多変量解析

第7回　重回帰分析

萩原・篠田
情報理工学部

回帰分析

マンション価格は広さと築年数から予測可能か？

量的変数　量的変数

回帰式：$\hat{y} = 1.02 + 0.067x_1 - 0.081x_2$

サンプル No.	広さ（m²） x_1	築年数（年） x_2	価格（千万円） y
1	51	16	3.0
2	38	4	3.2
3	57	16	3.3
4	51	11	3.9
5	53	4	4.4
6	77	22	4.5
7	63	5	4.5
8	69	5	5.4
9	72	2	5.4
10	73	1	6.0

keywords

目的変数、説明変数、
線形回帰、残差、
最小二乗法、
決定係数（寄与率）、
分散共分散行列

重回帰分析

重回帰分析：いくつかの<u>原因</u>と<u>結果</u>を結ぶもの

目的変数 y (結果)とそれに影響を与えるいくつかの
説明変数 $x_1, x_2, \cdots x_p$ (原因)とから

$$Y = a_1x_1 + a_2x_2 + \cdots + a_px_p + a_0$$

1次式作成。
この式を使って目的変数 y の予測や制御に役立てる。

目的変数　　　　　　　説明変数

$$y \underset{\text{制御}}{\overset{\text{予測}}{\rightleftarrows}} Y = a_1x_1 + a_2x_2 + \cdots + a_px_p + a_0$$

【説明変数が２個の場合】

$$Y = a_1 x_1 + a_2 x_2 + a_0$$

この式は立体座標上の平面(右図)に相当する

もし平面上にデータを表す点 P_1 があるならば

$$Y_1 = a_1 x_{11} + a_2 x_{21} + a_0$$

となるはずだが、実際はε_1だけ離れているので

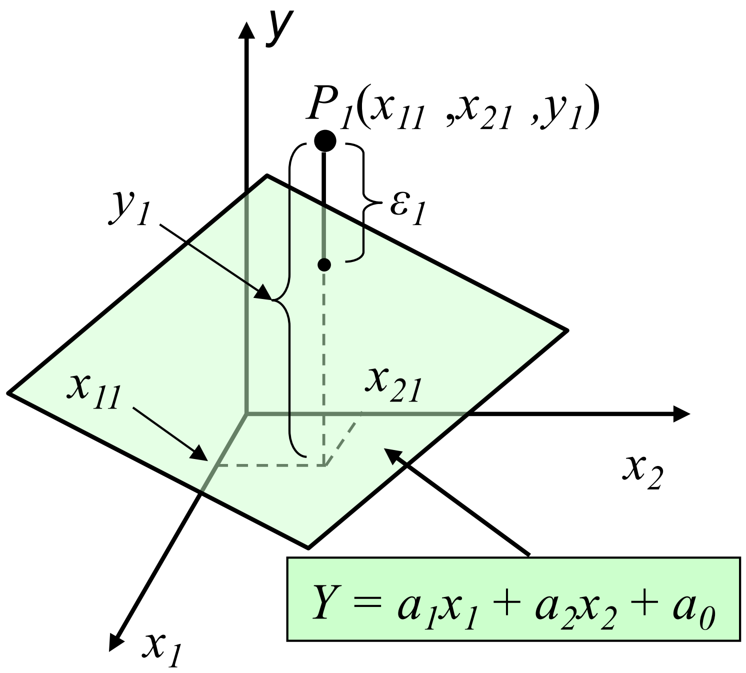
$$y_1 = a_1 x_{11} + a_2 x_{21} + a_0 + \varepsilon_1$$

従って

$$\varepsilon_1 = y_1 - a_1 x_{11} - a_2 x_{21} - a_0$$

どのデータにも当てはまるように書けば

$$\varepsilon_i = y_i - a_1 x_{1i} - a_2 x_{2i} - a_0$$



$Y = a_1 x_1 + a_2 x_2 + a_0$の平面がすべてのデータから最も近い距離にあるためには

$$S = \Sigma\varepsilon_i^2 = \Sigma(y_i - a_1 x_{1i} - a_2 x_{2i} - a_0)^2$$

が<u>最小</u>になるように a_1 と a_2 と a_0を求める（最小二乗法）

$$\boxed{\begin{aligned} \frac{\partial S}{\partial a_0} &= 0 \\ \frac{\partial S}{\partial a_1} &= 0 \\ \frac{\partial S}{\partial a_2} &= 0 \end{aligned}}$$

