Title	重回帰分析における多重共線性とRidge回帰について
Author(s)	吉田, 光雄
Citation	大阪大学人間科学部紀要. 1987, 13, p. 227-242
Version Type	VoR
URL	https://doi.org/10.18910/9749
rights	
Note	

重回帰分析における多重共線性と Ridge 回帰について

吉　田　光　雄

重回帰分析における多重共線性と Ridge 回帰について

I　序

重回帰分析は，主成分分析，因子分析等と並んで多変量解析各種手法のうちでも最も基本的なもののひとつである。多変量解析の成書の多くは，早い章でこの手法を取り上げ多変量解析の導入としているし，そこにさかれるページ数も多い (Harris 1975, Mardia 1979, Press 1972)。

周知のごとく，重回帰分析法は複数個の独立変量から従属変量を予測せんとするものであり，コンピュータとソフトウエアが目覚ましく普及した今日，あらゆる科学の領域で適用され成果をあげている。独立変量の数が大となるにつれ，計算量は加速度的に増大するが，現在では特殊な場合を除いて計算量自身はさほどに問題となることはない。

問題はむしろ逆に，容易に計算結果が得られるだけに，安易に適用され，事前のデータに対する吟味が十分になされないままに，しばしば，“とりあえず計算してみる”式のデータ処理が行われている，ということではなかろうか。計算をすれば何がしかの結果は得られるわけであるから，方法論的吟味が棚上げされたまま，ただ計算機に対するインプットとアウトプットのみに関心が払われるとすれば，それは問題と言わざるをえない。計算に先立って，変数や観測値のチェックは言うに及ばず，そもそも実験や調査のデザイン，数量化の是非，処理法の評価など，本来の計算以外にもこうした方法論に関する考察を欠かすことはできないのは勿論のことである。

しかしながら，多くの現象は多数の要因が複雑に絡みあっており，過度の吟味や人為的な条件統制がかえって現象を歪めることもまた否めない。多変量解析は，むしろそうした人為的条件統制を排除し，あるがままに得られたデータから，要因相互の関連や構造を吟味し，現象を単純化する方法として発展してきたとも言える。

分散分析のごとく，要因毎の変動をデータ処理によって分離するとか，主成分分析のごとく，座標軸を回転することによって，データに潜在する独立な軸（要因）を抽出する，等の作業が可能であるが，かつての心理学実験に見られたごとき，細かい，厳密な条件統制というよりも，それらを緩くしてむしろ多変量解析の手法を援用し，事後のデータ処理から要因を分離することも不可能ではない。

ともあれ，両者のバランスこそがまさに重要であるが，本稿では重回帰分析を例にとり，手法に限定してのことであるが，同法を適用する際に陥り易い問題のひとつとして，多重共

線性を取り上げ，その内容と対策について検討してみることとする。

以下本稿を進めるに際し，次のごとく記号を定めることとする。

n 個の観測値よりなる p 変量の独立変量を $\boldsymbol{X}(n\times p)$，従属変量を $\boldsymbol{y}(n\times 1)$ とし，回帰係数ベクトルを $\boldsymbol{\beta}$ とする時，重回帰式は

$$\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{u} \tag{1}$$

であるが，ここでは一般性を失なうことなく独立変量，従属変量ともに，変量毎に標準化されているものとする。また残差ベクトル $\boldsymbol{u}$ に関して

$$\boldsymbol{u}\sim \mathrm{N}(0,\ \sigma^2\boldsymbol{I}) \tag{2}$$

をも仮定しておく。

行列 $\boldsymbol{X'X}$ が相関行列となるよう標準化されているので，独立変量相互間の相関行列を $\boldsymbol{R}$，従属変量との相関ベクトルを $\boldsymbol{r}$ とすれば

$$\begin{aligned}\boldsymbol{R}&=\boldsymbol{X'X}\\ \boldsymbol{r}&=\boldsymbol{X'y}\end{aligned} \tag{3}$$

また，重相関係数 r_m は

$$r_m=\sqrt{\boldsymbol{r'R^{-1}r}} \tag{4}$$

として求められる。

いま，行列 $\boldsymbol{X'X}$ を固有値分解したとき，固有値を $\boldsymbol{l}_j$ $(l_1\geq l_2\geq\cdots\cdots l_p\geq 0)$，対応する固有ベクトルを $\boldsymbol{g}_j(p\times 1)$，さらにこれらを行列表示したものを $\boldsymbol{L}$，$\boldsymbol{G}$ とすると

$$\begin{aligned}&\boldsymbol{L}=\mathrm{diag}(l_j) \qquad\qquad j=1,\ \cdots\cdots,\ p\\ &\boldsymbol{G}=(\boldsymbol{g}_1,\ \boldsymbol{g}_2,\cdots\cdots,\ \boldsymbol{g}_p)\end{aligned} \tag{5}$$

であり，このとき

$$\boldsymbol{X'X}=\boldsymbol{GLG'} \tag{6}$$

$$\boldsymbol{L}=\boldsymbol{G'}(\boldsymbol{X'X})\boldsymbol{G} \tag{7}$$

である。

(1) 式における $\boldsymbol{\beta}$ の最小 2 乗解（最尤解とも一致）を $\hat{\boldsymbol{\beta}}$ とすれば，これは $\boldsymbol{X'X}$ が正則のとき，通常の最小 2 乗法で解くことができ（OLS：ordinary least square）

