

研究方法集中セミナー（因果推論）

第2回 マッチング

そん じえひよん

宋 財 況

関西大学総合情報学部

2021年8月7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
- Michael G. Findley, Kyosuke Kikuta, and Michael Denly. 2021. "External Validity," *Annual Review of Political Science*, 24:365-393.
- 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)

- Level1: **Association**
- $P(y|x)$
- Typical Activity: Seeing
- Typical Question:
 - What is?
 - How would seeing X change my belief in Y ?
- Examples:
 - What does a symptom tell me about a disease?
 - What does a survey tell us about the election results?

因果推論

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

- Level2: **Intervention**
- $P(y|do(x), z)$
- Typical Activity: Doing
- Typical Question:
 - What if?
 - What if I do X ?
- Examples:
 - What if I take aspirin, will my headache be cured?
 - What if we ban cigarettes?

因果推論

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

- Level3: **Counterfactual**
- $P(y_x|x', y')$
- Typical Activity: Imagining and Retrospection
- Typical Question:
 - Why?
 - What is X that caused Y ?
 - What if I had acted differently?
- Examples:
 - 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_(did) Y はどうなった_(would)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約265.333万円

ID	Income	Yaruki	Rishu	ID	Income	Yaruki	Rishu	ID	Income	Yaruki	Rishu
1	659	0	1	11	239	0	0	21	155	0	0
2	587	1	1	12	276	1	0	22	768	1	1
3	628	1	1	13	609	1	1	23	463	1	0
4	563	1	1	14	254	0	0	24	309	1	0
5	531	1	1	15	423	0	1	25	304	0	0
6	79	0	0	16	172	0	1	26	408	1	1
7	356	0	1	17	20	0	0	27	259	0	0
8	176	0	0	18	447	1	0	28	516	1	1
9	339	0	0	19	498	1	1	29	476	1	0
10	520	1	1	20	648	1	1	30	110	0	0

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(交絡要因)を揃える

- 「やる気」のない学生 (`Yaruki == 0`) 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

```
df1 %>%
  filter(Yaruki == 0) %>%
  group_by(Rishu) %>%
  summarise(
    Inc = mean(Income)
  ) %>%
  pull(Inc) %>%
  diff()

## [1] 209
```

ID	所得	やる気	履修	ID	所得	やる気	履修
1	659	0	1	6	79	0	0
7	356	0	1	8	176	0	0
15	423	0	1	9	339	0	0
16	172	0	1	11	239	0	0
				14	254	0	0
				17	20	0	0
				21	155	0	0
				25	304	0	0
				27	259	0	0
				30	110	0	0
Mean 402.5				Mean 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(交絡要因)を揃える

- 「やる気」のある学生 (`Yaruki == 1`) 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

```
df1 %>%
  filter(Yaruki == 1) %>%
  group_by(Rishu) %>%
  summarise(
    Inc = mean(Income)
  ) %>%
  pull(Inc) %>%
  diff()

## [1] 176.3455
```

ID	所得	やる気	履修	ID	所得	やる気	履修
2	587	1	1	12	276	1	0
3	628	1	1	18	447	1	0
4	563	1	1	23	463	1	0
5	531	1	1	24	309	1	0
10	520	1	1	29	476	1	0
13	609	1	1				
19	498	1	1				
20	648	1	1				
22	768	1	1				
26	408	1	1				
28	516	1	1				
Mean 570.5				Mean 394.2			

内生性への対処

やる気のない学生 ($Z_i = 0$)

履修 (T_i) 平均年収	
1	402.5
0	193.5

やる気のある学生 ($Z_i = 1$)

履修 (T_i) 平均年収	
0	570.5455
1	394.2000

講義履修の因果効果は約**191.6万円**

ID	N	やる気(Z_i)	平均年収 (Y_i)			ITE
			$Y_i(T_i = 1)$	$Y_i(T_i = 0)$	ITE	
1	14	0	402.5000	570.5455	209.0000	
2	16	1	193.5000	394.2000	176.3455	
加重平均						191.5843

マッチングの考え方

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、マッチングで因果効果の推定が可能
- **条件付き独立の仮定** (Conditional Independence Assumption; CIA)
 - $\{Y_i(T_i = 1), Y_i(T_i = 0)\} \perp T_i \mid X_i$
 - 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が満たされたためには、割当メカニズム上のすべての交絡要因が必要

条件付き独立の仮定とは

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) 0.429	(B) 0.429
T = 1	(C) 0.538	(D) 0.538

- 処置効果は $0.538 - 0.429 = 0.109$
- もし、統制群と処置群が同質なら
- A = C、そしてB = Dのはず
- 処置群がもし統制群になんて、今の統制群と同じ
- ⇒ 交換可能性が成立せず



条件付き独立の仮定とは

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 で条件づけた場合 ($Z = 0$)

	Y_0	Y_1
$T = 0$	(A) 0.250	(B) 0.250
$T = 1$	(C) 0.250	(D) 0.250

- 処置効果は $0.250 - 0.250 = 0.000$
- もし、統制群と処置群が同質なら
- $A = C$ 、そして $B = D$ が成立
- \Rightarrow 交換可能性が成立



条件付き独立の仮定とは

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 で条件づけた場合 ($Z = 0$)

	Y_0	Y_1
$T = 0$	(A) 0.667	(B) 0.667
$T = 1$	(C) 0.667	(D) 0.667

- 処置効果は $0.667 - 0.667 = 0.000$
- もし、統制群と処置群が同質なら
- $A = C$ 、そして $B = D$ が成立
- \Rightarrow 交換可能性が成立



条件付き独立の仮定とは

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するということは

- 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単位変化する時の応答変数の変化量
- 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 ○

- 手計算マッチング: 191.5843

```
library(modelsummary)
Fit1 <- lm(Income ~ Rishu, data = df1)
Fit2 <- lm(Income ~ Rishu + Yaruki,
           data = df1)
modelsummary(list("単回帰分析" = Fit1,
                  "重回帰分析" = Fit2))
```

	単回帰分析	重回帰分析
(Intercept)	260.400 (36.459)	198.595 (32.737)
Rishu	265.333 (51.561)	191.167 (44.915)
Yaruki		185.415 (45.015)
Num.Obs.	30	30
R2	0.486	0.684
R2 Adj.	0.468	0.661
AIC	386.1	373.5
BIC	390.3	379.1
Log.Lik.	-190.040	-182.726
F	26.481	29.273

重回帰分析との比較

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
 - 強い仮定を置いたマッチング
 - 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
 - マッチングはオーバーラップされないケースを分析から除外する
- 結論: **回帰分析より柔軟、拡張性がある**

ATE, ATT, ATC

3種類の因果効果

1. **ATE** (Average Treatment Effect): 平均処置効果
2. **ATT** (ATE for the Treated): 処置群における平均処置効果
 - 潜在結果: 処置群が処置を受けなかった場合の応答変数
1. **ATC** (ATE for the Control): 統制群における平均処置効果
 - 潜在結果: 統制群が処置を受けた場合の応答変数
 - 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 ↔ 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	$Y_i(T_i = 1)$	$Y_i(T_i = 0)$	ITE
1	0	659	193.5	465.5
2	1	587	394.2	192.8
3	1	628	394.2	233.8
4	1	563	394.2	168.8
5	1	531	394.2	136.8
7	0	356	193.5	162.5
...
20	1	648	394.2	253.8
22	1	768	394.2	373.8
26	1	408	394.2	13.8
28	1	516	394.2	121.8
平均				185.1

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。
- 統制群は $Y_i(T_i = 0)$ 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	$Y_i(T_i = 1)$	$Y_i(T_i = 0)$	ITE
6	0	402.5	79	323.5
8	0	402.5	176	226.5
9	0	402.5	339	63.5
11	0	402.5	239	163.5
12	1	570.5	276	294.5
14	0	402.5	254	148.5
...
25	0	402.5	304	98.5
27	0	402.5	259	143.5
29	1	570.5	476	94.5
30	0	402.5	110	292.5
平均				198.1

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 → 単純平均でOK
 - $\frac{1}{2}(185.1 + 198.1) = 191.6$
 - 手計算マッチングとと同じ結果

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

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(交絡要因)が完全に一致するケース同士の比較
- 共変量が少數、かつ、名目or順序変数の場合、使用可
- 共変量が多数、または連續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。
- 共変量が連續変数、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
 - 共変量を座標（超）平面に位置づけた場合、最も近いケースをマッチング

