

異質性分析

機械学習入門

川田恵介

Table of contents

1	異質性	1
1.1	Fukai, Ichimura, and Kawata (2021)	2
1.2	個人因果効果の予測	2
1.3	職業訓練の効果	2
2	バランス後の比較: 再訪	2
2.1	バランス後の比較	2
2.2	バランス後の平均差	3
2.3	再解釈	3
2.4	復習: 線型モデルの機能	3
2.5	R learner の拡張	3
2.6	解釈	4
2.7	例	4
2.8	発展	4
	Reference	4

1 異質性

- バランスさせた後の比較
 - 格差/因果分析の重要ツール
- ここまで: 平均的な差に焦点
 - ここから: 異質性も含めて理解したい

1.1 Fukai, Ichimura, and Kawata (2021)

- コロナ前 (2019 年) と後 (2020 年) の 4~6 月の就業率を比較したい
 - COVID とその対応策の影響を間接的に評価する
- $Y =$ 就業状態、 $D = 2020/2019$ 、 $X =$ 年齢等の基礎変数 + 前年の就業状態
- 平均的な差は限定的
 - 就業状態が流動的であると考えられる、一部の層に大きな影響がある

1.2 個人因果効果の予測

- 個人ごとに因果効果が予測できれば、
 - 医療行為の個別化: “きき方”/副反応の深刻さ、に応じて医療行為を変える
 - * ある医療の行為の因果効果を、 X から予測する
 - 単位取得計画のサジェスト: 学生の志望進路や興味関心などに応じて、おすすめ単位取得を示す
 - * 講義の受講の因果効果を、 X から予測する
- 実務の例: [EconML \(MicroSoft\)](#), [CausalML \(Uber\)](#)

1.3 職業訓練の効果

- フランスにおける実験: 職業訓練の提供主体を公的機関/民間企業にランダムに割り付ける
- Behaghel, Crépon, and Gurgand (2014): 平均的には公的機関の方が効果的
 - Kallus (2023): 異質性が大きい
 - 一部の労働者によって公的機関による訓練の影響が強いだけで、平均的には民間企業の訓練の方が効果的

2 バランス後の比較: 再訪

2.1 バランス後の比較

- 現実の Size の割合は、取引年に応じて異なる
 - Target とバランスさせる

Y の平均値	D	Size	N	Size の割合	Target
62.7	0	75	301	0.796	0.5
71.7	1	75	414	0.838	0.5
84.0	0	90	77	0.204	0.5
101.9	1	90	80	0.162	0.5

Y の平均差	Size	Target
9.0	75	0.5
17.9	90	0.5

2.2 バランス後の平均差

- 一旦 X 内で平均差を計算した後に、平均値を計算する
- ここまでの議論: バランス後の平均差 $= 90 \times 0.5 + 17.9 \times 0.5 = 13.5$
 - バランス後の平均差のみでは、Size によって平均差が異なる情報は廃棄される

2.3 再解釈

- ここまで: X 内での D 間の Y の平均値の差 を一つの数値 β_0 で近似する
- これから: X 内での D 間の Y の平均値の差 をシンプルなモデル $\beta_0 + \beta_1 X_1 + \dots + \beta_L X_L$ で近似する
 - 平均差が X_1, \dots, X_L に応じて、どのように異なるのか、知見をもたらす

2.4 復習: 線型モデルの機能

- $Y \sim X_1 + \dots + X_L$ で推定して、 Y と X の特徴を人間が理解するためのモデルを推定する
 - 人間の理解を要求しない予測モデルと比べて、単純なモデルである必要がある

2.5 R learner の拡張

- ここまで: $Y - E[Y|X] \sim D - E[D|X]$ を OLS で推定
- これから:

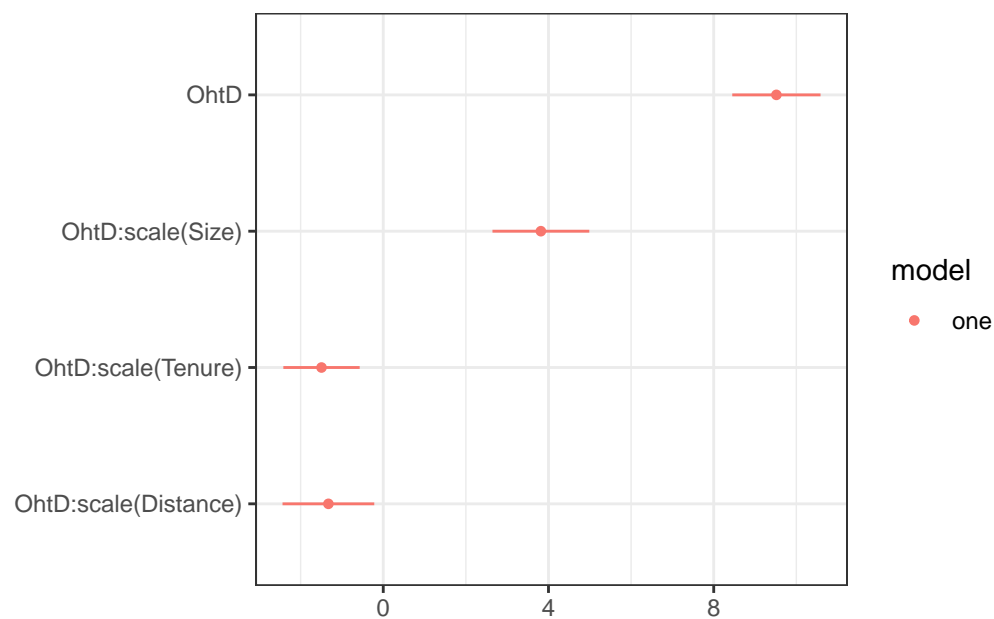
$$\begin{aligned}
Y - E[Y|X] &\sim \underbrace{(D - E[D|X])}_{\text{主効果}} \\
&+ \underbrace{X_1 \times (D - E[D|X]) + \dots + X_L \times (D - E[D|X])}_{\text{交差項}}
\end{aligned}$$

を OLS で推定

2.6 解釈

- 主効果 \simeq ここまでバランス後の平均差
- 交差項 $X_l \times (D - E[D|X]) \simeq$ 平均差が X_l が増加した場合に、どのように変化するか

2.7 例



2.8 発展

- 異質性の推定については、多くの手法が提案され、現在でも研究が続いている
 - 参照: [CausalML](#) : Chap. 14-15

Reference

- Behaghel, Luc, Bruno Crépon, and Marc Gurgand. 2014. “Private and Public Provision of Counseling to Job Seekers: Evidence from a Large Controlled Experiment.” *American Economic Journal: Applied Economics* 6 (4): 142–74.
- Fukai, Taiyo, Hidehiko Ichimura, and Keisuke Kawata. 2021. “Describing the Impacts of COVID-19 on the Labor Market in Japan Until June 2020.” *The Japanese Economic Review* 72 (3): 439–70.

Kallus, Nathan. 2023. “Treatment Effect Risk: Bounds and Inference.” *Management Science* 69 (8): 4579–90.