

オンライン補習授業の視聴と試験成績の関係：

学習履歴データに基づく因果推論アプローチの検討

田 中 頼 人¹

1. はじめに

統計的因果推論は、「原因」と「結果」の関係をデータに基づいて明らかにする手法であり、科学的な意思決定を支える重要な方法論である¹⁾。特に教育現場では、特定の教育施策や補習授業が学習成果に与える影響を定量的に評価することで、教育施策の改善や最適化を目指すことが可能となる。このような因果推論の方法論の中でも、ランダム化比較試験（Randomized Controlled Trial: RCT）は、その科学的信頼性が極めて高く、標準的な手法とされている²⁾。

RCTとは、対象をランダムに割り付けることで、施策を受けたグループと受けていないグループを比較し、因果関係を検証する手法である。このランダム化によって、交絡因子の影響を最小限に抑え、純粋な因果効果を測定することが可能となる。交絡因子とは、処置（独立変数）と結果（従属変数）の両方に影響を与える要因であり、その存在によって因果関係の推定が歪められることがある。例えば、教育施策の効果を評価する際、学生の「学習意欲」が交絡因子として働く場合がある（図1）。学習意欲は、補習を受講するかどうか（処置）にも影響し、試験成績（結果）にも影響を及ぼす。交絡因子を適切に調整しないと、補習そのものの効果を過大評価または過小評価するリスクが生じる。

RCTの大きな利点は、このような交絡因子の影響をランダム化によって排除できる点にある。そのため、RCTは医学や教育、政策評価など幅広い分野で活用されている。しかし、RCTにはいくつかの制約がある。その一つが倫理的制約である。例として、「タバ

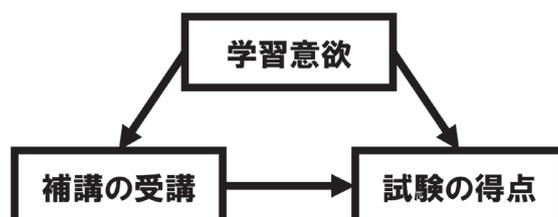


図1 交絡因子の存在

¹ サイバー大学 IT 総合学部・准教授

コを吸うと肺がんになるか」を明らかにする場合、ランダムに割り当てられた実験群の被験者に強制的にタバコを吸わせる実験は倫理的に許されない。同様に、教育現場で「新しい教材の効果があるか」を検証する場合、新しい教材を使う実験群と旧来の教材を使う対照群にランダムに分ける教育の実施は、公平性や倫理的観点から困難であることが多い。

さらに、RCTには時間とコストの課題も存在する。準備期間の長さや被験者の募集、データ収集の労力は大きく、現実的に実施が難しい場合が多い。特に教育現場のようにリソースが限られる環境では、RCTを実施することは非現実的であることが少なくない。そのため、既存の観察データを活用した因果推論手法が現実的な選択肢として注目されている。

観察データを用いた因果推論には、層別解析、回帰分析、傾向スコアマッチングなどが含まれる^{3) 4) 5)}。これらの手法では、交絡因子の影響を適切に調整することで、比較的信頼性の高い因果関係を導き出すことが可能である。特に層別解析は、データを交絡因子の値に応じて層に分け、その層内で因果効果を評価する手法である。層ごとに比較を行うことで、交絡因子の影響を排除し、因果関係をより正確に推定できる。

本稿では、観察データを基にした因果推論手法の中でも、比較的導入が容易で理解しやすい層別解析に焦点を当てる。この手法を用いて、サイバー大学（以降「本学」とする）オンライン補習授業の履修データを分析し、学習意欲や補習受講が試験成績に与える因果効果を検討する。層別解析は、データの特性に応じて柔軟に適用できる手法であり、教育現場で得られる多様なデータにも有効であることが期待される。

2. 層別解析の実践

2.1. 用いるデータ

本研究では、本学の専門科目「プログラミング入門」において、2023年春学期に実施された際のデータを使用する。取得されたデータは、期末試験の得点や教材の受講完了状況、補習授業の受講有無、学習者の属性などの情報を含む、473名分のものである。データ収集は学習管理システム（LMS）であるCloud Campusを通じて行い、データは個人を特定できないように匿名化し、統計処理のみに利用した。

補習授業は筆者が約10分間の動画形式で提供し、内容は期末試験に向けた復習の要点や心構えを説明するものであった（図2）。この補習授業は、期末試験開始前に科目内のお知らせ掲示板を通じて学習者に案内した。試験期間終了後、受講者の補習受講の有無、期末試験の得点、性別や居住地域などの属性データを収集した。

