解　説

調査データの多次元解析†

杉　山　明　子††

1. はじめに

世論調査のデータは，いわゆる測定データとは趣きを異にし，量的データ（定量的変数）より，質的データ（定性的属性）が中心になっている．世論調査では，“テレビを何時間ぐらいみているか”，“月収はいくらか”など数量で回答の得られる質問は限られており，むしろ，次に示すような，いくつかの選択肢の中から回答を選ぶ形式の質問が多い．

問……宗教とか信仰とかに関係すると思われることがらで，あなたが行っているものがありますか，ありましたら，リストの中からいくつでもあげてください．

回答……

A　ふだんから礼拝，お勤め，修業，布教など宗教的な行いをしている．またはおりにふれ，お祈りやお勤めをしている．

B　年に1，2回程度は墓参りをしている

C　聖書・経典など宗教関係の本を，おりにふれ読んでいる

D　この1，2年の間に，身の安全や商売繁盛，入試合格などを，祈願しに行ったことがある

E　お守りやおふだなど，魔除けや縁起ものを自分の身のまわりにおいている

F　この1，2年間に，おみくじを引いたり，易や占いをしてもらったことがある

この種の質問には，列挙したいくつかの回答選択肢のなかから，1つだけの回答を求める単一選択(Single Answer)形式と，該当するものをいくつでもあげさせる多肢選択（Multiple Answer）形式とがある．いずれの場合も，回答の記号A，B，……は，単なる符号であり，序列・数量などを意味していない質的データである．

世論調査の分析では，このような質的データの相互関係の追求が中心課題となる．質問ごとの単純集計，属性と質問・選択肢間のクロス集計，質問・選択肢相互間のクロス集計，あるいは3つの質問・選択肢間の三重クロス集計などを行う．さらに多くの質問間の関係をみるには，三重クロス集計では不十分になる．しかし，四重・五重の多重クロスをすると，たとえ計算したとしてもその結果を読み取るのは困難になる．たとえばさきの宗教の質問でみると，A～Fの選択肢相互の関係をみようとすると，AとB，AとC，……，EとF，AとBとC，……といろいろの組み合わせの多重集計をしなければならず，また，その多重集計を全部行ったとしても，とても繁雑で，全体の様子を俯瞰するのは難しい．そこに多次元解析の手法が威力を発揮するのである．

さて，多次元解析と一口で言っても，種々のモデルがあり，適用の条件，結果利用の方法などが異なるので，次章で詳述することにしよう．

2. 分析の方法

多次元解析の分析方法は，外的基準のある場合と，外的基準のない場合とに大きく分けられる．

外的基準のある場合の分析手法（表-1）は，問題が量の推定か，質の分類かで，さらに2つの種類に分かれる．「量の推定」とは，テレビ視聴時間量，視聴率，収入のように，定量的変数として表現できる既知のデータを外的基準とし，年令，制作費などの定量的変数，

表-1　外的基準のある場合の分析手法[1]

問　題	外的基準	説明要因	分析手法
量の推定	定量的変数(量的)	定量的変数	重線形回帰分析
		定性的属性	数量化理論　第Ⅰ類
質の分類	定性的属性(質的)	定量的変数	判別函数
		定性的属性	数量化理論　第Ⅱ類

† Multi-dimensional Analysis on Survey Data by Meiko SUGIYAMA (Public Opinion Research Institute of Japan Broadcasting Corporation).

†† NHK 放送世論調査所

あるいは性別，番組種目などの定性的属性を説明要因とした場合に，説明要因から外的基準の推定・予測をする問題である．説明要因が定量的変数の場合は，重線形回帰分析を用い，説明要因が定性的属性の場合には，数量化理論第Ⅰ類を用いる．

