計量経済学のための線形代数 攻略マニュアル

ゼロから $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ が読めるようになるまで

目次

序 なぜ計量経済学に線形代数が必要なのか **4**

序.1 回帰分析と「たくさんの式」 4

序.2 本書の目標と構成 5

序.3 最も大切なコツ：サイズを書き出す 5

1 スカラーとベクトル **7**

1.1 スカラー 7

1.2 ベクトル 7

1.3 ベクトルの計算 8

1.4 ベクトルを矢印として見る 9

2 内積・長さ・直交 **11**

2.1 和の記号 $\sum$ 11

2.2 内積 11

2.3 ベクトルの長さ（ノルム） 12

2.4 直交 13

3 行列とデータ行列 **15**

3.1 行列 15

3.2 データ行列 X：表計算ソフトの表 15

3.3 特別な行列 16

3.4 転置 17

4 行列の演算：積が最も重要 18

4.1 和とスカラー倍 18

4.2 行列の積 18

4.3 行列 × ベクトル：2つの見方 19

4.4 $Y = X\beta + u$：N 本の式を1本で書く 20

4.5 行列の積の性質 20

5 「横×縦」と「縦×横」、和の記号と行列 22

5.1 同じベクトルでも、掛ける順番で全く別物になる 22

5.2 回帰分析の $x_i^\top x_i$ 23

5.3 期待値をとる 23

5.4 和と行列の積：$X^\top X = \sum_i x_i^\top x_i$ 23

5.5 $X^\top X$ の中身：2乗和と積和 25

5.6 母集団の式と標本の式 26

6 逆行列・行列式・連立方程式 27

6.1 数の世界とのたとえ 27

6.2 逆行列 27

6.3 行列式 29

6.4 連立方程式を逆行列で解く 30

6.5 OLS の正規方程式 30

7 一次独立とランク：多重共線性 32

7.1 一次独立と一次従属 32

7.2 ランク 33

7.3 回帰分析での意味：完全多重共線性 33

8 射影行列：OLS の幾何学 35

8.1 対称行列とべき等行列 35

8.2 予測値と残差を行列で書く 35

8.3 P と M の性質 36

8.4 残差は説明変数と直交する 36
8.5 図で見る OLS 37
8.6 分散の分解と決定係数 38

9 二次形式と正定値行列 40
9.1 二次形式 40
9.2 正定値と半正定値 41
9.3 $X^\top X$ はいつも半正定値 41
9.4 残差平方和の「お椀の形」 42

10 ベクトルによる微分と OLS の導出 44
10.1 ベクトルによる微分の定義 44
10.2 2つの基本公式 44
10.3 OLS 推定量の導出 45

11 確率ベクトルと分散共分散行列 48
11.1 確率ベクトルと期待値 48
11.2 分散共分散行列 48
11.3 線形変換の公式 49
11.4 OLS 推定量の期待値：不偏性 50
11.5 OLS 推定量の分散 50

12 発展：トレース・固有値・2段階最小二乗法 52
12.1 トレース 52
12.2 固有値と固有ベクトル 53
12.3 2段階最小二乗法（操作変数法） 54

終 公式一覧と練習問題 55
終.1 よく使う公式一覧 55
終.2 練習問題 55

序章

なぜ計量経済学に線形代数が必要なのか

> **この章の問い**　回帰分析の教科書に出てくる $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ のような式は、何を表しているのか。なぜこんな書き方をするのか。
> **この章で分かること**　線形代数が「たくさんの数・たくさんの式を1つの記号にまとめて書き、まとめて計算するための道具」であること、本書の構成、そして行列の式を読むための最も大切なコツ（**サイズを書き出す**）です。

序.1　回帰分析と「たくさんの式」

賃金 y を、教育年数 x_2 と経験年数 x_3 で説明する回帰モデルを考えます。

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i, \qquad i = 1, 2, \ldots, N$$

i は人の番号、N は人数（標本の大きさ）、u_i は誤差項です。この1本の式は、実は N 人分の式をまとめたものです。

$$\begin{aligned} y_1 &= \beta_1 + \beta_2 x_{21} + \beta_3 x_{31} + u_1 \\ y_2 &= \beta_1 + \beta_2 x_{22} + \beta_3 x_{32} + u_2 \\ &\vdots \\ y_N &= \beta_1 + \beta_2 x_{2N} + \beta_3 x_{3N} + u_N \end{aligned}$$

また、推定したい係数は $\beta_1, \beta_2, \beta_3$ の3つで、それらを求めるには3本の連立方程式を解く必要があります。説明変数が10個なら、係数は11個、連立方程式も11本です。

このように、回帰分析では「たくさんの数」と「たくさんの式」を扱います。それを1本ずつ書いていたら大変です。**線形代数**（ベクトルと行列）は、たくさんの数を1つの記号にまとめ、たくさんの式を1本の式として書き、まとめて計算するための道具です。例えば上の N 本の式は、本書を読み終える頃には

$$Y = X\beta + u$$

という1本の式として読めるようになります。そして最小二乗法（OLS）の推定量は

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

と1行で書けます。

序.2 本書の目標と構成

本書の目標は、**線形代数をまったく知らない状態から、学部中級〜大学院入門の計量経済学（主に回帰分析）に出てくる行列の式を、記号の1つ1つまで意味が分かるように読めるようになる**ことです。証明よりも「この記号は何を表し、この式は何を計算しているのか」を重視します。

章	内容
第1〜2章	ベクトルとその計算、内積（掛けて足す）、長さと直交
第3〜4章	行列、データ行列 X、行列の積、$Y = X\beta + u$
第5章	「横×縦」と「縦×横」、和の記号 $\sum$ と行列の関係（$X^\top X = \sum x_i^\top x_i$）
第6〜7章	逆行列と行列式、連立方程式、一次独立とランク（多重共線性）
第8章	対称行列・べき等行列・射影行列（OLS の幾何学）
第9章	二次形式と正定値行列
第10章	ベクトルによる微分と OLS の導出
第11章	確率ベクトル、分散共分散行列、OLS の期待値と分散
第12章	発展：トレース、固有値、2段階最小二乗法
終章	公式一覧と練習問題

序.3 最も大切なコツ：サイズを書き出す

行列の式を見たら、まず**各記号のサイズ（何行何列か）を書き出してください**。例えば $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ なら、

$$\underbrace{\hat{\beta}}_{K\times 1} = \Big(\underbrace{X^\top}_{K\times N}\,\underbrace{X}_{N\times K}\Big)^{-1}\,\underbrace{X^\top}_{K\times N}\,\underbrace{Y}_{N\times 1}$$

のように書きます。サイズが分かれば、式がどんな計算をしているのか、計算が正しくできるのかが、ほぼ分かります。本書でも、新しい式には必ずサイズを併記します。

要点

- 線形代数は、たくさんの数や式をまとめて書き、まとめて計算するための道具である。
- 回帰分析では、N 人分の式が $Y = X\beta + u$ の1本に、K 本の連立方程式の解が $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ の1行にまとまる。
- 行列の式を読むときは、まず各記号のサイズを書き出す。

次の章へ　まずは線形代数の最も基本的な部品である「ベクトル」から始めます。

第1章

スカラーとベクトル

> **この章の問い**　「ベクトル」とは何で、どう書き、どう計算するのか。回帰分析の何がベクトルなのか。
> **この章で分かること**　スカラーとベクトルの区別、列ベクトルと行ベクトル、ベクトルの足し算とスカラー倍、そしてベクトルを「矢印」として図で見る方法です。

1.1　スカラー

ただの1つの数を**スカラー**と呼びます。例えば、ある人の賃金 $y_1 = 1500$（万円）、係数 $\beta_2 = 0.08$ などです。私たちが中学・高校で扱ってきた数は、全てスカラーです。

1.2　ベクトル

いくつかの数を1列（または1行）に並べたものを**ベクトル**と呼びます。並べる数の個数を**次元**と呼び、並べられた1つ1つの数を**成分**（要素）と呼びます。

1.2.1　列ベクトルと行ベクトル

数を縦に並べたものを**列ベクトル**、横に並べたものを**行ベクトル**と呼びます。

$$a = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \quad (3 \times 1), \qquad b = \begin{pmatrix} b_1 & b_2 & b_3 \end{pmatrix} \quad (1 \times 3)$$

括弧の後ろの (3×1) は「3行1列」という**サイズ**を表します。サイズは必ず「行の数 $\times$ 列の数」の順に書きます。

数式の読み方：回帰分析のベクトル

- **被説明変数のベクトル**：N 人の賃金を縦に並べたもの。

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \quad (N \times 1)$$

- **係数のベクトル**：推定したい K 個の係数を縦に並べたもの。$\beta = (\beta_1, \beta_2, \ldots, \beta_K)^\top \quad (K \times 1)$。
- **誤差項のベクトル**：$u = (u_1, \ldots, u_N)^\top \quad (N \times 1)$。
- **1人分の説明変数**：i さんの説明変数（定数項の1も含む）を横に並べたもの。

$$x_i = \begin{pmatrix} 1 & x_{2i} & x_{3i} & \cdots & x_{Ki} \end{pmatrix} \quad (1 \times K)$$

右肩の $^\top$ は**転置**（縦と横を入れ替える操作、第3章）を表します。$(\beta_1, \ldots, \beta_K)^\top$ は「横に書いたものを縦にしたもの」、つまり列ベクトルです。紙面を節約するために、列ベクトルをこのように書くことがよくあります。

補足：x_i を行で書くか列で書くか

本書では、1人分の説明変数 x_i を**行ベクトル**（$1 \times K$）とします。こうすると、x_i を上から N 人分積み重ねたものがそのままデータの表（第3章のデータ行列 X）になるので分かりやすいからです。ただし、教科書によっては x_i を**列ベクトル**（$K \times 1$）とするものもあります。その場合、本書の $x_i^\top x_i$ は $x_i x_i^\top$ と書かれます。教科書を読むときは、最初に x_i がどちらで定義されているかを確認してください。

1.3 ベクトルの計算

1.3.1 足し算

同じサイズのベクトルどうしは、**成分ごとに**足します。

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \end{pmatrix}$$

サイズが違うベクトルどうしは足せません。

1.3.2　スカラー倍

ベクトルにスカラー c を掛けると、**全ての成分が c 倍**になります。

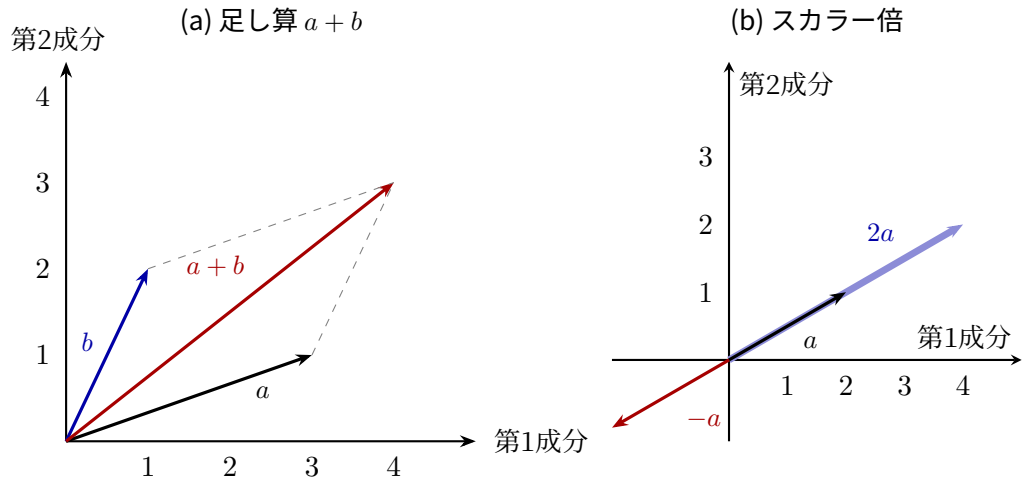
$$c\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} c\,a_1 \\ c\,a_2 \end{pmatrix}$$

例えば、全員の賃金を円からドルに換算する（全ての成分を同じ数で割る）のはスカラー倍です。

1.4　ベクトルを矢印として見る

2次元のベクトル $a = (a_1, a_2)^\top$ は、平面上の原点から点 (a_1, a_2) に向かう**矢印**として描けます。この見方は、第2章の内積や第8章の OLS の幾何学的な理解で役に立ちます。

図 1.1　ベクトルの足し算とスカラー倍



(a) $a = (3,1)^\top$、$b = (1,2)^\top$、$a+b = (4,3)^\top$。(b) $a = (2,1)^\top$。

図の読み方

1. **(a)**：$a+b$ は、a の矢印の先に b の矢印をつないだ先（または b の先に a をつないだ先）を指します。a と b を2辺とする平行四辺形の対角線になります。
2. **(b)**：$2a$ は a と同じ向きで長さが2倍の矢印、$-a$ は逆向きの矢印です。スカラー倍は、矢印の向きを変えずに（負なら逆向きにして）長さを伸び縮みさせる操作です。
3. 3次元以上のベクトルは図に描けませんが、同じように「多次元の空間の矢印」と考えることができます。N 人分のデータ Y は、N 次元空間の1本の矢印です。

