

Regression Discontinuity

労働経済学

川田恵介

2025-06-08

1 Get started

1.1 Regression Discontinuity

- 局所的な実験的状況を活用する代表的手法
 - ▶ “制度的な制約”による D の局所的な変化を、自然実験として活用する

1.2 Example: Kawaguchi et al. (2017)

- 労働時間規制は労働政策の大きな論点
 - ▶ 労働時間の短縮は、一見(労働者にとって)望ましいが、副作用はないか？
- 研究課題: Y = 実際の労働時間、賃金、ボーナス、従業員数
 - ▶ 労働時間への規制 ($= D$) が Y に与える因果的影響

1.3 Example. 識別

- 2017 年時点で、労働時間は、週 40 時間、一日 8 時間に規制
 - ▶ 超過するには、従業員代表との合意と残業代が必要
 - ▶ 1988 年以前は週 48 時間であったものが、徐々に短縮
 - 1994 年から 1997 年まで、従業員数が 300 以上の製造業の事業所では 40 時間、以下では 44 時間

1.4 Example: 識別

- 従業員 300 人付近の企業における効果

$$= \underbrace{E[Y \mid X = 300] - E[Y \mid X = 299]}_{\text{推定目標}}$$

- ▶ 300 人か 299 人で、法規制 D は大きく異なる
- ▶ たった一人の違いなので、背景属性の分布はほぼバランスしているはず

1.5 シンプルな実装

```
library(tidyverse)
library(rdrobust)
library(rddensity)

Data = readr::read_csv("Meyersson.csv")

Fit = rdrobust(Data$lpop1994, Data$X)

summary(Fit)
```

Sharp RD estimates using local polynomial regression.

Number of Obs.	2629	
BW type	mserd	
Kernel	Triangular	
VCE method	NN	
Number of Obs.	2314	315
Eff. Number of Obs.	400	233
Order est. (p)	1	1
Order bias (q)	2	2
BW est. (h)	13.320	13.320
BW bias (b)	21.368	21.368
rho (h/b)	0.623	0.623
Unique Obs.	2311	315

	Method	Coef.	Std. Err.	z	P> z	[95% C.I.]
	Conventional	0.012	0.278	0.045	0.964	[-0.532 , 0.557]
	Robust	-	-	0.001	0.999	[-0.644 , 0.645]

2 識別

2.1 研究対象

- D の局所的な因果効果

$$E[y(1) - y(0) \mid Z = z]$$

- Z = 実数 (running variable)

2.2 識別: Sharp discontinuity

- 仮定: Cutoff \bar{z} が存在し、

- $D = 1 \Leftrightarrow Z \geq \bar{z}$

- $D = 0 \Leftrightarrow Z < \bar{z}$

2.3 識別:

- Post-treatment variables 以外の観察できる/できない変数の分布はジャンプしない
- 潜在結果の分布が急変しない

$$\lim_{\epsilon \rightarrow 0} f(y(d) \mid \bar{z} + \epsilon) = \lim_{\epsilon \rightarrow 0} f(y(d) \mid \bar{z} - \epsilon)$$

2.4 Treatment effect on cutoff

- $X = \bar{X}$ を満たす集団内での因果効果 =

$$\begin{aligned} & \lim_{\epsilon \rightarrow 0} E[Y \mid \bar{z} + \epsilon] - E[Y \mid \bar{z} - \epsilon] \\ &= E[y(1) \mid \bar{z} + \epsilon] - E[y(0) \mid \bar{z} - \epsilon] \end{aligned}$$

2.5 Treatment effect on cutoff

- $$\begin{aligned} &= E[y(1) \mid \bar{z}] - E[y(0) \mid \bar{z}] \\ &+ \underbrace{E[y(1) \mid \bar{z} + \epsilon] - E[y(1) \mid \bar{z}]}_{\rightarrow 0} \\ &- \underbrace{(E[y(0) \mid \bar{z} - \epsilon] - E[y(0) \mid \bar{z}])}_{\rightarrow 0} \end{aligned}$$

2.6 解釈

- \bar{z} 上で実験が行われており、 Z は局所的にランダムに決まっている (Local randomization)
- あるいは、 Z が少し異なったとしても、結果変数は直接的な影響を受けない

