrespondenzanalyse	38
Mehrfachkorrespondenzanalyse: Diskretisierung	38
Mehrfachkorrespondenzanalyse: Fehlende Werte	38
Mehrfachkorrespondenzanalyse: Optionen	39
Mehrfachkorrespondenzanalyse: Ausgabe	40
Mehrfachkorrespondenzanalyse: Speichern	41
Mehrfachkorrespondenzanalyse: Objektdiagramme	41
Mehrfachkorrespondenzanalyse: Variablendiagramme	41
Zusätzliche Funktionen beim Befehl MULTIPLE CORRESPONDENCE	41

Kapitel 7. Multidimensionale Skalierung (PROXSCAL) 43

Ähnlichkeiten in mehreren Spalten von Matrizen	44
Ähnlichkeiten in Spalten	44
Ähnlichkeiten in einer Spalte	45
Erstellen von Ähnlichkeiten aus Daten	45
Erstellen von Distanzen aus Daten	45
Definieren eines Modells für die multidimensionale Skalierung	46
Multidimensionale Skalierung: Einschränkungen	47
Multidimensionale Skalierung: Optionen	47
Multidimensionale Skalierung: Diagramme, Version 1	48
Multidimensionale Skalierung: Diagramme, Version 2	49
Multidimensionale Skalierung: Ausgabe	49
Zusätzliche Funktionen beim Befehl PROXSCAL	49

Kapitel 8. Multidimensionale Entfaltung (PREFSCAL) 51

Definieren eines Modells für die multidimensionale Entfaltung 52
Multidimensionale Entfaltung: Einschränkungen . . 52
Multidimensionale Entfaltung: Optionen. 53
Multidimensionale Entfaltung: Diagramme 54
Multidimensionale Entfaltung: Ausgabe 55

Zusätzliche Funktionen beim Befehl PREFSCAL . . . 55

Bemerkungen. 57

Marken. 58

Index 61

Kapitel 1. Einführung in die Prozeduren für die optimale Skalierung von kategorialen Daten

Bei den Prozeduren in Categories wird die optimale Skalierung für die Analyse von Daten verwendet, die mit statistischen Standardprozeduren nur schwer oder gar nicht analysiert werden können. In diesem Kapitel werden die einzelnen Prozeduren, die Situationen, in denen die einzelnen Prozeduren am besten geeignet sind, die Beziehungen zwischen den Prozeduren und die Beziehungen der Prozeduren zu ihrem jeweiligen Gegenstück aus der Standardstatistik beschrieben.

Hinweis: Diese Prozeduren und ihre Implementierung in IBM® SPSS Statistics wurden von der Data Theory Scaling System Group (DTSS) entwickelt, die aus Mitgliedern der Institute für Pädagogik und Psychologie der sozialwissenschaftlichen Fakultät der Universität Leiden besteht.

Was ist optimale Skalierung?

Der Sinn der optimalen Skalierung ist es, den Kategorien jeder Variablen numerische Quantifizierungen zuzuweisen und dadurch zu ermöglichen, dass Standardprozeduren verwendet werden können, um eine Lösung für die quantifizierten Variablen zu erhalten.

Die optimalen Skalenwerte werden den Kategorien der einzelnen Variablen anhand des Optimierungskriteriums für die verwendete Prozedur zugewiesen. Anders als die ursprünglichen Beschriftungen der nominalen oder ordinalen Variablen in der Analyse haben diese Skalenwerte metrische Eigenschaften.

In den meisten Prozeduren in Categories wird die optimale Quantifizierung für die einzelnen skalierten Variablen über eine iterative Methode, **alternierende kleinste Quadrate** genannt, erzielt, bei der zunächst die aktuellen Quantifizierungen verwendet werden, um eine Lösung zu finden, und dann die Quantifizierungen anhand dieser Lösung aktualisiert werden. Anschließend werden die aktualisierten Quantifizierungen verwendet, um eine neue Lösung zu finden, anhand derer wiederum die Quantifizierungen aktualisiert werden, usw. Dieser Vorgang wird fortgesetzt, bis ein Stoppkriterium erreicht wird.

Warum sollte optimale Skalierung verwendet werden?

Kategoriale Daten finden sich oft bei der Marktforschung, der Auswertung von Umfragen und in der Sozial- und Verhaltensforschung. Zahlreiche Forscher haben fast ausschließlich mit kategorialen Daten zu tun.

Es gibt zwar von den meisten Standardmodellen Anpassungen für die Analyse kategorialer Daten, doch diese funktionieren häufig nicht gut für Datensets mit folgenden Eigenschaften:

- Zu wenige Beobachtungen
- Zu viele Variablen
- Zu viele Werte pro Variable

Bei der Quantifizierung von Kategorien können Verfahren zur optimalen Skalierung die Probleme in diesen Fällen vermeiden. Außerdem sind sie selbst dann nützlich, wenn spezialisierte Verfahren angebracht sind.

Anstatt auf der Interpretation von Parameterschätzungen beruht die Interpretation der Ausgabe bei der optimalen Skalierung häufig auf grafischen Darstellungen. Verfahren zur optimalen Skalierung bieten hervorragende explorative Analysen, die eine gute Ergänzung zu anderen IBM SPSS Statistics-Modellen darstellen. Durch Einengung des Fokus der Untersuchung kann die Visualisierung der Daten durch optimale Skalierung die Grundlage einer Analyse bilden, die sich auf die Interpretation von Modellparametern konzentriert.

Messniveau der optimalen Skalierung

Dieses Konzept kann bei der Verwendung der Prozeduren von Categories zunächst sehr verwirrend sein. Bei der Angabe des Niveaus geben Sie nicht das Niveau an, auf dem die Variablen *gemessen* werden, sondern das Niveau, auf dem sie *skaliert* werden. Der Grund dafür ist, dass die zu quantifizierenden Variablen möglicherweise nicht lineare Relationen aufweisen, unabhängig davon, wie sie gemessen werden.

Für Categories gibt es drei grundlegende Messniveaus:

- Beim **nominalen** Niveau wird davon ausgegangen, dass die Werte einer Variablen für nicht geordnete Kategorien stehen. Beispiele für möglicherweise nominale Variablen sind Bundesland, Postleitzahl, Religionszugehörigkeit und Multiple-Choice-Kategorien.
- Beim **ordinalen** Niveau wird davon ausgegangen, dass die Werte einer Variablen für geordnete Kategorien stehen. Beispiele sind Einstellungsskalen, die für den Grad der Zufriedenheit oder der Zuversicht stehen, sowie Scores für Präferenzen.
- Beim **numerischen** Niveau wird davon ausgegangen, dass die Werte einer Variablen für Kategorien mit einer sinnvollen Metrik stehen, sodass man sinnvolle Aussagen über die Abstände zwischen den Kategorien machen kann. Metrische Variablen sind beispielsweise Alter (in Jahren) oder Einkommen (in Geldeinheiten).

Nehmen wir beispielsweise an, die Variablen *Region*, *Tätigkeit* und *Alter* seien wie in der folgenden Tabelle codiert.

Tabelle 1. Codierungsschema für *Region*, *Tätigkeit* und *Alter*

Regionscode	Regionswert	Tätigkeitscode	Tätigkeitswert	Alter
E	Norden	E	Praktikant	20
Z	Süden	Z	Vertreter	22
3	Osten	3	Manager	25
4	Westen			27

Die angezeigten Werte stehen für die Kategorien der einzelnen Variablen. *Region* ist eine nominale Variable. Die Variable *Region* weist vier Kategorien ohne implizierte Reihenfolge auf. Die Werte 1 bis 4 stehen einfach für die vier Kategorien; das Codierungsschema ist völlig willkürlich. *Tätigkeit* dagegen könnte als ordinale Variable betrachtet werden. Die ursprünglichen Kategorien stellen eine Steigerung vom Praktikanten bis hin zum Manager dar. Höhere Codewerte stehen für eine Tätigkeit, die auf der innerbetrieblichen Karriereleiter weiter oben liegt. Es ist jedoch nur die Reihenfolge bekannt. Über den Abstand zwischen benachbarten Kategorien wird nichts ausgesagt. Im Gegensatz dazu kann *Alter* als numerische Variable betrachtet werden. Bei *Alter* haben die Abstände zwischen den Werten eine implizite Bedeutung. Der Abstand zwischen 20 und 22 ist genauso groß wie der Abstand zwischen 25 und 27, wohingegen der Abstand zwischen 22 und 25 größer als die beiden anderen ist.

Auswahl des Messniveaus der optimalen Skalierung

Es ist wichtig zu begreifen, dass es keine intrinsischen Eigenschaften einer Variablen gibt, die automatisch vordefinieren, welches Messniveau der optimalen Skalierung dafür angegeben werden sollte. Sie können die Daten so analysieren, wie es sinnvoll erscheint und die Interpretation erleichtert. Bei der Analyse einer Variablen mit eigentlich numerischem Niveau auf dem ordinalen Niveau kann durch die Verwendung einer nicht linearen Transformation eine Lösung in weniger Dimensionen möglich sein.

Die folgenden zwei Beispiele zeigen, dass das naheliegendste Messniveau nicht unbedingt das beste Niveau für die optimale Skalierung sein muss. Angenommen, eine Variable ordnet Objekte in Altersgruppen ein. Alter kann zwar als numerische Variable skaliert werden, aber dennoch steht Sicherheit für Personen

unter 25 Jahren möglicherweise in einer positiven Relation zum Alter, wohingegen sie für Personen über 60 Jahren in einer negativen Relation zum Alter steht. In diesem Fall kann es besser sein, das Alter als nominale Variable zu behandeln.

Zweites Beispiel: Eine Variable, die Personen nach politischen Neigungen ordnet, scheint ihrem Wesen nach nominal zu sein. Wenn Sie jedoch die Parteien von politisch links nach politisch rechts sortieren, wäre es sinnvoll, wenn die Quantifizierung der Parteien diese Reihenfolge berücksichtigen würde. Dazu müsste ein ordinales Niveau für die Analyse verwendet werden.

Auch wenn es keine vordefinierten Eigenschaften gibt, die eine Variable ausschließlich einem bestimmten Messniveau zuordnen, gibt es einige allgemeine Richtlinien, nach denen sich unerfahrene Benutzer richten können. Bei einzeln nominaler Quantifizierung kennen Sie normalerweise nicht die Reihenfolge der Kategorien, möchten jedoch, dass durch die Analyse eine festgelegt wird. Wenn die Reihenfolge der Kategorien bekannt ist, sollten Sie eine ordinale Quantifizierung verwenden. Wenn sich die Kategorien nicht ordnen lassen, sollten Sie eine mehrfach nominale Quantifizierung verwenden.

Transformationsdiagramme

Die verschiedenen Niveaus, auf denen die einzelnen Variablen skaliert werden können, bringen verschiedene Einschränkungen für die Quantifizierungen mit sich. Transformationsdiagramme illustrieren die Beziehung zwischen den Quantifizierungen und den ursprünglichen Kategorien, die sich aus dem jeweils ausgewählten Messniveau der optimalen Skalierung ergibt. So ergibt sich beispielsweise ein lineares Transformationsdiagramm, wenn eine Variable als numerisch behandelt wird. Als ordinal behandelte Variablen führen zu einem nicht fallenden Transformationsdiagramm. Transformationsdiagramme für als nominal behandelte Variablen, die die Form eines "U" oder eines umgedrehten "U" aufweisen, stellen eine quadratische Beziehung dar. Nominale Variablen können auch Transformationsdiagramme ohne offensichtliche Trends ergeben, indem die Reihenfolge der Kategorien vollständig geändert wird. Folgende Abbildung zeigt ein Beispieltransformationsdiagramm.

Transformationsdiagramme sind besonders gut geeignet, um zu bestimmen, wie gut das ausgewählte Messniveau für die optimale Skalierung funktioniert. Wenn mehrere Kategorien ähnliche Quantifizierungen erhalten, kann die Zusammenfassung dieser Kategorien zu einer einzigen Kategorie gerechtfertigt sein. Wenn umgekehrt eine als nominal behandelte Variable Quantifizierungen erhält, die einen steigenden Trend anzeigen, kann eine ordinale Transformation zu einer ähnlichen Anpassung führen. Wenn dieser Trend linear ist, ist möglicherweise eine numerische Behandlung geeignet. Wenn jedoch die Zusammenfassung von Kategorien oder die Änderung der Messniveaus für die Skalierung angebracht ist, ändert sich die Analyse nicht in signifikanter Weise.

Kategoriecodes

Bei der Codierung kategorialer Variablen sollte mit Vorsicht vorgegangen werden, da einige Codierungsschemas zu einer unerwünschten Datenausgabe oder zu unvollständigen Analysen führen können. Mögliche Codierungsschemas für *Tätigkeit* werden in der folgenden Tabelle angezeigt.

Tabelle 2. Alternative Codierungsschemas für die Tätigkeit

Kategorie	A	B	C	D
Praktikant	E	E	5	E
Vertreter	Z	Z	6	5
Manager	3	7	7	3

Für einige Prozeduren von Categories muss der Bereich jeder verwendeten Variablen definiert werden. Werte außerhalb dieses Bereichs werden als fehlende Werte betrachtet. Der minimale Kategoriewert ist immer 1. Der maximale Kategoriewert wird vom Benutzer angegeben. Dieser Wert gibt nicht die *Anzahl* der Kategorien für eine Variable an, sondern den *größten* Kategoriewert. In der Tabelle beispielsweise

weist Schema A einen maximalen Kategoriewert von 3 und Schema B einen maximalen Kategoriewert von 7 auf, obwohl beide Schemas dieselben drei Kategorien codieren.

Der Variablenbereich bestimmt, welche Kategorien aus der Analyse ausgeschlossen werden. Alle Kategorien mit Codes außerhalb des definierten Bereichs werden aus der Analyse ausgeschlossen. Dies ist eine einfache Methode für den Ausschluss von Kategorien, die jedoch zu unerwünschten Analysen führen kann. Wenn der maximale Kategoriewert falsch definiert ist, werden möglicherweise *gültige* Kategorien aus der Analyse ausgeschlossen. Beispiel: Bei Schema B wird durch die Festsetzung des maximalen Kategoriewerts auf den Wert 3 angegeben, dass für *Tätigkeit* Kategorien von 1 bis 3 codiert sind. Die Kategorie *Manager* wird als fehlend behandelt. Da tatsächlich keine Kategorie als 3 codiert wurde, enthält die dritte Kategorie in der Analyse keine Fälle. Wenn Sie alle Managerkategorien ausschließen möchten, ist diese Analyse geeignet. Wenn die Manager eingeschlossen sein sollen, muss der maximale Kategoriewert als 7 definiert sein, und fehlende Werte müssen mit Werten über 7 oder unter 1 codiert sein.

Bei als nominal oder ordinal behandelten Variablen hat die Reihenfolge der Kategorien keine Auswirkungen auf die Ergebnisse. Bei nominalen Variablen ist nur die Beschriftung und nicht der der Beschriftung zugeordnete Wert von Bedeutung. Bei ordinalen Variablen wird die Reihenfolge der Kategorien in den Quantifizierungen beibehalten; die Kategoriewerte selbst sind nicht von Bedeutung. Alle Codierungsschemas, die zur selben Kategoriereihenfolge führen, weisen identische Ergebnisse auf. Beispiel: Die ersten drei Schemas in der Tabelle sind funktional äquivalent, wenn *Tätigkeit* auf ordinaler Ebene analysiert wird. Die Reihenfolge der Kategorien ist in diesen Schemas identisch. Bei Schema D dagegen werden die zweite und die dritte Kategorie invertiert, und es ergeben sich andere Ergebnisse als bei den anderen Schemas.

Obwohl viele Codierungsschemas für eine Variable funktional äquivalent sind, werden Schemas mit kleinen Werten für die Differenz zwischen den Codes bevorzugt, da die Codes sich auf die Menge an Ausgabedaten auswirken, die sich aus der Prozedur ergeben. Alle Kategorien, die mit Werten zwischen 1 und dem benutzerdefinierten Höchstwert codiert sind, sind gültig. Wenn eine dieser Kategorien leer ist, sind die entsprechenden Quantifizierungen je nach Prozedur entweder systemdefiniert fehlend oder 0. Obwohl sich keine dieser Zuordnungen auf die Analysen auswirkt, wird für die betreffenden Kategorien eine Ausgabe erstellt. So weist *Tätigkeit* bei Schema B vier Kategorien auf, die systemdefiniert fehlende Werte erhalten. Bei Schema C gibt es außerdem vier Kategorien, die Indikatoren für systemdefiniert fehlende Werte erhalten. Im Gegensatz dazu gibt es bei Schema A keine systemdefiniert fehlenden Quantifizierungen. Die Verwendung aufeinander folgender Ganzzahlen als Codes für als nominal oder ordinal behandelte Variablen führt zu einer deutlich kleineren Ausgabe, ohne die Ergebnisse zu beeinträchtigen.

Die Codierungsschemas für Variablen, die als numerisch behandelt werden, unterliegen größeren Einschränkungen, als dies bei ordinalen Variablen der Fall ist. Bei diesen Variablen ist die Differenz zwischen aufeinander folgenden Kategorien von Bedeutung. In der folgenden Tabelle werden drei Codierungsschemas für *Alter* angezeigt.

Tabelle 3. Alternative Codierungsschemas für das Alter

Kategorie	A	B	C
20	20	E	E
22	22	3	Z
25	25	6	3
27	27	8	4

Bei allen Aufzeichnungen numerischer Variablen müssen die Differenzen zwischen den Kategorien erhalten bleiben. Die Verwendung der ursprünglichen Werte ist eine Methode, um die Beibehaltung der Differenzen zu gewährleisten. Dies kann jedoch dazu führen, dass viele Kategorien Indikatoren für systemdefiniert fehlende Werte aufweisen. Beispiel: Schema A verwendet die ursprünglich beobachteten Werte. Bei allen Prozeduren von Categories mit Ausnahme der Korrespondenzanalyse ist der maximale Kategoriewert 27, und der minimale Kategoriewert ist auf 1 gesetzt. Die ersten 19 Kategorien sind leer und erhal-

ten Indikatoren für systemdefiniert fehlende Werte. Die Ausgabe kann schnell sehr umfangreich und komplex werden, wenn der maximale Kategoriewert wesentlich größer als 1 ist und sehr viele leere Kategorien zwischen 1 und dem Höchstwert liegen.

Die Größe der Datenausgabe kann durch Umcodierung verringert werden. Bei numerischen Variablen sollte die Funktion für die automatische Umcodierung jedoch nicht verwendet werden. Die Codierung in aufeinander folgende Ganzzahlen führt zur Differenz 1 zwischen allen aufeinander folgenden Kategorien und damit dazu, dass alle Quantifizierungen dieselben Abstände aufweisen. Die metrischen Merkmale, die bei der Behandlung einer Variablen als nominal für bedeutsam erachtet werden, werden durch die Neucodierung in aufeinander folgende Ganzzahlen zerstört. Beispiel: Schema C in der Tabelle entspricht der automatischen Umcodierung von *Alter*. Die Differenz zwischen den Kategorien 22 und 25 hat sich von 3 auf 1 geändert, und die Quantifizierungen spiegeln die neue Differenz wider.

Ein alternatives Umcodierungsschema, bei dem die Differenzen zwischen den Kategorien erhalten bleiben, besteht darin, den kleinsten Kategoriewert von jeder Kategorie zu subtrahieren und zu der Differenz den Wert 1 zu addieren. Schema B ergibt sich aus dieser Transformation. Der kleinste Kategoriewert, 20, wurde von jeder Kategorie subtrahiert, und zu dem Ergebnis wurde der Wert 1 addiert. Die transformierten Codes weisen mindestens den Wert 1 auf, und alle Differenzen sind mit denen der ursprünglichen Daten identisch. Der maximale Kategoriewert ist nun 8, und die Nullquantifizierungen vor der ersten Nicht-Nullquantifizierung werden gelöscht. Allerdings sind die Nicht-Nullquantifizierungen für die einzelnen Kategorien, die sich aus Schema B ergeben, mit den Quantifizierungen aus Schema A identisch.

Welche Prozedur ist für Ihre Anwendung am besten geeignet?

Die in vier dieser Prozeduren (Korrespondenzanalyse, Mehrfachkorrespondenzanalyse, kategoriale Hauptkomponentenanalyse und nicht lineare kanonische Korrelationsanalyse) enthaltenen Verfahren fallen in den allgemeinen Bereich der multivariaten Datenanalyse, die als **Dimensionsreduktion** bekannt ist. Dies bedeutet, dass die Beziehungen zwischen den Variablen so oft wie möglich in wenigen Dimensionen, beispielsweise zwei oder drei, dargestellt werden. Dadurch können Sie Strukturen oder Muster in den Beziehungen beschreiben, die in ihrer ursprünglichen Komplexität schwer zu ergründen wären. Bei Marktforschungsanwendungen können diese Verfahren eine Art **Wahrnehmungszuordnung** sein. Ein wichtiger Vorteil dieser Prozeduren besteht darin, dass sie mit Daten mit verschiedenen Messniveaus der optimalen Skalierung umgehen können.

Die kategoriale Regression beschreibt die Beziehung zwischen einer kategorialen Antwortvariablen und einer Kombination kategorialer Prädiktorvariablen. Der Einfluss jeder Prädiktorvariablen auf die Antwortvariable wird durch die entsprechende Regressionsgewichtung beschrieben. Wie in den anderen Prozeduren können die Daten mit verschiedenen Messniveaus der optimalen Skalierung analysiert werden.

Mit multidimensionaler Skalierung und multidimensionaler Entfaltung werden Beziehungen zwischen Objekten in einem Raum mit wenigen Dimensionen anhand der Ähnlichkeiten zwischen den Objekten beschrieben.

Es folgen kurze Richtlinien für die einzelnen Prozeduren:

- Verwenden Sie die kategoriale Regression, um die Werte einer abhängigen kategorialen Variablen aus einer Kombination unabhängiger kategorialer Variablen vorherzusagen.
- Verwenden Sie die kategoriale Hauptkomponentenanalyse, um Variationsmuster in einem einzelnen Set von Variablen mit gemischten Messniveaus der optimalen Skalierung zu berücksichtigen.
- Verwenden Sie die nicht lineare kanonische Korrelationsanalyse, um zu bewerten, wie stark zwei oder mehr Sets von Variablen mit gemischten Messniveaus der optimalen Skalierung korrelieren.
- Verwenden Sie die Korrespondenzanalyse, um zweidimensionale Kontingenztabellen oder Daten zu analysieren, die als Zweifachtafel ausgedrückt werden können, beispielsweise Daten zur bevorzugten Marke oder zur soziometrischen Wahl.

- Verwenden Sie die Mehrfachkorrespondenzanalyse zur Analyse einer Matrix mit kategorialen multivariaten Daten, wenn Sie von keiner stärkeren Annahme ausgehen möchten als von der, dass alle Variablen auf dem nominalen Niveau analysiert werden.
- Verwenden Sie die multidimensionale Skalierung zur Analyse von Ähnlichkeitsdaten, um eine Darstellung eines einzelnen Sets von Objekten in einem schwachdimensionierten Raum mithilfe der Methode der kleinsten Quadrate zu finden.
- Verwenden Sie die multidimensionale Entfaltung zur Analyse von Ähnlichkeitsdaten, um eine Darstellung zweier Sets von Objekten in einem schwachdimensionierten Raum mithilfe der Methode der kleinsten Quadrate zu finden.

Kategoriale Regression

Die Verwendung der kategorialen Regression ist am besten geeignet, wenn das Ziel der Analyse darin besteht, eine abhängige (Antwort-)Variable aus einem Set unabhängiger (Prädiktor-)Variablen vorherzusagen. Wie bei allen Prozeduren für die optimale Skalierung werden jeder Kategorie jeder Variablen Skalenniveaus zugewiesen, dergestalt, dass diese Werte in Bezug auf die Regression optimal sind. Die Lösung einer kategorialen Regression maximiert die quadrierte Korrelation zwischen der transformierten Antwortvariablen und der gewichteten Kombination transformierter Prädiktoren.

Beziehung zu anderen Prozeduren von Categories. Die kategoriale Regression mit optimaler Skalierung ist vergleichbar mit der kanonischen Korrelationsanalyse mit optimaler Skalierung und zwei Sets, von denen ein Set jeweils nur eine Variable enthält. Bei letzterem Verfahren wird die Ähnlichkeit der Sets abgeleitet, indem jedes Set mit einer unbekanntem Variablen verglichen wird, die irgendwo zwischen allen Sets liegt. Bei der kategorialen Regression wird die Ähnlichkeit der transformierten Antwortvariablen und der gewichteten Kombination transformierter Prädiktoren unmittelbar bewertet.

Beziehung zu Standardverfahren. Beim Standardverfahren der linearen Regression können kategoriale Variablen entweder als Indikatorvariablen umcodiert oder genauso wie Variablen auf Intervallniveau behandelt werden. Beim ersten Ansatz enthält das Modell eine gesonderte Konstante und Steigung für die einzelnen Niveauekombinationen für die kategorialen Variablen. Dies führt zu einer großen Anzahl zu interpretierender Parameter. Im zweiten Ansatz wird für jede Variable nur ein einziger Parameter geschätzt. Durch die willkürliche Festlegung der Kategoriecodierungen werden jedoch Verallgemeinerungen unmöglich.

Wenn einige Variablen nicht stetig sind, stehen alternative Analysen zur Verfügung. Bei stetiger Antwortvariablen und kategorialen Prädiktoren wird häufig eine Varianzanalyse durchgeführt. Bei kategorialer Antwortvariablen und stetigen Prädiktoren kann eine logistische Regression oder eine Diskriminanzanalyse angemessen sein. Wenn sowohl die Antwortvariable als auch die Prädiktoren kategorial sind, werden häufig loglineare Modelle verwendet.

Die Regression mit optimaler Skalierung bietet drei Messniveaus für die Skalierung der einzelnen Variablen. Durch Kombinationen dieser Niveaus kann eine große Bandbreite an nicht linearen Beziehungen behandelt werden, für die eine einzelne "Standard"-Methode nicht geeignet ist. Folglich bietet die optimale Skalierung eine größere Flexibilität als die Standardansätze bei kaum höherer Komplexität.

Außerdem werden durch nicht lineare Transformationen der Prädiktoren normalerweise die Abhängigkeiten zwischen den einzelnen Prädiktoren verringert. Wenn Sie die Eigenwerte der Korrelationsmatrix für die Prädiktoren mit den Eigenwerten der Korrelationsmatrix für die optimal skalierten Prädiktoren vergleichen, ist die zweite Gruppe zumeist weniger variabel als die erste. Mit anderen Worten: bei der kategorialen Regression werden durch die optimale Skalierung die großen Eigenwerte der Korrelationsmatrix für die Prädiktoren kleiner und die kleinen Eigenwerte größer.

Kategoriale Hauptkomponentenanalyse

Die kategoriale Hauptkomponentenanalyse ist am besten geeignet, wenn Variationsmuster in einem einzelnen Set von Variablen mit gemischten Messniveaus der optimalen Skalierung berücksichtigt werden

sollen. Dieses Verfahren versucht, die Dimensionalität eines Variablensets zu verringern und dabei gleichzeitig die Variation soweit wie möglich zu berücksichtigen. Jeder Kategorie jeder Variablen werden Skalenwerte zugewiesen, dergestalt, dass diese Werte in Bezug auf die Hauptkomponentenlösung optimal sind. Die Objekte in der Analyse erhalten Komponentenscores auf der Grundlage der quantifizierten Daten. Diagramme der Komponentenscores lassen Muster in Bezug auf die Objekte in der Analyse sichtbar werden und können ungewöhnliche Objekte in den Daten kenntlich machen. Die Lösung einer kategorialen Hauptkomponentenanalyse maximiert die Korrelationen der Objektscores mit den einzelnen quantifizierten Variablen für die Anzahl der angegebenen Komponenten (Dimensionen).

