

# 人を対象とする医学系研究に おける統計解析の基本

大塚 俊昭

日本医科大学付属病院 臨床研究総合センター

日本医科大学 衛生学公衆衛生学

学校法人日本医科大学 研究統括センター

令和元年度 認定倫理講習会

2019年11月8日

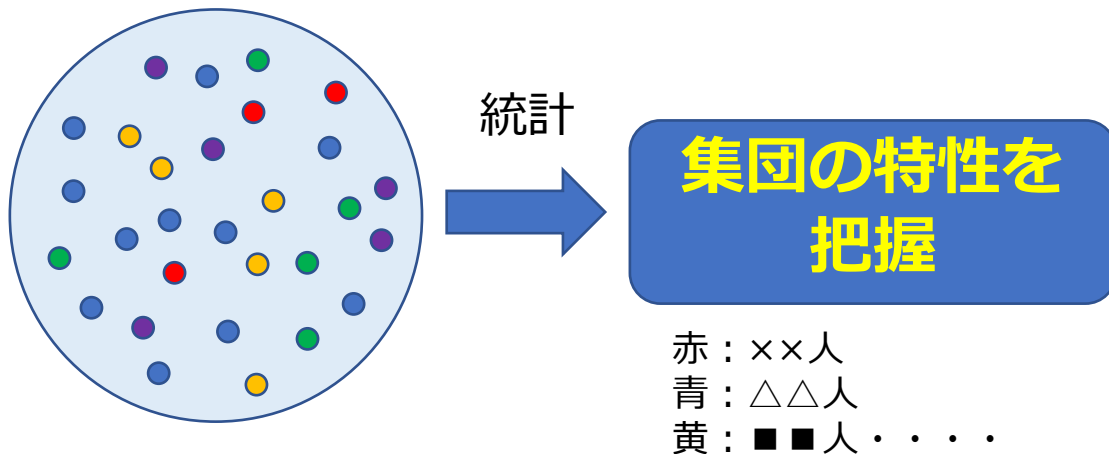
# なぜ医学研究に統計解析が必要か

- 統計解析は科学的検証の手段
- 研究結果およびその解釈を客観化、標準化する

# 「統計」とは

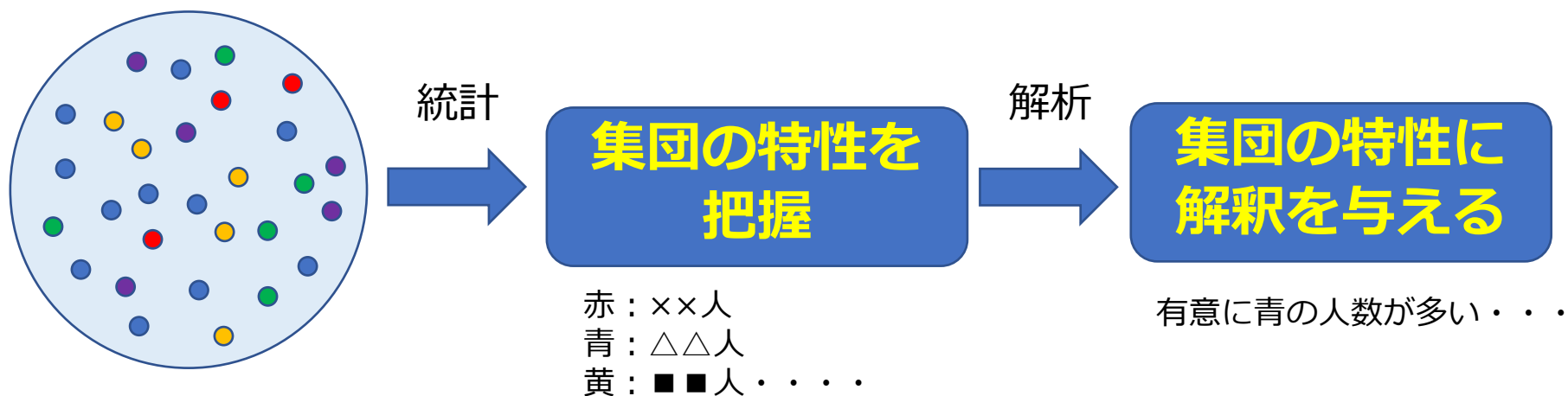
集団の個々の構成要素の分布を調べ、その集団の属性(特性)を数量的に把握すること。

