

## 多集合データのための階層的な主成分分析\*

村上 隆

### 1. 序—問題の所在

異なる内容領域に対応する、複数の項目群からなる調査の結果として得られるデータを多集合データ (multiset data) と呼ぶ (図1)。このようなデータの探索的因子分析の方法として、村上 (1985) は、多集合因子分析 (multiset factor analysis) を提案した。多集合因子分析は、共通因子モデル、すなわち、因子得点が不定性をもつタイプのモデルであったが、本研究では、因子得点が一義的にデータから確定するような、多集合データのための主成分モデルを提案する。

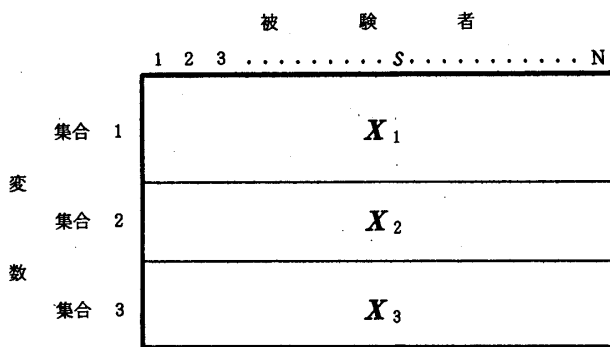


図1 多集合データ

村上 (1985) と若干の重複が生ずるが、本研究のモチベーションについて述べるところから始めよう。一般に、質問紙を用いた個人差測定の研究においては、幾つかの項目に対する反応を合成して尺度とするのが普通である。この際、基準として用いられるのは、通常、尺度の内的整合性である。具体的には、多集合データで、集合ごとに、主成分分析または因子分析を適用し、項目の分類、選択を行うことが多い。しかしながら、複数の集合が存在するという事は、調査の目的自体に集合間の関係を明らかにすることが含まれていると考えられ、その点も

\* 本研究の数値計算は、名古屋大学大型計算機センター、FACOM M-382 システムによった。

考慮すると、集合ごとに内的整合性だけを考慮していく分析には、不十分な面が出てくる。

この点について、図式を用いて、もう少し詳しく説明してみよう。集合の数は2とし、それぞれ6つずつの項目を含むとしよう。それぞれの集合に属する変数を、 $V_{11}$ — $V_{61}$ ,  $V_{12}$ — $V_{62}$  のように書くことにする。また、第1の集合において定義される因子、または合成変数を、 $F_{11}$ ,  $F_{21}$ 、第2の集合におけるそれを、 $F_{12}$ ,  $F_{22}$  のように示す。これらの変量間の関係は、図2のように表現することができる。

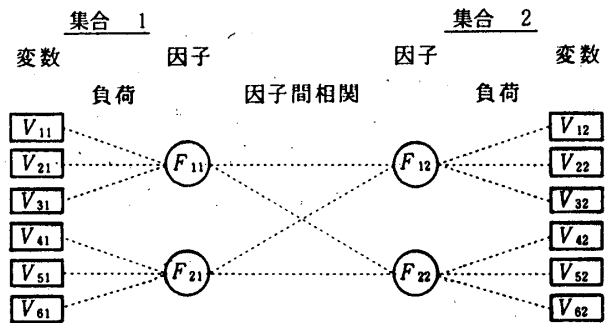


図2 2つの集合の因子と集合間関係(一般的表現)

Jöreskog (1974) 等にしたがって、直接観測可能な変数である  $V$  については  $\square$  で、そこから推定あるいは定義され、直接観測可能でない変数である  $F$  については、 $\circ$  で囲んで示す。変数の間を結ぶ直線は、相関関係を示すものとしよう。(この点で、この図は、因果関係を記述するパス・ダイアグラムとは異なる。)

ここで、各集合において尺度を定義するにあたっては、2つの極端な立場がありうる。1つは、上述の内的整合性を重視する立場、すなわち、集合ごとに個別に主成分分析を適用することである。これは、 $V$  と  $F$  の間の相関 (より正確にはその2乗和) を最大化しようとする (図3)。この結果、異なる集合間関係には、特別に注意がはられず、それらは結果として与えられる。

このやり方では、尺度の信頼性は一般に高くなるが、極端な場合、1つの尺度に含まれる項目 (1つの因子に

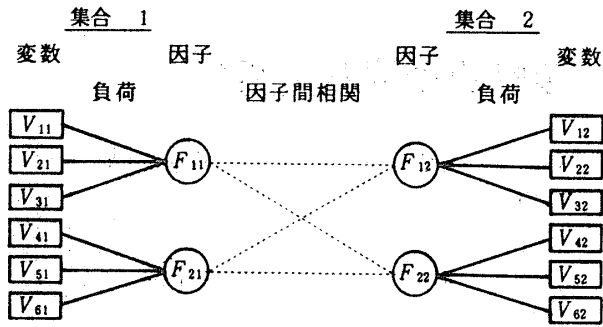


図3 主成分分析 (太線のところの最大化を目指す)

高く負荷する項目)は、多少、ワーディングは違うものの内容的にはほとんど同一ということになりかねない。このような尺度は、測定を目指す概念の、ごく狭い一側面を反映するにすぎないために、外在基準や関連を検討すべき他の変数と高い相関をもたず、有用なものとはならない可能性が大きい。すなわち、このやり方で定義される尺度は、信頼性のみを偏重し、妥当性を軽視したものとなる。

集合間の関係に関して言えば、このやり方では、興味ある関係が見落とされる危険性がある。

尺度の定義に関して、もう1つの極端な立場は、図4に示すように、2つの集合の合成変量間の相関を最大化しようとするものである。このような分析モデルとしては、正準相関分析がある。この方法(ただし、後述のような修正を施した形)では、合成変量間の相関(正準相関)の和が最大化される。

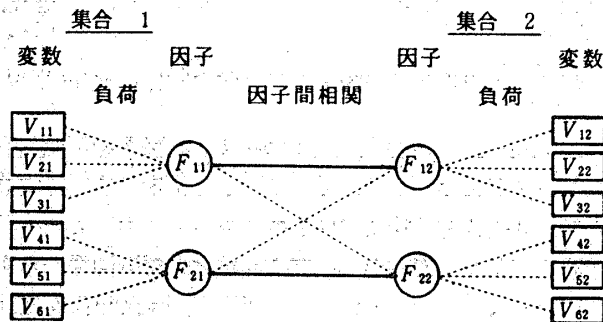


図4 正準相関分析 (太線のところの最大化を目指す)

こちらのやり方では、尺度は内容的に多様な、いわば幅をもったものとなる可能性がある。しかしながら、信頼性は低く内容は解釈困難なものとなりがちである。しかも、標本の大きさに比して項目数が多い場合には、この方法の基準は誤差成分をひろい上げやすいために、別の標本では集合間相関が一挙に低下してしまう、といったことが起こりうる。(重回帰分析において、shrinkage

とよばれている現象である。)

結局、正準相関分析は、主成分分析とは逆に、尺度の妥当性を(しかも、比較的小標本の場合には、不十分なやり方で)重視するあまり、信頼性を不当に軽視する方法ということになる。

標本論の問題は、主成分分析にもあるが、モデルの性質上、正準相関分析の方により問題が多い。

いずれにしても、問題は、個人差測定の研究、特にその探索的段階においては、不完全なデータに不完全なモデルをあてはめることによって、分析を行わざるを得ないところにある。母数に十分近い標本統計量が得られるような大サンプルで、かつ比較的少数の因子によって、ほぼ完全に記述できるようなデータが得られるような事態では、後に示唆するように、どんな基準にもとづく方法を用いたとしても、ほとんど同じ解釈に導くような結果となるはずである。しかし、測定すべき概念について、必ずしも明確な理論がなく、したがって、被験者数に比して多数の項目を含めざるを得ない、予備調査のデータの分析等においては、方法の選択はかなりクリティカルな意味をもつ。

