

特殊講義：ヨーロッパ国際政治経済 回帰不連続デザイン

ソンジェヒョン
宋財泫

(SONG Jaehyun)

同志社大学文化情報学部 助教

2020年12月28日

回帰不連続デザイン

回帰不連続デザイン

- Thistlewaite and Campbell (1960) で紹介
- 「RDD」と呼ばれる場合が多い
 - ▶ 「かいきふれんぞくでざいん」は長すぎる
- ある点を超えることで処置を受けるか否かが決まる
 - ▶ 人口によって選挙制度が決まる場合 (0 = 多数代表制; 1 = 比例代表制)

$$T_i = \begin{cases} 0 & \text{if Population} < 3500, \\ 1 & \text{if Population} \geq 3500. \end{cases}$$

RDDの考え方

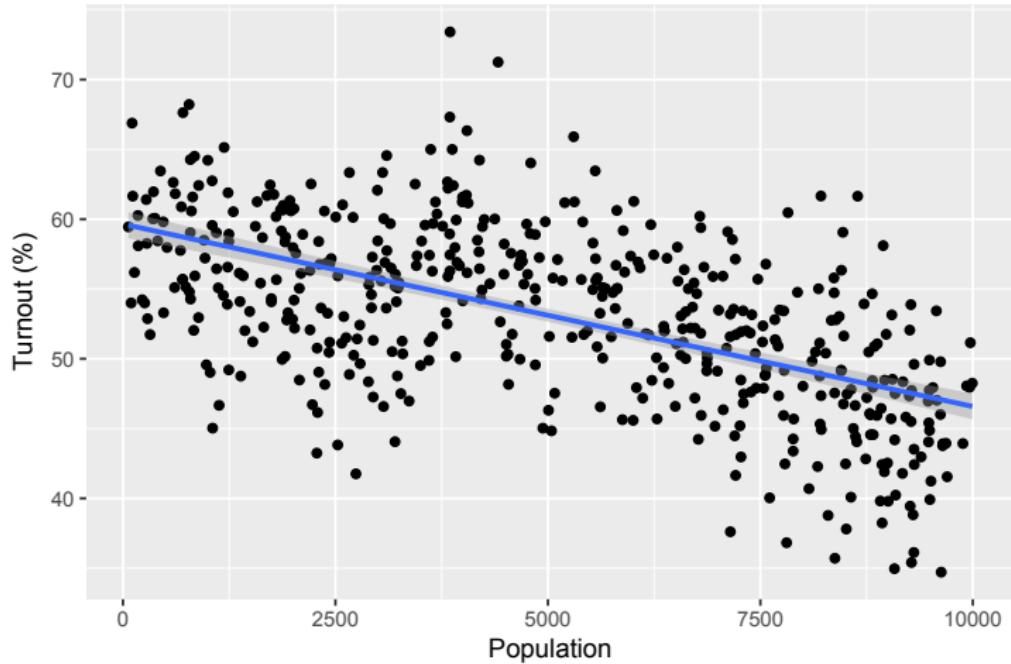
比例代表制と多数代表制の投票率の比較

- 選挙制度は国レベルで異なるため国家間比較になる
- 国固有の文脈により単純比較は困難
- フランスの場合、同一国家、レベルの選挙内で制度が異なる
 - = 国固有の政治的文脈はコントロールされる
- しかし、人口が多いところは都市部が多いし、投票率の低い都市部の特徴により、比例代表制の投票率は過小評価される可能性
 - = 「都市化度」という交絡要因が残存している
- 人口が 3450 人と村と 3550 人の村なら「都市化度」はほぼ同じであるのでは...?
 - ほぼ同条件で選挙制度だけが異なる状況
 - ⇒ 交絡要因が除去されているため、因果推論が可能に

RDDの考え方

比例代表制と多数代表制の投票率の比較

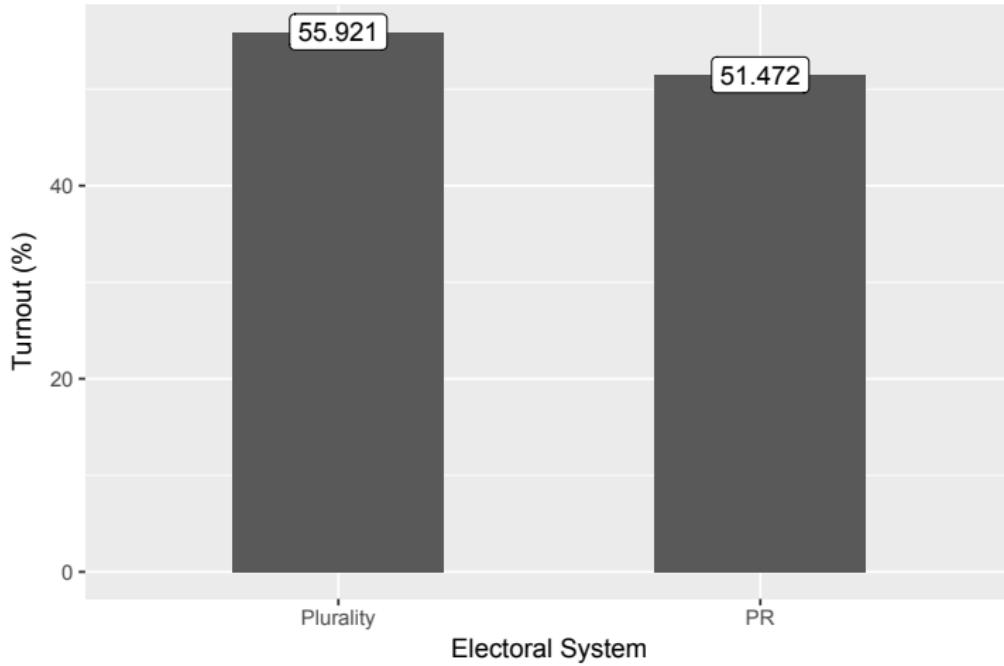
- 人口が多くなるにつれ、投票率が減少傾向



RDDの考え方

比例代表制と多数代表制の投票率の比較

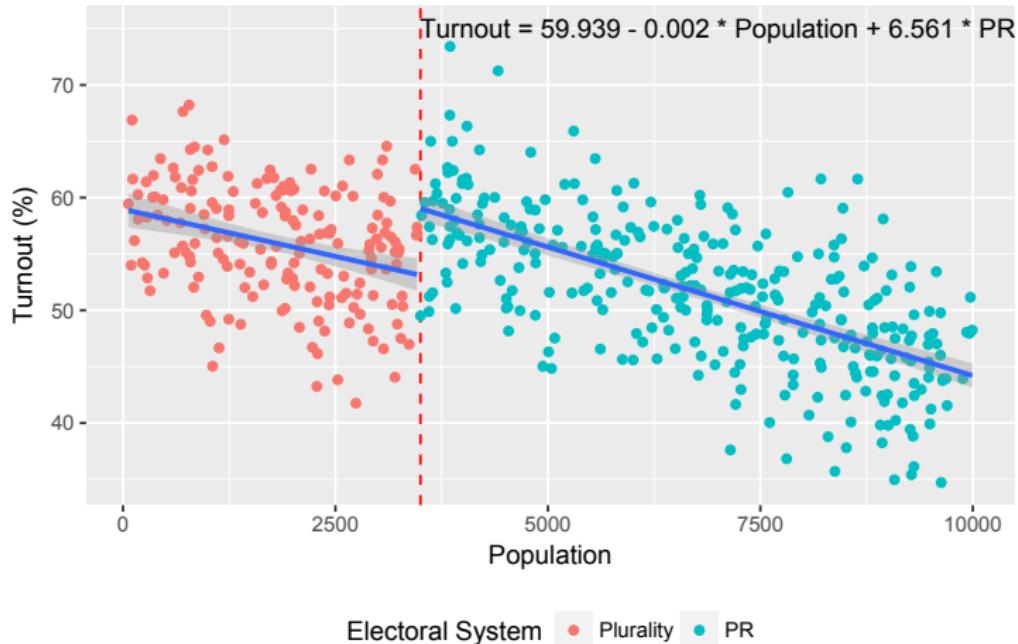
- 比例代表制は投票率を低下する制度だと解釈されてしまう



RDDの考え方

比例代表制と多数代表制の投票率の比較

- 実はそんなことはありませんでした!



2つのRDD

本講義で解説するのは SRD のみ

① Sharp Regression Discontinuity (SRD)

- ▶ 強制変数 (running variable) が閾値 (cut point) を超えると必ず処置を受ける。
- ▶ もっとも単純明快な RDD。実際、多く使われている。

例) 人口と選挙制度

② Fuzzy Regression Discontinuity (FRD)

- ▶ 強制変数が閾値を超えると処置を受ける可能性がある。
- ▶ 操作変数法の理解が必要 (実例あまり見ないような ...)

