高知工科大学 経済・マネジメント学群

計量経済学応用

3. ランダム化比較試験 (RCT)

やない ゆうき
矢内 勇生

https://yukiyanai.github.io

yanai.yuki@kochi-tech.ac.jp

このトピックの目標

- ランダム化比較試験 (RCT) とは何かを理解する
- なぜ「ランダム化」するのかを理解する
  - ランダム化によって何を解決したいのか？
  - ランダム化によってどんな良いことがあるのか？
- どうやって「ランダム化」するのかを理解する

ランダム化比較試験 とは？

RCT?

RCT (randomized controlled [clinical] trial)

- Randomized Controlled Trial; ランダム [無作為] 化比較試験

  or Randomized Clinical Trial; ランダム化臨床試験

  - KUT でよく実施されている「実験」

- RCT とは (Imbens and Rubin 2015: p.40)、処置の割り付け（割り当て）が

1. 確率的 かつ
2. 実験者が制御する関数によって実行される

例：病院と健康状態

- トピック2で扱った例を考える
- 何が問題だったか
  - データを使った単純比較：病院に行った人のほうが健康状態が悪い
  - セルフセレクション：健康状態が悪い人のほうが、病院に行きやすい
- ランダム化：病院に行くかどうかを、研究者がコイントスで決める
  - 表 → 病院に行く；　裏 → 病院に行かない
  - コイントスの結果に従って病院に行った人と病院に行かなかった人の健康状態を比べることで、病院の因果効果を調べる

なぜランダム化するのか

私たちが知りたいこと

処置が2通りしかない場合： $D_i \in \{0,1\}$

- $\text{ITE}_i = Y_i(1) - Y_i(0)$

individual treatment effect, 個体処置効果

- $\text{ATE} = \mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)]$

average treatment effect, 平均処置効果

- $\text{ATT} = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1] = \mathbb{E}[Y_i(1) \mid D_i = 1] - \mathbb{E}[Y_i(0) \mid D_i = 1]$

average treatment effect for the treated, 処置群における平均処置効果

- $\text{ATC} = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 0] = \mathbb{E}[Y_i(1) \mid D_i = 0] - \mathbb{E}[Y_i(0) \mid D_i = 0]$

average treatment effect for the control, 統制群における平均処置効果

★これらの効果は観察できる？

＊ さまざまな処置効果の違いについての詳細は、例えば 岩崎 (2015) 第3章を参照

観察された平均値の単純比較

- 観察された結果：Y_i
- 観察された処置：$D_i \in \{0,1\}$
- 処置群と統制群の間の結果変数 Y_i の平均値の差：

＊ 厳密には、1行目と2行目が等しくなるには因果的一致性の仮定が必要。これは、計量経済学の授業で扱った「一致性」 とは意味が異なるので注意。

$$
\begin{aligned}
& \mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0] \\
&= \mathbb{E}[Y_i(1) \mid D_i = 1] - \mathbb{E}[Y_i(0) \mid D_i = 0] \\
&= \boxed{\mathbb{E}[Y_i(1) \mid D_i = 1] - \mathbb{E}[Y_i(0) \mid D_i = 1]} \longleftarrow \textbf{ATT} \\
&\quad + \boxed{\mathbb{E}[Y_i(0) \mid D_i = 1] - \mathbb{E}[Y_i(0) \mid D_i = 0]} \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \textbf{バイアス} \\
&= \boxed{\mathbb{E}[Y_i(1) \mid D_i = 1] - \mathbb{E}[Y_i(1) \mid D_i = 0]} \\
&\quad + \boxed{\mathbb{E}[Y_i(1) \mid D_i = 0] - \mathbb{E}[Y_i(0) \mid D_i = 0]} \longleftarrow \textbf{ATC}
\end{aligned}
$$

セレクションバイアス (selection bias)