探索的段階の研究では、尺度にどの程度の内的整合性を期待すべきか、あるいは、どの程度の多様性を含めるべきかという点について、明瞭な基準を設定することが難しい。しかも、この2つの条件は、Cronbach(1970)が、bandwidth-fidelity dilemmaと呼んでいるように、本質的に両立させることが不可能である。主成分分析と正準相関分析は、このdilemmaを両極端の方向で解決しようとするものである。実際には、どこかでバランスをとって妥協する他はないが、そのバランス点は、両極の間に無数に設定し得る。結局は、研究の目的、測定しようとする概念の性質等を考慮して、解決されなければならないのであり、なんらかの絶対的な基準が設定できるような問題ではない。

しかしながら、ある程度の一般性をもつ基準が全くあ

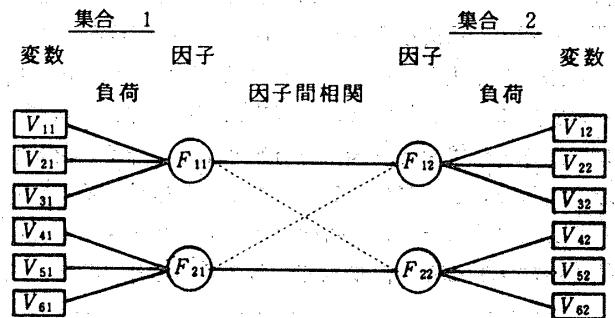


図5 多集合因子分析 (負荷と因子得点間相関をバランスをとってともに大きく)

り得ないわけではない。例えば、前述の多集合因子分析は、2つの集合の全項目間相関行列を最小2乗的な意味で最大限に説明する、という基準で、個々の集合の因子負荷と因子得点間相関行列を推定しようとするものであった(図5)。

本研究では、個々の集合における因子の背後にさらに高次の因子を仮定することによって、バランス点をもとめようとする(図6)。このようなやり方は、次の節で示すように、2段階の factoring に対応しており、階層的的主成分分析(hierarchical component analysis)と呼ぶことができる。通常、高次の因子は、斜交回転された因子得点間相関行列を再度 factoring することによって得られる(例えば、Harman, 1976)。しかし、本研究では、集合ごとの因子(それらは、各集合内では直交する)を factoring する形になっている。このような用法は3相因子分析をこの観点から検討した Bloxom (1984) にも見出される。

いずれにしても、階層的的主成分分析は、前述のように、素データの一次結合としての合成変量が明確に定義されるという意味で、因子分析というより、主成分分析的なモデルである。また、後に示すように、3相データのための探索的モデルである、準3相主成分分析の一般化でもある。

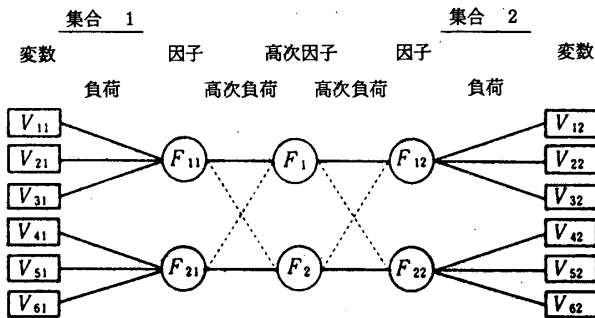


図6 階層的的主成分分析

## 2. モデル

次のような基本モデルを考える。

$$Z_k = A_k C_k F + E_k \quad k = 1, \dots, p \quad (1)$$

ここで、

$Z_k$ :  $n_k \times N$  のデータ行列。集合  $k$  における測定値(村上, 1985と行と列が入れ換わっていることに注意)

$A_k$ :  $n_k \times q_k$  の因子負荷行列

$C_k$ :  $q_k \times Q$  の高次因子負荷行列

$F$ :  $Q \times N$  の因子得点行列

$E_k$ :  $n_k \times N$  の誤差行列

ここで、 $n_k$  は集合  $k$  の変数、 $N$  は個体の数、 $p$  は集合の数、 $q_k$  は各集合の因子数、 $Q$  は高次因子の数を示す。データ行列  $Z_k$  について、

$$Z_k \mathbf{1} = \mathbf{0} \quad k = 1, \dots, p \quad (2)$$

$$\text{diag}\left(\frac{1}{N} Z_k Z_k'\right) = \mathbf{I} \quad k = 1, \dots, p \quad (3)$$

を満たすような前処理を行う。したがって、

$$R_{kl} = \frac{1}{N} Z_k Z_l' \quad \begin{matrix} k = 1, \dots, p \\ l = 1, \dots, p \end{matrix} \quad (4)$$

は、 $n_k \times n_l$  の2つの集合間の変数間相関行列となる。

伝統的な因子分析のスケーリングにしたがって、

$$\frac{1}{N} F F' = \mathbf{I} \quad (5)$$

$$C_k C_k' = \mathbf{I} \quad k = 1, \dots, p \quad (6)$$

の制約条件を課す。また、

$$\max(q_1, \dots, q_p) \leq Q \leq \sum_{k=1}^p q_k \quad (7)$$

とし、データ行列のランクは  $n_k$ 、行列  $A_k$  のランクは  $q_k$ 、行列  $[C_1', \dots, C_p']'$  のランクは  $Q$  とする。

このモデルは、直観的に言えば、次のように導出される。すなわち、図7Aのように、2つのデータ行列をまず個別に主成分分析したとする。次に、ここで得られた2つの因子得点行列を並べて、図7Bのようにもう一度主成分分析を行う。最初の個別の主成分分析により、因子負荷行列  $A_k$  が、二度目の主成分分析によって、高次因子負荷行列  $C_k$  と、因子得点行列  $F$  が得られる。

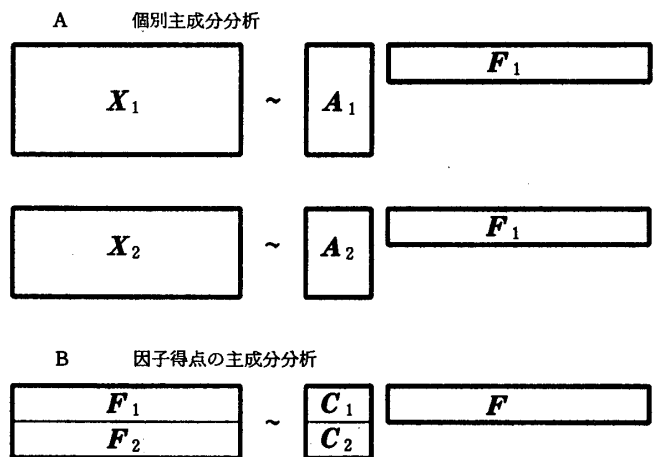


図7 階層的的主成分分析のモデルの直観的説明

ただし、この説明は、あくまでもモデルの形の説明にとどまる。このようなやり方では、当然、個別の主成分分析と何ら変わりのない結果しか得られない。実際のアルゴリズムは、素データ  $Z_k$  に対する最小2乗法によってなされ、その結果、得られる因子負荷行列は、個別の主成分分析によるそれとは異なったものとなる。次に、そのアルゴリズムについて説明しよう。

### 3. アルゴリズム

制約条件を含めて、 $\tilde{M}$ ,  $M_k$  を Lagrange の定数を要素とする行列として、

$$S = \sum_{k=1}^p \text{tr} \frac{1}{N} \mathbf{E}_k \mathbf{E}_k' - \text{tr} \tilde{M} \left( \frac{1}{N} \mathbf{F} \mathbf{F}' - \mathbf{I} \right) - \sum_{k=1}^p \text{tr} M_k (\mathbf{C}_k \mathbf{C}_k' - \mathbf{I}) \quad (8)$$

