高知工科大学 経済・マネジメント学群

計量経済学応用

5. 傾向スコア

やない ゆうき
矢内 勇生

https://yukiyanai.github.io

yanai.yuki@kochi-tech.ac.jp

分析計画の発表について

- 因果関係に関する問いを考える
  - ある特定の原因がもつ**因果効果**についての仮説を考える
  - プロジェクト研究に関連するテーマがおすすめ（指導教員に相談を）
- 問題によって、使うべき手法は異なりうる
- 分析するデータが必要なので、早めに考えはじめる

今日の目標

傾向スコアを使った因果推論の方法を身につける

- 傾向スコアとは何か？
  - 傾向スコアが因果推論に役立つのはなぜか？
  - 傾向スコアの利点とは何か？
- 傾向スコアをどのように利用するか？
  - 傾向スコアをどうやって推定するか？
  - 推定した傾向スコアをどのように使うか？

傾向スコアとは何か？

傾向スコア (propensity score)

傾向スコア $e_i(X_i)$ とは

- 観測された共変量 $X_i = (X_{1i}, X_{2i}, \ldots, X_{ki})$ で条件付けた、処置される$(D_i = 1)$ 確率 (Rosenbaum and Rubin 1983)

$$\boxed{e_i(X_i) = \Pr(D_i = 1 \mid X_i)}$$

ただし、$(0 \leq e_i(X_i) \leq 1)$

傾向スコアの表記についての注意

- 安井 (2020) は傾向スコアを

$$P(X_i)$$

と記しているが、傾向スコアは

$$\Pr(X_i = x)$$

ではないので注意!

.傾向スコアは、$\Pr(D_i = 1 \mid X_i)$

何が問題なのか？

- RCT だと群間比較で因果推論できるのに、調査・観察データだとできない理由
  - RCT：処置群と統制群が交換可能
  - 調査・観察データ：処置群と統制群が交換不可
    - セレクションバイアスがある
      - 解消法：処置群と統制群が交換可能になるように調整する

重回帰によるセレクションバイアスの解消

- 共変量 X の値で条件付ける
- 観測された共変量のみがセレクションの原因なら（つまり、強い意味での無視可能性が成り立つなら）
  - 処置群と統制群で共変量の値が同じ個体を2群間で比較すれば、共変量が特定の値をとる場合の処置効果が推定できる

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

  - 全体の処置効果は共変量ごとの期待値の期待値：重回帰の「傾き」

$$\mathbb{E}\left[\mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x]\right] = \mathbb{E}[Y_i(1) - Y_i(0)] = \text{ATE}$$

共変量の数が多い場合

- 共変量 $X_1, X_2, \ldots, X_k$ の値で条件付ける
- 観測された共変量のみがセレクションの原因なら（つまり、強い意味での無視可能性が成り立つなら）
  - 処置群と統制群で共変量の値が同じ個体を2群間で比較すれば、共変量が特定の値をとる場合の処置効果が推定できる

$$\mathbb{E}[Y_i(1) \mid D_i = 1, X_{1i}, X_{2i} \ldots, X_{ki}] - \mathbb{E}[Y_i(0) \mid D_i = 0, X_{1i}, X_{2i} \ldots, X_{ki}]$$

$$= \mathbb{E}[Y_i(1) - Y_i(0) \mid X_{1i}, X_{2i} \ldots, X_{ki}]$$

  - 全体の処置効果は共変量ごとの期待値の期待値：重回帰の「傾き」

$$\mathbb{E}_{X_1} \cdots \mathbb{E}_{X_k} \left[\mathbb{E}[Y_i(1) - Y_i(0) \mid X_{1i}, \ldots, X_{ki}]\right] = \mathbb{E}[Y_i(1) - Y_i(0)] = \mathrm{ATE}$$

共変量の数が多いときの問題

- サンプルサイズ（観測数; N）が小さいと、比較可能な個体がないかもしれない
- 例：共変量が $X_1, \ldots, X_{10}$ で、すべて二値変数の場合
  - 共変量の値のパタン：$2^{10} = 1024$ 通り
  - 観測数が $2 \times 1024 = 2048$ 以上ないと、すべてのパタンで処置 $(D = 1)$ と統制 $(D = 0)$ の比較ができない
- 共変量の数が増えると、重回帰が難しくなる
  - 共変量は、二値変数とは限らない（連続型変数かも）
  - 共変量の数は、もっと多いかもしれない（$k > N$ かも）
  - 結果変数と共変量の関係を、正しく定式化できないかも

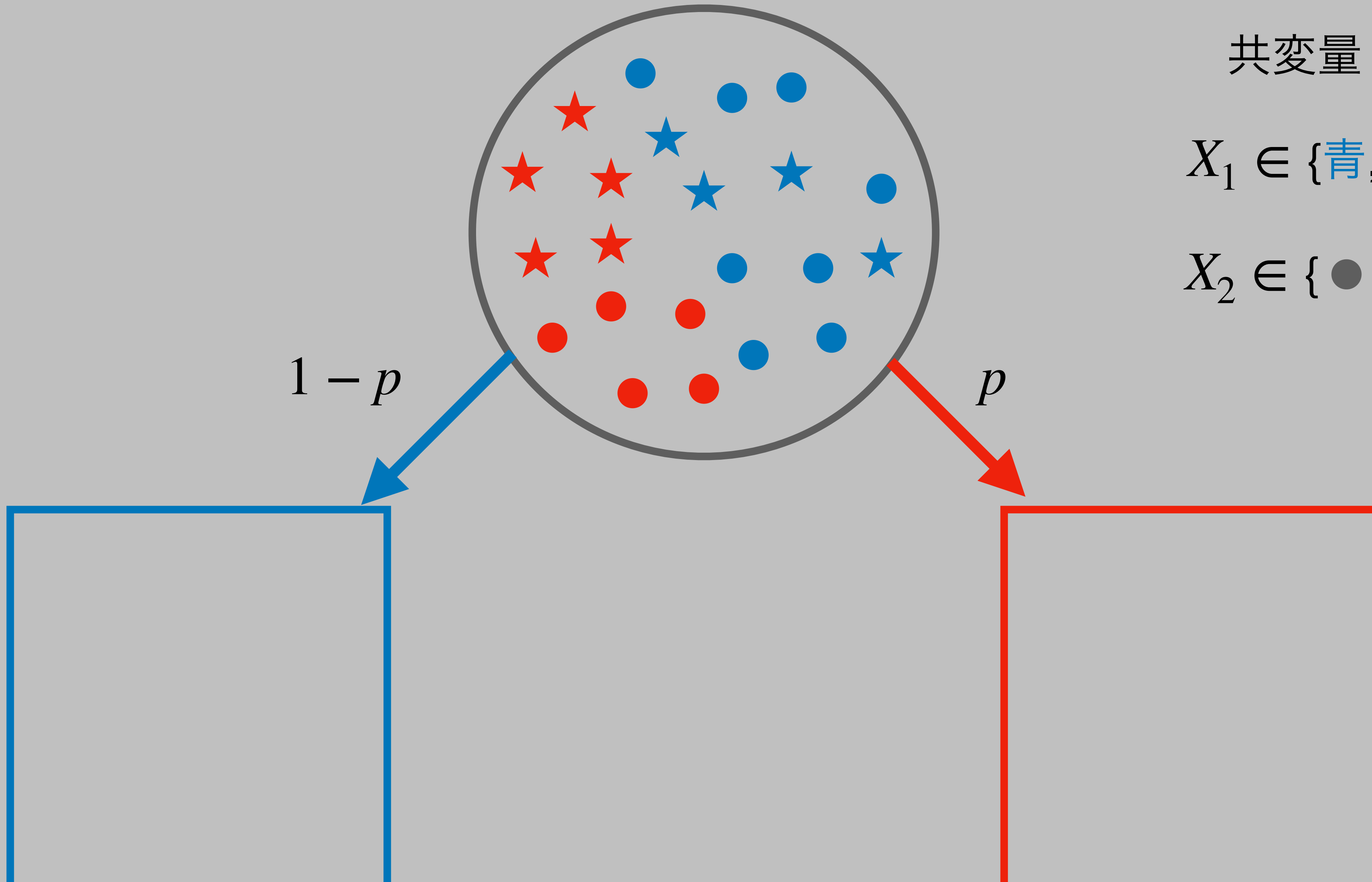
セレクションバイアスの問題点

- セレクションバイアス (SB) の何が問題だったか？
  - 処置を受ける確率が、共変量の値に依存して決まる
    - 共変量の値によって処置を受けやすい個体と、処置を受けにくい個体がいる
    - 例：RQ「病院に行くと、健康になるか？」
      - SB：元々不健康な人ほど病院に行きやすい
    - 例：RQ「計量経済学を受講すると、就職で有利か？」
      - SB：優秀な学生ほど計量経済学を受講しやすい

RCTの場合

- 個体が処置を受ける確率 p は、処置群と統制群で等しい

(　$p = 0.5$ という意味ではないので注意)



