

# ストレス1式によるSPSS・多次元尺度法の ALSCALとPROXSCALのモデル適合度比較

奥 喜 正・内 桶 誠 二

## 1. はじめに

SPSSにおける多次元尺度法 (Multi Dimensional Scaling : MDS) は、従来はALSCAL (Alternating Least squares SCALing) によるプログラムによってその計算が実行されている[1]。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。本項では、主要疾患粗死亡率データを[12]、INDSCALモデル (INDividual Difference SCALing) を中心に重みなしユークリッド距離モデルでもシミュレーション分析することで、PROXSCALアルゴリズムの最適性に検討を加える。

PROXSCALは、Majorization method (MA) を活用したガットマン変換を反復公式に利用したアルゴリズムであるので、その最適化においてALSCALなどの従来のそれよりも優れていることになっている[8]。そこで、本稿では主要死因別粗死亡率データのリサンプリングを行ったシミュレーション分析を実行し、クルスカルのストレス1式を評価基準にしてALSCALによる解とPROXSCAL解のストレス値をそれぞれ比較してPROXSCALアルゴリズムの最適性の優劣を考察する。結果として、INDSCALで解析した場合には順序尺度の変換では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たが、3次元布置では逆の結果を得た。

つぎに、死因から肺炎と脳血管疾患を除いたデータセットでも同様の適合度比較をINDSCALで行った。最後に、データ行列の階数が約2になるような人工データを作成して、重みなしユークリッド距離モデルでシミュレーション解析した [10]。以上の諸検討から、主要疾患粗死亡率データや人工データに対するALSCALとPROXSCALのモデル適合度の比較を行い、PROXSCALアルゴリズムの特性を調べた。

## 2-1. ALSICAL

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[4]. INDSCALプログラム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 (Alternating Least Squares) を利用して, 対象の座標 (共通空間) 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ではALSICALが採用されている.

ALSICAL (Alternating Least squares SCALing) は, 2元MDSモデルやINDSCAL, IDIOSCALなど3元MDSモデルを実行できる広範囲データが分析可能なプログラムである. 交互最小二乗法を利用して非類似性データ  $\delta_{ijk}$  から, ディスパリティ (擬似距離量)  $d_{ijk}^*$  を直接的に計算して二乗距離と二乗ディスパリティの剥離が小さくなるように (Squared Distances Approaches), すなわち

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

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適化するように解を求めるのがALSICALである[1].

具体的には, ALSICALでは最適尺度構成段階 (Optimal Scaling Phase) で非類似性データ  $\delta_{ijk}$  から  $d_{ijk}^*$  を推定する. そして, モデル推定段階 (Model Estimation Phase) において, まず布置行列Xと距離行列Dを固定してWを推定し, つぎにDと重み行列Wを固定してXを推定するという交互最小二乗法を使用するものである.

## 2-2. PROXSCAL

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[17]. PROXSCALでは良い布置を求めるために, モデル適合度を良好にするための最適化手法にMajorizing Algorithm (SMACOF: Scaling by MAjorizing a COmplicated Function) を利用するので, 局所最適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, 最急降下法を利用する従来のプログラム[15] よりも数値計算の観点からは局所最終解への収束が良好となるようである.

具体的には, PROXSCALはrawストレスの  $\sigma_r(X)$  を最小にするように最終解を求める[8]. Xを布置行列とすると  $\sigma_r(X)$  は以下のように定義されて

ストレス1式による SPSS・多次元尺度法の ALSCAL と PROXSCAL のモデル適合度比較

$$\begin{aligned}\sigma_r(X) &= \sum_{i < j} w_{ij} (\delta_{ij} - d_{ij}(X))^2 \\ &= \eta_\delta^2 + \text{tr } X' V X - 2 \text{tr } X' B(X) X \\ &\leq \eta_\delta^2 + \text{tr } X' V X - 2 \text{tr } X' B(Z) Z \equiv \tau(X, Z)\end{aligned}$$

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. ここで, 行列Bは  $B = [-w_{ij} \delta_{ij} / d_{ij}]$  と書かれて, Vは重み行列である.  $\tau(X, Z)$  を行列Xで微分すると

$$\begin{aligned}\nabla \tau(X, Z) &= \nabla(\text{tr } X' V X) - \nabla(2 \text{tr } X' B(Z) Z) \\ &= 2 V X - 2 B(Z) Z\end{aligned}$$

である. ここで,  $\nabla \tau(X, Z) = 0$  とおくと

$$VX = B(Z)Z$$

となり, 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  $V^+$  は一意に定まり, 上式を行列Xについて解くと

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

のように布置Xは求まる. よって, 反復式は

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

となり, これを利用して布置Xを更新する.