- 潜在的結果 $Y_i(D_i)$ の期待値が、$D_i = 1$ の群と $D_i = 0$ の群の間で異なる
  - $\mathbb{E}[Y_i(0) \mid D_i = 1] > \mathbb{E}[Y_i(0) \mid D_i = 0]$：もしすべての個体が処置を受けなかったとしたら、実際には処置された個体の Y_i のほうが大きくなる
  - $\mathbb{E}[Y_i(0) \mid D_i = 1] < \mathbb{E}[Y_i(0) \mid D_i = 0]$：もしすべての個体が処置を受けなかったとしたら、実際には処置された個体の Y_i のほうが小さくなる
  - $\mathbb{E}[Y_i(1) \mid D_i = 1] > \mathbb{E}[Y_i(1) \mid D_i = 0]$：もしすべての個体が処置を受けたとしたら、実際に処置された個体の Y_i のほうが大きくなる
  - $\mathbb{E}[Y_i(1) \mid D_i = 1] < \mathbb{E}[Y_i(1) \mid D_i = 0]$：もしすべての個体が処置を受けたとしたら、実際に処置された個体の Y_i のほうが小さくなる
- 因果推論のために、セレクションバイアスを取り除きたい！

セレクションバイアスを取り除く？

- 以下の条件を満たせば、群間比較で因果効果を推定できる

1. $\mathbb{E}[Y_i(0) \mid D_i = 1] = \mathbb{E}[Y_i(0) \mid D_i = 0]$：ATT が推定可能

2. $\mathbb{E}[Y_i(1) \mid D_i = 1] = \mathbb{E}[Y_i(1) \mid D_i = 0]$：ATC が推定可能

3. 1と2の両者を満たす： ATE が推定可能

★ 1と2を実現する方法を考えよう！

平均独立ならセレクションバイアスはない

- **平均独立** (mean independence; mean exchangeability)

$$\boxed{\begin{array}{c}\mathbb{E}[Y_i(1) \mid D_i = 1] = \mathbb{E}[Y_i(1) \mid D_i = 0] \\ \text{かつ} \\ \mathbb{E}[Y_i(0) \mid D_i = 1] = \mathbb{E}[Y_i(0) \mid D_i = 0]\end{array}}$$

＊ 平均独立についての詳細は、例えば Hernán and Robins (2020) chs. 2 & 3を参照

- 平均独立が成り立つとき：

$$\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0]$$

$$= \mathbb{E}[Y_i(1) \mid D_i = 1] - \mathbb{E}[Y_i(0) \mid D_i = 0]$$

$$= \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)]$$

$$= \text{ATE}$$

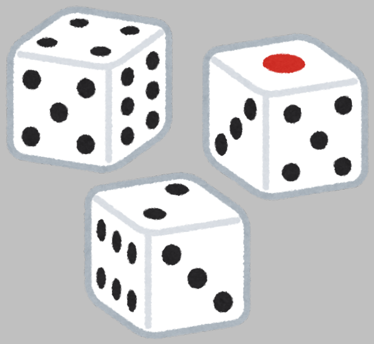
＊ 厳密には、1行目と2行目が等しくなるには因果的一致性の仮定が必要。

平均独立をどうやって確かめる？

- 平均独立が成り立っているかどうか、データから確かめられるか？
  - 確かめられない！（式に、観察できないモノが含まれる）
- よって、因果効果を推定するために平均独立を**仮定する**必要がある
  - $\mathbb{E}[Y_i(0) \mid D_i = 1] = \mathbb{E}[Y_i(0) \mid D_i = 0]$　という**仮定**が正当化できる：ATT が推定できる
  - $\mathbb{E}[Y_i(1) \mid D_i = 1] = \mathbb{E}[Y_i(1) \mid D_i = 0]$ という**仮定**が正当化できる：ATC が推定できる
    - 前者の仮定のほうが正当化しやすい場合が多いので、ATTが推定されることが多い

ただ仮定するより良い方法はないの？

ある！



Randomize!