〈重線形回帰分析〉

定量的変数 $Y(i)$ を外的基準とし，その説明要因として定量的変数 $X_j(i)$ が得られたとき，

$$Y'(i)=\sum_{j=1}^{m} a_j X_j(i)+c$$

とおき，

$$E=\sum_{i=1}^{n}(Y(i)-Y'(i))^2$$

が最小になるよう a_j, c を求める．

ここで，i は調査相手番号 $(i=1,2,\cdots,n$人$)$

j は要因番号 $(j=1,2,\cdots,m$アイテム$)$

〈数量化理論第Ⅰ類〉

定量的変数 $Y(i)$ を外的基準とし，その説明要因として定性的属性への回答 $\delta_{jk}(i)$ が得られたとき，

$$Y'(i)=\sum_{j=1}^{m}\sum_{k=1}^{k_j}\delta_{jk}(i)X_{jk}+c$$

とおき，

$$E=\sum_{i=1}^{n}(Y(i)-Y'(i))^2$$

が最小になるようカテゴリスコア X_{jk} と c を求める．

ここで，i は調査相手番号 $(i=1,2,\cdots,n$人$)$

j は要因番号 $(j=1,2,\cdots,m$アイテム$)$

k は選択肢番号 $(k=1,2,\cdots,k_j$カテゴリ$)$

$\delta_{jk}(i)$ は i 番目の調査相手の j アイテムの k カテゴリへの回答を示す．

$$\begin{cases}\delta_{jk}(i)=1\cdots\cdots\text{回答したとき}\\ \delta_{jk}(i)=0\cdots\cdots\text{回答しないとき}\end{cases}$$

つぎに，表-1 の「質の分類」とは，「保守一革新」「老年一中年一若年」のように，2個～数個の分類として与えられる既知の定性的属性を外的基準とし，定量的変数あるいは定性的属性を説明要因とした場合に，説明要因から外的基準の分類を行う問題である．ここでも，説明要因が定量的変数の場合と，定性的属性の場合とでは，分析手法が異なり，前者は判別函数を，後者は数量化理論第Ⅱ類を用いる．

〈判別函数〉

t 個の分類で示される定性的属性 $T(i)$ を外的基準とし，その説明要因として定量的変数 $X_j(i)$ が得られたとき，

$$\alpha(i)=\sum_{j=1}^{m} a_j X_j(i)$$

とおいて，α の T への相関比

$\eta^2=\sigma_b^2/\sigma^2$

が最大になるよう a_j を求める．

ここで，i は調査相手番号 $(i=1,\cdots,n$人$)$

j は要因番号 $(j=1,\cdots,m$アイテム$)$

σ^2 は $\alpha(i)$ の全分数

σ_b^2 は $\alpha(i)$ の t 個の分類の級間分散

〈数量化理論　第Ⅱ類〉

t 個の分類で示される定性的属性 $T(i)$ を外的基準とし，その説明要因として定性的属性への回答 $\delta_{jk}(i)$ が得られたとき，

$$\alpha(i)=\sum_{j=1}^{m}\sum_{k=1}^{k_j}\delta_{jk}(i)X_{jk}$$

とおいて，α の T への相関比

$\eta^2=\sigma_b^2/\sigma^2$

が最大になるよう X_{jk} を求める

(ここでの記号は前と同じ)

さきに述べたように，世論調査データは，いくつかの選択肢の中から回答をえらぶといった定性的属性が中心であるので，重線形回帰分析や判別函数より数量化理論第Ⅰ類，第Ⅱ類を用いることが多い．

外的基準のない場合の分析手法（表-2）は，i 要因

表-2　外的基準のない場合の分析手法[7]

要因相互の関係 R_{ij} の性質	分析手法
漠然とした親近性（数量表現だがメトリカルとはいえない）	数量化理論第Ⅳ類（e_{ij}-型）
大小関係（一対比較）	林の一対比較に基づく空間配置
ランクオーダ（N 個のものすべての順序がきまる）	ノンメトリックな方法→Coombs の方法
ランクオーダ（2個のもの同志の関係のランクオーダ）	Shepard の方法 Kruskal の方法 Guttman の SSA Young, de Leeuw, Takane の方法
ランクオーダのついた群分け	林の MDA-OR
ランクオーダのない単なる群分け	林の MDA-UO
親近性，非親近性（メトリカルな場合）	Togerson の方法 K-L 型数量化
頻度	潜在構造分析 数量化理論第Ⅲ類（パタン分類） (MSA, POSA を含む)
相関係数	成分分析法，因子分析法

と j 要因の関係 R_{ij} の性質によって種々の手法に分かれる．たとえば，R_{ij} が，漠然とした親近性を示す場合（数量表現だがメトリカルとはいえない）には，数量化理論第 Ⅳ 類の e_{ij}-型数量化を用いる．R_{ij} は，このほか大小関係，ランク・オーダ，ランク・オーダのついた群分け，ランク・オーダのない単なる群分け，親近性・非親近性，頻度，そして相関係数など，それぞれの R_{ij} の性質に応じて，各種の分析手法が開発されている．

