

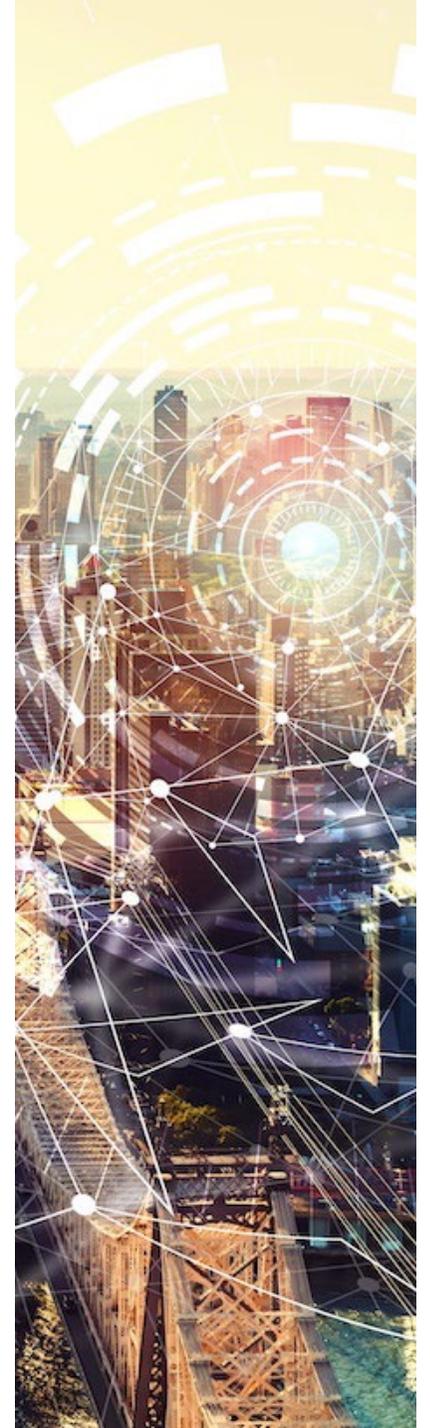
因果推論

株式会社野村総合研究所
金融ITイノベーション事業本部
金融デジタル企画一部

2020年6月22日

NRI

Share the Next Values!



ガイダンス

- 以下の講義内容で進めていきます。
- 質疑応答は、チャットなどでいつでも受け付けます。
 - 1コマ目：10時から11時半まで
 - ・ 統計分析の復習
 - 2コマ目：12時半から14時まで
 - ・ 相関と因果の考え方
 - ・ 効果量の測定方法
 - ・ DIDの考え方
 - 3コマ目：14時半から16時まで
 - ・ バイアスの除去方法
 - ・ 傾向スコアマッチング
 - ・ IPV法
 - ・ 効果量の測定方法
 - 4コマ目：16時半から18時まで
 - ・ 検証実験の設計
 - ・ 検出力解析
 - ・ 効果検証データの解析
 - ・ 検証結果の妥当性

講義 2（12時半から14時まで）の概要

- 以下の講義内容で進めていきます。
 - 相関と因果の考え方
 - 効果量の測定方法
 - ・ DIDの考え方

- 因果の考え方を持ち込むことで、統計分析がどのように変わるのか
 - どのようなことができるようになるのか
 - どのような点に注意しなければならないのか

因果推論のはじめ

因果推論のはじめ

- 適切な分析を行っても結論を読み間違えることはある。
- 相関と因果の読み取りを間違えると...

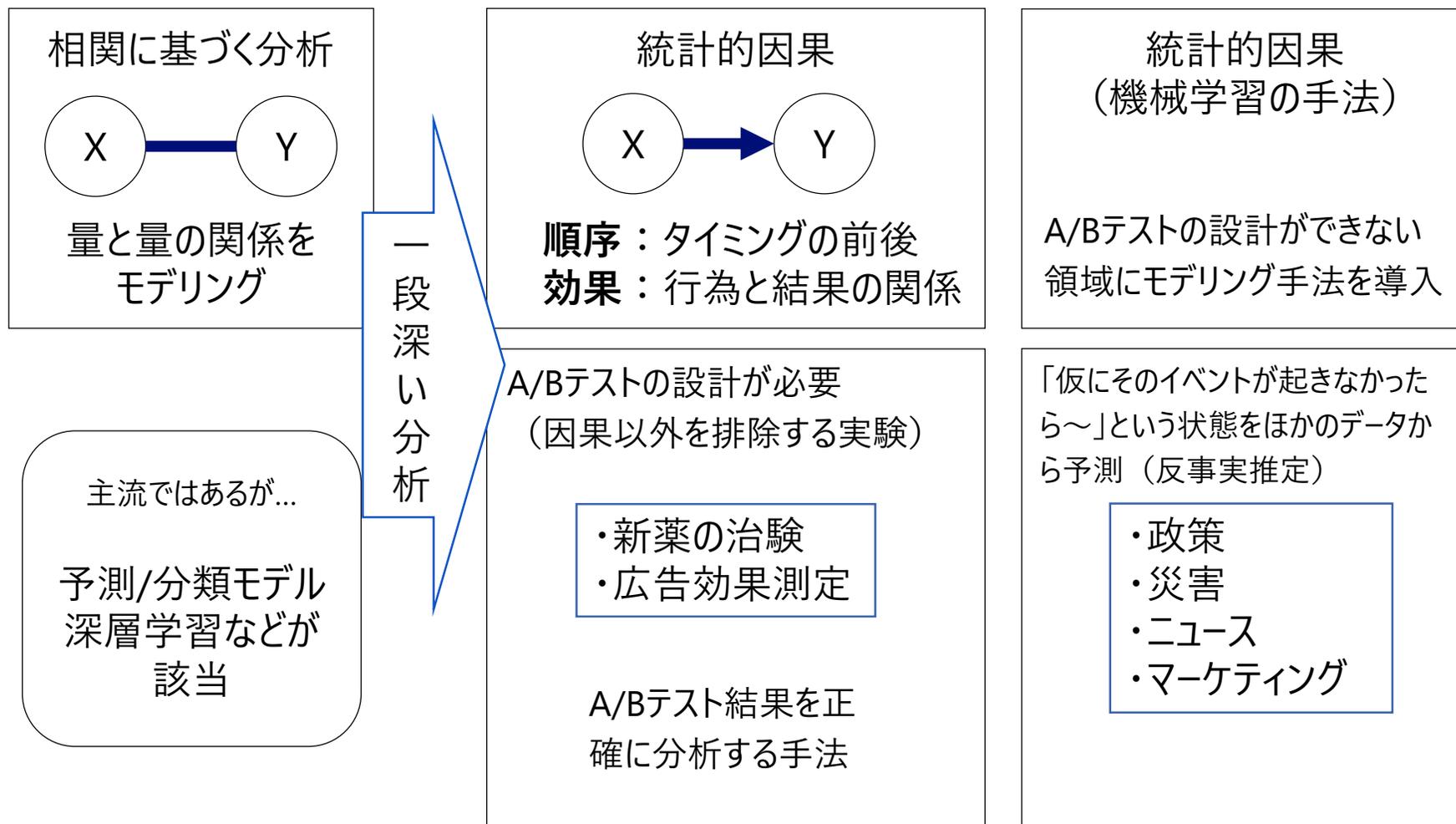
売上の大きな企業は一般に経費も大きい。つまり、売上と経費には強い相関関係があると考えられる。それでは、売上を増やすには経費を増やせばよいという判断は正しいか

