共通数量化の応用例

An Application of Common Quantification

岡　本　安　晴
Yasuharu OKAMOTO
（日本女子大学大学院人間社会研究科　心理学専攻）

要　約
共通数量化の応用例を報告する。共通数量化は，カテゴリ変数の数量化を変数ごとに行うものであり，数量化が変数ごとに行われるので数量化の解釈が容易であるとともに，得られた数量化変数は続いて行われる種々の分析において共通に用いられるので分析結果が解り易く，解釈に統一性が期待できるという特徴がある。共通数量化の有用性を示すための簡単な実データは，心理学に関する入学前の知識とか卒業後の進路希望などに関する簡単な質問（カテゴリ選択肢で答える形式のもの）からなる質問紙を心理学科の学生から一様に１年生１８人（１年生全体の約１／４）を選んで実施して得られたものである。共通数量化変数を構成して主成分分析を行うことにより，クロス表の情報が読み取り易くなることが確認された。最後に，共通数量化を行ためのソフトウェアなどの情報も記しておいた。

[Abstract]
In this study, an application of common quantification was reported, and its usefulness was demonstrated. Because common quantification quantifies categorical variables individually, interpreting the obtained quantification is easy. The quantification is performed prior to the analyses of the data, thus, the quantified variables are used commonly in the following analyses, and the results of various analyses can be interpreted based on the common quantification.

This study used data of 18 freshmen sampled from the department of psychology of a women's university. These freshmen completed a questionnaire with categorical items on their knowledge of psychology before entering the university as well as on the future course that they desire/desired. Common quantification was applied to the data, and the quantified variables were analyzed using principal components analysis, which resulted in better understanding of the data.

At the end of this study, internet information by the author concerning common quantification was provided.

はじめに

調査あるいは質問紙における質問項目は、カテゴリの選択で答える形式が多い。各カテゴリに適当な数値を割り当てると数値データとして分析できる。この数値の割り当てを、まず、カテゴリ変数ごとに行い（第１ステップ)、その後、割り当てられた数値を共通数量化データとして種々の分析を行う（第２ステップ）分析法が２段共通数量化分析法として提案されているが（Okamoto, 2015；岡本、2014，2015)、現実のデータに対する具体的な分析例は報告されていない。したがっ

て、共通数量化の現実のデータに対する応用例を報告することは、共通数量化の有用性の評価とその普及において重要であると考えられる。

カテゴリ変数の数量化あるいは質的データの分析法については、既に多くの研究があり、それらは例えば Nishisato（2007）に整理して紹介されている。これらの分析法は、カテゴリ変数の各カテゴリに数値を割り当てるものであるが、林の数量化法（林、1970；岩坪、1987）、双対尺度法（西里、1982；Nishisato、2007）、対応分析（Greenacre、2007）、非線型多変量解析（Gifi、1990）などと種々の名前で呼ばれている。これらの分析法における数量化（各カテゴリへの数値の割り当て）は分析ごとに目的関数を設定して行われているため、同じ変数であっても、分析ごとにカテゴリの数量化が行われることになり、分析間の比較が難しくなる。また、複数の変数のカテゴリがまとめて数量化されるため、数量化の意味付けが分かり難くなる。これに対して、共通数量化では、第１ステップにおいて数量化された値が、第２ステップにおける複数の分析において共通に用いられるので、分析間の比較が容易である。また、変数ごとにカテゴリの数量化が行われるので、意味付けも分かり易くなる。本研究では、共通数量化の解釈の容易さに焦点を当てて、簡単なデータを収集し分析を行ってみた。まず、次節において共通数量化について簡単に説明する。なお、共通数量化は、村上（1999）のカテゴリカル・データの分析法を簡単なものにしたいという動機から研究が進められたものであることを記しておく（岡本、2013）。

