

なぜ俺たちは回帰分析をするのか？

回帰分析勉強会 第1回

2023年3月10日



資源経済・政策と数理資源研究室
RESOURCE ECONOMICS
AND POLICY
AT MORIOKA, IWATE UNIVERSITY

議論：回帰分析に対するイメージ

- 「回帰分析」に対するイメージを教えてください
- なんのためにやるのか？
- なにをするのか？
- どのようなときに使うのか？
- など

研究とは？

研究とは？

- 研究とは、物事との事実や真理を明らかにすること
 - なんらかの事実や事象を、根拠を持って明らかにしていく作業

どうやって？

- 人間の知識を集めて考察し、実験、観察、調査 などを通して調べるなんらかの事実や事象を、根拠を持って明らかにしていく作業である。
- 研究の問い(Research Question)を立てて、答える

Research Questionとは？

- 研究対象についてのまだ明らかにされていない、しかし明らかにする価値のある問い
- 例：
 - キャッチシェアは水産資源の崩壊を止めるか？ (Costello et al., 2008)
 - 漁業のアクセス制限は水産資源の崩壊を防いだか？ (Bjørndal & Conrad 1987)
 - 魚のサイズは価格に影響するか？ (Zimmerman et al., 2013)
 - 個別漁獲割当（IFQ）は良い政策なのか？

Research Questionの3つのタイプ

1. 実証的問題
2. 規範的問題
3. 分析的問題

RQ：実証的問題

- 現実社会で何が真実で、何が真実ではないか？
- 例：Costello et al. (2008)
 - 世界の漁業のデータを集めて、キャッチシェアを行っている漁業とそうでない漁業の間で、水産資源の崩壊が起こっているかどうかを分析
- 数量的データを使う「定量的研究」の他に、インタビューや観察など「定性的研究」もある
 - RQや目的に応じて研究手法を選択

RQ：規範的問題

- 何が望ましく、何がそうでないか？
- 例：日本はITQを導入すべきか？
 - 導入した際に結果がよくなるか、悪くなるか、誰にとってそうか？という定義が必要
 - よい、悪いという価値判断やその基準が必要
- 客観的（科学的）に答えることが難しい
 - 研究者のバイアスが入りうる
- 実証的問題に変換することで、間接的に検証可能

RQ：規範的問題

1. 参照的枠組みを変える

- 日本はITQを導入すべきか？
- ⇒日本はITQの導入で資源を回復することができるか？

2. 背後にある前提条件に注目する

- ITQを導入すべき（したらよくなる）の背後にある条件は？

➤ITQによって資源が回復する

➤資源が回復すれば漁業者の収入が増える

➤漁業者の収入が増えれば、漁業者減少に歯止めがかかる

- それぞれの実証的問題に答えることは可能

RQ：分析的問題

- 観察できる事実を直接取り扱うのではなく、抽象度の高い命題の妥当性を検討する
 - 「フォーマル理論」と呼ばれる
 - 前提条件と定義に依存して、何らかの結論を導き出す
- 例：Gordon-Schaefer Model
 - 不変の限界コスト、バイオマスモデルなどを前提条件として、オープンアクセスの漁業では、漁業者の利益がゼロになる（オープンアクセス均衡）を示した

理論と仮説

- 研究において、重要なのは理論を見つけること
- 理論：原因と結果についての一般的な論述
 - 「原因」と「結果」が存在し、それを論理的に結びつける
 - 例：Nakamura et al. (2023)
 - 多魚種漁獲（原因）→リスクの減少（結果）
 - 理論：金融工学に基づく現代ポートフォリオ理論

Fisheries Science

<https://doi.org/10.1007/s12562-022-01656-4>

ORIGINAL ARTICLE

Social Science



A multi-species catch reduces risk and enhances stability in the fishery? Implications from a portfolio analysis of the Hokkaido setnet fishery

Kohsuke Nakamura¹  · Keita Abe² · Gakushi Ishimura¹

Received: 3 March 2022 / Accepted: 7 November 2022

現代ポートフォリオ理論

- 多様な資産を組み合わせて投資することで（原因）、リスクを減少させながらリターンを保つことができる（結果）
- なぜ？
 - 期待リターンとリスクを計算し、各資産のリスクを相殺しながら期待リターンを維持する最適な配分を計算することができるから
 - という論理的な結びつけがなされているので「理論」

理論仮説

- Nakamuraらはこう考えた

「日本の定置網漁業は様々な魚種を漁獲しているので、ポートフォリオ理論のように安定した漁獲収入を得ることができているのでは？」

- このように検証されるべき理論を理論仮説と呼ぶ

作業仮説

- Nakamuraらは理論仮説を検証すべく、定置漁業のデータを集めた。
- 理論仮説を検証するために、作業仮説を定義した

ツガイナカ中村漁場の定置漁業の漁業収入の分散は、魚種が多ければ多いほど低減する」

- 作業仮説は理論仮説から引き出された具体的に観察可能な特定の変数に関する論述である

理論

- つまり、科学における理論とはXが起これば、Yが起こるという関係を示している
- この原因(=X) と結果(=Y)の関係を因果関係と呼ぶ
- 数量で測る事のできる変数を用いてこの関係を表現する場合、Xを説明変数 (explanatory variable)、Yを応答変数(response variable)と呼ぶ*