デジタル大辞泉



# 「統計解析」とは

統計学の手法を用いて、集団属性(特性)について客観的な解釈を与えること



# 集団の特性を表す指標

## • 平均値

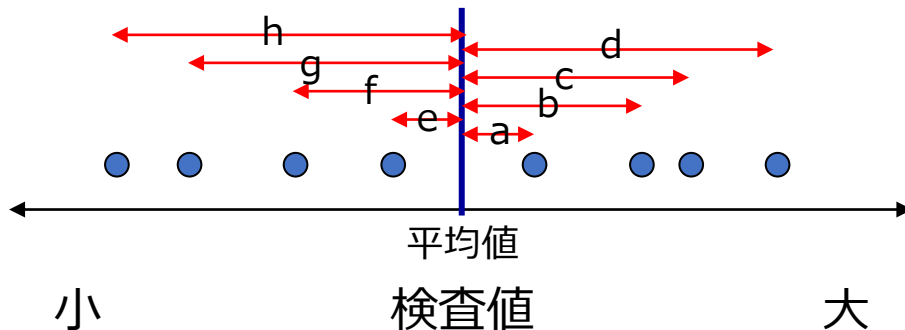
- ✓ いくつかの数や量の中間的な値
- ✓ 全部の値を足し合わせて、サンプル数で割った値 (算術平均)

## • 分散

- ✓ 個々の値が平均値からどのくらい散らばっているか？
- ✓  $= ((\text{個々の値} - \text{平均値}) \text{ の } 2 \text{ 乗}) \text{ の総和} \div \text{サンプル数}$