例題

セラミックスの配向度に対する温度と
時間が及ぼす影響を調べる

試料番号	加圧の条件			配向度
	温度 (℃) x_1	時間 (min) x_2	圧力 (MPa)	% y
1	1700	30	25	36
2	1800	25	15	39
3	1800	20	20	44
4	1850	30	20	44
5	1900	10	20	59
6	1930	10	20	51

$\bar{x}_1 = 1830$

$\bar{x}_2 = 20.83$

$\bar{y} = 45.5$

重回帰モデル

配向度 $= a_1$ ×温度 $+ a_2$ ×時間 + 定数

目的変数 $y =$ 配向度

説明変数 $x_1 =$ 温度、説明変数 $x_2 =$ 時間

配向度　　a_1 ×温度 $+ a_2$ ×時間 + 定数

$$y \longleftrightarrow Y = a_1 x_1 + a_2 x_2 + a_0$$

Y ：重回帰式

a_1, a_2 ：回帰係数

a_0 ：定数項

試料番号	実測値	重回帰式による予測値
1	36	$1700a_1 + 30a_2 + a_0$
2	39	$1800a_1 + 25a_2 + a_0$
3	44	$1800a_1 + 20a_2 + a_0$
4	44	$1850a_1 + 30a_2 + a_0$
5	59	$1900a_1 + 10a_2 + a_0$
6	51	$1930a_1 + 10a_2 + a_0$

$$S = \Sigma\varepsilon_i^2 = \Sigma(y_i - a_1x_{1i} - a_2x_{2i} - a_0)^2$$
$$= (36 - (1700a_1 + 30a_2 + a_0))^2$$
$$+ \cdots + (51 - (1930a_1 + 10a_2 + a_0))^2$$

偏微分してゼロとおくと

$$\left\{\begin{array}{ll} \dfrac{\partial S}{\partial a_0} = 2(6a_0 + 10980a_1 + 125a_2 - 273) = 0 & \cdots\cdots ① \\ \dfrac{\partial S}{\partial a_1} = 2(20127400a_1 + 225800a_2 + 10980a_0 - 502530) = 0 & \cdots\cdots ② \\ \dfrac{\partial S}{\partial a_2} = 2(3025a_2 + 225800a_1 + 125a_0 - 5355) = 0 & \cdots\cdots ③ \end{array}\right.$$

$\boldsymbol{a_0}$を消去し，$\boldsymbol{a_1}$と$\boldsymbol{a_2}$の式を得て

$$\left\{\begin{array}{ll} 34000a_1 - 2950a_2 - 2940 = 0 & \cdots\cdots ④ \\ -2950a_1 + 420.9a_2 + 332.5 = 0 & \cdots\cdots ⑤ \end{array}\right.$$

これを解くと

$$a_1 = 0.04573,\quad a_2 = -0.4695,\quad a_0 = -28.41$$

従って重回帰式は

$$Y = 0.04573x_1 - 0.4695x_2 - 28.41 \qquad \cdots\cdots ⑥$$

重回帰式 $Y = a_0 + a_1 x_1 + a_2 x_2$ を分散共分散行列を用いて求める方法

データ(説明変数x_1, x_2, 目的変数y)による分散共分散行列と
重回帰式(説明変数x_1, x_2, 重回帰式Y)による分散共分散行列を比較する

$$
\begin{array}{c}
 \\ x_1 \\ x_2 \\ y
\end{array}
\begin{array}{c}
\begin{array}{ccc} \quad x_1 \quad & \quad x_2 \quad & \quad y \quad \end{array} \\
\begin{bmatrix}
\text{分散} & \text{共分散} & \text{共分散} \\
\text{共分散} & \text{分散} & \text{共分散} \\
\text{共分散} & \text{共分散} & \text{分散}
\end{bmatrix}
\end{array}
$$

