岡山大学環境理工学部研究報告
第12巻，第1号，pp.29-40，2007年3月

# 合成変数の推定を利用した項目選択とその数値的検討[1]

森　裕一[2]　笛田　薫[3]　飯塚誠也[3]

# Variable selection besed on global score estimation and its numerical investigation

Yuichi Mori[2], Kaoru Fueda[3] and Masaya Iizuka[3]

(Received November 30, 2006)

abstract

A variable selection method using global score estimation is proposed, which is applicable as a selection criterion in any multivariate method without external variables such as principal component analysis. This method selects a reasonable subset of variables so that the global scores, e.g. principal component scores, which are computed based on the selected variables, approximate the original global scores as well as possible in the context of the least squares. Three computational steps are proposed to estimate the scores according to how to satisfy the restriction that the estimated global scores are mutually uncorrelated. Three different examples are analyzed to demonstrate the performance and usefulness of the proposed method numerically, in which three steps are evaluated and the results obtained using four cost-saving selection procedures are compared.

*Key words*: principal components, least square, orthogonalization, cost-saving selection.

## 1　はじめに

たとえば，予備調査では，重要な次元を拾い上げるために多数の調査項目を設定し，主成分分析によって次元を縮約していたが，本調査では，できるだけ調査項目を減らし，かつ全項目を調査した場合の主成分得点と同様の値を得たい場合がある。あるいは，継続的な調査において，調査項目は減らしたいが，調査の継続性は失われないようにしたい，つまり前回の調査において，今回調査する項目だけしか調査していなかったとしても，前回の調査結果が大きく変わらないように今回の調査項目を選択したいという場合もある。このような場合には，重回帰分析における応答変数のような外的変量がある変数選択手法は直接的には利用できないため，外的変量を特定しない多変量手法に特化した何らかの変数選択手法を考える必要が出てくる。

外的変量を特定しない多変量手法における変数選択は，これまでにもさまざまな手法や選択規準が提案されている。たとえば，主成分分析では，選ばれた変数から作られる主成分が元の全変数を最もよく予測するように変数群を選ぶ，あるいは，因子分析では，元の変数から算出される因子得点と選ばれた変数から算出される因子得点のそれぞれの空間上の相対的な布置が最も近くなるような変数群を選ぶ，といったことである。そのいくつかについて

[1]この報告は，笛田 他 (2005)，森 他 (2005a)，森 他 (2005b) を基に，笛田 他 (2003) と Mori et al. (2004) の計算過程をまとめ，新たな数値例を加える形で改訂したものである。

[2]岡山理科大学総合情報学部
[3]岡山大学大学院環境学研究科

は，http://mo161.soci.ous.ac.jp/vasmm/ に詳しくまとめられている。

これら既存の選択手法と選択規準は，たとえ同じ多変量手法での選択であっても，統一的な外的変量がないため，それぞれで異なる結果(異なる変数群)を提供する。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選択規準で選択を行ったり，検討が必要な場合はいくつかの手法を試して結果を比較したりすることが必要になる。また，実際に選択を実行するにあたって，総当たり法は時間がかかることから，変数減少法などの簡便法を用いるが，先行研究の多くは，変数減少法でしか選択を行っておらず，変数増加法やステップワイズな選択手順（変数減増法や変数増減法）などによる結果も検討したいところである。これらのことを考えると，提案されているどの規準でも変数選択が行え，しかも複数の選択手順が用意されているような計算環境があれば，選択結果が異なるという特性をカバーでき，さまざまな規準や選択手順を試すことができるので，非常に便利である。

このような背景から，外的変量を特定しない多変量解析における変数選択を行うための計算環境が1994年から提供されている。現在では，WebサイトVSAMM(Variable Selection in Multivariate Methods, http://mo161.soci.ous.ac.jp/vasmm/)で，変数選択に関連する情報とともに，Web上のオンライン解析と既存の数学・統計パッケージのマクロとして，その計算環境を提供している (Iizuka et al., 2002, など)。

一方，選択場面の目的に応じて，さまざまな選択規準が考案できるので，それぞれの場面に適した選択規準を提案・検討していくことも変数選択の研究として重要なことである。

そこで，本研究では，最初に述べたように，調査結果が主成分得点のような何らかの合成変数で評価されている継続的な調査において，調査の継続性を失わないように調査項目を選び出す場面での変数選択を考える。このような場面に対して，全調査項目から求めた合成変数を，選択された一部の調査項目のみを用いて近似する手法が笛田 他 (2003) で提唱され，それを変数選択の規準として利用する試みが笛田 他 (2003) で提案されている。この指標の性能について，数値的な検討を行うのがここでの目的である。すなわち，この一部の項目で元の合成変数を推定する手法には，推定された合成変数がもつ無相関性をどの程度考慮するかによって3つの計算ステップが用意されているが，この近似の精度が異なる3つのステップによって変数選択の結果にどの程度の差を生じるかを評価する。また，計算コストを削減するために森 他 (1988) で提唱されている変数減少法，変数増加法，変数減増法，変数増減法の4つの選択簡便法についても，今回の規準に対して適用したときの有効性も検討する。これらを通じて，合成変数の推定を選択規準に利用した変数選択の実際的利用の性能や有効性を検討していく。なお，計算環境として，VSAMMのオンライン解析およびRの関数に，今回の合成変数の推定に基づく変数選択規準を実装したので，その概要についても付録に記しておく。

## 2 合成変数推定を利用した変数選択規準

### 2.1 一部の変数による合成変数の推定

まず，予備調査あるいは前回の調査において観測された変数を $Y = \{\mathbf{y}_1, \ldots, \mathbf{y}_p\}$ とし，このデータに基づき，通常の外的変量を特定しない多変量手法により $r$ 個の次元に縮約された合成変数 $Z = \{\mathbf{z}_1, \ldots, \mathbf{z}_r\}$ $(1 \leq r \leq p)$ を得ていたとする。この合成変数とは，たとえば，多変量手法が主成分分析の場合は主成分得点であり，因子分析やコレスポンデンス分析ならば，それぞれ因子得点や個体スコアのことである。本調査あるいは今回の調査において，調査項目数を $q$ 個 $(r \leq q < p)$ に抑えたいならば，$Y$ の中の $q$ 個の変数 $Y_1 = \{\mathbf{y}_{1'}, \ldots, \mathbf{y}_{q'}\}$ ($1', \ldots, q'$ は $1, \ldots, p$ のうちの $q$ 個を示す) のみを観測し，$Z$ を近似する $\hat{Z} = \{\hat{\mathbf{z}}_1, \ldots, \hat{\mathbf{z}}_r\}$ を構成したい。ただし，合成変数 $\mathbf{z}_1, \ldots, \mathbf{z}_r$ は，一般には互いに

無相関であるため，$\hat{\mathbf{z}}_1,\ldots,\hat{\mathbf{z}}_r$ も互いに無相関である必要がある。この条件を制約として合成変数を推定するには，次の3つのステップを経ることになる。計算の詳細は，笛田 他 (2003) や Mori et al. (2004) を参照されたい。

