中京大学スポーツミュージアムの来館者による博物館評価
―カテゴリカルデータのための直交主成分分析によるデータの再分析―[1)]

村　上　　　隆[2)]

1．問題・目的

村上・谷岡・堀（2021）は、2019 年 10 月に開館した中京大学スポーツミュージアムへの来館者に対して行われたアンケート調査のデータを分析し、どのような属性の来館者によって、ミュージアムのどのような側面が評価されているか、現状の博物館展示において何が問題であるかを示した。

調査データは、あらかじめ定められた質問項目への応答と、自由記述に分けて分析されたが、前者については潜在クラス分析（latent class analysis; LCA）、後者については、計量テキスト分析のフリーソフトウェアである KHcoder を中核とする方法に依拠した分析が行われた。

それに対して本研究は、分析の対象を質問項目への応答に絞り、社会調査の領域では広く用いられてきた多重対応分析（multiple correspondence analysis; MCA; Greenacre, 1984; 2017）の拡張である、カテゴリカルデータのための正規直交主成分分析（orthonormal principal component analysis for categorical data; OPCA; 村上, 2016; 2019; Murakami, 2020）を用いた分析結果を報告する。結果としては、村上・谷岡・堀（2021）の結果が大筋で再現されるが、データを異なった側面から見ることになる結果として、若干の新しい知見もある。

OPCA は、LCA が調査の主要部分に的を絞って詳細な分析を行うのに適しているのに対し、調査の全体をおおまかにつかむのに適しており、どちらかと言えば予備的な分析のための方法である。そうした方法の特徴を明らかにすることが本研究の目的である。

さらに、本研究では、反応カテゴリーへの数量化の方法について、デザイン行列にもとづく新たな方式を試みることによって、主観的評定に関する興味ある知見を得ることも目指す。

なお、村上・谷岡・堀（2021）との内容の重複を極力避けるため、博物館評価に関する先行研究の検討、調査の方法の詳細についての記述は割愛する。ただし、アンケートの全体を概観できるように、反応パーセンテージを付した調査票全体を、文末に付録として掲げた。

2．分析方法とデータの概要

2.1　記述的多変量解析

多重対応分析は、記述的多変量解析（たとえば、大隅・ルバール・モリノウ・ワーウィック・馬場, 1994）と呼ばれる手法群の中に位置づけられるデータ分析の方法である。記述的多変量解析の対象と

なるデータは、量的なものと質的なものに分類される。量的データの記述的方法として本研究とかかわりが深いのは主成分分析（principal component analysis; PCA; たとえば、Jolliffe, 2002）である。主成分分析は、多数の量的変数を要約するために、それらに最適な重みをつけた和（1 次合成変量、linear composites）である主成分得点を求めることを目的とする。データの性質によっては、相互に独立な主成分得点を複数求めることもできる。社会科学分野では、これによるデータの縮約（data reduction）だけでなく、主成分得点による信頼性（reliability）の高い個人差測定尺度の構成や、複数の主成分の性質にもとづいた構成概念の探求といった機能が期待される。これらは、探索的因子分析に帰される機能であるが、ここでは主成分分析を広義の因子分析（factor analysis in generic sense）として議論を進めよう[3)]。

2.2　多重対応分析

他方、多重対応分析（たとえば、Benzécri, 1992; Greenacre, 1984; 2017 など）は質的なデータ、たとえば、何らかの質問文（statement）に対して、「1．反対」、「2．どちらかと言えば反対」、「3．どちらとも言えない」、「4．どちらかと言えば賛成」、「5．賛成」のような主観的評定の分析を問題にする。これらをそのまま、5 段階の離散的な量的データとして因子分析にかけることは、しばしば行われている。しかしながら、これらの数字の間が等間隔であるかどうか、さらには、迷いのない反応である 1 と 5 に対して、2〜4 は質的に異なるのではないかといった疑問が生じる。

さらに、職業や居住地域、好きな音楽の分野のように、順序さえつけられないような質問項目も社会調査には多い。こうしたデータはカテゴリカルデータと呼ばれ、量的データとは異なった扱いを必要とする。

多重対応分析では、各反応カテゴリーに対して、それらに対する回答者の反応を、全項目を通じて加算した時に、その分散が最大になるという意味で最適な重みを与える。通常、その重みは得点が独立になるようにしながら 2 組求められ、その値が平面上にプロットされ、主に視察によって解釈が進められる。

2.3　多重対応分析の適用上の問題点

たとえば、図 1 は、ここで分析の対象とするアンケートのうち、次の 4 項目の反応に多重対応分析を適用して得られたカテゴリーの重みと得点である。

Q. 9　あなたが自分でスポーツをした経験について教えてください
Q.10　あなたがスポーツを観戦した経験について教えてください
Q.11　来年開催される東京オリンピックにどのくらい関心をおもちですか
Q.12　全体としてこの Museum のご観覧に満足していただけましたか

Q.9〜Q.11 の反応選択肢（カテゴリー）は、「1．ほとんどない」、「2．あまりない」、「3．どちらとも言えない」、「4．かなりある」、「5．大いにある」の 5 段階、ただし、反応の少ない Q.11 の「1」は「2」に合併し、「ない」として 4 カテゴリーとした。また、Q.12 の選択肢は、「1．不満」、「2．やや不

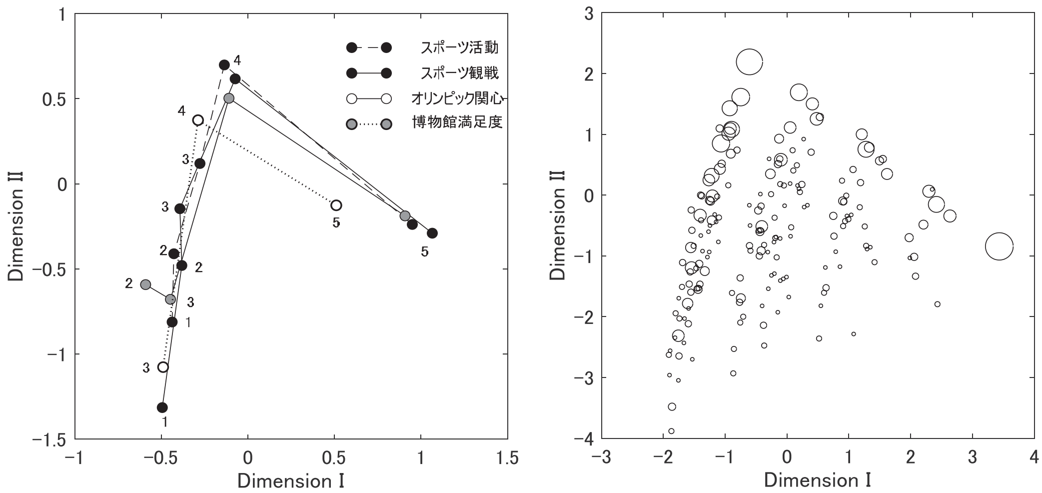

図１　多重対応分析の重み（左）と得点（右）。
得点の円の大きさは同点者の人数に面積が比例するように描かれている。

満」、「3．どちらとも言えない」、「4．ほぼ満足」、「5．満足」であるが、反応数の少ない「1」と「2」は「3」に合併し「満足でない」としている。

なお、以下においては分析方法の記述の慣例に従い、質問項目を変数、回答者を個体と呼んで議論を進めることにする。

ここで、横軸は左から右へ、「ない」から「ある」に向けた経験・関心の量を示している。他方、縦軸については、「かなりある」と「ない」が両極にくるが、その意味は明らかとは言えない。平面上の個人の布置は、しばしば2次曲線状を呈するが、これは馬蹄現象（Gifi, 1981）と呼ばれる（この例では形態はむしろ楔形に近いが）。この扱いには、単に縦軸に対応する次元を除去するというもの（Bekker & De Leeuw, 1989）から、何らかの個人の反応スタイルによると考えるもの（村上, 2019a）まで幅があるが、基本的に未解決の問題である。

以上の結果は、多重対応分析の2つの欠点を示唆している（村上, 2019b；Murakami, 2020)。すなわち、

1．多くの場合、2次元平面上での解釈が中心になり、社会調査項目の多様性を反映するはずの、より多次元の情報が読み取れない
2．馬蹄現象に代表される余剰次元が表れる。

第1点は、visual な representation の限界であるが、第2点については複雑な原因が考えられおり（Greenacre, 1984; Bekker & De Leeuw, 1989)、上述のように、データ解析上の取り扱いも異なるが、ここでは、それには立ち入らず、社会調査データとしての現実的な対処を考える。

2.4 カテゴリカルデータのための正規直交主成分分析

村上（2016; 2019b）、Murakami（2020）は、これらの欠点を補うための方法として、カテゴリカルデータのための正規直交主成分分析（Orthonormal principal component analysis for categorical data; OPCA）を提案した。

OPCA の分析の流れは、おおよそ以下のとおりである。

1）すべての変数の全カテゴリーに、c_k-1 次元の数量を与える。ここで c_k は k 番目の変数のカテゴリーの数とする。その際、数値は中心化（平均値が 0)、標準化（分散が 1)、相互に直交化(相関係数が 0）がなされているものとする。この数量の次元は、全体として $c-p$ となる。ただし、c は変数 k のカテゴリー数を c_k とするとき、これをすべての変数について総和した延べカテゴリー数、p は変数の数である。これらの新規に定義された数量は、質問項目に対応して用いてきた変数（variable）と区別して変量（variate）と呼ぶことにする。

2）得られた（あるいは、定義された）変量を量的データとみなして主成分分析する。

3）主成分得点計算のための重み行列（各列ベクトルの長さ（要素の 2 乗和）を 1 とする）が単純構造をなすように直交回転する。通常、Quartimax 回転（たとえば、芝, 1979）を用いる。

これによって、多重対応分析の解釈の手掛かりは、空間表現から、相対的に相互に独立した[4]軸に変わり、探索的因子分析と同様、3 次元以上の解の解釈も比較的容易にできるようになる。また、分析例で見るように、回転によって馬蹄現象のような余剰次元や（本論文では例示しないが）欠損値によって生じる余剰次元を分離し、解釈に影響を与えないようにすることができる。すなわち、記述的、探索的分析方法としての機能は格段に向上する。

しかし、1)～3）の手続きについては、いくつかの注釈が必要であろう。まず、1）の変量をどのように与えるかであるが、これについては、どのように与えても、中心化、標準化、直交化がなされている限り、主成分得点は全く同じであり、しかもその得点は多重対応分析によるそれと同一であることが証明できる（村上, 2019b；Murakami, 2020)。したがって、後述のように、変量は解釈に都合のよいように任意に決めてよい。

次に、主成分分析の解を回転してよいか、さらに通常、探索的因子分析において回転の対象となる負荷行列でなく、重み行列であることが許されるかと言う問題がある。この点については、最適化基準を、主成分分析の最適化基準を主成分得点の分散の総和の最大化とすることで正当化できる。分散の総和は重みの直交回転について不変だからである。

