

14 章の答え

練習問題 1

(a) 仮に苗字が出身地に関する情報を含んでいるならば、ランダムな割り当てになっていない可能性がある。たとえば、沖縄などは特有の名前が多く、これが内的妥当性に疑義を生じさせる可能性はある。

(b) 講師への陳情によって当初の割り当てが遵守されなかったケースであり、最終的な割り当てがランダムになっていない。このため、これでは内的妥当性は低くなる¹。たとえば、講師に陳情をするような人はやる気があると考えられるならば、プログラムの効果を高く評価してしまう。

練習問題 2

DID 分析において、対照群は平行トレンドの仮定を満たしていなければいけない。

練習問題 3

モデルは次のとおりである。

$$Y_{i,t} = \alpha + \alpha_2 D_i + \theta_2 Time_t + \beta X_{i,t} + u_{it}$$

$Time_t$ は 2 期(介入後)なら 1 となるダミー変数、 $X_{i,t}$ は交差項 $D_i \times Time_t$ となる。

2 時点の差 $\Delta Y_{i,2} = Y_{i,2} - Y_{i,1}$ は、次のようになる。

$$\begin{aligned} \Delta Y_{i,2} &= Y_{i,2} - Y_{i,1} = (\alpha + \alpha_2 D_i + \theta_2 + \beta X_{i,2} + u_{i,2}) - (\alpha + \alpha_2 D_i + u_{i,1}) \\ &= \theta_2 + \beta X_{i,2} + \Delta u_{i,2} \end{aligned}$$

ただし、 $\Delta u_{i,2} = u_{i,2} - u_{i,1}$ とする。つまり、被説明変数を $Y_{i,2} - Y_{i,1}$ 、説明変数を $X_{i,2}$ として推定すれば、処置効果 β を推定できる。

¹ 当初の割り当て結果は、実際の割り当て結果の良い操作変数となる。たとえば、 Z_i はプログラムに割り当てられたら 1 となるダミー変数とし、 D_i は、実際にプログラムに参加したら 1 となるダミー変数としよう(陳情などがあるため、 Z_i と D_i は必ずしも一致しない)。当初の割当はランダムであるから Z_i は外生変数(個人の属性などと無関係)となる一方、当初の割り当てが遵守されるケースも多いため Z_i と D_i は相関する。このため、 Z_i は D_i の良い操作変数となる。

練習問題 4

最低賃金の例で必要な変数は、店舗*i*がNJに属するなら1となるダミー変数 NJ_i 、2期(介入後)なら1となるダミー変数 $Time_t$ 、交差項 $X_{i,t} = D_i \times Time_t$ と、これらはパネルデータでなく、反復横断面データであっても定義できる変数となる。したがって、反復横断面データでも、DID分析は可能となる。ただし、反復横断面データでは、店舗ごとの固定効果を考慮することはできない。

練習問題 5

T大学合格者のうち最低点での合格者、最低点より1点低い不合格者の所得を比較することで、T大学に合格したことの効果を調べることができる。

合格者(最低点で合格した受験者)と不合格者(最低点から1点低い点数で不合格となった受験者)は本質的に同じ集団と考えられる。この場合、合格か不合格かはランダムに生じていると見なすことができ、合否は受験者の属性などとは無相関となっている。したがって、両集団の所得の平均を比較すれば、T大学に合格したことの効果を推定できるだろう。

練習問題 6

13章の練習問題8では、操作変数 Z_i がダミー変数の場合、2段階最小2乗推定量(2SLS)は、次のワルド推定量として表現できることを示した。

$$\hat{\beta}_{2SLS} = \frac{\bar{Y}_1 - \bar{Y}_0}{\bar{X}_1 - \bar{X}_0}$$

ただし、 \bar{Y}_1 は $Z_i = 1$ のときの Y の標本平均、 \bar{Y}_0 は $Z_i = 0$ のときの Y の標本平均である(同様に、 \bar{X}_1 と \bar{X}_0 は定義される)。