**ステップ1**　選択した $q$ 個の変数 $Y_1$ のみに着目し，全変数 $Y$ から求めた合成変数 $Z = \mathbf{z}_1,\ldots,\mathbf{z}_r$ を従属変数，$Y_1 = \{\mathbf{y}_{1'},\ldots,\mathbf{y}_{q'}\}$ を独立変数とみなし，一般的な重回帰と同様の方法で，$\mathbf{z}_1$ の推定値 $\hat{\mathbf{z}}_1$ から $\mathbf{z}_r$ の推定値 $\hat{\mathbf{z}}_r$ までを構成する。すなわち，$Y_1 = \{\mathbf{y}_{1'},\ldots,\mathbf{y}_{q'}\}$ に対する係数行列を，最小二乗の意味で，合成変数 $Z$ と $Y_1$ による推定値 $\hat{Z}$ との誤差平方和を最小にするものとして求める。この解を得る段階をステップ1とする。この方法は，推定値 $\hat{\mathbf{z}}_1,\ldots,\hat{\mathbf{z}}_r$ を個別に求めているだけで，推定された合成変数間の相関は考慮されていないので，$\mathbf{z}_1,\ldots,\mathbf{z}_r$ が無相関であっても，その推定値 $\hat{\mathbf{z}}_1,\ldots,\hat{\mathbf{z}}_r$ は，一般に無相関であることは保証されない。したがって，$r = 1$ の場合，あるいは，推定された合成変数間の無相関性が不要な場合は，計算はこれで終わる。

**ステップ2**　上記の通り，ステップ1では，推定値 $\hat{\mathbf{z}}_1,\ldots,\hat{\mathbf{z}}_r$ 間の無相関性は考慮されていないので，$r \geq 2$ の場合で，無相関性を考慮する必要がある場合は，さらに計算が必要である。そこで，たとえば，主成分分析においては，第1主成分が最も大きいため，誤差平方和を小さくするためには，第1主成分の推定値を最優先することが自然である。したがって，元の合成変数の重要さが $\mathbf{z}_1,\ldots,\mathbf{z}_r$ の順である場合，最も重要な $\mathbf{z}_1$ の推定値 $\hat{\mathbf{z}}_1$ を，ステップ1と同様に求め，2番目に重要な $\mathbf{z}_2$ の推定値 $\hat{\mathbf{z}}_2$ は，$\hat{\mathbf{z}}_1$ と無相関という制約の下で $\mathbf{z}_2$ を最もよく近似するもの，$\hat{\mathbf{z}}_3$ は，$\hat{\mathbf{z}}_1$ と $\hat{\mathbf{z}}_2$ と無相関という制約の下で $\mathbf{z}_3$ を最もよく近似するもの，$\cdots$，という順で，$\hat{\mathbf{z}}_r$ まで求めれば，合成変数の重要性に基づいた無相関性をもった推定値が求められる。実際には，これらは，ステップ1で求めた推定値に Gram-Schmidt の直交化を施すことで求めることができる。

**ステップ3**　ステップ2では，推定値が互いに無相関にはなっているが，$\mathbf{z}_1$ が最も重要であるという条件によって推定値が求められている。この条件は，無相関ということに対しては，本来要求されないことであるので，これをはずして，推定値を推定することをここで行う。すなわち，互いに無相関であるという制約のみでの推定値を求めるものである。具体的には，ステップ2で求めた推定値を無相関という関係を保ったまま，誤差平方和を小さくする方向へ Givens 変換を用いて回転させる。

## 2.2　合成変数推定を利用した変数選択

上記のステップは，どんな $q$ 個の変数が与えられても行うことができるが，$Z$ の推定値として最適なものがほしい場合は，${}_p\mathrm{C}_q$ 個の部分群のうち，$\mathbf{z}_1,\ldots,\mathbf{z}_r$ を近似したときの誤差平方和を最も小さくする $q$ 個の変数群 $\{\mathbf{y}_{1''},\ldots,\mathbf{y}_{q''}\}$ ($1'',\ldots,q''$ は $1,\ldots,p$ のうちの $q$ 個を示す) が最適な変数群であるということになる。ここで変数選択が実質的に行われたことになる。したがって，誤差平方和を最小にする $q$ 個の組合せを観測すべき変数として選択することで，この合成変数の推定が1つの変数選択の規準となりうることがわかる (笛田 他, 2003; Mori et al., 2004)。

そこで，これを規準として変数選択を行うことを考えるが，一般に，ステップ1，2は，直接法により推定値を求めることができるものの，ステップ3は，$r$ 個の推定値の中の2個ずつを選んで回転させることを反復するため，特に $r$ が大きいときには，収束までに時間がかかる。また，ステップ2の直交化も，$r$ が大きいときには時間がかかる。本来は，ステップ3まで経た推定値の誤差平方和を求めて，それを最小とする変数の組合せを選択すべきであるが，$q$ 個の変数の組合せすべてに関して推定値を求め，誤差平方和の小ささを比較する場合は，計算内の反復で行われる組合せの数 (${}_rC_2$) に加え，$p$ 変数から $q$ 変数を選ぶ変数の組合せの数 (${}_pC_q$) により，計算の数が非常に多くなる。そこで，簡便法として，ステップ1あるいはステップ2で推定値の計算を打ち切り (${}_rC_2$ の反復を避け),

誤差平方和を求めて，それを最小とする変数の組合せを選択する方法が考えられる。このように計算の詳しさによって計算時間を節約する方法は，本規準のように計算過程がいくつかのステップに分けられるものに採用が可能である。それでも，計算時間の短縮がむずかしい場合は，次にあげる変数減少法などの選択簡便法と組み合わせる（${}_pC_q$ の反復を避ける）方法をとることになる。

## 3 変数選択の簡便法

元の $p$ 個の変数の中から $q$ 個の変数を選ぶとき，変数の数が $q$ 個であるすべての変数の組合せの中から，ある規準の最大値（あるいは，最小値）をもつ変数群を選ぶことができればよいが，${}_pC_q$ 個すべての組合せを調べることは計算コストが高いので，次の4つの逐次選択法が提案されている (森 他, 1988)。

a. 変数減少法 (Backward elimination)

b. 変数増加法 (Forward selection)

c. 変数減増法 (Backward-forward stepwise selection)

d. 変数増減法 (Forward-backward stepwise selection)

それぞれ，1つずつ変数を落としたり足したりしながら，選択を行っていくものである。これらの手順によって，よりよい変数群が，自動的に選択される。

森 他 (1988) では，拡張主成分分析の規準を利用する変数選択手法において，すべての組合せの中から最適な変数群を探す総当たり法と上記の4つの簡便法を比較して，その差がそれほど大きくないことが数値的な検討によって示されている。また，同時に，単純系選択手順（a と b）よりステップワイズ系選択手順（c と d）の方が，後退系選択手順（a と c）より前進系選択手順（b と d）の方が，それぞれ性能がよいことも示されている。

## 4 数値例

ここでは，環境経営度調査，新国民生活指標，社会生活基本調査 (生活時間) の各データに対して，2節の合成変数推定を利用した変数選択を適用し，提案の手法の評価を行う。具体的には，環境経営度調査データで，調査そのものの評価を行い，新国民生活指標データでは，ステップの評価，すなわち，ステップ1，2の有効性を考察し，社会生活基本調査データでは，3つのステップのそれぞれに3節の4つの選択手順を適用したときの結果を評価する。いずれも元の解析では，得られたデータを主成分分析にかけ，その主成分得点で考察を行っているので，本数値例でも，合成変数を主成分得点として，その推定を一部の変数で行うことを規準にして変数選択を行う。なお，環境経営度調査データは，笛田 他 (2005)，森 他 (2005a)，森 他 (2005b) で，$r=1$ の場合の数値例として用いているが，本報告の目的から，その結果を再掲するとともに，$r=2$ の場合を追加して示すことにする。