を最小化する通常の最小2乗基準をとる。これを  $A_k$ ,  $C_k$ ,  $F$  で偏微分して  $\mathbf{0}$  とおくことにより、

$$\frac{1}{N} \mathbf{Z}_k \mathbf{F}' \mathbf{C}_k' = \mathbf{A} \quad k = 1, \dots, p \quad (9)$$

$$\mathbf{A}_k \frac{1}{N} \mathbf{Z}_k \mathbf{F}' - \mathbf{A}_k' \mathbf{A}_k \mathbf{C}_k = \mathbf{A}_k \mathbf{C}_k \quad \left( \mathbf{A}_k = \frac{\mathbf{M}_k + \mathbf{M}_k'}{2} \right) \quad k = 1, \dots, p \quad (10)$$

$$\sum_{k=1}^p \mathbf{C}_k' \mathbf{A}_k' \mathbf{Z}_k - \sum_{k=1}^p \mathbf{C}_k' \mathbf{A}_k' \mathbf{A}_k \mathbf{C}_k \mathbf{F} = \tilde{\mathbf{A}} \mathbf{F} \quad \left( \tilde{\mathbf{A}} = \frac{\tilde{\mathbf{M}} + \tilde{\mathbf{M}}'}{2} \right) \quad (11)$$

を得る。(10)の右から  $C_k'$  をかけ、(6), (9)を用いて、

$$\mathbf{A}_k = \mathbf{A}_k' \left( \frac{1}{N} \mathbf{Z}_k \mathbf{F}' \mathbf{C}_k' - \mathbf{A}_k \right) = \mathbf{0} \quad k = 1, \dots, p$$

である。したがって、(10)は、次のようにかきかえられる。

$$\mathbf{A}_k' \frac{1}{N} \mathbf{Z}_k \mathbf{F}' = \mathbf{A}_k' \mathbf{A}_k \mathbf{C}_k \quad k = 1, \dots, p \quad (12)$$

次に、(11)の右から、 $\frac{1}{N} \mathbf{F}'$  をかけ、(5), (12)を用いて、

$$\begin{aligned} \tilde{\mathbf{A}} &= \sum_{k=1}^p \mathbf{C}_k' \mathbf{A}_k' \frac{1}{N} \mathbf{Z}_k \mathbf{F}' - \sum_{k=1}^p \mathbf{C}_k' \mathbf{A}_k' \mathbf{A}_k \mathbf{C}_k \\ &= \sum_{k=1}^p (\mathbf{C}_k' \mathbf{A}_k' \mathbf{A}_k \mathbf{C}_k - \mathbf{C}_k' \mathbf{A}_k' \mathbf{A}_k \mathbf{C}_k) = \mathbf{0} \end{aligned}$$

となる。したがって、(11)は次のようにかきかえられる。

$$\left( \sum_{k=1}^p \mathbf{C}_k' \mathbf{A}_k' \mathbf{A}_k \mathbf{C}_k \right) \mathbf{F} = \sum_{k=1}^p \mathbf{C}_k' \mathbf{A}_k' \mathbf{Z}_k \quad (13)$$

行列  $\sum_{k=1}^p \mathbf{C}_k' \mathbf{A}_k' \mathbf{A}_k \mathbf{C}_k$  のランクは、 $Q$  だから、 $\tilde{\mathbf{L}}$  を  $Q$  次の正定符号の対角行列、 $\tilde{\mathbf{T}}$  を  $Q$  次の正規直交行列として、

$$\sum_{k=1}^p \mathbf{C}_k' \mathbf{A}_k' \mathbf{A}_k \mathbf{C}_k = \tilde{\mathbf{T}} \tilde{\mathbf{L}} \tilde{\mathbf{T}}' \quad (14)$$

のように分解できる。これを用いて、

$$\mathbf{F} = \tilde{\mathbf{T}} \tilde{\mathbf{L}}^{-1} \tilde{\mathbf{T}}' \left( \sum_{k=1}^p \mathbf{C}_k' \mathbf{A}_k' \mathbf{Z}_k \right) \quad (15)$$

(15)を(12)に代入して、(4)を用い、

$$\sum_{l=1}^p \mathbf{A}_k' \mathbf{R}_{kl} \mathbf{A}_l' \mathbf{C}_l' \tilde{\mathbf{T}} = \mathbf{A}_k' \mathbf{A}_k \mathbf{C}_k \tilde{\mathbf{T}} \tilde{\mathbf{L}} \quad k = 1, \dots, p$$

が得られるが、この左から  $(\mathbf{A}_k' \mathbf{A}_k)^{-\frac{1}{2}}$  をかけて、

$$\mathbf{A}_k^* = \mathbf{A}_k (\mathbf{A}_k' \mathbf{A}_k)^{-\frac{1}{2}} \quad k = 1, \dots, p$$

$$\mathbf{C}_k^* = (\mathbf{A}_k' \mathbf{A}_k)^{\frac{1}{2}} \mathbf{C}_k \tilde{\mathbf{T}} \quad k = 1, \dots, p$$

とおき、

$$\sum_{l=1}^p \mathbf{A}_k^* \mathbf{R}_{kl} \mathbf{A}_l^* \mathbf{C}_l^* = \mathbf{C}_k^* \tilde{\mathbf{L}} \quad k = 1, \dots, p \quad (16)$$

を得る。次に、(9)に(15)を代入し、 $\mathbf{A}_k \mathbf{C}_k \mathbf{T} = \mathbf{A}_k^* \mathbf{C}_k^*$  を用いて、

$$\mathbf{A}_k^* = \left( \sum_{l=1}^p \mathbf{R}_{kl} \mathbf{A}_l^* \mathbf{C}_l^* \right) \tilde{\mathbf{L}}^{-1} \mathbf{C}_k^* (\mathbf{A}_k' \mathbf{A}_k)^{-1} \quad k = 1, \dots, p$$

ここで、 $\mathbf{L}_k$  を  $q_k \times q_k$  の正定符号対角行列、 $\mathbf{T}_k$  を正規直交行列として、

$$\mathbf{A}_k' \mathbf{A}_k = \mathbf{T}_k \mathbf{L}_k \mathbf{T}_k' \quad k = 1, \dots, p$$

と分解できる。そこで、

$$\mathbf{A}_k^\dagger = \mathbf{A}_k^* \mathbf{T}_k \quad k = 1, \dots, p$$

$$\mathbf{C}_k^\dagger = \mathbf{T}_k' \mathbf{C}_k^* \quad k = 1, \dots, p$$

とおき、さらに、(16)から、

$$\mathbf{C}_k^\dagger = \tilde{\mathbf{L}}^{-1} \left( \sum_{l=1}^p \mathbf{C}_l^\dagger \mathbf{A}_l^\dagger \mathbf{R}_{lk} \right) \mathbf{A}_k^\dagger \quad k = 1, \dots, p$$

となるから、これを代入して整理し、

$$\left\{ \left( \sum_{l=1}^p \mathbf{R}_{kl} \mathbf{A}_l^\dagger \mathbf{C}_l^\dagger \right) \tilde{\mathbf{L}}^{-2} \left( \sum_{l=1}^p \mathbf{C}_l^\dagger \mathbf{A}_l^\dagger \mathbf{R}_{lk} \right) \right\} \mathbf{A}_k^\dagger = \mathbf{A}_k^\dagger \mathbf{L}_k \quad k = 1, \dots, p \quad (17)$$

を得る。(16)には、 $\mathbf{T}_k$  を左からかけて、

$$\sum_{l=1}^p (A_k^\dagger R_{kl} A_l^\dagger) C_l^\dagger = C_k^\dagger \tilde{L} \quad k=1, \dots, p \quad (18)$$

が得られる。かくして、(17), (18)を同時に満足する。

$A_k^\dagger, L_k, C_k^\dagger, \tilde{L}$  を得たとすれば、

$$A_k = A_k^\dagger L_k^{-\frac{1}{2}} T_k' \quad k=1, \dots, p \quad (19)$$

$$C_k = T_k L_k^{-\frac{1}{2}} C_k^\dagger T' \quad k=1, \dots, p \quad (20)$$

によって、最終的な解が得られる。これらは、直交回転  $T_k, \tilde{T}$  に関する不定性をもっており、例えば、バリマックス回転等によって、単純構造への回転を行う。すなわち、まず  $A_k$  が単純構造をなすように、バリマックス回転によって  $T_k$  を定め、次にこれによって回転された、行列  $(C_1^\dagger L_1^{-\frac{1}{2}} T_1', \dots, C_p^\dagger L_p^{-\frac{1}{2}} T_p')$  をバリマックス回転して  $\tilde{T}$  を定める。回転はこの順序で行うのが、解釈しやすい結果を与えることにつながるようである。

