

<紹介と展望>

多次元尺度分析<sup>†</sup>

——マーケティングにおける計量的一手法として——

倉谷好郎\*

1. 序

Paul E. Green と、Frank J. Carmone の *Multidimensional Scaling and Related Techniques in Marketing Analysis* (Allyn & Bacon, Mass., Marketing Science Institute Series) という著書が、1970年に出版されて以来、マーケティングの計量分析に興味のある人々の間で、多次元尺度分析（以下 MDS と略称する）に対する関心が、だいぶ深くなってきたようである。そこで、MDS を利用して、若干の需要構造の計量分析を試みた筆者として、MDS をとくに OR werker から見て、どのように考えるべきか、その手法と代表的な computer codes (Kruskal MDSAL Version III, May, 1967) の紹介を兼ねて、本論稿でその概要を述べてみたいと思う。もとより MDS の専門家でない筆者が、MDS の展望といった、大それたことを試みるものでないことは、あらかじめお断わりしておきたい。

2. MDS の基本概念

J. B. Kruskal は、その論文 (2) の中で「MDS の問題は、大ざっぱに言えば、 $n$  個の客体の実験的な非相似性 (dissimilarities) を、何らかの意味でマッチさせる多次元空間での空間距離を保持する  $n$  個の点を見つけ出すことである」と述べている。その場合、 $n$  個の客体相互の非相似性をインプットとし、多次元空間での  $n$  個のポイントの空間距離をアウトプットとして考えると、表 1 のように三つの場合を考えることができる。

この場合、 $n$  個の客体の非相似性を  $n-1$  またはそれ以上の次元空間の  $n$  点として表現すると、その関係を完全に再現 (recover) することができることはいうまでもない。MDS の目的は、 $n-1$  よりはるかに少ない  $k$  次元 ( $k < n$ ) で、できる限り  $n$  個の客体間の非相似性のすべてのペアごとの関係を忠実に表現することである。換言すれば、客体間の非相似性をはかるいくつかのインプリシットな  $k$  個の尺度を探り出すことである。仮に、 $n$  個の客体間の非相似性を決める主体 (subject) が、エクザクトに  $k$  個の尺度または基準を用いているとすれば、 $k$  次元における  $n$

† 1972年11月20日受理。1972年5月28日、春季研究発表会発表要旨。

\* Case Western Reserve University.

表1 MDS の分類

	インプット	アウトプット
Fully Metric MDS	METRIC (Interval Scale)	METRIC (Interval Scale)
Non-Metric MDS	NON METRIC (Ordinal)	METRIC (Interval Scale)
Fully Non-Metric MDS	NON METRIC (Ordinal)	NON METRIC (Ordinal)

点で、 $n$  個の非相似関係が完全に表現できることになる。したがって、仮に  $k$  次元の  $n$  点で完全に表現できぬとすれば、それは、主体の非相似性についての計量エラーによるものであるということとなる。

さて、前にかえて、MDS には、インプットとアウトプットが、METRIC かまたは NON-METRIC かによって fully metric, non-metric および fully non-metric の 3 種類あることを述べた。marketing における計量的接近の立場からは、non-metric MDS、すなわちインプットが、対象商品の非相似性という心理的変量であるところから、その関係はせいぜい ranking、すなわち ordinal scale でとらえられると考えるのが順当であり、これを  $k$  尺度による metric、すなわち interval scale で再現して、 $k$  尺度が具体的に何であるかを、統計的ではなく実体的に推理していくことが MDS の内容となる。対象商品の非相似性を計る尺度がその商品の物理的特性のような場合、すなわち metric な量としてとらえられる場合には、fully metric MDS も考えられるが、marketing への MDS の応用では、このような属性空間より消費者のレスポンス空間の分析が、より重要であるので、属性空間の分析は、レスポンス空間を探索するための有力な手がかりを提供するものとして取り扱われている。ここで、非相似性の再現 (recovery) が問題となったが、non-metric MDS の場合は、input が ordinal であるので、単調関係 (monotone relationship) の再現が要求されることになる。次に、 $n$  個の客体の非相似関係を、 $k$  次元の  $n$  個の点として recover した場合に、その recovery がどの程度行なわれたかを示す尺度が必要となる。そこで、Kruscal は、統計学における Chi Squares の goodness of fit の概念にならって、stress という統計量を提案している。

