

# 特殊講義：ヨーロッパ国際政治経済 マッチング

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25日

# マッチングの考え方

# 因果推論と内生性

**内生性:** 処置変数と誤差項間の相関関係

- 内生性は因果推論の敵

例)

- 処置変数 = ソンさんの講義を履修するか否か
- 結果変数 = 10年後の年収
- もし、やる気のある学生が履修する傾向があるとしたら？
- やる気のある学生は履修の有無と関係なく、高所得者になりやすい。  
⇒ 「やる気」は処置と結果、両方と連関している

内生性を除去する最良の手法 ⇒ **RCT**

# RCTの限界

- 高費用
  - ▶ 数万 ~ 数億円
- 倫理的な問題による実行不可能性
  - ▶ 喫煙と健康
  - ▶ Philip Zimbardo. 2008. *The Lucifer Effect: How Good People Turn Evil.* Rider.
- 外的妥当性の問題
- 回顧的因果推論には不向き
  - ▶ 主に介入 (intervention) の効果が推定対象

# 因果推論

## The Three Layer Causal Hierarchy (Pearl 2009: p. 29)

Level (Symbol)	Typical Activity	Typical Question	Examples
1. Association $P(y x)$	Seeing	What is? How would seeing $X$ change my belief in $Y$ ?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y do(x), z)$	Doing	What if? What if I do $X$ ?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x x', y')$	Imagining, Retrospection	Why? Was it $X$ that caused $Y$ ? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?

# 観察データを用いた因果推論

もし、 $X$  をしたら  $Y$  はどうなっただろうか

- 過去を対象に RCT を行うことは不可能
- 過去に収集された観察データが使用することに
- マッチング、回帰不連続デザイン、差分の差分法、操作変数法など

## 割当メカニズム (assignment mechanism)

- ユニットが処置を受けるか否かを規定するメカニズム
  - 例) 「やる気」が「履修」を規定
- 無作為割当なら無作為に処置を受けるか否かが決まるため、考える必要がない。

# 内生性への対処

Day2\_Data1.csv の例 (架空データ)

- 明らかに「やる気」と「履修」は連関
- 履修有無による平均年収の差は 266 万円

ID	やる気	履修	年収	ID	やる気	履修	年収	ID	やる気	履修	年収
1	x	○	659	11	x	×	239	21	x	×	155
2	○	○	587	12	○	×	276	22	○	○	768
3	○	○	628	13	○	○	609	23	○	×	463
4	○	○	563	14	x	×	254	24	○	×	309
5	○	○	531	15	x	○	423	25	x	×	304
6	x	x	79	16	x	○	172	26	○	○	408
7	x	○	356	17	x	×	20	27	x	×	259
8	x	x	176	18	○	×	447	28	○	○	516
9	x	x	339	19	○	○	498	29	○	×	476
10	○	○	520	20	○	○	648	30	x	×	110

# 内生性への対処

方法: 処置変数と結果変数に影響を与える要因を揃える

- 「やる気」のない学生だけに絞ってみる
- 履修有無による平均年収の差は 209 万円

ID	やる気	履修	年収	ID	やる気	履修	年収
1	×	○	659	6	×	×	79
7	×	○	356	8	×	×	176
15	×	○	423	9	×	×	339
16	×	○	172	11	×	×	239
				14	×	×	254
				17	×	×	20
				21	×	×	155
				25	×	×	304
				27	×	×	259
				30	×	×	110
平均		402.5				193.5	

# 内生性への対処

方法: 処置変数と結果変数に影響を与える要因を揃える

- 「やる気」のある学生だけに絞ってみる
- 履修有無による平均年収の差は 176.3 万円

ID	やる気	履修	年収	ID	やる気	履修	年収
2	○	○	587	12	○	×	276
3	○	○	628	18	○	×	447
4	○	○	563	23	○	×	463
5	○	○	531	24	○	×	309
10	○	○	520	29	○	×	476
13	○	○	609				
19	○	○	498				
20	○	○	648				
22	○	○	768				
26	○	○	408				
28	○	○	516				
平均			570.5				394.2

# 内生性への対処

履修 ( $T_i = 1$ )	
やる気	平均年収
×	402.5
○	570.5

非履修 ( $T_i = 0$ )	
やる気	平均年収
×	193.5
○	394.2

↓

ID	N	やる気	平均年収 ( $Y_i$ )		ITE
			$Y_i(T_i = 1)$	$Y_i(T_i = 0)$	
1	14	×	402.5	193.5	209.0
2	16	○	570.5	394.2	176.3
加重平均			191.5		

講義履修の因果効果は 191.5 万円

# Matching の考え方

割当メカニズムを想定し、交絡要因が同じユニット同士を比較

- **交絡要因**: 処置変数と結果変数、両方と関係のある変数
- 以下の条件が満たされる場合、マッチングで正確な因果効果の推定が可能

‣ **条件付き独立の仮定** (Conditional Independence Assumption; CIA)

$$\Rightarrow \{Y_i(T_i = 1), Y_i(T_i = 0)\} \perp T_i | X_i$$

★  $T_i$ : 学生  $i$  の履修有無、 $X_i$ : 学生  $i$  のやる気

★ やる気 (=交絡要因) が同じ場合、学生  $i$  がソンさんの講義を履修するか否か (=処置変数) は彼 (女) の将来収入 (=結果変数) と関係なく決まる。

★ **処置変数を外生変数として扱うことが可能**に

‣ CIA が満たされるためには、割当メカニズム上のすべての交絡要因が必要

# 条件付き独立とは？

ID	Z	T	$Y_0$	$Y_1$
1	0	0	0	1
2	0	0	1	0
3	0	0	0	0
4	0	0	0	0
5	0	1	0	0
6	0	1	1	0
7	0	1	0	0
8	0	1	0	1
9	1	0	1	1
10	1	0	1	0
11	1	0	0	1
12	1	1	1	1
13	1	1	1	1
14	1	1	0	1
15	1	1	0	1
16	1	1	0	1
17	1	1	1	1
18	1	1	1	0
19	1	1	1	0
20	1	1	1	0

	$Y_0$	$Y_1$
$T = 0$	(A) <b>0.429</b>	(B) 0.429
$T = 1$	(C) 0.538	(D) <b>0.538</b>

- 処置効果は  $0.538 - 0.429 = 0.109$  だが…
  - もし、2つ ( $T = 0$  or  $1$ ) の集団が同質なら
    - $T = 0$  群が処置を受けた場合の  $Y$  (=B) は (D) と一致すべき
    - $T = 1$  群が処置を受けなかった場合の  $Y$  (=C) は (A) と一致すべき
- ⇒ 交換可能性が成立せず



# 条件付き独立とは？

ID	Z	T	$Y_0$	$Y_1$
1	0	0	0	1
2	0	0	1	0
3	0	0	0	0
4	0	0	0	0
5	0	1	0	0
6	0	1	1	0
7	0	1	0	0
8	0	1	0	1
9	1	0	1	1
10	1	0	1	0
11	1	0	0	1
12	1	1	1	1
13	1	1	1	1
14	1	1	0	1
15	1	1	0	1
16	1	1	0	1
17	1	1	1	1
18	1	1	1	0
19	1	1	1	0
20	1	1	1	0

Z で条件付けた場合 ( $Z = 0$ )

	$Y_0$	$Y_1$
$T = 0$	(A) <b>0.250</b>	(B) 0.250
$T = 1$	(C) 0.250	(D) <b>0.250</b>

- 処置効果は  $0.250 - 0.250 = 0$
  - (A) と (C)、(B) と (D) が一致
- ⇒ 交換可能性が成立



# 条件付き独立とは？

ID	Z	T	$Y_0$	$Y_1$
1	0	0	0	1
2	0	0	1	0
3	0	0	0	0
4	0	0	0	0
5	0	1	0	0
6	0	1	1	0
7	0	1	0	0
8	0	1	0	1
9	1	0	1	1
10	1	0	1	0
11	1	0	0	1
12	1	1	1	1
13	1	1	1	1
14	1	1	0	1
15	1	1	0	1
16	1	1	0	1
17	1	1	1	1
18	1	1	1	0
19	1	1	1	0
20	1	1	1	0

Z で条件付けた場合 ( $Z = 1$ )

	$Y_0$	$Y_1$
$T = 0$	(A) <b>0.667</b>	(B) 0.667
$T = 1$	(C) 0.667	(D) <b>0.667</b>

- 処置効果は  $0.667 - 0.667 = 0$
  - (A) と (C)、(B) と (D) が一致
- ⇒ 交換可能性が成立



# 条件付き独立とは？

ID	Z	T	$Y_0$	$Y_1$
1	0	0	0	1
2	0	0	1	0
3	0	0	0	0
4	0	0	0	0
5	0	1	0	0
6	0	1	1	0
7	0	1	0	0
8	0	1	0	1
9	1	0	1	1
10	1	0	1	0
11	1	0	0	1
12	1	1	1	1
13	1	1	1	1
14	1	1	0	1
15	1	1	0	1
16	1	1	0	1
17	1	1	1	1
18	1	1	1	0
19	1	1	1	0
20	1	1	1	0

条件付き独立が成立

- 交換可能性が成立
- 処置群を統制群に、統制群を処置群にしても同じ結果が得られること

# 重回帰分析との比較

## 重回帰分析における回帰係数の解釈

- **他の変数すべてが同じ場合**、ある変数が 1 単位変化する時の結果変数の変化量
- マッチングと同じ?!
- 重回帰分析とマッチングの結果が近似することも ○
  - 単回帰分析
    - ★  $\widehat{\text{年収}} = 260.4 + 265.3 \text{ 履修}$
  - 重回帰分析
    - ★  $\widehat{\text{年収}} = 198.6 + 191.2 \text{ 履修} + 185.4 \text{ やる気}$
  - マッチングから推定された ATE: **191.5**

