TOHOKU UNIVERSITY

第4回 運動と健康：分野横断型勉会
2018年9月6日（木）
福井県県民ホールリハーサル室

研究における統計解析の役割

～ 「当たり前」を少しだけ丁寧に深く考えてみる ～

東北大学大学院
医学系研究科　運動学分野
門間 陽樹
h-momma@med.tohoku.ac.jp

本日の決意・目標・お断り

決意

- 普段何気なく行っている統計解析を **再考** してもらうきっかけをつくる。
- 可能なかぎり、**直感的・感覚的** な理解を目指す。

目標

- 普段何気なくやっている統計解析の意味を今より明確に **理解** してもらう。
- 統計の **守備範囲** を明確に示す。
- 理解できると、「**研究ってやっぱおもしろいな**」と思ってもらう。

お断り

- 統計の専門家ではございません。疫学研究（特に観察研究）をしています。
- 疫学研究を通して獲得した（個人的な）見解や理解の仕方を共有します。
- 疫学に偏った内容かもしれませんので、その場合は差し引いて解釈願います。

本日の内容

- 研究における統計の役割
- 論文における統計の記載
- 疫学マインドで眺める統計学

TOHOKU UNIVERSITY

研究における統計の役割

～ 基本的な話が中心です ～

まずはこの質問をきっかけに。

なぜ我々は研究で統計を使用しているのか？

統計ができること | 統計のそもそも

- 遠くのものを見たい。

望遠鏡

- 小さいものを見たい。

顕微鏡

- 複数あるものをまとめて見たい。

（そもそもの）

統計

（班田収授法や太閤検地）

記述統計 | データの記述、要約、圧縮

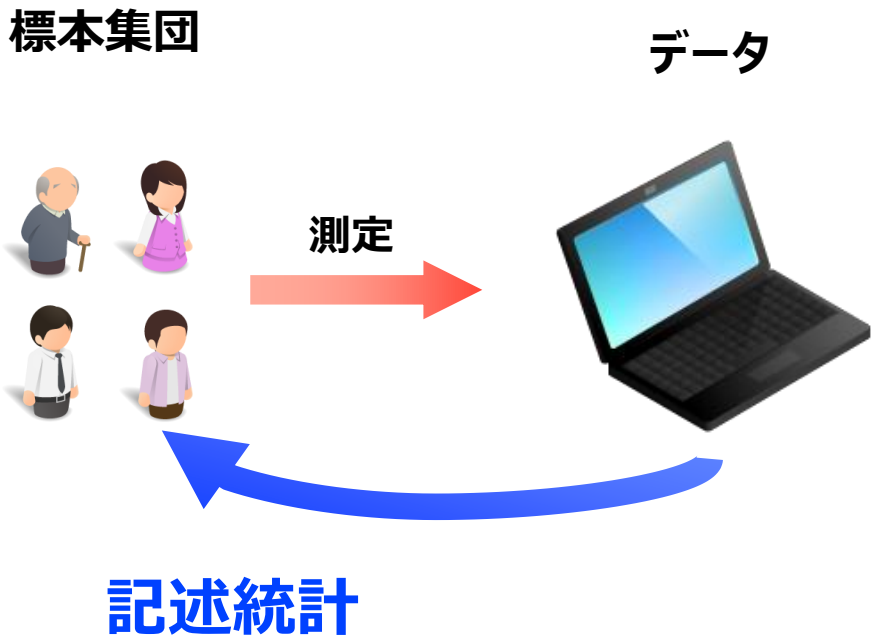
目的

- **データを与えてくれた集団** の **様子** を **簡単に** 表現すること。
  - ⇨ データを与えてくれた集団：**標本**
  - ⇨ 様子：**中心はどこか？　どのくらいバラついているか？ どのくらいいるか？**
  - ⇨ 簡単に：（可能なかぎり）**1つの数字** で

代表的な統計量（通称：記述統計量、要約統計量）

- **中心**
  - ⇨ **平均**、**中央値** など
- **バラつき**
  - ⇨ 分散、**標準偏差**、パーセンタイル、**四分位範囲**、歪度、尖度 など
- **頻度**
  - ⇨ 度数、**割合**