### 4.1 環境経営度調査

環境経営度調査は，日本経済新聞社により，1997年から毎年行われているもので，企業へ複数項目のアンケートを行い，その第1主成分を環境経営度とし，調査企業をランキングしたり，第1主成分と第2主成分で散布図を描き，それを環境影響度マップとして企業間の関係などを考察している。

アンケート項目は，毎年，新たに注目されるようになったものが加えられたり，不要となったものは削除されたりしていくため，異なる年度の調査結果を単純比較はできない。しかしながら，そのような調査方法に目を向けないまま結果だけが一人歩きし，異なる年度の調査結果が単純に比較されてしまうこともある。

そこで，もし前年度の結果と比較されてしまう場合，「前年度からの削除項目の選定は適切であったか」，言い換えると「前年度から削除されずに残された項目は，前年度の結果を有効に表していたか」という視点で，利用されている

調査項目を検証してみる。すなわち，合成変数が用いられ，かつ過去から現在にかけて調査項目が落とされていることから，合成変数の推定を利用した変数選択の典型的な例と考えられるので，数値的な検討のために，本データを利用することにする。初期の調査結果は残っていないため，第4回調査から第6回調査までについて調べた (表1)。

### 4.1.1 環境経営度指標

第4回調査と第5回調査では調査項目の名前がすべて変更されている。しかし，第4回の「4.$CO_2$ 対策」と第5回の「7. 温暖化対策」のように，名称が変わっただけのものも多く，実質的に削除されたのは，{1. リデュース，3. 廃棄物管理，5. 化学物質管理，8. 組織制度} の4つであると考えられる (表2)。そこで，第4回調査の11項目の中から4項目を削除するすべての組合せについて，残った項目で第4回の環境経営度，つまり，第1主成分の近似を行い，誤差平方和の大きさをみた。なお，第4回調査の相関行列による主成分の寄与率は，第1主成分0.6802，第2主成分0.0760，第3主成分0.0459，... である。また，分析に第1主成分のみを用いるので，主成分間の無相関性を考慮することはなく，したがって，前節のステップ2，3は不要である。

表3は，11項目から7項目を残す全330通りの組合せの一部を誤差平方和の小さい方から順に並べたものである。11項目から選択した7項目の中で，元の11項目による第1主成分を近似して最も誤差平方和が小さかった組合せは，{2. リサイクル，3. 廃棄物管理，4.$CO_2$ 対策，6. 汚染管理，9. 管理体制，10. 報告書会計，11. 教育社会貢献} であり，このときの誤差平方和は91.47であった。一方，第5回調査でも使われた7項目 {2. リサイクル，4.$CO_2$ 対策，6. 汚染管理，7. 商品対策，9. 管理体制，10. 報告書会計，11. 教育社会貢献} (項目番号と項目名は第4回調査のもの) を用いて近似したときの誤差平方和は115.12であり，これは全330通りの中で，79番目であった (表3中の *)。なお，330

表 1: 調査項目

| | 第4回 | 第5回 | 第6回 |
|---|---|---|---|
| 1 | リデュース | 運営体制 | 運営体制 |
| 2 | リサイクル | 情報公開 | 環境教育 |
| 3 | 廃棄物管理 | 環境教育 | ビジョン |
| 4 | $CO_2$ 対策 | 汚染リスク | 汚染リスク |
| 5 | 化学物質管理 | ビジョン | 資源循環 |
| 6 | 汚染管理 | 資源循環 | 製品物流対策 |
| 7 | 商品対策 | 温暖化対策 | 温暖化対策 |
| 8 | 組織制度 | 製品物流対策 | |
| 9 | 管理体制 | | |
| 10 | 報告書会計 | | |
| 11 | 教育社会貢献 | | |

表 2: 調査項目の変遷

| 第4回 | 第5回 | 第6回 |
|---|---|---|
| 1. リデュース | × | — |
| 2. リサイクル | 6. 資源循環 | 5. 資源循環 |
| 3. 廃棄物管理 | × | — |
| 4.$CO_2$ 対策 | 7. 温暖化対策 | 7. 温暖化対策 |
| 5. 化学物質管理 | × | — |
| 6. 汚染管理 | 4. 汚染リスク | 4. 汚染リスク |
| 7. 商品対策 | 8. 製品物流対策 | 6. 製品物流対策 |
| 8. 組織制度 | × | — |
| 9. 管理体制 | 1. 運営体制 | 1. 運営体制 |
| 10. 報告書会計 | 2. 情報公開 | × |
| 11. 教育社会貢献 | 3. 環境教育 | 2. 環境教育 |
| | 5. ビジョン | 3. ビジョン |

表 3: 第4回調査の11項目から7項目選択した結果 (一部)

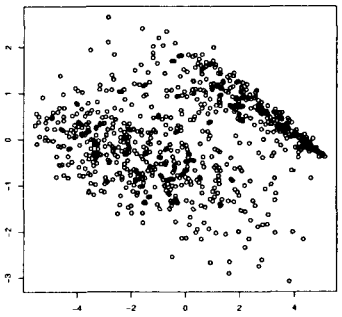
| 順位 | 選択された調査項目 | | | | | | | | | | 誤差 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 平方和 |
| 1 | | o | o | o | | o | | | o | o | o | 91.47 |
| 2 | | o | o | o | | o | | o | o | o | | 92.69 |
| 3 | | o | o | o | | o | o | | o | | o | 92.70 |
| 4 | | o | o | | o | o | | | o | o | o | 93.54 |
| 5 | | o | o | o | | o | | o | | o | o | 94.78 |
| | | | | | | | | | | | | |
| 76 | o | | o | | | o | o | | o | o | o | 114.06 |
| 77 | o | | o | o | o | o | | o | | o | | 114.11 |
| 78 | | o | o | | | o | o | o | o | | o | 114.67 |
| *79 | | o | | o | | o | o | | o | o | o | 115.12 |
| 80 | o | | o | | o | o | o | o | | | o | 115.42 |
| | | | | | | | | | | | | |
| 326 | | | o | | o | o | o | o | o | o | | 197.02 |
| 327 | | | o | | o | | o | o | o | o | o | 201.07 |
| 328 | o | o | o | | o | o | | o | o | | | 202.57 |
| 329 | | | o | | o | o | o | o | o | | o | 207.45 |
| 330 | | | | | o | o | o | o | o | o | o | 223.97 |
| 平均 | | | | | | | | | | | | 131.59 |