## • 標準偏差

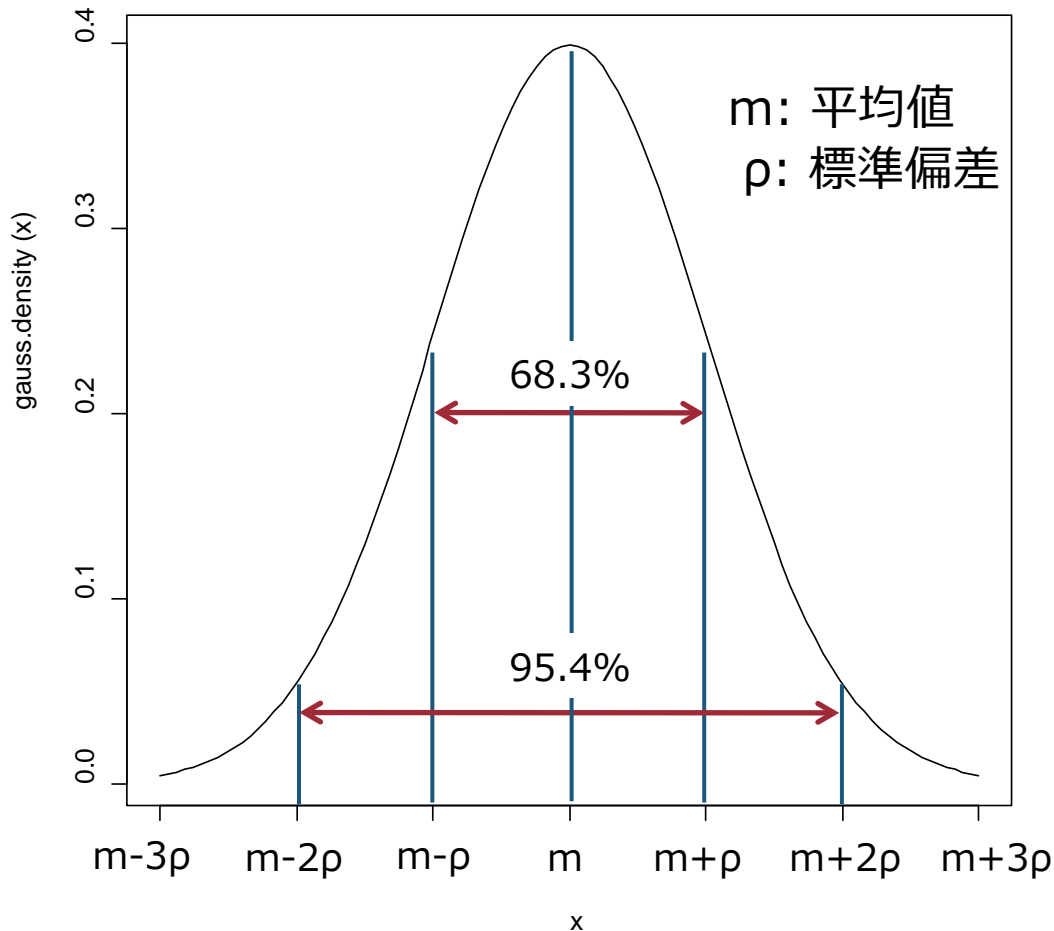
$$\checkmark = \sqrt{\text{分散}}$$



$$\text{分散} = \frac{a^2 + b^2 + c^2 + d^2 + e^2 + f^2 + g^2 + h^2}{8}$$

# 平均や標準偏差は、正規分布が前提

正規分布：左右対称の釣鐘型で、左右に裾を引く形状の分布



## 正規分布になりやすい

年齢、BMI、血圧など

## 正規分布になりにくい

入院日数、入院費用、  
CRPのようなマーカー値、  
薬剤の投与量など

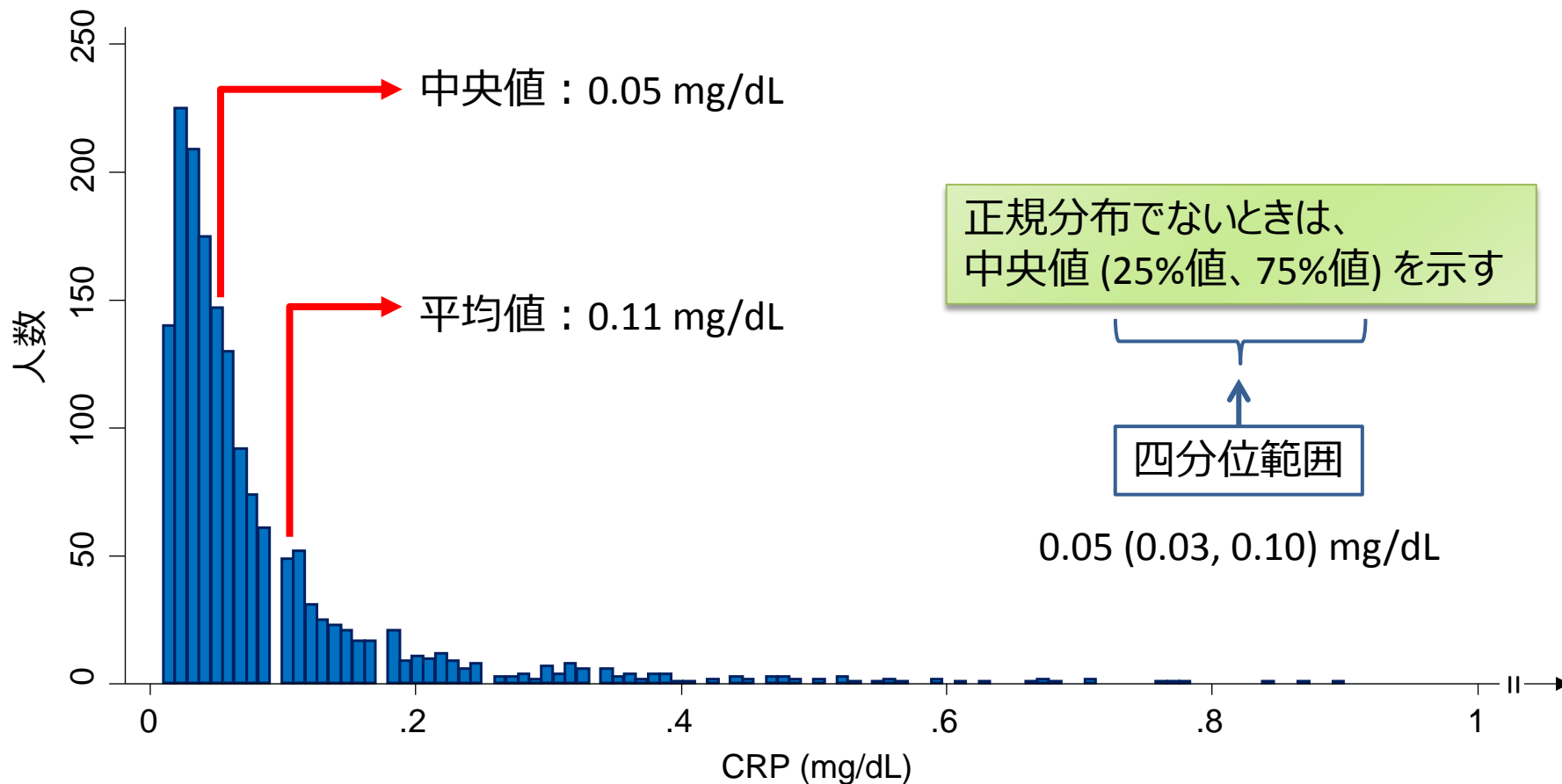
平均 $\pm 1p$ の範囲に、68.3%  
平均 $\pm 1.96p$ の範囲に、95%  
平均 $\pm 2p$ の範囲に、95.4%

