Bulletin of the School of Education,
Nagoya University (Educational Psychology)
1991, Vol. 38, 155－166.

多集合―多群データの階層的主成分分析

村　上　　隆

1．問題の所在

複数の質問項目やテスト項目を含む調査票やテストを用いて得られる変量（項目）×個体のデータ行列を考える（図1のa）。多くの場合，項目は概念的に異なる内容（たとえば，行動と興味，性格，読解力と聴解力等）に関係した複数の集合に分けることができる。こうしたデータを多集合データと呼ぶ（図1のb）。多集合データが収集される目的は，複数の概念間の（相関）関係をあきらかにすることであり，そのためには，各集合において，項目の（重みつき）総和による複数の尺度を構成した上で，集合間の相関をとることになる。この際，主成分分析を個々の集合ごとに実施するか，あるいは，全体を同時的に分析すべきかが問題になる。この場合，個別の分析では集合内の構造のみに関心が集中することになり，全体を同時に分析すると，個々の概念の（独立した）意味があいまいになる（村上，1987）。

また，個体の側も何らかの属性（性別，職業，学歴等）にもとづいて，複数の群に別けることができる場合がある。この形のデータを多群データと呼ぶ（図1のc）。これについても，群ごとに分析を行うべきか，全体を合併して分析すべきか迷うことがある。項目間，あるいは集合間の相関構造は，群によって異なっているかもしれないし，群間の平均差が大きい場合には，それも構造に影響するからである（たとえば，Flury，1988，51-66）。一般的に言って，群を無視すれば，群間の興味ある差が見逃がされることになり，群ごとに分析を行えばパラメータが増えすぎて，その推定精度に疑問が生ずる。

多集合データと多群データを，探索的に主成分分析する方法は，それぞれ，今日までにさまざまなものが提案されている（村上，1990 b）。ここでは，一つのデータ行列が，多集合データでも多群データでもある場合について考える。この形のデータ行列を多集合－多群データ（multiset-multigroup data）と呼ぶ（図2）。

本研究は，そうした多集合－多群データの全体を，それぞれの特徴の差異を見逃がすことなく縮約する方法を開発しようとする。結果的に，それは3相データの階層的主成分分析（村上，1990 a）の一種の一般化となる。

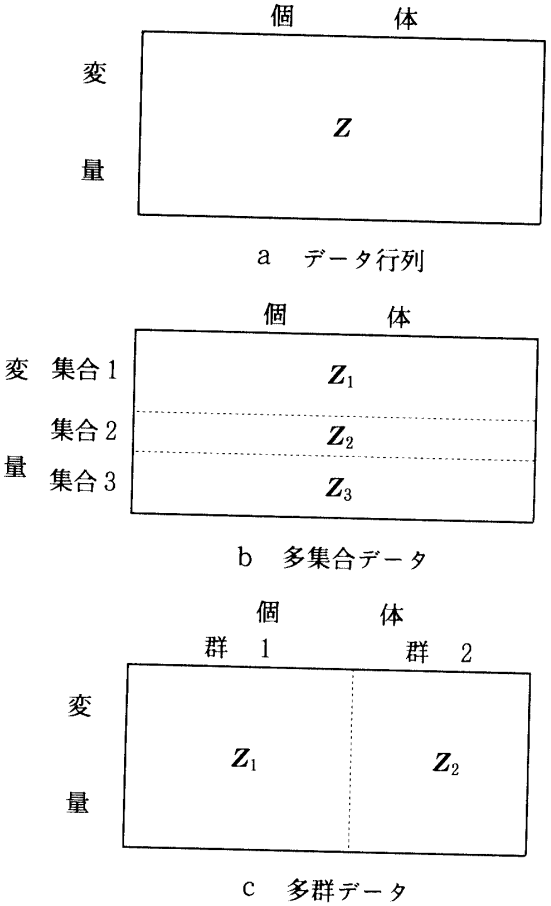


図1　多群データと多集合データ

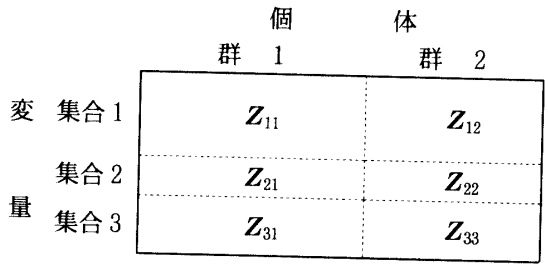


図2　多集合―多群データ

なお，本研究は第一報告であり，基本モデルの導出とアルゴリズムの記述，それに簡単な適用例を中心としている。方法全体の議論と適用上の問題点の詳細については稿を改めたい。

2．基本モデル

直観的導入　集合 k の項目に対する群 s の個体の反応を要素とする $p_k \times N_s$ のデータ行列を $\boldsymbol{X}_{ks}$ と書く。集合 k の項目数が p_k，群 s の個体数が N_s である。

$$p \equiv \sum_k p_k \qquad (2-1)$$

$$N \equiv \sum_s N_s \qquad (2-2)$$

と定義する。データ行列に，次のような変換を行った結果の行列 $\boldsymbol{Z}_{ks}$ を分析の対象とする。

$$\boldsymbol{Z}_{ks} \equiv \boldsymbol{D}_k^{-1/2} (\boldsymbol{X}_{ks} - \bar{\boldsymbol{x}}_{ks}\boldsymbol{1}') \qquad (2-3)$$

ここで，$\bar{\boldsymbol{x}}_{ks}$ は，集合 k の項目の群 s における平均値のベクトル $\boldsymbol{1}$ はすべての要素を1とする N_s 次元のベクトルであり，$\boldsymbol{D}_k$ は，

$$\boldsymbol{S}_{kk} \equiv \sum_s (\boldsymbol{X}_{ks} - \bar{\boldsymbol{x}}_{ks}\boldsymbol{1}')(\boldsymbol{X}_{ks} - \bar{\boldsymbol{x}}_{ks}\boldsymbol{1}')'/N \qquad (2-4)$$

の対角要素，すなわち，項目の級内分散を要素とする対角行列である。そこで，$\boldsymbol{Z}_{ks}$ は各項目を群ごとに中心化（平均値を0），全体として基準化（分散を1）したものである。