分散(Variance)　Var

共分散(Covariance)　Cov

重回帰式による分散共分散行列

データによる分散共分散行列

$$
\begin{array}{c|ccc}
 & x_1 & x_2 & Y \\
\hline
x_1 & \mathrm{Var}(x_1) & \mathrm{Cov}(x_1, x_2) & \mathrm{Cov}(x_1, Y) \\
x_2 & \mathrm{Cov}(x_1, x_2) & \mathrm{Var}(x_2) & \mathrm{Cov}(x_2, Y) \\
Y & \mathrm{Cov}(x_1, Y) & \mathrm{Cov}(x_2, Y) & \mathrm{Var}(Y)
\end{array}
=
\begin{array}{c|ccc}
 & x_1 & x_2 & y \\
\hline
x_1 & \mathrm{Var}(x_1) & \mathrm{Cov}(x_1, x_2) & \mathrm{Cov}(x_1, y) \\
x_2 & \mathrm{Cov}(x_1, x_2) & \mathrm{Var}(x_2) & \mathrm{Cov}(x_2, y) \\
y & \mathrm{Cov}(x_1, y) & \mathrm{Cov}(x_2, y) & \mathrm{Var}(y)
\end{array}
$$

重回帰式 $Y = a_0 + a_1 x_1 + a_2 x_2$ を求める方法　□ の部分を比較してみる

重回帰式による分散共分散行列　　データによる分散共分散行列

$$\begin{array}{c} \\ x_1 \\ x_2 \\ Y \end{array} \begin{array}{c} \begin{array}{ccc} x_1 & x_2 & Y \end{array} \\ \begin{bmatrix} \mathrm{Var}(x_1) & \mathrm{Cov}(x_1, x_2) & \boxed{\mathrm{Cov}(x_1, Y)} \\ \mathrm{Cov}(x_1, x_2) & \mathrm{Var}(x_2) & \boxed{\mathrm{Cov}(x_2, Y)} \\ \mathrm{Cov}(x_1, Y) & \mathrm{Cov}(x_2, Y) & \mathrm{Var}(Y) \end{bmatrix} \end{array} = \begin{array}{c} \\ x_1 \\ x_2 \\ y \end{array} \begin{array}{c} \begin{array}{ccc} x_1 & x_2 & y \end{array} \\ \begin{bmatrix} \mathrm{Var}(x_1) & \mathrm{Cov}(x_1, x_2) & \boxed{\mathrm{Cov}(x_1, y)} \\ \mathrm{Cov}(x_1, x_2) & \mathrm{Var}(x_2) & \boxed{\mathrm{Cov}(x_2, y)} \\ \mathrm{Cov}(x_1, y) & \mathrm{Cov}(x_2, y) & \mathrm{Var}(y) \end{bmatrix} \end{array}$$

$\mathrm{Cov}(x, ax + by + c) = a\mathrm{Var}(x) + b\mathrm{Cov}(x, y)$ より

$$\mathrm{Cov}(x_1, Y) = \mathrm{Cov}(x_1, a_0 + a_1 x_1 + a_2 x_2) = a_1 \mathrm{Var}(x_1) + a_2 \mathrm{Cov}(x_1, x_2) = \mathrm{Cov}(x_1, y)$$

$$\mathrm{Cov}(x_2, Y) = \mathrm{Cov}(x_2, a_0 + a_1 x_1 + a_2 x_2) = a_1 \mathrm{Cov}(x_1, x_2) + a_2 \mathrm{Var}(x_2) = \mathrm{Cov}(x_2, y)$$