また、(13)から、

$$F = \left( \sum_{k=1}^p C_k' A_k' A_k C_k \right)^{-1} \left( \sum_{k=1}^p C_k' A_k' Z_k \right) \quad (21)$$

によって因子得点が得られる。

したがって、アルゴリズムの主要部分は、(17), (18)を同時に満足するような  $A_k^\dagger$  と  $C_k^\dagger$  を求めることである。これにはまず、 $A_k^\dagger$  の適当な初期値を用いて、(18)の左辺の行列、

$$\begin{bmatrix} A_1^\dagger R_{11} A_1^\dagger & \dots & A_1^\dagger R_{1p} A_p^\dagger \\ \vdots & & \vdots \\ A_p^\dagger R_{p1} A_1^\dagger & \dots & A_p^\dagger R_{pp} A_p^\dagger \end{bmatrix}$$

を計算し、この行列の固有ベクトルを要素とする行列

$$C^\dagger = (C_1^\dagger, \dots, C_p^\dagger)' \text{ を、}$$

$$C^\dagger C^\dagger = \tilde{L}$$

と基準化したものを求める。つぎに、これと先の  $A_k^\dagger$  とをもちいて(17)の左辺の行列、

$$\left\{ \left( \sum_{l=1}^p R_{kl} A_l^\dagger C_l^\dagger \right) \tilde{L}^{-2} \left( \sum_{l=1}^p C_l^\dagger A_l^\dagger R_{lk} \right) \right\}$$

を算出し、その単位長さに基準化された固有ベクトルを要素とする行列を新しい  $A_k^\dagger$  とする。そして、以下このプロセスを収束するまで反復する(交互最小2乗法)。

なお、(17)には、一般に  $Q$  個、(18)には、 $\sum_{k=1}^p q_k$  個の 0 でない固有値が存在する。ここで、

$$S = \sum_{k=1}^p n_k - \sum_{k=1}^p \text{tr} L_k = \sum_{k=1}^p n_k - \text{tr} \tilde{L} \quad (22)$$

が成立することは容易に示せるから、それぞれ、最大  $q_k$  番目、 $Q$  番目までの固有値を採用すればよい。ここで、

$$\text{tr} \tilde{L} = \sum_{k=1}^p \text{tr} L_k \text{ の値は、モデルによって説明される}$$

分散の大きさと解釈される。

$A_k^\dagger$  の初期値は、行列  $R_{kk}$  の最大  $q_k$  番目までの固有値に対応する固有ベクトルを用いる。これは、個別の主成分分析の結果に等しく、比較的よい初期値となることが期待される。

#### 4. パラメータの解釈

集合ごとの因子得点  $\tilde{F}_k$  を、

$$\tilde{F}_k = C_k F \quad k=1, \dots, p \quad (23)$$

と定義する。このとき、(9)は、

$$A_k = \frac{1}{N} Z_k \tilde{F}_k \quad k=1, \dots, p \quad (24)$$

を意味しており、また、(23)の左から、 $\frac{1}{N} F'$  をかけて、(5)を用いることにより、

$$C_k = \frac{1}{N} \tilde{F}_k F' \quad (25)$$

が得られる。ここで、行列  $Z_k, F, \tilde{F}_k$  の各行はすべて、平均0分散1に基準化されていることが、(2), (3), (5), (6), (21)からわかるから、行列  $A_k$  と  $C_k$  はともに相関行列として解釈することができる。また、

$$C_k C_l' = \frac{1}{N} \tilde{F}_k \tilde{F}_l' \quad \begin{matrix} k=1, \dots, p \\ l=1, \dots, p \end{matrix} \quad (26)$$

だから、行列  $C_k C_l'$  は、集合間の因子得点間相関行列と解釈できる。

これらの性質は、(5), (6)の制約条件から生まれたものであり、結果の解釈に、大いに助けとなる。

#### 5. 階層的モデルの意味および他のモデルとの関係

まず、本研究における階層的成分分析が、通常の成分分析の自然な一般化であることを示すために、 $p=1$  の場合について3節のアルゴリズムが通常の成分分析と同一の結果を導くことを示そう。

基本モデルは、添え字を落として、

$$Z = ACF + E \quad (27)$$

のように書くことにする。なお、ここで、(7)式の制約条件から、 $Q=q$  であり、また  $C$  は、正方行列であり、かつ(6)式の条件から正規直交行列であることに注意する。最小化基準は、

$$S = \text{tr} \mathbf{E} \mathbf{E}' + \text{tr} \mathbf{M} \left( \frac{1}{N} \mathbf{F} \mathbf{F}' - \mathbf{I} \right) + \text{tr} \tilde{\mathbf{M}} (\mathbf{C} \mathbf{C}' - \mathbf{I}) \quad (28)$$

のようになる。以下、一般の場合と同様のプロセスを経て、(17)式は、

$$(\mathbf{R} \mathbf{A}' \mathbf{C}' \tilde{\mathbf{L}}^{-2} \mathbf{C}' \mathbf{A}' \mathbf{R}) \mathbf{A}' = \mathbf{A}' \mathbf{L} \quad (29)$$

となり、(18)式は、

$$\mathbf{A}' \mathbf{R} \mathbf{A}' \mathbf{C}' = \mathbf{C}' \tilde{\mathbf{L}} \quad (30)$$

となる。(30)を(29)に代入して整理すると、

$$\mathbf{R} \mathbf{A}' \mathbf{C}' \tilde{\mathbf{L}}^{-1} \mathbf{C}' = \mathbf{A}' \mathbf{L}$$

となる。ここで、 $\mathbf{C}' \mathbf{C}' = \tilde{\mathbf{L}}$  であることから、行列  $\mathbf{C}' \tilde{\mathbf{L}}^{-\frac{1}{2}}$  は、正規直交行列であることに注意すると、

$$\mathbf{C}' \tilde{\mathbf{L}}^{-1} \mathbf{C}' = \mathbf{I}$$

したがって、(29)は、

$$\mathbf{R} \mathbf{A}' = \mathbf{A}' \mathbf{L} \quad (31)$$

と書き替えられる。これは、 $\mathbf{A}'$  が、相関行列  $\mathbf{R}$  の固有ベクトルであることを意味している。(31)の左から  $\mathbf{A}'$  をかけ、 $\mathbf{A}' \mathbf{A}' = \mathbf{I}$  であることを利用して、

$$\mathbf{A}' \mathbf{R} \mathbf{A}' = \mathbf{L}$$

となる。したがって、(30)は、

$$\mathbf{L} \mathbf{C}' = \mathbf{C}' \tilde{\mathbf{L}} \quad (32)$$

となる。ここで、 $\mathbf{L}$  は対角行列だから、その固有ベクトルである  $\mathbf{C}'$  は、 $\mathbf{L}$  の対角要素がすべて異なると仮定すれば、制約条件に注意して、

$$\mathbf{C}' = \tilde{\mathbf{L}}^{\frac{1}{2}} \quad (33)$$

であり、固有値を要素とする行列  $\mathbf{L}$  は、 $\mathbf{L}$  そのもの、すなわち、

$$\tilde{\mathbf{L}} = \mathbf{L} \quad (34)$$

となる。かくして、因子負荷は、(19)と同様、

$$\mathbf{A} = \mathbf{A}' \mathbf{L}^{\frac{1}{2}} \mathbf{T} \quad (35)$$

によって得られるが、これは通常的主成分分析において、重みベクトルを構造ベクトルに変換して回転する公式そのものである。一方、高次因子負荷  $\mathbf{C}$  は、

$$\mathbf{C} = \mathbf{T} \tilde{\mathbf{T}}$$

のように、2つの直交行列の積となるが、ここで、 $\mathbf{T} = \tilde{\mathbf{T}}$  とすれば、

$$\mathbf{C} = \mathbf{I} \quad (36)$$

となって、基本モデルそのものが、

$$\mathbf{Z} = \mathbf{A} \mathbf{F} + \mathbf{E} \quad (37)$$

のように、通常的主成分分析のそれと一致する。(実際のアルゴリズムにおいて、 $\mathbf{C}$  を単純構造化しようとするとき、 $\mathbf{C}$  は結局は符号を除き単位行列となって、この結果が得られることになる。) このとき、因子得点をもとめるための式(21)は、

$$\mathbf{F} = (\mathbf{A}' \mathbf{A})^{-1} \mathbf{A}' \mathbf{Z} \quad (38)$$

のように、回転後の因子負荷を用いて、主成分得点を求める公式そのものとなる。

次に、第1節において述べた、多集合データにおいて、集合ごとに個別に主成分分析を適用した場合との関係について考えよう。個別の主成分分析は、

$$\mathbf{Z}_k = \mathbf{A}_k \mathbf{F}_k + \mathbf{E}_k \quad k = 1, \dots, p \quad (39)$$

