

IBM SPSS Categories 19

Jacqueline J. Meulman

Willem J. Heiser

SPSS Inc.



Note: Before using this information and the product it supports, read the general information under Notices p. 333 .

This document contains proprietary information of SPSS Inc, an IBM Company. It is provided under a license agreement and is protected by copyright law. The information contained in this publication does not include any product warranties, and any statements provided in this manual should not be interpreted as such.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

© Copyright SPSS Inc. 1989, 2010.

はじめに

IBM® SPSS® Statistics は、データ分析の包括的システムです。カテゴリは、このマニュアルで説明されている追加の分析手法を提供するオプションのアドオン モジュールです。カテゴリ アドオン モジュールは SPSS Statistics Core システムと組み合わせて使用し、Core システムに完全に統合されます。

SPSS Inc., an IBM Company について

SPSS Inc., an IBM Company は、余剰分析ソフトウェアおよびソリューションの世界的なリーディングカンパニーです。当社のデータ収集、統計、モデリング、展開という製品の包括的なポートフォリオによりお客様の考えや意見を収集、見込み客との対話の結果を予測、分析を業務プロセスに組み込むことによりこれらの見解に判断を下すことができます。SPSS Inc. のソリューションにより、分析、IT アーキテクチャ、業務プロセスの収束に焦点を当て、組織全体の相互接続した経営目標に取り組みます。世界中の民間、政府、学術分野のお客様が SPSS Inc. のテクノロジーを包括的に利用しています。お客様の関心を呼び、拡大する一方、不正やリスクを軽減、緩和します。2009 年 10 月、SPSS Inc. は IBM 社に買収されました。詳細は <http://www.spss.com> をご覧ください。

テクニカル サポート

テクニカル サポートのサービスをご利用いただけます。SPSS Inc. 製品の使用方法や、対応しているハードウェア環境へのインストールに関して問い合わせることもできます。テクニカル サポートにご連絡するには、<http://support.spss.com> の SPSS Inc. Web サイトを参照いただくか、<http://support.spss.com/default.asp?refpage=contactus.asp> の Web サイトでお近くの営業所にお問い合わせください。連絡の際は、所属団体名、サポート契約などを確認できるよう、あらかじめ手元にご用意ください。

カスタマ サービス

製品の発送やお支払いに関してご質問がある場合は、SPSS 社までお問い合わせください (SPSS Japan のホームページは <http://www.spss.co.jp> です)。お問い合わせの際には、シリアル番号をご用意ください。

トレーニング セミナー

SPSS Inc. では一般公開およびオンサイトでトレーニング セミナーを実施しています。セミナーでは実践的な講習を行います。セミナーは主要都市で定期的開催されます。セミナーの詳細については、SPSS 社までお問い合わせください (SPSS Japan のホームページは <http://www.spss.co.jp> です)。

追加の出版物

Marija Noruš による『SPSS Statistics: Guide to Data Analysis』、『SPSS Statistics: Statistical Procedures Companion』、『SPSS Statistics: Advanced Statistical Procedures Companion』が Prentice Hall から出版されました。補助的な資料としてご利用いただけます。これらの出版物には、SPSS Statistics Base モジュール、Advanced Statistics モジュール、Regression モジュールの統計的手続きについて記載されています。初めてデータ分析を行う場合、高度なアプリケーションを使用する場合に応じて、この本は IBM® SPSS® Statistics が提供している機能を効率よく使用するための手助けとなります。出版物の内容、サンプルの図表などの詳細は、作者の Web サイトを参照してください。 <http://www.norusis.com>

通知

最適尺度法手続きとその IBM® SPSS® Statistics への実装は、ライデン大学社会行動学部教育心理学科のメンバーで構成される、Data Theory Scaling System Group (DTSS) によって開発されました。

Willem Heiser、Jacqueline Meulman、Gerda van den Berg、および Patrick Groenen の各氏は、手続きの原型 (1990 年) の作成に参加しました。Jacqueline Meulman および Peter Neufeglise の各氏は、カテゴリ回帰分析、コレスポンデンス分析、カテゴリ主成分分析、および多次元尺度法のための手続きの開発に参加しました。Anita van der Kooij 氏は、CATREG、CORRESPONDENCE、および CATPCA 手続きについて多大な貢献をして頂きました。Willem Heiser、Jacques Commandeur、Frank Busing、Gerda van den Berg、および Patrick Groenen の各氏は、PROXSCAL 手続きの開発に参加しました。Frank Busing、Willem Heiser、Patrick Groenen、および Peter Neufeglise の各氏は、PREFSCAL 手続きの開発に参加しました。

パート I: ユーザー ガイド

1	カテゴリ データ用の最適尺度法手続きの概要	1
	最適尺度法とは	1
	最適尺度法を使用する理由	1
	最適尺度水準と尺度	2
	最適尺度水準の選択	3
	変換プロット	3
	カテゴリ コード	4
	アプリケーションに最適な手続きの選択	7
	カテゴリ回帰	8
	カテゴリ主成分分析	8
	非線型正準相関分析	9
	対応分析	10
	多重応答分析	12
	多次元尺度法	13
	多次元展開	13
	最適尺度図表の縦横比	14
	推奨参考文献	14
2	カテゴリ回帰分析 (CATREG)	16
	カテゴリ回帰分析の尺度の定義	18
	カテゴリ回帰分析: 離散化	19
	カテゴリ回帰: 欠損値	20
	カテゴリ回帰: オプション	21
	[カテゴリ回帰正規化]	23
	カテゴリ回帰: 出力	24
	カテゴリ回帰: 保存	26
	カテゴリ回帰: 変換プロット	27
	CATREG コマンドの追加機能	28

3 カテゴリ主成分分析 (CATPCA) 29

CATPCA の尺度と重み付けの定義	31
カテゴリ主成分分析: 離散化	33
カテゴリ主成分分析: 欠損値	34
カテゴリ主成分分析: オプション	36
カテゴリ主成分分析: 出力	39
カテゴリ主成分分析: 保存	40
カテゴリ主成分分析: オブジェクト プロット	41
カテゴリ主成分分析: カテゴリ プロット	43
カテゴリ主成分分析: 負荷プロット	44
CATPCA コマンドの追加機能	45

4 非線型正準相関分析 (OVERALS) 46

範囲と尺度の定義 (Categories)	49
範囲の定義 (プロビット分析/MANOVA)	50
非線型正準相関分析: オプション	50
OVERALS コマンドの追加機能	52

5 コレスポネンス分析 54

コレスポネンス分析: 行範囲の定義	56
コレスポネンス分析: 列範囲の定義	57
コレスポネンス分析: モデル	57
コレスポネンス分析: 統計	59
コレスポネンス分析: 作図	61
CORRESPONDENCE コマンドの追加機能	62

6 多重応答分析 63

多重応答分析での変数の重みの定義	65
多重応答分析での離散化	66
多重応答分析での欠損値	67

多重応答分析のオプション	68
多重応答分析での出力	71
多重応答分析での保存	72
多重応答分析でのオブジェクトプロット	73
多重応答分析での変数プロット	74
MULTIPLE CORRESPONDENCE コマンドの追加機能	76

7 多次元尺度法 (PROXSCAL) 77

複数列行列の近接	79
複数列の近接	80
単一系列の近接	81
データから近接の作成	81
データからの尺度の作成	83
多次元尺度法のモデルの定義	84
多次元尺度法の制限	85
多次元尺度法のオプション	86
多次元尺度法の作図、バージョン 1	88
多次元尺度法の作図、バージョン 2	89
多次元尺度法の出力	90
PROXSCAL コマンドの追加機能	92

8 多次元展開 (PREFSCAL) 93

多次元展開のモデルの定義	94
多次元展開の制限	96
多次元展開のオプション	98
多次元展開の作図	99
多次元展開の出力	101
PREFSCAL コマンドの追加機能	103

パート II: 例

9 カテゴリ回帰 105

例:カーペット専用洗剤のデータ	105
標準線型回帰分析	106
カテゴリ回帰分析	112
例:オゾンのデータ	126
変数の離散化	127
変換の種類を選択	127
数量化の最適性	141
変換の効果	143
推奨参考文献	153

10 カテゴリ主成分分析 155

例:社会システムの相互関係の調査	155
分析の実行	156
次元数	162
数量化 (カテゴリ / カテゴリの回帰)	163
オブジェクト スコア	165
成分負荷	166
追加の次元	168
例:摂食障害の症状	170
分析の実行	171
変換プロット	184
モデルの要約	187
成分負荷	188
オブジェクト スコア	189
病気の経過の構成の調査	191
推奨参考文献	206

11 非線型正準相関分析 209

例: 調査結果の分析	209
データの調査	210
グループ間の類似度の説明	217

成分負荷	220
変換プロット	221
単一カテゴリ座標と多重カテゴリ座標	223
重心と射影された重心	224
代替分析	227
一般的な提言	233
推奨参考文献	234

12 コレスポネンス分析 235

対応分析	235
正規化	236
例: コーヒー ブランドの印象	237
分析の実行	238
次元数	242
寄与率 (コレスポネンス分析)	243
作図	244
対称の正規化	246
推奨参考文献	247

13 多重応答分析 248

例: ハードウェアの特性	248
分析の実行	249
モデルの要約	252
オブジェクト スコア	253
判別測定	254
カテゴリ数量化	256
オブジェクト スコアの詳細な調査	257
外れ値の除外	260
推奨参考文献	264

14 多次元尺度法 266

例: 親族関係用語の調査	266
次元数の選択	267
3次元の解	274

デフォルト以外の変換による 3 次元の解	281
考察	284
推奨参考文献	284

15 多次元展開 286

例: 朝食メニューの嗜好	286
縮退解の生成	286
測定方法 (GLM 反復測定)	289
共通空間	290
非縮退分析の実行	291
測定方法 (GLM 反復測定)	292
共通空間	293
例: 朝食メニューの嗜好の三次展開	294
分析の実行	294
測定方法 (GLM 反復測定)	298
共通空間	299
個別空間	300
異なる初期布置の使い方	302
測定方法 (GLM 反復測定)	304
共通空間	305
個別空間	306
例: 「行動 - 状況」の適切性の調査	307
分析の実行	307
測定方法 (GLM 反復測定)	313
共通空間	314
近接変換	315
近接変換の変更 (順序)	315
測定方法 (GLM 反復測定)	317
共通空間	318
近接変換	319
推奨参考文献	319

付録

A サンプル ファイル	321
B Notices	333
参考文献	336
索引	343

パート I: ユーザー ガイド

カテゴリ データ用の最適尺度法 手続きの概要

カテゴリ手続きでは、標準統計手続きで分析が困難または不可能なデータを最適尺度法を使用して分析します。この章では、各手続きの実行内容、各手続きが最も適している状況、手続き同士の関係、手続きとそれに対応する標準統計手続きとの関係について説明します。

注:これらの手続きと IBM® SPSS® Statistics での実装はライデン大学の社会行動科学部の教育心理学科のメンバーで構成される Data Theory Scaling System Group (DTSS) によって開発されました。

最適尺度法とは

最適尺度法の背景には、各変数のカテゴリに数量化した数値を割り当てるという考え方があるので、標準手続きを使用することで、数量化された変数について解を求めることができます。

使用している手続きの最適化基準に基づいて、各変数のカテゴリに最適尺度値を割り当てます。分析内の名義変数または順序変数の元のラベルと異なり、これらの尺度値は計量プロパティを伴います。

ほとんどのカテゴリ手続きでは、各尺度変数に対する最適な数量化は、**交互最小 2 乗法**と呼ばれる反復法を利用して行われます。交互最小 2 乗法では、現在の数量化を使用して解を求めた後に、その解を使用して数量化を更新します。そして、更新された数量化を使用して新しい解を求め、さらにこの解を使用して数量化を更新する、というように、プロセスのなごらかの停止基準に達するまで繰り返します。

最適尺度法を使用する理由

カテゴリ データは、マーケティング リサーチ、サーベイ リサーチ、および社会行動科学におけるリサーチに多く現れます。実際、研究者の多くはカテゴリ データのみを使用しています。

ほとんどの標準モデルは特にカテゴリ データの分析に適しているのですが、次のようなデータ セットにはうまく働かないことがあります。

- 観測数が少なすぎる

- 変数が多すぎる
- 1 変数あたりの値が多すぎる

カテゴリを数量化すると、最適尺度法により、これらの状況で発生する問題を回避することができます。さらに、専用手法が適切な場合にも役立ちます。

パラメータ推定値を解釈するより、むしろグラフィック表示に基づいて最適尺度法の出力を解釈することがよくあります。最適尺度法には、優れた調査分析機能があり、他の IBM® SPSS® Statisticsモデルを適切に補完します。調査の対象を絞り、最適尺度法によりデータを視覚化することで、モデルパラメータの解釈を中心とする分析の基盤を構成することができます。

最適尺度水準と尺度

これは、初めてカテゴリ手続きを使用するときには、非常に混乱しがちな概念です。レベルを指定するとき、変数を測定するレベルではなく、尺度化するレベルを指定します。この考え方は、数量化する変数は、測定方法に関係なく、非線型関係を持っていてよいとするものです。

カテゴリを目的とする場合については、次のように 3 つの基本尺度があります。

- **名義**レベルでは、変数の値は順序付けられていないカテゴリを表します。名義変数としてよい変数の例としては、地区、郵便番号地域、宗教的所属関係、複数選択カテゴリがあります。
- **順序**レベルでは、変数の値は順序付けられたカテゴリを表します。例としては、満足度や信頼度を表す態度尺度と嗜好評価得点などがあります。
- **数値**レベルでは、変数の値は、有意な測定基準で順序付けられたカテゴリを表すので、このレベルは、カテゴリ間の距離の比較に適しています。例としては、年齢や、千ドル単位で表した所得があります。

たとえば、変数「人種」、「階級」、および「年齢」は、次の表に示されているようにコード化されます。

テーブル 1-1
人種、階級、年齢に対するコード化方式

地域		階級		年齢	
1	白人	1	低所得階級	20	20 歳
2	黒人	2	労働者階級/ 中流階級	22	22 歳
3	その他	3	マネージャ	25	25 歳
4	上流階級			27	27 歳

示されている値は、各変数のカテゴリを表しています。「人種」は名義変数となります。「人種」には4つのカテゴリがあり、固有の順序関係はありません。値1から3は、単に、3つのカテゴリを表しているだけです。コード方式は完全に任意です。一方、「階級」は、順序変数と仮定することができます。元のカテゴリは、下位の階級から上位の階級までを推移しています。大きなコードは、高い階級を表します。ただし、順序情報しか得られないので、隣接するカテゴリ間の距離については何もわかりません。対照的に、「年齢」は、数値型変数と仮定することができます。「年齢」の場合、値と値の間の距離には元々意味があります。20歳と22歳の距離は、25歳と27歳の距離と同じですが、22歳と25歳の距離は、これらのいずれよりも大きくなっています。

最適尺度水準の選択

指定すべき最適尺度水準を自動的にあらかじめ定義する特性を、変数は本来持っていないということを理解しておくことが重要です。データの探索には、意味があり、解釈が容易な方法であればどのような方法でも使用できます。たとえば、数値レベルの変数を順序レベルで分析する場合に、非線型変換を使用すると、解の次元を減らすことができます。

次の2つの例は、「明白な」尺度がなぜ最良の最適尺度水準とならない場合があるかを示しています。変数でオブジェクトを年齢グループに並べ替えると仮定します。年齢は、数値変数として尺度化することができますが、25歳未満の場合、安全性は年齢と正の関係を持ちますが、60歳を超える場合、安全性は年齢と負の関係を持つといえます。この場合、年齢を名義変数として扱う方が適しています。

他の例として、政治的関心を基準にして並べ替える変数は、本質的に名義変数のようです。しかし、左翼政党から右翼政党に政党を順序付ける場合、正当の数量化で順序レベルの分析を使用してこの順序を考慮することもできます。

排他的にどちらかのレベルにするための、あらかじめ定義されている特性が変数にないとしても、初心者役に役立つ一般的なガイドラインがあります。単一名義数量化では、通常、カテゴリの順序はわかりませんが、分析で順序を付けたいと考えます。カテゴリの順序がわかっている場合は、順序数量化を試してください。カテゴリが順序付けできない場合には、多重名義数量化を試してみるとよいでしょう。

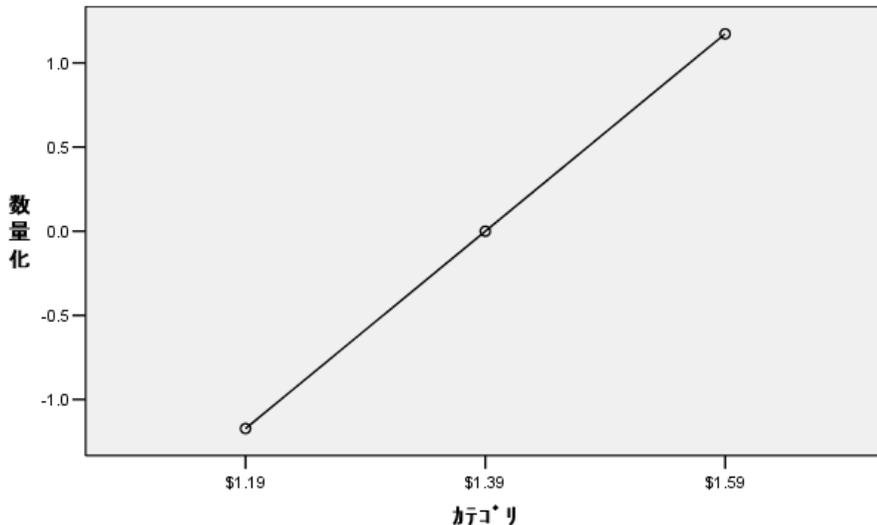
変換プロット

各変数を尺度化できるいろいろなレベルにより、数量化にさまざまな制限を加えます。変換プロットは、選択された最適尺度水準から得られる数量化と元のカテゴリとの関係を表します。たとえば、変数を数値変数として扱うと線型変換プロットが得られます。順序変数として扱われる変数からは、非減少変換プロットが得られます。U字形（または逆）になっている

名義変数として扱われる変数の変換プロットには、2次関係が示されます。名義変数は、カテゴリの順序を完全に変更することにより、見かけ上の傾向のない変換プロットを出力することもできます。次の図は、変換プロットの例を示しています。

変換プロットは、特に、選択された最適尺度水準がどれだけうまく機能するかを判別する場合に適しています。複数のカテゴリが類似の数量化を受ける場合、これらのカテゴリを1つのカテゴリに圧縮することも保証されます。それとは別に、名義変数として扱われる変数が増加トレンドを示す数量化を受ける場合、順序変換により類似の当てはめを得ることができます。そのトレンドが線型であれば、数値処理が適していると思われれます。ただし、カテゴリの圧縮や尺度レベルの変更が保証されていれば、分析が大幅に変わることはありません。

図 1-1
価格の変換プロット (数値)



カテゴリコード

カテゴリ変数をコード化するとき、コード化方式によっては不要な出力が生じたり、不完全な分析が得られることがあるため、いくつかの注意点が 있습니다。階級に使用できるコード化方式を次の表に示します。

テーブル 1-2
階級の代替コード化方式

カテゴリ	方式			
	A	B	C	D
低所得階級	1	1	5	1
労働者階級/中流階級	2	2	6	5
マネージャ	3	7	7	3

カテゴリ手続きによっては、使用するすべての変数の範囲を定義する必要がありますものもあります。この範囲を外れる値は欠損値として扱われます。最小カテゴリ値は常に 1 です。最大カテゴリ値はユーザーによって指定されます。この値は、変数に対するカテゴリの数ではなく、最大のカテゴリ値です。たとえば、表では、方式 A の最大カテゴリ値が 3 で、方式 B の最大カテゴリ値が 7 です。どちらの方式も、同じ 3 つのカテゴリをコード化します。

変数範囲は、分析から省かれるカテゴリを決定します。定義範囲を外れているコードを持つカテゴリが分析から除外されます。これは、カテゴリを省く単純な方法ですが、不要な分析が生じることがあります。最大カテゴリが間違っって定義されると、分析から有効なカテゴリが除外されることがあります。たとえば、方式 B では、最大カテゴリ値を 3 に定義すると、「階級」のカテゴリが 1 から 3 までの範囲でコード化されます。「上流階級」カテゴリは欠損値として扱われます。実際にはカテゴリは 3 とコード化されていないため、分析の第 3 のカテゴリにはケースが含まれません。すべての上位クラス カテゴリを除外すると、この分析は適切なものとなります。ただし、上位クラスを含める場合には、最大カテゴリを「10」と定義し、欠損値を 10 よりも大きい値、または 1 よりも小さい値でコード化しなければなりません。

名義変数または順序変数として扱われる変数では、カテゴリの範囲は結果に影響を及ぼしません。名義変数では、ラベルだけが重要であり、そのラベルと関連付けられている値は重要ではありません。順序変数では、カテゴリの順序は数量化で維持されます。カテゴリ値自体は重要ではありません。同じカテゴリ順序になるすべてのコード化方式は同一の結果を生み出します。たとえば、「階級」を順序レベルで分析した場合には、表の最初の 3 つの方式は機能的に同等です。カテゴリの順序はこれらの方式では同じです。一方、方式 D は、第 2 と第 3 のカテゴリを反転したものであり、他の方式とは異なる結果が得られます。

変数の多くのコード化方式は機能的には同等ですが、コードは手続きが出力する量に影響を及ぼすため、コードの違いが小さな方式の方が適しています。1 からユーザー定義の最大値までの範囲の値で、コード化したカテゴリはすべて有効です。これらのカテゴリが空であれば、対応する数量化の結果は、手続きしだいで、システム欠損値または 0 となります。これらの割り当てはいずれも分析に影響を及ぼしませんが、これらのカテゴリについては出力されます。したがって、方式 B では、「階級」に 4 つのカ

カテゴリがあり、システム欠損値を受け取ります。方式 C では、4 つのカテゴリがあり、システム欠損値指標を受け取ります。対照的に、方式 A では、システム欠損値の数量化がありません。名義変数または順序変数として扱われる変数のコードとして連続する整数を使用すると、結果に影響を及ぼすことなく、出力を減らすことができます。

数値変数として扱われる変数のコード化方式は、順序ケースに比べて制約が厳しくなっています。これらの変数では、連続するカテゴリ間の差が重要です。次の表では、「年齢」に対する 3 つのコード化方式を示しています。

テーブル 1-3
年齢の代替コード化方式

カテゴリ	方式		
	A	B	C
20	20	1	1
22	22	3	2
25	25	6	3
27	27	8	4

数値変数の値の再割り当てでは、カテゴリ間の差を保存しなければなりません。順序値を使用するのは、差を保存するための方法の 1 つです。ただし、この場合、システム欠損値指標を持つ多数のカテゴリができることがあります。たとえば、方式 A では、元の観測値を採用しています。コレスポネンス分析以外のすべてのカテゴリ手続きでは、最大カテゴリ値は 27、最小カテゴリ値は 1 に設定されています。最初の 19 件のカテゴリは空で、システム欠損値インジケータを受け取ります。最大カテゴリが 1 よりもかなり大きく、1 から最大値までの範囲に空のカテゴリが多数存在すると、出力は有用性の低いものとなってしまいます。

出力量を減らすには、値の再割り当てを実行します。ただし、数値ケースでは、[連続数への再割り当て] 機能を使用しないでください。連続整数に対してコード化すると、すべての連続カテゴリ間の差は 1 となり、そのため、すべての数量化は均等間隔となります。変数を数値変数として扱うときに重要とみなされる計量特性が、連続整数への値の再割り当てで破壊されます。たとえば、表の方式 C は、「年齢」に対する連続数への再割り当てに対応します。カテゴリ 22 と 25 の差は、3 から 1 に変更されており、数量化は後者の差を反映します。

カテゴリ間の差を保存する代替値再割り当て方式では、最小カテゴリ値をすべてのカテゴリから差し引き、それぞれの差に 1 を加えます。方式 B は、この変換から求められます。最小カテゴリ値 20 を各カテゴリから差し引き、1 をそれぞれの結果に加えました。変換されたコードの最小値は 1 であり、差はすべて元データと同じです。すると、最大カテゴリ値は 8 となり、最初の 0 でない数量化の前の 0 数量化はすべて取り除かれま

す。それでも、方式 B から得られる各カテゴリに対応する 0 でない数量化は、方式 A から得られる数量化と同じです。

アプリケーションに最適な手続きの選択

これらの手続きのうち 4 つ（コレスポネンス分析、多重応答分析、カテゴリ主成分分析、および非線型正準相関分析）で具現化されている手法は、**次元縮小**と呼ばれる多変量データ分析の一般分野に分類されます。つまり、変数間の関係は、できるだけ少ない次元（2 つまたは 3 つの次元）で表されます。そのため、データが多く複雑な元の形では、難しすぎて理解できない関係に含まれる構造やパターンを説明できません。マーケットリサーチの応用例では、これらの手法は**知覚マッピング**の形をとります。これらの手続きの大きな利点は、さまざまなレベルの最適尺度法でデータを処理できるという点にあります。

カテゴリ回帰分析は、カテゴリ応答変数とカテゴリ予測変数の組み合わせとの関係を説明する分析です。それぞれの予測変数が応答変数に及ぼす影響は、対応する回帰重み付けにより説明されます。他の手続きの場合と同様に、いろいろなレベルの最適尺度法でデータを分析することができます。

多次元尺度法と多次元展開により、近接度を使用して、低次元空間内のオブジェクト間の関係を記述します。

5 つの手続きについて以下に簡単なガイドラインを示します。

- カテゴリ従属変数の値をカテゴリ独立変数の組み合わせから予測するには、カテゴリ回帰分析を使用します。
- 混合最適尺度水準の 1 組の変数における変動のパターンを説明するには、カテゴリ主成分分析を使用します。
- 混合最適尺度水準の 2 組以上の変数の相関を求める範囲を評価するには、非線型正準相関分析を使用します。
- ブランドの好みや計量社会的選択データなど、2 元分割表または 2 元表として表すことができるデータを分析するには、コレスポネンス分析を使用します。
- すべての変数を名義レベルで分析するという、より強い仮定を行う場合にカテゴリ多変量データ行列を分析するには、多重応答分析を使用します。
- 低次元空間にあるオブジェクトの単一集合の最小 2 乗法表現を見つけるには、多次元尺度法を使用します。
- 低次元空間にあるオブジェクトの 2 集合の最小 2 乗法表現を見つけるには、多次元展開を使用します。

カテゴリ回帰

1 組の独立 (予測) 変数から従属 (応答) 変数を予測するのが分析の目標であれば、最適尺度法による回帰の使用が最も適しています。すべての最適尺度法手続きと同様、尺度値は、すべての変数の各カテゴリに割り当て、それらの値が回帰に関して最適になるようにします。カテゴリ回帰の解により、変換された応答と、変換された予測変数の重み付けされた組み合わせの平方相関が最大になります。

他のカテゴリ手続きとの関係。最適得点によるカテゴリ回帰は、2 組の変数を持ち、そのうち 1 つは従属変数のみを含む、最適尺度正準相関分析に相当します。後者の手法では、各組をすべての組の間のどこかに入る未知の変数と比較することで変数の組の類似度を導きます。カテゴリ回帰分析では、変換された応答変数と変換された予測変数の線型結合との類似度を直接評価します。

標準手法との関係。標準の線型回帰分析では、カテゴリ変数に対して指標変数として値の再割り当てを行ったり、間隔尺度変数と同じ方法でこの変数を扱うことができます。最初のアプローチでは、カテゴリ変数のレベルの組み合わせごとに異なる切片と傾きがモデルに含まれます。そのため、解釈すべきパラメータの数が多くなります。第 2 のアプローチでは、変数ごとに 1 つのパラメータしか推定しません。ただし、カテゴリのコード化は任意性があるため、一般化が不可能です。

変数の一部が量的でなければ、別の分析手法を使用することができます。応答変数が量的で、予測変数がカテゴリ的であれば、分散の分析がよく使われます。応答変数がカテゴリ的で、予測変数が量的であれば、ロジスティック回帰分析や判別分析が適しています。応答変数と予測変数が両方ともカテゴリ的であれば、対数線型モデルがよく使われます。

最適尺度法による回帰には、変数ごとに 3 つの尺度水準が用意されています。これらの水準を組み合わせると、1 つの「標準的な」方法が不適当な、さまざまな非線型関係を説明できます。その結果、最適尺度法では、複雑さを最小限に抑えながら、標準アプローチよりも柔軟性が高まっています。

さらに、予測変数に対し非線型変換を行うと、通常は、予測変数間の依存関係が減ります。また、予測変数の相関行列の固有値と最適尺度予測変数の相関行列の固有値を比較すると、通常、後者の変数の組は前者に比べて変動が少なくなります。つまり、カテゴリ回帰分析では、最適尺度法を使用することで、予測変数の相関行列の大きな固有値を小さくし、小さな固有値を大きくすることができるということです。

カテゴリ主成分分析

混合最適尺度水準の 1 組の変数における変動のパターンを説明するには、カテゴリ主成分分析を使用するのが最も適しています。この手法は、1 組の変数の次元を減らしながら、変動をできる限り説明しようという手法で

す。尺度値をすべての変数の各カテゴリに割り当て、それらの値が主成分分析の解に関して最適になるようにします。分析に含まれるオブジェクトは、数量化データに基づいて成分得点を受け取ります。成分得点の作図から、分析に含まれるオブジェクトのパターンがわかり、またデータ内の異常なオブジェクトも明らかにできます。カテゴリ主成分分析の解により、指定された成分（次元）の数に関して、オブジェクトスコアと数量化された各変数との相関が最大化されます。

カテゴリ主成分分析の重要な応用例として、選好データの調査があります。この調査では、応答者の順位付けや選好に関する項目数を評価します。通常の IBM® SPSS® Statistics データ布置では、行は個人、列は項目の測定結果、行にまたがる得点は選好得点（たとえば、0 ～ 10 の尺度で）で、データは行の条件に従います。選好データについては、個人を変数として扱うこともできます。データの行と列の入れ替えを行うには、入れ替え手続きを使用します。評価者が変数となり、すべての変数は順序変数として宣言されます。CATPCA で、オブジェクトよりも多い変数を使用することもできます。

他のカテゴリ手続きとの関係。すべての変数が多重名義変数として宣言されている場合、カテゴリ主成分分析を使用すると、同じ変数で実行した多重応答分析と同等の分析結果が得られます。したがって、カテゴリ主成分分析は、変数のうちいくつかは順序変数または数値変数として宣言されている多重応答分析の一種と見なすことができます。

標準手法との関係。すべての変数が数値レベルで尺度化されている場合、カテゴリ主成分分析は、標準主成分分析と同等です。

さらに一般的には、カテゴリ主成分分析は、非数値尺度の相関を計算し、標準主成分分析または因子分析のアプローチを使用して分析する方法の代わりになります。通常の Pearson の相関係数を順序データの関連付けの測度として使用する方法では、相関が推定値に非明示的な偏りを持つことがあります。

非線型正準相関分析

非線型正準相関分析は、いろいろな応用例のある非常に一般的な手続きです。非線型正準相関分析の目標は、主成分分析のように、変数自体の関係ではなく、変数の 2 つ以上の組の関係を分析することにあります。たとえば、2 組の変数があり、一方の組の変数は 1 組の回答者に関する人口統計学的な背景の項目であり、他方の組の変数は 1 組の態度項目に対する回答と考えられます。分析の尺度水準は、名義、順序、および数値の混合したものです。最適尺度正準相関分析では、同時に各組からの正準変数とオブジェクトに割り当てられている得点の折衷組とを比較することで、組の間の類似性を判別します。

他のカテゴリ手続きとの関係。2 つ以上の変数の組が存在するが、各組が持つ変数は実際には 1 つだけである場合、最適尺度正準相関分析は最適尺度主成分分析と同等になります。1 グループに 1 つの変数しかない分析でのすべての変数が多重名義変数の場合、最適尺度正準相関分析は多重応答分析に相当します。2 組の変数があり、そのうちの 1 つは変数を 1 つだけ含む場合、最適尺度正準相関分析は最適尺度法によるカテゴリ回帰分析と同等です。

標準手法との関係。標準の正準相関分析は、1 組の変数の線形結合と最大相関関係にある変数の第 2 の組変数の線型結合を見つける統計手法です。この線型結合の組が与えられると、正準相関分析により、小さな組の変数の個数に等しい最大数まで、正準変数と呼ばれる後続の線型結合の独立した組を見つけることができます。

分析に 2 組の変数があり、すべての変数が数値変数として定義されている場合、最適尺度正準相関分析は標準正準相関分析に相当します。IBM® SPSS® Statisticsでは正準相関分析手続きを用意していませんが、関係する統計量の多くは多変量分散分析から求めることができます。

最適尺度正準相関分析には、他にも多数の応用例があります。2 組の変数があり、一方の組の変数に単一名義変数として宣言されている名義変数が含まれている場合、最適尺度正準相関分析の結果を回帰分析と同様に解釈することができます。変数を多重名義とみなした場合、最適尺度分析は判別分析の代替手法となります。変数を 2 つ以上の組にグループ化すると、さまざまな方法でデータを分析できます。

対応分析

コレスポンデンス分析の目標は、コレスポンデンス テーブル用のパイロットを作成することにあります。コレスポンデンス テーブルでは、行変数と列変数は順序付けられていないカテゴリを表していると仮定します。したがって、名義最適尺度水準が常に使用されます。両変数とも、名義情報についてのみ検査されます。つまり、あるオブジェクトが同じカテゴリにあり、他のオブジェクトはそうではないという事実のみを考慮するという事です。同じ変数のカテゴリ間で距離または順序について何も仮定しません。

コレスポンデンス分析の具体的な使用例として、2 元分割表の分析があります。表に r 個のアクティブな行と c 個のアクティブな列がある場合、コレスポンデンス分析の解の次元数は r から 1 を引いた数値または c から 1 を引いた数値のうち、いずれか小さい方の最小値になります。つまり、分割表内の行カテゴリまたは列カテゴリを完全に表現できます。それには次元の空間を使用します。しかし、実際には、2 元表の行カテゴリと列カテゴリを低次元空間、たとえば 2 次元で表現する方が適しています。これは、2 次元の作図は、多次元空間表現に比べて理解しやすいからです。

使用できる次元の最大数が少ない場合、この分析で得られる統計量により、行カテゴリと列カテゴリが低次元表現でどれだけ正確に表されているかがわかります。2 次元の解の表現手法がよければ、行の点と列の点から

なるプロットを調べることで、行変数のどのカテゴリが似ているか、列変数のどのカテゴリが似ているか、そして列カテゴリと行カテゴリのどれが互いに似ているかがわかります。

他のカテゴリ手続きとの関係。単純コレスポンデンス分析は2元表に限定されます。対象となる変数が2つ以上ある場合は、変数を結合して、交互作用変数を作成することができます。たとえば、変数「人種」、「階級」、および「年齢」の場合、「人種」と「階級」を結合して、次の表に示す12個のカテゴリを持つ新しい変数「rejob」を作成できます。この新しい変数は、「年齢」との2元表(12行、4列)を構成します。この2元表は、コレスポンデンス分析で分析できます。

テーブル 1-4
人種と階級の結合

カテゴリ コード	カテゴリの定義	カテゴリ コード	カテゴリの定義
1	白人、低所得階級	7	黒人、中流階級
2	白人、労働階級	8	黒人、上流階級
3	白人、中流階級	9	その他、低所得階級
4	白人、上流階級	10	その他、労働階級
5	黒人、低所得階級	11	その他、中流階級
6	黒人、労働階級	12	その他、上流階級

このアプローチの欠点は、変数のペアのあらゆる組み合わせが発生するという点です。「階級」と「年齢」を結合して、別に12個のカテゴリ変数を作り出すことができます。または、「人種」と「年齢」を結合して、新しく16個のカテゴリ変数を作ることもできます。これらの交互作用変数はそれぞれ、残りの変数を含む2元表を構成します。これら3つの表のコレスポンデンス分析を行っても、同一の結果が得られませんが、それでも、有効なアプローチです。さらに、4つ以上の変数がある場合、一方の交互作用変数と他の交互作用変数を比較する2元表を構築することができます。分析に使用できる表の数は、変数の数が少なかったとしても、非常に多くなる可能性があります。1つの表を選択して分析することも、すべてを分析することもできます。それとは別に、多重応答分析手続きを使用することで、交互作用変数を作成しないで、すべての変数を同時に調べることができます。

標準手法との関係。クロス表手続きも、独立変数が分析の共通の対象として含まれる分割表の分析に使用できます。しかし、表がいくら小さくても、独立変数からの逸脱の原因を検出することが困難な場合があります。コレスポンデンス分析のユーティリティは、任意のサイズの2元表の、そのようなパターンを表示する機能を持っています。行変数と列変数に関連がある場合、つまり、カイ2乗値が有意な場合、コレスポンデンス分析では、関係の性質を簡単に明らかにできます。

多重応答分析

多重応答分析では、同じカテゴリのオブジェクト同士が近くにプロットされ、異なるカテゴリのオブジェクトは離れてプロットされる解を求めようとします。各オブジェクトは、そのオブジェクトに適用されるカテゴリのカテゴリポイントに、できる限り近い位置に配置されます。この方法で、カテゴリはオブジェクトを等質なサブグループに分割します。変数は、同じカテゴリ内のオブジェクトを同じサブグループに分類するときに等質であると考えられます。

1次元の解の場合、多重応答分析では、最適尺度値（カテゴリ数量化）を各変数の各カテゴリに割り当て、全体的に平均して、カテゴリが最大の広がりを持つようにします。2次元の解の場合、多重応答分析では、第1の組と無関係な各変数のカテゴリの数量化の第2の組を求め、さらに広がり最大化しようとする作業を、次々に続けます。変数のカテゴリは次元と同じ数の得点を受け取るので、分析の変数は最適尺度水準において多重名義変数と仮定されます。

多重応答分析では、さらに、得点を分析中のオブジェクトに割り当て、カテゴリ数量化がそのカテゴリ内のオブジェクトのオブジェクトスコアの平均、つまり重心となるようにします。

他のカテゴリ手続きとの関係。多重応答分析は、等質性分析または二重尺度法とも呼ばれています。変数が2つしかないときには、まったく同じではありませんが、コレスポンデンス分析に匹敵する結果が得られます。コレスポンデンス分析では、安定度の情報を含む、解の表現の当てはめおよび品質を要約した一意の出力が得られます。したがって、通常、コレスポンデンス分析は、変数が2つのケースでの多重応答分析に適しています。この2つの手続きの違いとしては他に、多重応答分析への入力、行がオブジェクト、列が変数であるデータ行列であるのに対して、コレスポンデンス分析への入力は、同じデータ行列、一般距離行列、または結合分割表である点が挙げられます。こちらは行と列の両方が変数のカテゴリを表す集計行列です。また、多重応答分析は、多重名義水準で尺度化したデータの主成分分析と見なすこともできます。

標準手法との関係。また、多重応答分析は多元分割表の分析と見なすこともできます。多元分割表は、クロス表手続きを使って分析することもできますが、クロス表は、各制御変数の各カテゴリについて集計統計量を別々に出力します。多重応答分析では、多くの場合、単一の2次元のプロットを使用して、すべての変数の関係を要約できます。多重応答分析の高度な使い方として、元のカテゴリ値を最初の次元からの最適尺度値で置き換えて、補助的な多変量分析を実行するという方法があります。多重応答分析ではカテゴリラベルを数値尺度値で置き換えるので、多重応答分析の後に数値データを必要とする多数の異なる手続きを適用することができます。たとえば、因子分析手続きでは、多重応答分析の最初の次元に相当する最初の主成分を出力します。最初の次元の成分得点は、オブジェクトスコア

に等しく、平方成分負荷は判別測定に等しい値です。しかし、2 番目の多重応答分析の次元は、因子分析の 2 番目の次元に等しくありません。

多次元尺度法

多次元尺度の使用は、分析の目標がオブジェクトまたはケースの単一集合間の 1 組の距離の構造を見つける場合に最も適しています。この手法では、観測値を概念上の低次元空間における特定の場所に割り当てます。この場合、空間における各ポイント間の距離が、指定した類似度（または非類似度）にできるだけ近くなるようにします。その結果得られるのが、その低次元空間におけるオブジェクトの最小 2 乗表現であり、多くの場合、データのより深い理解のために役立ちます。

他のカテゴリ手続きとの関係。距離を作成し、多次元尺度で分析する多変量データがある場合、分析結果は、オブジェクト主成分正規化とともにカテゴリ主成分分析を使用してデータを分析した結果と似ています。この種の PCA は、主座標分析とも呼ばれます。

標準手法との関係。カテゴリ多次元尺度手続き (PROXSCAL) は、Base System で利用できる尺度手続き (ALSCAL) を基にいくつか改良を加えたものです。PROXSCAL は、ある種のモデル用に加速アルゴリズムを用意しており、共通の空間に制限を課すことができます。さらに、PROXSCAL では、S ストレス (歪みとも呼ばれる) ではなく、正規化された原ストレスを最小限に抑えようとします。正規化された原ストレスは、距離に基づく測度であるため、一般にも好ましいのですが、S ストレスは平方距離に基づいています。

多次元展開

多次元展開の使用は、分析の目標がオブジェクトの 2 集合（行オブジェクトと列オブジェクトと呼びます）間の 1 組の距離の構造を見つける場合に最も適しています。この手法では、観測値を概念上の低次元空間における特定の場所に割り当てます。この場合、空間における各ポイント間の距離が、指定した類似度（または非類似度）にできるだけ近くなるようにします。その結果得られるのが、その低次元空間における、行と列の各オブジェクトの最小 2 乗表現であり、多くの場合、データのより深い理解のために役立ちます。

他のカテゴリ手続きとの関係。1 組のオブジェクト間の距離でデータが構成される場合は（2 次行列、対称行列）、多次元尺度法を使用してください。

標準手法との関係。カテゴリ多次元展開手続き (PREFSCAL) は、Base System で利用できる展開手続き (ALSCAL) を基にいくつか改良を加えたものです。PREFSCAL では、共通空間に制限を課すことができます。そのうえ、PREFSCAL では、ペナルティ付きストレス測定を最小限にすることで、（以前のアルゴリズムでは陥りがちであった）縮退解の回避を試みます。

最適尺度図表の縦横比

最適尺度法の作図の縦横比には、等方性があります。2次元作図では、次元1の1つの単位を示す距離は、次元2の1つの単位を示す距離と等しくなります。2次元作図の次元の範囲を変更する場合、他の次元のサイズも変更され、物理次元の同等性が保持されます。等方的縦横比は、最適尺度法手続きでは指定変更できません。

推奨参考文献

最適尺度法についての一般的な情報は、次のテキストを参照してください。

Barlow, R. E., D. J. Bartholomew, D. J. Bremner, および H. D. Brunk. 1972. *Statistical inference under order restrictions*. New York: John Wiley and Sons.

Benzécri, J. P. 1969. *Statistical analysis as a tool to make patterns emerge from data*. In: *Methodologies of Pattern Recognition*, S. Watanabe, ed. New York: Academic Press.

Bishop, Y. M., S. E. Feinberg, および P. W. Holland. 1975. *Discrete multivariate analysis: Theory and practice*. Cambridge, Mass.: MIT Press.

De Leeuw, J. 1984. *The Gifi system of nonlinear multivariate analysis*. In: *Data Analysis and Informatics III*, E. Diday, et al., ed.

De Leeuw, J. 1990. *Multivariate analysis with optimal scaling*. In: *Progress in Multivariate Analysis*, S. Das Gupta, および J. Sethuraman, 編集者. Calcutta: Indian Statistical Institute.

De Leeuw, J., および J. Van Rijckevorsel. 1980. *HOMALS and PRINCALS—Some generalizations of principal components analysis*. In: *Data Analysis and Informatics*, E. Diday, et al., ed. Amsterdam: North-Holland.

De Leeuw, J., F. W. Young, および Y. Takane. 1976. *Additive structure in qualitative data: An alternating least squares method with optimal scaling features*. *Psychometrika*, 41, .

Gifi, A. 1990. *Nonlinear multivariate analysis*. Chichester: John Wiley and Sons.

Heiser, W. J., および J. J. Meulman. 1995. *Nonlinear methods for the analysis of homogeneity and heterogeneity*. In: *Recent Advances in Descriptive Multivariate Analysis*, W. J. Krzanowski, ed. Oxford: Oxford University Press.

Israëls, A. 1987. *Eigenvalue techniques for qualitative data*. Leiden: DSWO Press.

- Krzanowski, W. J., および F. H. C. Marriott. 1994. Multivariate analysis: Part I, distributions, ordination and inference. London: Edward Arnold.
- Lebart, L., A. Morineau, および K. M. Warwick. 1984. Multivariate descriptive statistical analysis. New York: John Wiley and Sons.
- Max, J. 1960. Quantizing for minimum distortion. Proceedings IEEE (Information Theory), 6, .
- Meulman, J. J. 1986. A distance approach to nonlinear multivariate analysis. Leiden: DSWO Press.
- Meulman, J. J. 1992. The integration of multidimensional scaling and multivariate analysis with optimal transformations of the variables. Psychometrika, 57, .
- Nishisato, S. 1980. Analysis of categorical data: Dual scaling and its applications. Toronto: University of Toronto Press.
- Nishisato, S. 1994. Elements of dual scaling: An introduction to practical data analysis. Hillsdale, N. J.: Lawrence Erlbaum Associates, Inc.
- Rao, C. R. 1973. Linear statistical inference and its applications, 2nd ed. New York: John Wiley and Sons.
- Rao, C. R. 1980. Matrix approximations and reduction of dimensionality in multivariate statistical analysis. In: Multivariate Analysis, Vol. 5, P. R. Krishnaiah, ed. Amsterdam: North-Holland.
- Roskam, E. E. 1968. Metric analysis of ordinal data in psychology. Voorschoten: VAM.
- Shepard, R. N. 1966. Metric structures in ordinal data. Journal of Mathematical Psychology, 3, .
- Wolter, K. M. 1985. Introduction to variance estimation. Berlin: Springer-Verlag.
- Young, F. W. 1981. Quantitative analysis of qualitative data. Psychometrika, 46, .

カテゴリ回帰分析 (CATREG)

カテゴリ回帰分析は、カテゴリに数値を割り当ててカテゴリ データを数量化し、変換された変数の最適な線型回帰方程式を作ります。カテゴリ回帰分析は、CATREG という略語でも知られています。これは categorical regression の頭字語です。

標準線型回帰分析では、応答 (従属) 変数と、予測 (独立) 変数の重み付き組み合わせとの差を 2 乗したものの合計を最小化します。変数は標準的には量的変数で、(名義) カテゴリ データは値の再割り当てにより 2 値または対比変数に変換されます。その結果、カテゴリ変数をケース グループの分割に使用して、各グループの個々のパラメータのセットを推定できます。推定された係数は、予測変数の変化が応答変数にどのように影響するかを表します。応答の予測は、どのような予測変数の値の組み合わせについても行うことができます。

もう 1 つの手法として、応答変数をカテゴリ予測変数値自体で回帰させる方法があります。その結果、各変数に対して 1 つの係数が推定されます。しかし、カテゴリ変数の場合、カテゴリの値は任意です。異なる方法でカテゴリをコード化すると得られる係数も異なるため、同一の変数についての分析結果を比較することが難しくなります。

CATREG は、名義尺度、順序尺度、数値尺度を同時に使用することによって標準的なアプローチを拡張します。この手続きは、カテゴリ変数を元のカテゴリの特質を反映した形で数量化します。また、この手続きは、数量化されたカテゴリ変数を数値型変数と同様に扱います。非線型変換の使用により、変数をさまざまなレベルで分析して最適なモデルを見つけることができます。

例 カテゴリ回帰分析を使用して、仕事に対する満足感がどのように職種、地域、移動の量に依存しているかを記述できます。たとえば、高い満足度は、管理職と少ない移動量に対応していることがわかります。結果として得られる回帰式は、3 つの独立変数のあらゆる組み合わせに対する仕事の満足度を予測するために使用できます。

統計量と作図。度数分布、回帰係数、分散分析表、反復の記述、カテゴリ数量化、変換なしの予測変数間の相関、変換予測変数間の相関、残差プロット、変換プロット。

データ。 CATREG は、カテゴリ指標変数に対して実行されます。カテゴリ指標は、正の整数でなければなりません。小数値変数と文字型変数を正の整数に変換するには、[離散化] ダイアログ ボックスを使用します。

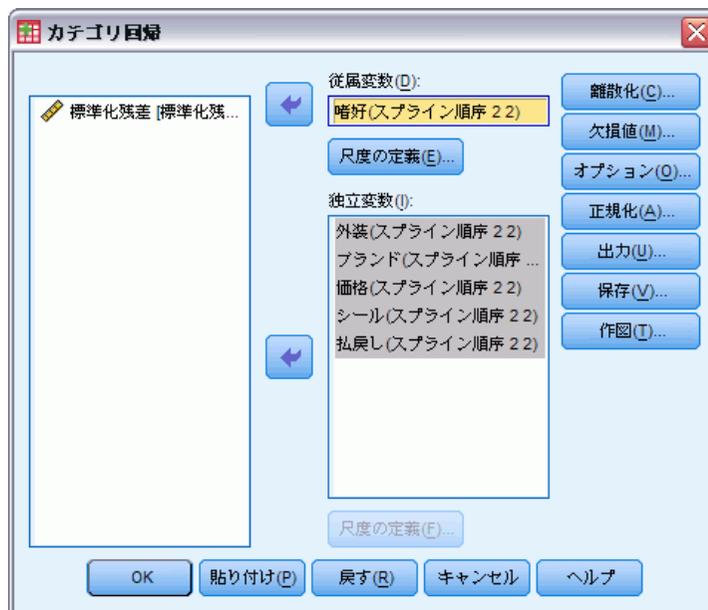
仮定。 応答変数は 1 つだけ使用できますが、予測変数の最大個数は 200 です。データには、少なくとも 3 つの有効なケースが含まれ、また、有効なケースの個数は予測変数の個数に 1 を加えた数よりも大きい必要があります。

関連手続き。 CATREG は、2 グループからなり、そのうちの 1 グループの変数が 1 つである最適尺度法のカテゴリ正準相関分析 (OVERALS) と同等です。すべての変数を数値レベルで尺度化した場合は、標準重回帰分析に相当します。

カテゴリ回帰分析を行うには

- ▶ メニューから次の項目を選択します。
分析(A) > 回帰 > 最適尺度法 (CATREG)...

図 2-1
[カテゴリ回帰] ダイアログ ボックス

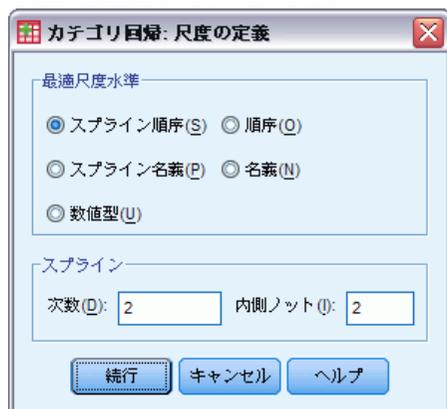


- ▶ 単一の従属変数と（単一または複数の）独立変数を選択します。
- ▶ [OK] をクリックします。
必要に応じて、各変数の尺度水準を変更します。

カテゴリ回帰分析の尺度の定義

従属変数と独立変数の最適尺度水準を設定できます。デフォルトでは、これらの変数は、2つの内側ノットを持つ2次の単調スプライン（順序）として尺度化されます。さらに、分析変数の重みを設定できます。

図 2-2
[尺度の定義] ダイアログ ボックス



最適尺度水準。各変数を数量化するための尺度水準を選択できます。

- **[スプライン順序]** 観測変数のカテゴリ順序は、最適尺度変数に保持されます。カテゴリ ポイントは、原点を通る直線（ベクトル）上にあります。ここで行われる変換は、選択された次数の滑らかな区分的単調多項式です。ユーザーが指定した内側ノットの数と手続きで決定されるその配置により、区分の数が決まります。
- **[スプライン名義]** 観測変数内の情報で最適尺度変数に保持されているのは、カテゴリ内のオブジェクトのグループだけです。観測変数のカテゴリ順序は保持されません。カテゴリ ポイントは、原点を通る直線（ベクトル）上にあります。ここで行われる変換は、選択された次数の滑らかな、場合によっては単調な区分的多項式です。ユーザーが指定した内側ノットの数と手続きで決定されるその配置により、区分の数が決まります。
- **順序。** 観測変数のカテゴリ順序は、最適尺度変数に保持されます。カテゴリ ポイントは、原点を通る直線（ベクトル）上にあります。ここで行われる変換は、スプライン順序変換よりもよく当てはまりますが、滑らかさが劣ります。
- **名義。** 観測変数内の情報で最適尺度変数に保持されているのは、カテゴリ内のオブジェクトのグループだけです。観測変数のカテゴリ順序は保持されません。カテゴリ ポイントは、原点を通る直線（ベクトル）

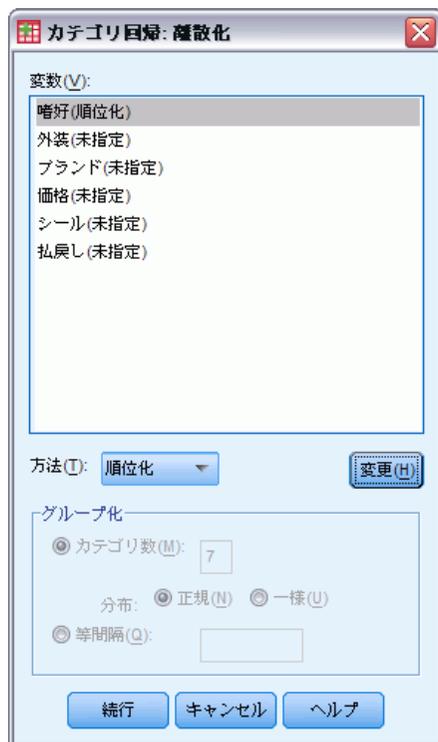
上にあります。ここで行われる変換は、スプライン名義変換よりもよく当てはまりますが、滑らかさが劣ります。

- **数値型**。カテゴリは順序化されており、カテゴリ間は等間隔である（間隔水準）とみなされます。カテゴリの順序とカテゴリ番号間の等間隔性は、最適尺度変数に保持されます。カテゴリポイントは、原点を通る直線（ベクトル）上にあります。すべての変数が数値水準である場合、この分析は標準の主成分分析に近くなります。

カテゴリ回帰分析: 離散化

[離散化] ダイアログボックスでは、変数の再割り当てを行う方法を選択できます。小数値変数は、別の方法で指定されない限り、ほぼ正規分布する7つのカテゴリ（またはその変数の持つ値の種類の数に7未満の場合はその数）に分類されます。文字型変数は、文字の昇順でカテゴリ指標を割り当てることにより、常に正の整数に変換されます。文字型変数の離散化では、これらの整数を適用します。デフォルトでは、他の変数はそのままです。これらの離散化された変数を分析で使用します。

図 2-3
[離散化] ダイアログボックス



方法。[グループ化]、[順位化]、[乗算] から選択します。

- **グループ化**。指定されたカテゴリ数に再割り当てするか、または等間隔に再割り当てします。
- **順位化**。変数は、ケースの順位化により離散化されます。
- **乗算**。変数の現在の値を標準化し、10 を掛けて丸めたものに、離散化された値の最小値が 1 となるように定数を加えます。

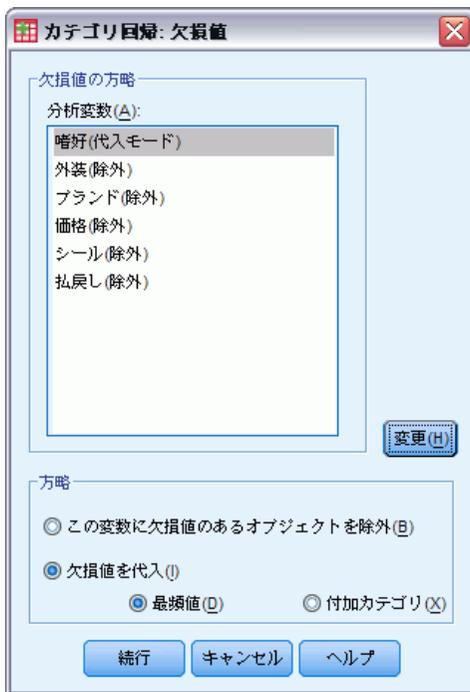
グループ化。グループ別で変数を離散化するときは、次のオプションを使用できます。

- **カテゴリ数**。カテゴリ数と、変数の値がカテゴリ全体にわたってほぼ正規分布または一様分布するかどうかを指定します。
- **等間隔**。変数は、等しい長さの間隔を持つカテゴリに再割り当てされます。間隔の長さを指定する必要があります。

カテゴリ回帰: 欠損値

[欠損値] ダイアログ ボックスでは、分析変数と補助変数の欠損値を処理するための方略を選択できます。

図 2-4
[欠損値] ダイアログ ボックス



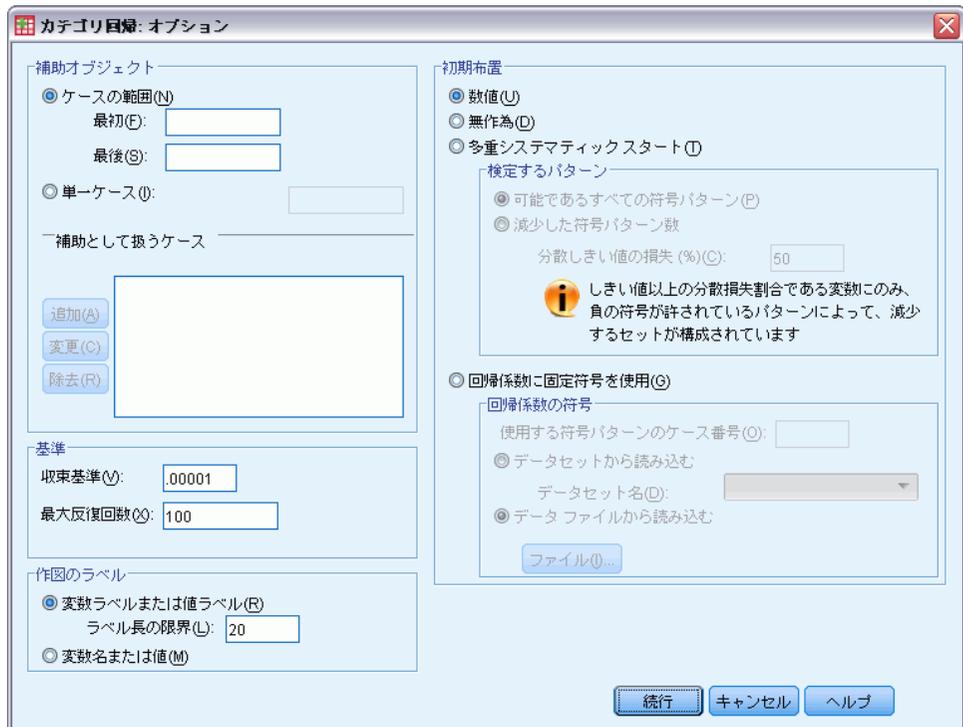
方略。欠損値のあるオブジェクトの除外（リストごとの削除）、または欠損値の代入（能動的な処理）を選択します。

- **この変数に欠損値のあるオブジェクトを除外(B)** 選択された変数に欠損値のあるオブジェクトが分析から除外されます。この方略は、補助変数には使用できません。
- **欠損値を代入(I)** 選択された変数に欠損値のあるオブジェクトでは、欠損値が代入されます。代入の方法を選択できます。欠損値を最頻カテゴリで置き換えるには、[最頻値]を選択します。最頻値が複数ある場合には、最小のカテゴリ指標を持つものを使用します。欠損値を同じ付加カテゴリの数量化で置き換えるには[付加カテゴリ]を選択します。これにより、その変数に欠損値のあるオブジェクトは、同じ(付加)カテゴリに属するとみなされます。

カテゴリ回帰: オプション

[オプション] ダイアログボックスには、初期布置の種類を選択するオプション、反復基準と収束基準を指定するオプション、補助(追加)オブジェクトを選択するオプション、プロットのラベル付けを設定するオプションがあります。

図 2-5
[オプション] ダイアログボックス



[補助オブジェクト] 補助オブジェクトとして扱うオブジェクトを指定できます。補助オブジェクト(またはケースの範囲を指定)の数を単純に入力し、[追加]をクリックします。補助オブジェクトに重みを付けることはできません(指定した重みは無視されます)。

初期布置。 名義変数として扱われる変数がない場合は、[数値]設定を選択します。名義変数として扱われる変数が1つでも存在する場合は、[無作為]設定を選択します。

代わりに、少なくとも1個の変数が、順序または順序スプライン尺度水準を持つときは、通常モデル当てはめアルゴリズムは最適な解を得られます。検定のすべての符号パターンと一緒に[多重システマティックスタート]を選択するとオプションの解が通常得られますが、データセットの順序または順序スプライン尺度水準の数の増加に伴い必要な処理時間が増えます。しきい値が高ければ多くの符号パターンが除外されますが、分散しきい値の損失のパーセントを指定して検定パターン数を少なくできます。オプションを使用すると、オプションの解の取得は保証できませんが、最適な解の取得の機会が減ります。また、オプションの解が見つからないと、最適な解がオプションの解と非常に相違することは減ります。多重システマティックスタートが要求されると、各スタートの回帰係数の符号が、現在のセッションの外部 IBM® SPSS® Statistics データまたはデータセットに書き込まれます。詳細は、p.26 [カテゴリ回帰: 保存](#)を参照してください。

多重システマティックスタートの前の実行の結果は、[回帰係数に固定符号を使用]を使用して取得できます。(1と-1により示される)符号は、指定されたデータセットまたはファイルの行の中にある必要があります。整数値開始番号は、使用される符号を含むファイルの行のケース番号です。

基準。 計算における回帰の最大反復数を指定できます。収束基準値の選択もできます。回帰の最後の2回の反復間の全体適合の差が収束値より小さくなるか、最大反復数に達した場合、回帰の反復は終了します。

作図のラベル。 変数ラベルと値ラベル、または変数名と値のどちらを作図で使用するかを指定できます。さらに、ラベルの最大長を指定することもできます。

[カテゴリ回帰正規化]

図 2-6
[正規化]ダイアログ ボックス

方法。 正規化の手法を使用すると、推定的小数部分を切り捨てて回帰係数の推定の多様性を削減し、モデルの予測誤差を改善します。Lasso および Elastic Net では、係数の推定値の小数部分を切り捨て、変数選択の形式を提供します。正規化方法が要求されると、各ペナルティ係数の正規化モデルと係数が、現在のセッションの外部 IBM® SPSS® Statistics データまたはデータセットに書き込まれます。詳細は、p. 26 [カテゴリ回帰: 保存](#) を参照してください。

- **Ridge 回帰。** Ridge 回帰は、**ペナルティ係数**を倍数にする平方係数の合計に等しいペナルティの項目を導入して係数を縮小できます。この係数は 0 (ペナルティなし) から 1 の範囲にでき、範囲と増分を指定して手順により「最良」のペナルティ値を検出します。

- **Lasso**。Lasso のペナルティ項目は絶対係数に基づいていて、ペナルティ係数の指定は Ridge 回帰の係数と似ていますが、Lasso はより集中的にコンピュータを使用することになります。
- **Elastic Net**。Elastic Net は Lasso と Ridge 回帰ペナルティを組み合わせたもので、「最良」の Lasso と Ridge 回帰ペナルティ係数を探索するために指定したグリッド値を検索します。Lasso と Ridge 回帰ペナルティを与えられると、Elastic Net は Lasso よりもコンピュータ的には高価ではありません。

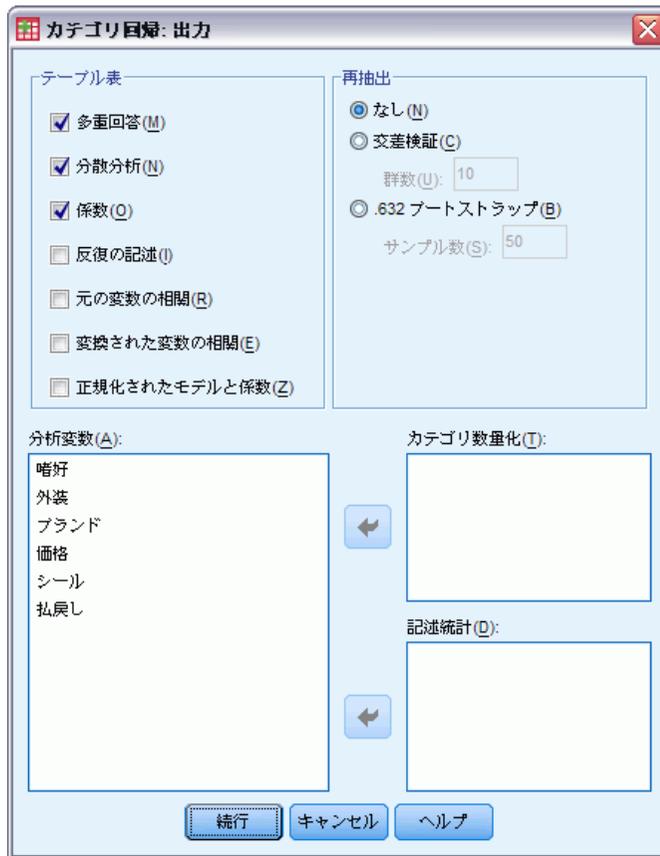
正規化プロットを表示します。 回帰係数を正常化ペナルティと対比するプロットがあります。「最良」のペナルティ係数の値の範囲を検索するとき、どのように回帰係数が範囲を変更するか表示します。

Elastic Net プロット。 Elastic Net 方法では、正常化プロットを分離すると Ridge 回帰ペナルティの値が生成されます。**すべてのプロット** は指定された最小と最大の Ridge 回帰ペナルティ値により決定された範囲のすべての値を使用します。**Ridge ペナルティ** では、最小と最大により決定された範囲の値のサブセットを指定できます。ペナルティ値(または値の範囲を指定)の数を単純に入力し、[追加] をクリックします。

カテゴリ回帰: 出力

[出力] ダイアログ ボックスでは、出力に表示する統計量を選択することができます。

図 2-7
[出力] ダイアログ ボックス



[テーブル] 次のテーブルが作成されます。

- **[重相関係数]** R^2 、調整済み R^2 、および最適尺度を考慮した調整済み R^2 が含まれます。
- **[分散分析]** 平方、平均平方、および F の回帰和および残差和が表示されます。2 つの [分散分析] テーブルが表示されます。一方のテーブルには予測変数の数に等しい回帰の自由度が表示され、もう一方のテーブルには最適尺度法を考慮した回帰の自由度が表示されます。
- **係数。** 次の 3 つのテーブルが指定されます。ベータ、ベータの標準誤差、t 値、および有意性を格納する係数テーブル。自由度の最適尺度レベルが考慮されたベータの標準誤差を格納する係数 - 最適尺度テーブル。0 次相関、部分相関、偏相関、変換後の予測変数に対する Pratt の相対的な重要度、および変換前後の許容度が格納されたテーブル。
- **反復の記述。** 反復ごとに、アルゴリズム、複数の R、および回帰誤差の初期値が表示されます。重相関係数の増加は、最初の反復から一覧されます。

- **元の変数の相関。**変換されていない変数間の相関を示す行列が表示されます。
- **変換された変数の相関。**変換された変数間の相関を示す行列が表示されます。
- **正規化モデルと係数。**各正規化モデルのペナルティ値、R-2 乗、回帰係数を表示します。再抽出方法が指定されたり、補助オブジェクト(検定ケース)が指定されると、予測誤差または MSE 検定を表示します。

再抽出。再抽出方法を使用するとモデルの予測誤差の推定ができます。

- **交差検証。**交差検証では、サンプルを群と呼ばれる複数のサブサンプルに分割します。分割の後、カテゴリ モデルが生成されますが、各サブサンプルのデータは除外されます。つまり、最初のモデルは最初のサブサンプル以外のすべてのケースを基に生成され、2 番目のモデルは 2 番目のサブサンプル以外のすべてのケースを基に生成されます。それぞれのモデルを、そのモデルの生成時に除外したサブサンプルに適用し、予測誤差のリスクを推定します。
- **.632 ブートストラップ。**ブートストラップを使用すると、1 つのブートストラップ サンプルを取得するためにこのプロセスを数回繰り返して、データを交換してランダムに抽出できます。モデルが各ブートストラップに適している、各モデルの予測誤差がこの適用モデルにより推定され、ブートストラップ サンプルにないケースに適用されます。

カテゴリ数量化。選択された変数の変換後の値を示すテーブルが表示されます。

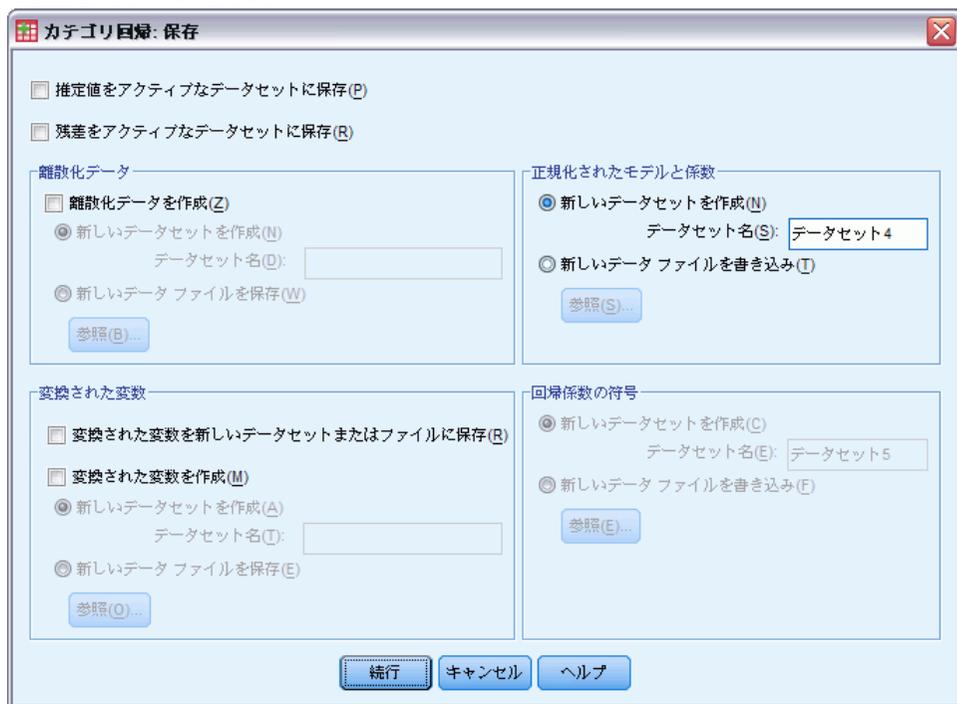
記述統計。選択された変数の度数、欠損値、および最頻値を示すテーブルが表示されます。

カテゴリ回帰: 保存

[保存] ダイアログ ボックスでは、予測値、残差、および変換された変数をアクティブなデータセットへ保存することができます。また、離散化データおよび変換された値、正規化モデルと係数、回帰係数を現在のセッションの外部 IBM® SPSS® Statistics データファイルやデータセットへ保存することもできます。

- データセットが利用可能なのは現在のセッション中ですが、今後のセッションで利用するには、データセットをデータ ファイルとして明示的に保存する必要があります。データセット名は、変数命名規則に従う必要があります。
- 保存されるデータの各タイプに応じて、ファイル名やデータセット名は異なる必要があります。

図 2-8
[保存] ダイアログ ボックス



正規化モデルと係数は、**[正規化]** ダイアログで正規化方法が選択されたときはいつでも保存されます。デフォルトでは、手続きが一意の名前の新規データセットを生成しますが、好きな任意の名前を選択できますし、外部のファイルに書き込みもできます。

多重システムティック スタートが**[オプション]** ダイアログで初期布置として使用されると回帰係数の符号が保存されます。デフォルトでは、手続きが一意の名前の新規データセットを生成しますが、好きな任意の名前を選択できますし、外部のファイルに書き込みもできます。

カテゴリ回帰: 変換プロット

[作図] ダイアログ ボックスでは、変換プロットおよび残差プロットを作成する変数を指定できます。

図 2-9
[作図] ダイアログ ボックス



変換プロット。 指定された各変数について、元のカテゴリの値に対する数量化をプロットします。空白のカテゴリは、横軸に表示されますが計算には影響を与えません。これらのカテゴリは、数量化された値を結ぶ線上の断絶として表されます。

残差プロット。 指定された各変数について、残差（問題の予測変数を除くすべての予測変数から予測された独立変数に対して計算されたもの）は、カテゴリ指標、およびカテゴリ指標に対するベータで乗算された最適なカテゴリ数量化に対して作図されます。

CATREG コマンドの追加機能

選択内容をシンタックス ウィンドウに貼り付け、CATREG コマンド シンタックスの結果を編集すると、カテゴリ回帰分析をカスタマイズできます。コマンド シンタックスを使用すると、次の作業も実行できます。

- 変換された変数をアクティブなデータセットに保存するときに、変数のルート名を指定 (SAVE サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

カテゴリ主成分分析 (CATPCA)

この手続きは、カテゴリ変数を数量化すると同時に、データの次元数を減らします。カテゴリ主成分分析は、CATPCA の略語でも知られています。これは categorical principal components analysis の頭字語です。

主成分分析の目的は、変数のグループ数を減らし、元の変数の持つ情報のほとんどを表す無相関の成分グループにまとめることです。この手法は、変数が大量にあるためにオブジェクト（被験者や単位）間の関係を効果的に解釈することができない場合に最も有効です。次元を減少させることで、多数の変数ではなく、2、3 の成分を解釈するだけで済ませることができます。

通常の主成分分析は、数値型変数間に線型関係があるものと仮定します。一方、最適尺度法では、変数を異なる水準で尺度化できます。カテゴリ変数は指定された次元で最適に数量化されます。その結果、変数間の非線型関係をモデルリングできます。

例。 カテゴリ主成分分析を使用して、仕事に対する満足感と、職種、地域、移動の量（大、中、小）との関係を視覚的に表現できます。たとえば、2 つの次元で分散のほとんどを説明できることがわかります。第 1 次元では職種と地域が離れています、第 2 次元では職種と移動の量が離れています。また、仕事の満足度の高さは移動の量が中程度であることに関連していることがわかります。

統計量と作図。 度数、欠損値、最適尺度水準、最頻値、重心座標、ベクトル座標およびその合計によって説明される分散（変数ごとおよび次元ごと）、ベクトル数量化変数の成分負荷、カテゴリ数量化と座標、反復の記述、変換された変数の相関と相関行列の固有値、元の変数の相関と相関行列の固有値、オブジェクトスコア、カテゴリプロット、結合カテゴリプロット、変換プロット、残差プロット、射影された重心のプロット、オブジェクトプロット、バイプロット、トリプロット、および成分負荷のプロット。

データ。 文字型変数値は、常に、文字の昇順で正の整数に変換されます。ユーザー定義の欠損値、システム欠損値、および 1 未満の値は、欠損値とみなされます。値が 1 未満である変数の場合は、値を再割り当てするか、または定数を加えることで非欠損値変数にできます。

仮定。 データには、少なくとも 3 つの有効なケースが含まれている必要があります。この分析は、正の整数データに基づいています。離散化オプションでは、値を「正規」に近い分布を持つカテゴリに分けることによ

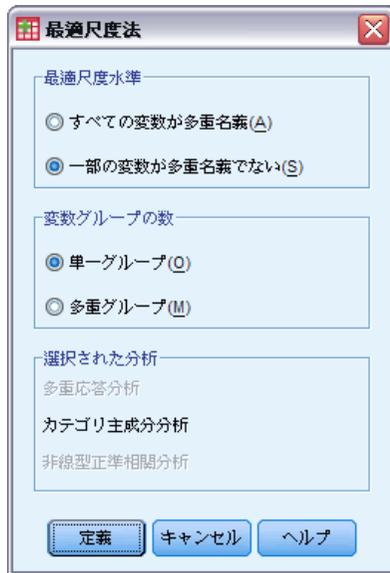
り、小数値変数を自動的に分類します。さらに、文字型変数の値を正の整数に自動的に変換します。他の離散化方式を指定することもできます。

関連手続き。すべての変数を数値水準で尺度化した場合は、標準の主成分分析に相当します。標準の線型主成分分析で変換された変数を使用すれば、代替の作図機能を使用できます。すべての変数が多重名義尺度水準を持つ場合、カテゴリ主成分分析は、多重応答分析と同一です。変数グループについて調べる場合は、カテゴリ（非線型）正準相関分析を使用します。

カテゴリ主成分分析を行うには

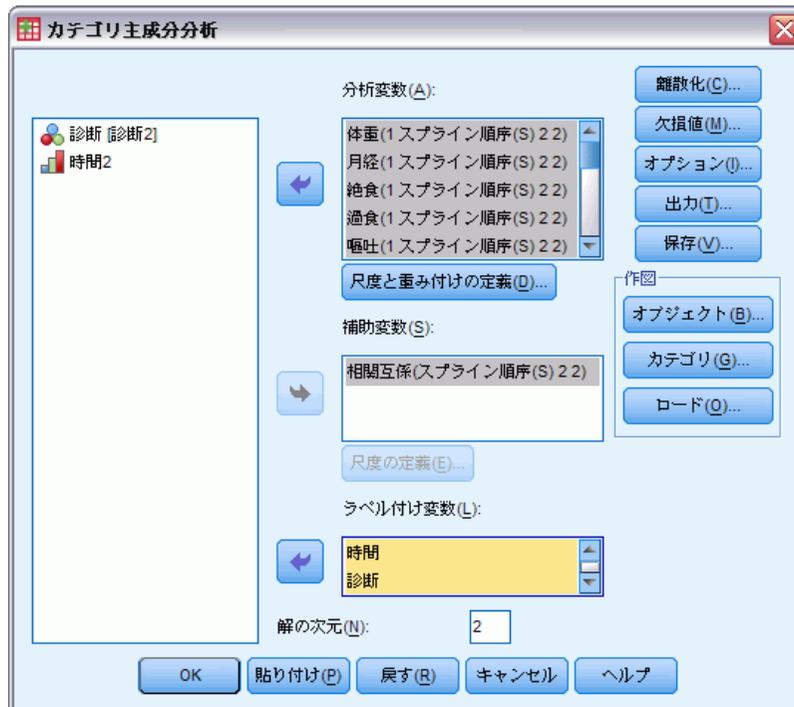
- ▶ メニューから次の項目を選択します。
[分析] > 次元縮小 > [最適尺度法]

図 3-1
[最適尺度法] ダイアログ ボックス



- ▶ [一部の変数が多重名義でない] を選択します。
- ▶ [単一グループ] を選択します。
- ▶ [定義] をクリックします。

図 3-2
[カテゴリ主成分分析] ダイアログ ボックス



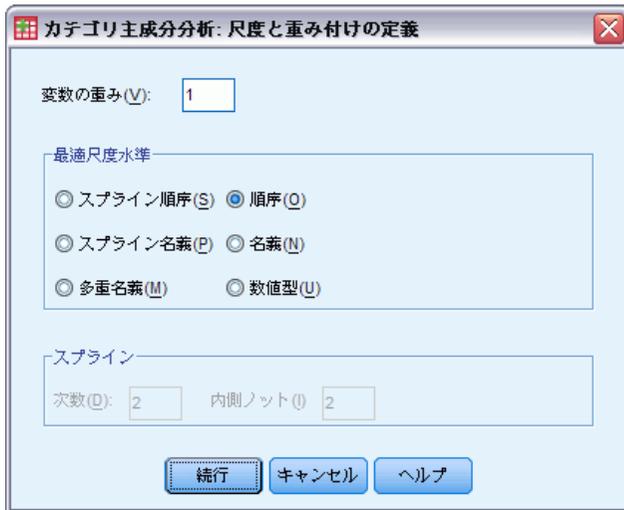
- ▶ 少なくとも 2 つの分析変数を選択し、解の次元の数を入力します。
- ▶ [OK] をクリックします。

必要に応じて、見つかった解に当てはめる補助変数、またはプロットのラベル付け変数を指定できます。

CATPCA の尺度と重み付けの定義

分析変数と補助変数の最適尺度水準を設定できます。デフォルトでは、これらの変数は、2 つの内側ノットを持つ 2 次の単調スプライン (順序) として尺度化されます。さらに、分析変数の重みを設定できます。

図 3-3
[尺度と重み付けの定義] ダイアログ ボックス



変数の重み。各変数の重みを定義できます。指定する値は、正の整数でなければなりません。デフォルト値は1です。

最適尺度水準。各変数を数量化するための尺度水準を選択できます。

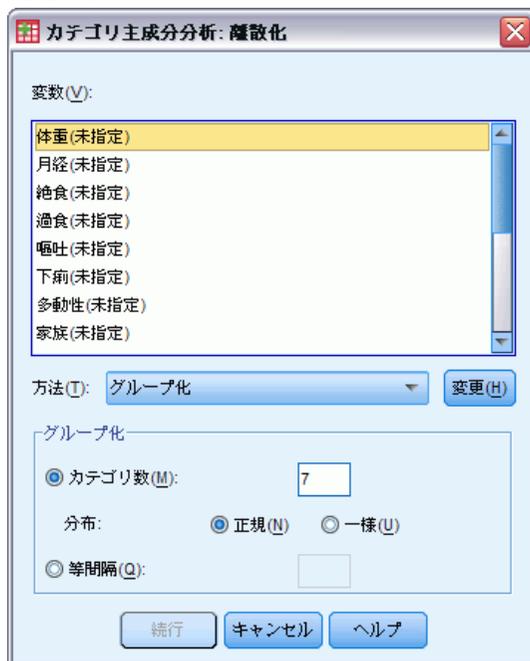
- **[スプライン順序]** 観測変数のカテゴリ順序は、最適尺度変数に保持されます。カテゴリ ポイントは、原点を通る直線（ベクトル）上にあります。ここで行われる変換は、選択された次数の滑らかな区分的単調多項式です。ユーザーが指定した内側ノットの数と手続きで決定されるその配置により、区分の数が決まります。
- **[スプライン名義]** 観測変数内の情報で最適尺度変数に保持されているのは、カテゴリ内のオブジェクトのグループだけです。観測変数のカテゴリ順序は保持されません。カテゴリ ポイントは、原点を通る直線（ベクトル）上にあります。ここで行われる変換は、選択された次数の滑らかな、場合によっては単調な区分的多項式です。ユーザーが指定した内側ノットの数と手続きで決定されるその配置により、区分の数が決まります。
- **[多重名義]** 観測変数内の情報で最適尺度変数に保持されているのは、カテゴリ内のオブジェクトのグループだけです。観測変数のカテゴリ順序は保持されません。カテゴリ ポイントは、そのカテゴリ内のオブジェクトの重心にあります。多重であるということは、各次元に対し異なる数量化グループが得られるということです。
- **順序。** 観測変数のカテゴリ順序は、最適尺度変数に保持されます。カテゴリ ポイントは、原点を通る直線（ベクトル）上にあります。ここで行われる変換は、スプライン順序変換よりもよく当てはまりますが、滑らかさが劣ります。

- **名義。** 観測変数内の情報で最適尺度変数に保持されているのは、カテゴリ内のオブジェクトのグループだけです。観測変数のカテゴリ順序は保持されません。カテゴリポイントは、原点を通る直線（ベクトル）上にあります。ここで行われる変換は、スプライン名義変換よりもよく当てはまりますが、滑らかさが劣ります。
- **数値型。** カテゴリは順序化されており、カテゴリ間は等間隔である（間隔水準）とみなされます。カテゴリの順序とカテゴリ番号間の等間隔性は、最適尺度変数に保持されます。カテゴリポイントは、原点を通る直線（ベクトル）上にあります。すべての変数が数値水準である場合、この分析は標準の主成分分析に近くなります。

カテゴリ主成分分析: 離散化

[離散化] ダイアログボックスでは、変数の再割り当てを行う方法を選択できます。小数値変数は、別の方法で指定されない限り、ほぼ正規分布する7つのカテゴリ（またはその変数の持つ値の種類数が7未満の場合はその数）に分類されます。文字型変数は、文字の昇順でカテゴリ指標を割り当てることにより、常に正の整数に変換されます。文字型変数の離散化では、これらの整数を適用します。デフォルトでは、他の変数はそのままです。これらの離散化された変数を分析で使用します。

図 3-4
[離散化] ダイアログボックス



方法。 [グループ化]、[順位化]、[乗算] から選択します。

- **グループ化**。指定されたカテゴリ数に再割り当てするか、または等間隔に再割り当てします。
- **順位化**。変数は、ケースの順位化により離散化されます。
- **乗算**。変数の現在の値を標準化し、10 を掛けて丸めたものに、離散化された値の最小値が 1 となるように定数を加えます。

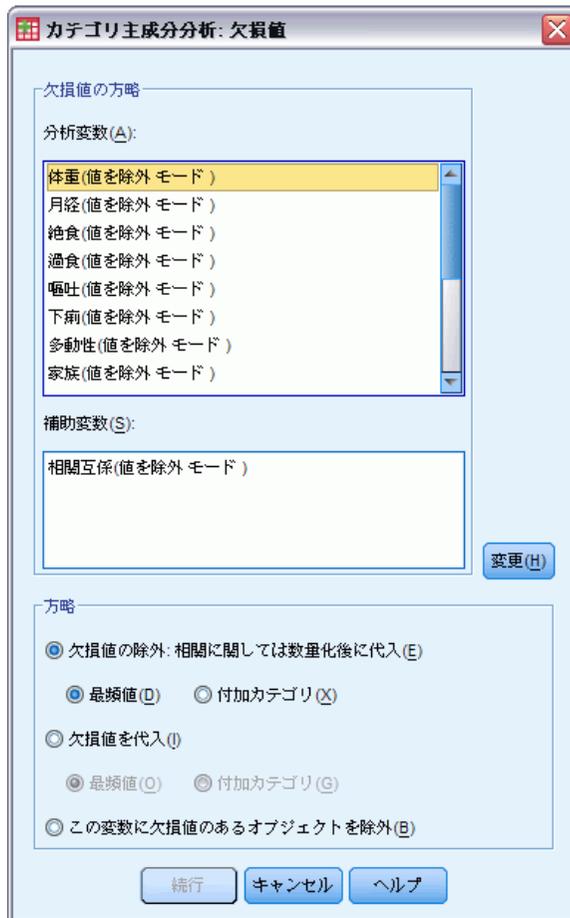
グループ化。グループ別で変数を離散化するときは、次のオプションを使用できます。

- **カテゴリ数**。カテゴリ数と、変数の値がカテゴリ全体にわたってほぼ正規分布または一様分布するかどうかを指定します。
- **等間隔**。変数は、等しい長さの間隔を持つカテゴリに再割り当てされます。間隔の長さを指定する必要があります。

