『津田塾大学紀要』第 55 号（2023 年）119-139 頁

「幾何学的データ解析（GDA）」では分散はどのように分解されるのか

―GDA で ANOVA の手法を用いるために押さえるべき事がある―

藤　本　一　男

概要

構造化データ解析（SDA）は、多重対応分析（MCA）を用いた幾何学的データ解析（GDA）における中心部分である。

SDA は、実験計画が不可能な観察データである社会調査データに対して、ANOVA の手法を用いた構造分析を行う。その手法を正しく活用するためには、MCA が元の表のデータをどのように分解するのかを、理解しておかなくてはならない。また、ANOVA の手法と MCA を結びつけることを数学的に保証している処理を理解しておかないと、「GDA 風」の分析にしかならない。対応分析（CA）も多重対応分析も、その処理の根底には、分析対象のデータ表の分散の分解がある。

そこで、本稿では、MCA が分散をどのように分解するのかを整理した。出発点（第 0 分解）は、分析対象表の持つ分散である。これが、分析の構造設計をへて、CA の処理（特異値分解）がおこなわれ次元縮減がおこなわれる（第 1 分解）。この過程で行空間と列空間という二つの空間が生成される。変数カテゴリ空間である列空間のカテゴリポイントの軸への寄与をみながら、軸の解釈、命名を行う（第 3 の分解 / 混合）。さらに、構造設計段階で追加変数として用意されたものを、個体空間に射影し、そこで ANOVA の技法を用いて、空間の構造分析を行う。

これらの過程で、LeRoux&Rouanet2010=2021 で MCA のバリアントと紹介されている speMCA、CSA について位置付けを述べた。

キーワード：幾何学的データ解析（GDA）、多重対応分析（MCA）、分散分析（ANOVA）、群内分散、群間分散、相関比、R、GDAtools

1　はじめに GDA と ANOVA

多重対応分析（MCA: Multiple Correspondence Analysis）を中心にすえた幾何学データ解析（GDA: Geometric Data Analysis）は、実験計画が不可能な条件下で採取された観察データに対して、分散分析（ANOVA）の手法を用いて構造分析を実現する。ここで「分散分析」という表記にコメントしておく。分散分析は、2つの処理にわけることができる。分析対象の分散（平方和）を分解する処理と、その分解を元にF検定を行う処理である。一般に分散分析というときには、この検定過程に注目している。これを、今扱おうとしているGDAにあてはめると、分析対象の構造化分析の部分（SDA: Structured Data Analysis）が、前者にあたり、そこで分解された分散をの様態をもとに検定を行うのが、帰納的データ解析（IDA: Inductive Data Analysis）と呼ばれる部分に相当する。本稿では、このSDAに注目する（図1）。

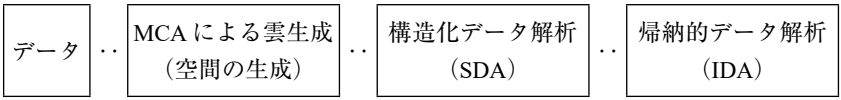


図1　GDAの遂行過程

入手したデータをMCAすることによって、二つの雲（ポイントCloud、空間）が生成される。データは、その空間の生成に寄与させる行（個体）や列（変数）と、寄与させずにその空間に射影させるものとに区分されている（構造化設計）。そして、そのように生成された雲の部分雲を形成し、その平均点（重心）および寄与率をもとに、データを部分集合に分割しその構造を分析していく。最後の帰納的データ解析（IDA）は、そうして取得された分析の検定を行う。

GDAのこの全体過程を通してデータが有する分散が分解され、変数カテゴリとの関係で評価されていく。その過程を理解する鍵は、対応分析、多重対応分析が、分析対象の分散をどのように分解するのか、を理解するところにある。そして、この分散の分解の理解とあわせて、それを実現する前提になっているコーディングの条件があり、それを踏まえて低頻度カテゴリの扱いなどを適切に行わないならば、「GDA風」の分析にしかならないので注意が必要である。

2　前提的確認

2. 1　CA と MCA の処理の原理的な同等性の確認

対応分析（CA）は、クロス表のような 2 変数データに対する処理で、多重対応分析（MCA）は、3 変数以上のデータで、行が回答者個体、列が変数という構成のデータに適用されると言われる。しかし､以下の説明においては､まず、CA における行変数は、MCA における個体、CA における列変数は、MCA における変数、に対応するものであることを確認しておきたい。

MCA において行は個体であるが、列の変数は、変数カテゴリに展開され、そのカテゴリが選択されている場合には、1 をたて、同時に、変数内には、必ず一つは 1 がたつ、つまり、行和は、変数の数になる、というルールのもとにコーディングされているという条件がある [1)]。

MCA とは、このように列を変数ごとに必ず 1 が一つたっている「変数カテゴリ」を列としてコーディングされた指示行列に対して CA を行うことである。そのため、空間の生成、その空間のポイント間の関係などについて理解する際には、CA での例でも MCA での例も同じ考え方でよい。

2. 2　CA/MCA の結果、2 つの空間が生成される

分析対象のデータは、行と列の 2 元のデータで、その交差したところに、CA であればクロス表の度数が入っており、MCA であれば 1 と 0 が入っている。こうした 2 元表の標準化残差行列（各要素から期待値要素を引き、標準偏差に相当する期待値要素の平方根で除した行列）の特異値分解（SVD：Singular Value Decompression）によって得られる三つの行列、U、D、V　の積に再構成される [2)]。