$$\begin{bmatrix} \mathrm{Var}(x_1) & \mathrm{Cov}(x_1, x_2) \\ \mathrm{Cov}(x_1, x_2) & \mathrm{Var}(x_2) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \mathrm{Cov}(x_1, y) \\ \mathrm{Cov}(x_2, y) \end{bmatrix} \Rightarrow \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \mathrm{Var}(x_1) & \mathrm{Cov}(x_1, x_2) \\ \mathrm{Cov}(x_1, x_2) & \mathrm{Var}(x_2) \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{Cov}(x_1, y) \\ \mathrm{Cov}(x_2, y) \end{bmatrix}$$

$$\begin{bmatrix} \mathrm{Var}(x_1) & \mathrm{Cov}(x_1, x_2) \\ \mathrm{Cov}(x_1, x_2) & \mathrm{Var}(x_2) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \mathrm{Cov}(x_1, y) \\ \mathrm{Cov}(x_2, y) \end{bmatrix} \Rightarrow \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \mathrm{Var}(x_1) & \mathrm{Cov}(x_1, x_2) \\ \mathrm{Cov}(x_1, x_2) & \mathrm{Var}(x_2) \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{Cov}(x_1, y) \\ \mathrm{Cov}(x_2, y) \end{bmatrix}$$

従って、データより得られる分散、共分散の値を用いた連立1次方程式から、回帰係数 a_1、a_2を求めることができる

例題で確認

平均

$$\overline{x} = \frac{\Sigma x_i}{n}$$

分散 Var(x)

$$s_x{}^2 = \frac{\Sigma(x_i - \overline{x})^2}{n - 1}$$

共分散 Cov(x, y)

$$s_{xy} = \frac{\Sigma(x_i - \overline{x})(y_i - \overline{y})}{n - 1}$$

温度 x_1 の平均 $\overline{x_1}$ と分散 $s_1{}^2$

$\overline{x_1} = 1830$, $s_1{}^2 = 6800$

時間 x_2 の平均 $\overline{x_2}$ と分散 $s_2{}^2$

$\overline{x_2} = 20.83$, $s_2{}^2 = 84.2$

配向度 y の平均 $\overline{y}$ と分散 $s_y{}^2$

$\overline{y} = 45.5$, $s_y{}^2 = 69.9$

温度 x_1 と時間 x_2 の共分散 s_{12}

$s_{12} = -590$

温度 x_1と配向度 y の共分散 s_{1y}

$s_{1y} = 588$

時間 x_2 と配向度 y の共分散 s_{2y}

$s_{2y} = -66.5$

分散共分散行列

$$\begin{array}{cc} & \begin{array}{ccc} \text{温度}\ x_1 & \text{時間}\ x_2 & \text{配向度}\ y \end{array} \\ \begin{array}{c} \text{温度}\ x_1 \\ \text{時間}\ x_2 \\ \text{配向度}\ y \end{array} & \begin{bmatrix} s_1^2 & s_{12} & s_{1y} \\ s_{12} & s_2^2 & s_{2y} \\ \hline s_{1y} & s_{2y} & s_y^2 \end{bmatrix} \end{array}$$

$$\begin{array}{cc} & \begin{array}{ccc} \text{温度}\ x_1 & \text{時間}\ x_2 & \text{配向度}\ y \end{array} \\ \begin{array}{c} \text{温度}\ x_1 \\ \text{時間}\ x_2 \\ \text{配向度}\ y \end{array} & \begin{bmatrix} 6800 & -590 & 588 \\ -590 & 84.2 & -66.5 \\ \hline 588 & -66.5 & 69.9 \end{bmatrix} \end{array}$$

4ページ前の例題の式

$$34000a_1 - 2950a_2 - 2940 = 0 \quad \cdots\cdots ④$$

$$-2950a_1 + 420.9a_2 + 332.5 = 0 \quad \cdots\cdots ⑤$$

を 5(= n - 1)で割ると、じつは同じ式

$$6800a_1 - 590a_2 - 588 = 0$$

$$-590a_1 + 84.2a_2 + 66.5 = 0$$

$$\begin{bmatrix} 6800 & -590 \\ -590 & 84.2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 588 \\ -66.5 \end{bmatrix}$$

