統計的因果探索の出発点

石垣ゼミ3年
名執祐矢

目次

1. 統計的因果探索
2.擬似相関
3.反事実モデル
4.構造方程式モデル
5.構造的因果モデル
6.ランダム化実験
7.まとめ

１．統計的因果探索

統計的因果推論

◎統計的因果推論：「因果関係」についてデータから推測する
ためのの方法論

◎統計的因果探索：因果グラフを「未知」として、どのような
条件で因果グラフが推測可能なのかを明らか
にする研究

→なぜ、統計的因果探索という分野が存在するのか？

統計的因果推論

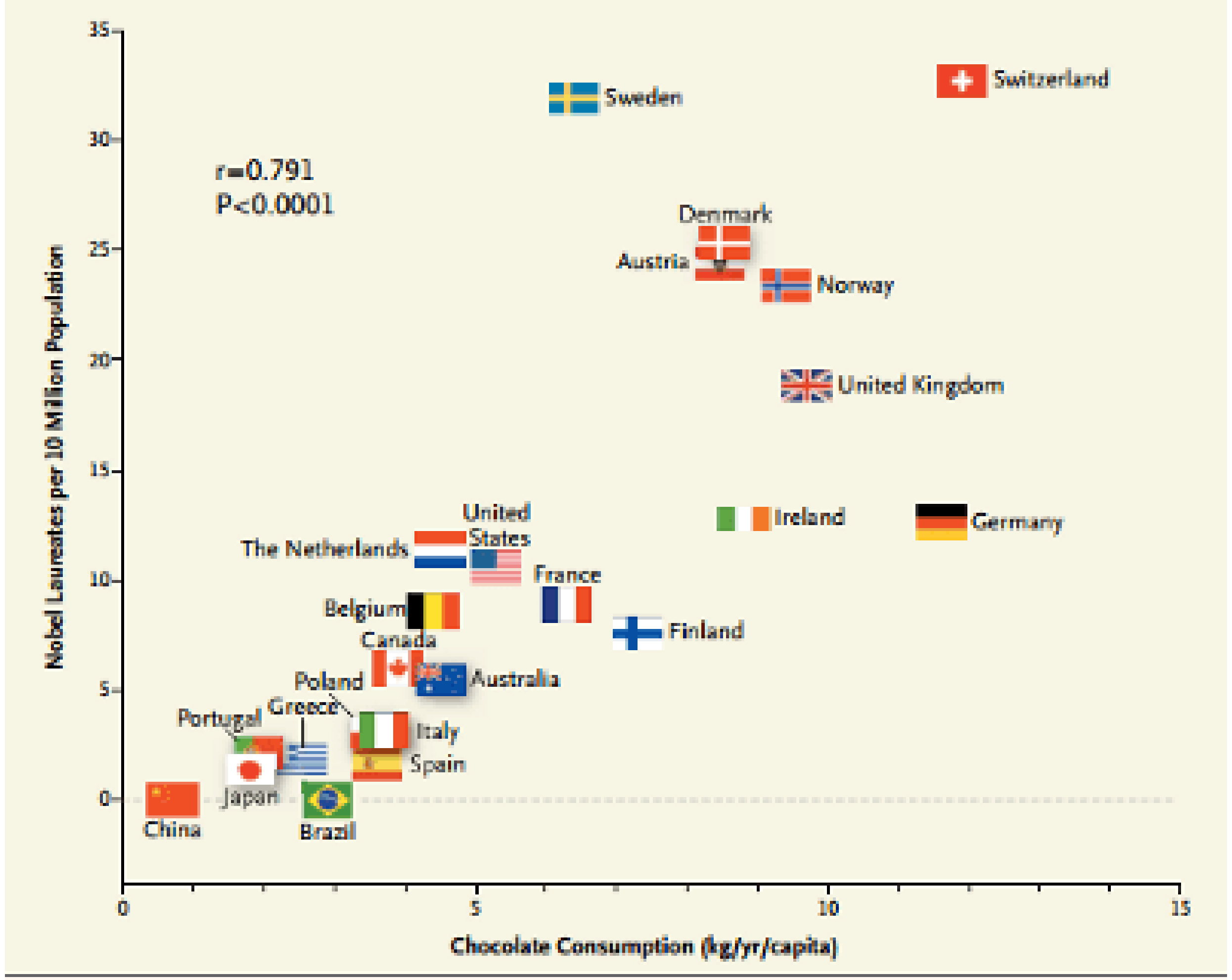


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

←「チョコレートの消費量」と
「ノーベル賞受賞者数」の散布図

※23カ国のデータ

※横軸：チョコレート消費量（年間1人あたり）

※縦軸：ノーベル賞受賞者数（人口1,000万人あたり）

チョコレートの消費量とノーベル賞の受賞者数の散布図． 出典：Franz H. Messerli, Chocolate Consumption, Cognitive Function, and Nobel Laureates. *New England Journal of Medicine*(367), 1563, Figure 1, Massachusetts Medical Society,2012.

統計的因果推論

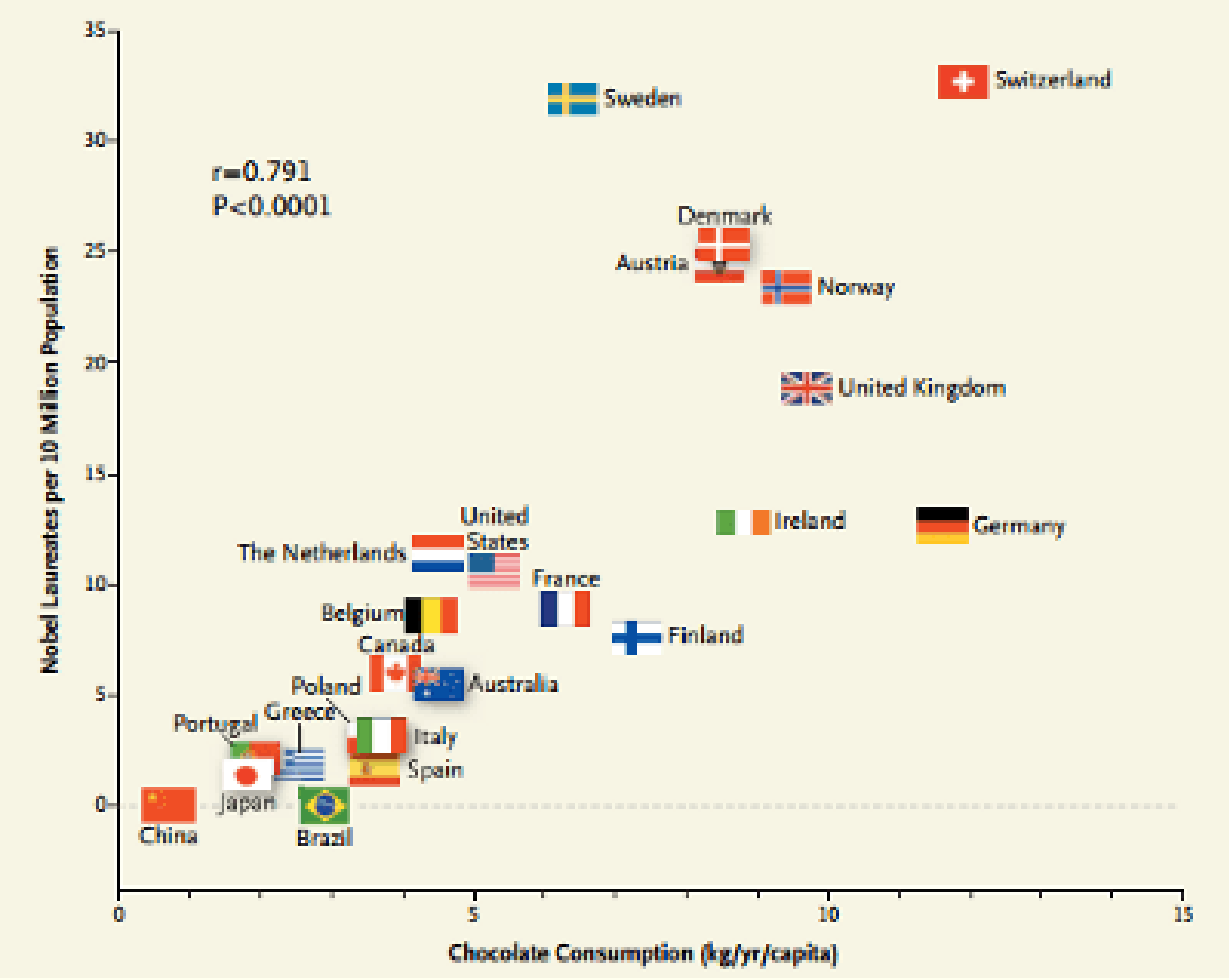


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

相関係数：0.79

→正の相関関係がある

予測1「チョコレートの消費量が多ければ、ノーベル賞受賞者数が多い」

予測2「ノーベル賞を多くとっているなら、チョコをたくさん食べている」

→深層学習でも予測可能

チョコレートの消費量とノーベル賞の受賞者数の散布図.　出典：Franz H. Messerli, Chocolate Consumption, Cognitive Function, and Nobel Laureates. *New England Journal of Medicine*(367), 1563, Figure 1, Massachusetts Medical Society,2012.

