線形回帰モデル（多重共線性，Ridge 回帰）

1　多重共線性

次の線形回帰モデルを考える。

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

ここで，$p < n$ として，

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{pn} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

とする。また，$\boldsymbol{\varepsilon}$ に関して，

条件 1: $E[\boldsymbol{\varepsilon}] = \boldsymbol{0}$
条件 2: $V[\boldsymbol{\varepsilon}] = \sigma^2 \boldsymbol{I}_n$

を仮定する。ただし，$\boldsymbol{I}_n$ は $n \times n$ 単位行列である。このとき，$\boldsymbol{Y}$ を目的変数，$\boldsymbol{X}$ を説明変数とし，$\boldsymbol{\beta}$ の推測問題を考える。$\boldsymbol{\beta}$ の最小二乗推定量 $\hat{\boldsymbol{\beta}}$ は，次式で与えられる。

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}$$

(1) 式の線形回帰モデルの $n \times (p+1)$ 行列 $\boldsymbol{X}$ について，$(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$ が正則であるために，$\mathrm{rank}(\boldsymbol{X}) = \mathrm{rank}((\boldsymbol{X}^\top \boldsymbol{X})^{-1}) = p+1$ であることが仮定されていた。この仮定を満たすためには，観測値の個数 n が変数の数 p よりも大きいことに加え，$\boldsymbol{X}$ の $p+1$ 個の列が 1 次独立であることが必要になる。この仮定が満たされない場合，$\mathrm{rank}((\boldsymbol{X}^\top \boldsymbol{X})^{-1}) < p+1$ となり，その逆行列は存在しないことになる。つまり，最小二乗推定量は定義されない。一般化逆行列を用いて，$\mathrm{rank}((\boldsymbol{X}^\top \boldsymbol{X})^{-1}) < p+1$ の場合でも最小二乗推定量の定義を拡張することは可能であるが，便宜的に拡張された最小二乗推定量の意味づけは，必ずしも明確ではない。

行列 $\boldsymbol{X} = (\boldsymbol{1}, \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_p)$ の階数が $p+1$ 未満であるということは，

$$a_0 \boldsymbol{1} + a_1 \boldsymbol{x}_1 + a_2 \boldsymbol{x}_2 + \cdots + a_p \boldsymbol{x}_p = \boldsymbol{0}$$

を満たす，0 ではない定数 $a_0, a_1, a_2, \dots, a_p$ が存在することと同値である。つまり，$\boldsymbol{1}, \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_p$ は一次従属であることと同値である。したがって，p 個の変数に**共線関係**が存在することを意味する。

簡単のために，$\boldsymbol{x}_1 = a\boldsymbol{x}_2$ という関係が成り立っていると仮定して，次の線形回帰モデルを考える。

$$\boldsymbol{Y} = \beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \boldsymbol{\varepsilon} \tag{2}$$

(2) 式の線形回帰モデルは，$\boldsymbol{x}_1 = a\boldsymbol{x}_2$ の仮定のもとで，次の線形回帰モデルと同等であり，与えられた観測値に基づいて，一方を他方から識別することは不可能である。

$$\boldsymbol{Y} = (\beta_1 a + \beta_2)\boldsymbol{x}_2 + \boldsymbol{\varepsilon}$$

$\boldsymbol{x}_1$ と $\boldsymbol{x}_2$ の各々が，$\boldsymbol{Y}$ の変動をどの程度説明できるかに関心があったとしても，$\boldsymbol{x}_1 = a\boldsymbol{x}_2$ の共線関係のために，どうしようもない。

一般には，説明変数間に厳密な共線関係が存在することはほぼないと考えられる。実際には，共線関係が近似的に生じる場合である。つまり，すべてが 0 ではない定数 $a_0, a_1, a_2, \ldots, a_p$ に対して，次式となる場合である。

$$a_0\mathbf{1} + a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_p\boldsymbol{x}_p \fallingdotseq \mathbf{0}$$

このような場合，$(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$ が正則であるために，$\boldsymbol{\beta}$ を最小二乗推定することは可能である。しかし，推定上の困難が生じることになる。(2) 式の線形回帰モデルのもとで，この問題について以下に述べていく。

簡単のために，$\boldsymbol{X}^\top\boldsymbol{X}$ と $\boldsymbol{X}^\top\boldsymbol{Y}$ を次のようにする。

$$\boldsymbol{X}^\top\boldsymbol{X} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}, \quad \boldsymbol{X}^\top\boldsymbol{Y} = \begin{pmatrix} m_{1y} \\ m_{2y} \end{pmatrix}$$