さらに重み行列の直交回転には、Harris & Kaiser（1964）の「直交変換による斜交解（oblique factor analytic solutions by orthogonal rotation)」の case 2 という先行事例があり、結果的に主成分得点は斜交する。さらに、この解から計算されるパターン行列（各変量から主成分への標準偏回帰係数の行列）には、優れた点が複数ある。まず、斜交回転であるにも関わらず、パターン行列の各列の 2 乗和は、対応する主成分得点による全変量の説明力（分散の大きさ）と解釈でき、さらにその

全主成分を通じた総和は最適化基準である主成分得点の分散の総和と一致する（村上, 2016; 2019b)。これはもちろん、変量間相関行列の固有値の和でもある。さらに、回転の前後にかかわらず、重み行列とパターン行列は列単位で比例する（Kiers & Ten Berge, 1994)。

2.5 カテゴリーの変量をどのように決めるか

変量の定義が任意である限り、主に解釈の対象となるパターン行列も一意に定まらず、結果を解釈することは難しい。解を一意に定めるための１つの方法は、重み行列を通常の後ろからの直交回転だけでなく、変数に対応するブロックごとに（つまり、特定の変数に与えられる変量ごとに）前からも回転することである。もちろん、変量にも同じ回転行列を後ろから掛ける。これによってパターン行列の単純構造の度合いは高まり、解釈はより容易になるはずである。

しかし、実際にやってみると、この方法による結果の解釈は大変手間のかかるものになる。一般に負荷行列の要素の解釈には、対応する変量の意味を解釈者が分かっていることが不可欠である。通常の因子分析では、質問文や物理量として与えられる変数の意味は自明である。しかし、回転によって変えられた変量は、カテゴリーとの関係をいちいち参照しない限りわからない。つまり、普通の探索的分析では不必要なステップが１つ必要となるのである。これを、すべての変数と主成分の組み合わせについて行うことは大変効率が悪い。適当な手掛かりがない場合には、乱数によって変量を与え、前述のように重み行列の前後からの回転によって変量を最終的に確定させるという手段もあり得るが、本研究ではその方法はとらない。

変量とカテゴリーの関係を参照しなくても済む方法として、現在のところ、次の 2 つの方法を試みている。

2.5.1 直交多項式による変量の定義

順序のあるカテゴリーに対しては、カテゴリーの番号（1, 2, 3, ..., c）に対して、1 次式、2 次式といった多項式を当てはめるのは自然である。カテゴリーの番号は、単なる順序にすぎず、等間隔とは言えない（間隔尺度でなく順序尺度）が、等間隔でない部分は高次の成分で補えることが期待できると考えるわけである。

このやり方で問題なのは、3 次以上の高次の成分の解釈がしばしば困難であることである。ただし、より詳細なグラフ化により、3 次以上の高次の成分にも意味付けが可能な場合もある（村上, 2019a)。

2.5.2 ダミー変数による変量の定義

必ずしも順序をもたないカテゴリーに対しては、より自由な設定ができることが望まれる。あるいは、順序はあるとしても個人の反応スタイル等の影響で、必ずしも単調な変動であるとは考えにくいケースにも対応できればよい。そうした手段として、実験計画法で用いられるデザイン行列を用いることが考えられる。

ただし、この方法はあまりにも自由度が大きすぎ、明確な仮説が考えられる場合以外には、かえって使いにくいものとなる。比較的穏当なところでは、重回帰分析等で用いられる特定のカテゴリーのみ 1、他のカテゴリーを 0 とするダミー変数を 個定義することであろう。以後、このやり方をダミー

コーディングと呼ぶ。これについては、あまり利用経験が蓄積されておらず、本研究が最初の適用例となる。

2.5.3 中心化を含む正規直交化

前述のように、OPCA を適用するためには、すべての変量は中心化、標準化、直交化されなければならない。そのためには、Gram-Schmidt の直交化法、あるいはそれと同値な QR 分解を用いる (村上, 2019b; 2020)。ただし、この変換の結果、それぞれ、0, 1, 0 となる平均値、分散、相関係数は、単に個々の値について単純に算出されるのではなく、それぞれに対応する度数で重みづけられる必要がある。したがって、既成の統計表にある正規直交多項式の数値をそのまま用いることはできないし、Gram-Schmidt の直交化法にも修正が必要である（村上, 2016; 2019b)。

また、正規直交化の順序も問題になる。多項式の場合は、低次の項から順に直交化するのが常識であろうが、ダミー変数の場合には判断が必要になる。

2.6 データの概要

中京大学スポーツミュージアムは 2019 年 10 月 24 日㈭から一般公開されたが、この調査は、その日から 27 日㈰までの 4 日間の間の来館者を対象に実施された。未就学児を除く来館者数は 1,112 名、そのうち、967 名が回答した（回収率 87.0%)。第 1 日目には、複数の授業の受講生が含まれ、2 日目以降は大学祭であったため、多数の学外者が来館したため、かなり多様な回答者を集めることができた。

データ収集の手続きの詳細については、村上・谷岡・堀（2021）に記したので、ここでは繰り返さない。調査票の全項目は、各カテゴリーへの反応率とともに、文末付録に示した。なお、この調査の実施については、「中京大学における人を対象とする研究に関する倫理審査委員会」の審査を受け承認された。

3．分析の結果

3.1 OPCA を適用するデータの範囲

OPCA はカテゴリカルなデータであれば、どのような内容のものであっても適用できる。従来の適用例（村上, 2019b; 2020）においては、性別、年齢、職業といったデモグラフィックな変数も分析に加えたが、今回はそうしたいわば「客観的」な変数は除き、回答者の主観にもとづくもの、すなわち、Q.1（スポーツミュージアムへの全般的評価、5 段階評定)、Q.2（印象に残った展示、チェックリスト形式)、Q.3（スポーツミュージアムのことをどこで知ったか、「その他」を含む 5 つの選択肢)、および、Q.9～12（2.3 で示した、スポーツの実行と観戦の頻度、オリンピックへの関心、スポーツミュージアムへの全体的満足度）の全 21 項目を対象とした。

Q.4（性別)、Q.5（年齢)、Q.7（職業）は、OPCA のスコアの分析の際の要因として使用する。

なお、文末付録から読み取れるように、特に評価に関わる Q.1 の 7 項目と Q.12 は著しく高評価

(「5」) の方向に分布が偏っている。OPCA において、極端に度数の少ないカテゴリーを含むことは、それほど有害ではないが、やはり解釈に困難をきたすことが多い。したがって、Q.1 の 7 項目については、「悪い」、「やや悪い」を「普通」と合併して、「普通以下」と表記し、すべて 3 段階で扱うことにした。これは、村上・谷岡・堀 (2021) において、潜在クラス分析を適用した際のデータの扱いと一致する。Q.11 と Q.12 については、2.3 で述べたように、Q.11 は 4 段階、Q.12 は 3 段階で扱っている。

以上の対処により、分析の対象となるデータに反応率が 5%を下回るものは存在しない。その結果、分析の対象となるのは 21 の変数の延べ 61 カテゴリーであり、したがって、変量の数は、40 (＝61－21) となる

以下の分析において、欠損値はリストワイズに除去されている。その結果、分析の対象となるのは、欠損値が存在しない 815 のケースに限定される[5)]。

3.2 変量の定義と予備的分析 (1)：直交多項式による数量化

上記の範囲の全変数を用いた分析を行う前に、一部の変数のみを用いた結果を示す。まず、2.5 で述べたカテゴリーに対する数値の付与、すなわち変量の定義について見ておきたい。

表 1 は、Q.9 の 5 つのカテゴリー (「1」：ほとんどない～「5」：大いにある) に対して当てはめられた数値であり、4 つの変量 (i～iv) である。「定義」では、1, 2, 3, 4, 5 の整数 (カテゴリーの番号) に対して、1 乗から 4 乗までの数値が与えられており、それを中心化 (平均が 0)、標準化 (分散が 1)、直交化 (相互相関が 0) がなされるように変換したものが「変換後」である。それぞれの統計測度は、度数で重みづけて算出される必要がある。変換後の i が公差約 0.78 の等間隔であること、ii が放物線の形状をとっていること等が読み取れるであろう。変換後の変量を以後、「1 次」～「4 次」と呼ぶ。

残りの Q.10～Q.12 についても多項式をあてはめるが、Q.11 については「1」を欠くので最大 3 次式まで、Q.12 には「1」と「2」がないので、2 次式までにとどまる。

表 2 は、以上のようにして定義された変量間の相関行列である。これは変数ごとの 4 つのブロックに分割することができ、対角ブロック (実線で囲んだ部分) は、標準化と直交化によって単位行列となっていることに注意。固有値は、2.14, 1.50, 1.29, 1.11, 1.02, … であるが、MCA との比較のため、主成分数は 2 として議論を進める。

表 1　Q.9 に対する直交多項式による変量の定義

カテゴリー	定義				変換後				度数
	i	ii	iii	iv	i	ii	iii	iv	
「1」	1	1	1	1	-1.83	2.04	-1.56	0.61	72
「2」	2	4	8	16	-1.05	-0.19	1.22	-1.00	176
「3」	3	9	27	81	-0.27	-1.10	0.21	1.74	152
「4」	4	16	64	256	0.51	-0.67	-1.14	-0.78	226
「5」	5	25	125	525	1.28	1.09	0.65	0.23	189

表 2　13 の変量間の相関行列

変　量		Q.9				Q.10				Q.11			Q.12	
		1 次	2 次	3 次	4 次	1 次	2 次	3 次	4 次	1 次	2 次	3 次	1 次	2 次
Q. 9　スポーツ	1 次	1.00	0.00	0.00	0.00	0.44	0.12	0.07	0.02	0.35	0.10	0.04	0.23	0.02
する	2 次	0.00	1.00	0.00	0.00	0.10	0.33	0.04	-0.02	0.09	0.15	0.03	0.16	0.15
	3 次	0.00	0.00	1.00	0.00	0.07	0.08	0.14	0.05	0.07	0.02	0.09	0.02	0.13
	4 次	0.00	0.00	0.00	1.00	0.03	0.03	0.03	0.15	0.04	0.04	-0.01	0.04	0.04
Q.10　スポーツ	1 次	0.44	0.10	0.07	0.03	1.00	0.00	0.00	0.00	0.37	0.09	-0.05	0.20	0.04
見る	2 次	0.12	0.33	0.08	0.03	0.00	1.00	0.00	0.00	0.12	0.19	0.06	0.16	0.10
	3 次	0.07	0.04	0.14	0.03	0.00	0.00	1.00	0.00	0.07	0.11	0.09	0.04	0.07
	4 次	0.02	-0.02	0.05	0.15	0.00	0.00	0.00	1.00	0.01	-0.03	0.14	-0.08	0.06
Q.11　オリンピック	1 次	0.35	0.09	0.07	0.04	0.37	0.12	0.07	0.01	1.00	0.00	0.00	0.32	0.00
関心	2 次	0.10	0.15	0.02	0.04	0.09	0.19	0.11	-0.03	0.00	1.00	0.00	0.07	0.12
	3 次	0.04	0.03	0.09	-0.01	-0.05	0.06	0.09	0.14	0.00	0.00	1.00	-0.02	0.18
Q.12　満足度	1 次	0.23	0.16	0.02	0.04	0.20	0.16	0.04	-0.08	0.32	0.07	-0.02	1.00	0.00
	2 次	0.02	0.15	0.13	0.04	0.04	0.10	0.07	0.06	0.00	0.12	0.18	0.00	1.00

