データ解析の基礎

酒井（5回）統計分析と仮説検定
鮫島（5回）ベイズ推論
田中（5回）デコーディング

解析？　分析？
analysis

なぜデータ解析が必要か？

データは確率的だから

確率とは？

未知の要因がある立場から決まっている現象を観測したときに、ある結果が再現される程度

未知の要因の可能性に十分な想像をしないと誤った理解につながる。

「統計」とは、これを避けるための手段

世の中には統計のまやかしが溢れている

この研究科で養って欲しいこと

- まやかしに騙されない目
- まやかしの背後にある構造を見抜く目
- 誤って、まやかしを出さないようにする知識
- データからどの程度のことが言えるのかを見抜く目

講義で使う統計ツール＆テキスト

Reactive Stat

https://www.emuyn.net/stats/

講義資料（酒井分）

https://brainsci.net/

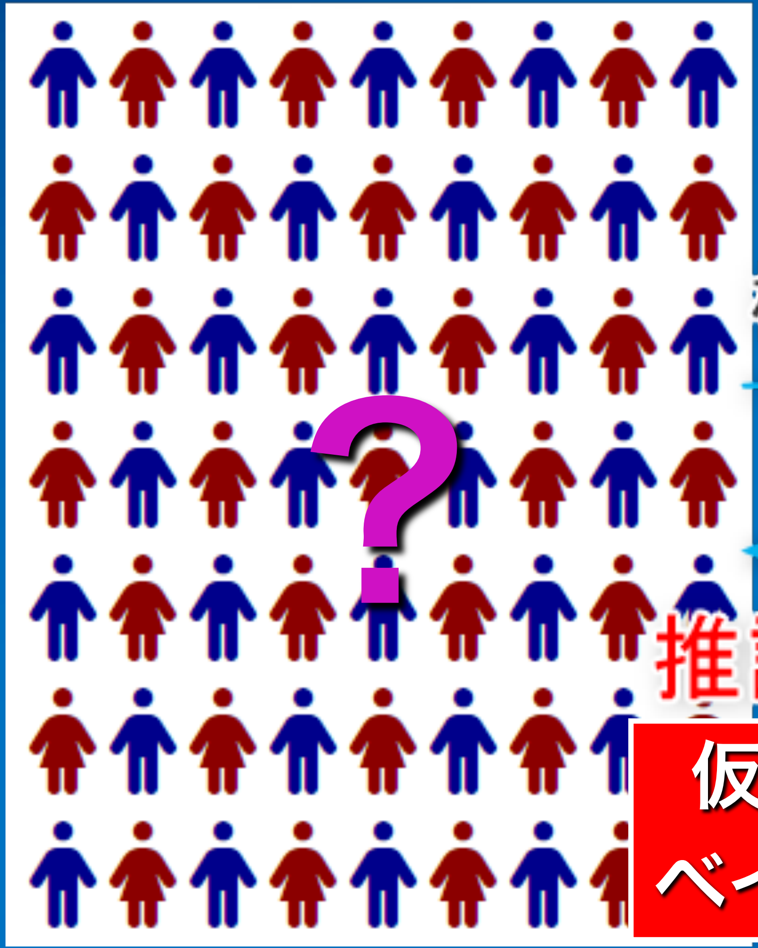
統計の種類　〜　記述統計・推論統計　〜

母集団

サンプル（観測データ）

未知の要因で起こりうる

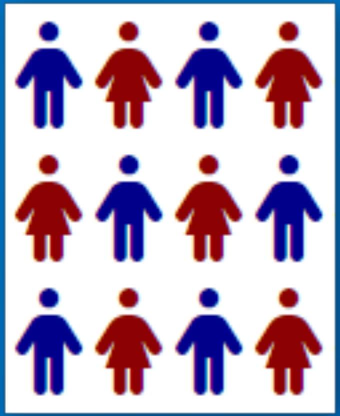
観測の全体（仮想）



標本抽出

推論統計

仮説検定
ベイズ推論



記述統計

統計量
（平均,分散,…）
可視化

信頼区間・標準誤差
デコーディング

(図出典: https://www.emuyn.net/)