統計的因果推論

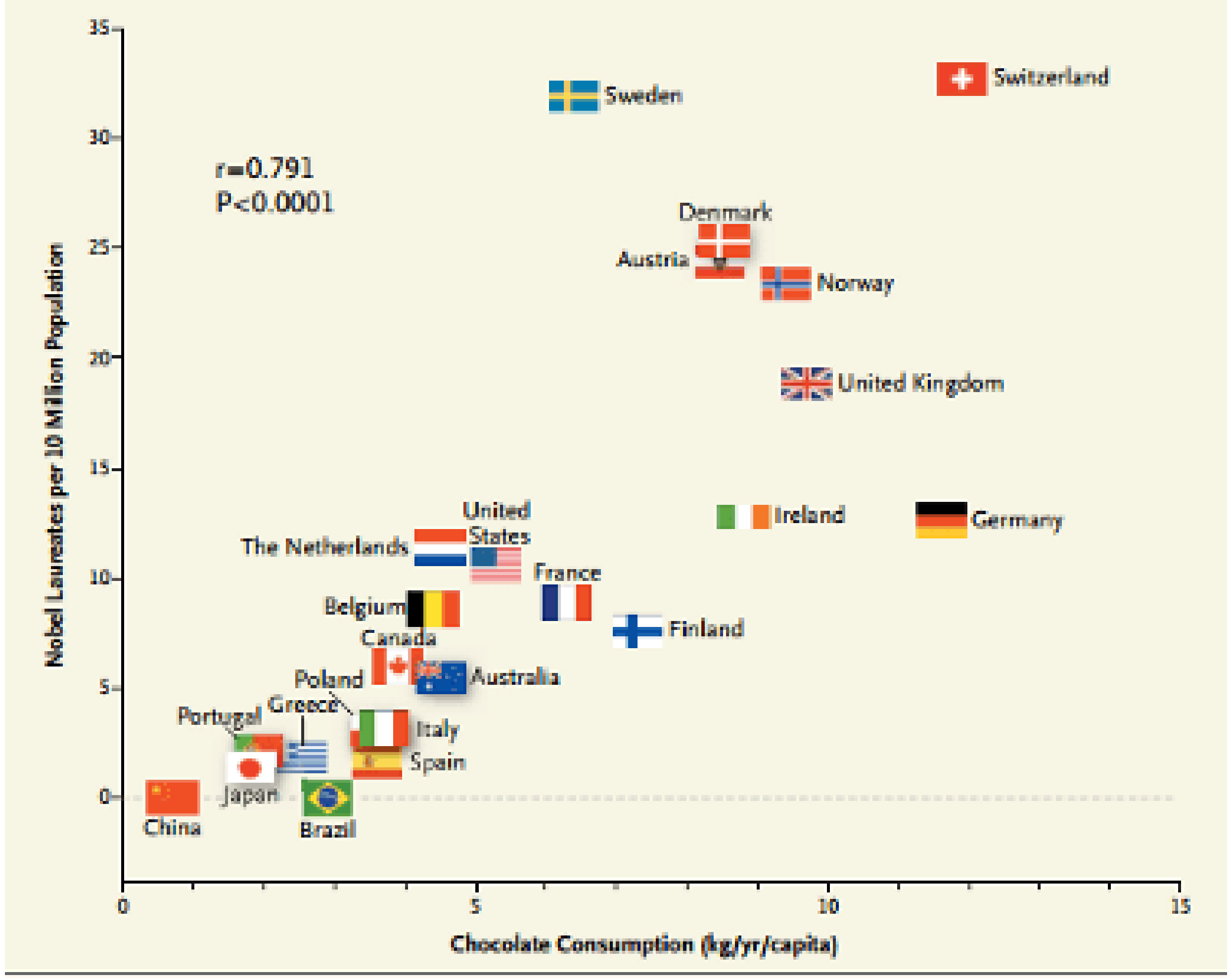


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

疑問「ノーベル賞受賞者数を増やすには
どうしたらよいか？」

「チョコをたくさん食べさせれば、
もっとノーベル賞をとれる？
どのくらい受賞が増える？」

→深層学習では因果推論を行わない

チョコレートの消費量とノーベル賞の受賞者数の散布図． 出典：Franz H. Messerli, Chocolate Consumption, Cognitive Function, and Nobel Laureates. *New England Journal of Medicine*(367), 1563, Figure 1, Massachusetts Medical Society,2012.

統計的因果推論

・因果関係において、適切な意思決定をするために、
　統計的因果推論が必要となる。

※ここでいう「因果関係」：変化を伴う関係

〈例〉

　「チョコレートの消費量が多ければ、ノーベル賞受賞者が多い」

≠「チョコレートの消費量を増やせば、ノーベル賞受賞者が増える」

統計的因果探索

◎統計的因果探索：因果関係が未知のときに、データから
因果関係の大きさを予測するための
機械学習技術

※古典的な統計的因果推論：因果関係が既知の場合を主に対象とする

→因果関係が未知である場合の困難は何か？

2. 擬似相関

擬似相関

◎擬似相関：「因果関係」にはないのに、「相関関係」は現れてしまうというギャップ（隔たり）のこと

〈例〉

「チョコレートの消費量が多ければ、ノーベル賞受賞者が多い」（相関関係がある）

≠「チョコレートの消費量を増やせば、ノーベル賞受賞者が増える」（因果関係がある）

→同じ相関関係を与えるような因果関係が複数あるから

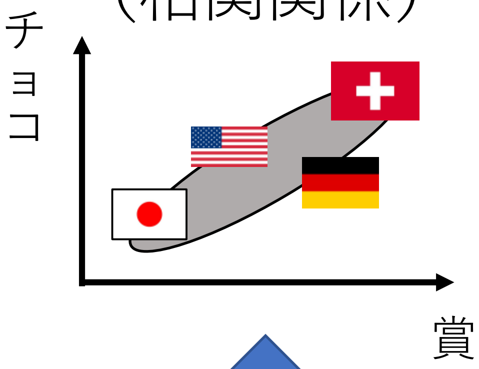
擬似相関

チョコレートをたくさん食べている国ほど受賞者が多い
(相関関係)

・四角で囲まれている変数：観測変数
（データが収集される変数）

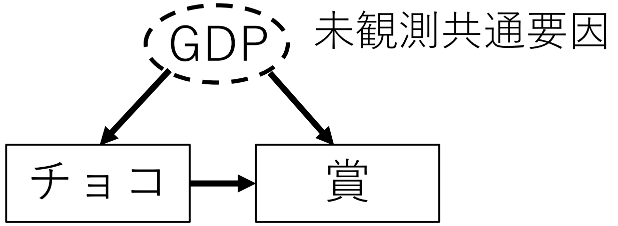
・点線の楕円で囲まれている変数
：未観測変数（データが収集されない変数）

・矢印の有無：因果関係の有無
（始点が原因、終点が結果の変数）

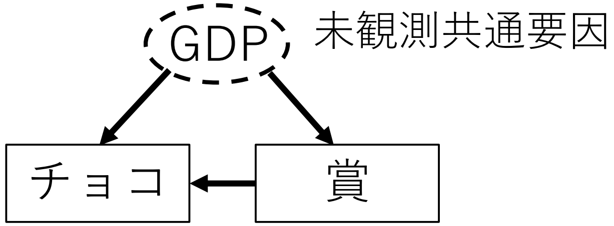


相関係数　0.79
P値　＜　0.0001

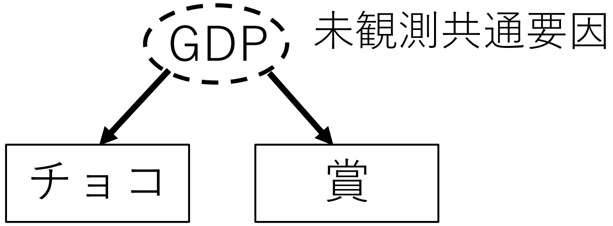
複数の因果関係が
同じ相関関係を与えうる



or



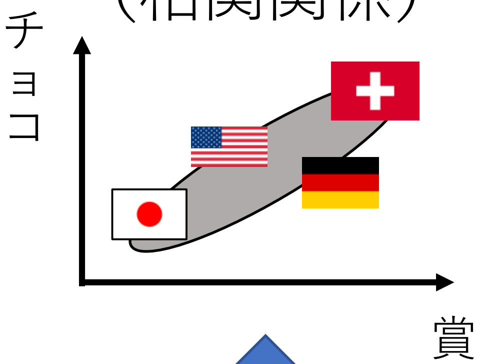
or



チョコレートをたくさん食べさせれば受賞者が増えるのか？
(因果関係)