表 3　主成分負荷行列（直交多項式による）

変　量		MCA		OPCA		R^2
		I	II	I	II	
Q. 9　スポーツ	1 次	0.65	-0.30	0.72	0.00	0.52
活動	2 次	0.38	0.42	0.09	-0.55	0.31
	3 次	0.18	0.30	-0.01	-0.35	0.13
	4 次	0.09	0.10	0.02	-0.13	0.02
Q.10　スポーツ	1 次	0.64	-0.35	0.73	0.05	0.54
観戦	2 次	0.42	0.43	0.11	-0.57	0.34
	3 次	0.18	0.26	0.02	-0.31	0.10
	4 次	0.00	0.21	-0.11	-0.19	0.05
Q.11　オリンピック	1 次	0.67	-0.29	0.73	-0.01	0.53
関心	2 次	0.30	0.34	0.07	-0.44	0.20
	3 次	0.08	0.44	-0.17	-0.44	0.22
Q.12　満足度	1 次	0.57	-0.15	0.57	-0.10	0.33
	2 次	0.20	0.55	-0.13	-0.59	0.37
主成分分散	1 次	1.61	0.32	1.91	0.01	1.93
	2 次	0.45	0.78	0.04	1.17	1.23
	3 次	0.07	0.35	0.03	0.42	0.42
	4 次	0.01	0.05	0.01	0.05	0.06
	合計	2.14	1.50	2.00	1.65	3.65
主成分間相関						
	I	1.00	0.00	1.00	-0.15	
	II	0.00	1.00	-0.15	1.00	

表 3 は、負荷行列（斜交解である OPCA についてはパターン行列）を示した。それぞれの変数に対応する変量は、付与された多項式の次数であらわしている（以後、1 次変量、2 次変量等と呼ぶ）。2.4 で述べたように、OPCA は MCA の解を直交回転したものにすぎないので、主成分分散の和は同

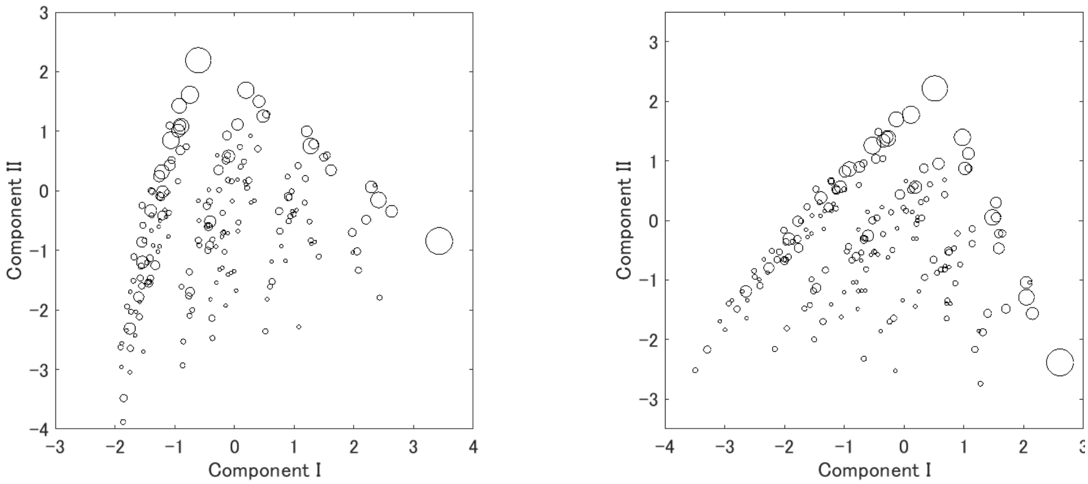

図 2　MCA の得点（左、再掲）と直交多項式による OPCA の主成分得点（右）

一である（3.65）。この値は、データ全体の分散 13 の 28.0% であり、通常の主成分分析で期待される値より相当小さいといわなければならない。この程度の小規模で比較的同質的な変数からなるデータでも、2 次元解に限定して用いられる MCA のデータ説明力は大きいとは言えない（説明されるイナーシアの比率（村上, 2019; p.114）では、約 44.6%）。

回転により、OPCA 負荷行列の単純構造の度合いは大きくなっている（1 つの変数が複数の主成分に高く負荷するいわゆる cross loadings が少ない）。第 I 主成分がほぼ 1 次変量のみからなるようになったが、このことは図 2 からも明らかである。2 つの散布図は回転によって重なる。第 I 次元と 4 つの項目反応の単純和（sum score：反応したカテゴリー番号の合計）との相関は MCA で 0.894、OPCA で 0.979 である。

第 II 主成分には、2 次変量が主に負荷しているが、3 次変量の負荷も比較的大きく、輪郭は、ややひずんだ放物線状を呈しており、ピークに来るのは、4 つの項目への反応がすべて「2」である回答者である。これと右端にある、すべての回答が「5」である回答者との間の線上には、反応が 1 つずつ「4」から「5」に変わっていく回答者が並ぶ（それぞれ、あり得るパターンの数である、4, 6, 4 個の円が区別できる）。ピークの左はそれほどはっきりした区別ができないが、左下隅に回答が「1」、「1」、「2」、「3」である回答者が 1 人いることが見てとれる。

3.3　変量の定義と予備的分析（2）：ダミーコーディングによる数量化

表 4 に、表 1 と同じ Q.9 に対するダミーコーディング（特定のカテゴリーのみ 1 とし、他は 0 とする）による変量の定義を示した。

2.5.3 で述べたように、直交化に際しては、その適用順序が問題になる。ここでは、両端の「5」、「1」から始めて、「4」、「2」という順序をとった。「3」はいわば reference point となる。中心化の結果、コード 1 が付与されたカテゴリーと残りのカテゴリー（同じ数値になる）との対比となるが、直

表 4　Q.9 に対するダミーコーディングによる変数定義

カテゴリー	定義				変換後				度数
	i	ii	iii	iv	i	ii	iii	iv	
「1」	0	1	0	0	-0.55	3.17	0.00	0.00	72
「2」	0	0	1	0	-0.55	-0.41	-1.01	1.46	176
「3」	0	0	0	0	-0.55	-0.41	-1.01	-1.70	152
「4」	0	0	0	1	-0.55	-0.41	1.46	0.00	226
「5」	1	0	0	0	1.82	0.00	0.00	0.00	189

表 5　主成分負荷行列（ダミーコーディングによる）

変量		MCA		OPCA		R^2
		I	II	I	II	
Q. 9　スポーツ活動	「5」	0.76	-0.16	0.78	0.00	0.61
	「1」	-0.07	-0.34	0.03	-0.35	0.12
	「4」	0.13	0.43	0.00	0.45	0.20
	「2」	-0.07	-0.21	-0.01	-0.22	0.05
Q.10 スポーツ観戦	「5」	0.76	-0.17	0.78	-0.02	0.60
	「1」	-0.08	-0.37	0.03	-0.38	0.14
	「4」	0.20	0.49	0.06	0.52	0.28
	「2」	0.01	-0.13	0.04	-0.12	0.02
Q.11 オリンピック関心	「5」	0.70	-0.12	0.71	0.02	0.50
	「2」	-0.13	-0.21	-0.07	-0.23	0.06
	「4」	0.20	0.58	0.03	0.61	0.37
Q.12　満足度	「5」	0.60	-0.12	0.61	0.00	0.38
	「4」	0.09	0.56	-0.07	0.57	0.32
主成分分散	「5」	2.00	0.09	2.09	0.00	2.09
	「1」	0.01	0.25	0.00	0.26	0.27
	「4」	0.11	1.06	0.01	1.16	1.17
	「2」	0.02	0.10	0.01	0.12	0.12
	合計	2.14	1.50	2.11	1.54	3.65
主成分間相関						
	I	1.00	0.00	1.00	0.08	
	II	0.00	1.00	0.08	1.00	

交化により、それ一旦コード 1 が付されたカテゴリーは以後の変量では 0 が付与されるので、順次 0 がふえていき、上の例で言えば最後の数量 iv はここで 1 が付与された「4」と reference の「3」との対比となる。適用順に関して、「4」と「1」を逆にすることも試みたが、結果には大きな違いは生じなかった。

異なった数量化が行われるから、変量間の相関行列も当然異なったものとなるが、同じデータの MCA と同値であるから、その固有値は、表 2 のものと同一である。

表 5 に負荷行列等を示した。こちらの方は、MCA とほとんど同じパターンを示している。すなわち、MCA の解は、そのままでダミーコーディングによる変量については、単純構造となるような解であることがわかる。

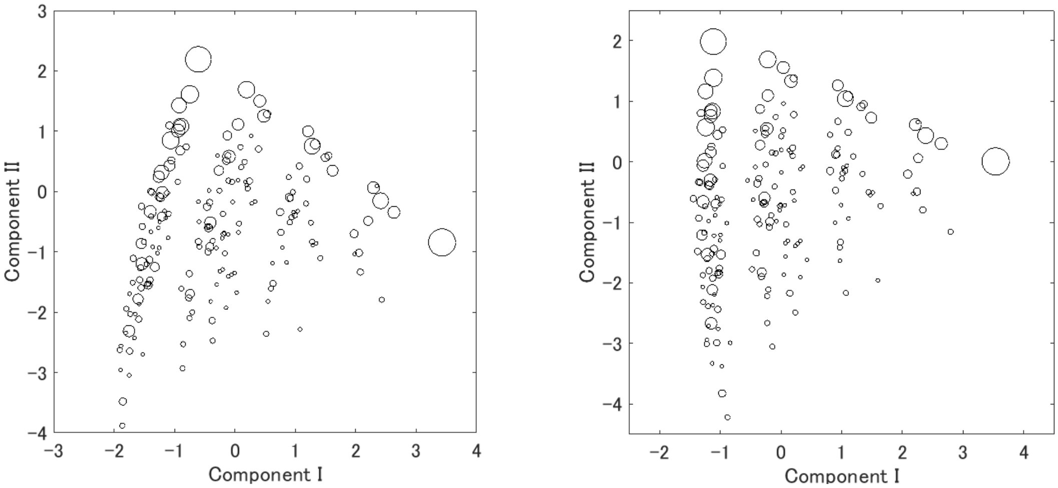

図 3　MCA の得点（左、再掲）とダミーコーディングによる OPCA の主成分得点（右）

図 3 からわかるように、回転の角度はわずかではあるが、それによって、第 I 次元は 4 項目中の「5」反応の数に対応し、第 II 次元は「1」から「4」への 1 次元の尺度となっている。このように見ると、第 II 次元を余剰と見ては、重要な変動を見逃すことになりそうである。

