

数値実験による線形回帰における 多重共線性の影響力評価

竹内 秀一

Assessment of the Influence of Multicollinearity in Linear Regression through Simulation Studies

Hidekazu TAKEUCHI

The presence of multicollinearity is a major problem of linear regression that cannot be ignored. The presence of multicollinearity results in abnormalities in the data structure of the linear regression model and in the correlative relationship between explanatory variables in particular, causing the results of data analysis to lack stability. The use of ridge regression is one method of handling cases that are thought to have this kind of strong multicollinearity. This method makes it possible to somewhat limit multicollinearity, in terms the condition number. However, in order to use ridge regression, it is necessary to consider the assessment of the influence of individual observations in the linear regression as a preliminary analysis of the data. Without confirming the way in which the dependency of the data structure on the value of the condition number is expressed in the results of the linear regression, it is impossible to carry out a proper data analysis.

Accordingly, the representative influence measures of Cook's distance and likelihood distance are adopted in order to investigate from a condition number standpoint the problem of assessing the influence (resulting from multicollinearity) of individual observations on the linear regression. Assessment of influence based on the leverages and residuals of the individual observations, which are the basic elements making up the various influence measures, is an investigation of what effects changes in the value of the condition number have. Because it is difficult to change the value of the condition number when actual data is used, simulation studies using artificial data structures are carried out. Simulation studies, for which five typical condition number cases are established, are carried out and the extent to which the influence measures vary is investigated. Moreover, the characteristics of the cut-off points at which the

influence of individual observations on two representative influence measures is judged to be large are compared from a point of view relating to changes in the value of the condition number.

1. はじめに

線形回帰 (linear regression) において、多重共線性 (muticollinearity) の存在は無視することができない大きな問題である。多重共線性が存在すると、線形回帰モデルにおけるデータ構造、とりわけ説明変数間の相関関係に異常が生じ、データ解析の結果が安定性を欠く (回帰係数の推定値や残差の分布などに大きなぶれが生じる) ことになる。

この多重共線性の強さを測る尺度規準の1つとして、条件数 (condition number) がある。一般には、条件数の値が大きいと、線形回帰で用いる説明変数間に多重共線性が強く現れる傾向がみられる。多重共線性が強いと考えられる場合の対処方法の1つとして、Hoerl and Kennard[3] によって提案されたリッジ回帰 (ridge regression) を適用することがある。リッジ回帰を適用することにより、条件数の値に基づく多重共線性の強さをある程度抑えることができるのである。

しかしながら、リッジ回帰を適用するためには、事前のデータ解析として、線形回帰における個々の観測値の影響力評価について検討しておく必要がある。データ構造が条件数の値に依存することにより、線形回帰の結果 (回帰係数の推定値や残差の分布など) にどのような効果として現れるのかを確認しておかなければ、適切なデータ解析ができないのである。つまり、リッジ回帰を適用したことにより、条件数を基にして考える多重共線性の強さがどのように調整されたのか (緩和されるのかされないのか) という点について、個々の観測値に分解してその原因を見つけ出すことができないのである。

そこで、条件数の視点から多重共線性が及ぼす線形回帰における個々の観測値の影響力評価の問題点を調べるために、代表的な診断統計量 (influence measure) である Cook の距離 (Cook's distance) および尤度距離 (likelihood distance) を取り上げる (Nurunnabi, Imon and Nasser[4] などを参照)。それぞれの診断統計量を構成する基本要素である個々の観測値のてこ比および残差 (誤差) に基づく影響力評価が、条件数の値の変化によりどのような効果をもたらすのかを調べる。実データでは条件数の値を変化させることが困難であるので、竹内 [10] と同じ人工的なデータ構造を利用して数値実験を行う。この数値実験では、典型的な5つの条件数のケースを設定している。それぞれのケースについて10,000回のシミュレーションを行い、その結果から算出された診断統計量の変動の度合いを調べる。また、代表的な2つの診断統計量に対して、個々の観測値のもつ影響力が大きいと判定する打ち切り点 (cut-off point あるいは calibration point) の特徴についても、条件数の値の変化に関する観点

から比較をする。

本論文の構成は以下のとおりである。第2節では線形回帰およびリッジ回帰における各種の基本的な統計量を与える。また、リッジ回帰における統計量の特別な場合が通常の線形回帰における統計量であることを利用して、Cookの距離および尤度距離に関する各種の統計量も定義する。第3節において、説明変数行列に基づく条件数の値により5つのデータ構造を設定し、その設定されたケースに対して数値実験を行う。第4節は全体のまとめと今後の課題である。

2. 定義

本節では、線形回帰モデルにおける各種の基本的な統計量を与える。つぎに、線形回帰は、リッジ回帰の特別な場合として含まれるので、リッジ回帰における診断統計量としてCookの距離および尤度距離を定義した上で、それぞれの特別な場合として線形回帰における診断統計量を与える。また、それら2つの診断統計量の関係式も示す。さらに、これら2つの診断統計量に基づく個々の観測値の影響力を評価するために、それぞれの打切り点を与える。最後に、多重共線性を測る一つの尺度規準として条件数を定義する。

2.1 線形回帰およびリッジ回帰

ここでは、線形回帰モデルとして、

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

を考える。このとき、 $\mathbf{y}$ は $n \times 1$ の目的変数ベクトル、 $\mathbf{X}$ は $n \times p$ のフルランクの説明変数行列、 $\boldsymbol{\beta}$ は $p \times 1$ の回帰係数ベクトル、そして $\boldsymbol{\varepsilon}$ は $n \times 1$ の誤差ベクトルであり、正規分布 $N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ に従うものとする。ただし、 $\mathbf{I}_n$ は n 次の単位行列を表す。また、 $\boldsymbol{\beta}$ の最小2乗推定量（線形回帰においては最尤推定量とも一致）は $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ として得られ、誤差分散 σ^2 の不偏推定量は $\hat{\sigma}^2 = \mathbf{e}'\mathbf{e}/(n-p)$ となる。ただし、「 $'$ 」は行列あるいはベクトルの転置を表し、 $\mathbf{e}$ は残差ベクトルであり、 $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{I}_n - \mathbf{H})\mathbf{y}$ である。このとき、 $\mathbf{H}$ は説明変数行列 $\mathbf{X}$ から構成されるハット行列（hat matrix） $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ であり、その第 i 対角成分 h_{ii} がてこ比である。このてこ比については、 $1/n \leq h_{ii} < 1$ とする。さらに、残差ベクトル $\mathbf{e}$ の第 i 成分 e_i を標準化した $t_i = e_i/(\hat{\sigma}\sqrt{1-h_{ii}})$ を標準化残差（内的スチューデント化残差）とし、 t_i を第 i 成分とする標準化残差ベクトルを $\mathbf{t} = [\text{diag}(\mathbf{I}_n - \mathbf{H})]^{-\frac{1}{2}}\mathbf{e}/\hat{\sigma}$ とする。ただし、 $\text{diag}(\mathbf{A})$ は正方行列 $\mathbf{A}$ の対角成分のみを取り出し、非対角成分をすべて0にした行列を表す。

線形回帰の1つの代替的方法としてリッジ回帰がある（Hoerl and Kennard[3] や Grob[2]

数値実験による線形回帰における多重共線性の影響力評価

などを参照)。リッジ回帰における回帰係数ベクトル β の推定量（以下、リッジ推定量）を、リッジパラメータ $k(\geq 0)$ を導入することにより、

$$\hat{\beta}^R \equiv (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}\mathbf{X}'\mathbf{y}$$