さて、 $n$  個の客体の非相似性を、各ペアごとのランキングで表現すると、 $n(n-1)/2$  個のランクで表現できることとなる。すなわち  $n \times n$  のマトリックスの対角エレメントを除いた half matrix ということになる。いま仮に説明を簡単にするために、 $n=4$  として次頁の dissimilarities matrix を実験的に得たとしよう (表2)。

$n=4$  であるから、この関係を回復する次元  $k=3$  とすれば、この非相似性関係を完全に recover できることは明らかであろう。いまそこで、 $k=2$  として、後述の Kruscal measure による最善の recovery をして、次の関係を得たと仮定する (表3)。

以上の表2と表3から、図1が得られる。

表 2 Dissimilarities matrix—Experimental Data  $\delta$  (input, ordinal scale)

Object	Object	A	B	C	D
A			3	1	6
B				2	5
C					4
D					

\* このマトリックスのランクによれば、客体のペア A, C が一番相似しており、客体のペア A, D が一番非相似であることになる。

表 3 Dissimilarities matrix\* ( $k=2$ , output... $d$  (interval scale))

	A	B	C	D
A		2.1	1.0	5.3
B			3.0	6.5
C				4.7
D				

\* この数字は単に説明の便宜のために人為的につけられたもので、MDS を適用して計算された結果ではない。

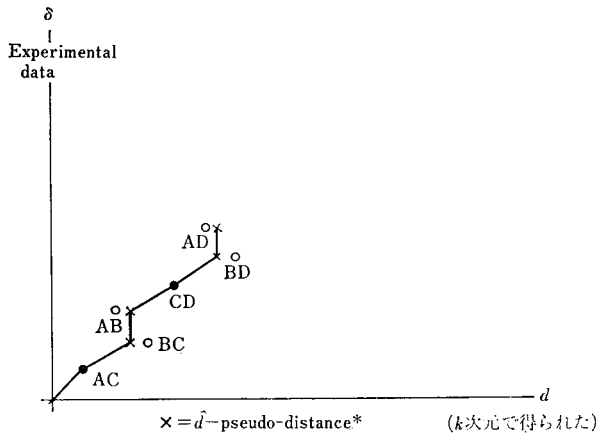


図 1 Measure of goodness of fit

\* pseudo distance というのは、単調増加関係を維持するため各客体相互間の距離のうち必要なものを修正したため、いわば似非距離で、この数値では客体相互間の距離が非斉合となるためこの名を使用したものである。

図で示した折線は、単調増加関係を保存するという条件下で、横方向へのエラーの二乗を極小化するように最小二乗法で計算された、回帰線の結合である。input である experimental data にエラーがないと仮定すれば、エラーは  $d$  方向にあると考えるのが妥当で、最小二乗法は水平方向に適用されたと考えるべきであろう。

Kruscal の stress は、上述の  $d$  と  $\hat{d}$  から次のごとく定義される。

$$S = \left[ \frac{\sum_{i < j}^n (d_{ij} - \hat{d}_{ij})^2}{\sum_{i < j} d_{ij}^2} \right]^{1/2}$$

stress S は零から 1 までの値をとることになり、stress が大であるほど fit が悪く、また元の非相似関係の recovery が不十分であることを示すことになる。Kruscal は stress の値から、recovery の状況を判定する基準として、次の評価を示している。