記述統計 | 絵にしてみると…



記述統計 | 全数（悉皆）調査の場合

ターゲット集団
（本当に知りたい人たち）

測定

データ

記述統計

ターゲット集団の人をすべて調査すれば、その様子がわかるが…

標本だけしかないと…

ターゲット集団
（本当に知りたい人たち）

標本集団

測定

データ

記述統計

標本の記述統計では、ターゲット集団の様子に近づけない。

ターゲット集団に近づく解決策 | 推測

ターゲット集団
（本当に知りたい人たち）

標本集団

データ

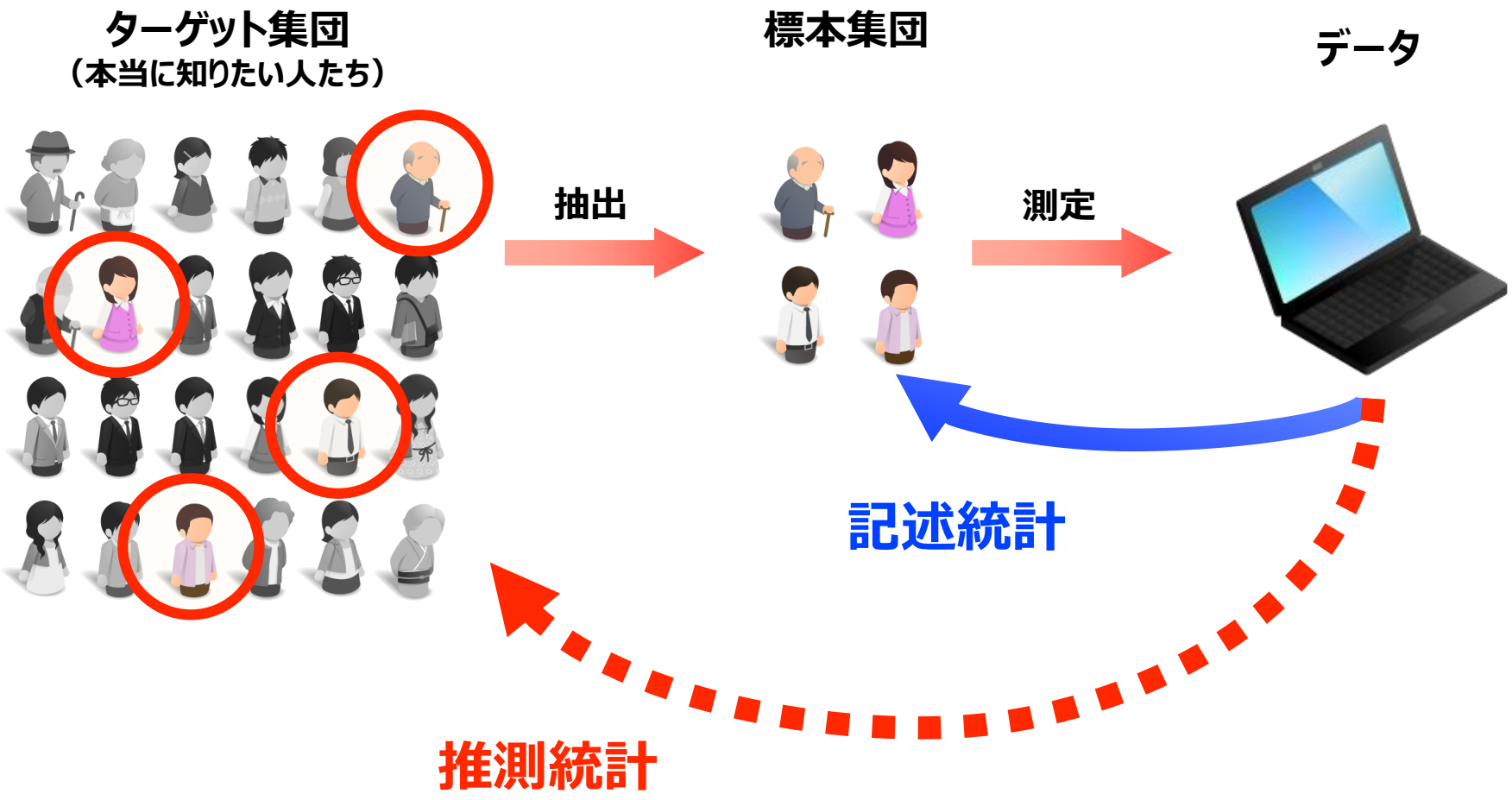
測定

記述統計

推測統計

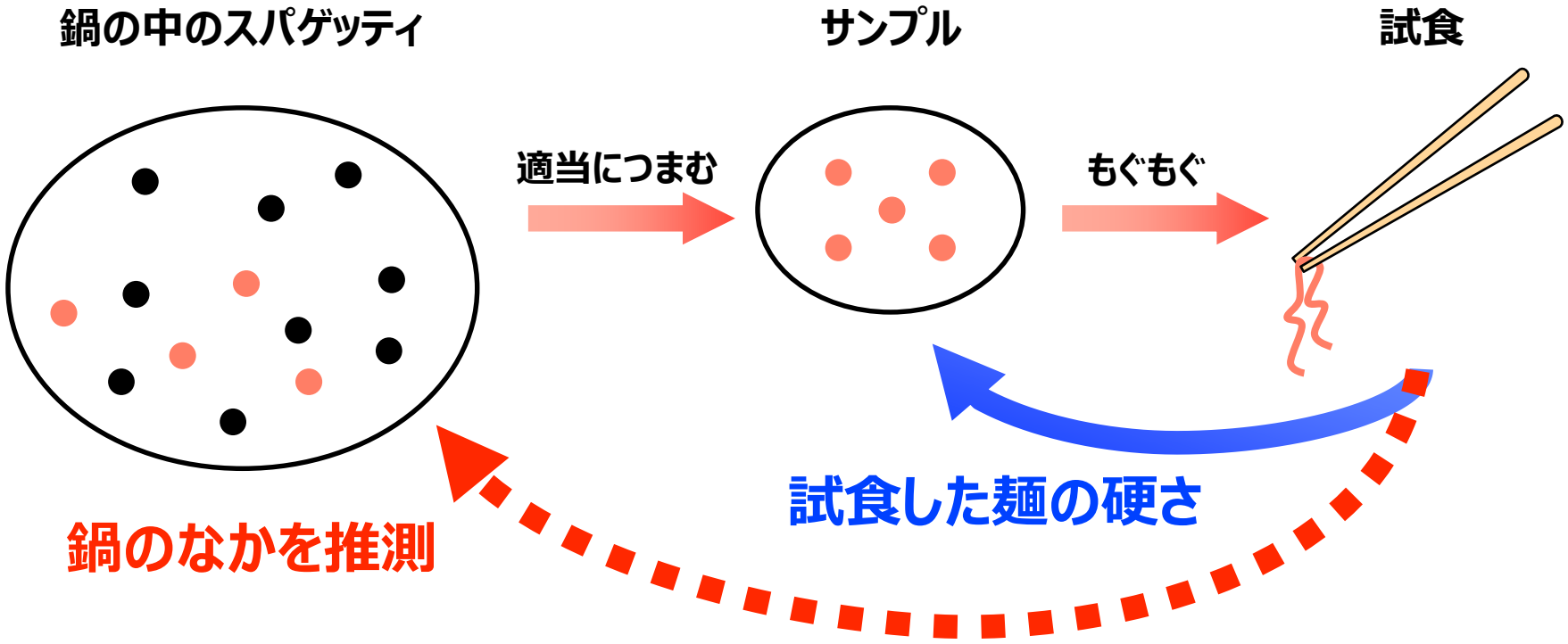
比江島欣愼. 全部絵で見る医療統計. 羊土社. 2017 を参考に作成

推測 | ランダムサンプリングを仮定



比江島欣愼. 全部絵で見る医療統計. 羊土社. 2017 を参考に作成

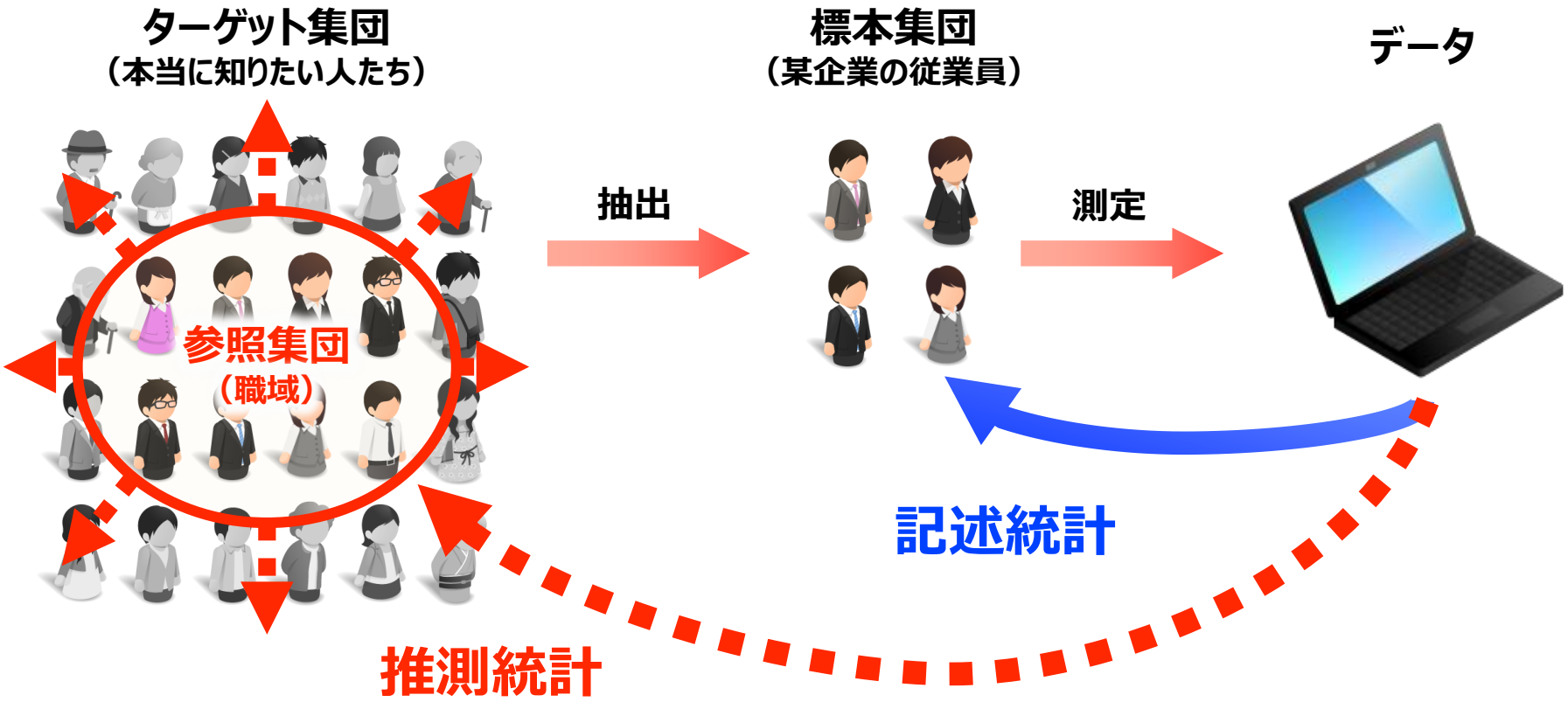
日常における推測｜スパゲッティのゆで加減



全数調査をしちゃー、おしめぇーよ。

週末の我が家を参考に作成

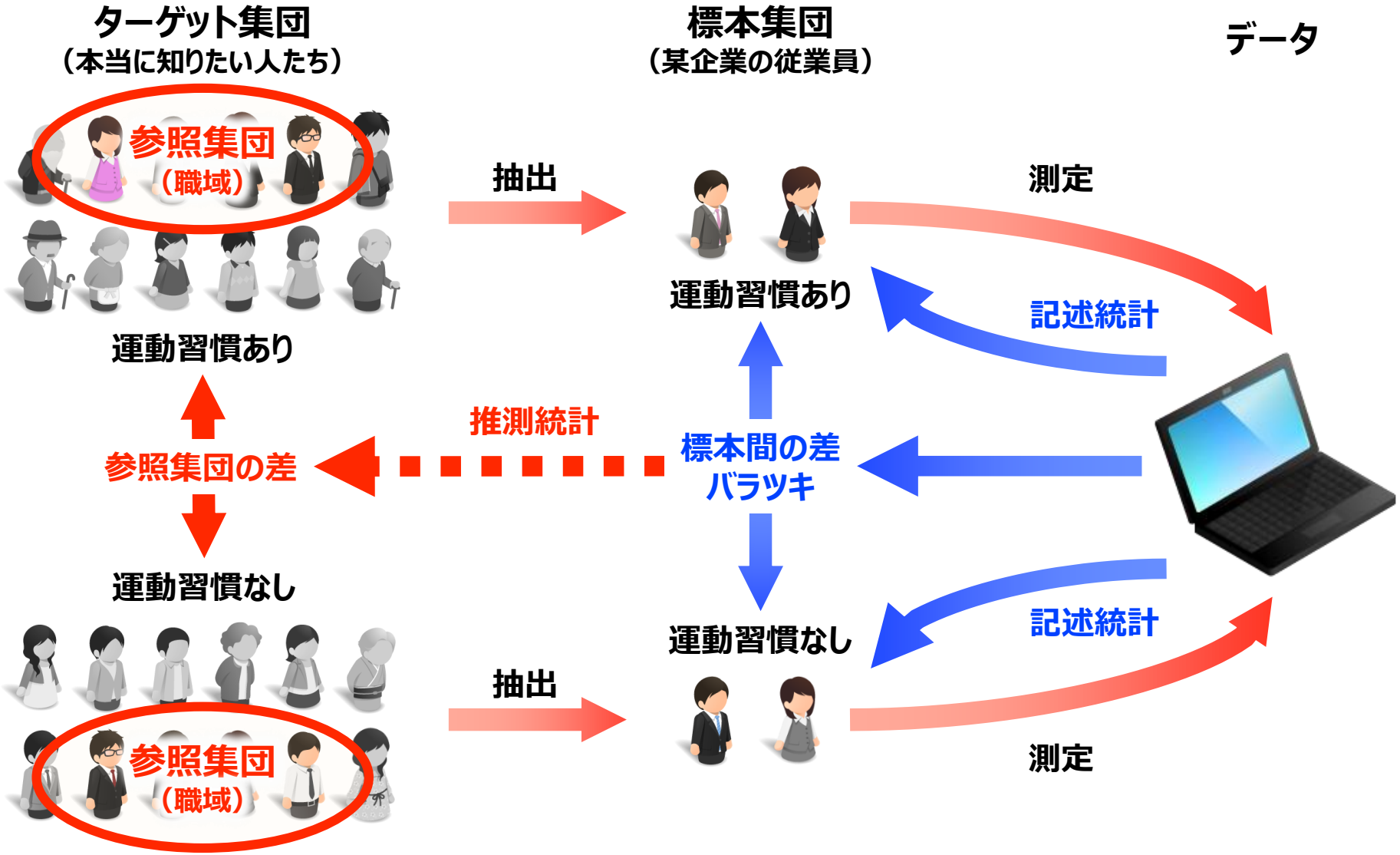
推測 | ランダムサンプリングを伴わない場合



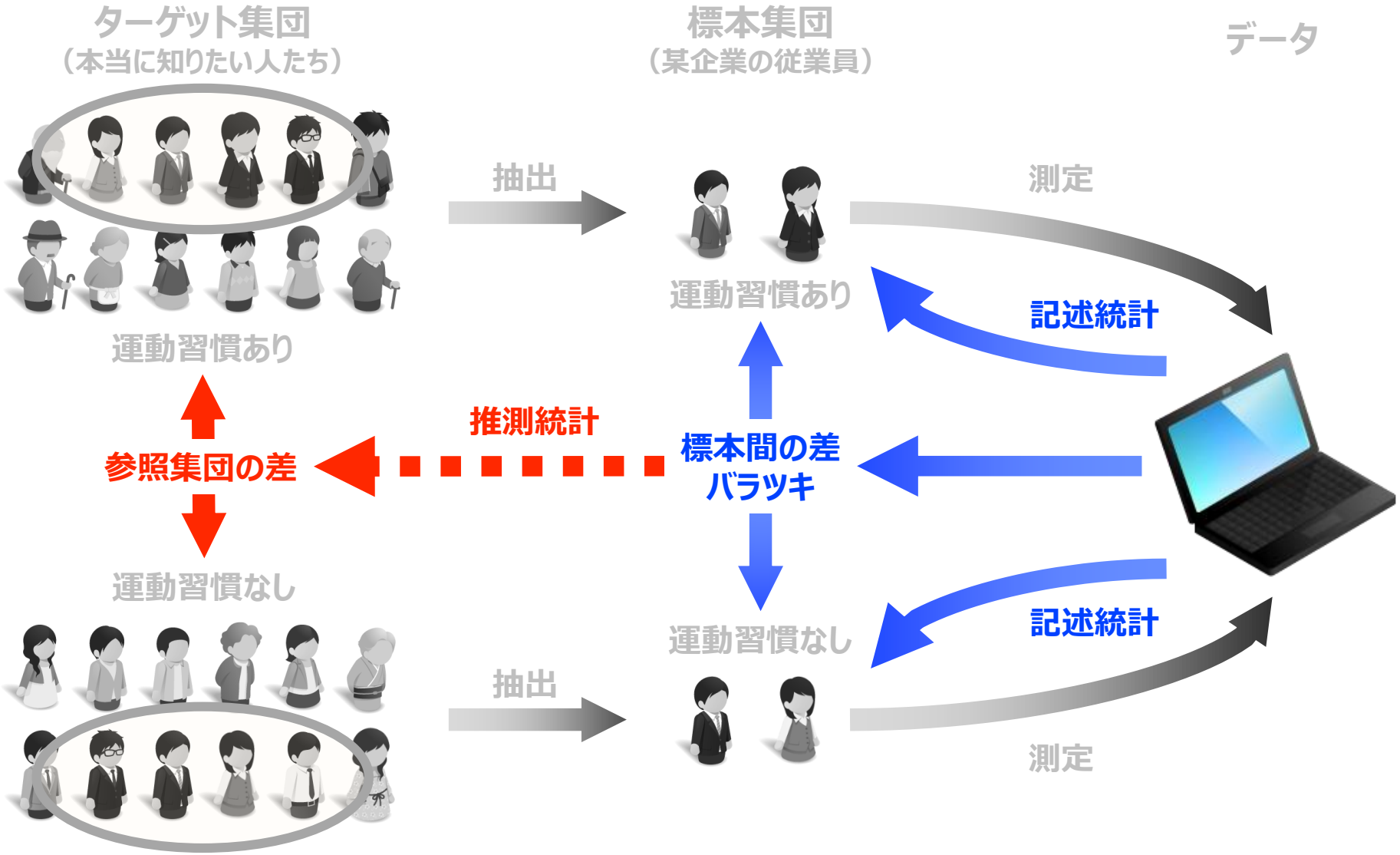
参照集団とターゲット集団の ズレ に注意
（一般化可能性の問題）

比江島欣愼. 全部絵で見る医療統計. 羊土社. 2017 を参考に作成

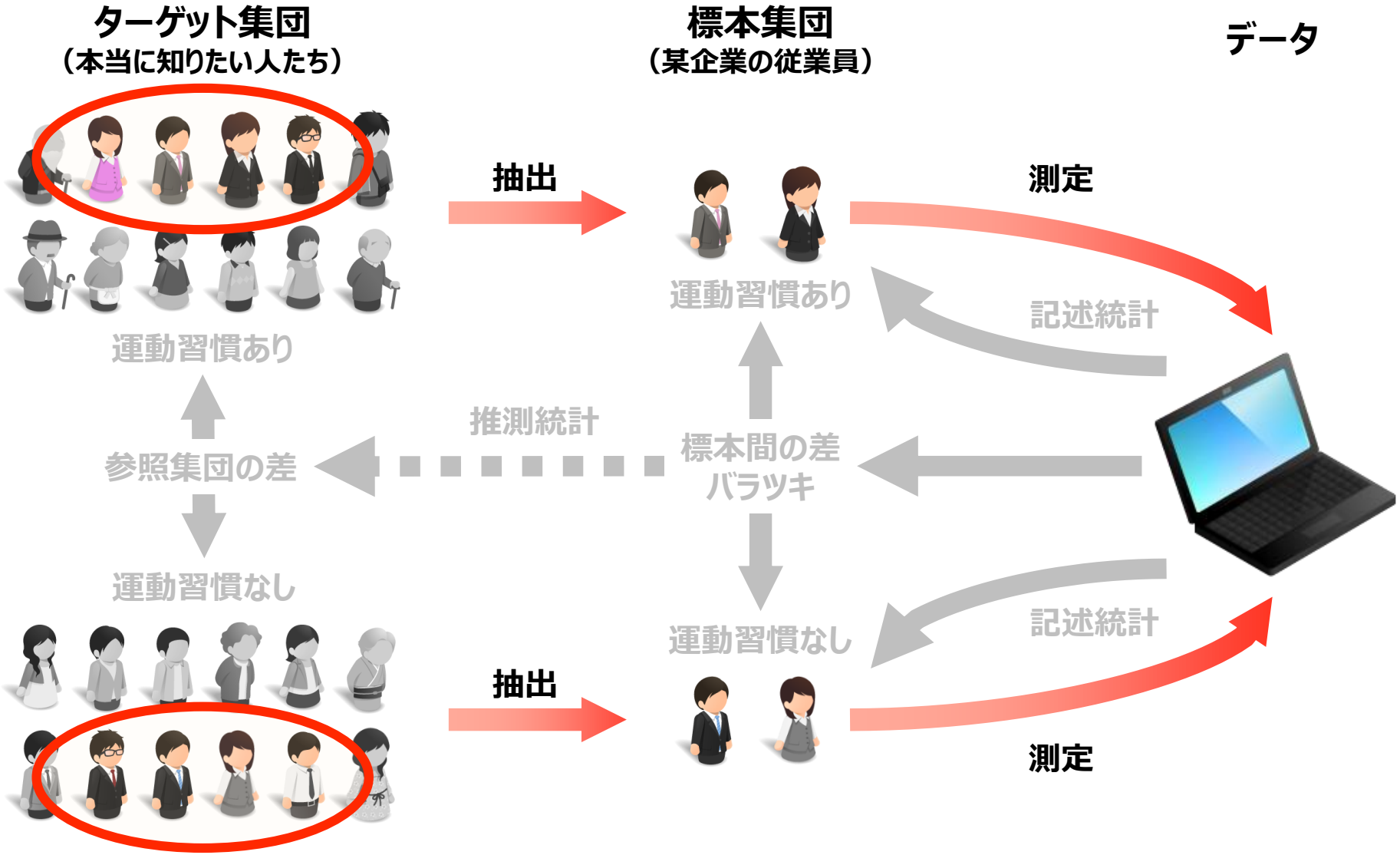
より実際の研究に沿うと… | 群間の差の検討