こうした傾向の解を産出するのが、このデータだけの特徴なのか、もっと一般的に見られるものであるのかについては、今後の検討課題である。ただし、次に示す Q.1 の 7 項目についての同様の分析は、この傾向をより明確に示しているように思われる。

3.4　スポーツミュージアムへの全般的評価（Q.1）の分析

Q.1 の 7 項目の質問文は、以下のとおりである。

1. 入館した時の印象
2. 展示物の価値・希少性
3. 展示のレイアウト
4. 説明のわかりやすさ
5. 提供されている情報の量
6. 館内の雰囲気・居心地のよさ
7. Museum へのアクセス・館内での動きやすさ

これらの項目については、すべて 5 段階の選択肢が付されているが、4.1 で述べたように、反応の少ない選択肢を合併して「普通以下」、「やや良い」、「良い」の 3 段階に統一されている。負荷行列等については省略するが、Q.9 の場合と同様、直交多項式とダミーコーディングによる OPCA のスコアを、MCA のスコアと並べて図示する（図 4，5）。

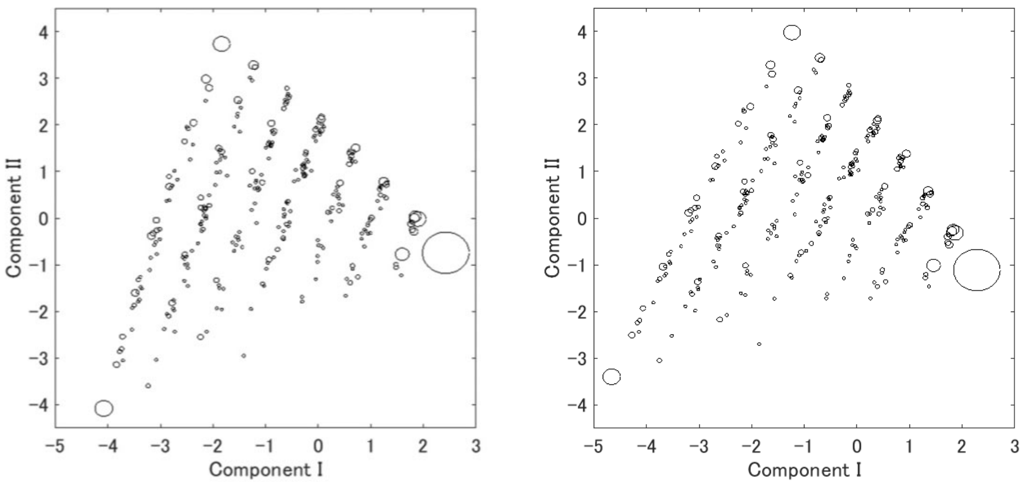

図 4　Q.1 の MCA（左）と直交多項式による OPCA（右）の得点

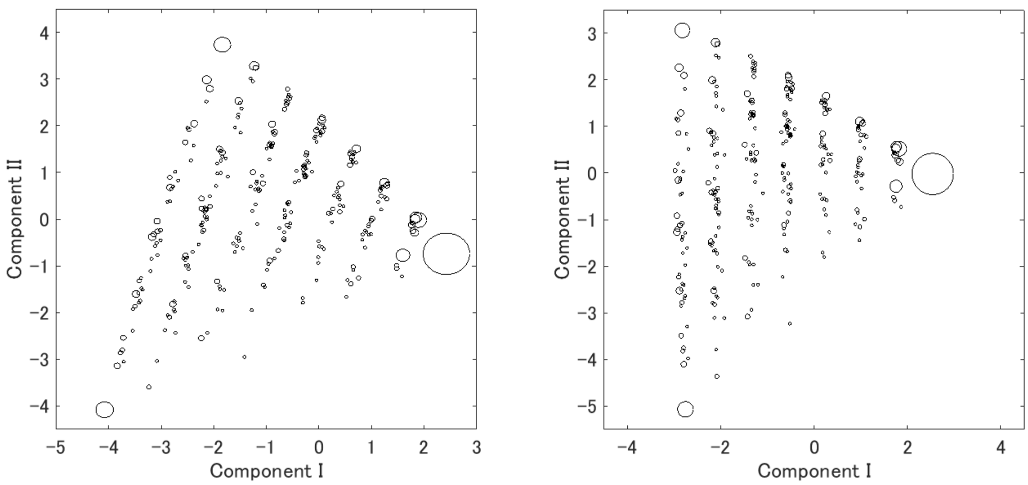

図 5　Q.1 の MCA（左、再掲）とダミーコーディングによる OPCA（右）の得点

図 4 の直交多項式による結果は、ほとんど MCA と違いがないが、反応が 3 段階に統一されていることも相俟って、右端の全部「良い」反応の回答者から、1 点刻みでピークのすべて「やや良い」の回答者につながり、さらに一点刻みですべて「普通以下」の回答者に至る過程は見やすい。

他方、図 5 のダミーコーディングによる変量定義にもとづくスコアは、直交多項式の場合とは逆方向のかなり大きな回転がなされ、図 3 の場合と同様、第 I 次元は評定「5」の数の 8 段階（0〜8）が明確に読み取れる。第 II 次元についても、第 I 次元ほど明瞭ではないが、左端の「5」が全くない回答について、上から下に「4」から「3」にかけて 8 段階の区別が認められる。この 2 本の線分が作る三角形の 2 辺の内側には、どのような割合にせよ「5」と「3」が共存する（もちろん、多くの場合「4」も）が共存する反応パターンが配列されている。

このように考えると、いわゆる馬蹄現象は余剰次元として捨て去られるべきでなく、複数の質問へ

の反応パターンをコンパクトに表現するための有効な手段と見ることができるのではないかとも考えられる。事実、Q.1 の質問項目への 3 段階の反応は、これらを数値とみなしても、相互に高い相関をもち、反応の単純和によって対象（この場合スポーツミュージアム）への評価を 1 次元で表現できるようにも思われる。しかしながら、5×7＝35 点（最高の評価）と 3×7＝21（最低の評価）以外は、複数の反応パターンが同じ得点を生み出す。たとえば、「5」が 3 項目、「3」が 3 項目、「4」が 1 項目といった反応と、すべての項目に「4」という場合は、ともに単純和は 28 となり、区別がつかない。この 2 次元の表現によれば、これらの区別が可能になる。

こうした性質が 3 値以上の場合にどこまで拡張可能性があるのか、また変数がどの程度異質的な (明確な複数次元があるといった) な場合に対する頑健性があるのかといった問題は、今後に残される。

4．全変数の同時分析

4.1 変量の定義

変量の定義、とくにダミーコーディングによる数量化の性質がある程度明らかになったところで、ここではデモグラフィックな変数を除く全変数の分析を行う。本文中で紹介していなかった Q.2 (「次のコーナーのから印象に残ったものの番号に〇をつけてください【〇はいくつでも】」) の全変数は次のものである（数字は選択率)。

1．スポーツ映像ウォール（エントランス）27.0%
2．真剣味の殿堂（中京大学のオリンピアン）34.9%
3．時代とスポーツのスラローム（夏）18.8%
4．時代とスポーツのスラローム（冬）18.4%
5．スポーツアーカイブス　11.8%
6．モスクワ五輪のメダル　29.5%
7．企画展示コーナー（金栗四三）22.0%
8．映像ブース　23.7%
9．その他（できれば具体的に書いてください）16.1%

これらの 2 値変数についての変量の定義は、中心化と標準化によって一意に定まるので、数量化の問題は生じない。具体的には、選択率を p (0＜p＜1) とすると、平均値は p，分散は p (1−p) となるので、変換後の値は、1（選択）と 0（非選択）に対して、それぞれ、

$$\frac{1-p}{\sqrt{p(1-p)}}, \qquad \frac{-p}{\sqrt{p(1-p)}}$$

となる（カテゴリー数が 2 なので、対応する変量は 1 つであり、直交化の必要はない)。

表 6　Q.3 に対するダミーコーディングによる変量の定義

カテゴリー	i	ii	iii	iv	i	ii	iii	iv	頻度
メディア	0	1	0	0	-0.76	1.29	0.00	0.00	248
ウェブサイト	0	0	0	0	-0.76	-1.22	-1.50	-3.31	49
キャンパスに来て	1	0	0	0	1.31	0.00	0.00	0.00	306
クチコミ	0	0	1	0	-0.76	-1.22	2.05	0.00	121
その他	0	0	0	1	-0.76	-1.22	-1.50	1.61	91

その他の変量については、すべてダミーコーディングによるものとした。反応カテゴリーに順序のある Q.1, 9〜12 についての直交化の順序は、「5」→「4」→「1」→「2」とした（カテゴリーの合併により、存在しなくなったカテゴリーは省く）。順序のない Q.13 については、反応数の多い順、「メディア」、「キャンパスに来てから」、「クチコミ」、「その他」の順とした（表 6）。

4.2　主成分の数の判断

解釈の対象とする主成分の数の判断は難しい。できる限り多くの興味ある変動を見出したいという要請がある一方で、そうした変動に再現性があることを、ある程度保証する必要がある。通常、判断の基準とされる固有値の推移（scree plot）は、図 6 に見られるように、elbow point として 4 を示唆しており、Velicer（1976）の MAP 基準も 4 が最適であることを示す。しかしながら、4 主成分の解だけでは、質問項目間の関係について、特に興味深いことが見いだされたとは言えなかった。主成分の数を 7、あるいは 8 とした場合に興味ある結果が得られるが、その再現性の確認のために、bootstrap 法（Efron & Tibshirani, 1993）を用いた検討を行った。

Timmerman, Kiers, & Smilde（2007）に従い、全サンプルによって算出した、回転した重み行列を target として、resampling によって得られた重み行列を Procrastes 回転し、target との内積

