Bulletin of the School of Education, Nagoya University (Psychology)
1997, Vol. 44, 87 − 105.

カテゴリカル・データの非計量的主成分分析の応用

村　上　　　隆

1．序

主成分分析（principal components analysis）にはさまざまな定式化があることが知られているが（たとえば，村上，1990），ここでは，個体×変数（$n \times m$）のデータ行列 Z に対して，

$$f(F,A) = n^{-1} \| Z - FA' \|^2 \tag{1}$$

を，$n^{-1}F'F = I$ の制約条件の下で最小化するような，$n \times r$ の主成分得点行列 F と，$m \times r$ の主成分負荷量行列 A を求めるという意味で考える。ただし，Z は列（変数）ごとに，平均値が 0，標準偏差が 1 となるように変換されているものとする。（このことを，以下，Z は列ごとに中心化，標準化されていると言う。）これはいわば，主成分分析の因子分析的用法とでも呼ぶべきものであるが，この意味の主成分分析（あるいは，いわゆる因子分析主成分解）は，質問紙調査を用いた心理学等の研究ではよく用いられている。その理由は，主に次の 3 点にあるものと考えられる。

1）多くの場合，m 個の質問項目への回答は，（4～7段階程度の）順序のついたカテゴリーでなされ，それらは通常，数値として扱ってよいと考えられている。

2）主成分得点行列 F には，上記の制約条件の形からして，直交回転に関する自由度があるとみなされる。その結果，主成分負荷行列 A に，単純構造に向けての回転（たとえば，varimax 回転）がなされる。このことは，主成分の数 r が多くなりがちなこの種の研究においては，解釈上極めて有利である。

3）項目数が大きい場合が多いから，結果が（主因子法等による）探索的因子分析と著しく異なることは希であり，かつ，共通性の推定，因子得点の不定性等の問題を免がれることができる。

心理学の研究では，単なる多変数のオーディネイション（ordination），すなわち，「元のデータのもつ構造を何らかの意味で保存するように，個体あるいは変数を低次元の空間上の点として付置する」（鷲尾・大橋，1989; p.34）だけでは意義が乏しく，次元の（自然言語による）解釈は必須であることも付け加えておこう。

しかしながら，1）の想定が妥当か否かについては，しばしば疑義が呈されるし，項目に対する反応が（項目ごとに異なる）順序のつかないカテゴリーによってなされる（方が自然な）場合もあり，その場合主成分分析の適用は不可能である。

そうした場合，（アイテム・カテゴリー型データに対する）数量化Ⅲ類（岩坪，1987; p.8），または，それと等価な方法である対応分析（correspondence analysis, 例えば，Greenacre, 1984），同質性分析（Homogeneity analysis, 例えば，Gifi, 1990），双対尺度法（Dual scaling, 例えば，Nishisato, 1994）等が用いられる。しかしながら，これには次のような問題がある。

1）個々のカテゴリーを，事実上質問項目と切り離して，「延べ」で扱わなければならない。項目が本来もっている内容的な「まとまり」が分析において失われる。

2）次元の数を増やすと，同じ次元数の数量がすべてのカテゴリーに付随することになり，結果は膨大なものになりがちである。

3）通常，回転が行われない。（ただし，アイテム・カテゴリー型データでは，原理的には可能である。たとえば，Kiers, 1991）そのため，グラフィカルな表示が不可能な 3 次元を超える解の解釈は困難である。

こうした問題に対処するためには，変数ごとに各カテゴリーに数量を割り当て，その結果を主成分分析することが考えられる。

従来，この趣旨の方法は数多く提案されている（例えば，Takane, Young, & De Leeuw, 1978; Saito &

The author is obliged to Dr. H.A.L. Kiers and Prof. J.M.F. ten Berge, University of Groningen, for their helpful comments and pleasant discussions with them in the preparatory stage of this paper. また，適用例 2 のデータの使用を許可していただいた関係各位に，心から謝意を表します。

Ohtsu, 1988; Gifi, 1990; 大津, 1993)。また，土屋（1995）は類似した定式化であるが，単純構造の解を直接目指している。

これらは，いずれも各項目に与えられる数量が1次元の場合に限定されているが，De Leeuw and Van Rijckevorsel（1988）は，項目に複数次元の数量を与える方法に初めて言及した（p.65)。しかしながら，実データに適用可能な分析モデルとしての具体化は，それ以後もなされていないと思われる。Murakami, Kiers, and Ten Berge（1997）は，この「複数次元の数量化を伴うカテゴリカル・データの非計量的（非線形）主成分分析」の実用的なアルゴリズムを開発し，その際，新たに発生する負荷行列の双方向からの回転について，新たな基準を提起しようとしている。その定式化は Gifi（1990）に近いものであり，アルゴリズムは，Ten Berge（1988）の方法の一般化を含む新しいものであるが，indicator 行列（後述）そのものでなく，その積和行列にもとづく点は，大津（1993)と共通している。

理論を扱った上記の論文の刊行までには，なお時間を要するが，本論文は，この方法の幾つかの適用例を示して，方法の実用性を検討することを目的とする。実際の適用に当たっては，各項目に与えられる数量の次元と主成分の数の決定法，本質的に非線型なアルゴリズムにおける極小（local minima）の問題等，検討すべき問題も多いが，本論文では，非計量的主成分分析で何がわかるかということを説明するための適用例を示すことに主眼をおく。適用例に関しても，後述するように，負荷行列の回転の方法に関して，われわれはまだ最終的結論に達しておらず，ここでは暫定的な手続きを用いているから，将来，同一のデータに対しても，若干異なった解釈を導く可能性もある。この点はあらかじめお断りしておきたい。

2. 方法の概要

2. 1　カテゴリカル・データのコーディング

Table 1 のようなデータを考えよう。

これは7人の被験者の3つの項目への反応を示したものであるとしよう。各項目には，それぞれ3つの反応選択肢があるが，それらは必ずしも自然な順序をもたないとする。これは，本論文で扱うカテゴリカル・データの典型的なケースである。仮に，3つの項目について，すべて，$a=1$，$b=2$，$c=3$と「数量化」して，項目間の相関行列を計算すると Table 2 のようになる。

項目1と2は全く無相関であり，項目3は項目1よりも項目2に高く相関しているとみることができる。次に，項目3についてのみ，$a=1$，$b=3$，$c=2$とコーディングしなおして，項目間相関行列を計算すると，Table 3 のようになる。

Table 1　Artificial categorical data

	Item		
Subject	1	2	3
1	a	a	a
2	a	b	a
3	b	b	b
4	b	b	b
5	c	a	b
6	c	c	c
7	a	c	c

Table 2　Correlation matrix of the coded data

	1	2	3
1	1.00	0.00	0.45
2	0.00	1.00	0.75
3	0.45	0.75	1.00

Table 3　Correlation matrix using another way of coding for Item 3

	1	2	3
1	1.00	0.00	0.65
2	0.00	1.00	0.00
3	0.65	0.00	1.00

今度は，項目3は項目2とは無相関，項目1と高い相関をもつようになった。

このように，恣意的なコーディングによって，主成分分析を含む，いわゆる相関分析の結果は，著しく左右されることがわかる。何らかの基準に基づいて，一義的な数量化を行うことが望ましいであろう。

2. 2　Indicator 行列

そこで，各カテゴリーに対して適切な数量を与える方法について考えよう。まず，indicator 行列を導入する。たとえば，項目3について言えば，その indicator 行列 G_3 の要素は，

$$g_{ij3} = \begin{cases} 1 & \text{被験者 } i \text{ が項目 3 で } j \text{ 番目の選択肢を選んだとき} \\ 0 & \text{そうでないとき} \end{cases}$$

のように定義される。すなわち，

$$G_3 = \begin{bmatrix} 1&0&0 \\ 1&0&0 \\ 0&1&0 \\ 0&1&0 \\ 0&1&0 \\ 0&0&1 \\ 0&0&1 \end{bmatrix}$$

となる。各行に1個所だけ1があることが特徴である。（この条件を満たす indicator 行列を，完全（complete）indicator 行列と言う。Gifi, 1990; p.67)。欠損値があると，この条件は満たされなくなる。3つの項目全体に対する indicator 行列を，Table 4 に示す。各行には項目数だけ1が存在する。

それぞれのカテゴリーに与える数量（重み）のベクトルを，$\boldsymbol{v}_3$ とすると，$\boldsymbol{x}_3 = G_3\boldsymbol{v}_3$ で定義される $\boldsymbol{x}_3$ は，変数3を数量化した結果を与える。例えば，$\boldsymbol{v}_3 = [1\ \ 2\ \ 3]'$ とすれば，

$$\boldsymbol{x}_3 = \begin{bmatrix} 1&0&0 \\ 1&0&0 \\ 0&1&0 \\ 0&1&0 \\ 0&1&0 \\ 0&0&1 \\ 0&0&1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \\ 2 \\ 3 \\ 3 \end{bmatrix}$$

である。これは，先に見た $a=1$，$b=2$，$c=3$ というコーディングの結果に等しい。

ここでは，(1) のような，主成分分析の対象となる，中心化，標準化された（すなわち標準得点化された）ベクトル $\boldsymbol{z}_3$ だけを対象とするとすれば，$\boldsymbol{v}_3 = [-1.32\ \ 0.00\ \ 1.32]'$ とする必要がある。ここで，$\boldsymbol{z}_3 = G_3\boldsymbol{v}_3 = [-1.32 \cdots 1.32]'$ が標準得点化されていることを確かめるのは容易である。

Table 4　Indicator matrix of the data in Table 1

	Item 1			Item 2			Item 3		
Subject	a	b	c	a	b	c	a	b	c
1	1	0	0	1	0	0	1	0	0
2	1	0	0	0	1	0	1	0	0
3	0	1	0	0	1	0	0	1	0
4	0	1	0	0	1	0	0	1	0
5	0	0	1	1	0	0	0	1	0
6	0	0	1	0	0	1	0	0	1
7	1	0	0	0	0	1	0	0	1

