統計学 Statistics
2021年度　後期
第14回　回帰分析

大久保　豪

suguruokubo@gmail.com

株式会社BMS横浜、立命館大学客員協力研究員

今日のテーマ

14-1. 回帰分析とは

14-2. この検定の注意点・論文の読み方

14-3. バイアスと交絡

14-1. 回帰分析とは

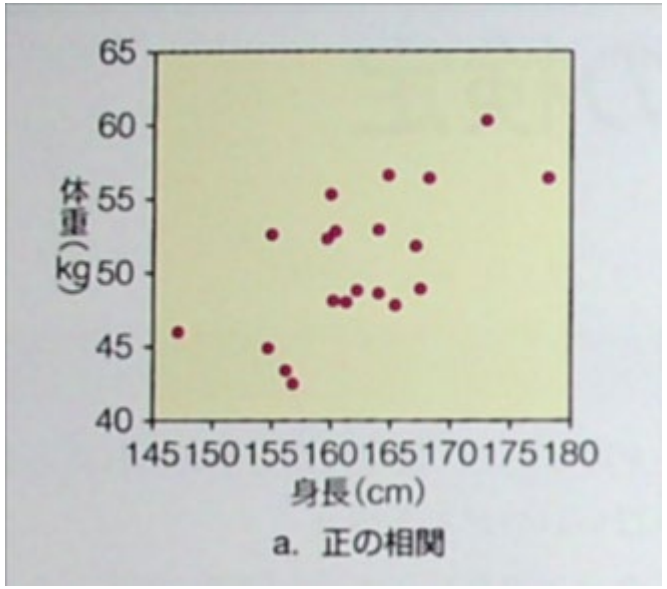
相関とは

わかりやすいのは身長と体重の例

個別にみれば身長に比してやせている人、太っている人がいるが、大まかには身長が高いほど体重も重い

表5-22　あるクラスの身長(cm)と体重(kg)

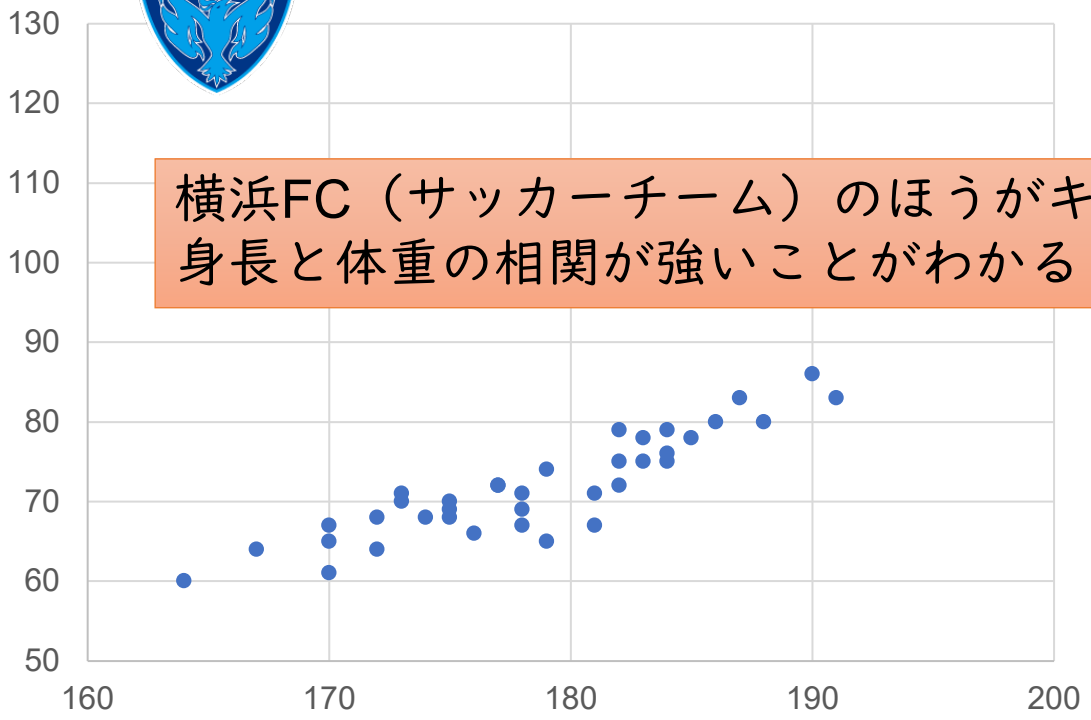
番号	1	2	3	4	5	6	7	8	9	10
身長	156.8	161.3	164.0	178.1	167.1	160.2	168.2	156.2	164.0	159.7
体重	42.5	48.0	52.9	56.4	51.8	48.1	56.4	43.4	48.6	52.3
番号	11	12	13	14	15	16	17	18	19	20
身長	167.5	147.1	173.0	160.0	164.8	160.4	165.4	155.0	154.7	162.2
体重	48.9	46.0	60.3	55.3	56.6	52.8	47.8	52.6	44.9	48.8



