重回帰分析における変数選択について

On the Selection Problem of Variables in Multiple Regression Analysis

慶応義塾大学理工学部
竹内寿一郎

１．はじめに

ずっと以前情報処理講座の授業で、「段階的重回帰分析」というテーマで APL 言語で書いたプログラムを紹介していたことがあった。その後理工学部の応用統計学第 2 を担当するようになって、J 言語でそれを書き直した。J で関数化したアルゴリズムは正確には「変数増加法」、もしくは「変数増減法」の特殊な場合といわれる手法で、いずれもスタート時に含まれる変数の数はゼロで、それから逐次変数を取り入れ、次に取捨選択を繰り返すという手法に限定されていた。

従属変数 (基準変数、または被説明変数ともいう) を最も良く説明する最適変数の組合せは、説明変数間に相関がある場合、多くのケースでは全ての変数の組合せを調べなければ知ることが出来ない。変数選択型回帰分析法を利用すると、ときには局所最適変数の組合せに到達してアルゴリズムが終了してしまうことがある。特に「変数増加法」、「変数減少法」では一本道で局所最適組合せに到達する可能性が高い。それを防ぐため変数の取捨選択はその時々の最適組合せを辿るのではなく、時にはランダムに強制的にある変数を取り入れる試みも為されている。これは一種の遺伝的アルゴリズムということが出来る。

以上のことを踏まえると、一般に「変数増減法」、「変数減増法」と呼ばれる、任意の変数組合せから出発する方法は、遺伝的アルゴリズムに似て変数の組合せをランダムに選ぶことにより、いくつかの局所最適組合せに到達させて、その中から最も良い組合せを選ぶ方法である。

ここではこれまでの変数選択型回帰分析アルゴリズムを一般化し、任意の変数組合せから出発する逐次型の「変数選択回帰分析」に拡張したのでそれについて述べることにする。

２．小行列で表された行列の逆行列

まず、逐次型変数選択型回帰分析で用いられる重要な関係式を見ておこう。
いま、従属変数 y の観測値ベクトルを $\boldsymbol{y}(n\times 1)$、$p$ 個の説明変数 $(x_1, x_2, \cdots, x_p)$ の観測値行列を $\mathbf{X}_p(n\times p)$ とし、簡単のためにこれらは平均を引いた変数として定義することにする。

$$\boldsymbol{X}_p = \begin{pmatrix} x_{11}-\bar{x}_1 & x_{21}-\bar{x}_2 & \cdots & x_{p1}-\bar{x}_p \\ x_{12}-\bar{x}_1 & x_{22}-\bar{x}_2 & \cdots & x_{p2}-\bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n}-\bar{x}_1 & x_{2n}-\bar{x}_2 & \cdots & x_{pn}-\bar{x}_p \end{pmatrix}, \qquad \boldsymbol{y} = \begin{pmatrix} y_1-\bar{y} \\ y_2-\bar{y} \\ \vdots \\ y_n-\bar{y} \end{pmatrix}$$

$$\boldsymbol{X}_{p+1} = \begin{pmatrix} x_{11}-\bar{x}_1 & x_{21}-\bar{x}_2 & \cdots & x_{p+1,1}-\bar{x}_{p+1} \\ x_{12}-\bar{x}_1 & x_{22}-\bar{x}_2 & \cdots & x_{p+1,2}-\bar{x}_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n}-\bar{x}_1 & x_{2n}-\bar{x}_2 & \cdots & x_{p+1,n}-\bar{x}_{p+1} \end{pmatrix}, \qquad S_{yy} = \boldsymbol{y}'\boldsymbol{y} = \sum_{\alpha=1}^{n}(y_\alpha-\bar{y})^2$$

$$\boldsymbol{X}'_p\boldsymbol{X}_p=(S_{ij})=\boldsymbol{S}_{pp},\qquad \boldsymbol{X}'_{p+1}\boldsymbol{X}_{p+1}=\boldsymbol{S}_{p+1,p+1}=\begin{pmatrix}\boldsymbol{S}_{pp} & \boldsymbol{s}_p\\ \boldsymbol{s}'_p & S_{p+1,p+1}\end{pmatrix}$$

$$\boldsymbol{u}_p=\boldsymbol{X}'_p\boldsymbol{y},\qquad \boldsymbol{u}_{p+1}=\begin{pmatrix}\boldsymbol{u}_p\\ S_{p+1,y}\end{pmatrix}=\begin{pmatrix}\boldsymbol{X}'_p\boldsymbol{y}\\ S_{p+1,y}\end{pmatrix}$$

