第 5 回：単回帰分析
【教科書第 5 章第 1 節～第 2 節，第 4 節】

北村　友宏

2025 年 10 月 28 日

本日の内容

1. 単回帰モデル

2. 通常の最小二乗法（OLS）

統計学と計量経済学での変数の表記の違い

- 統計学では確率変数を大文字で，実現値を小文字で表す．

$$\Updownarrow$$

- 計量経済学では確率変数と実現値を区別せず，行列を大文字で，ベクトルとスカラーを小文字で表す．

単回帰モデル

大きさ n の 2 変量データ
$((y_1, x_1), (y_2, x_2), \cdots, (y_n, x_n))$ を用いて，

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + u_i, \\ E(u_i \mid x_i) &= 0, \\ E(u_i u_j \mid x_i) &= 0 \quad (i \neq j), \\ V(u_i \mid x_i) &= \sigma^2, \\ i &= 1, 2, \cdots, n \end{aligned}$$

を推定することを考える．
これを推定すれば，2 つの変数間の関係（x_i が増加すると y_i はどの程度変化する傾向があるか？）を定量的に検証できる．

$E(u_i \mid x_i) = 0$ の仮定より，

$$E(y_i \mid x_i) = \beta_0 + \beta_1 x_i.$$

$\Rightarrow$ これは x_i が与えられたときの y_i の条件付き期待値.

- $E(y_i \mid x_i)$ を求めることを，y_i を x_i に回帰する（regress）という.
- $E(y_i \mid x_i)$ を与える式を回帰モデル（regression model）という.
- x_i の 1 次式で表される回帰モデルを線形回帰モデル（linear regression model）という.
- 定数項以外の説明変数が 1 つである回帰モデルを単回帰モデル（simple regression model）という.

- 説明される側の変数を被説明変数（explained variable）という．
  - 従属変数（dependent variable）ともいう．
- 説明する側の変数を説明変数（explanatory variable）という．
  - 独立変数（independent variable）ともいう．
- 回帰モデルにおける係数を回帰係数（regression coefficient）という．
- 個体や時点に依存せず，説明変数にも依存しない項を定数項（constant term）という．
- $u_i = y_i - E(y_i \mid x_i)$ を誤差項（error term）という．
  - 撹乱項（disturbance term）ともいう．

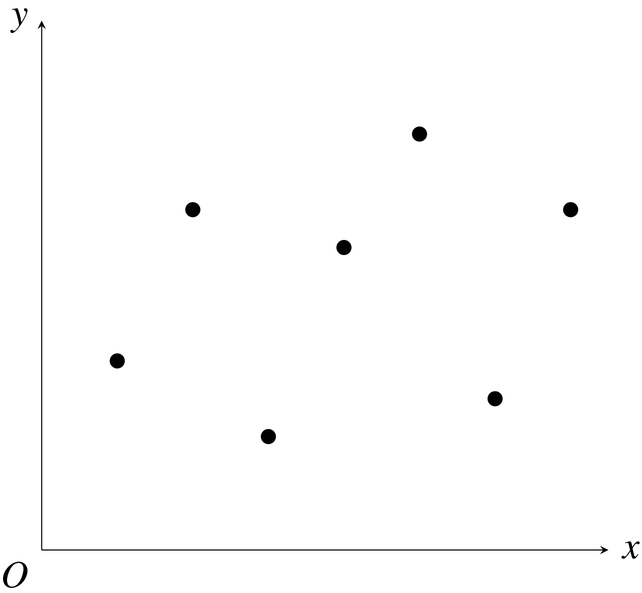
$y_i = \beta_0 + \beta_1 x_i + u_i$ において，

- y_i：被説明変数
  - e.g., 年収
- x_i：説明変数
  - e.g., 修学年数
- β_0, β_1：回帰係数
  - 特に，β_0 は定数項．
- u_i：誤差項

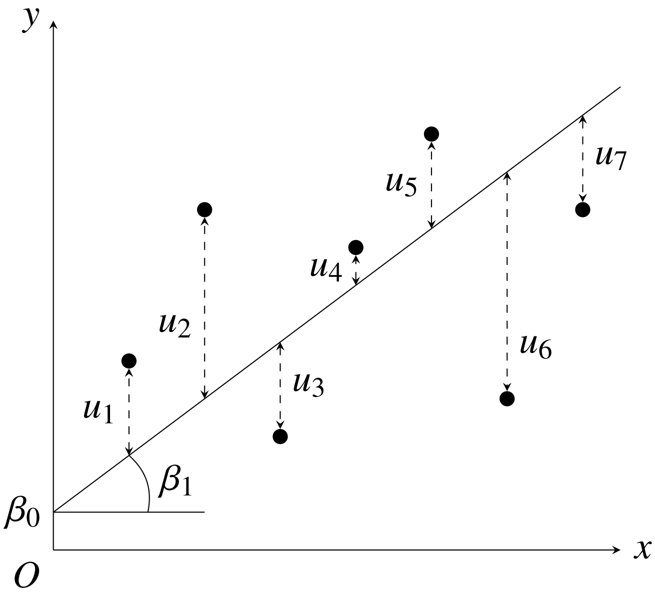
説明変数 x_i は確率的（stochastic）とする．

$$\Downarrow$$

β_0 と β_1 を求める（推定する）には？



y
x
O


y
x
O
β_0
β_1
u_1
u_2
u_3
u_4
u_5
u_6
u_7

モデルを
$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i$$
と書き換える.