一方が増加すればもう一方が増加する傾向があるばあい、「正の相関がみられる」という

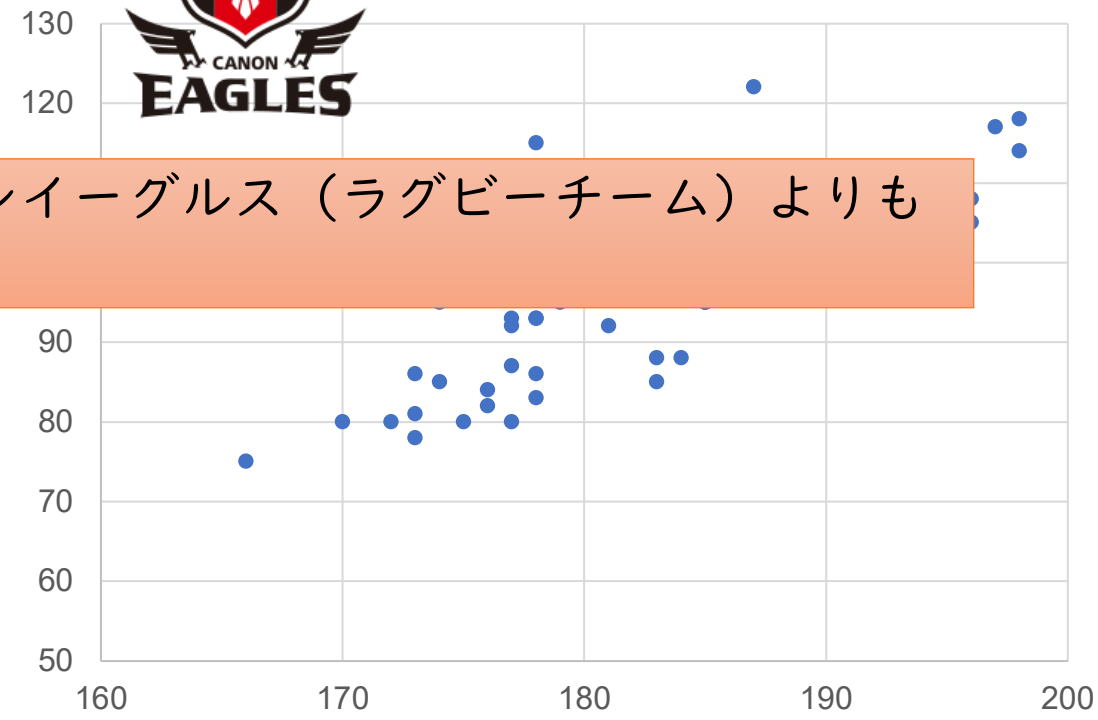
出典　高木晴良（2016）『系統看護学講座　基礎分野　統計学』医学書院, p151-152.

実際に身長と体重をみてみましょう



横浜FCの選手の身長（x）と体重（y）

$r_{横浜FC} = 0.90$



キヤノンイーグルスの選手の身長（x）と体重（y）

$r_{キヤノン} = 0.64$

横浜FC（サッカーチーム）のほうがキヤノンイーグルス（ラグビーチーム）よりも身長と体重の相関が強いことがわかる

単回帰分析

この検定のイメージ

- あの人とあの人は仲が良い、というだけじゃなく
- YくんがこうなったのはXさんの影響ではないか？と考えてしまうタイプ
- 話としてはおもしろいんだけれども、やっぱりよくよく話を聞いてみないとね...

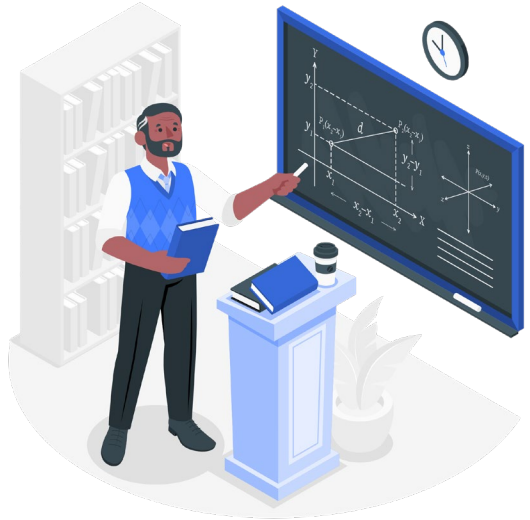
回帰分析 regression analysis とは

2つの変数x, yについて関連がみられるとき、その関係について、なんらかの関数で表すようにしたもの

この最も単純なものが　y=ax+b

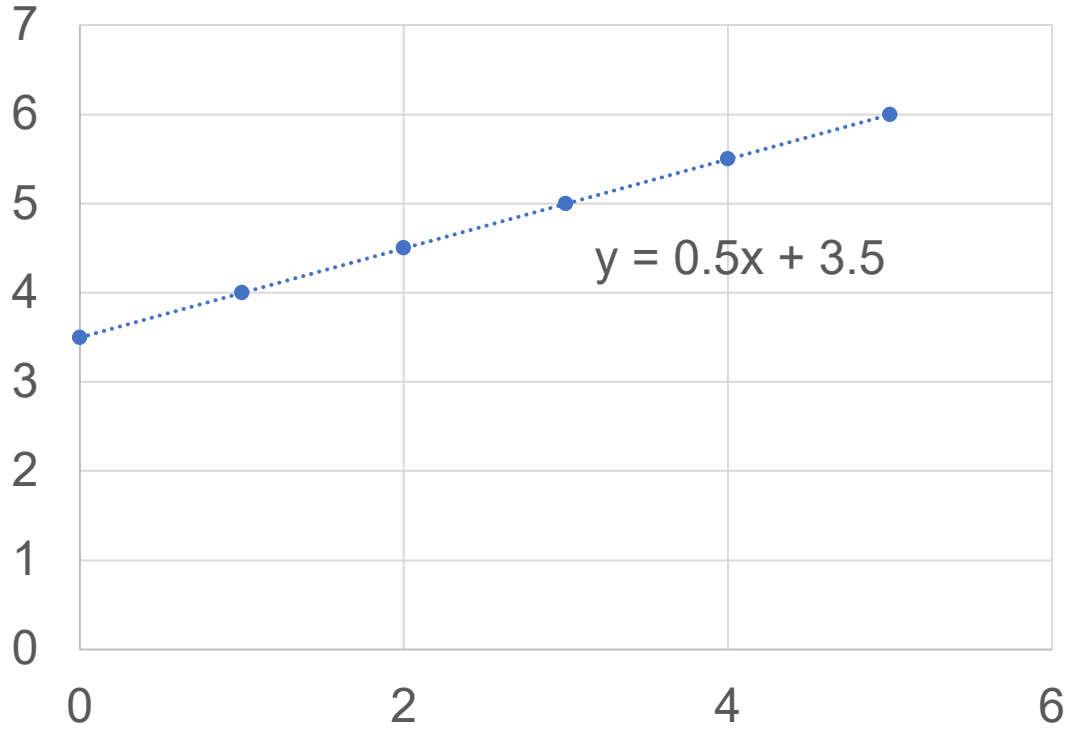
一次関数的に言えば、aが傾き（係数）、bが切片
回帰分析ではaを回帰係数ともいう。

グラフにすると一直線　→　この回帰分析を「線形回帰」という



出典　高木晴良（2016）『系統看護学講座　基礎分野　統計学』医学書院, p154-155.

