VI-2-3-1. 尺度レベルと対応分析の必要性

Stanley Smith Stevens (1946) はデータの表現に使われる尺度を、内容によって(名義尺度、順序策度、間隔尺度、比例尺度)の4つのレベルに分類しました。最もわかりやすい名義尺度の例は、性別、賛否、存否などです。こうしたデータを表現するにもコード番号が使われて、例えば、男性ならば1で女性ならば2、賛成ならば1、反対ならば0、存在すれば1、存在しなければ0のように数値的な表現がされます。順序尺度というのは、順番のことで、競走の到着順がもっともわかりやすいものです。下にあげた例は、質問票調査で使われる質問で、賛成の程度を5件法で尋ねています。このような尺度を Likert scale といいますが、これも順序尺度です。

例

質問:あなたは次の意見に賛成ですが反対ですか。貴方の意見に使いものを次の5つの 選択肢から選んでください。

意見:肉食は健康によくない。

1: 全くそうは思わない 2:そうは思わない, 3: どちらでもない (場合による)

4: そう思う 5: 強くそう思う

実験や観察によって測定されるデータの多くは、間隔尺度あるいは比例尺度で表されてい ます。これらはどちらも定量的なデータに分類されますが、その違いは少しわかりにくいと ころがあります。その違いの説明によく使われるのは摂氏で表された温度と絶対温度との 違いです。どちらも、1度の違いが標準状態の水の融点と沸点の百分の一であるということ は同じですが、0度の定義が違っていて、0℃は標準状態の水の融点です。ですから、0℃は 熱量がない状態を意味しません。0℃は-1℃よりも1度高い温度です。これに対して絶対温 度 0 度 (0K)は物質が熱量を持たない状態のことです。どちらも等間隔に数値がふられて 1 度の大きさに違いはありません。どちらも計量的なデータの表現に使われる尺度ですが、零 度の意味は全く違っています。絶対温度には 0 K 以下の温度はありません。 摂氏の温度は間 隔尺度、絶対温度は比例尺度です。 間隔尺度における 0 は測定のための基準点のようなもの ですが、比例尺度における0は空あるいはない状態です。ですから、負の値はありません。 比例尺度で表現されたものは、その数値の間で比を直接論ずることが出来ます。たとえば、 20℃ は 10℃, よりも 10 度温度が高いとは言えますが、20℃は 10℃よりも熱量が 2 倍ある とは言えません。これに対して、200K の物質は同じ大きさの 100K の物質の 2 倍の熱量を 持つということは言えます。Stevensの尺度レベルの4つの尺度は次のように要約できます。 名義尺度は同じものに属するということだけを表しています。大きさ、速さ、強さなどの比 較はできません。データ間の四則演算は意味を持ちません。したがって、平均、分散、偏差 などの記述統計的な代表値も意味を持ちません。唯一最頻値には意味があります。順序尺度 で記述されたデータも名義尺度で書かれたデータと同様にカテゴリカルデータに分類され ます。確かに、四則演算が意味を持たず、そのままでは計量的に扱えないという意味ではカテゴリカルなのですが、私たちは順序尺度から、あるデータあるいはデータ群は他のデータよりも大きいとか小さいということが出来ます。各データの大きさの相対的な関係を知ることが出来るのです。ですから、順序尺度は計量的な概念を含んでいます。筆者は、順序尺度で書かれたデータは準計量的データだと思っています。順序尺度で書かれたデータは最頻値に加えて中央値も記述統計的な代表値として使えます。間隔尺度で表現されたデータと比例尺度で表現されたデータはともに計量的データですが、間隔尺度では比には意味がありません。数学的にはこの違いが間隔尺度と比例尺度の本質的な違いです。間隔尺度では個々のデータの数値間での引き算や足し算には意味があります。しかし、データ間の掛け算は割り算には意味がありません。これに対して、比例データでは、四則演算すべてに意味があります。間隔尺度で書かれたデータは、記述統計的な代表値として、相加平均が使えて、分散、偏差などにもデータ分布を表す値として意味があります。また、比例尺度で表されたデータでは、これに加えて相乗平均も代表値として使えます。