のデータが存在する

# 正規分布でないときは中央値

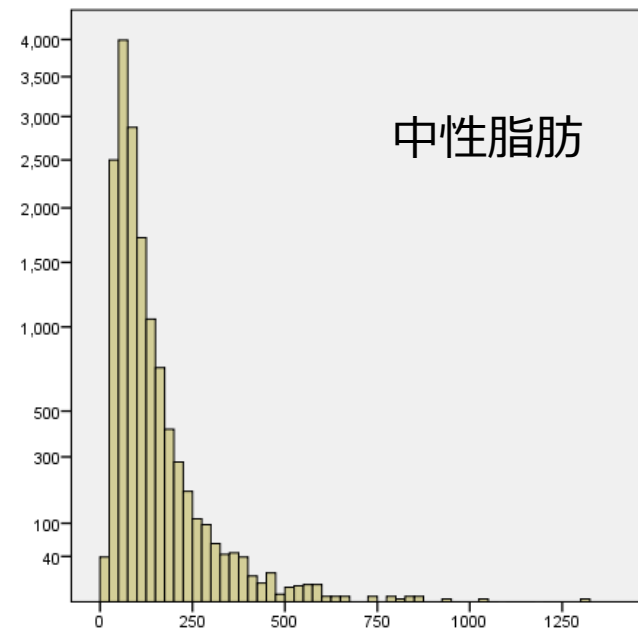
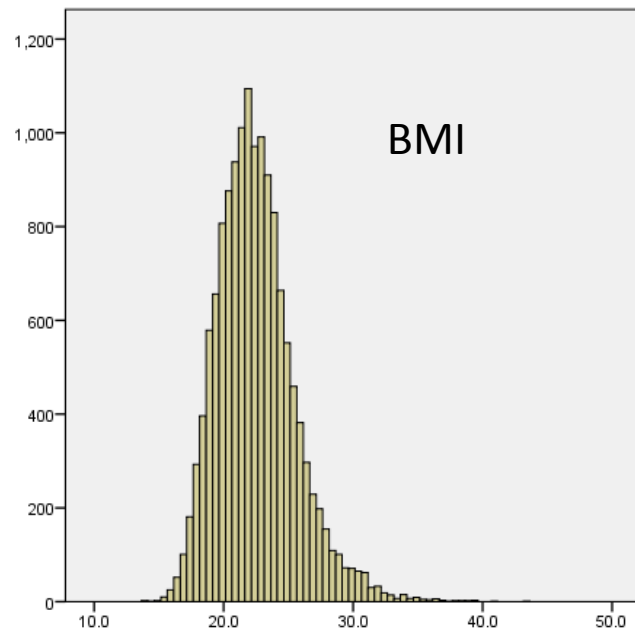
## 健康診断におけるCRPの分布

- 中央値・・・測定値を大きさの順に並べたとき、全体の中央にくる値



# 正規分布か否かの判断

正規性の判定： Kolmogorov-Smirnov (コロモゴロフ・スミノフ) 検定  
正規P-Pプロット、正規Q-Qプロット



実際には、厳密に判定する必要はない。  
視覚的に、分布に「歪み」がないかを判断すればよい。  
(私見)



# 統計学的検定

- その集団特性が単なる偶然なのか？意味のある現象なのか？統計学的に判定すること
- 「仮説」を検定する  
(統計学的仮説検定ともいう)
- 背理法を用いる

# 統計学的検定は「背理法」

- 命題の仮定のほかに、結論の否定を仮定して推論を行ない、そこから導かれる矛盾を示すことで命題を証明する方法
- A : この薬は効く
- B : この薬は効かない
- ~~• Aが正しいことを証明する~~
- Bが正しくないことを証明する