Stress	Goodness of Fit
20%	不良
10%	可
5%	良好
2.5%	優秀

0% 完全

この基準は客体の数  $n$  と独立に与えられているが、 $n$  の数が比較的大である場合は、評価基準を若干甘くすることも必要となるであろう。一応の基準としては妥当なものと思われる。

### 3. Non-Metric MDS の応用事例

non-metric MDS の応用事例として、まず、Green と Carmone が前掲の著書にあげた、1968 年製の 11 モデルの乗用自動車の例を借りることとし、第二の事例として、筆者がアメリカの大手のグリーンティングカード会社から委託されて行なった MDS の事例を示すことにする。

#### 3.1 乗用自動車の事例

顧客が乗用車に対して、どのようなイメージを持ち、そのイメージの背後にどのような要因が働いているかを知ることが、乗用車メーカーにとってきわめて興味があり、また販売戦略を立てる上で重要であることはいうまでもない。物理的な特性に対し、顧客が、購入という形でいかに反応するかは、物理的特性が計量的に表現できる限り、多変量回帰分析でこれをとらえることができる。もちろんこのような回帰分析による、因果構造を解明することは、マーケティングの上できわめて重要である。しかしさらにこの上に心理変量としての、商品イメージの実態が MDS その他の方法によって把握できれば、製品計画、販売計画の上で、きわめて重要な参考資料となるであろう。そこで、Green, Maheswari と Rao[1] は、1968 年型の下記の 11 種のモデルについて MDS の手法を用いて解析を試みた。

#### 1968 年型乗用車モデル

- 1 Ford Mustang 6
- 2 Mercury Cougar V 8
- 3 Lincoln Continental V 8
- 4 Ford Thunderbird V 8
- 5 Ford Falcon 6
- 6 Chrysler Imperial V 8
- 7 Jaguar Sedan
- 8 AMC Javelin V 8
- 9 Plymouth Barracuda V 8
- 10 Buick Le Sabre V 8
- 11 Chevrolet Corvair

以上の 11 のモデルについて、ある subject について実験した結果、ペアごとの非相似性によって、ランクづけを行なうと、 $11(11-1)/2=55$  のランクづけが可能となり、次頁の非相似性マトリックスが得られた。

この非相似性マトリックスとインプットとして二次元における scaling を行なった結果、次の scatter が得られた (図 2)。

モデル モデル	1	2	3	4	5	6	7	8	9	10	11
1		8	50	31	12	48	36	2	5	39	10
2			38	9	33	37	22	6	4	14	32
3				11	55	1	23	46	41	17	52
4					44	13	16	19	25	18	42
5						54	53	30	28	45	7
6							26	47	40	24	51
7								29	35	34	49
8									3	27	15
9										20	21
10											43
11											

上記の scatter からまず横軸を見ると、右端に Lincoln Continental (高級車) と Chrysler Imperial (高級車) があり、左端に Chevrolet Corvair (小型大衆車) と Ford Falcon (小型大衆車) があることから、横軸が奢侈性をあらわしているものと考えられる。また Jaguar (高級スポーツ車) が比較的上位に、Buick La Sabre V8 (中級大型車) が比較的下位にあることから、縦軸はスポーツ性をあらわしているものと考えられ、ほぼ二次元で乗用車のイメージ形成の根底にこの二つの要因があることが推定せられる。次元を増すことによってインプットデータの回復度が増すとはいえ、軸の解釈が困難となり、かえって良くない結果を生むこととなる場合が多い。この点たとえば、主成分分析がデータの相関性を利用して情報の集約化をねらうに対し、MDS はデータの蔭にかくれた要因を引き出そうとする点においてその目的は異なる。ついでながら、ここで主成分分析と MDS を対比して見ると右表のごときものとなる。