世論調査には，数量化理論第 Ⅲ 類（通称パタン分類）が古くからよく使用されており，意識の構造分析に役立っている．また数量化理論第 Ⅲ 類の計算から求まる個人スコアを用いて Guttman の P.O.S.A. (Partial Order Scalogram Analysis) 図を描き，いくつかの質問を通じての回答の流れを見い出す方法として利用するケースも増えている．

〈数量化理論　第 Ⅲ 類〉

定性的属性への回答 $\delta_{jk}(i)$ が得られたとき，

$$Y(i)=\sum_{j=1}^{m}\sum_{k=1}^{k_j}\delta_{jk}(i)X_{jk}\Big/\sum_{j=1}^{m}\sum_{k=1}^{k_j}\delta_{jk}(i)$$

とおいて，X と Y との相関係数

$$\rho=\frac{C_{XY}}{\sigma_X\sigma_Y}$$

が最大になるよう X_{jk}，$Y(i)$ を求める．

（ここでの記号は前と同じ）

また，Kruskal の方法 (MDSCAL) もよく使われているが，適合度の定義の仕方によって，新旧2つの方式があるが，ここには新 MDSCAL を紹介する．

〈Kruskal の方法……新 MDSCAL〉

n 個 $(1, 2, \cdots i, \cdots j, \cdots, n)$ の対象があり，対象 i, j 間の実測値 δ_{ij} が得られているとき，t 次元空間の n 個の点 x_1, x_2, …, x_n の空間布置を，実測値 δ_{ij} に最もよく適合するように求める．そのため適合度（ストレス）を

$$S=\sqrt{\sum_{i<j}(d_{ij}-\hat{d}_{ij})^2\Big/\sum_{i<j}(d_{ij}-\bar{d})^2}$$

$$\bar{d}=\frac{1}{\sum_{i<j}1}\sum_{i<j}d_{ij}$$

としたとき，適合度 S を最小にする空間布置を求める．実測値 δ_{ij} は非親近性，不一致性，混信度などの値であり，d_{ij} は t 次元空間での点 i, j 間の距離である．また $\hat{d}_{ij}$ は実測値 δ_{ij} とほぼ同じ順序をしている値とする．

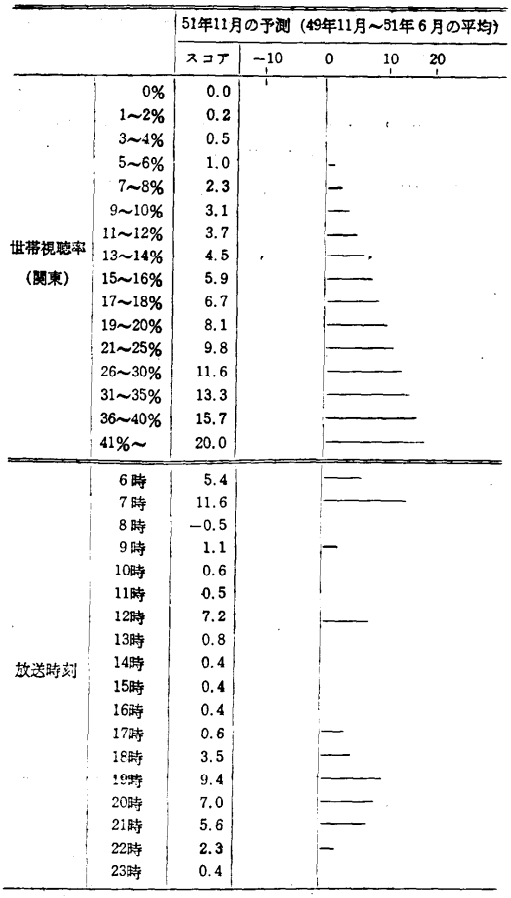

図-1　個人視聴率に対する各要因カテゴリ数量
—全国・NHK 総合テレビ—

3. 応　用　例

世論調査における多次元解析の応用例をいくつか紹介し，その手法の特徴を述べてみよう．

3.1 数量化理論第Ⅰ類…テレビ視聴率の予測[3]

テレビ番組の個人視聴率を，① 世帯視聴率，② 放送時刻，③ 放送種目，④ 放送曜日などの要因から予測する．これは，外的基準のある場合の「量の推定」の問題に該当し，要因はすべて定性的属性とする．なお，要因 ① の世帯視聴率は，% を単位とする数量データであるが，図-1 に示すように，0%，1～2%，3～4%……と 2% づつのカテゴリを作って，わざわざ定性的属性として扱っている．これは，一見，折角の情報を捨てているようであるが，その要因が外的基準

と線形関係にあるかどうか判らないときには有効な手段である.