例) 入試 (一定点数を超えると入学資格が与えられるが入学しない可能性もある)

RDDの前提

Moscoe and Barninnghausen (2015)

- ① 閾値周辺において交絡要因が変化しないこと
 - ▶ 飲酒年齢の事故の例: 20歳になった瞬間、アルコールを分解する酵素が劇的に増加するような悲しい現象は起きない
- ② 閾値のルールが明確であり、既知であること
 - ▶ SRDの場合、閾値は制度に起因するものが多いため、問題になるケースは少ない
- ③ 強制変数が閾値周辺において連続であること
- ④ 潜在的結果が閾値周辺において連続であること
 - ▶ 強制変数と結果変数が取りうる台(support)において連続であれば、第3・4の条件はだいたいクリアされる

RDDによる因果効果の推定

パラメトリック推定方法

(非) 線形回帰分析による推定

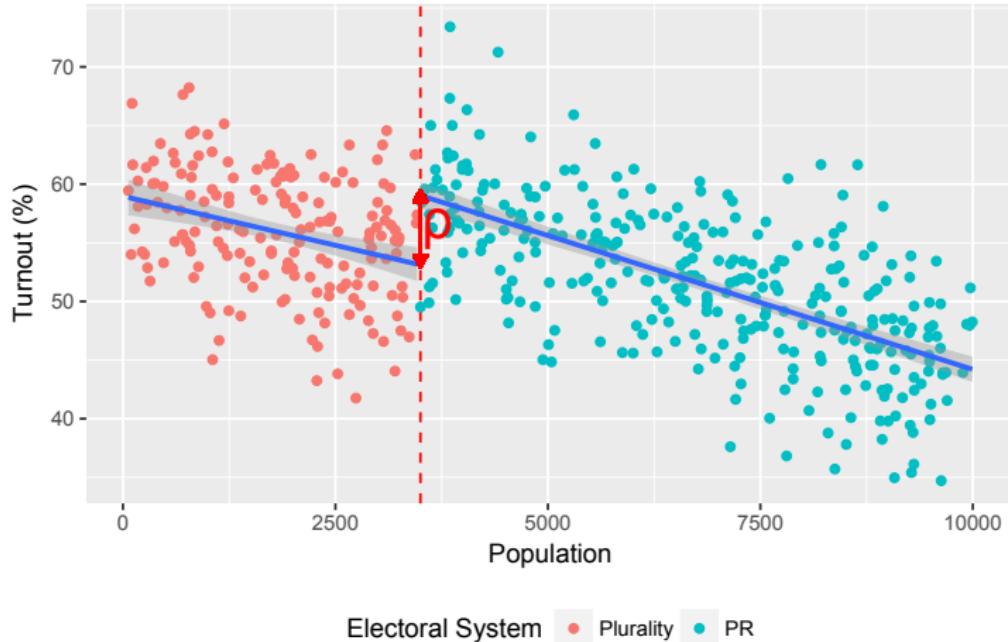
- フラン地方議会選挙の例

$$\widehat{\text{Turnout}} = \beta_0 + \beta_1 \text{Population} + \rho \cdot I(\text{Population} \geq 3500)$$

- 推定されるパラメータは β_0, β_1, ρ
 - $I(\cdot)$ は指示関数 (indicator function)
 - カッコ内の条件が満たされたら 1、それ以外の場合は 0 を返す関数
 - $1(\cdot)$ と表現する場合もあり
- 要は、処置変数 (=ダミー変数) を投入し、強制変数を統制した線形回帰モデル

パラメトリック推定の例

- $\widehat{\text{Turnout}} = \beta_0 + \beta_1 \text{Population} + \rho \cdot I(\text{Population} \geq 3500)$
 - ▷ ρ = 制度の因果効果



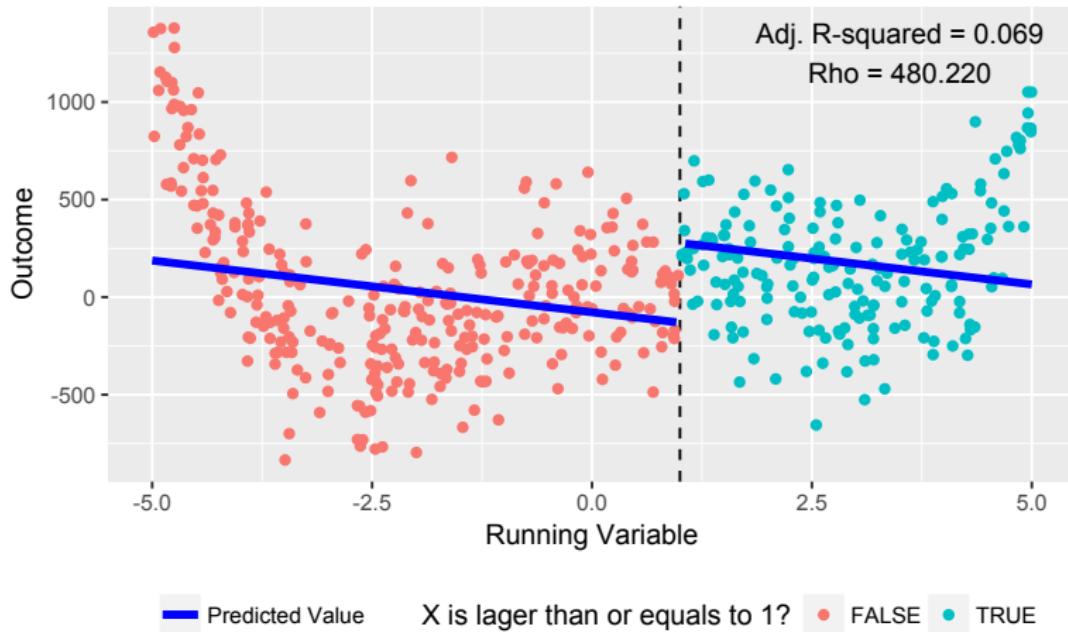
非線形の場合

- 強制変数の二乗、三乗、... を投入
 - どこまで多項式にするかはモデル比較などを通じて分析者が決める
 - R^2 , F 統計量, AIC, BIC, WAIC, LOO など
- モデルにおけるバイアス—分散のトレードオフ関係
 - High-order: バイアス ↓; 分散 ↑
 - High-order は直感に反する推定値が得られる場合も (Gelman and Imbens 2019)
 - ★ noisy estimates
 - ★ sensitivity to the degree of the polynomial
 - ★ poor coverage of confidence intervals

非線形回帰の例

$$\hat{Y} = \beta_0 + \rho \cdot I(X \geq 1) + \beta_1 X$$

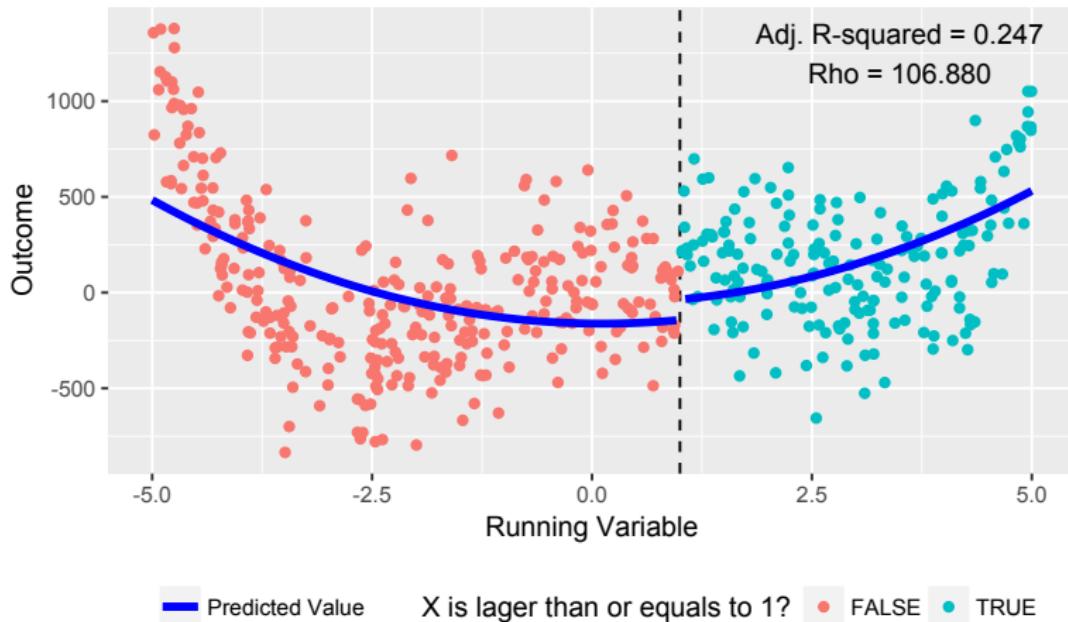
$$\beta_0 + \rho I(X \geq 1) + \beta_1 X$$



非線形回帰の例

$$\hat{Y} = \beta_0 + \rho \cdot I(X \geq 1) + \beta_1 X + \beta_2 X^2$$

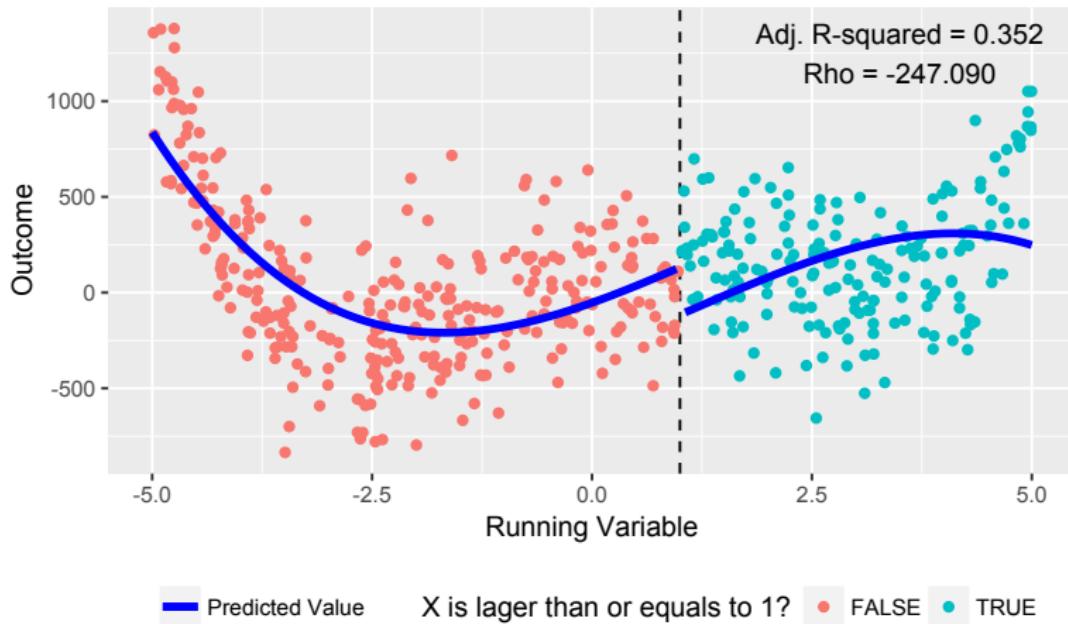
$$\beta_0 + \rho I(X \geq 1) + \beta_1 X + \beta_2 X^2$$



