

Balanced Comparison

労働経済学 2

川田恵介

Table of contents

1	Estimand	2
1.1	論点整理	2
1.2	例: “人種” 間格差	2
1.3	例: 平均格差	3
1.4	Estimand	3
1.5	準備: 繰り返し期待値の法則	3
1.6	正式な定義: 平均値の分解	4
1.7	例: $X = immigrant$	4
1.8	含意	4
1.9	Balanced mean	4
1.10	Target Weight	5
1.11	例: $X = immigrant$	5
1.12	例: Balanced Mean	5
1.13	仮定: Overlap	6
1.14	付論: 別解釈	6
2	推定方法	6
2.1	例: 移民、調査年	6
2.2	例: Balanced Comparison	7
2.3	例: 移民、調査年、年齢	7
2.4	例: 移民、調査年、年齢、その他	8
2.5	事例数問題	8
2.6	まとめ	8
	Reference	9

1 Estimand

- $E[Y|D = 1, X] - E[Y|D = 0, X]$ として定義される格差をどのように推定するのか？
 - X をバランスさせた比較
 - 因果推論/格差、両面で有効
 - * 理想的な因果推論に比べて、 X の分断が大きく、注意深い議論が必要な場合が多い

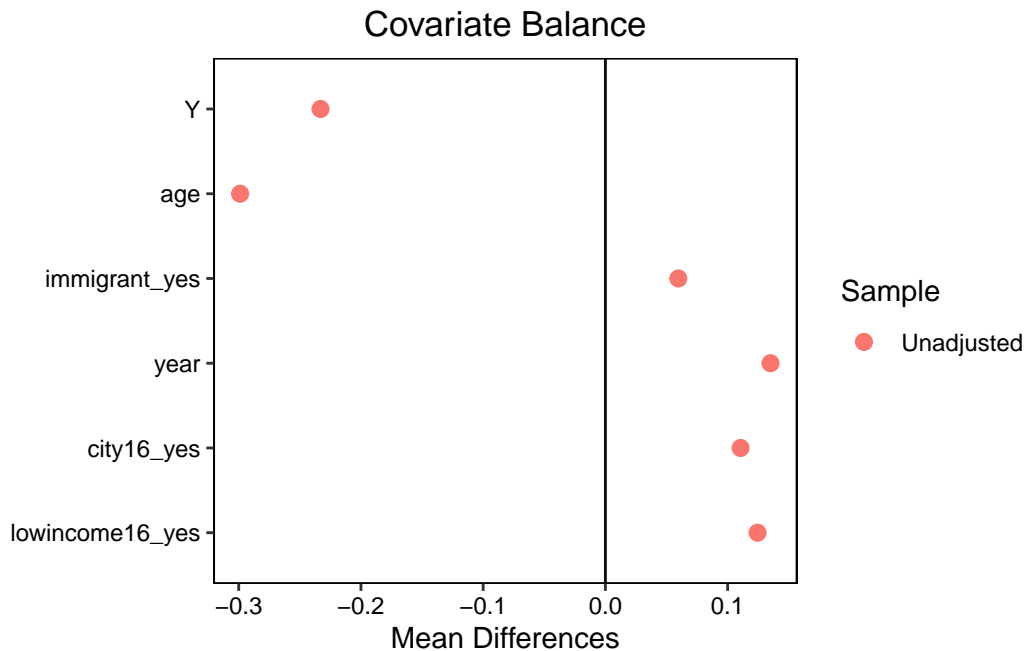
1.1 論点整理

- Balancing weights を用いて議論を整理
 - 因果効果や格差を推定するために、データから推定する必要がある
 - 傾向スコア、マッチング、IPW、Double Machine Learning、OLS など、Balancing weights を推定する手法として整理できる
 - * Chattopadhyay, Hase, and Zubizarreta (2020), Bruns-Smith et al. (2023), Ben-Michael et al. (2021)

1.2 例: “人種” 間格差

- “Race” 間での教育格差を推定
 - データ: US General Social Survey 1974-2002
 - $Y =$ 教育年数 / $D =$ “それ以外” / “白人”
 - $X =$ 年齢 / 移民かどうか / 何年の調査か
 - * 同じ社会 / コホート / 出身国内で差があれば、格差

1.3 例: 平均格差



- “白人”の方が、移民が少なく、年齢が高く、過去の調査に多い

1.4 Estimand

- 推定対象は、人種 (D) 間で、年齢、調査年、移民状態 (X) を Balance させた後の、教育年数 (Y) の平均差
 - Balanced Comparison
- Point: Balance とは何か?

1.5 準備: 繰り返し期待値の法則

- 繰り返し期待値の法則

大阪大学の平均身長 = 経済学部生の平均身長

×経済学部生の割合

+社会学部生の平均身長

×社会学部生の割合

+...