数量化理論第Ⅰ類では，ある i 番組の個人視聴率を Y_i とした時，その個人視聴率に寄与すると考えられるいくつかの要因 $(1,2\cdots j\cdots)$ のそれぞれのカテゴリ $(1,2\cdots jk\cdots)$ に，その寄与の程度に応じてカテゴリスコア (X_{jk}) を与えようというのである.

49 年 11 月，50年6月，50 年 12 月，51年6月と，4回それぞれ別々に数量化理論第Ⅰ類の計算をすると，実測個人視聴率と再現個人視聴率の相関係数 ρ は，0.96, 0.96, 0.95, 0.95 と非常に高い値が得られた.

つぎに過去4回の傾向から次回の個人視聴率の予測を試みることにし，49 年 11 月から51年6月までのスコアの平均を 51 年 11 月の予測スコアとしてみた.

このスコアから予測した個人視聴率と，実際に調査した 51 年 11 月の個人視聴率との相関図は，図-2 のとおりであり，実測と予測の相関係数は $\rho=0.97$ となり，予測の精度はかなり良いといえよう．なお，予測のはずれ方の大きい番組には，図-2 に番組名を記入してある.

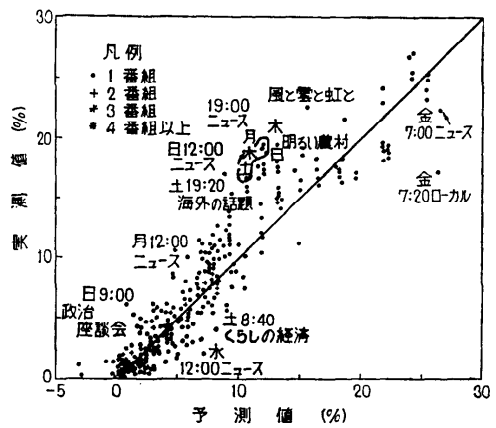

図-2　個人視聴率（関東・NHK 総合テレビ・51 年 11 月）の予測

テレビ番組の個人視聴率の予測をするにあたっての問題点は，

① 定式化の方法
② 番組特性の設定
③ 番組特性の分類

の3つであろう.

① 定式化の方法としては，重線形回帰分析が考えられ，かつて，実際に検討してみたことがある．その結果，数量化理論第Ⅰ類と比較してみると，実測値と再現値の相関係数は，重線形回帰分析で $\rho=0.87$ のところ，数量化理論第Ⅰ類では $\rho=0.95$ となり，後者の方が良い結果であった.

② 番組特性として何を設定したらよいか，要因をさらに追加する必要はないかについても，得られた相関係数で判断できる．また，どんな番組特性を追加するかの検討にあたっては，図-2 の相関図から大きく離れた番組の持つ特性をヒントにするとよい.

③ 番組特性の分類の仕方は，相関係数をあげるのに非常に重要であるが，内分散を小きく，外分散を大さくするように種々試行錯誤を重ねる以外にない.

3.2 数量化理論第Ⅲ類……婦人のニュース接触のパタン分類[5]

婦人調査*で，“どのニュース（またはニュースショー）をよくみているか”に対する回答（多肢選択）をもとに，数量化理論第 Ⅲ 類によるパタン分類を行った.

図-3 は，ニュース番組（27 番組）を，視聴状況の類似性によって，平面上に配置したものである．すなわち全体を通じて同一調査相手による視聴状況の類似しているニュース番組は近くに配置し，視聴状況の類似していないニュース番組は遠くに配置するよう，ニュース番組の位置を数量化理論第 Ⅲ 類によって求めた.

その結果，図-3 に円で囲んだように，婦人のテレビニュースの見方は，次の5つのタイプがあることが判った.