Eine wichtige Anwendung der kategorialen Hauptkomponentenanalyse ist die Untersuchung von Präferenzdaten, bei der die Befragten eine Reihe von Items nach Ihrer Präferenz einstufen oder bewerten. Bei der üblichen IBM SPSS Statistics-Datenkonfiguration stellen die Zeilen Einzelwerte und die Spalten Maße für die Items dar. Die zeilenübergreifenden Scores sind Präferenzscores (beispielsweise auf einer Skala von 0 bis 10). Dadurch werden die Daten zeilenabhängig. Bei Präferenzdaten kann es sinnvoll sein, die Einzelwerte als Variablen zu behandeln. Mit der Prozedur "Transponieren" können Sie die Daten transponieren. Die Bewerter werden die Variablen, und alle Variablen werden als ordinal deklariert. In CATPCA können problemlos mehr Variablen als Objekte verwendet werden.

Beziehung zu anderen Prozeduren von Categories. Wenn alle Variablen als mehrfach nominal deklariert werden, führt die kategoriale Hauptkomponentenanalyse zu einer Analyse, die einer Mehrfachkorrespondenzanalyse, die für dieselben Variablen durchgeführt wurde, entspricht. Daher kann die kategoriale Hauptkomponentenanalyse als eine Art Mehrfachkorrespondenzanalyse betrachtet werden, bei der einige der Variablen als ordinal oder numerisch deklariert sind.

Beziehung zu Standardverfahren. Wenn alle Variablen auf dem numerischen Niveau skaliert sind, entspricht die kategoriale Hauptkomponentenanalyse der normalen Hauptkomponentenanalyse.

Allgemeiner gesagt: Die kategoriale Hauptkomponentenanalyse ist eine Alternative zur Berechnung der Korrelationen zwischen nicht numerischen Skalen und ihrer Analyse mithilfe der normalen Hauptkomponentenanalyse oder der Faktoranalyse. Die unreflektierte Verwendung des üblichen Korrelationskoeffizienten nach Pearson als Zusammenhangsmaß für ordinale Daten kann zu einer nicht trivialen Verzerrung in der Schätzung der Korrelationen führen.

Nicht lineare kanonische Korrelationsanalyse

Die nicht lineare kanonische Korrelationsanalyse ist eine sehr allgemeine Prozedur mit vielen verschiedenen Anwendungen. Ziel der nicht linearen kanonischen Korrelationsanalyse ist die Analyse der Beziehungen zwischen zwei oder mehr Sets von Variablen, anstatt zwischen den Variablen selbst (wie bei der Hauptkomponentenanalyse). So können beispielsweise zwei Variablensets vorliegen, wobei es sich bei einem Variablenset um Items zum demografischen Hintergrund für eine Gruppe von Befragten handelt, während das zweite Set von Antworten auf eine Gruppe von Items zur Einstellung der Personen besteht. Die Skalierungsniveaus in der Analyse können aus einer beliebigen Mischung aus nominalen, ordinalen und numerischen Messniveaus bestehen. Die kanonische Korrelationsanalyse mit optimaler Skalierung bestimmt die Ähnlichkeit zwischen den Sets und vergleicht gleichzeitig die kanonischen Variablen aus den einzelnen Sets mit einem Kompromissset, das aus den Scores besteht, die den Objekten zugewiesen wurden.

Beziehung zu anderen Prozeduren von Categories. Wenn zwei oder mehr Variablensets mit jeweils nur einer einzigen Variablen vorliegen, entspricht die kanonische Korrelationsanalyse mit optimaler Skalierung der Hauptkomponentenanalyse mit optimaler Skalierung. Wenn alle Variablen in einer Analyse mit einer Variablen pro Set mehrfach nominal sind, entspricht die kanonische Korrelationsanalyse mit optimaler Skalierung der Mehrfachkorrespondenzanalyse. Bei zwei Variablensets, von denen eine nur eine einzige Variable enthält, entspricht die kanonische Korrelationsanalyse mit optimaler Skalierung der kategorialen Regression mit optimaler Skalierung.

Beziehung zu Standardverfahren. Die kanonische Standardkorrelationsanalyse ist ein statistisches Verfahren, bei der eine lineare Kombination eines Variablensets und eine lineare Kombination eines zweiten Variablensets ermittelt werden, die maximal korreliert sind. Ausgehend von diesem Set mit linearen Kombinationen kann die kanonische Korrelationsanalyse darauf folgende unabhängige Sets mit linearen Kombinationen ermitteln (als kanonische Variablen bezeichnet), bis zu einer Maximalzahl, die der Anzahl der Variablen im kleineren Set entspricht.

Wenn in der Analyse zwei Variablensets vorliegen und alle Variablen als numerisch definiert sind, entspricht die kanonische Korrelationsanalyse mit optimaler Skalierung einer kanonischen Standardkorrelationsanalyse. IBM SPSS Statistics beinhaltet zwar keine Prozedur für die kanonische Korrelationsanalyse, doch viele der relevanten Statistiken können aus einer multivariaten Varianzanalyse gewonnen werden.

Die kanonische Korrelationsanalyse mit optimaler Skalierung bietet noch viele weitere Anwendungsmöglichkeiten. Bei zwei Variablensets, von denen eines eine nominale Variable enthält, die als einfach nominal deklariert ist, können die Ergebnisse der kanonischen Korrelationsanalyse mit optimaler Skalierung ähnlich wie die Ergebnisse einer Regressionsanalyse interpretiert werden. Wenn Sie die Variable als mehrfach nominal betrachten, ist die Analyse mit optimaler Skalierung eine Alternative zur Diskriminanzanalyse. Durch die Gruppierung der Variablen in mehr als zwei Sets erhalten Sie eine Vielzahl von Möglichkeiten zur Analyse der Daten.

Korrespondenzanalyse

Ziel der Korrespondenzanalyse ist die Erstellung von Biplots für Korrespondenztafeln. In einer Korrespondenztabelle wird davon ausgegangen, dass die Zeilen- und Spaltenvariablen ungeordnete Kategorien darstellen; daher wird hier immer das nominale Messniveau der optimalen Skalierung verwendet. Beide Variablen werden hinsichtlich ihrer nominalen Informationen untersucht. Es wird also nur die Tatsache berücksichtigt, dass sich einige Objekte in derselben Kategorie befinden, andere jedoch nicht. Über die Distanz oder die Reihenfolge der verschiedenen Kategorien derselben Variablen werden keine Annahmen gemacht.

Eine spezielle Verwendung der Korrespondenzanalyse besteht in der Analyse von zweidimensionalen Kontingenztafeln. Wenn eine Tabelle r aktive Zeilen und c aktive Spalten aufweist, beträgt die Anzahl der Dimensionen in der Korrespondenzanalyse mindestens d r minus 1 bzw. c minus 1 (der kleinere der beiden Werte wird verwendet). Anders ausgedrückt: Die Zeilen- oder Spaltenkategorien einer Kontingenztafel könnten perfekt in einem mehrdimensionalen Raum dargestellt werden. In der Praxis würde man jedoch die Zeilen- und Spaltenkategorien einer Zweiwegetafel in einem Raum mit wenigen, beispielsweise zwei, Dimensionen darstellen wollen, da zweidimensionale Diagramme leichter verständlich sind als mehrdimensionale räumliche Darstellungen.

Wenn weniger als die Höchstzahl der möglichen Dimensionen verwendet werden, beschreibt die in der Analyse erstellte Statistik, wie gut die Zeilen- und Spaltenkategorien in der flachdimensionierten Darstellung repräsentiert werden. Vorausgesetzt, die Darstellung der zweidimensionalen Lösung weist eine gute Qualität auf, können Sie Plots der Zeilenpunkte und der Spaltenpunkte untersuchen, um herauszufinden, welche Kategorien der Zeilenvariablen ähnlich sind, welche Kategorien der Spaltenvariablen ähnlich sind und welche Zeilen- und Spaltenkategorien einander ähnlich sind.

Beziehung zu anderen Prozeduren von Categories. Die einfache Korrespondenzanalyse ist auf Zweiwegetafeln beschränkt. Liegen mehr als zwei relevante Variablen vor, können Sie die Variablen zu Interaktionsvariablen zusammenfassen. Bei den Variablen *Region*, *Tätigkeit* und *Alter* beispielsweise könnten *Region* und *Tätigkeit* zu der neuen Variablen *Regtätigkeit* zusammengefasst werden, die, wie in der untenstehenden Tabelle gezeigt, 12 Kategorien aufweist. Diese neue Variable führt zu einer Zweiwegetafel mit *Alter* (12 Zeilen, 4 Spalten), die mithilfe der Korrespondenzanalyse analysiert werden kann.

Tabelle 4. Kombinationen von *Region* und *Tätigkeit*

Kategoriecode	Kategoriedefinition	Kategoriecode	Kategoriedefinition
E	Norden, Praktikant	7	Osten, Praktikant

Tabelle 4. Kombinationen von Region und Tätigkeit (Forts.)

Kategoriecode	Kategoriedefinition	Kategoriecode	Kategoriedefinition
Z	Norden, Vertreter	8	Osten, Vertreter
3	Norden, Manager	9	Osten, Manager
4	Süden, Praktikant	10	Westen, Praktikant
5	Süden, Vertreter	11	Westen, Vertreter
6	Süden, Manager	12	Westen, Manager

Ein Nachteil bei diesem Ansatz besteht darin, dass beliebige Variablenpaare zusammengefasst werden können. *Tätigkeit* und *Alter* können kombiniert werden, was ebenfalls eine Variable mit 12 Kategorien ergibt. Oder man kann *Region* und *Alter* kombinieren, was zu einer neuen Variablen mit 16 Kategorien führt. Alle diese Interaktionsvariablen ergeben zusammen mit der verbleibenden Variablen eine Zweifachtafel. Die Korrespondenzanalysen dieser drei Tabellen führen nicht zu identischen Ergebnissen, aber dennoch ist jeder der Ansätze gültig. Außerdem können bei vier oder mehr Variablen Zweifachtafeln erstellt werden, die eine Interaktionsvariable mit einer weiteren Interaktionsvariablen vergleichen. Die Anzahl der möglichen Tabellen für die Analyse kann recht groß werden, selbst bei nur relativ wenigen Variablen. Sie können eine dieser Tabellen für die Analyse auswählen oder alle analysieren. Alternativ kann die Prozedur der Mehrfachkorrespondenzanalyse verwendet werden, um alle Variablen gleichzeitig zu untersuchen, ohne dass Interaktionsvariablen konstruiert werden müssen.

Beziehung zu Standardverfahren. Mit der Prozedur Crosstabs können ebenfalls Kontingenztafeln analysiert werden, wobei Unabhängigkeit einen gemeinsamen Fokus der Analysen darstellt. Selbst bei kleinen Tabellen kann es jedoch schwierig sein, die Ursache von Abweichungen von der Unabhängigkeit zu ermitteln. Der Nutzen der Korrespondenzanalyse besteht in der Anzeige derartiger Muster für Zweifachtafeln beliebiger Größe. Wenn ein Zusammenhang zwischen den Zeilen- und den Spaltenvariablen besteht, d. h., wenn der Chi-Quadrat-Wert signifikant ist, kann die Korrespondenzanalyse die Ermittlung der Art dieser Beziehung erleichtern.

Mehrfachkorrespondenzanalyse

Bei der Mehrfachkorrespondenzanalyse wird versucht, eine Lösung zu erstellen, bei der Objekte innerhalb derselben Kategorie eng zusammen und Objekte in verschiedenen Kategorien in großem Abstand zueinander geplottet werden. Jedes Objekt liegt so nah wie möglich an den Kategoriepunkten der Kategorien, die für dieses Objekt gelten. So unterteilen die Kategorien die Objekte in homogene Untergruppen. Variablen werden als homogen betrachtet, wenn sie Objekte, die in denselben Kategorien vorliegen, auch in dieselben Untergruppen einordnen.

Bei einer eindimensionalen Lösung weist die Mehrfachkorrespondenzanalyse jeder Kategorie jeder Variablen optimale Skalenwerte (Kategoriequantifizierungen) zu, dergestalt, dass die Kategorien insgesamt im Durchschnitt die maximale Streubreite aufweisen. Bei einer zweidimensionalen Lösung ermittelt die Mehrfachkorrespondenzanalyse ein zweites Set von Quantifizierungen der Kategorien jeder Variablen, das nicht mit dem ersten Set in Zusammenhang steht, wobei wieder versucht wird, die Streubreite zu maximieren, usw. Da die Kategorien einer Variablen so viele Scorings erhalten, wie Dimensionen vorhanden sind, wird davon ausgegangen, dass die Variablen in der Analyse auf dem Messniveau der optimalen Skalierung mehrfach nominal sind.

Die Mehrfachkorrespondenzanalyse weist auch den Objekten in der Analyse solche Werte zu, dass die Kategoriequantifizierungen die Durchschnittswerte bzw. Zentroide der Objektscores der Objekte in dieser Kategorie sind.

Beziehung zu anderen Prozeduren von Categories. Die Mehrfachkorrespondenzanalyse wird auch als Homogenitätsanalyse oder Doppelskalierung bezeichnet. Sie führt zu vergleichbaren (jedoch nicht identischen) Lösungen wie die Korrespondenzanalyse, wenn nur zwei Variablen vorhanden sind. Die Korres-

pondenzanalyse führt zu einer eindeutigen Ausgabe, bei der die Anpassung und die Qualität der Darstellung der Lösung, einschließlich der Stabilitätsinformationen, zusammengefasst werden. Daher ist bei zwei Variablen die Korrespondenzanalyse normalerweise der Mehrfachkorrespondenzanalyse vorzuziehen. Ein weiterer Unterschied zwischen den beiden Prozeduren besteht darin, dass es sich bei der Eingabe für die Mehrfachkorrespondenzanalyse um eine Datenmatrix handelt, bei der die Zeilen Objekte und die Spalten Variablen sind, wohingegen es sich bei der Eingabe für die Korrespondenzanalyse um dieselbe Datenmatrix, eine allgemeine Ähnlichkeitsmatrix oder eine verbundene Kontingenztabelle handeln kann, die eine aggregierte Matrix ist, bei der sowohl die Zeilen als auch die Spalten Kategorien von Variablen repräsentieren. Die Mehrfachkorrespondenzanalyse kann auch als Hauptkomponentenanalyse von Daten betrachtet werden, die auf dem mehrfach nominalen Niveau skaliert wurden.

Beziehung zu Standardverfahren. Die Mehrfachkorrespondenzanalyse kann als Analyse einer Mehrwegkontingenztabelle betrachtet werden. Mehrdimensionale Kontingenztabelle können auch mit der Prozedur Crosstabs analysiert werden, doch Crosstabs gibt getrennte Zusammenfassungen für die einzelnen Kategorien jeder Kontrollvariablen aus. Bei der Mehrfachkorrespondenzanalyse kann häufig die Beziehung zwischen allen Variablen in einem einzelnen zweidimensionalen Diagramm zusammengefasst werden. Eine erweiterte Verwendung der Mehrfachkorrespondenzanalyse besteht darin, die ursprünglichen Kategoriewerte mit den optimalen Skalenwerten aus der ersten Dimension zu ersetzen und eine sekundäre multivariate Analyse durchzuführen. Da die Mehrfachkorrespondenzanalyse die Kategoriebeschriftungen durch numerische Skalenwerte ersetzt, können nach der Mehrfachkorrespondenzanalyse viele verschiedene Prozeduren angewendet werden, für die numerische Daten erforderlich sind. Die Prozedur "Faktorenanalyse" beispielsweise erstellt zunächst eine erste Hauptkomponente, die der ersten Dimension der Mehrfachkorrespondenzanalyse entspricht. Die Komponentenscores in der ersten Dimension sind gleich den Objektscores, und die quadrierten Komponentenladungen sind gleich den Diskriminationsmaßen. Die zweite Dimension der Mehrfachkorrespondenzanalyse jedoch ist nicht gleich der zweiten Dimension der Faktorenanalyse.

Multidimensionale Skalierung

Die multidimensionale Skalierung ist am besten geeignet, wenn das Ziel der Analyse darin besteht, die Struktur eines Sets von Distanzmaßen innerhalb eines einzelnen Sets von Objekten oder Fällen zu ermitteln. Dies wird durch das Zuweisen von Beobachtungen zu bestimmten Positionen in einem konzeptuellen Raum mit wenigen Dimensionen erzielt, wobei die Abstände zwischen den Punkten des Raumes mit den vorgegebenen Unähnlichkeiten bzw. Ähnlichkeiten so gut wie möglich übereinstimmen. Als Ergebnis werden die Objekte in diesem Raum mit wenigen Dimensionen mithilfe der Methode der kleinsten Quadrate dargestellt, was häufig zu einem besseren Verständnis der Daten beiträgt.

Beziehung zu anderen Prozeduren von Categories. Wenn multivariate Daten vorliegen, aus denen Sie Distanzen erstellen und anschließend mit multidimensionaler Skalierung analysieren, erhalten Sie ein ähnliches Ergebnis wie bei der Analyse der Daten mithilfe der kategorialen Hauptkomponentenanalyse mit Objektprinzipalnormalisierung. Diese Art von PCA ist auch als Hauptkoordinatenanalyse bekannt.

Beziehung zu Standardverfahren. Die Prozedur "Multidimensionale Skalierung" in SPSS Categories (PROXSCAL) bietet zahlreiche Verbesserungen gegenüber der in der Statistics Base verfügbaren Skalierungsprozedur (ALSCAL). PROXSCAL bietet einen schnelleren Algorithmus für bestimmte Modelle und ermöglicht Ihnen, Einschränkungen für den gemeinsamen Raum festzulegen. Außerdem versucht PROXSCAL, den normalisierten Rohstress zu minimieren und nicht den S-Stress (auch als **Belastung** bezeichnet). Der normalisierte Rohstress wird im Allgemeinen bevorzugt, da es sich dabei um ein Maß handelt, das auf den Distanzen beruht, während S-Stress auf den quadrierten Distanzen beruht.

Multidimensionale Entfaltung

Die multidimensionale Entfaltung ist am besten geeignet, wenn das Ziel der Analyse darin besteht, die Struktur eines Sets von Distanzmaßen zwischen zwei Sets von Objekten (den Zeilen- und Spaltenobjekten) zu ermitteln. Dies wird durch das Zuweisen von Beobachtungen zu bestimmten Positionen in einem konzeptuellen Raum mit wenigen Dimensionen erzielt, wobei die Abstände zwischen den Punkten des Raumes mit den vorgegebenen Unähnlichkeiten bzw. Ähnlichkeiten so gut wie möglich übereinstimmen.

Als Ergebnis werden die Zeilen- und Spaltenobjekte in diesem schwachdimensionierten Raum mithilfe der Methode der kleinsten Quadrate dargestellt, was häufig zu einem besseren Verständnis der Daten beiträgt.

Beziehung zu anderen Prozeduren von Categories. Wenn die Daten aus Distanzen innerhalb eines einzelnen Sets von Objekten (eine quadratische und symmetrische Matrix) bestehen, verwenden Sie die multidimensionale Skalierung.

Beziehung zu Standardverfahren. Die Prozedur "Multidimensionale Entfaltung" in SPSS Categories (PREFSCAL) bietet zahlreiche Verbesserungen gegenüber der in der Statistics Base verfügbaren Entfaltungsprozedur (über ALSCAL). Mit PREFSCAL kann der gemeinsame Raum eingeschränkt werden. Außerdem wird in PREFSCAL versucht, ein penalisiertes Stressmaß zu minimieren, was dazu beiträgt, degenerierte Lösungen zu vermeiden (für die ältere Algorithmen anfällig sind).

Seitenverhältnis in Diagrammen mit optimaler Skalierung

Das Seitenverhältnis in Diagrammen mit optimaler Skalierung ist isotropisch. Bei einem zweidimensionalen Diagramm ist die Distanz, die in Dimension 1 für eine (1) Einheit steht, gleich der Distanz, die in Dimension 2 für eine (1) Einheit steht. Wenn Sie bei einem zweidimensionalen Diagramm den Bereich einer Dimension ändern, ändert das System die Größe der anderen Dimension, um die physischen Abstände gleich groß zu halten. Das isotropische Seitenverhältnis kann für die Prozeduren für die optimale Skalierung nicht außer Kraft gesetzt werden.

Kapitel 2. Kategoriale Regression (CATREG)

Durch die **Kategoriale Regression** werden kategoriale Daten quantifiziert, indem den Kategorien numerische Werte zugewiesen werden. Dadurch ergibt sich für die transformierten Variablen eine optimale lineare Regressionsgleichung. Die kategoriale Regression wird auch mit dem Akronym CATREG (*Categorical Regression*) bezeichnet.

Das Standardverfahren der linearen Regressionsanalyse beinhaltet die Minimierung der Summe von quadrierten Differenzen zwischen einer Antwortvariablen (abhängig) und einer gewichteten Kombination von Prädiktorvariablen (unabhängig). Variablen sind in der Regel quantitativ, wobei (nominale) kategoriale Daten in Binär- oder Kontrastvariablen umcodiert werden. Infolgedessen dienen kategoriale Variablen einer Aufteilung in verschiedene Gruppen von Fällen, sodass jeweils separate Parametersätze für jede Gruppe geschätzt werden. Die geschätzten Koeffizienten geben die Auswirkung einer Änderung in den Prädiktoren auf die Antwortvariable wieder. Die Antwort kann für jede beliebige Kombination von Einflusswerten vorhergesagt werden.

Eine andere Methode besteht darin, dass die Antwort auf die kategorialen Einflusswerte selbst einer Regression unterzogen wird. Folglich wird für jede Variable ein Koeffizient geschätzt. Bei kategorialen Variablen sind die Kategoriewerte jedoch willkürlich. Durch verschiedene Codierungsarten der Kategorien ergeben sich jeweils unterschiedliche Koeffizienten, wodurch ein analyseübergreifender Vergleich gleicher Variablen erschwert wird.

CATREG erweitert die Standardmethode durch eine gleichzeitige Skalierung nominaler, ordinaler und numerischer Variablen. Die Prozedur quantifiziert kategoriale Variablen, sodass in den Quantifizierungen die Merkmale der ursprünglichen Kategorien zum Ausdruck kommen. Dadurch werden quantifizierte kategoriale Variablen auf dieselbe Weise wie numerische Variablen behandelt. Durch die Verwendung nicht linearer Transformationen können Variablen auf einer Vielzahl von Ebenen analysiert und somit das jeweils geeignetste Modell gefunden werden.

Beispiel. Mithilfe der kategorialen Regression könnte beispielsweise beschrieben werden, wie die Zufriedenheit im Beruf von der Berufsgruppe, der geografischen Region und der Fahrtzeit abhängt. Unter Umständen könnte sich herausstellen, dass bei leitenden Angestellten und geringer Fahrtzeit zum Arbeitsplatz ein hohes Zufriedenheitsniveau zu verzeichnen ist. Mithilfe der resultierenden Regressionsgleichung könnte die Zufriedenheit im Beruf für eine beliebige Kombination der drei unabhängigen Variablen prognostiziert werden.

Statistiken und Diagramme. Häufigkeiten, Regressionskoeffizienten, ANOVA-Tabelle, Iterationsverlauf, Kategoriequantifizierungen, Korrelationen zwischen nicht transformierten Prädiktoren, Korrelationen zwischen transformierten Prädiktoren, Residuendiagramme und Transformationsdiagramme.

Erläuterungen der Daten für die kategoriale Regression

Daten. Mit CATREG werden Kategorieindikatorvariablen verarbeitet. Diese Kategorieindikatoren müssen positive Ganzzahlen sein. Im Dialogfeld "Diskretisierung" können Sie Variablen mit Dezimalbrüchen und Zeichenfolgevariablen in positive Ganzzahlen umwandeln.

Annahmen. Es ist nur eine Antwortvariable zulässig, die maximale Anzahl der Prädiktorvariablen beträgt jedoch 200. Die Daten müssen mindestens drei gültige Fälle enthalten. Die Anzahl gültiger Fälle muss die Anzahl der Prädiktorvariablen plus eins übersteigen.

Verwandte Prozeduren. CATREG entspricht der kategorialen kanonischen Korrelationsanalyse mit optimaler Skalierung (OVERALS) und zwei Sets, von denen ein Set jeweils nur eine Variable enthält. Eine Skalierung aller Variablen auf numerischem Niveau entspricht der normalen Mehrfachregressionsanalyse.

So aktivieren Sie eine kategoriale Regression:

1. Wählen Sie in den Menüs Folgendes aus:
Analysieren > Regression > Optimale Skalierung (CATREG)...
2. Wählen Sie die abhängige Variable und die unabhängige(n) Variable(n).
3. Klicken Sie auf **OK**.

Ändern Sie bei Bedarf das Skalierungsniveau für die einzelnen Variablen.

Definieren der Skala in der kategorialen Regression

Sie können das Messniveau der optimalen Skalierung für die abhängigen und unabhängigen Variablen festlegen. In der Standardeinstellung werden sie als monotone Splines zweiten Grades (ordinal) mit zwei inneren Knoten skaliert. Außerdem können Sie die Gewichtung der Analysevariablen bestimmen.

Messniveau der optimalen Skalierung. Sie können auch das Messniveau zur Quantifizierung jeder Variablen auswählen.

- **Spline ordinal.** Die Reihenfolge der Kategorien der beobachteten Variable wird in der quantifizierten Variable beibehalten. Die Kategoriepunkte werden auf einer geraden Linie (einem Vektor) durch den Koordinatenursprung platziert. Die resultierende Transformation ist ein glatter, stückweiser, monotoner Polynom mit dem gewählten Grad. Die Stücke werden durch die vom Benutzer festgelegte Anzahl und die durch die Prozedur bestimmte Lage der inneren Knoten definiert.
- **Spline nominal.** Die einzige Information aus der beobachteten Variable, die in der optimal skalierten Variable beibehalten wird, ist die Gruppierung der Objekte in Kategorien. Die Reihenfolge der Kategorien bei der beobachteten Variablen wird nicht beibehalten. Die Kategoriepunkte werden auf einer geraden Linie (einem Vektor) durch den Koordinatenursprung platziert. Die resultierende Transformation ist ein glatter, stückweiser, möglicherweise nicht monotoner Polynom mit dem gewählten Grad. Die Stücke werden durch die vom Benutzer festgelegte Anzahl und die durch die Prozedur bestimmte Lage der inneren Knoten definiert.
- **Ordinal.** Die Reihenfolge der Kategorien der beobachteten Variable wird in der quantifizierten Variable beibehalten. Die Kategoriepunkte werden auf einer geraden Linie (einem Vektor) durch den Koordinatenursprung platziert. Die resultierende Transformation ist besser angepasst, aber weniger glatt als bei "Spline ordinal".
- **Nominal.** Die einzige Information aus der beobachteten Variable, die in der optimal skalierten Variable beibehalten wird, ist die Gruppierung der Objekte in Kategorien. Die Reihenfolge der Kategorien bei der beobachteten Variablen wird nicht beibehalten. Die Kategoriepunkte werden auf einer geraden Linie (einem Vektor) durch den Koordinatenursprung platziert. Die resultierende Transformation ist besser angepasst, aber weniger glatt als bei "Spline nominal".
- **Numerisch.** Die Kategorien werden als geordnet mit gleichen Abständen behandelt (Intervallniveau). Die Reihenfolge der Kategorien und die gleichen Abstände zwischen den Kategoriennummern in der beobachteten Variablen bleiben in der optimal skalierten Variablen erhalten. Die Kategoriepunkte werden auf einer geraden Linie (einem Vektor) durch den Koordinatenursprung platziert. Wenn alle Variablen auf numerischem Niveau skaliert sind, entspricht die Analyse der normalen Hauptkomponentenanalyse.

Kategoriale Regression: Diskretisierung

Im Dialogfeld "Diskretisierung" können Sie eine Methode zum Umcodieren der Variablen festlegen. Sofern nicht anders angegeben, werden Variablen mit Dezimalwerten in sieben Kategorien gruppiert, die nahezu normalverteilt sind. (Wenn die Variable weniger als sieben unterschiedliche Werte aufweist, entspricht die Anzahl der Kategorien der Anzahl der unterschiedlichen Werte.) Werte von Zeichenfolgevariablen werden immer in positive ganze Zahlen umgewandelt, indem in aufsteigender alphanumerischer Reihenfolge Kategorieindikatoren zugewiesen werden. Die Diskretisierung von Zeichenfolgevariablen er-

folgt anhand dieser Ganzzahlen. Andere Variablen werden in der Standardeinstellung nicht modifiziert. Die diskretisierten Variablen werden dann in der Analyse verwendet.

Methode. Verfügbar sind die Optionen "Gruppieren", "Einstufen" oder "Multiplizieren".