データが一直線上に並んでくれれば…



いわゆる数学の一次関数

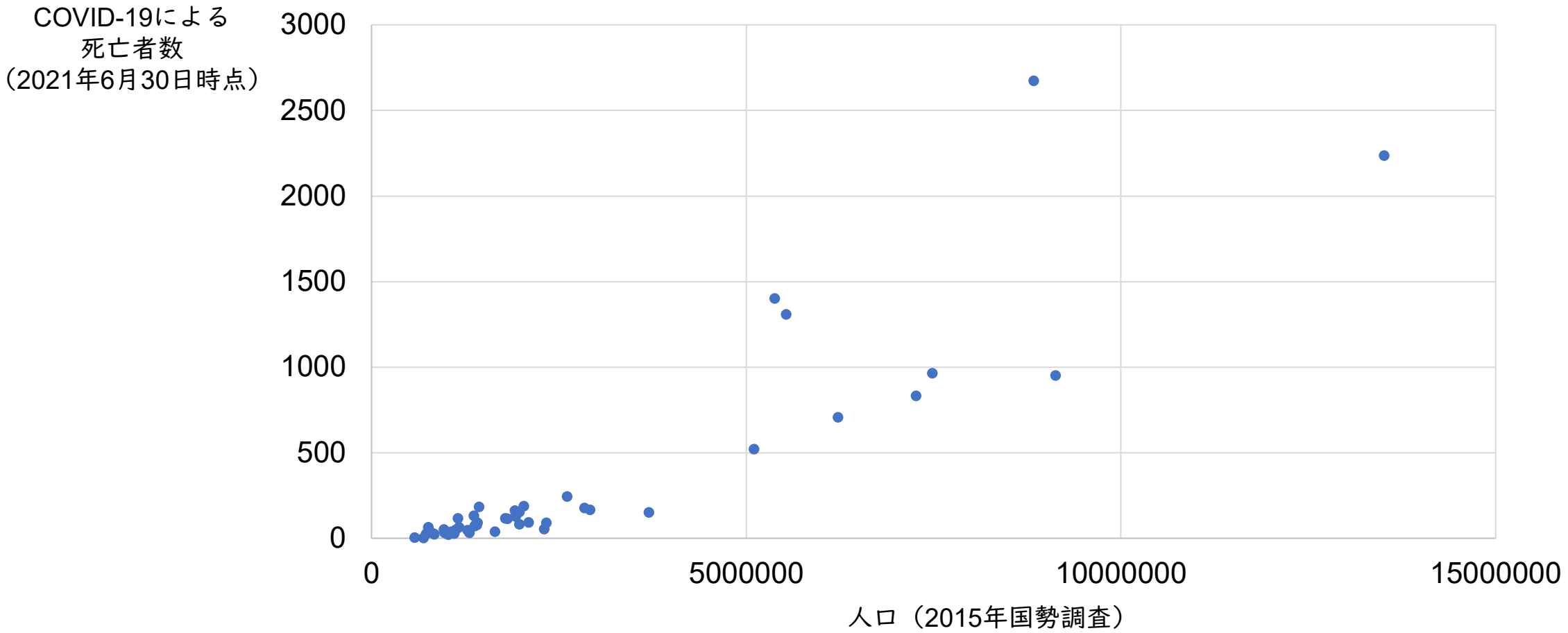
xとyの値から傾き（係数）と切片を求める

a　=（4.5-3.5）÷（2-0）
　　= 1÷2
　　= 0.5
b　= 3.5

しかし、現実のデータ（たとえば年齢と血糖値）が一直線上に並ぶことはありえない

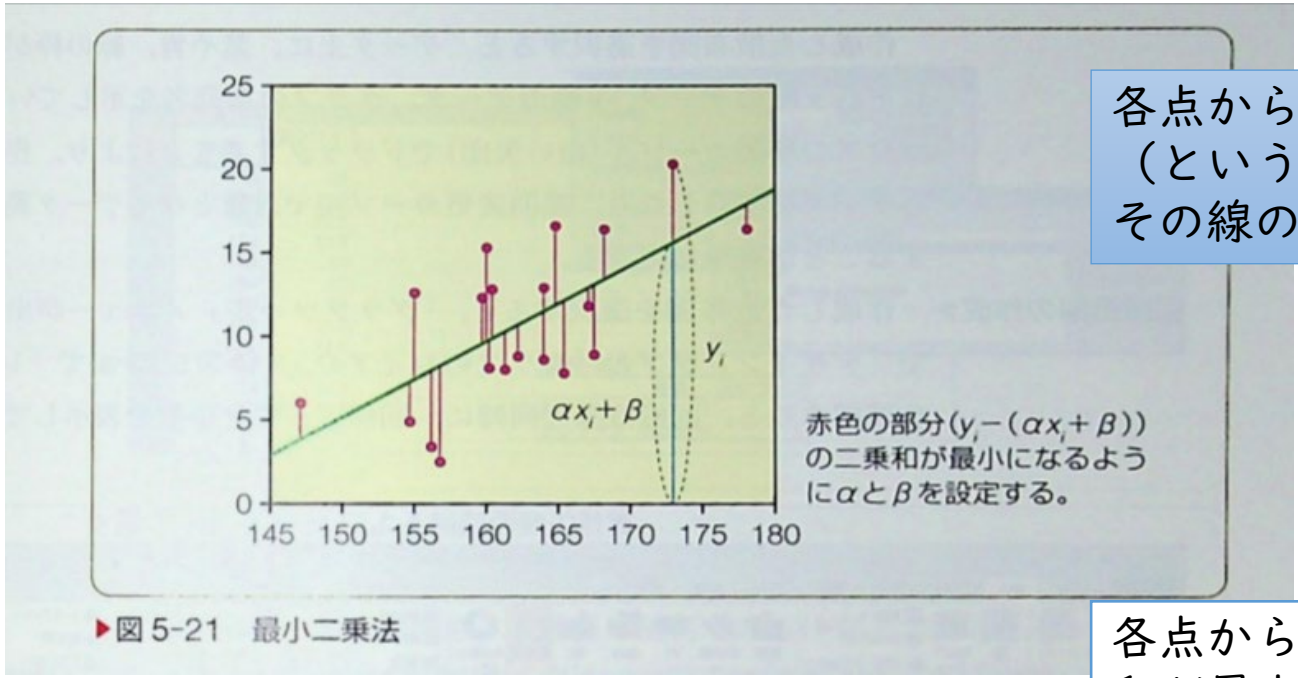
実際のデータは　ばらつく

なぜばらつくのか？については14-3．で



