

群集構造解析プログラムPACS (A Program for Analyzing Community Structures) とその応用について

品川汐夫・若本ゆかり・中尾久子*

1. 緒言

先に品川^{1,2)}は、種×地点の2次元表として表された各種の個体数データから地点と種を座標空間に配置して(新座標づけ法)、これにファジイクラスター解析を適用するという、生物群集数理解析の新手法を提案し、多くの計算例によりその有用性を検証した。^{3,4,5)}その後この方法は、生物群集だけでなく、女子短大生の食生活分析⁶⁾や職業ストレスと問題飲酒の関係の解析⁷⁾など、不規則に変動する、本質的にファジイな現象の調査データの解析に対しても効果的であることが報告されている。

しかしながら、この手法は一連の数値計算手順を必要とするため、市販されている一般的な統計解析のプログラムパッケージでは適用が困難である。

そこで今回、この手法に必用な一連の計算手続きを全て収録した一つのプログラムパッケージPACSを作成した。これは、必ずしも数理解析に明るくない人でも手法の適用が可能となることを目的としたものである。

以下においては、このプログラムの構成と機能、および使用法と使用例について解説する。

なお、このプログラムの作成に使用したプログラミング言語は、マイクロソフト社の Visual Basic Ver. 6.0で、Windows95以上のOSで作動する。研究上の使用目的に対しては著作権を主張しないが、断りなくプログラムを改ざんすることは禁止する。また、使用した場合は、研究内容とその結果について報告されることを希望する。

* 山口県立大学看護学部助教授

2. プログラムの構成と機能

2.1 各モジュールの構成と機能

PACSは、基本的に以下の単位機能を有する5つのモジュールから構成され、各モジュールは、さらにいくつかのプログラム群により構成されている（図1）。

PACSへの入力データは、エクセルのマクロによりエクセルのワークシートから出力されたテキストファイルを用いる。また、様々な計算結果も多くはテキストファイルとして出力され、エクセルのマクロによりワークシートに読み込んでエクセルの機能を用いて図表作成などの編集を行う。このために必要なマクロは、PACSに添付されているエクセルのファイル Macrolib-2 に収録されている。

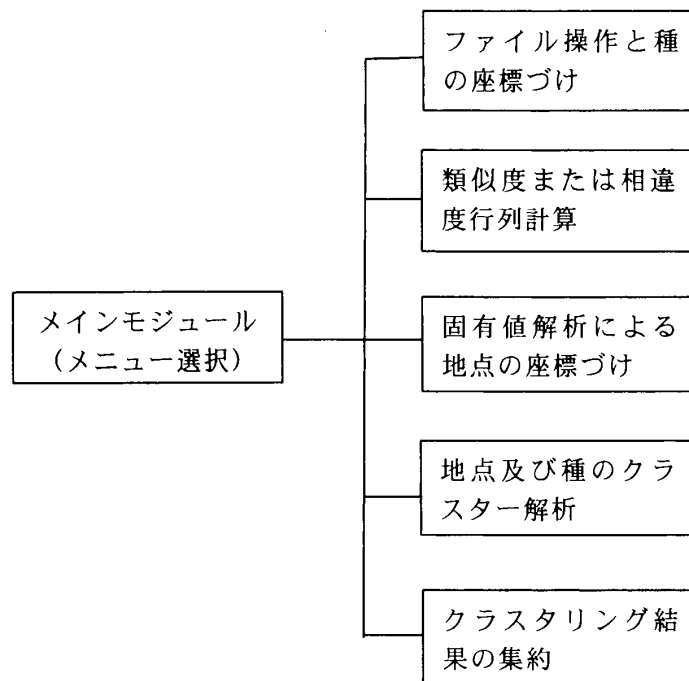


図1 PACSの基本モジュール構成

2.1.1 ファイル操作モジュール

このモジュールでは、エクセルとエクセルのマクロで作成されたテキスト形式のデータファイルから、出現地点数や個体数が希少な種および欠測データ等を除くことや、ファイルを合成することなど、入力データの編集を行うほか、バイナリ形式で出力された類似度行列や相違度行列の計算結果をエクセルに読み込める形式のテキスト形式に変換すること等を行う。また、便宜上、種（データの行）の座標づけ計算もこのモジュールで行う。

2.1.2 行列計算モジュール

ここでは、入力データファイルから地点（データの列）間の類似度や相違度行列、通常積率相関行列、および類似度行列から計算される品川^{1,2)}の関係類似度行列などを計算し、結果をバイナリ形式のファイルとして出力する。

計算可能な類似度行列には次の6種が用意されている。

- i Kimoto⁸⁾の $C\pi$
- ii 品川^{1,2)}の $C\pi'$
- iii 品川^{1,2)}のRs
- iv Bray-Curtis measure⁹⁾
- v Ruzicka's measure¹⁰⁾
- vi 偏差パターン類似率¹¹⁾

偏差パターン類似率は、豊川ら¹¹⁾によって、食生活のパターン分析に用いられた指数である。

また、計算可能な相違度行列には、次の2通りがある。

- i 通常のユークリッド距離
- ii Williams and Stephenson¹²⁾の群集間相違度

関係類似度 (Relational Similarity) を計算するためには、あらかじめ上の類似度行列または積率相関係数行列の計算を行っておいて、それを入力行列として指定する。

2.1.3 行列の固有値解析モジュール

このモジュールでは、2.1.2で出力された類似度または関係類似度行列の固有値 (Eigenvalue) と固有ベクトル (Eigenvector) を計算して出力し、さらに、固有ベクトルに固有値の平方根を乗じて要因ベクトル (Factor vector) として出力する。計算は反復代入法により行い、各反復回の最大誤差が 10^{-9} 未満になれば収束とみなす。固有ベクトルは固有値の大きい順に求め、その数 (次元数) はインプットボックスに入力して指定する。

2.1.4 クラスタ解析モジュール

上述の2.1.2で出力された類似度行列または相違度行列から群平均法によりデンドログラムを作成すること、および2.1.3で出力された要因ベクトルあるいは2.1.1で出力された種の座標づけ結果を用いて、地点（データの列）あるいは種（データの行）をファジィ c -means 法¹³⁾ (ファジィ k -means 法と呼ばれることもある) によりクラスタリングする。

デンドログラム作成を指定した場合、結果はテキスト形式のノードデータファイルとして出力される。そこでエクセルワークシートで、そのファイル名をフルパスで記入したセルにポイ

ンターを置いて Macrolib-2 中のマクロ, Dendgrm を実行すれば, このファイルを読み込んでワークシート内にデンドログラムが描かれる。

ファジイクラスタリングは, 地点と種では出力内容が少し異なる。すなわち, 地点のクラスタリングでは, クラスタリング結果と, もとの入力データからクラスタリングのあいまい度指数²⁾が計算され, 種のクラスタリングでは, 地点のクラスタリング結果を用いて種群と地点群の対応度指数²⁾が計算されて出力される。したがって, 種(データの行)のクラスタリングの前に地点(データの列)のクラスタリングを行っておく必要がある。

2.1.5 クラスタリング結果の集約モジュール

ここでは, ファジイクラスタリングされた地点群および種群の総合的評価を行う。そのために以下の4通りの機能が用意されている。

1) 種群の個体数と種類数およびそれらの特化係数¹⁴⁾の計算

種(データの行)のファジイクラスタリング結果を用いて, 種群の個体数と種類数およびそれらの特化係数¹⁴⁾を計算する。種群の個体数とは, 品川¹⁾が提案したファジイクラスタリング結果の集約法で, 各種の個体数を種群への所属率の2乗を重みとして種群へ比例配分し, それを全種について合計したものである。種類数は, 個体数の代りに在を1, 不在を0として計算したものである。また, 特化係数は, これらから行(種群)と列(地点)合計の影響を除いたものである。これらの結果は群集の連続的変化を表し, 物理化学的環境変数との回帰分析, 正準相関分析などに用いることが出来る。

