偏回帰モデルについて

研究ノート

偏回帰モデルについて

志村　健一

1．はじめに

偏回帰プロットの拡張については最初品質管理学会の研究会で報告[1]を行なった。その後、別のある研究会で紹介を行なったとき、後の質疑で偏回帰モデルの基本的性質の証明とその応用がPark and Wakimoto (1987)[2]にあることを教えて頂いた。そのとき「数学的にはPark and Wakimotoと同様であり、十分なオリジナリティは主張できないのでは・・・」という意見を頂き、志村(1991)[3]ではその点を考慮し、「またPark and Wakimotoでは変数グループについての偏回帰モデル（partial regression model）を考え、・・・関係式を導いている。その後、CookのDに対応する拡張を行なっているが、偏回帰プロットという立場からの議論は行なわれていない。」[4]あるいは、本論の(3)と(4)の関係式の証明に対して「・・・このことは掃き出しの過程を辿ることにより得られるが、証明としては例えばPark and Wakimotoがある。」[5]とした。ここでの意図は、偏回帰プロットの数学的証明は複雑なものではなく、これでオリジナリティを主張するものではないこと、この関係式の利用の仕方が問題であることを示すつもりであった。そのため当初考えた付録1に示したような証明は省略した。正規方程式を掃き出し法で解く過程の認識があれば、証明はなくてよいと考えていた。

しかしその後の議論を通して志村（1991）[3]の記述だけではこれらの意図は十分に伝わらないと感じるようになった。特に(2)から(4)の関係式の証明にあたる部分の説明は不十分であると感じ、本論においてこの関係式の証明とその応用について再論を試みる。

2．偏回帰モデルについて

ここで使用する回帰式を以下のように書く。nをデータ数とし、pを取り上げた説明変数の数とする。

$$y = b_0 \cdot x_0 + b_1 \cdot x_1 + \cdots + b_p \cdot x_p + e \tag{1}$$

ここでyは目的変数、x_i（$i=0,\cdots,p$）は説明変数、eは残差を表わし、これらはn次元ベクトルとする。b_i（$i=0,\cdots,p$）は推定された回帰係数とする。偏回帰プロットを拡張する場合、次の回帰式を使用する。

$$y_{\cdot I} = c_{m+1} \cdot x_{m+1 \cdot I} + c_{m+2} \cdot x_{m+2 \cdot I} + \cdots + c_p \cdot x_{p \cdot I} + e^* \tag{2}$$

ここで　Iは{0,…,m}なる添え字集合で、$y_{\cdot I}$、$x_{i \cdot I}$はそれぞれ添え字集合に対応する説明変数による回帰残差を表わすものとする。この式をPark and Wakimoto（1987）[2]は偏回帰モデルと呼んでいる。

回帰式(1)と(2)には次のような基本的な関係が成り立つ。

$$c_i = b_i \quad (i = m+1, \cdots, p) \tag{3}$$

$$e^* = e \tag{4}$$

3．関係式の証明とモデルの応用について

ここではまず(3)と(4)の関係式の証明について考え、(1)と(2)のモデルの意味をより明確にする証明方法を示す。次に偏回帰モデルとその応用について従来どのように考えられていたかをまとめる。

1）偏回帰モデルの基本関係式の証明について

まずPark and Wakimoto（1987）[2]論文のappendixにある証明方法について考える。同論文では(2)を偏回帰モデルと呼び、これを基本として議論を進めている。この点は特徴と考えられるが、appendixにある証明方法自体は、例えば佐和（1979）[6]などで知られている。つまり偏回帰モデルはその性質の証明にオ

リジナリティがあるのではなく、この関係式をどう使うかにオリジナリティがあると考えられる。

さらに(3)と(4)については、付録2に示すような方法で証明できる。これが「掃き出しの過程を辿ることによりわかる」[5]ということの意味である。つまり付録2の関係を前提とすれば(2)において(3)と(4)の関係はほとんど自明のことと考えられる。

