予測平方和による変数選択

1. まえがき

重回帰分析における変数選択は，古くて新しい問題である．表1に示したようなp個の説明変数$x_1, x_2, \cdots, x_p$とひとつの目的変数yについてのn組のデータが得られたとき，yの$(x_1, x_2, \cdots, x_p)$に対する重回帰式：

$$\hat{y}=b_0+b_1x_1+b_2x_2+\cdots+b_px_p \qquad (1)$$

の係数$\{b_i; i=0, 1, \cdots, p\}$は，最小二乗法によって求められる．すなわち，残差平方和 Residual Sum of Squares(RSS)を最小にする，または，重相関係数Rを最大にするという規準によって定まる．

$$RSS=\sum_{\alpha=1}^{n}(y_\alpha-\hat{y}_\alpha)^2, \quad R^2=1-RSS/S_{yy} \qquad (2)$$

ただし　$S_{yy}=\sum_{\alpha=1}^{n}(y_\alpha-\bar{y})^2$　(yの偏差平方和)

$$\hat{y}_\alpha=b_0+b_1x_{\alpha1}+b_2x_{\alpha2}+\cdots+b_px_{\alpha p} \qquad (3)$$

ところで，原料の諸特性や工程の諸条件を説明変数として製品の品質特性を予測したり制御したりしようとする工程解析や，種々のマクロ指標や社会指標から特定の商品の需要を予測しようとする試みにおいては，説明変数の数pは一般には20にも30にもなるであろう．そのうえ，式(1)は，計測された変数 $u_1, u_2, \cdots$ などについての1次式である必要はなく，$x_1=u_1$，$x_2=u_1^2$，$x_3=\log u_2$，$x_4=u_1u_3$，というように2乗や積やいろいろの関数であってもよい．したがって，式(1)にとりこまれる変数の数はますます増え，pは100以上にもなるケースが多い．

表1　重回帰分析のためのデータと統計量

No.	説明変数						目的変数
	x_1	x_2	$\cdots$	x_i	$\cdots$	x_p	y
1	x_{11}	x_{12}	$\cdots$	x_{1i}	$\cdots$	x_{1p}	y_1
2	x_{21}	x_{22}	$\cdots$	x_{2i}	$\cdots$	x_{2p}	y_2
⋮	⋮	⋮		⋮		⋮	⋮
α	$x_{\alpha1}$	$x_{\alpha2}$	$\cdots$	$x_{\alpha i}$	$\cdots$	$x_{\alpha p}$	y_α
⋮	⋮	⋮		⋮		⋮	⋮
n	x_{n1}	x_{n2}	$\cdots$	x_{ni}	$\cdots$	x_{np}	y_n
計	T_1	T_2	$\cdots$	T_i	$\cdots$	T_p	T_y
平均	$\bar{x}_1$	$\bar{x}_2$	$\cdots$	$\bar{x}_i$	$\cdots$	$\bar{x}_p$	$\bar{y}$

周知のように，説明変数を新しく追加すると，その変数がyに対して固有の説明力をもたなくても，式(2)の RSS は減少し，R^2 は増大する．$p=n-1$になると，$RSS=0$，$R^2=1$になり，n個の y の値の変動は，$(n-1)$個の変数によって完全に説明しつくされることになる．これはちょうど，$n=2$個のデータには $p=1$次式が，$n=3$個のデータには $p=2$ 次式が完全にあてはまるのと同じで，$n=10$の時系列データは九つの説明変数で完全に説明されるのである．このように多数の説明変数を用いた重回帰式は，書きおろすだけでも何十行にもわたり，それが将来の予測に役立つとは，すなわち，再現性があるとは到底考えられない．この手法のユーザーが期待するのは，たまたま手もとにあるn組のデータによくあてはまることではなく，将来出現するであろう値の予測がよくあたることである．とすると，このp個の変数は，重回帰式にとりこむべき説明変数の候補変数であり，われわれはこのなかから k 個 $(k\leq p)$ の変数を選択したいことになる．

この選択のアルゴリズムについては，p変数のあらゆる部分集合についての重回帰を効率的に計算する方法とか，変数を一つずつ増加または減少

させてゆく逐次選択法が工夫されている．部分集合の総数は 2^p-1 個あり，$p\leq 13$ の範囲では，1万以下であるから実行可能であるが，$p=30$ にもなると約10億とおりとなり，どうしても逐次選択によらざるをえない．このとき，従来の方法では，各ステップで best one を選ぶ局地的最適化を行なうので，あらゆる組合せを検討したときの最適解を見逃す危険が大きかった．各ステップで best five を選ぶというような修正を加え，かつ，変数増減または減増法[1]を採用すれば，多くの場合最適解に到達する．また，このような自動的選択の結果を参照して，固有技術・実質科学の立場から，より適切と思われるモデルを選択することが大切である．そのようなモデルの重相関係数 R が少々小さくなっても差支えないのである．

ここで，変数選択の規準についても考えなおしてみる必要がある．前にも述べたように，RSS や R^2 を用いる限り，説明変数は多ければ多いほど良いということになるから，その「良さ」にある限界を設けて，どこかで打切らねばならない．これについては，本誌小柳[2]を参照されたい．