共変量

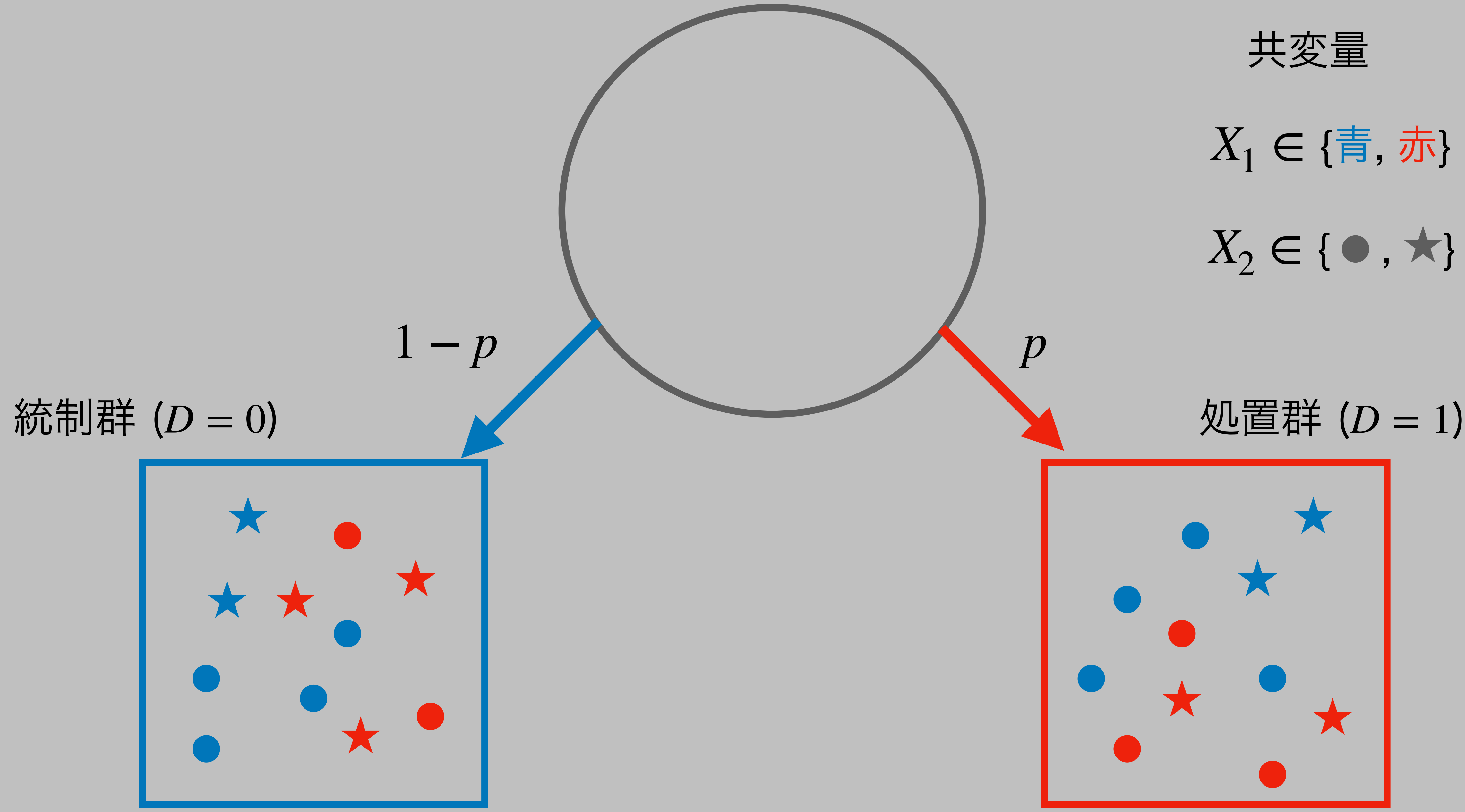
$X_1 \in \{$青, 赤$\}$

$X_2 \in \{$●, ★$\}$

RCTの場合

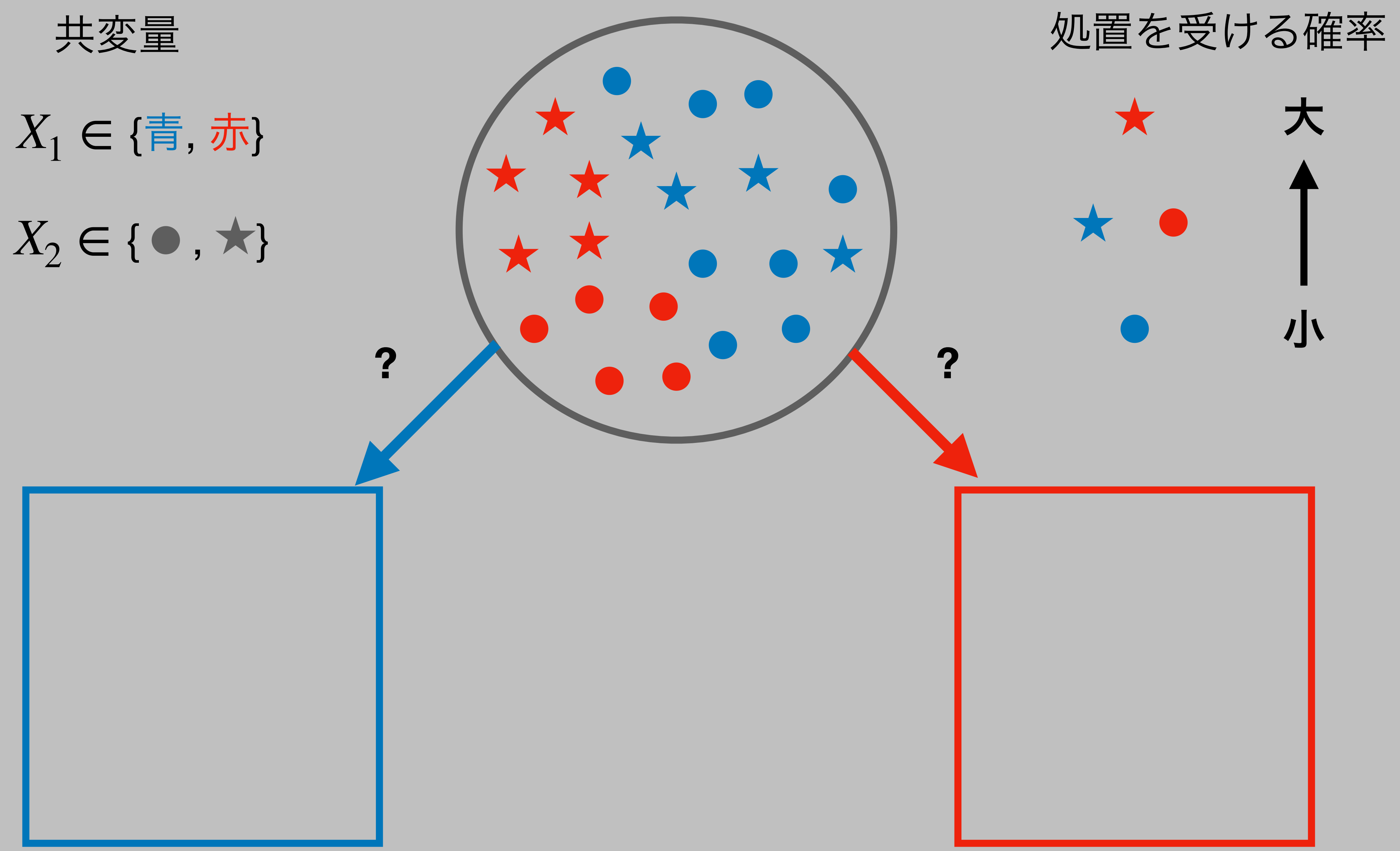
- 個体が処置を受ける確率 p は、処置群と統制群で等しい

($p = 0.5$ という意味ではないので注意)



