統計数理 (2000)
第48巻 第1号 121-145

UNISCALによる「日本人の国民性調査」データの分析

統計数理研究所　土　屋　隆　裕

(受付 1999年12月1日；改訂 2000年3月1日)

要　旨

本論文では，調査項目の最適尺度変換を行いつつ，同時に一次元性のある項目だけを選び出す方法であるUNISCALを提案する．UNISCALの基本的な考え方は土屋 (1996) で提案した．これを改良したUNISCALの特徴は，(1)項目を選び出す機能を持つパラメタ α_i を4種類用意し，様々な性質のデータセットに対応できること，(2)パラメタ k の値を変えることで選び出す項目の一次元性の程度を調整できること，(3)欠損値を含むデータセットに対応できること，である．人工データを使って数量化III類とUNISCALの結果を比較し，UNISCALの持つ特徴を説明する．また「日本人の国民性」第10次全国調査データに対しUNISCALを適用した結果を示す．

キーワード：UNISCAL，日本人の国民性調査，一次元尺度，数量化III類，等質性分析．

1. はじめに

1.1 本論文の目的

本論文の目的は，調査項目の最適尺度変換を行うと同時に一次元性のある項目だけを選び出す方法であるUNISCAL (**UNI**-dimensional **SCAL**ing) を提案し，「日本人の国民性調査」データへの適用例を示すことである．

項目の内容が多岐にわたる調査では，内容的に無関係な調査項目を同時に分析の対象とすることは，分析結果を読み解きにくくするため，必ずしも賢明な方法とは言えないことがある．例えば図1は，「日本人の国民性」第10次全国調査M型調査票に用いられた全ての調査項目のうち，複数回答の項目 (#4.16と#5.25) を除いた62項目，全223カテゴリに対して数量化III類を行い，その第I軸と第II軸を用いてカテゴリをプロットした図である．図が煩雑になるため各点に対応するカテゴリ名は表示していないが，この図から何らかの傾向を読み取ろうとすることは難しい．

「日本人の国民性調査」の内容は，基本属性や個人的態度に関するものから日本人・人種に関するものまで幅広く，数量化III類は，それら全てをまとめて分析の対象とすることに耐え得る方法ではないのである．むしろ，目的に照らし内容の上であるいはデータの上で関連性のある調査項目だけを取り出し分析の対象とする方が，豊かな知見が得られることが多い．

図2は，前記M型調査票のうち内容的に関連する“§6男女の差異”に関する項目だけを取り出し，数量化III類を行った結果である．‘#1.1 性別’と‘#6.2 男・女の生まれかわり’，‘#6.2d 楽しみどちらが多いか’，‘#6.2e 男の子と女の子’の各カテゴリは横方向に散らばっているのに対

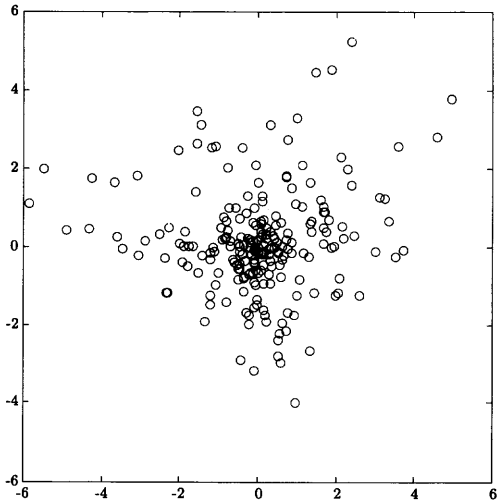

図1. M型調査項目の数量化III類の結果.

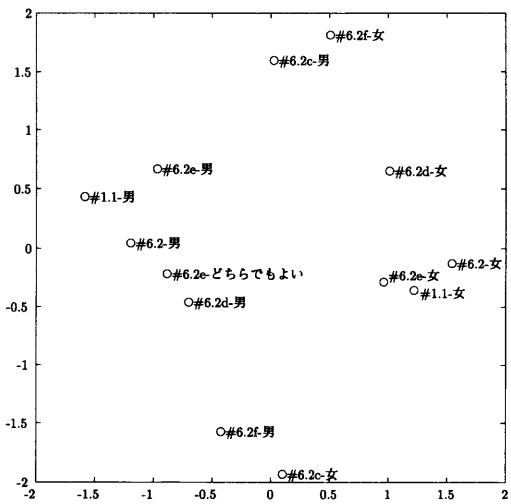

図2. M型調査項目のうち「男女の差異」に関する項目の数量化III類の結果.

し，'#6.2c 苦労どちらが多いか' と '#6.2f どちらがトクか'の各カテゴリは縦方向に分かれている．すなわち，'#6.2c 苦労どちらが多いか' と '#6.2f どちらがトクか' に対する回答は，'#1.1 性別' や '#6.2 男・女の生まれかわり' とは無関係であり，苦労が少ないから，あるいはトクであるから，といったことが生まれかわりたい性別を決める要因ではないことが読み取れる．

調査の設計段階では，何らかの仮説に基づき，図2で例示したような内容的に関連する項目を複数設けておくことがよくある．その場合，データが得られた段階で，それら内容的に関連する項目だけを取り出し，仮説に基づくモデルの検証を行う等の分析対象とすることになる．

その一方で，一見内容的には関連がないように思われる項目どうしが，データの上では関連しているといった知見が仮に得られれば，そこから新たな仮説が生まれる可能性もある．調査の設計段階の意図を考慮せず，探索的にデータ構造を見出そうとすることもまた意義深いこと

と言える．本論文では，一次元構造に着目し，どのような項目が一次元構造を成しているのか探ることとする．全ての調査項目の中から一次元構造を持つ項目だけを取り出すことで，多数の調査項目を分類・整理し，今後どのような項目群に着目して分析を進めればよいか，といった示唆が得られる．また，共分散構造モデル等さらに複雑なモデルを考える手がかりも得られると期待できる．

1.2　一次元構造の抽出法

量的データ，すなわち項目が間隔尺度や比尺度であれば，前述の目的を達する1つの方法は，単に項目間の積率相関係数を調べ，それが高い項目を選び出すことである．ところが「日本人の国民性調査」に代表される社会調査では，項目が名義尺度であることがほとんどである．この場合，χ^2値やKendallのτ_bといった名義尺度や順序尺度の連関係数（Liebetrau（1983））を調べるのでは不十分なだけでなく，誤った結論を導く可能性がある．これらの係数は，どのカテゴリどうしが関連しているのか，といったカテゴリの方向性に関して何ら情報を与えないからである．

例えば，項目Aについては，年齢が高くなるほどカテゴリA-1が選ばれ，年齢が低くなるほどカテゴリA-2が選ばれるとする．また，項目Bについては，年齢が低いか高いとカテゴリB-1が選ばれ，中間的な年齢ではカテゴリB-2が選ばれるとする．年齢と項目A，年齢と項目Bはどちらも連関係数が高くなるため，係数の値だけから判断すれば，年齢と項目Aと項目Bは全て1つのグループとして一次元性を持つと見なされる恐れがある．この例の場合は，項目Aと項目Bは関連しているとは言えないため，年齢と項目A，年齢と項目Bという2つのグループがあると考えるべきである．

CATDAP（坂元（1985））は，関連性のある項目を自動探索する方法であるが，やはりカテゴリの方向性については考慮していない．またAICを用いて分割表の比較をするため，'D.K.'等を欠損値として扱う場合には，それを含むサンプルは取り除く等の処理が必要となる（坂元（1981））．

主成分分析や数量化III類も，分析対象の項目を全て統合し1つの尺度値を与えるという意味では，一次元尺度を見出そうとする方法である．しかしそれらの方法により得られた尺度値は，全ての項目を統合して得られたものであり，どの項目間の関連が強いのかといったことを判断する根拠とはなり難い．関連の強い項目だけを選択するという機能が欠けているのである．

因子分析モデルには，相関の高い一部の項目だけを選択する機能がないため，通常因子分析においては，単純構造への因子の回転によってその機能を補っている．それに対し，図示に頼る数量化III類では，特に項目の数が非常に多い場合，一次元性を持つ項目の選択が不可能であることは，冒頭図1で示した通りである．項目が一次元性を持つ場合，カテゴリの布置は馬蹄形を示す．しかし逆に，馬蹄形に近い形が見出されたとしても，図に頼る限りどの項目を選び出すべきかという判断の基準はあいまいである上，必ずしもそれらの項目が一次元性を持つわけではないことは，土屋（1995，1996）で示されている．

土屋（1996）では，最適尺度変換を行いながら，一次元性を持つ項目を選び出す方法を提案している．この方法は，カテゴリの最適変換値が得られるため，カテゴリの方向性に関して情報が得られる．さらに項目選択の基準を図に頼る必要がないという利点も持つ．その一方で，欠損値を扱うことができない，方法の名称がない，といった不十分な点もある．本論文では，土屋の考え方を改良することで，項目選択の基準としていくつかのオプションを加えた方法を提案し，これをUNISCALと呼ぶこととする．UNISCALでは欠損値への対応も可能である．

本論文の構成は以下の通りである．まず，2章ではUNISCALの考え方と方法について述べる．3章では人工データを用いてUNISCALの特徴について例示する．4章ではUNISCALを

使った「日本人の国民性調査」データの分析例を示す．最後に5章ではUNISCALの適用に当たっての注意点等をまとめる．

2. UNISCALの方法

本章では，まず欠損値のない場合のUNISCALの方法について説明する．次に欠損値を含むデータに適用できるよう拡張したUNISCALの方法を提案する．最後に計算方法を述べる．

