回帰分析の基礎 2

crimsonbach

2004 年 9 月 8 日

まず前回の続きから始めよう。回帰係数の推定以外に β_1 および β_2 について、検定をしなければならない。それには推定量 $\hat{\beta}_1$ および $\hat{\beta}_2$ の標本分布を知る必要がある。誤差項 $u_1, \cdots, u_n$ が独立で正規分布に従うと仮定しよう。最初に $\hat{\beta}_2$ についてだが、 $\sum(X_i - \overline{X}) = 0$ より

$$\hat{\beta}_2 = \beta_2 + \frac{\sum(X_i - \overline{X})u_i}{\sum(X_i - \overline{X})^2}.$$

推定量 $\hat{\beta}_2$ は独立な正規分布に従う確率変数の和である。$\hat{\beta}_2$ の分布は正規分布で

$$N\Big(\beta_2, \frac{\sigma^2}{\sum(X_i - \overline{X})^2}\Big).$$

$(\hat{\beta}_2 - \beta_2)\sqrt{\sum(X_i - \overline{X})^2}\Big/\sigma$ は標準正規分布に従う。$\hat{\beta}_2$ の標準偏差（標準誤差）の推定量は

$$s.e.(\hat{\beta}_2) = \frac{s}{\sqrt{\sum(X_i - \overline{X})^2}}$$

$$t_2 = \frac{\hat{\beta}_2 - \beta_2}{s.e.(\hat{\beta}_2)}.$$

t_2 は自由度 $n-2$ の t 分布 $t(n-2)$ に従う。同様に $\hat{\beta}_1$ の分布は

$$N\Big(\beta_1, \frac{\sigma^2 \sum {X_i}^2}{n\sum(X_i - \overline{X})^2}\Big).$$

標準誤差は

$$s.e.(\hat{\beta}_1) = s\sqrt{\frac{\sum {X_i}^2}{n\sum(X_i - \overline{X})^2}}.$$

$$t_1 = \frac{\hat{\beta}_1 - \beta_1}{s.e.(\hat{\beta}_1)}.$$

続いて回帰係数の検定を行う。まず β_2 についてからだ。帰無仮説を

$$H_0 : \beta_2 = \alpha, \quad \alpha\text{は定数}.$$

対立仮説は状況に応じて

$$H_1 : \beta_2 \neq \alpha, (\text{両側検定})$$
$$H_1 : \beta_2 > \alpha, (\text{右片側検定})$$
$$H_1 : \beta_2 < \alpha, (\text{左片側検定})$$

まず $\hat{\beta}_2$ と $s.e.(\hat{\beta}_2)$ を求める。検定統計量

$$t_2 = \frac{(\hat{\beta}_2 - \alpha)}{s.e.(\hat{\beta}_2)}.$$

から、 t_2 、 $t_{\alpha/2}(n-2)$ 、 $t_\alpha(n-2)$ 、を比較して

$$\begin{array}{ll} H_1 : \beta_2 \neq \alpha & |t_2| > t_{\alpha/2}(n-2) \text{ の場合、帰無仮説を棄却} \\ H_1 : \beta_2 > \alpha & t_2 > t_\alpha(n-2) \text{ の場合、帰無仮説を棄却} \\ H_1 : \beta_2 < \alpha & t_2 < -t_\alpha(n-2) \text{ の場合、帰無仮説を棄却} \end{array}$$

最小二乗法以外の推定方法に最尤法（maximum likelihood method）があり、その推定量を最尤推定量（Maximun Likelihood Estimator : MLE）という。誤差項 $u_1, \cdots, u_n$ が独立で期待値が 0 、分散 σ^2 の正規分布 $N(0, \sigma^2)$ に従うとしよう。回帰モデルは

$$u_i = Y_i - \beta_1 - \beta_2 X_i.$$

推定量 $\beta_1, \beta_2, \sigma^2$ が既知なら、 $Y_i - \beta_1 - \beta_2 X_i$ は $N(0, \sigma^2)$ に従い、その確率密度関数は

$$f(X_i, Y_i, \beta_1, \beta_2, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\tfrac{(Y_i - \beta_1 - \beta_2 X_i)^2}{2\sigma^2}\right\}.$$

しかし、 β_1 、 β_2 、 σ^2 は未知なので、尤度関数と対数尤度はそれぞれ

$$L(\beta_1, \beta_2, \sigma^2) = \prod \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\tfrac{(Y_i - \beta_1 - \beta_2 X_i)^2}{2\sigma^2}\right\}$$
$$\ln L(\beta_1, \beta_2, \sigma^2) = -n(\ln\sqrt{2\pi} + \ln\sigma) - \sum\left\{\tfrac{(Y_i - \beta_1 - \beta_2 X_i)^2}{2\sigma^2}\right\}$$

β_1 および β_2 に関して $\ln L$ を最大にするということは、 $\sum(Y_i - \beta_1 - \beta_2 X_i)^2$ を最小にすることと同じである。この場合は β_1 および β_2 の最尤推定量は最小二乗推定量と一致し、

$$\hat{\beta}_2 = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})^2}, \qquad \hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X}.$$

また

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^2} \sum(Y_i - \beta_1 - \beta_2 X_i)^2.$$

より、 σ^2 の最尤推定量は

$$\hat{\sigma}^2 = \frac{\sum {e_i}^2}{n}$$

ただし、最小二乗法の場合は $s^2 = \frac{\sum {e_i}^2}{n-2}$ となる。対数最大尤度は

$$\ln L(\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2) = -\frac{n}{2}\left\{1 + \ln(2\pi) + \ln(\tfrac{\sum {e_i}^2}{n})\right\}.$$

実際の計算では $1 + \ln(2\pi)$ は一定なので、

$$\ln L^*(\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2) = -\frac{n}{2} \ln(\frac{\sum {e_i}^2}{n}).$$

を利用する。