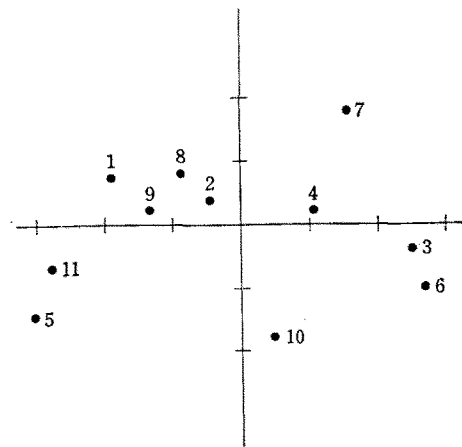


図 2 1968 年型乗用車 11 モデルの MDS によるアウトプット (k=二次元)

主成分分析と MDS との比較

	MDS	主成分分析
目的	Implicit な判断基準を探索発見	情報の集約化
インプット	metric または non-metric	metric
インプット相互の関係	monotone	linear
アウトプット	metric, 各 object の各軸に対する座標	metric 成分(合成変量)=インプット・データの一次結合

3.2 グリーティング・カードの事例

筆者は、1970 年末から 71 年春に至る約 6 カ月間に、米国中西部に本社を有するグリーティン

グ・カードの大手A社の研究委託を受けて、A社が1970年初頭から新規に発売を開始した新しいモードのグリーティング・カード36デザインの需要構造の解析を行なった。この新モードのカードは当時全国的に爆発的売行を示し、さらに販売高を伸ばすため新しく36デザインを追加するための基礎資料を提供してほしいとの要請があった。提供されたデータは、既発売の36のカードにつき、アメリカ全土に散在する13のoutletの約1カ月にわたる販売枚数と、カードのそれぞれの特性であった。特性として与えられたものは、(1)カードの形状(四角または卵型)、(2)カードのおもな色調(青、オリーブ、赤、紫、金色、茶)、(3)カードに現われている人間の数(多数、二人、一人、ゼロ)、(4)カードに書かれた文言の長短(短、中、長)、その他にも、たとえば文言が友情を表現するものか、または恋愛の感情を表現したものか等、他の特性もあったが、いずれも販売量に対して有意な変数ではなかったのでこれを分析の過程で除去して行った。筆者はまず、各outletにつき、それぞれダミー変数による販売関数を回帰式で推定し、それぞれの地域特性と、カード特性を表わす変量と販売量の関係を解明し、これによって、研究目的をほぼ達成したが、さらに洞察を深めるためMDSの手法を適用することに踏み切ったのである。MDSを適用する場合まず必要なインプットデータは、この場合objectとしてのカードの相似性を表わす数値である。この場合、実験者を選定して36のカードにつき相似性を計量させることも考えたが、それよりも13outletsにおける販売量ベクトルの相関度をもって相似性のmeasureとすることにした。計算はケース工大に所属する計算センター Chi Corp. のUNIVAC 1108を用い、programはKruscalのMDSCAL-Version IIIを用いた。

scalingの次元は、カード特性で有意なものが4変数であったので、基本的には四次元以下を目標とし、最高scaling次元を六次元とした。さらに、このプログラムのアウトプットであるfinal configurationを、BMDの因子分析プログラムに含まれたVARIMAX法によって軸を回転させて、各軸のidentificationに努めることとした。その結果を要約すると次のごとくである。

1) ストレスは、六次元で0.039、五次元で0.048で良好であったが、四次元では0.071、三次元では0.096としたいに悪くなり、二次元で0.169、一次元で0.273の結果が出て、Kruscalの標準に従うと、二次元以下は使用不可能となり、三次元以上を使わねばならなくなった。

2) 三次元以上の結果について、さまざまな角度から、各軸が何を表現しているか、懸命の努力を傾注し、またグリーティング・カードの専門家の協力も得たが、軸のidentificationには完全に失敗した。

3) 失敗の原因は、まず第一に、相似性の尺度に、各カードの13カ店における販売量の相関係数を用いたが、この尺度を採用したことの妥当性が疑わしい。第二に、objectとして36カード全部を用いたが、これは単に計算機による計算時間を長くしただけではなく、軸の解明をより複雑困難にした。この場合、36カードのうち先験的判断による類型の代表と見られるカード、10デザイン程度に絞るべきであったと思う。

