

# 1. 統計的因果推論の基本

## A. 統計的因果推論の目標

**統計的因果推論**（因果推論, statistical causal inference）とは、**施策**（処置, treatment）によって**結果変数**（outcome）がどれだけ**因果効果**（treatment effect）を受けたかをデータから推定するフレームワークを与えます。例えば

- 「ダイレクトマーケティングによって商品の売上がどれだけ向上したか？」
- 「広告の差し替えによってクリック率（Click Through Rate）がどれだけ改善したか？」

といったテーマは、マーケティング領域において常日頃から向き合われている因果推論の課題です。

## B. 因果推論がなぜ必要なのか

施策の影響を推論するなら、施策あり群と施策なし群で標本平均を比較すればよいと思うかもしれませんが。しかし、2群の標本平均の比較では**データの取り方（study）次第で共変量（covariate）によるbias**が生じることが知られています。

**例（職業訓練の例）**：例えば大学の希望者に職業訓練を実施し、それが初任給に及ぼす影響を考えます。私たちは、

- 職業訓練の受講：職業訓練を受けたか？
- 初任給：初任給はいくらか？

というデータを揃えることが出来たとしましょう。このデータがあれば、職業訓練を受けた群と受けなかった群で標本平均の比較を行うことはできます。しかし、得られる結果からは職業訓練が初任給に及ぼす影響を適切に推定できないと想定されるのです。なぜなら、**職業訓練の受講**と**初任給**が共に**卒業時のGPA**と独立でない可能性があるからです。卒業時のGPAが高い大学生ほど初任給は高く、職業訓練の受講も希望するという傾向があるのであれば、2群の標本平均の差は**職業訓練の因果効果**より**卒業時のGPAの影響**分だけ大きな値を推定していることになるでしょう。言い換えれば、**卒業時のGPAの影響**が**職業訓練の因果効果**を推定するうえでbiasになっているわけです。■



**Remark**：施策と結果変数の両方に対して独立でないような変数を共変量といいます。例えば、職業訓練の例で卒業時のGPAは共変量です。■

## C. 因果推論における因果効果の定式化

ここで、因果推論の推定対象である「因果効果」とはなにかを明確にしておきましょう。

- 個別因果効果
- 平均因果効果

という言葉を導入します。各被験者に対する**個別因果効果**とは「仮に施策を受けた場合の結果」と「仮に施策を受けなかった場合の結果」の差と定義します。また、個別因果効果の母平均を**平均因果効果**（average treatment effect）と言います。

因果推論を数理統計の言葉で議論するために、記号を導入しておきましょう。

変数名	変数の意味
$T = 0, 1$	施策
$Y$	結果変数
$X$	共変量
$Y^{t=1}$	仮に施策を受けた場合の結果変数（潜在変数）
$Y^{t=0}$	仮に施策を受けなかった場合の結果変数（潜在変数）

個別因果効果は  $\tau = Y^{t=1} - Y^{t=0}$ 、平均因果効果は  $\mathbb{E}[\tau]$  と表すことができます。

## D. 因果推論の基本的な問題

因果推論の興味は、個別因果効果・平均因果効果の推定にあります。この問題は、個別因果効果の実現値を得ることが出来れば、本来難しくない話でしょう。しかし、データの取られ方次第では個別因果効果の実現値が得られないことがあるのです。

職業訓練の例では、職業訓練を受けた大学生について  $Y^{t=1}$  の実現値を得ることは出来ますが、 $Y^{t=0}$  の実現値を得ることは出来ません。職業訓練を受けていない大学生にも同様のことが成り立ちます。個別因果効果が

$$\tau = Y^{t=1} - Y^{t=0}$$

と定義されるということは、職業訓練の例の場合には**どんな被験者に対しても個別因果効果の実現値を得ることが出来ない**ということが従うのです。因果推論では、このような**データの取られ方のために個別因果効果の実現値を得られない場合でも**、個別因果効果  $\tau$  や平均因果効果  $\mathbb{E}[\tau]$  を推定できないかを検討していきます。

**Remark** : 職業訓練の例のように、データのとられ方によっては個別因果効果の実現値を得られないことがあります。これを**因果推論の基本的な問題**と言います。■

## E. 演習

**演習問題1** : 因果推論の基本的な問題は、個別因果効果の実現値が得られるようにデータを取れば解決できると思っただ方もいるかもしれませんが。しかし、分析課題のシチュエーションによっては、そのようなデータの取り方が困難なことがあります。職業訓練の例を参考に、そのようなデータの取り方が可能であるか検討してみてください。