《論　文》

データ解析における SPSS・PROXSCALの問題点

奥　喜正・松本安司

1．はじめに

SPSSの多次元尺度法(Multi Dimensional Scaling: MDS)は，従来はALSCALに基づくプログラムによって構成されていた[4]．昨今，PROXSCALプログラムも追加されたことで[5][13][17]，ユーザの利便性が向上したように窺える．今回，主要疾患粗死亡率データを[12]，INDSCALモデル（INdividual Difference SCALing）でシミュレーション分析することで，PROXSCALの最適性を検討する[17]．

PROXSCALは，Majorization methodを活用したガットマン変換を反復公式に利用したプログラムであるので，その最適化においてALSCALなどの従来のアルゴリズムよりも優れているようである[8]．本稿では，優劣を考察するために主要死因別粗死亡率データのリサンプリングを行ってシミュレーション分析を実行した．結果として2次元布置ではPROXSCALのほうがALSCALよりもモデル適合度が有意に優れていたが，3次元布置では逆の結果を得た．このような解釈困難な結果から，PROXSCALという非線形最適化プログラムの利用法の難しさが想定できた．

2－1．MDSを実行するプログラム

多次元尺度法MDSを実行するプログラムとしては，初期の代表的なものはKruscalのMDSCAL(Multi Dimensional SCALing)である[15]．その後，Weighted Euclidian Modelのプログラムで，内積法によるINDSCALが歴史的に有名である[1]．さらに，高根らによるALSCALが開発され，これはS-Stressをモデル適合度の指標にして最適化するもの

である．SPSSのBaseでは多次元尺度法を実行するアルゴリズムとして，ALSCALを採用している．SMACOFというMajorization Algorithm[8]を非線形最適化アルゴリズムとして採用しているPROXSCALは，SPSSのCategoryオプションに搭載されている[13]．本稿では，１相２元データを解析するモデルと２相３元データに対応するINDSCALモデルを，それぞれALSCAL及びPROXSCALにて実行する．

２－２．ALSCALプログラム

重み付きユークリッドモデルを実用化した最初のプログラムはINDSCALである[1]．ここで，モデル名INDSCALとプログラム名INDSCALが同一であることに注意する．INDSCALは対象間距離をそれらの位置ベクトルで表現して，距離をベクトルの内積に変換してから交互最小二乗法を利用して（Inner Products Approach），対象の座標(共通対象空間)と被験者固有の重み(被検者空間)を推定するものである．それに対してSPSSが搭載している多次元尺度法アルゴリズムにはALSCALが採用されている．ALSCALは，非類似性データからディスパリティ(擬似距離量)を直接計算して，二乗距離と二乗ディスパリティの剥離を小さくなることを目的にして(Squared Distances Approaches)，すなわち

$$S-\text{stress} = \sqrt{\left(\frac{1}{m}\sum_{s}\sum_{ij}\Sigma_{ij}\left(d_{ij}^{(S)2} - d^{(S)}{}_{ij}^{2}\right)^2 \Big/ \Sigma_{ij} d^{(S)}{}_{ij}^{4}\right)} \rightarrow \min$$

という，クルスカルのストレス式ではない，S-Stressというモデル適合度指標を最小にするように交互最小二乗法を利用して解を求めるものである[4]．解の精度はINDSCALのほうが，一般的にはALSCALよりも優れている[7]．

ところで，INDSCALが30年前に開発されたにも関わらず，日本で1995年前後まであまり応用されなかった事情には，INDSCALプログラムには交互最小二乗法や最急降下法など反復アルゴリズムを含むために計算量が多いので，最近のPCに搭載されている高性能CPU，メモリの拡大によって，実行時間が大幅に短縮され数十秒以内に最終解が得られるようになったことで，一般ユーザのINDSCALの利用拡大がもたらされたのかもしれない．

３－１．Majorization Algorithm

多次元尺度法の数値計算には，最急降下法が伝統的に採用されてきた．いま，最適化すべきｎ変数関数を$S(\boldsymbol{x})$とおくと

$$S(x+\Delta x)-S(x)=\sum_j \Delta x_j \frac{\partial S}{\partial x_j}+o(|\Delta x|) \qquad (\Delta x \to 0)$$