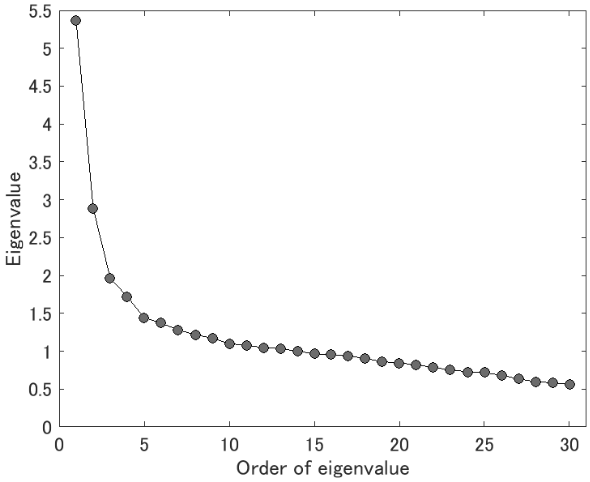

図 6　固有値の推移

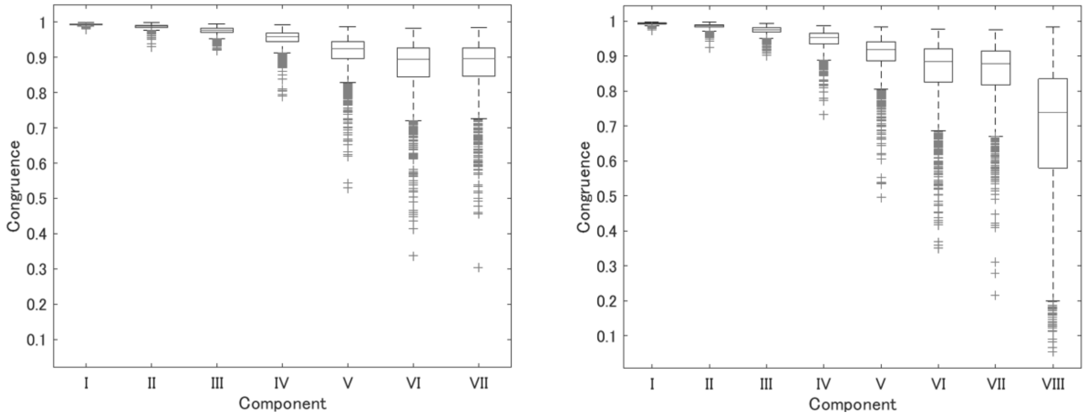

図 7　Bootstrap 法による 7 主成分（左）と 8 主成分（右）における target との一致度

(一致度；congruency）を求め、これを 2000 回反復して、各回の結果を記録した。図 7 に、7 主成分解と 8 主成分解の結果を box plot によって示した。明らかに 8 主成分解における第 VIII 主成分の一致度が低い。また、第 IV から第 VII 主成分についても、四分偏差（box）が大きく、主成分の数を大きくすると解が不安定化する傾向がみられる。これは、8 主成分解より 7 主成分解をとる 1 つの理由になる（7 主成分による変量の分散の説明率は 40.1%、イナーシアの説明率は 71.4%である)。

4.3　負荷量の有意性

そこで、暫定的に主成分数を 7 とした解の、各変数の主成分への標準偏回帰係数であるパターン行列を表 7 に示した。ここではいくつかの数値を、以下のような基準で、解釈の参考にするために太字にした。まず、各要素について、2000 回の resampling の反復による平均値と標準偏差を求めた。この標準偏差を標準誤差とみなし、どの要素も真の値が 0 であるという帰無仮説の下で、平均値を標準偏差で割って z 値を求めた。

このような同一の検定の反復に関しては、多重度を考慮する必要があるが、ここでは便宜的に 値を、(1) 3.6 以上（Bonferroni 流に有意水準（片側）0.05 をパターン行列のすべての要素の数で割って得られた水準に対応する臨界値に近い)、(2) 3 以上（単独の検定で 0.01%水準に近い)、(3) 1.7 以上（単独の検定で、5%水準に近い）(4) 1.7 未満（有意でない）の 4 種類に分けた。その上で、図 7 に見られるように、きわめて安定度が高いと考えられる主成分 I、II、III については（1）の基準を満たした上で（効果量という観点も考慮し）負荷量自体が 0.1 を下回らないもの、安定度が中程度の主成分 IV、V については（2）を満たすもの、さらに安定度の低い主成分 VI、VII については、(3)を満たすものまで含めた。結果的には、全主成分にわたり、ほぼ 0.3 を超える負荷量を選択することになった。

ただし、解釈にあたっては、太字で強調された項目が同等の「確かさ」をもつものでないことは考慮する必要がある。主成分 I、II、III においては、これらの「目立つ」負荷量が、(ほぼ）同一条件で反復された調査において現れないとは考えられない。主成分 IV、V ではその確実性はやや劣るも

表 7　全変数の負荷（パターン）行列（7 主成分）

		I	II	III	IV	V	VI	VII	R^2
Q.1-1	「5」	0.71	0.08	-0.06	0.02	0.07	0.02	-0.03	0.53
	「4」	-0.03	0.48	0.10	-0.13	-0.09	-0.03	0.04	0.25
Q.1-2	「5」	0.73	0.05	-0.02	0.05	0.04	-0.11	0.08	0.56
	「4」	-0.06	0.66	0.02	-0.06	0.00	0.07	-0.10	0.44
Q.1-3	「5」	0.82	0.01	-0.02	-0.02	0.02	-0.02	-0.01	0.66
	「4」	-0.02	0.69	-0.03	0.01	-0.16	0.08	0.10	0.48
Q.1-4	「5」	0.83	-0.04	0.00	0.02	-0.02	-0.04	0.03	0.67
	「4」	0.04	0.68	-0.06	-0.01	-0.04	0.04	0.02	0.46
Q.1-5	「5」	0.85	-0.08	-0.02	-0.02	-0.04	0.01	0.05	0.68
	「4」	0.03	0.66	0.01	0.00	0.06	0.00	-0.01	0.47
Q.1-6	「5」	0.84	0.03	-0.04	-0.06	0.00	0.05	-0.03	0.69
	「4」	-0.06	0.69	0.04	0.07	0.04	-0.09	-0.03	0.49
Q.1-7	「5」	0.77	-0.05	0.02	0.01	-0.05	0.08	-0.05	0.60
	「4」	0.04	0.49	0.02	0.06	0.05	-0.12	-0.04	0.28
Q.2-1		0.13	0.05	0.04	-0.08	-0.03	0.38	-0.12	0.18
Q.2-2		0.04	0.06	0.19	0.27	0.00	-0.16	0.34	0.26
Q.2-3		-0.05	0.00	0.03	0.76	0.07	0.06	0.02	0.60
Q.2-4		0.03	0.02	-0.03	0.78	-0.04	0.02	-0.05	0.61
Q.2-5		0.00	-0.06	0.03	0.45	-0.01	0.09	-0.05	0.22
Q.2-6		0.13	-0.02	-0.05	0.19	0.05	0.42	-0.07	0.25
Q.2-7		0.03	0.00	0.05	0.03	0.02	0.38	-0.12	0.16
Q.2-8		0.12	0.05	0.01	0.03	-0.12	-0.08	-0.06	0.04
Q.2-9		0.09	-0.03	-0.05	-0.14	0.05	-0.64	-0.10	0.47
Q.3	キャンパス	-0.03	0.03	-0.01	-0.04	0.05	0.34	0.32	0.23
	メディア	0.02	-0.11	-0.36	0.22	-0.02	-0.22	0.16	0.25
	クチコミ	0.07	-0.08	-0.09	-0.17	-0.10	0.42	-0.01	0.21
	その他	0.03	-0.08	-0.12	-0.06	0.14	0.15	-0.22	0.10
Q.9	「5」	-0.07	0.04	0.80	-0.02	0.03	0.06	-0.01	0.64
	「4」	0.00	-0.03	-0.02	0.01	0.58	0.17	-0.01	0.36
	「1」	-0.03	-0.01	0.03	0.07	-0.02	-0.02	-0.65	0.42
	「2」	0.01	-0.01	-0.02	0.29	-0.18	-0.19	-0.11	0.15
Q.10	「5」	-0.01	-0.04	0.79	0.00	0.00	-0.08	0.01	0.61
	「4」	-0.05	-0.04	0.10	-0.02	0.69	-0.15	-0.10	0.50
	「1」	0.03	0.04	0.02	0.08	-0.13	-0.03	-0.39	0.18
	「2」	0.03	0.12	0.01	0.02	-0.07	0.01	-0.40	0.18
Q.11	「5」	0.04	-0.03	0.67	0.07	0.03	0.06	0.02	0.49
	「4」	0.04	-0.01	-0.02	0.04	0.62	0.01	0.13	0.41
	「2」	-0.02	-0.03	-0.02	-0.02	0.18	-0.06	-0.49	0.27
Q.12	「5」	0.58	-0.01	0.36	-0.02	-0.04	-0.08	-0.01	0.56
	「4」	0.14	0.42	-0.20	0.05	0.34	0.00	0.03	0.41
主成分分散		4.83	3.02	2.11	1.74	1.51	1.44	1.39	16.03

のの、ほぼ確信をもって解釈を行うことができるであろう。主成分 VI、VII は、同様の意味をもつ主成分が出現するか否かも含めて、あくまでも仮説的なものとして受け取っておく必要がある。

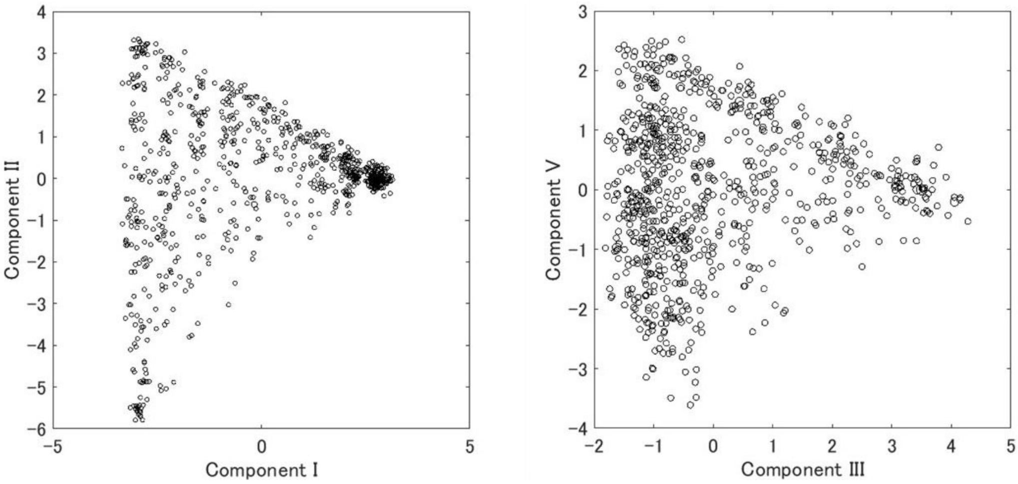

図 8　主成分 I と II（左）と主成分 III と V の散布図

4.4 主成分の解釈

その上で、主成分の解釈に移ろう。まず、個々の主成分についての解釈を行い、その後に主成分間相関行列等を用いて、それらの主成分間の関係を論じることにしたい。ただし、主成分 I と II、ならびに III と V は対として考えることができる。図 8 は、それぞれの散布図である。これらは Q.1（スポーツミュージアムへの評価）と Q.9〜11（スポーツへの関心）、Q.12（スポーツミュージアムへの満足度）の質問項目の対して、「5」という反応（「良い」、「大いにある」、「満足」）にコード 1 を与えるダミーコーディングと、「4」にコード 1 を与えるダミーコーディング（「やや良い」、「かなりある」、「ほぼ満足」）に対応している。したがって、図 8 の左右の散布図は、それぞれ図 5 左、および、図 3 右に対応しているわけであるが、それよりは、ややぼやけた印象を与える。それは、それぞれが高く負荷している Q.1（スポーツミュージアムへの評価）と Q.9〜11（スポーツへの関わり）の項目だけでなく、他の項目への反応からも多少の影響を受けているために、完全に同一な反応パターンが滅多に生じないからである。なお、Q.12（スポーツミュージアムへの満足度）の 2 つの変量は、主に主成分 I と II に負荷しているが、主成分 III と V にも無視できない負荷を示している。