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 6800 & -590 \\ -590 & 84.2 \end{bmatrix}^{-1} \begin{bmatrix} 588 \\ -66.5 \end{bmatrix}$$

$$\begin{bmatrix} \mathrm{Var}(x_1) & \mathrm{Cov}(x_1, x_2) \\ \mathrm{Cov}(x_1, x_2) & \mathrm{Var}(x_2) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \mathrm{Cov}(x_1, y) \\ \mathrm{Cov}(x_2, y) \end{bmatrix} \Rightarrow \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \mathrm{Var}(x_1) & \mathrm{Cov}(x_1, x_2) \\ \mathrm{Cov}(x_1, x_2) & \mathrm{Var}(x_2) \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{Cov}(x_1, y) \\ \mathrm{Cov}(x_2, y) \end{bmatrix}$$

説明変数がp個の場合

目的変数をy、説明変数を $x_1, x_2 \cdots, x_p$とすると

変数

	y	x_1	x_2	$\cdots$	x_p
1	y_1	x_{11}	x_{21}	$\cdots$	x_{p1}
2	y_2	x_{12}	x_{22}	$\cdots$	x_{p2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
n	y_n	x_{1n}	x_{2n}	$\cdots$	x_{pn}
平均	$\bar{y}$	$\bar{x}_1$	$\bar{x}_2$	$\cdots$	$\bar{x}_p$

個体

分散共分散行列

$$
\begin{array}{c} \\ x_1 \\ x_2 \\ \vdots \\ x_p \\ y \end{array}
\begin{array}{c}
\begin{array}{ccccc} x_1 & x_2 & \cdots & x_p & y \end{array} \\
\begin{bmatrix}
s_1{}^2 & s_{12} & \cdots & s_{1p} & s_{1y} \\
s_{12} & s_2{}^2 & \cdots & s_{2p} & s_{2y} \\
\vdots & \vdots & & \vdots & \vdots \\
s_{1p} & s_{2p} & \cdots & s_p{}^2 & s_{py} \\
s_{1y} & s_{2y} & \cdots & s_{py} & s_y{}^2
\end{bmatrix}
\end{array}
$$

重回帰式 $\boxed{Y = a_0 + a_1x_1 + a_2x_2 + \cdots + a_px_p}$

$$Y_i = a_0 + a_1x_{1i} + a_2x_{2i} + \cdots + a_px_{pi}$$

回帰係数 $a_1, a_2, \cdots, a_p$ と定数項 a_0 は次の連立方程式からもとまる

$$
\begin{bmatrix}
s_1^2 & s_{12} & \cdots & s_{1p} \\
s_{12} & s_2^2 & \cdots & s_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
s_{1p} & s_{12} & \cdots & s_p^2
\end{bmatrix}
\cdot
\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}
=
\begin{bmatrix} s_{1y} \\ s_{2y} \\ \vdots \\ s_{py} \end{bmatrix}
$$

$$
\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}
=
\begin{bmatrix}
s_1^2 & s_{12} & \cdots & s_{1p} \\
s_{12} & s_2^2 & \cdots & s_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
s_{1p} & s_{12} & \cdots & s_p^2
\end{bmatrix}^{-1}
\cdot
\begin{bmatrix} s_{1y} \\ s_{2y} \\ \vdots \\ s_{py} \end{bmatrix}
$$

$$\bar{y} = a_0 + a_1\bar{x}_1 + a_2\bar{x}_2 + \cdots + a_p\bar{x}_p$$

$$a_0 = \bar{y} - a_1\bar{x}_1 - a_2\bar{x}_2 - \cdots - a_p\bar{x}_p$$

重回帰は変数間の関係を
よく表しているか

重回帰式⑥は

$$Y = 0.04573x_1 - 0.4695x_2 - 28.41$$

実測値、予測値、誤差(残差)を
比較すると

試料番号	実測値y	予測値Y	誤差(残差)
1	36	35.25	0.75
2	39	42.17	-3.17
3	44	44.52	-0.52
4	44	42.11	1.89
5	59	53.79	5.21
6	51	55.16	-4.16
平均	45.5	45.5	0
偏差平方和	349.5	290.57	58.93
分散	69.9	58.11	11.79