要点

- スカラーは1つの数、ベクトルは数を縦（列ベクトル）または横（行ベクトル）に並べたもの。サイズは「行数 $\times$ 列数」で書く。
- ベクトルの足し算は成分ごと、スカラー倍は全成分を同じ数倍する。
- 回帰分析では Y（$N \times 1$）、β（$K \times 1$）、u（$N \times 1$）が列ベクトル、1人分の説明変数 x_i（$1 \times K$）が行ベクトル。

次の章へ　ベクトルどうしの「掛け算」の中で、回帰分析で最も基本になるのが内積です。第2章で学びます。

第2章

内積・長さ・直交

> **ここまでの流れ**　第1章で、ベクトルとその足し算・スカラー倍を学びました。
> **この章の問い**　2つのベクトルを「掛ける」とはどういうことか。それは回帰分析の何に対応するのか。
> **この章で分かること**　**内積**（成分どうしを掛けて足す）、**和の記号** $\sum$、ベクトルの**長さ**（ノルム）、そして2つのベクトルが**直交**するという性質です。直交は、OLSの残差の性質（第8章）を理解する鍵になります。

2.1　和の記号 $\sum$

線形代数に入る前に、和の記号を確認しておきます。

$$\sum_{i=1}^{N} a_i = a_1 + a_2 + \cdots + a_N$$

「i を1から N まで動かして、a_i を全部足す」という意味です。i は何番目かを表す番号（添字）で、別の文字（j や k）を使っても意味は同じです。例えば平均は $\bar{a} = \frac{1}{N}\sum_{i=1}^{N} a_i$ と書けます。

数式の読み方：$\sum$ の計算ルール

- 定数は外に出せる：$\sum_i c\,a_i = c\sum_i a_i$。
- 和は分けられる：$\sum_i (a_i + b_i) = \sum_i a_i + \sum_i b_i$。
- 定数を N 回足すと N 倍：$\sum_{i=1}^{N} c = Nc$。
- ただし、積の和は和の積ではない：$\sum_i a_i b_i \neq (\sum_i a_i)(\sum_i b_i)$。

2.2　内積

同じ次元の2つのベクトル a、b について、**対応する成分どうしを掛けて、全部足したもの**を**内積**と呼びます。

$$a \cdot b = a_1 b_1 + a_2 b_2 + \cdots + a_N b_N = \sum_{i=1}^{N} a_i b_i \tag{2.1}$$

結果は1つの数（スカラー）です。

行列の記法では、内積は「**行ベクトル × 列ベクトル**」として書きます。

$$a^\top b = \begin{pmatrix} a_1 & a_2 & \cdots & a_N \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{pmatrix} = a_1 b_1 + a_2 b_2 + \cdots + a_N b_N$$

$$\underbrace{a^\top}_{1\times N}\ \underbrace{b}_{N\times 1} = \underbrace{\text{スカラー}}_{1\times 1}$$

左の横長のベクトルの成分と、右の縦長のベクトルの成分を、順番に1つずつ組にして掛けて足す、と覚えてください。

数式の読み方：回帰分析の中の内積

- **1人分の予測値**：$x_i\beta = \beta_1 \cdot 1 + \beta_2 x_{2i} + \cdots + \beta_K x_{Ki}$。$x_i$（$1 \times K$）と β（$K \times 1$）の内積です。回帰式 $y_i = x_i\beta + u_i$ は、この内積を使って書かれています。
- **合計**：全ての成分が1のベクトル $\iota = (1, 1, \ldots, 1)^\top$（$N \times 1$）を使うと、$\iota^\top Y = \sum_i y_i$。平均は $\bar{y} = \frac{1}{N}\iota^\top Y$ と書けます。
- **2乗和**：$Y^\top Y = \sum_i y_i^2$。自分自身との内積は、成分の2乗の和です。
- **積和**：$a^\top b = \sum_i a_i b_i$ は、共分散の計算に出てくる「積の和」です。

2.3 ベクトルの長さ（ノルム）

2次元のベクトル $a = (a_1, a_2)^\top$ の矢印の長さは、三平方の定理から $\sqrt{a_1^2 + a_2^2}$ です。一般の次元でも同じように、ベクトルの**長さ（ノルム）**を

$$\|a\| = \sqrt{a_1^2 + a_2^2 + \cdots + a_N^2} = \sqrt{a^\top a} \tag{2.2}$$

と定義します。自分自身との内積の平方根です。

数式の読み方：残差平方和はベクトルの長さの2乗

OLS の残差 $\hat{u}_i = y_i - x_i\hat{\beta}$ を並べたベクトルを $\hat{u}$（$N \times 1$）とすると、

$$\sum_{i=1}^{N} \hat{u}_i^2 = \hat{u}^\top \hat{u} = \|\hat{u}\|^2.$$

つまり**残差平方和は、残差ベクトルの長さの2乗**です。OLS は「残差ベクトルを最も短くする $\hat{\beta}$ を選ぶ」方法だと言い換えられます。この見方は第8章で重要になります。

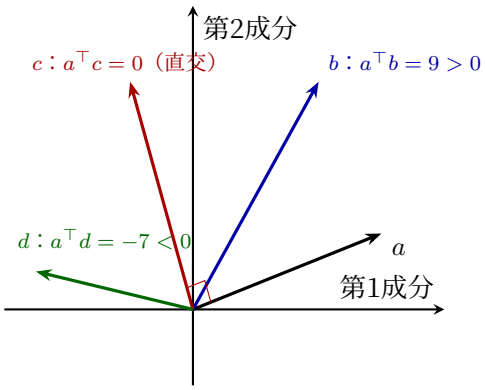
2.4 直交

内積がゼロになる2つのベクトルは**直交**すると言います。

$$a^\top b = 0 \quad \Longleftrightarrow \quad a \perp b$$

2次元・3次元では、矢印どうしが**垂直**に交わることを意味します。

図 2.1　内積の符号と直交



$a = (3,1)^\top$、$b = (2,3)^\top$、$c = (-1,3)^\top$、$d = (-2.5, 0.5)^\top$。c と a の間の小さな直角の印は、2本が垂直であることを示す。

図の読み方

1. a と b のように同じような方向を向いたベクトルの内積は正です。
2. a と c のように垂直なベクトルの内積はゼロです：$3 \times (-1) + 1 \times 3 = 0$。
3. a と d のように反対寄りの方向を向いたベクトルの内積は負です。
4. 一般に、2つのベクトルのなす角を θ とすると $a^\top b = \|a\|\,\|b\| \cos\theta$ が成り立ちます。$\cos\theta$ は、角度が90度より小さければ正、90度ならゼロ、大きければ負です。

発展：相関係数は2つのベクトルのなす角

2つの変数 a、b から平均を引いたベクトルを $\tilde{a} = (a_1 - \bar{a}, \ldots, a_N - \bar{a})^\top$、$\tilde{b}$ とすると、標本相関係数は

$$r_{ab} = \frac{\sum_i (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_i (a_i - \bar{a})^2}\sqrt{\sum_i (b_i - \bar{b})^2}} = \frac{\tilde{a}^\top \tilde{b}}{\|\tilde{a}\|\,\|\tilde{b}\|} = \cos\theta$$

と書けます。つまり**相関係数は、平均を引いた2つのデータのベクトルがなす角のコサイン**です。2つのベクトルが同じ向きなら相関は1、垂直なら0（無相関）、逆向きなら -1 です。「無相関」と「直交」が同じ考え方であることが分かります。

要点

- 内積 $a^\top b = \sum_i a_i b_i$ は「成分どうしを掛けて足す」計算で、結果はスカラー。行ベクトル（$1 \times N$）$\times$ 列ベクトル（$N \times 1$）として書く。
- ベクトルの長さは $\|a\| = \sqrt{a^\top a}$。残差平方和は残差ベクトルの長さの2乗。
- 内積がゼロなら直交（垂直）。相関係数は、平均を引いたベクトルのなす角のコサイン。

次の章へ　ベクトルを横にいくつも並べると「行列」になります。第3章では、行列と、データの表そのものである「データ行列 X」を学びます。

第3章

行列とデータ行列

> **ここまでの流れ**　第1・2章で、ベクトルと内積を学びました。
> **この章の問い**　「行列」とは何か。回帰分析のデータは、どのように行列として表されるのか。
> **この章で分かること**　行列の定義と記法、回帰分析の**データ行列** X（表計算ソフトの表そのもの）、よく出てくる特別な行列、そして縦と横を入れ替える**転置**です。

3.1　行列

数を長方形に並べたものを**行列**と呼びます。横の並びを**行**、縦の並びを**列**と呼び、m 行 n 列の行列を「$m \times n$ 行列」と言います。

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \quad (2 \times 3)$$

a_{ij} は「i 行目・j 列目」の成分を表します（**行が先、列が後**）。a_{23} は2行目・3列目の成分です。行列を $A = (a_{ij})$ と書くこともあります。

ベクトルは行列の特別な場合と考えられます。列ベクトルは $n \times 1$ 行列、行ベクトルは $1 \times n$ 行列、スカラーは 1×1 行列です。

3.2　データ行列 X：表計算ソフトの表

回帰分析のデータを表計算ソフトに入力すると、1行が1人、1列が1つの変数になります。この表から被説明変数の列を除いたものが**データ行列** X です。定数項のために、全ての成分が1の列を先頭に加えます。

$$X = \begin{pmatrix} 1 & x_{21} & x_{31} & \cdots & x_{K1} \\ 1 & x_{22} & x_{32} & \cdots & x_{K2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{2N} & x_{3N} & \cdots & x_{KN} \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \quad (N \times K) \tag{3.1}$$

数式の読み方：データ行列の読み方

- **行**：i 行目は i さんの説明変数を横に並べたもの、つまり第1章の行ベクトル x_i（$1 \times K$）です。X は $x_1, \ldots, x_N$ を上から積み重ねたものです。
- **列**：k 列目は、変数 k の N 人分の値を縦に並べた列ベクトル（$N \times 1$）です。1列目は全て1（定数項）です。
- 添字の注意：x_{ki} は「変数 k の i さんの値」で、行列の位置でいえば i 行目・k 列目です。行列の成分の書き方（行が先）と添字の順番が逆になっていることがあるので、教科書ごとに確認してください。

計算例：本書で繰り返し使う数値例

3人について、定数項と1つの説明変数（例えば教育年数を何らかの単位で測ったもの）、被説明変数のデータがあるとします。

人 i	定数項	x_{2i}	y_i
1	1	2	5
2	1	0	1
3	1	1	4

これを行列とベクトルで書くと

$$X = \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \ (3 \times 2), \qquad Y = \begin{pmatrix} 5 \\ 1 \\ 4 \end{pmatrix} \ (3 \times 1), \qquad x_1 = \begin{pmatrix} 1 & 2 \end{pmatrix} \ (1 \times 2).$$

3.3 特別な行列

- **正方行列**：行の数と列の数が等しい行列（$n \times n$）。逆行列（第6章）は正方行列についてだけ考えます。
- **零行列** O：全ての成分が0の行列。数の0に当たります。
- **対角行列**：正方行列で、左上から右下への対角線（**対角成分**）以外が全て0の行列。
- **単位行列** I_n：対角成分が全て1の対角行列。数の1に当たります（第4章）。

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- **対称行列**：対角線を挟んで対称な正方行列。$a_{ij} = a_{ji}$。例えば分散共分散行列（第11章）や $X^\top X$ は対称行列です。

3.4 転置

行列の行と列を入れ替える操作を**転置**と呼び、$A^\top$（A' と書く教科書もあります）で表します。A の i 行目が $A^\top$ の i 列目になります。

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \ (2 \times 3) \quad \Longrightarrow \quad A^\top = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} \ (3 \times 2)$$

数式の読み方：転置の性質

- サイズが入れ替わる：$m \times n$ 行列の転置は $n \times m$。
- 2回転置すると元に戻る：$(A^\top)^\top = A$。
- 列ベクトルの転置は行ベクトル：Y（$N \times 1$）なら $Y^\top$（$1 \times N$）。
- 対称行列は転置しても変わらない：$A^\top = A$。
- データ行列の転置 $X^\top$（$K \times N$）は、1行が1つの変数、1列が1人になります。数値例では $X^\top = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \end{pmatrix}$。

要点

- 行列は数を長方形に並べたもの。a_{ij} は i 行 j 列の成分で、サイズは「行数 × 列数」。
- データ行列 X（$N \times K$）は、1行が1人（x_i）、1列が1つの変数の表。先頭の列は定数項の1。
- 転置 $A^\top$ は行と列の入れ替え。$(A^\top)^\top = A$、対称行列では $A^\top = A$。