ここで、

$$\bar{x}_i=\frac{1}{n}\sum_{\alpha=1}^{n}x_{i\alpha},\qquad \bar{x}_j=\frac{1}{n}\sum_{\alpha=1}^{n}x_{j\alpha}\qquad i,j=1,2,\cdots,p+1\qquad \bar{y}=\frac{1}{n}\sum_{\alpha=1}^{n}y_\alpha$$

$$S_{ij}=\sum_{\alpha=1}^{n}(x_{i\alpha}-\bar{x}_i)(x_{j\alpha}-\bar{x}_j)\qquad S_{iy}=\sum_{\alpha=1}^{n}(x_{i\alpha}-\bar{x}_i)(y_\alpha-\bar{y})$$

太字の大文字は行列、太字の小文字はベクトル、その転置は $'$ で表す。従って太字でない記号は大文字でも小文字でもスカラ、また、下付き添え字は重要な要素数を表し、スカラでは要素名とする。上付き添え字は逆行列の小行列、ベクトル、スカラを表す。なお、上もしくは下に*印が付けられた変数は、逐次過程で新たに計算し直した結果得る変数を表す。そこで、これらの記号を使って変数選択の過程を追ってみる。

まず、変数を落とす場合。p 番目の変数を落とすとしても一般性を失わないのでその場合について述べる。さて、$\boldsymbol{S}_{pp}$ の逆行列はすでに計算されていて、

$$\boldsymbol{S}_{pp}^{-1}=\begin{pmatrix}\boldsymbol{S}_{p-1,p-1} & \boldsymbol{s}_{p-1}\\ \boldsymbol{s}'_{p-1} & S_{pp}\end{pmatrix}^{-1}=\begin{pmatrix}\boldsymbol{S}^{p-1,p-1} & \boldsymbol{s}^{p-1}\\ \boldsymbol{s}^{p-1\prime} & S^{pp}\end{pmatrix}=(S^{ij})$$

と書くとする。正規方程式を解くと、

$$\boldsymbol{b}^*_{p-1}=\boldsymbol{S}^{-1}_{p-1,p-1}\boldsymbol{u}_{p-1}$$

となる。逆行列に関する小行列式の関係式から、

$$\boldsymbol{S}^{-1}_{p-1,p-1}=\boldsymbol{S}^{p-1,p-1}-\frac{\boldsymbol{s}^{p-1}\boldsymbol{s}^{p-1\prime}}{S^{pp}}$$

から p を落とした場合の逆行列および回帰係数は簡単に計算することができる。この計算はピボットを S^{pp} として、$\boldsymbol{S}_{pp}^{-1}$ を掃き出す演算で、シンプレックス表における、λ_p を基底に入れる演算であり、その計算量は $\boldsymbol{S}_{p-1,p-1}$ の逆行列を直接計算する計算量に比べ、ずっと少ないことがわかる。

次に変数 x_{p+1} を加える場合。このとき出てくる元の平方和・積和行列 $\boldsymbol{S}_{pp}$ と $\boldsymbol{S}^{-1}_{p+1,p+1}$ との関係を求めてみる。

$$\boldsymbol{S}^{-1}_{p+1,p+1}=\begin{pmatrix}\boldsymbol{S}_{pp} & \boldsymbol{s}_p\\ \boldsymbol{s}'_p & S_{p+1,p+1}\end{pmatrix}^{-1}=\begin{pmatrix}\boldsymbol{S}^{pp}_* & \boldsymbol{s}^p_*\\ \boldsymbol{s}^{p\prime}_* & S^{p+1,p+1}_*\end{pmatrix}$$

とおき、小行列に関する定理を適用すると、

$$S^{p+1,p+1}_*=\frac{1}{S_{p+1,p+1}-\boldsymbol{s}'_p\boldsymbol{S}_{pp}^{-1}\boldsymbol{s}_p}$$

$$\boldsymbol{s}^p_*=-\boldsymbol{S}_{pp}^{-1}\boldsymbol{s}_pS^{p+1,p+1}_*$$

$$\boldsymbol{S}_{pp}^{-1}=\boldsymbol{S}^{pp}_*-\frac{\boldsymbol{s}^p_*\boldsymbol{s}^{p\prime}_*}{S^{p+1,p+1}_*}\qquad \text{すなわち、}\qquad \boldsymbol{S}^{pp}_*=\boldsymbol{S}_{pp}^{-1}+\frac{\boldsymbol{s}^p_*\boldsymbol{s}^{p\prime}_*}{S^{p+1,p+1}_*}$$

を得る。$\boldsymbol{S}_{pp}^{-1}$ は前の段階で既知であり、逆行列の演算をすることなしに、$\boldsymbol{S}_{p+1,p+1}$ の逆行列を計算することができる。またこの式は先の $\boldsymbol{S}_{p-1,p-1}^{-1}$ を求める式の逆演算、すなわち掃き出しをもとに戻す演算であり、シンプレックス表でいえば x_{p+1} を基底に入れる演算に相当することになる。