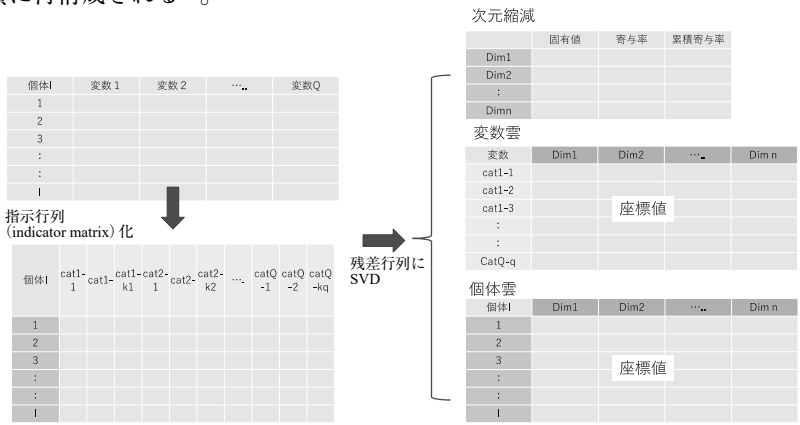


図 2　MCA による空間の生成を模式図で表示する

ここで、D は特異値（固有値の平方根）の大きさの順の対角行列であるが、この対角要素を選択することで、もとの行列の次元縮減をした近似値が得られることになる。D は、この次元変換の結果生成される固有値（座標軸）の行列であり、この行列 U は行変数 / 個体の座標に関係し、V は、列変数 / 変数カテゴリに関係する。

さらに、このように生成された「個体空間（雲）」「変数空間（雲）」は、それぞれ行が「個体」と「変数カテゴリ」になっており、列は分解され生成された空間の次元になったことを確認しておきたい（図２）。

3　第 0 の分散の分解

元データ行列 N を、I 行 J 列の行列としよう。この N が指示行列 注）である場合は、この元行列のもつ分散は、非常にシンプルに計算される。J 列の設問が Q 個、そのカテゴリ総数が K 個であるとすると、この行列の分散は、

$$Vcloud \quad = (K/Q) \quad -1$$

という簡単な数式で表現される Le Roux&Rouanet2010=2021:53。以下の CA 行列処理の過程の出発点は、この値である。特異値分解の結果計算されるすべての次元を総合すれば、元の行列の分散＝情報は、100% 回復されるわけだが、この分解が次元縮減を目的するので、分散率の大きなものから選択され、十分な分散率を確保できた時は、それよりも先の次元に体現されている分散は、誤差として扱われることになる。

4　第 1 の分散の分解

4. 1　CA の原理と特異値分解

ここで、CA がどのように元の行列から行空間、列空間を生成するのかを確認しておきたい。

処理は、以下のような段取りである（付録を参照）。

元行列（MCA の場合は指示行列）を、行列要素の総数で除して、対応行列 P とする。この P の行和（指示行列なので、変数の数 Q）と列和を計算し、そ

れをもとに、標準化残差行列 S を計算する。

このSに対して、特異値分解を行い、それで得られる3つ行列、U、D、Vから生成される座標の分散（固有値）、行と列の標準座標および主座標が計算される。この結果、図2にあるような二つの空間が生成される。

図2に示したように、行 / 個体、および列 / 変数カテゴリに、次元変換によって生成された空間の多次元の座標が割り当てられる。これが CA や MCA による「数量化」の実際である。つまり、CA や MCA による空間変換は、カテゴリカル変数の数量化処理である。

各軸（Dim1 … n）の下のセルに示されているものが、行列U、行列Vから計算される座標である[3)]。

次に、分解された座標軸が体現する分散である固有値についても確認しておく。これは、行列 D の対角成分として得られるが MCA においては、個体数も変数カテゴリ数も大きなものになるために、分解によって生成される軸数も多くなる。その結果、一つの軸あたりに配分される分散の値は小さいものになる。この問題に対処するために、寄与率を補正する方法が提案されている。この修正寄与率は、分析対象とする空間を何次元までを分析対象にするのか、を判断する時に用いられるが、本稿で扱っている分散の分解という場面での寄与率の計算の分母には補正前の値が用いられる[4)]。

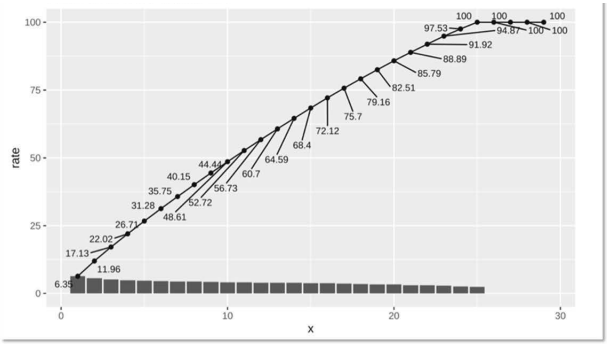


図3　生成された座標軸が有する分散（修正なし）

折れ線グラフは、累積寄与率。

寄与率とは、基準とする分散に対して、注目している軸、個体やカテゴリ点が体現している分散の割合を表している。

4. 2 生成された二つの雲：個体雲と変数雲

MCA によって生成された二つの空間を表示すると以下のようになる（データは、Le ROux&Rouanet2010=2021 で使用されている「嗜好データ」を用いた）。

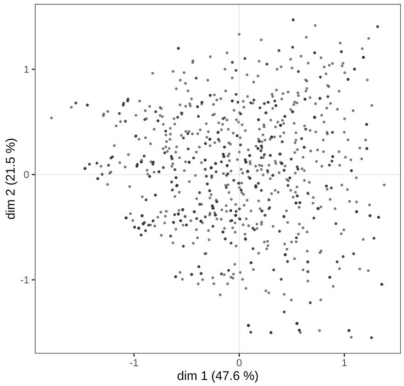


図 4-1　嗜好の個体雲

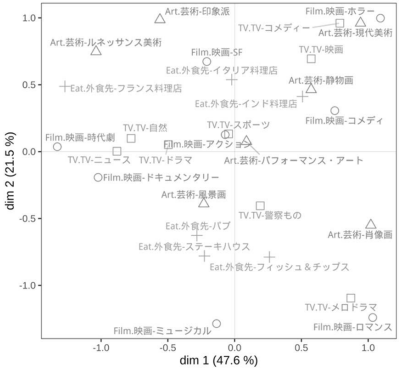


図 4-2　嗜好の変数雲

5　第２の分解（混合カテゴリの解釈）

分析の手順としては、まず、変数空間を構成している変数カテゴリをもとに分解された座標軸の解釈を行う。CA を行う前のデータであれば、それぞれの回答カテゴリが座標軸を形成し、それが張る空間に、個体が位置づく、という関係で考えることができる。

