

第 2 章: 単回帰モデル

Jeffrey Wooldridge (2018).
Introductory Econometrics: A Modern Approach
Seventh Edition. Cengage Learning.

2026-01-23

準備

必要なパッケージの読み込み

▶ wooldridge パッケージの読み込み

```
library(wooldridge)
```

2-1 単回帰モデルの定義

単回帰モデル

- ▶ **単回帰モデル** (simple regression model) は、2つの変数 x と y の関係を以下のように定義する。

$$y = \beta_0 + \beta_1 x + u$$

- ▶ y : **従属変数** (dependent variable)、被説明変数、目的変数
- ▶ x : **独立変数** (independent variable)、説明変数、共変量
- ▶ u : **誤差項** (error term)、攪乱項
- ▶ β_0 : **切片パラメータ** (intercept parameter)
- ▶ β_1 : **傾きパラメータ** (slope parameter)

誤差項 u について

- ▶ 誤差項 u は、 y に影響を与える x 以外のすべての要因を表す。
- ▶ 単回帰分析の主な目的は、他の要因 (u に含まれる要因) を一定に保ったまま (ceteris paribus)、 x が y に与える効果 β_1 を推定することである。
- ▶ しかし、 x と u に相関がある場合、因果関係の解釈は難しくなる。

ゼロ条件付き平均の仮定

- ▶ 因果関係を識別するための重要な仮定は、 x の値が与えられたときの u の平均値が 0 であることである。

$$E(u|x) = 0$$

- ▶ この仮定は、 x と u が無相関であることを意味する。
- ▶ この仮定が満たされるとき、**母回帰関数** (Population Regression Function: PRF) は以下のようなになる。

$$E(y|x) = \beta_0 + \beta_1 x$$

2-2 通常最小二乗法 (OLS) の導出

OLS 推定量

- ▶ 未知のパラメータ β_0, β_1 を、データの標本 $\{(x_i, y_i) : i = 1, \dots, n\}$ を用いて推定する。
- ▶ 残差 (residual) \hat{u}_i を以下のように定義する。

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

- ▶ 残差二乗和 (Sum of Squared Residuals: SSR) を最小化するように $\hat{\beta}_0, \hat{\beta}_1$ を決定する方法を通常最小二乗法 (Ordinary Least Squares: OLS) と呼ぶ。

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

(訳註: OLS は単に「最小二乗法」とも呼ばれるが、加重最小二乗法と区別するため、ここでは「通常」最小二乗法と訳す)

OLS 推定量の公式

- ▶ OLS 推定量 $\hat{\beta}_1$ と $\hat{\beta}_0$ は以下の式で計算される。

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\text{Cov}(x, y)}{\text{Var}(x)}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

- ▶ ここで、 \bar{x}, \bar{y} はそれぞれ標本平均である。

2-3 R による単回帰分析の実装

例: CEO の給与と自己資本利益率 (ceosal1)

- ▶ ceosal1 データセットを使用して、CEO の給与 (salary) と企業の自己資本利益率 (roe) の関係を分析する。

$$salary = \beta_0 + \beta_1 roe + u$$

データの確認

データの読み込みと確認

```
data(ceosal1)
```

```
head(ceosal1[, c("salary", "roe")])
```

```
##   salary  roe
## 1   1095 14.1
## 2   1001 10.9
## 3   1122 23.5
## 4    578  5.9
## 5   1368 13.8
## 6   1145 20.0
```

回帰分析の実行 (lm 関数)

- ▶ R では `lm()` 関数 (Linear Model) を使用して回帰分析を行う。

単回帰分析の実行

```
model_ceo <- lm(salary ~ roe, data = ceosal1)
```

```
# 結果の要約を表示
summary(model_ceo)
```

```
##
## Call:
## lm(formula = salary ~ roe, data = ceosal1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1160.2  -526.0  -254.0   138.8 13499.9
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   963.19     213.24   4.517 1.05e-05 ***
## roe           18.50       11.12   1.663  0.0978 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1367 on 207 degrees of freedom
## Multiple R-squared:  0.01319,    Adjusted R-squared:  0.008421
## F-statistic: 2.767 on 1 and 207 DF,  p-value: 0.09777
```