2）偏回帰プロットの拡張について

次式は（量的変数の）偏回帰プロットの基本的な関係式である。

$$y_{\cdot[j]} = b_j x_{j\cdot[j]} + e \quad (5)$$

Park and Wakimoto[2]では、関係式(5)の拡張として、変数グループについての偏回帰モデル(2)を考え、これを使用してCookのDの拡張を行なっている。つまり要約統計量に基づくアプローチを行なっている。しかし偏回帰プロットの拡張という立場からの考察は行なわれていない。実際「For the case of a single parameter, the partial regression plot is a very useful tool for detecting an influential observation. However, since it is difficult to employ a two-dimensional plot for the case of more than a single parameter …」[8]と述べている。

ちなみに関係式(5)は、偏回帰係数、偏相関係数との関係で以前から知られいる。しかしこれを偏回帰プロットと名付けて回帰診断に使用したのは1980年代である。[7]これも関係式をどう使用するかが重要であることを示す例と考えられる。

4．おわりに

偏回帰モデルの議論を通してここで取り上げた論文の特徴について整理した。また付録2として示した証明方法は、掃き出し法の過程に沿って偏回帰モデルの意味を示す形で証明を行なう方法であり、その意味で自然な方法と考えている。

回帰分析については多くのことが知られている。しかし実際に活用するという

観点から、いわゆる回帰診断という形では比較的最近検討が始められている。こうした視点から眺めてみて、従来の結果との関係を整理しておくことは、体系の整備の面からも重要であると考える。

【参考文献】

1) 志村健一・吉澤　正・椚　茂美（1987）：「数量化Ⅰ類における〝偏回帰プロット〟について」日本品質管理学会第17回年次大会研究発表要旨,pp.57-60.

2) Park,S.H. and Wakimoto,K. (1987) :"Detection of an influential observation on a subset of regression parameters in multiple regression", *J. Japan Statist. Soc.*,Vol. 17,No.1,pp.11−19.

3) 志村健一（1991）：「偏回帰プロットの質的変数への拡張とその変数選択への応用」琉球大学経済研究,42号,pp.1−18.

4) 同上 p.5（上から8行目）

5) 同上 p.12（上から11行目）

6) 佐和隆光（1979）：『回帰分析』朝倉書店,pp.81−82.

7) 志村健一（1994）：「偏回帰プロットについて」琉球大学経済研究,48号, pp.383−389.

8) 前出2) p.14（上から3行目）

9) Enslein,K.,Ralston,A. and Wilf,H.S. (1977) :*Statistical Methods for Digital Computer*,John Wiley & Sons.

付録1　(3)、(4)式の証明

ここでは　$X=(x_0,\ldots,x_p)$ はフルランクとして証明すれば十分である。$e=e^*$については、Park and Wakimoto (1987) [2] と同様であったので省略し(3)のみを示す。特徴は(3)とは独立に証明した(4)式（$e=e^*$）を使用することである。

X_Hを、Hに含まれる説明変数を並べた行列であり、Jは{m+1,…,p}なる添え字集合とする。$e=e^*$より

$$y-(X_I : X_J)\begin{pmatrix} b_I \\ b_J \end{pmatrix} = y_{.I}-X_{J.I}c_J$$
$$= y-(X_I : X_{J.I})\begin{pmatrix} c_I \\ c_J \end{pmatrix}$$

これは以下のように書ける。

$$X_{J.I}c_J = X_I(b_I-c_I)+X_Jb_J$$

さらにこの両辺に$X_{J.I}$の転置行列をかける。

$$X_{J.I}'X_{J.I}c_J = X_{J.I}'X_I(b_I-c_I)+X_{J.I}'X_Jb_J$$

$(X_I : X_J)$ がフルランクであることより、$X_{J.I}$もフルランクであること、$X_{J.I}'X_I=0$、$X_{J.I}'X_{J.I}=X_{J.I}'X_J$であることに注意すると次式を得る。