共通数量化法

共通数量化について簡単に説明する。詳しくは、Okamoto (2015)を参照されたい。

カテゴリ変数XがC個のカテゴリ値をとるものとする。カテゴリ値を１からCまでの自然数で表す。Xの i 番目（ i 人目）のデータをx_iで表す。Xを行列（縦ベクトル）として表せば

$$\mathrm{X} = (x_1 \quad \cdots \quad x_N)'$$

となる。

いま、カテゴリ変数Xの数量化変数をQとおき、カテゴリ j に数値v_jを与えるものとする。Xの i 番目の値x_iに対するQの値をq_iとおけば、

$$x_i = j \text{ のとき } q_i = v_j \qquad (1)$$

である。

カテゴリ変数Xの i 番目の値x_iに対してベクトル$\boldsymbol{g_i} = (g_{i1} \quad \cdots \quad g_{iC})$を

$$g_{ik} = \begin{cases} 1 & x_i = k \text{ のとき} \\ 0 & x_i \neq k \text{ のとき} \end{cases}$$

とおけば、式（１）は

$$q_i = v_j = g_{i1}v_1 + \cdots + g_{iC}v_C = (g_{i1} \quad \cdots \quad g_{iC})\begin{pmatrix} v_1 \\ \vdots \\ v_C \end{pmatrix}$$

と書ける。したがって、データ数が N 人分であるとき、

$$G = (g_{ij})、\quad \boldsymbol{\omega} = (v_1 \quad \cdots \quad v_C)'、\quad Q = \begin{pmatrix} q_1 \\ \vdots \\ q_N \end{pmatrix}$$

とおけば

$$Q = G\boldsymbol{\omega}$$

と書ける。Gはダミー変数のように機能しており、indicator 行列と呼ばれている。$\boldsymbol{\omega}$は数量化ベクトルである。

いま、数量化変数Qの分散が大きい程、カテゴリ変数Xに含まれる情報が数量化変数によく反映されていると考え、数量化ベクトル $\boldsymbol{\omega}$ を、$\boldsymbol{\omega}$ の長さが１の制約条件の下で数量化変数Qの分散を最大（極大）にするものとして求める。この $\boldsymbol{\omega}$ は、次式を満たすものとして求められる(Okamoto, 2015).

$$H'H\boldsymbol{\omega} = \lambda\boldsymbol{\omega} \qquad (2)$$

ここで

$$H = G - \frac{1}{N}\mathbf{1}_N\boldsymbol{f}、\quad \boldsymbol{f} = \mathbf{1}'_N G$$

であり、$\mathbf{1}_N$は１を要素とする$(N, 1)$型行列

$$\mathbf{1}_N = (1 \quad \cdots \quad 1)'$$

である。

式（２）を満たすλと $\boldsymbol{\omega}$ は、固有値とそれに対する固有ベクトルと呼ばれているが（岡本、2009）、式（２）の場合は、一般に０でない固有値λを$(C-1)$個もつ（Okamoto, 2015）。式（２）に対して Q の分散は

$$Var(Q) = \lambda/N$$

で与えられる。数量化変数 Q の平均を０、分散を１に標準化した値を共通数量化変数 $\boldsymbol{z}$ とする。

$$\boldsymbol{z} = (\lambda/N)^{-1/2}H\boldsymbol{\omega} \qquad (3)$$

である。

カテゴリ数Cのカテゴリ変数Xに対して、一般には$(C-1)$個の数量化変数（式（３））が与えられるが、それらはお互いに独立である（Okamoto、2015）。

応用例

データの収集

共通数量化法による実データの分析例を示すために、データ収集のための図１に示す簡単な質問紙を用意した。この質問紙をＮ大学心理学科の１年生１８人に配布し、回答に協力を求めて回

収した。１８人全員が全ての項目に回答していた。１８人は、１年生８２人から名簿順を基に４人目ごとに選んでグループ分けされた１グループに属する者で、一様にサンプリングされたとみなし得るものである。

質問紙調査

以下の問いについて、それぞれ最もよく当てはまるものを選んで（　）内に○を付けて下さい。

問１　心理学という分野があることを
（　　）　小学生の頃には知っていた。
（　　）　中学生の頃に知った。
（　　）　高校生の頃に知った。
（　　）　その他。

問２　大学受験のとき
（　　）　心理学≒臨床心理学と思っていた。
（　　）　心理学には臨床心理学以外の分野があることを知っていた。
（　　）　その他（「心理学とは何をする分野なのか知らなかった。」など）。