- **Gruppieren.** Hierbei werden die Daten in eine bestimmte Anzahl von Kategorien oder nach Intervall umcodiert.
- **Einstufen.** Die Variable wird durch Bilden einer Rangfolge diskretisiert.
- **Multiplizieren.** Die aktuellen Werte der Variablen werden standardisiert, mit 10 multipliziert und gerundet. Anschließend wird eine Konstante zu den Werten addiert, sodass der niedrigste diskretisierte Wert 1 beträgt.

Gruppieren. Folgende Optionen sind bei der Diskretisierung von Variablen durch Gruppieren verfügbar:

- **Anzahl Kategorien.** Geben Sie eine Anzahl von Kategorien an. Legen Sie außerdem fest, ob die Werte der Variablen über die Kategorien ungefähr normal- oder gleichverteilt werden sollen.
- **Gleiche Intervalle.** Die Variablen werden in Kategorien umcodiert, die durch diese gleich großen Intervalle definiert werden. Sie müssen die Größe der Intervalle angeben.

Kategoriale Regression: Fehlende Werte

Im Dialogfeld "Fehlende Werte" können Sie die Strategie für die Verarbeitung von fehlenden Werten in den Analyse- und Zusatzvariablen festlegen.

Strategie. Legen Sie fest, ob Objekte mit fehlenden Werten ausgeschlossen werden sollen (listenweiser Ausschluss) oder ob fehlende Werte imputiert (Aktivbehandlung).

- **Objekte mit fehlenden Werten für diese Variable ausschließen.** Objekte mit fehlenden Werten für die ausgewählte Variable werden aus der Analyse ausgeschlossen. Diese Strategie ist für Zusatzvariablen nicht verfügbar.
- **Fehlende Werte imputieren.** Fehlende Werte werden bei Objekten mit fehlenden Werten für die ausgewählte Variable gesetzt. Sie können die Methode auswählen, nach der die Werte gesetzt werden: Wählen Sie **Modalwert**, wenn fehlende Werte durch die häufigste Kategorie ersetzt werden sollen. Falls mehrere Modalwerte vorhanden sind, wird der Modalwert mit dem kleinsten Kategorieindikator verwendet. Wenn Sie **Zusatzkategorie** auswählen, werden fehlende Werte durch die gleiche Quantifizierung einer zusätzlichen Kategorie ersetzt. Dies bedeutet aber auch, dass Objekte mit einem fehlenden Wert für diese Variable als zur gleichen Kategorie (der zusätzlichen) zugehörig betrachtet werden.

Kategoriale Regression: Optionen

Im Dialogfeld Optionen werden festgelegt: die Anfangskonfiguration, Iterations- und Konvergenzkriterien, die Auswahl von zusätzlichen Fällen und die Beschriftung der Diagramme.

Zusatzobjekte. Mit dieser Option können Sie die Objekte angeben, die als Zusatzobjekte behandelt werden sollen. Geben Sie die Nummer eines Zusatzobjekts ein (oder geben Sie einen Bereich von Fällen an) und klicken Sie auf **Hinzufügen**. Zusatzobjekte können nicht gewichtet werden (Gewichtungen werden ignoriert).

Ausgangskonfiguration. Wenn keine Variablen als nominal behandelt werden, wählen Sie die Konfiguration **Numerisch**. Wenn mindestens eine Variable als nominal behandelt wird, wählen Sie die Konfiguration **Zufällig**.

Wenn mindestens eine Variable ein ordinales oder Spline-ordinales Messniveau besitzt, kann der normale Modellanpassungsalgorithmus zu einer suboptimalen Lösung führen. Die Auswahl von **Mehrere systematische Starts** mit allen möglichen zu testenden Vorzeichenmustern findet stets die optimale Lösung, aber die notwendige Verarbeitungszeit steigt rapide an, da die Anzahl der ordinalen und Spline-ordinalen Variablen im Dataset zunimmt. Sie können die Anzahl der Testmuster reduzieren, indem Sie einen Pro-

zentsatz an Verlust der Varianzgrenze angeben. Je höher die Grenze, umso mehr Vorzeichenmuster werden ausgeschlossen. Mit dieser Option ist eine optimale Lösung nicht garantiert, aber die Wahrscheinlichkeit, eine suboptimale Lösung zu erhalten, wird reduziert. Auch wenn zudem die optimale Lösung nicht gefunden wird, ist die Wahrscheinlichkeit, dass die suboptimale Lösung sich sehr von der optimalen Lösung unterscheidet, geringer. Wenn mehrere systematische Starts angefordert werden, werden die Vorzeichen der Regressionskoeffizienten für jeden Start in eine externe IBM SPSS Statistics-Datendatei oder ein Dataset in der aktuellen Sitzung geschrieben. Weitere Informationen finden Sie im Thema „Kategoriale Regression: Speichern“ auf Seite 17.

Die Ergebnisse eines früheren Laufs mit mehreren systematischen Starts ermöglichen Ihnen, **feste Vorzeichen für die Regressionskoeffizienten** zu verwenden. Die Vorzeichen (angegeben durch 1 und -1) müssen sich in einer Zeile des angegebenen Datasets bzw. der angegebenen Datei befinden. Die ganzzahlige Startzahl ist die Fallnummer der Zeile in dieser Datei, die die zu verwendenden Vorzeichen enthält.

Kriterien. Sie können die Höchstzahl an Iterationen festlegen, die bei den Regressionsberechnungen durchlaufen werden sollen. Außerdem können Sie einen Wert für das Konvergenzkriterium auswählen. Die Regression wird abgebrochen, wenn die Differenz der Gesamtanpassung zwischen den letzten zwei Iterationen kleiner als der Konvergenzwert ist oder die Maximalzahl von Iterationen erreicht wurde.

Diagramme beschriften mit. Hiermit können Sie festlegen, ob in den Diagrammen Variablen- und Wertbeschriftungen oder Variablenamen und Werte verwendet werden sollen. Sie können auch eine Höchstlänge für die Beschriftungen festlegen.

Kategoriale Regression: Regularisierung

Methode. Regularisierungsmethoden können den Vorhersagefehler des Modells verbessern, da die Variabilität in den Schätzungen des Regressionskoeffizienten durch die Verringerung der Schätzungen auf nahezu 0 verbessert wird. Lasso und Elastic Net verringern einige Koeffizientenschätzungen auf exakt 0 und stellen so eine Form der Variablenauswahl zur Verfügung. Wenn eine Regularisierungsmethode angefordert wird, werden das geregelte Modell und die Koeffizienten für jeden Penalisierungskoeffizientenwert in eine externe IBM SPSS Statistics-Datendatei oder ein Dataset in der aktuellen Sitzung geschrieben. Weitere Informationen finden Sie im Thema „Kategoriale Regression: Speichern“ auf Seite 17.

- **Ridge-Regression.** Ridge-Regression verringert Koeffizienten durch die Einführung eines Penalisierungsterms der Summe der Koeffizienten zum Quadrat mal eines **Penalisierungskoeffizienten**. Dieser Koeffizient kann zwischen 0 (keine Penalisierung) und 1 liegen. Das Verfahren sucht nach dem "besten" Wert der Penalisierung, wenn Sie einen Bereich und ein Inkrement eingeben.
- **Lasso.** Der Penalisierungsterm von Lasso basiert auf der Summe der absoluten Koeffizienten. Die Angabe eines Penalisierungskoeffizienten ähnelt der Ridge-Regression, Lasso ist aber rechenintensiver.
- **Elastic Net.** Elastic Net kombiniert einfach die Lasso- und Ridge-Regressionsspenalisierungen und sucht über das Raster der angegebenen Werte, um die "besten" Lasso- und Ridge-Regressionsspenalisierungskoeffizienten zu finden. Für ein gegebenes Paar von Lasso- und Ridge-Regressionsspenalisierungen ist Elastic Net wesentlich rechenintensiver als Lasso.

Regularisierungsdiagramme anzeigen. Dies sind Diagramme, die Regressionskoeffizienten und Regularisierungsspenalisierung gegenüberstellen. Wenn nach einem Bereich von Werten für den "besten" Penalisierungskoeffizienten gesucht wird, gibt es Einblick auf die Änderung der Regressionskoeffizienten in diesem Bereich.

Elastic Net-Diagramme. Für die Elastic Net-Methode werden getrennte Regularisierungsdiagramme durch die Werte der Ridge-Regressionsspenalisierung erzeugt. **Alle möglichen Diagramme** verwendet jeden Wert im von Minimum und Maximum angegebenen Bereich der Ridge-Regressionsspenalisierungswerte. **Für einige Ridge-Penalierungen** können Sie ein Subset an Werten im Bereich zwischen Minimum und Maximum angeben. Geben Sie die Nummer eines Penalisierungswerts ein (oder geben Sie einen Bereich von Werten an) und klicken Sie auf **Hinzufügen**.

Kategoriale Regression: Ausgabe

Im Dialogfeld Ausgabe wird festgelegt, welche statistischen Größen ausgegeben werden sollen.

Tabellen. Es werden die folgenden Tabellen erstellt:

- **Multiple R.** Umfasst R^2 , korrigiertes R^2 und korrigiertes R^2 unter Berücksichtigung der optimalen Skalierung.
- **ANOVA.** Diese Option enthält Quadratsummen für die Regression und die Residuen, Mittel der Quadrate und F . Es werden zwei ANOVA-Tabellen angezeigt: Eine Tabelle mit der Anzahl der Prädiktorvariablen entsprechenden Freiheitsgraden für die Regression und eine Tabelle mit Freiheitsgraden für die Regression unter Berücksichtigung der optimalen Skalierung.
- **Koeffizienten.** Diese Option bietet drei Tabellen: eine Koeffiziententabelle mit Beta-Koeffizienten, dem Standardfehler der Beta-Koeffizienten, t -Werten und der Signifikanz; eine Tabelle der optimalen Skalierung der Koeffizienten mit dem Standardfehler der Beta-Koeffizienten unter Berücksichtigung der optimalen Skalierungsfreiheitsgrade; und eine Tabelle mit Korrelationen nullter Ordnung, Teilkorrelationen und partiellen Korrelationen, Pratt's Maß für die relative Bedeutung der transformierten Prädiktoren und der Toleranz vor und nach der Transformation.
- **Iterationsverlauf.** Für jede Iteration, einschließlich der Anfangswerte des Algorithmus, werden das multiple R und der Regressionsfehler angezeigt. Die Erhöhung des multiplen R wird beginnend mit der ersten Iteration aufgeführt.
- **Korrelationen der Originalvariablen.** Eine Matrix mit den Korrelationen zwischen den nicht transformierten Variablen wird angezeigt.
- **Korrelationen der transformierten Variablen.** Eine Matrix mit den Korrelationen zwischen den transformierten Variablen wird angezeigt.
- **Geregelte Modelle und Koeffizienten.** Zeigt Penalisierungswerte, R -Quadrat und Regressionskoeffizienten für jedes geregelte Modell an. Wenn eine Resampling-Methode oder zusätzliche Objekte (Testfälle) angegeben sind, wird auch der Vorhersagefehler bzw. der mittlere quadratische Testfehler angezeigt.

Resampling. Mit Resampling-Methoden erhalten Sie eine Schätzung des Vorhersagefehlers des Modells.

- **Kreuzvalidierung.** Bei der Kreuzvalidierung wird die Stichprobe in mehrere Teilstichproben oder Aufteilungen gegliedert. Anschließend werden Modelle kategorialer Regression erzeugt; dabei werden nacheinander die Daten der einzelnen Stichproben ausgeschlossen. Das erste Modell beruht auf allen Fällen mit Ausnahme der Fälle in der ersten Stichprobenaufteilung, das zweite Modell auf allen Fällen mit Ausnahme der Fälle in der zweiten Stichprobenaufteilung usw. Bei jedem Modell wird jeweils der Vorhersagefehler geschätzt. Hierzu wird das Modell auf die Teilstichprobe angewendet, die beim Erstellen des Modells ausgeschlossen war.
- **.632 Bootstrap.** Mit Bootstrap werden Beobachtungen zufällig aus den Daten mit Zurücklegen gezogen. Dieser Vorgang wird mehrmals wiederholt, um eine Reihe von Bootstrap-Stichproben zu bekommen. Ein Modell eignet sich für jede Bootstrap-Stichprobe. Der Vorhersagefehler für jedes Modell wird durch sein angepasstes Modell geschätzt und dazu auf die Fälle angewendet, die sich nicht in der Bootstrap-Stichprobe befinden.

Kategoriequantifizierungen. Tabellen mit den transformierten Werten der gewählten Variablen werden angezeigt.

Deskriptive Statistiken. Tabellen mit den Häufigkeiten, fehlenden Werten und Modalwerten der gewählten Variablen werden angezeigt.

Kategoriale Regression: Speichern

Im Dialogfeld "Speichern" können Sie festlegen, dass die vorhergesagten Werte, Residuen und transformierten Werte im aktiven Dataset gespeichert werden. Außerdem können die diskretisierten Daten, die transformierten Werte, die geregelten Modelle und die Koeffizienten und Vorzeichen der Regressionskoeffizienten in einer IBM SPSS Statistics-Datendatei oder einem Dataset gespeichert werden.

- Datasets sind in der aktuellen Sitzung verfügbar. In nachfolgenden Sitzungen stehen Datasets nur zur Verfügung, wenn Sie sie ausdrücklich als Datendateien speichern. Die Namen von Datasets müssen den Regeln zum Benennen von Variablen entsprechen.
- Die Namen für die Dateien und Datasets müssen sich bei allen Arten von gespeicherten Daten unterscheiden.

Regularisierte Modelle und Koeffizienten werden immer dann gespeichert, wenn eine Regularisierungsmethode im Dialogfeld Regularisierung ausgewählt ist. Standardmäßig erstellt das Verfahren ein neues Dataset mit einem eindeutigen Namen, aber Sie können natürlich einen Namen eigener Wahl angeben oder eine externe Datei schreiben.

Die Vorzeichen der Regressionskoeffizienten werden immer dann gespeichert, wenn mehrere systematische Starts als anfängliche Konfiguration im Dialogfeld Optionen verwendet werden. Standardmäßig erstellt das Verfahren ein neues Dataset mit einem eindeutigen Namen, aber Sie können natürlich einen Namen eigener Wahl angeben oder eine externe Datei schreiben.

Kategoriale Regression: Transformationsdiagramme

Im Dialogfeld "Diagramme" können Sie die Variablen angeben, aus denen Transformationsdiagramme und Residuendiagramme erzeugt werden.

Transformationsdiagramme. Für jede dieser Variablen werden die Kategoriequantifizierungen anhand der ursprünglichen Kategoriewerte dargestellt. Leere Kategorien erscheinen auf der horizontalen Achse, haben jedoch keine Auswirkungen auf die Berechnungen. Diese Kategorien werden durch Unterbrechungen in der Verbindungslinie zwischen den Quantifizierungen gekennzeichnet.

Residuendiagramme. Für jede dieser Variablen werden Residuen (berechnet für die abhängige Variable, die aus allen Prädiktorvariablen mit Ausnahme der betreffenden Prädiktorvariable prognostiziert wurde) anhand der Kategorienummern und der optimalen Kategoriequantifizierungen multipliziert mit dem Beta über Kategorieindikatoren grafisch dargestellt.

Zusätzliche Funktionen beim Befehl CATREG

Sie können die kategoriale Regression Ihren Anforderungen entsprechend anpassen, wenn Sie ihre Auswahl in ein Syntaxfenster einfügen und die resultierende Befehlssyntax für den Befehl CATREG bearbeiten. Die Befehlssyntax ermöglicht außerdem Folgendes:

- Mit dem Unterbefehl SAVE können Sie Stammnamen für die transformierten Variablen festlegen, wenn Sie diese im aktiven Dataset speichern.

Vollständige Informationen zur Syntax finden Sie in der Befehlssyntaxreferenz.

Kapitel 3. Kategoriale Hauptkomponentenanalyse (CATPCA)

Bei diesem Verfahren werden kategoriale Variablen quantifiziert und gleichzeitig die Dimensionalität der Daten reduziert. Die kategoriale Hauptkomponentenanalyse ist auch unter dem Akronym CATPCA bekannt (*Categorical Principal Components Analysis*).

Das Ziel bei der Hauptkomponentenanalyse besteht darin, ein ursprüngliches Set von Variablen in ein kleineres Set von unkorrelierten Komponenten umzuwandeln, das soviel wie möglich von den Informationen aus den ursprünglichen Variablen enthält. Diese Methode ist besonders dann nützlich, wenn eine hohe Anzahl von Variablen eine effektive Interpretation der Beziehungen zwischen den Objekten (Subjekten und Einheiten) verhindert. Nach einer Reduktion der Dimensionalität werden nur einige Komponenten anstelle einer großen Anzahl von Variablen interpretiert.

Beim Standardverfahren der Hauptkomponentenanalyse wird von linearen Beziehungen zwischen numerischen Variablen ausgegangen. Bei der Methode der optimalen Skalierung können die Variablen jedoch auf verschiedenen Niveaus skaliert werden. Kategoriale Variablen werden optimal in der angegebenen Dimensionalität quantifiziert. Aufgrund dessen können nicht lineare Beziehungen zwischen Variablen modelliert werden.

Beispiel. Eine kategoriale Hauptkomponentenanalyse ließe sich verwenden, um den Zusammenhang zwischen Berufsgruppe, beruflicher Stellung, Region, Anzahl von Dienstreisen (hoch, mittel und gering) und der Zufriedenheit im Beruf grafisch darzustellen. Das Ergebnis könnte sein, dass ein Großteil der Varianz mit zwei Dimensionen erklärt werden kann. In der ersten Dimension ließen sich Berufsgruppe von Region trennen, und in der zweiten Dimension könnten berufliche Stellung und Häufigkeit von Dienstreisen getrennt werden. Außerdem könnte sich ein Zusammenhang zwischen hoher Zufriedenheit im Beruf und einer mittleren Häufigkeit von Dienstreisen ergeben.

Statistiken und Diagramme. Häufigkeiten, fehlende Werte, Messniveau der optimalen Skalierung, Modalwert, berücksichtigte Varianz nach Koordinaten der Zentroiden, Vektorkoordinaten, Gesamtsumme pro Variable und Dimension, Komponentenladungen für vektorquantifizierte Variablen, Quantifizierungen und Koordinaten der Kategorien, Iterationsverlauf, Korrelationen der transformierten Variablen und Eigenwerte der Korrelationsmatrix, Korrelationen der ursprünglichen Variablen und Eigenwerte der Korrelationsmatrix, Objektscores, Diagramme der Kategorien und verknüpften Kategorien, Transformationsdiagramme, Residuendiagramme, Diagramme der projizierten Zentroide, Objektdiagramme, Biplots, Triplots und Diagramme der Komponentenladungen.

Erläuterungen der Daten für die kategoriale Hauptkomponentenanalyse

Daten. Werte von Zeichenfolgevariablen werden immer in aufsteigender alphanumerischer Reihenfolge in positive Ganzzahlen umgewandelt. Benutzerdefinierte und systemdefiniert fehlende Werte sowie Werte kleiner als 1 werden als fehlend betrachtet. Sie können die Daten umcodieren oder eine Konstante zu Variablen mit Werten kleiner als 1 addieren, um die Werte als nicht fehlend zu deklarieren.

Annahmen. Die Daten müssen mindestens drei gültige Fälle enthalten. Die Analyse basiert auf Daten, die aus positiven ganzen Zahlen bestehen. Mit der Option zur Diskretisierung werden Variablen mit Dezimalwerten automatisch kategorisiert, indem die Werte in Kategorien gruppiert werden, die nahezu normalverteilt sind, und die Werte von Zeichenfolgevariablen automatisch in positive Ganzzahlen umgewandelt. Sie können andere Diskretisierungsmethoden bestimmen.

Verwandte Prozeduren. Eine Skalierung aller Variablen auf numerischem Niveau entspricht der normalen Hauptkomponentenanalyse. Wenn Sie die transformierten Variablen in einer normalen linearen Hauptkomponentenanalyse verwenden, stehen Ihnen andere Diagrammfunktionen zur Verfügung. Wenn alle Variablen ein mehrfach nominales Skalierungsniveau besitzen, ist die kategoriale Hauptkomponentenana-

lyse identisch mit der Mehrfachkorrespondenzanalyse. Wenn Variablensets untersucht werden sollen, verwenden Sie die kategoriale (nicht lineare) kanonische Korrelationsanalyse.

So aktivieren Sie eine kategoriale Hauptkomponentenanalyse:

1. Wählen Sie in den Menüs Folgendes aus:
Analysieren > Dimensionsreduktion > Optimale Skalierung...
2. Wählen Sie **Einige Variablen sind nicht mehrfach nominal** aus.
3. Wählen Sie die Option **Ein Set** aus.
4. Klicken Sie auf **Definieren**.
5. Wählen Sie mindestens zwei Analysevariablen aus und geben Sie die Anzahl der Dimensionen in der Lösung an.
6. Klicken Sie auf **OK**.

Sie können außerdem Beschriftungsvariablen für die Diagramme und Zusatzvariablen bestimmen, die in die gefundene Lösung eingepasst werden.

Definieren von Skala und Gewichtung in CATPCA

Sie können das Messniveau der optimalen Skalierung für die Analysevariablen und die Zusatzvariablen festlegen. In der Standardeinstellung werden sie als monotone Splines zweiten Grades (ordinal) mit zwei inneren Knoten skaliert. Außerdem können Sie die Gewichtung der Analysevariablen bestimmen.

Variablengewichtung. Sie können für jede Variable eine Gewichtung festlegen. Bei dem angegebenen Wert muss es sich um eine positive ganze Zahl handeln. Der Standardwert ist 1.

Messniveau der optimalen Skalierung. Sie können auch das für die Quantifizierung der einzelnen Variablen verwendete Skalierungsniveau auswählen.

- **Spline ordinal.** Die Reihenfolge der Kategorien der beobachteten Variable wird in der quantifizierten Variable beibehalten. Die Kategoriepunkte werden auf einer geraden Linie (einem Vektor) durch den Koordinatenursprung platziert. Die resultierende Transformation ist ein glatter, stückweiser, monotoner Polynom mit dem gewählten Grad. Die Stücke werden durch die vom Benutzer festgelegte Anzahl und die durch die Prozedur bestimmte Lage der inneren Knoten definiert.
- **Spline nominal.** Die einzige Information aus der beobachteten Variable, die in der optimal skalierten Variable beibehalten wird, ist die Gruppierung der Objekte in Kategorien. Die Reihenfolge der Kategorien bei der beobachteten Variablen wird nicht beibehalten. Die Kategoriepunkte werden auf einer geraden Linie (einem Vektor) durch den Koordinatenursprung platziert. Die resultierende Transformation ist ein glatter, stückweiser, möglicherweise nicht monotoner Polynom mit dem gewählten Grad. Die Stücke werden durch die vom Benutzer festgelegte Anzahl und die durch die Prozedur bestimmte Lage der inneren Knoten definiert.
- **Mehrfach nominal.** Die einzige Information aus der beobachteten Variable, die in der optimal skalierten Variable beibehalten wird, ist die Gruppierung der Objekte in Kategorien. Die Reihenfolge der Kategorien bei der beobachteten Variablen wird nicht beibehalten. Die Kategoriepunkte werden in den Zentroiden der Objekte in den einzelnen Kategorien platziert. Die Bezeichnung *Mehrfach* bedeutet, dass für jede Dimension andere Sets von Quantifizierungen berechnet werden.
- **Ordinal.** Die Reihenfolge der Kategorien der beobachteten Variable wird in der quantifizierten Variable beibehalten. Die Kategoriepunkte werden auf einer geraden Linie (einem Vektor) durch den Koordinatenursprung platziert. Die resultierende Transformation ist besser angepasst, aber weniger glatt als bei "Spline ordinal".
- **Nominal.** Die einzige Information aus der beobachteten Variable, die in der optimal skalierten Variable beibehalten wird, ist die Gruppierung der Objekte in Kategorien. Die Reihenfolge der Kategorien bei der beobachteten Variablen wird nicht beibehalten. Die Kategoriepunkte werden auf einer geraden Li-

nie (einem Vektor) durch den Koordinatenursprung platziert. Die resultierende Transformation ist besser angepasst, aber weniger glatt als bei "Spline nominal".

- **Numerisch.** Die Kategorien werden als geordnet mit gleichen Abständen behandelt (Intervallniveau). Die Reihenfolge der Kategorien und die gleichen Abstände zwischen den Kategoriennummern in der beobachteten Variablen bleiben in der optimal skalierten Variablen erhalten. Die Kategoriepunkte werden auf einer geraden Linie (einem Vektor) durch den Koordinatenursprung platziert. Wenn alle Variablen auf numerischem Niveau skaliert sind, entspricht die Analyse der normalen Hauptkomponentenanalyse.

Kategoriale Hauptkomponentenanalyse: Diskretisierung

Im Dialogfeld "Diskretisierung" können Sie eine Methode zum Umcodieren der Variablen festlegen. Sofern nicht anders angegeben, werden Variablen mit Dezimalwerten in sieben Kategorien gruppiert, die nahezu normalverteilt sind. (Wenn die Variable weniger als sieben unterschiedliche Werte aufweist, entspricht die Anzahl der Kategorien der Anzahl der unterschiedlichen Werte.) Werte von Zeichenfolgevariablen werden immer in positive ganze Zahlen umgewandelt, indem in aufsteigender alphanumerischer Reihenfolge Kategorieindikatoren zugewiesen werden. Die Diskretisierung von Zeichenfolgevariablen erfolgt anhand dieser Ganzzahlen. Andere Variablen werden in der Standardeinstellung nicht modifiziert. Die diskretisierten Variablen werden dann in der Analyse verwendet.

Methode. Verfügbar sind die Optionen "Gruppieren", "Einstufen" oder "Multiplizieren".

- **Gruppieren.** Hierbei werden die Daten in eine bestimmte Anzahl von Kategorien oder nach Intervall umcodiert.
- **Einstufen.** Die Variable wird durch Bilden einer Rangfolge diskretisiert.
- **Multiplizieren.** Die aktuellen Werte der Variablen werden standardisiert, mit 10 multipliziert und gerundet. Anschließend wird eine Konstante zu den Werten addiert, sodass der niedrigste diskretisierte Wert 1 beträgt.

Gruppieren. Folgende Optionen sind bei der Diskretisierung von Variablen durch Gruppieren verfügbar:

- **Anzahl Kategorien.** Geben Sie eine Anzahl von Kategorien an. Legen Sie außerdem fest, ob die Werte der Variablen über die Kategorien ungefähr normal- oder gleichverteilt werden sollen.
- **Gleiche Intervalle.** Die Variablen werden in Kategorien umcodiert, die durch diese gleich großen Intervalle definiert werden. Sie müssen die Größe der Intervalle angeben.

Kategoriale Hauptkomponentenanalyse: Fehlende Werte

Im Dialogfeld **Fehlende Werte** können Sie die Strategie für die Verarbeitung von fehlenden Werten in den Analyse- und Zusatzvariablen festlegen.

Strategie. Fehlende Werte können ausgeschlossen (Passivbehandlung) oder imputiert (Aktivbehandlung) werden oder Objekte mit fehlenden Werten können ausgeschlossen werden (listenweiser Ausschluss).

- **Fehlende Werte ausschließen; für Korrelationen nach Quantifizierung imputieren.** Objekte mit fehlenden Werten bei der ausgewählten Variablen tragen nicht zur Analyse dieser Variablen bei. Wenn alle Variablen passiv behandelt werden, werden Objekte mit fehlenden Werten für alle Variablen als Zusatzobjekte behandelt. Wenn im Dialogfeld "Ausgabe" Korrelationen ausgewählt sind, werden fehlende Werte nach der Analyse anhand der häufigsten Variablenkategorie (Modalwert) für die Korrelationen der ursprünglichen Variablen gesetzt. Sie können die Methode auswählen, nach der Werte für die Korrelationen der optimal skalierten Variablen gesetzt werden.
 - **Modalwert.** Ersetzen Sie fehlende Werte durch den Modalwert der optimal skalierten Variablen.
 - **Zusatzkategorie.** Ersetzen Sie fehlende Werte durch die Quantifizierung einer zusätzlichen Kategorie. Diese Einstellung impliziert, dass Objekte mit einem fehlenden Wert für diese Variable als zu derselben Kategorie (der zusätzlichen) zugehörig betrachtet werden.