のような基本モデルを、

$$S_s = \sum_{k=1}^p \text{tr} \mathbf{E}_k \mathbf{E}_k' + \sum_{k=1}^p \text{tr} \mathbf{M}_k \left( \frac{1}{N} \mathbf{F}_k \mathbf{F}_k' - \mathbf{I} \right) \quad (40)$$

を最小化するような基準で解いていることになる。こ

で、 $Q = \sum_k q_k$  , すなわち、高次因子の数を、(7)式の

上限に等しくとった場合についてみると、 $\mathbf{C}_k$  を  $q_k \times Q$  の高次因子負荷行列、 $\mathbf{F}$  を  $Q \times N$  の因子得点行列として、

$$\mathbf{F}_k = \mathbf{C}_k \mathbf{F} \quad (41)$$

のような分解が可能である。すなわち、この場合は行列  $\mathbf{F}$  と  $[\mathbf{F}_1', \dots, \mathbf{F}_p']'$  のランクが一致しているために、(41)の分解は全く誤差なしにできるわけである。したがって、この高次因子数を上限一杯にとった場合には、階層的主成分分析の因子負荷は、個別の主成分分析の結果として得られる因子負荷と完全に一致する。

ところで、この個別的主成分分析のモデルは、第1節で述べたように、個々の変数と因子との相関の2乗和を最大化するものであるとも言える。このことは、最小化基準が、

$$S_s = \sum_{k=1}^p n_k - \sum_{k=1}^p \text{tr} \mathbf{A}_k' \mathbf{A}_k \quad (42)$$

となることからわかる。一方、階層的主成分分析に関しても、最小化基準は、

$$S = \sum_{k=1}^p n_k - \sum_{k=1}^p \text{tr} \mathbf{A}'_k \mathbf{A}_k \quad (43)$$

のように、全く同一の形であらわすことができる。しかしながら、階層的主成分分析の場合、因子得点が、(23)式によって制約されているから、上述の  $Q = \sum q_k$  の場合は、個別の主成分分析と一致するものの、 $Q$  が小さくなるにしたがって、次第に集合間の関係が強く効くようになってくる。なぜならば、高次因子である  $\mathbf{F}$  の行が、個々の集合の因子に1つずつ対応するわけではなく、複数の集合で重複した部分を説明せざるを得ないからである。

こうして、階層的な主成分分析では、単に各集合の内的整合性を高めるだけでなく、集合間の関係についても配慮がなされ、bandwidth-fidelity dilemma をある程度バランスした解が得られることになる。ただし、そのバランスは、高次因子数の決め方によって、少しずつ異なったものとなり得る。

また、 $q_1 = \dots = q_k = Q$  のとき、全変数を一緒にして同時に主成分分析することと最適化基準としては同一になることも指摘しておこう。

もう1つ、階層的な主成分分析と正準相関分析との関係を論ずる必要がある。第1節でのべたような、主成分分析と正準相関分析との関係をより深く検討することにより、階層的な主成分分析の意味はさらに明らかになるが、それには、定式化の変更を含めてかなりの量の記述が必要になるので、この点については、適用例のところ、数値的な比較を行うにとどめ、理論的な検討は稿を改めることとしたい。

### 6. 3相データを含む場合への拡張

異なる集合に属する変数が同一のものである場合、そのような多集合データは3相データと呼ばれる(図8)。ここでは、その一部が3相データであるようなデータに対して適用可能な形にモデルを拡張することを考える。

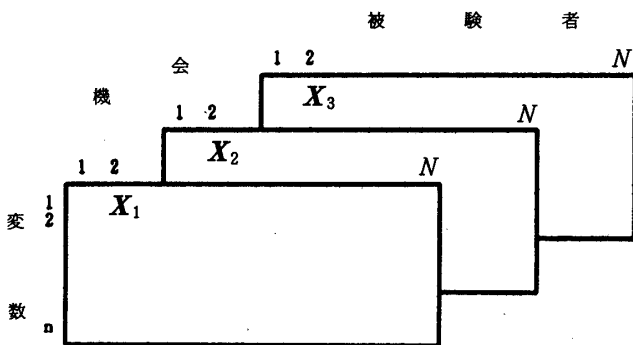


図8 3相データ

もっとも、ここまで述べてきた形のままだでも、3相データへの適用は可能である。ここで新たに考えるのは、

同一の変数に対しては同一の負荷が対応するような形への拡張である。

まず、変数が同一である(より正確には、負荷を同一にしようとしている)一連の集合をブロックと呼ぶことにしよう。ブロックの総数を  $M$  とする。そして、個々の集合がどのブロックに属するかを示す関数を  $\mu(k)$  とする。 $\mu$  は、 $1, \dots, p$  を定義域とし、 $1, \dots, M$  を値域とする整数値関数である。全集合は、 $\mu$  によって類別され、ブロック  $m$  に属する集合の数を  $p_m$  とする。当然、 $\sum_{m=1}^M p_m = p$  である。そこで、基本モデルは、

$$\mathbf{Z}_k = \mathbf{A}_{\mu(k)} \mathbf{C}_k \mathbf{F} + \mathbf{E}_k \quad k = 1, \dots, p \quad (44)$$

となる。制約条件(5)は、そのままであるが、(6)は、維持することが難しく、

$$\frac{1}{p_m} \sum_{\mu(k)=m} \mathbf{C}_k \mathbf{C}'_k = \mathbf{I} \quad m = 1, \dots, M \quad (45)$$

に変える。ここで  $\sum_{\mu(k)=m}$  は  $\mu(k) = m$  となる  $k$  についてのみ和をとることを意味するものとする。したがって、最小化基準は、

$$S_t = \sum_{k=1}^p \text{tr} \frac{1}{N} \mathbf{E}_k \mathbf{E}'_k - \text{tr} \tilde{\mathbf{M}} \left( \frac{1}{N} \mathbf{F} \mathbf{F}' - \mathbf{I} \right) - \sum_{m=1}^M \text{tr} \mathbf{M}_m \left( \frac{1}{p_m} \sum_{\mu(k)=m} \mathbf{C}_k \mathbf{C}'_k - \mathbf{I} \right) \quad (46)$$

となる。そこで、(9)、(10)、(11)は、それぞれ、

$$\frac{1}{N p_m} \sum_{\mu(k)=m} \mathbf{Z}_k \mathbf{F}' \mathbf{C}_k = \mathbf{A}_m \quad m = 1, \dots, M \quad (47)$$

$$\mathbf{A}'_{\mu(k)} \frac{1}{N} \mathbf{Z}_k \mathbf{F}' - \mathbf{A}'_{\mu(k)} \mathbf{A}_{\mu(k)} \mathbf{C}_k = \mathbf{A}_m \mathbf{C}_k \quad k = 1, \dots, p$$

$$\left( \mathbf{A}_m = \frac{\mathbf{M}_m + \mathbf{M}'_m}{2} \right) \quad (48)$$

$$\sum_{k=1}^p \mathbf{C}'_k \mathbf{A}'_{\mu(k)} \mathbf{Z}_k - \sum_{k=1}^p \mathbf{C}'_k \mathbf{A}'_{\mu(k)} \mathbf{A}_{\mu(k)} \mathbf{C}_k \mathbf{F} = \tilde{\mathbf{A}} \mathbf{F}$$

$$\left( \tilde{\mathbf{A}} = \frac{\tilde{\mathbf{M}} + \tilde{\mathbf{M}}'}{2} \right) \quad (49)$$