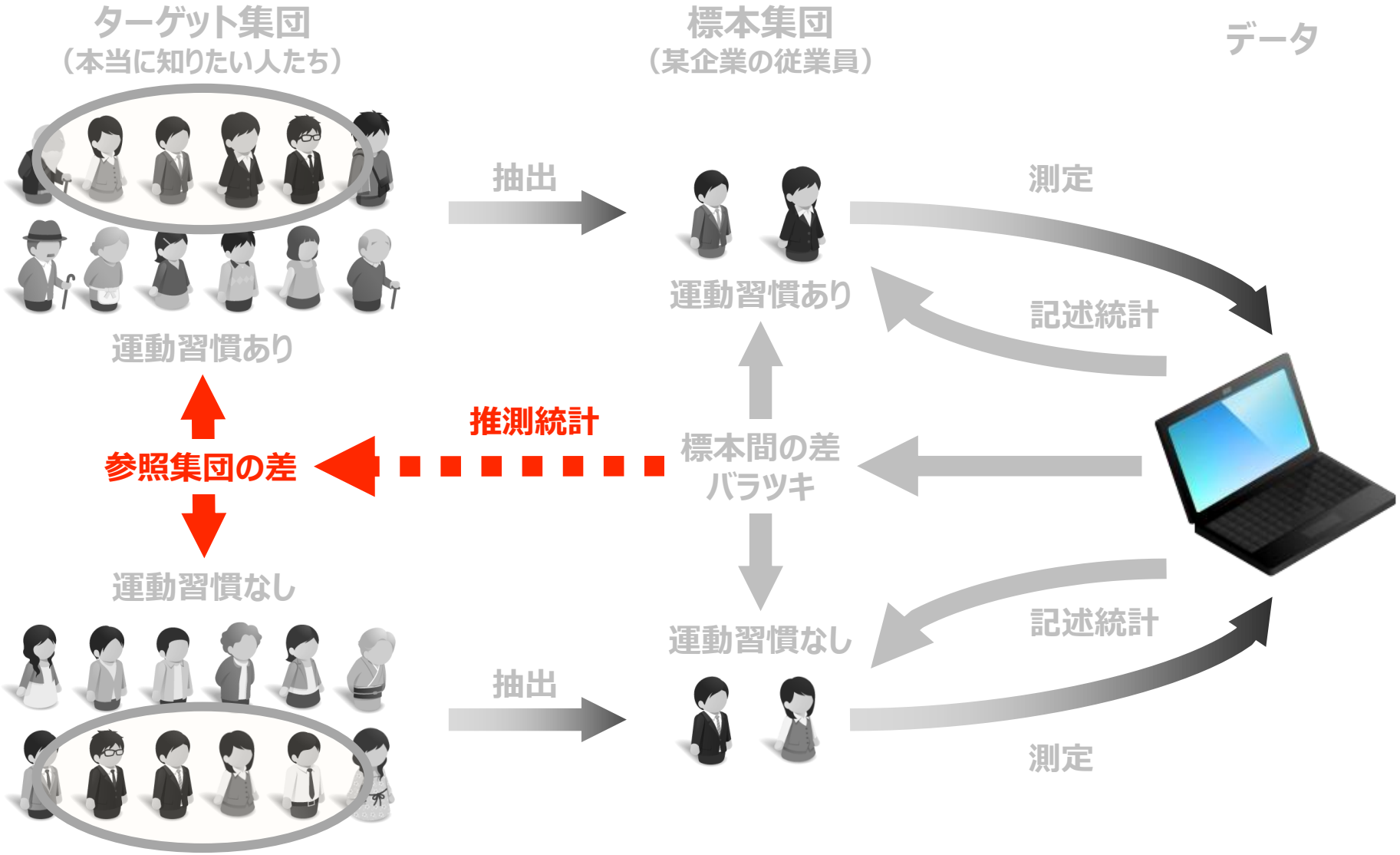
データを取得してから | 統計の守備範囲



対象者の選択と測定 | 疫学の守備範囲



推測について詳しく見ていきましょう。



推測の方法 | ターゲット or 参照集団に対して

□ 推定

- 具体的な **数値を示す** こと。
  - ⇨ **点推定**（1つの**数値**で）
    - ☞ 平均値、ハザード比、オッズ比 など
  - ⇨ **区間推定**（真値を含む確率と一緒に**範囲**で）
    - ☞ 95%信頼区間 など

□ 検定

~~統計学的な~~検定 | 何かを判断すること

- 簿記検定
- 英語検定
- 漢字能力検定
- 議員力検定
- 七転び八起き検定
- ホームパーティー検定
- サンタクロース検定
- 全国統一オタク検定
- 定年力検定
- お座敷遊びニスト検定
- 新日本プロレス検定
- こどもウルトラけんてい

Wikipedia（日本の検定試験一覧）等を参照

合格 or 不合格

□ 仮説検定

– 想定した（帰無）仮説が **誤っているか** を判断すること。**（正しいかではない）**

⇨「～がない」仮説 ➡ 検証すべき状況が**1つ**に絞られる。

☞ 両群間に差がない ➡ 差分は“ **0** ”の一点。

⇨「～がある」仮説では、想定できる状況が**無限**に存在する。

☞ 両群間に差がある ➡ 1？ 2？ 10？ 100？ −1？

仮説検定の流れ | ある研究室の日常

- よっしゃー！今日の夕方、やっと先生に博士論文をみてもらえる。
  あっても、今日は大人のサッカークラブの日だったような…。ちゃんと見てくれるかな…。

仮説を立てる	サッカーには行かない。 （サッカーより学生だよね！）