データ行列全体は，$p \times N$ の分割行列の形になる。これを集合ごとに，つまり $p_k \times N$ の部分ごとに，個別に主成分分析すると，個々のデータ行列は，$p_k \times q_k$ の負荷行列と，$q_k \times N_s$ の主成分得点行列によって次のように分解される（図3）。

$$\boldsymbol{Z}_{ks} \sim \boldsymbol{A}_k \boldsymbol{F}_{ks} \qquad (2-5)$$

ここで q_k は集合 k の主成分の数である。次に，m 個の $\boldsymbol{F}_{ks}$ を，$\sum_k q_k \times N_s$ の形に並べ直し，このデータ行列に，群別に第2段階の主成分分析する。すると $\boldsymbol{F}_{ks}$ は，$q_k \times r_s$ の2次負荷行列 $\boldsymbol{C}_{ks}$ と，$r_s \times N_s$ の2次主成分得点行列 $\boldsymbol{G}_s$ の積として，

$$\boldsymbol{F}_{ks} \sim \boldsymbol{C}_{ks}\boldsymbol{G}_s \qquad (2-6)$$

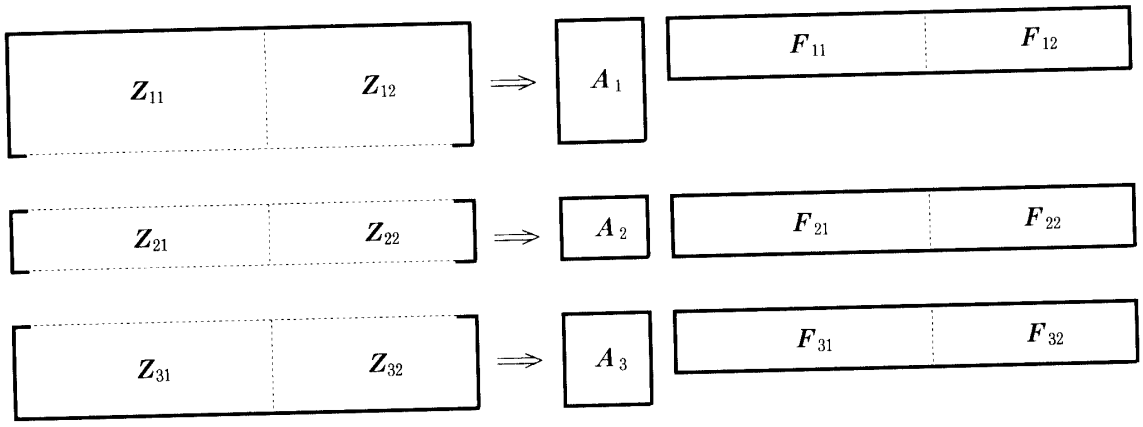


図3　第1段階の分析（集合ごとの個別主成分分析）

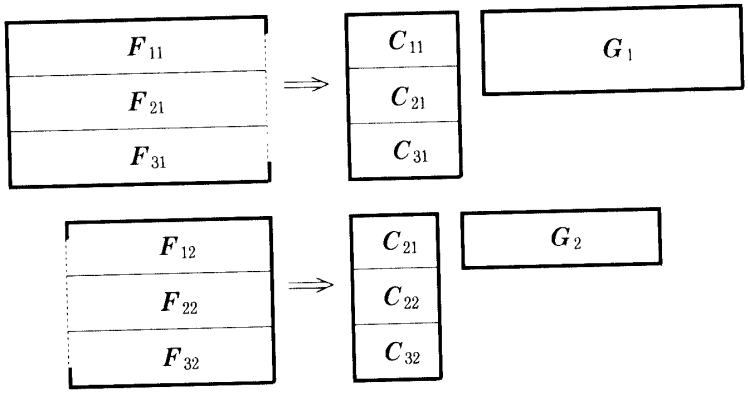


図4　第2段階の分析（主成分得点の群ごとの個別主成分分析）

と分解されることになる（図4）。

次に，(2−6) を (2−5) に「代入」すると，

$$\boldsymbol{Z}_{ks} \sim \boldsymbol{A}_k \boldsymbol{C}_{ks} \boldsymbol{G}_s \qquad (2-7)$$

という2段階の因子分解モデルが得られる（図5）。そこで，

$$\boldsymbol{Z}_{ks} = \boldsymbol{A}_k \boldsymbol{C}_{ks} \boldsymbol{G}_s + \boldsymbol{E}_{ks} \quad k = 1,\cdots, m \quad s = 1,\cdots, g \qquad (2-8)$$

を多集合−多群データの階層的主成分分析の基本モデルと呼ぶ。ただし，$\boldsymbol{E}_{ks}$ は $p_k \times N_s$ の行列であり，任意の k と s について，

$$q_k \leqq p_k \qquad (2-9)$$

$$r_s \leqq N_s \qquad (2-10)$$

$$q_k \leqq \sum_s r_s \qquad (2-11)$$

$$r_s \leqq \sum_k q_k \qquad (2-12)$$

とする。

もし，全ての集合について，群ごとに個別に主成分分析を行うとすれば，$(p \times g)$ 個の負荷行列と主成分得点行列を必要とすることになるが，このモデルによれば，それぞれ，$(p + g)$ 個が必要になるにすぎない。もちろん，それらに加えて $(p \times g)$ 個の2次負荷行列が必要になるわけであるが，(2−9) と (2−10) の不等式の左辺を右辺より，可能な限り小さくとることにより，必要なパラメータの数を大幅に減らすことは可能である。それは，データの縮約的な表現を可能にするとともに，パラメータの推定精度の改善にもつながる。同様の議論は，たとえば，Flury (1988) の pp. 53-54 等参照。

なお，3相データとは，すべての集合の変量が同一であるような多集合データと見ることができる。前述の3相データのための階層的主成分分析は，全集合について共通の負荷行列をもち，群の数 s が1である場合に相当する（図6）。すなわち，

$$\boldsymbol{Z}_k = \boldsymbol{A}\boldsymbol{C}_k\boldsymbol{G} + \boldsymbol{E}_k \qquad k = 1,\cdots, m \qquad (2-13)$$