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{X'X})^{-1}\boldsymbol{X'y}=\boldsymbol{R}^{-1}\boldsymbol{r} \tag{8}$$

であり，その期待値，分散は

$$\mathrm{E}(\hat{\boldsymbol{\beta}})=\boldsymbol{\beta} \tag{9}$$

$$\mathrm{V}(\hat{\boldsymbol{\beta}})=\sigma^2\ (\boldsymbol{X'X})^{-1} \tag{10}$$

である。これはまた先の固有値分解を用いて

$$\begin{aligned}\hat{\boldsymbol{\beta}}&=\ (\boldsymbol{GLG'})^{-1}\boldsymbol{X'y}\\ &=\ (\sum_j l_j\,\boldsymbol{g}_j\,\boldsymbol{g'}_j)^{-1}\boldsymbol{X'y}\end{aligned}$$

$$= \left(\sum_j \boldsymbol{g}_j \boldsymbol{g}'_j / l_j\right) \boldsymbol{X}'\boldsymbol{y} \tag{11}$$

$$\mathrm{V}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2 \sum \boldsymbol{g}_j \boldsymbol{g}'_j / l_j \tag{12}$$

とも書けるし，$\hat{\boldsymbol{\beta}}$ と $\boldsymbol{\beta}$ の距離を L_1 とすれば，$\hat{\boldsymbol{\beta}}$ の平均2乗誤差（MSE: Mean square error）も同様に

$$\begin{aligned} \mathrm{E}(L_1) &= \mathrm{MSE}\ (\hat{\boldsymbol{\beta}}) \\ &= \mathrm{E}\{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\} \\ &= \hat{\sigma}^2 \mathrm{tr}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\ &= \hat{\sigma}^2 \sum_j 1/l_j \end{aligned} \tag{13}$$

となる。

さらに，$\boldsymbol{G}$ を用いて $\boldsymbol{X}$ を変換したものを $\boldsymbol{W}$，その際の回帰係数を $\boldsymbol{\alpha}$ とすると

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X\beta}+\boldsymbol{u} \\ &= \boldsymbol{W\alpha}+\boldsymbol{u} \end{aligned} \tag{14}$$

である。ただし，

$$\begin{aligned} \boldsymbol{W} &= \boldsymbol{XG} \\ \boldsymbol{\alpha} &= \boldsymbol{G}'\boldsymbol{\beta} \end{aligned} \tag{15}$$

である。

II　多重共線性

$\boldsymbol{X}'\boldsymbol{X}$ が正則でないとき，(8) 式の $\hat{\boldsymbol{\beta}}$ の解は不安定となり求める事ができない。現実のデータの場合に精確に $|\boldsymbol{X}'\boldsymbol{X}|=0$ となることは，むしろ稀であろうが，0に近い，小さい値となることは十分起こりうる。たとえば，p 個の説明変量のうち極めて相互に相関の高い変量が含まれていたり，あるいはある変量が他の変量群と因果関係にあり，そのために近似的に一次式で表現できたりするような場合である。すなわち，すべてが0でない定数 a_j によって説明変量 x_j の間に

$$a_1 x_1 + \cdots\cdots + a_p x_p = 0 \tag{16}$$

なる関係が生起している時，一次従属となるため，$\boldsymbol{X}'\boldsymbol{X}$ のランク落ちが起こり，$\hat{\boldsymbol{\beta}}$ が不定ないしは不安定となる。

こうした状況は多重共線性 (multicollinearity) と呼ばれ (Cramer 1985)，データ処理時には注意すべき点のひとつである。Goldberger (1968) は，多重共線性を「説明変量のうちのいくつかが，相互に関連しており，そのために単独の影響を分離したり，効果を評価したりすることが，不可能ではないにしても，困難な状態」，と定義している。

心理学における数量化は，それ自身必ずしも常に確固たる基盤の上に構成されているとは

限らず，測定の不安定性が見られる分だけ信頼性を高める方法として反復測定が試みられる。たとえば，ある属性の測定に際して多数の下位項目の測定が行われ，それらの加算（必要に応じて重みつき加算）が行われるが，このとき下位項目に要請されることは一次元性であり，相互に相関の高い同一次元上の属性の測定がなされなければならない。そして，反復測定により，信頼性を向上させる仕組みである。

心理学における調査ないしは実験において独立変量のなかに，信頼性向上のために同一（類似）の現象が取り扱われるのは，発想としては極めて自然である。多変量解析を変量間相互の絡みを残して同時に処理する方法（Morrison 1976）として理解するとき，これはむしろ日常的である。しかし，重回帰分析の場合の説明変量としてはむしろ避けなければならない性質のものである。