2. 3　最適な数量化の方法

このようにして，各変数を，（今のところ未知の）重み $\boldsymbol{v}_k$ を用いて数量化して得られた変量ベクトル $\boldsymbol{z}_k$ をならべてできる $n \times m$ の行列 Z を，(1) のように主成分分析することを考えよう。主成分の数 r を所与としたとき，因子モデル FA' は，Z の列の分散の和，すなわち（各列の分散は標準化によって1となっているから）m を最大限説明するように求められる。そうであるとすれば，その因子モデルが説明する分散を最大にするように，Z の各列，$\boldsymbol{z}_k$ に数値をわりあてるのが，最適な数量化であると考えることが可能である。これは，次のような関数を最小化することに帰着する（例えば，Bekker & De Leeuw, 1988)。

$$f(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m, F, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_m) = n^{-1} \sum_{k=1}^{m} \| G_k\boldsymbol{v}_k - F\boldsymbol{a}_k \|^2 \quad (2)$$

式 (1) も，Z の列ごとに，

$$f(F, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_m) = n^{-1} \sum_{k=1}^{m} \| \boldsymbol{z}_k - F\boldsymbol{a}_k \|^2$$

と書くことができるから，$\boldsymbol{v}_k$ が未知であることを除けば，(2) は主成分分析と等価である。ここでは，負荷行列は，

$$A = \begin{bmatrix} \boldsymbol{a}_1' \\ \boldsymbol{a}_2' \\ \vdots \\ \boldsymbol{a}_m' \end{bmatrix}$$

と分割されて出現している。このことは，(2) においても同様である。

ただし，(2) の最小化においては，(1) の場合と同様の制約条件，$n^{-1}F'F = I$ に加えて，数量化されたデータを中心化，標準化するために，$\mathbf{1}'G_k'\boldsymbol{v}_k = \mathbf{0}$，$n^{-1}\boldsymbol{v}_k'D_k\boldsymbol{v}_k = 1$ という条件も課される必要がある。ここまでが，従来の議論である。

2. 4　複数次元の数量化

Table 2 と Table 3 において，項目3のコーディングの仕方を変えることにより，他の項目との相関のあり方が変化するのを見た。すなわち，項目3には項目1と項目2に関連する成分がともに含まれているとみることもできる。項目3に2通りの方法で数量を与えることによって，これらをともに取り出すことができるのではなかろうか。

一般に，項目 k のカテゴリーの数を p_k とするとき，p_k-1 次元の1次独立な数量を定義することが可能である。実際，項目3には，2つの重みベクトルを用いて，

次のような2つの相互に直交する変量を導き出すことができる。

$$Z_3 = \begin{bmatrix} 1&0&0\\1&0&0\\0&1&0\\0&1&0\\0&1&0\\0&0&1\\0&0&1 \end{bmatrix} \begin{bmatrix} 0.81 & -1.36 \\ 0.51 & 1.04 \\ -1.57 & -0.20 \end{bmatrix} = \begin{bmatrix} 0.81 & -1.36 \\ 0.81 & -1.36 \\ 0.51 & 1.04 \\ 0.51 & 1.04 \\ 0.51 & 1.04 \\ -1.57 & -0.20 \\ -1.57 & -0.20 \end{bmatrix}$$

このやり方は，項目3を「コピー」して新たに第4の項目を作り，それに項目3とは別の重みを適用して変量を作り出していると見ることもできる（たとえば，Van Rijckevorsel & De Leeuw，1988; p.65)。

こうして作り出された4つの変数の最適な主成分を，Murakami，Kiers & Ten Berge (1997) の方法で求めたものが，Table 5～8である。Table 5は，数量化の結果生み出された4つの変数であり，Table 6は，それらの間の相関係数である。項目3の2つの変数が，項目1と項目2(に対応する変数)にそれぞれ高く相関していることが見て取れる。

Table 7とTable 8が分析の主要なアウトプットである。Table 7は，負荷行列 A であり，Table 8は主成分得点 F である。それぞれ，Table 6とTable 5の簡明な要約となっていると見ることができるであろう。

変数Item 3−1の符号が，Item 2，および，主成分得点IIと逆であることに注意。結果的に負荷量も負になっている。われわれのアルゴリズムは，この符号をそろえる機能をもっていないから，必要なら，負荷量の行単位での符号反転を行う必要がある。

ここで，この方法のモデルを形式的に書いておこう。まず，改めて記号を定義する。G_k ($k = 1, \ldots, m$) を，項目 k の $n \times p_k$ の（所与の）indicator 行列（$G_k' 1 = 1$，すなわち，各行に1つの1がある)，V_k を，$p_k \times q_k$ ($p_k > q_k$) の重み（数量化）行列，F を $n \times r$ ($\max q_k \leq r \leq \Sigma q_k$) の主成分得点行列，$A_k$ (k=1, …, m) を $q_k \times r$ の項目 k の負荷量行列とする。なお，適用の対象となるデータは，$n \gg p = \sum_{k=1}^{m} p_k$ が成り立ち，$G = [G_1 | \ldots | G_m]$ のランクが p であるとする。ここで，p は延べカテゴリー数を表す。つまり，被験者数は，延べカテゴリー数に比して十分多く，各カテゴリーに少なくとも1人の被験者が反応しているような「まともな」データを考えるものとする。(この仮定を実現するためには，カテゴリーの抹消等が必要な場合があろう。)

そこで，関数，

$$f(V_1, \ldots, V_m, F, A_1, \ldots, A_m) = n^{-1} \sum_{k=1}^{m} \| G_k V_k - F A_k' \|^2 \tag{3}$$

を，$n^{-1} V_k' G_k' G_k V_k = I$ ($k = 1, \ldots, m$), $F' 1_n = 0$，および，$n^{-1} F' F = I$ の制約条件の下で最小化することを考える。(つまり，主成分得点は，中心化，標準化され，かつ相互に直交するように求められる。) ここで，q_k は各項目に対する数量化の次元の数，r は主成分の数であり，ともに分析に先立って与えられているものとする。したがって，(3) は，indicator 行列から，主成分数 r

Table 5 Quantified matrix of the data in Table 1

Subject	Item 1	Item 2	Item 3 - 1	Item 3 - 2
1	−1.10	−0.60	0.81	−1.36
2	−1.10	−0.65	0.81	−1.36
3	1.24	−0.65	0.51	1.04
4	1.24	−0.65	0.51	1.04
5	0.40	−0.60	0.51	1.04
6	0.40	1.58	−1.57	−0.20
7	−1.10	1.58	−1.57	−0.20

Table 6 Correlation matrix of quantified data

	Item 1	Item 2	Item 3−1	Item 3−2
Item 1	1.00	−0.23	0.11	0.87
Item 2	−0.23	1.00	−0.99	−0.13
Item 3−1	0.11	−0.99	1.00	0.00
Item 3−2	0.87	−0.13	0.00	1.00

Table 7 Matrix of Loadings

	I	II	h^2
Item 1	−0.11	**0.96**	0.94
Item 2	**0.99**	−0.13	1.00
Item 3−1	**−1.00**	−0.00	1.00
Item 3−2	0.00	**0.97**	0.94
Contribution	2.00	1.88	3.88

Table 8 Matrix of scores

Subject	I	II
1	−0.80	1.32
2	−0.82	1.32
3	−0.51	−1.15
4	−0.51	−1.15
5	−0.49	−0.72
6	1.57	−0.19
7	1.57	0.58

の主成分モデルが最大の説明力をもつような，$\sum_{k=1}^{m} q_k$ 個の，中心化，標準化された変量を作り出すことである，と解釈できるであろう。なお，数量化された変数が中心化されるという条件 $1_n' G_k V_k = \mathbf{0}$ は仮定されていないが，$F' \mathbf{1}_n = \mathbf{0}$ から導かれる。

また，$Z_k = G_k V_k$ とかくと，負荷行列は，

$$A_k = n^{-1} Z_k' F \qquad k = 1, \dots, m \tag{4}$$

が成立する。すなわち，数量化された変数と主成分得点との相関行列であり，通常の主成分分析と平行関係にあり，解釈上重要な性質である。

アルゴリズムの詳細については省略するが，重み V_k を所与として，変量 Z_k から算出される相関行列を固有値分解するステップと，因子負荷 A_k を所与として，重みを推定するステップを交互に反復する形になっている。

2. 5　数量化Ⅲ類との関係

ここで述べた方法と数量化Ⅲ類との関係について述べる。この関係は，それと等価な等質性分析（Gifi, 1990）の最適化基準と比較するのが見やすいであろう。その最適化基準は，X を $n \times r$ の未知数の行列とするとき，関数，

$$f(V_1, \dots, V_m, X) = n^{-1} \sum_{k=1}^{m} \| G_k V_k - X \|^2 \tag{5}$$

を，制約条件，$X'1 = 0$, $n^{-1} X'X = I$ の下で最小化することである。大きな違いは，(5) における，全項目について共通の得点行列 X が，(3) においては，項目ごとに異なる因子モデル FA_k' で置き換えられていることである。一見したところ，(3) の方が複雑であり，A_k の分だけ推定すべきパラメータも増加しているようにみえるが，q_k は r より小さく，かつ項目ごとに変えることができることから，V_k の次元が小さくなることが重要である。また，解釈の対象を，V_k から A_k に移すことができる点も（特に因子分析になれたユーザーには），大きな意味をもつ。V_k は，(4) のような記述統計測度としての意味をもたず，頻度の小さいカテゴリーに極端に大きな値が割り当てられる傾向があるため，解釈に困難を来すことがある。

なお，全ての項目 k について，$q_k = \min(r, p_k - 1)$ とすると，(3) は (5) と同じ最小値をもつ。すなわち，非計量的主成分分析は，アイテムカテゴリー型データの数量化Ⅲ類を，特殊な場合として含んでいる。

