第 III 部　線型モデル

4　直線の当てはまりの程度

ポイント

- 残差平方和
- 説明された平方和
- 総平方和
- 決定係数

4.1　はじめに

- 一番よく直線は当てはまりました。
- 一番よく当てはまってますがどの程度当てはまっているのでしょうか？
- 規準は何でしょう？

4.1.1 総和の公式

$$\sum_{i=1}^{n} c = nc \qquad ; \ c \text{ は定数} \tag{1}$$

$$\sum_{i=1}^{n} ax_i = a\sum_{i=1}^{n} x_i \qquad ; \ a \text{ は定数} \tag{2}$$

$$\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i \tag{3}$$

4.1.2 平均と総和

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{4}$$

$$n\bar{x} = \sum_{i=1}^{n} x_i \tag{5}$$

4.2　最小二乗法による推定値

4.2.1　最小二乗法

　もとめる直線を

$$\hat{y}_i = \alpha + \beta x_i \tag{6}$$

とする。ここで誤差を

$$e_i = y_i - \hat{y}_i \tag{7}$$

とする。このとき

$$\sum_{i=1}^{n} {e_i}^2 \tag{8}$$

を最小化するように α，β を求めることを最小二乗法という。

4.2.2 誤差の二乗和

(8) 式に (7) 式ならびに (6) 式を代入して整理すると誤差の二乗和の定義式を得る。

$$\sum_{i=1}^{n} {e_i}^2 = n\alpha^2 + 2\alpha\beta\sum_{i=1}^{n} x_i + \beta^2\sum_{i=1}^{n} {x_i}^2 - 2\alpha\sum_{i=1}^{n} y_i - 2\beta\sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} {y_i}^2 \tag{9}$$

誤差の二乗和 (9) は α, β の関数である。

4.2.3 誤差の二乗和の偏導関数

誤差の二乗和 (9) を α , β で偏微分し偏導関数を求める。
偏導関数は、それぞれのパラメータ軸における原始関数の傾きを返す関数である。

$$\frac{\partial f}{\partial \alpha} = 2n\alpha + 2\beta \sum_{i=1}^{n} x_i - 2 \sum_{i=1}^{n} y_i \tag{10}$$

$$\frac{\partial f}{\partial \beta} = 2\alpha \sum_{i=1}^{n} x_i + 2\beta \sum_{i=1}^{n} {x_i}^2 - 2 \sum_{i=1}^{n} x_i y_i \tag{11}$$

4.2.4　偏導関数を 0 と置いた連立方程式

　誤差の二乗和 (9) は非負なので下限をもつ。

したがって極値は最小値となる。

極値を求めるには偏導関数を 0 とおいた連立方程式を解けばよい。

$$
\begin{cases}
n\alpha + \beta \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i = 0 & (12) \\
\alpha \sum_{i=1}^{n} x_i + \beta \sum_{i=1}^{n} {x_i}^2 - \sum_{i=1}^{n} x_i y_i = 0 & (13)
\end{cases}
$$

4.3 β の推定値

連立方程式を解き、得られた推定値 $\hat{\beta}$ が (14) 式である。

$$\hat{\beta} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} {x_i}^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{14}$$

n で約分する

$$= \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} {x_i}^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} \tag{15}$$

分母分子それぞれの第二項に $n\frac{1}{n}$ を掛けると

$$\hat{\beta} = \frac{\displaystyle\sum_{i=1}^{n} x_i y_i - n\frac{1}{n}\frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\displaystyle\sum_{i=1}^{n} {x_i}^2 - n\frac{1}{n}\frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\displaystyle\sum_{i=1}^{n} x_i y_i - n\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)}{\displaystyle\sum_{i=1}^{n} {x_i}^2 - n\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)} \tag{16}$$

最右辺分母分子第二項をそれぞれ平均で置き換えると

$$\hat{\beta} = \frac{\displaystyle\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\displaystyle\sum_{i=1}^{n} {x_i}^2 - n\bar{x}\bar{x}} = \frac{\displaystyle\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\displaystyle\sum_{i=1}^{n} {x_i}^2 - n\bar{x}^2} \tag{17}$$