# 重回帰分析との比較

## 重回帰分析における回帰係数の解釈

- **他の変数すべてが同じ場合**、ある変数が 1 単位変化する時の結果変数の変化量
  - マッチングと同じ?!
- 
- 実質的にマッチングと回帰分析は同じという見解も (Angrist and Pischke 2009)
    - ▶ 具体的に言えば、回帰分析はマッチングの特殊な形態
  - 回帰分析は  $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$  の関数型を仮定 (parametric)
  - 回帰分析において誤差項の平均値は必ず 0 を仮定 ( $\mathbb{E}(\varepsilon|T, X) = 0$ )
    - ▶ マッチングの場合、( $\mathbb{E}(\varepsilon|T = 0, X) = \mathbb{E}(\varepsilon|T = 1, X)$ )
  - 回帰分析はオーバーラップ条件を無視する
    - ▶ マッチングされないケースでも、線形関数によって予測されてしまう
  - 結論: 回帰分析より柔軟、拡張性がある

# ATE, ATET, ATEC

## 3種類の因果効果

- ① **ATE** (Average Treatment Effect): 平均処置効果
  - ② **ATT** (Average Treatment Effect on Treated): 処置群における平均処置効果
    - ▶ 処置群が処置を受けなかった場合を想定した因果効果
  - ③ **ATC** (Average Treatment Effect on Control): 統制群における平均処置効果
    - ▶ 統制群が処置を受けた場合を想定した因果効果
- 因果効果は一般的に母集団ではなく、サンプルから推定されるため、「SATE/SATT/SATC」と呼ばれる場合も
  - 他にも RIE (Retrospective Intervention Effect) なども (Samii et al. 2016)
  
  - RCT では主に ATE が推定対象であるものの、マッチングでは区分される場合が多い。
    - ▶ 統計ソフトウェアによっては ATT を因果効果の推定値として表示する場合もある。

# ATE, ATT, ATC

ATT: 処置群における平均処置効果

- **処置群の潜在的結果を統制群から割り当てる。**
  - 処置群は  $Y_i(T_i = 1)$  が観察済みであり、潜在的結果は  $Y_i(T_i = 0)$
- やる気のない学生の  $Y_i(T_i = 0)$  は 193.5、ある学生は 394.2

ID	やる気	将来年収		ITE
		$Y_i(T_i = 1)$	$Y_i(T_i = 0)$	
1	×	659	193.5	465.5
2	○	587	394.2	192.8
3	○	628	394.2	233.8
4	○	563	394.2	168.8
5	○	531	394.2	136.8
7	×	356	193.5	162.5
10	○	520	394.2	125.8
13	○	609	394.2	214.8
15	×	423	193.5	229.5
:	:	:	:	:
26	○	408	394.2	13.8
28	○	516	394.2	121.8
平均			<b>185.1</b>	

# ATE, ATT, ATC

ATC: 統制群における平均処置効果

- 統制群の潜在的結果を処置群から割り当てる。
  - 処置群は  $Y_i(T_i = 0)$  が観察済みであり、潜在的結果は  $Y_i(T_i = 1)$
- やる気のない学生の  $Y_i(T_i = 1)$  は 402.5、ある学生は 570.5

ID	やる気	将来年収		ITE
		$Y_i(T_i = 1)$	$Y_i(T_i = 0)$	
6	×	402.5	79	323.5
8	×	402.5	176	226.5
9	×	402.5	339	63.5
11	×	402.5	239	163.5
12	○	570.5	276	294.5
14	×	402.5	254	148.5
17	×	402.5	20	382.5
18	○	570.5	447	123.5
21	×	402.5	155	247.5
:	:	:	:	:
29	○	570.5	476	94.5
30	×	402.5	110	292.5
平均			<b>198.1</b>	

# ATE, ATT, ATC

ATE: 平均処置効果

- ATT と ATC の加重平均
- 今回は処置群と統制群が 15:15 だったので、単純平均で 191.5
  - ▶ 先ほどと同じ結果

$$ATE = \frac{N_{\text{treated}}}{N_{\text{all}}} ATT + \frac{N_{\text{control}}}{N_{\text{all}}} ATC$$

# 様々なマッチング手法

# マッチングいろいろ

- ① Exact Matching
- ② Nearest-neighbor Matching
  - k-nearest Neighbor Matching
  - Caliper Matching (Radius Matching)
- ③ Coarsened Exact Matching
- ④ Propensity Score Matching
  - Inverse Probability Weighting Matching
  - Ensemble Matching

# Exact Matching

## Exact Matching

- 「正確マッチング」、「厳格なマッチング」などで訳される
- これまで見てきた方法が Exact Matching
  - データ内の共変量<sub>(交絡要因)</sub> が完全に一致するケース同士の比較
- 共変量が少数、かつ、名目 or 順序変数の場合、使用可
- 共変量が多数、または連続変数の場合は実質的に無理
  - 次元の呪い or 次元爆発



# Nearest-neighbor Matching

## Nearest-neighbor Matching

- 「最近傍マッチング」と訳される。
  - 共変量が連続変数、多次元の場合、「完全に一致」ケースはない場合がほとんど
- ⇒ 「一致」ではなく、「最も似ている」ケース同士と比較

「近さ」の基準

- ① Manhattan Distance
- ② Standardized Euclidean Distance
- ③ **Mahalanobis Distance** (← 最もよく使われる基準)

# Nearest-neighbor Matching

「近さ」の基準 1: Manhattan Distance / City-block Distance

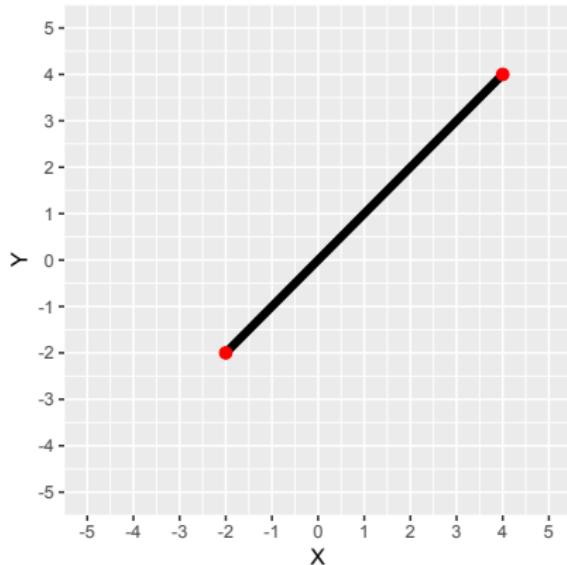
$$d(i,j) = |X_i - X_j| + |Y_i - Y_j|$$



# Nearest-neighbor Matching

「近さ」の基準 2: Standardized Euclidean Distance

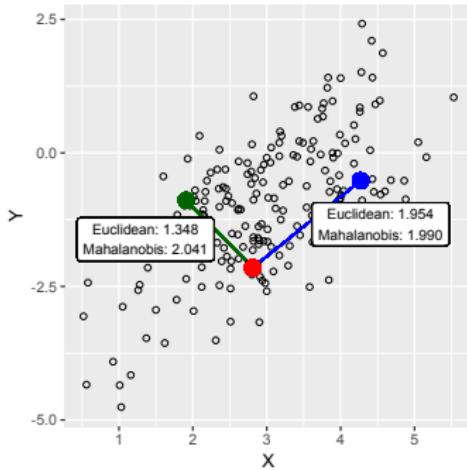
$$d(i,j) = \sqrt{\left(\frac{X_i - X_j}{\sigma_X}\right)^2 + \left(\frac{Y_i - Y_j}{\sigma_Y}\right)^2}$$



# Nearest-neighbor Matching

「近さ」の基準 3: Mahalanobis Distance (共変量間の相関が 0 の場合、正則化ユークリッド距離)

$$d(i,j) = \sqrt{\frac{1}{1-\rho_{XY}^2} \left[ \left( \frac{X_i - X_j}{\sigma_X} \right)^2 + \left( \frac{Y_i - Y_j}{\sigma_Y} \right)^2 - 2\rho_{XY} \left( \frac{X_i - X_j}{\sigma_X} \right) \left( \frac{Y_i - Y_j}{\sigma_Y} \right) \right]}$$



# マッチング方法

ATT の場合、処置群のケースに統制群の中で最も近いケース 1 個を割当

- 近さの測定はマハラノビス距離が一般的
- 処置群内の 1 ケースに複数の統制群ケースを割り当てる場合も ○
  - k-nearest Neighbor Matching
  - Caliper Matching

# k-nearest Neighbor Matching

## k-nearest Neighbor Matching

- 「k-最近傍マッチング」と訳される。
- 最も近い 1 個ケースを潜在的結果として使うのではなく、**最も近い  $k$  個のケースの平均値**を潜在的結果として用いる。

$j(m)$ :  $i$  から  $m$  番目に近いケース  $j$ .

$$Y_i(T_i = 0) \begin{cases} Y_i & \text{if } T_i = 0, \\ \frac{1}{K} \sum_{m=1}^K Y_{j(m)} & \text{if } T_i = 1. \end{cases}$$