都道府県別人口（2015年国勢調査）と新型コロナウイルス感染症による死亡者数（NHK、2021年6月30日時点）

回帰係数、切片をどのように求めるか？



▶図 5-21　最小二乗法

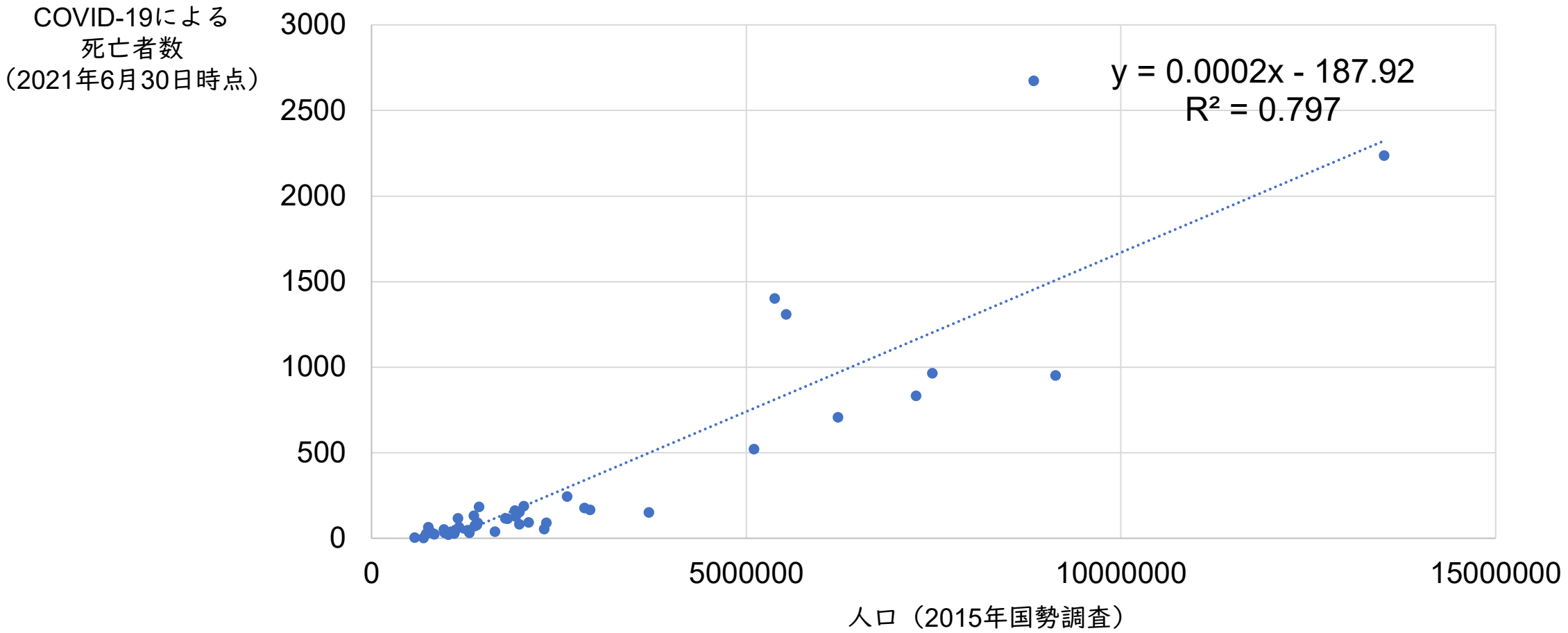
各点から直線に対して、x軸と垂直になるような（ということはy軸と平行するような）線を引き、その線の二乗和が最少になるようにする

各点から直線に対して垂線を引き、その線の二乗和が最少になるようにする

出典　高木晴良（2016）『系統看護学講座　基礎分野　統計学』医学書院, p154-155.