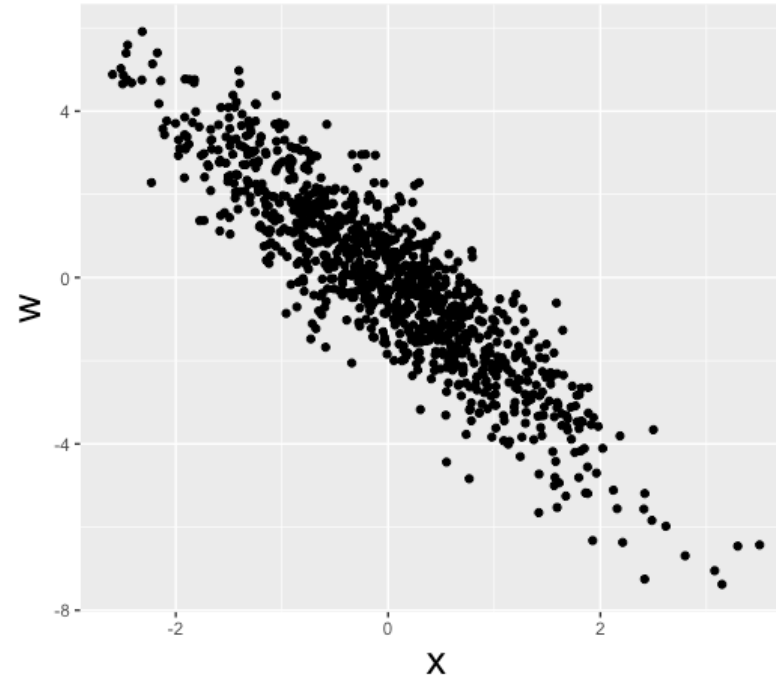
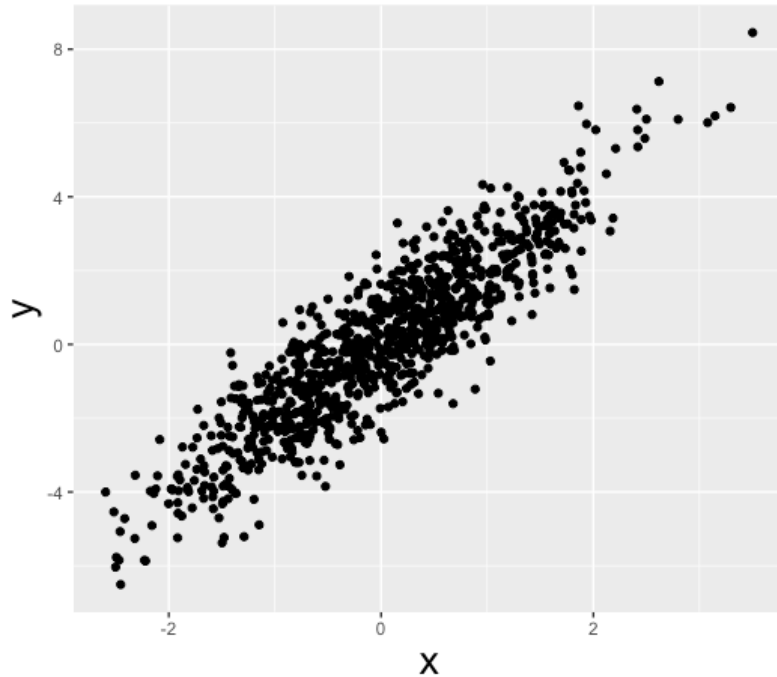
- Xは独立変数、Yは従属変数、被説明変数、目的変数などと呼ばれることもある。

前提条件

- 理論は前提条件が必要である
 - もしAであれば、Xが起これるとYが起これる
 - 例：もし魚種間の収入に負の相関があれば、魚種を増やすとリスクが低減される
- 前提条件がもっともらしいか
- 前提条件が多すぎないか？
 - 前提条件が少ない状態で成り立つ理論は良い理論

二つの変数の関係

- ある変数 X の数値が高い時に、別の変数 Y の数値が高い/低いという関係性があるとき、変数 X と Y は相関している



回帰分析と因果関係

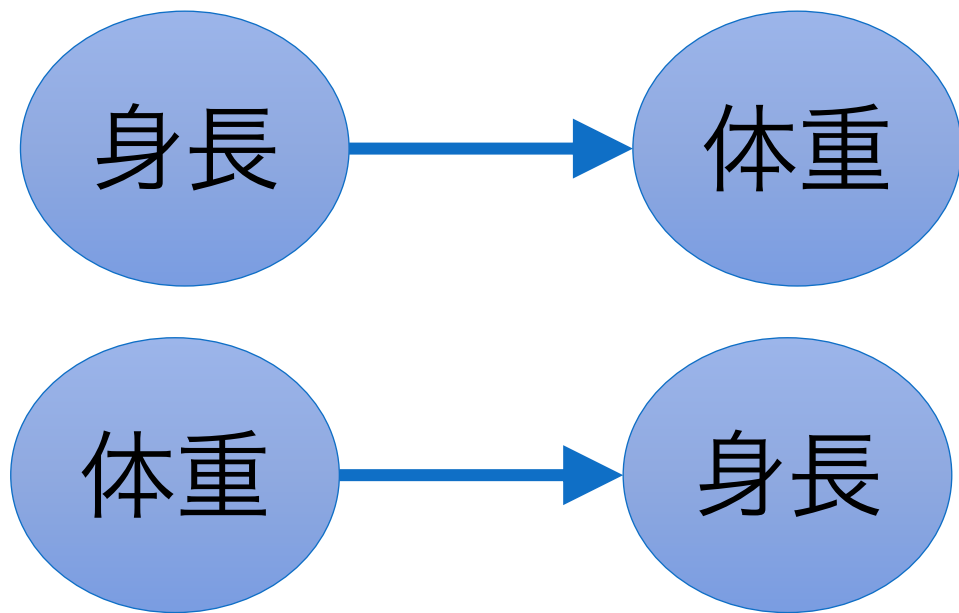
- 以下の単回帰モデルは、前スライドのような直線的な関係に対して、「直線を引く」ことができる

$$Y = \alpha + \beta X + \varepsilon$$

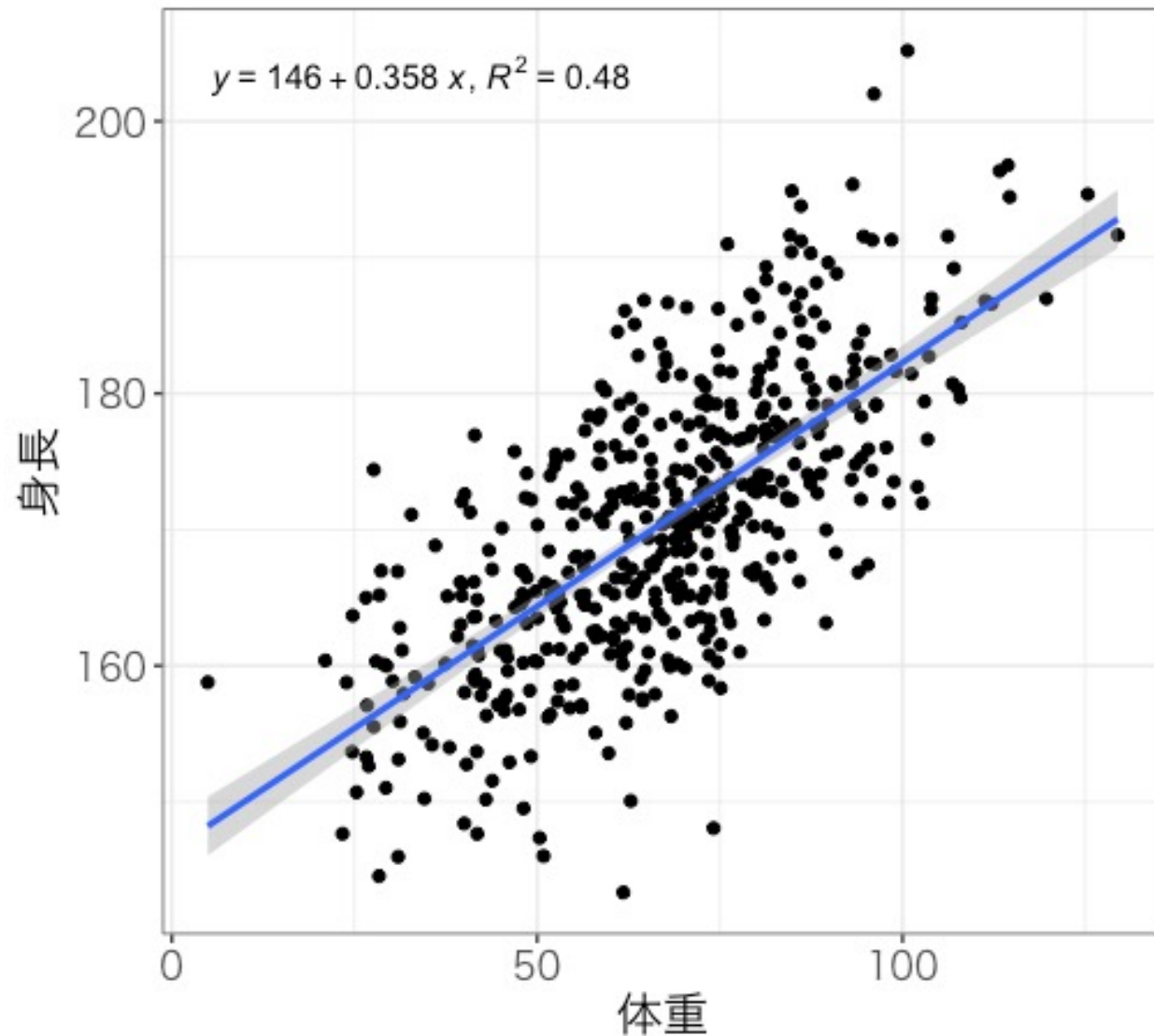
- α : 定数項
 - β : 係数
 - ε : 誤差項、Xで説明できないYの変動
-
- β を推定できた！
 - 理論を実証できた？

