

連載 (講義)

Common Data Processing System Version 10 の使用法

— (7) 多変量解析 —

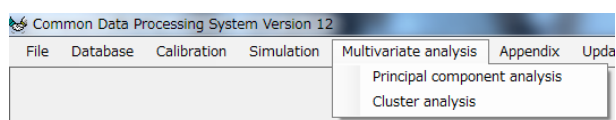
吉原 一紘*

オミクロンナノテクノロジー・ジャパン(株)
140-0002 東京都品川区東品川3-32-42 IS ビル
*k.yoshihara@omicron.oxinst.com

(2015年1月26日受理)

9. 多変量解析

COMPRO には多変量解析のために、主成分分析 [Principal component analysis] とクラスター分析 [Cluster analysis] が組み込まれている。メニュー画面の [Multivariate analysis] をクリックすると多変量解析の選択画面が現れる。

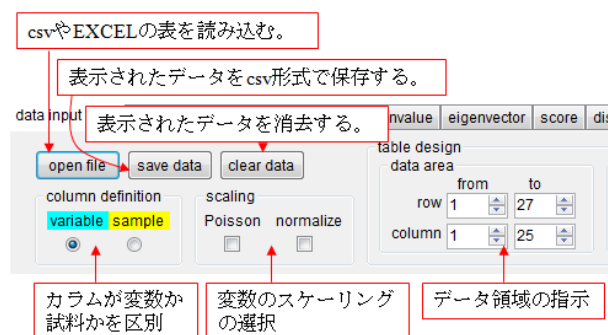


9.1. 主成分分析

[Principal component analysis] を選択するとデータの入力テーブルが現れる。(但し、下の画面は入力済みの画面である。)

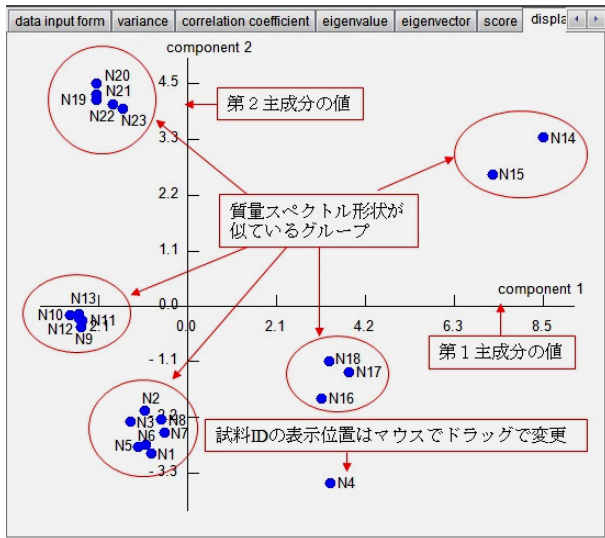
data input form	variance	correlation coefficient	eigenvalue	eigenvector	score	display score		
1	2	3	4	5	6	7	8	
1	質量	分子名	N1 ← 試料ID	N3	N4	N5	N6	
2	11.99898...	C	22323	14957	33431	26856	28653	1504
3	15.02343...	CH3	104604	124681	27365	103573	3748	
4	27.02280...	C2H3	217804	198699	31712	178575	1798	
5	27.97532...	Si	59895	25025	54543	580141	70614	4606
6	29.03927...	C2H5	145936	156847	128788	24196	122145	1519
7	30.99780...	CF	3696	3766	8173	15424	8106	4822
8	41.03921...	C3H5	303193	300003	283310	46246	261713	2723
9	46.99202...	COF	755	690	1183	4258	1598	870
10	49.99767...	CF2	1603	968	2079	4018	1734	1265
11	55.05583...	C4H7	174899	227247	167685	54250	169767	2402
12	68.99449...	CF3	3701	2215	3155	7060	2957	2090
13	73.06246...	SiC3H9	164682	77503	91343	593360	132789	9709
14	77.02280...	CF4	20753	25540	34434	5505	25200	2003

データ入力の表の下段にコントロールパネルが現れる。あらかじめデータが csv 形式か EXCEL で事前に作成されていれば、[open file] ボタンをクリックすると、それを読み込み、入力テーブルに表示できる。