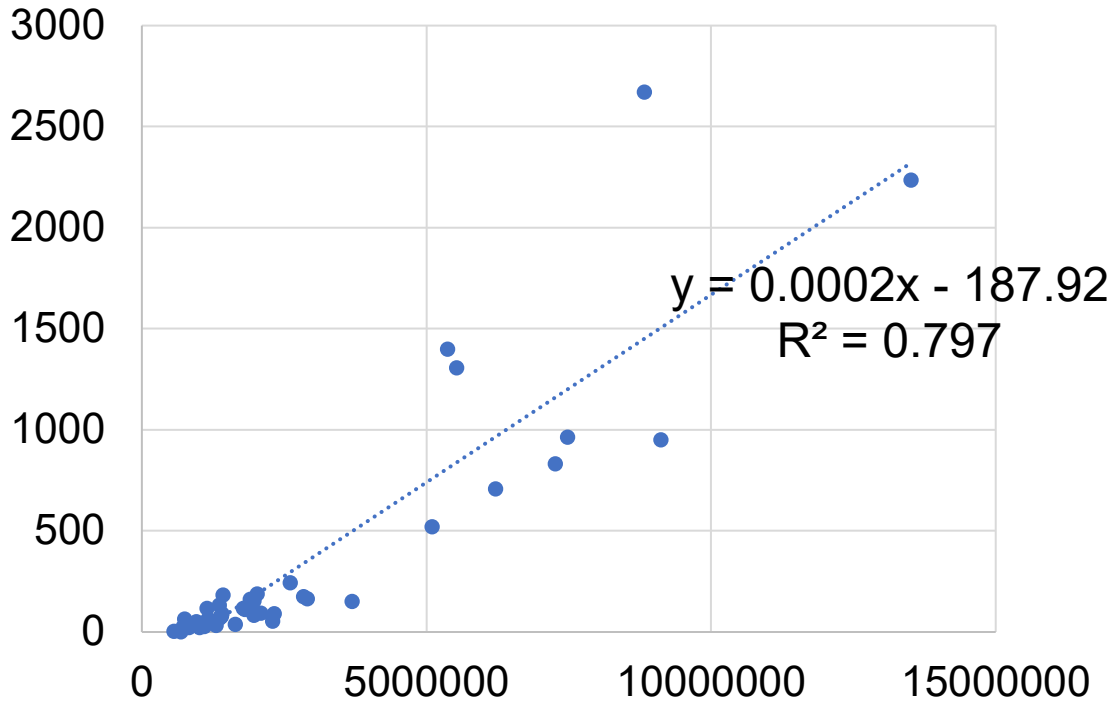
実際のデータは　ばらつく

なぜばらつくのか？については14-3．で



都道府県別人口（2015年国勢調査）と新型コロナウイルス感染症による死亡者数（NHK、2021年6月30日時点）

COVID-19による単回帰分析から何がみえてくるか？



都道府県別人口（2015年国勢調査）と新型コロナウイルス感染症による死亡者数（NHK、2021年6月30日時点）

何がみえてくるか？

（日本では）人口が多いほど、COVID-19による死亡者数が多い

ただし、第三の要因が関わっている可能性（交絡）は常に頭に入れておく

他には？

回帰直線から外れている都道府県は？

予測と評価

今日のテーマ

14-1. 回帰分析とは

14-2. この検定の注意点・論文の読み方

14-3. バイアスと交絡

14-2. この検定の注意点・論文の読み方

単回帰分析の注意点

1. 一方の変数（x; 独立変数）がもう一方の変数（y; 従属変数）を規定する関係が推論される場合に用いる。

2. 基本的な関係は　y=ax+b　直線的ではない関係が想定される場合、x^2などを独立変数とすることもある。

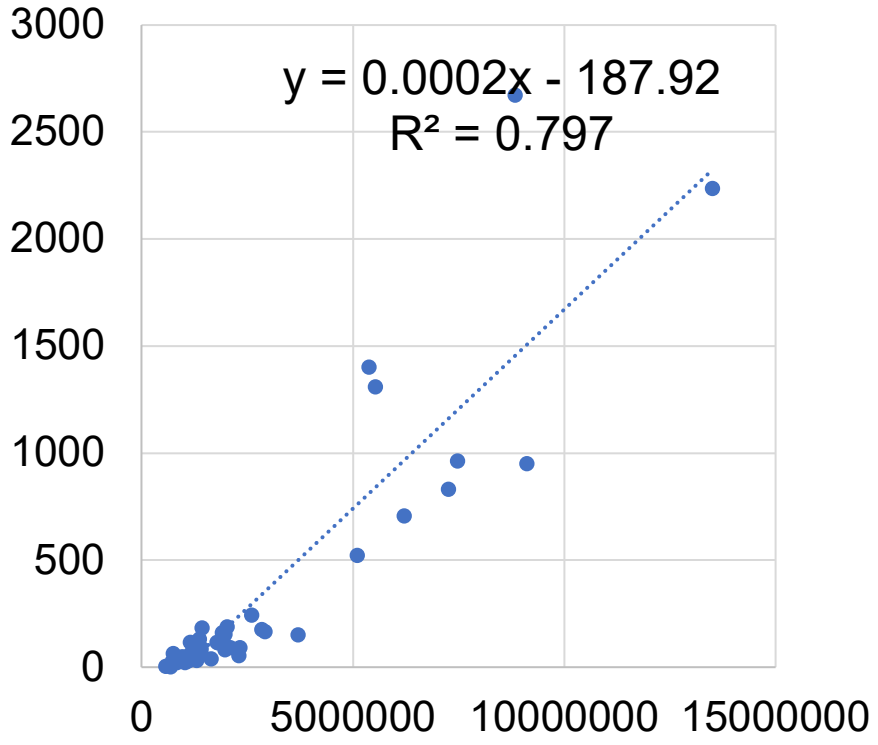
3. yとxを入れ替えると、係数が変わる。

4. 単回帰分析の結果は必ずしも両者の間の因果関係を意味しない。

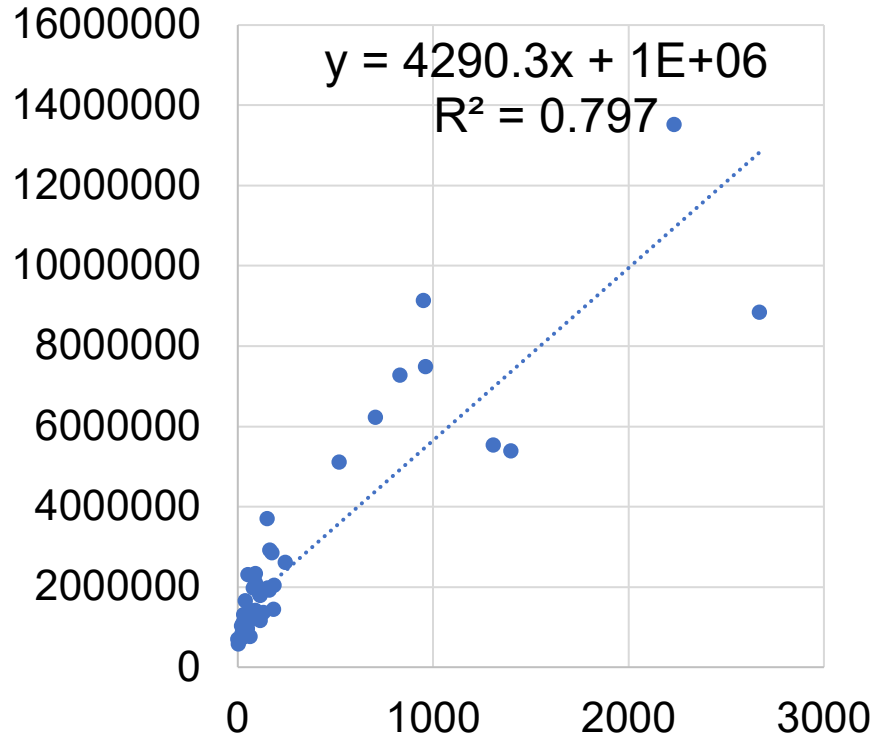
出典　中村 好一（2016）基礎から学ぶ　楽しい保健統計. 医学書院, p102-114.