ランダム化する (randomize)

- 処置と潜在的結果が独立 ⇒ 平均独立

$$\{Y(0), Y(1)\} \perp\!\!\!\perp D \quad \Rightarrow \quad \text{平均独立}$$

- どうやって処置と潜在的結果を独立にする？
  - 各個体に処置をランダムに割り付ける： ただし、$0 < \Pr(D_i = 1) < 1$

- D をランダムに割付けると：

$$\{Y(1), Y(0)\} \perp\!\!\!\perp D \qquad \textbf{（独立）}$$

$$\Rightarrow \begin{cases} \mathbb{E}[Y(0) \mid D = 1] = \mathbb{E}[Y(0) \mid D = 0] = \mathbb{E}[Y(0)] \\ \qquad \text{and} \\ \mathbb{E}[Y(1) \mid D = 1] = \mathbb{E}[Y(1) \mid D = 0] = \mathbb{E}[Y(1)] \end{cases} \qquad \textbf{（平均独立）}$$

$$\Rightarrow \mathbb{E}[Y(1) \mid D = 1] - \mathbb{E}[Y(0) \mid D = 0] = \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] = \text{ATE}$$

独立、交換可能性、外生性

- 独立 (independence)： $\{Y(0), Y(1)\} \perp\!\!\!\perp D$

- 交換可能性 (exchangeability)：
  $p(Y(0), Y(1) \mid D = 1) = p(Y(0), Y(1) \mid D = 0) = p(Y(0), Y(1))$
  - 処置群 ($D = 1$) に割り付けたはずの個体すべてに $D = 0$ を、統制群 ($D = 0$)に割付けたはずの個体すべてに $D = 1$ を与えても、予定どおりに処置を与えた場合と結果が同じ
  - 処置群と統制群は「群として」同じ：処置群と統制群を交換しても問題ない

- 処置群と統制群が交換可能なとき、その処置 D は外生 (exogenous) 変数
  - 潜在的変数の組 $\{Y(0), Y(1)\}$ と独立な D は外生性 (exogeneity) をもつ

独立と平均独立

- 独立（交換可能性）のほうが、平均独立より強い（厳しい）条件
  - 独立：母集団において2群の分布が同じ
  - 平均独立：2群の期待値が同じであれば、分布は異なっても良い（たとえば、分散は異なっても良い）
- ATEの推定には「平均独立」で十分なのに、より厳しい「独立」という性質を利用するのはなぜ？
  - 独立：処置を**ランダム化**すればOK

    ➡ 平均独立：**仮定**おかなくても「独立」から導かれる

何の独立を達成したいか

- 求めていない独立：$Y \perp\!\!\!\perp D$
  - $Y \perp\!\!\!\perp D$ なら $\mathbb{E}[Y_i \mid D_i = 1] = \mathbb{E}[Y_i \mid D_i = 0]$ となる：2群間に結果の差がない
  - つまり、$Y \perp\!\!\!\perp D$ は、 D とY に相関がないという意味
- 求めている独立：$\{Y(1), Y(0)\} \perp\!\!\!\perp D$
  - $\{Y(1), Y(0)\} \perp\!\!\!\perp D$ なら、 D とY の相関は因果効果：D が Y に与える処置効果

「ランダム化」は1度で良いのか？

- 1度のRCTでは、独立性は保証されない
  - 例：コイントスで通院を決めた結果、健康な人はほとんど病院に行かず、不健康な人ばかりが病院に行くと言う割り付けも、**偶然に**起こりうる
    - その場合、処置群と統制群は大きく異なる：交換可能ではない
- ランダム化による独立性：RCTを**何度も何度も繰り返すなら、平均的には**2群が交換可能
  - **1回（2回、3回）しか行われない実験で、2群が交換可能である保証はない**

大数の法則を利用する

- 「統計学」で勉強したことを思い出そう
- 大数の法則 (LLN)：標本サイズN が大きくなればなるほど、標本平均は母平均に近づく
  - 被験者の数（標本サイズ）が十分大きければ、各群の平均値は母平均に十分近いはず
  - ランダム化により、2群の母集団分布は同じ：母平均も同じ
  - よって、被験者の数が十分大きければ、平均独立は満たされる（だろうと信じやすい）

十分大きいとは？

- 「十分大きい」と言える標本サイズは、問題によって異なる
    - 大きな処置効果を識別したい：比較的小さいサイズでOK
    - 小さな処置効果を識別したい：比較的大きなサイズが必要
- 検出力分析 (power analysis) で標本サイズを決めることが必要 (参考：モファット 2018: 第2章)
    - 十分大きくないといけないが、**大きくすれば良いというものでもない**
        - 詳しくは「実験デザイン」の授業で

