応用計量経済分析 TA セッション
第 13 回：線形回帰モデルにおける仮説検定

TA：北村友宏 *

2016 年 1 月 26 日改訂

※以下では（行列でない）確率変数と実現値を小文字で表す。

1　線形回帰モデル

- 定数項を含む単回帰モデル（simple regression model）

$$y_i = \beta_0 + \beta_1 x_i + u_i,\ i = 1, 2, \ldots, n, \tag{1}$$
$$u_i \sim iidN\left(0, \sigma^2\right) \tag{2}$$

  を考える。
  - ⋆ y_i：被説明変数（explained variable）
    - ∗ e.g., 農産物の収穫量
  - ⋆ x_i：説明変数（explanatory variable）
    - ∗ e.g., 肥料の使用量
    - ∗ 簡単化のため、非確率変数とする。
  - ⋆ u_i：誤差項（disturbance term）
  - ⋆ β_0, β_1：係数（coefficient）
    - ∗ β_0 は定数項（constant term）、β_1 は傾き。
    - ∗ e.g., 肥料の使用量が 1 多くなると、農産物の収穫量は β_1 増加する。
  - ⋆ iid は独立に同一の確率分布に従う（independently and identically distributed）という意味。
    ⇒誤差項 $u_1, u_2, \ldots, u_n$ は互いに独立に $N\left(0, \sigma^2\right)$ に従っている。
- 確認
  - ⋆ y_i と x_i は既知（データとして与えられている）。
  - ⋆ β_0 と β_1 は未知（推定する必要がある）。
  - ⋆ u_i も未知。
- 誤差項の推定値を残差（residual）とよぶ。
- 残差平方和（残差二乗和）が最小となるようにパラメータ推定値（ここでは $\hat{\beta}_0, \hat{\beta}_1$）を求める方法を（通常の）最小二乗法（Ordinary Least Squares, OLS）とよぶ。

* Email: kitamu.tom@gmail.com　URL: http://tomkitamura.html.xdomain.jp

- OLS によって推定されたパラメータを **OLS 推定量**（OLS estimator）とよび、その実現値を **OLS 推定値**（OLS estimate）とよぶ。
  - ⋆ 上記の $\hat{\beta}_0$ と $\hat{\beta}_1$ が OLS 推定量（実現値であれば OLS 推定値）。
- x_i と y_i の間に相関があるかを調べたい。
  - ⋆ e.g.,「肥料の使用量」と「農産物の収穫量」の間に相関があるか？

  ⇒上記のモデルで、$\beta_1 = 0$ なら相関がなく、$\beta_1 \neq 0$ なら相関がある。

  ⇒ $\beta_1 = 0$ を H_0 とする仮説検定を行えば、β_1 が 0 かどうかを検証できる。

2　仮説検定

※ β_1 に関する検定のみ説明する。

両側検定

$$H_0 : \beta_1 = 0,$$
$$H_1 : \beta_1 \neq 0$$

を行うための検定統計量を導出する。

例題 1. (1) と (2) からなる単回帰モデルにおいて、β_1 の OLS 推定量 $\hat{\beta}_1$ が正規分布に従うことを証明しなさい。ただし、

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \overline{x})u_i}{(n-1)s_x^2},$$
$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

である。

（証明）(2) から、$u_i \sim N\left(0, \sigma^2\right)$（$iid$ の表記は省略）であり、また x_i は非確率変数である。（証明中断）

· ·

- **参考：正規分布の性質** 確率変数 $X \sim N\left(\mu, \sigma^2\right)$ と定数 a, b について、

$$aX + b \sim N\left(\underbrace{a\mu + b}_{\text{期待値}}, \underbrace{a^2\sigma^2}_{\text{分散}}\right)$$

  が成り立つ。この例題では、x_i が非確率変数であることから、$\dfrac{(x_i - \overline{x})}{(n-1)s_x^2}$ も非確率変数となる。非確率変数 $\dfrac{(x_i - \overline{x})}{(n-1)s_x^2}$ を定数として扱い、それに確率変数 $u_i \sim N\left(0, \sigma^2\right)$ を掛けたもの

$$\underbrace{\frac{(x_i - \overline{x})}{(n-1)s_x^2}}_{a\text{ に相当}} \cdot \underbrace{u_i}_{X\text{ に相当}}$$

  を考えると、

$$a \text{ を} \frac{(x_i - \overline{x})}{(n-1)s_x^2}, X \text{ を } u_i, b \text{ を } 0, \mu\text{を } 0, \sigma^2\text{を}\sigma^2$$