身長と体重

- 身長と体重の相関関係



$$\text{身長} = \alpha + \beta \text{体重} + \varepsilon$$

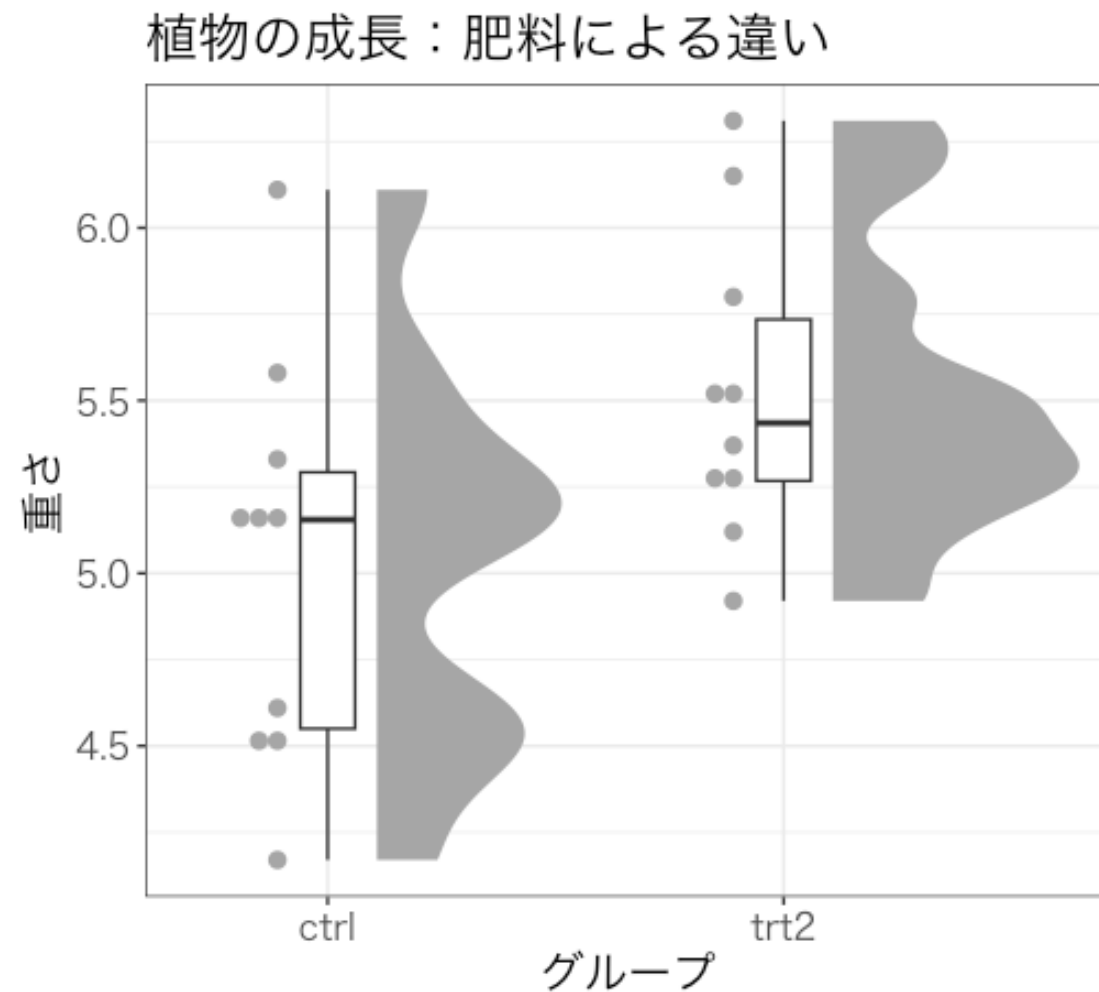


統計的結果と理論の乖離

- XとYの関係が因果関係であろうがなかろうが、相関関係があれば β は推定できる
- しかし、その β によって理論を実証できたとは言えない
 - 因果関係があるとは言えない
- データの構造に依存する
- ではどうすればよいのか？