多重共線性を発見する方法を，Mansfield (1982) は次の5つに整理している。

（1）　説明変量の対のうち極端に高い値をもつもの。

$$|r|_{\max}=\max_{i,j}\ (X'X_{ij}) \tag{17}$$

（2）　相関行列の行列式が非常に小さい。

$$|X'X| \fallingdotseq 0 \tag{18}$$

（3）　固有値に非常に小さいものがある。もし精確に $l_j=0$ となれば，完全な一次従属性がある。

（4）　大きい分散拡大要因（後述）VIF(j) の値。

$$\mathrm{VIF(j)} = (X'X)_{jj}^{-1} \tag{19}$$

（5）　大きい R_j^2 の値（1に近い）。

$$R_j^2 = 1-1/\mathrm{VIF(j)} \tag{20}$$

多重共線性を具体例で示そう。偏回帰係数 $\hat{\beta}$ の分散は（10）式の通りであるが，今 $p=2$ の場合について書けば

$$\mathrm{V}(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

$$= \sigma^2/n(1-r_{12}^2) \tag{21}$$

となり，x_1，x_2 の相関が高いとき，すなわち $r_{12}\to 1$ のとき，値は不安定となる。また，$X'X$ の固有値の中に0に近いものがあるとき，(12) 式では分母に l_j を含むので，$\hat{\beta}$ の分散は不安定であることがわかる。

こうした多重共線性を発見するために，計算プログラムの中にそのためのルーチンを組込むことが必要である。上述の（1）～（5）をすべて試みることは必ずしも必要ではないが，（1）は相関行列の要素をチェックすればよいし，（2）は $\hat{\beta}$ のみを求めるのではなく，

計算の過程で（$\boldsymbol{X'X}$）なり，$|\boldsymbol{X'X}|$ の値を見ればよい。多くのソフトないしはプログラム・ライブラリーでは $|\boldsymbol{X'X}|<\delta$ で0の判定がなされているが，δ のとり方如何では，計算が続行され $\hat{\boldsymbol{\beta}}$ がもとめられてしまう。(4)，(5) は（$\boldsymbol{X'X}$）の逆行列の要素から容易に検討することができる。

どの変量間に多重共線性が見られるかを発見するためには，更に手のこんだ計算を必要とする。2変量間の相関は $\boldsymbol{R}$ の要素をみればよいが，3変量間ないしはそれ以上にわたる線型性の判定は厄介である。

第1の方法は，p 個の説明変量それぞれを目的変量として，それをそれ以外の説明変量から予測すべく重回帰分析を反復することである。非常に高い重相関係数が得られたとき，変量間の1次従属性が示唆される。今，変量 x_j を目的変量としたときの重相関係数を R_j とすれば，

$$\mathrm{VIF}=(1-R_j^2)^{-1} \tag{22}$$

で求められる値がひとつの目安であり，通常 VIF>10 のとき，多重共線性が見られるとされる (Chatterjee 1977)。このＶＩＦは分散拡大要因（Variance inflation factor）と呼ばれ，この値は相関行列の逆行列の第 j 成分，すなわち $r_{jj}^{(-1)}$ と同じである。また $\hat{\boldsymbol{\beta}}$ の分散を $\hat{\sigma}^2$ で割ったものとも等しいので，$\hat{\boldsymbol{\beta}}$ の値のみならずその分散も同時に求めておけば多重共線性の有無を検討することができる。Gramer (1985) は標準偏回帰係数（標準化された $\boldsymbol{X}$ に用いる）は経験的に $-1\sim+1$ であるのでそれより大きい値，特に $|2|$ を越えるものは疑わしいと述べている。

第2の方法は，p 個の説明変量に主成分分析（ＰＣＡ）を施すことである。相関行列を用いてＰＣＡを行い，固有値の中に0に近いものがあるかどうかを見る。もし第 k 固有値が0に近いとすれば，その意味する所は，第 k 主成分の分散が0，すなわち定数であり，線型関係を示唆するものである。

何れの方法とも，本来の回帰分析以外に余分の計算を強いるものであるが，重回帰分析を正しく活用するためには不可避であろう。そして多重共線性が発見された時，それをとり除く作業が次いで必要となるが，その最も簡単な方法は相関する変数群の中から1つを残し，他を冗長なものとして除くことである。そうすることにより，$\boldsymbol{X'X}$ のランクが落ち正則な行列を得ることができる。しかし，単に計算結果の数字上の理由だけではなく，変量の内容からの検討が必要なことは論を俟たない。

しかしながら，単に予測ということのみに興味があるときはそれでもよいが，予測と同時にそれに貢献する変量の検討にも関心があるとき，ただ数量的に従属であるというだけの理由で，落とすことはできない変量もあるであろう。そうしたとき，次節で述べる Ridge 回

帰を適用すればこうした難点を克服することができる。

Ⅲ Ridge 回帰

正則でない行列 $\boldsymbol{X'X}$ に対するひとつの解決法として，リッジ回帰（ridge regression）なるものが紹介されている（Hoerl 1962，Hoerl & Kennard 1970，1975，1976，Williams 1968）。これは $\boldsymbol{X'X}$ の対角要素に一定の定数 k を加えて

$$\hat{\boldsymbol{\beta}}^* = (\boldsymbol{X'X}+k\boldsymbol{I})^{-1}\boldsymbol{X'y} \tag{23}$$

として $\boldsymbol{X'X}$ を正則化し回帰問題を解く方法である。ただし，標準化されていない $\boldsymbol{X}$ の場合には $k\boldsymbol{I}$ の代わりに

$$\boldsymbol{K}=\mathrm{diag}\ (k_j) \qquad j=1,\ \cdots\cdots,\ p \tag{24}$$

を用いることとなる。

こうして求められた $\hat{\boldsymbol{\beta}}^*$ は ridge 推定量と呼ばれ，後述の（29）式のごとく不偏推定量ではないが，変量を除去することなく回帰問題を解き，予測に貢献する変量等の検討に用いることができる。

