第 3 章: 重回帰分析：推定

Jeffrey Wooldridge (2016).
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

```r
library(wooldridge)
```

3-1 重回帰分析の動機

単回帰分析の限界

- 単回帰モデル $y = \beta_0 + \beta_1 x + u$ では、誤差項 u に y に影響を与える他のすべての要因が含まれている。
- x と u が相関している場合（つまり、重要な変数が欠落している場合）、OLS 推定量は**バイアス**を持つ（偏りがある）。
- 重回帰分析を用いることで、他の要因を**明示的に**モデルに組み込み、制御することができる。

多重回帰モデルの一般形

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

- β_0: 切片
- β_j: 変数 x_j の係数 $(j = 1, \ldots, k)$
- u: 誤差項

3-2 OLS 推定量の解釈

係数の解釈：セテリス・パリバス

- 多重回帰モデルにおける係数 β_j は、**他のすべての独立変数を一定に保ったまま**（ceteris paribus）、x_j を 1 単位変化させたときの y の変化量を表す。

$$\Delta y = \beta_j \Delta x_j \quad (\text{他の } x \text{ は固定})$$

- これにより、実験データでなくても、他の要因の影響を統計的に制御した上での因果的効果を推定できる可能性がある。

“部分” 回帰としての解釈

- β_j の推定値は、単に y を x_j で回帰したものではない。
- x_j と他の独立変数との相関関係を取り除いた上で、x_j が y に与える固有の影響を反映している。

3-3 R による重回帰分析の実装

例: 賃金方程式の拡張 (wage1)

- 単回帰モデル：

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

- 多重回帰モデル（経験年数 exper と勤続年数 tenure を追加）：

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u$$

R での実行

```r
data(wage1)

# 単回帰
model_simple <- lm(log(wage) ~ educ, data = wage1)

# 多重回帰
model_multi <- lm(log(wage) ~ educ + exper + tenure,
                  data = wage1)
```

結果の比較

Table 1: 係数の比較

	Variable	Simple	Multiple
(Intercept)	Intercept	0.584	0.284
educ	educ	0.083	0.092
exper	exper	NA	0.004
tenure	tenure	NA	0.022

- 教育年数（educ）の係数が、単回帰の 0.083 から多重回帰では 0.092 に変化した。
- 経験年数や勤続年数を一定に保つと、教育の収益率は少し高くなることが示唆される（ただし、この変化の大きさの評価は文脈による）。

3-4 決定係数 R^2

多重回帰における R^2

- 独立変数をモデルに追加すると、SSR（残差二乗和）は決して増加しないため、R^2（決定係数）は**決して減少しない**。

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

- そのため、変数を追加することの妥当性を判断するために R^2 だけを用いるのは不適切である場合がある（第 4 章以降で修正 R^2 などを扱う）。

決定係数の計算例

```r
summary(model_simple)$r.squared
```

```
## [1] 0.1858065
```

```r
summary(model_multi)$r.squared
```

```
## [1] 0.3160133
```

- 説明変数を追加することで、説明力が約 18.6%から、約 31.6%に向上した。

3-5 除外変数バイアス (Omitted Variable Bias)

除外変数バイアスとは

- 本来モデルに含めるべき変数（真のモデルにおいて $\beta \neq 0$ であり、かつ x と相関している変数）を除外して推定を行うと、OLS 推定量はバイアスを持つ。
- 真のモデル: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$
- 推定モデル: $y = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + v$

このとき、$\tilde{\beta}_1$ の期待値は以下のようになる。

$$E(\tilde{\beta}_1) = \beta_1 + \beta_2 \frac{\mathrm{Cov}(x_1, x_2)}{\mathrm{Var}(x_1)}$$

バイアスの方向

バイアスの符号（プラスかマイナスか）は、β_2 の符号と、x_1 と x_2 の相関（Corr(x_1, x_2)）によって決まる。

Table 2: 除外変数バイアスの方向

Corr.x1..x2.	beta2	Bias
正 (+)	正 (+)	正 (+)
正 (+)	負 (-)	負 (-)
負 (-)	正 (+)	負 (-)
負 (-)	負 (-)	正 (+)

3-6 OLS 推定量の分散

OLS 推定量の分散

- 独立変数が複数ある場合、$\hat{\beta}_j$ の分散は以下の式で与えられる。

$$\mathrm{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\mathrm{SST}_j(1 - R_j^2)}$$

- σ^2: 誤差項の分散
- SST_j: x_j の総変動 ($\sum(x_{ij} - \bar{x}_j)^2$)
- R_j^2: x_j を**他のすべての独立変数**で回帰したときの決定係数

多重共線性 (Multicollinearity)

- R_j^2 が 1 に近い場合（つまり、x_j が他の独立変数と強く相関している場合)、分母が 0 に近づき、Var$(\hat{\beta}_j)$ が非常に大きくなる。
- これを**多重共線性**の問題と呼ぶ。
- 分散が大きいと、推定精度が低くなり、t 検定などで有意な結果が得られにくくなる。

ガウス・マルコフの定理

- 以下の 5 つの仮定（MLR.1 ~ MLR.5）の下で、OLS 推定量は **最良線形不偏推定量**（BLUE: Best Linear Unbiased Estimator）となる。

1. **線形性**: パラメータについて線形である。
2. **ランダムサンプリング**: 無作為標本である。
3. **多重共線性なし**: 独立変数間に完全な線形関係がない。
4. **ゼロ条件付き平均**: $E(u|x_1, \dots, x_k) = 0$。
5. **均一分散**: $\text{Var}(u|x_1, \dots, x_k) = \sigma^2$。

まとめ

まとめ

- 重回帰分析は、他の要因を制御した上で変数の効果を推定できる。
- OLS 推定量は、セテリス・パリバス（他の条件一定）の効果として解釈する。
- 重要な変数が欠落していると、除外変数バイアスが生じる可能性がある。
- 独立変数間の相関が高いと、推定量の分散が大きくなる（多重共線性）。
- ガウス・マルコフの定理により、特定の仮定の下で OLS は最も効率的な不偏推定量である。