制約条件と最適化基準　上で述べたのは，あくまでも直観的イメージである。実際に主成分分析を2度行うのではなく，式 (2−8) の基本モデル全体の最小2乗解を求めることを考える。基本モデルには，大きな不定性が存在するので，解を一意に定めるためには，何らかの制約条件が必要になる。ここでは，(広義の) 因子分析に親しんでいるユーザーにとって戸惑いの少ない形の定式化にするために，基本モデルの行列の制約条件を以下のように設定する。まず，2次主成分得点行列は，群ごと

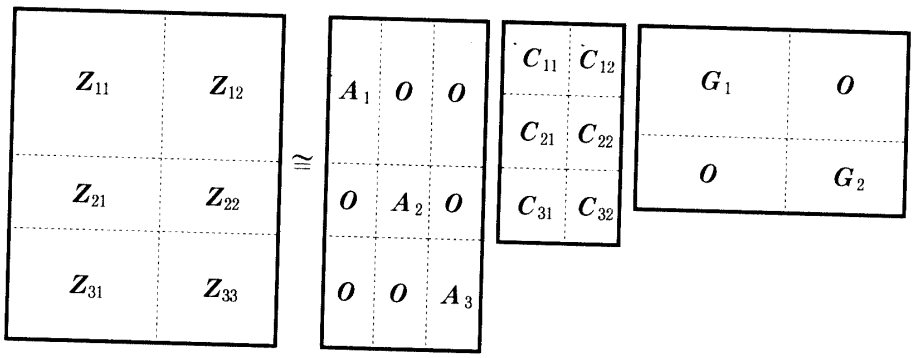


図5　多集合−多群データの階層的主成分分析の基本モデル

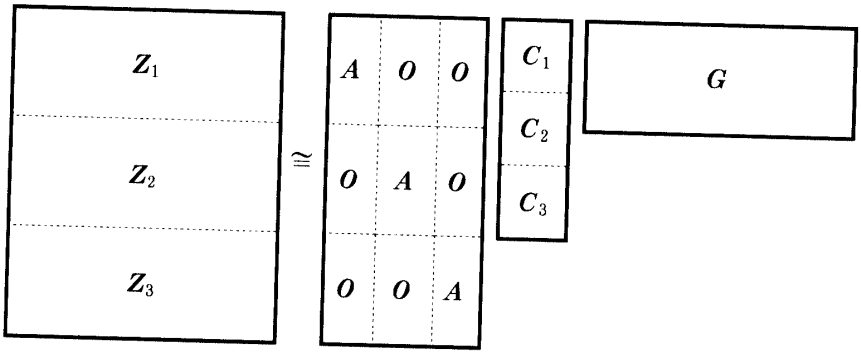


図6　3相データの階層的主成分分析の基本モデル

に基準化，直交化する。すなわち，

$$\boldsymbol{G}_s\boldsymbol{G}_s'/N_s=\mathrm{I}\quad s=1,\cdots,g \qquad (2-14)$$

を満たすものとする。この条件は，$\boldsymbol{G}_s$ の要素自体の解釈を容易にするものであると同時に，$\boldsymbol{A}_k$ や $\boldsymbol{C}_{ks}$ に記述統計測度としての意味を付与するにも役立つ。

つぎに，2次負荷行列 $\boldsymbol{C}_{ks}$ には，次の意味での基準化と直交化を行う。

$$\sum_s n_s\boldsymbol{C}_{ks}\boldsymbol{C}_{ks}'=\mathrm{I}\qquad k=1,\cdots,m \quad (2-15)$$

これも，$\boldsymbol{A}_k$ と $\boldsymbol{C}_{ks}$ 自身の解釈に資するものである。なお，n_s は群 s の個体の割合，

$$n_s\equiv N_s/N\qquad s=1,\cdots,g \quad (2-16)$$

とする。これでもなお，$\boldsymbol{A}_k$，$\boldsymbol{C}_{ks}$，$\boldsymbol{G}_s$ には直交回転に関する不定性が残る。

以上の制約条件の下で，次の最適化基準を達成することを考える。

$$\mathrm{tr}\sum_k\sum_k\boldsymbol{E}_{ks}\boldsymbol{E}_{ks}'/N\ \rightarrow\ \mathrm{Min.} \qquad (2-17)$$

次に，そのための アルゴリズムを導出しよう。

3．アルゴリズム

解の不定性と基本アルゴリズム　前述のように，基本モデルには大きな不定性が存在する。実際，$\boldsymbol{T}_k^*$ と $\boldsymbol{U}_q^*$ を非特異な正方行列とするとき，

$$\Upsilon_k\equiv\boldsymbol{A}_k\boldsymbol{T}_k^* \qquad k=1,\cdots,m$$

$$\Omega_{ks}^*\equiv\boldsymbol{T}_k^{*-1}\boldsymbol{C}_{ks}\boldsymbol{U}_s^* \qquad k=1,\cdots,m;\ s=1,\cdots,g$$

$$\Gamma_s\equiv\boldsymbol{U}_s^{*-1}\boldsymbol{G}_s \qquad s=1,\cdots,g$$

によって定義される，Υ_k，Ω_{ks}^*，Γ_s は，

$$\Upsilon_k\Omega_{ks}^*\Gamma_s=\boldsymbol{A}_k\boldsymbol{C}_{ks}\boldsymbol{G}_s \qquad k=1,\cdots,m;\ s=1,\cdots,g$$

であって，基本モデルの分解と同じ値を生み出す。これを利用して，アルゴリズムを導出しよう。

集合ごとに定義される，$p_k\times q_k$ の正規直交行列を Υ_k とし，群ごとに定義される $N_s\times r_s$ の正規直交行列を Γ_s' とする。すなわち，

$$\Upsilon_k'\Upsilon_k=\boldsymbol{I}\qquad k=1,\cdots,m \quad (3-1)$$

$$\Gamma_s\Gamma_s'=I\qquad s=1,\cdots,g \quad (3-2)$$

である。

つぎに，Ω_{ks}^* を $q_k\times r_s$，Δ_k と Φ_s を，それぞれ，q_k 次，r_s 次の対角行列として，

$$\sum_s\Omega_{ks}^*\Omega_{ks}^{*\prime}=\Delta_k\qquad k=1,\cdots,m \quad (3-3)$$

$$\sum_k\Omega_{ks}^{*\prime}\Omega_{ks}^*=\Theta_s\qquad s=1,\cdots,g \quad (3-4)$$