2.1 UNISCAL（欠損値のない場合）

数量化III類あるいは尺度変換を伴った主成分分析は，$\boldsymbol{m}$が平均0，分散1といった制約条件の下で，次式を最小化する尺度変換f_iおよび$\boldsymbol{m}$を求めることに等しい (Gifi (1990))．

$$S_1(f_i, \boldsymbol{m}) = \sum_{i=1}^{I} \|f_i(\boldsymbol{D}_i) - \boldsymbol{m}\|^2 = N\sum_{i=1}^{I} e_i^2 \tag{2.1}$$

ここで，Nはサンプル数，Iは項目数である．$\boldsymbol{D}_i$ $(i = 1, ..., I)$は項目iのデータ行列であり，$f_i(\boldsymbol{D}_i)$は尺度変換されたデータベクトルである．例えば，項目iが名義尺度の水準にあるならば，$\boldsymbol{D}_i$をダミー変数行列として，$f_i(\boldsymbol{D}_i) = \boldsymbol{D}_i\boldsymbol{w}_i$と表すことができる．また，項目$i$が間隔尺度であれば，$\boldsymbol{D}_i$をデータベクトル$\boldsymbol{d}_i$として，$f_i(\boldsymbol{D}_i) = w_i\boldsymbol{d}_i$となる．

社会調査では，ほとんどの項目が名義尺度や順序尺度である．また順序尺度は，その順序を考慮せず名義尺度として扱った方がよいことが多い．例えば年齢を順序尺度として扱うと，加齢とともに単調に増加するあるいは減少する効果しか捉えられないことになる．そこで以後は，$f_i(\boldsymbol{D}_i) = \boldsymbol{D}_i\boldsymbol{w}_i$とする．

(2.1)式で求まる$\boldsymbol{m}$は，I個全ての項目を統合した尺度値ということになる．I個全ての項目にわたって，$f_i(\boldsymbol{D}_i)$と$\boldsymbol{m}$との差を最小にしようとしているからである．$f_i(\boldsymbol{D}_i)$間の相関が低い項目どうしは，内容的にも異質であることが多く，それらを統合して得られた尺度値$\boldsymbol{m}$は，意味の解釈が困難となることもしばしばである．

一方，本論文の目標はお互いに関連の高い一部の項目のみを選び出し，それらを用いて尺度値$\boldsymbol{m}$を構成することである．すなわち，選択すべき一部の項目についてのみ$f_i(\boldsymbol{D}_i)$と$\boldsymbol{m}$との差を小さくしなければならない．

仮にどの項目を使うべきかあらかじめ知られている場合には，α_i (≥ 0)を項目iが$\boldsymbol{m}$の構成に寄与する程度を表す指標として，

$$S_2(\boldsymbol{w}_i, \boldsymbol{m}) = \sum_{i=1}^{I} \alpha_i\|\boldsymbol{D}_i\boldsymbol{w}_i - \boldsymbol{m}\|^2 = N\sum_{i=1}^{I} \alpha_i e_i^2 \tag{2.2}$$

を$\boldsymbol{w}_i$と$\boldsymbol{m}$に関して最小化することになる．ただし，$\boldsymbol{m}$に対する制約条件は

$$\|\boldsymbol{m}\|^2 = N, \qquad \mathbf{1}'\boldsymbol{m} = 0$$

である．項目jが$\boldsymbol{m}$の構成に寄与すべき場合，α_jが大きいため，(2.2)式を最小化するためには対応する$\|\boldsymbol{D}_j\boldsymbol{w}_j - \boldsymbol{m}\|^2$の値を小さくしなければならず，$\boldsymbol{m}$は$\boldsymbol{D}_j\boldsymbol{w}_j$との相関が高いものとなる．逆に項目$\ell$が$\boldsymbol{m}$の構成に寄与すべきでない場合，$\alpha_\ell$は小さいため$\|\boldsymbol{D}_\ell\boldsymbol{w}_\ell - \boldsymbol{m}\|^2$の値を小さくする必要はなく，$\boldsymbol{D}_\ell\boldsymbol{w}_\ell$と$\boldsymbol{m}$との相関は高くならない．西里 (1982) は，あらかじめ分析者が一部の項目に対してのみ重みを与えることで，その項目との関連が高い項目を探し出すこのような方法を強制分類法と呼んでいる．

一方ここでの目的は，α_iがあらかじめ知られておらず，むしろデータからこれを求めることである．そこで，α_iを$\alpha_i \geq 0$を満たす$e_i^2 = \|\boldsymbol{D}_i\boldsymbol{w}_i - \boldsymbol{m}\|^2/N$の減少関数

(2.3)　$$\alpha_i = g(e_i^2)$$

とする．(2.2) 式の α_i として (2.3) 式を用いる方法を UNISCAL と呼ぶこととする．減少関数 $g(\cdot)$ として本論文では次の 4 種類を示す．

(2.4)　$$g_1(e_i^2) = (1 - e_i^2)^k$$

(2.5)　$$g_2(e_i^2) = \left(\frac{1 - e_i^2}{1 + e_i^2}\right)^k$$

(2.6)　$$g_3(e_i^2) = \exp(-ke_i^2)$$

(2.7)　$$g_4(e_i^2) = (-\log(e_i^2))^k$$

k は正の実数であり，分析者があらかじめ与える値である．$0 \leq e_i^2 \leq 1$ であるので上記のどの $g(\cdot)$ を用いても $\alpha_i \geq 0$ は満たされる．なお，(2.7) 式の g_4 は $e_i = 0$ のとき定義できないが，普通 $e_i = 0$ となることはない．

一次元性を持つ項目のみを取り出すと言っても，実際のデータでは項目間相関が完全に 1 となることはあり得ない．どの程度までの関連を許容し，項目を選択するかということは分析者が決めなければならない．UNISCAL では，パラメタとして k を導入することでこのような要請に応えている．

このことを示すため，4 つの $g(\cdot)$ ごとに k の値を 0.1 から 5 まで変え，e_i^2 (横軸) と $\alpha_i e_i^2$ (縦軸) の関係を図示したものが，図 3 である．なお，縦軸のスケールは適宜修正してある．

k の値が小さい，例えば $k = 0.1$ の場合，基本的に曲線は右上がりであり，(2.2) 式の $\alpha_i e_i^2$ の値を小さくするためには e_i^2 そのものの値を小さくしなければならない．このため，得られる $\boldsymbol{m}$ はなるべく多くの項目との相関が高いものとなる．すなわち，多少関連が低い項目をも取り込むことになる．実際，$k = 0$ のとき (2.2) 式は (2.1) 式に一致する．一方 k の値が大きくなると，$\alpha_i e_i^2$ の値を小さくするためには，e_i^2 の値を小さくできる場合には小さくし，逆に小さくできない場合には大きくすればよい．このため，$\boldsymbol{m}$ は一部の項目とだけ相関が高いものとなり，内的一貫性の高い少数の項目だけが選択されることになる．

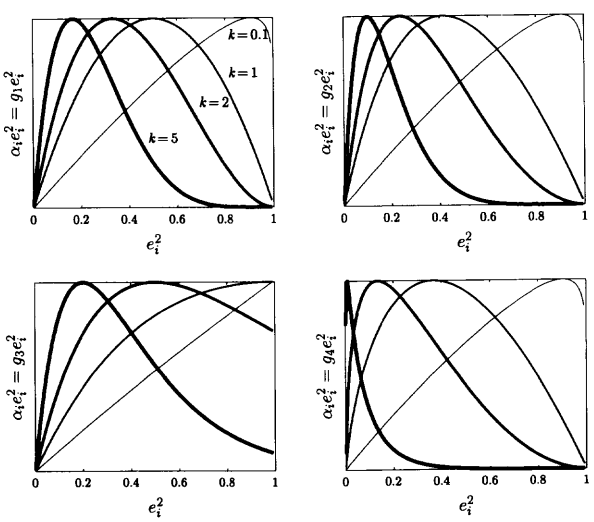

図 3.　4 種類の α_i.

図3では，用いる α_i の種類によって曲線の形が異なる．k の値が同じならば，g_1 に比べ g_2 や g_4 はピークが左に寄っており，g_3 は右に寄っている．したがって，g_3 を用いるとより多くの項目が取り込まれ，g_2 や g_4 ではより少数の項目だけが選択されることになる．土屋 (1996) の方法では，α_i として g_1 のみを用いている．UNISCAL では，g_1 から g_4 まで4つのオプションを選択できるようにし，項目選択の柔軟性を増した．このことについては，次章で分析例を示しながらさらに説明する．

2.2 UNISCAL（欠損値のある場合）

「日本人の国民性調査」では，調査項目への回答として 'その他' や 'D.K. (Don't Know)' がある．これらをカテゴリの1つとして扱うと，'その他' と 'D.K.' 対その他のカテゴリという一次元尺度が得られやすい．すなわち，提示されたカテゴリの中から回答を選ぶ人であるか否か，といったことを測る尺度ができてしまう．このような尺度は，現実にはあまり意味がなく，'その他' や 'D.K.' は欠損値として扱う方がよい．

(2.2) 式では，全てのサンプルの $\boldsymbol{D}_i\boldsymbol{w}_i$ と $\boldsymbol{m}$ との差の2乗和を考えた．欠損値がある場合には，欠損であるサンプルを除いて2乗和を求めればよい．そのために $N \times N$ の指示行列 $\boldsymbol{\Delta}_i$ を導入する (Meulmann (1982))．$\boldsymbol{\Delta}_i$ は個体 n が項目 i について欠損値でなければ1，欠損値であれば0を第 nn 対角要素とする対角行列である．このとき (2.2) 式に対応する基準式は以下で与えられる．

$$S_3(\boldsymbol{w}_i, \boldsymbol{m}) = \sum_{i=1}^{I} \alpha_i \|\boldsymbol{\Delta}_i(\boldsymbol{D}_i\boldsymbol{w}_i - \boldsymbol{m})\|^2 \tag{2.8}$$