さらに，全ての項目が2値データの場合には，必ず $q_k = 1$ で，(2) を用いることになるが，この場合，中心化，標準化された数量は，素点の標準得点化以外になく，これは2値データを標準化して行う主成分分析，（いわゆる ϕ 係数にもとづく主成分分析）と一致する。この場合は，上述の $q_k = \min(r, p_k - 1)$ の条件もみたしており，数量化Ⅲ類の結果とも一致することになる（たとえば，山田・西里，1993）。このことは，2値データを素点のまま分析する場合に，主成分分析を用いることの1つの justification になるであろうし，（あくまでもアイテムカテゴリー型のデータとして扱う場合に限るが，）数量化Ⅲ類の解釈に当たって負荷行列を用いたり，回転を行ったりすることも正当化するであろう。一方，ここで述べている非計量的主成分分析が，2値データの主成分分析の弱点を引き継いでいる可能性も示唆するものである。この点については，後に再度論じよう。

2. 6　回転方法についてのコメント

最後に負荷量行列の回転について述べておこう。通常の主成分分析の場合と同様，F と A_k には直交回転に関する自由度が存在するが，$q_k \geq 2$ の場合，さらに V_k と A_k' には，別に共通の直交回転が可能である。すなわち，$S_k (k = 1, \dots, m), T$ を任意の直交行列とすれば，$\| Z_k V_k - FA_k' \|^2 = \| Z_k V_k S_k - FTT' A_k' S_k \|^2$ である。したがって，最終的な解釈の対象となる行列は，$S_k' A_k T$, $V_k S_k$,（必要なら）FT である。（先の，人工データの例では，$V_k S_k$ のかわりに，数量化されたデータ行列 $G_k V_k S_k$ を示した。大きなデータでは，これは効率的な方法ではない。）

このような，負荷行列の両側からの回転は，3相主成分分析（three-mode principal component analysis）における核行列（core matrix）の回転（Kiers, 1997a）と類似しており，そこで用いられる基準をここにも導入することができる。その際，基準は複数考えることができるが，適切な基準は，データと分析の目的により異なっているように思われる（Kiers, 1997b）。前述のように，われわれはこの方法について，まだ結論を得るに至っていない。そこで，先の人工データの分析，および，以下の分析例では，まず，個々の A_k の varimax 回転によって S_k を定め，次に，これらをまとめた（左から回転された）負荷行列，

$$B = \begin{bmatrix} S_1' A_1 \\ S_2' A_2 \\ \vdots \\ S_m' A_m \end{bmatrix}$$

を varimax 回転するという方法を用いている。これによって，多くの場合，満足すべき単純構造が得られるのは事実であるが，これは，何らかの（単一の）基準の最

適化ではないから不満が残る。

3．適用例1――飲み物の特徴

3．1　データの概要

最初に適用の対象とするデータは，Kiers（1989）からとられた。これは，Table 12に示すように，34種類の飲み物を，7つの観点（項目）から評価したものである。それらの観点とは，1. アルコール濃度――0％(1)から30%以上(5)までの5段階，2. 糖分――あり(1)となし(2)の2カテゴリー，3. 炭酸――あり(1)となし(2)の2カテゴリー，4. 原材料――果物(1)，小麦(2)，ハーブ(3)，人工的香料(4)の4段階，5. 価格――安い(1)から高い(4)までの4段階，6. 味――非常に甘い(1)，甘い(2)，辛い(3)，苦い(4)の4カテゴリー，7. 色――無色(1)，赤(2)，薄赤(3)，黄色(4)，褐色(5)の5カテゴリーである。評定は，Kiers一人によって行われており，著者自身も指摘している通り，評定者の（限られた）知識だけにもとづくものであるため，データとしての信頼性・妥当性は高いとは言えない。

しかしながら，このデータは以下のような魅力的特

Table 9　Drinks data (Kiers, 1989)

Labels	Alcohol	Sugar	Carbon	Raw product	Price	Taste	Color
(1) Syrup	1	1	2	4	2	1	4
(2) Cola	1	1	1	4	1	1	5
(3) Seven-up	1	1	1	4	1	1	1
(4) Orangina	1	1	1	1	1	1	4
(5) Apple juice	1	2	2	1	1	1	4
(6) Orange juice	1	2	2	1	2	2	4
(7) Red Bordeaux	2	2	2	1	2	3	2
(8) White Bordeaux	2	2	2	1	2	3	4
(9) Red Lambrusco	2	2	1	1	2	1	2
(10) Rose	2	2	2	1	2	2	3
(11) Moselle wine	2	2	2	1	2	1	4
(12) Sekt	2	2	1	1	2	2	4
(13) Riesling	2	2	1	1	4	1	1
(14) Champagne ds	2	2	1	1	4	2	4
(15) Champagne br	2	2	1	1	4	3	4
(16) Sherry	3	2	2	1	3	2	5
(17) Port	3	2	2	1	3	1	2
(18) Cointreau	5	1	2	1	4	1	1
(19) Jenever	5	2	2	2	3	4	1
(20) Gin	5	2	2	2	4	4	1
(21) Whisky	5	2	2	2	4	4	4
(22) Beer	2	2	1	2	1	4	4
(23) Old-br. beer	2	1	1	2	2	4	5
(24) Guiness	2	2	1	2	2	4	5
(25) Cider	2	1	1	1	2	1	4
(26) Strawberry lq	4	1	2	4	3	1	3
(27) Banana lq	4	1	2	4	3	1	4
(28) Cherry brandy	4	1	2	4	3	1	3
(29) bl. Currant lq	4	1	2	4	3	1	2
(30) Slivovic	5	2	2	1	4	4	1
(31) Ouzo	5	1	2	3	4	1	1
(32) Pernod	5	1	2	3	4	1	1
(33) Jaegermeister	4	2	2	3	3	4	5
(34) Rum	5	1	2	2	4	2	1

徴をそなえている。

1）自然な順序をもつカテゴリカルな項目（1と5），2値の項目（2と3），順序のないカテゴリカルな項目（4，6，7）が混在しており，この方法がめざしているデータの1つの典型である。

2）特に，順序のないカテゴリカルな評定がなされている3項目に対して，複数の次元の数量を与える効果が示される可能性がある。

3）データ全体が提示できるほど小規模であり，（通常は被験者にあたる）対象である飲み物に，明確な個性があって結果の解釈可能性が判断しやすい。

そこで，このデータの分析を最初の適用例として提示しよう。

3. 2　分析の概要

このデータでは，特に複数次元の数量化というこの方法の特徴が，何らかの意味のある解釈を導くかどうかが目的なので，順序のないカテゴリーをもつ項目，4，6，7の数量の次元を2とした。自然な順序のある項目1と5の次元は1とした。2値項目である，2，3の次元は必然的に1である。$q_4=q_6=q_7=2$，$q_1=q_2=q_3=q_5=1$だから，$\sum_{k=1}^{7} q_k=10$，$\max_q q_k=2$であり，主成分の数は，$2 \le r \le 10$の範囲が可能である。そこで，q_kは上記の値に固定した上で，rの値を2～4の範囲で変えて，当てはまりを比較した。

式(3)のfは，$R=n^{-1}Z'Z$と定義したとき，$f=\sum_{k=1}^{m} q_k-\mathrm{tr}A'A$と書き換えることができる。通常の主成分分析の場合と同様，$\mathrm{tr}A'A$は，主成分によって説明される分散の大きさであり，Rの大小順にr番目までの固有値の和である。本論文では，この「説明力」を当てはまりの指標として用いることにしよう。式(3)で示される残差分散の大きさは，これを$\sum_{k=1}^{7} q_k=10$から引くことによって求められる。

「説明力」の値は，順に，5.79，7.35，8.21であった。それぞれの解の解釈を検討した上で，数量化で得られた10個の変数の全分散，すなわち，$\sum_{k=1}^{7} q_k=10$の73.5%を説明する$r=3$の解を提示することにした。対応する同質性分析の「説明力」は，$2 \le r \le 4$に対して，それぞれ，6.67，9.03，10.90であり，それぞれに対応する非計量的主成分分析の説明力の割合は，86.7%，81.4%，75.3%である。この面から見ても，$r=3$は，データに対する当てはまりの点で，特に不満のない解といえるであろう。

3. 3　結果の解釈

各項目に対する数量化（重み）行列をTable 10，負荷量行列をTable 11，主成分得点行列をTable 12に示した。解釈は（通常の主成分分析の場合と同様に）負荷量行列を中心に，適宜，重みと得点に言及しながら進め

Table 10 Matrices of Weights for Drinks data

	1	2
1. Alcohol		
	1	
(1) No alcohol	−0.82	
(2)	−0.89	
(3)	0.32	
(4)	0.98	
(5) Over 30%	1.36	
2. Sugar		
	1	
(1) Yes	−1.20	
(2) No	0.84	
3. Carbon		
	1	
(1) Yes	1.35	
(2) No	−0.74	
4. Raw product		
	1	2
(1) Fruit	−0.58	−0.78
(2) Grain	1.72	−0.26
(3) Herbs	1.06	1.31
(4) Artificial flavor	−0.78	1.60
5. Price		
	1	
(1) Cheap	1.34	
(2)	0.84	
(3)	−0.99	
(4) Expensive	−0.98	
6. Taste		
	1	2
(1) Very sweet	0.52	0.92
(2) Sweet	0.24	−1.09
(3) Dry	1.09	−1.80
(4) Bitter	−1.76	−0.13
7. Color		
	1	2
(1) Colorless	−0.77	−1.35
(2) Red	1.35	−0.28
(3) Light-red	1.15	−1.37
(4) Yellow	0.43	0.85
(5) Brown	−1.75	0.83