次の章へ　行列の計算の中で最も重要なのが「行列の積」です。第4章で、積の計算方法と、それを使って N 本の回帰式を1本にまとめる方法を学びます。

第4章

行列の演算：積が最も重要

ここまでの流れ　第3章で、行列とデータ行列 X を学びました。
この章の問い　行列どうしはどう計算するのか。とくに「行列の積」はどう計算し、何を意味するのか。
この章で分かること　行列の和とスカラー倍、**行列の積**の計算方法とサイズのルール、積の性質（順番を変えられない、転置の公式）、そして N 本の回帰式を1本で書く $Y = X\beta + u$ です。

4.1　和とスカラー倍

同じサイズの行列どうしの和は、ベクトルと同じく**成分ごと**に足します。スカラー倍は全ての成分を同じ数倍します。

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 6 & 8 \\ 10 & 12 \end{pmatrix}, \qquad 2\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 6 & 8 \end{pmatrix}$$

4.2　行列の積

行列の積 AB は、次のルールで計算します。

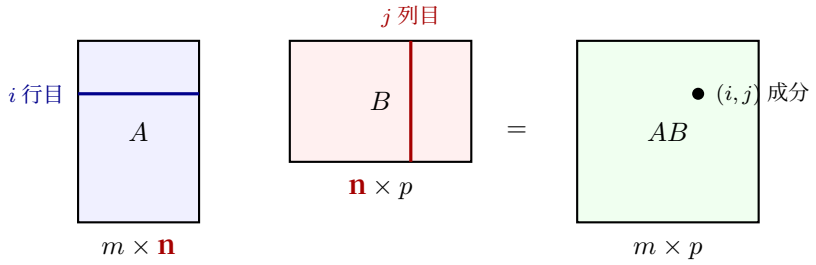
要点：行列の積のルール

AB の (i, j) 成分 $= A$ **の** i **行目と** B **の** j **列目**の内積

$$(AB)_{ij} = \sum_{k} a_{ik}\, b_{kj}$$

内積を計算するには、A の行の長さ（A の列の数）と、B の列の長さ（B の行の数）が等しくなければなりません。

図 4.1　行列の積のサイズのルール



内側のサイズ（赤の n）が一致しているときだけ積が計算でき、結果のサイズは外側の $m \times p$ になる。

図の読み方

1. A（$m \times n$）の i 行目は長さ n の横の並び、B（$n \times p$）の j 列目は長さ n の縦の並びです。この2つの内積が、AB の i 行 j 列の成分になります。
2. **内側**のサイズ（A の列数と B の行数）が一致しないと、内積が計算できず、積は定義されません。
3. 結果のサイズは**外側**のサイズ $m \times p$ です。「$(m \times n)(n \times p) = m \times p$」と、真ん中の n が消えると覚えてください。

計算例：2×2 行列の積

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 1\cdot 5 + 2\cdot 7 & 1\cdot 6 + 2\cdot 8 \\ 3\cdot 5 + 4\cdot 7 & 3\cdot 6 + 4\cdot 8 \end{pmatrix} = \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix}$$

(1,1) 成分は「1行目 $(1,2)$」と「1列目 $(5,7)^\top$」の内積 $1\cdot 5 + 2\cdot 7 = 19$、というように計算します。

4.3　行列 × ベクトル：2つの見方

行列にベクトルを掛ける計算は、回帰分析で最もよく出てきます。X（$N \times K$）と β（$K \times 1$）の積 $X\beta$（$N \times 1$）を、2つの見方で読んでみましょう。

見方1：行ごとの内積。$X\beta$ の i 番目の成分は、X の i 行目 x_i と β の内積です。

$$X\beta = \begin{pmatrix} x_1\beta \\ x_2\beta \\ \vdots \\ x_N\beta \end{pmatrix} = \begin{pmatrix} \beta_1 + \beta_2 x_{21} + \cdots + \beta_K x_{K1} \\ \vdots \\ \beta_1 + \beta_2 x_{2N} + \cdots + \beta_K x_{KN} \end{pmatrix}$$

つまり $X\beta$ は、N **人それぞれの「予測値」を縦に並べたもの**です。

見方2：列の組み合わせ。X の k 列目を $X_{(k)}$ と書くと

$$X\beta = \beta_1 X_{(1)} + \beta_2 X_{(2)} + \cdots + \beta_K X_{(K)}.$$

$X\beta$ は、**X の各列（各変数の N 人分のデータ）を係数で重みづけして足し合わせたもの**です。このような「スカラー倍して足し合わせたもの」を**線形結合**と呼びます。この見方は、第7章のランクや第8章の射影で重要になります。

4.4 $Y = X\beta + u$：N 本の式を1本で書く

序章の N 本の回帰式は、行列で書くと次の1本になります。

$$\underbrace{Y}_{N\times 1} = \underbrace{X}_{N\times K}\ \underbrace{\beta}_{K\times 1} + \underbrace{u}_{N\times 1} \tag{4.1}$$

数式の読み方：式(4.1)の確かめ方

- サイズ：$X\beta$ は $(N\times K)(K\times 1) = N\times 1$ で、Y と u と同じサイズなので、足し算ができます。
- 中身：両辺の i 番目の成分を比べると、$y_i = x_i\beta + u_i = \beta_1 + \beta_2 x_{2i} + \cdots + \beta_K x_{Ki} + u_i$。これは i さんの回帰式そのものです。
- つまり式(4.1)は、N 人分の回帰式を縦に積み重ねたものです。

4.5 行列の積の性質

- **順番を変えると一般に結果が変わる**：$AB \neq BA$（数の掛け算と最も違う点）。サイズが合わず片方しか計算できないことも多く、両方計算できても値が違うのが普通です。
- **結合法則**：$(AB)C = A(BC)$。どこから計算してもよい。
- **分配法則**：$A(B+C) = AB + AC$、$(A+B)C = AC + BC$。
- **単位行列は数の1**：$AI = IA = A$（サイズが合う単位行列で）。
- **積の転置は順番が逆**：$(AB)^\top = B^\top A^\top$。

計算例：$AB \neq BA$ の例

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix}.$$

右から掛けると列が入れ替わり、左から掛けると行が入れ替わります。

数式の読み方：$(AB)^\top = B^\top A^\top$ の理由

$(AB)^\top$ の (i,j) 成分は、AB の (j,i) 成分、つまり「A の j 行目と B の i 列目の内積」です。一方 $B^\top A^\top$ の (i,j) 成分は「$B^\top$ の i 行目（$= B$ の i 列目）と $A^\top$ の j 列目（$= A$ の j 行目）の内積」で、同じものです。サイズで考えても、A（$m \times n$）、B（$n \times p$）なら $(AB)^\top$ は $p \times m$ で、$B^\top A^\top$ は $(p \times n)(n \times m) = p \times m$ と合っています。
この公式は回帰分析で頻繁に使います。例えば $(X\beta)^\top = \beta^\top X^\top$、$(Y - X\beta)^\top = Y^\top - \beta^\top X^\top$ です。

要点

- AB の (i,j) 成分は「A の i 行目と B の j 列目の内積」。$(m \times n)(n \times p) = m \times p$。
- $X\beta$ は「各人の予測値を並べたもの」であり、「X の列の線形結合」でもある。
- $Y = X\beta + u$ は N 本の回帰式を積み重ねたもの。
- $AB \neq BA$、$(AB)^\top = B^\top A^\top$。

次の章へ　同じベクトルでも、「横 × 縦」と「縦 × 横」では全く違うものになります。第5章ではこの違いと、行列の積と和の記号 $\sum$ の関係を学びます。$X^\top X$ の正体が分かる章です。

第5章

「横×縦」と「縦×横」、和の記号と行列

ここまでの流れ　第4章で、行列の積とそのサイズのルールを学びました。
この章の問い　計量経済学の教科書に出てくる $X^\top X$ や $\sum_i x_i^\top x_i$ は何を計算しているのか。両者はどう関係しているのか。
この章で分かること　同じベクトルでも掛ける順番で**スカラー**にも**行列**にもなること、$X^\top X = \sum_i x_i^\top x_i$ という最重要の関係、$X^\top X$ の成分が「2乗和」と「積和」であること、そして母集団の式 $\beta = [E(x_i^\top x_i)]^{-1} E(x_i^\top y_i)$ の読み方です。

5.1　同じベクトルでも、掛ける順番で全く別物になる

列ベクトル $a = (1, 3)^\top$（2×1）を例にとります。

横 × 縦（内積）：

$$\underbrace{a^\top}_{1\times 2}\underbrace{a}_{2\times 1} = \begin{pmatrix} 1 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = 1 \cdot 1 + 3 \cdot 3 = 10 \quad (1 \times 1 \text{：スカラー})$$

縦 × 横（外積）：

$$\underbrace{a}_{2\times 1}\underbrace{a^\top}_{1\times 2} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \begin{pmatrix} 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 \cdot 1 & 1 \cdot 3 \\ 3 \cdot 1 & 3 \cdot 3 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 3 & 9 \end{pmatrix} \quad (2 \times 2 \text{：行列})$$

数式の読み方：なぜこうなるのか

- サイズのルール（第4章）で考えると、$(1\times 2)(2\times 1) = 1\times 1$、$(2\times 1)(1\times 2) = 2\times 2$ です。内側が消えて外側が残るので、「横 × 縦」はスカラー、「縦 × 横」は行列になります。
- 「縦 × 横」の (j, k) 成分は、左のベクトルの j 番目と右のベクトルの k 番目の**積**です（内積の「足す」がない）。全ての組み合わせの積を並べた表になります。
- 「縦 × 横」でできる行列は、(j, k) 成分と (k, j) 成分が等しい**対称行列**です。対角成分は各成分の2乗です。

5.2 回帰分析の $x_i^\top x_i$

本書では x_i は行ベクトル（$1 \times K$）なので、

- $x_i x_i^\top$（1×1）は「横 × 縦」で、スカラー $1 + x_{2i}^2 + \cdots + x_{Ki}^2$。
- $x_i^\top x_i$（$K \times K$）は「縦 × 横」で、行列。

回帰分析で重要なのは後者です。$K = 3$（定数項・x_2・x_3）なら

$$\underbrace{x_i^\top}_{3\times 1}\underbrace{x_i}_{1\times 3} = \begin{pmatrix} 1 \\ x_{2i} \\ x_{3i} \end{pmatrix} \begin{pmatrix} 1 & x_{2i} & x_{3i} \end{pmatrix} = \begin{pmatrix} 1 & x_{2i} & x_{3i} \\ x_{2i} & x_{2i}^2 & x_{2i}x_{3i} \\ x_{3i} & x_{3i}x_{2i} & x_{3i}^2 \end{pmatrix} \tag{5.1}$$

です。i さん1人分の「変数の値・2乗・積」を全部並べた表です。同様に

$$\underbrace{x_i^\top}_{3\times 1}\underbrace{y_i}_{1\times 1} = \begin{pmatrix} y_i \\ x_{2i}y_i \\ x_{3i}y_i \end{pmatrix} \quad (3 \times 1) \tag{5.2}$$

は、各説明変数と被説明変数の積を並べたベクトルです。

5.3 期待値をとる

行列やベクトルの**期待値**は、**成分ごとに**期待値をとったものと定義します。例えば式(5.1)の期待値は

$$E(x_i^\top x_i) = \begin{pmatrix} 1 & E(x_{2i}) & E(x_{3i}) \\ E(x_{2i}) & E(x_{2i}^2) & E(x_{2i}x_{3i}) \\ E(x_{3i}) & E(x_{3i}x_{2i}) & E(x_{3i}^2) \end{pmatrix} \quad (3 \times 3)$$

です。成分は、平均・2乗の平均・積の平均で、まとめて**2次のモーメント**と呼ばれます。同様に $E(x_i^\top y_i) = (E(y_i),\ E(x_{2i}y_i),\ E(x_{3i}y_i))^\top$ です。

5.4 和と行列の積：$X^\top X = \sum_i x_i^\top x_i$

ここが本章の最も重要な結果です。

要点：和と行列の積の関係

$$\underbrace{X^\top X}_{K\times K} = \sum_{i=1}^{N} \underbrace{x_i^\top x_i}_{K\times K}, \qquad \underbrace{X^\top Y}_{K\times 1} = \sum_{i=1}^{N} \underbrace{x_i^\top y_i}_{K\times 1} \tag{5.3}$$

行列の積 $X^\top X$ は、1人ずつ作った $x_i^\top x_i$ を N 人分足し合わせたものに等しい。

計算例：数値例で確かめる

第3章の数値例 $X = \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}$、$Y = (5, 1, 4)^\top$ を使います。

(i) 左辺：行列の積として計算。 $X^\top = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \end{pmatrix}$ (2×3) なので、$X^\top X$ は $(2 \times 3)(3 \times 2) = 2 \times 2$ です。

- (1,1) 成分：$X^\top$ の1行目 $(1,1,1)$ と X の1列目 $(1,1,1)^\top$ の内積 $= 3$。
- (1,2) 成分：$(1,1,1)$ と X の2列目 $(2,0,1)^\top$ の内積 $= 2+0+1 = 3$。
- (2,2) 成分：$(2,0,1)$ と $(2,0,1)^\top$ の内積 $= 4+0+1 = 5$。

