

データ分析を経営や政策に生かすには？ 「因果分析」と「予測」の適切な使い分け

伊藤公一朗

シカゴ大学公共政策大学院ハリススクール助教授

Email: ito@uchicago.edu

Web: www.koichiroito.com



THE UNIVERSITY OF
CHICAGO

本日お話したい内容

- 1) NHKスペシャル「AIに聞いてみた」から見る、データ分析への危険な誤解
- 2) データ分析で明らかにできる2つの（全く異なる）こと
 - 因果分析—AIだけでは解決が難しい
 - 予測—AIが威力を発揮
- 3) 企業や政府内で始まっているデータ分析の活用事例
- 4) データ・サイエンティスト争奪戦が始まっている米国の労働市場
 - 遅れを取る日本はどのようにデータ・サイエンティストを育成し活用するのか

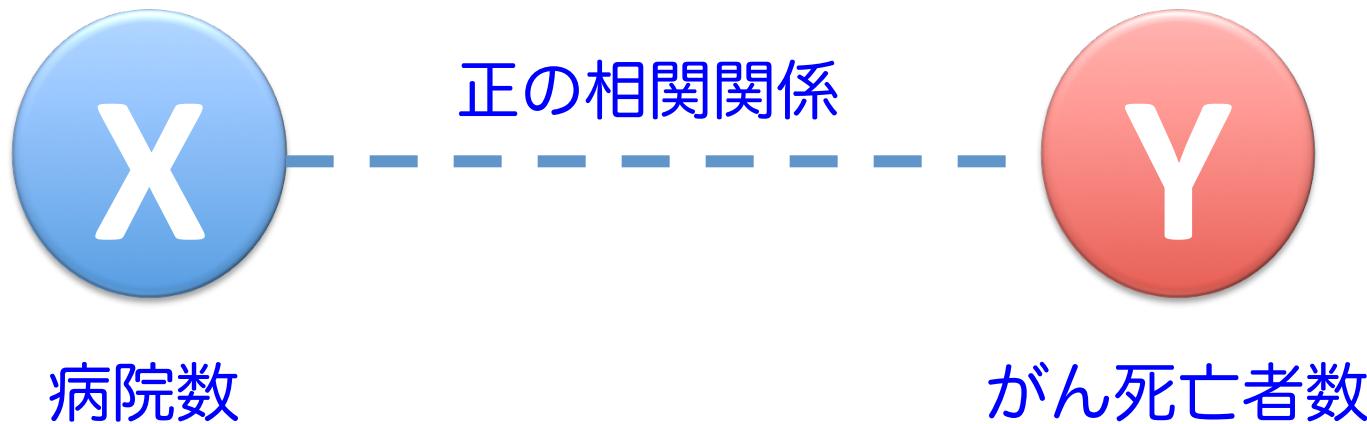
NHKスペシャル「AIに聞いてみた」



画像出典：<https://www.photo-ac.com>

- 人工知能（AI）「ひろしくん」がビッグデータを解析する画期的番組
- ところが、放送後、専門家からは「データ分析結果の解釈に誤りがある」との批判が相次いだ。その理由は？

- 人工知能（AI）が得意なのは「相関関係」の分析（少なくとも今のところ）
 - ある物事（X）とある物事（Y）がどのように「相関」しているかを解明
 - 番組では「病院数（X）」と「がん死亡者数（Y）」に正の相関関係を発見
 - 人工知能ひろしくんが行ったここまで分析に根本的な誤りはない

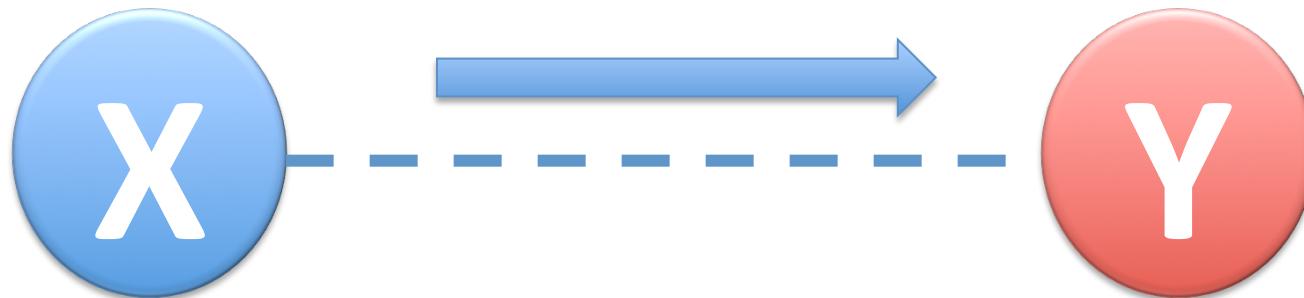


- この相関関係分析から言える正しい解釈は：
 - 「病院数が多い場所では、がん死亡者数が多い傾向がある」
 - 「がん死亡者数が多い場所では、病院数が多い傾向がある」

- 問題は、番組制作側が「相関関係」を「因果関係」として解釈していること
 - 「Xが起こると、Yが起こる」という日本語は因果関係を意味する
 - 「Yが欲しければ、Xをすべき」という日本語も因果関係を意味する
- 番組では相関関係分析結果をもとに、以下のような「提言」が行われた
 - 1) 健康になりたければ、病院を減らせ
 - 2) 少子化を食い止めるには、結婚よりもクルマを買え
 - 3) ラブホテルが多いと、女性が活躍する
 - 4) 男の人生のカギは、女子中学生のぽっちょり度
 - 5) 40代一人暮らしが、日本を滅ぼす

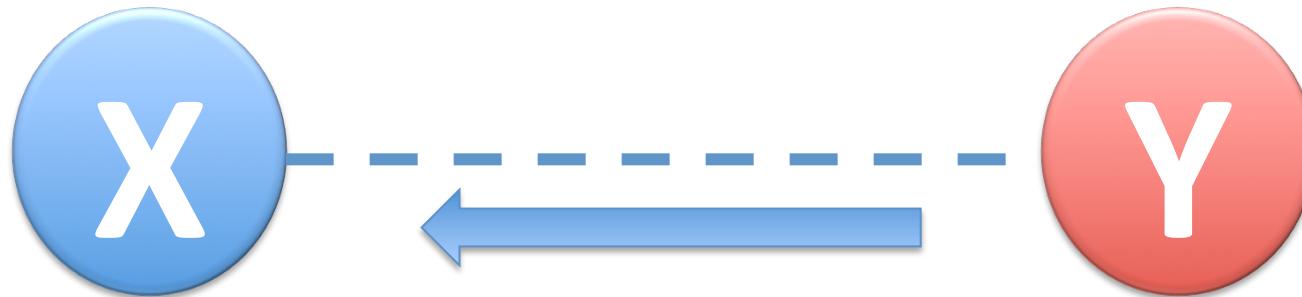
注：私は子供の頃からNHKスペシャル大好き少年だったので、特にこの番組だけを批判する意図はありません。むしろ「相関」を「因果」と誤解釈してしまう傾向は、新聞・雑誌・テレビ・ネットを含めた日本のメディア全体で見受けられる、より大きな問題です。

