

第8回 因果推論
相関関係と因果関係
計量経済学(2)

@東京工業大学 倪彬

本日の予定

- ▶ 因果推論
 - ▶ これまでの分析手法とエビデンス
 - ▶ 相関関係と因果関係
 - ▶ 因果推論と反事実

これまでの分析手法とエビデンス

- ▶ 統計学で用いられる代表的な分析手法
 - ▶ 単純集計・クロス表(統計学の基礎・経営統計学)
 - ▶ 散布図・相関係数
 - ▶ 回帰分析
- ▶ 何か議論や主張をしたい場合の根拠、**エビデンス**として有用
- ▶ エビデンス(Evidence)とはそもそも何？
 - ▶ Web辞典(英辞郎)によると、『証拠、証言』
 - ▶ 実務では、実行した証拠である資料などを指すことも
- ▶ 本来の意味としては、『**科学的根拠**』を指す
 - ▶ ある主張について、データを用いて、その主張を裏付ける結果

これまでの分析手法とエビデンス

▶ 誤解①: 数字やデータ分析を掲載する＝エビデンスではない

(百年の礎を築く) - [戦略14] 熊本アカデミズム～「知」の集積を「地」の活力につなげます～

(1)現状と課題		(2)概要		(3)施策体系		(4)県民アンケート結果	
<p>本県は、多くの大学が立地し、生命科学や半導体分野の企業や技術の集積もあります。このような特色を生かし、産学官連携により優秀な人材や技術を集積し、研究開発を活性化させる取組みが必要です。また、グローバル化が進化する中、将来の活力を生み出す若者が海外で学ぶチャンレンジへの支援などが求められています。さらに、熊本が留学生から選ばれるようなまちづくりを積極的に進める必要があります。</p>		<p>熊本には多くの大学が立地し、生命科学や半導体などの分野での企業や技術の集積もあります。このような特色を生かし、大学や企業の研究開発を活性化させる取組みを進めます。また、県内の大学などに海外から多くの留学生が集まる、世界に開かれた活気あるまちづくりを進めます。さらに、夢を持ち海外へ挑戦する若者を支援し、グローバルな人材を育成するなど、「知」の集積を図り、それが「地」の活力となって世界とつながり発展する熊本づくりを進めます。</p>		<p>◆「知」を集める ～世界的な知の集積～ 【担当部局:企画振興部・商工観光労働部】</p>		<p>【満足度】</p> <p>【今後の方向性】</p>	
<p>★戦略指標(単位)</p>		策定時	H24	H25	H26	目標値	
i 研究開発部門の企業立地件数 (件)		9 (H20～H23)	6 (<60.0%)	12 (<120.0%)	18 (<180.0%)	10 (件/4年(累計))	
ii 海外高校への留学生数 (人)		10 (H23)	19 (<19.0%)	34 (<34.0%)	44 (<44.0%)	100 (人/4年(累計))	
英語の学習が「好き」と回答した生徒 (中1～中3)の割合 (%)		46.9 (H23)	48.4 (<+1.5)	50.1 (<+3.2)	52.3 (<+5.4)	毎年度、前年度の割合を上回る [%/(毎年)]	
<p>企業の研究開発部門の新規の誘致及び既立地企業への研究開発部門の併設の働きかけを重点的に行い、結果、目標を大きく上回った。</p>		<p>◆研究開発部門と大学院を誘致する ～産学官連携による最先端技術の集積～ 【担当部局:総務部・企画振興部・環境生活部・商工観光労働部】</p>		<p>◆留学生から「選ばれる」まちを創る ～留学生とつながる仕組みづくり～ 【担当部局:企画振興部・商工観光労働部】</p>		<p>【満足度】</p> <p>【今後の方向性】</p>	
<p>本県の「留学生の受け皿」としては、海外大学進学や海外高校留学に対する意識は、まだ高くないと認識されるが、少しずつ、留学説明会や高校生モンタナ派遣事業(短期派遣)に加え、長期留学の促進の取組みの成果が出ている。</p>		<p>◆留学生から「選ばれる」まちを創る ～留学生とつながる仕組みづくり～ 【担当部局:企画振興部・商工観光労働部】</p>		<p>【満足度】</p> <p>【今後の方向性】</p>			
<p>英語音声O.D「I CAN DO IT!」や英語読み物資料「The Spirit of KUMAMOTO」、夏には「くまモン英語チャレンジ」などの取組により、「好き」は2.2ポイント増加した。</p>		<p>【満足度】</p> <p>【今後の方向性】</p>					