本研究では、交絡要因となる「学習意欲」の指標として、各受講者が科目内の全チャプター（章）のうち何割を完了したかを用いた。この指標は、学習者のLMS上の活動記録をもとに算出されたものである。学習への意欲は直接測定することが難しく、意欲について問う質問紙調査では十分な回答率を得られない問題があるため、LMS上の学習履歴データに基づいた間接的な指標を学習意欲として採用した。

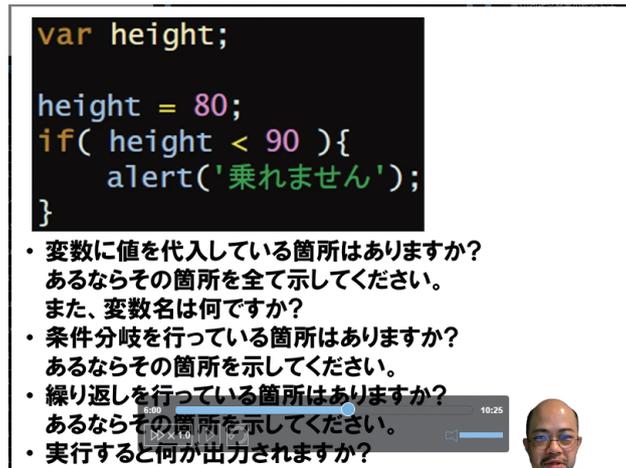


図 2 補習授業の様子

表 1 用いるデータのイメージ

学習意欲	補習受講ダミー	試験得点
1	0	88
0.973	1	80
0.986	1	92

分析においては、試験得点の単位を正規化し、100点満点に換算した値を用いた。また、補習受講の有無はダミー変数（1：受講、0：未受講）として扱った。後述する「交絡因子の数が増えるとグループ数が爆発的に増加」の問題を避けるため、本研究では「学習意欲、および補習受講ダミー」の2変数のみを独立変数とし「試験の得点」を従属変数とした。これらのデータを準備し、補習受講の有無が試験成績に与える影響を分析した（表1）。

2.2. 処理方法

本研究では、データ処理および分析を以下の手順で実施した。まず、期末試験得点が0点である学習者をデータから除外した。この操作は、試験を未受験の学習者や、特別な理由で得点が記録されていないケースを排除し、分析結果の精度を向上させる目的で行った。

次に、学習意欲を層別化するため、チャプター完了率を基に0.993を閾値として「高」「低」に分類した。この閾値は、データの分布を観察し、学習意欲の高い学習者と低い学習者をバランスよく区分するために設定したものである。閾値の0.993は、データ全体の学習意欲の上位層に位置する学習者を特定する目的で選定された。この設定により、「高」グループは学習意欲が顕著に高い学習者を代表し、「低」グループはそれ以外の学習者を含む分類が可能となった。

表2 層別解析の集計表

学習意欲層	補講未受講者の平均得点	補講受講者の平均得点
高	93.43	95.16
低	85.51	91.75

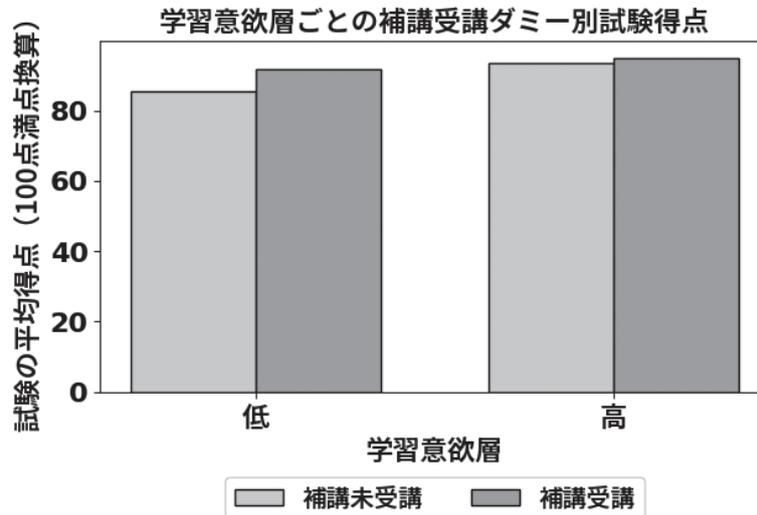


図3 層別解析のクラスター棒グラフ

最後に、層別解析を実施した。学習意欲の高低ごとに、補講受講ダミー（1: 補講受講、0: 未受講）別の試験得点平均を計算した。この層別解析は、学習意欲を交絡因子として調整しつつ、補講受講の効果を明確化することを目的としている。また、この解析によって、学習意欲の高低による補講受講の効果の違いを定量的に表した。出力結果は表2である。また、表2の結果を補完し、各学習意欲層内で補講の有無による得点の違いを可視化するために描画したクラスター棒グラフを図3に示す。本研究で用いた層別解析のPython プログラムは末尾の付録に示す。