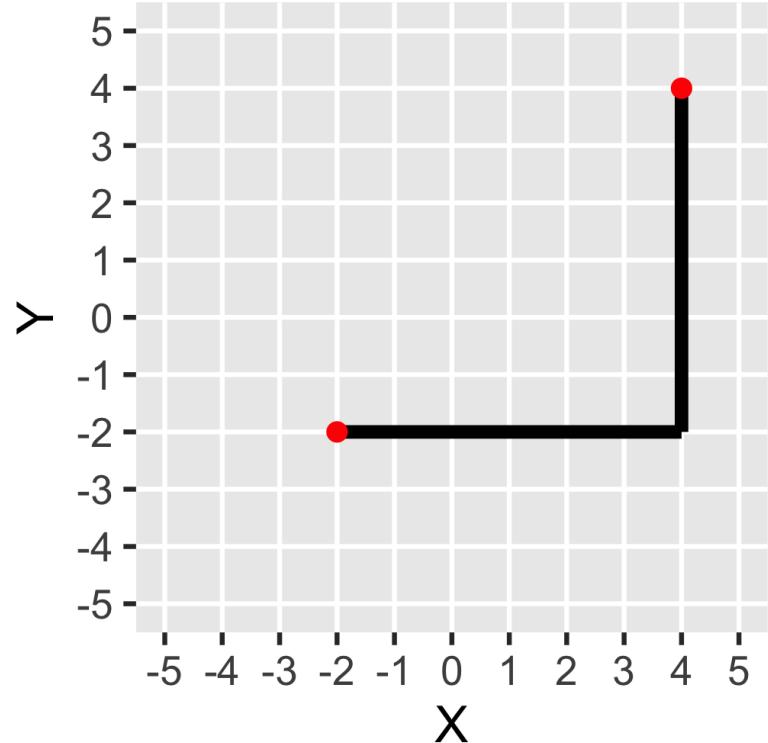
「近さ」の基準

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

Nearest-neighbor Matching

Manhattan Distance (City-block Distance)

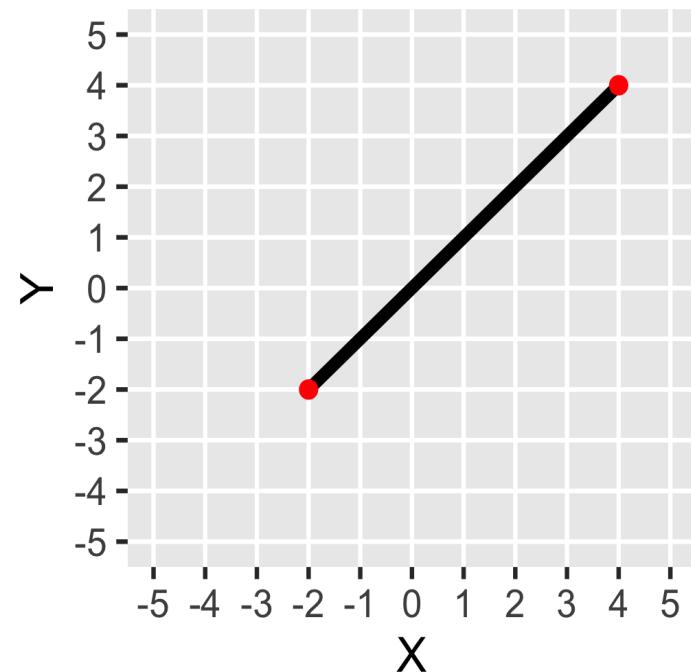
$$d(i, j) = |X_i - X_j| + |Y_i - Y_j| \text{ where } i \neq j.$$



Nearest-neighbor Matching

Standardized Euclidean Distance

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

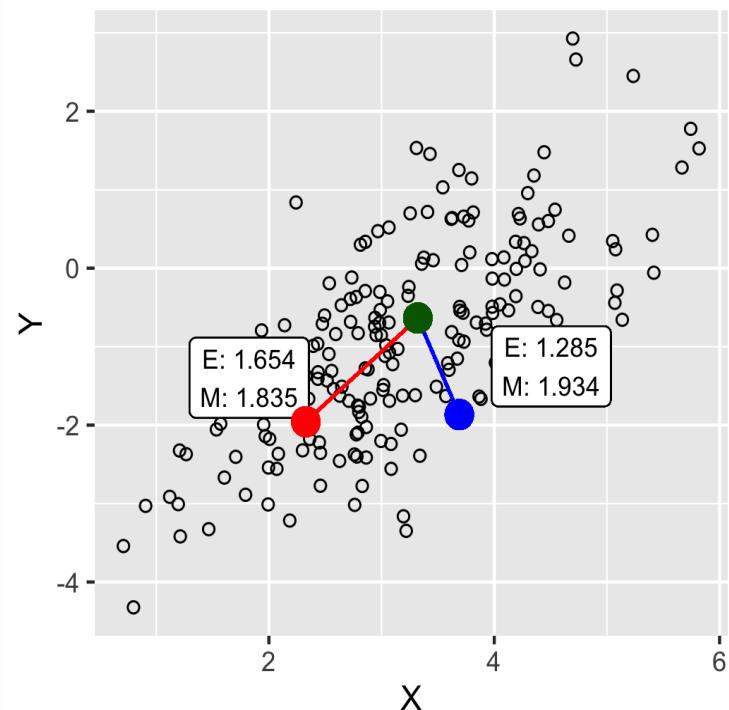


Nearest-neighbor Matching

Mahalanobis Distance

- 共変量間の相関が0の場合、Standardized Euclidean Distanceと同じ

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



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-最近傍マッチング

- 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

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

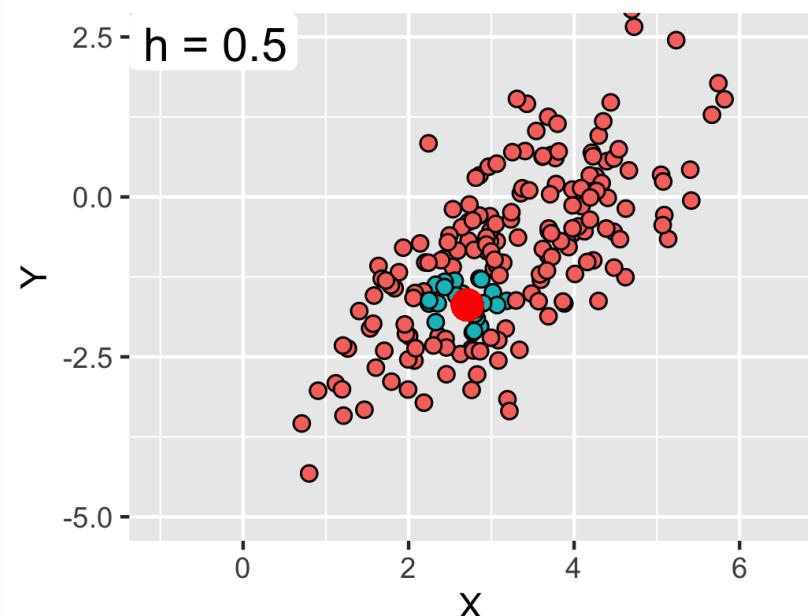
- 最適 k を決める理論的基準は無し
 - k を大きくすると、モデルの分散が小さくなる
 - ただし、モデルの分散が小さい = バイアスが拡大
 - Bias-variance trade-off**
 - k を変化させることによって結果がどのように変わるか観察

Caliper Matching

「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

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



Coarsened Exact Matching

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

- 定訳はなく、「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、すべての層に対して加重平均
- 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 (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	教育	処置	結果	ID	年齢	教育	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 (Day2_Data2.csv)

年齢は10歳刻み、学歴は大卒以上・未満に層化

- 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	教育	処置	結果	ID	年齢	教育	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 (Day2_Data2.csv)

層ごとにケースをマッチング (ペアが組めない層もあり)

年齢	教育	処置群			統制群		
		ID	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 (Day2_Data2.csv)

ペアが組めない層を除外

年齢	教育	処置群			統制群			
		ID	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	
	40代	19	1	9	2	0	3	
40代					16	0	2	
					22	0	9	
		23	1	6	8	0	4	
L	10	1	8	9	0	2		
	50代		15	1	9	13	0	4

CEMの例 (Day2_Data2.csv)

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	教育	処置群					統制群				
		ID	処置	結果	重み	ID	処置	結果	重み		
20代	H	11	1	5	1.000	1	0	6	2.200		
		14	1	5	1.000						
	L	12	1	6	1.000	17	0	2	1.100		
		21	1	8	1.000	20	0	4	1.100		
30代	H	3	1	7	1.000	4	0	5	1.100		
		18	1	4	1.000	24	0	8	1.100		
	H	19	1	9	1.000	2	0	3	0.367		

50代	L	10	1	8	1.000	9	0	2	1.100		
		15	1	9	1.000	13	0	4	1.100		

CEMの例 (Day2_Data2.csv)

重み付け回帰分析

- $W = \{2.200, 0.367, 1.000, 1.100, 0.000, 0.000, \dots, 1.100\}^\top$
 - 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
 - $\beta = (X^\top W X)^{-1} X^\top W Y$
- Rの場合、`lm(formula, data, weight = ...)`で推定可能
 - `{cem}` or `{MatchIt}`パッケージならもっと簡単
- $\widehat{\text{Outcome}} = 4.567 + 2.033 \cdot \text{Treat}$
 - 処置群における因果効果 (ATT) = 2.033

傾向スコア

傾向スコアとは

Propensity Score

- 簡単にいうと「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
 - 次元の問題 (dimension problem)
 - 恣意性
 - カテゴリ変数の扱い方
 - 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, ...)
 - 色々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ではオブジェクト名\$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

割り当てメカニズムを仮定

$$Pr(\text{処置}) \propto \beta_0 + \beta_1 \cdot \text{年齢} + \beta_2 \cdot \text{教育}$$

ID	年齢	教育	処置	結果	ID	年齢	教育	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

傾向スコアの算出

傾向スコアの算出

```
PS_Fit <- glm(処置 ~ 年齢 + 学歴, data = データ, family = binomial("logit"))
summary(PS_Fit)
```

