

第 5 章: 観察データによる因果効果の推定

Elena Llaudet & Kosuke Imai.

Data Analysis for Social Science: A Friendly and Practical
Introduction.

2026-03-09

5.1 観察研究と共変量

観察研究 (Observational Studies)

- ▶ 研究者が処置を無作為に割り当てない研究。
- ▶ メリット: 倫理的・物理的な理由で実験できない現象も分析可能。
- ▶ デメリット: 共変量 (Confounders) によるバイアスが発生しやすい。

共変量 (Confounders)

- ▶ 処置 (X) と結果 (Y) の両方に影響を与える変数。
- ▶ これにより、処置群と統制群が「平均的に同一」ではなくなり、単純な平均の差では因果効果を測定できなくなる。

5.2 重回帰分析 (Multiple Regression)

なぜ重回帰か？

- ▶ 単純な回帰 ($\hat{Y} = \hat{\alpha} + \hat{\beta}X$) では、他の要因を考慮できない。
- ▶ **重回帰**では、複数の独立変数を同時にモデルに含めることができる。
- ▶ 数式: $\hat{Y} = \hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots$
- ▶ これにより、「 X_2 などの**他の要因**を一定に保ったまま、 X_1 が Y に与える影響」を推定できる。

5.3 データの読み込みと準備

precincts データの読み込み (1)

▶ 2014 年ウクライナ議会選挙のデータを読み込みます。

1. ローカルに保存したデータの読み込み (推奨)

```
precincts <- read.csv("UA_precincts.csv")
```

(参考) URL から直接読み込むことも可能

```
# precincts <- read.csv("https://ayumu-tanaka.github.io/QSS/DSS_Data/UA_precincts.csv")
```

ウクライナの選挙データ (2)

- ▶ **関心:** ロシアのテレビ放送を視聴できることが、親ロシア派への投票に与える影響。

最初の数行を表示

```
head(precincts)
```

```
##   russian_tv pro_russian prior_pro_russian within_25km
## 1           0  2.7210884             25.14286           1
## 2           0  0.8928571             35.34483           0
## 3           1  1.6949153             20.53232           1
## 4           0 72.2689076             84.47761           1
## 5           0  1.2820513             28.99408           0
## 6           1  1.4285714             45.58824           0
```

5.4 共変量の制御

ロシア放送の影響

- ▶ 単純に比較すると、ロシア放送を受信できる地域は、もともと親ロシア的な傾向が強い可能性がある (交絡)。
- ▶ そこで、過去の投票率 (`prior_pro_russian`) を制御変数として追加する。

ロシア放送の影響: R コード

```
# 重回帰モデルの推定
```

```
model_mult <- lm(pro_russian ~ russian_tv + prior_pro_russian,  
                 data = precincts)
```

```
# 結果の表示
```

```
summary(model_mult)
```

```
##
```

```
## Call:
```

```
## lm(formula = pro_russian ~ russian_tv + prior_pro_russian, data = pr
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max  
## -38.776  -7.569   0.154   7.346  26.007
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept)   -13.795374   0.525786  -26.24  <2e-16 ***  
## russian_tv     5.004216   0.484348   10.33  <2e-16 ***  
## prior_pro_russian 0.767525  0.009919   77.38  <2e-16 ***
```

```
##
```

推定値の解釈

- ▶ `russian_tv` の係数: 過去の投票率を一定に保ったとき、ロシア放送が受信可能であることで、今回の親ロシア派の得票率が平均で何%変化したか。

5.5 回帰モデルの比較

モデル間の比較

- ▶ 制御変数を入れる前と後で係数がどう変わるかを確認する。

```
# 単純回帰
```

```
model_sim <- lm(pro_russian ~ russian_tv, data = precincts)
coef(model_sim)["russian_tv"]
```

```
## russian_tv
## 15.64044
```

```
# 重回帰 (過去の投票率を制御)
```

```
coef(model_mult)["russian_tv"]
```

```
## russian_tv
## 5.004216
```

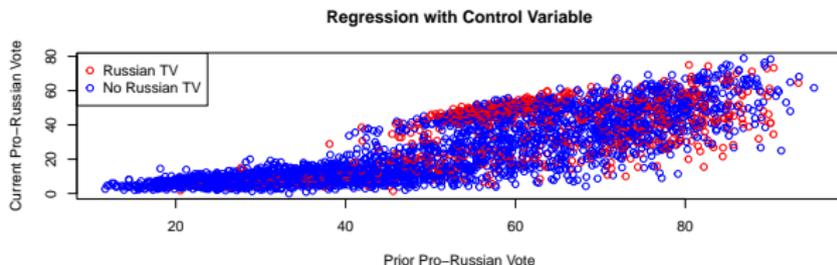
5.6 散布図と回帰平面 (視覚化)

3 次元的な関係の理解

▶ 複数の変数の関係をプロットする。

高度な可視化の代わりに、制御変数ごとの回帰直線を確認 (概念図)

```
plot(precincts$prior_pro_russian, precincts$pro_russian,  
     col = ifelse(precincts$russian_tv == 1, "red", "blue"),  
     xlab = "Prior Pro-Russian Vote",  
     ylab = "Current Pro-Russian Vote",  
     main = "Regression with Control Variable")  
legend("topleft", legend = c("Russian TV", "No Russian TV"),  
       col = c("red", "blue"), pch = 1)
```



5.7 まとめ

第 5 章のまとめ

- ▶ **観察研究**では、共変量（交絡因子）の存在を常に警戒する必要がある。
- ▶ **重回帰分析**を用いることで、他の要因を「一定に保つ（コントロールする）」ことができる。
- ▶ 適切な制御変数を含めることで、観察データからも、より信頼性の高い因果効果の推定が可能になる。
- ▶ ただし、**観察されていない共変量**によるバイアスの可能性は常に残る。