2.7 注意: Multiple treatment

- 他の”Treatment”の分布も変化しないことを仮定していることに注意
- 例: D = アルコール消費、 \bar{Z} = 20 歳、 Y = 健康状態
 - ▶ 他の Treatment も変化する

2.8 注意: Manipulation

- 背景属性分布も cutoff 前後で”jump”するのであれば、因果効果が識別できない

- Z を cutoff の前後に manipulate することで、 D を”操作”しているのであれば、背景属性分布の jump が疑われる
- 例: 労働時間規制を避けるために、一部の企業が従業員数を 300 名を超えないようにする
 - ▶ より”greedy”な企業が、 $X = 299$ に固まってしまう

2.9 Example: Shigeoka (2015)

- 「“早生まれはどの程度有利か?”」研究において、学年が切り替わる日 (日本の場合は、4月2日)が cutoff として使われてきた
 - ▶ 4月2日よりもう少し早く生まれた VS 4月2日に生まれた
- manipulation は無いのか?
 - ▶ 4月2日生まれが、その近傍に比べて、突出して多い

2.10 注意: Local effect

- 効果の異質性がある場合、 $Z = \bar{z}$ 内での因果効果のみが識別できていることに注意
 - ▶ 局所的な効果であり、他のグループでは効果が大きく異なっても不思議ではない
- 例: 300 人を cutoff にするのであれば、従業員数 300 名の企業への効果は識別される
 - ▶ 5 名の企業への効果と異なることが予想される

3 推定値

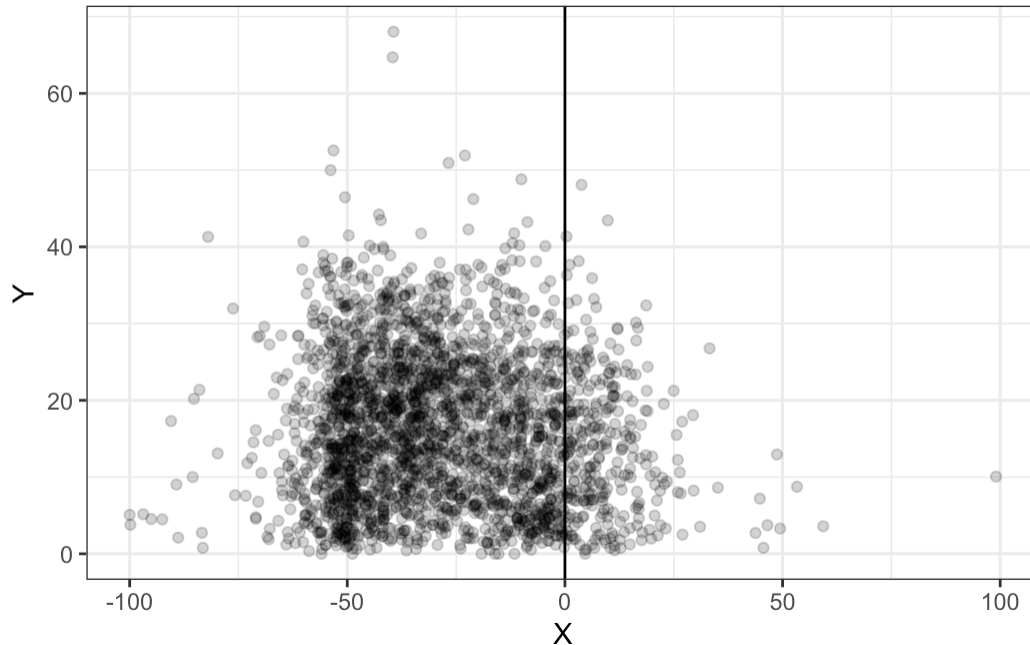
3.1 課題

- Z が \bar{z} よりも少し低い/大きい集団内での Y の平均値を推定する必要がある
 - ▶ \bar{z} 内で $D = 1/0$ を比較することは不可能であり、難しいチャレンジ
- $Y \sim Z$ を柔軟な方法で推定し、 $E[Y | \bar{z}]$ を補完する