2. 予測平方和による選択

"n 組のデータにもとづいて将来の値を予測する"ことを，手もとの n 組のデータのなかで模擬するためには，たとえば，No. 1 の y_1 を予測するのに，No. 1 を除く残りの $(n-1)$ 組のデータにもとづいて重回帰式を計算し，それへ $(x_{11}, x_{12}, \cdots, x_{1p})$ を代入すればよいと考えられる．No. 2 の y_2 を予測するには，No. 2 を除いた $(n-1)$ 組のデータから重回帰式を求める．こうすると，毎回少しずつ違った重回帰式が得られるであろうから，計算は n 回繰返すことになる．この方法で求めた y_α の予測値を $\hat{y}_\alpha{}^*$ であらわすと，

$$\hat{y}_\alpha{}^*=b_{0\alpha}+b_{1\alpha}x_{\alpha 1}+b_{2\alpha}x_{\alpha 2}+\cdots+b_{p\alpha}x_{\alpha p} \qquad (4)$$

と書ける．式(3)との相違は，係数 $b_{i\alpha}$ が α によって異なることである．このとき，予測平方和 Prediction Sum of Squares (PSS) は次式で定義される：

$$PSS=\sum_{\alpha=1}^{n}(y_\alpha-\hat{y}_\alpha{}^*)^2 \qquad (5)$$

概念としてこれは大変おもしろい規準であり，かつ後出の**図2**に見るように，変数の数を増してゆくと，PSS はあるところから先は増大しはじめる．つまり，極小値が存在するのである．したがって，打切り規準などをもちこまなくても，変数の数の少ないモデルが選ばれ，実用的には，それはおそらく再現性の高い，良いモデルになるであろうと期待される．しかし，この方法の数学的 justification はまだなされていない．

PSS がいかに興味ある規準であっても，n 回重回帰分析を行なうのでは，その計算量は尨大になる．$n=50$ で，計算時間は30倍くらいになった．しかし，この n 組のデータの manipulation によって，PSS をもっと簡単に求める方法があるはずである．D. Allen [4] の結果を知らないで，芳賀・竹内・奥野[5]は次式を導いた[1]：

$$PSS=\sum_{\alpha=1}^{n}\left(\frac{y_\alpha-\hat{y}_\alpha}{1-c_\alpha}\right)^2 \qquad (6)$$

ただし，

---

1) この証明を簡単に与えておこう．

$\hat{\boldsymbol{y}}=\boldsymbol{Z}\boldsymbol{b}$, $\hat{\boldsymbol{y}}^*=\boldsymbol{Z}\boldsymbol{b}_\alpha$ ただし $\boldsymbol{b}'=(b_0, b_1, \cdots, b_p)$

$\boldsymbol{M}=\boldsymbol{Z}'\boldsymbol{Z}$, $\boldsymbol{Z}$ の行ベクトル $\boldsymbol{z}_\alpha{}'=(1, x_{\alpha 1}, \cdots, x_{\alpha p})$

とおくと，両方の場合の正規方程式は

$$\boldsymbol{M}\boldsymbol{b}=\boldsymbol{Z}'\boldsymbol{y}$$

$$(\boldsymbol{M}-\boldsymbol{z}_\alpha\boldsymbol{z}_\alpha{}')\boldsymbol{b}_\alpha=\boldsymbol{Z}'\boldsymbol{y}-\boldsymbol{z}_\alpha y_\alpha$$

となり，辺々減算すると，($\boldsymbol{M}$ は正則とする)

$$\boldsymbol{M}(\boldsymbol{b}-\boldsymbol{b}_\alpha)=\boldsymbol{z}_\alpha(y_\alpha-\boldsymbol{z}_\alpha{}'\boldsymbol{b}_\alpha)=\boldsymbol{z}_\alpha(y_\alpha-\hat{y}_\alpha{}^*)$$

$$\therefore \hat{y}_\alpha-\hat{y}_\alpha{}^*=\boldsymbol{z}_\alpha{}'(\boldsymbol{b}-\boldsymbol{b}_\alpha)=\boldsymbol{z}_\alpha{}'\boldsymbol{M}^{-1}\boldsymbol{z}_\alpha(y_\alpha-\hat{y}_\alpha{}^*)$$

を得る．ここで，$c_\alpha=\boldsymbol{z}_\alpha{}'\boldsymbol{M}^{-1}\boldsymbol{z}_\alpha$ とおくと，

$$y_\alpha-\hat{y}_\alpha{}^*=(y_\alpha-\hat{y}_\alpha)/(1-c_\alpha)$$

となる．また，平方和・積和行列を $\boldsymbol{S}=(n-1)\boldsymbol{V}$ とおくと，

$$\boldsymbol{M}^{-1}=\begin{pmatrix}\frac{1}{n}+\bar{\boldsymbol{x}}'\boldsymbol{S}^{-1}\bar{\boldsymbol{x}} & -\bar{\boldsymbol{x}}'\boldsymbol{S}^{-1}\\ -\boldsymbol{S}^{-1}\bar{\boldsymbol{x}} & \boldsymbol{S}^{-1}\end{pmatrix}$$