なお、主成分 III には、Q.3（スポーツミュージアムのことを何で知ったか）の「メディア」に対して無視できない負の負荷を示している。スポーツへの関心が高い人々は、メディアが報道する以前に、ミュージアムのことを知っていた可能性が高いことを示している。

次に主成分 IV は、Q.2（印象に残った展示）の数項目が集中して負荷している。特に目立つのが「時代とスポーツのスラローム（夏）」、「時代とスポーツのスラローム（冬）」「スポーツアーカイブス」の 3 項目であり、これらは「その他」を除き、反応率が 10%台と他の項目より低いものである（4.1 参照）。Q.2 の項目群は、Q.1、Q.9〜12 とは内容的な差異もあるが、チェックリストという形式の差もあり、どうしても、評定型の項目とは別の主成分にまとまる傾向があると思われる。

主成分 VI は、やはり Q.2（印象に残った展示）のうち、「スポーツ映像ウォール（エントランス）」、

「モスクワ五輪のメダル」、「企画展示コーナー（金栗四三)」に正に負荷する一方、「その他」に大きな負の負荷をもっている。最初の3つは、反応率が20%台と比較的高いものである。また、Q.3（スポーツミュージアムのことを何で知ったか）では、「キャンパスに来てみて」と「クチコミ」に正の負荷がある。この意味については、注で述べたい[6]。

主成分VIIは、Q.2（印象に残った展示）から、「真剣味の殿堂（中京大学のオリンピアン)」(これは、Q.2の中で、最も反応率の高い項目である)、Q.3では「キャンパスに来てみて」、そして、Q.9〜11の「1」または「2」、つまり関心の低さを示す変量に、無視しえない負の負荷を示している。事前に調べてきたわけではないが、「中京大のオリンピアン」(の数多さ）に感銘を受ける人が多い一方で、そうでない人は、スポーツへの関心が低い傾向があることが示されている。

4.5 主成分得点の分布

図9に7つの主成分の度数分布を示した。分布の形状の比較のために、すべてbinの数は21とした。

順に見ていくと、主成分I（Component Iと表記、以下同じ）は、2つのピークをもつ双峰性の分布である。他方、これと対をなす主成分IIは中央の0のところに高いピークのある分布で、いずれも正規分布からは大きく隔たっている。この主成分に高く負荷するQ.1（スポーツミュージアムへの一般的評価）は「5」(良い）とそれ以外という2値変数であり、相互相関も高いため、主成分Iの高得点側にピークを作り、他の反応「3」と「4」は低得点側に集中してもう1つのピークを作る結果となると解される。つまり、ダミーコーディングで定義する限り、評価は2極化している。他方、主成分IIでは、その集中する「5」の値が0となって分布の中心に高いピークを作ることになり、対応して両極に広いすそ野をもつことになる。

主成分IIIと主成分Vは、5段階（Q.9, 10)、4段階（Q.11)、3段階（Q.12）と3つのタイプの項目が混在し、「5」(「大いに」）反応がそれほど多くなく、かつ項目間の相関もQ.1の7項目ほど高くないため、IIIはかなり正にひずんだ分布、Vは比較的普通のほぼ左右対称な単峰性の分布となる。

主成分IVは比較的反応率の低い2値変量が高く負荷しているため、得点は下位に集まり正に偏った分布となる。主成分VIは、比較的反応率の高い2値項目が正に負荷し、反応率の低い「その他」が負に負荷すること、相互相関はそれほど高くないことから、(中心極限定理に近い状況で）ほぼ左右対称な分布になる。主成分VIIは、Q.9〜Q.11のゆるやかに相関し、比較的低頻度な「1」(ほとんどない)、「2」(あまりない）が負に負荷するため、かなり負に偏った分布となっている。

理論的に扱いやすい正規分布からは、かなり隔たりの大きい分布が見られること、その理由がおおよそ推測できることは、データと解の関係が直接的な、記述的多変量解析の長所とみることができるであろう。

4.6 主成分間の相関関係

表8に主成分得点間の相関行列を示す。これらの相関係数についても、bootstrap法による 値を

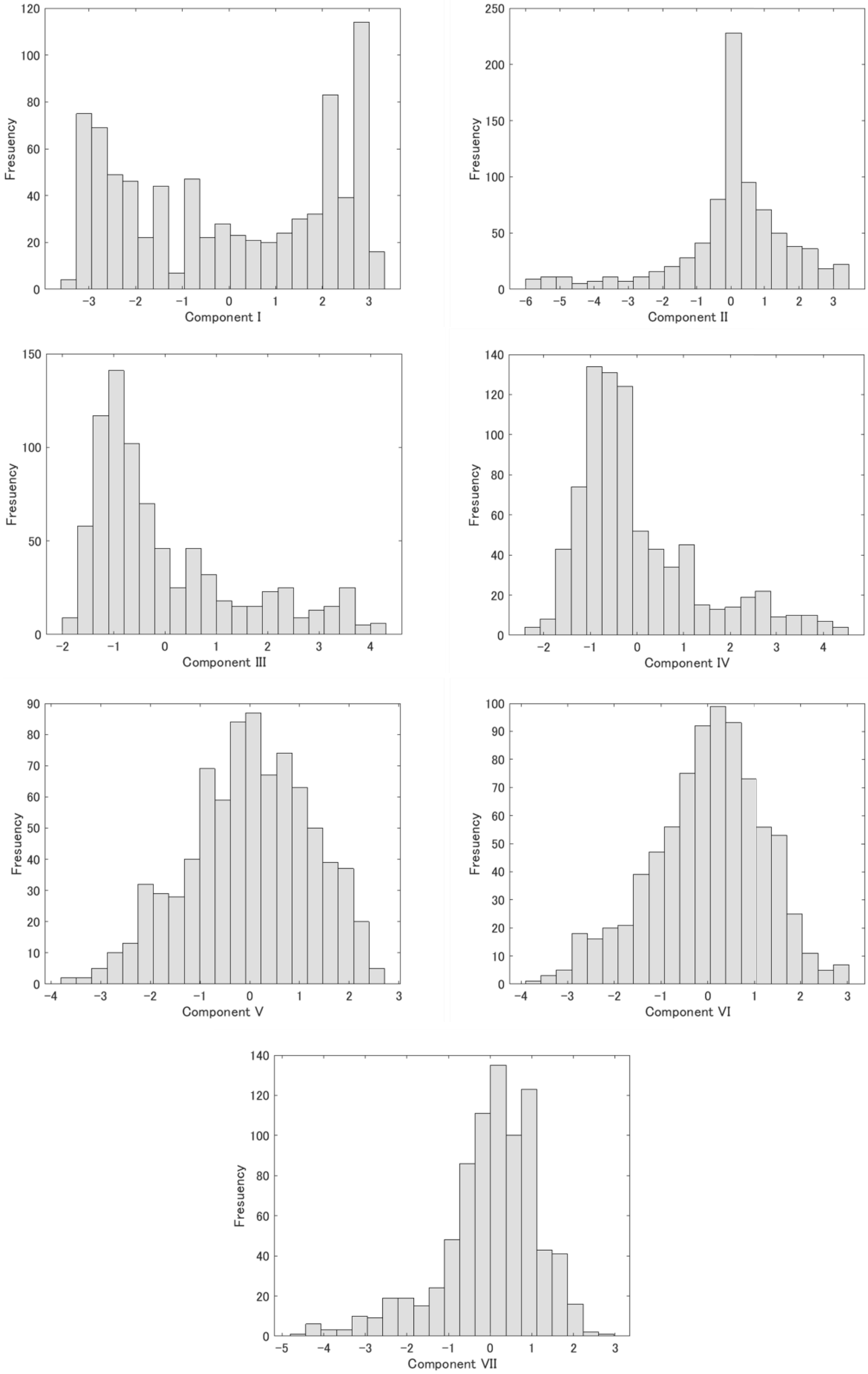

図 9　7 つの主成分の度数分布

表 **8**　主成分間の相関行列

	I	II	III	IV	V	VI
II	0.16					
III	0.27	0.08				
IV	0.11	0.06	0.07			
V	0.10	0.18	0.08	0.05		
VI	0.05	0.04	0.10	0.11	0.02	
VII	-0.04	0.01	0.04	0.02	0.05	0.06

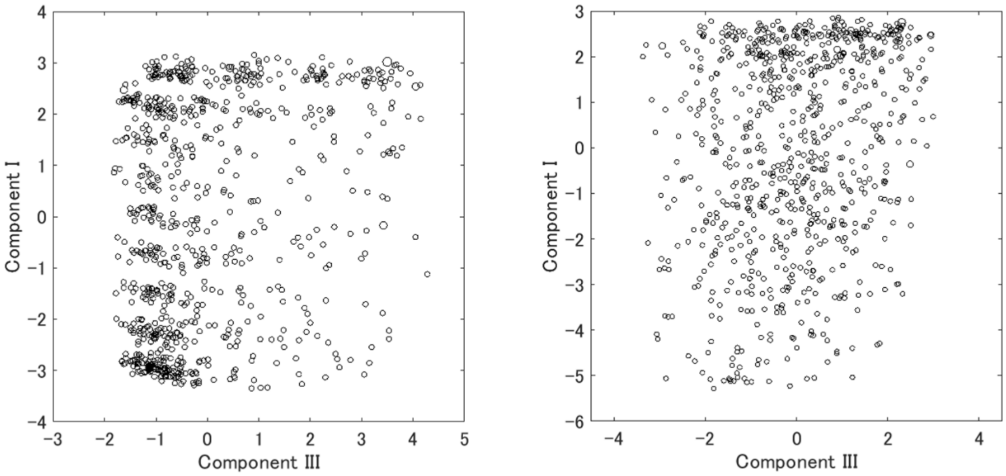

図 10　主成分 III と主成分 I の間の散布図。ダミーコーディング（左）と直交多項式（右）

算出した。21 個の係数のうち、3.6 を上回るものが 4 個、3 を超えるものが 5 個、1.7 より大きいものが 3 個あった。やはり効果量のことも考慮し、 値が 3.7 を超えるもののうち、さらに係数が 0.15 を上回るものを太字で、それ以外の 値が 3 を超えるものを斜体で示した。