（23）式は次のごとく求められる。今，$\boldsymbol{\beta}$ の任意の推定量を $\boldsymbol{b}$ とすると，残差平方和は

$$\begin{aligned} R(\boldsymbol{b}) &= (\boldsymbol{y}-\boldsymbol{Xb})'(\boldsymbol{y}-\boldsymbol{Xb}) \\ &= (\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}) + (\boldsymbol{b}-\hat{\boldsymbol{\beta}})'\,\boldsymbol{X'X}\,(\boldsymbol{b}-\hat{\boldsymbol{\beta}}) \\ &= R\mathrm{min}+\phi(\boldsymbol{b}) \end{aligned} \tag{25}$$

となり，右辺の第1項は最小2乗推定量 $\hat{\boldsymbol{\beta}}$ を用いたときの残差平方和，第2項はそれ以外の推定量を用いた場合の残差平方和の増分である。今，$\boldsymbol{b'b}=1$ のある一定の条件下でこの $\phi(\boldsymbol{b})$ を最小にすることを考えると，Lagrange の乗数を $1/k$ とすれば，

$$\phi(\boldsymbol{b}) = (\boldsymbol{b}-\hat{\boldsymbol{\beta}})'\,\boldsymbol{X'X}\,(\boldsymbol{b}-\hat{\boldsymbol{\beta}}) + \frac{1}{k}(\boldsymbol{b'b}-1) = \mathrm{min} \tag{26}$$

として，これを解けばよい。このとき，先の（23）式がえられる。さらに

$$\begin{aligned} \boldsymbol{Z} &= (\boldsymbol{I}+k(\boldsymbol{X'X})^{-1})^{-1} \\ &= (\boldsymbol{X'X}+k\boldsymbol{I})^{-1}\boldsymbol{X'X} \end{aligned} \tag{27}$$

とおけば，式の変形により

$$\begin{aligned} \hat{\boldsymbol{\beta}}^* &= (\boldsymbol{X'X}+k\boldsymbol{I})^{-1}\boldsymbol{X'y} \\ &= \boldsymbol{Z}(\boldsymbol{X'X})^{-1}\boldsymbol{X'y} \\ &= \boldsymbol{Z}\hat{\boldsymbol{\beta}} \end{aligned} \tag{28}$$

となるので，$\hat{\boldsymbol{\beta}}^*$ は $\hat{\boldsymbol{\beta}}$ の線型変換として得られることもわかる。その平均，分散は

$$\mathrm{E}(\hat{\boldsymbol{\beta}}^*) = \boldsymbol{Z\beta} \tag{29}$$

$$
\begin{aligned}
V(\hat{\boldsymbol{\beta}}^*) &= \boldsymbol{Z}V(\hat{\boldsymbol{\beta}})\boldsymbol{Z}' \\
&= \hat{\sigma}^2\boldsymbol{Z}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{Z} \\
&= \hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}
\end{aligned} \tag{30}
$$

である。また，$\hat{\boldsymbol{\beta}}^*$ と $\boldsymbol{\beta}$ の距離を L_2 とすれば

$$L_2 = \hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}$$

で $\hat{\boldsymbol{\beta}}^*$ の平均２乗誤差は

$$
\begin{aligned}
E(L_2) &= \mathrm{MSE}(\hat{\boldsymbol{\beta}}^*) \\
&= E\{(\hat{\boldsymbol{\beta}}^*-\boldsymbol{\beta})'(\hat{\boldsymbol{\beta}}^*-\boldsymbol{\beta})\} \\
&= E\{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{Z}'\boldsymbol{Z}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\} + (\boldsymbol{Z}\boldsymbol{\beta}-\boldsymbol{\beta})'(\boldsymbol{Z}\boldsymbol{\beta}-\boldsymbol{\beta}) \\
&= \hat{\sigma}^2 \mathrm{tr}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{Z} + \boldsymbol{\beta}'(\boldsymbol{Z}-\boldsymbol{I})'(\boldsymbol{Z}-\boldsymbol{I})\boldsymbol{\beta} \\
&= \hat{\sigma}^2 \sum_j l_j/(l_j+k)^2 + k^2\boldsymbol{\beta}'(\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-2}\boldsymbol{\beta} \\
&= \gamma_1(k) + \gamma_2(k) = \gamma
\end{aligned} \tag{31}
$$