これまでは説明変数として $(x_1, x_2, \cdots, x_p, x_{p+1})$ について述べてきたが、より一般的には $(x_1, x_2, \cdots, x_p, x_{p+1}, \cdots, x_k), (k > p)$ についても同様に議論することができる。そのときは $\boldsymbol{s}_p$ は $n \times (k-p)$ の行列、$S_{p+1,p+1}$ は $(k-p) \times (k-p)$ の行列となる。

３．変数選択型回帰分析

変数 x_i を回帰からはずしたときの残差平方和 S_e^* は、

$$S_e^* = S_e + \frac{b_i^2}{S^{ii}}$$

となるので、x_i を落すことにより残差平方和は b_i^2/S^{ii} だけ増加することがわかる。従って変数を減らすときは、$b_1^2/S^{11},\ b_2^2/S^{22},\ \cdots,\ b_p^2/S^{pp},$ を計算し、その中の最小値に対応する x_i が回帰から外される変数の候補と考えてよい。

変数 x_i を回帰に加えたときの残差平方和 S_e^* は、

$$S_e^* = S_e - \frac{b_i^{*2}}{S_*^{ii}}$$

となることから、b_i^{*2}/S_*^{ii} だけ減少することになる。$(x_{p+1},\ x_{p+2},\ \cdots,\ x_k)$ について $k-p+1$ 個の変数について減少分を計算し、その中で最大の x_i を選ぶことにする。そうすると $(x_1,\ x_2,\ \cdots,\ x_p)$ に x_i を加えた組による回帰式が S_e^* を最も小さくすることになる。

ところで、シンプレックス表において、$(x_1,\ x_2,\ \cdots,\ x_p)$ を基底に入れたとき、$S_{p+1,p+1}$ および $S_{p+1,y}$ があったところは、

$$\begin{pmatrix} \boldsymbol{S}_{pp}^{-1} & \boldsymbol{0} \\ -\boldsymbol{s}_p'\boldsymbol{S}_{pp}^{-1} & 1 \end{pmatrix} \left(\begin{array}{c|c|c|c} \boldsymbol{S}_{pp} & \boldsymbol{s}_p & \boldsymbol{I}_p & \boldsymbol{u} \\ \boldsymbol{s}_p' & S_{p+1,p+1} & \boldsymbol{0}' & S_{p+1,y} \end{array} \right)$$
$$= \left(\begin{array}{c|c|c|c} \boldsymbol{I}_p & \boldsymbol{S}_{pp}^{-1}\boldsymbol{s}_p & \boldsymbol{S}_{pp}^{-1} & \boldsymbol{S}_{pp}^{-1}\boldsymbol{u}_p \\ \boldsymbol{0}' & S_{p+1,p+1} - \boldsymbol{s}_p'\boldsymbol{S}_{pp}^{-1}\boldsymbol{s}_p & -\boldsymbol{s}_p'\boldsymbol{S}_{pp}^{-1} & S_{p+1,y} - \boldsymbol{s}_p'\boldsymbol{S}_{pp}^{-1}\boldsymbol{u}_p \end{array} \right)$$

より、$S_{p+1,p+1} - \boldsymbol{s}_p'\boldsymbol{S}_{pp}^{-1}\boldsymbol{s}_p$ および $S_{p+1,y} - \boldsymbol{s}_p'\boldsymbol{S}_{pp}^{-1}\boldsymbol{u}_p$ に変わっている。ここで、

$$\begin{pmatrix} \boldsymbol{S}_{pp}^{-1} & \boldsymbol{0} \\ -\boldsymbol{s}_p'\boldsymbol{S}_{pp}^{-1} & 1 \end{pmatrix}$$

は $(x_1,\ x_2,\ \cdots,\ x_p)$ を基底に入れるための掃き出し演算を行う行列で、シンプレックス表の $(x_1,\ x_2,\ \cdots,\ x_p)$ に対応する小行列を $(\boldsymbol{I}_p,\ \boldsymbol{0})'$ にする。一方、誤差の減少分は、$b_{p+1}^{*2}/S_*^{p+1,p+1}$ であるから、

$$\frac{b_{p+1}^{*2}}{S_*^{p+1,p+1}} = \frac{(S_{p+1,y} - \boldsymbol{s}_p'\boldsymbol{S}_{pp}^{-1}\boldsymbol{u}_p)^2}{S_{p+1,p+1} - \boldsymbol{s}_p'\boldsymbol{S}_{pp}^{-1}\boldsymbol{s}_p}$$