Model 1	
(Intercept)	3.839
	(2.364)
Age	-0.059
	(0.039)
Edu	-0.420
	(0.304)
Num.Obs.	24
AIC	35.5
BIC	39.0
Log.Lik.	-14.748

傾向スコアの算出

傾向スコアの抽出

```
データ$PS <- PS_Fit$fitted.value
```

ID	年齢	教育	処置	結果	PS	ID	年齢	教育	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

傾向スコアの算出

ATT: 傾向スコアが最も近い統制群を割り当てる

- 一回マッチングされたケースを除外する vs. しない
- 傾向スコアが同じケースが複数ある場合の対処

処置群			統制群			
ID	結果	PS	ID	結果	PS	差分
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

差分の平均値 (ATT): 1.818

傾向スコアの算出

ATC: 傾向スコアが最も近い処置群を割り当てる

処置群			統制群			差分
ID	結果	PS	ID	結果	PS	
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

差分の平均値 (ATC): 3.923

傾向スコアの算出

ATE: ATTとATCの加重平均

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

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

```
## [1] 2.958208
```

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
 - Neural 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 ($\in \{0, 1\}$)

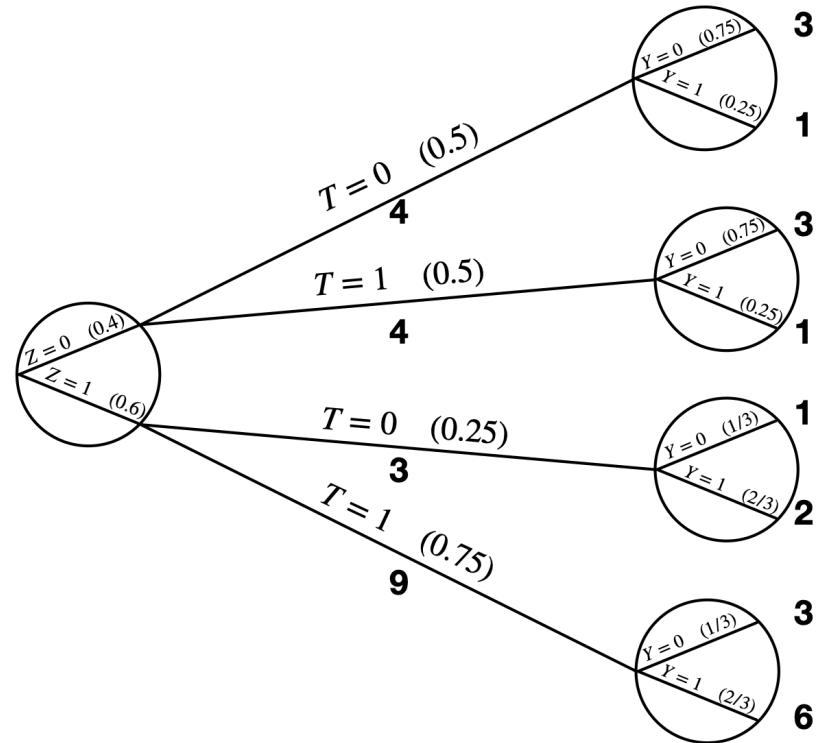
IPW の考え方

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

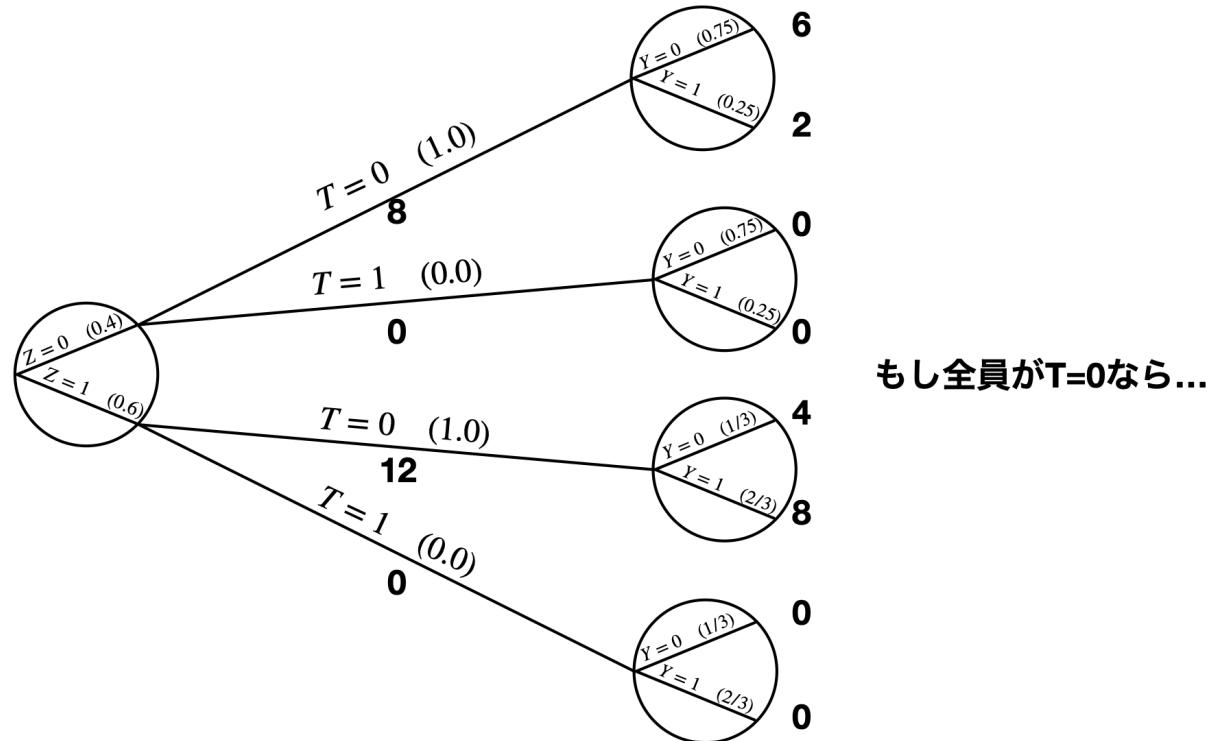
条件付き独立の例のデータ

- $Pr(Z = 0) = 0.4$ 、 $Pr(Z = 1) = 0.6$
- $Z = 0$ 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