多変量解析の目的に一つは、データ分布の背景にある構造の理解です。そのために、主成分 分析(PCA) や多次元尺度構成法(MDS)では、成分を直交軸とする空間に個々のデータを位 置付けました。これが出来るのは間隔尺度あるいは比例尺度で表されたデータです。間隔尺 度データ、比例尺度データでは、データ間の引き算が出来て、個々のデータ間の距離・方向 という概念が存在するからです。分布の中心と個々のデータの距離・方向がなければ、デー タの分布を論ずることが出来ません。順序尺度で表されたデータを計量的に数量化する方 法はいくつかあります。たとえば、Thurstone や Scheffe の一対比較法があります。これは、 様々な選択肢を組み合わせて総当たり的により好ましいものを選択させるという一対比較 を行い。各選択肢を1軸上に位置付けるものです。Likert scale のデータの場合にも、いく つかのデータ間の比較で計量的な尺度を作ることは可能だと思いますが、普通は、そのよう なことを行いません。 そんなことをしなくても、事前の情報によって、分析者が、Likert scale が等間隔だと確信できるのならば、何らかの変換をすることなくそのまま、多変量解析をし ても良いでしょう。多くの場合、そのような事前情報はないので、logit 変換して多変量解 析するということが行われます。それがいつでも有効とは限りませんが、logit 変換も等間 隔化のための一つの方法です。同じ、カテゴリカルデータでも、名義尺度で表現されたデー タはそれもできません。それは、名義尺度が大きさの概念を持たないからです。単純化して 言えば、対応分析とは、名義尺度で書かれたデータで行う PCA です。したがって、対応分 析で理解すべき本質は、名義尺度間での距離の定義です。 名義尺度で唯一計量的に使えるの は頻度です。もし、回答者を属性別にいくつかのグループに分けることが出来れば、そのグ ループごとのある質問に対する回答の選択肢が選択された頻度を数値として取ることが出 来ます(観測頻度)、また、全体的頻度とグループの大きさ、グループごとの周辺頻度から 頻度の期待値を計算することが出来ます(期待頻度)。計量データにおける平均値は一種の 期待値です。 期待頻度を名義尺度のおける分布中心と考えれば、 観測頻度と期待頻度の差を

距離として捉えることが出来ます。これが、対応分析の本質です。その他の計算上のアルゴリズムは、対応分析も主成分分析も同じです。すでに、主成分分析の章で、分散共分散行列あるいは相関行列の固有値分解という形で、そのアルゴリズムを説明しました。この解説では、特異値分解によって、対応分析のアルゴリズムを説明しますが、特異値分解は固有値分解を非正則な行列へと拡張したものです。主成分分析も特異値分解でアルゴリズムを説明する方が一般的です。いずれにしても、距離を定義して、その距離を使って、質問に対する回答を一次変換して直交するいくつかの主成分を作り、その主成分で構成される直交空間に回答を位置づけるというのが対応分析です。その本質である、名義データの距離とその標準化という考え方は、すでに、二項分布的なデータの差の有意差検定として紹介したカイニ乗分布で説明したものです。そこで使われているのは、期待値が分散の2分の1になるという関係ですが、その説明は、カイ二乗分布の章を参考にしてください。

VI-2-3-2. 対応分析の目的と結果の表現

対応分析に類する、カテゴリカルデータの数量化・多変量解析はいくつかあって、入力する データの形式や、結果の表現に微妙な違いがあり、それによってアルゴリズムの説明が大き く違います。この解説では、そうした混乱を避けて、それらを一括して、名義尺度データの 主成分分析だと割り切って説明しますが、いくつか異なる方法が成書で説明されているの で混乱があります。そこで、アルゴリズムの説明の前に、対応分析が何をしようとしている のかを例を挙げて確認します。

表 44 はある質問票調査の第一次のデータ集計表です。回答者は5人、質問は3つで、一つの質問について、3つの選択肢から一つの選択肢を選んで回答します。表 44 のように、一人一人の回答者を一つの軸として、5次元のベクトルとして、それぞれの選択肢をみると、Q1-A3 と Q2-A3 と Q3-A1、Q2-A1 と Q3-A3、Q2-A2 と Q3-A2 はそれぞれ同じベクトルです。また、表 45 のように一つの選択肢を1軸として、9次元のベクトルとして回答者を見ると、回答者2と4は同じベクトルです。次元数が多いとわかりにくいので、軸を一次結合して軸の数を減らして、2軸か3軸で、情報量を減らさずにわかりやすくこうした関係を空間的に認識したいというのが対応分析です。図88がその例で、同じベクトルは同じ位置に布置されています。つまり、できるだけ情報量を落とさずに、関連のある選択肢同士をできるだけ近くに布置した散布図を作るというのが対応分析です。