として、上記の正規分布の性質を使うことができる。$\dfrac{(x_i-\overline{x})}{(n-1)s_x^2}\cdot u_i$ に正規分布の性質を適用すると、

$$\underbrace{\frac{(x_i-\overline{x})}{(n-1)s_x^2}}_{a\text{ に相当}}\cdot\underbrace{u_i}_{X\text{ に相当}}+\underbrace{0}_{b\text{ に相当}}\sim N\left(\underbrace{\frac{(x_i-\overline{x})}{(n-1)s_x^2}}_{a\text{ に相当}}\cdot\underbrace{0}_{\mu\text{に相当}}+\underbrace{0}_{b\text{ に相当}},\underbrace{\left[\frac{(x_i-\overline{x})}{(n-1)s_x^2}\right]^2}_{a^2\text{に相当}}\cdot\underbrace{\sigma^2}_{\sigma^2\text{に相当}}\right)$$

が成り立つ。

....

（証明再開）よって、正規分布の性質より、任意の i について、

$$\frac{(x_i-\overline{x})}{(n-1)s_x^2}\cdot u_i\sim N\left(0,\left[\frac{(x_i-\overline{x})}{(n-1)s_x^2}\right]^2\cdot\sigma^2\right)$$

である。（証明中断）

....

- 参考：正規分布の再生性（確率変数が n 個の場合） 互いに独立な確率変数 $X_1,X_2,...,X_n$ がそれぞれ $N\left(\mu_1,\sigma_1^2\right),N\left(\mu_2,\sigma_2^2\right),...,N\left(\mu_n,\sigma_n^2\right)$ に従うとき、

$$X_1+X_2+\cdots+X_n\sim N\left(\underbrace{\mu_1+\mu_2+\cdots+\mu_n}_{\text{期待値}},\underbrace{\sigma_1^2+\sigma_2^2+\cdots+\sigma_n^2}_{\text{分散}}\right)$$

が成り立つ。この例題では、$u_1,u_2,...,u_n$ は互いに独立な確率変数なので、それらに非確率変数を掛けた

$$\frac{(x_1-\overline{x})}{(n-1)s_x^2}\cdot u_1,\frac{(x_2-\overline{x})}{(n-1)s_x^2}\cdot u_2,...,\frac{(x_n-\overline{x})}{(n-1)s_x^2}\cdot u_n$$

も互いに独立な確率変数。

$$X_1=\frac{(x_1-\overline{x})}{(n-1)s_x^2}\cdot u_1,X_2=\frac{(x_2-\overline{x})}{(n-1)s_x^2}\cdot u_2,...,X_n=\frac{(x_n-\overline{x})}{(n-1)s_x^2}\cdot u_n$$

とすれば、上記の正規分布の再生性を使うことができる。これら n 個の確率変数の期待値はそれぞれ、

$$\begin{aligned}
E(X_1)&=E\left[\frac{(x_1-\overline{x})}{(n-1)s_x^2}\cdot u_1\right]=\frac{(x_1-\overline{x})}{(n-1)s_x^2}\cdot\underbrace{E(u_1)}_{=0}=\underbrace{0}_{\mu_1\text{に相当}},\\
E(X_2)&=E\left[\frac{(x_2-\overline{x})}{(n-1)s_x^2}\cdot u_2\right]=\frac{(x_2-\overline{x})}{(n-1)s_x^2}\cdot\underbrace{E(u_2)}_{=0}=\underbrace{0}_{\mu_2\text{に相当}},\\
&\vdots\\
E(X_n)&=E\left[\frac{(x_n-\overline{x})}{(n-1)s_x^2}\cdot u_n\right]=\frac{(x_n-\overline{x})}{(n-1)s_x^2}\cdot\underbrace{E(u_n)}_{=0}=\underbrace{0}_{\mu_n\text{に相当}}
\end{aligned}$$

であり、分散はそれぞれ、

$$
\begin{aligned}
V(X_1) &= V\left[\frac{(x_1-\overline{x})}{(n-1)s_x^2}\cdot u_1\right] = \left[\frac{(x_1-\overline{x})}{(n-1)s_x^2}\right]^2 \cdot \underbrace{V(u_1)}_{=\sigma^2} = \underbrace{\left[\frac{(x_1-\overline{x})}{(n-1)s_x^2}\right]^2 \cdot \sigma^2}_{\sigma_1^2\text{に相当}}, \\
V(X_2) &= V\left[\frac{(x_2-\overline{x})}{(n-1)s_x^2}\cdot u_2\right] = \left[\frac{(x_2-\overline{x})}{(n-1)s_x^2}\right]^2 \cdot \underbrace{V(u_2)}_{=\sigma^2} = \underbrace{\left[\frac{(x_2-\overline{x})}{(n-1)s_x^2}\right]^2 \cdot \sigma^2}_{\sigma_2^2\text{に相当}}, \\
&\vdots \\
V(X_n) &= V\left[\frac{(x_n-\overline{x})}{(n-1)s_x^2}\cdot u_n\right] = \left[\frac{(x_n-\overline{x})}{(n-1)s_x^2}\right]^2 \cdot \underbrace{V(u_n)}_{=\sigma^2} = \underbrace{\left[\frac{(x_n-\overline{x})}{(n-1)s_x^2}\right]^2 \cdot \sigma^2}_{\sigma_n^2\text{に相当}}
\end{aligned}
$$