▶ 何も考えずに数字やアンケート集計結果を載せても、主張を支える確固たる「根拠」にはならない

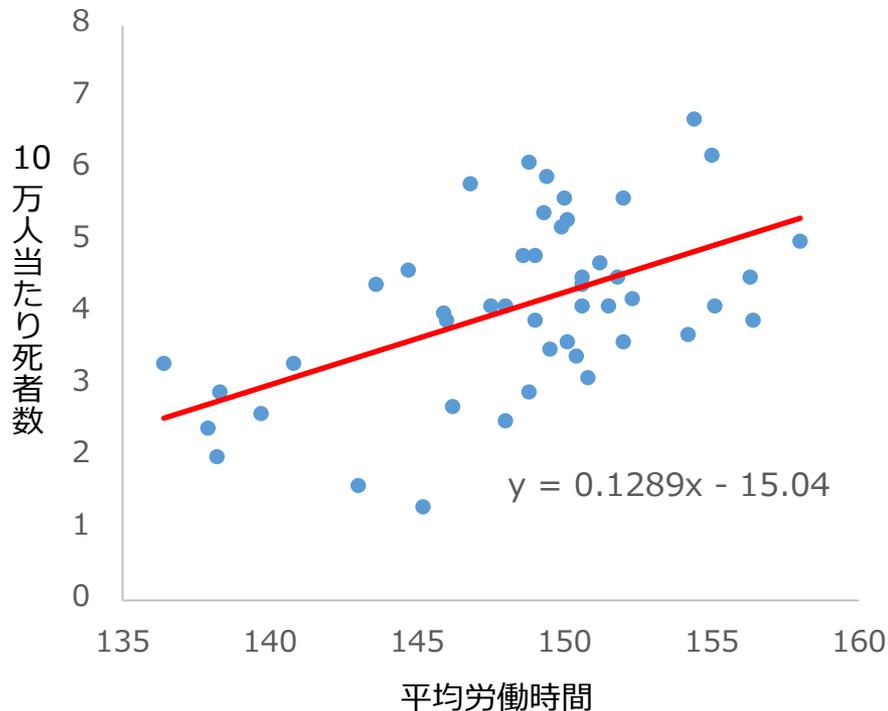
▶ エビデンスとして客観的な裏付けを示すには？

→ これまで学んできた分析手法が使える

これまでの分析手法とエビデンス

▶ 一次関数を用いて関係性を定式化する回帰分析

例：労働時間と死者数(47都道府県)