2) 各種(データの行)個体数の地点(データの列)群別平均

これは, 各種の個体数を, それが出現した地点の各地点群への所属率の2乗を重みとして平均したもので, その種がどの地点群で多く出現しているかを示す。

3) 各地点の種類数, 個体数, および多様度の計算

各地点における種類数, 個体数, および多様度は, その地点が属する地点群の性質を知る上で重要な手がかりとなる。そこでここでは, 種類数と個体数のほか, Simpsonの多様度 $S^{15)}$ および, Margalefの多様度 D_n と $D_m^{15)}$ を計算して出力する。その場合, データファイルが動物門別に作成されていれば, これらの指数を動物門別および全動物門について計算する。

4) Dufrene and Legendre¹⁶⁾のIndicator Valueの計算および有意性のテスト

Dufrene and Legendre¹⁶⁾は, クラスタリングされた地点群を特徴づける指標種を見出す方法として, 種別に計算されるIndicator Valueを提案している。この指数には2通りあり, 1つは特定の地点群への出現し易さを示し(IndVal 1), 他の1つは出現し難さを示す(IndVal 2)。ここでは, これらをファジイクラスタリングに拡張し, 通常平均と分散の代わりに, 地点(列)の各地点群への所属率の2乗で重み付けられた平均と分散を用いて計算する。

この指数の統計的有意性は、地点をランダムに並べ替えることによって繰り返し計算された指数の地点群別平均と標準偏差から、この指数を、地点群に差がない場合には標準正規分布に従う指数 z に変換することにより行う。 z の絶対値が1.96を超えれば5%の水準で平均からの偏りが有意と判定される。繰り返し計算の回数は、500回程度が必用である。また、もとのデータが著しく正規性から外れる場合には、対数変換などにより正規化しておくべきである。

2・2 メニュー構成

PACS を起動するとタイトル画面の次にメニュー画面が現われるこの画面上のメニューバーをクリックすることにより上に述べた5つの機能呼び出すことが出来るほか、主座標解析法 (PCO: 計量的多次元尺度法 MDS と呼ばれる)¹⁷⁾、対応分析法 (CA)¹⁸⁾、新座標づけ法、主成分分析法 (PCA) については、地点と種の座標づけからファジイクラスタリング、さらにクラスタリング結果の集約まで一貫して自動的に各モジュールを呼び出すメニューが用意されている。

以下では、各メニューについて説明する。

2・2・1 Fileメニュー

上に述べたファイル操作モジュールを呼び出すメニューで、以下のサブメニューから選択する。

1) Raw data to analyze

解析するデータファイルを読み込む。ここで読み込まれたデータは、各モジュールで共通して用いられるが、各モジュールを実行する際に読み込むことも出来る。

2) Matrix to analyze

解析するバイナリ形式の行列データファイルを読み込む。ここで読み込まれた行列は、各モジュールで共通して用いられるが、各モジュールを実行する際に読み込むことも出来る。

3) Select data to analyze

データファイルから、出現地点あるいは出現個体数が希少な種 (データの行) を削除し、また、9999と記入されたセル (欠測値) が一ヶ所でもあればその行を削除して、新しいデータファイルとして出力する。入力データファイルは10まで同時指定可能で、それらを一つのファイルに結合することが出来る。

4) Species ordination

データファイルと行列の固有ベクトルファイルを読み込んで、種 (データの行) の座標づけを行う。これを実行する前行列計算と固有ベクトル計算をしておく必要がある。

5) Binary to text

バイナリ形式の行列ファイルを読み込んで、テキスト形式に変換して出力する。

6) Cancel

入力された全てのファイル名をクリアする。ファイル名が残っていると、他のモジュールで自動的にそれが用いられるので、計算データを代える場合には必ずこれを実行しなければならない。

2.2.2 Methodsメニュー

座標づけからクラスタリング、さらにクラスタリング結果の集約までを一貫して自動的に行うルーチンを呼び出す。直下のサブメニューには In memory と Out memory があり、In memory を選択した場合には、全ての計算を内部メモリで計算する。計算速度は速いが、地点数が100数10を超えると計算機によってはメモリ不足になることがある。In memory で選択できる解析法には、主座標解析法 (PCO)、対応分析法 (CA)、主成分分析法 (PCA)、および品川の解析法 (New ordination) がある。PCO では、さらに類似度行列を用いる (Similarity) か、相違度行列を用いるか (Distance) を選択する。

Out memory を選択した場合には、作業領域にハードディスクを用いるので、地点数の制限はないが、計算時間が長くなる。選択できる解析法は、主座標解析法 (PCO)、対応分析法 (CA)、および品川の解析法 (New ordination) の3通りである。

2.2.3 MatrixCalメニュー

地点 (データ行列の列) 間の類似度または相違度行列を計算してファイル出力する。上と同様に、まず In memory か Out memory かを選択する。続いてその下位メニューで、S. Matrix (類似度行列)、Relational S. (関係類似度行列)、Distance (相違度行列)、Correlation (積率相関係数行列) のいずれかを選択する。S. Matrix と Distance にはさらに下位メニューがあり、2.1.2で述べた各種の類似度行列あるいは相違度行列を選択できる。

2.2.4 EigenAnalysisメニュー

上の2.2.3で出力された行列の固有値と固有ベクトルを計算してファイル出力する。ここでも In memory か Out memory かを選択し、反復計算によって固有値の大きい順に固有値と固有ベクトル、要因ベクトルが出力される。その際、入力に指定した行列ファイルは変形され保存されないため、必要な場合は2.2.1の Binary to text によりあらかじめテキストファイルに変換してエクセルのワークシートに読み込むか、別にコピーを保存しておく必要がある。

ここで計算された固有ベクトルは、2.2.1の種の座標づけ計算に、また、要因ベクトルは地点のクラスタリングに用いられる。

2・2・5 Clusteringメニュー

地点（列）と種（行）のクラスタリングを行う。クラスタリングの方法として、まず Dendrogram（群平均法によるデンドログラムの作成）か、Fuzzy（ファジイクラスタリング）かを指定する。

Dendrogram を指定した場合、さらに Similarity（類似度行列を用いる）か、Distance（相違度行列を用いる）かを選択する。したがって、あらかじめ2・2・3の MatrixCal メニューあるいは2・2・2の Methods メニューを実行して、類似度または相違度行列を計算しておく必要がある（2・1・4参照）。

Fuzzy を選択した場合、さらにその下位メニューで、Clustering of sites（地点あるいは列のクラスタリング）か、Clustering of species（種あるいは行のクラスタリング）か、あるいは Fuzziness of site groups（地点のクラスタリング結果と、もとのデータファイルから、品川²⁾による地点群のあいまい度を計算する）かを選択する。先に述べた（2・1・4）ように、Clustering of sites（地点のクラスタリング）を先に実行しておく必要がある。

その場合、2・2・4の EigenAnalysis で計算された要因ベクトルを入力データとして指定する。Clustering of sites を実行すれば、その中で Fuzziness of site groups（地点群のあいまい度）の計算は自動的に呼び出される。その際、元のデータファイル名が未だどのメニューでも読み込まれていなければ、その入力が必要される。読み込まれていればそれが自動的に用いられる。

Clustering of species（種のクラスタリング）を選択した場合、2・2・1の Species ordination で計算された種の座標づけ出力ファイルと、上の Clustering of sites で計算された地点群のあいまい度の出力ファイルを入力ファイルとして指定する。それらを用いて、種のファジイクラスタリング結果と、品川²⁾の、種群と地点群の対応度指数行列が計算されて出力される。