調査項目番号は第4回調査のもの

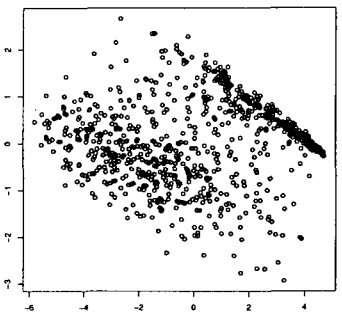
表 4: 第5回調査の8項目から7項目選択した結果

| 順位 | 選択された調査項目 | | | | | | | | 誤差 |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 平方和 |
| 1 | o | o | o | o | | o | o | o | 21.90 |
| 2 | o | o | o | o | o | o | | o | 23.14 |
| 3 | | o | o | o | o | o | o | o | 23.38 |
| *4 | o | | o | o | o | o | o | o | 24.01 |
| 5 | o | o | o | | o | o | o | o | 26.31 |
| 6 | o | o | | o | o | o | o | o | 28.28 |
| 7 | o | o | o | o | o | | o | o | 28.82 |
| 8 | o | o | o | o | o | o | o | | 29.40 |
| 平均 | | | | | | | | | 25.66 |

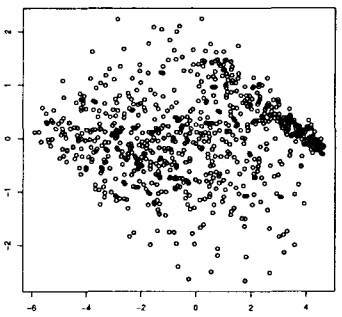
調査項目番号は第5回調査のもの



(i)



(ii)



(iii)

図 1: 第 1 主成分と第 2 主成分の散布図（環境経営度マップ）(i) 元の 11 変数から求めた主成分，(ii) 今回の手法で選択された 7 変数 {1, 2, 4, 6, 7, 9, 11} を用いて推定された主成分，(iii) 第 5 回の調査で用いられた 7 変数 {2, 4, 6, 7, 9, 10, 11} を用いて推定された主成分

通りの誤差平方和の平均は 131.59 であった。

次に，第 5 回調査と第 6 回調査では，「2. 情報公開」が削除されただけであり，他の項目は名称も含めて変更がなく，追加項目もない (第 5 回調査の主成分の寄与率は，第 1 主成分 0.7924，第 2 主成分 0.0485，第 3 主成分 0.0359，...)。8 項目から 7 項目を残す全 8 通りの組合せの中で，残った項目で元の 8 項目による第 1 主成分を近似して最も誤差平方和が小さかった組合せは，{1. 運営体制 2. 情報公開 3. 環境教育 4. 汚染リスク 6. 資源循環 7. 温暖化対策 8. 製品物流対策 }（項目番号と項目名は第 5 回調査のもの）であり，このときの誤差平方和は 21.90 であった。一方，第 6 回調査の 7 項目 {1. 運営体制 3. 環境教育 4. 汚染リスク 5. ビジョン 6. 資源循環 7. 温暖化対策 8. 製品物流対策 }（項目番号と項目名は第 5 回調査のもの）を用いて近似したときの誤差平方和は 24.01 で，これは全 8 通りの中で 4 番目であった (表 4 中の *)。なお，誤差平方和の平均は 25.66 であった。

### 4.1.2 環境経営度マップ

続いて，環境経営度マップに対して行った変数選択を考察する。環境経営度マップとは，第 1 主成分と第 2 主成分の布置で調査企業を観察するもので，第 1 軸は上記の通り環境経営度を，第 2 軸は環境経営のタイプを表していると解釈されている。これに対して，$q = 7$，$r = 2$ として，11 変数から 7 変数に落とす場合について，合成変数推定規準による変数選択を適用する。

このデータに対しては，3 つのステップとも同じ変数群 {1. リデュース，2. リサイクル，4.$CO_2$ 対策，6. 汚染管理，7. 商品対策，9. 管理体制，11. 教育社会貢献 } を選んだ。図 1 は，(i) 元の全 11 変数から求められた主成分，(ii) 今回の変数選択によって選ばれた変数群から求められた主成分の推定値，および (iii) 第 5 回調査で実際に用いられた 7 変数 {2. リサイクル，4.$CO_2$ 対策，6. 汚染管理，7. 商品対策，9. 管理体制，10. 報告書会計，11. 教育社会貢献 } の第 1 主成分と第 2 主成分の散布図（環境経営度マップ）である。この 3 つの図から，(iii) より (ii) の方

が元の布置 (i) をよく再現していることがわかる。2 つの布置の近さを測る $RV$ 係数 (Robert and Escoufier, 1976) で比較しても，(i) と (ii) の $RV$ 係数が 0.97787，(i) と (iii) の $RV$ 係数が 0.97705 と，今回の規準で選択した項目を使った方が環境経営度マップの意味でもよりよい変数群を選んでいることがわかる。

## 4.2　新国民生活指標

データは，都道府県の豊かさを表すために経済企画庁が策定した新国民生活指標のうち，「住む」ことに関する 23 項目で，平成 11 年に発表されたものを用いる。調査項目は，{ 1. 危険住宅, 2. 最低居住, 3. 借家家賃, 4. 持家比率, 5. 公害苦情, 6. 重要犯罪, 7. 重要窃盗, 8. 交通事故, 9. 建物火災, 10. ごみ処理率, 11. 歩道設置率, 12. 医療機関, 13. 居住水準, 14. 日照時間, 15. 畳数, 16. 敷地面積, 17. 交通機関, 18. 公園面積, 19. 下水普及率, 20. リサイクル率, 21. ごみ排出量, 22. 通勤通学時間, 23. 道路舗装率 } である。このデータも主成分を用いて評価されている。相関行列の主成分分析による寄与率は，0.4149, 0.1505, 0.0729, 0.0659, 0.0543, ... で，主成分数は 2 とする。

これらの主成分をより少ない調査項目で近似するとすれば，どの項目を調査すべきかを調べる。表 5 は，各 $q$ において，総当たり法により求めた最も誤差平方和を小さくする調査項目群，表 6 は，その誤差平方和である。表 5 をみると，3 つのステップの計算で，異なる項目群を選んだのは，$q = 17$ と $q = 4$ のときで，いずれもステップ 1 の選んだ項目群が他の 2 つのステップと違った。これら以外は，3 ステップとも同じ項目群を選び，また，ステップ 2 と 3 では，すべての $q$ において同じ項目群を選んだ。これより，減らしたい項目数が決まれば，どの項目にすればよいかの情報がすぐ得られることになる。

表 6 の左から 2 列目，4 列目，6 列目は，各ステップにより求められた誤差平方和で，3 列目と 5 列目は，それぞれステップ 1 と 2，ステップ 2 と 3 の誤差平方和の差である。計算をステップ 1，ステップ 2 で止める場合もステップ 3 まで実行する場合も選択される調査項目は，上記の通りほぼ同じであり，誤差平方和の差も非常に小さいので，調査項目の組合せを選ぶ場合は，組合せの数が多く，計算時間を短縮する必要がある場合は，ステップ 1 で計算を止めて，そこで選択された調査項目の組合せを選び，その後，より適切な推定値を得るために選択された調査項目に対してステップ 3 まで計算を行うことが，計算時間の短縮と推定値の適切さの両面を満たす変数選択手法であるということがいえる。