実測値の変動 ＝ 予測値の変動 ＋ 残差の変動

（実測値の偏差平方和 ＝ 予測値の偏差平方和 ＋ 残差平方和）

残差の変動が小さいということは重回帰式が良く当てはまっていることを示す。
従って予測値変動の大きさ（=残差変動の小ささ）が当てはまりの良さを示している。

$$R^2 = \frac{\text{予測値の偏差平方和}}{\text{実測値の偏差平方和}} = \frac{\text{予測値の分散}}{\text{実測値の分散}}$$

・・・（決定係数 or 寄与率）

$$R^2 = \frac{290.57}{349.5} = \frac{58.11}{69.9} = 0.831$$

決定係数 R^2 （$0 \le R^2 \le 1$）

R^2 が 1 に近いほど当てはまりが良い

決定係数の平方根 R を重相関係数と呼ぶ

重回帰は変数間の関係を
よく表しているか

重回帰式⑥は

$$Y = 0.04573x_1 - 0.4695x_2 - 28.41$$

実測値、予測値、誤差(残差)を
比較すると

試料番号	実測値y	予測値Y	偏差積
1	36	35.25	97.37
2	39	42.17	21.63
3	44	44.52	1.47
4	44	42.11	5.08
5	59	53.79	11.88
6	51	55.16	53.13
平均	45.5	45.5	偏差積和 = 290.57 共分散 =58.11
偏差平方和	349.5	290.57	
分散	69.9	58.11	

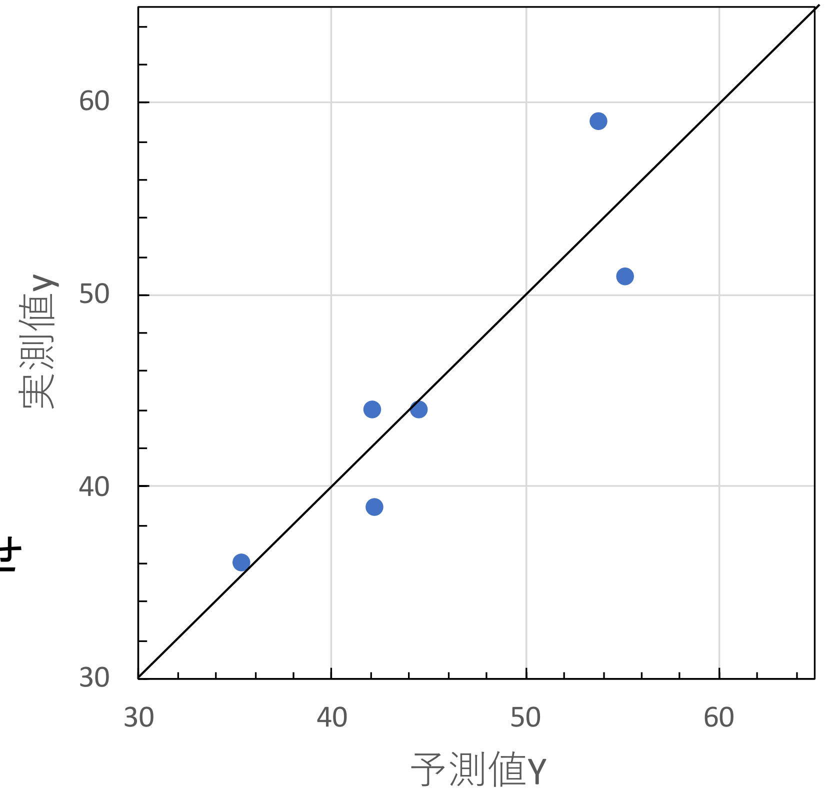
重相関係数*R*

$$R = \frac{\sum(y_i - \bar{y})(Y_i - \bar{Y})}{\sqrt{\sum(Y_i - \bar{Y})^2 \cdot \sum(y_i - \bar{y})^2}} = \frac{S_{yY}}{\sqrt{S_Y S_y}} = \frac{s_{yY}}{s_Y s_y}$$

$$R = \frac{290.57}{\sqrt{290.57 \cdot 349.5}} = \frac{58.11}{\sqrt{58.11 \cdot 69.9}} = 0.91$$