ただし，$\boldsymbol{m}$ に対する制約条件は

$$\sum_{i=1}^{I} \alpha_i \|\boldsymbol{\Delta}_i\boldsymbol{m}\|^2 = \sum_{i=1}^{I} \alpha_i \|\boldsymbol{\Delta}_i\mathbf{1}\|^2$$

$$\mathbf{1}'\sum_{i=1}^{I} \alpha_i \boldsymbol{\Delta}_i\boldsymbol{m} = 0$$

となる．また，(2.3) 式の e_i^2 としては次式を用いる．

$$e_i^2 = \frac{1}{\boldsymbol{m}'\boldsymbol{\Delta}_i\boldsymbol{m}} \|\boldsymbol{\Delta}_i(\boldsymbol{D}_i\boldsymbol{w}_i - \boldsymbol{m})\|^2 \tag{2.9}$$

$0 \leq e_i^2 \leq 1$ となることは，次節で示す．$\boldsymbol{\Delta}_i = \boldsymbol{I}$ とおけば，以上の式は全て欠損値のない場合の式に一致する．

2.3 UNISCAL の計算方法

$\boldsymbol{w}_i$ と $\boldsymbol{m}$ は，$\boldsymbol{\Delta}_i\boldsymbol{D}_i = \boldsymbol{D}_i$ であることを用いれば，以下の反復計算により求められる．欠損値のない場合は，$\boldsymbol{\Delta}_i = \boldsymbol{I}$ とする．

Step 0. $\boldsymbol{m}$ に初期値を与える．

Step 1. $\boldsymbol{w}_i$ を求める．

$$\boldsymbol{w}_i = (\boldsymbol{D}_i'\boldsymbol{D}_i)^{-1}\boldsymbol{D}_i'\boldsymbol{m} \tag{2.10}$$

Step 2. (2.4) 式から (2.7) 式のいずれかにより α_i を求める．

Step 3. $\boldsymbol{m}$ を求める．

$$\boldsymbol{y}=\left(\sum_{i=1}^{I}\alpha_i\boldsymbol{\Delta}_i\right)^{-1}\sum_{i=1}^{I}\alpha_i\boldsymbol{D}_i\boldsymbol{w}_i-\frac{\mathbf{1}'\sum_{i=1}^{I}\alpha_i\boldsymbol{D}_i\boldsymbol{w}_i}{\mathbf{1}'\sum_{i=1}^{I}\alpha_i\boldsymbol{\Delta}_i\mathbf{1}}\mathbf{1} \tag{2.11}$$

$$c=\left(\frac{\mathbf{1}'\sum_{i=1}^{I}\alpha_i\boldsymbol{\Delta}_i\mathbf{1}}{\boldsymbol{y}'\sum_{i=1}^{I}\alpha_i\boldsymbol{\Delta}_i\boldsymbol{y}}\right)^{1/2} \tag{2.12}$$

とすると

$$\boldsymbol{m}=c\boldsymbol{y} \tag{2.13}$$

Step 4. 収束していなければ Step 1 へ.

$0\leq e_i^2\leq 1$ となることは，Step 1 より

$$e_i^2=\frac{1}{\boldsymbol{m}'\boldsymbol{\Delta}_i\boldsymbol{m}}\boldsymbol{m}'(\boldsymbol{\Delta}_i-\boldsymbol{D}_i(\boldsymbol{D}_i'\boldsymbol{D}_i)^{-1}\boldsymbol{D}_i')\,\boldsymbol{m}$$
$$=\frac{1}{\boldsymbol{m}'\boldsymbol{\Delta}_i'\boldsymbol{\Delta}_i\boldsymbol{m}}\boldsymbol{m}'\boldsymbol{\Delta}_i'(\boldsymbol{I}-\boldsymbol{D}_i(\boldsymbol{D}_i'\boldsymbol{D}_i)^{-1}\boldsymbol{D}_i')\,\boldsymbol{\Delta}_i\boldsymbol{m}$$

であり，射影行列 $\boldsymbol{I}-\boldsymbol{D}_i(\boldsymbol{D}_i'\boldsymbol{D}_i)^{-1}\boldsymbol{D}_i'$ の固有値 λ が $0\leq\lambda\leq 1$ であることから明らかである.

この計算方法により求まる解は一意に定まるとは限らない．Step 0 の初期値を変えることで，いくつかの局所解が得られる可能性がある．しかしその中から唯一最適な解というものを選ぶ必要はない．分析の対象とする項目の中には，一次元性を持つ項目のグループが複数存在するのが普通であり，ある初期値により得られた解は，その中の1つに過ぎないからである．したがって，初期値を様々に変えて得られた全ての解が，考察の対象となるべきである．なお，初期値は最終解に近くなるよう設定する必要はない．次章以降では，Step 0 において一様乱数を発生させ，それを制約条件に合うよう基準化したものを $\boldsymbol{m}$ の初期値とした.

得られる解の数は，データと α_i 及び k の値に依存する．k の値が 0 に近ければ，相関の低い項目も一次元性を持つ項目と見なすことになるため，多くの場合1つの解しか得られない．逆に k の値を大きくすれば，極端な場合，一次元性を持つ項目はその項目のみということになり，項目の数だけ解が得られる．さらに異なる尺度変換を考えれば，項目の数よりも多い数の解が得られる．このことは次章で人工データを用いて詳解する.

3. UNISCALの分析例

UNISCALの特徴を例示するため，人工データによる分析例を紹介する.

3.1 扱うデータ

まず，平均 **0**，分散共分散行列 $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma}=\begin{pmatrix}
1.0 & .7 & .7 & .7 & & & & & & \\
.7 & 1.0 & .7 & .7 & & & & & & \\
.7 & .7 & 1.0 & .7 & & & & & & \\
.7 & .7 & .7 & 1.0 & & & & & & \\
 & & & & 1.0 & .7 & .7 & .6 & & \\
 & & & & .7 & 1.0 & .7 & .6 & & \\
 & & & & .7 & .7 & 1.0 & .6 & & \\
 & & & & .6 & .6 & .6 & 1.0 & .7 & \\
 & & & & & & & .7 & 1.0 & \\
 & & & & & & & & & 1.0
\end{pmatrix}$$

を用いて100個の10変量正規乱数を発生させる．ただし，Σの空白部分には共分散として0.2を与えた．すなわち，10の変数が，変数1から変数4，変数5から変数9，変数10のみという大きく3つのグループに分かれる構造を持つ．さらに変数5から変数9は，変数5から変数7と変数8から変数9という2つのグループに分かれ，変数8は変数5から変数7とも相関があることになる．

次に，各変数ごとに発生した乱数の値を小さい方から順に並べ，3つにカテゴリ化する．その際，各カテゴリの頻度はほぼ等しくなるようにした．このデータセットを人工データ1と呼ぶこととする．

3.2 数量化III類の場合

まず，人工データ1に対する数量化III類の結果を示す．表1は，(2.1)式の$f_i(\boldsymbol{D}_i)$と$\boldsymbol{m}$との相関係数を第I軸から第IV軸まで示したものである．0.50以上の相関係数は太字で示した．

第I軸はどの変数とも相関が高く，変数のグループは見出せない．第II軸は，変数5から変数8の相関は0.50以上であるものの他の変数も0.50に近く，変数10だけが異質であるということが分かるのみである．第III軸は，変数1から変数4の相関がやや高く，これらがグループ

表1. 人工データ1の数量化III類の結果．

		第I軸	第II軸	第III軸	第IV軸
$\boldsymbol{m}$との相関係数	変数1	**0.70**	0.43	**0.55**	0.24
	変数2	**0.65**	0.49	0.49	0.20
	変数3	**0.59**	0.47	**0.61**	0.11
	変数4	**0.66**	0.46	**0.62**	0.20
	変数5	**0.56**	**0.56**	0.24	**0.61**
	変数6	0.49	**0.56**	0.27	**0.56**
	変数7	**0.50**	**0.61**	0.30	0.41
	変数8	**0.61**	**0.68**	0.29	0.17
	変数9	0.49	0.48	0.46	**0.51**
	変数10	**0.58**	0.04	0.10	0.49

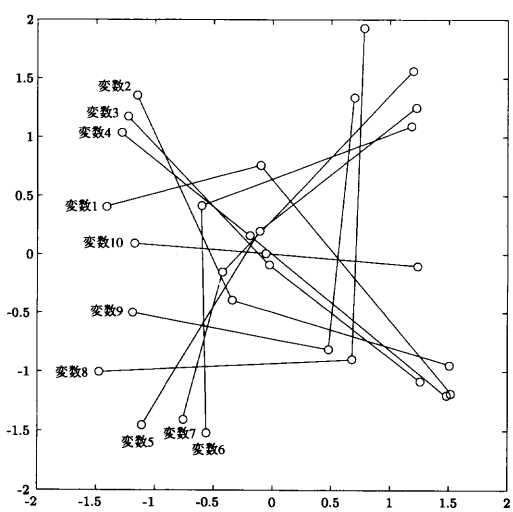

図4. 人工データ1の数量化III類の結果（第I軸と第II軸）．

になっていることが分かる．第IV軸では，変数5と変数6，変数9の相関が0.50以上であるが，変数7と変数8は相関が低くなってしまっている．

第IV軸までの結果を見る限り，本来の$\boldsymbol{\Sigma}$の構造についてはほとんど分からなかったと言える．仮に第V軸以降に，適切な変数のグループが現れていたとしても，そもそものデータ構造を知らない分析者にとっては，どの軸を採用すべきかは全く分からない．

数量化III類は，各カテゴリに複数の数値を与える多次元尺度法の1つであるため，結果の解釈に当たっては，図を用いる方が適切である．そこで，第I軸と第II軸のカテゴリ数量をプロットした結果を図4に示す．図では同じ変数の3つのカテゴリを順に線で結んである．