表 5: 新国民生活指標の「住む」の変数選択結果

| $q$ | 選択された調査項目 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 22 | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| 21 | ○ | | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| 20 | ○ | | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | | ○ | ○ | | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| 19 | ○ | | ○ | | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | | ○ | ○ | | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| 18 | ○ | | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | | ○ | ○ | | | ○ | ○ | | ○ | ○ | ○ |
| 17 | ○ | | ○ | | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | | ○ | ○ | | ○ | ○ | | | ○ | ○ | |
| | ○ | ○ | | | ○ | ○ | | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | | ○ | ○ | | | ○ | ○ | |
| 16 | ○ | ○ | | | ○ | ○ | | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | | | ○ | ○ | | | ○ | |
| 15 | ○ | | ○ | | ○ | ○ | | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | | | ○ | ○ | | | ○ | |
| 14 | ○ | | | | ○ | ○ | | ○ | ○ | ○ | ○ | | | ○ | | | | ○ | ○ | | | ○ | |
| 13 | ○ | | | | ○ | ○ | | ○ | ○ | ○ | ○ | | | ○ | | | | ○ | ○ | | | ○ | |
| 12 | | | | ○ | ○ | ○ | | ○ | | ○ | | | | ○ | ○ | | | ○ | ○ | ○ | ○ | | ○ |
| 11 | | ○ | | | ○ | ○ | | | | ○ | ○ | ○ | | ○ | | | | ○ | ○ | | ○ | | |
| 10 | | | | | ○ | ○ | | | | ○ | | | | ○ | ○ | | | ○ | ○ | ○ | ○ | | ○ |
| 9 | | | | | ○ | | | | | ○ | | | | ○ | ○ | | | ○ | ○ | ○ | ○ | | ○ |
| 8 | | | ○ | | ○ | | | | | ○ | ○ | ○ | | ○ | ○ | | | ○ | | | | | |
| 7 | | | ○ | | ○ | | | | | ○ | ○ | | | ○ | ○ | | | ○ | | | | | |
| 6 | | | ○ | | | | | | | ○ | ○ | | | ○ | ○ | | | ○ | | | | | |
| 5 | | ○ | | | | | | | | ○ | | ○ | | | | | | ○ | | | | ○ | |
| 4 | | | ○ | | | | | | | ○ | | | | ○ | ○ | | | | | | | | |
| | | ○ | | | | | | | | | | ○ | | ○ | | | | | | | | ○ | |
| 3 | | | ○ | | | | | | | | | | | ○ | ○ | | | | | | | | |
| 2 | | | | | | | | | | | | | | ○ | ○ | | | | | | | | |

$q = 17$ と $q = 4$ 以外では，選ばれた変数は 3 つのステップで同じなので，1 行で示してある。$q = 17$ と $q = 4$ では，ステップ 1 で選ばれた変数が異なっているので，上の行にステップ 1 で選ばれた変数，下の行にステップ 2 と 3 で選ばれた変数を示してある。

表 6: 新国民生活指標の「住む」の誤差平方和

| $q$ | 誤差平方和 ステップ 1 | 差 | ステップ 2 | 差 | ステップ 3 |
|---|---|---|---|---|---|
| 22 | 0.022 | 0 | 0.022 | 0 | 0.022 |
| 21 | 0.184 | 0 | 0.184 | 0 | 0.184 |
| 20 | 0.519 | 1E-05 | 0.519 | 0 | 0.519 |
| 19 | 0.838 | 3E-05 | 0.838 | -1E-05 | 0.838 |
| 18 | 1.324 | 0.0003 | 1.325 | -1E-04 | 1.325 |
| 17 | 1.846 | | | | |
| | | 0 | 1.846 | 0 | 1.846 |
| 16 | 2.416 | 0 | 2.416 | 0 | 2.416 |
| 15 | 3.317 | 0 | 3.317 | 0 | 3.317 |
| 14 | 4.222 | 0 | 4.222 | 0 | 4.222 |
| 13 | 5.395 | 0.0009 | 5.396 | -0.0002 | 5.396 |
| 12 | 6.614 | 0.0007 | 6.615 | -0.0002 | 6.615 |
| 11 | 8.074 | 0.0023 | 8.077 | -0.0006 | 8.076 |
| 10 | 10.590 | 0 | 10.590 | 0 | 10.590 |
| 9 | 13.458 | 0.001 | 13.459 | -0.001 | 13.458 |
| 8 | 17.023 | 0.002 | 17.025 | -0.001 | 17.024 |
| 7 | 21.274 | 0 | 21.274 | 0 | 21.274 |
| 6 | 27.974 | 0.025 | 27.999 | -0.006 | 27.993 |
| 5 | 39.601 | 0.014 | 39.615 | -0.004 | 39.611 |
| 4 | 56.079 | | | | |
| | | 0.11 | 56.189 | -0.025 | 56.164 |
| 3 | 77.437 | 0.023 | 77.460 | -0.005 | 77.455 |
| 2 | 137.350 | 0.4 | 137.750 | -0.07 | 137.680 |

$q = 17$ と $q = 4$ では，ステップ 1 で選ばれた変数が異なり，ステップ 2 と 3 で選ばれた変数が同じなので，上の行にステップ 1 で選ばれた変数による誤差平方和，下の行にステップ 2 と 3 で選ばれた変数による誤差平方和を示してある。

## 4.3　社会生活基本調査 (生活時間)

データは，国民の生活時間の配分について調査し，国民の社会生活の実態を明らかにすることにより，各種行政施策の基礎資料を得ること

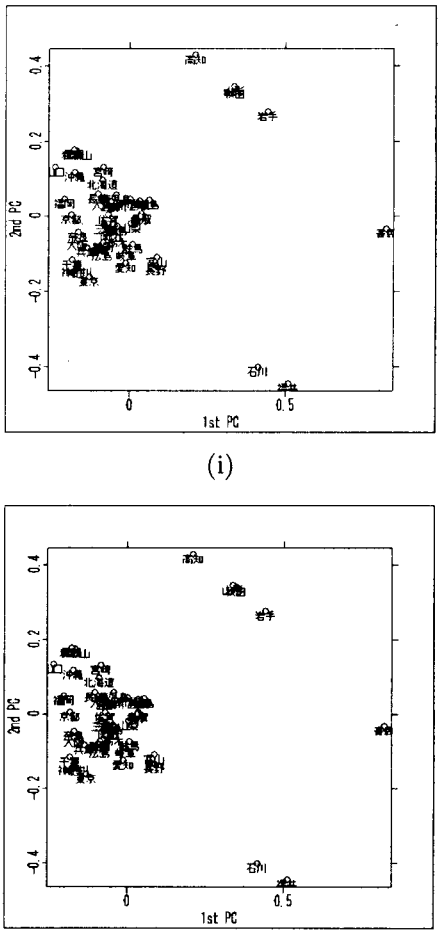
表 7: 社会生活基本調査 (生活時間) の変数選択結果 (ステップ 3 の場合の選択手順の比較)