とする。ただし，Δ_k と Θ_s の（対角）要素には等しいものはないとする。次に，$\boldsymbol{Z}_{ks}$ をさらに，

$$\tilde{\boldsymbol{Z}}_{ks}=\boldsymbol{Z}_{ks}/\sqrt{N} \qquad (3-5)$$

と変換する（$\sum_k\sum_s\mathrm{tr}\,\tilde{\boldsymbol{Z}}_{ks}\tilde{\boldsymbol{Z}}_{ks}'=\sum_k p_k$ に注意）。すると，

(2−17) の基準の必要条件は，

$$\begin{aligned}\sigma\equiv&\sum_k\sum_s\mathrm{tr}\,(\tilde{\boldsymbol{Z}}_{ks}-\Upsilon\Omega_{rs}^*\Gamma_s)\,(\tilde{\boldsymbol{Z}}_{ks}-\Upsilon_k\Upsilon_{ks}^*\Gamma_s)'\\&-\sum_k\mathrm{tr}\boldsymbol{H}_k(\Upsilon_k'\Upsilon_k-\boldsymbol{I})-\sum_k\mathrm{tr}H_k^*(\sum_s\Omega_{ks}^*\Omega_{ks}^{*\prime}-O)\\&-\sum_s\mathrm{tr}\boldsymbol{L}_s(\Gamma_s\Gamma_s'-\boldsymbol{I})-\sum_s\mathrm{tr}\boldsymbol{L}_s^*(\sum_k\Omega_{ks}^*\Omega_{ks}^{*\prime}-\boldsymbol{O})\end{aligned} \qquad (3-6)$$

を Υ_k，$\boldsymbol{C}_{ks}$，Γ_s で編微分してすべての要素を0とおいて得られる停留方程式として得られる。ただし，$\boldsymbol{H}_k$，$\boldsymbol{H}_k^*$，$\boldsymbol{L}_s$，$\boldsymbol{L}_s^*$ は，ラグランジュの定数の行列であって，制約条件の性質から，これらはすべて対称行列である。また，制約条件 (3−3) と (3−4) の性質から，$\boldsymbol{H}_k$ と $\boldsymbol{L}_s^*$ の対角要素は0である。停留方程式を，整理した形で示すと，

$$\sum_s\tilde{\boldsymbol{Z}}_{ks}\Gamma_s'\ \Gamma_s\tilde{\boldsymbol{Z}}_{ks}'\Upsilon_k=\Upsilon_k\Delta_k\qquad k=1,\cdots,m \quad (3-7)$$

$$\sum_k\tilde{\boldsymbol{Z}}_{ks}'\Upsilon_k\Upsilon_k'\tilde{\boldsymbol{Z}}_{ks}\Gamma_s'=\Gamma_s\Theta_s\qquad s=1,\cdots,g \quad (3-8)$$

$$\Omega_{ks}^*=\Upsilon_k'\tilde{\boldsymbol{Z}}_{ks}\Gamma_s' \qquad (3-9)$$

となり，また，最小化基準（残差分散）は，次のようになる。

$$\sigma=\sum_k p_k-\mathrm{tr}\sum_k\Delta_k=\sum_k p_k-\mathrm{tr}\,\Theta_s \quad (3-10)$$

式（3−7）と（3−8）は，それぞれ，行列 $\sum_s\tilde{\boldsymbol{Z}}_{ks}\Gamma_s'\Gamma_s\tilde{\boldsymbol{Z}}_{ks}'$ と $\sum_k\tilde{\boldsymbol{Z}}_{ks}'\Upsilon_k\Upsilon_k'\tilde{\boldsymbol{Z}}_{ks}$ の決定方程式と見ることができる。これは，交互最小2乗法であり，適切な初期値から出発して，(3−7) と (3−8) を交互に反復する（その際，(3−10)から，常に最大 q_k 番目，r_k 番目までの固有値と固有ベクトルをとる）ことによって解が得られる。

実際的なアルゴリズム　式（3−8）における固有値計算の対象となる行列の次数は N_s であり，これは，通

常の応用場面においては，扱い難い大きさになる。そこで，もっと小さい行列を対象にして実行できるアルゴリズムを考える。まず，

$$\Omega_{ks} \equiv \Omega^{*}_{ks}\,\Theta_s^{-1/2} \qquad k=1,\cdots,m \quad (3-11)$$
$$s=1,\cdots,g$$

と定義する。これは，

$$\sum_k \Omega'_{ks}\,\Omega_{ks}=I \qquad s=1,\cdots,g \quad (3-12)$$

を満たす。つぎに，

$$P_{kls} \equiv \tilde{Z}_{ks}\tilde{Z}'_{ks} \qquad k=1,\cdots,m \quad (3-13)$$
$$l=1,\cdots,m$$
$$s=1,\cdots,g$$

した上で若干の演算を行うと，(3—7) ～ (3−10) は，

$$\sum_l \Gamma_k' P_{kls}\Gamma_l\,\Omega_{ls}=\Omega_{ks}\Theta_s \quad k=1,\cdots,m \quad (3-14)$$
$$s=1,\cdots,g$$

$$\{\sum_s(\sum_l P_{kls}\Gamma_l\Omega_{ls})\,\Theta_s^{-1}(\sum_l \Omega'_{ls}\,\Gamma'_l P_{kls})\}\,\Gamma_k=\Gamma_k\Delta_k$$
$$k=1,\cdots,m \quad (3-15)$$

$$\Gamma_s=\Theta_s^{-1/2}\sum_k \Omega'_{ks}\Gamma'_k\tilde{Z}_{ks}$$
$$s=1,\cdots,g \quad (3-16)$$

と変形することができる。反復は (3−14) と (3−15) の間で行えばよく，固有値計算は，それぞれ $\sum_k q_k$ 次，p_k 次の行列について行えばよい。

実際の計算と分析の完成　まず，

$$S_{kls} \equiv P_{kls}/n_s \qquad k=1,\cdots,m \quad (3-17)$$
$$l=1,\cdots,m$$
$$s=1,\cdots,g$$

と定義すると，これは，

$$S_{kls}=Z_{ks}Z'_{ls}/N_s \qquad k=1,\cdots,m \quad (3-18)$$
$$l=1,\cdots,m$$
$$s=1,\cdots,g$$