$$c_J = b_J$$

付録2　掃き出し法の過程について

掃き出し法では次のことが知られている。ここでＩは｛0,…,m｝、Ｊは｛m＋１,…,ｐ,ｐ＋１｝なる添え字集合とする。ただしｐ＋１はｙを表わすこととする。またＡ$_{IJ}$を添え字集合ＩとＪに含まれる変数間の積和行列とする。回帰式(1)に対応する積和行列に対し、x_0からx_mについての掃き出しを行なうと、Ａ$_{JJ\cdot I}$としてx_{m+1}からx_pとｙを目的変数と考えたときの残差の積和行列を得る。

$$A_{JJ\cdot I}=A_{JJ}-A_{JI}\left(A_{II}\right)^{-1}A_{IJ} \tag{付1}$$

(上式についてはJennrich,R.I.（1977）[9] など参照。この本の62ページ(8)式を本論文の記法に従い書き改めたものが（付1）である。)

ｐを３として、掃き出しの各ステップを示せば次のようになる。ただし最初の積和行列と、Ａ$_{JJ\cdot I}$に対応する積和行列の要素は＊で示し、それ以外の要素は＋で示した。

①最初の積和行列

$$\left(\begin{array}{cccc:c} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ \hdashline * & * & * & * & * \end{array}\right) \qquad x_0 \quad x_1 \quad x_2 \quad x_3 \qquad y \quad \text{の積和行列}$$

②第 1 列掃き出し後の積和行列

$$\left(\begin{array}{cccc:c} 1 & + & + & + & + \\ 0 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \\ \hdashline 0 & * & * & * & * \end{array}\right) \qquad *\text{部分は} \quad x_{1\cdot0} \quad x_{2\cdot0} \quad x_{3\cdot0} \qquad y_{\cdot0} \quad \text{の積和行列}$$

③第 2 列の掃き出し後の積和行列

$$\left(\begin{array}{cccc:c} 1 & 0 & + & + & + \\ 0 & 1 & + & + & + \\ 0 & 0 & * & * & * \\ 0 & 0 & * & * & * \\ \hdashline 0 & 0 & * & * & * \end{array}\right) \qquad *\text{部分は} \quad x_{2\cdot01} \quad x_{3\cdot01} \qquad y_{\cdot01} \quad \text{の積和行列}$$

④第 3 列の掃き出し後の積和行列

$$\left(\begin{array}{cccc:c} 1 & 0 & 0 & + & + \\ 0 & 1 & 0 & + & + \\ 0 & 0 & 1 & + & + \\ 0 & 0 & 0 & * & * \\ \hdashline 0 & 0 & 0 & * & * \end{array}\right) \qquad *\text{部分は} \quad x_{3\cdot012} \qquad y_{\cdot012} \quad \text{の積和行列}$$

⑤第 4 列の掃き出し後の積和行列

$$\left(\begin{array}{cccc:c} 1 & 0 & 0 & 0 & + \\ 0 & 1 & 0 & 0 & + \\ 0 & 0 & 1 & 0 & + \\ 0 & 0 & 0 & 1 & + \\ \hdashline 0 & 0 & 0 & 0 & + \end{array}\right) \begin{array}{l} \rightarrow b_0 \quad \text{最初の正規方程式の解} \\ \rightarrow b_1 \\ \rightarrow b_2 \\ \rightarrow b_3 \\ \\ \rightarrow S_e \quad \text{最初の回帰式の残差平方和} \end{array}$$

各ステップにおける＊部分は、その右に示したような変数の積和行列になるというのが（付1）の結果である。つまり＊部分は右に示した変数による偏回帰モデルの正規方程式になっていることがわかる。最初の正規方程式の解はステップ⑤で得られる。よってこの掃き出しの途中で得られる、例えば②の$x_{2\cdot 01}$、$x_{3\cdot 01}$と$y_{\cdot 01}$の正規方程式の解もb_2、b_3となることがわかる。

同様に任意の添え字集合Ｉに対する偏回帰モデルの正規方程式は、最初の正規方程式を添え字集合Ｉの要素で掃き出したとき、すなわち最初の正規方程式を解く過程の１ステップとして得られることがわかる。よってその解は最初の正規方程式の対応する解に等しいことがわかる。