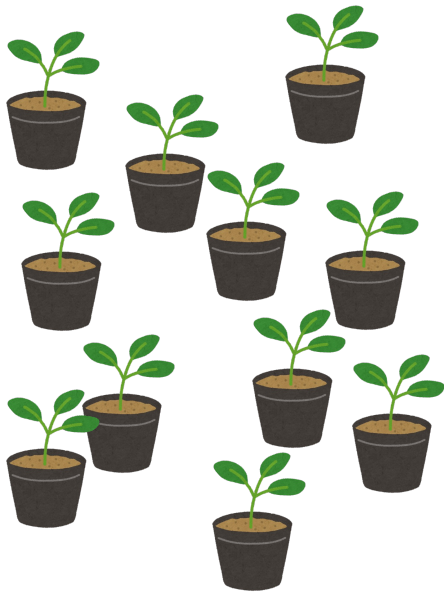
理想：実験

- 理想的なものは実験
- すべての条件を揃えて、Xのみ変えた結果を比較することができる



理想：実験

- 理想的なものは**実験**
- すべての条件を揃えて、Xのみ変えた結果を比較することができる



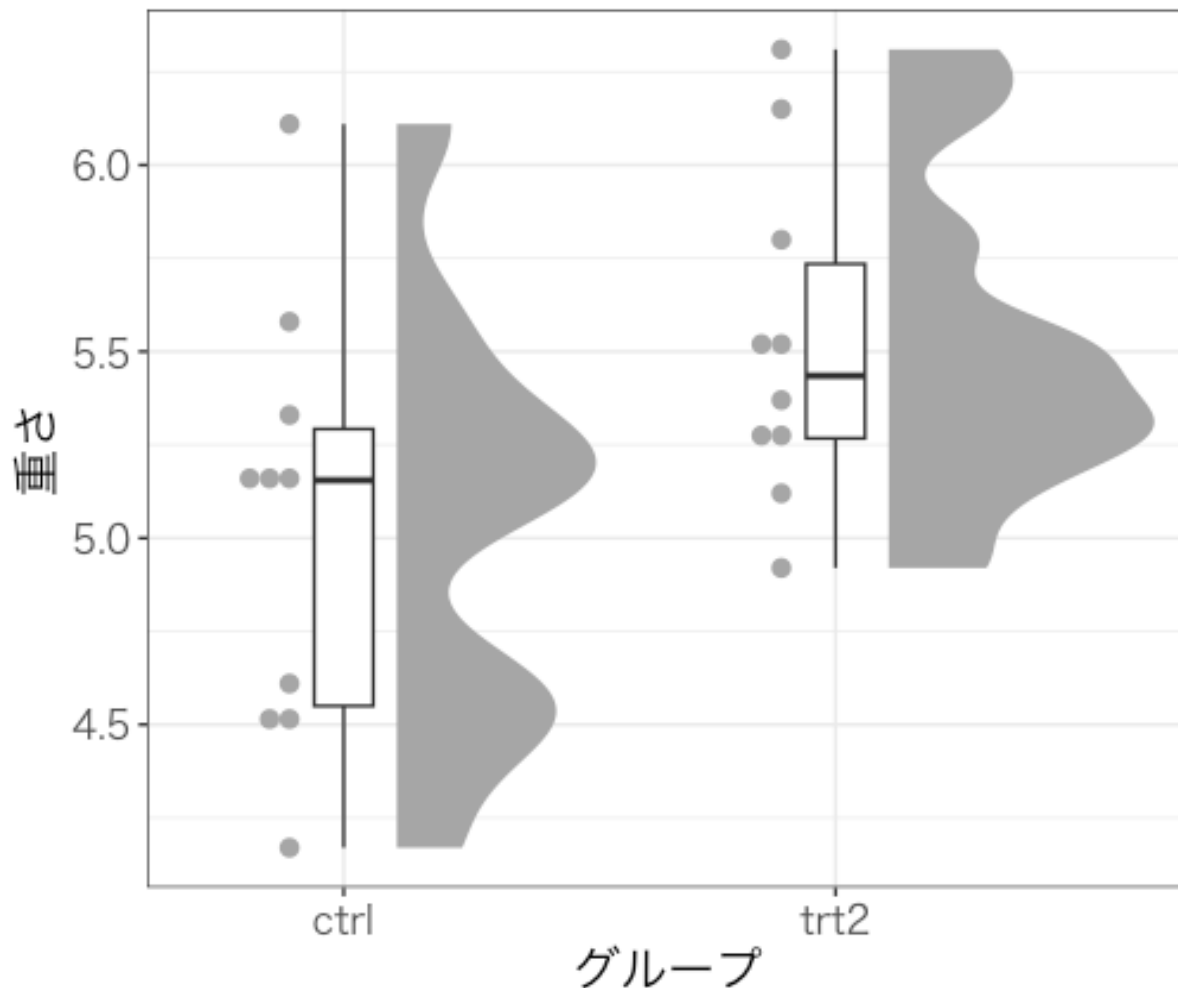
実験の例：

- 片方のグループをコントロール（対照群）
- もう一方のグループに処置Xを施す処置群
- 結果は？
 - 平均値が異なる
 - 二群の比較： t 検定
- 回帰で書くと？

$$\text{重さ} = \alpha + \beta \text{処置} + \varepsilon$$

処置という変数は対照群に属するなら0、処置群に属するなら1というダミー変数

植物の成長：肥料による違い



社会科学における実験

- ノーベル経済学賞（2019年）
- 「世界の貧困を改善するための実験的アプローチに関する功績」
- ランダム化比較試験(Randomized Controlled Trials: RCT)と呼ばれる手法で、途上国における人々を対象に実験を行って、様々なエビデンスを発見した。



社会科学における実験

- 例：ケニアの農家が高い収量の肥料を買わないのはなぜか？
- 収穫による収入を得た直後に買えばよいのに、後で買えると思って先延ばしにした結果、肥料が必要なシーズンになったときにもうお金がない
- 収穫後に期間限定で割引引換券を販売して先延ばしにくくすることで肥料購入率を上昇させられるのでは？
- →RCTによって、証明
 - 引換券を販売したグループのみの前後を比較しても他の要素（その年の収穫など）を排除できない
 - できる限り引換券を販売するグループとそうでないグループをランダムに決定

RCTの短所

- 小さいスケールで実験することが多いため、その政策を大規模したときに同じことが起こるかどうか不明
- 費用・労力がかかる
- 長期的な効果を調べるのが難しい
- 倫理的・政治的に実験をできないケースは多々ある

現実：観察データ

- 社会は研究のために存在しているのではない...
- 研究のための問いに対して、実験データを入手できることは少ない
- 我々が入手できるのは、実際に社会で起こったことを記録した観察データであることが多い

方法1：自然実験

- **自然実験**：制度やその変更、自然災害、紛争など、誰も予測できなかったが結果的に実験のような介入になった現象を用いて、その影響や効果を推定する*

*次に述べる疑似実験も含めて広義の自然実験と呼ぶこともある

自然実験の例 1

- 女性医師と男性医師の処置によって、処置の効果に違いはあるか？
 - アメリカでは病院に運び込まれた患者に対して、ランダムに担当医師が決まる。
 - ランダムに男女の意思が割り付けられる
 - 女性医師が担当したほうが0.4%、30日死亡率が低かった (津川ら, 2017)
 - 0.4%の死亡率の低下は、アメリカで10年間かけて低下させた死亡率と同等