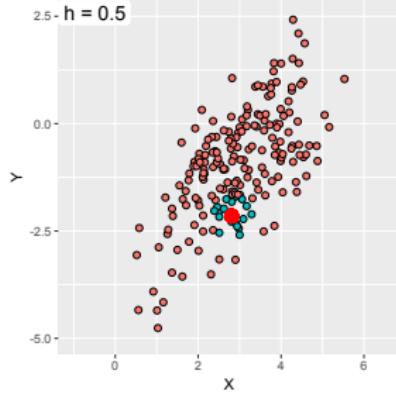
- 最適  $k$  を決める理論的基準は無し
  - $k$  を大きくすると、モデルの分散 ↓
  - ただし、モデルの分散が小さい = バイアスが拡大 ( $\Rightarrow$  Bias-variance tradeoff)
  - $k$  を変化させることによって結果がどのように変わるか観察

# Caliper Matching

## Caliper Matching

- 「カリパーマッチング」と訳される (訳されてない!)
- 半径  $h$  の中にある全てのケースの平均値を潜在的結果として使用

$$Y_i(T_i = 0) \begin{cases} Y_i & \text{if } T_i = 0, \\ \frac{\sum_{j=1}^N I(T_j=0, d(i,j) < h) \cdot Y_j}{\sum_{j=1}^N I(T_j=0, d(i,j) < h)} & \text{if } T_i = 1. \end{cases}$$



# Coarsened Exact Matching

## Coarsened Exact Matching (Iacus, King, and Porro 2011)

- 定説ではなく、「CEM」で呼ばれる（粗い正確マッチング？）
- アルゴリズムは簡単
  - ① 共変量をいくつかの層 (strata) へ分割する。
  - ② 各層にそれぞれ該当する処置・統制ユニットを入れる。
  - ③ 最低一つ以上の処置・統制ユニットがない層は捨てる。
  - ④ 各層の処置・統制ユニットの結果変数の差分を計算し、すべての層に對して加重平均
- 層を細かくするほど Exact Matching へ近づく
  - ただし、分析に使うケースが減少
  - バイアス ↓; 分散 ↑

# CEM の例

## Raw Data (Day2\_Data2.csv)

- 年齢は 10 歳刻み、学歴は大卒以上・未満に層化

ID	年齢	学歴	処置	結果	ID	年齢	学歴	処置	結果
1	29	院	0	6	13	57	高	0	4
2	41	大	0	3	14	25	院	1	5
3	31	院	1	7	15	55	中	1	9
4	39	院	0	5	16	48	院	0	2
5	53	大	0	6	17	23	専	0	2
6	59	大	0	1	18	34	大	1	4
7	37	高	1	8	19	42	大	1	9
8	44	中	0	4	20	23	高	0	4
9	51	中	0	2	21	22	高	1	8
10	59	小	1	8	22	49	大	0	9
11	21	大	1	4	23	45	高	1	6
12	24	中	1	6	24	33	大	0	8

# CEM の例

## Raw Data (Day2\_Data2.csv)

- 年齢は 10 歳刻み、学歴は大卒以上 (H)・未満 (L) に層化
  - カテゴリが少なくなり、マッチングされやすくなる

ID	年齢	学歴	処置	結果	ID	年齢	学歴	処置	結果
1	20 代	H	0	6	13	50 代	L	0	4
2	40 代	H	0	3	14	20 代	H	1	5
3	30 代	H	1	7	15	50 代	L	1	9
4	30 代	H	0	5	16	40 代	H	0	2
5	50 代	H	0	6	17	20 代	L	0	2
6	50 代	H	0	1	18	30 代	H	1	4
7	30 代	L	1	8	19	40 代	H	1	9
8	40 代	L	0	4	20	20 代	L	0	4
9	50 代	L	0	2	21	20 代	L	1	8
10	50 代	L	1	8	22	40 代	H	0	9
11	20 代	H	1	4	23	40 代	L	1	6
12	20 代	L	1	6	24	30 代	H	0	8

# CEM の例

層ごとにケースをマッチング

- ペアが組めない層も出現し得る (例: 30 代&L)

年齢	学歴	ID	処置 ○		処置 ×		
			処置	結果	ID	処置	結果
20 代	H	11	1	5	1	0	6
		14	1	5			
	L	12	1	6	17	0	2
		21	1	8	20	0	4
30 代	H	3	1	7	4	0	5
		18	1	4	24	0	8
	L	7	1	8			
40 代	H	19	1	9	2	0	3
					16	0	2
					22	0	9
	L	23	1	6	8	0	4
50 代	H				5	0	6
					6	0	1
	L	10	1	8	9	0	2
		15	1	9	13	0	4

# CEM の例

対のない層を除外

年齢	学歴	処置 ○			処置 ×		
		ID	処置	結果	ID	処置	結果
21-30	H	11	1	5	1	0	6
		14	1	5			
	L	12	1	6	17	0	2
		21	1	8	20	0	4
31-40	H	3	1	7	4	0	5
		18	1	4	24	0	8
41-50	H	19	1	9	2	0	3
					16	0	2
					22	0	9
	L	23	1	6	8	0	4
51-60	L	10	1	8	9	0	2
		15	1	9	13	0	4

# CEM の例

## 各ユニットの重みを計算

$m_{C,T}$ : 統制・処置ケースの数、 $m_{C,T}^s$ : 層  $s$  内の統制・処置ケースの数

$$w_i = \begin{cases} 1 & \text{if } T_i = 1, \\ \frac{m_C}{m_T} \cdot \frac{m_T^s}{m_C^s} & \text{if } T_i = 0. \end{cases}$$

年齢	学歴	処置 ○		処置 ×	
		ID	重み	ID	重み
21-30	H	11	1.000	1	2.200
		14	1.000		
	L	12	1.000	17	1.100
		21	1.000	20	1.100
31-40	H	3	1.000	4	1.100
		18	1.000	24	1.100
	H	19	1.000	2	0.367
				16	0.367
				22	0.367
41-50	L	23	1.000	8	1.100
	L	10	1.000	9	1.100
		15	1.000	13	1.100
51-60					

# CEM の例

## 重み付け回帰分析 (Weighted Least Squares; WLS)

- $w = \{2.200, 0.367, 1.000, 1.100, 0.000, 0.000, \dots, 1.100\}^T$ 
  - ▶ マッチングされないケースの重みは 0 にするか、分析から除外  
(上記の重みベクトルは、除外ケースの重みを 0 にした場合)
  - ▶  $\beta = (X^T W X)^{-1} X^T W Y$
- R だと `lm(formula, data, weight = ...)` で推定可能
  - ▶ `cem` or `MatchIt` パッケージならもっと簡単
- $\widehat{\text{Outcome}} = 4.567 + \mathbf{2.033} \cdot \text{Treat}$
- 処置群における因果効果 (ATT) = 2.033

傾向スコア

# Propensity Score とは

## Propensity Score

- 「傾向スコア」と訳される。
- 簡単にいふと「あるユニット  $i$  が**処置を受ける確率**」  
 $= e_i = \Pr(T_i = 1 | X_i)$

なぜ傾向スコア？

- マッチングの限界
  - ① 次元の問題 (dimension problem)
  - ② 恣意性
  - ③ カテゴリ変数の扱い方
  - ④ スケールの問題

# 傾向スコアの計算

処置変数 ( $T_i$ ) を応答変数とし、共変量 ( $X_i$ ) を説明変数とする

- 一般的に、ロジットやプロビット回帰分析で推定する。
  - 他にも色々ある (Support Vector Machine, Decision Tree, Neural Network, ...)
  - 色んな手法で算出した傾向スコアを重み付けして合成することも可能  
⇒ Ensemble Method (Samii, Paler, and Zukerman 2016)
- 推定された予測確率 ⇒ 傾向スコア
  - R では `$fitted.value` で抽出可
  - 傾向スコアは **多くの共変量を一つの変数に集約したもの**  
⇒ 傾向スコアを統制した回帰分析で因果効果を推定  
⇒ 傾向スコアを用いて最近傍マッチング

# 傾向スコア・マッチングの手順

## 1. 割り当てメカニズム

- $\text{Pr}(\text{処置の有無}) \propto \beta_0 + \beta_1 \cdot \text{年齢} + \beta_2 \cdot \text{学歴}$

ID	年齢	学歴	処置	結果	ID	年齢	学歴	処置	結果
1	29	院	0	6	13	57	高	0	4
2	41	大	0	3	14	25	院	1	5
3	31	院	1	7	15	55	中	1	9
4	39	院	0	5	16	48	院	0	2
5	53	大	0	6	17	23	専	0	2
6	59	大	0	1	18	34	大	1	4
7	37	高	1	8	19	42	大	1	9
8	44	中	0	4	20	23	高	0	4
9	51	中	0	2	21	22	高	1	8
10	59	小	1	8	22	49	大	0	9
11	21	大	1	4	23	45	高	1	6
12	24	中	1	6	24	33	大	0	8

# 傾向スコア・マッチングの手順

## 2. 傾向スコアの算出

- ① PS.Fit <- glm(処置 ~ 年齢 + 学歴, family = binomial("logit"))
- ② PS <- PS.Fit\$fitted.value

Coef.	Est.	S.E.
$\beta_0$ : Intercept	3.839	2.363
$\beta_1$ : Age	-0.059	0.039
$\beta_2$ : Education	-0.420	0.304

# 傾向スコア・マッチングの手順

## 2. 傾向スコアの算出

- ① PS.Fit <- glm(処置 ~ 年齢 + 学歴, family = binomial("logit"))
- ② PS <- PS.Fit\$fitted.value