カテゴリ主成分分析: 欠損値

[欠損値] ダイアログ ボックスでは、分析変数と補助変数の欠損値を処理するための方略を選択できます。

図 3-5
[欠損値] ダイアログ ボックス



方略。欠損値の除外（受動的な処理）、欠損値の代入（能動的な処理）、または欠損値のあるオブジェクトの除外（リストごとの削除）を選択します。

- 欠損値の除外: 相関に関しては数量化後に代入。** 選択された変数に欠損値のあるオブジェクトは、その変数の分析に関わりません。すべての変数に受動的な処理が指定されている場合、すべての変数に関して欠損値のあるオブジェクトは、補助変数として扱われます。[出力] ダイアログ ボックスで相関を指定した場合、元の変数の相関に関しては、分析後その変数の最頻カテゴリ、つまり最頻値が欠損値に代入されます。最適尺度変数の相関については、代入の方法を選択できます。欠損値を最適尺度変数の最頻値で置き換えるには、[最頻値] を選択します。欠損値を付加カテゴリの数量化で置き換えるには [付加カテゴリ] を選択します。これにより、その変数に欠損値のあるオブジェクトは、同じ (付加) カテゴリに属するとみなされます。

- **欠損値を代入。** 選択された変数に欠損値のあるオブジェクトでは、欠損値が代入されます。代入の方法を選択できます。欠損値を最頻カテゴリで置き換えるには、[最頻値]を選択します。最頻値が複数ある場合には、最小のカテゴリ指標を持つものを使用します。欠損値を同じ付加カテゴリの数量化で置き換えるには[付加カテゴリ]を選択します。これにより、その変数に欠損値のあるオブジェクトは、同じ（付加）カテゴリに属するとみなされます。
- **この変数に欠損値のあるオブジェクトを除外。** 選択された変数に欠損値のあるオブジェクトが分析から除外されます。この方略は、補助変数には使用できません。

カテゴリ主成分分析: オプション

[オプション] ダイアログ ボックスでは、初期布置の選択、反復および収束基準の指定、正規化方法の選択、作図のラベル付け方法の選択、補助オブジェクトの指定ができます。

図 3-6
[オプション] ダイアログ ボックス

補助オブジェクト。 補助オブジェクトにするオブジェクトのケース番号、またはオブジェクト範囲の先頭および最後のケース番号を指定し、[追加] をクリックします。すべての補助オブジェクトを指定するまで、処理を繰り返します。オブジェクトが補助オブジェクトとして指定された場合、そのオブジェクトに対してケースの重み付けは無視されます。

正規化。 オブジェクトスコアおよび変数の正規化を行う 5 つのオプションのうち、1 つを指定できます。1 つの分析で使用できる正規化の方法は 1 つだけです。

- **変数主成分。** このオプションを使用すると、変数間の関連が最適化されます。オブジェクト空間内の変数座標は、成分負荷（次元やオブジェクトスコアなどと主成分との相関）です。変数間の相関に主に関心がある場合に役立ちます。

- **オブジェクト主成分。**このオプションを使用すると、オブジェクト間の距離が最適化されます。オブジェクト間の相違または類似性に主に関心がある場合に役立ちます。
- **対称的。**オブジェクトと変数との関係に主に関心がある場合は、この正規化オプションを使用します。
- **独立。**オブジェクト間の距離、および変数間の相関を別々に調べる場合は、この正規化オプションを使用します。
- **ユーザー指定。**閉区間 $[-1, 1]$ 内の任意の実数値を指定できます。値 1 は「オブジェクト主成分」方法に相当し、値 0 は「対称的」方法に相当し、値 -1 は「変数主成分」方法に相当します。-1 よりも大きく 1 よりも小さい値を指定すると、固有値をオブジェクトと変数の両方に広げることができます。この方法は、目的に応じたバイプロットまたはトリプロットの作成に役立ちます。

基準。計算における手続きの最大反復数を指定できます。収束基準値の選択もできます。アルゴリズムの最後の 2 回の反復間における全体適合の差が収束値より小さくなるか、最大反復数に達した場合、アルゴリズムの反復は終了します。

作図のラベル。変数ラベルと値ラベル、または変数名と値のどちらを作図で使用するかを指定できます。さらに、ラベルの最大長を指定することもできます。

次元のプロット。出力表示する次元を制御できます。

- **[解のすべての次元の表示]** 解のすべての次元が散布図の行列に表示されます。
- **次元数の制限。**表示する次元をプロットされたペアに制限します。次元を制限する場合は、プロットする最低次元と最高次元を選択する必要があります。最低次元を、1 から解の次元数より 1 小さい値の間で指定できます。また、最低次元は、それよりも高い次元に対してプロットされます。次元の最高値には、2 から解の次元数までを指定できます。この値は、次元のペアのプロットで使用される次元の中での最高次元を示します。この指定は、要求されたすべての多次元プロットに適用されます。

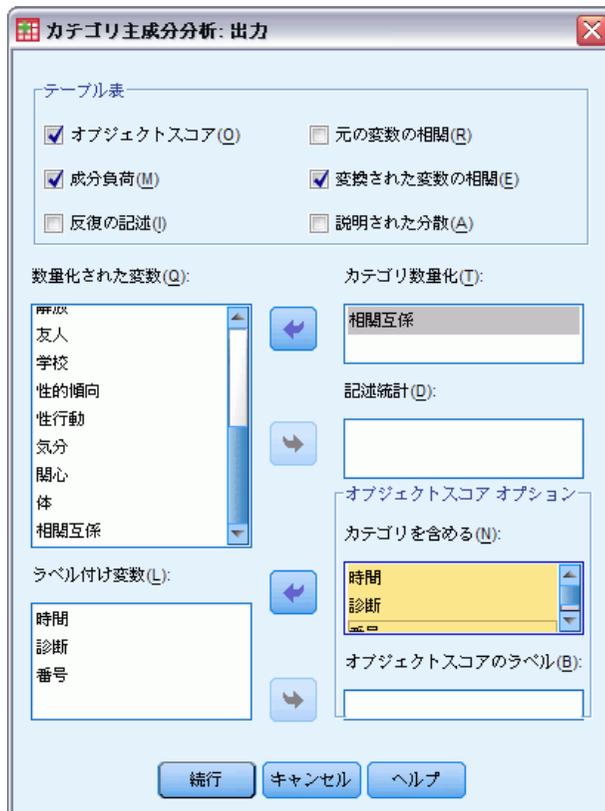
布置。布置の座標が記述されているファイルからデータを読み込むことができます。ファイルの最初の変数は最初の次元の座標を含み、2 番目の変数は 2 番目の次元の座標を含みます (以下同様です)。

- **初期。**指定されたファイル内の布置を分析の開始点として使用します。
- **固定。**指定されたファイル内の布置を変数への当てはめに使用します。当てはめられる変数は、分析変数として選択しなければなりません。布置は固定されているため、補助変数として扱われます (したがって、補助変数として選択する必要はありません)。

カテゴリ主成分分析: 出力

[出力] ダイアログ ボックスでは、オブジェクトスコア、成分負荷、反復の記述、元の変数の相関と変換された変数の相関、変数ごとおよび次元ごとの説明された分散、選択された変数のカテゴリ数量化、および選択された変数の記述統計量を表したテーブルを出力できます。

図 3-7
[出力] ダイアログ ボックス



オブジェクトスコア。オブジェクト スコアを表示します。次のオプションが用意されています。

- **カテゴリを含める。**選択された分析変数のカテゴリ指標を表示します。
- **オブジェクト スコアのラベル。**ラベル付け変数として指定されている変数のリストから、オブジェクトにラベル付けする変数を選択できます。

成分負荷。多重名義尺度水準を与えられてないすべての変数の成分負荷を表示します。

反復の記述。反復ごとに、説明された分散、損失、および説明された分散の増加量が表示されます。

元の変数の相関。元の変数の相関行列とその行列の固有値を表示します。

変換された変数の相関。変換（最適尺度）変数の相関行列とその行列の固有値を表示します。

説明された分散。変数ごとおよび次元ごとに、重心座標、ベクトル座標、および合計（重心座標とベクトル座標の結合）によって説明された分散の量を表示します。

カテゴリ数量化。選択された変数の次元ごとにカテゴリ数量化および座標を表示します。

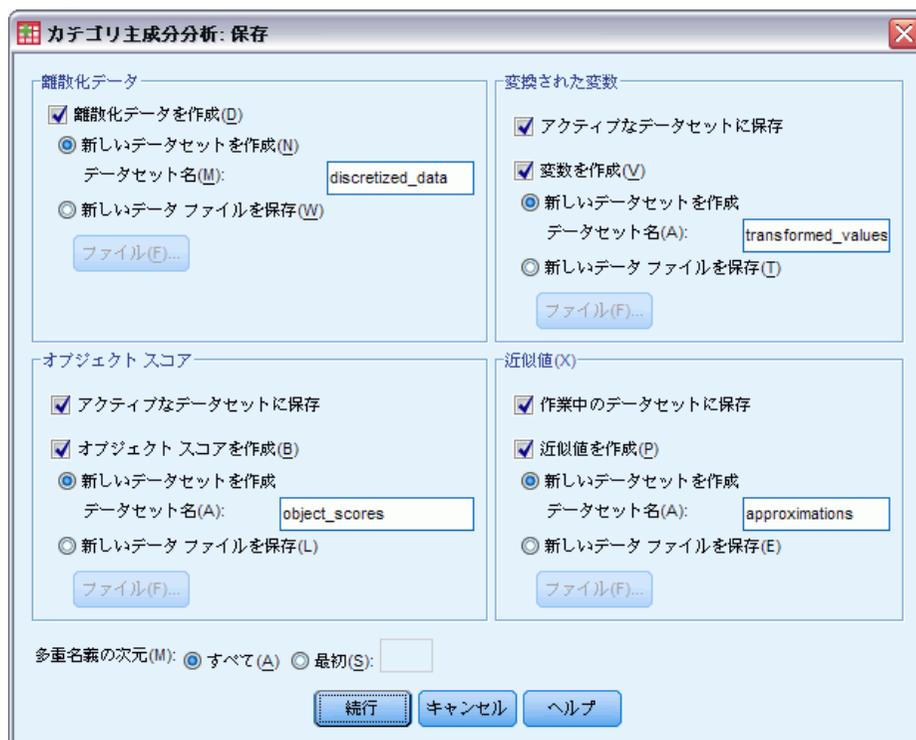
記述統計。選択された変数の度数、欠損値の数、および最頻値を表示します。

カテゴリ主成分分析: 保存

[保存] ダイアログ ボックスでは、離散化データ、オブジェクトスコア、変換された変数、および近似値を、現在のセッションの外部 IBM® SPSS® Statistics データファイルやデータセットに保存できます。変換された変数、オブジェクトスコア、および、近似値も、アクティブなデータセットに保存することができます。

- データセットが利用可能なのは現在のセッション中ですが、今後のセッションで利用するには、データセットをデータ ファイルとして明示的に保存する必要があります。データセット名は、変数命名規則に従う必要があります。
- 保存されるデータの各タイプに応じて、ファイル名やデータセット名は異なる必要があります。
- オブジェクトスコアや変換された値をアクティブなデータセットに保存する場合、多重名義尺度水準の番号を指定することができます。

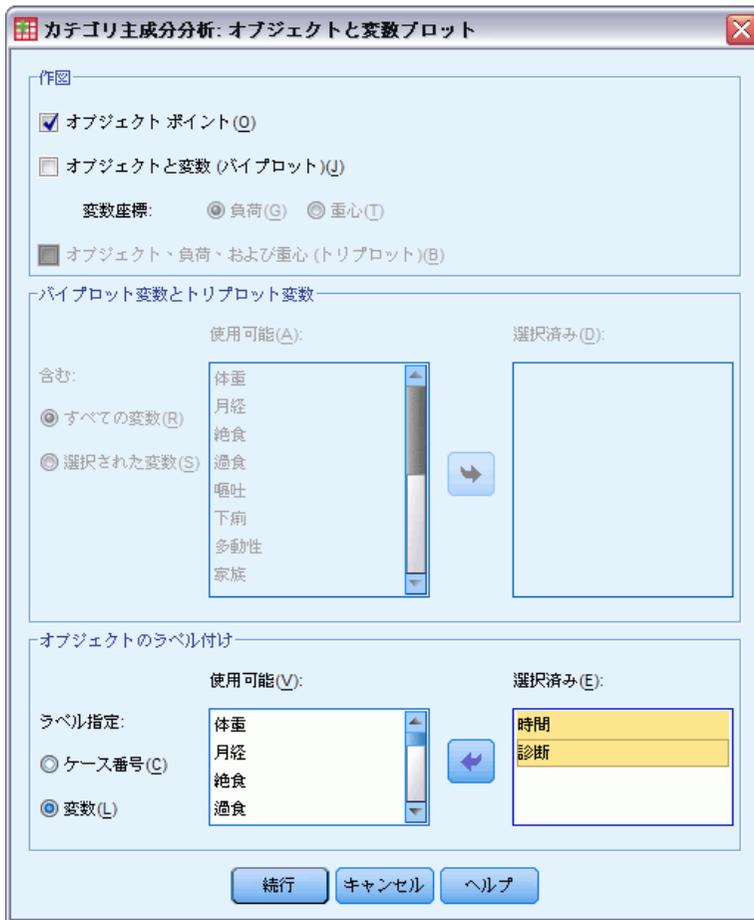
図 3-8
[保存] ダイアログ ボックス



カテゴリ主成分分析: オブジェクト プロット

[オブジェクトと変数プロット] ダイアログ ボックスで、使用したいプロットの種類とプロットを作成する変数を指定できます。

図 3-9
[オブジェクトと変数プロット] ダイアログ ボックス



オブジェクト ポイント。オブジェクト ポイントのプロットが表示されます。

オブジェクトと変数 (バイプロット)。オブジェクト ポイントは、変数座標として選択した成分負荷または変数の重心とともに作図されます。

オブジェクト、負荷、および重心 (トリプロット)。オブジェクト ポイントは、多重名義尺度水準の変数の重心および他の変数の成分負荷とともに作図されます。

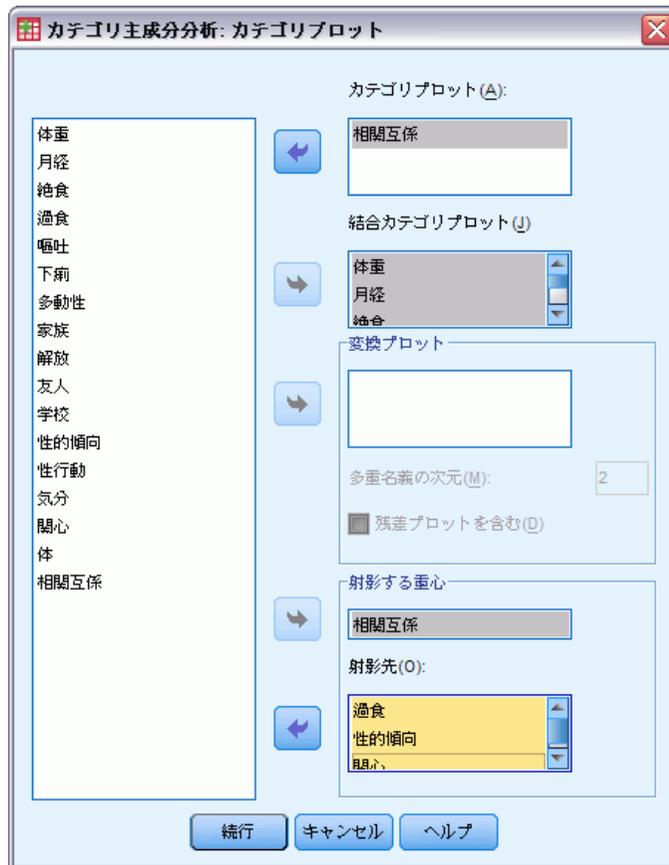
バイプロット変数とトリプロット変数。バイプロットおよびトリプロットに対するすべての変数の使用、またはサブグループを選択できます。

オブジェクトのラベル付け。選択された変数のカテゴリ (カテゴリ指標値か値ラベルかを [オプション] ダイアログ ボックスで選択できます) またはケース番号でオブジェクトにラベル付けできます。[変数] が指定されている場合は、変数ごとにプロットが 1 つ出力されます。

カテゴリ主成分分析: カテゴリ プロット

[カテゴリ プロット] ダイアログ ボックスで、使用したいプロットの種類とプロットを作成する変数を指定できます。

図 3-10
[カテゴリ プロット] ダイアログ ボックス



カテゴリプロット。選択した変数ごとに、重心とベクトル座標のプロットを作図します。多重名義尺度水準を持つ変数の場合、カテゴリはそのカテゴリ内のオブジェクトの重心にあります。他のすべての尺度水準では、カテゴリは原点を通るベクトル上にあります。

結合カテゴリプロット。これは、選択した各変数の重心とベクトル座標をひとつのプロットに示したものです。

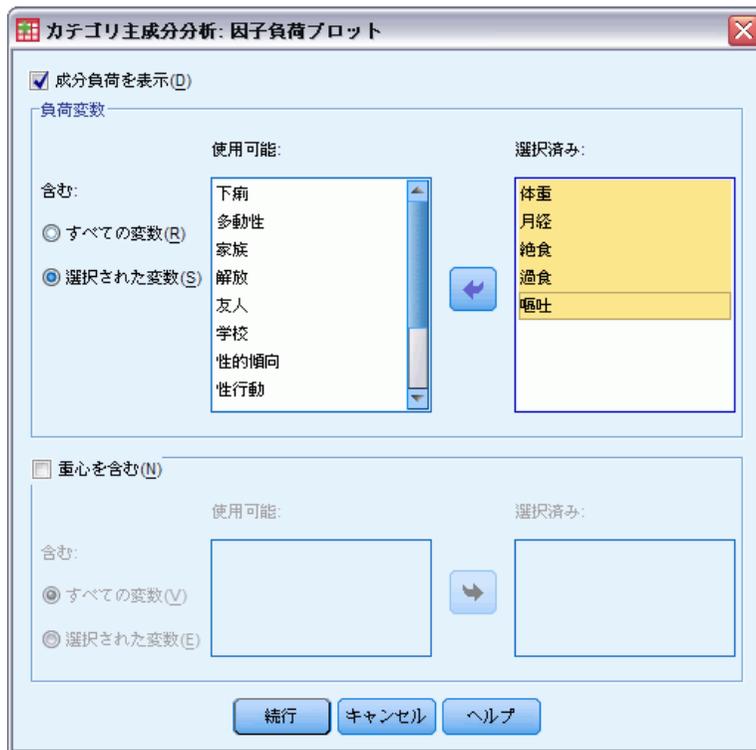
変換プロット。カテゴリ指標と対比して最適カテゴリ数量化のプロットを表示します。多重名義尺度水準を持つ変数に対し、使用したい次元の数を指定できます。次元ごとに1つのプロットが生成されます。選択された変数ごとに残差プロットを表示するようにもできます。

射影する重心。変数を選択し、その重心を選択された変数に射影できます。多重名義尺度水準を持つ変数は、射影先として選択することはできません。このプロットの出力を指定すると、射影された重心の座標のテーブルも表示されます。

カテゴリ主成分分析: 負荷プロット

[因子負荷プロット] ダイアログ ボックスでは、プロットに含める変数や、重心をプロットに含めるかどうかを指定できます。

図 3-11
[因子負荷プロット] ダイアログ ボックス



成分負荷を表示。選択すると、成分負荷のプロットが表示されます。

負荷変数。成分負荷プロットに対するすべての変数の使用、またはサブグループを選択できます。

重心を含む。多重名義尺度水準を持つ変数には成分負荷がありませんが、プロットにそれらの変数の重心を含めるようにすることはできます。すべての多重名義変数を使用するか、サブグループを選択できます。

CATPCA コマンドの追加機能

選択内容をシンタックス ウィンドウに貼り付け、CATPCA コマンド シンタックスを編集することで、カテゴリ主成分分析をカスタマイズできます。コマンド シンタックスを使用すると、次の作業も実行できます。

- 変換された変数、オブジェクト スコア、および近似値をアクティブなデータセットに保存するときに、変数のルート名を指定 (SAVE サブコマンドを使用)。
- 各プロットのラベルの最大長を別々に指定 (PLOT サブコマンドを使用)。
- 残差プロットの個別変数リストを指定 (PLOT サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

非線型正準相関分析 (OVERALS)

非線型正準相関分析は、最適尺度法のカテゴリ正準相関分析に相当します。この手続きの目的は、カテゴリ変数グループが互いにどのように類似しているかを明らかにすることです。非線型正準相関分析は、OVERALS の略語でも知られています。

標準正準相関分析は重回帰分析の拡張ですが、第 2 グループは単一の応答変数ではなく、複数の応答変数を含みます。この手法の目的は、低次元空間での 2 グループの数値型変数間の関係における分散を説明することです。まず、各グループの変数は、線型結合が最大の相関を持つように線型的に結合されます。この組み合わせが与えられると、先行する結合と相関がなく、かつ最大の相関を持つ次の線型結合が決定されます。

最適尺度法は、標準分析を次の 3 点で大きく拡張するものです。第 1 に、OVERALS では 3 つ以上の変数グループを使用できます。第 2 に、変数は名義、順序、数値のどの尺度でも取ることができます。したがって、変数間の非線型関係の分析ができます。最後に、変数グループ間の相関を最大化する代わりに、変数グループは、オブジェクト スコアによって定義された未知のグループと比較されます。

例。最適尺度法のカテゴリ正準相関分析を使用して、職種および教育年数を含む変数の 1 グループと居住地区および性別を含む他の変数グループとの関係を図示します。教育年数と居住地区がその他の変数よりも判別に役立つことがわかります。また、第 1 次元では、教育年数が最も判別性があることがわかります。

統計量と作図。度数分布、重心、反復の記述、オブジェクト スコア、カテゴリ数量化、重み付け、成分負荷、単一および多重適合、オブジェクト スコア プロット、カテゴリ軸プロット、成分負荷プロット、カテゴリ重心プロット、変換プロット。

データ。カテゴリ変数（名義または順位尺度レベル）をコード化するには整数を使用します。出力を最小にするには、各変数をコード化するとき、1 で始まる連続した整数を使用します。数値レベルで尺度化した変数を連続整数に再割り当てすることはできません。各変数が数値レベルで尺度化された場合に出力を最小にするには、各値から最小観測値を引いて 1 を加えます。小数点付きの値については、小数点以下を切り捨てます。

仮定。変数は2つ以上のグループに分けることができます。分析で使用する変数には、多重名義、単一名義、順序、または数値型の尺度を使用できます。手続きで使用される次元の最大数は、変数の最適尺度法レベルに依存します。すべての変数が順序、単一名義、数値のいずれかの型である場合に次元の最大数は、観測値の数から1を引いたものと変数の総数の小さい方です。しかし、定義された変数が2グループである場合、次元の最大数は小さい方のグループの変数の数です。多重名義変数が含まれる場合の次元の最大数は、多重名義カテゴリの総数に非多重名義変数の数を足したのから多重名義変数の数を引いたものです。たとえば、分析に5つの変数が含まれ、そのうち1つが4つのカテゴリを持つ多重名義変数である場合、次元の最大数は $(4 + 4 - 1)$ で7になります。最大値よりも大きな値を指定すると、最大値が使用されます。

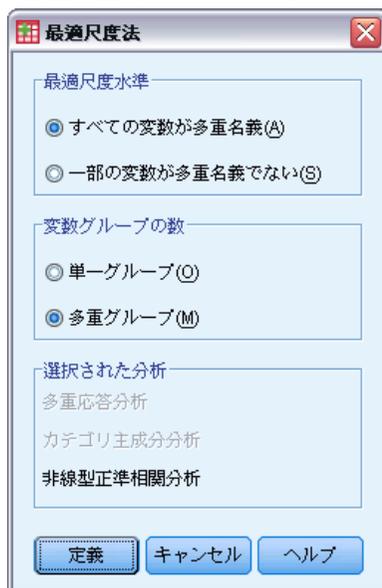
関連手続き。各グループに変数が1つ含まれる場合、非線型正準相関分析は、最適尺度法の主成分分析と同じです。これらの変数が多重名義の場合は、多重応答分析に相当します。変数グループが2つあり、グループの1つに変数が1つしかない場合、非線型正準相関分析は最適尺度法のカテゴリ回帰と同じです。

非線型正準相関分析を行うには

- ▶ メニューから次の項目を選択します。

[分析] > 次元縮小 > [最適尺度法]

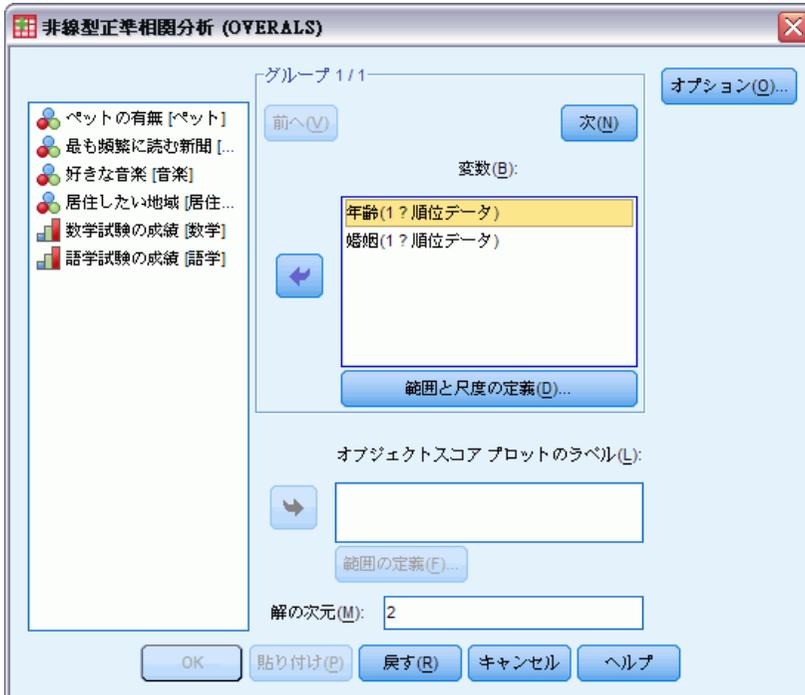
図 4-1
[最適尺度法] ダイアログ ボックス



- ▶ [すべての変数が多重名義] または [一部の変数が多重名義でない] を選択します。

- ▶ [多重グループ] を選択します。
- ▶ [定義] をクリックします。

図 4-2
[非線型正準相関分析 (OVERALS)] ダイアログ ボックス



- ▶ 最低 2 つの変数のグループを定義します。第 1 のグループに含める変数を選択します。次のグループへ移動するには [次] をクリックし、第 2 のグループに含める変数を選択します。さらにグループを追加することもできます。すでに定義した変数のグループに戻るには、[前] をクリックします。
- ▶ 選択された変数の範囲と測定尺度（最適尺度水準）を定義します。
- ▶ [OK] をクリックします。
- ▶ 次のオプションが選択できます。
 - 1 つ以上の変数を選択して、オブジェクトスコアプロットのポイントラベルを付与します。各変数は個別のプロットを作成し、ポイントにはその変数の値でラベル付けされます。作図のラベル付けに使用する変数には、それぞれ範囲を定義しなければなりません。ダイアログボックスを使用する際、分析とラベル付けの両方に同一の変数を指定することはできません。オブジェクトスコアのプロットを、分析で使用される変数でラベル付けする場合は、[変換] メニューの [計算] 機能を使用して、

その変数のコピーを作成します。図のラベル付けには新しい変数を使用します。または、コマンド シンタックスを使用することもできます。

- 適切な解の次元の数を指定します。一般に、変数の大半を説明するのに足りる最低限の次元を選択します。分析に 3 つ以上の次元が含まれる場合、最初の 3 つの次元の 3-D のプロットを作成します。その他の次元を表示するには、図を編集します。

範囲と尺度の定義 (Categories)

図 4-3
[範囲と尺度の定義] ダイアログ ボックス



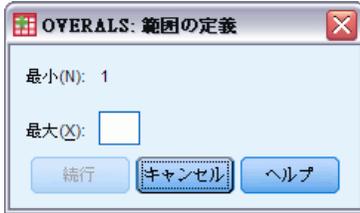
各変数には、範囲を定義する必要があります。指定する最大値は、整数でなければなりません。小数点付きデータ値は、分析では切り捨てられます。また、指定された範囲外のカテゴリ値は無視されます。出力を最小にするには、[変換] メニューの [連続数への再割り当て] 機能を使用して、名義または順序変数として扱われる変数に対して 1 で始まる連続的なカテゴリを作成します。数値型変数の連続整数に対して、値を再割り当てしないでください。数値型変数の出力を最小にするには、各変数のすべての値から最小値を引いたものに 1 を加えます。

各変数を数量化するには、使用する尺度を選択する必要があります。

- **順序**。観測変数のカテゴリ順序は、数量化された変数に保持されます。
- **単一名義**。数量化された変数では、同一カテゴリ内のオブジェクトの得点は同じです。
- **多重名義**。数量化は、各次元で異なる場合があります。
- **離散数値**。カテゴリは順序付けされ、カテゴリ間は等間隔であるとみなされます。カテゴリ間の差異と観測変数のカテゴリの順序は、数量化された変数においても保持されます。

範囲の定義 (プロビット分析/MANOVA)

図 4-4
[範囲の定義] ダイアログ ボックス



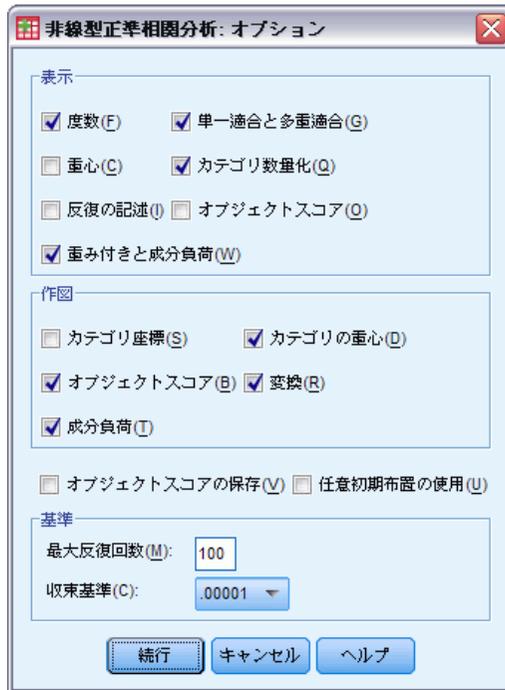
各変数には、範囲を定義する必要があります。指定する最大値は、整数でなければなりません。小数点付きデータ値は、分析では切り捨てられます。また、指定された範囲外のカテゴリ値は無視されます。出力を最小化するには、変換メニューの [連続数への再割り当て] を使用して、1 で始まる連続したカテゴリを作成します。

また、オブジェクト スコア プロットのラベル付けに使用する各変数の範囲を定義する必要があります。しかし、その変数について定義された範囲外のデータ値を持つカテゴリのラベルもプロットに表示されます。

非線型正準相関分析: オプション

[オプション] ダイアログ ボックスで、オプションの統計とプロットを選択し、オブジェクト スコアを新しい変数としてアクティブなデータセット中に保存できます。また、反復および収束基準の指定、分析の初期設定の指定ができます。

図 4-5
[オプション] ダイアログ ボックス



表示。 使用できる統計として、周辺度数（出現数）、重心、反復の記述、重み付け、成分負荷、カテゴリ数量化、オブジェクトスコア、単一適合および多重適合があります。

- **重心 (Categories).** カテゴリ数量化スコア。変数の同じカテゴリに属する対象に対して、各集合に含まれる対象（ケース）に対する対象スコアの予想された平均と、実際の平均。
- **重み付きと成分負荷.** セットで数量化されたすべての変数に対する各次元の回帰係数。対象得点は数量化された変数に回帰され、数量化された変数の投影は対象空間に回帰されます。各セット内の次元に対する各変数の寄与を示します。
- **単一適合と多重適合 (OVERALS).** オブジェクトを基準とした単一カテゴリおよび多重カテゴリ座標/カテゴリ数量化の適合度。
- **カテゴリ数量化.** 最適得点の変数カテゴリへの割り当て。
- **オブジェクトスコア.** 特定次元でオブジェクト（ケース）に割り当てられた最適な得点。

作図。 カテゴリ軸、オブジェクトスコア、成分負荷、カテゴリ重心、変換の各プロットを作成できます。

オブジェクト スコアの保存。オブジェクト スコアを新しい変数としてアクティブなデータセットに保存できます。オブジェクト スコアは、メインダイアログ ボックスで指定された次元の数だけ保存されます。

任意初期布置の使用。任意の初期布置は、変数のすべてまたは一部が単一名義の場合に使用します。このオプションが選択されなかった場合には、入れ子にされた初期布置が使用されます。

基準。計算における非線型正準相関分析の最大反復数を指定できます。収束基準値の選択もできます。最後の 2 回の反復間の全概適合の差が収束値より小さくなるか、最大反復数に達した場合、分析の反復は終了します。

OVERALS コマンドの追加機能

選択内容をシンタックス ウィンドウに貼り付け、OVERALS コマンド シンタックスを編集することで、非線型正準相関分析をカスタマイズできます。コマンド シンタックスを使用すると、次の作業も実行できます。

- すべての抽出された次元についてプロットする代わりに、プロットする次元のペアを指定 (PLOT サブコマンドで NDIM キーワードを使用)。
- プロットにポイントをラベル付けするために使用する値ラベルの文字数を指定 (PLOT サブコマンドを使用)。
- オブジェクト スコア プロットのラベル付け変数として 6 つ以上の変数を指定 (PLOT サブコマンドを使用)。
- オブジェクト スコア プロットのためのラベル付け変数として分析で使用する変数を選択 (PLOT サブコマンドを使用)。
- 数量化スコア プロットにポイント ラベルを付与する変数を選択 (PLOT サブコマンドを使用)。
- アクティブなデータセット内のすべてのケースを分析に使用しない場合、分析に含めるケースの数を指定 (NOBSERVATIONS サブコマンドを使用)。
- オブジェクト スコアを保存することで作成される変数のルート名を指定 (SAVE サブコマンドを使用)。
- 抽出されたすべての次元を保存せずに、保存する次元の数を指定 (SAVE サブコマンドを使用)。
- カテゴリ数量化を行列ファイルに書き込む (MATRIX サブコマンドを使用)。
- 通常の高解像度プロットよりも低解像度のプロットが読みやすい場合に、低解像度プロットを作成 (SET コマンドを使用)。
- 指定された変数のみの重心および変換プロットを作成 (PLOT サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

コレスポンデンス分析

コレスポンデンス分析の目的の 1 つは、低次元空間にコレスポンデンステーブル内の 2 つの名義変数間の関係を記述し、同時に各変数のカテゴリ間の関係も記述することです。図中の個々の変数のカテゴリポイント間の距離は、カテゴリ間の関係を反映し、類似したカテゴリは互いに近い位置に配置されます。また、ある変数の各ポイントを、原点から他の変数のカテゴリポイントに至るベクトル上に投影することで、変数間の関係を記述できます。

分割表の分析には、行と列のプロファイルの検討や、カイ 2 乗統計量を使用した独立性の検定が含まれます。しかし、プロファイルは、非常に多数である場合があるため、カイ 2 乗検定では従属性が明らかになりません。クロス集計表の手続きでは、連関の測定方法および連関の検定がいくつか提供されますが、変数間の関係を図表で表すことはできません。

因子分析は、低次元空間における変数間の関係を記述する標準的な手法です。しかし、因子分析は間隔データを必要とし、観測値の数は変数の数の 5 倍でなければなりません。一方、コレスポンデンス分析は名義変数を仮定しており、変数間の関係だけでなく各変数のカテゴリ間の関係も記述します。また、コレスポンデンス分析は、正の対応関係を測定するテーブルの分析にも使用されます。

例。コレスポンデンス分析を社員のカテゴリと喫煙量の関係の図示に使用できます。たとえば、中間管理職の喫煙量は秘書とは差があるかもしれませんが、上級管理職の喫煙量は秘書と同じである場合があります。また、中間管理職は喫煙量が多く、秘書は喫煙量が少ないことがわかります。

統計量と作図。コレスポンデンスの測定方法、行と列のプロファイル、特異値、行と列の得点、慣性、質量、行得点と列得点の信頼統計量、特異値の信頼統計量、変換プロット、行ポイントプロット、列ポイントプロット、およびバイプロット。

データ。分析するカテゴリ変数は、名義尺度化されます。集計データまたは度数以外のコレスポンデンスの測定方法の場合は、正の類似性値の重み付け変数を使用します。一方、テーブルデータについては、シンタックスを使用してデータを読み込みます。

仮定。この手続きで使用する次元の数の最大値は、アクティブな行と列のカテゴリの数および同等性の制約条件に依存します。同等性の制約条件が使用されず、すべてのカテゴリがアクティブな場合、次元の最大数は、最も

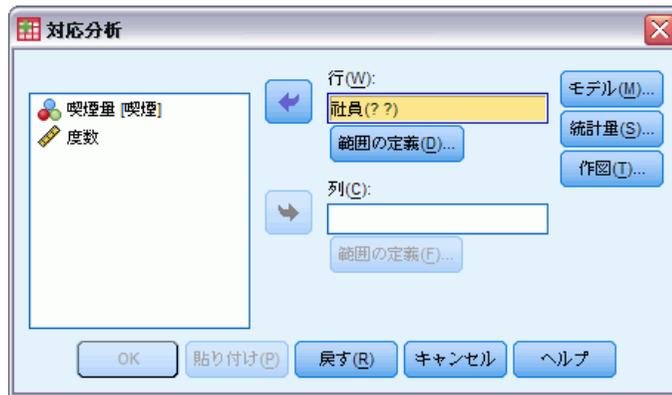
カテゴリの数の少ない変数のカテゴリ数より 1 少なくなります。たとえば、1 つの変数に 5 つのカテゴリがあり、他の変数に 4 つのカテゴリがある場合、最大の次元数は 3 です。補助カテゴリはアクティブではありません。たとえば、ある変数が 5 つのカテゴリを持ち、そのうちの 2 つが補助的で他の変数が 4 つのカテゴリを持つ場合、次元の最大数は 2 です。同等になるように制約されたすべてのカテゴリのグループは、1 つのカテゴリとして扱います。たとえば、ある変数が 5 つのカテゴリを持ち、そのうち 3 つが同等になるように正弦されている場合、次元の最大数を決定する際にその変数は、3 つのカテゴリを持つ変数として扱われなければなりません。カテゴリのうち 2 つは制約されておらず、第 3 のカテゴリが、制約された 3 つのカテゴリにあたります。次元の数の最大値よりも大きな数を指定した場合は、次元の数の最大値が使用されます。

関連手続き。 2 つ以上の変数が含まれる場合は、複数のコレスポネンス分析を使用します。変数を順序尺度化する必要がある場合は、カテゴリ主成分分析を使用します。

コレスポネンス分析を行うには

- ▶ メニューから次の項目を選択します。
[分析] > 次元縮小 > [コレスポネンス分析]

図 5-1
[コレスポネンス分析] ダイアログ ボックス

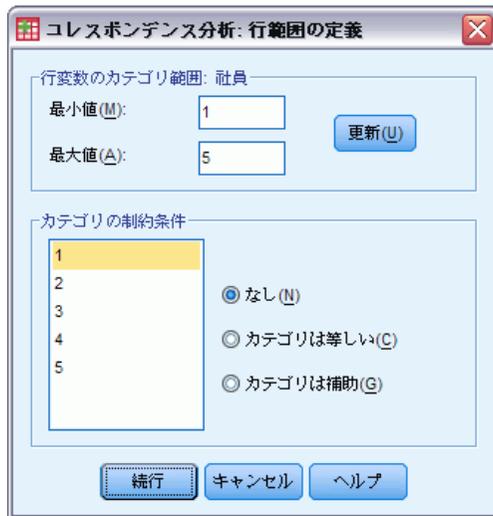


- ▶ 行変数を選択します。
- ▶ 列変数を選択します。
- ▶ 変数の範囲を定義します。
- ▶ [OK] をクリックします。

コレスポネンス分析: 行範囲の定義

行変数の範囲を定義する必要があります。指定する最小値および最大値は、整数でなければなりません。小数点付きデータ値は、分析では切り捨てられます。また、指定された範囲外のカテゴリ値は無視されます。

図 5-2
[行範囲の定義] ダイアログ ボックス



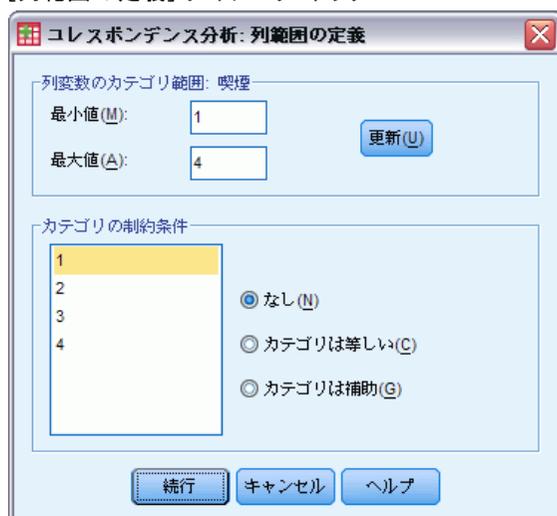
初期値では、すべてのカテゴリは制約されずアクティブです。行カテゴリを他の行カテゴリと同等として制約できます。また、行カテゴリを補助カテゴリとして定義することもできます。

- **カテゴリは等しい。** 各カテゴリの得点は同じでなければなりません。求められたカテゴリの順序が望ましくないか、直感に反したものである場合は、同等性の制約条件を使用します。同等であるとして制約された行カテゴリの最大数は、アクティブな行カテゴリの総数から 1 を引いたものです。異なる同等性の制約条件をいくつかのカテゴリグループに適用するには、シンタックスを使用します。たとえば、シンタックスを使用してカテゴリ 1 と 2 を同等とし、カテゴリ 3 と 4 を同等として制約します。
- **カテゴリは補助。** 補助的カテゴリは分析に影響しませんが、アクティブなカテゴリによって定義された範囲に表示されます。補助的カテゴリは、次元の定義には無関係です。補助的な行カテゴリの最大数は、行カテゴリの総数から 2 を引いたものです。

コレスポネンス分析: 列範囲の定義

列変数には、範囲を定義する必要があります。指定する最小値および最大値は、整数でなければなりません。小数点付きデータ値は、分析では切り捨てられます。また、指定された範囲外のカテゴリ値は無視されます。

図 5-3
[列範囲の定義] ダイアログ ボックス



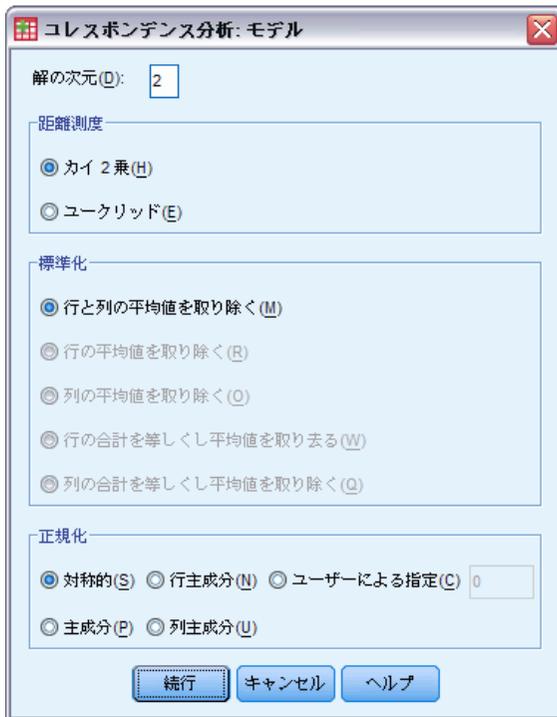
初期値では、すべてのカテゴリは制約されずアクティブです。ある列カテゴリを他の列カテゴリと同等であるとして制約できます。また、列カテゴリを補助カテゴリとして定義することもできます。

- **カテゴリは等しい。** 各カテゴリの得点は同じでなければなりません。求められたカテゴリの順序が望ましくないか、直感に反したものである場合は、同等性の制約条件を使用します。同等として制約された列カテゴリの最大数は、アクティブな列カテゴリの総数から 1 を引いたものです。異なる同等性の制約条件をいくつかのカテゴリ グループに適用するには、シンタックスを使用します。たとえば、シンタックスを使用してカテゴリ 1 と 2 を同等とし、カテゴリ 3 と 4 を同等として制約します。
- **カテゴリは補助。** 補助的なカテゴリは分析に影響しませんが、アクティブなカテゴリによって定義された範囲に表示されます。補助的なカテゴリは、次元の定義には無関係です。補助的な列カテゴリの最大数は、列カテゴリの総数から 2 を引いたものです。

コレスポネンス分析: モデル

[モデル] ダイアログ ボックスでは、次元の数、距離測度、標準化、正規化を指定できます。

図 5-4
[モデル] ダイアログ ボックス



[解の次元] 次元の数を指定します。一般に、変数の大半を説明するのに足りる最低限の次元を選択します。次元数の最大値は、分析で使用されるアクティブなカテゴリの数と同等性の制約条件に依存します。次元の数の最大値は、次の数よりも小さくなります。

- アクティブな行カテゴリ数から、同等として制約された行カテゴリ数を引いて、制約された行カテゴリのグループ数を足したもの。
- アクティブな列カテゴリ数から、同等として制約された列カテゴリ数を引いて、制約されたカテゴリのグループ数を足したもの。

距離尺度。 コレスポネンダンス テーブルの行間および列間の距離尺度を選択できます。次の選択肢の 1 つを選択します。

- **カイ 2 乗。** 重み付けプロファイルの距離を使用します。ただし、重みは、行または列のマス（質量）です。標準コレスポネンダンス分析では、この測定方法が必要とされます。
- **ユークリッド。** 行のペアと列のペアの差の平方和の平方根を使用します。

[標準化] 次の選択肢の 1 つを選択します。

- **行と列の平均値を取り除く。** 行と列が両方とも中心化されます。標準化コレスポネンダンス分析では、この方法を使用する必要があります。
- **[行の平均値を取り除く]** 行だけが中心化されます。

- **列の平均値を取り除く。**列だけが中心化されます。
- **行の合計を等しくし平均値を取り去る。**行を中心化する前に、行の周辺を同等にします。
- **[列の合計を等しくし平均値を取り除く]**列を中心化する前に、列の周辺を同等にします。

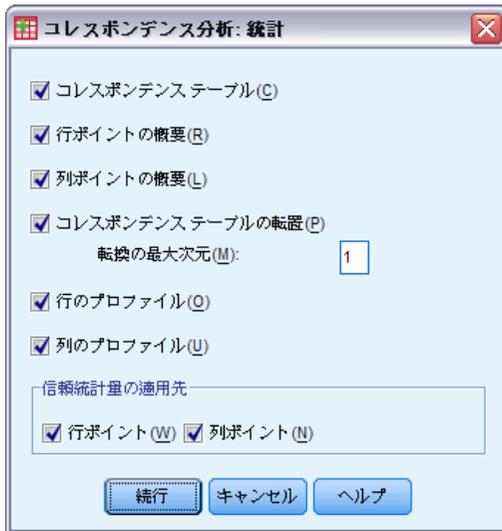
正規化。次の選択肢の1つを選択します。

- **[対称的]**各次元で、行の得点は、整合する特異値により割り当てられた列の得点の重み付き平均です。列の得点は、整合する特異値によって割り当てられた行の得点の重み付き平均です。2つの変数間の差または類似性を調べるには、この方法を使用します。
- **主成分。**行ポイントと列ポイントの距離は、選択された距離測度に従ったコレスポネンス テーブルでの距離にほぼ等しくなります。この方法は、2つの変数間の差の代わりにどちらかの変数または両方の変数のカテゴリ間の差を調べたい場合に使用します。
- **行主成分。**行ポイント間の距離は、選択された距離尺度に従ったコレスポネンス テーブルでの距離にほぼ等しくなります。行得点は、列得点の平均値を重み付けしたものです。行変数のカテゴリ間の差または類似度を調べたい場合に使用します。
- **列主成分。**列ポイント間の距離は、選択された距離尺度に従ったコレスポネンス テーブルでの距離にほぼ等しくなります。列得点は、行得点の平均値を重み付けしたものです。列変数のカテゴリ間の差または類似度を調べたい場合に使用します。
- **ユーザー指定。**-1 から 1 までの間の値を指定しなければなりません。値 -1 は列主成分に対応しています。1 は行主成分に対応します。0 は対称性による類似度に対応します。その他の値は、それぞれのイナーシャ（慣性）の程度に応じて、行および列得点全体にわたって広がります。この方法は、目的に応じたバイプロットの作成に役立ちます。

コレスポネンス分析: 統計

[統計] ダイアログ ボックスでは、数値出力を指定できます。

図 5-5
[統計] ダイアログ ボックス



コレスポネンステータブル。行と列の周辺合計付きの、入力変数のクロス集計表です。

行ポイントの概要。各行カテゴリについて、得点、マス（質量）、イナーシャ（慣性）、次元の慣性に対する寄与率、ポイントの慣性に対する次元の寄与率です。

列ポイントの概要。各列カテゴリについて、得点、マス（質量）、イナーシャ（慣性）、次元の慣性に対する寄与率、ポイントの慣性に対する次元の寄与率です。

行のプロファイル。各行のカテゴリについて、列変数のカテゴリに対する分布です。

列のプロファイル。各列のカテゴリについて、行変数のカテゴリに対する分布です。

コレスポネンステータブルの転置。行と列が第1次元上の得点について昇順となるように再割り当てされたコレスポネンステータブルです。必要に応じて、転置された表を作成する最大次元数を指定できます。1から指定された数までの各次元について転置された表が作成されます。

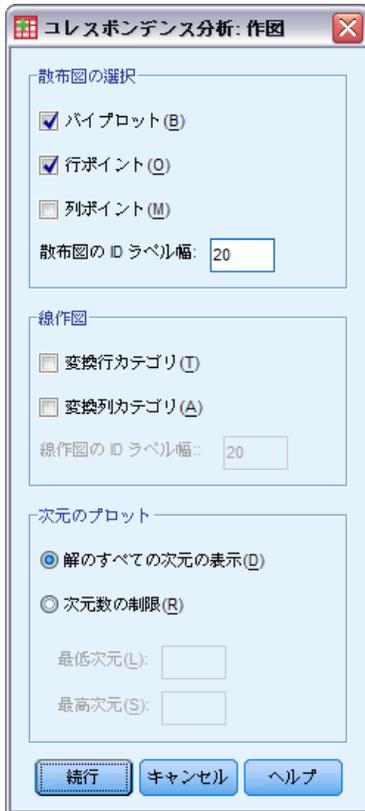
信頼統計量の適用先: 行ポイント。すべての非補助的な行ポイントについての標準偏差と相関を含みます。

信頼統計量の適用先: 列ポイント。すべての非補助的な列ポイントについての標準偏差と相関を含みます。

コレスポネンス分析: 作図

[作図] ダイアログ ボックスでは、作成する図を指定できます。

図 5-6
[作図] ダイアログ ボックス



[散布図] 次元のすべてのペアごとのプロットの行列を作成します。次の散布図が利用できます。

- **バイプロット**。行および列ポイントの結合プロットの行列を作成します。主成分の正規化が選択された場合、バイプロットは使用できません。
- **行ポイント**。行ポイントのプロットの行列を作成します。
- **列ポイント**。列ポイントのプロットの行列を作成します。

必要に応じて、ポイントをラベル付けするときに使用する値ラベルの文字数を指定できます。この値は、20 以下の負でない整数でなければなりません。

線作図。選択された変数の各次元のプロットを作成します。次の折れ線グラフができます。

- **変換行カテゴリ。** 対応する行得点に対する、もとの行カテゴリ値のプロットを作成します。
- **変換列カテゴリ。** 対応する列得点に対する、もとの列カテゴリ値のプロットを作成します。

必要に応じて、カテゴリ軸をラベル付けするときに使用する値ラベルの文字数を指定できます。この値は、20 以下の負でない整数でなければなりません。

次元のプロット。 出力表示する次元を制御できます。

- **【解のすべての次元の表示】** 解のすべての次元が散布図の行列に表示されます。
- **次元数の制限。** 表示する次元をプロットされたペアに制限します。次元を制限する場合は、プロットする最低次元と最高次元を選択する必要があります。最低次元を、1 から解の次元数より 1 小さい値の間で指定できます。また、最低次元は、それよりも高い次元に対してプロットされます。次元の最高値には、2 から解の次元数までを指定できます。この値は、次元のペアのプロットで使用される次元の中での最高次元を示します。この指定は、要求されたすべての多次元プロットに適用されます。

CORRESPONDENCE コマンドの追加機能

選択内容をシンタックス ウィンドウに貼り付け、CORRESPONDENCE コマンドシンタックスを編集することで、コレスポンデンス分析をカスタマイズできます。コマンドシンタックスを使用すると、次の作業も実行できます。

- ケースごとのデータを使用する代わりにテーブル データを入力データとして指定 (TABLE = ALL サブコマンドを使用)。
- 各種の散布図行列またはバイプロット行列の値ラベルの文字数を指定 (PLOT サブコマンドを使用)。
- 各種の線作図のポイントをラベル付けする値ラベルの文字数を指定 (PLOT サブコマンドを使用)。
- 行列データ ファイルに行得点および列得点の行列を書き込む (OUTFILE サブコマンドを使用)。
- 特異値および得点についての信頼統計 (分散および共分散) の行列を行列データ ファイルに書き込む (OUTFILE サブコマンドを使用)。
- 同等とみなすカテゴリの複数のグループを指定 (EQUAL サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

多重応答分析

多重応答分析では、数値をケース（オブジェクト）とカテゴリに割り当てることによって名義（カテゴリ）データを数量化します。この分析が適しているのは、同じカテゴリに属するオブジェクト同士の関係が近く、異なるカテゴリに属するオブジェクトは遠い関係にある場合です。各オブジェクトは、そのオブジェクトに適用されるカテゴリのカテゴリポイントに、できる限り近い位置に配置されます。この方法で、カテゴリはオブジェクトを等質なサブグループに分割します。変数は、同じカテゴリ内のオブジェクトを同じサブグループに分類するときに等質であると考えられます。

例 多重応答分析を使用して、職種、人種、および性別間の関係を視覚的に表示できます。それにより、人種と性別による特徴が示され、職種間には区別できるような特徴が見られないことに気付くかもしれません。また、ラテンアメリカ人とアフリカ系アメリカ人のカテゴリが互いに類似していることに気付くかもしれません。

統計量と作図。 オブジェクトスコア、判別測定、反復の記述、元の変数の相関と変換された変数の相関、カテゴリ数量化、記述統計、オブジェクトポイントのプロット、パイプロット、カテゴリプロット、結合カテゴリプロット、変換プロット、および判別測定プロット。

データ。 文字型変数値は、常に、文字の昇順で正の整数に変換されます。ユーザー定義の欠損値、システム欠損値、および1未満の値は、欠損値とみなされます。値が1未満である変数の場合は、値を再割り当てするか、または定数を加えることで非欠損値変数にできます。

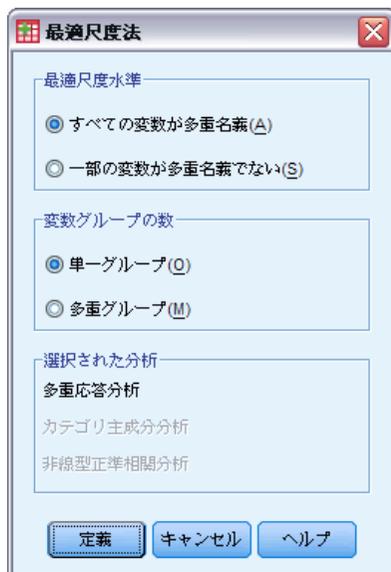
仮定。 すべての変数が多重名義尺度水準を持っています。データには、少なくとも3つの有効なケースが含まれている必要があります。この分析は、正の整数データに基づいています。離散化オプションでは、値を正規に近い分布を持つカテゴリに分けることにより小数値変数を自動的に分類します。さらに、文字型変数の値を正の整数に自動的に変換します。他の離散化方式を指定することもできます。

関連手続き。 変数が2つの場合、多重応答分析はコレスポンデンス分析に近くなります。変数に順序性または名義性があると考えられる場合は、カテゴリ主成分分析が使用されます。変数グループについて調べる場合は、非線形正準相関分析を使用します。

多重応答分析を行うには

- ▶ メニューから次の項目を選択します。
分析(A) > 次元分解 > 最適尺度法...

図 6-1
[最適尺度法] ダイアログ ボックス



- ▶ [すべての変数が多重名義] を選択します。
- ▶ [単一グループ] を選択します。
- ▶ [定義] をクリックします。

図 6-2
[多重応答分析] ダイアログ ボックス



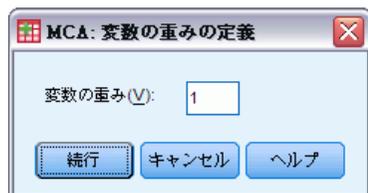
- ▶ 少なくとも 2 つの分析変数を選択し、解の次元の数を入力します。
- ▶ [OK] をクリックします。

必要に応じて、見つかった解に当てはめる補助変数、またはプロットのラベル付け変数を指定できます。

多重応答分析での変数の重みの定義

分析変数の重みを設定できます。

図 6-3
[変数の重みの定義] ダイアログ ボックス

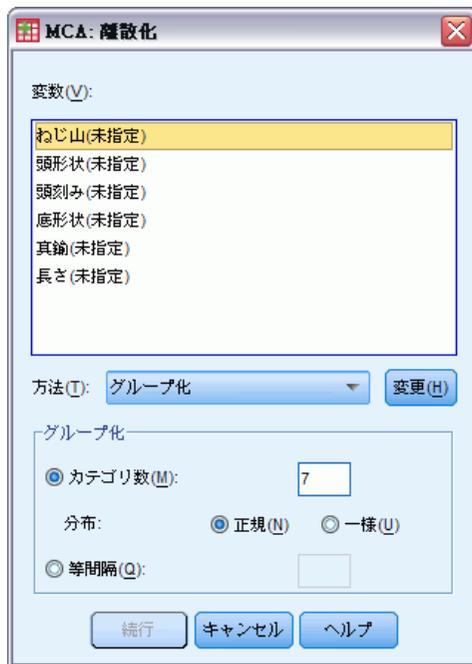


変数の重み。 各変数の重みを定義できます。指定する値は、正の整数でなければなりません。デフォルト値は 1 です。

多重応答分析での離散化

[離散化] ダイアログ ボックスでは、変数の再割り当てを行う方法を選択できます。小数値変数は、別の方法で指定されない限り、ほぼ正規分布する7つのカテゴリ（あるいはその変数の持つ値の種類の数¹が7未満の場合はその数）に分類されます。文字型変数は、文字の昇順でカテゴリ指標を割り当てることにより、常に正の整数に変換されます。文字型変数の離散化では、これらの整数を適用します。デフォルトでは、他の変数はそのままです。これらの離散化された変数を分析で使用します。

図 6-4
[離散化] ダイアログ ボックス



方法。 [グループ化]、[順位化]、[乗算] から選択します。

- **グループ化。** 指定されたカテゴリ数に再割り当てするか、または等間隔に再割り当てします。
- **順位化。** 変数は、ケースの順位化により離散化されます。
- **乗算。** 変数の現在の値を標準化し、10 を掛けて丸めたものに、離散化された値の最小値が 1 となるように定数を加えます。

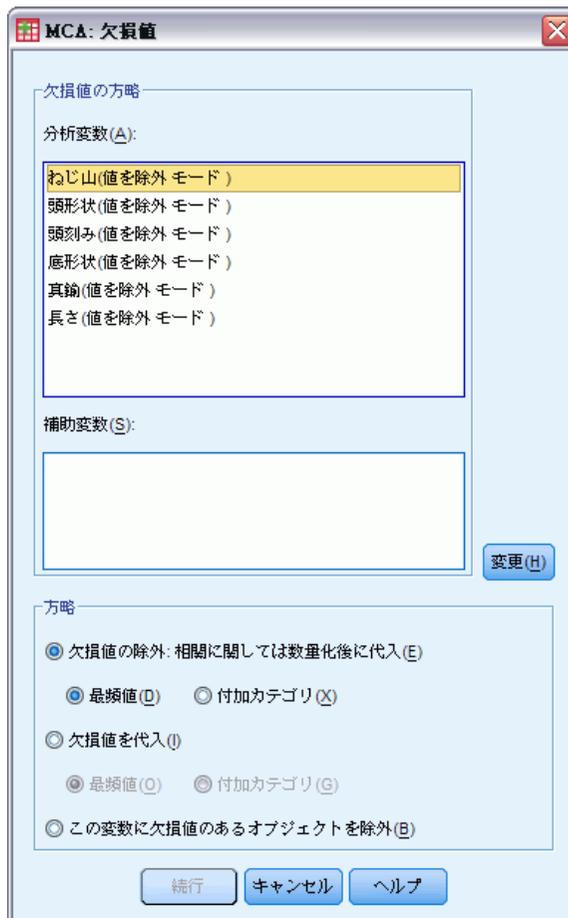
グループ化。 グループ別で変数を離散化するときは、次のオプションを使用できます。

- **カテゴリ数。** カテゴリ数と、変数の値がカテゴリ全体にわたってほぼ正規分布または一様分布するかどうかを指定します。
- **等間隔。** 変数は、等しい長さの間隔を持つカテゴリに再割り当てされます。間隔の長さを指定する必要があります。

多重応答分析での欠損値

[欠損値] ダイアログ ボックスでは、分析変数と補助変数の欠損値を処理するための方略を選択できます。

図 6-5
[欠損値] ダイアログ ボックス



欠損値の方略。 欠損値の除外（受動的な処理）、欠損値の代入（能動的な処理）、または欠損値のあるオブジェクトの除外（リストごとの削除）を選択します。

- **欠損値の除外: 相関に関しては数量化後に代入(E)** 選択された変数に欠損値のあるオブジェクトは、その変数の分析に関わりません。すべての変数に受動的な処理が指定されている場合、すべての変数に関して欠損値のあるオブジェクトは、補助変数として扱われます。[出力] ダイアログ ボックスで相関を指定した場合、元の変数の相関に関しては、分析後その変数の最頻カテゴリ、つまり最頻値が欠損値に代入されます。最適尺度変数の相関については、代入の方法を選択できます。欠損値を最適尺度変数の最頻値で置き換えるには、[最頻値] を選択します。欠損値を付加カテゴリの数量化で置き換えるには [付加カテゴリ] を選択します。これにより、その変数に欠損値のあるオブジェクトは、同じ (付加) カテゴリに属するとみなされます。
- **欠損値を代入(I)** 選択された変数に欠損値のあるオブジェクトでは、欠損値が代入されます。代入の方法を選択できます。欠損値を最頻カテゴリで置き換えるには、[最頻値] を選択します。最頻値が複数ある場合には、最小のカテゴリ指標を持つものを使用します。欠損値を同じ付加カテゴリの数量化で置き換えるには [付加カテゴリ] を選択します。これにより、その変数に欠損値のあるオブジェクトは、同じ (付加) カテゴリに属するとみなされます。
- **この変数に欠損値のあるオブジェクトを除外(B)** 選択された変数に欠損値のあるオブジェクトが分析から除外されます。この方略は、補助変数には使用できません。

多重応答分析のオプション

[オプション] ダイアログ ボックスでは、初期布置の選択、反復および収束基準の指定、正規化方法の選択、作図のラベル付け方法の選択、補助オブジェクトの指定ができます。

図 6-6
[オプション] ダイアログ ボックス

MCA: オプション

補助オブジェクト

ケースの範囲 (N)
 最初 (F):
 最後 (S):

単一ケース (I):

追加 (A)
 変更 (C)
 除去 (R)

正規化 (Z)

変数主成分
 ユーザー指定の値 (T):

基準

収束基準 (V):
 最大反復回数 (X):

作図のラベル

変数ラベルまたは値ラベル (R)
 ラベル長の限界 (L):
 変数名または値 (M)

次元のプロット

解のすべての次元の表示 (P)
 次元数の制限 (B)
 最低次元 (Q):
 最高次元 (D):

布置 (G)

なし

続行 キャンセル ヘルプ

補助オブジェクト。 補助オブジェクトにするオブジェクトのケース番号（またはオブジェクト範囲の先頭および最後のケース番号を指定）し、[追加] をクリックします。すべての補助オブジェクトを指定するまで、処理を繰り返します。オブジェクトが補助オブジェクトとして指定された場合、そのオブジェクトに対してケースの重み付けは無視されます。

正規化。 オブジェクトスコアおよび変数の正規化を行う 5 つのオプションのうち、1 つを指定できます。1 つの分析で使用できる正規化の方法は 1 つだけです。

- **変数主成分。** このオプションを使用すると、変数間の関連が最適化されます。オブジェクト空間内の変数座標は、成分負荷（次元やオブジェクトスコアなどと主成分との相関）です。変数間の相関に主に関心がある場合に役立ちます。

- **オブジェクト主成分。**このオプションを使用すると、オブジェクト間の距離が最適化されます。オブジェクト間の相違または類似性に主に関心がある場合に役立ちます。
- **対称的。**オブジェクトと変数との関係に主に関心がある場合は、この正規化オプションを使用します。
- **独立。**オブジェクト間の距離、および変数間の相関を別々に調べる場合は、この正規化オプションを使用します。
- **ユーザー指定。**閉区間 $[-1, 1]$ 内の任意の実数値を指定できます。値 1 は「オブジェクト主成分」方法に相当し、値 0 は「対称的」方法に相当し、値 -1 は「変数主成分」方法に相当します。-1 よりも大きく 1 よりも小さい値を指定すると、固有値をオブジェクトと変数の両方に広げることができます。この方法は、目的に応じたバイプロットまたはトリプロットの作成に役立ちます。

基準。計算における手続きの最大反復数を指定できます。収束基準値の選択もできます。アルゴリズムの最後の 2 回の反復間における全体適合の差が収束値より小さくなるか、最大反復数に達した場合、アルゴリズムの反復は終了します。

作図のラベル。変数ラベルと値ラベル、または変数名と値のどちらを作図で使用するかを指定できます。さらに、ラベルの最大長を指定することもできます。

次元のプロット。出力表示する次元を制御できます。

- **[解のすべての次元の表示]** 解のすべての次元が散布図の行列に表示されます。
- **次元数の制限。**表示する次元をプロットされたペアに制限します。次元を制限する場合は、プロットする最低次元と最高次元を選択する必要があります。最低次元を、1 から解の次元数より 1 小さい値の間で指定できます。また、最低次元は、それよりも高い次元に対してプロットされます。次元の最高値には、2 から解の次元数までを指定できます。この値は、次元のペアのプロットで使用される次元の中での最高次元を示します。この指定は、要求されたすべての多次元プロットに適用されます。

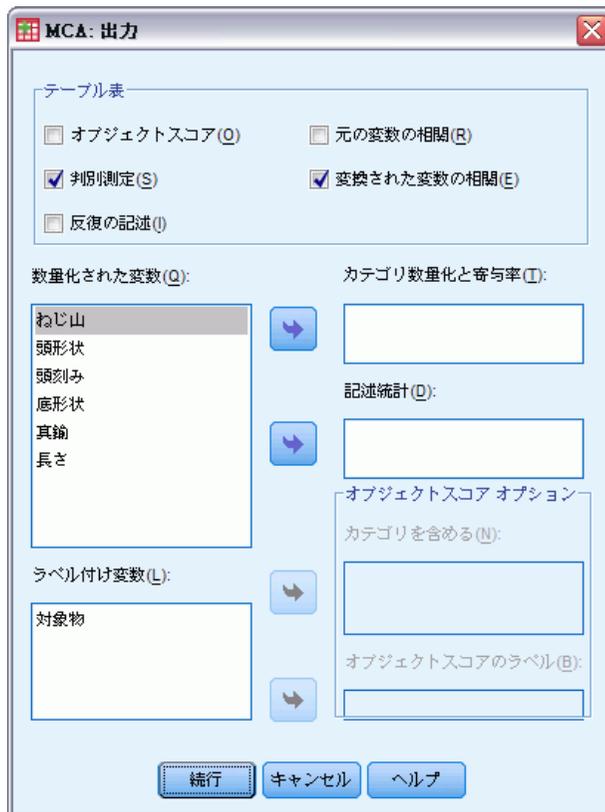
布置。布置の座標が記述されているファイルからデータを読み込むことができます。ファイルの最初の変数は最初の次元の座標を含み、2 番目の変数は 2 番目の次元の座標を含みます（以下同様です）。

- **初期。**指定されたファイル内の布置を分析の開始点として使用します。
- **固定。**指定されたファイル内の布置を変数への当てはめに使用します。当てはめられる変数は、分析変数として選択しなければなりません。布置は固定されているため、補助変数として扱われます（したがって、補助変数として選択する必要はありません）。

多重応答分析での出力

[出力] ダイアログ ボックスでは、オブジェクト スコア、判別測定、反復の記述、元の変数の相関と変換された変数の相関、選択された変数のカテゴリ数量化、および選択された変数の記述統計を表したテーブルを出力できます。

図 6-7
[出力] ダイアログ ボックス



オブジェクトスコア。マス（質量）、イナーシャ（慣性）、寄与率を含むオブジェクト スコアを表示します。次のオプションが用意されています。

- **カテゴリを含める。** 選択された分析変数のカテゴリ指標を表示します。
- **オブジェクト スコアのラベル。** ラベル付け変数として指定されている変数のリストから、オブジェクトにラベル付けする変数を選択できます。

判別測定。 変数および次数ごとに判別測定を表示します。

反復の記述。 反復ごとに、説明された分散、損失、および説明された分散の増加量が表示されます。

元の変数の相関。 元の変数の相関行列とその行列の固有値を表示します。

変換された変数の相関。変換（最適尺度）変数の相関行列とその行列の固有値を表示します。

カテゴリ数量化と寄与率。選択された変数の次元ごとに、マス（質量）、イナーシャ（慣性）、寄与率を含むカテゴリ数量化（座標）を表示します。

注：質量やイナーシャを含む座標および寄与率が、ピボットテーブル出力の各層に表示されます。座標はデフォルトで表示されます。寄与率を表示するには、テーブルをダブルクリックしてアクティブにし、[層] ドロップダウンリストから [寄与率] を選択します。

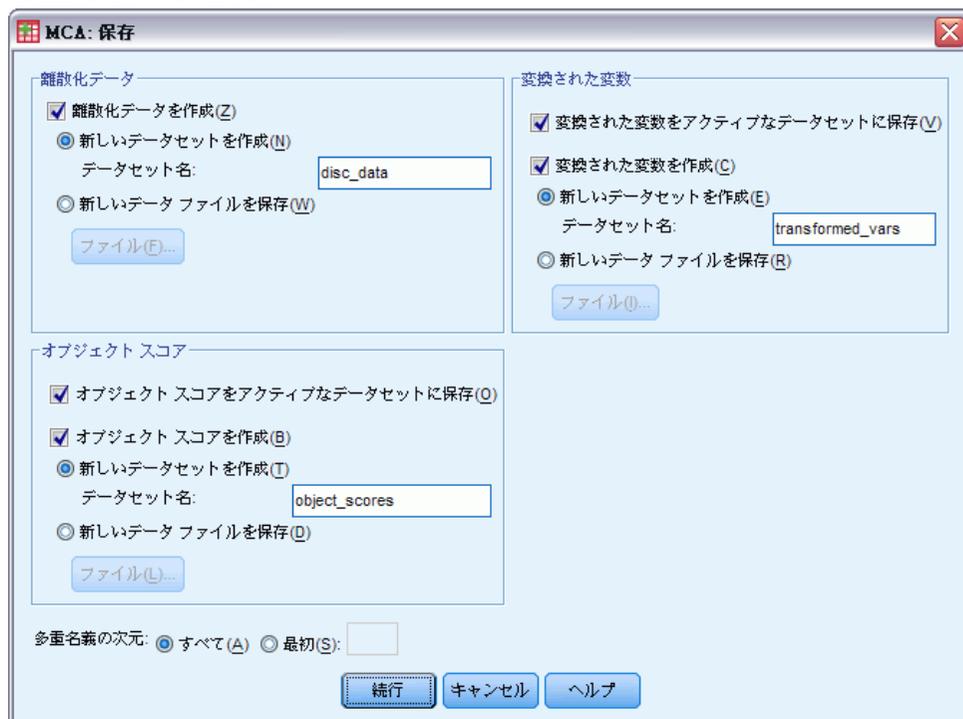
記述統計。選択された変数の度数、欠損値の数、および最頻値を表示します。

多重応答分析での保存

[保存] ダイアログ ボックスでは、離散化データ、オブジェクトスコア、および変換された変数を、現在のセッションの外部 IBM® SPSS® Statistics データファイルやデータセットに保存できます。変換された変数およびオブジェクトスコアも、アクティブなデータセットに保存することができます。

- データセットが利用可能なのは現在のセッション中ですが、今後のセッションで利用するには、データセットをデータ ファイルとして明示的に保存する必要があります。データセット名は、変数命名規則に従う必要があります。
- 保存されるデータの各タイプに応じて、ファイル名やデータセット名は異なる必要があります。
- オブジェクトスコアや変換された値をアクティブなデータセットに保存する場合、多重名義尺度水準の番号を指定することができます。

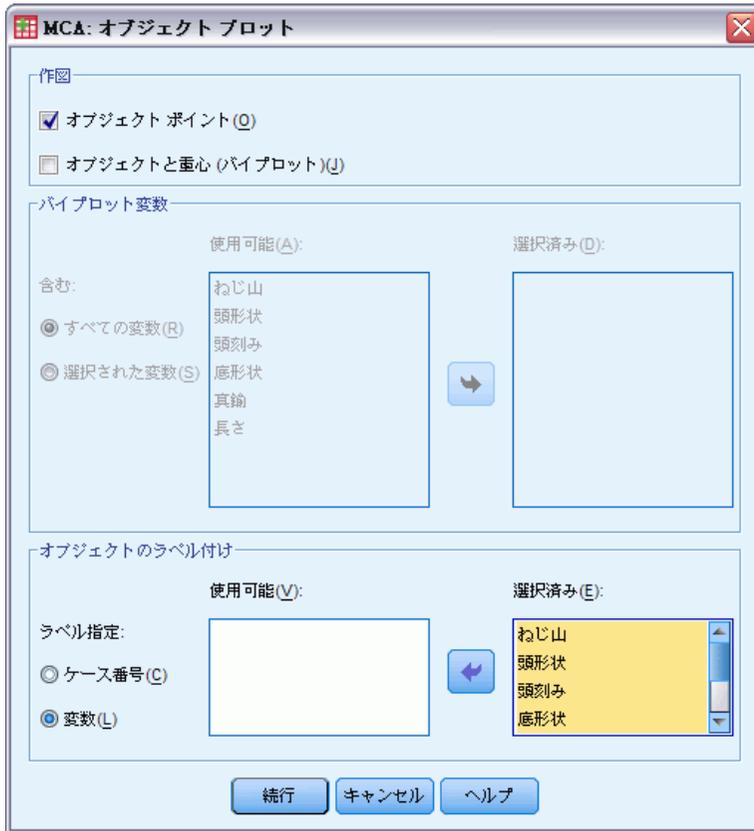
図 6-8
[保存] ダイアログ ボックス



多重応答分析でのオブジェクト プロット

[オブジェクト プロット] ダイアログ ボックスで、使用するプロットの種類とプロットする変数を指定できます。

図 6-9
[オブジェクト プロット] ダイアログ ボックス



オブジェクト ポイント。オブジェクト ポイントのプロットが表示されます。

オブジェクトと重心 (パイプロット)。オブジェクト プロットが変数の重心とともにプロットされます。

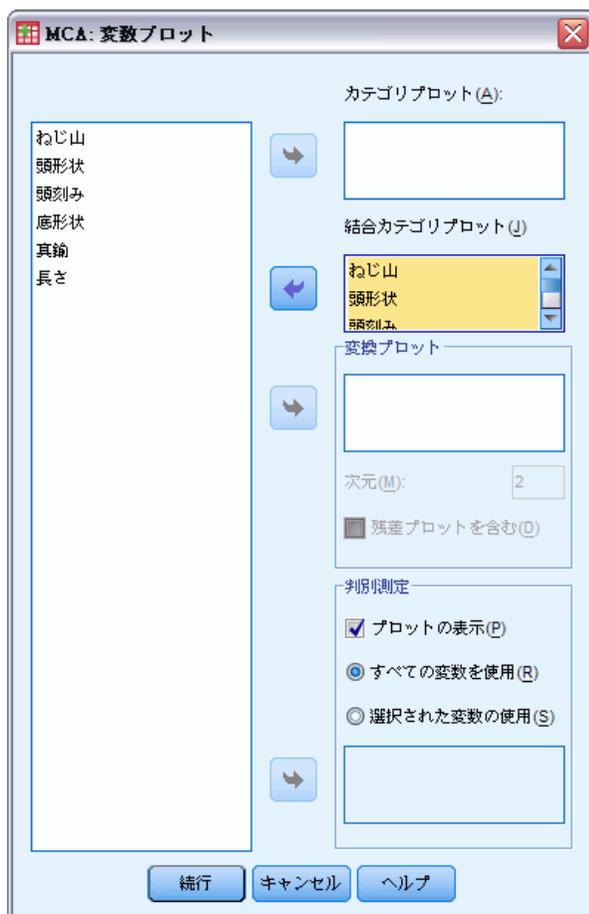
パイプロット変数。パイプロット用のすべての変数を使用するか、サブグループを選択できます。

オブジェクトのラベル付け。選択された変数のカテゴリ (カテゴリ指標値か値ラベルかを [オプション] ダイアログ ボックスで選択できます) またはケース番号でオブジェクトにラベル付けできます。[変数] が指定されている場合は、変数ごとにプロットが 1 つ出力されます。

多重応答分析での変数プロット

[変数プロット] ダイアログ ボックスで、使用するプロットの種類とプロットする変数を指定できます。

図 6-10
[変数プロット] ダイアログ ボックス



カテゴリプロット。 選択した変数ごとに、重心の座標のプロットが作図されます。カテゴリは、そのカテゴリに属するオブジェクトの重心に表示されます。

結合カテゴリプロット。 これは、選択した各変数の重心の座標を 1 つのプロットに示したものです。

変換プロット。 カテゴリ指標と対比して最適カテゴリ数量化のプロットを表示します。使用する次元数を指定できます。次元ごとに 1 つのプロットが作成されます。選択された変数ごとに残差プロットを表示するようにもできます。

判別測定。 選択された変数の判別測定を 1 つのプロットに作図します。

MULTIPLE CORRESPONDENCE コマンドの追加機能

選択内容をシンタックス ウィンドウに貼り付け、MULTIPLE CORRESPONDENCE コマンド シンタックスを編集することで、多重応答分析をカスタマイズできます。コマンド シンタックスを使用すると、次の作業も実行できます。

- 変換された変数、オブジェクト スコア、および近似値をアクティブなデータセットに保存するときに、変数のルート名を指定 (SAVE サブコマンドを使用)。
- 各プロットのラベルの最大長を別々に指定 (PLOT サブコマンドを使用)。
- 残差プロットの個別変数リストを指定 (PLOT サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

多次元尺度法 (PROXSCAL)

多次元尺度法は、オブジェクト間の距離を表す一群の測度の中に構造を見つける手法です。この手法では、観測値を概念上の低次元空間における特定の場所に割り当てます。この場合、空間における各ポイント間の距離が、指定した類似度（または非類似度）にできるだけ近くなるようにします。その結果得られるのが、その低次元空間におけるオブジェクトの最小2乗表現であり、多くの場合、データのより深い理解のために役立ちます。

例。多次元尺度法は、認知的な関係を調べる場合に非常に役立ちます。たとえば、製品のイメージを検討するときに調査を実施し、ライバル会社の製品と自社製品がどの程度の類似性（または近接性）を持つと認知されているかを示すデータセットを集めます。この近接と外部変数（価格など）を使用すると、製品に対する大衆の見方にとってどの変数が重要なかを判断することが可能となり、その結果に応じて自社製品のイメージを調整できることとなります。

統計量と作図。反復の記述、ストレス測定、ストレス分解、共通空間の座標、最終布置におけるオブジェクト距離、個別空間重み、個別空間、変換された近接、変換された独立変数、ストレスプロット、共通空間の散布図、個別空間の重み散布図、個別空間の散布図、変換プロット、Shepard 残差プロット、および独立変数の変換プロット。

データ。データは、近接行列や近接行列に変換される変数の形で入力できます。行列は列内、または列にまたがる形でフォーマットできます。近接は、比率、間隔、順序、またはスプラインの尺度法レベルで扱うことができます。

仮定。3 つ以上の変数を指定する必要があります。次元の数は、対象の数から 1 を引いた値以下です。多重ランダム スタートと組み合せた場合、次元数の縮小は省略されます。ソースを 1 つだけ指定した場合、モデルはすべて同一モデルに等しいため、分析のデフォルト モデルは同一モデルとなります。

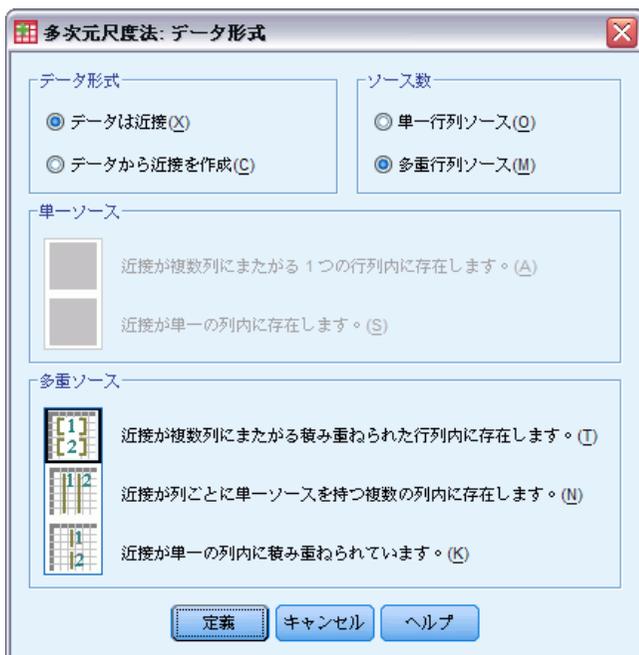
関連手続き。すべての変数を数値レベルで尺度化した場合は、標準の多次元尺度法に相等します。

多次元尺度法を使用するには

- ▶ メニューから次の項目を選択します。
分析 > 尺度 > 多次元尺度法 (PROXSCAL)...