もちろん今回のMDS応用の失敗は、経験の乏しい筆者の、リサーチデザインの失敗と、何と

いっても、グリーティング・カードに関する基本的な知識の欠如に求むべきであろう。一つには、グリーティング・カードの専門家も、MDS のような計量的接近に必ずしも信頼を置いていなかったという事実と、すでに多変量回帰分析によって、研究委託者である企業側から十分満足をかち得ていたため、MDS の応用には関係者の間に真剣味がたらなかったということもある。

#### 4. MDS のためのコンピュータ・プログラム

MDS のためのコンピュータ・プログラムについては、前述の Green と Carmone の著書の付録に要領よくまとめられているので、ここではもっぱらその中で代表的と考えられる Kruscal の MDSCAL-Version III について詳述してみることにする。

##### 4.1 MDSCAL-Version III

MDSCAL-Version III は、1964 年、ベル研究所の J. B. Kruscal によって書かれたもので、言語は FORTRAN IV が用いられ、IBM 以外のシステムにも、最小の修正によって使いうるよう配慮したきわめて汎用性の高いものである。さらに Kruscal は、1968 年に、MDSCAL-Version IV を発表しているが、これは Version III に若干の option を付加したもので、本質的に変わったところはなく、一般の用途には Version III で十分と考えられるので、Version IV はここではとり上げない。

MDSCAL-Version III は、下記の七つの routines から成り立っている。すなわち

1) MDSCAL 外生的に与えられた  $n$  points の初期 configuration より、steepest descent 法により、くり返し計算を行なうルーチンである。すなわち stress(S) を、 $t$  次元の  $n$  点、すなわち  $tn$  変数の関数として表現し、その各変数について、偏導関数を求め、ステップサイズを乗じて、しだいに  $n$  点を移動させ、実験主体のとらえた  $n$  点間の非相似性をできる限り  $t$  空間で忠実に再現しようとするものである。数学的には GRADIENT は

$$S=f(X_{11}, \dots, X_{1n}, \dots, X_{t1}, \dots, X_{tn})$$

$$g = \left[ -\frac{\partial S}{\partial x_{11}}, \dots, -\frac{\partial S}{\partial x_{tn}} \right]$$

で表現されることになる。なお、ステップサイズは次の数式の [ ] の表現によって求められる。

$$x'_{ts} = x_{ts} + \left[ \frac{\alpha}{\text{mag}(g)} \right] g_{ts}$$

まず  $\alpha$  の求め方であるが、 $\alpha$  の初期値は、Kruscal によれば、0.2 かその近傍の値がよいとされている。 $\alpha$  の数値は iteration のつど変化していくが、その変化は次式によって規定される。

$$\alpha_{\tau+1} = \alpha_{\tau} \text{ (angle factor) (relaxation factor) (good luck factor)}$$

angle factor は、 $\tau+1$  回目の iteration によって得られた gradient  $g_{ts}$  と、 $\tau$  回目に得られた  $g_{ts}''$  との間の角度  $\theta$  の cosine の関数として表現される。すなわち

$$\text{angle factor} = 4.0^{(\cos \theta)}$$

where

$$\cos \theta = r_{g_{is}g_{is}''} = \frac{\sum_{is} g_{is}g_{is}''}{\sqrt{\sum_{is} g_{is}^2} \sqrt{\sum_{is} g_{is}''^2}}$$

次に relaxation factor は次式によって求められる。

$$\text{Relaxation factor} = \frac{1.3}{1 + (5 \text{ step ratio})^{5.0}}$$

$$5 \text{ step ratio} = \min \left[ 1, \frac{\text{present stress}}{\text{stress 5 iterations ago}} \right]$$

$$\text{Good luck factor} = \min \left[ 1, \frac{\text{present stress}}{\text{previous stress}} \right]$$