すなわち，変換されたデータについて，群ごとに算出された，集合間の項目間共分散行列である（図7）。

S_{kls} を用いると，(3−14) と (3−15) は，

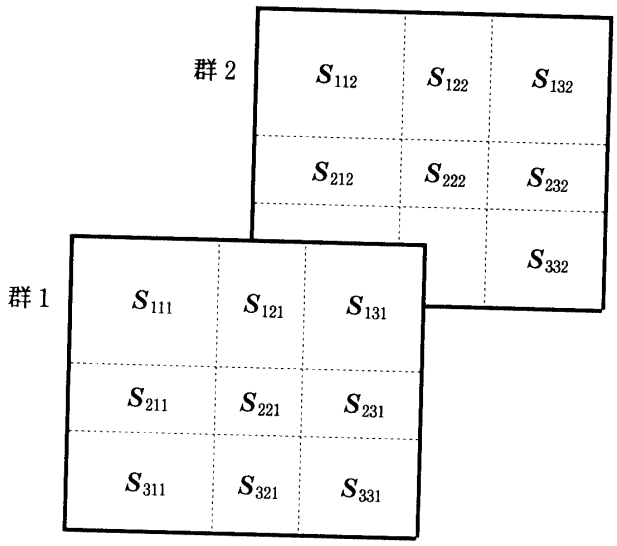


図7　群ごとの変量間共分散行列

$$\sum_l n_s\Gamma'_k S_{kls}\Gamma_l\,\Omega_{ls}=\Omega_{ks}\Theta_s$$
$$k=1,\cdots,m \quad (3-19)$$
$$s=1,\cdots,g$$

$$\{\sum_s n_s^2(\sum_l S_{kls}\Gamma_l\Omega_{ls})\,\Theta_s^{-1}\,(\sum_l \Omega'_{ls}\,\Gamma'_l S_{kls})\}\,\Gamma_k=\Gamma_k\Delta_k$$
$$k=1,\cdots,m \quad (3-20)$$

なお，初期値としては，各集合ごとの項目間共分散行列の，全群にわたる平均行列，

$$S_{kk} \equiv \sum_s n_s S_{kks} \qquad k=1,\cdots,m \quad (3-21)$$

の大小順に q_k 番目までの固有値に対応する固有ベクトルをとるのが適切であろう。

得られた Γ_k, Δ_k, Γ_s, Θ_s から，最終的な解を求めるためには，

$$A_k=\Gamma_k\Delta_k^{1/2}T'_k \qquad k=1,\cdots,m \quad (3-22)$$

$$C_{ks}=T_k\Delta_k^{-1/2}\Omega_{ks}\Theta_s^{1/2}U'_s/\sqrt{n_s}$$
$$k=1,\cdots,m \quad (3-23)$$
$$s=1,\cdots,g$$

$$G_s=(\sum_k C'_{ks}A'_kA_kC_{ks})^{-1}\sum_k C'_{ks}A'_kZ_{ks}$$
$$s=1,\cdots,g \quad (3-24)$$

のようにすればよい。T_k と U_s は任意の直交行列であり，これらは，A_k と C_{ks} が単純構造に近づくように決定する。

4．結果の解釈等

通常の「因子分析」における，bilinear な関係に帰着させた解釈を考える。第一に，

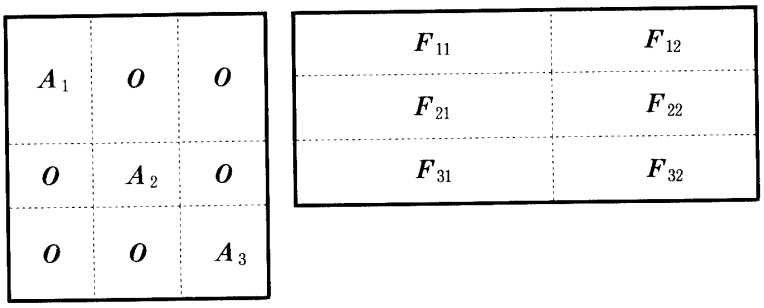
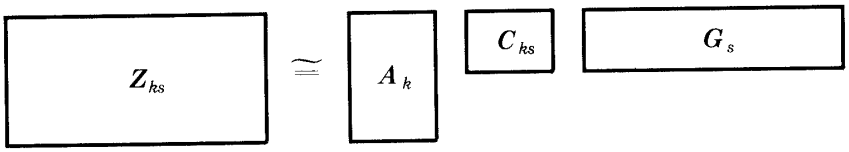


図8　モデルの第一の解釈



階層的主成分分析

図9　第二の解釈

$$F_{ks} \equiv C_{ks} G_s \qquad k = 1, \cdots, m \quad (4-1)$$
$$s = 1, \cdots, g$$

を1次主成分得点と呼ぶとすれば，基本モデルは，各集合，各群ごとに通常の因子モデルをあてはめたものと見ることもできる（図8）。さらに，A_kはこの各行と個々の項目との相関行列，すなわち，

$$A_k = \sum_s Z_{ks} F'_{ks} / N \qquad k = 1, \cdots, m \quad (4-2)$$

であることが証明できる。この性質は，(2－5) の観点から A_k を解釈する助けとなる。また，2次負荷行列 C_{ks} も，G_k と F_{ks} の群ごとの共分散行列と解釈できる。

第二に，(2－7) にもとづいて，A_k を集合 k の項目の負荷ベクトルの，G_s を群 s の個体の「因子得点」ベクトルの，それぞれ「プール」とみなし，C_{ks} によって定義されるそれらの適切な組み合わせが，集合 k の群 s における「因子モデル」であるとみなす（図9）。

さらに，G_s は，V_k を $p_k \times q_k$ の，W_{ks} を $q_k \times r_s$ のそれぞれ重み行列として，

$$G_s = \sum_k W'_{ks} V'_k Z_{ks} \qquad s = 1, \cdots, g \quad (4-3)$$

という合成変量の行列とみることができる。そこで，他の群や集合とは独立に定義される $V'_k Z_{ks}$（1次合成変量，村上，1990a，pp. 85-88）によって，集合間の相関関係とその群間の差異を検討することも考えられる。これを変換前の素データの行列 X_{ks} にもとづいて計算すれば，群間の平均差を検討することも可能である。これらについては，村上（1990）における3相データの場合と同様の定式化となる。