擬似相関

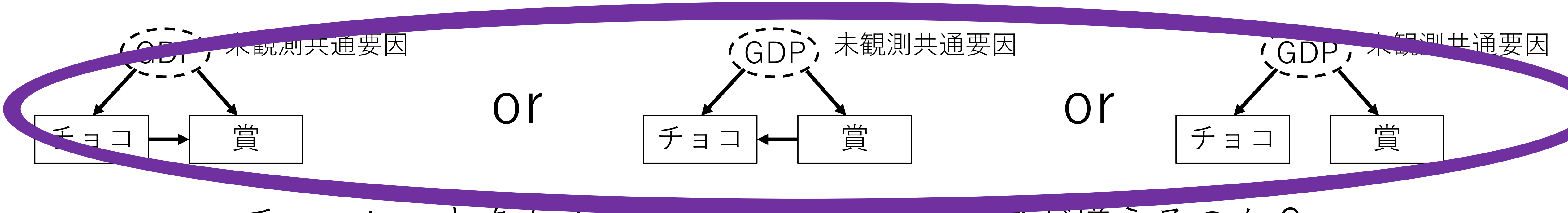
チョコレートをたくさん食べている国ほど受賞者が多い
（相関関係）



相関係数　0.79
P値　＜　0.0001

因果グラフ：定性的な因果関係を表す図
（因果関係の大きさの情報は含まない）

複数の因果関係が
同じ相関関係を与えうる

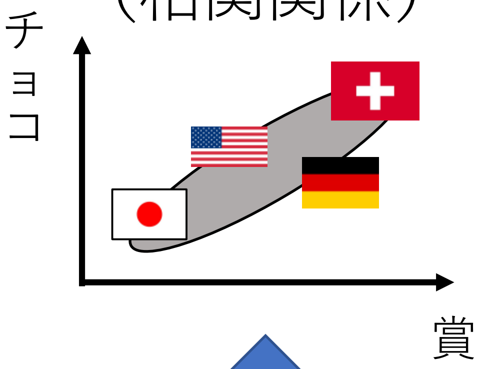


チョコレートをたくさん食べさせれば受賞者が増えるのか？
（因果関係）

擬似相関

チョコレートをたくさん食べている国ほど受賞者が多い

（相関関係）

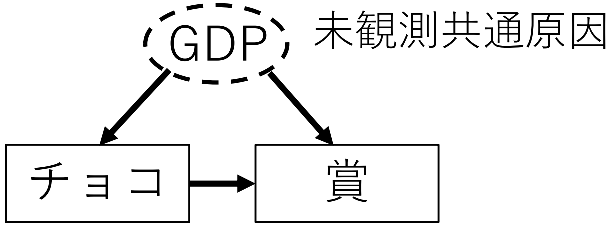


相関係数　0.79
P値　＜　0.0001

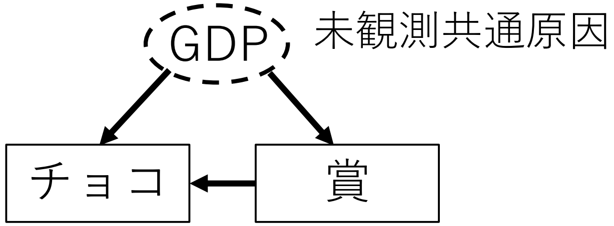
・GDP：チョコの原因、かつ、賞の原因
→共通原因

・GDPは未観測
→未観測共通原因

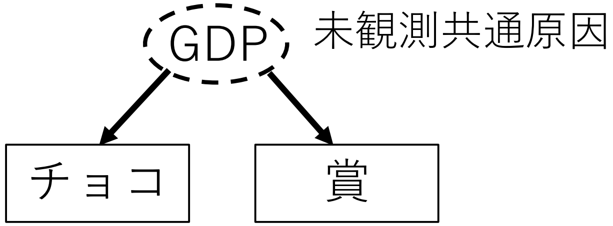
複数の因果関係が
同じ相関関係を与えうる



or



or



チョコレートをたくさん食べさせれば受賞者が増えるのか？

（因果関係）

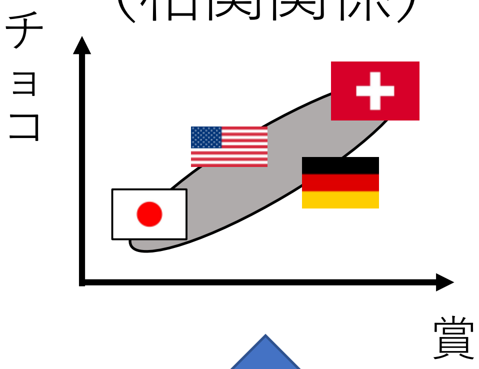
擬似相関

チョコレートをたくさん食べている国ほど受賞者が多い
(相関関係)

右の因果グラフ
→「GDPを増やせば、チョコの消費もノーベル賞受賞者数も増え」、

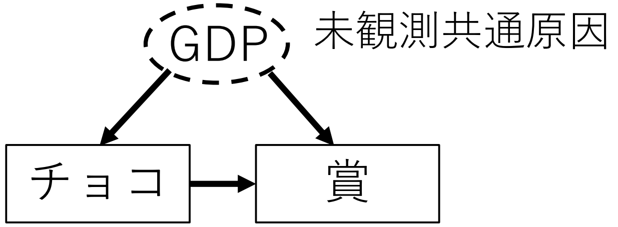
「チョコの消費が多ければ、ノーベル賞受賞者数が多い」

という相関関係が現れる

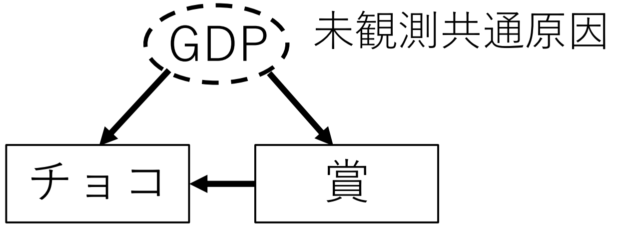


相関係数　0.79
P値　<　0.0001

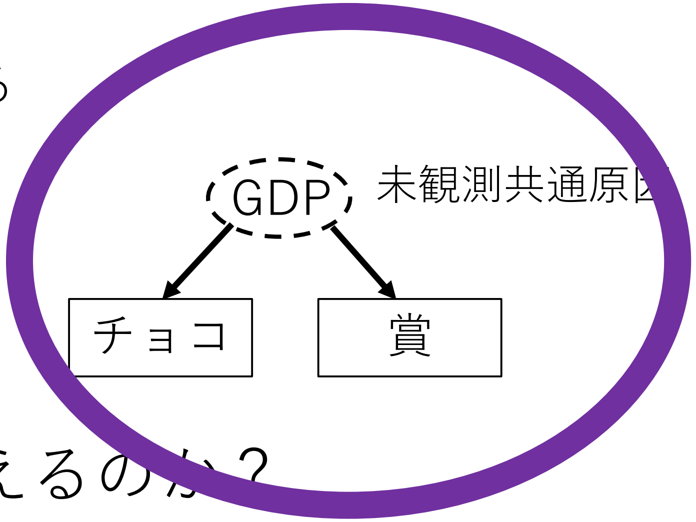
複数の因果関係が
同じ相関関係を与えうる



or



or



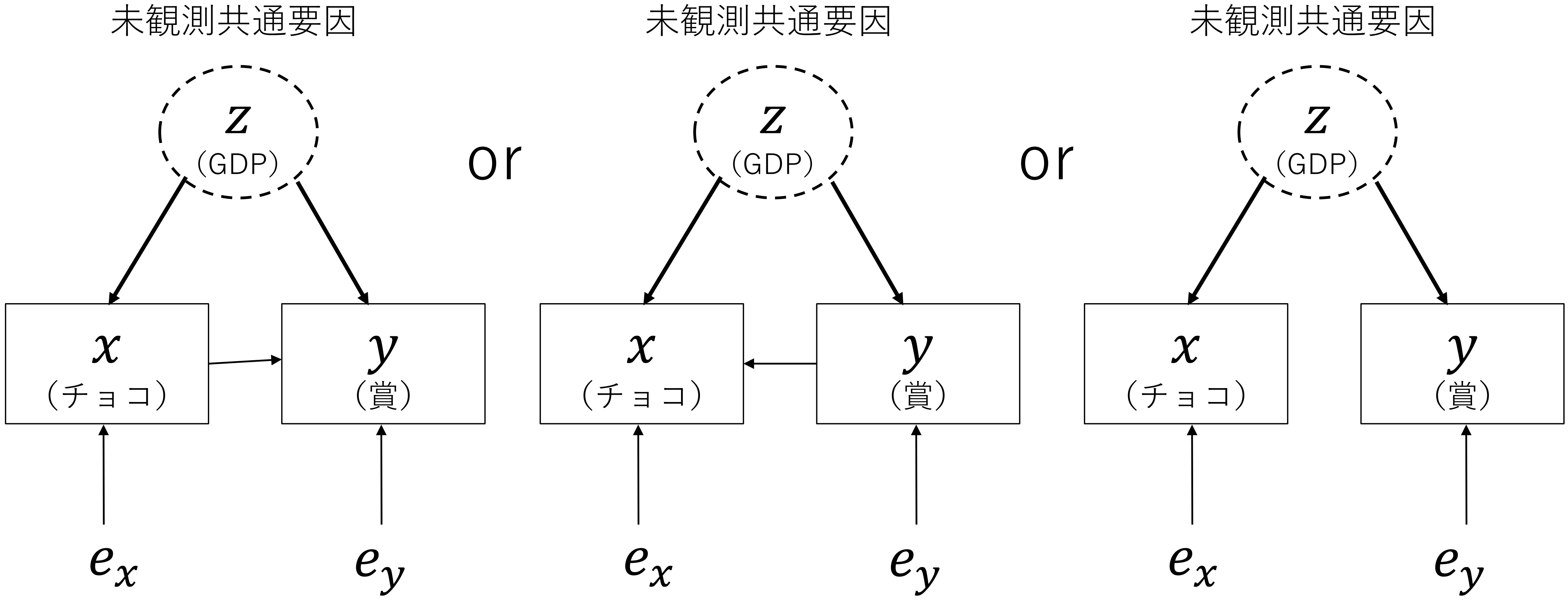
チョコレートをたくさん食べさせれば受賞者が増えるのか？
(因果関係)

擬似相関

・因果関係を相関係数の値だけから推測することはできない

→データから因果関係を推測することが、統計的因果探索の目的となる。

因果グラフに記号を導入



因果グラフに記号を導入

・e_x, e_y:誤差変数
（左グラフにおけるe_y: x, z以外にyの値を決める変数すべてを
1つにまとめて表したもの）

記号を使ったyのデータ生成過程

$$y = b_{yx}x + \lambda_{yz}z + e_y$$

・y,x,z,e_y ：確率変数

・b_{yx}, λ_{yz} ：定数（添字は、1つ目の文字が左辺の変数、2つ目の文字が対応する右辺の変数）

同様に $x = \lambda_{xz}z + e_x$

記号を使ったyのデータ生成過程

◎3種類の因果グラフのデータ生成過程

$$x \rightarrow y: \begin{cases} x = 0.3z + e_x \\ y = 0.7x + 0.3z + e_y \end{cases}$$

$$x \leftarrow y: \begin{cases} x = 0.7y + 0.3z + e_x \\ y = 0.3z + e_y \end{cases}$$