と定義する。すると、最小2乗推定量の場合と同じく、残差ベクトル $\mathbf{e}^R$ は、 $\mathbf{e}^R = \mathbf{y} - \mathbf{X}\hat{\beta}^R = (\mathbf{I}_n - \mathbf{H}^R)\mathbf{y}$ となる。ただし、リッジ回帰におけるハット行列 $\mathbf{H}^R$ は $\mathbf{H}^R = \mathbf{X}(\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}\mathbf{X}'$ であり、その第 i 対角成分 h_{ii}^R がリッジ回帰における t_i の比である。このとき、 $0 < h_{ii}^R < 1$ である。

また、第 i 番目の観測値を除去したときの回帰係数ベクトル β の最小2乗推定量は

$$\hat{\beta}_{(i)} = (\mathbf{X}_{(i)}'\mathbf{X}_{(i)})^{-1}\mathbf{X}_{(i)}'\mathbf{y}_{(i)}$$

と定義される。ただし、添字の $(\cdot)$ は n 個の観測値の中から除去される観測値の番号を表す。通常の最小2乗推定量の場合と同様に、第 i 番目の観測値を除去したときのリッジ推定量は

$$\hat{\beta}_{(i)}^R = (\mathbf{X}_{(i)}'\mathbf{X}_{(i)} + k\mathbf{I}_p)^{-1}\mathbf{X}_{(i)}'\mathbf{y}_{(i)}$$

となる。 σ^2 の不偏推定量 $\hat{\sigma}^2$ の場合についても、第 i 番目の観測値を除去したときの推定量 $\hat{\sigma}_{(i)}^2$ を

$$\hat{\sigma}_{(i)}^2 = \frac{\mathbf{y}_{(i)}'(\mathbf{I}_n - \mathbf{H}_{(i)})\mathbf{y}_{(i)}}{n - p - 1} = \frac{\mathbf{y}_{(i)}'\mathbf{y}_{(i)} - \mathbf{y}_{(i)}'\mathbf{X}_{(i)}(\mathbf{X}_{(i)}'\mathbf{X}_{(i)})^{-1}\mathbf{X}_{(i)}'\mathbf{y}_{(i)}}{n - p - 1} = \frac{n - p - t_i^2}{n - p - 1} \hat{\sigma}^2$$

と定義する。さらに、標準化残差 t_i の定義式において、 $\hat{\sigma}$ の代わりに、この $\hat{\sigma}_{(i)}$ を用いた残差として $t_i^* = e_i / \{\hat{\sigma}_{(i)} \sqrt{1 - h_{ii}}\}$ を利用することもある。 t_i^* はスチューデント化残差（外的スチューデント化残差）と呼ばれる。このとき、 t_i および t_i^* の関係式は、

$$t_i = t_i^* \sqrt{\frac{n - p}{n - p - 1 + t_i^{*2}}} \quad (2.1)$$

となる。

特に、 $k=0$ とすれば、 $\hat{\beta}^R = \hat{\beta}$ 、 $\hat{\beta}_{(i)}^R = \hat{\beta}_{(i)}$ 、あるいは $\mathbf{H}^R = \mathbf{H}$ など、リッジ回帰において定義される統計量が通常の線形回帰の統計量を包含していることがわかる。

2.2 診断統計量

通常の線形回帰における尤度距離（Cook and Weisberg[1] を参照）をリッジ推定量に基づく診断統計量として拡張したものを定義する。この尤度距離は、代表的な診断統計量である Cook の距離とも関連するのでその定義式も示す。

まず、代表的な診断統計量である Cook の距離を定義する。リッジ推定量に基づく Cook

の距離 CD_i^R は, Takeuchi[5] によって提案された通常の線形回帰における Cook の距離を拡張することにより導入され,

$$CD_i^R \equiv \frac{(\hat{\beta}^R - \hat{\beta}_{(i)}^R)'(\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)(\hat{\beta}^R - \hat{\beta}_{(i)}^R)}{p\hat{\sigma}^2} = \{s_i(c_i + d_i)\}^2 \quad (2.2)$$

と与えられる (Takeuchi[6] を参照)。ただし, c_i は $\mathbf{c} = \Pi^{\frac{1}{2}}\mathbf{t}/\sqrt{p}$ の第 i 成分であり, このとき, $\Pi = \text{diag}(\mathbf{H})[\text{diag}(\mathbf{I}_n - \mathbf{H})]^{-1}$ であり, この第 i 対角成分が, $\pi_i = h_{ii}/(1 - h_{ii})$ である。同様に, d_i は $\mathbf{d} = \Pi^{\frac{1}{2}}\mathbf{u}/\sqrt{p}$ の第 i 成分であり, このとき, $\mathbf{u} = [\text{diag}(\mathbf{I}_n - \mathbf{H})]^{-\frac{1}{2}}(\mathbf{e}^R - \mathbf{e})/\hat{\sigma}$ であり, この第 i 成分が u_i である。最後に, s_i は $\mathbf{S} = \text{diag}(\mathbf{I}_n - \mathbf{H})[\text{diag}(\mathbf{I}_n - \mathbf{H}^R)]^{-1}$ の第 i 対角成分であり, $s_i = (1 - h_{ii})/(1 - h_{ii}^R)$ である。特に, $k=0$ のとき, つまり, 線形回帰においては, 竹内[10] などから

$$CD_i^R|_{k=0} = CD_i = c_i^2$$

となる。

(2.2) 式で与えられるリッジ推定量に基づく Cook の距離 CD_i^R においては, $\hat{\beta}^R - \hat{\beta}_{(i)}^R$ の中に挟まれる行列の選び方によっていくつかの定義式が考えられるが, ここでは, $\hat{\beta}^R$ の分散共分散行列 $\text{Var}(\hat{\beta}^R) = \sigma^2(\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}$ の逆行列を選ぶものとする (たとえば, Walker and Birch[11] を参照)。

つぎに, リッジ推定量に基づく尤度距離 LD_i^R を以下のように定義する (竹内 [9] を参照)。

$$LD_i^R \equiv 2[L(\hat{\beta}^R) - L(\hat{\beta}_{(i)}^R)] \quad (2.3)$$

ただし, 通常の線形回帰における最小 2 乗推定量の場合と同じく,

$$L(\hat{\beta}^R) : \beta \text{ が } \hat{\beta}^R \text{ のときの対数尤度} \quad \text{および} \quad L(\hat{\beta}_{(i)}^R) : \beta \text{ が } \hat{\beta}_{(i)}^R \text{ のときの対数尤度}$$

とする。(2.3) 式を (2.2) 式の Cook の距離 CD_i^R を利用して表現すると,

$$\begin{aligned} LD_i^R = n \log T_i + \frac{1}{T_i} \cdot \frac{n}{n-p} \left[\frac{h_{ii}^{R2}}{h_{ii}} p CD_i^R + 2s_i(t_i + u_i)u_i^* \right] \\ + n \left(\frac{1}{T_i} - 1 \right) \left[1 + \frac{\mathbf{u}' \text{diag}(\mathbf{I}_n - \mathbf{H}) \mathbf{u}}{n-p} \right] \end{aligned} \quad (2.4)$$

