

# 分類データに順序情報がある場合の 解析法について

—大学生の行動とボランティア意識との関連—

野 口 博 司

## 要 約

本論文は、アンケート調査によって、「友達とのコミュニケーション手段」の行動と「大学生のボランティア活動」に対する意識との関係を探るために、得られた回答を分類データのままで数量化の方法3類によって解析した場合と、得られた回答に順序情報が含まれていたので順序情報を活かしたカテゴリカル主成分で解析した場合とを比較検討する。

その結果、順序情報を活かした解析の場合には、数量化の方法3類では発見できなかった因果関係が抽出できる可能性があることを示す。一方、この事例を通じて、制約下で求めた解の歪の存在についても論じる。

キーワード：行動と意識，数量化の方法3類，カテゴリカル主成分，  
順序情報

## 1. は じ め に

一般的に、社会科学調査においては、アンケートの回答である分類データから傾向をさぐるためにはクロス集計を行う。しかし、クロス集計はその組み合わせが多いことから、事前に数量化の方法3類にて全体的な傾向を把握することが多い。その場合に、分類データに順序情報が含まれているにもかかわらず、その順序情報を活かした解析を行っていないケースが多々見受けられる。分類データに順序情報がある場合には、その順序情報を活かす方法で解析するほうが、有意な情報を取り出せる場合がある。しかし、一方では、制約下における

分類データに順序情報がある場合の解析法について

解のために、解の歪の存在もあり解釈上の注意も必要である。今回、大学生の生活習慣とボランティア意識とはどのような関係にあるかについて調査を行い、どのような習慣行動を持っている学生がボランティアに関心があるのかを探ることにした。本事例を通じて一つの消費者行動と意識との関係を捉えるための方法論のあり方について言及したい。

## 2. 目 的

消費者行動と意識との関係を探るための解析法のあり方について提言する。

## 3. 方 法

今回の研究は以下の通りで進めた。

- ①目的のための簡単なアンケート調査を実施する。
- ②アンケート回答の全体的傾向を把握するために、数量化の方法3類にて解析する。
- ③アンケート回答の分類データに順序情報がある場合には、その情報を活かした最適な手法、即ち、カテゴリカル主成分を用いて解析する。
- ④②と③の解析結果について、社会科学上の解釈の立場からと方法論からの考察をする。
- ⑤得られた知見をまとめて今後の解析法のあり方について提言する。

## 4. 結 果 と 考 察

### 4.1 アンケート調査内容とデータについて

アンケート調査の内容は下記の通りであり、最終的には大学生100人に対してデータを収集した。

- (1)対象：神戸の大学へ通っている学生28名と、更に結果を確認するために追加調査した大学生72名の合計100名。
- (2)質問項目は次の通りである。

下記の設問に対して、貴方の回答に合う [ ] 内の番号に、○をつけて下さい。

問1. 社会貢献への関心はありますか。

[1.ない。 2.少しある。 3.やや高い。]

問2. ボランティア活動への関心はありますか。

[1.ほとんどない。 2.少しある。 3.関心がある。]

問3. 友達はすぐできますか。

[1.はい。 2.いいえ。]

問4. 友達とのコミュニケーション手段は主に下記の何を用いますか。

[1.メールで。 2.電話で。 3.直接会って。]

(3)学生28名により得られた当初のデータは表1の通りである。

表1. 学生28名のアンケート回答データ

氏名	性別	問1社会貢献	問2ボランティア	問3友達	問4手段
青木	2	3	2	2	1
橋本	2	3	2	2	1
藤原	2	3	2	2	1
杉山	1	3	2	1	3
塩沢	1	2	1	1	1
山本	1	2	2	2	1
尾崎	1	1	2	1	3
小藪	1	2	2	1	2
藤本ア	1	2	2	1	3
高岡	1	1	1	2	1
近藤	1	1	2	1	3
森岡	1	1	2	2	2
長井	1	2	2	2	1
鮫島	1	2	3	1	3
藤井ユ	1	2	1	1	1
宮崎	1	2	2	2	1
奥田	1	3	2	2	2
富上	1	3	2	2	1
藤井ト	1	2	2	2	2
中北	1	1	2	2	3
道満	1	2	2	2	2
藤本	1	2	1	1	3
篠原	1	2	1	1	1
榎本	1	1	1	1	1
脇田	1	1	1	1	3
浜島	1	2	2	2	2
小川	1	2	2	1	2
鈴木	1	1	1	1	3