仮説検定と信頼区間

仮説検定

棄却したい仮説（null hypothesis)を設定

↓

仮説にもとづいたら観測されるはずのデータの統計量の確率分布$\mathcal{P}$を算出

↓

$\mathcal{P}$のもとで、実測の統計量より外側になる確率P値を算出

↓

予め決めておいた有意水準αのもとで、

P値$<\alpha$ → 仮説を棄却

P値$\geq\alpha$ → 仮説を棄却しない

例：確率モデル：正規分布

A群とB群の平均は等しい

統計量：AとBの平均の差

よく使われるα：

α=0.05 *

α=0.01 **

仮説検定

設定した仮説の元でサンプリングした際、
　　ある統計量の平均操作に関して

$\boxed{\text{P値}}$ ＝真の平均との差が $\boxed{\text{実測の平均との差}}$ より大きい確率

信頼区間

$\boxed{\alpha}$ ＝真の平均との差が $\boxed{\text{信頼区間}}$ より大きい確率

サンプリングして求めた信頼区間の中に
真の平均が入るようなサンプリングの割合が1-α

「信頼区間の中に真の平均が入る確率」ではない　→　ベイズ推論

よく使う信頼区間

t 検定の場合：

<u>標準誤差</u>（SE）

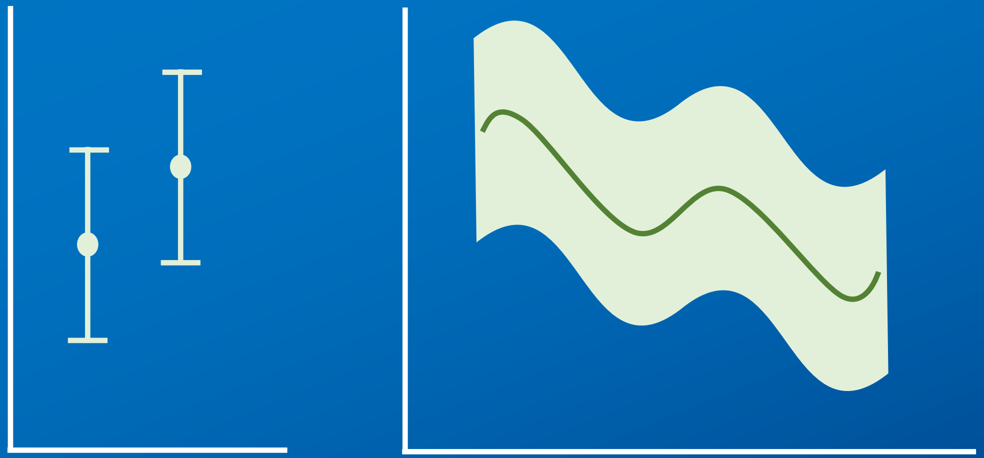
$$\text{信頼区間：}\ [\text{平均}] \pm t_{\alpha/2}(n-1)\frac{[\text{標準偏差}]}{\sqrt{n}}$$

nが十分大きいとき

±SE ≒ 68%信頼区間

±2xSE ≒ 95%信頼区間

±3xSE ≒ 99%信頼区間



基本的には可視化のツール（→記述統計？）

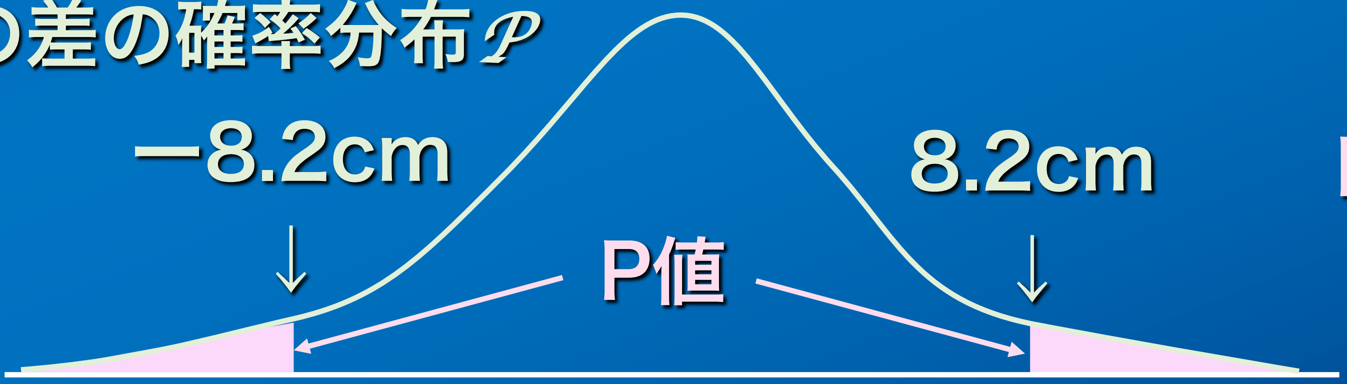
仮説検定の例

仮説：６歳の女子と男子の身長は平均が等しい正規分布
有意水準：α=0.01

ある小学校で６歳の女子と男子の身長データを取得

女子の身長データ ➡ 女子の平均身長 (123.1cm)	➡	平均の差 (8.2cm)
男子の身長データ ➡ 男子の平均身長 (114.9cm)		

仮説にもとづいたら出るはずの
平均の差の確率分布$\mathcal{P}$



女子平均ー男子平均 [cm]

P=0.0093<α **
仮説を棄却

信頼区間の例

仮説：６歳の女子と男子の身長は平均が等しい正規分布
有意水準：α=0.01

ある小学校で６歳の女子と男子の身長データを取得

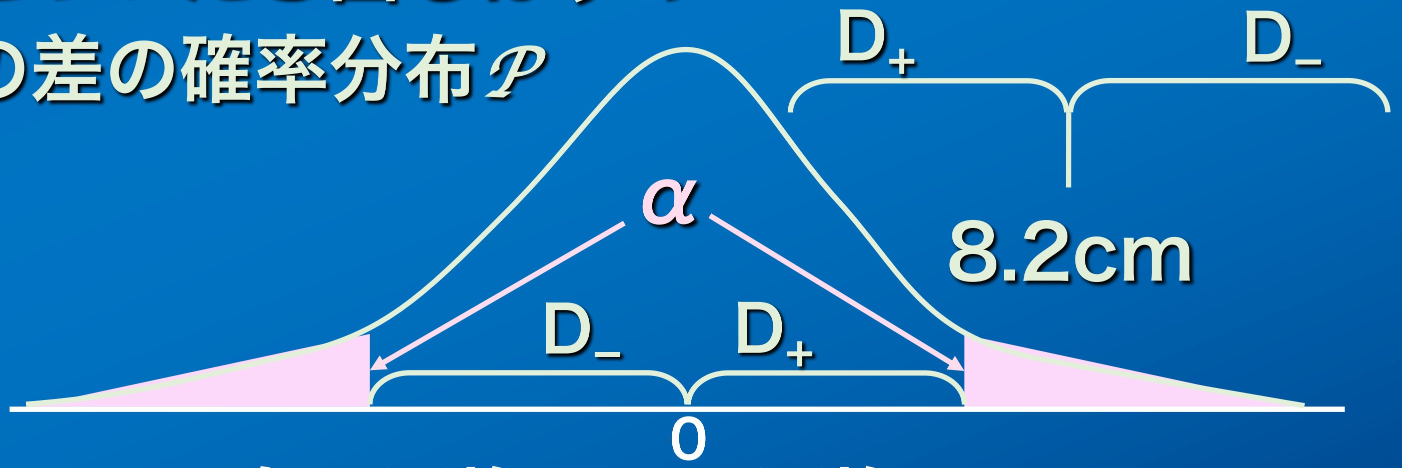
女子の身長データ ➡ 女子の平均身長 (123.1cm)