これで、[データ形式] ダイアログ ボックスが開きます。

図 7-1
[データ形式] ダイアログ ボックス



- ▶ 次の項目でデータ形式を指定します。

データ形式。使用するデータが近接測定値から成り立っているか、あるいはデータから近接を作成するかを指定します。

ソース数。データが近接である場合は、近接の測定値が単一ソースか、または多重ソースかを指定します。

単一ソース。近接のソースが1つある場合には、データセットの形式がいくつかの列にまたがる1つの行列の近接なのか、それとも各近接の行と列を識別するための2つの異なる変数を含む単一の近接なのかを指定します。

- **近接が複数列にまたがる1つの行列内に存在する。**近接行列は、オブジェクトの数と同じ数の列にまたがっています。これによって複数列行列の近接ダイアログ ボックスが表れます。
- **近接が単一の列内に存在する。**近接行列は、1つの列、つまり変数に集約されています。各セルの行と列を識別した2つの追加変数が必要になります。これによって単一の近接ダイアログ ボックスが表れます。

多重ソース。複数の近接ソースがある場合、データセットは積み重ねられた行列の近接ですが、その形式がいくつかの列にまたがるのか、列ごとに単一ソースを持つ複数列なのか、あるいは単一系列なのかを指定します。

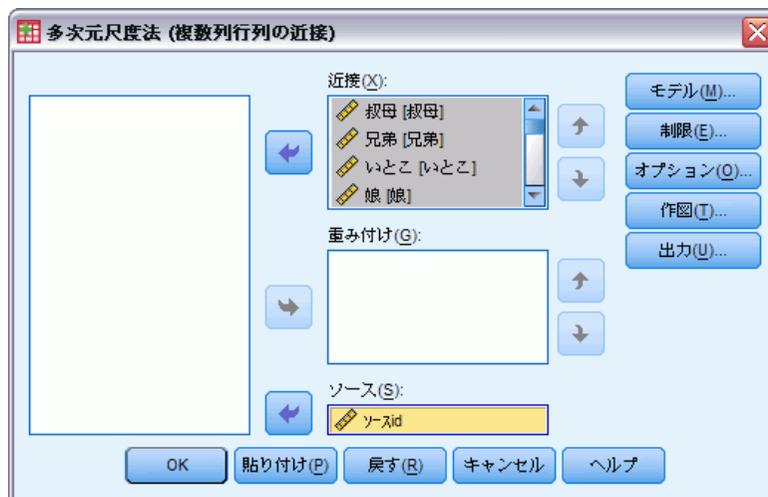
- **近接が複数列にまたがる積み重ねられた行列内に存在する。**近接行列は、オブジェクトの数と同じ数の列にまたがっていて、オブジェクト数にソース数を掛けたものと同じ数の行にわたって積み重なっています。これによって複数列行列の近接ダイアログボックスが表れます。
- **近接が列ごとに単一ソースを持つ複数の列内に存在する。**近接行列は、複数の列、つまり変数に集約されています。各セルの行と列を識別した2つの追加変数が必要になります。これによって列の近接ダイアログボックスが表れます。
- **近接が単一の列内に積み重ねられている。**近接行列は、1つの列、つまり変数に集約されています。各セルの列、カラムおよびソースを識別した3つの追加変数が必要になります。これによって単一系列の近接ダイアログボックスが表れます。

▶ [定義] をクリックします。

複数列行列の近接

[データ形式] ダイアログボックスで単一ソースまたは多重ソースのいずれかに対し、行列内の近接モデルを選択した場合は、次のようなメインダイアログボックスが表示されます。

図 7-2
[複数列行列の近接] ダイアログボックス



▶ 3つ以上の近接変数を選択します。(リストの変数の順序は、近接の列の順序に一致するようにしてください。)

- ▶ 必要に応じて、近接変数の個数に等しい数の重み付け変数を選択できます。(もう一度、重みの順序を、重み付けする近接の順序に一致するようにします。)
- ▶ さらに、多重ソースがある場合には、ソース変数を選択します。(各近接変数内のケースの数は、近接変数の数にソース数をかけた値に等しくなければなりません。)

さらに、多次元尺度法のモデルを定義し、共通空間に対し制限を設け、収束基準を設定することができます。また、使用する初期布置を指定し、作図と出力を選択できます。

複数列の近接

[データ形式] ダイアログ ボックスで多重ソースに対し複数列モデルを選択した場合は、次のようなメイン ダイアログ ボックスが表示されます。

図 7-3
[複数列の近接] ダイアログ ボックス



- ▶ 複数の近接変数を選択します。(それぞれの変数は、異なるソースからの近接の行列であると仮定します。)
- ▶ 行変数を選択し、各近接変数内の近接度の行位置を定義します。
- ▶ 列変数を選択し、各近接変数内の近接度の列位置を定義します。(行と列の定義が指定されていない近接行列のセルは、欠損値と見なされます)。
- ▶ 必要に応じて、近接変数の個数に等しい数の重み付け変数を選択できます。

さらに、多次元尺度法のモデルを定義し、共通空間に対し制限を設け、収束基準を設定することができます。また、使用する初期布置を指定し、作図と出力を選択できます。

単一系列の近接

[データ形式] ダイアログ ボックスで、単一ソースまたは多重ソースのいずれかに対して単一系列モデルを選択した場合は、次のようなメイン ダイアログ ボックスが表示されます。

図 7-4
[単一系列の近接] ダイアログ ボックス



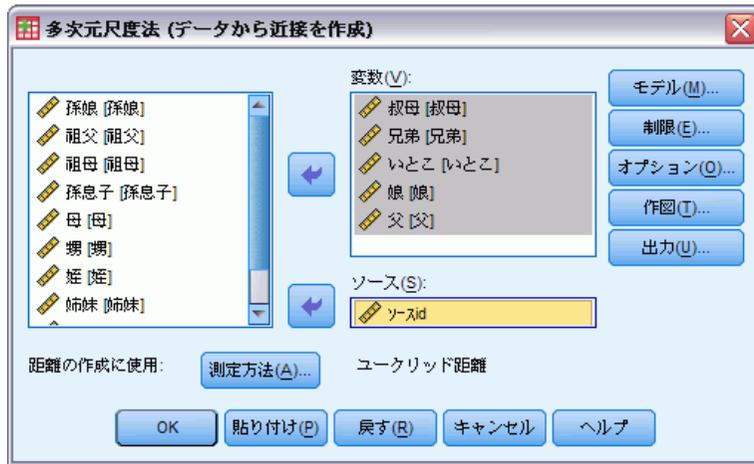
- ▶ 近接変数を選択します。(1 つ以上の近接行列であると仮定します。)
- ▶ 行変数を選択し、近接変数内の近接度の行位置を定義します。
- ▶ 列変数を選択し、近接変数内の近接度の列位置を定義します。
- ▶ 多重ソースがある場合には、ソース変数を選択します。(各ソースについて、行と列の定義がされていない近接行列のセルは、欠損値と見なされます)。
- ▶ 必要に応じて、重み付け変数を選択できます。

さらに、多次元尺度法のモデルを定義し、共通空間に対し制限を設け、収束基準を設定することができます。また、使用する初期布置を指定し、作図と出力を選択できます。

データから近接の作成

[データ形式] ダイアログ ボックスでデータから近接を作成する場合は、次のようなメイン ダイアログ ボックスが表示されます。

図 7-5
[データから近接を作成] ダイアログ ボックス



- ▶ 変数間の距離を作成する場合は、少なくとも 3 つの変数を選択します ([データからの尺度の作成] ダイアログ ボックスを参照)。これらの変数を使用して、1 つの近接行列 (または、ソースが複数ある場合は複数の行列) を作成します。ケース間の距離を作成する場合は、変数が 1 つだけ必要です。
- ▶ 多重ソースがある場合には、ソース変数を選択します。
- ▶ 必要に応じて、近接を作成する際の尺度を選択できます。

さらに、多次元尺度法のモデルを定義し、共通空間に対し制限を設け、収束基準を設定することができます。また、使用する初期布置を指定し、作図と出力を選択できます。

データからの尺度の作成

図 7-6
[データからの尺度の作成] ダイアログ ボックス



多次元尺度法では、非類似度データを使用して尺度解を作成します。データが多変量データ（測定された複数の変数の値）の場合は、非類似度データを作成し、多次元尺度法の解を計算します。データから非類似度の測度を作成する方法を詳細に指定できます。

測定方法。分析で使用する非類似度の測定方法を指定できます。データの種類に応じて [測定方法] のオプションを 1 つクリックします。選択したオプションのドロップダウン リストから測定方法を 1 つ選択します。使用可能な選択肢は次のとおりです。

- **間隔。**[ユークリッド距離]、[平方ユークリッド距離]、[Chebychev]、[都市ブロック]、[Minkowski]、または [カスタマイズ]。
- **度数。**[カイ 2 乗測度] または [ファイ 2 乗測度]。
- **2 値。**[ユークリッド距離]、[平方ユークリッド距離]、[サイズの差異]、[パターンの差異]、[分散]、または [Lance と Williams]。

距離行列の作成。分析の単位を選択します。[変数間] または [ケース間] をクリックします。

値の変換。たとえば、変数がまったく異なるスケールで測定されている場合などで、値を標準化してから近接を計算します (2 値データには適用できません)。[標準化] ドロップダウン リストから標準化の方法を選択します (標準化を行わない場合は、[なし] を選択します)。

多次元尺度法のモデルの定義

[モデル] ダイアログ ボックスでは、尺度モデル、次元の最大数および最小数、近接行列の構造、近接に対して使用する最適な変換、および近接を各ソース内で別個に変換するか、あるいはソース上で無条件に変換するかを指定できます。

図 7-7
[モデル] ダイアログ ボックス



尺度モデル。 次の選択肢の中から選択します。

- **同一。** ソースはすべて同じ布置をとります。
- **重み付きユークリッド。** このモデルは個人差モデルです。それぞれのソースは、共通空間のすべての次元の重みがそれぞれ異なるような個別空間を持ちます。
- **一般化ユークリッド。** このモデルは個人差モデルです。各ソースは、次元の異なる重み付けによる共通空間の回転に等しい個別空間を持っています。
- **縮小階数。** このモデルは、個別空間の階数を指定できる一般化ユークリッドモデルです。1 以上最大次元数未満の階数を指定する必要があります。

形。近接を近接行列の下三角か上三角のどちらから抽出されるかを指定します。完全行列の使用を指定できます。その場合は、近接行列の下三角と上三角の重みの合計が分析されます。いずれの場合も、指定された部分だけが使用されても、完全行列と対角成分が指定される必要があります。

近接。近接行列の測度が類似度なのか非類似度なのかを指定します。

近接変換。次の選択肢の中から選択します。

- **比。**変換された近接は、元の近接に比例します。これは、正の値を持つ近接にだけ使用できます。
- **間隔。**変換された近接は、元の近接と切片項を加えたものに比例します。切片により、変換されたすべての近接が正であることが保証されます。
- **順序。**変換された近接は元の近接と同じ順序です。同一順序の近接を結合したままにするか、同一でなくてもよいかを指定します。
- **スプライン。**変換された近接は、元の近接の滑らかな非減少区分的多項式変換です。多項式の次数と内側ノットの個数を指定します。

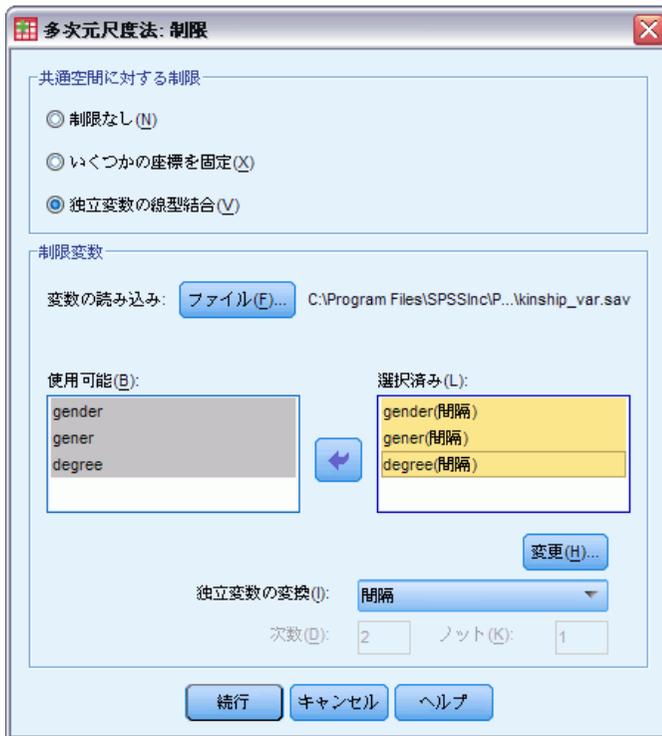
変換の適用。各ソース内の近接だけを互いに比較するのか、ソースを無条件に比較するのかを指定します。

次元。デフォルトでは、2次元で解を計算します（最小値 = 2、最大値 = 2）。整数の最小値および最大値として、1 からオブジェクトの個数 -1 までの範囲の値を選択します。ただし、最小値は最大値以下とします。分析手続きは、最大次元の解を計算し、それからいくつかのステップに分けて最低値に達するまで次元数を減らします。

多次元尺度法の制限

[制限] ダイアログ ボックスでは、共通空間への制限を設定します。

図 7-8
[制限] ダイアログ ボックス



共通空間に対する制限。使用する制限の種類を選択します。

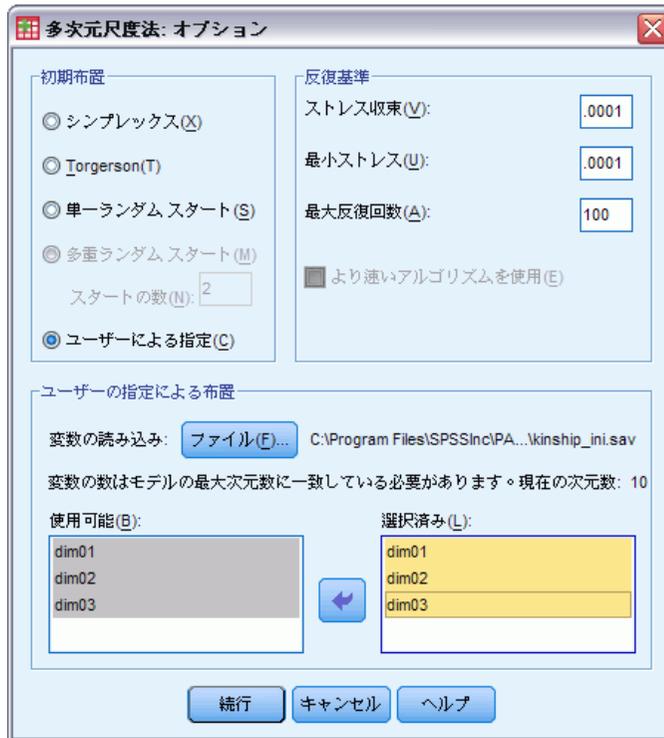
- **制限なし。**共通空間に制限を設けません。
- **いくつかの座標を固定。**選択された1番目の変数には1番目の次元のオブジェクトの座標、2番目の変数には2番目の次元のオブジェクトの座標、というようにデータが含まれていなければなりません。欠損値の場合、各次元の座標は自由となります。選択された変数の個数は、要求した次元の最大数に等しくなければなりません。
- **独立変数の線型結合。**共通空間は、選択された変数の線型結合であるように制限されます。

制限変数。共通空間に対する制限を定義する変数を選択します。線型結合を指定した場合は、制限変数に対して間隔、名義、順序、またはスプライン変換を指定します。いずれの場合も、各変数のケース数はオブジェクトの個数に等しくなければなりません。

多次元尺度法のオプション

[オプション] ダイアログ ボックスでは、初期布置のスタイルの選択、反復および収束基準の指定、標準の更新か緩和された更新の選択ができます。

図 7-9
[オプション] ダイアログ ボックス



初期布置。 次の選択肢の 1 つを選択します。

- **シンプレックス。** 各オブジェクトは、互いに等距離となる位置に置かれます。反復を 1 回行ってこの高次元の布置を改善し、その後次元数縮小の操作を行って、[モデル] ダイアログ ボックスで指定した最大次元数を持つ初期布置を得ます。
- **Torgerson。** 古典的な尺度解が初期布置として使用されます。
- **単一ランダム スタート。** 1 つの布置が無作為に選択されます。
- **多重ランダム スタート。** 複数の布置が無作為に選択され、最低の正規化された原ストレスを持つものが初期布置として使用されます。
- **ユーザー指定。** ユーザーが指定した初期布置の座標を含む変数を選択できます。最初の変数は次元 1 の座標に相当し、2 番目の変数は次元 2 の座標に相当するというように、選択された変数の数は、指定された次元の最大数に等しくなければなりません。各変数のケースの数は、オブジェクトの個数に等しくなければなりません。

反復基準。 反復基準値を指定します。

- **ストレス収束。** このアルゴリズムは、引き続き正規化原ストレス値間の差が指定した数値よりも小さくなったときに反復を停止します。この数値は、0.0 から 1.0 までの範囲になければなりません。

- **最小ストレス。**このアルゴリズムは、正規化された原ストレスがここで指定した数値よりも小さくなったときに反復を停止します。この数値は、0.0 から 1.0 までの範囲になければなりません。
- **最大反復回数。**このアルゴリズムは、上記の基準がどれも満たされない場合に、指定した回数まで反復を実行します。
- **より速いアルゴリズムを使用。**より速いアルゴリズムの使用を選択すると、アルゴリズムが高速化されます。同一モデル以外のモデルの場合、あるいは制限のある場合には使用できません。

多次元尺度法の作図、バージョン 1

[作図] ダイアログ ボックスでは、作成するプロットを指定できます。[複数列の近接] データ形式を使用すると、次の [作図] ダイアログ ボックスが表示されます。[個別空間の重み]、[変換前と変換後の近接]、および [変換後の近接と距離] の各プロットに対してそのソースを指定します。使用可能なソースのリストは、メイン ダイアログ ボックスの近接変数のリストです。

図 7-10
[作図] ダイアログ ボックス、バージョン 1



ストレス。正規化された原ストレスと次元を対比するプロットを出力します。このプロットは、最大次元数が最小次元数よりも大きい場合にだけ出力されます。

共通空間。共通空間の座標の行列散布図が表示されます。

個別空間。ソースごとに、個別空間の座標が行列散布図内に表示されます。これは、[モデル] ダイアログ ボックスで、個人差モデルの 1 つを指定した場合にだけ使用できます。

個別空間の重み。個別空間の重みの散布図が出力されます。これは、[モデル] ダイアログ ボックスで、個人差モデルの 1 つを指定した場合にだけ使用できます。重み付きユークリッドモデルでは、重みがプロット内に表示され、各軸に 1 つの次元が割り当てられます。一般化ユークリッドモデルでは、次元ごとに 1 つのプロットが出力され、その次元の回転と重みの両方が示されます。縮小階数モデルでは、一般化ユークリッドモデルと同じプロットを出力されますが、個別空間における次元の数は減らされます。

変換前と変換後の近接。元の近接と変換後の近接を対比したプロットが作成されます。

変換後の近接と距離。変換後の近接と距離を対比したプロットが作成されます。

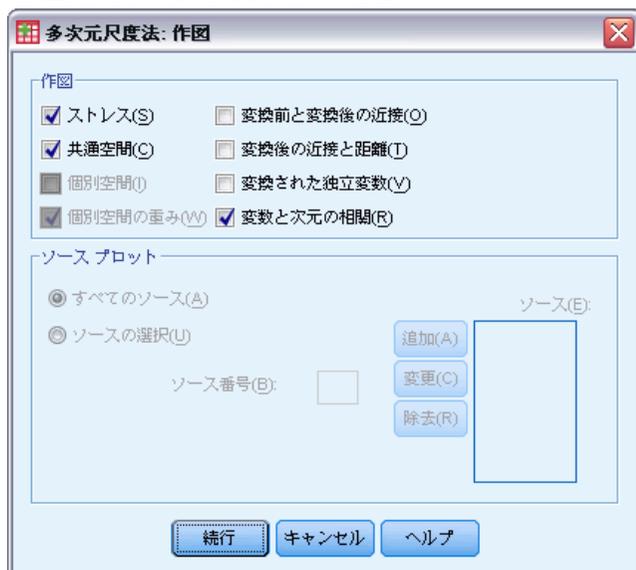
変換された独立変数。独立変数の変換プロットが出力されます。

変数と次元の相関。独立変数と共通空間の次元との相関のプロットが表示されます。

多次元尺度法の作図、バージョン 2

[作図] ダイアログ ボックスでは、作成するプロットを指定できます。「複数列の近接」以外のデータ形式では、次の [作図] ダイアログ ボックスが表示されます。[個別空間の重み]、[変換前と変換後の近接]、および [変換後の近接と距離] の各プロットに対してそのソースを指定します。入力するソース番号は、メイン ダイアログ ボックスで指定したソース変数の値で、1 からソース数の範囲内でなければなりません。

図 7-11
[作図] ダイアログ ボックス、バージョン 2



多次元尺度法の出力

[出力] ダイアログ ボックスでは、表示する出力の量を制御し、出力の一部を別のファイルに保存できます。

図 7-12
[出力] ダイアログ ボックス



表示。表示する項目を、次の中から 1 つ以上選択します。

- **共通空間の座標。**共通空間の座標を表示します。
- **個別空間の座標。**個別空間の座標は、モデルが同一モデルでない場合にだけ表示されます。
- **個別空間の重み。**個人差モデルのうちの 1 つを指定している場合にだけ、個別空間の重みを表示します。モデルによっては、空間の重みは回転重みと次元重みに分解され、これらも表示されます。
- **距離。**布置の中のオブジェクト間の距離を表示します。
- **変換された近接。**布置の中のオブジェクト間の変換された近接を表示します。
- **入力データ。**元の近接を含み、もし存在していれば、データの重み、初期設定、および独立変数の固定座標を含みます。
- **ランダム スタートのストレス。**ランダム シード数と各ランダム スタートのストレス値を表示します。
- **反復の記述。**メイン アルゴリズムの反復の履歴を表示します。

- **多重ストレス測定。**さまざまなストレス値を表示します。テーブルには、正規化された原ストレス、ストレス-I、ストレス-II、S-ストレス、説明された散らばり (DAF)、Tucker の適合係数が記述されます。
- **ストレス分解。**オブジェクトごとの平均とソースごとの平均を含む、オブジェクトとソースの最終正規化された原ストレスのストレス分解が表示されます。
- **変換された独立変数。**線型結合の制限を選択した場合は、変換された独立変数および対応する回帰重みが表示されます。
- **変数と次元の相関。**線型結合の制限を選択した場合は、独立変数と共通空間の次元との相関が表示されます。

新しいファイルに保存。共通空間の座標、個別空間の重み、距離、変換された近接、および変換された独立変数をそれぞれ別の IBM® SPSS® Statistics データ ファイルに保存できます。

PROXSCAL コマンドの追加機能

選択内容をシンタックス ウィンドウに貼り付け、PROXSCAL コマンドシンタックスを編集することで、多次元尺度法分析をカスタマイズできます。コマンドシンタックスを使用すると、次の作業も実行できます。

- 変換プロットと残差プロットの個別変数リストを指定 (PLOT サブコマンドを使用)。
- 個別空間の重みのプロット、変換プロット、および残差プロットの個別リストを指定 (PLOT サブコマンドを使用)。
- 表示する独立変数の変換プロットのサブグループを指定 (PLOT サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

多次元展開 (PREFSCAL)

[多次元展開] 手続きは共通の量的尺度を見つけるもので、2つのオブジェクト集合の関係を視覚的に調べることができるようになります。

例。21人に好みの朝食を15項目、好みの順番にランク付けするように依頼したとします。多次元展開を使用すると、個人が朝食の項目を2つの基本的な方法（柔らかいパンと硬いパン、および脂肪のある項目と脂肪のない項目）で判別していることがわかります。

また、ドライバのグループに対し、自動車26モデルの10の属性について6点スケール（1 = 「まったくそう思わない」から6 = 「本当にそう思う」）で採点するように依頼したとします。個人間の平均を取ったため、値は類似度として見なされます。多次元展開を使用すると、類似モデルのクラスター、および最も密接に関連しあう属性のクラスターが見つかります。

統計量と作図。多次元展開手続きで作成できるのは、反復の記述、ストレス測度、ストレス分解、共通空間の座標、最終布置におけるオブジェクト距離、個別空間重み、個別空間、変換後の近接、ストレスプロット、共通空間の散布図、個別空間の重み散布図、個別空間の散布図、変換プロット、および Shepard 残差プロットです。

データ。データは、長方近接行列の形式で入力できます。各列は、個別の列オブジェクトとして見なされます。近接行列の各行は、個別の行オブジェクトとして見なされます。複数の近接ソースがある場合、行列は積み重ねられます。

仮定。2つ以上の変数を指定する必要があります。解の次元数は、オブジェクトの数から1を引いた値以下でなければなりません。ソースを1つだけ指定した場合、モデルはすべて同一モデルに等しいため、分析のデフォルトモデルは同一モデルとなります。

多次元展開を使用するには

- ▶ メニューから次の項目を選択します。
分析 > 尺度 > 多次元尺度法 (PREFSCAL)...

図 8-1
[多次元展開] メイン ダイアログ ボックス



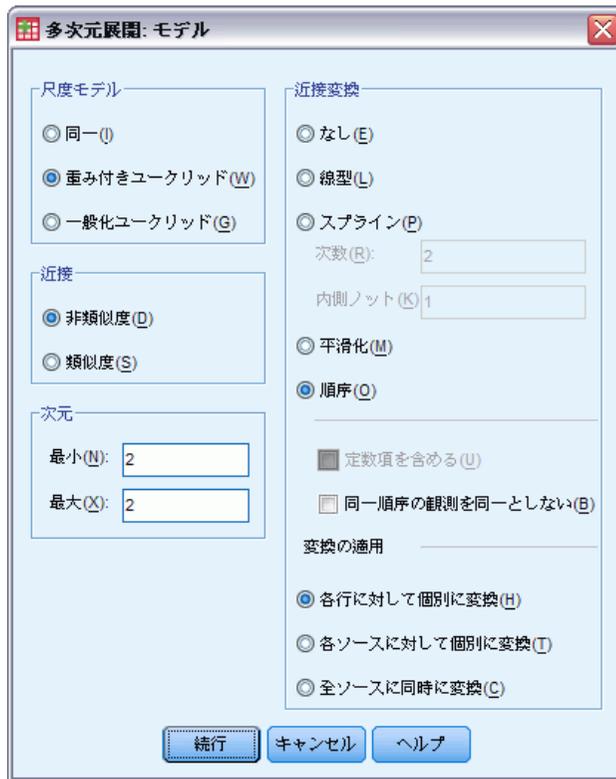
- ▶ 長方近接行列にある列を識別する、2 個以上の変数を選択します。各変数は、個別の列オブジェクトを表します。
- ▶ 必要に応じて、列オブジェクト変数の個数に等しい数の重み付け変数を選択できます。重み付け変数の順番は、重み付けの対象となる列オブジェクトの順番と一致している必要があります。
- ▶ 必要に応じて、行変数を選択します。この変数の値（または値ラベル）は、出力の行オブジェクトのラベル付けに使用されます。
- ▶ 多重ソースがある場合には、必要に応じてソース変数を選択します。データ ファイル内のケースの数は、行オブジェクトの数にソース数を掛けた値に等しくなければなりません。

さらに、多次元展開のモデルを定義し、共通空間に対し制限を設け、収束基準を設定することができます。また、使用する初期布置を指定し、作図と出力を選択できます。

多次元展開のモデルの定義

[モデル] ダイアログ ボックスでは、尺度モデル、次元の最大数および最小数、近接行列の構造、近接に対して使用する最適な変換、および近接の変換方法（行を条件とする変換、ソースを条件とする変換、またはソースとは無関係に変換）を指定できます。

図 8-2
[モデル] ダイアログ ボックス



尺度モデル。 次の選択肢の中から選択します。

- **同一。** ソースはすべて同じ布置をとります。
- **重み付きユークリッド。** このモデルは個人差モデルです。それぞれのソースは、共通空間のすべての次元の重みがそれぞれ異なるような個別空間を持ちます。
- **一般化ユークリッド。** このモデルは個人差モデルです。各ソースは、次元の異なる重み付けによる共通空間の回転に等しい個別空間を持っています。

近接。 近接行列の測度が類似度なのか非類似度なのかを指定します。

次元。 デフォルトでは、2次元で解を計算します（最小値 = 2、最大値 = 2）。整数の最小値および最大値として、1 からオブジェクトの個数 -1 までの範囲の値を選択できます。ただし、最小値は最大値以下とします。分析手続きは、最大次元の解を計算し、それからいくつかのステップに分けて最低値に達するまで次元数を減らします。

近接変換。 次の選択肢の中から選択します。

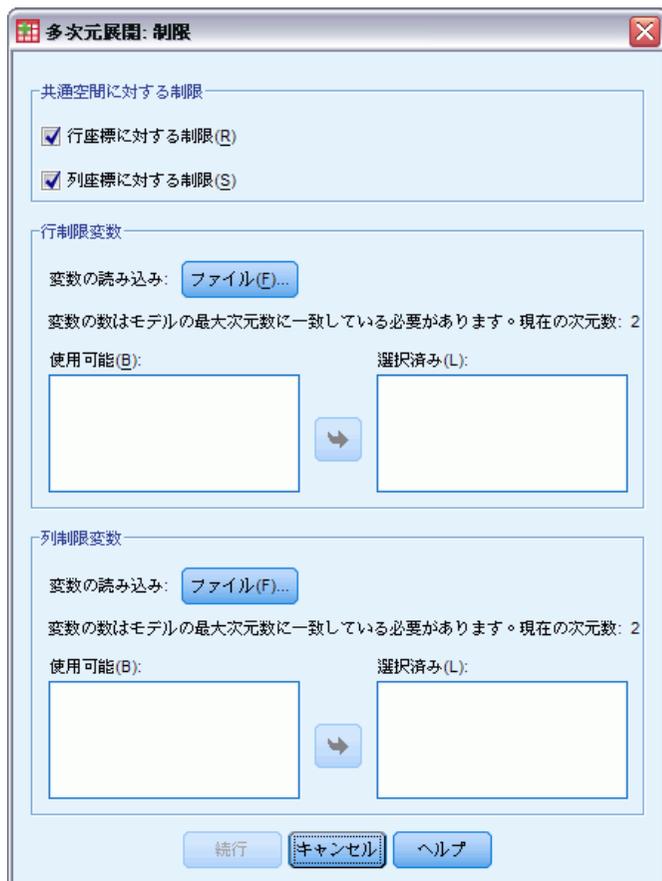
- **なし**。近接は変換されません。必要に応じて、[定数項を含める]を選択できます。この場合、近接を定数項分だけずらすことができます。
- **1次**。変換後の近接は元の近接に比例します。つまり、変換関数により傾きが計算され、切片は0に固定されます。この変換のことを、比率変換とも呼びます。必要に応じて、[定数項を含める]を選択できます。この場合、近接を定数項分だけずらすこともできます。この変換のことを、間隔変換とも呼びます。
- **スプライン**。変換された近接は、元の近接の滑らかな非減少区分的多項式変換です。多項式の次数と内側ノットの個数を指定できます。必要に応じて、[定数項を含める]を選択できます。この場合、近接を定数項分だけずらすこともできます。
- **平滑化**。変換後の近接は元の近接と同じ順番を取りますが、次の値との差を考慮に入れるよう制限されます。結果的に、「滑らかな順序」変換になります。同一順序の近接を結合したままにするか、同一でなくてもよいかを指定できます。
- **順序**。変換された近接は元の近接と同じ順序です。同一順序の近接を結合したままにするか、同一でなくてもよいかを指定できます。

変換の適用。近接同士の比較を、各行内でのみ行うか、各ソース内でのみ行うか、または行やソースなどの条件を付けずに行うかを指定します。つまり、変換の実行対象を、列ごとにするか、ソースごとにするか、またはすべての近接とするかを指定します。

多次元展開の制限

[制限] ダイアログ ボックスでは、共通空間への制限を設定します。

図 8-3
[制限] ダイアログ ボックス



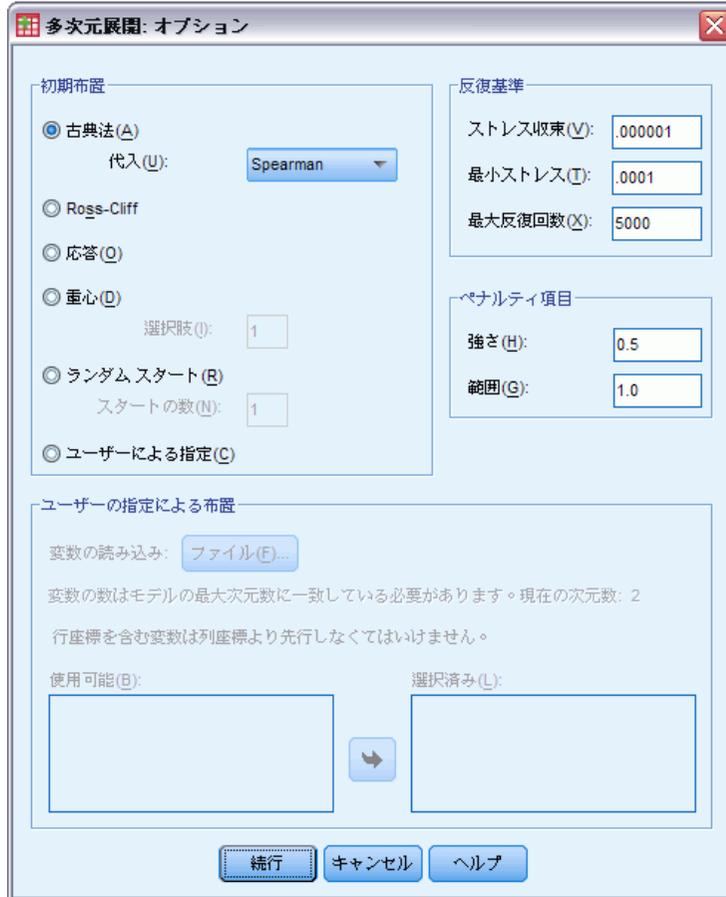
共通空間に対する制限。共通空間内にある行オブジェクトまたは列オブジェクト、あるいはその両方の座標を固定するよう選択できます。

行/列制限変数。制限を含むファイルを選択してから、共通空間での制限を定義する変数を選択します。選択された 1 番目の変数には 1 番目の次元のオブジェクトの座標、2 番目の変数には 2 番目の次元のオブジェクトの座標、というようにデータが含まれていなければなりません。欠損値の場合、各次元の座標は自由となります。選択された変数の個数は、要求した次元の最大数に等しくなければなりません。各変数のケースの数は、オブジェクトの個数に等しくなければなりません。

多次元展開のオプション

[オプション] ダイアログ ボックスでは、初期布置のスタイルの選択、反復および収束基準の指定、およびストレス値のペナルティ項目の設定を行うことができます。

図 8-4
[オプション] ダイアログ ボックス



初期布置。次の選択肢の1つを選択します。

- **古典法。**長方近接行列を使用して、完全対称 MDS 行列の内部ブロック（行と行、および列と列の間の値）を補います。完全行列を作成してから、古典法による尺度解が初期布置として使用されます。内部ブロックは、三角不等式や Spearman 距離を使用した代入によって埋められます。
- **Ross-Cliff。**Ross-Cliff スタートでは、行オブジェクトおよび列オブジェクトの初期値として二重中心化および平方化された近接行列上の特異値分解の結果を使用します。

- **応答。**コレスポンデンス スタートでは、行スコアおよび列スコアの対称の正規化と逆データ（非類似度ではなく類似度）のコレスポンデンス分析の結果を使用します。
- **重心。**この手続きは、固有値分解を使用した布置に行オブジェクトを配置して開始します。次に、列オブジェクトが、指定した選択の重心に配置されます。選択肢の数だけ、1 から近接変数の個数までの正整数を指定します。
- **多重ランダム スタート。**ランダムに選択された複数の初期布置に対して解を計算し、ペナルティ付きストレスが最も小さい解を最適解として示します。
- **ユーザー指定。**ユーザーが指定した初期布置の座標を含む変数を選択できます。最初の変数は次元 1 の座標に相当し、2 番目の変数は次元 2 の座標に相当するというように、選択された変数の数は、指定された次元の最大数に等しくなければなりません。各変数のケースの数は、行オブジェクトと列オブジェクトの個数の和に等しくなければなりません。行および列座標は、行座標の後に列座標が積み重ねられます。

反復基準。反復基準値を指定します。

- **ストレス収束。**このアルゴリズムは、連続するペナルティ付きストレス値間の相対差が、指定した数値よりも小さくなったときに反復を停止します。この数値は、負以外の値である必要があります。
- **最小ストレス。**このアルゴリズムは、ペナルティ付きストレスがここで指定した値よりも小さくなったときに反復を停止します。この数値は、負以外の値である必要があります。
- **最大反復回数。**このアルゴリズムは、上記の基準がどれも満たされない場合に、ここで指定した回数まで反復を実行します。

ペナルティ項目。このアルゴリズムは、ペナルティ付きストレスを最小にしようとしています。ペナルティ付きストレスとは、Kruskal ストレス 1 と、変換後の近接の変動係数に基づくペナルティ項目との積に等しい適合度です。これらのコントロールにより、ペナルティ項目の強度と範囲を設定することができます。

- **強さ。**強度パラメータの値を小さくするほど、ペナルティが強くなります。0.0 から 1.0 までの値を指定してください。
- **範囲。**このパラメータは、ペナルティがアクティブになるタイミングを設定します。0.0 に設定すると、ペナルティは非アクティブになります。値を大きくすると、アルゴリズムは、変換後の近接においてより大きな変動を持つ解を探索します。負でない値を指定してください。

多次元展開の作図

[作図] ダイアログ ボックスでは、作成するプロットを指定できます。

図 8-5
[作図] ダイアログ ボックス



作図。次の作図が利用できます。

- **多重開始値。**ペナルティ付きストレスの積み上げヒストグラムを表示します。ストレスとペナルティの両方が表示されます。
- **初期共通空間。**初期共通空間の座標の散布図行列を表示します。
- **次元ごとのストレス。**ペナルティ付きストレスと次元を対比する折れ線グラフが出力されます。このプロットは、最大次元数が最小次元数よりも大きい場合にだけ出力されます。
- **最終共通空間。**共通空間の座標の行列散布図が表示されます。

- **空間の重み。**個別空間の重みの散布図が出力されます。これは、[モデル] ダイアログ ボックスで、個人差モデルの 1 つを指定した場合にだけ使用できます。重み付きユークリッド モデルでは、全ソースの重みがプロット内に表示され、各軸に 1 つの次元が割り当てられます。一般化ユークリッド モデルでは、次元ごとに 1 つのプロットが出力され、各ソースに対する次元の回転と重みの両方が示されます。
- **個別空間。**各ソースの個別空間の座標の散布図行列が表示されます。これは、[モデル] ダイアログ ボックスで、個人差モデルの 1 つを指定した場合にだけ使用できます。
- **変換図。**元の近接と変換後の近接を対比する散布図が出力されます。変換の適用方法に応じて、行またはソースごとに別々の色が割り当てられます。無条件変換の場合は、単一色が使用されます。
- **Shepard 図。**元の近接と変換後の近接および距離とを対比した図です。距離は点で表示され、変換後の近接は線で表示されます。変換の適用方法に応じて、行またはソースごとに別々の線が作成されます。無条件変換の場合は、単一の線が作成されます。
- **適合度の散布図。**変換後の近接と距離を対比する散布図が表示されます。複数のソースを指定した場合、ソースごとに別々の色が割り当てられます。
- **残差図。**変換後の近接と残差の対比（変換後の近接から距離を引いたもの）の散布図が表示されます。複数のソースを指定した場合、ソースごとに別々の色が割り当てられます。

行オブジェクト スタイル。プロット中の行オブジェクトの表示をより細かく制御できます。オプションの色変数の値を使用することで、すべての色に変更することができます。オプションのマーカー変数の値を使用することで、すべての使用可能なマーカーに変更することができます。

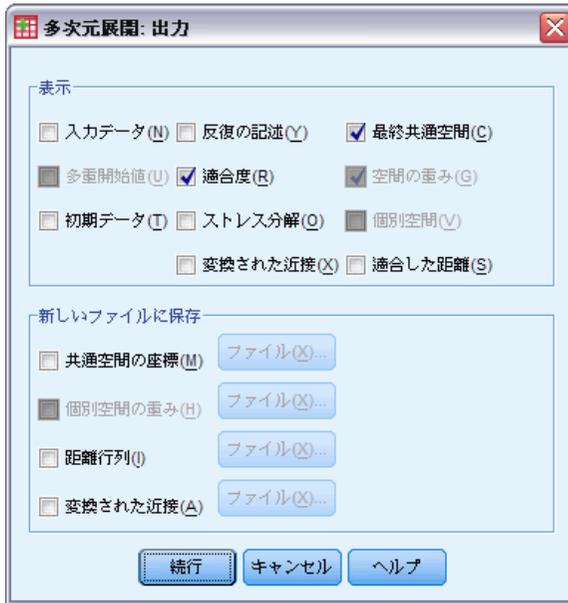
ソース プロット。[個別空間]、[適合度の散布図]、および [残差図]（変換がソースで適用される場合、[変換図] および [Shepard 図]）では、プロットの作成対象とするソースを指定できます。入力するソース番号は、メイン ダイアログ ボックスで指定したソース変数の値で、1 からソース数の範囲内であればなりません。

行の作図。変換が行で適用される場合、[変換図] および [Shepard 図] では、プロットの作成対象とする行を指定できます。入力する行番号は、1 から行の個数までの範囲内であればなりません。

多次元展開の出力

[出力] ダイアログ ボックスでは、表示する出力の量を制御し、出力の一部を別のファイルに保存できます。

図 8-6
[出力] ダイアログ ボックス



表示。表示するものを、次の中から 1 つ以上選択します。

- **入力データ。**元の近接を含み、もし存在していれば、データの重み、初期布置、および固定座標を含みます。
- **多重開始値。**ランダム シード数と各ランダム スタートのペナルティ付きストレス値を表示します。
- **初期データ。**初期共通空間の座標を表示します。
- **反復の記述。**メイン アルゴリズムの反復の履歴を表示します。
- **適合度。**異なる層を表示する。テーブルには、複数の適合度、不適合度、相関、変動、および非縮退の度数が含まれます。
- **ストレス分解。**オブジェクト、行、およびソースのペナルティ付きストレス分解を表示します。行、列、およびソースの平均と標準偏差が含まれます。
- **変換された近接。**[変換された近接] メニューが表示されています。
- **最終共通空間。**共通空間の座標を表示します。
- **空間の重み。**個別空間の重み。このオプションが使用可能となるのは、個別の異なるモデルの 1 つを指定した場合のみです。モデルによっては、空間の重みは回転重みと次元重みに分解され、これらも表示されます。
- **個別空間。**個別空間の座標が表示されます。このオプションが使用可能となるのは、個別の異なるモデルの 1 つを指定した場合のみです。
- **適合した距離。**布置の中のオブジェクト間の距離を表示します。

新しいファイルに保存。 共通空間の座標、個別空間の重み、距離、変換された近接をそれぞれ別の IBM® SPSS® Statistics データ ファイルに保存できます。

PREFSCAL コマンドの追加機能

選択内容をシンタックス ウィンドウに貼り付け、PREFSCAL コマンドシンタックスを編集することで、多次元展開分析をカスタマイズできます。コマンドシンタックスを使用すると、次の作業も実行できます。

- 複数ソースが利用可能である場合、個別空間、適合度の分布図、および残差図（行列の条件変換の場合は変換図および Shepard 図）用に複数のソース リストを指定 (PLOT サブコマンドを使用)。
- 行条件変換の場合、変換図および Shepard 図用に複数の行リストを指定 (PLOT サブコマンドを使用)。
- 列 ID 変数の代わりに列番号を指定 (INPUT サブコマンドを使用)。
- ソース ID 変数の代わりにソース番号を指定 (INPUT サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

パート II: 例

カテゴリ回帰

最適尺度法を使用したカテゴリ回帰分析の目的は、応答変数と一連の予測変数間の関係を記述することです。この関係を数量化することにより、任意の予測変数の組み合わせに対して応答の値を予測できます。

この章では、最適尺度法による回帰で行われる分析が2つの例で説明されています。1つ目の例では、小規模なデータセットが使用され、基本概念が説明されています。2つ目の例では、より大規模な一連の変数と観測値が実際の例で使用されます。

例:カーペット専用洗剤のデータ

一般的な例 (Green および Wind, 1973) としては、新しいカーペット専用洗剤を市販することに関心のある企業が消費者の嗜好に関する5種類の因子(パッケージのデザイン、ブランド名、価格、サービスシール、および料金の払い戻し)の影響について調べたい場合があります。パッケージのデザインには、3つの因子レベルがあります。それぞれ塗布用ブラシの位置が異なります。また、3つのブランド名 (K2R、Glory、および Bissell)、3つの価格水準があり、最後の2つの因子のそれぞれに対しては2つのレベル(「なし」または「あり」)があります。次の表に、カーペット専用洗剤に関する分析で使用される変数を、その変数ラベルと変数の値とともに示します。

テーブル 9-1
カーペット専用洗剤に関する分析の説明変数

変数名	変数ラベル(L)	値ラベル
パッケージ	パッケージのデザイン	A*、B*、C*
ブランド	ブランド名	K2R、Glory、Bissell
価格	Price	\$1.19、\$1.39、\$1.59
シール	サービスシール	なし、あり
料金	料金の払戻し	なし、あり

10人の消費者が、これらの因子により定義された22個のプロファイルに順位を付けます。変数「嗜好」には、各プロファイルの平均順位の序列が含まれています。順位が低いほど、嗜好度は高くなります。この変数には、各プロファイルの嗜好測定値がすべて反映されます。カテゴリ回

帰分析を使用して、5つの因子と嗜好との関係を探査します。このデータセットは、carpet.sav にあります。詳細は、A 付録 p.321 サンプルファイル を参照してください。

標準線型回帰分析

- ▶ 標準線型回帰出力を作成するには、メニューから次の項目を選択します。
分析(A) > Regression > 線型...

注：この機能は Statistics Base オプションが必要です。

図 9-1
[線型回帰] ダイアログ ボックス



- ▶ 従属変数として「嗜好」を選択します。
- ▶ 独立変数として「パッケージのデザイン」から「料金の払戻し」までを選択します。
- ▶ [作図] をクリックします。

図 9-2
[作図] ダイアログ ボックス



- ▶ y 軸変数として「*ZRESID」を選択します。
- ▶ x 軸変数として「*ZPRED」を選択します。
- ▶ [続行] をクリックします。
- ▶ [線型回帰] ダイアログ ボックスで [保存] をクリックします。

図 9-3
[保存] ダイアログ ボックス



- ▶ [残差] グループの [標準化] を選択します。
- ▶ [続行] をクリックします。
- ▶ [線型回帰] ダイアログ ボックスで [OK] をクリックします。

モデルの要約

図 9-4
標準線型回帰のモデルの要約

モデル	R	R2乗	調整済み R2乗	推定値の標準誤差
1	.841 ^a	.707	.615	3.998

a. 予測値: (定数)、料金の払戻し、価格、サービス シール、ブランド名、パッケージのデザイン。

b. 従属変数: 嗜好

この問題における関係を記述するための標準的なアプローチは線型回帰です。データに対して回帰モデルがどれだけ適合しているかを表す最も一般的な測定値は R^2 です。この統計量は、予測変数の重み付き組み合わせにより説明される応答の分散量を表します。 R^2 が 1 に近いほど、モデルはデータに適合しています。5 つの予測変数での「嗜好」の回帰の結果、 R^2 の値は 0.707 になりました。これは、嗜好順位の分散の約 71% が線型回帰の予測変数により説明されることを示しています。

係数(O)

表には、標準化係数が示されています。係数の符号は、ある予測変数の増加時に予測応答が増加するか減少するかにかかわらず、他のすべての予測変数は一定であることを示しています。カテゴリ データの場合、カテゴリ コードにより、予測変数の増加の意味が決まります。たとえば、「料金の払戻し」、「パッケージのデザイン」、「サービス シール」のいずれかが増加すると、予測嗜好順位が低下します。「料金の払戻し」では、[料金の払い戻しなし] を 1 とコード化し、[料金の払戻し] を 2 とコード化しています。「料金の払戻し」の増加は、返金保証の追加を意味します。したがって、返金保証が追加されると予測嗜好順位は低下します。これは予測嗜好度の増加を意味します。

図 9-5
回帰係数

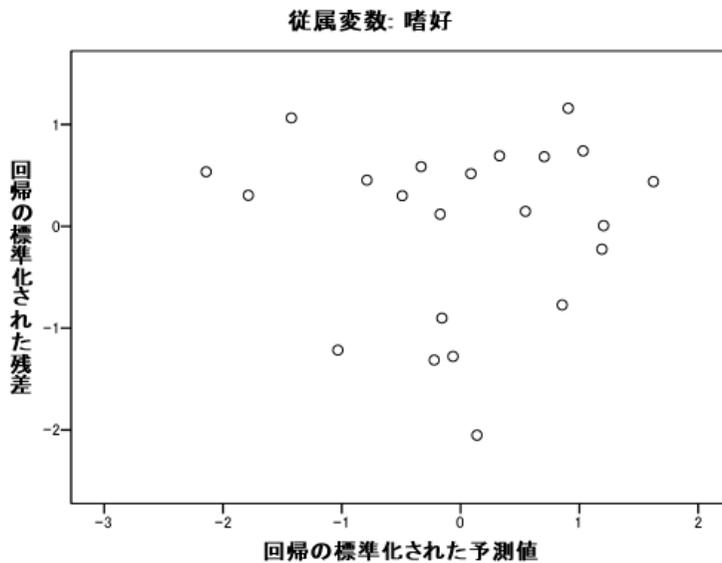
モデル		非標準化係数		標準化係数		有意確率
		B	標準誤差	ベータ	t	
1	(定数)	22.529	5.177		4.352	.000
	パッケージのデザイン	-4.159	1.036	-.560	-4.015	.001
	ブランド名	.429	1.054	.056	.407	.689
	価格	2.703	1.009	.366	2.681	.016
	サービス シール	-4.314	1.780	-.330	-2.423	.028
	料金の払戻し	-2.779	1.921	-.197	-1.447	.167

係数の値には、予測嗜好順位の変化量が反映されます。標準化係数を使用すると、変数の標準偏差に基づいて解釈が行われます。各係数が示す標準偏差の数値は、予測応答が予測変数内の 1 つの標準偏差に向かって変化し、それ以外のすべての予測変数が一定であることを表しています。たと

例えば、「ブランド名」の1つの標準偏差が変化すると、予測嗜好度の標準偏差が0.056に増加します。「嗜好」の標準偏差は6.44であるため、「嗜好」の増加率は、 $0.056 \times 6.44 = 0.361$ になります。「パッケージのデザイン」が変化すると、予測嗜好度は最も大きく変化します。

残差の散布図

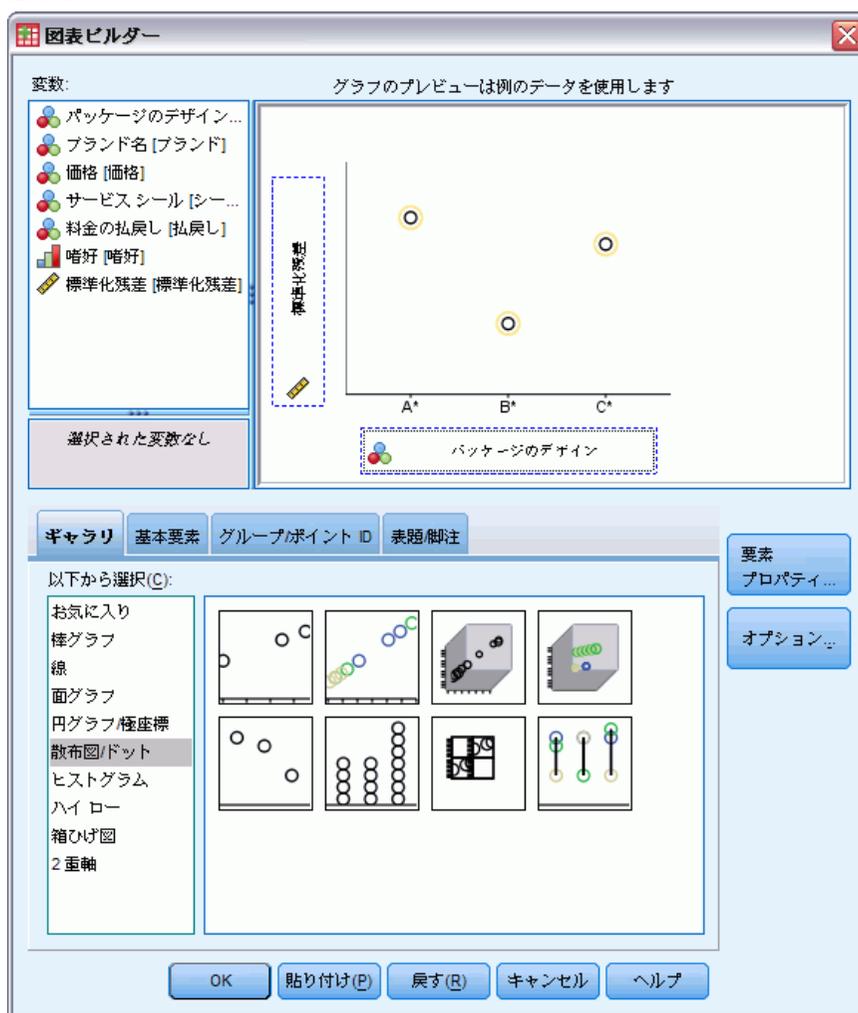
図 9-6
残差と予測値



標準化残差は、標準化予測値に対してプロットされます。モデルの適合度が高い場合、パターンは存在しません。ここでは、低い標準化予測値と高い標準化予測値の両方が正の残差を持つU字型が表示されます。0に近い標準化予測値は、負の残差を持つ傾向があります。

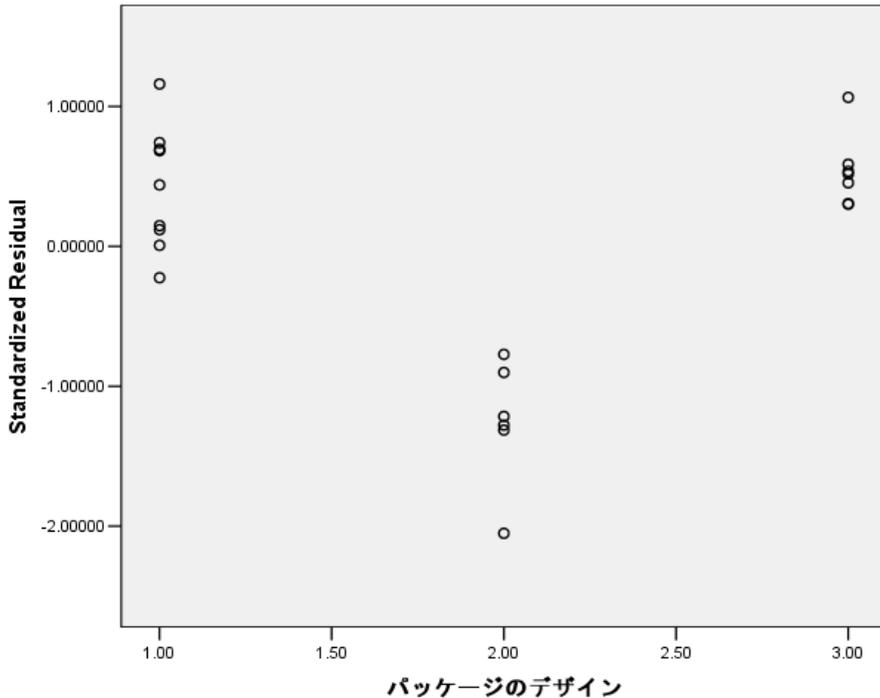
- ▶ 予測変数「パッケージのデザイン」により残差の散布図を作成するには、メニューから次の項目を選択します。
グラフ(G) > 図表ビルダー(C)...

図 9-7
図表ビルダー



- ▶ [散布図/ドット] ギャラリーを選択し、[単純散布図] を選択します。
- ▶ y 軸変数として「標準化残差」を選択し、x 軸変数として「パッケージのデザイン」を選択します。
- ▶ [OK] をクリックします。

図 9-8
残差とパッケージのデザイン



パッケージに対する標準化残差のプロットでは、より明確な U 字型が示されます。他の 2 つのデザインでは負の残差が 1 つだけであるのに対して、デザイン B* のすべての残差は負です。線型回帰モデルでは、各変数に適合するパラメータは 1 つなので、標準的なアプローチではその関係を明確にできません。

カテゴリ回帰分析

変数のカテゴリ特性、および「嗜好」と「パッケージのデザイン」間の非線型関係は、最適得点の回帰の方が標準回帰よりも優れていることを示しています。U 字型の残差プロットは、「パッケージのデザイン」を名義として使用する必要があることを示しています。他のすべての予測変数は、数値尺度水準で扱われます。

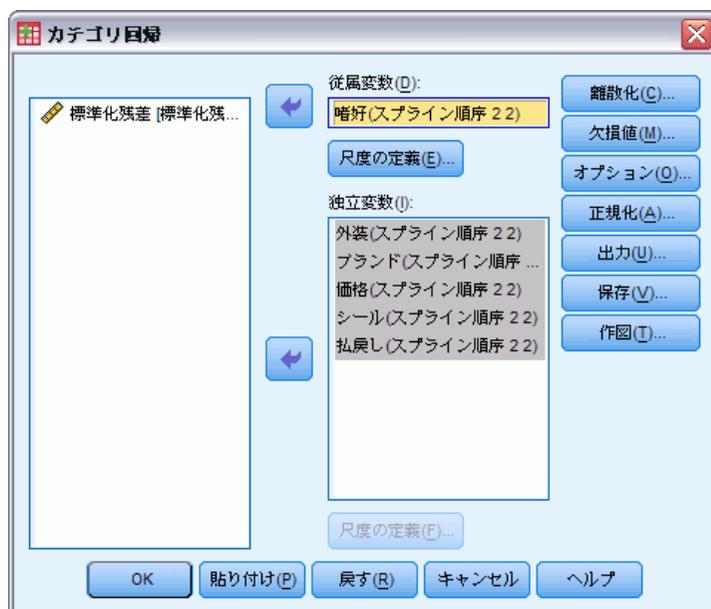
応答変数については、特別な考慮が必要です。たとえば、「嗜好」の値を予測する場合があります。その場合、数量化でカテゴリのできるだけ多くのプロパティを回復することが望まれます。順序尺度水準か名義尺度水準を使用すると、応答カテゴリ間の差は無視されます。ただし、応答カテゴリの線型変換を行うことにより、カテゴリ間の差を保持できます。こ

のため、応答は数値で尺度化することが一般に好ましいとされるので、ここではその方法を使用します。

分析の実行

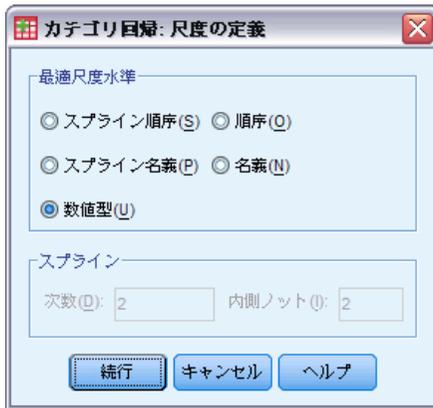
- ▶ カテゴリ回帰分析を実行するには、メニューから次の項目を選択します。
分析(A) > Regression > 最適尺度法 (CATREG)...

図 9-9
[カテゴリ回帰] ダイアログ ボックス



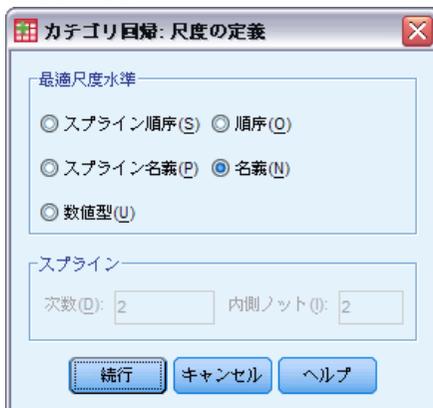
- ▶ 従属変数として「嗜好」を選択します。
- ▶ 独立変数として「パッケージのデザイン」から「料金の払戻し」までを選択します。
- ▶ 「嗜好」を選択し、[尺度の定義] をクリックします。

図 9-10
[尺度の定義] ダイアログ ボックス



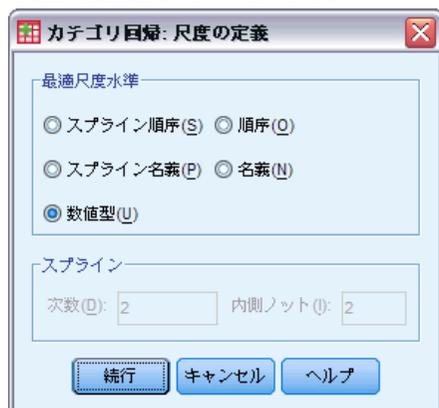
- ▶ 最適尺度水準として [数値] を選択します。
- ▶ [続行] をクリックします。
- ▶ 「パッケージのデザイン」を選択し、[カテゴリ回帰] ダイアログ ボックスの [尺度の定義] をクリックします。

図 9-11
[尺度の定義] ダイアログ ボックス



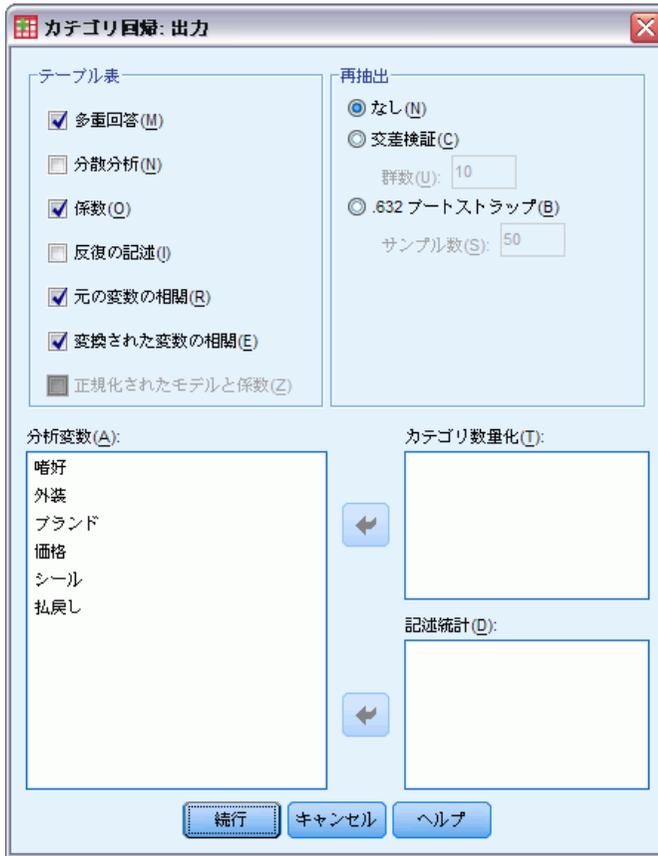
- ▶ 最適尺度水準として [名義] を選択します。
- ▶ [続行] をクリックします。
- ▶ 「ブランド名」から「料金の払戻し」までを選択し、[カテゴリ回帰] ダイアログ ボックスの [尺度の定義] をクリックします。

図 9-12
[尺度の定義] ダイアログ ボックス



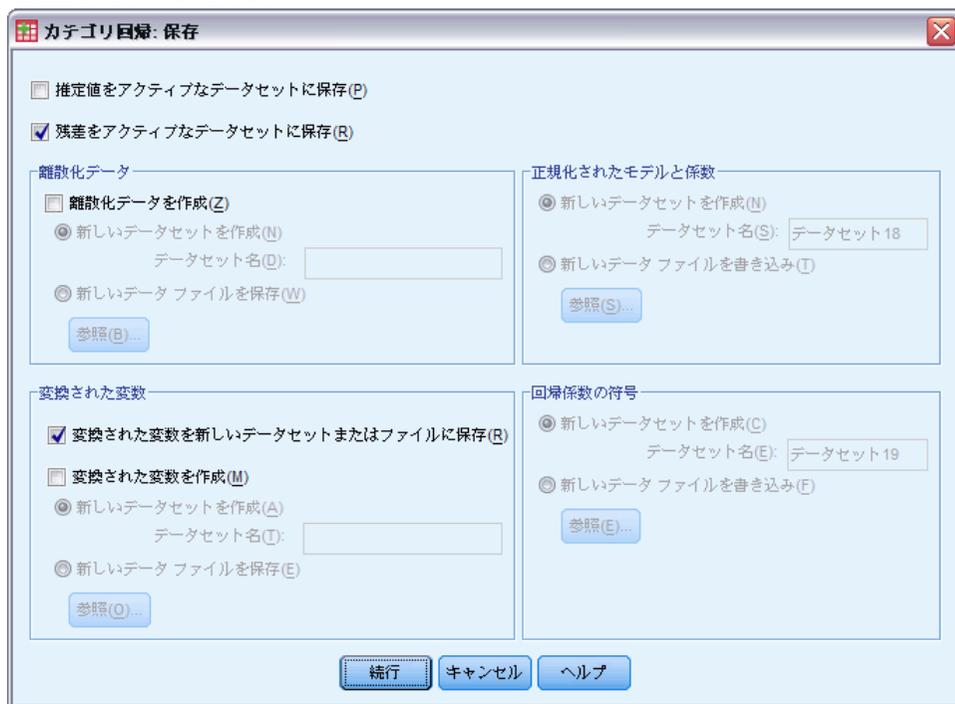
- ▶ 最適尺度水準として [数値] を選択します。
- ▶ [続行] をクリックします。
- ▶ [カテゴリ回帰] ダイアログ ボックスで [出力] をクリックします。

図 9-13
[出力] ダイアログ ボックス



- ▶ [元の変数の相関] と [変換された変数の相関] を選択します。
- ▶ [分散分析] の選択を解除します。
- ▶ [続行] をクリックします。
- ▶ [カテゴリ回帰] ダイアログ ボックスで [保存] をクリックします。

図 9-14
[保存] ダイアログ ボックス



- ▶ [残差をアクティブなデータセットに保存] を選択します。
- ▶ [変換された変数をアクティブなデータセットに保存] を [変換された変数] グループから選択します。
- ▶ [続行] をクリックします。
- ▶ [カテゴリ回帰] ダイアログ ボックスで [作図] をクリックします。

図 9-15
[作図] ダイアログ ボックス



- ▶ 「パッケージのデザイン」と「価格」の変換プロットの作成を選択します。
- ▶ [続行] をクリックします。
- ▶ [カテゴリ回帰] ダイアログ ボックスで [OK] をクリックします。

交互相関

予測変数間の交互相関は、回帰の多重共線性の識別に役立ちます。相関度の高い変数は、不安定な回帰推定の原因となります。ただし、その強い相関関係により、モデルから変数の1つを除外した際の予測への影響が最小限に抑えられます。除外された変数により説明できる応答の分散は、その変数と相関関係のある残りの変数によって依然として説明されます。ただし、0次相関は外れ値に敏感です。また、予測変数と他の予測変数の組み合わせ間は相関関係が強いので、多重共線性を識別することもできません。

図 9-16
変換前の予測変数の相関

	パッケージのデザイン	ブランド名	価格	サービスシール	料金の払戻し
パッケージのデザイン	1.000	-.189	-.126	.081	.066
ブランド名	-.189	1.000	.065	-.042	-.034
価格	-.126	.065	1.000	.000	.000
サービスシール	.081	-.042	.000	1.000	-.039
料金の払戻し	.066	-.034	.000	-.039	1.000
次元	1	2	3	4	5
固有値	1.291	1.038	.980	.905	.785

図 9-17
変換された予測変数の相関 (カテゴリ / カテゴリの回帰)

	パッケージのデザイン	ブランド名	価格	サービスシール	料金の払戻し
パッケージのデザイン	1.000	-.156	-.089	.032	.102
ブランド名	-.156	1.000	.065	-.042	-.034
価格	-.089	.065	1.000	.000	.000
サービスシール	.032	-.042	.000	1.000	-.039
料金の払戻し	.102	-.034	.000	-.039	1.000
次元	1	2	3	4	5
固有値	1.248	1.043	.983	.905	.821

変換前と変換後の予測変数の交互相関が示されています。すべての値は 0 に近くなります。これは、個々の変数間の多重共線性が懸案事項ではないことを示しています。

相関関係が変化するのは「パッケージのデザイン」のみであることに注目してください。他のすべての予測変数は数値として扱われるので、カテゴリ間の差とカテゴリの順序はこれらの変数では保持されます。したがって、相関関係は変化しません。

モデルの適合度と係数

カテゴリ回帰分析手続きで得られる R^2 の値は 0.948 です。これは、最適変換された予測変数の回帰により、変換された嗜好順位の分散の約 95% が説明されることを示しています。予測変数の変換により、標準的なアプローチへの適合度が向上します。

図 9-18
カテゴリ回帰分析のモデルの要約

多重 R	R ² 乗	調整済み R ² 乗
.974	.948	.927

従属変数 嗜好
予測変数 パッケージのデザイン ブランド名
価格 サービスシール 料金の払戻し

次の表には、標準回帰係数が示されています。カテゴリ回帰分析では変数が標準化されるので、標準化係数のみが報告されます。これらの値はそれぞれ対応する標準誤差で割られ、変数ごとに F 検定が行われます。ただ

し、各変数の検定は、他の予測変数がモデル内に存在することが条件となります。つまり、検定では、他のすべての予測変数が存在するモデルからある予測変数を除外することにより、モデルの予測機能が低下するかどうかが決まります。これらの値は、以降のモデルに対して一度に複数の変数を除外するのに使用すべきではありません。さらに、交互最小2乗法により、数量化が最適化されます。これは、これらの検定を控えめに解釈する必要があることを意味します。

図 9-19
変換された予測変数の標準化係数

	標準化係数		自由度	F	有意確率
	ベータ	標準誤差			
パッケージのデザイン	-.748	.060	2	155.289	.000
ブランド名	.045	.060	1	.578	.459
価格	.371	.059	1	39.312	.000
サービスシール	-.350	.059	1	35.299	.000
料金の払戻し	-.159	.059	1	7.175	.017

従属変数 嗜好

係数は、「パッケージのデザイン」の場合に最大となります。「パッケージのデザイン」の標準偏差が1つ増加すると、予測嗜好順位の標準偏差が0.748に減少します。ただし、「パッケージのデザイン」は名義として扱われるので、数量化における増加が元のカテゴリコードの増加に対応している必要はありません。

多くの場合、標準化係数は各予測変数の重要度を反映するものとして解釈されます。ただし、回帰係数では、予測変数の影響または予測変数間の関係を十分に説明することはできません。予測変数の効果を十分に探索するには、代替統計量を標準化係数と組み合わせて使用する必要があります。

相関係数と重要度

回帰に対する予測変数の寄与率を解釈するには、回帰係数を検定するだけでは不十分です。さらに、相関係数、偏相関、および部分相関を検定する必要があります。次の表には、各変数に対するこれらの相関測定法が含まれています。

0次相関は、変換後の予測変数と変換後の応答との相関です。このデータの場合、「パッケージのデザイン」で最も強い相関が発生します。ただし、予測変数か応答のいくつかの分散を説明できれば、予測変数の精度が高くなります。

図 9-20
0 次相関、部分相関、偏相関（変換された変数）

	相関			重要度	許容度	
	ゼロ次	偏	部分		変換後	変換前
パッケージのデザイン	-.816	-.955	-.733	.644	.959	.942
ブランド名	.206	.193	.045	.010	.971	.961
価格	.440	.851	.369	.172	.989	.982
サービスシール	-.370	-.838	-.349	.137	.996	.991
料金の払戻し	-.223	-.569	-.158	.037	.987	.993

従属変数 嗜好

モデル内の他の変数により、応答の予測時に特定の予測変数のパフォーマンスにばらつきが生じる可能性があります。偏相関係数により、予測変数と応答から、他の予測変数の線型効果を取り除かれます。この測定方法では、他の予測変数での予測変数の回帰による残差と、他の予測変数での応答の回帰による残差の相関が等しくなります。平方偏相関は、他の変数の効果を取り除いた後に残った応答の残差分散と比較して説明される分散の比率と一致します。たとえば、「パッケージのデザイン」の偏相関は -0.955 です。他の変数の効果を取り除くと、「パッケージのデザイン」により、嗜好順位の分散の $(-0.955)^2 = 0.91 = 91\%$ が説明されます。また、他の変数の効果を取り除いた場合、「価格」と「サービスシール」によっても分散の大部分が説明されます。

応答と予測変数の両方から変数の効果を取り除くことの代替策として、予測変数だけから効果を取り除くことができます。他の予測変数での予測変数の回帰による応答と残差間の相関が、部分相関です。この値を 2 乗すると、応答の全分散と比較して説明される分散の比率の測定値が得られます。「パッケージのデザイン」から、「ブランド名」、「サービスシール」、「料金の払戻し」、および「価格」の効果を取り除いた場合、「パッケージのデザイン」の残りの部分により、嗜好順位の分散の $(-0.733)^2 = 0.54 = 54\%$ が説明されます。

重要度

回帰係数と相関係数に加えて、Pratt の相対重要度 (Pratt, 1987) は、回帰に対する予測変数の寄与率の解釈に役立ちます。他の重要度と比べて個々の重要度が高い場合、その予測変数は回帰にとって重要です。また、重要な予測変数と同等のサイズの係数を持つ変数の重要度が低い場合、抑制変数の存在が示唆されます。

回帰係数とは対照的に、この測定方法では予測変数の重要度が相加的に定義されます。つまり、1 組の予測変数の重要度は、個々の予測変数の重要度の総和となります。Pratt の相対重要度は、回帰係数と予測変数の 0 次相関の積と等しくなります。これらの積は R^2 に追加するため、 R^2 で割って合計が 1 となります。たとえば、予測変数「パッケージデザイン」および「ブランド名」のセットの重要度は、 0.654 となります。重要度が最も高いのは「パッケージのデザイン」です。「パッケージのデ

ザイン」、「価格」、および「サービス シール」により、この予測変数の組み合わせの重要度の 95% が占められます。

多重共線性

予測変数間の相関係数が大きいと、回帰モデルの安定性が著しく低下します。予測変数間に相関関係があると、パラメータ推定値が不安定になります。許容度は、独立変数が互いにどれだけ線型関係にあるかを示します。この測定値は、方程式内の他の独立変数によって説明されていない変数の分散の割合です。他の予測変数によってある予測変数の分散のほとんどを説明できる場合、その予測変数はモデル内に必要ありません。許容度の値が 1 に近い場合は、他の予測変数からその変数をあまり正確に予測できないことを示します。逆に、許容度が非常に小さい変数では、モデルに提供される情報がわずかなの、計算上の問題が生じることがあります。さらに、Pratt の相対重要度が大きい負の値である場合は、多重共線性が存在することを示しています。

すべての許容度の測定値は非常に高くなります。どの予測変数も他の予測変数によっては正確に予測されません。また、多重共線性も存在しません。

変換プロット

変換前のカテゴリ値を対応する数量化に対してプロットすると、数量化のリストからはわからなかったトレンドが明らかになる可能性があります。そのようなプロットを一般に変換プロットといいます。数量化の結果が類似するカテゴリに注目する必要があります。これらのカテゴリが予測応答に与える影響は同一のものです。ただし、プロットの基本的な外観は、変換の種類によって決まります。

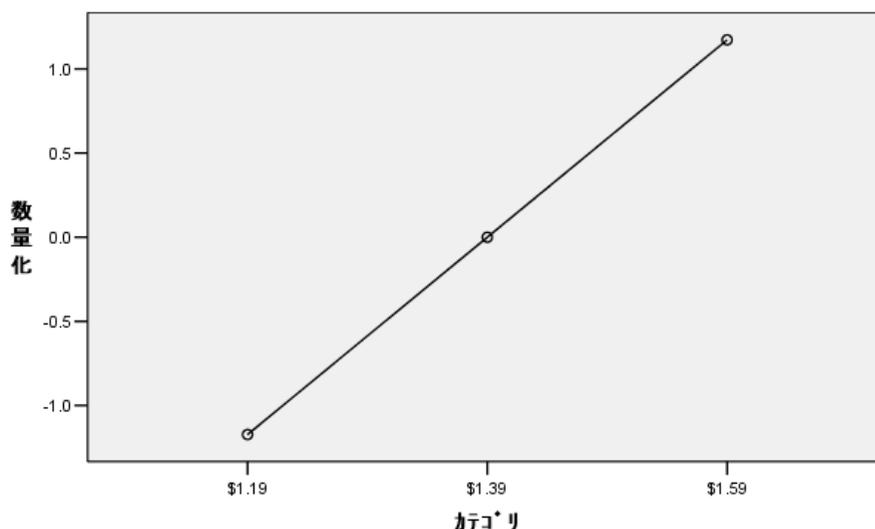
数値として扱われる変数の場合、数量化と変換前のカテゴリ間には線型関係が成立します。これは、変換プロットが直線であることからわかります。変換前のカテゴリの順番とカテゴリ間の差は、数量化において保持されます。

順序として扱われる変数の数量化の順序は、変換前のカテゴリの順序と一致します。ただし、カテゴリ間の差は保持されません。その結果、変換プロットは非減少となりますが、直線である必要はありません。連続したカテゴリの数量化が類似している場合、おそらくそれらのカテゴリは区別する必要がなく、結合される可能性があります。そのようなカテゴリの場合、変換プロットは水平線を示します。ただし、名義として扱う必要がある変数に順序の構造を使用した場合にも、このようなパターンが発生する可能性があります。それに続く変数を名義として扱い、同じパターンが発生した場合は、カテゴリの結合が保証されます。さらに、順序として扱った変数の数量化が直線に沿っている場合は、数値変換の方が適している可能性があります。

名義として扱われる変数の場合、横軸のカテゴリの順序は、カテゴリの表現に使用するコードの順序と一致します。カテゴリの順序またはカテゴリ間の距離の解釈はありません。プロットにより、すべての非線型形式または線型形式を推測できます。増加トレンドが存在する場合は、変数を順序として扱うことを試す必要があります。名義変換プロットにより線型トレンドが表示される場合は、数値変換の方が適している可能性があります。

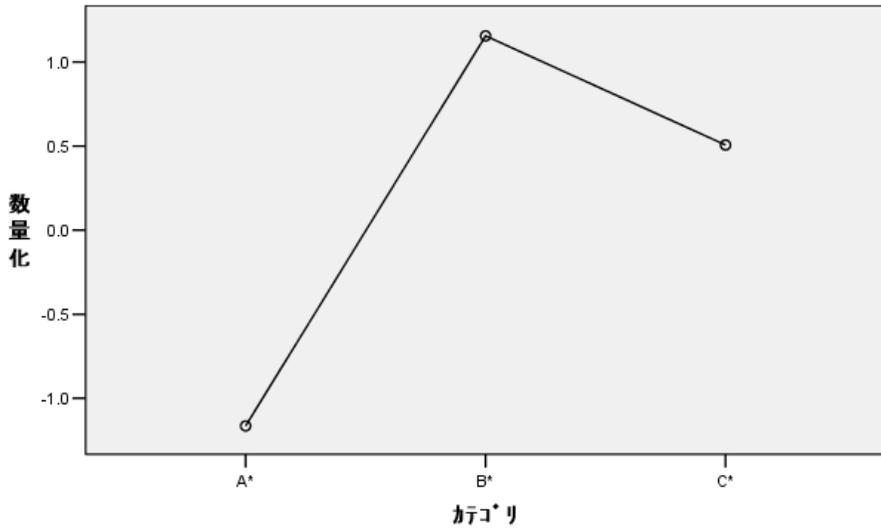
次の図に、数値として扱われた「価格」の変換プロットを示します。直線に沿ったカテゴリの順序が、変換前のカテゴリの順序と一致していることに注意してください。また、[\$1.19] と [\$1.39] の数量化の差 (-1.173 と 0) は、[\$1.39] と [\$1.59] の数量化の差 (0 と 1.173) と等しくなります。カテゴリ 1 とカテゴリ 3 がカテゴリ 2 から等距離にあるという事実は、数量化において保持されます。

図 9-21
価格の変換プロット (数値)



「パッケージのデザイン」の名義変換により、次の変換プロットが得られます。2番目のカテゴリが最大の数量化を示す、明確な非線型に注目してください。回帰に関しては、2番目のカテゴリでは予測嗜好順位が低下しますが、最初のカテゴリと3番目のカテゴリでは正反対の結果が得られます。

図 9-22
パッケージのデザイン（名義）の変換プロット



残差分析

アクティブなデータセットに保存した変換後のデータと残差を使用すると、変換された「パッケージのデザイン」の値により予測値の散布図を作成できます。

この散布図を作成するには、[図表ビルダー] ダイアログ ボックスを再表示し、[リセット] をクリックして、前回の選択内容のクリアとデフォルトのオプションの復元を行います。

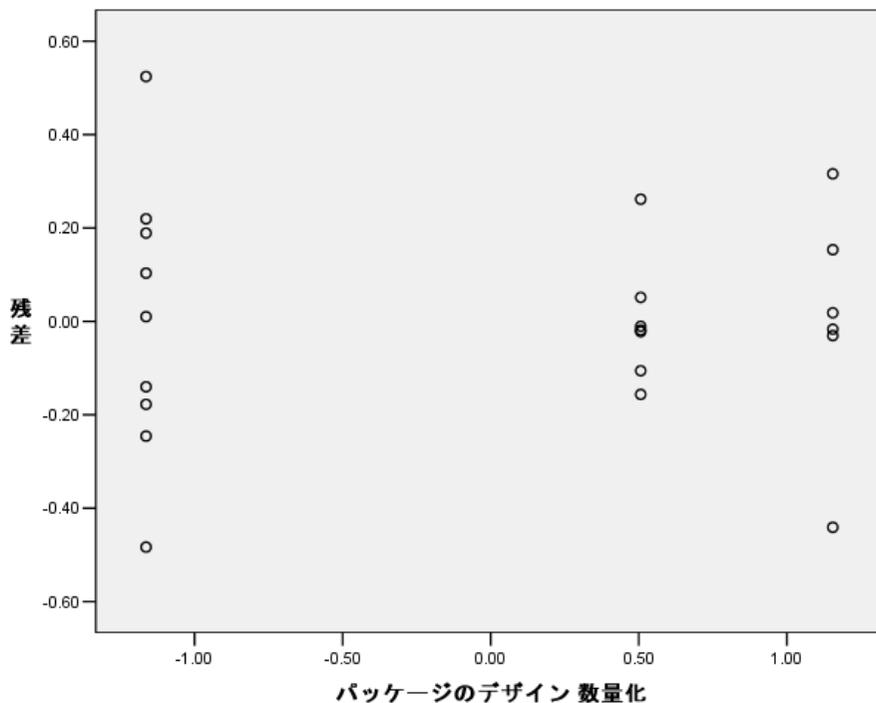
図 9-23
図表ビルダー



- ▶ [散布図/ドット] ギャラリーを選択し、[単純散布図] を選択します。
- ▶ y 軸変数として 残差 を選択します。
- ▶ x 軸変数として パッケージのデザイン を選択します。
- ▶ [OK] をクリックします。

散布図には、「パッケージのデザイン」の最適な得点に対してプロットされた標準化残差が示されます。すべての残差は、値が 0 の 2 つの標準偏差内に存在します。標準線型回帰分析で得られた散布図内に存在する U 字型は、無作為に散らばった点に置き換えられます。最適にカテゴリを数量化することにより、予測能力が向上します。

図 9-24
カテゴリ回帰分析の残差



例:オゾンのデータ

この例では、大規模なデータのセットを使用し、最適尺度変換の選択と効果について説明します。データには、主に Breiman と Friedman (Breiman および Friedman (F), 1985)、および Hastie と Tibshirani (Hastie および Tibshirani, 1990) によって以前に分析された 6 個の気象変数に対する 330 個の観測値が含まれています。次の表では、変換前の変数について説明します。カテゴリ回帰分析では、残りの変数からのオゾン濃度の予測を試行します。それまでの研究者により、これらの変数間に非線型性が確認されています。この場合、標準的な回帰アプローチは使用できません。

テーブル 9-2
変換前の変数

変数	説明
ozon	daily ozone level (38 種類のカテゴリに分類されます)
ibh	inversion base height
dpg	pressure gradient (mm Hg)

変数	説明
vis	visibility (miles)
temp	temperature (degrees F)
day	day of the year

このデータセットは ozone.sav にあります。詳細は、A 付録 p. 321 [サンプルファイル](#) を参照してください。

変数の離散化

変数のカテゴリが実際に解釈可能なカテゴリ数より多い場合、[離散化] ダイアログ ボックスを使用してカテゴリを変更し、カテゴリの範囲を管理可能な数字まで減らします。

変数 [Day of the year] の最小値は 3、最大値は 365 です。この変数をカテゴリ回帰分析で使用することは、365 個のカテゴリを持つ変数を使用するのと同じことです。同様に、[Visibility (miles)] の範囲は 0 ~ 350 です。分析の解釈を簡素化するには、これらの変数を長さが 10 ずつの等間隔に離散化します。

変数 [Inversion base height] の範囲は 111 ~ 5000 です。このような多くのカテゴリを持つ変数を変換すると、非常に複雑な関係が生成されます。ただし、この変数を長さが 100 ずつの等間隔に離散化すると、カテゴリは約 50 個になります。5,000 個のカテゴリを持つ変数ではなく 50 個のカテゴリを持つ変数を使用することにより、解釈が非常に簡素化されます。

[Pressure gradient (mm Hg)] の範囲は -69 ~ 107 です。この手続きでは、負の数でコード化されたカテゴリはすべて分析対象から除外されますが、この変数を長さが 10 ずつの等間隔に離散化すると、カテゴリは約 19 個になります。

「Temperature (degrees F)」の範囲は、25 ~ 93 (華氏目盛り) です。データを摂氏目盛りの場合と同様に分析するには、この変数を長さが 1.8 ずつの等間隔に離散化します。

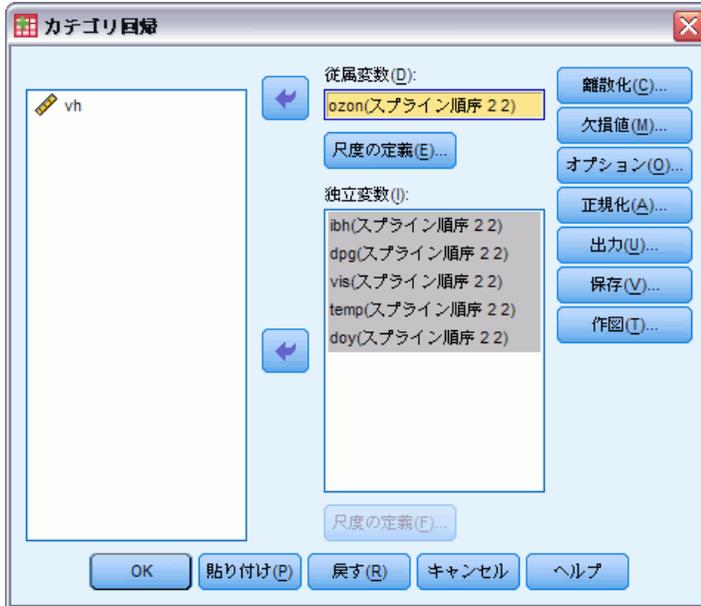
変数ごとに異なる離散化を実行することが望まれます。ここに挙げた例は、これ以降の都合にあわせて選んだものです。もっとカテゴリ数を減らす場合は、間隔を大きくします。たとえば、「Day of the year」は年間の月数や季節に分割できます。

変換の種類を選択

各変数は、いくつかの異なるレベルのいずれかで分析できます。ただし、応答の予測が目的なので、数値最適尺度水準を使用することにより、応答を「そのまま」尺度化する必要があります。その結果、カテゴリ間の順序と差は変換後の変数でも保持されます。

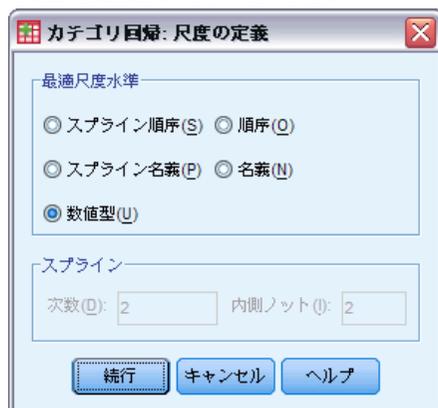
- ▶ カテゴリ回帰分析を実行するには、メニューから次の項目を選択します。
分析(A) > Regression > 最適尺度法 (CATREG)...