上に述べたように、地点群のあいまい度は、地点のクラスタリングで自動的に計算されるが、場合によってはクラスタリングで用いたデータとは別のデータを用いてあいまい度を計算する必要も生じる。その場合に Fuzziness of site groups メニューを実行する。

2・2・6 Summarize（クラスタリング結果の集約）メニュー

ここでは、地点群と種群のファジイクラスタリング結果の集約を行うほか、各地点の種類数、個体数、種多様度を計算するメニューが用意されている。

1) Sum within data group（種群の個体数）

ここでは、クラスタリングされた種（行）群の個体数と種類数およびそれらの特化係数を計算する（2・1・5の1）参照）。

2) Average within item group

ここでは、各種（行）毎に、その個体数の地点（列）群別平均を計算する（2・1・5の2）参照）。

3) Diversity

ここでは、各地点における種類数、個体数、および多様度を計算して出力する（2・1・5の3）参照）。

4) Cal. of IndVal and significance test

ここでは、Dufrene and Legendre¹⁶⁾によって提唱されている各地点群の2種類のIndValを各種別に計算すると共に、地点をランダムに並べ代える繰り返し計算によって、それらの値を標準正規分布に従う変数 z に変換して出力する（2・1・5の4）参照）。

3. 解析の流れとプログラムの使用法

ここでは、PACSの使用法を図2に示す品川の解析法の手順に沿って説明する。

3.1 データファイルの作成

先ずエクセルワークシートに、図3に示すような形式で種×地点の2次元表データを作成する。図3では、5行目に記入されたA~Eが地点名で、d1~d5が種名である。また、1行1列目にはデータファイル名をフルパスで記入し、2行1列目は地点（列）数を、3行1列目には種類数（行数）を記入する。さらに、4行1列目には適当なコメントを記入できる。次に、ファイル名を記入したセル（ここではA1）にセルポインターを置いてMacrolib-2中のマクロ、FileWrite-Bを実行すれば、指定したフォルダ（この例ではC:\My Documents）にデータファイルが作成される。作成されたデータファイルは、同じくファイル名を記入したセルにセルポインターを置いてマクロ、FileRead-Cを実行すれば読み込まれる。

3.2 類似度と関係類似度の計算

つぎに、スタートからPACSを起動すれば、タイトル画面になる。この中のPACSの文字付近をクリックするとメニューバーが現われる。このメニューから、Methods⇒In memory（またはOut memory）⇒New ordinationを選択すれば全ての手順が自動的に実行されるが、ここでは説明のため、各手順をメニューから実行する方法を述べる。

先ずメニューから、MatrixCal⇒In memoy（またはOut memory）⇒S.Matrix⇒Rsを選択して、Rs行列を計算して出力する。その際、出力ファイルの名前が要求されるが、任意のファイル名を指定する。続いて、同じメニューでRelational S.を選択し、関係類似度行列Rsnを

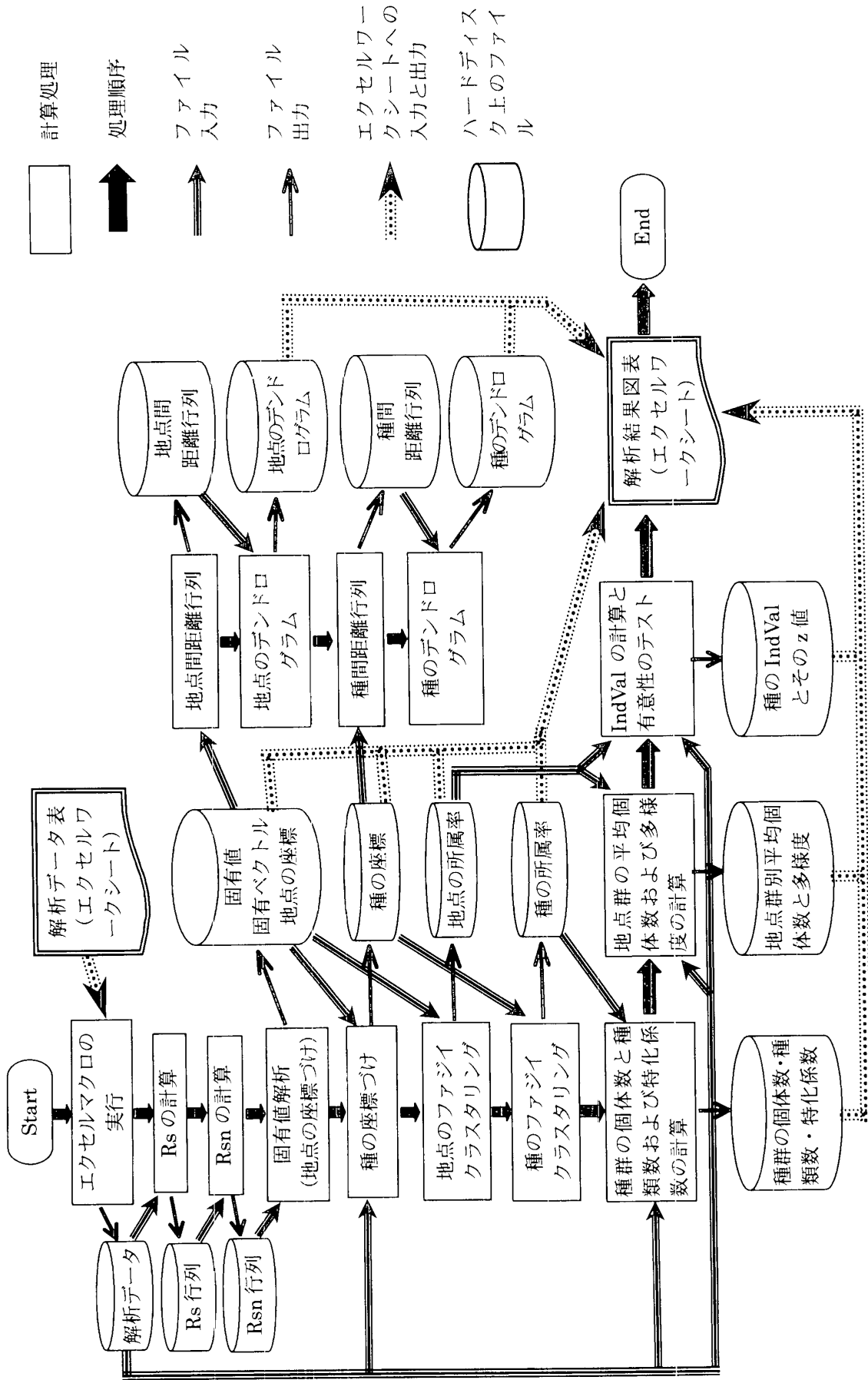


図2 PACSを用いた品川^{1,2)}の解析法の手順

	A	B	C	D	E	F	G
1	C:\My Documents\sample.sfl						
2	5						
3	5						
4	sample						
5	Id	data	A	B	C	D	E
6	1	d1	20	40	96	960	9600
7	2	d2	20	30	1	10	100
8	3	d3	20	15	1	10	100
9	4	d4	20	10	1	10	100
10	5	d5	20	5	1	10	100

図3 エクセルワークシートでのデータの作成例
A~E, 地点名; d1~d5, 種名

計算する。その際、入力ファイルとして、上で計算したRs 行列ファイルを指定し、出力ファイルとして任意のファイル名を指定する。

ここで出力した行列ファイルは、そのままではエクセルワークシートに読み込めないので、File メニューの Binary to text を実行し、行列の出力ファイル（バイナリ）をテキストファイルに変換する。その結果の出力ファイル名をエクセルワークシートの任意のセルにフルパスで記入して、Macrolib-2 中のマクロ FileRead-B を実行すれば行列が読み込まれる。