図4から，変数1から変数4，変数5から変数7，変数8と変数9，変数10のみ，という大きく4つのグループがあることが読み取れる．しかし，変数1や変数6については必ずしも当該グループにまとめるべきか否か迷うところである．また，変数8が変数5から変数7ともやや相関があるといったデータ構造は読み取れない．第I軸と第II軸を用いた結果だけでは，データ構造は十分に把握しきれないことになる．仮に3つ以上の軸を用いればデータ構造が分かるとしても，図に頼る数量化III類では同時に表現できる軸はせいぜい3つまでに限られる．さらに，カテゴリの数が増えると変数間の関連を読み取ることはほとんど不可能となる．

3.3 UNISCALの場合

次に，人工データ1に対して，UNISCALを行った結果を示す．ただし，α_iとしては(2.4)式の

$$\alpha_i = g_1(e_i^2) = (1 - e_i^2)^k$$

を用い，$k = 2$とした．初期値をランダムに100回変えて繰り返し計算を行い，得られた結果が表2である．

初期値を変えた結果，最終的に4種類の解が得られた．「回数」の欄に示した数値は，100回のうちそれぞれの解が何回得られたのかを示す．例えば100回のうち34回は，$\boldsymbol{w}_i$と$\boldsymbol{m}$の符号の向きを除いて完全に一致し，解1に示した結果に達したことになる．

表2．人工データ1のUNISCALの結果．

		解1	解2	解3	解4
	回数	34	30	27	9
	r	0.70	0.75	0.82	0.95
$\boldsymbol{m}$との相関係数	変数1	**0.84**	0.24	0.15	0.30
	変数2	**0.79**	0.10	0.25	0.30
	変数3	**0.69**	0.03	0.14	0.31
	変数4	**0.91**	0.14	0.12	0.29
	変数5	0.13	**0.84**	0.44	0.24
	変数6	0.20	**0.70**	0.25	0.21
	変数7	0.12	**0.95**	0.36	0.16
	変数8	0.28	**0.50**	**0.94**	0.29
	変数9	0.20	0.24	**0.91**	0.35
	変数10	0.35	0.21	0.31	**1.00**
$\boldsymbol{m}$間相関	解1	1.0	.13	.10	.35
	解2		1.0	.42	.20
	解3			1.0	.32
	解4				1.0

表3. 人工データ1の解1の $\boldsymbol{w}$.

	カテゴリ		
	1	2	3
変数1	0.83	0.32	−1.16
変数2	0.92	0.09	−1.01
変数3	0.86	−0.03	−0.83
変数4	1.05	0.13	−1.18
変数5	0.13	0.05	−0.18
変数6	−0.02	0.25	−0.23
変数7	0.06	0.10	−0.17
変数8	0.30	−0.37	0.09
変数9	0.25	−0.23	0.00
変数10	0.42	0.01	−0.43

「r」の欄の数値は，各解で得られた $\boldsymbol{w}_i$，$\boldsymbol{m}$ を用いて次式で与えられる．

$$r = 1 - \frac{1}{\sum_{i=1}^{I} \alpha_i \mathbf{1}'\boldsymbol{\Delta}_i\mathbf{1}} \sum_{i=1}^{I} \alpha_i \|\boldsymbol{\Delta}_i(\boldsymbol{D}_i\boldsymbol{w}_i - \boldsymbol{m})\|^2 \tag{3.1}$$

r は，ダミー変数行列において1の代わりに α_i を用いて数量化III類を行ったとき，最大化された相関係数の2乗に一致する（付記参照）．r の値が大きいほど，その尺度の内的一貫性が高いことを示す．

表中の「$\boldsymbol{m}$ との相関係数」は，各変数の最適変換尺度 $\boldsymbol{D}_i\boldsymbol{w}_i$ と $\boldsymbol{m}$ との相関係数を示す．解1の $\boldsymbol{m}$ は，変数1から変数4まで4つの変数との相関係数が高く，これら4つの変数から構成されていることが分かる．解2は，変数5から変数7との相関係数は0.70から0.95と高く，変数8との相関係数は0.50とやや高い．すなわち，変数5から変数7が1つのグループを成し，変数8はこのグループとやや関連がある，ということになる．解3は変数8と変数9の2つだけから成るグループがあることを示す．解4は変数10だけに高い相関係数を示すことから，変数10が他のどの変数とも関連がないことを示している．これらの結果は，データを発生させたときの分散共分散行列 $\boldsymbol{\Sigma}$ の特徴をよく捉えている．

なお，r の値は解4が0.95と最も大きく，解1が0.70と最も小さいが，必ずしも解4が重要で解1が重要でないということにはならない．r は内的一貫性の指標であるため，項目が少ないほどその値は高くなる傾向にある．r の値は解の重要性を示す指標ではないことに注意すべきである．

「$\boldsymbol{m}$ 間相関」は，4種類の $\boldsymbol{m}$ 間の相関係数である．解の間に何ら制約がないため，相関係数が0となることはほとんどない．解2と解3との相関係数が.42とやや高いが，これは変数8が両方の解に含まれているためである．

各解に対応して4種類の最適変換 $\boldsymbol{w}$ が得られるが，ここでは解1に対応する $\boldsymbol{w}$ のみを表3に示す．解1は基本的には変数1から変数4で構成されるため，この4つの変数のみに着目すればよい．4つの変数は全て，カテゴリ1が正の値をとりカテゴリ3は負の値をとっていることから，ある変数がカテゴリ1である個体は他の3つの変数でもカテゴリ1となっていることが分かる．一般に，得られた解が何を意味するのかは，この $\boldsymbol{w}$ を見て判断することになる．

3.4 一部の変数を削除した場合

UNISCALは関連のある変数だけを取り出す方法であるため，ある解とは無関係な変数が

表4．人工データ2の結果．

		解1	解2	解3
	回数	46	38	16
	r	0.70	0.75	0.89
mとの相関係数	変数1	**0.84**	0.24	0.16
	変数2	**0.79**	0.10	0.24
	変数3	**0.68**	0.03	0.10
	変数4	**0.91**	0.14	0.11
	変数5	0.13	**0.84**	**0.55**
	変数6	0.20	**0.70**	0.40
	変数7	0.12	**0.95**	0.48
	変数8	0.28	**0.50**	**1.00**
m間相関	解1	1.0	.13	.11
	解2		1.0	.54
	解3			1.0

データセットから取り除かれても，その解にはほとんど影響がない．このことを例示するため，人工データ1から変数9と変数10を取り除いた8変数のデータセット（人工データ2と呼ぶ）を作成し，UNISCALを行った．

表4はその結果であり，3つの解が得られている．取り除いた変数9及び変数10とは無関係であった解1と解2は，表2の解とほとんど同一の結果が得られている．表2の解3は変数8と変数9の2つから成っていたが，変数9がデータセットから取り除かれたため，表4の解3では，変数8との相関がやや高い変数5から変数7が取り込まれている．

このように，ある解とは無関係な変数をデータセットから取り除いてもその解には影響がない，逆に言えば，ある解とは無関係な変数をデータセットに加えてもその解には影響がないことがUNISCALの特徴の1つである．このため，UNISCALでは分析の対象とする変数の選択にさほど気を遣う必要がない．むしろ，まず全ての調査項目を同時にUNISCALの分析対象とし，そこから得た知見を基にさらに別の方法で分析をすすめる，といった探索的な方法の1つとしては非常に有効であると考えられるのである．

3.5　kの値を変えた場合

(2.4)式から(2.7)式におけるkは，その値が大きいほど関連の強い一部の項目のみを選択するようになる．このことを例示するため，人工データ1に対して，α_iとして(2.4)式を用いたときの$k=0.5$と$k=2.2$の場合の結果を表5及び表6に示す．

kの値が0に近づくほど，なるべく多くの変数を取り込んだグループができる．例えば，$k=0.5$のときには，2つの解が得られており，人工データ1の変数には，大雑把に言えば2つのグループがあることが分かる．変数9は，どちらかと言えば解2のグループに近く，変数10は解1のグループに近いことになる．各解のrの値は，$k=2$とした表2の各解のrに比べて小さく，kの値を小さくすると多くの変数が取り込まれ，結局内的一貫性は下がることが分かる．

逆にkの値を$k=2.2$と大きくした場合，7つの解が得られている．すなわち，より細かく見ていくと，人工データ1の変数には，7つのグループがあることになる．「m間相関」を見れば，解2と解5，解3と解6と解7はそれぞれ同じような変数から成り立つグループである．しかし，例えば変数3と変数4の相関は完全に1ではないため，変数1から変数4までのグループの中にも，より詳細に見れば，変数3を中心としたグループと変数4を中心としたグループがあることになる．表5の各解のrの値が表2に比べ大きくなっていることからも，より内的一

表5. 人工データ1の結果 ($k = 0.5$).

		解1	解2
回数		57	43
r		0.48	0.45
mとの相関係数	変数1	**0.83**	0.32
	変数2	**0.82**	0.19
	変数3	**0.77**	0.10
	変数4	**0.84**	0.18
	変数5	0.18	**0.84**
	変数6	0.24	**0.73**
	変数7	0.15	**0.83**
	変数8	0.35	**0.74**
	変数9	0.28	0.47
	変数10	0.45	0.34
m間相関	解1	1.0	.28
	解2		1.0

表6. 人工データ1の結果 ($k = 2.2$).