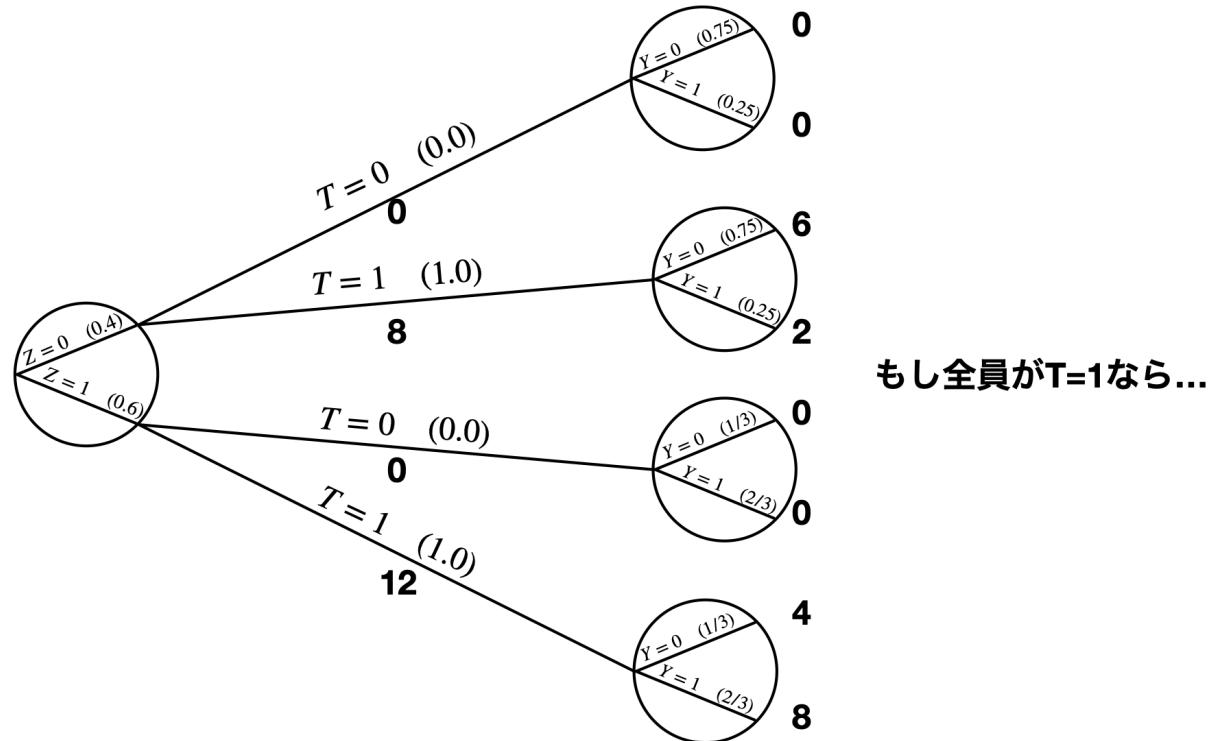
IPWの考え方



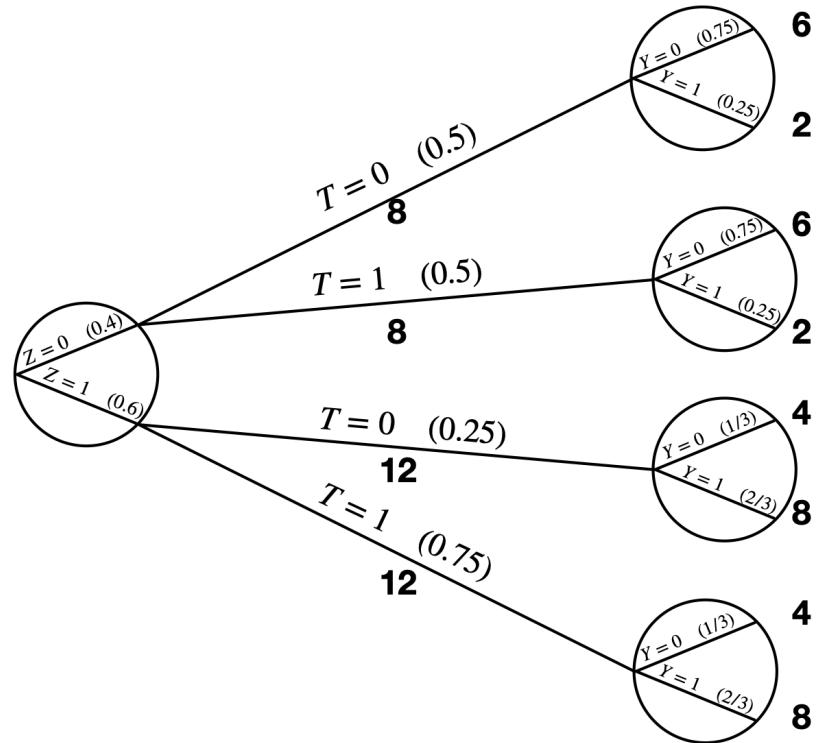
IPWの考え方



IPWの考え方

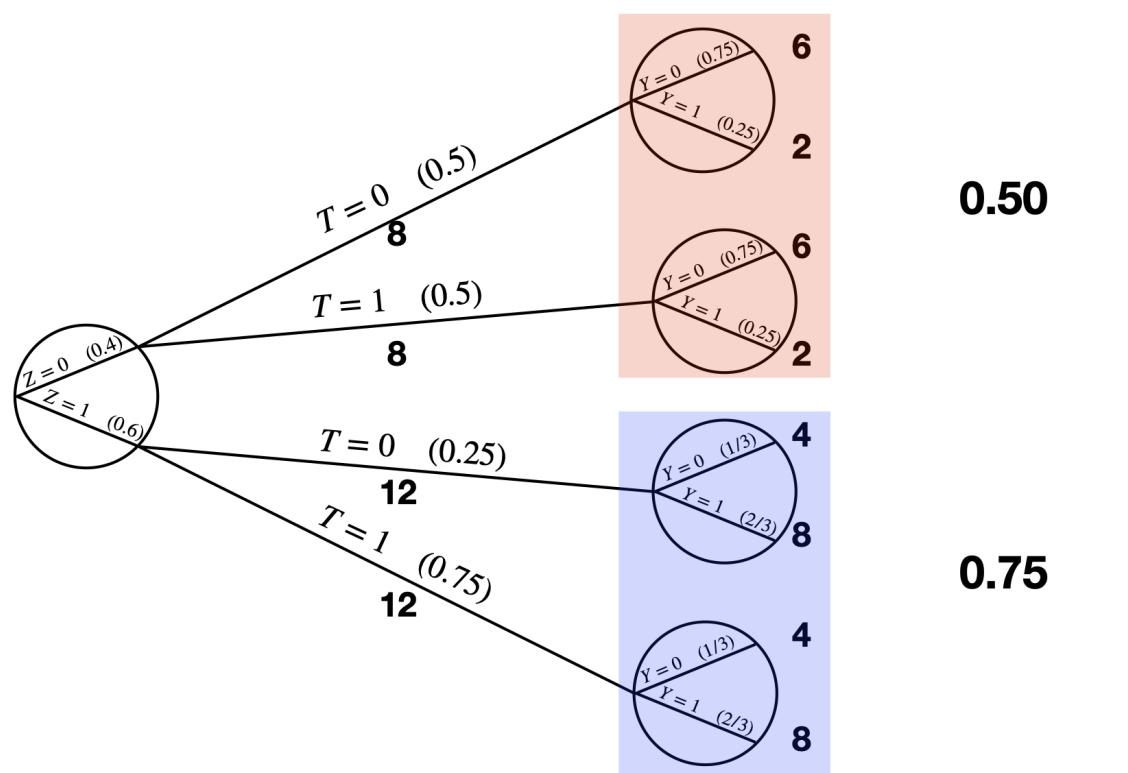


IPWの考え方

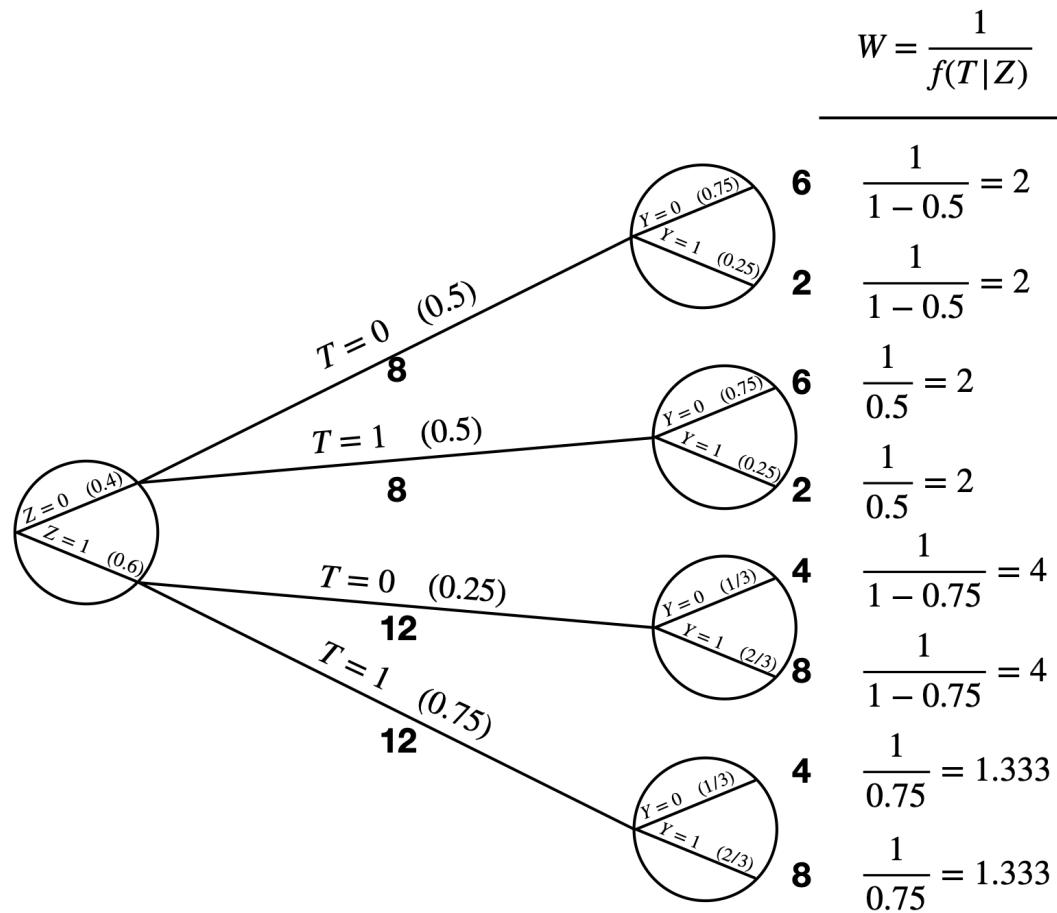


疑似母集団
(pseudo-population)

IPWの考え方



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	
6	0	1	?	0	0.50	
7	0	1	?	0	0.50	
8	0	1	?	1	0.50	
9	1	0	1	?	0.75	4.00
10	1	0	1	?	0.75	4.00
11	1	0	0	?	0.75	4.00
12	1	1	?	1	0.75	
13	1	1	?	1	0.75	
14	1	1	?	1	0.75	
15	1	1	?	1	0.75	
16	1	1	?	1	0.75	
17	1	1	?	1	0.75	
18	1	1	?	0	0.75	
19	1	1	?	0	0.75	
20	1	1	?	0	0.75	

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)
 - → ケース1~4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)
 - → ケース9~11の重みを4倍に

他の考え方

ID	Z	T	Y_0	Y_1	e	W
1	0	0	0	?	0.50	
2	0	0	1	?	0.50	
3	0	0	0	?	0.50	
4	0	0	0	?	0.50	
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	
10	1	0	1	?	0.75	
11	1	0	0	?	0.75	
12	1	1	?	1	0.75	1.33
13	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)
 - → ケース1~4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)
 - → ケース9~11の重みを4倍に

