岡山大学環境理工学部研究報告
第10巻，第1号，pp.49-56，2005年2月

コレスポンデンス分析における変数選択規準の検討

森　裕一*，杜　暁東**，飯塚誠也***

Considering variable selection criteria in correspondence analysis

Y. Mori, X. Du, M. Iizuka

(Received December 28, 2004)

Ordinary goodness of fit criteria in correspondence analysis are considered as variable selection criteria in case correspondence analysis which is one of multivariate methods without external variables can be applied. The goodness of fit criteria focused here are proportion of cumulative eigenvalues, proportion of cumulative squared-eigenvalues and proportion of cumulative off-diagonal fitness. Each criterion is applied to a couple of real data sets and evaluated with interpretation of the selection process and result (selected subset of variables). Four selection procedures such as backward elimination and forward-backward selection are also performed to compare with each other as well as with all possible selection procedure. These results illustrate that the criteria can be used as selection criteria to select a subset of variables in correspondence analysis and to assess categorical items (questions) in a survey (questionnaire).

1. はじめに

コレスポンデンス分析（数量化III類）を用いて分析を行う場合で，調査項目や検査項目を減らし，データのもつ全体的な様相をなるべく少ない情報損失で単純化したい，あるいは，無駄な項目を除去したいという場合を考える。データの特徴を測りとれる指標を低次元で作ろうという場面も，実施上の観点を考えるならば項目数は少ない方がよい。このような場面では，コレスポンデンス分析における変数選択を考えることになる。

コレスポンデンス分析は，外的変量をもたない多変量解析手法である。外的変量とは，たとえば，回帰分析における目的変数のように，他の変数から導かれる情報を基に予測あるいは説明される変数のことである。これをもたない多変量手法には，他に主成分分析や因子分析などがある。これらの手法における変数選択では，回帰分析のように外的変量を規準とした選択が行えないために，その選択の規準をそれぞれの手法に応じて考案していく必要がある。実際，その選択目的に応じて，いくつかの規準が考えられており，たとえば，主成分分析では，Robert and Escoufier (1976)，Krzanowski (1987)，Tanaka and Mori (1997)，森 他 (1999) などが，因子分析においては，Tanaka and Kodake (1981)，飯塚 他 (2002) などが，コレスポンデンス分析においては，Mori and Maehashi (1996)，Xia and Yang (1998)，Iizuka et al. (2002b) などがある。このように，選択規準は複数存在し，かつ，多くの場合，それぞれの規準によって選択結果が異なる（選ばれる変数群が選択規準で異なる）ということが起こる。したがって，実際の選択場面においては，選択の目的がはっきりしていればその目的に適した選択規準で選択を行えばよいし，選択の目的が明確でない場合は，いくつかの手法を試してみてその結果を比較・検討するということが必要になる。

このような状況下では，次の2つのことが必要となる。選択規準の検討・考察と各種の選択規準が実行できる解析環境の提供である。前者は，考えられる目的に対応した選択規準を考案したり，同時にその手法の安定性などを評価することである。後者は，既存の汎用統計パッケ

*　岡山理科大学総合情報学部
**　岡山理科大学大学院総合情報研究科
***　岡山大学環境理工学部

ージで変数選択を行うための関数を作成したり，ネットベースのシステムを含む変数選択ソフトウェアを開発したりすることである。すでに，主成分分析については，既存のものを含む主たる選択手法の検討がなされ（森 他, 1999; 森 他, 2000; Mori et al., 2000b; 森, 飯塚, 2002; Iizuka et al., 2003 など），実行環境が整備されている（Mori, 1998; Mori et al., 2000a; 飯塚 他, 2001）。因子分析では，飯塚 他（2002），コレスポンデンス分析も森 他（2004）などで検討がなされ，各手法の各規規準が変数選択パッケージVASMM（VAriable Selection in Multivariate Methods）に実装されている（http:// mo161.soci.ous.ac.jp/vasmm/index.html, Iizuka et al., 2002a など；主成分分析，因子分析，コレスポンデンス分析の各変数選択手法は，それぞれ VASpca, VASfa, VAScorres として VASMM に実装されている）。

本稿では，前者を目的として，コレスポンデンス分析における変数選択について検討する。コレスポンデンス分析における変数選択の研究は，先にあげたように Mori and Maehashi (1996)，Xia and Yang (1998)，Iizuka et al. (2002b) などで行われている。いずれも個体スコアの布置に注目し，全変数を用いた個体スコアの布置と選択後の変数群による個体スコアの布置を最も近づけるように変数を選んでいる。今回は，これと異なり，コレスポンデンス分析における一般的な適合度規準（GOF, Goodness of Fit）に着目し，これを選択規準として用いることを検討する。ここでは，固有値の累積寄与率，平方累積寄与率，非対角要素による累積寄与率の3つを取り上げ，その特徴や選択手順による違いなどを実データに適用して考察する。

