

IBM SPSS Categories 26

IBM

注記

本書および本書で紹介する製品をご使用になる前に、49 ページの『特記事項』に記載されている情報をお読みください。

本書は、IBM SPSS Statistics バージョン 26 リリース 0 モディフィケーション 0 および新しい版で明記されない限り、以降のすべてのリリースおよびモディフィケーションに適用されます。

お客様の環境によっては、資料中の円記号がバックスラッシュと表示されたり、バックスラッシュが円記号と表示されたりする場合があります。

原典： IBM SPSS Categories 26

発行： 日本アイ・ビー・エム株式会社

担当： トランスレーション・サービス・センター

目次

カテゴリー	1	多重応答分析: 変数の重みの定義	29
カテゴリー・データ用の最適尺度法手続きの概要	1	多重応答分析: 離散化	29
最適尺度法とは	1	多重応答分析: 欠損値	30
最適尺度法を使用する理由	1	多重応答分析: オプション	31
最適尺度水準と尺度	2	多重応答分析: 出力	32
最適な手続きの選択	5	多重応答分析: 保存	33
最適尺度法の図表の縦横比	10	多重応答分析: オブジェクト・プロット	33
カテゴリー回帰分析 (CATREG)	10	多重応答分析: 変数プロット	33
カテゴリー回帰: 尺度の定義	11	MULTIPLE CORRESPONDENCE コマンドの追 加機能	34
カテゴリー回帰: 離散化	11	多次元尺度法 (PROXSCAL)	34
カテゴリー回帰: 欠損値	12	複数列行列の近接	35
カテゴリー回帰: オプション	12	複数列の近接	36
カテゴリー回帰: 正規化	13	単一列の近接	36
カテゴリー回帰: 出力	14	データからの近接度の作成	36
カテゴリー回帰: 保存	15	データからの尺度の作成	36
カテゴリー回帰: 変換プロット	15	多次元尺度法: モデルの定義	37
CATREG コマンドの追加機能	16	多次元尺度法: 制限	38
カテゴリー主成分分析 (CATPCA)	16	多次元尺度法: オプション	38
カテゴリー主成分分析: 尺度と重み付けの定義	17	多次元尺度法: 作図、バージョン 1	39
カテゴリー主成分分析: 離散化	17	多次元尺度法: 作図、バージョン 2	40
カテゴリー主成分分析: 欠損値	18	多次元尺度法: 出力	40
カテゴリー主成分分析: オプション	19	PROXSCAL コマンドの追加機能	41
カテゴリー主成分分析: 出力	20	多次元展開 (PREFSCAL)	41
カテゴリー主成分分析: 保存	21	多次元展開: モデルの定義	42
カテゴリー主成分分析: オブジェクト・プロット	21	多次元展開: 制限	43
カテゴリー主成分分析: カテゴリー・プロット	22	多次元展開: オプション	44
カテゴリー主成分分析: 因子負荷プロット	22	多次元展開: 作図	45
CATPCA コマンドの追加機能	23	多次元展開: 出力	46
コレスポンデンス分析	23	PREFSCAL コマンドの追加機能	47
コレスポンデンス分析: 行範囲の定義	24	特記事項	49
コレスポンデンス分析: 列範囲の定義	25	商標	50
コレスポンデンス分析: モデル	25	索引	53
コレスポンデンス分析: 統計	26		
コレスポンデンス分析: 作図	27		
CORRESPONDENCE コマンドの追加機能	28		
多重コレスポンデンス分析	28		

カテゴリー

以下のカテゴリー機能が、SPSS® Statistics Professional Edition または Categories オプションに含まれています。

カテゴリー・データ用の最適尺度法手続きの概要

カテゴリー手続きは、標準統計手続きでは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、各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 でのこれらの手続きの実装は、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ある考え方は、数量化した数値を各変数のカテゴリーに割り当てるということです。これにより、標準手続きを使用して、数量化された変数について解を求めることができます。

使用する手続きの最適化基準に基づいて、最適尺度値が各変数のカテゴリーに割り当てられます。分析対象の名義変数または順序変数の元のラベルとは異なり、これらの尺度値は計量プロパティを持っています。

ほとんどのカテゴリー手続きには、交互最小二乗法と呼ばれる反復法を使用して、各尺度変数の最適な数量化が実行されます。交互最小二乗法では、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が更新されます。次に、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が更新されます。この処理は、特定の停止基準に達するまで繰り返されます。

最適尺度法を使用する理由

カテゴリー・データは、マーケティング・リサーチ、サーベイ・リサーチ、社会行動科学におけるリサーチでよく使用されます。実際に、研究者の多くは、ほとんどカテゴリー・データだけを使用しています。

多くの標準モデルは特にカテゴリー・データの分析に適していますが、次のようなデータ・セットに対しては適切に機能しない場合がよくあります。

- 観測数が少なすぎる
- 変数が多すぎる
- 1 つの変数あたりの値が多すぎる

最適尺度法でカテゴリーを数量化することにより、こうした状況で発生する問題を回避することができます。さらに、最適尺度法は、専門的な手法が必要な場合にも役立ちます。

最適尺度法の出力は、パラメーター推定値を解釈するのではなく、通常はグラフィック表示に基づいて解釈します。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。調査の対象を絞り込み、最適尺度法でデータを視覚化することにより、モデル・パラメーターの解釈を中心とする分析の基盤を形成することができます。

最適尺度水準と尺度

初めてカテゴリー手続きを使用する場合、これは非常に混乱する概念です。レベルを指定する場合、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。これは、数量化する変数は、測定方法に関係なく、非線型関係を持つ可能性があるという考え方です。

カテゴリーを目的とする場合、尺度の基本レベルには以下の3つがあります。

- 名義レベルの場合、変数の値は順序なしのカテゴリーを表します。名義変数として考えられる変数の例としては、地区、郵便番号地域、信仰している宗教、複数選択カテゴリーなどがあります。
- 順序レベルの場合、変数の値は順序付きカテゴリーを表します。例としては、満足度や信頼度を表す態度尺度、嗜好評価スコアなどがあります。
- 数値レベルの場合、変数の値は、有意な測定基準で順序付けられたカテゴリーを表すため、このレベルは、カテゴリー間の距離の比較に適しています。例としては、年齢や、千ドル単位の所得額などがあります。

例えば、地域、職種、年齢の各変数は、次の表のようにコード化されます。

表 1. 地域、職種、年齢のコード化方式

地域コード	地域値	職種コード	職種値	年齢
1	北部	1	実習生	20
2	南部	2	販売員	22
3	東部	3	管理職	25
4	西部			27

この表に示されている値は、各変数のカテゴリーを表しています。地域は名義変数です。地域には4つのカテゴリーがあり、固有の順序関係はありません。値1から4は、単純に4つのカテゴリーを表しています。コード化方式は完全に任意です。一方、職種は、順序変数と考えることができます。元のカテゴリーにより、実習生から管理職までが形成されます。コードが大きいほど、企業内での職位が高くなることを表します。ただし、わかっているのは順序の情報だけであるため、隣接するカテゴリー間の距離については何もわかりません。対照的に、年齢は、数値変数と考えることができます。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歳から22歳までの距離は、25歳から27歳までの距離と同じですが、22歳から25歳までの距離は、これらのいずれの距離よりも大きなものです。

最適尺度水準の選択

変数に対して指定すべき最適尺度水準を自動的に事前定義する固有のプロパティを、変数は持っていないということを理解する必要があります。合理的で、簡単に解釈できる方法であれば、任意の方法でデータを探索することができます。例えば、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、より少ない数の次元で解を得られる可能性があります。

次の2つの例で、なぜ「明白な」尺度が最善の最適尺度水準とならない場合があるのかについて説明します。ここでは、変数により、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できますが、25歳未満の場合、安全度は年齢と正の関係を持ち、60歳を超える場合、安全度は年齢と負の関係を持つ可能性があります。この場合は、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個人を並べ替える変数は、実質的に名義変数と見なされます。ただし、左翼政党から右翼政党まで各政党に順序を付けた場合、政党の数量化で分析の順序レベルを使用して、この順序を保持したい場合があります。

変数には、その変数を排他的にどちらのレベルにするかを決定する事前定義プロパティはありませんが、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、カテゴリーの順序は不明ですが、分析では順序が必要になります。カテゴリーの順序が判明したら、順序を数量化する必要があります。カテゴリーに順序を付けることができない場合は、多重名義の数量化を行ってください。

変換プロット

各変数を尺度化できる各種レベルにより、数量化に対して各種の制限が適用されます。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ーとの関係を表します。例えば、変数を数値変数として扱えば、線型変換プロットが作成されます。順序変数として扱われる変数の場合は、非減少変換プロットが作成されます。U 字形 (または逆 U 字形) の名義変数として扱われる変数の変換プロットでは、2 次関係が示されます。名義変数の場合、カテゴリーの順序を完全に変更することにより、明確なトレンド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、選択した最適尺度水準がどの程度適切に機能するかを判断する場合に特に適しています。複数のカテゴリーに類似した数量化が割り当てられた場合、これらのカテゴリーを 1 つのカテゴリーに圧縮できることがあります。また、名義変数として扱われる変数に、増加トレンドを示す数量化が割り当てられた場合、順序変換によって類似する適合が得られることがあります。そのトレンドが線型トレンドの場合、数値処理の方が適していることがあります。ただし、カテゴリーの圧縮や尺度水準の変更が行われる場合、分析は大幅には変わりません。

カテゴリー・コード

カテゴリー変数をコード化する場合、コード化方式によっては不要な出力が発生したり、分析が不完全になったりすることがあるため、いくつかの注意点があります。職種 に使用できるコード化方式を次の表に示します。

表 2. 職種の代替コード化方式

カテゴリー	A	B	C	D
実習生	1	1	5	1
販売員	2	2	6	5
管理職	3	7	7	3

一部のカテゴリー手続きでは、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。この範囲を外れる値は欠損値として扱われます。最小カテゴリー値は常に 1 です。最大カテゴリー値はユーザーが指定します。この値は、変数のカテゴリーの数ではなく、カテゴリーの最大値です。例えば上記の表では、方式 A の最大カテゴリー値が 3 で、方式 B の最大カテゴリー値が 7 ですが、どちらの方式も、同じ 3 つのカテゴリーをコード化します。

変数の範囲により、分析から除外されるカテゴリーが決まります。定義された範囲から外れたコードを持つカテゴリーが分析から除外されます。これは、カテゴリーを除外するための単純な方法ですが、不要な分析が発生する場合があります。最大カテゴリーの定義が誤っていると、分析から有効なカテゴリーが除外される場合があります。例えば方式 B の場合、最大カテゴリー値を 3 として定義すると、職種のカテゴリーが 1 から 3 までの範囲でコード化され、管理職カテゴリーは欠損値として処理されます。実際には 3 としてコード化されているカテゴリーはないため、分析の第 3 のカテゴリーにはケースは含まれません。すべての管理職カテゴリーを除外したい場合、この分析は適切です。ただし、管理職を分析に含める場合は、最大カテゴリーを 7 と定義し、欠損値を 7 よりも大きい値としてコード化するか、1 未満の値としてコード化する必要があります。

名義変数または順序変数として処理される変数の場合、カテゴリーの範囲は結果に影響しません。名義変数の場合、重要なのはラベルだけで、そのラベルに関連付けられている値は重要ではありません。順序変数の場合、カテゴリーの順序は数値化されて保存されるため、カテゴリーの値自体は重要ではありません。同じカテゴリー順序になるコード化方式の場合、すべて同じ結果になります。例えば、職種を順序レベルで分析した場合、上記のテーブルの最初の 3 つの方式は機能的に同じになります。これは、これらのすべての方式で、カテゴリーの順序が同じになっているためです。一方、方式 D の場合、第 2 と第 3 のカテゴリーが反転され、他の方式とは異なる結果になります。

変数のコード化方式は、その多くが機能的に同じものですが、コードは手続きによって生成される出力量に影響するため、コード間の差が小さな方式のほうが適しています。1 からユーザー定義の最大値までの範囲の値での範囲でコード化されたカテゴリーはすべて有効です。これらのいずれかのカテゴリーが空の場合、対応する数値化は、システム欠損値または 0 となります。どちらになるかは、手続きによって異なります。これらの割り当てはいずれも分析には影響しませんが、これらのカテゴリーに対して出力が生成されません。したがって、方式 B の場合、職種には、システム欠損値を受け取るカテゴリーが 4 つあります。方式 C にも、システム欠損値指標を受け取るカテゴリーが 4 つあります。一方、方式 A の場合、システム欠損値が数値化されることはありません。名義変数または順序変数として処理される変数のコードとして連続する整数を使用すると、結果に影響を与えることなく、出力を大幅に削減することができます。

数値変数として処理される変数のコード化方式は、順序変数の場合よりも制限が厳しくなります。これらの変数では、連続するカテゴリー間の差が重要になります。次の表に、年齢に対する 3 つのコード化方式を示します。

表 3. 年齢の代替コード化方式

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

数値変数の再割り当てでは、カテゴリー間の差を保持する必要があります。順序値を使用することは、この差を保持するための方法の 1 つです。ただし、これにより、多数のカテゴリーでシステム欠損値指標が発生する場合があります。例えば、方式 A では、元の観測値を使用しています。コレスポネンス分析を除くすべてのカテゴリー手続きでは、最大カテゴリー値は 27、最小カテゴリー値は 1 に設定されます。最初の 19 カテゴリーは空で、システム欠損値指標を受け取ります。最大カテゴリー値が 1 よりも大幅に大きく、1 から最大値までの範囲に空のカテゴリーが多数存在する場合、不要な出力が増えることとなります。

再割り当てを実行すると、出力量を削減することができます。ただし、数値変数の場合は、「連続数への再割り当て」機能は使用しないでください。連続整数としてコード化すると、すべての連続カテゴリー間の差が 1 となるため、すべての数値化が均等間隔になります。そのため、変数を数値変数として処理する場合に重要となる計量特性が、連続整数への再割り当てによって破棄されることとなります。例えば、上記のテーブルの方式 C は、年齢の自動再割り当てに対応しています。カテゴリー 22 と 25 の差は、3 から 1 に変更され、数値化では 1 が差として反映されます。

カテゴリー間の差を保持する代替値再割り当て方式では、すべてのカテゴリーから最小カテゴリー値を減算し、それぞれの差に 1 を加算します。方式 B は、この変換から算出された結果です。各カテゴリーから最小カテゴリー値である 20 が減算され、それぞれの結果に 1 が加算されています。変換後のコードの最小値は 1 で、差はすべて元のデータと同じになります。最大カテゴリー値は 8 になり、最初の 0 以外の

数量化前の 0 数量化はすべて除外されます。ただし、方式 B から生成される、各カテゴリに対応する 0 以外の数量化は、方式 A から生成される数量化と同じになります。

最適な手続きの選択

コレスポンデンス分析、多重応答分析、カテゴリ主成分分析、非線型正準相関分析という 4 つの手続き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。この場合、変数間の関係は、できる限り少ない次元 (2 次元または 3 次元) で表されます。これにより、元のデータ量が多くて複雑なために理解するのが困難な関係でも、その関係の構造やパターンを記述することができます。市場調査の応用例では、これらの手法は知覚マッピングの形をとることがあります。これらの手続きの主な利点は、さまざまなレベルの最適尺度法でデータを処理できる点です。