$$x \quad\quad y: \begin{cases} x = 0.89z + e_x \\ y = 0.89z + e_y \end{cases}$$

記号を使ったyのデータ生成過程

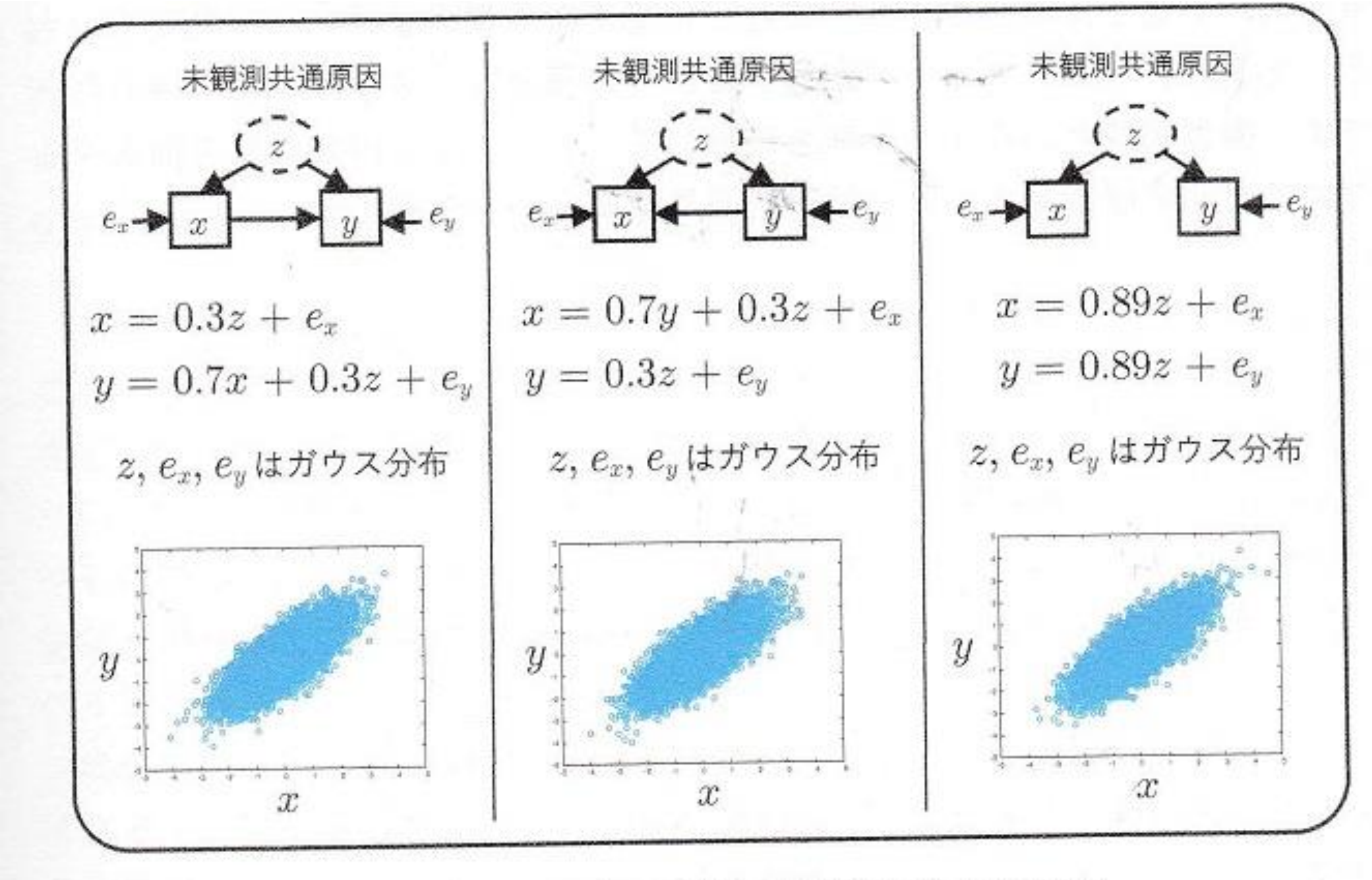


図 1.4 x と y に関する 3 種類の因果グラフとその散布図.

