

# Difference-in-Difference

Jaehyun SONG

Kobe University

2016/06/07

# Diff-in-Diff?

- 日本語では「差分の差分法」
- 集団単位における因果効果を確認する
- 数式で表すと引き算しか出てこない

$$\begin{aligned}\hat{\beta}_{DID} &= \underbrace{(\bar{Y}_{W=1,T=1} - \bar{Y}_{W=1,T=0})}_{\text{Difference}} - \underbrace{(\bar{Y}_{W=0,T=1} - \bar{Y}_{W=0,T=0})}_{\text{Difference}} \\ &= \underbrace{\Delta \bar{Y}_{W=1} - \Delta \bar{Y}_{W=0}}_{\text{Difference in Difference}}\end{aligned}$$

# 最初の DID の例

## John Snow の疫学 (コレラ) 研究

- 水道会社とコレラ率
- L 社が水源を移動させたことによる変化を測定
- 水源を変えなかった S 社に比べ、L 社のコレラによる死亡率が急激に低下  
(S 社: 0.503% ↔ L 社: 0.094%)



# 例: ほとんど無害なラーメン二郎

$$\begin{aligned}\hat{\beta}_{DID} &= \underbrace{(\bar{Y}_{W=1,T=1} - \bar{Y}_{W=1,T=0})}_{(1)} - \underbrace{(\bar{Y}_{W=0,T=1} - \bar{Y}_{W=0,T=0})}_{(2)} \\ &= \underbrace{\Delta \bar{Y}_{W=1} - \Delta \bar{Y}_{W=0}}_{(3)}\end{aligned}$$

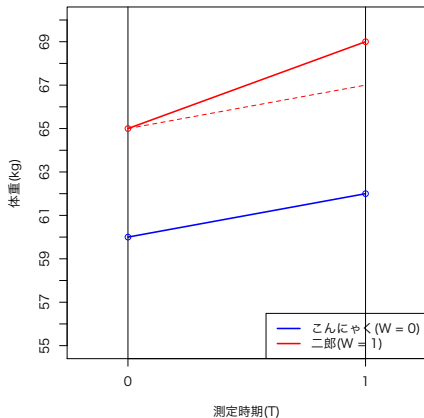
- ラーメン二郎を食べたら  $W = 1$ 、食べなかったら  $W = 0$
- $T = 0$ : 今日、 $T = 1$ : 明日、 $\bar{Y}$ : 体重の平均

(1) ラーメンを食べた場合の体重の増加量

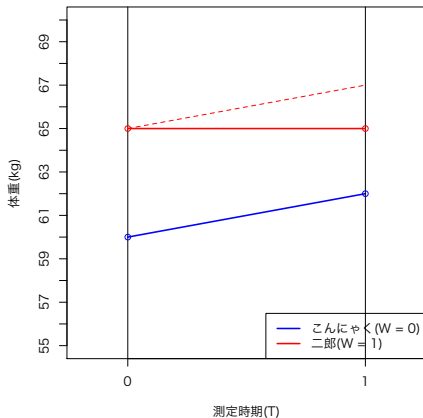
(2) ラーメンを食べなかった場合の体重の増加量

# 例: ほとんど無害なラーメン二郎

かなり有害なラーメン二郎



ほとんど無害なラーメン二郎(むしろ健康的)



- 処置が無作為に割り当てられたら、DID 推定量 ( $\hat{\beta}_{DID}$ ) は処置群における平均因果効果 (ACE) の不偏一致推定量
- 回帰分析でも推定可能
  - 推定値の標準誤差が簡単に計算可能
  - 回帰分析を用いることで共変量が投入可能  
ただし、共変量の選択には注意が必要