| $q$ | 選択された調査項目 | | | | | | | | | | | | | | | | | | | | 誤差平方和 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | |
| 20 | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | 0.00000 |
| 19 | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | 0.00000 |
| 18 | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | ○ | ○ | ○ | 0.00001 |
| 17 | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | | ○ | ○ | ○ | 0.00007 |
| 16 | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | | ○ | | ○ | 0.00017 |
| 15 | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | | ○ | ○ | ○ | ○ | ○ | ○ | | | ○ | | ○ | 0.00028 |
| 14 | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | | ○ | ○ | ○ | ○ | ○ | ○ | | | | | ○ | 0.00042 |
| 13 | ○ | | ○ | ○ | ○ | ○ | ○ | | | ○ | ○ | ○ | ○ | ○ | ○ | | | | | ○ | 0.00057 |
| 12 | ○ | | ○ | ○ | ○ | ○ | ○ | | | ○ | ○ | ○ | ○ | ○ | ○ | | | | | | 0.00077 |
| 11a | ○ | | ○ | ○ | ○ | ○ | ○ | | | ○ | ○ | ○ | ○ | | ○ | | | | | | 0.00104 |
| 11b | ○ | | ○ | ○ | ○ | ○ | ○ | | | | ○ | ○ | ○ | ○ | ○ | | | | | | 0.00108 |
| 11c | ○ | | ○ | ○ | ○ | ○ | ○ | | | ○ | ○ | ○ | ○ | | ○ | | | | | | 0.00104 |
| 11d | ○ | | ○ | ○ | ○ | ○ | ○ | | | ○ | ○ | ○ | ○ | | ○ | | | | | | 0.00104 |
| 10a | ○ | | ○ | ○ | ○ | ○ | ○ | | | ○ | | ○ | ○ | | ○ | | | | | | 0.00158 |
| 10b | ○ | | ○ | ○ | ○ | ○ | ○ | | | | | ○ | ○ | ○ | ○ | | | | | | 0.00157 |
| 10c | ○ | | ○ | ○ | ○ | ○ | ○ | | | | | ○ | ○ | ○ | ○ | | | | | | 0.00157 |
| 10d | ○ | | ○ | ○ | ○ | ○ | ○ | | | | | ○ | ○ | ○ | ○ | | | | | | 0.00157 |
| 9 | ○ | | ○ | ○ | ○ | ○ | ○ | | | | | ○ | ○ | | ○ | | | | | | 0.00226 |
| 8 | ○ | | ○ | ○ | ○ | ○ | | | | | | ○ | ○ | | ○ | | | | | | 0.00316 |
| 7 | ○ | | | ○ | ○ | ○ | | | | | | ○ | ○ | | ○ | | | | | | 0.00486 |
| 6 | ○ | | | ○ | ○ | ○ | | | | | | ○ | | | ○ | | | | | | 0.01326 |
| 5 | ○ | | | ○ | ○ | | | | | | | ○ | | | ○ | | | | | | 0.02298 |
| 4 | ○ | | | ○ | ○ | | | | | | | ○ | | | | | | | | | 0.05264 |

$q = 11$, $q = 10$ の a は Backward, b は Forward, c は Backward-forward, d は Forward-backward を示す。$q = 11$, $q = 10$ 以外では，a〜d のすべてで結果は同じであった。

を目的としたもので，総務省統計局により提供されている。昭和 51 年から 5 年ごとに実施されており，今回は平成 13 年の調査結果のうち，都道府県ごとの各生活時間の平均を示したデータを用いる。調査項目は，{1. 睡眠 2. 身の回りの用事 3. 食事 4. 通勤・通学 5. 仕事 6. 学業 7. 家事 8. 介護・看護 9. 育児 10. 買い物 11. 移動（通勤・通学以外）12. テレビ・ラジオ・新聞・雑誌 13. 休業・くつろぎ 14. 学習・研究（学業以外）15. 趣味・娯楽 16. スポーツ 17. ボランティア活動・社会参加 18. 交際・付き合い 19. 受診・療養 20. その他 } である。このデータに対しても合成変数を主成分として，ここでは，4 つの選択手順 (Backward, Forward, Backward-forward, Forward-backward) を適用し，その差をみることにする。分散共分散行列の主成分分析による寄与率は，0.5089, 0.2845, 0.0754, 0.0392, 0.0180, 0.0179, ... で，主成分数は 4 とする。

まず，選択手順を固定し，その手順で 3 つのステップを適用するといずれの選択手順においても選ばれる変数群に違いはなかった。次に，ステップを固定し，そのステップで 4 つの選択手順がどのような変数群を選ぶかをみてみると，いずれのステップにおいても $q = 11$ と $q = 10$ で選ばれる変数群に違いがあるだけで，他はすべて同じであった。表 7 にステップ 3 の場合を示す。このデータの場合は，手順間，ステップ間で選択結果の差が非常に小さいことがわかる。




(i)

(ii)

図 2: 第 1 主成分と第 2 主成分の布置 : (i) 全変数を用いたとき，(ii) 7 変数 {1. 睡眠，4. 通勤・通学，5. 仕事，6. 学業，12. テレビ・ラジオ・新聞・雑誌，13. 休業・くつろぎ，15. 趣味・娯楽 } による推定値を用いたとき

この結果を用いると，たとえば，第 1 主成分と第 2 主成分の布置を得たい場合，7 変数を例にとると，{1. 睡眠，4. 通勤・通学，5. 仕事，6. 学業，12. テレビ・ラジオ・新聞・雑誌，13. 休業・くつろぎ，15. 趣味・娯楽 } が選択されているので，この 7 変数で元の主成分の推定値を求め，その布置を描けば，図 2 にみる通り，元の主成分の布置とほぼ同じ結果が得られることがわかる。

## 5　まとめ

外的変量をもたない多変量手法の変数選択問題において，全変数を観測した場合に得られる合成変数を近似するという観点から変数を選択する規準を提案した。さらに，そのための計算手法として簡便なものから精密なものまで3つのステップを構成し，かつ簡便法としての4つの選択手順を実データに適用し，その性能を考察した。その結果，ステップ間，選択手順間において，選択される変数に大きな違いはなく，これらの計算手法を必要に応じて組み合わせることで，高速かつ高精度な結果が得られることが示唆された。

今後の課題としては，変数の数に関する情報を提供していくことがあげられる。重回帰分析における変数選択問題においては，最適な変数の数を定めることができるが，たとえば，主成分分析における変数選択問題においては，観測されなかった変数こそがノイズとなるため，最適な変数の数が定められず，観測される変数は多いほどよい，という結論になる。この変数の数の決定に対しては，$q$にともなう規準値の変化を考察し，その変化率が急に大きくなるところを1つの境目とする考え方や，ブートストラップ法やクロスバリデーション法により，理想的な変数の数を求めようとする試み (Iizuka et al., 2003) はあるが，まだ，示唆的な情報提供の域を超えていない。しかし，本規準に関しては，その性質を考慮すると，何らかの統計モデルの導入により，最適な変数の数を決定するための情報を提供できる可能性も考えられるので，新しい規準の提案や計算環境の整備とともに，変数の数に関する有効な情報提供についても継続的に研究していく必要がある。

## 参考文献


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*, 15: 337–345.

Mori, Y., Fueda, K. and Iizuka, M. (2004). Orthogonal score estimation with variable selection in multivariate methods. In:Antoch, J. (ed.), *COMPSTAT2004 Proceedings in Computational Statistics*, 1527-1534, Phisica-Verlag.

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

Robert, P. and Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: the RV-coefficient. *Appl. Statist.*, 25: 257–265.

笛田　薫，飯塚誠也，森 裕一 (2003). Orthogonal score estimation with variable selection in multivariate methods. 日本計算機統計学会第17回シンポジウム論文集, 129-132.

笛田　薫，飯塚誠也，森 裕一 (2005). 合成変数をベースにした項目選択手法とシステムへの実装. 日本計算機統計学会第19回シンポジウム論文集, 75-78.

森 裕一，垂水共之，田中 豊 (1998). 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，笛田　薫，飯塚誠也 (2005a). 主成分をベースにした項目選択と評価. 第33回日本行動計量学会大会論文集, 238-241.