となる。このことから各変数に対する減少分は改めて計算する必要はなく、シンプレックス表から分かるように、回帰に取り入れられていない変数に対応する対角要素、およびある段階でその変数と y との積和要素のある欄 (シンプレックス表の最右端) から要素を取り出して

$$\frac{(S_{p+1,y}-\boldsymbol{s}_p'\boldsymbol{S}_{pp}^{-1}\boldsymbol{u}_p)^2}{S_{p+1,p+1}-\boldsymbol{s}_p'\boldsymbol{S}_{pp}^{-1}\boldsymbol{s}_p}$$

を計算すればよい。

この量は増分 b_p^2/S^{pp} と対応していて、回帰に取り入れられていない変数に関しては、その変数が取り入れられたときの S_e のの減少分、回帰に取り入れられている変数に関してはその変数が除かれたときの S_e の増分となる。

４．ストッピングルール

いま、ある段階で y の回帰係数として $(x_1,\ x_2,\ \cdots,\ x_p)$ が用いられているとする。この段階で変数を１つ減少させるとすれば、b_i^2/S^{ii} を $i=1,2,\cdots,p$ について計算し、その最も小さい x_i を落す変数の候補とすることは前に述べた。ところで、

$$t_0=b_i/\sqrt{S^{ii}S_e/(n-k-1)}$$

は回帰係数の有意性の検定統計量で、自由度 $n-k-1$ の t 分布に従うことから、増分を $S_e/(n-p-1)$ でわった

$$(1)\qquad F_{OUT}=\frac{b_i^2}{S^{ii}S_e/(n-p-1)}$$

は自由度 $(1,n-p-1)$ のＦ分布に従うことが分かる。すなわちこの値ががＦ分布のあるパーセント点 F_{OUT}^* より小さければ仮説 $H_0:\beta_i=0$ は棄てられないので、x_i を変数の組から除外するのである。

一方、変数を増加させる場合は、変数を加えることにより残差が減少する減少分 b_i^{*2}/S_*^{ii} を $i=p+1,p+2,\cdots,k$ について計算し、その最大となる x_i を加える変数の候補とするのである。このとき、

$$(2)\qquad F_{IN}=\frac{b_i^{*2}}{S_*^{ii}(S_e^*)_i/\{n-(p+1)+1\}}$$

は自由度 $(1,n-p)$ のＦ分布に従うことから、この値が F 分布のあるパーセント点 F_{IN}^* より大きいとき、x_i を回帰変数として取り入れることになる。落としを先にするか、取り込みを先にするかが、変数減増型と変数増減型の違いであり、通常 $F_{IN}^*>F_{OUT}^*$ であれば適当な繰り返しで終了する。

$F_{IN}^*<F_{OUT}^*$ のように設定すると同じ変数が取り入れられては落とされるという、無限ループとなる場合があるので注意しなければならない。F_{IN}、および F_{OUT} の分布のパーセント点はパラメタとして自由度を含み、それにより値が異なるので、実際にはＦ分布のパーセント点の計算が必要である。しかし実用上は $F_{IN}^*=F_{OUT}^*$ とおき、自由度に関係なく 2.0〜2.5 ぐらいの値がよく使われている。

４．Ｊのプログラムと例題

任意の変数群から出発して変数増減、または変数減増しながら変数を選択して行く回帰分析をどのようにプログラムするかについて暫し考えた。ゼロから出発して指定された変数を逐次入れてゆき、その全てを取り入れたら取捨選択のアルゴリズムに入ることも考えた。す

でに出来上がっていた変数選択型回帰分析は各ステップ毎に分散分析表、重相関係数、AIC、Mallows の Cp などプリントされることになっていた。指定された変数を取り入れるステップでは何もプリントしないように修正するのも面倒であった。そこでまず、一気に指定された変数群で掃き出しを行い、その後増減のアルゴリズムに入ることにした。下の J の関数リストの中で sweepout がその関数である。指定された変数は当然飛び飛びに与えられるので、まずそれらを一箇所に集め、掃き出し演算を行った後、変数を元の位置に戻し、シンプレックス表 (Tab) の形になるようにした。関数 sweepout 内の変数 v1 は変数を抜き出すためのベクトル、v3 はそれを元に戻す指標である。

