データサイエンスの基礎
Rによる統計学独習: 補遺

地道 正行

2018 年 10 月 2 日

はじめに

拙著『R による統計学独習』をご購入いただき，まことに有り難うございます.

本文書では, 紙幅の関係上, 本文中には含めることが難しかった 2 つの話題について, 補遺として述べています.

まず, 最初の話題は「分散公式」です. この公式は, 分散が「2 乗平均」から「平均の 2 乗」を引くことによって得られることを表しています. この結果は, 単純な計算によって証明することができますが, 平方和を計算した結果として, 交差項が「消える」理由が判然としないことに疑問を感じる読者も多いのではないでしょうか. ここでは, この理由が「直交」という線形代数の知識を使って説明できることを解説しています.

次の話題は, いわゆる「重回帰分析」における一般的な説明です. 回帰分析に関する書籍の中には, 説明変数が単一の場合である「単回帰分析」を扱ったあとで, すぐに説明変数が一般の個数の場合の「重回帰分析」が扱われているものもありますが, このような書籍を読んだ初学者は, 戸惑いを感じることと思われます. この理由から, 本書では, 重回帰モデルにおける説明変数が 2 の場合について詳しく説明することを試みました. しかしながら, 3 以上の場合については, 残念ならが詳しく述べることができませんでしたので, 本文書に, 一般の場合についての重回帰分析を, より一般的な観点から「線形回帰分析」とよんで解説しています.

これらの話題は, 共に数学的には線形代数における「射影」が本質であり, 統計学的には「回帰」と同義であることを協調しておきたいと思います.

本補遺が本書の内容を補完し, 読者の皆さんが統計学を理解するために役立つものとなることを願います.

2018 年 初秋
筆者

目 次

第1章　分散公式

ここではデータの分散公式の証明と幾何学的説明を与える. 分散公式は線形代数における射影の概念を導入することによって明快に理解することができることに注意しよう.

1.1　分散公式の証明

データ $\{x_1, \ldots, x_n\}$ に対して, 平均が

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + \cdots + x_n)$$

であり, 分散が

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\left\{(x_1 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2\right\}$$

で定義されることを確認しておく. データの分散公式は以下の命題によって与えられる.

命題 1.1 (分散公式) 分散 s^2 は以下のように表すことができる:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2$$

証明: 等式

$$x_i = (x_i - \overline{x}) + \overline{x}$$

の両辺の 2 乗を $i = 1, \ldots, n$ について和をとることによって,

$$\begin{aligned}\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n}\left\{(x_i - \overline{x}) + \overline{x}\right\}^2 \\ &= \sum_{i=1}^{n}\left\{(x_i - \overline{x})^2 + \overline{x}^2 + 2\overline{x}(x_i - \overline{x})\right\} \\ &= \sum_{i=1}^{n}(x_i - \overline{x})^2 + n\overline{x}^2 + 2\overline{x}\sum_{i=1}^{n}(x_i - \overline{x}) \qquad (1.1)\end{aligned}$$

となる. ここで, 平均に関して,

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \Longleftrightarrow n\overline{x} = \sum_{i=1}^{n} x_i$$

が成り立つことを注意すると

$$\sum_{i=1}^{n}(x_i - \overline{x}) = \sum_{i=1}^{n} x_i - n\overline{x} = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0$$

となり, (1.1) 式にこの結果を代入すると,

$$\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 + n\overline{x}^2$$

が得られる. この式の両辺を n で割って整理すると,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2 + \frac{1}{n} n\overline{x}^2 \\
\Longleftrightarrow &\frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 \\
\Longleftrightarrow &s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2
\end{aligned}$$

となり, これは分散公式に他ならない. □

1.2 分散公式に関する幾何学的説明

分散公式が成り立つのは, その証明において等式

$$x_i = (x_i - \overline{x}) + \overline{x} \quad (i = 1, \ldots, n) \tag{1.2}$$

の両辺の 2 乗和が

$$\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 + n\overline{x}^2 \tag{1.3}$$

と表せることによる. すなわち, 等式 (1.2) の右辺の 2 乗を考えるときに, 以下のように交差項の和がゼロとなることが本質的である:

$$\sum_{i=1}^{n} (x_i - \overline{x})\overline{x} = 0 \tag{1.4}$$

このことは, 線形代数の知識を使うと以下のように幾何学的に説明することができる.

まず, 任意の実ベクトル (列ベクトル)

$$\boldsymbol{a} := \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{R}^n, \quad \boldsymbol{b} := \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \in \mathbb{R}^n$$

に対して内積を

$$\langle \boldsymbol{a}, \boldsymbol{b} \rangle := \sum_{i=1}^{n} a_i b_i$$

と定義することによって, ノルム (長さ)

$$\|\boldsymbol{a}\| := \sqrt{\langle \boldsymbol{a}, \boldsymbol{a} \rangle} = \sqrt{\sum_{i=1}^{n} a_i^2}$$

が定義されることに注意しよう.

次に, データ等を以下のように (列) ベクトルとして表現する:

$$\boldsymbol{x} := \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{1} := \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} (\in \mathbb{R}^n), \quad \overline{x}\mathbf{1} = \begin{bmatrix} \overline{x} \\ \vdots \\ \overline{x} \end{bmatrix}, \quad \boldsymbol{x} - \overline{x}\mathbf{1} = \begin{bmatrix} x_1 - \overline{x} \\ \vdots \\ x_n - \overline{x} \end{bmatrix}$$

ここで, $\overline{x}\mathbf{1}$ を**平均ベクトル**, $\boldsymbol{x} - \overline{x}\mathbf{1}$ を**偏差ベクトル**とよぶことにする.

これらのベクトルに関して, ノルムと内積を用いると以下のことが成り立つことに注意しよう:

$$\begin{aligned} \|\boldsymbol{x}\|^2 &= \langle \boldsymbol{x}, \boldsymbol{x} \rangle = \sum_{i=1}^n x_i^2, \\ \|\boldsymbol{x} - \overline{x}\mathbf{1}\|^2 &= \langle \boldsymbol{x} - \overline{x}\mathbf{1}, \boldsymbol{x} - \overline{x}\mathbf{1} \rangle = \sum_{i=1}^n (x_i - \overline{x})^2, \\ \|\overline{x}\mathbf{1}\|^2 &= \langle \overline{x}\mathbf{1}, \overline{x}\mathbf{1} \rangle = n\overline{x}^2, \\ \langle \boldsymbol{x} - \overline{x}\mathbf{1}, \overline{x}\mathbf{1} \rangle &= \sum_{i=1}^n (x_i - \overline{x})\overline{x} \end{aligned}$$

以上のことから, 分散公式に関する等式 (1.3) は,

$$\|\boldsymbol{x}\|^2 = \|\boldsymbol{x} - \overline{x}\mathbf{1}\|^2 + \|\overline{x}\mathbf{1}\|^2 \tag{1.5}$$

と表すことができ, 分散公式が成り立つために本質的に重要な等式 (1.4) は,

$$\langle \boldsymbol{x} - \overline{x}\mathbf{1}, \overline{x}\mathbf{1} \rangle = 0 \tag{1.6}$$

と表すことができる. すなわち, 偏差ベクトル $\boldsymbol{x} - \overline{x}\mathbf{1}$ と平均ベクトル $\overline{x}\mathbf{1}$ が直交することを示すことが定理の証明の本質といえる. このことを以下のように記号的に表す:

$$(\boldsymbol{x} - \overline{x}\mathbf{1}) \perp \overline{x}\mathbf{1} \tag{1.7}$$

この直交性を示すために以下のように行列を定義する:

$$\begin{aligned} \mathbf{P_1} &:= \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}' = \frac{1}{n}\mathbf{1}\mathbf{1}' = \frac{1}{n}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}, \\ \mathbf{Q_1} &:= \mathbf{I}_n - \mathbf{P_1} = \begin{bmatrix} 1 - \frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \ddots & \vdots \\ -\frac{1}{n} & \cdots & 1 - \frac{1}{n} \end{bmatrix} \end{aligned}$$