ここで，ヤコビ行列Mの負のベクトル －gradient＝－M＝（…，$-\frac{\partial S}{\partial x_j}$，…）が点$x$における局所的な最急降下方向である．最急降下法に続いて，一般的にはニュートン(ラフソン)法が数値計算法としては代表的であり，非線形最適化法の基礎をなすものである．多くの数値解法はこの方法の変種，改良型である．

つぎに，PROXSCALで利用されている Majorization Algorithm(MA) について，下記の整関数F(x)で説明する[5]．

閉区間［－1.5，2.0］を定義域とする4次関数

$$F(x)=6+3x+10x^2-2x^4 \qquad (1)$$

の最小値を与える点xを求めたいとする．このとき，関数

$$G(x,y)=6+3x+10x^2-8xy^3+6y^4$$

がF(x)のmajorizing functionで，(yを定数と見做せば) 解析が容易な関数G(x, y)をF(x)の代わりに考察することがMAの発想である．Majorizing function G(・)はG(・)≧F(・)という条件を満たすことが必要である．そこで，

$$G(x,y)-F(x)=2x^4-8xy^3+6y^4\equiv H(x,y)$$

とおいてこの条件を確認する．ここで，しばらくyを定数と考えて

$$\partial H(x,y)/\partial x=8x^3-8y^3=8(x-y)(x^2+xy+y^2)$$
$$=8(x-y)[(x+1/2y)^2+3/4y^2] \quad \cdots\cdots①$$

ゆえに，$x^2+xy+y^2>0$ for $\forall (x,y)\neq(0,0)$

また，$\partial^2 H/\partial x^2=12x^2>0$ for $\forall x\neq 0$ ……②

よって，①，②より関数Hはx＝y＝a(定数)で最小になり，かつH(x, x)＝0であるので

$$H(x,y)\geqq 0 \Leftrightarrow G(x,y)\geqq F(x)$$ (等号はx＝yのとき成立)

となる．関数Hを2変数関数と見做した場合の説明は付録を参照されたい．

そこで，関数Gでy＝a(定数) とおいて，4次関数(1)の定義域における最小点を探す代わりに，解析が容易な2次関数(2)の最小化問題を考える．

$$G(x,a)=6+3x+10x^2-8xa^3+6a^4 \qquad (2)$$

$$\partial G/\partial x=3+20x-8a^3$$

$\partial G/\partial x=0$とおいて，

$$x=1/20\ (8a^3-3) \qquad (3)$$

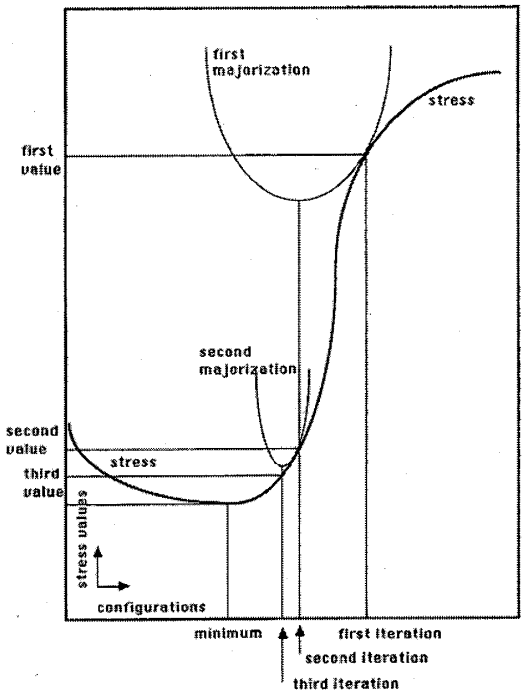
を得る．xをX(n+1)，aをX(n)と書きかえると式(3)は

$$X(n+1)=2/5X(n)-3/20 \quad (n=1,2,3,\cdots)$$

という漸化式になる．X(1)＝1.4と初期値を与えると

n	x(n)	x(n+1)
1	1.400000	0.947600
2	0.947600	0.190357
3	0.190357	-0.147241
4	-0.147241	-0.151277
5	-0.151277	-0.151385
6	-0.151385	-0.151388
7	-0.151388	-0.151388
8	-0.151388	-0.151388

のように点－0.151の近傍で関数G，すなわち，関数Fは区間［－1.5，2.0］において最小になる．この関係は下記の図からも理解できよう[8]．



Three iterations of the majorization algorithm. We have drawn a section of the stress loss function and two quadratic majorization functions. These touch the function at the current configuration, they are always above it, and their minimum provides the next configuration. 参考文献［8］.

