

社会システム分析のための統合化プログラム 25

ー パス解析・多次元尺度構成法・数量化IV類 ー

福井正康・尾崎誠・朝日亮太
福山平成大学経営学部経営学科

概要：我々は教育分野での利用を目的に社会システム分析に用いられる様々な手法を統合化したプログラム College Analysis を作成してきた。今回は変数間の影響の方向性と強さを求めるパス解析と定義された距離からデータの多次元空間での配置を求める多次元尺度構成法及び数量化IV類についてプログラムを作成した。この論文ではこれらの分析の理論を説明し、具体的なプログラムの利用法を紹介する。

キーワード：College Analysis, 多変量解析, パス解析, 多次元尺度構成法, 数量化IV類

URL：<http://www.heisei-u.ac.jp/ba/fukui/>

1. はじめに

これまで我々は多変量解析に含まれる多くの分析についてプログラム開発を行ってきたが、それに連れてメニューの分析名が増え、一度にすべてを表示することが難しくなってきた。そこで我々はメニューを検討し、分類分けを行うことにした。多変量解析には様々な用途があるが、我々はこれを、複数の変数で1つの量的な変数を予測するもの、複数の変数で1つの質的な変数を予測するもの、複数の変数から特徴的な量を見つけ出し、グループ分けをするもの、複数の変数間の関係を見出すもの、の4つのカテゴリに分け、これらをそれぞれ予測手法、判別手法、分類手法、関係分析手法と名付けて1段目のメニューとした。各分析は最も近いと思われるカテゴリに分類して、そのメニューの下に入れた。もちろんこの分類には問題があることも考えられるので、今後不都合が生じた場合、方針は継続しつつ、適宜変更して行く。これまでのメニューと少し異なるが、分析名はこれまで通り残すことになるので、混乱は生じないものと思う。

今回作成したプログラムは、まず変数間の線形関係を使って各変数の影響の方向や強さを考えるパス解析^[1]である。さらに、定義されたデータ間の距離または関連性からデータの多次元空間での配置を求め、データの間関係を調べる計量的及び非計量的多次元尺度構成法^{[1],[2]}である。また、準計量的多次元尺度構成法として知られる数量化IV類^[2]についてもプログラムを作成した。

2. パス解析

パス解析は観測変数間に線形の関係が仮定されるとき、因果関係の方向と強さを調べるために利用される手法で、共分散構造分析の特別な場合に相当する。ここではプログラムを実行し、動作を見ながら、理論についても解説する。

メニュー [分析ー多変量解析等ー関係分析手法ーパス解析] を選択すると、図 2.1 のようなパス解析実行メニューが表示される。

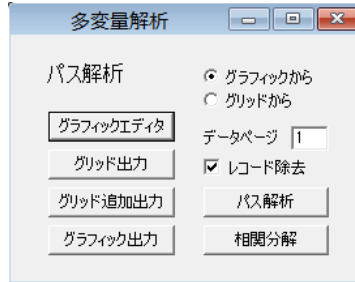


図 2.1 パス解析実行メニュー

「グラフィックエディタ」ボタンで、グラフィックエディタを起動し、例えば図 2.2 のような構造図を描く。

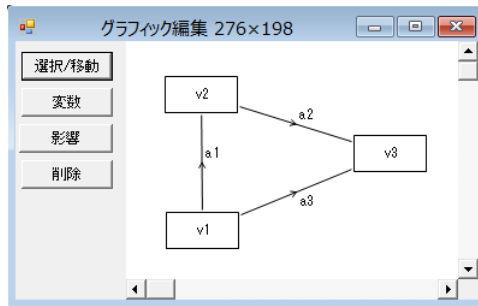


図 2.2 構造図

共分散構造分析の構造図では誤差変数についても描画するが、パス解析では誤差変数の入り方は明らかであるため描画しない。図は単純に観測変数とそれらの間の影響だけで描かれる。但し、影響はすべての変数を結ぶものとし、影響のループは含まないものとする。

これらの変数名のデータは、実行メニューの「データページ」テキストボックスに指定された、グリッドエディタのページに含まれるものとする。プログラムはデータページの変数の中で、変数名に合うデータを利用する。

変数間の構造データは、ラジオボタンにより、グリッドエディタとグラフィックエディタのどちらかを選ぶことができる。通常は、グラフィックエディタからの入力にしておき、良い構造が出来上がったら、「グリッド出力」か「グリッド追加出力」によって、構造データをグリッドエディタに移し、保存する。

実行メニューで「パス解析」ボタンをクリックすると図 2.3 のように、構造間の影響の強さが表示される。

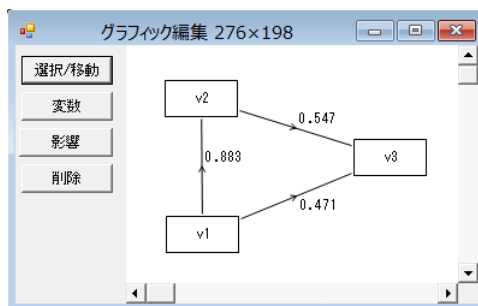


図 2.3 パス解析結果

これを見て我々は影響の強さ、影響の方向のよし悪しを判定する。これらの影響の強さの数値は以下のような標準化した重回帰式から求められる。

$$v2 = a1 * v1 + e2$$

$$v3 = a2 * v1 + a3 * v2 + e3$$

ここで、e2 と e3 は誤差項であり、自分自身を除いて他の変数との相関はないものとする。

これらの式から、各変数の相関について以下のような関係が分かる。

$$\text{cov}(v1, v2) = \text{cov}(v1, a1 * v1 + e2) = a1 * \text{cov}(v1, v1) + \text{cov}(v1, e2) = a1$$

$$\text{cov}(v1, v3) = \text{cov}(v1, a2 * v1 + a3 * v2 + e3) = a2 * \text{cov}(v1, v1) + a3 * \text{cov}(v1, v2) = a2 + a1 * a3$$

$$\text{cov}(v2, v3) = \text{cov}(v2, a3 * v2 + a2 * v1 + e3) = a3 * \text{cov}(v2, v2) + a2 * \text{cov}(v2, v1) + \text{cov}(e3, v2) = a3 + a1 * a2$$