カテゴリ回帰分析では、カテゴリ応答変数とカテゴリ予測変数の組み合わせ間関係が記述されます。応答変数に対する各予測変数の影響は、対応する回帰重みによって記述されます。他の手続きと同様に、さまざま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では、オブジェクト間の近接度を使用して、低次元空間内のオブジェクト間関係が記述されます。

各手続きの簡単なガイドラインを以下に示します。

- 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記述するには、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度を評価するには、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、コレスポンデ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強度の低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。
- 近接データを分析し、低次元空間で 1 組のオブジェクトの最小二乗法表現を検出するには、多次元尺度法を使用します。
- 近接データを分析し、低次元空間で 2 組のオブジェクトの最小二乗法表現を検出するには、多次元展開を使用します。

カテゴリ回帰

分析の目的として、一連の独立 (予測) 変数から従属 (応答) 変数を予測する場合、カテゴリ回帰分析を使用するのが最も適しています。すべての最適尺度法手続きと同様に、尺度値が回帰分析に関して最適になるように、すべての変数の各カテゴリに対して尺度値が割り当てられます。カテゴリ回帰分析の解により、変換後の応答変数と、変換後の予測変数の重み付きの組み合わせの平方相関が最大化されます。

他のカテゴリ手続きとの関係: 最適尺度法によるカテゴリ回帰分析は、2 組の変数を持ち、そのうちの 1 組には従属変数のみ含まれる、最適尺度正準相関分析に相当します。最適尺度正準相関分析では、すべての組のいずれかの場所にある不明な変数と各組を比較することにより、変数の組の類似度が取得されます。カテゴリ回帰分析では、変換後の応答変数と変換後の予測変数の線型結合の類似度が直接評価されます。

標準手法との関係: 標準の線型回帰分析では、カテゴリ変数を指標変数として再割り当てすることも、間隔レベル変数と同じ方法で処理することもできます。最初の方法では、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、解釈するパラメーターの数が多くなりま

す。2 番目の方法では、変数ごとに 1 つのパラメーターだけが推定されます。ただし、カテゴリは任意にコード化できるため、一般化することはできません。

一部の変数が連続型ではない場合、別の分析手法を使用することができます。応答変数が連続型で、予測変数がカテゴリ型の場合、通常は分散分析が使用されます。応答変数がカテゴリ型で、予測変数が連続型の場合、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型の場合、通常は対数線型モデルが使用されます。

最適尺度法による回帰分析では、変数ごとに 3 つの尺度水準を使用することができます。これらの水準を組み合わせると、いずれの単一の「標準的な」方法も適していない、幅広い範囲の非線型関係を記述することができます。そのため、最適尺度法は標準的な方法よりも柔軟で、付加される複雑性も最小限になります。

さらに、予測変数の非線型変換により、通常、予測変数間の依存関係が削減されます。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、最適尺度法により、予測変数の相関行列の大きな固有値が小さくなり、小さな固有値が大きくなります。

カテゴリ主成分分析

混合最適尺度水準の 1 組の変数の変動パターンを記述する場合、カテゴリ主成分分析を使用するのが最も適しています。この手法では、一連の変数の次元数を減らしながら、可能な限り変動を記述します。尺度値が主成分分析の解に関して最適になるように、すべての変数の各カテゴリに対して尺度値が割り当てられます。分析に含まれるオブジェクトは、数量化されたデータに基づいて成分スコアを受け取ります。成分スコアの作図から、分析に含まれるオブジェクト間のパターンが明らかになり、データ内の異常なオブジェクトも明らかになります。カテゴリ主成分分析の解により、指定された成分 (次元) の数について、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、回答者は、嗜好に関する多数の項目に順位を付けたり、項目を評価したりします。通常の IBM SPSS Statistics のデータ構成では、行は個人、列は項目の測定結果、行にまたがるスコアは選好スコア (例えば、0 から 10 までの尺度) であり、データは行の条件に従います。選好データでは、個人を変数として扱うことをお勧めします。入れ替え手続きを使用すると、データを入れ替えることができます。評価者が変数となり、すべての変数が順序変数として宣言されます。CATPCA では、オブジェクトよりも多数の変数を使用することができます。

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、同じ変数に対して実行された多重応答分析と同じ分析結果が生成されます。したがって、カテゴリ主成分分析は、一部の変数が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じ結果になります。

より一般的には、カテゴリ主成分分析は、非数値尺度間の相関を計算し、標準主成分分析または因子分析の方法を使用してそれらの相関を分析する方法の代替手段になります。通常の Pearson 相関係数を順序データの関連付けの測度として Naive で使用する方法では、相関の推定値で大きな偏りが発生する場合があります。

非線型正準相関分析

非線型正準相関分析は、さまざまな応用例が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的は、主成分分析のように変数自体の関係を分析することではなく、2 組以上の変数の関係を分析することです。例えば、2 組の変数があり、一方の組の変数は一連の回答者に関する人口統計学的な背景の項目で、他方の組の変数は一連の態度項目に対する回答という例が挙げられます。分析の尺度水準は、名義、順序、数値の任意の組み合わせにすることができます。最適尺度正準相関分析では、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スコアの折衷組を同時に比較することにより、組間の類似度が判別されます。

他のカテゴリー手続きとの関係: 変数の組が 2 つ以上あり、各組の変数が 1 つだけの場合、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。各組の変数が 1 つしかない分析で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と同等になります。変数の組が 2 つあり、そのうちの 1 組には変数が 1 つしかない場合、最適尺度正準相関分析は、最適尺度法によるカテゴリー回帰分析と同等になります。

標準手法との関係: 標準の正準相関分析は、1 組の変数の線型結合と、最大相関関係にある第 2 の組の変数の線型結合を検出する統計手法です。この線型結合が存在する場合、正準相関分析により、小さい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検出することができます。

分析内に 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と同等になります。IBM SPSS Statistics には正準相関分析手続きは用意されていませんが、関連する統計の多くは、多変量分散分析から取得することができます。

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、一方の組に単一の名義変数として宣言された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は、回帰分析と同様の方法で解釈することができます。変数を多重名義変数として見なす場合、最適尺度分析は判別分析の代替手法になります。変数を 3 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することができます。

コレスポンデンス分析

コレスポンデンス分析の目的は、コレスポンデンス・テーブルのバイプロットを作成することです。コレスポンデンス・テーブルでは、行変数と列変数が順序のないカテゴリーを表すことを前提とします。そのため、名義最適尺度水準が常に使用されます。どちらの変数も、名義情報についてのみ検査されます。つまり、一部のオブジェクトが同じカテゴリー内にあり、その他のオブジェクトは同じカテゴリー内にはないという事実だけが考慮されます。同じ変数のカテゴリー間の距離や順序については、何も想定されません。

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 を引いた数と c から 1 を引いた数のうちの小さい方の数になります。つまり、分割表の行カテゴリーまたは列カテゴリーは、次元空間で完全に表現できるということです。ただし、実際には、2 元表の行カテゴリーと列カテゴリーを低次元空間 (2 次元など) で表現する方が適しています。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。

使用可能な最大次元数より少ない次元数を使用すると、分析で生成された統計により、行カテゴリーと列カテゴリーが低次元表現でどの程度適切に表現されているかがわかります。2 次元の解の表現品質が高い場合、行の点と列の点から構成されるプロットを調べることにより、行変数のどのカテゴリーが類似しているか、列変数のどのカテゴリーが類似しているか、どの行カテゴリーと列カテゴリーが相互に類似しているかがわかります。

他のカテゴリ手続きとの関係: 単純コレスポネンス分析は 2 元表だけに制限されます。対象となる変数が 3 つ以上ある場合は、変数を組み合わせて交互作用変数を作成することができます。例えば、「地域」、「職種」、「年齢」という変数がある場合、「地域」と「職種」を組み合わせて、以下の表に示す 12 個のカテゴリを持つ新しい変数「地域職種」を作成することができます。この新しい変数により、「年齢」を持つ 2 元表 (12 行、4 列) が作成されます。この 2 元表は、コレスポネンス分析で分析することができます。

表 4. 地域と職種の組み合わせ

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	北部、実習生	7	東部、実習生
2	北部、販売員	8	東部、販売員
3	北部、管理職	9	東部、管理職
4	南部、実習生	10	西部、実習生
5	南部、販売員	11	西武、販売員
6	南部、管理職	12	西部、管理職

この方法の欠点は、変数のペアのすべての組み合わせが発生するという点です。「職種」と「年齢」を組み合わせて、別の 12 個のカテゴリ変数を作成することも、「地域」と「年齢」を組み合わせて、新しい 16 個のカテゴリ変数を作成することもできます。これらの各交互作用変数により、残りの変数を含む 2 元表が作成されます。これら 3 つの表のコレスポネンス分析では、同一の結果は生成されませんが、それぞれの分析は有効な方法です。また、4 つ以上の変数がある場合、一方の交互作用変数と他方の交互作用変数を比較する 2 元表を作成することができます。分析できる表の数は、変数の数が少ない場合でも、非常に多くなることがあります。1 つの表だけを選択して分析することも、すべての表を分析することもできます。または、多重応答分析手続きを使用して、すべての変数を同時に調べることもできます。その際、交互作用変数を作成する必要はありません。

標準手法との関係: クロス集計表手続きを使用して、独立変数を分析の共通の対象として持つ分割表を分析することもできます。ただし、小さな表であっても、独立変数からの逸脱の原因を検出するのは困難な場合があります。コレスポネンス分析のユーティリティには、任意のサイズの 2 元表について、こうしたパターンを表示する機能があります。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)、コレスポネンス分析を使用すると、関係の性質を明らかにするのに役立つことがあります。

多重コレスポネンス分析

多重応答分析では、同じカテゴリ内のオブジェクトは互いに近い位置にプロットされ、異なるカテゴリのオブジェクトは離れた位置にプロットされる解が求められます。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づくように配置されます。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。変数により、同じカテゴリ内のオブジェクトが同じサブグループに分類される場合、それらの変数は等質な変数と見なされます。

1 次元の解の場合、多重応答分析では、全体的に平均して、カテゴリが最大の広がりを持つような方法で、最適尺度値 (カテゴリ数量化) が各変数の各カテゴリに割り当てられます。2 次元の解の場合、多重応答分析では、第 1 の組とは無関係な各変数のカテゴリ数量化の第 2 組が求められ、さらに広がりを最大化しようとする処理が行われ、これらの処理が繰り返されます。変数のカテゴリは次元と同じ数のスコアを受け取るため、分析内の変数は、最適尺度水準の多重名義変数と仮定されます。

また、多重応答分析では、カテゴリの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) となるような方法で、分析内のオブジェクトにスコアが割り当てられます。

他のカテゴリー手続きとの関係: 多重応答分析は、等分散性分析または二重尺度法とも呼ばれます。変数が 2 つだけの場合、多重応答分析を実行すると、コレスポンデンス分析に類似する結果 (同じではありません) が得られます。コレスポンデンス分析を実行すると、安定度の情報を含む、解の表現の適合と品質を要約した一意の出力が得られます。したがって、変数が 2 つだけの場合、通常は、多重応答分析よりもコレスポンデンス分析の方が適しています。これらの 2 つの手続きのもう 1 つの相違点は、分析に対する入力です。多重応答分析に対する入力は、行がオブジェクトで列が変数になっているデータ行列ですが、コレスポンデンス分析に対する入力は、同じデータ行列にすることも、一般近接行列や結合分割表 (行と列の両方が変数のカテゴリーを表す集計行列) にすることもできます。また、多重応答分析は、多重名義レベルで尺度化され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: 多重応答分析は、多元分割表の分析と見なすことができます。多元分割表は、クロス集計表手続きで分析することもできますが、クロス集計表は、各制御変数の各カテゴリーについて個別の要約統計を出力します。多重応答分析では、多くの場合、単一の 2 次元プロットを使用して、すべての変数の関係を要約することができます。多重応答分析の高度な使用方法として、元のカテゴリー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、カテゴリー・ラベルが数値の尺度値で置き換えられるため、多重応答分析の実行後に、数値データを必要とするさまざまな手続きを適用することができます。例えば、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が出力されます。最初の次元の成分スコアはオブジェクト・スコアに等しく、平方成分負荷は判別測定に等しくなります。ただし、多重応答分析の 2 番目の次元と因子分析の 2 番目の次元は等しくありません。

多次元スケーリング法

分析の目的が、1 組のオブジェクトまたはケース間の 1 組の距離測度の構造を検出することである場合、多次元尺度法を使用するのが最も適しています。この手法では、空間内のポイント間の距離が、指定された類似度 (または非類似度) にできる限り近くなるように、概念上の低次元空間内の特定の場所に観測値が割り当てられます。これにより、その低次元空間で行オブジェクトと列オブジェクトの最小二乗表現が得られます。この表現は、多くの場合、データをより深く理解するのに役立ちます。

他のカテゴリー手続きとの関係: 距離の作成元となる、多次元尺度法で分析する多変量データがある場合、分析結果は、カテゴリー主成分分析とオブジェクト主成分正規化を使用してデータを分析した結果に類似したものになります。このような PCA は、主座標分析とも呼ばれます。

標準手法との関係: カテゴリー多次元尺度法手続き (PROXSCAL) は、Statistics Base Edition で使用できる尺度法手続きを基にいくつか改良を加えたものです (ALSCAL)。PROXSCAL は、特定モデル用の高速アルゴリズムを提供します。PROXSCAL により、共通空間に制限を適用することができます。また、PROXSCAL は、S ストレス (歪みとも呼ばれます) 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、多くの場合で適しています。一方の S ストレスは、平方距離に基づいています。

多次元展開

分析の目的が 2 組のオブジェクト (行オブジェクトと列オブジェクト) 間の 1 組の距離測度の構造を検出することである場合、多次元展開を使用するのが最も適しています。この手法では、空間内のポイント間の距離が、指定された類似度 (または非類似度) にできる限り近くなるように、概念上の低次元空間内の特定の場所に観測値が割り当てられます。これにより、その低次元空間で行オブジェクトと列オブジェクトの最小二乗表現が得られます。この表現は、多くの場合、データをより深く理解するのに役立ちます。

他のカテゴリー手続きとの関係: 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 (正方行列、対称行列)、多次元尺度法を使用します。

標準手法との関係: カテゴリー多次元展開手続き (PREFSCAL) は、Statistics Base Edition で使用できる展開機能を基にいくつか改良を加えたものです (ALSCAL を使用)。PREFSCAL では、共通空間に制限を適用することができます。また、PREFSCAL は、ペナルティー・ストレス測定の最小化を試行します。これにより、(従来のアルゴリズムでは多く発生した) 縮退解を回避することができます。

最適尺度法の図表の縦横比

最適尺度法の作図の縦横比には等方性があります。2次元の作図では、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次元の作図の次元の範囲を変更すると、他の次元のサイズも変更され、物理的な距離が等しく保たれます。最適尺度法手続きでは、等方的縦横比を上書きすることはできません。

カテゴリー回帰分析 (CATREG)

カテゴリー回帰分析は、カテゴリーに数値を割り当てることによってカテゴリー・データを数量化します。これにより、変換後の変数の最適な線型回帰方程式が作成されます。カテゴリー回帰分析の略語は CATREG (*categorical regression*) 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、それらの合計が最小化されます。変数は、通常は量的変数で、(名義) カテゴリー・データは、2 値変数または対比変数に再割り当てされます。その結果、カテゴリー変数によってケース・グループが分割され、各グループの個々のパラメーターのセットが推定されます。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でも可能です。

もう 1 つの方法として、応答変数をカテゴリー予測変数値自体で回帰させる方法があります。この場合、各変数について 1 つの係数が推定されます。ただし、カテゴリー変数の場合、カテゴリーの値は任意です。異なる方法でカテゴリーをコード化すると異なる係数が生成されるため、同一の変数について複数の分析結果を比較することが難しくなります。

CATREG は、名義変数、順序変数、数値変数を同時に尺度化することにより、標準的な方法を拡張します。この手続きにより、元のカテゴリーの特質を反映した形でカテゴリー変数が数量化されます。この手続きでは、数量化されたカテゴリー変数が数値変数と同じ方法で処理されます。非線型変換を使用すると、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: カテゴリー回帰分析を使用して、仕事の満足度が、職種、地域、移動の量とどのように関係しているかを記述することができます。例えば、高い満足度は、管理職と少ない移動量に対応している、などがわかります。結果として得られる回帰式を使用して、3 つの独立変数の任意の組み合わせについて、仕事の満足度を予測することができます。

統計と作図: 度数分布、回帰係数、分散分析表、反復の記述、カテゴリー数量化、未変換予測変数間の相関、変換済み予測変数間の相関、残差プロット、変換プロット。

カテゴリー回帰分析データの考慮事項

データ: CATREG は、カテゴリー指標変数に対して機能します。カテゴリー指標は、正の整数でなければなりません。[離散化] ダイアログ ボックスを使用して、小数値変数と文字列変数を正の整数に変換することができます。

仮定: 応答変数は 1 つしか使用できませんが、予測変数の最大数は 200 です。データには、3 つ以上の有効なケースを含める必要があります。また、有効なケースの数は、予測変数の数に 1 を加算した数よりも多くなければなりません。

関連プロシージャー: CATREG は、2 組の変数を持ち、そのうちの 1 組には変数が 1 つだけ含まれている、最適尺度法によるカテゴリー正準相関分析 (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、標準重回帰分析と同等になります。

カテゴリー回帰分析を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「回帰」 > 「最適尺度法 (CATREG)...