3.2 Visualization

- まずは可視化
- 例: Meyersson (2014)
 - ▶ トルコにおいて、宗教的保守派の市長が誕生することが、女性の就学に与える影響を推定
 - 政治的リーダーシップが社会に与える影響
- Y = 地域内の高等教育を受けている女性割合(15-20)
- Z = 宗教的保守派の得票率
 - ▶ cutoff $\bar{z} = 50\%$

3.3 Scatter plot



3.4 Global polynomial

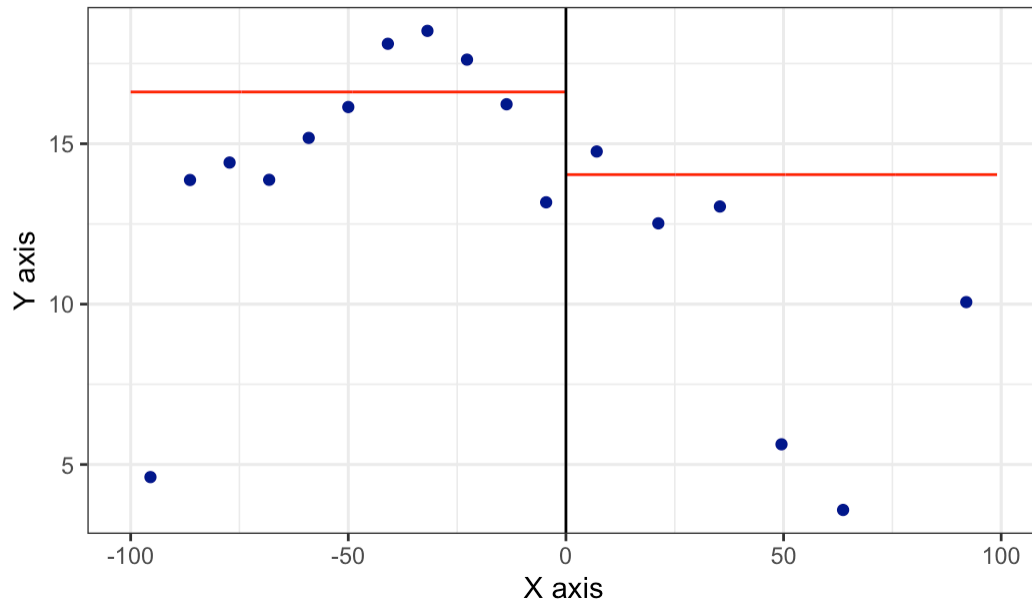
- (他の手法と同様に)何らかの集計が必要
- Cutoff \bar{z} の前後にサンプルを分割する $Data_- = \{Z \leq \bar{z}\}, Data_+ = \{Z > \bar{z}\}$
- Global polynomial:
 - ▶ $Data_-, Data_+$ それぞれについて、 $Y \sim poly(Z, p)$ を回帰する
 - ▶ $p =$ 次元数
 - $p = 3$ であれば、 Z, Z^2, Z^3 までをモデルに投入

3.5 Aggregation: Binplot

- $Data_-, Data_+$ それぞれについて、 Z についてさらにサブグループを作成
 - ▶ サブグループ内の Y の平均値を推定
- rdrobust package
 - ▶ サブグループの数を理論的基準に基づいて自動決定
 - 例えば「母平均を上手く近似する」ように決定
 - Default では、 $\bar{Z} = 0$ と基準化