ここで, プライム ($'$) はベクトル, 行列の転置を表す.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i = \frac{1}{n}(x_1 + \cdots + x_n) = \frac{1}{n}\mathbf{1}'\boldsymbol{x}$$

に注意すると, 行列 $\mathbf{P_1}$ と $\mathbf{Q_1}$ によって平均ベクトルと偏差ベクトルが以下のように与えられる:

$$\mathbf{P_1}\boldsymbol{x} = \frac{1}{n}\mathbf{1}\mathbf{1}'\boldsymbol{x} = \overline{x}\mathbf{1} \tag{1.8}$$

$$\mathbf{Q_1}\boldsymbol{x} = (\mathbf{I}_n - \mathbf{P_1})\boldsymbol{x} = \boldsymbol{x} - \mathbf{P_1}\boldsymbol{x} = \boldsymbol{x} - \overline{x}\mathbf{1} \tag{1.9}$$

よって, 等式 (1.2) は以下のように表現できることに注意しよう:

$$\boldsymbol{x} = (\boldsymbol{x} - \overline{x}\mathbf{1}) + \overline{x}\mathbf{1} = \mathbf{Q_1}\boldsymbol{x} + \mathbf{P_1}\boldsymbol{x} \tag{1.10}$$

ここで, 行列 $\mathbf{P_1}$, $\mathbf{Q_1}$ のそれぞれについて以下の性質が成り立つことに注意する:

$$\mathbf{P}_1' = \mathbf{P_1}, \quad \mathbf{Q}_1' = \mathbf{Q_1} \quad (: \text{対称性})$$
$$\mathbf{P}_1^2 = \mathbf{P_1}, \quad \mathbf{Q}_1^2 = \mathbf{Q_1} \quad (: \text{巾等性})$$

一般に, 対称性と巾等性をみたす行列は**射影行列** (projection matrix) とよばれることに注意しよう. さらに, 行列 $\mathbf{P_1}$ と $\mathbf{Q_1}$ について,

$$\mathbf{P_1}\mathbf{Q_1} = \mathbf{Q_1}\mathbf{P_1} = \mathbf{O} \quad (: \text{直交性})$$

が成り立つことにも注意しよう. このことから,

$$\langle \mathbf{P_1}\boldsymbol{x}, \mathbf{Q_1}\boldsymbol{x} \rangle = (\mathbf{P_1}\boldsymbol{x})'\mathbf{Q_1}\boldsymbol{x} = \boldsymbol{x}'\mathbf{P}_1'\mathbf{Q_1}\boldsymbol{x} = \boldsymbol{x}'\mathbf{P_1}\mathbf{Q_1}\boldsymbol{x} = 0$$

となり, 平均ベクトルと偏差ベクトルの直交性の等式 (1.6), すなわち, 等式 (1.4) が示されたことになる. よって, (1.10) 式よりベクトル $\boldsymbol{x}$ のノルムの 2 乗は,

$$\begin{aligned} \|\boldsymbol{x}\|^2 &= \|\mathbf{Q_1}\boldsymbol{x}\|^2 + \|\mathbf{P_1}\boldsymbol{x}\|^2 + 2\langle \mathbf{P_1}\boldsymbol{x}, \mathbf{Q_1}\boldsymbol{x} \rangle \\ &= \|\mathbf{Q_1}\boldsymbol{x}\|^2 + \|\mathbf{P_1}\boldsymbol{x}\|^2 \\ &= \|\boldsymbol{x} - \overline{x}\mathbf{1}\|^2 + \|\overline{x}\mathbf{1}\|^2 \end{aligned}$$

となり, (1.5) 式が示された.

これらのことは射影行列の役割から以下のように幾何学的に示すことができる. まず, ベクトル $\mathbf{1}$ によって張られる空間を

$$\mathcal{M}(\mathbf{1}) := \{\alpha\mathbf{1} \ ; \ \alpha \in \mathbb{R}\}$$

で定義する. ここで, α はスカラーである.

次に, この空間 $\mathcal{M}(\mathbf{1})$ に直交する空間 [1] を以下のように定義する:

$$\mathcal{M}^{\perp}(\mathbf{1}) := \{\boldsymbol{a} \in \mathbb{R}^n \ ; \ \langle \boldsymbol{a}, \boldsymbol{b} \rangle = 0, \ \boldsymbol{b} \in \mathcal{M}(\mathbf{1})\}$$

行列 $\mathbf{P_1}$ と $\mathbf{Q_1}$ は, n 次元実ベクトル空間 $\mathbb{R}^n$ から, それぞれ, $\mathcal{M}(\mathbf{1})$, $\mathcal{M}^{\perp}(\mathbf{1})$ へ (正) 射影する行列であることに注意しよう [2] . すなわち,

$$\mathbf{P_1} : \mathbb{R}^n \longrightarrow \mathcal{M}(\mathbf{1}),$$
$$\mathbf{Q_1} : \mathbb{R}^n \longrightarrow \mathcal{M}^{\perp}(\mathbf{1})$$

であり, かつ, 任意の n 次元実ベクトル $\boldsymbol{x}$ に対して,

$$\mathbf{P}_1^2\boldsymbol{x} = \mathbf{P_1}\boldsymbol{x} \in \mathcal{M}(\mathbf{1}), \quad \mathbf{Q}_1^2\boldsymbol{x} = \mathbf{Q_1}\boldsymbol{x} \in \mathcal{M}^{\perp}(\mathbf{1}),$$
$$\langle \mathbf{P_1}\boldsymbol{x}, \mathbf{Q_1}\boldsymbol{x} \rangle = 0 \quad (\Longleftrightarrow \mathbf{P_1}\boldsymbol{x} \perp \mathbf{Q_1}\boldsymbol{x})$$

---

[1] 正確には, **直交補空間**である.

[2] 行列 $\mathbf{P_1}$ は, 「空間 $\mathcal{M}(\mathbf{1})^{\perp}$ に沿った空間 $\mathcal{M}(\mathbf{1})$ への正射影行列」とよばれることがあり, 同様に, $\mathbf{Q_1}$ は, 「空間 $\mathcal{M}(\mathbf{1})$ に沿った空間 $\mathcal{M}^{\perp}(\mathbf{1})$ への正射影行列」とよばれることがあることに注意しよう. ここで**正射影** (orthogonal projection) の「正」とは「垂直に」という意味を持つことに注意しよう. また, 射影行列は**射影子** (predictor) ともよばれることに注意しよう.

が成り立つ.

以上の結果を幾何学的に説明したものが図 1.1 である.

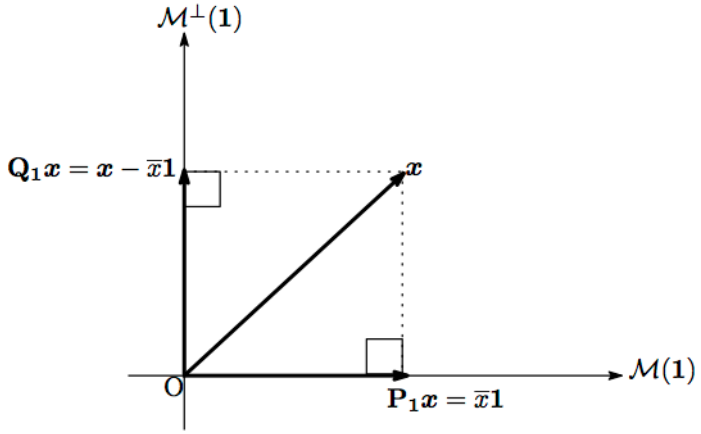

図 1.1: 射影行列の幾何学的説明

さらに, (1.7) 式と (1.10) 式は, 図 1.2 において説明されている. すなわち, $\boldsymbol{x}$ が直交する 2 つのベクトル (平均ベクトル $\bar{x}\mathbf{1}$ と偏差ベクトル $\boldsymbol{x} - \bar{x}\mathbf{1}$) の和によって分解されていることを示している. また, (1.5) 式は, この図における (n 次元実ベクトル空間内の) 直角三角形に関して一般化された「ピタゴラスの定理」が成り立つことを示していることに注意しよう.