難関大学へ合格した人は高級な文具を使っていた。合格するために同様の文具をそろえることは正しいか。

同業種のA社とB社は、競合他社でありながら売上を伸ばしている。このとき、2社の売上は相関関係にあることになるが、A社の営業活動がB社の売上に貢献しているのだろうか

因果分析

- 因果分析とは、「ある時点のイベントが、対象データに対して、どう影響したか大きさを測定すること」



因果関係推定

■ 因果関係推定で難しいと考えられる点。

データの持つ統計的性質や「因果」をどう考えるかによって、アプローチがさまざまにある。

● 因果の定義

- 「Aの後、必ずBが（直接的に）起こる」という順序性・直接性は、ノイズがあるデータから厳密に測定することは難しい。
- グレンジャー因果：説明変数にAを利用することで、Bを説明するモデルの精度が改善する場合、 $A \rightarrow B$ の因果がある。

● 因果関係推定の方法をデータの特성에応じて、使い分ける必要があること。

- ノンパラメトリック : データを0、1や上、下のように情報を落として、その記号の出方の連動を調べる。
- 定常性・線形性 : グレンジャー因果関係が代表的。時系列の定常性や線形性が必要で、難易が高い。
- 多変量対応 : 3つ以上の変量の時系列データに対応する。
- 予測モデル構築 : 予測モデルの構築が必要だが、モデルの過学習問題を防ぐことが課題となる。
- 予測モデルなし : 有向情報量など指標を導入して、時系列因果関係を調べる。

※<https://www.hottolink.co.jp/blog/20171010-1>

■ 本講義では、効果検証のような現場に即した部分に注目したいため、特定の因果の定義を定めず、すすめていく。

因果推論分野の動向

- マーケティング施策等の効果分析を中心に、因果推論を中心とした効果分析手法の発展が期待されている。
- 特に、データの複雑化や高速化が進み、機械学習やAIの手法が、因果推論に適用され始めている。
- 深層学習と因果推論を組み合わせ、AIの出力に対して解釈可能性や説明可能性を持たせ、AI技術の高度化に繋げる試みも行われている。

昨今の「因果推論」領域での取り組み

組織	取り組み
理化学研究所 革新知能統合研究センター (AIP)	2017年度センター創設に伴い、因果推論チームを創設 http://www.riken.jp/en/research/labs/aip/generic_tech/cause_infer/
IBM Research	2019年AI業界で注目している3つのトレンドについての発言：相関主体の分析が因果関係へと代わっていく。AIに説明性・解釈可能性を与える重要な技術になるだろう。 https://www.ibm.com/blogs/think/jp-ja/ai-year-review/
Uber	因果推論用のpythonライブラリをオープンソースとして公開。 https://github.com/uber/causalml
CyberAgent AI tech studio	GAN（AIの一種）で因果推論手法を開発 https://adtech.cyberagent.io/research/archives/2283
NEC X	NECのシリコンバレー子会社の3つの注力技術領域のひとつが「因果構造探索」。 https://tech.nikkeibp.co.jp/atcl/nxt/column/18/00141/032300045/ https://jpn.nec.com/press/201806/images/2001-01-01.pdf
Sony Computer Science Lab.	AIによる要因分析サービス「CALC 3.0」をリリース ～ 因果情報に基づく介入・施策効果をシミュレーションする新機能を搭載 ～ https://www.sonyosl.co.jp/press/prs20191106/

因果推論のはじめ

■ 因果分析の有名な事例

- 疫学：感染症の感染源の特定
- コレラのイギリス侵入（1831年10月）当時、コレラは空気感染すると考えられており恐れられていた。しかしスノウは同じ流行地域でも患者が出る家は飛び飛びである等の知見を得て空気感染説に疑問を持ち、「汚染された水を飲むとコレラになる」という「経口感染仮説」を立て、疫学的調査と防疫活動を行った