非線形回帰の例

$$\hat{Y} = \beta_0 + \rho \cdot I(X \geq 1) + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$$

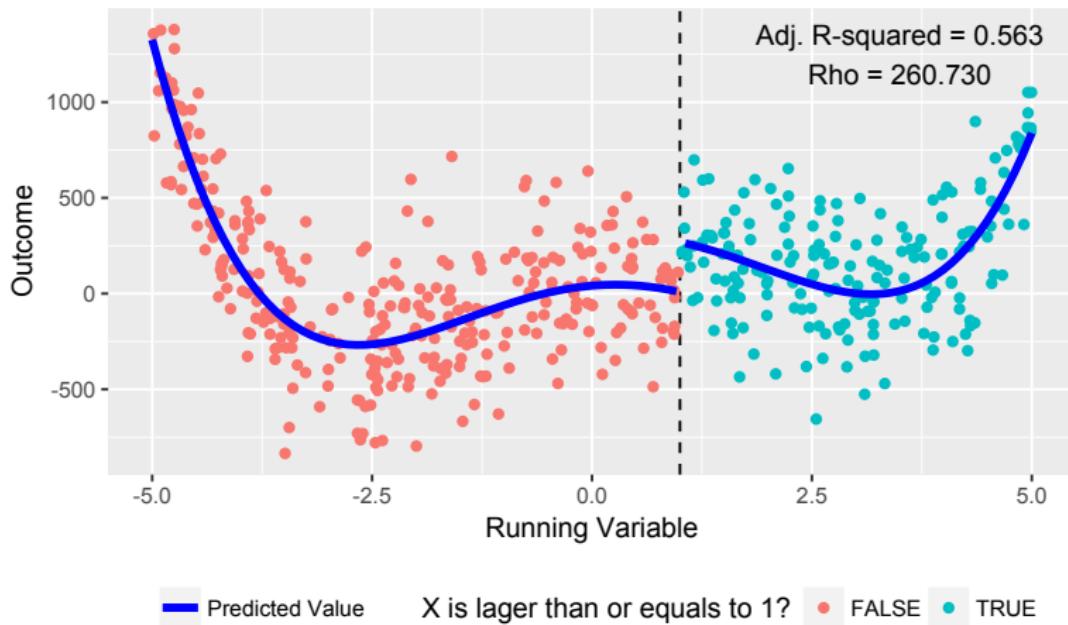
$$\beta_0 + \rho I(X \geq 1) + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$$



非線形回帰の例

$$\hat{Y} = \beta_0 + \rho \cdot I(X \geq 1) + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \beta_4 X^4$$

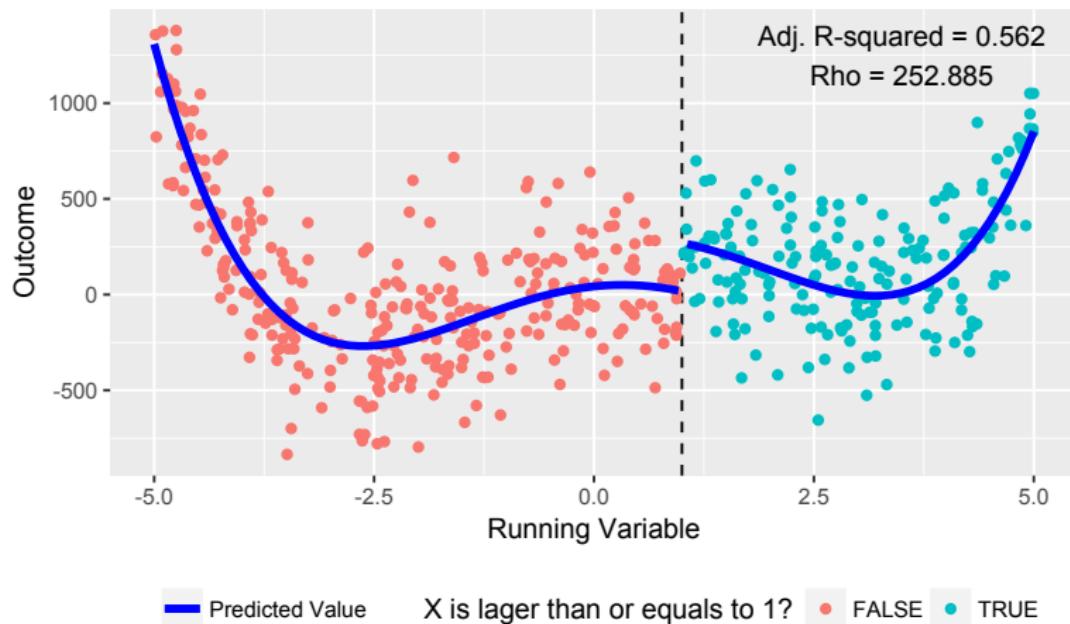
$$\beta_0 + \rho I(X \geq 1) + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \beta_4 X^4$$



非線形回帰の例

$$\hat{Y} = \beta_0 + \rho \cdot I(X \geq 1) + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \beta_4 X^4 + \beta_5 X^5$$

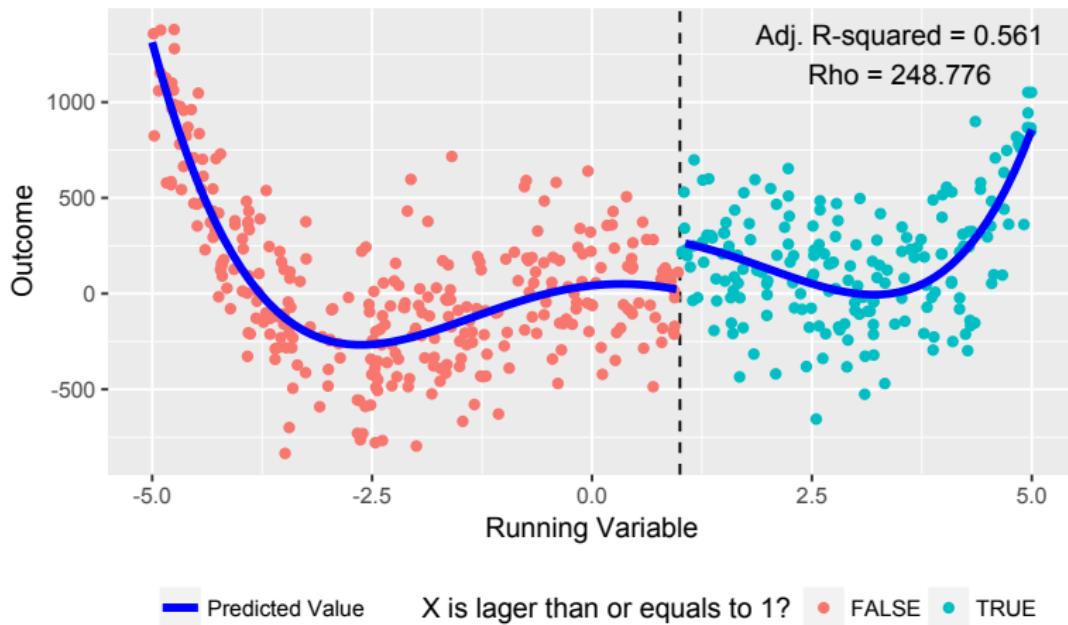
$$\beta_0 + \rho I(X \geq 1) + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \beta_4 X^4 + \beta_5 X^5$$



非線形回帰の例

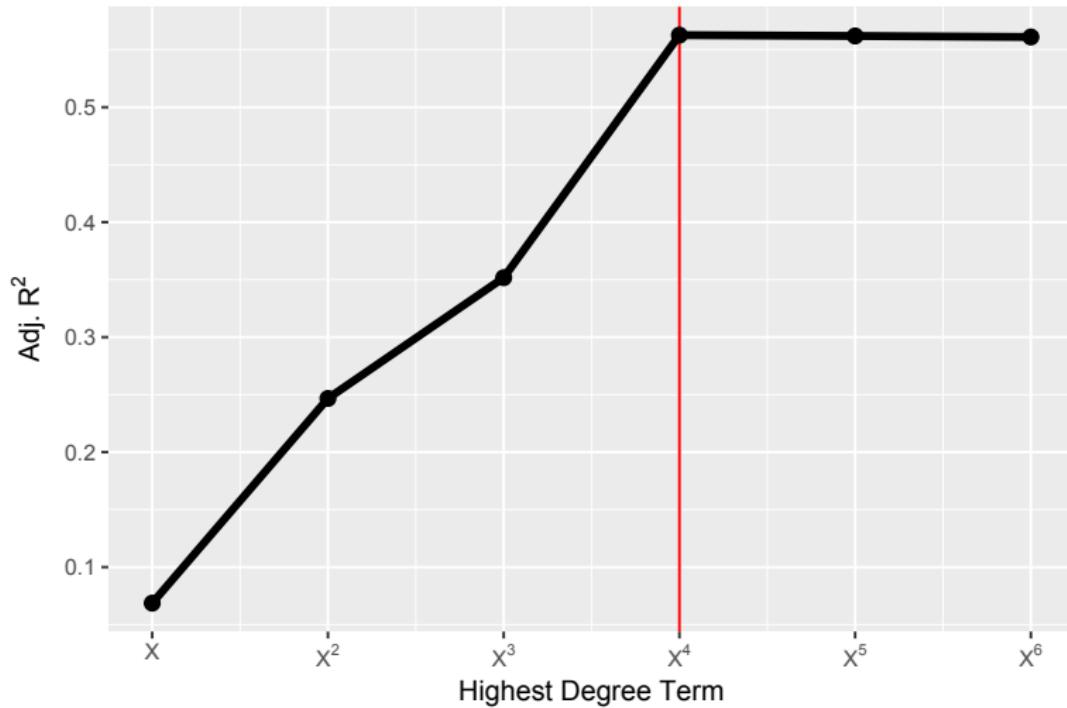
$$\hat{Y} = \beta_0 + \rho \cdot I(X \geq 1) + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \beta_4 X^4 + \beta_5 X^5 + \beta_6 X^6$$

$$\beta_0 + \rho I(X \geq 1) + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \beta_4 X^4 + \beta_5 X^5 + \beta_6 X^6$$