表 44. 多重対応分析で分析するデータの例 (ベクトルの選択肢についてい色分け)

回答者	質問1			質問2			質問3		
凹合有	選択肢 1	選択肢 2	選択肢3	選択肢1	選択肢 2	選択肢3	選択肢1	選択肢2	選択肢3
1	0	1	0	1	0	0	0	0	1
2	1	0	0	1	0	0	0	0	1
3	0	0	1	0	0	1	1	0	0
4	1	0	0	1	0	0	0	0	1
5	1	0	0	0	1	0	0	1	0

表 45. 多重対応分析で分析するデータの例(同一ベクトルの回答者ついてい色分け)

回答者	Q1-A1	Q1-A2	Q1-A3	Q2-A1	Q2-A2	Q2-A3	Q3-A1	Q3-A2	Q3-A3
1	0	1	0	1	0	0	0	0	1
2	1	0	0	1	0	0	0	0	1
3	0	0	1	0	0	1	1	0	0
4	1	0	0	1	0	0	0	0	1
5	1	0	0	0	1	0	0	1	0

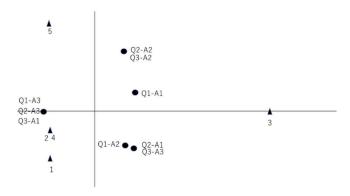


図 88.表 49 データの多重対応分析の結果の図示

別の例を、質問の内容とともに示します。

質問とその集計の例 (Selection from alternatives)
質問: どの食べ物が食べたいですか、食べたい物を下記の中から選んでチェックしてください。いくつ選んでも構いません。
□ 豚肉, □ 牛肉, □ 鶏肉, □ 羊肉, □ マグロ, □ サケ, □ リンゴ,□ ブドウ□ メロン, □ ナシ, □ サクランボ, □ キャベツ, □ レタス, □ キュウリ,

この質問に対する回答は典型的な名義尺度データです。分析者は、表 46 に示したように、たとえば、選択され場合に 1、選択されなかった場合に 0 のコード番号を与えて、データを集約します。回答者は性別、年齢、住所、教育レベル、政治信条、価値観、倫理観など様々な属性を持っています。全データ集団は様々な属性によって仕分けられる複数の小集団によって構成されています。分析者はそれらの属性がどのような関係を持っているのかを分析して、回答者たちの価値観や行動の背景になっている要因を調べようと考えます。そうすることによってまだわかっていない潜在的なメカニズムや理論が発見できるからです。表47 に小集団に分類して集計した集計表の例を示しました。この例で考えているのは比較的単純なグループ分けで、グループは互いに独立していて背反事象です。普通、質問票は複数の質問で構成されていて、質問に対する回答もまた属性の一つとなります。同一の回答をし

表 46. データの集計表の例(0:選択されない, 1:選択された)

	豚肉	牛肉	鶏肉		キュウリ	合計
回答者 1	1	0	1	•••	0	n_1 .
回答者 2	0	0	1		1	n_2 .
:	:	:	:	٠,	:	÷
回答者 n	1	0	0	•••	1	n_n .
合計	$n_{\cdot 1}$	n. ₂	n. ₃		$n_{\cdot m}$	$N = \sum_{j=1}^{m} n_{.j} = \sum_{i=1}^{n} n_{i}.$

表 47. 属性ごとのグループに分けた頻度データの集計表

	選択肢 1	選択肢 2	選択肢3		選択肢 m	Sum
グループ1	n_{11}	n_{12}	n ₁₃		n_{1m}	$\sum_{j=1}^{m} n_{1j}$
グループ 2	n_{21}	n ₂₂	n ₂₃		n_{2m}	$\sum_{j=1}^{m} n_{2j}$
:	:	:	:	N	:	
グループ n	n_{l1}	n_{l2}	n_{l2}		n_{lm}	$\sum_{j=1}^{m} n_{nj}$
Sum	$\sum_{i=1}^n n_{i1}$	$\sum_{i=1}^n n_{i2}$	$\sum_{i=1}^{n} n_{i3}$		$\sum_{i=1}^n n_{nm}$	$\sum_{i=1}^{n} \sum_{j=1}^{m} n_{ij}$