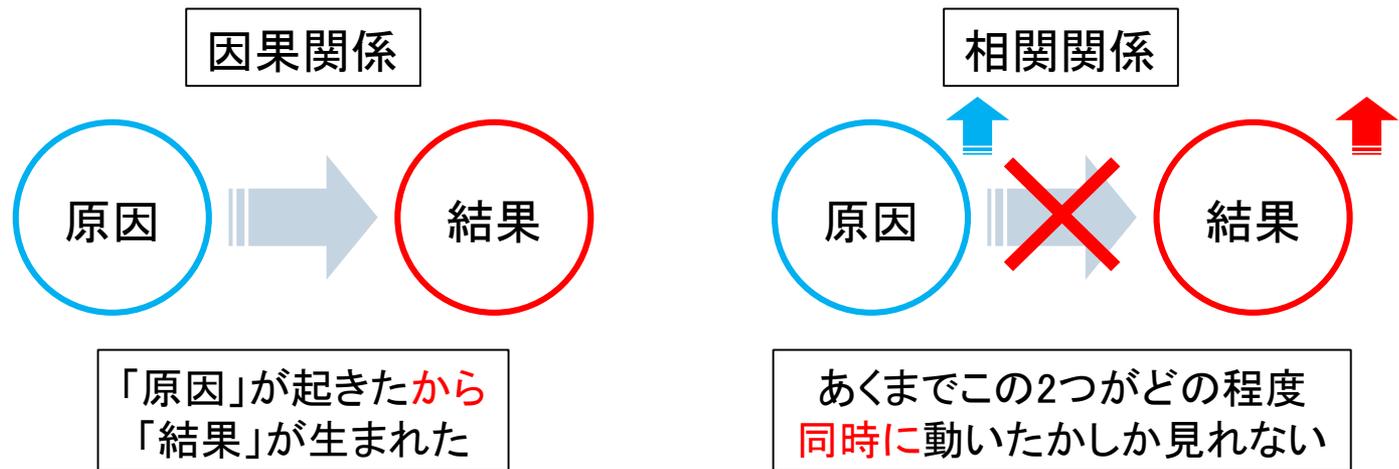


- ▶ 関係性の大きさ(傾き)を考慮
- ▶ 真のパラメータが0であるかどうかについて推論可能
- ▶ 複数の要因を同時に考慮できる形(重回帰分析)に拡張可能
- ▶ 幅広く応用でき、計量経済学分野では回帰分析を用いた研究が最も多い

→ 2つ以上の事柄の関係性・メカニズムの議論に便利かも

相関関係と因果関係

- ▶ これらの手法が明らかにしているのは、いずれも**相関関係**
- ▶ 誤解②: **相関関係は因果関係を意味しない**
 - ▶ 施策や政策の有効性を議論・主張するのに最も信頼性が高く使えるのは**因果関係**

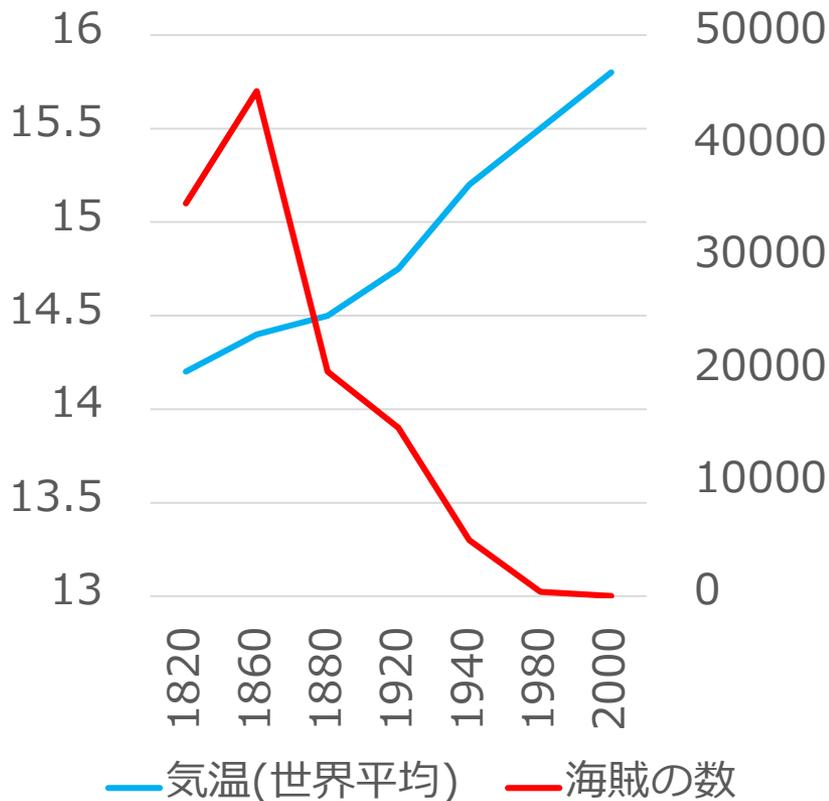


- ▶ 相関関係も材料としては使えるが、因果関係の解明が最も重要
- ▶ ここまでの手法では、因果関係を示すには不十分
 - ▶ 回帰分析でも、「原因」と「結果」をイメージはするが、**因果関係の有無までは主張するには弱い**