2. 1 つの従属変数と 1 つ以上の独立変数を選択します。

3. 「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。

カテゴリー回帰: 尺度の定義

従属変数と独立変数の最適尺度水準を設定することができます。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。また、分析変数の重みを設定することができます。

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することができます。

スプライン順序

観測変数のカテゴリーの順序は、最適尺度変数に格納されます。カテゴリー・ポイントは、原点を通る直線 (ベクトル) 上に配置されます。変換の結果は、選択された次数の滑らかな単調な区分的多項式になります。ユーザーが指定した内側ノットの数と手続きによって決定された内側ノットの配置により、区分が指定されます。

スプライン名義

観測変数内の情報のうち、カテゴリー内のオブジェクトのグループだけが最適尺度変数に格納されます。観測変数のカテゴリーの順序は格納されません。カテゴリー・ポイントは、原点を通る直線 (ベクトル) 上に配置されます。変換の結果は、選択された次数の滑らかな (場合によっては単調な) 区分的多項式になります。ユーザーが指定した内側ノットの数と手続きによって決定された内側ノットの配置により、区分が指定されます。

数値 カテゴリーは、順序付きで、等間隔 (間隔レベル) のカテゴリーとして扱われます。観測変数のカテゴリーの順序とカテゴリー番号間の等間隔性は、最適尺度変数に格納されます。カテゴリー・ポイントは、原点を通る直線 (ベクトル) 上に配置されます。すべての変数が数値レベルの場合、この分析は標準の主成分分析に類似したものになります。

順序 観測変数のカテゴリーの順序は、最適尺度変数に格納されます。カテゴリー・ポイントは、原点を通る直線 (ベクトル) 上に配置されます。変換の結果は、スプライン順序変換よりも適合性は優れていますが、滑らかさでは劣ります。

名義 観測変数内の情報のうち、カテゴリー内のオブジェクトのグループだけが最適尺度変数に格納されます。観測変数のカテゴリーの順序は格納されません。カテゴリー・ポイントは、原点を通る直線 (ベクトル) 上に配置されます。変換の結果は、スプライン名義変換よりも適合性は優れていますが、滑らかさでは劣ります。

カテゴリー回帰: 離散化

「離散化」ダイアログでは、変数の再割り当て方法を選択することができます。小数値変数は、特に指定されていない限り、ほぼ正規分布する 7 つのカテゴリー (または、その変数が持つ異なる値の数が 7 未満の

場合は、その数のカテゴリ)に分類されます。文字列変数は、英数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。文字列変数の離散化は、これらの整数に適用されます。デフォルトでは、その他の変数はそのままになります。これらの離散化された変数が分析で使用されます。

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、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って離散化されます。

乗算 離散化された値の最小値が 1 となるように、変数の現在の値を標準化し、10 を掛けて丸めた値に定数を加算します。

グループ化

グループ化によって変数を離散化する場合、次のオプションを使用できます。

カテゴリ数

カテゴリ数を指定し、変数の値をカテゴリ全体にわたってほぼ正規分布させるか、または一様分布させるかを指定します。

等間隔

変数は、ここで指定する等しいサイズの間隔で、定義された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することができます。

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

選択された変数に関して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、補助変数では使用できません。

欠損値の代入

選択された変数に関して欠損値のあるオブジェクトで、欠損値が代入されます。代入の方法を選択することができます。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、最小カテゴリ指標を持つ最頻値が使用されます。欠損値を付加カテゴリの同じ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、その変数に関して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見なされます。

カテゴリ回帰: オプション

「オプション」ダイアログでは、初期布置の形式の選択、反復基準と収束基準の指定、補助オブジェクトの選択、プロットのラベル付けの設定を行うことができます。

補助オブジェクト

補助オブジェクトとして処理したいオブジェクトを指定することができます。補助オブジェクトの数を入力します (またはケースの範囲を指定します)。補助オブジェクトに重みを付けることはできません (重みを指定しても無視されます)。

基準 計算における回帰の最大反復数を指定することができます。収束基準値を選択することもできます。最後の 2 回の反復間における全適合の差が収束値よりも小さくなるか、最大反復数に達すると、回帰の反復が終了します。

作図のラベル

変数ラベルと値ラベルを作図で使用するか、変数名と値を作図で使用するかを指定することができます。ラベルの最大長を指定することもできます。

初期布置

名義変数として処理される変数がない場合は、「数値」布置を選択します。1 つ以上の変数が名義変数として処理される場合は、「変量」布置を選択します。

または、1 つ以上の変数が順序尺度水準またはスプライン順序尺度水準を持っている場合、通常モデル適合アルゴリズムによって準最適解が取得されることがあります。検定対象のすべての有効な符号パターンとともに「多重システマティック スタート」を選択すると、常に最適解が取得されますが、データ・セットの順序変数とスプライン順序変数の数が増加するため、処理にかかる時間が急速に増加します。分散しきい値の損失のパーセントを指定することにより、検定パターン数を減らすことができます (しきい値が高くなるほど、除外される符号パターンが多くなります)。このオプションを使用すると、最適解の取得は保証されませんが、準最適解が取得される可能性は減少します。また、最適解が見つからない場合に、準最適解が最適解と大幅に異なるという可能性も減少します。多重システマティック・スタートが要求されると、スタートごとの回帰係数の符号が、外部の IBM SPSS Statistics データ・ファイル、または現行セッションのデータ・セットに書き込まれます。詳しくは、15 ページの『カテゴリー回帰: 保存』のトピックを参照してください。

多重システマティック・スタートを使用した以前の実行結果により、「回帰係数に固定符号を使用」することができます。符号 (1 と -1 で示されます) は、指定されたデータ・セットまたはファイルの行に含まれている必要があります。整数値の開始番号は、使用する符号が含まれているこのファイルの行のケース番号になります。

カテゴリー回帰: 正規化

方法 正規化の方法を使用すると、推定値を 0 に近づくように縮小して回帰係数の推定値の変動を低減することにより、モデルの予測誤差を改善することができます。Lasso と Elastic Net では、一部の係数推定値が正確に 0 に縮小されるため、変数選択の一形式が提供されます。正規化の方法が要求されると、正規化モデルと各ペナルティー係数値の係数が、外部の IBM SPSS Statistics データ・ファイル、または現行セッションのデータ・セットに書き込まれます。詳しくは、15 ページの『カテゴリー回帰: 保存』のトピックを参照してください。

Ridge 回帰

Ridge 回帰では、ペナルティー項 (平方した係数の合計にペナルティー係数を掛けた値) を導入することにより、係数が縮小されます。この係数の範囲は、0 (ペナルティーなし) から 1 までです。範囲と増分を指定すると、ペナルティーの「最良」の値が検索されます。

Lasso Lasso のペナルティー項は絶対係数の合計に基づいており、ペナルティー係数の仕様は Ridge 回帰の仕様に類似しています。ただし、Lasso ではより負荷の高い計算が実行されます。

Elastic Net

Elastic Net は、Lasso 回帰ペナルティーと Ridge 回帰ペナルティーを単純に組み合わせたもので、指定された値のグリッドから、「最良」の Lasso 回帰ペナルティー係数と

Ridge 回帰ペナルティー係数を検索します。Lasso 回帰ペナルティーと Ridge 回帰ペナルティーの特定ペアの場合、Elastic Net では Lasso よりも負荷の低い計算が実行されます。

Elastic Net プロット

Elastic Net 方式では、Ridge 回帰ペナルティーの値によって個別の正規化プロットが生成されます。「可能な Elastic Net プロットすべてを作成」を選択すると、指定された Ridge 回帰ペナルティーの最小値と最大値によって決定される範囲内のすべての値が使用されます。「Ridge ペナルティーの Elastic Net プロットを作成」では、最小値と最大値によって決定される範囲内の値のサブセットを指定することができます。ペナルティー値の数値を入力します (または値の範囲を指定します)。

正規化プロットを表示

これは、回帰係数と正規化ペナルティーを対比するプロットです。「最良」のペナルティー係数値の範囲を検索する場合に、回帰係数とその範囲においてどのように変化するかが示されます。

カテゴリー回帰: 出力

「出力」ダイアログでは、出力に表示する統計を選択することができます。

表 次の表が作成されます。

多重 R

R^2 、調整済み R^2 、および最適尺度法を考慮した調整済み R^2 が表示されます。

分散分析

このオプションでは、平方、平均平方、および F の回帰和と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、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 このオプションでは、次の 3 つのテーブルが表示されます。ベータ、ベータの標準誤差、 t 値、および有意性を持つ係数テーブル。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持つ係数 - 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を持つテーブル。

反復の記述

反復ごとに、アルゴリズム、重相関係数、および回帰誤差の開始値が表示されます。重相関係数の増加が、最初の反復からリストされます。

元の変数の相関

未変換の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

変換後の変数間の相関を示す行列が表示されます。

正規化されたモデルと係数

各正規化モデルのペナルティー値、 R^2 乗、および回帰係数が表示されます。再抽出方法または補助オブジェクト (検定ケース) を指定すると、予測誤差または検定 MSE も表示されます。

再抽出

再抽出手法により、モデルの予測誤差を推定することができます。

交差検証

交差検証では、サンプルが群と呼ばれる複数のサブサンプルに分割されます。分割後、カ

テゴリー回帰モデルが生成され、各サブサンプルからデータが順に除外されます。最初のモデルは、最初のサンプル群のケースを除くすべてのケースに基づき、2番目のモデルは、2番目のサンプル群のケースを除くすべてのケースに基づきます。3番目以降のモデルについても同様です。モデルの生成時に除外されたサブサンプルにそのモデルを適用することにより、モデルごとに予測誤差が推定されます。

.632 ブートストラップ

ブートストラップを使用すると、置換によってデータから観測値が無作為に抽出されるというプロセスが複数回繰り返されて、複数のブートストラップ・サンプルが取得されます。モデルは各ブートストラップ・サンプルに適合します。ブートストラップ・サンプル内に存在しないケースに適合モデルを適用することにより、各モデルの予測誤差が推定されます。

カテゴリー数量化

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、最頻値を示すテーブルが表示されます。

カテゴリー回帰: 保存

「保存」ダイアログでは、予測値、残差、変換後の値をアクティブなデータ・セットに保存したり、離散化データ、変換後の値、正規化モデルと係数、回帰係数の符号を、外部の IBM SPSS Statistics データ・ファイルまたは現行セッションのデータ・セットに保存したりすることができます。

- データ・セットは、現行セッションにおいて使用することができます。データ・セットをデータ・ファイルとして明示的に保存しない限り、データ・セットを以後のセッションで使用することはできません。データ・セット名は、変数の命名規則に準拠している必要があります。
- ファイル名またはデータ・セット名は、保存するデータのタイプごとに異なっている必要があります。

「正規化」ダイアログで正規化方法を選択するたびに、正規化モデルと係数が保存されます。デフォルトでは、固有の名前を持つ新規データ・セットが生成されますが、任意の名前を指定することも、外部ファイルに書き込むこともできます。

「オプション」ダイアログで、初期布置として多重システムティック・スタートを使用するたびに、回帰係数の符号が保存されます。デフォルトでは、固有の名前を持つ新規データ・セットが生成されますが、任意の名前を指定することも、外部ファイルに書き込むこともできます。

カテゴリー回帰: 変換プロット

「作図」ダイアログでは、変換プロットと残差プロットを生成する変数を指定することができます。

変換プロット

指定した各変数について、元のカテゴリー値に対するカテゴリー数量化がプロットされます。空白のカテゴリーは横軸に表示されますが、計算には影響しません。これらのカテゴリーは、数量化された値を結ぶ線の切れ目によって示されます。

残差プロット

指定した各変数について、カテゴリー指標と、カテゴリー指標に対するベータで乗算された最適カテゴリー数量化に対して、残差 (対象となる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残差) がプロットされます。

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・シンタックスを編集すると、カテゴリ回帰分析をカスタマイズすることができます。コマンド・シンタックス言語を使用すると、以下の処理を実行することもできます。

- 変換後の変数をアクティブなデータ・セットに保存するときに、それらの変数のルート名を指定する (SAVE サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」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、データの次元数を削減する一方で、複数のカテゴリ変数を同時に数量化します。カテゴリ主成分分析の略語は CATPCA (*categorical principal components analysis*) です。

主成分分析の目的は、元の変数の組数を減らし、元の変数が持っていた情報の大部分を表す無相関の成分の組にまとめることです。この手法は、多数の変数があるためにオブジェクト (被験者や単位) 間の関係を適切に解釈できない場合に最も有効です。次元数を削減することにより、多数の変数ではなく、少数の成分を解釈します。

標準の主成分分析では、数値変数間に線型関係があることを前提とします。一方、最適尺度法では、変数をさまざまなレベルで尺度化することができます。カテゴリ変数は、指定された次元数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るようになります。

例 カテゴリ主成分分析を使用して、仕事の満足度と、職種、部署、地域、移動量 (多、中、少) との関係視覚的に表現することができます。例えば、2 つの次元で分散のほとんどが示され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が、第 2 次元では部署と移動量が離れているような場合があります。また、高い仕事の満足度が中程度の移動量に関連していることがわかる場合もあります。

統計と作図

度数、欠損値、最適尺度水準、最頻値、変数と次元ごとに重心座標、ベクトル座標、および合計によって説明された分散、ベクトル数値変数の成分負荷、カテゴリ数値化と座標、反復の記述、変換後の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。