CA（同じく MCA）によって座標変換された空間では、座標軸には、各カテゴリが混在して影響している。最初の課題は、この軸の解釈、名づけにある。調査報告で我々が目にする「経済資本 -/ 文化資本＋　vs　経済資本 +/ 文化資本 -」（ブルデュー 1979=1990,2020:660）という「わかりやすい」（もしくは手にいれたい）座標軸が最初から見つかるものではない。

ここで、座標軸の解釈に必要になるのは、軸の生成に寄与している分散である。つまり、各変数カテゴリが、軸の分散のある量を体現している。つまり軸の生成に寄与している。その寄与の度合いをみながら、軸のプラス方向、マイナス方向の解釈を行い、名付ける。

この判定材料となる各カテゴリの分散は、分析対象の集計表から取得される相対頻度と、次元縮減によって計算された主座標から計算することができる。これを計算することは難しいことではないが、多重対応分析のパッケージでは、このように分析に必要な統計量を計算する function が提供されているので、その結果をつかうことにする。

ここで用いるのは、Nicolas Robette による GDAtools である。その名のとおり、GDA を行うためのツールがまとめられている。

これに含まれている tabcontib（寄与率表）を用いると、各軸の解釈で重視しなくてはならないカテゴリが列挙される（表 1 のデータは「嗜好データ」）。ctr1 が左、もしくは下、ctr2 はその反対側の右、もしくは上を意味している。weight は、MCA にかけられる前の集計表から得られる単純集計での度数である。ctrtot は、ctr1、ctr2 の分散を、weight を重みとして加算した値を設問ごとに示している。最右列には、こうして影響あり、と判定されたカテゴリが全体の分散の何パーセントにあたるかが表示されている。

表 1　第一軸に平均値（100/29）より大きく寄与しているカテゴリ

	var	moda	ctr1	ctr2	weight	ctrtot	cumctrtot
5	Film	映画-時代劇	-12.69		140	34.2	34.2
7		映画-ドキュメンタリー	-5.37		100		
6		映画-ホラー		3.8	62		
8		映画-ロマンス		5.55	101		
9		映画-コメディ		6.79	235		
13	TV	TV-ニュース	-8.78		220	26.91	61.11
11		TV-自然	-4.91		159		
10		TV-コメディー		4.85	152		
12		TV-メロドラマ		8.37	215		
3	Eat	外食先-フランス料理店	-8.21		99	13.55	74.66
4		外食先-インド料理店		5.34	402		
2	Art	芸術-現代美術		5.03	110	11.29	85.95
1		芸術-肖像画		6.26	117		

また、変数雲を表示する function である ggcloud_variables には、軸への寄与が基準値（平均値）よりも大きく検討すべき候補とこの表にあげられているカテゴリだけを表示した平面マップを表示するオプションが用意されている。

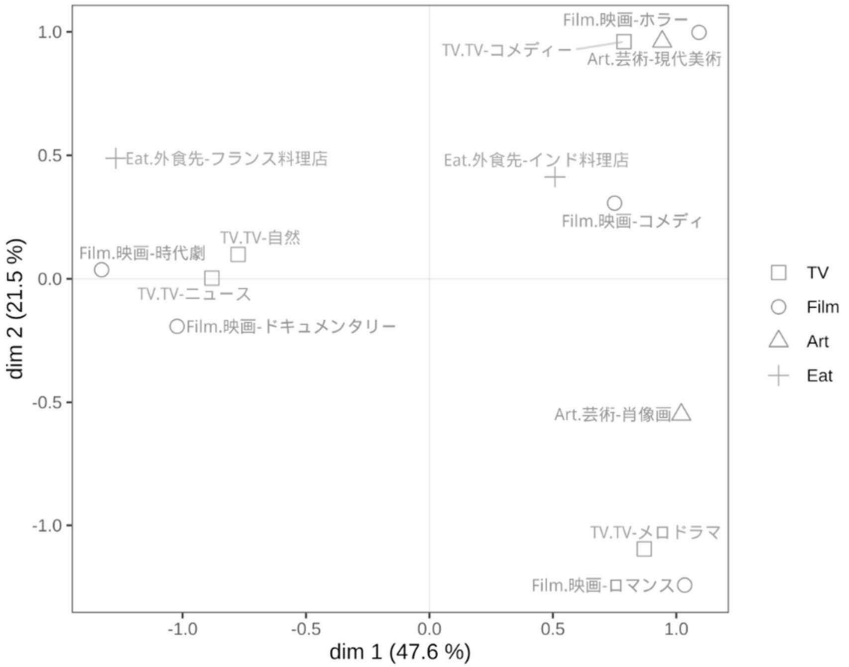


図5　第一軸に大きな寄与をしている点をプロット

2軸、3軸に対しても、同じような表が表示される。これをみながら、各軸の＋方向、－方向を解釈し軸を命名していく*。

* ちなみに、分析サンプルとしてこの嗜好データを用いたLe Roux & Rouanet 2010=2021は、以下のように解釈している。P72、p73、p74

- 第1軸
  - 「事実に即したものへの嗜好」（かつ伝統的なものへの嗜好）
  - 「架空の世界への嗜好」（かつ現代的なものへの嗜好）
- 第2軸
  - 「大衆的なもの」に対する嗜好
  - 「洗練されたもの」に対する嗜好
- 第3軸
  - 「硬いもの」（積極的、活発）
  - 「柔らかいもの」（消極的、穏やかさ）

このように、軸の命名は、機械的にできるものではなく、軸に寄与している変数カテゴリに対する（例えば）社会学的な知見が動員される必要がある。

6　追加変数による構造分析

6.1　空間を構造分析するための「追加変数」

次に、データの構造化を行う追加変数の機能を確認する。この変数は、空間の生成には寄与しないが、座標値をもつことができる、という性質をもっており、このことから、空間の構造分析に用いられる（構造化因子）（Le Roux & Rouanet 2010=2021: 95,168)。これは、先にふれた一方の空間のポイントは、他方の空間のすべての点の加重平均として表される、という関係（遷移公式）に依拠して取得される。