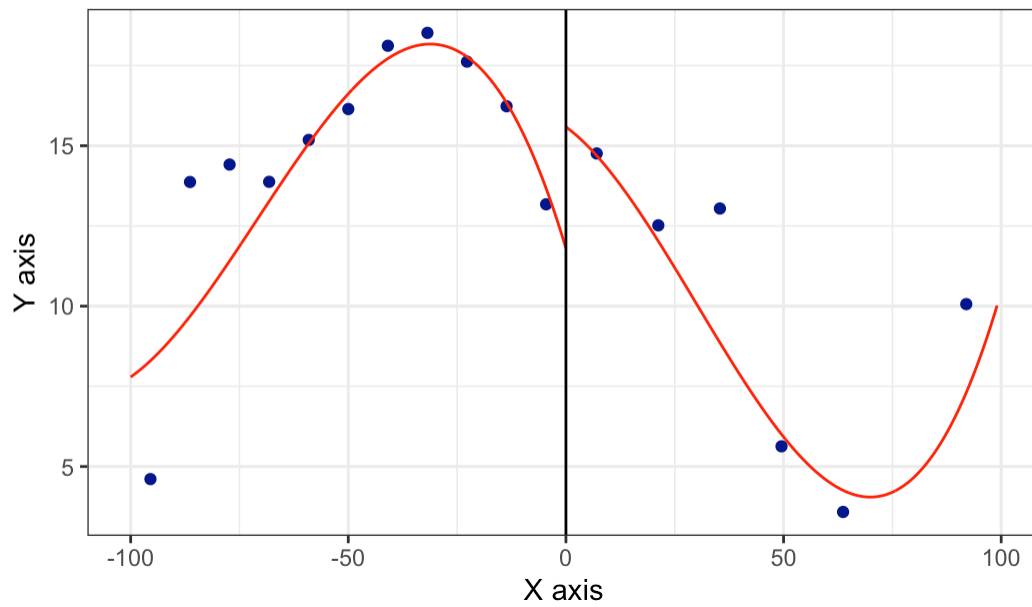
3.6 Example. 市長選挙

RD Plot



3.7 Example. 市長選挙

RD Plot



3.8 Example. 市長選挙

```
lm(Y ~ X,  
    Data,  
    subset = X >= 0) # 得票率が50%以上
```

```
Call:  
lm(formula = Y ~ X, data = Data, subset = X >= 0)  
  
Coefficients:  
(Intercept)          X  
    15.6453      -0.1565
```

```
lm(Y ~ X,  
    Data,  
    subset = X < 0)
```

```
Call:  
lm(formula = Y ~ X, data = Data, subset = X < 0)  
  
Coefficients:  
(Intercept)          X  
    16.19922     -0.01245
```

3.9 Example. 市長選挙

```
lm(Y ~ X + I(X^2),  
    Data,  
    subset = X >= 0) # 得票率が50%以上
```

```
Call:  
lm(formula = Y ~ X + I(X^2), data = Data, subset = X >= 0)  
  
Coefficients:  
(Intercept)          X      I(X^2)  
    16.110678    -0.230997     0.001361
```

```
lm(Y ~ X + I(X^2),  
    Data,  
    subset = X < 0)
```

```
Call:
lm(formula = Y ~ X + I(X^2), data = Data, subset = X < 0)

Coefficients:
(Intercept)          X          I(X^2)
  12.992676   -0.280431   -0.004025
```

3.10 定式化依存

- 一般に、推定結果は定式化に依存
 - ▶ X は 1 変数であることを活かした別推定戦略が有益

3.11 Local regression

- $E[Y | \bar{z}]$ のみ、正確に推定できれば良い
 - ▶ \bar{z} “付近”の事例のみを使えばいいのではないかな?
- Local regression

$$\min \sum \omega_i \times \left(Y - \beta_0 - \beta_1 Z - \beta_2 Z^2 - \dots \right)^2$$

- ω_i = (kernel) weight (\bar{Z}) 付近の事例について、大きな加重をつける

3.12 Kernel weight

- いくつか選択肢がある
- 最もシンプルなものとして、Uniform weight:
 - ▶ $= 1$ if $X \in [\bar{Z} - h, \bar{Z} + h]$
 - ▶ $= 0$ if $X \notin [\bar{Z} - h, \bar{Z} + h]$
 - ▶ h = bandwidth (何らかのやり方で選ぶ必要がある)
- Triangular/Epanechnikov weight をより推奨

3.13 Example. Local regression

```
lm(Y ~ X,
  Data,
  subset = X >= 0 & X <= 5) # 得票率が50%以上
```

```
Call:
lm(formula = Y ~ X, data = Data, subset = X >= 0 & X <= 5)

Coefficients:
```