観察する	教授室をこっそり覗いてみる。 サッカーのユニフォームを着てるんですけど！
観察結果が仮説と矛盾しないかを考える	ユニフォームを着ているということは……。
判断する	サッカーに行かないとは考えにくい。 （てか、絶対サッカー行くでしょ！）

TOHOKU UNIVERSITY

ただし、こういう場合もあります。

仮説を立てる	サッカーには行かない。 （サッカーより学生だよね！）
観察する	教授室をこっそり覗いてみる。 サッカーのユニフォームを着てるんですけど！
観察結果が仮説と矛盾しないかを考える	ユニフォームを着ているということは……。
判断する	サッカーに行かないとは考えにくい。 （てか、絶対サッカー行くでしょ！）
実際	あれ？ちゃんと論文見てくれた。 （先生、疑ってごめんなさい）

あるいは、こうも考えられます。

仮説を立てる	サッカーには行かない。 （サッカーより学生だよね！）
観察する	教授室をこっそり覗いてみる。 サッカーのユニフォームを着てるんですけど！
観察結果が仮説と矛盾しないかを考える	ユニフォームを着ていたとしても……。
判断する	先生はきっと行かない。 （先生だったら大丈夫）
実際	やっぱり行かないでちゃんと論文見てくれた。 （先生、信じてましたよ！）

TOHOKU UNIVERSITY

でも、間違う場合もあります。

仮説を立てる	サッカーには行かない。 （サッカーより学生だよね！）
観察する	教授室をこっそり覗いてみる。 サッカーのユニフォームを着てるんですけど！
観察結果が仮説と矛盾しないかを考える	ユニフォームを着ていたとしても……。
判断する	先生はきっと行かない。 （先生だったら大丈夫）
実際	え？行くんですか？サッカー優先ですか？ （まっ、しゃーないから明日またがんばろう）