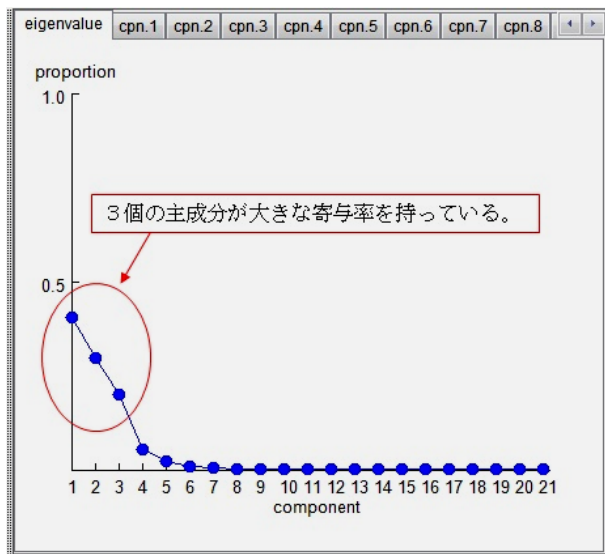


TOF-SIMS のデータを例にとって説明する。列 1 には質量数、列 2 には分子名となっており、列 3 からデータが入力されている。行 1 には試料 ID が入力され、行 2 からカウント数が入力されている。

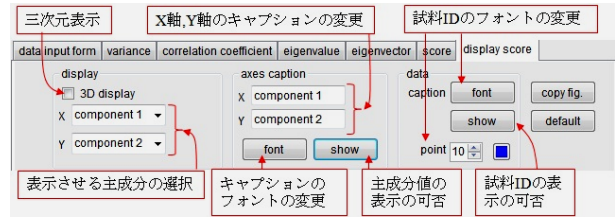
入力するデータの表形式はユーザーによって異なるため、ユーザーによる行と列の定義が必要である。入力した後、[column definition] のグループボックスで列が何を示しているかを指示する。この例の場合では、列は試料ごとのデータなので [sample] ボタンをクリックする。次にデータ領域を [data area] グループボックスの [row] と [column] の値を変更することにより決定する。通常の表では最終列、最終行は自動的に表示されるが、開始列、開始行は手動で指示する。この例の場合では、開始行は <2>、開始列は <3> を指示する。入力データの数値をあらかじめ [Poisson scaling] あるいは [normalize] することも可能である。必要があれば当該チェックボックスにチェックを入れると、入力テーブルのデータ表示が、scaling 処理された値に変わる。今回例示したデータには normalize 処理を行った。データの範囲を指定後、テーブルの [display score] タブをクリックすると主成分分析結果が表示される。



各試料の第1主成分の値と第2主成分の値を表示した結果が図示される。図からは、円で囲んだ試料群から得られた質量スペクトルの形状は似ていることが分かる。このグラフ表示画面の右側のパネルには主成分（固有ベクトル）とその寄与率（固有値の大きさ）が表示される。



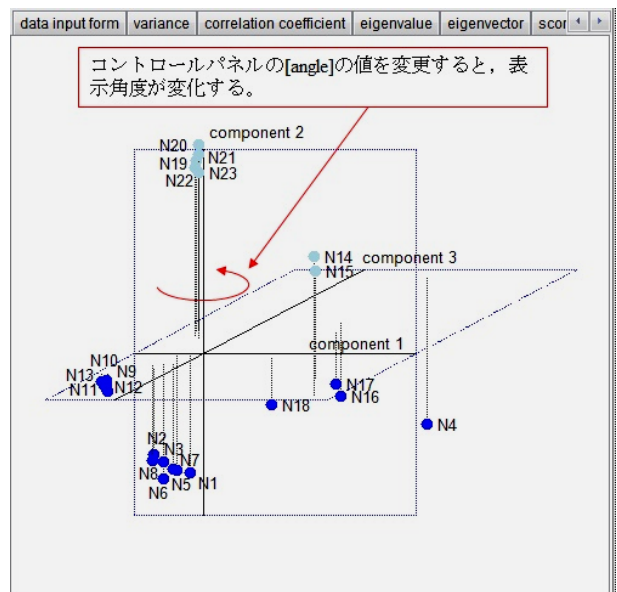
この TOF-SIMS のデータを分類するには、主成分の寄与率から主成分を3個選択する必要があると推定される。主成分分析では、二次元表示が普通であるが COMPRO では三次元表示まで可能である。分析結果表示画面の下にコントロールパネルが表示される。



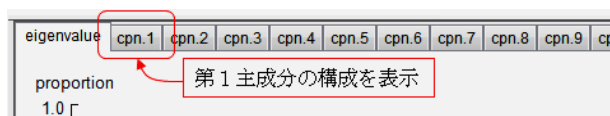
表示させる主成分は、コントロールパネル左端の [display] グループボックスの [X] と [Y] のコンボボックスから選択する。X 軸、Y 軸のキャプションの変更、フォントの変更など、グラフを見やすくさせるためのオプションがグループボックスにまとめて表示されている。