3.3 固有値解析による地点の座標づけ

次に、メニューから Eigen Analysis⇒In memory（または Out memory）を選択して、固有値と固有ベクトル、および要因ベクトル（地点の座標づけ）を計算する。このとき、求める固有ベクトルの数は地点数に一致する数まで指定できる（図3の例では列数5）。固有値の計算は反復代入法を用いるので、計算途中に収束誤差を表示して、Continue?と聞いてくるときがある。その場合、収束が十分でないとその後の計算誤差が大きくなるので、収束するまで何度も Yes を指定する方がよい。ベクトル成分の最大誤差が 10^{-9} 未満になれば収束と判定される。

図4は、図3のサンプルデータの固有値ファイル名を rsn.evl, 要因ベクトル（地点の座標づけ）ファイル名を rsn.fvc, 種の座標づけファイル名を spax.dat として、Macrolib-2 中のマクロ FileRead-B によりエクセルのワークシートに読み込んだ例を示す。

3.4 種を選択と座標づけ

種の座標づけでは、偶発的に出現する希少種は除く方がよい。そこで、まず File メニューの Select data to analyze (2.2.1の3)) を実行して、複数の入力データファイルから、条件に適するデータ（行）を選択して一つのファイルにして出力する。条件として、全地点での合計個体数が基準値以上、出現地点数が基準値以上、それらの結合（AND, OR）を選ぶことができる。対応分析法（CA）では、地点の座標づけと種の座標づけは同じデータを用いなければ

	A	B	C	D	E	F	G
1	C:\My Documents\rsn.evl						
2	5						
3	2						
4	関係類似度行列の固有値						
5	Id	Data	No. 1	No. 2	No. 3	No. 4	No. 5
6	1	Eigenvalue	3.907644	0.537306	0.533933	0.020224	0.000893
7	2	Cumulative	78.15288	88.899	99.57766	99.98214	100
8	C:\My Documents\rsn.fvc						
9	5						
10	5						
11	要因ベクトル(地点の座標づけ)						
12	Id	Data	A	B	C	D	E
13	1	FV. 1	0.994692	0.999242	-0.79962	-0.80059	-0.79962
14	2	FV. 2	-3.3E-07	-3.1E-07	0.518221	0.000194	-0.51841
15	3	FV. 3	0.000357	0.000426	-0.29839	0.596704	-0.29806
16	4	FV. 4	0.101875	0.029813	0.054674	0.054569	0.054674
17	5	FV. 5	-0.01445	0.025014	0.004432	0.004421	0.004432
18	C:\My Documents\spax.dat						
19	5						
20	5						
21	種の座標づけ						
22	Id	Data	d1	d2	d3	d4	d5
23	1	EV. 1	-0.44366	-0.18428	-0.26321	-0.2893	-0.31496
24	2	EV. 2	-0.69663	-0.65571	-0.67603	-0.68001	-0.68242
25	3	EV. 3	-0.32866	-0.30911	-0.31877	-0.32067	-0.32184
26	4	EV. 4	0.426889	0.592688	0.580689	0.573922	0.565746
27	5	EV. 5	0.166204	0.298747	0.186781	0.147234	0.106968

図4 計算結果をエクセルワークシートに読み込んだ例

ばならないが、新座標づけ法では、地点の座標づけにはできるだけ全種を用い、種の座標づけでは希少種を除くことができる。

こうして希少種を除外したデータファイルを作成後、FileメニューのSpecies ordinationを実行すれば、種の座標づけ結果が出力される。ただし、ここでは固有ベクトルの入力ファイルとして、3・3で出力したファイルを指定する。

3・5 地点および種のクラスタリング

地点および種の座標づけが計算されたら、その結果をエクセルのワークシートに読み込んで(図4)、固有値の大きい順にいくつかの座標成分を用いて散布図を作成する。また、Clusteringメニューにより地点と種のデンドログラムを作成する(2・1・4参照)。これらのデンドログラムあるいは散布図を視察することにより、用いる空間の次元数(座標成分の数)、および妥当なクラスター数を決定して、地点と種のファジイクラスタリングを行う。

3・6 クラスタリング結果の総合評価

最後に Summarize メニューにより、上のファジイクラスタリング結果と、元のデータから、

種群の個体数、種類数および特化係数を計算し、さらに地点群別各種の平均個体数や、各地点の種多様度の計算等を行う。その結果を、地点を各地点群ごとに、地点群への所属率の大きい順に並べてグラフ化すれば、地点群と種群の特性が鮮明に浮かび上がり、様々な考察に資することが出来る。

さらに Summarize メニューの、IndVal の計算と有意性テストを行えば、クラスタリングされた地点（列）の群の有意性を確認できるとともに、その群を特徴づける種（行）を見出す事ができる。

4. プログラムの使用例

PACS を用いた群集データの解析例は、すでに本紀要においても、過去に数編発表されている。そこでここでは、群集データ解析以外の分野での適用例について述べる。

4.1 女子短大生の食生活パターンについての解析

1998年、6月～7月の期間中の3日間、下関女子短大（現下関短大）学生を対象として食事調査が行われ、1日当たり平均摂取栄養素量と、食品別摂取量のデータが集められた。これらのデータについて PACS を用いた解析が行われ、興味ある結果が報告されている。⁶⁾

ここでは、その中から、食品別摂取量について解析した結果を、その後得られた新しい知見とともに紹介する。

4.2 回答者および食品類の座標布置とファジイクラスタリング

食品類（行）×回答者（列）の形式の食品別摂取量データから、品川¹⁾による回答者間の相関指数 R_s と関係類似度 R_{sn} を計算し、回答者と食品類の座標づけを行った。その結果、固有値の累積率は第Ⅲ軸成分までで61.5%となったので、第Ⅲ軸成分までを用いて回答者と食品類のファジイクラスタリングを行った。その場合、クラスター数は、回答者、食品類ともに、意味のある解釈が可能となる3群が妥当と考えられた。以下では、回答者の群を A1, B1, B2 群とし、食品類の群を 1-1, 2-1, 2-2 群と表す。図5には、回答者（左）と食品類（右）を第 I × II 軸平面に布置した結果を示す。図中のマークは、分割された3群への所属率が0.5以上の回答者と食品類を示し、どの群の所属率も0.5未満のあいまいな回答者と食品類は※印で示した。また、表1には、食品類を2分割した場合および3分割した場合それぞれのファジイクラスタリング結果を示す。表中の数字の1, 2は、各群への所属率が0.5以上であること示し、0は、3分割で所属率が0.5未満となったあいまいな食品類を示す。

表に示すように、食品類は、まず2分割で1群と2群に大別され、さらに3分割で2群が

表1 食品類のファジイクラスタリング結果

	2分割	3分割
いも類	1	1-1
緑黄色野菜	1	1-1
砂糖類	1	1-1
魚介類	1	1-1
その他穀類	1	1-1
海草類	1	1-1
果実類	1	1-1
嗜好飲料	1	1-0
油脂類	2	2-1
調味料	2	2-1
大豆・大豆製品	2	2-1
米	2	2-1
肉類	2	2-1
その他の野菜類	2	2-1
小麦類	2	2-2
乳・乳製品	2	2-2
卵	2	2-0
菓子類	2	2-0

食品類のクラスター1, 2 (2分割の場合) および 1-1, 2-1, 2-2 (3分割の場合) への所属率が 0.5 以上の食品類を所属率の大きい順に示す。ただし, 1-0, 2-0 は, 3分割で所属率が 0.5 未満となった食品類である。