となる。ただし,

$$T_i = \frac{n}{n-1} \cdot \frac{n-p-t_i^2}{n-p}$$

であり, h_{ii}^{R2} は $\mathbf{H}^{R2} = (\mathbf{H}^R)^2$ の第 i 対角成分であり, それに u_i^* は $\mathbf{u}^* = [\text{diag}(\mathbf{I}_n - \mathbf{H})]^{-\frac{1}{2}}(\mathbf{e}^{R2} - \mathbf{e}^R)/\hat{\sigma}$ の第 i 成分である。ここで, $\mathbf{e}^{R2} = (\mathbf{I}_n - \mathbf{H}^{R2})\mathbf{y}$ である。特に, $k=0$ のとき, つまり, 線形回帰においては,

$$LD_i^R|_{k=0} = LD_i = n \log T_i + \frac{1}{T_i} \cdot \frac{np}{n-p} CD_i + n \left(\frac{1}{T_i} - 1 \right)$$

である。

したがって、尤度距離は Cook の距離の関数として表現できる。特に、通常の線形回帰の場合、個々の観測値の影響力評価においては、定数を除き、残差の関数である T_i に関する項による効果が、Cook の距離とは別途、追加的に反映されることがわかる。

2.3 診断統計量の打ち切り点

診断統計量の打ち切り点を定義する。まず、通常の線形回帰における尤度距離については、竹内 [7] において導入されている打ち切り点として、

$$LD_i > 4n \cdot \frac{4n+p(n-p)}{(n-p)^2(n-p+4)} \quad (2.5)$$

がある。(2.5) 式は、 $h_{ii}=p/n$ (てこ比がバランスした状態) で $|t_i^*|>2$ (外れ値) となる場合を影響力の大きい観測値であるとみなすことを判断基準としている。

これに対して、通常の線形回帰における Cook の距離については、竹内 [8] において導入されている打ち切り点として、

$$CD_i > \frac{1}{n-p-2} \quad (2.6)$$

が示されているが、これは、標準化残差 t_i に基づいて導かれているため、(2.5) 式の尤度距離の打ち切り点と単純に比較することができない。そこで、(2.5) 式と同等に扱うことのできる打ち切り点とするために、(2.6) 式を修正し、付録 A のようにして導いた

$$CD_i > \frac{4}{n-p+3} \quad (2.7)$$

を利用する。

2.4 条件数

説明変数間の多重共線性を測る尺度はいくつかあるが、代表的な尺度規準として、説明変数行列 $\mathbf{X}$ に基づいた $\mathbf{X}'\mathbf{X}$ の条件数を定義しておく。 $\mathbf{X}'\mathbf{X}$ の最大固有値 δ_1 と最小固有値 δ_p の比、つまり、条件数 $=\delta_1/\delta_p$ として定義する。

通常は、 $\mathbf{X}$ の特異値の比である $\sqrt{\delta_1/\delta_p}$ を条件数に用いるが、本論文では、竹内 [10] と関連した検討をするために、上記の固有値の比に着目する。条件数の値が大きくなると多重共線性が高い(強い)と判断され、説明変数行列の構造としては不安定であるとみなされる。

よって、条件数の値が大きくなると通常の線形回帰よりもリッジ回帰を適用することが望ましいデータ構造をもっているともいえる。

3. 数値実験

多重共線性を測る尺度規準として導入した条件数に基づく検討をするために、いくつかのデータ構造を与える。本論文では、竹内 [10] と同じ人工的なデータ構造について検討するため、条件数としてつぎの5つのケースを設定する。条件数Ⅰのケースとしては条件数が1229.24（竹内 [10] における条件数Ⅰの実験 No. 7）の場合を、条件数Ⅱのケースとしては条件数が430.59（同条件数Ⅱの実験 No. 1）の場合を、条件数Ⅲのケースとしては条件数が146.49（同条件数Ⅲの実験 No. 1）の場合を、条件数Ⅳのケースとしては条件数58.43（同条件数Ⅳの実験 No. 7）の場合を、そして条件数Ⅴのケースとしては条件数22.96（同条件数Ⅴの実験 No. 1）の場合を、それぞれ設定する。そして、竹内 [10] と同じ説明変数行列 $\mathbf{X}$ を利用し、誤差ベクトル $\boldsymbol{\varepsilon}$ の標準偏差 σ については、標準的なケースである標準偏差として $\sigma=1.00$ を適用する（バラツキの極端に大きいケースである標準偏差 $\sigma=5.00$ の場合および極端に小さいケースである標準偏差 $\sigma=0.01$ については除外する）。したがって、竹内 [10] と同じ設定の数値実験とするために、観測値数 $n=30$ および説明変数の数 $p=5$ （定数項を除く）として、説明変数行列 $\mathbf{X}$ は上記の5つのケースに固定する。また、目的変数ベクトル $\mathbf{y}$ については、説明変数行列の最大固有値 δ_1 に対する固有ベクトルを回帰係数ベクトル $\boldsymbol{\beta}$ に代入し、平均0で標準偏差 $\sigma=1.00$ となる正規乱数を生成して誤差ベクトル $\boldsymbol{\varepsilon}$ として加える。各ケースとも説明変数行列に基づくてこ比は固定し、目的変数ベクトルを正規乱数の利用により生成して、30個の観測値を構成する。5つのケースについて30個の観測値（実質的には目的変数のみ）の生成を10,000回ずつ繰返すシミュレーションを行うことになる。なお、参考までに、竹内 [10] の数値実験において、平均0で標準偏差 $\sigma=1.00$ の正規乱数を生成した場合の誤差ベクトル $\boldsymbol{\varepsilon}$ に基づく計算結果を、付録Bの表B.1から表B.5に示しておく。

線形回帰における観測値の影響力を調べるために、上記5つのケースに対して数値実験を行った結果をまとめたものが、表3.1（条件数Ⅰのケース）から表3.5（条件数Ⅴのケース）である。いずれも、10,000回のシミュレーション結果から、観測値ごとのてこ比（ $|h_{ii}|$ ：これは10,000回のシミュレーションにおいて同じ値で一定）、診断統計量の「平均」、その「標準誤差」、これらに関する「変動係数」（＝「標準誤差」/「平均」）、それに打切り点を超過した回数（「超過数」）が、2つの診断統計量 CD_i^* および LD_i^* （診断統計量の定義式とシミュレーションにより算出された数値結果を区別するために「s」を右肩に付けた）について算出されている。ここで、「平均」は個々の観測値について算出された診断統計量10,000回分の平均であり、「標準誤差」はその診断統計量の標準誤差である。また、打切り点の値は、5つのすべ

てのケースにおいて $n=30$ および $p=5+1(\text{定数項})=6$ は同じであるので、Cook の距離については (2.7) 式から 0.148 であり、尤度距離については (2.5) 式から 1.964 である。なお、表内の小数表示の数値は、紙面の都合で、小数第四位を四捨五入しているため、「変動係数」の数値表記が「標準誤差」/「平均」を単純に計算した数値とは異なる場合がある。実際の数値計算においては小数第十五位まで（正確には有効数字が 15 桁で）算出してあり、その個別の結果を小数第三位までの表記に直している。このため、数値表記上の丸め誤差が大きくなっている場合があり、一見すると数値的なズレがかなり大きいように思われるかもしれないが、変動係数の数値自体に誤りはない。

表 3.1 から表 3.5 までは共通する特徴は、シミュレーションにより算出された Cook の距離 CD_i と尤度距離 LD_i を比較すると、前者は変動係数が一定値（約 1.3）で安定しているが、後者は 1～3 の間で観測値ごとにかなりばらついている点である。この原因は、両者の関係式からもわかるように、 LD_i が CD_i の項に加えて、標準化残差 t_i の関数である T_i の項があるため、残差の効果による影響をより強く受けているためであると考えられる。よって、線形回帰における観測値の影響力評価を行う場合には、こうした残差による効果の差異が強く現れることを考慮する必要がある。