第1式について a1 を直接相関、第2式について、a2 を直接相関、a1*a3 を間接相関、第3式について、a3 を直接相関、a1*a2 を擬似相関と呼ぶ。直接相関は変数間を直接的に結ぶ関係、間接相関は変数間の影響を、及ぼす方向通りにたどって行って2回以上でたどりつく関係、擬似相関は他の変数（ここでは v1）が両者に影響を及ぼしているような関係である。

左辺は相関係数であるので、これらの式は相関係数を、直接相関、間接相関、擬似相関に分解することに相当する。この関係は、実行メニューの「相関分解」をクリックすることで示される。結果を図2.4に示す。

	パス係数	相関係数	直接効果	間接効果	擬似相関
v1->v2	0.8833	0.8833	0.8833	0	0
v1->v3	0.4712	0.9544	0.4712	0.4832	0
v2->v3	0.547	0.9632	0.547	0	0.4162

図 2.4 相関分解

直接効果、間接効果、擬似相関の合計が相関係数になっていることが分かる。

次に、もう少しだけ複雑なモデルを使って、これらの計算法を考えてみる。図2.5にモデルを示すが、ここでは枠のウィンドウの表示は省略する。

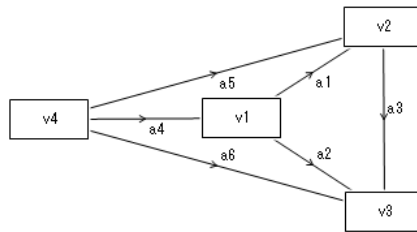


図 2.5 パスの例 2

ここではこの例を用いて v2, v3 への v4 の擬似相関を見てみよう。重回帰分析の計算を用いると、v2 と v3 の相関係数は以下のように与えられる。

$$\text{cov}(v2, v3) = \text{cov}(v2, a3 * v2 + a2 * v1 + a6 * v4)$$

$$= a3 + a1 * a2 + a5 * a6 + a5 * a2 * a4 + a6 * a1 * a4$$

最初の項は直接相関、次の項は v1 からの擬似相関、下線の項は v4 からの擬似相関とみる。ここで、v1 からの擬似相関はさらに、a4 を経由して v4 からの影響 a1*a2*a4*a4 があるように思われるが、同じ変数 v1 を通る場合は、以下に示すように、係数の掛け算 a1*a2*a4*a4 だけが現れるわけではない。

$$a1 * a2 = a1 * a2 * \text{cov}(v1, v1) = a1 * a2 * \text{cov}(a4 * v4 + e1, a4 * v4 + e1) = a1 * a2 * (a4 * a4 + \text{cov}(e1, e1))$$

cov(e1, e1) の値は直接には分からないが、観測値は標準化されており cov(v1, v1)=1 であるので、値を求めることができる。

3. 多次元尺度構成法

多次元尺度構成法 (MDS: Multi Dimensional Scaling) は個体間に与えられた、類似度または非類似度 (距離) を元に各個体の位置 (嗜好性等抽象的な位置関係も含む) を求める手法である。個体間の非類似度がユークリッド空間上の距離として与えられる場合を計量 MDS、非類似度が順序のみ意味を持つ場合を非計量 MDS と呼ぶ。我々はこれらの手法を順番に説明する。

3.1 計量 MDS

個体 i と個体 j ($1 \leq i, j \leq n$) の距離を d_{ij} とし、距離が以下の関係を満たすとき計量 MDS の手法が利用できる。

$$\begin{aligned} d_{ij} &\geq 0 \\ d_{ij} &= d_{ji} \\ d_{ij} + d_{jk} &\geq d_{ik} \end{aligned} \quad (3.1)$$

今、 p 次元のユークリッド空間中の個体 i の位置を $x_{i\alpha}$ ($1 \leq \alpha \leq p < n$) とする。個体 i と個体 j との距離 d_{ij} は以下のように求められる。

$$d_{ij} = \left(\sum_{\alpha=1}^p (x_{i\alpha} - x_{j\alpha})^2 \right)^{1/2} \quad (3.2)$$

この距離は原点の取り方に依存しないので、原点を個体の重心に設定するものとする。そのとき、

$$\bar{x}_\alpha = \frac{1}{n} \sum_{i=1}^n x_{i\alpha} = 0 \quad (3.3)$$

である。原点から個体 i, j へのベクトルの内積を z_{ij} とすると、これは余弦定理により、以下のように与えられる。

$$z_{ij} = \sum_{\alpha=1}^p x_{i\alpha} x_{j\alpha} = \frac{1}{2} (d_{i0}^2 + d_{j0}^2 - d_{ij}^2) \quad (3.4)$$