非線形回帰の例

モデル比較 (調整済み R^2)



閾値周辺で傾きが変化する場合

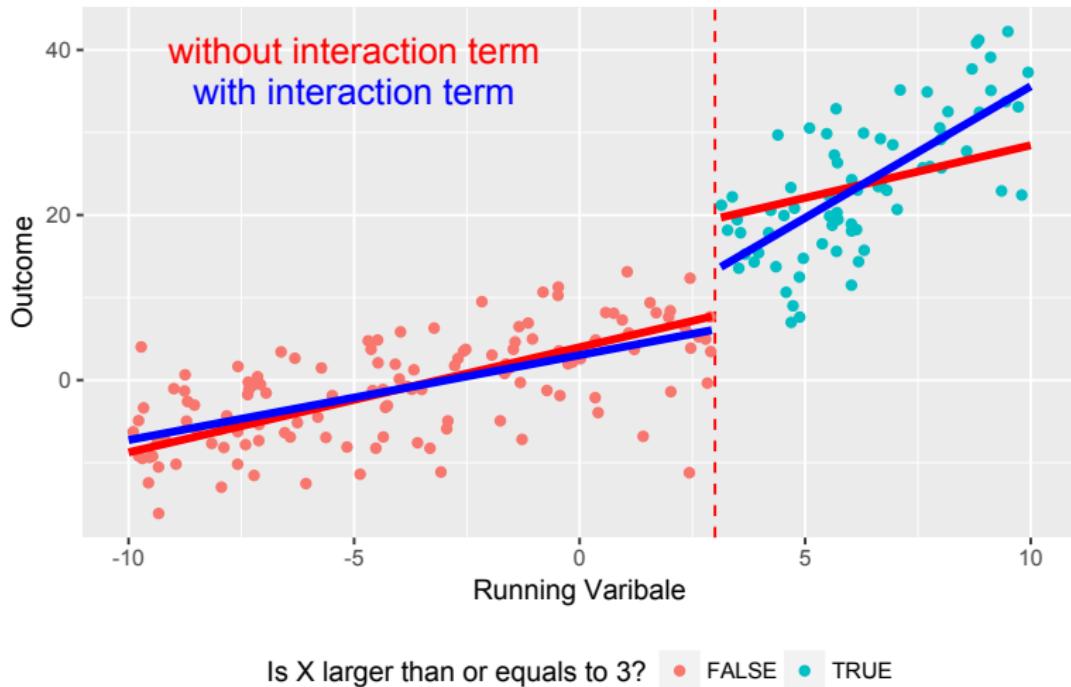
- 単純に $I(X > c)$ のみを回帰式に投入することは連続した回帰直線において切片のみが変化するという前提
 - 回帰直線が連続し、切片のみ変化 (jump) しているため、バイアスが生じる
- ⇒ 交差項を投入することで解決

$$\hat{Y} = \beta_0 + \beta_1 X + \rho \cdot I(X \geq 3) + \gamma \cdot X \cdot I(X \geq 3)$$

- $\beta_0 = 3, \beta_1 = 1, \rho = 1.5, \gamma = 2$ の場合
 - 因果効果は X が 3 になる場合に生じるため、 $1.5 \cdot I(X \geq 3) + 2 \cdot 3 \cdot I(X \geq 3) = 7.5 \cdot I(X \geq 3)$
- ⇒ 因果効果は 7.5

交互作用の例

因果効果 = **11.731** (交互作用なし), **7.172** (交互作用あり)



解釈をより分かりやすくするために ...

強制変数を閾値で中心化 (centering) しよう

$$\hat{Y} = \beta_0 + \beta_1 X + \rho \cdot I(X \geq 3) + \gamma \cdot X \cdot I(X \geq 3)$$

- 交差項がない場合、因果効果は ρ
- 交差項が含まれている場合、因果効果は $\rho + \gamma \cdot c$.
 - ▶ c は閾値 (cut point)
- 強制変数 (X) を c で中心化すると ...

解釈をより分かりやすくするためにには…

強制変数を閾値で中心化 (centering) しよう

$$\begin{aligned}\hat{Y} &= \beta_0 + \beta_1 X^c + \rho \cdot \mathbf{I}(X^c \geq 0) + \gamma \cdot X^c \cdot \mathbf{I}(X' \geq 0) \\ X^c &= X - c\end{aligned}$$

- 閾値が 0 になるため、閾値での因果効果は
 - ▶ $\rho + \gamma \cdot 0 = \rho$
- ρ を因果効果として解釈することが可能に
- 交差項が含まれていないモデルの場合、閾値で中心化してもしなくても ρ は同じ
 - ⇒ とりあえず、入れてみる
- パッケージで分析する際、中心化は気にしなくても良いが、パッケージを使う前に自分で探索的に分析をしてみよう

パラメトリック推定の問題点

強制変数と応答変数間の関数型 (functional form) が正しく設定できるか

- 閾値は制度などによって決まることが多いため、明確な場合が多い
 - 関数型が既知ならこれまでのよう~~に~~パラメトリックな推定が効率的
 - しかし、強制変数と応答変数間の関数型は、ほとんどの場合において未知
 - 関数型の誤設定 (misspecification) はバイアスの原因になりうる
- 例** 多項式モデルの3次項モデルの場合、因果効果が逆転

⇒ ノンパラメトリック推定

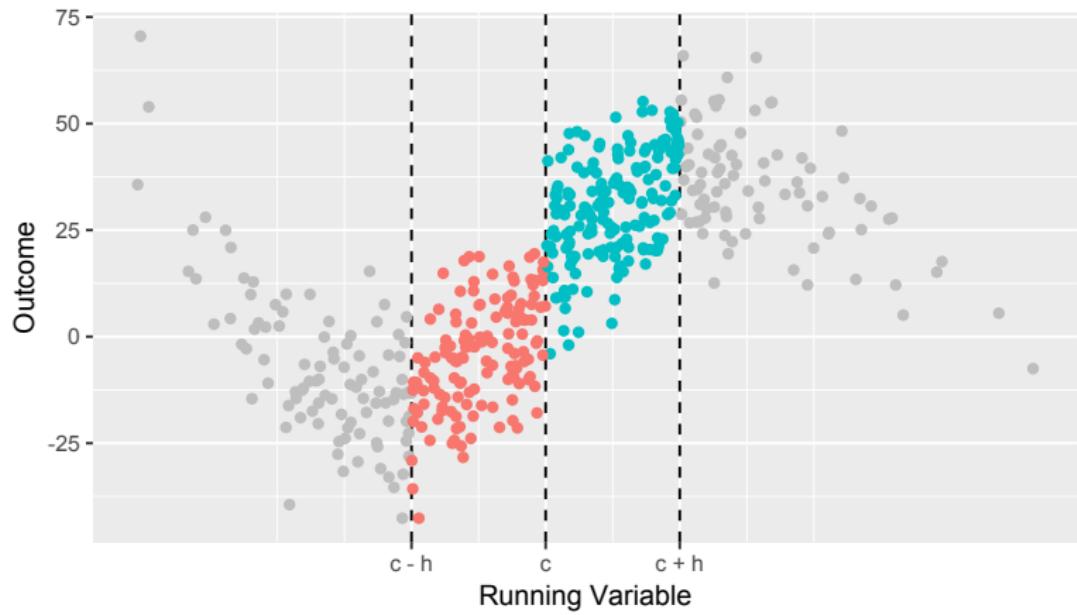
- 関数型と全く無関係ではないが、より柔軟な推定方法
- 関数型の特定が多少間違っても、そこまで大きく問題にならない推定法

ノンパラメトリック推定法

閾値 (c) から h より離れているケースは分析から除外

- 分析対象は $c - h \leq X \leq c + h$ のみ
 - 閾値を中心に中心化済みなら $-h \leq X^c \leq h$
 - h は 「**バンド幅** (bandwidth)」 と呼ばれる
-
- 代表的な因果効果の推定法は 2 つ
 - ① 平均値の差
 - ★ $\mathbb{E}(Y|0 \leq X^c \leq h) - \mathbb{E}(Y|-h \leq X^c < 0)$
 - ② 局所回帰分析 (Local Linear Regression)
 - ★ $-h \leq X^c \leq h$ の範囲内で交差項を含むカーネル回帰分析
 - ★ 定番のカーネル関数は三角 (triangular)
 - ★ カーネル関数が一様 (uniform, rectangular) なら線形回帰分析