ID	年齢	学歴	処置	結果	PS	ID	年齢	学歴	処置	結果	PS
1	29	院	0	6	0.406	13	57	高	0	4	0.317
2	41	大	0	3	0.339	14	25	院	1	5	0.463
3	31	院	1	7	0.378	15	55	中	1	9	0.443
4	39	院	0	5	0.275	16	48	院	0	2	0.183
5	53	大	0	6	0.203	17	23	専	0	2	0.692
6	59	大	0	1	0.152	18	34	大	1	4	0.437
7	37	高	1	8	0.601	19	42	大	1	9	0.326
8	44	中	0	4	0.603	20	23	高	0	4	0.774
9	51	中	0	2	0.502	21	22	高	1	8	0.784
10	59	小	1	8	0.489	22	49	大	0	9	0.243
11	21	大	1	4	0.624	23	45	高	1	6	0.485
12	24	中	1	6	0.831	24	33	大	0	8	0.451

# 傾向スコア・マッチングの手順

## 3. 傾向スコアが最も近い統制群を割り当てる

- ATT の場合

処置群			統制群			
ID	結果 (T)	PS	ID	結果 (C)	PS	(T) - (C)
3	7	0.378	1	6	0.406	1
7	8	0.601	8	4	0.603	4
10	8	0.489	9	2	0.502	6
11	4	0.624	8	4	0.603	0
12	6	0.831	20	4	0.774	2
14	5	0.463	24	8	0.451	-3
15	9	0.443	24	8	0.451	1
18	4	0.437	24	8	0.451	-4
19	9	0.326	13	4	0.317	5
21	8	0.784	20	4	0.774	4
23	6	0.485	9	2	0.502	4
平均 (=ATT)			1.818			

# 傾向スコア・マッチングの手順

## 3. 傾向スコアが最も近い処置群を割り当てる

- ATC の場合

処置群			統制群			
ID	結果 (T)	PS	ID	結果 (C)	PS	(T) - (C)
3	7	0.378	1	6	0.406	1
19	9	0.326	2	3	0.339	6
19	9	0.326	4	5	0.275	4
19	9	0.326	5	6	0.203	3
19	9	0.326	6	1	0.152	8
7	8	0.601	8	4	0.603	4
10	8	0.489	9	2	0.502	6
19	9	0.326	13	4	0.317	5
19	9	0.326	16	2	0.183	7
11	4	0.624	17	2	0.692	2
21	8	0.784	20	4	0.774	4
19	9	0.326	22	9	0.243	0
15	9	0.443	24	8	0.451	1
平均 (=ATC)			3.923			

# 傾向スコア・マッチングの手順

## 3. 傾向スコアが最も近い統制群を割り当てる

- ATE の場合
- ATT と ATC の加重平均

$$\begin{aligned} \text{ATE} &= \frac{N_{Treat}}{N_{All}} \cdot \text{ATT} + \frac{N_{Control}}{N_{All}} \cdot \text{ATC} \\ &= \frac{11}{24} 1.818 + \frac{13}{24} 3.923 = 2.958 \end{aligned}$$

▶ `weighted.mean(c(1.818, 3.923), c(11, 13))`

# 処置を受ける確率の計算

処置を受ける確率 = 傾向スコア

- 一般的にはロジスティック/プロビット回帰分析が使われる
- ただし、確率が予測できるなら他の手法でも良い
  - Covariate Balancing Propensity Score
  - Entropy Balancing
  - Nerural Network
  - Support Vector Machine
  - Random Forest
  - など
- 複数の手法を組み合わせる (= ensemble) することも可能  
⇒ Super Learner Algorithm

# 傾向スコアのもう一つの使い方

## IPW: Inverse Probability Weighting (Rubin 1985)

- 「逆確率重み付け」と訳される
- 実際に処置を受けた ( $T_i = 1$ ) にもかかわらず、処置を受ける傾向が小さい場合は分析において大きい重み
  - 傾向スコアを重み変数として用いる。
  - $e_i$  が 1 または 0 に近い場合、一部のケースに非常に大きい重みを付けることになるため、注意が必要

$$w_i = \begin{cases} \frac{1}{e_i} & \text{if } T_i = 1, \\ \frac{1}{1-e_i} & \text{if } T_i = 0. \end{cases}$$

$e_i$ :  $i$  の傾向スコア、 $T_i$ :  $i$  の処置有無  $\in \{0, 1\}$

# 傾向スコアのもう一つの使い方

## IPW: Inverse Probability Weighting (Rubin 1985)

- 「逆確率重み付け」と訳される
- 実際に処置を受けた ( $T_i = 1$ ) にもかかわらず、処置を受ける傾向が小さい場合は分析において大きい重み
  - 傾向スコアを重み変数として用いる。
  - $e_i$  が 1 または 0 に近い場合、一部のケースに非常に大きい重みを付けることになるため、注意が必要

$$w_i = T_i \frac{1}{e_i} + (1 - T_i) \frac{1}{1 - e_i}.$$

$e_i$ :  $i$  の傾向スコア、 $T_i$ :  $i$  の処置有無  $\in \{0, 1\}$

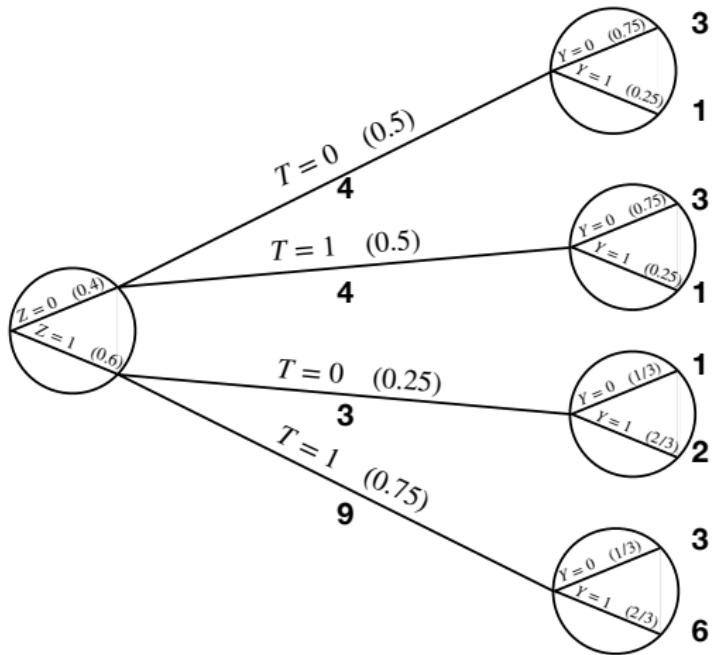
# IPW の考え方

ID	Z	T	$Y_0$	$Y_1$
1	0	0	0	1
2	0	0	1	0
3	0	0	0	0
4	0	0	0	0
5	0	1	0	0
6	0	1	1	0
7	0	1	0	0
8	0	1	0	1
9	1	0	1	1
10	1	0	1	0
11	1	0	0	1
12	1	1	1	1
13	1	1	1	1
14	1	1	0	1
15	1	1	0	1
16	1	1	0	1
17	1	1	1	1
18	1	1	1	0
19	1	1	1	0
20	1	1	1	0

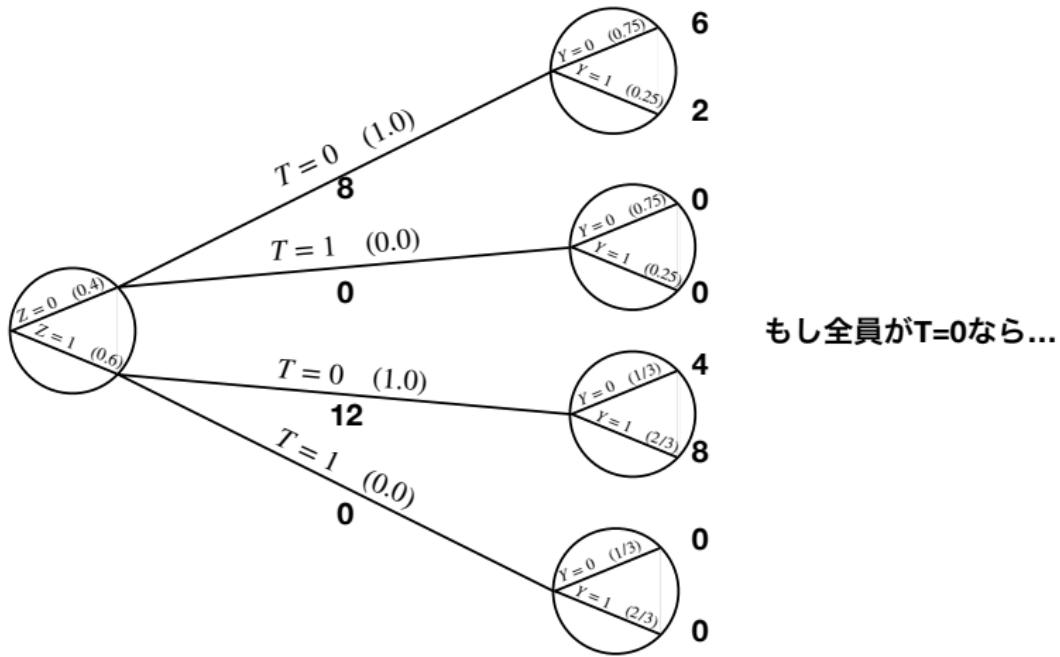
条件付き独立の例のデータ

- $Pr(Z = 0) = 0.4$ 、 $Pr(Z = 1) = 0.6$
- $Z = 0$  の場合、 $Pr(T = 0) = 0.5$ 、 $Pr(T = 1) = 0.5$
- $Z = 1$  の場合、 $Pr(T = 0) = 0.25$ 、 $Pr(T = 1) = 0.75$
- 観察されたデータからは約 0.1 の処置効果が推定されるが、本当の処置効果は 0