- **Zufallskategorie.** Imputieren Sie jeden fehlenden Wert für eine Variable mit dem quantifizierten Wert einer anderen Zufallskategorienummer auf der Basis der Randhäufigkeiten der Kategorien der Variablen.
- **Fehlende Werte imputieren.** Fehlende Werte werden bei Objekten mit fehlenden Werten für die ausgewählte Variable gesetzt. Sie können die Methode auswählen, nach der die Werte gesetzt werden:
 - **Modalwert.** Ersetzen Sie fehlende Werte durch die häufigste Kategorie. Falls mehrere Modalwerte vorhanden sind, wird der Modalwert mit dem kleinsten Kategorieindikator verwendet.
 - **Zusatzkategorie.** Ersetzen Sie fehlende Werte durch die gleiche Quantifizierung einer zusätzlichen Kategorie. Diese Einstellung impliziert, dass Objekte mit einem fehlenden Wert für diese Variable als zu derselben Kategorie (der zusätzlichen) zugehörig betrachtet werden.
 - **Zufallskategorie.** Ersetzen Sie jeden fehlenden Wert für eine Variable durch eine andere Zufallskategorienummer auf der Basis der Randhäufigkeiten der Kategorien.
- **Objekte mit fehlenden Werten für diese Variable ausschließen.** Objekte mit fehlenden Werten für die ausgewählte Variable werden aus der Analyse ausgeschlossen. Diese Strategie ist für Zusatzvariablen nicht verfügbar.

Kategoriale Hauptkomponentenanalyse: Optionen

Mit den Steuerelementen im Dialogfeld **Optionen** können Sie die anfängliche Konfiguration auswählen, Kriterien für die Iterationen und die Konvergenz festlegen, eine Normalisierungsmethode auswählen, die Methode zum Beschriften der Diagramme auswählen und Zusatzobjekte angeben.

Zusatzobjekte. Geben Sie die Fallnummer des Objekts bzw. die erste und letzte Fallnummer einer Reihe von Objekten an, die als Zusatzobjekte deklariert werden sollen, und klicken Sie dann auf **Hinzufügen**. Wenn ein Objekt als Zusatzobjekt deklariert ist, werden Fallgewichtungen für das Objekt ignoriert.

Normalisierungsmethode. Sie können eine von fünf Optionen für die Normalisierung der Objektscores und Variablen auswählen. In einer Analyse kann jeweils nur eine Normalisierungsmethode verwendet werden.

- **Variablenprinzipal.** Bei dieser Option wird die Assoziation zwischen Variablen optimiert. Die Komponentenladungen (Korrelationen mit Hauptkomponenten wie Dimensionen und Objektscores) stellen die Koordinaten der Variablen im Objektraum dar. Diese Methode ist nützlich, wenn Sie hauptsächlich an der Korrelation zwischen den Variablen interessiert sind.
- **Objektprinzipal.** Bei dieser Option werden die Distanzen zwischen den Objekten optimiert. Diese Methode ist nützlich, wenn Sie hauptsächlich an den Unterschieden oder Übereinstimmungen zwischen den Objekten interessiert sind.
- **Symmetrisch.** Verwenden Sie diese Normalisierungsmethode, wenn Sie hauptsächlich an den Beziehungen zwischen den Objekten und den Variablen interessiert sind.
- **Unabhängig.** Verwenden Sie diese Normalisierungsmethode, wenn Sie die Distanzen zwischen den Objekten und die Korrelationen zwischen den Variablen getrennt untersuchen möchten.
- **Benutzerdefiniert.** Sie können einen beliebigen reellen Wert aus dem abgeschlossenen Intervall $[-1, 1]$ eingeben. Der Wert 1 entspricht der Methode "Objektprinzipal". Der Wert 0 entspricht der Methode "Symmetrisch". Der Wert -1 entspricht der Methode "Variablenprinzipal". Bei Werten größer als -1 und kleiner als 1 wird der Eigenwert sowohl über Objekte als auch Variablen gestreut. Mit dieser Methode können Sie Biplots und Triplots genau an Ihre Bedürfnisse anpassen.

Kriterien. Sie können die maximale Anzahl von Iterationen festlegen, die in der Prozedur bei den Berechnungen durchlaufen werden. Außerdem können Sie einen Wert für das Konvergenzkriterium auswählen. Die Prozedur wird abgebrochen, wenn die Differenz der Gesamtanpassung zwischen den letzten zwei Iterationen kleiner als der Konvergenzwert ist oder die Maximalzahl von Iterationen erreicht wurde.

Diagramme beschriften mit. Sie können festlegen, ob in den Diagrammen Variablen- und Wertbeschriftungen oder Variablennamen und Werte verwendet werden sollen. Sie können auch eine Höchstlänge für die Beschriftungen festlegen.

Diagrammdimensionen. Sie können festlegen, wie die Dimensionen in der Ausgabe angezeigt werden.

- **Alle Dimensionen in der Lösung anzeigen.** Alle Dimensionen in der Lösung werden in einer Streudiagrammmatrix angezeigt.
- **Anzahl der Dimensionen beschränken.** Die angezeigten Dimensionen werden auf geplottete Paare beschränkt. Wenn Sie die Dimensionen beschränken, müssen Sie die niedrigste und die höchste Dimension auswählen, die jeweils geplottet werden soll. Die niedrigste Dimension muss im Bereich von 1 bis zur Anzahl der Dimensionen in der Lösung minus 1 liegen und wird gegen höhere Dimensionen geplottet. Der Wert für die höchste Dimension muss im Bereich von 2 bis zur Anzahl der Dimensionen in der Lösung liegen und gibt die höchste Dimension an, die beim Plotten der Dimensionspaare verwendet wird. Diese Spezifikation gilt für alle angeforderten multidimensionalen Plots.

Rotation. Sie können eine Rotationsmethode auswählen, um rotierte Ergebnisse zu erhalten.

Hinweis: Diese Rotationsmethoden sind nicht verfügbar, wenn Sie **Bootstrapping durchführen** im Dialogfeld **Bootstrap** auswählen.

- **Varimax.** Eine orthogonale Rotationsmethode, die die Anzahl der Variablen mit hohen Ladungen für jede Komponente minimiert. Sie vereinfacht die Interpretation der Komponenten.
- **Quartimax.** Eine Rotationsmethode, die die Anzahl der Komponenten minimiert, die zum Erklären aller Variablen benötigt werden. Sie vereinfacht die Interpretation der beobachteten Variablen.
- **Equamax.** Eine Rotationsmethode, die eine Kombination zwischen der Methode "Varimax" (vereinfacht die Komponenten) und der Methode "Quartimax" (vereinfacht die Variablen) darstellt. Die Anzahl der Variablen mit hohen Ladungen auf eine Komponente sowie die Anzahl der Komponenten, die benötigt werden, um eine Variable zu erklären, werden minimiert.
- **Oblimin.** Eine Methode für schiefe (nicht orthogonale) Rotation. Wenn Delta den Wert 0 annimmt, sind die Ergebnisse am schiefsten. Mit zunehmendem negativem Wert von Delta werden die Komponenten weniger schief. Positive Werte ermöglichen zusätzliche Komponentenkorrelation. Der Wert von **Delta** muss kleiner-gleich 0,8 sein.
- **Promax.** Eine schiefe (nicht orthogonale) Rotation, die eine Korrelation der Komponenten ermöglicht. Sie kann schneller berechnet werden als eine direkte Oblimin-Rotation und ist daher für große Datensets nützlich. Der zulässige Grad der Korrelation (Schiefe) wird vom Kappa-Parameter gesteuert. Der Wert von **Kappa** muss größer-gleich 1 und kleiner als 10.000 sein.

Konfiguration. Sie können die Koordinaten einer Konfiguration aus einer Datei einlesen. Die erste Variable in der Datei enthält die Koordinaten der ersten Dimension. Die zweite Variable enthält die Koordinaten der zweiten Dimension usw.

- **Initial.** Die in der Datei angegebene Konfiguration wird als Ausgangspunkt für die Analyse verwendet.
- **Fest.** Die in der Datei angegebene Konfiguration wird zum Anpassen der Variablen verwendet. Die Variablen, die angepasst werden, müssen als Analysevariablen ausgewählt werden. Da die Konfiguration aber fest ist, werden Sie als Zusatzvariablen verarbeitet (und müssen deshalb nicht als Zusatzvariablen ausgewählt werden).

Kategoriale Hauptkomponentenanalyse: Ausgabe

Das Dialogfeld **Ausgabe** steuert die Anzeige der Ergebnisse.

Objektscores. Hiermit werden die Objektscores angezeigt. Die folgenden Optionen sind verfügbar:

- **Diese Kategorien berücksichtigen.** Hiermit werden die Kategorieindikatoren der ausgewählten Analysevariablen angezeigt.

- **Objektscores beschriften nach.** Sie können eine der in der Liste der Beschriftungsvariablen enthaltene Variable für die Beschriftung der Objekte auswählen.

Komponentenladungen. Hiermit werden Komponentenladungen für alle Variablen angezeigt, bei denen keine mehrfach nominalen Skalierungsniveaus vorliegen. Sie können die Komponentenladungen nach Größe sortieren.

Iterationsverlauf. Für jede Iteration werden berücksichtigte Varianz, Verlust sowie Steigerung der berücksichtigten Varianz ausgegeben.

Korrelationen der Originalvariablen. Hiermit werden die Korrelationsmatrix der ursprünglichen Variablen und die Eigenwerte dieser Matrix ausgegeben.

Korrelationen der transformierten Variablen. Hiermit werden die Korrelationsmatrix der transformierten (optimal skalierten) Variablen und die Eigenwerte dieser Matrix ausgegeben.

Varianz berücksichtigt für. Hiermit wird die berücksichtigte Varianz nach Zentroidkoordinaten und nach Vektorkoordinaten sowie die Gesamtsumme (kombinierte Zentroid- und Vektorkoordinaten) pro Variable und Dimension ausgegeben.

Kategoriequantifizierungen. Dies sind die Kategoriequantifizierungen und -koordinaten pro Dimension der ausgewählten Variablen.

Deskriptive Statistiken. Hiermit werden Häufigkeiten, Anzahl der fehlenden Werte und Modalwert für die ausgewählten Variablen ausgegeben.

Kategoriale Hauptkomponentenanalyse: Speichern

Mit den Steuerelementen im Dialogfeld **Speichern** werden diskretisierte Daten, Objektscores, transformierte Werte und andere Ergebnisse des aktiven Datasets, ein neues Dataset in der aktuellen Sitzung oder eine externe Datei gespeichert.

- Datasets sind in der aktuellen Sitzung verfügbar. In nachfolgenden Sitzungen stehen Datasets nur zur Verfügung, wenn Sie sie ausdrücklich als Datendateien speichern. Die Namen von Datasets müssen den Regeln zum Benennen von Variablen entsprechen.
- Die Namen für die Dateien und Datasets müssen sich bei allen Arten von gespeicherten Daten unterscheiden.
- Wenn Sie die Objektscores oder die transformierten Werte im aktiven Dataset speichern, können Sie die Anzahl der mehrfach nominalen Dimensionen angeben.
- Die Optionen in der Gruppe **Bootstrap-Konfidenzellipsen** sind nur verfügbar, wenn Sie **Bootstrapping durchführen** im Dialogfeld **Bootstrap** auswählen.

Kategoriale Hauptkomponentenanalyse: Objekt- und Variablendiagramme

Im Dialogfeld "Objekt- und Variablendiagramme" können Sie festlegen, welche Diagramme und für welche Variablen die Diagramme erstellt werden sollen.

Objektpunkte. Es wird ein Diagramm der Objektpunkte angezeigt.

Objekte und Variablen (Biplot). Es werden Objektpunkte und Variablenkoordinaten ausgegeben. Hierbei bestehen die folgenden Möglichkeiten: Komponentenladungen oder Zentroide der Variablen.

Objekte, Ladungen und Zentroide (Triplot). Es werden die Objektpunkte, die Zentroide von Variablen mit mehrfach nominalem Skalierungsniveau und die Komponentenladungen von anderen Variablen ausgegeben.

Biplot- und Triplotvariablen. Die Biplots und Triplots können auf allen Variablen oder einem Subset der Variablen basieren.

Objekte beschriften. Objekte können mit den Kategorien ausgewählter Variablen (im Dialogfeld "Optionen" können Sie Kategorieindikatoren oder Wertbeschriftungen wählen) oder den Fallnummern beschriftet werden. Wenn **Variable** ausgewählt ist, wird pro Variable ein Diagramm erstellt.

Kategoriale Hauptkomponentenanalyse: Kategoriediagramme

Im Dialogfeld "Kategoriediagramme" können Sie festlegen, welche Diagramme und für welche Variablen die Diagramme erstellt werden sollen.

Kategoriediagramme. Für jede ausgewählte Variable wird ein Diagramm mit Zentroid und Vektorkoordinaten erstellt. Bei Variablen mit mehrfach nominalem Skalierungsniveau befinden sich die Kategorien in den Zentroiden der Objekte in der jeweiligen Kategorie. Bei allen anderen Skalierungsniveaus befinden sich die Kategorien auf einem Vektor, der durch den Koordinatenursprung verläuft.

Verbundene Kategoriediagramme. Hierbei handelt es sich um ein einzelnes Diagramm mit Zentroid und Vektorkoordinaten jeder ausgewählten Variablen.

Transformationsdiagramme. Hierbei wird ein Diagramm der optimalen Kategoriequantifizierungen gegen die Kategorieindikatoren erstellt. Bei Variablen mit mehrfach nominalem Skalierungsniveau können Sie die Anzahl von Dimensionen angeben. Pro Dimension wird ein Diagramm erstellt. Sie können außerdem festlegen, dass für jede ausgewählte Variable ein Diagramm der Residuen ausgegeben wird.

Zentroide projizieren von. Sie können eine Variable auswählen und ihre Zentroide auf ausgewählte Variablen projizieren. Es kann jedoch nicht auf Variablen mit mehrfach nominalem Skalierungsniveau projiziert werden. Wenn dieses Diagramm angefordert wird, wird außerdem eine Tabelle mit den Koordinaten der projizierten Zentroiden ausgegeben.

Kategoriale Hauptkomponentenanalyse: Ladungsdiagramme

Das Dialogfeld **Ladungsdiagramm** steuert die im Diagramm enthaltenen Variablen, die Anzeige von Zentroiden im Ladungsdiagramm und die Anzeige der Diagramme der berücksichtigten Varianz.

Varianz berücksichtigt für. Für jede Dimension wird ein Diagramm der berücksichtigten Varianz angezeigt.

Komponentenladungen anzeigen. Bei Auswahl dieser Option wird ein Diagramm der Komponentenladungen angezeigt.

Ladungsvariablen. Das Komponentenladungsdiagramm kann auf allen Variablen oder einem Subset der Variablen basieren.

Zentroide einschließen. Bei Variablen mit mehrfach nominalem Skalierungsniveau sind keine Komponentenladungen vorhanden. Bei diesen Variablen können jedoch die Zentroide in das Diagramm aufgenommen werden. Sie können alle mehrfach nominalen Variablen verwenden oder ein Subset auswählen.

Kategoriale Hauptkomponentenanalyse: Bootstrap

Im Dialogfeld **Bootstrap** werden die Parameter für die Bootstrap-Analyse angegeben.

Bootstrapping durchführen. Führt eine neue Ziehung von Bootstrap-Stichproben durch.

- Wenn Diagramme von Ladungen, Kategorien oder Komponentenscores angefordert werden, werden zusätzliche Diagramme angezeigt. Diese Diagramme zeigen die Punkte für die Datenstichprobe und die Bootstrap-Schätzungen an. Transformationsdiagramme enthalten Konfidenzbereiche. Es wird zudem ein Diagramm für die Eigenwerte angezeigt.
- Wenn eine zweidimensionale Lösung angegeben wird, werden Konfidenzellipsendiagramme für die Eigenwerte, die Komponentenladungen, die Kategoriepunkte und die Objektpunkte angezeigt.
- Die Ziehung von Bootstrap-Stichproben ist nicht verfügbar, wenn Sie im Dialogfeld **Optionen** eine Rotationsmethode angeben.

Sie können für das Bootstrapping **Ausgewogen** oder **Unausgewogen** auswählen.

Anzahl der Stichproben. Die Anzahl der Bootstrap-Stichproben, die zur Berechnung der Bootstrap-Schätzungen verwendet werden. Der Wert muss eine positive ganze Zahl sein.

Konfidenzniveau. Das Konfidenzniveau der Bootstrap-Schätzungen, ausgedrückt in Prozent. Der Wert muss eine positive Zahl unter 100 sein.

Abgleichmethode. Die beiden Alternativen sind **Procrustes-Wert** und **Reflexion**.

Konfidenzellipsen. Steuert den Schwellenwertbereich für Konfidenzellipsen in Diagrammen. Bei Bereichen, die größer als (Operator >) der angegebene Wert sind, nimmt die Anzahl der Ellipsen mit zunehmendem Schwellenwert ab.

- Die Einstellungen in dieser Gruppe sind nur verfügbar, wenn die Anzahl der im Hauptdialogfenster angegebenen Dimensionen 2 ist.
- Konfidenzellipsen für Ladungsdiagramme sind nur verfügbar, wenn **Komponentenladungen anzeigen** im Dialogfeld **Ladungsdiagramm** ausgewählt ist.
- Konfidenzellipsen für Objektdiagramme sind nur verfügbar, wenn **Objektpunkte** im Dialogfeld **Objekt- und Variablen-diagramme** ausgewählt ist.
- Konfidenzellipsen für Kategoriediagramme sind nur verfügbar, wenn in der Liste **Kategoriediagramme** im Dialogfeld **Kategoriediagramme** mehrere Variablen angegeben sind.

Anzahl Konturpunkte für Ellipse. Die Anzahl der Diagrammpunkte, die zum Zeichnen jeder Konfidenzellipse verwendet werden. Größere Werte erzeugen glattere Ellipsen. Der Wert muss eine positive ganze Zahl kleiner-gleich 100 sein.

Zusätzliche Funktionen beim Befehl CATPCA

Sie können die kategoriale Hauptkomponentenanalyse an Ihre Bedürfnisse anpassen, indem Sie Ihre Auswahl in ein Syntaxfenster einfügen und die daraus resultierende Befehlssyntax für CATPCA bearbeiten. Die Befehlssyntax ermöglicht außerdem Folgendes:

- Mit dem Unterbefehl **SAVE** können Sie Stammnamen für die transformierten Variablen, Objektscores und Approximationen festlegen, wenn Sie diese im aktiven Dataset speichern.
- Mit dem Unterbefehl **PL0T** kann die Höchstlänge für Beschriftungen für jedes Diagramm getrennt festgelegt werden.
- Mit dem Unterbefehl **PL0T** kann eine separate Variablenliste für Residuendiagramme festgelegt werden.

Vollständige Informationen zur Syntax finden Sie in der Befehlssyntaxreferenz.

Kapitel 4. Nicht lineare kanonische Korrelationsanalyse (OVERALS)

Die nicht lineare kanonische Korrelationsanalyse entspricht der kategorialen kanonischen Korrelationsanalyse mit optimaler Skalierung. Ziel dieser Prozedur ist es, den Grad der gegenseitigen Ähnlichkeit von Sets kategorialer Variablen zu bestimmen. Die nicht lineare kanonische Korrelationsanalyse ist auch unter der Akronym OVERALS bekannt.

Die kanonische Standardkorrelationsanalyse ist eine Erweiterung der mehrfachen Regression, wobei das zweite Set nicht nur eine Antwortvariable, sondern mehrere Antwortvariablen enthält. Ziel ist es, einen möglichst großen Teil der Varianz in der Beziehung zwischen zwei Sets von numerischen Variablen in einem Raum mit wenigen Dimensionen zu erklären. Zunächst werden die Variablen in jedem Set linear so kombiniert, dass die linearen Kombinationen maximal miteinander korrelieren. Nach diesen Kombinationen werden weitere lineare Kombinationen ermittelt, die nicht mit den vorherigen Kombinationen korrelieren und über die höchstmögliche Korrelation verfügen.

Der Einsatz der optimalen Skalierung erweitert die Standardanalyse in drei entscheidenden Punkten. Erstens ermöglicht OVERALS mehr als zwei Variablensets. Zweitens können Variablen nominal, ordinal oder numerisch skaliert werden. Als Folge davon können nicht lineare Beziehungen zwischen Variablen analysiert werden. Drittens werden nicht die Korrelationen zwischen den Variablensets maximiert, sondern die Sets werden mit einem unbekanntem, durch die Objektscores definierten Kompromissset verglichen.

Beispiel. Mit der kategorialen kanonischen Korrelationsanalyse mit optimaler Skalierung kann die Beziehung zwischen zwei Variablensets grafisch dargestellt werden, von denen das eine die Berufsgruppe und die Ausbildungsdauer in Jahren und das andere die Region des Wohnsitzes und das Geschlecht enthält. Die Analyse könnte beispielsweise ergeben, dass die Ausbildungsdauer und der Wohnsitz einen höheren Erklärungswert haben als die übrigen Variablen. Außerdem könnten Sie herausfinden, dass der Erklärungswert der Ausbildungsdauer in Jahren in der ersten Dimension am höchsten ist.

Statistiken und Diagramme. Häufigkeiten, Zentroide, Iterationsverlauf, Objektscores, Kategoriequantifizierungen, Gewichtungen, Komponentenladungen, einfache und mehrfache Anpassung, Diagramme der Objektscores, Diagramme der Kategoriekoordinaten, Diagramme der Komponentenladungen, Diagramme der Kategoriezentroide, Transformationsdiagramme.

Erläuterungen der Daten für die nicht lineare kanonische Korrelationsanalyse

Daten. Verwenden Sie zum Codieren von kategorialen Variablen ganze Zahlen (nominales oder ordinales Skalierungsniveau). Um die Ausgabe zu minimieren, verwenden Sie zum Codieren jeder Variablen aufeinander folgende ganze Zahlen, mit 1 beginnend. Variablen, die auf numerischem Niveau skaliert sind, dürfen nicht in aufeinander folgende ganze Zahlen umcodiert werden. Um die Ausgabe zu minimieren, subtrahieren Sie für jede auf numerischem Niveau skalierte Variable den kleinsten beobachteten Wert von jedem Wert, und addieren Sie 1. Gebrochene Werte werden hinter der Dezimalstelle abgeschnitten.

Annahmen. Variablen können in ein oder mehr Sets klassifiziert werden. Variablen in der Analyse werden mehrfach nominal, einzeln nominal, ordinal oder numerisch skaliert. Die in der Prozedur verwendete maximale Anzahl von Dimensionen hängt vom Messniveau der optimalen Skalierung der Variablen ab. Wenn alle Variablen als ordinal, einzeln nominal oder numerisch angegeben sind, entspricht die maximale Anzahl von Dimensionen dem kleineren der folgenden beiden Werte: die Anzahl der Beobachtungen minus 1 oder die Gesamtanzahl der Variablen. Wenn jedoch nur zwei Variablensets definiert sind, entspricht die maximale Anzahl von Dimensionen der Anzahl von Variablen im kleineren Set. Wenn einige Variablen mehrfach nominal sind, entspricht die maximale Anzahl von Dimensionen der Gesamtanzahl von mehrfach nominalen Kategorien plus der Anzahl von nicht mehrfach nominalen Variablen minus der

Anzahl von mehrfach nominalen Variablen. Wenn die Analyse beispielsweise fünf Variablen umfasst, von denen eine mehrfach nominal mit vier Kategorien ist, ist die maximale Anzahl von Dimensionen gleich $(4 + 4 - 1)$, also 7. Wenn Sie eine Zahl angeben, die größer als der Höchstwert ist, wird der Höchstwert verwendet.

Verwandte Prozeduren. Wenn jedes Set eine Variable enthält, ist die nicht lineare kanonische Korrelationsanalyse äquivalent zur Hauptkomponentenanalyse mit optimaler Skalierung. Wenn jede dieser Variablen mehrfach nominal ist, entspricht die Analyse der Mehrfachkorrespondenzanalyse. Wenn zwei Variablensets vorhanden sind und eines dieser Sets nur eine Variable enthält, ist die Analyse identisch mit der kategorialen Regression mit optimaler Skalierung.

So aktivieren Sie eine nicht lineare kanonische Korrelationsanalyse:

1. Wählen Sie in den Menüs Folgendes aus:
Analysieren > Dimensionsreduktion > Optimale Skalierung...
2. Wählen Sie entweder **Alle Variablen sind mehrfach nominal** oder **Einige Variablen sind nicht mehrfach nominal** aus.
3. Wählen Sie **Mehrere Sets**.
4. Klicken Sie auf **Definieren**.
5. Definieren Sie mindestens zwei Variablensets. Wählen Sie die Variablen aus, die in das erste Set aufgenommen werden sollen. Um zum nächsten Set zu wechseln, klicken Sie auf **Weiter** und wählen Sie die Variablen aus, die in das zweite Set aufgenommen werden sollen. Sie können weitere Sets hinzufügen. Klicken Sie auf **Zurück**, um zum zuvor definierten Variablenset zurückzukehren.
6. Legen Sie für jede Variable den Wertebereich und das Messniveau (Niveau der optimalen Skalierung) fest.
7. Klicken Sie auf **OK**.
8. Die folgenden Optionen sind verfügbar:
 - Wählen Sie eine oder mehrere Variablen aus, um Punktbeschriftungen für Objektscorediagramme anzugeben. Jede Variable erzeugt ein separates Diagramm, dessen Punkte mit den Werten dieser Variablen beschriftet sind. Für jede dieser Variablen zur Diagrammbeschriftung muss ein Bereich festgelegt werden. In diesem Dialogfeld kann eine Variable nicht gleichzeitig in der Analyse und als Beschriftungsvariable verwendet werden. Wenn Sie das Objektscorediagramm mit einer in der Analyse verwendeten Variablen beschriften möchten, verwenden Sie die Funktion "Berechnen" (verfügbar im Menü "Transformieren"), um eine Kopie dieser Variablen zu erstellen. Verwenden Sie die neue Variable, um das Diagramm zu beschriften. Wahlweise können Sie dazu auch die Befehlsyntax verwenden.
 - Geben Sie die in der Lösung gewünschte Anzahl der Dimensionen an. Im Allgemeinen sollten Sie nur so viele Dimensionen auswählen, wie zum Erklären des größten Teils der Variation notwendig ist. Wenn die Analyse mehr als zwei Dimensionen umfasst, werden dreidimensionale Diagramme der ersten drei Dimensionen erzeugt. Weitere Dimensionen können durch die Bearbeitung des Diagramms angezeigt werden.

Definieren von Bereich und Skala

Sie müssen für jede Variable einen Bereich definieren. Der angegebene Höchstwert muss eine Ganzzahl sein. Gebrochene Datenwerte werden in der Analyse abgeschnitten. Ein Kategoriewert, der außerhalb des angegebenen Bereiches liegt, wird in der Analyse ignoriert. Um eine möglichst kleine Ausgabe zu erzielen, erstellen Sie mithilfe der Funktion "Automatisch umcodieren" (verfügbar im Menü "Transformieren") aufeinander folgende, mit 1 beginnende Kategorien für Variablen, die als nominale oder ordinale Variablen behandelt werden. Das Umcodieren in fortlaufende Ganzzahlen empfiehlt sich nicht für Variablen, die auf numerischem Niveau skaliert sind. Um bei Variablen, die als numerische Variablen behandelt werden, eine möglichst kleine Ausgabe zu erzielen, subtrahieren Sie für alle Variablen von jedem Wert den Minimalwert, und addieren Sie 1.

Sie müssen außerdem die für die Quantifizierung der einzelnen Variablen verwendete Skalierung auswählen.

- **Ordinal.** Die Reihenfolge der Kategorien der beobachteten Variable wird in der quantifizierten Variable beibehalten.
- **Einzel nominal.** Bei der quantifizierten Variablen erhalten Objekte der gleichen Kategorie den gleichen Score.
- **Mehrfach nominal.** Die Quantifizierungen können bei den einzelnen Dimensionen unterschiedlich sein.
- **Diskret numerisch.** Die Kategorien werden als geordnet mit gleichen Abständen behandelt. Die Unterschiede zwischen den Kategoriennummern und die Reihenfolge der Kategorien der beobachteten Variablen bleiben bei der quantifizierten Variablen erhalten.