Table 11 Matrix of loadings for Drinks data

	I	II	III	h^2
1. Alcohol	**0.90**	0.26	0.25	0.94
2. Sugar	−0.07	0.15	**−0.89**	0.82
3. Carbon	**−0.81**	0.14	0.06	0.69
4. Raw product 1	0.19	**0.87**	0.00	0.80
5. Raw product 2	0.22	−0.05	**0.81**	0.72
6. Price	**−0.86**	−0.13	−0.02	0.75
7. Taste 1	−0.12	**−0.88**	0.09	0.80
8. Taste 2	0.09	−0.08	**0.79**	0.64
9. Color 1	0.05	**−0.77**	−0.24	0.65
10. Color 2	**−0.69**	0.03	−0.24	0.54
Contribution	2.80	2.27	2.28	7.35

Table 12 Matrix of scores for Drinks data

	I	II	III
(1) Syrup	−0.57	−0.82	1.27
(2) Cola	−1.66	0.23	1.90
(3) Seven-up	−1.02	−0.25	1.83
(4) Orangina	−1.43	−0.46	0.73
(5) Apple juice	−0.59	−0.64	−0.37
(6) Orange juice	−0.33	−0.50	−1.15
(7) Red Bordeaux	0.06	−1.22	−1.48
(8) White Bordeaux	−0.31	−0.83	−1.39
(9) Red Lambrusco	−0.79	−0.77	−0.24
(10) Rose	0.27	−0.89	−1.18
(11) Moselle wine	−0.45	−0.64	−0.41
(12) Sekt	−1.06	−0.24	−0.88
(13) Riesling	−1.02	−0.56	−1.12
(14) Champagne ds	−0.47	−0.25	−1.04
(15) Champagne br	−0.43	−0.57	−1.27
(16) Sherry	0.40	0.35	−0.96
(17) Port	0.89	−0.98	−0.64
(18) Cointreau	1.12	−0.30	0.56
(19) Jenever	1.31	1.57	−0.50
(20) Gin	1.31	1.57	−0.50
(21) Whisky	0.89	1.26	−0.75
(22) Beer	−1.27	1.40	−0.34
(23) Old-br. beer	−1.46	2.10	0.77
(24) Guiness	−1.32	2.19	−0.06
(25) Cider	−1.30	−0.46	0.69
(26) Strawberry lq	1.20	−1.11	1.13
(27) Banana lq	0.58	−0.73	1.16
(28) Cherry brandy	1.20	−1.11	1.13
(29) bl. Currant lq	0.95	−1.13	1.06
(30) Slivovic	1.29	0.69	−0.71
(31) Ouzo	1.13	0.31	1.33
(32) Pernod	1.13	0.31	1.33
(33) Jaegermeister	0.55	1.76	0.11
(34) Rum	1.24	0.72	0.03

ることにしよう。

Table 11 の第1列に示した第1主成分には,「アルコール」,「炭酸」(負),「価格」(負),「色」の第2次元(負)が高い負荷をもっている。この主成分が,「アルコールが強く,炭酸を含まず,価格が高いもの」と,そうでないものを両極にもつものであることは,Table 12 の第1列で正の大きい値をもつ飲み物が,Gin, Cherry Brandy, Rum 等であり,負の目立つ値をとるのが,Cola, Beer, Cider 等であることからあきらかであろう。(「価格」の数量化は,Table 12 の重みからわかるように,安いほど大きくなるようになされている。)「色」の第2次元は,「黄色」と「褐色」に正,「薄赤」と「無色」に負となるように付されている。主成分得点上で正負の値をとる飲み物の特徴を見れば,このことは納得できるであろう。他の特徴からすれば,この主成分上で正の大きい値をとるはずの Whisky の値がさほどでない (0.81) のは,色(黄色と評価されている)がここで対象となった強アルコール飲料の中で例外的だからである。

Table 11 にもどって,第2の主成分に高く負荷する変数を見ると,「原材料」の第1次元,「味」の第1次元(負),「色」の第1次元(負)となっている。奇しくも,自然な順序のないカテゴリーだけが負荷するものとなっており,解釈には Table 10 の重みを検討する必要がある。「原材料」の第1次元は,「小麦」,「ハーブ」が正,「果物」と「人工的香料」が負となっている。「味」の第1次元は,「辛み」(正)と「苦み」(負)が対立する。「色」の第1次元は,「赤」,「薄赤」が正,「褐色」と「無色」が負となっている。第2の主成分は,小麦,ハーブを原材料とし,苦く,褐色(または無色)の飲み物と,果物,または人工的香料を材料とし,辛く,赤い色の飲み物を両極にもつはずである。Table 12 を見ると,大きな正の値は,Old-brown beer と Guiness 等がとり,負の極には,Red Bordeau, Strawberry liquor 等が並ぶから,その解釈は裏付けられる。

Table 11 で第3の主成分に高く負荷しているのは,「糖分」(負),「原材料」の第2次元,「味」の第2次元である。Table 10 から,「糖分」はない場合に正となる方向となっており,「原材料」の第2次元は,「ハーブ」と「人工的香料」が正,「果物」が負となっている。また,「味」の第2次元は,「非常に甘い」(正)対「甘い」と「辛い」という方向である。したがって,この主成分

は，主に人工的な材料による，極端に甘い飲み物と，果物を原材料とする相対的に薄味の飲み物が対立する内容であろう。Table 12 から，この主成分で大きな正の値を示すのが，Cola と Seven-up であり，負の側にあるのはワインとシャンパン系のドリンクだから，納得がいくであろう。

この例は，先に述べた信頼性・妥当性の点以外にも，対象（飲み物）の数の点で，統計解析の対象としては十分なものとは言えず，ここから何らかの経験的結論を引き出すべきではない。しかしながら，自然な順序をもたないカテゴリーに，複数の次元の数量を与えて因子分析的に取り扱うという方法に，一定の必然性があることを示した点で，意義を認めてよいと思われる。

4．適用例 2 —— 愛知の長期展望

4．1　データの概要

第 2 のデータは，「愛知の長期展望有識者アンケート調査」の一部である。この調査は，「21世紀社会を展望し，愛知県の今後の長期的な地域づくりの方向を検討するための基礎資料をうるため，本県の地域づくりにかかわる事柄について県内外の有識者から幅広く意見を聴取」することを目的に，1994年11月から12月にかけて行われたものである。対象者は，特定の方法で抽出された愛知県内，および県外在住の有識者各300名であり，有効回収数は306（回収率51%）である。調査全体は，述べ項目数約400に及ぶ膨大なものであるが，ここで分析の対象とするのは，その中の問29に含まれる17項目である。

問29の質問は，「地域づくりの各分野において，その推進主体として中心的な役割を担うのは，どの主体であるとお考えですか。」というものであり，17の分野（1. 都市開発，2. 先端技術開発・研究，3. マルチメディア・高度情報通信システム，4. 産業基盤整備・地域産業おこし，5. 研究学園都市整備，6. イベント・コンベンション，7. 国際開発・国際協力，8. リゾート・観光産業，9. 文化・スポーツ事業，10. 空港整備，11. 高速道路整備，12. 鉄道整備，13. 住宅整備，14. 福祉事業，15. 環境保全・緑化事業，16. 教育施設整備，17. 医療施設整備）がならべられ，そのすべてに対して，9 つの選択肢（国，県，市町村，広域行政体，民間企業，経済団体・労働組合，地域コミュニティ・ボランティアグループ，家庭・個人，第三セクター）が付されている。実際には，1 番大きな役割をはたすものと，2 番目に大きな役割を果たすものという 2 つの反応が求められているが，ここで分析の対象とするのは，「最も重要な役割を果たすもの」の方だけである。なお，分析の対象となったのは，有効回答のうちでも，この17項目のすべてに回答が記されていた274名分である。

これは典型的な，自然な順序のないカテゴリカル・データである。ただし，非計量的主成分分析を適用するのに適切なデータであるかどうかは，多少疑問もある。この調査の場合，有識者の代表的反応がどのあたりにあるか，すなわち，各項目において反応の集中する選択肢は何か，

Table 13　The numbers of responses for each category in each item of Aichi data

	国	県	市町村	広域行政	民間企業	各種団体	ボランティア	家庭個人	第三セクター
1．都市開発	10	83	108	48	17	1	0	1	6
2．先端技術	57	38	3	6	158	2	0	0	10
3．マルチメディア	54	37	3	32	135	3	0	0	10
4．産業基盤	5	102	55	36	52	18	4	0	2
5．学園都市	55	137	19	50	5	2	0	0	6
6．イベント	0	97	62	30	45	19	10	0	11
7．国際交流	38	97	38	20	29	13	25	10	4
8．リゾート	2	25	39	28	148	2	1	2	27
9．文化	4	81	82	18	43	5	26	8	7
10．空港整備	138	71	0	38	2	1	0	0	24
11．高速道路	193	33	1	30	1	0	0	0	16
12．鉄道整備	126	24	3	30	62	1	0	0	28
13．住宅整備	20	83	58	13	67	0	0	24	9
14．福祉事業	30	62	134	17	5	0	21	3	2
15．環境保全	28	105	90	33	4	1	5	5	3
16．教育施設	38	118	95	10	5	0	0	5	3
17．医療施設	27	100	100	26	13	0	1	1	6

という点に目的があると見るべきであろう。集計の結果をTable 13に示したが，このように，国，地方自治体，民間企業の間の分業のあり方について，有識者の意見は，おおよそ常識的なところに集中していることが確認できるであろう。

しかし見方を変えれば，対象者の過半数が集中して選択したカテゴリーは，「高速道路整備」に対する「国」(193) だけであり，説明されるべきかなりの個人差が存在するともいえる。また，項目間の連関性の分析からは，項目と選択肢の意味について，意味のある解釈が得られる可能性もある。したがって，このデータを第2の適用例として用いることにしよう。

4. 2　分析の概要

上述のような事情からして，各項目に複数の次元の数量を付することは，解釈を必要以上に複雑なものとする恐れがある。したがって，ここでは全ての項目について，数量化の次元を1とした。すなわち，特に新しいものではない(2)のモデルを試みる。従来，このモデルの適用例もほとんどないから，ここで試みる意味はあろう。