2. コレスポンデンス分析における変数選択

2.1 適合度規準

Yをn個の個体とp個のカテゴリー変数をもつデータ行列とする。このYの第k列のカテゴリーを (0, 1) に指標化した行列をG_kとして，$G=(G_1,G_2,\cdots,G_p)$，$B=G'G$，$D=diag(B)=diag(G'G)$とする。ここで，$GD^{-1/2}$の特異値分解$GD^{-1/2}=U\Delta V'$を考える。そのランクを$R+1$とすると，Δは特異値の対角行列$\Delta=diag(\lambda_0,\lambda_2,\cdots,\lambda_R)$，対応する固有ベクトルは，それぞれ$U=(\boldsymbol{u}_0,\boldsymbol{u}_1,\cdots,\boldsymbol{u}_R)$，$V=(\boldsymbol{v}_0,\boldsymbol{v}_1,\cdots,\boldsymbol{v}_R)$となる。なお，$\lambda_j$ $(j=0,1,\cdots,R)$は大きさの順に並んでいるとする。

このとき，GOF として，次のような固有値の累積寄与率（PCE, Proportion of Cumulative Eigenvalues），平方累積寄与率（PCS, Proportion of Cumulative Squared-eigenvalues），非対角要素による累積寄与率（PCO, Proportion of Cumulative Off-diagonal fitness）を計算することができる（Greenacre, 1994; Adachi, 2004）。

$$PCE=\sum_{j=1}^{r}\lambda_j^2\Big/\sum_{j=1}^{R}\lambda_j^2$$

$$PCS=\sum_{j=1}^{r}\lambda_j^4\Big/\sum_{j=1}^{R}\lambda_j^4$$

$$PCO=\frac{p}{\gamma^2(p-1)}\sum_{j=1}^{r}(\lambda_j^2-1)^2$$

ただし，rは対象とする次元，PCO については，

$\gamma^2=\sum_{j=1}^{p}\sum_{l=1,l\neq j}^{p}\left\|D_j^{-1/2}(B_{jl}-C_{jl}^{(0)})D_l^{-1/2}\right\|^2$，$C_{jl}^{(0)}$は$C^{(0)}=\lambda_0^2D^{1/2}\boldsymbol{v}_0\boldsymbol{v}_0'D^{1/2}$の$B_{jl}$に対応する（$j$, l）部分行列，$\|\bullet\|$はユークリッドノルム，$\lambda_r^2>1$である。

2.2 適合度規準を利用した変数選択

変数選択では，先のYをq個の変数をもつ$n\times q$部分行列Y_1と，残りの$p-q$個の変数をもつ$n\times(p-q)$部分行列Y_2に分割することになる。Y_1を選択変数群，Y_2を削除変数群とすると，q個の変数をもつすべての組み合わせの中で，2.1 にあげた GOF の値を最も大きくするようなY_1を見つけることが目的となる。ただし，すべての組み合わせを調べる総当たり法（AllPossible, All possible combinations）が最善の方法であるが，計算コストを低減するために，一連の変数選択手法の研究では，変数減少法（Back, Backward elimination），変数増加法（For, Forward selection），および変数減少法と変数増加法を 1 変数に関して交互に繰り返していく変数減増法（Back-For, Backward-forward stepwise selection）と変数増減法（For-Back, Forward-backward stepwise selection）の4つの選択手順を採用している（各手順の詳細は，Mori, 1997, 森 他, 1999 を見よ）。今回もこの4手順で変数選択を行うと同時に，4手順それぞれの選択結果の比較，および総当たり法との結果の比較を行い，今回検討する GOF 規準の評価と4選択手順の利用可能性を考察する。

表1　授業アンケート質問項目（「はい」から「いいえ」までの5肢択一式）

問題番号	アンケート内容
1	教員の声は十分聞き取れる大きさでしたか。
2	教員の話し方は適当な早さでしたか。
3	教員の話し方は明瞭でしたか。
4	黒板の字や図形や適当な大きさでしたか。
5	黒板の字や図形は丁寧に書かれていましたか。
6	シラバスに沿って授業がなされました。
7	教員は授業時間を守っていましたか。
8	教科書やテキストは適切でしたか。
9	教員は私語をしている学生に十分な注意をしていると思いますか。
10	授業内容をわかりやすく工夫が感じられましたか。
11	授業の進度は適当でしたか。
12	教員は学生の理解力を配慮していると思いますか。
13	教員は熱心に教えてくれたと思いますか。
14	総合的に見て，この授業を履修してよかったですか。
15	あなたはこの授業によく出席しましたか。
16	あなたはシラバスをよく活用しましたか。
17	あなたは時間外にこの授業の学習をしましたか。
18	総合的に見て，あなたはこの授業に熱心に取り組みましたか。