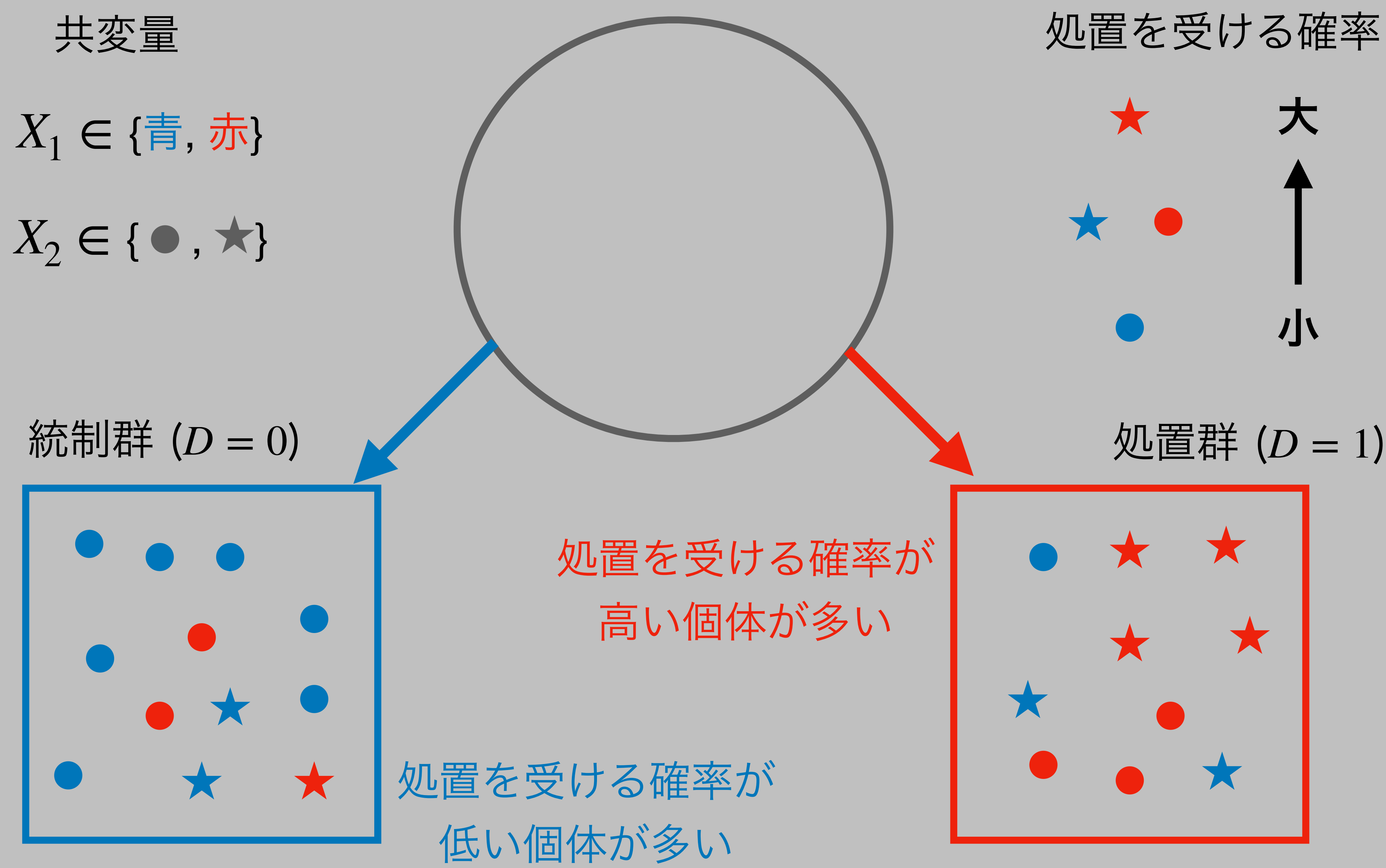
調査・観察データの場合

- 個体が処置を受ける確率が、処置群と統制群で異なる



調査・観察データの場合

- 個体が処置を受ける確率が、処置群と統制群で異なる



調査・観察データの処置群と統制群

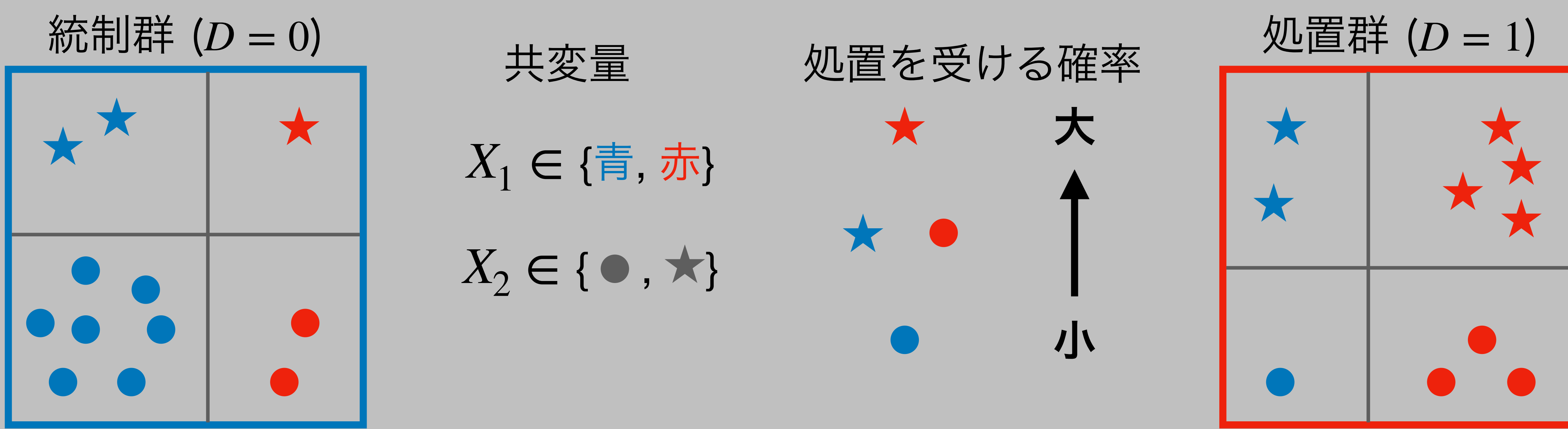
- 調査・観察データにおける処置群と統制群の違い
  - 処置を受ける確率が異なる
    - 処置群：処置を受ける確率が**高い**個体が「多い」集団
    - 統制群：処置を受ける確率が**低い**個体が「多い」集団

共変量の数が少ない場合

- 共変量によって処置を受ける確率が変わる
  - 共変量の値に条件付けて分析する

処置を受ける確率が
低い個体が多い集合

処置を受ける確率が
高い個体が多い集合

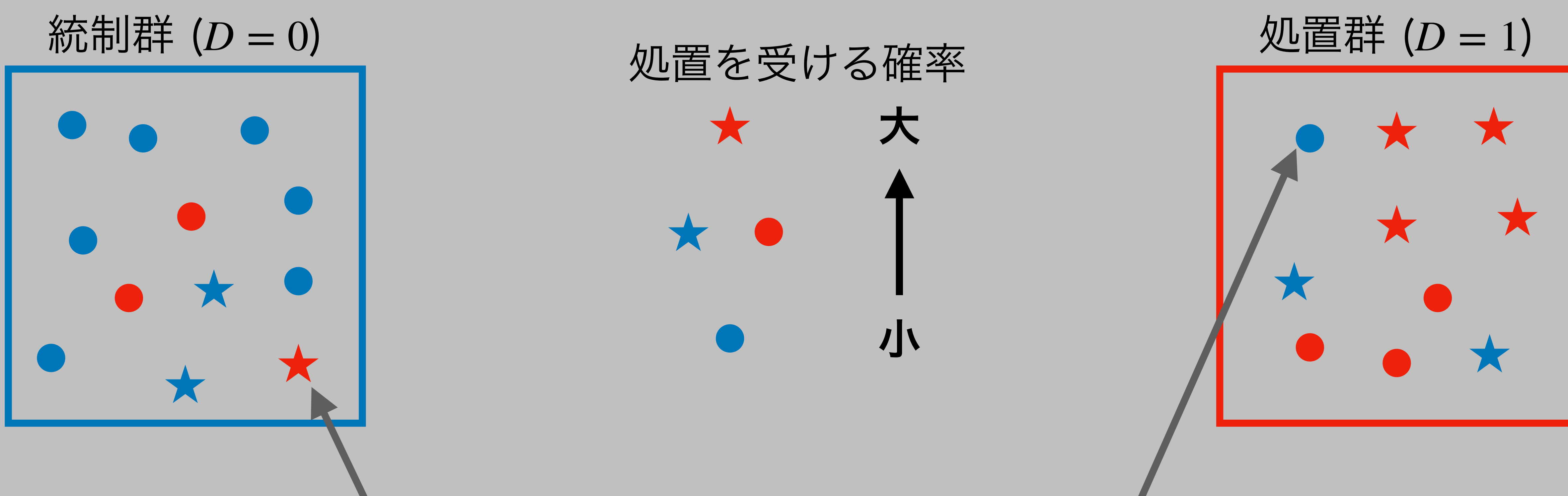


共変量の数が多い場合

- 共変量によって処置を受ける確率が変わる
  - 処置を受ける確率を揃えて比較をすればいいのでは？

処置を受ける確率が
低い個体が多い集合

処置を受ける確率が
高い個体が多い集合



処置を受ける確率が高いのに、
統制群にいる個体：
処置群の比較対象として有力

処置を受ける確率が低いのに、
処置群にいる個体：
統制群の比較対象として有力

傾向スコア (propensity score)

傾向スコア $e(X)$ とは

- 観測された共変量 $X = (X_1, X_2, \ldots, X_k)$ で条件付けた、処置される($D = 1$) 確率 (Rosenbaum and Rubin 1983)

$$\boxed{e_i(X_i) = \Pr(D_i = 1 \mid X_i)}$$

ただし、$(0 \leq e_i(X_i) \leq 1)$

★ 傾向スコアが同じ（近い）もの同士を処置群と統制群で比較すれば、因果効果が推定できる

傾向スコアによる条件付け

- 強い意味での無視可能性が**仮定**できるなら、すべての共変量で条件付けを行う代わりに、$e(X)$ による条件付けでセレクションバイアスを除去できる

$$\{Y(0), Y(1)\} \perp\!\!\!\perp D \mid X$$

$$\Rightarrow \{Y(0), Y(1)\} \perp\!\!\!\perp D \mid e(X) \quad \text{かつ } D \perp\!\!\!\perp X \mid e(X)$$

傾向スコアを使ってATEを推定する

$$\{Y(0), Y(1)\} \perp\!\!\!\perp D \mid e(X)$$

$$\Rightarrow \begin{cases} \mathbb{E}[Y(1) \mid e(X), D=1] = \mathbb{E}[Y(1) \mid e(X)] \\ \quad \text{and} \\ \mathbb{E}[Y(0) \mid e(X), D=0] = \mathbb{E}[Y(0) \mid e(X)] \end{cases}$$

$$\begin{aligned} \text{ATE} &= \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] \\ &= \boxed{\mathbb{E}}\left[\mathbb{E}[Y(1) \mid e(X)] - \mathbb{E}[Y(0) \mid e(X)]\right] \\ &= \mathbb{E}\left[\boxed{\mathbb{E}[Y(1) \mid e(X), D=1]} - \boxed{\mathbb{E}[Y(0) \mid e(X), D=0]}\right] \end{aligned}$$

傾向スコアによる条件付けを消去する期待値

傾向スコアが$e(X)$ で、$D=1$ のときに観測される結果の平均値

傾向スコアが$e(X)$ で、$D=0$ のときに観測される結果の平均値

傾向スコアを使うメリット

- 次元を縮約できる
  - 共変量が $X = (X_1, X_2, \ldots, X_k)$：$k$ 次元
  - $e(X)$：1次元
- 結果変数と共変量の回帰モデルを仮定する必要がない
  - 詳細については、星野（2009）の 第3章を参照