1.6 正式な定義: 平均値の分解

- 一般にデータ上の Y の平均値は以下のように書き換えられる

$$d \text{ における } Y \text{ の平均値} = \left\{ (x \ \& \ d) \text{ における } Y \text{ の平均値} \right. \\ \left. \times d \text{ における } x \text{ の割合} \right\} \text{ の } x \text{ についての総和}$$

1.7 例: $X = immigrant$

$E[Y D,X]$	D	immigrant	$f(x d)$
11.9	1	no	0.843
13.0	1	yes	0.157
12.8	0	no	0.902
12.3	0	yes	0.098

- $D = 1$ の平均教育年数

$$\underbrace{12.0727}_{E[Y|1]} = \underbrace{0.843}_{f(no|1)} \times \underbrace{11.9}_{E[Y|1,no]} + \underbrace{0.157}_{f(yes|1)} \times \underbrace{13.0}_{E[Y|1,yes]}$$

1.8 含意

- D 間での格差を生み出す要因は 2 種類に分解できる
 - $(x \ \& \ d)$ における Y の平均値 の違い (X 内での格差)
 - d における x の割合 の違い (X の格差)
- Balanced Comparison: 後者を排除

1.9 Balanced mean

- “ X の格差” を排除した平均値

$$Balanced \ Mean = \left\{ (x \ \& \ d) \text{ における } Y \text{ の平均値} \right. \\ \left. \times \underbrace{x \text{ の仮想的な割合}}_{Target \ Weight} \right\} \text{ の } x \text{ についての総和}$$

1.10 Target Weight

- Target Weight は D 間で共通
 - 研究者が設定する必要がある
- 代表例として
 - データ全体での x の割合 $f(X)$
 - $D = 1$ または $D = 0$ における x の割合 $f(X|D = 1), f(X|D = 0)$

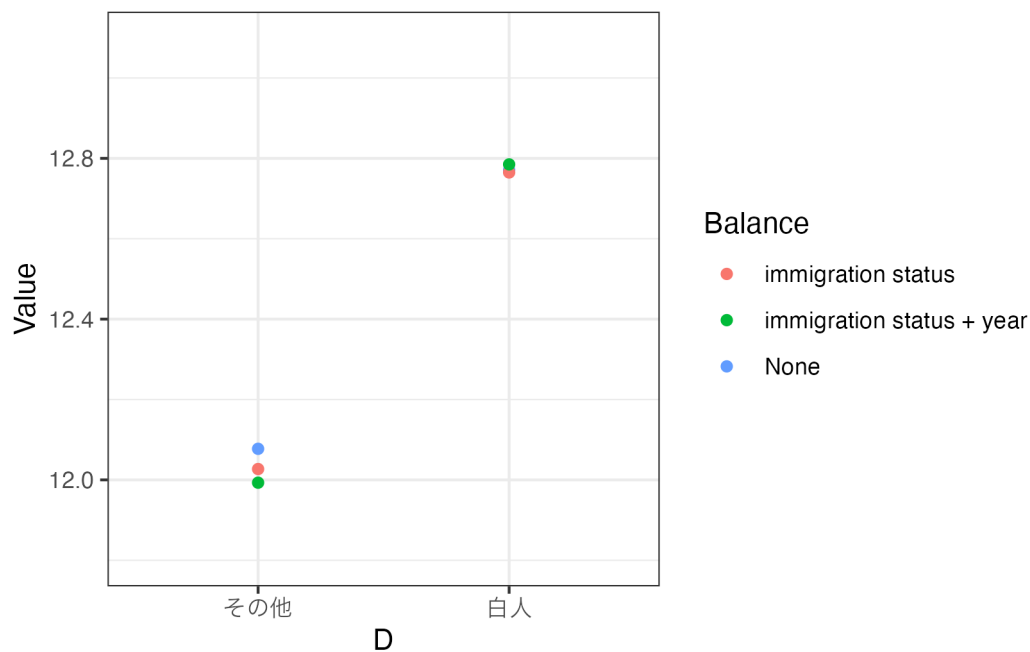
1.11 例: $X = immigrant$

$E[Y D,X]$	D	immigrant	$f(x d)$	Target= $f(x)$	Target= $f(x 1)$	Target= $f(x 0)$
11.9	1	no	0.843	0.891	0.843	0.902
13.0	1	yes	0.157	0.109	0.157	0.098
12.8	0	no	0.902	0.891	0.843	0.902
12.3	0	yes	0.098	0.109	0.157	0.098

- $f(x)$ を使用した $D = 1$ の Balanced Mean

$$12.01 = \underbrace{0.89}_{f(no)} \times \underbrace{11.9}_{E[Y|1,no]} + \underbrace{0.11}_{f(yes)} \times \underbrace{13.0}_{E[Y|1,yes]}$$