なお、関数 sweepout の作成時に expand を使用したが、

```
   0 0 0 0 1 0 1 0 expand 100 200
```

のように、左引数の先頭にゼロが多く並ぶとエラーとなるので、

```
   }.1,0 0 0 0 1 0 1 0 expand 1,100 200
```

としてエラーを避けることにした。
これは expand=:(1: j. # ;. _1)@[#] の思いもよらなかったバグである。
まお、expand=:#^:_1 とすれば問題なく結果を得ることができる。

```
NB.
NB. Stepwise Regression Analysis
NB.
NB. Usage1 : step_a x;y
NB. Usage2 : 4 7 step_a x;y
print=:(1!:2)&2
input=:1!:1   NB.  Usage   data=.input 1
mean=:+/%#
mp=:+/ .*
mvar=:((|:mp])@(-"1 mean))%(<:@#)
mss=:(|:mp])@(-"1 mean)
corrv=:3 : 0
y.%*/~%:(<0 1)|:y.
)
NB.  テストデータ
 x1=.20.6  11.5  2.9  7.0  1.7  0.873  0.297  0.083  0.586  0.609  23.6
 x2=.22.8  15.8  6.2  8.7  4.3  0.860  0.328  0.189  0.694  0.551  26.5
 x3=.18.3  13.0  3.7  6.1  3.4  0.762  0.254  0.186  0.919  0.469  24.0
 x4=.16.5  15.9  5.6  9.7  5.1  0.868  0.511  0.309  0.911  0.610  19.0
 x5=.17.0  11.1  2.1  7.0  1.7  0.783  0.323  0.100  0.810  0.631  21.7
 x6=.22.1  16.1  5.0  7.0  3.5  0.965  0.306  0.158  0.700  0.435  22.9
 x7=.18.9  15.0  6.0  7.3  3.5  0.727  0.281  0.185  0.583  0.487  26.0
 x8=.19.3  11.8  2.1  6.1  1.7  0.684  0.216  0.088  0.810  0.517  28.2
```

```
 x9=.19.3  15.2  7.0  5.2  3.5  0.757  0.204  0.181  0.500  0.342  25.5
x10=.15.4  15.6  4.1  5.2  3.6  0.670  0.226  0.234  0.878  0.333  23.0
x11=.18.8  17.2  6.8  8.7  2.6  0.870  0.403  0.138  0.382  0.506  21.6
x12=.18.3  14.5  4.4  4.2  3.3  0.871  0.200  0.180  0.750  0.290  21.0
x13=.18.1  16.1  4.6  7.9  2.6  0.691  0.302  0.144  0.565  0.491  26.2
x14=.23.9  16.6  6.0  7.9  4.4  0.885  0.293  0.184  0.733  0.476  27.0
x15=.22.4  13.0  6.0  7.3  3.5  0.933  0.304  0.156  0.583  0.562  24.0
XX=:15 11$x1,x2,x3,x4,x5,x6,x7,x8,x9,x10,x11,x12,x13,x14,x15
y=:{:"1 XX
x=:}:"1 XX
X=:x;y

NB.----------------------関数定義-------------------------------------
   step_a=:3 :0
(i.0)step_a y.
:
'y x'=.y.
k=.#Tab=._1}. S=:mss y,.x
(av=.(+/y,.x)%n),.D=.%:((<0 1)|:S)%<:n=.#y
print'Means and Standard deviations'
print'Variable  Mean       Std.dev.'
print (fmt=:4.0 12.4 12.4)":(>:i.>:k),.av,.D
print'Correlation Coefficients'
print 8.4":corrv S
TL=:2.2 2.2
b=.x%.xa=.1,.y      NB. for Mallows' Cp