分析に際して，誰も選択していないカテゴリーは前述のように削除しなければならないが，選択頻度の低いカテゴリーも結果に悪影響を与える恐れがある。したがって，幾つかのカテゴリーを事前に合併しておくことにした。暫定的に，頻度20未満のカテゴリーを他のカテゴリーに合併することにした。それでも，頻度10未満のカテゴリーを完全になくすことはできなかったが，そのまま実行した。結果的に，各変数のカテゴリーの数は，4～7の範囲にある（どのカテゴリーが合併されたかは，後にTable 14から，読み取れるようにした。）

その上で，$1 \leq r \leq 7$ の範囲で，当てはまりを検討した。前節同様，R の最大 r 番目までの固有値の和を求めると，順に，3.82，5.70，7.28，8.63，9.73，10.72，11.62となり，とくに大きな増加の「切れ目」は見出せないし，最大の $r = 7$ でも変数の全分散（項目数が17だから，17.0）の65.3%を説明するにすぎない。他方，同一次元の同質性分析に対する効率は，順に，100%，81.3%，75.7%，71.1%，67.5%，65.3%，63.4%となり，こちらにも決め手となるようなポイントは存在しないように見える。そこで，結果を検討したところ，$r = 5$ が比較的解釈しやすいことがわかった。通常の主成分分析の用法からすれば，17変数に対して5つの主成分はほぼ上限に近いといってよいであろう。（負荷量行列を用いて，多次元の解を解釈することができることが，非計量的主成分分析のメリットなので，主成分数を多くする方向へのバイアスは，幾分働いている。）説明力（57.9%），同質性分析への効率（67.5%）とも必ずしも満足のいくものではないが，この範囲では一応のバランスは取れているとみるべきかもしれない。このような結果になるのは，このデータがもともと組織的に個人差を測定する目的で収集されていないためでもあろう。

4. 3　結果の解釈

重みのベクトル（$\boldsymbol{v}_k$）を横に倒した形で（すなわち，転置して）Table 14に示し，負荷量行列をTable 15に示した。

まず，Table 14の重みを，一瞥しておこう。*** で示したのは，選択数0のカテゴリーである。他のカテゴリーと合併されたものは斜体で示した。重み（数量化）の方向は任意だから，「県」の値が正になるように揃えた。これは県が実施主体となった調査であり，もともと県が担当することが著しく自然でないような項目はふくまれていないから，「県」が他のカテゴリーと合併されることはなかった。実際には，県と同符号のカテゴリーと異符号のカテゴリーを見ることによって，解釈が容易になるであろう。

次に，Table 15の負荷量行列を見る。これは，極めて明快な構造をとっている。第1の主成分は，高く負荷している5つの項目から，「住民福祉」と呼ぶことができるであろう。同様に，第2の主成分は「文化事業」，第3は「運輸・交通インフラ整備」，第4は「先端技術開発」，第5は「地域開発」と名づけることができよう。数量化Ⅲ類では，こうした明快な結果を得ることは難しいから，この方法の意義は小さくないと認められよう。ただし，カテゴリカル・データでは，このような単純な構造の根拠を問うことは必ずしも単純ではない。

通常の主成分分析では，同じ主成分に高く負荷する項目は内的に一貫している，すなわち，それらの項目の間では，ある項目に肯定的に反応する被験者は他の項目にも肯定的に反応する傾向があり，ある項目に否定的に答える被験者は別の項目にも否定的に解答すると予想することができる。この意味で，主成分分析は，個人差の分析である。しかしながら，自然な順序のないカテゴリーをもつ項目では，肯定的－否定的という表現自体に意味がないから，Table 14にもとづいてカテゴリーの方向を検討することを通じて，一貫性の意味を考え直す必要がある。

まず，第1主成分，「住民福祉」に高く負荷する項目13～17の重みについて，Table 14を検討すると，「国」が「県」と同様正の値をとっているのに対し，負の値をとっているのは，「市町村」を除く少数意見，すなわち，現在の行政担当組織以外に対応するカテゴリーである。

Table 13 に見られるように，これらの項目では「県」と「市町村」の2つのカテゴリーにほぼ反応が集中している。しかし，数量化の結果はこの2つを対立させてはいない。「県」には高い値が付与されるが，「市町村」の値はほぼ0である。この理由は，恐らく，「県」を選択する対象者は5つの項目すべてにおいて一貫して「県」または「国」を選択し続けるが，「市町村」の選択者は，項目によって，「県」，「国」の方向と，「広域行政体」をふくむ民間団体の方向とに，反応が動揺する傾向があるせいであろう。

Table 14　Vectors of Weights for Aichi data

	国	県	市町村	広域行政	民間企業	各種団体	ボランティア	家庭個人	第三セクター
1. 都市開発	*0.49*	1.30	−0.53	−1.41	*0.49*	*0.49*	***	*0.49*	*0.49*
2. 先端技術	0.11	2.39	*−0.90*	*−0.90*	−0.50	*−0.90*	***	***	0.49
3. マルチメディア	0.40	2.25	*−0.69*	0.10	−0.72	*−0.69*	***	***	*−0.69*
4. 産業基盤	*0.21*	0.32	0.71	−2.53	0.26	*0.21*	*0.21*	***	*0.21*
5. 学園都市	0.37	0.53	*0.40*	−2.11	*0.40*	*0.40*	***	***	*0.40*
6. イベント	***	0.59	0.74	−2.41	0.11	*−0.90*	*−0.90*	***	*−0.90*
7. 国際交流	0.47	0.56	0.76	−3.05	−0.14	*−0.61*	−0.80	*−0.61*	*−0.61*
8. リゾート	*0.02*	0.59	1.37	−2.60	0.08	*0.02*	*0.02*	*0.02*	−0.26
9. 文化	*−2.12*	0.89	0.44	*−2.12*	−0.20	*−2.12*	−0.41	*−2.12*	*−2.12*
10. 空港整備	−0.86	0.23	***	1.76	*0.60*	*0.60*	***	***	1.43
11. 高速道路	−0.62	0.90	*1.51*	2.09	*1.51*	***	***	***	*1.51*
12. 鉄道整備	−0.91	0.67	*1.33*	1.85	0.01	*1.33*	***	***	*1.33*
13. 住宅整備	0.86	1.16	−0.29	*−1.92*	−0.26	***	***	−1.55	*−1.92*
14. 福祉事業	0.93	1.01	0.00	*−2.35*	*−2.35*	***	−1.31	*−2.35*	*−2.35*
15. 環境保全	0.43	0.94	−0.22	−1.49	*−2.34*	*−2.34*	*−2.34*	*−2.34*	*−2.34*
16. 教育施設	0.25	0.82	−0.45	*−2.75*	*−2.75*	***	***	*−2.75*	*−2.75*
17. 医療施設	1.01	0.94	−0.39	−1.88	*−1.56*	***	*−1.56*	*−1.56*	*−1.56*

Table 15　Matrix of Loadings for Aichi data

	I	II	III	IV	V	h^2
1. 都市開発	0.30	−0.16	−0.13	−0.04	**0.67**	0.58
2. 先端技術	0.09	−0.03	−0.01	**0.83**	0.06	0.71
3. マルチメディア	−0.01	0.10	−0.02	**0.83**	−0.05	0.70
4. 産業基盤整備	0.08	0.33	0.06	0.06	**0.62**	0.51
5. 研究学園都市	0.08	0.25	0.01	0.02	**0.60**	0.42
6. イベント・コンベンショ	0.07	**0.75**	−0.07	0.03	0.12	0.59
7. 国際交流	0.02	**0.76**	−0.08	0.07	0.10	0.60
8. リゾート・観光	−0.27	**0.44**	0.00	0.00	0.32	0.37
9. 文化・スポーツ	0.34	**0.61**	0.04	0.02	−0.18	0.52
10. 空港整備	−0.04	−0.05	**0.79**	−0.05	−0.16	0.66
11. 高速道路整備	−0.08	−0.01	**0.85**	0.06	0.08	0.74
12. 鉄道整備	−0.10	−0.05	**0.78**	−0.05	0.01	0.63
13. 住宅整備	**0.66**	0.01	−0.02	0.09	0.33	0.55
14. 福祉事業	**0.65**	0.27	−0.07	0.05	0.15	0.52
15. 環境保全・緑化	**0.61**	0.21	−0.07	−0.16	−0.02	0.45
16. 教育施設整備	**0.74**	0.10	−0.05	0.11	−0.04	0.58
17. 医療施設整備	**0.72**	−0.13	−0.20	0.03	0.15	0.60
Contribution	2.62	2.07	2.05	1.45	1.54	9.73

第2主成分「文化事業」は，行政の役割を重く見る人は一貫した反応を維持しており，民間企業・団体や個人を重視する人々と対峙している。しかし，第3主成分の「運輸・交通インフラ整備」では，「国」とそれ以外という対立が明確であり，「県」を主体と考える人は少ない。第4主成分「先端技術」は，このような，通常は地方自治体の取り組むべきテーマとは考えにくい分野ですら「県」を中心とした推進を支持する（サービス的反応？）少数派が，「民間企業」を重視する多数派の対極にある。第5主成分の「地域開発」では，「広域行政体」が唯一，他のカテゴリーと対立する一貫した反応である。面白いことに，「市町村」は，「都市開発」では「広域行政体」側の，「産業基盤整備」ではそれと対立する側の反応とみなされ，「研究学園都市整備」では反応自体が激減してしまう。

このあたりは，カテゴリーの合併の仕方によって異なった結果が得られる可能性もあるから，解釈には慎重を期する必要があるが，この種のデータの多変量解析としては，比較的明快な結果が得られたと言えよう。なお，ほとんどの主成分は，Table 13 から読み取れる多数意見と少数意見の対立の観を呈するから，主成分得点の分布は偏ったものになる。実際，主成分得点の歪度は，順に，−0.77，−1.30，0.84，1.41，−0.68 である。県が主体となることが多数意見の主成分では負の歪度が，そうでない主成分では正の歪度が得られていることがわかる。