ここに、 d_{i0} は原点から個体 i までの距離である。(3.3) の関係式を使うと以下となるが、

$$\sum_{i=1}^n z_{ij} = \sum_{j=1}^n z_{ij} = \sum_{i=1}^n \sum_{j=1}^n z_{ij} = 0$$

これに(3.4)式を代入して、 d_{i0} についての関係式を求め、 z_{ij} を以下のように書き換えることができる。

$$z_{ij} = \frac{1}{2} \left(\sum_{k=1}^n \frac{d_{kj}^2}{n} + \sum_{k=1}^n \frac{d_{ik}^2}{n} - \sum_{k=1}^n \sum_{l=1}^n \frac{d_{kl}^2}{n^2} - d_{ij}^2 \right) \quad (3.5)$$

我々は求められた距離行列から、(3.5)式によってこの内積で作られた対称行列 \mathbf{Z} を求め、(3.4)の最初の等号関係を用いて、後に示す方法で位置 $x_{i\alpha}$ を求める。

対称行列 \mathbf{Z} は p 個の固有値 λ_α (≥ 0) を対角成分に並べた対角行列 $\mathbf{\Lambda}(p \times p)$ とその固有値に対する固有ベクトル \mathbf{y}_α を横に並べた行列 $\mathbf{Y} = (\mathbf{y}_1 \cdots \mathbf{y}_p)$ によって以下のように分解できる。

$$\mathbf{Z} = \mathbf{Y} \mathbf{\Lambda}' \mathbf{Y} \quad (3.6)$$

今、固有値の平方根を対角成分に並べた対角行列を $\Lambda^{1/2}$ として、 $\mathbf{X} = \mathbf{Y}\Lambda^{1/2}$ とおくと、上の関係式は以下ようになる。

$$\mathbf{Z} = \mathbf{X}'\mathbf{X} \quad (3.7)$$

これを(3.4)式と比較すると、以下の関係を得る。

$$x_{i\alpha} = \sqrt{\lambda_\alpha} y_{i\alpha} \quad (3.8)$$

これで個体の空間配置が決まる。

3.2 非計量 MDS

非計量 MDS では、非類似度 s_{ij} ($s_{ij} = s_{ji}$) を用いるが、これをディスペリティと呼ばれる量 \hat{d}_{ij} に変換して利用する。これらは以下の関係を満たすようにする。

$$\begin{aligned} s_{ij} > s_{kl} &\Rightarrow \hat{d}_{ij} \geq \hat{d}_{kl} \\ s_{ij} = s_{kl} &\Rightarrow \hat{d}_{ij} = \hat{d}_{kl} \end{aligned} \quad (3.9)$$

ディスペリティの生成は参考文献[2] に示された以下の手順で行う。ある手法で(我々のプログラムでは非類似度 s_{jk} を用いた計量 MDS の手法)、位置が求まっているとする。その位置から距離 d_{jk} を求める。非類似度 s_{jk} ($1 \leq j < k \leq n$) を小さい順に並べ、それを s_1, s_2, \dots, s_l とする。 s_i に対応する非類似度 s_{jk} に対応する距離 d_{jk} についても s_i と同じ順番で d_1, d_2, \dots, d_l としておく。但し、 s_i に同順位のものがある場合、それに対応する d_i について、平均をとっておくものとする。

この準備を行った後、以下の手順を実行する。

- 1) $\hat{d}_1 = d_1$ とする。
- 2) $(k-1)$ 番目までの $\{\hat{d}_i\}$ を作ったとする。
- 3) $d_k \geq \hat{d}_{k-1}$ のとき、 $\hat{d}_k = d_k$ と定める。2) に行き、 \hat{d}_{k+1} に移る。
- 4) $d_k < \hat{d}_{k-1}$ のとき \hat{d}_k とその前の値を以下のように決定、変更する。2) に行き、 \hat{d}_{k+1} に移る。
 $i = 1, 2, \dots, k-1$ と順に変えて、以下を満たす最小の i を見つける。

$$\hat{d}_k = \frac{1}{i+1} \left(d_k + \sum_{j=1}^i \hat{d}_j \right) \geq \hat{d}_i$$

見つければ、 $\hat{d}_k = \hat{d}_{k-1} = \dots = \hat{d}_{k-i} = \hat{d}_i$ とする。

見つからなければ、 $\hat{d}_k = \hat{d}_{k-1} = \dots = \hat{d}_1 = \frac{1}{k} \sum_{j=0}^{k-1} \hat{d}_{k-j}$ とする。

- 5) \hat{d}_l を定めたとき、プロセスを終了する。

ディスペリティ \hat{d}_{ij} の導入は、矛盾を含む非類似度 s_{ij} を、矛盾なく求められる距離 d_{ij} を使って、(3.9)の順序関係を保持しながら、できるだけ実現可能な値に近づける操作と考えられる。

ディスペリティが求められたら、その値にできるだけ近づけるように再度 d_{ij} を構成しなおす。その基準をストレスと呼び、以下のように定義する。

$$S = \sqrt{\sum_{i=1}^n \sum_{j=i+1}^n (d_{ij} - \hat{d}_{ij})^2 / \sum_{i=1}^m \sum_{j=i+1}^m d_{ij}^2} \quad \text{ここに、} d_{ij} = \left(\sum_{\alpha=1}^p |x_{i\alpha} - x_{j\alpha}|^t \right)^{1/t} \quad (3.10)$$