つぎに、打ち切り点の超過数については、 CD_i の打ち切り点を (2.7) 式で与えられるように従来よりも少し大きめ（影響力評価の点からは評価基準を緩め）に変更したが、(2.5) 式で与えられる LD_i の打ち切り点の場合よりも超過数の回数がすべての条件数のケースにおいて平均的には多くなっている（表 3.6 を参照）。また、超過数のばらつきである標準偏差も LD_i の打ち切り点の方が CD_i の打ち切り点よりも小さいので安定しているといえる。個々の観測値の目的変数を構成する誤差（項）については両者とも同じ乱数の値を用いているので、打ち切り点については、 CD_i がこれでも厳しいと考えるのか、 LD_i が緩いと考えるのかについて、厳密な検討を要する。ただ、観測値数 $n=30$ の 5% 程度が影響力の大きい観測値であると仮定すると、1 個か 2 個程度が検出されれば適正であると考えられるので、超過数の回数の多い観測値数が少ない LD_i の打ち切り点の方が妥当な評価をしているといえる。

さらに、詳しく検討するために、表 3.7 および表 3.8 のように「超過数」のみを取り出し、てこ比に基づく比較（広い意味では、条件数に基づく比較を含む）ができるように大小順に並べ替える。観測値ごとの目的変数を構成する誤差の入り方はすべて同一の確率分布（ここでは正規分布）であるので、残差の効果については同等であると想定できる。つまり、この数値実験における個々の観測値の影響力の大きな差異が生じる原因は、基本的にてこ比の値の大小によるものであると仮定する。超過数は各観測値とも繰返した 10,000 回中に、何回打ち切り点を超過したかの回数であるので、10,000 回中で超過数が 1,000 回を超えると 10% 超過している、つまり影響力が大きい観測値であると判定された比率（確率）がシミュレーションにおいて 10% を超えるということになる。一つの目安として、10%（超過数が 1,000 回）

を超える場合を比較してみると、条件数Ⅰのケースでは CD_i^* が8個（てこ比の順位で上位8位まで）と LD_i^* が3個であり、以下同様に、条件数Ⅱのケースでは6個と3個、条件数Ⅲのケースでは8個と3個（てこ比の順位で第3位となる観測値は10%未満なので除外）、条件数Ⅳのケースでは8個と2個、そして条件数Ⅴのケースでは10個と4個となっている。ケースにより若干のばらつきはあるが、概ね CD_i^* が約30%の観測値数である8個を、 LD_i^* が10%の3個を、それぞれ影響力の大きい観測値であると判定している。影響力評価の判定規準を厳しくして、超過数が5%を超える場合に変更すると、概ね CD_i^* が約50%の観測値数である14個を、 LD_i^* が30%の9個を影響力の大きい観測値と判定する。したがって、 CD_i^* の打ち切り点が確率10%であり、 LD_i^* の打ち切り点が確率5%である場合に対応しているともみなせ、観測値の影響力が大きいと判定する打ち切り点としてはCookの距離が尤度距離よりも厳しく判定することがわかる。 CD_i^* と LD_i^* は、診断統計量として観測値の影響力の大きさを測る視点が異なるので、どちらが良いとか悪いとかは議論できないが、単純な打ち切り点に基づく影響力評価を行った場合には、このような大きな差異があることに注意を要することがわかる。

最後に、条件数に着目した別の視点から、表3.7および表3.8を検討してみる。これらの表の各超過数の右に「*」が付いているのは、てこ比の大小順と超過数の回数の大小順において順位が異なるものであることを示している。これをみると、条件数が小さくなる（条件数Ⅰのケースから条件数Ⅴのケースの順）につれて、大小順が異なる観測値が増える傾向にあることがわかる（ただし、条件数Ⅳのケースでは減っている）。特に、てこ比の値が大きい（てこ比の順位が上位の）観測値において、条件数が小さくなると超過数との対応関係にずれが数多く生じている。これは、多重共線性が原因であると考えられる。条件数Ⅰや条件数Ⅱのケースのように条件数の値が大きい場合は、てこ比の値が大きい観測値がそのまま影響力の大きい観測値として判定されるが、条件数Ⅲから条件数Ⅴまでのケースのように条件数の値が約150以下になるような場合は、てこ比の値の大きさだけではなく残差の大きさによる効果も反映されるため、てこ比の値の大小だけで観測値の影響力評価をすることができないことが読み取れる。つまり、条件数の数値が概ね400を超えるような場合には、てこ比による影響力が強くなり過ぎる可能性が高いことを示しているので、リッジ回帰などの手法を適用することにより、てこ比の影響力を適度に調整する必要があるといえる。こうした改善あるいは改良をしなければ、データ解析における観測値の影響力評価を説明変数行列の主要な成分であるてこ比からの視点だけで判定してしまう危険性が高くなる。各観測値のもつてこ比および残差の効果をバランス良く取り上げ、適切な影響力評価をするためにも、条件数の値が大きいケースについては最大限の配慮をする必要があるといえる。

表 3.1 シミュレーションの結果 (条件数 I のケース)

No.	h_{ii}	CD_i^s				LD_i^s			
		平均	標準誤差	変動係数	超過数	平均	標準誤差	変動係数	超過数
1	0.165	0.033	0.043	1.319	278	0.333	0.625	1.875	224
2	0.161	0.032	0.043	1.352	267	0.333	0.735	2.204	237
3	0.302	0.072	0.096	1.331	1,535	0.663	1.150	1.734	806
4	0.098	0.018	0.024	1.326	14	0.210	0.457	2.175	107
5	0.153	0.030	0.041	1.344	253	0.322	0.634	1.970	253
6	0.375	0.100	0.134	1.344	2,261	0.896	1.480	1.652	1,278
7	0.195	0.040	0.053	1.321	518	0.392	0.695	1.770	337
8	0.085	0.015	0.020	1.323	2	0.186	0.424	2.281	83
9	0.305	0.076	0.100	1.323	1,653	0.701	1.229	1.752	862
10	0.258	0.059	0.079	1.338	1,126	0.556	0.987	1.777	626
11	0.087	0.016	0.021	1.327	0	0.198	0.427	2.154	102
12	0.203	0.042	0.057	1.333	616	0.416	0.755	1.814	388
13	0.210	0.045	0.059	1.307	685	0.439	0.778	1.771	406
14	0.216	0.045	0.060	1.323	647	0.434	0.804	1.852	365
15	0.126	0.024	0.032	1.326	88	0.259	0.523	2.018	154
16	0.174	0.035	0.046	1.321	356	0.351	0.659	1.876	276
17	0.342	0.086	0.116	1.346	1,960	0.784	1.367	1.745	1,032
18	0.093	0.017	0.023	1.347	6	0.213	0.662	3.115	122
19	0.101	0.019	0.025	1.322	18	0.220	0.458	2.075	121
20	0.212	0.045	0.059	1.309	689	0.435	0.765	1.757	381
21	0.238	0.052	0.069	1.327	874	0.499	0.938	1.880	483
22	0.295	0.069	0.092	1.339	1,436	0.631	1.093	1.733	760
23	0.265	0.060	0.078	1.308	1,203	0.555	0.929	1.675	636
24	0.133	0.026	0.034	1.336	121	0.279	0.562	2.011	181
25	0.124	0.024	0.032	1.331	86	0.271	0.594	2.194	162
26	0.392	0.108	0.142	1.320	2,471	0.958	1.583	1.653	1,423
27	0.138	0.026	0.035	1.344	152	0.281	0.568	2.022	187
28	0.248	0.055	0.073	1.310	976	0.523	0.913	1.745	524
29	0.170	0.034	0.045	1.332	328	0.349	0.698	1.998	267
30	0.135	0.026	0.035	1.352	153	0.286	0.600	2.097	208