3－2．PROXSCALプログラム

PROXSCALは，SPSSのCategoryオプションに搭載されているプログラムである．PROXSCALでは良好な布置を求めるために，モデル適合度を最小化するための最適化手法に Majorizing Algorithm（SMACOF: Scaling by MAjorizing a COmplicatd Function）を利用するので，局所最小解への収束が保証されている[8][14]．そして，PROXSCALはSMACOFを利用して，最急降下法を使用するALSCALなどのモデルよりも数値計算の観点からは最終解への収束が良好となる．

具体的には，PROXSCALはrawストレス，$\sigma_r(X)$を最小にするように最終解を求める[8]．

Xを布置として$\sigma_r(X)$は以下のように定義されて

$$\sigma_r(X) = \sum_{i<j} w_{ij}\ (\delta_{ij} - d_{ij}(X))^2$$

$$= \eta_\delta^2 + \mathrm{tr}\ X'VX - 2\mathrm{tr}\ X'B(X)X$$

$$\leqq \eta_\delta^2 + \mathrm{tr}\ X'VX - 2\mathrm{tr}\ X'B(Z)Z \equiv \tau(X, Z)$$

となる[5]．このように，$\sigma_r(X)$を最小にする代わりに$\tau(X, Z)$を最小にするというmajorization algorithm が採用されている．Vは重み行列である．$\tau(X, Z)$を行列Xで微分すると

$$\nabla\tau(X, Z) = \nabla(\mathrm{tr}\ X'VX) - \nabla(2\mathrm{tr}\ X'B(Z)Z) = 2VX - 2B(Z)Z$$

となる．ここで，$\nabla\tau(X, Z) = 0$とおくと

$$2VX - 2B(Z)Z = 0$$

$$\therefore \quad VX = B(Z)Z$$

行列Vは一般にランク落ちをするので，Vのムーアペンローズ(一般)逆行列をV^+とするとV^+は一意に定まって

$$X = V^+B(Z)Z$$

と布置Xは求まる．よって，反復式は

$$X_{k+1} = V^+B(X_k)X_k$$

となり，ガットマン変換$\Gamma(X)$となる．すなわち，

$$X_{k+1} = \Gamma(X_k)$$

$$= V^+B(X_k)X_k$$

を利用して布置Xを反復更新するのがSMACOF (Scaling by MAjorizing a COmplicated Function) というアルゴリズムである．

4．分析方法と結果

今回，2相2元データの主要疾患粗死亡率データの行データの各年度を1相とみなして，2相3元データに変換してINDSCAL で解析する．そして，ALSCALアルゴリズム，及びPROXSCALアルゴリズムでそれぞれ分析した場合の解析結果の相違検討をする．0から4までの範囲の一様乱数を発生させて，主要疾患粗死亡率データの75%を無作為に取り出してサブデータセットを作成し，このサブデータを利用したシミュレーションによって両アルゴリズムのモデル適合度を検討する．なお，主要疾患粗死亡率データを解析対象に採用したのは，当該データの有用性が文献[6]で十分に考察されているからである．そして，同様な手順を24回繰り返し（リサンプリング），24個の作成されたサ

ブデータセットに各々 PROXSCAL 及びALSCALを順序尺度レベルで適用して，２次元布置及び３次元布置を得る．そして，それぞれ得られた布置のクルスカル第一式のストレス値を記録して，両アルゴリズムの優劣を検討する[15]．