問３　大学受験のとき、卒業後は何になりたいと考えていましたか？
（　　）　カウンセラーになりたい。
（　　）　カウンセラーではないが、心理学を活かした職業に就きたい。
（　　）　その他。

問４　いま、卒業後は何になりたいと考えていますか？
（　　）　カウンセラーになりたい。
（　　）　カウンセラーではないが、心理学を活かした職業に就きたい。
（　　）　その他。

問５　自分の得意分野は
（　　）　文系であると思う。
（　　）　理系であると思う。
（　　）　その他。

図１　質問紙

分析１（共通数量化）

質問紙の各問いに対する回答の度数分布と共通数量化ベクトルおよび共通数量化変数を表１に示す。問１のカテゴリ４「その他」および問５のカテゴリ３「その他」に回答した者はいなかった。カテゴリの数量化においては、これら度数０のカテゴリを除いて共通数量化変数を算出した。数量化ベクトルは、有効回答カテゴリ数が３個である問１から問４までは２個（$\boldsymbol{\omega_1}$と$\boldsymbol{\omega_2}$）あり、有効カテゴリ数が２個である問５については数量化ベクトルは１個（$\boldsymbol{\omega_1}$）である。なお、カテゴリ数が２個の場合は、数量化変数の平均値が０、分散が１という条件で数量化の値が決まる。また、数量化ベクトル $\boldsymbol{\omega}$ もデータの分布によらず符号を除いて一意に決まる。このことは、附記で説明した。

各数量化ベクトル$\boldsymbol{\omega_j}$に対応する固有値λ_jは、ベクトルの右側の（　）内に示した。数量化ベクトルに対応する共通数量化（式（３））を表す変数名はベクトルの左側に示した。問 j の k 番目の数量化変数を$\mathrm{Qj}-\mathrm{h}$（ただし、$h=k-1$）で表している。h が k より１小さく０から始まっているのは、使用したプログラムでは順序を表す整数値が０から始まるためである。

表１　質問紙回答の度数分布と共通数量化

数量化ベクトル$\boldsymbol{\omega_j}$の右側の丸カッコ内の数値は対応する固有値λ_jである。共通数量化変数名$\mathrm{Qj}-\mathrm{h}$を対応する数量化ベクトル$\boldsymbol{\omega_j}$の左側に示した。通し番号 h が０から始まり、対応する数量化ベクトルの添え字 j より１小さいのは、使用したプログラムの出力における順序が０から始まる整数値になっていることに合わせたものである。

問１．心理学を知った時期	カテゴリ	１．小学生	２．中学生	３．高校生	４．その他
	度数	3	１２	3	0
Q1 − 0	$\boldsymbol{\omega_1}$(6.00)	−0.408	0.816	−0.408	−
Q1 − 1	$\boldsymbol{\omega_2}$(3.00)	−0.707	0.000	0.707	−
問２．心理学とは（受験時）	カテゴリ	１．　心理学 ≒ 臨床心理学	２．　心理学 ＞ 臨床心理学	３．　その他（知らない等）	
	度数	2	１４	2	
Q2 − 0	$\boldsymbol{\omega_1}$(4.67)	−0.408	0.816	−0.408	
Q2 − 1	$\boldsymbol{\omega_2}$(2.00)	0.707	0.000	−0.707	
問３．卒業後の進路希望（受験時）	カテゴリ	１．カウンセラー	２．心理学を活かした職業	３．その他	
	度数	4	6	8	
Q3 − **0**	$\boldsymbol{\omega_1}$(6.95)	−0.136	−0.630	0.765	
Q3 − **1**	$\boldsymbol{\omega_2}$(4.60)	0.805	−0.520	−0.285	
問４．卒業後の進路希望（今、考える）	カテゴリ	１．カウンセラー	２．心理学を活かした職業	３．その他	
	度数	3	7	8	
Q4 − **0**	$\boldsymbol{\omega_1}$(7.48)	−0.032	−0.691	0.722	
Q4 − **1**	$\boldsymbol{\omega_2}$(3.74)	0.816	−0.435	−0.381	
問５．自分は文系／理系	カテゴリ	１．文系	２．理系	３．その他	
	度数	１６	2	0	
Q5 − **0**	$\boldsymbol{\omega_1}$(3.56)	0.707	−0.707	−	