となり，これを用いると，つぎを得る：

$$c_\alpha=\frac{1}{n}+(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'\boldsymbol{S}^{-1}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})=\frac{1}{n}+\frac{D_\alpha{}^2}{n-1}.$$

$$c_\alpha=\frac{1}{n}+\frac{D_\alpha^2}{n-1}, \tag{7}$$

$D_\alpha^2=(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'\,\boldsymbol{V}^{-1}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})$

D_α: $(x_{\alpha1}, x_{\alpha2}, \cdots, x_{\alpha p})$ と $(\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_p)$ とのマハラノビス汎距離

$\boldsymbol{V}$：分散・共分散行列

この式は，PSS が残差 $y_\alpha-\hat{y}_\alpha$ の重みづき2乗和で，その重みは，観測点 $\boldsymbol{x}_\alpha$ と重心 $\bar{\boldsymbol{x}}$ とのマハラノビス汎距離 D_α が大きいほど大きいことを示している．つまり，重心より離れた点へのあてはまりの悪いモデルはより強く排除されるのである．

3. 他の選択規準との関係

式(1)に対応する回帰モデルとして，

$$y_\alpha=\beta_0+\sum_{i=1}^{p}\beta_i x_{\alpha i}+\epsilon_\alpha,\quad \epsilon_\alpha\sim N(0,\sigma^2) \tag{8}$$
$$=\boldsymbol{z}_\alpha'\,\beta+\epsilon_\alpha$$

を考える．すると，

$$E[RSS]=E\Big[\sum_{\alpha=1}^{n}(y_\alpha-\hat{y}_\alpha)^2\Big]$$
$$=\sum_{\alpha=1}^{n}(1-c_\alpha)\sigma^2=(n-p-1)\sigma^2 \tag{9}$$

であることはよく知られている[2]．PSS の期待値はつぎのように評価される．

$$E[PSS]=E\Big[\sum_{\alpha=1}^{n}(y_\alpha-\hat{y}_\alpha)^2/(1-c_\alpha)^2\Big]$$
$$=\sigma^2\sum_{\alpha=1}^{n}1/(1-c_\alpha) \tag{10}$$
$$\geq\sigma^2\sum_{\alpha}(1+c_\alpha)=(n+p+1)\sigma^2$$
$$=E[RSS]+2(p+1)\sigma^2$$

(1) 自由度二重調整重相関係数　doubly-adjusted multiple correlation coefficient

重相関係数 R^2 は，式(2)で定義した．この右辺第2項の分母・子を自由度で調整して分散に直したものを，自由度調整重相関係数とよび，R^* であらわす[3]．

$$R^{*2}=1-\frac{RSS/(n-p-1)}{S_{yy}/(n-1)}$$
$$=1-\frac{n-1}{n-p-1}(1-R^2) \tag{11}$$

ここでは，式(9)と(10)を比較して，PSS の期待値の下限は，$E[RSS]$ の $(n+p+1)/(n-p-1)$ 倍であることに注意し，これを式(11)の右辺第2項の分子に，またそこで $p=0$ としたときの $(n+1)/(n-1)$ 倍を S_{yy} に掛けてその分母においたものを，自由度二重調整として R^{**2} であらわす[3]．

$$R^{**2}=1-\Big(\frac{n+p+1}{n-p-1}RSS\Big)\Big/\Big(\frac{n+1}{n-1}S_{yy}\Big)$$
$$=1-\frac{n+p+1}{n+1}\,\frac{n-1}{n-p-1}(1-R^2) \tag{12}$$

明らかに $R^2\geq R^{*2}\geq R^{**2}$ で，R^*, R^{**} は変数を増しても大きくなるとは限らない．

(2) F 統計量による打切り規準

すでに取り入れられている $(k-1)$ 変数にさらに1変数を加えて k 変数とするとき，または，k 変数のなかから一つを除いて $(k-1)$ 変数にするとき，その取捨は，つぎの F_k 統計量によって判断されるのがふつうである．

$$F_k=\frac{(RSS)_{k-1}-(RSS)_k}{(RSS)_k/(n-k-1)}=\frac{R^2_k-R^2_{k-1}}{(1-R_k^2)/(n-k-1)} \tag{13}$$

この F_k が，自由度 $(1, n-k-1)$ の F 分布に従うことを利用して，有意性の検定を行なうことができる．

また，このとき，つぎの関係が成立する：

$$\text{①}\quad R_k^2\geq R^2_{k-1}\quad\Leftrightarrow F_k\geq 0\ \text{（いつでも成立）}$$
$$\text{②}\quad R_k^{*2}\geq R_{k-1}^{*2}\quad\Leftrightarrow F_k\geq 1.0 \tag{14}$$
$$\text{③}\quad R_k^{**2}\geq R_{k-1}^{**2}\Leftrightarrow F_k\geq\frac{2n}{n+k}$$

この関係は容易に証明することができる，n が k より充分大きいとき，③は $F_k\geq 2.0$ を示している．変数の逐次選択で，F の有意点をいちいち参照しないで，$F_{IN}=F_{OUT}=2.00$ [1] ととっているのは，R_k^{**2} が増大するという規準に対応している．