分類データに順序情報がある場合の解析法について

#### 4.2 分類データとした場合の数量化の方法3類による解析結果

表2. 数量化の方法3類の成分ベクトルと固有値 (n=28)

変数名	成分1	成分2	成分3	成分4	成分5
性別: 1	-0.147	0.127	-0.061	-0.036	0.045
2	0.425	-0.367	0.177	0.103	-0.131
問1社会: 1	-0.237	-0.223	0.002	-0.554	0.413
2	-0.087	0.338	-0.172	0.427	-0.110
3	0.406	-0.259	0.260	-0.012	-0.309
問2ボラ: 1	-0.295	-0.327	-0.426	0.103	-0.021
2	0.230	0.198	0.146	-0.186	-0.099
3	-0.168	0.062	0.567	0.521	0.492
問3友達: 1	-0.351	-0.109	0.065	0.113	-0.430
2	0.351	0.109	-0.065	-0.113	0.430
問4手段: 1	0.201	-0.303	-0.356	0.299	0.213
2	0.103	0.573	-0.049	-0.182	-0.126
3	-0.323	-0.156	0.454	-0.185	-0.135
固有値	0.502	0.312	0.239	0.229	0.127
累積寄与率	0.314	0.509	0.658	0.801	0.880

表2は、最初の調査28名による回答データの表1から、数量化の方法3類で解析した結果である。そして、表2は、求められた各成分の成分ベクトル値と固有値を示している。数量化の方法3類とは、成分1で各設問の回答番号（各変数名のカテゴリー）に与える数値（成分ベクトル値）と、対象者の回答に従って合計される対象者の合計数値（成分ベクトル合計値）との相関が最大になるようにして、各設問の回答番号に最適な数値を与える方法である。

成分2は、成分1で与えた数値と成分2で与える数値との内積が0となる（成分1と成分2は直交する）条件下で、2番目に相関係数が大きくなるような数値を各設問の回答番号に与えられる。以下順に成分5までの与えられた数値（ベクトル値）を表2に示す。固有値は各成分の上記の相関係数の2乗を示す。今回の解析では、成分1と2との累積寄与率は0.509であり、この2つの成分で全体の情報の半分が再現できている。

一般的に、数量化の方法3類の解析結果は、固有値の出方がなだらかになり、集約した成分を抽出するのが困難であることが多い。また、相関関係を最大にする条件下で各設問の回答番号に対して数値を与えることから、表1のように、

成分1の問2におけるボランティア意識のカテゴリーに与えられた数値、即ち、関心が「1.ほとんどない。」が-0.295、「2.少しある。」が0.230、「3.関心がある。」が-0.168となり、数値が高くなるほど関心があるという解釈に適合する結果にならない場合がある。従って、分類データに方向性があっても、得られた数値の解釈では各成分の意味が捉えられないことが生じる。表2のグレー部分の成分ベクトルはいずれも対応がとれていないことを示している。

今回は、友達とのコミュニケーション手段行動とボランティア活動への関心との関係を調査した事例を通じて、消費者行動と意識との関係を捉えるための方法論について、以下に詳細に考察していく。

#### 4.2.1 クロス集計と数量化3類による詳細解析

友達とのコミュニケーション手段行動とボランティア活動への関心度合いとのクロス集計を行い、それらの独立性について $\chi^2$ 乗検定した結果を表3に示す。

表3. 友達とのコミュニケーション手段とボランティア意識

変数名		問1 社会貢献	問2 ボランティア	問4 手段
問1 社会貢献	$\chi^2$ 乗	—	5.316	5.939
	P値上側	—	0.256	0.204
問2 ボランティア	$\chi^2$ 乗	5.316	—	6.330
	P値上側	0.256	—	0.176
問4 手段	$\chi^2$ 乗	5.939	6.330	—
	P値上側	0.204	0.176	—