相関関係と因果関係

▶ なぜ「相関関係」だけではまずいのか？

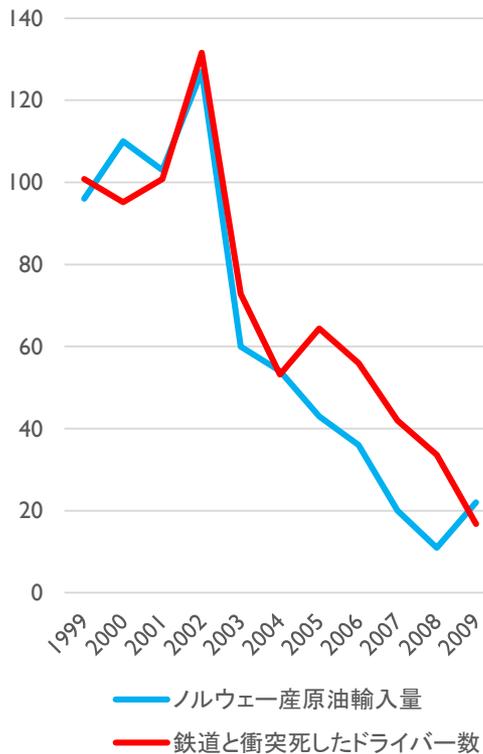
1. ただの偶然の可能性



- ▶ 海賊の数が減ると同時に、気温が上昇
- 相関関係あり
- ▶ 海賊が減ったから、気温が上がったとは考えられない
- 因果関係あり？
- おそらく偶然
- ▶ 海賊の数が減ると海面水位が変化して...とかいう「こじつけ」「トンデモ説」が生まれたりしている