遷移公式は、行列表記では以下のように表現される。ここで用いられる諸要素は、以下の通りである（Clausen1989=2015:102、Greenacre1984:64）。

$$F = RGD_{\alpha}^{-1/2}$$
$$G = CFD_{\alpha}^{-1/2}$$

F:行主座標、G列主座標
R行プロファイル、C列プロファイル

$D_{\alpha}^{-1/2}$　固有値の平方根（特異値）分の１を要素とする対角行列

今､列に新たな回答カテゴリを加えてみる｡この列和は計算できる｡つまり、列比率は計算できる。しかし、行和には変更はない。こういう関係のもとで、その列プロファイルを重みとして、すでに計算されている、つまり空間として生成されている行主座標の加重和をもとめ、それを固有値の平方根でスケーリングする。こうすることで、生成された空間上に射影される追加された変数カテゴリの座標が求められる。行列表記を模試図として描けば図６のようになる。

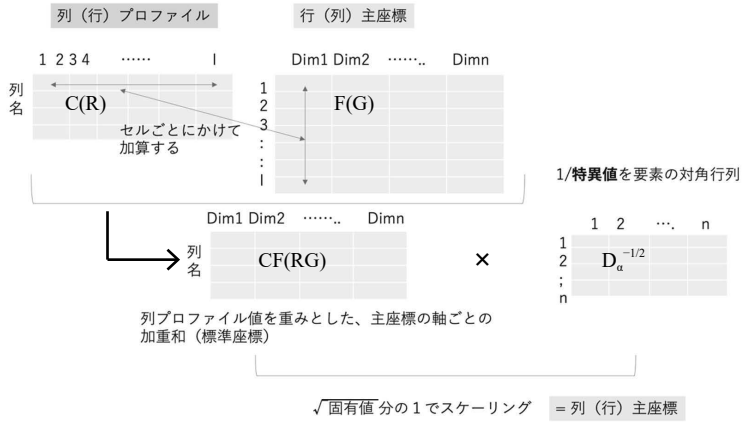


図６　遷移公式を用いて、列追加変数から列種座標を計算する

こうして、追加変数の座標がもとめられ、それらは、変数空間にプロットされる。

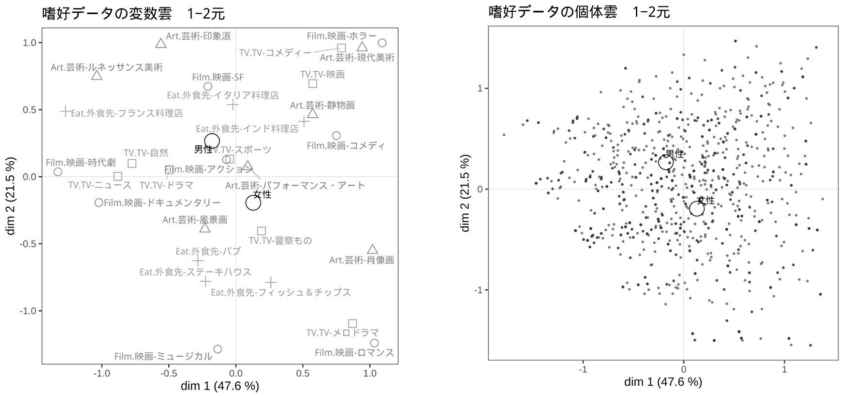


図7　追加変数をプロットした変数雲と個体雲

そして、この追加ポイント（性別、男性 / 女性の平均点）は、個体空間にも位置をもつことができる。この追加変数による構造化が構造分析の第一歩である。これを用いて、生成された空間に、追加変数の平均点をプロットすることができるが、Le Roux と Rouanet は、この構造分析をさらに進化させていった。

7　第３の分解：構造分析（平均点、η^2、集中楕円、ANOVA、CSA）

7. 1　追加変数による個体表の拡張と分析

今、追加変数を二つの空間にプロットした。この追加変数は、もう一つの使い方ができる。空間生成された個体雲のデータ（図 2）に、変数を追加してみよう。つまり、「性別」という列を追加し、そこに男性 / 女性という選択肢が入る状態である。このような表によって、個体雲は、この変数によって部分雲へと分割することが可能になる。つまり、個体全体雲は、男性雲と女性雲への 2 分割される。これは同様に、6 つの年齢区分（カテゴリ）を有する年齢変数を投入すれば、個体全体雲は、年齢カテゴリ（6）に分割されることになる。さらに、合成変数（性別 - 年齢）を考えれば、それに応じて 12 の部分雲への分割することが可能になるのである。

その部分雲は、平均点をもち、また、その平均点を中心とした分散と、平均点同士の分散を考えることができる。この平均点を中心とした分散は、Vwithin 群内分散、また、（分散をもつ）平均点の原点からの分散を、Vbetween 群間分散と呼んで、次のような関係があることが確認されている[5)]。

$$\mathrm{Vtotal} = \Sigma\,\mathrm{Vwithin} + \mathrm{Vbetween}$$

そして、相関比 η^2 が　Vbetween/Vtotal　として定義される。

幾何学的データ解析とは、このように分解した分散の大きさをもとに、変数の関係を分析される。

7. 2 個体雲を分割した部分雲を集中楕円で表示する

分割された個体雲のちらばりを集中楕円（Concentration ellipses）で表現できる。この楕円は、部分雲の重心を中心として、部分雲が正規分布にしたがっていると仮定した場合に、その範囲に、86.47 %が含まれている範囲を示している。これによって部分雲の構造因子による分散を幾何学的に確認することができる。（注：数理的には、Cramer1946 = 1973: 131、林 1993:39、Le Roux & Rouanet 2010 = 2021:97 を参照。）