社会科学データで「causality（因果関係）」を言うのはとても難しい



- XとYのデータが、一緒に動いている様子が観察できたとする（相関関係）
- 多くの場合、私たちは相関関係を見ると、以下のことを言ってしまいがち
- しかし、これは「3つの可能性のうちの1つでしかない」

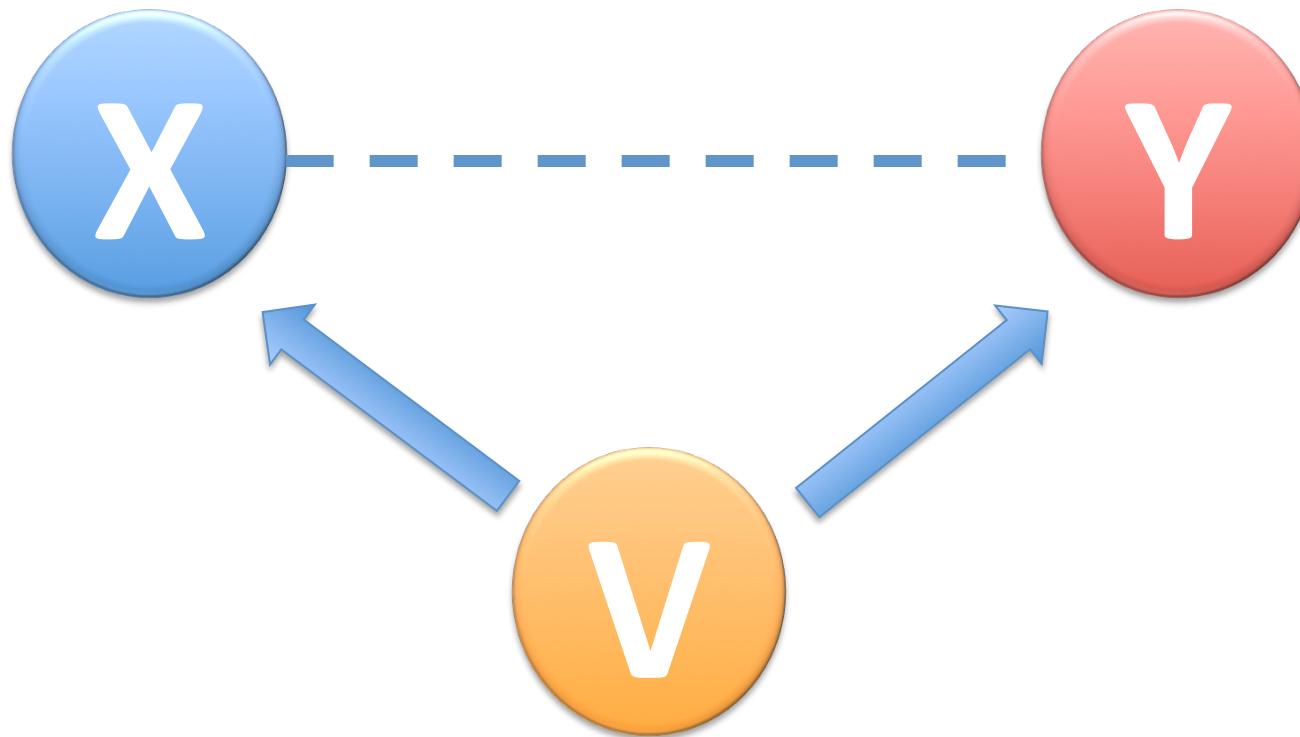
可能性1) 実際にXがYを引き起こした



- ところが、実験室で作られたデータでない場合、別の可能性もあり得る

可能性（2）YがXを引き起こした
(逆の因果関係)