### 2-3. 両プログラムによるモデル適合度相違の具体例

同一データをALSCALとPROXSCALで解析した場合のモデル適合度の相違を例証するために, つぎの具体例を示す. このアンケートデータは各種アルコール飲料に対して被験者が感じるイメージを, 居酒屋とホテルバーという異なる2か所で調査し, 5点満点で得点化したアルコール飲料特性調査を目的とするアンケート用紙を使用して得られたものである[16]. そして, アルコール飲料を嗜む場所の違いでどの程度, 消費者のイメージ認知が変動するかを視覚的に把握できる知覚マップ(製品マップ)の作成を目的としている.

この製品マップの横軸は(図1)「かるい-アルコール度がきつい」, 縦軸は「男性的な-スマートな」と解釈できる. 2次元布置のモデル適合度を示すストレス1式の値は, ALSCALでは0.0496, 一方, PROXSCALでは0.0429となって, PROXSCALのほうがALSCALよりも優れた最小二乗近似布置を与えている様子がこの具体例からは推定できる.

つぎにアルコール飲料を嗜む場所の違いの影響を読み取る. バーで嗜んだ場合のビール, 日本酒の点が居酒屋の対象点から左下に移動していることから, バーでビールなど大衆酒を飲むとスマートでややアルコール度を強く感じるようである. 他方, ウイスキーや赤ワインなどはバーの対象点が右下に移動しているため, バーでウイスキーを飲んだほうがさらにスマートでアルコール度を軽く感じて飲めるようである.

## オブジェクト ポイント

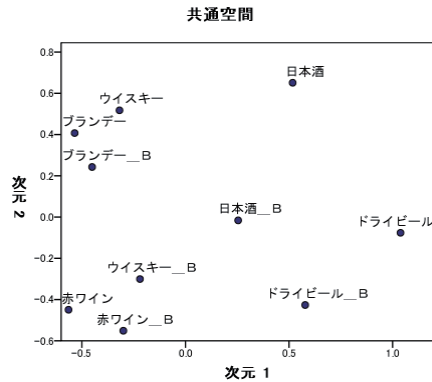


図 1. アルコール飲料の製品マップ (添え字Bはバーで嗜んだ場合を示す)

### 3. 分析方法

本稿では、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行う。現実には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約75%を無作為に取り出してサブデータセットを作成し、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の相違を検討する。また、初期値にはデフォルトのシンプレックスを採用してソフトウェアを実行する。

対象データは、主要死因の結核、悪性新生物、糖尿病、心疾患、脳血管疾患、肺炎、肝疾患、老衰、不慮の事故、自殺など10個の変数で、1940年度から2005年度（1944～1946年度はデータなし）までがケースの粗死亡率データからなる63x10型のデータ行列である。

最初に、上記のような手順を24回繰り返し（リサンプリング）、作成された24個のサブデータに対して、順序尺度レベルの変換でPROXSCAL及びALSCALを適用して、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1式のストレス値を記録して、両アルゴリズムの最適性を検討する[15]。本項ではモデル適合度の評価基準としてストレス第1式を採用した。ALSCALがS-ストレスという適合度基準を最適化するプログラムであるのに対して、PROXSCALはRawストレスという別の評価基準を最適化する。すなわち、最適化プロセスにおいて両プログラムは異なった評価基準を最適化している。そこで、双方のプログラムにより得られた布置の評価基準に本稿では、距離 $d_{ij}$ と擬似距離 $d_{ij}^*$ の差の二乗和を評価するという最小二乗基準の発想

ストレス1式による SPSS・多次元尺度法の ALSCAL と PROXSCAL のモデル適合度比較

であるストレス第1式を採用したのである。

$$\text{ストレス1式} : \sqrt{\frac{\sum_{i<j} \sum (d_{ij} - d^*_{ij})^2}{\sum_{i<j} d_{ij}^2}}$$

つぎに、脳血管疾患及び肺炎による死亡率データを除いた63x8型サブデータ行列を解析して、順序尺度の変換でPROXSCAL及びALSCALを適用して、3次元布置を得て適合度比較を行う。最後に、データ行列の階数が大よそ2になるような、すなわち、モデルによる解析結果の布置が2次元になるような63x8型人工データ行列を作成した(表0)。そして、ユークリッド距離モデルでこの人工データを解析して、ALSCALとPROXSCALによる2次元布置の適合度を比較検討する。