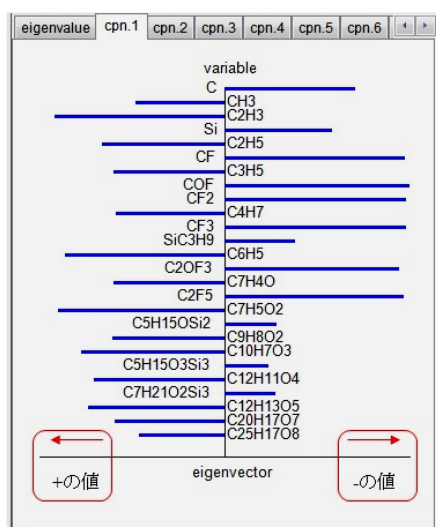
[3D display] をチェックすると、次図に示すように3次元表示され、表示できる主成分は2個から3個に拡大される。X 軸に第1主成分、Y 軸に第3主成分、Z 軸に第2主成分を表示させて、3次元表示したグラフを示す。表示グラフの下部にコントロールパネルが現れる。コントロールパネルの [angle] の値を変更すると3次元表示の表示角度が変わる。X 軸、Y 軸、Z 軸にどの主成分を表示させるかは自由に選択できる。



二次元表示では同一グループに属していた(N16, N17)と N18 は、3次元表示から、別グループと判断した方が良いということが分かる。



主成分（固有ベクトル）の構成は、たとえば第一主成分については[eigenvalue]タブに続いて現れる[cpn.1]タブをクリックすると表示される。



各試料の質量スペクトルが第1主成分の軸上で、どの位置にあるかは、主成分（固有ベクトル）の各変量(C, CH3, C2H3, Si, ...)の比率から計算できる。このグラフから、CやSiを含むスペクトルは第1主成分の値が小さく、C-Hを含むスペクトルは逆に値が大きくなることを示している。この構成図と比較することにより、データの分類理由が推定できる。

COMPROでは、主成分分析の計算過程も表示される。データ入力画面で[variance]タブを選択すると、各変量の分散が表示される。[correlation coefficient]タブをクリックすると相関係数の計算結果が表示される。[eigenvalue]タブをクリックすると、主成分ごとの寄与率、寄与率の積算値、固有値が表示される。[eigenvector]タブをクリックすると主成分ごとの構成比が表示される。この表が[cpn.1], [cpn.2]などのタブをクリックするときに表示される図の元データである。[score]タブをクリックすると、各試料の主成分ごとの値が表示される。[display score]タブで表示さ

れる図の元データである。

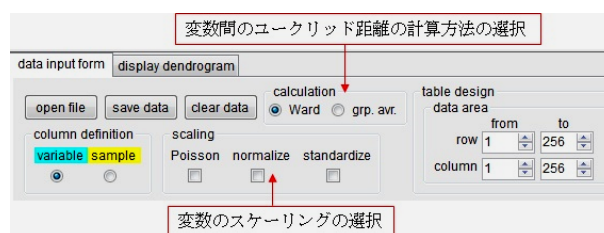
分析を終了するときには[Return]タブをクリックする。

ここでは、主成分分析法の原理の説明は行わなかったが、簡単な解説はJSA誌に掲載されているので参照されたい[1]。

9.2. クラスタ分析

[Cluster analysis]を選択するとデータの入力テーブルが現れる。入力テーブルは[Principal component analysis]と全く同じ形式であり、あらかじめデータがcsv形式かEXCELで事前に作成されていれば、[open file]ボタンをクリックすると、それを読み込み、入力テーブルに表示できる。

入力テーブルの下部にコントロールパネルが現れる。入力データの読み込み方法やデータ領域の定義方法などは[Principal component analysis]と全く同じなので、繰り返しては説明しないが、前項の説明と異なる箇所は計算方法の選択があることと、scaling方法の選択である。



クラスタ分析では変量間の相違をユークリッド距離で判断するが、その計算方法にWard法と群平均法がある。Ward法を選択する場合には[calculation]グループボックスの中の[Ward]にチェックを入れ、群平均法を選択する場合には[grp. avr.]にチェックを入れる。なお、デフォルトではWard法を選択することになっている。

