

2 最小二乗法：直線の当てはめ

2.1 データの整理

■ 単純回帰モデル

- 基本モデル $Y = \alpha + \beta X$ (2.1)
- データ
 - クロス・セクション（横断面）データ Cross-Section Data
 - タイム・シリーズ（時系列）データ Time Series Data
 - パネルデータ（プーリングデータ）Panel Data (Pooling Data)
- 用語
 - Y: 従属変数 Dependent variable、被説明変数 Explained variable
 - X: 独立変数 Independent variable、説明変数 Explanatory variable

■ データ（標本）の平均と分散

- 標本平均 $\bar{X} = \frac{\sum X_i}{n}$ (2.2)

- 標本分散 $s_x^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$ (2.3)

- 標本標準偏差 $s_x = \sqrt{s_x^2}$ (2.4)

- 標本共分散 $s_{xy} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$ (2.5)

■ 重要表現

- 標本平均から $\sum X_i = n\bar{X}$ (2.6) $\rightarrow \sum (X_i - \bar{X}) = 0$ (2.7)

*平均周りの偏差の総和は 0

- 標本分散から $\sum (X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2$ (2.8)

$$= \sum (X_i - \bar{X})X_i \quad (2.9)$$

- 標本共分散から $\sum (X_i - \bar{X})(Y_i - \bar{Y}) = \sum X_i Y_i - n\bar{X}\bar{Y}$

$$= \sum (X_i - \bar{X})Y_i$$

$$= \sum (Y_i - \bar{Y})X_i \quad (2.10)$$

2.2 最小 2 乗法と回帰直線

■ 最小 2 乗法

- 当てはまり基準 残差 residual: $\tilde{u}_i = Y_i - \tilde{\alpha} - \tilde{\beta}X_i$ (2.11)

- 残差 2 乗和 $J = \sum \tilde{u}_i^2 = \sum (Y_i - \tilde{\alpha} - \tilde{\beta}X_i)^2$ (2.12)

*J を最小化問題として、正規方程式(2.16)から最小 2 乗推定値（推定量）を算出

[微分：(2.13)&(2.14)→最小化：(2.15)→正規方程式：(2.16)]

- ・ 回帰直線 「Y の X への回帰 (regress Y on X)」——単回帰 simple regression

$$\text{理論値 } \hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i \quad (2.18)$$

$$\text{実績値 } Y_i = \hat{Y}_i + \hat{u}_i = \hat{\alpha} + \hat{\beta}X_i + \hat{u}_i \quad (2.20)$$

- ・ 計算公式
$$\hat{\beta} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\text{Cov}(X_i, Y_i)}{\text{Var}(X_i)} \quad (2.23), (2.25)$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} \quad * \text{回帰直線は必ず変数の標本平均値を通る}$$

2.3 回帰直線の当てはまりの尺度：決定係数

- 回帰モデルの説明力
$$Y_i - \bar{Y} = \hat{Y}_i - \bar{Y} + \hat{u}_i \quad (2.26)$$

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum \hat{u}_i^2 \quad (2.31)$$

$$\text{SST} = \text{SSE} + \text{SSR}$$

- ・ 残差の性質
$$\sum \hat{u}_i = 0, \sum X_i \hat{u}_i = 0, \sum \hat{Y}_i \hat{u}_i = 0 \quad (2.27), (2.28), (2.29)$$

- 決定係数
$$R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} = 1 - \frac{\sum \hat{u}_i^2}{\sum (Y_i - \bar{Y})^2} \quad (2.32), (2.33)$$

$$= \frac{\text{SSE}}{\text{SST}} = 1 - \frac{\text{SSR}}{\text{SST}}$$

$$0 \leq R^2 \leq 1 \quad (2.35)$$

2.4 計算手順のまとめ