以下は、性別の部分雲を集中楕円で表示したものである。

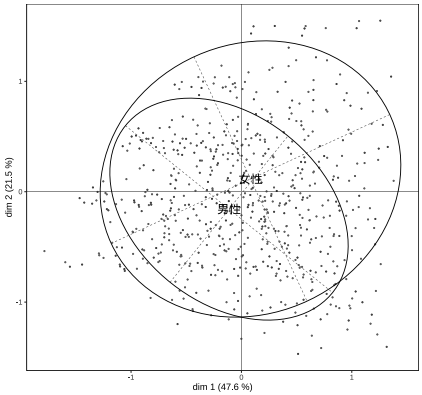


図８　性別＝男性 / 女性で分割された個体雲の集中楕円

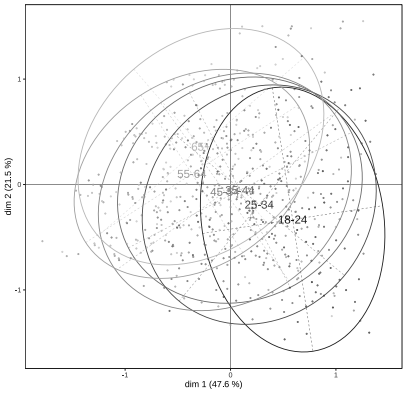


図９　個体雲を、追加変数として年齢で分割した集中楕円

この部分雲に関する、平均点の座標値、平均点（重心）の分散、群内分散、群間分散、相関比（η^2）を算出する function も用意されている。varsup()である。それを用いると以下の出力が得られる。

表２　追加変数「性別」に関連する分散関連の表

	dim.1	dim.2	dim.3
男性	0.291503	0.252755	0.256650
女性	0.463911	0.391608	0.261315
within	0.391117	0.332981	0.259345
between	0.009239	0.018185	0.065664
total	0.400355	0.351166	0.325009
eta2	0.023076	0.051784	0.202037

この表から確認できることを見ておこう。total の行にあるのは、最初に確認した固有値そのもの、つまり軸が体現している分散である。

分散を評価するにあたって、注目したいのは、η2 である。この表をみると、第 3 軸の値が一番大きい（0.202037）。つまり「男女の差は第 3 軸にある」ということを読み取ることができる。

7. 3　年齢を構造因子として投入する

今、性別でみたのと同じように、変数：年齢を空間に射影してみよう。個体雲に集中楕円で年齢をプロットすると図 9 のようになる。

性別と同様に、各平均点の分散、群内分散、群間分散、相関比を表示すると以下のようになる。

表３　追加変数「年齢」に関連する分散関連の表

	dim.1	dim.2	dim.3
18-24	0.191606	0.394644	0.258083
25-34	0.308269	0.322495	0.293415
35-44	0.337113	0.288032	0.340576
45-54	0.360430	0.317632	0.312032
55-64	0.312070	0.245902	0.409496
65+	0.340083	0.314282	0.307831
within	0.320573	0.306763	0.324007
between	0.079782	0.044403	0.001002
total	0.400355	0.351166	0.325009
eta2	0.199278	0.126444	0.003084

この表の相関比（η^2、eta2）からわかることは、第 1 軸が、年齢カテゴリでの変化が一番大きいということである。第 3 軸ではほとんど影響がない。

7. 4　相関比を検討する

変数の「ちらばり」が軸に対してどのようになっているかを η2 は表している。図 10 に示したように、追加変数（性別：Gender、年齢：Age）は、空間生成に寄与した変数にくらべてちらばり度合いはすくない。しかし、すでに検討したように、性別の η2 は、1、2 軸より第 3 軸で大きいことがわかる。また 1-2 軸でみれば、性別よりも年齢による「ちらばり」が大きいことがわかる。

グラフのメモリは、maxが100%=1になっている。このグラフはFactoMineR::plot.MCAで作成

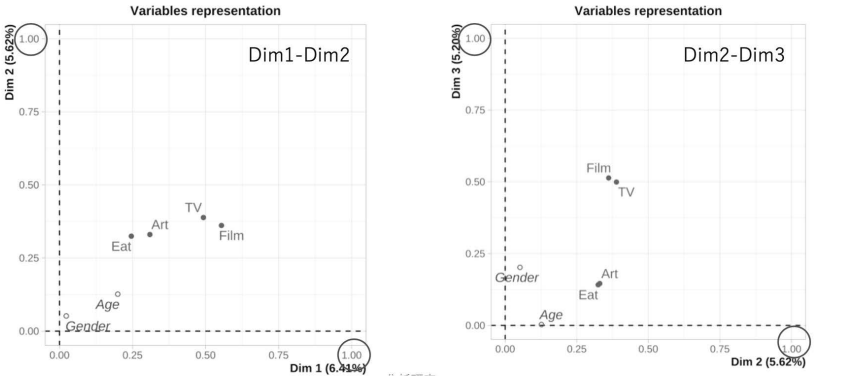


図 10　変数ごとに相関比を比較する

7. 5 平均値の差、分散の差、相関比 η^2 の比較

以下にこうして計算された分割された個体雲の平均値（重心）のへだたり、分散値の比較、と合わせて、相関比 η^2 を図示して比較してみる。

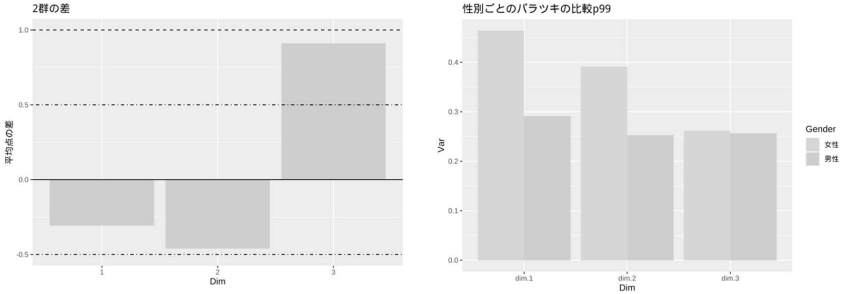


図 11　平均点の座標の差、部分雲の分散の比較

このように、分散の違いは、1、2 軸で顕著であるが、その中心のずれは、第三軸において大きいことがわかる。図 12 に、軸の分散値における群間分散の割合（η^2）を示した。