分析２（主成分分析）

カテゴリデータに含まれる情報を取り出すために、共通数量化変数$Q1-0$から$Q5-0$に対して主成分分析を行った。特異値は、以下のようであった。

$\lambda_1 = 1.69241$、$\lambda_2 = 1.35123$、$\lambda_3 = 1.17223$、$\lambda_4 = 0.98757$、$\lambda_5 = 0.78868$、$\lambda_6 = 0.73413$、$\lambda_7 = 0.59491$、$\lambda_8 = 0.51055$、$\lambda_9 = 0.43006$

ここで、特異値の２乗 λ_p^2 は、主成分の空間へのデータの正射影の分散を表す(Okamoto, 2006)。説明される分散が１以上という基準で主成分を３つ選び、対応する特異値の大きさの順に$\mathrm{Comp.1}$、$\mathrm{Comp.2}$、$\mathrm{Comp.3}$と呼ぶことにする。主成分と数量化変数との相関係数を構造行列の形式で表したものを表２に、主成分との相関係数を座標値として共通数量化変数を図示したものを図２、図３に示す。

表２　構造行列（主成分と数量化変数との相関係数）

	Comp.1	Comp.2	Comp.3
Q1-0	-0.76589	-0.25081	0.24787
Q1-1	0.26294	-0.27262	0.78077
Q2-0	-0.78028	0.24055	0.07405
Q2-1	0.07392	0.14331	-0.59162
Q3-0	0.88571	0.03780	0.01603
Q3-1	-0.08038	0.76336	-0.00627
Q4-0	0.86111	0.23767	0.01277
Q4-1	-0.22851	0.78145	-0.01248
Q5-0	-0.09763	-0.59909	-0.58905

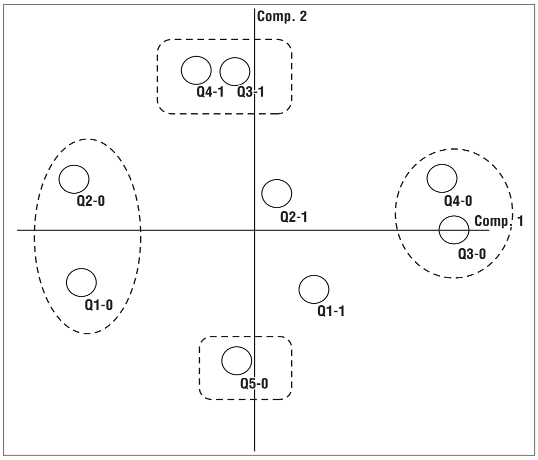

図２　構造行列（表２）の図示
第１主成分$\mathrm{Comp.1}$と第２主成分$\mathrm{Comp.2}$に
対する各共通数量化変数$\mathrm{Qj-h}$の関係を表す

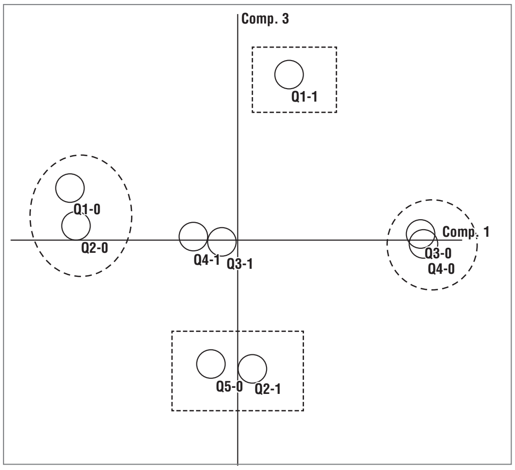

図３　構造行列（表２）の図示
第１主成分Comp. 1と第３主成分Comp. 3に
対する各共通数量化変数Qj − hの関係を表す。

図２から、共通数量化変数Q3 − 0とQ4 − 0のまとまりとQ1 − 0とQ2 − 0のまとまりが主成分Comp. 1上に見られる。数量化変数Q3 − 0は、問３「大学受験のとき、卒業後は何になりたいと考えていましたか？」に対するカテゴリの数量化を