傾向スコアを利用した因果推論

傾向スコアを使う手順

1. 傾向スコアの推定に使う変数の選定
2. 傾向スコアの推定
3. 傾向スコアを用いたバランス調整
    - 重み付け
    - 層別
    - マッチング*：傾向スコアによるマッチングは非推奨（King & Nielsen 2019）
4. 共変量のバランスチェック
5. 処置効果（因果効果）の推定
6. 感度分析*：この授業では説明しない（参考：Carnegie et al. 2016）

傾向スコアは観察できない

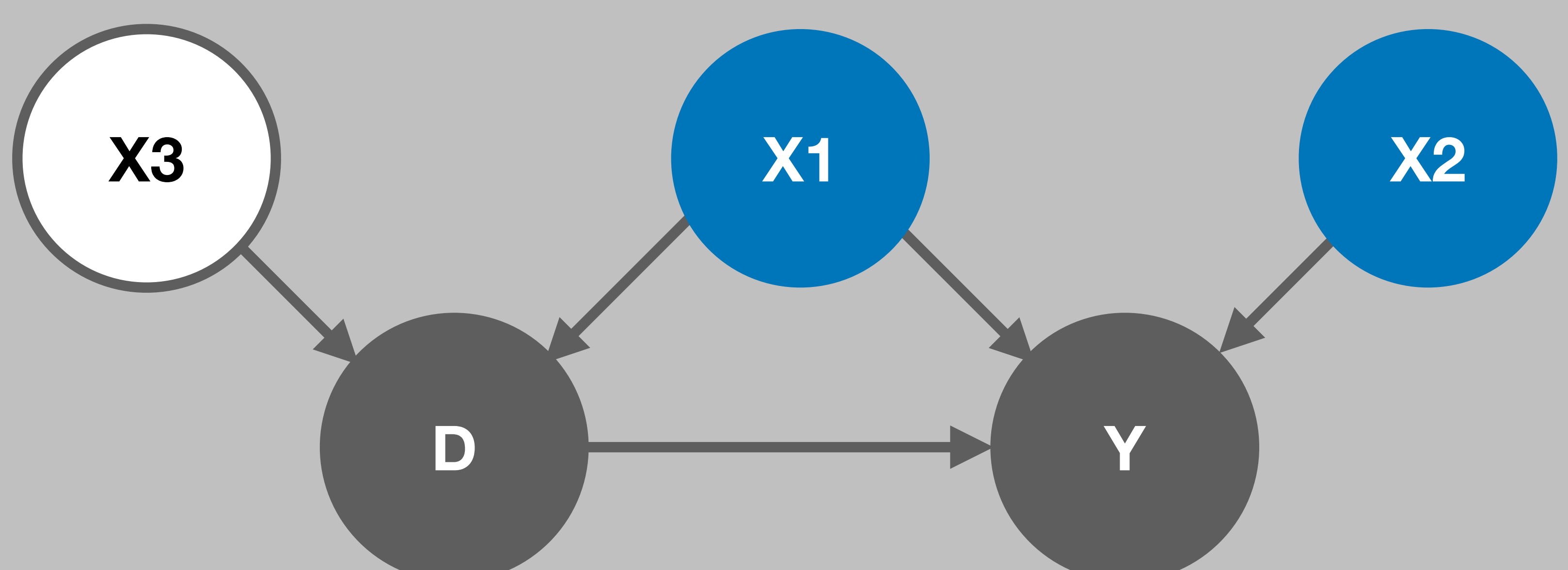
- $e_i(X_i) = \Pr(D_i = 1 \mid X_i)$：観察できない
- 観察できるのは
  - D_i
  - $X_i = (X_{1i}, X_{2i}, \ldots, X_{ki})$
- 観察できるものを使って、傾向スコアを推定する

1. 共変量の選定 (1)

- 傾向スコアの推定に使う共変量を選ぶ
  - 交絡因子：絶対に使う！
    - 使わないとバイアスが除去できない
  - 処置には影響しないが、結果に影響するもの：使ったほうが良い
    - 使わなくてもバイアスは生じないが、使うと処置効果の推定精度が向上する
  - 処置には影響するが、結果には影響しないもの：使わないほうが良い
    - 使ってもバイアスは生じないが、処置効果の推定精度が落ちる
  - その他の変数は、重回帰と同じように考える
    - 参考：[Brookhart et al. 2006]

1. 共変量の選定 (2)

- 傾向スコアの推定に使う共変量の選定
  - X1 に該当するもの：絶対使う
  - X2に該当するもの：使ったほうが良い
  - X3に該当するもの：使わないほうが良い
  - **処置後変数：使わない**
- X1, X2, X3 に該当する変数はそれぞれ複数ありうる
- 各分野（経済学, 経営学, etc.）の**専門知識をいかして見極める**



2. 傾向スコアの推定

- 推定方法は色々ある
  - 一般化線形モデル (GLM)
    - **ロジットモデル（ロジスティック回帰）**
    - プロビットモデル
  - 機械学習の手法
    - ランダムフォレスト (random forest)
    - ブースティング (boosting)
- この講義では処置が二値の場合について説明する。処置が二値でない場合については、Hirano & Imbens (2014) を参照

ロジスティック回帰

- 結果変数が二値 (0 or 1) の場合に使う回帰の1つ
- 一般化線形モデル (generalized linear model; GLM)
- 傾向スコアの推定では、処置を結果変数として扱う
  - Rでの推定方法は補助教材を参照
    - 詳しくは、浅野・矢内 (2018) 第15章を参照

ロジスティック回帰モデル

$$D_i \sim \text{Bernoulli}(\theta_i)$$

$$\text{logit}(\theta_i) = \alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_k X_{ik}$$

- θ_i：ベルヌーイ試行の成功確率
  - 傾向スコアとして知りたいのは θ_i
- α_j（$j = 0,1,\ldots,k$）を推定する
  - 通常のロジスティック回帰では α を知りたいが、傾向スコアの推定では興味の対象ではない

ロジスティック回帰のリンク関数

$$D_i \sim \text{Bernoulli}(\theta_i)$$

リンク関数

線形予測子 (linear predictor)

$$\boxed{\text{logit}(\theta_i)} = \boxed{\alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_k X_{ik}}$$

- 結果変数を生成するパラメタと線形予測子を結びつける関数をリンク関数という
  - $0 < \theta_i < 1$
  - $-\infty < \alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_k X_{ik} < \infty$
- ロジスティック回帰のリンク関数：ロジット (logit)

ロジット*

$$\text{logit}(\theta_i) = \log\left(\frac{\theta_i}{1-\theta_i}\right) = \alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_k X_{ki}$$

$$\Leftrightarrow \exp\left[\log\left(\frac{\theta_i}{1-\theta_i}\right)\right] = \frac{\theta_i}{1-\theta_i} = \exp[\alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_k X_{ki}]$$

$$\Leftrightarrow \theta_i = (1-\theta_i)\exp[\alpha_0 + \alpha_1 X_{1i} + \cdots \alpha_k X_{ki}]$$

$$\Leftrightarrow \theta_i \left(\exp[\alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_k X_{ki}] + 1\right) = \exp[\alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_k X_{ki}]$$

$$\Leftrightarrow \theta_i = \frac{\exp[\alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_k X_{ki}]}{\exp[\alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_k X_{ki}] + 1}$$

ロジスティック関数* (1)

$$\text{logistic}(x) = \text{logit}^{-1}(x) = \frac{\exp(x)}{\exp(x) + 1} = \frac{1}{1 + \exp(-x)}$$

- ロジスティック関数は、ロジットの逆関数

$$\text{logit}(\theta_i) = \alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_k X_{ik}$$

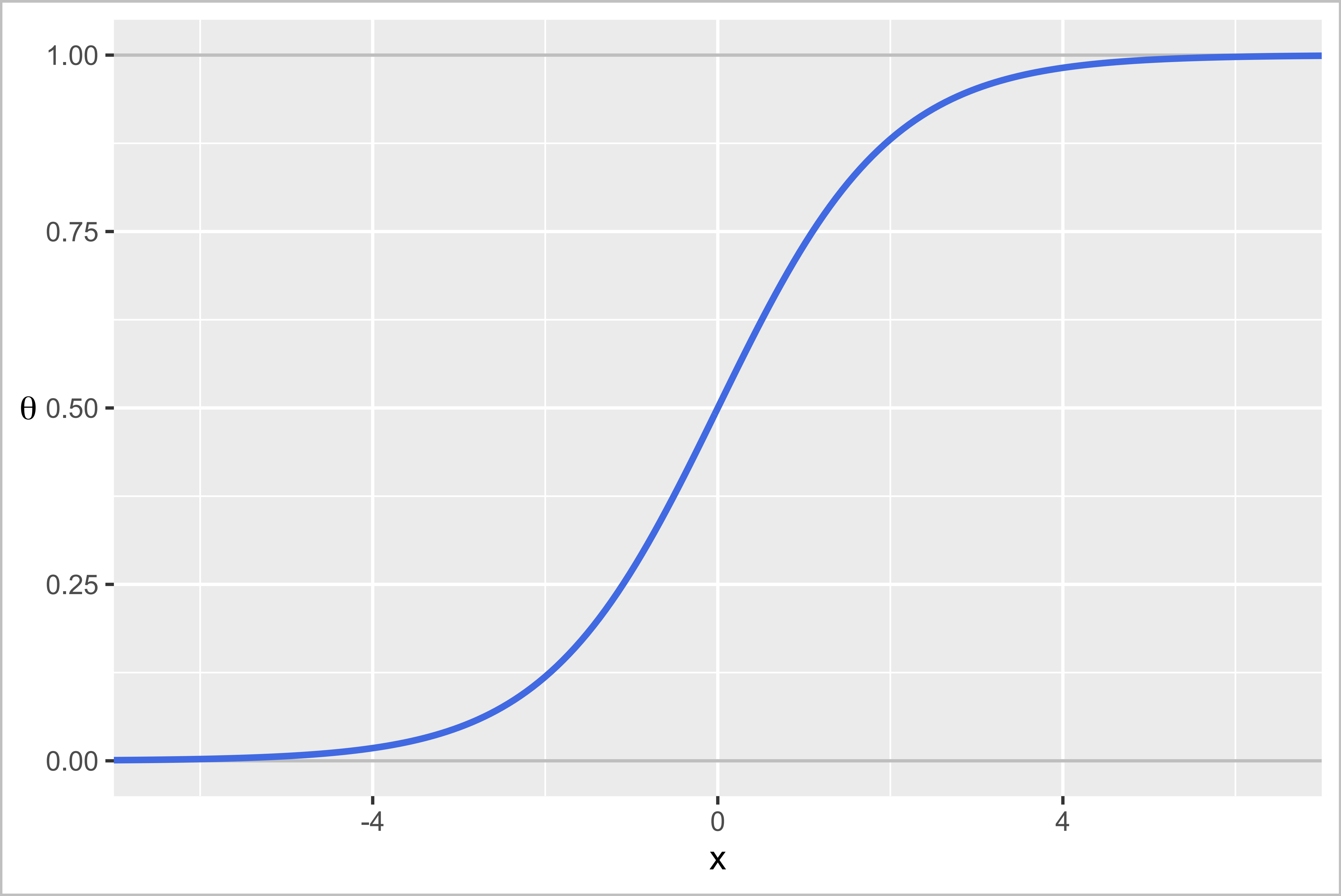
$$\Leftrightarrow \theta_i = \text{logit}^{-1}\left(\alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_k X_{ki}\right)$$

