様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科研費 KAKENHI

科学研究費助成事業　　研究成果報告書

平成 ３０ 年　 ６ 月 １９ 日現在

機関番号：３３９０８
研究種目：基盤研究(C)（一般）
研究期間：2015～2017
課題番号：１５Ｋ０４１９７

研究課題名（和文）完全単純構造主成分分析とその個人差測定尺度構成への応用に関する計量心理学的研究

研究課題名（英文）Psychometric study on perfect simple structure principal component analysis and its applications to scale construction for measuring individual diffrences

研究代表者
村上　隆（MURAKAMI，Takashi）

中京大学・現代社会学部・教授

研究者番号：７００９３０７８

交付決定額（研究期間全体）：（直接経費）　　2,300,000円

研究成果の概要（和文）：多重対応分析は，しばしば質的データの主成分分析と呼ばれる。本研究は，この陳述を文字通りの意味で成立させることを目指した。より具体的には，直交多項式主成分分析と名づけられた方法を開発した。この方法においては，中心化，精気直交化されたカテゴリー数ー1次元のベクトルによって数量化が行われる。驚いたことに，このベクトルの数値は全く任意であり，かつ同一の個体スコアと固有値を生み出すという意味で，多重対応分析と等価であることが示される。数量化ベクトルの不定性を利用して，解を一意に定める２つの方法が導入されるとともに，一般的な調査データに対する多次元の分析が可能となることが示された。

研究成果の概要（英文）：Multiple correspondence analysis is often called principal component analysis for categorical data. This project was aiming at making the statement hold literally. More concretely, a method named orthogonal polynomial principal component analysis was developed, where each category was quantified using number of categories minus one centered and orthonormal weight vectors. Astonishingly enough, it was shown that the weight vectors are entirely arbitrary, Owing yo the arbitrariness, two ways of procedures was employed; one is the coding of categories by orthogonal polynomials, the other is the quartimax rotations of the weight matrix from both sides. The method was applied to various psyhometirc and survey data, and several new findings have been obtained.

研究分野：計量心理学

キーワード：主成分分析　直交回転　斜交回転　多重対応分析　クラスター分析　単純構造　心理測定尺度　調査データ

１．研究開始当初の背景

複数の質問項目に対する反応を集計して個人の特性を測定するという心理測定尺度を構成するための統計的な方法論を扱う学問領域が計量心理学（psychometrics）である。その中でも，探索的因子分析（exploratory factor analysis）と古典的テスト理論（classical test theory）とは，そうした分析の最も基本のところ，すなわち研究の最初のステップにおいて，多くの貢献をなしてきた（Nunnally, 1978）。

今や，潜在変量（latent variate）を基本にすえた進んだ理論が大いに開発されているものの，項目反応に重みをつけて加算するという形で定義される線形合成変量（linear composite）にもとづく方法が，研究のより初期の段階でもつ重要性は失われることがないと考える（Levy & Mislevy, 2016）。

研究の出発点において問題となるのは，個人差の次元として（ある程度のあいまいさをもって）構想されている個人差の次元（心理学的特性 psychological trait）を測定するための適切な質問項目群はどのような statements からなるべきかという内容的妥当性の判断であり，それらの項目が統計的には幾つの次元からなり，かつそれらの次元をどの程度の信頼性をもって査定（assess）できるかの見当をつけることであろう。そのためには，項目を複数の下位概念に沿って統計的に分類するための方法を改めて確立する必要がある（村上，2014）。

それとともに，形式的には心理測定尺度と同様，個人の質問への応答の形をとる社会調査データへの応用も考慮されるべきである。社会調査の目標は，個人差の測定（個人の査定）というよりは，個人の目に映じている社会の描像を明らかにすることであろうが，これも多数の次元にそって個人（の見方）を位置づけるという考え方に立てば，類似の方法論が有用であるとも考えられる（Greenacre, 2017）。

以上のような認識の下に，計量心理学と社会統計学の伝統に依拠しつつ，多数の個人から得られた反応を分析するための方法論の開発は重要な課題である。

２．研究の目的

こうした状況認識から，本研究の目的は，多数の質問項目からなる尺度を探索的に，かつ筋の通った形で，かつ，算出されるパラメータについて，推測統計学的な標準誤差の大きさを推定できるような方法を確立することであった。

その際，主として念頭に置くデータの形態は，回答者の意識や行動に関する statements に対して 3～7 段階で応答することを求めるという Likert 型項目であった。そこでは，古くから反応カテゴリーが等間隔であるかどうかが問われてきたこと，さらに，回答者の反応の構えとして知られている極端反応傾向を，特に査定においてどのように考慮に入れるかという問題への一定の解答を与えることが目指された。