となる。以下、3節のアルゴリズムの導出過程とほぼ同様に、

$$\mathbf{A}_m^\dagger = \mathbf{A}_m (\mathbf{A}'_m \mathbf{A}_m)^{-\frac{1}{2}} \mathbf{T}_m \quad m = 1, \dots, p$$

$$\mathbf{C}_k^\dagger = \mathbf{T}'_{\mu(k)} (\mathbf{A}'_{\mu(k)} \mathbf{A}_{\mu(k)})^{\frac{1}{2}} \mathbf{C}_k \tilde{\mathbf{T}} \quad k = 1, \dots, p$$

とおいた上で、(17)、(18)に相当する式は、それぞれ、

$$\sum_{\mu(k)=m}^p \{ (\sum_{l=1}^p R_{kl} A_{\mu(l)}^\dagger C_l^\dagger) L^{-2} (\sum_{l=1}^p C_l^\dagger A_{\mu(l)}^\dagger R_{lk}) \} A_m^\dagger = A_m^\dagger L_m \quad m = 1, \dots, M \quad (50)$$

$$\sum_{l=1}^p (A_{\mu(k)}^\dagger R_{kl} A_{\mu(l)}^\dagger) C_l^\dagger = C_k^\dagger \tilde{L} \quad (51)$$

となる。最終的な解は、

$$A_m = \frac{1}{\sqrt{p_m}} A_m^* L_m^{\frac{1}{2}} T_m' \quad m = 1, \dots, M \quad (52)$$

$$C_k = \sqrt{p_m} T_{\mu(k)} L_{\mu(k)}^{-\frac{1}{2}} C_k^* \tilde{T}' \quad k = 1, \dots, p \quad (53)$$

によって得られる。

パラメータの解釈に関しては、(24)は、ほぼそのまま成立し、

$$A_m = \frac{1}{p_m} \sum_{\mu(k)=m} Z_k \tilde{F}_k' \quad m = 1, \dots, M \quad (54)$$

となり、(23)も形のうえではそのまま、

$$C_k = \frac{1}{N} \tilde{F}_k F' \quad (55)$$

であるが、(55)は、もはや相関行列という意味をもたず、共分散行列にすぎない。集合ごとの因子得点について、

$$\frac{1}{N} \tilde{F}_k \tilde{F}_k' = C_k C_k' \neq I \quad (56)$$

だからである。行列  $C_k C_k'$  も因子得点間共分散行列である。

なお、集合全部が3相データであるという図9のようなデータでは、上記のアルゴリズムは、そのまま準3相主成分分析 (Murakami, 1983) のそれに一致する。これはまた、Kroonenberg & de Leeuw (1977) の TUCKALS2 の結果と、パラメータのスケールングを除いて一致した結果を与える。すなわち、階層的な主成分分析の一般化は、準3相主成分分析をその一部として含む一般化であったわけである。

### 7. 適用例

斎藤・村上・若林 (1986) のデータの一部を分析する。このデータは組織イメージと組織活動の関連を検討するためのかなり大規模なものであるが、ここで分析の対象とするのは、表3と表4に示されたような各10項目からなる2つの集合である。集合1は、職業志向性を問うも

表1 変数間の分割相関行列\*

	集 合 1										集 合 2									
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
1	100	-11	44	00	30	02	22	07	29	07	28	19	12	16	46	31	13	23	10	11
2	-11	100	-02	58	-07	50	-13	39	-05	49	-05	14	10	05	-19	-00	-03	-10	10	-04
集 3	44	-02	100	06	47	15	30	10	35	15	41	30	13	19	25	31	23	23	04	05
4	00	58	06	100	-04	52	-13	41	-03	48	02	18	09	07	-10	09	-05	-01	07	-01
合 5	30	-07	47	-04	100	09	49	07	36	09	24	09	09	11	15	14	33	19	04	06
6	02	50	15	52	09	100	-01	46	06	67	06	27	21	10	-09	08	06	00	14	01
1 7	-22	-13	30	-13	49	-01	100	05	33	00	18	08	05	-00	16	11	26	13	-01	04
8	07	39	10	41	07	46	05	100	09	61	-03	18	13	10	-06	-01	01	01	15	04
9	29	-05	35	-03	36	06	33	09	100	11	21	17	10	16	17	15	20	18	02	08
10	07	49	15	48	09	67	00	61	11	100	-01	22	21	13	-08	05	06	01	15	01
1	28	-05	41	02	24	06	18	-03	21	-01	100	52	16	12	33	51	24	18	-01	11
2	19	14	30	18	09	27	08	18	17	22	52	100	34	24	25	40	14	17	15	12
集 3	12	10	13	09	09	21	05	13	10	21	16	34	100	34	24	15	21	23	16	07
4	16	05	19	07	11	10	-00	10	16	13	12	24	34	100	30	17	15	41	22	17
合 5	46	-19	25	-10	15	09	16	-06	17	-08	33	25	24	30	100	46	19	41	17	25
6	31	-00	31	09	14	08	11	-01	15	05	51	40	15	17	46	100	18	26	15	17
2 7	13	-03	23	-05	33	06	26	01	20	06	24	14	21	15	19	18	100	30	10	12
8	23	-10	23	-01	19	00	13	01	18	01	18	17	23	41	41	26	30	100	20	25
9	10	10	04	07	04	14	-01	15	02	15	-01	15	16	22	17	15	10	20	100	29
10	11	-04	05	-01	06	01	04	04	08	01	11	12	07	17	25	17	12	25	29	100

\* 小数点省略



のであり、集合2はライフ・スタイルを問うものである。被験者は、愛知県下の大学生1456名であった。

最初に、表1に全変数間の相関行列を示す。また表2に、幾つかの因子数の場合について、階層的主成分分析のモデルによって説明される分散の大きさ、すなわち、

$$\text{tr } L = \sum_{k=1}^2 \text{tr } L_k \text{ の値を示す。ここで、このデータの}$$

全分散は、 $n_1+n_2 = 20$  だから、例えば、因子数が  $q_1 = q_2 = Q = 2$  の場合、 $7.36/20 \times 100 = 37\%$  の分散が説明されていることになる。以下、他の方法との比較を中心に、階層的主成分分析の結果を示す。因子数の選択は、この観点からなされており、このデータに対して最上のものとは必ずしも言えないかもしれない。

まず、階層的主成分分析(以下、HCAと略記)、正準相関分析(以下、CANCORと略記)、および、集合ごと個別になされる主成分分析(以下、CAと略記)の結果を示す。この3つは、どれも素データの線型結合として合成変数が求められる形のモデルであるという共通

表2 モデルによって説明される分散の大きさ

$q_1$	$q_2$	1	2	3	Q	4	5	6
1	1	4.20	6.23					
1	2		6.38	7.59				
2	1		7.17	8.69				
2	2		7.36	8.92	10.05			
3	2			9.00	10.18	10.93		
2	3			9.03	10.27	11.19		
3	3			9.11	10.39	11.27	11.98	

点をもつ。合成変数の数は、どのモデルにおいても2とし、HCAの高次因子の数は2の場合と4の場合をともに示す。前述のように、高次因子の数が4の場合の因子負荷はCAのそれと一致する。

