

## 統計的因果推論による数学の学習支援効果の定量的分析

永井朋子<sup>1,2</sup>, 奥田喬之<sup>2</sup>, 中村友哉<sup>2</sup>, 佐藤雄一郎<sup>2</sup>, 佐藤悠介<sup>2</sup>, 金城謙作<sup>2</sup>,  
河村建吾<sup>3</sup>, 菊田伸<sup>2</sup>, 熊ノ郷直人<sup>2</sup>

国際医療福祉大学<sup>1</sup> 工学院大学<sup>2</sup> 大阪産業大学<sup>3</sup>

## 1. はじめに

工学院大学学習支援センター(以後“センター”)数学科は決定木により, 対面と遠隔による個別指導と基礎講座, およびセンター数学科の学修支援システムのオンデマンド教材などによる教育効果を定量的に分析した[1]. 一方, 統計的因果推論と呼ばれるデータから因果関係を推定する理論が発展した. 処置を施したときの処置効果を定量化するための枠組みが確立され, 「ルービンの反事実モデル」と呼ばれている[2]. 近年は因果推論と機械学習を融合した手法も提案されている[3]. 我々の決定木の教育効果は統計的因果推論における処置効果の定義とは異なる. 本研究の目的は統計的因果推論の視点から, 2022年度のデータを用いてセンター数学科の処置効果を推定することである.

工学院大学では4学期制(1Q~4Q)を導入しており, 各学期末に定期試験を実施している. 4月入学時にテスト形式の習熟度調査も実施された.

## 2. 統計的因果推論と機械学習を融合した枠組み

反事実モデルと機械学習を融合した T-learner[3]と呼ばれる枠組みを適用し, 本研究では処置をセンター数学科の「対面個別」指導の利用と設定する. 目的変数を 1Q「微分」定期試

験の偏差値とする. 定期試験の偏差値に対する「対面個別」の処置効果を定量化する. 説明変数 $\mathbf{X}$ については,  $X_1$ を習熟度調査の偏差値,  $X_2$ を1Qの「対面個別」の利用回数として, 2.1では1次元ベクトル $\mathbf{X} = (X_1)$ , 2.2では2次元ベクトル $\mathbf{X} = (X_1, X_2)$ とした.

学生 $k$ が「対面個別」を利用したときの定期試験の偏差値を $Y^k(1)$ とする. 同一の学生 $k$ が「対面個別」を利用しないときの定期試験の偏差値を $Y^k(0)$ とする. 反事実モデルでは, 学生 $k$ の個体処置効果は $D^k = Y^k(1) - Y^k(0)$ によって定義される. しかし $Y^k(1)$ と $Y^k(0)$ どちらか一方は必ず欠測値であるため,  $D^k$ を計算できない.

T-learner[3]では, 1Qの「対面個別」利用者91人の説明変数 $\mathbf{X}$ と定期試験の偏差値から成る実測データの組 $(\mathbf{X}^{i,obs}, Y^{i,obs}(1))$  ( $i = 1, \dots, 91$ )から機械学習によって,  $\mathbf{X}$ と目的変数 $Y(1)$ を対応させる関数 $\mu_1(\mathbf{x}) = \mathbb{E}[Y(1)|\mathbf{X} = \mathbf{x}]$ を推定し,  $\hat{\mu}_1(\mathbf{x})$ とする. 非利用者1298人の同様の実測値,  $(\mathbf{X}^{j,obs}, Y^{j,obs}(0))$  ( $j = 92, \dots, 1389$ )から, 関数 $\mu_0(\mathbf{x}) = \mathbb{E}[Y(0)|\mathbf{X} = \mathbf{x}]$ を推定し,  $\hat{\mu}_0(\mathbf{x})$ とする.

2.1 T-learnerによる結果( $\mathbf{X}, \mathbf{x} \in \mathbb{R}^1$ )

ここでは $\mathbf{X} = (X_1)$ は1次元ベクトルとする. 1Q「微分」定期試験受験者のうち, 習熟度調査偏差値が $x_1$ の値の学生の集合を $S_{x_1}$ とする. 全受験者数は $N = 1389$ である. 機械学習で推定した関数 $\hat{\mu}_1(\mathbf{x})$ と $\hat{\mu}_0(\mathbf{x})$ を用いて, 条件付き処置効果の推定量 $\bar{\tau}(x_1)$ [3]を次のように計算する.

$$\bar{\tau}(x_1) = \frac{1}{|S_{x_1}|} \sum_{k \in S_{x_1}} \{\hat{\mu}_1(X_1^{k,obs}) - \hat{\mu}_0(X_1^{k,obs})\} \quad (1)$$

$X_1^{k,obs}$ は学生 $k$ の実測の習熟度調査偏差値である. 用いる機械学習はランダムフォレスト(回帰, 深さ2)とした. Pythonのscikit-learnにて実装した.