た人々をまとめると、それを属性として小集団を作れます。質問と回答の数だけの属性が考えられますから、グループ分けの組み合わせはもっと複雑なものになります。また、回答として示された属性は、分析によって説明しようとする目的変数であるとともに、空間構造を作り出す説明変数でもあります。主成分分析では、各項目のデータの一次結合で得られる成分を主成分として、複数の主成分で構成される直交空間に、データを布置して散布図を作ります。それが出来るのは、PCAが計量的なデータで、データ間の違い(差)を距離として捉えることが出来るからです。対応分析(CA)は名義尺度データで行う主成分分析(PCA)だと考えて良いでしょう。各選択肢に対する反応を一次結合して得られる成分度構成される直交空間に選択肢を布置します。違いはデータを表現している尺度です。ですから、対応分析の理解で一番基本的なことは、名義尺度データにおける距離の定義です。

VI-2-3-3. 名義尺度データの多変量解析における距離

主成分分析(PCA) は計量的なデータの分析で、データがもともと計量的な空間に存在し、その空間が距離という概念を持っています。ですから、分布の中心を平均によって求めることが可能です。データの分布は平均として求めた中心からの距離と方向で表せます。分散や偏差も分布特性を表す代表値として使えます。これらを使って分散・共分散行列を作り、分散や標準偏差によって、標準化された距離でデータの分布を表すこともできます。名義尺度には距離の概念がないので、これが出来ないということが問題なのですが、名義尺度でも頻度は計量的に扱うことが出来ます。頻度は量的に扱えますし、周辺頻度から頻度の期待値

(期待頻度)を計算することもできます。カイ二乗検定では、期待頻度と実測頻度の差を距離として捉えて、その距離が得られる確率をカイ二乗分布として計算します。表 47 のデータの期待値は、周辺頻度を使って以下のように計算できます。

$$e_{ij} = \frac{\sum_{i=1}^{k} n_{ji} \cdot \sum_{j=1}^{p} n_{ji}}{\sum_{i=1}^{k} \sum_{j=1}^{p} n_{ij}}$$

eii:期待頻度

 $\sum_{i=1}^k n_{ii}$: 列の周辺頻度

 $\sum_{i=1}^{p} n_{ii}$: 行の周辺頻度

$$\sum_{i=1}^k \sum_{j=1}^p n_{ij}$$
: 全度数

すでに平均というのはデータから求めたデータ分布中心の期待値だという説明はしました。 この場合にも期待値は分布を代表する中心値と考えられますから、分布中心からの距離は 以下の式で表せます。

$$d_{ij} = n_{ij} - e_{ij}$$

そこで、表48のように期待値の集計表を作ります。

統計処理のためにはこの距離を標準偏差のような何らかの基準で標準化する必要があります。カイ二乗値は次の式で求めます。

$$\chi^2_{ij} = \frac{\left(n_{ij} - e_{ij}\right)^2}{e_{ij}}$$

カイ二乗分布の説明で、期待値は分散の2分の1だという説明をすでにしました。ここでは その説明を繰り返しませんので、詳しく知りたい人はカイ二乗分布の章を読んでください。 これを使ってカイ二乗に式を書き換えると以下のようになります。

$$\chi^2 = \sum \frac{(f_i - e_i)^2}{e_i} = \sum \frac{2(f_i - e_i)^2}{{\sigma_i}^2}$$

表 48. 期待値の集計表

cubnonulation	alternative							
subpopulation	1	2	3	•••	р			
1	e_{11}	e_{12}	e_{13}	•••	e_{1p}			
2	e_{21}	e_{22}	e_{23}	•••	e_{2p}			
3	e_{31}	e_{32}	e_{33}	•••	e_{3p}			
:	:	:	:	•	÷			
k	e_{k1}	e_{k2}	e_{k3}	•••	e_{kp}			