ところで，このような適用場面では，主成分得点自体は解釈の対象とならず，その群差や他の測度との関係を検討することに意味がある場合が多い。ここでは，そのような用法も示しておこう。Table 16 は，愛知県居住者と（岐阜・三重両県を除く）愛知県外居住者に関して，主成分得点の平均差を求めたものである。はっきりとした差は，第1主成分にある。つまり，愛知県居住者は，「住民福祉」を「国」，「県」を中心に推進することを望む度合いが強い。

Table 17 は，同じ調査の問28「地域づくりを担う主体としては，国・県・市町村をはじめ民間企業など広範に及びますが，それらの各主体の果たす役割は今後どう変わっていくでしょうか。それぞれの主体ごとに，1つずつ選んで下さい。」に対する反応と，5つの主成分得点との相関を示したものである。Table 17 に示した9つの主体について，「1－大いに高まる，2－やや高まる，3－変わらない，4－やや低下する，5－大いに低下する」の5段階で評定されているが，この数値を逆転（5を1に，4を2に，…，1を5に変換）して，相関をもとめた。ここでも目立つ値は，第1主成分にある。「県」の役割が「高まる」という反応は，「住民福祉」の「県」，「国」による推進ということと結びついているようである。いずれも目覚しい結果とは言えないが，結果の利用方法の一例として見ていただきたい。

Table 16 Means of Component Scores for Two Respondents' Groups

	I	II	III	IV	V
愛知県内居住者	**0.18**	0.08	0.02	0.08	0.08
愛知県外居住者	**−0.27**	−0.03	−0.03	−0.07	−0.09

Table 17 Correlations of Component Scores with Other Questiions

	I	II	III	IV	V
国	**0.19**	−0.00	−0.09	0.11	0.11
県	**0.35**	−0.05	0.09	0.10	**0.18**
市町村	0.04	0.09	0.15	0.04	0.05
広域行政体	−0.04	−0.12	−0.06	0.04	−0.10
民間企業	−0.05	0.03	0.02	−0.05	0.06
各種団体	0.10	0.03	−0.01	0.07	0.08
ボランティア等	−0.08	−0.02	0.14	0.04	−0.02
家庭・個人	−0.00	−0.02	0.01	0.03	0.11
第三セクター	0.07	0.13	0.10	0.08	−0.04

5．適用例3 —— 自尊心尺度

5．1　データの概要

第3のデータは，Rosenberg（1965）による自尊心尺度（Self Esteem Scale, SES）である。被験者は日本人大学生1630名であり，項目の翻訳と実施は，1991年から1992年にかけて，他の13のパーソナリティ尺度とともに，中村陽吉学習院大学教授を代表者とする PSL 研究会によって行われた。自尊心尺度の原版は，Table 18 と 19 に示された10項目からなるが，日本語版の項目は，「1．私は概して自分に満足している。」，「2．私はときどき，自分がてんでだめだと思う。」，「3．自分には良いところがたくさんあると思う。」，「4．私はたいていの人と同じ程度にものごとができる。」，「5．私には誇りに思えることがあまりない。」，「6．私はときどき自分が確かに役立たずだと感じることがある。」，「7．私は少なくとも他人と同じ条件ならば，値打ちがある人間だと感じている。」，「8．もう少し自分の値打ちを認めることができればいいのにと感じる。」，「9．私は概して自分を落伍者だと思いがちである。」，「10．私は自分に対して前向きの態度をとっている。」である。

自尊心尺度は，すべての項目に対して，「全くあてはまらない」，「ややあてはまる」，「ややあてはまる」，「まったくあてはまらない」の4段階で反応を求めるものであ

り，自然な順序のあるカテゴリーをもっている。したがって，非計量的主成分分析を適用する理由は，第1に，各項目に1次元の数量を割り当てた場合，非計量的主成分分析は，4つのカテゴリーに1～4の数値を当てはめて行う通常の主成分分析の結果を再現するか，第2に，各項目に2次元の数量を割り当てたとき，通常の方法では得られないような，なんらかの意味のある主成分を生み出すか，ということに絞られるであろう。

5. 2　データ分析の概要

まず，上述の4つの反応カテゴリーを，順に，1～4とコーディングし，$r=2$として，通常の主成分分析を行うと，その説明力は4.94であった。次に，すべての項目に1次元の数量を割り当て（$q_k=1$, $k=1,\ldots,10$），やはり$r=2$として非計量的主成分分析（(2)の最小化）を行うと，その説明力は5.00であり，その差はほとんどない。実際，得られるアウトプットもほとんど同一であることは，後に Table 20 に示す。次に，$r=3$, $q_k=2$($k=1,\ldots,10$) として分析を行った。説明力は，8.34に一見増加するが，実際には変数の「コピー」によって，変数の数は20に増加するので，説明される分散の割合は，50.0%から41.7%に減少する。この解は興味深い側面を含んでいるので，次節で詳しく検討する。同質性分析の説明力は，$r=2$では7.21，$r=3$では9.02であった。$q_k=1$, $r=2$の非計量的主成分分析の解の同質性分析に対する効率は69.3%，$q_k=2$, $r=3$の効率は92.4%である。

5. 3　結果の解釈

この例では，やはり，各カテゴリーに複数の数量を与えた場合の結果に興味があるから，$q_k=2$, $r=3$の解から検討しよう。Table 18 は重み行列，Table 19 が負荷行列である。

ここでは，まず，重みから検討しよう。すべての項目について，第1の数量は4つのカテゴリーへの数量は，1から4へと単調に増加しており，しかもその間隔はお

Table 18　Matrices of Weights for SES data for the case of $r=3$

		1	2	3	4
(1) On the whole, I am satisfied with myself.	1	−2.00	−0.50	0.61	1.78
	2	−1.43	0.67	0.34	−2.71
(2) At times I think I am no good at all.	1	−2.28	−1.12	−0.19	1.41
	2	−3.53	−0.06	0.73	−0.77
(3) I feel that I have a number of good qualities.	1	−2.65	−0.60	0.61	1.72
	2	−2.32	0.58	0.24	−2.51
(4) I am able to do things as well as most other people.	1	−3.55	−1.22	0.12	1.58
	2	−3.86	−0.14	0.60	−1.61
(5) I feel I do not have much to be proud of.	1	−1.44	−0.65	0.57	2.19
	2	−2.13	0.35	0.68	−1.73
(6) I certainly feel useless at times.	1	−2.19	−0.72	0.17	1.67
	2	−2.48	0.45	0.61	−1.32
(7) I feel that I'm a person of worth, at least on an equal plane with others.	1	−2.77	−0.77	0.57	1.68
	2	−2.74	0.46	0.34	−2.43
(8) I wish I could have more respect for myself.	1	−2.08	−0.83	−0.13	1.62
	2	−3.49	0.15	0.69	−0.80
(9) All in all, I am inclined to feel that I am a failure.	1	−1.15	−0.12	1.17	2.89
	2	−1.17	0.84	0.06	−2.59
(10) I take a positive attitude toward myself.	1	−3.03	−1.14	0.15	1.11
	2	−2.60	0.10	0.86	−1.18

Table 19 Matrices of Loadings for SES data in the case of $r = 3$

		I	II	III	h^2
(1) On the whole, I am satisfied with myself.	1	**0.59**	0.37	0.04	0.48
	2	0.05	−0.01	**0.57**	0.32
(2) At times I think I am no good at all.	1	−0.29	**−0.68**	−0.02	0.55
	2	0.04	0.02	**0.55**	0.31
(3) I feel that I have a number of good qualities.	1	**0.78**	0.10	−0.02	0.62
	2	−0.02	0.02	**0.63**	0.39
(4) I am able to do things as well as most other people.	1	**0.61**	0.17	0.01	0.40
	2	0.01	0.07	**0.54**	0.30
(5) I feel I do not have much to be proud of.	1	**−0.58**	−0.38	0.00	0.48
	2	−0.01	0.02	**0.60**	0.36
(6) I certainly feel useless at times.	1	−0.29	**−0.63**	0.00	0.48
	2	0.07	−0.01	**0.58**	0.34
(7) I feel that I'm a person of worth, at least on an equal plane with others.	1	**0.75**	0.02	−0.00	0.57
	2	−0.00	−0.02	**0.65**	0.43
(8) I wish I could have more respect for myself.	1	0.06	**−0.72**	0.05	0.52
	2	0.02	−0.03	**0.55**	0.30
(9) All in all, I am inclined to feel that I am a failure.	1	−0.32	**−0.62**	−0.00	0.49
	2	−0.01	0.00	**0.56**	0.32
(10) I take a positive attitude toward myself.	1	**0.49**	0.37	−0.04	0.38
	2	−0.11	0.14	**0.53**	0.31
Contribution		2.77	2.24	3.33	8.34

およそ1である。(ただし，これは出力そのものではなく，適用例2と同様，後で方向をそろえたものである。) すなわち，(量的な制約は一切なされていないないにもかかわらず，) 与えられた数量はだいたい等間隔でその標準偏差はほぼ1となっている。これは，カテゴリーの「ラベル」をそのまま数値として分析したのとほぼ一致することを意味している。他方，第2の次元は，これもほぼ例外なく，カテゴリー1と4に対応する数量が負，2と3が正となっている。(方向をそろえたことは，第1の数量の場合と同様である。) この形は馬蹄 (horse-shoe) とよばれる形態であり，本質的にデータが1次元である場合に現れるとされているものである (たとえば，岩坪，1987; Pp.125-137, Gifi, 1990; p.147)。