スノウは患者発生マップと各水道会社の給水地域との比較照合を行い、特定の水道会社の給水地域においてコレラ患者が多発していることを突き止めた。

→地区の井戸（のハンドル）が原因



(参考) 分析疫学での因果の考え方

■ Surgeon General (米国公衆衛生局長諮問委員会) の5基準

- <https://ja.wikipedia.org/wiki/%E7%96%AB%E5%AD%A6>
- 関連の一致性(consistency)
 - 違う国、違う時代でも同じ事が起こるか (人、場所、時間の関連に普遍性があるか)
- 関連の強固性(strength)
 - 効果が定量的か (量-反応関係が成立するか)
- 関連の特異性(specificity)
 - 原因のある所に結果があり、結果のある所に原因があるか
- 関連の時間性(temporality)
 - 原因→結果の順になっているか
- 関連の整合性(coherence)
 - 既知の知識体系と矛盾しないか

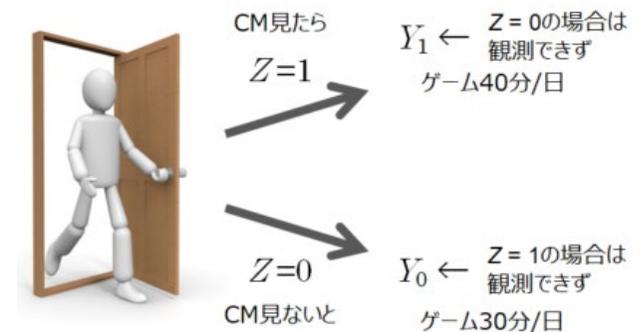
因果推論の利用

因果推論のはじめ

■ 因果分析の最近の事例

■ スマホゲームTVCMの効果測定

- 同一人物に対して、TVCMを見なかった場合と、見た場合を、実験することはできない。
- TVCMを見た人と、見なかった人を、単純に比較することはできない。
- なぜなら、TVCMを見た人は、もともとテレビ視聴時間が長く、スマホ利用度合いが少ない集団に属してる可能性が高い。
- 集団の違いなのかCMの効果なのかを分離するための調整が必要。



観測できるデータはどちらかだけ

出所：<http://www.intageholdings.co.jp/rd/blog/academicprograms/contents201901281400.html>

因果推論のはじめ

■ 因果分析の最近の事例

■ 時系列売上データから、販促施策の効果を測定

- 販促施策を打った場合と、打たなかった場合を実験することができない場合がある。（全国規模の施策など）
- もし施策を打たなかった場合をシミュレートし、実際のデータと比較することで、効果を測定（Synthetic Control法）。

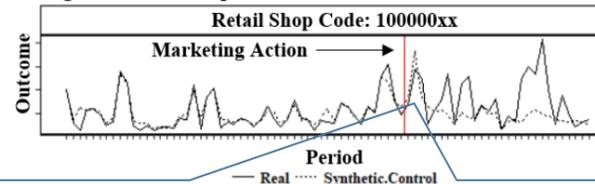
大手消費財メーカーのマーケティング施策
(広告クリエイティブの出し分け)をSCMで解析

既に取得された売上の時系列データにのみ基づいて…

Table1. Data example

Region	Retail Shop Code	Period	Sales
A	10000xx	20xx/xx/x	2,688
A	10000xy	20xx/xx/x	328
B	10000yy	20xx/xx/x	2,952
:	:	:	:

Figure1. SCM example



「もし施策をしなかった場合(ここでは、Aというクリエイティブの広告を出さなかった場合)」のKPI(=売上)をシミュレート。

実線の“Real”に対して、点線の“Synthetic Control”がシミュレート値

* 齊藤・星野・中野(2018, 日本行動計量学会)

出所： <http://www.intageholdings.co.jp/rd/blog/academicprograms/contents201901281400.html>

因果推論のはじめ

■ ワークマンの事例

- 詳細なABテスト（AグループとBグループに分けて実験）設計によって、店舗展開や利益幅の向上に成功。
- 詳細に設計されたデータや専門性のある人が設計したデータ取得プロセスのため、高度な分析者が必要なく効果検証がはやい。