発展課題：決定係数 = 重相関係数² となることを示せ

$$R^2 = \frac{\sum(Y_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = \frac{S_Y}{S_y} = \frac{{s_Y}^2}{{s_y}^2}$$



実測値（目的変数のデータ）y、予測値 Y、残差 εとしたときに

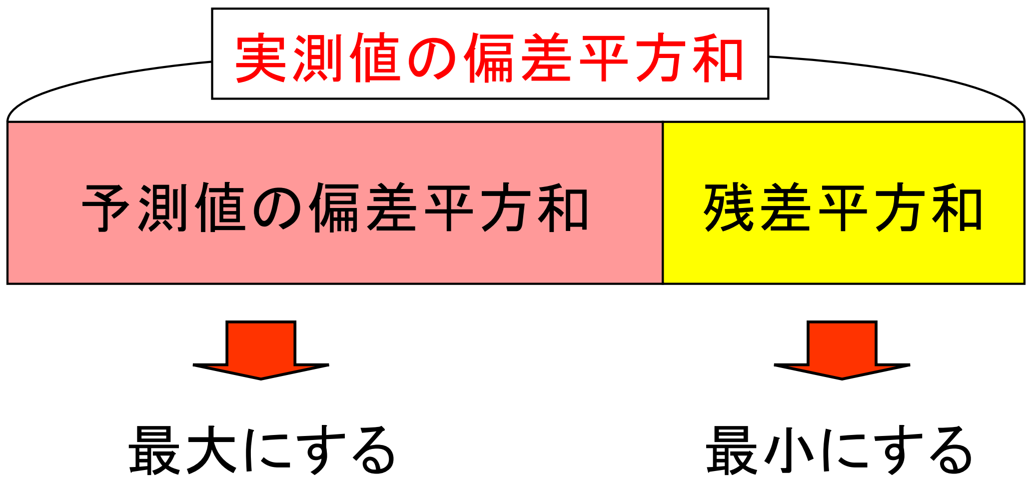
$S_y = S_Y + S_\varepsilon$が成立する

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

S_y 実測値の変動（偏差平方和）
S_Y 予測値の変動（偏差平方和）
S_ε 残差の変動（偏差平方和）

重回帰分析 ＝ 残差平方和を最小にする

予測値の偏差平方和を最大にする



$S_y = S_Y + S_e$の確認

$$\underbrace{\sum_{i=1}^{n}(y_i-\bar{y})^2}_{S_y} = \sum_{i=1}^{n}(y_i - Y_i + Y_i - \bar{y})^2 = \underbrace{\sum_{i=1}^{n}(y_i - Y_i)^2}_{S_\varepsilon} + \underbrace{\sum_{i=1}^{n}(Y_i-\bar{y})^2}_{S_Y} + \underbrace{2\sum_{i=1}^{n}(y_i - Y_i)(Y_i - \bar{y})}_{0}$$

$$\boxed{\sum_{i=1}^{n}(y_i - Y_i)(Y_i - \bar{y}) = \sum_{i=1}^{n}\varepsilon_i(a_0 + a_1 x_{1i} + a_2 x_{2i} - \bar{y}) = (a_0 - \bar{y})\underbrace{\sum_{i=1}^{n}\varepsilon_i}_{0} + a_1\underbrace{\sum_{i=1}^{n}x_{1i}\varepsilon_i}_{0} + a_2\underbrace{\sum_{i=1}^{n}x_{2i}\varepsilon_i}_{0}}$$

$$S_\varepsilon = \sum_{i=1}^{n}\varepsilon_i^2 = \sum_{i=1}^{n}(y_i - Y_i)^2 = \sum_{i=1}^{n}(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i})^2$$