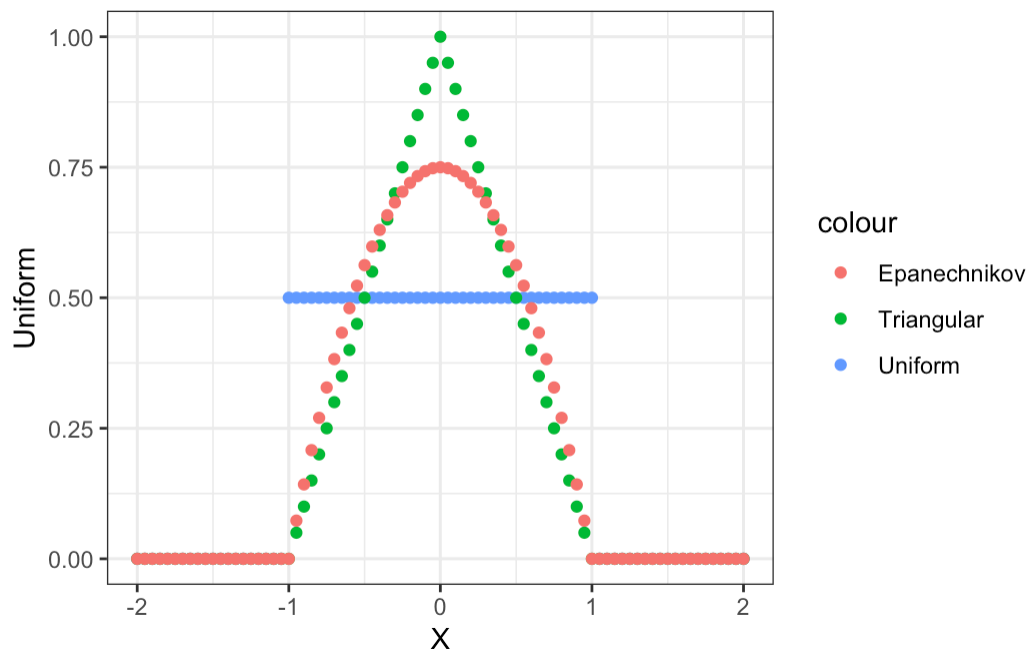

(Intercept)	X
15.6259	-0.1137

```
lm(Y ~ X,
  Data,
  subset = X < 0 & X >= -5)
```

Call:
lm(formula = Y ~ X, data = Data, subset = X < 0 & X >= -5)

Coefficients:
(Intercept) X
13.2466 -0.2175

3.14 Epanechnikov/Triangular weight



3.15 Bandwidth selection

- bandwidth をデータ主導で選ぶ方法については、多くの研究蓄積が存在
 - ▶ Local average treatment effect の推定精度を最大化するように決定(したい)

$$\min E \left[\left(\underbrace{\bar{\tau}}_{\text{推定値}} - \underbrace{\tau}_{\text{真の値}} \right)^2 \right]$$

- rdrobust package が提供 (詳細は、package の HP 参照)
 - 信頼区間の修正も提供

3.16 Example. 市長選挙

```
Fit = rdrobust(
  Data$Y,
  Data$X,
  p = 1) # Defaultでは、Triangular kernelを使用

summary(Fit)
```

Sharp RD estimates using local polynomial regression.

Number of Obs.	2629	
BW type	mserd	
Kernel	Triangular	
VCE method	NN	
Number of Obs.	2314	315
Eff. Number of Obs.	529	266
Order est. (p)	1	1
Order bias (q)	2	2
BW est. (h)	17.240	17.240
BW bias (b)	28.576	28.576
rho (h/b)	0.603	0.603
Unique Obs.	2311	315

	Method	Coef.	Std. Err.	z	P> z	[95% C.I.]
	Conventional	3.020	1.427	2.116	0.034	[0.223 , 5.816]
	Robust	-	-	1.776	0.076	[-0.309 , 6.276]

3.17 Example. 市長選挙

```
Fit = rdrobust(
  Data$Y,
```

```
Data$X,
p = 2) # Defaultでは、Triangular kernelを使用

summary(Fit)
```

Sharp RD estimates using local polynomial regression.

