Chapter4 : Potential Outcomes Causal Model（潜在アウトカム因果モデル）

富木良　平井藍里　鈴木涼太

目次

4 Potential Outcomes Causal Model

因果関係は数世紀に渡って経済学者の大きな関心であった

- 因果関係に関する現代的な概念の発展は、何人かの哲学者の著作から見られる
- ヒューム（1993）
  - 因果関係を「最初の事象が起こらなかったら、その後の事象も起こらないこと」という一時的事象の連続として説明
- ミル（2010）
  - 因果関係を推論するための5つの方法①一致法、②差分法、③結合法、④随伴変動法、⑤残差法を考案した。
- **経済学者たちは、因果関係を考える上で使うことのできるツール（数式など）を探していた**

統計的推論の登場

- そんな因果関係に関する研究は、現代的な統計学の発達とともに大きな発展を遂げる
  - 統計学が、メインのツールとなる！
- 特に、確率論と統計学は19世紀、天文学の分野を皮切りに科学に革命をもたらした
  - Carl Friedrich Gaussが18 歳で OLS を発見し、1809 年に 24 歳で OLS の導出を発表した

OLS発明の父

学問での応用の始まり

- 統計学者のG. Udny Yuleは、社会科学で早くから回帰分析を使用した
- 当時のイギリスの、英国における貧困に目を向ける。
- 公的支援が貧困者の数をどう変化させるか知りたかった。
- 最小二乗法を用いて、貧困と公的支援の関係を推測
- 1871年と1881年の国勢調査からデータを抽出
- データを用いた回帰の一例

$$\boxed{\mathrm{Pauper} = \alpha + \delta \mathrm{Outrelief} + \beta_1 \mathrm{Old} + \beta_2 \mathrm{Pop} + u}$$

学問での応用の始まり

- 回帰の結論として、「**公的扶助が貧困の増加率を高める**」と結論
- パッと聞いておかしいと思いましたか？？？
  - 公的扶助と貧困の増加率の間のバックドアが全て閉じられている？
  - この二つに影響を与える要因はあるのか？
  - 経済的な要因はコントロールしていない
  - 何よりも、因果関係が逆！貧困の増大が公的扶助を増やすであって、その逆ではない

統計的推論はまだまだ未熟だった

学問での応用の始まり

- 100年前の研究者、つまり回帰に代わるものがイデオロギー的な作り話であった時代の研究者を叩くのは、簡単なことである。ただ、彼らがその主張に反論をしてくれるわけではない。
- この事例のように、重要な政策課題に対する情報を提供する因果関係の主張方法として、回帰分析の素朴な利用は長らく疑いも無くされてきている。
  - そしてこれからもしばらくはされるだろう。

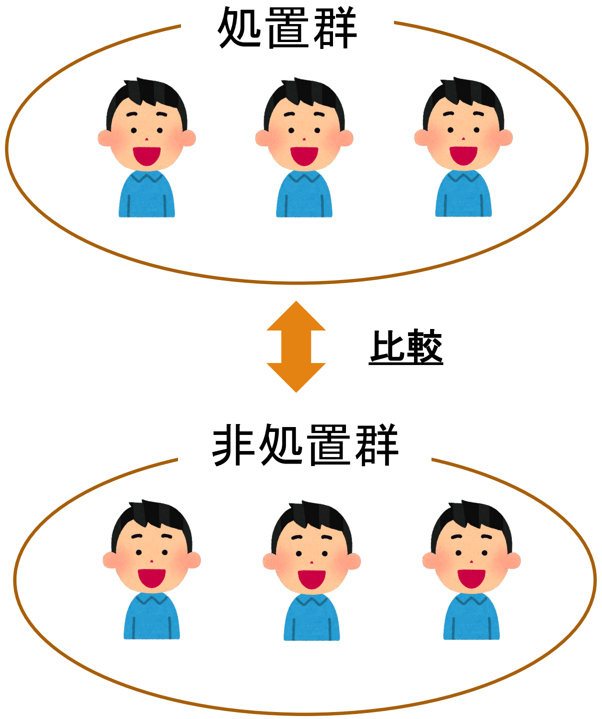
4.1 Physical Randomization

ここまでのおさらい

- 因果関係の研究に関するこれまでの歴史
  - 統計学の登場がキーとなり大きく発展
- ここからは、因果関係を考える上で重要な理論に触れる
  - ランダム化を元に、因果関係を探る

Physical Randomization

- 因果推論の基礎となるPhysical Randomization の概念は19世紀と20世紀には存在していた
  - 要するにランダム化実験のこと
  - 無作為にある人には処置を行なって、ある人には行わないで、結果を比較する
- 例：1971年から1982年にかけて、ランド研究所は医療保険が医療利用に及ぼす因果関係を研究する大規模な無作為化実験を行った。
  - 参加者は幾つかの医療保険プランのうちの1つに無作為に割り当てられた。自由診療、さまざまなレベルの費用負担を伴う3つのプラン、HMOプランである。費用負担のあるプランでは、無料プランに比べ、医師の診察回数や入院回数が少なかった。



なぜ無作為化は重要なのか？

無作為化を特徴とする実験計画は、応用ミクロ経済学、政治学、社会学、心理学などの分野で特徴的なものとなっている。しかし、なぜ無作為化は、因果関係を分析する上で重要なのか？

これを理解するためには、Splawa-Neyman（1923）が開発した "potential outcomes "（潜在アウトカム）について学ぶ必要がある

潜在アウトカム

- 潜在アウトカムは、因果関係の記述について考え、表現するための多かれ少なかれ共通語（lingua franca）となっている
- 因果関係の定義とは？
  - 世界の二つの状態の比較として定義される
  - 世界の最初の状態（世界の「実際」の状態と呼ばれることもある）において、ある人が頭痛のためにアスピリンを飲み、1時間後に頭痛のひどさを報告する。第二の世界（「反実仮想」と呼ばれることもある）では、同じ男が頭痛のために何も飲まず、一時間後に頭痛の程度を報告する。アスピリンの因果関係はどうだったのだろうか？potential outcomes tradition (Splawa-Neyman 1923; D. Rubin 1974)によれば、アスピリンの因果効果は、彼がアスピリンを飲んだ状態（世界の現実の状態）とアスピリンを飲まなかった状態（世界の反実仮想）の2つの世界の間の頭痛の深刻さの差である。この2つの世界の状態の間の頭痛の重症度の差が、それ以外は同じ時点で測定された、彼の頭痛に対するAspirinの因果関係である。
  - 簡単でしょ？

潜在アウトカム

- 前のスライドで説明した因果関係の定義には問題がある
  - 因果関係を分析するためには、実際に取らなかった選択肢の仮想現実からのデータが必要になる。
    →しかし、選択をしてしまったらもう一方の選択肢は消えてしまう
- これを具体的に式を使って表現する