3. 考察

本研究の層別解析では、学習意欲を「高」「低」の2層に分け、それぞれの層内で補講受講の有無による期末試験得点の平均差を評価した。学習意欲が「低」の層では、補講受講者の平均得点が未受講者を上回る結果が得られ、補講がこの層の学習者に一定の効果をもたらしている可能性が示唆された。一方、学習意欲が「高」の層においても、得点差が比較的小さいものの、補講を受講した学習者の方が試験成績が良好であることが観察された。

層別解析は、学習意欲という交絡因子を考慮しつつ補講の効果を分析する有用な手法である。しかし、この手法にはいくつかの限界も存在する。例えば、層内での比較において

も、他の交絡因子（例：学習者の学力、時間的制約、受講環境など）が試験成績に影響を及ぼしている可能性がある。これらの因子が補講受講や学習意欲と関連している場合、解析結果が偏るリスクがある。

また、層別解析は学習意欲を基準とした単純な分類に依存するため、連続的な変数を扱う場合には情報が損なわれる可能性がある。本研究では学習意欲の閾値を 0.993 としたが、層の境界設定が異なると結果が変わる可能性がある。これを補うために、回帰分析や傾向スコアマッチングといったより精密な手法を導入することが、補講の効果をより厳密に評価する上で有効である。

以上の結果から、補講の効果は学習意欲による層ごとに異なる可能性が高く、今後の研究では交絡因子をより精緻に調整した分析が必要であると考えられる。

4. まとめと今後の展望

本研究では、学習意欲を交絡因子とした層別解析を用い、補講受講が試験得点に与える影響を評価した。層別解析の結果、補講受講が試験得点に一定の効果を及ぼす可能性が示唆された。しかし、交絡因子としての学習意欲以外の要因（例：学習環境や事前学力）についての調整が不十分である点や、層別解析における一般的な課題が残っている。

層別解析の大きな課題の一つは、交絡因子の数が増えるにつれてグループ数が指数的に増加し、解析が複雑化する点である⁶⁾。この問題は、交絡因子が多次的に関与する教育データにおいて特に顕著である。また、サンプルサイズの制約により、一部のグループで十分なデータが得られず、有効な解析が困難になる場合がある。さらに、層別解析は基本的に線形の関係を仮定しており、非線形関係や交互作用を適切に考慮するには限界がある。

今後の研究では、これらの課題に対処するためのより高度な解析手法の導入が求められる。具体的には、回帰分析や傾向スコアマッチングを用いた多変量解析が有用である。また、機械学習手法を活用することで、非線形関係や交互作用を考慮しつつ、解析の精度を向上させることが期待できる。

注および参考文献

- 1) 中室牧子. 「原因と結果」の経済学. ダイヤモンド社, 2017, 208p.
- 2) アンドリュー・リー. RCT 大全 ランダム化比較試験は世界をどう変えたのか」みすず書房, 2020, 312p.
- 3) Scott Cunningham. 因果推論入門～ミックステープ：基礎から現代的アプローチまで, 技術評論社, 2023, 415p.
- 4) 岩崎学. 統計的因果推論. 朝倉書店, 2015, 204p.
- 5) 高橋将宜, 石田基広, 市川太祐. 統計的因果推論の理論と実装. 共立出版, 2022, 320p.
- 6) Nathan Kallus, Xiaojie Mao, Madeleine Udell. Causal Inference with Noisy and Missing Covariates via Matrix Factorization. arXiv, 2018, arXiv:1806.00811, <https://arxiv.org/abs/1806.00811> (accessed 2024-12-09).

【付録】層別解析のPythonプログラム

```
# 必要なライブラリを導入
import pandas as pd

# データの読み込み
data = pd.read_csv('因果推論入力データ.csv')

# 試験の得点が0である者を除外
filtered_data = data[data['試験の得点'] > 0]

# 試験の得点を100点満点に換算
filtered_data['試験の得点_100点満点'] = (filtered_data['試験の得点'] / filtered_data['試験の得点'].max()) * 100

# 学習意欲を基に「高」「低」を設定
threshold = 0.993
filtered_data['学習意欲層'] = filtered_data['学習意欲'].apply(lambda x: '高' if x >= threshold else '低')

# 層別:学習意欲の層別(高・低)ごとに補講受講ダミー別の試験得点平均を計算
result = filtered_data.groupby(['学習意欲層', '補習受講ダミー'])['試験の得点_100点満点'].mean().unstack()

# 分析結果の出力
print("層別解析結果:")
print(result)
```