まとめ

・統計的因果探索の出発点は、相関関係と因果関係のギャップ（擬似相関）

・↑のギャップが起こる最大の理由は、未観測共通原因の存在

・擬似相関というギャップを懐柔し、データから因果関係を推測する方法を研究するのが、統計的因果探索

統計的因果推論

◎統計的因果推論：「因果関係」についてデータから推測する
　ためのの方法論

◎因果関係：何かを「変化」させたときに、何かほかのものが
「変化」すれば、その2つは「因果関係にある」と
いう

→数学的枠組みで考えてみる
（因果関係は数学的にはどう表せるのか？）

統計的因果推論の数学的枠組み

反事実モデル
（因果のモデル）

構造方程式モデル
（データ生成過程のモデル）

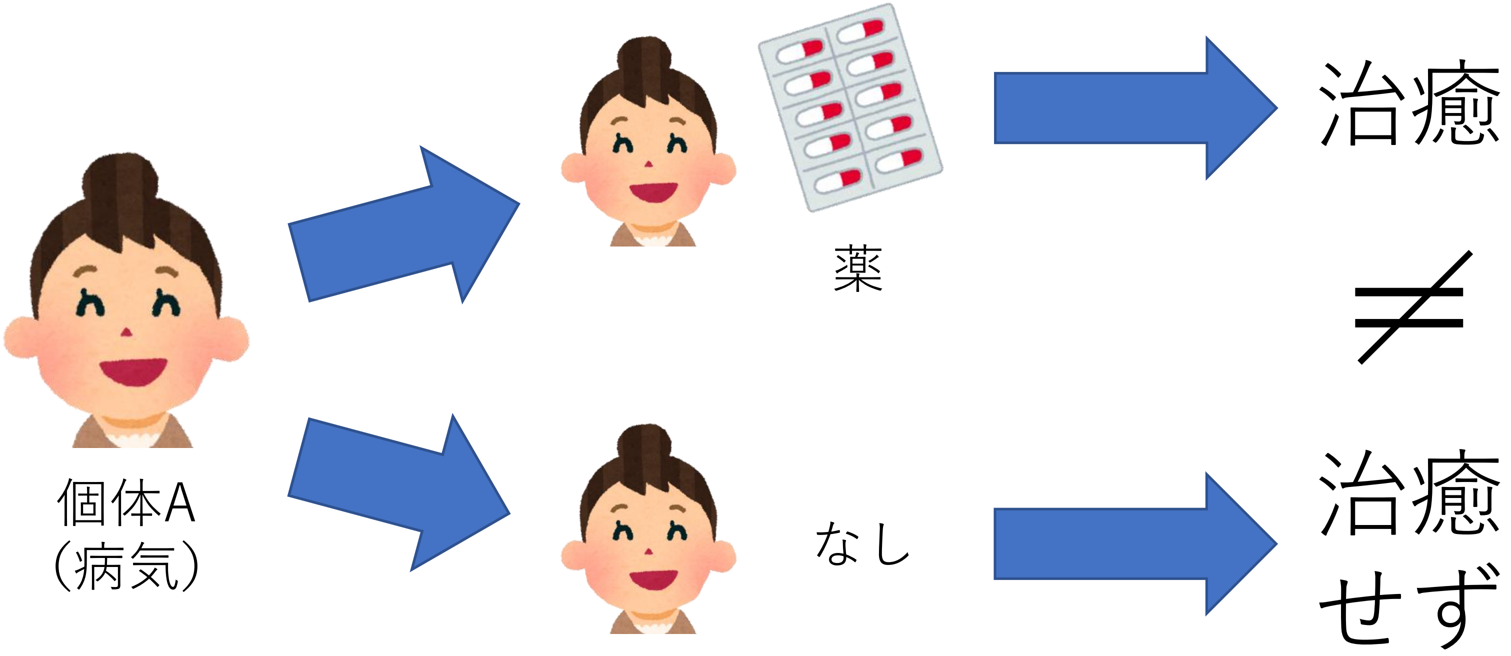
構造的因果モデル
（因果推論のための枠組み）

＋ランダム化実験
（因果関係を推測する方法）

3．反事実モデル

個体レベルの因果

◎個人レベルの因果：個体について考える因果



・個体Aにとって、投薬は治癒の原因となるか？

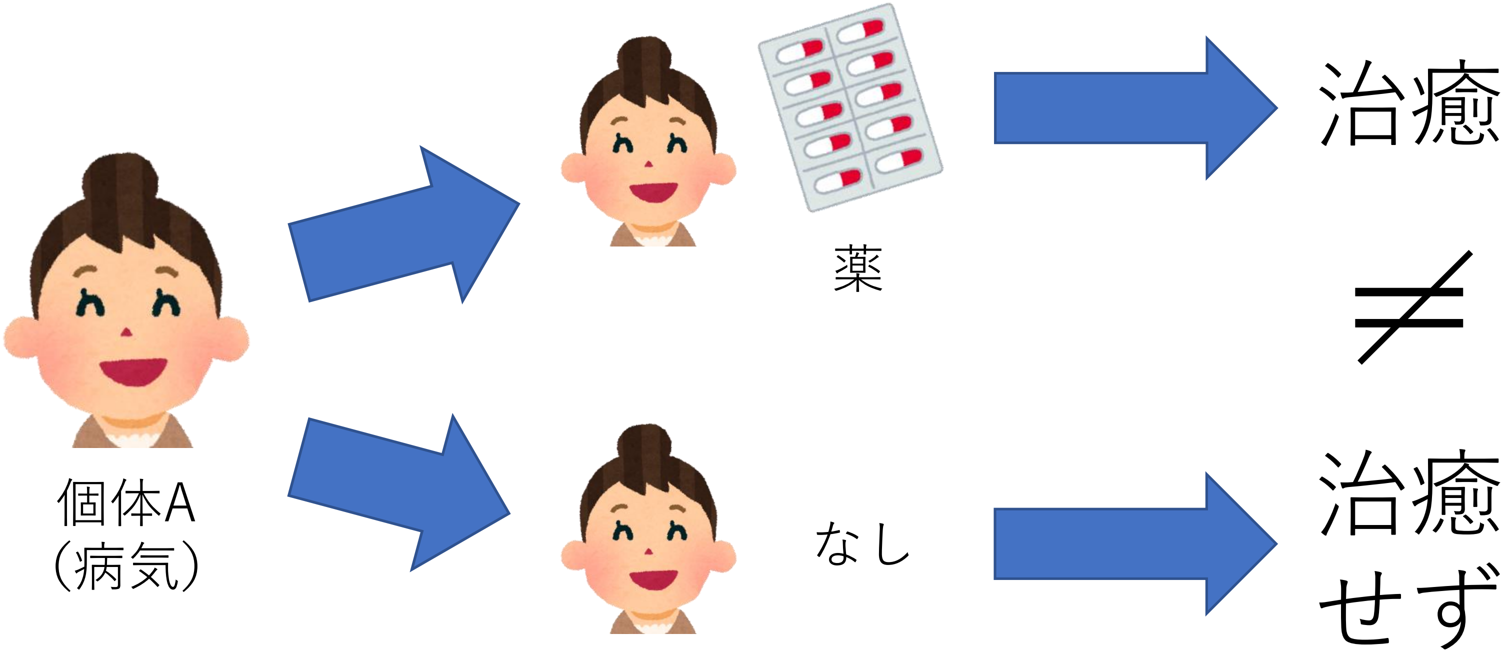
→次の2つを比較

ーもしも薬を飲んでもらったとしたら、治癒するか

ーもしも薬を飲まないでもらったとしたら、治癒するか

個体レベルの因果

◎個人レベルの因果：個体について考える因果



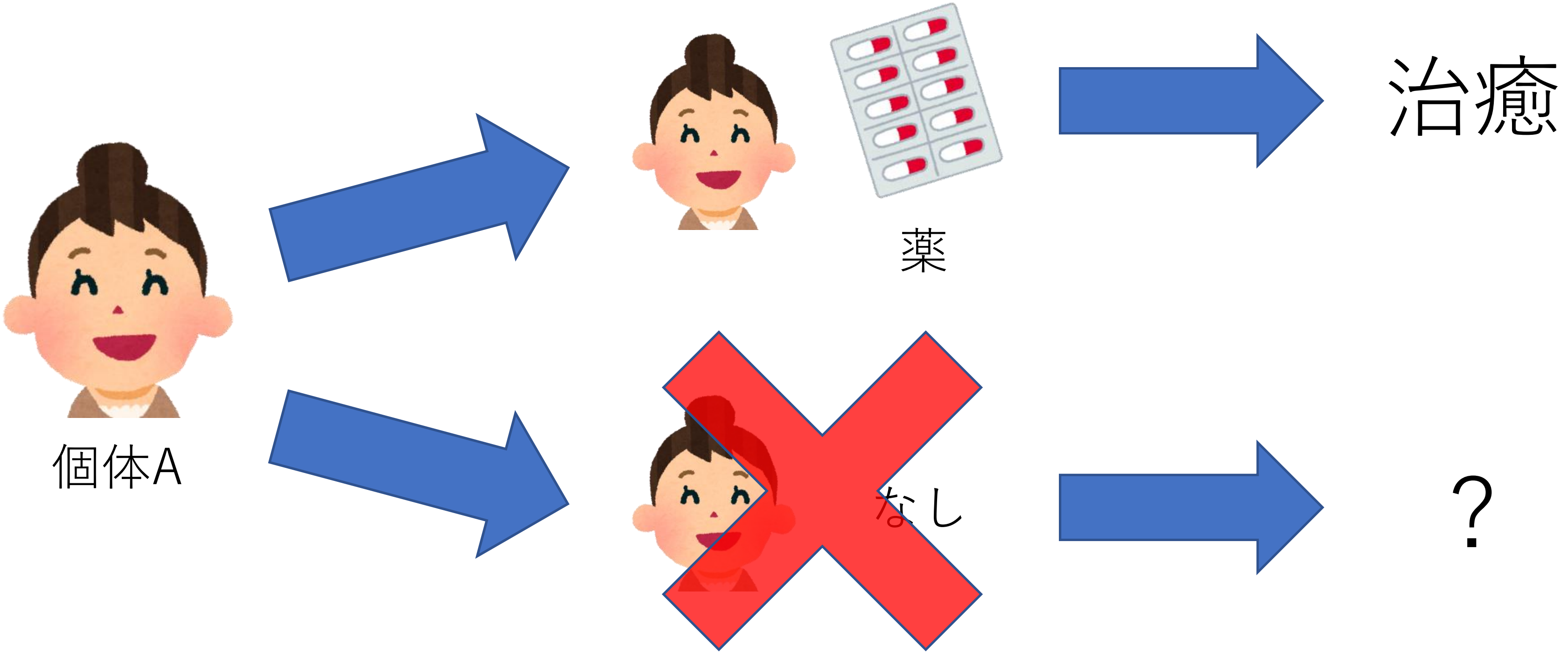
・左の例：2つの行動の結果に違い

→個体Aにおいて、薬を飲むかどうかが、病気が治るかどうかの原因となる

→個体Aにおいては、薬を飲むという行為に、病気を治す効果がある

因果推論の根本問題

◎因果推論の根本問題：個体のデータに基づいて、個体レベルの因果に関する結論が導けない

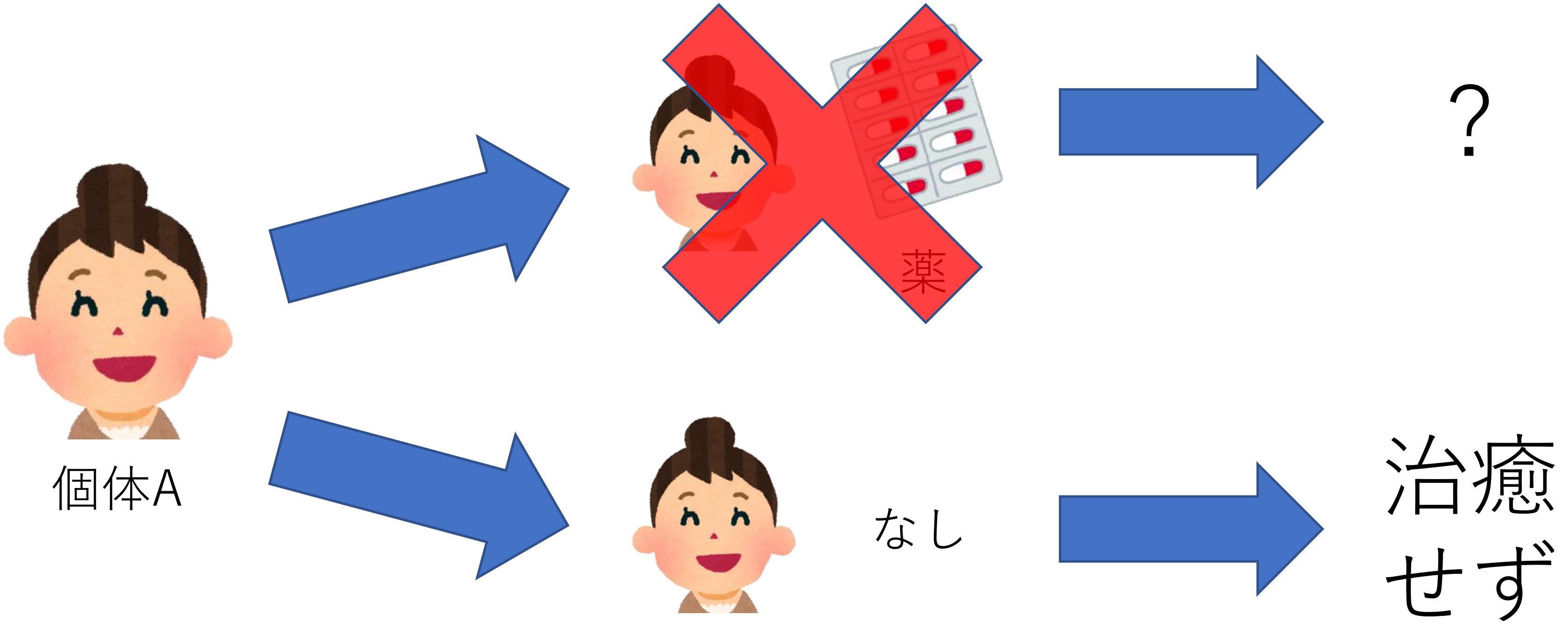


・観測できるのは
どちらか一方

・薬を飲んでもらってしまったら、薬を飲まないでもらった場合にどうなるか不明

因果推論の根本問題

◎因果推論の根本問題：個体のデータに基づいて、個体レベルの
因果に関する結論が導けない



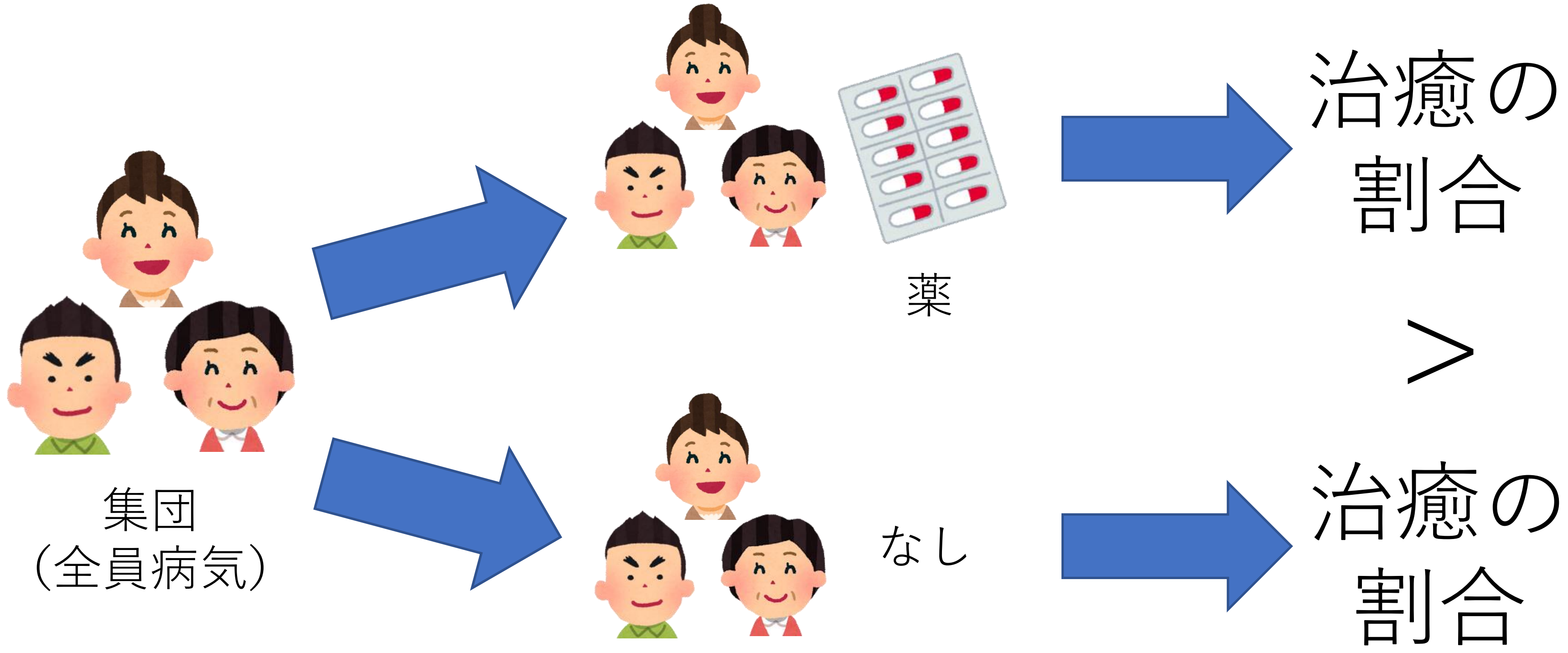
←の例では

・「薬を飲まないで
もらう」
⇒事実（行動の結果）

・「薬を飲んでもらう」
⇒反事実（実際に起こらない行動の結果）

集団レベルの因果

◎集団レベルの因果：集団について考える因果



・この「集団」にとって、
投薬は治癒の原因か？

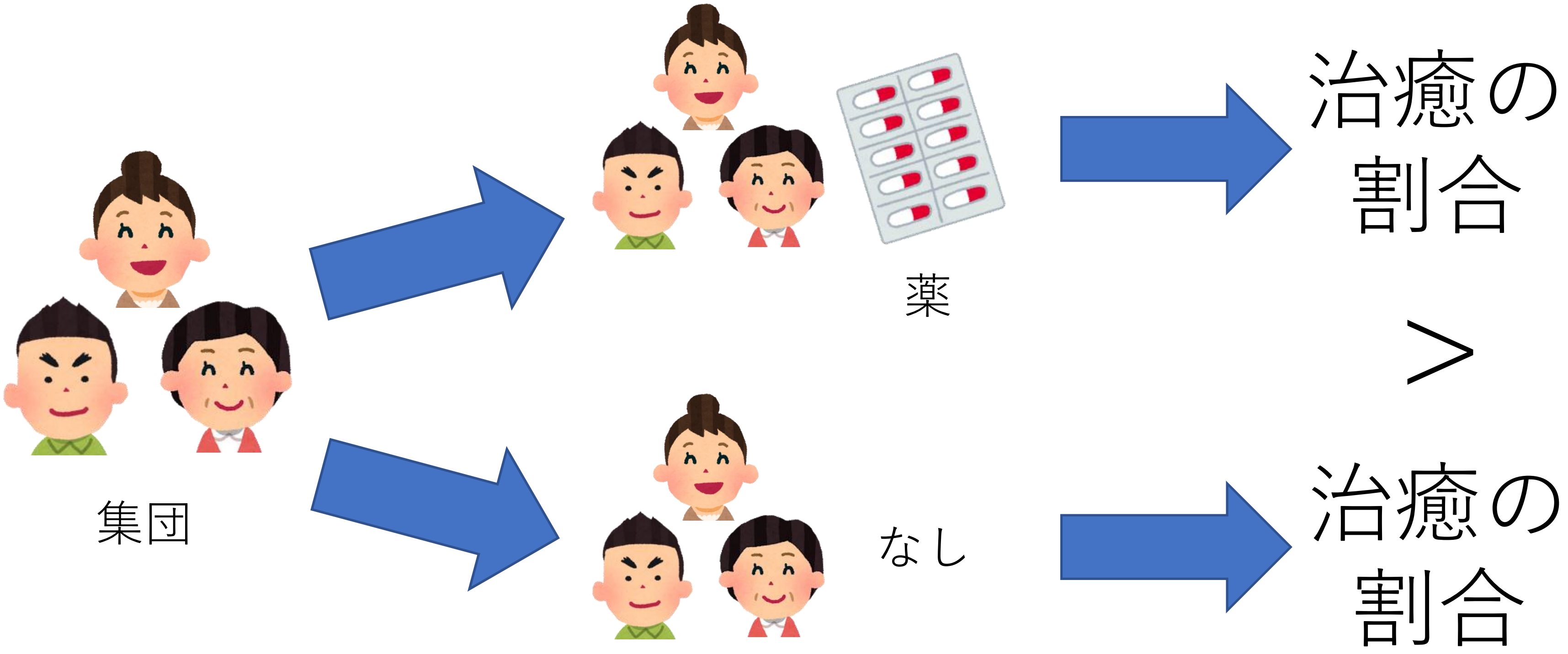
→次の2つを比較

－集団の個体全員に薬を
飲んでもらった場合に
治癒する割合

－集団の個体全員に薬を
飲まないでもらった場合に
治癒する割合

集団レベルの因果

◎集団レベルの因果：集団について考える因果