図 9-25
[カテゴリ回帰] ダイアログ ボックス



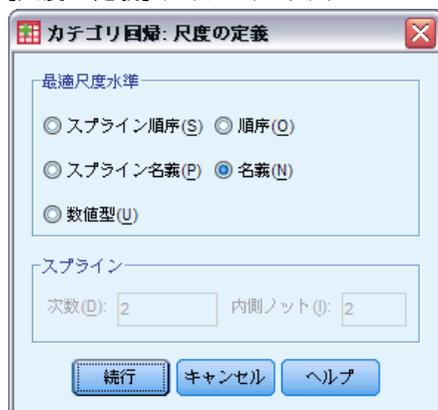
- ▶ 従属変数として「Daily ozone level」を選択します。
- ▶ 独立変数として「Inversion base height」から「Day of the year」までを選択します。
- ▶ 「Daily ozone level」を選択し、[尺度の定義] をクリックします。

図 9-26
[尺度の定義] ダイアログ ボックス



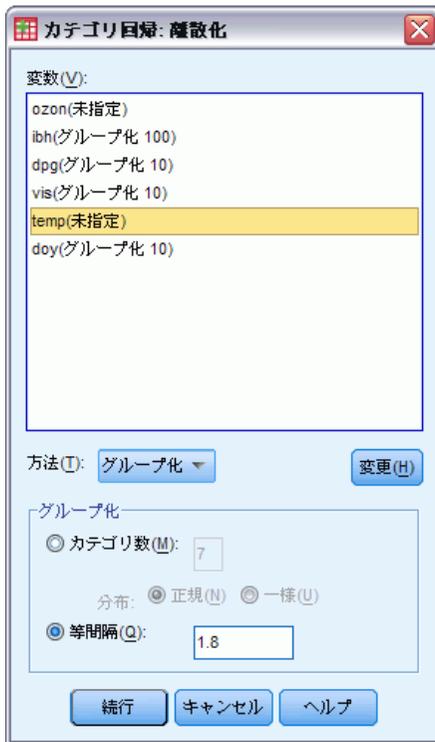
- ▶ 最適尺度水準として [数値] を選択します。
- ▶ [続行] をクリックします。
- ▶ 「Inversion base height」から「Day of the year」までを選択し、[カテゴリ回帰] ダイアログ ボックスの [尺度の定義] をクリックします。

図 9-27
[尺度の定義] ダイアログ ボックス



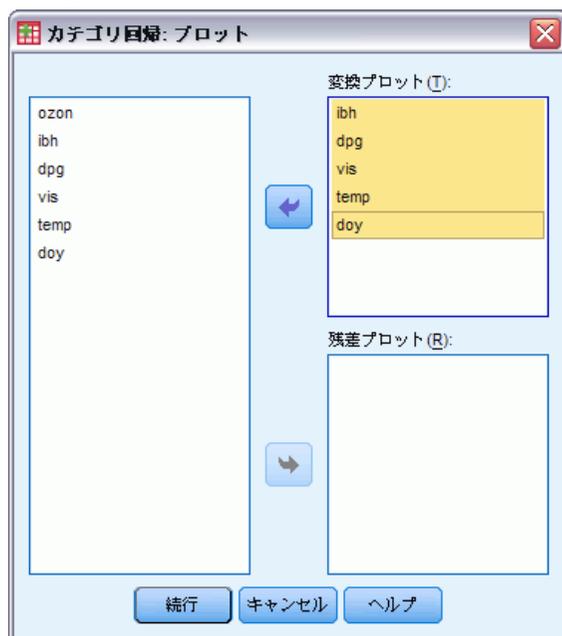
- ▶ 最適尺度水準として [名義] を選択します。
- ▶ [続行] をクリックします。
- ▶ [カテゴリ回帰] ダイアログ ボックスで [離散化] をクリックします。

図 9-28
[離散化] ダイアログ ボックス



- ▶ 「ibh」を選択します。
- ▶ [等間隔] を選択し、間隔の長さとして「100」と入力します。
- ▶ [変更] をクリックします。
- ▶ 「dpg」、「vis」、および「doy」を選択します。
- ▶ 間隔の長さとして「10」と入力します。
- ▶ [変更] をクリックします。
- ▶ 「temp」を選択します。
- ▶ 間隔の長さとして「1.8」と入力します。
- ▶ [変更] をクリックします。
- ▶ [続行] をクリックします。
- ▶ [カテゴリ回帰] ダイアログ ボックスで [作図] をクリックします。

図 9-29
[作図] ダイアログ ボックス



- ▶ 「Inversion base height」から「Day of the year」までの変換プロットを選択します。
- ▶ [続行] をクリックします。
- ▶ [カテゴリ回帰] ダイアログ ボックスで [OK] をクリックします。

図 9-30
モデルの要約(M)

	多重 R	R2 乗	調整済み R2 乗	見かけ上の予測誤差
標準化データ	.938	.880	.785	.120

従属変数 ozon
予測: ibh dpg vis temp doy

すべての予測変数を名義として扱うと、 R^2 の値は 0.880 になります。このような大量の分散が説明されることは驚くべきことではありません。予測変数を名義として扱っても数量化には制限が課されないからです。ただし、結果の解釈は非常に難しくなる可能性があります。

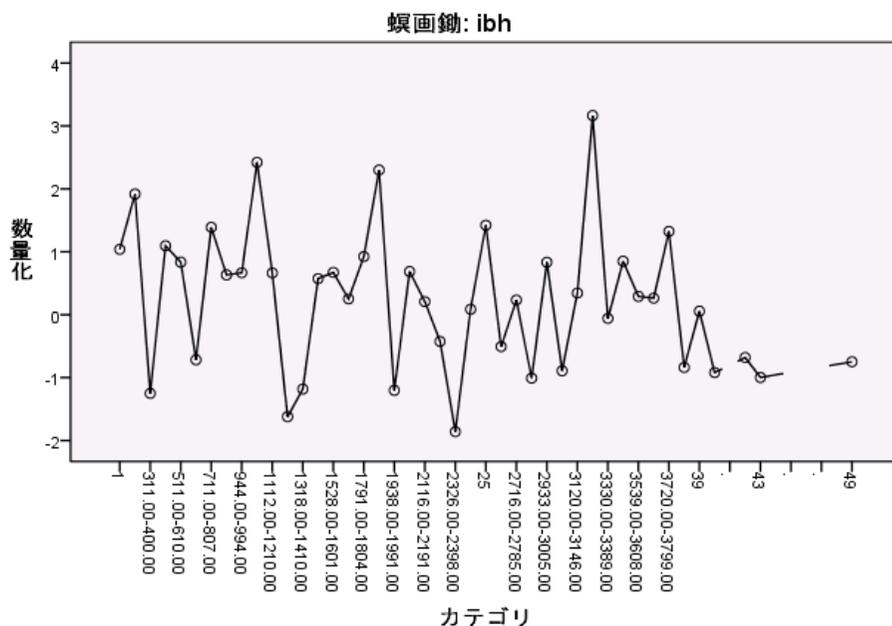
図 9-31
 回帰係数 (すべての予測変数を名義として扱う)

	標準化係数		自由度	F	有意確率
	ベータ	標準誤差の ブートストラップ ラップ(1000) 推定値			
ibh	.297	.053	42	31.047	.000
dpg	.326	.055	16	34.793	.000
vis	.229	.050	17	21.465	.000
temp	.577	.091	35	40.562	.000
day	.420	.070	36	36.171	.000

従属変数: ozon

この表には、予測変数の標準化回帰係数が示されています。これらの値の解釈においてよく起こる間違いは、数量化を無視して係数に注目してしまうことです。「Inversion base height」の正の値が、たとえば予測値が大きくなると「オゾンの予測が大きくなることを示すと単純に断定することはできません。すべての解釈は変換変数に関連する必要があるため、「Inversion base heightの数量が大きくなると、予測される「オゾン」は大きくなります。変換前の変数の効果を調べるには、カテゴリを数量化に関連付ける必要があります。

図 9-32
Inversion base height (名義) の変換プロット



「Inversion base height」の変換プロットでは、明確なパターンは示されません。プロットがギザギザであることからわかるように、低いカテゴリから高いカテゴリに移行すると、数量化に上下のばらつきが見られます。したがって、この変数の効果を説明するには、個々のカテゴリに注目する必要があります。この変数の数量化に順序制限か線型制限を課すと、適合度が大きく低下する場合があります。

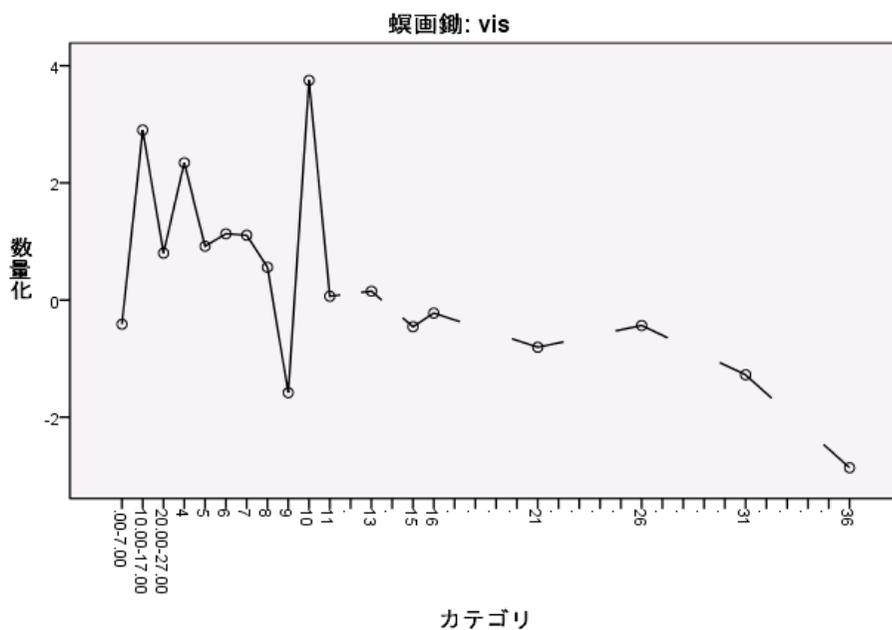
図 9-33
Pressure gradient (名義) の変換プロット



この図には、「Pressure gradient」の変換プロットが示されています。離散化された最初のカテゴリ (1 から 6) は小規模な数量化を受けるので、予測応答に対する寄与率は最小になります。次の 3 つのカテゴリでは若干大きな正の値を受け取るので、予測されたオゾン是中程度増加します。

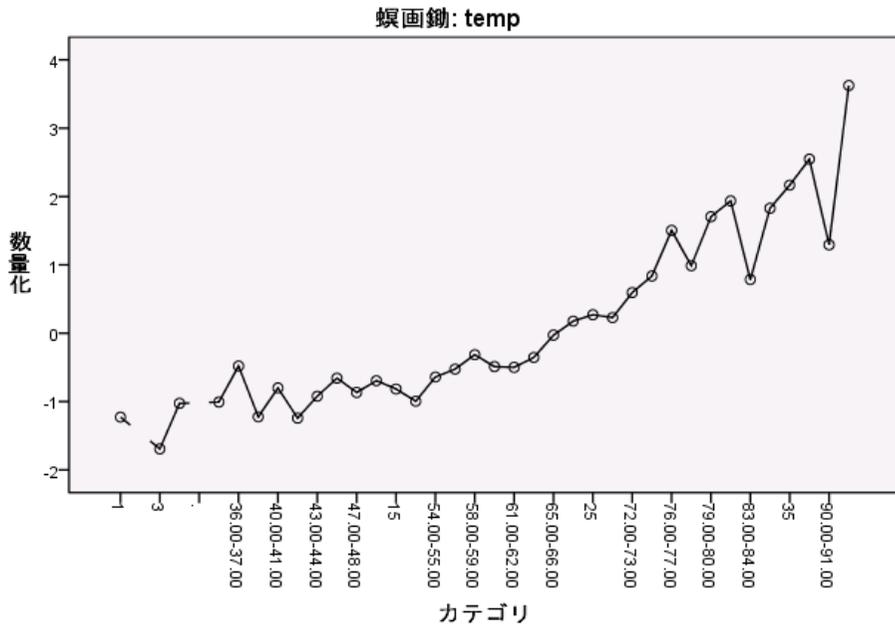
数量化はカテゴリ 16 まで減少します。このカテゴリでは、「Pressure gradient」が予測されたオゾンに対してもたらず減少効果が最大になります。このカテゴリの後は数量化が増加しますが、「Pressure gradient」の順序尺度水準を使用しても適合度が大きく低下することはなく、効果の解釈が簡素化されます。ただし、重要度が 0.04 であることと「Pressure gradient」の回帰係数は、この変数が回帰ではあまり役立たないことを示しています。

図 9-34
Visibility (名義) の変換プロット



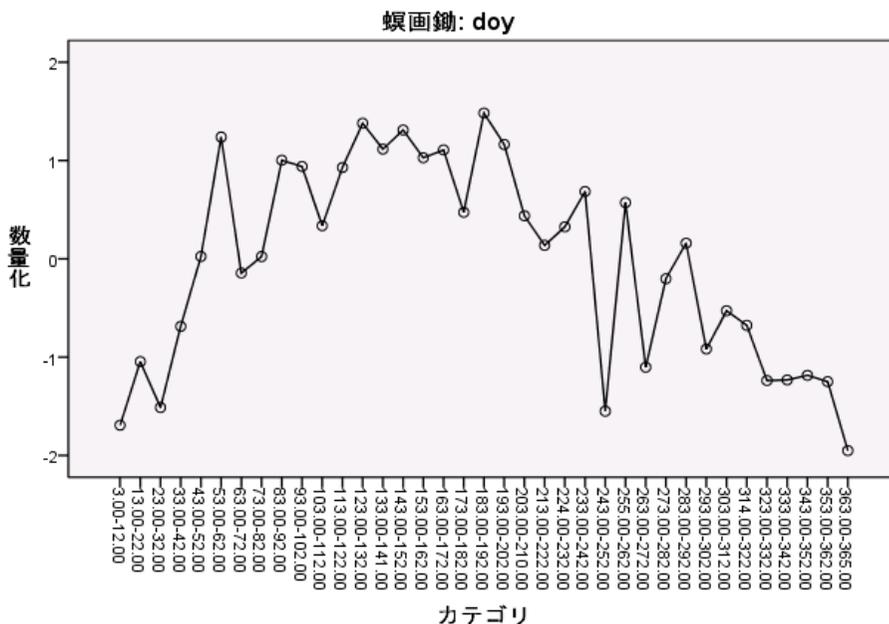
「Inversion base height」の場合と同様に、「Visibility」の変換プロットでも、明確なパターンは示されません。この変数の数量化に順序制限か線型制限を課すと、適合度が大きく低下する場合があります。

図 9-35
Temperature (名義) の変換プロット



「Temperature」の変換プロットでは、代替パターンが示されます。カテゴリーが増加すると、数量化も増加する傾向があります。その結果、「Temperature」が増加すると、予測されたオゾンも増加する傾向があります。このパターンは、順序水準で「Temperature」を尺度化する必要があることを示しています。

図 9-36
Day of the year (名義) の変換プロット



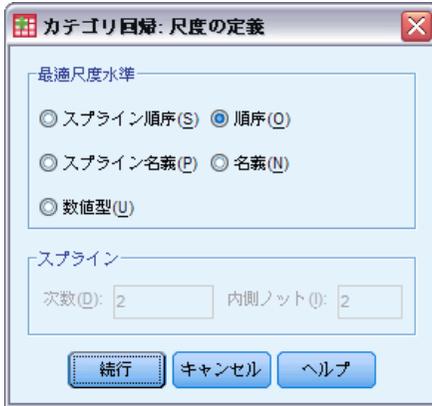
この図には、「Day of the year」の変換プロットが示されています。数量化は、グラフの中心点まで増加し、そこから減少して逆U字型になる傾向があります。「Day of the year」の回帰係数の符号を考慮すると、初期のカテゴリでは、オゾンの予測値に減少効果をもたらす数量化を受けます。中央のカテゴリについては、予測されるオゾンの数量化の効果は増加し、グラフの中心点あたりの最大値に達します。

中心点以降は、オゾンの予測値は数量化により減少する傾向があります。線はギザギザになりますが、全体的な形は依然として識別できます。したがって、変換プロットは、「Temperature」を順序水準で尺度化し、他のすべての予測変数をそのまま名義として尺度化する必要があることを示しています。

「Temperature」を順序水準で尺度化し、回帰を再計算するには、[カテゴリ回帰] ダイアログ ボックスを再度表示します。

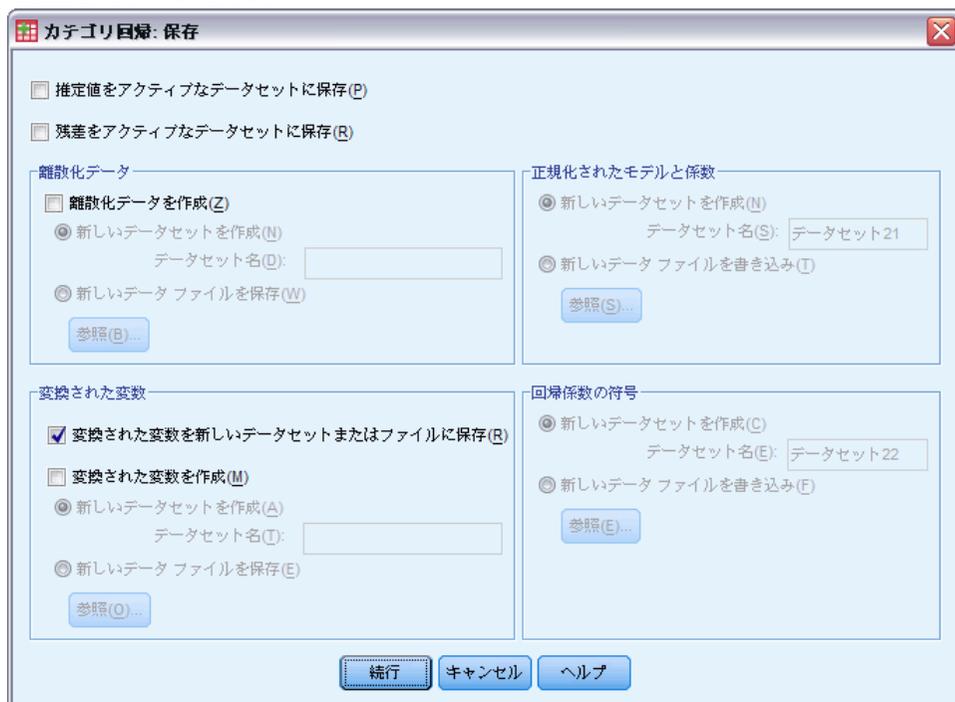
図 9-37

[尺度の定義] ダイアログ ボックス



- ▶ 「Temperature」を選択し、[尺度の定義]をクリックします。
- ▶ 最適尺度水準として[順序]を選択します。
- ▶ [続行]をクリックします。
- ▶ [カテゴリ回帰] ダイアログ ボックスで[保存]をクリックします。

図 9-38
[保存] ダイアログ ボックス



- ▶ [変換された変数をアクティブなデータセットに保存] を [変換された変数] グループから選択します。
- ▶ [続行] をクリックします。
- ▶ [カテゴリ回帰] ダイアログ ボックスで [OK] をクリックします。

図 9-39
Temperature (順序) を使用した回帰のモデルの要約

モデル集計

	多重 R	R2 乗	調整済み R2 乗	見かけ上の予測誤差
標準化データ	.934	.872	.787	.128

従属変数: ozon
予測: ibh dpg vis temp doy

このモデルでは、 R^2 の値は 0.872 になるので、「Temperature」の数量化が順序付けられるように制限を課した場合でも、説明される分散の割合はほとんど低下しません。

図 9-40
Temperature (順序) での回帰係数

	係数				
	標準化係数		自由度	F	有意確率
	ベータ	標準誤差の ブートストラップ ラップ(1000) 推定値			
ibh	.298	.042	42	51.376	.000
dpg	.301	.047	16	41.882	.000
vis	.224	.044	17	25.659	.000
temp	.609	.084	21	52.113	.000
doy	.373	.053	36	50.330	.000

従属変数: ozon

この表は、「Temperature」が順序として尺度化されたモデルの係数を示しています。これらの係数を [Temperature] が名義として尺度化されたモデルの係数と比較した場合、あまり大きな変化は見られません。

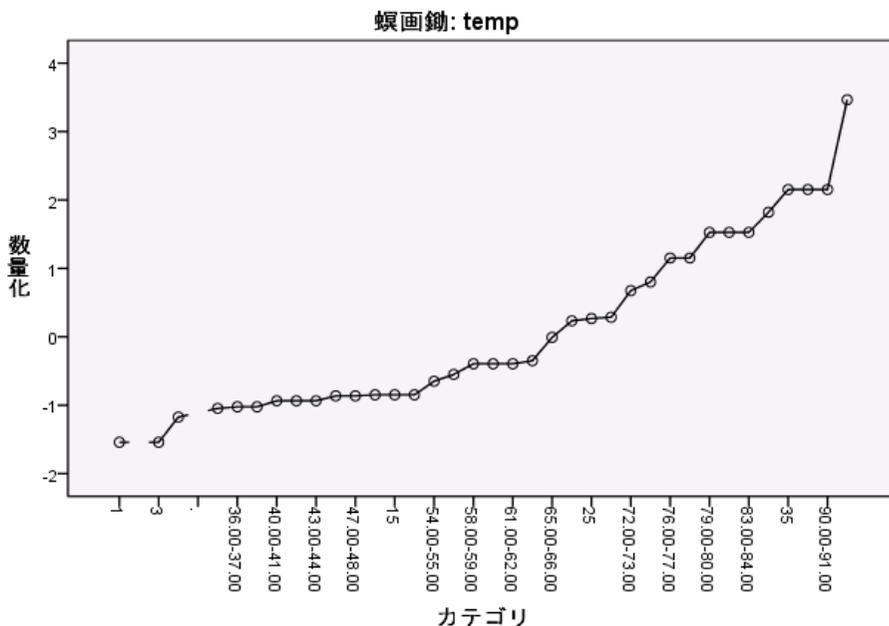
図 9-41
相関係数、重要度、および許容度

	相関および許容度					
	相関			重要度	許容度	
	ゼロ次	偏	部分		変換後	変換前
ibh	.438	.627	.288	.150	.930	.596
dpg	.128	.606	.272	.044	.815	.858
vis	.365	.518	.216	.094	.933	.752
temp	.804	.843	.559	.562	.842	.580
doy	.352	.677	.329	.151	.777	.802

従属変数: ozon

さらに、重要度の測定値は、「Temperature」が他の変数よりも依然として回帰にとって重要であることを示しています。ただし、ここでは、「Temperature」で順序尺度水準が使用されていることと、回帰係数が正であることから、「Temperature」が増加すると、オゾンの予測値も増加すると断定できます。

図 9-42
Temperature (順序) の変換プロット



変換プロットには、「Temperature」の数量化で課された順序制限が図示されています。名義変換のギザギザの線は、ここではなだらかな増加線に置き換えられます。さらに、長い水平状態は存在しないので、カテゴリの集約は必要ありません。

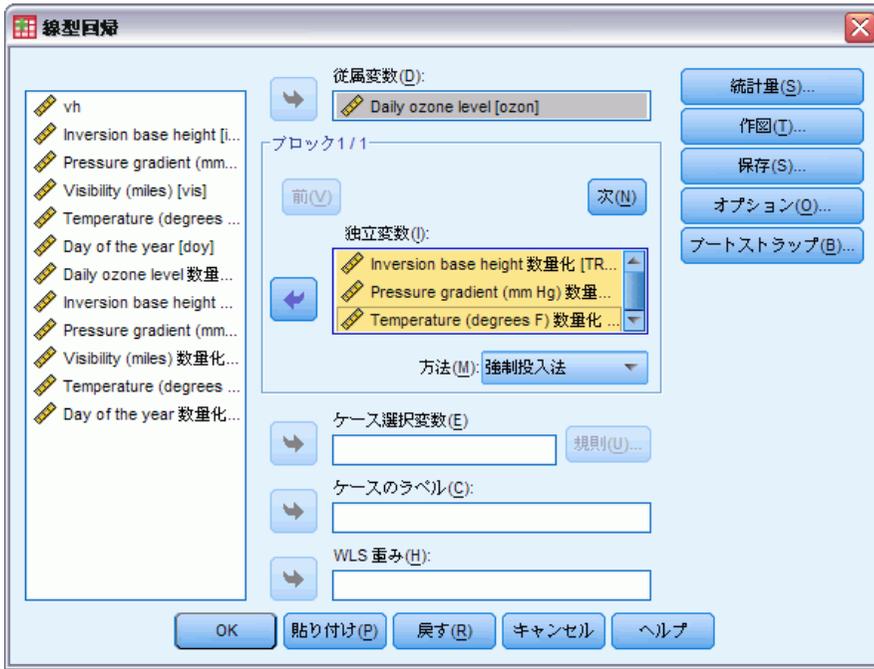
数量化の最適性

カテゴリ回帰分析で変換された変数は、標準線型回帰で使用でき、同一の結果が得られます。ただし、数量化はそれらの変数を作成したモデルに対してのみ最適化されます。ある予測変数のサブセットを線型回帰で使った場合の結果と、そのサブセットを最適尺度法による回帰で使った場合の結果は一致しません。

たとえば、計算したカテゴリ回帰分析の R^2 の値は 0.875 です。変換後の変数を保存したので、「Temperature」、「Pressure gradient」、および「Inversion base height」のみを予測変数として使用する線型回帰を適合させるには、メニューから次の項目を選択します。

分析(A) > Regression > 線型...

図 9-43
[線型回帰] ダイアログ ボックス



- ▶ 「Daily ozone level 数量化」を従属変数として選択します。
- ▶ 「Inversion base height 数量化」、「Pressure gradient (mm Hg) 数量化」、および「Temperature (degrees F) 数量化」を独立変数として選択します。
- ▶ [OK] をクリックします。

図 9-44
最適尺度予測変数のサブセットを使用した回帰のモデルの要約

モデル集計

モデル	R	R2 乗	調整済み R2 乗	推定値の標準誤差
1	.856 ^a	.732	.729	4.16711

a. 予測値: (定数)、Temp 数量化、Dpg 数量化、lbh 数量化。

応答の数量化を使用すると、標準線型回帰での「Temperature」、「Pressure gradient」、および「Inversion base height」の適合度は 0.732 となります。この適合度を、これら 3 つの予測変数のみを使用したカテゴリ回帰分析の適合度と比較するには、[カテゴリ回帰] ダイアログ ボックスを再度表示します。

図 9-45
[カテゴリ回帰] ダイアログ ボックス



- ▶ 「Visibility (miles)」と「Day of the year」の独立変数としての選択を解除します。
- ▶ [OK] をクリックします。

図 9-46
3 つの予測変数でのカテゴリ回帰分析のモデルの要約

モデル集計				
	多重 R	R2 乗	調整済み R2 乗	見かけ上の予測誤差
標準化データ	.892	.796	.735	.204

従属変数 ozon
予測: ibh dpg temp

カテゴリ回帰分析の適合度は 0.796 であり、0.732 を上回っています。これにより、元の回帰における数量化で得られた尺度のプロパティは 5 つの変数すべてがモデル内に含まれている場合のみ最適であることが実証されます。

変換の効果

変数を変換すると、変換前は非線型であった応答と予測変数のセットとの関係が線型になります。ただし、複数の予測変数が存在する場合、モデル内の他の変数によりペアごとの関係が混乱します。

「Daily ozone level」と「Day of the year」の関係を分析することに重点を置くには、まず散布図を参照します。メニューから次の項目を選択します。

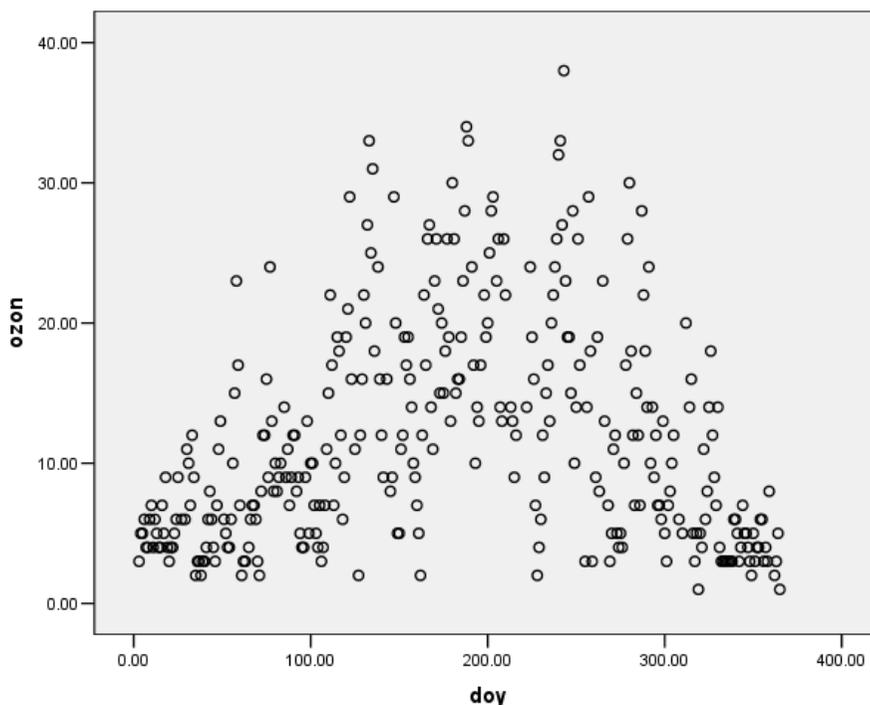
グラフ(G) > 図表ビルダー(C)...

図 9-47
[図表ビルダー] ダイアログ ボックス



- ▶ [散布図/ドット] ギャラリーを選択し、[単純散布図] を選択します。
- ▶ y 軸変数として「Daily ozone level」を選択し、x 軸変数として「Day of the year」を選択します。
- ▶ [OK] をクリックします。

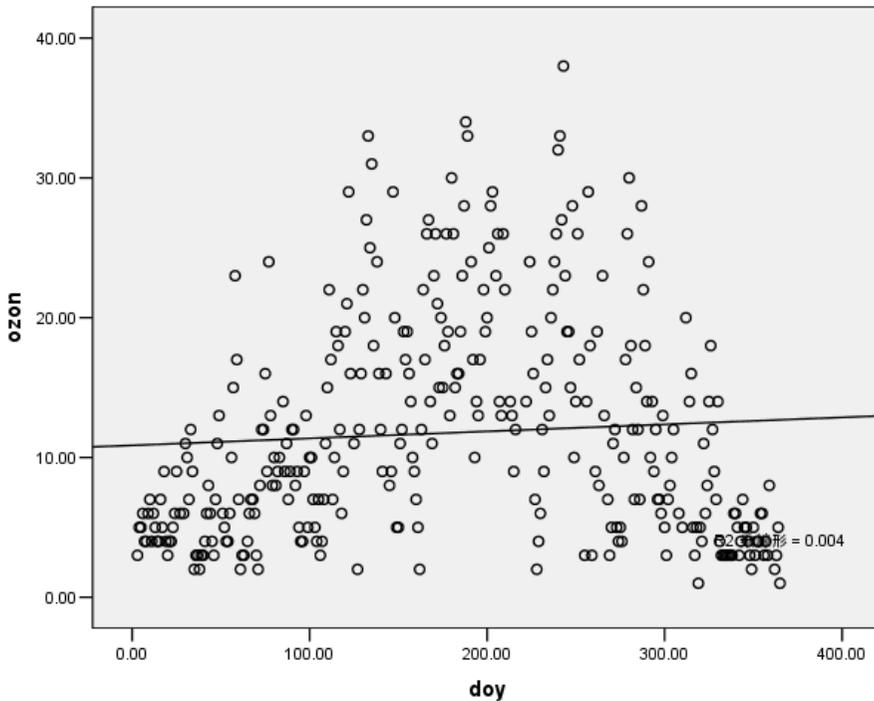
図 9-48
[Day of the year] に対する [Daily ozone level] の散布図



この図には、「Daily ozone level」と「Day of the year」の関係が示されています。「Day of the year」が200前後に達するまでは、その増加に従って、「Daily ozone level」も増加します。しかし、「Day of the year」の値が200を超えると、「Daily ozone level」は減少します。この反転したU字型のパターンは、2つの変数間の2次関係を示しています。線型回帰では、この関係を発見することはできません。

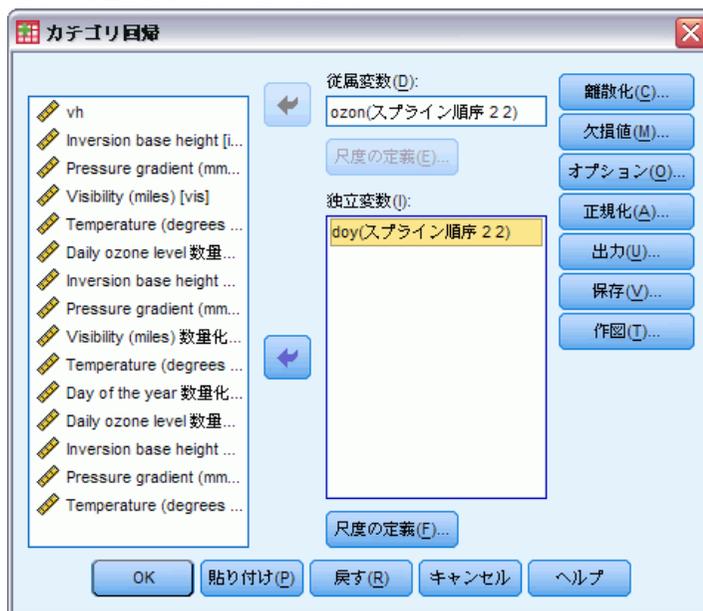
- ▶ 散布図の点を通る最適線を表示するには、グラフをダブルクリックしてアクティブにします。
- ▶ 図表エディタ内の点を選択します。
- ▶ [追加 合計での線の当てはめ] ツールをクリックし、図表エディタを閉じます。

図 9-49
最適な線の当てはめを表示する散布図



「Day of the year」に対する「Daily ozone level」の線型回帰を行うと、 R^2 の値は0.004になります。この適合度は、「Day of the year」が「Daily ozone level」に対して予測値を持たないことを示しています。これは、図に示されているパターンを考えると驚くべきことではありません。ただし、最適尺度法を使用することにより、2次関係を線型化したり、変換された「Day of the year」を使用して応答を予測できます。

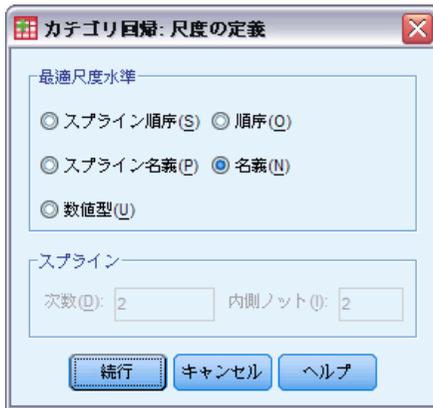
図 9-50
[カテゴリ回帰] ダイアログ ボックス



「Day of the year」に対する「Daily ozone level」のカテゴリ回帰分析を行うには、次の手順を実行して [カテゴリ回帰] ダイアログ ボックスを再表示します。

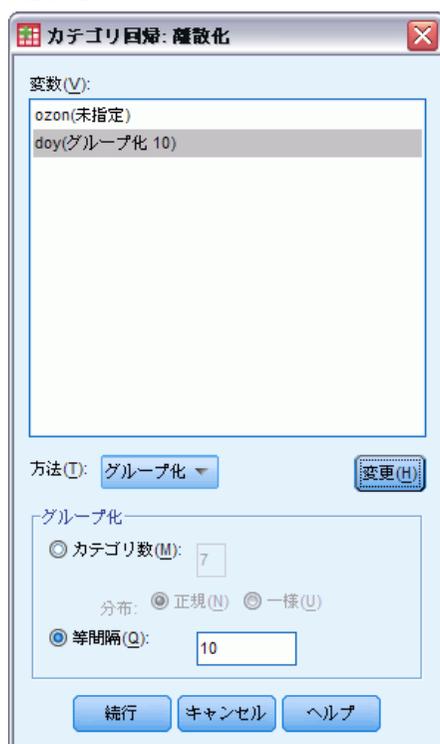
- ▶ 「Inversion base height」から「Temperature (degrees F)」までの独立変数としての選択を解除します。
- ▶ 「Day of the year」を独立変数として選択します。
- ▶ [尺度の定義] をクリックします。

図 9-51
[尺度の定義] ダイアログ ボックス



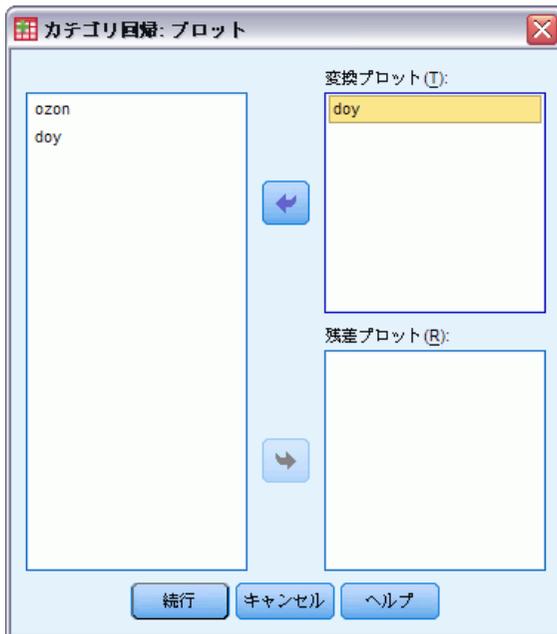
- ▶ 最適尺度水準として [名義] を選択します。
- ▶ [続行] をクリックします。
- ▶ [カテゴリ回帰] ダイアログ ボックスで [離散化] をクリックします。

図 9-52
[離散化] ダイアログ ボックス



- ▶ 「doy」を選択します。
- ▶ [等間隔]を選択します。
- ▶ 間隔の長さとして「10」と入力します。
- ▶ [変更]をクリックします。
- ▶ [続行]をクリックします。
- ▶ [カテゴリ回帰] ダイアログ ボックスで [作図] をクリックします。

図 9-53
[作図] ダイアログ ボックス



- ▶ 変換プロットとして「doy」を選択します。
- ▶ [続行] をクリックします。
- ▶ [カテゴリ回帰] ダイアログ ボックスで [OK] をクリックします。

図 9-54
[Day of the year] に対する [Daily ozone level] のカテゴリ回帰分析のモデルの要約

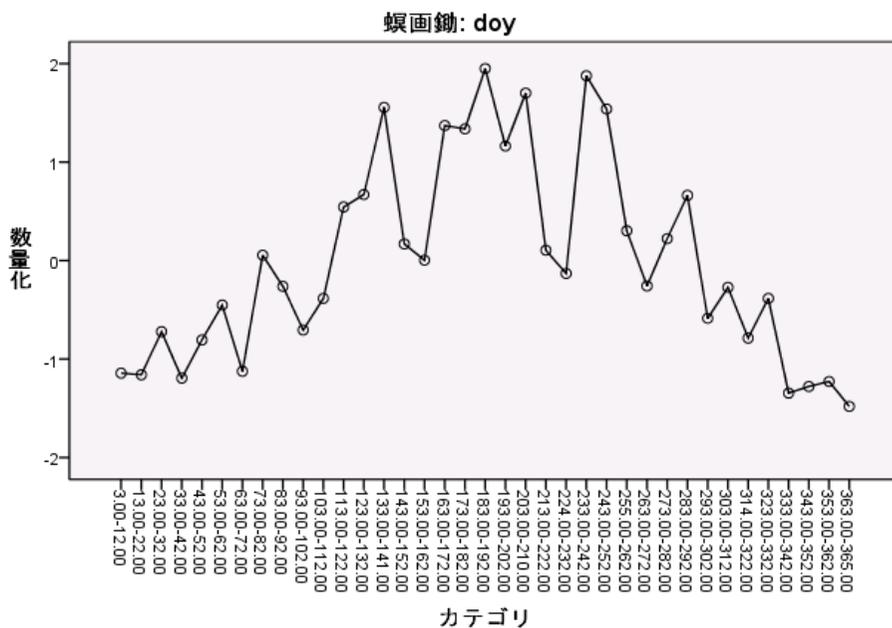
モデル集計

	多重 R	R ² 乗	調整済み R ² 乗	見かけ上の予測誤差
標準化データ	.741	.549	.494	.451

従属変数: ozon
予測変数: doy

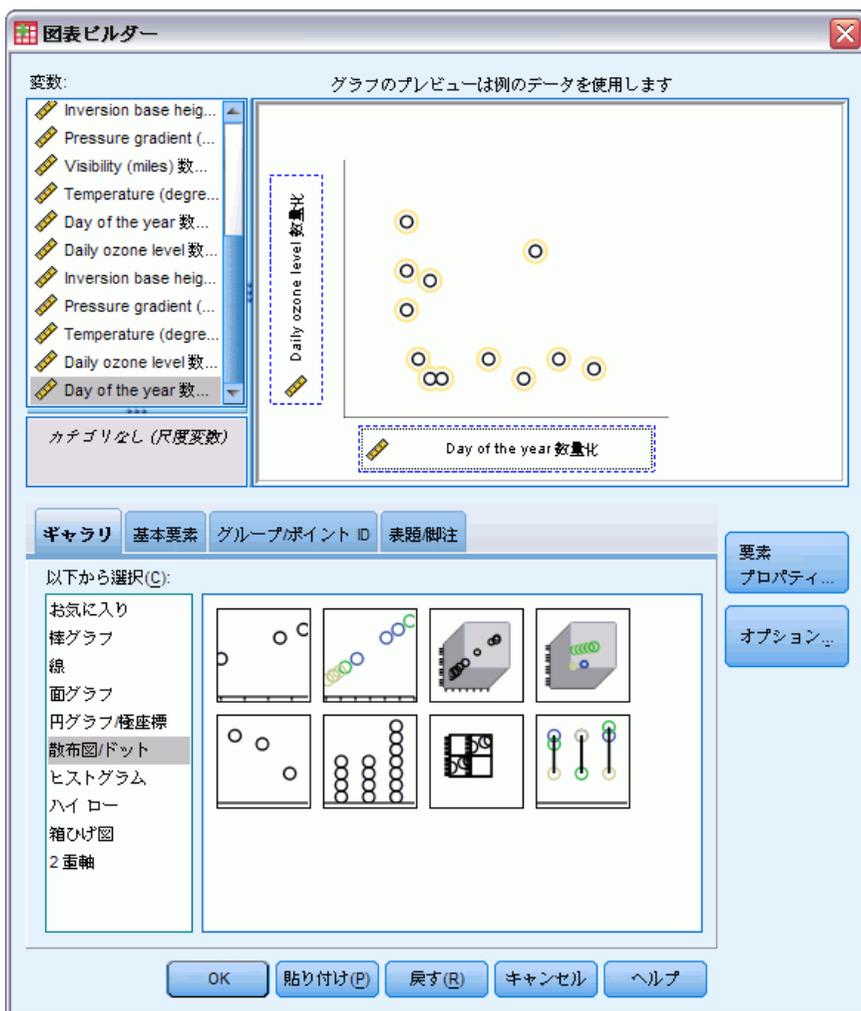
最適尺度法による回帰では、「Daily ozone level」が数値として扱われ、「Day of the year」が名義として扱われます。この結果、 R^2 の値は 0.549 になりました。カテゴリ回帰分析によって説明される「Daily ozone level」の分散はわずか 55% ですが、元の回帰よりは実質的に向上しています。「Day of the year」を変換すると、「Daily ozone level」を予測できます。

図 9-55
Day of the year (名義) の変換プロット



この図には、「Day of the year」の変換プロットが示されています。「Day of the year」の極値は、どちらも負の数量化を受けますが、中央値は正の数量化のみを受けます。この変換を適用すると、「Day of the year」の上限値と下限値は、予測された「Daily ozone level」に類似した効果をもたらします。

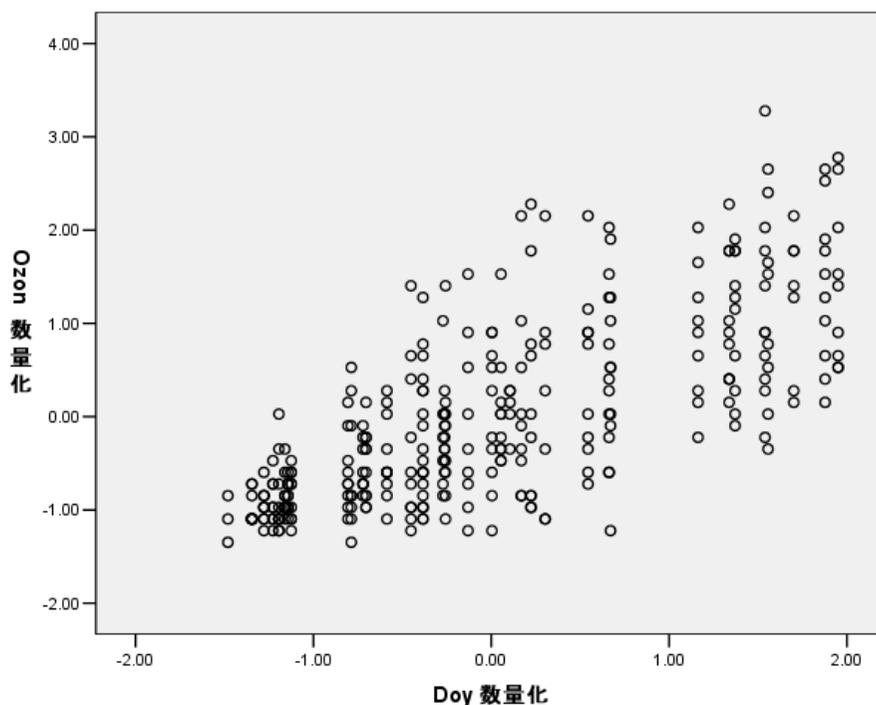
図 9-56
図表ビルダー



変換された変数の散布図を表示するには、[図表ビルダー] を再表示し、[リセット] をクリックして前回の選択内容をクリアします。

- ▶ [散布図/ドット] ギャラリーを選択し、[単純散布図] を選択します。
- ▶ y 軸変数として [Daily ozone level 数量化 [TRA1_3]] を選択し、x 軸変数として [Day of the year 数量化 [TRA2_3]] を選択します。
- ▶ [OK] をクリックします。

図 9-57
変換された変数の散布図



この図には、変換された変数間の関係が示されています。逆U字型は、増加トレンドに置き換えられます。回帰線は正の傾きを持っており、変換された [Day of the year] が増加すると、[Daily ozone level] の予測値が増加することを示しています。最適尺度法を使用すると、関係が線型化され、通常は気付かないような解釈が可能になります。

推奨参考文献

カテゴリ回帰分析の詳細は、次のテキストを参照してください。

Buja, A. 1990. Remarks on functional canonical variates, alternating least squares methods and ACE. *Annals of Statistics*, 18, .

Hastie, T., R. Tibshirani, および A. Buja. 1994. Flexible discriminant analysis. *Journal of the American Statistical Association*, 89, .

Hayashi, C. 1952. On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics*, 2, .

Kruskal, J. B. 1965. Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society Series B*, 27, .

Meulman, J. J. 2003. Prediction and classification in nonlinear data analysis: Something old, something new, something borrowed, something blue. *Psychometrika*, 4, .

Ramsay, J. O. 1989. Monotone regression splines in action. *Statistical Science*, 4, .

Van der Kooij, A. J., および J. J. Meulman. 1997. MURALS: Multiple regression and optimal scaling using alternating least squares. In: *Softstat '97*, F. Faulbaum, および W. Bandilla, 編集者. Stuttgart: Gustav Fisher.

Winsberg, S., および J. O. Ramsay. 1980. Monotonic transformations to additivity using splines. *Biometrika*, 67, .

Winsberg, S., および J. O. Ramsay. 1983. Monotone spline transformations for dimension reduction. *Psychometrika*, 48, .

Young, F. W., J. De Leeuw, および Y. Takane. 1976. Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features. *Psychometrika*, 41, .

カテゴリ主成分分析

カテゴリ主成分分析は、次元縮小の方法と見なすことができます。一連の変数を分析すると、変動の主な次元が明確になります。元のデータセットは、情報の損失を最小限にしつつ、より小さい新しいデータセットで置き換えることができます。この方法により、変数間、ケース間、およびその両方間の関係が明らかになります。

観測データを数量化する場合にカテゴリ主成分分析で使用される基準は、オブジェクト スコア (成分得点) が、数量化された各変数との大きな相関関係を持つ必要があることを示します。解はこの基準が満たされる範囲において適切になります。

カテゴリ主成分分析の 2 つの例を紹介します。最初の例では、手続きに関連付けられた基本的な概念および解釈を説明するのに役立つ少し小さいデータセットを採用します。2 つ目の例では、実際の応用について調査します。

例:社会システムの相互関係の調査

この例では、Bell (Bell, 1961) の表に対する Guttman (Guttman, 1968) の解釈を検証しています。また、このデータについては、Lingoes (Lingoes, 1968) も論じています。

Bell は、予想される社会グループを示す表を作成しました。Guttman は、この表の一部を使用しました。この表では、社会交相互作用、グループへの帰属感、メンバとの物理的な近接性、関係の形式化などを表す 5 個の変数が、理論上の 7 つの社会グループと交差しています。このグループには、観衆 (例、フットボールの試合の観戦者)、視聴者 (例、映画館または授業の参加者)、公衆 (例、新聞やテレビの視聴者)、暴徒 (観衆に似ているが、より強い交相互作用がある)、第一次集団 (親密な関係)、第二次集団 (自発的な集団)、および近代コミュニティ (物理的により密接した近接性と特化されたサービスの必要性によるゆるい同盟関係) があります。

次の表では、Guttman-Bell データで使用された 7 つの社会グループへの分類の結果として生じるデータセット内の変数を、各変数のレベルに関連する変数ラベルおよび値ラベル (カテゴリ) と一緒に示しています。このデータセットは `guttman.sav` にあります。詳細は、A 付録 p. 321 [サンプルファイル](#) を参照してください。カテゴリ主成分分析の計算に含める変数を

選択する以外に、プロット内のオブジェクトにラベルを付ける際に使用される変数を選択できます。この例では、データ内の最初の 5 個の変数は分析に含まれますが、cluster はラベル付け変数として単独で使用されます。カテゴリ主成分分析を指定する場合、分析変数ごとに最適尺度水準を指定する必要があります。この例では、すべての分析変数に対して順序レベルが指定されています。

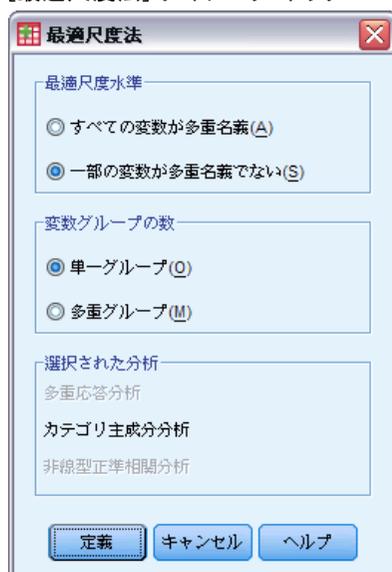
テーブル 10-1
Guttman-Bell データセットの変数

変数名	変数ラベル (L)	値ラベル
intnsity	相互作用の強さ	非常に弱い、弱い、適度、強い
frquency	相互作用の頻度	わずか、繰り返しがない、少ない、頻繁
blonging	グループへの所属感覚	ない、わずか、変化する、高い
proxmity	物理的的近接性	隔たりがある、緊密
formlity	関係性の公私	関係性なし、公式、非公式
cluster		CR、AU、PU、MB、PG、SG、MC

分析の実行

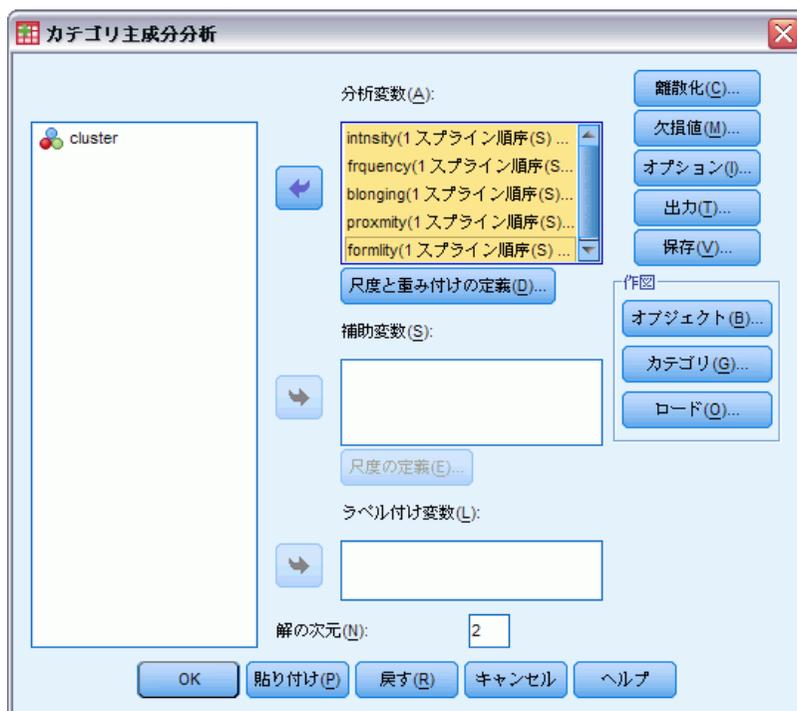
- ▶ このデータセットのカテゴリ主成分分析の出力を作成するには、メニューから次の項目を選択します。
分析(A) > 次元分解 > 最適尺度法...

図 10-1
[最適尺度法] ダイアログ ボックス



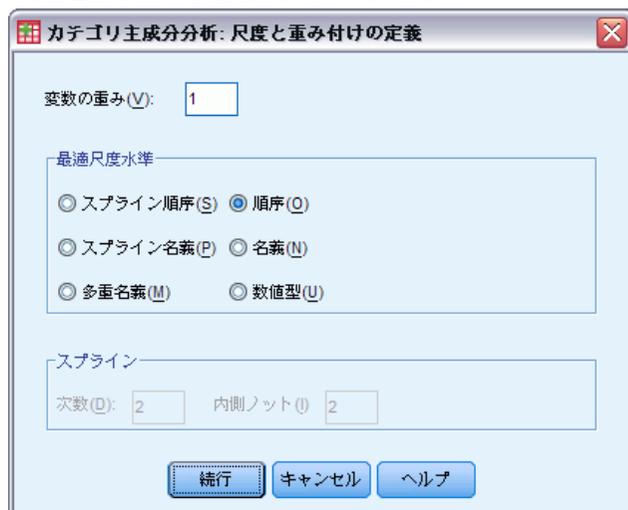
- ▶ [最適尺度水準] グループから [一部の変数が多重名義でない] を選択します。
- ▶ [定義] をクリックします。

図 10-2
[カテゴリ主成分分析] ダイアログ ボックス



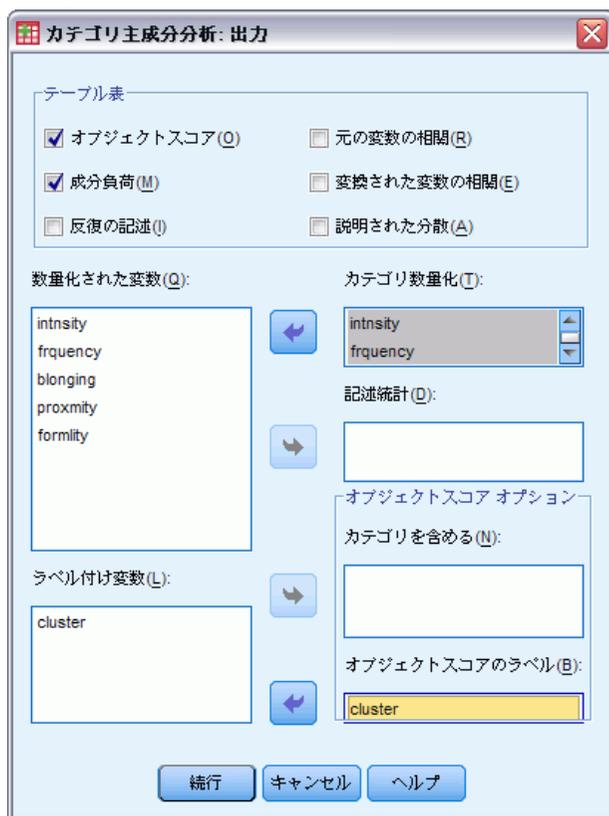
- ▶ 「intensity」から「formality」までを分析変数として選択します。
- ▶ [尺度と重み付けの定義] をクリックします。

図 10-3
[尺度と重み付けの定義] ダイアログ ボックス



- ▶ [最適尺度水準] グループで [順序] を選択します。
- ▶ [続行] をクリックします。
- ▶ [カテゴリ主成分分析] ダイアログ ボックスで、ラベル付け変数として「cluster」を選択します。
- ▶ [出力] をクリックします。

図 10-4
[出力] ダイアログ ボックス



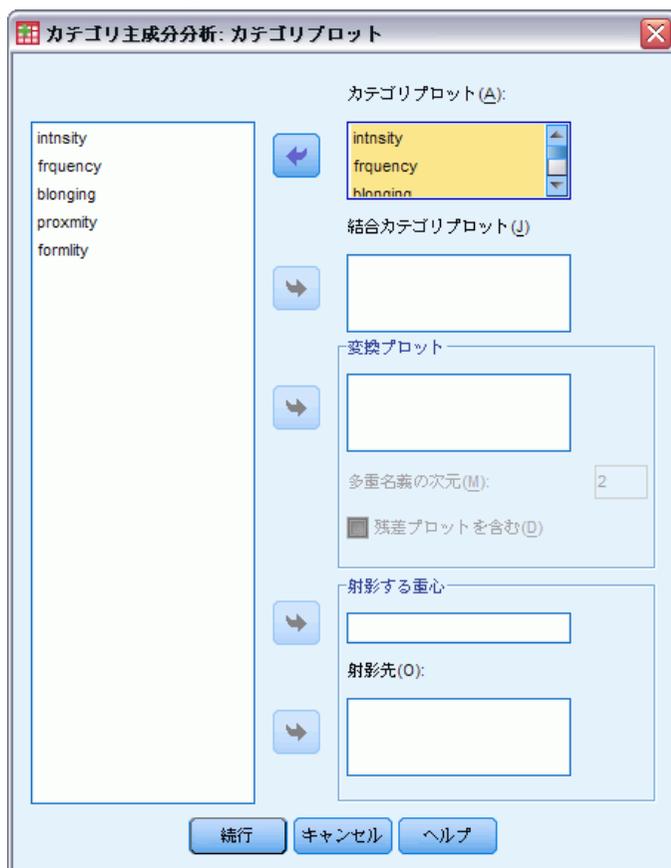
- ▶ [テーブル] グループで、[オブジェクトスコア] を選択し、[変換された変数の相関] の選択を解除します。
- ▶ 「intnsity」から「formlity」まで、カテゴリ数量化を行うように選択します。
- ▶ [オブジェクトスコアのラベル] として「cluster」を選択します。
- ▶ [続行] をクリックします。
- ▶ [カテゴリ主成分分析] ダイアログ ボックスの [作図] グループで、[オブジェクト] をクリックします。

図 10-5
[オブジェクトと変数プロット] ダイアログ ボックス



- ▶ [作図] グループで、[オブジェクトと変数 (バイプロット)] を選択します。
- ▶ [オブジェクトのラベル付け] グループで、オブジェクトのラベル付けに [変数] を選択し、オブジェクトにラベルを付ける変数として「cluster」を選択します。
- ▶ [続行] をクリックします。
- ▶ [カテゴリ主成分分析] ダイアログ ボックスの [作図] グループで、[カテゴリ] をクリックします。

図 10-6
[カテゴリ プロット] ダイアログ ボックス



- ▶ 「intnsity」から「formlity」まで、結合カテゴリ プロットを作成するように選択します。
- ▶ [続行] をクリックします。
- ▶ [カテゴリ主成分分析] ダイアログ ボックスで、[OK] をクリックします。

次元数

次の図では、カテゴリ主成分分析の初期出力の一部を示しています。アルゴリズムの反復の記述の後、各次元の固有値を含む、モデルの要約が表示されます。これらの固有値は、古典的な主成分分析の固有値に相当します。固有値は、次元ごとに分散が占める割合を示す測定値です。

図 10-7
反復の記述(Y)

反復回数	説明された分散		損失		
	合計	増加	合計	重心座標	重心のベクトル座標への制限
0 ^a	4.515315	.000000	5.484685	4.075583	1.409101
31 ^b	4.726009	.000008	5.273991	4.273795	1.000196

- a. 反復が 0 回だと、数値として取り扱われる、最適尺度水準多重名義を持つ変数を除く、すべての変数による解の統計量が表示されます。
b. 収束テスト値に達したため、反復処理を終了しました。

図 10-8
モデルの要約(M)

次元	Cronbach のアルファ	説明された分散	
		合計(固有値)	分散の%
1	.881	3.389	67.774
2	.315	1.337	26.746
合計	.986 ^a	4.726	94.520

- a. Cronbach のアルファ合計は、固有値合計に基づいています。

固有値は、必要な次元数を示す指標として使用できます。この例では、デフォルトの次元数 2 を使用しました。これは適切な数値なのでしょう。一般に、すべての変数が単一名義、順序、数値のいずれかになる場合、次元の固有値を 1 より大きくする必要があります。2 次元の解は分散の 94.52% を占めているので、おそらく、3 番目の次元ではそれほど多くの情報は追加されません。

多重名義変数の場合、適切な次元数を決定するための簡単な目安はありません。変数の数を、カテゴリの総数から変数の数を引いた数に置き換えると、上記の規則がそのまま適用されます。ただし、必要以上の次元が許可されるのは、この規則だけと考えられます。次元数を選択する場合の最も有用なガイドラインは、意味のある解釈が可能になる小さな数をそのまま使用することです。また、モデルの要約表では、Cronbach のアルファ（信頼度）も示します。これは、手続きで最大化されます。

数量化 (カテゴリ / カテゴリの回帰)

変数ごとに、各次元の数量化、ベクトル座標、および重心の座標が表示されます。数量化は、各カテゴリに割り当てられる値です。重心の座標は、同一カテゴリにあるオブジェクトのオブジェクト スコアの平均です。ベクトル座標は、カテゴリを同じ水準にする必要がある場合のカテゴリの座標で、対象空間の変数を表します。これは、尺度水準が順序および数値の変数に必要です。

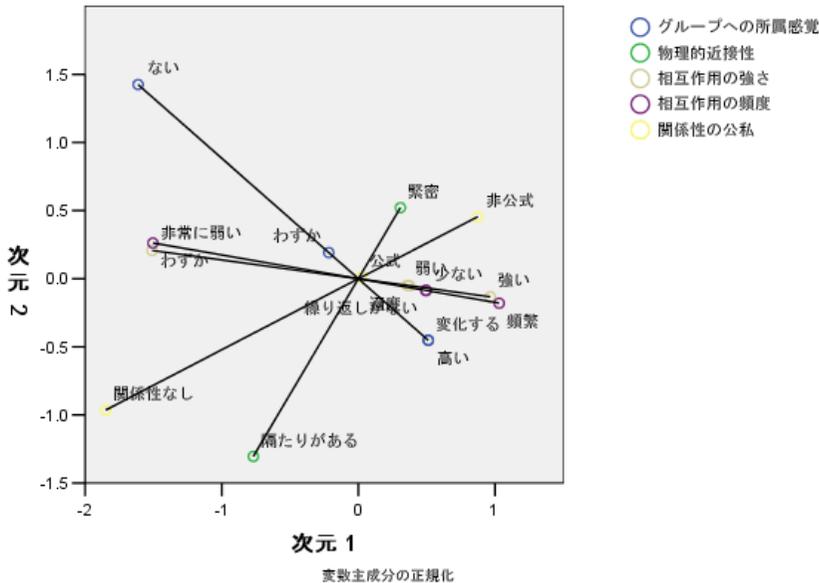
図 10-9
交互作用の強度の数量化

カテゴリ	度数	数量化	重心座標		ベクトル座標	
			次元		次元	
			1	2	1	2
非常に弱い	2	-1.530	-1.496	.308	-1.510	.208
弱い	2	.362	.392	.202	.358	-.049
適度	1	.379	.188	-1.408	.374	-.051
強い	2	.978	1.010	.194	.965	-.133

変数主成分の正規化

カテゴリ ポイントの結合プロット内の数量化を確認すると、カテゴリ主成分分析では一部の変数のカテゴリが、水準が正しい順序だった場合に予測されるほど明確には区別されなかったことがわかります。たとえば、変数「intensity」および「frequency」は、中央にある2つのカテゴリの数量化と等しいか、またはほぼ等しくなります。このような結果では、集約された一部のカテゴリ、または(多重)名義などの別の水準の分析で、別のカテゴリ主成分分析を行うように示されることがあります。

図 10-10
結合プロットのカテゴリ ポイント



カテゴリ ポイントの結合プロットは、成分負荷のプロットに似ていますが、最低の数量化に対応する終点の場所も示します(たとえば、「intensity」の場合は [slight]、「belonging」の場合は [none])。交互作用を測定する2つの変数 [intensity] および [frequency] は、互いに近

くに表示され、次元 1 の分散の多くを占めています。また、[formality] も [proximity] の近くに表示されます。

カテゴリ ポイントに注目することで、関係がさらに明確になります。

「intensity」と「frequency」は近いだけではなく、尺度の方向が類似しています。つまり、低い強度は低い頻度に近くなり、交互作用の頻度は高い交互作用の強度に近くなります。また、物理的に接近することは、形式張らない関係と密接な関係があるように見えることがわかります。物理的な距離に関連する関係はありません。

オブジェクト スコア

オブジェクト スコアの一覧とプロットも要求できます。オブジェクト スコアのプロットは、外れ値の検出、一般的なオブジェクトのグループの検出、または一部の特殊なパターンの明示に使用できます。

オブジェクト スコアの表では、Guttman-Bell データの社会グループでラベル付けしたオブジェクト スコアの一覧が示されます。オブジェクト ポイントの値を調査することにより、プロット内の特定のオブジェクトを識別できます。

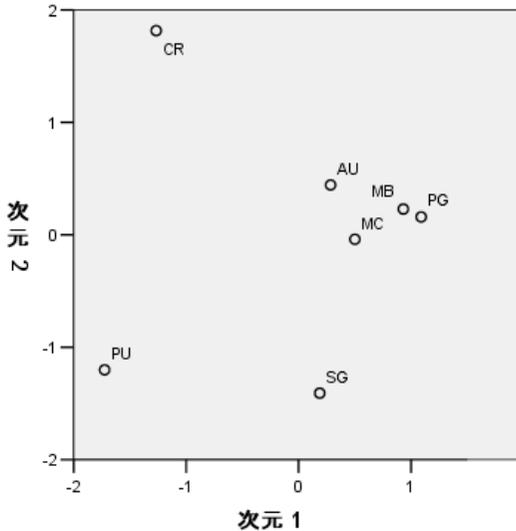
図 10-11
オブジェクトスコア(B)

クラス	次元	
	1	2
CR	-1.266	1.816
AU	.284	.444
PU	-1.726	-1.201
MB	.931	.229
PG	1.089	.159
SG	.188	-1.408
MC	.500	-.039

変数主成分の正規化

最初の次元では、比較的大きな負の得点を持つ「CR」と「PU」が、比較的大きな正の得点を持つ「MB」と「PG」とは区別されているように見えます。2 番目の次元には、3 つの集合があります。大きい負の値を持つ「PU」と「SG」、大きい正の値を持つ「CR」、その間には他の社会グループがあります。オブジェクト スコアのプロットを検査することにより、わかりやすくなります。

図 10-12
オブジェクト スコア プロット

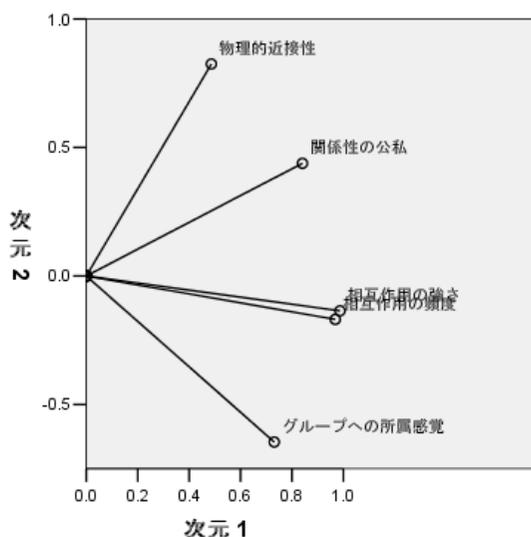


プロットでは、「PU」および「SG」が一番下に、「CR」が一番上に、他の社会グループが中央に表示されます。個々のオブジェクト間のパターンの調査は、分析の単位に使用できる追加情報によって異なります。この場合、オブジェクトの分類がわかっています。そうでない場合は、補助変数を使用して、オブジェクトにラベルを付けることができます。また、カテゴリ主成分分析では、「MB」が「PG」から区別されないこともわかります。一般的には家族を暴徒とは見なしません。使用した変数では、この2つのグループは、5個の変数のうち4個の変数の得点が同じでした。当然、使用した変数およびカテゴリで考えられる欠点を見つけ出すことが必要になります。たとえば、交互作用の強度および形式張らない関係の強度が高いことは、この2つのグループでは別のことを意味します。これらの代わりに、より高い次元の解を検討してもよいでしょう。

成分負荷

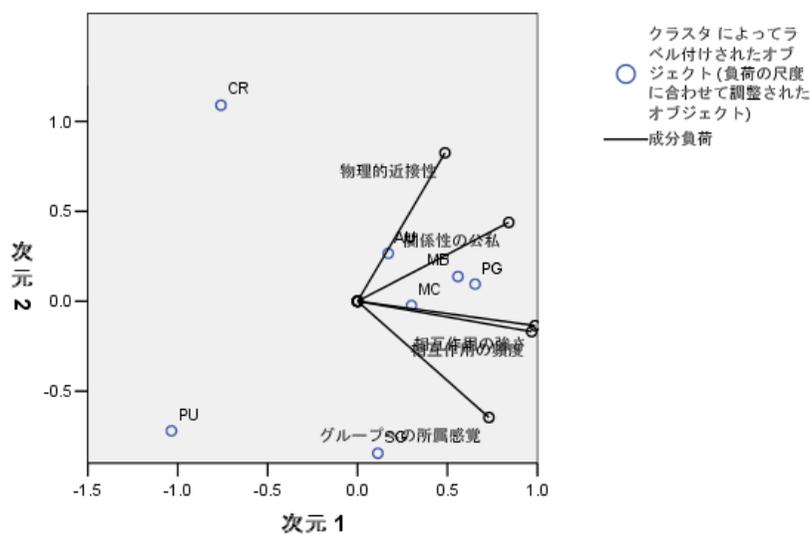
次の図では、成分負荷のプロットを示します。ベクトル（行）は比較的長く、数量化されたすべての変数の分散の大半を、最初の2つの次元が占めていることを再度示しています。最初の次元では、すべての変数に高い（正の）成分負荷があります。2番目の次元は、主に、数量化された変数「グループへの所属感覚」と「物理的近接性」とは反対方向の相関関係にあります。つまり、次元2で負の値が大きい得点を持つオブジェクトは、feeling of belonging（帰属感）に高得点、物理的な physical proximity（近接性）に低得点を持ちます。したがって、2番目の次元は、この2つの変数の対比を明確にしますが、数量化された変数「intensity」および「frequency」との関係はほとんどありません。

図 10-13
成分負荷(M)



オブジェクトと変数間の関係を調べるには、オブジェクトと成分負荷のバイプロットを確認します。変数のベクトルは、変数の最高カテゴリの方向を示しています。たとえば、「proximity」と「belonging」の場合、最高カテゴリは、それぞれ、[緊密]と[高い]となります。したがって、「CR」は、高い物理的近接性があり、帰属感がないことで特徴付けられ、「SG」は、物理的な距離が離れており、帰属感が高いことで特徴付けられています。

図 10-14
バイプロット(B)



追加の次元

次元数を増やすと、変動の占める量が増加するため、より低次元の解に隠れている相異が明らかになることがあります。すでに説明したとおり、2つの次元では、「MB」および「PG」を区別できません。しかし、次元数を増やすと、2つのグループを区別できる場合があります。

分析の実行

- ▶ 3次元の解を取得するには、もう一度、[カテゴリ主成分分析] ダイアログボックスを開きます。
- ▶ 解の次元の数として「3」と入力します。
- ▶ [カテゴリ主成分分析] ダイアログボックスで、[OK] をクリックします。

モデルの要約

図 10-15
モデルの要約(M)

次元	Cronbach のアルファ	説明された分散	
		合計 (固有値)	分散の %
1	.885	3.424	68.480
2	-.232	.844	16.871
3	-.459	.732	14.649
合計	1.000 ^a	5.000	99.999

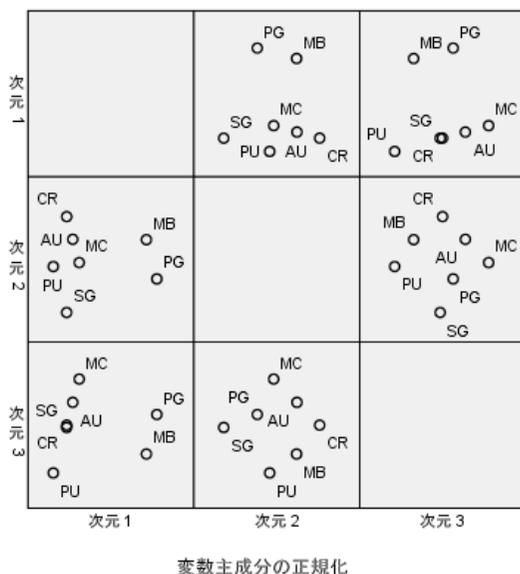
a. Cronbach のアルファ合計は、固有値合計に基づいています。

3次元の解には、ほぼすべての分散を占める、3.424、0.844、および0.732という固有値があります。

オブジェクト スコア

3次元の解のオブジェクト スコアは、散布図の行列にプロットされます。散布図の行列では、一連の2次元の散布図にある他のすべての次元に対して、すべての次元がプロットされます。3つの次元にある最初の2つの固有値は、2次元の解の固有値と等しいことに注意してください。つまり、解は入れ子になりません。次元2および次元3の固有値は1より小さいので（負のCronbachのアルファを指定）、2次元の解が望まれます。3次元の解は、説明の目的で含まれています。

図 10-16
3次元のオブジェクトスコアの散布図の行列



プロットの一番上の行は、最初の次元では、「PG」および「MB」が他のグループとは区別されていることを示します。縦軸に沿ったオブジェクトの順序は、一番上の行のどのプロットでも変化しないことに注意してください。そのため、各プロットでは、y 軸として次元 1 を採用します。

プロットの中央の行では、次元 2 の解釈が可能になります。2 番目の次元は、2 次元の解からわずかに変化しています。以前は 2 番目の次元には 3 つの異なる集合がありましたが、現在、オブジェクトは、軸に沿って拡散されています。

3 番目の次元は、「PG」と「MB」を区別するのに役立ちます。これは、2 次元の解では行われません。

次元 2 と次元 3、次元 1 と次元 2 のプロットの比較をよく確認してください。次元 2 および次元 3 で定義された平面上で、オブジェクトは、頂点に「CR」、「MC」、「SG」、「PU」を持ついびつな四辺形を構成します。この平面上で、「MB」と「PG」は、それぞれ、「PU-CR」および「SG」-「MC」という凸結合のように見えます。ただし、すでに説明したとおり、これらは次元 1 上では他のグループから区別されています。

「AU」は、次元 1 上では他のグループから区別されず、「CR」と「MC」の結合のように見えます。

成分負荷

図 10-17
3次元の成分負荷

	次元		
	1	2	3
相互作用の強さ	.980	-.005	-.201
相互作用の頻度	.521	-.643	.561
グループへの所属感覚	.980	-.002	-.197
物理的近接性	.519	.656	.549
関係性の公私	.981	.004	-.193

オブジェクトがどのように区別されているかわかっても、どの変数がどの次元に対応するかは明らかになりません。これは、成分負荷を使用することにより解決します。最初の次元は、主に、「グループへの所属感覚」、「相互作用の強さ」、および「関係性の公私」に対応します。2番目の次元は、「相互作用の頻度」と「物理的近接性」を区別します。3番目の次元は、これらの変数を他の変数と区別します。

例:摂食障害の症状

摂食障害とは、摂食行動の障害、ボディイメージの重度の歪み、心身両方に影響を与える体重への執着に関連する消耗性疾患です。毎年、大勢の人々が影響を受け、特に若者には危険が伴います。複数の治療法が使用可能であり、その多くは、病気が早期に発見された場合に役立ちます。

医療専門家は、心理学鑑定および医学的評価を通じて、摂食障害の診断を行うことができます。ただし、拒食行動または過食行動の標準的な症状はないので、摂食障害のさまざまな異なるクラスのいずれかに患者を割り当てることは困難です。患者を4つのグループに明確に区分する症状はあるのでしょうか。どの症状が4つのグループで共通するのでしょうか。

このような質問に回答するために、調査員 (Van der Ham, Meulman, Van Strien, および Van Engeland, 1997) は、次の表に示すとおり、摂食障害の傾向のある55人の若者について調査を実施しました。

テーブル 10-2
患者の診断

分析	患者数
拒食症	25
過食症を併発した拒食症	9
拒食症後の過食症	14
特殊な摂食障害	7
合計	55

各患者が4年間で4回診察を受けたので、観測値は合計で220になりました。各観測では、患者は、16の各症状の得点が記録されました。各症状については、次の表で説明します。患者71(2回目)、患者76(2回目)、患者47(3回目)の症状の得点が見つからなかったため、残っている217回分の観測値が有効です。このデータは anorectic.sav にあります。詳細は、A付録 p.321 [サンプルファイル](#) を参照してください。

テーブル 10-3
健康を測定する変更された Morgan-Russell 下位尺度

変数名	変数ラベル(L)	下限(得点1)	上限(得点3または4)
weight	体重	標準範囲外	正規(N)
mens	月経	無月経	定期的
fast	食事量の制限	1200 カロリ未満	標準/通常の食事
binge	過食症	1週間に2回以上	なし
vomit	嘔吐	1週間に2回以上	なし
purge	下痢	1週間に2回以上	なし
hyper	多動性、衝動性	静止してられない	なし
fami	家族関係	悪い	良い
eman	家族の束縛からの解放	依存性が高い	十分
frie	友人	親友がいない	2人以上の親友がいる
school	学校/職歴	退学/退職	並みの成績から良い成績
satt	性的傾向態度	不十分	十分
sbeh	性的行動	不十分	性行為を楽しむことができる
mood	精神状態(気分)	ひどく落ち込んでいる	正規(N)
preo	食べ物と体重に対する関心	完全	なし
body	体知覚	異常	正規(N)

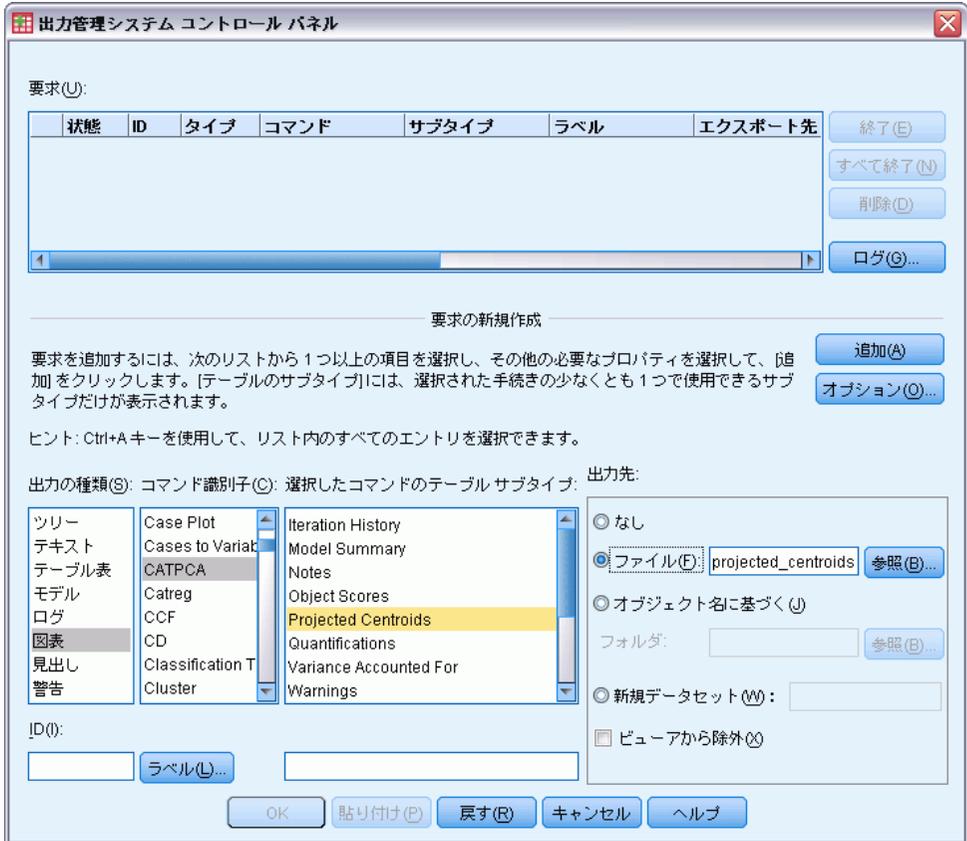
調査は、摂食障害の症状とさまざまなクラス間の関係を解明することを目的としているので、この場合は、主成分分析が理想的です。さらに、症状には順序尺度で得点が付けられるので、カテゴリ主成分分析は、古典的な主成分分析よりも役に立つ場合があります。

分析の実行

診断ごとに病気の経過の構成を正確に調査するために、散布図用のデータとして使用できる、射影された重心のテーブルの結果を作成します。これは、出力管理システムを使用することによって可能になります。

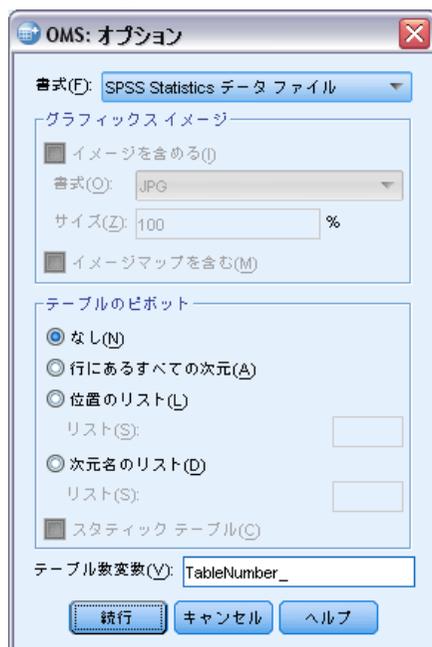
- ▶ OMS 要求を開始するには、メニューから次の項目を選択します。
ユーティリティ > OMS コントロール パネル...

図 10-18
出力管理システム コントロール パネル



- ▶ 出力の種類として [テーブル表] を選択します。
- ▶ コマンドとして [CATPCA] を選択します。
- ▶ テーブルの種類として [Projected Centroids] を選択します。
- ▶ [出力先] グループの [ファイル] を選択し、ファイル名として「projected_centroids.sav」と入力します。
- ▶ [オプション] をクリックします。

図 10-19
[オプション] ダイアログ ボックス



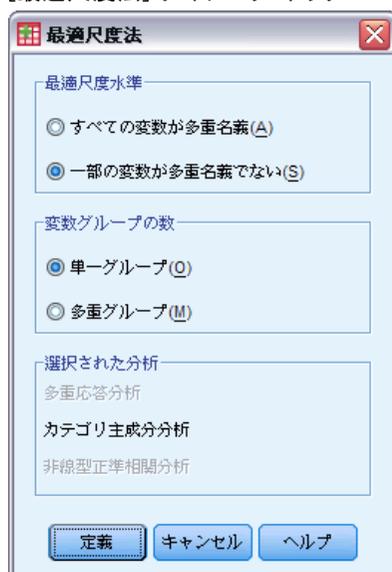
- ▶ 出力形式として [IBM® SPSS® Statistics データ ファイル] を選択します。
- ▶ テーブル数変数として「TableNumber_1」と入力します。
- ▶ [続行] をクリックします。

図 10-20
出力管理システム コントロール パネル



- ▶ [追加] をクリックします。
- ▶ [OK] をクリックし、さらに OMS セッションを確認するための [OK] をクリックします。
この段階で、出力管理システムは、射影された重心のテーブルの結果を projected_centroids.sav ファイルに書き込むように設定されます。
- ▶ このデータセットのカテゴリ主成分分析の出力を作成するには、メニューから次の項目を選択します。
分析(A) > 次元分解 > 最適尺度法...