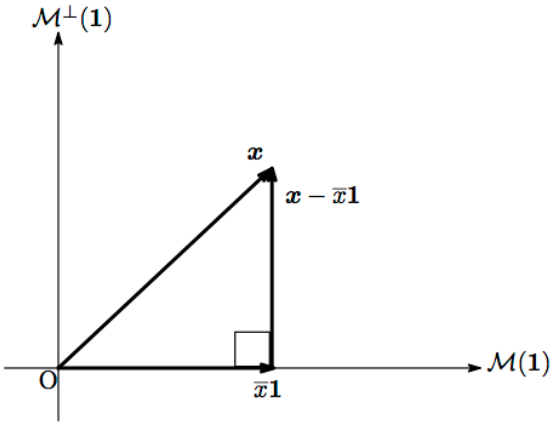

図 1.2: 分散公式に関する幾何学的説明

注意 1.1 ここで扱った分散公式に限らず統計学 (特に分散分析, 回帰分析) では, ある変量の 2 乗をそれらの和において交差項がなくなるような項を加減したものを展開 (分解) することによってその変動を調べる方法がある. これらは, 数学的 (幾何学的) には, あるベクトルを直交するいくつかの空間へ射影することによって, 分解されたもののノルム (の 2 乗) をその変動と見なして比較する手法であることに注意しよう.

注意 1.2 一般に, 1 次元の場合には, 実数 $a, b \in \mathbb{R}^1$ に関して,

$$(a+b)^2 = a^2 + b^2 + 2ab$$

が成り立ち,

$$(a+b)^2 = a^2 + b^2$$

が成り立つのは, $a = 0$ または $b = 0$ が成り立つ場合に限られる.

一方, n 次元 ($n \geq 2$, すなわち 2 次元以上) の場合に, 実ベクトル $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$ に対して,

$$\|\boldsymbol{a} + \boldsymbol{b}\|^2 = \|\boldsymbol{a}\|^2 + \|\boldsymbol{b}\|^2 + 2\langle \boldsymbol{a}, \boldsymbol{b} \rangle$$

が成り立つ. この等式において, ベクトル $\boldsymbol{a}$ と $\boldsymbol{b}$ が直交する場合, すなわち,

$$\langle \boldsymbol{a}, \boldsymbol{b} \rangle = 0$$

となるときは,

$$\|\boldsymbol{a} + \boldsymbol{b}\|^2 = \|\boldsymbol{a}\|^2 + \|\boldsymbol{b}\|^2 \tag{1.11}$$

となる. よって, 1 次元の場合のように $\boldsymbol{a} = \boldsymbol{0}$ または $\boldsymbol{b} = \boldsymbol{0}$ が成り立つ場合に限られるわけでなく, 直交するベクトルは無数に存在するため, (1.11) 式が成り立つ場合は無数に存在することに注意しよう. ただし, 分散公式の場合は, ベクトル $\boldsymbol{x}$ と射影する空間が固定されているため, 直交するベクトルは一意に定まることにも注意しよう.

第2章　線形回帰分析

ここでは, 説明変数が一般に p 個の場合の重回帰分析を考える. なお, ここでは説明変数の線形結合で応答変数をモデル化したもので分析を行うため, より一般的な観点からの名称として**線形回帰分析** (linear regression analysis) とよぶ.

2.1　線形回帰分析

線形回帰モデル

変量 $(X_1, \ldots, X_p, Y)$ を考え, その実現値を $(x_1, \ldots, x_p, y)$ とする. $(x_1, \ldots, x_p)$ と y の間に以下の関係が成り立つと仮定する:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon \tag{2.1}$$

これは, **線形回帰モデル** (linear regression model) とよばれ, 各項は重回帰の場合と同様によばれる. 特に, 説明変数 $(x_1, \ldots, x_p)$ が p 個あり, それに伴って, 回帰係数も切片項 β_0 に加え, $\beta_1, \ldots, \beta_p$ となっている.

線形回帰モデルは入力を $x_1, \ldots, x_p$ とし, 出力を y とする (線形) システムという観点から説明すると以下の図のように理解できる:

$$\overset{\text{Input}}{x_1, \ldots, x_p} \rightarrow \overbrace{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p}^{\text{Linear System}} \rightarrow \overset{\text{Output}}{y}$$

$$\uparrow$$

$$\underset{\epsilon}{\text{Error}}$$

図 2.1: 線形システムと線形回帰モデル

実際に, 変数のセット $(x_1, \ldots, x_p, y)$ に関する n 組の観測値 (多変量データ) $(x_{i1}, \ldots, x_{ip}, y_i)$ $(i = 1, \ldots, n)$ が与えれた場合は, 線形回帰モデル (2.1) は以下のように成分で表される:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i, \ i = 1, \ldots, n \tag{2.2}$$

線形回帰モデルの成分表現 (2.2) は表 2.1 のようにまとめることができる:

表 2.1: 線形回帰モデルの成分表現

i	y_i	$=$	β_0	$+$	$\beta_1 x_{i1}$	$+$	$\cdots$	$+$	$\beta_p x_{ip}$	$+$	ϵ_i
1	y_1	$=$	β_0	$+$	$\beta_1 x_{11}$	$+$	$\cdots$	$+$	$\beta_p x_{1p}$	$+$	ϵ_1
2	y_2	$=$	β_0	$+$	$\beta_1 x_{21}$	$+$	$\cdots$	$+$	$\beta_p x_{2p}$	$+$	ϵ_2
$\vdots$							$\vdots$				
n	y_n	$=$	β_0	$+$	$\beta_1 x_{n1}$	$+$	$\cdots$	$+$	$\beta_p x_{np}$	$+$	ϵ_n
ベクトル表現	$\boldsymbol{y}$	$=$	$\beta_0\mathbf{1}$	$+$	$\beta_1\boldsymbol{x}_1$	$+$	$\cdots$	$+$	$\beta_p\boldsymbol{x}_p$	$+$	$\boldsymbol{\epsilon}$

最小自乗法と最小自乗推定値

線形回帰モデル (2.2) における回帰係数 $\beta_0, \beta_1, \ldots, \beta_p$ の最小自乗法による推定を考える. すなわち, 誤差平方和

$$\Delta^2(\beta_0, \beta_1, \ldots, \beta_p) := \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip})^2$$

を $\beta_0, \beta_1, \ldots, \beta_p$ に関して最小化する:

最小自乗法の定式化

$$\Delta^2(\beta_0, \beta_1, \ldots, \beta_p) \longrightarrow \min_{\beta_0, \beta_1, \ldots, \beta_p}$$

回帰係数で誤差平方和を偏微分したものをゼロとおくことによって,

$$\begin{cases} \dfrac{\partial \Delta^2(\beta_0, \beta_1, \ldots, \beta_k)}{\partial \beta_0} &= -2\displaystyle\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip}) = 0 \\ \dfrac{\partial \Delta^2(\beta_0, \beta_1, \ldots, \beta_k)}{\partial \beta_1} &= -2\displaystyle\sum_{i=1}^{n} x_{i1}(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip}) = 0 \\ & \vdots \\ \dfrac{\partial \Delta^2(\beta_0, \beta_1, \ldots, \beta_p)}{\partial \beta_p} &= -2\displaystyle\sum_{i=1}^{n} x_{ip}(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip}) = 0 \end{cases}$$

となり, 以下の正規方程式を得る:

$$\begin{cases} n\beta_0 + \left(\displaystyle\sum_{i=1}^{n} x_{i1}\right)\beta_1 + \cdots + \left(\displaystyle\sum_{i=1}^{n} x_{ip}\right)\beta_p = \displaystyle\sum_{i=1}^{n} y_i \\ \left(\displaystyle\sum_{i=1}^{n} x_{i1}\right)\beta_0 + \left(\displaystyle\sum_{i=1}^{n} x_{i1}^2\right)\beta_1 + \cdots + \left(\displaystyle\sum_{i=1}^{n} x_{i1}x_{ip}\right)\beta_p = \displaystyle\sum_{i=1}^{n} x_{i1}y_i \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \quad \vdots \\ \left(\displaystyle\sum_{i=1}^{n} x_{ip}\right)\beta_0 + \left(\displaystyle\sum_{i=1}^{n} x_{ip}x_{i1}\right)\beta_1 + \cdots + \left(\displaystyle\sum_{i=1}^{n} x_{ip}^2\right)\beta_p = \displaystyle\sum_{i=1}^{n} x_{ip}y_i \end{cases}$$