男子の身長データ ➡ 男子の平均身長 (114.9cm)

➡ 平均の差 (8.2cm)

仮説にもとづいたら出るはずの
平均の差の確率分布$\mathcal{P}$

1−αの信頼区間



女子平均ー男子平均 [cm]

「６歳の女子は男子より身長が高い」？

・正規分布じゃなかったのかも
・たまたま偏った集団を計測しただけかも
・たまたま4・５月生まれの女子が多かったせいかも

正しい帰結

ある小学校の６歳の身長を調べた結果

$\boxed{\text{６歳の女子と男子の身長は平均が等しい正規分布}}$ という仮説が

$\boxed{\text{有意水準：}\alpha=0.01}$ で棄却された。

標準的な有意水準を共有している分野では

「小学校の６歳の女子は男子より身長が有意に高い（～検定: P=…）」

という表現は許される

- 「有意に」というのは棄却の程度の範囲を意味しているに過ぎない
- ～検定が何を仮定しているのかを知っている人に対するメッセージ
- 表面的な帰結だけ伝わってしまう弊害

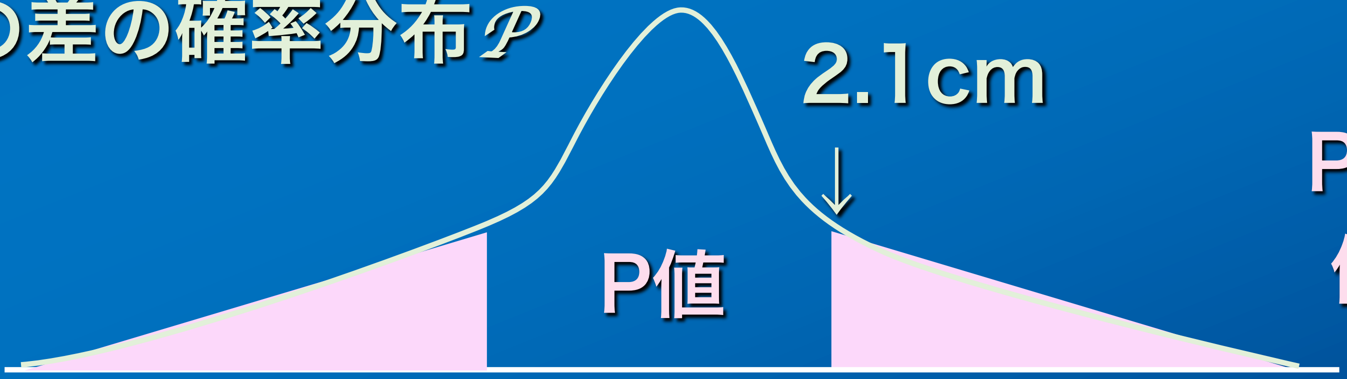
仮説検定の例

仮説：６歳の女子と男子の身長は平均が等しい正規分布
有意水準：α=0.05

ある小学校で６歳の女子と男子の身長データを取得

女子の身長データ → 女子の平均身長 (120.3cm)
男子の身長データ → 男子の平均身長 (118.2cm)
→ 平均の差 (2.1cm)

仮説にもとづいたら出るはずの
平均の差の確率分布$\mathcal{P}$



女子平均ー男子平均 [cm]