# DID と回帰分析

$$\hat{Y} = \mu + \gamma W + \phi T + \delta(W \cdot T)$$

$Y$ : 目的変数、 $W$ : 処置、 $T$ : 時間

- 処置群の差分:  $\hat{Y}_{W=1,T=1} - \hat{Y}_{W=1,T=0}$   
$$\delta + \phi = \underbrace{(\mu + \gamma + \phi + \delta)}_{\hat{Y}_{W=1,T=1}} - \underbrace{(\mu + \gamma)}_{\hat{Y}_{W=1,T=0}}$$
- 統制群の差分:  $\hat{Y}_{W=0,T=1} - \hat{Y}_{W=0,T=0}$   
$$\phi = \underbrace{(\mu + \phi)}_{\hat{Y}_{W=0,T=1}} - \underbrace{(\mu)}_{\hat{Y}_{W=0,T=0}}$$
- 差分の差分:  $\delta = (\phi + \delta) - (\phi)$

**DID 推定値 = 処置変数と時間変数の交差項の係数の推定値**

# DID と固定効果

複数のユニットと時間を用いる場合は固定効果でコントロール

$$\hat{Y} = \mu + \delta \text{LEGAL}_{st} + \sum_{k=\text{Alaska}}^{\text{Wyoming}} \beta_k \text{STATE}_{ks} + \sum_{j=1971}^{1983} \gamma_j \text{YEAR}_{jt}$$

(Angrist and Pischke(2015. p. 194) より)

- $\text{LEGAL}_{st}$ :  $t$  年における  $s$  州の飲酒可能年齢
- $\text{STATE}$ : 州 (ダミー変数)
- $\text{YEAR}$ : 年 (ダミー変数)
- $Y$ : 死亡者数

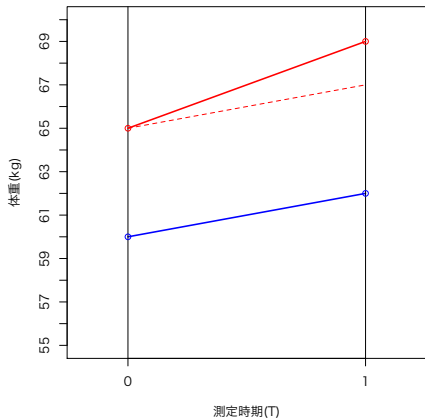
各州と年度固有の傾向および初期値を除去し、  
 $\text{LEGAL}$  のみの効果を推定



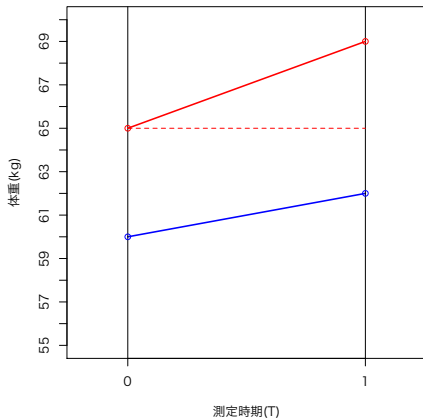
# DID の前提: 並行トレンド

処置群がもし処置を受けなかった場合の効果量 (差分) は統制群のそれと一致する。

並行トレンドを満たす例



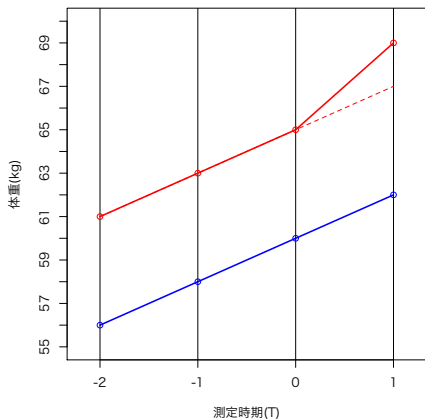
並行トレンドを満たしていない例



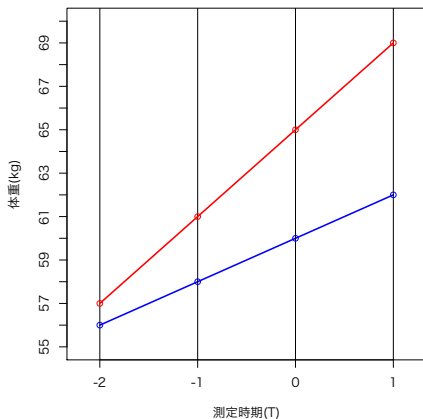
# 並行トレンドのチェック (1)