バランスチェック

- 実践的には、「バランスチェック」で平均独立と言えそうかどうか判断する
- バランスチェック
  - 結果変数（応答変数）と相関がありそうな複数の変数の平均値が、2群で等しいことを確認する
  - バランスチェックの問題点
    1. 測定していない（観測できない）変数のバランスはわからない
    2. 「違いがない（バランスしている）」ことを示したいが、検定で積極的に示せるのは「違いがある（バランスしていない）」ことのみ

処置をどのように
ランダム化するか

RCTとその目的

- RCTのポイント: 処置をランダムに割り付ける
- 目的：各群を交換可能にする（処置と潜在的結果の独立性）
  - 各群の特徴を表す変数の確率分布を同じにする
    - 大数の法則を利用して、平均独立を実現する
      - 因果効果を推定可能にする

どうやってランダム化する？

- 実験者が事前に決めた方法で、確率的に処置を割り付ける
- たとえば、
  - コイントス：表が出たら処置、裏が出たら統制
  - サイコロ：1の目が出たら処置、その他の目は統制
  - 実験室に到着した順に：処置、 統制、処置、統制、…
  - 誕生日の日付が奇数なら処置、偶数なら統制

  etc.

局所管理 (local control)

- 社会科学の研究： 個体が均質ではない
- できるだけ同質的な個体のブロック（あるいは層）を作り、ブロック内でランダム化した方がよい
  - **ブロッキング** (blocking)、 層別化 (stratification)
    - ブロッキングにより、推定の精度が高まる
    - ブロッキングしなくても ATE は推定できるが、ブロッキングできるならしたほうが良い

ブロッキングの例

- 男女50人ずつ、計100人の被験者に処置を割付ける
  - ブロッキングせず、ランダムに50人を処置群に割り付ける
    - 最悪の場合、男だけの群と、女だけの群ができる
    - そこまで極端でなくても、男女比にばらつきが出ると、性別による差が生じてしまう
  - 性別でブロッキングし、各**ブロック内**でランダムに半数を処置群に割付ける
    - ブロックした要素（性別）について、処置群と統制群の違いがなくなる

基本的なランダム化法

1. ベルヌーイ実験 (Bernoulli trial for each unit): 各個体について、ベルヌーイ試行（コイントス）を行って処置するかどうか決める

2. 完全ランダム化実験 (completely randomized experiment)：N 人から処置群に割当てる n_1 人をランダムに選び、残りを統制

3. ブロック別（層別）ランダム化実験 (stratified randomized experiment)：あらかじめ決めた各ブロック内で、完全ランダム化実験を行う — 全体の平均処置効果は、ブロックサイズを重みにした加重平均

4. ペア毎のランダム化実験 (paired randomized experiment)：個体をできるだけ同質なペアに分け、各ペア内でランダムに選んだ一方を処置（3の特殊ケースである）

各方法におけるパタンの数

男女50人ずつ、計100人の被験者を、処置群と統制群に割り付ける

1. ベルヌーイ実験： $2^{100} \approx 1.27 \times 10^{30}$

2. 完全ランダム化実験（100人から処置群に割り付ける50人をランダムに選ぶとする）： $\binom{100}{50} \approx 1.01 \times 10^{29}$

3. ブロック別ランダム化実験（性別でブロッキングし、各ブロック内で半数を処置群に）： $\binom{50}{25} \cdot \binom{50}{25} \approx 1.60 \times 10^{28}$