ここで操作変数 Z は、割当変数 W が閾値 c 以上なら1となるダミー変数であるため、2SLS推定量はワルド推定量となる。このとき、分子 $\bar{Y}_1 - \bar{Y}_0$ は、閾値前後での Y の平均の変化となる(\bar{Y}_1 は閾値以上の Y の平均、 \bar{Y}_0 は閾値未満での Y の平均である)。同様に、分母 $\bar{X}_1 - \bar{X}_0$ は、閾値前後での X の平均の変化となる。ここで、 X は処置が割り当てられたら1となるダミー変数であるため、その平均は処置割合(いわば処置確率)となる。つまり、閾値前後のデータを用いた2SLS推定量は

$$\hat{\beta}_{2SLS} = \frac{\text{閾値前後での}Y\text{の平均の変化}}{\text{閾値前後での処置確率の変化}}$$

となることが確認できる。

以下は、ファジーな RDD に関心がある方への補足となる。一般には、割当変数 W が Y に影響を与えると考えられる。ここで、割当変数が 3 乗まで影響すると考えるなら、回帰モデルは次のようになる。

$$Y_i = \alpha + \beta X_i + \theta_1 W_i + \theta_2 W_i^2 + \theta_3 W_i^3 + \gamma_1 W_i X_i + \gamma_2 W_i^2 X_i + \gamma_3 W_i^3 X_i + u_i$$

ここで、 X_i は内生変数なので、交差項 ($W_i X_i$ 、 $W_i^2 X_i$ 、 $W_i^3 X_i$) も内生変数となる。操作変数 Z は、割当変数 W が閾値 c 以上なら 1 となるダミー変数である。操作変数は Z だけでなく、交差項 (ZX 、 ZX^2 、 ZX^3) も操作変数となる。閾値前と後 ($c-h$ から $c+h$ までの区間) のデータを用いて、このモデルを 2SLS で推定すれば処置効果が推定できる。

練習問題 7

ウェブサイトから、再現に必要なデータと STATA の do file をダウンロードできる。STATA 再現コードは以下のとおり。

STATA の再現コード

```
use star_data.dta

**14.2.1 節の推定結果
reg total_score small regular_aid i.school, r
display e(r2_a)
reg total_score small regular_aid total_experience i.school, r
display e(r2_a)

**学校をクラスターと考えると、クラスターロバスト標準誤差を用いる
reg total_score small regular_aid total_experience i.school, cluster(school)
```

これらは通常の OLS 推定である。学校をクラスターと考えると、クラスターロバスト標準誤差を用いたいなら、

```
reg total_score small regular_aid total_experience i.school, cluster(school)
```

とすればよい。なお、`i.school` は学校ごとに異なるダミー変数をいれることを意味する。

練習問題 8

ウェブサイトから、再現に必要なデータと STATA の do file をダウンロードできる。STATA 再現コードは以下のとおり。

```

STATA の再現コード

use mwage_data.dta
xtset store time
gen timestate = time*state

** 14.3.1 節の推定結果
mean fulltime if(state==1 & time==0)
mean fulltime if(state==1 & time==1)
mean fulltime if(state==0 & time==0)
mean fulltime if(state==0 & time==1)
reg fulltime state time timestate, vce(cluster store)
reg fulltime time timestate i.store, vce(cluster store)

** 練習問題の推定結果
reg fulltime state time timestate hours register, vce(cluster store)
reg fulltime time timestate hours register i.store, vce(cluster store)
```

これらは通常の OLS 推定である。店舗をクラスターと考えて、クラスターロバスト標準誤差を使うと、

```
reg fulltime state time timestate, vce(cluster store)
```

となる。また、店舗ごとに異なるダミー変数を入れると、

```
reg fulltime time timestate i.store, vce(cluster store)
```