Table 19 の負荷行列においては，面白いことに，第1の数量は，IとIIの2つの主成分に別れて負荷している。その際，項目(5)を除き，肯定したとき自尊心が高いと判断される項目がIに，逆転項目がIIに高く負荷している。このことは，Rosenberg (1965) においても見出された傾向である。すなわち，Iは自尊心の高さに，IIは低さに対応する。この2つの特性が一致しないのは，黙従傾向のような反応傾向の存在を示唆するものであるが，このデータだけからそのような判断を下すことはできない。他方，第2の数量は，すべて主成分IIIに高く負荷しており，この主成分は，I，IIより多くの分散を説明している。

ここで問題は，「馬蹄」の形を呈する第IIIの主成分が解釈可能か否かである。重みから解釈する限り，この主成分は極端な反応 (1か4) をする被験者と中庸の反応 (2と3) を選ぶ被験者が対極に位置するものであろう。このことについて検討するために，まず3つの主成分のヒストグラムを見る (Figure 1)。ここで，主成分IとIIはほぼ左右対称の分布となっているにもかかわらず，主成分IIIの分布は著しく負に偏っている。このことは，一貫して極端な反応をする被験者は相対的に少なく，多

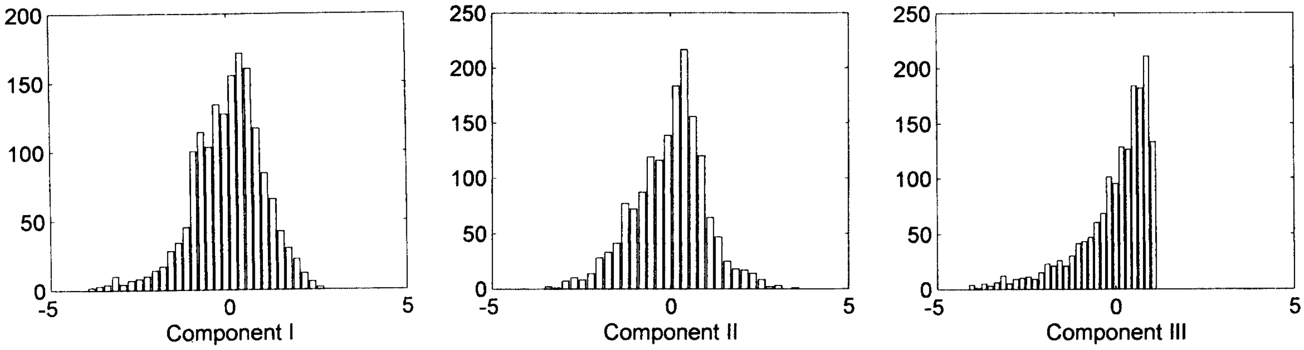

Figure 1　Histogram of 3 component scores for SES data

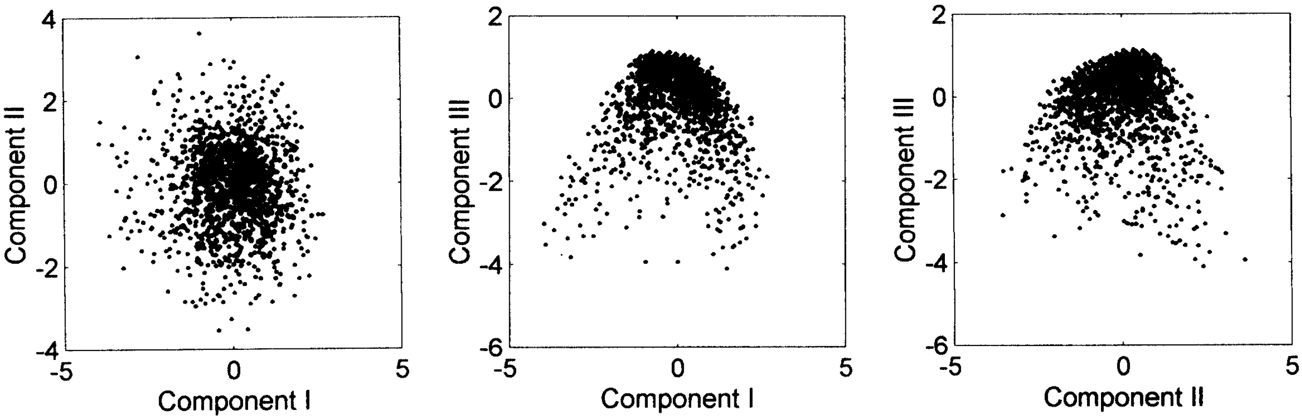

Figure 2　Scatter diagram showing the relationships between 2 component scores for SES data

くの被験者は中庸の反応を選んでいることが分かる。またこのことは，Table 18 の第 2 の重みの値が，極端に負の方向に偏っていることからも見てとれる。

次に，これらの 3 つの主成分の間の関係を示す散布図を，Figure 2 に示す。主成分 I と II の関係は，ほぼ円状であり，普通の意味で独立な次元である。しかし，主成分 III は，最初の 2 つの主成分と明確な非線型の関係にある。換言すれば，これらの間には再度「馬蹄」の関係がある。すなわち，主成分 III において，極端な値をとる被験者は，他の 2 つの主成分においてもどちらかの極に位置している。このことは，主成分 III が，第 I と第 II の主成分と並ぶ第 III の特性の次元を表現しているというよりは，反応の極端さという反応の仕方にかかわる個人差の次元であることを示唆している。

通常の数量化の分析では，「馬蹄」の出現をデータの 1 次元性の証拠，あるいは，データに内在する本質的な非線型性を，実質的に 1 次元のモデルである数量化 III 類で分析した結果生ずる「余剰情報」とみなし，あえて解釈しない考え方もあるが，この場合の主成分 III にもその考え方を適用すべきかもしれない。こうした余剰情報を生み出す点は，前述した，この方法が 2 値変数の主成分分析から受け継いだ欠点と見なされうる。2 値変数の主成分分析でしばしば問題になる難しさの因子（difficulty factor）は，ここで言う馬蹄と同じ意味をもつと解される。

ほとんど，同じ説明力をもっていた，$q_k = 1$，$r = 2$ の負荷行列と，カテゴリーをそのまま数値とみなして行った主成分分析のそれを，Table 20 にならべて示した。やはり，ほとんど同一の結果となっており，この結果は，少なくともこのデータに関する限り，カテゴリーをそのまま数量とみなして主成分分析することの正当性を示すものと言えよう。また，先の $q_k = 2$，$r = 3$ の解における主成分 III を解釈の対象としないとすれば，非計量的主成分分析は，このデータから特に意味のある情報を新たに抽出できなかったと言わざるを得ないであろう。

それでは，$r = 2$ で Table 20 の 2 つの解より格段に大きな説明力を示す，数量化 III 類の結果はどのようなものなのであろうか。これは，$q_k = 2$，$r = 2$ の解と等価であるから，それを示そう（Table 21）。ここでは，わざと重みの符号も，順序も変えずに生の出力をそのまま表にした。ここでは，項目に与えられた 2 次元の数量が，きれいに 2 つの主成分に分離して負荷している。第 1 の

Table 20 Matrices of Loadings for SES data by Nonmetric PCA and Ordinary PCA

	Nonmetric PCA			Ordinary PCA		
	I	II	h^2	I	II	h^2
(1)	**0.58**	0.39	0.49	**0.59**	0.35	0.47
(2)	−0.27	**−0.70**	0.56	−0.26	**−0.70**	0.55
(3)	**0.78**	0.12	0.62	**0.79**	0.09	0.64
(4)	**0.60**	0.20	0.40	**0.59**	0.19	0.39
(5)	**−0.56**	−0.40	0.48	**−0.58**	−0.37	0.47
(6)	−0.28	**−0.63**	0.48	−0.25	**−0.66**	0.49
(7)	**0.76**	0.04	0.57	**0.75**	0.04	0.56
(8)	0.10	**−0.72**	0.53	0.03	**−0.70**	0.49
(9)	−0.30	**−0.63**	0.49	−0.32	**−0.63**	0.49
(10)	**0.48**	0.38	0.37	**0.52**	0.34	0.39
Contribution	2.67	2.33	5.00	2.73	2.21	4.94

Table 21 Matrices of Loadings for SES in the case of $r = 2$

		I	II	III
(1) On the whole, I am satisfied with myself.	1	**−0.69**	−0.02	0.48
	2	−0.02	**−0.56**	0.31
(2) At times I think I am no good at all.	1	**−0.66**	−0.02	0.44
	2	0.02	**0.54**	0.30
(3) I feel that I have a number of good qualities.	1	**0.66**	−0.02	0.44
	2	−0.02	**0.63**	0.40
(4) I am able to do things as well as most other people.	1	**−0.59**	−0.02	0.35
	2	−0.02	**−0.53**	0.28
(5) I feel I do not have much to be proud of.	1	**0.68**	−0.01	0.47
	2	−0.01	**0.60**	0.36
(6) I certainly feel useless at times.	1	**0.63**	0.02	0.40
	2	0.02	**0.57**	0.33
(7) I feel that I'm a person of worth, at least on an equal plane with others.	1	0.01	**−0.66**	0.43
	2	**0.59**	−0.01	0.35
(8) I wish I could have more respect for myself.	1	−0.00	**−0.55**	0.30
	2	**0.42**	0.00	0.17
(9) All in all, I am inclined to feel that I am a failure.	1	**0.65**	−0.01	0.42
	2	0.01	**−0.57**	0.32
(10) I take a positive attitude toward myself.	1	**0.61**	−0.04	0.37
	2	−0.03	**0.54**	0.29
Contribution		3.89	3.32	7.21

主成分には，単調に増加する数量が，第２の主成分には「馬蹄」形の数量が一貫して負荷しており，主成分ⅠとⅡの関係も典型的な非線型関係である。（項目 (7) と (8) で負荷する数量が逆になっているように見えるが，回転による非本質的な順序の逆転によるものであり，重みはこれらの変数では第１次元が「馬蹄」，第２次元が「単調」で，負荷している主成分と一致している。）第１次元と２次の関係にある数量は棄却するという（通常の）扱いによれば，自尊心尺度は１次元であるとみなされるであろう。