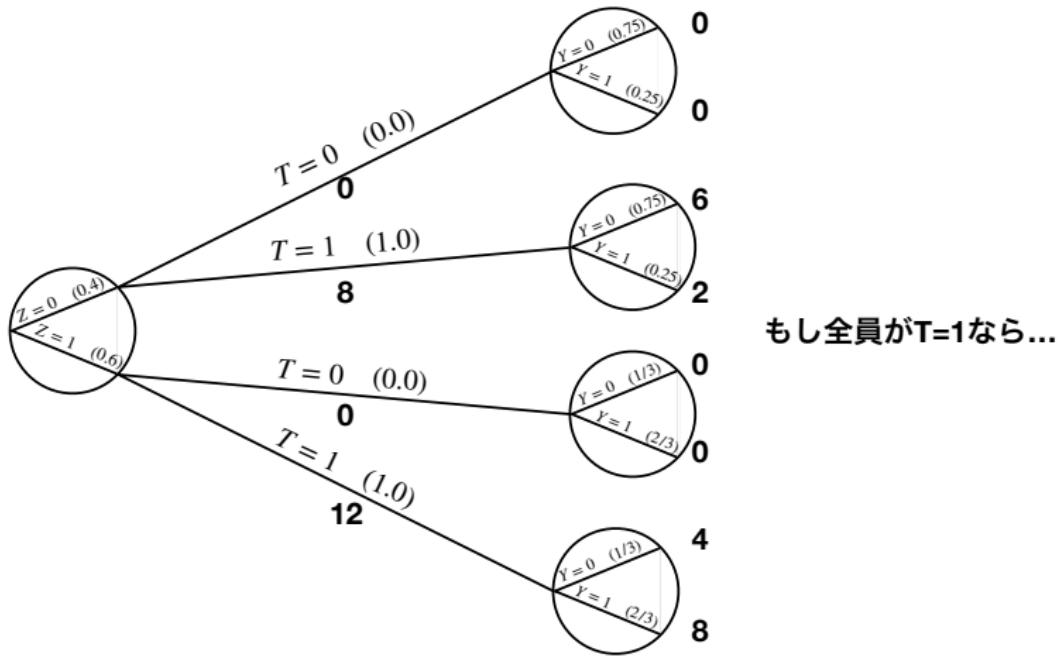
# IPW の考え方



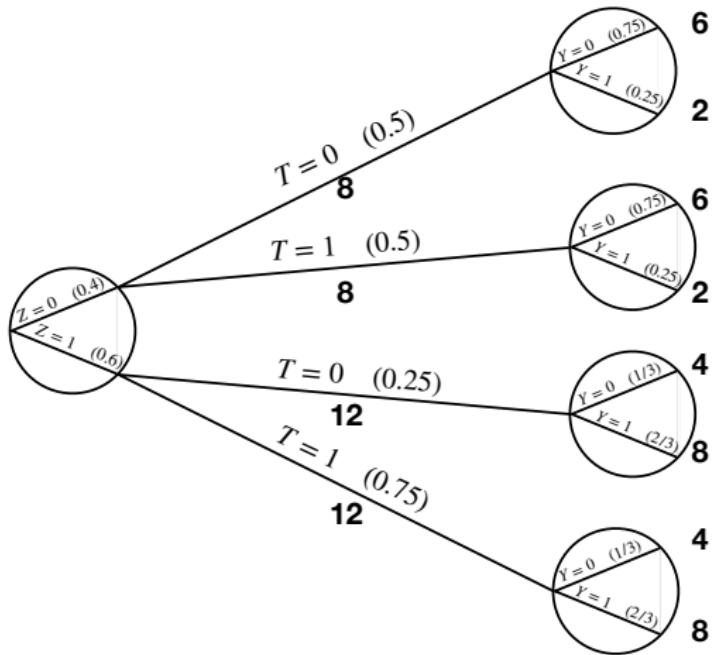
# IPW の考え方



# IPW の考え方



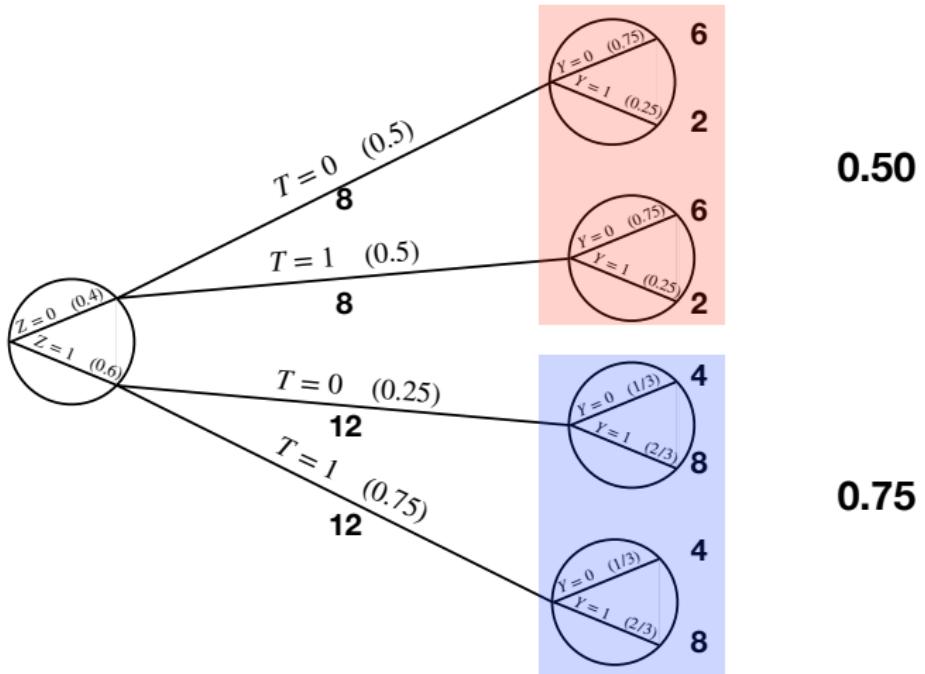
# IPW の考え方



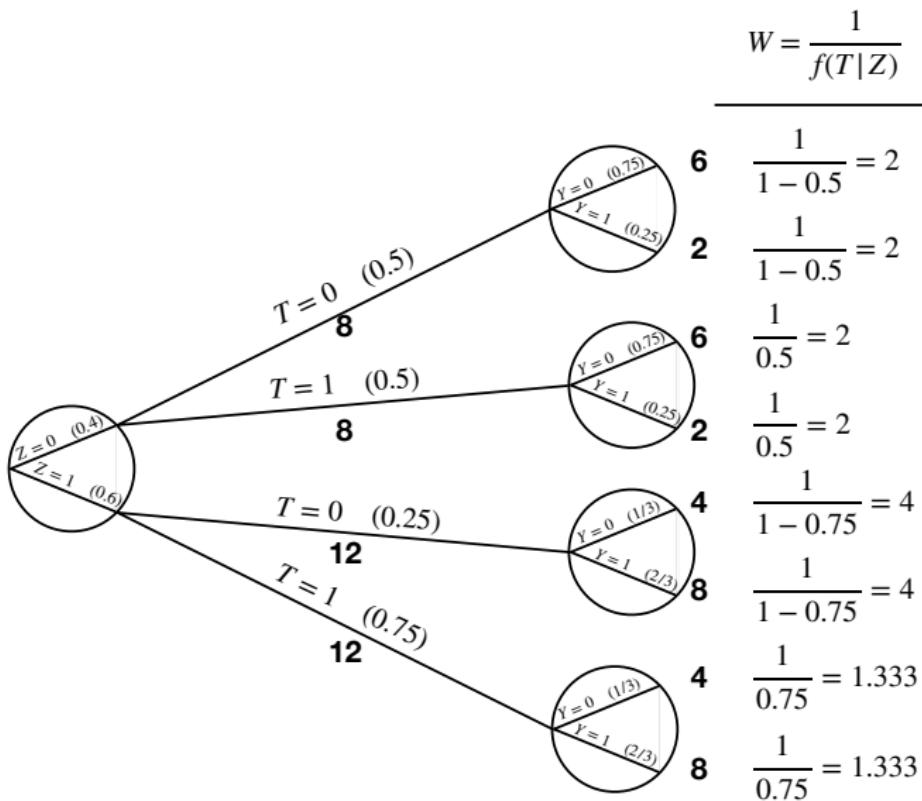
疑似母集団  
(pseudo-population)

# IPW の考え方

処置を受ける確率



# IPW の考え方



# 他の考え方

ID	Z	T	$Y_0$	$Y_1$	e	W
1	0	0	0	?	0.50	2.00
2	0	0	1	?	0.50	2.00
3	0	0	0	?	0.50	2.00
4	0	0	0	?	0.50	2.00
5	0	1	?	0	0.50	2.00
6	0	1	?	0	0.50	2.00
7	0	1	?	0	0.50	2.00
8	0	1	?	1	0.50	2.00
9	1	0	1	?	0.75	4.00
10	1	0	1	?	0.75	4.00
11	1	0	0	?	0.75	4.00
13	1	1	?	1	0.75	1.33
12	1	1	?	1	0.75	1.33
14	1	1	?	1	0.75	1.33
15	1	1	?	1	0.75	1.33
16	1	1	?	1	0.75	1.33
17	1	1	?	1	0.75	1.33
18	1	1	?	0	0.75	1.33
19	1	1	?	0	0.75	1.33
20	1	1	?	0	0.75	1.33

もし、全ケースが統制群なら？

- $Z = 0$  の統制群は 4 ケース (1~4)
- $Z = 0$  は全部で 8 ケース (2 倍)  
⇒ 1~4 ケースの重みを 2 倍に
- $Z = 1$  の統制群は 3 ケース (9~11)
- $Z = 1$  は全部で 12 ケース (4 倍)  
⇒ 9~11 ケースの重みを 4 倍に
- 全ケースが処置群なら？