- ロジスティック回帰モデルの書き換え

$$D_i \sim \text{Bernoulli}(\theta_i)$$

$$\theta_i = \text{logit}^{-1}\left(\alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_k X_{ki}\right)$$

ロジスティック関数* (2)



3. 傾向スコアを用いたバランス調整

- I. 重み付けによる調整
  - 推定したい対象に応じた「重み (weight)」を計算し、その重みを使った加重平均で期待値を推定する
- II. 層別
  - 傾向スコアに基づいて、観測個体を複数の層に分け、それぞれの層で処置群と統制群を比較する

I. 重み付けによるバランス調整

- 傾向スコアを使って、観測された各個体に重み (weight) を与える
- **推定対象 (estimand)** によって、重みが異なる
- 主な推定対象
  - ATT（処置群における平均処置効果）

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

  - ATE（平均処置効果）

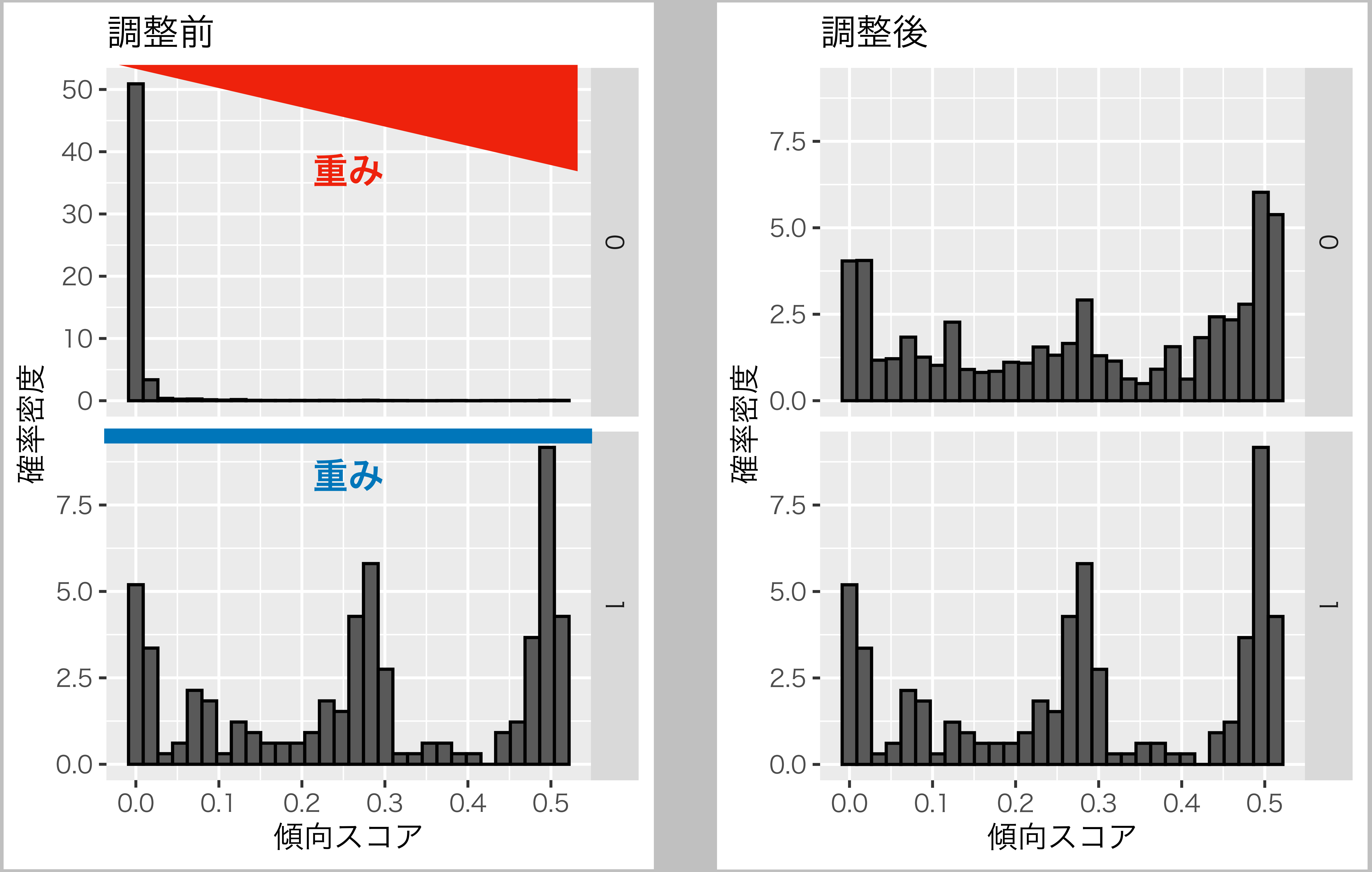
$$\text{ATE} = \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$$

ATT を推定するための重み

$$w_i^{\text{ATT}} = D_i + (1 - D_i)\frac{e_i(X)}{1 - e_i(X)}$$

- 処置群の個体の重みは 1
- 統制群の個体の重みは、傾向スコアとともに大きくなる
  - 傾向スコアが大きい：同種の個体が**処置群に**多いと思われるので、割り増し
  - 傾向スコアが小さい：同種の個体が**処置群に**少ないと思われるので、割り引き

重み付けによる調整：ATTの推定



LaLonde (1986) のデータ

ATEを推定するための重み

$$w_i^{\mathrm{ATE}} = \frac{D_i}{e_i(X)} + \frac{1 - D_i}{1 - e_i(X)}$$

- 各群に割付けられる確率（傾向スコア）の逆数で重み付け
  - 逆確率重み付け法 (inverse probability weighting; IPW)
- 処置群：傾向スコアが大きいほど、重みが小さくなる
- 統制群：傾向スコアが大きいほど、重みが大きくなる
  - 各群で、珍しい個体ほど重みが大きくなる