特集 **イノベーション**

デカトン&ワークマン 売り切る技 第4回/全6回

ワークマン大量出店の裏にABテスト 実はデータ 経営企業だった

2019年03月14日 読了時間：9分



酒井 大輔 日経クロストrend 記者

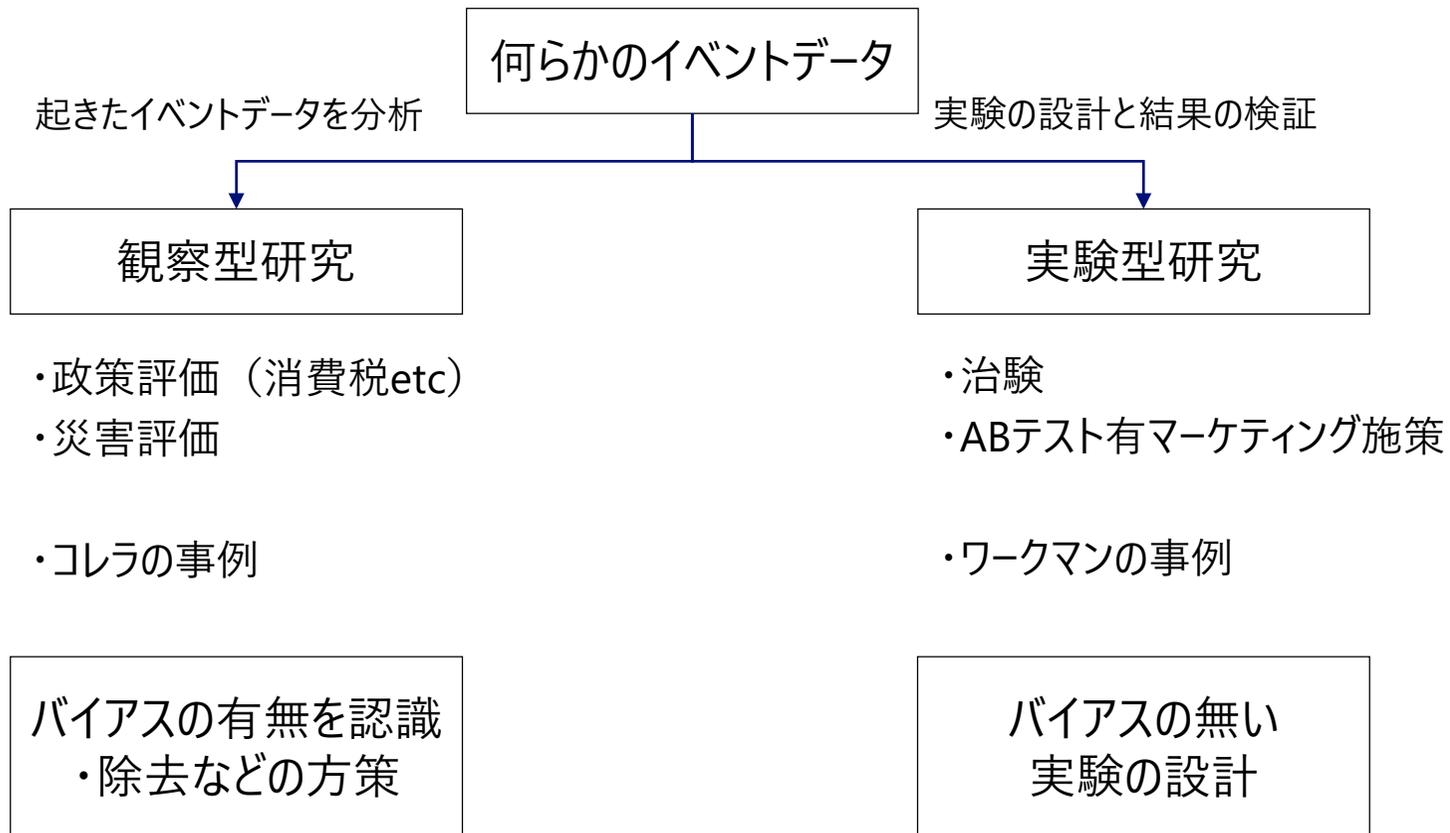


<https://xtrend.nikkei.com/atcl/contents/18/00122/00004/>

バイアスの存在

実際にデータを分析するときには...

- 因果分析を現場で使うときには、2つのアプローチがある。
 - それぞれでデータ分析の品質をあげるために注意する点がことなる。
- 実際のデータ分析の場では、観察型研究が多いと思われる。



分析の品質を高めるためのガイドライン

- 分析や結果の読み取り、他人との共有で必要となることをガイドラインとして作成。
 - 多くはまず観察型として、あたえられたデータで注意することは何か→そもそもそういうデータを作らないようにするためにはという実験型の考えに昇華する。

本編：報告書において記述すべき事項
データに関する事項

- データの取得方法
 - どこから、どのような方法でデータを取得したか
- データの質・量
 - データのサイズ（レコード数、カラム数）
 - 欠損、異常値の存在、割合
 - 観測値・推定値の存在（+正しく計測、推定されているか）
 - 表記ゆれの存在
 - データの型が適切か
 - 各キーとなるデータの正当性（他データとの実合率など）
 - 数量の単位（統一されているか等）
- データの意味把握
 - 選択バイアス、生存バイアスの存在
 - 変数の十分性（交絡）、過剰性（多重共線）
 - リークがあるかどうか（予測）
- データに対して与えた操作・フロー
 - （例）欠損の取り扱い、異常値の取り扱い、データ型の取り扱いなど
- データの基礎集計値
 - データの不均衡（例：0,1Flagデータで1が99%など）
 - 要約統計量
 - ・ データごとに特徴となる量の計算
 - ・ 公差/同時発生/相関

データの質や量に懸念点や不足点があれば、以下についても記載

- ・なぜ存在するのか
- ・どういう意味があるのか
- ・主張の根拠
- ・解決策はあるか。どう方法か。

必ずしも正解があるわけではないので、**手法は共有しないと乱立してしまう**。他グループと足並みをそろえていることも述べる。（これ以降のモデリングや読み取りについても同様）

Copyright (C) Nomura Research Institute, Ltd. All rights reserved. **NRI** 6

分析する前にデータを眺める

■ データに潜むバイアスの存在

- バイアス：データの抽出過程の設計に不備があって、データが特定の傾向に偏ってしまうこと
- バイアスがあることに気付かないことも多い。
- **気付いたとしても、対処できないことがある。**ということを含めて、適切に対処しないと実態にそぐわない結果を導いてしまう