3. 数値例

GOF を用いたコレスポンデンス分析における変数選択を，大学2年生56名に実施した授業アンケートのデータと，軽症意識障害87例に実施された軽症意識障害検査のデータに適用する。それぞれのデータに対して各規準による変数選択を実施し，Y_1 として用いる変数の数 q ごとの規準値（*PCE*, *PCS*, *PCO*）と選択された変数群の比較を行う。なお，前者のデータに対しては，4手順の選択過程の考察と総当たり法との比較も行う。

3.1 授業アンケートデータ

データは，実際に利用されていた授業アンケート（表1に示した5肢択一の18問，現在は利用されていない）を，本研究のために著者の1人の授業で実施して得られたものである。このうち，教師に対する評価部分である最初の14問に対して変数選択を試みる。

表2は，14変数を $r=2$ とし，変数減少法によって選択した過程の要約である。各 q において選択された変数群（Y_1）から求められる（選択の規準となった）規準値 *PCE*, *PCS*, *PCO* と削除される変数である（図1は各規準値の変化をグラフ化したもの）。

この3つの規準によって選択された変数群を見ると，PCE規準とPCS規準の選択結果は，$q=5\sim7$ で削除される変数 {V6, V 8, V 12} の順番が入れ替わる以外は同じである。最初に削除される3変数 {V4, V7, V9} は，削除された順番は異なるが3つの規準で同じであること

表2　*PCE*, *PCS*, *PCO* の変化（授業アンケートデータの14変数，変数減少法，$r=2$）

q	*PCE*	削除変数	*PCS*	削除変数	*PCO*	削除変数
14	0.24669		0.48825		0.55856	
13	0.26026	V7	0.50805	V7	0.58386	V9
12	0.27457	V9	0.53011	V9	0.60969	V7
11	0.28891	V4	0.54996	V4	0.63615	V4
10	0.30658	V5	0.57033	V5	0.66593	V2
9	0.32338	V1	0.59103	V1	0.69035	V12
8	0.34176	V2	0.61166	V2	0.72477	V5
7	0.36597	V8	0.63206	V6	0.75446	V6
6	0.39678	V6	0.65716	V12	0.79238	V10
5	0.43976	V12	0.69456	V8	0.84269	V8
4	0.50785	V11	0.73928	V11	0.87322	V3
3	0.59016	V3	0.80765	V3	0.96332	V1
2	0.74545	V10	0.88978	V10	1.00000	V11
		V13, V14		V13, V14		V13, V14

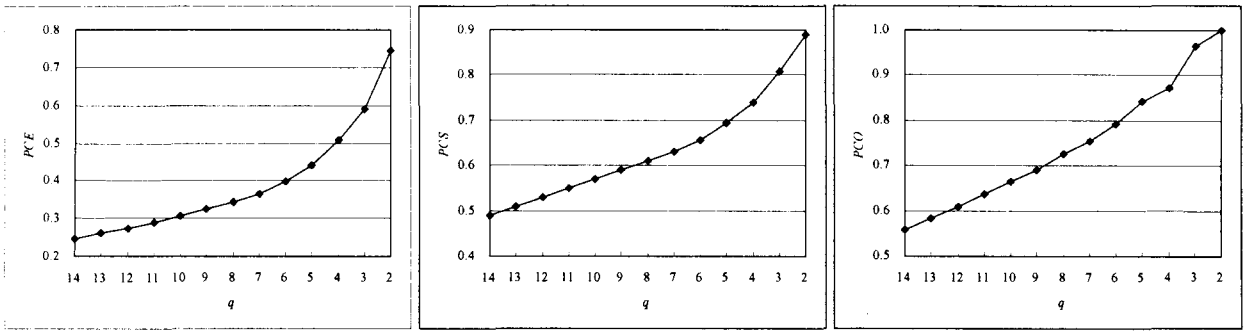


図1 q による *PCE*, *PCS*, *PCO* の変化（授業アンケートデータの 14 変数，変数減少法，$r=2$）

から，これらの質問項目は，教師の評価に対する寄与が低いということになる。さらに，3 変数を削除した $q=8$ のところで見ると，{V1} と {V12} が入れ替わっているが，{V2, V4, V5, V7, V9} の 5 変数が共通で削除されたことがわかる。一方，最後まで含まれていた変数は V13 と V14 であった。

次に，真の選択変数群として，総当たり法の結果を示す。表 3〜5 の「$Y_1 \mid Y_2$」の列が総当たり法により選ばれた変数群と削除された変数群で，「AllPossible」の列がそのときの規準値である。