表3より、 $\chi^2$ 乗値=6.330（有意確率0.176）となり、友達とのコミュニケーション手段行動とボランティア活動への関心意識とは有意でなく、独立している。

さらに、友達との各コミュニケーション手段とボランティア活動への関心度合いとの関係を確認するために、今回得られた数量化3類の各成分ベクトル値を図1に図示してみた。その結果、友達と直接会うことが多い学生はボランティアに関心があるケースとないケースに二分される。また、メールで連絡する

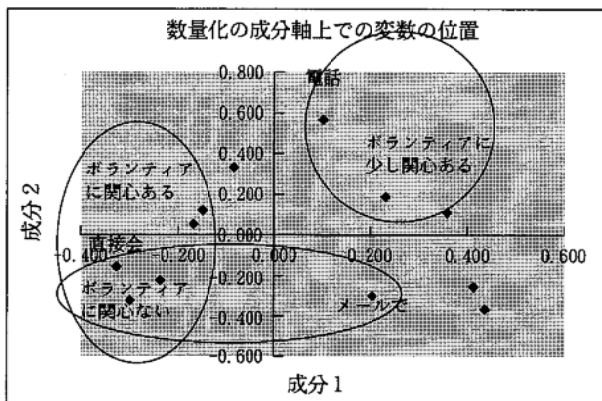


図1. コミュニケーション手段とボランティア意識の位置

学生はボランティアに関心がない様に見える。しかし、 $\chi^2$ 乗検定の結果から、学生の行動と意識には関係がないので、このことは有意であるとはいえなかった。

そこで、ボランティア活動の関心に対する回答には、1.ほとんどない。<2.少しある。<3.関心がある。という様に順序があるので、この順序情報を活かす制約付きカテゴリカル主成分分析を行った。

#### 4.2.2 制約付カテゴリカル主成分の結果

カテゴリカル主成分の原型は主成分分析と同じである。即ち、表1のデータを数値データとして、制約を付与しないでカテゴリカル主成分分析を行うと、相関行列から導いた主成分分析の結果と一致する。表4は、SPSSで求めた主成分分析の結果を示している。主成分1と主成分2とで全体の69.439%説明できている。

次に、社会貢献への関心とボランティア活動への関心に対する回答には、順序関係が存在するので、その条件を活かしてカテゴリカル主成分を行った。得られた結果が表5である。表4の固有値の2.282よりも固有値が2.222と少なくなり、制約が入ることにより、1次元の第1主成分の説明力は落ちているこ

表4. 表1をデータとした制約無しの主成分分析の結果  
モデルの要約

次元	Cronbach のアルファ	説明された分散	
		合計(固有値)	分散の%
1	.702	2.282	45.637
2	.220	1.190	23.802
合計	.890 <sup>a</sup>	3.472	69.439

a. Cronbachのアルファ合計は、固有値合計に基づいている。

成分負荷

	次元	
	1	2
性別	.751	-.080
問1 社会貢献	.779	.124
問2 ボランティア	.421	.854
問3 友達	.729	.032
問4 手段	-.634	.662

変数主成分の正規化

表5. 順序情報を入れた制約下で求めたカテゴリカル主成分の結果

モデルの要約

次元	Cronbach のアルファ	説明された分散	
		合計(固有値)	分散の%
1	.688	2.222	44.438
2	.392	1.457	29.144
合計	.910 <sup>a</sup>	3.679	73.582

a. Cronbachのアルファ合計は、固有値合計に基づいている。

成分負荷

	次元	
	1	2
問1 社会貢献	.833	-.176
問2 ボランティア	.599	.671
問3 友達	.700	.280
問4 手段	-.176	.874
性別	.805	-.367