Definieren eines Bereichs

Sie müssen für jede Variable einen Bereich definieren. Der angegebene Höchstwert muss eine Ganzzahl sein. Gebrochene Datenwerte werden in der Analyse abgeschnitten. Ein Kategoriewert, der außerhalb des angegebenen Bereiches liegt, wird in der Analyse ignoriert. Um eine möglichst kleine Ausgabe zu erzielen, erstellen Sie mithilfe der Funktion "Automatisch umcodieren" (verfügbar im Menü "Transformieren") aufeinander folgende Kategorien, die mit 1 beginnen.

Ferner müssen Sie einen Bereich für jede Variable definieren, mit deren Hilfe die Diagramme der Objektscores beschriftet werden. Beschriftungen für Kategorien mit Datenwerten, die außerhalb des für die Variable definierten Bereichs liegen, werden in den Diagrammen jedoch angezeigt.

Nicht lineare kanonische Korrelationsanalyse: Optionen

Das Dialogfeld "Optionen" ermöglicht Ihnen die Auswahl von optionalen Statistiken und Diagrammen, die Speicherung von Objektscores als neue Variablen im aktiven Dataset, die Angabe von Iterations- und Konvergenzkriterien sowie die Angabe einer Ausgangskonfiguration für die Analyse.

Anzeigen. Die verfügbaren Statistiken umfassen: marginale Häufigkeiten (Zähler), Zentroide, Iterationsverlauf, Gewichtungen und Komponentenladungen, Kategoriequantifizierungen, Objektscores und einfache und mehrfache Anpassung.

- *Zentroide.* Kategoriequantifizierungen sowie die projizierten und tatsächlichen Durchschnitte der Objektscores für die in jedem Set enthaltenen Fälle, die zu derselben Kategorie wie die Variable gehören.
- *Gewichtungen und Komponentenladungen.* Die Regressionskoeffizienten in jeder Dimension für jede quantifizierte Variable in einem Set, wobei für die Objektscores eine Regression auf die quantifizierten Variablen durchgeführt wird, sowie die Projektion der quantifizierten Variablen im Objektraum. Liefert einen Hinweis über den Beitrag, den jede Variable zu der Dimension innerhalb eines Sets liefert.
- *Einfache und mehrfache Anpassung.* Maße für die Anpassungsgüte der einfachen und multiplen Kategoriekoordinaten bzw. Kategoriequantifizierungen hinsichtlich der Objekte.
- *Kategoriequantifizierungen.* Zuordnung optimaler Skalenwerte zu den Kategorien einer Variablen.
- *Objektscores.* Der optimale Score, der einem Objekt (Fall) in einer bestimmten Dimension zugewiesen wird.

Diagramm. Sie können Diagramme von Kategoriekoordinaten, Objektscores, Komponentenladungen, Kategoriezentroiden und Transformationen erzeugen.

Objektscores speichern. Sie können die Objektscores als neue Variablen im aktiven Dataset speichern. Objektscores werden für die im Hauptdialogfeld angegebene Anzahl von Dimensionen gespeichert.

Zufällige Ausgangskonfiguration verwenden. Wenn einige oder alle Variablen einzeln nominal sind, sollte eine zufällige Ausgangskonfiguration verwendet werden. Wenn diese Option nicht ausgewählt wird, wird eine verschachtelte Ausgangskonfiguration verwendet.

Kriterien. Sie können die maximale Anzahl von Iterationen festlegen, die in der nicht linearen kanonischen Korrelationsanalyse bei den Berechnungen durchlaufen werden. Außerdem können Sie einen Wert für das Konvergenzkriterium auswählen. Die Analyse wird abgebrochen, wenn die Differenz der Gesamtanpassung zwischen den letzten zwei Iterationen kleiner als der Konvergenzwert ist oder die Maximalzahl von Iterationen erreicht wurde.

Zusätzliche Funktionen beim Befehl OVERALS

Sie können die kategoriale Hauptkomponentenanalyse an Ihre Bedürfnisse anpassen, indem Sie Ihre Auswahl in ein Syntaxfenster einfügen und die daraus resultierende Befehlssyntax für OVERALS bearbeiten. Die Befehlssyntax ermöglicht außerdem Folgendes:

- Festlegen der in Diagrammen darzustellenden Dimensionspaare anstelle der Darstellung aller extrahierten Dimensionen (unter Verwendung des Schlüsselwortes `NDIM` für den Unterbefehl `PLOT`).
- Festlegen der Zeichenanzahl für Wertbeschriftungen zur Beschriftung der Punkte in den Diagrammen (mit dem Unterbefehl `PLOT`).
- Festlegen von mehr als fünf Variablen für die Diagramme der Objektscores (mit dem Unterbefehl `PLOT`).
- Auswahl von in der Analyse verwendeten Variablen als Beschriftungsvariablen für die Diagramme der Objektscores (mit dem Unterbefehl `PLOT`).
- Auswahl von Variablen als Punktbeschriftungen für die Diagramme der Quantifizierungsscores (mit dem Unterbefehl `PLOT`).
- Festlegen der Anzahl von Fällen, die in die Analyse einbezogen werden sollen, falls Sie nicht alle Fälle im aktiven Dataset verwenden möchten (mit dem Unterbefehl `NOBSERVATIONS`).
- Festlegen der Stammnamen für Variablen, die durch Speichern der Objektscores erstellt wurden (mit dem Unterbefehl `SAVE`).
- Festlegen der zu speichernden Anzahl an Dimensionen, sodass nicht alle extrahierten Dimensionen gespeichert werden (mit dem Unterbefehl `SAVE`).
- Schreiben von Kategoriequantifizierungen in eine Matrixdatei (mit dem Unterbefehl `MATRIX`).
- Erzeugen von Diagrammen mit geringer Auflösung, die lesbarer sind als die üblichen hochauflösenden Diagramme (mit dem Befehl `SET`).
- Erstellen von Zentroid- und Transformationsdiagrammen für bestimmte Variablen (mit dem Unterbefehl `PLOT`).

Vollständige Informationen zur Syntax finden Sie in der Befehlssyntaxreferenz.

Kapitel 5. Korrespondenzanalyse

Ein Ziel der Korrespondenzanalyse besteht darin, die Beziehungen zwischen zwei nominalen Variablen in einer Korrespondenztabelle in einem flachdimensionierten Raum und gleichzeitig die Beziehungen zwischen den Kategorien für jede Variable zu beschreiben. Für jede Variable werden die Beziehungen zwischen den Kategorien durch die Distanzen zwischen den Kategoriepunkten in einem Diagramm dargestellt, wobei ähnliche Kategorien jeweils nahe beieinander liegen. Durch die Projizierung von Punkten für eine Variable auf dem Vektor vom Ursprung zu einem Kategoriepunkt einer anderen Variablen wird die Beziehung zwischen den Variablen beschrieben.

Bei einer Analyse der Kontingenztabelle werden häufig Zeilen- und Spaltenprofile untersucht und Tests auf Unabhängigkeit mithilfe der Chi-Quadrat-Statistik durchgeführt. Die Anzahl der Profile kann jedoch unter Umständen relativ groß sein, wobei durch die Chi-Quadrat-Statistik die Abhängigkeitsstruktur nicht erkennbar wird. Die Prozedur "Kreuztabelle" bietet verschiedene Zusammenhangsmaße und -tests, kann jedoch keine Beziehungen zwischen den Variablen darstellen.

Die Faktorenanalyse ist ein Standardverfahren zur Beschreibung von Beziehungen zwischen Variablen in einem flachdimensionierten Raum. Zur Faktorenanalyse werden jedoch Intervalldaten benötigt. Außerdem muss die Anzahl der Beobachtungen das Fünffache der Anzahl der Variablen betragen. Andererseits wird bei der Korrespondenzanalyse von nominalen Variablen ausgegangen, sodass die Beziehungen zwischen den Kategorien jeder Variablen sowie die Beziehungen zwischen den Variablen beschrieben werden können. Zudem kann die Korrespondenzanalyse zur Untersuchung einer beliebigen Tabelle mit positiven Korrespondenzmaßen verwendet werden.

Beispiel. Mithilfe der Korrespondenzanalyse kann die Beziehung zwischen einer Mitarbeitergruppe und den Rauchgewohnheiten grafisch dargestellt werden. Es könnte sich herausstellen, dass die Rauchgewohnheiten bei Angestellten im mittleren Management und beim Sekretariatspersonal jeweils unterschiedlich, bei Sekretariatsmitarbeitern im Vergleich zu Angestellten im gehobenen Management jedoch identisch sind. Als Ergebnis könnte sich ebenfalls die Erkenntnis herausstellen, dass starkes Rauchen mit Angestellten des mittleren Managements und gelegentliches Rauchen mit Sekretärinnen in Verbindung gebracht werden kann.

Statistiken und Diagramme. Korrespondenzmaße, Zeilen- und Spaltenprofile, Einzelwerte, Zeilen- und Spaltenscores, Trägheit, Masse, Konfidenzstatistik für Zeilen- und Spaltenscores, Konfidenzstatistik für Einzelwerte, Transformationsdiagramme, Zeilenpunktdiagramme, Spaltenpunktdiagramme und Biplots.

Erläuterungen der Daten für die Korrespondenzanalyse

Daten. Die kategorialen Variablen, die analysiert werden sollen, werden nominal skaliert. Verwenden Sie für aggregierte Daten oder für ein Korrespondenzmaß (außer Häufigkeiten) eine Gewichtungsvariable mit positiven Ähnlichkeitswerten. Sie können für Tabellendaten auch die Syntax zum Lesen der Tabelle verwenden.

Annahmen. Die maximale Anzahl der in der Prozedur verwendeten Dimensionen ist von der Zahl der aktiven Zeilen- und Spaltenkategorien und der Anzahl der Gleichheitsbedingungen abhängig. Wenn keine Gleichheitsbedingungen verwendet werden und alle Kategorien aktiv sind, ist die maximale Dimensionalität um eins kleiner als die Anzahl der Kategorien in der Variablen mit den wenigsten Kategorien. Wenn beispielsweise eine Variable über fünf und eine andere über vier Kategorien verfügt, beträgt die maximale Anzahl an Dimensionen drei. Zusätzliche Kategorien sind nicht aktiv. Wenn beispielsweise eine Variable über fünf Kategorien (davon zwei Zusatzkategorien) und eine andere über vier Kategorien verfügt, beträgt die maximale Anzahl an Dimensionen zwei. Behandeln Sie alle Kategorie-Sets, die auf Gleichheit beschränkt sind, als eine Kategorie. Wenn beispielsweise eine Variable über fünf Kategorien verfügt, von denen drei auf Gleichheit beschränkt sein müssen, muss diese Variable bei der Bestimmung der maxima-

len Dimensionalität so behandelt werden, als ob sie drei Kategorien besitzen würde. Für zwei der Kategorien gelten keine Nebenbedingungen, und die dritte Kategorie entspricht den drei Kategorien mit der Gleichheitsbedingung. Wenn Sie für die Anzahl der Dimensionen eine über dem Maximalwert liegende Zahl angeben, wird in jedem Fall der Maximalwert verwendet.

Verwandte Prozeduren. Falls Sie mit mehr als zwei Variablen arbeiten, verwenden Sie die Mehrfachkorrespondenzanalyse. Wenn die Variablen ordinal skaliert werden sollen, verwenden Sie die kategoriale Hauptkomponentenanalyse.

So aktivieren Sie eine Korrespondenzanalyse:

1. Wählen Sie in den Menüs Folgendes aus:
Analysieren > Dimensionsreduktion > Korrespondenzanalyse...
2. Wählen Sie eine Zeilenvariable aus.
3. Wählen Sie eine Spaltenvariable aus.
4. Definieren Sie die Bereiche für die Variablen.
5. Klicken Sie auf **OK**.

Definieren des Zeilenbereichs in der Korrespondenzanalyse

Sie müssen einen Bereich für die Zeilenvariable angeben. Die angegebenen Minimal- und Maximalwerte müssen ganzzahlig sein. Gebrochene Datenwerte werden in der Analyse abgeschnitten. Ein Kategorie-wert, der außerhalb des angegebenen Bereiches liegt, wird in der Analyse ignoriert.

Alle Kategorien sind zunächst aktiv und nicht durch Nebenbedingungen eingeschränkt. Sie können festlegen, dass bestimmte Zeilenkategorien anderen Zeilenkategorien entsprechen müssen. Außerdem können Sie eine Zeilenkategorie als ergänzende Kategorie definieren.

- **Kategorien müssen gleich sein.** Die Kategorien müssen gleiche Werte haben. Verwenden Sie Gleichheitsbedingungen, wenn die erhaltene Ordnung der Kategorien nicht erwünscht oder nicht intuitiv ist. Die maximale Anzahl der Zeilenkategorien, für die Gleichheitsbedingungen festgelegt werden können, entspricht der Gesamtzahl aktiver Zeilenkategorien minus 1. Zum Definieren anderer Gleichheitsbedingungen für Kategoriesets können Sie die Syntax verwenden. Beispielsweise können Sie mithilfe der Syntax festlegen, dass die Kategorien 1 und 2 sowie die Kategorien 3 und 4 gleich sein müssen.
- **Ergänzende Kategorie.** Ergänzende Kategorien haben keine Auswirkungen auf die Analyse, werden jedoch in dem durch die aktiven Kategorien definierten Raum dargestellt. Ergänzende Kategorien spielen keine Rolle bei der Definition von Dimensionen. Die maximale Anzahl der ergänzenden Zeilenkategorien entspricht der Gesamtanzahl der Zeilenkategorien minus 2.

Definieren des Spaltenbereichs in der Korrespondenzanalyse

Sie müssen einen Bereich für die Spaltenvariable angeben. Die angegebenen Minimal- und Maximalwerte müssen ganzzahlig sein. Gebrochene Datenwerte werden in der Analyse abgeschnitten. Ein Kategorie-wert, der außerhalb des angegebenen Bereiches liegt, wird in der Analyse ignoriert.

Alle Kategorien sind zunächst aktiv und nicht durch Nebenbedingungen eingeschränkt. Sie können festlegen, dass bestimmte Spaltenkategorien anderen Spaltenkategorien entsprechen müssen. Außerdem können Sie eine Spaltenkategorie als ergänzende Kategorie definieren.

- **Kategorien müssen gleich sein.** Die Kategorien müssen gleiche Werte haben. Verwenden Sie Gleichheitsbedingungen, wenn die erhaltene Ordnung der Kategorien nicht erwünscht oder nicht intuitiv ist. Die maximale Anzahl der Spaltenkategorien, für die Gleichheitsbedingungen festgelegt werden können, entspricht der Gesamtzahl aktiver Spaltenkategorien minus 1. Zum Definieren anderer Gleichheitsbedingungen für Kategoriesets können Sie die Syntax verwenden. Beispielsweise können Sie mithilfe der Syntax festlegen, dass die Kategorien 1 und 2 sowie die Kategorien 3 und 4 gleich sein müssen.

- **Ergänzende Kategorie.** Ergänzende Kategorien haben keine Auswirkungen auf die Analyse, werden jedoch in dem durch die aktiven Kategorien definierten Raum dargestellt. Ergänzende Kategorien spielen keine Rolle bei der Definition von Dimensionen. Die maximale Anzahl der ergänzenden Spaltenkategorien entspricht der Gesamtanzahl der Spaltenkategorien minus 2.

Korrespondenzanalyse: Modell

Im Dialogfeld "Modell" können Sie die Anzahl der Dimensionen, die Distanzmaße sowie die Standardisierungsmethode und die Normalisierungsmethode angeben.

Dimensionen in der Lösung. Legen Sie die Anzahl der Dimensionen fest. Im Allgemeinen sollten Sie nur so viele Dimensionen auswählen, wie zum Erklären des größten Teils der Variation notwendig ist. Die maximal zulässige Dimensionszahl hängt von der Anzahl der in der Analyse und in den Gleichheitsbedingungen verwendeten aktiven Kategorien ab. Die maximale Anzahl an Dimensionen ist jeweils der kleinere von folgenden Werten:

- Die Anzahl der aktiven Zeilenkategorien minus der Anzahl der Zeilenkategorien, die gleich sein müssen, plus der Anzahl der eingeschränkten Zeilenkategoriegruppen.
- Die Anzahl der aktiven Spaltenkategorien minus der Anzahl der Spaltenkategorien, die gleich sein müssen, plus der Anzahl der eingeschränkten Spaltenkategoriegruppen.

Distanzmaß. Sie können das Distanzmaß für Zeilen und Spalten der Korrespondenztabelle auswählen. Wählen Sie eine der folgenden Optionen:

- **Chi-Quadrat.** Verwendet eine gewichtete Profildistanz, wobei die Gewichtung der Masse der Zeilen oder Spalten entspricht. Dieses Maß ist für die Standardkorrespondenzanalyse erforderlich.
- **Euklidisch.** Verwendet die Quadratwurzel der Summe aus den quadrierten Differenzen von Zeilen- und Spaltenpaaren.

Standardisierungsmethode. Wählen Sie eine der folgenden Optionen:

- **Zeilen- und Spaltenmittel werden entfernt.** Sowohl Zeilen als auch Spalten werden zentriert. Diese Methode ist für die Standardkorrespondenzanalyse erforderlich.
- **Zeilenmittel werden entfernt.** Nur die Zeilen werden zentriert.
- **Spaltenmittel werden entfernt.** Nur die Spalten werden zentriert.
- **Zeilensummen werden angeglichen und Mittelwerte entfernt.** Vor dem Zentrieren der Zeilen werden die Zeilenränder angeglichen.
- **Spaltensummen werden angeglichen und Mittelwerte entfernt.** Vor dem Zentrieren der Spalten werden die Spaltenränder angeglichen.

Normalisierungsmethode. Wählen Sie eine der folgenden Optionen:

- **Symmetrisch.** Für jede Dimension sind die Zeilenscores der gewichtete Durchschnitt der Spaltenscores geteilt durch den entsprechenden Einzelwert, und die Spaltenscores sind der gewichtete Durchschnitt der Zeilenscores geteilt durch den entsprechenden Einzelwert. Verwenden Sie diese Methode, wenn Sie die Unterschiede oder Ähnlichkeiten zwischen Kategorien der beiden Variablen untersuchen möchten.
- **Prinzipal.** Die Distanzen zwischen Zeilenpunkten und Spaltenpunkten sind Approximationen der Distanzen in der Korrespondenztabelle entsprechend dem gewählten Distanzmaß. Verwenden Sie diese Methode, wenn Sie die Unterschiede zwischen Kategorien einer oder in beiden Variablen anstelle der Unterschiede zwischen den beiden Variablen untersuchen möchten.
- **Zeilenprinzipal.** Die Distanzen zwischen Zeilenpunkten sind Approximationen der Distanzen in der Korrespondenztabelle entsprechend dem gewählten Distanzmaß. Die Zeilenscores entsprechen dem gewichteten Durchschnitt der Spaltenscores. Verwenden Sie diese Methode, wenn Sie die Unterschiede oder Ähnlichkeiten zwischen Kategorien der Zeilenvariable untersuchen möchten.
- **Spaltenprinzipal.** Die Distanzen zwischen Spaltenpunkten sind Approximationen der Distanzen in der Korrespondenztabelle entsprechend dem gewählten Distanzmaß. Die Spaltenscores entsprechen dem

gewichteten Durchschnitt der Zeilenscores. Verwenden Sie diese Methode, wenn Sie die Unterschiede oder Ähnlichkeiten zwischen Kategorien der Spaltenvariablen untersuchen möchten.

- **Benutzerdefiniert.** Sie müssen einen Wert zwischen 1 und -1 angeben. Der Wert -1 entspricht dem Spaltenprinzipal. Der Wert 1 entspricht dem Zeilenprinzipal. Der Wert 0 entspricht der Einstellung "Symmetrisch". Bei allen übrigen Werten wird die Trägheit in unterschiedlichem Maße über Zeilen- und Spaltenscores gestreut. Mit dieser Methode können Sie Biplots genau an Ihre Bedürfnisse anpassen.

Korrespondenzanalyse: Statistiken

Im Dialogfeld "Statistiken" können Sie festlegen, welche numerische Ausgabe erzeugt werden soll.

Korrespondenztabelle. Eine Kreuztabelle für die Eingabevariable mit Randsummen der Zeilen und Spalten.

Übersicht der Zeilenpunkte. Für jede Zeilenkategorie werden die Scores, die Masse, die Trägheit, der Anteil der Dimension an der Trägheit und der Anteil der Dimension an der Trägheit des Punkts angezeigt.

Übersicht der Spaltenpunkte. Für jede Spaltenkategorie werden die Scores, die Masse, die Trägheit, der Anteil der Dimension an der Trägheit und der Anteil der Dimension an der Trägheit des Punkts angezeigt.

Zeilenprofile. Für jede Zeilenkategorie wird die Verteilung über die Kategorien der Spaltenvariablen angezeigt.

Spaltenprofile. Für jede Spaltenkategorie wird die Verteilung über die Kategorien der Zeilenvariablen angezeigt.

Permutationen der Korrespondenztabelle. Die Korrespondenztabelle wird so umstrukturiert, dass die Zeilen und Spalten entsprechend den Scores der ersten Dimension in aufsteigender Reihenfolge angezeigt werden. Optional können Sie auch eine maximale Anzahl Dimensionen angeben, für die Permutationstabellen erzeugt werden sollen. Dadurch wird für jede Dimension von 1 bis zur angegebenen Zahl eine Permutationstabelle erzeugt.

Konfidenzstatistiken für Zeilenpunkte. Bezieht die Standardabweichung und Korrelationen für alle nicht ergänzenden Zeilenpunkte mit ein.

Konfidenzstatistiken für Spaltenpunkte. Bezieht die Standardabweichung und Korrelationen für alle nicht ergänzenden Spaltenpunkte mit ein.

Korrespondenzanalyse: Diagramme

Im Dialogfeld "Diagramme" können Sie festlegen, welche Diagramme erzeugt werden sollen.

Streudiagramme. Erzeugt eine Matrix mit allen paarweisen Diagrammen der Dimensionen. Folgende Streudiagramme sind verfügbar:

- **Biplot.** Erzeugt eine Matrix mit verbundenen Diagrammen der Zeilen- und Spaltenpunkte. Bei Wahl der Normalisierungsmethode "Prinzipal" ist die Option "Biplot" nicht verfügbar.
- **Zeilenpunkte.** Erzeugt eine Matrix mit Diagrammen der Zeilenpunkte.
- **Spaltenpunkte.** Erzeugt eine Matrix mit Diagrammen der Spaltenpunkte.

Optional können Sie angeben, wie viele Zeichen in Wertbeschriftungen zur Beschriftung der Punkte verwendet werden sollen. Dieser Wert muss eine nicht negative Ganzzahl sein, die kleiner oder gleich 20 ist.

Liniendiagramme. Erzeugt ein Diagramm für jede Dimension der gewählten Variable. Folgende Liniendiagramme sind verfügbar:

- **Transformierte Zeilenkategorien.** Erzeugt ein Diagramm unter Verwendung der ursprünglichen Zeilenkategoriewerte und der entsprechenden Zeilenscores.
- **Transformierte Spaltenkategorien.** Erzeugt ein Diagramm unter Verwendung der ursprünglichen Spaltenkategoriewerte und der entsprechenden Spaltenscores.

Optional können Sie angeben, wie viele Zeichen in Wertbeschriftungen zur Beschriftung der Kategorieachse verwendet werden sollen. Dieser Wert muss eine nicht negative Ganzzahl sein, die kleiner oder gleich 20 ist.

Diagrammdimensionen. Hiermit legen Sie fest, wie die Dimensionen in der Ausgabe angezeigt werden.

- **Alle Dimensionen in der Lösung anzeigen.** Alle Dimensionen in der Lösung werden in einer Streudiagrammmatrix angezeigt.
- **Anzahl der Dimensionen beschränken.** Die angezeigten Dimensionen werden auf geplottete Paare beschränkt. Wenn Sie die Dimensionen beschränken, müssen Sie die niedrigste und die höchste Dimension auswählen, die jeweils geplottet werden soll. Die niedrigste Dimension muss im Bereich von 1 bis zur Anzahl der Dimensionen in der Lösung minus 1 liegen und wird gegen höhere Dimensionen geplottet. Der Wert für die höchste Dimension muss im Bereich von 2 bis zur Anzahl der Dimensionen in der Lösung liegen und gibt die höchste Dimension an, die beim Plotten der Dimensionspaare verwendet wird. Diese Spezifikation gilt für alle angeforderten multidimensionalen Plots.

Zusätzliche Funktionen beim Befehl CORRESPONDENCE

Sie können die Korrespondenzanalyse an Ihre Bedürfnisse anpassen, wenn Sie Ihre Auswahl in ein Syntaxfenster einfügen und die resultierende Befehlssyntax für den Befehl CORRESPONDENCE bearbeiten. Die Befehlssyntax ermöglicht außerdem Folgendes:

- Angeben der Tabellendaten als Eingabedaten anstelle der Verwendung fallweise gebundener Daten (Unterbefehl TABLE = ALL).
- Angeben der Anzahl Zeichen für Wertbeschriftungen zur Beschriftung von Punkten für jeden Typ der Streudiagramm- oder Biplot-Matrix (mit dem Unterbefehl PLOT).
- Angeben der Anzahl Zeichen für Wertbeschriftungen zur Beschriftung von Punkten für jeden Typ von Liniendiagramm (mit dem Unterbefehl PLOT).
- Schreiben einer Matrix mit Zeilen- und Spaltenscores in eine Matrixdatendatei (mit dem Unterbefehl OUTFILE).
- Schreiben einer Matrix der Konfidenzstatistiken (Varianzen und Kovarianzen) für die Einzelwerte und die Scores in eine Matrixdatendatei (mit dem Unterbefehl OUTFILE).
- Festlegen der Gleichheit für mehrere Categoriesets (mit dem Unterbefehl EQUAL).

Vollständige Informationen zur Syntax finden Sie in der Befehlssyntaxreferenz.

Kapitel 6. Mehrfachkorrespondenzanalyse

Die Mehrfachkorrespondenzanalyse quantifiziert nominale (kategoriale) Daten durch Zuweisung numerischer Werte zu den Fällen (Objekten) und Kategorien, sodass Objekte in derselben Kategorie eng beieinander und Objekte in verschiedenen Kategorien weit voneinander entfernt liegen. Jedes Objekt liegt so nah wie möglich an den Kategoriepunkten der Kategorien, die für dieses Objekt gelten. So unterteilen die Kategorien die Objekte in homogene Untergruppen. Variablen werden als homogen betrachtet, wenn sie Objekte, die in denselben Kategorien vorliegen, auch in dieselben Untergruppen einordnen.

Beispiel. Mithilfe der Mehrfachkorrespondenzanalyse könnte die Beziehung zwischen Berufsgruppe, Minderheit und Geschlecht grafisch dargestellt werden. Unter Umständen stellt sich dabei heraus, dass Minderheit und Geschlecht im Gegensatz zur Berufsgruppe klare Unterscheidungsmerkmale für Personen darstellen. Außerdem könnte sich ergeben, dass die Kategorien "Latino" und "Afro-Amerikaner" einander sehr ähnlich sind.

Statistiken und Diagramme. Objektscores, Diskriminationsmaße, Iterationsverlauf, Korrelationen der ursprünglichen und transformierten Variablen, Kategoriequantifizierungen, deskriptive Statistiken, Objektpunktdiagramme, Biplots, Kategoriediagramme, verbundene Kategoriediagramme, Transformationsdiagramme und Diskriminationsmaßprotokolle.

Erläuterungen der Daten für die Mehrfachkorrespondenzanalyse

Daten. Werte von Zeichenfolgevariablen werden immer in aufsteigender alphanumerischer Reihenfolge in positive Ganzzahlen umgewandelt. Benutzerdefinierte und systemdefiniert fehlende Werte sowie Werte kleiner als 1 werden als fehlend betrachtet. Sie können die Daten umcodieren oder eine Konstante zu Variablen mit Werten kleiner als 1 addieren, um die Werte als nicht fehlend zu deklarieren.