「３．その他」(0.765)
「１．カウンセラー」(-0.136)
「２．心理学を活かした職業」(-0.630)

と行うものであり、数量化変数Q4 − 0は、問４「いま、卒業後は何になりたいと考えていますか？」に対するカテゴリの数量化を

「３．その他」(0.722)
「１．カウンセラー」(-0.032)
「２．心理学を活かした職業」(-0.691)

と行うものである（表１）。それぞれの数量化が、カテゴリ３＞カテゴリ１＞カテゴリ２の大きさの順になっていてよく対応している。問３と問４のカテゴリのクロス表（表３）における度数の分布も対角線上に集まっている。これらは、卒業後の進路について受験時と今もほぼ同じであることを示すものである。

表３　問３と問４のクロス表

		問４　いま、卒業後は？			
		１．カウンセラー	２．心理学を活かす	３．その他	計
問３ 受験のとき、 卒業後は？	１．カウンセラー	2	1	1	4
	２．心理学を活かす	1	5	0	6
	３．その他	0	1	7	8
	計	3	7	8	１８

問１と問２の関係については、以下の関係が読み取れる。数量化変量Q1－0は、問１「心理学という分野があることを知った時期」に対するカテゴリの数量化を

「２．中学生」(0.816)

「１．小学生」(-0.408)

「３．高校生」(-0.408)

と行うものであり、数量化変量Q2－0は、問２「受験のとき、心理学とはどういう学問であると考えていたか」を問うものに対してのカテゴリの数量化を

「２．心理学には臨床心理学以外の分野があることを知っていた」(0.816)

「１．心理学≒臨床心理学と思っていた」(-0.408)

「３．その他」(-0.408)

と行うものである（表１）。問１および問２ともカテゴリ１と３に対する重みが同じ値-0.408 であるので、数量化Q1－0およびQ2－0は、カテゴリ１および３とカテゴリ２を区別するものと考えられる。このことは、問１と問２のクロス表（表４）における度数の分布が、「問１（心理学を知った時期）：カテゴリ２（中学生）」で「問２（進路）：カテゴリ２（心理学には臨床心理学以外もある）」である者に集中していることに対応している。

表４　問１と問２のクロス表

		問２ 大学受験のとき、心理学とは			
		１．心理学≒臨床心理学	２．心理学には臨床心理学以外もある	３．その他	計
問１ 心理学を 知ったのは？	１．小学生の頃には知っていた	1	2	0	3
	２．中学生の頃に知った	0	１１	1	１２
	３．高校生の頃に知った	1	1	1	3
	計	2	１４	2	１８

数量化Q1－0とQ2－0のグループと、数量化変量Q3－0とQ4－0のグループが第１主成分 Comp.1 の軸上で正負反対側に位置していることは、問３の数量化

「３．その他」(0.765)

「１．カウンセラー」(-0.136)

「２．心理学を活かした職業」(-0.630)

および、問４の数量化

「３．その他」(0.722)

「１．カウンセラー」(-0.032)

「２．心理学を活かした職業」(-0.691)

の順序と、問１の数量化

「２．中学生」(0.816)

「１．小学生」(-0.408)

「３．高校生」(-0.408)

および問２の数量化

「２．心理学には臨床心理学以外の分野があることを知っていた」(0.816)

「１．心理学≒臨床心理学と思っていた」(-0.408)

「３．その他」(-0.408)

の順序が逆であることを意味する。これは、問１と問３とのクロス表（受験生の動向を念頭に置いた分析を行っているので、問３とほぼ重なった分布である問４とのクロス表はここでは取り上げない）における次の傾向として現れている（表５）。すなわち、問１においてカテゴリ２（中学生のとき）を選んだものが多いが、このカテゴリを選んだものは問３においてカテゴリ２（心理学を活かした職業）を選んだものが多い。問１においてカテゴリ１（小学生）あるいはカテゴリ３（高校生）を選んだ者は問３においてカテゴリ３（その他）を選んだ者が多い。いずれにせよ、カウンセラーを選んだ者は少ない。