変数主成分の正規化

とが判る。ただし、2次元の解は1次元と直交する条件で導かれるので説明力は落ちるとは限らない。

次に、この制約下で導かれた各カテゴリーの数量結果を表6に示す。

表6. 各カテゴリーに与えられた数量の結果

質問	カテゴリー	度数	制約数量化	1次元ベクトル座標	2次元ベクトル座標
問1 社会貢献	ない	8	-0.633	-0.527	0.112
	少しある	14	-0.457	-0.381	0.081
	やや高い	6	1.909	1.591	-0.337
問2 ボランティア	ほとんどなし	8	-1.581	-0.947	-1.060
	少しある	19	0.632	0.379	0.424
	関心がある	1	0.632	0.379	0.424
問3 友達	いいえ	14	-1.000	-0.700	-0.280
	はい	14	1.000	0.700	0.280
問4 コミュ手段	メールで	12	-0.973	0.172	-0.850
	電話で	7	1.537	-0.271	1.343
	直接会って	9	0.102	-0.018	0.089
性別	男性	25	-0.346	-0.279	0.127
	女性	3	2.887	2.325	-1.058

分類データに順序情報がある場合の解析法について

$i$ 次元成分の  $p$  質問の成分負荷  $r_{ip}$  と、 $p$  質問  $k$  のカテゴリの数量化  $x_{pk}$  との積は  $r_{ip} \times x_{pk} = l_{ipk}$  となり、 $i$ 次元成分のベクトル座標  $l_{ipk}$  となる。また、 $i$ 次元成分の質問  $p$  における主成分負荷の2乗和は  $\sum r_{ip}^2 = \lambda_i$  と  $i$ 次元成分の固有値と一致する。

いま、簡単に、被験者  $n$  の反応カテゴリ Yes, No の即ち 1, 0 のデータ行列を  $D$  とおき、求めたい数量化行列を  $X$  とおくと、被験者の成分得点  $Z$  は  $Z = XD$  となる。

即ち、カテゴリカル主成分は主成分分析をより一般化したものであり、

$$\text{Var}\{Z\} = \text{Var}\{XDD^T X^T\} = \text{Tr}\{\lambda_i\}$$

となる。

制約は、制約のある質問  $p$  において、一次元解である数量には  $x_{p1} \leq x_{p2} \leq \dots \leq x_{pk}$  と順序制約が入る。具体的には、順序整合をとるためには単調変換を行うことになる。また、求める数量  $x_{pk}$  は、固有ベクトルを、カテゴリの反応数  $f_{pk}$  と  $x_{pk}$  数量とにおいて、

$$\sum f_{pk} x_{pk} = 0, \sum f_{pk} x_{pk}^2 = 1$$

と規準化して導かれる。得られた数量において、被験者の一次元成分得点の分散が再度計算される。この計算された分散が、カテゴリカル主成分の一次元の分散となる。制約があるので、必ず一次元の分散は、制約のない主成分分析の解の第1固有値より小さくなる。

二次元解は、一次元解の固有ベクトルとの内積を0にする条件下で、カテゴリの反応数  $f_{pk}$  とカテゴリ数量  $x_{pk}$  とにおいて、再び

$$\sum f_{pk} x_{pk} = 0, \sum f_{pk} x_{pk}^2 = 1$$

と規準化して導かれる。従って、制約のない主成分分析結果の第2固有値よりも大きくなる場合がある。SPSS 解析ソフトでは、三次元以上は複雑になるので解を導くことを推奨していない。

制約下で求めた「友達のコミュニケーション手段」と「ボランティア活動への意識」とのベクトル座標の関係を図示すると図2のようになった。



問2 ボランティアベクトル座標		
	1次元	2次元
ほとんどない	0.947	1.060
少しある	0.379	0.424
関心がある	0.379	0.424
問4 手段ベクトル座標		
	1次元	2次元
メールで	0.172	-0.850
電話で	-0.271	1.343
直接会って	-0.018	0.089
固有値	2.222	1.457
累積寄与率	44.438	73.582