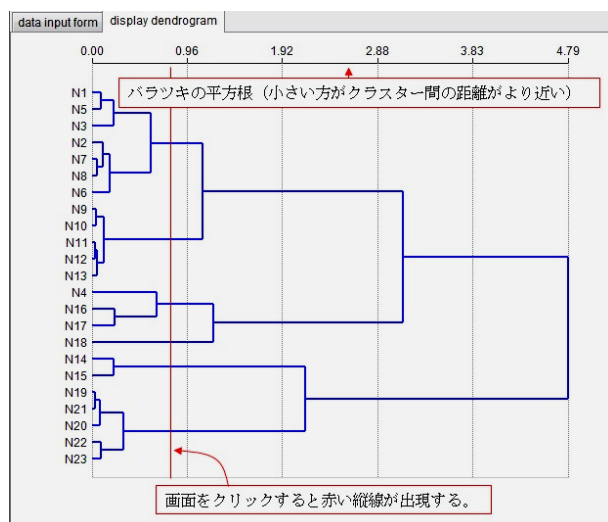
入力データの数値をあらかじめ[Poisson scaling]あるいは[normalize]することが可能であることは、主成分分析の場合と同じであるが、クラスタ分析にはそれに加えて[standardize]の選択が可能である。データの大きさが変量によって大きく異なる場合には、ユークリッド距離をそのまま用いるとデータ群を正しく分離することが出来ないことがある。このような場合にはデータをあらかじめ標準化(standardize処理)しておくことが望ましい。

例として、前出の主成分分析の項で説明に用いたTOF-SIMSのデータを用いる。この場合も同様に

データは normalize 処理しておく。

data input form		display dendrogram					
1	2	3	4	5	6	7	8
1	質量	分子名	試料ID	N3	N4	N5	N6
2	11.99898...	C	0.072947...	0.048692...	0.117054...	0.043926...	0.108504...
3	15.02343...	CH3	0.34452...	0.439485...	0.044785...	0.395086...	0.131...
4	27.02280...	C2H3	0.71816...	0.701027...	0.052122...	0.681982...	0.65...
5	27.97532...	Si	0.196959...	0.082293...	0.191653...	0.977690...	0.269013...
6	29.03927...	C2H5	0.480950...	0.522233...	0.453997...	0.039437...	0.466128...
7	30.99780...	CF	0.011466...	0.011343...	0.027805...	0.024633...	0.029909...
8	41.03921...	C3H5	1	1	1	0.076650...	1
9	46.99202...	COF	0.001759...	0.001077...	0.003105...	0.005788...	0.005014...
10	49.99767...	CF2	0.004558...	0.002005...	0.006271...	0.005383...	0.005535...
11	55.05583...	C4H7	0.576546...	0.757185...	0.591439...	0.090158...	0.648290...
12	68.99449...	CF3	0.011482...	0.006167...	0.010073...	0.010517...	0.010213...
13	73.06246...	SIC3H9	0.542824...	0.257432...	0.321685...	1	0.506843...
14	77.02280...	C2H5	0.004470...	0.004044...	0.100007...	0.000013...	0.000074...

データの入力と計算方法の選択と scaling が終われば[display dendrogram]タブをクリックすると、結果がデンドログラムで表示される。縦軸は試料名、横軸は、Ward 法で計算を実施したため、バラツキの平方根(小さい方がクラスター間の距離がより近い)で表示されている。



グルーピングをどの距離で判定するかは原則として解析者に依存する。グルーピングの判定の便宜のために、画面をクリックすると赤い縦線が出現するようになっている。この赤線の位置でグルーピングすると(N1, N5, N3, N2, N7, N8, N6), (N9, N10, N11, N12, N13), (N4, N16, N17), (N18), (N14, N15), (N19, N21, N20, N22, N23)の6グループに分類できる。なお、横軸の小さな値の部分を広大表示するために、グラフ表示画面の下部に現れるコントロールパネルの中の[log scale]にチェックを入れると、横軸の値が対数に変換されて表示される。

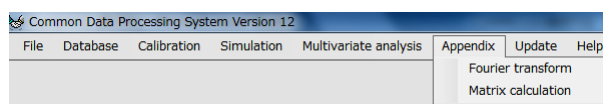
分析を終了するときには[Return]タブをクリック

する。

クラスター分析法の原理の説明は行わなかったが、簡単な解説はJSA誌に掲載されているので参照されたい[2]。

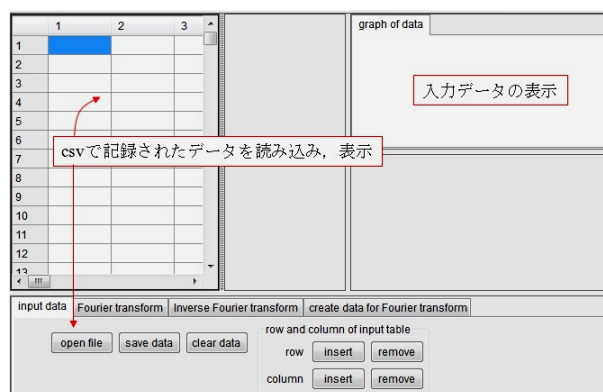
10. 付録

COMPRO ではフーリエ変換や行列計算をしばしば利用しているが、これらの計算方法を独立して利用できるようにしてある。ただし、これはフーリエ変換や行列の概念を把握することが主目的であるので、ここでは簡単な説明に留める。メニュー画面の[Appendix]をクリックすると付録に登録されている[Fourier transform]か[Matrix calculation]の選択画面が現れる。