カテゴリ主成分分析データの考慮事項

データ

文字列変数の値は、英数字の昇順で常に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 未満の変数は、再割り当てを行うか定数を加算することにより、非欠損値変数にすることができます。

仮定 データには、3 つ以上の有効なケースを含める必要があります。この分析は、正の整数データに基づきます。離散化オプションは、小数値変数の値を「正規」分布に近いカテゴリにグループ化することによってその小数値変数を自動的に分類し、さらに、文字列変数の値を正の整数に自動的に変換します。他の離散化方式も指定することができます。

関連プロシージャ

すべての変数を数値レベルで尺度化した場合は、標準の主成分分析と同等になります。標準の線型主成分分析で変換された変数を使用することにより、代替の作図機能を使用することができます。

すべての変数に多重名義尺度水準が含まれている場合、カテゴリー主成分分析は、多重応答分析と同じになります。変数の組について調べる場合、カテゴリー（非線型）正準相関分析を使用する必要があります。

カテゴリー主成分分析：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することができます。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。また、分析変数の重みを設定することができます。

変数の重み付け

各変数の重みを定義することができます。正の整数を指定する必要があります。デフォルト値は1です。

最適尺度水準

各変数を数値化するための尺度水準を選択することができます。

- **スプライン順序:** 観測変数のカテゴリーの順序は、最適尺度変数に格納されます。カテゴリー・ポイントは、原点を通る直線（ベクトル）上に配置されます。変換の結果は、選択された次数の滑らかな単調な区分的多項式になります。ユーザーが指定した内側ノットの数と手続きによって決定された内側ノットの配置により、区分が指定されます。
- **スプライン名義:** 観測変数内の情報のうち、カテゴリー内のオブジェクトのグループだけが最適尺度変数に格納されます。観測変数のカテゴリーの順序は格納されません。カテゴリー・ポイントは、原点を通る直線（ベクトル）上に配置されます。変換の結果は、選択された次数の滑らかな（場合によっては単調な）区分的多項式になります。ユーザーが指定した内側ノットの数と手続きによって決定された内側ノットの配置により、区分が指定されます。
- **多重名義:** 観測変数内の情報のうち、カテゴリー内のオブジェクトのグループだけが最適尺度変数に格納されます。観測変数のカテゴリーの順序は格納されません。カテゴリー・ポイントは、特定のカテゴリー内のオブジェクトの重心に配置されます。多重とは、次元ごとに異なる数値化セットが取得されることを示します。
- **順序:** 観測変数のカテゴリーの順序は、最適尺度変数に格納されます。カテゴリー・ポイントは、原点を通る直線（ベクトル）上に配置されます。変換の結果は、スプライン順序変換よりも適合性は優れていますが、滑らかさでは劣ります。
- **名義:** 観測変数内の情報のうち、カテゴリー内のオブジェクトのグループだけが最適尺度変数に格納されます。観測変数のカテゴリーの順序は格納されません。カテゴリー・ポイントは、原点を通る直線（ベクトル）上に配置されます。変換の結果は、スプライン名義変換よりも適合性は優れていますが、滑らかさでは劣ります。
- **数値型:** カテゴリーは、順序付きで、等間隔（間隔レベル）のカテゴリーとして扱われます。観測変数のカテゴリーの順序とカテゴリー番号間の等間隔性は、最適尺度変数に格納されます。カテゴリー・ポイントは、原点を通る直線（ベクトル）上に配置されます。すべての変数が数値レベルの場合、この分析は標準の主成分分析に類似したものになります。

次数 多項式の次数。

内側ノット

内側ノットの数。

カテゴリー主成分分析：離散化

「離散化」ダイアログでは、変数の再割り当て方法を選択することができます。小数値変数は、特に指定されていない限り、ほぼ正規分布する7つのカテゴリー（または、その変数の持つ異なる値の数が7未満の

場合は、その数のカテゴリー)に分類されます。文字列変数は、英数字の昇順でカテゴリー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。文字列変数の離散化は、これらの整数に適用されます。デフォルトでは、その他の変数はそのままになります。これらの離散化された変数が分析で使用されます。

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ー数に再割り当てするか、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って離散化されます。

乗算 離散化された値の最小値が 1 となるように、変数の現在の値を標準化し、10 を掛けて丸めた値に定数を加算します。

グループ化

グループ化によって変数を離散化する場合、次のオプションを使用できます。

カテゴリー数

カテゴリー数を指定し、変数の値をカテゴリー全体にわたってほぼ正規分布させるか、または一様分布させるかを指定します。

等間隔

変数は、ここで指定する等しいサイズの間隔で、定義されたカテゴリー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。

カテゴリー主成分分析: 欠損値

「欠損値」ダイアログでは、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することができます。

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。

方略

欠損値の除外: 相関に関しては数量化後に代入

選択された変数に関して欠損値のあるオブジェクトは、その変数の分析には関係しません。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、元の変数の相関について、分析後に、その変数の最頻カテゴリー (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の場合、代入の方法を選択することができます。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ーの数量化で置き換えるには、「付加カテゴリー」を選択します。これにより、その変数に関して欠損値のあるオブジェクトは、同じ (付加) カテゴリーに属すると見なされます。

欠損値の代入

選択された変数に関して欠損値のあるオブジェクトで、欠損値が代入されます。代入の方法を選択することができます。欠損値を最頻カテゴリーで置き換えるには、「最頻値」を選択します。最頻値が複数ある場合、最小カテゴリー指標を持つ最頻値が使用されます。欠損値を付加カテゴリーの同じ数量化で置き換えるには、「付加カテゴリー」を選択します。これにより、その変数に関して欠損値のあるオブジェクトは、同じ (付加) カテゴリーに属すると見なされます。

この変数に欠損値のあるオブジェクトを除外

選択された変数に関して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、補助変数では使用できません。

カテゴリー主成分分析: 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、反復基準と収束基準の指定、正規化方法の選択、プロットのラベル付け方法の選択、補助オブジェクトの指定を制御することができます。

補助オブジェクト

補助オブジェクトにしたいオブジェクトのケース番号 (または、オブジェクト範囲の最初と最後のケース番号) を指定します。すべての補助オブジェクトを指定するまで、この処理を繰り返します。補助オブジェクトとして指定されているオブジェクトの場合、そのオブジェクトに対するケースの重み付けは無視されます。

正規化

オブジェクト・スコアと変数を正規化するための 5 つのオプションのうち、いずれかのオプションを指定することができます。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

このオプションを指定すると、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の座標は、成分負荷 (次元やオブジェクト・スコアなどの主成分との相関) です。このオプションは、主に変数間の相関を調べる場合に役立ちます。

オブジェクト主成分

このオプションを指定すると、オブジェクト間の距離が最適化されます。このオプションは、主にオブジェクト間の相違や類似性を調べる場合に役立ちます。

対称的

主にオブジェクトと変数との関係を調べる場合は、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と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

「ユーザー指定の値」フィールドで、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適用することができます。この方法は、目的に応じたバイプロットやトリプロットを作成する場合に役立ちます。

基準 計算における手続きの最大反復数を指定することができます。収束基準値を選択することもできます。最後の 2 回の反復間における全適合の差が収束値よりも小さくなるか、最大反復数に達すると、アルゴリズムの反復が終了します。

作図のラベル

変数ラベルと値ラベルを作図で使用するか、変数名と値を作図で使用するかを指定することができます。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

出力に表示する次元を制御することができます。

解のすべての次元の表示

解のすべての次元が散布図行列に表示されます。

次元数の制限

表示される次元を、プロットされたペアだけに制限します。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。最低次元の最小値は 1 で、最大値は、解の次元数から 1 を減算した数です。この範囲で最低次元を指定することができます。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。最高次元の最小値は 2 で、最大値は解の次元数です。この範囲で最高次元を指定することができます。この値は、次元ペアのプロットで使用される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。

回転 回転法を選択して、回転した結果を取得できます。

バリマックス

直交回転法の 1 つ。各成分の負荷量が高い変数の個数を最小化します。成分の解釈が単純化されます。

オブリミン

斜交 (非直交) 回転法の 1 つ。デルタが 0 のときに、成分の斜交度は最大になります。デルタが負で絶対値が大きくなるに従って、成分の斜交度は下がります。正の値では、追加の成分相関が許可されます。「デルタ」の値は、0.8 以下でなければなりません。

クォーティマックス

回転法の 1 つ。各変数を説明するために必要な成分の数を最小化します。観測された変数の解釈が単純化されます。

エカマックス

回転法の 1 つ。成分を単純化するバリマックス法と、変数を単純化するコーティマックス法を組み合わせたもの。成分に高い負荷を加える変数の数と、変数を説明するために必要な成分の数が最小化されます。

プロマックス

成分を相関付けることができる斜交 (非直交) 回転。直接オブリミン回転法よりも高速に計算できるため、大きなデータ セットの場合に役立ちます。許可される相関 (斜交) 量は、カッパ パラメータによって制御されます。「カッパ」の値は、1 以上 10,000 未満でなければなりません。

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。ファイル内の最初の変数には、最初の次元の座標を含める必要があります、2 番目の変数には、2 番目の次元の座標を含める必要があります (3 番目以降の変数についても同様です)。

なし 構成ファイルは使用されません。

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使用して、変数への適合を行います。適合対象の変数は、分析変数として選択する必要がありますが、布置は固定されているため、これらの変数は補助変数として扱われます (そのため、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」ダイアログでは、オブジェクト・スコア、判別測定、反復の記述、元の変数と変換後の変数の相関、選択された変数のカテゴリ数量化、選択された変数の記述統計を示すテーブルを生成することができます。

表

オブジェクト・スコア

マス (質量)、イナーシャ (慣性)、および寄与率を含むオブジェクト・スコアを表示します。以下のオプションが使用可能です。

オブジェクトスコア オプション: カテゴリを含める

選択した分析変数のカテゴリー指標が表示されます。

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を付けるための変数を選択することができます。

成分負荷

多重名義尺度水準が指定されていないすべての変数の成分負荷が表示されます。成分負荷をサイズで並べ替えできます。

反復の記述

反復ごとに、説明された分散、損失、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が表示されます。

変換された変数の相関

変換後の変数 (最適尺度化された変数) の相関行列とその行列の固有値が表示されます。

説明された分散

変数と次元ごとに、重心座標、ベクトル座標、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が表示されます。

カテゴリー数量化

選択した変数の次元ごとに、マス (質量) およびイナーシャ (慣性) を含むカテゴリー数量化 (座標) が表示されます。

注: マスとイナーシャを含む座標は、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。

記述統計

選択した変数について、度数、欠損値の数、最頻値が表示されます。

カテゴリー主成分分析: 保存

「保存」ダイアログでは、離散化データ、オブジェクト・スコア、変換後の変数を、外部の IBM SPSS Statistics データ・ファイルまたは現行セッションのデータ・セットに保存することができます。変換後の値とオブジェクト・スコアは、アクティブなデータ・セットに保存することもできます。

- データ・セットは、現行セッションにおいて使用することができます。データ・セットをデータ・ファイルとして明示的に保存しない限り、データ・セットを以後のセッションで使用することはできません。データ・セット名は、変数の命名規則に準拠している必要があります。
- ファイル名またはデータセット名は、保存するデータの型ごとに異なっている必要があります。
- オブジェクト・スコアや変換後の値をアクティブなデータ・セットに保存する場合、多重名義の次元の数を指定することができます。

カテゴリー主成分分析: オブジェクト・プロット

「オブジェクト プロット」ダイアログでは、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。

プロット

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・ポイントが、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プロットされます。

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・ポイントが、多重名義尺度水準の変数の重心と他の変数の成分負荷とともにプロットされます。

バイプロット変数とトリプロット変数

バイプロットとトリプロットに対してすべての変数を使用することも、変数のサブセットを選択することもできます。

オブジェクトのラベル付け

オブジェクトに、選択した変数のカテゴリーをラベルとして付けることも (「オプション」ダイアログで、カテゴリー指標値または値ラベルを選択できます)、ケース番号をラベルとして付けることもできます。「変数」を選択した場合、変数ごとに 1 つのプロットが作成されます。

カテゴリー主成分分析: カテゴリー・プロット

「カテゴリプロット」ダイアログでは、使用する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。

カテゴリー・プロット

選択された変数ごとに、重心座標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、カテゴリーはそのカテゴリー内のオブジェクトの重心に表示されます。その他のすべての尺度水準では、カテゴリーは原点を通るベクトル上に表示されます。

結合カテゴリー・プロット

これは、選択した各変数の重心座標とベクトル座標の単一プロットです。

変換プロット

最適カテゴリー数量化とカテゴリー指標を対比したプロットを表示します。多重名義尺度水準の変数に対して次元の数を指定することができます。次元ごとに 1 つのプロットが生成されます。選択した変数ごとに残差プロットを表示することもできます。

射影する重心

変数を選択し、その変数の重心を、選択した変数に射影することができます。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。このプロットを要求すると、射影された重心の座標を持つテーブルも表示されます。

カテゴリー主成分分析: 因子負荷プロット

「因子負荷プロット」ダイアログでは、プロットに含まれる変数、因子負荷プロット内の重心の表示、および説明された分散のプロットの表示を制御します。

説明された分散

各次元について、説明された分散のプロットを表示します。

成分負荷を表示

これを選択すると、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してすべての変数を使用することも、サブセットを選択することもできます。

重心を含む

多重名義尺度水準を持つ変数には成分負荷はありませんが、プロットにそれらの変数の重心を含めることができます。すべての多重名義変数を使用することも、サブセットを選択することもできます。

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と、カテゴリ主成分分析をカスタマイズすることができます。コマンド・シンタックス言語を使用すると、以下の処理を実行することもできます。

- 変換後の変数、オブジェクト・スコア、近似値をアクティブなデータ・セットに保存するときに、それらのルート名を指定する (SAVE サブコマンドを使用)。
- 各プロットのラベルの最大長を個別に指定する (PLOT サブコマンドを使用)。
- 残差プロットの個別の変数リストを指定する (PLOT サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

コレスポネンス分析

コレスポネンス分析の目的の 1 つは、低次元空間でコレスポネンス・テーブル内の 2 つの名義変数間の関係を記述することと、各変数のカテゴリ間関係を同時に記述することです。各変数について、プロットのカテゴリ・ポイント間の距離は、カテゴリ間関係を反映しています。類似したカテゴリは、相互に近い位置にプロットされます。また、1 つの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射影すると、変数間の関係が記述されます。

分割表の分析には、通常、行と列のプロファイルの検討と、カイ 2 乗統計を使用した独立性の検定が含まれます。ただし、プロファイルの数が非常に多くなる場合があるため、カイ 2 乗検定では従属構造は明らかになりません。クロス集計表手続きでは、関連度の指標と関連度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示すことはできません。

因子分析は、低次元空間で変数間の関係を記述するための標準的な手法です。ただし、因子分析では間隔データが必要で、変数の 5 倍の数の観測値が必要になります。一方、コレスポネンス分析は名義変数を前提とし、変数間の関係のほかに、各変数のカテゴリ間関係も記述することができます。また、コレスポネンス分析を使用して、正のコレスポネンス指標の任意のテーブルを分析することができます。

例: コレスポネンス分析を使用して、社員の категория と喫煙習慣の関係を図示することができます。例えば、喫煙に関して、中間管理職と秘書では異なるが、秘書と上級管理職では同じであることがわかる場合があります。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る場合もあります。