表５　問１と問３のクロス表

		問３　受験のとき、卒業後は？			
		１．カウンセラー	２．心理学を活かす	３．その他	計
問１ 心理学を 知ったのは？	１．小学生の頃には知っていた	1	0	2	3
	２．中学生の頃に知った	3	6	3	１２
	３．高校生の頃に知った	0	0	3	3
	計	4	6	8	１８

問２と問３とのクロス表（表６）では、問２においてカテゴリ２（心理学には臨床心理学以外の分野もある）を選んだ者が多いが、彼女らの多くは問３においてカテゴリ２（心理学を活かした職業）を選んでいる。問２においてカテゴリ１（心理学≒臨床心理学）あるいはカテゴリ３（その他）を選んだ者は、問３においてカテゴリ３（その他）を選んでいる。

表６　問２と問３のクロス表

		問３　受験のとき、卒業後は？			
		１．カウンセラー	２．心理学を活かす	３．その他	計
問２ 大学受験のとき、心理学とは	１．心理学≒臨床心理学	0	0	2	2
	２．心理学には臨床心理学以外もある	4	6	4	１４
	３．その他	0	0	2	2
	計	4	6	8	１８

以上より、第１主成分の負の方向は、卒業後の進路選択においては「心理学を活かした職業を希望している」、心理学については「心理学には臨床心理学以外の分野があることを知っている」、心理学という分野があることは「中学生の頃」という傾向を表していると解釈される。

次に、第２主成分について見てみる（図２）。問３と問４の共通数量化変数Q3－1とQ4－1のグループと問５の共通数量化変数Q5－0に分かれている。共通数量化変数Q3－1は、問３「大学受験のとき、卒業後は何になりたいと考えていましたか？」に対するカテゴリの数量化を

「１．カウンセラー」(0.805)

「３．その他」(-0.285)

「２．心理学を活かした職業」(-0.520)

と行うものであり、数量化変量Q4－1は、問４「いま、卒業後は何になりたいと考えていますか？」に対するカテゴリの数量化を

「１．カウンセラー」(0.816)

「３．その他」(-0.381)

「２．心理学を活かした職業」(-0.435)

と行うものである（表１）。それぞれの数量化が、カテゴリ１＞カテゴリ３＞カテゴリ２の大きさの順になっていてよく対応している。卒業後の進路について受験時と今もほぼ同じであることを示している。

数量化変量Q5－0は、問５「自分の得意分野？」に対するカテゴリの数量化を

「１．文系」(0.707)

「２．理系」(-0.707)

と行うものである（表１）。

表７　問３と問５のクロス表

		問５　自分の得意分野は？		
		１．文系	２．理系	計
問３ 受験のとき、 卒業後は？	１．カウンセラー	２	１	３
	２．心理学を活かす	７	０	７
	３．その他	７	１	８
	計	１６	２	１８

問３と問５のクロス表（表７）を見ると、理系の２人がカテゴリ２「心理学を活かした職業」を選んでいないことが第２主成分において互いに反対側に位置していることに表されていると解釈できる。問４と問５との関係も同じである。第２主成分は、理系の２人が卒業後の進路において心理学を活かした職業以外を希望していることを表すものであると解釈される。

第３主成分では（図３）、問１の共通数量化変数Q1－1に対して問２と問５の共通数量化変数Q2－1とQ5－0のグループが反対側にある。共通数量化変数Q1－1は、問１「心理学という分野があることを知った時期」に対するカテゴリの数量化を

「３．高校生」(0.707)

「２．中学生」(0.000)

「１．小学生」(-0.707)

と行うものであり、共通数量化変数Q2－1は、問２「受験のとき、心理学とはどういう学問であると考えていたか」を問うものに対してのカテゴリの数量化を

「１．心理学≒臨床心理学と思っていた」(0.707)