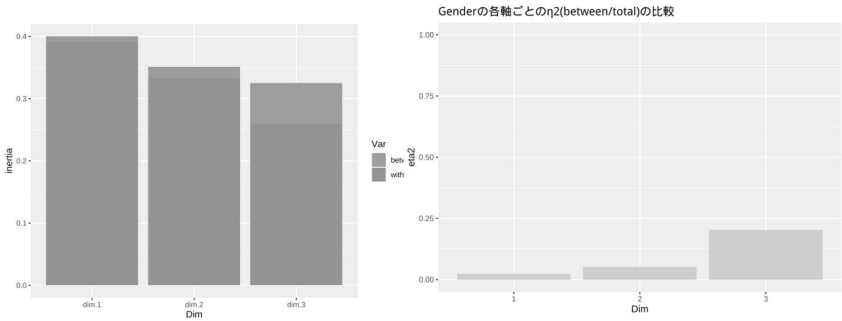


図 12　軸ごとの群内分散、群間分散の比較と、η^2 値

7. 6　交互作用の検討

次に、今見てきた性別（Gender）と年齢（Age）の合成変数を作成し、それを個体空間にプロットしてみる。これまでと同じやり方を使えば、「性別 - 年齢」を構成したのちに、それを、追加変数として、個体雲を分割することができる。次の図（左）は、その分割された個体雲の重心（平均）をプロットしたもので、右の図は、年齢のみでのプロットである。

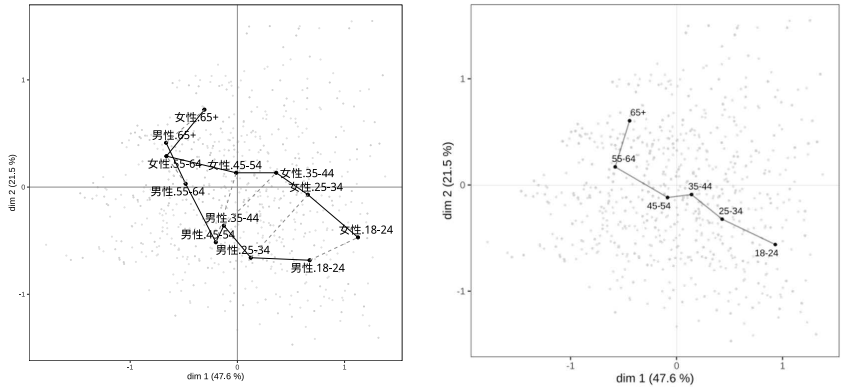


図 13　合成変数による交互作用の表示

合成変数（これは、交互作用を分析するための組み合わせなので、交互作用コーディングと呼ぶこともある。Greenace 2017=2020:122）をプロットすることで、男性と女性で、年齢効果内の性別変化が異なっていることがわかる。

この変化の方向を解釈するためには、変数空間での軸の命名が重要になってくる。右の図からわかることは、年齢が上になるにつれて、1-2 次元平面では、右下の象限から左上の象限へと推移している傾向である。左の図をみれば、男性は、ほぼその傾向であるものの、女性は、55 − 64、65+ の年齢区分で、それまでの年齢区分とは異なる傾向を有していることが示唆される動きを見せている。

なお、Le Roux&Rouanet は、構造化分析による分析のために、比較分析（Analysis of Comparison）という概念のもと、第 6 章 STRUCTURED（構造化データ解析）で以下の比較方法を整理している（Le Roux & Rouanet2004: 256）。

- 入れ子構造（Nesting Structure 6.2.3）
- 交差構造（Crossing Structure 6.2.4）
- 分散の二重分解（Double Breakdown of variance 6.2.5）
- 加算雲（Additive Cloud 6.3.1）
- 交互作用雲（Interaction Cloud 6.3.2）
- 構造効果（Structural Effect 6.3.3）

8　MCA と ANOVA を接続できる条件

8.1　指示行列というコーディングルール

以上、MCA において分散の分解がどのようにおこなわれるのか、また、その分解した分散を用いて、どのように分析がおこわなれるのか、ということを概観してきた。

ここで、こうした手法が成立するために、遵守する必要がある要件を整理しておく。なぜなら、以下に述べる条件を無視しても、ソフトウェアはなんらかの計算結果を出力するグラフを描画するからである。

上に述べた幾何学的データ解析で、分散分析の手法を活用するためには、まず、以下のコーディング要件がみたされていなくてはならない。

Disjuntive coding（完備排他形式：大隈・ルバールほか 1994、指示行列。LeRoux&Rouanet2010=2021、p172 の用語集の図）が行われているかどうかである。冒頭、確認したように、MCA は、このタイプのデータに対する CA 処理である。この形式を維持していないデータ行列を入力しても、なにかしら計算結果は得られる。しかし、本稿で整理してきたような分散分析（ANOVA）との接続の数理的な保証はなりたたない。

その保証とは、分散の分解式である：

全体の分散 = 群間分散 + 群内分散
（Vtotal = V_between + ΣV_within）

の関係である。この式は、平方和の分解として、回帰分析でも分散分析でも前提とされるが、このようにシンプルに分解できるためには、要素平均が厳密にゼロであることが必要であった。それがなければ、このようなシンプルな構成にはならない。

指示行列は、こうした関係を維持するためのもっとも基本的な前提である。

8. 2　specificMCA による要素平均ゼロの保持

調査データに MCA を実行すると次のような場面に直面することがある。度数が少ないカテゴリが、外れ値として軸をひっぱってしまい、分析対象にしたい領域が密集してしまい評価ができない、という問題である。

こうした事態に直面すると、そうした小度数のカテゴリを外して CA を行うことが考えられるが、このようにした場合、そのカテゴリを選択した個体と選択してない個体の「距離」の違いが表現されてなくなってしまうということが発生する。また、機械的にそのカテゴリを除去することになると、先にのべた指示行列化によって維持されていた、すべての行で同じ周辺度数をもつ、という前提が壊れてしまい、各変数内カテゴリの平均ポイントがゼロではなくなってしまうのである。

ここでふれた低度数カテゴリの扱いは、Le Roux & Rouanet2004 によって、SpecificMCA として定式化されている。GDAtools::speMCA や FavctoMineR::MCA では、excl= で除去するカテゴリ位置を指示することで specificMCA が実行されるだけでよい。