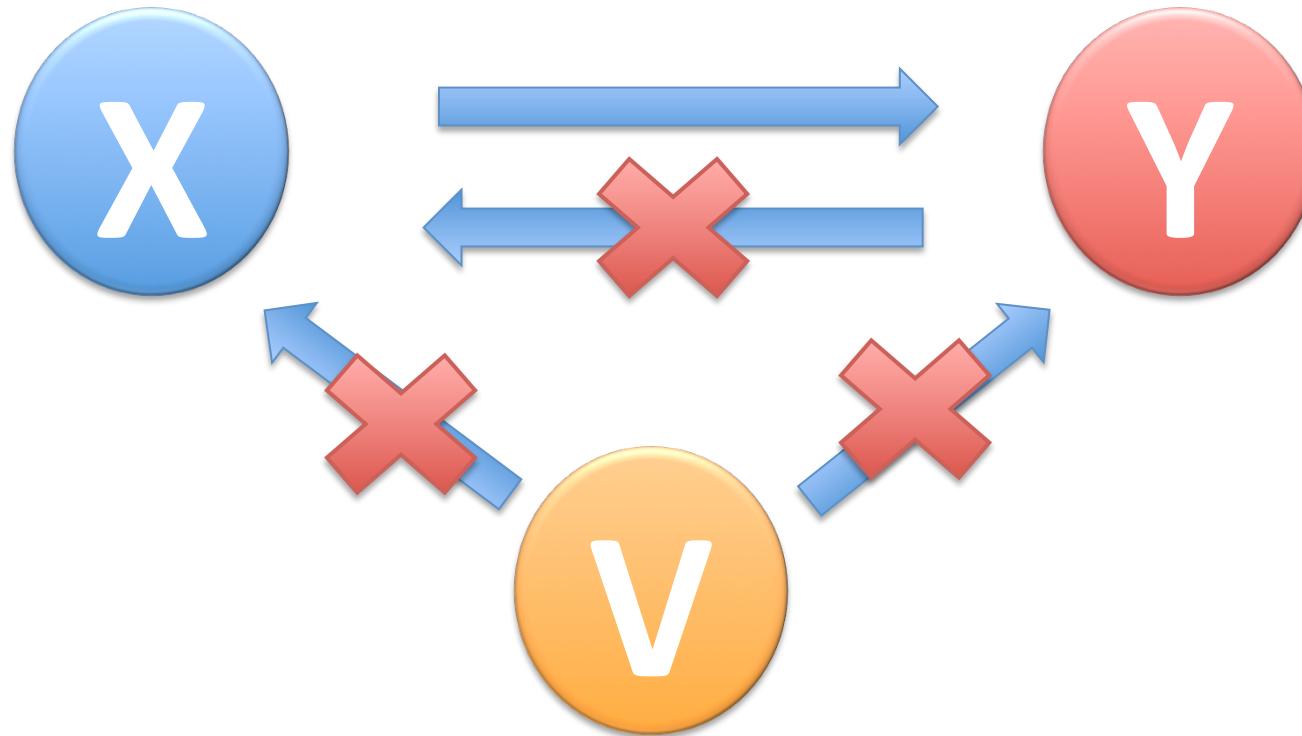
社会科学データで「causality（因果関係）」を言うのはとても難しい



可能性（3）VがXとYの両方に影響を与えただけで、
XとYに直接の関係性はない

（第3の要素の存在）

社会科学データで「causality（因果関係）」を言うのはとても難しい



これらの矢印の可能性を排除できないと

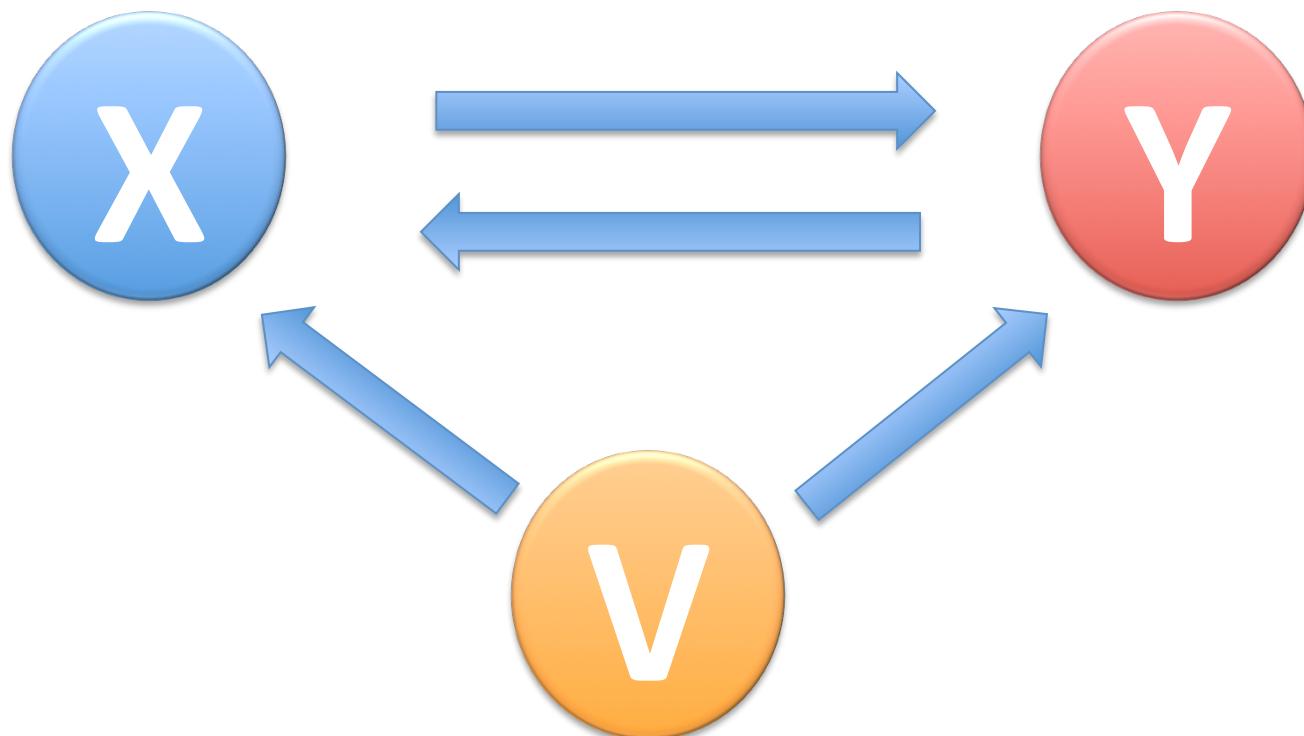
XからYへのCausality（因果関係）は言えない

相関関係の発見だけでは、3つの可能性のどれが本当に生じているか特定できない

- AIが発見した「XとYの相関関係」から言える、3つの可能性

- 1) 病院数 (X) →がん死亡者数 という因果関係が本当に存在
- 2) がん死亡者数 (Y) →病院数 (X) という逆の因果関係（番組内では否定）
- 3) 第3の要素 (V) がXとYに影響している