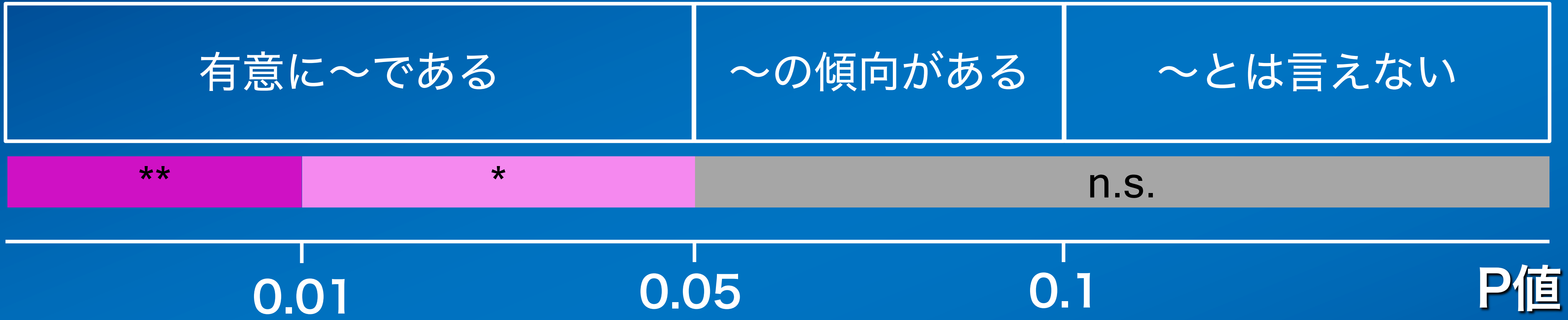
P=0.26>α (n.s.)
仮説を棄却しない

「６歳の女子と男子の身長に違いはない」？

違うとは言えなかっただけ

例えば、脳科学・神経科学では

許される表現

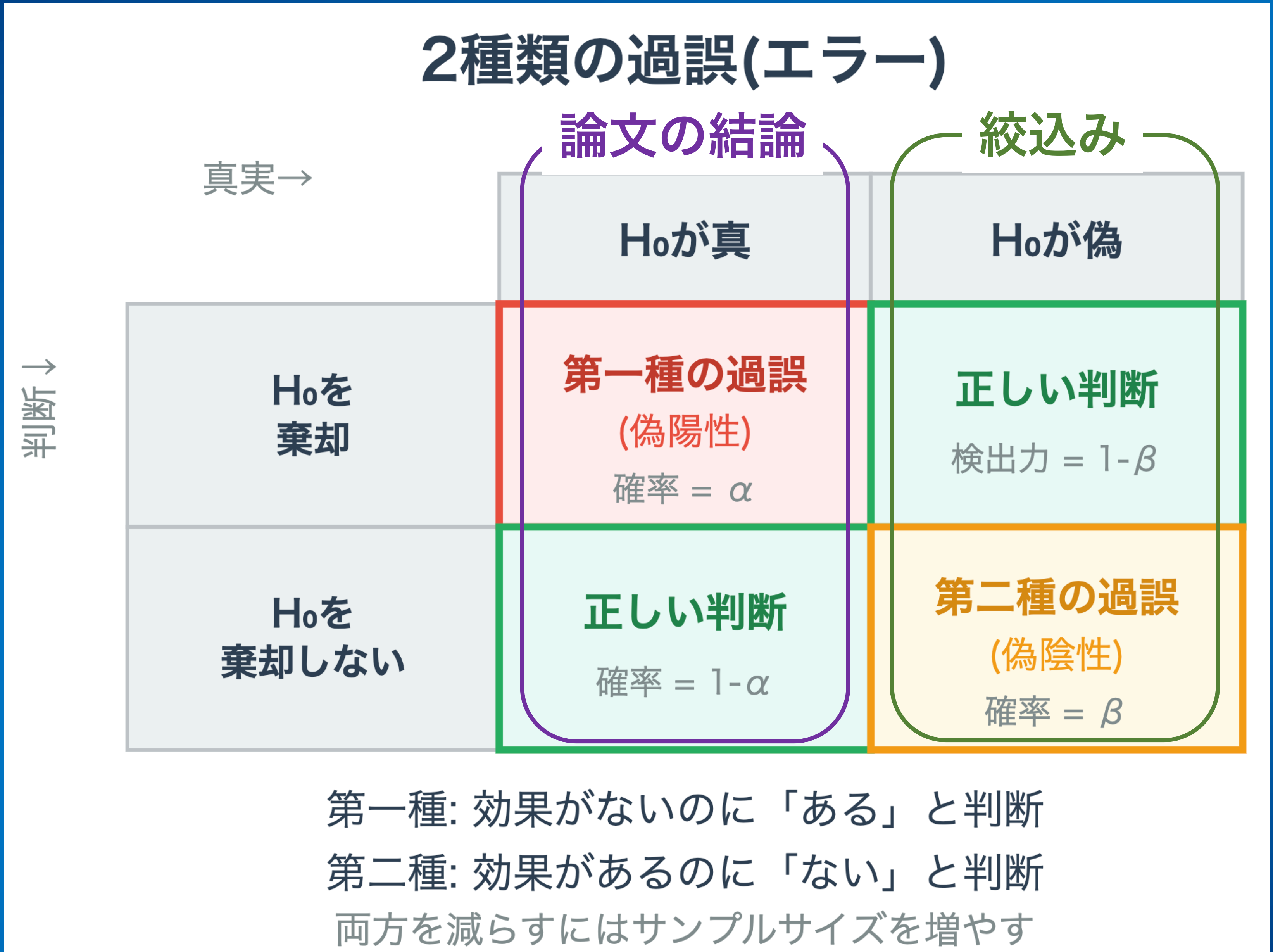


おかしな慣習であるが、まともな研究をスクリーニングするのに役立ってきた。

一般に対する（過度な）アウトリーチに役立ってきた。

長年、この慣習が続いたことによる弊害はでてきている。

仮説検定における２種類の過誤



H_0：仮説

(図出典: https://www.emuyn.net/)

「仮説検定とは科学的結論を導くツール」？

ある意味正しいが、
導ける結論はかなり限定される

導ける結論

〜の確率分布について〜を満たすとしたら、〜であるという仮説が
有意水準〜で棄却された

アメリカ統計協会のP値と統計的有意性に関する声明（2016)

https://doi.org/10.1080/00031305.2016.1154108

P値に関する6つの基本原則

1. P値はモデルとデータの矛盾の程度を示す一指標である。
2. P値は仮説の正しさや偶然性の確率を示さない。
3. P値だけに基づいて科学的結論や決定を下してはならない。
4. 透明性が必要であり、すべてのデータと解析を報告すべきである。
5. P値は効果の大きさや結果の重要性を意味しない。
6. P値だけでは仮説やモデルに関する強い証拠にはならない。