9　個体に対するサブセット MCA としての CSA

また、同様の措置を行（個体）に適用する部分個体 MCA は、CSA（Class Specific Analysis）として定式化されている。

この CSA は、ブルデュー派の社会空間分析において「界」分析のツールとして用いられている（Hjellbrekke2018）。SpecificMCA も CSA も、全体行列に対して、変数カテゴリ列に対する選択的 MCA 処理、また個体行に対する選

択的 MCA 処理である。処理的には、全体との関係を無視してそれぞれ選択した範囲で MCA を実行すればいいように思えるかもしれないが、それでは、specificMCA の場合は、行和一定の関係がくずれ、カテゴリの要素平均が厳密にゼロがなりたたなくなり、ANOVA 的手法を適用する数学的根拠が崩れる。また、選択個体行を独立して MCA することももちろん可能であるが、その場合は、その選択されたサブセットと全体の関係、もしくは異なるサブセット間の関係は崩れることになる。ここは、分析の目的に依存する領域になるが、全体の雲（Gloabal cloud）との関係を論じるアプローチでは CSA は必須となる。

このように構造化を経て、分析対象の解析がすすんでいくが、必要に応じて、次のステップ、分散分析における「F 検定」に相当する領域が用意されている。これは GDA における帰納的データ解析（Inductive data analysis）と呼ばれている。IDA については稿をあらためて説明したい。

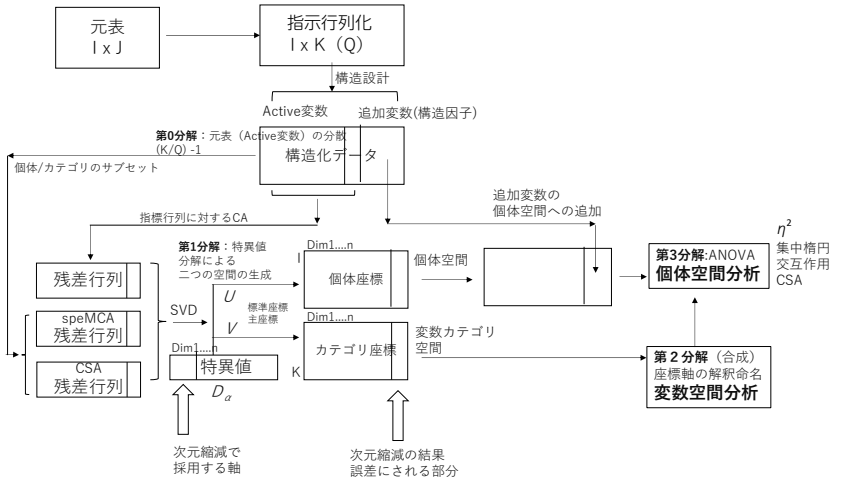


図 14　GDA における分散の分解過程

謝辞

本稿は、拙訳本『対応分析の理論と実践』および大隈・小野・鳰訳の『多重対応分析』を検討素材とした「対応分析研究会」（磯直樹先生主催）での発表、報告を基礎にまとめたものである。1 ヶ月に 1 回というペースでの研究会に

参加してくださり、さまざまなご意見をいただいたことを、参加者の皆様に感謝いたします。

また、本稿は、日本社会学会 95 回全国大会での「研究法・調査法」部会での筆者による発表「幾何学的データ解析（GDA）の中で多重対応分析（MCA）と分散分析（ANOVA）の連携を見る」をもとに加筆修正したものである。大会での発表に対する質問、ご意見に感謝いたします。

なお、本研究は、JSPS 科研費 JP20K0216「データの幾何学的配置に注目したカテゴリカルデータ分析手法の研究」の助成を受けたものです。

注：

1）indicator 行列（指示行列）。ここでのコーディング方式は、Disjunctive Codeing, Crisp coding とも呼ばれる。Le Roux&Rouanet2010=2021 の用語解説 p170 に詳しい説明がある。

2）SVD を用いた次元縮減の事例は、”Metric Scaling”にわかりやすい例が提示されている。以下に仮訳を提示。https://wp.me/p70mJn-s1

3）行列表記の計算式は、付録の 3.1 式～ 3.4 式。および、実際の展開については次の文献を参照のこと。

Clausen1998=2015:188、Greenacre2017=2020:244

4）ベンゼクリの補正については Le Roux&Rouanet2010=2021:56、Greeancre の補正については Greenacre2017=2020:149 を参照。

5）この関係は、Le Roux & Rouanet 2020=2021 では、ホイヘンスの原理と呼ばれているが、回帰分析や分散分析で行われる平方和の分解と原理は同じである。

参考文献

Bourdieu, P.（1979）. *La Distinction: Critique Social du Judgment*, Paris: Editions de Minuit（English translation: Distinction（198A）Boston: Harvard University Press）（石井洋二郎（訳）（1990）「ディスタンクシオン・社会的判断力批判―社会的判断力批判（Ⅰ, Ⅱ）」（藤原書店）

Blasius, Jörg, Frédéric Lebaron, Brigitte Le Roux, Andreas Schmitz, ed.. 2019. *Empirical Investigations of Social Space*. Methodos Series, volume 15. Cham: Springer.

Cramér, H. ,1946, *Mathematical Methods of Statistics*. P〈meton, NY: Princeton University Press.（H. クラメール（著）, 池田貞雄（監訳）, 前田功雄 , 松井敬 訳）（1972, 1973）「統計学の数学的方法（1）,（2）,（3）」（東京図書））

Hjellbrekke, Johs. 2018. *Multiple correspondence analysis for the social sciences*. Routledge.