それとともに，社会調査においてよく用いられる check list 型の項目（選択肢を単位に考えるとそうした項目への応答は 1 か 0 かの 2 値的なものとなる），全く順序のないカテゴリー（これには，回答者の職業や居住地域等のいわゆる demographic な属性に関する質問が含まれる）を扱うことのできるような方法であることも望まれる。

こうした特長を備えた方法を，従来から計量心理学と社会統計学の分野において使い込まれた方法論を基礎として開発することが目指された。

３．研究の方法

以上のような目的を達成するためには，２つの方向での研究活動を必要とする。

第 1 は，目的に見合った新しい分析方法の数理を明らかにし，１つの分析のフローを確立する理論的な研究である。第 2 は，そのように構想され定式化された数理を分析プログラムとして作成し，実際の（心理学的，社会学的）調査によって得られたデータに適用して結果を解釈し，開発された分析方法の実用可能性を確認することである。

第 1 の目的に関しては，カテゴリカルな反応をあくまでも質的なものと見なして分析する多重対応分析（multiple correspondence analysis; MCA）の検討から開始する。この方法が，Likert 型の項目反応に適用される場合には，カテゴリーの等間隔性を仮定しない（たとえば 5 段階の反応をそのまま数値と見なして加算することはしない）という点で，２で述べた問題を解決する基礎として適切である。

その一方で，この方法は解の提示に主として平面上の map という視覚的な方法を用いているために，3 次元以上の構造を扱うことが難しい点に問題がある。

他方，解の単純構造（simple structure）への回転を伴う主成分分析（principal component analysis; PCA）は，解の表現を空間内の軸を基準に行っているために，多くの次元を扱うのに適している。ただし，この方法は，カテゴリーの間隔を等しいと仮定し，回答者の反応傾向にも配慮が払われていない。

したがって，方法開発の目標は，多重対応分析を，単純構造をもつ軸単位の解釈が可能な方法へと作り変えること，すなわち，多重対応分析の解を，主成分分析で解釈に用いられる単純構造をもつパターン行列で表現できるように改訂（revise）する，ということになる。

従来，MCA も PCA も記述的な方法とされ，（少なくとも心理・社会学への応用に際しては，）多変量正規分布等を基盤にした推測統計学的な扱いをうけることがなかった。確率

分布を用いると，そこから分析方法に一定の制約が課されることになり，適用されるデータがその前提条件を満たすかどうかが問題になる等，心理・社会学的データの分析には必ずしも有用でない面が出てくる。しかしながら近年，得られたデータを母集団と見なし，そこから復元抽出によって同じ大きさの標本を反復抽出して分析を繰り返し，推定されるパラメータの変動にもとづいて，標本誤差を数値的に推定する手法である，Bootstrap 法が開発されたことにより，記述的とされていた方法の推測統計学的な扱いが可能になってきた。

ただし，ここで対象としている MCA と PCA には, パラメータの符号に関する不定性があり，Bootstrap 法の単純な適用によっては，標準誤差を正しく求めることができない。そのための工夫も必要になる。

第 2 のプログラム作成と実データによる方法の検証であるが，プログラムについては，研究代表者の主要言語である MATLAB による従来の蓄積を用い，さまざまな実験的試みも含めて，多様な検討を行う準備は整っていた。プログラムの作成と実データへの適用，出力結果の整理と視覚化等の作業は，すべて研究代表者がひとりで行った（その種の作業のためのアルバイター雇用は行わなかった)。

検証のためのデータとしては，すでに 20 種に近いパーソナリティ尺度のデータが手元にあり，それを状況に応じて使用することができた。また社会調査に関しては，研究代表者が中心になって進めてきた野球場での観客の意識調査のデータが 10 年分以上あり，これら自分が関わって収集と分析を進めた経験のあるデータが，結果を詳細に検討するために非常に有用であった。