(2) 式の線形回帰モデルの β_1 と β_2 の最小二乗推定量は，次の正規方程式の解によって与えられる。

$$\begin{aligned} m_{11}\hat{\beta}_1 + m_{12}\hat{\beta}_2 &= m_{1y} \\ m_{12}\hat{\beta}_1 + m_{22}\hat{\beta}_2 &= m_{2y} \end{aligned}$$

この連立方程式の解を求めると，次式を得る。

$$\hat{\beta}_1 = \frac{m_{22}m_{1y} - m_{12}m_{2y}}{m_{11}m_{22} - m_{12}^2}, \quad \hat{\beta}_2 = \frac{m_{11}m_{2y} - m_{12}m_{1y}}{m_{11}m_{22} - m_{12}^2}$$

$\boldsymbol{x}_1$ と $\boldsymbol{x}_2$ の間に近似的な共線関係 $\boldsymbol{x}_1 \fallingdotseq a\boldsymbol{x}_2$ が成り立っている場合，$\boldsymbol{X}^\top\boldsymbol{X}$ の行列式は近似的に 0 になる。つまり，

$$m_{11}m_{22} - m_{12}m_{21} = m_{11}m_{22} - m_{12}^2 \fallingdotseq 0$$

となるため，最小二乗推定値は求まるが，分母が 0 に近いため，その値は不安定になることが考えられる。実際，分母が 0 に近いということは，観測誤差や丸めの誤差のために，分子の値がごくわずか変化したとき，推定値が大幅に変動する可能性を示唆している。また，$\hat{\beta}_1$ と $\hat{\beta}_2$ の分散共分散行列は次式となる。

$$V[\hat{\boldsymbol{\beta}}] = \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1} = \frac{\sigma^2}{m_{11}m_{22} - m_{12}^2}\begin{pmatrix} m_{22} & -m_{12} \\ -m_{12} & m_{11} \end{pmatrix}$$

したがって，σ^2 が小さいとしても，$\boldsymbol{x}_1$ と $\boldsymbol{x}_2$ の間に近似的な共線関係 $\boldsymbol{x}_1 \fallingdotseq a\boldsymbol{x}_2$ が成り立っていれば，推定量の分散・共分散は著しく大きくなることがわかる。$\boldsymbol{x}_1$ と $\boldsymbol{x}_2$ が強い正の相関をもっている $(a > 0)$ ならば，$\hat{\beta}_1$ と $\hat{\beta}_2$ の間に強い負の相関が生じる。つまり，$\boldsymbol{x}_1$ の係数 $\hat{\beta}_1$ を過大（または過少）推定したとすれば，$\boldsymbol{x}_2$ の係数 $\hat{\beta}_2$ を過少（または過大）推定する傾向があることを示している。

2　多重共線への対処

多重共線に対処するための一つの可能な推定上の工夫として，後に述べる Ridge 回帰がある。通常よく用いられるのは，多重共線関係にある変数群の一部を除去する方法である。仮に，$\boldsymbol{x}_1$ と $\boldsymbol{x}_2$

が強度の多重共線関係にあるとする。つまり，$\boldsymbol{x}_1$ と $\boldsymbol{x}_2$ の間に近似的な共線関係 $\boldsymbol{x}_1 \fallingdotseq a\boldsymbol{x}_2$ が成り立っているとする。$\boldsymbol{x}_1$ が既に線形回帰モデルに含まれているとき，$\boldsymbol{x}_2$ を追加することによりもたらされる $\boldsymbol{Y}$ への説明力の増加は大きくないと考えられる。回帰分析の目的が予測であり，$\boldsymbol{x}_1$ と $\boldsymbol{x}_2$ の多重共線が構造的なものであれば，$\boldsymbol{x}_2$ を除去することにより，精度高く予測できることが見込まれる。多重共線の問題は，**共線関係にある変数の一部を除去する**ことにより処理されている，というのが現状である。

3　多重共線の数値例

多重共線性が線形回帰モデルの推定にどのような影響をもたらすのかを例示するために，表 1 のデータ (E. Hilleboe. *New York State Journal of Medicine*, **57**, 1957, 2243–2254) 用いる。$\boldsymbol{x}_1$ は，総カロリーに占める脂肪のカロリーの比率，$\boldsymbol{x}_2$ は総カロリーに占める動物性蛋白のカロリーの比率，$\boldsymbol{Y}$ は $100 \times [\log(55$ 歳から 59 歳までの男子 10 万人当たりの心臓疾患による死亡者数$) - 2]$ である。