IPWによる因果効果の推定

- 処置が1の場合の潜在的結果の期待値の推定値

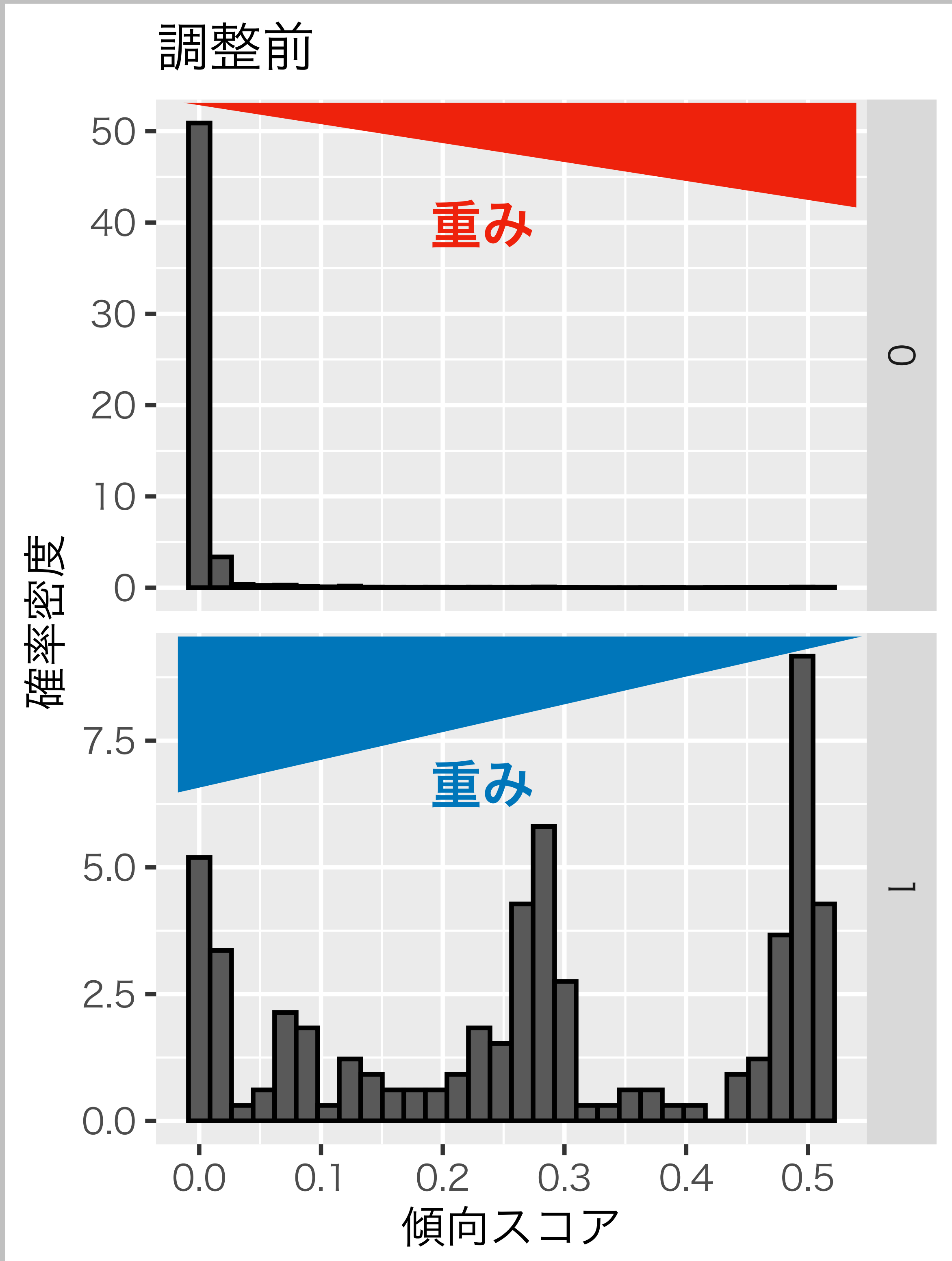
$$\hat{\mathbb{E}}[Y(1)] = \sum \frac{D_i / e_i(X)}{\sum \left[D_i / e_i(X) \right]} Y_i$$

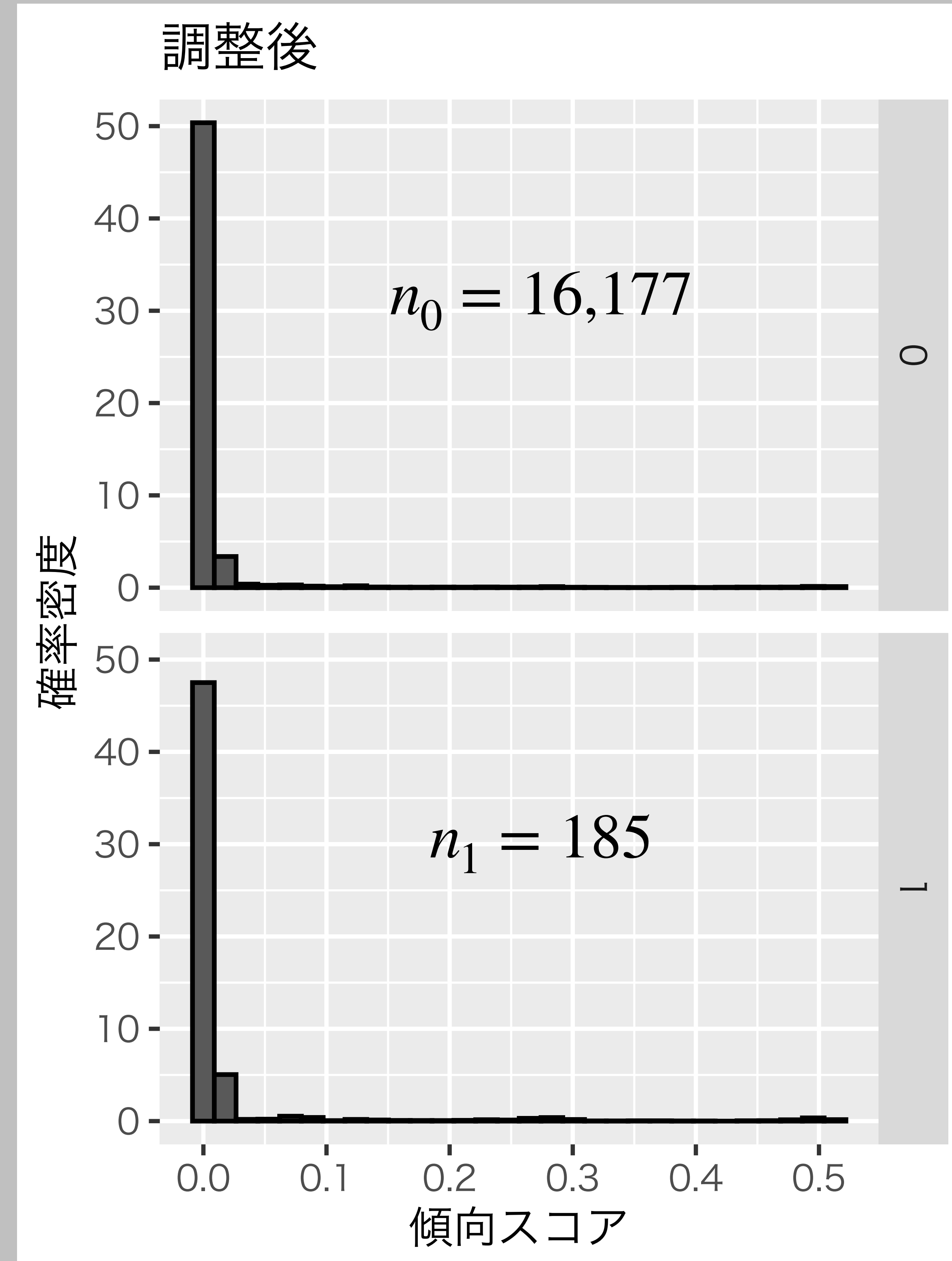
- 処置が0の場合の潜在的結果の期待値の推定値

$$\hat{\mathbb{E}}[Y(0)] = \sum \frac{(1 - D_i) / \left(1 - e_i(X)\right)}{\sum \left[(1 - D_i) / \left(1 - e_i(X)\right) \right]} Y_i$$

- これらの差で、ATE が推定できる

IPWによる調整：ATEの推定





LaLonde (1986) のデータ

II. 層別によるバランス調整

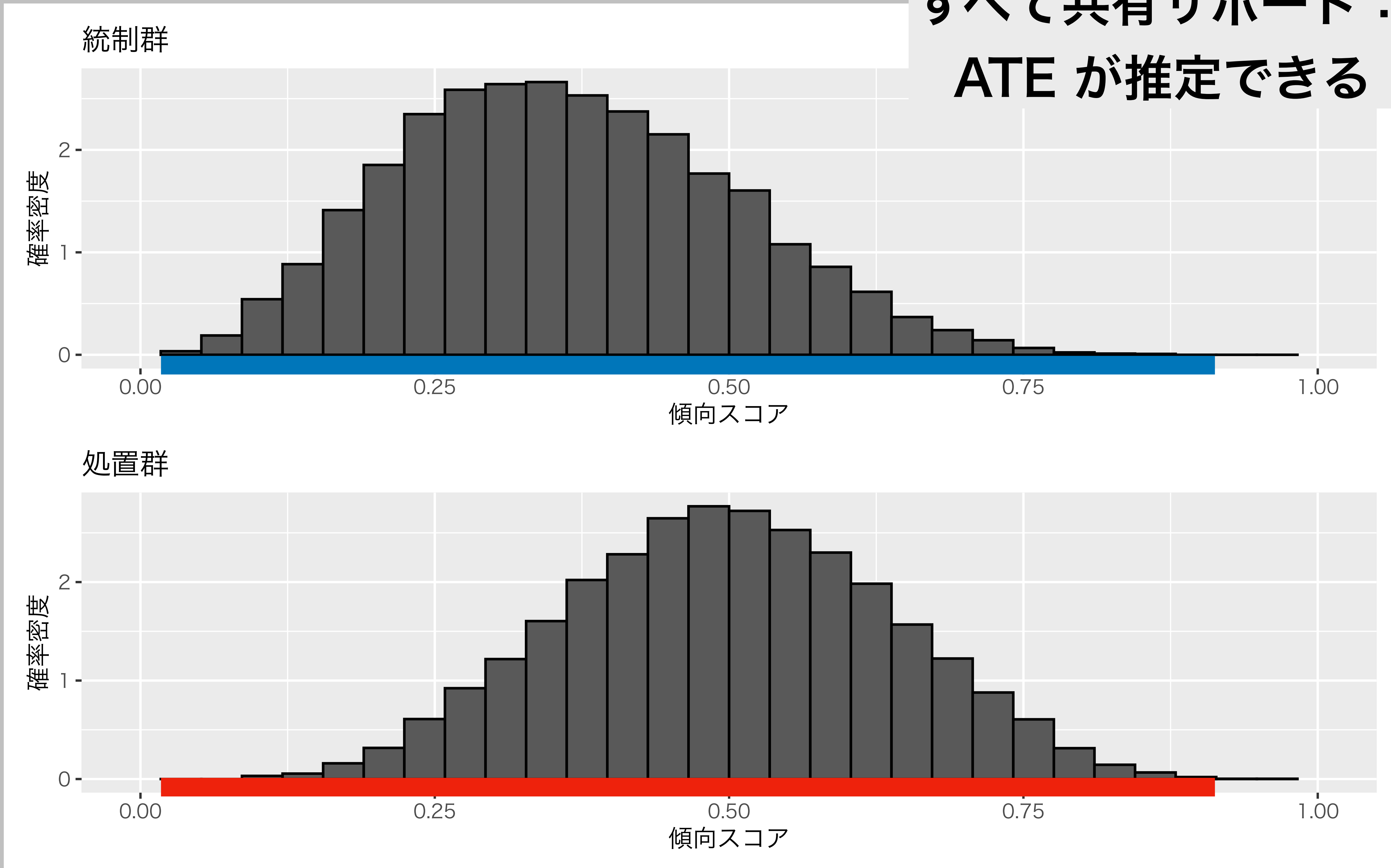
- 傾向スコアの値ごとに、観測個体を複数の層 (staratum, subclass) に分ける
  - よく使われる層の数は5
- 処置群と統制群の合計数が均等になるように分ける
  - ATEの推定
- 処置群の個体数が均等になるように分ける
  - ATTの推定