■ バイアスを含むデータ分析で注意するデータ設計

- 生存バイアス
- 選択バイアス

物件番号	顧客番号	契約フラグ	見学日	初見学フラグ	会員種別	住所	年齢	家族数	性別
294712	17733	契約	2018/3/23	1	一般	埼玉県	37	2	男
294712	178208	契約	2018/1/20	1	一般	神奈川県	33	3	男
294712	342810	契約	2018/3/21	1	一般	東京都	64	3	男
294712	386884	なし	2018/2/11	1	一般	埼玉県	52	2	男
294712	501588	なし	2017/9/9	1	一般	埼玉県	42	3	男
294712	508377	なし	2017/11/18	1	一般	埼玉県	35	2	男
294712	540403	契約	2017/8/26	1	一般	埼玉県	28	2	男
294712	713048	なし	2017/10/9	1	一般	埼玉県	33	2	男
294712	805675	なし	2018/3/18	1	会員	東京都	29	3	男
294712	879134	なし	2017/12/10	1	一般	埼玉県	61	3	男
294712	894516	契約	2017/8/19	1	一般	千葉県	38	4	男
294712	900575	なし	2017/8/20	1	一般	埼玉県	36	3	男
294712	994787	なし	2017/8/29	1	一般	埼玉県	70	3	男
294712	999240	なし	2017/9/28	1	一般	埼玉県	45	3	男
294712	1002233	なし	2017/11/21	1	一般	埼玉県	40	4	男

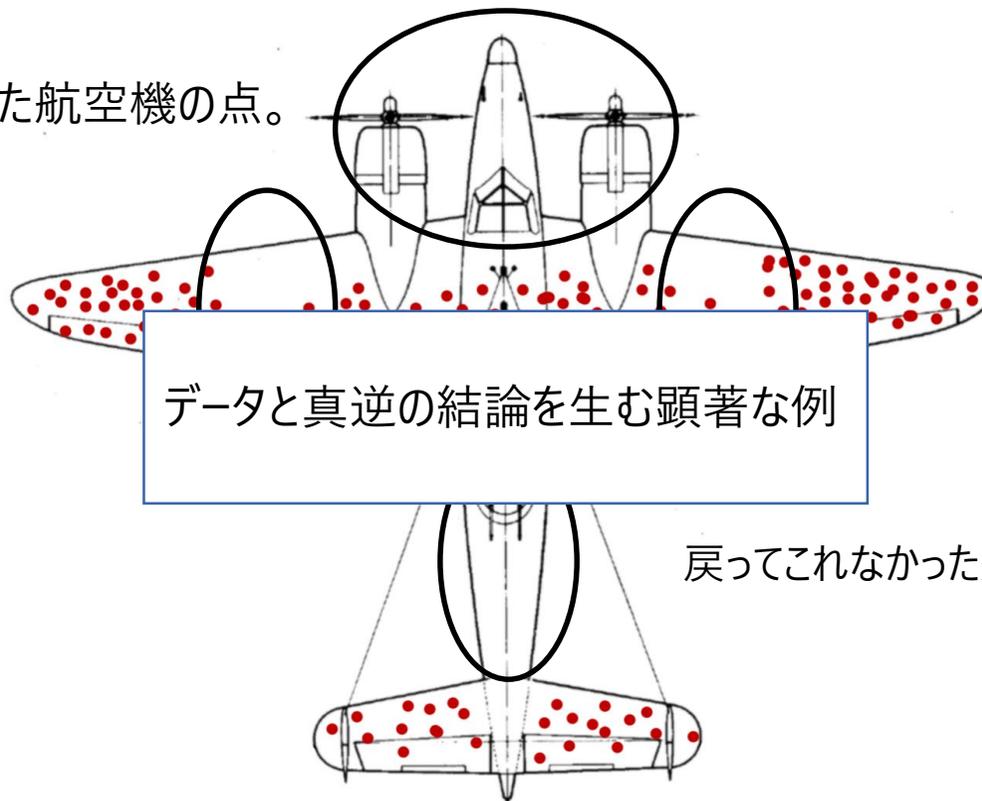
データサイエンティスト



生存バイアス 1

- 第二次世界大戦中、任務から戻った航空機が受けた損傷の研究を行い、最も損傷が多かった部位に装甲を施すよう推奨した。

赤点は戻ってこれた航空機の点。



戻ってこれなかった航空機は黒丸を損傷した？

生存バイアス 2

- 生存バイアスは、データの抽出過程で「失敗」が除外されることで生じる。
 - 抽出過程の例：どのようなデータを取得しようか、どのような指標を作成しようかetc
- 特に失敗側のデータが取得できないことが多いので、非常に厄介。

適切なデータ？

- Webサイトへの集客数アップ施策。
- GoogleAnalyticsデータ？

- 商品の評判を知りたい。
- 購入客のアンケートデータ？

適切な指標？

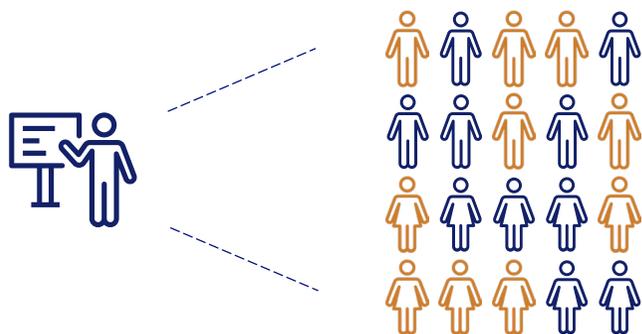
- 10年以上利用している証券会社顧客の平均利益

- 10年以上在籍している営業員の平均成績

- 10年以上上場している企業の平均売上

選択バイアス

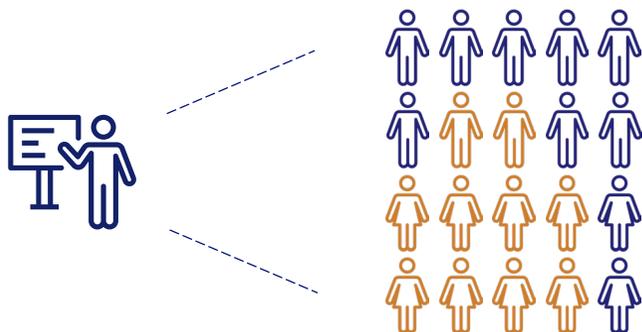
- データから一部を選び出すプロセスで、発生する偏り。
- 施策効果を推定したいというキャンペーンデータの分析で頻発。
 - 理想は、キャンペーン対象は無作為に選ばれてほしい...