図 10-21
[最適尺度法] ダイアログ ボックス



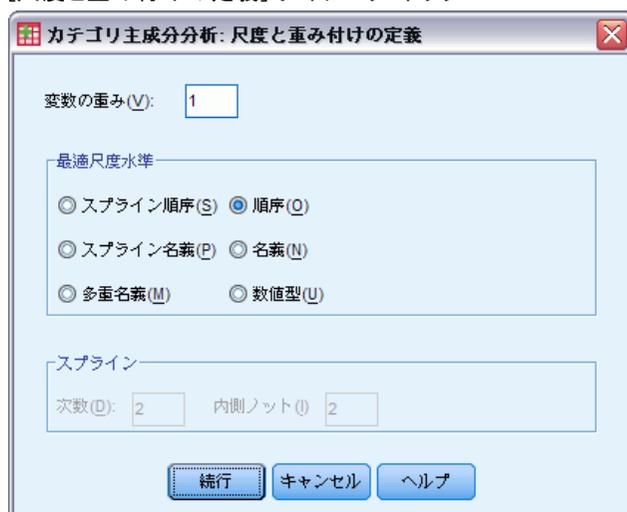
- ▶ [最適尺度水準] グループから [一部の変数が多重名義でない] を選択します。
- ▶ [定義] をクリックします。

図 10-22
[カテゴリ主成分分析] ダイアログ ボックス



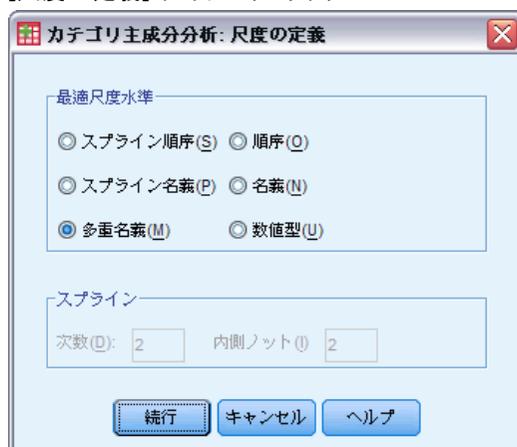
- ▶ 分析変数として「体重」から「体知覚」までを選択します。
- ▶ [尺度と重み付けの定義] をクリックします。

図 10-23
[尺度と重み付けの定義] ダイアログ ボックス



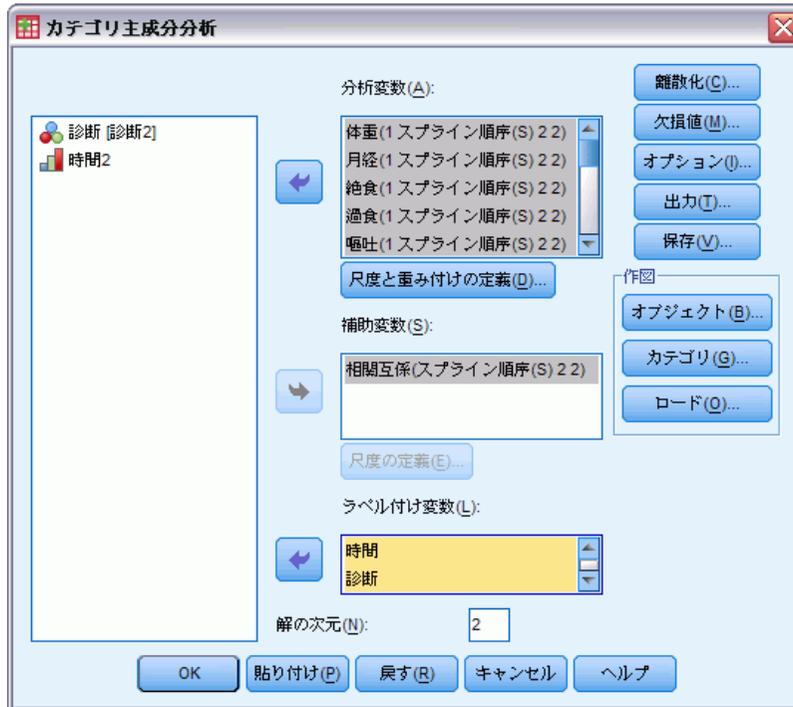
- ▶ 最適尺度水準として [順序] を選択します。
- ▶ [続行] をクリックします。
- ▶ 補助変数として「時間/診察の相互関係」を選択し、[カテゴリ主成分分析] ダイアログ ボックスの [尺度の定義] をクリックします。

図 10-24
[尺度の定義] ダイアログ ボックス



- ▶ 最適尺度水準として [多重名義] を選択します。
- ▶ [続行] をクリックします。

図 10-25
[カテゴリ主成分分析] ダイアログ ボックス



- ▶ ラベル付け変数として「インタビューの時間」から「患者番号」までを選択します。
- ▶ [オプション] をクリックします。

図 10-26
[オプション] ダイアログ ボックス

カテゴリ主成分分析: オプション

補助オブジェクト

ケースの範囲 (N)
最初 (F):
最後 (S):

単一ケース (I):

追加 (A)
変更 (C)
除去 (R)

正規化 (Z)
変数主成分
ユーザー指定の値 (T):

基準
収束基準 (V):
最大反復回数 (X):

作図のラベル
 変数ラベルまたは値ラベル (R)
ラベル長の限界 (L):
 変数名または値 (M)

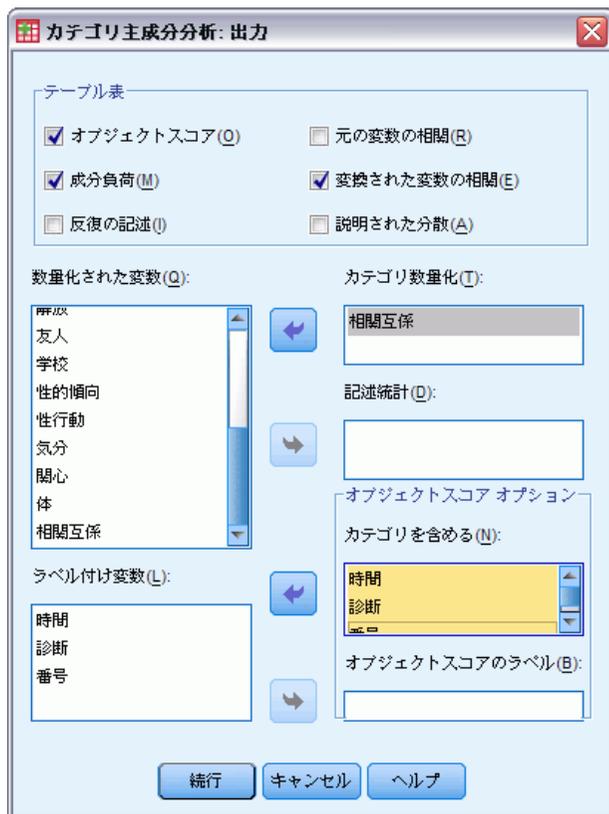
次元のプロット
 解のすべての次元の表示 (P)
 次元数の制限 (B)
最低次元 (Q):
最高次元 (D):

布置 (G)
なし
ファイル (F)...

続行 キャンセル ヘルプ

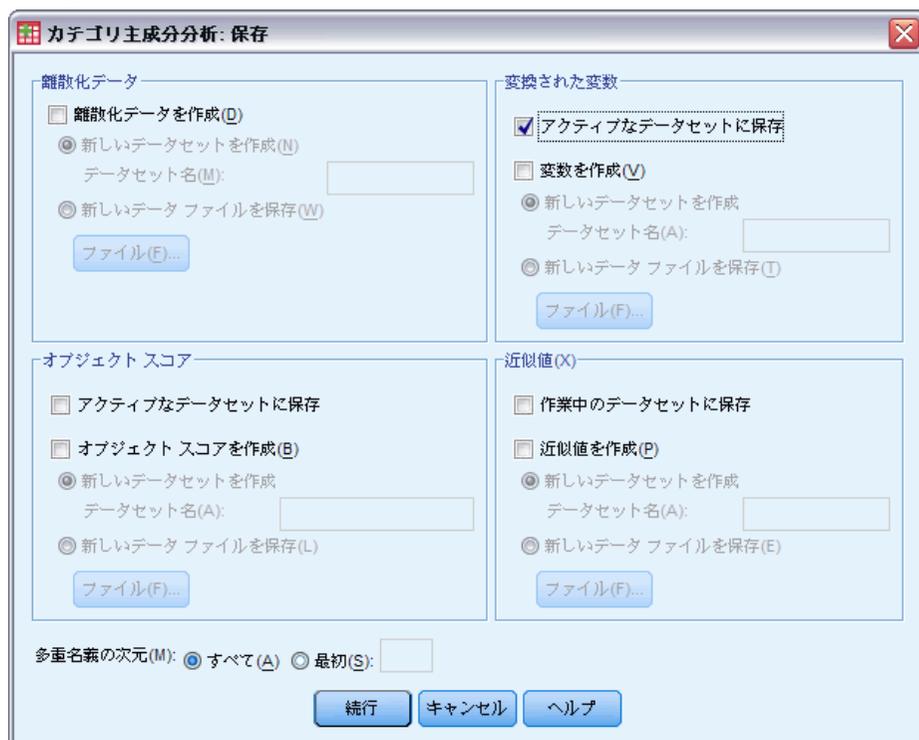
- ▶ プロットのラベル付けに [変数名または値] を選択します。
- ▶ [続行] をクリックします。
- ▶ [カテゴリ主成分分析] ダイアログ ボックスの [出力] をクリックします。

図 10-27
[出力] ダイアログ ボックス



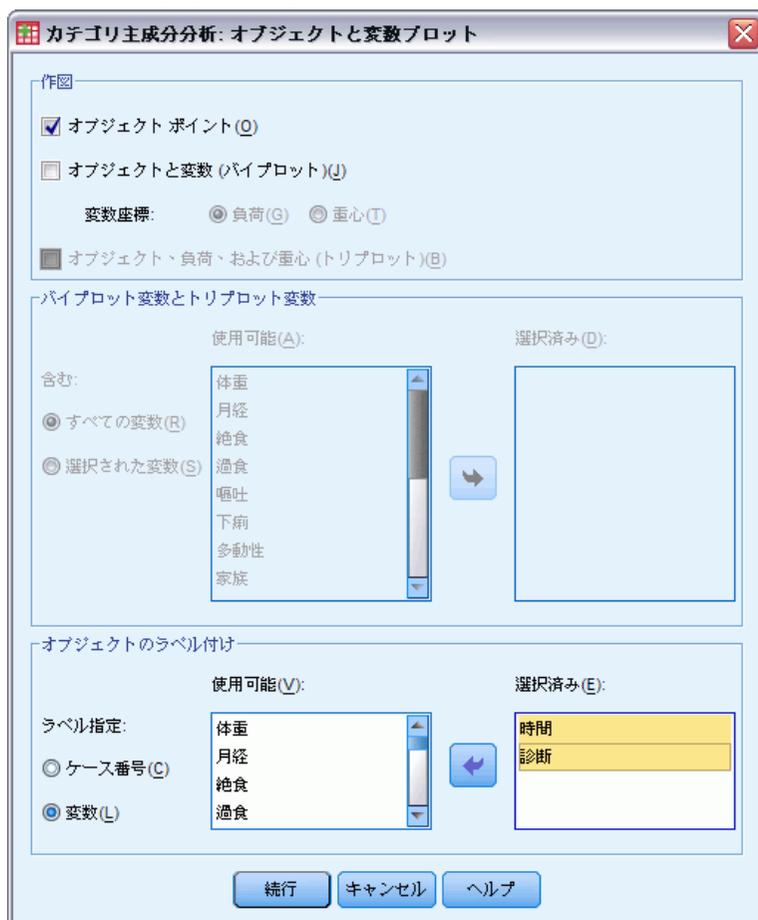
- ▶ [テーブル] グループで、[オブジェクトスコア] を選択します。
- ▶ 「相互関係」のカテゴリ数量化を要求します。
- ▶ 「時間」、「診断」、および「番号」のカテゴリを含めるように選択します。
- ▶ [続行] をクリックします。
- ▶ [カテゴリ主成分分析] ダイアログ ボックスの [保存] をクリックします。

図 10-28
[保存] ダイアログ ボックス



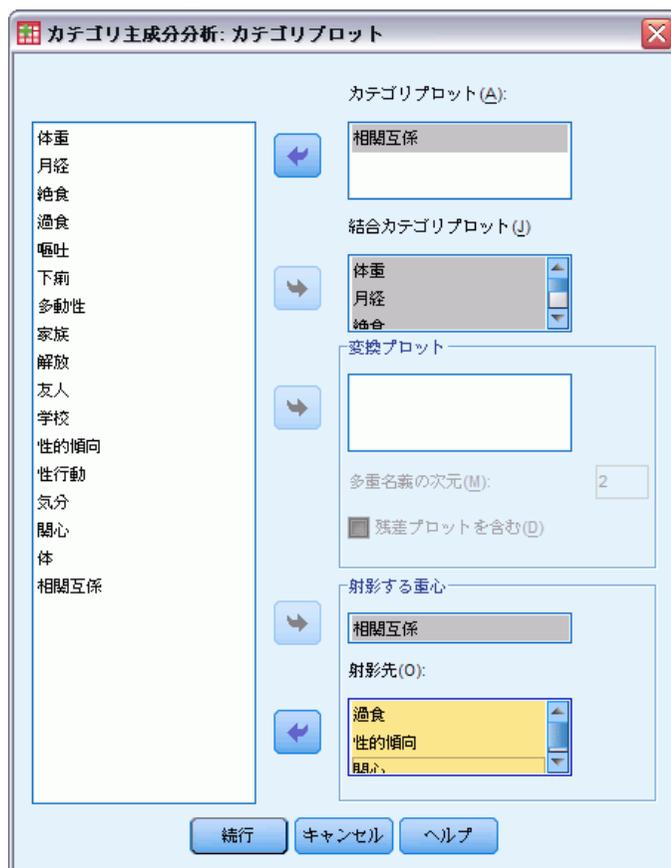
- ▶ [変換された変数] グループで、[アクティブなデータセットに保存] を選択します。
- ▶ [続行] をクリックします。
- ▶ [カテゴリ主成分分析] ダイアログ ボックスで、[オブジェクト] をクリックします。

図 10-29
[オブジェクトと変数プロット] ダイアログ ボックス



- ▶ [オブジェクトのラベル付け] で [変数] を選択します。
- ▶ オブジェクトにラベルを付ける変数として「時間」および「診断」を選択します。
- ▶ [続行] をクリックします。
- ▶ [カテゴリ主成分分析] ダイアログ ボックスで、[カテゴリ] をクリックします。

図 10-30
[カテゴリ プロット] ダイアログ ボックス

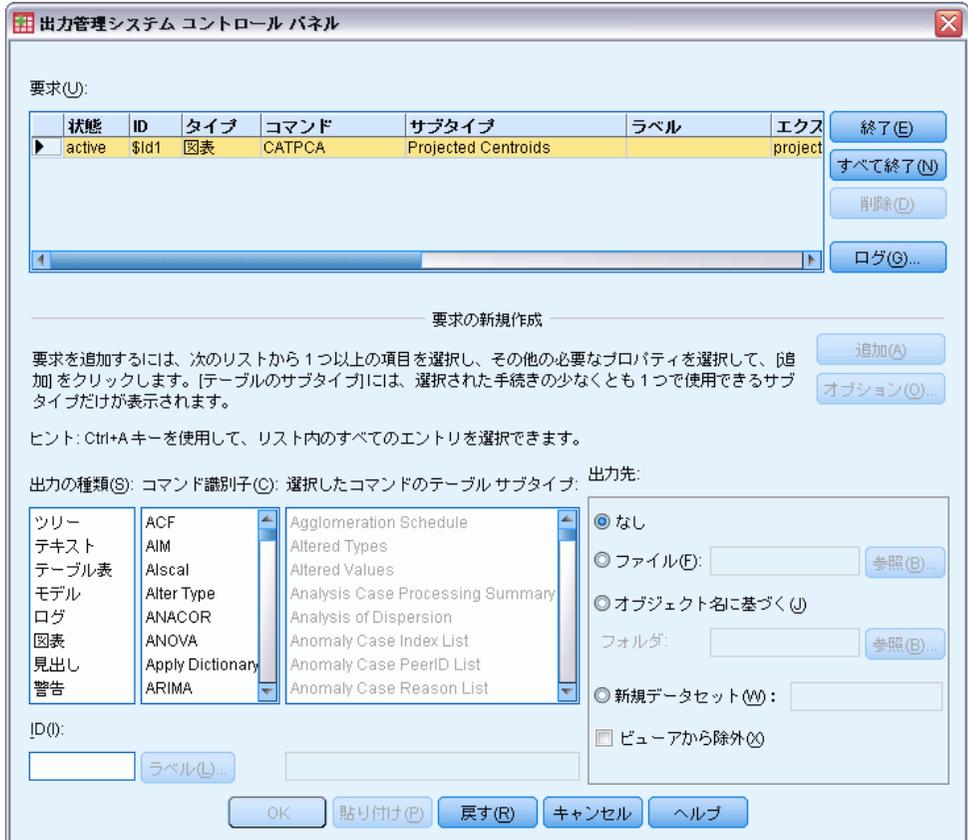


- ▶ 「相互関係」のカテゴリプロットを要求します。
- ▶ 「体重」から「体」までの変換プロットを要求します。
- ▶ 「相互関係」を射影の重心として選択し、「過食」、「性的傾向」、および「関心」を射影先として選択します。
- ▶ [続行] をクリックします。
- ▶ [カテゴリ主成分分析] ダイアログ ボックスで、[OK] をクリックします。

この手続きでは、結果として、被験者得点と変換された変数の平均平方の相関を最大化するカテゴリの（平均 0 で単位分散を持つ）被験者および数量化の得点を示します。現在の分析では、カテゴリ数量化は、順序の情報を反映するように制約されています。

最終的に、射影された重心のテーブル情報を projected_centroids.sav に書き込むには、OMS 要求を終了する必要があります。OMS コントロールパネルをもう一度表示します。

図 10-31
出力管理システム コントロール パネル

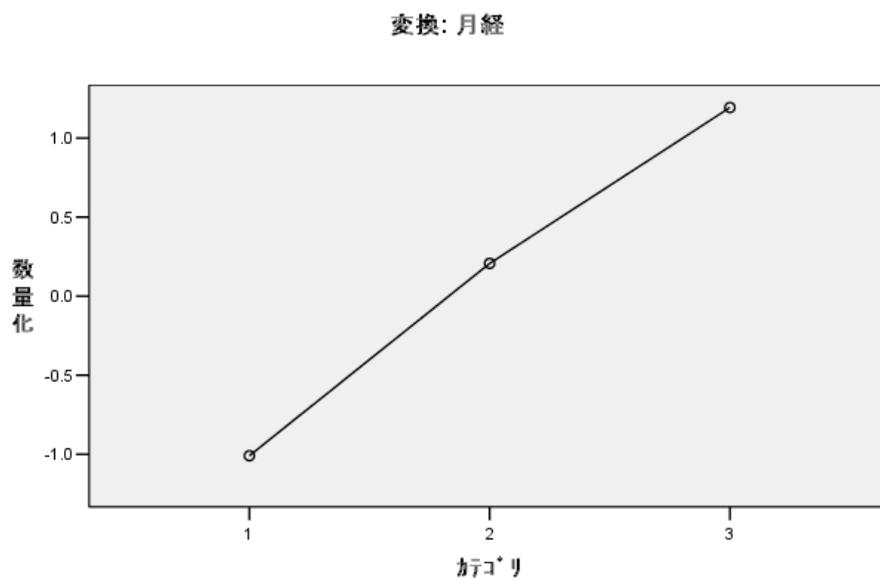


- ▶ [終了] をクリックします。
- ▶ [OK] をクリックし、さらに確認の [OK] をクリックします。

変換プロット

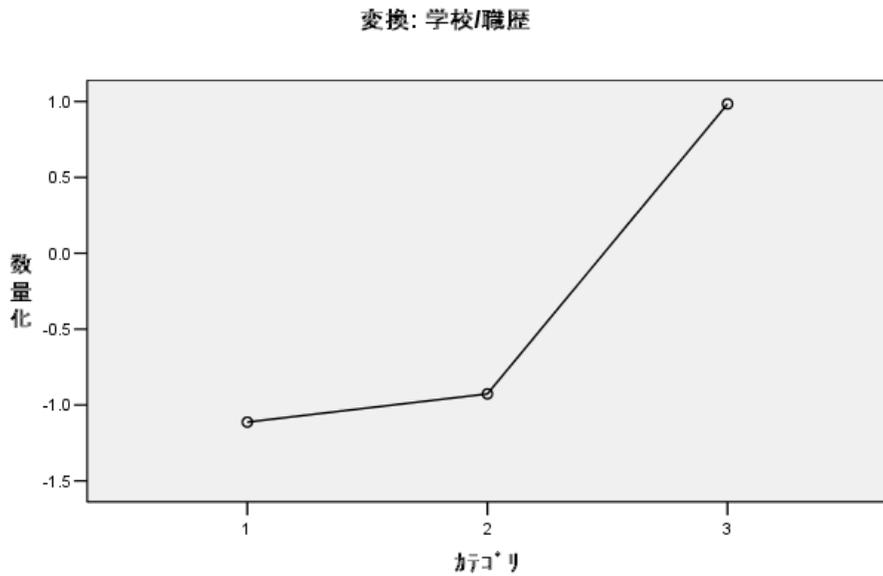
変換プロットでは、横軸に元のカテゴリ番号が表示され、縦軸に最適な数量化が示されます。

図 10-32
月経の変換プロット



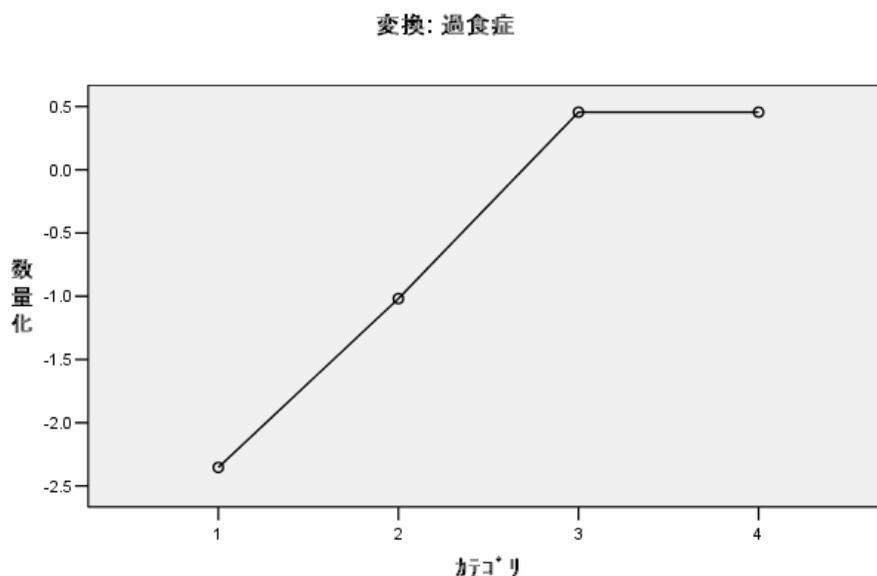
「月経」などの変数の一部は、線型変換をほぼ実行しました。そのため、この分析では、これらの変数を数値として解釈することがあります。

図 10-33
学校/職歴の変換プロット



「学校/職歴」などの他の変数の数値化は、線型変換を実行しませんでした。この数値化は、順序尺度水準で解釈する必要があります。2 番目のカテゴリと 3 番目のカテゴリの差は、最初のカテゴリと 2 番目のカテゴリの差よりも重要です。

図 10-34
過食症の変換プロット



「過食症」の数量化では、興味深い状況が発生します。実行した変換は、カテゴリ 1 からカテゴリ 3 までは線型ですが、カテゴリ 3 とカテゴリ 4 の数量化された値は等しくなります。この結果は、3 と 4 の得点では、患者が区別されないことを示し、4 の得点を 3 の得点として記録することにより、2 成分の解で数値尺度水準を使用できることを示しています。

モデルの要約

図 10-35
モデルの要約(M)

次元	Cronbach のアルファ	説明された分散	
		合計 (固有値)	分散の %
1	.874	5.550	34.690
2	.522	1.957	12.234
合計	.925 ^a	7.508	46.924

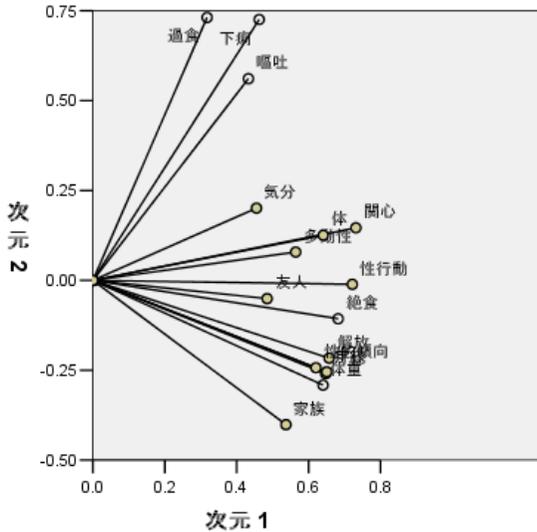
a. Cronbach のアルファ合計は、固有値合計に基づいています。

モデルがデータにどの程度適合しているかを確認するには、モデルの要約を参照してください。全分散の約 47% は 2 成分のモデルで、35% は最初の次元で、12% は 2 番目の次元で説明されています。そのため、個々のオブジェクトレベルでの変動のほぼ半数は、2 成分のモデルで説明されています。

成分負荷

解の2つの次元の解釈を開始するには、成分負荷を参照します。すべての変数は、最初の次元で正の成分負荷を持っています。つまり、すべての変数と正の相関関係を持つ共通因子が存在します。

図 10-36
成分負荷プロット



2番目の次元は、変数を区別します。変数「過食症」、「嘔吐」、および「下痢」は、2番目の次元で大きい正の負荷を持つバンドルを構成します。一般的に、これらの症状は、典型的な過食行動と見なされます。

変数「家族の束縛からの解放」、「学校/職歴」、「性的傾向態度」、「体重」、「月経」は、別のバンドルを構成します。さらに、「食事量の制限」および「家族関係」をこのバンドルに含めることができます。これは、これらの変数のベクトルが主要なクラスターの近くに存在し、変数が拒食症の症状（絶食、体重、月経）と見なされたり、心理社会的な性質（解放、学歴/経歴、性的態度、家族関係）を示したりするからです。このバンドルのベクトルは、過食、嘔吐、下痢のベクトルに対して直交（垂直に交差）しています。つまり、この変数のグループは、過食症の変数のグループとは相関関係がありません。

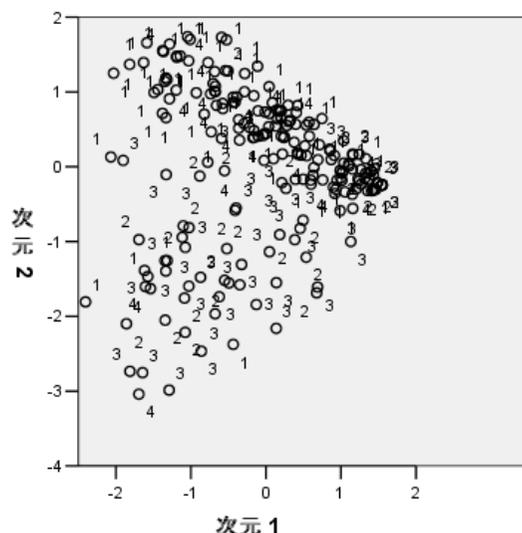
変数「友人」、「精神状態（気分）」、および「多動性、衝動性」は、解にあまり適合しているように見えません。これは、各ベクトルの長さを観測することにより、プロット内で確認できます。与えられた変数のベクトルの長さは、適切に対応しています。これらの変数は、最も短いベクトルを持っています。2成分の解に基づいて、摂食障害に提起された症状からこれらの変数をドロップすることもできます。ただし、この変数は、より高い次元の解にうまく適合する場合があります。

変数「性的行動」、「食べ物と体重に対する関心」、および「体知覚」は、患者の身にどのような症状が起きるかに関して、もう1つ別の症状の理論グループを構成します。これらの変数は、直交する2個の変数のバンドルと相関関係にあると同時に、非常に長いベクトルを持ち、最初の次元と強い相関関係にあります。そのため、「共通」因子に関する有用な情報を提供する場合があります。

オブジェクト スコア

次の図では、オブジェクトスコアのプロットを示しています。このプロットでは、被験者が診断カテゴリでラベル付けされます。

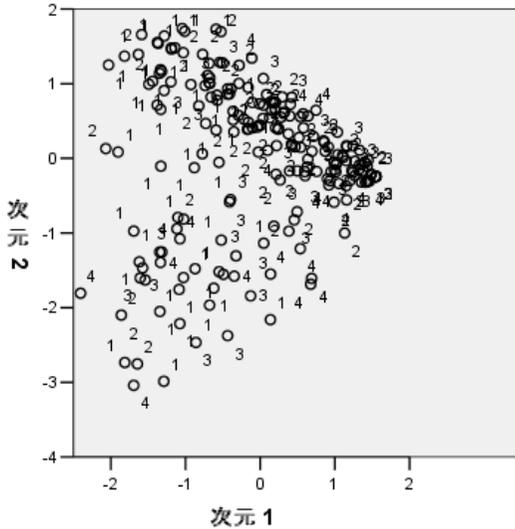
図 10-37
診断でラベル付けされたオブジェクトスコア プロット



最初の次元上では、患者が診断で区別されていないので、このプロットは最初の次元の解釈に役立ちません。ただし、2番目の次元に関する情報が含まれている場合があります。拒食症の被験者(1)および特殊な摂食障害を持つ患者(4)は、なんらかの形で過食症を伴う被験者(2および3)の上にグループを構成します。したがって、前の項での説明(過食症のバンドル内の変数が、大きな正の成分負荷を2番目の次元に持っていること)にあるように、2番目の次元では、過食症患者と他の患者が区別されます。従来、過食症に関連付けられていた症状の成分負荷が2番目の次元で大きな値を持っていることを考えると、これは当然のことです。

次の図では、オブジェクトスコアのプロットを示しています。このプロットでは、被験者が診断回数でラベル付けされます。

図 10-38
診断回数でラベル付けされたオブジェクトスコア



回数でオブジェクトスコアにラベルを付けると、最初の次元が回数に関係していることがわかります。これは、診断回数が 1 の場合は主に左に、それ以外の場合は右に進行するように見えるからです。オブジェクトスコアを保存し、x 軸上の次元 1 の得点、y 軸上の次元 2 の得点、および患者数によるマーカーの設定を使用して散布図を作成することにより、このプロット内の時点に接続できるように注意してください。

回数でラベル付けしたオブジェクトスコアプロットと診断でラベル付けしたオブジェクトスコアプロットを比較すると、異常なオブジェクトを明らかにできます。たとえば、回数でラベル付けしたプロットでは、4 回目の診断結果を持つ患者が、プロット内の他のすべての点より左に位置します。一般的な点の傾向では、最後の回数はさらに右に位置することになるので、これは例外的です。興味深いことに、患者を過食症のクラスターに配置する得点を持つ患者が拒食症である場合には、回数に適していないように見えるこの点でも、例外的な診断が行われています。オブジェクトスコアの表を参照すると、これは、拒食症と診断された患者 43 であることがわかります。この患者のオブジェクトスコアを次の表に示します。

テーブル 10-4
患者 43 のオブジェクトスコア

時刻	次元 1	次元 2
1	-2.031	1.250
2	-2.067	0.131

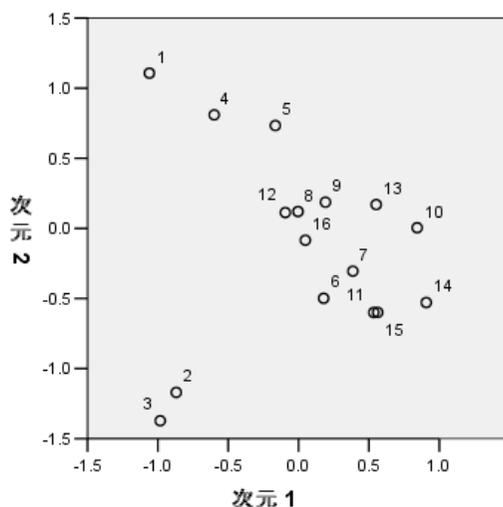
時刻	次元 1	次元 2
3	-1.575	-1.467
4	-2.405	-1.807

1 回目の患者の得点は拒食症のプロトタイプであり、次元 1 の大きな負の得点はボディ イメージが悪いことを示し、次元 2 の正の得点は拒食症の症状があったり、心理社会的な行動が良くないことを示します。ただし、大多数の患者とは異なり、次元 1 では進行がほとんどないか、まったくありません。次元 2 では、「正常」(0 付近、つまり、拒食行動と過食行動の中間)に向かって進んでいるように見えますが、その後、患者は、過食症の症状を示すようになります。

病気の経過の構成の調査

2 つの次元を 4 個の診断カテゴリと 4 つの時点に関連付けた方法についてより詳しく調べるために、「患者の診断」の 4 個のカテゴリと「インタビューの時間」の 4 個のカテゴリのクロス分類で、補助変数「時間/診断の相互関係」を作成しました。したがって、「時間/診断の相互関係」には 16 個のカテゴリがあり、この場合、最初のカテゴリは、初診時の拒食症患者を示します。5 番目のカテゴリは 2 回目での拒食症患者などを示し、16 番目のカテゴリは 4 回目での特殊な摂食障害の患者を示します。補助変数 [時間/診断の相互関係] を使用すると、さまざまなグループの病気の経過を時間とともに調査できます。変数に多重名義尺度水準が与えられたので、カテゴリ ポイントは次の図に表示されます。

図 10-39
時間/診断の相互関係のカテゴリ ポイント



構成の一部はこのプロットから明確になります。1 回目の診断カテゴリは、2 番目の次元で、拒食症と特殊な摂食障害を、過食症を伴う拒食症および拒食症後の過食症から明確に区別します。その後、パターンは少し見にくくなります。

ただし、数量化に基づいた散布図を作成することにより、より見やすいパターンを作成できます。これを行うためには、メニューから次の項目を選択します。

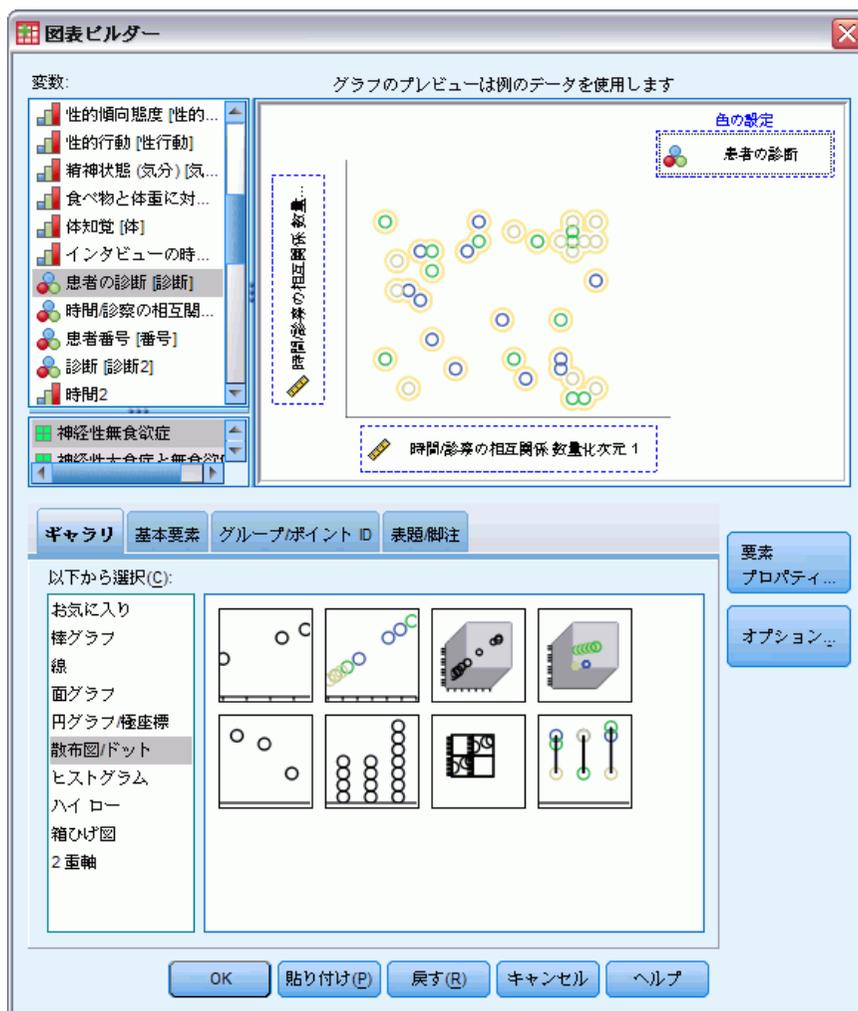
グラフ(G) > 図表ビルダー(C)...

図 10-40
[散布図/ドット] ギャラリー



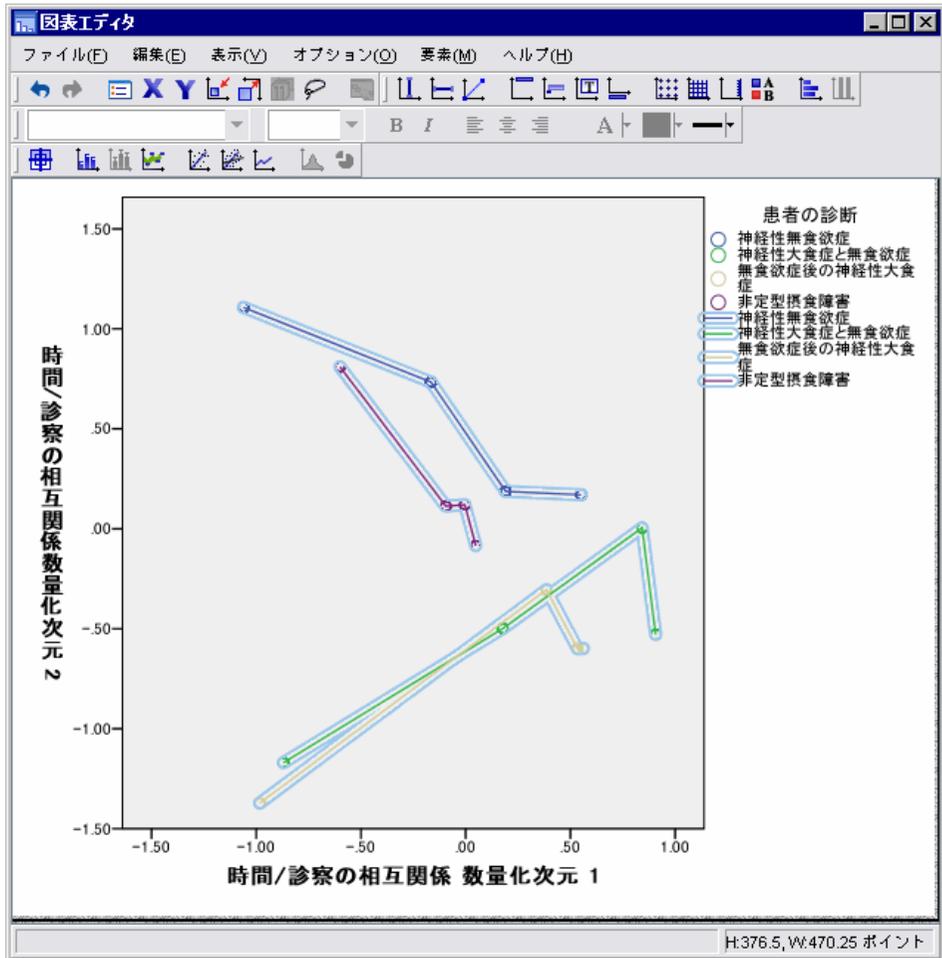
- ▶ [散布図/ドット] ギャラリーを選択し、[グループ化散布図] を選択します。

図 10-41
図表ビルダー



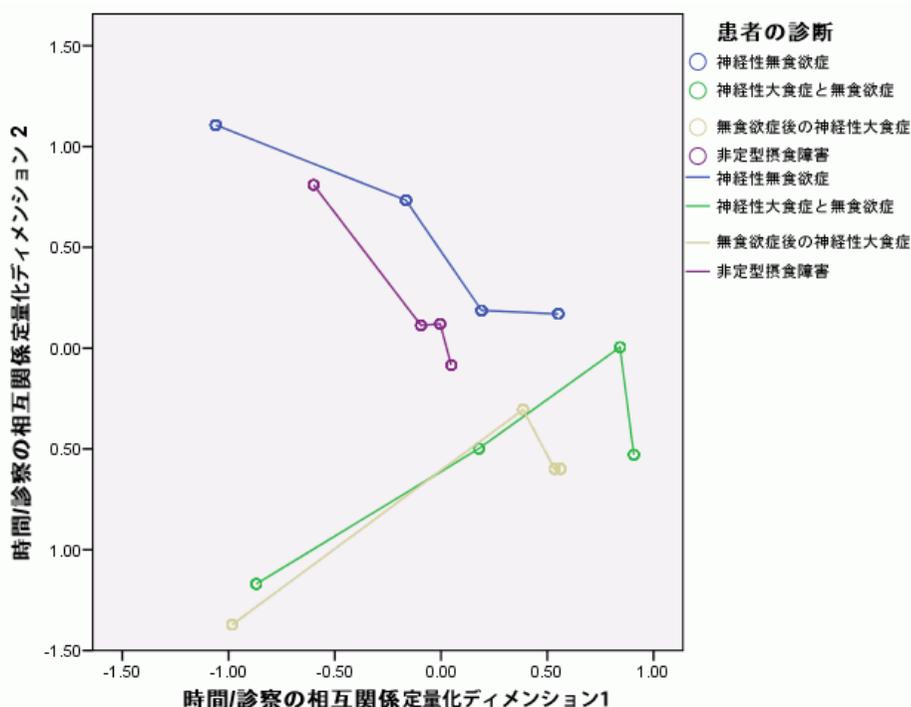
- ▶ y 軸変数として「時間/診察の相互関係 数量化次元 2」を選択し、x 軸変数として「時間/診察の相互関係 数量化次元 1」を選択します。
- ▶ 「患者の診断」で色を設定するように選択します。
- ▶ [OK] をクリックします。

図 10-42
病気の経過の構成



- ▶ その後、その点に接続するには、グラフ上をダブルクリックし、図表エディタで [参照線を追加] ツールをクリックします。
- ▶ 図表エディタを閉じます。

図 10-43
病気の経過の構成



複数の回数にわたって各診断カテゴリのカテゴリ ポイントを接続すると、すでにオブジェクト スコア プロットから決定したとおり、パターンでは、最初の次元が回数に関連し、2 番目の次元が診断に関連していることがすぐに示されます。

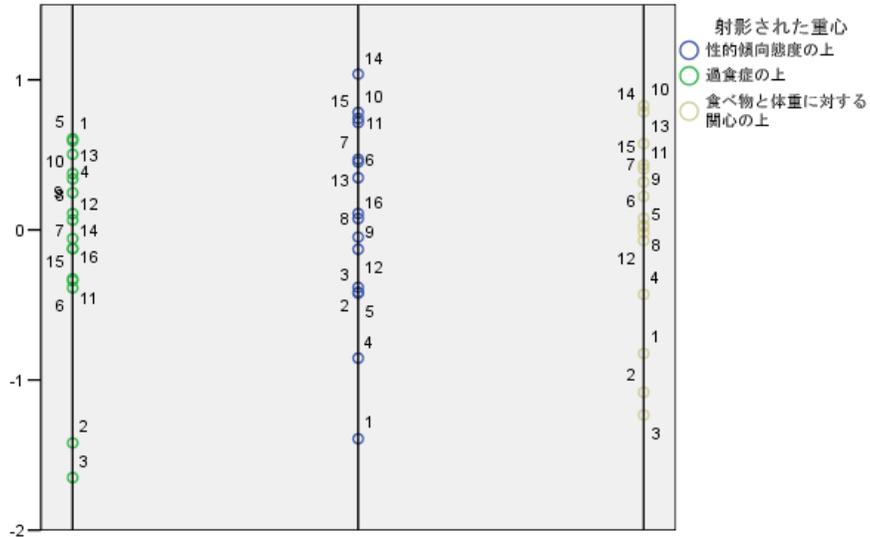
ただし、このプロットでは、時間の経過とともに、病気が似てくる傾向があることも示されます。さらに言えば、すべてのグループでは、1 回目と 2 回目の間で最も大きく進行します。拒食症患者は、2 回目から 3 回目にかけてさらに進行することを示しますが、他のグループは、あまり進行しないことを示します。

選択された変数の差分作成

成分負荷で識別された症状の各バンドルから、1 つの変数をそのバンドルの「代表」として選択します。過食症のバンドルから摂食障害、拒食症/心理社会的行動のバンドルから性的態度、3 番目のバンドルから身体への偏見が選択されました。

考えられる病気の経過の差を調査するために、「時間/診断の相互関係」の「過食症」、「性的傾向態度」、および「食べ物と体重に対する関心」への投影は、計算され、次の図にプロットされました。

図 10-44
[過食症]、[性的傾向態度]、および [食べ物と体重に対する関心] に射影された [時間/診断の相互関係] の重心



このプロットでは、最初の時点で、摂食障害が過食症患者 (2 と 3) を他の患者 (1 と 4) と区別し、性的態度が拒食症患者と特殊な患者 (1 と 4) を他の患者 (2 と 3) と区別し、身体への偏見では実際には患者を区別しないことが示されます。多くのアプリケーションでは、このプロットだけで症状と診断間の関係を説明できます。しかし、複数の時点により複雑になると、画像は混乱した状態になります。

時間の経過とともに投影を表示するには、射影された重心のテーブルの内容をプロットできるようにする必要があります。これは、この情報を `projected_centroids.sav` に保存した OMS 要求によって可能になります。

図 10-45

Projected_centroids.sav

	Label_	Var1	過食症	性的傾向態度	食べ物と体重に対する関心
1	射影された重心	1	0.593	-1.391	-0.823
2	射影された重心	2	-1.419	-0.383	-1.082
3	射影された重心	3	-1.650	-0.415	-1.233
4	射影された重心	4	0.504	-0.854	-0.430
5	射影された重心	5	0.607	-0.421	-0.018
6	射影された重心	6	-0.386	0.347	0.077
7	射影された重心	7	-0.126	0.471	0.319
8	射影された重心	8	0.109	-0.048	0.019
9	射影された重心	9	0.247	0.109	0.224
10	射影された重心	10	0.340	0.783	0.827
11	射影された重心	11	-0.337	0.716	0.406
12	射影された重心	12	0.066	-0.129	-0.071

変数「過食症」、「性的傾向態度」、および「食べ物と体重に対する関心」には、該当する各症状に射影された重心の値が含まれています。ケース番号 (1 ~ 16) は、Time/diagnosis interaction に対応しています。時間の値と診断の値を分離する新しい変数を計算する必要があります。

- ▶ メニューから次の項目を選択します。
変換(T) > 変数の計算(C)...

図 10-46
[変数の計算] ダイアログ ボックス



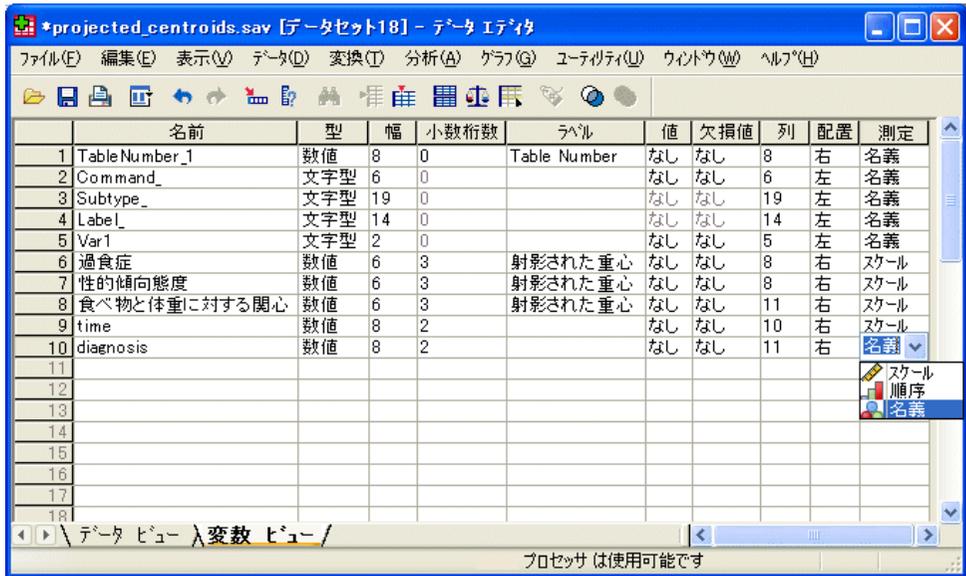
- ▶ 目標変数として「time」と入力します。
- ▶ 数式として「 $\text{trunc}((\$casenum-1) / 4) + 1$ 」と入力します。
- ▶ [OK] をクリックします。

図 10-47
[変数の計算] ダイアログ ボックス



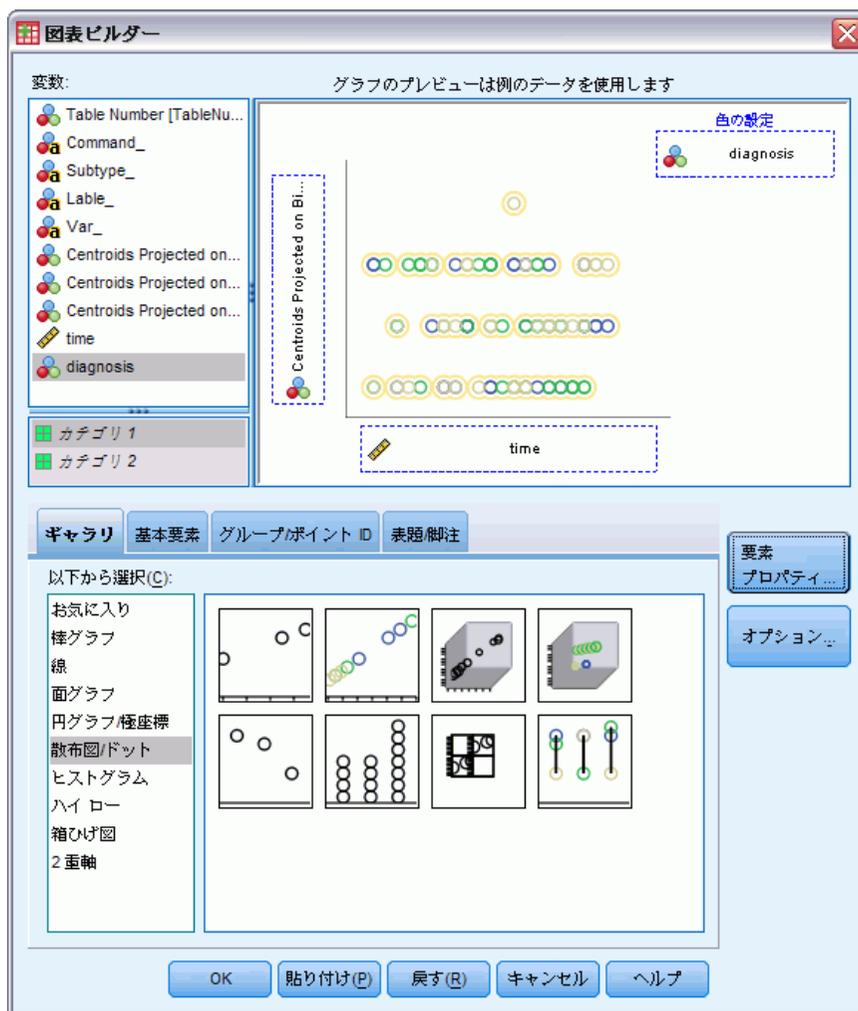
- ▶ もう一度、[変数の計算] ダイアログ ボックスを表示します。
- ▶ 目標変数として「diagnosis」と入力します。
- ▶ 数式として「mod(\$casenum-1, 4) + 1」と入力します。
- ▶ [OK] をクリックします。

図 10-48
Projected_centroids.sav



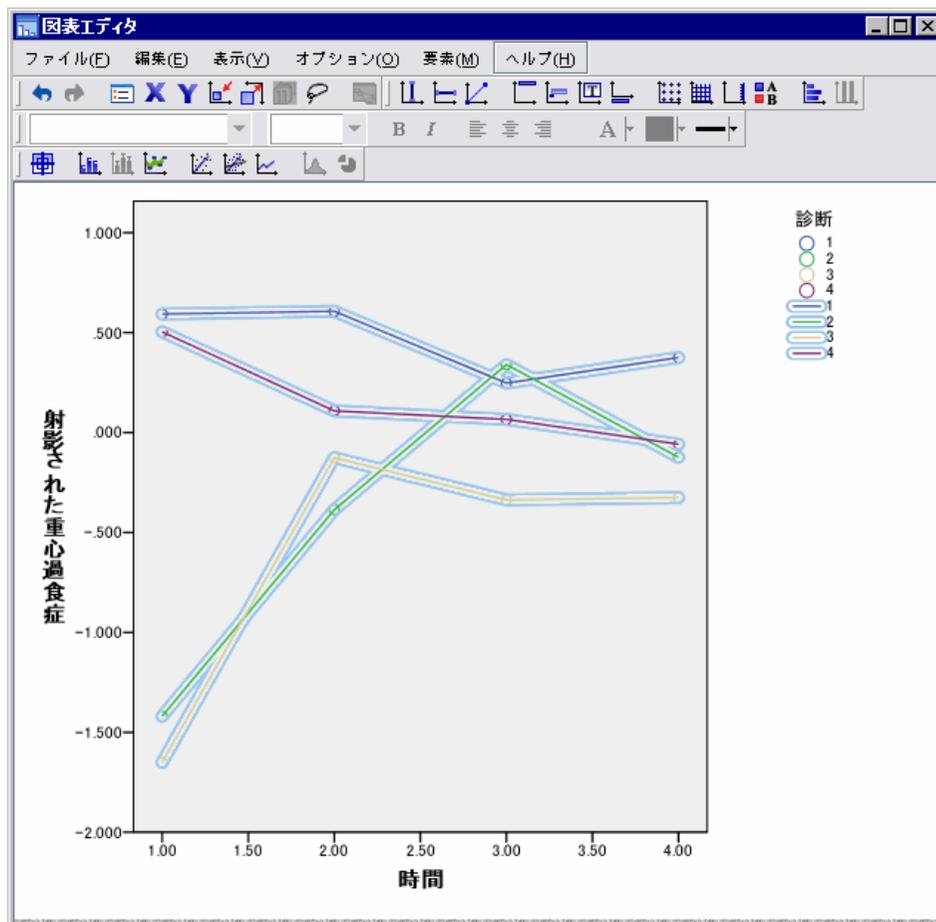
[変数ビュー] で、「diagnosis」の測定方法を [尺度] から [名義] に変更します。

図 10-49
図表ビルダー



- ▶ 最終的に、時間の経過とともに過食に射影された診断時の重心を表示するには、図表ビルダーをもう一度表示し、[戻す] をクリックして前回の選択内容をクリアします。
- ▶ [散布図/ドット] ギャラリーを選択し、[グループ化散布図] を選択します。
- ▶ y 軸変数として「過食症」を選択し、x 軸変数として「時間」を選択します。
- ▶ 「診断」で色を設定することを選択します。
- ▶ [OK] をクリックします。

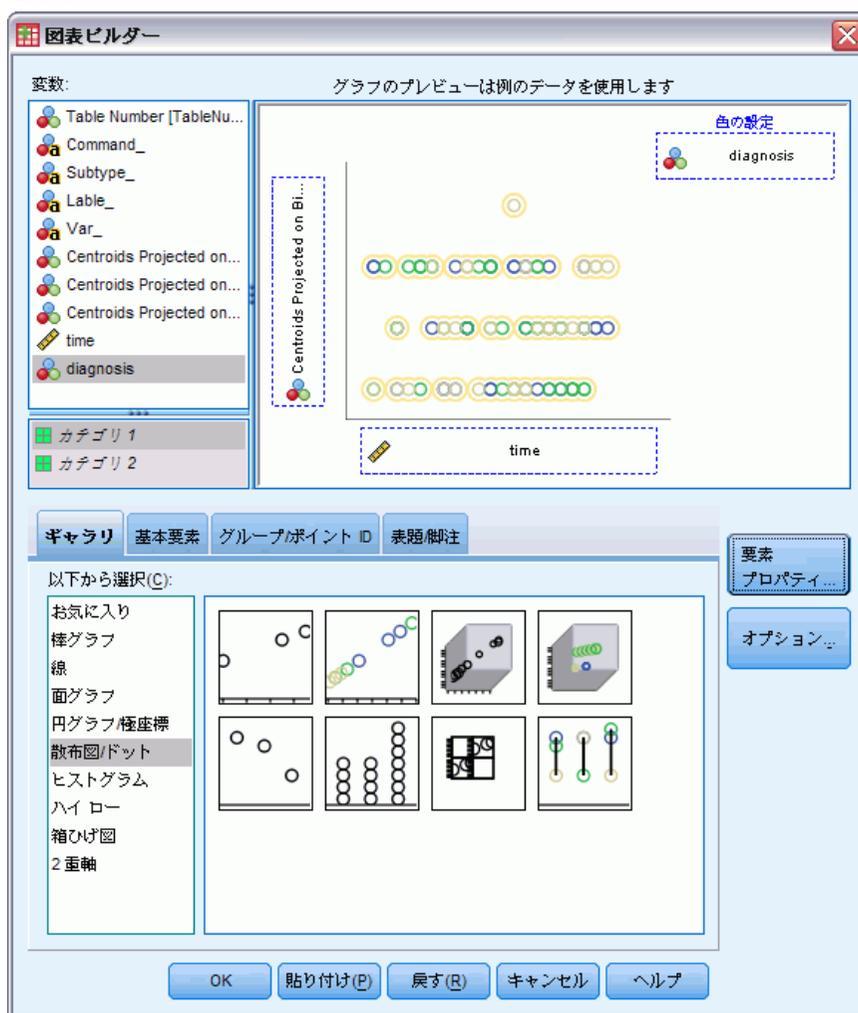
図 10-50
時間の経過とともに摂食障害に射影された診断時の重心



- ▶ その後、その点に接続するには、グラフ上をダブルクリックし、図表エディタで [参照線を追加] ツールをクリックします。
- ▶ 図表エディタを閉じます。

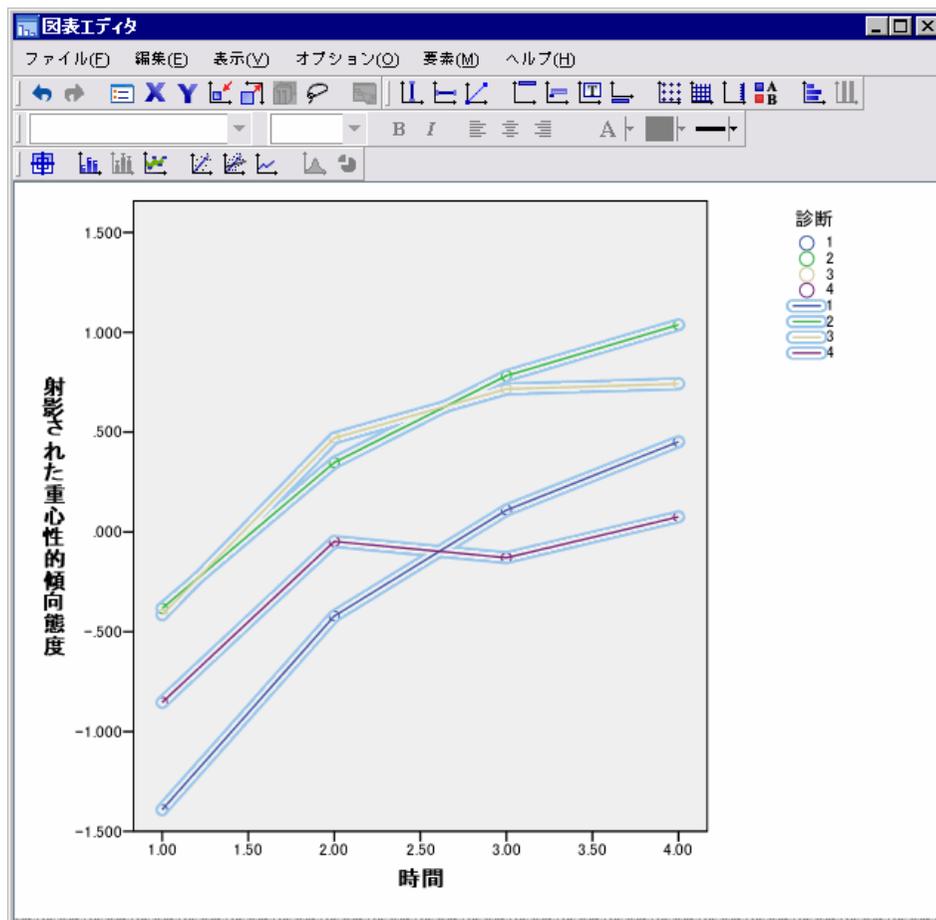
摂食障害に関して、拒食症グループの初期値が過食症グループの初期値とは異なることがわかります。拒食症グループがほとんど変化せず、過食症グループが増加を示す場合、この差は時間の経過とともに小さくなります。

図 10-51
図表ビルダー



- ▶ 図表ビルダーをもう一度表示します。
- ▶ y 軸変数として選択した「過食症」の選択を解除し、y 軸変数として「性的傾向態度」を選択します。
- ▶ [OK] をクリックします。

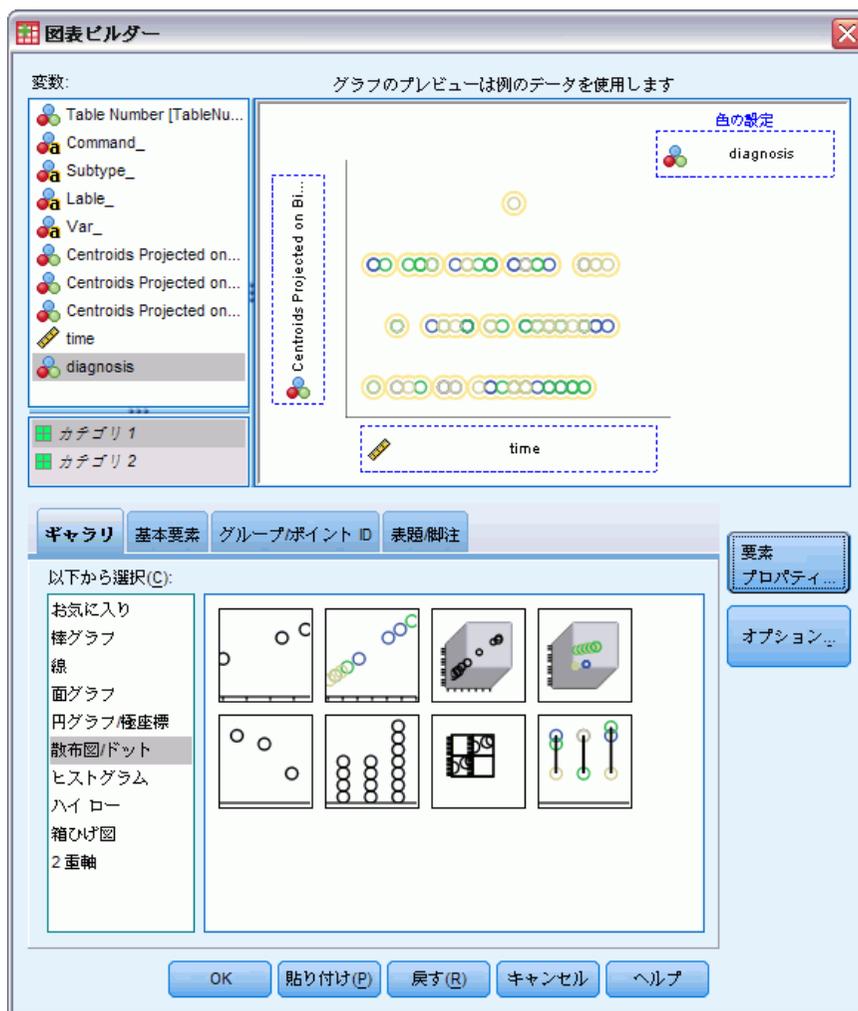
図 10-52
時間の経過とともに性的態度に射影された診断時の重心



- ▶ その後、その点に接続するには、グラフ上をダブルクリックし、図表エディタで [参照線を追加] ツールをクリックします。
- ▶ 図表エディタを閉じます。

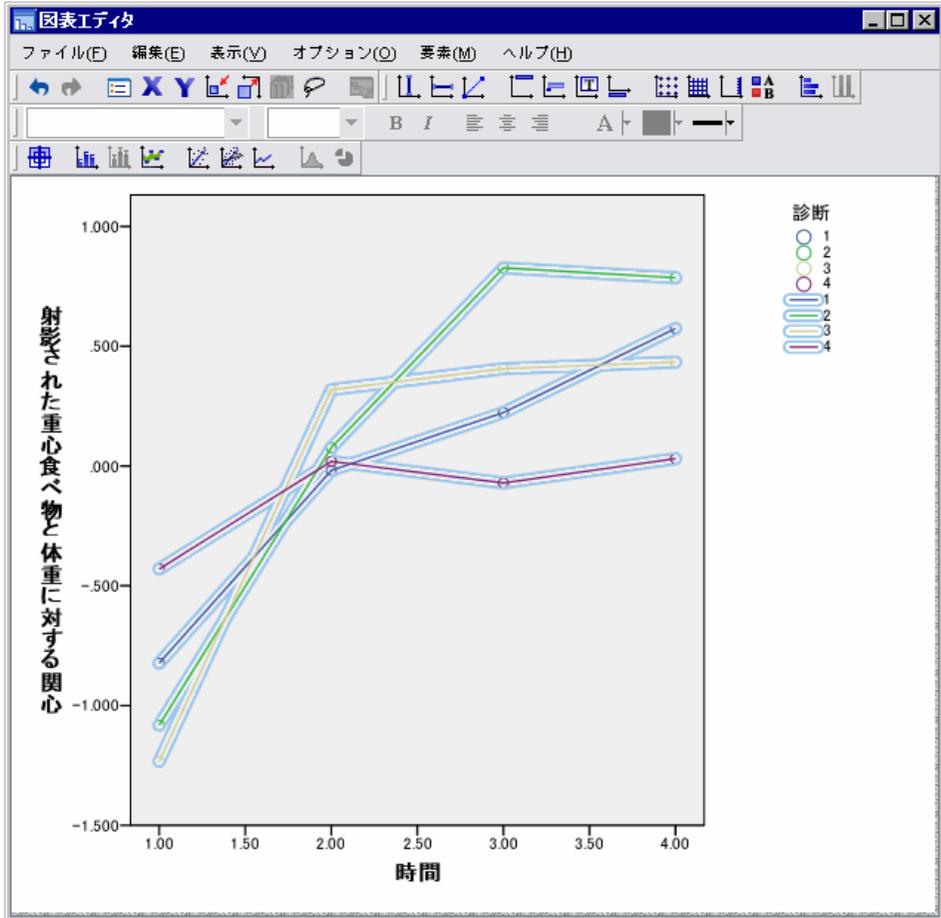
性的態度に関して、4つの曲線が時間の経過とともにほぼ並行になり、すべてのグループが進行を示します。ただし、過食症グループの得点は、拒食症グループの得点よりも高く（良く）なっています。

図 10-53
図表ビルダー



- ▶ 図表ビルダーをもう一度表示します。
- ▶ y 軸変数として選択した「性的傾向態度」の選択を解除し、y 軸変数として「食べ物と体重に対する関心」を選択します。
- ▶ [OK] をクリックします。

図 10-54
時間の経過とともに身体への偏見に射影された診断時の重心



- ▶ その後、その点に接続するには、グラフ上をダブルクリックし、図表エディタで [参照線を追加] ツールをクリックします。
- ▶ 図表エディタを閉じます。

Body preoccupation は、中心的な症状を表す変数で、4 つの異なるグループによって共有されています。特殊な摂食障害の患者は別として、拒食症グループおよび 2 つの過食症グループの水準は、最初と最後の両方で非常に似ています。

推奨参考文献

カテゴリ主成分分析の詳細は、次のテキストを参照してください。

- De Haas, M., J. A. Algera, H. F. J. M. Van Tuijl, および J. J. Meulman. 2000. Macro and micro goal setting: In search of coherence. *Applied Psychology*, 49, .
- De Leeuw, J. 1982. Nonlinear principal components analysis. In: *COMPSTAT Proceedings in Computational Statistics*, Vienna: Physica Verlag.
- Eckart, C., および G. Young. 1936. The approximation of one matrix by another one of lower rank. *Psychometrika*, 1, .
- Gabriel (G), K. R. 1971. The biplot graphic display of matrices with application to principal components analysis. *Biometrika*, 58, .
- Gifi, A. 1985. PRINCALS. Research Report UG-85-02. Leiden: Department of Data Theory, University of Leiden.
- Gower, J. C., および J. J. Meulman. 1993. The treatment of categorical information in physical anthropology. *International Journal of Anthropology*, 8, .
- Heiser, W. J., および J. J. Meulman. 1994. Homogeneity analysis: Exploring the distribution of variables and their nonlinear relationships. In: *Correspondence Analysis in the Social Sciences: Recent Developments and Applications*, M. Greenacre, および J. Blasius, 編集者. New York: Academic Press.
- Kruskal, J. B. 1978. Factor analysis and principal components analysis: Bilinear methods. In: *International Encyclopedia of Statistics*, W. H. Kruskal, および J. M. Tanur, 編集者. New York: The Free Press.
- Kruskal, J. B., および R. N. Shepard. 1974. A nonmetric variety of linear factor analysis. *Psychometrika*, 39, .
- Meulman, J. J. 1993. Principal coordinates analysis with optimal transformations of the variables: Minimizing the sum of squares of the smallest eigenvalues. *British Journal of Mathematical and Statistical Psychology*, 46, .
- Meulman, J. J., および P. Verboon. 1993. Points of view analysis revisited: Fitting multidimensional structures to optimal distance components with cluster restrictions on the variables. *Psychometrika*, 58, .
- Meulman, J. J., A. J. Van der Kooij, および A. Babinec. 2000. New features of categorical principal components analysis for complicated data sets, including data mining. In: *Classification, Automation and New Media*, W. Gaul, および G. Ritter, 編集者. Berlin: Springer-Verlag.

Meulman, J. J., A. J. Van der Kooij, および W. J. Heiser. 2004. Principal components analysis with nonlinear optimal scaling transformations for ordinal and nominal data. In: Handbook of Quantitative Methodology for the Social Sciences, D. Kaplan, ed. Thousand Oaks, Calif.: Sage Publications, Inc..

Theunissen, N. C. M., J. J. Meulman, A. L. Den Ouden, H. M. Koopman, G. H. Verrips, S. P. Verloove-Vanhorick, および J. M. Wit. 2003. Changes can be studied when the measurement instrument is different at different time points. Health Services and Outcomes Research Methodology, 4, .

Tucker, L. R. 1960. Intra-individual and inter-individual multidimensionality. In: Psychological Scaling: Theory & Applications, H. Gulliksen, および S. Messick、編集者. New York: John Wiley and Sons.

Vlek, C., および P. J. Stallen. 1981. Judging risks and benefits in the small and in the large. Organizational Behavior and Human Performance, 28, .

Wagenaar, W. A. 1988. Paradoxes of gambling behaviour. London: Lawrence Erlbaum Associates, Inc.

Young, F. W., Y. Takane, および J. De Leeuw. 1978. The principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling features. Psychometrika, 43, .

Zeijl, E., Y. te Poel, M. du Bois-Reymond, J. Ravestloot, および J. J. Meulman. 2000. The role of parents and peers in the leisure activities of young adolescents. Journal of Leisure Research, 32, .

非線型正準相関分析

非線型正準相関分析は、2 つ以上の変数グループが互いにどの程度類似しているかを判断することを目的としています。線型正準相関分析と同様に、低次元空間でのグループ間の関係における分散をできる限り多く説明することが目的です。ただし、線型正準相関分析と異なり、非線型正準相関分析では、間隔尺度の測定または関係が線型であることが想定されていません。もう 1 つ重要な違いとして、非線型正準相関分析では、各グループに含まれている変数の線型結合を不明なグループ (オブジェクト スコア) と同時に比較することで、グループ間の類似度が確立されます。

例: 調査結果の分析

この章で説明する例は、調査 (Verdegaal, 1985) からの引用です。8 つの変数に対する 15 人の被験者の回答を記録しました。次の表は、データセットに含まれている変数、変数ラベル、および値ラベル (カテゴリ) を示しています。

テーブル 11-1
調査データ

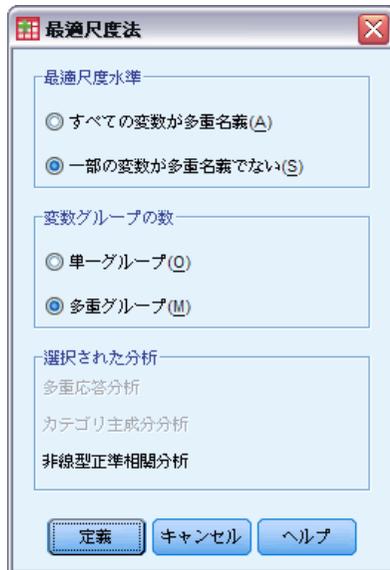
変数名	変数ラベル	値ラベル
年齢	年齢	20-25, 26-30, 31-35, 36-40, 41-45, 46-50, 51-55, 56-60, 61-65, 66-70
婚姻	婚姻状況	独身、既婚、その他
ペット	ペットの有無	なし、猫、犬、犬猫以外、家畜類
新聞	最も頻繁に読む新聞	なし、Telegraaf、Volkskrant、NRC、その他
音楽	好きな音楽	クラシック、ニュー ウェーブ、ポピュラー、その他、なし
居住地域	居住したい地域	都市、町、村
数学	数学試験の成績	0-5, 6-10, 11-15
言語	語学試験の成績	0-5, 6-10, 11-15, 16-20

このデータセットは verd1985.sav にあります。詳細は、A 付録 p.321 サンプル ファイル を参照してください。対象となる変数は最初の 6 つの変数で、それらは 3 つのグループに分類されます。グループ 1 には「年齢」と「婚姻」、グループ 2 には「ペット」と「新聞」、グループ 3 には「音楽」と「居住地域」がそれぞれ含まれます。「ペット」は多重名義として尺度化され、「年齢」は順序として尺度化されます。また、その他のすべての変数は単一名義として尺度化されます。この分析では、任意の初期配置が必要になります。デフォルトでは、初期配置は数値です。ただし、一部の変数が順序付けの可能性がない単一名義として扱われる場合は、任意の初期配置を選択することが最善です。このことは、この調査の大部分の変数に当てはまります。

データの調査

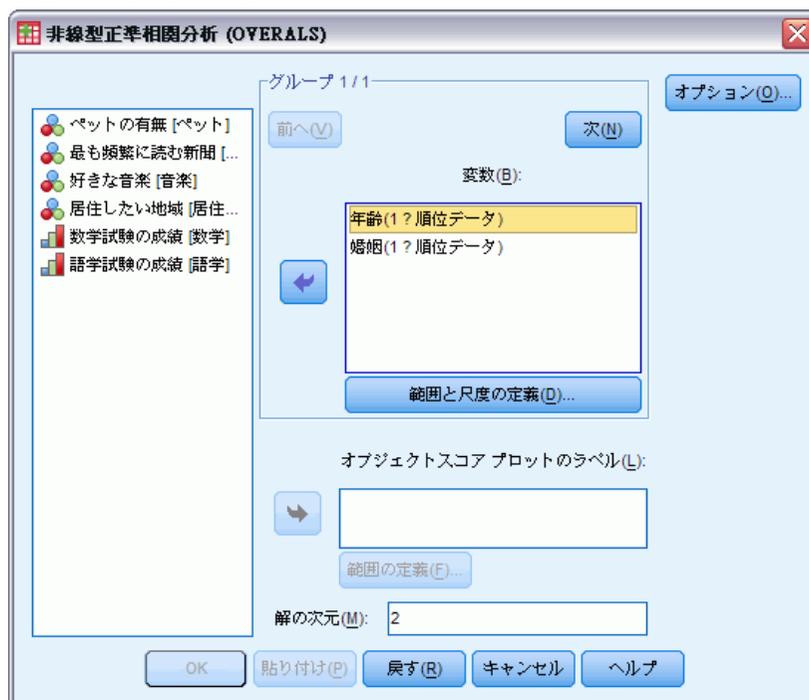
- ▶ このデータセットの非線型正準相関分析を実行するには、メニューから次の項目を選択します。
[分析] > [次元分解] > [最適尺度法]

図 11-1
[最適尺度法] ダイアログ ボックス



- ▶ [最適尺度水準] グループから [一部の变数が多重名義でない] を選択します。
- ▶ [変数グループの数] グループから [多重グループ] を選択します。
- ▶ [定義] をクリックします。

図 11-2
[非線型正準相関分析] ダイアログ ボックス



- ▶ 第 1 グループの変数として、「年齢」と「婚姻状況」を選択します。
- ▶ 「年齢」を選択し、[範囲と尺度の定義] をクリックします。

図 11-3
[範囲と尺度の定義] ダイアログ ボックス



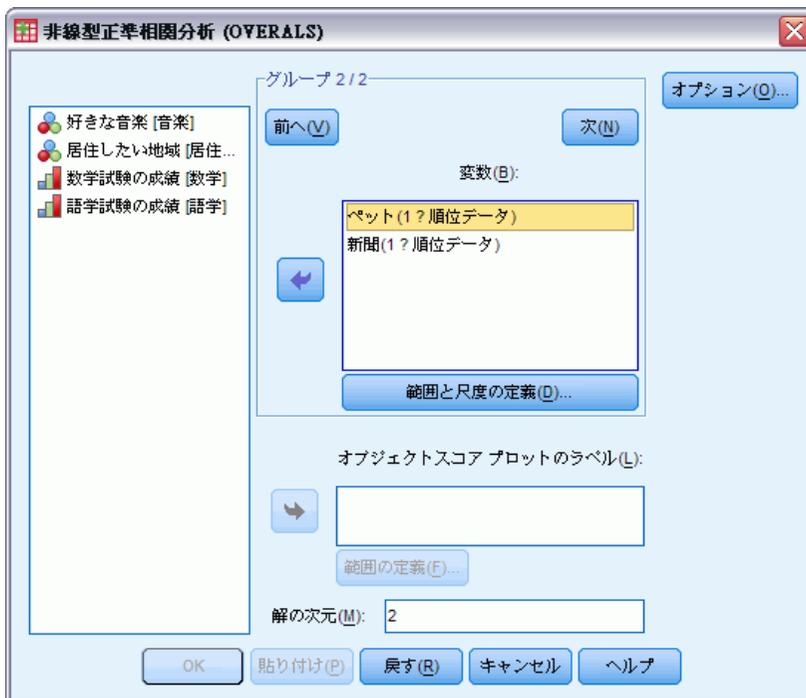
- ▶ この変数の最大値として「10」と入力します。
- ▶ [続行] をクリックします。
- ▶ 「婚姻」を選択し、[非線型正準相関分析] ダイアログ ボックスの [範囲と尺度の定義] をクリックします。

図 11-4
[範囲と尺度の定義] ダイアログ ボックス



- ▶ この変数の最大値として「3」と入力します。
- ▶ 測定尺度として [単一名義] を選択します。
- ▶ [続行] をクリックします。
- ▶ [非線型正準相関分析] ダイアログ ボックスで [次] をクリックし、次の変数グループを定義します。

図 11-5
[非線型正準相関分析] ダイアログ ボックス



- ▶ 第 2 グループの変数として、「ペットの有無」と「最も頻繁に読む新聞」を選択します。

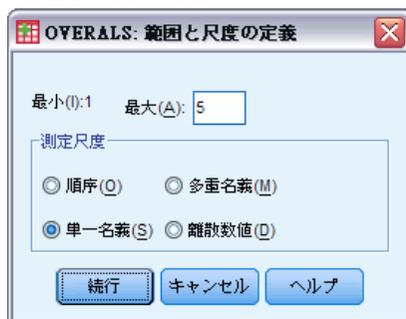
- ▶ 「ペット」を選択し、[範囲と尺度の定義]をクリックします。

図 11-6
[範囲と尺度の定義] ダイアログ ボックス



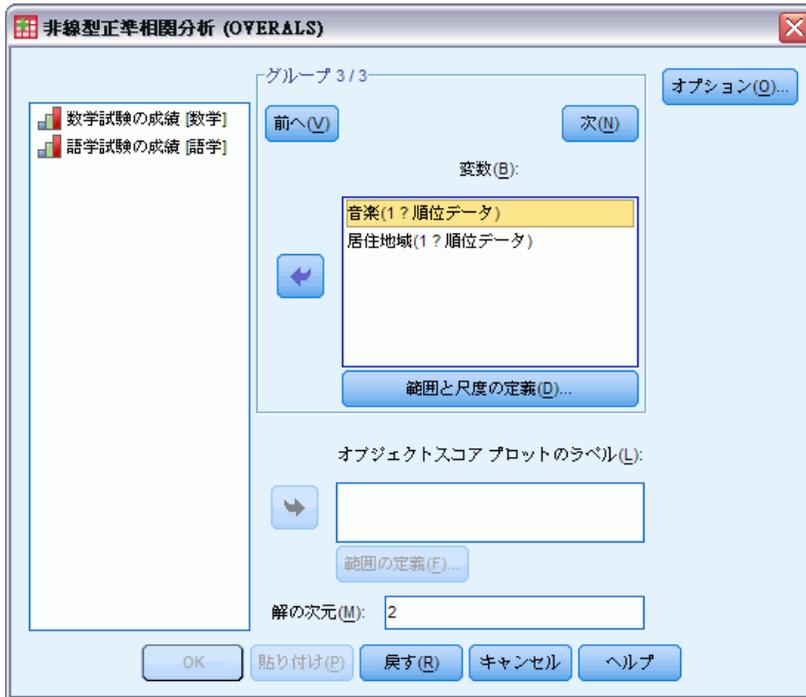
- ▶ この変数の最大値として「5」と入力します。
- ▶ 測定尺度として [多重名義] を選択します。
- ▶ [続行] をクリックします。
- ▶ 「新聞」を選択し、[非線型正準相関分析] ダイアログ ボックスの [範囲と尺度の定義] をクリックします。

図 11-7
[範囲と尺度の定義] ダイアログ ボックス



- ▶ この変数の最大値として「5」と入力します。
- ▶ 測定尺度として [単一名義] を選択します。
- ▶ [続行] をクリックします。
- ▶ [非線型正準相関分析] ダイアログ ボックスで [次] をクリックし、次の変数グループを定義します。

図 11-8
[非線型正準相関分析] ダイアログ ボックス



- ▶ 第 3 グループの変数として、「好きな音楽」と「居住したい地域」を選択します。
- ▶ 「音楽」を選択し、[範囲と尺度の定義] をクリックします。

図 11-9
[範囲と尺度の定義] ダイアログ ボックス



- ▶ この変数の最大値として「5」と入力します。
- ▶ 測定尺度として [単一名義] を選択します。
- ▶ [続行] をクリックします。

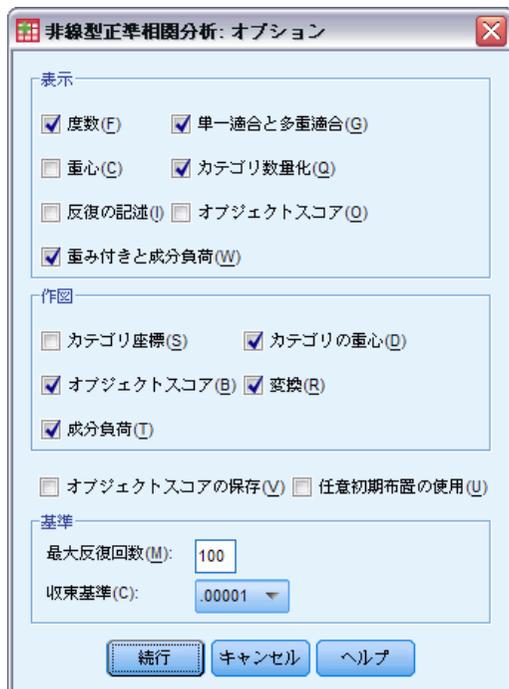
- ▶ 「居住地」を選択し、[非線型正準相関分析] ダイアログ ボックスの [範囲と尺度の定義] をクリックします。

図 11-10
[範囲と尺度の定義] ダイアログ ボックス



- ▶ この変数の最大値として「3」と入力します。
- ▶ 測定尺度として [単一名義] を選択します。
- ▶ [続行] をクリックします。
- ▶ [非線型正準相関分析] ダイアログ ボックスの [オプション] をクリックします。

図 11-11
[オプション] ダイアログ ボックス



- ▶ [表示] グループで [重心] の選択を解除し、[重み付きと成分負荷] を選択します。

- ▶ [作図] グループで [カテゴリの重心] と [変換] を選択します。
- ▶ [任意初期布置の使用] を選択します。
- ▶ [続行] をクリックします。
- ▶ [非線型正準相関分析] ダイアログ ボックスの [OK] をクリックします。