表0. 人工データの概要

人工データ (1 ≤ X ≤ 63を満たす整数)			
		関数型	区間
変数1	減少関数	400 - 200LOG(X)	
変数2	減少関数	$X_{N+1} = X_N - 0.5\sqrt{X}$	
変数3	減少関数	60 - 0.1X	
変数4	増加関数	70 + √X	
変数5	増加関数	0.1X <sup>3/2</sup>	
変数6	増加関数	20√X	
変数7	増加減少関数	10 + 2(X - 1)	X ≤ 31
		70 - 2(X - 31)	X ≥ 32
変数8	増加減少関数	60 + 1.2X	X ≤ 31
		96 - (X - 32) <sup>1.1</sup>	X ≥ 32
変数9	減少増加関数	300 - X <sup>1.2</sup>	X ≤ 31
		238 + (X - 31) <sup>1.2</sup>	X ≥ 32
変数10	減少増加関数	200 - X <sup>3/2</sup>	X ≤ 31
		36 + (X - 31) <sup>3/2</sup>	X ≥ 32

#### 4. 結果

粗死亡率データ行列Zから相関行列を計算して主成分分析を行い、データ行列Zのデータ構造を調べた。表1から当該データ行列の階数は2ないし3と見做せるので、当該データの多次元尺度法の解析結果は2次元布置あるいは3次元布置が妥当であると想定できる。

表1. 粗死亡率データ行列に関する主成分分析の解析結果

成分	説明された分散の合計			抽出後の負荷量平方和		
	合計	分散の%	累積%	合計	分散の%	累積%
1	5.720	57.196	57.196	5.720	57.196	57.196
2	2.020	20.199	77.395	2.020	20.199	77.395
3	1.123	11.226	88.621	1.123	11.226	88.621
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: 主成分分析

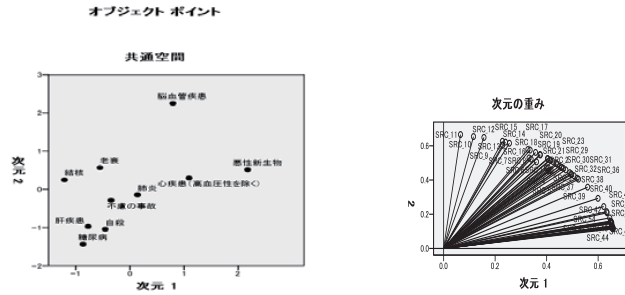


図2. PROXSCALによる2次元INDSCALの布置 共通空間（左図）と被験者空間（右図）

つぎに、2個のアルゴリズムの当該データへのモデル適合度の優劣を検討する。最初に、順序尺度での変換の場合で、ALSCALおよびPROXSCALそれぞれのアルゴリズムを使用した際の2次元布置、および3次元布置のストレス1式の値を示す（表2）。

表の値についてt-検定を実行すると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は、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。

表2. ALSCALおよびPROXSCALのストレス1式の値（順序尺度レベルの変換）

NO	2次元布置		NO	3次元布置	
	ALSCAL	PROXSCAL		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	0.227	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 3. 左図：肺炎及び脳血管疾患を除いた場合のデータ  
右図：人工データのユークリッド距離モデル解析時のストレス1式の値

3次元布置			人工データ2次元布置		
NO	PROXSCAL	ALSCAL	NO	ALSCAL	PROXSCAL
乱数1	0.0699	0.0738	乱数25	0.00508	0.00988
乱数2	0.0666	0.0729	乱数26	0.0021	0.009111
乱数3	0.0712	0.0697	乱数27	0.00171	0.00964
乱数4	0.0666	0.0729	乱数28	0.00268	0.00997
乱数5	0.0797	0.0726	乱数29	0.00244	0.0255
乱数6	0.0729	0.102	乱数30	0.00035	0.00999
乱数7	0.0729	0.102	乱数31	0.0018	0.00528
乱数8	0.0618	0.0698	乱数32	0.00173	0.00327
乱数9	0.0656	0.0701	乱数33	0.00323	0.00957
乱数10	0.0718	0.0727	乱数34	0.00201	0.00914
乱数11	0.0674	0.071	乱数35	0.00126	0.0263
乱数12	0.0728	0.0731	乱数36	0.0026	0.0071
乱数13	0.0685	0.0738	乱数37	0.00058	0.00697
乱数14	0.0584	0.0649	乱数38	0.00128	0.00643
乱数15	0.0662	0.0716	乱数39	0.00104	0.00574
乱数16	0.0768	0.0729	乱数40	0.00168	0.00952
乱数17	0.0631	0.0716	乱数41	0.0026	0.0098
乱数18	0.0614	0.0676	乱数42	0.00122	0.00928
乱数19	0.073	0.0753	乱数43	0.0098	0.0144
乱数20	0.0623	0.065	乱数44	0.00227	0.00922
乱数21	0.00683	0.0717	乱数45	0.0024	0.00948
乱数22	0.0675	0.0682	乱数46	0.003	0.00981
乱数23	0.0763	0.0754	乱数47	0.00314	0.00923
乱数24	0.0625	0.067	乱数48	0.00124	0.00949

