単回帰モデルの確率論的取り扱い

Ver.2022.09.05
松浦 真也

単回帰モデルでは，目的変数 y の値は，説明変数 x の値から確定的に決まる部分に，誤差（ノイズ）が加わっていると考えることによって，確率論的取り扱いが可能となる．具体的には，目的変数に対応して，下記で定まる確率変数 $Y_1, Y_2, \ldots, Y_n$ を考える．

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad (i = 1, 2, \ldots, n). \tag{a}$$

ここで，β_0, β_1 は定数とする．また，説明変数に相当する $x_1, x_2, \ldots, x_n$ も，値が確定している場合を考え，定数として扱う．さらに，ノイズに相当する $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ は，正規分布 $N(0, \sigma^2)$ に従う確率変数で，互いに独立とする．

確率変数 $Y_1, Y_2, \ldots, Y_n$ の実現値（標本）を $y_1, y_2, \ldots, y_n$ とし，確率変数 $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ の実現値（標本）を $e_1, e_2, \ldots, e_n$ とするとき，(a) 式は，

$$y_i = \beta_0 + \beta_1 x_i + e_i \qquad (i = 1, 2, \ldots, n) \tag{b}$$

となる．

回帰分析において，β_1 の推定値 $\widehat{\beta}_1$ は，次のように計算される．

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2, \quad S_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(Y_i - \overline{Y}), \quad \widehat{\beta}_1 = \frac{S_{xy}}{S_{xx}}.$$

このとき，$\widehat{\beta}_1 \sim N\left(\beta_1, \dfrac{\sigma^2}{S_{xx}}\right)$ であることを，以下に示す．

ステップ 0 まず，以下のことに留意する．

- $x_i, \overline{x}, S_{xx}$ は定数，$Y_i, \overline{Y}, S_{xy}, \widehat{\beta}_1$ は確率変数である．
- $\varepsilon_i \sim N(0, \sigma^2)$ なので，(a) 式より，$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ である．
- $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ は互いに独立なので，$Y_1, Y_2, \ldots, Y_n$ も互いに独立である．
- $\displaystyle\sum_{i=1}^{n}(x_i - \overline{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\overline{x} = \sum_{i=1}^{n} x_i - n\overline{x} = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0.$
- $\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})\overline{Y} = \left\{\sum_{i=1}^{n}(x_i - \overline{x})\right\}\overline{Y} = 0 \cdot \overline{Y} = 0$ なので，$S_{xy} = \displaystyle\sum_{i=1}^{n}(x_i - \overline{x})Y_i$ である．

  従って，$\widehat{\beta}_1 = \dfrac{1}{S_{xx}}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})Y_i$ である．

ステップ 1 $Y_1, Y_2, \ldots, Y_n$ は正規分布に従い，互いに独立なので，正規分布の再生性より，これらの線形和である $\widehat{\beta}_1$ も正規分布に従う．

ステップ 2 期待値の線形性より，

$$
\begin{aligned}
E[\widehat{\beta}_1] &= E\left[\frac{1}{S_{xx}}\sum_{i=1}^{n}(x_i-\overline{x})Y_i\right] = \frac{1}{S_{xx}}\sum_{i=1}^{n}(x_i-\overline{x})E[Y_i] = \frac{1}{S_{xx}}\sum_{i=1}^{n}(x_i-\overline{x})(\beta_0+\beta_1 x_i) \\
&= \frac{1}{S_{xx}}\sum_{i=1}^{n}(x_i-\overline{x})\{\beta_0+\beta_1(x_i-\overline{x})+\beta_1\overline{x}\} \\
&= \frac{1}{S_{xx}}\left\{\beta_0\sum_{i=1}^{n}(x_i-\overline{x})+\beta_1\sum_{i=1}^{n}(x_i-\overline{x})^2+\beta_1\overline{x}\sum_{i=1}^{n}(x_i-\overline{x})\right\} = \frac{\beta_1 S_{xx}}{S_{xx}} = \beta_1.
\end{aligned}
$$

ステップ 3 $Y_1, Y_2, \ldots, Y_n$ は互いに独立なので，

$$
\begin{aligned}
V[\widehat{\beta}_1] &= V\left[\frac{1}{S_{xx}}\sum_{i=1}^{n}(x_i-\overline{x})Y_i\right] = \frac{1}{S_{xx}^2}\sum_{i=1}^{n}(x_i-\overline{x})^2 V(Y_i) = \frac{1}{S_{xx}^2}\sum_{i=1}^{n}(x_i-\overline{x})^2\sigma^2 \\
&= \frac{S_{xx}\sigma^2}{S_{xx}^2} = \frac{\sigma^2}{S_{xx}}.
\end{aligned}
$$

以上より，$\widehat{\beta}_1$ は期待値が β_1，分散が $\dfrac{\sigma^2}{S_{xx}}$ の正規分布に従うことが分かった．すなわち，$\widehat{\beta}_1 \sim N\left(\beta_1, \dfrac{\sigma^2}{S_{xx}}\right)$ である．