sE0=:+/*:x-xa mp b  NB. for Mallows' Cp
E=.(L=.1),1}.J=.k#C=.0*sT=.sE=:{:{:S
if.0<#x.do.
print x.
b=.x%.xa=.(ID=.1),.((J=.}:v1=.+/(<:,x.)=/i.#S)#"1 y)
sE=:+/*:x-xa mp b
Tab=.sweepout v1;S
end.
label_in.
PI=.(<Q;Q=.(-.J)#i.k){Tab+0*SN=.1 NB. SN=1 variable in
F=.T2%(sE-T2=.(*:Q{{:"1 Tab)%(2#&.><"0 i.#PI){PI)%_2+n-+/J
goto_judge.
label_out.
PI=.(<Q;Q=.J#i.k){Tab+0*SN=._1 NB. SN=_1 variable out
```

```
F=.(T2=.SN*(*:Q{{:"1 Tab)%(2#&.><"0 i.#PI){PI)%sE%_1+n-+/J
label_judge.
if.(>./F)>SN*(ID=.2-(3+SN)%2){TL do. goto_sweep. end.
if.0=(k=+/J)+3<:L=.L+1 do. goto_cont. else. goto_owari. end.
label_sweep.
J=.(SN=L=.1)(<(M=.(T2 i.W=.>./T2){Q))}J NB.
P0=.M{PP=.M{"1 Tab
Tab=.((-(>:M)-1)|.E)M}"0 1 Tab
Tab=.Tab-((-.M=i.k)*PP) */ M{Tab=.((M{Tab)%P0)M}Tab
sE=:sE-W
dB=.%:(((2#&.><"0 i.#Wk){Wk=.(<Q;Q=.J#i.k){Tab))*RES=.sE%nF=._1+n-kk=:+/J
b0=.((J,1)#av)mp(-b=.J#{:"1 Tab),1
T=.b%dB
print 'Step : ',":C=.C+1
print '   Varibles Coefficients Std.Coeff. Stand.Dev.  t-Valus'
print (11.0":,.(>:J#i.k)),"1[11.4":b,.((b*J#}:D)%{:D),.dB,.T
print '   Constant ',12.4":b0
print '   Coeff. of Det.',(8.5":r2),'   Adjusted R2   ',
     8.5":((r2=.(sR=.sT-sE)%sT)*(nT=._1+n)%nF)-kk%nF
print ('   AIC   ',12.5":((n*^.sE%n)+2*kk+2)),'       Mallows Cp   ',
     12.5":(sE%sE0)+(2*kk)+2-n
print '                    Analysis of Variance'
print U=.'---------------------------------------------------------'
print '   Source      Sum.of Squ. df    Mean Squ.  F-value'
print U
print '  Regression',(12.4,fmt)":sR,kk,(sR%kk),(sR%kk)%RES
print '  Error     ',12.4 4.0 12.4":sE,nF,RES
print U
print '  Total     ',12.4 4.0":sT,nT
print U
label_cont.
if. ID=0 do. goto_out. else. goto_in. end.
label_owari.
'END'
)
NB.-----------------------------------------------------------------
NB.  2 6 4 a_step x;y
NB.  sweepout v1;S
NB.-----------------------関数定義--------------------------------
   sweepout=:3 : 0
```

```
'v1 S'=.y.
v2=.-.v1
a11=.v1#v1#"1 S
a12=.|:a21=:v2#v1#"1 S
a22=.v2#v2#"1 S
NB. Sweepout
A1=.ia,-a21 mp ia=.%.a11
A2=.(ia mp a12),a22-a21 mp ia mp a12
A=.A1,"1 A2
v3=.(i.l)(v1#(i.#A))}}.(1,v2)expand 1,(l=.+/v1)}.i.#A NB. Replacement Indices
Tab=.}:v3{v3{"1 A  NB. Replace
)
```