よって $X^\top X = \begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix}$。

(ii) 右辺：1人ずつ作って足す。

$$x_1^\top x_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}, \quad x_2^\top x_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad x_3^\top x_3 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

を足すと $\begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix}$ で、(i) と一致します。

同様に、$X^\top Y = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 5 \\ 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 10 \\ 14 \end{pmatrix}$ と、$\sum_i x_i^\top y_i = \begin{pmatrix} 5 \\ 10 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 4 \\ 4 \end{pmatrix} = \begin{pmatrix} 10 \\ 14 \end{pmatrix}$ も一致します。

数式の読み方：なぜ一般に成り立つのか

$X^\top X$ の (j,k) 成分は、「$X^\top$ の j 行目（$=X$ の j 列目、変数 j の N 人分）」と「X の k 列目（変数 k の N 人分）」の内積なので

$$(X^\top X)_{jk} = \sum_{i=1}^{N} x_{ji}\, x_{ki}$$

です。一方、$x_i^\top x_i$ の (j,k) 成分は $x_{ji}x_{ki}$（i さんの変数 j と変数 k の積）なので、$\sum_i x_i^\top x_i$ の (j,k) 成分も $\sum_i x_{ji}x_{ki}$ です。全ての成分が等しいので、2つの行列は等しくなります。

言葉で言えば、$X^\top X$ は「**変数ごとに**（列どうしで）内積をとる」計算、$\sum_i x_i^\top x_i$ は「**人ごとに**表を作って足す」計算で、同じ積和を違う順番で足しているだけです。

5.5 $X^\top X$ の中身：2乗和と積和

定数項と1つの説明変数 x の場合、$X^\top X$ と $X^\top Y$ を成分で書くと

$$X^\top X = \begin{pmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix}, \qquad X^\top Y = \begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{pmatrix}$$

です。N で割ると

$$\frac{1}{N}X^\top X = \begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \overline{x^2} \end{pmatrix}, \qquad \frac{1}{N}X^\top Y = \begin{pmatrix} \bar{y} \\ \overline{xy} \end{pmatrix}$$

（$\overline{x^2}$ は x_i^2 の平均、$\overline{xy}$ は $x_i y_i$ の平均）となり、**標本の2次のモーメント**を並べたものになります。$X^\top X$ と $X^\top Y$ は、データの「平均・2乗の平均・積の平均」を全部まとめた表なのです。

発展：平均を引く行列

$\iota = (1, \ldots, 1)^\top$（$N \times 1$）を使って、$N \times N$ の行列

$$M_0 = I_N - \frac{1}{N}\iota\iota^\top$$

を作ると、$M_0 Y = (y_1 - \bar{y}, \ldots, y_N - \bar{y})^\top$ となり、データから平均を引いたベクトルが得られます。$\frac{1}{N}\iota\iota^\top Y = \iota\bar{y}$（全ての成分が $\bar{y}$ のベクトル）だからです。この行列を使うと、偏差平方和は $Y^\top M_0 Y = \sum_i (y_i - \bar{y})^2$ と書け、決定係数 R^2 などの式に登場します。

5.6 母集団の式と標本の式

大学院の計量経済学では、OLS の係数を次の形で書くことがあります。

$$\underbrace{\beta}_{K\times 1} = \underbrace{\left[E(x_i^\top x_i)\right]}_{K\times K}^{-1} \underbrace{E(x_i^\top y_i)}_{K\times 1} \tag{5.4}$$

これは**母集団**の係数で、「2次のモーメントの行列の逆行列（第6章）$\times$ x と y の積のモーメント」という形をしています。データから推定するときは、期待値 $E(\cdot)$ を標本の平均 $\frac{1}{N}\sum_i(\cdot)$ に置き換えます（**類推原理**、モーメント法）。

$$\hat{\beta} = \left[\frac{1}{N}\sum_{i=1}^{N} x_i^\top x_i\right]^{-1} \frac{1}{N}\sum_{i=1}^{N} x_i^\top y_i = (X^\top X)^{-1} X^\top Y$$

最後の等号は、式(5.3)と、$\frac{1}{N}$ が逆行列の中と外で打ち消し合う（$(\frac{1}{N}A)^{-1} = NA^{-1}$）ことから出ます。この式が何をしているかは、第6章（逆行列）と第10章（微分による導出）で詳しく見ます。

要点

- 「横 $\times$ 縦」はスカラー（内積）、「縦 $\times$ 横」は行列（外積）。x_i が行ベクトルなら $x_i^\top x_i$ は $K \times K$ 行列。
- 行列の期待値は成分ごとの期待値。
- $X^\top X = \sum_i x_i^\top x_i$、$X^\top Y = \sum_i x_i^\top y_i$。変数ごとの内積と、人ごとの表の和は同じもの。
- $X^\top X/N$、$X^\top Y/N$ は標本の2次のモーメントの表。母集団の式の期待値を標本平均に置き換えると OLS 推定量になる。

次の章へ　$\hat{\beta} = (X^\top X)^{-1} X^\top Y$ の中の「$^{-1}$」（逆行列）が何を意味するのかを、第6章で学びます。

第6章

逆行列・行列式・連立方程式

> **ここまでの流れ**　第5章で、$X^\top X$ と $X^\top Y$ が何を計算しているかを学びました。
> **この章の問い**　$(X^\top X)^{-1}$ の「-1」は何を意味するのか。行列で「割り算」はできるのか。
> **この章で分かること**　数の世界の「1」と「逆数」に当たる**単位行列**と**逆行列**、逆行列が存在するかを判定する**行列式**、逆行列を使った**連立方程式**の解き方、そしてOLSの**正規方程式**です。

6.1　数の世界とのたとえ

数の方程式 $3\beta = 6$ を解くには、両辺に 3 の逆数 $3^{-1} = 1/3$ を掛けて $\beta = 3^{-1} \times 6 = 2$ とします。逆数は「掛けると1になる数」（$3^{-1} \times 3 = 1$）です。

行列の世界にも同じ仕組みがあります。

	数の世界	行列の世界
「1」に当たるもの	1（$1 \times a = a$）	単位行列 I（$IA = AI = A$）
「逆数」に当たるもの	a^{-1}（$a^{-1}a = 1$）	逆行列 A^{-1}（$A^{-1}A = AA^{-1} = I$）
逆数がない場合	$a = 0$	行列式 $\lvert A\rvert = 0$
方程式の解	$a\beta = c \Rightarrow \beta = a^{-1}c$	$A\beta = c \Rightarrow \beta = A^{-1}c$

行列には「割り算」の記号はありません。その代わりに、逆行列を**掛ける**ことで割り算と同じ働きをさせます。行列の積は順番を変えられない（第4章）ので、左から掛けるか右から掛けるかに注意が必要です。

6.2　逆行列

正方行列 A（$n \times n$）に対して

$$A^{-1}A = AA^{-1} = I_n$$

を満たす行列 A^{-1} を、A の**逆行列**と呼びます。逆行列は正方行列についてだけ考え、存在しない場合もあります。逆行列が存在する行列を**正則行列**と呼びます。

6.2.1　2×2 行列の逆行列

2×2 行列の逆行列には、覚えておくと便利な公式があります。

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \Longrightarrow \quad A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \tag{6.1}$$

数式の読み方：公式の読み方と確かめ

- 作り方：対角成分 a と d を**入れ替え**、非対角成分 b と c の**符号を変え**、全体を $ad - bc$ で**割る**。
- 確かめ：$\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} da - bc & db - bd \\ -ca + ac & -cb + ad \end{pmatrix} = (ad - bc) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$。$ad - bc$ で割れば単位行列になります。
- $ad - bc = 0$ なら割れないので、逆行列は存在しません。この $ad - bc$ が次の節の**行列式**です。

計算例：数値例の $(X^\top X)^{-1}$

第5章で求めた $X^\top X = \begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix}$ の逆行列は、$ad - bc = 3 \cdot 5 - 3 \cdot 3 = 6$ なので

$$(X^\top X)^{-1} = \frac{1}{6} \begin{pmatrix} 5 & -3 \\ -3 & 3 \end{pmatrix}.$$

確かめ：$\frac{1}{6} \begin{pmatrix} 5 & -3 \\ -3 & 3 \end{pmatrix} \begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix} = \frac{1}{6} \begin{pmatrix} 15 - 9 & 15 - 15 \\ -9 + 9 & -9 + 15 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$。

6.2.2　逆行列の性質

- $(A^{-1})^{-1} = A$。
- **積の逆行列は順番が逆**：$(AB)^{-1} = B^{-1}A^{-1}$（$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AA^{-1} = I$ で確かめられます）。
- **転置と逆行列は入れ替えられる**：$(A^\top)^{-1} = (A^{-1})^\top$。とくに A が対称なら A^{-1} も対称です。
- スカラー倍：$(cA)^{-1} = \frac{1}{c}A^{-1}$。

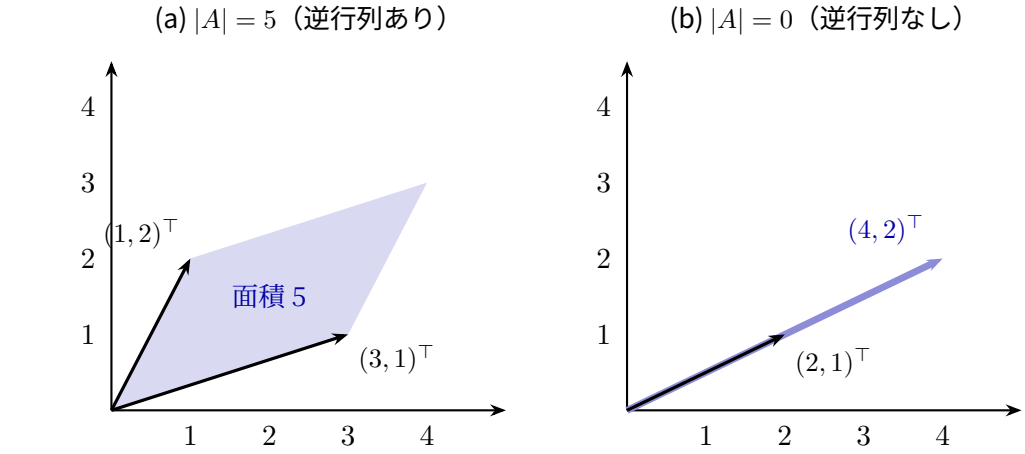
6.3 行列式

2×2 行列 $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ の**行列式**を

$$|A| = \det A = ad - bc$$

と定義します。逆行列が存在するのは $|A| \neq 0$ のときだけです。3×3 以上の行列にも行列式が定義され、同じく「$|A| \neq 0 \Leftrightarrow$ 逆行列が存在する」が成り立ちます（計算は手では面倒なので、計量経済学では統計ソフトに任せます）。

図 6.1　行列式の意味：列ベクトルがつくる平行四辺形の面積



(a) $A = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$、(b) $A = \begin{pmatrix} 2 & 4 \\ 1 & 2 \end{pmatrix}$。各図の矢印は A の2つの列ベクトル。

図の読み方

1. 2×2 行列の2つの列を矢印として描くと、行列式の絶対値は、その2本がつくる**平行四辺形の面積**に等しくなります。
2. **(a)**：2本が違う方向を向いているので、平行四辺形ができ、面積（行列式）は $3 \cdot 2 - 1 \cdot 1 = 5$ です。
3. **(b)**：2本目の列が1本目のちょうど2倍で、同じ直線の上に乗っています。平行四辺形がつぶれて面積は0、行列式も $2 \cdot 2 - 4 \cdot 1 = 0$ です。
4. 行列式が0になるのは、**列ベクトルのどれかが他の列で表せてしまう**（「つぶれている」）ときです。このとき逆行列は存在しません。これは第7章の「一次従属」と同じことです。

6.4 連立方程式を逆行列で解く

次の連立方程式を考えます。

$$\begin{cases} 3\beta_1 + 3\beta_2 = 10 \\ 3\beta_1 + 5\beta_2 = 14 \end{cases} \quad \Longleftrightarrow \quad \underbrace{\begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}}_{\beta} = \underbrace{\begin{pmatrix} 10 \\ 14 \end{pmatrix}}_{c}$$

左の連立方程式を行列で書いたのが右の式です（左辺を行列の積で計算すると、元の2本の式の左辺になることを確かめてください）。両辺に**左から** A^{-1} を掛けると

$$\begin{aligned} A^{-1}A\beta &= A^{-1}c && \text{（両辺に左から } A^{-1}\text{）} \\ I\beta &= A^{-1}c && \text{（} A^{-1}A = I \text{）} \\ \beta &= A^{-1}c && \text{（} I\beta = \beta \text{）} \end{aligned}$$

数値を入れると

$$\beta = \frac{1}{6}\begin{pmatrix} 5 & -3 \\ -3 & 3 \end{pmatrix}\begin{pmatrix} 10 \\ 14 \end{pmatrix} = \frac{1}{6}\begin{pmatrix} 50 - 42 \\ -30 + 42 \end{pmatrix} = \begin{pmatrix} 4/3 \\ 2 \end{pmatrix}.$$