2) $\sum_{\alpha=1}^{n}c_\alpha=\mathrm{tr}(\boldsymbol{Z}\boldsymbol{M}^{-1}\boldsymbol{Z}')=\mathrm{tr}(\boldsymbol{M}^{-1}\boldsymbol{M})=\mathrm{tr}(I_{p+1})=p+1$

(3) 予測平均二乗誤差 Mean Square Error of Prediction (*MSEP*)

将来の観測点が $z_0'=(1, x_{01}, x_{02}, \cdots, x_{0p})$ であり，そこでの実現値を y_0 とし，かつ，採用したモデルは変数の一部だけをとったものとして，式(8)に「偏り」の項 γ_α を加え，

$$y_\alpha=\boldsymbol{z}_\alpha'\boldsymbol{\beta}+\gamma_\alpha+\epsilon_\alpha \tag{15}$$

ただし $\boldsymbol{z}_\alpha'\boldsymbol{\gamma}=0$, $\boldsymbol{\gamma}'=(\gamma_1, \gamma_2, \cdots, \gamma_n)$

$$\epsilon_\alpha \sim \mathrm{N}(0, \sigma^2)$$

とする．このとき，

$$\hat{y}_0=\boldsymbol{z}_0'\boldsymbol{b}$$

とすると，

$$\begin{aligned} MSEP_0&=E[(y_0-\hat{y}_0)^2] \\ &=V[y_0]+V[\hat{y}_0]+(E[y_0]-E[\hat{y}_0])^2 \\ &=(1+c_0)\sigma^2+\gamma_0 \end{aligned} \tag{16}$$

いま，$\boldsymbol{z}_0$ として，手もとの n 個の観測点の全体 $\boldsymbol{Z}$ をとり，その $MSEP$ の和を $TMSEP$ と書くと，

$$\begin{aligned} TMSEP&=\sigma^2\sum(1+c_\alpha)+\sum\gamma_\alpha{}^2 \\ &=(n+p+1)\sigma^2+\Gamma(\text{とおく}) \end{aligned} \tag{17}$$

を得る．一方，モデル(15)の下では，

$$E[RSS]=(n-p-1)\sigma^2+\Gamma \tag{18}$$

と書けるから，

$$TMSEP=E[RSS]+2(p+1)\sigma^2 \tag{19}$$

となり，式(10)で求めた $E[PSS]$ の下限と一致する．

(4) Mallows の C_p 統計量

これについては，本誌佐和[6]にくわしい．モデル(15)の下で，$\hat{y}_\alpha(\alpha=1, 2, \cdots, n)$ の二乗誤差の和 Total Squared Error (TSE) を求めると，

$$\begin{aligned} TSE&=\sum_\alpha\{V[\hat{y}_\alpha]+(E[\hat{y}_\alpha]-\boldsymbol{z}'_\alpha\beta-\gamma_\alpha)^2\} \\ &=\sum(c_\alpha\sigma^2+\gamma_\alpha{}^2)=(p+1)\sigma^2+\Gamma \\ &=E[RSS]+\{2(p+1)-n\}\sigma^2 \\ &\qquad\text{(式(18)より)} \end{aligned} \tag{20}$$

となる．これを σ^2 で割って標準化したものを Γ_p とおくと，その推定値が

$$C_p=\frac{RSS}{\sigma^2}+2(p+1)-n \tag{21}$$

となる．ここでは，σ^2 をどのようにして推定するかが問題となるが，σ^2 を既知とすれば，この期待値は前項の $TMSEP$ や $E[PSS]$ の下限と本質的に一致する．

(5) 情報量規準 AIC

AIC は次式で与えられる[3]．

$AIC=-2\ln$(最大尤度)
$\quad+2$(推定したパラメータの数)

ここで，式(8)の回帰モデルの下で，σ^2 を既知とすれば，

$$\begin{aligned} \ln(\text{最大尤度})&=\max_{\boldsymbol{\beta}}\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{Z\beta})'(\boldsymbol{y}-\boldsymbol{z\beta})\right. \\ &\left.-\frac{n}{2}\ln\sigma^2+\text{const}\right\}=-\frac{RSS}{2\sigma^2}+\text{const} \end{aligned} \tag{22}$$

よって，この場合には

$$AIC=\frac{RSS}{\sigma^2}+2(p+1)=\frac{1}{\sigma^2}\,TMSEP \tag{23}$$

となる．

以上から本節で与えた五つの規準は漸近的には，まったく等しく，かつ，予測平方和の期待値 $E[PSS]$ の下限にあたることがわかった．よって，つぎの計算例では，これらの代表として R^{**} のみを示し，これと PSS 自身を選択規準に用いる場合とを比較する．

4. 数値例と実施例

［例 1］ 多項式回帰——**図 1** に示す6本の曲線(これをケース[1]～[6]とよぶ)に多項式をあてはめる場合を考える．$n=10$ 点であるから9次の多項式をあてはめれば，10点を完全に通るが，その曲線は上下に大きく振動し，観測点以外のところでは予測誤差が大きくなると予想される．

