http://www.math.sci.ehime-u.ac.jp/~matsuura/

単回帰分析

Ver.2022.09.04

愛媛大学
データサイエンスセンター（CDSE）
理工学研究科／理学部

まつうら まさや
松浦　真也

2020年4月設立

CDSE
Center for Data Science, Ehime University

単回帰分析（目的）

【例】「宿題の点数」と「試験の点数」はどういう関係にあるのか？

データ出典：The Data And Story LibraryのMidterms

https://dasl.datadescription.com/datafile/midterms/

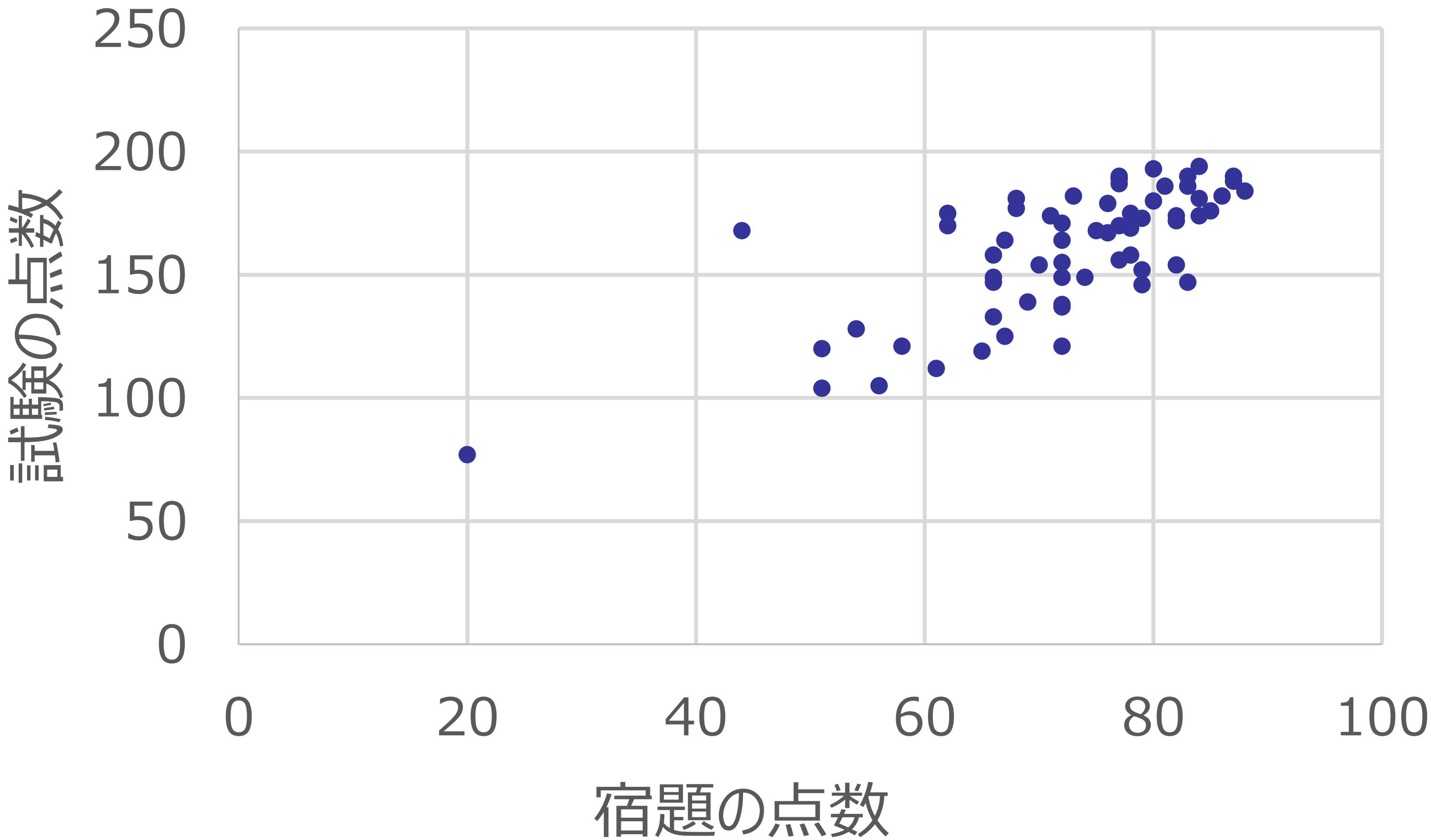
学生	宿題の点数	試験の点数
Adam	62	175
Alexandra	78	169
Alexis	68	181
Amandeep	54	128
Annie	68	181
Benjamin	72	149
Brian	76	167
Brian	82	172
⋮	⋮	⋮
Yvon	82	154