次に mag(g) は gradient の相対的な大きさを示す変量で、次の表現によって示される。

$$\text{mag}(g) = \frac{\sqrt{\sum_{is} g_{is}^2}}{\sqrt{\sum_{is} x_{is}^2}}$$

ただし  $x_{is}$  は、標準化が行なわれて  $\sum_{is} x_{is}^2 = 1$  とされているから、結局

$$\text{mag}(g) = \sqrt{\sum_{is} g_{is}^2}$$

となる。

以上数式で示されたステップサイズは heuristic に求められたもので、必ずしも最適サイズとはいわれないが、実験の結果満足な結果が得られたとされている。以上の方法によって、所与の  $k$  次元でくり返し計算の 1 ラウンドが修了すると、そのつど stress が計算せられ、その値が収斂したところで計算が終了し、その結果がアウトプットされることになる。ただし収斂しない場合でも、計算時間の浪費を防ぐため、くり返し計算のラウンド数が一定回数に達すると計算を終了させることもできる。収斂したストレスの値がはたして global optimum かあるいは local optimum か疑問がある場合には、代替的な初期 configuration から、同じプロセスをくり返して、同一のアウトプットに帰着するか否かを確かめればよい。この問題はすべての gradient 法のもつ共通の問題である。アウトプットとしては、くり返し計算ごとにストレス、ステップサイズ・ストレスレシオ、その他、くり返し計算の進行状況を示す数値が出力され、くり返し計算が終了したところで、 $k$  次元における  $n$  点を示す final configuration が、出力される。さらに、その configuration が、

1. 軸の転換 (rotation)
2. 原点の移動 (translation)
3. 軸の一樣な伸縮 (uniform stretch or contraction)
4. 対象投影 (reflection)

にも不変 (invariant) であることから、二次処理のための configuration データのカードパンチも可能となっている。このようにして  $k$  次元での scaling が終わると、 $k-1$  次元での scaling が始まり、最終的には一次元に至るまで scaling が継続される。そして最後に図 3 のようなストレスを次元の関数としてグラフが打ち出されて計算は終了する。



図3の曲線のパターンによっても、 $k$ 次元の $k$ の値をどの数値に決めるか一つの手掛りを得ることができる。

2) FIT このルーチンは、最小二乗法により  $\hat{d}_{ij}$  (pseudo-distance) を算出するためのもので、この際、必要に応じて加重 (weighted) 回帰を用いることもできる。回帰は単調関係を維持することを条件として行なわれる (前掲図1参照) ので、 $\delta_{ij}$  の sorting が  $d_{ij}$  の値の順序に行なわれなくてはならず、多くの sorting のサブルーチンが使われることになる。

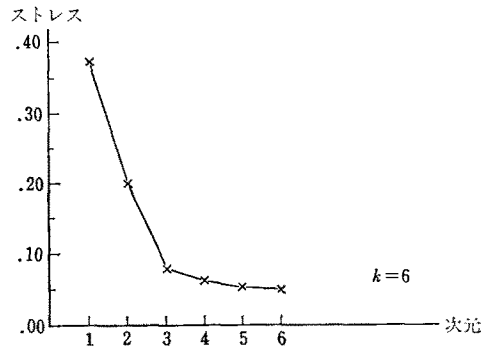


図3 ストレス推移の一例

3) NEWSTP このルーチンは、前述の数式によって示された angle factor  $\alpha$  と  $\text{mag}(g)$  を計算してステップサイズを算出するルーチンである。サブルーチンとして、MDSCAL から call されることになっている。

4) CCACT このルーチンは、プログラムコントロール・カードを読み込んで、カードに与えられた数値によってそれぞれのパラメータの値を決めていくサブルーチンである。コントロール・カードでは、scaling の最大次元  $k$  の値や、Gradient 法によるくり返し計算度数の極大値や、また読み込まれるデータが非相似性マトリックスか、相似性マトリックスか等を指示することになっている。