統計と作図: コレスポネンス指標、行と列のプロファイル、特異値、行と列のスコア、イナーシャ (慣性)、マス (質量)、行スコアと列スコアの信頼統計、特異値の信頼統計、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。

コレスポンデンス分析データの考慮事項

データ: 分析対象のカテゴリ変数は、名義変数として尺度化されます。集計データの場合、または度数以外のコレスポンデンス指標の場合、正の類似値を含む重み付け変数を使用します。一方、テーブル・データの場合は、シンタックスを使用してテーブルを読み込みます。

前提条件: この手続きで使用される次元の最大数は、アクティブな行と列のカテゴリの数と、同等性の制約条件の数によって異なります。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、最もカテゴリ数の少ない変数のカテゴリ数から 1 を減算した数になります。例えば、ある変数に 5 つのカテゴリがあり、他の変数には 4 つのカテゴリがある場合、次元の最大数は 3 になります。補助カテゴリは、アクティブなカテゴリではありません。例えば、ある変数に 5 つのカテゴリがあり、そのうちの 2 つが補助カテゴリで、他の変数には 4 つのカテゴリがある場合、次元の最大数は 2 になります。同等になるように制約されたカテゴリの組は、すべて 1 つのカテゴリとして扱う必要があります。例えば、ある変数に 5 つのカテゴリがあり、そのうち 3 つが同等になるように制約されている場合、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う必要があります。この場合、2 つのカテゴリは制約されておらず、第 3 のカテゴリが、制約された 3 つのカテゴリに対応します。最大値を超える次元数を指定しても、最大値が使用されます。

関連手続き: 3 つ以上の変数を使用する場合は、多重応答分析を使用します。変数を順序変数として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。

コレスポンデンス分析を実行するには

1. メニューから次の項目を選択します。

「分析」 > 「次元分解」 > 「コレスポンデンス分析...」

2. 行変数を選択します。
3. 列変数を選択します。
4. 変数の範囲を定義します。
5. 「OK」をクリックします。

コレスポンデンス分析: 行範囲の定義

行変数の範囲を定義する必要があります。最小値と最大値は、整数で指定する必要があります。小数点以下のデータ値は、分析では切り捨てられます。指定された範囲内に含まれていないカテゴリ値は、分析では無視されます。

すべてのカテゴリについて、初期には制約がなくアクティブになっています。行カテゴリが他の行カテゴリと同等になるように制約することも、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

各カテゴリについて、同じスコアが必要です。生成されたカテゴリの順序が望ましくない場合や、直感的でない場合は、同等性の制約条件を使用します。同等になるように制約できる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 を減算した数です。異なる同等性の制約条件をカテゴリの組に適用するには、シンタックスを使用します。例えば、シンタックスを使用して、カテゴリ 1 と 2、カテゴリ 3 と 4 がそれぞれ同等になるように制約することができます。

カテゴリーは補助

補助カテゴリーは分析には影響しませんが、アクティブなカテゴリーによって定義された空間に表示されます。補助カテゴリーは、次元の定義には無関係です。補助行カテゴリーの最大数は、行カテゴリーの総数から 2 を減算した数です。

コレスポンデンス分析: 列範囲の定義

列変数の範囲を定義する必要があります。最小値と最大値は、整数で指定する必要があります。小数点以下のデータ値は、分析では切り捨てられます。指定された範囲内に含まれていないカテゴリー値は、分析では無視されます。

すべてのカテゴリーについて、初期には制約がなくアクティブになっています。列カテゴリーが他の列カテゴリーと同等になるように制約することも、列カテゴリーを補助カテゴリーとして定義することもできます。

カテゴリーは等しい

各カテゴリーについて、同じスコアが必要です。生成されたカテゴリーの順序が望ましくない場合や、直感的でない場合は、同等性の制約条件を使用します。同等になるように制約できる列カテゴリーの最大数は、アクティブな列カテゴリーの総数から 1 を減算した数です。異なる同等性の制約条件をカテゴリーの組に適用するには、シンタックスを使用します。例えば、シンタックスを使用して、カテゴリー 1 と 2、カテゴリー 3 と 4 がそれぞれ同等になるように制約することができます。

カテゴリーは補助

補助カテゴリーは分析には影響しませんが、アクティブなカテゴリーによって定義された空間に表示されます。補助カテゴリーは、次元の定義には無関係です。補助列カテゴリーの最大数は、列カテゴリーの総数から 2 を減算した数です。

コレスポンデンス分析: モデル

「モデル」ダイアログでは、次元の数、距離測度、標準化の方法、正規化の方法を指定することができます。

解の次元

次元の数を指定します。通常、変動の大部分を記述するために必要な最小限の次元数を選択します。次元の最大数は、分析で使用されるアクティブなカテゴリーの数と同等性の制約条件によって異なります。次元の最大数は、以下に示す数のうち小さい方の数になります。

- アクティブな行カテゴリーの数から、同等になるように制約された行カテゴリー数を減算し、制約された行カテゴリーの組の数を加算した数。
- アクティブな列カテゴリーの数から、同等になるように制約された列カテゴリー数を減算し、制約された列カテゴリーの組の数を加算した数。

距離測度

コレスポンデンス・テーブルの行間と列間の距離測度を選択することができます。以下のいずれかを選択します。

カイ 2 乗

重み付けプロファイルの距離を使用します。ここでの重みは、行または列のマス (質量) です。標準コレスポンデンス分析では、この測度が必要になります。

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

以下のいずれかを選択します。

行と列の平均値を取り除く

行と列の両方が中心化されます。標準コレスポネンス分析では、この方法が必要になります。

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、行の周辺が同等化されます。

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、列の周辺が同等化されます。

正規化

以下のいずれかを選択します。

対称的

各次元において、行スコアは、列スコアの加重平均を対応する特異値で除算した数値です。列スコアは、行スコアの加重平均を対応する特異値で除算した数値です。2 つの変数のカテゴリー間の差異または類似性を調べる場合は、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い、コレスポネンス・テーブルの距離の近似値になります。2 つの変数間の差ではなく、一方の変数または両方の変数のカテゴリー間の差を調べる場合は、この方法を使用します。

行主成分

行ポイント間の距離は、選択された距離測度に従い、コレスポネンス・テーブルの距離の近似値になります。行スコアは、列スコアの加重平均です。行変数のカテゴリー間の差異または類似性を調べる場合は、この方法を使用します。

列主成分

列ポイント間の距離は、選択された距離測度に従い、コレスポネンス・テーブルの距離の近似値になります。列スコアは、行スコアの加重平均です。列変数のカテゴリー間の差異または類似性を調べる場合は、この方法を使用します。

ユーザー指定

-1 から 1 までの値を指定する必要があります。値 -1 は、列主成分に対応します。値 1 は、行主成分に対応します。値 0 は、対称に対応します。その他のすべての値では、行スコアと列スコアの両方において、イナーシャ (慣性) がさまざまな度合いで適用されます。この方法は、目的に応じたバイプロットを作成する場合に役立ちます。

コレスポネンス分析: 統計

「統計」ダイアログでは、生成する数値出力を指定することができます。

コレスポネンス テーブル

行と列の周辺合計を含む、入力変数のクロス集計表。

行ポイントの概要

行カテゴリごとの、スコア、マス (質量)、イナーシャ (慣性)、次元のイナーシャに対する寄与率、ポイントのイナーシャに対する次元の寄与率。

列ポイントの概要

列カテゴリごとの、スコア、マス (質量)、イナーシャ (慣性)、次元のイナーシャに対する寄与率、ポイントのイナーシャに対する次元の寄与率。

コレスポンデンス テーブルの転置

最初の次元のスコアに従って行と列が昇順となるように再編成されたコレスポンデンス・テーブル。必要に応じて、転置されたテーブルを作成する最大次元数を指定することができます。1 から、指定された数までの各次元について、転置されたテーブルが作成されます。

行のプロファイル

行カテゴリごとの、列変数のカテゴリに対する分布。

列のプロファイル

列カテゴリごとの、行変数のカテゴリに対する分布。

信頼統計量の適用先

行ポイント

すべての非補助行ポイントについての標準偏差と相関が含まれます。

列ポイント

すべての非補助列ポイントについての標準偏差と相関が含まれます。

コレスポンデンス分析: 作図

「作図」ダイアログでは、作成する図を指定することができます。

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。以下の散布図を使用できます。

バイプロット

行ポイントと列ポイントの結合プロットの行列を作成します。主成分の正規化を選択した場合、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。

散布図の ID ラベル幅

必要に応じて、ポイントにラベルを付けるときに使用する値ラベルの文字数を指定することができます。この値には、20 以下の負でない整数を指定する必要があります。

線作図

選択された変数のすべての次元のプロットを作成します。以下の折れ線グラフを使用できます。

変換行カテゴリ

対応する行スコアに対して、元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

対応する列スコアに対して、元の列カテゴリ値のプロットを作成します。

線作図の ID ラベル幅

必要に応じて、カテゴリ軸にラベルを付けるときに使用する値ラベルの文字数を指定することができます。この値には、20 以下の負でない整数を指定する必要があります。

次元のプロット

出力に表示する次元を制御することができます。

解のすべての次元の表示

解のすべての次元が散布図行列に表示されます。

次元数の制限

表示される次元を、プロットされたペアだけに制限します。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。最低次元の最小値は 1 で、最大値は、解の次元数から 1 を減算した値です。この範囲内で、最低次元を指定することができます。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。最高次元値の最小値は 2 で、最大値は解の次元数です。この範囲内で、最高次元値を指定することができます。この値は、次元ペアのプロットで使用される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と、コレスポンデンス分析をカスタマイズすることができます。コマンド・シンタックス言語を使用すると、以下の処理を実行することもできます。

- ケースごとのデータを使用する代わりに、テーブル・データを入力として指定する (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ポイントにラベルを付けるために使用される値ラベルの文字数を指定する (PLOT サブコマンドを使用)。
- 各種の折れ線グラフのポイントにラベルを付けるために使用される値ラベルの文字数を指定する (PLOT サブコマンドを使用)。
- 行列データ・ファイルに行スコアと列スコアの行列を書き込む (OUTFILE サブコマンドを使用)。
- 特異値とスコア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)。
- 同等にする複数の組のカテゴリを指定する (EQUAL サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

多重コレスポンデンス分析

多重応答分析では、同じカテゴリ内のオブジェクトは距離的に近く、異なるカテゴリのオブジェクトは距離的に遠くなるような形で、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り、名義 (カテゴリ) データが数量化されます。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づくように配置されます。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。変数により、同じカテゴリ内のオブジェクトが同じサブグループに分類される場合、それらの変数は等質な変数と見なされます。

例: 多重応答分析を使用して、職種、人種、性別間の関係を図示することができます。これにより、例えば、人種と性別では人々の間に差異があるが、職種では差異がないことなどがわかります。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などもわかります。

統計と作図: オブジェクト・スコア、判別測定、反復の記述、元の変数と変換後の変数の相関、カテゴリー数量化、記述統計、オブジェクト・ポイントのプロット、バイプロット、カテゴリー・プロット、結合カテゴリー・プロット、変換プロット、および判別測定プロット。

多重応答分析データの考慮事項

データ: 文字列変数の値は、英数字の昇順で常に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 未満の変数は、再割り当てを行うか定数を加算することにより、非欠損値変数にすることができます。

仮定: すべての変数が、多重名義尺度水準を持ちます。データには、3 つ以上の有効なケースを含める必要があります。この分析は、正の整数データに基づきます。離散化オプションは、小数値変数の値を正規分布に近いカテゴリーにグループ化することによってその小数値変数を自動的に分類し、さらに、文字列変数の値を正の整数に自動的に変換します。他の離散化方式も指定することができます。

関連プロシージャー: 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類似したものになります。順序特性または数値特性を持っている変数の場合は、カテゴリー主成分分析を使用する必要があります。変数の組について調べる場合は、非線型正準相関分析を使用する必要があります。

多重応答分析を行うには

1. 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...」

2. 「すべての変数が多重名義」を選択します。

3. 「単一グループ」を選択します。

4. 「定義」をクリックします。

5. 2 つ以上の分析変数を選択し、解の次元の数を指定します。

6. 「OK」をクリックします。

必要に応じて、検出された解に適合させる補助変数を指定することも、プロットのラベル付け変数を指定することもできます。

多重応答分析: 変数の重みの定義

分析変数の重みを設定することができます。

変数の重み

各変数の重みを定義することができます。正の整数を指定する必要があります。デフォルト値は 1 です。

多重応答分析: 離散化

「離散化」ダイアログでは、変数の再割り当て方法を選択することができます。小数値変数は、特に指定されていない限り、ほぼ正規分布する 7 つのカテゴリー (または、その変数の持つ異なる値の数が 7 未満の場合は、その数のカテゴリー) に分類されます。文字列変数は、英数字の昇順でカテゴリー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。文字列変数の離散化は、これらの整数に適用されます。デフォルトでは、その他の変数はそのままになります。これらの離散化された変数が分析で使用されます。

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ー数に再割り当てするか、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って離散化されます。

乗算 離散化された値の最小値が 1 となるように、変数の現在の値を標準化し、10 を掛けて丸めた値に定数を加算します。

グループ化

グループ化によって変数を離散化する場合、次のオプションを使用できます。

カテゴリー数

カテゴリー数を指定し、変数の値をカテゴリー全体にわたってほぼ正規分布させるか、または一様分布させるかを指定します。

等間隔

変数は、ここで指定する等しいサイズの間隔で、定義されたカテゴリー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: 欠損値

「欠損値」ダイアログでは、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することができます。

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。

方略

欠損値の除外: 相関に関しては数量化後に代入

選択された変数に関して欠損値のあるオブジェクトは、その変数の分析には関係しません。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、元の変数の相関について、分析後に、その変数の最頻カテゴリー (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の場合、代入の方法を選択することができます。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ーの数量化で置き換えるには、「付加カテゴリー」を選択します。これにより、その変数に関して欠損値のあるオブジェクトは、同じ (付加) カテゴリーに属すると見なされます。

欠損値の代入

選択された変数に関して欠損値のあるオブジェクトで、欠損値が代入されます。代入の方法を選択することができます。欠損値を最頻カテゴリーで置き換えるには、「最頻値」を選択します。最頻値が複数ある場合、最小カテゴリー指標を持つ最頻値が使用されます。欠損値を付加カテゴリーの同じ数量化で置き換えるには、「付加カテゴリー」を選択します。これにより、その変数に関して欠損値のあるオブジェクトは、同じ (付加) カテゴリーに属すると見なされます。

この変数に欠損値のあるオブジェクトを除外

選択された変数に関して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、補助変数では使用できません。

多重応答分析: オプション

「オプション」ダイアログでは、初期布置の選択、反復基準と収束基準の指定、正規化方法の選択、プロットのラベル付け方法の選択、補助オブジェクトの指定を行うことができます。

補助オブジェクト

補助オブジェクトにしたいオブジェクトのケース番号 (または、オブジェクト範囲の最初と最後のケース番号) を指定します。すべての補助オブジェクトを指定するまで、この処理を繰り返します。補助オブジェクトとして指定されているオブジェクトの場合、そのオブジェクトに対するケースの重み付けは無視されます。

正規化

オブジェクト・スコアと変数を正規化するための 5 つのオプションのうち、いずれかのオプションを指定することができます。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

このオプションを指定すると、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の座標は、成分負荷 (次元やオブジェクト・スコアなどの主成分との相関) です。このオプションは、主に変数間の相関を調べる場合に役立ちます。