なお，主成分の数に関する「境界条件」である，すべての k について，$q_k = \sum_s r_s$ の場合は，群ごとの群ごとの個別の主成分分析と一致し，$\sum_k q_k = r_s$ の場合は，集合ごとの個別の主成分分析の結果と一致する（ただし，事前にデータは (2－3) にしたがって変換してあるものとする）。そうした意味で，この論文で述べた方法は，通常の主成分分析と，多群データ，多集合データの階層的主成分分析（村上，1990b）の自然な一般化であると見ることができる。

5．適用例

村上（1985）のデータを再分析する。被験者は私立C大学の心理学科の学生132名（男子77名，女子55名）であった。被験者は，5～6名のグループに分かれ，表1の12の尺度の上で相互に評定しあう。同時に自分自身の評定も行う。被験者ごとに他者からの評定値を平均したものを集合1，自分自身の評定値を集合2とする。なお，ここで採用されたグループは，心理学の基礎実験の「班」であって，被験者は調査時点までに，約3箇月間，一緒に実験を行ってきており，相互の熟知度はかなり高い。調査は，1984年7月上旬になされた。

表1は，男女を込みにした単なる多集合データとして，それを集合ごとに個別に主成分分析した結果として得られた負荷行列である。主成分の数は3としており，第Ⅰの主成分は「安定性」，第Ⅱの主成分は「成熟性」，第Ⅲの主成分は「向社会性」と名づけることができよう。な

表1　個別主成分分析による負荷行列（男女込みのデータ）

	他者評定				自己評定			
	Ⅰ	Ⅱ	Ⅲ	平方和	Ⅰ	Ⅱ	Ⅲ	平方和
1）子どもっぽい—おとなっぽい	.22	.78	−.10	.67	.02	.81	.14	.68
2）慎重な—軽率な	−.87	−.34	.02	.87	−.68	−.31	−.00	.56
3）感情的—理性的	.83	.05	.02	.69	.56	.50	.01	.56
4）面倒見のよい—面倒見の悪い	−.31	−.25	−.74	.71	−.21	−.07	−.75	.61
5）説得力のある—説得力のない	−.22	−.74	−.33	.70	−.00	−.43	−.59	.54
6）頼りない—しっかりした	.33	.68	.38	.72	.29	.51	.40	.51
7）融通のきく—融通のきかない	.15	.12	−.81	.69	.07	−.17	−.70	.53
8）未熟な—成熟した	.23	.85	.09	.77	.08	.72	.34	.65
9）思いやりのある—思いやりのない	−.48	−.02	−.74	.78	−.43	.01	−.76	.75
10）空想家—現実家	.04	.61	−.12	.39	.22	.57	−.07	.38
11）常識のある—非常識な	−.75	−.34	−.30	.77	−.62	.18	−.45	.61
12）情緒の安定している—情緒不安定である	−.52	−.27	−.27	.81	−.83	−.14	−.20	.75
平方和	3.28	3.10	2.20	8.58	2.22	2.38	2.52	7.12
寄与率	.273	.258	.183	.715	.185	.198	.210	.594

表2　主成分得点間相関行列

	他者評定			自己評定		
	Ⅰ	Ⅱ	Ⅲ	Ⅰ	Ⅱ	Ⅲ
他者評定						
Ⅰ	1.00	.00	.00	.21	.00	−.29
Ⅱ	.00	1.00	.00	−.13	.33	.11
Ⅲ	.00	.00	1.00	.05	−.19	.07
自己評定						
Ⅰ	.21	−.13	.05	1.00	.00	.00
Ⅱ	.00	.33	−.19	.00	1.00	.00
Ⅲ	−.29	.11	.07	.00	.00	1.00

お，このデータでは，2つの集合が同じ変量からなっているから，3相データとしての扱いも可能であるが，ここでは，一般の多集合データとして扱う。ただし，本研究の方法を，3相データに適用可能なように拡張することは容易である（村上，1986）。

表2は，2つの集合の主成分間の相関行列を算出したものである。主成分分析は個別になされているから，主成分は集合ごとに分散1で直交している。第Ⅰと第Ⅱの主成分については，一応自己評定と他者評定が対応している。しかしながら，第Ⅲの主成分については，自己評定と他者評定が対応せず，しかも自己評定は，他者評定の第Ⅰ主成分と，はっきりした逆相関となっている。すなわち，自分自身を「向社会的」と評定する人は，他者からは「不安定」と評定される傾向がある。

さて，こうした傾向が男女を通じて共通して見られるものであるか否かを検討しよう。しかし，$N_1 = 77$，$N_2 = 55$ という個体数は，個別の主成分を行うことをためらわせる。そこで，男子を群1，女子を群2として，本研究の方法を適用してみよう。

まず，主成分の数を決める必要がある。1次主成分の数（q_1 と q_2）は，先の分析に合わせてそれぞれ3とする。2次主成分の数は，ここでは一応 $r_1 = r_2$ とし，それを4，5，6の3段階で変えてみる。表3は，グループ別に集合ごとの，データの分散の n_s 倍（これらの総和は $\sum_k p_k = 12$ となる），説明される分散の n_s 倍（これらのグループごとの総和は $\mathrm{tr}\,\Theta_s$ に等しい）および，前者の後者に対する割合（寄与率）を，それぞれの場合について示したものである。この表から，適切な主成分の数を決定することは困難である。しかし，$r_1 = r_2 = 4$ では，男子の自己評定の説明力がかなり小さく，一方，$r_1 = r_2 = 6$ は，(2−12) で等号が成立する境界条件であり，この方法の特徴が示されない。そこで，ここでは，$r_1 = r_2 = 5$ を選択して，結果を示すことにする。なお，$r_1 \neq r_2$ とすることも可能であることを念のために指摘

表3 寄与の大きさ（$q_1 = q_2 = 2$）

	データの分散 $\times n_s$	説明される分散 $\times n_s$ $r_1 = r_2 = 4$	$r_1 = r_2 = 5$	$r_1 = r_2 = 6$
男　子				
他者評定	7.36	4.91 (0.67)	5.17 (0.70)	5.29 (0.72)
自己評定	6.95	3.38 (0.49)	3.70 (0.53)	4.00 (0.58)
女　子				
他者評定	4.63	2.91 (0.63)	3.07 (0.66)	3.22 (0.70)
自己評定	5.05	2.70 (0.54)	3.00 (0.60)	3.11 (0.62)