なお、ここで言う正準相関分析は、Hakstian(1973)におけるCase IV、すなわち、2つの集合の正準変量に同一の直交回転を施した形である。また、示されているのは、各変数と正準変量との相関、すなわち、正準負荷(canonical loading)である。正準ウェイトのかわりに、これを用いる方が解釈に便利であるというのが最近の考え方である(Levine, 1977)。回転は、正準負荷を単純構造化するように、多少拡張されたバリマックス回転が用いられている。

表3に、職業志向性に関する因子負荷(これらのモデルでは、正確には構造ベクトルとよぶべきであるが)を、表4にライフ・スタイルのそれを示した。

表3では、3つの方法の結果は、定性的に見る限り一致していると言ってよい。すなわち、第I因子には偶数番号の項目、第II因子には奇数番号の項目がそれぞれ高く負荷しており、命名するとすれば、第I因子は「挑戦志向」、第II因子は「安定志向」とでもいうことになる。それに対して表4では、HCAとCANCORの結果はほぼ一致するものの、CAの結果は全く異なっている。HCAとCANCORの第I因子は「体制型のライフ・スタイル」、第II因子は「自己実現型のライフ・スタイル」と呼ぶことができよう。CAの第I因子は「公的生活重視のライフ・スタイル」と「私的生活重視のライフ・スタイル」とでもいうことになるだろうか。ここで、HCAとCANCORで、CAと異なる因子構造が得られている

表3 職業志向性の因子負荷行列のモデル間比較 I\*

項 目	HCA(Q=2)			CANCOR			CA		
	I	II	$h^2$	I	II	$h^2$	I	II	$h^2$
1. 安定した会社や勤め先	-02	60	37	15	81	69	00	63	39
2. 困難な仕事に挑戦する機会	75	-16	59	64	-24	47	76	-17	60
3. 高い給与やボーナス	11	65	43	38	81	79	12	74	56
4. 仕事上の責任の重さ	75	-06	57	57	-06	32	77	-09	60
5. 休日多い・勤務時間短い	-00	55	30	17	55	33	02	77	59
6. 能力がためされる機会	81	07	66	84	01	70	81	09	67
7. 仕事の気楽さ	-12	45	22	-01	44	20	-10	68	47
8. 仕事を通じ勉強・成長	71	04	50	55	-09	31	72	12	53
9. 通勤の便利さ	03	50	25	22	49	29	03	66	43
10. 自力で成し遂げる	83	07	69	76	-04	57	84	13	72
2 乗 和	299	157	456	255	212	467	307	249	556

\* 小数点省略

表4 ライフ・スタイルの因子負荷行列のモデル間比較 I\*

項 目	HCA(Q=2)			CANCOR			CA		
	I	II	$h^2$	I	II	$h^2$	I	II	$h^2$
1. お金持ちになること	54	30	38	77	23	64	85	-03	73
2. 仕事で成功すること	29	54	38	46	68	67	73	13	55
3. 自分にあった仕事	20	43	22	22	47	27	30	43	27
4. 親友をもつこと	30	36	22	34	32	22	16	65	45
5. 安定した仕事につく	66	18	46	78	-34	73	49	48	47
6. 大きな組織で高い地位	51	34	37	66	18	47	74	17	58
7. 好きなことをする暇	42	23	23	46	24	27	29	34	20
8. よい結婚相手・幸福な家庭	51	25	32	53	03	28	22	68	51
9. 世の中の不平等なくす	10	31	11	08	28	08	-07	63	40
10. 両親や親類のそばに住む	27	16	10	18	-01	03	04	57	32
2 乗 和	171	108	280	254	113	367	232	217	449

\* 小数点省略

表5 高次因子負荷行列 (HCA)\*

		Q=2		Q=4			
		1	2	1	2	3	4
1. 職業指向性	I	-24	97	100	-01	03	03
2.	II	97	-24	-01	97	20	11
3. ライフ・	I	97	-24	03	20	98	-01
4. スタイル	II	-24	97	04	10	-01	99

\* 小数点省略

表7 因子得点間相関行列 (CANCOR)\*

		因子得点間相関			
		1	2	3	4
1. 職業指向性	I	100	00	-01	44
2.	II	00	100	55	-01
3. ライフ・	I	-01	55	100	00
4. スタイル	II	44	-01	00	100

\* 小数点省略

表6 因子得点間相関行列 (HCA Q=2)\*

		因子得点間相関			
		1	2	3	4
1. 職業指向性	I	100	00	-47	88
2.	II	00	100	88	-47
3. ライフ・	I	-47	88	100	00
4. スタイル	II	88	-47	00	100

\* 小数点省略

表8 因子得点間相関行列 (CA)\*

		因子得点間相関			
		1	2	3	4
1. 職業指向性	I	100	00	06	07
2.	II	00	00	39	21
3. ライフ・	I	06	39	100	00
4. スタイル	II	07	21	00	100

\* 小数点省略

のは、表1から見て取れるように、集合2の変数2と3が、集合1の第II因子に高く相関する変数とやや高い相関をもっていることが原因であろう。

3列目の $h^2$ の欄は共通性、すなわち、各々2つの因子によって説明される、その変数の分散の大きさを表しているが、その和についてみると、CA、CANCOR、HCAの順になっている。ここで、集合内分散の説明を目指すCAが、最大の説明力をもつことは当然であるが、集合間関係に的を絞っているCANCORが、HCAより大きな説明力をもつのは少々意外に思われるかもし

れない。

これは、HCAにおける集合ごとの因子得点が、(2)式によって定義されるように、かなり厳しい制約が課されているためである。

表5は、HCAの高次因子負荷である。Q=2の場合について言えば、集合1の第I因子と集合2の第II因子、集合2の第II因子と集合1の第I因子が相互に高く相関していることが示唆される。このことは、表6の、集合間の因子得点間相関からも明らかである。

一方、Q=4の場合、2つの集合の因子間関係は、

余り認められない。この場合における因子得点間相関行列は、表8のCAに対するそれと一致するが、この表からも同様のことが確認できる。

さらに、CANCORとCAにおける因子得点（合成変量）間相関を表7と表8に示す。CANCORにおいては、HCAの高次因子数2の場合とほぼ同様の傾向である。他方、CAでは前述のように、集合1の第II因子と集合2の第I因子の間に、やや関係が認められるものの、その他はあまり明瞭な関係が見出されない。

次に、多集合因子分析（以下、MSFAと略記）、および2つの集合を全体として因子分析した結果（FAと略記）について述べよう。（後者の方法について、本研究

ではほとんど触れなかったが、その問題点については、村上、1985で述べた。）これらは、ともに共通性の反復推定を含むモデルであるという共通点をもつ。FAについては、因子数が2の場合と3の場合をともに示した。

表9の職業志向性についての因子負荷行列は、先の表2の場合と同様の、2つの因子があらわれている。FAの3因子の場合、第III因子に対する負荷はどの変数もすべて低く、結局第I因子と第II因子が、2因子の場合の2つの因子と定性的には同じパターンを示している。それに対して、表10のライフ・スタイルでは、各場合の結果がかなり異なっている。すなわち、MSFAでは、さきのHCAの高次因子数2の場合、および、CANCOR

表9 職業志向性の因子負荷行列のモデル間比較 II\*

項 目	MSFA			FA(q=2)			FA(q=3)			
	I	II	h <sup>2</sup>	I	II	h <sup>2</sup>	I	II	III	h <sup>2</sup>
1. 安定した会社や勤め先	00	63	40	-01	56	32	-03	42	36	30
2. 困難な仕事に挑戦する機会	68	-16	49	69	-15	50	68	14	-02	49
3. 高い給与やボーナス	12	72	54	11	61	38	10	61	28	46
4. 仕事上の責任の重さ	68	-05	46	70	-05	49	67	-09	07	46
5. 休日多い・勤務時間短い	01	61	37	00	51	26	03	74	03	55
6. 能力がためされる機会	78	09	61	78	08	61	77	08	09	61
7. 仕事の気楽さ	-08	49	24	-10	42	19	-07	62	-01	39
8. 仕事を通じ勉強・成長	63	08	41	67	05	45	63	08	03	41
9. 通勤の便利さ	04	52	27	03	45	20	03	50	14	28
10. 自力で成し遂げる機会	81	11	67	80	08	64	81	11	06	67
2 乗 和	261	186	447	267	136	403	259	179	25	462