5) RFUNCT MDS においては、通常 Euclidian distance を、distance measure  $d_{ij}$  に使用するが、このルーチンでは、一般的な Minkowski, distance を使うことも可能となっている。Minkowski's distance は次式によって表現される。

$$d_{ij} = \left[ \sum_{s=1}^t |x_{is} - x_{js}|^r \right]^{1/r}$$

6) SORT このルーチンは、rank order を行なうためのサブルーチンで、簡単ではあるが、速度が遅い。したがって CPU で、高速の SORTING が可能なライブラリープログラムを使用できる場合は、その sorting を使うほうが効率的であろう。

7) WTRAN このルーチンは、 $n$  個の客体に不均等の重みづけを与えようとする場合に、その重みを決定する関数を外部から与えられるようにするためのダミーサブルーチンである。この関数そのものは、プログラムの使用者によって作成されねばならない。

MDSCAL-Version III が処理できる最大の客体 (object) 数は 75 であり、最大の scaling の次元は 10 次元までとなっている。この数は MDS の実用的見地から考えてきわめて十分と考えられる。

## 5. MDS の課題

まず第一の問題は、MDS によって求められた final configuration の各軸が、現実には、いかなる基準を計量しているかということである。この問題は、ひとり MDS だけでなく主成分分析における成分や、因子分析における因子の実体が、何を表現しているのかという問題と軌を一にしている。いわゆる identification の問題で、統計数学的な問題ではなく、subject matter の問題である。したがって対象となっている主題に関する知識が十分でなければ、この軸の labeling は不可能である。Green と Carmone は前掲の著書で、この問題の解決に若干の示唆を与えているので、そのいくつかについて述べてみたい。

1) 客体 (objects) には通常 physical な特性があるので、この特性との連関において、軸の実体を探索する。すなわち物理的特性空間と心理的反応空間との関連を追求することによって、反応空間の軸の実体を究める方法で、MDS を特性空間と反応空間の両方に適用して、反応空間の軸を両者の相関 (canonical correlation) が極大化するように回転させてみるのもおもしろい方法であろう。

2) すでに基準となっている軸の実体がわかっている objects をいく組か投入して、他の objects に対する基準点として考える方法。

3) MDS で得られた final configuration は相似性変換 (similarity transformation) についてユニークであるので、軸の変換を数個試みることによって軸の基準を探索する方法。普通、因子分析における VARIMAX 法による軸の回転が軸の labeling を容易にするといわれ、広く使われている。

以上は、軸の identification の問題であるが、その他、実験主体 (subjects) が複数個ある場合のデータの aggregation の問題や、またインプットデータに誤りがある場合の stress の measure や fitting に横方向の最小二乗法を使う代わりに直交最小二乗法 (orthogonal least squares) を使う問題等、統計的に興味ある問題も今後研究を要する課題であろう。MDS のマーケティング面における応用は歴史も浅く、その有用性が確立したとはとうていいいえないが、その潜在的有用性は否定することができないので、一つの計量的ツールとしてマーケティング・アナリストが MDS に関する知識を持っておくことは必要であろう。

### 参考文献

- [1] Green, Paul E. and Frank J. Carmone, *Multidimensional Scaling and Related Techniques in Marketing Analysis*, Allyn and Bacon, Boston, 1970.
- [2] Kruscal, J. B., "Multidimensional Scaling by Optimizing Goodness of Fit to a Non-metric Hypothesis," *Psychometrika*, **29**, 1 (March 1964).
- [3] Kruscal, J. B., "Nonmetric Multidimensional Scaling : A Numerical method," *Psychometrika*, **29**, 2 (June 1964).
- [4] Kruscal, J. B., How to use MDSCAL, A Multidimensional Scaling Program (Version III, May 1967, all in FORTRAN IV) (mimeographed).