		解1	解2	解3	解4	解5	解6	解7
回数		27	25	17	9	8	8	6
r		0.83	0.74	0.79	0.97	0.80	0.83	0.78
mとの相関係数	変数1	0.15	**0.79**	0.22	0.29	**0.63**	0.22	0.28
	変数2	0.25	**0.74**	0.11	0.29	**0.62**	0.05	0.08
	変数3	0.14	**0.64**	0.03	0.31	**0.99**	0.12	0.06
	変数4	0.12	**0.96**	0.14	0.29	**0.63**	0.13	0.13
	変数5	0.44	0.13	**0.79**	0.24	0.04	**0.61**	**0.96**
	変数6	0.26	0.20	**0.68**	0.20	0.19	**0.99**	**0.64**
	変数7	0.36	0.12	**0.97**	0.16	0.13	**0.69**	**0.82**
	変数8	**0.94**	0.28	0.48	0.29	0.19	0.43	**0.51**
	変数9	**0.91**	0.19	0.24	0.35	0.17	0.10	0.24
	変数10	0.31	0.33	0.20	**1.00**	0.33	0.16	0.24
m間相関	解1	1.0	.09	.40	.32	.08	.30	.42
	解2		1.0	.13	.34	.74	.12	.14
	解3			1.0	.19	.05	.78	.92
	解4				1.0	.33	.16	.23
	解5					1.0	.12	.04
	解6						1.0	.73
	解7							1.0

貫性の高い解が得られていることが分かる.

kの値をいくつにすべきか,ということは,分析者がどの程度の詳細さで変数間の関連性を調べたいのか,ということに依存するため,一概には言えない.大雑把な変数の括りを知りたいのであれば,kの値は小さくすればよいし,変数間のより詳細な異同を見たいのであれば,kの値は大きくすべきである.kの値を定める明確な指針が存在しないことは必ずしもUNISCALの欠点ではない.kをある1つの値に定めた結果だけを見るよりもむしろ,kの値を様々に変え,その結果を比較し吟味することで,データの構造をより深く知ることができるからである.

3.6 a_iを変えた場合

a_iとしては(2.4)式のg_1から(2.7)式のg_4まで4種類の形を提案した.a_iの定義の仕方による結果の違いを例示するため,人工データ1を用い,kを0.1から2.2まで0.1ずつ変えたとき,得られる解の数を示したものが図5である.横軸はkの値を表し,縦軸は得られた解の数を表す.図中,例えばg_1の折れ線はa_iとして

$$a_i = g_1(e_i^2) = (1 - e_i^2)^k$$

を用いた場合の結果である.

kの値が同じならば,g_1に比べg_2の方がより多くの解が得られやすい.このことは,図3において,g_1に比べg_2ではグラフの山がより左に寄っていることからも分かる.kの値が同じならばg_2の方が,e_i^2の値が小さい,すなわち内的一貫性の高い項目だけを選別するからである.

図3において,g_3はg_1に比べ右裾が緩やかである.このため,g_3ではg_1に比べなるべく多くの項目を取り込もうとする.その結果,図5に見られるように,g_3ではkの値を大きくしても解の数はなかなか増えない.

逆に図3のg_4では,kの値が大きくなると,グラフの山が急激に左側に寄る.このため,図5のようにkの値を少し大きくしただけで解の数は爆発的に増加する.

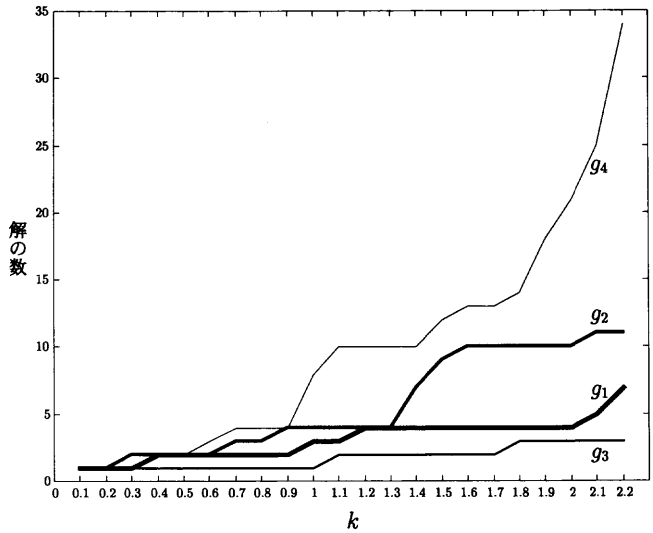

図5. 各 k に対する解の数 (α_i の種類別).

α_i としてどの形を用いるかは，データの状況に依る．経験的には，g_1 で十分なことが多い．しかし，小さな k の値でも解の数が多くなりすぎてしまう場合や，急激に解の数が増えてしまう場合には，g_3 を使えばよい．例えば，g_1 では $k = 2.0$ のとき 4 つの解が得られていたのに，$k = 2.2$ とするだけで 7 つもの解が得られていた．逆に，なるべく細かいグルーピングを行いたいにもかかわらず，k の値を大きくしても解の数が増えない場合には，g_2 や g_4 を用いるべきである．

3.7 欠損値を含む場合

次に，人工データ 1 の (100 × 10) 個のデータの中からランダムに 100 個のデータを欠損値と

表7. 人工データ 3 の結果.

		解 1	解 2	解 3	解 4
	回数	39	29	23	9
	r	0.72	0.75	0.83	0.94
mとの相関係数	変数 1	**0.82**	0.21	0.07	0.29
	変数 2	**0.81**	0.08	0.26	0.31
	変数 3	**0.69**	0.08	0.10	0.29
	変数 4	**0.93**	0.15	0.11	0.28
	変数 5	0.14	**0.94**	0.43	0.29
	変数 6	0.27	**0.70**	0.28	0.33
	変数 7	0.10	**0.87**	0.34	0.22
	変数 8	0.28	**0.53**	**0.93**	0.35
	変数 9	0.13	0.28	**0.93**	0.37
	変数 10	0.33	0.30	0.39	**1.00**
m間相関	解 1	1.0	.11	.08	.33
	解 2		1.0	.42	.30
	解 3			1.0	.39
	解 4				1.0

し(人工データ3と呼ぶ),これに対してUNISCALを適用した.ただし表2との比較のため,α_iは(2.4)式のg_1であり,$k=2$とした.その結果を表7に示す.

表中の「$\boldsymbol{m}$との相関係数」の値は,それぞれ欠損値のデータを除いた$\boldsymbol{D}_i\boldsymbol{w}_i$と$\boldsymbol{m}$との相関係数を表す.若干の数値の違いはあるものの,基本的には表2に示された欠損値のない場合との違いはない.

4. UNISCALによる「日本人の国民性調査」データの分析

UNISCALによる国民性調査データの分析例を示す.対象とした項目は,自由回答及び複数回答を除いた全項目であり,'その他'及び'D.K.'は欠損値として扱った.

4.1 第10次全国調査M型調査票の分析例

まず,1998年に実施された第10次全国調査のうちM型調査票の62項目を用いた結果を示す.なお,このデータセットの数量化III類の結果は図1に示した通りである.計算は,α_iとして(2.4)式のg_1を用い,kの値を様々に変えて行ったが,ここでは$k=0.8$としたときの結果のみを示す.初期値を1000回変えた結果,12個の解が得られた.

以下では,回数の多かった解1から解3と解6,解7を順に示す.解4と解5はともに,'#1.5市郡別'と'#1.6地方別'の相関が大きい解であった.

表9の解1は1000回の繰り返し計算のうち,412回得られたものである.「相関」の欄に示した数値は,各項目の$\boldsymbol{D}_i\boldsymbol{w}_i$と$\boldsymbol{m}$との相関係数であり,表では相関が0.30よりも大きな項目についてのみ結果を示してある.「$\boldsymbol{w}_i$」の欄に示したのは,各カテゴリの最適変換値である.

'#1.2年齢'の各カテゴリに対しては,年齢が上がるほど負の数値が与えられており,'#1.3学歴'では'小学校'や'新制中学'に負の数値が与えられている.すなわち,高齢者の最終学歴は小学校や新制中学であることが多い,ということになる.'#1.4c職業'では'農林水産業'や'学生,無職','#1.2b婚姻・子'では'死別'に負の大きな数値が与えられており,高齢者は農林水産業に従事していたり無職であることが多く,年齢を重ねるにつれ死別が増えることが分かる.勤め先には'ながく勤めるのがよい','家族的な雰囲気'の会社に勤めたい,外国人との結婚には'反対する'といった保守的な意見は,年齢が上がるほど増えることになる.また,高齢者ほど重い病気や戦争に対する不安感を感じるようになる.'#2.30e不安感 失業'に関しては'非常に感じる'と'まったく感じない'に負の数値が与えられており,高年齢ほど,不安感が強い人と同時にもはや退職し失業に対する懸念の不要な人が混じっていることが分かる.

表10の解2は245回得られたものである.不安感に関する7項目と'#1.2年齢'が選択されている.'#1.2年齢'の各カテゴリに与えられる数値は,必ずしも順序通りとはなっていないものの,一般的に高年齢ほど不安感が強いといった傾向が見られる.また,'#1.2年齢'以外には相関が0.30以上の項目がなく,M型調査票の中では不安感に関する項目群は独立していると言える.

表11の解3は207回得られたものである.満足感に関する6項目のうち,'#2.3d社会に満足か'以外の個人的な生活に関する満足感を尋ねた5項目が上位5つを占めている.'#7.29くらし

表8. 第10次M型調査票のUNISCALの結果.

	解											
	1	2	3	4	5	6	7	8	9	10	11	12
回数	412	245	207	50	26	25	22	5	3	2	2	1
r	.31	.40	.27	.52	.51	.46	.25	.42	.42	.44	.49	.57

表9. 第10次M型調査票の解1.