となる。

$$
\frac{(x_1-\overline{x})}{(n-1)s_x^2}\cdot u_1, \frac{(x_2-\overline{x})}{(n-1)s_x^2}\cdot u_2, \ldots, \frac{(x_n-\overline{x})}{(n-1)s_x^2}\cdot u_n
$$

を合計し、正規分布の再生性を適用すると、

$$
\begin{aligned}
&\underbrace{\frac{(x_1-\overline{x})}{(n-1)s_x^2}\cdot u_1}_{X_1\text{に相当}} + \underbrace{\frac{(x_2-\overline{x})}{(n-1)s_x^2}\cdot u_2}_{X_2\text{に相当}} + \cdots + \underbrace{\frac{(x_n-\overline{x})}{(n-1)s_x^2}\cdot u_n}_{X_n\text{に相当}} \\
&\sim N\left(\underbrace{0}_{\mu_1\text{に相当}} + \underbrace{0}_{\mu_2\text{に相当}} + \cdots + \underbrace{0}_{\mu_n\text{に相当}}, \underbrace{\left[\frac{(x_1-\overline{x})}{(n-1)s_x^2}\right]^2 \cdot \sigma^2}_{\sigma_1^2\text{に相当}} + \underbrace{\left[\frac{(x_2-\overline{x})}{(n-1)s_x^2}\right]^2 \cdot \sigma^2}_{\sigma_2^2\text{に相当}} + \cdots + \underbrace{\left[\frac{(x_n-\overline{x})}{(n-1)s_x^2}\right]^2 \cdot \sigma^2}_{\sigma_n^2\text{に相当}}\right)
\end{aligned}
$$

が成り立つ。総和記号を用いると、

$$
\sum_{i=1}^{n} \frac{(x_i-\overline{x})}{(n-1)s_x^2}\cdot u_i \sim N\left(\underbrace{\sum_{i=1}^{n} 0}_{=0}, \sum_{i=1}^{n}\left[\frac{(x_i-\overline{x})}{(n-1)s_x^2}\right]^2 \cdot \sigma^2\right)
$$

と表すことができる。

....

（証明再開）$\dfrac{(x_i-\overline{x})}{(n-1)s_x^2}\cdot u_i$ を $i=1,2,\ldots,n$ について合計すると、$u_1,u_2,\ldots,u_n$ は互いに独立なので、正規分布の再生性より、

$$
\sum_{i=1}^{n} \frac{(x_i-\overline{x})}{(n-1)s_x^2}\cdot u_i \sim N\left(0, \sum_{i=1}^{n}\left[\frac{(x_i-\overline{x})}{(n-1)s_x^2}\right]^2 \cdot \sigma^2\right)
$$

となる。ここで、

$$\sum_{i=1}^{n}\frac{(x_i-\overline{x})}{(n-1)s_x^2}\cdot u_i=\frac{1}{(n-1)s_x^2}\cdot\sum_{i=1}^{n}(x_i-\overline{x})u_i=\frac{\sum_{i=1}^{n}(x_i-\overline{x})u_i}{(n-1)s_x^2},$$

$$\sum_{i=1}^{n}\left[\frac{(x_i-\overline{x})}{(n-1)s_x^2}\right]^2\cdot\sigma^2=\frac{1}{\left[(n-1)s_x^2\right]^2}\cdot\sigma^2\underbrace{\sum_{i=1}^{n}(x_i-\overline{x})^2}_{=(n-1)s_x^2}=\frac{\sigma^2(n-1)s_x^2}{\left[(n-1)s_x^2\right]^2}=\frac{\sigma^2}{(n-1)s_x^2}$$

なので、

$$\frac{\sum_{i=1}^{n}(x_i-\overline{x})u_i}{(n-1)s_x^2}\sim N\left(0,\frac{\sigma^2}{(n-1)s_x^2}\right)$$

である。（証明中断）

……………………………………………………………………………………

- $\boxed{\text{参考：正規分布の性質・再掲}}$確率変数 $X\sim N\left(\mu,\sigma^2\right)$ と定数 a,b について、

$$aX+b\sim N\left(\underbrace{a\mu+b}_{\text{期待値}},\underbrace{a^2\sigma^2}_{\text{分散}}\right)$$