共有サポート (common support)

- 共有サポート（共通サポート）
- 処置群と統制群で、傾向スコアが同じ範囲に分布していること
- 共有サポートがない
  - 相手の群に似ている個体がないということ
  - 比較できない
  - 因果推論できない

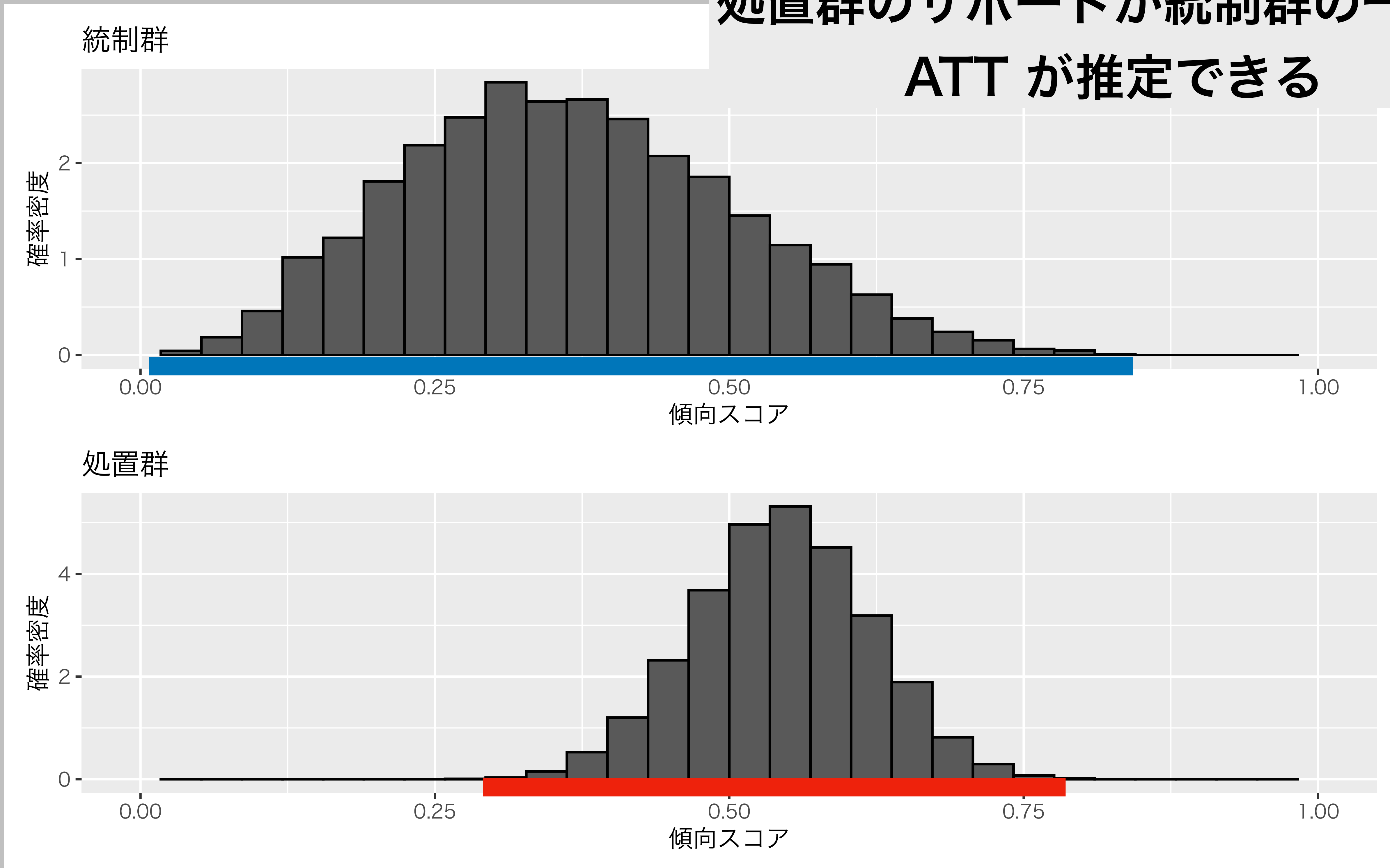
共有サポート：ケース1

すべて共有サポート：ATE が推定できる



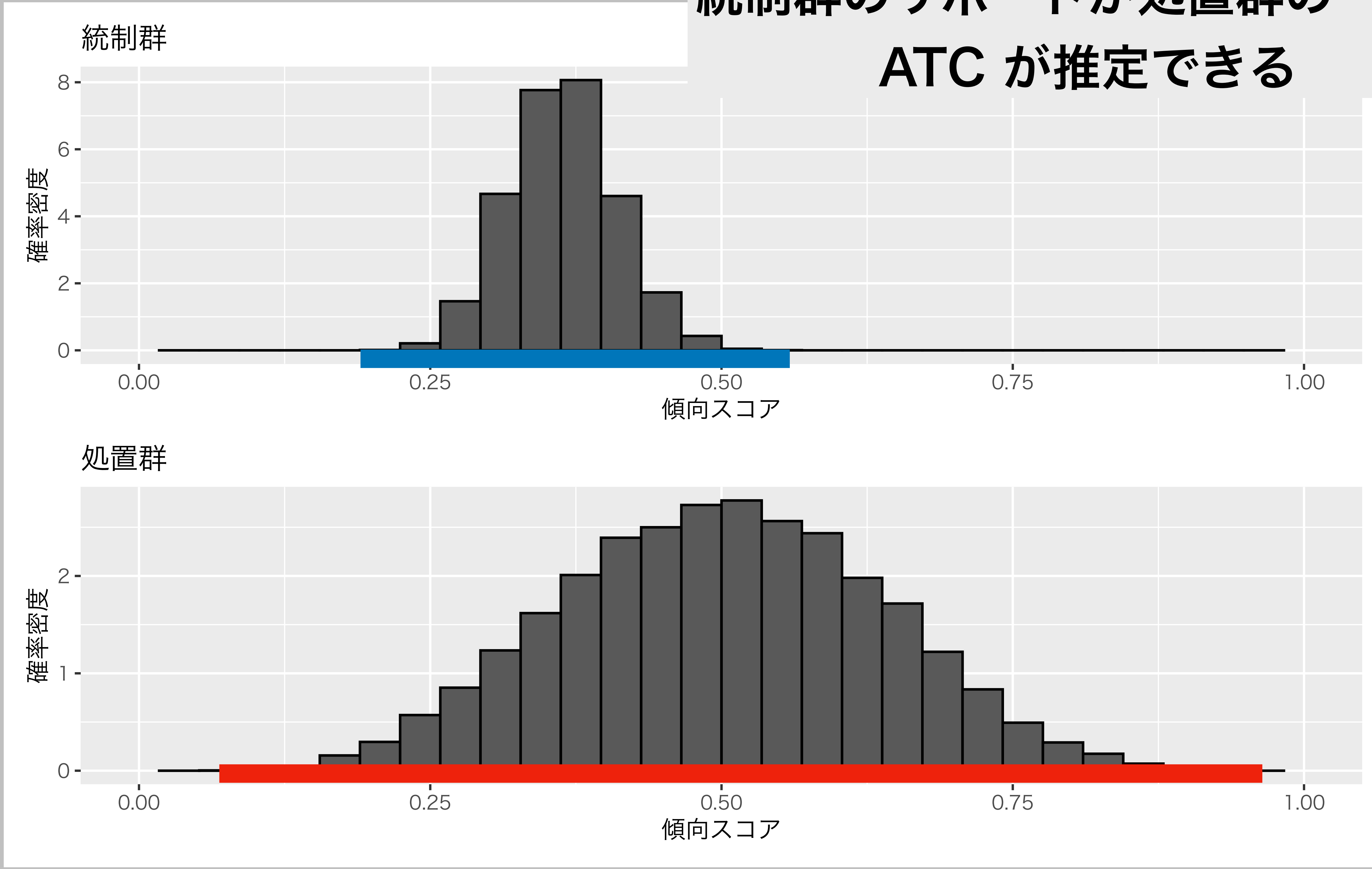
共有サポート：ケース2

処置群のサポートが統制群の一部：ATT が推定できる

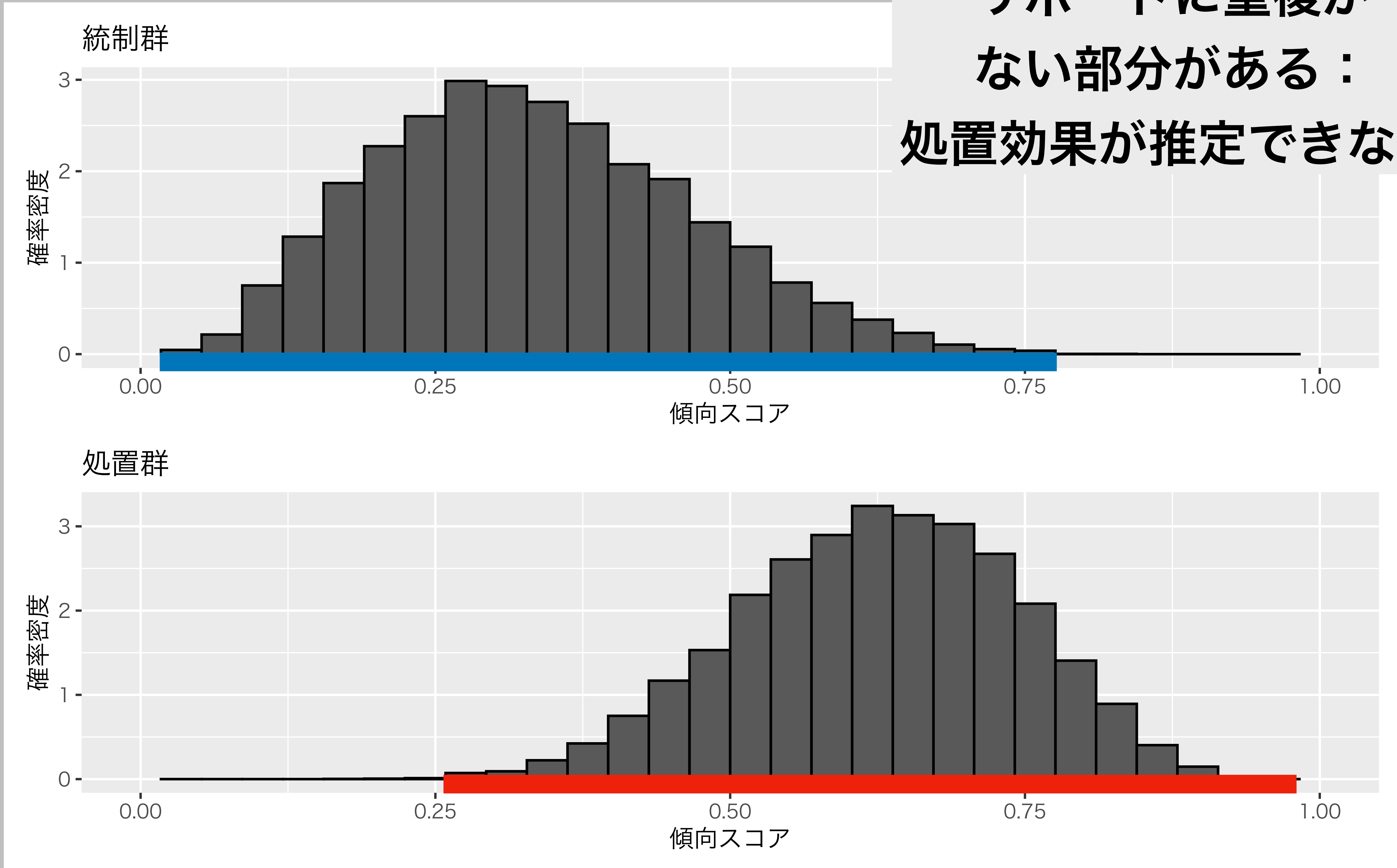


共有サポート：ケース3

統制群のサポートが処置群の一部：ATC が推定できる

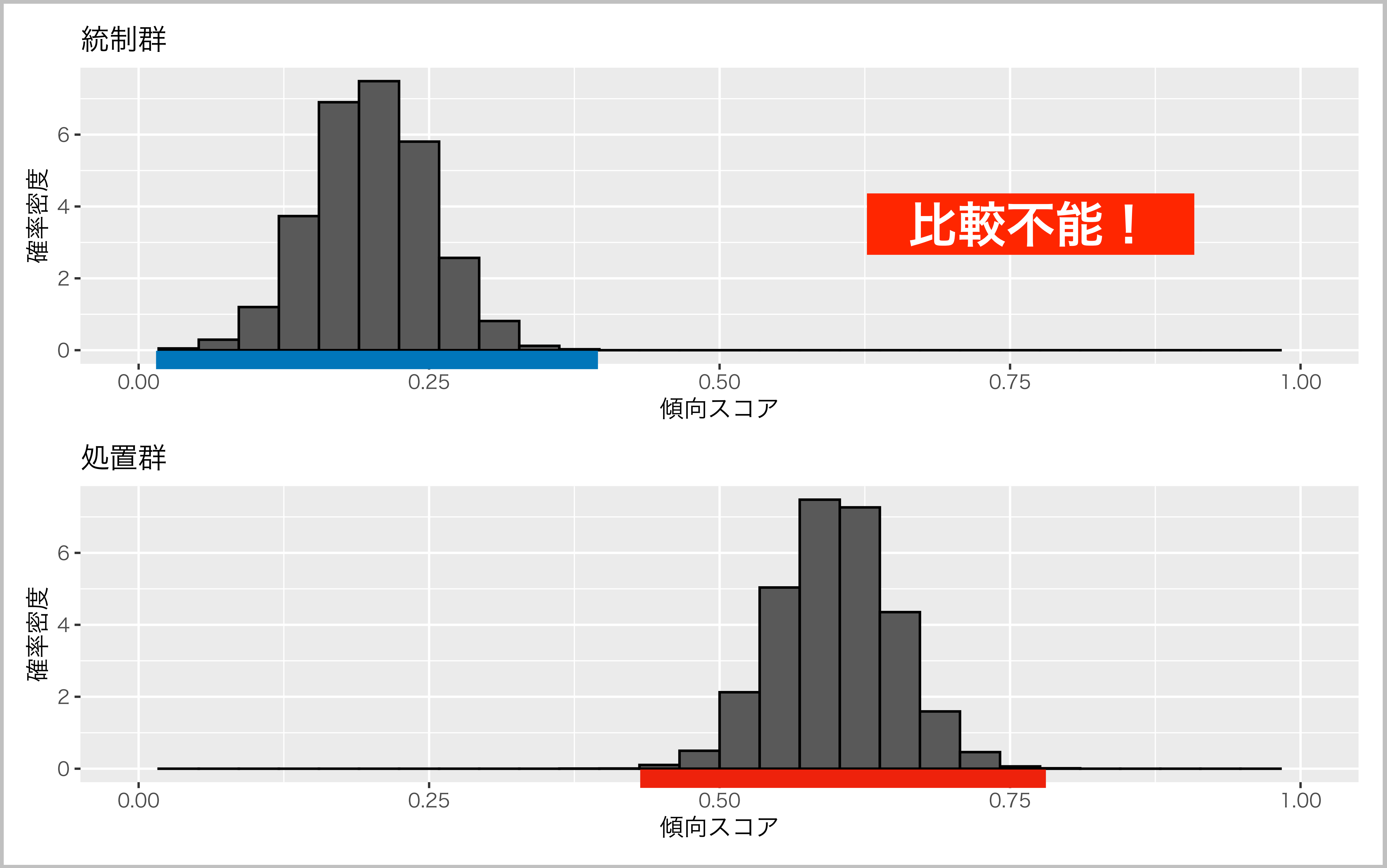


共有サポート：ケース4



サポートに重複がない部分がある：処置効果が推定できない

共有サポート：ケース5



統制群
確率密度
6
4
2
0
0.00
0.25
0.50
0.75
1.00
傾向スコア
比較不能！
処置群
確率密度
6
4
2
0
0.00
0.25
0.50
0.75
1.00
傾向スコア

共有サポートの確認が必要

- 片方の群のサポートが他の群サポートの一部なら、サポートが狭いほうの群における処置効果が推定できる
    - 観測数が小さいほうが、サポートが狭くなりやすい
    - 観察データでは、処置群のほうが標本サイズが小さい場合が多い（処置 [介入] にコストがかかることが多いため）
        - ATT を推定する場合が多い
- 共有サポートは図を描いて確かめる
    - ヒストグラム、密度曲線
    - 箱ひげ図、バイオリン図、蜂群図