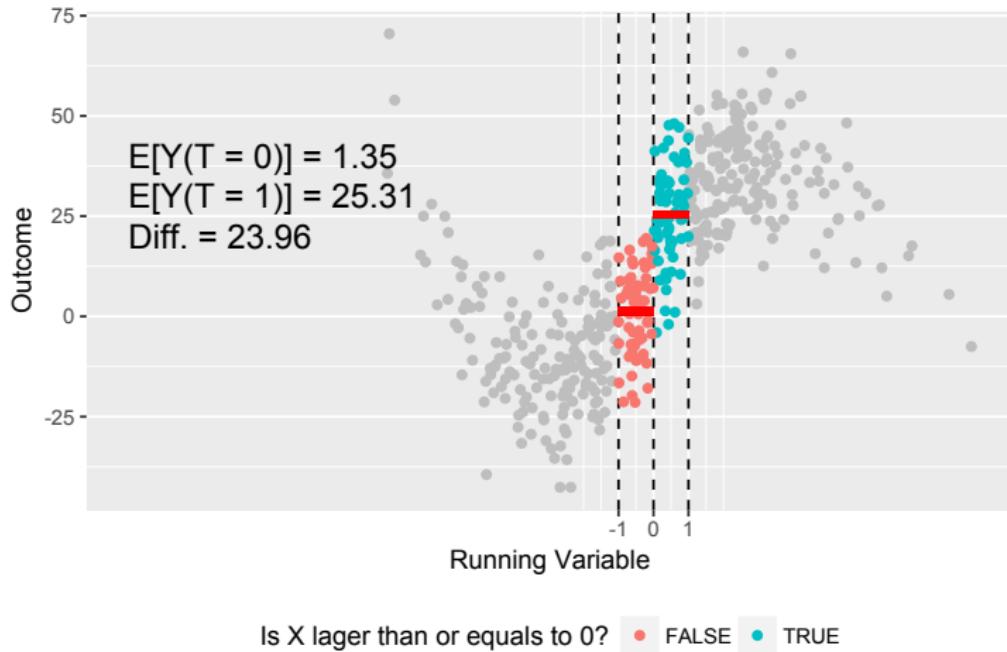
ノンパラメトリック推定法



Is X lager than or equals to c ? ● FALSE ● TRUE

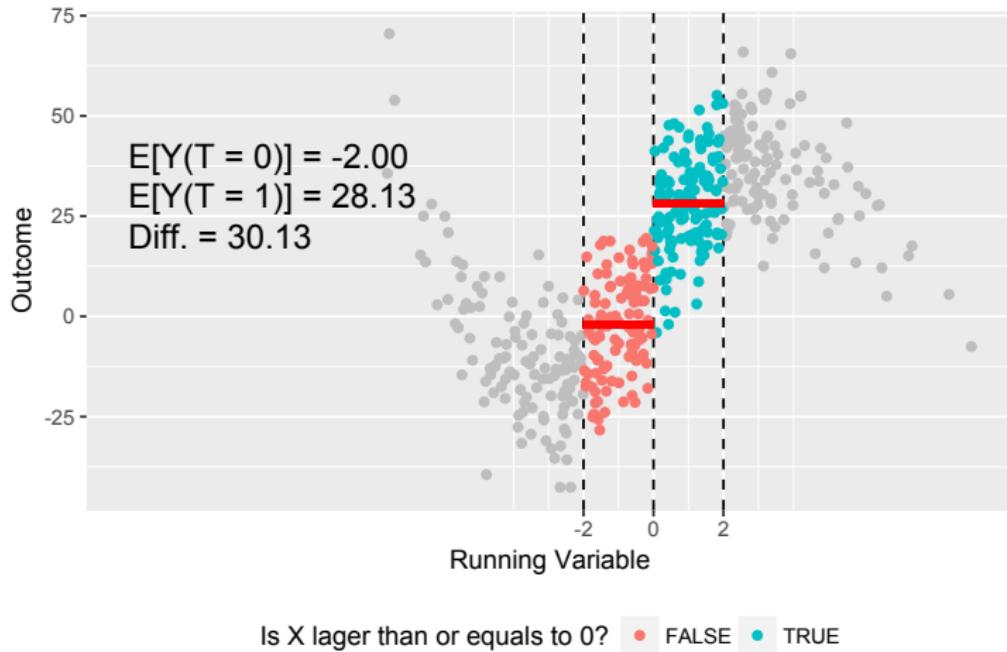
平均値の差分

$h = 1$ の場合、因果効果は 23.96



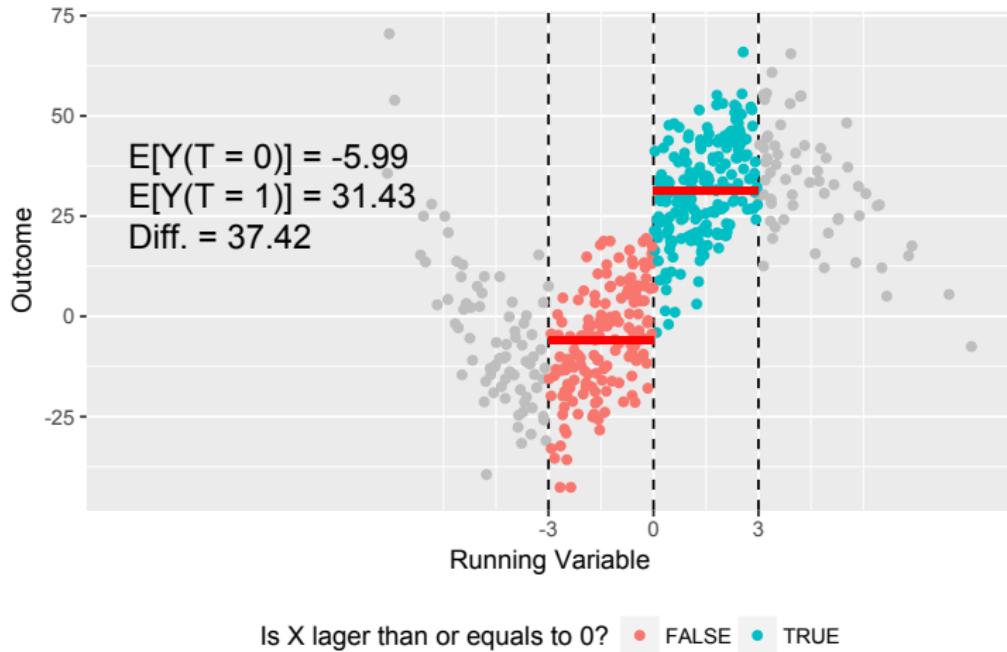
平均値の差分

$h = 2$ の場合、因果効果は 30.13



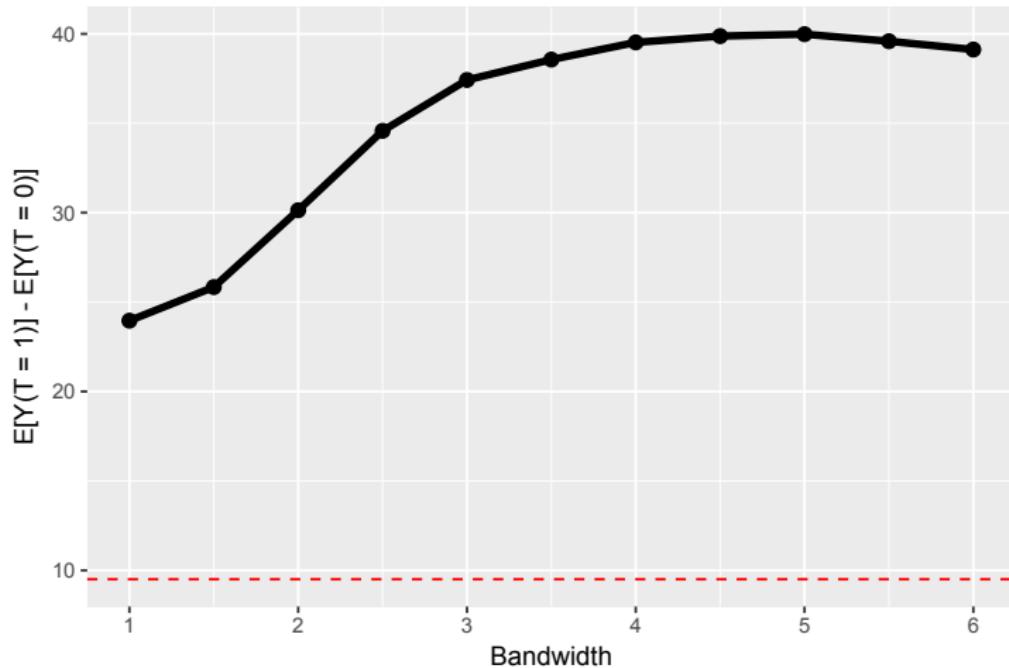
平均値の差分

$h = 3$ の場合、因果効果は 37.42



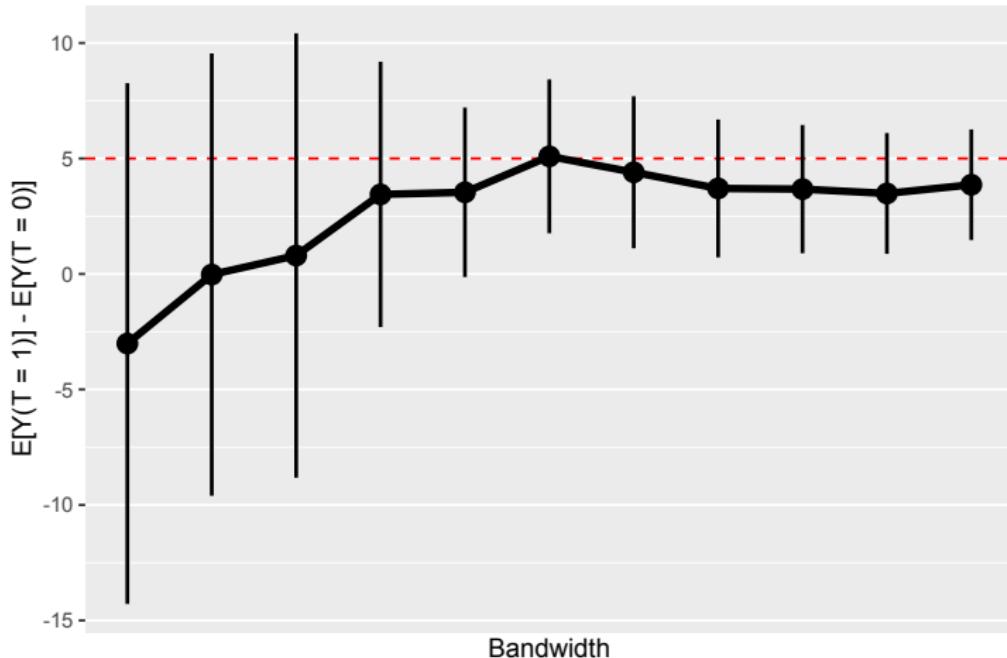
平均値の差分

h の調整による因果効果の変動 (真の因果効果は 9.5)