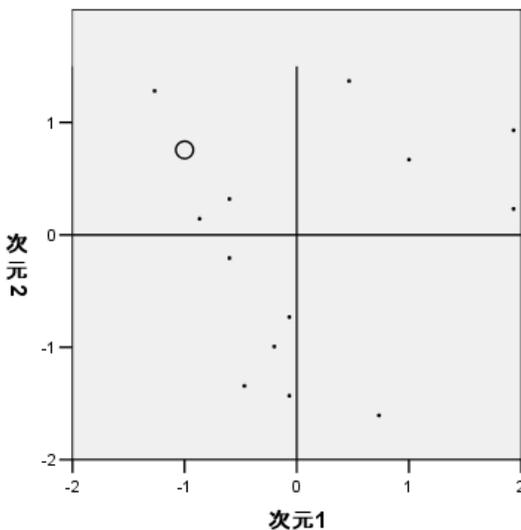
最適尺度法のカテゴリ正準相関分析による最適尺度水準の変数リストの後に、カテゴリ内のオブジェクトの度数を示すテーブルが作成されます。ほぼ空のカテゴリは解を支配する可能性が高いので、このテーブルは、欠損データがある場合に特に重要になります。この例では、欠損データはありません。

2 つ目の予備確認として、外れ値に関してオブジェクトスコアのプロットを調べます。外れ値には、他のオブジェクトと非常に異なる数量化があるので、外れ値はプロットの境界に位置し、1 つ以上の次元を支配します。

外れ値を見つけた場合は、次のいずれかの方法で処理できます。単に外れ値をデータから削除して、非線型正準相関分析を再実行できます。または、一部のカテゴリを集約（結合）して、外れ値のオブジェクトの極値応答数を記録できます。

オブジェクトスコアのプロットで示すように、調査データには外れ値がありません。

図 11-12
オブジェクトスコア



オブジェクト数により重み付けされたケース

グループ間の類似度の説明

非線型正準相関分析でグループ間の連関を測る方法はいくつかあります（各方法については、個別のテーブルまたはテーブルのセットで詳しく説明します）。

分析の要約

当てはめ値と損失値により、グループ間の連関について、非線型正準相関分析解が最適に数量化されたデータにどの程度適合しているかがわかります。分析の要約テーブルは、調査例の当てはめ値、損失値、および固有値を示します。

図 11-13
分析の要約

		次元		合計
		1	2	
損失	グループ 1	.240	.183	.423
	グループ 2	.184	.408	.593
	グループ 3	.171	.205	.376
	平均値	.199	.265	.464
固有値		.801	.735	
当てはめ				1.536

損失は、次元間とグループ間で分割されます。損失は次元とグループごとに、グループ内の変数の重み付き組み合わせでは説明できないオブジェクトスコアの分散の比率を表します。平均損失は、平均値とラベル付けされます。この例では、2つのグループの平均損失は0.464です。2番目の次元では、最初の次元よりも多くの損失が発生することに注意してください。

各次元の固有値は、1から対象の次元の平均損失を引いた値に相当し、次元ごとに説明される関係の程度を示します。固有値は、合計すると全体適合になります。Verdegaalのデータによると、実際の適合の $0.801 / 1.536 = 52\%$ が最初の次元で説明されます。

最大当てはめ値は次元数に相当し、関係が成り立っている場合は、その関係が完全なものであることを示します。2つのグループの平均損失値と次元により、最大当てはめ値と実際の当てはめ値との差異がわかります。当てはめ値に平均損失値を加えると、次元数になります。完全に一致することはほとんどないので、通常はデータの小さな相違点を利用することになります。

2つの変数グループを使用した別の一般的な統計量は、正準相関です。正準相関は固有値に関連し、追加情報は提供しないので、非線型正準相関分析の出力には含まれません。2つの変数グループの場合、次元ごとの正準相関は次の公式で得られます。

$$\rho_d = 2 \times E_d - 1$$

d は次元数で、E は固有値を表します。

次の公式を使用すると、2 つ以上のグループの正準相関を一般化できます。

$$\rho_d = ((K \times E_d) - 1) / (K - 1)$$

d は次元数、K はグループ数、E は固有値を表します。例をあげると、次のようになります。

$$\rho_1 = ((3 \times 0.801) - 1) / 2 = 0.702$$

かつ

$$\rho_2 = ((3 \times 0.735) - 1) / 2 = 0.603$$

重み付きと成分負荷 (Categories)

別の連関度として、各グループからの線型結合とオブジェクトスコア間の重相関があります。グループ内に多重名義の変数がない場合は、グループ内の各変数の重み付けと成分負荷を乗算し、これらの積を加算して、その合計の平方根を取得することで、この測定値を計算できます。

図 11-14
重み付け

グループ°	次元	
	1	2
1 年齢 ^a	.680	.789
婚姻状況 ^{a,b}	.296	-1.016
2 最も頻繁に読む新聞	-.845	-.361
3 好きな音楽 ^{c,b}	.631	-.749
居住したい地域 ^{c,b}	-.484	-.780

図 11-15
成分負荷

グループ°	次元	
	1	2
1 年齢 ^{a,b}	.834	.259
婚姻状況 ^{a,b}	.651	-.604
2 ペットの有無 ^{d,e}	.397	-.431
次元 1	-.277	.680
次元 2	-.667	-.391
最も頻繁に読む新聞 ^{a,b}	.786	-.500
3 好きな音楽 ^{c,b}	-.687	-.540
居住したい地域 ^{c,b}	-.687	-.540

- a. 最適尺度水準 順序
- b. オブジェクト空間内の単一数量化変数の射影
- c. 最適尺度水準 単一名義
- d. 最適尺度水準 多重名義
- e. オブジェクト空間内の多重数量化変数の射影

上図は、この例の変数の重み付けと成分負荷を示しています。オブジェクトスコアの最初の次元での最適尺度変数（年齢と婚姻状況）の最初の重み付き合計の重相関（R）は、次の公式で求めます。

$$R = \sqrt{(0.701 \times 0.841 + (-0.273 \times -0.631))} \\ = \sqrt{(0.5895 + 0.1723)} \\ = 0.873$$

次元ごとに、 $1 - \text{損失} = R^2$ になります。たとえば、分析の要約テーブルでは、 $1 - 0.238 = 0.762$ になります。これは、0.873 を 2 乗した数です（丸め誤差を加えています）。したがって、小さな損失値は、最適尺度変数の重み付き合計と次元間の重相関が大きいことを示します。重み付けは、多重名義変数に固有のものではありません。多重名義変数の場合は、グループごとに $1 - \text{損失}$ を使用します。

当てはめと損失の分割

各グループの損失は、非線型正準相関分析によって、いくつかの方法で分割されます。当てはめテーブルには、多重適合、単一適合、および単一損失のテーブルが示されます。これらのテーブルは、この調査例では、非線型正準相関分析で作成されています。多重適合から単一適合を引いた数は、単一損失になることに注意してください。

図 11-16
当てはめと損失の分割

グループ	多重当てはめ			単一当てはめ			単一損失		
	次元		合計	次元		合計	次元		合計
	1	2		1	2		1	2	
1 年齢	.494	.676	1.170	.462	.622	1.085	.032	.054	.085
婚姻状況	.089	1.033	1.122	.088	1.033	1.120	.001	.000	.001
2 ペットの有無	.402	.439	.841						
最も頻繁に読む新聞	.724	.187	.911	.714	.130	.844	.010	.057	.067
好きな音楽	.421	.577	.998	.398	.561	.960	.022	.016	.039
居住したい地域	.234	.609	.843	.234	.608	.843	.000	.000	.000

- a. 最適尺度水準 順序
b. 最適尺度水準 単一名義
c. 最適尺度水準 多重名義

単一損失は、変数を 1 つのグループの数量化（つまり、単一名義、順序、または名義）に制限することで生じる損失を示しています。単一損失が大きい場合は、変数を多重名義として扱う方が適しています。ただし、この例では、単一適合と多重適合はほとんど等しくなります。つまり、多重座標は、重み付けによって指定される方向の直線上にほとんど位置します。

多重適合は、各変数の多重カテゴリ座標の分散と等しくなります。このような測定方法は、等質性分析で使用される判別測定に類似しています。多重適合テーブルを調べると、どの変数が最も判別力が大きいかわかります。たとえば、「婚姻状況」と「最も頻繁に読む新聞」の多重適合表を参照してください。2 つの次元を合計した当てはめ値は、「婚姻状況」で

は 1.122 で、「最も頻繁に読む新聞」では 0.911 です。この情報により、ある人の婚姻状況は、購読新聞よりも判別力が大きいことがわかります。

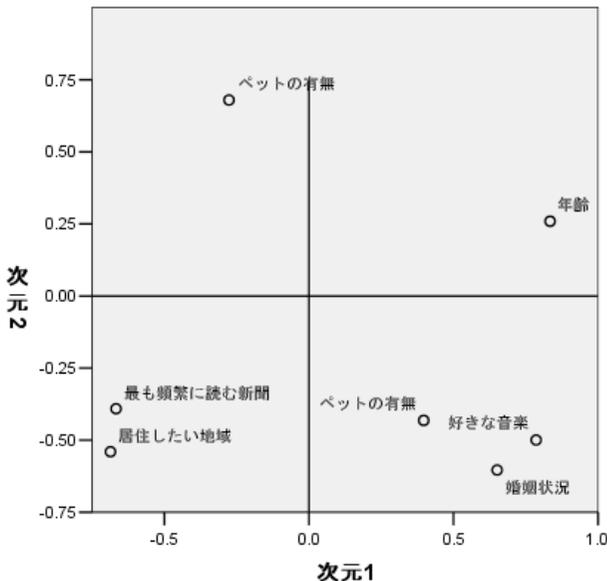
単一適合は、各変数の重み付けの 2 乗に一致し、単一カテゴリ座標の分散と等しくなります。したがって、重み付けは単一カテゴリ座標の標準偏差と等しくなります。単一適合が次元間でどのように分類されるかを調べると、変数「最も頻繁に読む新聞」は主に最初の次元で判別し、変数「婚姻状況」はほぼ全体的に 2 番目の次元で判別することがわかります。つまり、「最も頻繁に読む新聞」のカテゴリは、2 番目の次元よりも最初の次元で離れたところに位置し、「婚姻状況」の場合はパターンが逆になります。対照的に、「年齢」は最初の次元と 2 番目の次元の両方で判別します。したがって、カテゴリの広がりは両方の次元に沿って等しくなります。

成分負荷

次の図は、調査データの成分負荷のプロットを示しています。欠損データがない場合、成分負荷は、数量化された変数とオブジェクトスコア間で Pearson の相関と等しくなります。

原点から各変数ポイントまでの距離は、その変数の重要度に近くなります。正準変数はプロットされませんが、原点を通る水平線と垂直線で表すことができます。

図 11-17
成分負荷



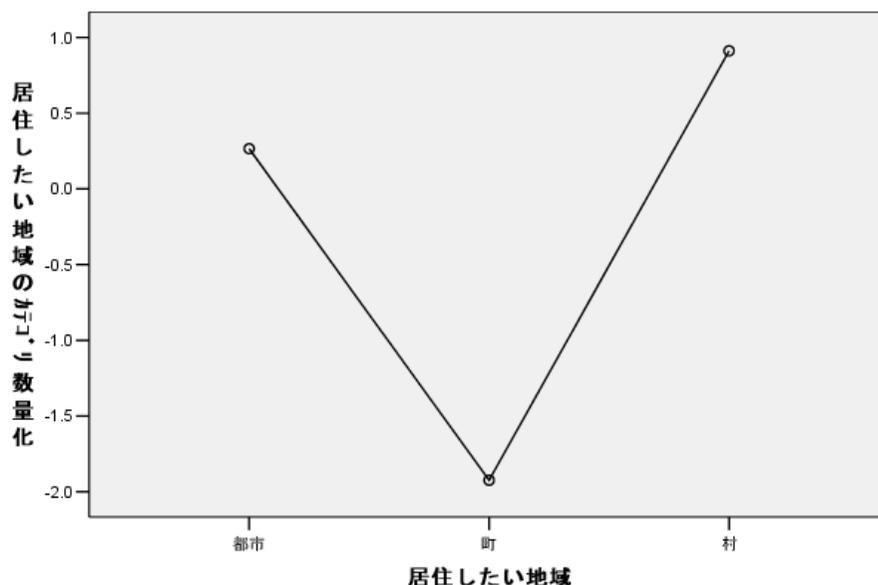
変数間の関係は明白です。水平軸や垂直軸と重ならない方向が2つあります。1つの方向は、「年齢」、「最も頻繁に読む新聞」、および「居住したい地域」で特定されます。もう一つの方向は、変数「婚姻状況」、「好きな音楽」、および「ペットの有無」で定義されます。「ペットの有無」変数は多重名義変数なので、プロットされるポイントが2つあります。各数量化は、1つの変数として解釈されます。

変換プロット

各変数を尺度化できるさまざまなレベルにより、数量化に制限を加えることができます。変換プロットは、選択された最適尺度水準から得られる数量化と元のカテゴリとの関係を表します。

名義として扱われた「居住したい地域」の変換プロットには、U字形のパターンが示されます。U字形のパターンでは、中央にあるカテゴリが最低の数量化を受け取り、極値カテゴリは互いに類似した値を受け取ります。このパターンは、元の変数と変換後の変数間の2次関係を示します。「居住したい地域」に対して代替の最適尺度水準を使用することはお勧めしません。

図 11-18
居住したい地域の変換プロット(名義)



対照的に、「最も頻繁に読む新聞」の数量化は、ケースを観測した3つのカテゴリにまたがる増加トレンドに対応します。最初のカテゴリは最低の数量化を受け取り、2番目のカテゴリは最初のカテゴリよりも高い値を受

け取ります。また、3番目のカテゴリは最高値を受け取ります。変数は名義として尺度化されますが、カテゴリの並び順は数量化で取得されます。

図 11-19
最も頻繁に読む新聞(名義)の変換プロット

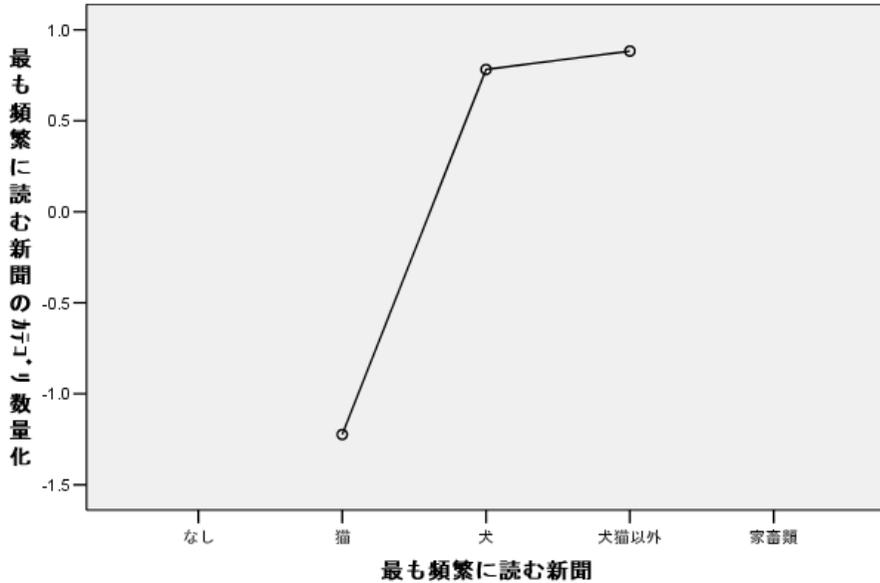
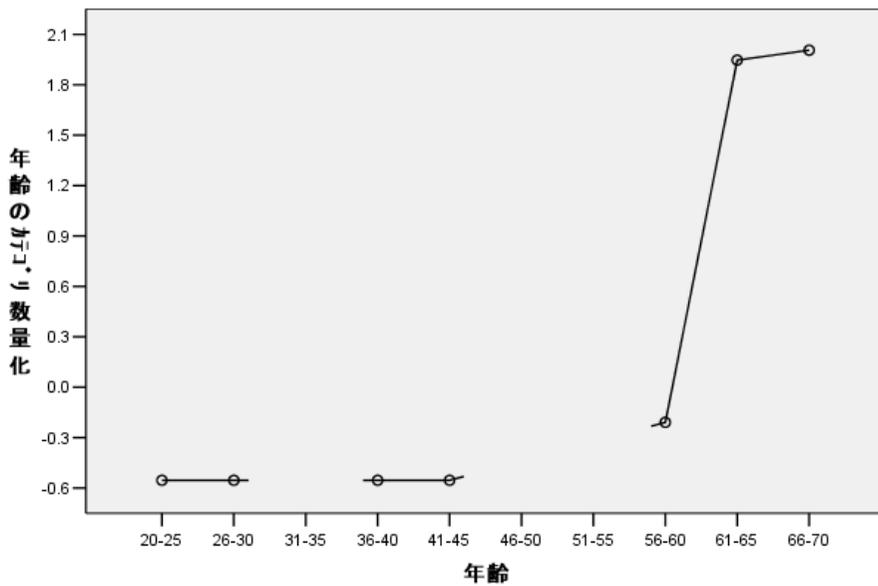


図 11-20
年齢(順序)の変換プロット



「年齢」の変換プロットには、S字型の曲線が示されます。最も年齢の若い4つの観測カテゴリはすべて同じ負の数量化を受け取り、最も年齢の高い2つのカテゴリは類似した正の値を受け取ります。したがって、すべての年齢の若いカテゴリを1つの共通カテゴリ（つまり、50歳未満）に集約し、最も年齢の高い2つのカテゴリを1つに集約するように試みることができます。ただし、年齢の若いグループの数量化に正確な等質性があることは、数量化の順序を元のカテゴリの順序に制限することは適さない場合があることを示しています。26～30、36～40、および41～45のグループの数量化を20～25グループの数量化より下げることはできないので、これらの値は境界値と等しくなるように設定します。このような値を最も若いグループの数量化よりも小さくできるようにすると（つまり、年齢を名義として扱うと）、適合が向上する場合があります。したがって、年齢は順序変数と見なされることがありますが、ここでは、年齢は適切でないものとして扱います。さらに、年齢を数値型として扱うと、カテゴリ間の距離が保たれるので、適合が実質的に減少します。

単一カテゴリ座標と多重カテゴリ座標

単一名義、順序、または数値型として扱われるすべての変数で、数量化、単一カテゴリ座標、および多重カテゴリ座標が特定されます。「年齢」に関するこのような統計量を示します。

図 11-21
「年齢」の座標

	周辺度数	数量化	単一カテゴリ座標		多重カテゴリ座標	
			次元		次元	
			1	2	1	2
20-25	3	-.554	-.377	-.437	-.192	-.139
26-30	5	-.554	-.377	-.437	-.404	-.623
31-35	0	.000				
36-40	1	-.554	-.377	-.437	-.318	-.733
41-45	1	-.554	-.377	-.437	-.356	-.534
46-50	0	.000				
51-55	0	.000				
56-60	2	-.209	-.142	-.165	-.435	.087
61-65	1	1.947	1.324	1.536	1.710	1.204
66-70	2	2.006	1.364	1.583	1.215	1.711
欠損値	0					

ケースが記録されなかったカテゴリはすべて0という数量化を受け取ります。[年齢]の場合は、このようなカテゴリに31-35、46-50、および51-55のカテゴリが含まれます。このようなカテゴリは、他のカテゴリとともに順序付けされるように制限されないため、計算には影響しません。

多重名義変数の場合、各カテゴリは各次元で異なる数量化を受け取ります。その他のすべての変換の種類では、解の次元にかかわらず、カテゴリは数量化を1つだけ受け取ります。単一カテゴリ座標の各グループは、対象空間内の直線上のカテゴリの位置を表します。特定のカテゴリの座標は、重み付けを掛けた数量化と等しくなります。たとえば、「年齢」の

テーブルでは、カテゴリ 56 ~ 60 の単一カテゴリ座標 (-0.142, -0.165) は、重み付け (0.680, 0.789) に数量化 (-0.209) を掛けたものになります。

単一名義、順序、または数値型として扱われる変数の多重カテゴリ座標は、順序または線型制約が適用される前に、対象空間内のカテゴリの座標を表します。これらの値は、制約されていない損失の最小値です。多重名義変数の場合、これらの座標はカテゴリの数量化を表します。

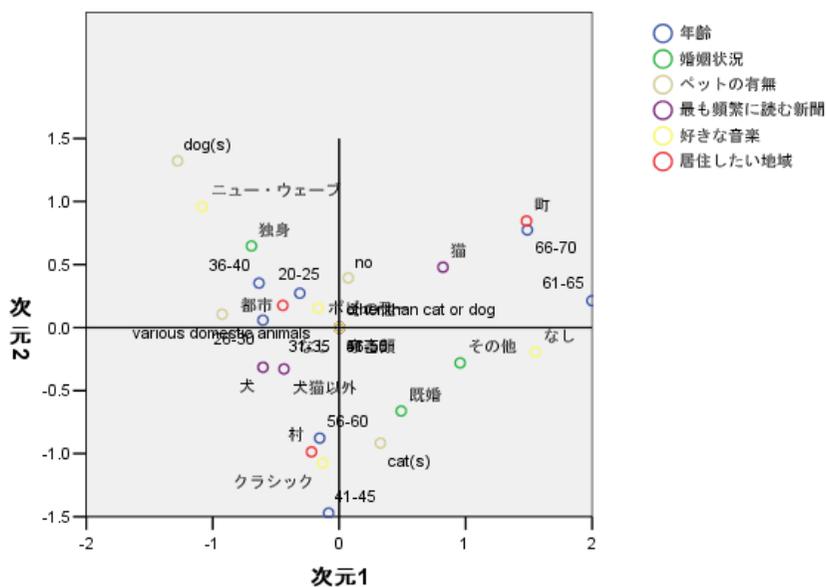
カテゴリとカテゴリの数量化との関係に制約条件を設けた場合の効果は、単一カテゴリ座標と多重カテゴリ座標を比較することで明らかになります。最初の次元で、「年齢」の多重カテゴリ座標は、カテゴリ 2 まで減少し、カテゴリ 9 まで相対的に同レベルの状態を保って、カテゴリ 9 で飛躍的に増加します。2 番目の次元でも同様のパターンが証明されます。このような関係は、単一カテゴリ座標では取り除かれます。単一カテゴリ座標では、順序に関する制約条件が適用されます。座標は、両方の次元で非減少になりました。2 組の座標の異なる構造は、名義として扱う方がより適切だということを示しています。

重心と射影された重心

変数でラベル付けされた重心のプロットは、等質性分析のカテゴリ数量化プロットまたは非線型主成分分析の多重カテゴリ座標と同じ方法で解釈される必要があります。このようなプロットは、それ自体で、オブジェクトのグループが変数によってどの程度適切に分離されるかを示します（重心はオブジェクトの質量の中心に位置します）。

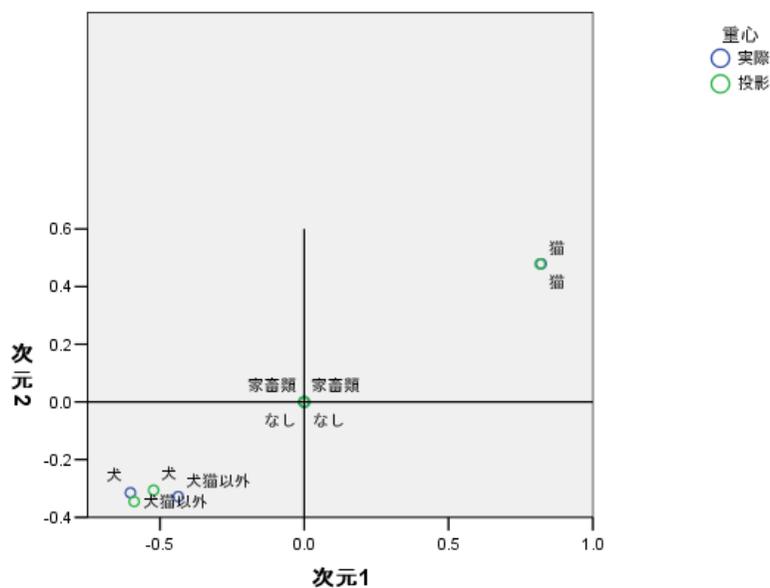
「年齢」のカテゴリは、ほとんど明確に分離されていないことに注意してください。年齢の若いカテゴリは、プロットの左側にまとめてグループ化されます。先に説明したように、順序は非常に厳密なので、「年齢」に尺度水準を適用できないことがあります。

図 11-22
変数でラベル付けされた重心



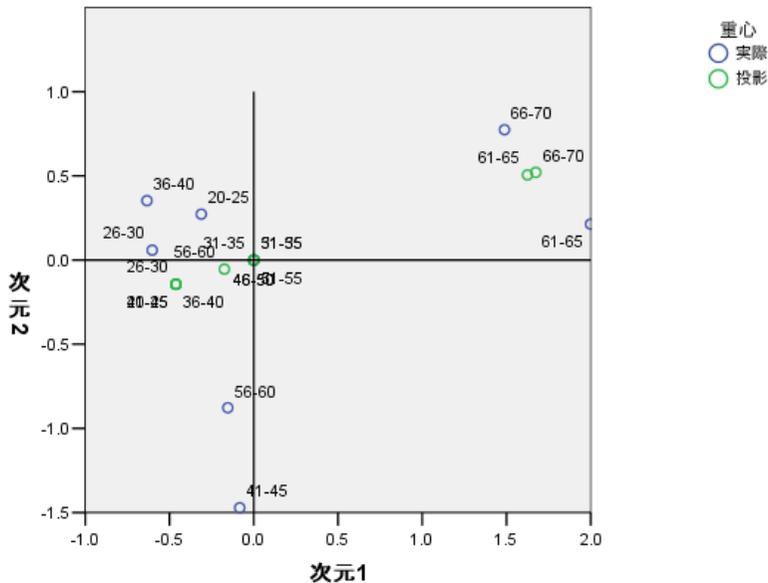
重心のプロットを要求すると、値ラベルでラベル付けされた各変数の個別の重心のプロットと射影された重心のプロットも作成されます。射影された重心は、対象空間の直線上に位置します。

図 11-23
「最も頻繁に読む新聞」の重心と射影された重心



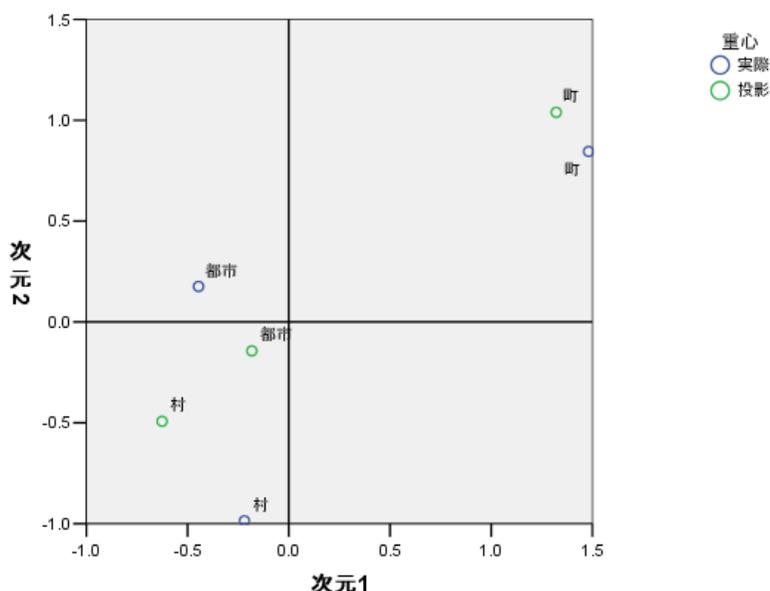
実際の重心は、成分負荷で定義されたベクトルに射影されます。これらのベクトルは、実際の重心から射影された重心を識別するときに役立つように、重心のプロットに追加されました。射影された重心は、原点を通る2つの垂直な参照線を伸ばすと形成される4象限のうちのいずれかに分類されます。単一名義、順序、または数値型の変数の方向は、射影された重心の位置から解釈できます。たとえば、変数「最も頻繁に読む新聞」を単一名義として指定したとします。射影された重心は、「Volkskrant」と「NRC」が「Telegraaf」と対比されることを示します。

図 11-24
「年齢」の重心と射影された重心



「年齢」に関する問題は、射影された重心から明らかになります。「年齢」を順序として扱うことは、年齢グループの順序を保持する必要があることを示します。この制限を満たすために、45歳未満のすべての年齢グループが同じポイントに射影されます。「年齢」、「最も頻繁に読む新聞」、および「居住したい地域」で定義された方向に沿って見てみると、年齢の若いグループは分離されません。したがって、変数を名義として扱うことが望まれます。

図 11-25
「居住したい地域」の重心と射影された重心



変数間の関係を理解するには、重心のプロットでカテゴリのクラスタに対する特定のカテゴリ（値）を確認します。「年齢」、「最も頻繁に読む新聞」、および「居住したい地域」間の関係は、プロットの右上と左下に注目すると説明できます。右上には、年齢グループの高い回答者が示されます。このグループの人たちは、Telegraaf 新聞を読み、村に居住することを好みます。各プロットの左下角に注目すると、若者から中高年までの回答者は Volkskrant または NRC を読み、地方または都市に居住することを望んでいることがわかります。ただし、より年齢の若いグループを分離することは非常に困難です。

重心のプロットの左上と右下に注目することで、他の方向（好きな音楽、婚姻状況、およびペットの有無）についても同じように解釈できます。左上角を見てみると、独身者は犬を飼い、新しいスタイルの音楽を好む傾向にあることがわかります。婚姻状況に関する「既婚」カテゴリや「その他」カテゴリに属する人は猫を飼っています。前者のグループはクラシック音楽を好み、後者のグループは音楽が好きではありません。

代替分析

分析結果から、「年齢」を順序として扱うことは適切でないことがわかります。「年齢」は順序レベルで測定されますが、他の変数との関係は単調ではありません。最適尺度水準を単一名義に変更する効果を調べるために、分析を再実行できます。

分析を実行するには

- ▶ [非線型正準相関分析] ダイアログ ボックスをリコールし、第 1 グループに移動します。
- ▶ 「年齢」を選択し、[範囲と尺度の定義] をクリックします。
- ▶ [範囲と尺度の定義] ダイアログ ボックスで [単一名義] を測定尺度として選択します。
- ▶ [続行] をクリックします。
- ▶ [非線型正準相関分析] ダイアログ ボックスの [OK] をクリックします。

2 次元の解の固有値はそれぞれ 0.806 と 0.757 で、全体適合は 1.564 です。

図 11-26
2 次元の解の固有値

		次元		合計
		1	2	
損失	グループ 1	.249	.115	.363
	グループ 2	.176	.408	.584
	グループ 3	.157	.205	.363
	平均値	.194	.243	.436
固有値		.806	.757	
当てはめ				1.564

多重適合テーブルと単一適合テーブルは、多重適合値の合計からもわかるように、「年齢」が依然として判別力の高い変数であることを示します。ただし、以前の結果とは対照的に、単一適合値を調べると、判別がほぼ完全に 2 番目の次元に沿っていることが明らかになります。

図 11-27
当てはめと損失の分割

グループ	多重当てはめ			単一当てはめ			単一損失		
	次元		合計	次元		合計	次元		合計
	1	2		1	2		1	2	
1 年齢 ^a	.246	1.197	1.443	.195	1.188	1.384	.051	.008	.059
婚姻状況 ^b	.273	1.136	1.409	.272	1.135	1.407	.001	.000	.002
2 ペットの有無 ^b	.530	.392	.921						
最も頻りに読む新聞 ^a	.639	.185	.824	.631	.149	.780	.008	.036	.044
3 好きな音楽 ^a	.604	.438	1.041	.603	.437	1.040	.000	.001	.001
居住したい地域 ^a	.075	.822	.897	.075	.822	.897	.000	.000	.000

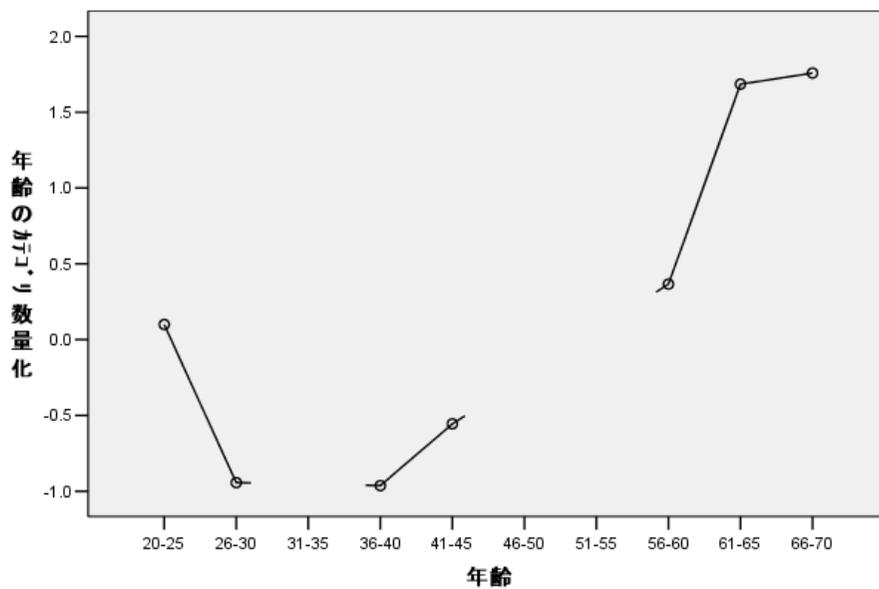
a. 最適尺度水準 単一名義

b. 最適尺度水準 多重名義

次に、「年齢」の変換プロットについて考えます。名義変数の数量化は制限されていないので、「年齢」を順序変数として扱ったときに示された非減少トレンドは、示されなくなりました。U 字型 (2 次) 関係に対応する、40 歳までの減少トレンドと、それに続く増加トレンドが存在します。年齢

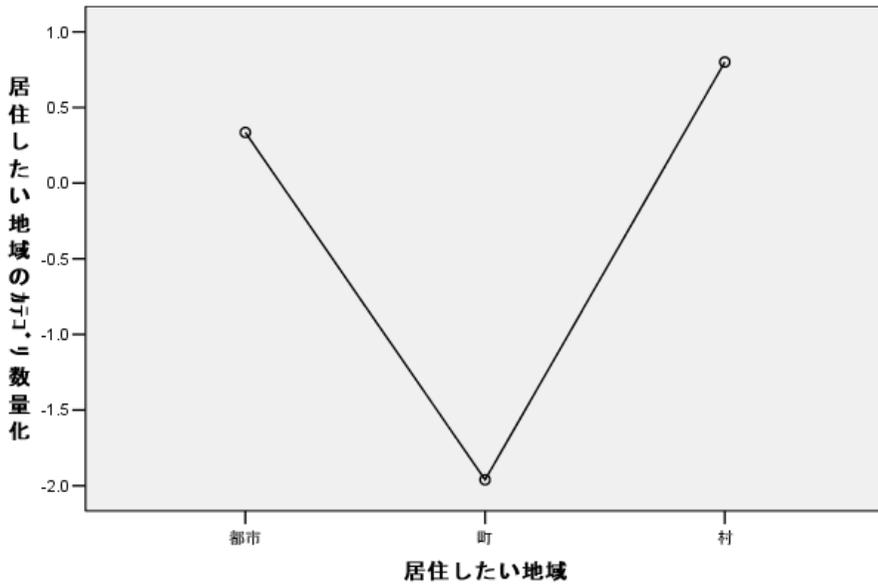
の高い2つのカテゴリは依然として同様の得点を受け取り、それ以降の分析に、これらのカテゴリの結合が含まれることがあります。

図 11-28
年齢(名義)の変換プロット



「居住したい地域」の変換プロットを以下に示します。「年齢」を名義として扱っても、「居住したい地域」の数量化には重大な影響を及ぼしません。中央のカテゴリは最小の数量化を受け取り、極値カテゴリは大きな正の値を受け取ります。

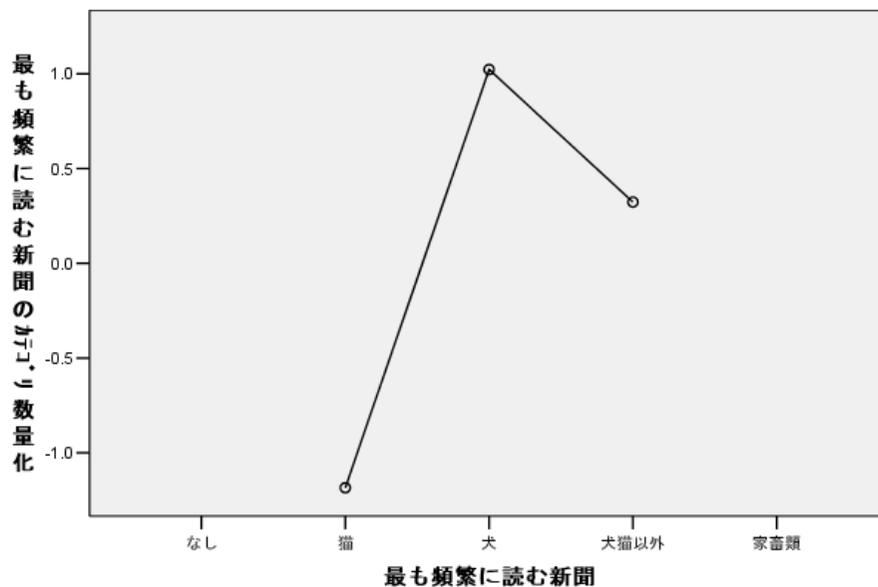
図 11-29
居住したい地域の変換プロット（「年齢」を名義として扱う）



「最も頻繁に読む新聞」の変換プロットには、変化が見られます。前回の分析では、増加トレンドが数量化に存在しました。このことは、この変数が順序として扱われていることを示している可能性があります。ただ

し、「年齢」を名義として扱うと、このトレンドがニュースの数量化から取り除かれます。

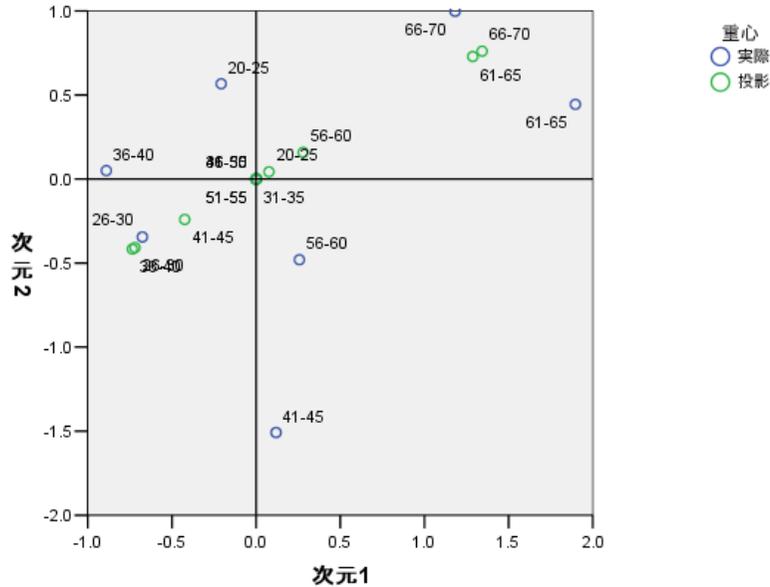
図 11-30
最も頻繁に読む新聞の変換プロット（「年齢」を名義として扱う）



以下のプロットは、「年齢」の変換プロットです。カテゴリは、射影された重心を結合する線に沿って、時系列順に分類されないことに注意してください。20～25のグループは、端ではなく中央に位置します。カ

テゴリの広がり、前回の分析で示された順序に対応するものよりもはるかに改善されています。

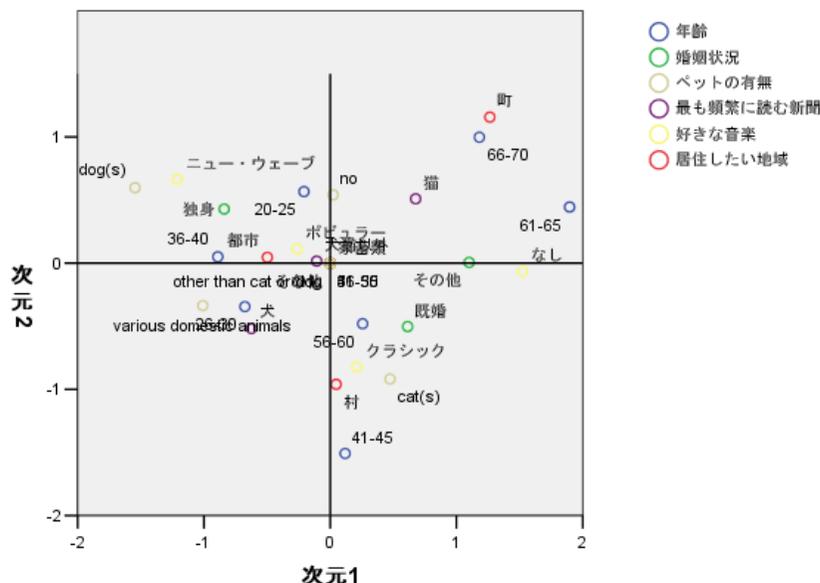
図 11-31
年齢の重心と射影された重心(名義)



年齢の若いグループの解釈は、重心のプロットで可能になりました。また、「Volkskrant」と「NRC」のカテゴリは、前回の分析よりも離れているので、それぞれ個別に解釈できます。26～45歳までのグループはVolkskrantを読み、地方に居住することを好みます。20～25歳のグループと56～60歳のグループはNRCを読みます。また、前者のグループは都市に居住することを好み、後者のグループは地方に居住することを好みます。最高齢のグループはTelegraafを読み、村に居住することを好みます。

他の方向（好きな音楽、婚姻状況、およびペットの有無）の解釈は、基本的に前回の分析と変わりありません。唯一、婚姻状況が「その他」の人は猫を飼っているか、ペットを飼っていないという明らかな違いがあります。

図 11-32
変数でラベル付けされた重心（「年齢」を名義として扱う）



一般的な提言

最初の結果の調査が終わったら、非線型正準相関分析に関する指定を一部変更することで、分析の精度向上が望まれる場合があります。ここで、分析を構造化するためのヒントをいくつか示します。

- 可能な限り多くのグループを作成する。個別のグループに、予測する重要な変数を単独で追加します。
- 予測変数と見なす変数は1つのグループにまとめて追加する。予測変数が多数ある場合は、複数のグループに分割します。
- それぞれの多重名義変数を個別のグループに単独で追加する。
- 変数の相互相関度が高く、この関係によって解が影響を受けないようにする場合は、そのような変数を同じグループにまとめて追加する。

推奨参考文献

非線型正準相関分析の詳細は、次のテキストを参照してください。

Carroll, J. D. 1968. Generalization of canonical correlation analysis to three or more sets of variables. In: Proceedings of the 76th Annual Convention of the American Psychological Association, 3, Washington, D.C.: American Psychological Association.

De Leeuw, J. 1984. Canonical analysis of categorical data, 2nd ed. Leiden: DSWO Press.

Horst, P. 1961. Generalized canonical correlations and their applications to experimental data. *Journal of Clinical Psychology*, 17, .

Horst, P. 1961. Relations among m sets of measures. *Psychometrika*, 26, .

Kettenring, J. R. 1971. Canonical analysis of several sets of variables. *Biometrika*, 58, .

Van der Burg, E. 1988. Nonlinear canonical correlation and some related techniques. Leiden: DSWO Press.

Van der Burg, E., および J. De Leeuw. 1983. Nonlinear canonical correlation. *British Journal of Mathematical and Statistical Psychology*, 36, .

Van der Burg, E., J. De Leeuw, および R. Verdegaal. 1988. Homogeneity analysis with k sets of variables: An alternating least squares method with optimal scaling features. *Psychometrika*, 53, .

Verboon, P., および I. A. Van der Lans. 1994. Robust canonical discriminant analysis. *Psychometrika*, 59, .

コレスポネンス分析

対応分析

コレスポネンス テーブルは、セルに、行と列間の対応度が記された 2 元表です。対応度により、行変数と列変数間の類似度、類似点、曖昧さ、相互作用をすべて示すことができます。非常によく使用されるコレスポネンス テーブルは、セルに度数が含まれているクロス集計表です。

このようなテーブルを、クロス集計表手続きを使用して簡単に作成できます。しかし、クロス集計表を使用しても、2 つの変数の間に存在する関係の性質を適切に説明できない場合があります。これは特に、対象の変数が名義型（本質的に次数や順位がない型）であり、膨大な数のカテゴリを持っている場合に当てはまります。クロス集計表では、観測セル度数が、「職業」および「朝食用シリアル」を示す 10×9 のクロス集計表の期待値と大きく異なることが示される場合があります。しかし、似たような味を好む職業グループや好まれた味はどれなのか、ということを理解するのは困難な場合があります。

コレスポネンス分析では、2 つの名義変数間の関係を、図表を使用して多次元空間で調べることができます。コレスポネンス分析では行得点と列得点が計算され、得点に基づいてプロットが作図されます。プロットでは、互いに類似したカテゴリが近い位置に表示されます。このように、ある変数のカテゴリが互いに類似しているカテゴリや、2 つの変数のカテゴリの間に関連が見られるカテゴリを簡単に表示できます。また、コレスポネンス分析手続きを使用して、アクティブなポイントによって定義された空間に補助ポイントを当てはめることができます。

得点に従ってカテゴリを並べ替えることが望ましくなく、直感的ではない場合、カテゴリの得点が等しくなるように制約を設けることにより、並べ替えを強制的に制限できます。たとえば、変数「喫煙量」がカテゴリ [吸わない]、[少]、[中]、および [多] を持ち、この順序に対応する得点が与えられているとします。ただし、[吸わない]、[少]、[多]、および [中] のカテゴリを分析で並べる場合には、[多] と [中] の得点が等しくなるように制限して、カテゴリの並び順を得点順に保ちます。

コレスポネンス分析における解釈の距離は、使用する正規化の方法によって異なります。コレスポネンス分析手続きでは、1 つの変数が属する複数のカテゴリ間の相違、または、複数の変数間の相違のいずれかを分析できます。デフォルトの正規化では、行変数と列変数の相違を分析できます。

コレスポンデンス分析のアルゴリズムを使用してさまざまな種類の分析を行えます。行と列を中心化しカイ 2 乗距離を使用することは、標準のコレスポンデンス分析に相当します。ただし、ユークリッド距離と組み合わせた中心化方法を選択すると、低次元空間の行列を異なる形で表現できます。

ここでは 3 つの例を示します。最初の例では、比較的小さなコレスポンデンス テーブルを使用してコレスポンデンス分析特有の概念を説明します。2 つ目の例では、マーケティングでの実用化について実例を挙げて説明します。最後の例では、多次元尺度法での距離を示した表を使用します。

正規化

正規化は行得点と列得点全体にイナーシャを広げるために使用されます。さまざまな正規化がありますが、コレスポンデンス分析の解には、特異値、次元ごとのイナーシャ、寄与率など、正規化後も変化しない側面があります。行得点と列得点、およびそれらの分散は影響を受けます。コレスポンデンス分析で、複数の方法でイナーシャを広げることができます。最も一般的な 3 つの方法として、行得点のみにイナーシャを広げる方法、列得点のみにイナーシャを広げる方法、および行得点と列得点の両方に対称的にイナーシャを広げる方法があります。

行主成分。行主成分の正規化では、行ポイント間のユークリッド距離がコレスポンデンス テーブルの行間のカイ 2 乗距離に近似します。行得点は、列得点の平均値を重み付けしたものです。列得点は標準化され、重心 1 までの距離を 2 乗した和に重み付けされます。この方法では行カテゴリ間の距離が最大になるので、行変数のカテゴリ間の相違を調べる場合は、行主成分の正規化を使用する必要があります。

列主成分。一方、コレスポンデンス テーブルの列間のカイ 2 乗距離を近づけたい場合があります。この場合、列得点は行得点の平均値に重み付けしたものになります。行得点は、重心 1 までの距離を 2 乗した和に重みを付けた値を持つように標準化されます。この方法では列カテゴリ間の距離が最大になるので、列変数のカテゴリ間の相違を調べる場合に使用します。

対称的。行と列を対称的に扱うこともできます。この正規化では、行得点と列得点全体に均一にイナーシャを広げます。この場合は、行ポイント間の距離も列ポイント間の距離もカイ 2 乗距離の近似ではないことに注意してください。2 つの変数間の相違または類似性を重点的に調べるには、この方法を使用します。この方法は通常、バイプロットを作成する場合に適しています。

主成分。4 つ目の方法は主成分の正規化と呼ばれ、1 回は行得点全体で、1 回は列得点全体でというように、イナーシャを解の中で 2 回広げます。行ポイントと列ポイント間相互の関係を調べるのではなく、行ポイント間の距離と列ポイント間の距離を個別に調べるには、この方法を使用します。

バイプロットはこの正規化方法に適さないので、主成分の正規化方法を指定した場合にはバイプロットを使用できません。

例:コーヒー ブランドの印象

先ほどの例では架空のデータで構成された小さなテーブルを使用しました。実用では、通常、はるかに大規模なテーブルを使用します。この例では、6つのアイス コーヒー ブランド (Kennedy, Riquier, および Sharp, 1996) について受けた印象に関連するデータを使用します。このデータセットは coffee.sav にあります。詳細は、A 付録 p. 321 [サンプル ファイル](#) を参照してください。

回答者は、アイス コーヒーに対する 23 の各印象属性に対して、その属性が言い表していると思われるすべてのブランドを選択しました。機密保持のため、6つのブランドを [AA]、[BB]、[CC]、[DD]、[EE]、および [FF] で表しています。

テーブル 12-1
アイス コーヒーの属性

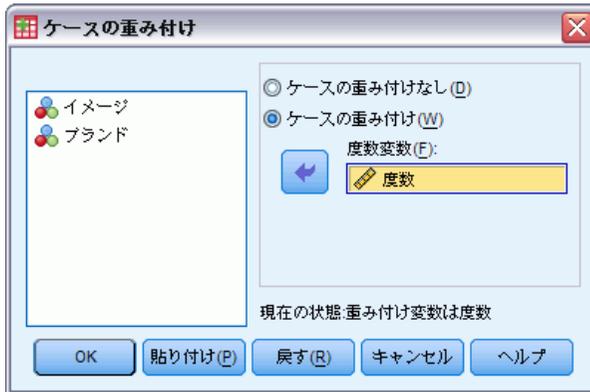
イメージ	ラベル	イメージ	ラベル
二日酔いに効果がある	回復	太るイメージ	太る
低脂肪/低カロリー	低脂肪	男性受けがよい	男性
子供向け	子供	南オーストラリア風	南オーストラリア人
労働者向け	実用的	伝統的/古風	伝統的
香り豊か/甘い	甘い	プレミアム/高品質	プレミアム
不人気なイメージ	不人気	健康的なイメージ	健康的
太った/ハンサムでないイメージ	面倒	高カフェイン	カフェイン
とても新鮮	フレッシュ	新しいイメージ	新規作成
都会的なイメージ	都会的	魅力的なイメージ	魅力的
栄養価の高いイメージ	栄養価の高い	タフなイメージ	タフ
女性向け	女性	人気のあるイメージ	人気がある
マイナーなイメージ	マイナー		

最初に、属性間の関係とブランド間の関係に注目します。主成分の正規化を使用すると一度にイナーシャ総計が行と列全体に広がります。これはバイプロットの解釈を妨げることとなりますが、こうすることでカテゴリ間の距離を変数ごとに調べることができます。

分析の実行

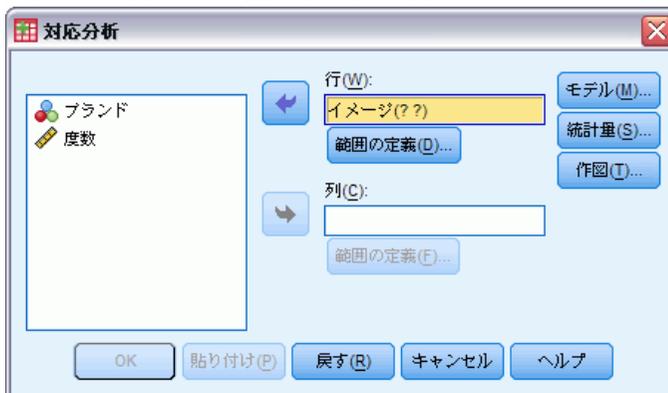
- ▶ データをセットアップするには、ケースに変数「度数」で重みを付ける必要があります。これを行うためには、メニューから次の項目を選択します。
データ > ケースの重み付け(W)...

図 12-1
[ケースの重み付け] ダイアログ ボックス



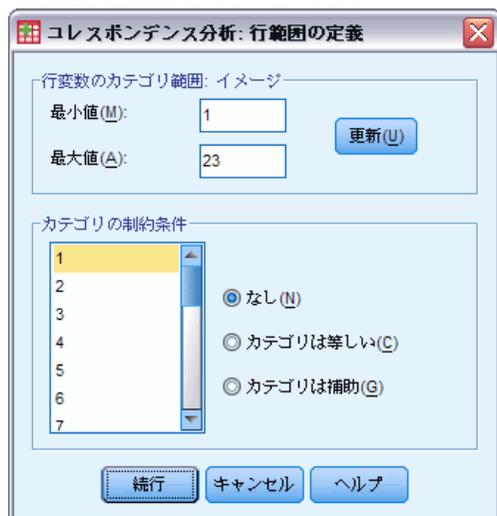
- ▶ 「度数」でケースに重みを付けます。
- ▶ [OK] をクリックします。
- ▶ 主成分の正規化を使用して 5 つの次元で初期の解を求めるには、メニューから次の項目を選択します。
分析(A) > 次元分解 > コレスポネンス分析...

図 12-2
[コレスポネンス分析] ダイアログ ボックス



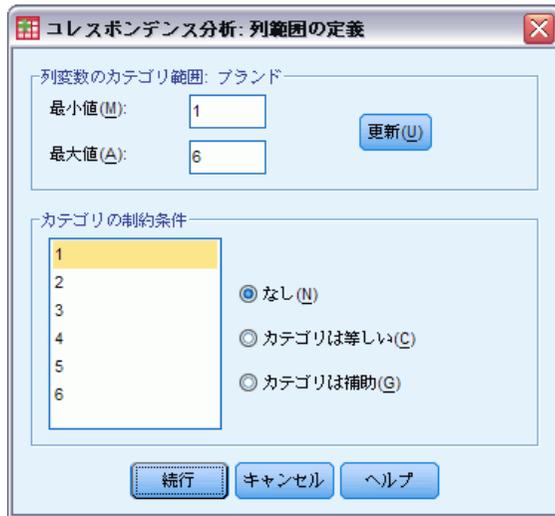
- ▶ 行変数として「イメージ」を選択します。
- ▶ [範囲の定義] をクリックします。

図 12-3
[行範囲の定義] ダイアログ ボックス



- ▶ 最小値として「1」を入力します。
- ▶ 最大値として「23」と入力します。
- ▶ [更新] をクリックします。
- ▶ [続行] をクリックします。
- ▶ 列変数として「ブランド」を選択します。
- ▶ [コレスポネンス分析] ダイアログ ボックスの [範囲の定義] をクリックします。

図 12-4
[列範囲の定義] ダイアログ ボックス



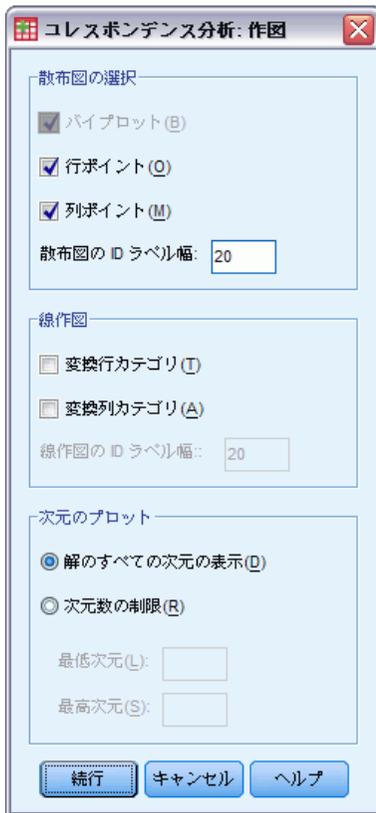
- ▶ 最小値として「1」を入力します。
- ▶ 最大値として「6」と入力します。
- ▶ [更新] をクリックします。
- ▶ [続行] をクリックします。
- ▶ [コレスポネンダ分析] ダイアログ ボックスで [モデル] をクリックします。

図 12-5
[モデル] ダイアログ ボックス



- ▶ 正規化方法として [主成分] を選択します。
- ▶ [続行] をクリックします。
- ▶ [コレスポネンス分析] ダイアログ ボックスで [作図] をクリックします。

図 12-6
[作図] ダイアログ ボックス



- ▶ [散布図の選択] グループで [行ポイント] と [列ポイント] を選択します。
- ▶ [続行] をクリックします。
- ▶ [コレスポネンダ分析] ダイアログ ボックスで [OK] をクリックします。

次元数

次元ごとのイナーシャは、各次元におけるイナーシャ総計の分割結果を示します。2つの次元でイナーシャ総計の83%を占めています。第3次元を追加しても、計上されるイナーシャは8.6%しか増えません。したがって、2次元表現を使用します。

図 12-7
次元ごとのイナーシャ

次元	特異値	要約イナ	加2乗	有意確率	イナ		信頼特異値	
					説明	累積	標準偏差	相関 2
1	.711	.506			.629	.629	.009	.132
2	.399	.159			.198	.827	.014	
3	.263	.069			.086	.913		
4	.234	.055			.068	.982		
5	.121	.015			.018	1.000		
要約合計		.804	3746.968	.000 ^a	1.000	1.000		

a. 自由度110

寄与率 (コレスポネンス分析)

行ポイントの概要には、次元のイナに対する行ポイントの寄与率と、行ポイントのイナに対する次元の寄与率が示されています。すべてのポイントがイナに等しく寄与していると、寄与率は0.043になります。[健康的]と[低脂肪]は、どちらも大部分が第1次元のイナへ寄与しています。[男性]と[タフ]は、第2次元のイナに最も多く寄与しています。[面倒]と[フレッシュ]は、両方ともどちらの次元に対してもほとんど寄与していません。

図 12-8
属性の寄与率

イメージ	λ	次元の得点		要約イナ	寄与率				概要合計
		1	2		次元のイナに対するポイント		ポイントのイナに対する次元		
					1	2	1	2	
太る	.080	-.514	-.265	.033	.042	.035	.652	.173	.825
男性	.051	-.852	.825	.072	.073	.219	.512	.480	.992
南オーストラリア	.057	-.303	-.350	.046	.010	.044	.114	.152	.266
トラディショナル	.040	-.703	-.532	.043	.039	.071	.454	.260	.715
プレミアム	.042	-.444	-.582	.028	.016	.090	.296	.509	.805
ヘルシー	.053	1.200	.174	.081	.152	.010	.953	.020	.973
カフェイン	.047	-.452	.124	.014	.019	.005	.702	.053	.755
新しい	.047	.960	.147	.048	.086	.006	.893	.021	.914
魅力がある	.041	.657	-.056	.019	.035	.001	.911	.007	.918
タフ	.039	-.850	1.002	.070	.056	.246	.404	.560	.964
人気	.060	-.697	-.042	.038	.058	.001	.771	.003	.774
癒し	.026	-.389	.266	.009	.008	.011	.446	.209	.655
ローファット	.052	1.305	.196	.094	.175	.013	.941	.021	.962
子供	.024	-.352	-.513	.017	.006	.041	.179	.380	.559
労働	.045	-.785	.477	.040	.055	.064	.693	.255	.948
スウィート	.038	-.519	-.683	.048	.020	.112	.212	.368	.580
不評	.024	.489	.186	.010	.011	.005	.585	.085	.670
不快	.030	.006	-.109	.003	.000	.002	.000	.131	.131
フレッシュ	.036	-.096	-.100	.002	.001	.002	.196	.214	.410
都会的	.034	.380	-.301	.012	.010	.019	.392	.246	.637
滋養	.040	.722	.055	.022	.041	.001	.946	.006	.951
女性	.054	.758	-.063	.032	.062	.001	.965	.007	.972
マイナー	.040	.579	.063	.023	.027	.001	.593	.007	.600
合計	1.000			.804	1.000	1.000			

2つの次元が、大部分の行ポイントのイナーシャに対して大きく寄与しています。[健康的]、[新しい]、[魅力的]、[低脂肪]、[栄養価の高い]、および[女性]に対する第1次元の大きな寄与は、これらのポイントが1つの次元で非常によく表現されていることを示します。その結果、第2次元以上では、横軸に非常に近い位置にあるこれらのポイントのイナーシャに対してほとんど寄与しません。第2次元は[男性]、[プレミアム]、および[タフ]に最も多く寄与しています。両方の次元は[南オーストラリア人]と[面倒]のイナーシャに対してはほとんど寄与していないので、これらのポイントは適切に表現されていません。

列ポイントの概要には、列ポイントに含まれている寄与率が表示されます。ブランド[CC]と[DD]が最も大きく第1次元に寄与しており、ブランド[EE]と[FF]は第2次元の大量のイナーシャを説明しています。[AA]と[BB]はどちらの次元に対してもほとんど寄与していません。

図 12-9
ブランドの寄与率

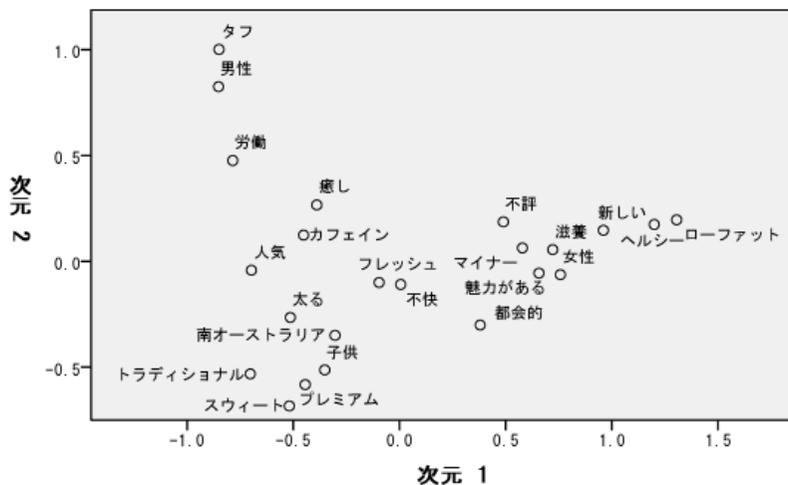
ブランド	Σ	次元の得点		概要イシャ	寄与率				
		1	2		次元のイシャに対するポイント		ポイントのイシャに対する次元		概要合計
					1	2	1	2	
AA	.217	-.659	.046	.127	.187	.003	.744	.004	.748
BB	.131	-.284	-.404	.078	.021	.134	.135	.272	.407
CC	.185	.996	.076	.193	.362	.007	.951	.006	.957
DD	.162	.915	.101	.146	.267	.010	.928	.011	.939
EE	.152	-.651	.706	.153	.127	.477	.420	.494	.914
FF	.153	-.343	-.618	.107	.036	.369	.169	.550	.718
合計	1.000			.804	1.000	1.000			

2つの次元で、[BB]を除くすべてのブランドが適切に表現されています。[CC]と[DD]は1つの次元で適切に表現されています。第2次元は[EE]と[FF]に最も大きく寄与しています。[AA]が第1次元で適切に表現されており、かつ、第1次元への寄与率がそれほど高くないことに注意してください。

作図

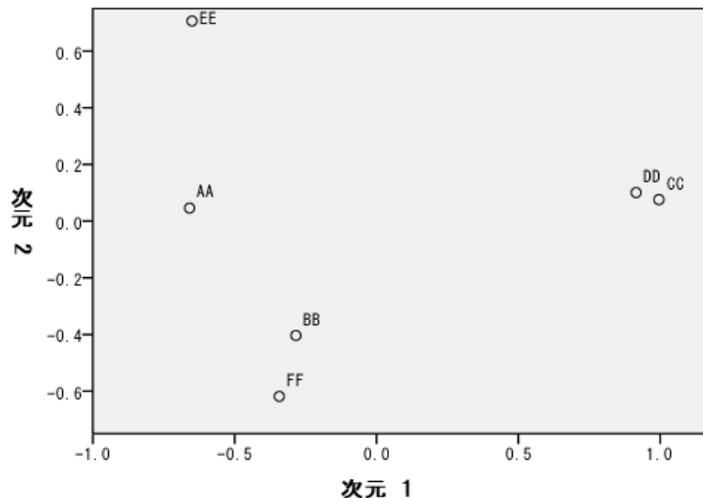
行得点のプロットでは、[フレッシュ]と[面倒]がどちらも原点近くに表示され、平均行のプロファイルとほとんど変わらないことを示しています。3つの一般的な区分が表示されます。プロットの左上の[タフ]、[男性]、および[実用的]は、どれも互いに類似しています。左下には、[甘い]、[太る]、[子供]、および[プレミアム]が示されています。一方、[健康的]、[低脂肪]、[栄養価の高い]、および[新しい]のグループがプロットの右側に示されています。

図 12-10
印象属性のプロット (主成分の正規化)



列ポイントのプロットではすべてのブランドが原点から遠くに示されているので、全体の重心に類似するブランドがないことに注意してください。ブランド [CC] と [DD] のグループは、右に一緒に示されており、[BB] と [FF] のグループはプロットの下半分に示されています。ブランド [AA] と [EE] は他のどのブランドとも類似しません。

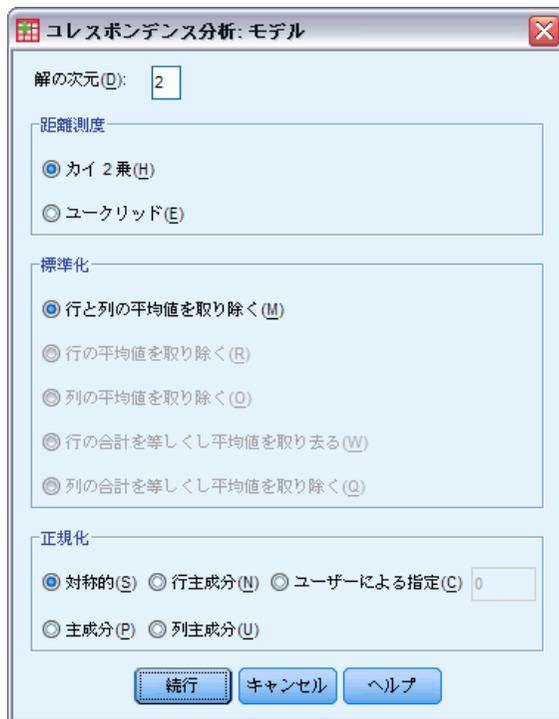
図 12-11
ブランドのプロット (主成分の正規化)



対称の正規化

ブランドは印象属性とどのように関わっているのでしょうか。主成分の正規化ではこのような関係を扱えません。変数が互いにどのような関係にあるかを調べるには、対称の正規化を使用します。対称の正規化では、主成分の正規化のようにイナーシャを2倍に広げるのではなく、行と列の両方でイナーシャを均一に分割します。1つの変数に対するカテゴリ間の距離は解釈できませんが、異なる変数に対するカテゴリ間の距離には意味があります。

図 12-12
[モデル] ダイアログ ボックス

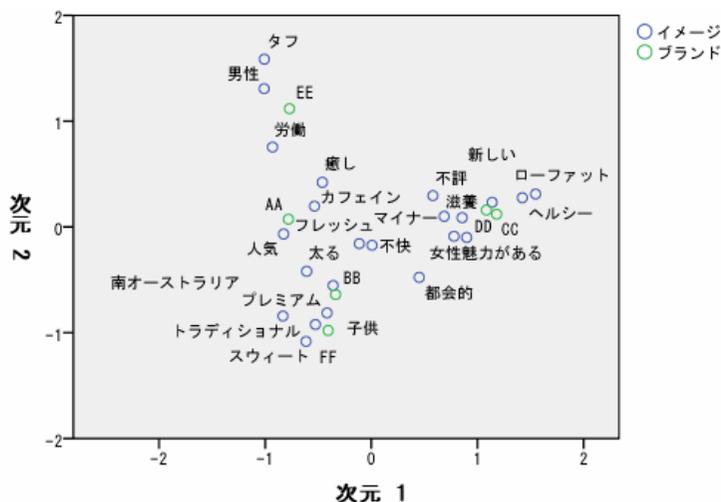


- ▶ 対称の正規化を使用して次の解を求めるには、[コレスポネンス分析] ダイアログ ボックスを再表示し、[モデル] をクリックします。
- ▶ 正規化方法として [対称的] を選択します。
- ▶ [続行] をクリックします。
- ▶ [コレスポネンス分析] ダイアログ ボックスで [OK] をクリックします。

作成されたバイプロットの左上では、ブランド [EE] が唯一力強く、労働者階級向けブランドで、かつ、男性に魅力的であることが示されます。ブランド [AA] は最も人気があり、かつ、カフェインが多いと思われる

ます。甘く、太りそうなブランドには [BB] と [FF] が入っています。ブランド [CC] と [DD] は新しく健康的と受け取られている一方で、最も人気がないという面もあります。

図 12-13
ブランドと属性のパイプロット (対称の正規化)



さらに詳しく解釈するために、原点と2つの印象属性 [男性] および [都会的] を通る線を引き、ブランドをこの線に射影します。2つの属性が互いに反対側にあり、[男性] のブランドの関連パターンが [都会的] のパターンの逆になっていることが示されます。つまり、男性はブランド [EE] と最も強く関連付けられ、かつ、ブランド [CC] との関連度は弱いのですが、若い都会派はブランド [CC] と最も強く関連付けられ、かつ、ブランド [EE] との関連度が弱くなっています。

推奨参考文献

コレスポネンス分析の詳細は、次のテキストを参照してください。

Fisher, R. A. 1938. Statistical methods for research workers. Edinburgh: Oliver and Boyd.

Fisher, R. A. 1940. The precision of discriminant functions. Annals of Eugenics, 10, .

Gilula, Z., および S. J. Haberman. 1988. The analysis of multivariate contingency tables by restricted canonical and restricted association models. Journal of the American Statistical Association, 83, .

多重応答分析

多重応答分析は等質性分析とも呼ばれ、カテゴリを可能な限り他のカテゴリから分離するための最適な数量化を見つけることを目的としています。このことは、同一カテゴリ内のオブジェクトは互いに近くにプロットされ、異なるカテゴリ内のオブジェクトは可能な限り離れた位置にプロットされることを表します。また、[等質性] という用語は、変数が等質のとき、つまり、変数によってオブジェクトが同一または同様のカテゴリに属するクラスターに分割されるときに、分析が最も成功するという事実を表します。

例:ハードウェアの特性

多重コレスポンデンス分析がどのように機能するかを調査するには、Hartigan (Hartigan, 1975)からデータを取得して使用します。このデータは、screws.sav ファイルの中にあります。詳細は、A 付録 p. 321 [サンプルファイル](#) を参照してください。このデータセットには、ねじ、ボルト、ナット、鋷（びょう）の特性に関する情報が含まれています。次の表は、Hartigan の金具データセットにおける変数、変数ラベル、および各変数カテゴリに割り当てられている値ラベルを示しています。

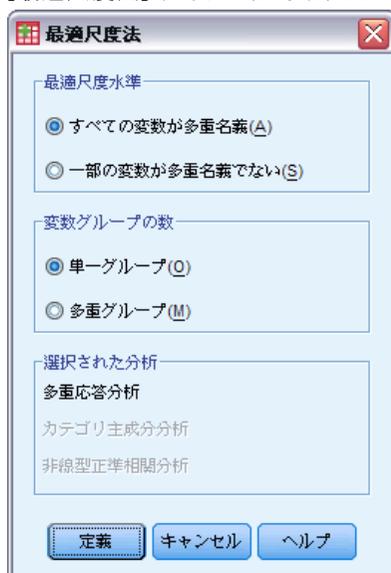
テーブル 13-1
Hartigan の金具データセット

変数名	変数ラベル(L)	値ラベル
ねじ山	ねじ山の有無	ねじ山あり、ねじ山なし
頭形状	頭部の形状	平ら、カップ、円錐、円、円筒
頭刻み	頭部の刻み目	なし、星型、スリット
底形状	底部の形状	尖っている、平ら
長さ	半インチ単位の長さ	0.5 インチ、1 インチ、1.5 インチ、2 インチ、2.5 インチ
真鍮	真鍮	真鍮でできている、真鍮でできていない
対象物	オブジェクト	釘 1、釘 2、釘 3、釘 4、釘 5、釘 6、釘 7、釘 8、釘 b、ねじ 1、ねじ 2、ねじ 3、ねじ 4、ねじ 5、ねじ b、ボルト 1、ボルト 2、ボルト 3、ボルト 4、ボルト 5、ボルト 6、鋷 1、鋷 2、鋷 3

分析の実行

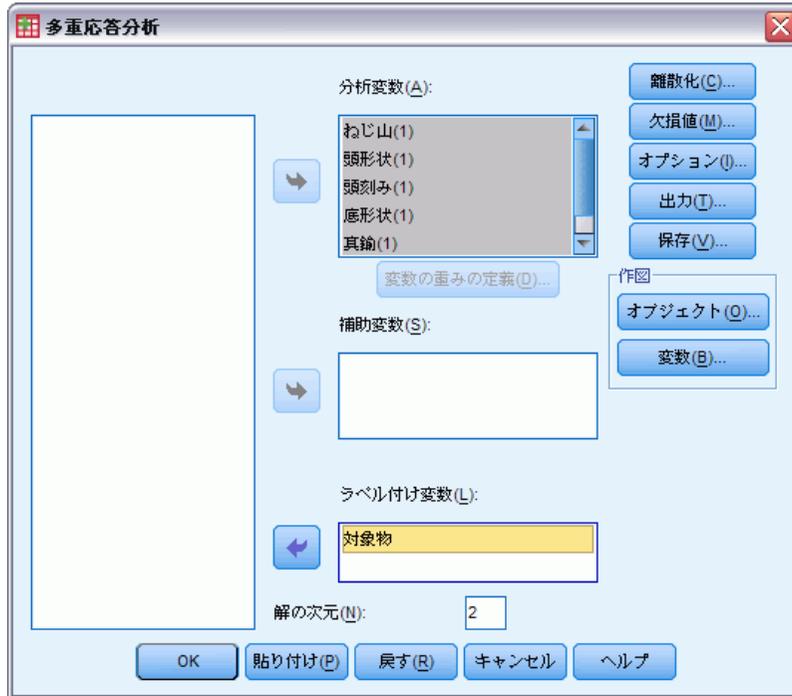
- ▶ 多重応答分析を実行するには、メニューから次の項目を選択します。
分析(A) > 次元分解 > 最適尺度法...