一般にこの距離をミンコフスキー距離、 t の値をミンコフスキー定数と呼ぶ。特ミンコフスキー定数が2の場合がユークリッド距離である。我々のプログラムでは、 S の最適化の方法は最急降下法を用い、 $x_{i\alpha}$ の初期値には、元の計量MDSから求めた値を使っている。ストレスの定義は参考文献[2]で別の定義を示しているが、ここでは参考文献[1]の定義に従っている。

次元数 p を増やして行く際のストレスの変化を表す折れ線グラフ(ストレスプロット)を描き、どの次元から適合度が良くなるか調べる。また s_{ij} の値を横軸に取り、縦軸にその値に対応する \hat{d}_{ij} 及び d_{ij} の値を2種類のマーカーでポイントする。これをシェパードダイアグラムと呼ぶ。 \hat{d}_{ij} の値は同じ値を取るものがあるので、 \hat{d}_{ij} の上下に d_{ij} が散らばる傾向があるが、これらの点が \hat{d}_{ij} に近く、 s_{ij} の大きさによる逆転が起こらないほど適合度は高い。推測されたデータの点 $x_{i\alpha}$ の α の2つの次元($\alpha=1,2$ の場合が多いが)について平面上に点を描いて、位置を確かめることも多い。

3.3 プログラムの動作

メニュー[分析-多変量解析等-関係分析手法-多次元尺度構成法]を選択すると図3.1のような多次元尺度構成法実行メニューが表示される。デフォルトは、計量MDSになっている。

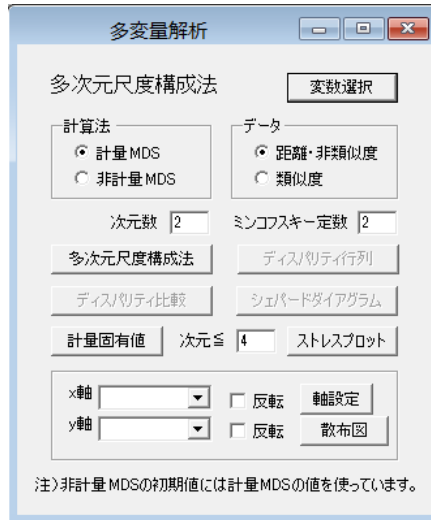


図 3.1 計量MDS 多次元尺度構成法実行メニュー

データには、図3.2のように、類似度が低いほど大きな値を取る非類似度データ(または距離データ)か、類似度が高いほど大きな値を取る類似度データを用いる。これらの選択は「データ」グループボックスで指定する。

	a	b	c	d	e
a					
b		3.8			
c		4.3	5.4		
d		4.3	3.0	3.0	
e		6.5	5.7	3.0	2.8

図 3.2 非類似度データ

非類似度データの場合、対角成分は空欄か 0 にする。類似度データの場合、対角成分は空欄か、最も大きな値を取るものとする。類似度データの場合はこの最大の値（またはデータ中の最大の値）から各セルの値を引いたものを非類似度データの値として用いている。データは図 3.2 のように三角データか、対称データを用いる。非対称データの場合の処理もできるが、我々のプログラムでは、2つの対応するデータの平均を取ることで対称化して利用している。

計量 MDS か非計量 MDS かは「計算法」グループボックスで指定する。計量 MDS の場合、ミンコフスキー定数は通常 2 で考える。このデータでは次元数を 2 として、「変数選択」した後、「多次元尺度構成法」ボタンをクリックすると図 3.3 のような位置座標に関する実行結果が表示される。

	1次元	2次元
a	2.844	-1.747
b	2.308	2.019
c	-1.463	-1.846
d	-0.473	0.974
e	-3.216	0.600

図 3.3 計量 MDS の実行結果

計算途中の非類似度行列の固有値と固有ベクトルは、「計量固有値」ボタンをクリックすることで、図 3.4 のように表示される。

	1次元	2次元
固有値	26.124	11.840
a	0.556	-0.508
b	0.452	0.587
c	-0.286	-0.536
d	-0.093	0.283
e	-0.629	0.174

図 3.4 非類似度行列の固有値と固有ベクトル

「次元 ≤」を 4 で、「ストレスプロット」ボタンをクリックすると、図 3.5 のようなグラフが表示される。

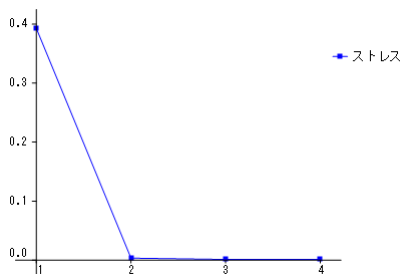


図 3.5 ストレスプロット

ストレス値の評価は、0.2 : 良くない、0.1 : 悪くはない、0.05 : 良い、0.025 : 非常に良い、というように言われている。この例の場合だと、2次元の段階で評価が非常に良くなっているため、2次元の結果を受け入れる。

2次元の実行結果の位置を図として表示するために、「軸設定」ボタンで軸を選択し（この場合は自動的に2つの次元）、「散布図」ボタンをクリックすると、図 3.6 のような結果が表示される。

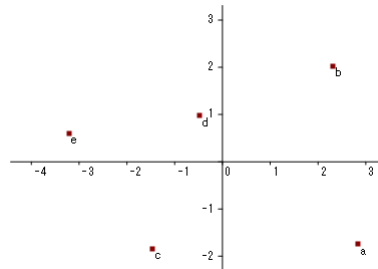


図 3.6 位置関係結果

次に、参考文献[1]にある例題を用いて非計量 MDS の操作を説明する。実行メニューの「計算法」で「非計量 MDS」を選ぶと実行メニューは図 3.7 のように変わり、いくつかのボタンが利用可能になっている。