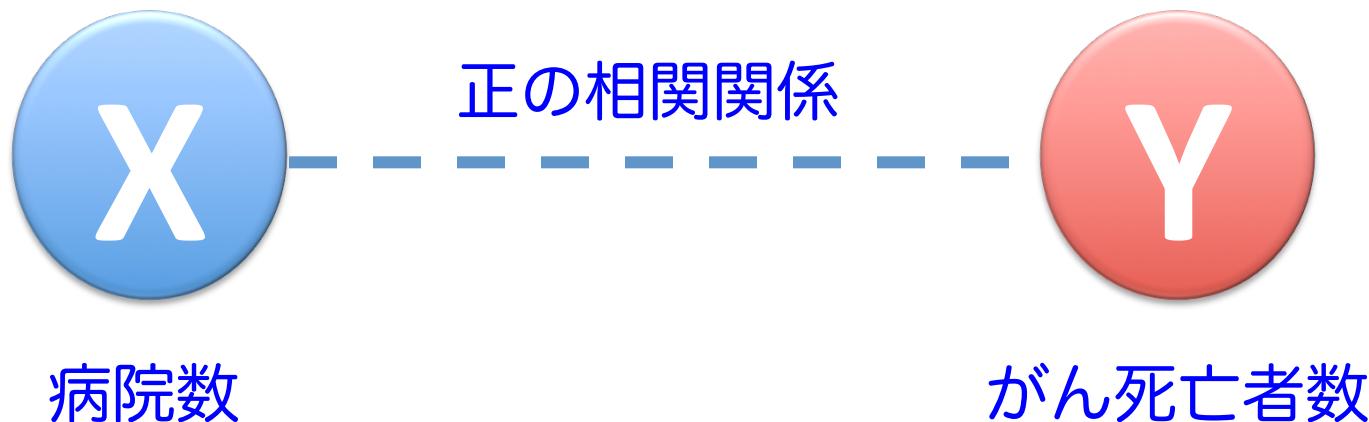
例：高度医療を必要とするがん患者は街を出て行った (V) など



一方で、「人口知能による相関分析が全く有用でない」というのも大きな誤解

- **人工知能（AI）による相関分析は「予測」に役立つ**

- 「がん死亡者数」のデータから「病院数」を予測
- 「病院数」のデータから「がん死亡者数」を予測



- **本日の狙い**

- 経営や政策にデータ分析を生かしたい場合、どのような場面では「因果分析」有効で、どのような場面では「相関分析」が有効かを解説したい

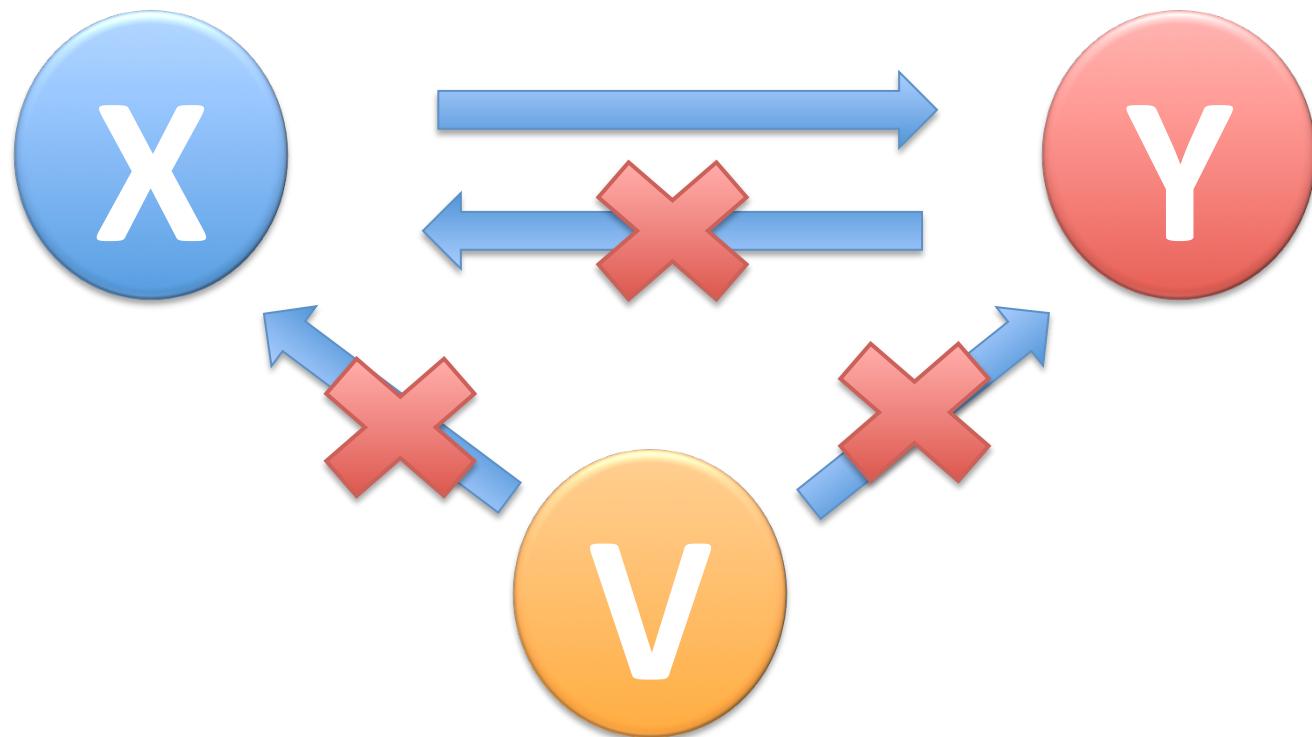
本日お話したい内容

- 1) NHKスペシャル「AIに聞いてみた」から見る、データ分析への危険な誤解
- 2) データ分析で明らかにできる2つの（全く異なる）こと
 - 因果分析—AIだけでは解決が難しい
 - 予測—AIが威力を発揮
- 3) 企業や政府内で始まっているデータ分析の活用事例
- 4) データ・サイエンティスト争奪戦が始まっている米国の労働市場
 - 遅れを取る日本はどのようにデータ・サイエンティストを育成し活用するのか

データ分析で明らかにできる2つの（全く異なる）こと

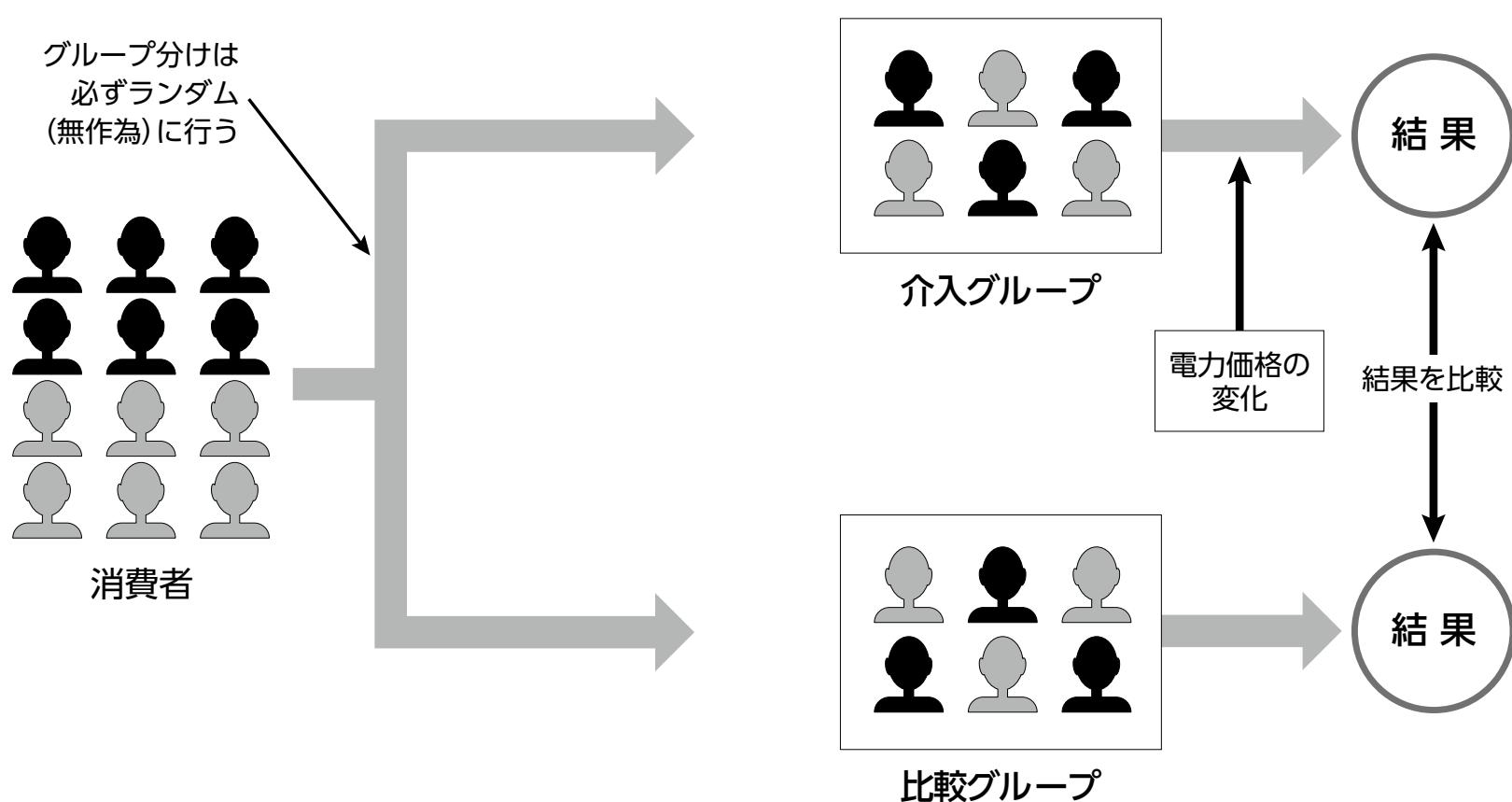
● 1) 因果分析 ($X \rightarrow Y$ という因果関係の特定)

- ・ X を変化させる「介入」を行うと、この介入による影響で Y がどう変化するか
- ・ 例：商品の価格（ X ）を下げるとき、売れ行き（ Y ）がどれだけ上がるのか？
- ・ RCT（ランダム化比較試験）や「自然実験」という手法を用いる必要あり
 - 拙著『データ分析の力（光文社新書）』を参照



因果分析を行う上で最良の方法であるRCT（ランダム化比較試験）

図表2-3 ランダム化比較試験(RCT)



ランダム化比較試験(RCT)では、介入グループと比較グループのグループ分けをランダム(無作為)に行います。

出典：『データ分析の力 因果関係に迫る思考法』伊藤公一朗（光文社新書、2017年）

データ分析で明らかにできる2つの（全く異なる）こと

● 2) 予測

- あるデータ (X) を用いて、関心事項 (Y) の発生や存在の予測する
- 例：消費者の過去の購買データ (X) から、次に買いそうな商品 (Y) を予測
- この場合「相関関係」の分析で十分であるため、人工知能が威力を發揮する