単純な比較でOK

ただ、理論上の施策、こんなのない...

- もちろん、実際は**施策効果の大きそうな対象**を優先して行っている。（**選択バイアス**）



単純な比較だと、**施策効果を大きく見積もってしまう**

選択バイアスは技術的に取り除けることも多い。
除去できるバイアスであることを認識して現実的な施策を設計する。

因果推論のはじめ

■ ランダム化比較実験（RCT）

- 実験設計のときには、ABテストなどはRCTを目指して設計される。
 - ただし、そうでないデータや実験がほとんどなので、補正方法などが用意されている。
- 以下の要素がそろっていることが求められる。
- エンドポイント : 効果量に関する客観的指標。
- 比較対照 : 影響を受ける群と、影響を受けない群。
- ランダム化 : 集団からのランダムな抽出と、2つの群へのランダムな割り当て。
- 盲検化 : 実験実施者と被験者は、各個人がどちらの群に属しているかを分らないようにする。

因果効果を測定する方法

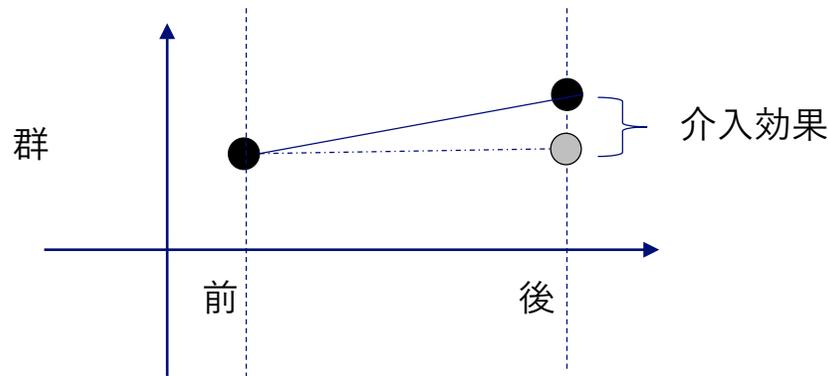
因果効果を測定する方法

- 因果効果を測定するためのアプローチ：デザイン
 - 前後比較デザイン
 - DID
 - 中断時系列デザイン
 - 回帰不連続デザイン
 - 合成コントロール法
- それぞれ特徴的な考えや仮定をおき、効果量を求めている。

因果効果を測定する方法

■ 因果効果を測定する方法

- **前後比較デザイン**：最もシンプルな方法。
- イベントの前後でどのような変化があったか（= 効果量）を差分として求める方法
 - ・ 各個体の**計測対象となる数値をアウトカム**という。
- イベントがなければ、アウトカムに変化はないとの仮定が置かれている。



例えば、N人の被験者がいて、イベント前のアウトカム $y_{i,0}$ とイベント後のアウトカム $y_{i,1}$ が組で与えられるとする。

$$Y_i = (y_{i,0}, y_{i,1})$$

このときの、効果量 τ は

$$\tau = \frac{1}{N} \sum_{i=1}^N y_{i,1} - \frac{1}{N} \sum_{i=1}^N y_{i,0}$$

因果効果を測定する方法

■ 因果効果を測定する方法

例えば、N人の被験者がいて、イベント前のアウトカム $y_{i,0}$ とイベント後のアウトカム $y_{i,1}$ が組
で与えられるとする。

$$Y_i = (y_{i,0}, y_{i,1})$$

このときの、効果量 τ は

$$\tau = \frac{1}{N} \sum_{i=1}^N y_{i,1} - \frac{1}{N} \sum_{i=1}^N y_{i,0}$$

介入後の平均アウトカム 介入前の平均アウトカム

$$Y_1^t$$

$$Y_0^t$$

- τ はとくに平均処置効果という。

$$\tau = Y_1^t - Y_0^t$$

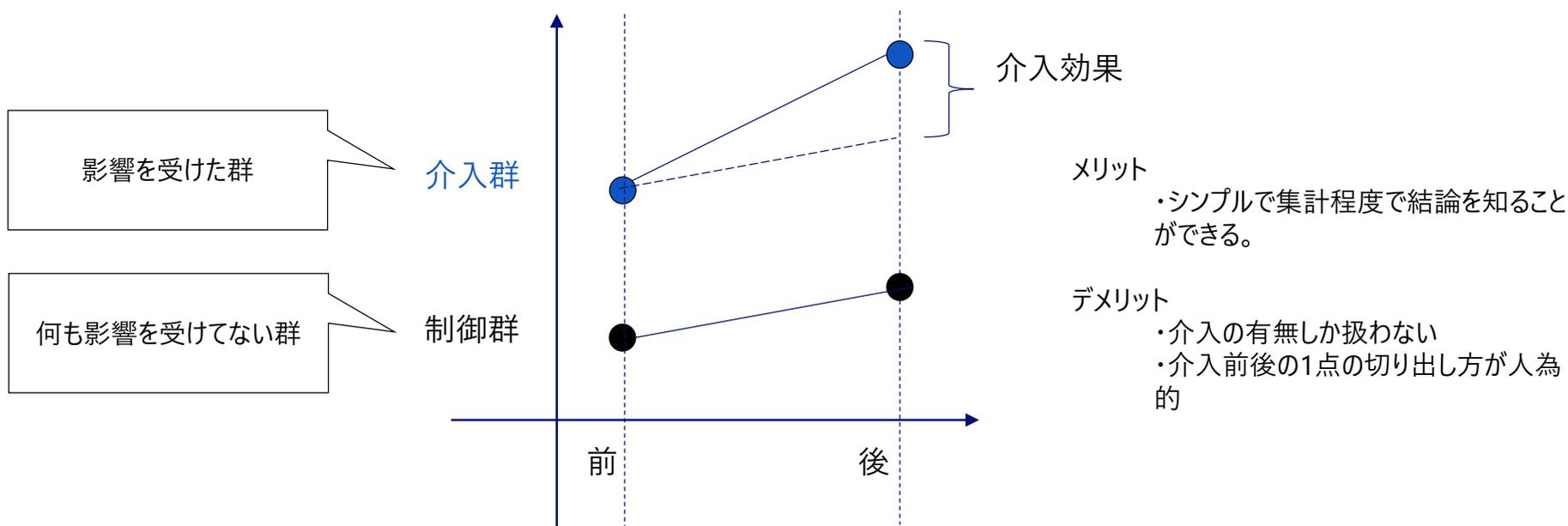
実験以外・データにはない環境の変化による影響が除去できていないので
正確なイベントの効果量ではない。