表 3.2 シミュレーションの結果 (条件数 II のケース)

No.	h_{ii}	CD_i^s				LD_i^s			
		平均	標準誤差	変動係数	超過数	平均	標準誤差	変動係数	超過数
1	0.173	0.035	0.047	1.336	365	0.361	0.697	1.934	279
2	0.171	0.034	0.045	1.315	312	0.349	0.680	1.950	240
3	0.265	0.061	0.081	1.339	1,216	0.569	0.995	1.749	657
4	0.238	0.052	0.070	1.344	922	0.498	0.934	1.876	502
5	0.302	0.072	0.096	1.333	1,538	0.665	1.172	1.763	826
6	0.204	0.043	0.058	1.345	650	0.424	0.773	1.823	398
7	0.293	0.068	0.089	1.325	1,430	0.618	1.036	1.678	719
8	0.187	0.039	0.052	1.345	460	0.396	0.794	2.006	331
9	0.137	0.027	0.034	1.293	125	0.281	0.524	1.865	178
10	0.247	0.054	0.072	1.338	959	0.510	0.907	1.778	509
11	0.070	0.013	0.017	1.338	0	0.173	0.417	2.408	94
12	0.085	0.016	0.021	1.331	4	0.197	0.456	2.310	107
13	0.074	0.013	0.018	1.359	2	0.186	0.508	2.726	94
14	0.372	0.097	0.129	1.333	2,207	0.860	1.434	1.668	1,231
15	0.224	0.048	0.065	1.350	794	0.468	0.873	1.866	455
16	0.212	0.045	0.060	1.324	712	0.443	0.790	1.783	387
17	0.172	0.034	0.045	1.345	323	0.343	0.712	2.078	253
18	0.164	0.033	0.044	1.344	313	0.337	0.655	1.943	266
19	0.507	0.170	0.228	1.338	3,602	1.478	2.364	1.599	2,262
20	0.226	0.049	0.063	1.297	776	0.461	0.805	1.748	407
21	0.220	0.047	0.063	1.343	732	0.460	0.861	1.871	443
22	0.122	0.023	0.030	1.340	67	0.249	0.529	2.126	141
23	0.136	0.026	0.035	1.333	133	0.278	0.578	2.078	183
24	0.104	0.019	0.026	1.340	21	0.226	0.483	2.134	132
25	0.072	0.013	0.017	1.365	0	0.174	0.500	2.872	79
26	0.107	0.020	0.027	1.352	36	0.236	0.537	2.278	151
27	0.350	0.090	0.120	1.335	2,073	0.814	1.375	1.690	1,110
28	0.181	0.037	0.050	1.349	440	0.376	0.733	1.950	309
29	0.225	0.049	0.064	1.321	801	0.469	0.829	1.767	458
30	0.161	0.032	0.042	1.309	266	0.332	0.605	1.819	241

表 3.3 シミュレーションの結果 (条件数 III のケース)

No.	h_{ii}	CD_i^s				LD_i^s			
		平均	標準誤差	変動係数	超過数	平均	標準誤差	変動係数	超過数
1	0.096	0.018	0.024	1.326	11	0.212	0.441	2.085	108
2	0.168	0.034	0.044	1.318	317	0.342	0.634	1.851	247
3	0.245	0.053	0.069	1.308	925	0.494	0.833	1.685	504
4	0.234	0.051	0.066	1.295	868	0.477	0.798	1.674	444
5	0.101	0.018	0.025	1.333	20	0.217	0.463	2.136	118
6	0.347	0.088	0.118	1.339	2,002	0.799	1.356	1.698	1,081
7	0.165	0.033	0.044	1.328	286	0.335	0.641	1.913	251
8	0.242	0.053	0.070	1.318	938	0.503	0.893	1.774	487
9	0.109	0.020	0.027	1.321	26	0.231	0.493	2.131	131
10	0.346	0.085	0.111	1.313	1,948	0.751	1.205	1.605	990
11	0.184	0.038	0.051	1.328	457	0.386	0.736	1.908	329
12	0.166	0.034	0.045	1.345	291	0.354	0.737	2.082	249
13	0.189	0.039	0.051	1.323	466	0.386	0.740	1.915	305
14	0.396	0.108	0.145	1.341	2,477	0.960	1.625	1.691	1,404
15	0.154	0.030	0.040	1.340	219	0.310	0.601	1.943	216
16	0.114	0.021	0.028	1.338	51	0.244	0.541	2.222	140
17	0.272	0.063	0.085	1.339	1,290	0.595	1.067	1.792	677
18	0.100	0.019	0.024	1.304	18	0.219	0.470	2.145	112
19	0.146	0.028	0.038	1.325	164	0.300	0.592	1.970	191
20	0.103	0.019	0.025	1.317	17	0.221	0.449	2.037	120
21	0.161	0.032	0.043	1.328	271	0.338	0.687	2.031	240
22	0.173	0.035	0.046	1.329	360	0.350	0.642	1.832	276
23	0.112	0.021	0.029	1.354	57	0.246	0.557	2.266	146
24	0.306	0.072	0.095	1.323	1,544	0.655	1.093	1.669	785
25	0.224	0.049	0.065	1.323	811	0.474	0.847	1.787	480
26	0.296	0.069	0.093	1.350	1,431	0.636	1.127	1.773	751
27	0.256	0.059	0.078	1.317	1,118	0.559	0.975	1.743	607
28	0.344	0.088	0.117	1.333	2,041	0.797	1.354	1.699	1,095
29	0.122	0.023	0.031	1.321	71	0.259	0.541	2.090	149
30	0.133	0.025	0.033	1.326	104	0.273	0.597	2.192	166

表 3.4 シミュレーションの結果 (条件数 IV のケース)

No.	h_{ii}	CD $_i^s$				LD $_i^s$			
		平均	標準誤差	変動係数	超過数	平均	標準誤差	変動係数	超過数
1	0.237	0.051	0.069	1.336	911	0.491	0.883	1.800	484
2	0.199	0.042	0.056	1.342	590	0.414	0.786	1.900	360
3	0.183	0.037	0.049	1.320	387	0.366	0.716	1.954	269
4	0.224	0.047	0.062	1.333	718	0.450	0.866	1.926	424
5	0.120	0.023	0.030	1.339	66	0.255	0.559	2.192	157
6	0.326	0.080	0.108	1.351	1,766	0.727	1.228	1.690	954
7	0.214	0.045	0.060	1.342	680	0.433	0.779	1.799	412
8	0.094	0.018	0.024	1.339	13	0.218	0.505	2.316	120
9	0.263	0.060	0.078	1.314	1,146	0.557	0.955	1.715	596
10	0.350	0.090	0.119	1.328	2,036	0.809	1.346	1.663	1,109
11	0.099	0.018	0.024	1.333	20	0.219	0.507	2.322	110
12	0.342	0.086	0.114	1.333	1,969	0.773	1.307	1.692	1,048
13	0.243	0.054	0.073	1.351	985	0.516	0.932	1.806	539
14	0.260	0.059	0.079	1.339	1,120	0.557	0.989	1.775	617
15	0.141	0.028	0.037	1.331	168	0.297	0.584	1.966	209
16	0.192	0.041	0.054	1.326	534	0.407	0.774	1.903	347
17	0.197	0.041	0.054	1.334	533	0.400	0.754	1.886	343
18	0.219	0.047	0.064	1.346	772	0.463	0.856	1.849	449
19	0.126	0.023	0.032	1.350	80	0.256	0.526	2.055	160
20	0.169	0.034	0.046	1.348	325	0.350	0.688	1.965	261
21	0.208	0.044	0.057	1.297	633	0.427	0.741	1.737	397
22	0.138	0.027	0.036	1.352	148	0.291	0.587	2.013	207
23	0.163	0.032	0.043	1.344	289	0.334	0.654	1.957	248
24	0.113	0.021	0.028	1.353	58	0.244	0.554	2.274	148
25	0.089	0.016	0.021	1.320	3	0.195	0.436	2.235	90
26	0.275	0.063	0.083	1.309	1,283	0.586	1.013	1.729	626
27	0.089	0.017	0.022	1.331	0	0.207	0.485	2.350	116
28	0.325	0.080	0.107	1.338	1,770	0.727	1.248	1.717	937
29	0.147	0.029	0.039	1.343	201	0.310	0.688	2.218	229
30	0.253	0.057	0.076	1.338	1,083	0.542	0.958	1.768	597