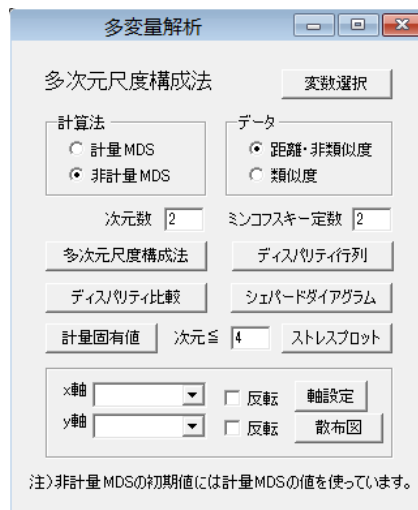


図 3.7 非計量 MDS 多次元尺度構成法実行メニュー

図 3.8 に距離行列を示す。これは類似度データである。

	クラウン	セドリック	サニー	マークII	カローラ	スカイライン	マーチ	ウィッツ	RAV4	パジェロ
クラウン	10									
セドリック	9	10								
サニー	6	7	10							
マークII	7	9	8	10						
カローラ	5	6	8	8	10					
スカイライン	2	3	6	3	6	10				
マーチ	2	3	5	4	7	6	10			
ウィッツ	1	2	4	3	5	5	9	10		
RAV4	1	1	2	1	3	3	7	8	10	
パジェロ	2	3	3	4	5	2	5	5	4	10

図 3.8 非計量、類似度データ

実行メニューの「データ」グループボックスで「類似度」を選び、「次元数」を2にして、すべての変数を選択し、「多次元尺度構成法」ボタンをクリックすると、図 3.9 のような結果が表示される。

	1次元	2次元
ク라운	4.494	-0.955
セドリック	3.983	-0.576
サニー	2.131	1.772
マークII	3.335	-0.779
カローラ	0.806	0.610
スカイライン	-1.225	4.126
マーチ	-2.914	0.601
ヴィッツ	-4.009	0.105
RAV4	-4.889	-1.177
パジェロ	-1.713	-3.727

図 3.9 非計量 MDS の実行結果

実行メニューの「ディスパリティ行列」ボタンをクリックすると、図 3.8 の類似度データに対応するディスパリティを図 3.10 のように表示する。但し、類似度データは、非類似度 = 類似度最大値 - 類似度、によって、非類似度に変更して計算される。

	ク라운	セドリック	サニー	マークII	カローラ	スカイライン	マーチ	ヴィッツ	RAV4	パジェロ
ク라운	0.0000	0.8391	3.8135	2.6355	4.6990	7.5870	7.5870	8.7703	8.7703	7.5870
セドリック	0.8391	0.0000	2.6355	0.8391	3.8135	6.7177	7.5870	8.7703	8.7703	6.7177
サニー	3.8135	2.6355	0.0000	2.2557	2.2557	3.8135	4.6990	5.6701	7.5870	6.7177
マークII	2.6355	0.8391	2.2557	0.0000	2.2557	6.7177	5.6701	6.7177	8.7703	5.6701
カローラ	4.6990	3.8135	2.2557	2.2557	0.0000	3.8135	2.6355	4.6990	6.7177	4.6990
スカイライン	7.5870	6.7177	3.8135	6.7177	3.8135	0.0000	3.8135	4.6990	6.7177	7.5870
マーチ	7.5870	6.7177	4.6990	5.6701	2.6355	3.8135	0.0000	0.8391	2.6355	4.6990
ヴィッツ	8.7703	7.5870	5.6701	6.7177	4.6990	4.6990	0.8391	0.0000	2.2557	4.6990
RAV4	8.7703	8.7703	7.5870	8.7703	6.7177	6.7177	2.6355	2.2557	0.0000	5.6701
パジェロ	7.5870	6.7177	6.7177	5.6701	4.6990	7.5870	4.6990	4.6990	5.6701	0.0000

図 3.10 ディスパリティ行列

「ディスパリティ比較」ボタンをクリックすると、図 3.11 のように非類似度、ディスパリティ、距離を非類似度の昇順に並べた表が示される。

	S	D.P.	Dist
s2,1	1.0000	0.8391	0.6361
s4,2	1.0000	0.8391	0.6796
s8,7	1.0000	0.8391	1.2015
s4,3	2.0000	2.2557	2.8210
s5,3	2.0000	2.2557	1.7614
s5,4	2.0000	2.2557	2.8851
s9,8	2.0000	2.2557	1.5551
s4,1	3.0000	2.6355	1.1728
s3,2	3.0000	2.6355	2.9910
s7,5	3.0000	2.6355	3.7209
s9,7	3.0000	2.6355	2.6574
s3,1	4.0000	3.8135	3.6088
s5,2	4.0000	3.8135	3.3913

図 3.11 非類似度、ディスパリティ、距離比較表

この関係を図で表したものがシェパードダイアグラムである。「シェパードダイアグラム」ボタンをクリックすると図 3.12 のようなグラフが表示される。横軸は元の類似度の値である。距離の点の散らばり方で、適合の良し悪しをみることができる。