*次に述べる疑似実験も含めて広義の自然実験と呼ぶこともある

自然実験の例 2

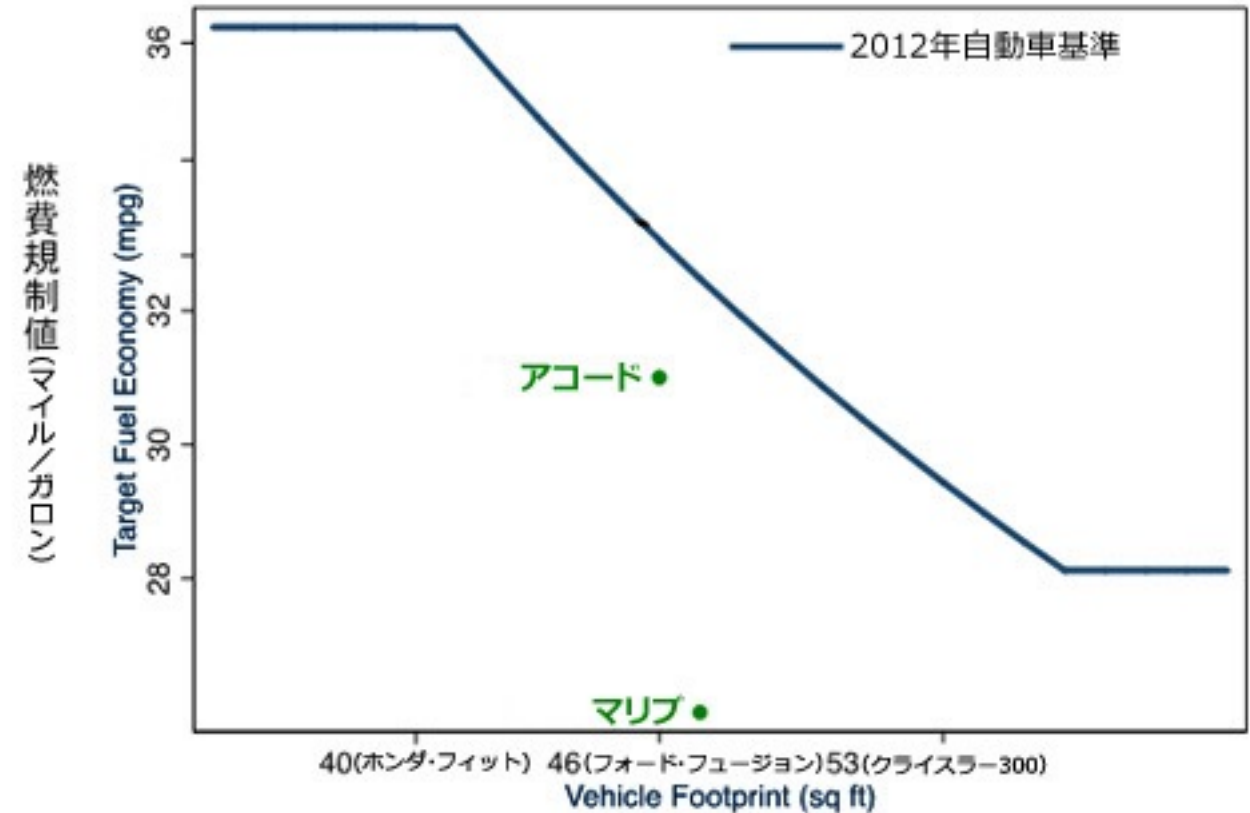
- ベトナム戦争での従軍経験は退役後の所得に影響を与えるのか？
- ベトナム戦争に従軍した経験のある者とない者を単純比較？
 - 民間で仕事を得にくい人・所得が低い人が従軍しやすい？（選択バイアス）
- ベトナム戦争の従軍の一部は、誕生日に基づくくじ引きで決まっていた（自然実験）
- 退役軍人の所得は、そうでない人に比べて15%低いという結果

方法2：疑似実験

- 自然実験のようなタイミングの良い状況はなかなかない
- 実験に近いような状況を作り出す（見つけ出す）ことはできる
- 疑似実験には例えば次のようなものがある
 - 差の差分析 (Difference-in-Differences: DID)
 - 操作変数法 (Instrumental Variable)
 - 回帰不連続デザイン (Regression Discontinuity Design)
 - マッチング法 (Matching)