これらの相関係数は大部分、通常の評価基準では「相関がない」と判断されるほど小さい値をとっているが、有意性に関してはいわゆる「高度に有意」な水準にあるものも多い。標本の大きさが 800 を超えるこのデータでは、「標本が大きければ、実質的に意味のない相関も有意になる」として、無視されることが多いが、これは (調査) データの他の集計方法との関連で、改めて議論される必要がある。

そうした低い係数の中でも、値として目立つのは主成分 I と主成分 III の間の 0.27 という値である。主成分 III の方を横軸にとって図示したのが、図 9 左である。図 7 におけるそれぞれの度数分布から想像されるように、通常想定される 2 変量正規分布とは著しく異なった散布図であり、相関係数の値が示すよりも高い連関があるとみることができる。すなわち、スポーツへの関心が高ければミュージアムの評価は高く、評価が低い人はスポーツへの関心が低い。ただし、関心は低くても評価は高い人はあって、それが相関係数の値を低下させている。換言すれば、スポーツに関心がなくても、スポーツミュージアムを楽しむことはできる。

ただし、この結論は「5」という評価の数で関心度と評価を定義するダミーコーディングを使用した場合であって、直交多項式を用いるとこの関係は図 10 右に見られるように明瞭でなくなり、相関係数も 0.24 に低下する。

あと 2 つの「高い」相関係数（太字）を生み出しているのが主成分 I と II、II と V であるが、前者は図 8 左のような同一変数の 2 つの変量間の関係であり特に解釈の対象とはならず、後者も「4」を 1 とする 2 つめのダミーコーディング間の関係であり、あまり重視できるものではない。

4.7 デモグラフィックな変数との関係

デモグラフィックな変数のうち、性別、年齢、職業との関係を検討した。これらの要因を同時的に扱う一般線形モデルの利用も考えられたが、図 9 にみられるような得点分布の形状や、年齢と職業の強い連関（30 代以上の学生はいない等）を考慮し、1 要因ずつ検討する方針をとった。これは、geometric data analysis（Le Roux, & Rouanet, 2010）などでもとられている方針である。ただし、参考程度ではあるが、主成分ごとの 1 要因分散分析の p 値と、効果量として相関比 η を併記した。

表 9 は性差であるが、ここでは一般的評価にかかわる主成分 I、II に大きな差がみられない。これは、村上・谷岡・堀（2021）における潜在クラスの比率に関する結論と一致する。III から VI の 4 つの主成分については、多少とも差が認められるが、すべて男性の平均値が上回っている。男性の方が（平均的に）スポーツへの関心が高いことを示していると思われる。

表 10 は年齢差であるが、まず、主成分 I におけるスポーツミュージアムへの一般的評価に関して、年齢段階が高くなるほど評価が低下していく傾向が見られる。これも、村上・谷岡・堀（2021）の結果と一致する。さらに、主成分 III や VI のようなスポーツや展示への関心に関しても、概して若い

表 9　主成分の平均差（性別）

性別	N	I	II	III	IV	V	VI	VII
男性	277	-0.109	0.120	0.338	0.168	0.327	0.330	0.072
女性	533	0.070	-0.070	-0.178	-0.084	-0.171	-0.176	-0.040
p 値		0.274	0.140	0.000	0.010	0.000	0.000	0.200
η		0.032	0.055	0.167	0.089	0.192	0.200	0.045

表 10　主成分の平均差（年齢）

年齢	N	I	II	III	IV	V	VI	VII
10 代	106	0.729	0.172	0.531	0.115	0.127	0.509	-0.037
20 代	136	0.304	0.266	0.651	-0.457	0.070	0.428	-0.355
30 代	47	0.164	-0.009	0.064	-0.329	-0.011	-0.316	0.167
40 代	176	-0.107	-0.148	-0.353	-0.013	-0.076	-0.279	0.056
50 代	256	-0.266	-0.031	-0.298	0.124	0.063	-0.210	0.101
60 代以上	91	-0.388	-0.222	-0.155	0.420	-0.294	0.072	0.119
p 値		0.001	0.217	0.000	0.000	0.154	0.000	0.005
η		0.164	0.095	0.281	0.197	0.100	0.266	0.141

表 11　主成分の平均差（職業）

職　業	N	I	II	III	IV	V	VI	VII
学生	194	0.636	0.173	0.715	-0.119	0.164	0.470	-0.214
公務員・会社員	275	-0.239	-0.041	-0.066	-0.047	0.088	-0.063	0.132
パート・アルバイト	129	-0.090	-0.068	-0.606	-0.014	-0.148	-0.333	0.045
専門職・会社経営・自営	67	-0.572	0.281	0.197	-0.100	-0.005	0.002	0.053
専業主婦・主夫	105	-0.437	-0.410	-0.499	0.244	-0.263	-0.334	0.058
無職等	33	0.925	0.163	-0.245	0.437	-0.280	0.116	-0.271
p 値		0.000	0.073	0.000	0.087	0.021	0.000	0.032
η		0.207	0.114	0.326	0.110	0.126	0.247	0.122

年齢層ほど高得点を示す傾向がある。他方、やはり展示への関心である主成分 IV では、20～30 代と比較して、60 代以上の関心が高い。このあたりは展示内容により踏み込んだ解釈が必要になるが、現時点で筆者にはそれを行うだけの知識が不足していると言わざるを得ない。

表 11 は職業間の差である。これはもともと 9 つのカテゴリーからの選択の形式をとっていたが、「公務員」と「会社員」を合併、さらに「専門職」、「会社経営」、「自営」を併せて、6 カテゴリーに整理している。比較的大きな差は、主成分 I、III、VI にあり、いずれも「学生」の平均値が高いことが目立つ。このうち主成分 I については年齢差、III と VI については性差（「パート・アルバイト」、「専業主婦」が低い）との交絡があり、このデータだけからそれぞれの効果を分離して評価することは難しい。

4.8　データ全体を同時に分析することの意味

調査データ全体の様相を大づかみにとらえることが多重対応分析の応用における 1 つのメリットであったが、そこには 2.3 でも述べたように、次元数の制約という問題があった。本研究は、解釈の対象を軸に移す手法である OPCA を用いて、問題の解決を図った。

調査データの分析における重要な目的の 1 つは、事前に設定された質問項目のまとまり、たとえば、今回の調査で言えば Q.1（一般的な調査）と Q.2（印象に残った展示）の間の関連を明らかにすることがあげられる。そうした目的を達成するために、それぞれの間の個々の項目間でクロス集計をすべて行うことがあるが、それではまとまった（多少とも抽象度の高い）結論を得ることは難しい。

村上・谷岡・堀（2021）では、Q.1 の 7 項目を潜在クラス分析（LCA）によって 5 つのクラスに要約し、Q.2 の個別項目ごとにクラス間の反応率を比較することによって興味ある結果を得た。ここで適用した OPCA において、こうした複数項目群間の関係を検討する手段としては、4.6 で検討した主成分得点間の相関関係がある。Q.1 と Q.2 の関係については、主成分 I と IV の関係がそれに当たる。ただ、この相関係数は（有意ではあるものの）0.11 と値そのものが小さいことと、主成分 IV に高く負荷する変数は、潜在クラスとの関係で有意であったものとそうでないものの区別ができていない。

項目群間の関係を読み取るもう 1 つの手段として、同一の主成分への目立つ負荷量に注目することがあり得る。実際、表 7 の主成分 V、VI、VII においては、そうした負荷量による解釈を行うこと

表 12　Q.2 の諸項目の統計測度

項　　目	主成分 I の 負荷行列の要素	主成分 I の 構造行列の要素	潜在クラス間の 差に関する p 値
1．スポーツ映像ウォール（エントランス）	0.131	0.158	0.000
2．真剣味の殿堂（中京大学のオリンピアン）	0.040	0.110	0.012
3．時代とスポーツのスラローム（夏）	-0.053	0.049	0.495
4．時代とスポーツのスラローム（冬）	0.033	0.113	0.001
5．スポーツアーカイブス	-0.003	0.053	0.280
6．モスクワ五輪のメダル	0.127	0.158	0.000
7．企画展示コーナー（金栗四三）	0.026	0.068	0.449
8．映像ブース	0.123	0.125	0.013
9．その他（できれば具体的に書いてください）	0.093	0.038	0.966

ができた。そこで主成分 I における Q.2 の諸項目の負荷量（表 12 に再掲）のうち、1, 6, 8 の負荷行列の要素は、bootstrap 法によって 5% 水準では有意な結果に到達しているが、これも潜在クラスとの関係を尽くしていない。

負荷行列の要素は、前述のように標準偏回帰係数であるが、表 12 の次の列に、単純相関係数である構造行列の要素と、潜在クラス間の差に関する p 値と並べて示した。構造行列で 0.1 を超える要素と、潜在クラス間の差の p 値が 0.05 を下回る要素とは一致している。すなわち、潜在クラス分析を用いて行った結果得られた情報は、OPCA の中に再現してはいたのである。

なお、通常は問題にされない 0.1 のオーダーの相関係数を用いた議論ついては、統計的に有意ではあっても効果量としては無視し得るという疑義は当然生じるであろう。しかし、この程度の関係であっても、潜在クラス間では、反応率に 10%を超える差異があるのである（村上・谷岡・堀、2021)。このあたりは、個人の assessment を目的として発達してきた計量心理学と集団の様相の記述が目的となる社会調査の違いとして論じられるべき問題である（村上、2020)。

5．おわりに

本研究は、村上・谷岡・堀（2021）において分析された、中京大学スポーツミュージアムの一般公開直後に実施された、来館者調査のデータを再分析した。用いられた方法は、村上（2019b)、Murakami（2020）による正規直交主成分分析（OPCA）であった。この方法は、カテゴリカルな変数について、カテゴリー数より 1 だけ少ない数の正規直交な変量を定義し、それらの変量を重みの回転を伴う主成分分析にかけるというものである。これは同じデータを対象とする多重対応分析（MCA）と（直交回転を除いて）同一の個体スコアを生み出し、MCA の許容される変換を施して、解釈の対象を平面上の視覚的表現から、相対的に独立な軸への負荷量に変え、3 次元以上のデータの変動を表現し解釈することができるようにするものであった。

本研究を通じて、デモグラフィックな変数を除く全変数の同時分析によって、一定のまとまりをもつ質問項目群ごとに行う MCA よりも 1 つの調査全体の構造が明らかいなることが期待された。この

