データ解析 (第14回)

安藤 和敏

静岡大学システム工学科

残差平方和 S_e の式変形 (ホワイトボードで)

$$
\begin{aligned}
S_e &= \sum_{i=1}^{n} e_i^2 \\
&= \sum_{i=1}^{n} \{y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2})\}^2 \\
&= \sum_{i=1}^{n} \{y_i - (\bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2})\}^2 \\
&= \sum_{i=1}^{n} \{y_i - \bar{y} - \hat{\beta}_1 (x_{i1} - \bar{x}_1) - \hat{\beta}_2 (x_{i2} - \bar{x}_2)\}^2
\end{aligned}
$$

残差平方和 S_e の式変形 (続き)

$$
\begin{aligned}
= & \sum_{i=1}^{n}(y_i - \bar{y})^2 + \hat{\beta}_1^2 \sum_{i=1}^{n}(\bar{x}_1 - x_{i1})^2 \\
& + \hat{\beta}_2^2 \sum_{i=1}^{n}(\bar{x}_2 - x_{i2})^2 - 2\hat{\beta}_1 \sum_{i=1}^{n}(y_i - \bar{y})(x_{i1} - \bar{x}_1) \\
& - 2\hat{\beta}_2 \sum_{i=1}^{n}(y_i - \bar{y})(x_{i2} - \bar{x}_2) \\
& + 2\hat{\beta}_1\hat{\beta}_2 \sum_{i=1}^{n}(x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)
\end{aligned}
$$

残差平方和 S_e の式変形 (続き)

$$
\begin{aligned}
&= S_{yy} + \hat{\beta}_1^2 S_{11} + \hat{\beta}_2^2 S_{22} - 2\hat{\beta}_1 S_{y1} - 2\hat{\beta}_2 S_{y2} \\
&\quad +2\hat{\beta}_1\hat{\beta}_2 S_{12} \\
&= S_{yy} + \hat{\beta}_1(\hat{\beta}_1 S_{11} + \hat{\beta}_2 S_{12}) + \hat{\beta}_2(\hat{\beta}_2 S_{22} + \hat{\beta}_1 S_{12}) \\
&\quad -2\hat{\beta}_1 S_{1y} - 2\hat{\beta}_2 S_{2y} \\
&= S_{yy} + \hat{\beta}_1 S_{1y} + \hat{\beta}_2 S_{2y} - 2\hat{\beta}_1 S_{1y} - 2\hat{\beta}_2 S_{2y} \\
&= S_{yy} - \hat{\beta}_1 S_{1y} - \hat{\beta}_2 S_{2y}
\end{aligned}
$$

母分散 σ^2 の推定量

$$\hat{\sigma}^2 = V_e = \frac{S_e}{n - 3}.$$

(2) 寄与率と自由度調整済み寄与率

平方和の分解(ホワイトボードで)

$$
\begin{aligned}
S_{yy} &= \sum_{i=1}^{n}(y_i - \bar{y})^2 \\
&= \sum_{i=1}^{n}\{y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}) \\
&\qquad + (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}) - \bar{y}\}^2 \\
&= \sum_{i=1}^{n}\{e_i + (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}) - \bar{y}\}^2 \\
&= \sum_{i=1}^{n} {e_i}^2 + \sum_{i=1}^{n}\{(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}) - \bar{y}\}^2 \\
&\quad + 2\sum_{i=1}^{n} e_i\{(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}) - \bar{y}\}
\end{aligned}
$$

$$\sum_{i=1}^{n} e_i\{(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}) - \bar{y}\} = 0$$

$$\begin{aligned}
&\sum_{i=1}^{n} e_i\{(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}) - \bar{y}\} \\
&= \sum_{i=1}^{n} e_i(\hat{\beta}_0 - \bar{y}) + \sum_{i=1}^{n} e_i \hat{\beta}_1 x_{i1} + \sum_{i=1}^{n} e_i \hat{\beta}_2 x_{i2} \\
&= (\hat{\beta}_0 - \bar{y}) \sum_{i=1}^{n} e_i + \hat{\beta}_1 \sum_{i=1}^{n} e_i x_{i1} + \hat{\beta}_2 \sum_{i=1}^{n} e_i x_{i2}.
\end{aligned}$$