〈視聴のタイプ〉

情報型
- 1　NHK ニュース型
- 2　民放ショー型

娯楽型
- 3　深夜型
- 4　娯楽追随型
- 5　特殊型……接触量極めて小

しかし，このうち「NHK ニュース型」と「民放ショー型」との位置は近く，それをまとめることにすると，情報型と名付けられるタイプになる．それ以外のタイプは，どちらかといえば，前後の番組，主として娯楽番組にひかれてついでにスポットニュースをみたり，生活時間との関連で早朝，深夜にニュース番組をみるタイプであり，娯楽型と呼ぶことができよう．図-3 の番組配置をみると，大雑把に言って，横軸

* 東京都の 20～59 才の婦人，有効 358 人，個人面接法，昭和 53 年3月調査.

(X) は，NHK か民放かの軸，縦軸 (Y) は，情報か娯楽の軸といえよう．

ニュース番組について，視聴状況の類似によって図-3 が描けたと同様，調査相手(358 人)についても，視聴状況の類似によって，似ている人を近くに，似ていない人を遠くに配置する図が描ける．それを，調査相手の属性別（年層，有職か家庭婦人か，既婚か未婚か，職業，報道型か非報道型か，NHK か民放か）に平均値を算出し図-4 を描いた．この図-4 と，図-3 の相対的位置関係から，次のようなことがよみとれる．

① 何らかの職業をもっている人は，未婚者を除いて，とにかく“ニュース情報に接したい”と考えている．

② 家庭婦人は，NHK 寄りである．

③ NHK 型の人達は，ストレート・ニュース，ニュースタジオ 102 などの硬派番組を好み，民放型の人たちは，娯楽番組前後のニュースをみている．

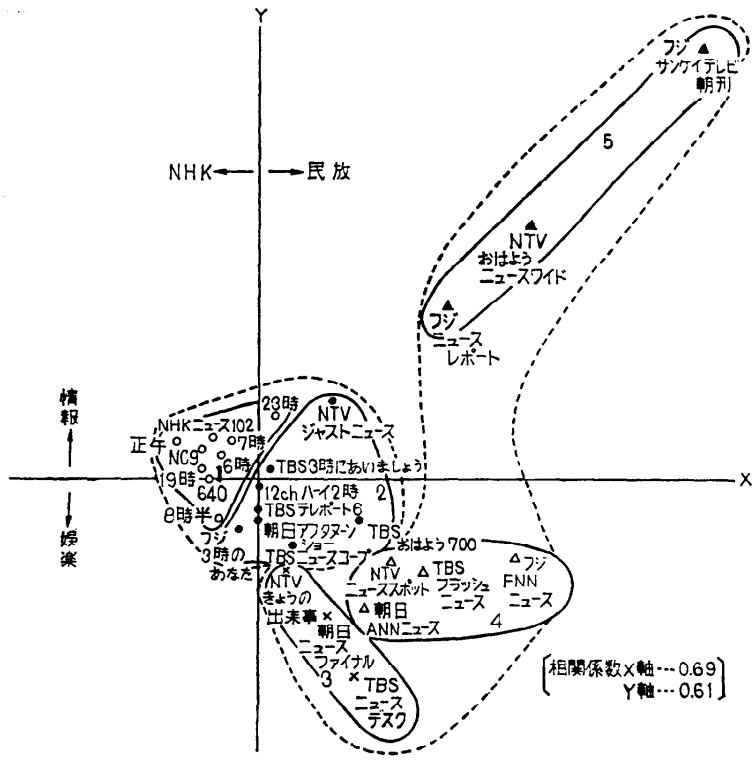

図-3　ニュース番組視聴のパタン分析

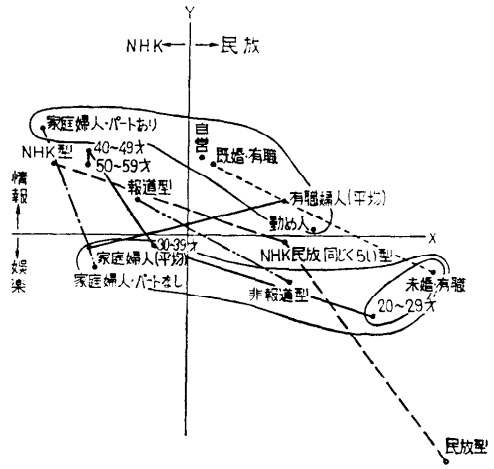

図-4　ニュース番組視聴のパタン分析——婦人の属性別

この例は，多肢選択の場合であるが，数量化理論第Ⅲ類は，単一選択の質問についても扱える．したがって，世論調査の質問相互の関係をみるのに適応範囲の広い手法である．

一方，注意しなければいけないのは，図-3 にもみられるように，回答数の少ないカテゴリが，どのカテゴリとも類似していないという意味で，端の方に外れる傾向があることである．あらかじめそのようなことが判った場合には，そのカテゴリを除外するか，他のカテゴリと合併する方がよい．さらに，職業別の「学生」と学歴別の「在学中」のように，質問は違っても，調査相手でみると全く同じカテゴリは，同時に扱うことは出来ないので，どちらか一方をとるようにする必要がある．