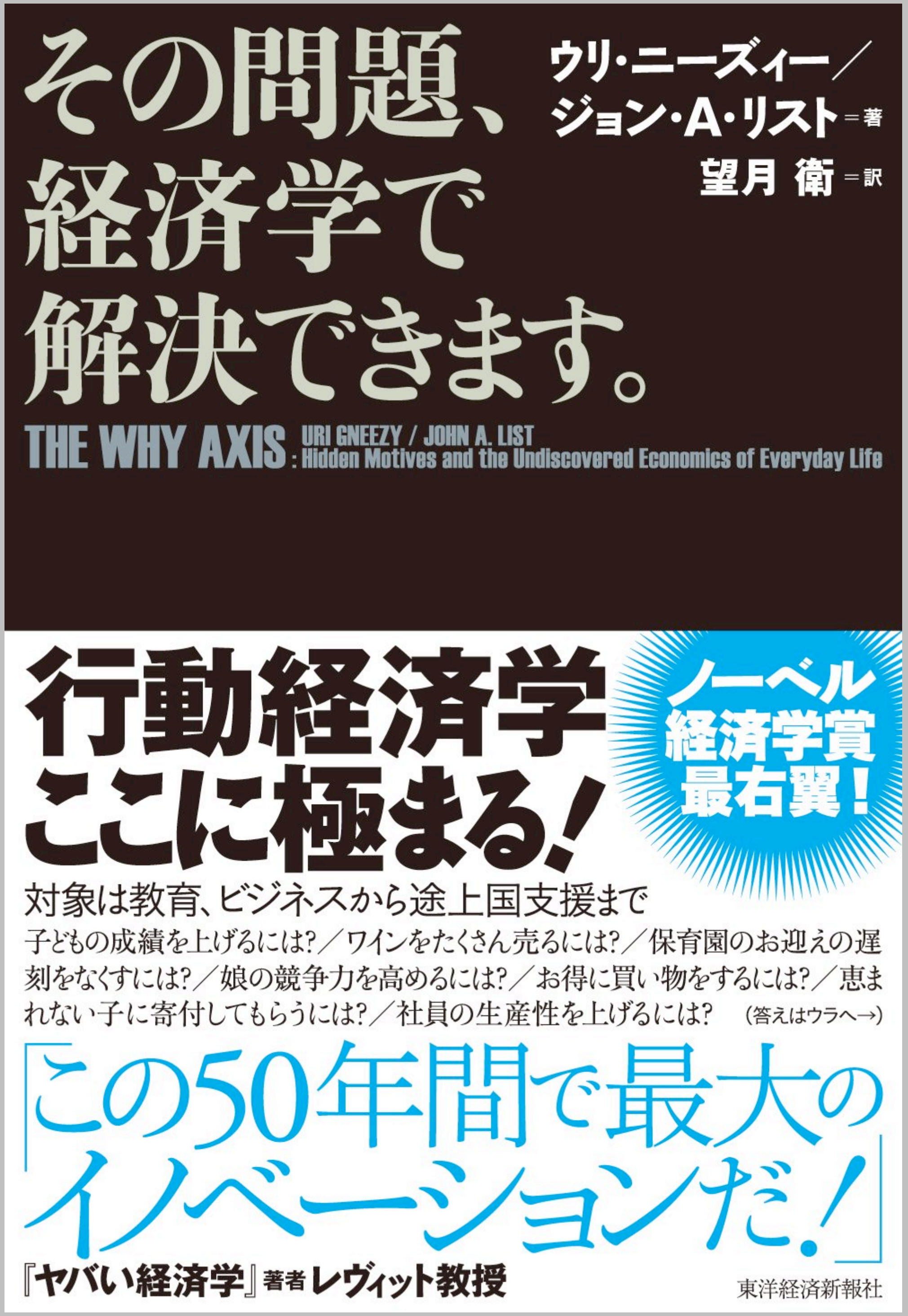
4. ペア毎のランダム化実験： $2^{50} \approx 1.13 \times 10^{15}$

- 処置の割り付けパタンは1が最も多く、4が最も少ない：4による因果効果の推定が、最も精度が高い

トピック3のまとめ

- ランダム化比較試験（RCT）
  - 処置をランダムに割り付け、平均処置効果を推定する
- RCTによって、セレクションバイアスを取り除くことができる
  - ただし、1つひとつの実験で平均処置効果が正しく推定できる保証はない
    - 大数の法則を利用する：標本サイズを「十分大きく」する
- ランダム化の方法は色々ある：ブロッキングしたほうが良い

参考文献：RCTの利用例



RCTへの疑問

- どんな処置でもランダム化していいのか？
  - 病院に行くかどうか、研究者がコイントスで決めていいのか？
  - どんな処置（介入）を与えてもいいのか？
- ランダム化できない問題もあるのでは？
  - RCT ができない問題は研究できない・すべきでないのか？
  - 実験外の観察からしか得られない情報（データ）もあるのでは？
  - 実験外のデータを使わないのはもったいないのでは？

次回予告

4. 回帰分析