同表では，変数減少法，変数増加法，変数減増法，変数増減法の 4 手順がその真の選択結果とどの程度異なるかも示している。すなわち，「Back」〜「For-Back」の各列に，各手順による規準値と総当たり法による規準値の差を示した（差のあるセルに網掛けをしてある）。差が 0 であるということは，総当たり法と同じ規準値を示したということだけでなく，規準値は Y_1 から算出されるので，規準値が同じならば，選ばれた変数群も総当たり法と同じであることが容易に想像される（なお，PCO 規準に関しては，$q=2$ のときのみ，規準値は一致するが選択された変数は異なっている）。これらの表から，3 つの規準とも，変数減増法（Back-For）と変数増減法（For-Back）では総当たり法と同じ結果が得られること，変数減少法（Back）は，一部総当たり法と異なる変数を選んでいること，変数増加法（For）については，PCE 規準と PCS 規準では違いは少しであるが，PCO 規準では異なる部分が多いことがわかる。これらのことから，最も時間を節約するには変数減少法を用いることができるが，真の変数群に近い変数群を得るには Stepwise 系の変数減増法か変数増減法を用いればよいことが示唆される。主成分分析における変数選択の数値的検討でも変数増減法が総当たり法に最も近い結果が得られることがわかっている（森 他, 1999）。

最後に，$q=9$ の場合で，元の変数と選択された変数によるコレスポンデンス分析の結果を個体スコアの布置で比較してみる。図 2(a) は，全 14 変数を用いたときの個体スコアの 1 軸と 2 軸のプロット，図 2(b), (c)は，それぞれ PCE/PCS 規準により選択された 9 変数 {V2, V3, V6, V8, V10, V11, V12, V13, V14} と PCO 規準により選択された 9 変数 {V1, V3, V5, V6, V8, V10, V11, V13, V14} による個体スコアのプロットである。(a)−(b), (a)−(c)を見ると，それほど大きな違いは見られず，個体スコアに関しては，選択された変数でも元の変数と同等の考察が可能であることがわかる。なお，(a)と(b)および(c)との布置の近さを *RV* 係数 (Robert and Escoufier, 1976) で評価すると，*RV*[(a), (b)]=0.9093, *RV*[(a), (c)]=0.8611 となり，PCE/PCS 規準により選ばれた変数による個体スコアの布置が PCO 規準のそれより元の全変数による布置に近いことがわかる。q が 9 以外の場合でも同様であり，個体スコアの布置に関しては，PCE/PCS 規準の方が良い変数群を選んでいることがわかる。

3.2 軽症意識障害データ

データは，表 6 に示すように 25 項目からなる検査で，各問 4 肢択一または 2 肢択一である（佐野 他, 1977）。このうち，佐野 他（1977）などの先行研究にしたがって，全体的な構造への寄与が少ないとされる 21 番と 22 番を除いた 23 項目に対して，変数選択を実施する。以下，23 番以降の変数番号は 2 を引いて表記（旧 23＝新 21，旧 24＝新 22，旧 25＝新 23）する。