具体例を使って理解する「因果分析」と「予測」の違い

- 子供と成績の関係について、下記の仮想的な例を考える（事実には基づきません）
 - X：ピアノを習っている
 - Y：成績が良い
 - V：親が教育熱心
 - 「XとYに因果関係はなく、VがXとYに影響している」と仮定する
 - この場合でも、XとYに「相関関係」は発生する



● 1) 因果分析

- 適切な因果分析（RCTや自然実験）を行えば以下のことことがわかる
- ピアノを習う（X）→成績が上がる（Y）という因果関係はない
- つまり「成績をあげたいなら、ピアノを習わせよう！」という提言は誤り
- 相関分析から因果関係を主張することが間違いとなる例である



● 2) 予測

- ・一方、相関分析結果を「予測」へ利用することに問題はない
- ・データ分析の目的が「成績の高い子供を発見したい」だとする
- ・ピアノを習っている (X) というデータがあれば、この予測ができる
- ・この際、「本当の因果関係はVから発生している」ことに問題はない
- ・なぜなら、理由が何であれXとYが連動してれば、予測には役立つからである



本日お話したい内容

- 1) NHKスペシャル「AIに聞いてみた」から見る、データ分析への危険な誤解
- 2) データ分析で明らかにできる2つの（全く異なる）こと
 - 因果分析—AIだけでは解決が難しい
 - 予測—AIが威力を発揮
- 3) 企業や政府内で始まっているデータ分析の活用事例
- 4) データ・サイエンティスト争奪戦が始まっている米国の労働市場
 - 遅れを取る日本はどのようにデータ・サイエンティストを育成し活用するのか



● 1) 因果分析

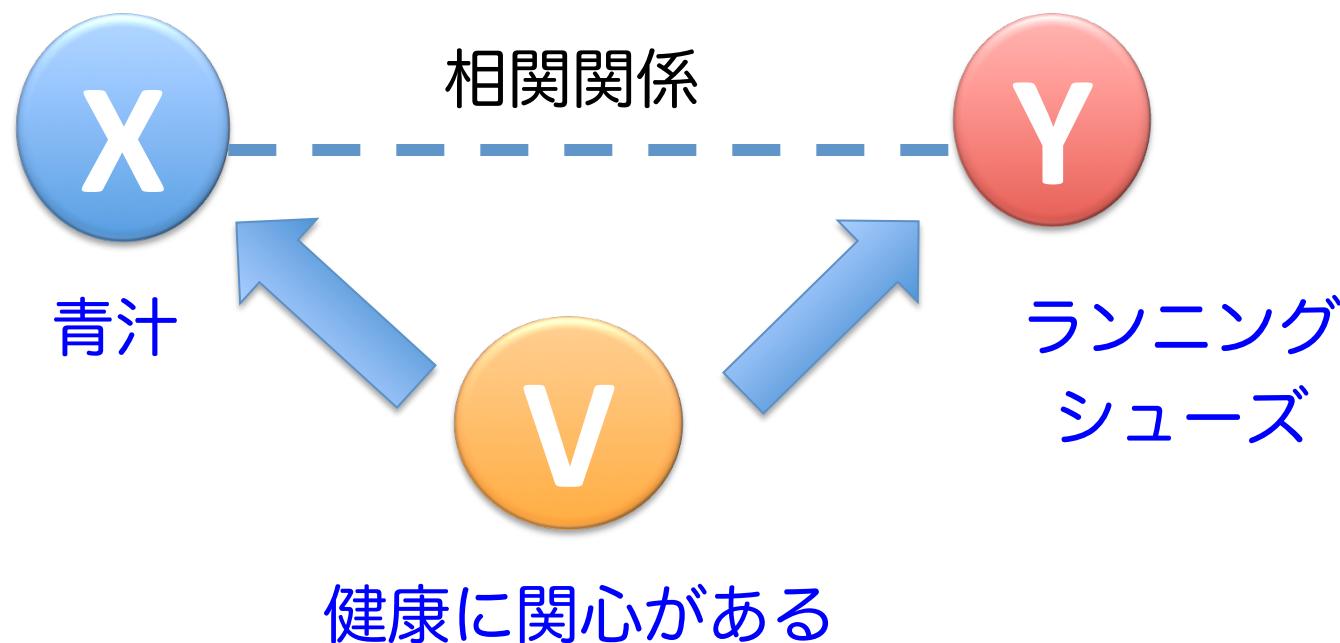
- 広告 (X) を消費者に見せると、売上 (Y) がどの程度伸びるのか？
- 価格 (X) を変化させると、売上 (Y) にはどの程度の影響があるのか
- アマゾンはランダム化比較試験 (RCT) を日常的に用いて答えを出している

● 2) 予測

- 顧客の購買履歴などのビッグデータを人工知能に分析させる
- ここで得られる相関関係は「顧客が次に買いそうな物を予測する」ことに有用
- 「あなたにオススメの商品」という顧客一人一人に向けた広告が可能になる

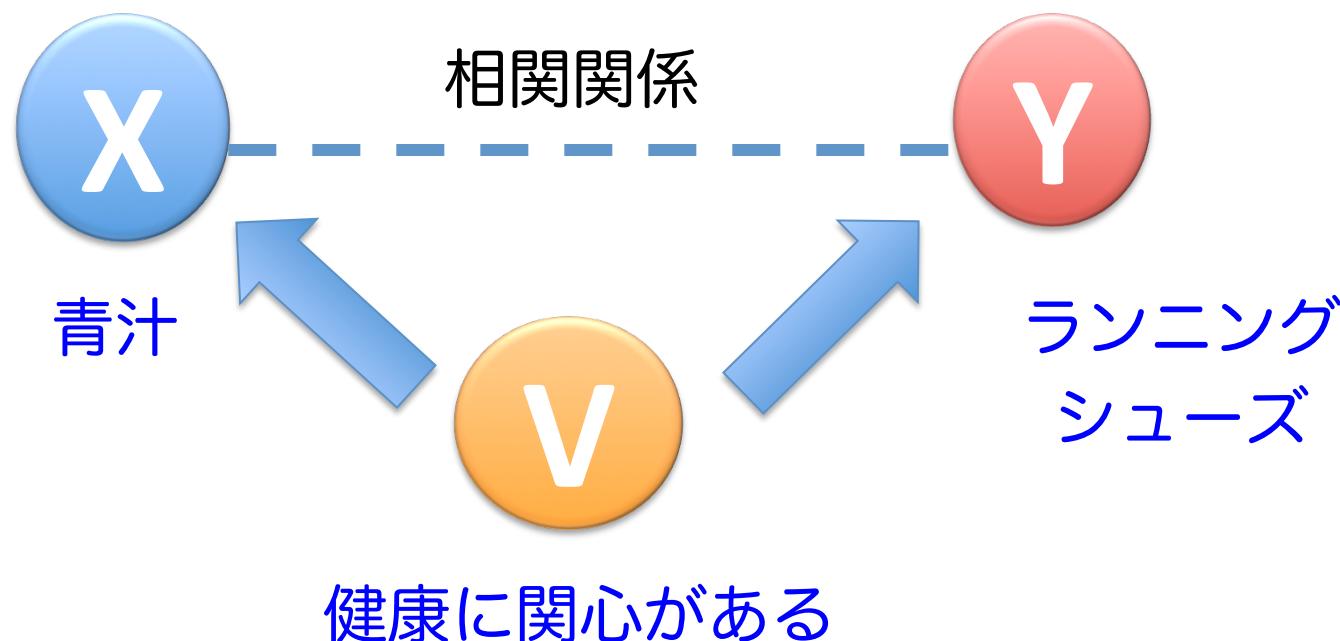
- 消費者の購買行動について、下記の仮想的な例を考える（事実には基づきません）

- X：青汁を買う
- Y：ランニング・シューズを買う
- V：健康に関心がある
- 「XとYに因果関係はなく、VがXとYに影響している」と仮定する
- この場合でも、XとYに「相関関係」は発生する