表4 1次負荷行列 $\mathbf{A}_1$ と $\mathbf{A}_2$（$r_1 = r_2 = 5$ の場合）

	他者評定				自己評定			
	I	II	III	平方和	I	II	III	平方和
1) 子どもっぽい—おとなっぽい	.20	**.80**	−.09	.68	.26	**.77**	.06	.66
2) 慎重な—軽率な	**−.86**	−.35	−.06	.86	**−.68**	−.22	−.14	.53
3) 感情的—理性的	**.77**	.13	.09	.62	**.68**	.32	.08	.57
4) 面倒見のよい—面倒見の悪い	−.28	−.24	**−.74**	.68	−.12	−.18	**−.74**	.59
5) 説得力のある—説得力のない	−.26	**−.69**	−.28	.62	.00	**−.55**	−.48	.53
6) 頼りない—しっかりした	.32	**.71**	.29	.69	.36	.49	.37	.51
7) 融通のきく—融通のきかない	.13	.13	**−.81**	.70	.14	.31	**−.28**	.45
8) 未熟な—成熟した	.18	**.87**	.05	.79	.21	**.72**	.22	.62
9) 思いやりのある—思いやりのない	−.43	−.05	**−.75**	.75	−.27	−.03	**−.82**	.74
10) 空想家—現実家	.14	**.53**	−.10	.31	.40	.42	.10	.35
11) 常識のある—非常識な	**−.77**	−.33	−.26	.77	−.30	.13	**−.66**	.54
12) 情緒の安定している—情緒不安定である	**−.79**	−.30	−.22	.78	**−.67**	−.04	−.42	.63
平　方　和	3.08	3.06	2.10	8.25	1.98	2.12	2.61	6.70
寄　与　率	.255	.175	.257	.687	.165	.176	.217	.558

表5 2 次 負 荷 行 列

	男子					女子				
	1	2	3	4	5	1	2	3	4	5
他者評定										
I	**1.05**	.15	−.24	−.12	−.10	**.78**	.19	−.13	.13	−.20
II	.03	**.95**	.21	.05	−.02	.13	−.11	**.86**	−.09	.36
III	−.05	.05	−.06	**.97**	.13	.12	−.03	−.06	**1.01**	.02
自己評定										
I	**.72**	−.24	**.53**	.09	−.11	.24	**1.02**	−.20	−.05	.17
II	−.07	.27	**.76**	−.08	.05	−.37	−.03	**1.14**	−.04	−.21
III	−.12	−.02	.03	.14	**1.03**	−.19	.13	−.02	−.03	**.90**
平 方 和	1.64	0.96	1.06	0.99	0.96	0.87	1.27	2.10	1.06	1.05

しておく。

表4は，2つの集合の1次負荷行列である。表1と比較することによって明らかなように，これらは，個別主成分分析の場合と非常によく似ている。3つの主成分はこの順に，「安定性」，「成熟性」，「向社会性」と呼ぶことにする。

次に，表5は2次負荷行列を，群（男女）別に示したものである。これは興味深い形を示している。男子の第1の2次主成分は，他者評定と自己評定のそれぞれ第Iの1次主成分が高く負荷しており，「安定性」の他者評定と自己評定の間の相関が高いことを示唆している。第2の2次主成分は，他者評定の第IIの1次主成分が，第

3の2次主成分は，自己評定の第Ⅰと第Ⅱの1次主成分が高く負荷している。すなわち，「成熟性」の他者評定と自己評定の間の相関が高くないこと，および（男子においては，「安定性」と「成熟性」が正に相関していることを示している。第4と第5の2次主成分は，それぞれ，第Ⅲの1次主成分の他者評定と自己評定が高く負荷し，「向社会性」も他者評定と自己評定が対応しないことを示している。

女子においては，第3の2次主成分が，第Ⅱの1次主成分の他者評定と自己評定の両方が高く負荷しているが，残りの2次主成分は，1次主成分の他者評定か自己評定が1つずつ高く負荷している。すなわち，女子では，男子と異なり，他者と自己の「成熟性」の評定の間の一致度が高く，「安定性」の対応はよくないことを示している。「向社会性」については，男女とも自己評定と他者評定が一致していない。この点は，広い意味では表2の結果が確認されているが，相関の大小に関しては，男女差があることがわかる。

1次主成分間の関係をよりはっきりと示すために，各群の1次主成分間の共分散行列，すなわち，

$$\tilde{S}_{kls} = C_{ks} C'_{ls} \qquad \begin{matrix} k = 1, 2 \\ l = 1, 2 \\ s = 1, 2 \end{matrix} \qquad (5-1)$$

を算出する。表6－1に男子，表6－2に女子の結果を示した。他者評定と自己評定の間の共分散の大きさからは，2次負荷行列に示唆された事実が確認できる。表2で示された，「向社会性」の自己評定と「安定性」の逆相関は男女とも見られるが，これは，女子の方により顕著であることがわかる。

これ以外にも，若干の興味深い傾向が見られる。しかし，たまたま選択した主成分数に関する結果をこれ以上詮索することは止めよう。集合間の相関については，1次主成分によるより，個々の集合の変量だけの1次結合である1次合成変量による方がよい。前述のように，それによれば，3つの次元上での男女間の平均差も検討できる。しかし，それらはもはや本研究の範囲を越えている。

表6－1　1次主成分間共分散行列（男子，$r_1 = r_2 = 5$）

	他者評定			自己評定		
	Ⅰ	Ⅱ	Ⅲ	Ⅰ	Ⅱ	Ⅲ
他者評定						
Ⅰ	1.20	.12	−.16	.59	−.21	−.25
Ⅱ	.12	.95	.08	−.09	.41	−.03
Ⅲ	−.16	.08	.97	−.01	−.10	.26
自己評定						
Ⅰ	.59	−.09	−.01	.87	.27	−.17
Ⅱ	−.21	.41	−.10	.27	.66	.07
Ⅲ	−.25	−.03	.26	−.17	.07	1.10