「２．心理学には臨床心理学以外の分野があることを知っていた」(0.000)

「３．その他」(-0.707)

と行うものである（表１）。第３主成分においてQ1－1とQ2－1が反対側にあることは、問１と問２のクロス表（表４）を見れば、問１においてカテゴリ１（小学生）と答えた者は、問２においてカテゴリ１（心理学≒臨床心理学）あるいはカテゴリ２（心理学＞臨床心理学）と答えており、問１においてカテゴリ２（中学生）と答えた者は問２においてカテゴリ２（心理学＞臨床心理学）あるいはカテゴリ３（その他）と答えていることに対応していると解釈される。共通数量化変数Q1－1とQ5－0が第３主成分において反対側にあることは、問１と問５のクロス表（表８）において理系の２人が問１においてカテゴリ２（中学生）あるいはカテゴリ３（高校生）と答えていることに対応していると考えられる。第３主成分は、心理学という分野を知った時期（問１）の時間の流れによる問２（心理学とは）あるいは問５（文系／理系）の答えの傾向を表すものと解釈される。

表８　問１と問５のクロス表

		問５　自分の得意分野は？		
		1. 文系	2. 理系	計
問１ 心理学を知ったのは？	１．小学生の頃には知っていた	3	0	3
	２．中学生の頃に知った	１１	1	１２
	３．高校生の頃に知った	2	1	3
	計	１６	2	１８

まとめ

図１の質問紙を用いて心理学科１年生から１８人のカテゴリデータを収集し、共通数量化とその主成分分析を行った。第１主成分から、卒業後の進路選択として「心理学を活かした職業」を希望し、心理学という分野があることを知ったのは「中学生の頃」で、大学受験時には「心理学に臨床心理学以外の分野があることを知っていた」という傾向が見いだされた。第２主成分は、理系が２人いるが、彼女らが、卒業後の進路として「心理学を活かした職業」以外のカテゴリ、すなわち「カウンセラー」あるいは「その他」を選んでいるという傾向を反映していると解釈できる。第３主成分は、心理学という分野があることを知った時期（問１）の時間の流れに沿う問２（心理学とは）あるいは問５（文系／理系）の答えの傾向を表すものと理解される。以上のように、共通数量化と主成分分析によって、カテゴリ変数のクロス表に含まれている情報をより詳しく読み取ることができる。

カテゴリ変数の分析法として現在よく知られている他の方法では、対象とするカテゴリ変数の各カテゴリ値がまとめて同時に数量化されるが（例えば、西里（1982)、p. 69)、この場合、得られた各カテゴリの数量化された値の解釈が難しくなりやすい。しかし、共通数量化の場合は、まず、カテゴリ変数ごとに数量化が行われるので、得られた数量化の解釈がより容易になり、また、変数間の関係も扱いやすくなる。すなわち、共通数量化では、数量化（第１ステップ）と変数間の分析（第２ステップ）に分かれているので、第１ステップで行われる数量化の解釈は、数量化がカテゴリ変数ごとに行われることにより容易になり、その後、この数量化変数の分析が行われるので、変数間の関係が見やすくなる（例えば、図２，３)。また、数量化変数は複数の分析において共通のものが用いられるので、分析の解釈に一貫性を保つことが期待される。さらに、第１ステップで構成された共通数量化変数は、第２ステップの分析において他の量的変数とともに分析することができる。本研究は、主成分分析のみの報告であるが、複数の分析を共通数量化変数に適用した場合、および他の量的変数とともに分析した場合については、仮想データを用いた分析例が Okamoto (2015)に示されている。

なお、本研究の分析において用いたプログラムは、著者のウェブサイト
http://y-okamoto-psy1949.la.coocan.jp/booksetc/TwoStepCmnQExmpl/PrgsExpln/

から入手可能である。共通数量化は、SPSS などによっても可能であるが、SPSS による方法を説明した PDF ファイルはウェブサイト上に
http://y-okamoto-psy1949.la.coocan.jp/booksetc/jwu_2016_2015/HowTo_SPSS.pdf
として用意されている。

［引用文献］

Gifi, A. (1990). *Nonlinear multivariate analysis*. Chichester: John Wiley & Sons.