3. yとxを入れ替えると、係数が変わる

COVID-19による死亡者数（2021年6月30日時点）



人口（2015年国勢調査）



都道府県別人口（2015年国勢調査）と新型コロナウイルス感染症による死亡者数（NHK、2021年6月30日時点）

相関 correlation と回帰 regression

- 同じように散布図を使って求めるため、混同されがち

- 相関係数はxとyを入れ替えても値が変わらない。
  （2変数の関係が対等なので）

- 回帰分析ではxとyを入れ替えると（通常は）値が変わる。
  yが従属変数（被説明変数）、xが独立変数（説明変数）

4. 単回帰分析の結果は必ずしも両者の間の因果関係を意味しない

原因があったから、結果が起きた

要因Aがあると現象Bが起こっているようにみえるが、
原因と結果の関係にはない

今日のテーマ

14-1. 回帰分析とは

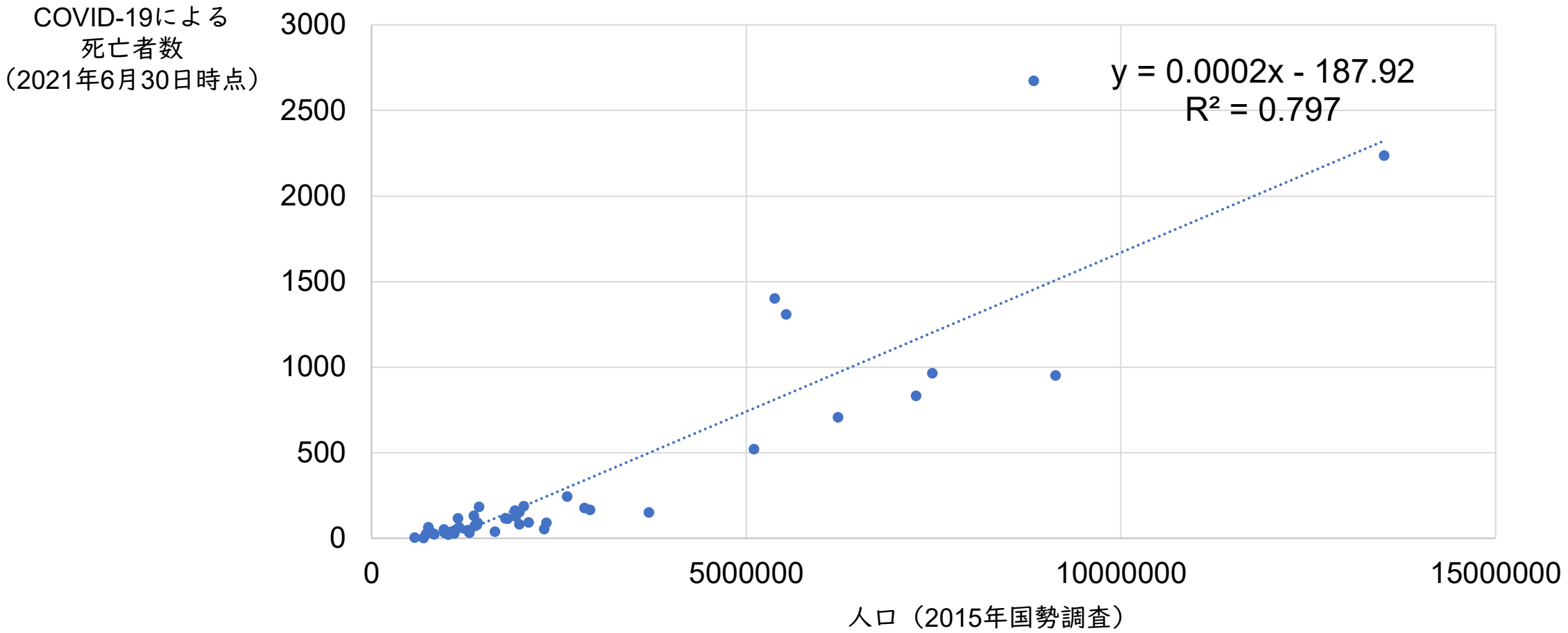
14-2. この検定の注意点・論文の読み方

14-3. バイアスと交絡

14-3. バイアスと交絡

実際のデータは　ばらつく

なぜばらつくのか？については14-3．で

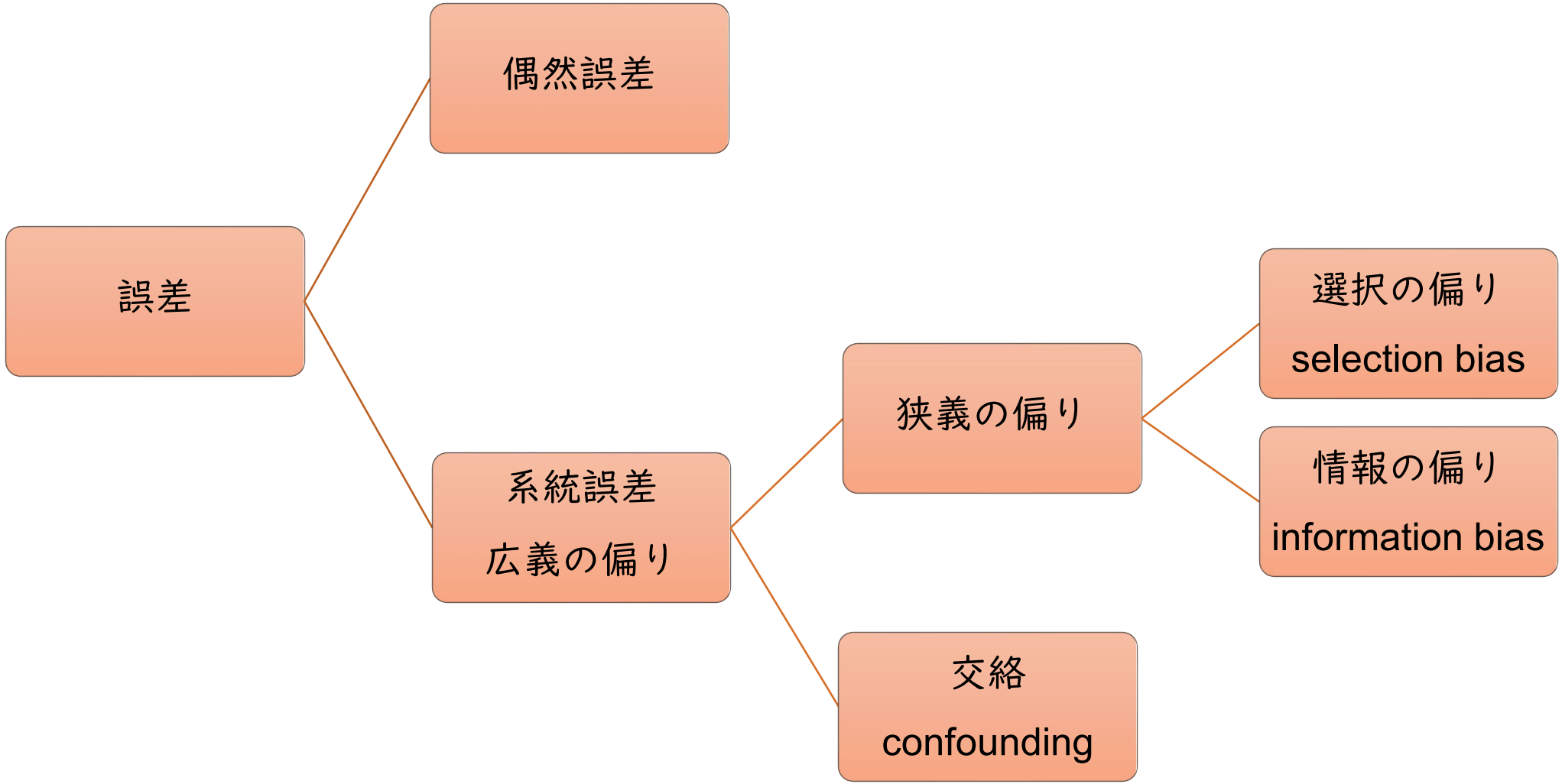


都道府県別人口（2015年国勢調査）と新型コロナウイルス感染症による死亡者数（NHK、2021年6月30日時点）

疫学研究における課題

- そもそも観察結果って一致しないよね・・・ 偶然誤差
- 本当に正確に答えてくれているのかな・・・ 情報バイアス
- だいぶ偏った人達ばかり集まっちゃったな・・・ 選択バイアス
- もし調べていない曝露が影響していたらどうしよう・・・ 未知の交絡