オブジェクト主成分

このオプションを指定すると、オブジェクト間の距離が最適化されます。このオプションは、主にオブジェクト間の相違や類似性を調べる場合に役立ちます。

対称的

主にオブジェクトと変数との関係を調べる場合は、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と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

「ユーザー指定の値」フィールドで、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適用することができます。この方法は、目的に応じたバイプロットやトリプロットを作成する場合に役立ちます。

基準 計算における手続きの最大反復数を指定することができます。収束基準値を選択することもできます。最後の 2 回の反復間における全適合の差が収束値よりも小さくなるか、最大反復数に達すると、アルゴリズムの反復が終了します。

作図のラベル

変数ラベルと値ラベルを作図で使用するか、変数名と値を作図で使用するかを指定することができます。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

出力に表示する次元を制御することができます。

解のすべての次元の表示

解のすべての次元が散布図行列に表示されます。

次元数の制限

表示される次元を、プロットされたペアだけに制限します。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。最低次元の最小値は 1 で、最大値は、解の次元数から 1 を減算した数です。この範囲で最低次元を指定することができます。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。最高次元の最小値は 2

で、最大値は解の次元数です。この範囲で最高次元を指定することができます。この値は、次元ペアのプロットで使用される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。ファイル内の最初の変数には、最初の次元の座標を含める必要があります、2番目の変数には、2番目の次元の座標を含める必要があります (3番目以降の変数についても同様です)。

なし 構成ファイルは使用されません。

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使用して、変数への適合を行います。適合対象の変数は、分析変数として選択する必要がありますが、布置は固定されているため、これらの変数は補助変数として扱われます (そのため、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: 出力

「出力」ダイアログでは、オブジェクト・スコア、判別測定、反復の記述、元の変数と変換後の変数の相関、選択された変数のカテゴリ-数量化、選択された変数の記述統計を示すテーブルを生成することができます。

表

オブジェクト・スコア

マス (質量)、イナーシャ (慣性)、および寄与率を含むオブジェクト・スコアを表示します。以下のオプションが使用可能です。

オブジェクトスコア オプション: カテゴリを含める

選択した分析変数のカテゴリ-指標が表示されます。

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を付けるための変数を選択することができます。

判別測定

変数と次元ごとに判別測定が表示されます。

反復の記述

反復ごとに、説明された分散、損失、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が表示されます。

変換された変数の相関

変換後の変数 (最適尺度化された変数) の相関行列とその行列の固有値が表示されます。

カテゴリ-数量化と寄与率

選択した変数の次元ごとに、マス (質量)、イナーシャ (慣性)、および寄与率を含むカテゴリ-数量化 (座標) が表示されます。

注: マスとイナーシャを含む座標と寄与率が、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

選択した変数について、度数、欠損値の数、最頻値が表示されます。

多重応答分析: 保存

「保存」ダイアログでは、離散化データ、オブジェクト・スコア、変換後の変数を、外部の IBM SPSS Statistics データ・ファイルまたは現行セッションのデータ・セットに保存することができます。変換後の値とオブジェクト・スコアは、アクティブなデータ・セットに保存することもできます。

- データ・セットは、現行セッションにおいて使用することができます。データ・セットをデータ・ファイルとして明示的に保存しない限り、データ・セットを以後のセッションで使用することはできません。データ・セット名は、変数の命名規則に準拠している必要があります。
- ファイル名またはデータセット名は、保存するデータの型ごとに異なっている必要があります。
- オブジェクト・スコアや変換後の値をアクティブなデータ・セットに保存する場合、多重名義の次元の数を指定することができます。

多重応答分析: オブジェクト・プロット

「オブジェクト プロット」ダイアログでは、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。

プロット

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 (バイプロット)

オブジェクト・ポイントが変数の重心とともにプロットされます。

バイプロット変数

バイプロットに対してすべての変数を使用することも、サブグループを選択することもできます。

オブジェクトのラベル付け

オブジェクトに、選択した変数のカテゴリーをラベルとして付けることも (「オプション」ダイアログで、カテゴリー指標値または値ラベルを選択できます)、ケース番号をラベルとして付けることもできます。「変数」を選択した場合、変数ごとに 1 つのプロットが作成されます。

多重応答分析: 変数プロット

「変数プロット」ダイアログでは、使用するプロットの種類とプロットする変数を指定することができます。

カテゴリー・プロット

選択した変数ごとに、重心座標のプロットが作図されます。カテゴリーは、そのカテゴリーに属するオブジェクトの重心に配置されます。

結合カテゴリー・プロット

これは、選択した各変数の重心座標の単一プロットです。

変換プロット

最適カテゴリー数量化とカテゴリー指標を対比したプロットを表示します。次元数を指定することができます。次元ごとに 1 つのプロットが作成されます。選択した変数ごとに残差プロットを表示することもできます。

判別測定

選択した変数について、判別測定の単一のプロットが作成されます。

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と、多重応答分析をカスタマイズすることができます。このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 変換後の変数、オブジェクト・スコア、近似値をアクティブなデータ・セットに保存するときに、それらのルート名を指定する (SAVE サブコマンドを使用)。
- 各プロットのラベルの最大長を個別に指定する (PLOT サブコマンドを使用)。
- 残差プロットの個別の変数リストを指定する (PLOT サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」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 1 組の近傍度の構造を検出するための手法です。この処理は、概念上の低次元空間内のポイント間の距離が、指定された類似度 (または非類似度) に可能な限り近くなるように、この空間内の特定の場所に観測値を割り当てることにより、実行されます。これにより、その低次元空間内のオブジェクトの最小二乗表現が得られます。この表現は、多くの場合、データをより深く理解するのに役立ちます。

例: 多次元尺度法は、認知的な関係を調べる場合に非常に便利です。例えば、製品のイメージを検討する場合は、調査を実施して、競合他社の製品と自社製品がどの程度の類似性 (または近傍度) を持つと考えられているかを示すデータ・セットを取得することができます。この近傍度と独立変数 (価格など) を使用すると、製品に対するユーザーの見方に対してどの変数が重要なかを判断し、その結果に応じて自社製品のイメージを調整することができます。

統計と作図: 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、個別空間の重み、個別空間、変換後の近傍度、変換後の独立変数、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。

多次元尺度法データの考慮事項

データ: データは、近傍行列の形式で提供することも、近傍行列に変換された変数の形式で提供することもできます。行列は、列形式にすることも、複数列にまたがる形式にすることもできます。近傍度は、比率、間隔、順序、またはスプライン尺度水準で処理することができます。

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、オブジェクトの数から 1 を減算した数を超えてはいけません。多重ランダム・スタートと組み合せた場合、次元数の削減は省略されます。ソースを 1 つだけ指定した場合、すべてのモデルが同一モデルに等しくなるため、分析のデフォルト・モデルは同一モデルになります。

関連プロシージャー: 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分析と同等になります。

多次元尺度法を使用するには

1. メニューから次の項目を選択します。

「分析」 > 「多次元尺度法」 > 「多次元尺度法 (PROXSCAL)...

これにより、「データ形式」ダイアログ・ボックスが開きます。

2. データの形式を指定します。

データ形式

データが近傍度で構成されているか、データから近傍度を作成するかを指定します。

ソース数

使用するデータが近傍度の場合は、その近傍度のソースが単一か複数かを指定します。

単一ソース

近傍度のソースが 1 つだけの場合、近傍度を含むデータ・セットの形式が、複数の列にまたがる 1 つの行列なのか、各近傍度の行と列を識別する 2 つの個別の変数を持つ単一の列なのかを指定します。

The proximities are in a matrix across columns (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存在します。「複数列行列の近接」ダイアログ・ボックスに遷移します。

The proximities are in a single column (- 演算子)

近接行列を 1 つの列 (変数) に集約します。各セルの行および列を識別する 2 つの追加の変数が必要です。「単一系列の近接」ダイアログ・ボックスに遷移します。

多重ソース: 近接度のソースが複数ある場合、近接度を含むデータ・セットの形式が、複数の列にまたがる積み重ね行列なのか、列ごとに 1 つのソースを持つ複数の列なのか、または単一の列なのかを指定します。

- *The proximities are in stacked matrices across columns* (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存在し、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ります。「複数列行列の近接」ダイアログ・ボックスに遷移します。
- *The proximities are in columns, one source per column* (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 (変数) に集約されています。各セルの行および列を識別する 2 つの追加の変数が必要です。「複数列の近接」ダイアログ・ボックスに遷移します。
- *The proximities are stacked in a single column* (近接が単一の列内に積み重ねられている). 近接行列を 1 つの列 (変数) に集約します。各セルの行、列、およびソースを識別する 3 つの追加の変数が必要です。「単一系列の近接」ダイアログ・ボックスに遷移します。

複数列行列の近接

「変数」ダイアログで、単一ソースまたは複数ソースの行列データ・モデルの近傍度を選択した場合、以下の手順を実行します。

1. 3 つ以上の近接変数を選択します (リスト内の変数の順序は、近接度の列の順序に一致している必要があります)。
2. 必要に応じて、近接変数の数に等しい数の重み付け変数を選択します (重み付け変数の順序は、重み付け対象の近接変数の順序に一致している必要があります)。
3. 複数のソースがある場合、必要に応じて、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の数を乗算した数値に一致している必要があります)。

また、多次元尺度法のモデルの定義、共通空間に対する制限の設定、収束基準の設定、使用する初期布置の指定、プロットと出力の選択を行うことができます。

複数列の近接

「変数」ダイアログで複数ソースの複数列モデルを選択した場合、以下の手順を実行します。

1. 2 つ以上の近接変数を選択します (各変数は、個別ソースからの近接度の行列であることが前提です)。
2. 行変数を選択して、各近接変数内の近接度の行位置を定義します。
3. 列変数を選択して、各近接変数内の近接度の列位置を定義します (行と列の指定がない近接行列のセルは、欠損値として扱われます)。
4. 必要に応じて、近接変数の数に等しい数の重み付け変数を選択します。

また、多次元尺度法のモデルの定義、共通空間に対する制限の設定、収束基準の設定、使用する初期布置の指定、プロットと出力の選択を行うことができます。

単一系列の近接

「変数」ダイアログで、単一ソースまたは複数ソースの単一系列モデルを選択した場合、以下の手順を実行します。

1. 近接変数を選択します (1 つ以上の近接行列であることが前提です)。
2. 行変数を選択して、近接変数内の近接度の行位置を定義します。
3. 列変数を選択して、近接変数内の近接度の列位置を定義します。
4. 複数のソースがある場合、ソース変数を選択します (各ソースについて、行と列の指定がない近接行列のセルは、欠損値として扱われます)。
5. 必要に応じて、重み付け変数を選択します。

また、多次元尺度法のモデルの定義、共通空間に対する制限の設定、収束基準の設定、使用する初期布置の指定、プロットと出力の選択を行うことができます。

データからの近接度の作成

「変数」ダイアログで、データから近傍度を作成することを選択した場合は、以下の手順を実行します。

1. 変数間の距離を作成する場合は、3 つ以上の変数を選択します。これらの変数を使用して、1 つの近傍行列 (複数のソースがある場合は複数の近傍行列) が作成されます。ケース間の距離を作成する場合は、1 つの変数のみ必要です。
2. 複数のソースがある場合、ソース変数を選択します。
3. 必要に応じて、近接度を作成するための測定方法を選択します。

また、多次元尺度法のモデルの定義、共通空間に対する制限の設定、収束基準の設定、使用する初期布置の指定、プロットと出力の選択を行うことができます。

データからの尺度の作成

多次元尺度法では、非類似度データを使用して尺度解が作成されます。使用するデータが多変量データ (測定変数の値) の場合は、多次元尺度法の解を計算するために、非類似度データを作成する必要があります。データからの非類似度の尺度の作成について、詳細を指定することができます。

測定方法

分析で使用する非類似度の測定方法を指定することができます。データの種類に応じて、「測定方法」グループからいずれかのオプションを選択し、その測定方法の種類に対応するドロップダウン・リストから測定方法を 1 つ選択します。選択肢は次のとおりです。

- 区間 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」、「パターンの違い」、「分散」、または「Lance-Williams」。

値の変換

大きく異なる尺度で変数が測定されている場合などは、近傍度を計算する前に、値を標準化する必要があります (2 値データの場合は該当しません)。このような場合は、「標準化」ドロップダウン・リストで標準化の方法を選択してください (標準化が不要な場合は、「なし」を選択します)。

距離行列の作成

分析の単位を選択できます。「変数間」または「ケース間」のいずれかを選択してください。

多次元尺度法: モデルの定義

「モデル」ダイアログでは、尺度モデル、尺度モデル次元の最小数と最大数、近傍行列の構造、近傍度で使用する変換を指定することができます。また、近傍度を各ソース内で個別に変換するか、ソース上で無条件に変換するかを指定することができます。

尺度モデル

以下のオプションから選択します。

同一 すべてのソースが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。各ソースは個別空間を持ち、その個別空間では共通空間のすべての次元に異なる重みが付けられます。

一般化ユークリッド

このモデルは個人差モデルです。各ソースは個別空間を持ち、その個別空間は共通空間を回転したものと同じで、各次元に異なる重みが付けられます。

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。この階数には、1 以上で、最大次元数未満の数値を指定する必要があります。

近接変換

以下のオプションから選択します。

比データ

変換後の近傍度は、元の近傍度に比例します。このオプションは、正の値を持つ近傍度の場合のみ選択することができます。

区間 変換後の近傍度は、元の近傍度に切片項を加算したものに比例します。この切片項により、変換後のすべての近傍度が正の値になります。

順序 変換後の近傍度は元の近傍度と同じ順序になります。近傍度の順序を固定したままにするか、固定の解除を許可するかを指定します。

スプライン

変換後の近傍度は、元の近傍度を滑らかな非減少区分的多項式で変換したものになります。多項式の次数と内側ノットの数を選択します。

形状 近傍行列の下三角部分と上三角部分のどちらから近傍度を取得するかを指定します。完全行列を使

用することもできます。その場合、上三角部分と下三角部分の重み付けの合計が分析されます。いずれの場合も、指定した部分だけが使用されますが、対角成分を含む完全行列を指定する必要があります。

変換の適用

各ソース内の近傍度だけを相互に比較するか、ソース上で無条件に比較を行うかを指定します。

近傍 類似度と非類似度のどちらの指標を近傍行列に含めるかを指定します。

次元 デフォルトでは、解は 2 つの次元で計算されます (最小値 = 2、最大値 = 2)。最小値は 1 で、最大値は、オブジェクトの数から 1 を減算した数です。この範囲内の整数を指定することができます。ただし、最小値には最大値以下の値を指定する必要があります。この手続きでは、最大次元で解が計算され、その後、最小値に達するまで段階的に次元数が減っていきます。

多次元尺度法: 制限

「制限」ダイアログでは、共通空間に対する制限を設定することができます。

共通空間に対する制限

使用する制限の種類を指定します。

制限なし

共通空間に対する制限は設定されません。

いくつかの座標を固定

選択した最初の変数には最初の次元のオブジェクトの座標が格納され、2 番目の変数には 2 番目の次元の座標が格納されます (3 番目以降についても同様です)。欠損値は、次元の座標が空いていることを示します。選択した変数の数と要求した次元の最大数は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にな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、間隔変換、名義変換、順序変換、またはスプライン変換を指定します。いずれの場合も、各変数のケースの数とオブジェクトの数は一致している必要があります。