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
12	1	1	?	1	0.75	1.33
13	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
12	1	1	?	1	0.75	1.33
13	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
 - $0 \cdot 2 + 1 \cdot 2 + 0 \cdot 2 + 0 \cdot 2 + \cdots + 0 \cdot 4$
 - $\mathbb{E}^w[Y_0] = 10$
- 統制群の加重平均
 - $0 \cdot 2 + 0 \cdot 2 + 0 \cdot 2 + 1 \cdot 2 + \cdots + 0 \cdot 1.33$
 - $\mathbb{E}^w[Y_1] = 10$
- $\mathbb{E}^w[Y_1] - \mathbb{E}^w[Y_0] = 0$

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、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

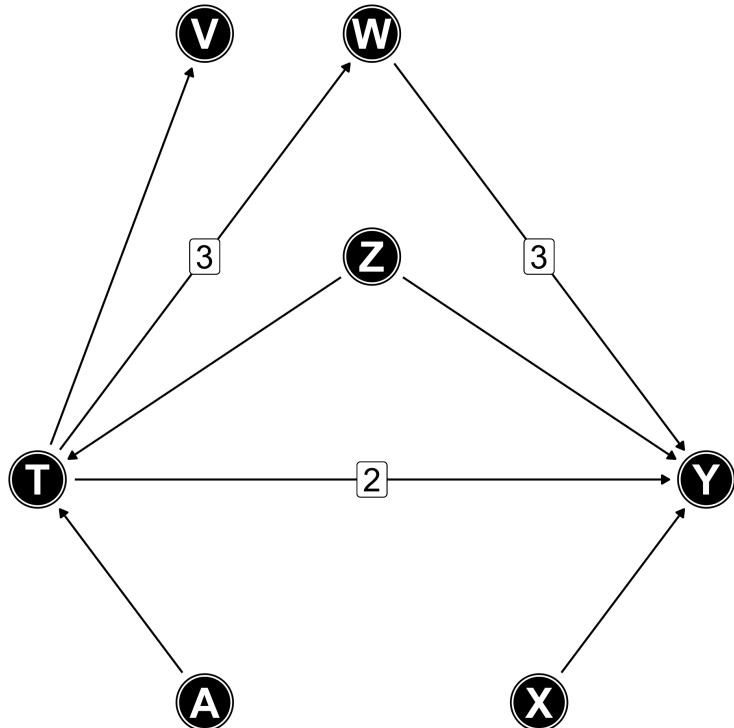
共変量の選択

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統制しても良い
 - 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

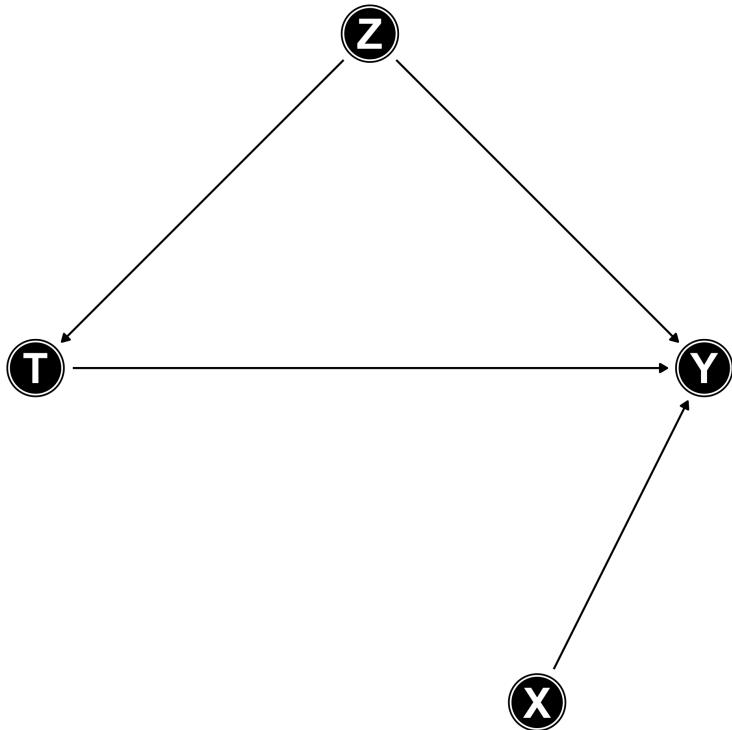
詳細は <https://www.slideshare.net/tintstyle/ss-141543274> を参照

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 ← 投入
- A: 操作変数 ← 除外
- X: Y の原因 ← 投入
- V:Tの結果 ← 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
 - 全効果: 直接効果 + 間接効果
 - 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なし
 - 直接効果のみ推定する場合、 W も統制
- 結論: Z と X のみ統制
 - 実は X は入れなくてもOK

ダイアグラムのツール

DAGitty – draw and analyze causal diagrams

- ウェーブページ or Rパッケージ({dagitty})
 - <http://www.dagitty.net/>

```
library(dagitty)

DAG1 <- dagitty(
  "dag {
    T -> Y
    T -> W -> Y
    T <- Z -> Y
    A -> T
    X -> Y
    T -> V
  }"
)

# Total Effect推定のための共変量
adjustmentSets(DAG1, exposure = "T", outcome = "Y",
                 effect = "total")

## { Z }

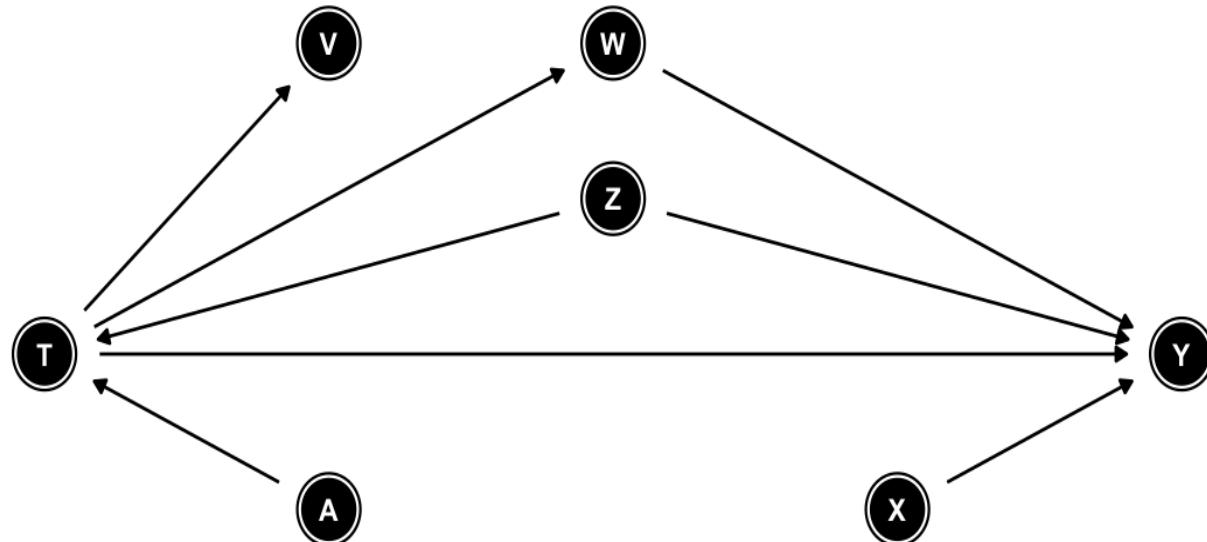
# Direct Effect推定のための共変量
adjustmentSets(DAG1, exposure = "T", outcome = "Y",
                 effect = "direct")

## { W, Z }
```

ダイアグラムのツール

{ggdag}を用いた可視化: 『私たちのR』 第20章も参照

```
library(ggdag)
coordinates(DAG1) <- list(
  x = c(T = 1, Y = 5, Z = 3, W = 3, A = 2, V = 2, X = 4),
  y = c(T = 2, Y = 2, Z = 3, W = 4, A = 1, V = 4, X = 1)
)
ggdag(DAG1, node_size = 12, stylized = TRUE) +
  theme_dag_blank()
```



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

変数名	説明
1 treat	職業訓練の履修有無 (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 (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

使用パッケージ

- 1～3: {MatchIt}
 - 事前に{cem}パッケージをインストールしておく
- 4: {WeightIt}
- バランスチェック: {cobalt}