誤差の分類



出典　中村好一（2013）「第5章　偏りと交絡」『楽しい疫学』p83-96.

選択の偏り　selection bias

- 標的集団から観察対象集団を抽出する場合に、偏った抽出方法を行った結果生じるバイアス

- 標的対象集団から無作為抽出（random sampling）を行うのが原則

- 以下の場合は偏ったサンプルとなる
  - 無作為抽出が行えなかった場合
  - 無作為抽出が行えても回収率、回答率に著しい差がみられた場合

出典　中村好一（2013）「第5章　偏りと交絡」『楽しい疫学』p83-117.

情報の偏り　information bias

- 別名　観察の偏り observation bias、あるいは誤分類 misclassification。
- 思い出しバイアス recall bias
- 面接者バイアス interview bias

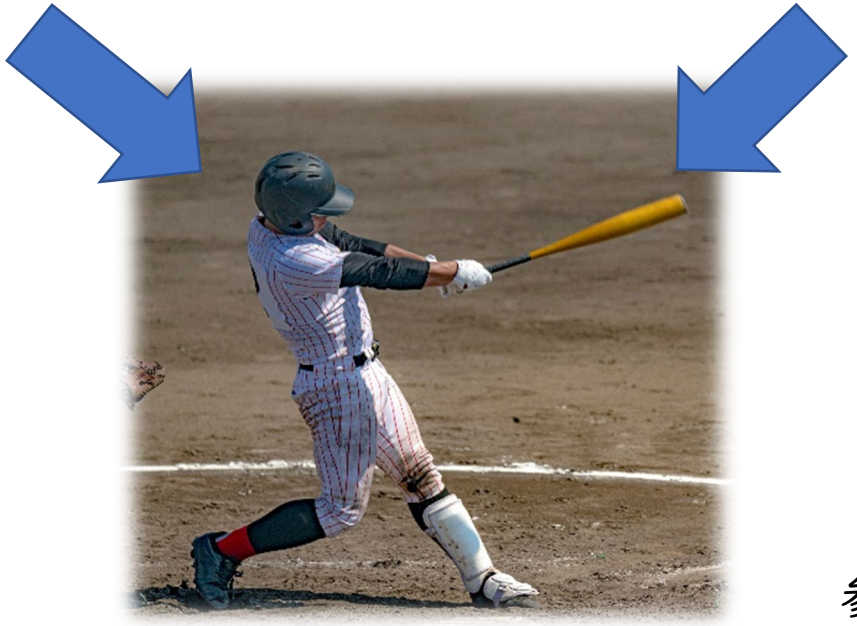
出典　中村好一（2013）「第5章　偏りと交絡」『楽しい疫学』p83-117.

大きく影響を与えるのが社会的望ましさ social desirability

the tendency of some respondents to report an answer in a way they deem to be more socially acceptable than would be their "true" answer

“Social Desirability” Encyclopedia of Survey Research Methods (edited by Paul J. Lavrakas) Sage Publications, 2008