仮説検定における判断の誤り

		真 実	
		～ない	～ある
判断	～ない	正しい	第2種の過誤 （β）
	～ある	第1種の過誤 （α）	正しい （検出力）

あわてんぼの（α） & ぼんやりものの（β）

判断の根拠 | 統計による仮説検定の必要性

仮説検定は統計がなくても行える。

では、なぜ統計学的仮説検定が行われるのか？

判断基準に

客観性 と **共通性**

が必要なため。

さっきの例 | 判断材料 ＝ ユニフォーム

- よっしゃー！今日の夕方、やっと先生に博士論文をみてもらえる。
  あっても、今日は大人のサッカークラブの日だったような…。ちゃんと見てくれるかな…。

仮説を立てる	サッカーには行かない。 （サッカーより学生だよね！）
観察する	教授室をこっそり覗いてみる。 サッカーのユニフォームを着てるんですけど！
観察結果が仮説と矛盾しないを考える	ユニフォームを着ているということは……。
判断する	サッカーに行かないとは考えにくい。 （てか、絶対サッカー行くでしょ！）

統計学的仮説検定｜確率で判断する

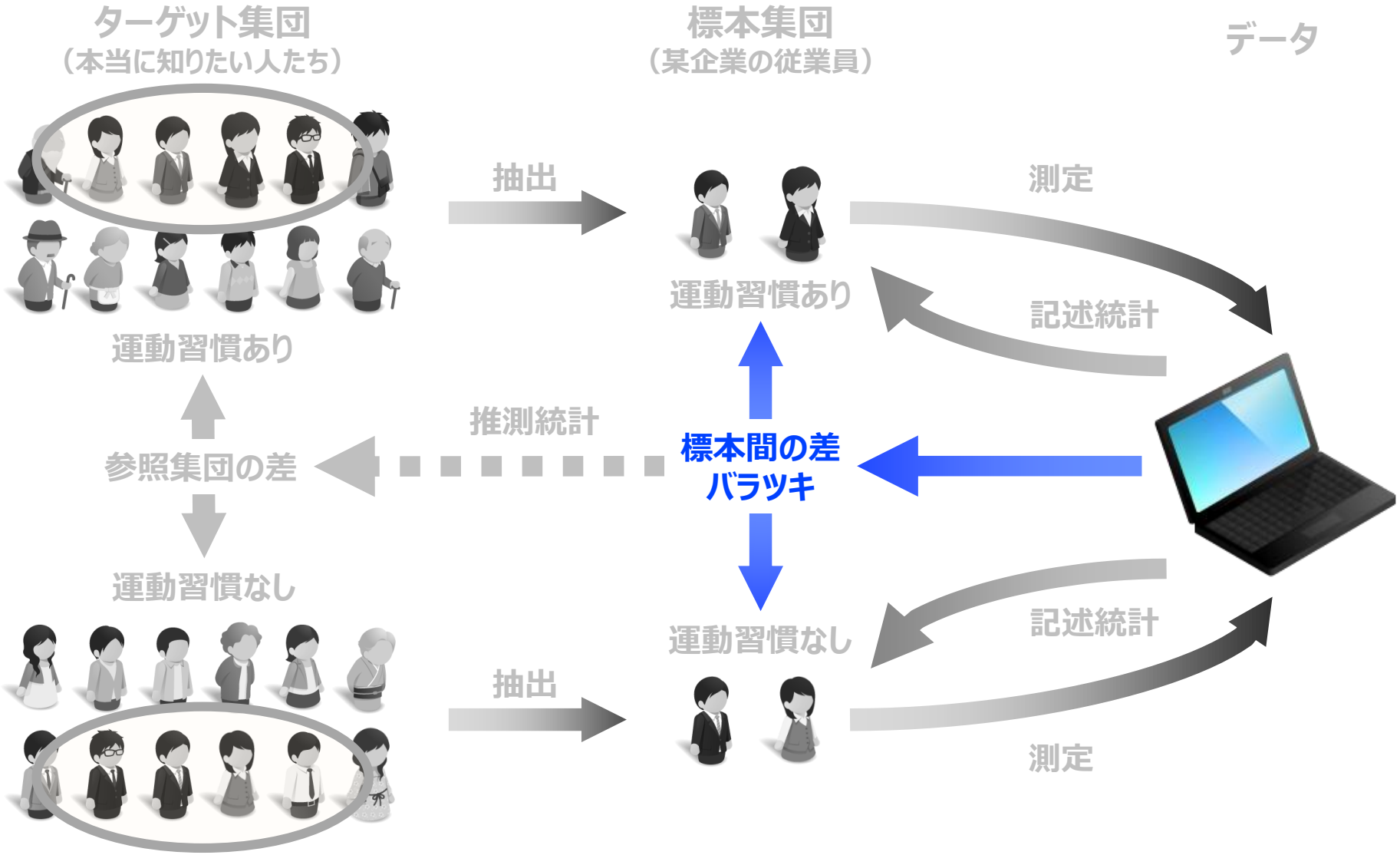
- **先生はサッカーに行かないとの立場で、先生が、普段どのくらいの確率でユニフォームを着ているかを考える。**

- **ケース1：普段から着ている確率：50%**
  - ユニフォームを着ているのはよくあること。珍しくもなんともない。
    ⇨「サッカーに行かなくたって、おかしくはない」と判断。

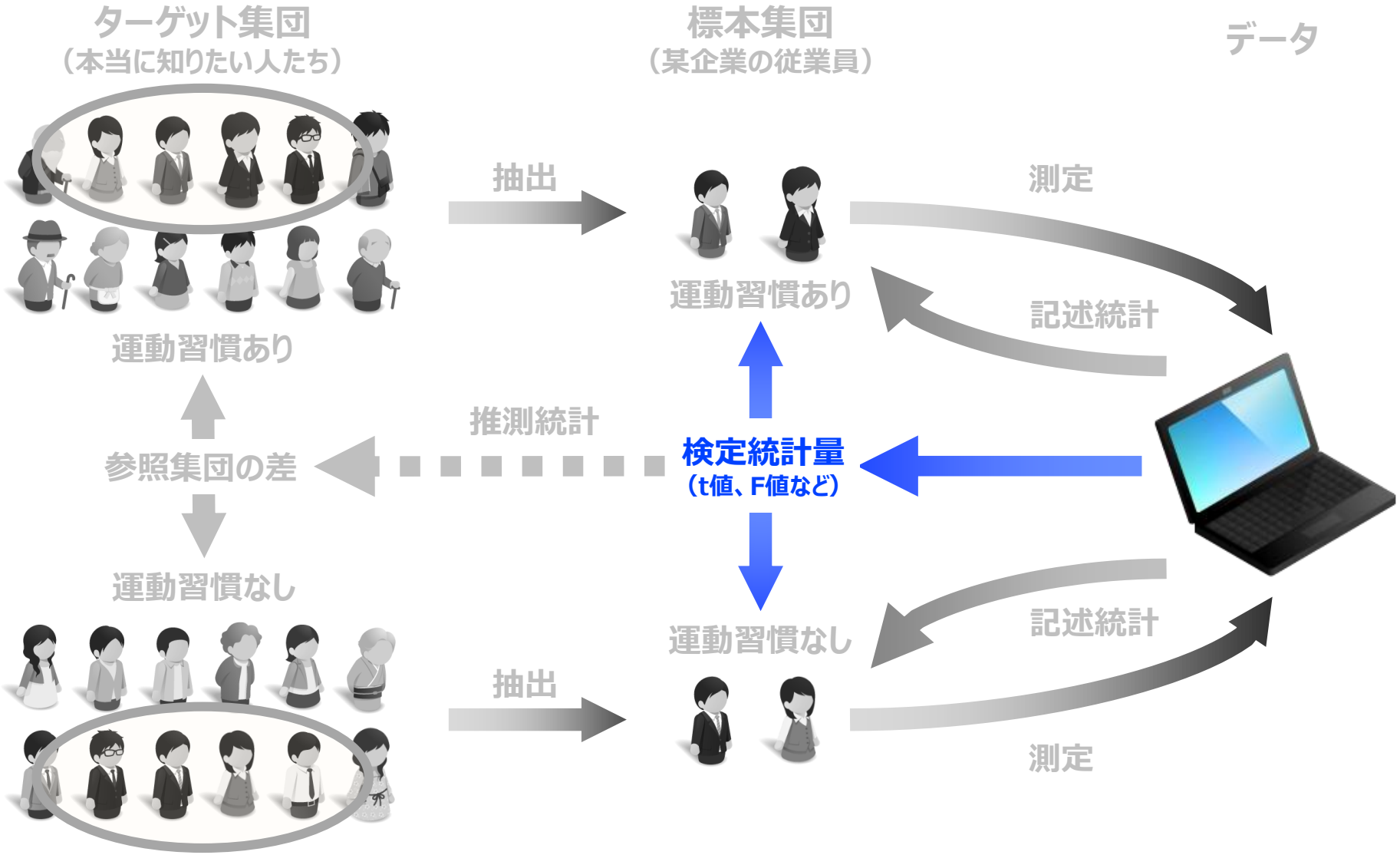
- **ケース2：普段から着ている確率：10%**
  - ユニフォームを着ているなんてめったにない。きっとサッカークラブがあるに違いない。
    ⇨「サッカーに行かないとは考えにくい」と判断。