である。第１項は $\hat{\boldsymbol{\beta}}^*$ の分散の和であり，第２項は $\hat{\boldsymbol{\beta}}$ よりも $\hat{\boldsymbol{\beta}}^*$ を用いた際の偏りの平方和と考えられるので，これらを比較することにより最良の k を選ぶことができる。$\gamma_1(k)$ は k の減少関数，$\gamma_2(k)$ は k の増加関数であるので，平均２乗誤差を最小にする k を $0<k<1$ の範囲で探せばよい。さらに，第２項は $\boldsymbol{X}'\boldsymbol{X}$ を正準形式に変換すると，計算はより容易になる。今，$\boldsymbol{X}'\boldsymbol{X}$ の固有値を (5) 式のごとく定めれば，$\boldsymbol{X}'\boldsymbol{X}$ は (6) 式のごとく分解され，(15) 式を用いて

$$
\begin{aligned}
\hat{\boldsymbol{\alpha}}^* &= (\boldsymbol{W}'\boldsymbol{W}+k\boldsymbol{I})^{-1}\boldsymbol{W}'\boldsymbol{y} \\
&= (\boldsymbol{L}+k\boldsymbol{I})^{-1}\boldsymbol{L}\boldsymbol{G}'\hat{\boldsymbol{\beta}} \\
&= (\boldsymbol{L}+k\boldsymbol{I})^{-1}\boldsymbol{L}\hat{\boldsymbol{\alpha}}
\end{aligned} \tag{32}
$$

で，$\hat{\boldsymbol{\alpha}}$ ないしは $\hat{\boldsymbol{\beta}}$ の線型変換で求めることができる。このとき，

$$\gamma_2(k) = k^2 \sum_j \alpha_j^2/(l_j+k)^2 \tag{33}$$

である。

さらに，一般的に $\boldsymbol{X}'\boldsymbol{X}$ が相関行列ではなく，すなわち diag $(\boldsymbol{X}'\boldsymbol{X}) = \boldsymbol{I}$ となっていないときには (24) 式のごとく $\boldsymbol{K}$ を定めることになる。このとき

$$\hat{\boldsymbol{\alpha}}^*(\boldsymbol{W}'\boldsymbol{W}+\boldsymbol{K})^{-1}\boldsymbol{W}'\boldsymbol{y} \tag{34}$$

で，$\hat{\boldsymbol{\alpha}}^*$ の第 j 要素は

$$\hat{\alpha}^*_j = (l_j/(l_j+k_j))\,\alpha_j \tag{35}$$

で，また

$$\gamma_1(k)=\hat{\sigma}^2\sum_j l_j/\ (l_j+k_j)^2 \tag{36}$$

$$\gamma_2(k)=\sum_j (\alpha_j k_j/\ (l_j+k_j)^2) \tag{37}$$

である。$\hat{\boldsymbol{\beta}}^*$ ないしは $\hat{\boldsymbol{\alpha}}^*$ の分散は当然 $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\alpha}}$ のそれよりは小さい筈であり，計算の過程でチェックしておくことが望ましい。

以上の外にも，k の推定方法は色々と報告されており，一覧すれば，たとえば次の通りである（ただし，⑥，⑦については Yoshioka (1986) による)。

①	OLS	
②	$\gamma=\min$	(Hoerl, Kennard 1970)
③	$k=p\hat{\sigma}^2/\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}}$	(Hoerl, Kennard 1975)
④	$k=p\hat{\sigma}^2/\sum\hat{\alpha}_j^2 l_j$	(Lawless, Wang 1976)
⑤	$k(i)=p\hat{\sigma}^2/\hat{\boldsymbol{\beta}}'(i)\hat{\boldsymbol{\beta}}(i)$	(Hoerl, Kennard 1976)
⑥	$\sum\hat{\alpha}_j^2\ (\frac{1}{k}+\frac{1}{l_j})^{-1}=p\hat{\sigma}^2$	(Dempster 1977)
⑦	$\sum\hat{\alpha}_j^2\ (\frac{1}{k}+\frac{1}{l_j})^{-1}=p\hat{\sigma}^2\ (\frac{n-p}{n-p-2})$	(Scolve 1973)
⑧	$k_j(i)=\hat{\sigma}^2/\hat{\alpha}^{*2}_j(i)$	(Hemmerle 1975)

③，④は代数的に解くことができるので容易であるが，⑤，⑧は反復法により解くことになる。(i) を第 i 番目の反復とすれば，$|k(i)-k(i+1)|<\varepsilon$ で k の収束を見ればよい。Hoerl は③に加えてこれを反復法で解く方法⑤を紹介している。⑧は一定の値 k を求めるのではなく，$\hat{\alpha}$ を用いて k を求めるものである。⑥，⑦は k を動かし，最も右辺の値に近づくところを探せばよい。0～1の範囲からスタートし，次々とその範囲を狭め，桁を落とし，必要な精度の桁数を求めることができる。

いずれの方法が優れているかについての結論は得られていないが，次節では，数値例について上式を適用し，k の値を求めてみることにする。

Ridge 回帰を巡る諸問題は，Hoerl を始めとして多くの統計学者によって取り上げられている (Banerjee & Carr 1971, Marquadt & Snee 1975, McDonald & Galarneau 1975, Lawless 1978, Wichern & Churchill 1978, Gibsons 1981, 他)。後藤 (1976) は研究の系譜を，(1)一般化行列回帰，主成分回帰，(2) Ridge 推定量の性質，(3)最適推定量の選択の3つに分類して整理しているが，今後ますます研究されていくテーマのように思われる。