「仮説が正しいことを示すことができる」？

肯定は原理的に不可能！
できるのは否定だけ

（科学全般の話）

「仮説検定で〜が誤りだと示すことができる」？

できない！

示すことができるのは誤りの可能性の程度だけ

アメリカ統計協会のP値と統計的有意性に関する声明（2016)

https://doi.org/10.1080/00031305.2016.1154108

P値に関する6つの基本原則

1. P値はモデルとデータの矛盾の程度を示す一指標である。
2. P値は仮説の正しさや偶然性の確率を示さない。
3. P値だけに基づいて科学的結論や決定を下してはならない。
4. 透明性が必要であり、すべてのデータと解析を報告すべきである。
5. P値は効果の大きさや結果の重要性を意味しない。
6. P値だけでは仮説やモデルに関する強い証拠にはならない。

サンプル数とP値　～大数の法則～

サンプル数を増やすとP値は小さくなりやすい

背景　～　大数の法則
数を増やせば平均値の精度が良くなる

サンプル数とP値　～大数の法則～

確率変数 X, Y

定数 a

$$E[X+Y] = E[X] + E[Y], \quad E[aX]=a\,E[X]$$

$$E[XY] = E[\,E[X|Y]\,Y\,] = E[\,E[Y|X]\,X\,]$$

$$Var[X] = E[\,(X-E[X])^2\,] = E[\,X^2 - 2X\,E[X] + E[X]^2\,]$$

$$= E[X^2] - 2E[X]E[X] + E[X]^2 = E[X^2] - E[X]^2$$

$$Var[aX] = E[a^2X^2] - E[aX]^2 = a^2\,E[X^2] - a^2\,E[X]^2 = a^2\,Var[X]$$

$$Var[X+Y] = E[(X+Y)^2] - E[X+Y]^2$$

$$= E[X^2 + Y^2 + 2XY] - (E[X]^2 + 2\,E[X]E[Y] + E[Y]^2)$$

$$= E[X^2] - E[X]^2 + E[Y^2] - E[Y]^2 + 2E[XY] - 2E[X]E[Y]$$

$$= Var[X] + Var[Y] + 2(E[XY] - E[X]E[Y])$$

$$= Var[X] + Var[Y] + 2\,Cov[X,Y]$$

XとYが独立なら

$$E[X|Y] = E[X],\ E[XY] = E[X]E[Y],\ \ Var[X+Y] = Var[X] + Var[Y]$$

サンプル数とP値　～大数の法則～

サンプル数を増やすとP値は小さくなりやすい

背景　～　大数の法則

数を増やせば平均値の精度が良くなる

同じ確率分布から独立にサンプルした確率変数 $X_1, X_2, \cdots, X_n$ の平均値の分散

$$\mathrm{Var}[\,(X_1+X_2+\cdots+X_n)/n\,] = (\,\mathrm{Var}[X_1]+\mathrm{Var}[X_2]+\cdots+\mathrm{Var}[X_n]\,)/n^2$$

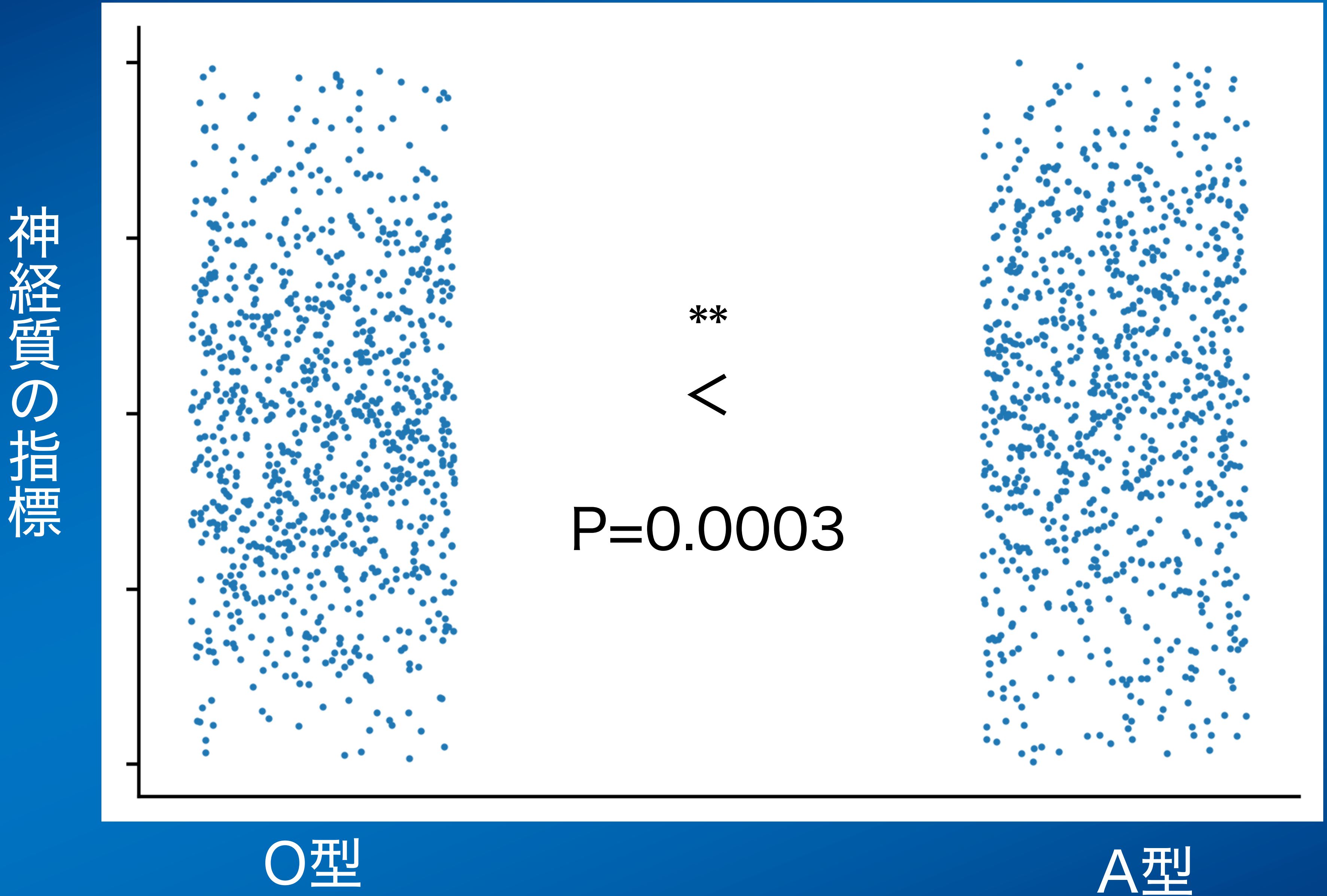
$$= n\,\mathrm{Var}[X]\,/n^2 = \mathrm{Var}[X]\,/n$$