多項式あてはめは，重回帰式あてはめの特殊の場合であるから，多項式の次数の順序にこだわらず，1次から7次までの7変数をとり，このなかから変数選択を行なってみる．このとき，直交多項式 $\phi_1(x), \phi_2(x), \cdots, \phi_7(x)$ をとれば，説明変数が相互に完全に無相関の場合 (case A とよぶ) にあたり，通常のベキ，$x, x^2, \cdots, x^7$ をとれば，説

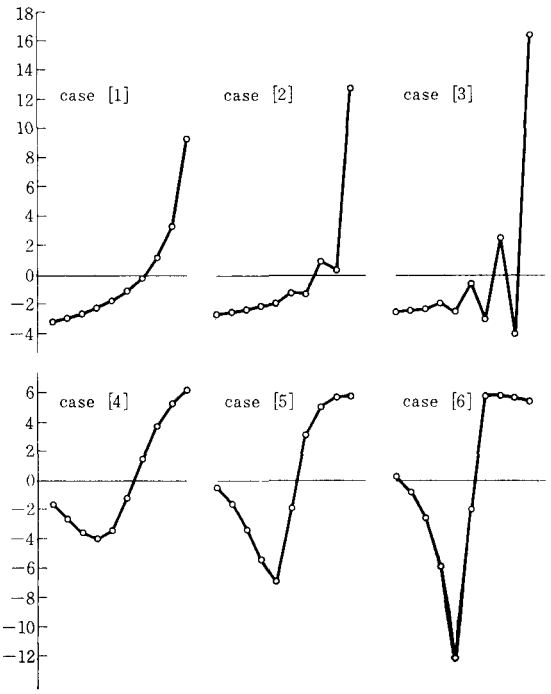

図 1 ケース［1］～［6］の曲線

明変数間に高い相関がある場合（case B とよぶ）に相当する．その相関係数を**表 2** に示す．

case A で $\phi_i(x)$ $(i=1,2,\cdots,7)$ と y との相関を見れば，case［1］と［4］では，1/2 ずつ，［2］と［5］で $1/\sqrt{2}$ ずつ，［3］と［6］では $1/2^{1/4}$ ずつ次数とともに小さくなっている．これに対応して，分散分析表は，**表 3** のようになる．case B では，説明変数間の相関は相隣る次数の間では大体0.98以上であり，次数が離れるにつれて減少するが，いちばん遠い x と x^7 の間にも0.770という高い相関がある．y との相関も case A の場合より高く，最高の相関は，case［1］では x^5 と，［2］［3］では x^7 と，［4］［5］［6］では x^3 との間に見られる．

さて，このようなデータについて，変数の数 k を $1,2,\cdots,7$ と増しながら，各 k について RSS または PSS の最小な組合せを選んで**表 4**，**表 5** にまとめる．表 3 と，表 4 を比べながら，説明変数間に相関がないとき（caseA）を検討すると，つぎのようなことがわかる．

表 2(a) $\phi_i(x)$ と y との相関係数 case A

	case	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7
	［1］(［4］)	.866	.433	±.217	±.108	.054	.027	±.014
y	［2］(［5］)	.708	.500	±.354	±.250	.177	.125	±.088
	［3］(［6］)	.554	.466	±.391	±.329	.277	.233	±.196

表 2(b) x^i 相互の間および x^i と y との相関 case B

		x	x^2	x^3	x^4	x^5	x^6	x^7
	x	1						
	x^2	.975	1					
	x^3	.928	.987	1				
	x^4	.882	.961	.993	1			
	x^5	.839	.933	.977	.995	1		
	x^6	.802	.904	.958	.985	.997	1	
	x^7	.770	.877	.938	.972	.990	.998	1
	［1］	.866	.941	.975	.988	.992	.989	.983
	［2］	.708	.802	.860	.898	.924	.941	.954
y	［3］	.554	.644	.706	.752	.786	.813	.835
	［4］	.866	.941	.952	.934	.905	.874	.843
	［5］	.708	.802	.822	.807	.777	.743	.710
	［6］	.554	.644	.664	.649	.620	.588	.556

① PSS 規準で選ぶと，case［1］,［4］では5次式，［2］,［5］では3次式，［3］では0次式（x を用いない，定数のみ），［6］では ϕ_1 と ϕ_3 の二つのみを用いればよいという結論になる．

②これに反して，R^2，R^{*2}，R^{**2} のどれを用いてもいつも7次式をあてはめるのが最適であるとの結論に達して，変数を減らすことができない．

③表3を見れば，F 値が5％水準で有意なのは，case［1］,［4］で6次まで，［2］,［5］で4次まで，［3］,［6］では0次ということになっているから，PSS による選択はこれに近い．実際［1］,［4］での6次の項は，7次項と残差をプールした誤差分散

表 3 分 散 分 析 表 (case [A])

変動因	自由度	ケース：[1][4] 平方和 S	分散 V	F	ケース：[2][5] S	V	F	ケース：[3][6] S	V	F
全 体	9	133.334			199.610			326.333		
1 次	1	100.000		***	100.000		170.6**	100.000		13.26
2 次	1	25.000		***	50.000		85.3*	70.711		9.37
3 次	1	6.250		***	25.000		42.7*	50.000		6.63
4 次	1	1.563		391**	12.500		21.3*	35.355		4.69
5 次	1	0.391		97.8*	6.250		10.7	25.000		3.31
6 次	1	0.098		24.5*	3.125		5.33	17.678		2.34
7 次	1	0.024		6.0	1.563		2.67	12.500		1.66
残 差	2	0.008	0.004		1.172	0.586		15.089	7.544	