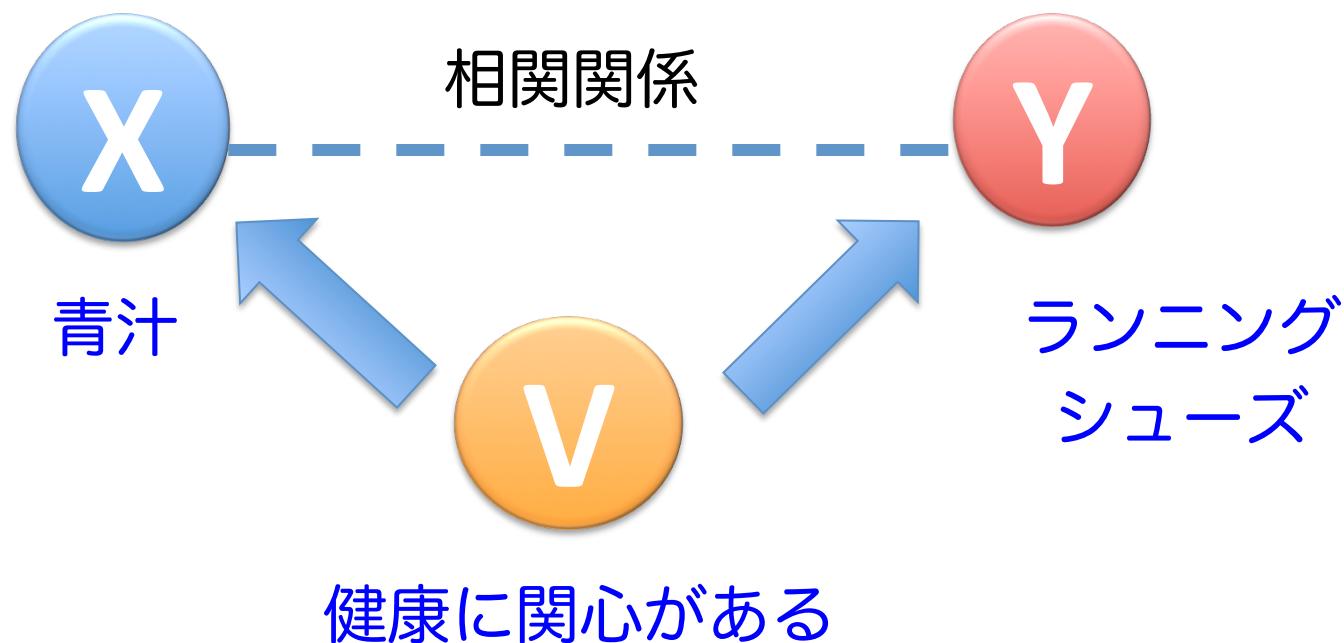
● 1) 因果分析

- 適切な因果分析（RCTや自然実験）を行えば以下のことわかる
- 青汁を買う（X）→ランニング・シューズを買う（Y）という因果関係はない
- そのため、「ランニング・シューズの売上を上げるために、青汁の価格を下げる」というマーケティング戦略は効果がゼロであることがわかる
- 相関分析から因果関係を主張することが間違いとなる例である



● 2) 予測

- ・一方、相関分析結果を「予測」へ利用することに問題はない
- ・データ分析の目的が「ランニング・シューズを買う可能性の高い消費者を探し出したい」だとする
- ・青汁を買っている（X）というデータがあれば、この予測ができる
- ・この場合、「青汁を買っている消費者にターゲットを絞り、ランニング・シューズの広告を出す」というマーケティング戦略は正しい判断と言える



事例2：シカゴ大学と米国EPA（環境保護庁）の協力



- 米国の大気汚染・水質汚染に関する費用対効果の高い監視・監査方法の設計
 - 政策担当者：米国EPA
 - 政策目標：費用対効果の高い監視・監査方法の設計
 - より少ない監視費用・監査費用で環境規制の遵守を100%に近づける方法はないか検討（予測）
 - 効果測定法：ビッグデータを人工知能を用いたデータ分析方法で解析
 - 政策効果：監査方法を変更することにより費用対効果が大幅に向上

事例3：シカゴ大学 Crime Labの取り組み



右：Jens Ludwig（シカゴ大学教授・Crime Lab所長）

- シカゴ大学Crime Lab
 - 政策担当者：シカゴ市、シカゴ警察、シカゴ市内の学校
 - 研究者：シカゴ大学の研究者を中心とする研究チーム
 - 協力者：犯罪防止を目指すNPO
- この3者が協力し、RCTなどの科学的分析手法を用いて政策評価を行っている