2-1群と2-2群に細分される。このことは、食品の嗜好に階層構造の存在することを示唆するものと考えられ、興味深い。

4.2.1 嗜好を決定する2つの要因

この解析では、回答者間の類似度に基づいて座標づけを行っているので、回答者の座標布置の場合には原点の位置に意味はある（原点からの位置ベクトル間の余弦が R_{sn} を表す）が、食品類の場合には意味がなく、その位置は任意に変え、さらに回転して結果を解釈してもよい。そこで、図5の座標軸原点を、所属があいまいな卵および菓子類の付近に移動して、なおかつ直交座標系を右回りに45°回転して第1軸および第2軸（図中の矢線）とすると、食品類は第1軸によって1群と2群に2分され、さらに第2軸によって2-1群と2-2群が分離される。すなわち、回答者の食品に対する嗜好は、この2つの軸で示される要因によって大きく決定づけられているといえる。

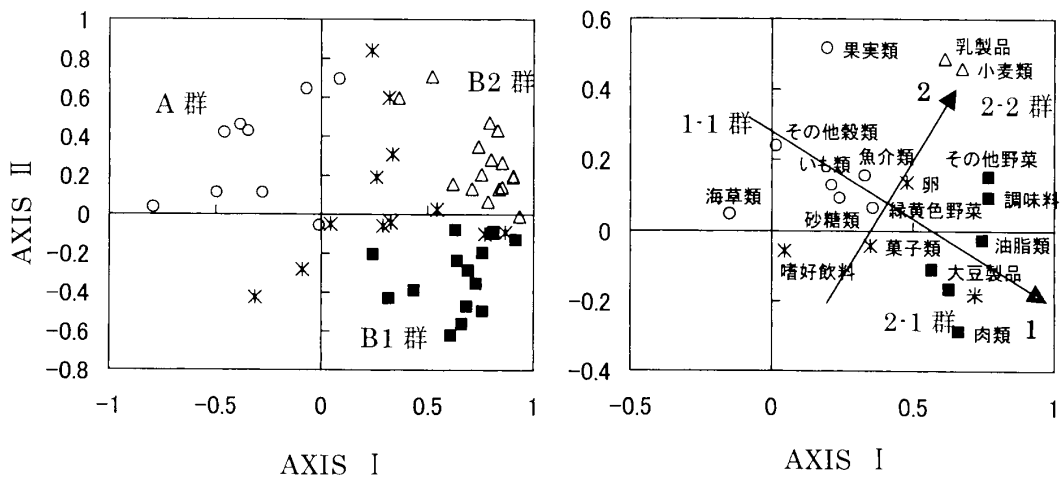


図5 食品類別摂取量の類似度に基づく対象者（左）と食品類（右）の座標布置

4・2・2 回答者各群の食品摂取パターンと有意性テスト

各回答者の食品類摂取データから、回答者の各群について、Dufrene and Legendre の IndVal 1 を計算し、さらに回答者をランダムに並べ替える500回の繰り返し計算によって有意性テストを行った。その結果得られた各食品類のz値（標準正規変数）を図6に示す。図の横軸では、食品類をA1, B1, B2群の順に、かつそれらへの所属率が大きい順に並べている。z値の性質から、その絶対値が1.96以上であれば、有意水準5%で、その回答者群が全体の平均値から有意に偏っていると判定される。

図から、回答者のA1群は、食品類の1-1群で全体の平均値より正の方向に偏る一方、食品類の2-1群では負の方向に偏っている。また、食品類の2-2群についてはほぼ全体の平均値である。回答者のB1群は、A1群とは逆に食品類の2-1群では正の方向に、食品類の1-1群と2-2群では負の方向に偏っている。回答者のB2群では、食品類の2-1群と2-2群の中間付近に位置するその他野菜類、および2-2群の小麦類と乳・乳製品で正の方向への偏りが大きい。

上述の食品類摂取パターンの特徴としてまず言えることは、A1群は主食となる食品の摂取量が少なく、B1群は主食として米類を多く摂取し、B2群は小麦類を多く摂取しているということである。このことから、主食の摂取の仕方が、食生活のパターンを決める重要な要因になっていると考えられる。すなわち、食生活のパターンは、まず主食をあまり摂らない群と摂る群に2分され（第1軸）、さらに主食を摂る群が米食中心と小麦食中心に細分される（第2軸）ことによって、パターンの階層構造が形成される。また、主食の少ない群では嗜好飲料や果実類を多く摂る傾向があり、米を多く摂る群では肉類、大豆製品などの蛋白質食品と同時に油脂類と調味料も多く摂る傾向がある。さらに、小麦類を多く摂る群では、乳・乳製品とともに、緑黄色ではない、その他の野菜類を多く摂る傾向がみられる。

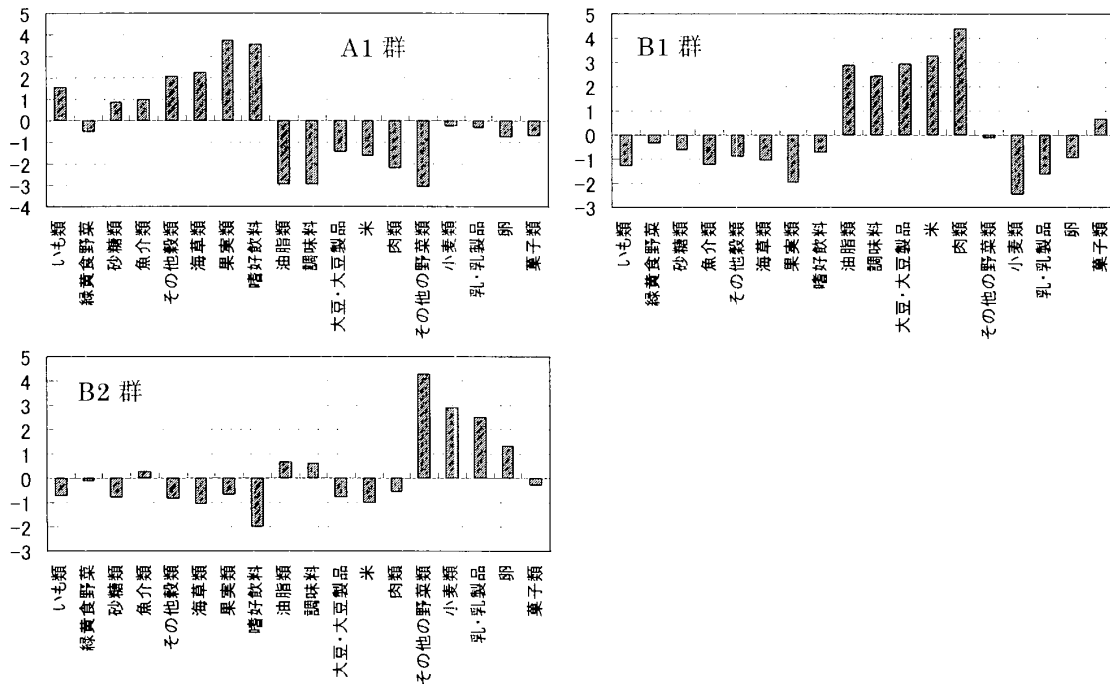


図6 クラスタリングされた回答者群A1, B1, B2群の, 各食品類別指標値
500回のランダムアロケーションによる繰り返し計算で, ファジイクラスタリングに拡張したDufrene and Legendre¹⁶⁾のIndicator value を標準正規変数に変換した値. 数値が大きいかほどその食品類を全体の平均より多く摂取していることを示す.