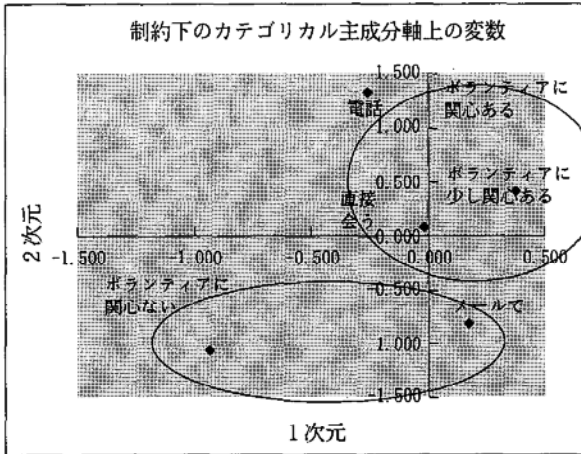


図2. コミュニケーション手段とボランティア意識との位置

図2から友達とメールでコミュニケーションを図る学生はボランティアに関心がなく、友達と電話や直接会う学生はボランティアに少し関心がある兆候が伺える。図1の数量化方法3類の結果よりも、今回の解析結果である図2の方が、友達と直接会う行動とボランティアの関心がある度合いとの対応が取れているようである。

更に、これら二つの項目間の関係を図示して捉えるのには、コレスポンデンス分析が適するので、その解析を行った。

分類データに順序情報がある場合の解析法について

#### 4.2.3 コレスポネンス分析の結果—項目間の関係確認—

表7は、「友達のコミュニケーション手段」のカテゴリーと「ボランティア活動へ意識」のカテゴリーとにおけるクロス集計の結果である。この表7より、コレスポネンス分析を行い、得た結果が表8である。

表7. 「コミュニケーション手段」と「ボランティア活動へ意識」とのクロス集計  
コレスポネンステーブル

問2 ボランティア	問4 手段			
	メールで	電話で	直接会って	周辺
ほとんどない	5	0	3	8
少しある	7	7	5	19
関心がある	0	0	1	1
周辺	12	7	9	28

表8. 表7のコレスポネンステーブルから計算した結果  
要約

次元	特異値	要約イナーシャ	カイ2乗	有意確率	イナーシャの寄与率		信頼特異値	
					説明	累積	標準偏差	相関 2
1	.399	.159			.703	.703	.094	.147
2	.259	.067			.297	1.000	.135	
要約合計		.226	6.330	.176 <sup>a</sup>	1.000	1.000		

a. 自由度 4

表8の(特異値)<sup>2</sup>は固有値に対応し、(特異値)<sup>2</sup>=要約イナーシャとなる。イナーシャの寄与率は、(要約イナーシャ)/(要約イナーシャの合計)であり、何次元までで全体の何パーセントが説明できているかを示す。今回は、行のカテゴリ

表9. 行の項目についての次元の得点  
行ポイントの概要<sup>a</sup>

問2 ボランティア	マス	次元の得点		概要 イナーシャ	寄与率				
		1	2		次元のイナー シャに対する ポイント		ポイントのイナーシャ に対する次元		
					1	2	1	2	概要合計
ほとんどない	.286	-.887	.370	.100	.563	.151	.899	.101	1.000
少しある	.679	.433	-.025	.051	.320	.002	.998	.002	1.000
関心がある	.036	-1.143	-2.479	.075	.117	.847	.247	.753	1.000
合計	1.000			.226	1.000	1.000			