Annahmen. Alle Variablen weisen ein mehrfach nominales Skalierungsniveau auf. Die Daten müssen mindestens drei gültige Fälle enthalten. Die Analyse basiert auf Daten, die aus positiven ganzen Zahlen bestehen. Mit der Option zur Diskretisierung werden Variablen mit Dezimalwerten automatisch kategorisiert, indem die Werte in nahezu normalverteilte Kategorien gruppiert werden, und Werte von Zeichenfolgevariablen werden automatisch in positive Ganzzahlen umgewandelt. Sie können andere Diskretisierungsmethoden bestimmen.

Verwandte Prozeduren. Bei zwei Variablen entspricht die Mehrfachkorrespondenzanalyse der Korrespondenzanalyse. Wenn Sie annehmen, dass die Variablen ordinale oder numerische Eigenschaften besitzen, sollte die kategoriale Hauptkomponentenanalyse verwendet werden. Wenn Variablensets untersucht werden sollen, verwenden Sie die nicht lineare kanonische Korrelationsanalyse.

So aktivieren Sie eine Mehrfachkorrespondenzanalyse:

1. Wählen Sie in den Menüs Folgendes aus:
Analysieren > Dimensionsreduktion > Optimale Skalierung...
2. Wählen Sie die Option **Alle Variablen sind mehrfach nominal** aus.
3. Wählen Sie die Option **Ein Set** aus.
4. Klicken Sie auf **Definieren**.
5. Wählen Sie mindestens zwei Analysevariablen aus und geben Sie die Anzahl der Dimensionen in der Lösung an.
6. Klicken Sie auf **OK**.

Sie können außerdem Beschriftungsvariablen für die Diagramme und Zusatzvariablen bestimmen, die in die gefundene Lösung eingepasst werden.

Definieren der Variablengewichtung in der Mehrfachkorrespondenzanalyse

Sie können die Gewichtung für die Analysevariablen festlegen.

Variablengewichtung. Sie können für jede Variable eine Gewichtung festlegen. Bei dem angegebenen Wert muss es sich um eine positive ganze Zahl handeln. Der Standardwert ist 1.

Mehrfachkorrespondenzanalyse: Diskretisierung

Im Dialogfeld "Diskretisierung" können Sie eine Methode zum Umcodieren der Variablen festlegen. Sofern nicht anders angegeben, werden Variablen mit Dezimalwerten in sieben Kategorien gruppiert, die nahezu normalverteilt sind. (Wenn die Variable weniger als sieben unterschiedliche Werte aufweist, entspricht die Anzahl der Kategorien der Anzahl der unterschiedlichen Werte.) Werte von Zeichenfolgevariablen werden immer in positive ganze Zahlen umgewandelt, indem in aufsteigender alphanumerischer Reihenfolge Kategorieindikatoren zugewiesen werden. Die Diskretisierung von Zeichenfolgevariablen erfolgt anhand dieser Ganzzahlen. Andere Variablen werden in der Standardeinstellung nicht modifiziert. Die diskretisierten Variablen werden dann in der Analyse verwendet.

Methode. Verfügbar sind die Optionen "Gruppieren", "Einstufen" oder "Multiplizieren".

- **Gruppieren.** Hierbei werden die Daten in eine bestimmte Anzahl von Kategorien oder nach Intervall umcodiert.
- **Einstufen.** Die Variable wird durch Bilden einer Rangfolge diskretisiert.
- **Multiplizieren.** Die aktuellen Werte der Variablen werden standardisiert, mit 10 multipliziert und gerundet. Anschließend wird eine Konstante zu den Werten addiert, sodass der niedrigste diskretisierte Wert 1 beträgt.

Gruppieren. Folgende Optionen sind bei der Diskretisierung von Variablen durch Gruppieren verfügbar:

- **Anzahl Kategorien.** Geben Sie eine Anzahl von Kategorien an. Legen Sie außerdem fest, ob die Werte der Variablen über die Kategorien ungefähr normal- oder gleichverteilt werden sollen.
- **Gleiche Intervalle.** Die Variablen werden in Kategorien umcodiert, die durch diese gleich großen Intervalle definiert werden. Sie müssen die Größe der Intervalle angeben.

Mehrfachkorrespondenzanalyse: Fehlende Werte

Im Dialogfeld "Fehlende Werte" können Sie die Strategie für die Verarbeitung von fehlenden Werten in den Analyse- und Zusatzvariablen festlegen.

Strategie für fehlenden Wert. Fehlende Werte können ausgeschlossen (Passivbehandlung) oder imputiert (Aktivbehandlung) werden oder Objekte mit fehlenden Werten können ausgeschlossen werden (listenweiser Ausschluss).

- **Fehlende Werte ausschließen; für Korrelationen nach Quantifizierung imputieren.** Objekte mit fehlenden Werten bei der ausgewählten Variablen tragen nicht zur Analyse dieser Variablen bei. Wenn alle Variablen passiv behandelt werden, werden Objekte mit fehlenden Werten für alle Variablen als Zusatzobjekte behandelt. Wenn im Dialogfeld "Ausgabe" Korrelationen ausgewählt sind, werden fehlende Werte nach der Analyse anhand der häufigsten Variablenkategorie (Modalwert) für die Korrelationen der ursprünglichen Variablen gesetzt. Sie können die Methode auswählen, nach der Werte für die Korrelationen der optimal skalierten Variablen gesetzt werden. Wenn Sie **Modalwert** auswählen, werden fehlende Werte durch den Modalwert der optimal skalierten Variablen ersetzt. Wenn Sie **Zusatzkategorie** auswählen, werden fehlende Werte durch die Quantifizierung einer zusätzlichen Kategorie ersetzt. Dies bedeutet aber auch, dass Objekte mit einem fehlenden Wert für diese Variable als zur gleichen Kategorie (der zusätzlichen) zugehörig betrachtet werden.

- **Fehlende Werte imputieren.** Fehlende Werte werden bei Objekten mit fehlenden Werten für die ausgewählte Variable gesetzt. Sie können die Methode auswählen, nach der die Werte gesetzt werden: Wählen Sie **Modalwert**, wenn fehlende Werte durch die häufigste Kategorie ersetzt werden sollen. Falls mehrere Modalwerte vorhanden sind, wird der Modalwert mit dem kleinsten Kategorieindikator verwendet. Wenn Sie **Zusatzkategorie** auswählen, werden fehlende Werte durch die gleiche Quantifizierung einer zusätzlichen Kategorie ersetzt. Dies bedeutet aber auch, dass Objekte mit einem fehlenden Wert für diese Variable als zur gleichen Kategorie (der zusätzlichen) zugehörig betrachtet werden.
- **Objekte mit fehlenden Werten für diese Variable ausschließen.** Objekte mit fehlenden Werten für die ausgewählte Variable werden aus der Analyse ausgeschlossen. Diese Strategie ist für Zusatzvariablen nicht verfügbar.

Mehrfachkorrespondenzanalyse: Optionen

Im Dialogfeld "Optionen" können Sie die anfängliche Konfiguration auswählen, Kriterien für die Iterationen und die Konvergenz festlegen, eine Normalisierungsmethode auswählen, die Methode zum Beschriften der Diagramme auswählen und Zusatzobjekte angeben.

Zusatzobjekte. Geben Sie die Fallnummer des Objekts (oder die erste und letzte Fallnummer einer Reihe von Objekten) an, die als Zusatzobjekte deklariert werden sollen, und klicken Sie dann auf **Hinzufügen**. Wiederholen Sie diesen Vorgang, bis Sie alle Zusatzobjekte festgelegt haben. Wenn ein Objekt als Zusatzobjekt deklariert ist, werden Fallgewichtungen für das Objekt ignoriert.

Normalisierungsmethode. Sie können eine von fünf Optionen für die Normalisierung der Objektscores und Variablen auswählen. In einer Analyse kann nur eine Normalisierungsmethode gleichzeitig verwendet werden.

- **Variablenprinzipal.** Bei dieser Option wird die Assoziation zwischen Variablen optimiert. Die Komponentenladungen (Korrelationen mit Hauptkomponenten wie Dimensionen und Objektscores) stellen die Koordinaten der Variablen im Objektraum dar. Dies ist nützlich, wenn Sie hauptsächlich an der Korrelation zwischen den Variablen interessiert sind.
- **Objektprinzipal.** Bei dieser Option werden die Distanzen zwischen den Objekten optimiert. Dies ist nützlich, wenn Sie hauptsächlich an den Unterschieden oder Übereinstimmungen zwischen den Objekten interessiert sind.
- **Symmetrisch.** Verwenden Sie diese Normalisierungsmethode, wenn Sie hauptsächlich an den Beziehungen zwischen den Objekten und den Variablen interessiert sind.
- **Unabhängig.** Verwenden Sie diese Normalisierungsmethode, wenn Sie die Distanzen zwischen den Objekten und die Korrelationen zwischen den Variablen getrennt untersuchen möchten.
- **Benutzerdefiniert.** Sie können einen beliebigen reellen Wert aus dem abgeschlossenen Intervall $[-1, 1]$ eingeben. Der Wert 1 entspricht der Methode "Objektprinzipal", der Wert 0 der Methode "Symmetrisch" und der Wert -1 der Methode "Variablenprinzipal". Bei Werten größer als -1 und kleiner als 1 wird der Eigenwert sowohl über Objekte als auch Variablen gestreut. Mit dieser Methode können Sie Biplots und Triplots genau an Ihre Bedürfnisse anpassen.

Kriterien. Sie können die maximale Anzahl von Iterationen festlegen, die in der Prozedur bei den Berechnungen durchlaufen werden. Außerdem können Sie einen Wert für das Konvergenzkriterium auswählen. Die Prozedur wird abgebrochen, wenn die Differenz der Gesamtanpassung zwischen den letzten zwei Iterationen kleiner als der Konvergenzwert ist oder die Maximalzahl von Iterationen erreicht wurde.

Diagramme beschriften mit. Hiermit können Sie festlegen, ob in den Diagrammen Variablen- und Wertbeschriftungen oder Variablennamen und Werte verwendet werden sollen. Sie können auch eine Höchstlänge für die Beschriftungen festlegen.

Diagrammdimensionen. Hiermit legen Sie fest, wie die Dimensionen in der Ausgabe angezeigt werden.

- **Alle Dimensionen in der Lösung anzeigen.** Alle Dimensionen in der Lösung werden in einer Streudiagrammmatrix angezeigt.

- **Anzahl der Dimensionen beschränken.** Die angezeigten Dimensionen werden auf geplottete Paare beschränkt. Wenn Sie die Dimensionen beschränken, müssen Sie die niedrigste und die höchste Dimension auswählen, die jeweils geplottet werden soll. Die niedrigste Dimension muss im Bereich von 1 bis zur Anzahl der Dimensionen in der Lösung minus 1 liegen und wird gegen höhere Dimensionen geplottet. Der Wert für die höchste Dimension muss im Bereich von 2 bis zur Anzahl der Dimensionen in der Lösung liegen und gibt die höchste Dimension an, die beim Plotten der Dimensionspaare verwendet wird. Diese Spezifikation gilt für alle angeforderten multidimensionalen Plots.

Konfiguration. Sie können die Koordinaten einer Konfiguration aus einer Datei einlesen. Die erste Variable in der Datei muss die Koordinaten der ersten Dimension, die zweite Variable die Koordinaten der zweiten Dimension usw. enthalten.

- **Initial.** Die in der Datei angegebene Konfiguration wird als Ausgangspunkt für die Analyse verwendet.
- **Fest.** Die in der Datei angegebene Konfiguration wird zum Anpassen der Variablen verwendet. Die Variablen, die angepasst werden, müssen als Analysevariablen ausgewählt werden. Da die Konfiguration aber fest ist, werden Sie als Zusatzvariablen verarbeitet (und müssen deshalb nicht als Zusatzvariablen ausgewählt werden).

Mehrfachkorrespondenzanalyse: Ausgabe

Im Dialogfeld "Ausgabe" können Sie festlegen, dass Tabellen für Objektscores, Diskriminationsmaße, der Iterationsverlauf, Korrelationen der ursprünglichen und transformierten Variablen, Kategoriequantifizierungen für ausgewählte Variablen und deskriptive Statistiken für ausgewählte Variablen erstellt werden sollen.

Objektscores. Hiermit werden die Objektscores (z. B. Masse, Trägheit, Beiträge) angezeigt. Die folgenden Optionen sind verfügbar:

- **Diese Kategorien berücksichtigen.** Hiermit werden die Kategorieindikatoren der ausgewählten Analysevariablen angezeigt.
- **Objektscores beschriften nach.** Sie können eine der in der Liste Beschriftungsvariablen enthaltenen Variablen auswählen, um die Objekte zu beschriften.

Diskriminationsmaße. Zeigt die Diskriminationsmaße pro Variable und Dimension an.

Iterationsverlauf. Für jede Iteration werden berücksichtigte Varianz, Verlust sowie Steigerung der berücksichtigten Varianz ausgegeben.

Korrelationen der Originalvariablen. Hiermit werden die Korrelationsmatrix der ursprünglichen Variablen und die Eigenwerte dieser Matrix ausgegeben.

Korrelationen der transformierten Variablen. Hiermit werden die Korrelationsmatrix der transformierten (optimal skalierten) Variablen und die Eigenwerte dieser Matrix ausgegeben.

Kategoriequantifizierungen und Beiträge. Dies sind die Kategoriequantifizierungen und (Koordinaten), einschließlich Masse, Trägheit und Beiträge, pro Dimension jeder ausgewählten Variable.

Hinweis: Die Koordinaten und Beiträge (unter anderem Masse und Trägheit) werden in separaten Schichten der Ausgabe von Pivot-Tabellen angezeigt, wobei die Koordinaten standardmäßig eingebledet werden. Um die Beiträge anzuzeigen, aktivieren Sie die Tabelle (durch Doppelklicken) und wählen Sie "Beiträge" aus der Dropdown-Liste "Schicht" aus.

Deskriptive Statistiken. Hiermit werden Häufigkeiten, Anzahl der fehlenden Werte und Modalwert für die ausgewählten Variablen ausgegeben.

Mehrfachkorrespondenzanalyse: Speichern

Im Dialogfeld "Speichern" können Sie festlegen, dass diskretisierte Daten, Objektscores und transformierte Werte in einer IBM SPSS Statistics-Datendatei oder in einem Dataset gespeichert werden. Außerdem können die transformierten Werte und die Objektscores im aktiven Dataset gespeichert werden.

- Datasets sind in der aktuellen Sitzung verfügbar. In nachfolgenden Sitzungen stehen Datasets nur zur Verfügung, wenn Sie sie ausdrücklich als Datendateien speichern. Die Namen von Datasets müssen den Regeln zum Benennen von Variablen entsprechen.
- Die Namen für die Dateien und Datasets müssen sich bei allen Arten von gespeicherten Daten unterscheiden.
- Wenn Sie die Objektscores oder die transformierten Werte im aktiven Dataset speichern, können Sie die Anzahl der mehrfach nominalen Dimensionen angeben.

Mehrfachkorrespondenzanalyse: Objektdiagramme

Im Dialogfeld "Objektdiagramme" können Sie festlegen, welche Diagramme und für welche Variablen die Diagramme erstellt werden sollen.

Objektpunkte. Es wird ein Diagramm der Objektpunkte angezeigt.

Objekte und Zentroide (Biplot). Die Objektpunkte werden mit den Variablenzentroiden geplottet.

Biplotvariablen. Die Biplots können auf allen Variablen oder einem Subset der Variablen basieren.

Objekte beschriften. Objekte können mit den Kategorien ausgewählter Variablen (im Dialogfeld "Optionen" können Sie Kategorieindikatoren oder Wertbeschriftungen wählen) oder den Fallnummern beschriftet werden. Wenn **Variable** ausgewählt ist, wird pro Variable ein Diagramm erstellt.

Mehrfachkorrespondenzanalyse: Variablendiagramme

Im Dialogfeld "Variablendiagramme" können Sie festlegen, welche Diagramme und für welche Variablen die Diagramme erstellt werden sollen.

Kategoriediagramme. Für jede ausgewählte Variable wird ein Diagramm mit Zentroidkoordinaten erstellt. Die Kategorien werden in den Zentroiden der Objekte in den einzelnen Kategorien platziert.

Verbundene Kategoriediagramme. Hierbei handelt es sich um ein einzelnes Diagramm mit Zentroidkoordinaten jeder ausgewählten Variablen.

Transformationsdiagramme. Hierbei wird ein Diagramm der optimalen Kategoriequantifizierungen gegen die Kategorieindikatoren erstellt. Sie können die Anzahl der Dimensionen angeben. Pro Dimension wird ein Diagramm erstellt. Sie können außerdem festlegen, dass für jede ausgewählte Variable ein Diagramm der Residuen ausgegeben wird.

Diskriminationsmaße. Erstellt ein einzelnes Diagramm der Diskriminationsmaße für die ausgewählten Variablen.

Zusätzliche Funktionen beim Befehl MULTIPLE CORRESPONDENCE

Sie können die Mehrfachkorrespondenzanalyse an Ihre Bedürfnisse anpassen, wenn Sie Ihre Auswahl in ein Syntaxfenster einfügen und die resultierende Befehlssyntax für den Befehl MULTIPLE CORRESPONDENCE bearbeiten. Die Befehlssyntax ermöglicht außerdem Folgendes:

- Mit dem Unterbefehl SAVE können Sie Stammnamen für die transformierten Variablen, Objektscores und Approximationen festlegen, wenn Sie diese im aktiven Dataset speichern.

- Mit dem Unterbefehl PLOT kann die Höchstlänge für Beschriftungen für jedes Diagramm getrennt festgelegt werden.
- Mit dem Unterbefehl PLOT kann eine separate Variablenliste für Residuendiagramme festgelegt werden.

Vollständige Informationen zur Syntax finden Sie in der Befehlssyntaxreferenz.

Kapitel 7. Multidimensionale Skalierung (PROXSCAL)

Bei der multidimensionalen Skalierung wird versucht, die Struktur in einem Set von Ähnlichkeitsmaßen zwischen Objekten zu ermitteln. Dies wird durch das Zuweisen von Beobachtungen zu bestimmten Positionen in einem konzeptuellen Raum mit wenigen Dimensionen erzielt, wobei die Abstände zwischen den Punkten des Raumes mit den vorgegebenen Unähnlichkeiten bzw. Ähnlichkeiten so gut wie möglich übereinstimmen. Als Ergebnis werden die Objekte in diesem Raum mit wenigen Dimensionen mithilfe der Methode der kleinsten Quadrate dargestellt, was häufig zu einem besseren Verständnis der Daten beiträgt.

Beispiel. Die multidimensionale Skalierung kann das Ermitteln von Wahrnehmungsbeziehungen erheblich erleichtern. Der Hersteller eines Produkts kann zum Beispiel eine Umfrage durchführen, um das Produktimage zu ermitteln und ein Dataset zu erhalten, das die wahrgenommene Ähnlichkeit (Nähe) des Produkts mit Konkurrenzprodukten beschreibt. Unter Verwendung dieser Ähnlichkeitswerte und unabhängiger Variablen (z. B. des Preises) kann der Hersteller dann feststellen, welche Variablen für die Produktwahrnehmung durch die Verbraucher wichtig sind, und das Image des Produkts entsprechend ändern.

Statistiken und Diagramme. Iterationsverlauf, Stressmaße, Stresszerlegung, Koordinaten des gemeinsamen Raums, Objektdistanzen innerhalb der endgültigen Konfiguration, individuelle Raumgewichtungen, individuelle Räume, transformierte Ähnlichkeiten, transformierte unabhängige Variablen, Stressdiagramme, Streudiagramme des gemeinsamen Raums, Streudiagramme der individuellen Raumgewichtungen, Streudiagramme des individuellen Raums, Transformationsdiagramme, Shepard-Residuumdiagramme und Transformationsdiagramme für unabhängige Variablen.

Erläuterungen der Daten für die multidimensionale Skalierung

Daten. Daten können in Form von Distanzmatrizen oder Variablen, die in Distanzmatrizen umgewandelt werden, bereitgestellt werden. Die Matrizen können in einer Spalte oder über mehrere Spalten formatiert sein. Die Ähnlichkeiten können auf dem Skalierungsniveau der Verhältnisskala, Intervallskala, Ordinalskala oder Spline-Skala behandelt werden.

Annahmen. Es müssen mindestens drei Variablen angegeben werden. Die Anzahl der Dimensionen darf die Anzahl der Objekte minus 1 nicht übersteigen. Bei der Kombination mit mehreren Zufallsstarts wird die Reduktion der Dimensionalität ausgelassen. Wenn nur eine Quelle angegeben ist, sind alle Modelle äquivalent zum Identitätsmodell. In diesem Fall wird für die Analyse standardmäßig das Identitätsmodell verwendet.

Verwandte Prozeduren. Eine Skalierung aller Variablen auf numerischem Niveau entspricht der standardmäßigen multidimensionalen Skalierung.

So aktivieren Sie eine multidimensionale Skalierung:

1. Wählen Sie in den Menüs Folgendes aus:

Analysieren > Skala > Multidimensionale Skalierung (PROXSCAL)...

Daraufhin wird das Dialogfeld "Datenformat" geöffnet.

2. Geben Sie das Format der Daten an:

Datenformat. Geben Sie an, ob die Daten aus Ähnlichkeitsmaßen bestehen oder ob Sie aus den Daten Ähnlichkeitsmaße erstellen möchten.

Anzahl der Quellen. Wenn es sich bei den Daten um Ähnlichkeiten handelt, geben Sie an, ob eine Quelle oder mehrere Quellen für Ähnlichkeitsmaße vorhanden sind.

Eine Quelle. Wenn nur eine Quelle für Ähnlichkeiten vorhanden ist, geben Sie die Art der Formatierung des Datensets an: Die Ähnlichkeiten befinden sich in einer Matrix über mehrere Spalten verteilt oder in einer einzigen Spalte mit zwei verschiedenen Variablen zur Kennzeichnung der Zeile und Spalte jeder Ähnlichkeit.

- *Ähnlichkeiten stammen aus mehreren Spalten einer Matrix.* Die Ähnlichkeitsmatrix erstreckt sich über mehrere Spalten, deren Anzahl der Anzahl der Objekte entspricht. Hiermit öffnen Sie das Dialogfeld "Ähnlichkeiten in Matrizen über mehrere Spalten".
- *Ähnlichkeiten stammen aus einer einzigen Spalte.* Die Ähnlichkeitsmatrix befindet sich in einer einzigen Spalte (bzw. Variablen). Es sind zwei weitere Variablen erforderlich, welche die Zeile und Spalte jeder Zelle angeben. Hiermit öffnen Sie das Dialogfeld "Ähnlichkeiten in einer einzigen Spalte".

Mehrere Quellen. Wenn mehrere Quellen für Ähnlichkeiten vorhanden sind, geben Sie die Art der Formatierung des Datensets an: Die Ähnlichkeiten befinden sich über mehrere Spalten verteilt in gestapelten Matrizen, in mehreren Spalten mit einer Quelle pro Spalte oder in einer einzigen Spalte.

- *Ähnlichkeiten stammen aus mehreren Spalten in gestapelten Matrizen.* Die Ähnlichkeitsmatrizen erstrecken sich über mehrere Spalten, deren Anzahl der Anzahl der Objekte entspricht, und sind übereinandergestapelt in mehreren Zeilen, deren Anzahl der Anzahl der Objekte mal der Anzahl der Quellen entspricht. Hiermit öffnen Sie das Dialogfeld "Ähnlichkeiten in Matrizen über mehrere Spalten".
- *Ähnlichkeiten stammen aus Spalten (eine Quelle pro Spalte).* Die Ähnlichkeitsmatrizen sind in mehreren Spalten (bzw. Variablen) zusammengefasst. Es sind zwei weitere Variablen erforderlich, welche die Zeile und Spalte jeder Zelle angeben. Hiermit öffnen Sie das Dialogfeld "Ähnlichkeiten in Spalten".
- *Ähnlichkeiten sind in einer Spalte gestapelt.* Die Ähnlichkeitsmatrizen sind in einer einzigen Spalte (bzw. Variablen) zusammengefasst. Es sind drei weitere Variablen erforderlich, welche die Zeile, Spalte und Quelle jeder Zelle angeben. Hiermit öffnen Sie das Dialogfeld "Ähnlichkeiten in einer einzigen Spalte".

3. Klicken Sie auf **Definieren**.

Ähnlichkeiten in mehreren Spalten von Matrizen

Wenn Sie im Dialogfeld "Datenformat" das Datenmodell mit Ähnlichkeiten in Matrizen auswählen (aus einer oder mehreren Quellen), führen Sie die folgenden Schritte aus:

1. Wählen Sie mindestens drei Ähnlichkeitsvariablen aus. (Achten Sie darauf, dass die Reihenfolge der Variablen in der Liste der Reihenfolge der Spalten mit den Ähnlichkeiten entspricht.)
2. Wählen Sie optional GewichtungsvARIABLEN aus. Die Anzahl muss der Anzahl der ausgewählten Ähnlichkeitsvariablen entsprechen. (Achten Sie darauf, dass die Reihenfolge der Gewichtungen der Reihenfolge der damit gewichteten Ähnlichkeiten entspricht.)
3. Wenn mehrere Quellen vorhanden sind, wählen Sie optional eine Quellenvariable aus. (Die Anzahl der Fälle in jeder Ähnlichkeitsvariablen muss der Anzahl der Ähnlichkeitsvariablen multipliziert mit der Anzahl der Quellen entsprechen.)

Darüber hinaus können Sie ein Modell für die multidimensionale Skalierung angeben, Einschränkungen für den gemeinsamen Raum festlegen, Konvergenzkriterien festlegen, die zu verwendende Ausgangskonfiguration angeben sowie Diagramme und die Ausgabe auswählen.

Ähnlichkeiten in Spalten

Wenn Sie im Dialogfeld "Datenformat" das Modell mit mehreren Spalten für mehrere Quellen auswählen, führen Sie die folgenden Schritte aus:

1. Wählen Sie mindestens zwei Ähnlichkeitsvariablen aus. (Es wird davon ausgegangen, dass es sich bei den Variablen um Distanzmatrizen aus verschiedenen Quellen handelt.)
2. Wählen Sie eine Zeilenvariable zum Festlegen der Zeilenpositionen für die Ähnlichkeiten in jeder Ähnlichkeitsvariablen aus.

3. Wählen Sie eine Spaltenvariable zum Festlegen der Spaltenpositionen für die Ähnlichkeiten in jeder Ähnlichkeitsvariablen aus. (Zellen der Ähnlichkeitsmatrix, denen keine Zeilen-/Spaltenposition zugewiesen ist, werden als fehlend behandelt.)
4. Wählen Sie optional GewichtungsvARIABLEN aus. Die Anzahl muss der Anzahl der ausgewählten Ähnlichkeitsvariablen entsprechen.

Darüber hinaus können Sie ein Modell für die multidimensionale Skalierung angeben, Einschränkungen für den gemeinsamen Raum festlegen, Konvergenzkriterien festlegen, die zu verwendende Ausgangskonfiguration angeben sowie Diagramme und die Ausgabe auswählen.

Ähnlichkeiten in einer Spalte

Wenn Sie im Dialogfeld "Datenformat" das Ein-Spalten-Modell auswählen (entweder aus einer oder mehreren Quellen), führen Sie folgende Schritte aus:

1. Wählen Sie eine Ähnlichkeitsvariable aus. (Es wird davon ausgegangen, dass es sich dabei um eine oder mehrere Distanzmatrizen handelt.)
2. Wählen Sie eine Zeilenvariable zum Festlegen der Zeilenpositionen für die Ähnlichkeiten in der Ähnlichkeitsvariablen aus.
3. Wählen Sie eine Spaltenvariable zum Festlegen der Spaltenpositionen für die Ähnlichkeiten in der Ähnlichkeitsvariablen aus.
4. Wenn mehrere Quellen vorhanden sind, wählen Sie eine Quellenvariable aus. (Zellen der Ähnlichkeitsmatrix, denen keine Zeilen-/Spaltenposition zugewiesen ist, werden für jede Quelle als fehlend behandelt.)
5. Wählen Sie optional eine GewichtungsvARIABLEN aus.

Darüber hinaus können Sie ein Modell für die multidimensionale Skalierung angeben, Einschränkungen für den gemeinsamen Raum festlegen, Konvergenzkriterien festlegen, die zu verwendende Ausgangskonfiguration angeben sowie Diagramme und die Ausgabe auswählen.