平均値の差分

参考) 選挙制度と投票率の例 (真の因果効果は 5)：
⇒ 傾きが緩やかな場合、バイアス小

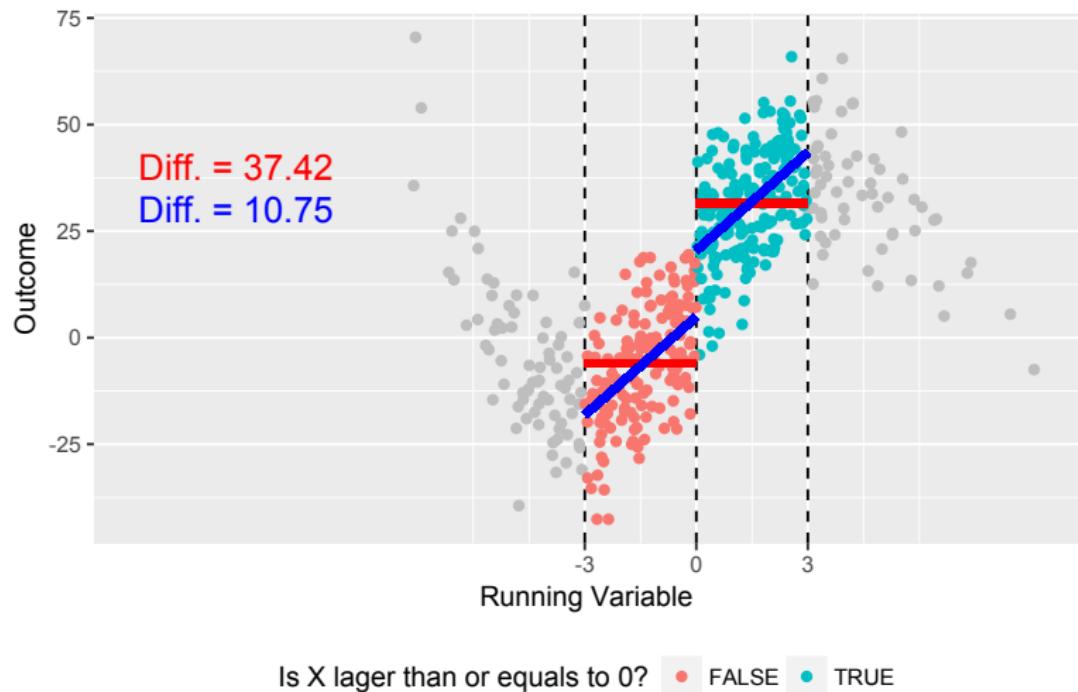


限界

パラメトリック推定に比べてむしろバイアスが大きい!

- バンド幅内のデータのみが分析対象になるため、必然的にサンプルサイズが小さくなる
 - 少数のケースによって平均値の変動が大きい
 - バンド幅内に十分に大きいサンプルサイズが確保されている必要
- 平均値を用いることは、バンド幅内のデータにおいて強制変数の傾きが 0 という前提を置く
- より前提を緩めた推定法
⇒ 局所回帰分析 (local regression)

限界



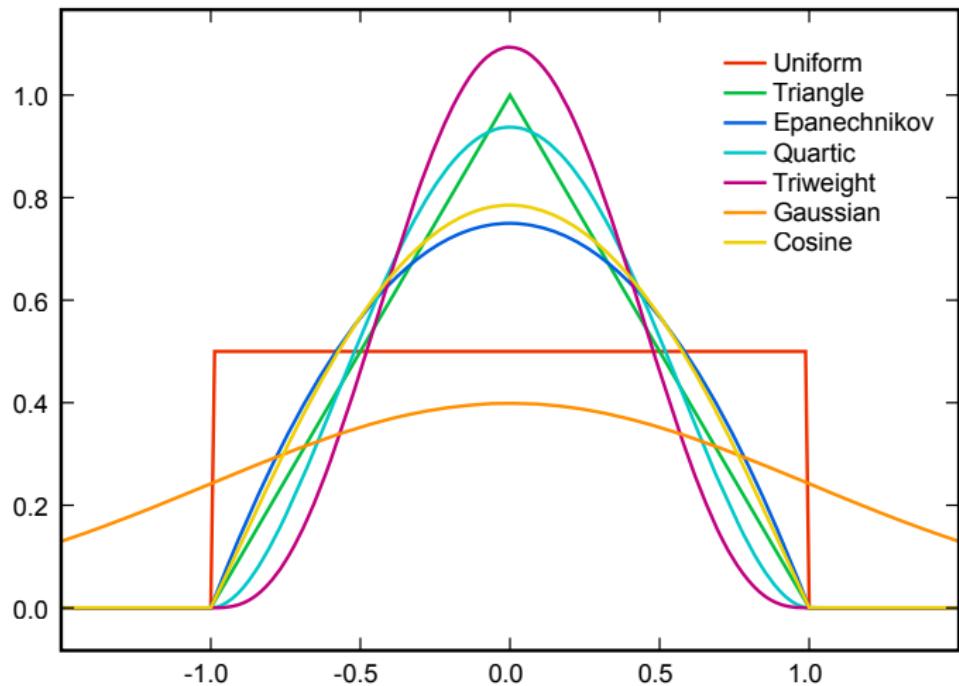
局所回帰分析

バンド幅内データを対象にした線形回帰分析

(Hahn et al. 2001, Poter 2003, Imbens and Lemieux 2008)

- パラメトリック推定と同様、**交差項**や**多項式**も投入可能
 - R の rdd パッケージの場合、交差項 ○・多項式 ×
- 閾値 (c) に近いほど、ケースに**重み**を付ける
 - 重み関数はカーネル関数を用いる
 - Imbens and Lemieux (2008) は一様 (uniform; rectangular) カーネル関数を用いたが、この場合、重みを付けない普通の回帰分析と一致
 - ★ バンド幅内の全てのケースに同じ重みを付ける
 - 三角 (triangular) カーネル関数が統計学的観点からは最適 (optimal) とも (Fan and Gijbels 1996)
 - ★ ほとんどのパッケージは三角カーネル関数がデフォルト

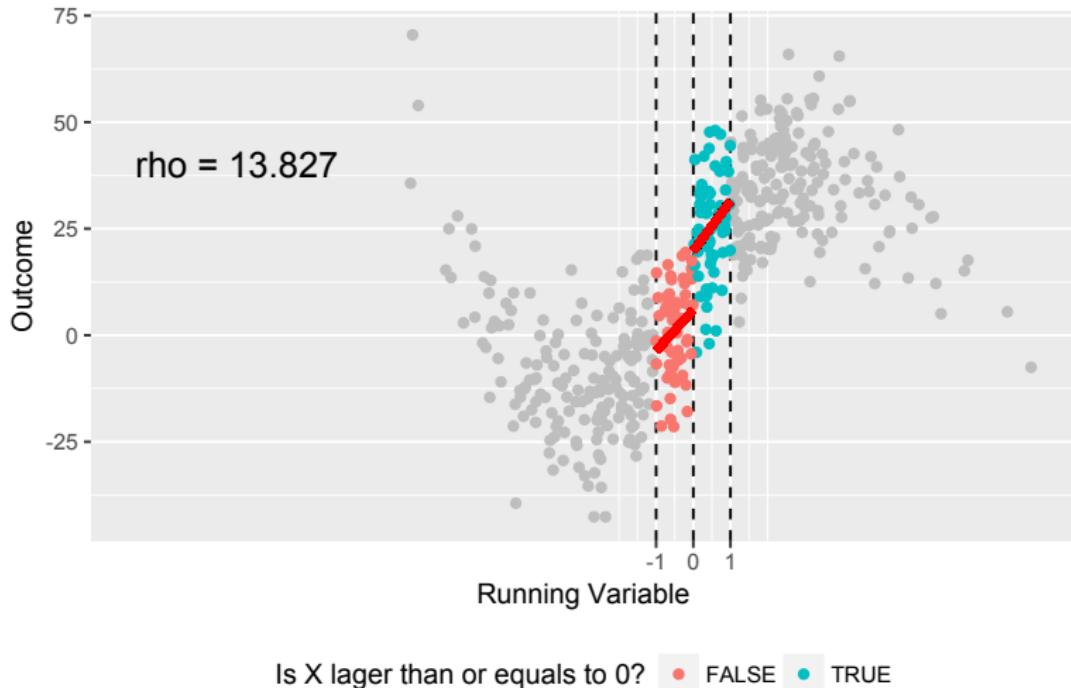
様々なカーネル関数



Source: [https://en.wikipedia.org/wiki/Kernel_\(statistics\)](https://en.wikipedia.org/wiki/Kernel_(statistics))

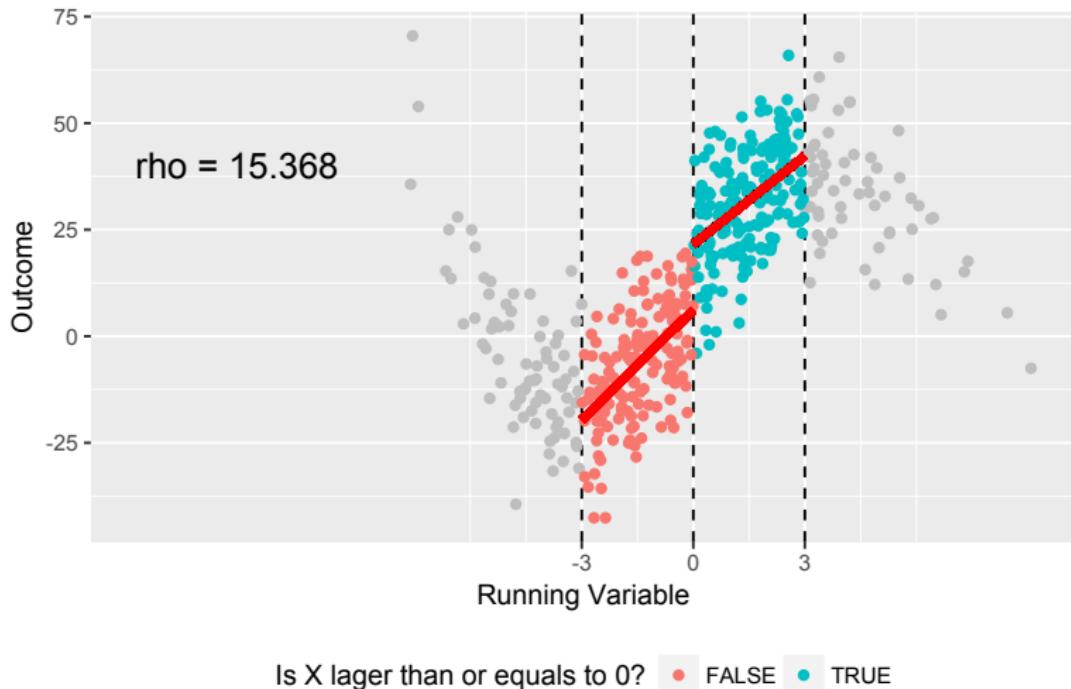
局所回帰の例 (一様カーネル)