４．研究成果

(1) 分析手法の開発

一般的なカテゴリカルデータ　分析の対象は，n 人の回答者の p 個の多肢選択形式の質問項目への反応（整数）からなるデータ行列とする。項目 k の反応カテゴリーの数 $c_k\ (\geq 2, k=1,\ldots p)$ は項目ごとに異なってもよい。また，$c = c_1 + c_2 + \cdots + c_p$ とする。全変数を 2 値にコード化した $n \times c$ のダミー変数行列を $\mathbf{G} = [\mathbf{G}_1 \quad \mathbf{G}_2 \quad \ldots \quad \mathbf{G}_p]$ とする。なお，$\mathbf{D}_k = \mathrm{dg}\,\mathbf{G}'_k\mathbf{G}_k$，この行列と対角要素にもつ対角行列を $\mathbf{D}$ とする。また，$\mathbf{d}_k = \mathbf{D}_k\mathbf{1}_{c_k} = \mathbf{G1}_n$ と定義する。$\mathbf{D}$ と $\mathbf{d}_k$ の非ゼロの要素は，各カテゴリーへの反応頻度である。

多重対応分析（再定式化）　$\mathbf{V}$ を $\mathbf{V}'\mathbf{DV} = \mathbf{I}$ を満たす $c \times r$ の行列（$r < c$）とする。個体スコアを $\mathbf{F} = \mathbf{GV}$，ただし $\mathbf{F}'\mathbf{1}_n = \mathbf{0}$ と定義する。最大化基準は個体スコアの分散の和 $\mathrm{tr}\, n^{-1}\mathbf{F}'\mathbf{F}$ である。この解は，$\mathbf{D}^{-1/2}\mathbf{G}'\mathbf{GD}^{-1/2} - \mathbf{D}^{1/2}\mathbf{1}_c\mathbf{1}'_c\mathbf{D}^{1/2} = \mathbf{K\Lambda K}'$ という固有値分解において，$\mathbf{\Lambda}$ の大小順に r 番目までの固有値に対応する正規直交な固有ベクトルの行列を $\mathbf{K}^{(r)}$ とするとき，$\mathbf{V} = \mathbf{K}^{(r)}\mathbf{T}$ によって与えられる。なお，r は（原則的に）分析に先立って定められる解の次元である。

通常の定式化と異なる点は，個体スコアが $\sqrt{p}$ 倍であることと，最大化基準を 2 乗和で定義した結果，重み行列 $\mathbf{V}$ に（結果的に個体スコア $\mathbf{F}$ に）直交回転に関する不定性が生じたことである。

一般化正規直交多項式主成分分析（村上, 2018）　$\mathbf{U}_k$ を, $\mathbf{U}'_k\mathbf{d}_k = \mathbf{0}$, $n^{-1}\mathbf{U}'_k\mathbf{D}_k\mathbf{U}_k = \mathbf{I}$ を満たす任意の $c_k \times (c_k - 1)$ の行列とする。$\mathbf{Z}_k = \mathbf{G}_k\mathbf{U}_k$ と定義すると，この各列は平均値が 0，分散が 1 の標準得点となり，これから算出される相関行列 $\mathbf{R}_{kl} = n^{-1}\mathbf{Z}'_k\mathbf{Z}_l$ を要素行列とする c - p 次のブロック正方行列を $\mathbf{R}$ とする。その上で，$\mathbf{W}'\mathbf{W} = \mathbf{I}$ を満たすような $(c-p) \times r$ の行列 $\mathbf{W}$ により主成分得点 $\ddot{\mathbf{F}} = \mathbf{ZW}$ を定義する。この分散の和，$\mathrm{tr}\, n^{-1}\ddot{\mathbf{F}}'\ddot{\mathbf{F}} = \mathrm{tr}\,\mathbf{W}'\mathbf{RW}$ を最大化する $\mathbf{W}$ は，固有値分解 $\mathbf{W}'\mathbf{RW} = \mathbf{L\Delta L}'$ において，大小順に r 番目までの固有値に対応する固有ベクトルの列からなる行列 $\mathbf{L}^{(r)}$ によって $\mathbf{W}$=$\mathbf{L}^{(r)}\mathbf{T}$ と与えられる。ここで，（負荷行列でなく）重み行列が回転の対象となる点が，通常の PCA の定式化とは異なる。$\mathbf{R}$ の対角ブロックは $\mathbf{U}_k$ の制約条件からすべて単位行列となるので，この方法は正準相関分析の一般化（の一種）と見ることもできる。