このような傾向は, いずれも最近の食生活の問題点として指摘される傾向であるが, 座標づけによる解析の結果は, それが主食の摂り方と密接に関わって生じていることを示すものであり, 栄養指導を行う上での重要な視点になると思われる.

4.3 職業ストレスと問題飲酒に関するアンケート調査データの解析

社会現象における非数量的なデータ解析の方法として, 今日一般的に行われている方法に対応分析法(数量化Ⅲ類とも呼ばれる)がある. この方法は, アンケート調査など, 0又は1からなるダミー変数に対して主成分分析に代る有効性があるとされている.¹⁹⁾

そこでここでは, 中尾ら⁷⁾によって行われた職業ストレスと問題飲酒の関係についてのアンケート調査データを新座標づけ法と対応分析法により解析して比較した結果を紹介する.

4.3.1 アンケートの内容

このアンケート調査は, 1998年9月~11月に某企業の男性労働者, 日勤者のスタッフ系98名と交替勤務者の技能系155名を対象として行われた.⁷⁾ 質問項目は, 勤務内容(事務系, 技術系), 勤務形態(常昼, 交代制), 労働時間など, 勤務条件に関わる項目と, 作業環境ストレス, 業務関連ストレス, 蓄積疲労徴候, 鬱尺度など, 職業ストレスに関わる項目, および飲酒習慣に関わる項目であった.

4・3・2 回答カテゴリーと回答者の座標布置

アンケートの質問に対する各回答カテゴリーに該当する場合は1, 該当しない場合は0として各人の回答をダミー変数に変換し, PACS を用いて新座標づけ法と対応分析法による座標づけを行った. ただし, 新座標づけ法における各回答カテゴリー間の相関指数 R_s の計算では, 多様度の類似度はこの場合は無意味と考えられるので省略した.

各回答カテゴリーと回答者を, 新座標づけ法および対応分析法で座標づけした結果を図7と図8に示す. 左図の中で, 各回答カテゴリーに添えられた数字は, 図の説明に示す勤務条件と,

3

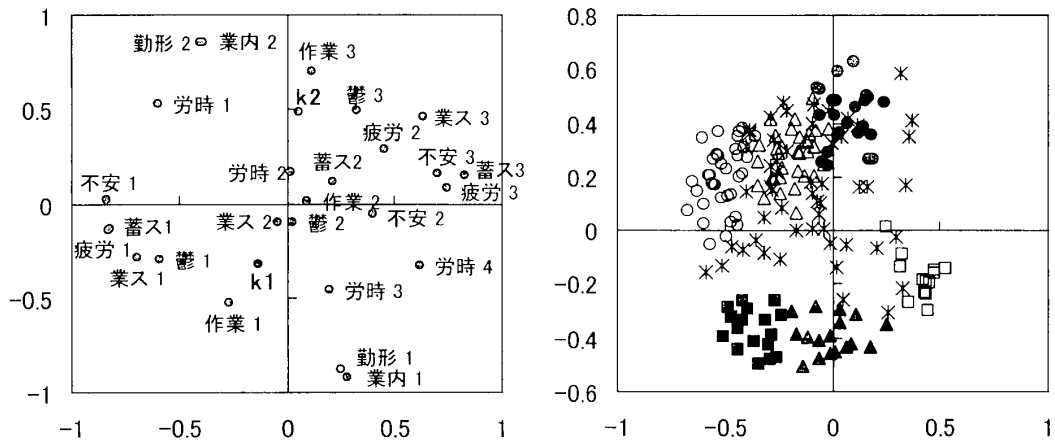


図7 新座標づけ法による, 各質問項目への回答カテゴリー (左) と回答者 (右) の座標布置
 左図中の文字は質問項目を, 数字は回答カテゴリーを表し, ストレスと鬱度の項目については, 数字が高いほどストレスと鬱の度合いが高いことを示す.
 勤形 1, 常昼勤務; 勤形 2, 交代勤務; 業内 1, 事務職; 業内 2, 技術職;
 k 1, 正常飲酒習慣; k 2, 問題飲酒習慣;
 労時 1~4, 労働時間; 作業 1~3, 作業環境ストレス;
 業ス 1~3, 業務ストレス; 蓄ス 1~3, 蓄積ストレス;
 不安 1~3, 不安感; 疲労 1~3, 慢性疲労; 鬱 1~3, 抑鬱.

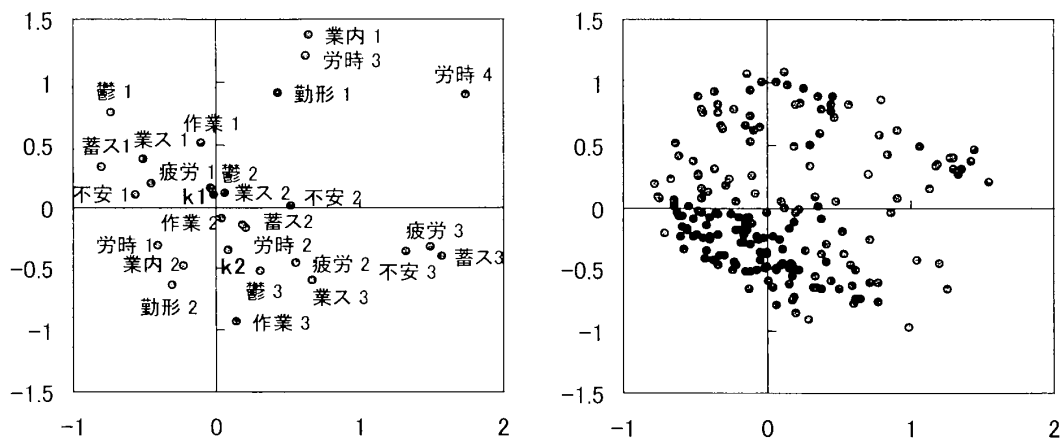


図8 対応分析法による, 各質問項目への回答カテゴリー (左) と回答者 (右) の座標布置
 左図中の文字と数字は図7を参照.

職業ストレスのレベルを表し、また、k1は正常な飲酒習慣、k2は問題飲酒習慣を表す。また、図7右図中のマークは、4・2と同様に、回答者のファジイクラスタリング結果を示す。

4・3・3 新座標づけ法と対応分析法による結果の比較

新座標づけ法の結果と対応分析法の結果を比較すると、新座標づけ法の結果では、問題飲酒習慣(k2)と正常飲酒習慣(k1)が明瞭に分離し、それに対応する回答者の群も識別される。一方、対応分析法の結果ではそれほど明瞭には分離されない。この新座標づけ法の結果からは、問題飲酒習慣が作業環境ストレスや鬱尺度が高いことと深く関連し、また、それは技術系の交代勤務とも関連していることを読み取ることが出来る。一方、対応分析法の結果では、k1とk2の分離が悪く、このために問題飲酒者と正常者との群の識別が明瞭ではない。こうしたことから、新座標づけ法による結果の方が、クラスター解析に適した結果を与えているといえる。

5. 結 言

生物や、人の社会生活に関わる現象は、確率現象とは異なる本質的なファジイさを特徴とする。そのような混沌としたあいまいさの中に一定の規則性を見いだすことは、データ解析の課題であり、かつまた醍醐味である。

今日、経済的・社会的・文化的な大きな変化を背景として益々多様化する人々の食生活は、一層ファジイな性質を強めている。その中に潜む食生活のパターンを識別し、その要因を明らかにすることは、社会問題ともいえる成人病予防のための食生活の改善指導に重要な指針を与えるものとなるであろう。