単回帰分析（目的）

【例】「宿題の点数」と「試験の点数」はどういう関係にあるのか？

データ出典：The Data And Story LibraryのMidterms

https://dasl.datadescription.com/datafile/midterms/

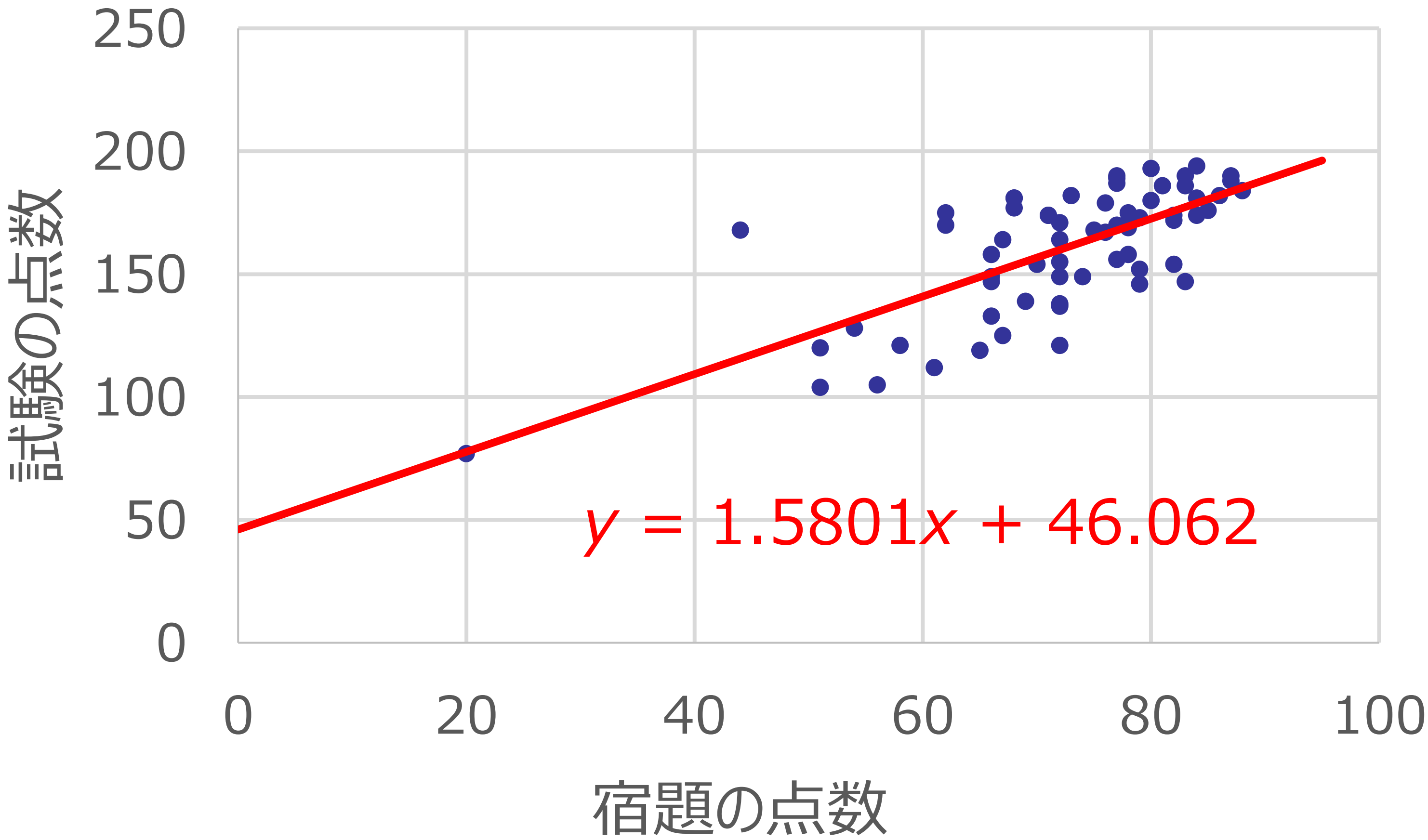


単回帰分析（目的）

【例】「宿題の点数」と「試験の点数」はどういう関係にあるのか？

データ出典：The Data And Story LibraryのMidterms

https://dasl.datadescription.com/datafile/midterms/



単回帰分析（回帰モデル）

【例】「宿題の点数」と「試験の点数」はどういう関係にあるのか？

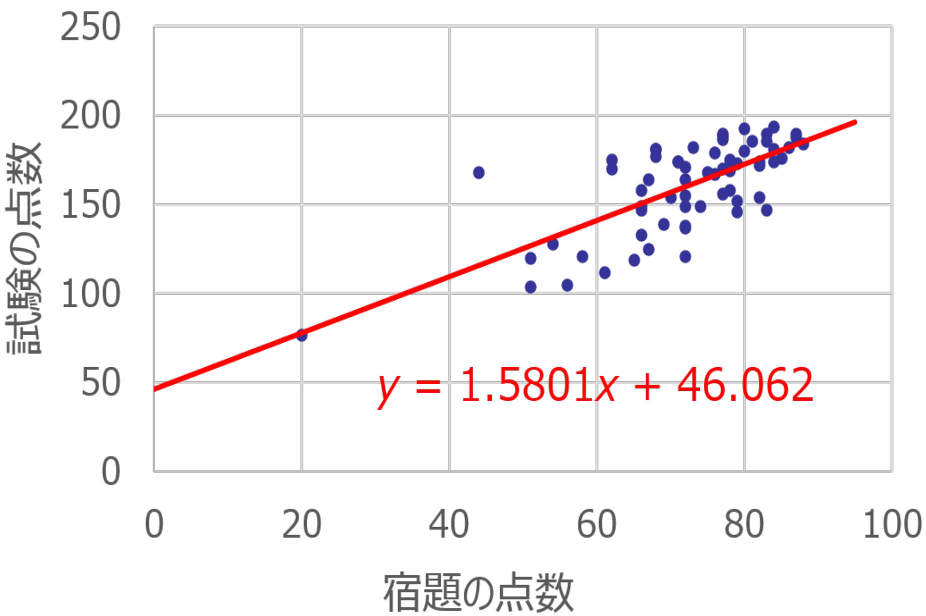
回帰係数

回帰式

$$\underbrace{y}_{\text{目的変数}} = 1.5801\underbrace{x}_{\text{説明変数}} + 46.062$$

目的変数
※従属変数とも言う

説明変数
※独立変数とも言う



単回帰分析（回帰モデル）

【例】「宿題の点数」と「試験の点数」はどういう関係にあるのか？

データ出典：The Data And Story LibraryのMidterms

https://dasl.datadescription.com/datafile/midterms/

学生	宿題の点数	試験の点数
Adam	62	175
Alexandra	78	169
Alexis	68	181
Amandeep	54	128
Annie	68	181
Benjamin	72	149
Brian	76	167
Brian	82	172
⋮	⋮	⋮
Yvon	82	154

単回帰分析（最小２乗法）

名前 i	宿題の点数 x_i	試験の点数 y_i
1	62	175
2	78	169
3	68	181
⋮	⋮	⋮
n	68	181

学生に依存しない未知の定数（これを推定）

回帰モデル

$$\underbrace{y_i}_{\text{学生 } i \text{ の試験の点数}} = \beta_0 + \beta_1 \underbrace{x_i}_{\text{学生 } i \text{ の宿題の点数}} + \underbrace{e_i}_{\text{残差}} \quad (1 \leq i \leq n)$$