期待が満たされたかどうかについては、4.8 で述べたように、複数の質問群が共通に負荷するような主成分が出現するかどうかで決まる。今回の場合、主成分 III、VI、VII にはそうした性質が認められたが、主成分 VI と VII は再現性を保証する evidence がやや弱いうらみがあった。

他方で、図 3、図 5 と図 8 を見比べてみれば、少数の意味的まとまりのある項目群の分析から、より明確な構造と次元が得られるのも事実であり、複数の項目群間の関係は、群ごとの分析から得た次元の間の相関関係の分析によって明らかにする方が、手続き的に見通しがよいという考え方もあり得る。これについては、実データにもとづく研究の積み重ねによってより有効な手順を選ぶことができるようになると考えられる。

ただし、比較的少数の内容的に（ある程度）一貫した質問項目の分析においても、OPCA は回答者の反応スタイル等の解明に役立つことも示されている（村上、2019a)。

もう一つ、OPCA には、MCA にあらわれる余剰次元を、解釈可能の次元と分離して、結果の解釈をより純正なものにすることが期待された。村上（2020）は、順序のあるカテゴリカルな反応に、直交多項式を適用して 2 次以上の項を解釈において無視することにより、そのことを実現しえたと考えていたのであるが、本研究でダミーコーディングを用いて得られる複数の次元は、「最も強い肯定を意味する反応」対「その他の反応」、「次に強い肯定的反応」対「強い 2 つの反応以外の反応」といった形で、1 つの概念に関する異なる「強度」の次元とみることもできた。しかも、それらは同じ MCA による次元の変換にすぎないのである。この点も、今後いろいろなデータにおいて確認しながら、意味づけを考えていかなければならない。

注

1）本研究は、学術振興会科学研究費　基盤（c）20K03303 の支援を受けた。

2）本研究のための調査の実施と論文執筆にあたり、多大なご援助をいただいた、中京大学現代社会学部国際文化専攻の教員・学生の皆さん、ならびに、同大学スポーツ振興室のスタッフの方々に、心からお礼申し上げます。また、村上・谷岡・堀（2021）における共同研究者、谷岡謙、堀兼大朗のお二人にも、深く感謝いたします。

3）因子分析は、現在は構造方程式モデルの中の一手法として分類される。主成分分析が変数の 1 次結合としての合成変量にもとづく分析であるのに対し、因子分析は直接観測できない潜在変量に依拠した分析である。しかし、変数間の相関関係にもとづいて、少数の変量に要約しようとする目的については、同一と考えられる。

4）1 次独立ではあるが、必ずしも無相関ではないという意味でこの表現を用いた。

5）回転により、欠損値による余剰次元の分離の可能性についても、現在検討中であるが、それについては、機会を改めて論じることにしたい。

6）このやや特異な負荷のパターンを示す主成分の出現の理由は、調査期間の後半において、アイドル的な人気があった男子フィギュアスケーターの衣装が、いわゆるツーショットの自撮りが可能な形で展示された影響が大きいと思われる（村上・谷岡・堀, 2021)。Q.2 の「その他」の大きな負の負荷量は、自由記述にそれを記した人たちによって生じており、Q.3 でやや目立つ「その他」は、SNS による拡散が含まれる。この主成分の外部変数との関係（表 9、11）にも注意。

文献

Bekker, P., & 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 (pp. 1-31). Chichester: Wiley.

Benzécri, J.P. (1992). Correspondence analysis handbook. New York, NY: Dekker.

Efron, B. & Tibshirani, R. J. (1993). An introduction to the Bootstrap. New York: Chapman & Hall.

Greenacre, M.J. (1984). Theory and applications of correspondence analysis. London: Academic Press.

Greenacre, M.J. (2017). Correspondence analysis in practice. 3rd ed. New York: CRC Press.

Harris, C. H., & Kaiser, H. F. (1964). Oblique factor analytic solutions by orthogonal transformations. Psychometrika, 29, 347-362.

Jolliffe, I.T. (2002). Principal component analysis 2nd ed. New York: Springer.

Kiers, H.A.L. & Ten Berge, J.M.F. (1994). The Harris-Kaiser independent cluster rotation to simple component weights. Psychometrika, 59, 81-90.

Le Roux, B. & Rouanet, H. (2010). Geometric data analysis: From correspondence analysis to structured data analysis. Dordrecht: Kluwer Academic Publishers.

村上 隆（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　データ分析の理論と応用，5(1)，27-47.

村上 隆（2019a）．正規直交多項式主成分による Likert 型項目への反応スタイルの研究　日本行動計量学会第 47 回大会抄録集　144-147.

村上 隆（2019b）．多重対応分析の因子分析的使用ー正規直交主成分分析とそのプロ野球の観客調査データへの適用ー　中京大学社会学部紀要，12(2)，95-162.

Murakami, T. (2020). Orthonormal principal component analysis for categorical data as a transformation of multiple correspondence analysis. In T. Imaizumi, A. Nakayama, S. Yokoyama (Eds.) Advanced studies in behavior metrics and data science : Essays in honor of Akinori Okada. Singapore: Springer Nature, 211-231.

村上 隆・谷岡 謙・堀 兼大朗（2021）．大学博物館の来館者による評価　岡部真由美・亀井哲也・斉藤尚文・栂正行（編）大学教育と博物館　中京大学文化科学研究所叢書 22（印刷中）

大隅昇・ルバール，L.・モリノウ，A.・ワーウィック，K.M.・馬場康維（1994）．記述的多　変量解析法　日科技連

芝祐順（1979）．因子分析法　第 2 版　東京大学出版会

Timmerman, M., Kiers, H.A.L. & Smilde, A.K. (2007). Estimating confidence intervals for principal component loadings: A comparison between the bootstrap and asymptotic results. British Journal of Mathematical and Statistical Psychology, 60, 295-314

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. Psychometrika, 41, 321-327.

付録

Sports museum **来館者アンケート**

Museum へのご来館、ありがとうございます。以下のアンケートにお答えいただければ幸いです。

結果は今後の運営の参考にさせていただきます。なお、統計的分析にもとづいて論文等を刊行する可能性のあることもご了承ください。

Q.1　Museum の次の点について評価してください

【全部の項目について、当てはまる数字に〇をつけてください。】

	悪い	**やや悪い**	**普通**	**やや良い**	**良い**
1．入館した時の印象	1 0.0%	2 0.0%	3 15.0%	4 23.3%	5 61.6%
2．展示物の価値・希少性	1 0.1%	2 0.3%	3 8.0%	4 27.4%	5 64.2%
3．展示のレイアウト	1 0.1%	2 1.6%	3 14.7%	4 29.6%	5 54.1%
4．説明のわかりやすさ	1 0.2%	2 2.6%	3 22.2%	4 30.7%	5 44.2%
5．提供されている情報の量	1 0.3%	2 3.1%	3 22.7%	4 31.0%	5 42.9%
6．館内の雰囲気・居心地のよさ*	1 0.1%	2 1.4%	3 14.2%	4 28.4%	5 56.0%
7．Museum へのアクセス・館内での動きやすさ*	1 0.4%	2 7.2%	3 30.1%	4 23.6%	5 38.7%

*を付した項目には、複数回答が 1 ずつある。集計には含めていない。

Q.2　次のコーナーのから印象に残ったものの番号に〇をつけてください【〇はいくつでも】

1．スポーツ映像ウォール（エントランス） 27.0%	2．真剣味の殿堂（中京大学のオリンピアン） 34.9%
3．時代とスポーツのスラローム（夏） 18.8%	4．時代とスポーツのスラローム（冬） 18.4%
5．スポーツアーカイブス 11.8%	6．モスクワ五輪のメダル 29.5%
7．企画展示コーナー（金栗四三） 22.0%	8．映像ブース 23.7%
9．その他（できれば具体的に書いてください） 16.1%	

Q.3　この Museum のことを何によってお知りになりましたか【○は１つだけ】

1．新聞記事・テレビ番組などメディア 30.3%	2．大学のホームページ 6.2%
3．キャンパスに来てみて 38.6%	4．友人・知人らに聞いて（クチコミ） 14.7%
5．その他（できれば具体的に書いてください） 11.3% 複数回答 41（上記集計外）	

Q.4　性別：

1．男　33.4%	2．女　66.4%	3．その他　0.2%	複数回答　1（集計外）

Q.5　年齢：

10 歳未満 0.0%	10 代 13.1%	20 代 16.3%	30 代 5.6%	40 代 21.4%	50 代 31.3%	60 代 7.9%	70 代以上 4.3%

Q.6　現在お住まいの場所は次のどれですか

1．豊田市 20.6%	2．名古屋市 18.3%	3．その他の愛知県内 36.8%	4．愛知県外 24.3%

Q.7　職業：【○は１つだけ】

1．学生・生徒 23.3%	2．公務員 5.0%	3．会社員 28.5%	4．パート・アルバイト 16.1%
5．専門職（教員・勤務医など） 4.0%	6．会社経営・会社役員 2.0%	7．自営業（農業含む） 2.5%	
8．専業主婦・主夫 12.8%	9．無職・家事手伝い 4.1%	10．その他（　　　　　　　　） 1.6%	

Q.8　中京大学とのご関係を教えてください【○はいくつでも】

1．学生（豊田） 18.0%	2．学生（名古屋） 1.9%	3．生徒（中京大中京高校） 0.5%	4．卒業生（大学・高校） 15.4%
5．保護者 25.3%	6．教職員（旧職員含む） 1.8%	7．その他（　　　　　　　　） 10.2%	8．特に関係はない 26.0%

Q.9　あなたが自分でスポーツを**した経験**について教えてください【○は１つだけ】

1．ほとんどない 10.1%	2．あまりない 22.2%	3．どちらとも言えない 18.3%	4．かなりある 27.6%	5．大いにある 21.8%

Q.10　あなたがスポーツを**観戦した経験**について教えてください*【○は１つだけ】

1．ほとんどない 5.3%	2．あまりない 17.1%	3．どちらとも言えない 21.9%	4．かなりある 37.0%	5．大いにある 18.5%

Q.11　来年開催される東京オリンピックにどのくらい関心をおもちですか【○は１つだけ】

1．ほとんどない 2.7%	2．あまりない 6.2%	3．どちらとも言えない 23.5%	4．かなりある 47.0%	5．大いにある 20.7%

Q.12　全体としてこの Museum のご観覧に満足していただけましたか【○は１つだけ】

1．不満 0.2%	2．やや不満 1.8%	3．どちらとも言えない 11.0%	4．ほぼ満足 49.7%	5．満足 37.3%

入館日時

24 日午前	24 日午後	25 日午前	25 日午後	26 日午前	26 日午後	27 日午前	27 日午後
6.9%	5.4%	0.4%	4.8%	38.0%	30.5%	5.0%	9.1%
67	52	4	46	367	295	48	88

回収数　967

最後にご意見をお書きください。この Museum に限らず、オリンピックのことスポーツ全般のことなど何でも自由に書いてください。

ご協力、ありがとうございました。