(5.7) より $\sum_{i=1}^{n} e_i = 0$,
(5.8) より $\sum_{i=1}^{n} e_i x_{i1} = 0$,
(5.9) より $\sum_{i=1}^{n} e_i x_{i2} = 0$
であるから,

$$\sum_{i=1}^{n} e_i\{(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}) - \bar{y}\} = 0$$

が示された. この式は, 以下のようにも書けることに注意.

$$\sum_{i=1}^{n} e_i(\hat{y}_i - \bar{y}) = 0.$$

$$S_{yy} = \sum_{i=1}^{n} e_i{}^2 + \sum_{i=1}^{n} \{(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}) - \bar{y}\}^2 \tag{5.34}$$

$S_R = \hat{\beta}_1 S_{1y} + \hat{\beta}_2 S_{2y}$ と置くと，

$$S_e = S_{yy} - \hat{\beta}_1 S_{1y} - \hat{\beta}_2 S_{2y} \tag{5.32}$$

より，

$$S_{yy} = S_e + S_R \tag{5.35}$$.

この式と, (5.34) 式より,

$$S_R = \sum_{i=1}^{n} \{(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}) - \bar{y}\}^2$$

(5.36)
$$= \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2.$$

S_R を回帰による平方和と呼ぶ.

重相関係数

$\hat{y}_i$ と y_i の相関係数

$$R = \frac{\displaystyle\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\displaystyle\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\displaystyle\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}} \tag{5.37}$$

を重相関係数と呼ぶ.

$\sum_{i=1}^{n} e_i = 0$であったから,

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - \hat{y}_i) = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{y}_i = n\bar{y} - n\bar{\hat{y}}.$$

したがって,

$$\bar{\hat{y}} = \bar{y}.$$

$$
\begin{aligned}
R^2 &= \frac{\left(\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2} \\
&= \frac{\left(\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{y})\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2} \\
&= \frac{\left(\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{y})\right)^2}{S_{yy} S_R}
\end{aligned}
$$

$$
\begin{aligned}
\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{y}) &= \sum_{i=1}^{n}(y_i - \hat{y}_i + \hat{y}_i - \bar{y})(\hat{y}_i - \bar{y}) \\
&= \sum_{i=1}^{n}(e_i + \hat{y}_i - \bar{y})(\hat{y}_i - \bar{y}) \\
&= \sum_{i=1}^{n}\{e_i(\hat{y}_i - \bar{y}) + (\hat{y}_i - \bar{y})^2\} \\
&= \sum_{i=1}^{n} e_i(\hat{y}_i - \bar{y}) + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \\
&= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = S_R
\end{aligned}
$$

平方和の分解と寄与率

$$R^2 = \frac{S_R^2}{S_{yy}S_R} = \frac{S_R}{S_{yy}} = 1 - \frac{S_e}{S_{yy}} \qquad (5.38).$$

$$\underbrace{\sum_{i=1}^{n}(y_i - \bar{y})^2}_{S_{yy}} = \underbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}_{S_e} + \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}_{S_R}$$

y の全変動 ＝ 誤差による変動 + 回帰による変動

R^2 は, 寄与率 (または, 決定係数) と呼ばれ, 1 に近いほどよい.

自由度と自由度調整済寄与率 (補正 R2)

各平方和に対して, 以下のように自由度が対応する

平方和	自由度
S_{yy}	$\phi_T = n - 1$
S_R	$\phi_R = 2$
S_e	$\phi_e = n - 3$

$$R^{*2} = 1 - \frac{S_e/\phi_e}{S_{yy}/\phi_T} \tag{5.39}$$

は, 自由度調整済寄与率と呼ばれる.