$(0.024+0.008)/3=0.0107$ を用いると $F=9.19$ となって $F(1,3;0.05)=10.13$に達しない．また[2],[5]の4次項も，5,6,7次と残差をこみにした誤差分散に対しては，$F=5.16$ となって5％有意点 $F(1,5;0.05)=6.61$ に達しない．一方，case [3],[6]では，高次の項をどのようにプールしても，1次項すら5％有意とならないから，定数項(水平線)をあてはめるのがもっとも良いのであるが，[6]のとき PSS は ϕ_1 または ϕ_1 と ϕ_3 の組をとることを勧めている．しかし，R^{**2}を用

表 4 RSS および PSS で選ばれた最適の組合せ——case A 直交多項式 $\phi_i(x)$を用いる場合

cases 規準 / 変数番号	[1]と[4] R^2	R^{*2}	R^{**2}	[1] PSS	[4] PSS	[2]と[5] R^2	R^{*2}	R^{**2}	[2] PSS	[5] PSS	[3]と[6] R^2	R^{*2}	R^{**2}	[3] PSS	[6] PSS
0 (定数項のみ)	—	—	—	164.61	164.61	—	—	—	246.43	246.43	—	—	—	402.88	402.88
1	75.0	71.9	69.3	66.97	56.91	50.1	43.9	38.8	207.33	153.63	30.6	22.0	14.9	465.18	322.11
7	—	—	—	—	—	—	—	—	—	—	—	—	—	402.36	—
1,2	93.8	92.0	90.5	33.13	25.40	75.2	68.1	62.2	194.03	123.38	52.3	38.7	27.5	555.59	332.18
6,7	—	—	—	—	—	—	—	—	—	—	9.3	—	—	414.29	—
1,3	—	—	—	—	—	—	—	—	—	—	46.0	—	—	—	313.17
1,2,3	98.4	97.7	97.0	18.13	14.07	87.7	81.5	76.5	192.57	121.64	67.6	51.5	38.2	705.00	397.03
4,6,7	—	—	—	—	—	—	—	—	—	—	20.1	—	—	473.64	—
1,3,5	—	—	—	—	—	—	—	—	—	—	53.6	—	—	—	341.46
1,2,3,4	99.6	99.3	99.0	12.52	10.00	93.9	89.1	85.1	247.51	161.79	78.5	61.2	47.2	1185.1	674.99
1,2,3,6	—	—	—	—	—	89.2	80.6	73.6	197.48	—	—	—	—	—	—
1,2,3,7	—	—	—	—	—	88.5	—	—	—	150.15	—	—	—	—	—
1,4,5,7	—	—	—	—	—	—	—	—	—	—	53.0	—	—	577.63	—
1,2,4,6	—	—	—	—	—	—	—	—	—	—	68.6	—	—	—	400.80
1,2,3,4,5	99.9	99.8	99.7	11.57	9.90	97.1	93.4	90.4	422.05	312.12	86.1	68.8	54.6	2655.7	1754.5
1,2,3,6,7	—	—	—	—	—	90.0	77.5	67.3	219.85	—	76.9	—	—	709.88	—
1,2,3,5,7	—	—	—	—	—	91.6	—	—	—	184.62	79.1	—	—	—	539.80
1,2,3,4,5,6	99.98	99.9	99.9	15.33	14.08	98.6	95.9	93.7	1054.2	894.68	91.6	74.6	60.8	8929.3	7124.7
1,2,3,4,5,7	99.92	—	—	—	13.67	—	—	—	—	—	—	—	—	—	—
1,2,3,4,6,7	—	—	—	—	—	96.3	88.9	82.8	338.62	338.62	—	—	—	—	—
1,2,3,5,6,7	—	—	—	—	—	—	—	—	—	—	84.5			1234.5	1234.5
1,2,3,4,5,6,7	99.99	99.97	99.96	32.30	32.30	99.4	97.4	95.7	4196.7	4196.7	95.4	79.2	66.0	48061.0	48061.0