Erstellen von Ähnlichkeiten aus Daten

Wenn Sie im Dialogfeld "Datenformat" die Option zum Erstellen von Ähnlichkeiten aus Daten auswählen, führen Sie die folgenden Schritte aus:

1. Wenn Sie Distanzen zwischen Variablen erzeugen (siehe Dialogfeld "Distanzen aus Daten erstellen"), wählen Sie mindestens drei Variablen aus. Diese Variablen dienen zum Erstellen der Ähnlichkeitsmatrix (oder Distanzmatrizen, wenn mehrere Quellen vorhanden sind). Bei der Erzeugung von Distanzen zwischen Fällen wird nur eine Variable benötigt.
2. Wenn mehrere Quellen vorhanden sind, wählen Sie eine Quellenvariable aus.
3. Wählen Sie bei Bedarf ein Maß für die Erstellung der Ähnlichkeiten aus.

Darüber hinaus können Sie ein Modell für die multidimensionale Skalierung angeben, Einschränkungen für den gemeinsamen Raum festlegen, Konvergenzkriterien festlegen, die zu verwendende Ausgangskonfiguration angeben sowie Diagramme und die Ausgabe auswählen.

Erstellen von Distanzen aus Daten

Die multidimensionale Skalierung verwendet Unähnlichkeitsdaten, um eine Skalierungslösung zu erstellen. Wenn Ihre Daten multivariate Daten darstellen (Werte gemessener Variablen), müssen Sie Unähnlichkeitsdaten erstellen, um eine multidimensionale Skalierungslösung berechnen zu können. Sie können Optionen für das Erstellen von Unähnlichkeitsmaßen aus Ihren Daten festlegen.

Maß. Hier können Sie das Unähnlichkeitsmaß für Ihre Analyse festlegen. Wählen Sie aus dem Gruppenfeld "Maß" die Option aus, die Ihrem Datentyp entspricht. Wählen Sie dann aus dem Dropdown-Listefeld ein Maß aus, das diesem Messwerttyp entspricht. Die folgenden Optionen sind verfügbar:

- **Intervall.** Euklidische Distanz, quadrierte euklidische Distanz, Tschebyscheff, Block, Minkowski oder ein benutzerdefiniertes Maß.
- **Häufigkeiten.** Chi-Quadrat-Maß oder Phi-Quadrat-Maß.
- **Binär.** Euklidische Distanz, quadrierte euklidische Distanz, Größendifferenz, Musterdifferenz, Varianz und Distanzmaß nach Lance und Williams.

Distanzmatrix erstellen. Mit dieser Funktion können Sie die Einheit der Analyse wählen. Zur Auswahl stehen "Zwischen den Variablen" oder "Zwischen den Fällen".

Werte transformieren. In bestimmten Fällen, zum Beispiel wenn die Variablen mit sehr unterschiedlichen Skalen gemessen werden, empfiehlt sich das Standardisieren der Werte vor dem Berechnen der Ähnlichkeiten (nicht auf binäre Daten anwendbar). Wählen Sie eine Standardisierungsmethode aus dem Dropdown-Listefeld "Standardisieren" aus. (Wenn keine Standardisierung erforderlich ist, wählen Sie **Keine** aus).

Definieren eines Modells für die multidimensionale Skalierung

Im Dialogfeld "Modell" können Sie ein Skalierungsmodell, die minimale und maximale Anzahl seiner Dimensionen, die Struktur der Ähnlichkeitsmatrix und die für die Ähnlichkeiten zu verwendende Transformation angeben sowie festlegen, ob Ähnlichkeiten für jede Quelle getrennt oder unabhängig von der Quelle transformiert werden.

Skalierungsmodell. Wählen Sie eine der folgenden Alternativen aus:

- **Identität.** Alle Quellen weisen dieselbe Konfiguration auf.
- **Euklidisch gewichtet.** Hierbei handelt es sich um ein Modell für individuelle Differenzen. Jede Quelle hat einen individuellen Raum, in dem jede Dimension des gemeinsamen Raums unterschiedlich gewichtet ist.
- **Euklidisch verallgemeinert.** Hierbei handelt es sich um ein Modell für individuelle Differenzen. Jede Quelle besitzt einen individuellen Wahrnehmungsraum, der einer Rotation des gemeinsamen Raums mit einer anschließenden unterschiedlichen Gewichtung der Dimensionen entspricht.
- **Reduzierter Rang.** Bei diesem Modell handelt es sich um ein verallgemeinertes euklidisches Modell, für das Sie den Rang des individuellen Raums angeben können. Der angegebene Rang muss größer oder gleich 1 und kleiner als die Höchstanzahl von Dimensionen sein.

Form. Geben Sie an, ob die Ähnlichkeiten der unteren oder oberen Dreiecksmatrix der Ähnlichkeitsmatrix entnommen werden sollen. Sie können angeben, dass die gesamte Matrix verwendet werden soll. In diesem Fall wird die gewichtete Summe der unteren und oberen Dreiecksmatrix analysiert. Obwohl nur die angegebenen Abschnitte verwendet werden, muss in jeden Fall die gesamte Matrix einschließlich der Diagonalen angegeben werden.

Ähnlichkeiten. Geben Sie an, ob die Ähnlichkeitsmatrix Ähnlichkeits- oder Unähnlichkeitsmaße enthält.

Ähnlichkeitstransformationen. Wählen Sie eine der folgenden Alternativen aus:

- **Verhältnis.** Die transformierten Ähnlichkeiten sind proportional zu den ursprünglichen Ähnlichkeiten. Diese Option ist nur für Ähnlichkeiten mit positiven Werten zulässig.
- **Intervall.** Die transformierten Ähnlichkeiten sind proportional zu den ursprünglichen Ähnlichkeiten plus einem konstanten Term. Durch den konstanten Term wird sichergestellt, dass alle Ähnlichkeiten positive Werte haben.

- **Ordinal.** Die transformierten Ähnlichkeiten weisen dieselbe Ordnung wie die ursprünglichen Ähnlichkeiten auf. Sie geben an, ob die Bindung von gebundenen Ähnlichkeiten erhalten bleiben muss oder aufgehoben werden darf.
- **Spline.** Die transformierten Ähnlichkeiten sind eine geglättete, nicht fallende, stückweise, polynomiale Transformation der ursprünglichen Ähnlichkeiten. Sie geben den Grad des Polynoms und die Anzahl der inneren Knoten an.

Transformationen zuweisen. Geben Sie an, ob nur Ähnlichkeiten innerhalb jeder Quelle miteinander verglichen werden oder ob die Vergleiche unabhängig von der Quelle sein sollen.

Dimensionen. Standardmäßig wird eine Lösung in zwei Dimensionen berechnet (Minimum = 2, Maximum = 2). Geben Sie für das Minimum und Maximum eine beliebige ganze Zahl zwischen 1 und der Anzahl der Objekte minus 1 an (solange das Minimum kleiner oder gleich dem Maximum ist). Die Prozedur berechnet eine Lösung in der höchsten Dimension und reduziert die Dimensionalität dann schrittweise, bis die niedrigste Dimensionalität erreicht ist.

Multidimensionale Skalierung: Einschränkungen

Im Dialogfeld "Einschränkungen" können Sie Einschränkungen für den gemeinsamen Raum festlegen.

Einschränkungen im gemeinsamen Raum. Geben Sie die gewünschte Art der Einschränkung an.

- **Keine Einschränkungen.** Es werden keine Einschränkungen für den gemeinsamen Raum festgelegt.
- **Einige Koordinaten sind fest.** Die erste ausgewählte Variable enthält die Koordinaten der Objekte in der ersten Dimension, die zweite Variable entspricht den Koordinaten der zweiten Dimension usw. Ein fehlender Wert zeigt an, dass eine Koordinate in einer Dimension frei ist. Die Anzahl der ausgewählten Variablen muss der angegebenen Höchstanzahl von Dimensionen entsprechen.
- **Lineare Kombination der unabhängigen Variablen.** Der gemeinsame Raum muss eine lineare Kombination der ausgewählten Variablen sein.

Einschränkungsvariablen. Wählen Sie die Variablen aus, die die Einschränkungen für den gemeinsamen Raum definieren. Wenn Sie eine lineare Kombination angegeben haben, geben Sie für die EinschränkungsvARIABLEN eine Intervall-, Nominal-, Ordinal- oder Spline-Transformation an. Die Anzahl der Fälle muss in jedem Fall der Anzahl der Objekte entsprechen.

Multidimensionale Skalierung: Optionen

Das Dialogfeld "Optionen" ermöglicht Ihnen die Auswahl der Anfangskonfiguration, die Angabe von Iterations- und Konvergenzkriterien und die Wahl zwischen standardmäßigen und gelockerten Aktualisierungen.

Ausgangskonfiguration. Wählen Sie eine der folgenden Optionen:

- **Simplex.** Objekte werden in der höchsten Dimension in gleichen Abständen voneinander platziert. Eine Iteration wird durchgeführt, um diese Konfiguration der höchsten Dimension zu optimieren, und anschließend wird eine Dimensionsreduktion vorgenommen, um eine Ausgangskonfiguration mit der maximalen, im Dialogfeld "Modell" angegebenen Anzahl von Dimensionen zu erhalten.
- **Torgerson.** Als Ausgangskonfiguration wird eine klassische Skalierungslösung verwendet.
- **Einzelner Zufallsstart.** Eine Konfiguration wird zufällig ausgewählt.
- **Mehrere Zufallsstarts.** Es werden mehrere Konfigurationen nach dem Zufallsprinzip ausgewählt. Die Konfiguration mit dem niedrigsten normalisierten Rohstress wird als Ausgangskonfiguration verwendet.
- **Benutzerdefiniert.** Wählen Sie Variablen aus, die die Koordinaten Ihrer eigenen Ausgangskonfiguration enthalten. Die Anzahl der ausgewählten Variablen muss der angegebenen Höchstanzahl von Dimensionen entsprechen.

nen entsprechen, wobei die erste Variable den Koordinaten von Dimension 1, die zweite Variable den Koordinaten von Dimension 2 entspricht usw. Die Anzahl der Fälle in jeder Variablen muss der Anzahl der Objekte entsprechen.

Iterationskriterien. Geben Sie Werte für Iterationskriterien an.

- **Stresskonvergenz.** Der Algorithmus führt keine weiteren Iterationen durch, wenn die Differenz von aufeinander folgenden normalisierten Rohstresswerten unter dem hier angegebenen Wert liegt, der zwischen 0,0 und 1,0 liegen muss.
- **Mindeststress.** Der Algorithmus wird beendet, wenn der normalisierte Rohstress unter den hier angegebenen Wert fällt, der zwischen 0,0 und 1,0 liegen muss.
- **Maximale Anzahl der Iterationen.** Der Algorithmus führt die angegebene Anzahl von Iterationen durch, falls nicht vorher eines der oben genannten Kriterien erfüllt wird.
- **Gelockerte Aktualisierungen.** Gelockerte Aktualisierungen beschleunigen den Algorithmus. Diese Aktualisierungen können mit keinem anderen Modell als dem Identitätsmodell und nicht zusammen mit Einschränkungen verwendet werden.

Multidimensionale Skalierung: Diagramme, Version 1

Im Dialogfeld "Diagramme" können Sie die zu erstellenden Diagramme festlegen. In diesem Abschnitt wird das Dialogfeld "Diagramme" beschrieben, wenn das Datenformat "Ähnlichkeiten in Spalten" ausgewählt wurde. Geben Sie für die Diagramme **Individuelle Raumgewichtungen**, **Ursprüngliche über transformierte Ähnlichkeiten** und **Transformierte Ähnlichkeiten über Distanzen** die Quellen an, für die die Diagramme erzeugt werden sollen. Die Liste der verfügbaren Quellen entspricht der Liste der Ähnlichkeitsvariablen im Hauptdialogfeld.

Stress. Ein Diagramm wird erstellt, in dem der normalisierte Rohstress über die Dimensionen abgebildet wird. Dieses Diagramm wird nur erstellt, wenn die Höchstanzahl von Dimensionen größer als die Mindestanzahl von Dimensionen ist.

Gemeinsamer Raum. Eine Streudiagrammmatrix der Koordinaten des gemeinsamen Raums wird angezeigt.

Individueller Raum. Für jede Quelle werden die Koordinaten der individuellen Räume in Streudiagrammmatrizen angezeigt. Dies ist nur möglich, wenn im Dialogfeld "Modell" eines der Modelle für individuelle Differenzen angegeben ist.

Individuelle Raumgewichtungen. Ein Streudiagramm der individuellen Raumgewichtungen wird erstellt. Dies ist nur möglich, wenn im Dialogfeld "Modell" eines der Modelle für individuelle Differenzen angegeben ist. Für das Modell "Euklidisch gewichtet" werden die Gewichtungen in Diagrammen mit einer Dimension pro Achse ausgegeben. Für das Modell "Euklidisch verallgemeinert" wird ein Diagramm pro Dimension erstellt, das Rotation und Gewichtung der jeweiligen Dimension anzeigt. Das für das Modell "Reduzierter Rang" erstellte Diagramm gleicht dem Diagramm für das Modell "Euklidisch verallgemeinert", die Anzahl der Dimensionen für die individuellen Räume ist jedoch reduziert.

Ursprüngliche über transformierte Ähnlichkeiten. Es werden Diagramme der ursprünglichen Ähnlichkeiten im Vergleich zu den transformierten Ähnlichkeiten erstellt.

Transformierte Ähnlichkeiten über Distanzen. Die transformierten Ähnlichkeiten werden über die Distanzen abgebildet.

Transformierte unabhängige Variablen. Es werden Transformationsdiagramme für die unabhängigen Variablen erstellt.

Variablen- und Dimensionskorrelationen. Es wird ein Diagramm der Korrelationen zwischen den unabhängigen Variablen und den Dimensionen des gemeinsamen Raums angezeigt.

Multidimensionale Skalierung: Diagramme, Version 2

Im Dialogfeld "Diagramme" können Sie die zu erstellenden Diagramme festlegen. In diesem Abschnitt wird das Dialogfeld "Diagramme" beschrieben, wenn ein beliebiges Datenformat außer dem Datenformat "Ähnlichkeiten in Spalten" ausgewählt wurde. Geben Sie für die Diagramme **Individuelle Raumgewichtungen, Ursprüngliche über transformierte Ähnlichkeiten** und **Transformierte Ähnlichkeiten über Distanzen** die Quellen an, für die die Diagramme erzeugt werden sollen. Die eingegebenen Quellennummern müssen Werte der im Hauptdialogfeld angegebenen Quellenvariable sein und im Bereich von 1 bis zur Anzahl der Quellen liegen.

Multidimensionale Skalierung: Ausgabe

Im Dialogfeld "Ausgabe" können Sie die angezeigte Ausgabe steuern und Auszüge davon in eigenen Dateien speichern.

Anzeigen. Wählen Sie mindestens eine der folgenden Anzeigeoptionen aus:

- **Gemeinsame Raumkoordinaten.** Zeigt die Koordinaten des gemeinsamen Raums an.
- **Individuelle Raumkoordinaten.** Zeigt nur dann die Koordinaten der individuellen Räume an, wenn das Modell kein Identitätsmodell ist.
- **Individuelle Raumgewichtungen.** Zeigt nur dann die individuellen Raumgewichtungen an, wenn eines der Modelle für individuelle Differenzen angegeben ist. Je nach Modell werden die Raumgewichtungen in Rotations- und Dimensionsgewichtungen zerlegt, die ebenfalls angezeigt werden.
- **Distanzen.** Zeigt die Distanzen zwischen den Objekten in der Konfiguration an.
- **Transformierte Ähnlichkeiten.** Zeigt die transformierten Ähnlichkeiten zwischen den Objekten in der Konfiguration an.
- **Eingabedaten.** Schließt die ursprünglichen Ähnlichkeiten und, sofern vorhanden, die Datengewichtungen, die Ausgangskonfiguration und die festen Koordinaten der unabhängigen Variablen ein.
- **Stress für Zufallsstarts.** Zeigt den Startwert für Zufallszahlen und den normalisierten Rohstresswert für jeden Zufallsstart an.
- **Iterationsverlauf.** Zeigt den Iterationsverlauf des Hauptalgorithmus an.
- **Multiple Stressmaße.** Zeigt verschiedene Stresswerte an. Die Tabelle enthält Werte für normalisierten Rohstress, Stress-I, Stress-II, S-Stress, erklärte Streuung und den Kongruenzkoeffizienten nach Tucker.
- **Stresszerlegung.** Zeigt eine Objekte- und Quellenzerlegung des endgültigen normalisierten Rohstress einschließlich des Durchschnitts pro Objekt und des Durchschnitts pro Quelle an.
- **Transformierte unabhängige Variablen.** Wenn eine Einschränkung der linearen Kombination ausgewählt wurde, werden die transformierten unabhängigen Variablen und die entsprechenden Regressionsgewichtungen angezeigt.
- **Variablen- und Dimensionskorrelationen.** Wenn eine Einschränkung der linearen Kombination ausgewählt wurde, werden die Korrelationen zwischen den unabhängigen Variablen und den Dimensionen des gemeinsamen Raums angezeigt.

In neuer Datei speichern. Sie können die Koordinaten des gemeinsamen Raums, die individuellen Raumgewichtungen, die Distanzen, die transformierten Ähnlichkeiten und die transformierten unabhängigen Variablen in eigenen IBM SPSS Statistics-Datendateien speichern.

Zusätzliche Funktionen beim Befehl PROXSCAL

Sie können die kategoriale Hauptkomponentenanalyse individuell anpassen, indem Sie die Auswahl in ein Syntaxfenster einfügen und die daraus resultierende Befehlssyntax für PROXSCAL bearbeiten. Die Befehlssyntax ermöglicht außerdem Folgendes:

- Geben Sie eigene Variablenlisten für Transformationen und Residuendiagramme an (mit dem Unterbefehl PLOT).

- Geben Sie eigene Quellenlisten für individuelle Raumgewichtungen, Transformationen und Residuen-
diagramme an (mit dem Unterbefehl PLOT).
- Geben Sie ein Subset der Transformationsdiagramme für unabhängige Variablen an, die angezeigt wer-
den sollen (mit dem Unterbefehl PLOT).

Vollständige Informationen zur Syntax finden Sie in der Befehlssyntaxreferenz.

Kapitel 8. Multidimensionale Entfaltung (PREFSCAL)

In der Prozedur "Multidimensionale Entfaltung" wird versucht, eine gemeinsame, quantitative Skala zu finden, mit der Sie die Beziehung zwischen zwei Sets von Objekten visuell untersuchen können.

Beispiele. Sie haben 21 Personen gebeten, 15 Frühstücksartikel in der Vorzugsreihenfolge von 1 bis 15 zu ordnen. Mit der Prozedur "Multidimensionale Entfaltung" können Sie ermitteln, dass die Personen die Frühstücksartikel in erster Linie auf zwei Arten unterscheiden: zwischen Weiß- und Vollkornbrot und zwischen kalorienhaltigen und -armen Artikeln.

In einer anderen Umfrage haben Sie eine Gruppe von Fahrern gebeten, 26 Fahrzeugmodelle anhand von 10 Attributen auf einer 6-Punkte-Skala von "1=trifft überhaupt nicht zu" bis "6=trifft vollständig zu" zu bewerten. Die Werte werden über die einzelnen Personen gemittelt und als Ähnlichkeiten verwendet. Mit der Prozedur "Multidimensionale Entfaltung" finden Sie Gruppen von ähnlichen Modellen sowie von Attributen, mit denen diese am ehesten in Verbindung gebracht werden.

Statistiken und Diagramme. Mit der Prozedur "Multidimensionale Entfaltung" können Sie einen Iterationsverlauf, Stressmaße, Stresszerlegung, Koordinaten des gemeinsamen Raums, Objektdistanzen innerhalb der endgültigen Konfiguration, individuelle Raumgewichtungen, individuelle Räume, transformierte Distanzen, Stressdiagramme, Streudiagramme des gemeinsamen Raums, Streudiagramme der individuellen Raumgewichtungen, Streudiagramme des individuellen Raums, Transformationsdiagramme und Shepard-Residuumdiagramme erstellen.

Erläuterungen der Daten für die multidimensionale Entfaltung

Daten. Die Daten werden in Form von rechtwinkligen Distanzmatrizen bereitgestellt. Jede Spalte wird als separates Spaltenobjekt betrachtet. Jede Zeile einer Ähnlichkeitsmatrix wird als separates Zeilenobjekt betrachtet. Wenn es mehrere Quellen für die Distanzen gibt, werden die Matrizen gestapelt.

Annahmen. Es müssen mindestens zwei Variablen angegeben werden. Die Anzahl der Dimensionen in der Lösung darf die Anzahl der Objekte minus 1 nicht übersteigen. Wenn nur eine Quelle angegeben ist, sind alle Modelle äquivalent zum Identitätsmodell. In diesem Fall wird für die Analyse standardmäßig das Identitätsmodell verwendet.

So aktivieren Sie eine multidimensionale Entfaltung:

1. Wählen Sie in den Menüs Folgendes aus:
Analysieren > Skala > Multidimensionale Entfaltung (PREFSCAL)...
2. Wählen Sie zwei oder mehrere Variablen aus, mit denen die Spalten in der rechtwinkligen Ähnlichkeitsmatrix identifiziert werden. Jede Variable stellt ein separates Spaltenobjekt dar.
3. Wählen Sie optional Gewichtungsvariablen aus. Die Anzahl der Gewichtungsvariablen muss der Anzahl der ausgewählten Spaltenobjektvariablen entsprechen. Die Reihenfolge der Gewichtungsvariablen muss mit der Reihenfolge der Spaltenobjekte übereinstimmen, die sie gewichten.
4. Wählen Sie optional eine Zeilenvariable aus. Die Werte (oder Wertbeschriftungen) dieser Variable werden verwendet, um die Zeilenobjekte in der Ausgabe zu beschriften.
5. Wenn mehrere Quellen vorhanden sind, wählen Sie bei Bedarf eine Quellenvariable aus. Die Anzahl der Fälle in der Datendatei muss der Anzahl der Zeilenobjekte multipliziert mit der Anzahl der Quellen entsprechen.

Darüber hinaus können Sie ein Modell für die multidimensionale Entfaltung angeben, Einschränkungen für den gemeinsamen Raum festlegen, Konvergenzkriterien festlegen, die zu verwendende Ausgangskonfiguration angeben sowie Diagramme und die Ausgabe auswählen.

Definieren eines Modells für die multidimensionale Entfaltung

Im Dialogfeld "Modell" können Sie ein Skalierungsmodell, die minimale und maximale Anzahl seiner Dimensionen, die Struktur der Ähnlichkeitsmatrix und die für die Ähnlichkeiten zu verwendende Transformation angeben sowie festlegen, ob Ähnlichkeiten abhängig von der Zeile, abhängig von der Quelle oder unabhängig von der Quelle transformiert werden.

Skalierungsmodell. Wählen Sie eine der folgenden Alternativen aus:

- **Identität.** Alle Quellen weisen dieselbe Konfiguration auf.
- **Euklidisch gewichtet.** Hierbei handelt es sich um ein Modell für individuelle Differenzen. Jede Quelle hat einen individuellen Raum, in dem jede Dimension des gemeinsamen Raums unterschiedlich gewichtet ist.
- **Euklidisch verallgemeinert.** Hierbei handelt es sich um ein Modell für individuelle Differenzen. Jede Quelle besitzt einen individuellen Wahrnehmungsraum, der einer Rotation des gemeinsamen Raums mit einer anschließenden unterschiedlichen Gewichtung der Dimensionen entspricht.

Ähnlichkeiten. Geben Sie an, ob die Ähnlichkeitsmatrix Ähnlichkeits- oder Unähnlichkeitsmaße enthält.

Dimensionen. Standardmäßig wird eine Lösung in zwei Dimensionen berechnet (Minimum = 2, Maximum = 2). Sie können für das Minimum und Maximum eine beliebige ganze Zahl zwischen 1 und der Anzahl der Objekte minus 1 angeben, solange das Minimum kleiner oder gleich dem Maximum ist. Die Prozedur berechnet eine Lösung in der höchsten Dimension und reduziert die Dimensionalität dann schrittweise, bis die niedrigste Dimensionalität erreicht ist.

Ähnlichkeitstransformationen. Wählen Sie eine der folgenden Alternativen aus:

- **Keine.** Die Ähnlichkeiten werden nicht transformiert. Sie können optional **Konstanten Term einschließen** auswählen. In diesem Fall werden die Ähnlichkeiten durch einen konstanten Term verschoben.
- **Linear.** Die transformierten Ähnlichkeiten sind proportional zu den ursprünglichen Ähnlichkeiten. Hierbei wird in der Transformationsfunktion eine Steigung geschätzt und der konstante Term wird auf 0 festgelegt. Dies wird auch als Verhältnistransformation bezeichnet. Sie können optional **Konstanten Term einschließen** auswählen. In diesem Fall werden die Ähnlichkeiten durch einen konstanten Term verschoben. Dies wird auch als Intervalltransformation bezeichnet.
- **Spline.** Die transformierten Ähnlichkeiten sind eine geglättete, nicht fallende, stückweise, polynomiale Transformation der ursprünglichen Ähnlichkeiten. Sie können den Grad des Polynoms und die Anzahl der inneren Knoten angeben. Sie können optional **Konstanten Term einschließen** auswählen. In diesem Fall werden die Ähnlichkeiten durch einen konstanten Term verschoben.
- **Glatt.** Die transformierten Ähnlichkeiten weisen dieselbe Ordnung wie die ursprünglichen Ähnlichkeiten auf. Es gilt jedoch eine Einschränkung, die die Unterschiede zwischen aufeinander folgenden Werten berücksichtigt. Das Ergebnis ist eine "glatte ordinale" Transformation. Sie können angeben, ob die Bindung von gebundenen Ähnlichkeiten erhalten bleiben muss oder aufgehoben werden darf.
- **Ordinal.** Die transformierten Ähnlichkeiten weisen dieselbe Ordnung wie die ursprünglichen Ähnlichkeiten auf. Sie können angeben, ob die Bindung von gebundenen Ähnlichkeiten erhalten bleiben muss oder aufgehoben werden darf.

Transformationen zuweisen. Geben Sie an, ob nur die Ähnlichkeiten innerhalb jeder Zeile oder nur die Ähnlichkeiten innerhalb jeder Quelle miteinander verglichen werden sollen oder ob die Vergleiche unabhängig von Zeile oder Quelle sind, also ob die Transformationen für einzelne Zeilen, für einzelne Quellen oder für alle Ähnlichkeiten gleichzeitig durchgeführt werden sollen.

Multidimensionale Entfaltung: Einschränkungen

Im Dialogfeld "Einschränkungen" können Sie Einschränkungen für den gemeinsamen Raum festlegen.

Einschränkungen im gemeinsamen Raum. Sie können die Koordinaten von Zeilen- und/oder Spaltenobjekten im gemeinsamen Raum festlegen.

Variablen für Zeilen-/Spalteneinschränkungen. Wählen Sie die Datei aus, die die Einschränkungen enthält, und wählen Sie die Variablen aus, die die Einschränkungen im gemeinsamen Raum definieren. Die erste ausgewählte Variable enthält die Koordinaten der Objekte in der ersten Dimension, die zweite Variable entspricht den Koordinaten der zweiten Dimension usw. Ein fehlender Wert zeigt an, dass eine Koordinate in einer Dimension frei ist. Die Anzahl der ausgewählten Variablen muss der angegebenen Höchstanzahl von Dimensionen entsprechen. Die Anzahl der Fälle jeder Variablen muss der Anzahl der Objekte entsprechen.

Multidimensionale Entfaltung: Optionen

Im Dialogfeld "Optionen" können Sie die Ausgangskonfiguration auswählen, Iterations- und Konvergenzkriterien angeben und den Penalisierungsterm für den Stress festlegen.