「珍しい・めったにない」と判断する確率 ＝ 有意水準
（もし、30%に設定したとすると…）

研究でユニフォームに該当するものは？



研究でユニフォームに該当するものは？



研究における統計学的仮説検定

- ある企業の従業員を対象に、運動習慣の有無によって血糖値に差がないか検討する。

検定用の仮説を立てる	運動習慣の有無によって血糖値に差はない。 （あり群の血糖値－なし群の血糖値＝0）
検定統計量を算出する	平均値の差、全体のバラツキ、サンプル数から検定統計量（t値）を算出する。
検定統計量が出現する確率を算出する	t値の確率が5%以上 ➡ 珍しくない。信じられる。 t値の確率が5%未満 ➡ 珍しい！信じられない！
判断する	t値の確率が5%以上 ➡ 帰無仮説を信じる。 t値の確率が5%未満 ➡ 帰無仮説は信じない！

研究における統計の役割

なぜ我々は研究で統計を使用しているのか？

- 集団の特徴をまとめるため。
  - ☞ **記述統計（情報の要約・圧縮）**
- 観察された結果を客観的に判断するため。
  - ☞ **推測統計（研究結果の保証）**

TOHOKU UNIVERSITY

論文における統計の記載

～ 実践的なお話が中心です ～

次はこの質問をきっかけに。

統計は論文でどのように記載されているか？

参照論文

Journal of Epidemiology

Original Article J Epidemiol 2018

Importance of Achieving a "Fit" Cardiorespiratory Fitness Level for Several Years on the Incidence of Type 2 Diabetes Mellitus: A Japanese Cohort Study

Haruki Momma[1,2], Susumu S. Sawada[2], Robert A. Sloan[3], Yuko Gando[2], Ryoko Kawakami[4], Shin Terada[5], Motohiko Miyachi[2], Chihiro Kinugawa[6], Takashi Okamoto[6], Koji Tsukamoto[6], Cong Huang[1], Ryoichi Nagatomi[1], and Steven N. Blair[7,8]

[1]Division of Biomedical Engineering for Health and Welfare, Tohoku University Graduate School of Biomedical Engineering, Sendai, Japan
[2]Department of Health Promotion and Exercise, National Institutes of Biomedical Innovation, Health and Nutrition, Tokyo, Japan
[3]Department of Psychosomatic Internal Medicine, Graduate Medical and Dental School, Kagoshima University, Kagoshima, Japan
[4]Faculty of Sport Sciences, Waseda University, Tokorozawa, Japan
[5]Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan
[6]Tokyo Gas Health Promotion Center, Tokyo, Japan
[7]Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA
[8]Department of Exercise Science, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA

Received January 10, 2017; accepted May 5, 2017; released online November 25, 2017

全身持久力の基準の継続的達成と2型糖尿病の関連

論文における記載 | 統計解析のセクション

- 対象者のベースライン特性は全身持久力のグループ同士で比較した。
  連続変数は**中央値（四分位範囲）**、カテゴリー変数は**割合**を記載した。

  記述統計

- 2型糖尿病はアウトカム、全身持久力のグループは曝露要因とした。
  Cox比例ハザードモデルを用いて、全身持久力のグループ間における2型糖尿病の発症率を比較し、年齢調整**ハザード比**と**95%信頼区間**を算出した。
  さらに、以下の変数で調整を行った：（中略）。

  推定

- さらに、追跡後3年以内に糖尿病を発症した対象者を除外する感度分析を実施した。

（中略）

- 比例ハザード性の仮定はログマイナスログプロットで確認した。すべての解析はSPSSバージョン22で解析した。**両側検定**で**P値**が**0.05未満**である場合、**統計学的に有意**と見なした。

  検定

Momma et al. J Epidemiol. 2017

表1 対象者特性 | 対象集団の記述統計

	All	Fit_{AUC}	$Unfit_{AUC}$	
n	2235	971	1264	P 値
Age, years	43.0 (38.0–49.0)	44.0 (38.0–49.0)	43.0 (38.0–49.0)	
BMI, kg/m²	23.2 (21.7–25.0)	22.4 (20.9–23.9)	24.0 (22.3–25.6)	
・	・	・	・	
・	・	・	・	
・	・	・	・	
・	・	・	・	
・	・	・	・	

値は中央値（四分位範囲）もしくはn（%）で示す。

記述統計

推測統計
（検定）

P値の算出 ＝ 参照（もしくはターゲット）集団の差を検定

参照（もしくはターゲット）集団の差を検定する意味は？

Momma et al. J Epidemiol. 2017 を改変

メインの結果 | 記述統計 と 推測統計

グループ	人数	人年	罹患数	年齢調整 ハザード比	多変量調整 ハザード比[a]
Fit_{AUC}	971	13,980	128	1.00（Reference）	1.00（Reference）
$Unfit_{AUC}$	1,264	17,428	272	1.72（1.39-2.12）	1.33（1.06-1.65）
P値					

記述統計

推定

AUC, 1979年～1986年までの全身持久力の曲線下面積。
[a]年齢、BMI、収縮期血圧、喫煙習慣、飲酒習慣、糖尿病家族歴、デスクワーク、血糖値、全身持久力の測定回数。

さて、P値（検定結果）はどのように記載するか？

Momma et al. J Epidemiol. 2017 を改変

古典的の表記 | 検定（判断）するためのP値

グループ	人数	人年	罹患数	年齢調整 ハザード比	多変量調整 ハザード比[a]
Fit_{AUC}	971	13,980	128	1.00（Reference）	1.00（Reference）
$Unfit_{AUC}$	1,264	17,428	272	1.72（1.39-2.12）	1.33（1.06-1.65）
P値				P < 0.05	P < 0.05

記述統計

推定

AUC, 1979年～1986年までの全身持久力の曲線下面積。
[a]年齢、BMI、収縮期血圧、喫煙習慣、飲酒習慣、糖尿病家族歴、デスクワーク、血糖値、全身持久力の測定回数。

“検定” に関する情報であればこれで十分。

でも、今はこれだけでは不十分とされている。

Momma et al. J Epidemiol. 2017 を改変

現代的な表記 | 実値を記載する

グループ	人数	人年	罹患数	年齢調整 ハザード比	多変量調整 ハザード比[a]
Fit_{AUC}	971	13,980	128	1.00（Reference）	1.00（Reference）
$Unfit_{AUC}$	1,264	17,428	272	1.72（1.39-2.12）	1.33（1.06-1.65）
P値				<0.001	0.012

記述統計

推定

AUC, 1979年～1986年までの全身持久力の曲線下面積。
[a]年齢、BMI、収縮期血圧、喫煙習慣、飲酒習慣、糖尿病家族歴、デスクワーク、血糖値、全身持久力の測定回数。

P < 0.05：白か黒か？　（2値変数）

P値の実値：グレーの程度（連続変数）

Momma et al. J Epidemiol. 2017 を改変

改めて統計とは？ | 記述統計 と 推測統計

グループ	人数	人年	罹患数	年齢調整 ハザード比	多変量調整 ハザード比[a]
Fit_{AUC}	971	13,980	128	1.00（Reference）	1.00（Reference）
$Unfit_{AUC}$	1,264	17,428	272	1.72（1.39-2.12）	1.33（1.06-1.65）
P値				<0.001	0.012

記述統計

検定

推定

AUC, 1979年～1986年までの全身持久力の曲線下面積。
[a]年齢、BMI、収縮期血圧、喫煙習慣、飲酒習慣、糖尿病家族歴、デスクワーク、血糖値、全身持久力の測定回数。

記述：**データを与えてくれた集団** の様子を **要約** すること。

推定：**参照（もしくはターゲット）集団** の様子を具体的な **数値** で示すこと。

検定：**参照（もしくはターゲット）集団** の様子を確率に基づいて **判断** すること。

Momma et al. J Epidemiol. 2017 を改変

統計に関する見解｜お恥ずかしいですが…

Q：統計についてどう思ってました？

正直、論文書いていても、そこまで**気にかけなかった**んです。

とりあえず、結果とP値を書けばいいかなって。

みんなやっているから、私もやろうかなって。

気にかけるようになったきっかけ。

Journal of Epidemiology

Original Article J Epidemiol 2019

Physical Fitness Tests and Type 2 Diabetes Among Japanese: A Longitudinal Study From the Niigata Wellness Study