過去のデータを見る。

並行トレンドを満たす例



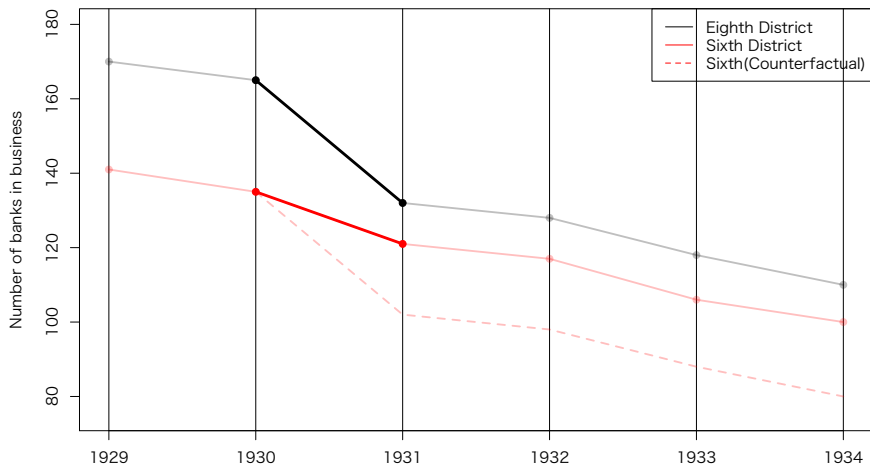
並行トレンドを満たしていない例



回帰分析の際にトレンド変数の投入でも確認可能

# 並行トレンドのチェック (1): 例

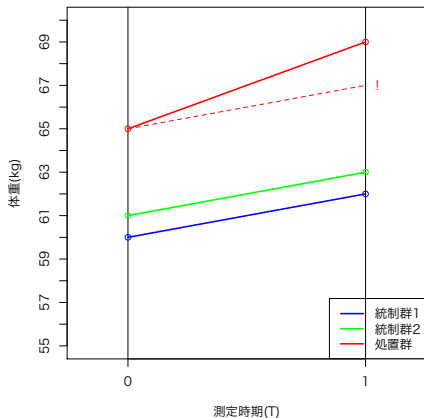
Angrist and Pischke(2015. p.186) より



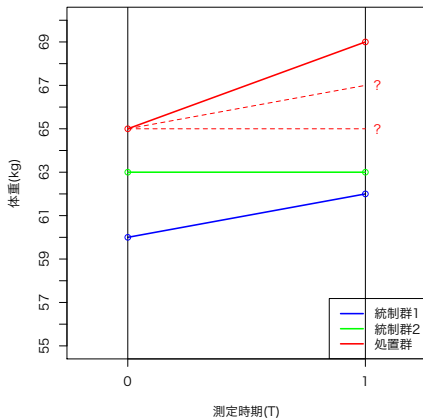
# 並行トレンドのチェック (2)

プラセボ・テスト: もう一つの統制群を用意する

プラセボテスト(O)



プラセボテスト(X)



- 推定が簡単
- ユニット固有の属性がコントロール可能 (固定効果の発想)

- 並行トレンド
  - 過去のデータ、プラセボテストなど
- 逆の因果 (主に政策効果)
  - ある政策の実行が予告されることによって並行トレンドが崩れる可能性
  - $t$  と  $-t$  の変数を同時投入することで確認?
- 共変量選択
  - 因果推論における回帰分析が持つ限界