n を増やせば平均値の分散は小さくなる

真の平均に僅かでも差があれば、平均値が逆転する確率は

n を増やすと、どんどん小さくなる　～　P値は0に近づく

サンプル数とP値



サンプル数とP値

サンプル数を増やすとP値は小さくなりやすい

僅かな差でも有意になる

⋮

適切なサンプル数

⋮

サンプル数が少ないとP値は大きくなりやすい

大きな差でも有意にならない

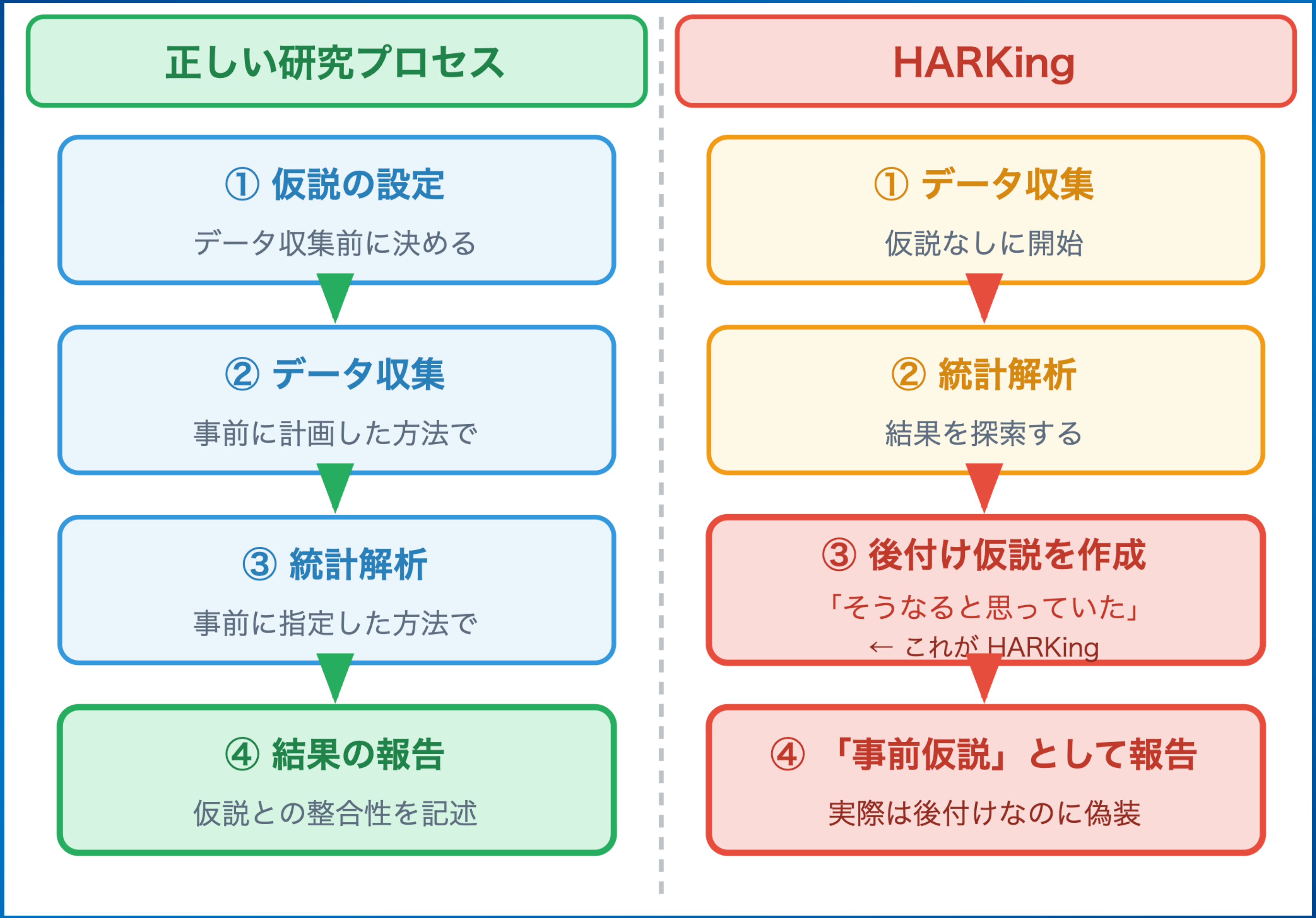
P値ハッキング

- ある仮説Aを示すために、ヒトを対象にした行動実験を始めた。
- 予備的に10人の被験者を集めて実験をし、データを分析してみたら、仮説Aの傾向が確認できたが、有意にはならなかった。
- その後、データ分析をしながら、実験を繰り返した。
- 被験者が30人になったところで、有意差が確認できたので、そこで実験を終了し、結果をまとめた。

