

第 13 章 実験と準実験

James H. Stock 著・Mark W. Watson 著・宮尾 龍蔵 訳
『入門計量経済学』（共立出版、2016）

<https://www.kyoritsu-pub.co.jp/book/b10003746.html>

2026-04-10

1. 理想的な実験と因果関係の効果

ランダム化比較試験 (RCT) の理論的基礎

- ▶ **因果関係の定義:** 他の条件をすべて一定としたとき、処置 X が結果 Y に与える影響。
- ▶ **ランダム化の役割:** 処置 X をランダムに割り当てることで、 X と誤差項 u (他のすべての決定要因) の相関を完全に断ち切る。
- ▶ **OLS の仮定:** 理想的な実験では $E(u_i|X_i) = 0$ が担保され、OLS 推定量は不偏かつ一致推定量となる。

階差推定量 (Differences Estimator)

- ▶ 単純な比較: トリートメント群とコントロール群の平均の差。

$$\hat{\beta}_1 = \bar{Y}_{treatment} - \bar{Y}_{control}$$

- ▶ 回帰分析の利点:
 - ▶ 標準誤差の算出が容易。
 - ▶ コントロール変数 (W) の追加が可能。
 - ▶ W を加えても $\hat{\beta}_1$ の一貫性は保たれるが、モデルの残差分散が減るため、 $\hat{\beta}_1$ の精度 (標準誤差) が改善する。

2. 現実の実験における内部正当性の脅威

脅威 1：ランダム化の失敗と不完全な遵守

- ▶ **ランダム化の失敗**: 実際には特定の属性（やる気など）に基づいて割り振られてしまうケース。
- ▶ **不完全な遵守 (Partial Compliance)**:
 - ▶ 処置群に選ばれたが処置を受けない (**Never-takers**)。
 - ▶ 対照群に選ばれたが自力で処置を受ける (**Always-takers**)。
- ▶ **解決策**: 当初の割り当て（意図）を操作変数として用いる。

脅威 2：人員の脱落 (Attrition)

- ▶ 実験の途中で参加者が辞めてしまう。
- ▶ **問題:** もし「成績が悪い生徒ほど少人数クラスから脱落しやすい」といった傾向があると、残ったサンプルの平均成績は不当に高く見えてしまう (バイアス)。
- ▶ **対策:** 脱落した理由を調査し、脱落者が非ランダムでないかを確認する。

脅威 3：実験効果 (Hawthorne Effect)

- ▶ **ホーソン効果**: 自分が実験の対象であり、観察されているという意識が、本来の処置の効果とは別に、参加者の行動やパフォーマンスに影響を与えてしまう現象。
- ▶ これにより、推定された効果が「処置」そのものの効果なのか、「観察されていること」の効果なのか区別できなくなる。

3. テネシー州 STAR プロジェクトの深掘り

STAR プロジェクトの実験デザイン

- ▶ 1980 年代後半にテネシー州の 79 校で実施。
- ▶ 生徒と教師を以下の 3 群にランダムに割り当て：
 1. 少人数クラス (13-17 人)
 2. 通常クラス (22-25 人)
 3. 補助教員付き通常クラス (22-25 人 + 補助教員)
- ▶ 幼稚園から 3 年生までの 4 年間追跡。

推定結果の大きさの解釈

- ▶ 少人数クラスに入ることの効果（幼稚園時）：テスト成績の **0.19 標準偏差 (SD)** 分の上昇。
- ▶ この大きさの意味：
 - ▶ 教師の経験年数による効果との比較：少人数クラスの効果は、**20 年以上の経験を持つベテラン教師**に教わる効果にほぼ匹敵する。
 - ▶ 性別による差との比較：女子と男子の平均スコアの差（約 12 ポイント）よりも大きい（約 13.9 ポイント）。