Ⅳ 数 値 例

これまでのところを数値例で検討しよう。Morrison (1983) は多重共線性の例としてドルの対外貨レートから金の価格を予測する例をあげている。それによると，目的変量 y は1オンスあたりの金の価格，これを予測するための外貨レートは，x_1（米ドル／ドイツマルク），x_2（米ドル／スイスフラン）であり，1973～1979年の27四半期のデータが用いられた。相関行列，y との相関ベクトルは

$$\boldsymbol{X}'\boldsymbol{X}=\begin{pmatrix}1.0000 & 0.9538\\ .9538 & 1.0000\end{pmatrix}$$

$$\boldsymbol{r}=(.8740,\quad .8418)'$$

であり，これらよりＯＬＳによる解として $\hat{\boldsymbol{\beta}}=(0.7876,\ 0.0906)'$，$r_m=0.6744$ をうる。単独に y を予測せんとしたときの予測の精度は $r_{m_1}=0.7638$，$r_{m_2}=0.7086$ であり，いずれの変数も高い精度で予測が可能であることが示されている。しかし，それにも拘らず，同時に x_1 と x_2 を用いた場合には，r_m は下がり，さらに $\hat{\beta}_2$ は $\hat{\beta}_1$ に比し極端に小さい値であり，不合理な解となっている。しかも $\hat{\beta}_1$ と $\hat{\beta}_2$ の相関が $r=-0.9538$ となっており，高い負の相関である。これはx_1 と x_2 の高い相関に起因しており，多重共線性の一つの例である。ちなみに，$|\boldsymbol{X}'\boldsymbol{X}|=0.0903$であり，ＶＩＦは共に11.08であった。

さらに，他の例を示そう。表1は Chatterjee & Price (1977) に引用されている例であ

表1　Economic Data of France

Year	y	x_1	x_2	x_3
1949	15.9	149.3	4.2	108.1
50	16.4	161.2	4.1	114.8
51	19.0	171.5	3.1	123.2
52	19.1	175.5	3.1	126.9
53	18.8	180.8	1.1	132.1
54	20.4	190.7	2.2	137.7
55	22.7	202.1	2.1	146.0
56	26.5	212.4	5.6	154.1
57	28.1	226.1	5.0	162.3
58	27.6	231.9	5.1	164.3
59	26.3	239.0	0.7	167.6
Mean	21.891	194.591	3.30	139.736
SD	4.332	28.603	1.572	19.674

るが，フランスの経済データであり，x_1（国民総生産），x_2（株式出来高），x_3（国民消費量）の年次推移から，y（輸入額）を予測せんとするものである。データを標準化し，$\boldsymbol{X'X}, \boldsymbol{X'y}$ を求めると

$$\boldsymbol{X'X}=\boldsymbol{R}=\begin{pmatrix} 1.0000 & 0.0258 & 0.9973 \\ 0.0258 & 1.0000 & 0.0357 \\ 0.9973 & 0.0357 & 1.0000 \end{pmatrix}$$

$$\boldsymbol{X'y}=\boldsymbol{r}=(0.9653 \quad 0.2507 \quad 0.9719)'$$

を得る。x_1 と x_3 の相関が高く多重共線性を暗示している。ちなみに，$|\boldsymbol{X'X}|=0.00529$ であった。これは事実上 0 とみなしても差し支えないような値である。逆行列は

$$(\boldsymbol{X'X})^{-1}=\begin{pmatrix} 188.806 & 1.8533 & -188.363 \\ 1.8533 & 1.0195 & -1.8847 \\ -188.363 & 1.8847 & 188.921 \end{pmatrix}$$

であるので，分散拡大要因は

$$\mathrm{VIF}=(188.81, \quad 1.02, \quad 188.92)'$$

で，VIF>10 の条件ははるかに越えている。重回帰式を機械的に解いて得られる通常の回帰係数（ＯＬＳ）は

$$\hat{\boldsymbol{\beta}}=(-0.3503, \quad 0.2128, \quad 1.3137)'$$

$$\hat{\boldsymbol{b}}=(-10.137(b_0), \quad -0.0531, \quad 0.5864, \quad 0.2893)'$$

で，このとき重相関数は $r_m=0.9924$，$r_m^2=0.9848$ をうる。一応は高い精度の予測であるが，解は不安定である。第１に，x_1 の回帰係数が y との相関 $r_{01}=0.9653$ であるにもかかわらず，マイナスになるのは解釈を苦しくする。第２に，x_3 の回帰係数が１をこえ，Cramer (1985) によれば，これはむしろ高い値である。

固有値，固有ベクトルは

$$\boldsymbol{l}=(1.9992, \quad 0.99816, \quad 0.002651)'$$

$$\boldsymbol{G}=\begin{pmatrix} 0.7063 & -0.0357 & -0.7070 \\ 0.0435 & 0.9990 & -0.0070 \\ 0.7066 & -0.0258 & 0.7072 \end{pmatrix}$$

であり，これらを用いると正準形式の回帰係数は

$$\hat{\boldsymbol{\alpha}}^*=(0.6901, \quad 0.1912, \quad 1.1752)'$$

となる。$\hat{\boldsymbol{\beta}}$ より $\hat{\boldsymbol{y}}$ を求め誤差分散の推定値を算出すると，

$$\hat{\sigma}^2=0.013049$$

を得た。

さて次に ridge 推定量を求める作業であるが，②では γ を最小にすればよいから，$\gamma=\gamma_1+\gamma_2=\min$ となるような k をさがせばよい。凡その見当として 0 から 1 までの k について，まず 0.1 の幅で（これを 0 (0.1) 1 と略記する (McDonald & Galarneau 1975)）γ_1, γ_2 を求めたところ，0〜0.1 の間にあることがわかった。以後，0(0.01)0.1，0(0.001)0.01，0.009 (0.0001) 0.010 と計算を反復し k の幅を狭めていけば，最終的に $k=0.0094$ あたりで，$\gamma=1.097$ の最小値が得られた。γ_1，γ_2 は何れも k の単調関数であるので，これ以外に極値は存在しない。