項　目	相関	w_i
#1.2　年齢	0.92	20～24(1.06)　25～29(1.05)　30～34(0.91)
		35～39(0.78)　40～44(0.53)　45～49(0.30)
		50～54(0.10)　55～59(−0.29)　60～64(−0.46)
		65～69(−1.09)　70以上 (−2.02)
#1.3　学歴	0.74	小学校 (−2.24)　新制中学 (−0.86)
		新制高校 (0.18)　大学 (0.72)
#1.4c　職業	0.68	農林水産業 (−1.03)　自営の商工業 (−0.03)
		専門，自由業 (0.85)　管理職 (0.36)
		事務系の勤め人 (0.85)　作業系の勤め人 (0.30)
		主婦 (−0.09)　学生，無職 (−1.27)
#1.2b　婚姻・子	0.67	未婚 (1.09)　死別 (−2.08)
		離別 (−0.36)　既婚 (−0.01)
#5.24　勤め先を変えるか	0.40	かわった方がよい (0.46)
		ながく勤めるのがよい (−0.34)
#5.6b　つとめたい会社	0.35	給料が多い会社 (0.48)
		家族的な雰囲気 (−0.24)
#9.14　外国人との結婚	0.35	賛成する (0.24)
		反対する (−0.55)
		場合による (0.17)
#2.30　不安感 重い病気	0.31	非常に感じる (−0.45)　かなり感じる (−0.14)
		少しは感じる (0.27)　まったく感じない (0.31)
#2.30e 不安感 失業	0.31	非常に感じる (−0.26)　かなり感じる (0.07)
		少しは感じる (0.41)　まったく感じない (−0.32)
#2.30f 不安感 戦争	0.31	非常に感じる (−0.47)　かなり感じる (−0.08)
		少しは感じる (0.31)　まったく感じない (0.13)

表10. 第10次M型調査票の解2.

項　目	相関	w_i
#2.30f 不安感 戦争	0.84	非常に感じる (1.25)　かなり感じる (0.27)
		少しは感じる (−0.40)　まったく感じない (−1.01)
#2.30c 不安感 街での暴力	0.79	非常に感じる (1.33)　かなり感じる (0.61)
		少しは感じる (−0.33)　まったく感じない (−0.97)
#2.30g 不安感 原子力施設の事故	0.78	非常に感じる (1.13)　かなり感じる (0.13)
		少しは感じる (−0.52)　まったく感じない (−1.12)
#2.30d 不安感 交通事故	0.71	非常に感じる (0.95)　かなり感じる (−0.03)
		少しは感じる (−0.72)　まったく感じない (−1.32)
#2.30e 不安感 失業	0.68	非常に感じる (1.05)　かなり感じる (0.34)
		少しは感じる (−0.42)　まったく感じない (−0.83)
#2.30　不安感 重い病気	0.61	非常に感じる (0.83)　かなり感じる (0.18)
		少しは感じる (−0.38)　まったく感じない (−0.94)
#2.30h 不安感 経済面の不安	0.53	非常に感じる (0.87)　かなり感じる (0.26)
		少しは感じる (−0.35)　まったく感じない (−0.78)
#1.2　年齢	0.33	20～24(−0.52)　25～29(−0.43)　30～34(−0.38)
		35～39(−0.34)　40～44(−0.28)　45～49(−0.05)
		50～54(0.27)　55～59(0.25)　60～64(0.46)
		65～69(0.36)　70以上 (0.17)

表11. 第10次M型調査票の解3.

項 目	相関	w_i
#2.3l 生活全体に満足か	0.87	満足 (−1.26) やや満足 (−0.07)
		やや不満 (1.17) 不満 (2.05)
#2.3c 家庭に満足か	0.65	満足 (−0.63) やや満足 (0.27)
		やや不満 (1.21) 不満 (2.30)
#2.3i 仕事や職場に満足か	0.61	満足 (−1.04) やや満足 (−0.21)
		やや不満 (0.56) 不満 (1.06)
#2.3j 余暇に満足か	0.60	満足 (−0.86) やや満足 (−0.06)
		やや不満 (0.67) 不満 (1.27)
#2.3k 健康状態に満足か	0.58	満足 (−0.75) やや満足 (−0.02)
		やや不満 (0.52) 不満 (1.26)
#7.29 くらしむき	0.50	非常に豊か (−1.13)
		やや豊か (−0.70)
		ふつう (−0.13)
		やや貧しい (0.88)
		非常に貧しい (1.64)
#2.3d 社会に満足か	0.43	満足 (−1.20) やや満足 (−0.48)
		やや不満 (0.10) 不満 (0.57)
#1.8 帰属階層	0.42	上 (−0.95) 中の上 (−0.55) 中の中 (−0.20)
		中の下 (0.51) 下 (0.98)
#7.30a 生活水準10年の変化	0.41	よくなった (−0.76)
		ややよくなった (−0.21)
		変らない (−0.19)
		ややわるくなった (0.37)
		わるくなった (1.13)
#2.80c 病気 いらいら	0.37	かかったことあり (0.41)
		かかったことなし (−0.35)
#2.30h 不安感 経済面の不安	0.36	非常に感じる (0.52) かなり感じる (0.18)
		少しは感じる (−0.14) まったく感じない (−0.69)

むき', '#1.8 帰属階層', '#7.30a 生活水準10年の変化' といった経済的な生活水準に関する3項目や, '#2.30h 不安感 経済面の不安' も選択されている. また, '#2.80c 病気 いらいら' も相関が0.37となっており, 不満感が高じるといらいらが増す傾向にある. すなわち, 個人生活上の満足感は個人の生活水準や健康状態と関連していることが分かる.

表12の解6では, まず '#1.1 性別' と '#6.2 男・女の生まれかわり' が相関が大きい. すなわち '男' は '男に', '女' は '女に' 生まれ変わりたいと考えていることになる. '#1.4c 職業' の相関が0.59と高いのは '管理職' には '男' が多く, '主婦' には '女' が多いといった関連が強くあるためであろう. また, '#6.2e 男の子と女の子' では同性の子どもを望み, '#6.2d 楽しみどちらが多いか' では同性の方が多いと答える傾向があるが, 相関が0.40と0.38とあまり高くなく, 強い傾向ではない. また, 男性の方が, #6.2eで 'どちらでもよい' と答える傾向にあるようである. 参考のため '#6.2f どちらがトクか' と '#6.2c 苦労どちらが多いか' の相関も示したが, 0.15や0.05と低く, こういったことが生まれ変わりたい性別を決める要因ではないことが分かる.

本論文の冒頭では, 図2によって表12で得られた項目の数量化III類の結果を示した. UNISCALを行うことで, 図2で分析対象として取り上げた項目は, これ以外にはせいぜい '#1.4c 職業' しかないという意味で, 適切なものであったということができる.

表12.　第10次M型調査票の解6.

項　目	相関	w_i
#1.1　性別	0.92	男(1.03)
		女(−0.83)
#6.2　男・女の生まれかわり	0.87	男に(0.75)
		女に(−1.01)
#1.4c 職業	0.59	農林水産業(0.14)　自営の商工業(0.42)
		専門，自由業(−0.05)　管理職(1.33)
		事務系の勤め人(−0.06)　作業系の勤め人(0.16)
		主婦(−1.10)　学生，無職(0.28)
#6.2e 男の子と女の子	0.40	男の子(0.44)
		女の子(−0.41)
		どちらでもよい(0.29)
#6.2d 楽しみどちらが多いか	0.38	男が多い(0.30)
		女が多い(−0.45)
#6.2f　どちらがトクか	0.15	男がトクだ(0.13)
		女がトクだ(−0.16)
#6.2c 苦労どちらが多いか	0.05	男が多い(0.01)
		女が多い(−0.09)

表13の解7は22回得られたものである．日本の水準に関する5項目のうち，'#9.12d日本の「生活水準」'，'#9.12c日本の「経済力」'，'#9.12e日本の「心の豊かさ」'という3項目の相関が高い．参考のため'#9.12日本の「科学技術の水準」'と'#9.12b日本の「芸術」'についてもその相関を示したが，0.25と0.16と低い上，#9.12ではカテゴリの順序通りの数値が与えられていない．

第5次調査(1973年)のM型調査票で，これら日本の水準に関する5項目ははじめて用いられた．そこで比較のため，第5次調査データのUNISCALの結果のうち，これら5項目が選択された解を表14と表15に示す．1973年の時点では，日本の「生活水準」や「経済力」と「科学技術の水準」は関連していたと見なすことができる．また，「心の豊かさ」は「生活水準」や「経済力」とは無関係であったと言える．しかし1998年には，「生活水準」や「経済力」に対する評価は，「科学技術の水準」とは無関係となり，「心の豊かさ」に対する評価と強く関連するようになっている．

表13では，他に'#2.3d社会に満足か'，'#7.40社会は公平か'といった社会に対する満足感の項目や，'#7.30a生活水準10年の変化'，'#2.30h不安感 経済面の不安'，'#1.8帰属階層'，'#7.18d生活は豊かになるか'，'#7.29くらしむき'といった個人の生活水準に関する項目が選択されている．すなわち，日本の豊かさに対する評価は，社会に対する満足感や個人の生活水準と関連していることが分かる．前田(1995)は，前回の第9次全国調査のデータを基に，個人生活への満足感と社会への満足感を，日本の豊かさや個人の生活水準，健康状態によって説明しようとしている．これは本論文の表11と表13の結果を組み合わせたモデルと考えられ，前田(1995)で分析対象として取り上げた項目は適切なものであったと言えるのである．

4.2　第10次全国調査K型調査票の分析例

次に，1998年に実施された第10次全国調査のうちK型調査票の44項目を用いた結果を示す．M型調査票と同様にα_iとしてはg_1を用い，$k=0.8$とした．M型調査票と同じく，初期値を1000回変えた結果，18個の解が得られた．

表13. 第10次M型調査票の解7.