表 3.5 シミュレーションの結果 (条件数 V のケース)

No.	h_{ii}	CD_i^s				LD_i^s			
		平均	標準誤差	変動係数	超過数	平均	標準誤差	変動係数	超過数
1	0.130	0.025	0.033	1.345	106	0.270	0.535	1.984	182
2	0.349	0.087	0.116	1.332	1,986	0.785	1.345	1.714	1,033
3	0.433	0.127	0.169	1.330	2,906	1.114	1.806	1.621	1,695
4	0.090	0.017	0.022	1.321	1	0.204	0.430	2.110	115
5	0.291	0.069	0.093	1.340	1,480	0.645	1.128	1.750	773
6	0.145	0.028	0.038	1.318	168	0.299	0.571	1.906	191
7	0.185	0.037	0.049	1.332	422	0.368	0.683	1.859	278
8	0.110	0.021	0.027	1.323	38	0.237	0.521	2.197	119
9	0.163	0.033	0.044	1.335	269	0.342	0.696	2.037	239
10	0.170	0.034	0.045	1.325	348	0.346	0.639	1.847	274
11	0.251	0.057	0.075	1.320	1,073	0.538	0.976	1.816	567
12	0.198	0.041	0.055	1.346	522	0.400	0.751	1.875	350
13	0.132	0.025	0.033	1.308	99	0.272	0.523	1.926	172
14	0.099	0.019	0.025	1.332	20	0.225	0.509	2.260	142
15	0.113	0.021	0.028	1.330	44	0.237	0.485	2.044	138
16	0.293	0.069	0.091	1.309	1,468	0.636	1.079	1.698	760
17	0.134	0.025	0.034	1.320	106	0.273	0.533	1.955	160
18	0.170	0.034	0.045	1.318	317	0.343	0.652	1.901	252
19	0.215	0.045	0.060	1.326	680	0.438	0.792	1.808	399
20	0.274	0.063	0.084	1.344	1,243	0.589	1.101	1.870	661
21	0.153	0.030	0.040	1.331	205	0.312	0.607	1.945	206
22	0.070	0.013	0.017	1.329	0	0.170	0.402	2.361	80
23	0.141	0.027	0.037	1.346	149	0.294	0.595	2.026	193
24	0.156	0.031	0.041	1.341	252	0.321	0.625	1.949	243
25	0.282	0.066	0.087	1.328	1,346	0.611	1.059	1.734	722
26	0.349	0.088	0.118	1.330	2,006	0.797	1.401	1.758	1,058
27	0.174	0.035	0.047	1.353	362	0.359	0.718	1.998	280
28	0.108	0.020	0.027	1.329	40	0.238	0.540	2.263	141
29	0.274	0.063	0.083	1.326	1,269	0.583	0.993	1.701	657
30	0.348	0.089	0.120	1.339	2,000	0.807	1.372	1.699	1,101

表 3.6 打ち切り点の超過数

	条件数 I		条件数 II		条件数 III		条件数 IV		条件数 V	
	CD_i^s	LD_i^s	CD_i^s	LD_i^s	CD_i^s	LD_i^s	CD_i^s	LD_i^s	CD_i^s	LD_i^s
平均	692.4	433.0	709.3	448.1	686.6	426.6	676.2	418.8	697.5	439.4
標準偏差	714.2	353.1	809.3	445.8	729.8	349.8	618.1	287.8	783.7	387.6

表 3.7 シミュレーションの結果（てこ比に基づく総括表）前半

順位	条件数 I			条件数 II			条件数 III		
	h_{ii}	CD 超過数	LD 超過数	h_{ii}	CD 超過数	LD 超過数	h_{ii}	CD 超過数	LD 超過数
1	0.392	2,471	1,423	0.507	3,602	2,262	0.396	2,477	1,404
2	0.375	2,261	1,278	0.372	2,207	1,231	0.347	2,002*	1,081*
3	0.342	1,960	1,032	0.350	2,073	1,110	0.346	1,948*	990*
4	0.305	1,653	862	0.302	1,538	826	0.344	2,041*	1,095*
5	0.302	1,535	806	0.293	1,430	719	0.306	1,544	785
6	0.295	1,436	760	0.265	1,216	657	0.296	1,431	751
7	0.265	1,203	636	0.247	959	509	0.272	1,290	677
8	0.258	1,126	626	0.238	922	502	0.256	1,118	607
9	0.248	976	524	0.226	776*	407*	0.245	925*	504
10	0.238	874	483	0.225	801*	458*	0.242	938*	487
11	0.216	647*	365*	0.224	794*	455*	0.234	868	444*
12	0.212	689*	381*	0.220	732	443*	0.224	811	480*
13	0.210	685*	406*	0.212	712	387*	0.189	466	305*
14	0.203	616	388*	0.204	650	398*	0.184	457	329*
15	0.195	518	337	0.187	460	331	0.173	360	276
16	0.174	356	276	0.181	440	309	0.168	317	247*
17	0.170	328	267	0.173	365	279	0.166	291	249*
18	0.165	278	224*	0.172	323	253*	0.165	286	251*
19	0.161	267	237*	0.171	312*	240*	0.161	271	240
20	0.153	253	253*	0.164	313*	266*	0.154	219	216
21	0.138	152*	187*	0.161	266	241*	0.146	164	191
22	0.135	153*	208*	0.137	125*	178*	0.133	104	166
23	0.133	121	181	0.136	133*	183*	0.122	71	149
24	0.126	88	154*	0.122	67	141*	0.114	51*	140*
25	0.124	86	162*	0.107	36	151*	0.112	57*	146*
26	0.101	18	121*	0.104	21	132	0.109	26	131
27	0.098	14	107*	0.085	4	107	0.103	17*	120
28	0.093	6	122*	0.074	2	94	0.101	20*	118
29	0.087	0*	102	0.072	0	79*	0.100	18*	112
30	0.085	2*	83	0.070	0	94*	0.096	11	108