このとき, $\mathbf{\Lambda}$ と $\mathbf{\Delta}$ の非ゼロの要素は一致し，回転を除いて $\mathbf{F} = \ddot{\mathbf{F}}$ であることも示される（村上，2016a)。すなわち，この方法は実質的に多重対応分析と同等である。また，主成分得点は，ダミー変数行列から算出するなら（回転を除いて)，$\mathbf{F} = \sum_k \mathbf{G}_k\mathbf{U}_k\mathbf{W}_k$ となるが，ここには $\mathbf{U}_k\tilde{\mathbf{T}}_k\tilde{\mathbf{T}}'_k\mathbf{W}_k\mathbf{T}$ という意味で 2 種の直交回転についての不定性がある。$\mathbf{U}_k$ の要素は，上記の制約条件を満たす限り任意であるが，2 種の直交回転によりこれを一意に定められるともいえる。回転に Orthomax 基準を用いるとすれば，$\mathbf{W}$ は正規直交行列なので，Quartimax 基準に帰着する。$\mathbf{W}$ の要素の 4 乗和を最大化基準とし，$\mathbf{T}$ と $\tilde{\mathbf{T}}_k$ を交互に更新する。

なお，解釈にあたっては，パターン行列 $\mathbf{A} = \mathbf{WT}(\mathrm{dg}\,\mathbf{T}'\mathbf{\Lambda T})^{1/2}$ を用いる。この要素は，各変量 $\mathbf{G}_k\mathbf{U}_k$ の主成分得点への標準偏回帰係数として解釈できるのみならず，式の形から明らかなように，回転後の重み行列と列単位に比例すること，各列の 2 乗和が主成分得点の分散と一致し，いわゆる主成分寄与としても使用できるといった有用な性質をもっている。主成分間相関行列は，共分散行列 $n^{-1}(\mathbf{FT})'\mathbf{FT}$ から得られる。

双方向からのプロクラステス回転　広義の因子分析において負荷量行列の要素の標準誤差を Bootstrap 法によって推定しようとすると，負荷量の符号の任意性が問題になる。この問題の解決には，Bootstrap 標本から計算された負荷行列を基準になる（全サンプルによる）解に負荷量の符号を合わせるか，プ

ロクラステス法により基準負荷行列に近似させる方法が知られている（Timmerman et al., 2007）。上記の方法では，$\mathbf{U}_k$ と $\mathbf{W}$ の 2 つの行列の符号がからみ，符号を一致させる手続きは相当に煩雑なものとなる。したがって，プロクラステス法によることを考える。全標本によるターゲットとなる重み行列を $\vec{\mathbf{W}}_k$ として，$\sum_k \| \vec{\mathbf{W}}_k - \tilde{\mathbf{T}}_k \mathbf{W}_k \mathbf{T} \|^2$ を最小化する。通常の特異値分解によるアルゴリズムを用いた交互最小 2 乗法は，そのままではかなりの頻度で local minima に収束することがわかった。この問題は，現在までの経験では，順序のあるカテゴリカル変数のみによる予備回転を行うことで 100%解消できた。

直交多項式のあてはめ　通常の探索的因子分析のやり方でパターン行列を解釈しようとすると困難に遭遇する。A の列にあたる各変量は，通常，質問項目あるいは尺度であり，その意味を解釈するのに困難はない。しかし，この方法では，量は $\mathbf{U}_k$ によって定義されるから，解釈にあたっては，その要素にもどって検討しなければならない。全変数についてこれを行うのはきわめて煩雑である。

そのための対応として，順序のあるカテゴリカル変数については，$\mathbf{U}_k$ の定義は上記の方法によらず，直交多項式によってコーディングするものとした。整数 $1^l, 2^i \ldots, c_k^l, (l = 1, \ldots, c_k - 1)$ をカテゴリー頻度を考慮して中心化した上で，同じく反応頻度を考慮した Gram-Schmidt の直交化法により正規直交化する（村上，2016）。

(2) 実データへの適用

この方法はまず，定型的な Likert 尺度を対象にして行い，ほぼ所期の成果を得た（村上，2016）。この際，すべての質問項目について整数値による正規直交多項式の当てはめを行ったが，内容的な２つの次元（ポジティブ感情とネガティブ感情）について，すべて高次の成分が伴っていた。特に，2 次成分については，個人の極端反応の程度を表すという解釈ができ，これは個人の assessment にあたっては，なんらかの形で調整すべき成分であることが確認できた。

従来，主要な 2 つの次元が作る放物線状の布置については馬蹄現象（horse-shoe phenomenon）と呼ばれ，データの作る高次元多面体の頂点を低次元のユークリッド空間に射影することから生じるアーチファクトと考えられてきたが，本研究では，実は馬蹄現象にも回答者の反応の構えという経験的事実が対応することが明らかになったわけであり，これは本研究の大きな成果と言える。