このデータに，簡便法として最も良い結果が得られる変数増減法を用いて選択を行い，先行研究との比較を考慮し，$q=10$ の前後の結果を表 7 に表した。

表3　PCE規準による総当たり法の結果（規準値と選択・削除変数）と4選択手順の比較（授業アンケートデータ）

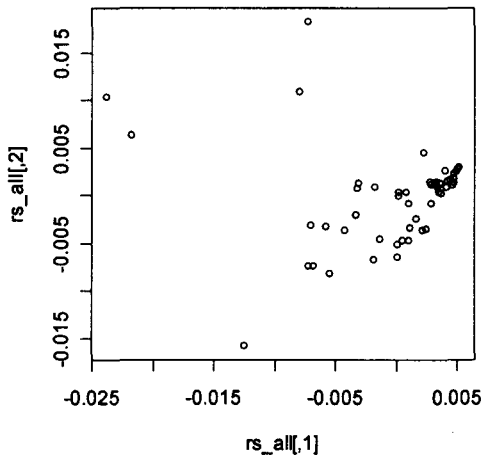
q	*PCE*					Y_1 | Y_2													
	Back	For	Back-For	For-Back	AllPossible														
14	0.00000	0.00000	0.00000	0.00000	0.24669	1	2	3	4	5	6	7	8	9	10	11	12	13	14 |
13	0.00000	0.00000	0.00000	0.00000	0.26026	1	2	3	4	5	6	8	9	10	11	12	13	14	| 7
12	0.00000	0.00000	0.00000	0.00000	0.27457	1	2	3	4	5	6	8	10	11	12	13	14	| 7	9
11	0.00000	0.00000	0.00000	0.00000	0.28891	1	2	3	5	6	8	10	11	12	13	14	| 4	7	9
10	0.00000	0.00000	0.00000	0.00000	0.30658	1	2	3	6	8	10	11	12	13	14	| 4	5	7	9
9	0.00000	0.00026	0.00000	0.00000	0.32338	2	3	6	8	10	11	12	13	14	| 1	4	5	7	9
8	0.00127	0.00000	0.00000	0.00000	0.34303	1	3	6	10	11	12	13	14	| 2	4	5	7	8	9
7	0.00273	0.00000	0.00000	0.00000	0.36870	1	3	10	11	12	13	14	| 2	4	5	6	7	8	9
6	0.00632	0.00000	0.00000	0.00000	0.40310	1	3	10	11	13	14	| 2	4	5	6	7	8	9	12
5	0.00000	0.00000	0.00000	0.00000	0.43976	3	10	11	13	14	| 1	2	4	5	6	7	8	9	12
4	0.00000	0.00000	0.00000	0.00000	0.50785	3	10	13	14	| 1	2	4	5	6	7	8	9	11	12
3	0.00000	0.00000	0.00000	0.00000	0.59016	10	13	14	| 1	2	3	4	5	6	7	8	9	11	12
2	0.00000	0.00000	0.00000	0.00000	0.74545	13	14	| 1	2	3	4	5	6	7	8	9	10	11	12

表4　PCS規準による総当たり法の結果（規準値と選択・削除変数）と4選択手順の比較（授業アンケートデータ）

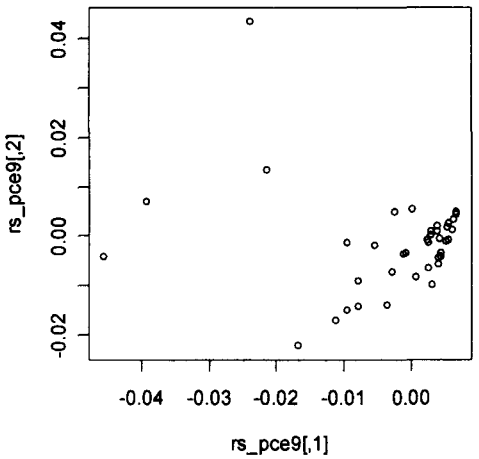
q	*PCS*					Y_1 | Y_2													
	Back	For	Back-For	For-Back	AllPossible														
14	0.00000	0.00000	0.00000	0.00000	0.48825	1	2	3	4	5	6	7	8	9	10	11	12	13	14 |
13	0.00000	0.00000	0.00000	0.00000	0.50805	1	2	3	4	5	6	8	9	10	11	12	13	14	| 7
12	0.00000	0.00000	0.00000	0.00000	0.53011	1	2	3	4	5	6	8	10	11	12	13	14	| 7	9
11	0.00000	0.00000	0.00000	0.00000	0.54996	1	2	3	5	6	8	10	11	12	13	14	| 4	7	9
10	0.00000	0.00000	0.00000	0.00000	0.57033	1	2	3	6	8	10	11	12	13	14	| 4	5	7	9
9	0.00000	0.00006	0.00000	0.00000	0.59103	2	3	6	8	10	11	12	13	14	| 1	4	5	7	9
8	0.00000	0.00253	0.00000	0.00000	0.61166	3	6	8	10	11	12	13	14	| 1	2	4	5	7	9
7	0.00287	0.00000	0.00000	0.00000	0.63473	1	3	8	10	11	13	14	| 2	4	5	6	7	9	12
6	0.00700	0.00000	0.00000	0.00000	0.66416	1	3	10	11	13	14	| 2	4	5	6	7	8	9	12
5	0.00000	0.00000	0.00000	0.00000	0.69456	3	10	11	13	14	| 1	2	4	5	6	7	8	9	12
4	0.00000	0.00000	0.00000	0.00000	0.73928	3	10	13	14	| 1	2	4	5	6	7	8	9	11	12
3	0.00000	0.00000	0.00000	0.00000	0.80765	10	13	14	| 1	2	3	4	5	6	7	8	9	11	12
2	0.00000	0.00000	0.00000	0.00000	0.88978	13	14	| 1	2	3	4	5	6	7	8	9	10	11	12