前提の確認

Y_i^1 事象が発生した場合

Y_i^0 事象が発生しなかった場合

まったく同じ瞬間に、2つの別々の世界の状態にあることを示している
→二つの潜在的なアウトカム

Y_i 実際に観察される事象

実際に起こった現実
→仮想ではないので、右上の数字はない

実際に起こった結果はこのように表記が可能である

$$Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0$$

処置効果を２つの状態の差として表現可能（治療の効果など）

$$\delta_i = Y_i^1 - Y_i^0$$

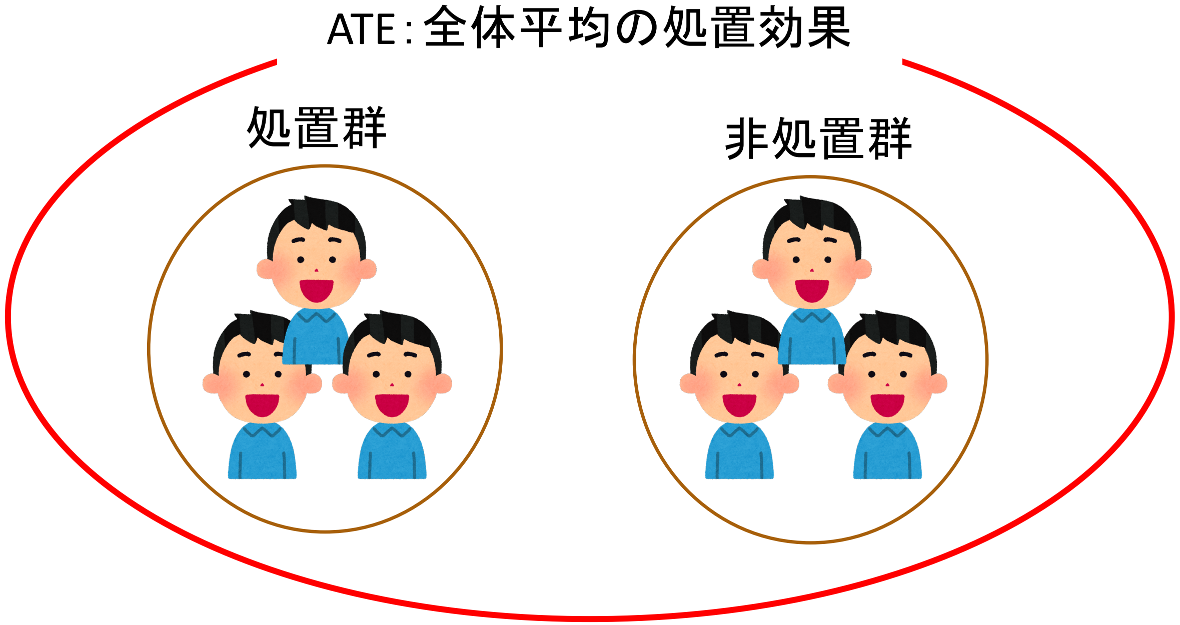
因果推論の根本的問題：必ず必要な片方のデータを集めることが不可能
（片方しか起きないから）

- 処置効果はどんなリサーチャーも気になることである
  - でもこのままでは仮想現実（潜在アウトカム）からデータを入手する必要があり、調べることが不可能である
- この回では、そんな処置効果を求める方法について学ぶ
  - ここに無作為化も関わってくる
- まずは、そんな求めたい処置効果を考える上で役立つATE、ATT、ATUの定義を確認する
  - そこからなぜ処置効果が求められるかの説明

平均的な処置効果

- リサーチャーが最も気になる指標
- 前ページの治療効果の単純な定義から平均治療効果を以下のように記述可能
- ただ、前述の通り、ATEを求めるために必要な数を両方とも取得することは不可能である
    - ATEは計算する量ではなく、推定する量である

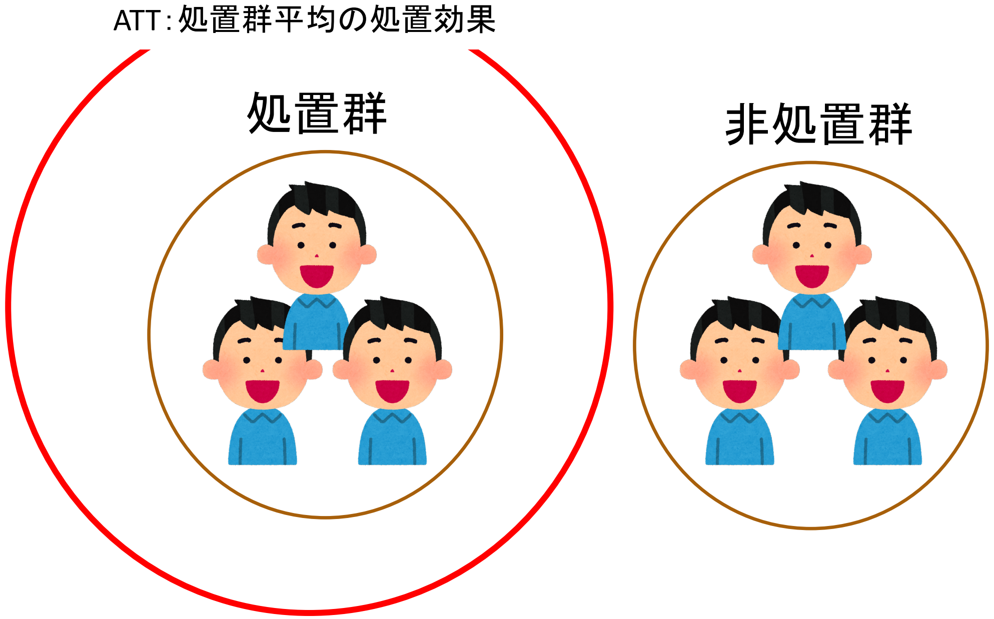
$$
\begin{aligned}
ATE &= E[\delta_i] \\
&= E[Y_i^1 - Y_i^0] \\
&= E[Y_i^1] - E[Y_i^0]
\end{aligned}
$$