表 1　心臓疾患による死亡率とカロリー摂取

国名	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{Y}$
オーストラリア	33	8	81
オーストリア	31	6	55
セイロン	17	2	24
デンマーク	39	6	52
フィンランド	30	7	88
フランス	29	7	45
ドイツ	35	6	50
アイルランド	31	5	69
イスラエル	23	4	66
イタリア	21	3	45
日本	8	3	24
メキシコ	23	3	43
オランダ	37	6	38
ニュージーランド	40	8	72
ノルウェー	38	6	41
ポルトガル	25	4	38
スウェーデン	39	7	52
スイス	33	7	52
イギリス	38	6	66
アメリカ	39	8	89
カナダ	38	8	80

心臓病の発症率が動物性蛋白や脂肪の摂取と関係あることは，医学的に良く知られた事実であ

る。つまり，動物性蛋白や脂肪の摂取が多い人ほど，心臓病の発症率が高くなることが予想される。この関係を検証するために，(3) 式の線形回帰モデルにより，表 1 のデータを回帰分析してみる。

$$\boldsymbol{Y} = \beta_0 + \beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \boldsymbol{\varepsilon} \tag{3}$$

(3) 式の線形回帰モデルにより，線形回帰分析をすると表 2 の結果が得られた。

表 2　表 1 に対する線形回帰分析の結果

変数	推定値	標準誤差	t 値
$\boldsymbol{x}_1$	-0.2230	0.6563	-0.340
$\boldsymbol{x}_2$	8.0437	3.0046	2.677
定数項	16.6217	11.9115	1.395

動物性蛋白や脂肪の摂取が多い人ほど心臓病の発症率が高くなる，つまり，$\boldsymbol{x}_1$ と $\boldsymbol{x}_2$ の回帰係数 β_1 と β_2 は正となることが予想される。しかし，予想に反して，表 2 の結果では，β_1 の推定値は負となった。この推定結果は，脂肪の摂取が多いほど心臓病による死亡率が低い，あるいは脂肪の摂取量と心臓病による死亡率との間には有意な関係が認められない，という医学的に良く知られた事実とは異なる結論を示唆している。表 2 の結果の妥当性を検討するために，より詳しくデータを調べてみる。

3 変数の相関係数を表 3 にまとめた。表 3 からわかるように，$\boldsymbol{x}_1$ と $\boldsymbol{Y}$ には，正の相関があることがわかる。つまり，脂肪の摂取が多い人ほど，心臓病の発症率が高くなることが予想されるという，医学的に良く知られた事実と一致する結果となっている。$\boldsymbol{x}_1$ と $\boldsymbol{x}_2$ の間に強い正の相関があることがわかる。一般に，動物性蛋白の含有量の高い食品は，脂肪も多く含む傾向がある。したがって，$\boldsymbol{x}_1$ と $\boldsymbol{x}_2$ の間に強い正の相関があることは妥当な結果である。表 2 の結果のように，脂肪の摂取が多いほど心臓病による死亡率が低い，あるいは脂肪の摂取量と心臓病による死亡率との間には有意な関係が認められない，という医学的に良く知られた事実とは異なる結果が得られたのは，$\boldsymbol{x}_1$ と $\boldsymbol{x}_2$ の共線関係によるものである可能性が高いと考えられる。もっとはっきりとした結論を得るためには，観測値の個数を増やす必要がある。

表 3　相関行列

	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{Y}$
$\boldsymbol{x}_1$	1.0	0.823	0.547
$\boldsymbol{x}_2$		1.0	0.704
$\boldsymbol{Y}$			1.0

4　Ridge 回帰

説明変数間に多重共線が存在するとき，一つの有用な対応策として，最小二乗推定法に次のような改良をする方法がある。近似的な共線関係が存在するとき，推定が不安定になるそもそもの原因は，$\boldsymbol{X}^\top \boldsymbol{X}$ の行列式が近似的に 0 になることであった。そこで，$\boldsymbol{X}^\top \boldsymbol{X}$ の対角成分に正定数 k を

加えて次のような推定量を考える。

$$\hat{\boldsymbol{\beta}}_k = (\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1}\boldsymbol{X}^\top \boldsymbol{Y}$$