3.3 P.O.S.A.……宗教・信仰行動の分析[6]

L. Guttman の P.O.S.A. は，複雑な現象を整理するのに有効な手段といわれているが，その解法を具体的に知り得ないので，数量化理論第 Ⅲ 類を利用して，P.O.S.A. の分析をする．P.O.S.A. では，多少再現率が低下しても，質的に著しく異なりかつ頻度の少ない回答パタンを除外することによって，大多数の回答パタンをできるだけ次元の少ない空間に配置する

ことを目的としている．これと同じことを，数量化理論第 Ⅲ 類の手法を用いて求める．

データは，前述の日本人の「宗教・信仰」に関する世論調査である．

宗教・信仰行動のうち，実利的信仰を意味している D. 商売・合格祈願，E. お守り・おふだ，F. おみくじ・占いの3カテゴリについて，数量化理論第 Ⅲ 類の計算をし，その結果得られた調査相手のスコア，すなわち $2^3=8$ 通りの回答パタンを配置したのが**図-5**である．(YES を 1，NO を2で表示)

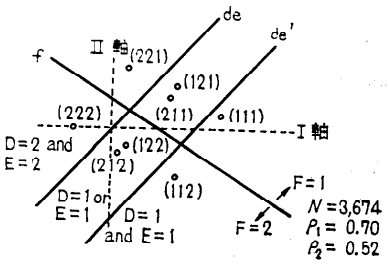

図-5　回答パタンの分布……実利的信仰 (D. E. F)

図-5 を手がかりに，頻度の比較的少ない (1, 2, 1) 109人を除いて，POSA 図を描いたのが**図-6**である．P. O. S. A. の考えでは，変化が1つだけの回答パタン間を細線で結んでいる．

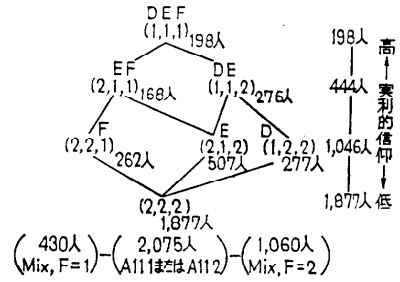

図-6　実利的信仰の POSA (D. E. F)

D, E, F とも全部 NO の (2, 2, 2) から，D, E, F とも全部 YES の (1, 1, 1) までに，いくつかの流れがある．その1つは，図-6 の左回りでの F-E-D の順に YES になる流れであり，もう1つは，右回りの D-E-F の順に YES になる流れである．人数として大きいのは，真中の E が YES になり，ついで D または F が YES になり，最後にすべてが YES になる流れである．

この P. O. S. A. の再現率は，3,674 人中 3,565 人で再現率 97% である．

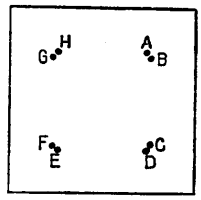

図-7　8項目の原配置図

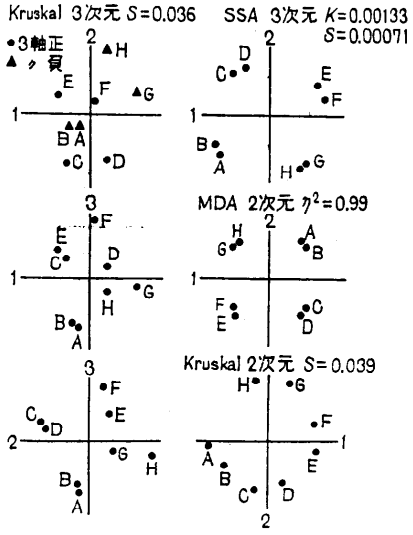

図-8　8項目の空間布置図

このように，実利的信仰3項目については，数量化理論第 Ⅲ 類を用いて，比較的簡単に P. O. S. A. が描けたが，いつもこうとはいえない．項目数が増えれば，回答パタンの組み合わせが増えるので，手作業では 5〜6 項目が精一杯であろう．また，どの回答パタンを除くことによって，複雑さがどのくらい単純化されるかについては，名人芸的なところがあり，経験を重ねる以外にないようである．