Ausgangskonfiguration. Wählen Sie eine der folgenden Optionen:

- **Klassisch.** Die rechtwinklige Ähnlichkeitsmatrix wird verwendet, um die Zwischenblöcke (Werte zwischen Zeilen und zwischen Spalten) der vollständig symmetrischen MDS-Matrix zu ergänzen. Sobald die vollständige Matrix gebildet ist, wird als Ausgangskonfiguration eine klassische Skalierungslösung verwendet. Die Zwischenblöcke können durch Ableiten von Werten mithilfe der Ungleichung im Dreieck oder der Spearman-Distanzen aufgefüllt werden.
- **Ross-Cliff.** Beim Ross-Cliff-Start werden die Ergebnisse einer Einzelwertzerlegung auf der zweifach zentrierten und quadratischen Ähnlichkeitsmatrix als Ausgangswerte für die Zeilen- und Spaltenobjekte verwendet.
- **Korrespondenz.** Beim Korrespondenzstart werden die Ergebnisse einer Korrespondenzanalyse der Daten in umgekehrter Reihenfolge (Ähnlichkeiten anstelle von Unähnlichkeiten) mit einer symmetrischen Normalisierung von Zeilen- und Spaltenscores verwendet.
- **Zentroide.** Die Prozedur beginnt mit der Positionierung von Zeilenobjekten in der Konfiguration auf der Grundlage einer Eigenwertzerlegung. Anschließend werden die Spaltenobjekte auf dem Zentroid der angegebenen Auswahl positioniert. Geben Sie für die Anzahl der Auswahl eine positive ganze Zahl zwischen 1 und der Anzahl der Ähnlichkeitsvariablen an.
- **Mehrere Zufallsstarts.** Die Lösungen werden für verschiedene zufällig ausgewählte Ausgangskonfigurationen berechnet. Die Lösung mit dem kleinsten penalisierten Stress wird als beste Lösung angezeigt.
- **Benutzerdefiniert.** Sie können Variablen auswählen, die die Koordinaten Ihrer eigenen Ausgangskonfiguration enthalten. Die Anzahl der ausgewählten Variablen muss der angegebenen Höchstanzahl von Dimensionen entsprechen, wobei die erste Variable den Koordinaten von Dimension 1, die zweite Variable den Koordinaten von Dimension 2 entspricht usw. Die Anzahl der Fälle in jeder Variablen muss der Gesamtzahl der Zeilen- und Spaltenobjekte entsprechen. Die Zeilen- und Spaltenkoordinaten müssen gestapelt sein, wobei die Spaltenkoordinaten den Zeilenkoordinaten folgen.

Iterationskriterien. Geben Sie Werte für Iterationskriterien an.

- **Stresskonvergenz.** Der Algorithmus beendet die Iterationen, wenn die relative Differenz von aufeinander folgenden penalisierten Stresswerten unter dem hier angegebenen Wert liegt. Dieser Wert darf nicht negativ sein.
- **Mindeststress.** Der Algorithmus wird beendet, wenn der penalisierte Stress unter den hier angegebenen Wert fällt. Dieser Wert darf nicht negativ sein.
- **Maximale Anzahl der Iterationen.** Der Algorithmus führt die hier angegebene Anzahl von Iterationen durch, falls nicht vorher eines der oben genannten Kriterien erfüllt wird.

Penalisierungsterm. Der Algorithmus versucht, den penalisierten Stress zu minimieren. Der penalisierte Stress ist ein Maß für die Güte der Anpassung, das dem Produkt des Kruskal-Stress-I und einem Penalisierungsterm auf der Basis des Variationskoeffizienten der transformierten Ähnlichkeiten entspricht. In diesen Feldern können Sie die Stärke und den Bereich des Penalisierungsterms festlegen.

- **Stärke.** Je kleiner der Wert des Stärkeparameters ist, desto stärker wird die Penalisierung. Geben Sie einen Wert zwischen 0,0 und 1,0 an.
- **Bereich.** Mit diesem Parameter legen Sie den Moment fest, ab dem die Penalisierung erfolgt. Bei 0,0 erfolgt keine Penalisierung. Höhere Wert führen dazu, dass der Algorithmus eine Lösung mit einer größeren Variation innerhalb der transformierten Ähnlichkeiten sucht. Geben Sie einen nicht negativen Wert an.

Multidimensionale Entfaltung: Diagramme

Im Dialogfeld "Diagramme" können Sie die zu erstellenden Diagramme festlegen.

Diagramme. Folgende Diagramme sind verfügbar:

- **Mehrere Starts.** Zeigt ein gestapeltes Histogramm des penalisierten Stress mit Stress und Penalisierung an.
- **Anfänglicher gemeinsamer Raum.** Zeigt eine Streudiagrammmatrix der Koordinaten des anfänglichen gemeinsamen Raums an.
- **Stress pro Dimension.** Erstellt ein Liniendiagramm des penalisierten Stress im Vergleich zur Dimensionalität. Dieses Diagramm wird nur erstellt, wenn die Höchstanzahl von Dimensionen größer als die Mindestanzahl von Dimensionen ist.
- **Endgültiger gemeinsamer Raum.** Eine Streudiagrammmatrix der Koordinaten des gemeinsamen Raums wird angezeigt.
- **Raumgewichtungen.** Ein Streudiagramm der individuellen Raumgewichtungen wird erstellt. Dies ist nur möglich, wenn im Dialogfeld "Modell" eines der Modelle für individuelle Differenzen angegeben ist. Für das Modell "Euklidisch gewichtet" werden die Gewichtungen für alle Quellen mit einer Dimension auf jeder Achse in einem Diagramm angezeigt. Für das Modell "Euklidisch verallgemeinert" wird ein Diagramm pro Dimension erstellt, das für jede Quelle die Rotation und Gewichtung der jeweiligen Dimension anzeigt.
- **Individueller Raum.** Eine Streudiagrammmatrix der Koordinaten des individuellen Raums jeder Quelle wird angezeigt. Dies ist nur möglich, wenn im Dialogfeld "Modell" eines der Modelle für individuelle Differenzen angegeben ist.
- **Transformationsdiagramme.** Es wird ein Streudiagramm der ursprünglichen Ähnlichkeiten im Vergleich zu den transformierten Ähnlichkeiten erstellt. Abhängig davon, wie Transformationen angewendet werden, wird jeder Zeile oder Quelle eine separate Farbe zugewiesen. Eine unabhängige Transformation erzeugt eine Farbe.
- **Shepard-Diagramme.** Die ursprünglichen Ähnlichkeiten im Vergleich zu sowohl transformierten Ähnlichkeiten als auch Distanzen. Die Distanzen werden durch Punkte angezeigt, die transformierten Ähnlichkeiten durch eine Linie. Abhängig davon, wie Transformationen angewendet werden, wird für jede Zeile oder Quelle eine separate Linie gezeichnet. Eine unabhängige Transformation erzeugt eine Linie.
- **Streudiagramm der Anpassung.** Es wird ein Streudiagramm der transformierten Ähnlichkeiten im Vergleich zu den Distanzen angezeigt. Wenn mehrere Quellen angegeben werden, erhält jede Quelle eine separate Farbe.
- **Residuendiagramme.** Es wird ein Streudiagramm der transformierten Ähnlichkeiten im Vergleich zu den Residuen (transformierte Ähnlichkeiten minus Distanzen) angezeigt. Wenn mehrere Quellen angegeben werden, erhält jede Quelle eine separate Farbe.

Stile für Zeilenobjekte. Hiermit können Sie umfassender festlegen, wie Zeilenobjekte in den Diagrammen dargestellt werden. Die Werte der optionalen Farbvariablen werden verwendet, um alle Farben zu durchlaufen. Die Werte der optionalen Markierungsvariablen werden verwendet, um alle möglichen Markierungen zu durchlaufen.

Diagramme der Quellen. Für **Individueller Raum**, **Streudiagramm der Anpassung** und **Residuendiagramme** (sowie, falls Transformationen nach Quellen angewendet werden, für **Transformationsdiagramme** und **Shepard-Diagramme**) können Sie angeben, für welche Quellen Diagramme erstellt werden sol-

len. Die eingegebenen Quellennummern müssen Werte der im Hauptdialogfeld angegebenen Quellenvariable sein und im Bereich von 1 bis zur Anzahl der Quellen liegen.

Zeilendiagramme. Wenn Transformationen für Zeilen erfolgen, können Sie für **Transformationsdiagramme** und **Shepard-Diagramme** die Zeile angeben, für die die Diagramme erstellt werden sollen. Die eingegebenen Zeilenanzahl muss im Bereich von 1 bis zur Anzahl der Zeilen liegen.

Multidimensionale Entfaltung: Ausgabe

Im Dialogfeld "Ausgabe" können Sie die angezeigte Ausgabe steuern und Auszüge davon in eigenen Dateien speichern.

Anzeigen. Wählen Sie mindestens eine der folgenden Anzeigeoptionen aus:

- **Eingabedaten.** Zeigt die ursprünglichen Ähnlichkeiten und, sofern vorhanden, die Datengewichtungen, die Ausgangskonfiguration und die festen Koordinaten an.
- **Mehrere Starts.** Zeigt den Startwert für Zufallszahlen und den penalisierten Stresswert für jeden Zufallsstart an.
- **Anfängliche Daten.** Zeigt die Koordinaten des anfänglichen gemeinsamen Raums an.
- **Iterationsverlauf.** Zeigt den Iterationsverlauf des Hauptalgorithmus an.
- **Anpassungsmaße.** Zeigt verschiedene Maße an. Die Tabelle enthält mehrere Maße für die Güte bzw. mangelhafte Güte der Anpassung, Korrelation, Variation und Nichtdegeneration.
- **Stresszerlegung.** Zeigt die Zerlegung des penalisierten Stress von Objekten, Zeilen und Quellen an, einschließlich der Mittelwerte und Standardabweichungen von Zeilen, Spalten und Quellen.
- **Transformierte Ähnlichkeiten.** Zeigt die transformierten Ähnlichkeiten an.
- **Endgültiger gemeinsamer Raum.** Zeigt die Koordinaten des gemeinsamen Raums an.
- **Raumgewichtungen.** Zeigt die individuellen Raumgewichtungen an. Diese Option ist nur verfügbar, wenn eines der Modelle für individuelle Differenzen angegeben wird. Je nach Modell werden die Raumgewichtungen in Rotations- und Dimensionsgewichtungen zerlegt, die ebenfalls angezeigt werden.
- **Individueller Raum.** Zeigt die Koordinaten der individuellen Räume an. Diese Option ist nur verfügbar, wenn eines der Modelle für individuelle Differenzen angegeben wird.
- **Angepasste Distanzen.** Zeigt die Distanzen zwischen den Objekten in der Konfiguration an.

In neuer Datei speichern. Sie können die Koordinaten des gemeinsamen Raums, die individuellen Raumgewichtungen, die Distanzen und die transformierten Ähnlichkeiten in eigenen IBM SPSS Statistics-Datendateien speichern.

Zusätzliche Funktionen beim Befehl PREFSCAL

Sie können die multidimensionale Entfaltung der Ähnlichkeiten individuell anpassen, indem Sie die Auswahl in ein Syntaxfenster einfügen und die daraus resultierende Befehlssyntax für PREFSCAL bearbeiten. Die Befehlssyntax ermöglicht außerdem Folgendes:

- Wenn mehrere Quelle verfügbar sind, können Sie für individuelle Räume, Streudiagramme der Anpassung und Residuendiagramme sowie bei Transformationen von Matrizen für Transformationsdiagramme und Shepard-Diagramme Listen mit mehreren Quellen angeben (mit dem Unterbefehl PLOT).
- Bei Transformationen von Zeilen können Sie für Transformationsdiagramme und Shepard-Diagramme Listen mit mehreren Zeilen angeben (mit dem Unterbefehl PLOT).
- Sie können anstelle einer Zeilen-ID-Variablen eine Anzahl von Zeilen angeben (mit dem Unterbefehl INPUT).
- Sie können anstelle einer Quellen-ID-Variablen eine Anzahl von Quellen angeben (mit dem Unterbefehl INPUT).

Vollständige Informationen zur Syntax finden Sie in der Befehlssyntaxreferenz.

Bemerkungen

Die vorliegenden Informationen wurden für Produkte und Services entwickelt, die auf dem deutschen Markt angeboten werden. IBM stellt dieses Material möglicherweise auch in anderen Sprachen zur Verfügung. Für den Zugriff auf das Material in einer anderen Sprache kann eine Kopie des Produkts oder der Produktversion in der jeweiligen Sprache erforderlich sein.

Möglicherweise bietet IBM die in dieser Dokumentation beschriebenen Produkte, Services oder Funktionen in anderen Ländern nicht an. Informationen über die gegenwärtig im jeweiligen Land verfügbaren Produkte und Services sind beim zuständigen IBM Ansprechpartner erhältlich. Hinweise auf IBM Lizenzprogramme oder andere IBM Produkte bedeuten nicht, dass nur Programme, Produkte oder Services von IBM verwendet werden können. Anstelle der IBM Produkte, Programme oder Services können auch andere, ihnen äquivalente Produkte, Programme oder Services verwendet werden, solange diese keine gewerblichen oder anderen Schutzrechte von IBM verletzen. Die Verantwortung für den Betrieb von Produkten, Programmen und Services anderer Anbieter liegt beim Kunden.

Für in diesem Handbuch beschriebene Erzeugnisse und Verfahren kann es IBM Patente oder Patentanmeldungen geben. Mit der Auslieferung dieses Handbuchs ist keine Lizenzierung dieser Patente verbunden. Lizenzanforderungen sind schriftlich an folgende Adresse zu richten (Anfragen an diese Adresse müssen auf Englisch formuliert werden):

*IBM Director of Licensing
IBM Europe, Middle East & Africa
Tour Descartes
2, avenue Gambetta
92066 Paris La Defense
France*

Trotz sorgfältiger Bearbeitung können technische Ungenauigkeiten oder Druckfehler in dieser Veröffentlichung nicht ausgeschlossen werden. Die hier enthaltenen Informationen werden in regelmäßigen Zeitabständen aktualisiert und als Neuausgabe veröffentlicht. IBM kann ohne weitere Mitteilung jederzeit Verbesserungen und/oder Änderungen an den in dieser Veröffentlichung beschriebenen Produkten und/oder Programmen vornehmen.

Verweise in diesen Informationen auf Websites anderer Anbieter werden lediglich als Service für den Kunden bereitgestellt und stellen keinerlei Billigung des Inhalts dieser Websites dar. Das über diese Websites verfügbare Material ist nicht Bestandteil des Materials für dieses IBM Produkt. Die Verwendung dieser Websites geschieht auf eigene Verantwortung.

Werden an IBM Informationen eingesandt, können diese beliebig verwendet werden, ohne dass eine Verpflichtung gegenüber dem Einsender entsteht.

Lizenznehmer des Programms, die Informationen zu diesem Produkt wünschen mit der Zielsetzung: (i) den Austausch von Informationen zwischen unabhängig voneinander erstellten Programmen und anderen Programmen (einschließlich des vorliegenden Programms) sowie (ii) die gemeinsame Nutzung der ausgetauschten Informationen zu ermöglichen, wenden sich an folgende Adresse:

*IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
USA*

Die Bereitstellung dieser Informationen kann unter Umständen von bestimmten Bedingungen - in einigen Fällen auch von der Zahlung einer Gebühr - abhängig sein.

Die Lieferung des in diesem Dokument beschriebenen Lizenzprogramms sowie des zugehörigen Lizenzmaterials erfolgt auf der Basis der IBM Rahmenvereinbarung bzw. der Allgemeinen Geschäftsbedingungen von IBM, der IBM Internationalen Nutzungsbedingungen für Programmpakete oder einer äquivalenten Vereinbarung.

Die angeführten Leistungsdaten und Kundenbeispiele dienen nur zur Illustration. Die tatsächlichen Ergebnisse beim Leistungsverhalten sind abhängig von der jeweiligen Konfiguration und den Betriebsbedingungen.

Alle Informationen zu Produkten anderer Anbieter stammen von den Anbietern der aufgeführten Produkte, deren veröffentlichten Ankündigungen oder anderen allgemein verfügbaren Quellen. IBM hat diese Produkte nicht getestet und kann daher keine Aussagen zu Leistung, Kompatibilität oder anderen Merkmalen machen. Fragen zu den Leistungsmerkmalen von Produkten anderer Anbieter sind an den jeweiligen Anbieter zu richten.

Aussagen über Pläne und Absichten von IBM unterliegen Änderungen oder können zurückgenommen werden und repräsentieren nur die Ziele von IBM.

Diese Veröffentlichung enthält Beispiele für Daten und Berichte des alltäglichen Geschäftsablaufs. Sie sollen nur die Funktionen des Lizenzprogramms illustrieren und können Namen von Personen, Firmen, Marken oder Produkten enthalten. Alle diese Namen sind frei erfunden; Ähnlichkeiten mit tatsächlichen Namen und Adressen sind rein zufällig.

COPYRIGHTLIZENZ:

Diese Veröffentlichung enthält Beispielanwendungsprogramme, die in Quellsprache geschrieben sind und Programmier Techniken in verschiedenen Betriebsumgebungen veranschaulichen. Sie dürfen diese Beispielprogramme kostenlos kopieren, ändern und verteilen, wenn dies zu dem Zweck geschieht, Anwendungsprogramme zu entwickeln, zu verwenden, zu vermarkten oder zu verteilen, die mit der Anwendungsprogrammierschnittstelle für die Betriebsumgebung konform sind, für die diese Beispielprogramme geschrieben werden. Diese Beispiele wurden nicht unter allen denkbaren Bedingungen getestet. Daher kann IBM die Zuverlässigkeit, Wartungsfreundlichkeit oder Funktion dieser Programme weder zusagen noch gewährleisten. Die Beispielprogramme werden ohne Wartung (auf "as-is"-Basis) und ohne jegliche Gewährleistung zur Verfügung gestellt. IBM übernimmt keine Haftung für Schäden, die durch die Verwendung der Beispielprogramme entstehen.

Kopien oder Teile der Beispielprogramme bzw. daraus abgeleiteter Code müssen folgenden Copyrightvermerk beinhalten:

© (Name Ihrer Firma) (Jahr). Teile des vorliegenden Codes wurden aus Beispielprogrammen der IBM Corporation abgeleitet.

© Copyright IBM Corp. _Jahr/Jahre angeben_. Alle Rechte vorbehalten.

Marken

IBM, das IBM Logo und ibm.com sind Marken oder eingetragene Marken der IBM Corp in den USA und/oder anderen Ländern. Weitere Produkt- und Servicennamen können Marken von IBM oder anderen Unternehmen sein. Eine aktuelle Liste der IBM Marken finden Sie auf der Webseite "Copyright and trademark information" unter www.ibm.com/legal/copytrade.shtml.

Adobe, das Adobe-Logo, PostScript und das PostScript-Logo sind Marken oder eingetragene Marken der Adobe Systems Incorporated in den USA und/oder anderen Ländern.

Intel, das Intel-Logo, Intel Inside, das Intel Inside-Logo, Intel Centrino, das Intel Centrino-Logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium und Pentium sind Marken oder eingetragene Marken der Intel Corporation oder ihrer Tochtergesellschaften in den USA oder anderen Ländern.

Linux ist eine eingetragene Marke von Linus Torvalds in den USA und/oder anderen Ländern.

Microsoft, Windows, Windows NT und das Windows-Logo sind Marken der Microsoft Corporation in den USA und/oder anderen Ländern.

UNIX ist eine eingetragene Marke von The Open Group in den USA und anderen Ländern.

Java und alle auf Java basierenden Marken und Logos sind Marken oder eingetragene Marken der Oracle Corporation und/oder ihrer verbundenen Unternehmen.

Index

A

- Ähnlichkeitstransformationen
 - in "Multidimensionale Entfaltung" 52
- Allgemeines euklidisches Modell
 - in "Multidimensionale Entfaltung" 52
- Anfängliche Diagramme des gemeinsamen Raums
 - in "Multidimensionale Entfaltung" 54
- ANOVA
 - in "Kategoriale Regression" 17
- Anpassung
 - in "Nicht lineare kanonische Korrelationsanalyse" 29
- Ausgangskonfiguration
 - in "Kategoriale Regression" 15
 - in "Multidimensionale Entfaltung" 53
 - in "Multidimensionale Skalierung" 47
 - in "Nicht lineare kanonische Korrelationsanalyse" 29

B

- Biplots
 - in der kategorialen Hauptkomponentenanalyse 24
 - in der Korrespondenzanalyse 34
 - in der Mehrfachkorrespondenzanalyse 41
- Bootstrapping
 - Kategoriale Hauptkomponentenanalyse 26

D

- Deskriptive Statistiken
 - in "Kategoriale Regression" 17
- Diagramme des gemeinsamen Raums
 - in "Multidimensionale Entfaltung" 54
 - in "Multidimensionale Skalierung" 48
- Diagramme des individuellen Raums
 - in "Multidimensionale Entfaltung" 54
 - in "Multidimensionale Skalierung" 48
- Diagramme für mehrere Starts
 - in "Multidimensionale Entfaltung" 54
- Diagramme für Raumgewichtungen
 - in "Multidimensionale Entfaltung" 54
- Diagramme mit individuellen Raumgewichtungen
 - in "Multidimensionale Entfaltung" 54
 - in "Multidimensionale Skalierung" 48
- Diagramme mit projizierten Zentroiden
 - in der kategorialen Hauptkomponentenanalyse 25
- Dimensionen
 - in der Korrespondenzanalyse 33
- Diskretisierung
 - in "Kategoriale Regression" 14
 - in der kategorialen Hauptkomponentenanalyse 21

- Diskretisierung (*Forts.*)
 - in der Mehrfachkorrespondenzanalyse 38
- Diskriminationsmaßdiagramme
 - in der Mehrfachkorrespondenzanalyse 41
- Diskriminationsmaße
 - in der Mehrfachkorrespondenzanalyse 40
- Distanzen
 - in "Multidimensionale Entfaltung" 55
 - in "Multidimensionale Skalierung" 49
- Distanzmaße
 - in der Korrespondenzanalyse 33

E

- Einschränkungen
 - in "Multidimensionale Skalierung" 47
- Einschränkungen im gemeinsamen Raum.
 - in "Multidimensionale Entfaltung" 52
- Elastic Net
 - in "Kategoriale Regression" 16
- Endgültige Diagramme des gemeinsamen Raums
 - in "Multidimensionale Entfaltung" 54

F

- Fehlende Werte
 - in "Kategoriale Regression" 15
 - in der kategorialen Hauptkomponentenanalyse 21
 - in der Mehrfachkorrespondenzanalyse 38

G

- Gelockerte Aktualisierungen
 - in "Multidimensionale Skalierung" 47
- Gemeinsame Raumkoordinaten
 - in "Multidimensionale Entfaltung" 55
 - in "Multidimensionale Skalierung" 49
- Gewichtetes euklidisches Modell
 - in "Multidimensionale Entfaltung" 52
- Gewichtungen
 - in "Nicht lineare kanonische Korrelationsanalyse" 29
- Grafik
 - in "Kategoriale Regression" 18
 - in "Multidimensionale Skalierung" 48, 49
 - in "Nicht lineare kanonische Korrelationsanalyse" 29
 - in der Korrespondenzanalyse 34

I

- Identitätsmodell
 - in "Multidimensionale Entfaltung" 52
- Individuelle Raumgewichtungen
 - in "Multidimensionale Entfaltung" 55
 - in "Multidimensionale Skalierung" 49
- Individuelle Raumkoordinaten
 - in "Multidimensionale Entfaltung" 55
- Iterationskriterien
 - in "Multidimensionale Entfaltung" 53
 - in "Multidimensionale Skalierung" 47
- Iterationsverlauf
 - in "Multidimensionale Entfaltung" 55
 - in "Multidimensionale Skalierung" 49
 - in der kategorialen Hauptkomponentenanalyse 23
 - in der Mehrfachkorrespondenzanalyse 40

K

- Kategoriale Hauptkomponentenanalyse 19, 22
 - Bootstrapping 26
 - Messniveau der optimalen Skalierung 20
 - Variablen speichern 24
 - zusätzliche Funktionen beim Befehl 26
- Kategoriale Regression 13
 - Grafik 13
 - Messniveau der optimalen Skalierung 14
 - Regularisierung 16
 - Speichern 17
 - Statistik 13
 - zusätzliche Funktionen beim Befehl 18
- Kategoriediagramme
 - in der kategorialen Hauptkomponentenanalyse 25
 - in der Mehrfachkorrespondenzanalyse 41
- Kategoriequantifizierungen
 - in "Kategoriale Regression" 17
 - in "Nicht lineare kanonische Korrelationsanalyse" 29
 - in der kategorialen Hauptkomponentenanalyse 23
 - in der Mehrfachkorrespondenzanalyse 40
- Komponentenladungen
 - in "Nicht lineare kanonische Korrelationsanalyse" 29
 - in der kategorialen Hauptkomponentenanalyse 23
- Komponentenladungsdigramme
 - in der kategorialen Hauptkomponentenanalyse 25

Konfidenzstatistik
in der Korrespondenzanalyse 34
Korrelationen
in "Multidimensionale Skalierung" 49
Korrelationsdiagramme
in "Multidimensionale Skalierung" 48
Korrelationsmatrix
in der kategorialen Hauptkomponentenanalyse 23
in der Mehrfachkorrespondenzanalyse 40
Korrespondenzanalyse 31, 32, 33, 34
Grafik 31
Statistik 31
zusätzliche Funktionen beim Befehl 35

L

Lasso
in "Kategoriale Regression" 16

M

Mehrfachkorrespondenzanalyse 37, 39
Messniveau der optimalen Skalierung 38
Variablen speichern 41
zusätzliche Funktionen beim Befehl 41
Messniveau der optimalen Skalierung
in der kategorialen Hauptkomponentenanalyse 20
in der Mehrfachkorrespondenzanalyse 38
Multidimensionale Entfaltung 51
Ausgabe 55
Einschränkungen im gemeinsamen Raum. 52
Grafik 51, 54
Modell 52
Optionen 53
Statistik 51
zusätzliche Funktionen beim Befehl 55
Multidimensionale Skalierung 43, 44, 45
Ausgabe 49
Einschränkungen 47
Grafik 43, 48, 49
Modell 46
Optionen 47
Statistik 43
zusätzliche Funktionen beim Befehl 49
Multiples R
in "Kategoriale Regression" 17

N

Nicht lineare kanonische Korrelationsanalyse 27, 28, 29
Grafik 27
Statistik 27
zusätzliche Funktionen beim Befehl 30

Normalisierung
in der Korrespondenzanalyse 33

O

Objektpunktdiagramme
in der kategorialen Hauptkomponentenanalyse 24
in der Mehrfachkorrespondenzanalyse 41
Objektscores
in "Nicht lineare kanonische Korrelationsanalyse" 29
in der kategorialen Hauptkomponentenanalyse 23
in der Mehrfachkorrespondenzanalyse 40

P

Penalisierungsterm
in "Multidimensionale Entfaltung" 53
PREFSCAL 51

R

Regressionskoeffizienten
in "Kategoriale Regression" 17
Residuumdiagramme
in "Multidimensionale Entfaltung" 54
Ridge-Regression
in "Kategoriale Regression" 16

S

Shepard-Diagramme
in "Multidimensionale Entfaltung" 54
Skalierungsmodell
in "Multidimensionale Entfaltung" 52
Standardisierung
in der Korrespondenzanalyse 33
Stressdiagramme
in "Multidimensionale Entfaltung" 54
in "Multidimensionale Skalierung" 48
Stressmaße
in "Multidimensionale Entfaltung" 55
in "Multidimensionale Skalierung" 49
Streudiagramm der Anpassung
in "Multidimensionale Entfaltung" 54

T

Trägheit
in der Korrespondenzanalyse 34
Transformationsdiagramme
in "Multidimensionale Entfaltung" 54
in "Multidimensionale Skalierung" 48
in der kategorialen Hauptkomponentenanalyse 25
in der Mehrfachkorrespondenzanalyse 41
Transformierte Ähnlichkeiten
in "Multidimensionale Entfaltung" 55
in "Multidimensionale Skalierung" 49

Transformierte unabhängige Variablen
in "Multidimensionale Skalierung" 49
Triplots
in der kategorialen Hauptkomponentenanalyse 24

V

VariablenGewichtung
in der kategorialen Hauptkomponentenanalyse 20
in der Mehrfachkorrespondenzanalyse 38
Varianz berücksichtigt für
in der kategorialen Hauptkomponentenanalyse 23
Verbundene Kategoriediagramme
in der kategorialen Hauptkomponentenanalyse 25
in der Mehrfachkorrespondenzanalyse 41

Z

Zentroide
in "Nicht lineare kanonische Korrelationsanalyse" 29
Zusätzliche Fälle
in "Kategoriale Regression" 15