処置群の平均処置効果

- 処置を受けたグループだけの平均処置効果

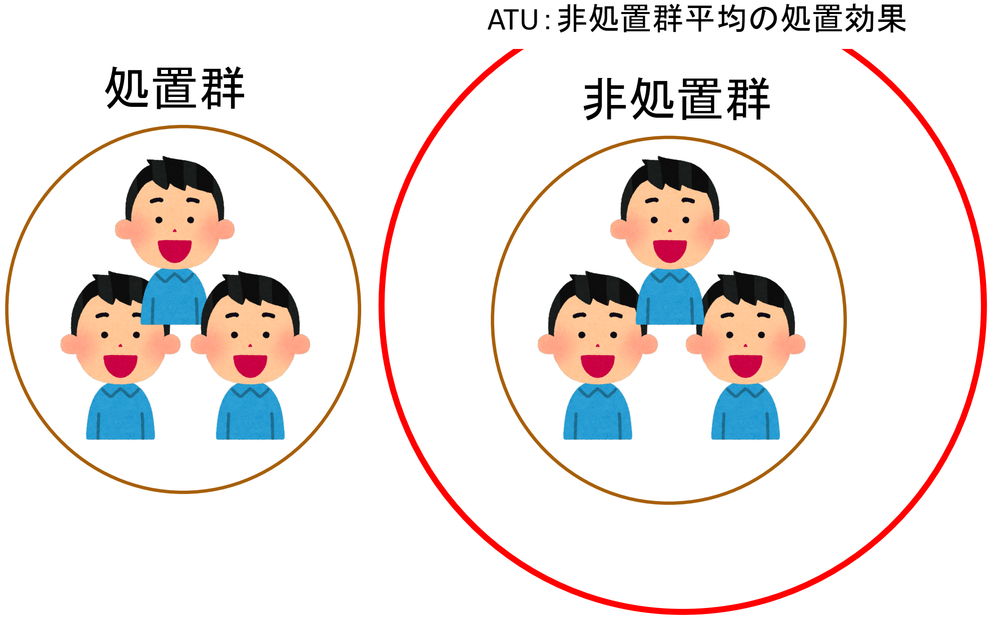
$$
\begin{aligned}
ATT &= E[\delta_i \mid D_i = 1] \\
&= E[Y_i^1 - Y_i^0 \mid D_i = 1] \\
&= E[Y_i^1 \mid D_i = 1] - E[Y_i^0 \mid D_i = 1]
\end{aligned}
$$



非処置群の平均処置効果

- 処置を受けなかったグループだけの平均処置効果

$$
\begin{aligned}
ATU &= E\left[\delta_i \mid D_i = 0\right] \\
&= E\left[Y_i^1 - Y_i^0 \mid D_i = 0\right] \\
&= E\left[Y_i^1 \mid D_i = 0\right] - E\left[Y_i^0 \mid D_i = 0\right]
\end{aligned}
$$



- ATE、ATT、ATUの中で全体の処置効果であるATEを求めることは可能なのか？
  - そもそも、もう一方で起きなかった仮想現実のデータを取るのは難しいし。。
  - でも、ATEを計算できれば、因果関係を調べる上で強力なツールになるはず！
  - そんなATEの計算方法を次で説明します！

4.1.3 SDOの分解　で理解すべきポイント！

最重要課題：ATEを求めること

①ATEを観察できるデータから求めることは不可能

②観察できるデータから求められるのはSDOのみ（ATT、ATU、ATEは不可能）

③SDOを用いて、ATEを推定したいがSelection Biasや異質介入バイアスが原因で推定できない

└SDOとは、処置群の処置による結果と無処置群の無処置による結果の差

$$
\begin{aligned}
SDO &= E[Y^1 \mid D=1] - E[Y^0 \mid D=0] \\
&= \frac{1}{N_T}\sum_{i=1}^{n}(y_i \mid d_i = 1) - \frac{1}{N_C}\sum_{i=1}^{n}(y_i \mid d_i = 0)
\end{aligned}
$$

4.1.3 SDOの分解
(STEP1)仮定をおいてATEの値を求める

仮定

・10人のがん患者（i＝1,2,3,4,5,6,7,8,9,10）

・各患者は手術を受ける（D=1）か、化学療法を受ける（D=0）かを選択できる

・潜在アウトカム

└ある患者iが　・手術を受けた場合の延命年数：Y_i^1
　　　　　　　・化学療法を受けた場合の延命年数：Y_i^0

└通常はどちらかのデータしか観察できないが、この仮定の下ではどちらのデータも分かってはいるとされている

（STEP1）
仮定のもとでのATE

（memo）
・手術を受けた患者iの延命年数：Y_i^1
・化学療法を受けた患者iの延命年数：Y_i^0

Patients	Y^1	Y^0
1	7	1
2	5	6
3	5	1
4	7	8
5	4	2
6	10	1
7	1	10
8	5	6
9	3	7
10	9	8

（STEP1）
仮定のもとでのATE

（memo）
・手術を受けた患者iの延命年数：Y_i^1
・化学療法を受けた患者iの延命年数：Y_i^0

ATE：平均処置効果(the average treatment effect)

$$ATE = E[Y_i^1] - E[Y_i^0]$$

ある患者iが
手術を受けた時の平均延命年数

ある患者iが
化学療法を受けた時の平均延命年数

・ATEは「手術を受けた場合、化学療法を受けたより何年長く生きられか」を示している

・ATE＞0→手術をした方が延命年数が長い
・ATE＜0→化学療法をした方が延命年数が長い（手術をした方が延命年数が短い）
・ATE＝0→手術と化学療法後の延命年数に違いがない

（STEP1）
仮定のもとでのATE

（memo）
・手術を受けた患者iの延命年数：Y_i^1
・化学療法を受けた患者iの延命年数：Y_i^0

Patients	Y^1	Y^0
1	7	1
2	5	6
3	5	1
4	7	8
5	4	2
6	10	1
7	1	10
8	5	6
9	3	7
10	9	8

ATE
$=E[Y_i^1]-E[Y_i^0]$
$=(7+5+5+7+4+10+1+5+3+9)/10$
$\quad -(1+6+1+8+2+1+10+6+7+8)/10$
$=5.6-5$
$=0.6$

→（結論）手術をした方が延命年数0.6年長い

（STEP1）
仮定のもとでのATE

（memo）
・手術を受けた患者iの延命年数：Y_i^1
・化学療法を受けた患者iの延命年数：Y_i^0

Patients	Y^1	Y^0
1	7	1
2	5	6
3	5	1
4	7	8
5	4	2
6	10	1
7	1	10
8	5	6
9	3	7
10	9	8

ATE
$= E[Y_i^1] - E[Y_i^0]$
$= (7+5+5+7+4+10+1+5+3+9)/10$
$\quad - (1+6+1+8+2+1+10+6+7+8)/10$
$= 5.6 - 5$
$= 0.6$

→（結論）手術をした方が延命年数0.6年長い

→（注意）
この仮定の下でのATEはただの平均であり、
異質な処置の効果も含んでしまっている

（STEP2）
新しい仮定をおいてATEを求める

新しい仮定

・10人のがん患者（i＝1,2,3,4,5,6,7,8,9,10）

・完璧な医者の存在

└各患者のY^1とY^0を知っており、
＝各患者の手術を受けた場合の延命年数と化学療法を受けた場合の延命年数を知っており、
処置後の寿命を最大にする治療を選択する完璧な医者