この方程式の解は,

$$\begin{cases} \beta_0 = \overline{y} - \widehat{\boldsymbol{\beta}}_p' \overline{\boldsymbol{x}}_p =: \widehat{\beta}_0 \\ \boldsymbol{\beta}_p = \left(\widetilde{\mathbf{X}}_p' \widetilde{\mathbf{X}}_p\right)^{-1} \widetilde{\mathbf{X}}_p' \boldsymbol{y} =: \widehat{\boldsymbol{\beta}}_p \end{cases}$$

によって与えられ, 最小自乗推定値とよばれる. ここで,

$$\boldsymbol{\beta}_p := \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \widehat{\boldsymbol{\beta}}_p := \begin{bmatrix} \widehat{\beta}_1 \\ \vdots \\ \widehat{\beta}_p \end{bmatrix}, \quad \overline{\boldsymbol{x}}_p := \begin{bmatrix} \overline{x}_1 \\ \vdots \\ \overline{x}_p \end{bmatrix}, \quad \overline{x}_j := \frac{1}{n}\sum_{i=1}^{n} x_{ij},$$

であり,

$$\widetilde{\mathbf{X}}_p := \mathbf{X}_p - \mathbf{1}\overline{\boldsymbol{x}}_p' = \begin{bmatrix} x_{11} - \overline{x}_1 & \dots & x_{1p} - \overline{x}_p \\ \vdots & & \vdots \\ x_{n1} - \overline{x}_1 & \dots & x_{np} - \overline{x}_p \end{bmatrix}, \quad \boldsymbol{y} := \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

とおいた. さらに,

$$\mathbf{S}_p := \frac{1}{n}\widetilde{\mathbf{X}}_p'\widetilde{\mathbf{X}}_p = \begin{bmatrix} s_1^2 & \cdots & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{p1} & \cdots & s_p^2 \end{bmatrix}, \quad \boldsymbol{s}_{py} := \frac{1}{n}\widetilde{\mathbf{X}}_p'\boldsymbol{y} = \begin{bmatrix} s_{1y} \\ \vdots \\ s_{py} \end{bmatrix},$$

$$s_j^2 := \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2, \quad s_{jy} := \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)(y_i - \overline{y}), \quad j = 1, \dots, p,$$

$$s_{jk} := \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)(x_{ik} - \overline{x}_k), \quad j \neq k = 1, \dots, p,$$

とおくことによって,

$$\widehat{\boldsymbol{\beta}}_p = \mathbf{S}_p^{-1}\boldsymbol{s}_{py}$$

と表すことができる. これは, 単回帰の場合に回帰係数 β_1 (傾き) の最小自乗推定値が $\widehat{\beta}_1 = s_{xy}/s_x^2 (= (s_x^2)^{-1}s_{xy})$ であることのアナロジーである.

標本回帰超平面, 当てはめ値, 残差

最小自乗推定値 $\widehat{\beta}_0, \widehat{\beta}_1, \dots, \widehat{\beta}_p$ を係数とする超平面 (hyper plane)

$$y = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \cdots + \widehat{\beta}_p x_p$$

は**標本回帰超平面** (sample regression hyper plane) とよばれ, 説明変数の観測値の組 $(x_{i1}, \dots, x_{ip})$ における超平面上の値を当てはめ値とよぶ:

$$\widehat{y}_i := \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \cdots + \widehat{\beta}_p x_{ip}$$

当てはめ値と応答変数の観測値の差を残差とよぶ:

$$\begin{aligned} e_i &:= y_i - \widehat{y}_i \\ &= y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \cdots + \widehat{\beta}_p x_{ip}) \end{aligned}$$

応答変数の観測値 y_i と当てはめ値 $\widehat{y}_i$, そして残差 e_i の間には以下の関係が成り立つ:

$$\begin{aligned} y_i &= \widehat{y}_i + e_i \\ &= \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \cdots + \widehat{\beta}_p x_{ip} + e_i \end{aligned}$$

最小自乗推定量の性質

線形回帰モデルの成分表現は,

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

と表すことでき, 誤差の正規性

$$\varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0, \sigma^2) \quad (i = 1, \ldots, n) \tag{2.3}$$

のもとで, 応答変数 Y_i は, 独立に以下の正規分布に従う:

$$Y_i \overset{\text{ind.}}{\sim} \mathsf{N}(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}, \sigma^2), \quad i = 1, \ldots, n$$

この仮定のもとで最小自乗推定量は,

$$\begin{cases} \widehat{\beta}_{0n} := \overline{Y}_n - \widehat{\boldsymbol{\beta}}'_{pn} \overline{\boldsymbol{x}}_p, \\ \widehat{\boldsymbol{\beta}}_{pn} := \left(\widetilde{\mathbf{X}}'_p \widetilde{\mathbf{X}}_p \right)^{-1} \widetilde{\mathbf{X}}'_p \boldsymbol{Y} = \mathbf{S}_p^{-1} \boldsymbol{s}_{pY} \end{cases}$$

となる. ここで,

$$\boldsymbol{Y} := \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad \widehat{\boldsymbol{\beta}}_{pn} := \begin{bmatrix} \widehat{\beta}_{1n} \\ \vdots \\ \widehat{\beta}_{pn} \end{bmatrix}, \quad \boldsymbol{s}_{pY} := \frac{1}{n} \widetilde{\mathbf{X}}'_p \boldsymbol{Y} = \begin{bmatrix} s_{1Y} \\ \vdots \\ s_{pY} \end{bmatrix}, \quad s_{jY} := \frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \overline{x}_j)(Y_i - \overline{Y}_n)$$

とおいた.

最小自乗推定量は, 確率変数であり回帰係数を理論的に推定することになる:

$$\beta_0 \overset{\text{推定}}{\longleftarrow} \widehat{\beta}_{0n}, \quad \boldsymbol{\beta}_p \overset{\text{推定}}{\longleftarrow} \widehat{\boldsymbol{\beta}}_{pn} \quad \text{(理論的)}$$

よって, これらの推定量の精度を評価するためには分布を考察する必要がある. 誤差に関する仮定 (2.3) のもとで最小自乗推定量に関して以下の命題 (性質) が成り立つ:

(LSL-1) 最小自乗推定量 $\widehat{\beta}_{0n}, \widehat{\boldsymbol{\beta}}_{pn}$ の同時分布は以下のように与えられる:

$$\begin{bmatrix} \widehat{\beta}_{0n} \\ \widehat{\boldsymbol{\beta}}_{pn} \end{bmatrix} \sim N_{p+1} \left(\begin{bmatrix} \beta_0 \\ \boldsymbol{\beta}_p \end{bmatrix}, \begin{bmatrix} \dfrac{\sigma^2}{n} \left(1 + \overline{\boldsymbol{x}}'_p \mathbf{S}_p^{-1} \overline{\boldsymbol{x}}_p \right) & -\dfrac{\sigma^2}{n} \overline{\boldsymbol{x}}'_p \mathbf{S}_p^{-1} \\ -\dfrac{\sigma^2}{n} \mathbf{S}_p^{-1} \overline{\boldsymbol{x}}_p & \dfrac{\sigma^2}{n} \mathbf{S}_p^{-1} \end{bmatrix} \right) \quad \text{(多変量正規性)} \tag{2.4}$$

ここで,

$$\begin{aligned} \mathrm{V}(\widehat{\beta}_{jn}) &= \frac{\sigma^2}{n} c_j, \quad j = 0, 1, \ldots, p \quad \text{(分散)} \\ \mathrm{Cov}(\widehat{\beta}_{0n}, \widehat{\boldsymbol{\beta}}_{pn}) &= -\frac{\sigma^2}{n} \overline{\boldsymbol{x}}'_p \mathbf{S}_p^{-1}, \quad \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{pn}, \widehat{\beta}_{0n}) = -\frac{\sigma^2}{n} \mathbf{S}_p^{-1} \overline{\boldsymbol{x}}_p \quad \text{(共分散ベクトル)} \\ \mathrm{V}(\widehat{\boldsymbol{\beta}}_{pn}) &= \frac{\sigma^2}{n} \mathbf{S}_p^{-1} \quad \text{(分散共分散行列)} \end{aligned}$$