6.5 OLS の正規方程式

上の連立方程式は、実は第3章の数値例の OLS の係数を求める式です。OLS の係数 $\hat{\beta}$ は、次の連立方程式（**正規方程式**）を満たします（導出は第10章）。

$$\underbrace{X^\top X}_{K\times K}\ \underbrace{\hat{\beta}}_{K\times 1} = \underbrace{X^\top Y}_{K\times 1} \tag{6.2}$$

K 個の未知数 $\hat{\beta}_1, \ldots, \hat{\beta}_K$ についての K 本の連立方程式です。$X^\top X$ に逆行列があれば、両辺に左から $(X^\top X)^{-1}$ を掛けて

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y \tag{6.3}$$

が得られます。数値例では $\hat{\beta} = (4/3,\ 2)^\top$：切片が $4/3$、傾きが 2 です。

数式の読み方：式(6.3)を読む

- サイズ：$(K \times K)^{-1}(K \times N)(N \times 1) = K \times 1$。係数の数だけの縦ベクトルになります。
- 意味：$X^\top Y$ は「各説明変数と y の積和」、$(X^\top X)^{-1}$ は「説明変数どうしの積和の表の逆行列」です。説明変数が1つで平均を引いたデータなら、$\hat{\beta} = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$（共分散 ÷ 分散）になり、学部で習った単回帰の傾きの公式と一致します。式(6.3)は「共分散 ÷ 分散」を、複数の説明変数に拡張したものだと読めます。
- よくある間違い：$(X^\top X)^{-1} = X^{-1}(X^\top)^{-1}$ とは**できません**。X は一般に正方行列ではなく、逆行列をもたないからです。

要点

- 単位行列 I は数の1、逆行列 A^{-1} は逆数に当たる。$A^{-1}A = AA^{-1} = I$。
- 2×2 の逆行列は「対角を入れ替え、非対角の符号を変え、$ad - bc$ で割る」。
- 行列式 $|A|$ は列ベクトルがつくる平行四辺形の面積（体積）。$|A| = 0$ なら逆行列は存在しない。
- OLS の係数は正規方程式 $X^\top X\hat{\beta} = X^\top Y$ の解で、$\hat{\beta} = (X^\top X)^{-1}X^\top Y$。

次の章へ　$(X^\top X)^{-1}$ が存在しないのはどんな場合でしょうか。第7章では、それが「説明変数のどれかが他の説明変数で表せてしまう」場合（完全多重共線性）であることを学びます。

第7章

一次独立とランク：多重共線性

ここまでの流れ　第6章で、$X^\top X$ に逆行列があれば OLS の係数が求まることを学びました。
この章の問い　$X^\top X$ に逆行列がないのはどんなときか。そのとき回帰分析では何が起こるのか。
この章で分かること　**一次独立**と**一次従属**、行列の**ランク**、そして回帰分析の**完全多重共線性**と**ダミー変数の罠**です。

7.1　一次独立と一次従属

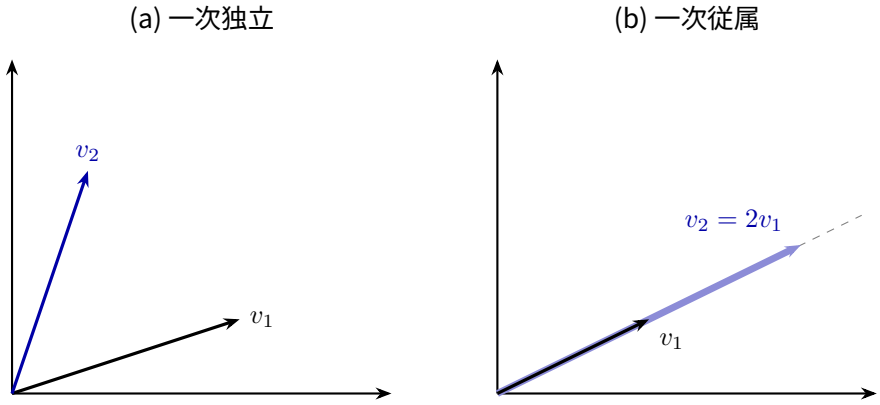
ベクトル $v_1, \dots, v_K$ について、どれか1つが**他のベクトルの線形結合（定数倍して足したもの）で表せる**とき、これらは**一次従属**であると言います。どれも他で表せないとき、**一次独立**であると言います。

数式の読み方：定義を式で書くと

$$c_1 v_1 + c_2 v_2 + \cdots + c_K v_K = 0$$

を満たす $c_1, \dots, c_K$ が「全て0」以外にないとき、一次独立です。もし0でない c_k があれば、その式を v_k について解くことで、v_k を他のベクトルで表せてしまう（一次従属）からです。

図 7.1　一次独立と一次従属



(b) の破線は、2本のベクトルが同じ直線の上に乗っていることを示す。

図の読み方

1. **(a)**：2本のベクトルは違う方向を向いており、一方を他方の定数倍で表せません（一次独立）。2本の線形結合で、平面上のどの点にも行けます。
2. **(b)**：$v_2 = 2v_1$ で、同じ直線の上にあります（一次従属）。2本あっても、線形結合で行けるのはその直線の上だけです。v_2 は新しい「方向」の情報を何も加えていません。

7.2 ランク

行列の**ランク**（階数）は、その行列の列ベクトルのうち、**一次独立なものの最大の本数**です。言い換えると、「本当に別々の情報をもっている列が何本あるか」です（行ベクトルで数えても同じ数になることが知られています）。

- $N \times K$ 行列のランクは、K と N のうち小さい方を超えません。
- 全ての列が一次独立なとき（ランクが列の数 K に等しいとき）、**列フルランク**と言います。
- $K \times K$ の正方行列については、「ランクが K」「行列式が0でない」「逆行列が存在する」は全て同じことです。

7.3 回帰分析での意味：完全多重共線性

要点：OLS が計算できる条件

データ行列 X（$N \times K$）が列フルランク（$\text{rank}(X) = K$）であることと、$X^\top X$ に逆行列が存在することは同値である。つまり、OLS 推定量 $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ が計算できるのは、**どの説明変数も他の説明変数の線形結合で表せない**ときである。

説明変数のどれかが他の説明変数の線形結合で表せる状態を**完全多重共線性**と呼びます。このとき $X^\top X$ の行列式は0で、逆行列が存在せず、OLS の係数を1つに決めることができません。

数式の読み方：なぜ係数が決まらないのか

例えば $x_3 = 2x_2$ だとします。回帰式 $y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + u$ に代入すると $y = \beta_1 + (\beta_2 + 2\beta_3) x_2 + u$ となり、データから分かるのは $\beta_2 + 2\beta_3$ の値だけです。$\beta_2 = 1$、$\beta_3 = 0$ でも、$\beta_2 = 0$、$\beta_3 = 0.5$ でも、全く同じ予測値になるので、データからは区別できません。

計算例：ダミー変数の罠

性別のダミー変数を、男性 $= 1$ の D^m と女性 $= 1$ の D^f の両方、さらに定数項と一緒に回帰に入れるとします。3人のデータ（男性、女性、男性）なら

$$X = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{matrix} \leftarrow \text{男性} \\ \leftarrow \text{女性} \\ \leftarrow \text{男性} \end{matrix}$$

です。どの人も $D^m + D^f = 1$ なので、**1列目 $=$ 2列目 $+$ 3列目**が成り立ち、一次従属です。ランクは2で、$X^\top X$ は逆行列をもちません。これを**ダミー変数の罠**と呼びます。解決策は、ダミー変数を1つ除く（または定数項を除く）ことです。

補足：不完全な多重共線性

説明変数どうしが完全には一次従属でなくても、**ほとんど**同じ方向を向いている（相関が非常に高い）場合があります。このとき $X^\top X$ の行列式は0に近く、逆行列の成分は非常に大きくなります。第11章で見るように、OLS の分散は $\sigma^2(X^\top X)^{-1}$ なので、係数の推定値が不安定になり、標準誤差が大きくなります。これが通常「多重共線性の問題」と呼ばれるものです。図6.1でいえば、平行四辺形がつぶれかけて面積が小さい状態です。

要点

- どれかのベクトルが他の線形結合で表せるなら一次従属、表せないなら一次独立。
- ランクは一次独立な列の最大本数。正方行列ではランクが最大 $\Leftrightarrow$ 行列式 $\neq 0 \Leftrightarrow$ 逆行列が存在。
- X が列フルランクのとき、かつそのときに限り $(X^\top X)^{-1}$ が存在し、OLS が計算できる。完全多重共線性やダミー変数の罠では計算できない。

次の章へ　OLS の計算ができるようになったので、次はその結果の「形」を見ます。第8章では、OLS が「データのベクトルを説明変数の空間に垂直に落とす（射影する）」操作であることを、射影行列を使って学びます。

第8章

射影行列：OLS の幾何学

ここまでの流れ　第6・7章で、OLS の係数 $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ の計算方法と、それが計算できる条件を学びました。
この章の問い　OLS の予測値と残差は、どんな性質をもつのか。OLS は図で見ると何をしているのか。
この章で分かること　**対称行列**と**べき等行列**、OLS の予測値をつくる**射影行列** P と残差をつくる**残差行列** M、そして「OLS はデータのベクトルを説明変数の空間に垂直に落とす操作である」という幾何学的な見方です。

8.1　対称行列とべき等行列

- **対称行列**：$A^\top = A$ を満たす正方行列（第3章）。
- **べき等行列**：$AA = A$ を満たす正方行列。「2回掛けても1回掛けたのと同じ」行列です。数でいえば $a^2 = a$ を満たす0と1に当たります。

8.2　予測値と残差を行列で書く

OLS の係数 $\hat{\beta}$ を使った予測値（当てはめ値）のベクトルは

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

です。そこで

$$\underbrace{P}_{N \times N} = X(X^\top X)^{-1} X^\top \tag{8.1}$$

とおくと、$\hat{Y} = PY$ と書けます。Y に P を掛けると予測値が得られる、という意味です。同様に、残差のベクトルは

$$\hat{u} = Y - \hat{Y} = Y - PY = (I_N - P)Y$$

なので、

$$\underbrace{M}_{N \times N} = I_N - P = I_N - X(X^\top X)^{-1} X^\top \tag{8.2}$$

とおくと $\hat{u} = MY$ です。P を**射影行列**（ハット行列）、M を**残差行列**（残差をつくる行列）と呼びます。

8.3 P と M の性質

要点：射影行列と残差行列の性質

1. P、M は対称：$P^\top = P$、$M^\top = M$。
2. P、M はべき等：$PP = P$、$MM = M$。
3. $PX = X$、$MX = O$。
4. $PM = MP = O$。

数式の読み方：性質の確かめ方

- **対称**：$P^\top = \big(X(X^\top X)^{-1}X^\top\big)^\top = X\big((X^\top X)^{-1}\big)^\top X^\top = X(X^\top X)^{-1}X^\top = P$。積の転置は順番が逆（第4章）、$X^\top X$ は対称なのでその逆行列も対称（第6章）であることを使いました。
- **べき等**：$PP = X(X^\top X)^{-1}\underbrace{X^\top X(X^\top X)^{-1}}_{=I}X^\top = X(X^\top X)^{-1}X^\top = P$。
- $PX = X$：$PX = X(X^\top X)^{-1}X^\top X = X$。意味：$Y$ の代わりに X の列そのものを被説明変数にして回帰すると、完全に当てはまる（予測値が元のデータそのもの）。
- $MX = O$：$MX = X - PX = X - X = O$。説明変数そのものを回帰すると残差はゼロ。
- M のべき等性：$MM = (I-P)(I-P) = I - 2P + PP = I - P = M$。

8.4 残差は説明変数と直交する

残差ベクトルと説明変数の関係を見ます。$MX = O$ の転置をとると $X^\top M = O$（M は対称）なので

$$X^\top \hat{u} = X^\top M Y = 0 \quad (K \times 1) \tag{8.3}$$

です。これは、**残差ベクトルが X の全ての列と直交する**（内積が0、第2章）ことを意味します。第6章の正規方程式 $X^\top X\hat{\beta} = X^\top Y$ を $X^\top(Y - X\hat{\beta}) = 0$ と書き直したものと同じです。