・完璧な医者が各患者に対して、<u>手術群</u>（D=1）か<u>化学療法群</u>（D=0）かを割り当てる

医者に手術を割り当てられた患者

医者に化学療法を割り当てられた患者

（memo）
・観察できる処置後の延命年数：Y
・医者が手術を割り当てた：D＝1
・医者が化学療法を割り当てた：D＝ 0

（STEP2）新しい仮定でのATE

Patients	Y	D
1	7	1
2	6	0
3	5	1
4	8	0
5	4	1
6	10	1
7	10	0
8	6	0
9	7	0
10	9	1

（STEP2）
新しい仮定をおいてATEを求める

・患者を手術群と化学治療群に分けなかった場合（前述の仮定）

→$ATE = E[Y_i^1] - E[Y_i^0]$

・患者を手術群と化学治療群に分けた場合（新しい仮定）

→$ATE = \pi \cdot ATT + (1 - \pi) \cdot ATU$

手術群の
手術後の平均寿命

化学療法群の
手術後の平均寿命

手術を受けた患者の割合

化学療法を受けた患者の割合

→新しい仮定の下でATEを求めるには、
<u>ATTとATUを求める</u>ことが必要

（STEP2-1） 新しい仮定でのATTとATU

（memo）
- 手術を受けた場合の延命年数：Y^1
- 化学療法を受けた場合の延命年数：Y^0
- 医者が手術を割り当てた：D＝1
- 医者が化学療法を割り当てた：D＝ 0

- ATT＝$E[Y_i^1 | Di=1] - E[Y_i^0 | Di=1]$

手術群のうち、

手術を選んだ人の平均延命年数

化学療法を選んだ人の平均延命年数

└患者は期待される利益に基づいて、医者に割り当てられた処置以外を選択することもできる

- ATU＝$E[Y_i^1|Di=0] - E[Y_i^0|Di=0]$

(STEP2-1)
新しい仮定でのATTとATU

(memo)
・処置後の延命年数：Y
・医者が手術を割り当てた：D＝1
・医者が化学療法を割り当てた：D＝ 0

Patients	Y	D
1	7	1
2	6	0
3	5	1
4	8	0
5	4	1
6	10	1
7	10	0
8	6	0
9	7	0
10	9	1

・ATT $= E[Y_i^1 \mid Di=1] - E[Y_i^0 \mid Di=1]$

$= 7 - ?$

・ATU $= E[Y_i^1 \mid Di=0] - E[Y_i^0 \mid Di=0]$

$= ? - 7.4$

（STEP2-1）
新しい仮定でのATTとATU

（memo）
・手術を受けた場合の延命年数：Y^1
・化学療法を受けた場合の延命年数：Y^0
・医者が手術を割り当てた：D＝1
・医者が化学療法を割り当てた：D＝0

Patients	Y^1	Y^0	D
1	7	1	1
2	5	6	0
3	5	1	1
4	7	8	0
5	4	2	1
6	10	1	1
7	1	10	0
8	5	6	0
9	3	7	0
10	9	8	1

・$ATT = E[Y_i^1 | Di=1] - E[Y_i^0 | Di=1]$
$= 7 - 2.6 = \underline{4.4}$
※患者1,3,5,6,10が対象

・$ATU = E[Y_i^1 | Di=0] - E[Y_i^0 | Di=0]$
$= 4.2 - 7.4 = \underline{-3.2}$
※患者2,4,7,8,9が対象

（STEP2）新しい仮定をおいてATEを求める

・ATT＝4.4

→手術群の手術後平均寿命は4.4年長くなる

・ATU＝ー3.2

→化学療法群の手術後平均寿命は3.2年短くなる

・ATE＝π・ATT＋（１ーπ）・ATU

＝0.5×4.4＋（1ー0.5）×（ー3.2）＝0.6

→手術をした方が平均寿命が0.6年長くなる

※1つ目の仮定のもとで求めたATEと同じ値になっている

（結論）
手術の効果は
一部の人にとってはマイナスだが、
全体としてはプラスになっている

（STEP2）新しい仮定をおいてATEを求める

→実際にATTとATUを観察することは不可能でATEも求められない。
　なぜなら、ATTとATUを求めるには観察不可能なデータが含まれているから。

ATE＝π・ATT＋（1－π）・ATU

観察不可能！

・ATT＝$E[Y_i | Di=1] - E[Y_i^0 | Di=1]$

・ATU＝$E[Y_i^1 | Di=0] - E[Y_i | Di=0]$

観察可能なデータのみを用いて求めるSDOを使って、ATEを推定する

(STEP3-1)
ATEの推定値SDOを求める

・SDOはATEの推定値として機能する
・観察可能なデータのみを用いて求められる
・手術群が実際に手術をし、化学療法群が実際に化学療法を受けた場合の平均寿命を比較する

SDO：平均値の単純差(the simple differnce in outcomes)

$$SDO = E[Y^1 | D=1] - E[Y^0 | D=0]$$

医者に手術を割り当てられ、
手術を受けた患者の平均延命年数

医者に化学療法を割り当てられ、
化学療法を受けた患者の平均延命年数

（STEP3-1）
ATEの推定値SDOを求める

Patients	Y	D
1	7	1
2	6	0
3	5	1
4	8	0
5	4	1
6	10	1
7	10	0
8	6	0
9	7	0
10	9	1

SDO＝7ー7.4＝ー0.4
→化学療法の方が
　平均寿命は0.4年長くなる

本当にSDOはATEの推定値になっているのか？ATEは0.6だったのではないか？
→SDOを分解して、SDO＝ATEが成り立っているのかを確かめていく。

(STEP3-2) SDOを分解する

（memo）

ATT＝E[Yi1|Di=1]－E[Y0|Di=1]
ATU＝E[Yi1|Di=0]－E[Y0|Di=0]

①ATEの構成要素を分解し、SDO＝ATE＋{E[Y0|Di=1]－E[Y0|Di=0]}＋(1－π)[ATT－ATU]を導く

ATE ＝π･ATT＋（1－π）ATU

＝π{E[Yi1|Di=1]－E[Y0|Di=1]} ＋（1－π）{E[Yi1|Di=0]－E[Y0|Di=0]}

ATE ＝{π･E[Yi1|Di=1]＋（1－π）E[Yi1|Di=0]}－{π･E[Y0|Di=1]＋（1－π）E[Y0|Di=0]}

②文字に置き換える

e ＝π･a＋(1－π)b－π･c＋(1－π)d

e ＝π･a＋(1－π)b－π･c＋(1－π)d +(a-a)+(b-b)+(c-c)

⋮

a－d＝e＋(c－d)＋(1－π)(a－c)－(1－π)(b－d)

a－d＝e＋(c－d)＋(1－π)[(a－c)－(b－d)]

(STEP3-2)
SDOを分解する

③置き換えていた文字を戻す

a－d＝e＋(c－d)＋(1－π)[(a－c)－(b－d)]

E[Yi1|Di=1]－E[Y0|Di=0]＝ATE＋{E[Y0|Di=1]－E[Y0|Di=0]}＋(1－π)[ATT－ATU]

SDO＝ATE＋{E[Y0|Di=1]－E[Y0|Di=0]}＋(1－π)[ATT－ATU]　□

(memo)