結果の解釈

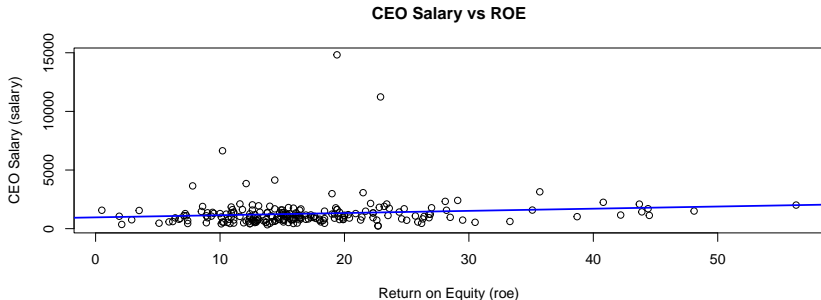
- ▶ 推定された回帰式は以下の通り。

$$\widehat{salary} = 963.19 + 18.50 \cdot roe$$

- ▶ 切片 963.19 は、 $roe = 0$ のときの平均給与（千ドル）を表す。
- ▶ 傾き 18.50 は、 roe が 1 ポイント上昇すると、給与が平均で 18.5 千ドル（1 万 8500 ドル）増加することを意味する。

散布図と回帰直線のプロット

```
plot(ceosal1$roe, ceosal1$salary,  
     xlab = "Return on Equity (roe)",  
     ylab = "CEO Salary (salary)",  
     main = "CEO Salary vs ROE")  
abline(model_ceo, col = "blue", lwd = 2)
```



2-4 適合度と決定係数

決定係数 (R^2)

- ▶ **決定係数** (R-squared) は、従属変数の変動のうち、独立変数によって説明される割合を示す。

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

- ▶ SSE (Explained Sum of Squares): 回帰変動
- ▶ SSR (Residual Sum of Squares): 残差変動
- ▶ SST (Total Sum of Squares): 全変動
- ▶ $0 \leq R^2 \leq 1$ であり、1に近いほどモデルの当てはまりが良いとされるが、低いからといってモデルが無意味なわけではない。

ceosal1 の R^2

- ▶ 先ほどの分析結果 (`summary(model_ceo)`) を見ると、
Multiple R-squared: 0.01319 となっている。
- ▶ これは、CEO の給与の変動の約 1.3% しか `roe` で説明できていないことを意味する。
- ▶ 給与には `roe` 以外の多くの要因が影響していることが示唆される。

2-5 関数形: 対数を用いたモデル

対数モデルの利点

- ▶ 変数を対数変換 (log) することで、非線形な関係を線形モデルとして扱うことができる。
- ▶ 係数の解釈が「変化量」から「変化率 (%)」に変わる。

モデル	方程式	解釈
Level-level	$y = \beta_0 + \beta_1 x$	$\Delta y = \beta_1 \Delta x$
Log-level	$\log(y) = \beta_0 + \beta_1 x$	$\% \Delta y \approx (100\beta_1) \Delta x$
Level-log	$y = \beta_0 + \beta_1 \log(x)$	$\Delta y \approx (\beta_1 / 100) \% \Delta x$
Log-log	$\log(y) = \beta_0 + \beta_1 \log(x)$	$\% \Delta y \approx \beta_1 \% \Delta x$ (弾力性)

例: 賃金方程式 (wage1)

- ▶ 賃金 (wage) の対数を取り、教育年数 (educ) で回帰する (Log-level モデル)。

$$\log(\text{wage}) = \beta_0 + \beta_1 \text{educ} + u$$

例: 賃金方程式 (wage1): R による実装

```
# データの読み込み
data(wage1)
# 対数賃金の回帰分析
model_wage <- lm(log(wage) ~ educ, data = wage1)
# 結果の表示
summary(model_wage)$coefficients
```

```
##              Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) 0.58377267 0.097335834   5.99751 3.736702e-09
## educ        0.08274437 0.007566694  10.93534 3.270645e-25
```

結果の解釈 (賃金方程式)

$$\log(\widehat{wage}) = 0.584 + 0.083 \cdot educ$$

- ▶ $\beta_1 = 0.083$
- ▶ これは、教育年数が1年増えると、賃金が約 8.3% 増加することを意味する (教育の収益率)。
- ▶ $100 \times 0.083 = 8.3\%$

まとめ

まとめ

- ▶ 単回帰モデルは、2 変数間の線形関係を分析する基本的な手法である。
- ▶ OLS（最小二乗法）は、残差二乗和を最小化することでパラメータを推定する。
- ▶ R^2 （決定係数）はモデルの当てはまりの良さを測る指標である。
- ▶ 対数変換を用いることで、弾力性や成長率（%変化）として係数を解釈できる。