第 4 回　単回帰モデル②

[3] 決定係数

（復習） y_i，フィット $\hat{y}_i$，残差 $\hat{u}_i$ の関係

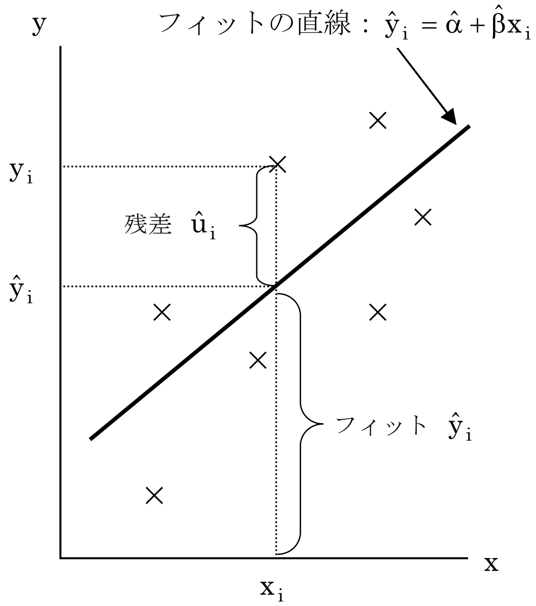
残差 $\hat{u}_i$ の定義は次のように書ける。

$$y_i = \hat{y}_i + \hat{u}_i \quad \cdots ①$$

つまり， y_i は二つの部分に分解できる。

$\hat{y}_i$ ： y_i の中で， x_i で説明できる部分

$\hat{u}_i$ ： y_i の中で， x_i で説明できない部分



①式は標本iについて成り立つ関係であるが，すべての標本について，直線の「あてはまりのよさ」の尺度を求めたい。①の両辺から $\overline{y}$ を引き，両辺を二乗した後，その結果をn個の標本について合計すると，次の関係が導ける。

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 + \sum_{i=1}^{n}\hat{u}_i^2 \quad \cdots ②$$

TSS	=	ESS	+	RSS
y の総変動		x で説明できる部分		x で説明できない部分

※ ①から②を導くときに必要な関係

$$\sum_{i=1}^{n}(\hat{y}_i - \overline{y})\hat{u}_i = \sum_{i=1}^{n}\hat{y}_i\hat{u}_i - \overline{y}\sum_{i=1}^{n}\hat{u}_i = \sum_{i=1}^{n}\hat{y}_i\hat{u}_i = \sum_{i=1}^{n}(\hat{\alpha} + \hat{\beta}x_i)\hat{u}_i = \hat{\alpha}\sum_{i=1}^{n}\hat{u}_i + \hat{\beta}\sum_{i=1}^{n}x_i\hat{u}_i = 0$$

決定係数 $R^2 = \dfrac{ESS}{TSS} = 1 - \dfrac{RSS}{TSS}$　　（$0 \le R^2 \le 1$）

R^2 が 1 に近い　⇒　x は y をうまく説明する

R^2 が 0 に近い　⇒　x は y をうまく説明できない

（決定係数についての注意事項）

・定数項を含まないモデルでは，同じ解釈ができない（②式が成り立たないため）

⇒ 特別の理由がなければ，いつもモデルに定数項を含める。

・従属変数 y が違うモデルについて，決定係数は比較できない。

（例）モデル 1：$y_i = \alpha + \beta x_i + u_i$

モデル 2：$\log(y_i) = \alpha + \beta x_i + u_i$

モデル 3：$\log(y_i) = \alpha + \beta \log(x_i) + u_i$

・計量モデルのよさは決定係数のよさだけで決まるわけではない（目的による）。

（例）時給 y と教育年数 x について，526 人のクロスセクションデータを使って，次のような推定結果を得た。これは非常に悪い結果なのだろうか？

$$\log(y_i) = 0.584 + 0.083\, x_i, \quad R^2 = 0.19$$

（例）先週の数値例について決定係数 R^2 を計算する。

[4] 最小二乗推定量の平均

まず，最小二乗推定量 $\hat{\alpha}$ と $\hat{\beta}$ は確率変数であることを確かめる。

y_i が入っているので，確率変数？

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

x_i は確率変数ではないので，非確率変数

$\hat{\beta}$ は確率変数 u_i から影響を受けるため，確率変数である。また，$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$ なので，$\hat{\alpha}$ もやはり確率変数である。したがって，$\hat{\alpha}$ と $\hat{\beta}$ は確率分布をもつ。

次の三つの仮定を置くとき，$\hat{\alpha}$ と $\hat{\beta}$ の平均は，定理 2.1 のように求められる。

(S1) $E(u_i) = 0$

(S2) x_i は確率変数ではない。

(S3) $S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 > 0$

定理 2.1（最小二乗推定量の不偏性）

仮定(S1)～(S3)が成り立つとき，$E(\hat{\alpha}) = \alpha$，$E(\hat{\beta}) = \beta$

（$E(\hat{\beta}) = \beta$ の意味：$\hat{\beta}$ の不偏性）

$\hat{\beta}$ は確率変数なので，n 個の標本から $\hat{\beta}$ を計算しても，真の値 β を当てられないかもしれない。しかし，誤差項 u が 0 のまわりに現れ，平均的には 0 であれば，$\hat{\beta}$ も β のまわりに現れ，平均的には β に等しいといえる。

（定理の証明）