a. 対称的正規化

リーが3なので二次元の図示で100パーセント示せる。また、行と列の3×3分割表の $\chi^2$ 乗値は6.330であり、表3の結果と同じである。

表9は、行の項目についての次元の得点であり、ボランティアに関心が『殆どない』 $=(-0.887, 0.370)$ となる。

表10. 列の項目についての次元の得点  
列ポイントの概要<sup>a</sup>

問4手段	マス	次元の得点		概要 イナーシャ	寄与率				
		1	2		次元のイナーシャに対するポイント		ポイントのイナーシャに対する次元		概要合計
					1	2	1	2	
メールで	.429	-.292	.538	.047	.092	.480	.312	.688	1.000
電話で	.250	1.087	-.098	.118	.741	.009	.995	.005	1.000
直接会って	.321	-.456	-.642	.061	.167	.511	.437	.563	1.000
合計	1.000			.226	1.000	1.000			

a. 対称的正規化

表10は、列の項目についての次元の得点であり、友達とのコミュニケーション手段の『メールで』 $=(-0.292, 0.538)$ となる。

行ポイントと列ポイント  
対称的正規化

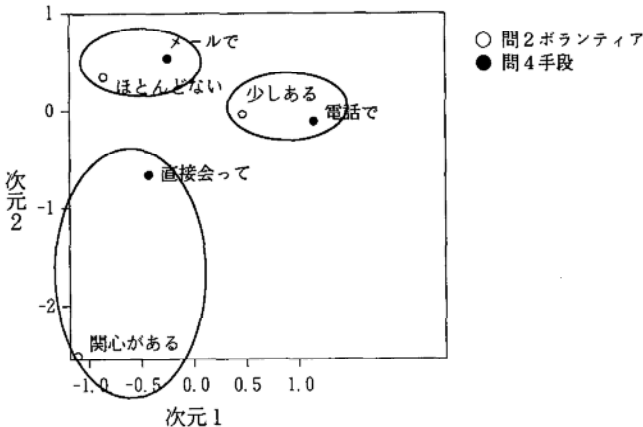


図3. 行次元の得点と列次元の得点との同時布置



項目間のカテゴリーにおけるクロス集計を行うと表12のようになり、 $\chi^2$ 乗検定の結果統計量は

$$\chi^2_{0.2} \text{乗値} = 7.677547 \text{ (有意確率 } 0.10413)$$

と、やはり有意にはならなかった。

表 12. 追加調査による  $\chi^2$ 乗検定 (n=100) 結果

	ボランティア 関心なし	少しある	関心ある	計
メールで	29	17	1	47
少しある	14	14	3	31
直接会って	8	10	4	22
計	51	41	8	100

行変数の数	3
列変数の数	3
自由度	4
帰無仮設 H0	$P_{ij} = P_{ij0}$
統計量 $\hat{\chi}^2$	7.677547
P 値 (上側)	0.10413
帰無仮設 H0 は有意水準 5% で棄却されない	

もともと、ボランティア意識の回答には関心度合いの順序があるので、順序情報を扱うノンパラメトリック手法のクラスカル・ウォリスの検定を再度実施した。

その結果、表13のようになり、 $\chi^2_{0.2}$ 乗値=6.064 (有意確率0.048) で有意とな

表 13. 順位データとして、クラスカル・ウォリスの検定結果

帰無仮設 H0	
各母集団の中心位置は等しい	
対立仮設 H1	
各母集団の中心位置は等しくない	
データ数	47, 31, 22
順位和 Ri	2074.5, 1661.5, 1314.0
クラスカル・ウォリス検定統計量 KW	6.064
検定統計量 $\hat{\chi}^2_0$	6.064
P 値 (上側)	0.048
統計量 $\hat{\chi}^2$	5.991
$\hat{\chi}^2_0 \geq \hat{\chi}^2$ より有意水準 5.0% で帰無仮設 H0 は棄却される	

分類データに順序情報がある場合の解析法について

った。即ち、「友達とのコミュニケーション手段」と「ボランティアへの意識」とは独立であるとはいえない。即ち、メールが手段のものはボランティアに関心がなく、直接会うものはボランティアに関心があるという傾向になる。即ち人との会話を重んじる社会的な学生は、社会的なボランティア活動にも関心を示すようだ。