・左の例：2つの行動の結果に違い（治癒の割合に差）

→この集団において、薬を飲むかどうかが、病気が治るかどうかの原因となる

→この集団においては、薬を飲むという行動には、病気を治す効果がある

個体レベルの因果と集団レベルの因果

◎個体レベル：「治る」「治らない」の2択

一般に調べることはできない

◎集団レベル：「何割」の個体が治るかという集団としての

ふるまいに着目

調べることができる場合がある（後述）

4． 構造方程式モデル

構造方程式モデル

◎構造方程式モデル：どのように変数の値が決定されるかを表現

→構造方程式という等式を使う

構造方程式モデル（薬と病気の例）

〈薬と病気の例〉

・病気にかかっているかを表す変数yの値がどのように決定されるのか

$$y = f_y(x, e_y)$$ （←構造方程式）

y:病気にかかっているか（1:かかっている、0:かかってない）

x:薬を飲むかどうか（1:飲む、0:飲まない）

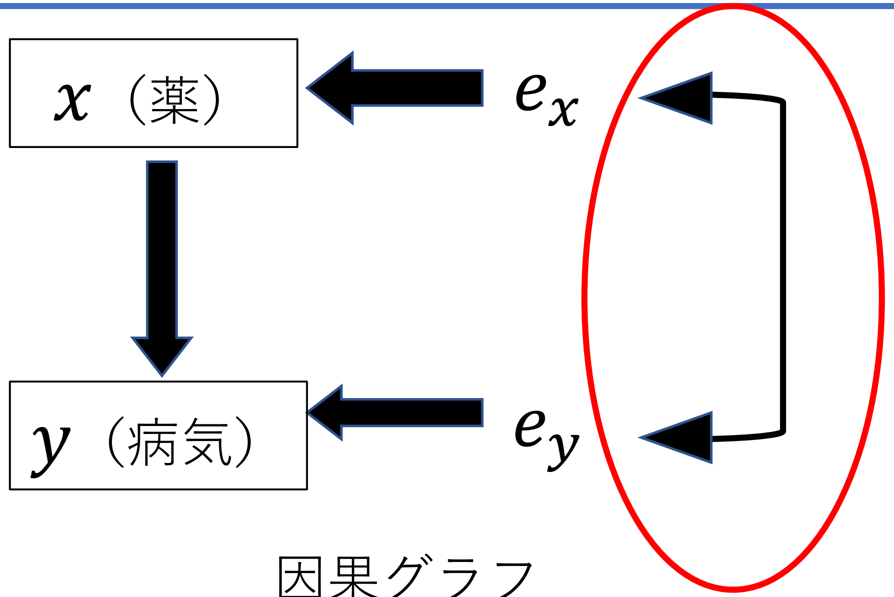
e_y: yの値を決定するために寄与しうるx以外の全ての変数（誤差変数）

観測変数：x, y　　未観測変数：e_y

構造方程式モデル（薬と病気の例）

$$x = e_x$$
$$y = f_y(x, e_y)$$

構造方程式
（↑yがどのように決定されるか）



因果グラフ

e_x, e_yが独立ではなく従属している可能性があることを表す

・データ生成過程のモデル

ー変数の「値」が、どういう過程を経て生成されるか

・構造方程式：変数の「値」の決定関係を表す

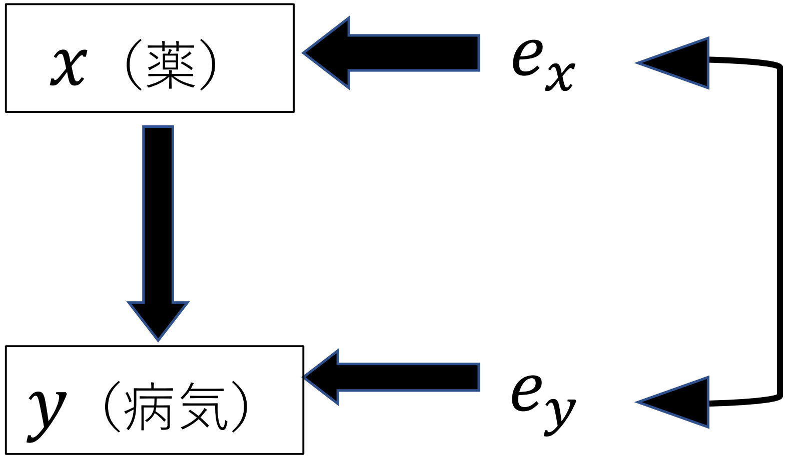
ー$y = f_y(x, e_y)$　》》単なる等式ではない：左辺を右辺で定義

》》e_y:yの値を決定するために必要なx以外の変数全て

構造方程式モデル（薬と病気の例）

$$x = e_x$$
$$y = f_y(x, e_y)$$

構造方程式



因果グラフ

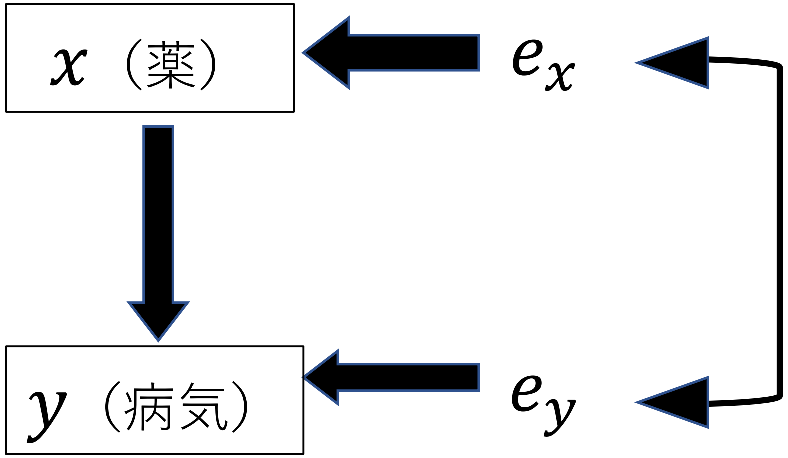
・y:内生変数（構造方程式の左辺に登場、
ほかのどの変数の値によって決まるのかが記述されている）

・e_x, e_y:外生変数（構造方程式の右辺にのみ登場、
どんな変数からどのような手順で生成されるか記述されていない）

構造方程式モデル（薬と病気の例）

$$x = e_x$$
$$y = f_y(x, e_y)$$

構造方程式



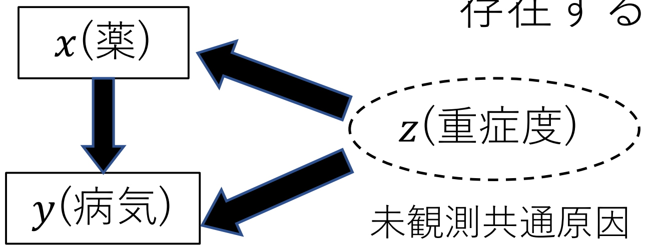
因果グラフ

片方向矢印（有向辺）

・構造方程式の左辺の値を計算するために
必要かもしれないとき

両方向矢印付き円弧（有向円弧）

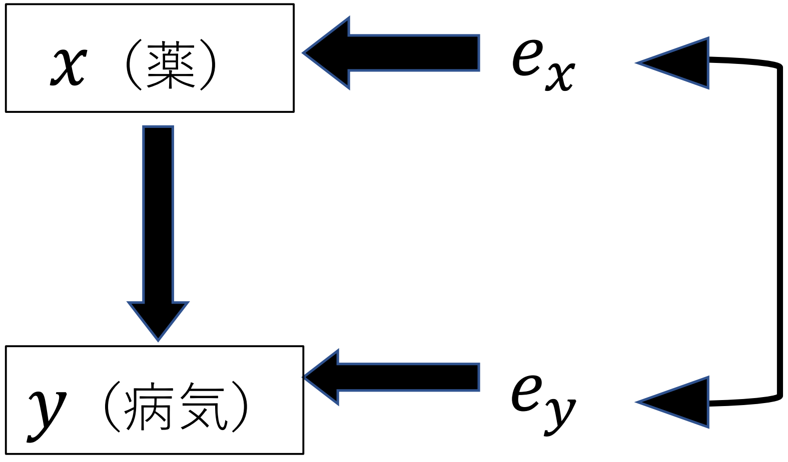
・未観測原因が
存在するかもしれないとき