バランスチェック

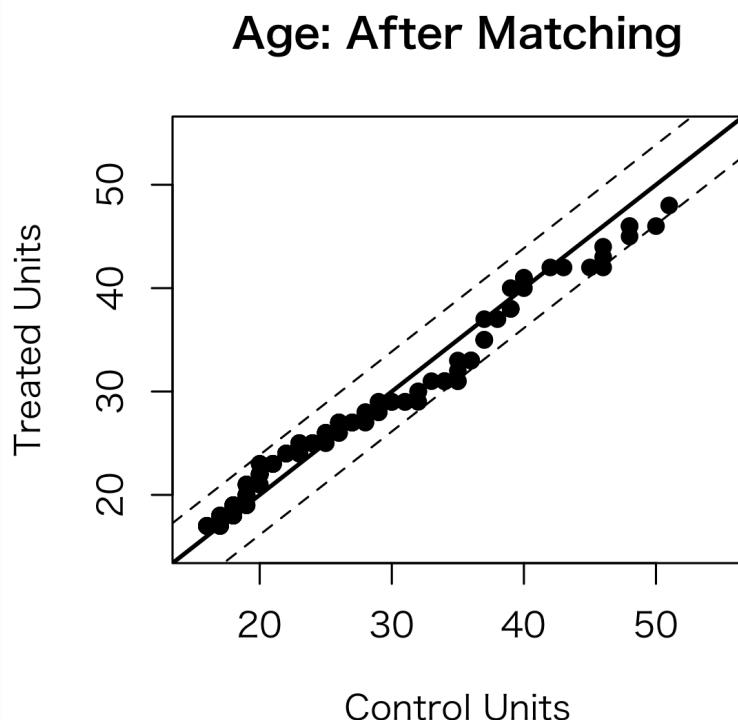
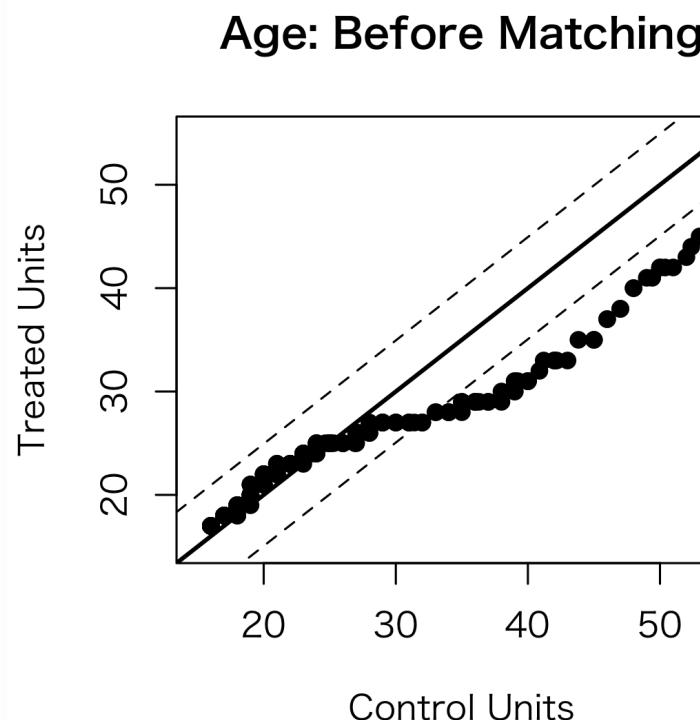
{MatchIt}パッケージが提供する方法

- 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}パッケージなど
- 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

QQ Plotの出し方

plot(マッチング後のオブジェクト名)で出力

- 点が45度線上に位置する場合、バランス



QQ Plotとは

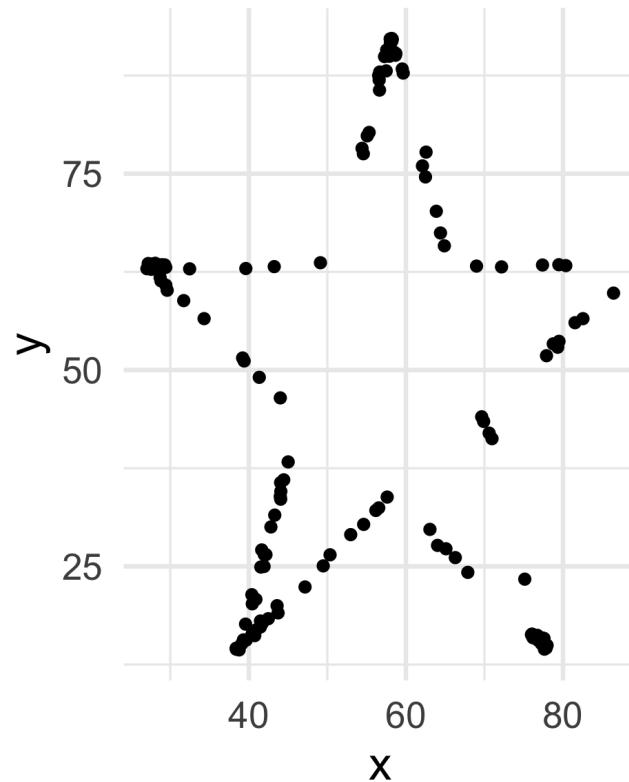
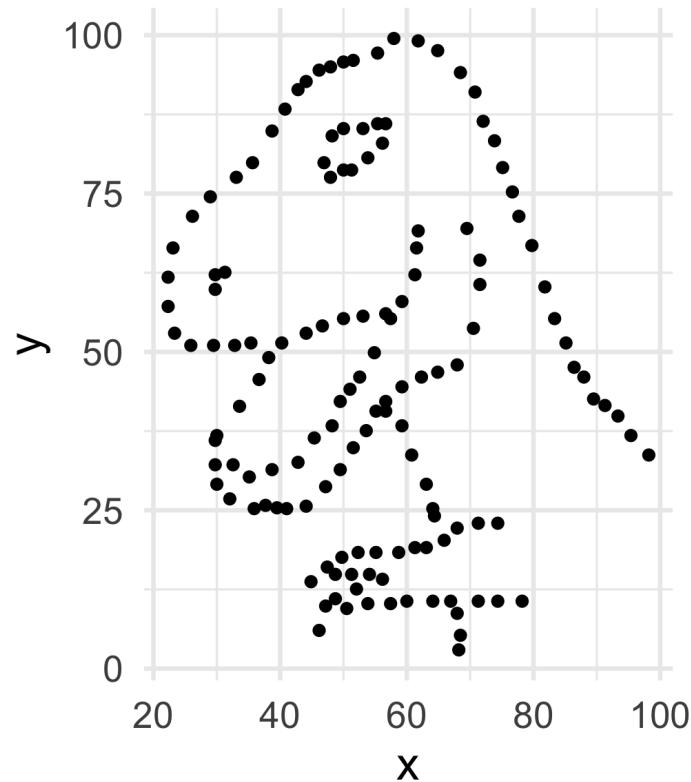
Quantile-Quantile Plot

- (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
 - 分布の形まで比較可能
 - 平均値と分散のみに依存するため、 サンプルサイズに鈍感
- 短所
 - 離散変数においては使いにくい

同じ平均値・分散・相関、異なる分布

Alberto CairoのDatasaurus dataset

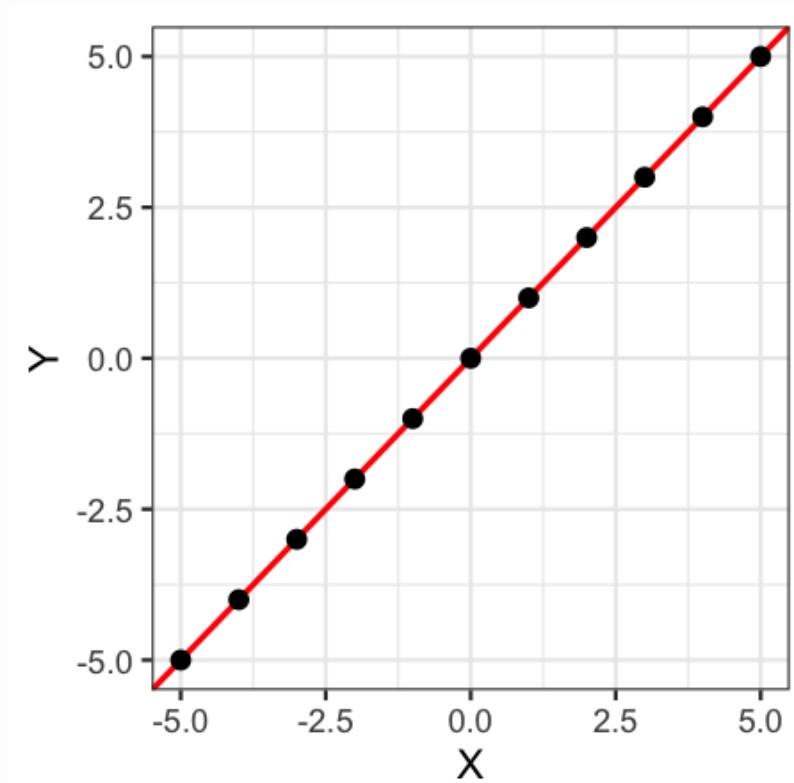
- 以下の2つの散布図のXとYは平均値、分散、相関係数が同じ
 - 記述統計量だけでなく、可視化も必要



QQ Plotの例 (1)

もし2つのデータが**完全に同じ**なら...