多次元尺度法: オプション

「オプション」ダイアログでは、初期布置の形式を選択したり、反復基準と収束基準を指定したり、標準アルゴリズムやより速いアルゴリズムを選択したりすることができます。

初期布置

以下のいずれかを選択します。

シンプレックス

各オブジェクトは、最大次元で相互に等距離となる位置に配置されます。反復が 1 回実行されてこの高次元の布置が改善され、次に次元数縮小操作が実行されて、「モデル」ダイアログで指定した最大次元数を持つ初期布置が取得されます。

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。

ユーザー指定

(「ユーザーの指定による布置」セクションで) 独自の初期布置の座標を格納する変数を選択します。最初の変数は次元 1 の座標に対応し、2 番目の変数は次元 2 の座標に対応するというように、選択した変数の数と指定した次元の最大数が等しくなっている必要があります。各変数のケースの数は、オブジェクトの数と一致している必要があります。

多重ランダム スタート

複数の布置が無作為に選択され、正規化された最小原ストレスを持つ布置が初期布置として使用されます。

反復基準

反復基準の値を指定します。

ストレス収束

連続する正規化された原ストレス値の差が、ここで指定した数値よりも小さくなると、このアルゴリズムは反復を停止します。この数値は、0.0 から 1.0 までの範囲で指定する必要があります。

最小ストレス

正規化された原ストレス値がここで指定した数値よりも小さくなると、このアルゴリズムが停止します。この数値は、0.0 から 1.0 までの範囲で指定する必要があります。

最大反復回数

上記の基準がどれも満たされなかった場合、ここで指定した回数の反復処理がこのアルゴリズムによって実行されます。

より速いアルゴリズムを使用

より速いアルゴリズムによってアルゴリズムが高速化されます。このアルゴリズムは、同一モデル以外のモデルの場合や制限がある場合は使用できません。

多次元尺度法: 作図、バージョン 1

「作図」ダイアログでは、作成するプロットを指定することができます。このトピックでは、「複数列の近傍」データ形式を使用する場合の「作図」ダイアログについて説明します。「個別空間の重み」、「変換前と変換後の近傍」、および「変換後の近傍と距離」の各プロットについて、プロットを作成する必要があるソースを指定します。使用可能なソースのリストは、「変数」ダイアログの近傍変数のリストです。

プロット

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が作成されます。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のみ作成されます。

共通空間

共通空間の座標の散布図行列が表示されます。

個別空間

ソースごとに、個別空間の座標が散布図行列内に表示されます。「モデル」ダイアログ・ボックスでいずれかの個人差モデルを指定した場合のみ、このプロットを使用できます。

個別空間の重み

個別空間の重みの散布図が作成されます。「モデル」ダイアログ・ボックスでいずれかの個人差モデルを指定した場合のみ、このプロットを使用できます。重み付きユークリッド・モデルの場合、重みがプロット内に出力され、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・モデルの場合、次元ごとに 1 つのプロットが作成され、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が作成されますが、個別空間の次元数は削減されます。

変換前と変換後の近接

元の近傍度と変換後の近傍度を対比するプロットが作成されます。

変換後の近接と距離

変換後の近傍度と距離を対比する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作成されます。

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: 作図、バージョン 2

「作図」ダイアログでは、作成するプロットを指定することができます。このトピックでは、データ形式が「複数列の近傍」以外の場合の「作図」ダイアログについて説明します。「個別空間の重み」、「変換前と変換後の近傍」、および「変換後の近傍と距離」の各プロットについて、プロットを作成する必要があるソースを指定します。ソース番号には、メイン・ダイアログ・ボックスで指定したソース変数の値を入力する必要があります。また、1 からソース数までの範囲で入力する必要があります。

多次元尺度法: 出力

「出力」ダイアログでは、表示される出力量を制御したり、出力の一部を個別ファイルに保存したりすることができます。

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

同一モデル以外のモデルの場合のみ、個別空間の座標を表示します。

個別空間の重み

いずれかの個人差モデルが指定されている場合のみ、個別空間の重みを表示します。モデルによっては、空間の重みが回転の重みと次元の重みに分解され、これらの重みも表示されます。

距離 布置のオブジェクト間の距離を表示します。

変換された近接

布置のオブジェクト間の変換後の近傍度を表示します。

入力データ

元の近傍度を表示します。データの重み、初期布置、固定座標がある場合は、それらも表示します。

ランダム・スタートのストレス

各ランダム・スタートの乱数シードと正規化された原ストレスを表示します。

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

各種ストレス値を表示します。テーブルに、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、説明された分散 (DAF)、Tucker の適合係数が表示されます。

ストレス分解

正規化された最終的な原ストレスの、オブジェクトとソースのストレス分解を表示します (オブジェクト当たりの平均とソース当たりの平均を含む)。

変換された独立変数

線型結合の制限が選択されている場合に、変換後の独立変数と対応する回帰重みを表示します。

変数と次元の相関

線型結合の制限が選択されている場合に、独立変数と共通空間の次元との相関を表示します。

共通空間の座標

共通空間の座標を表示します。共通空間の座標を個別の IBM SPSS Statistics データ・ファイルに保存することができます。

距離 布置のオブジェクト間の距離を表示します。距離を個別の IBM SPSS Statistics データ・ファイルに保存することができます。

変換された独立変数

線型結合の制限が選択されている場合に、変換後の独立変数と対応する回帰重みを表示します。変換後の独立変数を個別の IBM SPSS Statistics データ・ファイルに保存することができます。

個別空間の重み

いずれかの個人差モデルが指定されている場合のみ、個別空間の重みを表示します。モデルによっては、空間の重みが回転の重みと次元の重みに分解され、これらの重みも表示されます。個別空間の重みを個別の IBM SPSS Statistics データ・ファイルに保存することができます。

変換された近接

布置のオブジェクト間の変換後の近傍度を表示します。変換後の近傍度を個別の IBM SPSS Statistics データ・ファイルに保存することができます。

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と、近接度の多次元尺度法分析をカスタマイズすることができます。このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 変換プロットと残差プロットについて、個別の変数リストを指定する (PLOT サブコマンドを使用)。
- 個別空間の重みプロット、変換プロット、残差プロットについて、個別のソース・リストを指定する (PLOT サブコマンドを使用)。
- 表示する独立変数変換プロットのサブグループを指定する (PLOT サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」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ためのもので、2 組のオブジェクト間の関係を視覚的に調べることができます。

例: 21 人の個人に対して、15 項目の朝食を好みの順に 1 から 15 まで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主に 2 つの方法 (柔らかいパンと硬いパン、脂肪のある項目と脂肪のない項目) で区別していることがわかります。

また、運転手のグループに対して、26 モデルの自動車の 10 個の属性について 6 段階の尺度 (1 = 「まったくそう思わない」から 6 = 「本当にそう思う」) で評価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、値は類似度と見なされます。多次元展開を使用することにより、類似するモデル群と、互いに最も密接に関連する属性群が見つかります。

統計と作図: 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、変換後の近傍、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。

多次元展開データの考慮事項

データ: データは、長方近傍行列の形式で提供されます。各列は、個別の列オブジェクトと見なされます。近傍行列の各行は、個別の行オブジェクトと見なされます。複数の近傍度ソースがある場合、行列が積み重ねられます。

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 を減算した数値を超えることはできません。ソースを 1 つだけ指定した場合、すべてのモデルが同一モデルに等しくなるため、分析のデフォルト・モデルは同一モデルになります。

多次元展開を行うには

1. メニューから次の項目を選択します。

「分析」 > 「尺度」 > 「多次元展開 (PREFSCAL)...

2. 長方近傍行列の列を識別する変数を 2 つ以上選択します。各変数は、個別の列オブジェクトを表します。
3. 必要に応じて、列オブジェクト変数の数に等しい数の重み付け変数を選択します。重み付け変数の順序は、重み付けの対象となる列オブジェクトの順序と一致している必要があります。
4. 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、出力内の行オブジェクトのラベル付けに使用されます。
5. 複数のソースがある場合は、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、行オブジェクト数にソース数を乗算した数値と一致している必要があります。

また、多次元展開のモデルを定義したり、共通空間に対して制限を設定したり、収束基準を設定したり、使用する初期布置を指定したり、プロットと出力を選択したりすることができます。

多次元展開: モデルの定義

「モデル」ダイアログでは、尺度モデル、尺度モデルの次元の最小数と最大数、近傍行列の構造、近傍度に対して使用する変換、近傍度の変換条件 (行を条件として変換するか、ソースを条件として変換するか、ソースとは無関係に変換するか) を指定することができます。

尺度モデル

以下のオプションから選択します。

同一 全てのソースが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。各ソースは個別空間を持ち、その個別空間では共通空間のすべての次元に異なる重みが付けられます。

一般化ユークリッド

このモデルは個人差モデルです。各ソースは個別空間を持ち、その個別空間は共通空間を回転したのと同じで、各次元に異なる重みが付けられます。

近傍 類似度と非類似度のどちらの指標を近傍行列に含めるかを指定します。

次元 デフォルトでは、解は 2 つの次元で計算されます (最小値 = 2、最大値 = 2)。最小値は 1 で、最大値は、オブジェクトの数から 1 を減算した数です。この範囲内の整数を指定することができます。ただし、最小値には最大値以下の値を指定する必要があります。この手続きでは、最大次元で解が計算され、その後、最小値に達するまで段階的に次元数が減っていきます。

近接変換

以下のオプションから選択します。

なし 近傍度は変換されません。必要に応じて、「定数項を含める」を選択することができます。これを選択すると、近傍度を定数項分だけずらすことができます。

線型 変換後の近傍度は元の近傍度に比例します。つまり、変換関数によって傾きが計算され、切片は 0 に固定されます。この変換は、比率変換とも呼ばれます。必要に応じて、「定数項を含める」を選択することができます。これを選択すると、近傍度を定数項分だけずらすことができます。この変換は、間隔変換とも呼ばれます。

スプライン

変換後の近傍度は、元の近傍度を滑らかな非減少区分的多項式で変換したものになります。多項式の次数と内側ノットの数指定することができます。必要に応じて、「定数項を含める」を選択することができます。これを選択すると、近傍度を定数項分だけずらすことができます。

平滑化

後続の値との差を考慮する制限が適用され、変換後の近傍度は元の近傍度と同じ順序になります。結果的に、「滑らかな順序」という変換になります。近傍度の順序を固定したままにするか、固定の解除を許可するかを指定することができます。

順序 変換後の近傍度は元の近傍度と同じ順序になります。近傍度の順序を固定したままにするか、固定の解除を許可するかを指定することができます。

変換の適用

近傍度の相互比較を、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、変換を列ごとに実行するか、ソースごとに実行するか、またはすべての近傍度に対して同時に実行するかを指定します。

多次元展開: 制限

「制限」ダイアログでは、共通空間に対する制限を設定することができます。

共通空間に対する制限

共通空間で、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) の座標を固定することができます。

行/列制限変数

制限が記述されたファイルを選択し、共通空間に対する制限を定義する変数を選択します。選択した最初の変数には最初の次元のオブジェクトの座標が格納され、2 番目の変数には 2 番目の次元

の座標が格納されます (3 番目以降についても同様です)。欠損値は、次元の座標が空いていることを示します。選択した変数の数と要求した次元の最大数は等しくなければなりません。また、各変数のケースの数は、オブジェクトの数と等しくなければなりません。

多次元展開: オプション

「オプション」ダイアログでは、初期布置の形式の選択、反復基準と収束基準の指定、ストレスのペナルティ項目の設定を行うことができます。

初期布置

以下のいずれかを選択します。

古典法

長方近傍行列を使用して、完全対称 MDS 行列の内部ブロック (行間の値と列間の値) が補完されます。この完全行列の作成後、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ることができます。

Ross-Cliff

Ross-Cliff スタートでは、行オブジェクトと列オブジェクトの初期値として、二重中心化されて平方化された近傍行列上の特異値分解の結果が使用されます。

応答 応答スタートでは、行スコアと列スコアの対称的正規化とともに、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が使用されます。

重心 固有値分解を使用して布置に行オブジェクトを配置することにより、手続きが開始されます。次に、列オブジェクトが、指定された選択枝の重心に配置されます。選択枝の数として、1 から近傍変数の数までの正の整数を指定します。

ランダム スタート

無作為に選択された複数の初期布置に対して解が計算され、最も小さいペナルティ付きストレスを持つ解が最適解として表示されます。

ユーザー指定

独自の初期布置の座標を格納する変数を選択することができます。最初の変数は次元 1 の座標に対応し、2 番目の変数は次元 2 の座標に対応するというように、選択した変数の数と指定した次元の最大数が等しくなる必要があります。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数を加算した数値に等しくなる必要があります。行と列の座標は、行座標の後に列座標が配置される形で積み重ねる必要があります。

反復基準

反復基準の値を指定します。

ストレス収束

連続するペナルティ付きストレス値における相対差が、ここで指定した数値よりも小さくなると、このアルゴリズムは反復を停止します。この数値には、負ではない値を指定する必要があります。

最小ストレス

ペナルティ付きストレスがここで指定した数値よりも小さくなると、このアルゴリズムは停止します。この数値には、負ではない値を指定する必要があります。

最大反復回数

上記の基準がどれも満たされなかった場合、ここで指定した回数の反復処理がこのアルゴリズムによって実行されます。

ペナルティー項目

このアルゴリズムは、ペナルティー付きストレスを最小にしようとします。ペナルティー付きストレスとは、Kruskal のストレス I と、変換後の近傍度の変動係数に基づくペナルティー項目との積に等しい適合度のことです。これらのコントロールにより、ペナルティー項目の強さと範囲を設定することができます。

強さ このパラメーターの値を小さくするほど、ペナルティーの強度が高くなります。0.0 から 1.0 までの範囲の値を指定してください。

範囲 このパラメーターにより、ペナルティーがアクティブになるタイミングを設定します。これを 0.0 に設定すると、ペナルティーはアクティブになりません。この値を大きくすると、アルゴリズムにより、変換後の近傍度においてより大きな変動を持つ解が検索されず、負ではない値を指定してください。

多次元展開: 作図

「作図」ダイアログでは、作成するプロットを指定することができます。

プロット

使用可能な作図は以下のとおりです。

多重開始値

ペナルティー付きストレスの積み上げヒストグラムが表示されます。ストレスとペナルティー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が表示されます。

次元ごとのストレス

ペナルティー付きストレスと次元を対比する折れ線グラフが作成されます。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のみ作成されます。

最終共通空間

共通空間の座標の散布図行列が表示されます。

空間の重み

個別空間の重みの散布図が作成されます。「モデル」ダイアログ・ボックスでいずれかの個人差モデルを指定した場合のみ、このプロットを使用できます。重み付きユークリッド・モデルの場合、すべての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・モデルの場合、次元ごとに 1 つのプロットが作成され、ソースごとにその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。「モデル」ダイアログでいずれかの個人差モデルを指定した場合のみ、このプロットを使用できます。

変換図

元の近傍度と変換後の近傍度を対比する散布図が作成されます。変換の適用方法に応じて、行ごとまたはソースごとに個別の色が割り当てられます。無条件変換の場合は、1 つの色だけが使用されます。

Shepard 図

元の近傍度と、変換後の近傍度と距離の両方を対比する図が表示されます。距離は点で示され、変換後の近傍度は線で示されます。変換の適用方法に応じて、行ごとまたはソースごとに個別の線が作成されます。無条件変換の場合は、1 つの線だけが作成されます。