森　裕一，笛田　薫，飯塚誠也 (2005b). 変数の一部を用いた合成変数の推定とその変数選択. 平成17年度科学研究費シンポジウム「多変量同時解析モデルと関数データに関する研究会」.

# 付録: 変数選択環境VASMMによる計算

## A VASMMの仕様

VASMM (http://mo161.soci.ous.ac.jp/vasmm/) では，今回提唱の手法に対して，Web版とR版を提供している。

Web版については，日本語版と英語版を用意しており，国内外両方からのアクセスに対応できる。このシステムは，スクリプト言語として，主に，日本語処理に優れているRuby (http://www.ruby-lang.org/) を用いたCGIで制御されている。実際の計算には，統計エンジンとして，R (http://cran.r-project.org/) を用いている。

R版は，主にローカルマシン上で実行するために提供するユーザ向けのRの関数である。ここでは，ただ関数を提供するだけでなく，誰でも使えることを念頭に，Rに用意されているGUI関係の関数を用いて，できる限り，コマンドを打つことなく，データやオプションを指定していくだけで計算結果が得られるように配慮した関数にしている。これにより，変数選択を行いたいユーザは，他に新たなインストールをすることもなく，容易に自分のコンピュータで変数選択を実行することが可能になる。

これらに今回提唱の合成変量をベースにした変数選択のモジュールを加えたものが最新版として提供されている。

### A.1 実際の動作（合成変数推定を利用した変数選択）

#### A.1.1 Web版

主成分分析の場合で説明する。VASMMのサイトhttp://mo161.soci.ous.ac.jp/vasmm/にアクセスし，Vaspca/Webへのリンクhttp://mo161.soci.ous.ac.jp/vaspca/をクリックすると，実行画面に移動する。ユーザは実行画面では，CGI版かXQS版を選択することができる。ここで，CGI版をクリックすると，主成分分析における変数選択Vaspca/Webで利用できる変数選択規準の一覧が表示される（図3)。このうち，[F. 合成変量の推定をベースにした変数選択] を指定することにより，今回新たに加えられた規準で変数選択が実行できる。この規準についての簡単な説明ページも設けているので，論文などを参照することなく，どのような理論で計算されているかは，説明へのリンクをクリックするだけで確認できるようになっている。

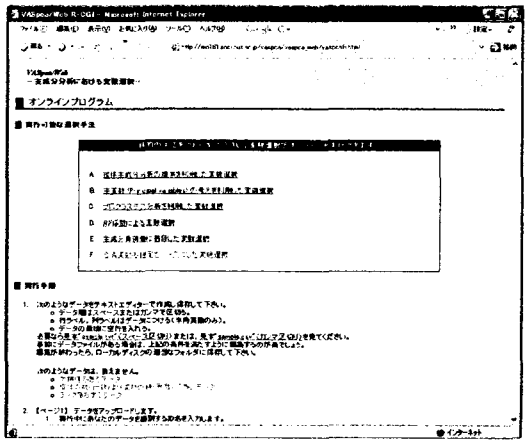


図 3: Web版：選択規準の指定画面

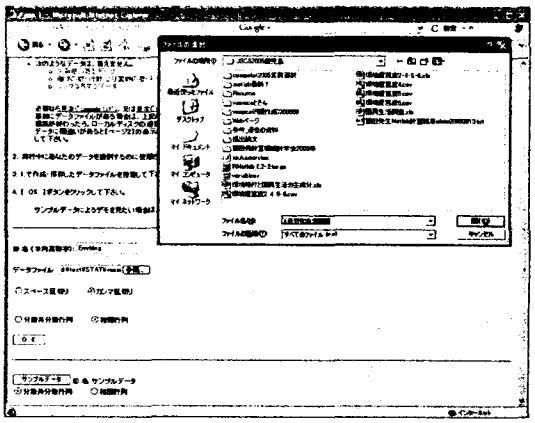


図 4: Web版：データセットの選択

実際の選択では，まず，ファイル（第4回環境経営度調査のデータセット）を選択し，ID名や処理に使う行列のタイプなどを指定し実行する（図4)。すると，通常の主成分分析の結果が表示される（図5）ので，これを基に主成分数$r$を決め，条件指定画面（図6）へ進む。ここで，主成分数，適用するステップ，選択手順を指定して，[変数選択開始] ボタンをクリックす

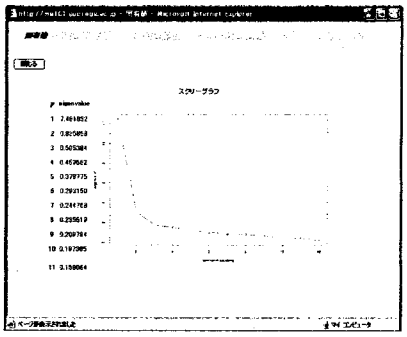


図 5: Web 版：通常の主成分分析による固有値の表示

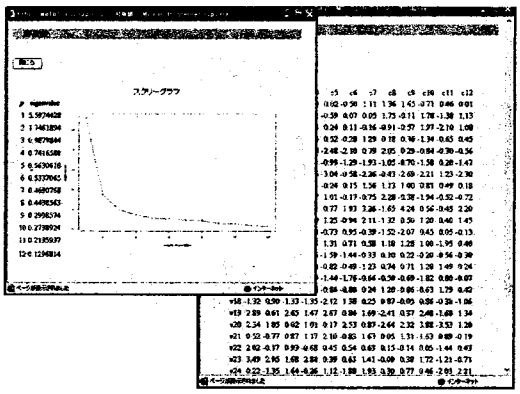


図 6: Web 版：変数選択のためのパラメータ指定画面

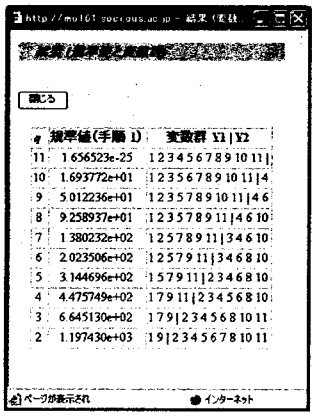


図 7: Web 版：選択結果（$q$ ごとの規準値と選択変数）

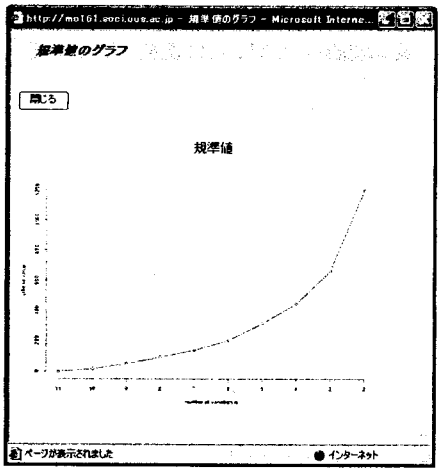



図 8: Web 版：選択結果（規準値の変化）

ると，計算の後，変数選択結果が表示される。選択結果は，$q$ ごとの規準値と選択変数を一覧表にしたもの（図 7）と，$q$ による規準値の変化を示すグラフ（図 8）の 2 つのサブウィンドウで表示される。このグラフは，変数の数を決めるときなどの参考にする。メインウィンドウは図 6 に戻っているので，必要に応じて，条件を変えて選択を繰り返すことができる。