表 5 *RSS* および *PSS* で選ばれた最適組合せ——case B 説明変数間に高い相関がある場合

cases / 規準 / 変数番号	[1] R^2	[1] R^{*2}	[1] R^{**2}	[1] *PSS*	[2] R^2	[2] R^{*2}	[2] R^{**2}	[2] *PSS*	[3] R^2	[3] R^{*2}	[3] R^{**2}	[3] *PSS*
0(定数項のみ)	—	—	—	164.61	—	—	—	246.43	—	—	—	402.88
5	98.3	88.1	97.9	4.89	—	—	—	—	—	—	—	—
6	97.8	97.5	97.3	3.92	—	—	—	—	—	—	—	—
7	—	—	—	—	91.0	89.8	88.9	92.17	69.7	66.0	62.9	486.2
1	—	—	—	—	—	—	—	—	30.6	22.0	14.9	465.2
1,7	99.8	99.7	99.6	1.71	91.1	88.6	86.5	139.90	—	—	—	—
6,7	—	—	—	—	93.4	91.6	90.0	273.09	78.9	72.9	68.0	1226.8
1,2	—	—	—	—	—	—	—	—	52.3	38.7	27.5	555.6
3,6,7	99.9	99.9	99.83	2.91	—	—	—	—	—	—	—	—
5,6,7	99.9	99.8	99.75	0.44	97.4	96.1	95.1	289.73	86.6	79.9	74.4	2183.6
1,2,5	—	—	—	—	91.0	86.4	82.7	182.51	—	—	—	—
1,2,3	—	—	—	—	—	—	—	—	67.6	51.5	38.2	705.0
1,5,6,7	99.95	99.90	98.87	6.85	—	—	—	—	—	—	—	—
4,5,6,7	—	—	—	—	98.1	96.6	95.4	593.30	89.7	81.5	74.8	4485.6
1,2,3,5	—	—	—	—	94.6	90.3	86.8	246.00	—	—	—	—
1,2,3,4	—	—	—	—	—	—	—	—	78.5	61.2	47.2	1185.1
3,4,5,6,7	99.98	99.97	99.95	7.19	98.8	97.2	95.9	840.17	92.0	82.1	73.9	7768.5
1,2,3,4,5	—	—	—	—	97.1	93.4	90.4	422.05	86.1	68.8	54.6	2655.7
2,3,4,5,6,7	99.99	99.96	99.94	14.22	99.1	97.3	95.8	1400.3	93.6	80.9	70.5	14281
1,2,4,5,6,7	99.99	99.96	99.93	13.30	—	—	—	—	—	—	—	—
1,2,3,4,5,6	—	—	—	—	98.6	95.9	93.7	1054.2	91.6	74.6	60.8	8929
1,2,3,4,5,6,7	99.99	99.97	99.96	30.70	99.8	98.2	97.0	4191.1	95.4	79.2	66.0	47995
	[4]				[5]				[6]			
0(定数項のみ)	—	—	—	164.61	—	—	—	246.43	—	—	—	402.88
3	90.5	89.4	88.4	23.56	67.6	63.5	60.2	103.39	44.1	37.1	31.4	270.17
2	88.6	—	—	22.98	64.3	—	—	102.60	41.5	—	—	264.06
6,7	95.6	94.4	93.3	24.27	78.1	71.8	66.7	94.54	55.8	43.1	32.8	303.22
5,6	95.3	—	—	9.82	—	—	—	—	53.8	—	—	221.24
5,7	—	—	—	—	77.1	—	—	71.19	—	—	—	—
2,3,4	99.6	99.4	99.2	4.41	93.6	90.4	87.8	53.85	77.1	65.7	56.3	249.12
1,3,4	99.4	—	—	2.75	91.8	—	—	51.51	73.9	—	—	235.26
3,4,5,6	99.97	99.95	99.93	0.25	—	—	—	—	—	—	—	—
2,4,6,7	99.96	—	—	0.16	—	—	—	—	—	—	—	—
4,5,6,7	—	—	—	—	98.3	96.9	95.7	217.91	89.3	80.7	73.6	588.32
3,4,6,7	—	—	—	—	97.6	—	—	16.04	—	—	—	—
3,5,6,7	—	—	—	—	—	—	—	—	87.7	—	—	132.70
1,3,4,5,7	99.97	99.94	99.91	1.49	—	—	—	—	—	—	—	—
1,3,4,5,6	99.97	—	—	0.62	98.3	—	—	41.77	89.9	—	—	356.51
2,3,4,5,7	—	—	—	—	98.5	96.5	94.9	111.76	90.6	78.9	69.4	873.41
1,2,3,4,5,6	99.98	99.93	99.89	14.08	98.6	95.9	93.7	894.68	91.6	74.6	60.8	7124.7
1,2,3,4,5,6,7	99.99	99.97	99.96	32.28	99.4	97.4	95.7	4193.5	95.4	79.2	66.0	48027

いても $\phi_1 \sim \phi_7$ を全部採用せよというのであるから，それに比べれば，変数の数をずっとしぼっている．このような曲線に高次の多項式をはめて予測に用いると大変な失敗をすることは，その場合の PSS の大きい値からも容易に読みとれる．

表4の[2]，[5]について，PSS と RSS の変化を**図2**に示す．

case B については，比較した規準の間にさらに興味ある相違が見られる．——を引いた最適組合せにふくまれる変数の数に注目すると，つぎのようになる．

① R^{*2}, R^{**2}, PSS の順にその数は，case[1]では，5,5,3，[2]では，7,5,1，[3]では 5,4,0 になる．つまり，R^{**2} は R^{*2} よりも変数の数が0ないし2個少ない組合せを選ぶが，PSS はこれらよりもかなり少ない数の変数を選ぶ．

②ところが，case[4][5][6]では，R^{*2}, R^{**2}, PSS はいずれも4変数の組を選ぶ．しかし，選ばれた変数の組は，PSS によると，R^{*2}, R^{**2} とすこし異なる．