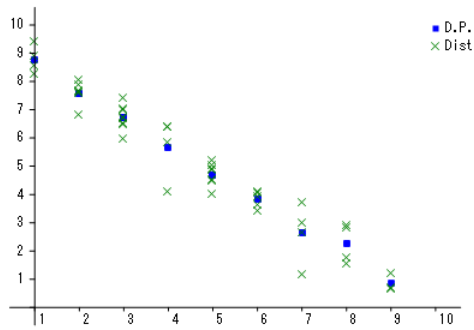


図 3.12 シェパードダイアグラム

軸を設定して「散布図」ボタンをクリックすると、図 3.13 のような位置表示のグラフが表示される。

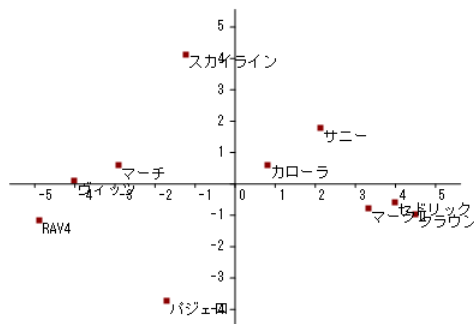


図 3.13 位置関係結果

次元を多くして行った際の、ストレスの変化は「ストレスプロット」ボタンをクリックすることで図 3.14 のように得られる。

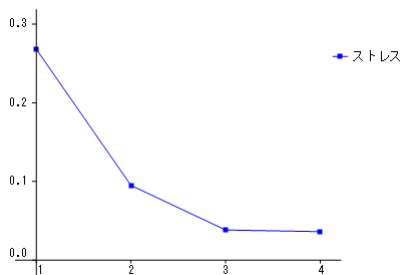


図 3.14 ストレスプロット

4. 数量化IV類

4.1 数量化IV類の理論

林の数量化IV類はデータ間の親近性を仮定し、その中に内在するパターンをデータの空間配置として表現する手法である^[2]。 r 次元ユークリッド空間中に m 個のデータがあり、データ i とデータ j との親近性（類似度）を e_{ij} とする。親近性には正負の符号の制限はないが、親近性が高いほど大きな値を取るものとする。また一般に対称性 $e_{ij} = e_{ji}$ を仮定しない。同一のデータ同士の親近性 e_{ii} は、後の議論から定義する必要はないが、取り敢えず 0 としておく。 r 次元の空間中のデータ i の位置座標を

$x_{i\alpha}$ ($\alpha=1,2,\dots,r$) とし、この α 次元目のデータをベクトルで表し $\mathbf{x}_\alpha = (x_{1\alpha}, x_{2\alpha}, \dots, x_{m\alpha})'$ とする。

データ i とデータ j の距離 $d_{ij} (\geq 0)$ を位置座標 $x_{i\alpha}$ と $x_{j\alpha}$ を使って、以下のように定義する。

$$d_{ij}^{(r)2} = \sum_{\alpha=1}^r (x_{i\alpha} - x_{j\alpha})^2 \quad (4.1)$$

今、親近性の高いデータ同士は近い距離に位置するように配置したいが、これを実現するために、以下の量 Q を最大化することを考える。

$$Q = -\sum_{i=1}^m \sum_{j=1}^m e_{ij} d_{ij}^{(r)2} = -\sum_{i=1}^m \sum_{j=1}^m \sum_{\alpha=1}^r e_{ij} (x_{i\alpha} - x_{j\alpha})^2 \quad (4.2)$$

ここで、

$$(\mathbf{G})_{ij} = g_{ij} = h_{ij} - \delta_{ij} \sum_{k=1}^m h_{ik}, \quad h_{ij} = e_{ij} + e_{ji} \quad (4.3)$$

と定義とすると、 Q は以下のように書ける。

$$Q = \sum_{\alpha=1}^r \sum_{i=1}^m \sum_{j=1}^m g_{ij} x_{i\alpha} x_{j\alpha} = \sum_{\alpha=1}^r \mathbf{x}'_\alpha \mathbf{G} \mathbf{x}_\alpha \quad (4.4)$$

ここで、 \mathbf{x}_α の値によって Q の値はいくらでも大きくできるため、以下の条件を付けることにする。

$$\sum_{i=1}^m x_{i\alpha}^2 = 1 \quad (4.5)$$

制約条件を付けたラグランジュの未定定数法を用いて、 Q の式を以下のように変更する。

$$L = \sum_{\alpha=1}^r \mathbf{x}'_\alpha \mathbf{G} \mathbf{x}_\alpha - \sum_{\alpha=1}^r \lambda_\alpha (\mathbf{x}'_\alpha \mathbf{x}_\alpha - 1) \quad (4.6)$$

これを \mathbf{x}_α で微分して以下の固有値方程式を得る。

$$\mathbf{G} \mathbf{x}_\alpha = \lambda_\alpha \mathbf{x}_\alpha \quad (4.7)$$

固有値方程式を成分で書き換えると以下のようなになる。

$$\sum_{k=1}^m g_{ik} x_{k\alpha} = \lambda_\alpha x_{i\alpha} \quad (4.8)$$

これより以下となる。

$$\lambda_\alpha \sum_{i=1}^m x_{i\alpha} = \sum_{i=1}^m \sum_{k=1}^m g_{ik} x_{k\alpha} = 0 \quad (4.9)$$