であり,

$$c_j := \begin{cases} 1 + \overline{\boldsymbol{x}}'_p \mathbf{S}_p^{-1} \overline{\boldsymbol{x}}_p, & (j = 0) \\ \dfrac{1}{s_j^2 (1 - R_j^2)}, & (j = 1, \ldots, p) \end{cases} \tag{2.5}$$

とおいた. ただし,

$$R_j^2 := \boldsymbol{z}_j' \mathbf{P}_{(j)} \boldsymbol{z}_j, \quad j = 1, \ldots, p$$

であり, ここで標準化説明変数行列:

$$\begin{aligned} \mathbf{Z}_p &:= [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_p] \\ &:= \begin{bmatrix} z_{11} & \cdots & z_{1p} \\ \vdots & & \vdots \\ z_{n1} & \cdots & z_{np} \end{bmatrix} := \begin{bmatrix} \frac{x_{11}-\overline{x}_1}{\sqrt{n}s_1} & \cdots & \frac{x_{1p}-\overline{x}_p}{\sqrt{n}s_p} \\ \vdots & & \vdots \\ \frac{x_{n1}-\overline{x}_1}{\sqrt{n}s_1} & \cdots & \frac{x_{np}-\overline{x}_p}{\sqrt{n}s_p} \end{bmatrix} \end{aligned}$$

に対して,

$$\mathbf{Z}_{(j)} := [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{j-1}, \boldsymbol{z}_{j+1}, \ldots, \boldsymbol{z}_p], \quad \mathbf{P}_{(j)} := \mathbf{Z}_{(j)} (\mathbf{Z}_{(j)}' \mathbf{Z}_{(j)})^{-1} \mathbf{Z}_{(j)}'$$

とおいた.

(LSL-2) 最小自乗推定量の周辺分布は以下のように与えられる:

$$\widehat{\beta}_{jn} \sim N\left(\beta_j, \frac{\sigma^2}{n} c_j\right), \quad j = 0, 1, \ldots, p \quad \text{(正規性)} \tag{2.6}$$

(LSL-3) 最小自乗推定量の平均 (期待値) は以下のように与えられる:

$$\mathrm{E}(\widehat{\beta}_{jn}) = \beta_j, \quad j = 0, 1, \ldots, p \quad \text{(不偏性)}$$

(LSL-4) 最小自乗推定量の標準誤差は以下のように与えられる:

$$\mathrm{s.e.}(\widehat{\beta}_{jn}) := \sqrt{\frac{\sigma^2}{n} c_j}, \quad j = 0, 1, \ldots, p$$

これらの命題のうち, (LSL-1) からその他のものは導くことができる. また, (LSL-2), (LSL-3), (LSL-4) の結果から,

$$\frac{\widehat{\beta}_{jn} - \mathrm{E}(\widehat{\beta}_{jn})}{\mathrm{s.e.}(\widehat{\beta}_{jn})} \sim \mathsf{N}(0, 1), \quad j = 0, 1, \ldots, p \tag{2.7}$$

となり, 最小自乗推定量を標準化することができる.

誤差分散の推定

未知の誤差分散 σ^2 は残差

$$\widehat{\varepsilon}_i := Y_i - \widehat{\beta}_{0n} - \widehat{\beta}_{1n} x_{i1} - \cdots - \widehat{\beta}_{pn} x_{ip}, \quad i = 1, \ldots, n$$

の平方和

$$\Delta^2(\widehat{\beta}_{0n}, \widehat{\beta}_{1n}, \ldots, \widehat{\beta}_{pn}) := \sum_{i=1}^n \widehat{\varepsilon}_i^2 = \sum_{i=1}^n (Y_i - \widehat{\beta}_{0n} - \widehat{\beta}_{1n} x_{i1} - \cdots - \widehat{\beta}_{pn} x_{ip})^2$$

を利用して, 以下のように推定量を定義する:

$$\widehat{\sigma}_n^2 := \frac{1}{n-p-1} \Delta^2(\widehat{\beta}_{0n}, \widehat{\beta}_{1n}, \ldots, \widehat{\beta}_{pn}) = \frac{1}{n-p-1} \sum_{i=1}^n \widehat{\varepsilon}_i^2 \tag{2.8}$$

この推定量に関して, 以下の命題が成り立つ:

(EVL-1)

$$\mathrm{E}(\widehat{\sigma}_n^2) = \sigma^2 \quad \text{(不偏性)}$$

(EVL-2)

$$\frac{(n-p-1)\widehat{\sigma}_n^2}{\sigma^2} \sim \chi^2_{n-p-1} \quad \text{(自由度 } n-p-1 \text{ のカイ自乗分布)}$$

(EVL-3)

$$\widehat{\beta}_{jn} \perp\!\!\!\perp \widehat{\sigma}_n^2 \quad j = 0, 1, \ldots, p \text{ (独立性)}$$

また, 実際に誤差分散を数値として推定するためには線形回帰モデルをデータに対して当てはめた結果として得られる残差

$$e_i = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_{i1} - \cdots - \widehat{\beta}_p x_{ip}, \quad i = 1, \ldots, n$$

を使って計算された残差平方和を自由度 $n-p-1$ で割った

$$\widehat{\sigma}^2 := \frac{1}{n-p-1}\Delta^2(\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_p) = \frac{1}{n-p-1}\sum_{i=1}^{n} e_i^2$$

を誤差分散の推定値として用いる. また, この推定値を利用することによって, 最小自乗推定量の標準誤差を以下のように数値的に推定することができる:

$$\widehat{\mathrm{s.e.}}(\widehat{\beta}_{jn}) := \sqrt{\frac{\widehat{\sigma}^2}{n} c_j}$$

これらの推定量を最小自乗推定量の標準化 (2.7) における標準誤差の部分に代入し, 誤差分散に関する命題 (EVL-2), (EVL-3) を利用すると,

$$\frac{\widehat{\beta}_{jn} - \mathrm{E}(\widehat{\beta}_{jn})}{\widehat{\mathrm{s.e.}}_n(\widehat{\beta}_{jn})} = \frac{\widehat{\beta}_{jn} - \beta_j}{\widehat{\sigma}_{jn}} = \frac{\widehat{\beta}_{jn} - \beta_j}{\sqrt{\frac{\sigma^2}{n} c_j}} \Bigg/ \sqrt{\frac{\widehat{\sigma}_n^2}{\sigma^2}} \sim \frac{\mathsf{N}(0,1)}{\sqrt{\chi^2_{n-p-1}/(n-p-1)}} \stackrel{\mathrm{d}}{=} \mathsf{t}_{n-p-1} \tag{2.9}$$

となり, これらの標準化に関する統計量は共に自由度 $n-p-1$ のティー分布に従うことがわかる.