表１．ALSCALおよびPROXSCALのストレス１式の値

２次元布置			３次元布置		
NO	ALSCAL	PROXSCAL	NO	ALSCAL	PROXSCAL
乱数1	0.222	0.193	乱数1	0.102	0.138
乱数2	0.218	0.193	乱数2	0.103	0.137
乱数3	0.188	0.164	乱数3	0.1	0.122
乱数4	0.226	0.194	乱数4	0.103	0.138
乱数5	0.222	0.207	乱数5	0.104	0.139
乱数6	0.218	0.193	乱数6	0.103	0.138
乱数7	0.22	0.193	乱数7	0.103	0.138
乱数8	0.184	0.163	乱数8	0.1	0.122
乱数9	0.219	0.193	乱数9	0.103	0.138
乱数10	0.217	0.183	乱数10	0.097	0.128
乱数11	0.22	0.193	乱数11	0.103	0.137
乱数12	0.219	0.195	乱数12	0.1	0.151
乱数13	0.212	0.181	乱数13	0.1	0.132
乱数14	0.277	0.182	乱数14	0.093	0.127
乱数15	0.219	0.193	乱数15	0.103	0.138
乱数16	0.184	0.163	乱数16	0.099	0.122
乱数17	0.219	0.193	乱数17	0.103	0.138
乱数18	0.184	0.164	乱数18	0.1	0.122
乱数19	0.2	0.174	乱数19	0.102	0.126
乱数20	0.217	0.183	乱数20	0.094	0.128
乱数21	0.2	0.173	乱数21	0.102	0.126
乱数22	0.212	0.181	乱数22	0.1	0.132
乱数23	0.225	0.195	乱数23	0.102	0.154
乱数24	0.2	0.175	乱数24	0.103	0.128

表２．２方法のストレス値の差に関する検定

２次元布置　検定統計量

	PROXSCAL_2 － ALSCAL_2
Z	－4.291a
漸近有意確率(両側)	.00002

a. 正の順位に基づく
b. Wilcoxonの符号付き順位検定

３次元布置　検定統計量

	PROXSCAL_3 － ALSCAL_3
Z	－4.302a
漸近有意確率(両側)	.00002

a. 負の順位に基づく
b. Wilcoxonの符号付き順位検定

さて，ALSCALの解析結果を吟味すると２次元布置の共通空間の横軸は，「成人性疾患－感染性疾患」，縦軸は「死亡率の時代変動の大きい疾患－時代変動の小さい疾患」と解釈できる（図１）．被験者空間では，以前は「死亡率時代変動」が重要因子であったことが判り，最近の死亡構造は，「成人性―感染性」という因子で把握できることが判る．

誘導された刺激布置個人差（重み付き）ユークリッド距離モデル

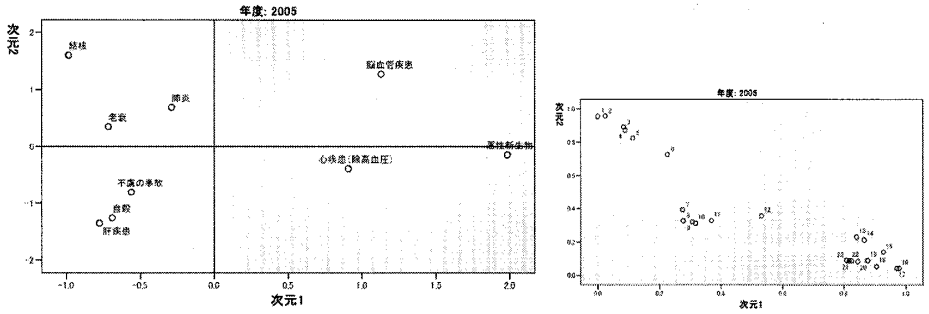


図1．ALSCALによる2次元INDSCALの布置　共通空間（左図）と被験者空間（右図）

つぎに表1，2から2個のアルゴリズムの当該データに対するモデル適合度を検討する．Wilcoxsonの検定結果から，2次元布置ではPROXSCALのほうが有意にストレス値が小さいが，3次元布置では逆の結果になった．当該データ解析結果に限って，PROXSCALがALSCALに比較して，必ずしも優れた解をもたらすとは結論できなかった．データによってはその特有のデータ構造によって必ずしもPROXSCALが優れているとは言えない場合があり，当該データ構造をさらに吟味した．

そこで，データ構造を詳細に調べるために因子分析を実行した（表3）．

表3．因子分析によるデータ構造の考察

説明された分散の合計

因子	初期の固有値			抽出後の負荷量平方和			回転後の負荷量平方和		
	合計	分散の%	累積%	合計	分散の%	累積%	合計	分散の%	累積%
1	5.720	57.196	57.196	5.584	55.843	55.843	4.472	44.722	44.722
2	2.020	20.199	77.395	1.885	18.854	74.696	2.734	27.338	72.060
3	1.123	11.226	88.621	.854	8.536	83.232	1.117	11.171	83.232
4	.591	5.913	94.534						
5	.308	3.082	97.615						
6	.148	1.482	99.097						
7	.058	.575	99.672						
8	.017	.170	99.842						
9	.010	.097	99.940						
10	.006	.060	100.000						