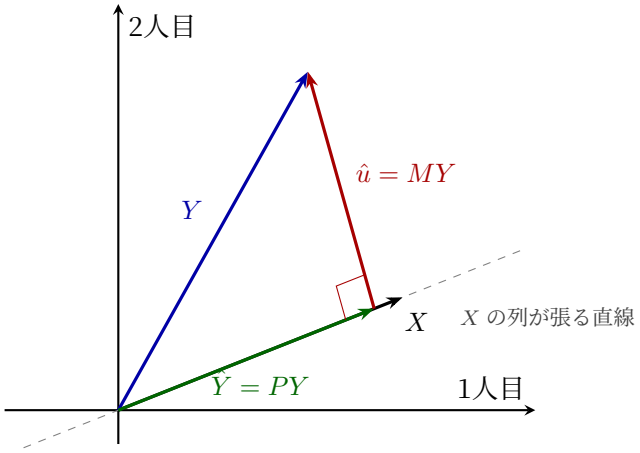
数式の読み方：直交条件の読み方

- X の1列目が定数項（全て1）なら、式(8.3)の1行目は $\sum_i \hat{u}_i = 0$：**残差の合計はゼロ**。
- 2行目以降は $\sum_i x_{ki}\hat{u}_i = 0$：**残差と各説明変数の積和はゼロ**（標本の中で無相関）。
- 数値例：残差は $\hat{u} = Y - X\hat{\beta} = (5,1,4)^\top - (16/3,\ 4/3,\ 10/3)^\top = (-1/3,\ -1/3,\ 2/3)^\top$。合計は0、$x_2 = (2,0,1)^\top$ との内積は $-2/3+0+2/3=0$ です。

8.5 図で見る OLS

説明変数が1つで定数項がない場合（$K=1$）、データが2人分（$N=2$）なら、全てを平面上の矢印として描けます。

図 8.1 OLS は Y を X の方向に垂直に射影する



$X = (3,1)^\top$、$Y = (2,3)^\top$ の場合。$\hat{\beta} = X^\top Y / X^\top X = 9/10$、$\hat{Y} = (2.7,\ 0.9)^\top$、$\hat{u} = (-0.7,\ 2.1)^\top$。

図の読み方

1. 縦軸・横軸は1人目・2人目のデータの値です。データのベクトル Y（青）と説明変数のベクトル X（黒）が、それぞれ1本の矢印になります。

2. $X\beta$ は β をどう選んでも、X の方向の直線（灰色の破線）の上にあります。OLS は、この直線上で Y に最も近い点を探します。

3. 最も近い点は、Y の先から直線に**垂直に下ろした足**です。これが予測値 $\hat{Y} = PY$（緑）で、「Y を直線に**射影**した」と言います。P が射影行列と呼ばれるのはこのためです。

4. 残差 $\hat{u} = MY$（赤）は、Y の先と $\hat{Y}$ を結ぶ矢印で、X と直交しています（直角の印）。これが式(8.3)の意味です。

5. 残差平方和は赤い矢印の長さの2乗です（第2章）。垂直に下ろしたときに最も短くなるので、これが最小二乗法になっています。

一般の場合（N 人、K 個の説明変数）でも同じです。Y は N 次元空間の矢印、X の K 本の列は K 次元の「平面」（**列空間**）を張り、OLS は Y をその平面に垂直に射影します。

8.6 分散の分解と決定係数

$Y = \hat{Y} + \hat{u}$ で、$\hat{Y}$ と $\hat{u}$ は直交しているので（$\hat{Y}^\top \hat{u} = Y^\top PMY = 0$）、三平方の定理と同じ関係

$$\underbrace{Y^\top Y}_{\|Y\|^2} = \underbrace{\hat{Y}^\top \hat{Y}}_{\|\hat{Y}\|^2} + \underbrace{\hat{u}^\top \hat{u}}_{\|\hat{u}\|^2}$$

が成り立ちます。平均を引いたデータで同じことを考えると、「全変動 = 回帰で説明される変動 + 残差の変動」となり、**決定係数** $R^2 = 1 - \hat{u}^\top \hat{u} / \sum_i (y_i - \bar{y})^2$ が定義されます。図でいえば、R^2 は Y と $\hat{Y}$ のなす角のコサインの2乗（平均を引いた場合）です。

発展：FWL 定理（部分回帰）

説明変数を2つのグループに分けて $Y = X_1\beta_1 + X_2\beta_2 + u$ と書きます。X_1 についての残差行列 $M_1 = I - X_1(X_1^\top X_1)^{-1}X_1^\top$ を使うと、β_2 の OLS 推定量は

$$\hat{\beta}_2 = (X_2^\top M_1 X_2)^{-1} X_2^\top M_1 Y$$

と書けます（フリッシュ＝ウォー＝ラヴェルの定理）。$M_1 Y$ は「Y から X_1 で説明できる部分を除いた残り」、$M_1 X_2$ は「X_2 から X_1 で説明できる部分を除いた残り」なので、**重回帰の係数は、他の変数の影響を取り除いた残差どうしの回帰の係数**に等しい、という意味です。「他の変数を制御する」ことの正体を表す重要な定理です。

要点

- 予測値 $\hat{Y} = PY$、残差 $\hat{u} = MY$。$P = X(X^\top X)^{-1}X^\top$、$M = I - P$。
- P、M は対称かつべき等。$PX = X$、$MX = O$、$PM = O$。
- 残差は全ての説明変数と直交する（$X^\top \hat{u} = 0$）。定数項があれば残差の合計は 0。
- OLS は Y を X の列が張る空間に垂直に射影する操作。残差平方和はその距離の2乗。

次の章へ　残差平方和 $\hat{u}^\top \hat{u}$ や分散のように、「ベクトル × 行列 × ベクトル」の形をした式がよく出てきます。第9章では、この「二次形式」と、それが常に正になる条件（正定値）を学びます。

第9章

二次形式と正定値行列

ここまでの流れ　第8章で、残差平方和が残差ベクトルの長さの2乗であることを見ました。
この章の問い　$c^\top Ac$ のような式は何を表すのか。それが常に正になるのはどんなときか。
この章で分かること　**二次形式**、**正定値・半正定値**の行列、$X^\top X$ が常に半正定値であること、そして残差平方和を係数の関数として見たときの「お椀の形」です。分散が負にならないことや、OLS が最小値を与えることの数学的な裏付けになります。

9.1　二次形式

$n \times 1$ のベクトル c と $n \times n$ の対称行列 A について

$$\underbrace{c^\top}_{1\times n}\ \underbrace{A}_{n\times n}\ \underbrace{c}_{n\times 1} \quad (1\times 1:\ \text{スカラー})$$

を**二次形式**と呼びます。サイズのルールから、結果は1つの数です。

計算例：2×2 の場合の展開

$c = (c_1, c_2)^\top$、$A = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$ なら

$$c^\top Ac = \begin{pmatrix} c_1 & c_2 \end{pmatrix} \begin{pmatrix} ac_1 + bc_2 \\ bc_1 + dc_2 \end{pmatrix} = ac_1^2 + 2bc_1c_2 + dc_2^2.$$

全ての項が c の成分の2次（2乗または2つの積）になるので「二次形式」と呼びます。対角成分が2乗の項の係数、非対角成分が積の項の係数（の半分）です。

数式の読み方：回帰分析の中の二次形式

- 平方和：$Y^\top Y = Y^\top I Y = \sum_i y_i^2$（$A = I$ の二次形式）。
- 残差平方和：$\hat{u}^\top \hat{u} = Y^\top M Y$（$M$ が対称・べき等なので $\hat{u}^\top \hat{u} = Y^\top M^\top M Y = Y^\top M Y$）。
- 偏差平方和：$Y^\top M_0 Y = \sum_i (y_i - \bar{y})^2$（第5章の発展）。
- 線形結合の分散：確率ベクトル z の分散共分散行列を Σ とすると、$\mathrm{Var}(c^\top z) = c^\top \Sigma c$（第11章）。

9.2 正定値と半正定値

対称行列 A が

- 0でない全てのベクトル c について $c^\top A c > 0$ を満たすとき、**正定値**、
- 全てのベクトル c について $c^\top A c \geq 0$ を満たすとき、**半正定値**

であると言います。数でいえば「正の数」「0以上の数」に当たる概念です。

計算例：判定の例

- $A = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$：$c^\top A c = 2c_1^2 + 3c_2^2 > 0$（$c \neq 0$）なので正定値。
- $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$：$c^\top A c = c_1^2 + 2c_1 c_2 + c_2^2 = (c_1 + c_2)^2 \geq 0$ なので半正定値。ただし $c = (1, -1)^\top$ のとき0になるので、正定値ではありません。この行列は行列式が0です。
- $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$：$c = (1, -1)^\top$ なら $c^\top A c = 1 - 4 + 1 = -2 < 0$ なので、半正定値ではありません。

9.3 $X^\top X$ はいつも半正定値

要点：$X^\top X$ の性質

どんなデータ行列 X についても $X^\top X$ は半正定値である。さらに X が列フルランクなら、$X^\top X$ は正定値である。

数式の読み方：証明

任意のベクトル c（$K \times 1$）について、

$$c^\top X^\top X c = (Xc)^\top (Xc) = \|Xc\|^2 \geq 0.$$

1つ目の等号は積の転置の公式 $(Xc)^\top = c^\top X^\top$、2つ目は「自分自身との内積 = 長さの2乗」（第2章）です。長さの2乗は負にならないので、半正定値です。
さらに X が列フルランク（第7章）なら、$c \neq 0$ のとき Xc（X の列の線形結合）は0になりません。したがって $\|Xc\|^2 > 0$ で、正定値です。

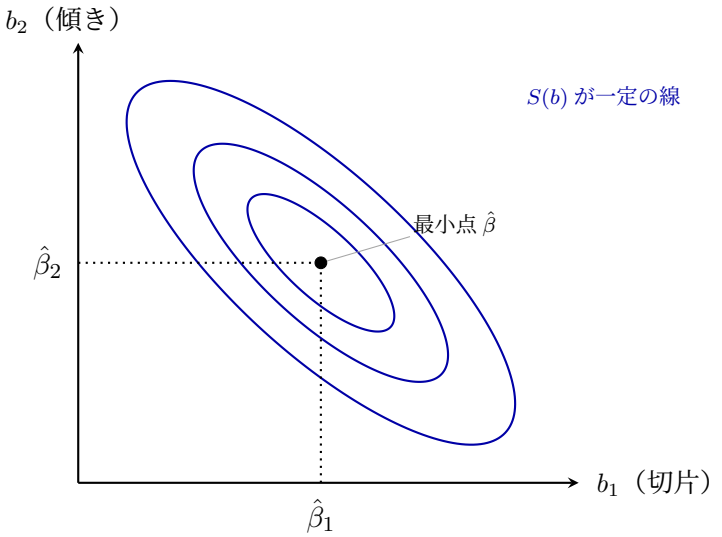
9.4 残差平方和の「お椀の形」

係数の候補 b についての残差平方和 $S(b) = (Y - Xb)^\top (Y - Xb)$ は、OLS の係数 $\hat{\beta}$ を使って次のように書き直せます（$X^\top \hat{u} = 0$ を使って展開すると確かめられます）。

$$S(b) = \underbrace{\hat{u}^\top \hat{u}}_{\text{最小値}} + (b - \hat{\beta})^\top X^\top X \, (b - \hat{\beta}) \tag{9.1}$$

第2項は $X^\top X$ の二次形式です。$X^\top X$ が正定値なら、第2項は $b \neq \hat{\beta}$ のとき必ず正なので、$S(b)$ **は** $b = \hat{\beta}$ **でただ1つの最小値をとります**。

図 9.1　残差平方和の等高線



第3章の数値例。内側から $S(b) - S(\hat{\beta}) = 0.5,\ 1.5,\ 3.5$ となる b の集まり。

図の読み方

1. 横軸と縦軸は、係数の候補 $b = (b_1, b_2)^\top$ です。曲線は残差平方和 $S(b)$ が同じ値になる b を結んだ**等高線**です。

2. 式(9.1)の第2項が正定値の二次形式なので、等高線は $\hat{\beta}$ を中心とする**楕円**になります。$S(b)$ は $\hat{\beta}$ を底とする「お椀」の形をしており、外側の楕円ほど残差平方和が大きくなります。

3. 楕円が斜めに傾いているのは、$X^\top X$ の非対角成分（$\sum_i x_i$）が0でないからです。切片を大きくしても傾きを小さくすれば同じくらいの当てはまりになる、という2つの係数のトレードオフを表しています。

4. もし X が列フルランクでない（完全多重共線性）と、$X^\top X$ は半正定値にすぎず、楕円が無限に伸びた「谷」になって、最小点が1つに決まりません（第7章）。

要点

- 二次形式 $c^\top A c$ はスカラーで、c の成分の2乗と積の和。平方和・残差平方和・分散が二次形式で書ける。
- 正定値：$c \neq 0$ なら $c^\top A c > 0$。半正定値：$c^\top A c \geq 0$。
- $X^\top X$ はいつも半正定値で、X が列フルランクなら正定値。
- 残差平方和は $\hat{\beta}$ を底とするお椀の形で、$X^\top X$ が正定値なら最小点は1つ。

次の章へ　お椀の底（最小点）を求めるには、微分して0とおけばよいはずです。第10章では、ベクトルで微分する方法と、それを使った OLS の導出を学びます。

第10章

ベクトルによる微分と OLS の導出

ここまでの流れ　第9章で、残差平方和が $\hat{\beta}$ を底とする「お椀」の形をしていることを見ました。
この章の問い　お椀の底を求めるには、ベクトル b でどう微分すればよいのか。OLS の公式はどう導かれるのか。
この章で分かること　ベクトルによる微分（**勾配**）の定義、計量経済学で使う2つの公式、そしてそれを使った**OLS 推定量の導出**と、最小であることの確認（**ヘッセ行列**）です。