相関関係と因果関係

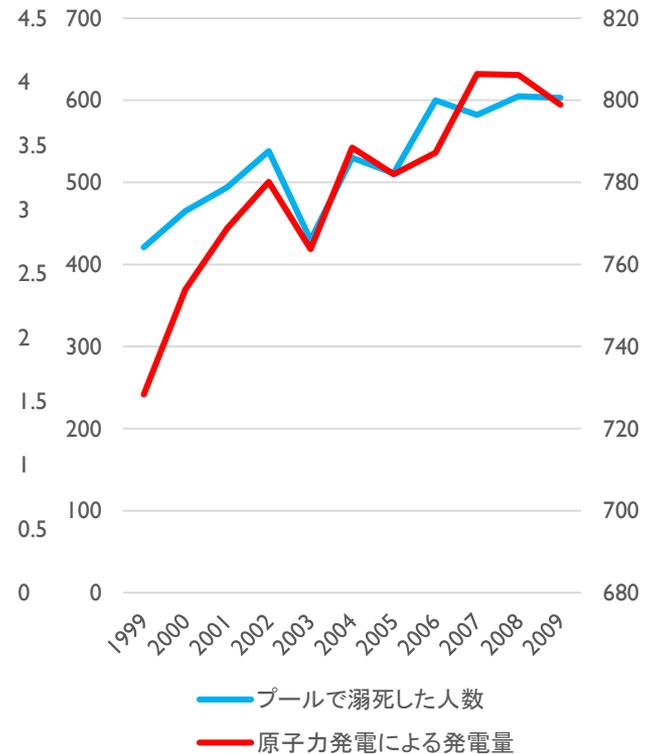
- ▶ 他にも以下のようなものが
- ▶ 青が左軸、赤が右軸



$r = 0.95$



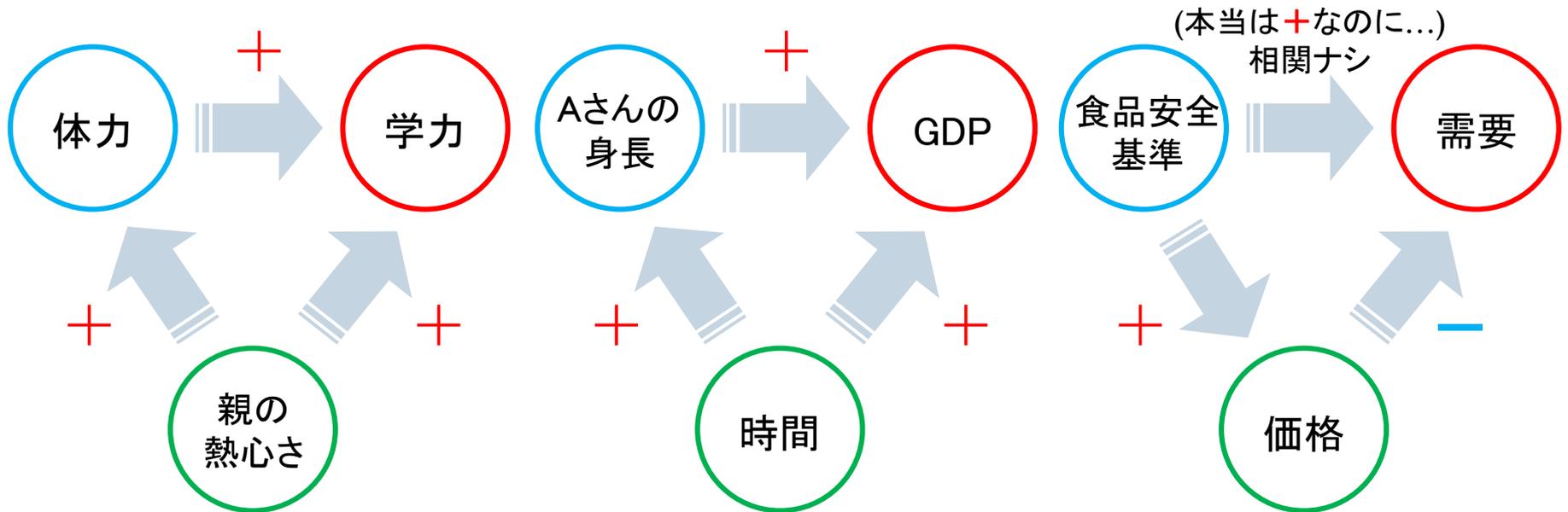
$r = 0.67$



$r = 0.90$

相関関係と因果関係

2. 第3の要素(交絡因子)が介在している可能性



▶ 関係がありそうに見える(疑似相関)可能性や、相関を打ち消す場合がある

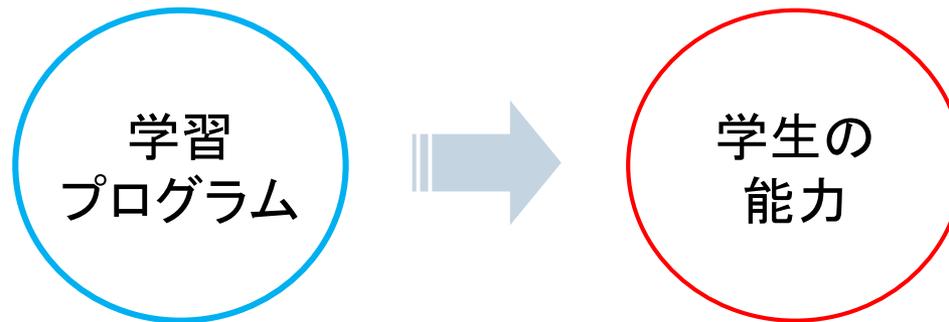
→ いずれも、相関関係から因果関係を考えるのを邪魔している

Actively learning (1)

- ▶ 以下は真の相関/疑似相関か、そして疑似相関の場合、間にある交絡因子はどのようなものが考えられるかを述べなさい。
 1. 1年生から6年生までの複数学年で調査し、身長が高い人ほど数学の点数が高かった
 2. 仕事のやりがいについてアンケートを行い、自分の仕事は専門性が高いと回答している人ほど、仕事に対する満足度が高かった
 3. 授業を多く履修している人ほど、GPAが高かった

相関関係と因果関係

3. 本当の因果関係は実は逆である可能性



- ▶ 任意の学習プログラム参加によって、学生がより優秀になったように見えるが...
- ▶ 優秀な学生ほど、自主的に学習プログラムへ参加しやすい
- 学生の能力の高さは学習プログラムのおかげ？
 - ▶ 逆の因果関係しかない場合もあるし、双方向に影響しあっている可能性もある

相関関係と因果関係

- ▶ **誤解②: 因果関係の検証に、ビッグデータも高度なテクニックも不要**
 - ▶ 近年流行しているビッグデータは、サンプルサイズが膨大で様々なことができる
 - ▶ 本当は母集団における真のパラメータを知りたいが、一部データを用いて、その結果から真のパラメータを推論(統計的有意性)
 - ▶ ビッグデータを用いれば、ほぼ母集団に近いサンプルを用いることができるので、統計的有意性を用いた推論を経ずとも母集団の特性を考えることができる
 - ▶ **膨大なデータがあっても、扱い方がわかっていない(先ほどの問題の可能性)と無用の長物**
 - ▶ むしろ情報の「ゴミ」が多すぎて、別途高度な知識や設備が必要になる

因果推論と反事実

- ▶ これらに対処して、因果関係を発見を試みる
 - **因果推論**
 - ▶ 近年社会科学で注目されている解決方法
 - 「もし〇〇が起きなかったら」というifの状況、**反事実(counterfactual)**を想定して比較する(**ルービンの因果モデル**)