因子抽出法：重みなし最小二乗法

	因子		
	1	2	3
心疾患（高血圧性を除く）	.987	-.159	-.025
悪性新生物	.896	-.082	-.319
糖尿病	.844	-.260	.013
不慮の事故	-.742	.310	.031
老衰	-.712	.651	.160
脳血管疾患	-.706	-.109	.402
肺炎	.086	.971	.077
結核	-.517	.808	.193
肝疾患	.329	-.702	.516
自殺	.209	-.113	-.719

因子抽出法：重みなし最小二乗法
回転法：Kaiserの正規化を伴うバリマックス法
a. 5回転の反復で回転が収束しました．

相関行列の固有値が5.72，2.20，1.12，0.59であるので因子を３個とすると，第一因子は「成人性疾患」，第二因子は「感染性疾患」第三因子は「自殺」と解釈できるが，表３から因子が２個でも，すなわち，２次元空間で十分に当該データ情報を説明できると言える．

さらに，図２を参考に,当該データに改良を加えたデータ，すなわち脳血管疾患を除き，かつ年度を1948年から1996年度までに絞ったデータで改めてデータ解析を実行した．この改良されたデータでは３次元布置のストレス１式の値がPROXSCALでは 0.0869，ALSCALでは0.0896となって PROXSCALのほうが，モデル適合度はやや良好になった．微細なデータ構造の変動で，モデル適合度の優劣が変化してしまう．

２次元空間での解析で十分であるデータ構造を３次元空間で冗長な解析を行うと，このようなALSCALのほうのモデル適合度が良いという逆転現象が生ずるのかもしれないとも想像できる．また，両者ともストレス１式の値が0.15以下であるので，必ずしもPROXSCALのモデル適合度が劣るとも言い切れないという見解もあろう．