回帰係数の検定

回帰係数 β_j $(j = 0, 1, \ldots, p)$ に対する以下の仮説を検定することを考えよう:

$$\begin{cases} H_0: & \beta_j = 0 \\ H_1: & \beta_j \neq 0 \end{cases}, \quad j = 0, 1, \ldots, p \tag{2.10}$$

有意水準を α とし, これらの検定に関しては以下の検定統計量 (ティー統計量) とその実現値 (ティー値) がそれぞれ利用される:

$$T_{jn} := \frac{\widehat{\beta}_{jn}}{\widehat{\sigma}_{jn}}, \quad t_j := \frac{\widehat{\beta}_j}{\widehat{\sigma}_j}, \quad j = 0, 1, \ldots, p$$

これらの検定統計量は, 標準化された統計量の分布に関する結果 (2.9) から, 帰無仮説 H_0 のもとで以下のように自由度 $n-p-1$ のティー分布に従う:

$$T_{jn} \sim \mathsf{t}_{n-p-1}, \quad j = 0, 1, \ldots, p$$

棄却域は, 自由度 $n-p-1$ のティー分布の上側 $100(\alpha/2)\%$ 点を $t_{n-p-1}(\alpha/2)$ とすると,

$$R = (-\infty, -t_{n-p-1}(\alpha/2)] \quad \text{または} \quad [t_{n-p-1}(\alpha/2), \infty)$$

であり, 検定は検定統計量の実現値 t_j $(j = 0, 1, \ldots, p)$ の値に応じて以下のように実行される:

$$t_j \in R \Longrightarrow \text{帰無仮説 } H_0 \text{ を棄却},$$
$$t_j \notin R \Longrightarrow \text{帰無仮説 } H_0 \text{ を棄却できない (採択)}$$

回帰係数ベクトルの検定

切片項 β_0 以外の回帰係数のベクトルを

$$\boldsymbol{\beta}_p := \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

で表し, 仮説

$$\begin{cases} H_0 : \boldsymbol{\beta}_p = \mathbf{0} \\ H_1 : \boldsymbol{\beta}_p \neq \mathbf{0} \end{cases}$$

を検定することを考える. ここで $\mathbf{0} := [0, \ldots, 0]'$ は (p 次元) ゼロベクトル (列ベクトル) である. この検定は, 平方和の分解によって, 全体の変動を回帰によって説明される変動と説明されない変動に分解し, それらを比較することによって行われる. 平方和の分解を観測 Y_i で置き換えたものに関して帰無仮説 $H_0 : \boldsymbol{\beta}_p = \mathbf{0}$ のもとで以下のような分布に関する特性が成り立つ:

$$\begin{array}{ccccc} \dfrac{\sum_{i=1}^n (Y_i - \overline{Y})^2}{\sigma^2} & = & \dfrac{\sum_{i=1}^n (\widehat{Y}_i - \overline{Y})^2}{\sigma^2} & + & \dfrac{\sum_{i=1}^n (Y_i - \widehat{Y}_i)^2}{\sigma^2} \\ \wr & & \wr & & \wr \\ \chi^2_{n-1} & = & \chi^2_p & + & \chi^2_{n-p-1} \end{array}$$

この結果の右辺第 1 項と第 2 項が独立であることから, 検定統計量 (エフ統計量) に関して以下の結果が成り立つ:

$$F_{pn} := \frac{\displaystyle\sum_{i=1}^n (\widehat{Y}_i - \overline{Y})^2 / p}{\displaystyle\sum_{i=1}^n (Y_i - \widehat{Y}_i)^2 / (n-p-1)} = \frac{\left(\dfrac{\sum_{i=1}^n (\widehat{Y}_i - \overline{Y})^2}{\sigma^2} \right) \Big/ p}{\left(\dfrac{\sum_{i=1}^n (Y_i - \widehat{Y}_i)^2}{\sigma^2} \right) \Big/ (n-p-1)}$$

$$\sim \frac{\chi^2_p / p}{\chi^2_{n-p-1} / (n-p-1)} \stackrel{\mathrm{d}}{=} \mathsf{F}^p_{n-p-1} \quad (; \text{自由度 } (p, n-p-1) \text{ のエフ分布}) \tag{2.11}$$

この結果を利用し, 有意水準 α に対して棄却域を

$$R = \left[F^p_{n-p-1}(\alpha), \infty \right)$$

とし, 検定統計量 F_{pn} の実現値 f_p の値に応じて以下のように検定が実行される:

$$f_p \in R \Longrightarrow \text{帰無仮説 } H_0 \text{ を棄却},$$
$$f_p \notin R \Longrightarrow \text{帰無仮説 } H_0 \text{ を棄却できない (採択)}$$

ここで, $F^{p}_{n-p-1}(\alpha)$ は自由度 $(p, n-p-1)$ のエフ分布の上側 $100\alpha\%$ 点である.

以上の検定を行う際に利用される分散分析表は以下のように与えられる:

表 2.2: 回帰分析における分散分析表

変動要因	自由度	平方和	平均平方和	エフ比
回帰	p	$\sum_{i=1}^{n}(\widehat{y}_i-\overline{y})^2$	$\sum_{i=1}^{n}(\widehat{y}_i-\overline{y})^2/p$	$\frac{\sum_{i=1}^{n}(\widehat{y}_i-\overline{y})^2/p}{\sum_{i=1}^{n}(y_i-\widehat{y}_i)^2/(n-p-1)}$
残差	$n-p-1$	$\sum_{i=1}^{n}(y_i-\widehat{y}_i)^2$	$\sum_{i=1}^{n}(y_i-\widehat{y}_i)^2/(n-p-1)$	
全体	$n-1$	$\sum_{i=1}^{n}(y_i-\overline{y})^2$		

第 5 列は, エフ比

$$f_p := \frac{\sum_{i=1}^{n}(\widehat{y}_i-\overline{y})^2/p}{\sum_{i=1}^{n}(y_i-\widehat{y}_i)^2/(n-p-1)} \tag{2.12}$$

である.

モデルの当てはまりの程度をみるための指標

当てはまりの程度をみるための指標に関しては, 単回帰, 重回帰の場合と同様であるが, 自由度調整済み決定係数が以下で与えられる:

$$\begin{aligned}\overline{R}^2 &= 1-\frac{\sum_{i=1}^{n}(y_i-\widehat{y}_i)^2/(n-p-1)}{\sum_{i=1}^{n}(y_i-\overline{y})^2/(n-1)}\\ &= 1-\frac{\mathrm{RSS}/(n-p-1)}{\mathrm{TSS}/(n-1)}\end{aligned}$$

2.2 最小自乗法の幾何学的説明

これまでは, 回帰分析と最小自乗法をできる限り「成分」で表現してきたけれども, ここからは線形代数の知識を利用し, 空間的な概念を導入することによって幾何学的な説明を行う.

線形回帰モデルの成分表現 (2.2) を行列とベクトルを使って, 以下のように表すことができる:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2.13}$$

ここで,

$$\boldsymbol{y} := \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},\ \mathbf{X} := \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix},\ \boldsymbol{\beta} := \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix},\ \boldsymbol{\epsilon} := \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

であり, 各項は以下のようによばれる:

$\boldsymbol{y}$; **応答変数ベクトル** (vector of response variable)
$\mathbf{X}$; **説明変数行列** (matrix of explanatory variables)
$\boldsymbol{\beta}$; **回帰係数ベクトル** (vector regression coefficients)
$\boldsymbol{\epsilon}$; **誤差ベクトル** (vector of error)

ここで, 説明変数行列はフルランクとする. すなわち, $\mathrm{rank}(\mathbf{X}) = p+1(\leq n)$ が成り立つものとする [1)].

最小自乗法は誤差平方和

$$\Delta^2(\boldsymbol{\beta}) := \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})'(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})$$

を回帰係数ベクトル $\boldsymbol{\beta}$ に関して最小化することによって以下のように定式化される:

最小自乗法の定式化 (行列・ベクトル表現)

$$\Delta^2(\boldsymbol{\beta}) \longrightarrow \min_{\boldsymbol{\beta}}$$

この最小化問題は極値問題に置き換えることによって解かれる. すなわち, 残差平方和を回帰係数ベクトル $\boldsymbol{\beta}$ によって微分したものを $\mathbf{0}$ とおき,

$$\frac{\partial \Delta^2(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}$$

より導かれる方程式 (正規方程式)

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\boldsymbol{y}$$

を $\boldsymbol{\beta}$ について解くことによって,

$$\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} =: \widehat{\boldsymbol{\beta}}$$

が得られ, 解 $\widehat{\boldsymbol{\beta}}$ を **最小自乗推定値ベクトル** (vector of least square estimates) とよぶ.