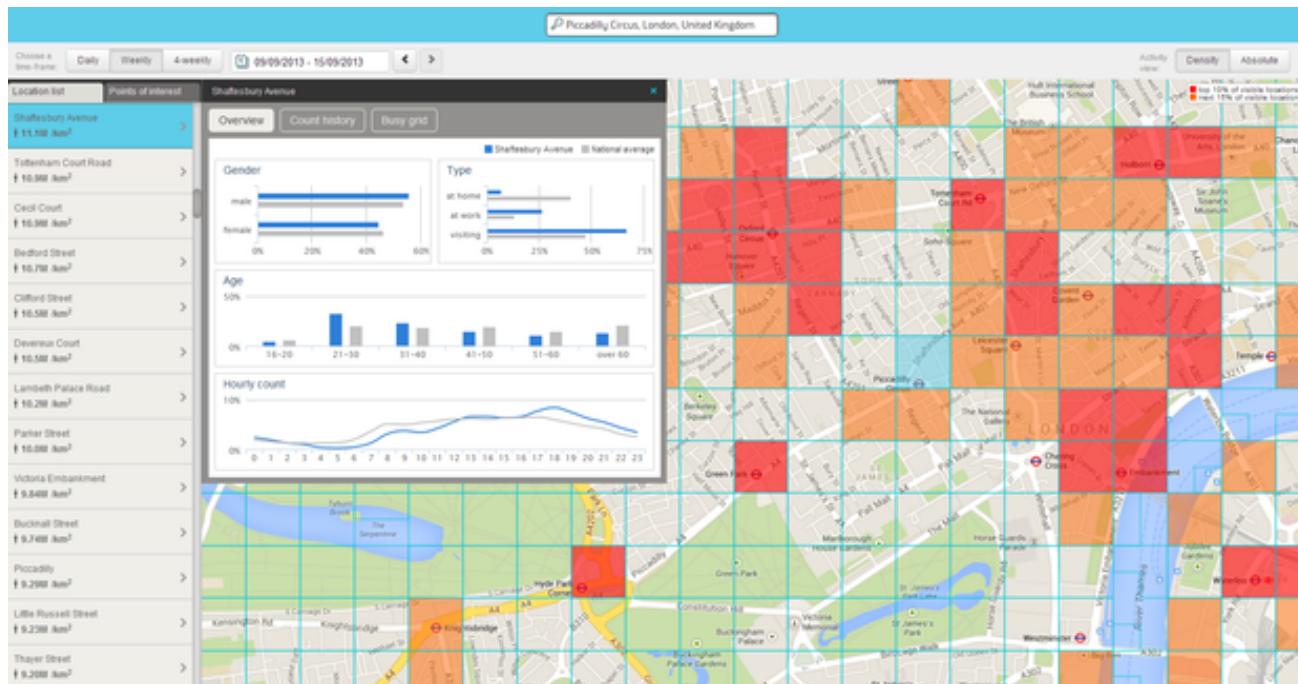
事例3：シカゴ大学 Crime Labの取り組み



● Becoming a Man Program (因果分析)

- 政策目標：青少年の犯罪をいかに予防するか
- 政策介入：中高に通う学生、及び少年院にいる青少年への犯罪防止教育
- 効果測定法：RCT（ランダムに選ばれた子供だけがこの教育を受ける）
- 政策効果：犯罪を犯す確率が大幅に下がり、高校卒業率が飛躍的に向上

事例3：シカゴ大学 Crime Labの取り組み



● Predictive policing (予測)

- 政策目標：警察のパトロールの効率化する
- データ分析：ビッグデータと人工知能を用いて、犯罪発生率を予測
- 予測された発生率に応じてパトロールの配置を決定
- 効果：膨らんだ警察費用の最小化

事例4：政府、企業、大学研究者が主導した日本でのRCTによる政策評価

- 次世代エネルギー・社会システム実証実験

- 資源エネルギー庁省エネルギー・新エネルギー部新産業・社会システム推進室
- 地方自治体
- 企業
- 大学研究者：依田高典（京大）、田中誠（GRIPS）、伊藤公一朗（シカゴ大）

- 日本の4地域でスマート・グリッドに関する実証実験

- 大学研究者がRCTに基づいた実験設計のお手伝いと統計分析を行った
- 1) 横浜、2) けいはんな、3) 豊田、4) 北九州
- 本日は、けいはんな地域で行われた実験結果を簡単に紹介

- 原論文：“Moral Suasion and Economic Incentives: Field Experimental Evidence from Energy Demand.” Koichiro Ito, Takanori Ida, and Makoto Tanaka, *American Economic Journal: Economic Policy*, 10 (1): 240-67, 2018.

事例4：政府、企業、大学研究者が主導した日本でのRCTによる政策評価

- 原発事故以降、電力が足りない日本

- 経済学的には、「強制的に企業や世帯の電力を止める」政策はナンセンス
- 2つの政策があり得る。一つは、①価格変動で節電を促す
- もう一つは、②良心に訴えて、自発的な省エネをお願いすること

- 京都けいはんな地域で行った実験（因果分析）

- 京都けいはんな地域で行った実験の紹介
- 700世帯にスマートメーターを入れ各世帯30分ごとの電力消費データ入手

- 700世帯を3つのグループへランダムに分ける

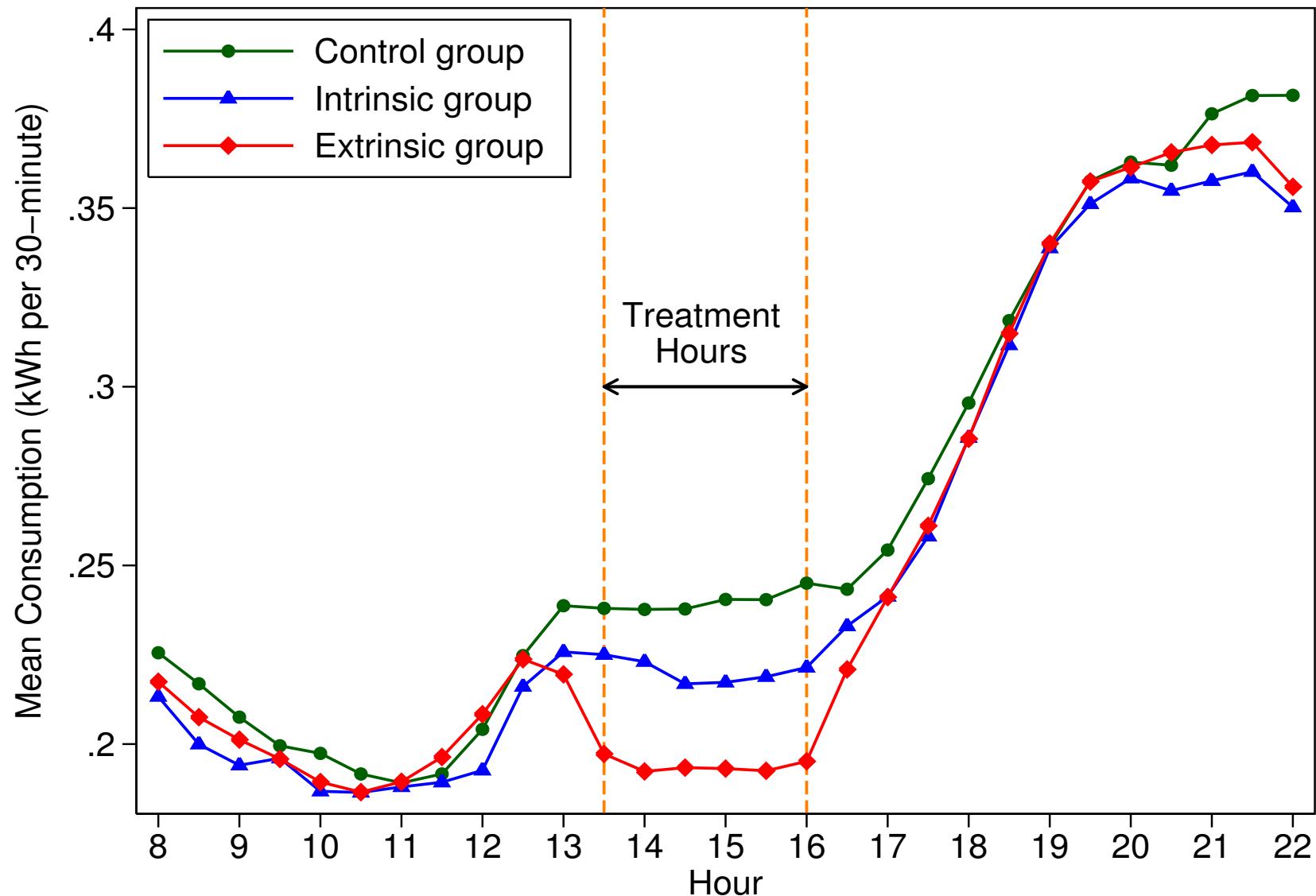
- 1) 価格変動グループ
- 2) 自発的な省エネをお願いする節電要請グループ
- 3) 何も行わない比較グループ