Analysis of Educational Effects in Mathematics by means of Causal Inference

1,2 Tomoko Nagai, International University of Health and Welfare, Kogakuin University

2 Takayuki Okuda, Tomoya Nakamura, Yuichiro Sato, Yusuke Sato, Kensaku Kinjo, Shin Kikuta, and Naoto Kumano-go, Kogakuin University

3 Kengo Kawamura, Osaka Sangyo University

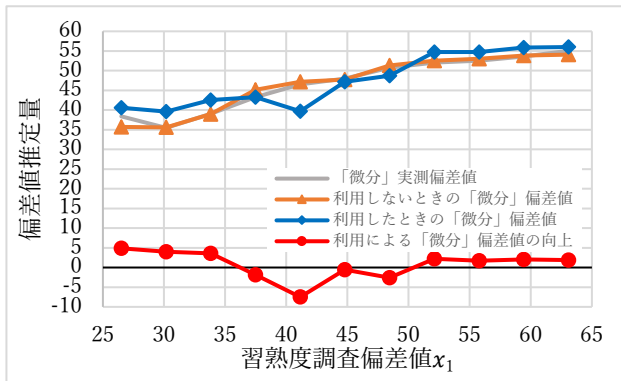


図1 偏差値推定量の習熟度調査偏差値依存性

図1に各偏差値推定量の習熟度調査偏差値依存性を示す。対面個別を利用しないときの1Q「微分」定期試験の偏差値推定量 $\hat{\mu}_0(x_1)$ (図1の▲)は概ね緩やかに単調増加する傾向が見られる。対面個別を利用したときの定期試験の偏差値推定量 $\hat{\mu}_1(x_1)$ (◆)は $\hat{\mu}_0(x_1)$ に対して上下する。式(1)の $\bar{\tau}(x_1)$ は対面個別を利用したことによる定期試験偏差値の向上(●)を意味し、習熟度調査偏差値35以下と50以上で正の値を示し、それぞれ4~5と2程度の向上であった。

### 2.2 T-learner による結果( $X, x \in \mathbb{R}^2$ )

次に「対面個別」の利用回数を説明変数に加え、 $X$ を2次元ベクトル $X = (X_1, X_2)$ とした。2.1同様に実測値から機械学習により関数を推定する。非利用者の「対面個別」利用回数実測値は $X_2^{j,obs} = 0$ である。さらに式(1)を拡張して本研究で新たに利用回数が $x_2$ のときの処置効果推定量を

$$\varphi(x_1, x_2) = \frac{1}{|S_{x_1}|} \sum_{k \in S_{x_1}} \{ \hat{\mu}_1(X_1^{k,obs}, x_2) - \hat{\mu}_0(X_1^{k,obs}, X_2^{k,obs}) \} \quad (x_2=1, 2, \dots) \quad (2)$$

によって定義する。推定された関数 $\hat{\mu}_0(x)$ が $\hat{\mu}_0(X_1^{k,obs}, X_2^{k,obs}) = \hat{\mu}_0(X_1^{k,obs}, 0) = \hat{\mu}_0(X_1^{k,obs}, x_2)$ を満たすことを確認した。図2に処置効果推定量の「対面個別」利用回数依存性を示す。習熟度調査の偏差値 $x_1 \leq 35$ では利用回数 $x_2$ が1回(図2の▲)から2回(◆)で式(2)の処置効果推定量 $\varphi(x_1, x_2)$ が小さくなり、3回(●)以降から $\varphi(x_1, x_2)$ が増加する。習熟度調査の偏差値 $x_1 \geq 50$ では $x_2$ が3回(●)までは $\varphi(x_1, x_2)$ は2程度である

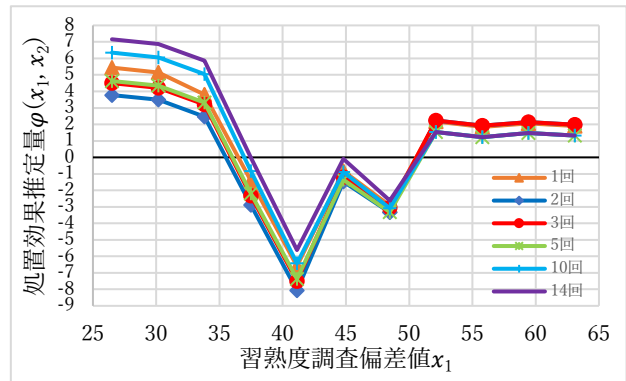


図2 処置効果推定量の対面個別利用回数依存性

が、5回(✕)以降で $\varphi(x_1, x_2)$ は1.4程度になった。

### 3. まとめ

統計的因果推論によりセンター数学科の対面個別指導の処置効果の定量的な分析を行った。

- 習熟度調査の偏差値が35以下と50以上で、「対面個別」の利用により定期試験の偏差値の向上が推定された。
- T-learnerの枠組みを拡張し、処置の回数、すなわち「対面個別」の利用回数に対応した処置効果推定量を提案した。習熟度調査の偏差値が35以下では、「対面個別」利用回数が3回以上から定期試験の偏差値の変化が大きくなる傾向が予測された。習熟度調査の偏差値が50以上では、「対面個別」利用回数の増加により定期試験の偏差値の変化が大きくなる傾向は予測されなかった。

### 参考文献

[1] 永井朋子, 金城謙作, 河村建吾, 中村友哉, 奥田喬之, 佐藤雄一郎, 棕野純一, 菊田伸, 熊ノ郷直人, 工学院大学学習支援センター数学科の教育効果-決定木の利用-, 工学教育, 71巻, 第3号, pp.112-116, 2023.

[2] 高橋将宜, 統計的因果推論の理論と実装 潜在的結果変数と欠測データ, 共立出版, 2022.

[3] S. R. Künzel *et al.*, Metalearners for Estimating Heterogeneous Treatment Effects Using Machine Learning, Proc. Natl. Acad. Sci., Vol. 116, No.10, pp. 4156-4165, 2019.