4.3.1 $\hat{\alpha}$ と $\hat{\beta}$ の関係

(12) 式より

$$n\alpha + \hat{\beta}\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i = 0 \tag{18}$$

左辺第二項、第三項を右辺へ移項

$$n\alpha = \sum_{i=1}^{n} y_i - \hat{\beta}\sum_{i=1}^{n} x_i \tag{19}$$

n を払い、$\hat{\alpha}$ を得る。

$$\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{\beta}\frac{1}{n}\sum_{i=1}^{n} x_i \tag{20}$$

右辺各項を平均で置き換え

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \tag{21}$$

4.4　平均からの偏差の積の総和

4.4.1　S_{xy}

x と y のそれぞれの『平均からの偏差』を掛け合わせた値の総和を S_{xy} という記号であらわす。

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \tag{22}$$

問題 III–4 – 1

次の式を展開しなさい。

$$\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)$$

解例 III–4 − 1

$$
\begin{aligned}
\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) &= \sum_{i=1}^{n}(x_iy_i-\bar{x}y_i-\bar{y}x_i+\bar{x}\bar{y}) \\
&= \sum_{i=1}^{n}x_iy_i-\bar{x}\sum_{i=1}^{n}y_i-\bar{y}\sum_{i=1}^{n}x_i+n\bar{x}\bar{y} \\
&= \sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}-\cancel{n\bar{y}\bar{x}}+\cancel{n\bar{x}\bar{y}} \\
&= \sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}
\end{aligned}
$$

4.4.2　S_{xx} と S_{yy}

同様に、x の『平均からの偏差』の二乗の総和を S_{xx} という記号であらわす。

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{23}$$

また、y の『平均からの偏差』の二乗の総和を S_{yy} という記号であらわす。

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{24}$$

問題 III–4 – 2

次の式を展開しなさい。

(1)

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

(2)

$$\sum_{i=1}^{n}(y_i-\bar{y})^2$$

解例 III–4 − 2 (1)

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

解例 III–4 − 2 (2)

$$
\begin{aligned}
&\sum_{i=1}^{n}(y_i-\bar{y})^2\\
&=\sum_{i=1}^{n}\left({y_i}^2-2\bar{y}y_i+\bar{y}^2\right)\\
&=\sum_{i=1}^{n}{y_i}^2-2\bar{y}\sum_{i=1}^{n}y_i+n\bar{y}^2\\
&=\sum_{i=1}^{n}{y_i}^2-2n\bar{y}^2+n\bar{y}^2\\
&=\sum_{i=1}^{n}{y_i}^2-n\bar{y}^2
\end{aligned}
$$

4.4.3　S_{xy} と S_{xx} と S_{yy}

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} \tag{25}$$

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} {x_i}^2 - n\bar{x}^2 \tag{26}$$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} {y_i}^2 - n\bar{y}^2 \tag{27}$$

4.4.4 $\hat{\beta}$ と $\boldsymbol{S_{xy}}$ $\boldsymbol{S_{xx}}$

(17) 式再掲

$$\hat{\beta} = \frac{\displaystyle\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\displaystyle\sum_{i=1}^{n} {x_i}^2 - n\bar{x}^2} \tag{17}$$

(17) 式に (25) 式と (26) 式を代入

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} \tag{28}$$

4.4.5 $\boldsymbol{\alpha}$ と $\boldsymbol{S_{xy}}$ $\boldsymbol{S_{xx}}$

(21) 式再掲

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \tag{21}$$

(21) 式に (28) 式を代入

$$\hat{\alpha} = \bar{y} - \frac{S_{xy}}{S_{xx}}\bar{x} \tag{29}$$

4.5　残差平方和（Residual Sum of Squares）

ここで、$\hat{\alpha}, \hat{\beta}$ によって推定された直線上の値を $\tilde{y}_i$ であらわす。

$$\tilde{y}_i = \hat{\alpha} + \hat{\beta} x_i \tag{30}$$

y_i と $\tilde{y}_i$ の差

$$y_i - \tilde{y}_i \tag{31}$$

を『残差』と呼ぶ。

残差の二乗和を『残差平方和』と呼び RSS とあらわす。

$$RSS = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 \tag{32}$$

$$= \sum_{i=1}^{n} \left(y_i - (\hat{\alpha} + \hat{\beta} x_i)\right)^2 \tag{33}$$

$$= \sum_{i=1}^{n} \left(y_i - \hat{\alpha} - \hat{\beta} x_i\right)^2 \tag{34}$$