効果の持続性とタイミング

- ▶ **発見:** クラス規模の効果は、低学年（幼稚園・1年生）の時に最も顕著に現れる。
- ▶ その後、少人数クラスに留まり続けても、追加的なプラス効果（累積的な拡大）は見られなかった。
- ▶ しかし、初期に得られたアドバンテージは高学年になっても維持される。
- ▶ **補助教員の効果:** 通常クラスに補助教員をつけても、テスト成績の向上にはほとんど寄与しないことが判明した。

4. 準実験 (Quasi-Experiments)

準実験の定義：「あたかも」ランダムな状況

- ▶ **準実験（自然実験）**：法改正、地理的境界、自然災害などの外部的な要因によって、処置の割り当てが「あたかも」ランダムに行われたかのような状況。
- ▶ **メリット**：倫理的・コスト的理由で RCT が不可能な大規模な社会問題（最低賃金、増税、教育制度）を分析できる。

階差の階差 (Differences-in-Differences, DiD)

- ▶ 処置を受けた群と受けていない群の「前後比較の差」を比較する。
- ▶ モデル:
$$Y_{it} = \beta_0 + \beta_1(Treat_i \times Post_t) + \beta_2Treat_i + \beta_3Post_t + u_{it}$$
- ▶ **平行トレンド仮定**: 処置がなければ、トリートメント群とコントロール群は同じトレンドを辿っていたはずであるという仮定。
- ▶ この仮定が成立しない場合、推定値にはバイアスが生じる。

不連続回帰 (Regression Discontinuity, RD)

- ▶ 割り当てが特定の閾値（カットオフ）に基づいている状況。
- ▶ 例: テストの点数が 60 点以上の生徒にのみ奨学金を与える。
- ▶ ロジック: 59 点の人と 61 点の人は、能力的には「ほぼ同じ」はずだが、制度によって処置の有無が峻別される。
- ▶ 閾値付近のサンプルを比較することで、ランダム化に近い状況を作り出すことができる。

5. 異質な処置効果 (Heterogeneous Effects)

処置効果の個人差

- ▶ すべての人に同じ効果 (β_1) があるとは限らない。
- ▶ 例: 職業訓練プログラムの効果は、元々のスキルが低い人には大きい、高い人には小さいかもしれない。
- ▶ 処置効果が個人 i に依存する場合 (β_{1i})、通常の OLS や IV で何を推定しているのかが問題となる。

局所平均処置効果 (LATE)

- ▶ 操作変数 (IV) 法を用いて推定されるのは、**LATE (Local Average Treatment Effect)** である。
- ▶ これは、操作変数の変化によって「自身の処置の有無を切り替えた人々」(**Compliers**) における平均的な効果である。
- ▶ 操作変数に反応しない人々 (常に処置を受ける、あるいは絶対に受けない人) の効果は推定値に含まれないことに注意が必要である。

6. 外部正当性の評価

STAR プロジェクトの一般化

- ▶ **母集団の代表性:** テネシー州の結果は、ニューヨークやカリフォルニアのような都市部でも通用するか？
- ▶ **設定の普遍性:** 1980 年代の教育環境（IT 化以前）の結果は、現在のデジタル化された教室でも有効か？
- ▶ **比較:** カリフォルニア州やマサチューセッツ州の観察データとの比較により、異なる地域でも概ね整合的な結果が得られることが示されている（外部正当性の補強）。

7. 結論

実証分析の評価フレームワーク

- ▶ **内部正当性:**
 - ▶ 割り当ては本当にランダムか？
 - ▶ 脱落者や不遵守者の影響を適切に処理しているか（IV法の活用）。
- ▶ **外部正当性:**
 - ▶ その結果は「誰に」「いつ」「どこで」適用できるのか。
- ▶ 実験と準実験は、因果関係という計量経済学の核心に迫るための最も強力なツールである。
- ▶ しかし、その結果を解釈する際には、実験のプロセスで生じた不備や、対象となった集団の特殊性を常に批判的に吟味する必要がある。