Haruki Momma[1,2,3], Susumu S Sawada[2], Kiminori Kato[4], Yuko Gando[2], Ryoko Kawakami[5], Motohiko Miyachi[2], Cong Huang[6,7], Ryoichi Nagatomi[1,7], Minoru Tashiro[8], Masahiro Ishizawa[3], Satoru Kodama[4], Midori Iwanaga[3], Kazuya Fujihara[3], and Hirohito Sone[3]

[1]Division of Biomedical Engineering for Health and Welfare, Tohoku University Graduate School of Biomedical Engineering, Miyagi, Japan
[2]Department of Health Promotion and Exercise, National Institutes of Biomedical Innovation, Health and Nutrition, Tokyo, Japan
[3]Department of Hematology, Endocrinology and Metabolism, Niigata University Faculty of Medicine, Niigata, Japan
[4]Department of Laboratory Medicine and Clinical Epidemiology for Prevention of Noncommunicable Diseases, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan
[5]Faculty of Sport Sciences, Waseda University, Saitama, Japan
[6]Department of Physical Education and Sports Science, Zhejiang University, Zhejiang, China
[7]Department of Medicine and Science in Sports and Exercise, Tohoku University Graduate School of Medicine, Miyagi, Japan
[8]Niigata Association of Occupational Health, Niigata, Japan

Received November 14, 2017; accepted March 5, 2018; released online July 28, 2018

6つの体力テストと2型糖尿病の関連

ありがたいある査読コメント

複数の体力を扱っているため、**検定の多重性**が気になる。

多重検定の問題 | 数撃ちゃ当たる

P値

- 観察結果をもとに「～がある」と判断するとき、その判断が間違っている確率。
  - ⇨「まぐれ」で結果が得られる確率。

例えば…

- 打率0割5分（5%）のバッターが1試合（5打席）で1本でもヒットを打つ確率。
  - ⇨ 1本でもヒットを打つ確率 ＝ $1.00 - \text{打てない確率}^{\text{打席数}}$
    - ☞ 1打席： 5.0% ＝ $1.00 - (0.95^1)$
    - ☞ 2打席： 9.7% ＝ $1.00 - (0.95^2)$
    - ⋮ ⋮ ⋮
    - ☞ 5打席：22.6% ＝ $1.00 - (0.95^5)$
- 打席に多く立てば、まぐれでヒットが打てる確率も高くなる。
  - ⇨ **検定をいっぱいすれば、まぐれで有意と判断する確率も高くなる。**

一般的な対処｜打席数を考えようぜ

◻ 有意水準（α）の補正

- 有意水準を検定数で割る。
  - ⇨ 0.05 ÷ 6＝ **0.008** → P < 0.008である場合、帰無仮説を棄却する。

◻ P値（出現確率）の補正

- 算出されたP値に検定数を掛ける。
  - ⇨ 0.01 × 6 ＝ **0.06** → 帰無仮説は棄却されない。

これらの補正を行っても研究結果に変わりはなかった。

でも、本当に研究者としてこの回答だけでいいのか？

どこからどこまでが多重検定なのか？

- **本研究の体力測定は6つだけど、割る数は “6” でいいのか？**
  - 同じコホートで報告されているこれまでの解析は考慮しなくてよいのか？それとも、
  - 報告されている解析数でよいのか？結果が公表されなかった検定は？それとも、
  - 自分がこれまで実施した検定数？

- **もし、“6” でよいとしたら…**
  - 体力テストを別々の論文で発表すれば、多重検定に該当しないのか？
    - ⇨ 同じことをやっているのに、一貫性がない。
    - ⇨ サラミ論文という観点からも不適切。

みなさん、どう思いますか？

悩んだ挙げ句の発想

検定をやめれば、多重検定の問題は生じないはず。

（判断するから、間違える？）

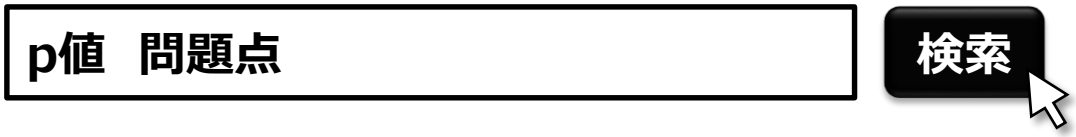
判断は読者に任せよう | 補正はしない方向で

本文

- 統計学的仮説検定に関する記述はすべてやめた。
  - ⇨「significantly」や「significant」という単語はすべて削除。
  - ⇨「All statistical tests were 2-sided.」も削除。有意水準の記載なし。

返答書

- 検定より推定のほうがより重要である立場を宣言。
- そのため、信頼区間を算出していることを強調。
- さらにP値は読者にとって馴染みがあるため、一応、正確な値を記載したことを付記。
- 検定ではなく推定を支持する声明を引用して補足。
  - ⇨ Wasserstein and Lazar. The American Statistician. 2016. 70:129-133
  - ⇨ Greenland et al. Eur J Epidemiol. 2016. PMID: 27209009



論文における統計の記載

統計は論文でどのように記載されているか？

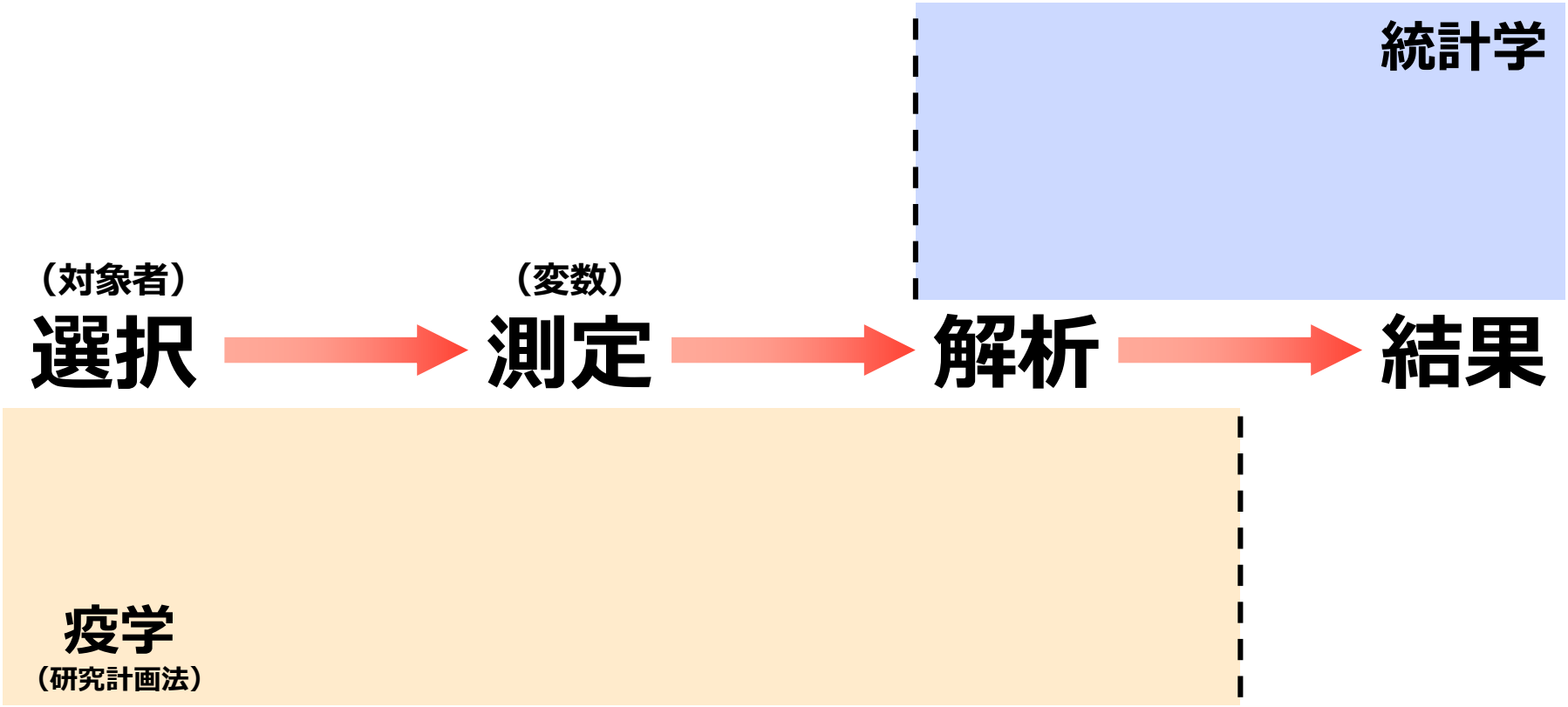
- 記述統計（標本集団の状態を要約する）
  - ☞ **推定値やP値以外。**
- 推定（ターゲット集団の状態を数値で示す）
  - ☞ **モデルから算出された推定値。**
- 検定（ターゲット集団の状態を判断する）
  - ☞ **P値、significant、有意水準、両側 などなど**