# 他の考え方

ID	Z	T	$Y_0$	$Y_1$	e	W
1	0	0	0	?	0.50	2.00
2	0	0	1	?	0.50	2.00
3	0	0	0	?	0.50	2.00
4	0	0	0	?	0.50	2.00
5	0	1	?	0	0.50	2.00
6	0	1	?	0	0.50	2.00
7	0	1	?	0	0.50	2.00
8	0	1	?	1	0.50	2.00
9	1	0	1	?	0.75	4.00
10	1	0	1	?	0.75	4.00
11	1	0	0	?	0.75	4.00
13	1	1	?	1	0.75	1.33
12	1	1	?	1	0.75	1.33
14	1	1	?	1	0.75	1.33
15	1	1	?	1	0.75	1.33
16	1	1	?	1	0.75	1.33
17	1	1	?	1	0.75	1.33
18	1	1	?	0	0.75	1.33
19	1	1	?	0	0.75	1.33
20	1	1	?	0	0.75	1.33

- 統制群における W の和: 20
  - 処置群における W の和: 20
- ⇒ 各群における W の和はサンプルサイズと一致する
- ⇒ 全サンプルが統制/処置群の場合の結果変数の期待値を計算 (加重平均)

# 他の考え方

ID	Z	T	$Y_0$	$Y_1$	e	W
1	0	0	0	?	0.50	2.00
2	0	0	1	?	0.50	2.00
3	0	0	0	?	0.50	2.00
4	0	0	0	?	0.50	2.00
5	0	1	?	0	0.50	2.00
6	0	1	?	0	0.50	2.00
7	0	1	?	0	0.50	2.00
8	0	1	?	1	0.50	2.00
9	1	0	1	?	0.75	4.00
10	1	0	1	?	0.75	4.00
11	1	0	0	?	0.75	4.00
13	1	1	?	1	0.75	1.33
12	1	1	?	1	0.75	1.33
14	1	1	?	1	0.75	1.33
15	1	1	?	1	0.75	1.33
16	1	1	?	1	0.75	1.33
17	1	1	?	1	0.75	1.33
18	1	1	?	0	0.75	1.33
19	1	1	?	0	0.75	1.33
20	1	1	?	0	0.75	1.33

- 処置群の加重平均

- $0 \cdot 2 + 1 \cdot 2 + 0 \cdot 2 + 0 \cdot 2 + \dots + 0 \cdot 4$
- $\mathbb{E}^w[Y_0] = 10$

- 統制群の加重平均

- $0 \cdot 2 + 0 \cdot 2 + 0 \cdot 2 + 1 \cdot 2 + \dots + 0 \cdot 1.33$
- $\mathbb{E}^w[Y_1] = 10$

- $\mathbb{E}^w[Y_1] - \mathbb{E}^w[Y_0] = 0$

# 共変量の選択

# 共変量の選択

共変量選択の基準は (星野 2009; Imbens and Rubin 2015 など)

- ① 処置変数と結果変数、両方と連関があること
  - ▶ OVB と関係
- ② 処置前変数と処置後変数の区別
  - ▶ 処置変数に時間的に先行しているか否か
- ③ 処置前変数 (pre treatment variable) は必ず投入する
- ④ 処置後変数 (post treatment variable) は目的による
  - ▶ というものの、基本的に投入しない
  - ▶ 応答変数よりも時間的に後なら絶対に投入しない

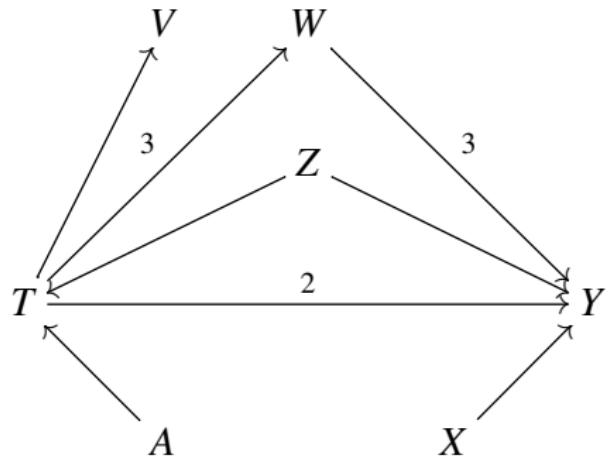
# 共変量の選択

VanderWeele (2019) の modified disjunctive cause criterion

- ① 処置変数と応答変数どちらかの原因となる変数
- ② 処置変数と応答変数両方の原因となる変数
- ③ 操作変数は共変量として投入しない
- ④ 上記の基準を満たさない場合でも、観察されていない共変量の代理変数は ○
  - ▶ しかし、慎重に選択しないとバイアスが拡大
  - ▶ (2) の該当する変数の代理変数が望ましい

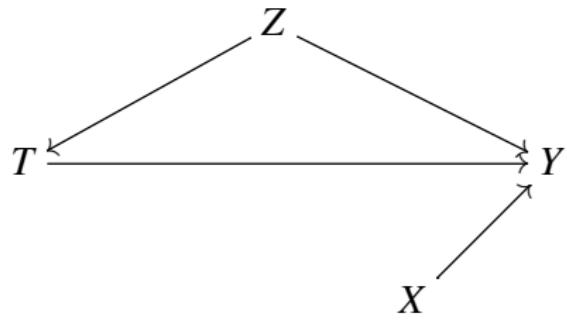
詳細は <https://www.slideshare.net/tintstyle/ss-141543274>

# ダイアグラムを使った例



- $T \rightarrow Y$  の効果は 11
- $Z$ :  $T$  と  $Y$  の原因  $\leftarrow$  投入
- $A$ : 操作変数  $\leftarrow$  除外
- $X$ :  $Y$  の原因  $\leftarrow$  投入
- $V$ :  $T$  の結果  $\leftarrow$  除外
- $W$  は ... ?

# ダイアグラムを使った例



- $T \rightarrow Y$ : 直接効果
- $T \rightarrow W \rightarrow Y$ : 間接効果
  - $W$  は中間変数 (mediate variable)
- 因果推論では主に全効果 (total effect) に関心があるため  $W$  は投入しない
  - $T$  が変動したら  $W$  も必ず変わるため、 $T$  のみの効果はあまり意味 ×
- **結論:**  $Z, X$  のみ統制するだけで十分

実習

# 利用データ

## 定番の lalonde データセット

- 職業訓練の有無と所得
- `data("lalonde", package = "cobalt")` で読み込み
  - その前に cobalt パッケージを読み込む
- 変数の詳細
  - ① `treat`: 職業訓練の履修有無 ( $\leftarrow$  Treatment)
  - ② `age`: 年齢
  - ③ `educ`: 教育年数
  - ④ `race`: 人種 (白人、黒人、ヒスパニック)
  - ⑤ `married`: 結婚有無
  - ⑥ `nodegree`: 学位有無
  - ⑦ `re74`: 1974 年の所得
  - ⑧ `re75`: 1975 年の所得
  - ⑨ `re78`: 1978 年の所得 ( $\leftarrow$  Outcome)

# Matching 手法

Exact Matching は現実的に不可能であるため省略

- ① Mahalanobis Matching
- ② Coarsened Exact Matching
- ③ Propensity Score Matching
- ④ Inverse Probability Weighting

使用するパッケージ

- 1~3: MatchIt
- 4: WeightIt
- バランスチェック: BalanceR、cobalt

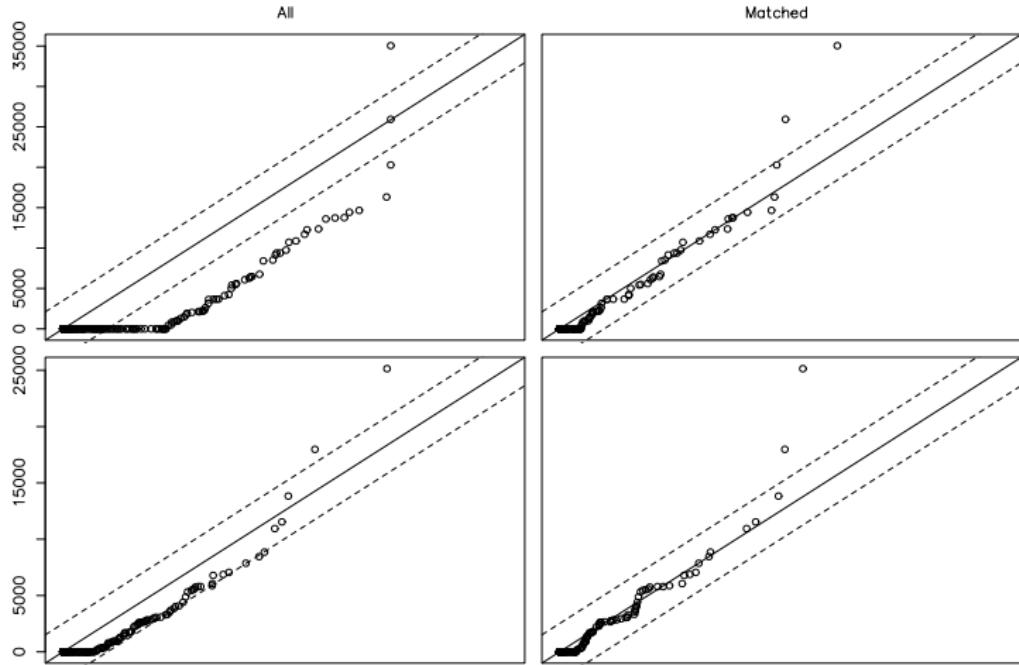
# バランスチェック

MatchIt パッケージが提供する方法

- RCT の講義で紹介したように、標準化差分を利用
  - MatchIt の `summary(..., standardize = TRUE)` を利用
  - cobalt、BalanceR パッケージ
- QQ Plot
  - 統制—処置群の共変量の分布に違いがあるか否か
- Histogram
  - マッチング前後ごとに共変量のヒストグラム
- Jitter Plot
  - ヒストグラムと似ているが棒の高さより密度で判断