図 13-1
[最適尺度法] ダイアログ ボックス



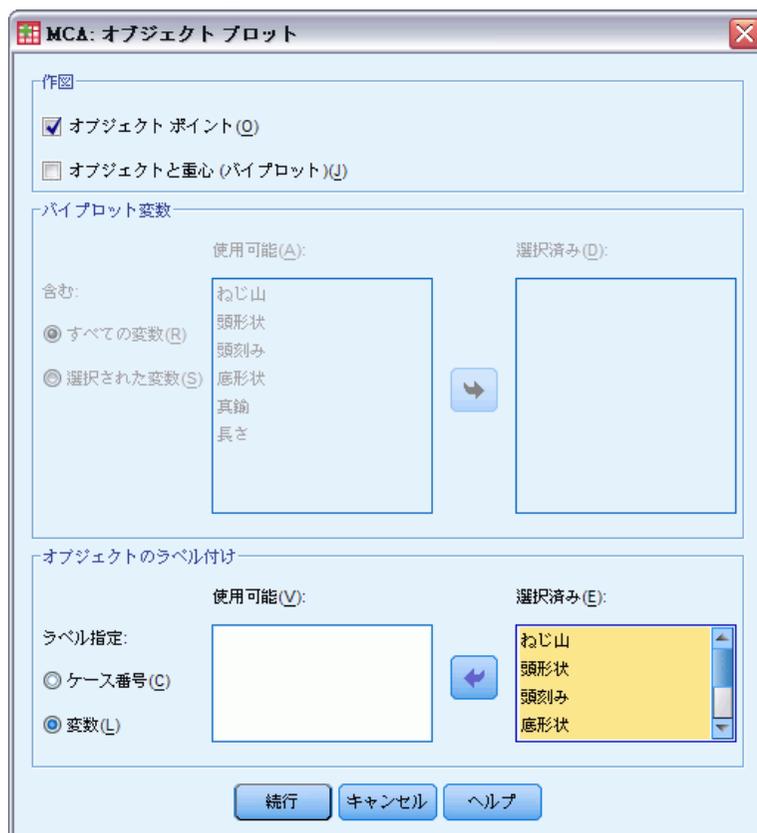
- ▶ [全ての変数が多重名義] と [単一グループ] が選択されていることを確認し、[定義] をクリックします。

図 13-2
[多重応答分析] ダイアログ ボックス



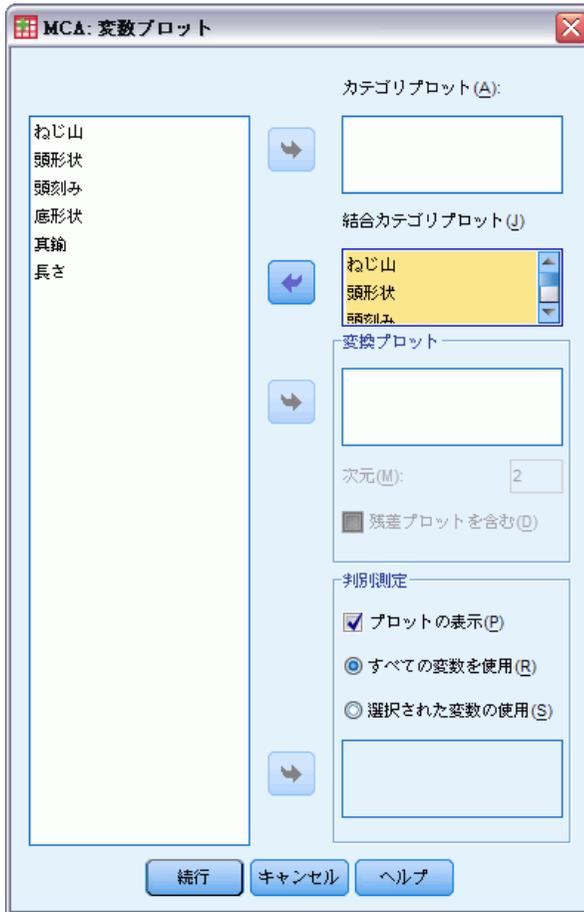
- ▶ 分析変数として、「ねじ山の有無」から「半インチ単位の長さ」までを選択します。
- ▶ ラベル付け変数として「対象物」を選択します。
- ▶ [作図] グループの [オブジェクト] をクリックします。

図 13-3
[オブジェクト プロット] ダイアログ ボックス



- ▶ [オブジェクトのラベル付け] で [変数] を選択します。
- ▶ ラベル付け変数として「ねじ山」から「対象物」までを選択します。
- ▶ [続行] をクリックして、[多重応答分析] ダイアログ ボックスの [作図] グループで [変数] をクリックします。

図 13-4
[変数プロット] ダイアログ ボックス



- ▶ [結合カテゴリ プロット] の変数として「ねじ山」から「長さ」までを選択します。
- ▶ [続行] をクリックします。
- ▶ [多重応答分析] ダイアログ ボックスで、[OK] をクリックします。

モデルの要約

等質性分析では、複数の次元の解を計算できます。次元の最大数は、カテゴリ数から欠損値を含まない変数の数を引いた数、または観測の数から1を引いた数のどちらか小さい方と同じになります。ただし、最大数の次元は原則として使用しないでください。次元数は少ないほど解釈しやすく、一定の次元の後には、説明される新たな連関の量が無視できるほど

少なくなります。等質性分析では1次元、2次元、または3次元の解が非常に一般的です。

図 13-5
モデルの要約(M)

次元	Cronbachの アルファ	説明された分散		
		合計 (固有値)	イナーシャ	分散の %
1	.878	3.727	.621	62.123
2	.657	2.209	.368	36.809
総計		5.936	.989	
平均値	.796 ^a	2.968	.495	49.466

a. Cronbachのアルファ平均値は、固有値平均値に基づいています。

データのほとんどすべての分散は解で説明されます。62.1%は第1次元で説明され、36.8%は第2次元で説明されます。

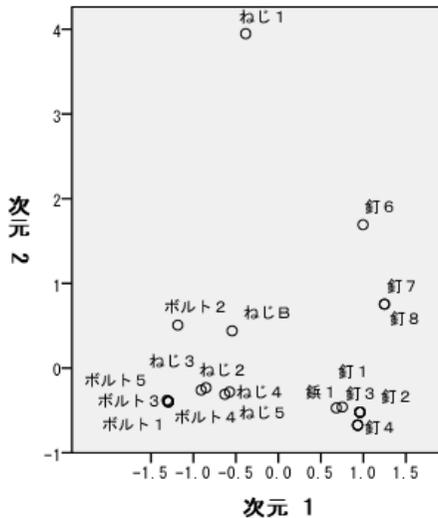
どちらの次元からも、距離に関する解釈が得られます。変数が適切に判別すると、オブジェクトは属するカテゴリの近くに配置されます。理想的には、同一カテゴリ内のオブジェクトは互いに近くに配置され（つまり、オブジェクト同士が類似したスコアを持ち）、異なる変数のカテゴリは、変数同士が同一オブジェクトに属している場合は近くに配置されます（つまり、1つの変数に対して類似したスコアを持っている2つのオブジェクトは、解の他の変数に対しても互いに近いスコアを得る必要があります）。

オブジェクト スコア

モデルの要約の調査後、オブジェクト スコアを確認する必要があります。1つ以上の変数を指定して、オブジェクト スコア プロットのラベル付けを行うことができます。各ラベル付け変数により、その変数の値でラベル付けされた個別のプロットが作成されます。変数オブジェクトでラベル付けされたオブジェクト スコアのプロットについて確認していきます。これは単なるケース識別用変数なので、計算では使用されません。

オブジェクトから原点までの距離は、「平均」応答パターンとの差を反映します。この平均応答パターンは、各変数の最頻カテゴリに対応しています。最頻カテゴリに対応している多くの特性を持つオブジェクトは、原点の近くに配置されます。一方、独自の特性を持つオブジェクトは、原点から離れた位置に配置されます。

図 13-6
オブジェクトでラベル付けされたオブジェクト スコア プロット



プロットを調べると、第1次元（横軸）によって、（ねじ山を持つ）ねじおよびボルトと、（ねじ山を持たない）釘および鋸とが識別されることがわかります。このことは、ねじとボルトが横軸の一端にあり、鋸と釘が他端にあるので、プロット上で容易に理解できます。また、あまり大きな特徴ではありませんが、第1次元では、（底が平らな）ボルトと（底が尖っている）その他のオブジェクトが離れています。

第2次元（縦軸）では、[ねじ1]と[釘6]が他のオブジェクトから離れています。[ねじ1]と[釘6]の共通点は、変数の長さに関する値です。これら2つのオブジェクトは、データの中で最長のオブジェクトです。さらに、[ねじ1]は、他のオブジェクトよりも原点からはるかに離れたところに位置します。このことは、全体として見ると、このオブジェクトの多くの特性が他のオブジェクトと共有されていないことを示しています。

オブジェクトスコアプロットは、外れ値を見つける際に特に役立ちます。[ねじ1]は、外れ値と見なすことができます。後ほど、このオブジェクトを対象から外すとどのような変化が生じるか検討します。

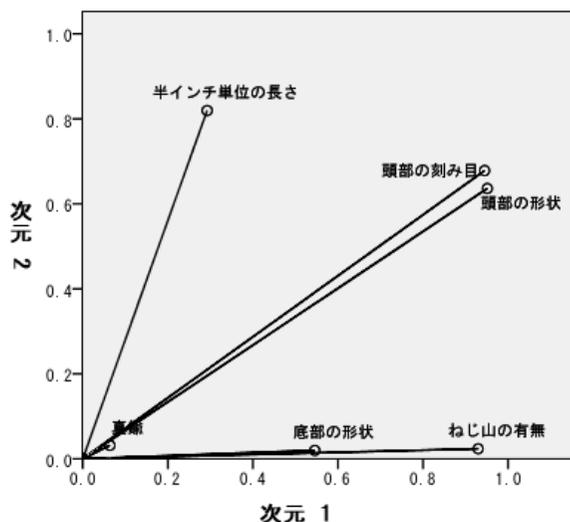
判別測定

残りのオブジェクトスコアプロットを調べる前に、判別測定がこれまで述べてきた内容と合致するかどうかを確認しましょう。判別測定は、各変数で平方成分負荷と見なすことができ、次元ごとに計算されます。この測定は、その次元における数量化された変数の分散でもあります。この測定の最大値は1です。オブジェクトスコアが相互に排他的なグループに分かれ、カテゴリ内のすべてのオブジェクトスコアが同一の場合は、1に達します。（注：欠損値がある場合は、この測定の値は1より大きくなります。）大き

な判別測定は、変数のカテゴリ間での大きな広がりに対応するので、対象の次元に沿った変数のカテゴリ間で判別の度合いが高いことを示します。

すべての次元の判別測定の平均は、その次元で説明される分散のパーセントに相当します。したがって、次元は平均判別に基づいて順序付けられます。解のすべての次元で、第1次元には最大の平均判別があり、第2次元には2番目に大きな平均判別があります。第3次元以降も同様になります。

図 13-7
判別測定のプロット



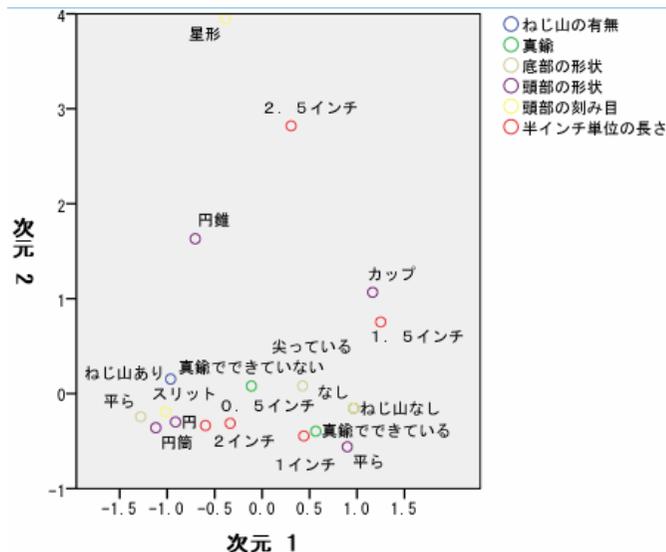
オブジェクトスコアプロットで説明したように、判別測定プロットは、第1次元が変数「ねじ山の有無」と「底部の形状」に関連していることを示します。これらの変数の判別測定は、第1次元では大きく、第2次元では小さくなります。したがって、これらのどちらの変数でも、カテゴリは第1次元だけに沿って広がります。「半インチ単位の長さ」の値は、第2次元では大きく、第1次元では小さくなります。その結果、「長さ」は第2次元に最も近くなっており、第2次元が長いオブジェクトと他のオブジェクトとを切り離しているというオブジェクトスコアプロットの観測と一致します。「頭部の刻み目」と「頭部の形状」は、両方の次元で比較的大きな値を持っています。このことは、第1次元と第2次元の両方で判別が行われたことを示しています。原点に非常に近いところに位置する変数「真鍮」では、第1次元でも第2次元でもまったく判別が行われません。すべての部品は、真鍮で作ることも、真鍮以外の材料で作ることもできるので、これは当然のことです。

カテゴリ数量化

判別測定は、特定の次元に沿った数量化された変数の分散であることを思い出してください。判別測定プロットには、どの変数で、どの次元に沿って判別が行われるかを示すこれらの分散が含まれています。ただし、同じ分散は、適度に離れて広がるカテゴリすべて、または近くに配置されるカテゴリの多くに対応し、それにはこのグループとは異なる一部のカテゴリが含まれます。判別プロットでは、これらの2つの条件を区別できません。

カテゴリ数量化プロットには、カテゴリ間の関係を識別できる変数の判別を示す他の方法が用意されています。このプロットでは、各次元上の各カテゴリの座標が示されます。したがって、変数ごとに類似しているカテゴリを特定できます。

図 13-8
カテゴリ数量化(Q)



「半インチ単位の長さ」には5つのカテゴリがあり、その中の3つのカテゴリはプロットの最上部付近にグループ化されます。残りの2つのカテゴリはプロットの下半分に位置し、[2.5インチ]カテゴリはそのグループから非常に離れたところに位置します。第2次元に沿って[長さ]の判別測定が大きいのは、[2.5インチ]カテゴリが他の[長さ]のカテゴリと大きく異なっているためです。同様に、「頭部の形状」の場合、カテゴリ[星形]が他のカテゴリから非常に離れたところに位置しているため、第2次元に沿って判別測定が大きくなります。このようなパターンは、判別測定のプロットで図示できません。

変数のカテゴリ数量化の広がりや分散を反映するので、その変数が各次元でどの程度適切に判別されたかを示します。第1次元に注目すると、「ねじ山の有無」のカテゴリは離れたところに位置します。しかし、この

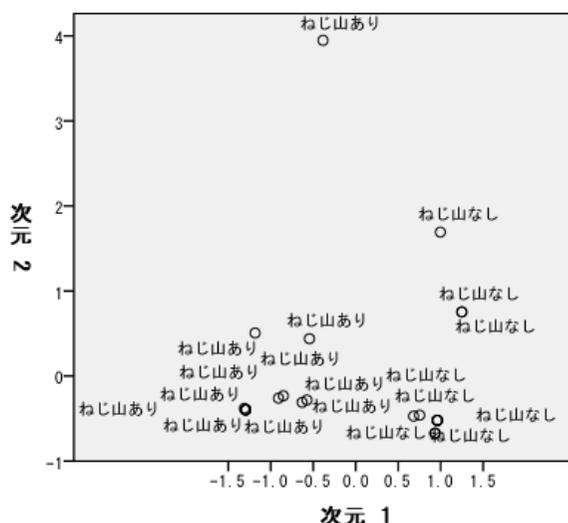
変数のカテゴリは、第2次元に沿って見てみると非常に近くに位置します。したがって、「ねじ山の有無」は、第2次元よりも第1次元でより適切に判別します。一方、「頭部の形状」のカテゴリは、両方の次元に沿って広がります。このことは、この変数が両方の次元で適切に判別することを示します。

カテゴリ数量化プロットでは、変数が判別を行う次元や方法を決定するだけでなく、変数の判別も比較されます。離れたところに位置するカテゴリを持つ変数は、近くに位置するカテゴリを持つ変数よりも適切に判別します。たとえば、第1次元に沿って見てみると、「真鍮」の2つのカテゴリは、「ねじ山の有無」の2つのカテゴリよりも互いに近くに位置します。このことは、「ねじ山の有無」は、第1次元に沿って「真鍮」よりも適切に判別することを示しています。ただし、第2次元に沿って見てみると、距離は非常に似ています。このことは、2つの変数が第2次元に沿って同程度に判別することを示しています。先に説明した判別測定プロットでは、カテゴリの広がりを反映する分散を使用して、これらの同一の関係が識別されます。

オブジェクト スコアの詳細な調査

各変数でラベル付けされたオブジェクト スコア プロットを調査すると、データをより詳しく理解できます。理想的な状況は、類似したオブジェクト がはっきりと区別できるグループを形成しており、そのグループが互いに離れている状況です。

図 13-9
[ねじ山の有無] でラベル付けされたオブジェクト スコア



「ねじ山の有無」でラベル付けされたプロットは、第1次元で[ねじ山あり]と[ねじ山なし]が完全に分離していることを示します。ねじ山を持つすべてのオブジェクトには負のオブジェクトスコアがあり、ねじ山を持たないすべてのオブジェクトには正のオブジェクトスコアがあります。これら2つのカテゴリではコンパクトなグループは形成されませんが、通常、カテゴリ間の完全な差は適切な結果と見なされます。

図 13-10
[頭部の形状]でラベル付けされたオブジェクトスコア

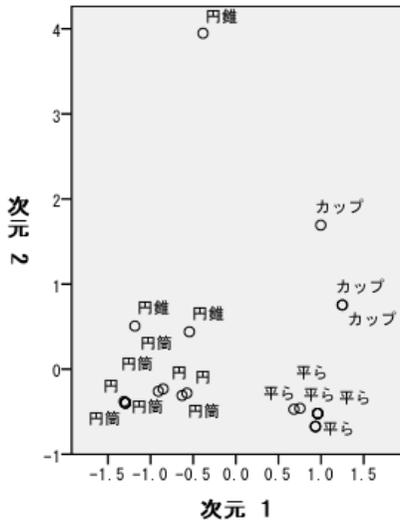
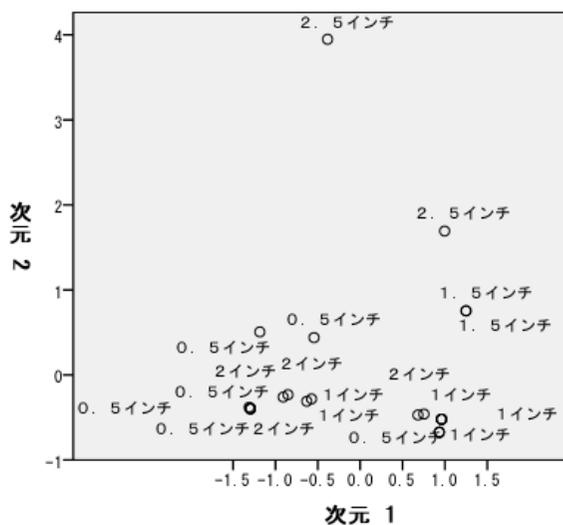
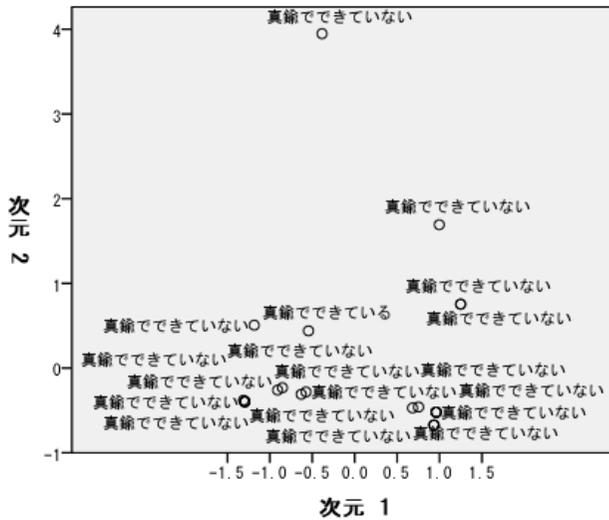


図 13-11
[半インチ単位の長さ] でラベル付けされたオブジェクト スコア



「半インチ単位の長さ」でラベル付けされたプロットは、この変数が第1次元で判別しないことを示します。この変数のカテゴリでは、水平線に射影されたときにグループ化が示されません。ただし、「半インチ単位の長さ」は、第2次元で判別します。短いオブジェクトは正のスコアに対応し、長いオブジェクトは大きな負のスコアに対応します。

図 13-12
[真鍮] でラベル付けされたオブジェクト スコア



「真鍮」でラベル付けされたプロットは、この変数の各カテゴリが、第 1 次元でも第 2 次元でもあまり分離できないことを示します。このオブジェクト スコアは空間全体にわたって広範に広がります。[真鍮] オブジェクトは、[真鍮] ではないオブジェクトと区別できません。

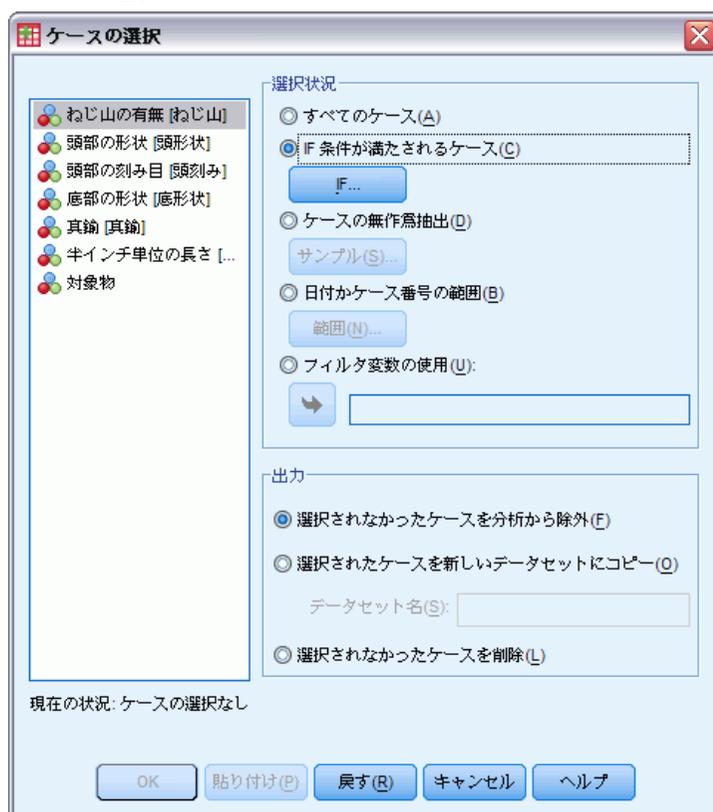
外れ値の除外

等質性分析でいう外れ値とは、非常に多くの独自の特性をもつオブジェクトです。先に説明したように、[ねじ 1] は、外れ値と見なすことができます。

このオブジェクトを削除して分析を再実行するには、メニューから次の項目を選択します。

データ > ケースの選択(S)...

図 13-13
[ケースの選択] ダイアログ ボックス



- ▶ [IF 条件が満たされるケース] を選択します。
- ▶ [IF] をクリックします。

図 13-14
[IF 条件の定義] ダイアログ ボックス



- ▶ 条件として「対象物 ~= 16」と入力します。
- ▶ [続行] をクリックします。
- ▶ [ケースの選択] ダイアログ ボックスで、[OK] をクリックします。
- ▶ そして、[多重応答分析] ダイアログ ボックスを再び開き、[OK] をクリックします。

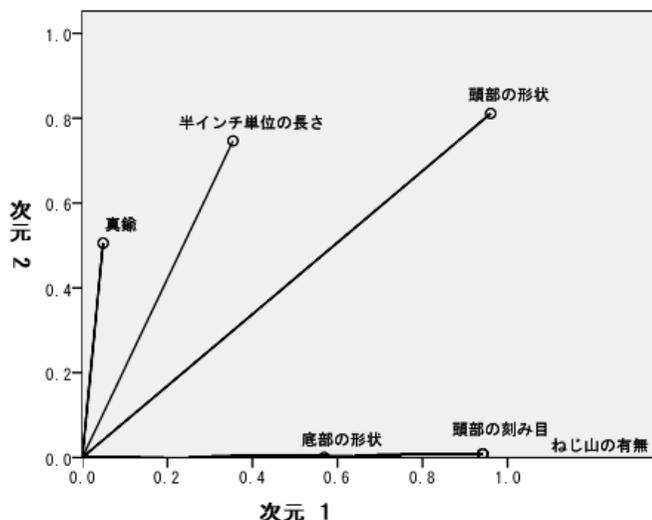
図 13-15
モデルの要約 (外れ値は除外)

次元	Cronbachの アルファ	説明された分散		
		合計 (固有値)	イナーシャ	分散の %
1	.885	3.815	.636	63.591
2	.623	2.081	.347	34.676
総計		5.896	.983	
平均値	.793 ^a	2.948	.491	49.133

a. Cronbach のアルファ平均値は、固有値平均値に基づいています。

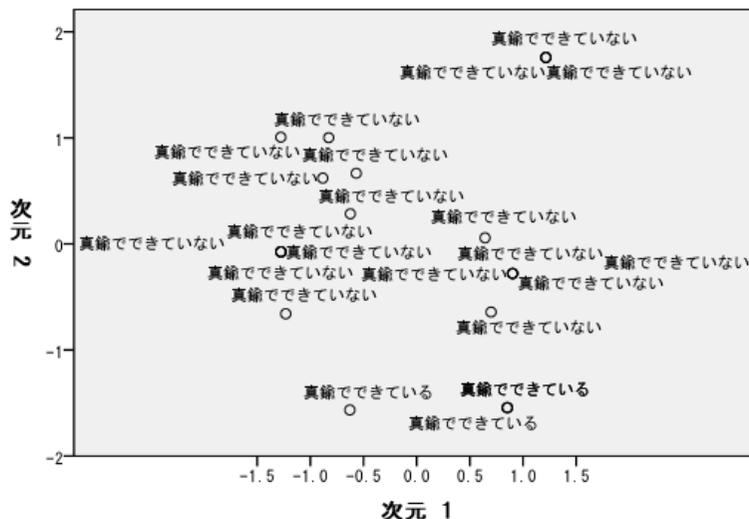
固有値がわずかに変動します。第 1 次元で、以前よりも分散が説明されるようになりました。

図 13-16
判別測定(S)



判別プロットで示されるように、「頭部の刻み目」は第2次元で判別しなくなりましたが、どちらの次元でも判別しなかった「真鍮」は、第2次元で判別するようになりました。他の変数の判別は、それほど変化していません。

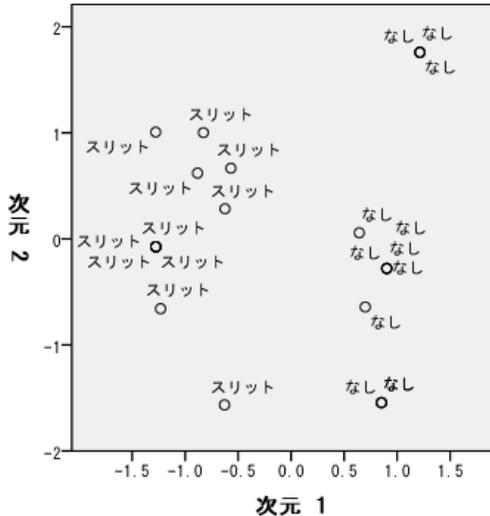
図 13-17
[真鍮]でラベル付けされたオブジェクトスコア(外れ値は除外)



「真鍮」でラベル付けされたオブジェクトスコアプロットでは、4つの[真鍮]オブジェクトがすべて、プロットの最下部付近に表示されています(3つのオブジェクトが同じ場所に位置する)。このことは、第2次元に沿っ

て判別の度合いが高いことを示しています。前回の分析での「ねじ山の有無」の場合と同様に、オブジェクトはコンパクトなグループを形成していませんが、カテゴリごとのオブジェクトの差異は完全に判別できます。

図 13-18
[頭部の刻み目] でラベル付けされたオブジェクト スコア (外れ値は除外)



「頭部の刻み目」でラベル付けされたオブジェクト スコア プロットは、前回の分析と同様に、第 1 次元で、刻み目のないオブジェクトと刻み目のあるオブジェクトとが完全に判別されることを示しています。ただし、前回の分析とは異なり、第 2 次元では 2 つのカテゴリを判別できなくなりました。

このように、星型の頭部を持つ唯一のオブジェクトである [ねじ 1] を除外すると、第 2 次元の解釈に大きな影響を及ぼします。この次元では、「真鍮」、「頭部の形状」、および「半インチ単位の長さ」に基づいてオブジェクトが判別されるようになりました。

推奨参考文献

多重応答分析の詳細は、次のテキストを参照してください。

Benzécri, J. P. 1992. Correspondence analysis handbook. New York: Marcel Dekker.

Guttman, L. 1941. The quantification of a class of attributes: A theory and method of scale construction. In: The Prediction of Personal Adjustment, P. Horst, ed. New York: Social Science Research Council.

Meulman, J. J. 1982. Homogeneity analysis of incomplete data. Leiden: DSWO Press.

Meulman, J. J. 1996. Fitting a distance model to homogeneous subsets of variables: Points of view analysis of categorical data. *Journal of Classification*, 13, .

Meulman, J. J., および W. J. Heiser. 1997. Graphical display of interaction in multiway contingency tables by use of homogeneity analysis. In: *Visual Display of Categorical Data*, M. Greenacre, および J. Blasius、編集者. New York: Academic Press.

Nishisato, S. 1984. Forced classification: A simple application of a quantification method. *Psychometrika*, 49, .

Tenenhaus, M., および F. W. Young. 1985. An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis, and other methods for quantifying categorical multivariate data. *Psychometrika*, 50, .

Van Rijckevorsel, J. 1987. The application of fuzzy coding and horseshoes in multiple correspondence analysis. Leiden: DSWO Press.

多次元尺度法

多次元尺度法の目的は、オブジェクトの集合を使用して、低次元空間でオブジェクトを表現することです。多次元尺度法の解は、オブジェクト間の近接を使用して見つけ出します。多次元尺度法の手続きでは、元のオブジェクト（変換された可能性があります）の近接と、低次元空間でのユークリッド距離の間の平方偏差が最小化されます。

低次元空間を使用するのは、オブジェクトの関係を明らかにするためです。解を独立変数の線型結合に制限することで、各変数について、解の次元を解釈できる場合があります。次の例では、15個の異なる親族関係用語を3つの次元で表現する方法について説明します。また、各用語の性別、世代、および親等について、空間を解釈する方法について説明します。

例：親族関係用語の調査

Rosenberg と Kim (Rosenberg および Kim, 1975) は、15種類の親族関係用語（祖父、祖母、父、母、叔父、叔母、兄弟、姉妹、いとこ、息子、娘、甥、姪、孫息子、孫娘）の分析を行いました。Rosenberg と Kim は、大学生の4つのグループ（女性2組、男性2組）に、類似性に基づいて上記の用語を並べ替えるよう依頼しました。2つのグループ（女性1組、男性1組）には、1回目と違う条件に基づいて、2回目の並べ替えをするように頼みました。このようにして、次のテーブルに示すとおり、合計で6つの「ソース」を取得できました。

テーブル 14-1
親戚関係データのソース構造

ソース	性別	条件	サンプル サイズ
1	女性	1 回限りの並べ替え	85
2	男性	1 回限りの並べ替え	85
3	女性	1 回目の並べ替え	80

ソース	性別	条件	サンプル サイズ
4	女性	2 回目の 並べ替え	80
5	男性	1 回目の 並べ替え	80
6	男性	2 回目の 並べ替え	80

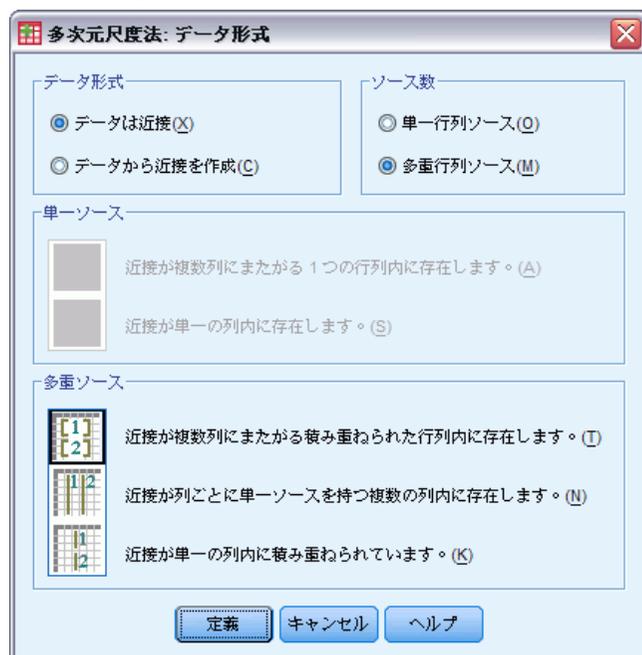
各ソースは、 15×15 の近接行列に対応します。この近接行列のセルの数は、ソースの人数から、ソース内でオブジェクトを分割した回数を引いたものです。このデータセットは kinship_dat.sav にあります。詳細は、A 付録 p. 321 [サンプルファイル](#) を参照してください。

次元数の選択

解に含める次元数を決定します。次元数を決定する場合は、スクリープロットを利用できます。

- ▶ スクリープロットを作成するには、メニューから次の項目を選択します。
分析 > 尺度 > 多次元尺度法 (PROXSCAL)...

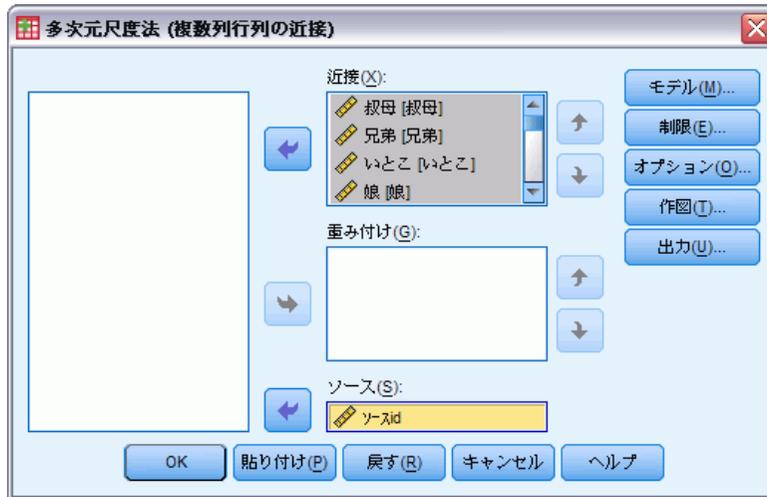
図 14-1
[データ形式] ダイアログ ボックス



- ▶ [ソース数] グループの [多重行列ソース] を選択します。

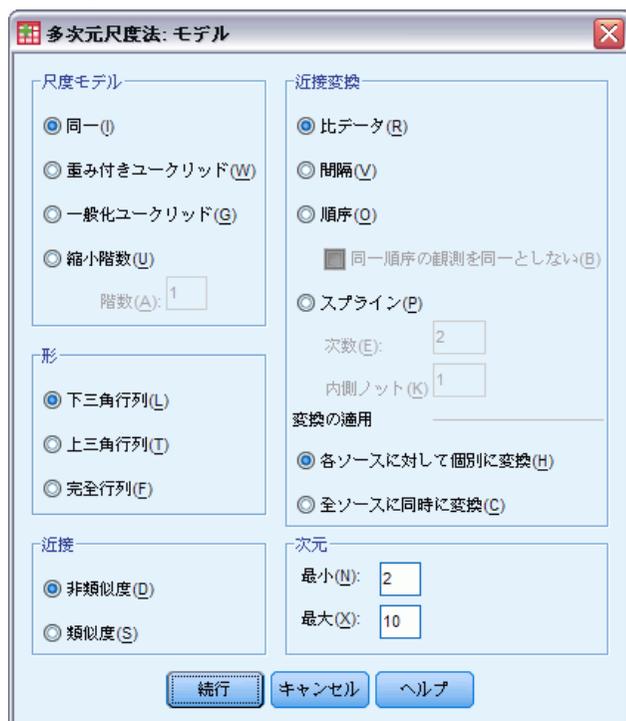
- ▶ [定義] をクリックします。

図 14-2
[多次元尺度法] ダイアログ ボックス



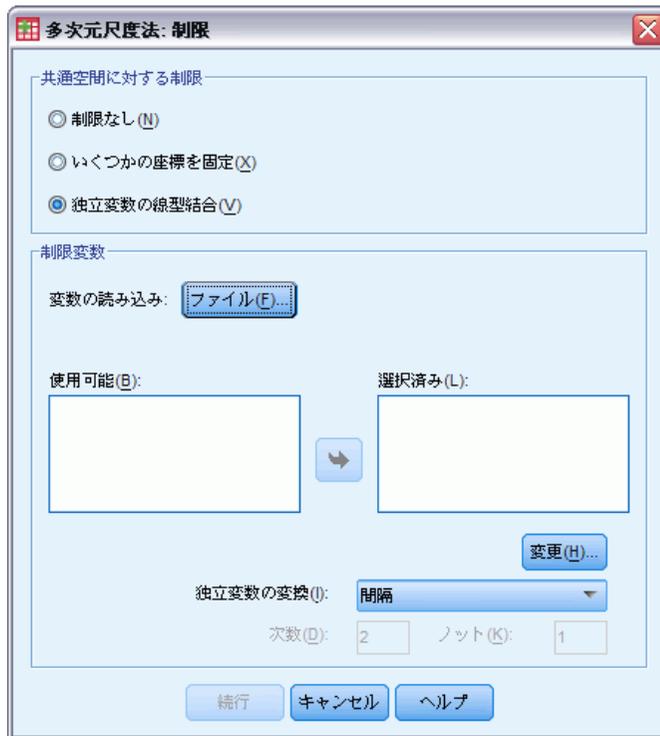
- ▶ 近接変数として、「叔母」から「叔父」までを選択します。
- ▶ ソース変数として、「ソース id」を選択します。
- ▶ [モデル] をクリックします。

図 14-3
[モデル] ダイアログ ボックス



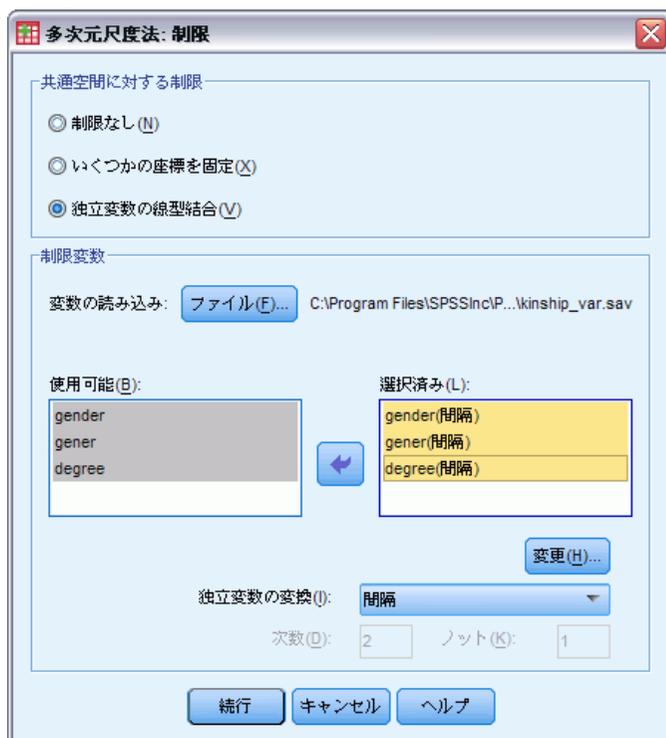
- ▶ 最大次元数として「10」と入力します。
- ▶ [続行] をクリックします。
- ▶ [多次元尺度法] ダイアログ ボックスで [制限] をクリックします。

図 14-4
[制限] ダイアログ ボックス



- ▶ [独立変数の線型結合] を選択します。
- ▶ [ファイル] をクリックして、独立変数の読み込み元を選択します。
- ▶ kinship_var.sav を選択します。

図 14-5
[制限] ダイアログ ボックス



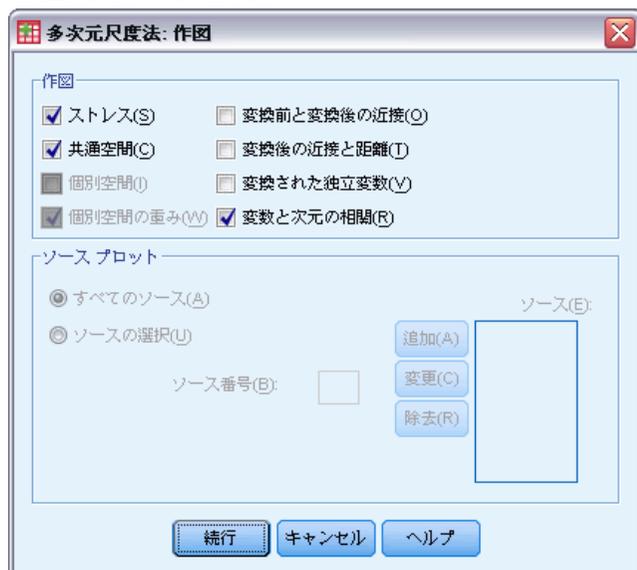
- ▶ 制限変数として、「性別」、「世代」、および「親等」を選択します。
変数「性別」にはユーザー欠損値である 9 = 欠損値 (いどこ) があることに注意します。手続きでは、この欠損値が有効なカテゴリとして処理されます。したがって、デフォルトの線型変換が適切であるとは考えられません。代わりに、名義変換を使用します。

図 14-6
[制限] ダイアログ ボックス



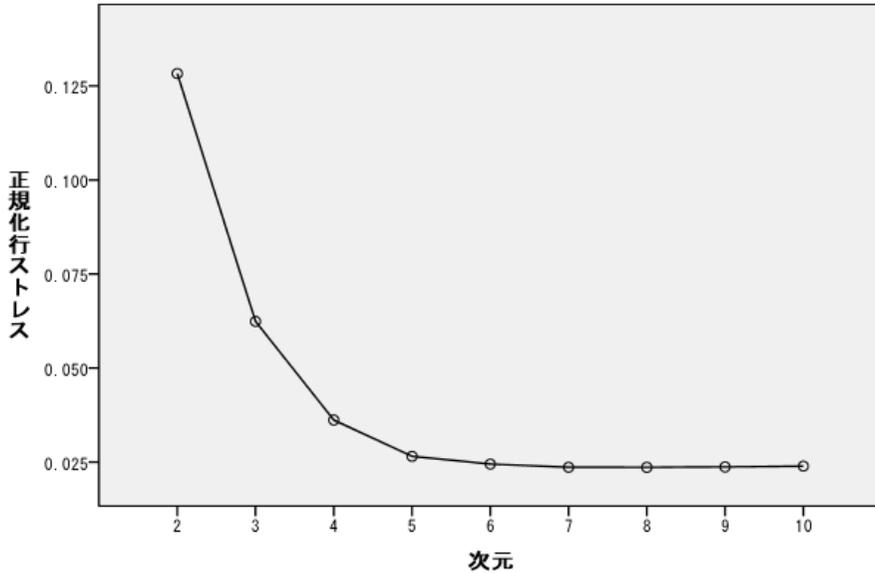
- ▶ 「性別」を選択します。
- ▶ [独立変数の変換] ドロップダウン リストから、[名義]を選択します。
- ▶ [変更] をクリックします。
- ▶ [続行] をクリックします。
- ▶ [多次元尺度法] ダイアログ ボックスで [作図] をクリックします。

図 14-7
[作図] ダイアログ ボックス



- ▶ [作図] グループの [ストレス] をクリックします。
- ▶ [続行] をクリックします。
- ▶ [多次元尺度法] ダイアログ ボックスで [OK] をクリックします。

図 14-8
スクリー プロット



手続きは 10 次元の解から開始して、2 次元の解まで実行されます。スクリープロットは、各次元における、正規化された解の原ストレスを示します。次元数を 2 から 3、3 から 4 に増やすことで、ストレスが大幅に改善されることが、プロットからわかります。4 次元より後は、改善の度合いが小さくなります。結果が解釈しやすくなるので、3 次元の解を使用してデータを分析します。

3 次元の解

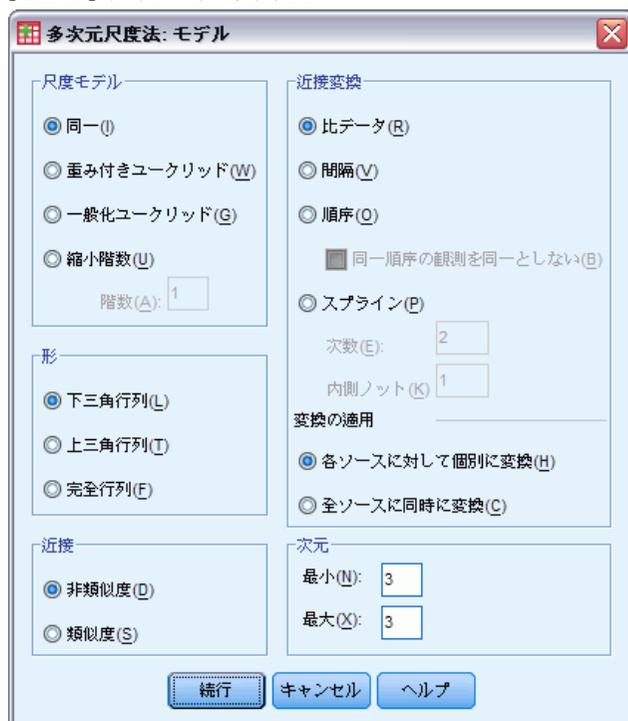
独立変数の「性別」、「世代」、および「親等」は、解の次元の解釈に使用するために作成されました。独立変数は、次のように作成されました。

性別	1 = 男性、2 = 女性、9 = 欠損値（いとこ）
世代	自分を中心にして親族を示す世代の数値。数値が小さくなるほど、古い世代に対応します。祖父母は -2、孫は 2、兄弟/姉妹は 0 になります。
親等	家系図に従って表す親等の数値。両親は 1 ノード上で、子供は 1 ノード下になります。兄弟/姉妹の場合、親まで 1 ノード上がり、1 ノード下がると兄弟/姉妹になるので、2 親等になります。いとこの場合、祖父母まで 2 ノード上がり、叔父や叔母を経由して 2 ノード下がるといとこになるので、4 親等になります。

外部変数は kinship_var.sav にあります。さらに、上の分析の初期布置は kinship_ini.sav に用意されています。詳細は、A 付録 p.321 サンプルファイル を参照してください。

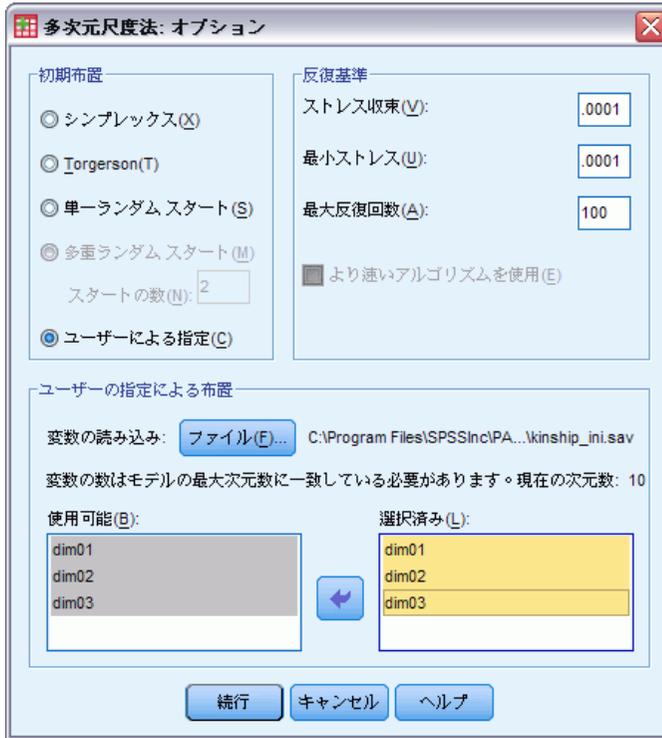
分析の実行

図 14-9
[モデル] ダイアログ ボックス



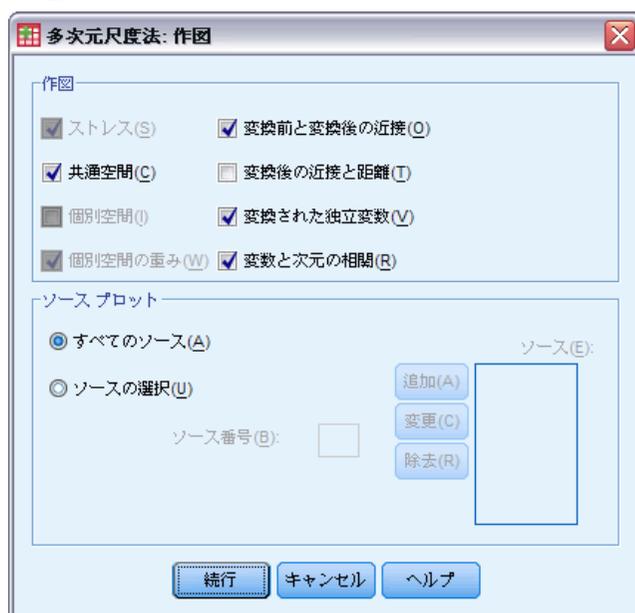
- ▶ 3次元の解を取得するには、[多次元尺度法] ダイアログ ボックスを再表示し、[モデル] をクリックします。
- ▶ [次元] の [最小値] および [最大値] に、「3」と入力します。
- ▶ [続行] をクリックします。
- ▶ [多次元尺度法] ダイアログ ボックスで [オプション] をクリックします。

図 14-10
[オプション] ダイアログ ボックス



- ▶ [初期布置] として [ユーザー指定による] を選択します。
- ▶ [変数の読み込み] で [ファイル] をクリックし、kinship_ini.sav を選択します。
- ▶ 変数として、「次元 1」、「次元 2」、および「次元 3」を選択します。
- ▶ [続行] をクリックします。
- ▶ [多次元尺度法] ダイアログ ボックスで [作図] をクリックします。

図 14-11
[作図] ダイアログ ボックス



- ▶ [変換前と変換後の近接] と [変換された独立変数] を選択します (チェックを付けます)。
- ▶ [続行] をクリックします。
- ▶ [多次元尺度法] ダイアログ ボックスで [出力] をクリックします。

図 14-12
[出力] ダイアログ ボックス



- ▶ [入力データ]、[ストレス分解]、および [変数と次元の相関] を選択します (チェックを付けます)。
- ▶ [続行] をクリックします。
- ▶ [多次元尺度法] ダイアログ ボックスで [OK] をクリックします。

ストレス測定

ストレスと適合の測度により、解の距離が元の距離にどの程度近接しているかがわかります。

図 14-13
ストレスと適合の測度

正規化された原ストレス	.06234
ストレス-I	.24968 ^a
ストレス-II	.87849 ^a
S-ストレス	.14716 ^b
説明された散らばり (D.A.F.)	.93766
Tuckerの適合係数	.96833

PROXSCALは正規化された原ストレスを最小化します。

a. 最適尺度因子 = 1.066。

b. 最適尺度因子 = .984。

4つのストレス統計量により、データの不適合度が測定され、説明された散らばり、および Tucker の適合係数により、適合度が測定されます。ストレス測度が小さく（最低 0）、適合度が大きい（最高 1）ほど、より適切な解であることを示します。

図 14-14
正規化された原ストレスの分解

	ソース						平均値
	SRC 1	SRC 2	SRC 3	SRC 4	SRC 5	SRC 6	
オブジェクト							
叔母	.0991	.0754	.0629	.0468	.0391	.0489	.0620
兄弟	.1351	.0974	.0496	.0813	.0613	.0597	.0807
いとこ	.0325	.0336	.0480	.0290	.0327	.0463	.0370
娘	.0700	.0370	.0516	.0229	.0326	.0207	.0391
父	.0751	.0482	.0521	.0225	.0272	.0298	.0425
孫娘	.1410	.0736	.0801	.0707	.0790	.0366	.0802
祖父	.1549	.1057	.0858	.0821	.0851	.0576	.0952
祖母	.1550	.0979	.0858	.0844	.0816	.0627	.0946
孫息子	.1374	.0772	.0793	.0719	.0791	.0382	.0805
母	.0813	.0482	.0526	.0229	.0260	.0227	.0423
甥	.0843	.0619	.0580	.0375	.0317	.0273	.0501
姪	.0850	.0577	.0503	.0353	.0337	.0260	.0480
姉妹	.1361	.0946	.0496	.0816	.0629	.0588	.0806
息子	.0689	.0373	.0456	.0242	.0337	.0253	.0392
叔父	.0977	.0761	.0678	.0489	.0383	.0498	.0631
平均値	.1035	.0681	.0613	.0508	.0496	.0407	.0623

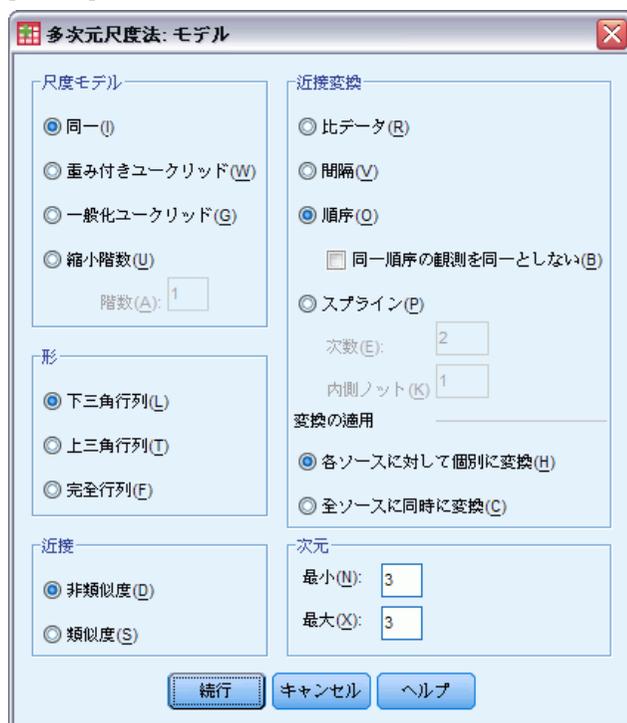
ストレスの分解により、解の全体的ストレスに対し最も影響の大きいソースとオブジェクトを識別できます。今回の分析では、ソースのストレスのほとんどが、ソース 1 と 2 に起因します。オブジェクトでは、ほとんどのストレスが「兄弟」、「孫娘」、「祖父」、「祖母」、「孫息子」、および「姉妹」に起因します。

ほとんどのストレスを説明する 2 つのソースは、「1 回限りの並べ替え」を行った 2 つのグループです。この情報は、用語を並べ替える際に、学生が複数の因子を考慮したことを示します。そして、2 回並べ替えることを許可された学生は、最初の並べ替えではそういった因子の一部分に集中し、2 回目の並べ替えで残りの因子を考慮した、と言えます。

デフォルト以外の変換による 3 次元の解

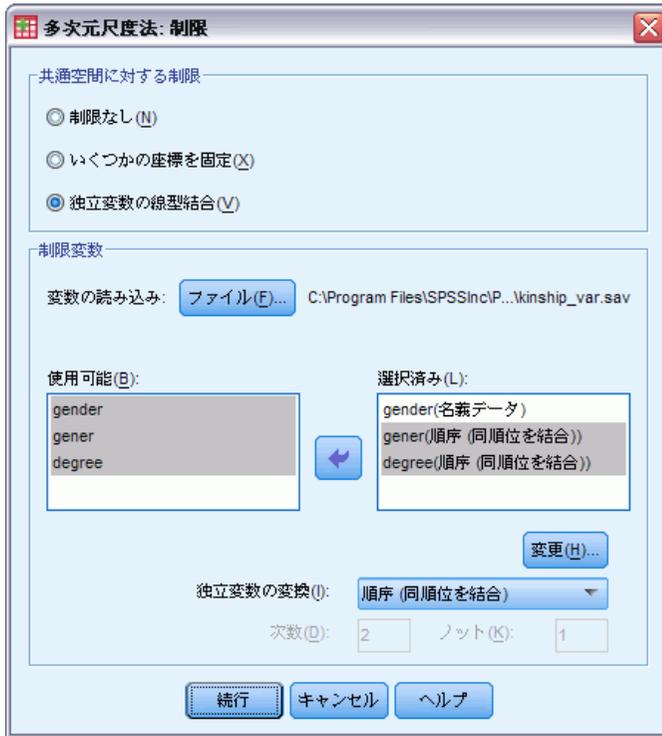
最初の分析の解は、近接のデフォルトの比率変換、および独立変数の「世代」と「親等」の間隔変換を使用して計算しました。結果はかなり適切なものですが、他の変換を使用することで、さらに適切な結果が得られる可能性があります。たとえば、近接「世代」と「親等」はいずれも自然順をもちますが、線型変換ではなく順序変換を使用することで、モデルの精度が向上する場合があります。

図 14-16
[モデル] ダイアログ ボックス



- ▶ 順序レベル（同順位を結合）で近接「世代」と「親等」を尺度化して、分析を再実行するには、[多次元尺度法] ダイアログ ボックスを再表示し、[モデル] をクリックします。
- ▶ 近接変換として [順序] を選択します。
- ▶ [続行] をクリックします。
- ▶ [多次元尺度法] ダイアログ ボックスで [制限] をクリックします。

図 14-17
[制限] ダイアログ ボックス



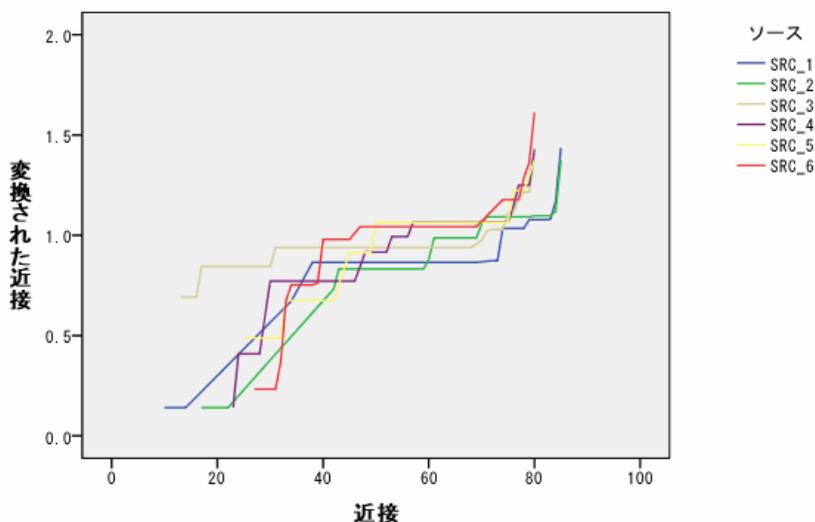
- ▶ 「世代」と「親等」を選択します。
- ▶ [独立変数の変換] ドロップダウン リストから、[順序 同順位を結合] を選択します。
- ▶ [変更] をクリックします。
- ▶ [続行] をクリックします。
- ▶ [多次元尺度法] ダイアログ ボックスで [OK] をクリックします。

変換プロット

変換プロットは、元の変換が適切であったかどうかを最初にチェックするのに適しています。変換プロットがほとんど線型に近いのであれば、線型仮定は適切です。そうでない場合は、ストレス測度をチェックして、適合度の改善ができるかどうか確認し、共通空間プロットをチェックして、解釈がより有益になっているかどうかを確認します。

それぞれの独立変数がほぼ線型の変換になっている場合、各変数を数値型として解釈することも適切といえます。しかし、変換後の近接が線型でない場合、順序変換のほうが近接に適している可能性があります。

図 14-18
変換された近接



ストレス測定

今回の解のストレスからすると、順序レベルで近接を尺度化することは適切であるといえます。

図 14-19
ストレスと適合の測定

正規化された原ストレス	.07976
ストレス-I	.28241 ^a
ストレス-II	.79034 ^a
S-ストレス	.21220 ^b
説明された散らばり (D.A.F.)	.92024
Tuckerの適合係数	.95929

PROXSCALIは正規化された原ストレスを最小化します。

a. 最適尺度因子 = 1.087。

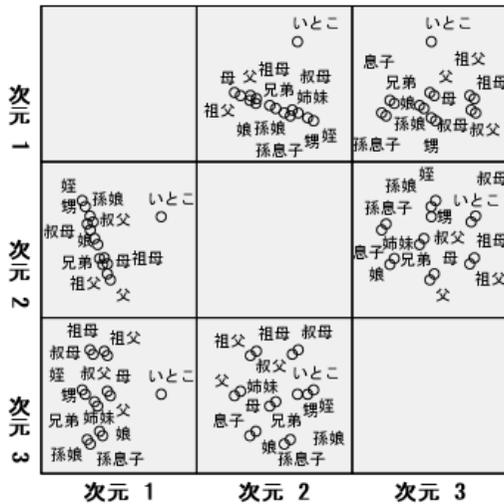
b. 最適尺度因子 = .931。

以前の解の正規化された原ストレスは、0.06234 です。デフォルト以外の変換を使用して変数を尺度化すると、ストレスは半分の 0.03137 になります。

共通空間の最終座標

共通空間プロットでは、各次元について、最初の分析の解と基本的に同じ解釈ができます。

図 14-20
共通空間の座標



考察

ストレス測度が大幅に改善されるため、近接は順序変数として扱うのが最適です。次の段階として、順序変数の「同順位をブレイク」することもできます。つまり、元の変数が等しい値でも異なる変換値を取得できるようになります。たとえば、1 番目のソースでは、[叔母] と [息子]、および [叔母] と [孫息子] の間の近接は 85 です。順序変数で「同順位を結合」するアプローチでは、この近接の変換値が等しくなりますが、変換値が等しいと想定する理由は特にありません。この場合、近接の「同順位をブレイク」することで、不必要な制限から解放されます。

推奨参考文献

多次元尺度法の詳細は、次の文献を参照してください。

Commandeur, J. J. F., および W. J. Heiser. 1993. Mathematical derivations in the proximity scaling (PROXSCAL) of symmetric data matrices. Leiden: Department of Data Theory, University of Leiden.

De Leeuw, J., および W. J. Heiser. 1980. Multidimensional scaling with restrictions on the configuration. In: Multivariate Analysis, Vol. V, P. R. Krishnaiah, ed. Amsterdam: North-Holland.

Heiser, W. J. 1981. Unfolding analysis of proximity data. Leiden: Department of Data Theory, University of Leiden.

Heiser, W. J., および F. M. T. A. Busing. 2004. Multidimensional scaling and unfolding of symmetric and asymmetric proximity relations. In: Handbook of Quantitative Methodology for the Social Sciences, D. Kaplan, ed. Thousand Oaks, Calif.: Sage Publications, Inc..

Kruskal, J. B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29, .

Kruskal, J. B. 1964. Nonmetric multidimensional scaling:A numerical method. *Psychometrika*, 29, .

Shepard, R. N. 1962. The analysis of proximities:Multidimensional scaling with an unknown distance function I. *Psychometrika*, 27, .

Shepard, R. N. 1962. The analysis of proximities:Multidimensional scaling with an unknown distance function II. *Psychometrika*, 27, .

多次元展開

[多次元展開] 手続きは共通の量的尺度を見つけるもので、2つのオブジェクト集合の関係を視覚的に調べることができるようになります。

例: 朝食メニューの嗜好

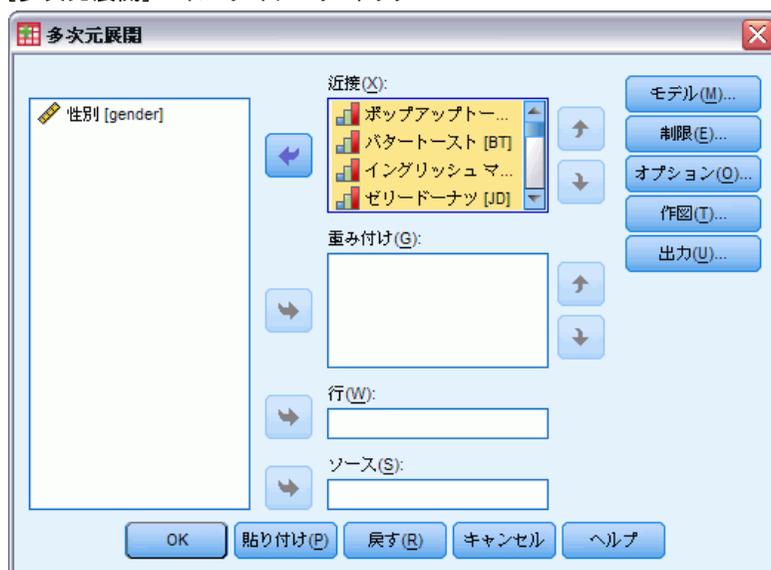
21人の Wharton School MBA の学生およびその配偶者に、15種類の朝食を好みの順に (1 = 「最も好き」から 15 = 「最も嫌い」まで) ランク付けするよう依頼した研究があります (Green および Rao, 1972)。この情報は breakfast_overall.sav に収集されています。詳細は、A 付録 p. 321 [サンプルファイル](#) を参照してください。

調査結果は、縮退の問題が見られる典型例です。縮退の問題は、多次元展開アルゴリズムにおいては避けることが難しいものですが、変換後の近接の変動係数にペナルティを課すことによって解決できます (Busing, Groenen, および Heiser, 2005)。調査結果の分析を通じて、縮退解について知り、多次元展開を使った問題解決方法を理解し、個人がどのように朝食のメニューを判別するか判断することができるようになります。調査結果の分析を再生するシンタックスは、prefscal_breakfast-overall.sps にあります。

縮退解の生成

- ▶ 多次元展開分析を実行するには、メニューから次の項目を選択します。
[分析] > [スケール] > [多次元展開 (PREFSCAL)...]

図 15-1
[多次元展開] メイン ダイアログ ボックス



- ▶ 「ポップアップトースト」から「コーン マフィンとバター」までを近接変数として選択します。
- ▶ [オプション] をクリックします。

図 15-2
[オプション] ダイアログ ボックス

多次元展開: オプション

初期布置

古典法(A)
代入(U): Spearman

Ross-Cliff

応答(Q)

重心(D)
選択肢(I): 1

ランダム スタート(R)
スタートの数(N): 1

ユーザーによる指定(C)

反復基準

ストレス収束(V): .000001

最小ストレス(T): .0001

最大反復回数(X): 5000

ペナルティ項目

強さ(H): 1.0

範囲(G): 0.0

ユーザーの指定による布置

変数の読み込み: ファイル(F)...

変数の数はモデルの最大次元数に一致している必要があります。現在の次元数: 2
行座標を含む変数は列座標より先行しなくてはなりません。

使用可能(B):

選択済み(L):

続行 キャンセル ヘルプ

- ▶ [初期布置] は [古典法] を選び、[代入] の方法として [Spearman] を選択します。
- ▶ [ペナルティ項目] グループで、[強さ] のパラメータとして 1.0 を入力し、[範囲] のパラメータとして 0.0 を入力します。これで [ペナルティ項目] はオフになります。
- ▶ [続行] をクリックします。
- ▶ [多次元展開] ダイアログ ボックスの [OK] をクリックします。

以上の選択により、次のコマンド シンタックスが生成されます。

```
PREFSCAL
VARIABLES=TP BT EMM JD CT BMM HRB TMd BTJ TMn CB DP GD CC CMB
/INITIAL=CLASSICAL (SPEARMAN)
/TRANSFORMATION=NONE
/PROXIMITIES=DISSIMILARITIES
```

```

/CRITERIA=DIMENSIONS(2,2) DIFFSTRESS(.000001) MINSTRESS(.0001)
MAXITER(5000)
/PENALTY=LAMBDA(1.0) OMEGA(0.0)
/PRINT=MEASURES COMMON
/PLOT=COMMON .

```

- このシンタックスは、変数 TP (ポップアップトースト) から変数 CMB (コーン マフィンとバター) までの分析を指定します。
- INITIAL サブコマンドは、Spearman 距離を使って初期値を代入するよう指定します。
- PENALTY サブコマンドで指定された値は、基本的にペナルティの項目を非表示にするため、その手続きにより Kruskal のストレス-I が最小化されます。結果的に縮退解となります。
- PLOT サブコマンドは、共通空間の作図を要求します。
- 他のすべてのパラメータは、デフォルト値へ戻ります。

測定方法 (GLM 反復測定)

図 15-3
縮退解の測定方法

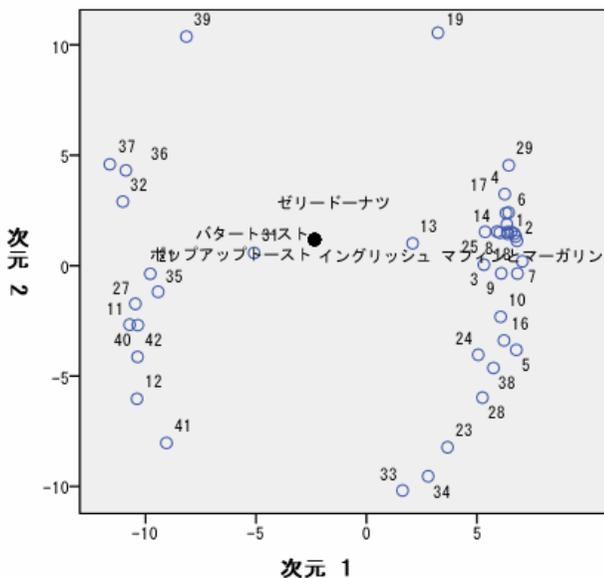
反復回数		154
最終関数値		.0000990
関数値の部分	ストレス部分	.0000990
	ペナルティ部分	1.0000000
当てはめの不適合度	正規化されたストレス	.0000000
	Kruskal の S-ストレス-I	.0000990
	Kruskal の S-ストレス-II	.6129749
	Young の S-ストレス-I	.0001980
	Young の S-ストレス-II	.7703817
当てはめの適合度	説明された分散	1.0000000
	説明された分散	.6230788
	回復された嗜好順序	.7074830
	Spearman のロー	.7450748
	Kendall のタウ b	.6218729
変動係数	変動近接	.5590170
	変動変換近接	.0000924
	変動距離	.1808765
退化指数	DeSarbo の混合指数の平方和	1.1731154E2
	Shepard の粗非退化指数	.0000000

アルゴリズムは 154 回反復した後、0.0000990 のペナルティ付きストレス (マークの付いた最後の関数値) で解に収束します。ペナルティ項目がオフになっているため、ペナルティ付きストレスは Kruskal のストレス-I に等しくなります (関数値のストレス部分は Kruskal 不適合度と一致します)。一般的にストレスの値が低いと解がデータに正確に適合しますが、縮退解の警告信号がいくつかあります。

- 変換後の近接の変動係数は、変換前の近接の変動係数と比べると非常に小さくなっています。これは、各行に対する変換後の近接がほぼ定数であるため、解がオブジェクト間の判別を行わないことを示しています。
- DeSarbo の混合指数の平方和は、さまざまな集合のポイントがどの程度混合しているかを示す測定値です。混合していない場合は、解が縮退する可能性があるという警告信号です。0 に近いほど、解はより混合したものになります。表示された値は非常に大きく、解が混合していないことを示しています。
- 異なる距離のパーセントとして表示される Shepard の粗非退化指数は、0 です。これは、異なる距離が不十分で、解が縮退する可能性があることを明確に数値で示したものです。

共通空間

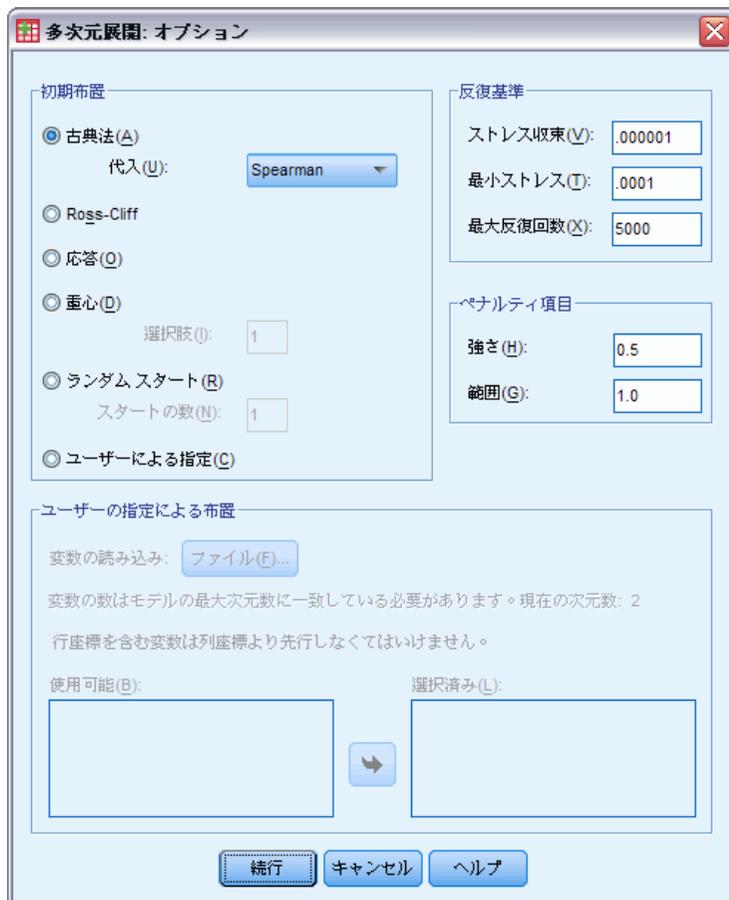
図 15-4
縮退解の共通空間の結合プロット



解の縮退は、行オブジェクトと列オブジェクトの共通空間の結合プロットで図示して確認できます。列オブジェクト（朝食メニュー）の座標は単一の点に集約されており、行オブジェクト（個人）は列オブジェクトを中心とする円周上に位置しています。

非縮退分析の実行

図 15-5
[オプション] ダイアログ ボックス



- ▶ 非縮退解を生成するには、[ダイアログのリコール] ツールをクリックして、[多次元展開] を選択します。
- ▶ [多次元展開] ダイアログ ボックスの [オプション] をクリックします。
- ▶ [ペナルティ項目] グループで、[強さ] のパラメータとして 1.0 を入力し、[範囲] のパラメータとして 0.0 を入力します。これで [ペナルティ項目] はオフになります。
- ▶ [続行] をクリックします。
- ▶ [多次元展開] ダイアログ ボックスの [OK] をクリックします。

以上の選択により、次のコマンド シンタックスが生成されます。

PREFSCAL

```

VARIABLES=TP BT EMM JD CT BMM HRB TmD BTJ TMn CB DP GD CC CMB
/INITIAL=CLASSICAL (SPEARMAN)
/TRANSFORMATION=NONE
/PROXIMITIES=DISSIMILARITIES
/CRITERIA=DIMENSIONS(2,2) DIFFSTRESS(.000001) MINSTRESS(.0001)
MAXITER(5000)
/PENALTY=LAMBDA(0.5) OMEGA(1.0)
/PRINT=MEASURES COMMON
/PLOT=COMMON .

```

- 変わったのは PENALTY サブコマンドの部分だけです。LAMBDA に 0.5 が、OMEGA に 1.0 がそれぞれのデフォルト値として設定されています。

測定方法 (GLM 反復測定)

図 15-6
非縮退解の測定方法

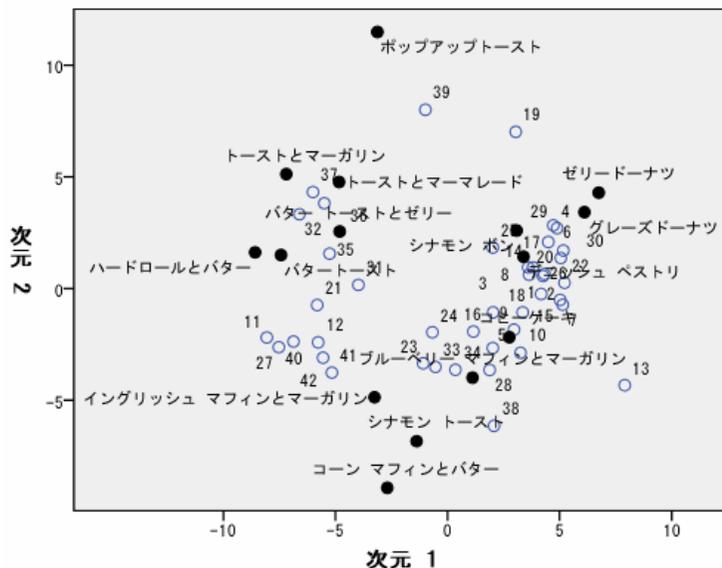
反復回数		157
最終関数値		.6848930
関数値の部分	ストレス部分	.2428268
	ペナルティ部分	1.9317409
当てはめの不適合度	正規化されたストレス	.0583589
	Kruskal の S-ストレス-I	.2415758
	Kruskal の S-ストレス-II	.5875599
	Young の S-ストレス-I	.3446361
	Young の S-ストレス-II	.5030127
当てはめの適合度	説明された分散	.9416411
	説明された分散	.7651552
	回復された嗜好順序	.7818594
	Spearman のロー	.8179181
	Kendall のタウ b	.6916725
変動係数	変動近接	.5590170
	変動変換近接	.6006156
	変動距離	.4833617
退化指数	DeSarbo の混合指数の平方和	.1590979
	Shepard の粗非退化指数	.7895692

縮退解の測定方法で示した問題点は、ここでは修正されています。

- 正規化されたストレスが、0 ではなくなっています。
- 変換後の近接の変動係数は、変換前の近接の変動係数に似た値になります。
- DeSarbo の混合指数は、はるかに 0 に近くなっています。これは解の混合の程度がはるかに高いことを示しています。
- 異なる距離のパーセントとして表示されている Shepard の粗非退化指数は、ほぼ 80% になっています。異なる距離が十分にあり、解は非縮退と考えられます。

共通空間

図 15-7
非縮退解の共通空間の結合プロット



共通空間の結合プロットでは、次元の解釈が可能になっています。水平次元には、柔らかいパン/トーストと、硬いパン/トーストとの判別が検出され、軸に沿って右へ移動するほどより柔らかいメニューになります。垂直次元には明確な解釈はありませんが、おそらくは利便性に基づいて判別しているようであり、軸に沿って下へ移動するほどより「フォーマル」なメニューになっています。

こうして、朝食メニューにいくつかのクラスタが作成されます。たとえば、ドーナツ、シナモンパン、デニッシュペストリーは柔らかく幾分正式なメニューのクラスタを形成します。マフィンおよびシナモントーストは、やや「硬め」でより「フォーマル」なメニューのクラスタを形成します。その他のトーストおよびハードロールは、「硬め」で若干「インフォーマル」なメニューのクラスタを形成します。ポップアップトーストは「硬め」で非常に「インフォーマル」なメニューです。

行オブジェクトで表示される各個人は、「硬め」か「柔らかめ」のメニューの好みに基づいて明確にクラスタに分割されますが、垂直次元では、大きなクラスタ内変動が見られます。

例: 朝食メニューの嗜好の三次展開

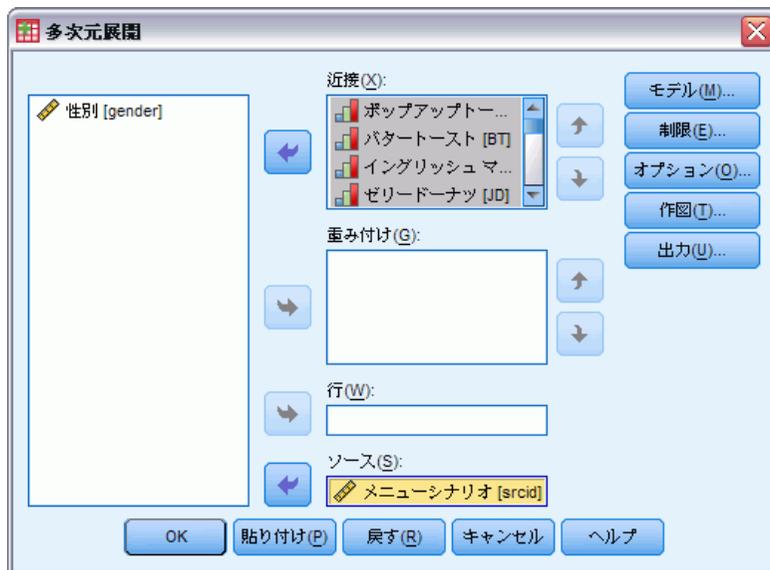
21 人の Wharton School MBA の学生およびその配偶者に、15 種類の朝食を好みの順に (1 = 「最も好き」 から 15 = 「最も嫌い」 まで) ランク付けするよう依頼した研究があります (Green など, 1972)。調査対象者の嗜好は、「すべて」から「スナックとドリンクのみ」まで、6 つの異なるシナリオに基づいて記録されました。この情報は breakfast.sav に収集されています。詳細は、A 付録 p. 321 [サンプル ファイル](#) を参照してください。

6 つのシナリオは、別々のソースとして扱うことができます。PREFSCAL を使用して行、列、およびソースの三次展開を実行します。調査結果の分析を再生するシンタックスは、prefscal_breakfast.sps にあります。

分析の実行

- ▶ 多次元展開分析を実行するには、メニューから次の項目を選択します。
[分析] > [スケール] > [多次元展開 (PREFSCAL)...]

図 15-8
[多次元展開] メイン ダイアログ ボックス



- ▶ 「ポップアップトースト」から「コーン マフィンとバター」までを近接変数として選択します。
- ▶ 「メニューシナリオ」をソース変数として選択します。
- ▶ [モデル] をクリックします。

図 15-9
[モデル] ダイアログ ボックス



- ▶ 尺度モデルとして [重み付きユークリッド] を選択します。
- ▶ [続行] をクリックします。
- ▶ [多次元展開] ダイアログ ボックスの [オプション] をクリックします。

図 15-10
[オプション] ダイアログ ボックス

多次元展開: オプション

初期布置

古典法(A)
代入(U): Spearman

Ross-Cliff

応答(Q)

垂心(D)
選択肢(I): 1

ランダム スタート(R)
スタートの数(N): 1

ユーザーによる指定(C)

反復基準

ストレス収束(V): 0.000001

最小ストレス(T): 0.0001

最大反復回数(X): 5000

ペナルティ項目

強さ(H): 0.5

範囲(G): 1.0

ユーザーの指定による布置

変数の読み込み: ファイル(F)...

変数の数はモデルの最大次元数に一致している必要があります。現在の次元数: 2
行座標を含む変数は列座標より先行しなくてはなりません。

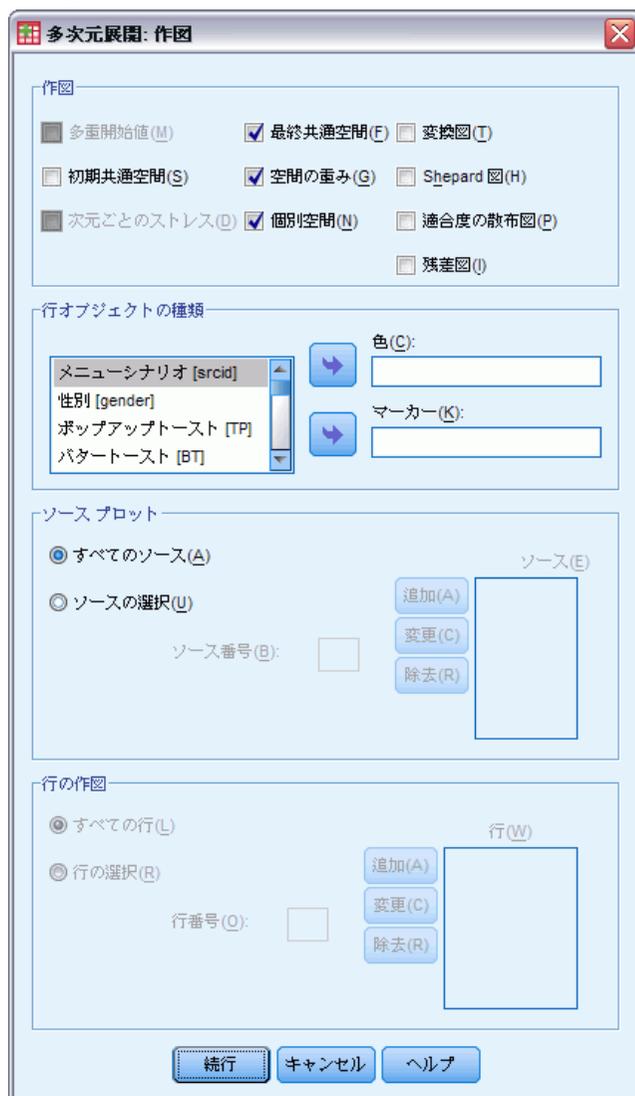
使用可能(B):

選択済み(L):

続行 キャンセル ヘルプ

- ▶ [初期布置] は [古典法] を選び、[代入] の方法として [Spearman] を選択します。
- ▶ [続行] をクリックします。
- ▶ [多次元展開] ダイアログ ボックスの [作図] をクリックします。

図 15-11
[作図] ダイアログ ボックス



- ▶ [作図] グループの [個別空間] を選択します。
- ▶ [続行] をクリックします。
- ▶ [多次元展開] ダイアログ ボックスの [OK] をクリックします。

以上の選択により、次のコマンド シンタックスが生成されます。

```
PREFSCAL
VARIABLES=TP BT EMM JD CT BMM HRB TMd BTJ TMn CB DP GD CC CMB
/INPUT=SOURCES(srcid )
/INITIAL=CLASSICAL (SPEARMAN)
```

```

/CONDITION=ROW
/TRANSFORMATION=NONE
/PROXIMITIES=DISSIMILARITIES
/MODEL=WEIGHTED
/CRITERIA=DIMENSIONS(2,2) DIFFSTRESS(.000001) MINSTRESS(.0001)
MAXITER(5000)
/PENALTY=LAMBDA(0.5) OMEGA(1.0)
/PRINT=MEASURES COMMON
/PLOT=COMMON WEIGHTS INDIVIDUAL ( ALL ) .

```

- このシンタックスは、変数 TP (ポップアップトースト) から変数 CMB (コーンマフィンとバター) までの分析を指定します。ソースを指定するために変数 srcid を使用します。
- INITIAL サブコマンドは、Spearman 距離を使って初期値を代入するよう指定します。
- MODEL サブコマンドは、重み付きユークリッドモデルを指定します。個別空間ごとに共通空間の次元を重み付けできるようになります。
- PLOT サブコマンドは、共通空間のプロット、個別空間のプロット、および個別空間の重み付けのプロットを要求します。
- 他のすべてのパラメータは、デフォルト値へ戻ります。

測定方法 (GLM 反復測定)

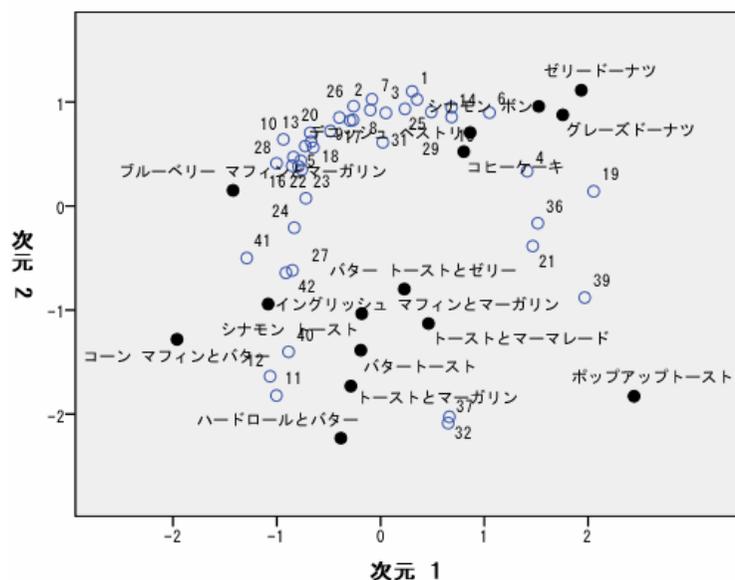
図 15-12
測定方法 (GLM 反復測定)

反復回数		481
最終関数値		.8199642
関数値の部分	ストレス部分	.3680994
	ペナルティ部分	1.8265211
当てはめの不適合度	正規化されたストレス	.1335343
	Kruskal の S-ストレス-I	.3654234
	Kruskal の S-ストレス-II	.9780824
	Young の S-ストレス-I	.4938016
	Young の S-ストレス-II	.6912352
当てはめの適合度	説明された分散	.8664657
	説明された分散	.5024853
	回復された嗜好順序	.7025321
	Spearman のロー	.6271702
	Kendall のタウ b	.4991188
変動係数	変動近接	.5590170
	変動変換近接	.6378878
	変動距離	.4484515
退化指数	DeSarbo の混合指数の平方和	.2199287
	Shepard の粗非退化指数	.7643613

アルゴリズムは 481 回反復した後、0.8199642 のペナルティ付きストレスを最後に収束します。変動係数および Shepard 指数が十分に大きく、DeSarbo 指数が十分小さいので、縮退に関する問題はないと考えられます。

共通空間

図 15-13
共通空間の結合プロット



共通空間の結合プロットは、すべての嗜好に基づく二次分析に非常によく似た最終布置を示しており、ちょうど解を 45° 線で反転させた状態です。したがって、垂直次元には、柔らかいパン/トーストと、硬いパン/トーストとの判別が検出され、軸に沿って上へ移動するほどより柔らかいメニューになります。今回は水平次元には明確な解釈がありませんが、おそらくは利便性に基づいて判別しているようであり、軸に沿って左へ移動するほどより「フォーマル」なメニューになっています。

行オブジェクトで表示される各個人は、やはり「硬め」か「柔らかめ」のメニューの好みに基づいて明確にクラスタに分割されますが、水平次元では、大きなクラスタ内変動が見られます。

個別空間

図 15-14
次元の重みのプロット

		次元		
		1	2	特異性 ^a
ソース	すべて	3.235	4.297	.186
	モーニングセットと ジュース、ベーコン、 卵、ドリンク	4.883	2.193	.457
	モーニングセットと ジュース、冷たいシリア ル、ドリンク	4.131	3.438	.109
	モーニングセットと ジュース、パンケーキ、 ソーセージ、ドリンク	4.291	3.267	.164
	モーニングセットとドリ ンクのみ	3.124	4.413	.223
	スナックとドリンクのみ	2.750	4.541	.313
	重要度 ^b	.504	.496	

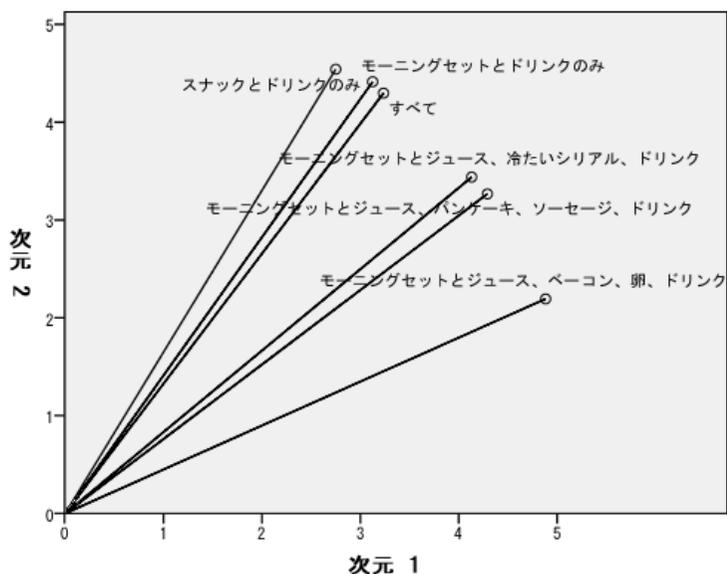
a. 特異性はソースの典型性を示します。特異性の範囲は0～1です。0は同一の次元の重みの平均的なソースを示し、1は1つの例外的な大きい次元の重みとその他の0に近い重みの非常に特殊なソースを示します。

b. その次元の平方和と全平方和の比率として示される各次元の相対的な重要度。

個別空間はソースごとに計算されます。次元の重みは、個別空間が共通空間の次元にどのように負荷をかけているかを示します。重みが大きいほど個別空間における差異が大きいことを意味し、したがってその個別空間では、対象次元におけるオブジェクト間の判別が大きくなります。

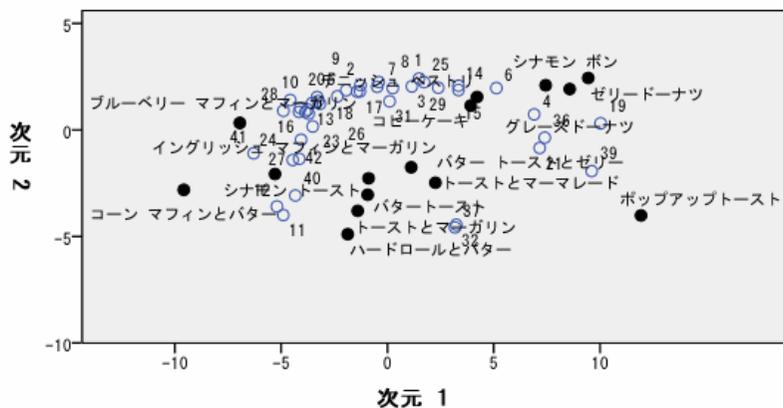
- **特異性** は、個別空間が共通空間とどれくらい異なるかを示す測定値です。共通の空間と同一の個々の空間は同じ次元の重みと、0の特異度を持ちます。特定次元に特有の個々の空間は単一の大きい次元の重みと1の特異度を持ちます。この場合、共通空間からの相違が最も大きいソースは「モーニングセット（ジュース、ベーコン エッグ、ドリンク付）」および「スナックとドリンクのみ」です。
- **重要度** は、解に対する各次元の相対的寄与の測定値です。今回の分析では、どちらの次元も重要度では差がありません。

図 15-15
次元の重みのプロット



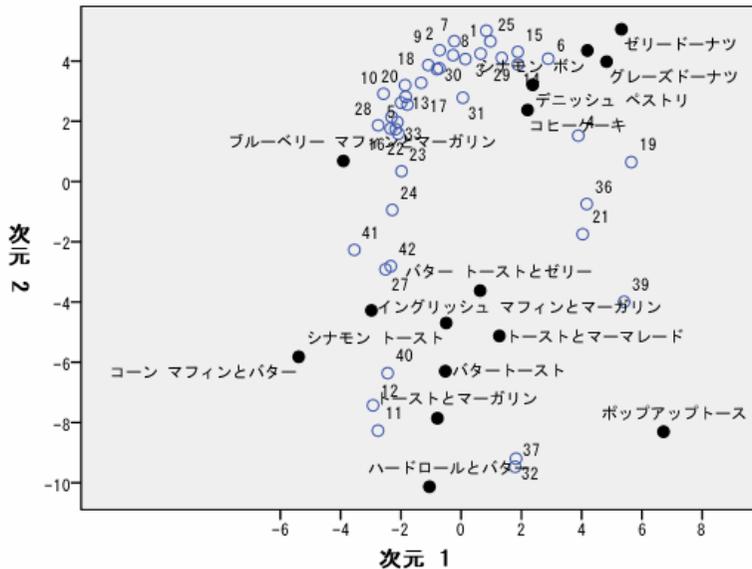
次元の重みのプロットにより、次元の重みの表を図示できます。「モーニングセット (ジュース、ベーコン エッグ、ドリンク付)」および「スナックとドリンクのみ」が次元軸に最も近くなっていますが、いずれも特定の次元に強く偏っているというほどではありません。

図 15-16
個別空間の結合プロット「モーニングセット (ジュース、ベーコン エッグ、ドリンク付)」



個別空間の結合プロット「モーニングセット (ジュース、ベーコン エッグ、ドリンク付)」は、各嗜好に基づいたシナリオの効果を示しています。このソースは第 1 次元により大きな負荷をもたらしており、そのため、メニュー間の差はほとんど第 1 次元によるものとなっています。

図 15-17
個別空間の結合プロット「スナックとドリンクのみ」



個別空間の結合プロット「スナックとドリンクのみ」は、嗜好に基づいたシナリオの効果を示しています。このソースは第2次元により大きな負荷をもたらしており、そのため、メニュー間の差はほとんど第2次元によるものとなっています。ただし、第1次元においてもかなりの差があります。これは、ソースの特異性がそれほど高くないためです。

異なる初期布置の使い方

最終布置はアルゴリズムが受け取る開始点によって変わる場合があります。解の全体的な構造がそのまま同じであるというのが理想的な状況です。そうでないと、場合によってはどちらが正しいか確定するのが困難になります。しかし、朝食データの三次分析におけるコレスポネンススタートを使用するなど、異なる初期布置を実行することにより、分析結果がより詳細で明確になる場合があります。

- ▶ コレスポネンス スタートを使用して解を生成するには、[ダイアログのリコール] ツールをクリックして、[多次元展開] を選択します。

- ▶ [多次元展開] ダイアログ ボックスの [オプション] をクリックします。

図 15-18
[オプション] ダイアログ ボックス

- ▶ [初期布置] グループの [応答] を選択します。
- ▶ [続行] をクリックします。
- ▶ [多次元展開] ダイアログ ボックスの [OK] をクリックします。

以上の選択により、次のコマンド シンタックスが生成されます。

PREFSCAL

```
VARIABLES=TP BT EMM JD CT BMM HRB TmD BTJ TMn CB DP GD CC CMB
/INPUT=SOURCES(srcid )
/INITIAL=CORRESPONDENCE
/TRANSFORMATION=NONE
/PROXIMITIES=DISSIMILARITIES
/CRITERIA=DIMENSIONS(2,2) DIFFSTRESS(.000001) MINSTRESS(.0001)
MAXITER(5000)
/PENALTY=LAMBDA(0.5) OMEGA(1.0)
/PRINT=MEASURES COMMON
```

/PLOT=COMMON WEIGHTS INDIVIDUAL (ALL) .