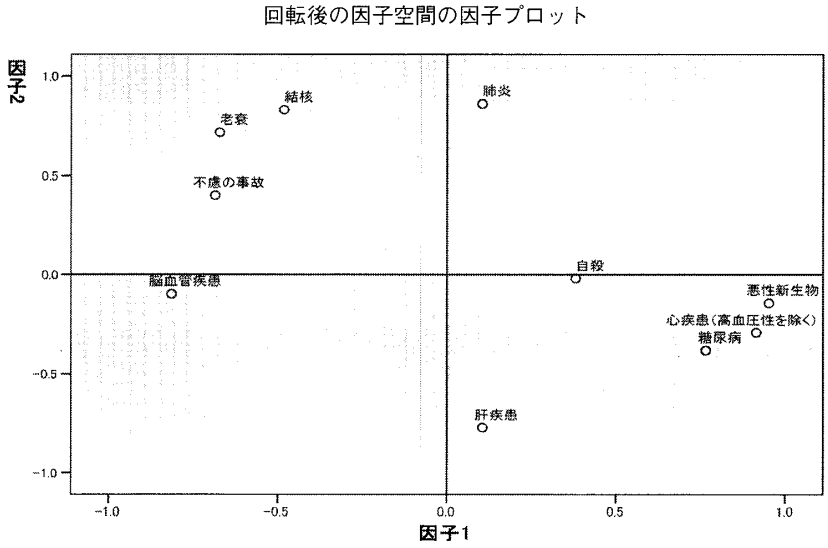


図２．因子分析による二次元疾患マップ

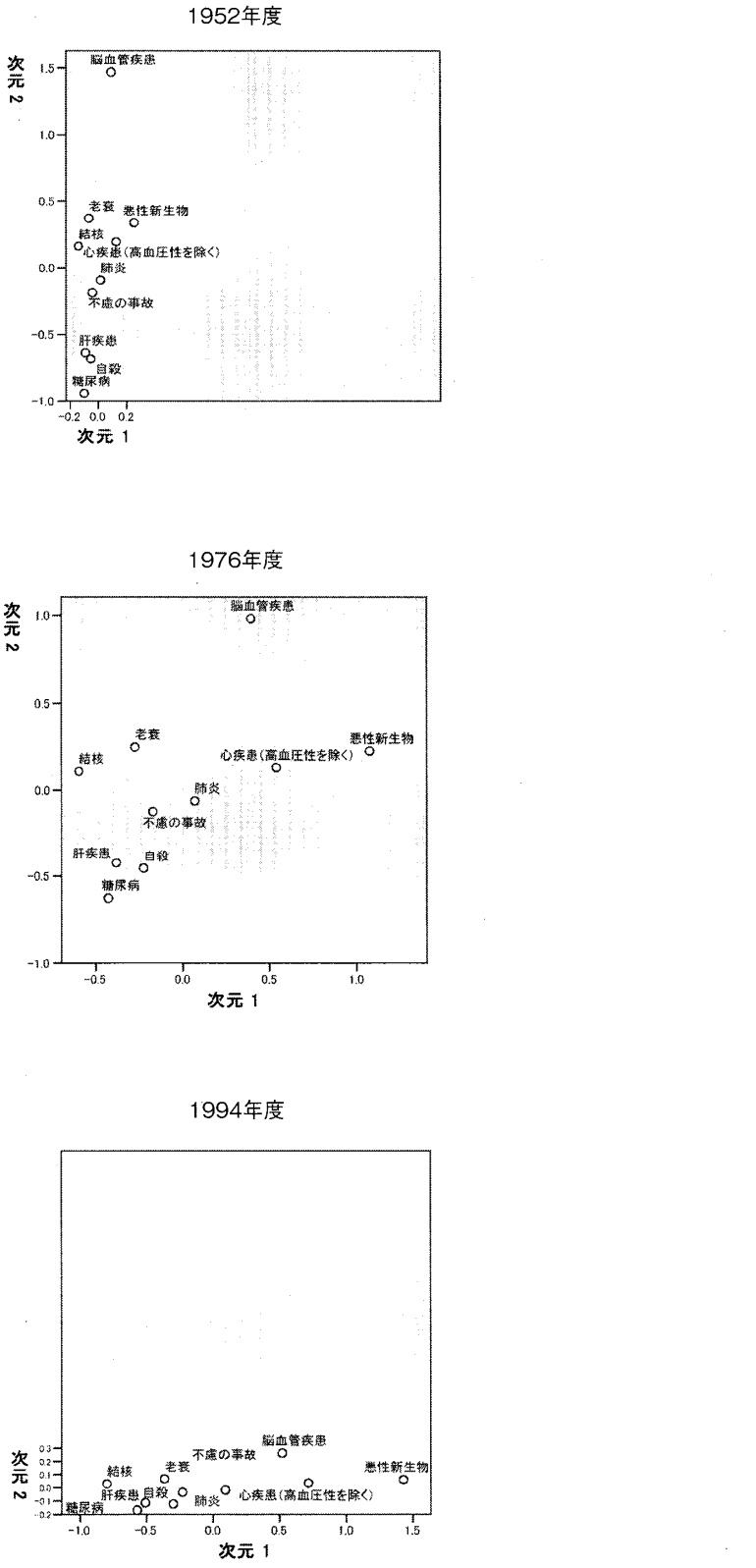


図 3 . PROXSCALによる 2 次元の個別空間 (上から1952年, 1975年, 1994年度疾患マップ)

最後に，若年女性の被験者25人がイメージするアルコール飲料に関するマーケティングデータを一元二相データと見做して，ALSCAL及びPROXSCALで解析した．このアンケートデータが集計されたプロセスと，布置の詳細な解釈については参考文献[16]を参照されたい．ストレス１式の値は，２次元布置ではALSCALでは0.125，他方，PROXSCALは0.113となり，PROXSCALのほうが適合度が優れていた．アルコール飲料の認知マップ，すなわち，製品マップを以下に示す．