表 4. 人工データに関する主成分分析の解析結果

成分	説明された分散の合計								
	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	5.827	58.265	58.265	5.827	58.265	58.265	5.802	58.020	58.020
2	4.060	40.597	98.862	4.060	40.597	98.862	4.084	40.842	98.862
3	.101	1.006	99.868						
4	.011	.114	99.982						
5	.001	.012	99.994						
6	.000	.004	99.998						
7	.000	.002	100.000						
8	1.357E-6	1.357E-5	100.000						
9	7.365E-7	7.365E-6	100.000						
10	6.126E-15	6.126E-14	100.000						

因子抽出法: 主成分分析

つぎに、データ行列  $\Delta$  から肺炎と脳血管疾患による死亡率データを除いたデータ行列  $\Delta_{NEW}$  を3次元かつ順序尺度変換で解析した。それは、肺炎と脳血管疾患の粗死亡率は時代的変動が単調ではなく複雑な変動パターンを示す疾患であるので、解析データの情報構造を単純化するために当該疾患群を削除した。 $\Delta_{NEW}$  はデータ行列  $\Delta$  の列空間を変化させたもので、列空間の微小変動でどの程度モデル適合度が変化するかを調べた。このとき  $p\text{-value} = 0.023$  で有意にPROXSCALのほうがストレス1式の値が小さくモデル適合度が良好といえた(表3)。2疾患除去によるデータ構造の単純化というデータ行

列の列空間微小変化によって、PROXSCALのほうがモデル適合度が良好になるという納得しやすい結果が得られたことは当該アルゴリズムの頑健性の弱さを暗示するとみなせる。また、2疾患の除去でPROXSCALのモデル適合度が改善したことからデータ行列 $\Delta$ の情報構造の特異性は、肺炎と脳血管疾患によるものとも推定できる。

さらに、人工データを2次元空間でユークリッド距離モデルを使用して解析した場合にもストレス1式の値はALSCALのほうが小さくモデル適合度良好という結果を得た(表3の右図)。以上からPROXSCALがALSCALに比べて、ストレス1式を評価基準とした場合、一般的に最適な解をもたらすとは一概に言い難いことと頑健性の弱さが本稿のシミュレーション分析から示唆された。

## 参考文献

- [1] 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)
- [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] Carroll, J. D. and Chang, J. J. Analysis of Individual Differences in Multi-Dimensional Scaling via an N-way Generalization of Eckart-Young Decomposition. *Psychometrika*, Vol 35, pp.283-319 (1970)
- [5] Borg, I and Groenen, P. Modern Multidimensional Scaling: Theory and Applications 2<sup>nd</sup> ed. Springer. (2005)
- [6] Cox, T. F. & Cox, M. AA. Multidimensional Scaling 2<sup>nd</sup>, Chapman & Hall. (2001)
- [7] Weinberg, S. L. and Menil, V. C. The Recovery of Structure in Linear and Ordinal Data. *Multivariate Behavioural Research*, 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] 奥喜正 SPSS多次元尺度法オプション ALSCALとPROXSCALのモデル適合度比較, 日本計算機統計学会第23回シンポジウム論文集, pp.131-134 (2009)
- [11] Kruskal, 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 厚生指標: 国民衛生の動向 厚生統計協会 (2007)



ストレス 1 式による SPSS・多次元尺度法の ALSICAL と PROXSCAL のモデル適合度比較

- [13] Commandeur, J. J. F & Heiser, W. J. Mathematical 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)