項 目	相関	w_i
#9.12d 日本の「生活水準」	0.83	非常によい (−1.42) ややよい (−0.61)
		ややわるい (0.64) 非常にわるい (2.02)
#9.12c 日本の「経済力」	0.78	非常によい (−1.61) ややよい (−0.92)
		ややわるい (0.17) 非常にわるい (1.05)
#9.12e 日本の「心の豊かさ」	0.61	非常によい (−1.72) ややよい (−0.55)
		ややわるい (−0.01) 非常にわるい (0.80)
#2.3d 社会に満足か	0.51	満足 (−1.17) やや満足 (−0.55)
		やや不満 (0.05) 不満 (0.76)
#7.40 社会は公平か	0.45	公平だ (−0.95) だいたい公平だ (−0.64)
		あまり公平でない (0.07) 公平でない (0.61)
#7.30a 生活水準10年の変化	0.42	よくなった (−0.68)
		ややよくなった (−0.19)
		変らない (−0.20)
		ややわるくなった (0.33)
		わるくなった (1.24)
#2.30h 不安感 経済面の不安	0.41	非常に感じる (0.70) かなり感じる (0.17)
		少しは感じる (−0.25) まったく感じない (−0.61)
#1.8 帰属階層	0.41	上 (−1.00) 中の上 (−0.48) 中の中 (−0.19)
		中の下 (0.47) 下 (1.07)
#7.18d 生活は豊かになるか	0.39	豊かに (−0.65)
		貧しく (0.33)
		変わらない (−0.27)
#7.18e 幸福になるか	0.35	幸福に (−0.49)
		不幸に (0.43)
		変わらない (−0.11)
#7.29 くらしむき	0.34	非常に豊か (−0.18)
		やや豊か (−0.46)
		ふつう (−0.09)
		やや貧しい (0.47)
		非常に貧しい (1.41)
#2.3i 仕事や職場に満足か	0.32	満足 (−0.35) やや満足 (−0.20)
		やや不満 (0.32) 不満 (0.68)
#2.3l 生活全体に満足か	0.32	満足 (−0.35) やや満足 (−0.08)
		やや不満 (0.42) 不満 (1.00)
#9.12 日本の「科学技術の水準」	0.25	非常によい (−0.23) ややよい (−0.04)
		ややわるい (0.65) 非常にわるい (−0.13)
#9.12b 日本の「芸術」	0.16	非常によい (−0.32) ややよい (−0.07)
		ややわるい (0.13) 非常にわるい (0.47)

以下では，回数の多かった解1，解4から解7，解9を順に示す．解2と解3はともに，'#1.5 市郡別' と '#1.6 地方別' の相関が大きい解であった．また解8は，'#2.1 しきたりに従うか' のみの相関が大きい解であった．

表17の解1では，相関の高い項目としては，'#1.2 年齢'，'#1.3 学歴'，'#1.4c 職業'，'#1.2b 婚姻・子' があり，これらの項目はM型調査票においては表9で選択されている．その他，年齢と関連が強い項目としては '#3.9 首相の伊勢参り'，'#4.11 先祖を尊ぶか'，'#4.5 子供に「金は大切」と教える'，'#8.7h 支持政党'，'#3.1 宗教を信じるか'，'#8.6 選挙への関心'，'#4.10 他人の

表 14. 第 5 次 M 型調査票の解 1.

項 目	相関	w_i
#9.12 日本の「科学技術の水準」	0.97	非常によい (1.24)
		ややよい (−0.59)
		ややわるい (−1.38)
		非常にわるい (−2.66)
#9.12c 日本の「経済力」	0.54	非常によい (0.79)
		ややよい (−0.13)
		ややわるい (−0.49)
		非常にわるい (−0.98)
#9.12b 日本の「芸術」	0.36	非常によい (0.74)
		ややよい (−0.09)
		ややわるい (−0.26)
		非常にわるい (−0.65)
#9.12d 日本の「生活水準」	0.34	非常によい (0.82)
		ややよい (0.08)
		ややわるい (−0.18)
		非常にわるい (−0.53)
#9.12e 日本の「心の豊かさ」	0.18	非常によい (0.64)
		ややよい (0.01)
		ややわるい (−0.05)
		非常にわるい (−0.20)

表 15. 第 5 次 M 型調査票の解 2.

項 目	相関	w_i
#9.12c 日本の「経済力」	0.95	非常によい (−1.25)
		ややよい (0.03)
		ややわるい (1.03)
		非常にわるい (2.33)
#9.12d 日本の「生活水準」	0.57	非常によい (−1.17)
		ややよい (−0.22)
		ややわるい (0.27)
		非常にわるい (1.15)
#9.12 日本の「科学技術の水準」	0.50	非常によい (−0.60)
		ややよい (0.26)
		ややわるい (0.78)
		非常にわるい (1.44)
#7.18d 生活は豊かになるか	0.33	豊かに (−0.36)
		貧しく (0.35)
		変わらない (0.18)
#9.12b 日本の「芸術」	0.30	非常によい (−0.56)
		ややよい (0.00)
		ややわるい (0.36)
		非常にわるい (0.55)
#9.12e 日本の「心の豊かさ」	0.25	非常によい (−0.73)
		ややよい (−0.14)
		ややわるい (0.09)
		非常にわるい (0.38)

表16. 第10次K型調査票のUNISCALの結果.

	解										
	1	2	3	4	5	6	7	8	9	10	11～18
回数	833	65	37	20	18	5	5	4	3	2	1
r	.34	.56	.53	.55	.62	.62	.53	.46	.61	.58	.53～.75

表17. 第10次K型調査票の解1.

項 目	相関	w_i
#1.2 年齢	0.92	20～24(−1.15) 25～29(−1.03) 30～34(−0.84) 35～39(−0.78) 40～44(−0.64) 45～49(−0.42) 50～54(−0.19) 55～59(0.11) 60～64(0.67) 65～69(0.95) 70以上(1.83)
#1.3 学歴	0.74	小学校(2.21) 新制中学(0.76) 新制高校(−0.25) 大学(−0.67)
#1.4c 職業	0.65	農林水産業(1.15) 自営の商工業(0.06) 専門，自由業(−0.73) 管理職(−0.20) 事務系の勤め人(−0.84) 作業系の勤め人(−0.34) 主婦(0.13) 学生，無職(1.09)
#1.2b 婚姻・子	0.58	未婚(−1.04) 死別(1.57) 離別(0.23) 既婚(0.03)
#3.9 首相の伊勢参り	0.52	行かねばならぬ(1.26) 行った方がよい(1.02) 本人の自由(−0.25) 行かない方がよい(−0.09) 行くべきではない(−0.43)
#4.11 先祖を尊ぶか	0.46	尊ぶ方(0.38) 普通(−0.46) 尊ばない方(−0.73)
#4.5 子供に「金は大切」と教える	0.43	賛成(0.59) 反対(−0.38) いちがいにはいえない(0.02)
#8.7h 支持政党	0.39	自民党(0.68) 民主党(−0.10) 新党平和・公明(0.23) 自由党(0.06) 共産党(−0.01) 社民党(0.33) 改革クラブ(−0.25) 新党さきがけ(−0.67) その他の政党(−1.13) 支持政党なし(−0.27)
#3.1 宗教を信じるか	0.32	信じている(0.51) 信じていない(−0.20)
#8.6 選挙への関心	0.32	なにをおいても投票(0.36) なるべく投票(−0.10) あまり投票する気にならない(−0.53) ほとんど投票しない(−0.60)
#4.10 他人の子供を養子にするか	0.32	つがせる(0.58) つがせない(−0.17) 場合による(−0.16)
#7.4b 国の繁栄と国民の生活	0.30	よくならない(−0.40) よくなる(0.21)

子供を養子にするか', '# 7.4b 国の繁栄と国民の生活' が選択されており, いずれも頷ける結果である. K 型調査票の 44 項目のうち, 表 17 で相関が 0.30 以上の項目は 12 項目ある上, この解は 1000 回の反復計算のうち 833 回得られたものであり, 過去からの継続質問項目の多い K 型調査票では, 項目のグループが他にはあまりないことになる.

表 18 の解 4 では, '# 1.1 性別' と関連がある項目として, '# 6.2 男・女の生まれかわり' と '# 1.4c 職業' が選択されているが, これは M 型調査票の結果における表 12 に対応する. 表の 3 項目以外には相関の高い項目は見当たらず, 生まれかわりたい性別を決める要因は不明である. なお K 型調査票には, '# 6.2e 男の子と女の子' や '# 6.2d 楽しみどちらが多いか' といった項目は含まれていない.

表 19 の解 5 では, '# 7.24 就職の第 1 の条件' と '# 7.24b 就職の第 2 の条件' が関連していることが分かる. すなわち, 就職の条件として '気の合った人たち' と 'やりとげたという感じ' は同時に選ばれやすいことになる.

表 20 の解 6 で挙げられた 2 項目は, '賛成' あるいは '反対' に対して正の数量が与えられ, 'い

表 18. 第 10 次 K 型調査票の解 4.

項　目	相関	w_i
#1.1 性別	0.92	男 (1.00)
		女 (−0.85)
#6.2 男・女の生まれかわり	0.88	男に (0.71)
		女に (−1.08)
#1.4c 職業	0.59	農林水産業 (0.06)　自営の商工業 (0.39)
		専門, 自由業 (0.02)　管理職 (1.24)
		事務系の勤め人 (0.01)　作業系の勤め人 (0.22)
		主婦 (−1.12)　学生, 無職 (0.26)

表 19. 第 10 次 K 型調査票の解 5.

項　目	相関	w_i
#7.24 就職の第 1 の条件	0.91	よい給料 (0.39)
		失業の恐れがない (−0.10)
		気の合った人たち (1.19)
		やりとげたという感じ (−0.92)
#7.24b 就職の第 2 の条件	0.90	よい給料 (−0.04)
		失業の恐れがない (0.19)
		気の合った人たち (−1.04)
		やりとげたという感じ (1.30)

表 20. 第 10 次 K 型調査票の解 6.