表6－2　1次主成分間共分散行列（女子，$r_1 = r_2 = 5$）

	他者評定			自己評定		
	Ⅰ	Ⅱ	Ⅲ	Ⅰ	Ⅱ	Ⅲ
他者評定						
Ⅰ	.71	−.17	.22	.37	−.40	−.31
Ⅱ	−.17	1.07	−.11	−.50	.88	.26
Ⅲ	.22	−.11	1.05	−.04	−.17	.02
自己評定						
Ⅰ	.37	−.50	−.04	1.18	−.38	.24
Ⅱ	−.40	.88	−.17	−.38	1.47	−.10
Ⅲ	−.31	.26	.02	.24	−.10	.86

6．おわりに

最後に，この方法の心理測定における意味を述べておこう。もし，質問項目や問題項目の「意味」が事前に明らかであるならば，それらの項目間に因果的構造の存在を想定し，その構造を明らかにするという問いには意味がある。また，その場合には，各集合に含まれる項目の選択に理論的な必然性があるはずであり，項目間，集合間に共有される分散を最も節約的に説明するという目的は正当化されよう。しかしながら，少なくとも，心理・教育測定の分野ではそうしたことは稀であり，項目の「意味」は，分析の結果として得られた構造から明らかになると考えざるを得ない場合が多い。つまり個々の項目の「意味」と，全体の構造の意味は，同時平行的に明らかにされていく。その過程において，項目の集合も，個体の群も，必ずしも通常のデータ解析における外部基準として機能するわけではなく，（通常あいまいな形で存在する）理論にもとづく結果の解釈可能性という意味での，測定の妥当性を検討するために用いられる（村上，1989）。こうした心理・教育測定尺度の構成過程において，ここで提案した方法は一定の有用性をもつであろう。

文　　献

Flury, B. 1988 *Common Principal Components & Related Multivariate Models.* New York : Wiley.

Kroonenberg, P. M. & de Leeuw, J. 1980 Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika,* **45**, 69-97.

村上　隆　1985　変量間の関係の構造を探る―因子分析　海保博之（編著）心理・教育データの解析法10講　基礎編　福村出版

村上　隆　1986　多集合データのための階層的主成分分析　名古屋大学教育学部紀要―教育心理学科，**33**，35-48.

村上　隆　1987　複数の変数集合の主成分と正準変量　名古屋大学教育学部紀要―教育心理学科，**34**，235-253.

村上　隆　1989　心理測定の理論と家族的類似の概念　名古屋大学教育学部紀要―教育心理学科，**36**，149-156.

村上　隆　1990a　3相データの階層的主成分分析　柳井晴夫・岩坪秀一・石塚智一（編）人間行動の計量分析　東京大学出版会，71-94.

村上　隆　1990b　多群データ，多集合データ，多層データの主成分分析について　行動計量学，**18**，28-40.

（1991年8月31日　受稿）

ABSTRACT

Hierarchical Principal Component Analysis of Multiset-Multigroup Data

Takashi MURAKAMI

Consider a data matrix $\boldsymbol{Z}$, which is of the form : variables $\times$ subjects, and it can be partitioned into m sets of rows and g groups of columns. $\boldsymbol{Z}$ can be seen as an $m \times g$ super matrix, the elements of which are $p_k \times N_s$ matrices $\boldsymbol{Z}_{ks}$'s ($k = 1, \cdots, m$; $s = 1, \cdots, g$). Assume that rows of $\boldsymbol{Z}_{ks}$ have zero mean for each group, and have unit variance across all the groups. Let us call the data which can be arranged as $\boldsymbol{Z}$ multiset-multigroup one.

This paper proposed a method for component analysis of multiset-multigroup data. The basic model is written as

$$\boldsymbol{Z}_{ks} = \boldsymbol{A}_k \boldsymbol{C}_{ks} \boldsymbol{G}_s + \boldsymbol{E}_{ks}, \qquad k = 1, \cdots, m, \quad s = 1, \cdots, g ; \qquad (1)$$

where $\boldsymbol{A}_k$ is the $p_k \times q_k$ first order loading matrix for variables of k-th sets, $\boldsymbol{C}_{ks}$ is the $q_k \times r_s$ second order loading matrix of s-th group on k-th first order components which are defined as

$$\boldsymbol{F}_{ks} \equiv \boldsymbol{C}_{ks} \boldsymbol{G}_s, \qquad (2)$$

and $\boldsymbol{G}_s$ is the $r_s \times N_s$ second order component score matrix for s-th group ; $\boldsymbol{E}_{ks}$ denotes the $p_k \times N_s$ residual matrix. The basic model is a natural extension of Kroonenberg & de Leeuw (1977)'s *TUCKER 2* model for three-mode data which can be written as

$$\boldsymbol{Z}_k = \boldsymbol{A}\, \boldsymbol{C}_k \boldsymbol{G} + \boldsymbol{E}_k, \qquad k = 1, \cdots, m,$$

The criterion to be minimized is

$$\sigma = \mathrm{tr} \sum_k \sum_s \boldsymbol{E}_{ks} \boldsymbol{E}_{ks}', \qquad (3)$$

under the constraints

$$\sum_s N_s \boldsymbol{C}_{ks} \boldsymbol{C}_{ks}' / \sum_s N_s = \boldsymbol{I}, \qquad k = 1, \cdots, m, \qquad (4)$$

and,

$$\boldsymbol{G}_s \boldsymbol{G}_s' / N_s = \boldsymbol{I}, \qquad s = 1, \cdots, g, \qquad (5)$$

An alternating least squares algorithm, which is also a slight modification of *TUCKALS 2* solving *TUCKER 2* problem, is derived and it is adapted to handle the data with large N_s's. One of the most distinct feature of the output of this method is that the first ordr loading matrices can be interpreted as correlation matrices between variables and first order components such as

$$\boldsymbol{A}_k = \sum_s \boldsymbol{Z}_{ks} \boldsymbol{F}_{ks} / \sum_s N_s, \qquad k = 1, \cdots, m, \qquad (6)$$

This hierarchical component model is not only able to explain the data more parsimoniously than individual analysis of each element matrix but also more sensitive to the group differences of loadings than analysis of all groups as a whole. An application for the data with two sets—Peer and Self ratings, and two groups—males and females was demonstrated as an illustrative example.