ここで、分子の2を定数として取り除くと、以下の式が出来ます。この式は、標準偏差で標準化した距離です。

$$\left(\frac{f_{ij}-e_{ij}}{\sigma_{ij}}\right)^2$$

これを使えば、標準化された距離は以下のように表せます。

$$\frac{f_{ij} - e_{ij}}{\sigma_{ij}} = \frac{f_{ij} - e_{ij}}{\sqrt{e_{ij}}}$$

筆者はこの式で表された数値を何と呼ぶのか正しい名前を知りません。個人的にはカイ二 乗距離と呼んでいます。いずれにしてもこれによって、名義尺度のデータを頻度データとし て計量的に扱うことが出来るようになります。「カイ二乗距離」の定義式は以下の通りです。

$$z_{ij} = \frac{f_{ij} - e_{ij}}{\sqrt{e_{ij}}}$$

$$f_{ij} : 觀測頻度$$

$$e_{ij} : 期待頻度$$

$$z_{ij} : 力イ二乗距離$$

これで、選択肢に対する反応の一次結合で作られた主成分の直交空間にグループの実測値と選択肢を布置することが出来ます。これを受け入れると、個々の回答者を直交空間に位置付けることもできます。何故ならば、個々の回答者は最小単位の小グループだと考えることができるからです。そうすると、集計表は表 49 のように、0 と 1 のコード番号だけで作られることになります。表 49 の集計表から、表 50 期待値の集計表ができ、最終的に表 51 のカイ二乗距離の集計表が作れます。表 51 から行列**Z**を作ります。

回答者	質問1				質問 2			質問3			
四合有	Q1-A1	Q1-A2	Q1-A3	Q2-A1	Q2-A2	Q2-A-3	Q3-A1	Q3-A2	Q3-A3	合計	
1	0	1	0	1	0	0	0	0	1	3	
2	1	0	0	1	0	0	0	0	1	3	
3	0	0	1	0	0	1	1	0	0	3	
4	1	0	0	1	0	0	0	0	1	3	
5	1	0	0	0	1	0	0	1	0	3	
合計	3	1	1	3	1	1	1	1	3	15	

表 49. 回答者の反応の集計 (表 44 のデータ)

表 50. 期待値の集計表

回答者		質問 1		質問2 質問3					
凹合有	Q1-A1	Q1-A2	Q1-A3	Q2-A1	Q2-A2	Q2-A-3	Q3-A1	Q3-A2	Q3-A3
1	0.6	0.2	0.2	0.6	0.2	0.2	0.2	0.2	0.6
2	0.6	0.2	0.2	0.6	0.2	0.2	0.2	0.2	0.6
3	0.6	0.2	0.2	0.6	0.2	0.2	0.2	0.2	0.6
4	0.6	0.2	0.2	0.6	0.2	0.2	0.2	0.2	0.6
5	0.6	0.2	0.2	0.6	0.2	0.2	0.2	0.2	0.6

表 51. カイ二乗距離の集計表

回答者		質問 1		質問 2 質問 3					
凹合有	Q1-A1	Q1-A2	Q1-A3	Q2-A1	Q2-A2	Q2-A-3	Q3-A1	Q3-A2	Q3-A3
1	-0.77	1.789	-0.45	0.516	-0.45	-0.447	-0.45	-0.45	0.516
2	0.516	-0.45	-0.45	0.516	-0.45	-0.447	-0.45	-0.45	0.516
3	-0.77	-0.45	1.789	-0.77	-0.45	1.789	1.789	-0.45	-0.77
4	0.516	-0.45	-0.45	0.516	-0.45	-0.447	-0.45	-0.45	0.516
5	0.516	-0.45	-0.45	-0.77	1.789	-0.447	-0.45	1.789	-0.77

$$\mathbf{Z} = \begin{pmatrix} -0.77 & 1.789 & -0.45 & 0.516 & -0.45 & -0.45 & -0.45 & 0.516 \\ 0.516 & -0.45 & -0.45 & 0.516 & -0.45 & -0.45 & -0.45 & -0.45 & 0.516 \\ -0.77 & -0.45 & 1.789 & -0.77 & -0.45 & 1.789 & 1.789 & -0.45 & -0.77 \\ 0.516 & -0.45 & -0.45 & 0.516 & -0.45 & -0.45 & -0.45 & -0.45 & 0.516 \\ 0.516 & -0.45 & -0.45 & -0.77 & 1.789 & -0.45 & -0.45 & 1.789 & -0.77 \end{pmatrix}$$