●計算例１：変数の個数ゼロからスタート

```
   step_a X
Means and Standard deviations
Variable  Mean       Std.dev.
   1     19.4467      2.4556
   2     14.5600      1.9769
   3      4.8333      1.6007
   4      7.0200      1.4833
   5      3.2267      1.0103
   6      0.8133      0.0942
   7      0.2965      0.0804
   8      0.1677      0.0572
   9      0.6936      0.1584
  10      0.4873      0.1030
  11     24.0133      2.5551
Correlation Coefficients
  1.0000  0.1125  0.3474  0.2370  0.1814  0.6143 _0.0567 _0.2627 _0.3316  0.1537  0.4881
  0.1125  1.0000  0.7724  0.3442  0.6504  0.1605  0.3025  0.5725 _0.2730 _0.3812 _0.0502
  0.3474  0.7724  1.0000  0.3426  0.6739  0.3463  0.2728  0.4908 _0.5324 _0.2265  0.0151
  0.2370  0.3442  0.3426  1.0000  0.3013  0.3315  0.8904  0.2018 _0.1635  0.7334 _0.0817
  0.1814  0.6504  0.6739  0.3013  1.0000  0.3059  0.3317  0.8970  0.2407 _0.1981 _0.1457
  0.6143  0.1605  0.3463  0.3315  0.3059  1.0000  0.4107  0.0472 _0.1857  0.2168 _0.3853
 _0.0567  0.3025  0.2728  0.8904  0.3317  0.4107  1.0000  0.3708 _0.0162  0.6558 _0.5099
 _0.2627  0.5725  0.4908  0.2018  0.8970  0.0472  0.3708  1.0000  0.3896 _0.2405 _0.3754
 _0.3316 _0.2730 _0.5324 _0.1635  0.2407 _0.1857 _0.0162  0.3896  1.0000  0.0127 _0.1961
  0.1537 _0.3812 _0.2265  0.7334 _0.1981  0.2168  0.6558 _0.2405  0.0127  1.0000 _0.0527
  0.4881 _0.0502  0.0151 _0.0817 _0.1457 _0.3853 _0.5099 _0.3754 _0.1961 _0.0527  1.0000
Step : 1
   Varibles Coefficients Std.Coeff. Stand.Dev.  t-Valus
          7    _16.1940    _0.5099     7.5780    _2.1370
   Constant      28.8154
   Coeff. of Det. 0.25996   Adjusted R2    0.20304
   AIC      28.59165        Mallows Cp      377.86041
                    Analysis of Variance
```

```
-------------------------------------------------------
   Source      Sum.of Squ. df    Mean Squ.  F-value
-------------------------------------------------------
  Regression     23.7600   1     23.7600      4.5667
  Error          67.6374  13      5.2029
-------------------------------------------------------
  Total          91.3973  14
-------------------------------------------------------
Step : 2
   Varibles Coefficients Std.Coeff. Stand.Dev.  t-Valus
          4     3.0942     1.7963     0.2917    10.6090
          7   _66.9918    _2.1092     5.3778   _12.4572
   Constant      22.1577
   Coeff. of Det. 0.92870   Adjusted R2    0.91682
   AIC      _4.50553       Mallows Cp       28.46503
                    Analysis of Variance
-------------------------------------------------------
   Source      Sum.of Squ. df    Mean Squ.  F-value
-------------------------------------------------------
  Regression     84.8808   2     42.4404     78.1523
  Error           6.5166  12      0.5430
-------------------------------------------------------
  Total          91.3973  14
-------------------------------------------------------
Step : 3
   Varibles Coefficients Std.Coeff. Stand.Dev.  t-Valus
          4     3.0548     1.7734     0.2695    11.3357
          6    _3.7627    _0.1387     2.1195    _1.7753
          7   _64.5365    _2.0319     5.1416   _12.5518
   Constant      24.7661
   Coeff. of Det. 0.94458   Adjusted R2    0.92946
   AIC      _6.28448       Mallows Cp       22.12153
                    Analysis of Variance
-------------------------------------------------------
   Source      Sum.of Squ. df    Mean Squ.  F-value
-------------------------------------------------------
  Regression     86.3320   3     28.7773     62.4937
  Error           5.0653  11      0.4605
-------------------------------------------------------
  Total          91.3973  14
-------------------------------------------------------
Step : 4
   Varibles Coefficients Std.Coeff. Stand.Dev.  t-Valus
          1     1.3175     1.2662     0.1652     7.9767
          4    _0.2685    _0.1559     0.4294    _0.6253
          6   _32.2638    _1.1891     3.6657    _8.8014
          7     6.0025     0.1890     9.0637     0.6623
   Constant      24.7365
   Coeff. of Det. 0.99247   Adjusted R2    0.98946
```

```
   AIC     _34.23116       Mallows Cp       _1.04479
                    Analysis of Variance
--------------------------------------------------------
   Source      Sum.of Squ. df    Mean Squ.  F-value
--------------------------------------------------------
  Regression     90.7094   4     22.6773    329.6325
  Error           0.6880  10      0.0688
--------------------------------------------------------
  Total          91.3973  14
--------------------------------------------------------
Step : 5
   Varibles Coefficients Std.Coeff. Stand.Dev.  t-Valus
          1     1.2173     1.1699     0.0389    31.2642
          6   _30.0861    _1.1088     1.1118   _27.0610
          7     0.3739     0.0118     1.0285     0.3636
   Constant      24.6979
   Coeff. of Det. 0.99218   Adjusted R2    0.99005
   AIC     _35.65586       Mallows Cp       _2.89015
                    Analysis of Variance
--------------------------------------------------------
   Source      Sum.of Squ. df    Mean Squ.  F-value
--------------------------------------------------------
  Regression     90.6825   3     30.2275    465.1319
  Error           0.7149  11      0.0650
--------------------------------------------------------
  Total          91.3973  14
--------------------------------------------------------
Step : 6
   Varibles Coefficients Std.Coeff. Stand.Dev.  t-Valus
          1     1.2112     1.1641     0.0339    35.7631
          6   _29.8575    _1.1004     0.8832   _33.8069
   Constant      24.7411
   Coeff. of Det. 0.99208   Adjusted R2    0.99077
   AIC     _37.47670       Mallows Cp       _4.84077
                    Analysis of Variance
--------------------------------------------------------
   Source      Sum.of Squ. df    Mean Squ.  F-value
--------------------------------------------------------
  Regression     90.6739   2     45.3369    752.0168
  Error           0.7234  12      0.0603
--------------------------------------------------------
  Total          91.3973  14
--------------------------------------------------------
END
```

