

第 2 章: 因果関係 (2.3. 因果効果と反事実 / 2.4. ランダム化比較 試験)

今井耕介 著

『社会科学のためのデータ分析入門 (QSS)』

2026-03-09

2.3 因果効果と反事実

因果関係の定義

- ▶ **因果関係**: ある変数 (処置) の変化が別の変数 (結果) の変化を引き起こすこと。
- ▶ **反事実 (Counterfactual)**: 「もし処置が行われなかったら、結果はどうなっていたか?」という、現実には観察できないシナリオ。
- ▶ **因果効果**: 「処置が行われた場合の結果 ($Y(1)$)」と「反事実の結果 ($Y(0)$)」の差。
 - ▶ 因果効果 = $\tau_i = Y_i(1) - Y_i(0)$

因果推論の根本問題

- ▶ **根本問題:** 同じ個人について、処置を受けた場合と受けなかった場合の両方の結果を同時に観察することは不可能。
- ▶ **解決策:** 個人の効果ではなく、グループ全体の**平均因果効果 (ATE)** を推定することを目指す。
 - ▶ **標本平均因果効果 (SATE):** 標本における $Y(1)$ の平均と $Y(0)$ の平均の差。

2.4 ランダム化比較試験 (RCT)

社会的圧力と投票率の実験

- ▶ **問い:** 「あなたの投票状況を近所に知らせる」というメッセージは投票率を上げるか？
- ▶ **実験デザイン:**
 - ▶ 多数の有権者を、異なるメッセージを送るグループに**無作為 (ランダム)** に割り当てる。
 - ▶ 「無作為化」により、各グループの特徴が平均的に等しくなるため、グループ間の差を因果効果として解釈できる。

social データの読み込み (1)

▶ 実験で収集されたデータを R に読み込みます。

1. ローカルに保存した *social* データの読み込み (推奨)

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

(参考) URL から直接読み込むことも可能

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

2. データの次元 (行数と列数) を確認

```
dim(social)
```

```
## [1] 305866      6
```

social データの読み込み (2)

- ▶ データの最初の数行を表示して、変数の並びや中身を確認します。

```
# 最初の 5 行を表示  
# messages: 処置の種類 (Control, Civic Duty, Neighbors 等)  
# primary2006: 2006 年予備選挙で投票したか (1=Yes, 0=No)  
head(social, n = 5)
```

```
##      sex yearofbirth primary2004  messages primary2006 hhsize  
## 1  male      1941           0 Civic Duty           0         2  
## 2 female      1947           0 Civic Duty           0         2  
## 3  male      1951           0 Hawthorne          1         3  
## 4 female      1950           0 Hawthorne          1         3  
## 5 female      1982           0 Hawthorne          1         3
```

各グループの投票率の計算

```
# 1. グループごとの平均（投票率）を計算する  
# tapply (結果変数, 分類変数, 関数) を使用  
tapply(social$primary2006, social$messages, mean)
```

```
## Civic Duty    Control Hawthorne Neighbors  
## 0.3145377 0.2966383 0.3223746 0.3779482
```

```
# 2. 全体の平均投票率  
mean(social$primary2006)
```

```
## [1] 0.3122446
```

因果効果 (SATE) の推定

- ▶ 「Neighbors (近所に知らせる)」メッセージの因果効果を、何も送らない「Control」グループと比較して求めます。

```
# Neighbors グループの投票率
mean_neighbors <- mean(social$primary2006[social$messages == "Neighbors"])

# Control グループの投票率
mean_control <- mean(social$primary2006[social$messages == "Control"])

# SATE (平均因果効果) の推定値
mean_neighbors - mean_control

## [1] 0.08130991
```

- ▶ 解釈: 「近所に知らせる」という社会的圧力をかけることで、投票率が約 8.1 パーセントポイント上昇した。

2.4.1 グラフによる比較

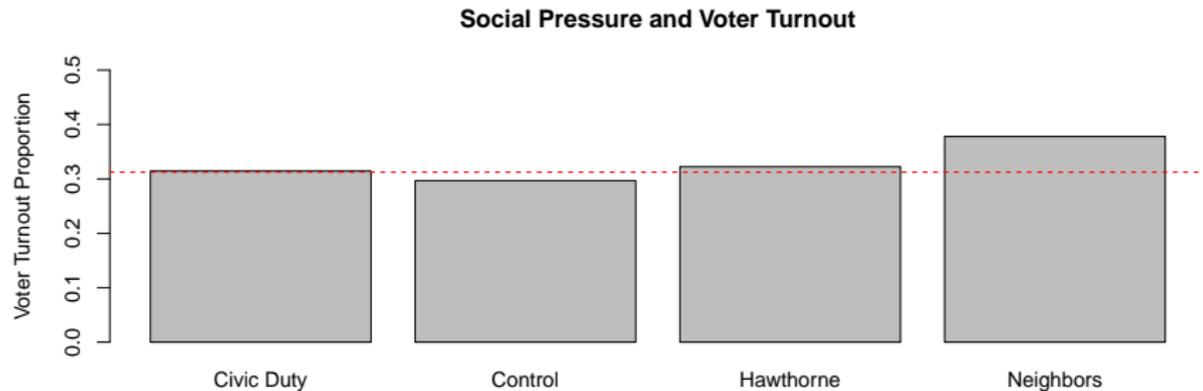
グループ別投票率の可視化: コード

```
# 1. 棒グラフ (barplot) の作成
# 各グループの投票率を計算してプロット
out <- tapply(social$primary2006, social$messages, mean)

barplot(out,
        names.arg = c("Civic Duty", "Control", "Hawthorne", "Neighbors"),
        ylim = c(0, 0.5),
        ylab = "Voter Turnout Proportion",
        main = "Social Pressure and Voter Turnout")

# 2. 全体平均の線を追加
abline(h = mean(social$primary2006), lty = "dashed", col = "red")
```

グループ別投票率の可視化: プロット



2.5 まとめ

このセクションのまとめ

- ▶ **因果効果:** $Y(1) - Y(0)$ で定義されるが、片方は常に観察できない。
- ▶ **ランダム化比較試験 (RCT):**
 - ▶ 無作為割り当てによって「処置が行われなかった場合の世界 (反事実)」を代表する対照群 (Control Group) を作り出す。
 - ▶ 社会的圧力の実験では、Neighbors グループにおいて非常に強力な因果効果が確認された。
- ▶ **R の操作:**
 - ▶ `tapply()` を使ったグループごとの統計量の一括計算。
 - ▶ `barplot()` を用いたカテゴリ別の可視化。
 - ▶ `abline()` による基準線の追加。