VI-2-3-4. 特異値分解(SVD)を使った多重対応分析(MCA)のアルゴリズム

特異値分解 (SVD) は非正則行列を特異値を対角成分とする対角行列に変換する方法で、次の式で表せます。

$$\boldsymbol{U}^T \boldsymbol{Z} \boldsymbol{V} = \boldsymbol{\Sigma}$$

 $Z: k \times p$ 非正則行列

Σ: 特異値を対角成分とする対角行列

U:左特異射影子

V:r右特異射影子

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_r & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}_{p \times k}$$

$$\sigma_j \colon \cancel{F} \cancel{E} \cancel{E}$$

$$r \colon \cancel{D} \cancel{D} \cancel{D} \cancel{D}$$

各回答者に対する主成分の負荷は、左特異射影子の行列を縦ベクトルを横に並べた形で取り出すことが出来ます。

$$\mathbf{u} = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1k} \\ u_{21} & u_{22} & \cdots & u_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ u_{k1} & u_{k2} & \cdots & u_{kk} \end{pmatrix}_{k \times k}$$

$$\downarrow$$

$$\begin{pmatrix} u_{11} \\ u_{21} \\ \vdots \\ u_{k1} \end{pmatrix}, \begin{pmatrix} u_{12} \\ u_{22} \\ \vdots \\ u_{k2} \end{pmatrix}, \cdots, \begin{pmatrix} u_{1k} \\ u_{2k} \\ \vdots \\ u_{kk} \end{pmatrix}$$

各選択肢に対する主成分の負荷は、右特異射影子の行列を縦ベクトルを横に並べたものと

して取り出すことが出来ます。

$$V = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1p} \\ v_{21} & v_{22} & \cdots & v_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ v_{p1} & v_{p2} & \cdots & v_{pp} \end{pmatrix}_{p \times p}$$

$$\downarrow$$

$$\begin{pmatrix} v_{11} \\ v_{21} \\ \vdots \\ v_{n1} \end{pmatrix}, \begin{pmatrix} v_{12} \\ v_{22} \\ \vdots \\ v_{n2} \end{pmatrix}, \cdots, \begin{pmatrix} v_{1p} \\ v_{2p} \\ \vdots \\ v_{nn} \end{pmatrix}$$

数学的には、ランク数 (r) 個の主成分があることになりますが、必要な情報だけに絞って簡略化して情報を集約することが目的なので、小さな特異値は近似的に 0 として、大きな特異値に属する主成分に着目して、2 つないし 3 つの主成分で、二次元、三次元の分布図を作るのが普通です。

付記 (SVD での射影子の求め方)

特異値分解は2つの固有値分解の組み合わせで、射影子は固有値分解で求めることが出来 ます。これはすでに、特異値分解の章で説明したことですが、ここで例を挙げてもう一度説 明します。

具体例:具体亭な例として、非正則行列**Z**を特異値分解の例を示します。

$$\mathbf{Z} = \begin{pmatrix} -0.77 & 1.789 & -0.45 & 0.516 & -0.45 & -0.45 & -0.45 & 0.516 \\ 0.516 & -0.45 & -0.45 & 0.516 & -0.45 & -0.45 & -0.45 & -0.45 & 0.516 \\ -0.77 & -0.45 & 1.789 & -0.77 & -0.45 & 1.789 & 1.789 & -0.45 & -0.77 \\ 0.516 & -0.45 & -0.45 & 0.516 & -0.45 & -0.45 & -0.45 & 0.516 \\ 0.516 & -0.45 & -0.45 & -0.77 & 1.789 & -0.45 & -0.45 & 1.789 & -0.77 \end{pmatrix}$$

$$\mathbf{Z}^{T} = \begin{pmatrix}
z_{11} & z_{12} & \cdots & z_{2j} & \cdots & z_{2p} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
z_{i1} & z_{i2} & \cdots & z_{ij} & \cdots & z_{ip} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
z_{k1} & z_{k2} & \cdots & z_{kj} & \cdots & z_{kp}
\end{pmatrix}_{k \times p}$$

$$\mathbf{Z}^{T} = \begin{pmatrix}
z_{11} & z_{21} & \cdots & z_{i1} & \cdots & z_{k1} \\
z_{12} & z_{22} & \cdots & z_{i2} & \cdots & z_{k2} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
z_{1j} & z_{2j} & \cdots & z_{ij} & \cdots & z_{kj} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
z_{1p} & z_{2p} & \cdots & z_{ip} & \cdots & z_{kp}
\end{pmatrix}_{k \times p}$$

$$\mathbf{Z}^{T} = \begin{pmatrix}
-0.77 & 0.516 & -0.77 & 0.516 & 0.516 \\
1.789 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & 1.789 & -0.45 & -0.45 \\
-0.45 & -0.45 & 1.789 & -0.45 & -0.45 \\
-0.45 & -0.45 & 1.789 & -0.45 & -0.45 \\
-0.45 & -0.45 & 1.789 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 \\
-0.45 & -0.45 & -0.45 & -0.45 \\$$