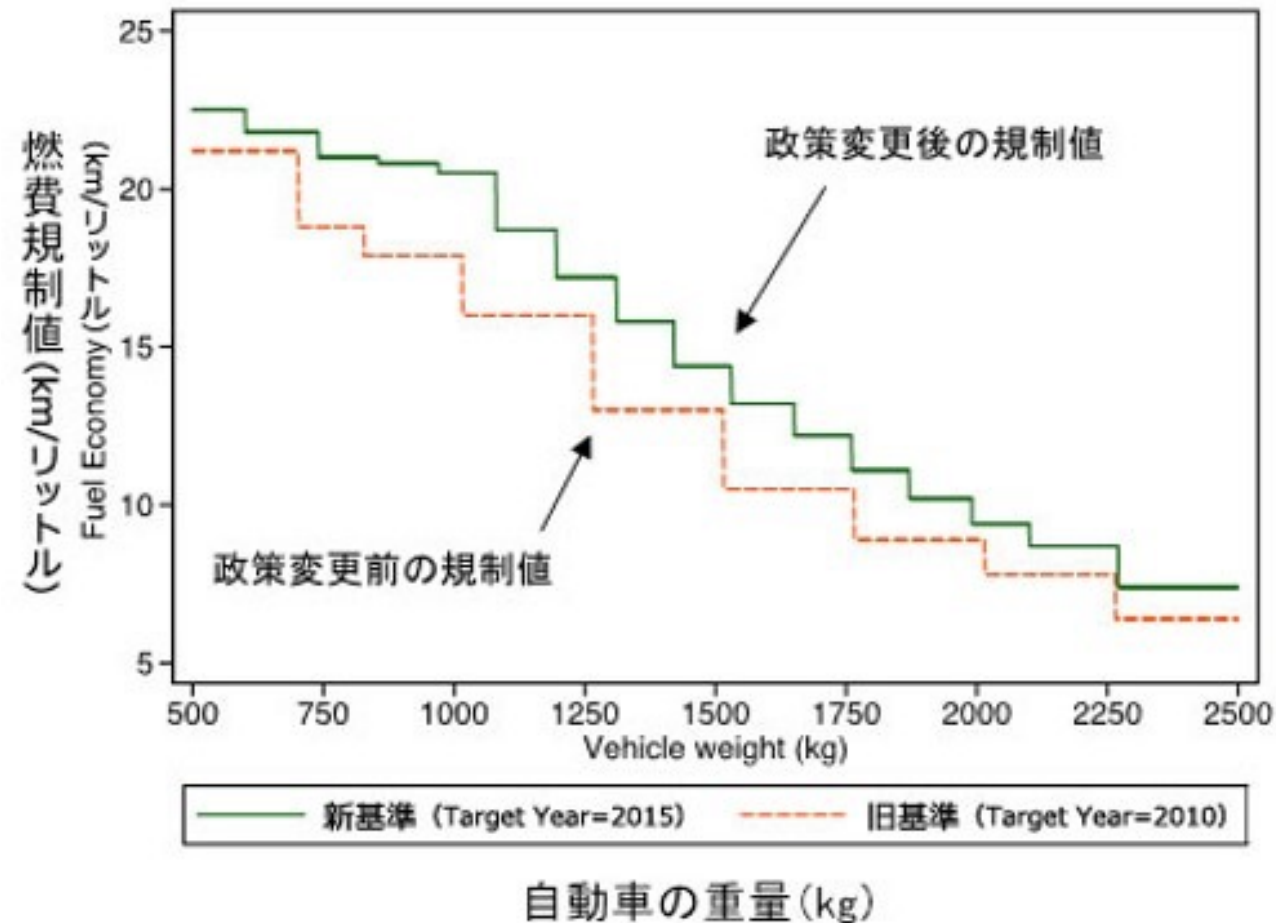
疑似実験の例：集積分析

- 燃費規制は車の大きさに比例する
- 燃費規制の目的としては、技術を向上させて図の上に行ってほしい
- しかし、大きくなるほど燃費がゆるくなっているので、車を大きくすることで燃費規制を達成するという「裏技」がまかり通っている可能性があるのでは??



疑似実験の例：集積分析

- 2008年に日本の自動車の燃費規制が変化

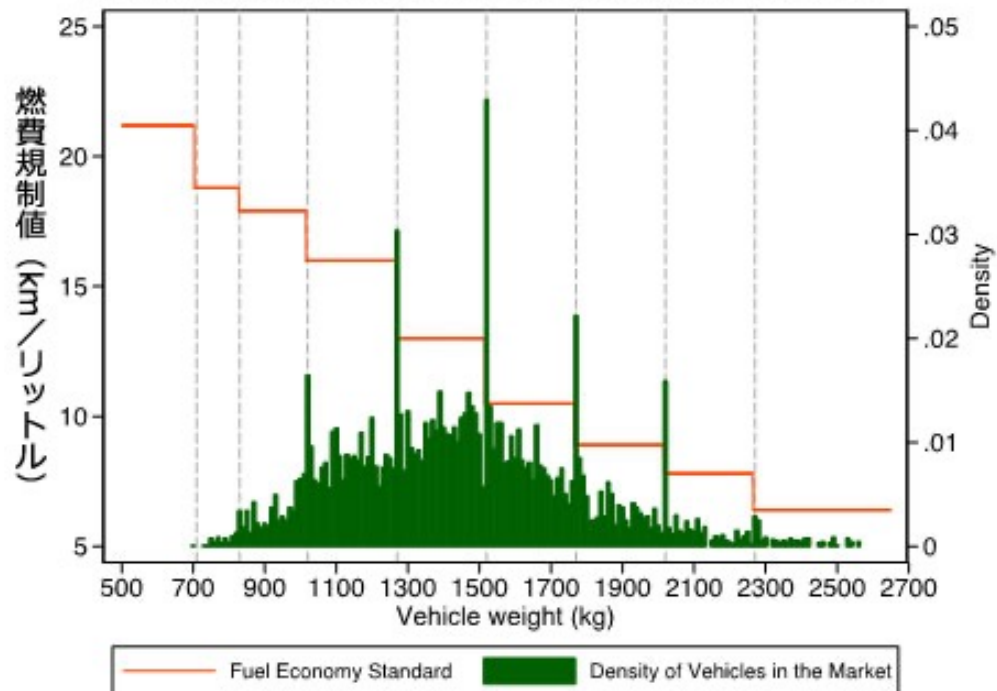


疑似実験の例：集積分析

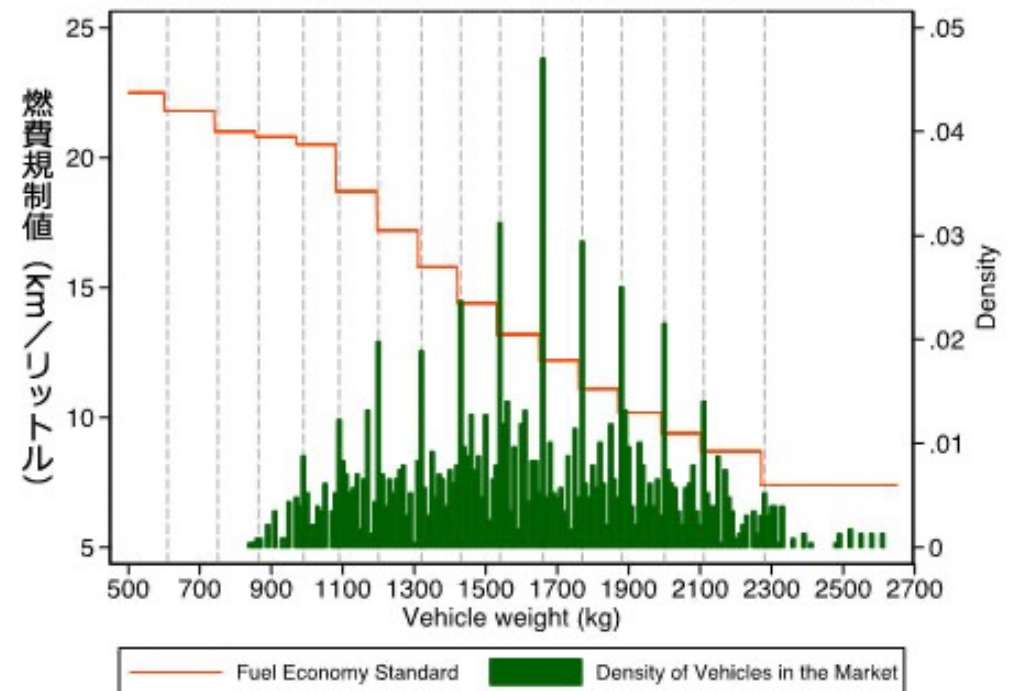
- 多くの自動車境界点のギリギリ右側に集中
- 制度変更によって集中点が変化→政策の影響
- 政策変更によって「車が重く」なった→事故時の死亡率が上昇

[Ito & Saltee, 2018](#)

(1) 政策変更前の規制値(オレンジの実線)と、自動車の分布(緑)



(2) 政策変更後の規制値(オレンジの実線)と、自動車の分布(緑)



疑似実験の例：回帰不連続デザイン

- 高齢者の医療費の自己負担割合は70歳で変化：3割→1割*

Figure 2: Age Profile for Outpatient Visits
2A. Overall Outpatient Visits (Log Scale)