Le Roux, Brigitte, Rouanet, Henry, 1998, *Interpreting Axes in Multiple Correspondence Analysis: Method of the Contributions of Points and Deviations*, Blasius, Jörg, Michael J Greenacre.ed, 1998, *Visualization of Categorical Data*, CRC press

Le Roux, Brigitte, Rouanet, Henry, 2004, *Geometric Data Analysis: From Correspondence Analysis to Structured Data Analysis*. Dordrecht: Kluwer Academic Publishers

Rouanet, Henry., 2006, *The Geometric Analysis of Structured Individuals x Variables Tables*, “Greenacre, Michael J., Jörg Blasius, ed, 2006, Multiple correspondence analysis and related methods”, pp138-159

Le Roux, Brigitte, Rouanet, Henry, 2010, *Multiple correspondence analysis*. Quantitative applications in the social sciences 163. Thousand Oaks, Calif: Sage Publications（訳：大隅昇・小野裕亮・鳰真紀子 , 2021,『多重対応分析』, オーム社）

Greenacre, Michael J., Jörg Blasius, ed,2006, *Multiple correspondence analysis and related methods*. Statistics in the social and behavioral sciences series. Boca Raton: Chapman & Hall/CRC

Greenacre, Michael, 2017, “*Correspondence Analysis in Practice 3rd Edition*”, CRC press（訳：藤本一男 , 2020,『対応分析の理論と実践 : 基礎・応用・展開』, 東京 : オーム社）

Lebart, L., Morneau, A., & Warwick, KX. M.（1984）. *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices*, New York: Wiley（訳：大隅昇 , L. ルバール , A. モリノウ , K.M. ワーウィック , 馬場康維 , 1994.『記述的多変量解析』（日科技連出版社）

R Core Team（2022）. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

藤本一男 ,

2017,「対応分析のグラフを適切に解釈する条件－ Standard Coordinate, Principal Coordinate を理解する」『津田塾大学紀要』第 49 号、pp141-153

2018,「プログラミング言語 R における 2 つの mosaic plot と日本語、多言語表示」『津田塾大学紀要』第 50 号、pp129-146

2019,「『Supplymentary』変数から多重対応分析（MCA）を考える―幾何学的データ解析（GDA）と多重対応分析（MCA）―」『津田塾大学紀要』第 51 号、pp156-167

2020,「対応分析は〈関係〉をどのように表現するのか　―CA/MCA の基本特性と分析フレームワークとしての GDA ―」『津田塾大学紀要』第 52 号、pp169-184

2022,「日本における「対応分析」受容の現状を踏まえて、EDA（探索的データ解析）の中に対応分析を位置付け、新たなデータ解析のアプローチを実現する」『津田塾大学紀要』第 54 号、pp172-193

付録 MCA、speMCA、CSA の標準化残差行列

共通

- データ行列（個体×変数）を指示行列化する。それをNとする。
- Nの要素をNの総数で割り、対応行列Pとする。

基本形

rをPの行和ベクトル、cをPの列和ベクトルとすると、rc^tが期待値行列を表す。これから、残差は、$P - rc^t$ であらわされるでの、標準化要素としての標準偏差（ポアソン分布を考えるので期待値要素が分散であるため、期待値の平方根）で割って、残差行列 S が得られる。つまり、

$$S = D_r^{-1/2}(P - rc^t)D_c^{-1/2}$$

となる。これを特異値分解（SVD）して得られる行列をU, D, Vとすれば、

$$S = UDV^t$$

である。これをもとに、

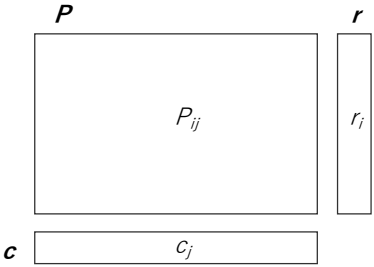
標準座標（Φ, Γ）、主座標(F, G)が求められる。

$$\Phi = D_r^{-1/2}U$$

$$\Gamma = D_c^{-1/2}V$$

$$F = D_r^{-1/2}UD_\alpha = \Phi D_\alpha$$

$$G = D_c^{-1/2}VD_\alpha = \Gamma D_\alpha$$



図A-1 対応行列、行和、列和の基本形

SpecificMCA (列：回答カテゴリ選択 MCA)

除去するカテゴリをはずした対応行列をP'とする。列和ベクトルcの構成は変わるが要素の値に変更はない。しかし、行和ベクトルrについてみれると、カテゴリが除去されているので、行和に変更が生じてうる。これをr'とする。

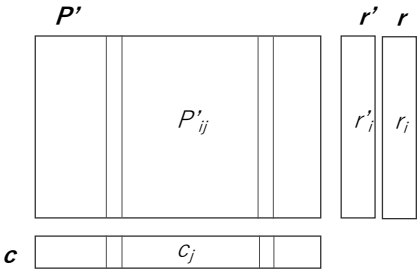
残差を計算する時の期待値要素は、$r'c^t$である。これをP'からひくことで残差が得られる。しかし、標準化する分母に用いるものは、元のPの行和ベクトルのrを用いる。つまり、残差の大きさの評価は、あくまで元のPの値が基準になっているということである。こうして、specificMCA の標準化残差行列は、以下のようになる。

$$S' = D_r^{-1/2}(P' - r'c^t)D_c^{-1/2}$$

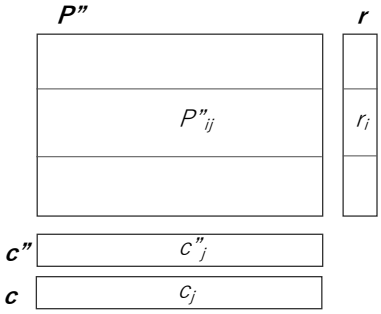
CSA （行：個体選択 MCA）

CSA の場合は、MCA の対象とする個体が選択されている。その対応行列を P"とする。この P"の期待値要素は、行和ベクトルの要素値に変更はない。しかし、行が少なくなっているので、列和ベクトルは変更されてる。これを、c"とする。ここでも、期待値は、rc"t で掲載し、残差は、specificMCA の場合と同じように、それをPから引けばよい。ここでの標準化要素である標準偏差は、もとの列和要素を用いたものを使う。それゆえ、CSA での標準化残差行列は以下のようになる。

$$S' = D_r^{-1/2}(P' - rc^{"t})D_c^{-1/2}$$



図A-2 speMCAでの 対応行列、行和、列和



図A-3 CSAでの 対応行列、行和、列和