･ATT＝E[Yi1|Di=1]－E[Yi0|Di=1]=a－c
･ATU＝E[Yi1|Di=0]－E[Yi0|Di=0]=b－d
・ATE＝e
・SDO=E[Y1|D=1]－E[Y0|D=0]

(STEP3-2)
SDOの右辺を考える

(memo)

ATE $=\boxed{E[Yi1]}-\boxed{E[Yi0]}$

ATT $=E[Yi1|Di=1]-\boxed{E[Yi0|Di=1]}$

ATU $=\boxed{E[Yi1|Di=0]}-E[Yi0|Di=0]$

観察不可能

SDOは以下の3つの部分に分解できる(スライド28、29の証明を参照)

SDO $=$ ATE $+\{\boxed{E[Y^0|Di=1]}-E[Y^0|Di=0]\}+(1-\pi)[\boxed{ATT-ATU}]$

観察できないデータを用いて計算される値だから、
SDOから差し引いてATEを求めることも出来なくさせる邪魔者＝バイアス

→結論:SDO＝ATEはバイアスがあるため成り立たない

(STEP3-3)
SDOの右辺を考える

$$SDO = ATE + \{E[Y^0 | Di=1] - E[Y^0 | Di=0]\} + (1-\pi)[ATT - ATU]$$

Selection Bias

└一般的には、

処置群と無処置群間における個体自体の固有の差

(STEP3-3)
SDOの右辺を考える

SDO＝ATE＋{E[Y^0|Di=1]－E[Y^0|Di=0]}＋(1－π)[ATT－ATU]

Selection Bias

異質介入効果バイアス（heterogeneous treatment effect bias）

└処置群と無処置群における処置効果の差に、
無処置群の割合を掛けたもの
└観察できないデータに基いているからからバイアスになる

（memo）

SDO＝ATE＋{$E[Y^0|Di=1]$－$E[Y^0|Di=0]$}＋(1－π)[ATT－ATU]

(STEP3-4)
バイアスを除くために均質性を仮定する

均質性を仮定する
＝処置群の患者であっても無処置群の患者であっても、
　すべての患者における処置効果の大きさ（$Yi^1-Yi^0=di$）が等しいとする

→ATT=ATUになるため、ATT－ATU＝0になる

（memo）

$ATT = E[Y_i^1|Di=1] - E[Y_i^0|Di=1]$

$ATU = E[Y_i^1|Di=0] - E[Y_i^0|Di=0]$

(STEP3-4)
バイアスを除くために均質性を仮定する

SDO＝ATE＋{E[Y^0|D=1]－E[Y^0|D=0]}＋$\underbrace{(1－π)[ATT－ATU]}_{=0}$

→異質介入効果バイアスは消えるが、Selection Biasは残り続ける

※均質性を仮定するとは？
患者１の処置効果：$Y1^1$-$Y1^0$=d1
患者2の処置効果：$Y2^1$-$Y2^0$=d2
d1＝d2を仮定しているだけであり、$Y1^0$と$Y2^0$が等しいことを表しているわけではないから、Selection Biasは均質性を仮定しても存在し続ける

（＋α）
均質性を仮定しなくてもSelection Biasは邪魔者！

SDOを上記とは別の形で分解する

$$\mathrm{SDO} = E[Y^1|D=1] - E[Y^0|D=0]$$

$$= E[Yi^1|Di=1] - E[Yi^0|Di=0] + E[Yi^0|Di=1] - E[Yi^0|Di=1]$$

$$= E[Yi^1|Di=^1] - E[Yi^0|Di=1] + E[Yi^0|Di=1] - E[Yi^0|Di=0]$$

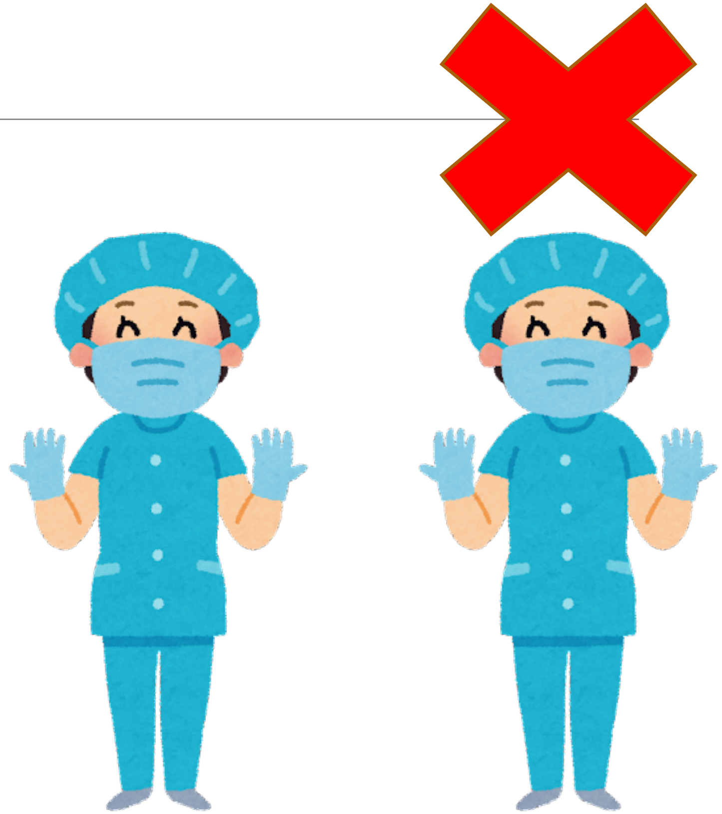
$$= \mathrm{ATT} + \text{Selection Bias}$$

手術群の手術効果（処置群の処置効果ATT）はATEほどではないが、まだ興味がある。
しかし、Selection BiasがあるからATTさえも求めることが出来なくなっている。

4.1.4 独立性の仮定

$$(Y^1, Y^0) \perp D$$ と仮定する

<ins>処置するかは潜在アウトカムとは無関係に決める</ins>



D=1なら処置あり　D=0なら処置な

Y^1: 処置するときの余命　Y^0: 処置しないの余命

独立性の仮定

$(Y^1、Y^0) \perp D$ が成立しない例

- 医者が患者ごとに適した治療を選ぶ

<u>$(Y^1、Y^0) \perp D$ が成立する例</u>

- 場合1．患者を名前順で並べ、交互に治療を割り当てる
- 場合2．時計の秒針が1-30にあったら手術し、31-60にあったら手術しない

独立性の仮定

(Y^1、Y^0) ⊥ D を仮定すると以下の2式が成立

$$E[Y^1 \mid D=1] - E[Y^1 \mid D=0] = 0$$
$$E[Y^0 \mid D=1] - E[Y^0 \mid D=0] = 0$$

(式の意味)

