

第 2 章: 因果関係 (2.5. 観察研究)

今井耕介 著

『社会科学のためのデータ分析入門 (QSS)』

2026-03-09

2.5 觀察研究 (Observational Studies)

観察研究とは

- ▶ **観察研究:** 研究者が処置を無作為に割り当てない研究。
 - ▶ 実世界のデータ（自然に発生したデータ）を分析する。
- ▶ **課題:** 交絡バイアス (Confounding Bias)。
 - ▶ 処置群と対照群が、処置以外の要因（共変量）についても異なっている可能性がある。
- ▶ **解決策のデザイン:**
 1. 前後比較 (Before-and-After) デザイン
 2. 差の差 (Difference-in-Differences: DiD) デザイン

2.5.1 最低賃金と雇用

カード=クルーガー研究 (1994)

- ▶ **問い:** 最低賃金を引き上げると、雇用（失業）はどう変化するか？
- ▶ **事例:**
 - ▶ 1992 年にニュージャージー州 (NJ) が最低賃金を引き上げた。
 - ▶ 隣接するペンシルベニア州 (PA) は引き上げなかった。
- ▶ **データ:** 両州のファストフード店の雇用データ (minwage.csv)。

minwage データの読み込み (1)

▶ カード=クルーガー研究のデータを R に読み込みます。

1. ローカルに保存した *minwage* データの読み込み (推奨)

```
minwage <- read.csv("minwage.csv")
```

(参考) URL から直接読み込むことも可能

```
# minwage <- read.csv("https://ayumu-tanaka.github.io/QSS/QSS_Data/minwage.csv")
```

2. データの次元 (行数と列数) を確認

```
dim(minwage)
```

```
## [1] 358 8
```

州別のデータ分割 (2)

- ▶ ニュージャージー州 (NJ) とペンシルベニア州 (PA) にデータを分け、規模を確認します。

```
# 1. location 列が "PA" かどうかで判定して抽出
```

```
minwageNJ <- subset(minwage, subset = (location != "PA"))
```

```
minwagePA <- subset(minwage, subset = (location == "PA"))
```

```
# 2. データの規模を確認
```

```
dim(minwageNJ) # NJ の店舗数
```

```
## [1] 291 8
```

```
dim(minwagePA) # PA の店舗数
```

```
## [1] 67 8
```

賃金引き上げの確認

▶ 実際に NJ 州だけで賃金が上がったかを確認します。

```
# 時給が $5.05 (以前の最低賃金) 未満の店舗の割合を計算
```

```
# 引き上げ前 (wageBefore)
```

```
mean(minwageNJ$wageBefore < 5.05)
```

```
## [1] 0.9106529
```

```
mean(minwagePA$wageBefore < 5.05)
```

```
## [1] 0.9402985
```

```
# 引き上げ後 (wageAfter)
```

```
mean(minwageNJ$wageAfter < 5.05) # NJ は 0 に近いはず
```

```
## [1] 0.003436426
```

```
mean(minwagePA$wageAfter < 5.05) # PA は高いままのはず
```

```
## [1] 0.9552239
```

2.5.2 交絡バイアス (Confounding Bias)

単純な比較の危険性

- ▶ 引き上げ後の NJ 州と PA 州の雇用率を単純に比較します。

```
# フルタイム雇用率 (fullPropAfter) を計算
# フルタイム数 / (フルタイム数 + パートタイム数)
minwageNJ$fullPropAfter <- minwageNJ$fullAfter /
  (minwageNJ$fullAfter + minwageNJ$partAfter)
minwagePA$fullPropAfter <- minwagePA$fullAfter /
  (minwagePA$fullAfter + minwagePA$partAfter)

# NJ 州と PA 州の平均の差を計算
mean(minwageNJ$fullPropAfter) - mean(minwagePA$fullPropAfter)

## [1] 0.04811886
```

- ▶ **問題:** NJ と PA は州そのものが異なる (経済状況など)。この差が「最低賃金」によるものか「州の違い」によるものか区別できない。これが**交絡**である。