Number of Obs.	2629	
BW type	mserd	
Kernel	Triangular	
VCE method	NN	
Number of Obs.	2314	315
Eff. Number of Obs.	702	291
Order est. (p)	2	2
Order bias (q)	3	3
BW est. (h)	23.121	23.121
BW bias (b)	35.191	35.191
rho (h/b)	0.657	0.657
Unique Obs.	2311	315

Method	Coef.	Std. Err.	z	P> z	[95% C.I.]
Conventional	2.772	1.808	1.533	0.125	[-0.772 , 6.315]
Robust	-	-	1.325	0.185	[-1.276 , 6.600]

3.18 実戦への推奨

- Local regression を使用
 - ▶ Triangular kernel の使用
 - ▶ データ主導で Bandwidth を選択 (+ 値を変えた robustness check)
 - ▶ 推定結果を安定させるために、小さめの $p(= 0, 1)$ を使用 (Gelman & Imbens, 2019)

4 Diagnostic

- Regression Discontinuity の仮定の一部は、データから診断できる

4.1 Placebo test

- Manipulation の存在が大きな課題

- ▶ 背景変数 (Treatment の影響を受けない変数)を Placebo として Y 変数に使用し、RD を適用
- ▶ 顕著な”因果効果”が見られたら、Manipulation の存在を示唆
- 他にも cutoff の水準を変える (placebo cutoff)なども有益

4.2 Example. 市長選挙

```
Fit = rdrobust(
  Data$lpop1994, # 選挙前の人口
  Data$X,
  p = 1) # Defaultでは、Triangular kernelを使用

summary(Fit)
```

Sharp RD estimates using local polynomial regression.

Number of Obs.	2629	
BW type	mserd	
Kernel	Triangular	
VCE method	NN	
Number of Obs.	2314	315
Eff. Number of Obs.	400	233
Order est. (p)	1	1
Order bias (q)	2	2
BW est. (h)	13.320	13.320
BW bias (b)	21.368	21.368
rho (h/b)	0.623	0.623
Unique Obs.	2311	315

	Method	Coef.	Std. Err.	z	P> z	[95% C.I.]
	Conventional	0.012	0.278	0.045	0.964	[-0.532 , 0.557]
	Robust	-	-	0.001	0.999	[-0.644 , 0.645]

4.3 Density test

- Cutoff の前後で、Density が jump していれば、manupulation の証拠

```
Fit = rddensity(Data$X) # Use rddensity package
```

```
summary(Fit)
```

Manipulation testing using local polynomial density estimation.

Number of obs =	2629	
Model =	unrestricted	
Kernel =	triangular	
BW method =	estimated	
VCE method =	jackknife	
c = 0	Left of c	Right of c
Number of obs	2314	315
Eff. Number of obs	965	301
Order est. (p)	2	2
Order bias (q)	3	3
BW est. (h)	30.539	28.287
Method	T	P > T
Robust	-1.3937	0.1634

P-values of binomial tests ($H_0: p=0.5$).

Window Length / 2	<c	>=c	P> T
0.874	20	26	0.4614
1.748	42	49	0.5296
2.622	70	63	0.6030
3.496	95	81	0.3271
4.370	131	98	0.0342
5.245	155	112	0.0100
6.119	183	131	0.0039
6.993	209	148	0.0015
7.867	229	160	0.0005
8.741	257	173	0.0001

4.4 まとめ

- 優れた紹介文献
 - <https://titiunik.scholar.princeton.edu/publications2>
 - cutoff が複数あるケース/control 変数の導入、などへの拡張されている
- 大量の Runding variable と Cutoff がある状況への拡張 (Abdulkadiroğlu et al., 2022)

4.5 Reference

Bibliography

- Abdulkadiroğlu, A., Angrist, J. D., Narita, Y., & Pathak, P. (2022). Breaking ties: Regression discontinuity design meets market design. *Econometrica*, 90(1), 117–151.
- Gelman, A., & Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, 37(3), 447–456.
- Kawaguchi, D., Naito, H., & Yokoyama, I. (2017). Assessing the effects of reducing standard hours: Regression discontinuity evidence from Japan. *Journal of the Japanese and International Economies*, 43, 59–76.
- Meyersson, E. (2014). Islamic Rule and the Empowerment of the Poor and Pious. *Econometrica*, 82(1), 229–269.
- Shigeoka, H. (2015). School entry cutoff date and the timing of births.