TOHOKU UNIVERSITY

疫学マインドで眺める統計学

～ 統計の守備範囲を再確認します ～

疫学と統計学の境界線 | 実際の守備範囲



守備とは？ | 敵の攻撃に備えて守ること

研究における 敵 とは何ぞや？

誤 差

疫学 と **統計学** は

誤差から研究を守っている。

研究 | 誤差の海を泳ぐようなもの。

誤差

誤差

- 偶然誤差
- 系統誤差

□ 偶然誤差

- **ランダム**な（方向を持たない）誤差
  - ⇨ 測定精度や個人差が影響

□ 系統誤差

- **一方向**に偏った誤差
  - ⇨ 対象者選択や測定方法が影響
  - ⇨ 他の因子の影響

誤差の海 | 海には波と潮の流れがある。

誤差

誤差
- 偶然誤差（波の高さ）
- 系統誤差（潮の流れ）

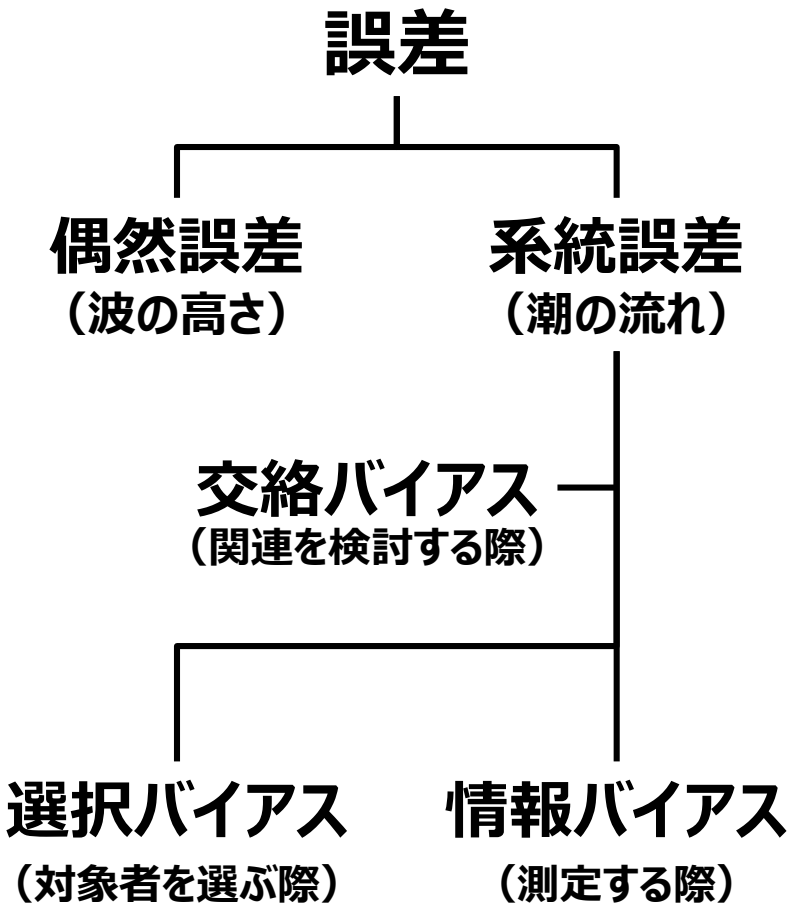
波が高ければ溺れやすい。

ただ、波が低くても…

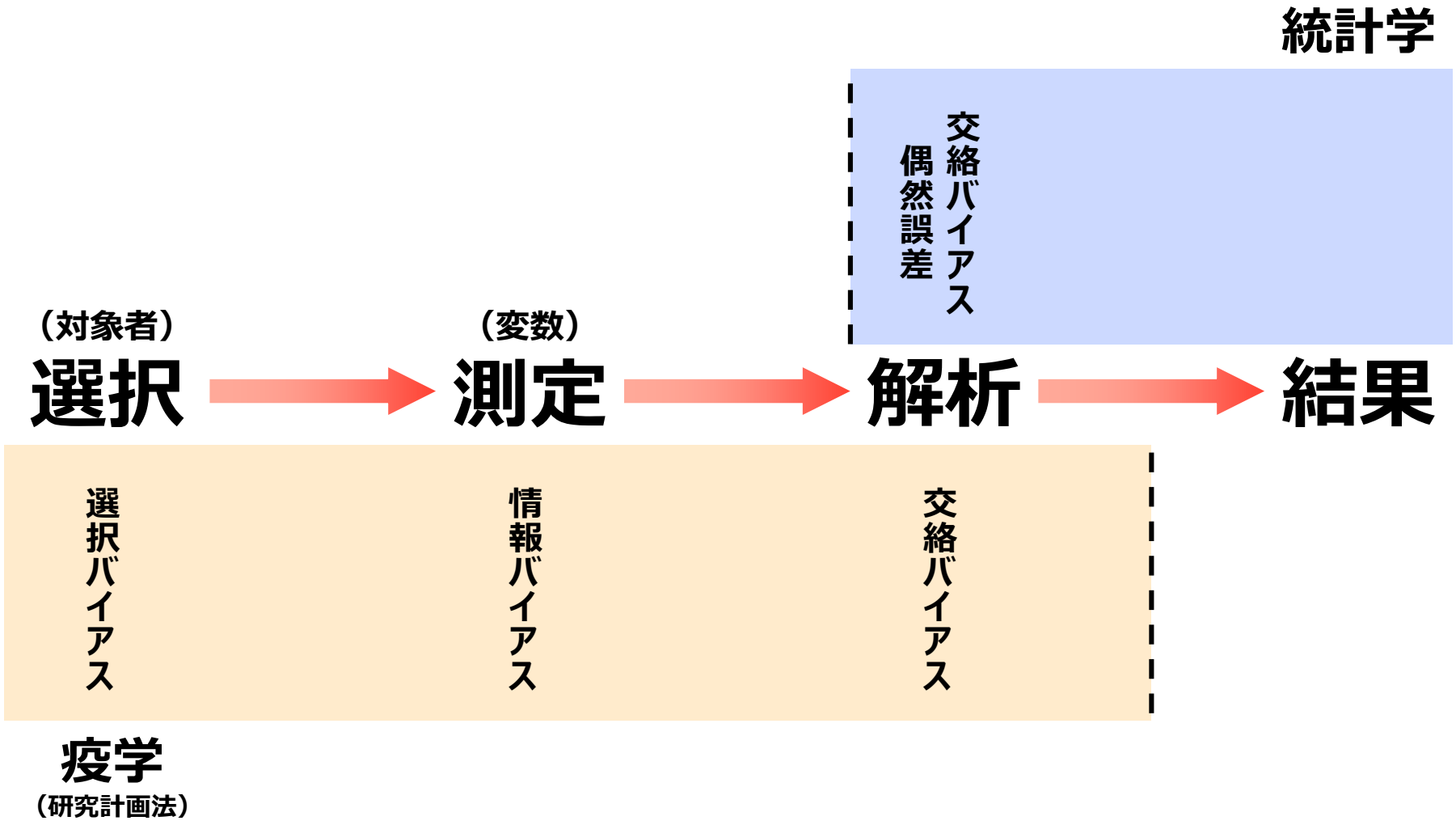
潮の流れがきついと、
真なる目的値まで辿り着けない。

系統誤差の種類 | 混入するタイミング

誤差



誤差に対する守備範囲 | 疫学と統計学



疫学マインドで眺める統計学

- **統計は情報の要約や結果の保証するためだけではなく、交絡への対応も（少なからず）できるため、重要な役割を果たす。**
- **ただし、統計が力を発揮できるのは、データを取得してからである。**
- **統計の力を十分に発揮させるためには、系統誤差の混入を可能な限り抑える必要がある。**
  - 偶然誤差は、人数を増やしたり、精度のよい指標にすれば（勝手に）小さくなる。
- **疫学は系統誤差への対処を一生懸命考える学問であり、研究においては、統計に先立つものである。**

本日のまとめ

- **研究における統計の役割は全部で3つある。**
  - 情報の要約。
  - 結果の保証。
  - 誤差に対する守備。

- **統計も疫学も誤差から研究を守ってくれているが、統計はデータを取得してから勝負で、疫学はデータを取る前が勝負。**

- **「ただ何となく」ではなく、やる理由を考えてみてはどうだろうか？**

- **何気なくやっている統計解析の意味を少し理解すると、研究も少し理解できて、おもしろくなってくるはず。**