# Balance Check: Q-Q Plot

plot (MatchIt オブジェクト名, type = "QQ")



(上段は re74、下段は re75)

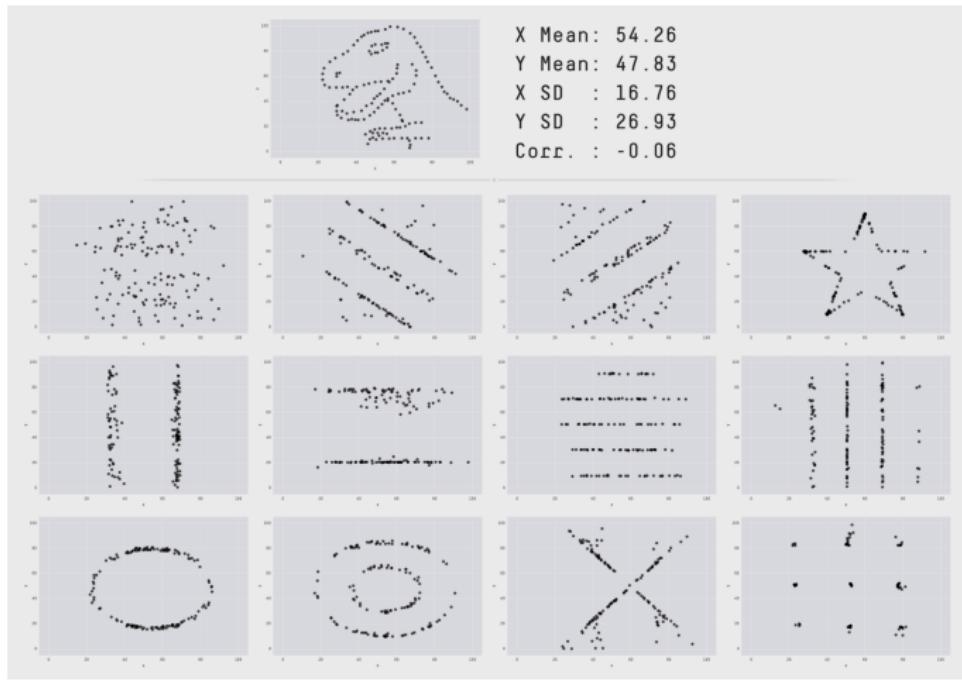
# Q–Q Plot とは

## Q–Q Plot

- Quantile–Quantile Plot
- (簡単にいうと)2つの変数を小さい値から大きい値の順に並び替え、散布図を作成し、45度線を引いたもの
- もし2つのデータが完全に同じ分布をしているのなら、全ての点は45度線上に位置
- 長所
  - 分布の形まで確認可能
  - 平均値と分散のみに依存するため、分布の形に鈍感
- 短所
  - 離散変数においては使いにくい

# 同じ平均値、同じ分散、同じ相関、異なる分布

二変数の平均、分散、相関係数が同じである 12 の散布図

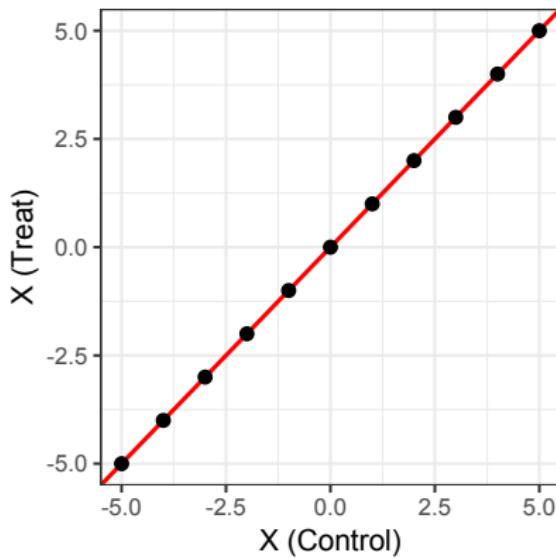


Source: <https://twitter.com/albertocairo/status/765167969139765250>

# Q-Q Plot の例

もし 2 つのデータが完全に同じ分布をしているのなら ...

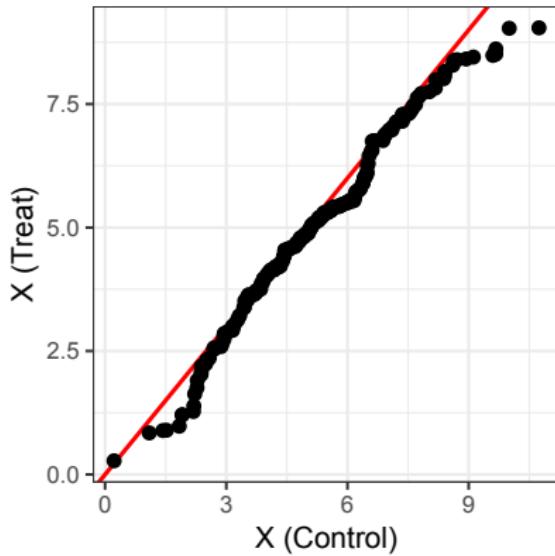
- $X_{\text{Control}}: -5, -4, -2, -1, 0, 1, 2, 3, 4, 5$
- $X_{\text{Treat}}: -5, -4, -2, -1, 0, 1, 2, 3, 4, 5$



# Q-Q Plot の例

もし 2 つのデータが**同じ分布**から生成されたなら ...

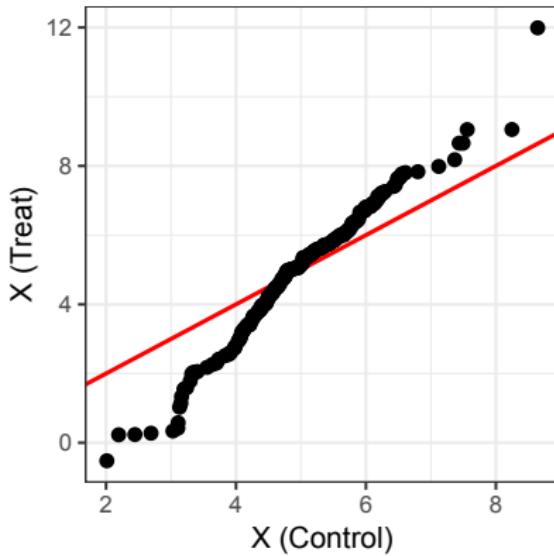
- $X_{\text{Control}} \sim \text{Normal}(5, 2)$
- $X_{\text{Treat}} \sim \text{Normal}(5, 2)$



# Q-Q Plot の例

もし 2 つのデータが**分散が異なる**分布から生成されたなら ...

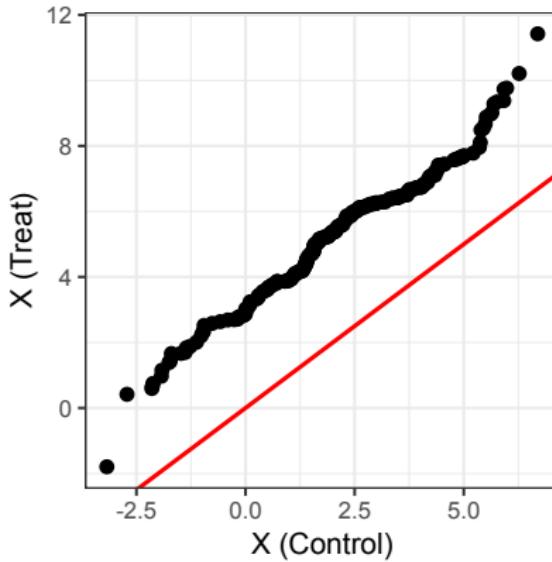
- $X_{\text{Control}} \sim \text{Normal}(5, 1)$
- $X_{\text{Treat}} \sim \text{Normal}(5, 2)$



# Q-Q Plot の例

もし 2 つのデータが**平均が異なる**分布から生成されたなら ...