3.4 Kruskal, SSA, MDA の次元の問題[7]

多次元尺度構成といわれる手法のうち Kruskal の MDSCAL, Guttman の SSA, 林の MDA-OR の3つの方法について，解を求めるに際して，次元をどこまでにするか，それによって空間布置がどう変るかを，モデル・データで検討した．

モデル・データは，**図-7** に原配置図を示した，A〜H の8項目である．AB, CD, EF, GH が2つづつ組んで，4カ所にまとまっているデータであり，この図-7 に基づく，8項目間の距離行列は与えられているとする．

Kruskal と SSA とは，この項目間の距離をそのま

ま用いずに距離順位数 (distance ranking number) に直して用いる．MDA-OR の場合は，さらにグループ順位 (rank group relation) に直して用いる．

グループ順位では，全般的にみて小さな差は無視されるが，距離順位数では，0.01 の差も 0.99 の差も同じ1順位の差となることがある．

このような，各インプット行列の特有の性質を前提にして，得られた空間布置 (**図-8**) を眺めてみよう．

Kruskal では，最小のストレスを示した次元を解とするが，1次元 $S=0.383$，2次元 $S=0.039$，3次元 $S=0.036$ となり，3次元が解となる．3次元の場合の空間布置は，1×2 軸，1×3 軸，2×3 軸どれも原配置図に似ていない．これに対し，2次元の空間布置の方が，原配置図にやや近い．

SSA では，逸脱係数 K の最小のものである3次元での空間布置を描くと，1×2 軸の場合，やや原配置図に似ている（1次元 $K=0.33638$，2次元 $K=0.00136$，3次元 $K=0.00133$）．

MDA-OR では，1次元 $\eta_1^2=0.415$，2次元 $\eta_2^2=0.993$ と，2次元でほぼ η^2 の最大値である 1.0 に近くなる．2次元の MDA-OR の空間布置は完全に，原配置図と一致している．

以上の検討してきたように，次元の決定に際して Kruskal の方法には少し難点がある．一般の場合には，モデルの場合とは異なり，当然，原配置図を知らないで分析をするのであるから，このように誤った結論に到達する例が1つでもあるのは，心配なことである．

4. おわりに

電子計算機の発達，わけても，この種の多次元解析のプログラムが開発・整備されるにしたがって，調査データの多次元解析をした報告が増えてきている．その結果，データの全貌がつかめるようになり，データの全体構造や，質問の相互関連を見通せるようになり，分析に厚みを増しているのは，好ましい傾向である．

一方，データを入れさえすれば，何らかのアウトプットが得られるという手軽さから，手法を充分に消化せずに適用しているケースが時折みうけられるのは残念である．これらの手法はいずれも，ある種の条件，ある種の仮定を置いて解いているのであるから，その条件・仮定を理解した上で，結果の解釈にあたらなければならないのである．

それよりも，調査データの性質，分析の目的に合致した手法を選択する必要があろう．また，この種の計算は，一回で，良い結果が得られるとは限らない．要因を追加し，入れかえ，さらに，要因の分類基準を修正し，カテゴリをまとめるなどの試行錯誤をくりかえすうちに，良い結果に到達することが多い．電子計算機が，スマートに解いているかのようであるが，調査データの取り扱いには，この手づくりの要素が多いのである．

参考文献

1) 飽戸 弘：数量化理論，年報社会心理学，第5号，pp. 73-103 (1964)．
2) 林 知己夫・飽戸 弘共編：多次元尺度解析法，サイエンス社，p. 5 (1976)．
3) 高宮義雄・杉山明子：個人視聴率と世帯視聴率との関係，NHK 文研月報，第28巻10号(1978)．
4) 杉山明子：聴視率の予測研究，NHK 放送文化研究年報 10 (1965)．
5) テレビ報道評価の研究 ～phase I～，1978年5月テレビ報道研究会（未公刊）．
6) 前掲 2) pp. 219-235.
7) 前掲 2) pp. 186-192.

(昭和54年1月16日受付)