因果効果を測定する方法

■ 因果効果を測定する方法

- 差分の差分法：ABテストの最もシンプルな形式。
- 点線部分：並行トレンド仮定がある。

「差分の差分(DID法)」比較

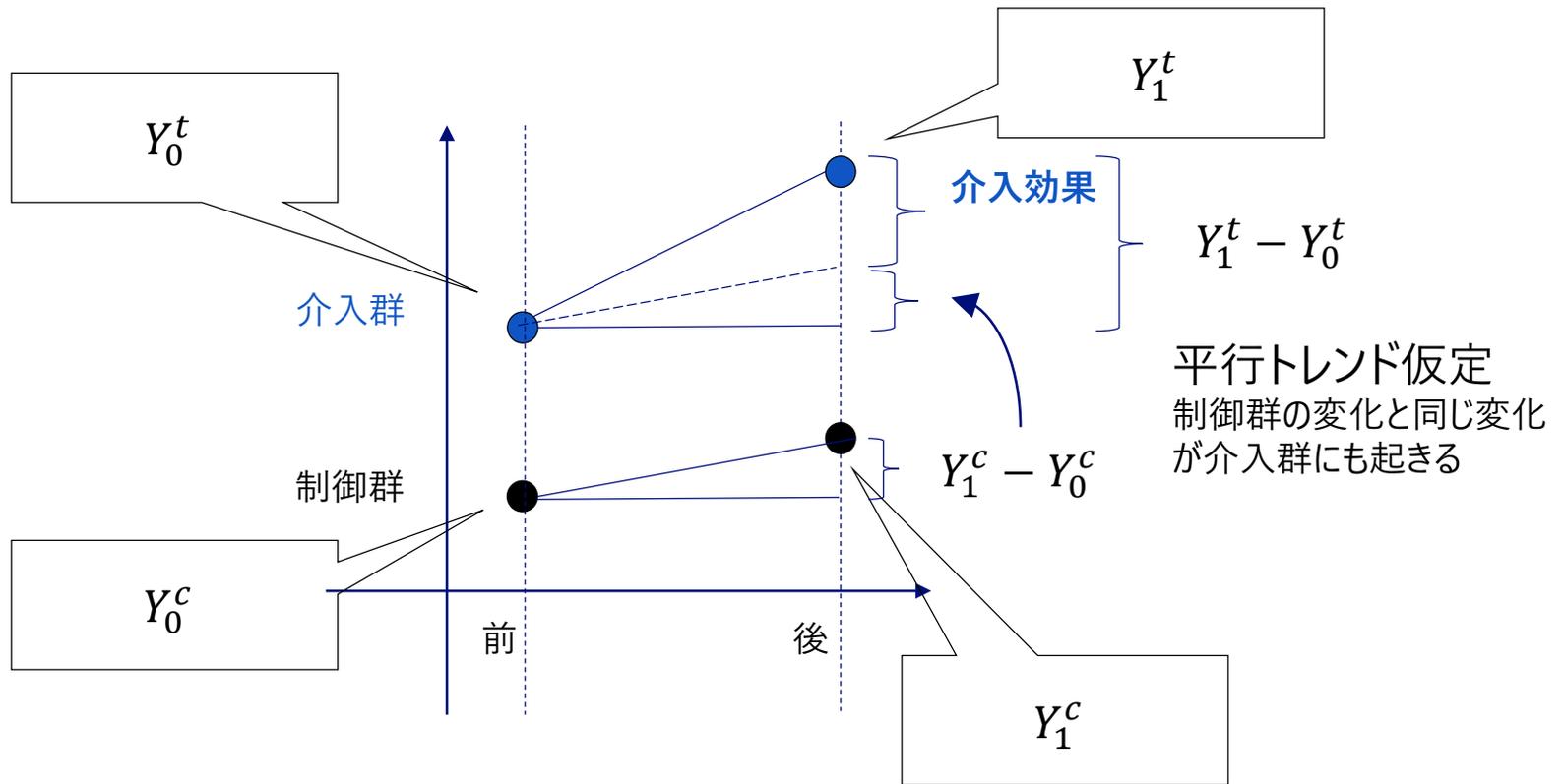


群を表す2点とその介入前後を表す計4点から効果量を計算する。

因果効果を測定する方法

■ 因果効果を測定する方法

- それぞれ平均アウトカムとすると、差分の差分法における平均処置効果は次のようになる。



因果推論のはじめ

■ 結果の比較

- 差分の差分法では、制御群の変化を差し引くことで、環境（マクロトレンド）の変化による前後変化を見積もることができる。
 - ・ ただし、2群の間で同じトレンドがあるという仮定（平行トレンド仮定）はある。