\* 小数点省略

表10 ライフ・スタイルの因子負荷行列のモデル間比較 II\*

項 目	MSFA			FA(q=2)			FA(q=3)			
	I	II	h <sup>2</sup>	I	II	h <sup>2</sup>	I	II	III	h <sup>2</sup>
1. お金持ちになること	52	29	36	01	59	35	-02	35	45	32
2. 仕事で成功すること	31	57	43	31	49	34	25	17	51	35
3. 自分にあった仕事	20	44	23	24	34	18	20	08	40	20
4. 親友をもつこと	29	37	22	15	40	19	10	07	48	24
5. 安定した仕事につく	64	18	44	-14	61	40	-20	20	66	52
6. 大きな組織で高い地位	50	34	36	06	58	34	01	21	57	37
7. 好きなことをする暇	38	19	18	01	43	18	00	35	24	18
8. よい結婚相手・幸福な家庭	48	24	29	-02	52	27	-06	19	52	30
9. 世の中の不平等なくす	11	31	11	19	21	08	14	-04	33	13
10. 両親や親類のそばに住む	24	17	08	01	28	08	-02	02	35	12
2 乗 和	159	111	270	23	215	238	18	41	216	275

\* 小数点省略

表 11 因子得点間相関行列 (MSFA)\*

		1	2	3	4
1.	職業指向性 I	100	00	-40	68
2.	II	00	100	70	07
3.	ライフ・I	-40	70	100	00
4.	スタイル II	68	07	00	100

\* 小数点省略

の結果と定性的には等しいパターンとなっているのに対し、FAでは、因子数にかかわらず、全項目が同一の因子に高く負荷する、1因子的な結果となっている。

最後に、MSFAの因子得点間相関行列を表11に示す。結果は、HCAとCANCORの場合に近いが、特に、集合1の第I因子と、集合2の第I因子が逆相関となっているのが興味深い。この部分は、HCA、CANCOR、MSFAでかなり異なる結果となっているが、これは、それぞれのモデルの制約条件によるものである。(HCAでは、異なる集合間の相関行列が直交行列となり、CANCORでは対称行列となる。)

これらは、いずれも何らかの基準にもとづく結果であって、それぞれ少しずつ違った観点からもとのデータを要約するものとなっている。その点で、どの結果が正しいか、という問には意味がない。ただし、HCAとMSFAの方が、集合間の関係についてCAより面白い結果を示しているとは言えよう。

また、このデータでは、HCAとMSFAは、ともに、CAよりは、CANCORの方に近い結果をうみだしたが、これは必ずしもどんなデータでもこうなるというわけではない。ただ、HCAとMSFAの結果が似ていることは、これらが必ずしも同じ基準によるものではないだけに興味深い。

なお、この分析からある程度一般的に言えそうなことは、項目がある程度精選され、かつ標本が大きい場合には、方法によらずほぼ同一の結果が得られるということである。集合1の職業指向性の10項目は、もともと31項目からなる項目群から因子分析によって選択されたものであった。したがって、表1からも見て取れるように、これらは、5項目ずつ、グループ内では相互相関が高く、グループ間では相互相関の低い、かなり明瞭な2群にもともと分かれていた。一方、集合2の10項目は、それぞれ単独で、平均評定値を国際比較する、という目的でつくられており、事前に何らかの構造をなすことは想定されていなかった。この点が、前記の分析において、集合1では、いかなる分析においても、ほぼ同様の因子負荷

のパターンがうみだされ、集合2では、方法によって異なる結果が得られるという相違となってあらわれた、と考えられる。

## 文 献

- Bloxom, B. 1984 Tucker's three mode factor analysis model. In H. G. Hattie, C. W. Snyder Jr., J. A. Hattie, & R. P. McDonald (Eds.) *Research Methods for Multimode Data Analysis*. New York: Prager Publishers.
- Cronbach, L. J. 1970 *Essentials of Psychological Testing*, 3rd ed. New York: Harper & Row.
- Hakstian, A. R. 1973 Procedures for the factor analytic treatment of measures obtained on different occasions. *British Journal of Mathematical and Statistical Psychology*, 26, 219-239.
- Harman, H. H. 1976 *Modern Factor Analysis*, 3rd ed. Chicago: The University of Chicago Press.
- Jöreskog, K. G. 1974 Analyzing psychological data by structural analysis of covariance matrices. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.) *Contemporary Developments in Mathematical Psychology*. San Francisco: Freeman & Company.
- Kroonenberg, P. M. & de Leeuw, J. 1977 TUCKALS2: A principal component analysis of thress mode data. *Research Bulletin* RB 001-77 Leiden: Department of Data Theory, University of Leiden.
- Levine, M. S. 1977 *Canonical Analysis and Factor Comparison*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-006. Biverly Hills: SAGE Publications.
- Murakami, T. 1983 Quasi three-mode principal Component analysis — A method for assessing factor change —. *Behaviormetrika*, 14, 27-48.

村上 隆 1985 多集合データのための探索的因子分析  
名古屋大学教育学部紀要——教育心理学科——, 32、  
75-93.

斎藤和志・村上 隆・若林 満 1986 組織イメージと  
組織活動との関連の分析(2) 東海心理学会第35会  
大会発表論文抄録集 37.

(1986年7月31日 受稿)

## ABSTRACT

### HIERARCHICAL COMPONENT ANALYSIS OF MULTISSET DATA

Takashi MURAKAMI

#### SUMMARY

Let us consider  $p$  sets of variables which correspond to different measurement domains. The data consisting of scores of  $N$  individuals on the variables of all the sets are called multiset data. Let us assume that  $n_k$  variables of each set are essentially multidimensional and several linearly independent composite scores can be constructed from them. Through the analysis of multiset data, one may wish to find out both composite scores for each set and their relationships across sets.

There are two extreme methods for analyzing multiset data. One is the principal component analysis applied to each set separately followed by calculation of correlations between component scores; the other the canonical correlation analysis. The former is concentrated on the internal consistency; the latter on the correlations across sets.

The problem which we confront with is almost equivalent to what Cronbach (1970) called the bandwidth-fidelity dilemma. As there is no optimal unique solution for the problem, one must find a compromise between opposite principles. We will propose a method, *hierarchical component analysis*, by which one can choose a solution which is the most balanced for his objective.

The basic model is written as

$$Z_k = A_k C_k F + E_k \quad k = 1, \dots, P, \quad (1)$$

where  $Z_k$  is  $n_k$  by  $N$  data matrix for set  $k$ ,  $A_k$  is  $n_k$  by  $q_k$  loading matrix for set  $k$ ,  $C_k$  is  $q_k$  by  $Q$  higher order loading matrix for set  $k$ ,  $F$  is  $Q$  by  $N$  component score matrix, and  $E_k$  is  $n_k$  by  $N$  error matrix. Between the number of component of each set  $q_k$  and the number of higher order component  $Q$ , following restrictions are imposed;

$$\max(q_1, \dots, q_p) \leq Q \leq \sum q_k \quad (2)$$

Moreover, the model is subjected to following two constraints;

$$C_k C_k' = I \quad k = 1, \dots, p, \quad (3)$$

and

$$\frac{1}{N} F F' = I \quad (4)$$

The model is equivalent to separate component analysis when  $Q = \sum q_k$ , and to simultaneous component analysis of all the sets when  $Q = q_1 = \dots = q_p$ . When  $Q$  is set to be large, the components for each set become

internally consistent. On the other hand, when  $Q$  is set to be small, correlations of components between sets tend to be large. Therefore, one can seek the balanced solution by changing the number of higher order components.

The alternating least squares algorithm is formulated and tested. And the model is generalized to be applicable to the partially three-mode data, that is, the data obtained on the variable sets some of which include the same variables. Through the generalization, it is shown that the *quasi three-mode component analysis* (Murakami, 1983) is a special case of the model proposed here.