ランダム化によって、グループは統計的に均一になる

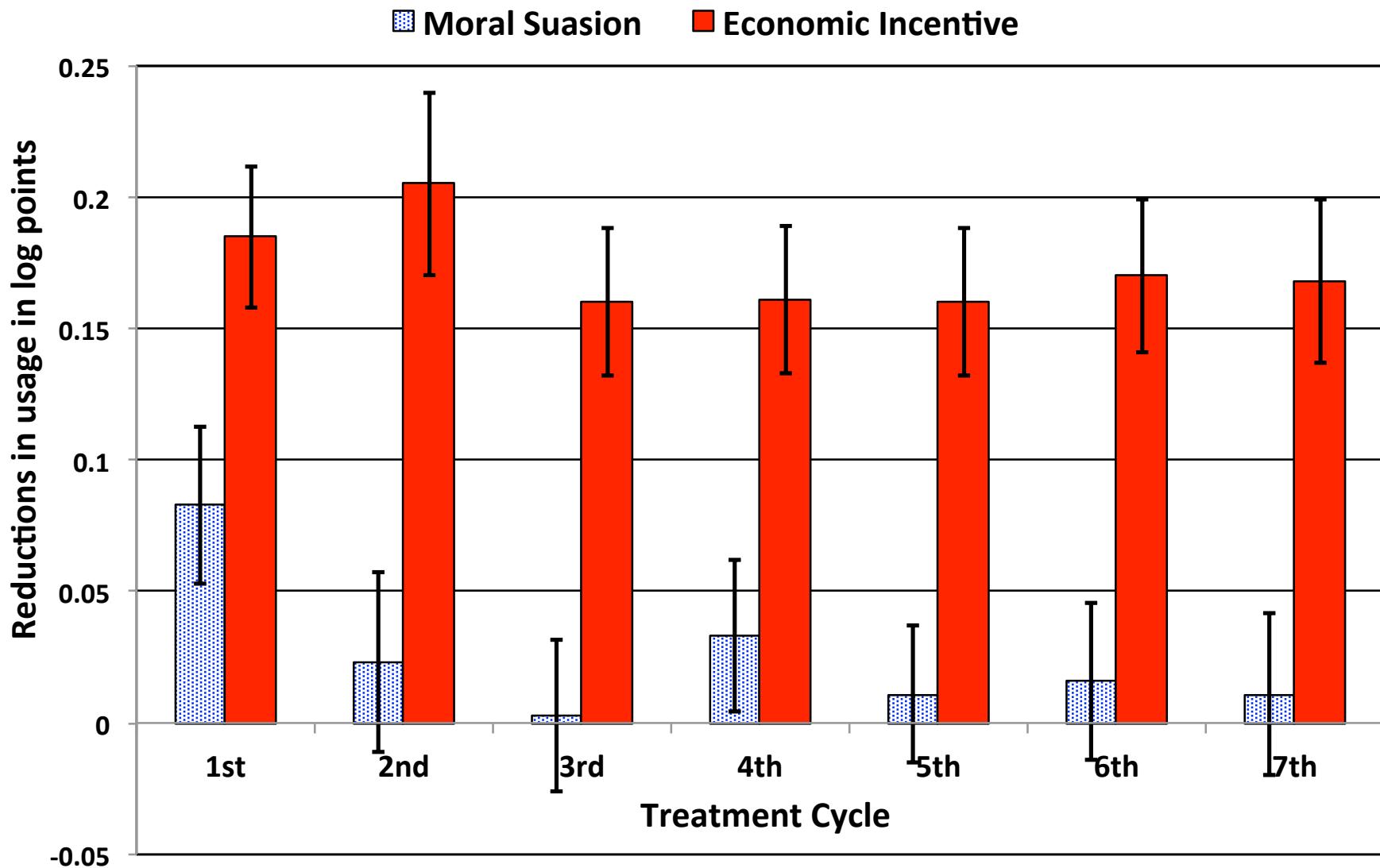
	Sample in the field experiment			Random sample of population (P)	Difference between sample and population
	Moral suasion (M)	Economic incentive (E)	Control group (C)		
Electricity use (kWh/day)	15.14 (6.91)	15.76 (8.49)	15.92 (8.47)	16.23 (7.97)	-0.45 [0.62]
Household income (1,000USD)	66.74 (31.49)	66.59 (31.34)	67.06 (31.01)	66.83 (41.81)	-1.69 [3.68]
Square meter of the house	121.49 (57.54)	113.08 (46.92)	122.15 (46.52)	125.90 (59.65)	-8.95 [4.66]
Number of AC	3.46 (1.93)	3.50 (1.67)	3.68 (1.64)	3.95 (1.71)	-0.43 [0.23]
Mean age of the household	42.26 (17.67)	42.22 (19.07)	40.31 (17.38)	41.91 (16.76)	-0.11 [0.48]
Age of building (years)	13.83 (8.25)	13.39 (7.54)	13.12 (8.20)	15.05 (8.11)	-1.62 [0.28]
Household Size	3.21 (1.18)	3.14 (1.23)	3.32 (1.25)	2.98 (1.41)	0.22 [0.09]

RCTのもう一つのメリットは分析結果が非専門家にも理解しやすいこと

Figure 3: Effects of Intrinsic and Extrinsic Incentives on Electricity Consumption



政策設計への重要な発見：価格効果は持続するが、モラルへ訴える効果は持続せず



本日お話したい内容

- 1) NHKスペシャル「AIに聞いてみた」から見る、データ分析への危険な誤解
- 2) データ分析で明らかにできる2つの（全く異なる）こと
 - 因果分析—AIだけでは解決が難しい
 - 予測—AIが威力を発揮
- 3) 企業や政府内で始まっているデータ分析の活用事例
- 4) データ・サイエンティスト争奪戦が始まっている米国の労働市場
 - 遅れを取る日本はどのようにデータ・サイエンティストを育成し活用するのか

データ・サイエンティスト争奪戦が始まっている米国の労働市場

- 米国の学部新卒市場：データ分析ができる学部生は高待遇で各種民間企業へ
 - シカゴ大学でも研究補助員を毎年募集するが、企業や他大学と人材の争奪戦に
 - 米国の学部生は、理系文系に関わらずデータ分析の授業を履修済み
 - 加えて、学部時代にも教授のデータ分析補助などの経験を豊富に積んでいる
 - そのような経験ができる学部教育現場は日本にどれだけあるのか
- 博士号を持ったデータ・サイエンティストの価値が上がっている
 - 各業種の企業が博士号を保持した専門家を高待遇で引き抜く
 - 学部卒のデータ・サイエンティストを率いるマネージャーとして活躍
 - こういったデータ分析の専門家集団に対抗する術は何か、日本企業も真剣に考える時が来ているのでは？

以上です。質問・コメントはこちらへお寄せください。

シカゴ大学公共政策大学院ハリススクール助教授
伊藤公一朗

Email: ito@uchicago.edu

Web: www.koichiroito.com