表 3.8 シミュレーションの結果（てこ比に基づく総括表）後半

順位	条件数 IV			条件数 V		
	h_{ii}	CD 超過数	LD 超過数	h_{ii}	CD 超過数	LD 超過数
1	0.350	2,036	1,109	0.433	2,906	1,695
2	0.342	1,969	1,048	0.349	2,006	1,058*
3	0.326	1,766*	954	0.349	1,986*	1,033*
4	0.325	1,770*	937	0.348	2,000*	1,101*
5	0.275	1,283	626	0.293	1,468*	760*
6	0.263	1,146	596*	0.291	1,480*	773*
7	0.260	1,120	617*	0.282	1,346	722
8	0.253	1,083	597*	0.274	1,269	657*
9	0.243	985	539	0.274	1,243	661*
10	0.237	911	484	0.251	1,073	567
11	0.224	718*	424*	0.215	680	399
12	0.219	772*	449*	0.198	522	350
13	0.214	680	412	0.185	422	278*
14	0.208	633	397	0.174	362	280*
15	0.199	590	360	0.170	348	274
16	0.197	533*	343*	0.170	317	252
17	0.192	534*	347*	0.163	269	239*
18	0.183	387	269	0.156	252	243*
19	0.169	325	261	0.153	205	206
20	0.163	289	248	0.145	168	191*
21	0.147	201	229	0.141	149	193*
22	0.141	168	209	0.134	106	160*
23	0.138	148	207	0.132	99*	172*
24	0.126	80	160	0.130	106*	182*
25	0.120	66	157	0.113	44	138*
26	0.113	58	148	0.110	38*	119*
27	0.099	20	110*	0.108	40*	141*
28	0.094	13	120*	0.099	20	142*
29	0.089	0*	116*	0.090	1	115
30	0.089	3*	90	0.070	0	80

4. まとめと今後の課題

本論文では、線形回帰における影響力評価について、人工的なデータに基づく数値実験により検討を行った。具体的には、代表的な診断統計量として Cook の距離 CD_i および尤度距離 LD_i の2つを取り上げ、てこ比と残差（誤差）の影響力の評価が条件数の値の変化に対してどのような効果を与えるのかを検討した。10,000回のシミュレーション結果から算出された Cook の距離 CD_i および尤度距離 LD_i の比較においては、観測値の影響力が大きいと判定する打ち切り点としては Cook の距離が尤度距離よりも厳しく判定することが判明した。また、条件数については、その値が大きい場合には、てこ比の値により影響力の大きさが過大評価されてしまう可能性があるという問題点が浮き彫りになった。条件数の値が大きい場合には、リッジ回帰等を適用することにより、てこ比の効果を適度に調整する（抑える）必要性があることもわかった。

今後の課題としては、線形回帰における観測値の影響力評価において上記のような問題点を念頭に置き、リッジ回帰における影響力評価を適正に行うことを考える必要がある。とりわけ、個々の観測値の影響力評価を行う場合に、条件数の値が大きいケースにおいてはリッジ回帰を適用することにより、どの程度の影響力が調整されるのかを検討したい。

付録 A：(2.7) 式の導出

線形回帰における Cook の距離の打ち切り点を導出する。線形回帰における Cook の距離は、第2節の定義式から

$$CD_i = c_i^2 = \frac{1}{p} \cdot \frac{h_{ii}}{1-h_{ii}} t_i^2 \quad (\text{A.1})$$

と書き直すことができる。標準化残差 t_i および学生化残差 t_i^* の関係式は、(2.1) 式のように表されるので、(A.1) 式の Cook の距離の表現を t_i^* を用いて表せば

$$CD_i = \frac{n-p}{p} \cdot \frac{h_{ii}}{1-h_{ii}} \cdot \frac{t_i^{*2}}{n-p-1+t_i^{*2}} \quad (\text{A.2})$$

となる。

(A.2) 式に対する打ち切り点を尤度距離の場合（竹内 [7] を参照）と同じく、 $h_{ii}=p/n$ （てこ比がバランスした状態）で、 $|t_i^*|>2$ （外れ値）という場合で設定する。これらを (A.2) 式に代入すれば、Cook の距離の打ち切り点は

$$CD_i > \frac{n-p}{p} \cdot \frac{p/n}{1-p/n} \cdot \frac{2^2}{n-p-1+2^2} = \frac{4}{n-p+3}$$

となり、(2.7) 式を導くことができる。

(2.6) 式と (2.7) 式の大小関係を調べると、

$$n-p > \frac{11}{3}$$

の場合であれば、(2.7) 式の打ち切り点が (2.6) 式よりも大きな値になる。つまり、個々の観測値の影響力を調べるときに、(2.7) 式の打ち切り点の方が (2.6) 式よりも大きな値になるので、影響力の大きい観測値であると判定する割合が小さくなる緩い打ち切り点ということになる。

付録 B：線形回帰分析の結果

付録として、5つの条件数のケースそれぞれについて、標準偏差 $\sigma=1.00$ のケースのみに線形回帰を適用した場合の計算結果を参考までに提示する。

表 B. 1 線形回帰分析の結果（条件数 I のケース）

No.	$k=0$				
	h_{ii}	t_i	T_i	CD_i	LD_i
1	0.165	-0.015	1.034	0.000	0.017
2	0.161	0.350	1.029	0.004	0.041
3	0.302	-0.318	1.030	0.007	0.066
4	0.098	1.041	0.988	0.020	0.152
5	0.153	0.752	1.010	0.017	0.128
6	0.375	2.474	0.771	0.611	7.060
7	0.195	0.599	1.019	0.014	0.112
8	0.085	-1.382	0.952	0.029	0.269
9	0.305	-0.049	1.034	0.000	0.018
10	0.258	-2.360	0.794	0.323	3.910
11	0.087	0.632	1.017	0.006	0.051
12	0.203	-0.813	1.006	0.028	0.210
13	0.210	-0.305	1.030	0.004	0.044
14	0.216	0.060	1.034	0.000	0.018
15	0.126	-0.625	1.018	0.009	0.074
16	0.174	-0.979	0.993	0.034	0.255
17	0.342	-0.526	1.023	0.024	0.183
18	0.093	-0.013	1.034	0.000	0.017
19	0.101	0.114	1.034	0.000	0.018
20	0.212	1.071	0.985	0.052	0.396
21	0.238	-1.464	0.942	0.112	0.945
22	0.295	-1.272	0.965	0.113	0.897
23	0.265	0.624	1.018	0.023	0.177
24	0.133	1.346	0.956	0.046	0.392
25	0.124	0.322	1.030	0.002	0.031
26	0.392	-0.201	1.033	0.004	0.047
27	0.138	0.959	0.995	0.024	0.185
28	0.248	1.744	0.903	0.167	1.549
29	0.170	-1.046	0.987	0.037	0.286
30	0.135	-0.714	1.012	0.013	0.100

表 B.2 線形回帰分析の結果（条件数 II のケース）

No.	$k=0$				
	h_{ii}	t_i	T_i	CD_i	LD_i
1	0.173	-0.781	1.008	0.021	0.160
2	0.171	0.897	1.000	0.028	0.208
3	0.265	1.855	0.886	0.207	1.979
4	0.238	0.242	1.032	0.003	0.037
5	0.302	-0.839	1.004	0.051	0.379
6	0.204	-1.074	0.985	0.049	0.378
7	0.293	0.378	1.028	0.010	0.083
8	0.187	-0.824	1.005	0.026	0.195
9	0.137	0.384	1.028	0.004	0.040
10	0.247	-2.327	0.801	0.296	3.572
11	0.070	0.433	1.026	0.002	0.027
12	0.085	-0.910	0.999	0.013	0.096
13	0.074	-0.154	1.033	0.000	0.018
14	0.372	0.643	1.017	0.041	0.305
15	0.224	-1.079	0.984	0.056	0.431
16	0.212	-0.600	1.019	0.016	0.124
17	0.172	-0.222	1.032	0.002	0.027
18	0.164	0.278	1.031	0.003	0.032
19	0.507	-0.250	1.032	0.011	0.093
20	0.226	-0.500	1.024	0.012	0.097
21	0.220	0.829	1.005	0.032	0.242
22	0.122	0.285	1.031	0.002	0.028
23	0.136	0.059	1.034	0.000	0.018
24	0.104	-0.232	1.032	0.001	0.022
25	0.072	-0.158	1.033	0.000	0.018
26	0.107	2.227	0.821	0.099	1.530
27	0.350	-0.833	1.005	0.062	0.466
28	0.181	0.034	1.034	0.000	0.017
29	0.225	-0.569	1.021	0.016	0.121
30	0.161	2.546	0.755	0.207	3.358