ここで定義式によって成り立つ以下の関係を使った。

$$\sum_{j=1}^m g_{ij} = 0 \quad (4.10)$$

(4.9)式より $\lambda_\alpha \neq 0$ の場合、以下となる

$$\sum_{i=1}^m x_{i\alpha} = 0 \quad (4.11)$$

また、(4.10)式が成り立つことから方程式の1つの解として

$$\lambda_\alpha = 0, \quad x_{i\alpha} = 1/\sqrt{m} \quad (4.12)$$

を持つことも分かる。この場合(4.9)式の関係から、(4.11)式は成り立たなくてもよい。

最後に、方程式(4.8)を用いると、(4.4)の定義と(4.5)の制約より以下となる。

$$Q = \sum_{\alpha=1}^r \sum_{i=1}^m \lambda_\alpha x_{i\alpha}^2 = \sum_{\alpha=1}^r \lambda_\alpha \quad (4.13)$$

親近性 e_{ij} の線形変換に対する固有値と固有ベクトルの変化を調べてみる。

$$e'_{ij} = ae_{ij} + b \quad (4.14)$$

の変換に対して、

$$h'_{ij} = ah_{ij} + 2b$$

$$g'_{ij} = ag_{ij} - 2b(m\delta_{ij} - 1), \quad \sum_{j=1}^m g'_{ij} = 0 \quad (4.15)$$

これにより、固有方程式は以下となる。

$$\sum_{k=1}^m (ag_{ik} + 2b)y_{k\alpha} = (\lambda'_\alpha + 2mb)y_{i\alpha} \quad (4.16)$$

これは $y_{k\alpha} = x_{k\alpha}$ とすると以下の関係を得る。

$$\lambda'_\alpha = a\lambda_\alpha - 2mb \quad \text{for } \lambda_\alpha \neq 0, \sum_{i=1}^m x_{i\alpha} = 0$$

$$\lambda'_\alpha = 0 \quad \text{for } \lambda_\alpha = 0, x_{i\alpha} = \text{const.} \quad (4.17)$$

即ち、固有値も線形の変換を受ける。これより、0でない固有値の分布の間隔比

$$\gamma(\alpha) = \frac{\lambda_{\max} - \lambda_\alpha}{\lambda_{\max} - \lambda_{\min}} \quad (4.18)$$

は変換(4.14)に対して不変である。これにより、データに固有の親近性の特徴を調べることができると考えられる。最後に、数量化の適合度の1つの指標として、距離 $-e_{ij}$ と(4.1)で与えられる r 次元の距離 $d_{ij}^{(r)}$ との順位相関係数を考えることもある。しかし、これは次元数を増やせば必ず適合度が上がるとは限らず、注意が必要である。

4.2 プログラムの利用法

数量化IV類のデータは、変数間の親近性（類似度）または距離（非類似度）を表すデータである。その例を図4.1に示す。

果物の非類似性	みかん	りんご	いちご	ぶどう	なし	メロン
みかん	0	2.70	3.00	2.65	2.60	3.10
りんご	2.30	0	2.80	2.90	2.40	3.50
いちご	3.00	2.80	0	2.25	3.05	3.40
ぶどう	2.65	2.90	2.25	0	3.20	3.25
なし	2.60	2.40	3.05	3.20	0	3.30
メロン	3.10	3.50	3.40	3.25	3.30	0

図 4.1 距離を表すデータ

メニュー [分析-多変量解析等-関係分析手法-数量化IV類] を選択すると図 4.2 のような数量化IV類分析メニューが表示される。

数量化IV類

変数選択

データ

距離(非類似度)

親近性(類似度)

次元数 ≤ 100

数量化IV類

変換(eijは親近性)

なし

eij-max|eij|

線形変換

1 eij - 0

注)親近性=-距離、変換は親近性を使っています。

X軸 [] 反転 軸設定

Y軸 [] 反転 散布図

図 4.2 数量化IV類分析メニュー

変数選択ですべての変数を選択し、データによって「距離」か「親近性」かを選択する。距離の場合はデータの符号を変えて親近性にして分析を進める。変数の変換が必要な場合は変換ラジオボタンで指定する。特に「eij-max|eij|」は固有値をすべて正にするための設定であり、「線形変換」は他の多次元尺度構成法と合わせるための設定である。「次元数」は大きな値を設定しておけば、変数数・1の値に設定し直して計算する。もちろん見やすくするため小さな値に設定することもできる。

「数量化IV類」ボタンをクリックすると、図 4.3 のような実行結果が表示される。

	1次元	2次元	3次元	4次元	5次元
固有値	39.788	36.280	33.808	33.202	32.522
みかん	0.111	-0.160	0.564	-0.070	-0.687
りんご	0.242	-0.275	0.130	0.734	0.378
いちご	0.218	0.539	-0.585	0.185	-0.346
ぶどう	0.158	0.495	0.372	-0.418	0.501
なし	0.179	-0.603	-0.426	-0.494	0.111
メロン	-0.908	0.004	-0.056	0.062	0.043
間隔比	0.000	0.483	0.823	0.907	1.000
ed順位相関	0.570	0.953	0.770	0.511	

図 4.3 分析結果

固有値、固有ベクトルが表示され、その下に固有値の間隔比と親近性と予測距離との順位相関が表示される。

「軸設定」をして、「散布図」ボタンをクリックすると、パラメータ（固有ベクトル）の値が散布図として図 4.4 のように表示される。軸の向きは「反転」チェックボックスによって変更できる。