となる。営業時間 `hours` とレジの台数 `registers` を入れた推定は、以下の通り。

```
reg fulltime state time timestate hours register, vce(cluster store)
```

```
reg fulltime time timestate hours register i.store, vce(cluster store)
```

練習問題 9

ウェブサイトから、再現に必要なデータと STATA の do file をダウンロードできる。STATA 再現コードは以下のとおり。

STATA の再現コード

```
**パッケージをインストールしていないなら、
net install rdrobust, from(https://raw.githubusercontent.com/rdpackages/rdrobust/master/stata) replace
net install rddensity, from(https://raw.githubusercontent.com/rdpackages/rddensity/master/stata) replace
net install rdlocrand, from(https://raw.githubusercontent.com/rdpackages/rdlocrand/master/stata) replace
net install lpdensity, from(https://raw.githubusercontent.com/nppackages/lpdensity/master/stata) replace

use election_data.dta
gen W = diff_share
gen X = W>=0
gen Y = myoutcomenext

**14.5 節の推定結果
rdplot Y W if W>=-0.2 & W<=0.2, nbins(20 20) c(0) p(3) graph_options(xtitle("t 期の得票率差")
ytitle("t+1 期の当選確率"))

gen W2 = W^2
gen W3 = W^3
gen WX = W*X
gen WX2 = W2*X
gen WX3 = W3*X
reg Y X W W2 W3 WX WX2 WX3 if(W>=-0.2 & W<=0.2), r

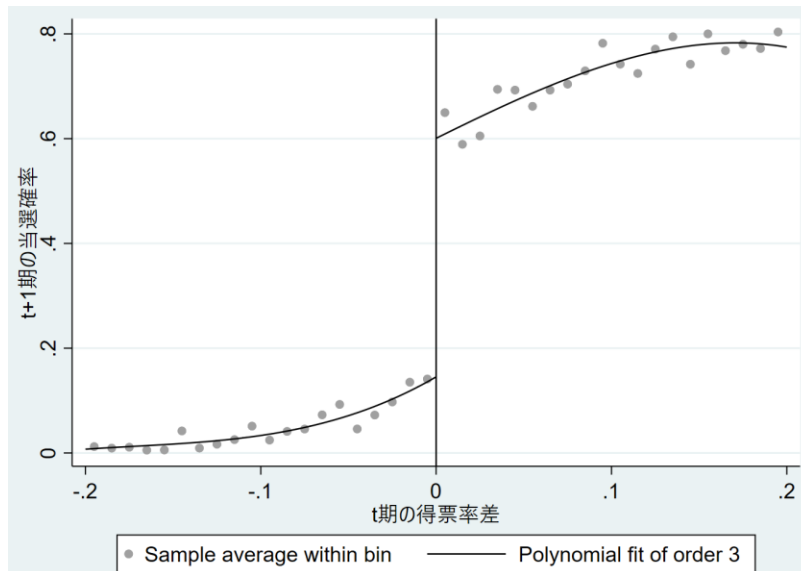
**練習問題
reg Y X W WX if(W>=-0.15 & W<=0.15), r
```

**バンド幅と乗数を最適に選択するコマンド

```
rdrobust Y W, kernel(triangular) c(0)
```

パッケージを使って推定するため、パッケージをインストールする。そして、
`rdplot Y W if W>=-0.2 & W<=0.2, nbins(20 20) c(0) p(3) graph_options(xtitle("t
期の得票率差") ytitle("t+1 期の当選確率"))`

とすると、以下の画面が表示される。ここで、`W>=-0.2 & W<=0.2`とは、`W`が-0.2から0.2までを用いることを意味する。`c(0)`とは閾値が0、`p(3)`とは`W`の`W`乗まで推定に用いることを意味する。



もしバンド幅、`W`の何乗まで入れるかを選択しないで、最適な値を選ばせるなら、次のようにすればよい。

```
rdrobust Y W, kernel(triangular) c(0)
```