構造方程式モデル（薬と病気の例）

$$x = e_x$$
$$y = f_y(x, e_y)$$

構造方程式



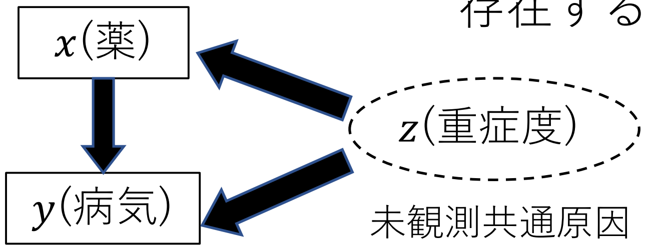
因果グラフ

片方向矢印（有向辺）

・構造方程式の左辺の値を計算するために
必要かもしれないとき

両方向矢印付き円弧（有向円弧）

・未観測原因が
存在するかもしれないとき



未観測共通原因

5．構造的因果モデル

構造的因果モデル

◎構造的因果モデル：反事実モデルで定義される因果関係を、
構造方程式モデルを用いて数学的に表現

反事実モデル
（因果のモデル）

構造方程式モデル
（データ生成過程のモデル）

構造的因果モデル
（因果推論のための枠組み）

集団レベルの因果の表現

◎ある変数 x に介入する：「ほかのどの変数がどんな値をとろうとも、変数 x の値を定数 c にとる」

→ $do(x = c)$　と表す

〈例〉薬と病気

変数x（薬を飲むかどうかを表す）に介入する

：年齢や性別、病気の重症度などにかかわらず、

xの値を1にとる（必ず薬を飲んでもらう）　or

xの値を0にとる（決して飲まないでもらう）

集団レベルの因果の表現

$$x = e_x$$
$$y = f_y(x, e_y)$$

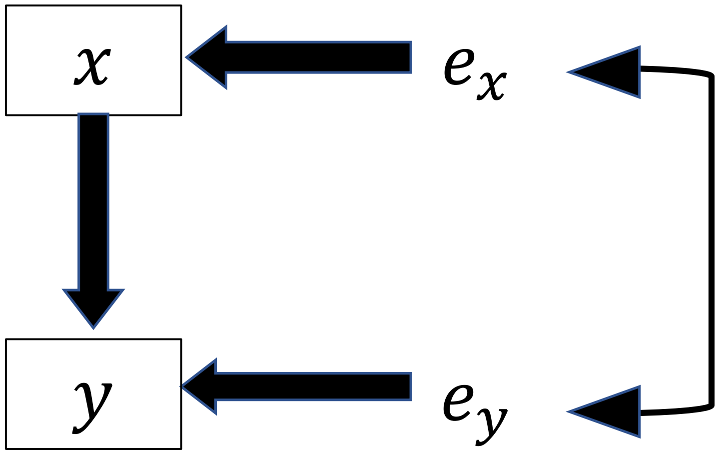
xに介入
（xの値をcに定める）

$$\boxed{\begin{aligned} x &= c \\ y &= f_y(x, e_y) \end{aligned}}$$

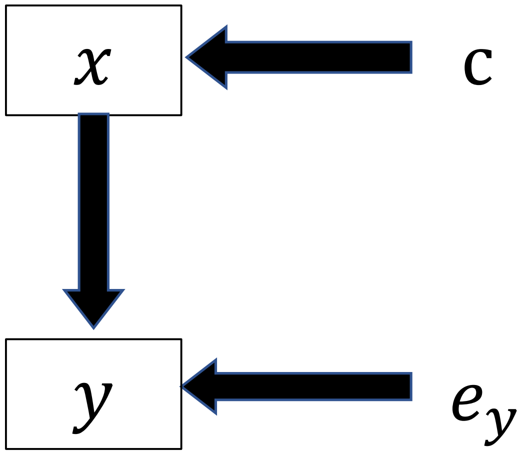
$M_{x=c}$ とおく。

・$M_{x=c}$：xに介入して、その値をcに定めた構造方程式モデル
（ただし、仮想的な集団）

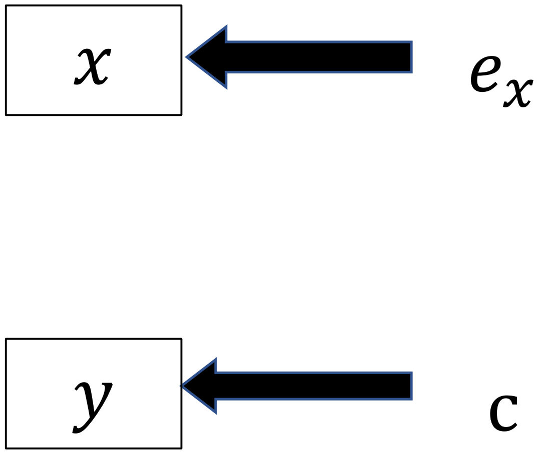
集団レベルの因果の表現



自然にお任せ



xに介入
（$M_{x=c}$の因果グラフ）



yに介入

集団レベルの因果の表現

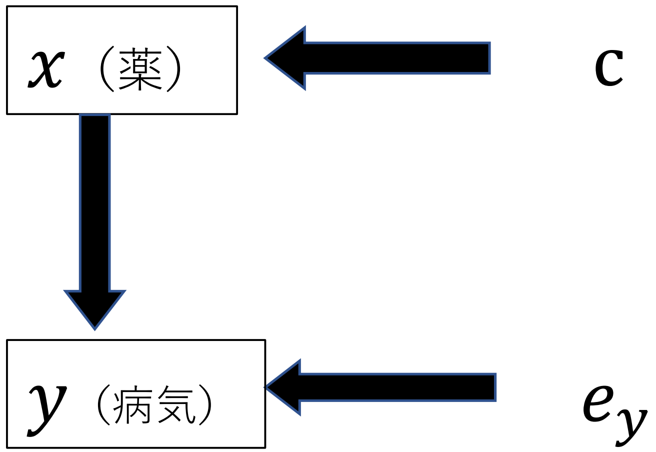
- 介入後のyの分布:= 介入後のモデル$M_{x=c}$におけるyの分布

$$p(y|do(x=c)) \coloneqq p_{M_{x=c}}(y)$$

- 介入後のモデル$M_{x=c}$

$$x = c$$
$$y = f_y(x, e_y)$$

構造方程式



集団レベルの因果の表現

・薬を飲むかどうかが病気が治るかどうかの原因となる

$$\rightarrow p(y|do(x=1)) = p(y|do(x=0))$$

→薬を飲んでもらう場合と飲まないでもらう場合とで、3日後に病気にかかっているかどうかの分布が異なれば（↑を満たせば）、「この集団において、薬を飲むかどうかが病気が治るかどうかの原因となる」という

