第 1 章: 計量経済学と経済データの性質

Jeffrey Wooldridge (2016).
Introductory Econometrics: A Modern Approach
6rd ed. Cengage Learning.

2026-01-19

準備

必要なパッケージのインストールと読み込み

- `wooldridge` パッケージが未インストールの場合は実行

```
install.packages("wooldridge")
```

- `wooldridge` パッケージの読み込み

```
library(wooldridge)
```

- https://cran.r-project.org/web/packages/wooldridge/index.html

1-1 計量経済学とは何か

計量経済学とは

- 経済関係を推定し、経済理論を検証し、政府や企業の政策を評価・実施するための統計的手法の開発に基づいている。
- 応用経済学のほぼすべての分野で重要な役割を果たす。
- 重要なマクロ経済変数（金利、インフレ率、GDP など）の予測は、計量経済学の一般的な応用例である。

非実験データ

- 計量経済学は、個人、企業、または経済のセグメントに対する制御された実験を通じて蓄積されない**非実験データ**（観測データまたは遡及的データとも呼ばれる）の収集と分析に固有の問題に焦点を当てているため、数理統計学とは別の学問分野として発展してきた。

1-2 実証的経済分析のステップ

実証分析のステップ（1）関心のある問題の注意深い定式化

- 経済理論の特定の側面を検証すること、あるいは政府政策の効果を検証することなど。

実証分析のステップ（2）経済モデルの構築

- **経済モデル**は、様々な関係を記述する数式で構成される。

例：犯罪の経済モデル

$$crime = f(\text{wages}, \text{other_income}, \text{freq_arrest},$$
$$\text{freq_convict}, \text{sentence}, \text{age})$$

ここで、$crime$：犯罪率、$wages$：賃金、$other_income$：その他の収入、$freq_arrest$：逮捕頻度、$freq_convict$：有罪判決頻度、$sentence$：刑罰の厳しさ、age：年齢。

実証分析のステップ（3）計量経済モデルの特定

- 経済モデルを**計量経済モデル**に変換する。

$$crime = \beta_0 + \beta_1 \text{wages} + \beta_2 \text{other_income} + ... + u$$

ここで、u は誤差項または撹乱項であり、観測されない要因を含む。

1-3 経済データの構造

データの種類

- クロスセクションデータ
- 時系列データ
- プールされたクロスセクション
- パネルデータ（または縦断的データ）

クロスセクションデータ

- 個人、世帯、企業、都市、州、国など、様々な単位のサンプルを**ある特定の時点**で採取したもの。
- クロスセクションデータでは、データの順序は重要ではない。

例: wage1 データセット

- https://cran.r-project.org/web/packages/wooldridge/refman/wooldridge.html#wage1

```r
# wage1 データセットの最初の数行を表示
head(wage1[, c("wage", "educ", "female")])
```

```
##   wage educ female
## 1 3.10   11      1
## 2 3.24   12      1
## 3 3.00   11      0
## 4 6.00    8      0
## 5 5.30   12      0
## 6 8.75   16      0
```

時系列データ

- 1 つまたは複数の変数の観測値を**時間の経過とともに**集めたもの。
- 過去の出来事が未来の出来事に影響を与える可能性があるため、時系列データセットでは時間の次元が重要である。
- データの**頻度**（日次、週次、月次、四半期、年次）に注意が必要。

例: prminwge データセット (プエルトリコ)

- https://cran.r-project.org/web/packages/wooldridge/refman/wooldridge.html#prminwge

```r
# prminwge データセットの最初の数行を表示
data(prminwge)
head(prminwge[, c("year", "avgmin")])
```

```
##   year avgmin
## 1 1950  0.198
## 2 1951  0.209
## 3 1952  0.225
## 4 1953  0.311
## 5 1954  0.313
## 6 1955  0.369
```

プールされたクロスセクション

- クロスセクションデータと時系列データの両方の特徴を持つ。
- 例えば、1985 年と 1990 年に家計の無作為標本を調査する。
- 新しい政府政策の効果を分析するのに有効な方法である。

例: hprice3 データセット (1)

- https://cran.r-project.org/web/packages/wooldridge/refman/wooldridge.html#hprice3

```r
# 1978年の hpriceデータセットを抽出
hprice1978 <- subset(hprice3, year == 1978)
# hpriceデータセットの最初の数行を表示
head(hprice1978[, c("year", "price", "rooms")])
```

```
##   year price rooms
## 1 1978 60000     7
## 2 1978 40000     6
## 3 1978 34000     6
## 4 1978 63900     5
## 5 1978 44000     5
## 6 1978 46000     6
```

例: hprice3 データセット (2)

```r
# 1981年の hpriceデータセットを抽出
hprice1981 <- subset(hprice3, year == 1981)
# hpriceデータセットの最初の数行を表示
head(hprice1981[, c("year", "price", "rooms")])
```

```
##     year price rooms
## 180 1981 49000     6
## 181 1981 52000     5
## 182 1981 68000     6
## 183 1981 54000     6
## 184 1981 70000     6
## 185 1981 47000     6
```

パネルデータ（または縦断的データ）

- データセット内の各クロスセクションメンバーに対する時系列で構成される。
- **同じ**クロスセクション単位（個人、企業、郡など）が一定期間追跡される。
- 観測されない特性を制御することができるという利点がある。

例: crime3 データセット

- https://cran.r-project.org/web/packages/wooldridge/refman/wooldridge.html#crime3

```r
# crime3データセットの最初の数行を表示
head(crime3[, c("district", "year", "crime")])
```

```
##   district year crime
## 1        1   72 49.54
## 2        1   78 71.32
## 3        2   72 14.77
## 4        2   78 17.85
## 5        3   72 17.35
## 6        3   78 24.57
```

1-4 因果関係、セテリス・パリバス、反事実的推論

因果関係とセテリス・パリバス

- 経済理論のほとんどの検証や公共政策の評価において、経済学者の目標は、ある変数（例：教育）が別の変数（例：労働者の生産性）に**因果的効果**を持つことを推論することである。
- **ceteris paribus**（「他の（関連する）要因を等しく保つ」という意味）の概念は、因果分析において重要な役割を果たす。

反事実的推論

- セテリス・パリバスの概念は、**反事実的推論**を通じても説明できる。
- ある経済単位（個人や企業など）を、2 つ以上の異なる世界の状態で想像する。
- 例：ある労働者が、職業訓練プログラムに参加した場合と参加しなかった場合の 2 つの世界の状態で、その後の賃金がどうなるかを想像する。

まとめ

- 計量経済学は、経済理論の検証、政策立案者への情報提供、経済時系列の予測に使用される。
- 経済データには、クロスセクション、時系列、プールされたクロスセクション、パネルデータなど、様々な種類がある。
- 社会科学における仮説のほとんどは、セテリス・パリバスの性質を持っている。
- ほとんどのデータは非実験的な性質を持つため、因果関係を明らかにすることは非常に困難である。