4.5.1 RSS と $\hat{\alpha}$

RSS に (21) 式を代入

$$RSS = \sum_{i=1}^{n} \left(y_i - \left(\bar{y} - \hat{\beta}\bar{x} \right) - \hat{\beta}x_i \right)^2 \tag{35}$$

$$= \sum_{i=1}^{n} \left(y_i - \bar{y} + \hat{\beta}\bar{x} - \hat{\beta}x_i \right)^2 \tag{36}$$

$$= \sum_{i=1}^{n} \left((y_i - \bar{y}) - \hat{\beta}\left(x_i - \bar{x}\right) \right)^2 \tag{37}$$

『平均からの偏差』を一つの記号とみなしてシグマ括弧内を展開

$$RSS = \sum_{i=1}^{n} \left((y_i - \bar{y})^2 + \hat{\beta}^2 \left(x_i - \bar{x}\right)^2 - 2\hat{\beta}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right) \right) \tag{38}$$

総和の公式を使い整理

$$= \sum_{i=1}^{n} (y_i - \bar{y})^2 + \hat{\beta}^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 - 2\hat{\beta} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \tag{39}$$

(25) 式、(26) 式、(27) 式より

$$RSS = S_{yy} + \hat{\beta}^2 S_{xx} - 2\hat{\beta} S_{xy} \tag{40}$$

(28) 式より

$$
\begin{aligned}
RSS &= S_{yy} + \left(\frac{S_{xy}}{S_{xx}}\right)^2 S_{xx} - 2\frac{S_{xy}}{S_{xx}} S_{xy} && (41)\\
&= S_{yy} + \frac{S_{xy}}{S_{xx}} \frac{S_{xy}}{\cancel{S_{xx}}} \cancel{S_{xx}} - 2\frac{S_{xy} S_{xy}}{S_{xx}} && (42)\\
&= S_{yy} + \frac{{S_{xy}}^2}{S_{xx}} - 2\frac{{S_{xy}}^2}{S_{xx}} && (43)\\
&= S_{yy} - \frac{{S_{xy}}^2}{S_{xx}} && (44)
\end{aligned}
$$

4.5.2 総平方和 (Total Sum of Squares)

S_{yy} を『総平方和』といい、TSS とあらわす。

$$TSS = S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y}_i)^2 \tag{45}$$

4.5.3 説明された平方和 (Explained Sum of Squares)

$\dfrac{{S_{xy}}^2}{S_{xx}}$ を『説明された平方和』といい、ESS とあらわす。

$$ESS = \frac{{S_{xy}}^2}{S_{xx}} \tag{46}$$

4.6 決定係数 $\boldsymbol{R^2}$

4.6.1 ESS と TSS と RSS の関係

(44) を TSS ならびに ESS であらわすと

$$RSS = TSS - ESS \tag{47}$$

$$TSS = ESS + RSS \tag{48}$$

4.6.2 ESS と TSS の比

ESS と TSS の比を決定係数と呼び、R^2 であらわす。決定係数は直線の当てはまりの尺度として利用される。

$$R^2 = \frac{ESS}{TSS} = ESS\frac{1}{TSS} = \frac{{S_{xy}}^2}{S_{xx}}\frac{1}{S_{yy}} = \frac{S_{xy}}{S_{xx}}\frac{S_{xy}}{S_{yy}} = \hat{\beta}\frac{S_{xy}}{S_{yy}} \tag{49}$$

4.7　まとめ

- 誤差の二乗和を最小化するのが最小二乗法である。
- 当てはまりの基準が決定係数 (R^2) である。
- 決定係数 (R^2) の最大値は 1、最小値は 0 である。
- 説明された平方和 (ESS) と総平方和 (TSS) の比率が決定係数 (R^2) である。