左特異値射影子

$$\mathbf{Z}\mathbf{Z}^T = \begin{pmatrix} 5.33 & 0.33 & -3 & 0.33 & -3 \\ 0.33 & 2 & -3 & 2 & -1.33 \\ -3 & -3 & 12 & -3 & -3 \\ 0.33 & 2 & -3 & 2 & -1.33 \\ -3 & -1.33 & -3 & -1.33 & 8.67 \end{pmatrix}$$

対角化

$$U^T Z Z^T U = \Lambda$$

Λ: 固有值行列

右特異射影子

$$\mathbf{Z}^T\mathbf{Z} = \begin{pmatrix} 2 & -1.73 & -1.73 & 0.33 & 1.15 & -1.73 & -1.73 & 1.15 & 0.33 \\ -1.73 & 4 & 1 & 1.15 & 1 & 1 & 1 & 1 & 1.15 \\ -1.73 & 1 & 4 & -1.73 & 1 & 4 & 4 & 1 & -1.73 \\ 0.33 & 1.15 & -1.73 & 2 & -1.73 & -1.73 & -1.73 & 2 \\ 1.15 & 1 & 1 & -1.73 & 4 & 1 & 1 & 4 & -1.73 \\ -1.73 & 1 & 4 & -1.73 & 1 & 4 & 4 & 1 & -1.73 \\ -1.73 & 1 & 4 & -1.73 & 1 & 4 & 4 & 4 & -1.73 \\ 1.15 & 1 & 1 & -1.73 & 1 & 4 & 4 & 4 & -1.73 \\ 0.33 & 1.15 & -1.73 & 2 & -1.73 & -1.73 & -1.73 & 2 \end{pmatrix}$$

 Z^TZ の対角化

$$V^T Z^T Z V = \Lambda$$

$$\mathbf{V} = \begin{pmatrix} 0.22 & 0.18 & -0.46 & 0.48 & 0 & 0 & 0 & 0 & 0.69 \\ 0.13 & -0.32 & 0.80 & 0.28 & 0 & 0 & 0 & 0 & 0.40 \\ -0.52 & 0 & 0 & -0.06 & 0.48 & -0.70 & -0.07 & 0.06 & 0.21 \\ 0.22 & -0.33 & -0.13 & -0.46 & -0.45 & -0.32 & -0.11 & 0.48 & 0.25 \\ 0.13 & 0.57 & 0.23 & -0.14 & -0.16 & -0.25 & 0.70 & 0.07 & 0.06 \\ -0.52 & 0 & 0 & 0.49 & -0.65 & -0.16 & -0.01 & 0.09 & -0.17 \\ -0.52 & 0 & 0 & 0.44 & -0.19 & 0.40 & 0.10 & -0.32 & 0.48 \\ 0.13 & 0.57 & 0.23 & -0.14 & -0.21 & -0.17 & -0.67 & -0.24 & 0.06 \\ 0.22 & -0.33 & -0.13 & -0.03 & -0.19 & -0.41 & 0.16 & -0.77 & -0.05 \end{pmatrix}$$

Λ: 固有值行列

$$A = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix} = \mathbf{\Sigma}^2 = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 & \cdots \\ 0 & \sigma_2^2 & \cdots & 0 & \cdots \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots & \bullet \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots & \bullet \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots & \bullet \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots & \bullet \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots & \bullet \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots & \bullet \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots & \bullet \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots & \bullet \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots & \bullet \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots & \bullet \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots & \bullet \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots & \bullet \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots & \bullet \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots & \bullet \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots & \bullet \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots & \bullet \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots & \bullet \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots & \bullet \\ 0 & 0 & \cdots & \sigma_T^2 & \cdots$$