社会調査データについては，分析の緒についたばかりであるが，野球場での観客の多次元的表現について，従来得られなかった次元がいくつか明らかになった。１つは，ひいきのチームを応援することが目的のサポーターと，球場での観戦回数が多いファンとが別次元であることがわかったこと，もう１つは，会社や友人からもらったチケットで，球場内の特等席で観戦しながら，飲食や一緒に来場した同僚らとのおしゃべりが楽しみという一群の人々の存在である（これは球団の営業方針に起因するようだが）。

また，Bootstrap 法による推測統計学的扱いに関しては，ターゲットとの内積の集計によって，発見された次元の安定性といったものが評価できた。

結果を下の表に示す。

主成分	下 2.5%	下 5%	中央値	平均値	上 2.5%	最大値
Ⅰ（サポーター）	0.945	0.952	0.971	0.970	0.984	0.989
Ⅱ	0.928	0.935	0.958	0.956	0.974	0.982
Ⅲ（ファン）	0.858	0.883	0.941	0.935	0.968	0.979
Ⅳ（ヤング）	0.826	0.847	0.928	0.920	0.963	0.977
Ⅴ（レディース）	0.801	0.829	0.914	0.905	0.953	0.968
ⅤⅠ（ホーム）	0.833	0.863	0.928	0.922	0.963	0.978
ⅤⅡ	0.789	0.820	0.915	0.905	0.956	0.970
ⅤⅢ（フリー）	0.704	0.759	0.903	0.885	0.954	0.968
Ⅸ	0.479	0.543	0.816	0.786	0.927	0.954

パターン行列の要素の標準誤差についても，適切と思われる推定値が得られている。

なお，上の表の第 II, VII, IX 主成分については，Likert 型項目の 2 次以上の成分であり，個人の査定を目的としない社会調査では，解釈の対象とはしなかった。従来，MCA においてこうしたアーチファクトに順ずる（経験的事実ではあっても，内容的な解釈はできない）次元を軸ごと分離する手法はなかった。

＜引用文献＞

Greenacre, M. (2017). *Correspondence analysis in practice. 3rd ed.* Boca Raton NW: CRC Press

Levy, R. & Mislevy, R.J. (2016). *Bayesian psychometric modeling*. Boca Raton, NW: CRC Press.

村上 隆 (2014). 完全単純構造・主クラスター成分分析・resampling による確認：心理測定尺度構成のための単純な分析方法試論　中京大学現代社会学部紀要, 8 (1), 47-90.

Nunnaly, J.C. (1978). *Psychometric theory. 2nd ed.* New York: McGraw Hill.

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 1 件）

①<u>村上 隆</u> (2016). 多重対応分析の変換としての正規直交多項式主成分分析　－Likert 型項目の探索的分析のための新たな手続き－ データ分析の理論と応用, 5, No.1, 27-47.

〔学会発表〕（計 10 件）

①<u>村上 隆</u>（2018）. 多重対応分析の「因子分析」としての使用　日本分類学会第 37 回大会

② Murakami, T. (2017). Generalized Orthonormal Polynomials Principal Component Analysis. International forum of classification society 2017.
③**村上 隆（2017）．回転を含む主成分分析におけるブーツストラップ法の使用について 日本行動計量学会第 45 回大会**
④ Murakami, T. (2017). Orthonormal polynomials principal component analysis. International meeting of psychometric society 2017.
⑤**村上 隆（2016）．Likert 尺度再考 正規直交多項式主成分分析による検討 日本行動計量学会第 44 回大会**
⑥Murakami, T. (2016). Two causes of the horseshoe phenomenon in multiple correspondence analysis. International meeting of psychometric society 2016.
⑦**村上 隆（2015）．Likert 尺度と数量化理論：多重対応分析を主成分分析として使う 日本テスト学会学会賞受賞記念講演**
⑧**村上 隆（2015）．順序のあるカテゴリカルデータの多重対応分析と主成分分析 2015 年度統計学関連学会連合大会**
⑨ Murakami, T. (2015). Applications of principal cluster components analysis to scale development and the validation of clustering by a resampling method. International meeting of psychometric society 2015.
⑩**村上 隆（2015）．多重対応分析における馬蹄現象はなぜ発生するのか　３値データについての検討 日本行動計量学会第 43 回大会**

〔図書〕（計 0 件）

〔産業財産権〕

○出願状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況（計　　件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等

６．研究組織
(1)研究代表者
村上 隆　（MURAKAMI, Takashi）
中京大学・現代社会学部・教授
研究者番号：70093078

(2)研究分担者
（　　　）

研究者番号：

(3)連携研究者
（　　　）

研究者番号：

(4)研究協力者
（　　　）