HARKing (Hypothesizing After the Results are Known)

- ある仮説Aを示すために、ヒトを対象にした行動実験を始めた。
- 念の為、仮説Aとは直接関係ない生理指標や行動指標を計測しておいた。
- 予定の被験者数の実験を終え、データ分析をしたら、仮説Aは有意とはならなかった。
- 念の為計測しておいた指標についても分析した結果、ある生理指標に関して、仮説Bを有意に支持する結果がでた。
- そこで、仮説Bを中心にして論文をまとめた。

HARKing (Hypothesizing After the Results are Known)



(https://www.emuyn.net/)

公表されている論文の研究が
どう進められたかは、絶対わからない

P値ハッキングやHARKingの可能性を前提に
どの結果が確からしいのかを見抜く

自分はどう研究を進めるべきか？

(https://www.emuyn.net/)

- 左の正しいプロセスにしたがって、仮説と実験計画を立て、実験した。
- 予定のデータ分析をしたら、仮説は有意とはならなかった。

どうする？

確証的研究・探索的研究

確証的研究	探索的研究
検証すべき仮説が予めあり、検証する研究（仮説検証型）	新たな仮説を探索し、発見する研究（仮説生成型）

仮説検定はこのため

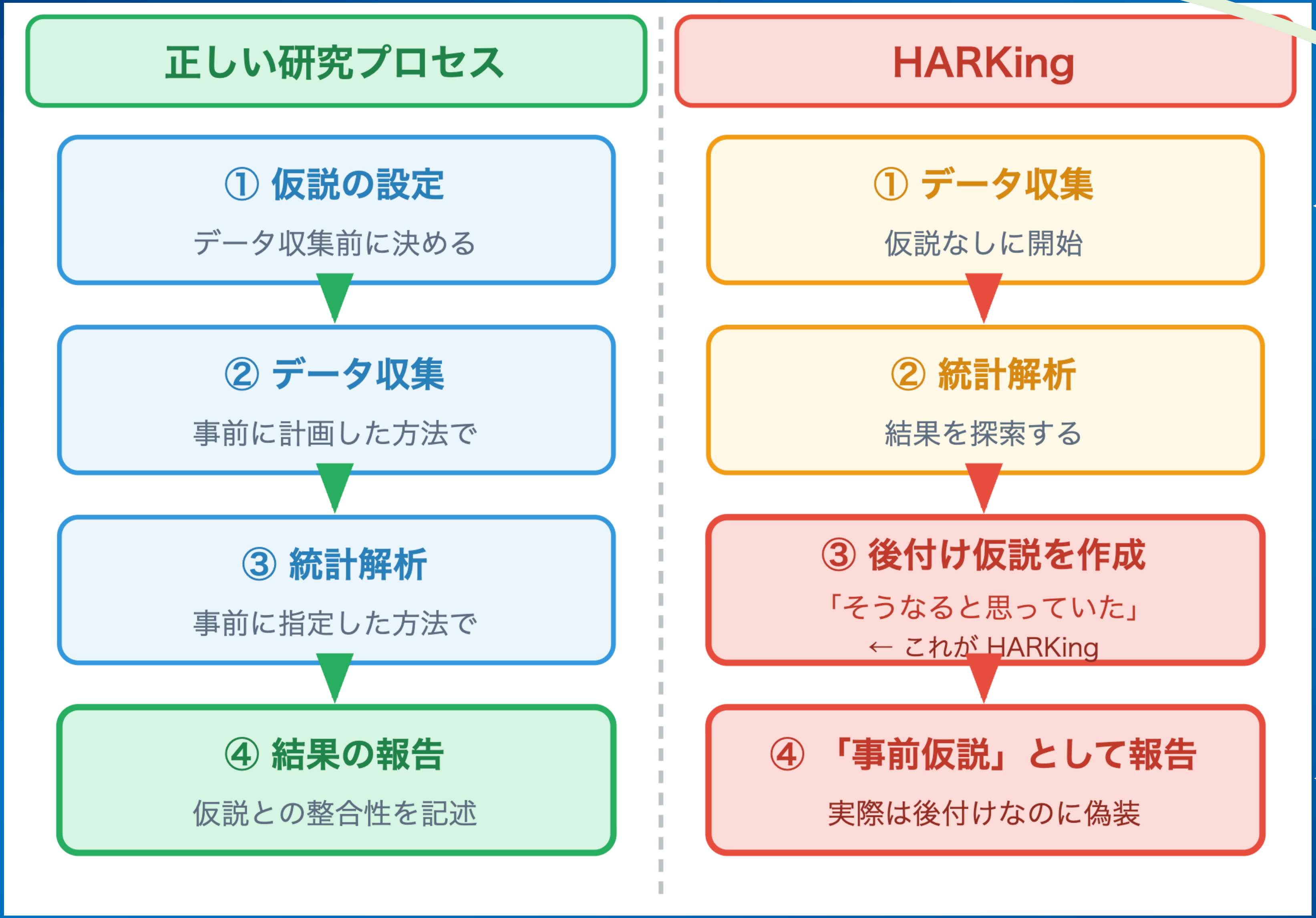
本来、仮説検定ではないが検定手法をスクリーニングに使える

素粒子物理学実験
大規模な社会科学実験
etc.