Greenacre, M. (2007). *Correspondence analysis in practice, second edition.* Boca Raton: Chapman & Hall/CRC.

林知己夫・樋口伊佐夫・駒沢勉（1970）．情報処理と統計数理．産業図書.

岩坪秀一（1987）．数量化法の基礎．朝倉書店.

村上　隆（1999）．カテゴリカル・データの主成分分析の心理計量学的研究．科学研究費研究成果報告書（名古屋大学：課題番号 09610114）.

西里静彦（1982）．質的データの数量化――双対尺度法とその応用――．朝倉書店.

Nishisato, S. (2007). *Multidimensional nonlinear descriptive analysis.* Boca Raton: Chapman & Hall/CRC.

Okamoto, Y. (2006). A justification of rotation in principal component analysis: Projective viewpoint of PCA. *Faculty of Integrated Arts and Social Sciences journal: Japan Women's University*, 17, 59-71.

岡本安晴（2009）．統計学を学ぶための数学入門［下］．培風館.

岡本安晴（2013）．カテゴリカルデータの単純２段主成分分析：カテゴリ型と連続型の同時主成分分析．日本行動計量学会第 41 回大会抄録集、192－195.

岡本安晴（2014）．２段数量化分析法：Two step analysis with unified quantification. 日本行動計量学会第 42 回大会抄録集、72－75.

岡本安晴（2015）．２段共通数量化分析法による分析例：ビッグデータ分析のために．日本行動計量学会第 43 回大会抄録集、298－301.

Okamoto, Y. (2015). A two-step analysis with common quantification of categorical data. *Faculty of Integrated Arts and Social Sciences journal: Japan Women's University*, 26, 99-112.

付　記

カテゴリ変数のカテゴリ数が２のときの数量化は、それぞれのカテゴリに異なる数値を与えるという条件の他に、平均が０、分散が１という標準化の制約により、符号の反転を除いて決まる。また、数量化のウェイトベクトル$\boldsymbol{\omega}$は、以下に示すようにデータによらず反転を除いて同じベクトル$\boldsymbol{\omega} = 1/\sqrt{2}\,(1 \quad -1)'$である。

いま、カテゴリ数が２のカテゴリ変数のデータにおいて、第１カテゴリが選ばれたものがN_1個、第２カテゴリが選ばれたものがN_2個であるとする。このとき indicator 行列Gは次のように書くことができる。

$$G = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} N_1\text{行} \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} N_2\text{行} \end{matrix}$$

これより、次式を得る。

$$H = G - \frac{1}{N}\mathbf{1}_N\boldsymbol{f} = \begin{bmatrix} N_2/N & -N_2/N \\ \vdots & \vdots \\ N_2/N & -N_2/N \\ -N_1/N & N_1/N \\ \vdots & \vdots \\ -N_1/N & N_1/N \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} N_1\text{行} \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} N_2\text{行} \end{matrix}$$

ただし、

$$N = N_1 + N_2$$

である。

したがって、

$$H'H = \begin{bmatrix} N_1 \cdot N_2^2/N^2 + N_2 \cdot N_1^2/N^2 & -N_1 \cdot N_2^2/N^2 - N_2 \cdot N_1^2/N^2 \\ -N_1 \cdot N_2^2/N^2 - N_2 \cdot N_1^2/N^2 & N_1 \cdot N_2^2/N^2 + N_2 \cdot N_1^2/N^2 \end{bmatrix}$$

となる。

$$C = N_1 \cdot N_2^2/N^2 + N_2 \cdot N_1^2/N^2$$

とおけば、

$$H'H = \begin{bmatrix} C & -C \\ -C & C \end{bmatrix}$$

である。これより、

$$H'H\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 2C \\ -2C \end{bmatrix} = 2C\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

が成り立つことがわかる。すなわち、N_1およびN_2の値に関わらず$H'H$の固有ベクトルは

$$\boldsymbol{\omega} = \pm\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

である。