10.1　ベクトルによる微分の定義

スカラーの値をとる関数 $f(b)$ が、K 個の変数 $b = (b_1, \ldots, b_K)^\top$ に依存するとします。このとき、f を各変数で偏微分したもの（第1章の偏微分）を縦に並べたベクトルを、f の b による微分（**勾配**）と呼びます。

$$\underbrace{\frac{\partial f}{\partial b}}_{K\times 1} = \begin{pmatrix} \partial f/\partial b_1 \\ \partial f/\partial b_2 \\ \vdots \\ \partial f/\partial b_K \end{pmatrix} \tag{10.1}$$

新しい計算規則ではなく、「偏微分を K 回して、結果を縦に並べただけ」です。最小値（最大値）では、全ての偏微分が0になるので、一階条件は $\partial f/\partial b = 0$（$K \times 1$ のゼロベクトル）と1本で書けます。

10.2　2つの基本公式

計量経済学でよく使うのは、次の2つの公式です。

要点：ベクトル微分の基本公式

a を $K \times 1$ の定数ベクトル、A を $K \times K$ の定数の対称行列とすると、

$$\frac{\partial\,(a^\top b)}{\partial b} = a \tag{10.2}$$

$$\frac{\partial\,(b^\top A b)}{\partial b} = 2Ab \tag{10.3}$$

数の微分 $\dfrac{d(ab)}{db} = a$、$\dfrac{d(ab^2)}{db} = 2ab$ と同じ形をしていることに注目してください。

数式の読み方：$K = 2$ で成分ごとに確かめる

式(10.2)：$a^\top b = a_1 b_1 + a_2 b_2$ なので、$\partial / \partial b_1 = a_1$、$\partial / \partial b_2 = a_2$。縦に並べると a です。

式(10.3)：$A = \begin{pmatrix} a & c \\ c & d \end{pmatrix}$ とすると、第9章の展開から $b^\top A b = a b_1^2 + 2 c b_1 b_2 + d b_2^2$。偏微分すると

$$\frac{\partial}{\partial b_1} = 2ab_1 + 2cb_2, \qquad \frac{\partial}{\partial b_2} = 2cb_1 + 2db_2.$$

縦に並べると $\begin{pmatrix} 2ab_1 + 2cb_2 \\ 2cb_1 + 2db_2 \end{pmatrix} = 2 \begin{pmatrix} a & c \\ c & d \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = 2Ab$ です。

補足：A が対称でない場合

A が対称でないときは $\dfrac{\partial (b^\top A b)}{\partial b} = (A + A^\top) b$ となります。A が対称なら $A + A^\top = 2A$ で式(10.3)に戻ります。計量経済学に出てくる二次形式の行列（$X^\top X$ など）はほぼ全て対称なので、式(10.3)で十分です。

10.3 OLS 推定量の導出

係数の候補 b（$K \times 1$）についての残差平方和

$$S(b) = \sum_{i=1}^{N} (y_i - x_i b)^2 = (Y - Xb)^\top (Y - Xb)$$

を最小にする b が OLS 推定量 $\hat{\beta}$ です。

10.3.1 ステップ1：展開する

$$\begin{aligned} S(b) &= (Y^\top - b^\top X^\top)(Y - Xb) && \text{（転置の公式 } (Xb)^\top = b^\top X^\top\text{）} \\ &= Y^\top Y - Y^\top Xb - b^\top X^\top Y + b^\top X^\top Xb && \text{（分配法則で4項に）} \\ &= Y^\top Y - 2 b^\top X^\top Y + b^\top X^\top Xb && (Y^\top Xb = b^\top X^\top Y) \end{aligned} \tag{10.4}$$

数式の読み方：$Y^\top Xb = b^\top X^\top Y$ の理由

$Y^\top Xb$ のサイズは $(1 \times N)(N \times K)(K \times 1) = 1 \times 1$ で、スカラーです。スカラーは転置しても変わらないので、$Y^\top Xb = (Y^\top Xb)^\top = b^\top X^\top Y$ です（積の転置は順番が逆）。2つの項は同じ数なので、まとめて $-2 b^\top X^\top Y$ になります。

10.3.2　ステップ2：b で微分して0とおく

式(10.4)の各項を微分します。

- $Y^\top Y$：b を含まないので0。
- $-2b^\top X^\top Y$：$a = X^\top Y$（$K \times 1$）とすると $-2a^\top b$ の形（$b^\top a = a^\top b$）なので、式(10.2)より $-2X^\top Y$。
- $b^\top X^\top X b$：$A = X^\top X$（対称）の二次形式なので、式(10.3)より $2X^\top X b$。

したがって一階条件は

$$\frac{\partial S}{\partial b} = -2X^\top Y + 2X^\top X b = 0 \quad \Longrightarrow \quad X^\top X \hat{\beta} = X^\top Y \tag{10.5}$$

となります。これが第6章の**正規方程式**です。

10.3.3　ステップ3：解く

X が列フルランクなら $X^\top X$ に逆行列があるので、左から $(X^\top X)^{-1}$ を掛けて

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

10.3.4　ステップ4：最小であることの確認

一階条件だけでは、最小か最大か分かりません。1変数の場合の二階条件（第1章）に当たるのが、勾配をもう一度微分した行列（**ヘッセ行列**）です。

$$\frac{\partial^2 S}{\partial b \, \partial b^\top} = 2X^\top X \quad (K \times K)$$

ヘッセ行列が正定値なら、関数は下に凸（お椀の形）で、一階条件を満たす点は最小点です。第9章で見たように、X が列フルランクなら $X^\top X$ は正定値なので、$\hat{\beta}$ は確かに残差平方和を最小にします。

要点

- ベクトルによる微分は、各変数の偏微分を縦に並べたもの。一階条件は $\partial f/\partial b = 0$。
- 基本公式：$\partial(a^\top b)/\partial b = a$、$\partial(b^\top Ab)/\partial b = 2Ab$（$A$ 対称）。
- 残差平方和を展開して微分すると、正規方程式 $X^\top X\hat{\beta} = X^\top Y$ が得られ、$\hat{\beta} = (X^\top X)^{-1}X^\top Y$。
- ヘッセ行列 $2X^\top X$ が正定値なので、これは最小点。

次の章へ　OLS 推定量の式が導けました。次は、その推定量が「平均的に正しいか」「どれくらいばらつくか」という統計的な性質を考えます。そのために、確率変数をベクトルにまとめた「確率ベクトル」と「分散共分散行列」を第11章で学びます。

第11章

確率ベクトルと分散共分散行列

> **ここまでの流れ**　第10章で、OLS 推定量 $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ を導きました。
> **この章の問い**　$\hat{\beta}$ の期待値や分散は、行列でどう書けるのか。教科書の $\mathrm{Var}(\hat{\beta} \mid X) = \sigma^2 (X^\top X)^{-1}$ はどう導かれるのか。
> **この章で分かること**　**確率ベクトル**の期待値、**分散共分散行列**、線形変換の公式、そしてそれを使った OLS の**不偏性**と**分散**の導出、さらに不均一分散のもとでの分散（**サンドイッチ型**）です。

11.1　確率ベクトルと期待値

確率変数を並べたベクトルを**確率ベクトル**と呼びます。例えば誤差項のベクトル $u = (u_1, \ldots, u_N)^\top$ は確率ベクトルです。確率ベクトルの期待値は、成分ごとの期待値を並べたものです（第5章）。

$$E(z) = \begin{pmatrix} E(z_1) \\ \vdots \\ E(z_n) \end{pmatrix}$$

11.2　分散共分散行列

確率ベクトル z（$n \times 1$）の期待値を $\mu = E(z)$ とすると、**分散共分散行列**を

$$\underbrace{\mathrm{Var}(z)}_{n \times n} = E\Big[\underbrace{(z-\mu)}_{n \times 1} \underbrace{(z-\mu)^\top}_{1 \times n}\Big] \tag{11.1}$$

と定義します。「縦 $\times$ 横」の外積（第5章）の期待値です。

数式の読み方：中身を見る（$n=2$）

$$(z-\mu)(z-\mu)^\top = \begin{pmatrix} (z_1-\mu_1)^2 & (z_1-\mu_1)(z_2-\mu_2) \\ (z_2-\mu_2)(z_1-\mu_1) & (z_2-\mu_2)^2 \end{pmatrix}$$

なので、期待値をとると

$$\mathrm{Var}(z) = \begin{pmatrix} \mathrm{Var}(z_1) & \mathrm{Cov}(z_1,z_2) \\ \mathrm{Cov}(z_2,z_1) & \mathrm{Var}(z_2) \end{pmatrix}.$$

対角成分が各変数の分散、非対角成分が2つの変数の共分散です。共分散は対称なので、分散共分散行列は対称行列です。

計算例：誤差項の分散共分散行列

回帰分析の標準的な仮定「各人の誤差の分散は等しく σ^2（**均一分散**）、異なる人の誤差は無相関」は、行列で1行に書けます。

$$\mathrm{Var}(u \mid X) = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix} = \sigma^2 I_N$$

対角成分が全て σ^2（均一分散）、非対角成分が全て0（無相関）です。分散が人によって違う（**不均一分散**）なら、対角成分が $\sigma_1^2, \ldots, \sigma_N^2$ と異なる対角行列になります。

11.3 線形変換の公式

確率ベクトル z（$n \times 1$）に定数の行列 A（$m \times n$）を掛け、定数ベクトル c（$m \times 1$）を足したもの $Az + c$ について、次の公式が成り立ちます。

要点：期待値と分散の線形変換

$$E(Az+c) = A\,E(z) + c \tag{11.2}$$

$$\underbrace{\mathrm{Var}(Az+c)}_{m\times m} = \underbrace{A}_{m\times n}\,\underbrace{\mathrm{Var}(z)}_{n\times n}\,\underbrace{A^\top}_{n\times m} \tag{11.3}$$

数式の読み方：公式の意味と導出

- 式(11.2)：期待値は「足し算と定数倍」を通り抜けられる（数の場合の $E(aX+b)=aE(X)+b$ と同じ）。Az の各成分は z の成分の線形結合なので、成分ごとにこの性質を使えば出ます。
- 式(11.3)：定義(11.1)に代入すると、$Az+c-E(Az+c)=A(z-\mu)$ なので

  $$\mathrm{Var}(Az+c)=E\big[A(z-\mu)(A(z-\mu))^\top\big]=E\big[A(z-\mu)(z-\mu)^\top A^\top\big]=A\,\mathrm{Var}(z)\,A^\top.$$

  定数 c は分散に影響しません。数の場合の $\mathrm{Var}(aX)=a^2\mathrm{Var}(X)$ に当たり、a^2 が「A と $A^\top$ で挟む」形になっています。
- 特に A が $1\times n$ の行ベクトル $c^\top$ なら、$\mathrm{Var}(c^\top z)=c^\top\mathrm{Var}(z)c$。線形結合の分散は二次形式（第9章）で、分散は負にならないので、**分散共分散行列は必ず半正定値**です。

11.4 OLS 推定量の期待値：不偏性

$Y=X\beta+u$ を $\hat\beta$ の式に代入します。

$$\begin{aligned}\hat\beta&=(X^\top X)^{-1}X^\top(X\beta+u) && (Y\text{ を代入})\\&=(X^\top X)^{-1}X^\top X\beta+(X^\top X)^{-1}X^\top u && (\text{分配法則})\\&=\beta+(X^\top X)^{-1}X^\top u && ((X^\top X)^{-1}X^\top X=I)\end{aligned}\tag{11.4}$$

推定量 = 真の値 + 誤差項の線形結合という形です。X を所与（条件付け）とすると、$(X^\top X)^{-1}X^\top$ は定数の行列 A（$K\times N$）とみなせるので、式(11.2)より

$$E(\hat\beta\mid X)=\beta+(X^\top X)^{-1}X^\top E(u\mid X)=\beta$$

です（$E(u\mid X)=0$ という仮定を使いました）。つまり OLS 推定量は**不偏**です。

11.5 OLS 推定量の分散

式(11.4)で $A=(X^\top X)^{-1}X^\top$ とすると、式(11.3)より

$$\begin{aligned}\mathrm{Var}(\hat\beta\mid X)&=A\,\mathrm{Var}(u\mid X)\,A^\top\\&=(X^\top X)^{-1}X^\top\,\mathrm{Var}(u\mid X)\,X(X^\top X)^{-1}\end{aligned}\tag{11.5}$$

です（$A^\top=X\big((X^\top X)^{-1}\big)^\top=X(X^\top X)^{-1}$）。

均一分散の場合：$\mathrm{Var}(u \mid X) = \sigma^2 I_N$ を代入すると

$$\mathrm{Var}(\hat{\beta} \mid X) = \sigma^2 (X^\top X)^{-1} \underbrace{X^\top X (X^\top X)^{-1}}_{=I} = \sigma^2 (X^\top X)^{-1} \tag{11.6}$$