→　≠が「>」であれば、「この集団においては、薬を飲むという行動には、病気を治す効果がある」といえる（逆なら、この薬は有害だといえる）

因果効果の大きさの定量化

・介入後の分布が異なるかを調べれば、
　因果関係にあるかどうかは分かる

→因果関係の大きさはどのくらいあるのか？

因果効果の大きさの定量化

・変数xから変数yへの因果効果の大きさを定量化
→平均的な差を評価する

$$E(y|do(x=d))-E(y|do(x=c))$$（←平均因果効果）

：$\mathrm{M}_{\mathrm{X=d}}$と$\mathrm{M}_{\mathrm{X=c}}$における$y$の期待値の差

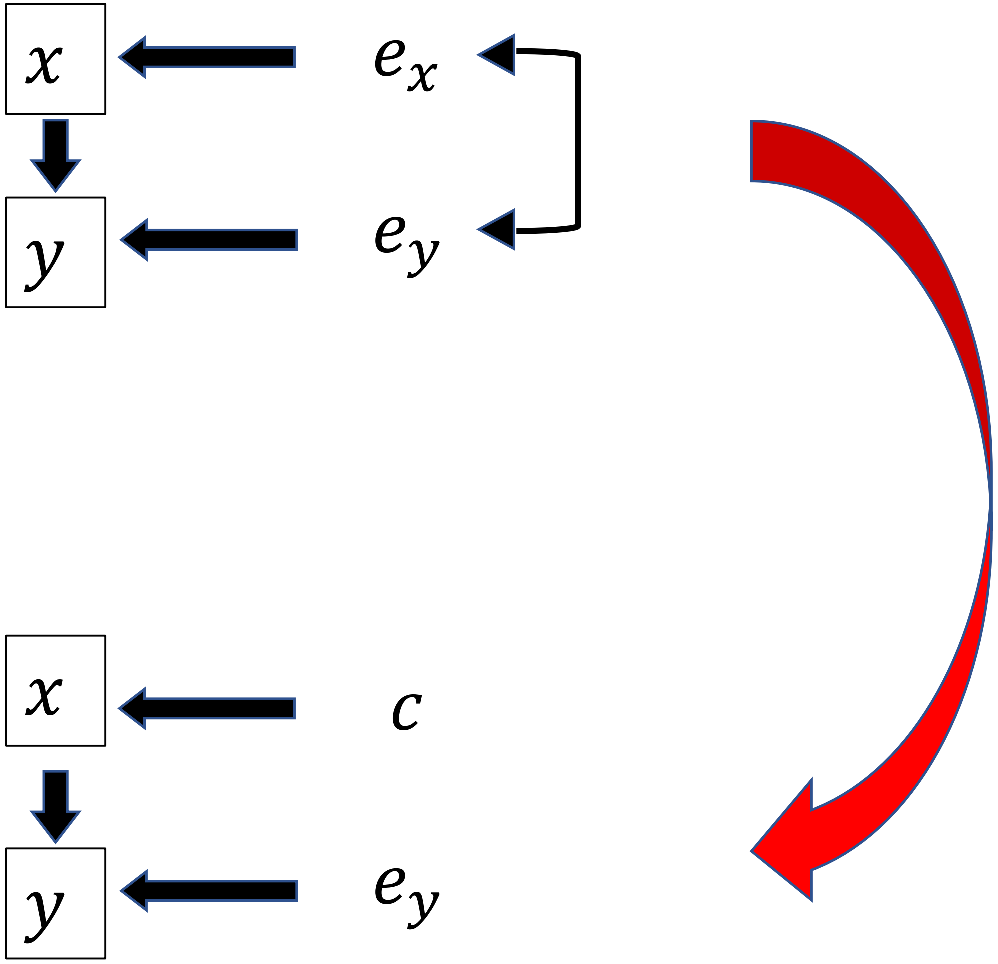
因果効果の大きさの定量化

・介入前のデータ生成過程（自然にお任せ）

$$x = e_x$$
$$y = b_{yx}x + e_y$$

・介入後のデータ生成過程$M_{x=c}$

$$x = c$$
$$y = b_{yx}x + e_y$$



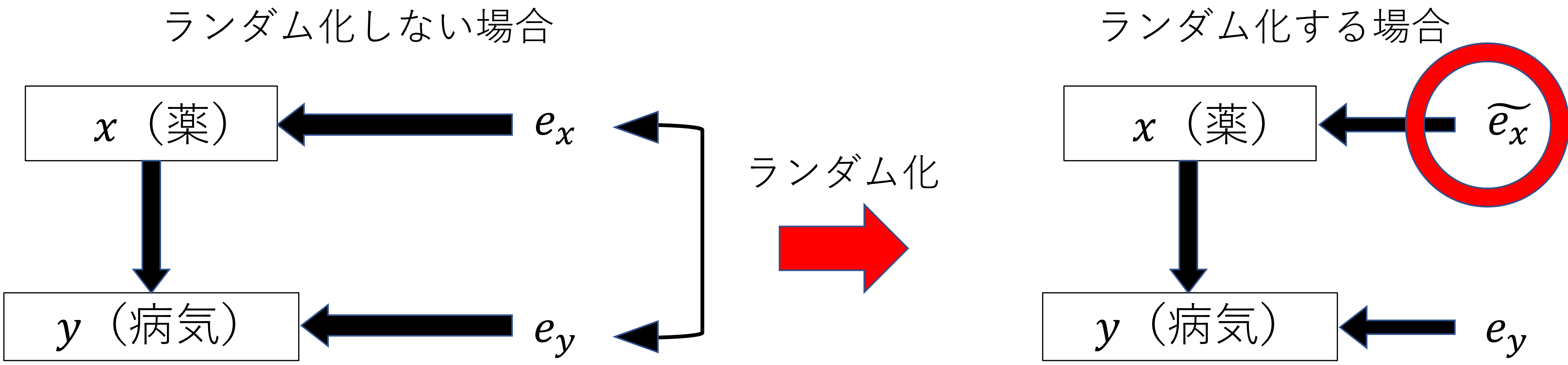
このデータ生成過程におけるyの平均が$E(y|do(x = c))$

6．ランダム化実験

ランダム化実験

◎ランダム化実験：因果関係を推測する上で、最も分析が単純になる方法

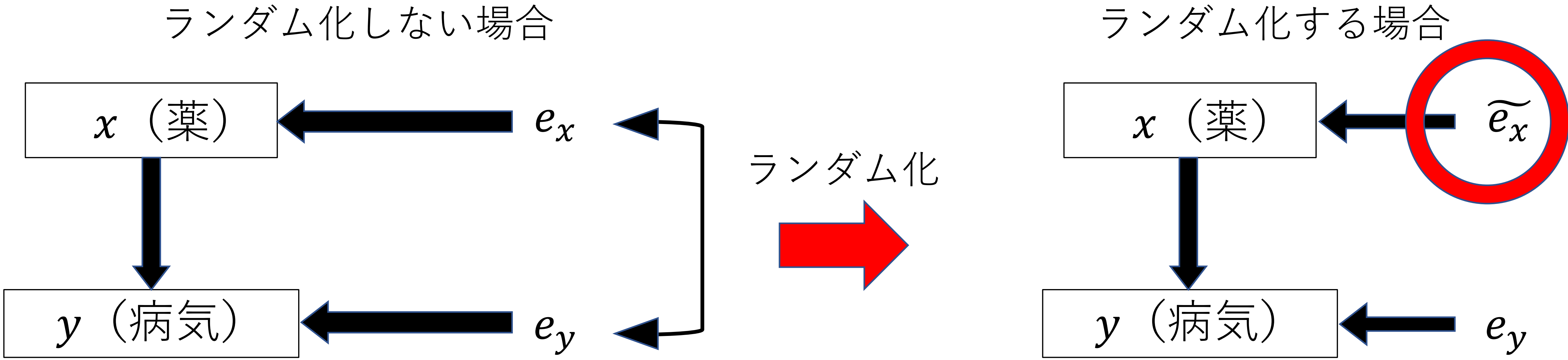
ランダム化実験



$\widetilde{e_x}$：成功確率1/2のベルヌーイ分布に従う確率変数

→薬を飲むかどうかをランダムに決めることで、誤差変数が独立になる
（未観測共通原因がない）

ランダム化実験



・仮定1（事前知識）：時間的先行性

→ありうる因果の向きが決まる（病気→薬はありえない）

・仮定2：ランダム化

→誤差変数が独立になる：未観共通原因がない

ランダム化実験

ランダム化実験
：平均因果効果$E(y|do(x=1))-E(y|do(x=0))$を

通常の条件付き期待値の差$E(y|x=1)-E(y|x=0)$によって

推定可能

→集団の個体すべてのxの値に1または0を定めるという介入をする必要が無い

ランダム化実験

ランダム化実験：一般的に、

$$p(y|do(x = c)) = p(y|x = c)$$

$$p(y|x = 1) \neq p(y|x = 0)$$

→この集団において、薬を飲むかどうかが、
病気が治るかどうかの原因となる

まとめ

- 統計的因果推論の代表的な枠組みとして「構造的因果モデル」がある
- 因果効果の大きさを測るためには、「平均因果効果」を計算する
- 平均因果効果の推定法の1つに「ランダム化実験」があり、平均因果効果をデータから推定することが可能になる