●計算例２：ランダムに７変数を選んでスタート

```
   (print >:7?10)step_a X
8 2 1 5 6 7 4
Means and Standard deviations
```

```
Variable  Mean       Std.dev.
   1     19.4467      2.4556
   2     14.5600      1.9769
   3      4.8333      1.6007
   4      7.0200      1.4833
   5      3.2267      1.0103
   6      0.8133      0.0942
   7      0.2965      0.0804
   8      0.1677      0.0572
   9      0.6936      0.1584
  10      0.4873      0.1030
  11     24.0133      2.5551
Correlation Coefficients
  1.0000  0.1125  0.3474  0.2370  0.1814  0.6143 _0.0567 _0.2627 _0.3316  0.1537  0.4881
  0.1125  1.0000  0.7724  0.3442  0.6504  0.1605  0.3025  0.5725 _0.2730 _0.3812 _0.0502
  0.3474  0.7724  1.0000  0.3426  0.6739  0.3463  0.2728  0.4908 _0.5324 _0.2265  0.0151
  0.2370  0.3442  0.3426  1.0000  0.3013  0.3315  0.8904  0.2018 _0.1635  0.7334 _0.0817
  0.1814  0.6504  0.6739  0.3013  1.0000  0.3059  0.3317  0.8970  0.2407 _0.1981 _0.1457
  0.6143  0.1605  0.3463  0.3315  0.3059  1.0000  0.4107  0.0472 _0.1857  0.2168 _0.3853
 _0.0567  0.3025  0.2728  0.8904  0.3317  0.4107  1.0000  0.3708 _0.0162  0.6558 _0.5099
 _0.2627  0.5725  0.4908  0.2018  0.8970  0.0472  0.3708  1.0000  0.3896 _0.2405 _0.3754
 _0.3316 _0.2730 _0.5324 _0.1635  0.2407 _0.1857 _0.0162  0.3896  1.0000  0.0127 _0.1961
  0.1537 _0.3812 _0.2265  0.7334 _0.1981  0.2168  0.6558 _0.2405  0.0127  1.0000 _0.0527
  0.4881 _0.0502  0.0151 _0.0817 _0.1457 _0.3853 _0.5099 _0.3754 _0.1961 _0.0527  1.0000
8 2 1 5 6 7 4
Step : 1
   Varibles Coefficients Std.Coeff. Stand.Dev.  t-Valus
          1     1.4440     1.3878     0.1587     9.0963
          2    _0.6786    _0.5250     0.2640    _2.5701
          4     0.9079     0.5271     0.6481     1.4009
          5    _0.5546    _0.2193     0.6441    _0.8610
          6   _33.1476    _1.2216     3.2785   _10.1105
          7    10.0285     0.3157     8.5810     1.1687
          8     5.6451     0.1264    11.7786     0.4793
         10   _20.0480    _0.8080     7.2604    _2.7613
   Constant      34.0351
   Coeff. of Det. 0.99739   Adjusted R2    0.99391
   AIC     _42.12336       Mallows Cp        4.37102
                    Analysis of Variance
--------------------------------------------------------
   Source      Sum.of Squ. df    Mean Squ.  F-value
--------------------------------------------------------
  Regression     91.1589   8     11.3949    286.6979
  Error           0.2385   6      0.0397
--------------------------------------------------------
  Total          91.3973  14
--------------------------------------------------------
Step : 2
   Varibles Coefficients Std.Coeff. Stand.Dev.  t-Valus
```

```
          1     1.4114     1.3564     0.1353    10.4314
          2    _0.6855    _0.5304     0.2487    _2.7560
          4     0.8484     0.4925     0.6001     1.4138
          5    _0.2490    _0.0984     0.0850    _2.9277
          6   _33.6417    _1.2398     2.9360   _11.4584
          7    11.5925     0.3650     7.4868     1.5484
         10   _20.2129    _0.8146     6.8416    _2.9544
   Constant      35.1661
   Coeff. of Det. 0.99729   Adjusted R2    0.99458
   AIC     _43.55984       Mallows Cp       2.42350
                    Analysis of Variance
--------------------------------------------------------
   Source      Sum.of Squ. df    Mean Squ.  F-value
--------------------------------------------------------
  Regression     91.1497   7     13.0214    368.1324
  Error           0.2476   7      0.0354
--------------------------------------------------------
  Total          91.3973  14
--------------------------------------------------------
Step : 3
   Varibles Coefficients Std.Coeff. Stand.Dev.  t-Valus
          1     1.5311     1.4715     0.1119    13.6847
          2    _0.3920    _0.3033     0.1453    _2.6972
          5    _0.2416    _0.0955     0.0900    _2.6837
          6   _36.3319    _1.3390     2.3713   _15.3213
          7    18.2561     0.5748     6.1694     2.9591
         10   _12.4076    _0.5000     4.2855    _2.8953
   Constant      30.9050
   Coeff. of Det. 0.99652   Adjusted R2    0.99391
   AIC     _41.79220       Mallows Cp       0.82997
                    Analysis of Variance
--------------------------------------------------------
   Source      Sum.of Squ. df    Mean Squ.  F-value
--------------------------------------------------------
  Regression     91.0790   6     15.1798    381.5236
  Error           0.3183   8      0.0398
--------------------------------------------------------
  Total          91.3973  14
--------------------------------------------------------
END
}
```