- ▶ 例：広告の売上効果



企業が2016年に広告を出した → 売上1500万円
+500万円



もし(if)同じ企業が
2016年に**広告を出さなかったら** → 売上1000万円

- 事実と反事実の比較であれば、その違いである事柄によって引き起こされたものとして、因果関係を主張できる

因果推論と反事実

- ▶ 反事実によって先ほどの問題は解決するか？
 - ▶ 第3の要素がある(問題2)のでは？因果関係が逆(問題3)では？
 - 同じ人・主体のifなので、他要素は同じ条件＝無視できる
 - ▶ 偶然(問題1)では？
 - なるべく多くの人・主体で調査して、平均的な効果で検証すればクリア

- ▶ 唯一にして最大の弱点
 - 事実は観測できるが、反事実は観測できない

因果推論と反事実

- ▶ 例：メタボ健診と平均寿命
- ▶ 本来欲しい情報

個人番号	メタボ健診を受診		因果効果 (A-B)	
	した場合の寿命 (A)	していない 場合の寿命 (B)		
受診した	1	85	80	5
	2	79	75	4
	3	83	77	6
していない	4	79	74	5
	5	73	69	4
	6	77	71	6
平均	79.3	74.3	5	

因果推論と反事実

▶ 実際に入手可能な情報

個人番号	メタボ健診を受診		因果効果 (A-B)	
	した場合の寿命 (A)	していない 場合の寿命 (B)		
受診した	1	85	?	?
	2	79	?	?
	3	83	?	?
していない	4	?	74	?
	5	?	69	?
	6	?	71	?
平均	82.3	71.3	11?	

- ▶ 観測できない反事実部分(???)を、もっともらしい値で穴埋めする必要がある

因果推論と反事実

- ▶ **誤解③: 前後比較やありもののデータの単純比較では反事実との比較にならない**
 - ▶ 前後比較やグループ間の単純比較は、突き詰めるとこの反事実比較を行いたくてやっているはず
 - ▶ 施策前や対象外グループは、反事実の穴埋め役として適切か？
- 多くの場合で、反事実として不適切な特性・違いを持っている

Active learning (2)

- ▶ 以下の問題意識について、反事実として不適切なものを
用いているが、なぜ不適切か自分の考えを述べなさい。
 1. 目的: 広告効果の検証
反事実: 広告実施前データ

 2. 目的: メタボ健診の効果の検証
反事実: 未受診の人たちの健康状態

因果推論と反事実

▶ 穴埋めの方法

1. 実験によって反事実を作り出す

→ ランダム化比較試験(Randomized Controlled Trial, RCT)

2. 施策の前後比較を応用して計算する

→ 差の差法(Difference In Difference, DID)

▶ 他にもいくつか手法があるが、共通している目標は、比較可能なグループを作り出し、反事実をもっともらしい値で置き換えること

▶ 少なくとも上記2つは、考え方や概念を理解するだけで、実施・応用が比較的容易

因果推論と重回帰分析

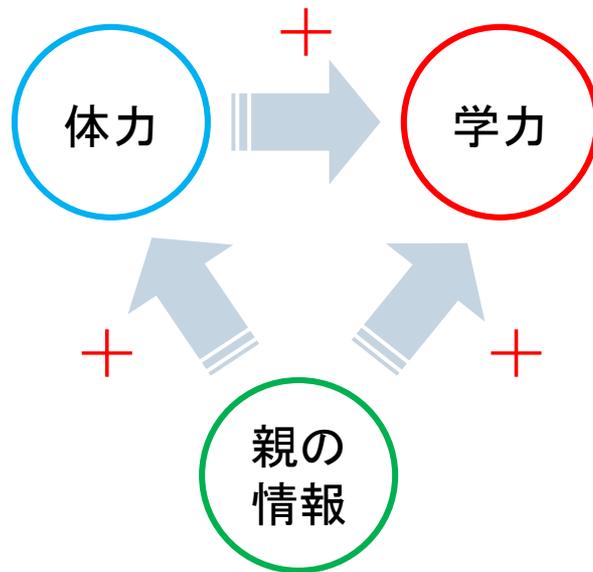
- ▶ 因果推論への最短ルートは、反事実を扱うこと
- ▶ 因果推論のために収集されたわけではない「ありもの」のデータでは、なかなか反事実を扱うのは難しい
 - ▶ RCTは専用の実験を行う必要がある
 - ▶ DIDは仮定を満たすデータが見当たらないことが多い