表5　PCO規準による総当たり法の結果（規準値と選択・削除変数）と4選択手順の比較（授業アンケートデータ）

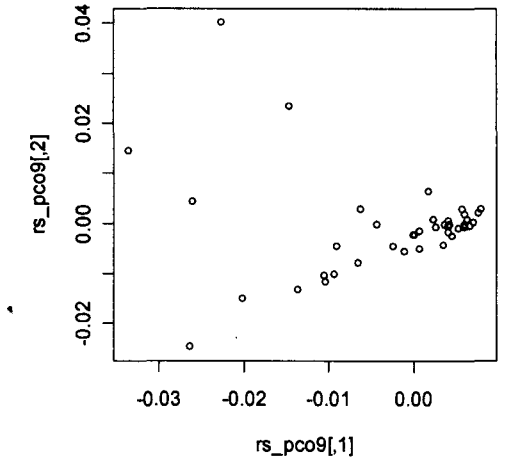
q	*PCO*					Y_1 | Y_2													
	Back	For	Back-For	For-Back	AllPossible														
14	0.00000	0.00000	0.00000	0.00000	0.55856	1	2	3	4	5	6	7	8	9	10	11	12	13	14 |
13	0.00000	0.00229	0.00000	0.00000	0.58386	1	2	3	4	5	6	7	8	10	11	12	13	14	| 9
12	0.00000	0.02280	0.00000	0.00000	0.60969	1	2	3	4	5	6	8	10	11	12	13	14	| 7	9
11	0.00000	0.03295	0.00000	0.00000	0.63615	1	2	3	5	6	8	10	11	12	13	14	| 4	7	9
10	0.00000	0.03537	0.00000	0.00000	0.66593	1	3	5	6	8	10	11	12	13	14	| 2	4	7	9
9	0.00000	0.03429	0.00000	0.00000	0.69035	1	3	5	6	8	10	11	13	14	| 2	4	7	9	12
8	0.00000	0.07804	0.00000	0.00000	0.72477	1	3	6	8	10	11	13	14	| 2	4	5	7	9	12
7	0.00000	0.08525	0.00000	0.00000	0.75446	1	3	8	10	11	13	14	| 2	4	5	6	7	9	12
6	0.00000	0.08241	0.00000	0.00000	0.79238	1	3	8	11	13	14	| 2	4	5	6	7	9	10	12
5	0.00000	0.10632	0.00000	0.00000	0.84269	1	3	11	13	14	| 2	4	5	6	7	8	9	10	12
4	0.03302	0.09283	0.00000	0.00000	0.90624	10	11	13	14	| 1	2	3	4	5	6	7	8	9	12
3	0.00915	0.07263	0.00000	0.00000	0.97247	10	13	14	| 1	2	3	4	5	6	7	8	9	11	12
2	0.00000	0.00000	0.00000	0.00000	1.00000	6	10	| 1	2	3	4	5	7	8	9	11	12	13	14



(a) 全 14 変数



(b) PCE/PCS 規準で選ばれた 9 変数
{V2, V3, V6, V8, V10, V11, V12, V13, V14}



(c) PCO 規準で選ばれた 9 変数
{V1, V3, V5, V6, V8, V10, V11, V13, V14}

図 2 個体スコアの散布図（授業アンケートデータ）

このデータは，すでに，佐野 他（1977, 1979），Tanaka and Mori（1997），Iizuka et al.（2003）などで数値例として変数選択が行われている。佐野 他（1977, 1979）では，データを計量データとして扱い，全 25 変数の因子負荷量をプロットし，{V4, V5, V6, V11, V14, V22}，{V7, V8, V9, V10, V13}，{V1, V2, V3, V12, V15, V16, V18, V19, V20}，{V21, V23} と，寄与のない旧 21 番と旧 22 番の計 5 つのクラスターのうち，前の 3 つを主要なクラスターとして分析を行っている。また，同時に検査した「意識レベル」を目的変数として通常の回帰分析による変数選択を行い，{V1, V4, V8, V14, V19, V21} の 6 変数を選択したり，因子分析の共通性に基づいて，{V3, V4, V5, V8, V10, V11, V14, V15, V17, V18, V19, V20} の 12 変数を選択したりしている。一方，Tanaka and Mori（1997）と Iizuka et al.（2003）でも同様に計量データとして扱い，拡張主成分分析の寄与率（Tanaka and Mori, 1997）を選択規準として用いた変数選択を行い，前者では，1 つの選択結果として 13 変数 {V1, V3, V4, V5, V6, V7, V9, V11, V12, V16, V17, V19, V22} を考察し，後者では，最低必要な変数の数として 6 を示唆している。今回の選択は，佐野 他（1977, 1979）や Tanaka and Mori（1997）

表6　軽症意識障害検査の検査項目

	項　目		項　目		項　目		項　目		項　目
1	食　事	6	見当識（月日）	11	知　識	16	表　情	21	抑制欠如（多弁）
2	尿失禁	7	見当識（時間）	12	命令への反応	17	診察中の態度	22	抑制欠如（発動）
3	呼名・挨拶への反応	8	見当識（人）	13	1から20まで数える	18	自発動作	23	保持傾向
4	見当識（場所）	9	自分の病識の程度	14	計算力	19	自発発語	24	生年月日が言える
5	見当識（季節）	10	意　欲	15	声の調子	20	注　意	25	名前が言える