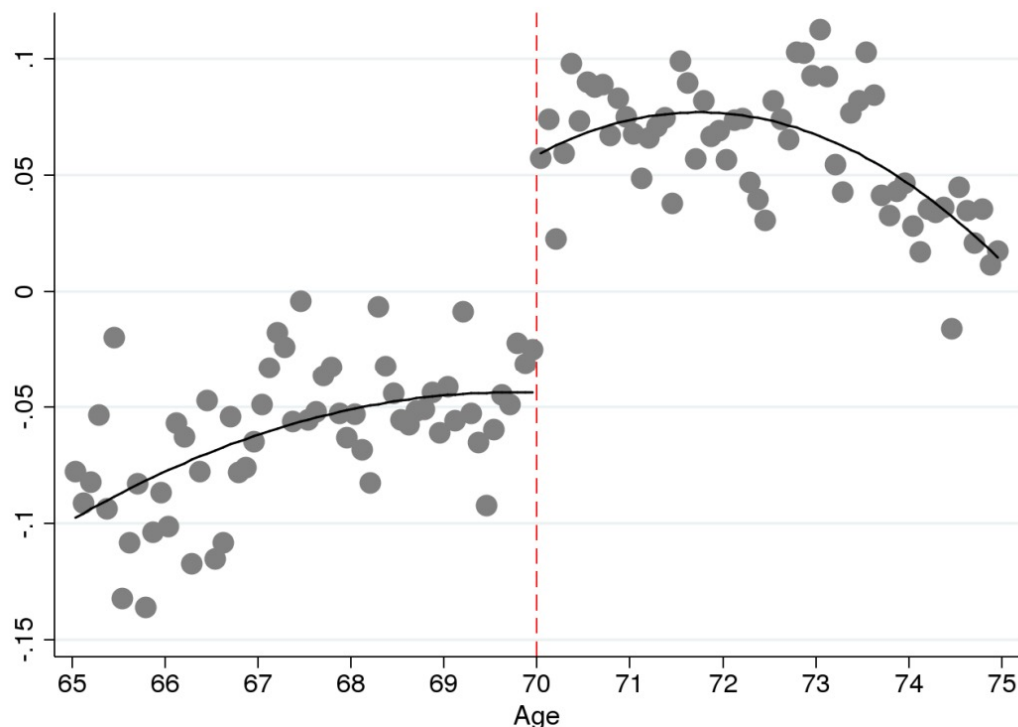
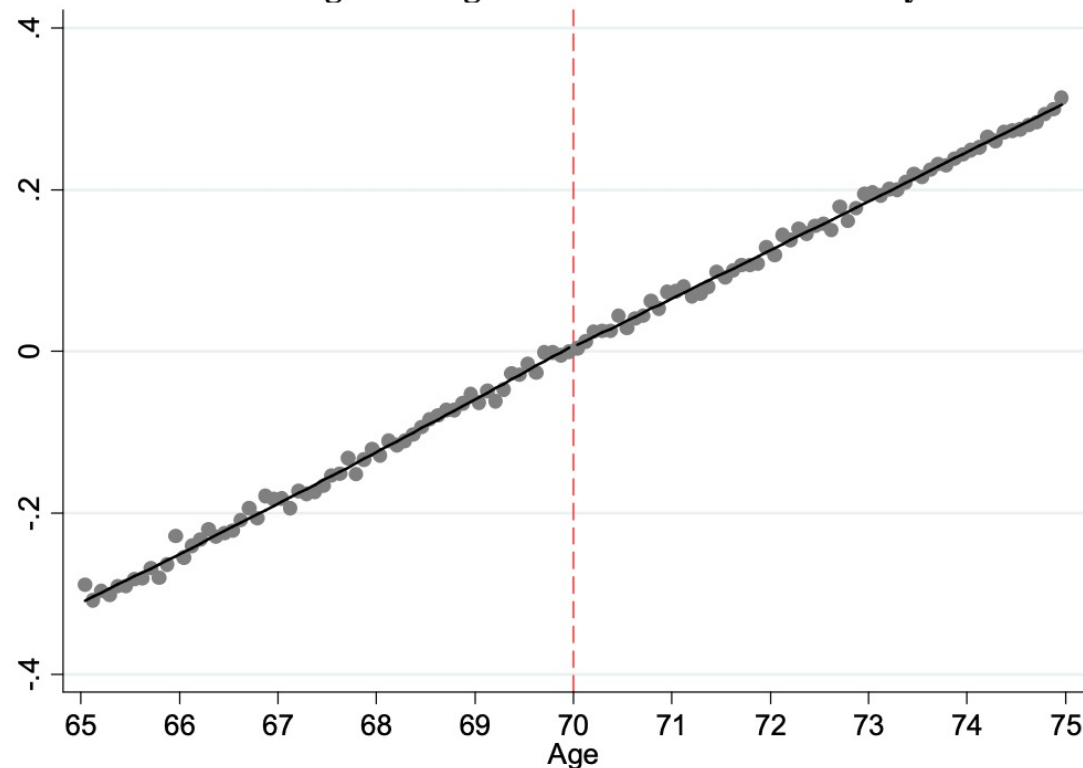


Figure 6: Age Profile of Overall Mortality



方法3：（普通の）回帰分析

- 以上で説明した手法は、特定の状況に当てはまるデータを用いて、回帰分析を行っている
 - 推定されたパラメータ β が因果関係を説明している、という信頼度が高い
- しかし、自然実験や疑似実験を行えそうな状況にないときはどうするか？
 - 観察データを使って、研究を行う