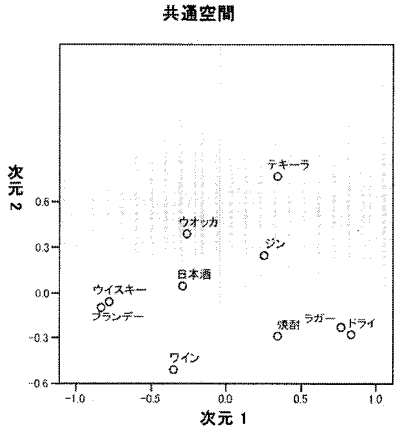


図４．PROXSCALによるアルコール飲料の２次元製品マップ

本稿では，主要疾患粗死亡率データを中心にアルコール飲料データも分析してつぎのことを考察した．つまり，PROSXCALは最適な布置を求めるために，SMACOFを利用するためにALSCALに比べて最終解への収束が一般的に良好であり，さらに，PROXSCALでは個別空間を表示されるので，認知個人差が理解し易い(図３)．PROXSCALはSMACOFとrawストレスを利用している点が他のプログラムと比較した場合に，その特徴となる．特徴を鑑みると，データによっては固有な構造によって必ずしもPROXSCALが優れた解をもたらすとは判定できない場合が生ずることが想像できるが，本稿ではシミュレーション分析からこの点を実証した．

(付録) 関数Hを２変数関数と見做すと以下のように説明できる．

$Hx = 8(x^3 - y^3)$　　$Hxx = 24x^2$　　$Hxy = -24y^2$

$Hy = 24y^2(y - x)$　　$Hyy = 36(2y^2 - xy)$

$Hx = Hy = 0$ より $x = y$ (＝aとおく)

$Hxx = 24x^2 > 0$　$\Delta = 288\{3x^2(2y^2 - xy) - (-y^2)(-2y^2)\}$

$= 288(6x^2y^2 - 3x^3y - 2y^4)$

$\Delta(x=y) = 288x^4 > 0$　よって，$x=y=a_0$で関数Hは最小になる．

参考文献

[1] Carroll, J.D. and Chang, J.J. Analysis of Indivisual Differences in Multi-Dimensional Scaling via an N-way Generalization of Eckart-Young Decompossion. *Psychometrika*, Vol 35, pp.283-319(1970)

[2] Everitt, B.S. and Rabe-Hesketh, S. The Analysis of Proximity Data. Kendall's Library of Statistics 4. Arnold.(1997)

[3] Arabie, P., Carroll, J.D. and Desarbo, W.S. Three-way Scaling and Clustering. Newbery Park, CA: Sage(1987)（邦訳：岡太・今泉（共訳）3元データの分析　共立出版（1990））

[4] Takane,Y., Young, F.W. and De Leeuw, J. Nonmetric Individual Differences Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling Features. *Psychometrika*, Vol 47,pp.7-67(1977)

[5] Borg, I and Groenen, P. Modern Multidimensional Scaling: Theory and Applications 2nd ed. Springer.(2005)

[6] Oku Yoshimasa Statistical Analysis of the Variations in the Death Structure over the Past Forty Years by Means of the MDPREF, *St. Marianna Medical Journal*, Vol.18, pp.274-281.(1990)

[7] Weinberg, S.L. and Menil, V.C. The Recovery of Structure in Linear and Ordinal Data. *Multivariate Behaviournal Reasearch*, Vol 28, pp.215-233.(1993)

[8] De Leeuw, J. Convergence of the Majorization Method for Multidimensional Scaling. *Journal of Classification*, 5, 163-180.(1988)

[9] Guttman, L. A General Nonmetric Technique for Finding the Smallest Coordinate Space for A Configuration of Points. *Psychometrika*, 36, 469-506.(1968)

[10] 中川徹・小柳義夫　最小二乗法による実験データ解析　東大出版会（1982）

[11] Kruskall, J.B. Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. *Psychometrika*, Vol 29, pp.1-27(1964)

[12] 厚生統計協会　厚生の指標：国民衛生の動向　厚生統計協会（2007）

[13] Commandeur, J.J.F & Heiser, W.J. Matematical Derivations in the Proximity Scaling (Proxscal)of Symmetric Data Matrices(Tech. Rep. No.RR 93-03)Leiden, The Netherlands: Department of Data Theory.(1993)

[14] Kier, H.A.L. Majorization as a tool for optimizing a class of matrix functions. *Psychometrika*, 55, 417-428.(1990)

[15] Kruskal, J.B. Multidimensional Scaling by Optimizing Goodness of fit to a Nonmetric Hypothesis, *Psychometrika*, 29, 1-27.(1964)

[16] 奥喜正・前鶴政和　INDSCALによる重み係数を利用した市場細分化，日本経営数学会誌，Vol28, No2, pp.61-72.（2007）

[17] Meulman, J.J., Heiser, W.J. & SPSS. SpssCategories 10.0. Chicago: SPSS.(1999)