と同じ変数を選ぶことが目的ではないので，単純に比較することはできないが，参考として，PCE規準とPCS規準について，佐野 他（1977, 1979）による因子負荷量のクラスターと表7の結果を照らし合わせてみると，どのqにおいても，その主要な3つのクラスターから変数が選ばれており，さらに，もう1つのクラスターである {V21, V23} からも選ばれていることがわかる。また，佐野 他（1977, 1979）による12変数やTanaka and Mori（1997）による13変数と比較すると，PCE規準とPCS規準では，5つの変数は異なっているが他は同じであることがわかり，先行研究の規準と比較してもそれほど大きく変わらないということが参考としていえる。

ここで，あらためて表7に注目すると，{V5, V6, V7, V20, V22} は，どの規準においてもほとんどのqで選ばれており，寄与が高いことがわかるが，$q<10$ では，{V8, V9, V11} と {V15, V17, V19} が選ばれるかどうかで，PCE／PCS規準とPCO規準の結果に差が出ている。全体としては，授業アンケートデータと同様に，PCE規準とPCS規準では選ばれた変数の違いが少なく，PCO規準ではそれらとやや異なる変数が選ばれていることがみてとれる。

4. おわりに

外的変量をもたない多変量解析手法であるコレスポンデンス分析における変数選択として，3つの適合度規準を選択規準として利用し，実データに適用して検討を行った。その結果，選択結果の解釈や選択過程を考察することができた。これらの規準を選択規準として採用するかどうかについては，コレスポンデンス分析の変数選択において考えられる他の規準との比較を待たなくてはならないが，規準としての大きな問題はないといえる。ただし，今回の手法は，規準の算出にあたって，基とすべき全p変数を使った規準値と比較をしているのではなく，利用する変数の数qを決めたときに，最も良いGOFを提供するq変数を求める方法をとっている。したがって，全p変数が提供する結果にいかに近づけるかとか，$q+1$ 変数や $q-1$ 変数と逐次的な比較を行うといった選択

表7　選択された変数群（軽症意識障害検査，変数増減法，$r=3$）

q	規　準	選択された変数
14	*PCE*	3 5 6 7 8 10 12 15 17 18 19 20 22 23
	PCS	3 5 6 7 8 10 12 15 17 18 19 20 22 23
	PCO	2 3 5 6 7 10 11 12 13 15 17 19 20 22
13	*PCE*	3 5 6 7 8 12 15 17 18 19 20 22 23
	PCS	3 5 6 7 8 12 15 17 18 19 20 22 23
	PCO	2 3 5 7 10 11 12 13 15 17 19 20 22
12	*PCE*	3 5 6 7 12 15 17 18 19 20 22 23
	PCS	3 5 6 7 8 12 15 17 19 20 22 23
	PCO	2 3 5 7 10 11 12 13 15 17 19 20
11	*PCE*	3 5 6 7 15 17 18 19 20 22 23
	PCS	3 5 6 7 12 15 17 19 20 22 23
	PCO	2 3 5 7 10 12 13 15 17 19 20
10	*PCE*	3 5 6 7 15 17 19 20 22 23
	PCS	3 5 6 7 15 17 19 20 22 23
	PCO	4 5 6 7 8 9 11 18 20 22
9	*PCE*	5 6 7 15 17 19 20 22 23
	PCS	5 6 7 15 17 19 20 22 23
	PCO	5 6 7 8 9 11 18 20 22
8	*PCE*	5 6 7 17 19 20 22 23
	PCS	5 6 7 15 17 19 20 22
	PCO	5 6 7 8 9 11 20 22
7	*PCE*	5 6 7 19 20 22 23
	PCS	5 6 7 8 10 22 23
	PCO	5 6 7 9 11 20 22
6	*PCE*	5 6 7 8 22 23
	PCS	5 6 7 8 22 23
	PCO	5 6 7 9 11 22

手法ではないので，これらの観点で比較が行える規準の考案も今後は必要になろう。

一方，今回特に触れなかったが，たとえば，授業アンケートデータでは，寄与が高いものとしてV13とV14が最後まで残っている。それらの質問文は「教員は熱心に教えてくれたと思いますか」と「総合的に見て，この授業を履修してよかったですか」という総合的な評価をうながす文面であることから，変数選択手法は，単に変数を選ぶだけでなく，その選択過程から，質問が意味のあるものかどうかなどを検討する観点を提供するものともいえる。特にアンケート調査などでは，冗長な質問や似た性質をもった質問などは避けるべきであり，この観点での研究も今後進めていきたい。

コレスポンデンス分析における変数選択については，各種のデータや選択後の変数での再調査を行ったりすることで，評価を加えていくとともに，計算環境としてのVASMM（コレスポンデンス分析における変数選択VAScorresについては，`http://mo161.soci.ous.ac.jp/vascorres/index.html`）への規準の実装も考えていく予定である。

参考文献

Adachi, K. (2004). Correct classification rates in multiple correspondence analyss, *J. Japanese Soc. Comp. Statist.*, **17**, 1-20.