つぎに，選ばれた変数に着目しよう．

① PSS で選ぶ変数の組と RSS, R^{*2}, R^{**2}(これらは共通の変数の組を選ぶ）のそれとは，一般に異なる．case[1]では，1変数を選ぶとき y との相関が最高の x^5 は，RSS, R^{*2}, R^{**2} によって選ばれるが，PSS では x^6 のほうが選ばれる．2変数を選ぶときは，どちらも (x, x^7) であるが，3変数では RSS による (x^3, x^6, x^7) に比べて PSS による (x^5, x^6, x^7) のほうが値はかなり小さい．

② case[3]では，PSS は定数項のとき最小で，どの変数も選ぶなという結論であるが，取りこむ変数の数をふやすと，1，(1,2)，(1,2,3)，(1,2,3,4) と昇べきの順序に選ばれる．しかし，これらの組は，変数の数を決めたとき，RSS ではもっとも悪い組合せである．RSS(または R^{*2}, R^{**2}) では，反対に，7，(7,6)，(7,6,5)，(7,6,5,4)…という順序に選んでいる．

③ case[4][5][6]では，いずれの選択規準でも4変数を選ぶが，RSS の最小のもの(たとえば[5][6]の (4,5,6,7)の組）は PSS がかなり大きいという結果を示している．

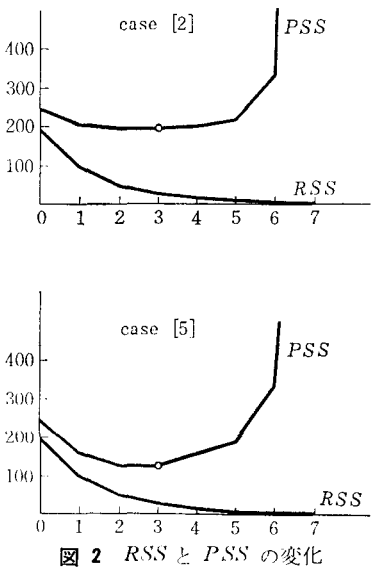

図2 RSS と PSS の変化

以上から，R^{*2}, R^{**2} は変数の数 k を大きくするときの打切り規準を変えるだけで，選ぶ変数の組には差異がないが，PSS はこれらとはまったく別の組の変数を選ぶことがあり，かつ，PSS 最小の組の変数の数は非常に小さくなることがあることがわかった．いろいろの適用例について，いちいちその結果を述べる余裕はないが，つぎの数値は一つの傾向を示している．

これは，メキシコにおけるトウモロコシの試験例で，1962—65年に実施された72カ所の試験で，各試験とも窒素肥料の量を4段階に変えた合計288個のデータにもとづいている．候補として取上げた説明変数は $p=33$ 個で，それらは，施肥窒素量の1次Nおよび2次N^2の項，土壌中窒素量の1次Aおよび2次A^2の項，1次同士の交互作用項N×A，前作物に施用した窒素量Bの1次および2次の項，交互作用N×B，A×B，土壌水分量CとC×N，C×A，C×B，葉の萎凋した日数DとD×N，D×A，D×B，根の深さE，土地の勾配F，土性の指標GとG²，ひょう害H，H×N，H×A，H×B，胴枯れ病J，J×N，J×A，J×B，雑草量L，L×N，L×A，L×

Bであった.

このデータを最初の3年分の $n=228$ と第4年目の $n'=60$ に分け，$n=228$ について，$p=33$ 変数全部を用いた場合と，ふつうの変数増減法で $F_{IN}=F_{OUT}=2.5$ として選んだ $k=15$ 変数を用いた場合，および *PSS* で増減法を適用した $k=9$ 変数の場合の *RSS* と *PSS* を次表に示す．また，この三つのモデルを $n'=60$ の次年度のデータに適用したときの予測二乗誤差を示す．

変数の数 k	33	15	9
RSS	67,521	80,134	91,806
PSS	93,819	92,868	100,794
予測二乗誤差	67,661	42,395	30,846

PSS によって選ばれた9変数は，N, N^2, A, C, D, D×N, F, H, J で，技術的には非常に解釈しやすいものであった．$k=15$ の場合は A^2 と A×N をふくむのに A がなかったり，H の係数が正になったりして，その解釈に困惑するような変数をいくつも含んでいた．これからも *PSS* 選択の良さが示唆された．

参考文献

[1] 奥野・芳賀・久米・吉沢(1971)：多変量解析法，日科技連出版社

[2] 小柳義夫(1978)：ロバスト推定法とデータ解析への応用（本誌）

[3] 奥野ほか(1976)：続多変量解析法，日科技連出版社

[4] D. Allen (1971) : The Prediction sum of squares as a criterion for selecting predictor variables, Univ.of Kentucky, Dept. of Statistics, Technical Report. No. 23. (1977年に入手)

[5] 芳賀敏郎・竹内啓・奥野忠一(1976)：重回帰分析における変数選択の新しい規準，「品質」 vol. 6, No. 2, pp. 35—40.

[6] 佐和隆光(1978)：回帰分析における説明変数選択のための諸基準（本誌）

おくの・ただかず　1922年生
1944年　東京大学理学部数学科卒業
農林省農業技術研究所を経て，現在東京大学工学部計数工学科教授