表 B.3 線形回帰分析の結果 (条件数 III のケース)

No.	$k=0$				
	h_{ii}	t_i	T_i	CD_i	LD_i
1	0.096	1.364	0.954	0.033	0.292
2	0.168	0.843	1.004	0.024	0.179
3	0.245	-1.155	0.977	0.072	0.560
4	0.234	-0.040	1.034	0.000	0.018
5	0.101	1.003	0.991	0.019	0.143
6	0.347	1.165	0.976	0.120	0.931
7	0.165	0.708	1.013	0.016	0.125
8	0.242	-1.789	0.896	0.170	1.612
9	0.109	-0.637	1.017	0.008	0.065
10	0.346	-0.646	1.016	0.037	0.275
11	0.184	-0.711	1.013	0.019	0.143
12	0.166	2.904	0.671	0.279	5.855
13	0.189	0.049	1.034	0.000	0.018
14	0.396	-0.427	1.027	0.020	0.156
15	0.154	-0.878	1.001	0.023	0.175
16	0.114	1.165	0.976	0.029	0.231
17	0.272	-0.336	1.030	0.007	0.064
18	0.100	0.473	1.025	0.004	0.039
19	0.146	-0.729	1.012	0.015	0.114
20	0.103	-0.038	1.034	0.000	0.017
21	0.161	-0.678	1.015	0.015	0.112
22	0.173	0.759	1.010	0.020	0.150
23	0.112	-0.034	1.034	0.000	0.017
24	0.306	-0.692	1.014	0.035	0.263
25	0.224	-0.419	1.027	0.008	0.072
26	0.296	1.040	0.988	0.076	0.578
27	0.256	-0.896	1.000	0.046	0.345
28	0.344	-0.919	0.998	0.074	0.554
29	0.122	-1.418	0.948	0.046	0.412
30	0.133	0.572	1.020	0.008	0.067

表 B. 4 線形回帰分析の結果 (条件数 IV のケース)

No.	$k=0$				
	h_{ii}	t_i	T_i	CD_i	LD_i
1	0.237	-1.798	0.895	0.167	1.591
2	0.199	-0.424	1.027	0.007	0.065
3	0.183	-0.988	0.992	0.036	0.276
4	0.224	1.515	0.936	0.111	0.955
5	0.120	-1.016	0.990	0.024	0.180
6	0.326	-1.454	0.943	0.171	1.410
7	0.214	0.724	1.012	0.024	0.179
8	0.094	-0.275	1.031	0.001	0.024
9	0.263	0.404	1.027	0.010	0.082
10	0.350	-0.120	1.034	0.001	0.026
11	0.099	0.216	1.032	0.001	0.021
12	0.342	0.504	1.024	0.022	0.169
13	0.243	-1.472	0.941	0.116	0.980
14	0.260	0.963	0.995	0.054	0.410
15	0.141	0.770	1.009	0.016	0.122
16	0.192	-0.567	1.021	0.013	0.100
17	0.197	0.188	1.033	0.001	0.026
18	0.219	-0.438	1.026	0.009	0.075
19	0.126	0.238	1.032	0.001	0.025
20	0.169	0.898	1.000	0.027	0.205
21	0.208	-1.345	0.957	0.079	0.653
22	0.138	1.500	0.937	0.060	0.544
23	0.163	1.796	0.895	0.105	1.069
24	0.113	2.113	0.842	0.094	1.312
25	0.089	-0.431	1.026	0.003	0.032
26	0.275	0.279	1.031	0.005	0.050
27	0.089	0.065	1.034	0.000	0.017
28	0.325	-0.627	1.018	0.032	0.237
29	0.147	-0.392	1.028	0.004	0.043
30	0.253	-1.160	0.976	0.076	0.592

表 B.5 線形回帰分析の結果（条件数 V のケース）

No.	$k=0$				
	h_{ii}	t_i	T_i	CD_i	LD_i
1	0.130	-0.976	0.993	0.024	0.180
2	0.349	-1.084	0.984	0.105	0.805
3	0.433	-2.054	0.853	0.536	5.119
4	0.090	0.316	1.030	0.002	0.025
5	0.291	1.560	0.930	0.167	1.427
6	0.145	-0.229	1.032	0.001	0.026
7	0.185	-0.306	1.030	0.004	0.039
8	0.110	1.675	0.914	0.058	0.600
9	0.163	-0.472	1.025	0.007	0.062
10	0.170	-0.968	0.994	0.032	0.242
11	0.251	2.744	0.710	0.420	6.422
12	0.198	-0.612	1.018	0.015	0.119
13	0.132	-0.908	0.999	0.021	0.157
14	0.099	-0.641	1.017	0.007	0.059
15	0.113	-0.978	0.993	0.020	0.154
16	0.293	-0.371	1.029	0.009	0.081
17	0.134	0.127	1.034	0.000	0.019
18	0.170	1.465	0.942	0.073	0.636
19	0.215	-0.700	1.013	0.022	0.168
20	0.274	-0.184	1.033	0.002	0.031
21	0.153	1.509	0.936	0.069	0.615
22	0.070	0.009	1.034	0.000	0.017
23	0.141	1.026	0.989	0.029	0.221
24	0.156	-0.111	1.034	0.000	0.019
25	0.282	-0.848	1.003	0.047	0.353
26	0.349	0.007	1.034	0.000	0.017
27	0.174	0.908	0.999	0.029	0.217
28	0.108	0.373	1.028	0.003	0.032
29	0.274	-0.082	1.034	0.000	0.020
30	0.348	-0.643	1.017	0.037	0.276

参 考 文 献

- [1] Cook, R. D. and Weisberg, S. (1982), *Residuals and Influence in Regression*, New York: Chapman and Hall.
- [2] Groß, J. (2003), *Linear Regression*, Berlin: Springer.
- [3] Hoerl, A. E. and Kennard, R. W. (1970), Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, **12**, 55-67.
- [4] Nurunnabi, A. A. M., Imon, A. H. M. R. and Nasser, M. (2011), A diagnostic measure for influential observations in linear regression, *Communications in Statistics — Theory and Methods*, **40**, 1169-1183.
- [5] Takeuchi, H. (1991), Detecting influential observations by using a new expression of Cook's distance, *Communications in Statistics — Theory and Methods*, **20**, 261-274.
- [6] Takeuchi, H. (1994), Sensitivity analysis with an extension of Cook's distance in ridge regression, *Journal of the Japan Statistical Society*, **24**, 221-236.
- [7] 竹内秀一 (2005a), 線形回帰分析における尤度距離による影響力評価, 人文自然科学論集, **119** 号, 19-30.
- [8] 竹内秀一 (2005b), 線形回帰分析における部分影響力評価, 人文自然科学論集, **120** 号, 53-64.
- [9] 竹内秀一 (2007), リッジ回帰における尤度距離による影響力評価, 人文自然科学論集, **123** 号, 3-16.
- [10] 竹内秀一 (2011), 数値実験に基づくリッジパラメータの推定量の比較, 人文自然科学論集, **130** 号, 71-93.
- [11] Walker, E. and Birch, J. B. (1988), Influence measures in ridge regression, *Technometrics*, **30**, 221-227 (Correction **30**, 469-470).