となります。

数式の読み方：$\sigma^2 (X^\top X)^{-1}$ の読み方

- サイズは $K \times K$。対角成分が各係数の分散で、その平方根が**標準誤差**です（σ^2 は残差から推定した $s^2 = \hat{u}^\top \hat{u}/(N-K)$ で置き換えます）。非対角成分は係数の推定値どうしの共分散です。
- $X^\top X$ は2乗和の表なので、**データが多いほど、説明変数のばらつきが大きいほど、$X^\top X$ が「大きく」なり、その逆行列が小さくなって、推定が精密になります**。
- 説明変数どうしの相関が高い（多重共線性）と $X^\top X$ の行列式が0に近づき、逆行列の成分が大きくなって、標準誤差が大きくなります（第7章）。

不均一分散の場合：式(11.5)は、$X^\top \mathrm{Var}(u \mid X)\, X$ を $(X^\top X)^{-1}$ で両側から挟んだ形をしているので、**サンドイッチ型**の分散と呼ばれます。$\mathrm{Var}(u \mid X)$ の対角成分を残差の2乗 $\hat{u}_i^2$ で置き換えて推定したものが、統計ソフトの「頑健な標準誤差」（ホワイトの標準誤差）です。

要点

- 分散共分散行列 $\mathrm{Var}(z) = E[(z-\mu)(z-\mu)^\top]$ は、対角が分散、非対角が共分散の対称・半正定値行列。
- $E(Az+c) = AE(z)+c$、$\mathrm{Var}(Az+c) = A\,\mathrm{Var}(z)A^\top$。
- $\hat{\beta} = \beta + (X^\top X)^{-1} X^\top u$ なので、$E(u \mid X) = 0$ なら不偏。
- 均一分散なら $\mathrm{Var}(\hat{\beta} \mid X) = \sigma^2 (X^\top X)^{-1}$、一般にはサンドイッチ型 $(X^\top X)^{-1} X^\top \mathrm{Var}(u \mid X) X (X^\top X)^{-1}$。

次の章へ　回帰分析の基本的な行列の式は、これで一通り読めるようになりました。第12章では、もう一歩進んだ内容で出てくるトレース・固有値・2段階最小二乗法を簡潔に紹介します。

第12章

発展：トレース・固有値・2段階最小二乗法

ここまでの流れ　第11章までで、OLS の推定量・幾何学・統計的性質を行列で読めるようになりました。
この章の問い　一歩進んだ計量経済学の教科書に出てくる「トレース」「固有値」「P_Z」とは何か。
この章で分かること　誤差分散の推定量の自由度 $N-K$ を導く**トレース**、正定値性やべき等行列を特徴づける**固有値**、そして操作変数法の**2段階最小二乗法**の行列表現です。最初に読むときは、必要になったときに戻ってくる形でも構いません。

12.1　トレース

正方行列の対角成分の和を**トレース**と呼びます。

$$\mathrm{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn}$$

- $\mathrm{tr}(A+B) = \mathrm{tr}(A) + \mathrm{tr}(B)$、$\mathrm{tr}(cA) = c\,\mathrm{tr}(A)$。
- **順番を入れ替えられる**：$\mathrm{tr}(AB) = \mathrm{tr}(BA)$（サイズが違っても、両方の積が定義されていれば成り立つ）。
- スカラーはそれ自身のトレース：$a = \mathrm{tr}(a)$。

発展：なぜ $s^2 = \hat{u}^\top \hat{u}/(N-K)$ なのか

1. 射影行列のトレース：$\mathrm{tr}(P) = \mathrm{tr}\big(X(X^\top X)^{-1}X^\top\big) = \mathrm{tr}\big((X^\top X)^{-1}X^\top X\big) = \mathrm{tr}(I_K) = K$。順番の入れ替えで $N \times N$ の行列のトレースを $K \times K$ のトレースにしています。よって $\mathrm{tr}(M) = \mathrm{tr}(I_N) - \mathrm{tr}(P) = N - K$。
2. $\hat{u} = MY = M(X\beta + u) = Mu$（$MX = O$）なので、$\hat{u}^\top \hat{u} = u^\top M u$。
3. スカラーはトレースに等しく、順番を入れ替えて $u^\top M u = \mathrm{tr}(u^\top M u) = \mathrm{tr}(Muu^\top)$。
4. 期待値をとると $E(\hat{u}^\top \hat{u} \mid X) = \mathrm{tr}\big(M\,E(uu^\top \mid X)\big) = \mathrm{tr}(M\sigma^2 I) = \sigma^2(N-K)$。

したがって $\hat{u}^\top \hat{u}/(N-K)$ は σ^2 の不偏推定量です。N ではなく $N-K$ で割るのは、K 個の係数を推定したことで残差が K 本の制約（$X^\top \hat{u} = 0$）を受けているからで、$N-K$ を**自由度**と呼びます。

12.2 固有値と固有ベクトル

正方行列 A について、0でないベクトル v と数 λ が

$$Av = \lambda v$$

を満たすとき、λ を**固有値**、v を**固有ベクトル**と呼びます。「A を掛けても向きが変わらず、長さが λ 倍になるだけの方向」です。

- 対称行列の固有値は全て実数です。
- **対称行列が正定値** $\Leftrightarrow$ **全ての固有値が正**、半正定値 $\Leftrightarrow$ 全ての固有値が0以上。
- 行列式は固有値の積、トレースは固有値の和です。行列式が0（逆行列なし）$\Leftrightarrow$ 固有値に0がある。
- べき等行列の固有値は0か1だけで、1の個数がランク（$=$ トレース）に等しい。P は固有値1が K 個、M は $N-K$ 個です。

第9章の図9.1の楕円の軸の方向は $X^\top X$ の固有ベクトル、軸の長さは固有値の逆数の平方根に比例します。固有値に0に近いものがあると楕円が細長く伸び、多重共線性の問題が生じます。

12.3 2段階最小二乗法（操作変数法）

説明変数 X（$N \times K$）が誤差項と相関している（内生性）とき、誤差項と無相関で X と相関している**操作変数** Z（$N \times L$、$L \geq K$）を使います。Z についての射影行列

$$P_Z = Z(Z^\top Z)^{-1}Z^\top \quad (N \times N)$$

を使うと、**2段階最小二乗法**（2SLS）の推定量は

$$\hat{\beta}_{2SLS} = (X^\top P_Z X)^{-1} X^\top P_Z Y \tag{12.1}$$

と書けます。

数式の読み方：式(12.1)の読み方

- **第1段階**：X の各列を Z に回帰した予測値は $\hat{X} = P_Z X$（第8章の射影）。X のうち操作変数で説明できる部分（誤差項と無相関な部分）だけを取り出しています。
- **第2段階**：Y を $\hat{X}$ に回帰すると、OLS の公式から $(\hat{X}^\top \hat{X})^{-1}\hat{X}^\top Y$。
- P_Z は対称・べき等なので $\hat{X}^\top \hat{X} = X^\top P_Z^\top P_Z X = X^\top P_Z X$、$\hat{X}^\top Y = X^\top P_Z Y$ となり、式(12.1)に一致します。

OLS の式の $X^\top$ を $X^\top P_Z$ に置き換えただけの形をしており、射影行列の性質がそのまま使われています。

要点

- トレースは対角成分の和で、$\mathrm{tr}(AB) = \mathrm{tr}(BA)$。$\mathrm{tr}(M) = N - K$ から誤差分散の推定量の自由度が出る。
- 固有値は「向きを変えない方向の伸び率」。正定値 $\Leftrightarrow$ 全ての固有値が正。
- 2SLS は、X を操作変数の空間に射影してから OLS をする方法で、$\hat{\beta}_{2SLS} = (X^\top P_Z X)^{-1} X^\top P_Z Y$。

公式一覧と練習問題

終.1 よく使う公式一覧

公式	意味・使いどころ
$(m \times n)(n \times p) = m \times p$	積は内側のサイズが一致するときだけ計算できる（第4章）
$a^\top b = \sum_i a_i b_i$、$\|a\|^2 = a^\top a$	内積と長さ。残差平方和 $= \|\hat{u}\|^2$（第2章）
$AB \neq BA$、$(AB)^\top = B^\top A^\top$	積は順番を変えられない、転置は順番が逆（第4章）
$Y = X\beta + u$	N 本の回帰式をまとめたもの（第4章）
$X^\top X = \sum_i x_i^\top x_i$、$X^\top Y = \sum_i x_i^\top y_i$	行列の積と和の記号の関係（第5章）
$(AB)^{-1} = B^{-1}A^{-1}$、$(A^\top)^{-1} = (A^{-1})^\top$	逆行列の性質（第6章）
$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$	2×2 の逆行列（第6章）
$X^\top X\hat{\beta} = X^\top Y$、$\hat{\beta} = (X^\top X)^{-1}X^\top Y$	正規方程式と OLS 推定量（第6・10章）
$\mathrm{rank}(X) = K \Leftrightarrow (X^\top X)^{-1}$ が存在	完全多重共線性がない条件（第7章）
$P = X(X^\top X)^{-1}X^\top$、$M = I - P$	射影行列と残差行列。対称・べき等（第8章）
$PX = X$、$MX = O$、$X^\top \hat{u} = 0$	残差は説明変数と直交（第8章）
$c^\top X^\top X c = \|Xc\|^2 \geq 0$	$X^\top X$ は半正定値（第9章）
$\partial(a^\top b)/\partial b = a$、$\partial(b^\top Ab)/\partial b = 2Ab$	ベクトル微分の基本公式（第10章）
$E(Az+c) = AE(z)+c$、$\mathrm{Var}(Az+c) = A\mathrm{Var}(z)A^\top$	線形変換の公式（第11章）
$\mathrm{Var}(\hat{\beta} \mid X) = \sigma^2(X^\top X)^{-1}$	均一分散のもとでの OLS の分散（第11章）
$\mathrm{tr}(AB) = \mathrm{tr}(BA)$、$\mathrm{tr}(M) = N - K$	トレースと自由度（第12章）

終.2 練習問題

1. $A = \begin{pmatrix} 1 & 0 & 2 \\ -1 & 3 & 1 \end{pmatrix}$、$B = \begin{pmatrix} 2 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$ のとき、AB と BA のサイズを答え、AB を計算せよ。

（答：AB は 2×2、BA は 3×3。$AB=\begin{pmatrix}4 & 1\\ -1 & 2\end{pmatrix}$）

2. $a=(2,-1,3)^\top$ について、$a^\top a$ と $aa^\top$ を計算せよ。
   （答：$a^\top a=14$。$aa^\top=\begin{pmatrix}4 & -2 & 6\\ -2 & 1 & -3\\ 6 & -3 & 9\end{pmatrix}$）

3. $X=\begin{pmatrix}1 & 1\\ 1 & 2\\ 1 & 3\\ 1 & 4\end{pmatrix}$、$Y=(2,3,5,6)^\top$ のとき、$X^\top X$、$X^\top Y$、$(X^\top X)^{-1}$、$\hat\beta$ を求めよ。
   （答：$X^\top X=\begin{pmatrix}4 & 10\\ 10 & 30\end{pmatrix}$、$X^\top Y=\begin{pmatrix}16\\ 47\end{pmatrix}$、$(X^\top X)^{-1}=\frac{1}{20}\begin{pmatrix}30 & -10\\ -10 & 4\end{pmatrix}$、$\hat\beta=\frac{1}{20}\begin{pmatrix}480-470\\ -160+188\end{pmatrix}=\begin{pmatrix}0.5\\ 1.4\end{pmatrix}$）

4. 問3の残差 $\hat u$ を求め、$X^\top\hat u=0$ を確かめよ。
   （答：$X\hat\beta=(1.9,\ 3.3,\ 4.7,\ 6.1)^\top$、$\hat u=(0.1,\ -0.3,\ 0.3,\ -0.1)^\top$。合計 0、$1\cdot 0.1+2\cdot(-0.3)+3\cdot 0.3+4\cdot(-0.1)=0$）

5. 定数項・男性ダミー・女性ダミーを全て含むデータ行列のランクが列の数より小さくなる理由を説明せよ。

6. $P=X(X^\top X)^{-1}X^\top$ が対称かつべき等であることを示せ。

7. $S(b)=(Y-Xb)^\top(Y-Xb)$ を展開し、b で微分して正規方程式を導け。

8. $\mathrm{Var}(u\mid X)=\sigma^2 I$ のとき、$\mathrm{Var}(\hat\beta\mid X)=\sigma^2(X^\top X)^{-1}$ を導け。

9. 確率ベクトル $z=(z_1,z_2)^\top$ の分散共分散行列が $\begin{pmatrix}4 & 1\\ 1 & 2\end{pmatrix}$ のとき、z_1-z_2 の分散を求めよ。
   （答：$c=(1,-1)^\top$ として $c^\top\Sigma c=4-2\cdot 1+2=4$）

最後に　行列の式で迷ったら、(1) 各記号のサイズを書き出す、(2) 小さい数値例（2人・2変数など）で成分を計算してみる、(3) 和の記号 $\sum$ で書き直してみる、の3つを試してください。本書の式は全て、この3つの方法で中身を確かめることができます。