4. バランスチェック

- 傾向スコアによる調整後に、処置群と統制群の間のバランスが改善したかどうか確かめる
  - 傾向スコアのバランス
  - 各共変量のバランス
- 処置群と統制群がバランスしていないと、バイアスが残っていると考えられる
- 傾向スコアと各共変量について、標準化平均差ができるだけ小さくなるように調整を行う

標準化平均差

- 標準化平均差 (standardized mean difference)

$$d = \frac{\bar{X}_{D=1} - \bar{X}_{D=0}}{\sqrt{\frac{\mathrm{Var}(X_{D=1}) + \mathrm{Var}(X_{D=0})}{2}}}$$

- 「バランスした」といえる標準平均差の絶対値
    - 厳しい基準：$|d| < 0.1$ (Austin 2011)
    - 緩い基準：$|d| < 0.25$ (Stuart 2010)

5. 因果効果の推定

- バランス調整がうまくいけば、あとは結果変数の平均を処置群と統制群で比較するだけ
- 何を推定しているのか (**estimand** は何か）を明確に
  - ATE, ATT, ATC, あるいはその他？
- 重み付けを使う場合：重み付き回帰
- 層別を使う場合：層別に求めた推定値の加重平均

まとめ

- 傾向スコアによって、セレクションバイアスを除去できる [こともある]
- メリット：1次元の条件付け
- 傾向スコアの使い方：重み付け、層別
  - 共有サポートの確認とバランスチェックが重要
  - 推定対象を明確にすることが必要
- 観測された共変量だけがセレクションバイアスの原因であること（強い意味での無視可能性）が**仮定**されている
  - 観測されていない共変量がセレクションバイアスの原因なら、推定はうまくいかない

次回予告

6. パネルデータの分析