### A.1.2　R 版

R 版については，Vaspca/Win のページから関数をダウンロードし，ローカルマシンの上で，R を立ち上げ，この関数を読み込み，R Console で `vasmm()` と入力し，そこに順に表示される指示に従い，パラメータを指定していけば，各種の変数選択が実行できるようになっている。

まず，`vasmm()` を実行すると，解析対象のデータセットを指定するダイヤログボックスが開かれる。ここでデータセット（第 4 回環境経営度調査のデータセット）を選択し，Console 上でそのデータの形式をテキスト，CSV，Excel の中から指定する。指定された形式でデータセットが読み込まれると，対象とする多変量手法の指定に移る。主成分分析，因子分析，コレスポンデンス分析から 1 つを選択し，用いる行列のタイプを指定する。すると，通常の多変量手法が行われ，合成変量の計算に必要な軸数（$r$）の指定に移る。主成分分析の場合，固有値とそのスクリープロッ

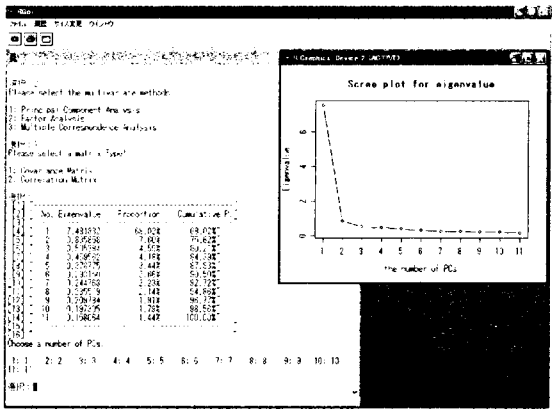



図 9: R 版：通常の主成分分析による固有値の表示

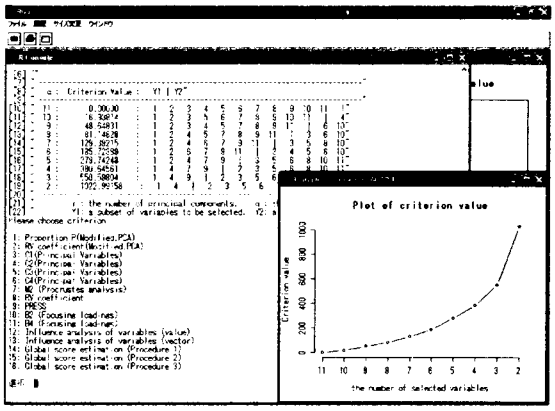



図 10: R 版：選択結果（$q$ ごとの規準値と選択変数および規準値の変化を示すグラフ）

トが表示される（図 9）。これを参考に $r$ を指定すると，次に，その多変量手法に適用可能な選択規準が示される。主成分分析では，16 個の規準が表示されているので，ここでは，今回の規準の [16: Global score estimation (Step 3)] を選ぶ。最後に，総当たり法も含めた選択手順の指定が促されるので，6 つの選択手法から 1 つを指定する。変数の数は 11 と，十分計算可能な数なので，[6: Allpossible selection] を選んで，実行する。結果は，図 10 のように，Console 上には $q$ ごとの規準値と選択変数の一覧表が出力され，同時に，$q$ による規準値の変化を示すグラフが表示される。Console 上の一連の入出力は図 11 のようになる。

```
> vasmm()
Please select file format

1: Text File
2: CSV File
3: Excel File

選択：2
Please select the multivariate methods

1: Principal Component Analysis
2: Factor Analysis
3: Multiple Correspondence Analysis

選択：1
Please select a matrix Type!

1: Covariance Matrix
2: Correlation Matrix

選択：2
 [1] "-------------------------------------------------"
 [2] "  No. Eigenvalue    Proportion    Cumulative P."
 [3] "-------------------------------------------------"
 [4] "   1    7.481832       68.02%        68.02%"
 [5] "   2    0.835858        7.60%        75.62%"
 [6] "   3    0.505384        4.59%        80.21%"
 [7] "   4    0.459562        4.18%        84.39%"
 [8] "   5    0.378775        3.44%        87.83%"
 [9] "   6    0.293150        2.66%        90.50%"
[10] "   7    0.244768        2.23%        92.72%"
[11] "   8    0.235519        2.14%        94.86%"
[12] "   9    0.209784        1.91%        96.77%"
[13] "  10    0.197305        1.79%        98.56%"
[14] "  11    0.158064        1.44%       100.00%"
[15] "-------------------------------------------------"
[16] "                                               "
Choose a number of PCs.

 1: 1   2: 2   3: 3   4: 4   5: 5   6: 6   7: 7   8: 8   9: 9  10: 10
11: 11

選択：2
Please choose criterion

 1: Proportion P(Modified.PCA)
 2: RV coefficient(Modified.PCA)
 3: C1(Principal Variables)
 4: C2(Principal Variables)
 5: C3(Principal Variables)
 6: C4(Principal Variables)
 7: M2 (Procrustes analysis)
 8: RV coefficient
 9: PRESS
10: B2 (Focusing loadings)
11: B4 (Focusing loadings)
12: Influence analysis of variables (value)
13: Influence analysis of variables (vector)
14: Global score estimation (Step 1)
15: Global score estimation (Step 2)
16: Global score estimation (Step 3)

選択：16
Please choose selection procedure

1: Backward elimination
2: Forward selection
3: Backward-Forward stepwise selection
4: Forward-Backward stepwise selection
5: Allpossible selection at specified q
6: Allpossible selection

選択：6
 [1] " ************************************************************$
 [2] "     Variable Selection in Principal Component Analysis    "
 [3] "    using selection criteria in Global score estimation (Proced$
 [4] " ************************************************************$
 [5] "     Correlation Matrix,  Global score estimation (Procedure 3)$
 [6] ""
 [7] " ------------------------------------------------------------$
 [8] "   q :  Criterion Value :   Y1 | Y2"
 [9] " ------------------------------------------------------------$
[10] "  11 :       0.00000    :  1  2  3  4  5  6  7  8  9 $
[11] "  10 :      16.93814    :  1  2  3  5  6  7  8  9 10 $
[12] "   9 :      48.64831    :  1  2  3  4  5  7  8  9 11 $
[13] "   8 :      81.14828    :  1  2  4  5  7  8  9 11  | $
[14] "   7 :     129.39215    :  1  2  4  6  7  9 11  | 3 $
[15] "   6 :     185.72389    :  1  2  6  7  9 11  | 3  4 $
[16] "   5 :     279.74248    :  1  2  4  7  9  | 3  5  6 $
[17] "   4 :     380.64561    :  1  4  7  9  | 2  3  5  6 $
[18] "   3 :     550.59804    :  1  4  9  | 2  3  5  6  7 $
[19] "   2 :    1022.99158    :  1  4  | 2  3  5  6  7  8 $
[20] " ------------------------------------------------------------$
[21] "         r : the number of principal components,   q : the num$
[22] "         Y1: a subset of variables to be selected,  Y2: a subse$
```

図 11: R Console 上の入力と出力の遷移（上から 3 つ目の選択により，図 9 のように通常の主成分分析の結果が表示され，最後の選択によって，図 10 のように結果が得られる。）