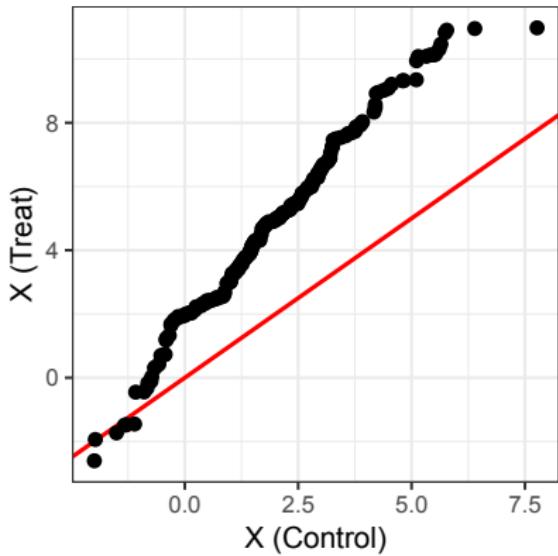
- $X_{\text{Control}} \sim \text{Normal}(2, 2)$
- $X_{\text{Treat}} \sim \text{Normal}(5, 2)$



# Q-Q Plot の例

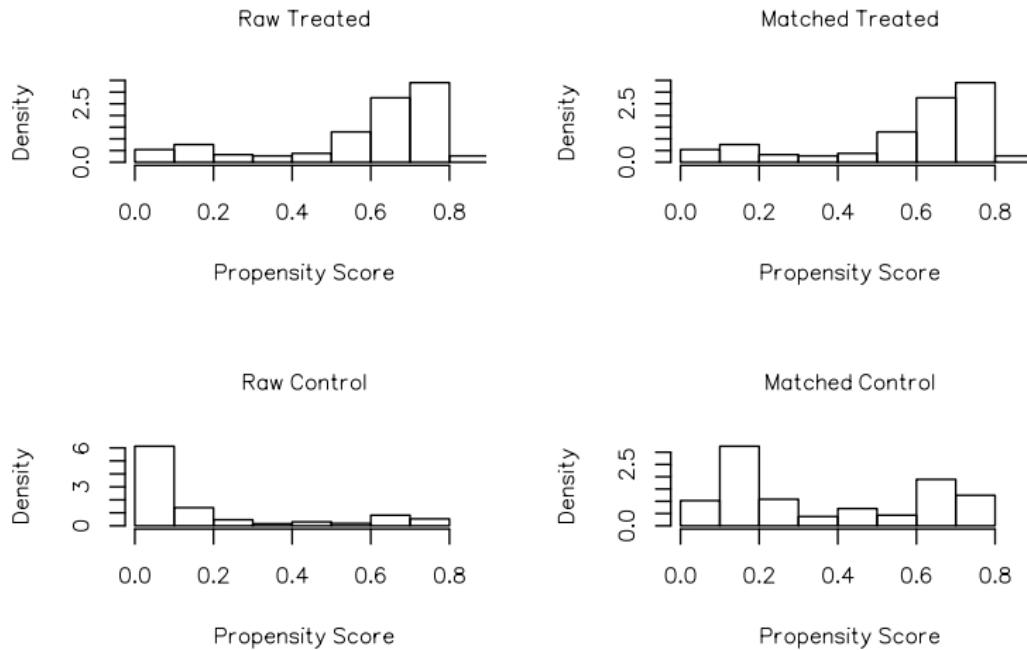
もし 2 つのデータが**平均と分散が異なる**分布から生成されたなら …

- $X_{\text{Control}} \sim \text{Normal}(2, 2)$
- $X_{\text{Treat}} \sim \text{Normal}(5, 3)$



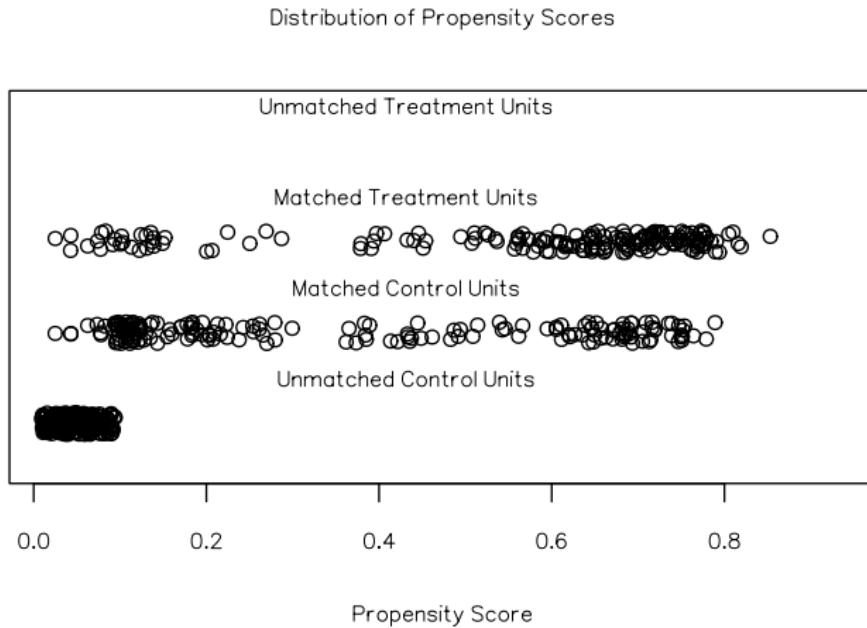
# Balance Check: Histogram

plot (MatchIt オブジェクト名, type = "hist")



# Balance Check: Jitter

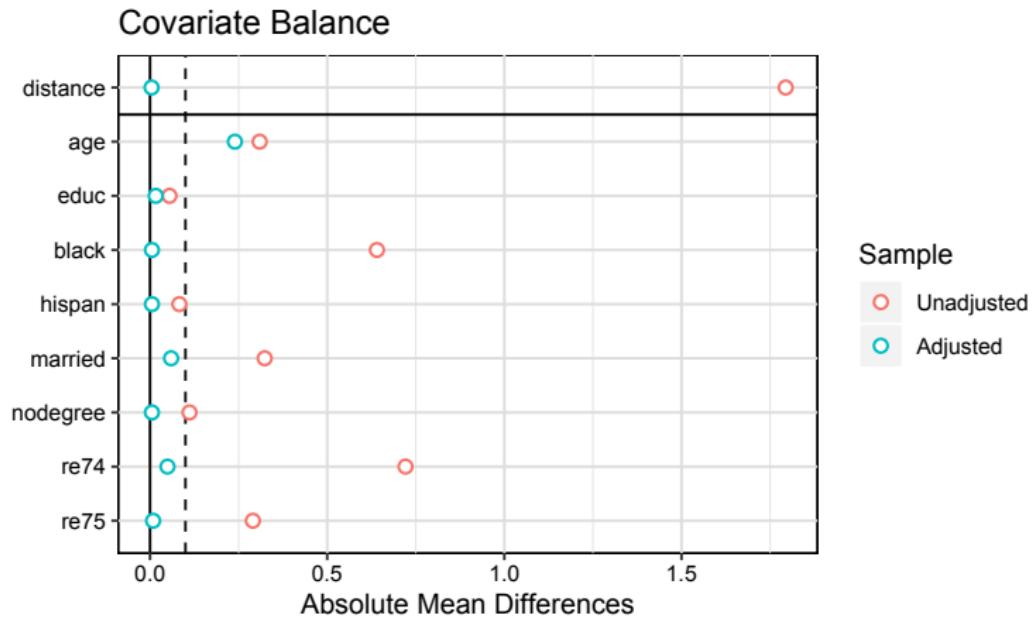
```
plot(MatchIt オブジェクト名, type = "jitter")
```



# Balance Check: Standardized Bias

標準化差分を利用

- cobalt パッケージが便利

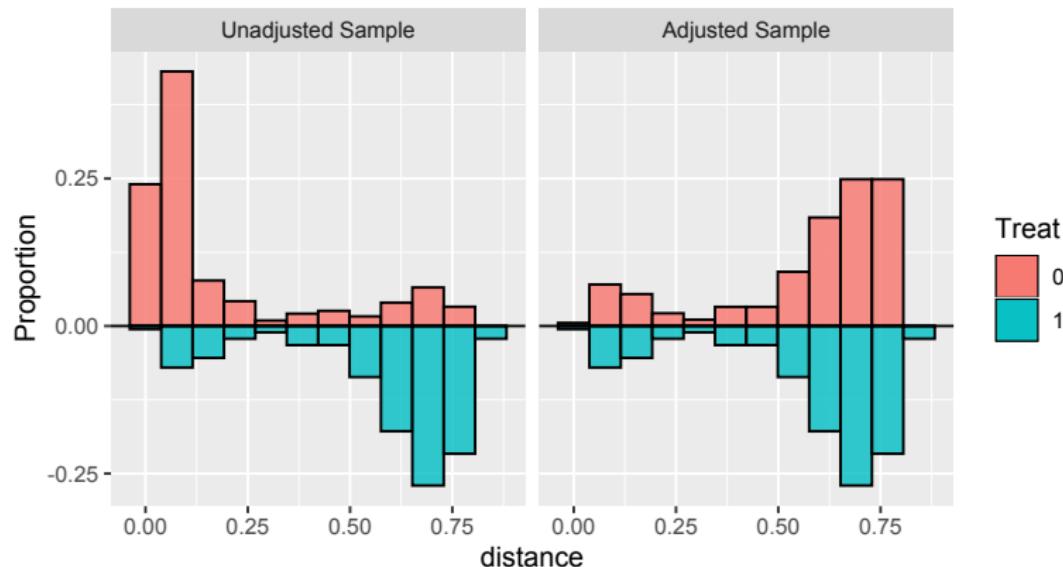


# Balance Check: Histogram

傾向スコアのヒストグラムを利用

- cobalt パッケージが便利

Distributional Balance for "distance"



# 真の因果効果は？

手法ごとに異なる処置効果

- 観察データでは真の因果効果が分からぬいため、どの推定値が正しいかは分からぬ  
    ▸ 最近傍マッチングはランダム要素があるため、分析の度に変化
- 様々な手法で推定値、または傾向が安定していることを示す

Method	Treatment Effect		
	Est.	95% LL	95% UL
Single Reg.	-635.03	-1923.00	652.94
Multiple Reg.	1548.23	16.96	3079.52
Mahalanobis	742.75	-724.32	2209.83
CEM	415.45	-1545.64	2376.54
PS Matching	1874.12	9.59	3738.66
IPW	224.67	-907.51	1356.86

# 真の因果効果は?

## 手法ごとに異なる処置効果

- 観察データでは真の因果効果が分からぬいため、どの推定値が正しいかは分からぬ  
▪ 最近傍マッチングはランダム要素があるため、分析の度に変化
- 様々な手法で推定値、または傾向が安定していることを示す