食生活を規定する最大因子の一つとされる食嗜好は若年期に確立するといわれ、⁶⁾この時期における食嗜好の決定要因を明らかにすることはとりわけ重要である。しかしながら、個々人の食嗜好は、これまでその人が過ごしてきた様々な社会的・経済的・文化的環境の中で形成されてきたものであるから、まさしくファジイであり、そのデータの解析には特有の困難さがある。

これまでも摂取食品類の構造解析に、池田ら²⁰⁾による主成分分析法や重回帰分析法、さらに豊川ら¹¹⁾による親和性指数や偏差パターン類似率による解析などが試みられているが、これらを我々のデータに適用した結果では、いずれも意味のある結果を得る事が出来なかった。その理由は、個々人の食品類摂取データには極端な偏りや0の数値が多く含まれており、たとえ対数変換などのデータ変換を行っても、その非正規性が改善されないからである。

このように、食品類摂取量のデータは不均質性とランダム性が大きく、均質性と正規性を前提とする解析法には向かない。こうした問題が生じることは、生物群集の個体数データ解析に

においても同じであり、これを克服するために開発された生物群集解析のための手法を食品類摂取パターンの解析に適用することは、当然の成り行きである。

今回、女子短大生の食品類摂取パターンの解析に PACS を用いたことはそうした試みの一つであり、食生活の実態を解明し、その改善指導を行う方法に新しい展望を開くものといえる。

PACS の適用可能性は、食品類摂取のような数量データだけでなく、アンケート調査データのようなカテゴリカルなデータ解析においても効果的であることが、職業ストレスと問題飲酒に関するアンケートデータの解析によって示された。このようなデータの場合、一般には対応分析法（数量化Ⅲ類）が適用されることが多い。対応分析法は、データの行と列を同時に座標づけして、その中にいくつかの軸を見出すことを目的としており、その点では主成分分析法や因子分析法と同様である。一方、PACS に収録されている新座標づけ法は、座標づけ空間の中に自然な群を見出すことを目的としている。

ここに紹介したアンケートデータの解析結果によれば、新座標づけ法は、問題飲酒者の群と正常者の群を明瞭に識別できたという点で、対応分析法より優れているといえる。新座標づけ法が、対応分析法に比べてランダムな変動の影響を受け難く、データの中に潜む自然な群を見いだすのに優れていることは、すでに数値実験結果²¹⁾により指摘されているが、今回行った、データ形式が異なる2つの解析例においても、そのことが実証された。

これらのことは、ここに紹介したプログラムパッケージ PACS が、広範囲な分野のファジイな現象の探索的解析に、大きく貢献できることを示すものといえよう。

6. 謝 辞

本プログラムの作成に際しては、東京大学海洋研究所 海洋科学国際共同センター研究生、橋詰和慶博士に、解析法等について多くのご助言を頂いた。ここに記して、厚く感謝の意を表する。

文 献

- 1) 品川汐夫：日本ベントス研究会誌，26，49～65(1984)
- 2) 品川汐夫(1998)：ファジイクラスター解析を用いた新しい生物群集解析法，長崎大学学位請求論文(学術)，170pp.
- 3) 品川汐夫・多部田修：日本水産学会誌，64，418～426(1998)
- 4) 品川汐夫・多部田修：日本水産学会誌，64，796～806(1998)
- 5) 品川汐夫：下関女子短期大学紀要，No.18，1～18(1999)
- 6) 若本ゆかり・品川汐夫・他：体力・栄養・免疫学雑誌，10，21～28(2000)
- 7) 中尾久子・品川汐夫・他：臨床環境医学，10，78～84(2001)

- 8) S. Kimoto. : Esakia, 5, 1~20(1967)
- 9) J.R. Bray and J.T. Curtis, *Ecol. Monogr.*, 27, 325~349(1957)
- 10) D.W. Goodall : Sample similarity and species correlation. In "Ordination of Plant Communities" (R. H. Whittaker, Ed.), W. Junk, The Hague, 1978, pp. 99~149.
- 11) 豊川裕行・西川浩昭・城田知子他 : 日本栄養・食糧学会誌, 48, 253~270(1995)
- 12) W.T. Williams and W. Stephenson : *J. Exp. Mar. Biol. Ecol.*, 11, 207~227(1973)
- 13) 宮本定明 : クラスタ分析入門, ファジィクラスタリングの理論と応用, 初版, 森北出版, 東京, 1999, 176pp.
- 14) 上田尚一 : データ解析の方法, 初版, 朝倉書店, 東京, 1982, pp. 5~10.
- 15) 木元新作 : 動物群集研究法 I, 初版, 共立出版, 東京, 1976, pp. 54-94, pp. 131~166.
- 16) M. Dufrene and P. Legendre : *Ecol. Monogr.*, 67, pp. 345~366(1997)
- 17) 柳井晴夫 : 多変量データ解析法, 初版, 朝倉書店, 東京, 1991, pp. 46~51.
- 18) 柳井晴夫 : 多変量データ解析法, 初版, 朝倉書店, 東京, 1991, pp. 108~124.
- 19) 西里静彦 : 質的データの数量化, 第1版, 朝倉書店, 東京, 1982, pp. 1~60.
- 20) 池田順子・浅野弘明・木村みさか・永田久紀 : 日本公衛誌, 29, 616~625(1982)
- 21) 品川汐夫・多部田修 : 日本水産学会誌, 64, 56~64(1998)
- 22) 品川汐夫 : 下関女子短期大学紀要, No.17, 1~20(1998)

付録 新座標づけ法^{1, 2)} について

1. 主成分分析法の基本原則

p 次元空間における n 個の点の座標を $p \times n$ の行列 \mathbf{Y} で表し, \mathbf{Y} の各列が各点の座標を表すものとする. つぎに行列 $\mathbf{Y}\mathbf{Y}$ を考え, $\mathbf{Y}\mathbf{Y}$ のランク数を r ($r \leq p$) とする. 行列の理論から, $\mathbf{Y}\mathbf{Y}$ のランク数は \mathbf{Y} のランク数に一致する.

ランク数とは, \mathbf{Y} の中の独立な座標ベクトルの個数で, 言い換えれば, n 個の点が存在する空間の最低次元数である. 例えば, n 個の点が全て一つの平面上に散布していれば, ランク数は2である.

このような, もとの p 次元空間の中に全ての点が載っている r 次元の部分空間を求めることは次元の縮約となり, データ解析の有力な武器となる.

この空間は, 行列 $\mathbf{Y}\mathbf{Y}$ を固有値分解することによって得られる. すなわち, 行列 $\mathbf{Y}\mathbf{Y}$ の r 個の固有ベクトルが張る空間がそのような次元を縮約した空間で, この空間内における各2点間の距離は, もとの p 次元空間内における距離に等しくなっている. なおかつ, 新しい空間における軸に関する成分の2乗和は, 対応する固有値に等しくなっているから, 固有値の大きい順に適切な個数の軸を選べば, 同じ軸の数(次元数)で最大の情報を保存する空間になる. またこの軸の方向は, もとの空間内で最も点が大きく散らばっている方向(主軸)に一致する. この方法が主成分分析と呼ばれる所以である.