が成り立つ。この例題では、すでに見たように、$u_1,u_2,\ldots,u_n$ は確率変数なので、それらに非確率変数を掛けた

$$\frac{(x_1-\overline{x})}{(n-1)s_x^2}\cdot u_1,\frac{(x_2-\overline{x})}{(n-1)s_x^2}\cdot u_2,\ldots,\frac{(x_n-\overline{x})}{(n-1)s_x^2}\cdot u_n$$

も確率変数。よって、これら（確率変数）を合計した

$$\sum_{i=1}^{n}\frac{(x_i-\overline{x})}{(n-1)s_x^2}\cdot u_i=\frac{\sum_{i=1}^{n}(x_i-\overline{x})u_i}{(n-1)s_x^2}\sim N\left(0,\frac{\sigma^2}{(n-1)s_x^2}\right)$$

も確率変数。これに β_1 を足したもの

$$\underbrace{\frac{\sum_{i=1}^{n}(x_i-\overline{x})u_i}{(n-1)s_x^2}}_{X\text{ に相当}}+\underbrace{\beta_1}_{b\text{ に相当}}=\underbrace{\beta_1}_{b\text{ に相当}}+\underbrace{\frac{\sum_{i=1}^{n}(x_i-\overline{x})u_i}{(n-1)s_x^2}}_{X\text{ に相当}}$$

は、問題文から $\hat{\beta}_1$ となる。上式を考えると、

$$a\text{ を }1,X\text{ を}\frac{\sum_{i=1}^{n}(x_i-\overline{x})u_i}{(n-1)s_x^2},b\text{ を}\beta_1,\mu\text{を }0,\sigma^2\text{を}\frac{\sigma^2}{(n-1)s_x^2}$$

として、上記の正規分布の性質を使うことができる。$\hat{\beta}_1=\beta_1+\frac{\sum_{i=1}^{n}(x_i-\overline{x})u_i}{(n-1)s_x^2}$ に正規分布の性質を適用すると、

$$\underbrace{1}_{a\text{ に相当}}\cdot\underbrace{\frac{\sum_{i=1}^{n}(x_i-\overline{x})u_i}{(n-1)s_x^2}}_{X\text{ に相当}}+\underbrace{\beta_1}_{b\text{ に相当}}\sim N\left(\underbrace{1}_{a\text{ に相当}}\cdot\underbrace{0}_{\mu\text{に相当}}+\underbrace{\beta_1}_{b\text{ に相当}},\underbrace{1^2}_{a^2\text{に相当}}\cdot\underbrace{\frac{\sigma^2}{(n-1)s_x^2}}_{\sigma^2\text{に相当}}\right)$$

が成り立つ。

(証明再開)よって、正規分布の性質より、

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \overline{x})u_i}{(n-1)s_x^2} \sim N\left(\beta_1, \frac{\sigma^2}{(n-1)s_x^2}\right)$$

となる。
したがって、$\hat{\beta}_1$ は正規分布に従う(証明終)。

$\hat{\beta}_1$ を標準化すると、

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\sigma^2}{(n-1)s_x^2}}} \sim N(0,1)$$

となる(証明略)。
⇒ u_i の分散(**誤差分散**)σ^2 が入っている。σ^2 は通常、**未知**。
⇒ σ^2 の不偏推定量を $\hat{\sigma}^2$ とする。$\hat{\sigma}^2$ を用いた式で上記の統計量を書き換えると、

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{(n-1)s_x^2}}} \sim t(n-2)$$

となる(証明略)。

- $\sqrt{\dfrac{\hat{\sigma}^2}{(n-1)s_x^2}}$ は標準誤差。
- 自由度 $n-2$ の 2 は、推定するパラメータの数(ここでは β_0 と β_1)。
  - ⋆ もし定数項がなければ自由度は $n-1$ となり、定数項以外に説明変数が 2 つあれば自由度は $n-3$ となる。

⇒標準誤差を $SE\left(\hat{\beta}_1\right)$ として、<u>$H_0: \beta_1 = 0$ **が真であると仮定すると、**</u>H_0 のもとでの検定統計量は、

$$\frac{\hat{\beta}_1 - 0}{SE\left(\hat{\beta}_1\right)} \sim t(n-2),$$

つまり、

$$\frac{\hat{\beta}_1}{SE\left(\hat{\beta}_1\right)} \sim t(n-2).$$

- 定数項 β_0 に関する検定統計量も同様の方法で導出できる。

⇒続いて有意水準の設定、受容域・棄却域・有意点の導出、検定統計値の計算をすれば、β_1 が 0 かどうかを検定できる。