- 処置群と対照群に同じ治療をしても潜在アウトカムは同じになる
  - そもそもランダムに分けてるので当たり前

独立性の仮定

$$E[Y^1 \mid D=1] - E[Y^1 \mid D=0] = 0$$

$$E[Y^0 \mid D=1] - E[Y^0 \mid D=0] = 0$$

の成立時

$$\underbrace{E[Y^0 \mid D=1] - E[Y^0 \mid D=0]}_{\text{Selection bias}} = 0$$

$$\underbrace{(1-\pi)(ATT - ATU)}_{\text{Heterogeneous treatment effect bias}} = 0$$

【結論】2つのbias=0

独立性の仮定

$$\underbrace{(1-\pi)(ATT-ATU)}_{\text{Heterogeneous treatment effect bias}} = 0$$

の理由

$$\begin{aligned} ATT - ATU &= \mathbf{E}[\mathbf{Y}^1 \mid \mathbf{D}=\mathbf{1}] - E[Y^0 \mid D=1] \\ &\quad - \mathbf{E}[\mathbf{Y}^1 \mid \mathbf{D}=\mathbf{0}] + E[Y^0 \mid D=0] \\ &= 0 \end{aligned}$$

*仮定より、2つの$\boxed{}$ =0

独立性の仮定

結果（Y^1、Y^0）⊥　D　という仮定の下で

SDO＝ATE

SDO＝ATEのために必要なもの

(a)観測可能なデータについての結果

(b)治療割り当てのデータ

(c)(Y^1、Y^0) ⊥　D　という仮定

そもそも推定したいのはATEなので

SDOをそのままATEとして扱えるという結果はすごい！

独立性の仮定

しかし$(Y^1、Y^0) \perp D$ という仮定はあまり現実的ではない

- 意思決定は合理的選択と関係しているから

Ex. 親が子供にとってベストになる学校へ入学させる。

- 現実では、SDO＝ATEは成り立ちにくい
  - 平均値の単純比較では因果関係を正しく推定できないことが多い！

Chapter4 : Potential Outcomes Causal Model（潜在アウトカム因果モデル）

富木良　平井藍里　鈴木涼太

目次

4.2 無作為化推論

因果推論の根本問題（the fundamental problem of causal inference）

→反実仮想は事実として存在しないから、因果関係を直接的に求めることはできない

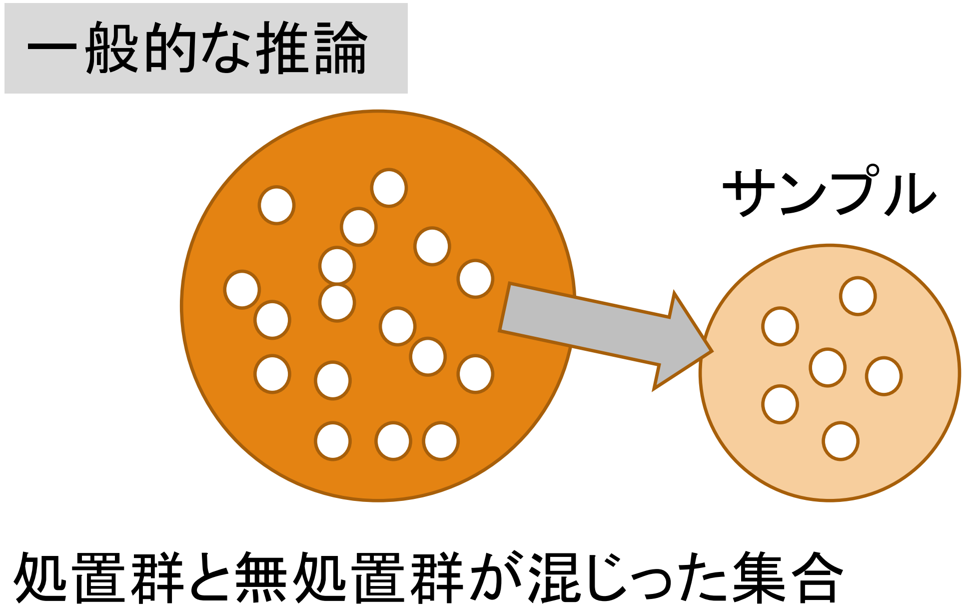
私たちに出来ることは推定することのみであるため、
私たちが計算した推定値が本当に因果関係になり得ているのかを示すことが必要