- 誤差項の推定値を残差（residual）という.
- e_i：残差
    - 誤差項 u_i とは別物.

そして，残差二乗和
$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$
が最小になるような $\hat{\beta}_0$ と $\hat{\beta}_1$ を求める.

- 残差二乗和 $\sum_{i=1}^{n} e_i^2$ が最小になるように回帰係数を求める方法を通常の最小二乗法（Ordinary Least Squares, OLS）という.

- OLS によって推定される統計量を OLS 推定量（OLS estimator）といい，その実現値を OLS 推定値（OLS estimate）という．

この場合の OLS 推定量は，

$$
\begin{aligned}
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x}, \\
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}.
\end{aligned}
$$

- $\bar{y} = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} y_i,\ \bar{x} = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} x_i.$

（導出方法は付録参照）

OLS 推定における仮定（単回帰の場合）

- 「説明変数を所与とした誤差項の条件付き期待値」と，「誤差項の条件なし期待値」が等しく，その値は 0.
  - $E(u_i \mid x_i) = E(u_i) = 0.$

  $\Longrightarrow$ 説明変数と誤差項は平均独立で，誤差項の期待値は 0.

- 説明変数を所与として，誤差項の条件付き分散は一定で，異なる個体の誤差項同士は無相関.
  - $V(u_i \mid x_i) = \sigma^2.$
  - $E(u_i u_j \mid x_i) = 0 \quad (i \neq j).$

- 説明変数を所与として，誤差項は正規分布に従う.
  - $u_i \mid x_i \sim N\left(0, \sigma^2\right).$

実証分析例：ミンサー方程式の推定

「修学年数が増えると，年収がどれだけ増えるのか」を分析するためのモデル（ミンサー方程式）

$$income_i = \beta_0 + \beta_1 yeduc_i + u_i$$

- $income_i$：年収（万円）
- $yeduc_i$：修学年数（年）
- i：個人番号

を推定する．

➥ 「年収」を「修学年数」に回帰する．

- 対数変換していないものをレベル（level）という.
- 被説明変数と説明変数がともにレベルであるモデルを「レベル＝レベル・モデル」ということがある.

➥ 前スライドのモデルは，被説明変数も説明変数も対数変換していないため，「レベル＝レベル・モデル」である.

レベル＝レベル・モデル推定結果

gretl: モデル1

ファイル 編集(E) 検定(T) 保存(S) グラフ(G) 分析(A) LaTeX

```
モデル 1: 最小二乗法(OLS), 観測: 1-4327
従属変数: income

                 係数      標準誤差     t値         p値
  ----------------------------------------------------------
  const       -56.8928    19.3568    -2.939    0.0033     ***
  yeduc        23.1510     1.38425   16.72     6.85e-061  ***

Mean dependent var    263.9040   S.D. dependent var   176.5552
Sum squared resid     1.27e+08   回帰の標準誤差              171.1286
R-squared             0.060744   Adjusted R-squared   0.060527
F(1, 4325)            279.7095   P-value(F)           6.85e-61
Log-likelihood       -28389.98   Akaike criterion     56783.96
Schwarz criterion     56796.70   Hannan-Quinn         56788.46
```

出力結果の見方

- 係数: 回帰係数推定値
- 標準誤差: 回帰係数の標準誤差
  - 第 8 回授業で説明
- t 値: 「回帰係数が 0」という帰無仮説の両側 t 検定における検定統計値（t 値）
  - 第 8 回授業で説明
- p 値: 両側 p 値
  - 第 8 回授業で説明
- 回帰の標準誤差：誤差項の標準偏差の推定値
- R-squared: 決定係数

誤差項の分散の推定

定数項のある単回帰モデルの場合，誤差項 u_i の（条件付き）分散

$$V(u_i \mid x_i) = \sigma^2,$$

は，以下のように推定できる．

- 誤差項の分散の推定量：

$$s^2 = \frac{\sum_{i=1}^n e_i^2}{n-2}.$$

この s^2 は不偏性をもつ，すなわち $E(s^2) = \sigma^2$ となることが知られている．（証明は省略）

誤差項の標準偏差の推定

定数項のある単回帰モデルの場合，誤差項 u_i の（条件付き）標準偏差 σ は，以下のように推定できる．

- 標準偏差の推定値を標準誤差（standard error）という．
- 誤差項の標準偏差の推定量：

$$s = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}}.$$

この式の n と $e_1, e_2, \cdots, e_n$ に具体的な値を代入すれば，誤差項の標準偏差の推定値（誤差項の標準誤差）を計算できる．

決定係数

決定係数（R-squared）は，

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}.$$

- 定数項ありの単回帰の場合，$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$.
- 意味 モデルの当てはまりの良さ（説明変数で，被説明変数の動きのうち，どの程度の割合を説明できているか）
- $0 \leq R^2 \leq 1$.
  - $R^2 = 0$: 全く説明できていない.
  - $R^2 = 1$: 完全に説明できている.