因果推論と重回帰分析

- ▶ 因果推論への最短ルートは、反事実を扱うこと
 - ▶ 因果推論のために収集されたわけではない「ありもの」のデータでは、なかなか反事実を扱うのは難しい
 - ▶ RCTは専用の実験を行う必要がある
 - ▶ DIDは仮定を満たすデータが見当たらないことが多い
 - ▶ 相関関係から因果関係を考えるのを邪魔している3つの理由のうち、最もやっかいなのは**第3の要因(交絡因子)の存在**
 - ▶ 少なくとも交絡因子だけでも取り除いて、相関関係 \doteq 因果関係に近づきたい
- 重回帰分析の出番

因果推論と重回帰分析

▶ 例: 体力と学力の関係



▶ 単回帰分析

$$\text{学力} = \alpha + \beta \text{体力}$$

→ おそらく β は有意となるが、この2つが同時に動いているだけで、因果関係は主張できない

▶ 重回帰分析

$$\text{学力} = \alpha + \beta_1 \text{体力} + \beta_2 \text{親の情報} \dots$$

→ 交絡因子を取り除いてなお、 β_1 が有意であれば、体力から学力への因果関係の説得力が増す

▶ 交絡因子を完璧に取り除けるのはレアケースだが、交絡因子を取り除くためにアンケートを事前に丁寧に練る、あるいは積極的に多くの変数を追加するのを強く推奨

因果推論まとめ

- ▶ 何か主張や議論を行いたい、施策を行いたい場合は、**因果関係を示せるエビデンス**を意識しながら制度設計や評価を行うことが理想
 - ▶ エビデンスには様々あるが、従来型のものはとても因果関係を主張できるものではない
 - ▶ 厳密な費用対効果が欲しい場合や、失敗できない場合の調査としては不十分
- ▶ 因果関係を考える上で重要なのは**反事実**
 - ▶ 反事実を求めるため、RCTやDIDが有用
 - ▶ 特にRCTは本格的にやろうとすると、アンケート構築や調査対象、協力自治体、資金面など多くの制限があり、なかなか難しい
 - ▶ いずれも考え方そのものは難しくないので、それぞれのゼミ・団体の活動にも取り入れることが可能
 - ▶ 因果関係に近づく次善の策として、交絡因子を取り除ける**重回帰分析**も使える
- ▶ イベントやプロダクトを実施・作成するだけ、あるいは形だけのアンケートを行って満足するのではなく、以上のような考えに基づいて工夫しながら効果の検証を行うことで、より成果や主張の有効性・説得力が増す

これから自学するために

▶ 自分で学ぶ際にオススの書籍

1. 中室牧子・津川友介、『「原因と結果」の経済学』、2017年、ダイヤモンド社
2. 伊藤公一朗『データ分析の力 因果関係に迫る思考法』、2017年、光文社新書
3. 『岩波データサイエンス Vol.3 [特集]因果推論』、2016年、岩波書店

▶ 最新の政策評価の動向に関する資料

1. 朝井・神林・山口(2016)「保育所整備と母親の就業率」『経済分析』第191号、内閣府社会総合研究所
2. 家子・小林・松岡・西尾(2016)「エビデンスで変わる政策形成—イギリスにおける「エビデンスに基づく政策」の動向、ランダム化比較試験による実証、及び日本への示唆—」三菱UFJリサーチ&コンサルティング 政策研究レポート
3. 伊藤公一朗(2016)「政策の効果をどう測定するか？海外における「エビデンスに基づく政策」の最新動向」経済産業研究所(RIETI)セミナー資料
<http://www.rieti.go.jp/jp/events/bbl/16102501.html>
4. 小林庸平(2016)「エビデンスを政策にどう使うか—イギリスの動向と日本への適用事例からの示唆—」経済産業研究所(RIETI)セミナー資料

おわりに

- ▶ 計量経済学で使う手法に関して
 - ▶ OLSが適切でない場合の別の手法、ここから先もまだまだ広がる学問
 - ▶ 重回帰分析とRCT・DIDの合わせ技などもある
 - ▶ 最近では機械学習・ディープラーニングとの連携も出てきている

- ▶ 本気でもっと勉強したい人、より正確な手法を学びたい人は、別途相談すること

Last but not least

Stay hungry, stay foolish.

—by Steve Jobs at the
commencement of
Stanford University,
2005