残差平方和 $S_e = \sum_{i=1}^{n} e_i^2$ が最小になるように β_0, β_1 を定める

単回帰分析（ベクトル表記）

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad (1 \leq i \leq n)$$

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \vec{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \vec{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

$$\vec{y} = \beta_0 \vec{1} + \beta_1 \vec{x} + \vec{e}$$

（$\vec{y}$ を $\vec{1}$ と $\vec{x}$ の一次結合で近似）

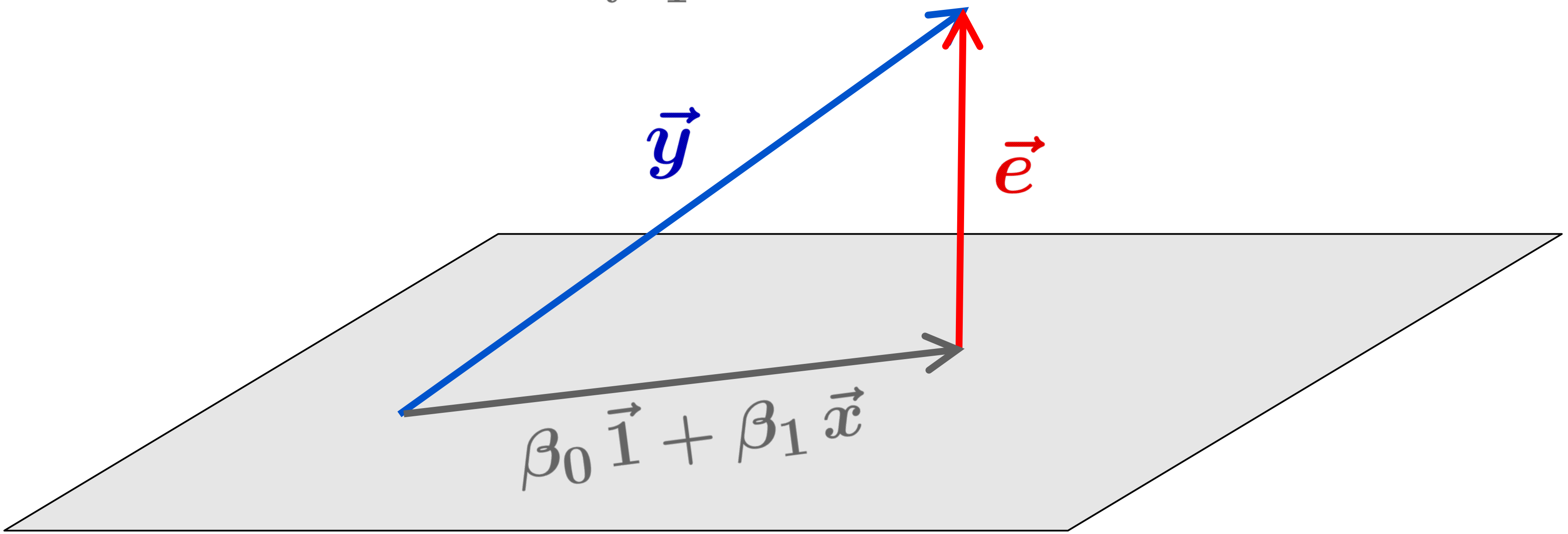
残差平方和 $S_e = \sum_{i=1}^{n} e_i^2 = |\vec{e}|^2$ を最小化