脳科学・神経科学

分野全体の研究をスクリーニングするために仮説検定の体裁を取らせている

脳科学・神経科学で行うべき研究プロセス



(https://www.emuyn.net/)

理想的

◀ ここから
開始していい

何度も繰り
返していい

仮説N

◀ 悪いのは
ここだけ

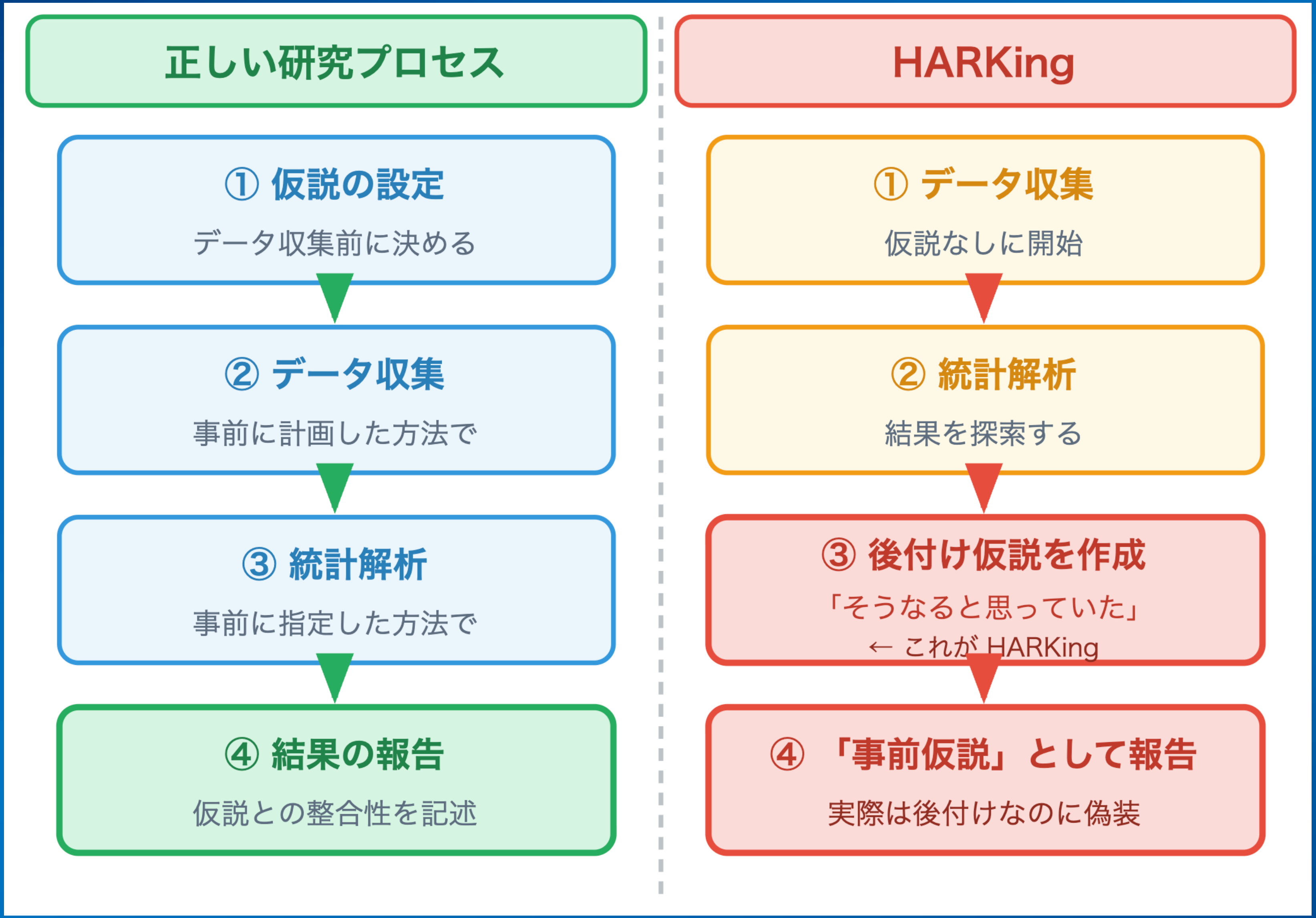
新たな仮説検証実験が難しい場合

- 得られているデータをどう分割しても同じ結果が出るかどうか確かめる（微妙に不十分であるが確証は上がる）
- 仮説探索の繰り返しをするのは、得られたデータの半分だけ使い、尤もらしい仮説を見つけたら、まだ解析していない半分で検証する

脳科学・神経科学分野で論文を通すためにHARKingはせざるを得ない。

主張した結論が嘘でないことを、可能な限り確かめる姿勢が大事。

脳科学・神経科学で行うべき研究プロセス



(https://www.emuyn.net/)

何度も繰り返していい

仮説N

脳科学・神経科学で行うべき研究プロセス

・生データをよく見てください。

・できるだけ平均処理をしていない生データに近いところから
　可視化してください。（散布図、時系列のオーバーレイ、etc.）

・想定していなかった要因の様々な可能性を想像して
　それを確かめる可視化をしてください。

・こうした観察と仮説検定の結果の対応関係を肌で感じてください。

他人の研究の結果を見ただけで、
どの程度、確からしいのかを見抜く力が養われます。