  $\Rightarrow$ $R^2 = 0$ や $R^2 = 1$ になることは，実際の実証分析ではまず起こり得ない.

モデル推定結果

- 修学年数の係数
  - 23.151（符号は正）
    ➡ 修学年数が 1 年長くなると，年収が平均して 23.151 万円（231,510 円）高くなる傾向がある．
- 定数項
  - −56.8928（符号は負）
- 誤差項の標準誤差
  - 171.1286
- 決定係数
  - $R^2 = 0.060744.$
    ➡ 「年収」の動きの約 6.1%を「修学年数」の動き で説明できている．

今日のキーワード

回帰する，回帰モデル，線形回帰モデル，単回帰モデル，被説明変数，説明変数，回帰係数，定数項，誤差項，残差，通常の最小二乗法（OLS），OLS 推定量，OLS 推定値，レベル，標準誤差，決定係数

次回までの準備

- 今回の講義スライドを読み直す.
- 「提出課題 2」に取り組む.
- 教科書第 5 章第 3 節，第 5 節を読む.

付録：「（定数項を含む）単回帰モデル」の OLS 推定量の導出

残差二乗和最小化問題は，

$$\min_{(\hat{\beta}_0,\hat{\beta}_1)\in\mathbb{R}^2} \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2.$$

1 階条件は，

$$
\begin{aligned}
&\frac{\partial \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{\partial \hat{\beta}_0} = 0 \\
\Leftrightarrow &\sum_{i=1}^{n} 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \cdot (-1) = 0 \\
\Leftrightarrow &\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0, \qquad (1) \\
&\frac{\partial \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{\partial \hat{\beta}_1} = 0 \\
\Leftrightarrow &\sum_{i=1}^{n} 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \cdot (-x_i) = 0 \\
\Leftrightarrow &\sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0. \qquad (2)
\end{aligned}
$$

(1) より，

$$
\begin{aligned}
\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 &\Leftrightarrow \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{\beta}_0 - \sum_{i=1}^{n} \hat{\beta}_1 x_i = 0 \\
&\Leftrightarrow \sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} x_i = 0 \\
&\Leftrightarrow \sum_{i=1}^{n} y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i = n\hat{\beta}_0 \\
&\Leftrightarrow \hat{\beta}_0 = \frac{1}{n} \sum_{i=1}^{n} y_i - \hat{\beta}_1 \cdot \frac{1}{n} \sum_{i=1}^{n} x_i .
\end{aligned}
$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

とすると，$\hat{\beta}_0$ は，

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \tag{1'}$$

(2) と (1) より，

$$
\begin{aligned}
&\sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \\
\Leftrightarrow &\sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) - \bar{x} \underbrace{\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)}_{\text{(1) より，0 となる}} = 0 \\
\Leftrightarrow &\sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) - \sum_{i=1}^{n} \bar{x}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \\
\Leftrightarrow &\sum_{i=1}^{n} \left\{ x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) - \bar{x}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \right\} = 0 \\
\Leftrightarrow &\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0.
\end{aligned}
$$

(1') を代入すると，

$$
\begin{aligned}
& \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \\
\Leftrightarrow & \sum_{i=1}^{n}(x_i - \bar{x})\left\{y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i\right\} = 0 \\
\Leftrightarrow & \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y} + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i) = 0 \\
\Leftrightarrow & \sum_{i=1}^{n}(x_i - \bar{x})\left\{y_i - \bar{y} - \hat{\beta}_1 (x_i - \bar{x})\right\} = 0
\end{aligned}
$$

$$
\begin{aligned}
&\Leftrightarrow \sum_{i=1}^{n} \left\{(x_i - \bar{x})(y_i - \bar{y}) - (x_i - \bar{x})\hat{\beta}_1(x_i - \bar{x})\right\} = 0 \\
&\Leftrightarrow \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) - \sum_{i=1}^{n}(x_i - \bar{x})\hat{\beta}_1(x_i - \bar{x}) = 0 \\
&\Leftrightarrow \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) - \hat{\beta}_1 \sum_{i=1}^{n}(x_i - \bar{x})^2 = 0 \\
&\Leftrightarrow \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \hat{\beta}_1 \sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&\Leftrightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.
\end{aligned}
$$