1.12 例: Balanced Mean



1.13 仮定: Overlap

- Balancing weight が母集団において”存在する”前提
- “すべての X の組み合わせについて、 $D = d$ もそれ以外も存在する”
 - $f(X)$ を Target Weight とするのであれば、

$$1 > f(D = d|X = x) > 0$$

- $f(X|D = d)$ を Target Weight とするのであれば、

$$1 > f(D = d|X = x) \geq 0$$

1.14 付論: 別解釈

- Stratified estimation としても解釈できる
1. すべての X の組み合わせについて、 D 間での Y の平均差 $\tau(X)$ を計算
 2. $\tau(X)$ の”平均値”を計算

2 推定方法

- シンプルな方法で推定できる
 - Balancing Weight を直接計算 (MatchIt package など)
 - Stratified estimation
- X の組み合わせが増えると、事例数の問題から推定できない

2.1 例: 移民、調査年

```
Weight = MatchIt::matchit(  
  D ~ immigrant + year,  
  data = Data,  
  method = "Exact", # Balancing Weight  
  estimand = "ATE"  
)  
  
Weight
```

A matchit object

- method: Exact matching
- number of obs.: 9120 (original), 9120 (matched)
- target estimand: ATE
- covariates: immigrant, year
 - number of obs. = 事例数
 - original: 元の事例数
 - matched: weights を計算できた事例数

2.2 例: Balanced Comparison

```
lm(  
  Y ~ D,  
  Data,  
  weights = Weight$weights  
)
```

Call:

```
lm(formula = Y ~ D, data = Data, weights = Weight$weights)
```

Coefficients:

(Intercept)	D
12.785	-0.792

2.3 例: 移民、調査年、年齢

```
MatchIt::matchit(  
  D ~ immigrant + year + age,  
  data = Data,  
  method = "Exact",  
  estimand = "ATE"  
)
```

A matchit object

- method: Exact matching
- number of obs.: 9120 (original), 7788 (matched)
- target estimand: ATE

- covariates: immigrant, year, age
- 年齢も加えると Weight が計算できない事例が増える

2.4 例: 移民、調査年、年齢、その他

```
MatchIt::matchit(
  D ~ age + lowincome16 + city16 + immigrant + siblings + year,
  data = Data,
  method = "Exact",
  estimand = "ATT"
)
```

A matchit object

- method: Exact matching
- number of obs.: 9120 (original), 1261 (matched)
- target estimand: ATT
- covariates: age, lowincome16, city16, immigrant, siblings, year
- 居住地なども加えると Weight が計算できない事例がさらに増える

2.5 事例数問題

- 母集団において Overlap が成立していたとしても、事例数が限られるデータにおいては、 $D = 1/0$ のどちらかしか存在しない X が発生する

D	immigrant	year	age
0	no	2002	57
0	no	2002	81
0	no	2002	69
0	yes	2002	75
0	no	2002	64
0	yes	2002	58
0	yes	2002	56
0	yes	2002	76
0	no	2002	79
0	yes	2002	70

2.6 まとめ

- X をバランスさせたもとの比較 (Balanced comparison) は、Balancing weights の算出が要求する

- X の組み合わせに対して、事例数が十分あれば、Overlap の仮定のもとで、 X を "Exact" にバランスさせる Weight を計算できる
- 多くの実践で不十分
 - * OLS (Imbens 2015; Chattopadhyay and Zubizarreta 2022) や傾向スコア、明示的な最適化 (Hainmueller 2012; Zubizarreta 2015) を使用する必要がある

Reference

- Ben-Michael, Eli, Avi Feller, David A Hirshberg, and José R Zubizarreta. 2021. "The Balancing Act in Causal Inference." *arXiv Preprint arXiv:2110.14831*.
- Bruns-Smith, David, Oliver Dukes, Avi Feller, and Elizabeth L Ogburn. 2023. "Augmented Balancing Weights as Linear Regression." *arXiv Preprint arXiv:2304.14545*.
- Chattopadhyay, Ambarish, Christopher H Hase, and José R Zubizarreta. 2020. "Balancing Vs Modeling Approaches to Weighting in Practice." *Statistics in Medicine* 39 (24): 3227–54.
- Chattopadhyay, Ambarish, and José R Zubizarreta. 2022. "On the Implied Weights of Linear Regression for Causal Inference." *Biometrika*, asac058.
- Hainmueller, Jens. 2012. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis* 20 (1): 25–46.
- Imbens, Guido W. 2015. "Matching Methods in Practice: Three Examples." *Journal of Human Resources* 50 (2): 373–419.
- Zubizarreta, José R. 2015. "Stable Weights That Balance Covariates for Estimation with Incomplete Outcome Data." *Journal of the American Statistical Association* 110 (511): 910–22.