この推定量を **Ridge 推定量**という。Ridge 推定量は，$\boldsymbol{\beta}$ の不偏推定量ではない。

$$\begin{aligned}
E[\hat{\boldsymbol{\beta}}_k] &= E[(\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1}\boldsymbol{X}^\top \boldsymbol{Y}] \\
&= E[(\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1}\boldsymbol{X}^\top (\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})] \\
&= (\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} \\
&\neq \boldsymbol{\beta}
\end{aligned}$$

しかし，$(\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1}$ は $(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$ よりも，推定量の分散共分散行列が小さくなるという意味で，推定が安定すると予想される。このことは，次式から容易に確かめることが可能である。

$$\begin{aligned}
V[\hat{\boldsymbol{\beta}}_k] &= V[(\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1}\boldsymbol{X}^\top \boldsymbol{Y}] \\
&= V[(\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1}\boldsymbol{X}^\top (\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})] \\
&= V[(\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1}\boldsymbol{X}^\top \boldsymbol{\varepsilon}] \\
&= (\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1}\boldsymbol{X}^\top V[\boldsymbol{\varepsilon}]((\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1}\boldsymbol{X}^\top)^\top \\
&= (\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1}\boldsymbol{X}^\top (\sigma^2 \boldsymbol{I}_n)\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1} \\
&= \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1}\boldsymbol{X}^\top \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1} \\
&\leq \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \quad (\because \text{補足 3 を参照}) \\
&= V[\hat{\boldsymbol{\beta}}]
\end{aligned}$$

補足 3

任意の列ベクトル $\boldsymbol{a}$ に対して次式が成り立つ。

$$\begin{aligned}
&\boldsymbol{a}^\top \left[(\boldsymbol{X}^\top \boldsymbol{X})^{-1} - (\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1}\boldsymbol{X}^\top \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1}\right] \boldsymbol{a} \\
&= \left[\boldsymbol{a}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} - \boldsymbol{a}^\top (\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1}\boldsymbol{X}^\top \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1}\right] \boldsymbol{a} \\
&= \boldsymbol{a}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{a} - \boldsymbol{a}^\top (\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1}\boldsymbol{X}^\top \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1}\boldsymbol{a} \\
&\geq 0
\end{aligned}$$

したがって，対称行列 $(\boldsymbol{X}^\top \boldsymbol{X})^{-1} - (\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1}\boldsymbol{X}^\top \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_{p+1})^{-1}$ は非負定値行列である。
以上の結果から，任意の列ベクトル $\boldsymbol{a}$ に対して次式が成り立つ。

$$\begin{aligned}
\boldsymbol{a}^\top \left(V[\hat{\boldsymbol{\beta}}] - V[\hat{\boldsymbol{\beta}}_k]\right) \boldsymbol{a} \geq 0 &\Leftrightarrow \boldsymbol{a}^\top V[\hat{\boldsymbol{\beta}}]\boldsymbol{a} \geq \boldsymbol{a}^\top V[\hat{\boldsymbol{\beta}}_k]\boldsymbol{a} \\
&\Leftrightarrow V[\boldsymbol{a}^\top \hat{\boldsymbol{\beta}}] \geq V[\boldsymbol{a}^\top \hat{\boldsymbol{\beta}}_k]
\end{aligned}$$

列ベクトル $\boldsymbol{a}$ を単位ベクトルとすれば，次式を得る。

$$V[\hat{\boldsymbol{\beta}}] \geq V[\hat{\boldsymbol{\beta}}_k]$$

したがって，k をうまく選択することで，Ridge 推定量を用いることで，推定値や予測値の平均二乗誤差を小さくすることができる。k の選択方法は，Mallows の C_p 規準に代表される情報量規

準の最小化により求められることが多い。それらの情報量規準は，予測の平均二乗誤差に基づくリスク関数の推定量であり，それを最小にする k を選ぶことで，あてはめ値の予測平均二乗誤差が小さくなることが期待できる。しかし，情報量規準を最小にする解は陽な形で求めることができないため，実際の最適化には計算機による繰り返し計算が必要になる。

参考文献

[1] 佐和隆光．(2020)．回帰分析（新装版）．朝倉書店．

[2] 柳原宏和，永井勇，佐藤健一．(2009)．多変量一般化リッジ回帰におけるリッジパラメータ最適化のためのバイアス補正 C_p 規準．応用統計学，**38**，151–172．