単回帰分析（幾何学的考察）

$$\vec{y} = \beta_0 \vec{1} + \beta_1 \vec{x} + \vec{e}$$

（ $\vec{y}$ を $\vec{1}$ と $\vec{x}$ の一次結合で近似）

残差平方和 $S_e = \sum_{i=1}^{n} e_i^2 = |\vec{e}|^2$ を最小化

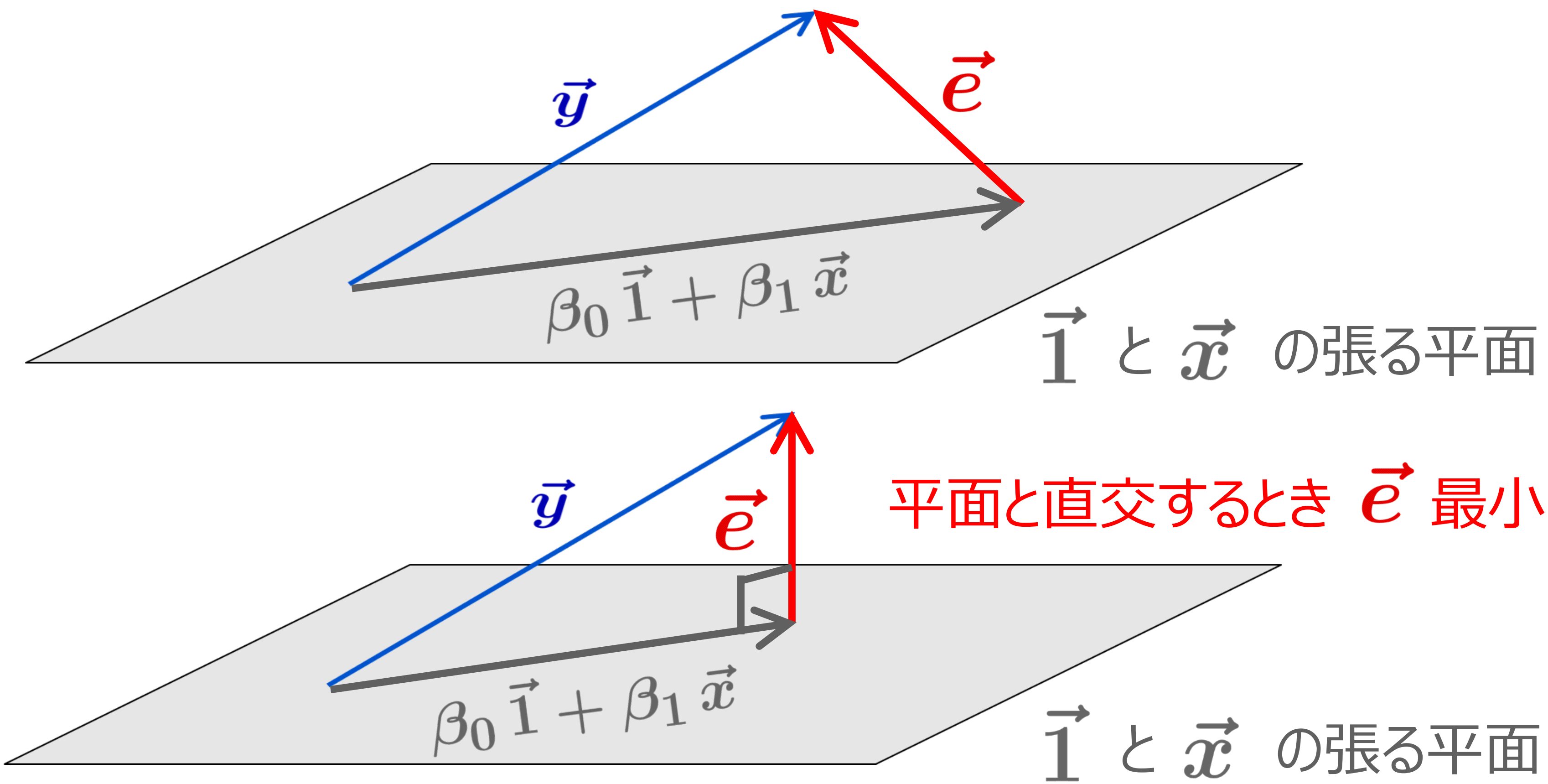


$\vec{1}$ と $\vec{x}$ の張る平面

単回帰分析（幾何学的考察）

$$\vec{y} = \beta_0 \vec{1} + \beta_1 \vec{x} + \vec{e}$$

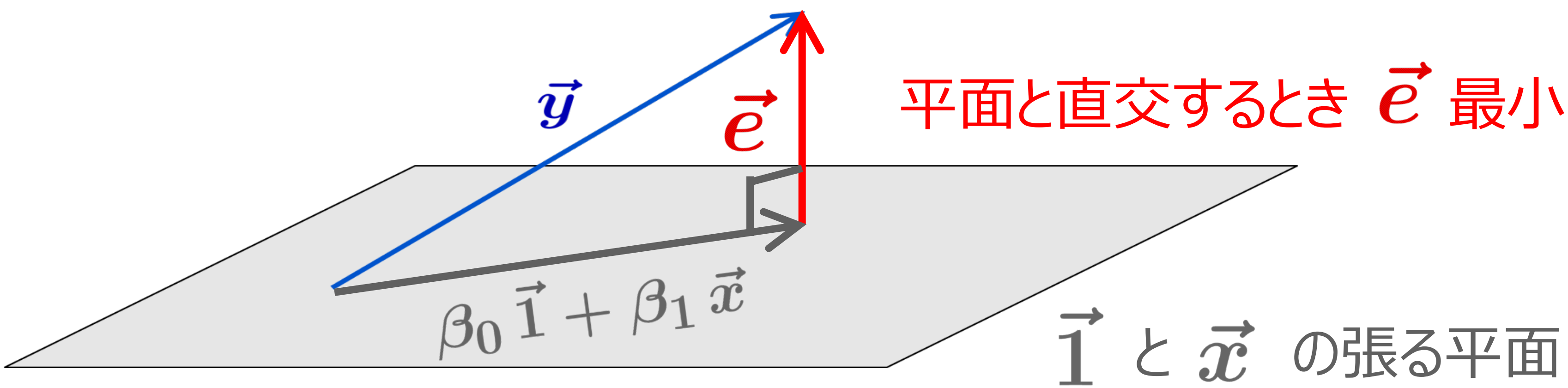
残差平方和 $S_e = \sum_{i=1}^{n} e_i^2 = |\vec{e}|^2$ を最小化



単回帰分析（幾何学的考察）

$$\vec{y} = \beta_0\,\vec{1} + \beta_1\,\vec{x} + \vec{e}$$

残差平方和 $S_e = \sum_{i=1}^{n} e_i^2 = |\vec{e}\,|^2$ を最小化



⇩

内積： $(\vec{1}, \vec{e}) = 0,\ (\vec{x}, \vec{e}) = 0$

単回帰分析（代表値のベクトル内積表記）

平均

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

偏差平方和

$$S_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2, \quad S_{yy} = \sum_{i=1}^{n} (y_i - \overline{y})^2$$

偏差積和

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

$$\Downarrow$$

内積

$$\frac{(\vec{x}, \vec{1})}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}, \qquad \frac{(\vec{y}, \vec{1})}{n} = \frac{1}{n}\sum_{i=1}^{n} y_i = \overline{y},$$

$$\left(\vec{x} - \overline{x}\vec{1}, \vec{x} - \overline{x}\vec{1}\right) = S_{xx}, \ \left(\vec{x} - \overline{x}\vec{1}, \vec{y} - \overline{y}\vec{1}\right) = S_{xy}$$

単回帰分析（回帰係数の計算）

単回帰モデル $\vec{y} = \beta_0 \vec{1} + \beta_1 \vec{x} + \vec{e}$

$$\Downarrow$$

$$\overline{y} = \frac{(\vec{y}, \vec{1})}{n} = \beta_0 \frac{(\vec{1}, \vec{1})}{n} + \beta_1 \frac{(\vec{x}, \vec{1})}{n} + \frac{(\vec{e}, \vec{1})}{n} = \beta_0 + \beta_1 \overline{x}$$

$$\Downarrow$$

$$\vec{y} - \overline{y}\vec{1} = \beta_1 (\vec{x} - \overline{x}\vec{1}) + \vec{e}$$