③，④による推定量は代数的に解けばよいから容易に求められる。それぞれ，$k=0.02065$, 0.03941 である。⑤では 9 回の反復で $k=0.08305$ を得た。収束の判定には k のみを用いたが，$\hat{\beta}$ も同様の速度で収束しているようであった。⑥，⑦は②と同様の方法で求めればよい。解はそれぞれ，$k=0.07205$, $k=0.09978$ である。問題は⑧であるが，$\hat{\alpha}$ の値が反復の 4〜5 回目あたりから0.01以下になり，従って $k>1$ となりますます発散して，本データに関しては求めることができなかった。

以上，各々の場合について k を推定し ridge 回帰を行った結果を一覧にすれば，表 2 の通りである。同表では k の値の小さい順にならべたが，興味ある事実として，ＯＬＳから遠ざかるにつれ，重相関係数は減少，残差平方和（誤差分散推定値）は増大し，その意味ではＯＬＳが最小 2 乗推定であることを裏づけている。また，ＯＬＳでは $\hat{\beta}_1<0$ であったが ridge 推定量ではそれが見事に克服され，いずれも正で，かつ 0〜1 の範囲であった。そして予測の貢献度としては $\hat{\beta}_1^*$，$\hat{\beta}_3^*$ が良く，$\hat{\beta}_2^*$ が悪いことがわかるが，これは独立変量と従属変量の相関からも頷けるところである。

表 2　Results of Ridge Regressions

	① OLS	② $\gamma=\min$	③ Hoerl	④ Lawless	⑥ Dempster	⑤ Hoerl	⑦ Scolve
k	0	.00945	.02065	.03941	.07205	.08305	.09978
r_m	.9923	.9841	.9780	.9684	.9525	.9473	.9395
r_m^2	.9847	.9684	.9564	.9379	.9073	.8974	.8827
$\hat{c}^2$	.01305	.01632	.01744	.01841	.02013	.02082	.02201
$\hat{\beta}^*$	−.3503	.2963	.3812	.4190	.4345	.4359	.4365
	.2128	.2174	.2160	.2127	.2069	.2049	.2021
	1.3137	.6624	.5722	.5257	.4954	.4892	.4814
γ	4.9420	1.1465	1.1754	1.2441	1.2985	1.3091	1.3217
γ_1	4.9420	.3197	.1105	.0535	.0365	.0349	.0338
γ_2	0	.8269	1.0649	1.1905	1.2621	1.2742	1.2879

何れの方法の k を選択するかについては議論が分かれるが，これは数値例からだけでは決められない。第②法ですでに，正則の問題は解決されており，k を 0 から僅かに動かすだけで妥当な $\hat{\beta}^*$ が得られている。②〜⑦で $\hat{\beta}^*$ の値は動いているので，独立変量の貢献度を何れで見るかは難しい。

そもそも正則でないデータに遭遇した場合，考えられる解決法としては，

（1） 観測値の吟味——新しいサンプルを追加するとか，異常値（outlier）を除外する等，またデータを下位グループに分け，異質のサンプルの分離を試みることが必要となる場合もあろう

（2） 説明変量の選択——一次従属の関係にある変数（群）を削除したり，また必要に応じて新しい変数を追加する

等が考えられるが，本稿で見てきたごとく ridge 回帰を試みることにより，説明変量を削除することなく回帰の問題を解くことができる。特に，単なる予測だけが問題ではなく，説明変量の吟味をも目的とするとき，この手法は不可避であろう。

そしてそのときには，さらに AIC，Mallow の Cp 基準等を用いた説明変量の検討も必要であろうが，それらについては本稿の目的ではないので，ここでは割愛する。

V 要　　約

重回帰分析に用いられる説明変量の中に相互に関連するものが組込まれているとき，相関行列が正則でなくなり，解が求められなかったり，求められても非常に不安定であったりする場合がある。

これは統計学で多重共線性と呼ばれる現象であるが，このようなとき説明変量を削除したり，追加したりすることなくこれを解決する一つの方法として，ridge 回帰なる方法が提案されている。すなわち，相関行列の対角要素に定数（極めて小さい）を加えて解く方法であるが，本稿では，これら多重共線性と種々の ridge 推定量を求める方法をあげ，数値例を用いて，それらを検討した。その結果，通常の最小2乗法を僅かに修正するだけで，安定した解が得られることがわかった。

こうした ridge 回帰の方法は，今後ますます科学の領域で適用され，効果を発揮するものと思われる。

References

Banerjee, K.S. & Carr, R.N. 1971 A commet on ridge regression, Biased estimation for non-orthogonal problems, *Technometrics*, ***13***, 895-98.

Chatterjee, S. & Price, B. 1977 *Regression analysis by examples.* John-Wiley.