このことは，数量化Ⅲ類と主成分分析で拾い上げているデータの変動の側面が異なっていることを示している。反応の極端性を示す数量化で見出される第２の次元は，カテゴリーを数量とみなす観点からは非線型の成分であり，主成分分析では隠されてしまう。しかし，この変動は極めて大きくカテゴリーの順序を無視する数量化では，第２の変動として取り出される。どちらか一方を「正しい」分析であるとすることはできないと考えるが，多数の項目から多様な「意味」を見出そうとする立場からは，主成分分析の方が好ましいように感じられる。

このデータでは，非計量的主成分分析自体は，なんら新たな発見をもたらさなかったが，同一のデータを量的なものとして扱う主成分分析と，質的なものとして扱う数量化Ⅲ類，あるいは同等の方法による結果を比較するための視点を提供した点で，この適用自体には意味があったと言えるであろう。

6．討　論

6．1　結果の要約

3つのデータ・セットについて，拡張された非計量的主成分分析を適用した結果を示してきた．ここではまず，得られた結果を要約的に示しておこう。

第１の飲み物の評定のデータでは，１つの変数に複数次元の数量を割り当てることの効果が，明らかになったと思われる。そこで用いられていた「原材料」，「味」，「色」といったカテゴリーは自然な順序を持たないものであり，多次元的に表現することが適切である。結果には，おおよそ納得のいく解釈が可能であった。

第２の有識者による地域の開発の主体の選択に関するデータは，従来数量化によって，比較的少次元で扱われたと思われる順序のないカテゴリーをもつ項目群に，主成分分析的 feature を通じて，比較的明快な多次元的構造が見出された例と言えるであろう。これは，既成の統計パッケージに含まれるレベルの非線形主成分分析の方法が，数量化，重対応分析系統の方法の代替案として十分考慮に値することを示していると思われる。

第３の自尊心尺度のデータでは，非計量的主成分分析自体は，必ずしも興味ある新たな結果を生み出したわけではないが，順序のついたカテゴリーをそのまま数値的に扱って主成分分析（因子分析）することと，順序のないカテゴリーとして数量化Ⅲ類，対応分析系の方法で分析した場合との関係の一端を明らかにすることができたと思われる。２値データの場合には，これらの方法の結果は一致することが証明されており（たとえば，山田・西里，1993），問題は，解釈に構造（負荷量）行列を用いるか重み行列を用いるか，あるいは，回転を行うかどうか，という一種の事後分析の違いに帰されるが，３つ以上のカテゴリーをもつデータについて，そのような単純な定理は成立しない。むしろ，今回の結果は，同一の次元数であっても，２つの方法は多次元データの異なる側面を強調しているようことがわかった。もちろん，こうした特性はデータによって異なった相違を生み出すであろうから，今後も，いろいろなデータで同種の分析を行って比較してみる必要がある。

また，３つの適用例を通じて，直交回転を通じて単純構造に近づけられた負荷行列を解釈に用いることの効果は明らかになったと思われる。このことは，本論文で扱った制約のついたモデルだけでなく，制約のない一般の対応分析，あるいは，アイテム・カテゴリー型のデータの数量化Ⅲ類においても，試みられてよいことだと思われる。

6．2　残された問題

以上のように，拡張された非計量的主成分分析の一定の実用性は，確認されたと思われるが，この方法がさらに広範に利用されるためには，なお解決すべき問題が残されていると思われる。

第１に，各項目に与える数量の次元をどのように決定するべきかという問題がある。これについて，本論文では，モデルが与える説明力自体と，同じ次元数 r で実施された同質性分析（アイテム・カテゴリー型データの数量化Ⅲ類）への効率という２つの指標を示したが，ともに解を１つに絞るには十分と言えず，決定は解釈に依存するものにならざるを得ないであろう。特に，主成分得点間の「馬蹄」には，余分な次元を抽出しないために注意が必要である。しかし，１つの数値についての決定である通常の主成分分析の場合と異なり，項目ごとに異なる次元数を決めることができるこの方法では，困難が格段に大きいことは覚悟しなければならないであろう。

第２に，カテゴリーの合併の基準の問題がある。この方法では，選択数０のカテゴリーは抹消されるが，そのことは解に実質的な影響を与えない。しかしながら，反

応数が余り少ないカテゴリーは解に悪影響を与える懸念があり，合併は必須であろう。ただし，前述のように，重みそのものが解釈される数量化の場合と異なり，負荷量を主に解釈の対象とするこの方法では，さほどこの点に神経質になる必要はないかもしれない。適用例2での扱いは，幾分慎重でありすぎたかもしれず，カテゴリーを全く合併しない分析も試みる余地があったかもしれない。

第3に，負荷量の双方向からの回転については，引き続き検討する余地がある。本論文の簡便法でも，多くの場合，解釈の容易な単純構造は得られたが，なお改善の余地はあるからである。

6.3 結　語

なお，改善と検討の余地はあるとは言え，ここで紹介した，「複数次元の数量化を伴う非計量的主成分分析」は，心理測定的研究において有望な実用化可能性をもつ方法と言うことができるであろう。質問項目の作成に当たって，従来，後の分析を容易にするためだけのために，質問の内容からすればいささか無理な数量的反応を求めることがないとは言えなかったように思われるが，そのような場合には，むしろ質的な反応をとるようにすべきであろう。その意味で，この方法の含意は得られたデータの単なる分析にはとどまらない含意をもつ可能性もあると考えられる。

なお，量的なカテゴリーをもつ項目と質的反応選択肢をもつ項目が混在している場合のために，量的カテゴリーについては，数量化を行わず，質的項目のみを数量化する形に方法を修正することは容易である。実際の研究への適用は，まずそのようなケースから始めるべきかもしれない。

文　献

Bekker, P. and De Leeuw, J. (1988). Relations between variants of non-linear principal component analysis. In J.L.A. van Rijckevorsel & J. de Leeuw (Eds.) *Component and Correspondence Analysis: Dimension Reduction by Functional Approximation.* New York: Wiley. Pp.1-31.

De Leeuw, J. & Van Rijckevorsel, J.L.A. (1988). Beyond homogeneity analysis. In J.L.A. van Rijckevorsel & J. de Leeuw (Eds.) *Component and Correspondence Analysis: Dimension Reduction by Functional Approximation.* New York: Wiley. Pp.55-80.

Gifi, A. (1990). *Nonlinear Multivariate Analysis.* New York: Wiley.

Greenacre, M.J. (1984). *Theory and Applications of Correspondence Analysis.* London: Academic Press.

岩坪秀一（1987）数量化法の基礎　朝倉書店

Kiers, H.A.L. (1989). *Three-way Methods for the Analysis of Qualitative and Quantitative Two-way Data.* Leiden: DSWO Press.

Kiers, H.A.L. (1991). Simple structure in component analysis techniques for mixture of qualitative and quantitative variables. *Psychometrika,* **56**, 197-212.

Kiers, H.A.L. (1997a). Three-mode orthomax rotation. *Psychometrika* (in press).

Kiers, H.A.L. (1997b). Personal communication.

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

Murakami, T., Kiers, H.A.L., and Ten Berge, J.M.F. (1997). Nonlinear principal component analysis of categorical variables with multiple quantifications. Unpublished manuscript.

Nishisato, S. (1994). *Elements of Dual Scaling: An Introduction to Practical Data Analysis.* Hilsdale: Lawrence Erlbaum.

大津　起夫（1993）OSMOD とその拡張：人工データによる特性の検討　行動計量学，**20**，No.2, 9-23.

Rosenberg, M. (1065). *Society and the adolescent self-image.* Princeton: Princeton University Press.

Saito, T. & Ohtsu, T. (1988). A method of optimal scaling for multivariate ordinal data and its extension. *Psychometrika,* **53**, 5-25.

Takane, Y., Young, F.W. & De Leeuw, J. (1978). The principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling features. *Psychometrika,* **43**, 278-281.

Ten Berge, J.M.F. (1988). Generalized approach to the MAXBET problem and the MAXDIFF problem, with application to canonical correlations. *Psychometrika,* **53**, 487-494.

Ten Berge, J.M.F. (1993). *Least Squares*

Optimization in Multivariate Analysis. Leiden: DSWO Press.
土屋隆裕（1995）項目分類のための数量化法　行動計量学，22，95-109.
鷲尾泰俊・大橋靖雄（1989）多次元データの解析　岩波書店
山田文康・西里静彦（1993）双対尺度法に関するいくつかの特性 —— 2値形式のアイテム・カテゴリー型データに対する適用 —— 行動計量学，20，No. 1，56-63.

（1997年9月16日　受稿）

ABSTRACT

Applications of Nonmetric Principal Components Analysis for Categorical Data

Takashi MURAKAMI

An extended method of nonmetric (nonlinear) principal components analysis (PCA) for (unordered) categorical data, which gives multiple quantifications to categories of an item maximizing the size of explained variances by the factor analytical model with the specified number of components, is applied to three data sets; ratings of 34 drinks on 7 categorical items, choices of the most important agent from the list including 9 candidates by 274 experienced persons in terms of promoting developments of local districts on 17 domains, and self ratings of 1630 university students on 10 self-esteem items with 4 ordered categories. The main findings are as follows: (1) multidimensional quantifications of an item by the new method described some interesting features of drinks in the first data set which could not be found by the unidimensional quantification. (2) Nonmetric PCA gave a compact and interpretable representation of the five dimensional structure of nominal variables in the second data set. (3) Relationships between results of Multiple Correspondence Analysis of ordered categorical variables, which is an unconstrained version of nonmetric PCA, and those of classical PCA treating the same variables as interval scales were clarified by nonmetric PCA of the third data set. (4) Rotational freedom of the model and availability of the matrix of loadings in the method facilitated interpretations greatly through the three cases of applications.

Key words; categorical data, principal component analysis, nonlinear multivariate analysis, quantification of qualitative data, multiple correspondence analysis