交絡（confounding）とは？



子どもがプロ野球選手になった

父親と母親の両方が
「自分に似たんだ」と主張

どちらの影響もあるはず

こういった状況を
「交絡」confounding
と呼ぶ

参考：石村貞夫『すぐわかる統計用語』

交絡（confounding）とは？

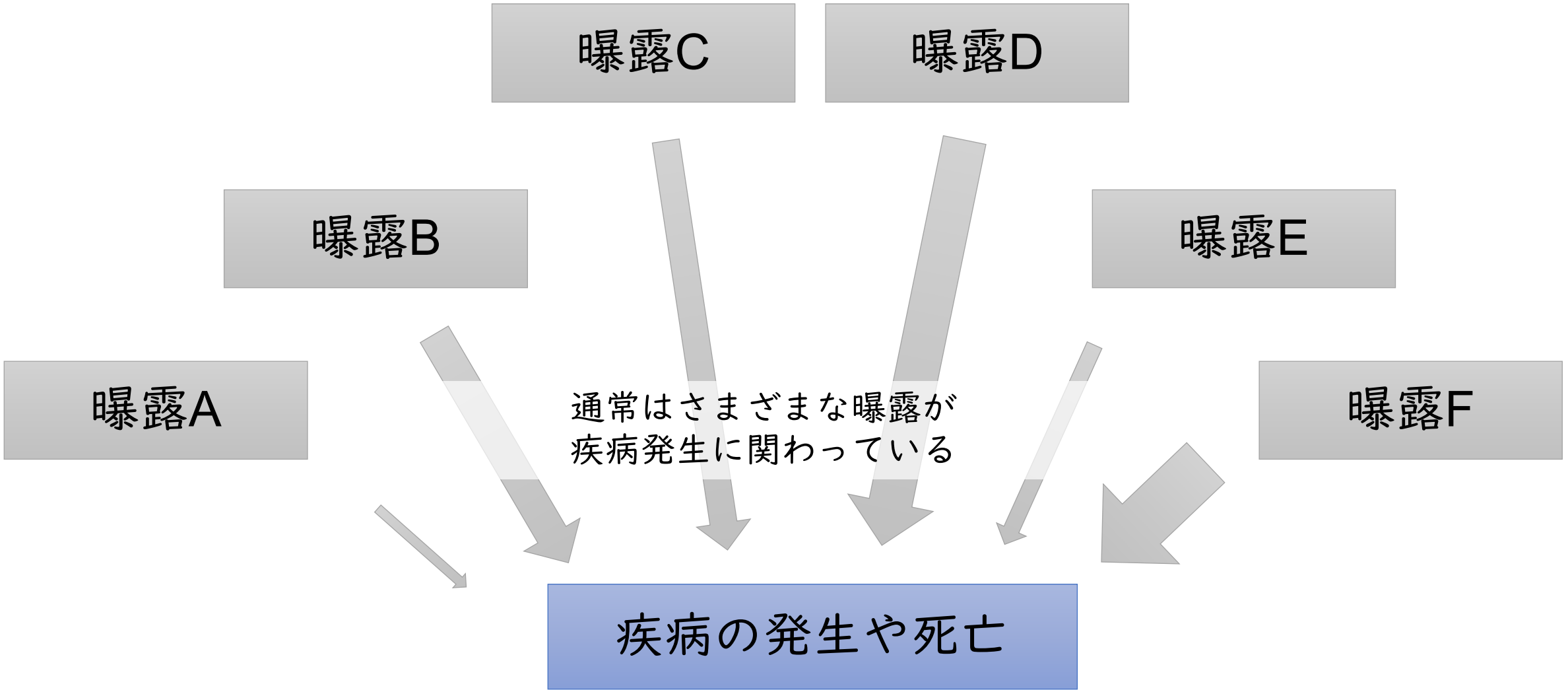
実際に父親も母親も影響を
与えているだろうけれど・・・

支えてくれたチームメイト？

他にもっと重要な要因があるかも

優秀なコーチ？

交絡とは



そこで行われるのが、、、伝家の宝刀

無作為割付

Randomization

Random allocation

無作為抽出（random sampling）とは違います。気をつけて。

そこで行われるのが、、、伝家の宝刀

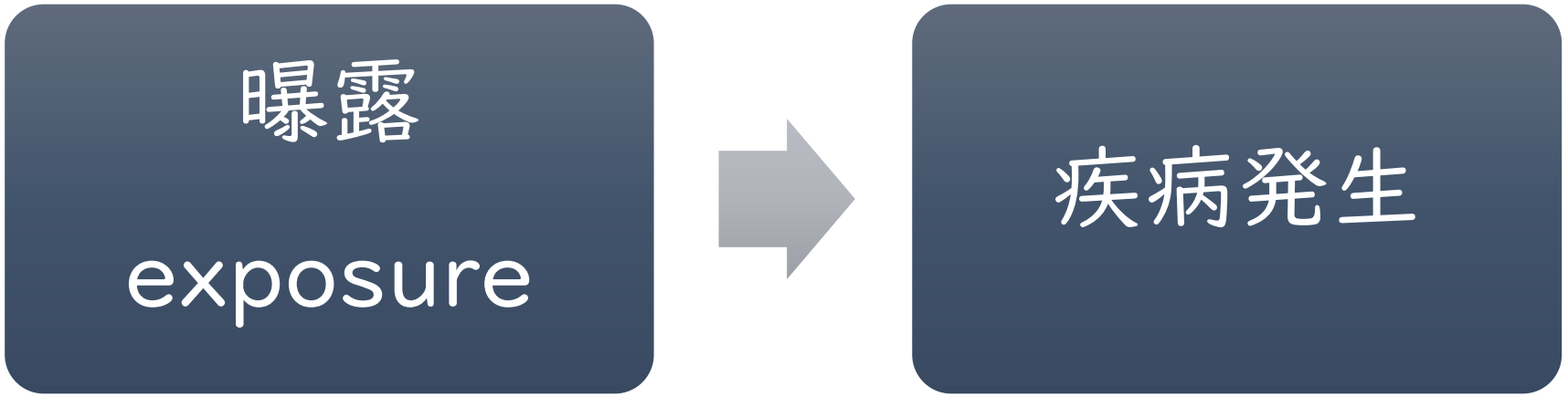
無作為割付 Random allocation を行う介入研究を

Randomized Controlled Trial (RCT)
あるいは
Randomized Clinical Trial (RCT)

などと呼ぶ

無作為抽出（random sampling）とは違います。気をつけて。

なぜ無作為割付（random allocation）が重要なのか？



曝露因子以外の因子が、ほぼ同じ状態の集団にすることができる（と考えられている）から

臨床研究・疫学研究における理想と現実

出典　杉山雄大, 井上浩輔, 後藤温（2021）「因果推論レクチャー　第1回　因果推論で医学研究を身近で素敵なものに」『週刊医学会新聞』3415: 5.

今日のテーマ

14-1. 回帰分析とは

14-2. この検定の注意点・論文の読み方

14-3. バイアスと交絡