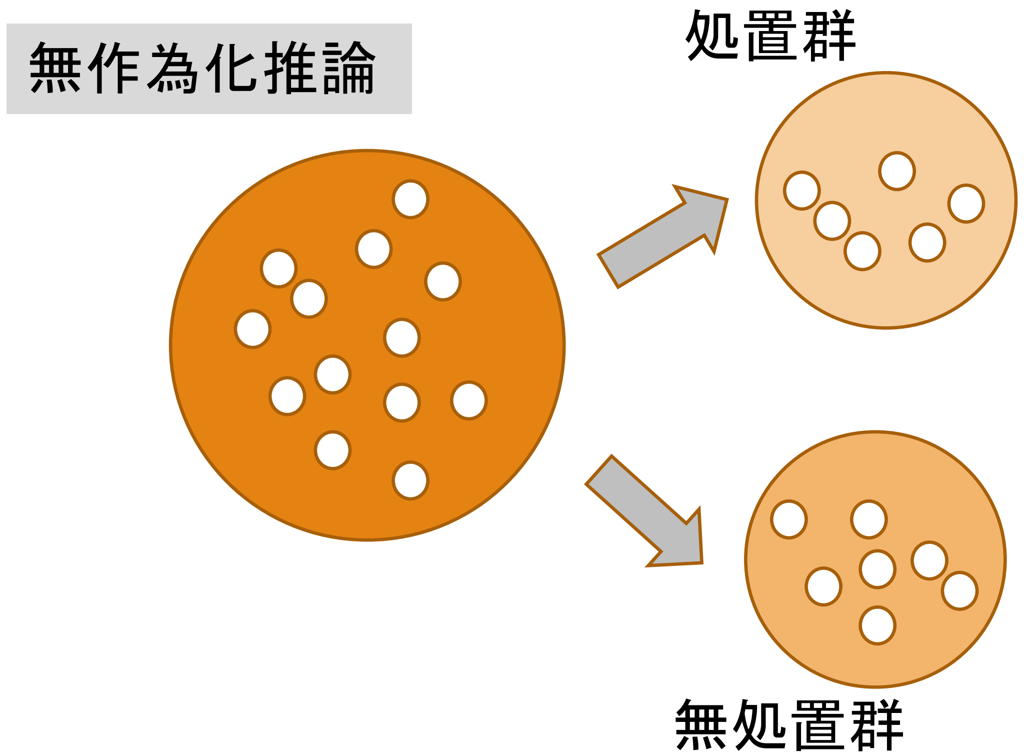
R. A. Fisher

推定値の信頼性が高いことは、
「偶然からこの推定値が生じた確率（p値）」を求めてその値が十分に小さいことを理由に示せる

→この単元のテーマである無作為化推論は、p値を求めるために使われている

（復習）無作為化推論とは？





無作為化推論の特徴

①無作為化推論では、潜在アウトカムによる不確実性を扱う

（一般的な推論で扱う不確実性は、サンプリングによる不確実性）

②Young（2019）によれば、

一般的な推論は外れ値に引っ張られやすいが、

それに比べて無作為化推論は外れ値に結果が引っ張られにくい

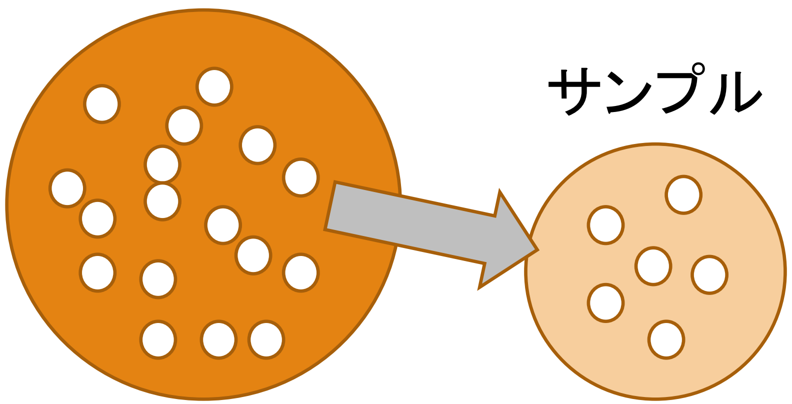
③直感的に理にかなっていると捉えられている

無作為化推論の特徴

①無作為化推論では、潜在アウトカムによる不確実性を扱う

（一般的な推論で扱う不確実性は、サンプリングによる不確実性）

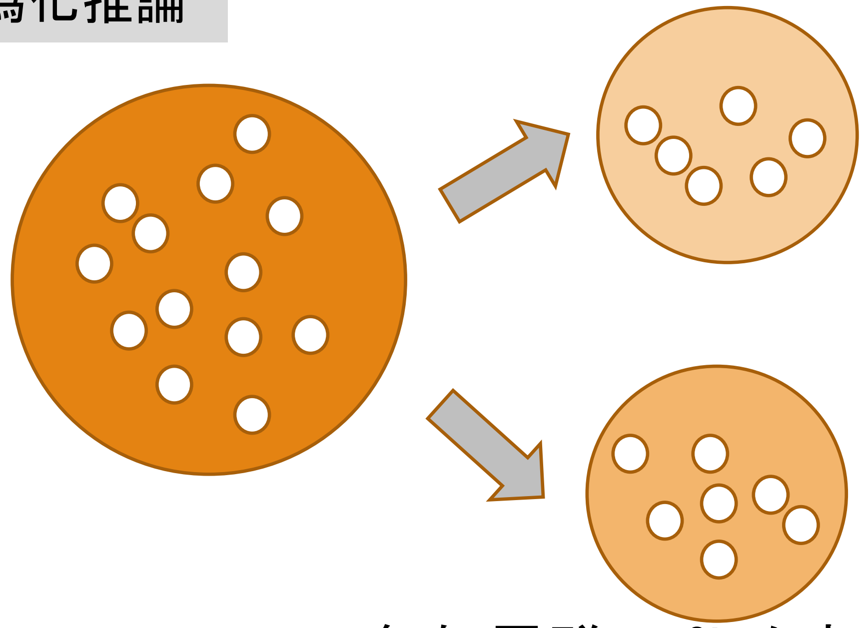
一般的な推論



処置群と無処置群が混じった集合

無作為化推論

処置群→Y^1しかわからない



無処置群→Y^0しかわからない

無作為化推論の特徴

①一般的な推論で扱う不確実性は、サンプリングによる不確実性

　無作為化推論で扱う不確実性は、潜在アウトカムの不確実性

②Young（2019）によれば、

　一般的な推論は外れ値に引っ張られやすいが、

　それに比べて無作為化推論は外れ値に結果が引っ張られにくい

③直感的に理にかなっていると捉えられている

4.2.1 お茶を飲む女性（無作為化推論の起源）

実験（1935）

Muriel Bristol（女性の名前）

└女性が博士号を取得することがほとんどできなかった時代に博士号を取った科学者

└ミルク入りの紅茶を与えられれば、ミルクと紅茶のどちらが先に注がれたかを見分けることができると主張した

→　R. A. Fisherは、Murielが主張する才能を確かめるために実験を行った

実験内容：Fisher’s exact test

→ミルクを先に入れた紅茶

→紅茶を先に入れた紅茶

Murielが
（ミルクが先に入れた紅茶）or（紅茶が先に入れた紅茶）
のどちらかを4つ分すべて選ぶ

実験内容：Fisher's exact test

もし、4つ分の紅茶を正しく選べた場合
①本当に才能があって判別できた可能性
②たまたま偶然に正しい判別をしていた可能性　　　がある

Murielが本当に才能を持っているのかを調べるために

4つ分の紅茶を正しく選ぶことが偶然によって生じた確率を求める
②が生じる確率が十分に低ければ、確率的にはMurielは才能を持っているといえる

偶然4つ分の紅茶を正しく選べる確率は？

・8つの紅茶から4つ分を選ぶには何通りあるか？　→8×7×6×5＝1680通り

・4つ分の紅茶を過不足なく選ぶには何通りあるか？　→4×3×2×1＝24通り

・8つの紅茶から4つ分の（ミルクor紅茶）が先に入れられた紅茶のみを過不足なく選ぶ確率は？　→24/1680＝1/70

（結論）
ランダムに、偶然に、4つ全てを正しく選ぶ確率は十分に低いため、
もしMurielが4つ分を正しく選べたのなら、彼女は才能を持っているといえる

→ある事象が偶然に生じる確率を求めることで、その事象の信頼性を確かめられる

4.2.2 Fisherのシャープナルの仮定

通常

$$H_0 : E[Y_i^1 - Y_i^0] = 0$$

Di	Yi	Y^1	Y^0	Y^1 - Y^0
1	2	2	?	2 - ?
0	3	?	3	? - 3
0	6	?	6	? - 6
1	7	1	?	1 - ?

→平均が0

シャープナルの仮定

$$H_0 : \delta_i = Y_i^1 - Y_i^0 = 0 \forall i$$

Di	Yi	Y^1	Y^0	Y^1 - Y^0
1	2	2	?	0
0	3	?	3	0
0	6	?	6	0
1	7	1	?	0

4.2.3 無作為化推論の手順

用いるデータ

Name	D	Y	Y^0	Y^1
Andy	1	10	.	10
Ben	1	5	.	5
Chad	1	16	.	16
Daniel	1	3	.	3
Edith	0	5	5	.
Frank	0	7	7	.
George	0	8	8	.
Hank	0	10	10	.