$h = 1$ の場合、因果効果は 13.827



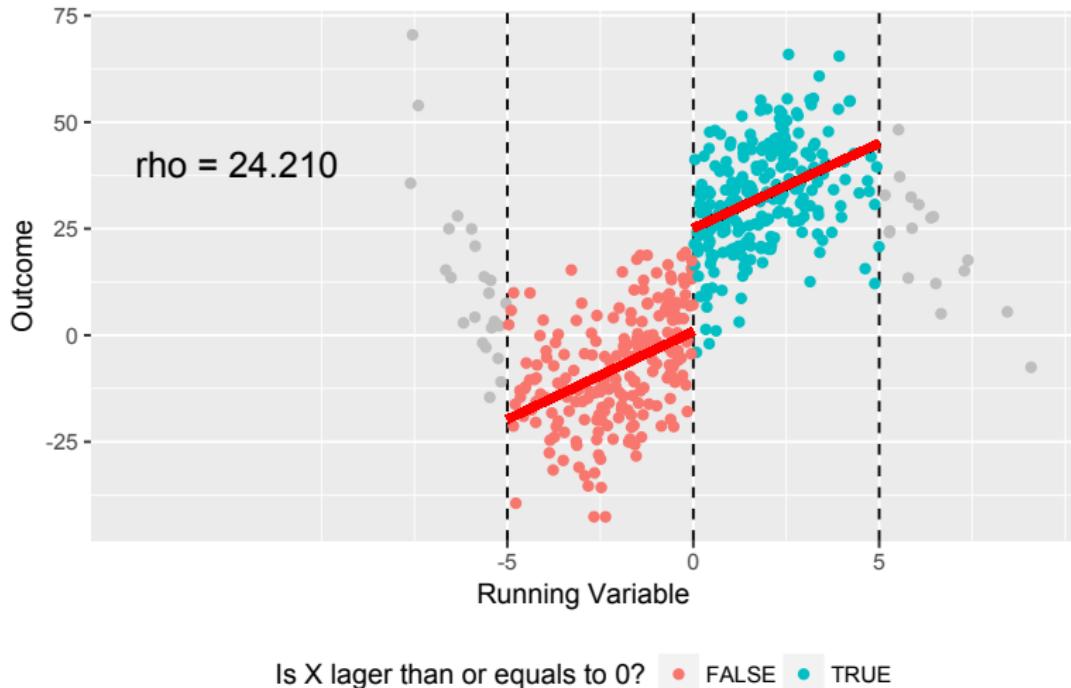
局所回帰の例 (一様カーネル)

$h = 3$ の場合、因果効果は 15.368



局所回帰の例 (一様カーネル)

$h = 5$ の場合、因果効果は 23.96



カーネルを変えてみると…

今回の例だと Triangular 関数がバイアスが小さい傾向 (真の因果効果: 9.5)



バンド幅の話

ノンパラメトリック推定の場合、バンド幅の設定が大事

- 今回の例だと、 h が大きいほどバイアスが大きい
 - あくまでも今回の例だけであって、逆の傾向や凹関数の形をしている場合もある
 - h が小さいとサンプルサイズが小さいため、標準誤差が大きくなるため、 h が適切なサンプルサイズを維持しながら推定の頑健性を確保する必要がある
- **最適バンド幅** (optimal bandwidth)
 - 残念ながら、**バンド幅を決めるルールは存在しない**
 - バンド幅を決めるのは分析者の仕事
 - いろいろと h を変えながら分析を繰り返す
 - 多く使われているバンド幅の決め方
⇒ **Imbens-Kalyanaraman Optimal Bandwidth**

Imbens-Kalyanaraman (IK) Optimal Bandwidth

簡単に計算可能な最適バンド幅の一つ (Imbens and Kalyanaraman 2009)

- 他に CCT バンド幅 (Calonico, Cattaneo, and Titiunik 2014)、CV バンド幅 (Ludwig, and Mill 2007) など

Imbens-Kalyanaraman (IK) Optimal Bandwidth

簡単に計算可能な最適バンド幅の一つ (Imbens and Kalyanaraman 2009)

- 他に CCT バンド幅 (Calonico, Cattaneo, and Titiunik 2014)、CV バンド幅 (Ludwig, and Mill 2007) など

$$\hat{h}_{\text{opt}} = C_K \cdot \left(\frac{2 \cdot \hat{\sigma}^2(c) / \hat{f}(c)}{\left(\hat{m}_+^{(2)}(c) - \hat{m}_-^{(2)}(c) \right)^2 + (\hat{r}_+ + \hat{r}_-) } \right)^{\frac{1}{5}} \cdot N^{-\frac{1}{5}}.$$

簡単！ 簡単！



Imbens-Kalyanaraman (IK) Optimal Bandwidth

簡単に計算可能な最適バンド幅の一つ (Imbens and Kalyanaraman 2009)

- 他に CCT バンド幅 (Calonico, Cattaneo, and Titiunik 2014)、CV バンド幅 (Ludwig, and Mill 2007) など

$$\hat{h}_{\text{opt}} = C_K \cdot \left(\frac{2 \cdot \hat{\sigma}^2(c) / \hat{f}(c)}{\left(\hat{m}_+^{(2)}(c) - \hat{m}_-^{(2)}(c) \right)^2 + (\hat{r}_+ + \hat{r}_-) } \right)^{\frac{1}{5}} \cdot N^{-\frac{1}{5}}.$$

簡単！ 簡単！



- パソコンが勝手に計算してくれるので安心して良い

カーネルとの関係

カーネル選択は推定値に大きな影響を与えない (Lee and Lemieux 2010)

カーネル	h_{opt}	因果効果		
		LATE	Half BW	Double BW
Rectangular	2.964	15.166	17.847	28.919
Triangular	1.886	16.070	12.812	16.341
Epanechnikov	1.755	16.643	13.306	16.010
Quartic	2.004	16.317	13.027	16.224
Triweight	2.227	16.201	12.948	16.312
Tricube	2.023	16.551	13.154	16.024
Gaussian	0.690	15.799	12.794	16.561
Cosine	1.787	16.524	13.205	16.096

実習

実習用データ

講師が作成した架空データ

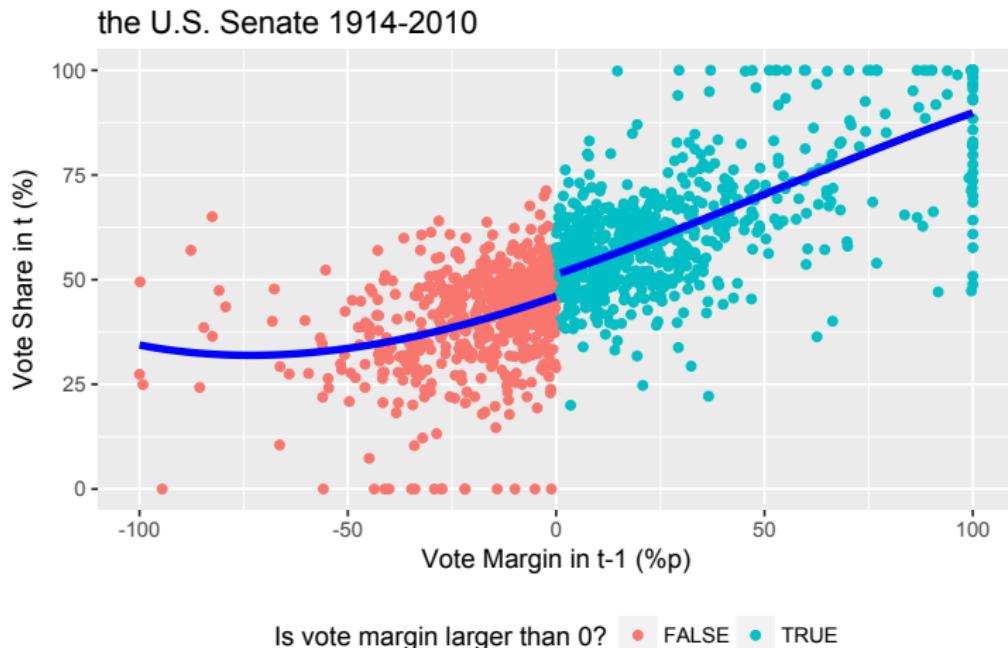
- Day4_Data1.csv: スライド 3 ページ
 - ▶ 真の因果効果: 5
- Day4_Data2.csv: スライド 10 ページ
 - ▶ 真の因果効果: 300
- Day4_Data3.csv: スライド 12 ページ
 - ▶ 真の因果効果: 7.5
- Day4_Data4.csv: スライド 15 ページ
 - ▶ 真の因果効果: 9.5