適合度の散布図

変換後の近傍度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、ソースごとに個別の色が割り当てられます。

残差図

変換後の近傍度と残差 (変換後の近傍度から距離を減算した値) 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、ソースごとに個別の色が割り当てられます。

行オブジェクトの種類

プロット内の行オブジェクトの表示をより詳細に制御することができます。オプションの色変数の値を使用すると、任意の色に切り替えることができます。また、オプションのマーカー変数の値を使用すると、任意の使用可能なマーカーに切り替えることができます。

ソース・プロット

「個別空間」、「適合度の散布図」、および「残差プロット」(変換がソースによって適用される場合は、これらのプロットのほかに「変換プロット」と「Shepard 図」) では、プロットを作成する必要があるソースを指定することができます。ソース番号には、メイン・ダイアログ・ボックスで指定したソース変数の値を入力する必要があります。また、1 からソース数までの範囲で入力する必要があります。

行の作図

変換が行によって適用される場合、「変換プロット」と「Shepard 図」では、プロットを作成する必要がある行を指定することができます。行番号には、1 から行の数までの範囲で値を入力する必要があります。

多次元展開: 出力

「出力」ダイアログでは、表示される出力量を制御したり、出力の一部を個別ファイルに保存したりすることができます。

表示 表示する項目を、次の中から 1 つ以上選択します。

入力データ

元の近傍度を表示します。データの重み、初期布置、固定座標がある場合は、それらも表示します。

多重開始値

各ランダム・スタートの乱数シードとペナルティー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

さまざまな測度を表示します。テーブルに、複数の適合度、不適合度、相関、変動、非縮退の度数が表示されます。

ストレス分解

オブジェクト、行、ソースのペナルティー付きストレス分解を表示します。行、列、ソースの平均と標準偏差も表示します。

変換された近接

変換後の近傍度を表示します。

最終共通空間

共通空間の座標を表示します。

空間の重み

個別空間の重みを表示します。いずれかの個人差モデルを指定した場合のみ、このオプションを指定することができます。モデルによっては、空間の重みが回転の重みと次元の重みに分解され、これらの重みも表示されます。

個別空間

個別空間の座標を表示します。いずれかの個人差モデルを指定した場合のみ、このオプションを指定することができます。

適合した距離

布置のオブジェクト間の距離を表示します。

共通空間の座標、個別空間の重み、距離、変換後の近傍度を、個別の IBM SPSS Statistics データ・ファイルに保存することができます。

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と、近接度の多次元展開分析をカスタマイズすることができます。このコマンド・シンタックス言語では、次の作業を行うこともできます。

- 複数のソースを使用できる場合、個別空間プロット、適合度の散布図、残差プロット (行列の条件変換の場合は、これらのほかに変換プロットと Shepard 図) 用に複数のソースのリストを指定する (PLOT サブコマンドを使用)。
- 行条件変換の場合、変換プロットと Shepard 図用に複数の行リストを指定する (PLOT サブコマンドを使用)。
- 行 ID 変数の代わりに行番号を指定する (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する (INPUT サブコマンドを使用)。

シンタックスについて詳しくは、「コマンド・シンタックス・リファレンス」を参照してください。

特記事項

本書は米国 IBM が提供する製品およびサービスについて作成したものです。この資料は、IBM から他の言語でも提供されている可能性があります。ただし、これを入手するには、本製品または当該言語版製品を所有している必要がある場合があります。

本書に記載の製品、サービス、または機能が日本においては提供されていない場合があります。日本で利用可能な製品、サービス、および機能については、日本 IBM の営業担当員にお尋ねください。本書で IBM 製品、プログラム、またはサービスに言及していても、その IBM 製品、プログラム、またはサービスのみが使用可能であることを意味するものではありません。これらに代えて、IBM の知的所有権を侵害することのない、機能的に同等の製品、プログラム、またはサービスを使用することができます。ただし、IBM 以外の製品とプログラムの操作またはサービスの評価および検証は、お客様の責任で行っていただきます。

IBM は、本書に記載されている内容に関して特許権 (特許出願中のものを含む) を保有している場合があります。本書の提供は、お客様にこれらの特許権について実施権を許諾することを意味するものではありません。実施権についてのお問い合わせは、書面にて下記宛先にお送りください。

〒103-8510

東京都中央区日本橋箱崎町19番21号

日本アイ・ビー・エム株式会社

法務・知的財産

知的財産権ライセンス渉外

IBM およびその直接または間接の子会社は、本書を特定物として現存するままの状態を提供し、商品性の保証、特定目的適合性の保証および法律上の瑕疵担保責任を含むすべての明示もしくは黙示の保証責任を負わないものとします。国または地域によっては、法律の強行規定により、保証責任の制限が禁じられる場合、強行規定の制限を受けるものとします。

この情報には、技術的に不適切な記述や誤植を含む場合があります。本書は定期的に見直され、必要な変更は本書の次版に組み込まれます。IBM は予告なしに、随時、この文書に記載されている製品またはプログラムに対して、改良または変更を行うことがあります。

本書において IBM 以外の Web サイトに言及している場合がありますが、便宜のため記載しただけであり、決してそれらの Web サイトを推奨するものではありません。それらの Web サイトにある資料は、この IBM 製品の資料の一部ではありません。それらの Web サイトは、お客様の責任でご使用ください。

IBM は、お客様が提供するいかなる情報も、お客様に対してなら義務も負うことのない、自ら適切と信ずる方法で、使用もしくは配布することができるものとします。

本プログラムのライセンス保持者で、(i) 独自に作成したプログラムとその他のプログラム (本プログラムを含む) との間での情報交換、および (ii) 交換された情報の相互利用を可能にすることを目的として、本プログラムに関する情報を必要とする方は、下記に連絡してください。

IBM Director of Licensing

IBM Corporation

North Castle Drive, MD-NC119

Armonk, NY 10504-1785

US

本プログラムに関する上記の情報は、適切な使用条件の下で使用することができますが、有償の場合もあります。

本書で説明されているライセンス・プログラムまたはその他のライセンス資料は、IBM 所定のプログラム契約の契約条項、IBM プログラムのご使用条件、またはそれと同等の条項に基づいて、IBM より提供されます。

記載されている性能データとお客様事例は、例として示す目的でのみ提供されています。実際の結果は特定の構成や稼働条件によって異なります。

IBM 以外の製品に関する情報は、その製品の供給者、出版物、もしくはその他の公に利用可能なソースから入手したものです。IBM は、それらの製品のテストは行っておりません。したがって、他社製品に関する実行性、互換性、またはその他の要求については確認できません。IBM 以外の製品の性能に関する質問は、それらの製品の供給者にお願いします。

IBM の将来の方向または意向に関する記述については、予告なしに変更または撤回される場合があります、単に目標を示しているものです。

本書には、日常の業務処理で用いられるデータや報告書の例が含まれています。より具体性を与えるために、それらの例には、個人、企業、ブランド、あるいは製品などの名前が含まれている場合があります。これらの名前はすべて架空のものであり、名前や住所が類似する個人や企業が実在しているとしても、それは偶然にすぎません。

著作権使用許諾:

本書には、様々なオペレーティング・プラットフォームでのプログラミング手法を例示するサンプル・アプリケーション・プログラムがソース言語で掲載されています。お客様は、サンプル・プログラムが書かれているオペレーティング・プラットフォームのアプリケーション・プログラミング・インターフェースに準拠したアプリケーション・プログラムの開発、使用、販売、配布を目的として、いかなる形式においても、IBM に対価を支払うことなくこれを複製し、改変し、配布することができます。このサンプル・プログラムは、あらゆる条件下における完全なテストを経ていません。従って IBM は、これらのサンプル・プログラムについて信頼性、利便性もしくは機能性があることをほのめかしたり、保証することはできません。これらのサンプル・プログラムは特定物として現存するままの状態を提供されるものであり、いかなる保証も提供されません。IBM は、お客様の当該サンプル・プログラムの使用から生ずるいかなる損害に対しても一切の責任を負いません。

それぞれの複製物、サンプル・プログラムのいかなる部分、またはすべての派生的創作物にも、次のように、著作権表示を入れていただく必要があります。

© IBM 2019. このコードの一部は、IBM Corp. のサンプル・プログラムから取られています。

© Copyright IBM Corp. 1989 - 20019. All rights reserved.

商標

IBM、IBM ロゴおよび [ibm.com](http://www.ibm.com) は、世界の多くの国で登録された International Business Machines Corporation の商標です。他の製品名およびサービス名等は、それぞれ IBM または各社の商標である場合があります。現時点での IBM の商標リストについては、<http://www.ibm.com/legal/copytrade.shtml> をご覧ください。

Adobe、Adobe ロゴ、PostScript、PostScript ロゴは、Adobe Systems Incorporated の米国およびその他の国における登録商標または商標です。

インテル、Intel、Intel ロゴ、Intel Inside、Intel Inside ロゴ、Centrino、Intel Centrino ロゴ、Celeron、Xeon、Intel SpeedStep、Itanium、および Pentium は、Intel Corporation または子会社の米国およびその他の国における商標または登録商標です。

Linux は、Linus Torvalds の米国およびその他の国における登録商標です。

Microsoft、Windows、Windows NT および Windows ロゴは、Microsoft Corporation の米国およびその他の国における商標です。

UNIX は The Open Group の米国およびその他の国における登録商標です。

Java およびすべての Java 関連の商標およびロゴは Oracle やその関連会社の米国およびその他の国における商標または登録商標です。

索引

日本語, 数字, 英字, 特殊文字の順に配列されています。なお, 濁音と半濁音は清音と同等に扱われています。

[ア行]

- 一般化ユークリッド・モデル
 - 多次元展開 42
- イナーシャ (慣性)
 - コレスポネンス分析 26
- オブジェクト・スコア
 - カテゴリー主成分分析 20
 - 多重応答分析 32
- オブジェクト・ポイントのプロット
 - カテゴリー主成分分析 21
 - 多重応答分析 33
- 重み付きユークリッド・モデル
 - 多次元展開 42

[カ行]

- 回帰係数
 - カテゴリー回帰 14
- カテゴリー回帰 10
 - コマンドの追加機能 16
 - 最適尺度水準 11
 - 作図 10
 - 正規化 13
 - 統計 10
 - 保存 15
- カテゴリー主成分分析 16, 19
 - コマンドの追加機能 23
 - 最適尺度水準 17
 - 変数の保存 21
- カテゴリー数量化
 - カテゴリー回帰 14
 - カテゴリー主成分分析 20
 - 多重応答分析 32
- カテゴリー・プロット
 - カテゴリー主成分分析 22
 - 多重応答分析 33
- 記述統計
 - カテゴリー回帰 14
- 共通空間に対する制限
 - 多次元展開 43
- 共通空間の座標
 - 多次元尺度法 40
 - 多次元展開 46

- 共通空間プロット
 - 多次元尺度法 39
 - 多次元展開 45
- 距離
 - 多次元尺度法 40
 - 多次元展開 46
- 距離測定
 - コレスポネンス分析 25
- 近接変換
 - 多次元展開 42
- 空間の重みプロット
 - 多次元展開 45
- 結合カテゴリー・プロット
 - カテゴリー主成分分析 22
 - 多重応答分析 33
- 欠損値
 - カテゴリー回帰 12
 - カテゴリー主成分分析 18
 - 多重応答分析 30
- 個別空間の重み
 - 多次元尺度法 40
 - 多次元展開 46
- 個別空間の重みプロット
 - 多次元尺度法 39
 - 多次元展開 45
- 個別空間の座標
 - 多次元展開 46
- 個別空間プロット
 - 多次元尺度法 39
 - 多次元展開 45
- コレスポネンス分析 23, 24, 25, 26, 27
 - コマンドの追加機能 28
 - 作図 23
 - 統計 23

[サ行]

- 最終共通空間プロット
 - 多次元展開 45
- 最適尺度水準
 - カテゴリー主成分分析 17
 - 多重応答分析 29
- 作図
 - カテゴリー回帰 15
 - コレスポネンス分析 27
 - 多次元尺度法 39, 40
- 残差プロット
 - 多次元展開 45
- 次元
 - コレスポネンス分析 25

- 射影する重心のプロット
 - カテゴリー主成分分析 22
- 尺度モデル
 - 多次元展開 42
- 初期共通空間プロット
 - 多次元展開 45
- 初期布置
 - カテゴリー回帰 12
 - 多次元尺度法 38
 - 多次元展開 44
- 信頼統計量
 - コレスポネンス分析 26
- ストレス測定
 - 多次元尺度法 40
 - 多次元展開 46
- ストレス・プロット
 - 多次元尺度法 39
 - 多次元展開 45
- 正規化
 - コレスポネンス分析 25
- 制限
 - 多次元尺度法 38
- 成分負荷
 - カテゴリー主成分分析 20
- 成分負荷プロット
 - カテゴリー主成分分析 22
- 説明された分散
 - カテゴリー主成分分析 20
- 相関
 - 多次元尺度法 40
- 相関行列
 - カテゴリー主成分分析 20
 - 多重応答分析 32
- 相関プロット
 - 多次元尺度法 39

[タ行]

- 多次元スケーリング法 34, 35, 36
 - オプション 38
 - コマンドの追加機能 41
 - 作図 34, 39, 40
 - 出力 40
 - 制限 38
 - 統計 34
 - モデル 37
- 多次元展開 41
 - オプション 44
 - 共通空間に対する制限 43
 - コマンドの追加機能 47
 - 作図 41, 45

- 多次元展開 (続き)
 - 出力 46
 - 統計 41
 - モデル 42
- 多重開始プロット
 - 多次元展開 45
- 多重回答
 - カテゴリー回帰 14
- 多重コレスポネンス分析 28, 31
 - コマンドの追加機能 34
 - 最適尺度水準 29
 - 変数の保存 33
- 適合度の散布図
 - 多次元展開 45
- 同一モデル
 - 多次元展開 42
- トリプロット
 - カテゴリー主成分分析 21

[ハ行]

- バイプロット
 - カテゴリー主成分分析 21
 - コレスポネンス分析 27
 - 多重応答分析 33
- 反復基準
 - 多次元尺度法 38
 - 多次元展開 44
- 反復履歴
 - カテゴリー主成分分析 20
 - 多次元尺度法 40
 - 多次元展開 46
 - 多重応答分析 32
- 判別測定
 - 多重応答分析 32
- 判別測定プロット
 - 多重応答分析 33
- 標準化
 - コレスポネンス分析 25
- 分散分析
 - カテゴリー回帰 14
- ペナルティー項目
 - 多次元展開 44
- 変換された近接
 - 多次元尺度法 40
 - 多次元展開 46
- 変換された独立変数
 - 多次元尺度法 40
- 変換プロット
 - カテゴリー主成分分析 22
 - 多次元尺度法 39
 - 多次元展開 45
 - 多重応答分析 33
- 変数の重み
 - カテゴリー主成分分析 17
 - 多重応答分析 29

- 補助オブジェクト
 - カテゴリー回帰 12

[ヤ行]

- より速いアルゴリズム
 - 多次元尺度法 38

[ラ行]

- 離散化
 - カテゴリー回帰 11
 - カテゴリー主成分分析 17
 - 多重応答分析 29

E

- Elastic Net
 - カテゴリー回帰 13

L

- Lasso
 - カテゴリー回帰 13

P

- PREFSCAL 41

R

- Ridge 回帰
 - カテゴリー回帰 13

S

- Shepard 図
 - 多次元展開 45



Printed in Japan