特異値分解になっていることを確認します。

$$Z = U \Sigma V^T$$

0

2.015956/

$$\mathbf{Z} = \begin{pmatrix} -0.77 & 1.789 & -0.45 & 0.516 & -0.45 & -0.45 & -0.45 & -0.45 & 0.516 \\ 0.516 & -0.45 & -0.45 & 0.516 & -0.45 & -0.45 & -0.45 & -0.45 & 0.516 \\ -0.77 & -0.45 & 1.789 & -0.77 & -0.45 & 1.789 & 1.789 & -0.45 & -0.77 \\ 0.516 & -0.45 & -0.45 & 0.516 & -0.45 & -0.45 & -0.45 & -0.45 & 0.516 \\ 0.516 & -0.45 & -0.45 & -0.77 & 1.789 & -0.45 & -0.45 & 1.789 & -0.77 \end{pmatrix}$$

$$=\begin{pmatrix} -0.22 & -0.47 & 0.73 \\ -0.22 & -0.18 & -0.46 \\ 0.89 & 0 & 0 \\ -0.22 & -0.18 & -0.46 \\ -0.22 & 0.84 & 0.20 \end{pmatrix}\begin{pmatrix} -3.87298 & 0 & 0 \\ 0 & 3.30695 & 0 \\ 0 & 0 & 2.015956 \end{pmatrix}$$

$$\begin{pmatrix} 0.22 & 0.13 & -0.52 & 0.22 & 0.13 & -0.52 & -0.52 & 0.13 & 0.22 \\ 0.18 & -0.32 & 0 & -0.33 & 0.57 & 0 & 0 & 0.57 & -0.33 \\ -0.46 & 0.80 & 0 & -0.13 & 0.23 & 0 & 0 & 0.23 & -0.13 \end{pmatrix}$$

VI-2-3-4. 散布図に選択肢を布置する.

ランク数から、例に挙げたデータには3つの主成分があり、3次元的な構造で分布していることがわかります。各選択肢の各主成分への負荷は表8のように表せます。また、各回答者の負荷は表9のように表せます。この結果を二次平面の散布図で表すには、3つの内から2つの主成分を取り出さなくてはなりませんが、その場合、特異値の大きな主成分を選ぶのが普通でしょう。この例では全体の分散(固有値の総和)は30であり、第一主成分、第二主成分、第三主成分の分散(固有値)はそれぞれ15、10.94、4.06です。全体分散に対する比をとると、全分散の86%は第一、第二主成分で説明できることがわかります。それを根拠に、第三主成分を無視して、第一、第二主成分で二次元平面上の散布図で選択肢間の関係や回答者の関係を表すことは、十分な妥当性があると筆者は思います。この考えに基づいて、第一主成分を横軸、第二主成分を縦軸とする二次平面上での各選択肢の座標は以下の通りとなります。

Q1-A1: (0.22 0.18), Q1-A2: (0.13 -0.32), Q1-A3: (-0.52 0), Q2-A1: (0.22 -0.33), Q2-A2: (0.13 0.57), Q1-A3: (0.52 0), Q3-A1: (-0.52 0), Q3-A2: (0.13 0.57), Q1-A3: (0.22 -0.33)

表 52. 各選択肢の主成分への負荷

選択肢	第一主成分	第二主成分	第三主成分
Q1-A1	0.22	0.18	-0.46
Q1-A2	0.13	-0.32	0.80
Q1-A3	-0.52	0	0
Q2-A1	0.22	-0.33	-0.13
Q2-A2	0.13	0.57	0.23
Q2-A3	-0.52	0	0
Q3-A1	-0.52	0	0
Q3-A2	0.13	0.57	0.23
Q3-A3	0.22	-0.33	-0.13

表 53. 回答者の主成分への負荷

回答者	第一主成分	第二主成分	第三主成分
1	-0.22	-0.47	0.73
2	-0.22	-0.18	-0.46
3	0.89	0	0
4	-0.22	-0.18	-0.46
5	-0.22	0.84	0.20

追加情報

この解説では、表 49 の集計表を最初の行列としてそれを起点に特異値分解によって MCA を行いました。表 49 は極めて単純な集計表ですが、質問項目や選択肢が多い場合には大きな表になってしまいます。質問項目が多い場合は表 54 の形式でデータを集約する方が一般的かもしれません。一般に使われている多変量解析のコンピュータ・ソフトウェアーたとえば R のパッケージ FactoMineR の MCA では表 54 の形式のデータを読み込んで直接 MCA をすることが出来ます。

表 54.データの集計例

回答者	質問1	質問2	質問3
1	Q1-A2	Q2-A1	Q3-A3
2	Q1-A1	Q2-A1	Q3-A3
3	Q1-A3	Q2-A3	Q3-A1
4	Q1-A1	Q2-A1	Q3-A3
5	Q1-A1	Q2-A2	Q3-A2