Cramer, E. M. 1985 Multicollinearity, in Kotz, S. & Johnson, N. L. ed, *Encyclopedia of statistical sciences*, ***5***, 639-643, John Wiley.

Dempster, A. P., Schatzoff, M. & Wermut, N. 1977 A simulation study of alternatives to ordinary least squares, *Journal of the American Statistical Association*, ***72***, 77-91.

Gibsons, D.G 1981 A Simulation study of some ridge estimators, *Journal of the American Statistical Association*, ***76***, 131-139.

Goldberger, A.S. 1968 *Topics in regression analysis*, Macmillan.

後藤昌司, 1976 回帰係数の推定と調整, 日本統計学会誌, 6, 39-61.

Harris, R.J. 1975 *A primer of multivariate statistics*, Academic Press.

Hemmerle, W. J. 1975 An explicit solution for generalized ridge regression, *Technometrics*, ***17***, 309-314.

Hoerl, A.E. 1962 Application of ridge analysis to regression problems, *Chemical Engineeering Progress*, ***58***, 54-59.

Hoerl, A.E. & Kennard, R.B. 1970 Ridge regression: biased estimation fot nonorthgonal problems, *Technometrics*, ***12***, 55-67.

Hoerl, A.E. & Kennard, R.B. 1970 Ridge regression: applications to nonorthogonal problems, *Technometrics*, ***12***, 69-82.

Hoerl, A. E., Kennard, R.W. & Baldwin, K.F. 1975 Ridge regression: some simulations, *Communications in statistics*, ***4***, 105-123.

Hoerl, A.E. & Kennard, R.W. 1976 Ridge regression: iterative estimation of the biased parameter, *Communications in Statistics*, ***5***, 77-88.

Lawless, J.F. & Wang, P. 1976 A simulation study of ridge and other regression estimators, *Communications in statics*, ***5***, 307-323.

Lawless, J.F. 1978 Ridge and related estimation procedure: theory and practice, *Communications in Statistics*, ***7***, 139-164.

Mansfield, E.R. & Helms, B.P. 1982 Detecting multicollinearity, *The American Statistician*, ***36***, 158-160.

Mardia, K.V., Kent, J.T. & Bibby, J.M. 1979 *Multivariate analysis.* Academic Press.

Marquardt, D.W. & Snee, R.D. 1975 Ridge regression in practice, *The American Statistician*, ***29***, 3-20.

McDonald, G. C. & Galarneau, D.I. 1975 A monte calro evaluation of some ridge-type estimators, *Jounal of the American Statistical Association*, ***70***, 407-416.

Morrison, D.F. 1976 *Multivariate statistical methods*, 2nd ed. McGraw-Hill.

Morrison, D.F. 1983 *Applied linear statistical methods*, Prentice-Hall.

Press, S.J. 1972 *Applied multivariate analysis*, Holt.

Scolve, S.L. 1973 Least squares with random regression coefficient, *Technical Reports*, ***87***, Economic Series, Stanford Univ.

Wichern, D.W. & Churchill, G.A. 1978 A comparison of ridge estimators, *Technometrics*, ***20***, 301-311.

Williams, E.J. 1968 Ridge regression, in Kruskal, W.H. & Tanner, J.M. ed, *International encyclopedia of statistics*, 539-541, The Free Press.

Yoshioka, S. 1986 Multicollinearity and avoidance in regression analysis, *Behaviormetrics*, ***19***, 103-120.

MULTICOLLINEARITY AND RIDGE REGRESSION IN REGRESSION ANALYSIS

Mitsuo YOSHIDA

Examination of a set of data for the existence of "multicollinearity" should always be performed as an initial step in any multiple regression analysis for every researcher who tries to employ regression analysis. The term multicollinearity is used in statistics in situation where one variable is nearly a linear combination of other variables. It might be an inevitable pitfall for a mechanical user of statistical computer programs. Checking the latent roots of the matrix, $X'X$ the correlation matrix if standardized, or the VIFs (Variance Inflation Factors) is effective for detecting multicollinearity among the regressor variables.

An easy way to avoid multicollinearity is to delete or to transform the variables so that they are uncorrelated. It is less serious when aims of regression analysis is the only prediction by the independent variables, but it is very serious when emphasis lies on the estimation of individual parameters and contribution of the variables in the prediction system.

In such a case, a method of ridge regression, first proposed by Hoerl, is a useful method for overcoming multicollinearity without deleting any variables. To calculate ridge regression, the variables are first scaled so that the matrix, $X'X$, is a correlation matrix with unit diagonal elements, and then a small positive constant, k, is added to every diagonal element, which makes the matrix less singular. After that, regression coefficients, $\hat{\beta}^* = (X'X+kI)^{-1}X'y$, are obtained by the ordinary least square method.

A ridge estimate is not unbiased, but mean square error is smaller than that of least square estimate: the bias is conpensated by reduction in variance. The central question of ridge regression is how to choose the constant to achieve satisfactory balance between bias and variance. Eight methods of Hoerl's and others' to estimate the constant are introduced in this paper, and compared how they improved shortcomings of the results of the ordinary least square regression.

As far as a numerical example is concerned, the method that choose the constant, k, at the point where mean square error (summation of variance and square of bias) becomes minimum was a desireable one. This (new) method of ridge regression is expected to be applied increasingly in every field of behavioral sciences.