Greenacre, M. (1994). Multiple and joint correspondence analysis. In Greenacre, M. and Blasius, J. (eds), *Correspondence analysis in the social sciences*, 141-161, Academic Press.

Iizuka, M., Mori, Y., Tarumi, T. and Tanaka, Y. (2002a). Statistical software VASMM for variable selection in multivariate methods. In Härdle, W. and Rönz, B. (eds), *COMPSTAT2002 Proceedings in Computational Statistics*, 563-568, Physica-Verlag.

Iizuka,M., Mori,Y., Tarumi,T. and Tanaka,Y.(2002b). Some New Modules in Variable Selection Software VASMM. *Proceedings of the 4th ARS Conference of the IASC*, 166-169.

Iizuka, M., Mori, Y., Tarumi, T. and Tanaka, Y. (2003). Computer intensive trials to determine the number of variables in PCA. *Journal of the Japanese Society of Computational Statistics: Special Issue of ICNCB*, **14**(2): 337-345.

Krzanowski, W. J. (1987). Selection of variables to preserve multivariate data structure, using principal components. *Appl. Statist.*, **36**, 22-33.

Mori,Y. (1998). Statistical Software VASPCA - Variable Selection in PCA -. 岡山理科大学紀要, **33** (A), 329-340.

Mori, Y. and Maehashi, A .(1996). Variable Selection for Categorical Data： A Numerical Investigation on Fatigue Data. *Bulletin of Kurashiki City College*, **26**, 25-38.

Mori, Y., Iizuka, M., Tarumi, T. and Tanaka, Y. (2000a). Statistical software "VASPCA" for variable selection in principal component analysis. *The 14th Symposium on Computational Statistics*, Short Communications, 73-74.

Mori, Y., Iizuka, M., Tarumi, T. and Tanaka, Y. (2000b). Study of variable selection criteria in data analysis. *Proceedings of the Tenth Japan and Korea Joint Conference of Statistics*, 119-124.

Robert, P. and Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: the RV-coefficient. *Appl. Statist.*, **25**, 257-65.

Tanaka, Y. and Kodake, K. (1981). A method of variable selection in factor analysis and its numerical investigation. *Behaviormetrika*, **10**, 49-61.

Tanaka, Y and Mori, Y. (1997). Principal component analysis based on a subset of variables: Variable selection and sensitivity analysis. Amer. *J. Mathematical and Management Sciences*, **17**, 1 & 2, 61-89.

Xia, L. and Yang, Y. (1988). A Method of Variable Selection in Hayashi's Third Method of Quantification. *J. Japanese Soc. Comp. Statist.*, **1**, 27-43.

飯塚誠也，森　裕一，垂水共之，田中　豊 (2001). 主成分分析における変数選択プログラムのWWWへの実装．文部科学省統計数理研究所「統計数理」, **49**(2): 277-292.

飯塚誠也，森　裕一，垂水共之，田中　豊 (2002). 因子分析における変数選択規準の考察．日本計算機統計学会第16回大会論文集，68-71.

佐野圭司 他 (1977). 軽症意識障害の評価方法に関する統計的研究－断面調査による特徴的臨床像の抽出．神経進歩, **21**, 1052-65.

佐野圭司 他 (1979). 軽症意識障害の評価方法に関する統計的研究－経時調査および項目選択．神経進歩, **23**(6), 1207-18.

森　裕一，飯塚誠也 (2002). 主成分分析における変数選択手法の考察．岡山理科大学紀要, **38**(A): 105-112.

森　裕一，飯塚誠也，垂水共之，田中　豊 (2000). 変数の影響分析を利用した変数選択．日本行動計量学会第28回大会論文集, 301-302.

森　裕一，垂水共之，田中　豊 (1999). 変数の一部に基づく主成分分析：変数選択手法の数値的検討．日本計算機統計学会「計算機統計学」, **11**(1): 1-12.

森　裕一，杜　暁東，飯塚誠也 (2004). コレスポンデンス分析における変数選択規準の検討．日本計算機統計学会第18回シンポジウム論文集，9-12.