$$\Downarrow$$

$$\begin{aligned} S_{xy} &= (\vec{x} - \overline{x}\vec{1}, \vec{y} - \overline{y}\vec{1}) \\ &= \beta_1 (\vec{x} - \overline{x}\vec{1}, \vec{x} - \overline{x}\vec{1}) + (\vec{x} - \overline{x}\vec{1}, \vec{e}) = \beta_1 S_{xx} \end{aligned}$$

$$\Downarrow$$

$$\beta_1 = \frac{S_{xy}}{S_{xx}}, \quad \beta_0 = \overline{y} - \beta_1 \overline{x} = \overline{y} - \frac{S_{xy}}{S_{xx}} \overline{x}$$

単回帰分析（結論）

回帰モデル

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad (1 \leq i \leq n)$$

残差平方和 $S_e = \sum_{i=1}^{n} e_i^2$ が最小になるように β_0, β_1 を定める。

このときの β_0, β_1 を $\widehat{\beta}_0, \widehat{\beta}_1$ とすると、

$$\widehat{\beta}_0 = \overline{y} - \frac{S_{xy}}{S_{xx}} \overline{x}, \quad \widehat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

ただし、$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i,\ \overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i,$

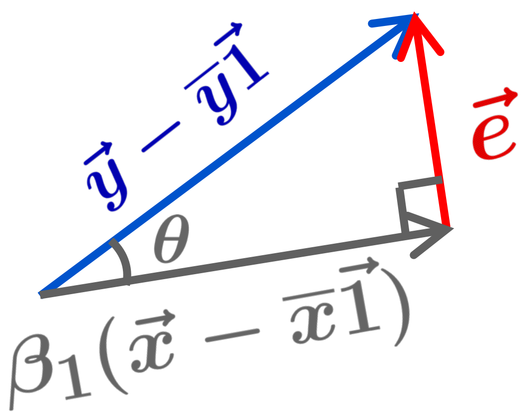
$$S_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2, \quad S_{xy} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$$

単回帰分析（決定係数と相関係数）

単回帰式（の変形版）

$$\underbrace{\vec{y}-\overline{y}\vec{1}}_{\vec{y}\text{の変動}} = \underbrace{\beta_1(\vec{x}-\overline{x}\vec{1})}_{\vec{x}\text{の変動で決まる部分}} + \underbrace{\vec{e}}_{\text{残差}}, \quad \beta_1 = \frac{S_{xy}}{S_{xx}}$$

$$(\vec{1},\vec{e}) = 0, \ (\vec{x},\vec{e}) = 0$$



決定係数（寄与率）の定義

$$R^2 = \frac{(\vec{x}\text{の変動で決まる部分の大きさ})^2}{(\vec{y}\text{の変動の大きさ})^2}$$

$$R^2 = \beta_1^2 \frac{(\vec{x}-\overline{x}\vec{1}, \vec{x}-\overline{x}\vec{1})}{(\vec{y}-\overline{y}\vec{1}, \vec{y}-\overline{y}\vec{1})} = \frac{S_{xy}^2}{S_{xx}^2}\frac{S_{xx}}{S_{yy}} = \frac{S_{xy}^2}{S_{xx}S_{yy}}$$

一方、図より $R^2 = \cos^2\theta$　　※ $0 \leq R^2 \leq 1$

単回帰分析（決定係数と相関係数）

標本分散

$$V_x = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{S_{xx}}{n-1},$$

$$V_y = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2 = \frac{S_{yy}}{n-1}$$

標本共分散

$$C_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \frac{S_{xy}}{n-1}$$

標本相関係数

$$r_{xy} = \frac{C_{xy}}{\sqrt{V_x V_y}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

$$\Downarrow$$

決定係数は相関係数の2乗

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} = r_{xy}^2$$

単回帰分析（参考文献）

参考文献

・成績のデータ
The Data And Story Library, Midterms,
https://dasl.datadescription.com/datafile/midterms/
（最終閲覧日 2022年9月3日）