\mathbf{Y} として, 生データ行列 \mathbf{X} を中心化(平均値を差し引くこと)した行列を用いれば $\mathbf{Y}\mathbf{Y}$ は

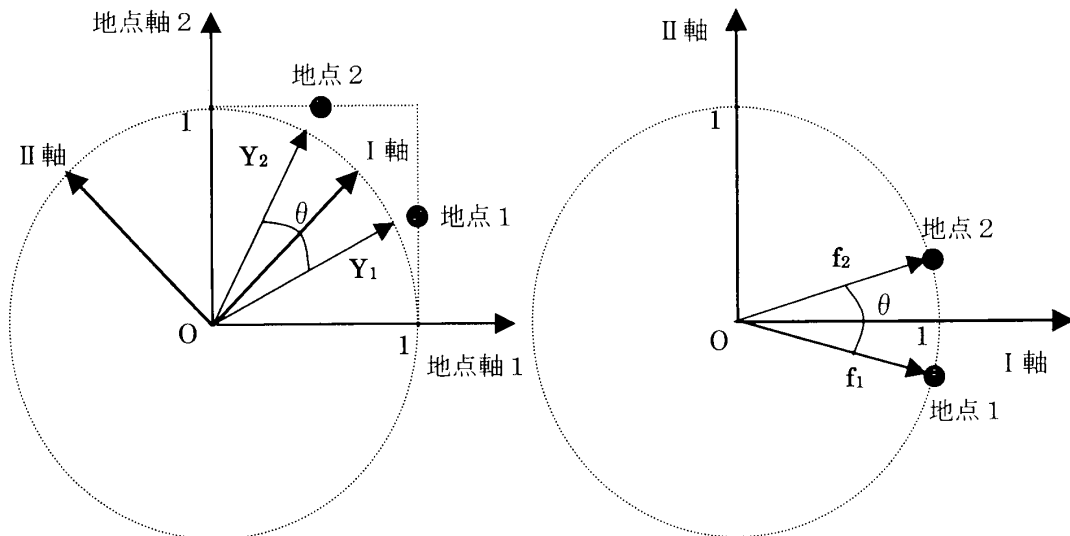
分散行列を表し（定数倍を別として），中心化後さらに標準偏差で除した行列を用いれば $Y'Y$ は相関行列になる．従って，通常の（中心化した）主成分分析は，分散行列又は相関行列の固有値解析として一般には説明される．

2. 新座標づけ法と主成分分析法の関係

品川^{1,2)}の地点間相関指数を成分とする行列を R_s とし，その各列ベクトルの長さの2乗を対角成分とする対角行列を N とする．このとき， $Y=RsN^{-1/2}$ とすれば， $Y'Y$ は品川^{1,2)}が定義した関係類似度行列 R_{sn} になる．したがって，関係類似度行列 R_{sn} の固有値分解は，行列 R_s に，中心化を行わず基準化だけを行った主成分分析を適用することを意味している．中心化しない座標づけの利点については，品川²⁾が述べている．ちなみに， R_s を中心化後のデータ行列とすれば， N の対角項は分散を表し（定数倍を別として），関係類似度行列 $Y'Y$ は相関行列そのものである．したがって，新座標づけ法は行列 $R_sN^{-1/2}$ の各列ベクトルで示される点群を，各点間の関係を最もよく保存する低次元の空間に直交射影することを意味している．

3. 座標づけ空間の座標軸について

主成分分析では，直接各変数軸空間における原データの点群を低次元空間に射影するので，軸を変数軸空間に描くことが出来るが，新座標づけ法で強いてそれをするならば，地点を軸とする空間内に R_s で表される点群を描き，その空間内に新しい（回転した）軸を描くことになる．たとえば，

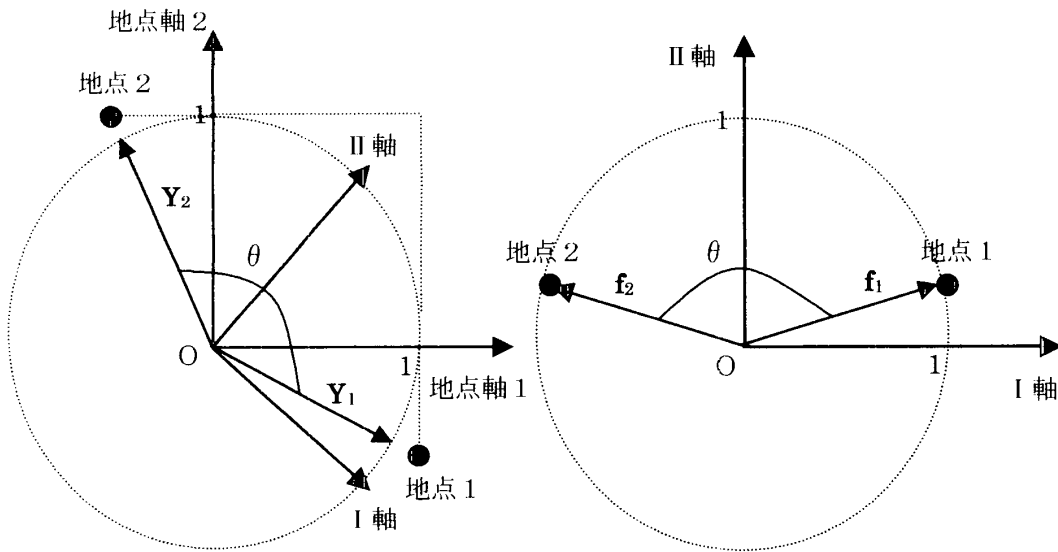


図A-1 相関指数行列 R_s による2地点の地点軸空間での布置（左）と固有ベクトル空間での布置（右）その1

$$R_s = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \text{ の場合 (図 A-1),}$$

$$Y = R_s N^{-1/2} = \begin{bmatrix} 0.894 & 0.447 \\ 0.447 & 0.894 \end{bmatrix}, \quad Y^* Y = R_s n = \begin{bmatrix} 1 & \cos \theta \\ \cos \theta & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$

これから、 $R_s n$ の固有値分解の結果、座標づけ位置ベクトル (要因ベクトル) として、 $f_1 = [0.949, -0.316]$ 、 $f_2 = [0.949, 0.316]$ が得られる。この場合の2点間の距離は0.632で、 Y の2点間の距離に一致する。 f_1 、 f_2 の長さは1に等しく、その間の角度はもとの空間内のそれと一致し、その余弦が関係類似度を表す。



図A-2 相関指数行列 R_s による2地点の地点軸空間での布置 (左) と固有ベクトル空間での布置 (右) その2

$$R_s = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix} \text{ の場合 (図 A-2),}$$

$$Y = R_s N^{-1/2} = \begin{bmatrix} 0.894 & -0.447 \\ -0.447 & 0.894 \end{bmatrix}, \quad Y^* Y = R_s n = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}$$

これから、要因ベクトルとして、

$$f_1 = [0.949, 0.316], \quad f_2 = [-0.949, 0.316]$$

が得られる。この場合の2点間の距離は1.897で、 Y の2点間の距離に一致する。 f_1 、 f_2 の長さは1 (単位球面上) で、位置ベクトル間の角度はもとの空間内のそれと一致し、その余弦が関係類似度を表す。

これらのことから、新座標づけ法は、行列 R_s に基準化を伴う主成分分析を適用する方法で、 R_s で表される地点間の関係を、グラフィカルに表現する方法といえる。具体的には、行列 R_s を座標行列とみなして、各点を先ず単位球面上に射影 (基準化) し、それらの各2点間の距離

が最もよく保存される低次元の空間を求める方法である。その場合の軸は、地点軸空間（行列 \mathbf{R}_s の各列を座標とした空間）内で、点群の原点からの散らばり（重心からではない）が最も大きくなる方向（中心化した主成分分析では重心からの散らばり）である。その意味については、軸方向に分布する地点の性質から判断しなければならないが、主成分分析のように変数軸空間内ではないので、変数（ここでは種）の性質から直接的に説明することはできない。またそれは、この座標づけの主目的ではないので、主成分分析のように軸の意味にこだわる必要はない。この座標づけの主目的は、最適な群を見出すことにある。すなわち、クラスタリングを主目的とした座標づけで、その点が、従来の環境傾度（の軸）を見出すことを主目的とする座標づけとはやや異なる。