- 変わったのは INITIAL サブコマンドの部分だけです。開始布置は CORRESPONDENCE に設定されており、行スコアおよび列スコアの対称的正規化と、逆データ（非類似度の代わりに類似度）上で行うコレスポネンデンス分析の結果を使用します。

測定方法 (GLM 反復測定)

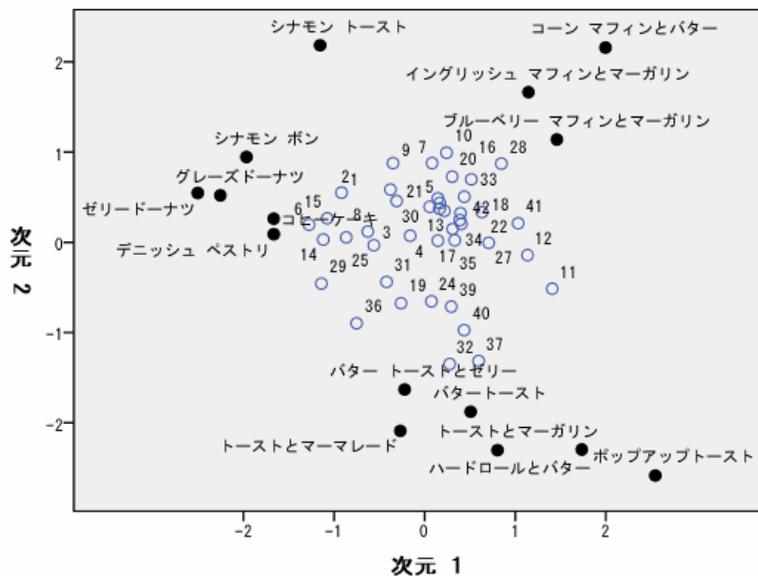
図 15-19
コレスポネンデンス初期布置の測定方法

反復回数		385
最終関数値		.8140741
関数値の部分	ストレス部分	.3493640
	ペナルティ部分	1.8969229
当てはめの不適合度	正規化されたストレス	.1212145
	Kruskal の S-ストレス-I	.3481587
	Kruskal の S-ストレス-II	1.0770522
	Young の S-ストレス-I	.4812632
	Young の S-ストレス-II	.6871733
当てはめの適合度	説明された分散	.8787855
	説明された分散	.5183498
	回復された嗜好順序	.7174981
	Spearman のロー	.6446272
	Kendall のタウ b	.5165230
変動係数	変動近接	.5590170
	変動変換近接	.6122308
	変動距離	.4043695
退化指数	DeSarbo の混合指数の平方和	1.7571887
	Shepard の粗非退化指数	.7532124

アルゴリズムは 385 回反復した後、0.8140741 のペナルティ付きストレスを最後に収束します。この統計量、不適合度、適合度、変動係数、および Shepard のインデックスはすべて、古典的な Spearman の開始を使った解の導き方と非常によく似ています。DeSarbo 指数は少し異なっており、値が 0.2199287 から 1.7571887 に変化しています。これは、コレスポネンデンス スタートを使用した解が相対的に混合されていないことを示しています。これが解にどのような影響を与えるかを確認するには、共通空間の結合プロットを参照してください。

共通空間

図 15-20
コレスポンデンス初期布置の共通空間の結合プロット



共通空間の結合プロットは、古典法の Spearman の初期布置による分析によく似た最終布置を示しています。ただし、列オブジェクト（朝食メニュー）は行オブジェクト（個人）と混合せずにその周囲に位置しています。

個別空間

図 15-21
コレスポンデンス初期布置に対する次元の重み

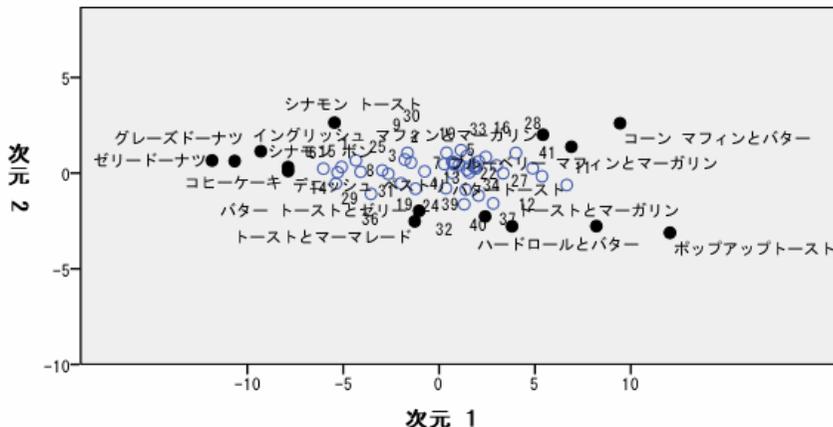
		次元		
		1	2	特異性 ^a
ソース	すべて	2.836	3.877	.279
	モーニングセットと ジュース、ベーコン、 卵、ドリンク	4.727	1.207	.636
	モーニングセットと ジュース、冷たいシリア ル、ドリンク	4.183	2.377	.263
	モーニングセットと ジュース、パンケーキ、 ソーセージ、ドリンク	4.412	1.993	.389
	モーニングセットとドリ ンクのみ	2.605	4.050	.351
	スナックとドリンクのみ	1.864	4.415	.552
	重要度 ^b	.556	.444	

a. 特異性はソースの典型性を示します。特異性の範囲は0～1です。0は同一の次元の重みの平均的なソースを示し、1は1つの例外的な大きい次元の重みとその他の0に近い重みの非常に特殊なソースを示します。

b. その次元の平方和と全平方和の比率として示される各次元の相対的な重要度。

コレスポンデンス初期布置を使用すると、各個別空間の特異性が増しています。つまり、調査対象者が朝食の各メニューをランク付けした状況が、特定の次元により強く関連しています。共通空間からの相違が最も大きいソースは依然として「モーニングセット（ジュース、ベーコン エッグ、ドリンク付）」および「スナックとドリンクのみ」です。

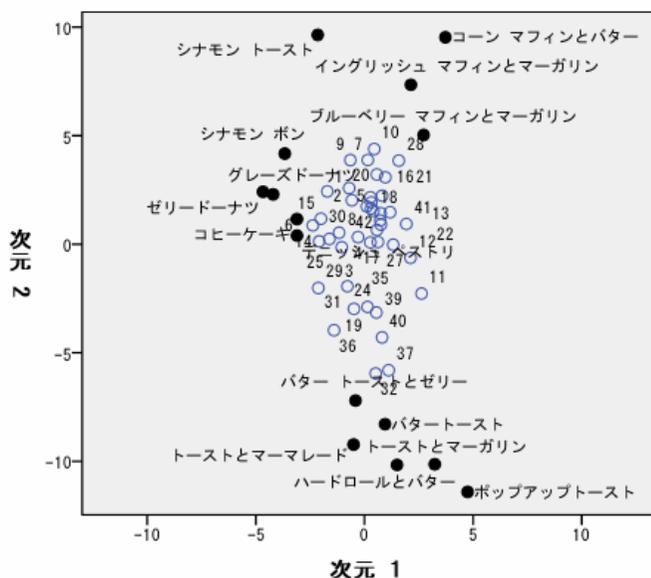
図 15-22
コレスポンデンス初期布置での個別空間の結合プロット「モーニングセット（ジュース、ベーコン エッグ、ドリンク付）」



個別空間の結合プロット「モーニングセット（ジュース、ベーコン エッグ、ドリンク付）」では、明らかに特異性が増しています。このソースは、古典法の Spearman の開始点の場合よりもさらに第 1 次元に負荷をもたらし

ており、そのため、行オブジェクトおよび列オブジェクトは、垂直軸では変動が小さくなり、水平軸では変動が大きくなっています。

図 15-23
コレスポンデンス初期布置での個別空間の結合プロット「スナックとドリンクのみ」



個別空間の結合プロット「スナックとドリンクのみ」では、行オブジェクトおよび列オブジェクトが古典法の Spearman の開始点を使用した場合よりも垂直線により近い配置で示されています。

例:「行動 - 状況」の適切性の調査

52 人の学生に 15 の状況と 15 の行動の組み合わせについて、0 = 「非常に適切」から 9 = 「非常に不適切」までの 10 段階でランク付けするよう依頼した研究があります (Price および Bouffard, 1974)。個人間の平均を取ったため、値は非類似度としてみなされます。

この情報は、behavior.sav に収集されています。詳細は、A 付録 p. 321 [サンプル ファイル](#) を参照してください。多次元展開を使用して、類似した状況およびその状況と最も密接に関連付けられた行動のクラスタを見つけます。調査結果の分析を再生するシンタックスは、prefscal_behavior.sps にあります。

分析の実行

- ▶ 多次元展開分析を実行するには、メニューから次の項目を選択します。
[分析] > [スケール] > [多次元展開 (PREFSCAL)...]

図 15-24
[多次元展開] メイン ダイアログ ボックス



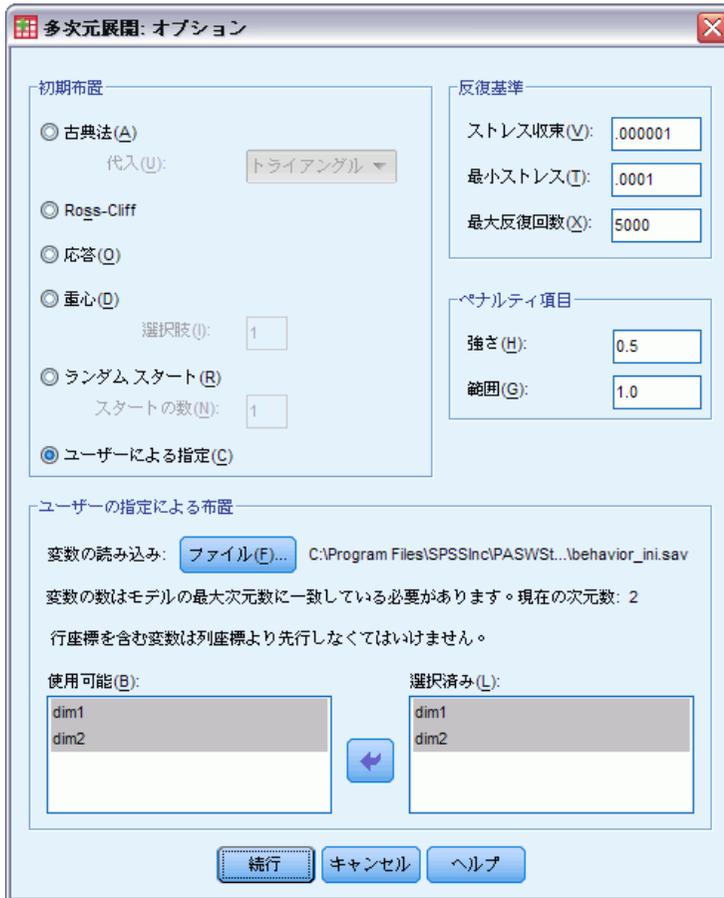
- ▶ 「Run」から「Shout」までを近接変数として選択します。
- ▶ 「ROWID」を行変数として選択します。
- ▶ [モデル] をクリックします。

図 15-25
[モデル] ダイアログ ボックス



- ▶ 近接変換として [線型] を選択し、[定数項を含める] を選択します。
- ▶ [変換の適用] で [全ソースに同時に変換] を選択します。
- ▶ [続行] をクリックします。
- ▶ [多次元展開] ダイアログ ボックスの [オプション] をクリックします。

図 15-26
[オプション] ダイアログ ボックス



- ▶ [初期布置] グループの [ユーザーの指定による] を選択します。
- ▶ [ユーザーの指定による布置] で [ファイル] をクリックし、behavior_ini.sav を参照し、選択します。詳細は、A 付録 p. 321 サンプル ファイル を参照してください。
- ▶ 「dim1」と「dim2」を、初期布置を指定する [選択された変数] として選択します。
- ▶ [続行] をクリックします。
- ▶ [多次元展開] ダイアログ ボックスの [作図] をクリックします。

図 15-27
[作図] ダイアログ ボックス



- ▶ [作図] グループの [変換図] を選択します。
- ▶ [続行] をクリックします。
- ▶ [多次元展開] ダイアログ ボックスの [OK] をクリックします。

以上の選択により、次のコマンド シンタックスが生成されます。

PREFSCAL

```
VARIABLES=Run Talk Kiss Write Eat Sleep Mumble Read Fight Belch Argue Jump
Cry Laugh Shout
/INPUT=ROWS (ROWID )
```

```
/INITIAL=( 'samplesDirectory/behavior_ini.sav' )
dim1 dim2
/CONDITION=UNCONDITIONAL
/TRANSFORMATION=LINEAR (INTERCEPT)
/PROXIMITIES=DISSIMILARITIES
/MODEL=IDENTITY
/CRITERIA=DIMENSIONS(2,2) DIFFSTRESS(.000001) MINSTRESS(.0001)
MAXITER(5000)
/PENALTY=LAMBDA(0.5) OMEGA(1.0)
/PRINT=MEASURES COMMON
/PLOT=COMMON TRANSFORMATIONS .
```

- このシンタックスは変数 Run から変数 Shout までの分析を指定します。行を指定するために変数 ROWID を使用します。
- INITIAL サブコマンドは、behavior_ini.sav ファイルから開始値を取得するように指定します。行および列座標は、行座標の後に列座標が積み重ねられます。
- CONDITION サブコマンドは、すべての近接が互いに比較されるよう指定します。今回の分析においてなぜこの指定が当てはまるのかというと、公園内を走ることと教会内を走ることの近接を比較することで、一方の行動が他方の行動よりも適切でないと考えられていることがわかるからです。
- TRANSFORMATION サブコマンドは、定数項を含む近接の線型変換を指定します。この指定が適切になるのは、近接の 1 段階ごとの差が 10 段階の全範囲において等しい場合です。すなわち、0 と 1 の差が 5 と 6 の差と等しくなるように学生がスコアを割り当てた場合、線型変換は適切な方法となります。
- PLOT サブコマンドは、共通空間のプロットと変換プロットを要求します。
- 他のすべてのパラメータは、デフォルト値へ戻ります。

測定方法 (GLM 反復測定)

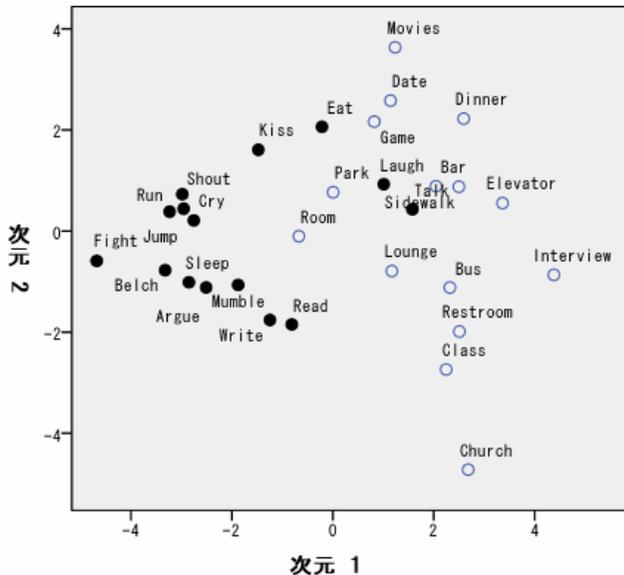
図 15-28
測定方法 (GLM 反復測定)

反復回数		169
最終関数値		.6427725
関数値の部分	ストレス部分	.1900001
	ペナルティ部分	2.1745069
当てはめの不適合度	正規化されたストレス	.0361000
	Kruskal の S-ストレス-I	.1900001
	Kruskal の S-ストレス-II	.5224668
	Young の S-ストレス-I	.2760971
	Young の S-ストレス-II	.4525933
当てはめの適合度	説明された分散	.9639000
	説明された分散	.8082862
	回復された嗜好順序	.8608333
	Spearman のロー	.8981120
	Kendall のタウ b	.7202452
変動係数	変動近接	.5138436
	変動変換近接	.4751934
	変動距離	.3912592
退化指数	DeSarbo の混合指数の平方和	.4957969
	Shepard の粗非退化指数	.7173810

アルゴリズムは 169 回反復した後、0.6427725 のペナルティ付きストレスを最後に収束します。変動係数および Shepard 指数が十分に大きく、DeSarbo 指数が十分小さいので、縮退に関する問題はないと考えられます。

共通空間

図 15-29
共通空間の結合プロット

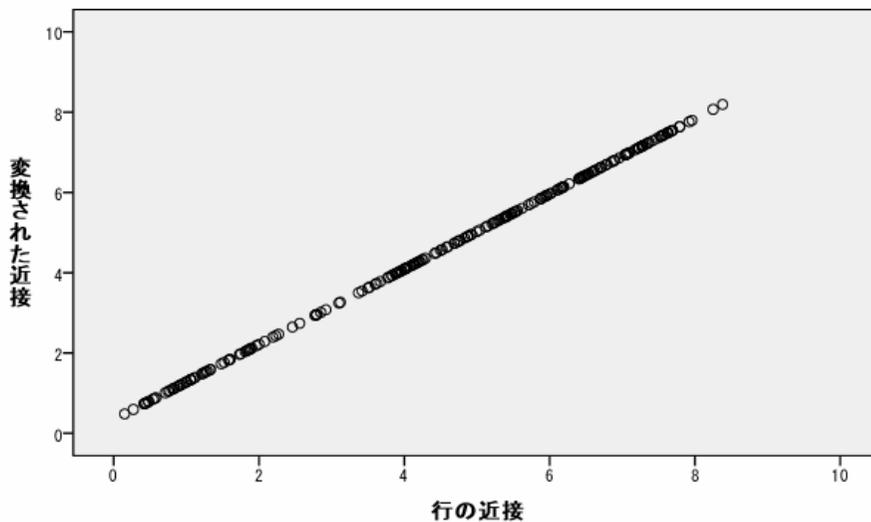


水平次元は列オブジェクト（行動）とより強い相関関係にあるように見え、不適切な行動（喧嘩する (Fight)、げっぷする (Belch)) と、より適切な行動とを判別します。垂直次元は行オブジェクト（状況）とより強い相関関係にあるように見え、さまざまな「状況 - 行動」の制限を定義します。

- 垂直次元の下方にある状況（教会 (Church)、教室 (Class)) では、適切な行動は、より静かな/内省的な行動（読む (Read)、書く (Write)) に制限されます。そのために、より静かな/内省的な行動は垂直軸の下方に位置しています。
- 垂直次元の上方にある状況（映画 (Movies)、ゲーム (Game)、デート (Date)) では、適切な行動は、より社会的な/外向的な行動（食べる (Eat)、キスする (Kiss)、笑う (Laugh)) に制限されます。そのために、より社会的な/外向的な行動は垂直軸の上方に位置しています。
- 垂直次元の中央では、状況は、各状況の通常制限に基づき水平次元に沿って分かれています。行動から離れている状況（インタビュー (Interview)) は制限度が最も高く、行動に近い状況（部屋 (Room)、公園 (Park)) は一般的に制限度が低いものです。

近接変換

図 15-30
変換プロット



無条件 線型 変換 定数項を含む

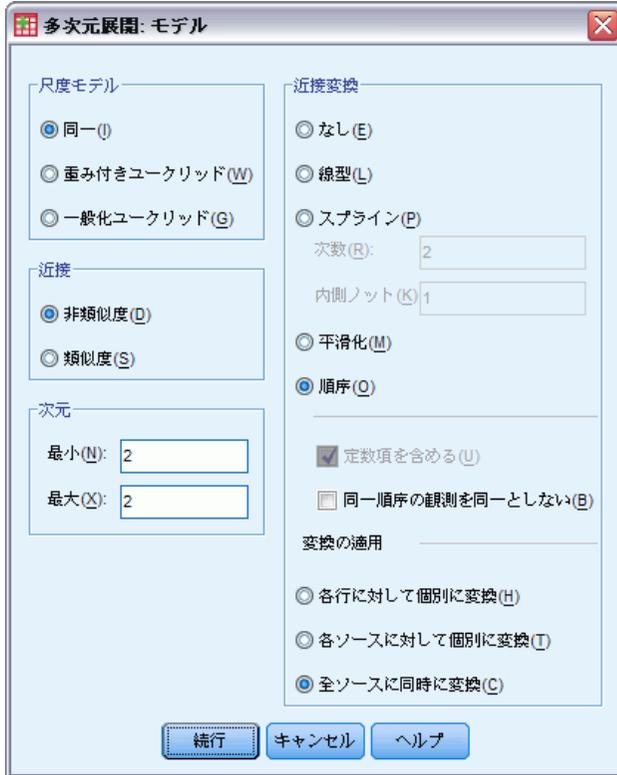
この分析では、近接を線型と見なすため、変換後の近接と変換前の近接を対比するプロットは直線となります。この解の当てはめは適切ですが、近接に異なる変換を行うとさらに適切に当てはめることができると思われます。

近接変換の変更(順序)

- ▶ 近接の順序変換による解を生成するには、[ダイアログのリコール] ツールをクリックして、[多次元展開]を選択します。

- ▶ [多次元展開] ダイアログ ボックスの [モデル] をクリックします。

図 15-31
[モデル] ダイアログ ボックス



- ▶ 近接変換として [順序] を選択します。
- ▶ [続行] をクリックします。
- ▶ [多次元展開] ダイアログ ボックスの [OK] をクリックします。

以上の選択により、次のコマンド シンタックスが生成されます。

```

PREFSCAL
VARIABLES=Run Talk Kiss Write Eat Sleep Mumble Read Fight Belch Argue Jump
Cry Laugh Shout
/INPUT=ROWS (ROWID )
/INITIAL=( 'samplesDirectory/behavior_ini.sav' )
dim1 dim2
/CONDITION=UNCONDITIONAL
/TRANSFORMATION=ORDINAL (KEEP TIES)
/PROXIMITIES=DISSIMILARITIES
/MODEL=IDENTITY
/CRITERIA=DIMENSIONS (2,2) DIFFSTRESS (.000001) MINSTRESS (.0001)
MAXITER (5000)
/PENALTY=LAMBDA (0.5) OMEGA (1.0)
/PRINT=MEASURES COMMON

```

/PLOT=COMMON TRANSFORMATIONS .

- 変わったのは TRANSFORMATION サブコマンドの部分だけです。変換は ORDINAL に設定されており、近接の順序は保持しますが、変換後の値が変換前の値の比を保つことは要求しません。

測定方法 (GLM 反復測定)

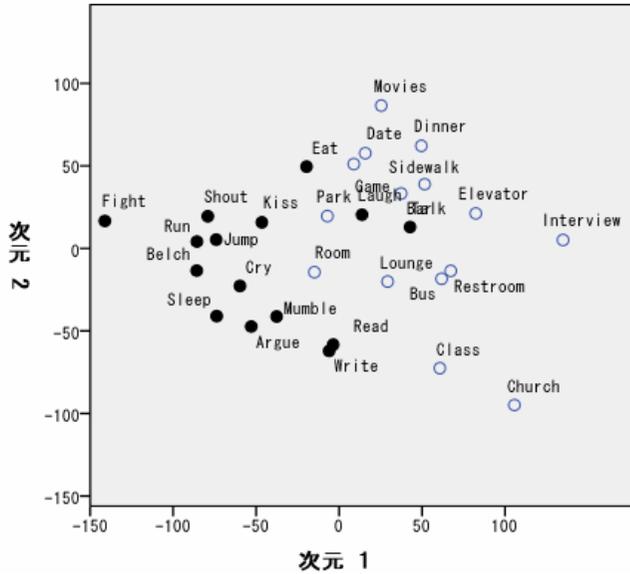
図 15-32
順序変換による解の測定方法

反復回数		268
最終関数値		.6044671
関数値の部分	ストレス部分	.1747239
	ペナルティ部分	2.0911875
当てはめの不適合度	正規化されたストレス	.0305285
	Kruskal の S-ストレス-I	.1747239
	Kruskal の S-ストレス-II	.4444641
	Young の S-ストレス-I	.2707147
	Young の S-ストレス-II	.3978003
当てはめの適合度	説明された分散	.9694715
	説明された分散	.8454488
	回復された嗜好順序	.8574206
	Spearman のロー	.9032676
	Kendall のタウ b	.7532788
変動係数	変動近接	.5138436
	変動変換近接	.4930018
	変動距離	.4284849
退化指数	DeSarbo の混合指数の平方和	.3610680
	Shepard の粗非退化指数	.7469048

アルゴリズムは 268 回反復した後、0.6044671 のペナルティ付きストレスを最後に収束します。この統計量および他の測定方法は、この解に対しては近接の線型変換を使用するよりも少し適切です。

共通空間

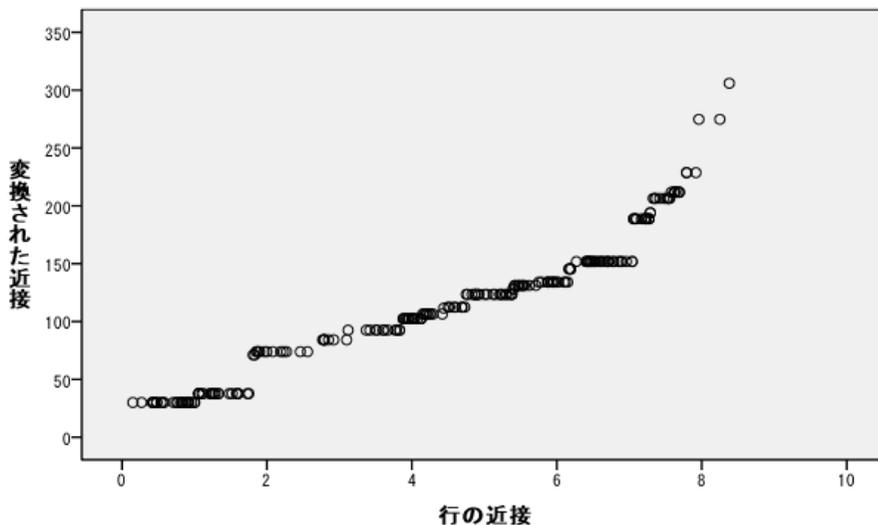
図 15-33
順序変換による解の共通空間の結合プロット



共通空間の解釈は、いずれの解においても同じです。線型変換による解の場合、水平次元よりも垂直次元の方が明らかに大きい変動を示していましたが、この解（順序変換による）では、そのような変動の差は相対的に小さいようです。

近接変換

図 15-34
順序変換による解の変換プロット



無条件 順序 変換 同順位を同一のままとする

大きな近接を持つ値は、他の値と比べ、折れ曲がったように上方に位置しているため、これを除くとすれば、近接の順序変換はほぼ線型です。線型になっていない大きな近接が、順序解と線型解の差の大部分であるように思われます。ただし、今の段階では十分な情報がないため、より大きな値における非線型トレンドが真のトレンドであるのか異常であるのかを判断することはできません。

推奨参考文献

詳細は、次のテキストを参照してください。

Busing, F. M. T. A., P. J. F. Groenen, および W. J. Heiser. 2005. Avoiding degeneracy in multidimensional unfolding by penalizing on the coefficient of variation. *Psychometrika*, 70, .

Green, P. E., および V. Rao. 1972. *Applied multidimensional scaling*. Hinsdale, Ill.: Dryden Press.

Price, R. H., および D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, 30, .

サンプル ファイル

製品とともにインストールされるサンプル ファイルは、インストール ディレクトリの Samples サブディレクトリにあります。[サンプル] サブディレクトリ内に次の各言語の別のフォルダがあります。英語、フランス語、ドイツ語、イタリア語、日本語、韓国語、ポーランド語、ロシア語、簡体字中国語、スペイン語、そして繁体中国語です。

すべてのサンプル ファイルが、すべての言語で使用できるわけではありません。サンプル ファイルがある言語で使用できない場合、その言語のフォルダには、サンプル ファイルの英語バージョンが含まれています。

説明

以下は、このドキュメントのさまざまな例で使用されているサンプル ファイルの簡単な説明です。

- **accidents.sav**。与えられた地域での自動車事故の危険因子を年齢および性別ごとに調べている保険会社に関する架空のデータ ファイルです。各ケースが、年齢カテゴリと性別のクロス分類に対応します。
- **adl.sav**。脳卒中患者に提案される治療の効果を特定するための取り組みに関する架空のデータ ファイルです。医師団は、女性の脳卒中患者たちを、2つのグループのいずれかにランダムに割り当てました。一方のグループは標準的な理学療法を受け、もう一方のグループは感情面の治療も追加で受けました。治療の3か月後に、各患者が日常生活の一般的な行動をどの程度とることができるかを、順序変数として得点付けしました。
- **advert.sav**。広告費とその売上成果の関係を調べるための小売業者の取り組みに関する架空のデータ ファイルです。この小売業者は、そのために、過去の売上と、それに関する広告費のデータを収集しました。
- **aflatoxin.sav**。収穫物によって濃度が大きく異なる毒物であるアフラトキシンを、トウモロコシの収穫物に関して検定することに関する架空のデータ ファイルです。ある穀物加工業者は、8つそれぞれの収穫物から16のサンプルを受け取って、10億分の1単位でアフラトキシン レベルを測定しました。

- **aflatoxin20.sav.** このデータ ファイルには、aflatoxin.sav データ ファイルの収穫物 4 および 8 の 16 個のサンプルに関して、アフラトキシンの測定結果が含まれています。
- **anorectic.sav.** 拒食行動または過食行動の標準的な症状の特定を目指して、調査員 (Van der Ham, Meulman, Van Strien, および Van Engeland, 1997) が、摂食障害を持つ大人 55 人の調査を行いました。各患者が 4 年間で 4 回診察を受けたので、観測値は合計で 220 になりました。観測値ごとに、16 種類の症状に関して患者の得点が記録されました。患者 71 (2 回目)、患者 76 (2 回目)、患者 47 (3 回目) の症状の得点が見つからなかったため、残っている 217 回分の観測値が有効です。
- **autoaccidents.sav.** 運転手の年齢や性別も考慮したうえで運転手ごとの自動車事故の数をモデリングするための、ある保険アナリストの取り組みに関する架空のデータ ファイルです。各ケースが別々の運転手を表し、運転手の性別、年齢、最近 5 年間の自動車事故の回数が記録されています。
- **band.sav.** このデータ ファイルには、あるバンドの音楽 CD の架空の週間売上数が含まれています。3 つの予測変数のデータも含まれています。
- **bankloan.sav.** 債務不履行率を低減させるための銀行の取り組みに関する架空のデータ ファイルです。このファイルには、過去の顧客および見込み客 850 人に関する財務情報と人口統計情報が含まれています。最初の 700 ケースは、以前に貸付を行った顧客です。残りの 150 ケースは見込み顧客で、これらの顧客に関して銀行は信用リスクの良し悪しを分類する必要があります。
- **bankloan_binning.sav.** 過去の顧客 5,000 人に関する財務情報と人口統計情報を含む架空のデータ ファイルです。
- **behavior.sav.** 52 人の学生に 15 の状況と 15 の行動の組み合わせについて、0 = 「非常に適切」から 9 = 「非常に不適切」までの 10 段階でランク付けするよう依頼した研究があります (Price および Bouffard, 1974)。個人間の平均を取ったため、値は非類似度としてみなされます。
- **behavior_ini.sav.** このデータ ファイルには、behavior.sav の 2 次元の解の初期配置が含まれています。
- **brakes.sav.** 高性能自動車のディスク ブレーキを生産している工場での品質管理に関する架空のデータ ファイルです。このデータ ファイルには、8 台の機械で生産した 16 個のディスクの直径測定値が含まれています。ブレーキの目標の直径は 322 ミリメートルです。
- **breakfast.sav.** 21 人の Wharton School MBA の学生およびその配偶者に、15 種類の朝食を好みの順に (1 = 「最も好き」から 15 = 「最も嫌い」まで) ランク付けするよう依頼した研究があります (Green および Rao, 1972)。調査対象者の嗜好は、「すべて」から「スナックとドリンクのみ」まで、6 つの異なるシナリオに基づいて記録されました。
- **breakfast-overall.sav.** このデータ ファイルには、最初のシナリオ (「すべて」) のみの朝食の好みが含まれています。

- **broadband_1.sav**。全国規模のブロードバンド サービスの地域ごとの契約者数を含む架空のデータ ファイルです。このデータ ファイルには、85 地域の月々の契約者数が 4 年間分含まれています。
- **broadband_2.sav**。このデータ ファイルは `broadband_1.sav` と同じですが、データが 3 か月分追加されています。
- **car_insurance_claims.sav**。他の場所 (McCullagh および Nelder, 1989) で表示および分析される、自動車の損害請求に関するデータセットです。逆リンク関数を使用して従属変数の平均値を保険契約者の年齢、車種、製造年の線型結合と関連付けることにより、平均請求数はガンマ分布としてモデリングできます。申請された請求の数は、尺度重み付けとして使用できます。
- **car_sales.sav**。このデータ ファイルには、自動車のさまざまな車種やモデルの架空の売上推定値、定価、仕様が含まれています。定価と仕様はそれぞれ、`edmunds.com` と製造元のサイトから入手しました。
- **car_sales_uprepared.sav**。変換したバージョンのフィールドを含まない `car_sales.sav` の修正したバージョンです。
- **carpet.sav**。一般的な例 (Green および Wind, 1973) としては、新しいカーペット専用洗剤を市販することに関心のある企業が消費者の嗜好に関する 5 種類の因子 (パッケージのデザイン、ブランド名、価格、サービスシール、料金の払い戻し) の影響について調べたい場合があります。パッケージのデザインには、3 つの因子レベルがあります。それぞれ塗布用ブラシの位置が異なります。また、3 つのブランド名 (K2R、Glory、および Bissell)、3 つの価格水準があり、最後の 2 つの因子のそれぞれに対しては 2 つのレベル (「なし」または「あり」) があります。10 人の消費者が、これらの因子により定義された 22 個のプロファイルに順位を付けます。変数「嗜好」には、各プロファイルの平均順位の序列が含まれています。順位が低いほど、嗜好度は高くなります。この変数には、各プロファイルの嗜好測定値がすべて反映されます。
- **carpet_prefs.sav**。このデータ ファイルは `carpet.sav` と同じ例に基づいていますが、10 人の消費者それぞれから収集した実際のランキングが含まれています。消費者は、22 種類の製品プロファイルを、一番好きなものから一番嫌いなものまで順位付けすることを依頼されています。変数 `PREF1` から `PREF22` には、`carpet_plan.sav` で定義されている、関連するプロファイルの ID が含まれています。
- **catalog.sav**。このデータ ファイルには、あるカタログ会社が販売した 3 つの製品の、架空の月間売上高が含まれています。5 つの予測変数のデータも含まれています。
- **catalog_seasfac.sav**。このデータ ファイルは `catalog.sav` と同じですが、季節性の分解手続きとそれに付随する日付変数から計算した一連の季節因子が追加されています。

- **cellular.sav**。解約率を削減するための携帯電話会社の取り組みに関する架空のデータ ファイルです。解約の傾向スコアは、0 ~ 100 の範囲でアカウントに適用されます。スコアリングが 50 以上のアカウントはプロバイダの変更を考えている場合があります。
- **ceramics.sav**。新しい上質の合金に標準的な合金より高い耐熱性があるかどうかを特定するための、ある製造業者の取り組みに関する架空のデータ ファイルです。各ケースが 1 つの合金の別々のテストを表し、軸受けの耐熱温度が記録されます。
- **cereal.sav**。880 人を対象に、朝食の好みについて、年齢、性別、婚姻状況、ライフスタイルが活動的かどうか（週 2 回以上運動するか）を含めて調査した、架空のデータ ファイルです。各ケースが別々の回答者を表します。
- **clothing_defects.sav**。ある衣料品工場での品質管理工程に関する架空のデータ ファイルです。工場で生産される各ロットから、調査員が衣料品のサンプルを取り出し、不良品の数を数えます。
- **coffee.sav**。このデータ ファイルは、6 つのアイスコーヒー ブランド（Kennedy, Riquier, および Sharp, 1996）について受けた印象に関連しています。回答者は、アイス コーヒーに対する 23 の各印象属性に対して、その属性が言い表していると思われるすべてのブランドを選択しました。機密保持のため、6 つのブランドを AA、BB、CC、DD、EE、および FF で表しています。
- **contacts.sav**。企業のコンピュータ営業グループの担当者リストに関する架空のデータ ファイルです。各担当者は、所属する会社の部門および会社のランクによって分類されています。また、最新の販売金額、最後の販売以降の経過時間、担当者の会社の規模も記録されています。
- **creditpromo.sav**。最近のクレジット カード プロモーションの有効性を評価するための、あるデパートの取り組みに関する架空のデータ ファイルです。このために、500 人のカード所有者がランダムに選択されました。そのうち半分には、今後 3 か月間の買い物に関して利率を下げることをプロモーションする広告を送付しました。残り半分には、通常どおりの定期的な広告を送付しました。
- **customer_dbase.sav**。自社のデータ ウェアハウスにある情報を使用して、反応がありそうな顧客に対して特典を提供するための、ある会社の取り組みに関する架空のデータ ファイルです。顧客ベースのサブセットをランダムに選択して特典を提供し、顧客の反応が記録されています。
- **customer_information.sav**。名前や住所など、顧客の連絡先情報を含む架空のデータ ファイルです。
- **customer_subset.sav**。customer_dbase.sav の 80 件のケースのサブセット。
- **customers_model.sav**。このファイルには、あるマーケティング キャンペーンの対象になった個人に関する架空のデータが含まれています。人口統計情報、購入暦の概要、各個人がキャンペーンに反応したかどうか、などのデータが含まれています。各ケースが別々の個人を表します。

- **customers_new.sav**。このファイルには、あるマーケティング キャンペーンの候補である個人に関する架空のデータが含まれています。各個人の人口統計情報および購入暦の概要のデータが含まれています。各ケースが別々の個人を表します。
- **debate.sav**。政治討論の出席者に対して行った調査の、討論の前後それぞれの回答に関する架空のデータ ファイルです。各ケースが別々の回答者に対応します。
- **debate_aggregate.sav**。debate. sav 内の回答を集計する、架空のデータ ファイルです。各ケースが、討論前後の好みのクロス分類に対応しています。
- **demo.sav**。月々の特典を送付することを目的とした、購入顧客のデータベースに関する架空のデータ ファイルです。顧客が特典に反応したかどうか、さまざまな人口統計情報と共に記録されています。
- **demo_cs_1.sav**。調査情報のデータベースをコンパイルするための、ある会社の取り組みの最初のステップに関する架空のデータ ファイルです。各ケースが別々の都市に対応し、地域、地方、地区、および都市の ID が記録されています。
- **demo_cs_2.sav**。調査情報のデータベースをコンパイルするための、ある会社の取り組みの第 2 のステップに関する架空のデータ ファイルです。各ケースが、最初のステップで選択した都市の別々の世帯単位に対応し、地域、地方、地区、都市、区画、および単位の ID が記録されます。計画の最初の 2 つの段階からの抽出情報も含まれています。
- **demo_cs.sav**。コンプレックス サンプル計画を使用して収集された調査情報を含む架空のデータ ファイルです。各ケースが別々の世帯単位に対応し、さまざまな人口統計情報および抽出情報が記録されています。
- **dmdata.sav**。これは、人口統計およびダイレクト マーケティングの購入情報のデータを含む架空のデータ ファイルです。dmdata2. sav には、テスト メールを受け取った連絡先のサブセットに関する情報が含まれ、dmdata3. sav には、テスト メールを受け取っていない残りの連絡先に関する情報が含まれています。
- **dietstudy.sav**。この架空のデータ ファイルには、” Stillman diet” (Rickman, Mitchell, Dingman, および Dalen, 1974) の研究結果が含まれています。各ケースが別々の被験者に対応し、被験者のダイエット前後の体重 (ポンド単位) と、トリグルセリド レベル (mg/100 ml 単位) が記録されています。
- **dvdplayer.sav**。新しい DVD プレーヤーの開発に関する架空のデータ ファイルです。プロトタイプを使用して、マーケティング チームはフォーカス グループ データを収集しました。各ケースが別々の調査対象ユーザーに対応し、ユーザーの人口統計情報と、プロトタイプに関する質問への回答が記録されています。

- **german_credit.sav**。このデータ ファイルは、カリフォルニア大学アーバイン校の Repository of Machine Learning Databases (Blake および Merz, 1998) にある “German credit” データセットから取ったものです。
- **grocery_1month.sav**。この架空のデータ ファイルは、grocery_coupons.sav データ ファイルの週ごとの購入を「ロールアップ」して、各ケースが別々の顧客に対応するようにしたものです。その結果、週ごとに変わっていた変数の一部が表示されなくなり、買物の総額が、調査を行った 4 週間の買物額の合計になっています。
- **grocery_coupons.sav**。顧客の購買習慣に関心を持っている食料雑貨店チェーンが収集した調査データを含む架空のデータ ファイルです。各顧客を 4 週間に渡って追跡し、各ケースが別々の顧客の週に対応しています。その週に食料品に費やした金額も含め、顧客がいつどこで買物をするかに関する情報が記録されています。
- **guttman.sav**。Bell (Bell, 1961) は、予想される社会グループを示す表を作成しました。Guttman (Guttman, 1968) は、この表の一部を使用しました。この表では、社会交相互作用、グループへの帰属感、メンバとの物理的な近接性、関係の形式化などを表す 5 個の変数が、理論上の 7 つの社会グループと交差しています。このグループには、観衆 (例、フットボールの試合の観戦者)、視聴者 (例、映画館または授業の参加者)、公衆 (例、新聞やテレビの視聴者)、暴徒 (観衆に似ているが、より強い交相互作用がある)、第一次集団 (親密な関係)、第二次集団 (自発的な集団)、および近代コミュニティ (物理的により密接した近接性と特化されたサービスの必要性によるゆるい同盟関係) があります。
- **health_funding.sav**。医療用資金 (人口 100 人あたりの金額)、罹患率 (人口 10,000 人あたりの人数)、医療サービス機関への訪問率 (人口 10,000 人あたりの人数) のデータを含む、架空のデータ ファイルです。各ケースが別々の都市を表します。
- **hivassay.sav**。HIV 感染を発見する迅速な分析方法を開発するための、ある製薬研究所の取り組みに関する架空のデータ ファイルです。分析の結果は、8 段階の濃さの赤で表現され、色が濃いほど感染の可能性が高くなります。研究所では 2,000 件の血液サンプルに関して試験を行い、その半数が HIV に感染しており、半分は感染していませんでした。
- **hourlywagedata.sav**。管理職から現場担当まで、またさまざまな経験レベルの看護師の時給に関する架空のデータ ファイルです。
- **insurance_claims.sav**。不正請求の恐れがある、疑いを区別するためにモデルを作成する必要がある保険会社の仮説データ ファイルです。各ケースがそれぞれの請求を表します。
- **insure.sav**。10 年満期の生命保険契約に対し、顧客が請求を行うかどうかを示す危険因子を調査している保険会社に関する架空のデータ ファイルです。データ ファイルの各ケースは、年齢と性別が一致する、請求を行った契約と行わなかった契約のペアを表します。

- **judges.sav.** 訓練を受けた審判（および 1 人のファン）が 300 件の体操の演技に対して付けた得点に関する架空のデータ ファイルです。各行が別々の演技を表し、審判たちは同じ演技を見ました。
- **kinship_dat.sav.** Rosenberg と Kim (Rosenberg および Kim, 1975) は、15 種類の親族関係用語（祖父、祖母、父、母、叔父、叔母、兄弟、姉妹、いとこ、息子、娘、甥、姪、孫息子、孫娘）の分析を行いました。Rosenberg と Kim は、大学生の 4 つのグループ（女性 2 組、男性 2 組）に、類似性に基づいて上記の用語を並べ替えるよう依頼しました。2 つのグループ（女性 1 組、男性 1 組）には、1 回目と違う条件に基づいて、2 回目の並べ替えをするように頼みました。このようにして、合計で 6 つの「ソース」が取得できました。各ソースは、 15×15 の近接行列に対応します。この近接行列のセルの数は、ソースの人数から、ソース内でオブジェクトを分割した回数を引いたものです。
- **kinship_ini.sav.** このデータ ファイルには、kinship_dat.sav の 3 次元の解の初期配置が含まれています。
- **kinship_var.sav.** このデータ ファイルには、kinship_dat.sav の解の次元の解釈に使用できる独立変数である性別、世代、および (ation), and 親等が含まれています。特に、解の空間をこれらの変数の線型結合に制限するために使用できます。
- **marketvalues.sav.** 1999 ~ 2000 年間の、イリノイ州アルゴンキンの新興住宅地での住宅売上に関するデータ ファイルです。個の売り上げは、公的レコードの一種です。
- **nhis2000_subset.sav.** National Health Interview Survey (NHIS) は、米国民を対象とした人口ベースの大規模な調査です。全国の代表的な世帯サンプルについて対面式で調査が行われます。各世帯のメンバーに関して、人口統計情報、健康に関する行動および状態の観測値が得られます。このデータ ファイルには、2000 年の調査から得られた情報のサブセットが含まれています。National Center for Health Statistics. National Health Interview Survey, 2000。一般使用データおよびドキュメント。ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/。2003 年にアクセス。
- **ozone.sav.** データには、残りの変数からオゾン濃度を予測するための、6 個の気象変数に対する 330 個の観測値が含まれています。それまでの研究者 (Breiman および Friedman (F), 1985)、(Hastie および Tibshirani, 1990) が、他の研究者と共に、これらの変数間に非線型性を確認しています。この場合、標準的な回帰アプローチは使用できません。
- **pain_medication.sav.** この架空のデータ ファイルには、慢性関節炎を治療する抗炎症薬の臨床試験の結果が含まれています。特に興味深いことは、薬の効果が出るまでの時間と、既存の薬剤との比較です。

- **patient_los.sav**。この架空のデータ ファイルには、心筋梗塞 (MI、または「心臓発作」) の疑いで入院した患者の治療記録が含まれています。各ケースが別々の患者に対応し、入院に関連する多くの変数が記録されています。
- **patlos_sample.sav**。この架空のデータ ファイルには、心筋梗塞 (MI、または「心臓発作」) の治療中に血栓溶解剤を投薬された患者のサンプルの治療記録が含まれています。各ケースが別々の患者に対応し、入院に関連する多くの変数が記録されています。
- **polishing.sav**。これは、Data and Story Library の “Nambeware Polishing Times” データ ファイルです。生産スケジュールを立てるための、金属食器製造業者 (ニューメキシコ州サンタフェの Nambe Mills) の取り組みに関連しています。各ケースが、生産ラインの別々の製品を表します。各製品に関して、直径、研磨時間、価格、製品タイプが記録されています。
- **poll_cs.sav**。市民の法案支持率を議会開会前に特定するための、世論調査員の取り組みに関する架空のデータ ファイルです。各ケースは登録有権者に対応しています。ケースごとに、有権者が居住している郡、町、区域が記録されています。
- **poll_cs_sample.sav**。この架空のデータ ファイルには、poll_cs.sav の有権者のサンプルが含まれています。サンプルは、poll_csplan 計画ファイルで指定されている計画に従って抽出され、このデータ ファイルには包含確率およびサンプル重み付けが記録されています。ただし、抽出計画では確率比例 (PPS) 法を使用するため、結合選択確率を含むファイル (poll_jointprob.sav) もあります。サンプル抽出後、有権者の人口統計および法案に関する意見に対応する追加の変数が収集され、データ ファイルに追加されました。
- **property_assess.sav**。限られたリソースで資産価値評価を最新に保つための、郡の評価担当者の取り組みに関する架空のデータ ファイルです。各ケースは、前年に郡内で売却された資産に対応します。データ ファイル内の各ケースでは、資産が存在する町、最後に訪問した評価担当者、その評価からの経過時間、当時行われた評価、および資産の売却価値が記録されています。
- **property_assess_cs.sav**。限られたリソースで資産価値評価を最新に保つための、州の評価担当者の取り組みに関する架空のデータ ファイルです。各ケースは州内の資産に対応します。データ ファイル内の各ケースでは、資産が存在する郡、町、および区域、最後の評価からの経過時間、および当時行われた評価が記録されています。
- **property_assess_cs_sample.sav**。この架空のデータ ファイルには、property_assess_cs.sav の資産のサンプルが含まれています。サンプルは、property_assess_csplan 計画ファイルで指定されている計画に従って抽出され、このデータ ファイルには包含確率およびサンプル重み付けが記録されています。サンプル抽出後、現在の価値変数が収集され、データ ファイルに追加されました。

- **recidivism.sav.** 管轄地域での累犯率を把握するための、政府の法執行機関の取り組みに関する架空のデータ ファイルです。各ケースは元犯罪者に対応し、人口統計情報、最初の犯罪の詳細、初犯から 2 年以内の場合には 2 回目の逮捕までの期間が記録されています。
- **recidivism_cs_sample.sav.** 管轄地域での累犯率を把握するための、政府の法執行機関の取り組みに関する架空のデータ ファイルです。各ケースは 2003 年の 7 月に最初の逮捕から釈放された元犯罪者に対応し、人口統計情報、最初の犯罪の詳細、2006 年 7 月までの 2 回目の逮捕のデータが記録されています。犯罪者は `recidivism_cs.csplan` で指定された抽出計画に従って抽出された部門から選択されます。調査では確率比例 (PPS) 法を採用したため、結合選択確率を保持したファイル (`recidivism_cs_jointprob.sav`) も用意されています。
- **rfm_transactions.sav.** 購入日、購入品目、各取引のマネタリー量など、購買取引データを含む架空のデータ ファイルです。
- **salesperformance.sav.** 2 つの新しい販売トレーニング コースの評価に関する架空のデータ ファイルです。60 人の従業員が 3 つのグループに分けられ、全員が標準のトレーニングを受けます。さらに、グループ 2 は技術トレーニングを、グループ 3 は実践的なチュートリアルを受けます。トレーニング コースの最後に各従業員がテストを受け、得点が記録されました。データ ファイルの各ケースは別々の訓練生を表し、割り当てられたグループと、テストの得点が記録されています。
- **satisf.sav.** ある小売業者が 4 箇所の店舗で行った満足度調査に関する架空のデータ ファイルです。合計で 582 人の顧客を調査し、各ケースは 1 人の顧客からの回答を表します。
- **screws.sav.** このデータ ファイルには、ねじ、ボルト、ナット、鋸 (びょう) (Hartigan, 1975) の特性に関する情報が含まれています。
- **shampoo_ph.sav.** あるヘアケア製品工場での品質管理に関する架空のデータ ファイルです。定期的に、6 つの異なる製品が測定され、pH が記録されます。目標範囲は 4.5 ~ 5.5 です。
- **ships.sav.** 他の場所 (McCullagh など, 1989) で表示および分析される、波による貨物船への損害に関するデータセットです。件数は、船舶の種類、建造期間、およびサービス期間によって、ポワゾン率で発生するものとしてモデリングできます。因子のクロス分類によって形成されたテーブルの各セルのサービス月数の集計によって、危険にさらされる確率の値が得られます。
- **site.sav.** 業務拡大に向けて新たな用地を選択するための、ある会社の取り組みに関する架空のデータ ファイルです。2 人のコンサルタントを雇って、用地を別々に評価させました。広範囲のレポートに加えて、各用地を「良い」、「普通」、「悪い」のいずれかで集計しました。
- **smokers.sav.** このデータ ファイルは、1998 年の National Household Survey of Drug Abuse から抜粋したものであり、アメリカの世帯の確率サンプルです。 (<http://dx.doi.org/10.3886/ICPSR02934>) したがっ

て、このデータ ファイルを分析する場合は、まず人口の傾向を反映させてデータを重み付けする必要があります。

- **stroke_clean.sav**。この架空のデータ ファイルには、[データの準備] オプションの手続きを使用して整理した後の、医療データベースの状態が含まれています。
- **stroke_invalid.sav**。この架空のデータ ファイルには、医療データベースの初期状態が含まれており、データ入力にいくつかエラーがあります。
- **stroke_survival**。この架空のデータ ファイルは、虚血性脳卒中で数回の困難に直面した後リハビリ プログラムを終えた患者の生存時間に関するものです。脳卒中後、心筋梗塞の発生、虚血性脳卒中、または出血性脳卒中が注意され、イベントの時間が記録されます。脳卒中後に実施されたリハビリ プログラムの最後まで生存した患者のみが含まれるため、サンプルは左側が切り捨てられます。
- **stroke_valid.sav**。この架空のデータ ファイルには、[データの検証] 手続きを使用して確認した後の、医療データベースの状態が含まれています。異常である可能性のあるケースが含まれています。
- **survey_sample.sav**。このデータ ファイルには、人口統計データおよびさまざまな態度指標などの調査データが含まれています。これは「1998 NORC General Social Survey」の変数のサブセットに基づいていますが、いくつかのデータ値が変更され、追加の架空変数がデモの目的で追加されています。
- **telco.sav**。顧客ベースにおける解約率を削減するための電気通信会社の取り組みに関する架空のデータ ファイルです。各ケースが別々の顧客に対応し、人口統計やサービス利用状況などのさまざまな情報が記録されています。
- **telco_extra.sav**。このデータ ファイルは telco.sav データ ファイルに似ていますが、「期間」および対数変換された顧客支出の属性が削除され、標準化された対数変換顧客支出の変数に置き換えられています。
- **telco_missing.sav**。このデータ ファイルは telco.sav データ ファイルのサブセットですが、一部の人口統計データ値が欠損値に置き換えられています。
- **testmarket.sav**。この架空のデータ ファイルは、新しいメニューを追加しようというファーストフードチェーンの計画に関連しています。新製品をプロモーションするためのキャンペーンには3つの候補があるため、新メニューはいくつかのランダムに選択した市場のある場所で紹介されます。場所ごとに別々のプロモーションを使用し、最初の4週間の新メニューの週間売上高が記録されます。各ケースが場所と週に対応します。
- **testmarket_1month.sav**。この架空のデータ ファイルは、testmarket.sav データ ファイルの週ごとの売上を「ロールアップ」して、各ケースが別々の場所に対応するようにしたものです。その結果、週ごとによって

ていた変数の一部が表示されなくなり、売上高が、調査を行った 4 週間の売上高の合計になっています。

- **tree_car.sav**。これは、人口統計および自動車購入価格のデータを含む架空のデータ ファイルです。
- **tree_credit.sav**。これは、人口統計および銀行ローン履歴のデータを含む架空のデータ ファイルです。
- **tree_missing_data.sav**。これは、人口統計および銀行ローン履歴のデータと、多数の欠損値を含む架空のデータ ファイルです。
- **tree_score_car.sav**。これは、人口統計および自動車購入価格のデータを含む架空のデータ ファイルです。
- **tree_textdata.sav**。尺度および値ラベルを割り当てる前の、変数のデフォルトの状態を示すことを主な目的とする、変数を 2 つだけ含む単純なデータ ファイルです。
- **tv-survey.sav**。テレビスタジオで実施された、ヒットした番組の放送期間を延長するかどうかを検討する調査に関する架空のデータ ファイルです。906 人の回答者に、さまざまな条件下でこの番組を視聴するかどうかを質問しました。各行は別々の回答者を表し、各列は別々の条件を表します。
- **ulcer_recurrence.sav**。このファイルには、潰瘍の再発を防ぐための 2 つの治療の有効性を比較するように計画された調査の情報の一部が含まれています。これは区間調査の良い例であり、他の場所 (Collett, 2003) で表示および分析されています。
- **ulcer_recurrence_recoded.sav**。このファイルでは、ulcer_recurrence.sav の情報が、単に調査終了時のイベント確率ではなく調査の区間ごとのイベント確率をモデリングできるように再編成されています。これは他の場所 (Collett など, 2003) で表示および分析されています。
- **verd1985.sav**。このデータ ファイルは調査 (Verdegaal, 1985) に関連しています。8 つの変数に対する 15 人の被験者の回答を記録しました。対象となる変数が 3 つのグループに分類されます。グループ 1 には「年齢」と「婚姻」、グループ 2 には「ペット」と「新聞」、グループ 3 には「音楽」と「居住地域」がそれぞれ含まれます。「ペット」は多重名義として尺度化され、「年齢」は順序として尺度化されます。また、その他のすべての変数は単一名義として尺度化されます。
- **virus.sav**。自社のネットワーク上のウィルスの影響を特定するための、インターネット サービス プロバイダ (ISP) の取り組みに関する架空のデータ ファイルです。この ISP は、ネットワーク上の感染した E メールトラフィックの (およその) パーセンテージを、発見の瞬間から脅威が阻止されるまで追跡しました。
- **wheeze_steubenville.sav**。これは、子供 (Ware, Dockery, Spiro III, Speizer, および Ferris Jr., 1984) に対する大気汚染の健康上の影響の長期調査から得られたサブセットです。このデータには、オハイオ州スビューベンビルの 7 歳、8 歳、9 歳、10 歳の子供を対象に行った、喘鳴

の状態の反復 2 値測定と、調査の初年に母親が喫煙していたかどうかの固定記録が含まれています。

- **workprog.sav**。体の不自由な人をより良い仕事に就かせようとする政府の事業プログラムに関する架空のデータ ファイルです。プログラムの参加者候補のサンプルが追跡されました。その中には、ランダムに選ばれてプログラムに登録された人と、そうでない人がいました。各ケースが別々のプログラム参加者を表します。

Notices

Licensed Materials – Property of SPSS Inc., an IBM Company. © Copyright SPSS Inc. 1989, 2010.

Patent No. 7,023,453

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: SPSS INC., AN IBM COMPANY, PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. SPSS Inc. may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-SPSS and non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this SPSS Inc. product and use of those Web sites is at your own risk.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Information concerning non-SPSS products was obtained from the suppliers of those products, their published announcements or other publicly available sources. SPSS has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-SPSS products. Questions on the capabilities of non-SPSS products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to SPSS Inc., for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. SPSS Inc., therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided “AS IS”, without warranty of any kind. SPSS Inc. shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and [ibm.com](http://www.ibm.com) are trademarks of IBM Corporation, registered in many jurisdictions worldwide. A current list of IBM trademarks is available on the Web at <http://www.ibm.com/legal/copytrade.shtml>.

SPSS is a trademark of SPSS Inc., an IBM Company, registered in many jurisdictions worldwide.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

This product uses WinWrap Basic, Copyright 1993-2007, Polar Engineering and Consulting, <http://www.winwrap.com>.

Other product and service names might be trademarks of IBM, SPSS, or other companies.

Adobe product screenshot(s) reprinted with permission from Adobe Systems Incorporated.

Microsoft product screenshot(s) reprinted with permission from Microsoft Corporation.



参考文献

- Barlow, R. E., D. J. Bartholomew, D. J. Bremner, および H. D. Brunk. 1972. *Statistical inference under order restrictions*. New York: John Wiley and Sons.
- Bell, E. H. 1961. *Social foundations of human behavior: Introduction to the study of sociology*. New York: Harper & Row.
- Benzécri, J. P. 1969. *Statistical analysis as a tool to make patterns emerge from data*. In: *Methodologies of Pattern Recognition*, S. Watanabe, ed. New York: Academic Press.
- Benzécri, J. P. 1992. *Correspondence analysis handbook*. New York: Marcel Dekker.
- Bishop, Y. M., S. E. Feinberg, および P. W. Holland. 1975. *Discrete multivariate analysis: Theory and practice*. Cambridge, Mass.: MIT Press.
- Blake, C. L., および C. J. Merz. 1998. "UCI Repository of machine learning databases." Available at <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- Breiman, L., および J. H. Friedman(F). 1985. *Estimating optimal transformations for multiple regression and correlation*. *Journal of the American Statistical Association*, 80, .
- Buja, A. 1990. *Remarks on functional canonical variates, alternating least squares methods and ACE*. *Annals of Statistics*, 18, .
- Busing, F. M. T. A., P. J. F. Groenen, および W. J. Heiser. 2005. *Avoiding degeneracy in multidimensional unfolding by penalizing on the coefficient of variation*. *Psychometrika*, 70, .
- Carroll, J. D. 1968. *Generalization of canonical correlation analysis to three or more sets of variables*. In: *Proceedings of the 76th Annual Convention of the American Psychological Association*, 3, Washington, D. C.: American Psychological Association.
- Collett, D. 2003. *Modelling survival data in medical research*, 2 ed. Boca Raton: Chapman & Hall/CRC.
- Commandeur, J. J. F., および W. J. Heiser. 1993. *Mathematical derivations in the proximity scaling (PROXSCAL) of symmetric data matrices*. Leiden: Department of Data Theory, University of Leiden.
- De Haas, M., J. A. Algera, H. F. J. M. Van Tuijl, および J. J. Meulman. 2000. *Macro and micro goal setting: In search of coherence*. *Applied Psychology*, 49, .
- De Leeuw, J. 1982. *Nonlinear principal components analysis*. In: *COMPSTAT Proceedings in Computational Statistics*, Vienna: Physica Verlag.

- De Leeuw, J. 1984. Canonical analysis of categorical data, 2nd ed. Leiden: DSWO Press.
- De Leeuw, J. 1984. The Gifi system of nonlinear multivariate analysis. In: Data Analysis and Informatics III, E. Diday, et al., ed. .
- De Leeuw, J., および W. J. Heiser. 1980. Multidimensional scaling with restrictions on the configuration. In: Multivariate Analysis, Vol. V, P. R. Krishnaiah, ed. Amsterdam: North-Holland.
- De Leeuw, J., および J. Van Rijckevorsel. 1980. HOMALS and PRINCALS—Some generalizations of principal components analysis. In: Data Analysis and Informatics, E. Diday, et al., ed. Amsterdam: North-Holland.
- De Leeuw, J., F. W. Young, および Y. Takane. 1976. Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, 41, .
- De Leeuw, J. 1990. Multivariate analysis with optimal scaling. In: Progress in Multivariate Analysis, S. Das Gupta, および J. Sethuraman, 編集者. Calcutta: Indian Statistical Institute.
- Eckart, C., および G. Young. 1936. The approximation of one matrix by another one of lower rank. *Psychometrika*, 1, .
- Fisher, R. A. 1938. Statistical methods for research workers. Edinburgh: Oliver and Boyd.
- Fisher, R. A. 1940. The precision of discriminant functions. *Annals of Eugenics*, 10, .
- Gabriel(G), K. R. 1971. The biplot graphic display of matrices with application to principal components analysis. *Biometrika*, 58, .
- Gifi, A. 1985. PRINCALS. Research Report UG-85-02. Leiden: Department of Data Theory, University of Leiden.
- Gifi, A. 1990. Nonlinear multivariate analysis. Chichester: John Wiley and Sons.
- Gilula, Z., および S. J. Haberman. 1988. The analysis of multivariate contingency tables by restricted canonical and restricted association models. *Journal of the American Statistical Association*, 83, .
- Gower, J. C., および J. J. Meulman. 1993. The treatment of categorical information in physical anthropology. *International Journal of Anthropology*, 8, .
- Green, P. E., および V. Rao. 1972. Applied multidimensional scaling. Hinsdale, Ill.: Dryden Press.
- Green, P. E., および Y. Wind. 1973. Multiattribute decisions in marketing: A measurement approach. Hinsdale, Ill.: Dryden Press.

参考文献

- Guttman, L. 1941. The quantification of a class of attributes: A theory and method of scale construction. In: *The Prediction of Personal Adjustment*, P. Horst, ed. New York: Social Science Research Council.
- Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points. *Psychometrika*, 33, .
- Hartigan, J. A. 1975. *Clustering algorithms*. New York: John Wiley and Sons.
- Hastie, T., および R. Tibshirani. 1990. *Generalized additive models*. London: Chapman and Hall.
- Hastie, T., R. Tibshirani, および A. Buja. 1994. Flexible discriminant analysis. *Journal of the American Statistical Association*, 89, .
- Hayashi, C. 1952. On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics*, 2, .
- Heiser, W. J. 1981. *Unfolding analysis of proximity data*. Leiden: Department of Data Theory, University of Leiden.
- Heiser, W. J., および F. M. T. A. Busing. 2004. Multidimensional scaling and unfolding of symmetric and asymmetric proximity relations. In: *Handbook of Quantitative Methodology for the Social Sciences*, D. Kaplan, ed. Thousand Oaks, Calif.: Sage Publications, Inc..
- Heiser, W. J., および J. J. Meulman. 1994. Homogeneity analysis: Exploring the distribution of variables and their nonlinear relationships. In: *Correspondence Analysis in the Social Sciences: Recent Developments and Applications*, M. Greenacre, および J. Blasius, 編集者. New York: Academic Press.
- Heiser, W. J., および J. J. Meulman. 1995. Nonlinear methods for the analysis of homogeneity and heterogeneity. In: *Recent Advances in Descriptive Multivariate Analysis*, W. J. Krzanowski, ed. Oxford: Oxford University Press.
- Horst, P. 1961. Generalized canonical correlations and their applications to experimental data. *Journal of Clinical Psychology*, 17, .
- Horst, P. 1961. Relations among m sets of measures. *Psychometrika*, 26, .
- Israëls, A. 1987. *Eigenvalue techniques for qualitative data*. Leiden: DSWO Press.
- Kennedy, R., C. Riquier, および B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. *Journal of Targeting, Measurement, and Analysis for Marketing*, 5, .
- Kettenring, J. R. 1971. Canonical analysis of several sets of variables. *Biometrika*, 58, .
- Kruskal, J. B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29, .

- Kruskal, J. B. 1964. Nonmetric multidimensional scaling:A numerical method. *Psychometrika*, 29, .
- Kruskal, J. B. 1965. Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society Series B*, 27, .
- Kruskal, J. B. 1978. Factor analysis and principal components analysis:Bilinear methods. In: *International Encyclopedia of Statistics*, W. H. Kruskal, および J. M. Tanur、編集者. New York: The Free Press.
- Kruskal, J. B., および R. N. Shepard. 1974. A nonmetric variety of linear factor analysis. *Psychometrika*, 39, .
- Krzanowski, W. J., および F. H. C. Marriott. 1994. *Multivariate analysis:Part I, distributions, ordination and inference*. London: Edward Arnold.
- Lebart, L., A. Morineau, および K. M. Warwick. 1984. *Multivariate descriptive statistical analysis*. New York: John Wiley and Sons.
- Lingoes, J. C. 1968. The multivariate analysis of qualitative data. *Multivariate Behavioral Research*, 3, .
- Max, J. 1960. Quantizing for minimum distortion. *Proceedings IEEE (Information Theory)*, 6, .
- McCullagh, P., および J. A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.
- Meulman, J. J. 1982. *Homogeneity analysis of incomplete data*. Leiden: DSWO Press.
- Meulman, J. J. 1986. *A distance approach to nonlinear multivariate analysis*. Leiden: DSWO Press.
- Meulman, J. J. 1992. The integration of multidimensional scaling and multivariate analysis with optimal transformations of the variables. *Psychometrika*, 57, .
- Meulman, J. J. 1993. Principal coordinates analysis with optimal transformations of the variables:Minimizing the sum of squares of the smallest eigenvalues. *British Journal of Mathematical and Statistical Psychology*, 46, .
- Meulman, J. J. 1996. Fitting a distance model to homogeneous subsets of variables:Points of view analysis of categorical data. *Journal of Classification*, 13, .
- Meulman, J. J. 2003. Prediction and classification in nonlinear data analysis:Something old, something new, something borrowed, something blue. *Psychometrika*, 4, .

参考文献

- Meulman, J. J., および W. J. Heiser. 1997. Graphical display of interaction in multiway contingency tables by use of homogeneity analysis. In: *Visual Display of Categorical Data*, M. Greenacre, および J. Blasius, 編集者. New York: Academic Press.
- Meulman, J. J., および P. Verboon. 1993. Points of view analysis revisited: Fitting multidimensional structures to optimal distance components with cluster restrictions on the variables. *Psychometrika*, 58, .
- Meulman, J. J., A. J. Van der Kooij, および A. Babinec. 2000. New features of categorical principal components analysis for complicated data sets, including data mining. In: *Classification, Automation and New Media*, W. Gaul, および G. Ritter, 編集者. Berlin: Springer-Verlag.
- Meulman, J. J., A. J. Van der Kooij, および W. J. Heiser. 2004. Principal components analysis with nonlinear optimal scaling transformations for ordinal and nominal data. In: *Handbook of Quantitative Methodology for the Social Sciences*, D. Kaplan, ed. Thousand Oaks, Calif.: Sage Publications, Inc..
- Nishisato, S. 1980. *Analysis of categorical data: Dual scaling and its applications*. Toronto: University of Toronto Press.
- Nishisato, S. 1984. Forced classification: A simple application of a quantification method. *Psychometrika*, 49, .
- Nishisato, S. 1994. *Elements of dual scaling: An introduction to practical data analysis*. Hillsdale, N. J.: Lawrence Erlbaum Associates, Inc.
- Pratt, J. W. 1987. Dividing the indivisible: Using simple symmetry to partition variance explained. In: *Proceedings of the Second International Conference in Statistics*, T. Pukkila, および S. Puntanen, 編集者. Tampere, Finland: University of Tampere.
- Price, R. H., および D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, 30, .
- Ramsay, J. O. 1989. Monotone regression splines in action. *Statistical Science*, 4, .
- Rao, C. R. 1973. *Linear statistical inference and its applications*, 2nd ed. New York: John Wiley and Sons.
- Rao, C. R. 1980. Matrix approximations and reduction of dimensionality in multivariate statistical analysis. In: *Multivariate Analysis*, Vol. 5, P. R. Krishnaiah, ed. Amsterdam: North-Holland.
- Rickman, R., N. Mitchell, J. Dingman, および J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. *Journal of the American Medical Association*, 228, .

- Rosenberg, S., および M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 10, .
- Roskam, E. E. 1968. Metric analysis of ordinal data in psychology. Voorschoten: VAM.
- Shepard, R. N. 1962. The analysis of proximities: Multidimensional scaling with an unknown distance function I. *Psychometrika*, 27, .
- Shepard, R. N. 1962. The analysis of proximities: Multidimensional scaling with an unknown distance function II. *Psychometrika*, 27, .
- Shepard, R. N. 1966. Metric structures in ordinal data. *Journal of Mathematical Psychology*, 3, .
- Tenenhaus, M., および F. W. Young. 1985. An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis, and other methods for quantifying categorical multivariate data. *Psychometrika*, 50, .
- Theunissen, N. C. M., J. J. Meulman, A. L. Den Ouden, H. M. Koopman, G. H. Verrips, S. P. Verloove-Vanhorick, および J. M. Wit. 2003. Changes can be studied when the measurement instrument is different at different time points. *Health Services and Outcomes Research Methodology*, 4, .
- Tucker, L. R. 1960. Intra-individual and inter-individual multidimensionality. In: *Psychological Scaling: Theory & Applications*, H. Gulliksen, および S. Messick, 編集者. New York: John Wiley and Sons.
- Van der Burg, E. 1988. *Nonlinear canonical correlation and some related techniques*. Leiden: DSWO Press.
- Van der Burg, E., および J. De Leeuw. 1983. Nonlinear canonical correlation. *British Journal of Mathematical and Statistical Psychology*, 36, .
- Van der Burg, E., J. De Leeuw, および R. Verdegaal. 1988. Homogeneity analysis with k sets of variables: An alternating least squares method with optimal scaling features. *Psychometrika*, 53, .
- Van der Ham, T., J. J. Meulman, D. C. Van Strien, および H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry*, 170, .
- Van der Kooij, A. J., および J. J. Meulman. 1997. MURALS: Multiple regression and optimal scaling using alternating least squares. In: *Softstat '97*, F. Faulbaum, および W. Bandilla, 編集者. Stuttgart: Gustav Fisher.
- Van Rijckevorsel, J. 1987. *The application of fuzzy coding and horseshoes in multiple correspondence analysis*. Leiden: DSWO Press.
- Verboon, P., および I. A. Van der Lans. 1994. Robust canonical discriminant analysis. *Psychometrika*, 59, .

参考文献

- Verdegaal, R. 1985. Meer sets analyse voor kwalitatieve gegevens (in Dutch). Leiden: Department of Data Theory, University of Leiden.
- Vlek, C., および P. J. Stallen. 1981. Judging risks and benefits in the small and in the large. *Organizational Behavior and Human Performance*, 28, .
- Wagenaar, W. A. 1988. Paradoxes of gambling behaviour. London: Lawrence Erlbaum Associates, Inc.
- Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, および B. G. Ferris Jr.. 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities. *American Review of Respiratory Diseases*, 129, .
- Winsberg, S., および J. O. Ramsay. 1980. Monotonic transformations to additivity using splines. *Biometrika*, 67, .
- Winsberg, S., および J. O. Ramsay. 1983. Monotone spline transformations for dimension reduction. *Psychometrika*, 48, .
- Wolter, K. M. 1985. Introduction to variance estimation. Berlin: Springer-Verlag.
- Young, F. W. 1981. Quantitative analysis of qualitative data. *Psychometrika*, 46, .
- Young, F. W., J. De Leeuw, および Y. Takane. 1976. Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features. *Psychometrika*, 41, .
- Young, F. W., Y. Takane, および J. De Leeuw. 1978. The principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling features. *Psychometrika*, 43, .
- Zeijl, E., Y. te Poel, M. du Bois-Reymond, J. Ravesloot, および J. J. Meulman. 2000. The role of parents and peers in the leisure activities of young adolescents. *Journal of Leisure Research*, 32, .

- 係数
 - カテゴリ回帰分析, 119, 124
 - カテゴリ回帰分析, 27
 - コレスポネンス分析, 242
- 離散化
 - カテゴリ回帰分析, 19–21
- 重要度
 - カテゴリ回帰分析, 120
 - カテゴリ主成分分析, 168, 184, 187
 - カテゴリ主成分分析, 163
 - コレスポネンス分析, 236, 243
- 交互相関
 - カテゴリ回帰分析, 118, 120
 - カテゴリ主成分分析, 166, 170
 - カテゴリ回帰分析, 24
 - 多重コレスポネンス分析, 254
- 偏相関分析
 - カテゴリ回帰分析, 120
 - カテゴリ回帰分析, 24
- 多重応答分析, 63
 - 最適尺度水準, 65
 - コマンドの追加機能, 76
 - 多重コレスポネンス分析, 65
- 多重応答分析, 248
 - 判別測定, 254
 - オブジェクト スコア, 253, 257
 - カテゴリ数量化, 256
 - モデルの要約, 252
 - 外れ値, 260
- 固有値
 - カテゴリ主成分分析, 162
 - 多重コレスポネンス分析, 66–67
- 成分負荷
 - カテゴリ主成分分析, 188
- 多重応答分析, 68
- 判別測定
 - 多重コレスポネンス分析, 71
- 多重応答分析
 - 変数の保存, 72
- 分散分析 (N)
 - カテゴリ回帰分析, 24
- 0 次相関
 - カテゴリ回帰分析, 120
- Cronbach のアルファ
 - カテゴリ主成分分析, 162
- DeSarbo の混合指数
 - 多次元展開, 289, 292, 298, 304, 313
- Elastic Net
 - カテゴリ回帰分析, 23
- Lasso
 - カテゴリ回帰分析, 23
- legal notices, 333
- PREFSCAL, 93
- R^2
 - カテゴリ回帰分析, 119
- Ridge 回帰
 - カテゴリ回帰分析, 23
- Shepard 図
 - 多次元展開, 99
- Shepard の粗非退化指数
 - 多次元展開, 289, 292, 298, 304, 313
- trademarks, 334
- 当てはめ
 - 非線型正準相関分析, 50
- 当てはめ値
 - 非線型正準相関分析, 217
- 一般化ユークリッド モデル
 - 多次元展開, 94
- イナーシャ
 - コレスポネンス分析, 59
- 補助オブジェクト
 - カテゴリ回帰分析, 21
- オブジェクト スコア
 - カテゴリ主成分分析, 165, 168, 189
 - 多重コレスポネンス分析, 71, 253, 257
- オブジェクト ポイントのプロット
 - 多重コレスポネンス分析, 73
- オブジェクト スコア
 - カテゴリ主成分分析, 39
 - 非線型正準相関分析, 50
- オブジェクト ポイントのプロット
 - カテゴリ主成分分析, 41
- 重み付きと成分負荷
 - 非線型正準相関分析, 50, 218
- 重み付きユークリッド モデル
 - 多次元展開, 94
- カテゴリ回帰, 16
- 統計量, 16, 18

索引

- カテゴリ回帰, 105
 - 保存, 26, 124
 - 重要度, 120
 - 交互相関, 118–120
 - 正規化, 23
 - コマンドの追加機能, 28
 - 変換プロット, 122
 - モデルの適合度, 119
- カテゴリ数量化
 - カテゴリ回帰分析, 24
 - 多重コレスポンデンス分析, 256
- カテゴリ主成分分析, 155, 170
 - 数量化, 184
 - 成分負荷, 166, 170
 - 数量化, 163
 - オブジェクト スコア, 165, 168
 - 反復の記述, 162
 - モデルの要約, 168, 187
 - モデルの要約, 162
- カテゴリ主成分分析
 - カテゴリ ポイント, 191
- カテゴリ主成分分析
 - 成分負荷, 188
 - オブジェクト スコア, 189
- カテゴリ数量化
 - 多重コレスポンデンス分析, 71
- カテゴリ プロット
 - 多重コレスポンデンス分析, 74
- 結合カテゴリ プロット
 - 多重コレスポンデンス分析, 74
- カテゴリ ポイント
 - カテゴリ主成分分析, 191
- カテゴリ座標
 - 非線型正準相関分析, 223
- カテゴリ主成分分析, 29, 36
 - コマンドの追加機能, 45
 - 最適尺度水準, 31
 - 変数の保存, 40
- カテゴリ数量化
 - カテゴリ主成分分析, 39
 - 非線型正準相関分析, 50
- カテゴリ プロット
 - カテゴリ主成分分析, 43
- 分散が占める度合い
 - カテゴリ主成分分析, 187
 - カテゴリ主成分分析, 162
- 共通空間の結合プロット
 - 多次元展開, 290, 293, 299, 305, 314, 318
- 共通空間
 - 多次元尺度法, 280, 283
 - 多次元展開, 290, 293, 299, 305, 314, 318
- 共通空間に対する制限
 - 多次元展開, 96
- 共通空間の座標
 - 多次元尺度法, 90
 - 多次元展開, 101
- 共通空間プロット
 - 多次元尺度法, 88
 - 多次元展開, 99
- 距離
 - 多次元尺度法, 90
 - 多次元展開, 101
- 距離測度
 - コレスポンデンス分析, 57
- 近接変換
 - 多次元展開, 94
- 空間の重みプロット
 - 多次元展開, 99
- 結合カテゴリ プロット
 - カテゴリ主成分分析, 43
- 欠損値
 - カテゴリ主成分分析, 34
- 個別空間
 - 多次元展開, 300, 306
- 個別空間の重み
 - 多次元尺度法, 90
 - 多次元展開, 101
- 個別空間の重みプロット
 - 多次元尺度法, 88
 - 多次元展開, 99
- 個別空間の結合プロット
 - 多次元展開, 300, 306
- 個別空間の座標
 - 多次元展開, 101
- 個別空間プロット
 - 多次元尺度法, 88
 - 多次元展開, 99
- 固有値
 - 非線型正準相関分析, 217
- コレスポンデンス分析, 235
 - 正規化, 236
- コレスポンデンス分析
 - 列得点のプロット, 244
- コレスポンデンス分析, 237
 - 次元, 242
 - 寄与率, 243
 - 行得点のプロット, 244
- コレスポンデンス分析, 54, 56–57, 59, 61
 - コマンドの追加機能, 62

- 作図, 54
- 統計量, 54
- 最終共通空間プロット
 - 多次元展開, 99
- 最適尺度水準
 - カテゴリ主成分分析, 31
- 作図
 - コレスポンデンス分析, 61
 - 多次元尺度法, 88–89
 - 非線型正準相関分析, 50
- 残差図
 - 多次元展開, 99
- 三次展開
 - 多次元展開, 294
- サンプル ファイル
 - 位置, 321
- 次元
 - コレスポンデンス分析, 57
- 次元の重み
 - 多次元展開, 300, 306
- 射影された重心
 - カテゴリ主成分分析, 43
 - 非線型正準相関分析, 224
- 尺度モデル
 - 多次元展開, 94
- 重心
 - 非線型正準相関分析, 50, 224
- 初期共通空間プロット
 - 多次元展開, 99
- 初期布置
 - 多次元尺度法, 86
 - 多次元展開, 98
 - 非線型正準相関分析, 50
- 信頼統計量
 - コレスポンデンス分析, 59
- 数量化
 - 非線型正準相関分析, 221
- ストレス測定
 - 多次元尺度法, 90, 278, 283
 - 多次元展開, 101
- ストレスプロット
 - 多次元尺度法, 88
 - 多次元展開, 99
- 正規化
 - コレスポンデンス分析, 57
- 制限
 - 多次元尺度法, 85
- 成分負荷
 - カテゴリ主成分分析, 39
 - 非線型正準相関分析, 50, 220
- 成分負荷プロット
 - カテゴリ主成分分析, 44
- 相関行列
 - カテゴリ主成分分析, 39
- 相関係数
 - 多次元尺度法, 90
- 相関プロット
 - 多次元尺度法, 88
- 損失値
 - 非線型正準相関分析, 217
- 多次元尺度法, 77, 79–81, 83, 266
 - オプション, 86
 - 共通空間, 280, 283
 - コマンドの追加機能, 92
 - 作図, 77, 88–89
 - 出力, 90
 - ストレス測定, 278, 283
 - 制限, 85
 - 統計量, 77
 - 変換プロット, 282
 - モデル, 84
- 多次元展開, 93, 286, 307
 - オプション, 98
 - 共通空間, 290, 293, 299, 305, 314, 318
 - 共通空間に対する制限, 96
 - 近接変換, 315, 319
 - 個別空間, 300, 306
 - コマンドの追加機能, 103
 - 作図, 93, 99
 - 三次展開, 294
 - 縮退解, 286
 - 出力, 101
 - 測定方法, 289, 292, 298, 304, 313, 317
 - 統計量, 93
 - モデル, 94
- 多重開始プロット
 - 多次元展開, 99
- 適合度の散布図
 - 多次元展開, 99
- 同一モデル
 - 多次元展開, 94
- トリプロット
 - カテゴリ主成分分析, 41

索引

- 反復の記述
 - 多重コレスポンデンス分析, 71
- 対称の正規化
 - コレスポンデンス分析, 236
- 反復の記述
 - カテゴリ主成分分析, 162
- 主成分の正規化
 - コレスポンデンス分析, 236
- 行主成分の正規化
 - コレスポンデンス分析, 236
- 行得点のプロット
 - コレスポンデンス分析, 244
- 変数の重み
 - 多重コレスポンデンス分析, 65

- バイプロット
 - カテゴリ主成分分析, 41
 - 多重コレスポンデンス分析, 73
 - コレスポンデンス分析, 61
- 反復基準
 - 多次元尺度法, 86
 - 多次元展開, 98
- 反復の記述
 - カテゴリ主成分分析, 39
 - 多次元尺度法, 90
 - 多次元展開, 101

- 非線型正準相関分析, 46, 49–50, 209
 - 重み付きと成分負荷, 218
 - カテゴリ座標, 223
 - コマンドの追加機能, 52
 - 作図, 46
 - 重心, 224
 - 数量化, 221
 - 成分負荷, 218, 220
 - 統計量, 46
 - 分析の要約, 217
- 標準化
 - コレスポンデンス分析, 57

- 変換プロット
 - カテゴリ回帰分析, 122
 - 多重コレスポンデンス分析, 74
- 判別測定プロット
 - 多重コレスポンデンス分析, 74
- 分散が占める度合い
 - カテゴリ主成分分析, 39

- ペナルティ項目
 - 多次元展開, 98
- ペナルティ付きストレス
 - 多次元展開, 289, 298, 304, 313, 317

- 変換された近接
 - 多次元尺度法, 90
 - 多次元展開, 101
- 変換された独立変数
 - 多次元尺度法, 90
- 変換プロット
 - カテゴリ主成分分析, 43
 - 多次元尺度法, 88, 282
 - 多次元展開, 99, 315, 319
- 変数の重み
 - カテゴリ主成分分析, 31
- 変動係数
 - 多次元展開, 289, 292, 298, 304, 313

- モデルの要約
 - 多重コレスポンデンス分析, 252

- より速いアルゴリズム
 - 多次元尺度法, 86

- 離散化
 - カテゴリ主成分分析, 33

- 外れ値
 - 多重コレスポンデンス分析, 260