#### 4.2.5 追加調査によるコレスポンデンス分析の結果

追加の調査データによりコレスポンデンス分析を実施した。その結果が表14である。

表14. 追加調査表13のクロス集計から計算したコレスポンデンス分析の結果

要 約

次元	特異値	要約 イナーシャ	カイ2乗	有意確率	イナーシャの寄与率		信頼特異値	
					説明	累積	標準偏差	相関 2
1	.275	.075			.982	.982	.092	.252
2	.037	.001			.018	1.000	.106	
要約合計		.077	7.678	.104 <sup>a</sup>	1.000	1.000		

a. 自由度 4

表15のボランティアへの関心度合いカテゴリーの次元の得点と、表16の友達とのコミュニケーション手段カテゴリーの次元の得点とを2次元に図示したものが図4である。カテゴリー間の関係が図3の図示よりもより密接になっている

表15. 行項目についての次元の得点

行ポイントの概要<sup>a</sup>

ボランティア	マス	次元の得点		概要 イナーシャ	寄与率				
		1	2		次元のイナーシャに対するポイント		ポイントのイナーシャに対する次元		概要合計
					1	2	1	2	
ほとんどない	.510	-.393	.122	.022	.286	.204	.987	.013	1.000
少しある	.410	.195	-.221	.005	.057	.533	.852	.148	1.000
関心がある	.080	1.501	.351	.050	.657	.263	.993	.007	1.000
合 計	1.000			.077	1.000	1.000			

a. 対称的正規化

表 16. 列項目についての次元の得点  
列ポイントの概要<sup>a</sup>

友達手段	マス	次元の得点		概要 イナーシャ	寄与率				
		1	2		次元のイナー シャに対する ポイント		ポイントのイナーシャ に対する次元		概要合計
					1	2	1	2	
メールで	.470	-.509	.083	.033	.443	.087	.996	.004	1.000
電話で	.310	.205	-.279	.004	.047	.643	.799	.201	1.000
直接会って	.220	.798	.214	.039	.510	.270	.990	.010	1.000
合計	1.000			.077	1.000	1.000			

a. 対称的正規化

行ポイントと列ポイント  
対称的正規化

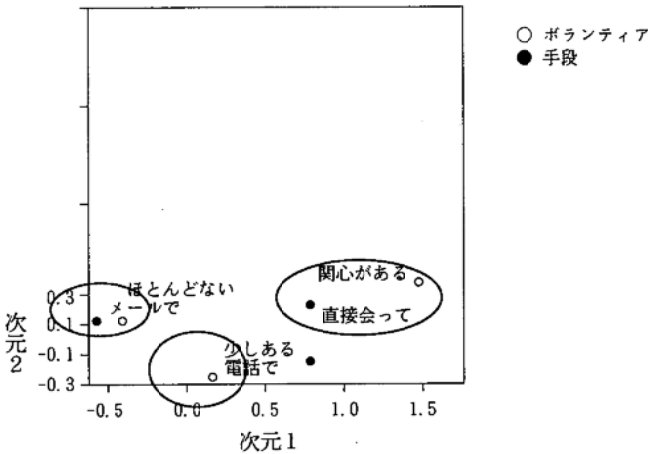


図 4. 追加調査後の行次元の得点と列次元得点との同時布置

ることがわかる。即ち、メールでコミュニケーションをとるものはボランティアに関心がなく、直接会うものはボランティアに関心を持っている。

## 5. 考 察

### 5.1 社会科学調査結果としての考察

今回の大学生における「友達とのコミュニケーション手段」と「ボランティ

分類データに順序情報がある場合の解析法について

「活動に対する意識」との調査では、回答データの持つ順序情報を活かしたカテゴリカル主成分で解析を行えば、行動と意識とに関連がありそうな傾向が伺えた。そこで、その関係を確認するために、サンプル数を増やして再度確認追加調査を行って調べた。その結果、「友達とのコミュニケーション手段」と「ボランティア活動に対する意識」とは有意となった。即ち、友達と直接会う学生はボランティアに関心があり、メールで連絡する学生はボランティアに関心がないという、1つの興味を引く結果が導かれた。