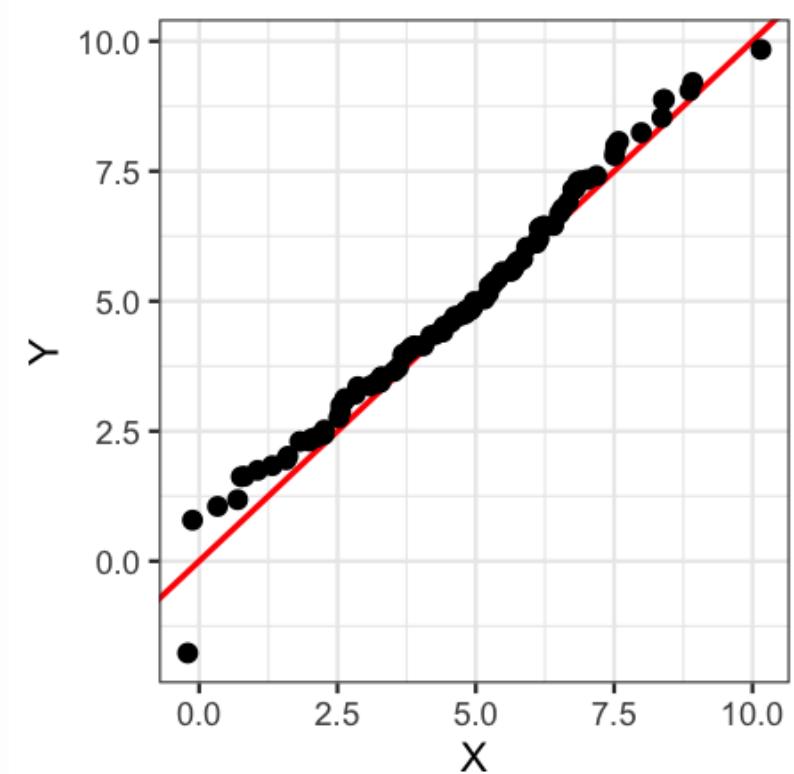
- X: -5, -4, -2, -1, 0, 1, 2, 3, 4, 5
- Y: -5, -4, -2, -1, 0, 1, 2, 3, 4, 5



QQ Plotの例 (2)

もし2つのデータが**同じ分布**から生成されたなら...

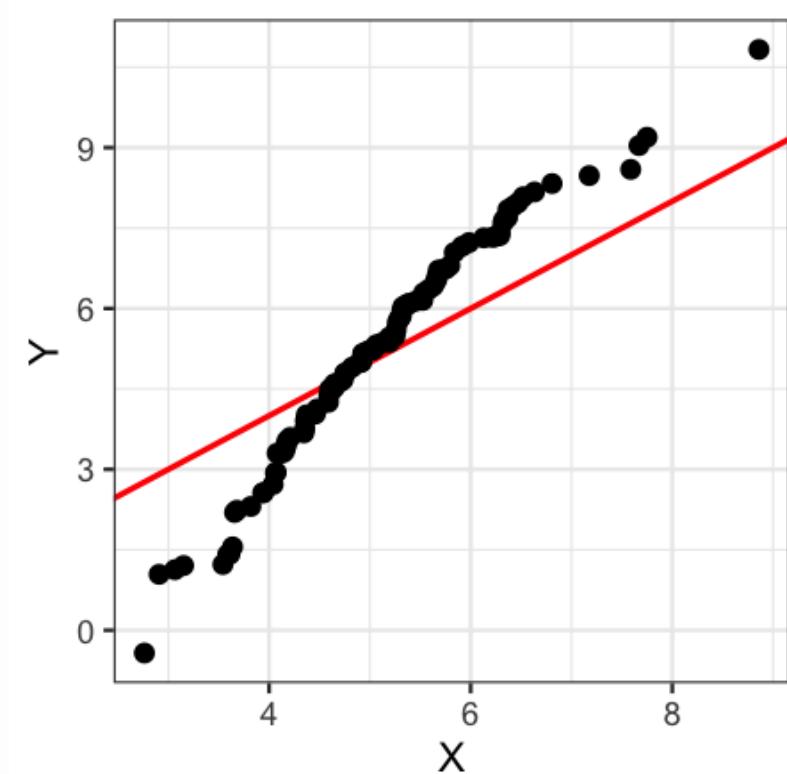
- $X \sim \text{Normal}(5, 2)$
- $Y \sim \text{Normal}(5, 2)$



QQ Plotの例 (3)

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...

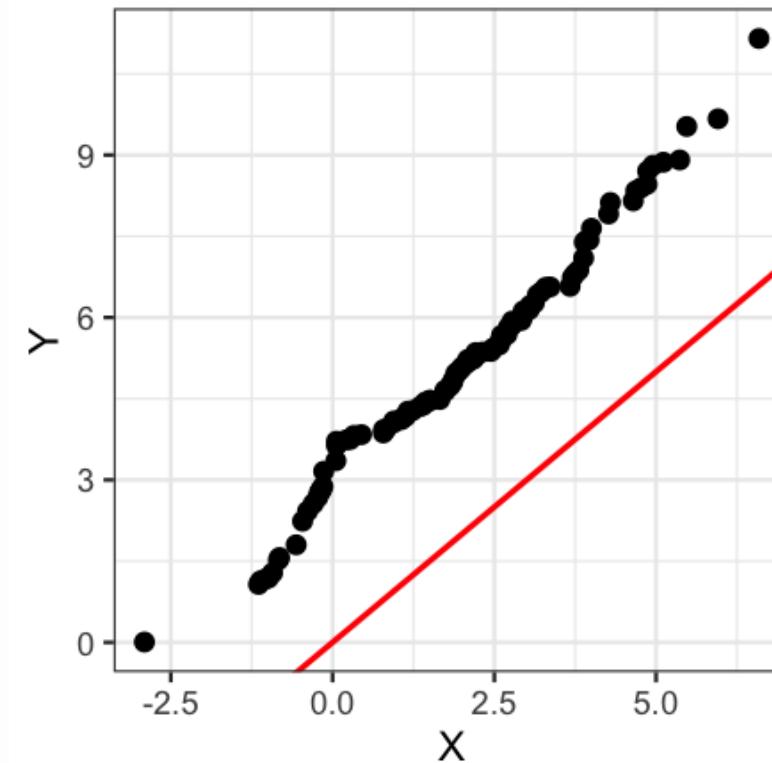
- $X \sim \text{Normal}(5, 1)$
- $Y \sim \text{Normal}(5, 2)$



QQ Plotの例 (4)

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

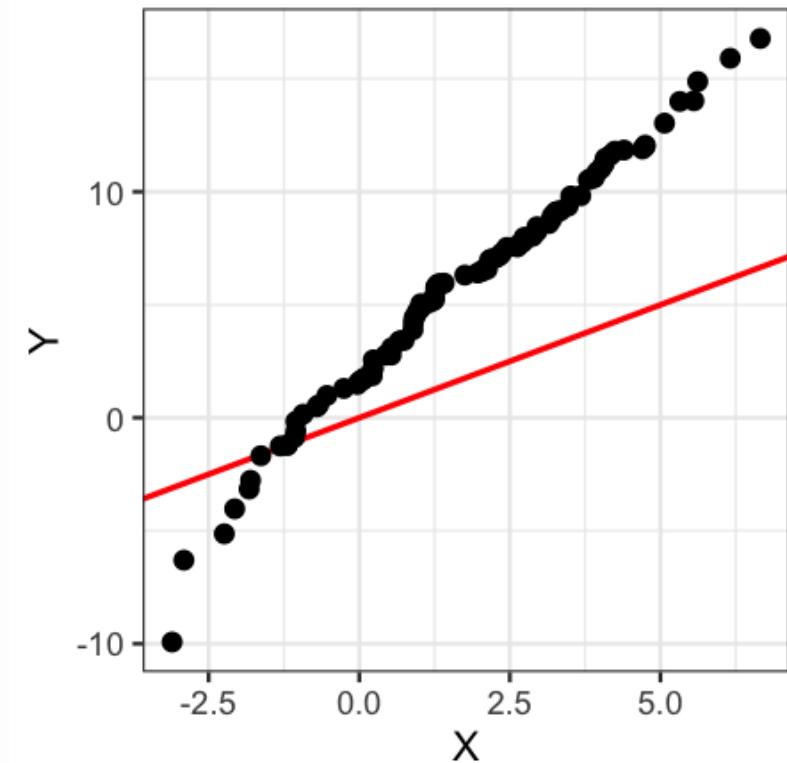
- $X \sim \text{Normal}(2, 2)$
- $Y \sim \text{Normal}(5, 2)$



QQ Plotの例 (5)

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...