2.5.3 前後比較と差の差 (DiD) デザイン

1. 前後比較 (Before-and-After) デザイン

- ▶ 同じ NJ 州の中で、引き上げ前と後を比較します。

```
# NJ 州の引き上げ前の雇用率を計算
```

```
minwageNJ$fullPropBefore <- minwageNJ$fullBefore /  
  (minwageNJ$fullBefore + minwageNJ$partBefore)
```

```
# 前後の差を計算
```

```
NJdiff <- mean(minwageNJ$fullPropAfter) - mean(minwageNJ$fullPropBefore)  
NJdiff
```

```
## [1] 0.02387474
```

- ▶ **問題:** 時間の経過とともに、最低賃金とは無関係に経済全体が変化した可能性がある。

2. 差の差 (Difference-in-Differences: DiD)

- ▶ **アイデア:** NJ 州の変化から、PA 州 (対照群) で起きた変化 (自然なトレンド) を差し引く。
- ▶ **数式:** $DiD = (NJ_{after} - NJ_{before}) - (PA_{after} - PA_{before})$

PA 州の引き上げ前後の雇用率の変化を計算

```
minwagePA$fullPropBefore <- minwagePA$fullBefore /  
  (minwagePA$fullBefore + minwagePA$partBefore)  
PAdiff <- mean(minwagePA$fullPropAfter) - mean(minwagePA$fullPropBefore)
```

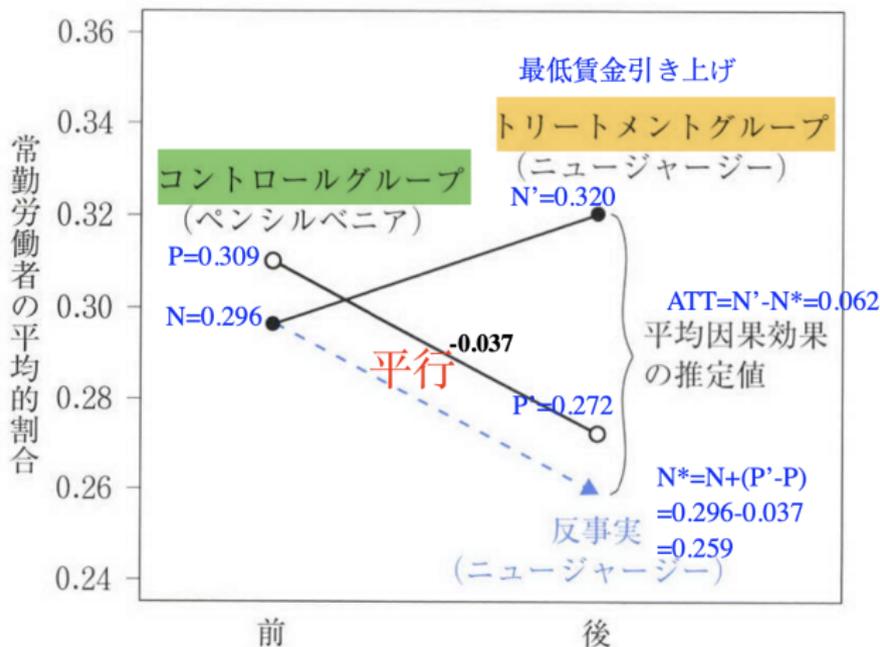
差の差 (DiD) を計算

```
NJdiff - PAdiff
```

```
## [1] 0.06155831
```

- ▶ **結論:** 最低賃金引き上げ後、NJ 州のフルタイム雇用率は PA 州に比べて相対的に約 6.2 パーセントポイント上昇した (雇用は減らなかった)。

差の差 (DiD) の概念図



- ▶ **平行トレンド仮定:** もし処置がなければ、処置群と対照群は同じトレンドを辿るといふ仮定が重要。

2.5.4 まとめ

このセクションのまとめ

- ▶ **観察研究:** 現実世界のデータには交絡因子が潜んでいるため、慎重なデザインが必要。
- ▶ **交絡バイアス:** 処置以外の要因が結果に影響を与え、因果推論を妨げること。
- ▶ **差の差 (DiD) デザイン:**
 - ▶ 処置群の変化から、比較可能な対照群の変化を差し引くことで、共通のトレンドを制御する。
 - ▶ 社会科学において非常に多用される強力な手法。
- ▶ **R の操作:**
 - ▶ 複雑な数式（雇用率の計算など）をデータフレームの列同士の演算で行う。
 - ▶ `subset()` を用いた州や時期の切り分け。