では, 線形代数の基本的な知識, 特に**射影** (projection) を使って最小自乗法を幾何学的に説明しよう. まず当てはめ値をベクトルとして表したものは**回帰ベクトル** (regression vector) とよばれ, 以下のように定義される:

$$\widehat{\boldsymbol{y}} := \mathbf{X}\widehat{\boldsymbol{\beta}}$$

このベクトルがどのような性質を持つかということを明らかにするために, 説明変数行列の列ベクトルによる分割 $\mathbf{X} = [\mathbf{1}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_p]$ に注意して, これらのベクトルの線形結合で構成される以下の空間を定義する:

$$\mathcal{M}(\mathbf{X}) := \left\{\mathbf{X}\boldsymbol{\xi} \mid \boldsymbol{\xi} \in \mathbb{R}^{p+1}\right\} = \left\{\xi_0\mathbf{1} + \xi_1\boldsymbol{x}_1 + \cdots + \xi_p\boldsymbol{x}_p \mid \xi_0, \xi_1, \ldots, \xi_p \in \mathbb{R}\right\} (\subset \mathbb{R}^n)$$

この空間を行列 $\mathbf{X}$ の列ベクトルによって張られる空間 (spaned space by column vectors) とよぶ. 回帰ベクトル $\widehat{\boldsymbol{y}}$ は

$$\widehat{\boldsymbol{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} = \widehat{\beta}_0\mathbf{1} + \widehat{\beta}_1\boldsymbol{x}_1 + \cdots + \widehat{\beta}_p\boldsymbol{x}_p \in \mathcal{M}(\mathbf{X})$$

[1)] 記号 $\mathrm{rank}(\mathbf{A})$ は行列 $\mathbf{A}$ の**ランク** (rank) を表し, 行列のランクとは一次独立な列ベクトル (または行ベクトル) の個数を表す.

となり, 説明変数行列 $\mathbf{X}$ の列ベクトルによって張られる空間に属していることがわかる.

さらに, 以下で定義される行列を導入する:

$$\mathbf{P} := \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

この行列は以下の性質をもつ:

(P1) $\mathbf{P}' = \mathbf{P}$ (対称性)
(P2) $\mathbf{P}^2 = \mathbf{P}$ (巾等性)
(P3) $\mathbf{PX} = \mathbf{X}$ (不動性)

一般に, 行列が (P1) と (P2) の性質 (対称性と巾等性) をもつとき, **射影行列** (projection matrix) とよばれる [2].

射影行列 $\mathbf{P}$ を使って回帰ベクトル $\widehat{\boldsymbol{y}}$ が以下のように表される:

$$\mathbf{P}\boldsymbol{y} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} = \mathbf{X}\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{y}} \in \mathcal{M}(\mathbf{X})$$

この結果から, 射影行列 $\mathbf{P}$ は応答変数ベクトル $\boldsymbol{y}$ を説明変数行列 $\mathbf{X}$ の列ベクトルによって張られる空間 $\mathcal{M}(\mathbf{X})$ へ移す写像であることがわかる:

$$\begin{array}{rccc} \mathbf{P}: & \mathbb{R}^n & \longrightarrow & \mathcal{M}(\mathbf{X}) \\ & \cup & & \cup \\ & \boldsymbol{y} & \mapsto & \widehat{\boldsymbol{y}} \end{array}$$

一方, **残差ベクトル** (residual vector) は応答変数ベクトル $\boldsymbol{y}$ と回帰ベクトル $\widehat{\boldsymbol{y}}$ の差で定義される:

$$\boldsymbol{e} := \boldsymbol{y} - \widehat{\boldsymbol{y}} = \boldsymbol{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$$

このベクトルの性質を調べるために以下の空間を定義する:

$$\mathcal{M}(\mathbf{X})^{\perp} := \{\boldsymbol{\zeta} \mid \langle \boldsymbol{\zeta}, \boldsymbol{\eta} \rangle = 0,\ \boldsymbol{\eta} \in \mathcal{M}(\mathbf{X})\} = \{\boldsymbol{\zeta} \mid \boldsymbol{\zeta} \perp \boldsymbol{\eta},\ \boldsymbol{\eta} \in \mathcal{M}(\mathbf{X})\}$$

この空間は, $\mathcal{M}(\mathbf{X})$ の**直交補空間** (orthogonal complement) とよばれる. ここで, $\langle \boldsymbol{a}, \boldsymbol{b} \rangle := \sum_{i=1}^{n} a_i b_i = \boldsymbol{a}'\boldsymbol{b}$ はベクトル $\boldsymbol{a} := [a_1, \ldots, a_n]'$ と $\boldsymbol{b} := [b_1, \ldots, b_n]'$ の**内積** (inner product) であり, 内積が 0 となる場合を**直交** (orthogonal) するという. なお, $\boldsymbol{a} \perp \boldsymbol{b}$ はベクトルが直交することを表す記号である. よって, 直交補空間 $\mathcal{M}(\mathbf{X})^{\perp}$ に属している任意のベクトルは $\mathcal{M}(\mathbf{X})$ に属している任意のベクトルと直交している.

ここで, 以下のような行列を導入する:

$$\mathbf{Q} := \mathbf{I}_n - \mathbf{P} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

この行列は, 射影行列 $\mathbf{P}$ の性質 (P1), (P2), (P3) を利用することによって, 以下の性質を持つことがわかる:

(Q1) $\mathbf{Q}' = \mathbf{Q}$ (対称性)
(Q2) $\mathbf{Q}^2 = \mathbf{Q}$ (巾等性)
(Q3) $\mathbf{QX} = \mathbf{O}$ (直交性)

[2] 特に, 性質 (P2) (巾等性) が何らかの空間への射影することを意味する.

性質 (Q1), (Q2) から行列 $\mathbf{Q}$ も射影行列であることがわかる.

残差ベクトル $\boldsymbol{e}$ は射影行列 $\mathbf{Q}$ を使って以下のように表すことができる:

$$\boldsymbol{e} = \boldsymbol{y} - \mathbf{P}\boldsymbol{y} = (\mathbf{I}_n - \mathbf{P})\boldsymbol{y} = \mathbf{Q}\boldsymbol{y}$$

残差ベクトル $\boldsymbol{e}$ がどのような空間に属しているかを調べるために, 任意の $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\xi} \in \mathcal{M}(\mathbf{X})$ との内積を考えると, 性質 (Q1), (Q3) より,

$$\langle \boldsymbol{e}, \boldsymbol{\eta} \rangle = \boldsymbol{e}'\boldsymbol{\eta} = \boldsymbol{y}'\mathbf{Q}'\mathbf{X}\boldsymbol{\xi} = \boldsymbol{y}'\mathbf{Q}\mathbf{X}\boldsymbol{\xi} = 0$$

となり, 残差ベクトル $\boldsymbol{e}$ は直交補空間 $\mathcal{M}(\mathbf{X})^{\perp}$ に属していることがわかる:

$$\boldsymbol{e} = \mathbf{Q}\boldsymbol{y} \in \mathcal{M}(\mathbf{X})^{\perp}$$

よって, 射影行列 $\mathbf{Q}$ は応答変数ベクトル $\boldsymbol{y}$ を直交補空間 $\mathcal{M}(\mathbf{X})^{\perp}$ へ移す写像であることがわかる:

$$\begin{array}{rccc} \mathbf{Q}: & \mathbb{R}^n & \longrightarrow & \mathcal{M}(\mathbf{X})^{\perp} \\ & \cup & & \cup \\ & \boldsymbol{y} & \mapsto & \boldsymbol{e} \end{array}$$

回帰ベクトル $\widehat{\boldsymbol{y}}$ は説明変数行列の列ベクトルによって張られる空間 $\mathcal{M}(\mathbf{X})$ に属しているため, その直交補空間に属している残差ベクトル $\boldsymbol{e}$ と直交することが予想される. 実際,

$$\langle \widehat{\boldsymbol{y}}, \boldsymbol{e} \rangle = \boldsymbol{y}'\mathbf{P}'\mathbf{Q}\boldsymbol{y} = \boldsymbol{y}'\mathbf{P}(\mathbf{I}_n - \mathbf{P})\boldsymbol{y} = \boldsymbol{y}'(\mathbf{P} - \mathbf{P}^2)\boldsymbol{y} = 0$$

となり直交する ($\widehat{\boldsymbol{y}} \perp \boldsymbol{e}$) ことがわかる. ここで, 射影行列の性質 (P1), (P2) を利用している.