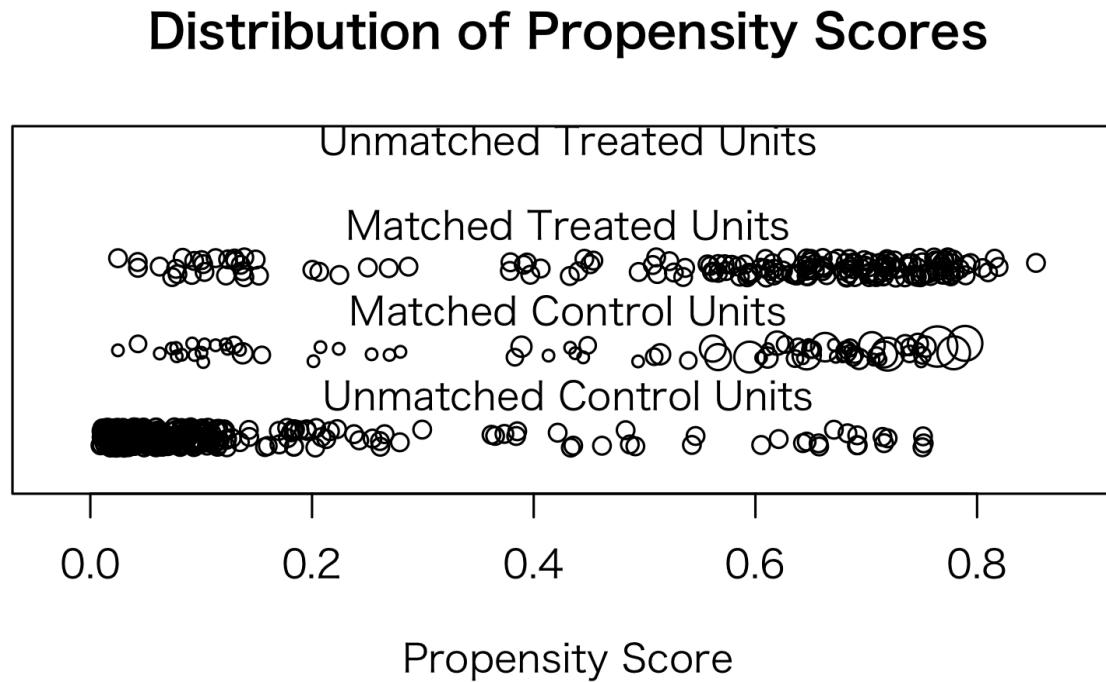
- $X \sim \text{Normal}(2, 2)$
- $Y \sim \text{Normal}(5, 5)$



Jitter Plot

傾向スコアの分布を確認する

- `plot(マッチング・オブジェクト名, type = "jitter")` で出力

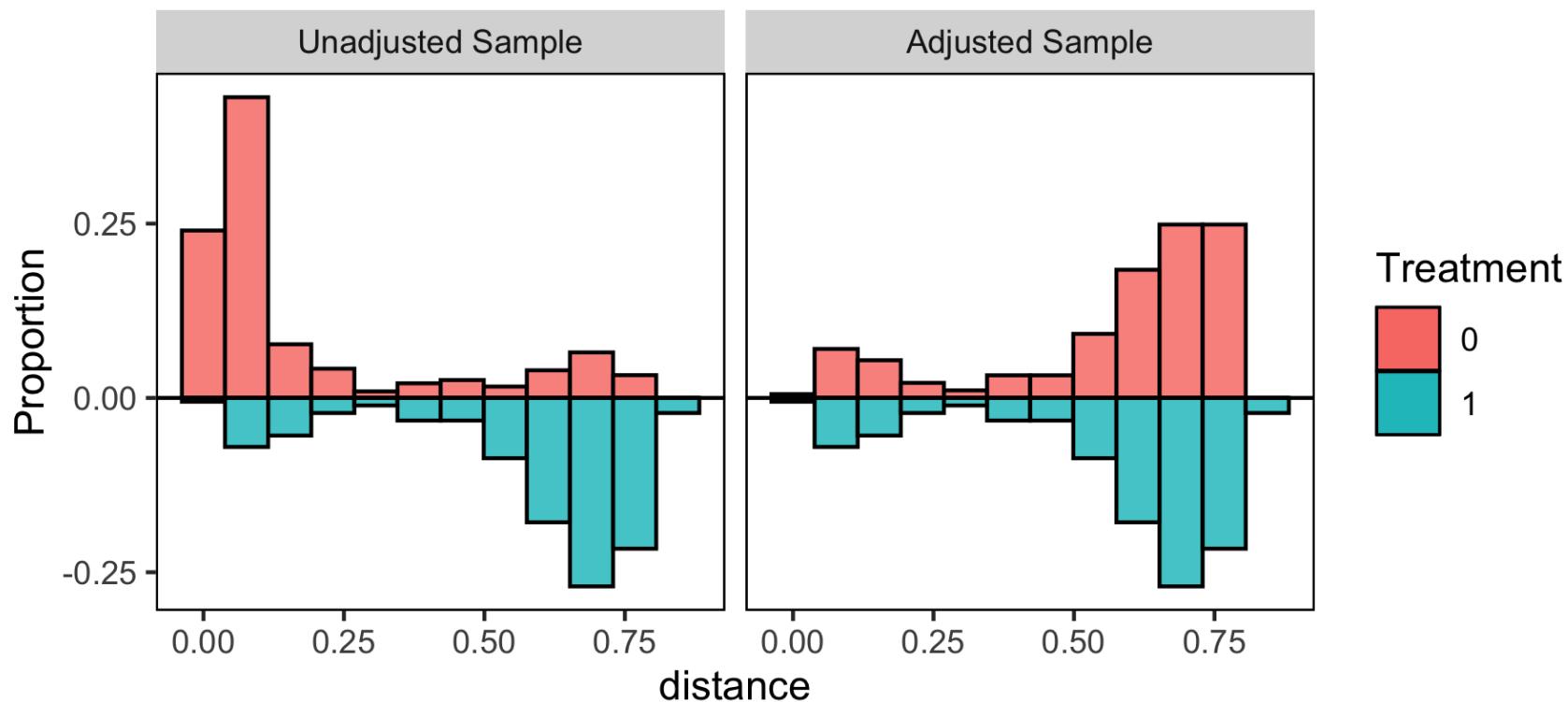


{cobalt}の利用 (1)

傾向スコアのヒストグラム: `cobalt::bal.plot()`

- ヒストグラムが上下対称ならバランス

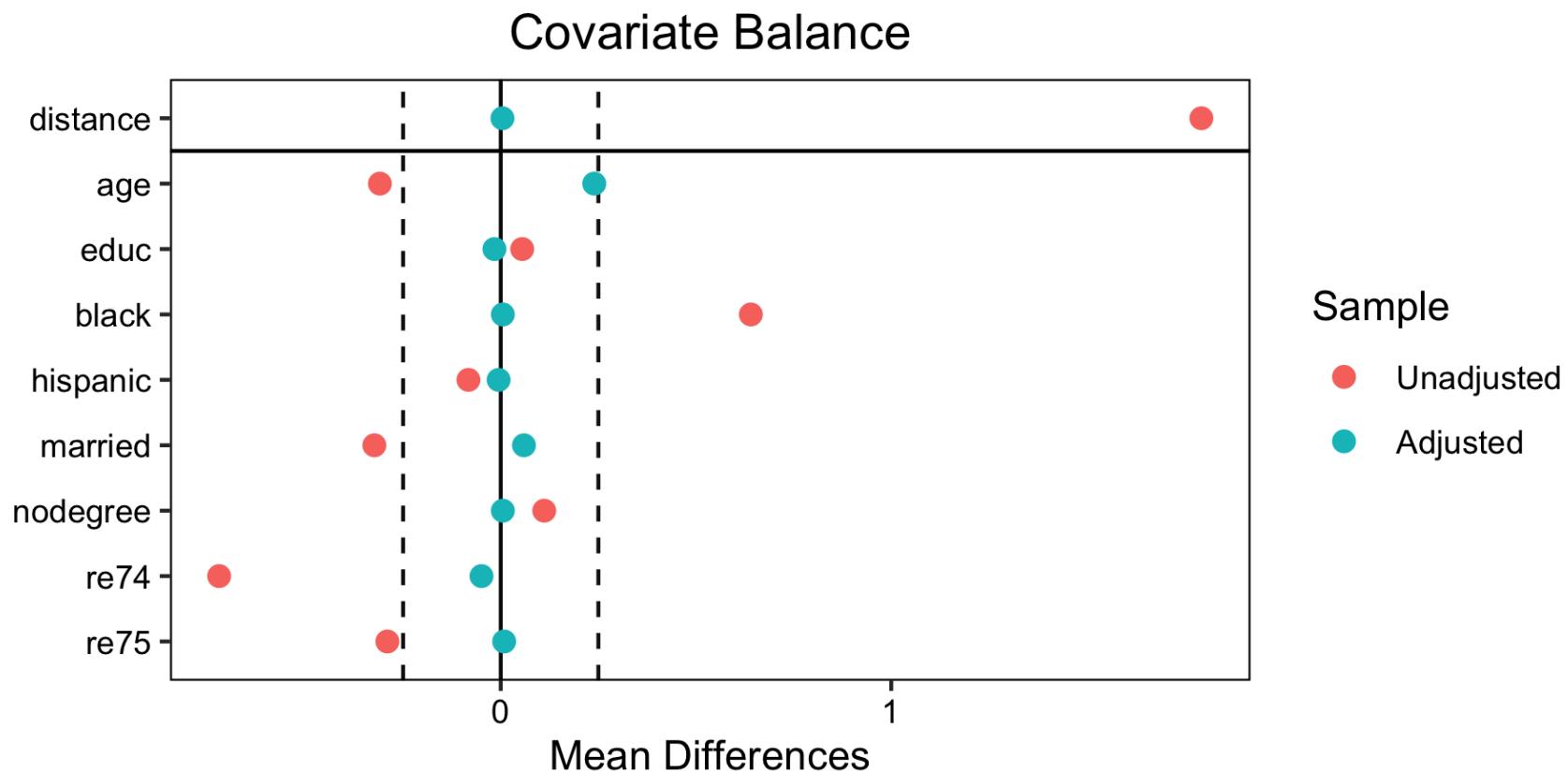
Distributional Balance for "distance"



{cobalt}の利用 (2)

ラブ・プロット: `cobalt::love.plot()`

- 標準化差分によるバランス・チェック



真の因果効果は?

手法ごとに異なる処置効果

- 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ない
- 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かを確認

Matching Methods