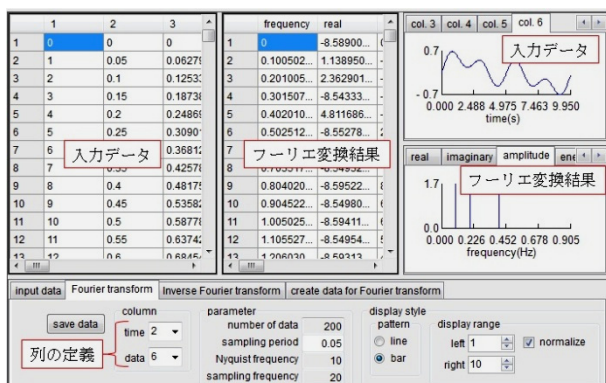
10.1. フーリエ変換

[Fourier transform]タブをクリックすると、データの入力画面が現れる。



フーリエ変換したいデータがあらかじめ csv 形式で保存されていれば、画面下部のコントロールパネルの[open file]ボタンをクリックすることにより画面左の入力テーブルに転記できる。入力データは[graph of data]タブページにも表示される。

データを入力後、コントロールパネルの[Fourier transform]タブをクリックすると、フーリエ変換結果が表示される。

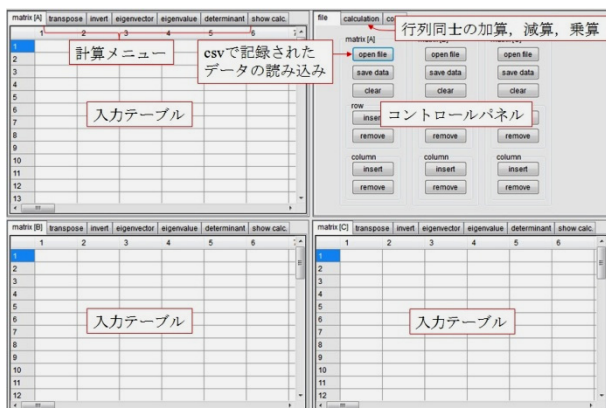


[Inverse Fourier transform]タブをクリックすれば逆フーリエ変換が行われる。また、[create data for Fourier transform]タブをクリックすると、あらかじめ組み込まれている三角関数や指数関数を合成した関数が作成でき、それをフーリエ変換することで、変換の意味が理解できるようにしてある。

解析を終了するときには[Return]タブをクリックする。

10.2. 行列計算

[Matrix calculation]タブをクリックすると、データの入力画面が現れる。



入力テーブルが3画面 (A, B, C), コントロールパネルが1画面現れる。C画面はA画面の行列とB画面の行列の加算, 減算, 乗算結果の表示に使用される。行列データがあらかじめ csv 形式で保存されていれば, [open file]ボタンで読み出して入力テーブルに転記できる。行列の転置[transpose], 逆[inverse], 固有ベクトル[eigenvector], 固有値[eigenvalue], 行列式[determinant]は入力テーブルの上部のタブを選択すると表示される。

解析を終了するときには[Return]タブをクリックする。

11. ヘルプ

[Help]メニューには[Tip], [Help], [Mail to COMPRO], [Version]が含まれる。

[Tip]: 各画面に対応して, ボタンの使い方などの簡単なヒントが出現する。

[Help]: マニュアルや計算原理などを紹介する予定であるが, まだ完成していない。

[Mail to COMPRO]: 自分の使用しているメールソフトが起動し, 質問や注文を筆者まで送信できる。

[Version]: 現在のCOMPROのversionを表示する。

12. 終わりに

7回に渡ってCOMPROの使用方法を解説してきましたが, 今回が最終回となります。このような貴重な機会を与えていただきましたJSA編集委員の方々には心より感謝申し上げます。COMPROはこれからも会員諸氏の意見を取り入れて改良していきますので, ぜひCOMPROをご活用ください。

13. 文献

[1] 吉原一紘, 徳高平蔵, *J. Surf. Anal.*, **20**, No2, A-68 (2013).

[2] 吉原一紘, 徳高平蔵, *J. Surf. Anal.*, **21**, 10 (2014).