重回帰分析：複数の変数

- 重回帰式

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

- なぜXがたくさんあるのか？

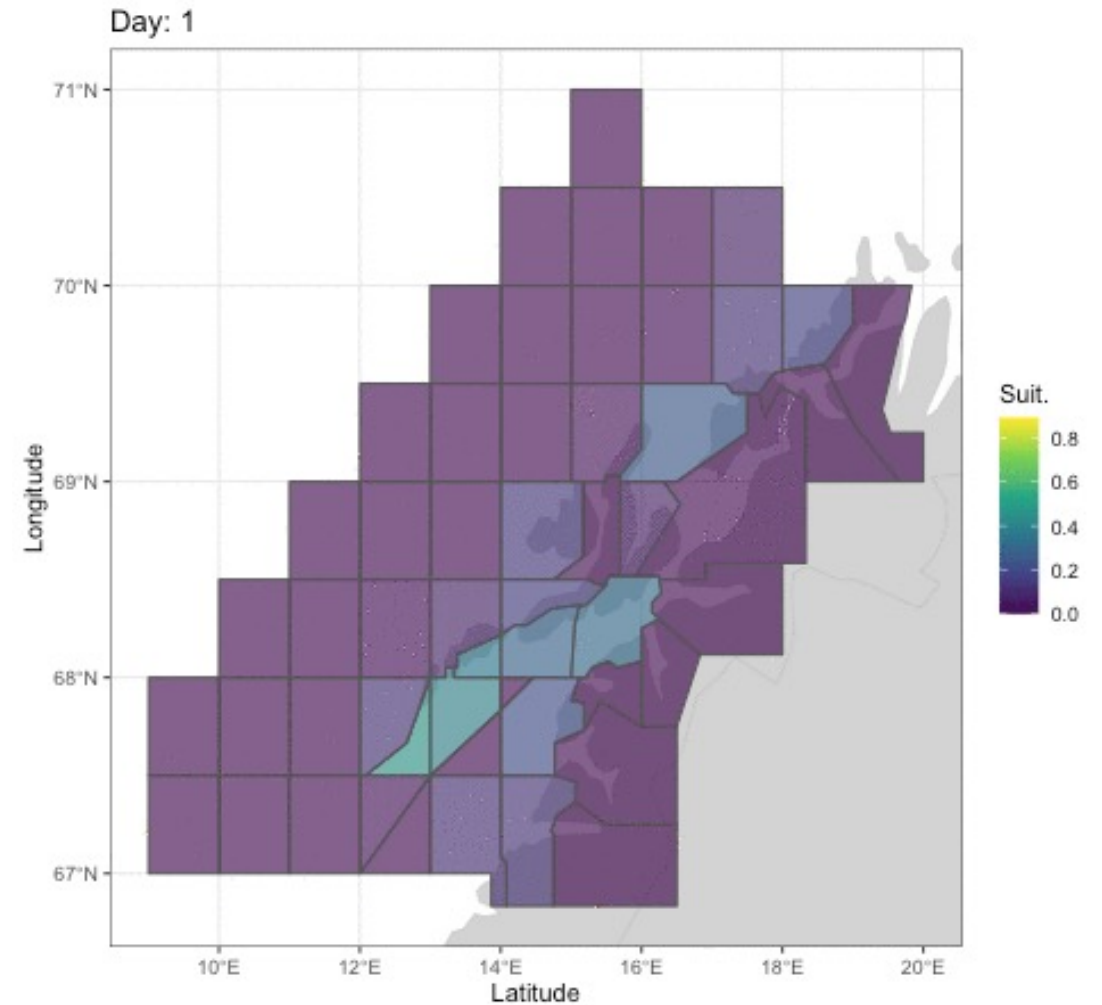
Xがたくさんあるとうれしいから

△ Yの変動をできるだけ多く説明できるから

◎ XのYに対する(真の)影響を測れるから

例：漁業者は魚を追っているのか？

- 気候変動による魚の移動に対して、漁業も移動すると予測されている
- 漁業は本当に魚を追えているのか？
- 海洋学モデルによって、「魚のいる場所」を予測
 - 産卵適正地を計算
 - 産卵適正は、漁船の数を予測できるか？



例：漁業者は魚を追っているのか？

Dependent Variable: Model:	(1)	(2)	Num. of ves.		
			(3)	(4)	(5)
<i>Variables</i>					
Suitability	0.3599*** (0.1183)	0.3696*** (0.1177)	-0.2896 (0.1699)	-0.2884 (0.1708)	-0.1347 (0.2346)
Num. ves. w-1	0.5800*** (0.0216)	0.5801*** (0.0216)	0.5785*** (0.0217)	0.5757*** (0.0219)	0.5757*** (0.0219)
Num. ves. y-1	0.3215*** (0.0161)	0.3211*** (0.0161)	0.3196*** (0.0162)	0.3170*** (0.0163)	0.3170*** (0.0163)
Suitability × Time trend (years)					-0.0137 (0.0164)
<i>Fixed-effects</i>					
Year		Yes	Yes	Yes	Yes
Week		Yes	Yes	Yes	Yes
Area			Yes	Yes	Yes
Main Area			Yes	Yes	Yes
Gear category				Yes	Yes
Length Group				Yes	Yes
<i>Fit statistics</i>					
Observations	1,183,200	1,183,200	1,183,200	1,183,200	1,183,200
R ²	0.73837	0.73853	0.73880	0.73923	0.73923
Within R ²		0.73804	0.72906	0.71568	0.71569

Clustered (Area & Year) standard-errors in parentheses

Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

古典的な回帰分析の考え方

- 最も良いモデルはなにか？
 - いかにYを説明できるか？
 - データに対するフィットのよさ
 - モデルフィットの良さを測る
 - 残差平方和, 決定係数, AICなど
- フィットの良いモデルを見つけて予測
 - 機会学習の登場
 - より複雑で予測の良いモデルが作成可能

新しい回帰分析の考え方

- データの構造に注目
- シンプルなモデルを用いて、因果関係を推定できる「設定」により注力
- モデルのフィットよりも、パラメーターの頑健性がより重要
 - モデルの仮定を変えても（緩めても）結果は変わらないか？
- 複数のモデルを比較しながら、興味のあるパラメーターから得られる理論への知見を重視

議論：回帰分析に対するイメージ

- 今日の話聞いて、どういうイメージの変化があったか？

課題：練習問題 1

- 田中 P.11

1-1 次の2つの事柄の関係は因果関係でしょうか、それとも相関関係でしょうか。あなたの考えを述べてみてください。なお、相関関係と因果関係のどちらかが正しいというわけではありませんので、自分はどう思うのかを説明してみてください。

(1) 2つの事柄：両親の所得，子どもの学力

関係：両親の所得が高いと，子どもの学力が高い

(2) 2つの事柄：クラブ活動への参加，友だちの数

関係：クラブ活動へ参加している人は，友だちの数が多い

(3) 2つの事柄：一国内の所得格差，経済成長率

関係：所得格差の小さな国は，経済成長率が高い

(4) 2つの事柄：友人の喫煙率，自身の喫煙

関係：友人の喫煙率が高い人は，喫煙しやすい

(5) 2つの事柄：都市の貧困率，犯罪率

関係：貧困率の高い都市の犯罪率が高い

(6) 2つの事柄：都市の凶悪犯罪発生率，1人当たり警察官数

関係：1人当たり警察官数の多い都市の犯罪発生率が高い

1-2 **1-1** で見た関係が因果関係とすると、次の目標のためにどのような対策（政策）をとることができるでしょうか。また、これらの関係が相関関係であるときに、これらの対策は効果を持つでしょうか。6 つそれぞれの関係について、あなたの考えを説明してみてください。

- (1) 目標：子どもの学力を高める
- (2) 目標：友だちの数を増やす
- (3) 目標：一国の経済成長率を高める
- (4) 目標：喫煙率を下げる
- (5) 目標：都市の犯罪率を下げる