[STEP1]　帰無仮説をたてる

・シャープナルの仮定 $H_0 : \delta_i = Y_i^1 - Y_i^0 = 0 \forall i$

Name	D	Y	Y^0	Y^1
Andy	1	10	**10**	10
Ben	1	5	**5**	5
Chad	1	16	**16**	16
Daniel	1	3	**3**	3
Edith	0	5	5	**5**
Frank	0	7	7	**7**
George	0	8	8	**8**
Hank	0	10	10	**10**

[step2] 検定統計量　τ　の構築

τ＝τ (D,Y)

- 割り当てDとアウトカムYから計算する値

  Ex. [処置群のyの平均]ー[対照群のyの平均]

- STEP１で設定した帰無仮説が正しいか判定するのに用いる

$$\boxed{H_0 : \delta_i = Y_i^1 - Y_i^0 = 0 \forall i}$$

[step2] 検定統計量　τ　の構築

Name	D	Y	Y^0	Y^1
Andy	1	10	**10**	10
Ben	1	5	**5**	5
Chad	1	16	**16**	16
Daniel	1	3	**3**	3
Edith	0	5	5	**5**
Frank	0	7	7	**7**
George	0	8	8	**8**
Hank	0	10	10	**10**

□の平均ー□の平均＝1

これを検定統計量τとする

割り当てにより、偶然に生じた差(帰無仮説)

Or

処置効果から生じた差(対立仮説)

[step3]　τの分布を調べる

帰無仮説が正しいなら、τ＝0になるはず

実際にはτ＝1になったが、これは「割り当て」により生じた誤差？

τ＝1が他の割り当てでのτと比べ上位何%か(p値)を調べたい

今回はτの分布が完全に分かるので、p値計算できる

[step3] τの分布を調べる(続き)

８C４＝７０通りの割り当てを考え、それぞれの場合でτを計算する

その中で検定統計量τ＝１が極端に上位だった場合

✕「小さな確率でしか起こらないことが偶然起こった」

〇「そもそも帰無仮説が間違っている」と考えた方が自然

[STEP4] p値を計算する

p値の計算方法

$$\Pr\left(t(D', Y) \geq t(D_{observed}, Y) \mid \delta_i = 0, \forall i)\right) = \frac{\sum_{D' \in \Omega} I(t(D', Y) \geq t(D_{observed}, Y))}{K}$$

4.2.4例

使うデータ

Name	D	Y	Y^0	Y^1
Andy	1	10	.	10
Ben	1	5	.	5
Chad	1	16	.	16
Daniel	1	3	.	3
Edith	0	5	5	.
Frank	0	7	7	.
George	0	8	8	.
Hank	0	10	10	.

８人の精神疾患を持つ人から４人選び、認知行動療法(CBT)プログラムを行う。その後、８人の精神的健康度を０－２０の数値で示す。

例

[step1] 帰無仮説を立てる

$$H_0 : \delta_i = Y_i^1 - Y_i^0 = 0 \forall i$$

「CBTプログラム」の効果はすべての個人に対して0である」

[step2] 検定統計量τの構築

検定統計量τ=「処置群のyの平均－非処置群のyの平均」

例

[step3]τの分布を調べる

	Andy	Ben	Chad	Daniel	Edith	Frank	George	Hank
D	1	0	1	1	0	1	0	0
y	10	5	16	3	5	7	8	10

割り当てを上の表のようにする

- この割り当ての下で計算されるτはτ＝2
- 8C4＝70通りの全てについて同様にτを計算する

例

[Step４]

τの分布が分かったら、τ＝１が上位何％か(p値)を計算する

棄却する水準(例えばp=0.05)と比較する

それより小さければ帰無仮説を棄却する

その他の検定統計量

検定統計量として、
τ=「処置群のyの平均－対照群のyの平均」
を選んだ

<u>**他にどのような検定統計量が考えられる？**</u>
<u>**（例）外れ値に頑健な検定統計量**</u>

log変換

- log変換された平均的な処置効果の差
- これは、元データが歪んでいる場合に効果がある
  - データの歪みは、収益などの正の値の場合や、処置効果が相加的ではなく乗法的である場合に発生する。

$$T_{\log} = \left| \frac{1}{N_T} \sum_{i=1}^{N} D_i \ln(Y_i) - \frac{1}{N_C} \sum_{i=1}^{N} (1 - D_i) \ln(Y_i) \right|$$

中間値の差の絶対値

- 中間値の差の絶対値
  - 外れ値に対して頑健となる

$$T_{\text{median}} = \left| \text{median}(Y_T) - \text{median}(Y_C) \right|$$

差ではなくランキングを用いる

- 外れ値が懸念される場合は、**差を用いるのではなく、ランキングを用いる**ことも一つの手である
  - 外れ値が多数ある場合、結果が連続的である場合、またはデータセットが小さい場合に役立つ。
- ランク統計の基本的な考え方は、結果をランク付けしてから、治療群と対照群の平均ランクを比較することとなる。

差ではなくランキングを用いる

Name	D	Y	Y^0	Y^1	Rank
Andy	1	10	**10**	10	6.5
Ben	1	5	**5**	5	2.5
Chad	1	16	**16**	16	8
Daniel	1	3	**3**	3	1
Edith	0	5	5	**5**	2.5
Frank	0	7	7	**7**	4
George	0	8	8	**8**	5
Hank	0	10	10	**10**	6.5

Table 4.12: Illustrating ranks using the example data.

- 実際の例
  - 数字が小さい順にランキングが高くなる。
  - 2位でランキングが等しい時は、2.5とする（平均を取る）

差ではなくランキングを用いる

Name	D	Y	Y^0	Y^1	Rank	R_i
Andy	1	10	**10**	10	6.5	2
Ben	1	5	**5**	5	2.5	-2
Chad	1	16	**16**	16	8	3.5
Daniel	1	3	**3**	3	1	-3.5
Edith	0	5	5	**5**	2.5	-1
Frank	0	7	7	**7**	4	-0.5
George	0	8	8	**8**	5	0.5
Hank	0	10	10	**10**	6.5	2

Table 4.12: Illustrating ranks using the example data.

- 一番右の列では、ランキングを標準化している

$$\widetilde{R}_i = \widetilde{R}_i(Y_1, \dots, Y_N) = \sum_{j=1}^{N} I(Y_j \leq Y_i) - \frac{N+1}{2}$$

- このランクを使って、absolute value of the simple difference in mean outcomes

$$T_{\text{rank}} = |0 - 1/4| = 1/4$$

問題

	処置あり Y1	処置なしY0
Aさん	8	
Bさん	2	
Cさん		1
Dさん		4

処置の効果を調べるため、調査がなされた。上のような
結果が観測された。
帰無仮説を $H_0 : \delta_i = Y_i^1 - Y_i^0 = 0 \forall i$
として無作為化推論を行うとき、p値を求めよ