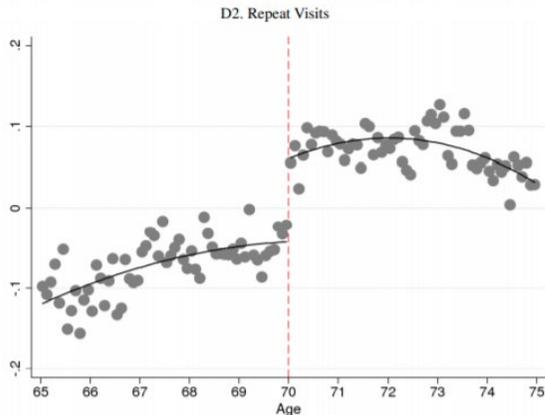
方法	効果量
前後比較デザイン	$\tau = Y_1^t - Y_0^t$
差分の差分法	$\tau = (Y_1^t - Y_0^t) - (Y_1^c - Y_0^c)$

- 因果推論は、本当の効果は何なのかを適切に抜き出す手法の集合であり、個々で利用する統計手法が特別に難しいというわけではない。（今の段階では...）
- データの中、あるいは比較対象のなかにどのような効果が紛れ込んでいるかを気にかけていく必要がある。

因果推論のはじめ

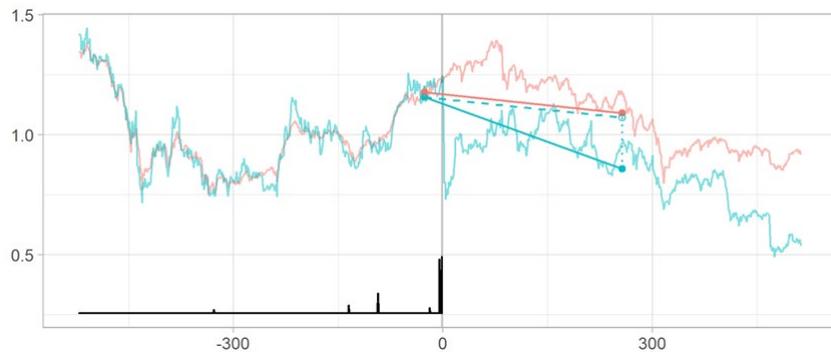
■ その他の測定方法（観察型研究）

● 不連続時系列デザイン



Shigeoka NBER 19726

● 反事実推定による方法



前後比較デザインの発展

- ・時系列モデルでのフィッティングとともに介入時の「ギャップ」から効果量を推定する。
- ・制御群データがないときに行うことが多い。

起きてない事実を推定する方法

- ・合成コントロール法
- ・イベントが仮に起きなかったとしたときの時系列を推定する。（ニュースの株価へのインパクト推定 etc）

差分の差分法の回帰分析での取り扱い

- 実際には、パネルデータが与えられて差分の差分法を行うことになる。
- イベント前後、制御群or対照群のカテゴリ変数とその交互作用項（掛け算項）を導入したデータを作る。

アウトカム	イベント前後E 前=0、後=1	制御or対照G 制御=0、対照=1	E×G
200	0	0	0
250	1	0	0
210	0	0	0
235	1	0	0
250	0	1	0
280	1	1	1
...			

つぎの交互作用項を含むモデルの回帰によって、 β_3 が効果量 τ として求めることができる。

$$\text{アウトカム} = \beta_0 + \beta_1 \times \text{イベント前後} + \beta_2 \times \text{制御対照フラグ} + \beta_3 \times \text{イベント前後} \times \text{制御対照フラグ}$$

DID法を利用した主要な成果について

■ ニュージャージーにおける最低賃金の引き上げに関する研究

- 最低賃金の引き上げに際して、失業率が上がったのではないかという政策評価研究
- ニュージャージーとペンシルベニアを比較して、景気が下降トレンドにあった時期において、フルタイム等量にプラスの効果をもたらしていたことがわかった。

Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania

By DAVID CARD AND ALAN B. KRUEGER*

On April 1, 1992, New Jersey's minimum wage rose from \$4.25 to \$5.05 per hour. To evaluate the impact of the law we surveyed 410 fast-food restaurants in New Jersey and eastern Pennsylvania before and after the rise. Comparisons of employment growth at stores in New Jersey and Pennsylvania (where the minimum wage was constant) provide simple estimates of the effect of the higher minimum wage. We also compare employment changes at stores in New Jersey that were initially paying high wages (above \$5) to the changes at lower-wage stores. We find no indication that the rise in the minimum wage reduced employment. (JEL J30, J23)

効果量の測定手法について

- ランダム化比較実験で効果量を測定することが望ましく、DIDは最もシンプルな方法である。
- ここでは、理想的なデータにおいてDIDで効果量が簡単に抽出することができた。
 - 一方で、DIDであれば選択バイアスを、群比較することである程度除去できることが分かった。
 - 平行トレンドが成立しないような選択が行われると、バイアスの影響が出る。
- 実際に行う上での難しさとは...
 - RCTを担保することは非常に難しい。
 - ・ 影響を受ける群と、影響を受けない群を用意することが難しい。
 - ・ 集団からのランダムな抽出と、2つの群へのランダムな割り当て。
 - ・ 改善幅が大きいところを集中してほしいという現場の要望
 - 実験通りに従事してくれない被験者が一定割合いる。
- 以上のようなデータにはバイアスが残存し、分析結果をゆるがす。
 - この後、バイアスがどのようにデータから見えるか、修正するかを講義する。

The text is framed by two decorative swooshes. The top swoosh is a gradient bar transitioning from blue on the left to red on the right. The bottom swoosh is a solid blue bar.

Share the Next Values!