■回帰直線

変数 y を x の一次式で説明するモデルを考える(単回帰モデル).

$$y = \alpha + \beta x$$

観測値 $(x_i, y_i)\ (i=1,2,\cdots,n)$ に対して

$$Q(\alpha,\beta) = \sum_{i=1}^{n} \{y_i - (\alpha + \beta x_i)\}^2$$

を最小にする α, β として以下の値が求まることが知られている(最小二乗法，問05-2*(1)).

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{T_{xy}}{T_{xx}} \left(= \frac{s_{xy}}{{s_x}^2} \right)$$ および $$\hat{\alpha} = \overline{y} - \hat{\beta}\,\overline{x}$$

回帰直線が点 $(\overline{x}, \overline{y})$ を通ることを意味する

$y = \hat{\alpha} + \hat{\beta} x$ を回帰式または回帰直線という．回帰式による y_i の推定値を $\hat{y}_i (= \hat{\alpha} + \hat{\beta} x_i)$ と表す.

(決定係数)

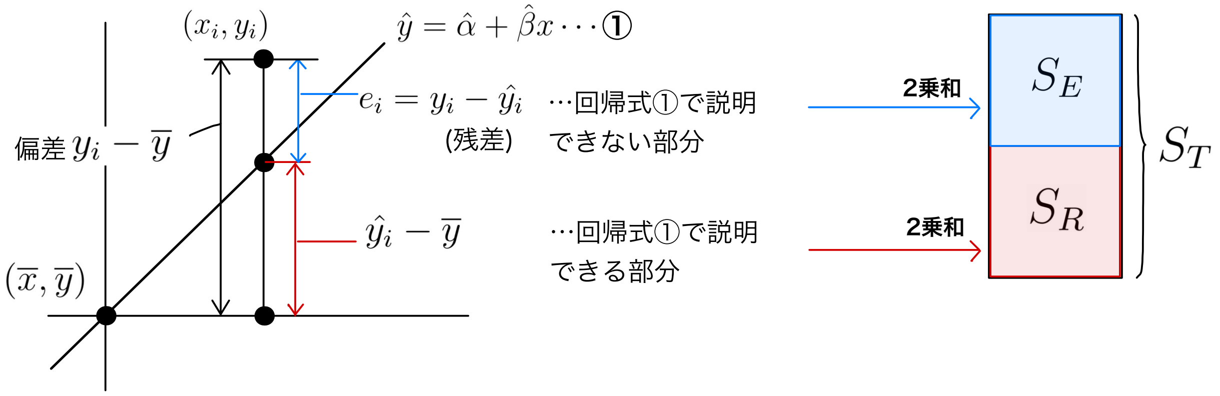
総平方和 $S_T = \sum_{i=1}^{n}(y_i - \overline{y})^2$ は、次のように分解できることが知られている(問05-2*(2)).

$$S_T = \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}_{\text{回帰平方和 } S_R} + \underbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}_{\text{残差平方和 } S_E}$$

S_T に対する S_R の割合を決定係数といい、推定された直線の当てはまり具合を表す(0以上1以下)

$$\text{決定係数 } R^2 = \frac{\text{回帰平方和 } S_R}{\text{総平方和 } S_T} = 1 - \frac{S_E}{S_T}$$

←これは x と y の相関係数の二乗 ${r_{xy}}^2$ に等しいことが知られている(問05-2*(3))



(自由度修正済み決定係数)

変数 y をk個の変数 $x_1, x_2, \cdots, x_k$ で説明するモデルを考える(重回帰モデル).

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

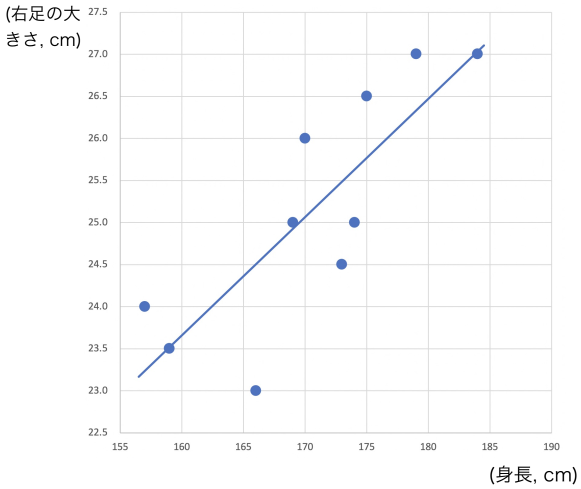
一般に，説明変数を増やすと決定係数は増大する．説明変数の個数の異なるモデルの比較のためには，次の自由度修正済み決定係数 R^{*2} が用いられる.

$$R^{*2} = 1 - \frac{S_E/(n-k-1)}{S_T/(n-1)}$$

(05-1) 10人の男子大学生の身長(x)と右足の大きさ(y)を測定すると，次のようになった.

	1	2	3	4	5	6	7	8	9	10	平均	分散
身長(xcm)	175	184	166	173	174	170	169	157	159	179	170.6	63.04
右足の大きさ(ycm)	26.5	27.0	23.0	24.5	25.0	26.0	25.0	24.0	23.5	27.5	25.2	2.06

また，xとyの共分散の値は9.28である.



(1) (x,y) の散布図および回帰直線は右のようになる．与えられたデータから求められる回帰直線として最も適当なものを選択肢から選べ.

【選択肢】

① y=0.076+0.1572x

② y=0.086+0.1472x

③ y=0.096+0.1372x

(2) 決定係数を求めよ．ただし，小数第4位を四捨五入せよ.

(3)回帰直線によって，身長190cmの男子大学生の右足の大きさを推定せよ．ただし，小数第2位を四捨五入せよ.

(05-2*)

$Q(\alpha,\beta)=\sum\{y_i-(\alpha+\beta x_i)\}^2$ は非負の2次式なので，これを最小化する値 $\hat{\alpha},\hat{\beta}$ は $Q(\alpha,\beta)$ を α,β を偏微分した式が0になるような値として求められる (この問で $\sum$ は $\sum_{i=1}^{n}$ を意味するとする).

$$\begin{cases}\dfrac{\partial Q}{\partial \alpha}=-2\sum\{y_i-(\alpha+\beta x_i)\}=0\\ \dfrac{\partial Q}{\partial \beta}=-2\sum x_i\{y_i-(\alpha+\beta x_i)\}=0\end{cases}\quad \text{すなわち}\quad \begin{cases}\sum\left\{y_i-(\hat{\alpha}+\hat{\beta}x_i)\right\}=0\cdots①\\ \sum x_i\left\{y_i-(\hat{\alpha}+\hat{\beta}x_i)\right\}=0\cdots②\end{cases}$$

(1) ①, ② を $\hat{\alpha},\hat{\beta}$ について解いて $\hat{\beta}=T_{xy}/T_{xx}$，$\hat{\alpha}=\overline{y}-\hat{\beta}\overline{x}$ を導け.

(2) $S_T=S_R+S_E$ を導け (①, ② を利用するとよい).

(3) $R^2={r_{xy}}^2$ を導け.