社会科学の調査研究では、あまり統計的な考えや手法の特徴を考慮しないで、解釈の面白さを優先させて結果を導くことが多い。今回のような場合には、最初からカテゴリカル主成分を用いていれば、このような結果を支持するだろう。しかし、制約下で求めた解には歪があることに注意しなければならない。今回の結果では、解の歪は小さく、また追加調査の結果から、「友達とのコミュニケーション手段」と「ボランティア活動に対する意識」とは有意になったので、問題はない。

しかしながら、制約下で求めた解の場合には、解析上の癖があるので、そのことを考慮した上での結果の解釈が必要である。次に、今回の事例を通じて、解析上についての注意を示す。

## 5.2 解析上からの考察

今回用いた制約下での解析上から考察すると、以下のことがいえる。

- ①得られたデータと回答者とのマッチ度を単に重んじるのなら、数量化3類による解析で、回答者の回答パターンに傾向があるかどうかを考察するだけでよい。ただし、関係ある項目間は、必ず $\chi^2$ 乗検定で有意性の確認をする。
- ②回答データに順序がある場合は、カテゴリカル主成分で求めるとよい。ただし、制約を課すことにより解釈に都合のよさが加わる場合がある。特に、二次元解は制約下の一次元解との直交性下で導かれるので、元データの構造を変える可能性がある。従って、制約により、どのような構造に変化が生じた



のか、その変化を調べておくことが必要である。今回の事例では、固有値の変化は小さかったのでその影響は少なかったといえる。

③制約を加えた場合のメリットは、元データに対する説明が容易になり、解析の対象にしなかった外部データに対しての予測を可能にする。常に②で述べたような元データの構造の変化を考慮して、仮説構造変化の経験的妥当性を常に吟味して解釈することが大切である。

④制約を加えた場合は、その制約を活かした統計的検定法を併用して確認する。

⑤また関連性があるという傾向が出たのなら、サンプル数を増やして、追加調査して結果の安定性を確認する。等が提言できる。

## 6. ま と め

最近の社会科学やマーケティングの研究では、データマイニングと称して、制約を課すことにより得られる新しい知見が「知の情報」としてもはやされることがある。予測を主に行う場合は、制約下で得られた結果を用いる方が頑強といえる。しかし、果たして得られた知見は確かなのかを解析上の癖も考慮して確認すべきである。特に、市販の著名な解析ソフトは活用のためのマニュアルは充実しているが、解析アルゴリズムは示されていない。従って、解析上表れる癖と調査結果の傾向とを分けて把握する姿勢が必要である。

今回、大学生の行動と意識との関係について、SPSSのカテゴリカル主成分で解析を行ったところ、「知の情報」結果が読み取れた。そこで、その情報を確認するために、サンプル数を増やして再度考察検討を行った。その結果、得られた「知の情報」を支持できたが、制約下の解析で得られる知見には、数理的な影響もあるので、得られた結果が常に安定したものかどうかを今後も確認していく姿勢が必要である。

## 参 考 文 献

- [1] Pralo Giudici, "Applied Data Mining", Principles of Data Mining MIT Pre-

分類データに順序情報がある場合の解析法について

- ss, Cambridge MA, Wiley, (2003).
- [ 2 ] Rissanen.J., "Stochastic complexity in learning", Jr. Computer System and Sciences, 55(1), (1997), pp. 89-95.
  - [ 3 ] Benzecri, J. P., "Statistical analysis as a tool to make patterns emerge from data", In Methodologies of Pattern Recognition (S.Watanabe, ed, New York: Academic (1969)), pp. 35-60.
  - [ 4 ] Hair, Anderson, Tatham, Black., "Multivariate Data Analysis", Fifth Edition, Prentice Hall, (1998).
  - [ 5 ] Isogai.T and Noguchi.H., "Singular value decomposition and the application", Journal of Social Science Research, Osaka University, No. 40, (1992), pp. 63-101.