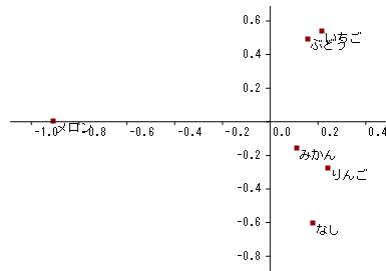


図 4.4 パラメータ散布図

5. おわりに

パス解析は観測変数間の因果関係を簡単に調べる方法として興味深い。本格的に内生変数も含めて、因果関係を調べるには共分散構造分析が優れているであろうが、パス解析では相関を、直接相関、間接相関、擬似相関に分けて考えることができ、モデルを簡単に作れるところが優れている。影響の方向性を仮定して分析を実行し、係数の値を見て影響の方向性を部分的に考え直すという使い方が一般的な利用法であろうが、モデル構築はすべての変数間を結び、その方向性のみ考えるというところが非常に簡単である。共分散構造分析に移る前段階として利用することを考えたい。

数量化IV類は多次元尺度構成法の1つの方法と考えてもよい。ここでは別々にプログラムしているが、目的は基本的に同じである。そこで、同じデータを用いて、計量及び非計量MDSと数量化IV類の結果を比較してみる。計量MDSの図3.6を数量化IV類と比べた結果が図5.1である。

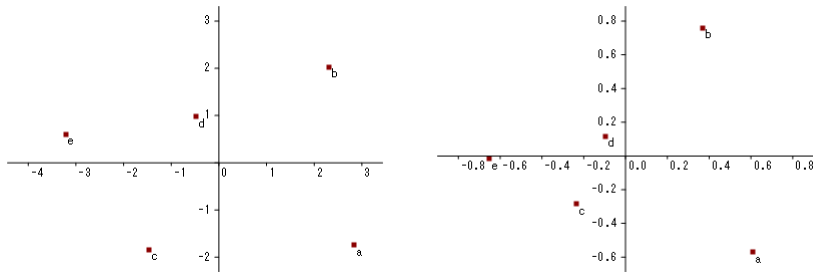


図 5.1 計量MDSと数量化IV類

計量MDSのデータは2次元上の点の配置から作ったデータで、散布図はその位置を正確に反映している。これに対して数量化IV類の散布図は点c, dが横軸に寄っている。

非計量MDSの図3.13を数量化IV類と比べた結果が図5.2である。

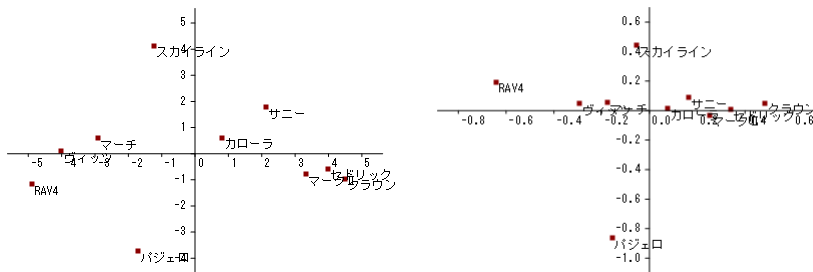


図 5.2 非計量MDSと数量化IV類

これは矛盾を含むデータであり、散布図を見るとパジェロに対して他の個体は横軸に寄っているように感じられる。縦軸方向は非計量 MDS の方が均一に分かれている。

最後に数量化IV類の図 4.4 を非計量 MDS と比べた結果が図 5.3 である。

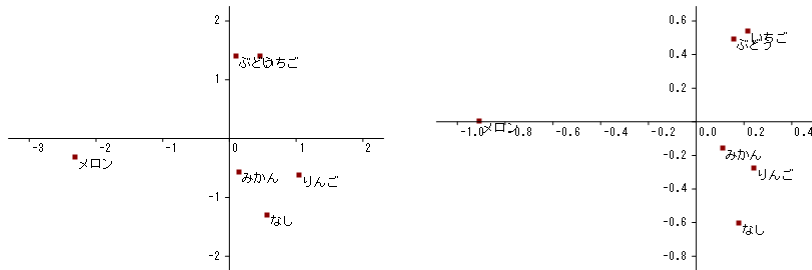


図 5.3 非計量 MDS と数量化IV類

これらの結果を見ると、非計量 MDS では全体的に丸く点が配置されるのに対して、数量化IV類では軸に沿って直線的に点が配置される傾向があるように思われる。どのような場合にどちらの手法を使うべきか、今後調べてみる必要があろう。

参考文献

- [1] 多変量解析法入門, 永田靖, 棟近雅彦, サイエンス社, 2001.
- [2] 関連性データの解析法 多次元尺度構成法とクラスター分析法, 齋藤堯幸, 宿久洋, 共立出版, 2006.

Multi-purpose Program for Social System Analysis 25 - Path Analysis, Multi-Dimensional Scaling, Quantification Method Type IV -

Masayasu FUKUI, Makoto OZAKI and Ryota ASAHI
Department of Business Administration, Faculty of Business Administration,
Fukuyama Heisei University

Abstract: We have been constructing a unified program on the social system analysis for the purpose of education. This time, we create a program of path analysis which predicts the influence between variables. Also, we create programs of multi-dimensional scaling and quantification method type IV which determine the placement of data in a multi-dimensional space from distances between data.

Keywords: College Analysis, multivariate analysis, path analysis, multi-dimensional scaling, MDS, quantification method type IV

URL: <http://www.heisei-u.ac.jp/ba/fukui/>