実習用データ

- Day4_Data5.dta

アメリカにおける現職効果

$$\rho = 9.645, S.E. = 2.115, h_{\text{opt.}} = 7.550$$



Ariga et al. データについて

詳細は以下の論文を参照

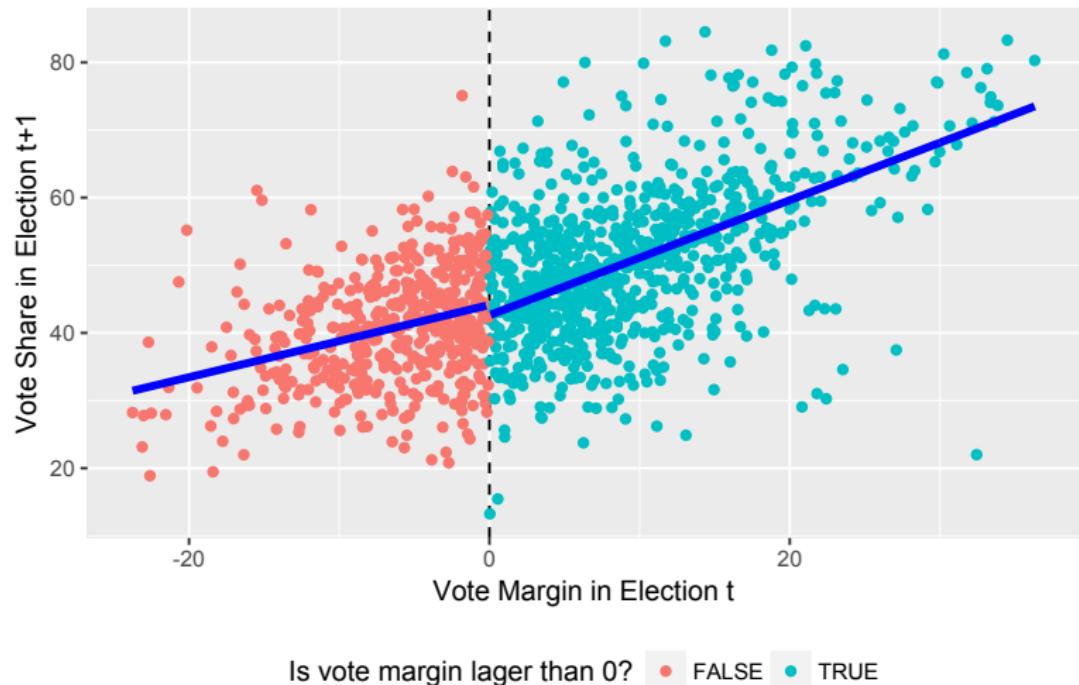
- Kenichi Ariga, Yusaku Horiuchi, Roland Mansilla, and Michio Ueda. 2016. "No Sorting, No Advantage: Regression Discontinuity Estimates of Inc incumbency Advantage in Japan." *Electoral Studies*, Vol. 43, pp. 21–31

変数の説明

- $vm1$: t 期選挙における自民党候補者の vote margin
 - ▶ vote margin: 自分の得票率 - 非自民候補の最高得票率
- F_1dpv_smd : $t+1$ 期選挙における自民党候補者の得票率
- $1dp_LCF$: $t-1$ から $t+1$ 期まで選挙区割が変化せず、自民党候補者がいる選挙区ダミー

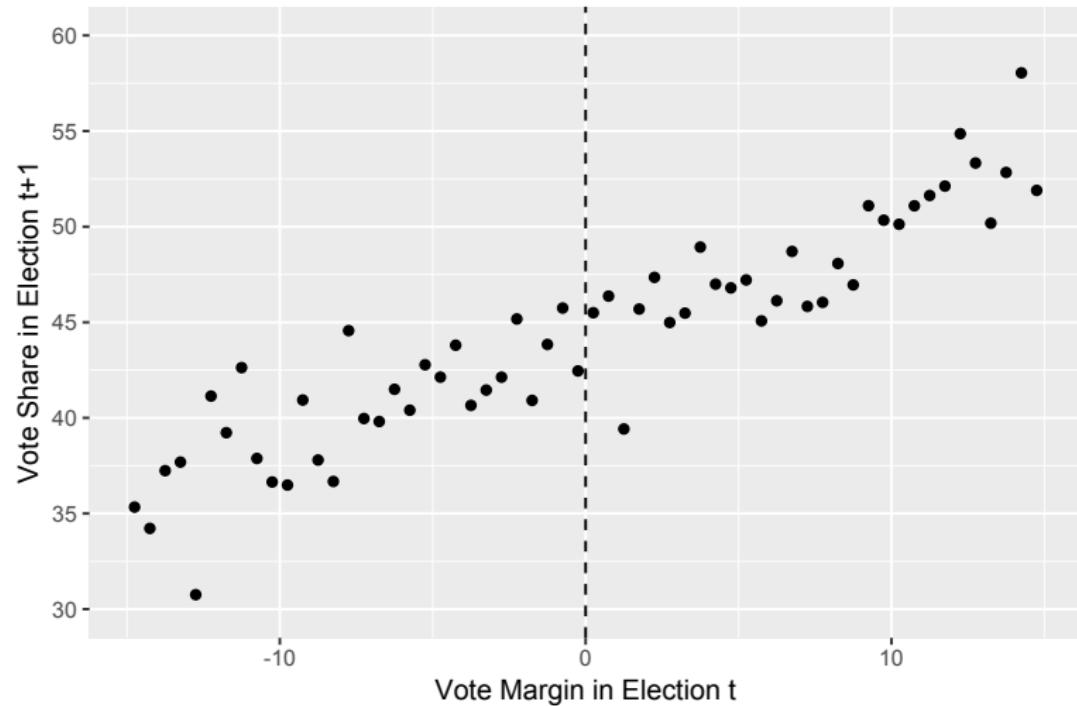
散布図

全ケースの散布図



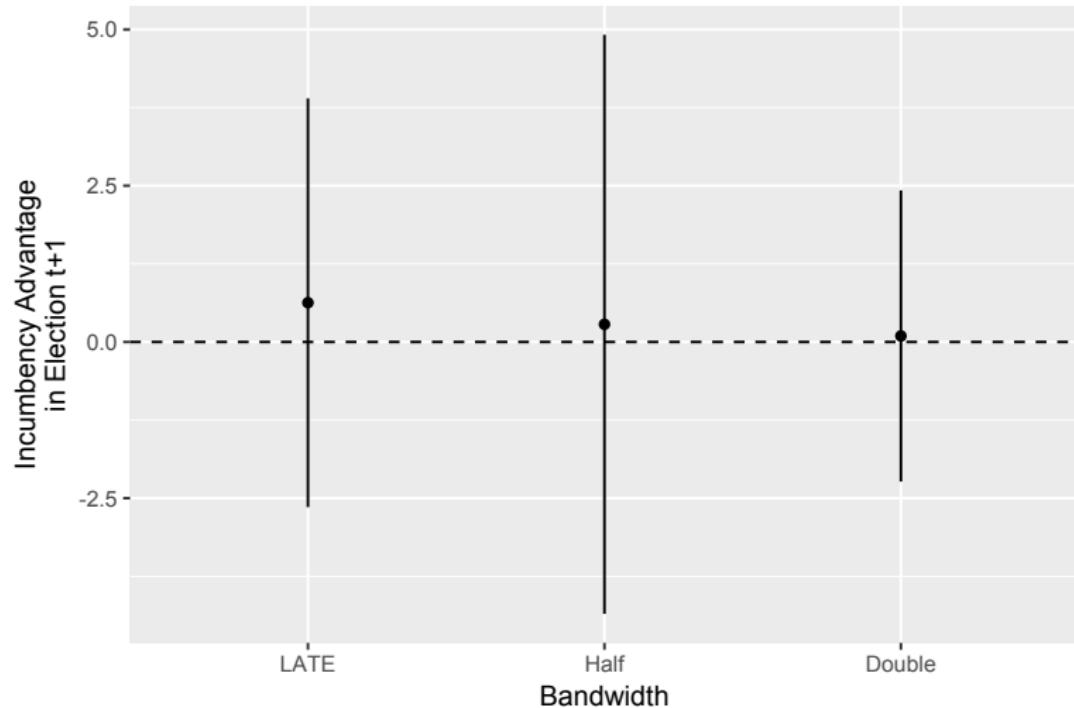
散布図

区間ごとの平均値の散布図



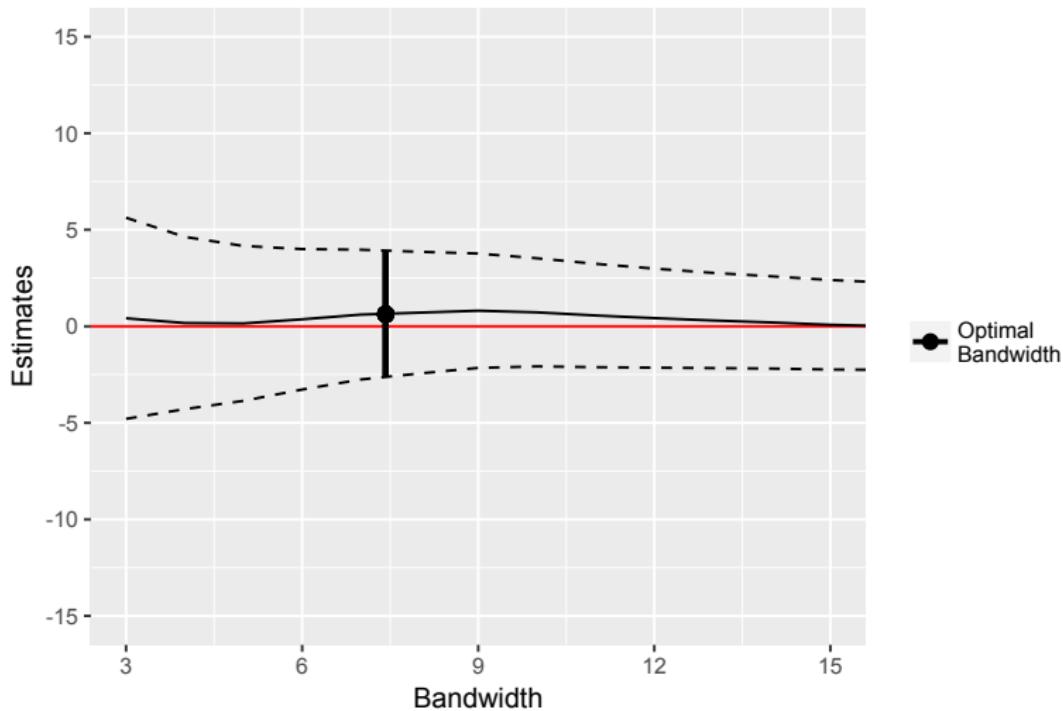
因果効果

現職効果は見られない!



頑健性チェック

バンド幅を動かしても推定値は大きく変化しない



頑健性チェック

局所多項式回帰でも同様