項　目	相関	w_i
#7.1 人間らしさはへるか	0.99	賛成 (0.63)
		いちがいにいえない (−1.68)
		反対 (0.51)
#7.2 心の豊かさはへらないか	0.33	反対 (0.28)
		いちがいにいえない (−0.58)
		賛成 (0.10)

表21. 第10次K型調査票の解7.

項 目	相関	w_i
#5.1c2 入社試験（恩人の子）	0.88	1番の人 (−0.81)
		恩人の子 (0.97)
#5.1c1 入社試験（親戚）	0.87	1番の人 (−0.50)
		親戚の人 (1.56)
#5.1 恩人がキトクのとき	0.04	故郷へ帰る (−0.03)
		会議に出る (0.04)
#5.1b 親がキトクのとき	0.01	故郷へ帰る (0.00)
		会議に出る (0.02)

表22. 第10次K型調査票の解9.

項 目	相関	w_i
#5.1 恩人がキトクのとき	0.92	故郷へ帰る (0.91)
		会議に出る (−0.94)
#5.1b 親がキトクのとき	0.91	故郷へ帰る (0.95)
		会議に出る (−0.89)
#5.1c1 入社試験（親戚）	0.02	1番の人 (0.00)
		恩人の子 (−0.05)
#5.1c2 入社試験（恩人の子）	0.02	1番の人 (0.00)
		恩人の子 (−0.04)

ちがいにいえない'に対しては負の数量が与えられている．すなわち，人間らしさに関する問いに対しはっきりとした回答をするか否かということを調べる項目としては，#7.1と#7.2は似ているということになる．

表21の解7及び表22の解9はどちらも義理人情に関する質問項目である．参考のため，相関は低いもののお互いの項目を示している．これらの結果から，対象が親戚か恩人かということよりも，状況が入社試験かキトクのときかということによって回答が決まることが分かる．

おわりに

本論文では，調査項目の最適尺度変換を行いつつ，多数の調査項目の中から一次元性のある項目だけを選び出す方法であるUNISCALを提案し，「日本人の国民性」第10次全国調査データへの適用例を示した．国民性調査では，広い領域にわたって質問がなされているため，お互いに無関連な項目が多い．関連のある項目を拾い出すUNISCALでは，必ずしも目新しい知見が見出されたとは言えないが，得られた結果は全て納得のいくものであったと言える．

「日本人の国民性調査」は継続調査であるため，各次の調査データにUNISCALを適用し，その結果を比較すれば，より興味深い結果が得られるかもしれない．しかしそのためには，大量の結果を見比べなければならず，効率的とは言えない．継続調査データに対応したUNISCALの開発が望まれる．

本論文では質的データを念頭に置いてUNISCALを説明したが，当然ながらUNISCALは量的データへの対応も可能である．ただし間隔尺度をそのまま扱うと，変数間の非線形な関係は見つけ出すことができない．その場合は，データをカテゴリ化する等の工夫が必要である．

UNISCALではkの値を変えることで，変数の数が多いグループから少ないグループまで

様々な変数のグルーピングができる．そのためkの値を段階的に変化させることで，変数の階層的なクラスタリング手法として発展する可能性がある．

最後に，UNISCALの適用に当たっての補足を述べる．

まず，先に述べたように，分析の対象とする項目の選定には気を遣う必要はなく，むしろ全ての調査項目を投入した方がよい．調査データはあるものの検証すべき明確な仮説等がない，といった場合にUNISCALを適用することで，その後どのような観点から分析を進めればよいか，といったヒントが得られることが多い．また，UNISCALはそのような使い方こそが最も適しているのである．ただし，複数選択の項目には対応していないため，今後この点については改良が必要である．

α_iとkの値は分析者が定めなければならないが，その選択方法に明確な指針はない．しかし，それはUNISCALの欠点とは言えない．なぜなら先に述べたとおり，α_iやkの値をいくつか変えながらその結果を見比べることで，データ構造に対する理解がより深まるからである．UNISCALは探索的なデータ解析の方法であり，その特徴を活かすためにはα_iやkの値は一意に定めない方がよい，ということである．

UNISCALでは反復計算を行うことは必須である．データの中に含まれる一次元構造の種類が多いほど，反復回数は大きくする必要がある．すなわち，変数の数が多い場合やkの値を大きくした場合である．通常は，まず数十回の反復計算を行って得られる解の数を見極めた後，多くの解が得られそうであれば，反復計算の数を大きくすればよい．

また，反復計算を行うことで，普通複数の解が得られる．経験的には，得られる回数の少ない解は不安定であったり，カテゴリの最適変換値が不自然で意味を解釈できないことが多い．また，1つの項目だけに相関が高いこともしばしばである．そのような解は無理に採用する必要はなく，意味があると思われる解だけを採用したり，α_iやkの値を変えて再分析することが有効である．

数量化III類では，反応頻度の少ないカテゴリがあると，解はその影響を大きく受けることが知られている．UNISCALでも，反応頻度が極端に偏っているときには，その影響を受ける場合がある．例えば，'D.K.'等のように，頻度が少なく，ある項目で'D.K.'を選んだ個体が他の項目でもそれを選ぶ傾向がある場合には，'D.K.'対その他のカテゴリという解のみが多数回得られてしまう．そのときは，カテゴリを併合する，個体を削除する，欠損値として扱う，等の処理が必要である．一般的には，頻度の少ないカテゴリを選んだ個体が，他の項目の特定のカテゴリのみを選ぶという傾向がなければ，反応頻度が多少偏っているとしても，反復計算の結果に与える影響は大きくないと思われるが，この点に関しては今後の検討が必要である．

謝　　辞

本研究所の坂元慶行教授，中村隆教授，前田忠彦助手には，本論文の草稿を読み有益なコメントをいただきました．またレフェリーの方には丁寧に査読をしていただきました．心より感謝いたします．

付　　記

(3.1)式のrが，ダミー変数行列において1の代わりにα_iを用いて数量化III類を行ったとき，最大化された相関係数の2乗に一致することは，以下のように明らか．

$\boldsymbol{w}_i=(\boldsymbol{D}_i'\boldsymbol{D}_i)^{-1}\boldsymbol{D}_i'\boldsymbol{m}$，$\sum_{i=1}^{I}\alpha_i\|\boldsymbol{\Delta}_i\boldsymbol{m}\|^2=\sum_{i=1}^{I}\alpha_i\|\boldsymbol{\Delta}_i\boldsymbol{1}\|^2$であることから，

$$
\begin{aligned}
r &= 1 - \frac{1}{\sum_{i=1}^{I} \alpha_i \mathbf{1}' \boldsymbol{\Delta}_i \mathbf{1}} \sum_{i=1}^{I} \alpha_i \| \boldsymbol{\Delta}_i (\boldsymbol{D}_i \boldsymbol{w}_i - \boldsymbol{m}) \|^2 \\
&= 1 - \frac{1}{\sum_{i=1}^{I} \alpha_i \boldsymbol{m}' \boldsymbol{\Delta}_i \boldsymbol{m}} \sum_{i=1}^{I} \alpha_i \boldsymbol{m}' (\boldsymbol{\Delta}_i - \boldsymbol{D}_i (\boldsymbol{D}_i' \boldsymbol{D}_i)^{-1} \boldsymbol{D}_i') \boldsymbol{m} \\
&= \frac{\boldsymbol{m}' (\sum_{i=1}^{I} \alpha_i \boldsymbol{D}_i (\alpha_i \boldsymbol{D}_i' \boldsymbol{D}_i)^{-1} \alpha_i \boldsymbol{D}_i') \boldsymbol{m}}{\boldsymbol{m}' (\sum_{i=1}^{I} \alpha_i \boldsymbol{\Delta}_i) \boldsymbol{m}}
\end{aligned}
$$

となる.

参 考 文 献

Gifi, A. (1990). *Nonlinear Multivariate Analysis*, Wiley, Chichester.

Liebetrau, A.M. (1983). *Measures of Association*, Sage, Beverly Hills.

前田忠彦 (1995). 日本人の満足感の構造とその規定因に関する因果モデル ── 共分散構造分析の「日本人の国民性調査」への適用 ──, 統計数理, **43**(1), 141-160.

Meulman, J. (1982). *Homogeneity Analysis of Incomplete Data*, DSWO Press, Leiden.

西里静彦 (1982). 『質的データの数量化』, 朝倉書店, 東京.

坂元慶行 (1981). カテゴリカルデータにおける変数選択 ── プログラム CATDAP を中心に ──, 統数研彙報, **28**(1), 135-155.

坂元慶行 (1985). 『カテゴリカルデータのモデル分析』, 共立出版, 東京.

土屋隆裕 (1995). 項目分類のための数量化法, 行動計量学, **22**(2), 95-109.

土屋隆裕 (1996). 質的項目の選択による一次元尺度の構成法, 応用心理学研究, **21**, 21-30.

An Application of UNISCAL to the Survey of the Japanese National Character

Takahiro Tsuchiya

(The Institute of Statistical Mathematics)

This paper proposes UNISCAL (UNI-dimensional SCALing), which is a method to perform simultaneously both optimal scaling of items and selection of uni-dimensional items. The basic idea of UNISCAL was proposed in Tsuchiya (1996). The features of improved UNISCAL in this paper are (1) to provide four variants of α_i, which serves item selection, in order to treat various kinds of data sets, (2) to control the degree of uni-dimensionality of selected items by changing parameter k, and (3) to treat data sets with missing values.

Some artificial data sets are analyzed with both Hayashi's quantification method III and UNISCAL to illustrate the characteristics of UNISCAL in detail. An application of UNISCAL to the tenth nation-wide survey of the Japanese national character is also presented to demonstrate the performance of UNISCAL.

Key words: UNISCAL, survey of the Japanese national character, uni-dimensional scale, Hayashi's quantification method III, homogeneity analysis.