応答変数ベクトル $\boldsymbol{y}$ は回帰ベクトル $\widehat{\boldsymbol{y}}$ と残差ベクトル $\boldsymbol{e}$ の和で表すことができ, それぞれのベクトルが属している空間との関係と共に以下のように書かれる:

$$\begin{array}{ccccc} \mathbb{R}^n & = & \mathcal{M}(\mathbf{X}) & \oplus & \mathcal{M}(\mathbf{X})^{\perp} \\ \cup & & \cup & & \cup \\ \boldsymbol{y} & = & \widehat{\boldsymbol{y}} & + & \boldsymbol{e} \\ & = & \mathbf{P}\boldsymbol{y} & + & \mathbf{Q}\boldsymbol{y} \end{array}$$

ここで $\oplus$ は空間の**直和** (direct sum) を表す [3].

以上のことから, 回帰ベクトル $\widehat{\boldsymbol{y}}$ は応答変数ベクトル $\boldsymbol{y}$ を射影行列 $\mathbf{P}$ によって空間 $\mathcal{M}(\mathbf{X})$ へ垂直に射影 (正射影) しており, 残差ベクトル $\boldsymbol{e}$ は応答変数ベクトル $\boldsymbol{y}$ を射影行列 $\mathbf{Q}$ によって空間 $\mathcal{M}(\mathbf{X})^{\perp}$ へ正射影している.

さらに, 応答変数ベクトル $\boldsymbol{y}$ のノルムと回帰ベクトル $\widehat{\boldsymbol{y}}$ と残差ベクトル $\boldsymbol{e}$ のノルムの間には, 直交性を使うと以下の関係が成り立つ:

$$\begin{aligned} \|\boldsymbol{y}\|^2 &:= \langle \boldsymbol{y}, \boldsymbol{y} \rangle = \langle \widehat{\boldsymbol{y}} + \boldsymbol{e}, \widehat{\boldsymbol{y}} + \boldsymbol{e} \rangle \\ &= \|\widehat{\boldsymbol{y}}\|^2 + \|\boldsymbol{e}\|^2 + 2\langle \widehat{\boldsymbol{y}}, \boldsymbol{e} \rangle \\ &= \|\widehat{\boldsymbol{y}}\|^2 + \|\boldsymbol{e}\|^2 \end{aligned}$$

この関係は直角三角形に関するピタゴラスの定理 (三平方の定理) を n 次元へ拡張したものと見なすことができる.

---

[3] 一般に, 空間 X, Y が直和であるとは, それらの交わり部分が原点 O のみの場合をいう. すなわち, $X \cap Y = \{\mathrm{O}\}$. ここでは, より正確には「直交直和」である.

以上のことをまとめると, 最小自乗法とは, n 次元ユークリッド空間 $\mathbb{R}^n$ に属している応答変数ベクトル $\boldsymbol{y}$ を射影行列 $\mathbf{P}$ によって空間 $\mathcal{M}(\mathbf{X})$ へ正射影 (回帰) することによって, 空間 $\mathcal{M}(\mathbf{X})$ の中で応答変数ベクトル $\boldsymbol{y}$ に最も距離が近いベクトル $\widehat{\boldsymbol{y}} = \mathbf{P}\boldsymbol{y} = \mathbf{X}\widehat{\boldsymbol{\beta}}$ を見つけ, その係数ベクトル (最小自乗推定値ベクトル) $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$ で未知の回帰係数ベクトル $\boldsymbol{\beta}$ を推定する手法であると幾何学的に説明できる [4]. なお, 図 2.2 は最小自乗法を幾何学的に説明するためのイメージを与えている.

また, 回帰分析とは, 幾何学的には, n 次元ユークリッド空間 $\mathbb{R}^n$ を二つの直交する空間 $\mathcal{M}(\mathbf{X})$, $\mathcal{M}(\mathbf{X})^{\perp}$ に直和分解し, それぞれの射影されたベクトル $\widehat{\boldsymbol{y}}$ と $\boldsymbol{e}$ のノルム (の 2 乗) を比較 (分散分析) することによって説明変数行列 $\mathbf{X}$ が応答変数ベクトル $\boldsymbol{y}$ をどれだけ「説明」できるかを考察する手法といえる [5].

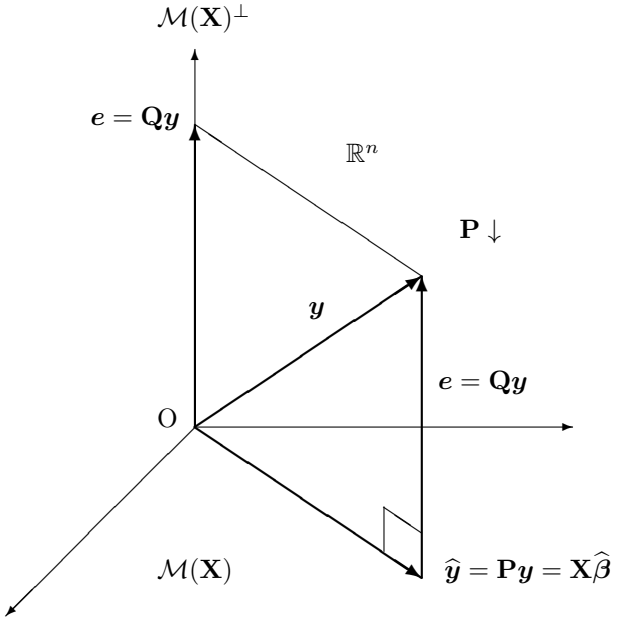

図 2.2: 最小自乗法の幾何学的説明

注意 2.1 付録 1.2 で扱った分散公式の幾何学的説明は, データを列ベクトル $\boldsymbol{x} := [x_1, \ldots, x_n]'$ で表したとき, $\boldsymbol{x} = \beta\mathbf{1} + \boldsymbol{\epsilon}$ なる線形回帰モデルに最小自乗法を適用した場合と理解できる. ただし, $\mathbf{1} = [1, \ldots, 1]' \in \mathbb{R}^n$, $\beta \in \mathbb{R}$ であり, $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_n]'$ は誤差ベクトルである.

2.3 最小自乗推定量と誤差分散の推定量の性質

誤差を確率変数と考え, ベクトル化したもの $\boldsymbol{\varepsilon} := [\varepsilon_1, \ldots, \varepsilon_n]'$ を考えると, 誤差の正規性に関する仮定 (2.3) は以下のように表すことができる:

$$\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$$

[4] 影絵遊びのように「もの」に「光源」から光を当てることによって「スクリーン」に影を写すことを想像すると, 「もの」に対応するものが「応答変数ベクトル $\boldsymbol{y}$」であり, 「光源」に対応するものが「射影行列 $\mathbf{P}$ 」である. さらに, 「スクリーン」に対応するものが「空間 $\mathcal{M}(\mathbf{X})$ 」である.

[5] 回帰分析における実際の分散分析は, 応答変数ベクトルの偏差 (平均を引いたもの) に関する平方和の分解に関して行われる.

このとき, 線形回帰モデル

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

における応答ベクトル $\boldsymbol{Y}$ は以下の分布に従う:

$$\boldsymbol{Y} \sim \mathsf{N}_{p+1}\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n\right)$$

この結果から, 最小自乗推定量ベクトル:

$$\widehat{\boldsymbol{\beta}}_n = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{Y}$$

は以下の分布に従う:

$$\widehat{\boldsymbol{\beta}}_n \sim N_{p+1}\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right)$$

残差ベクトル

$$\widehat{\boldsymbol{\varepsilon}} := \boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_n$$

に関する平方和

$$\Delta^2(\widehat{\boldsymbol{\beta}}_n) := \widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}} = (\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_n)'(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_n)$$

を利用した誤差分散の推定量は

$$\widehat{\sigma}_n^2 := \frac{1}{n-p-1}\Delta^2(\widehat{\boldsymbol{\beta}}_n) = \frac{1}{n-p-1}\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}$$

と表すことができる. これは, 誤差分散の推定量 (2.8) を残差ベクトルによって表現したものであり, 性質 (EVL-1), (EVL-2), (EVL-3) が成り立つことも思い出そう.