$$\begin{cases} \dfrac{\partial S_\varepsilon}{\partial a_0} = -2\sum_{i=1}^{n}(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i}) = -2\sum_{i=1}^{n}\varepsilon_i = 0 & (1) \\ \dfrac{\partial S_\varepsilon}{\partial a_1} = -2\sum_{i=1}^{n}x_{1i}(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i}) = -2\sum_{i=1}^{n}x_{1i}\varepsilon_i = 0 & (2) \\ \dfrac{\partial S_\varepsilon}{\partial a_2} = -2\sum_{i=1}^{n}x_{2i}(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i}) = -2\sum_{i=1}^{n}x_{2i}\varepsilon_i = 0 & (3) \end{cases}$$

$$\frac{\text{予測値の偏差平方和}}{\underbrace{\text{予測値の偏差平方和}\quad\text{残差平方和}}_{\text{実測値の偏差平方和}}} = R^2$$

決定係数（寄与率）

$$\frac{S_Y}{S_y} = 1 - \frac{S_\varepsilon}{S_y} = R^2$$

大きいほど実測値と予測値が近づく
回帰方程式の当てはまりの良さを示す
指標になる

次の関係が成り立つ

$$0 \leq R^2 \leq 1$$

決定係数（寄与率）が1に近いほど残差平方和も小さく、
回帰方程式は資料の情報を十分に吸収している。

また一般的な決定係数（寄与率）の評価の目安として

$0.8 \leq R^2$ ：かなり良い精度

$0.5 \leq R^2 < 0.8$ ：まあまあ良い精度

といわれている。

自由度調整済み決定係数 $\hat{R}^2$

決定係数は重回帰式の当てはまりの良さを表す統計量
しかしながら　“説明変数の数を増やすと決定係数も増加する”

試料番号	加圧の条件			配向度
	温度 (℃)	時間 (min)	圧力 (MPa)	%
1	1700	30	25	36
2	1800	25	15	39
3	1800	20	20	44
4	1850	30	20	44
5	1900	10	20	59
6	1930	10	20	51

目的変数 y = 配向度
説明変数 x_1 = 温度、説明変数 x_2 = 時間

$$Y = 0.04573x_1 - 0.4695x_2 - 28.41$$

$$R^2 = 0.831$$

説明変数 x_3 = 圧力　を加えると

$$Y = 0.05751x_1 - 0.4157x_2 + 0.4829x_3 - 60.73$$

$$R^2 = 0.853 > 0.831$$

圧力は配向度にあまり関係がなさそうな変数

目的変数にあまり影響を与えない説明変数を加えても決定係数は単調に増加する
そこで自由度調整済み決定係数 $\hat{R}^2$が考え出された

自由度調整済み決定係数 $\hat{R}^2 = 1 - \dfrac{\frac{S_\varepsilon}{n-p-1}}{\frac{S_y}{n-1}}$

$S_\varepsilon = \sum_{i=1}^{n}(y_i - Y_i)^2$

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

p: 説明変数の個数
n: データ数

今までの例題で計算する

$$S_\varepsilon = \sum_{i=1}^{n}(y_i - Y_i)^2$$

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

p: 説明変数の個数

n: データ数

自由度調整済み決定係数 $\hat{R}^2 = 1 - \dfrac{\frac{S_\varepsilon}{n-p-1}}{\frac{S_y}{n-1}}$

説明変数に圧力を加える場合と加えない場合を比較する

説明変数	温度　時間	温度　時間　圧力
決定係数R^2	0.831	0.853
自由度調整済み決定係数$\hat{R}^2$	0.719	0.646
差	0.112	0.207

説明変数を増やすと決定係数は増加する。
しかしながら自由度調整済み決定係数では値が逆転する。
決定係数と自由度調整済み決定係数の差の小さい説明変数が良い。

重回帰分析

①重回帰分析とは何か。
　式も用いて説明せよ。
②説明変数が2個の場合を図示して重回
　帰式の求め方を説明せよ。
③重回帰分析における決定係数(寄与率)
　とは何か。
　実測値の変動、予測値の変動、
　残差の変動の関係を述べ
　決定係数について説明せよ。