# なぜ「背理法」なのか

- 仮説を支持するためのデータはいくらあってもきりが無い。
- 仮説を棄却するには一つのデータで十分

## • この世のカラスは黒い

仮説を支持するためには…

→ 生息する全てのカラスを観察して確かめる（実現不可能！）

## • この世のカラスは白い

仮説を棄却するためには…

→ 黒いカラスを（1羽でも）見つければ良い

# 背理法による統計学的検定の進め方

新薬Xの治癒率8割、偽薬の治癒率5割

## 1 仮説を設定する

**帰無仮説**：新薬Xは効かない → 治癒率に**差がない**（差は偶然）

**対立仮説**：新薬Xは効く → 治癒率に**差がある**

## 2 帰無仮説の立場で、その差が生じる確率を算出する

治癒率の差（8割と5割）は、1%の確率でしか起こらない現象だった！

## 3 事前に、帰無仮説を棄却する(偶然とはみなせない)確率を設定しておく(有意水準)：通常は5%に設定

## 4 算出された確率と、事前に設定した有意水準とを比較し、帰無仮説が成立するか(否か)を判定

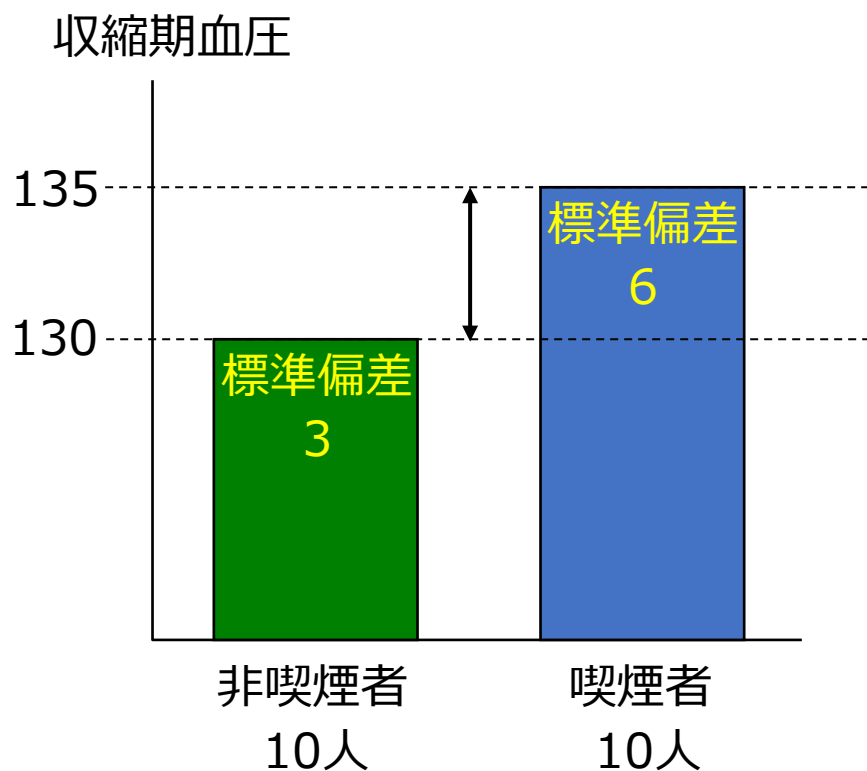
1%の確率でしか起こらない現象であれば、これは偶然ではない！  
→ 帰無仮説を棄却、対立仮説を採択

# 5%の根拠は？

- サイコロを振って偶数、奇数が出る確率は各々50% (0.5)
  - 2回連続で奇数が出る確率  $0.5^2 = 0.25$
  - 3回連続で奇数が出る確率  $0.5^3 = 0.125$
  - 4回連続で奇数が出る確率  $0.5^4 = 0.0625$
  - 5回連続で奇数が出る確率  $0.5^5 = 0.03125$
- 
- どこまでを「偶然」と考えるか？
  - どこから、「理由がある（偶然ではない）」と考えるか？（サイコロに細工？）

# 統計学的検定と帰無仮説

- 喫煙者は血圧が高い？



この差は偶然なのか？

- 帰無仮説：差はあるとはいえない(偶然)
- 対立仮説：差がある(偶然ではない)

↓

ある分布(**確率分布**)に従う変数(**確率変数**)  
を求める

t 検定であれば、t 分布と確立変数 t

↓

その変数が分布のどこに位置するかを確認し、帰無仮説が成立する確率を求める。

↓

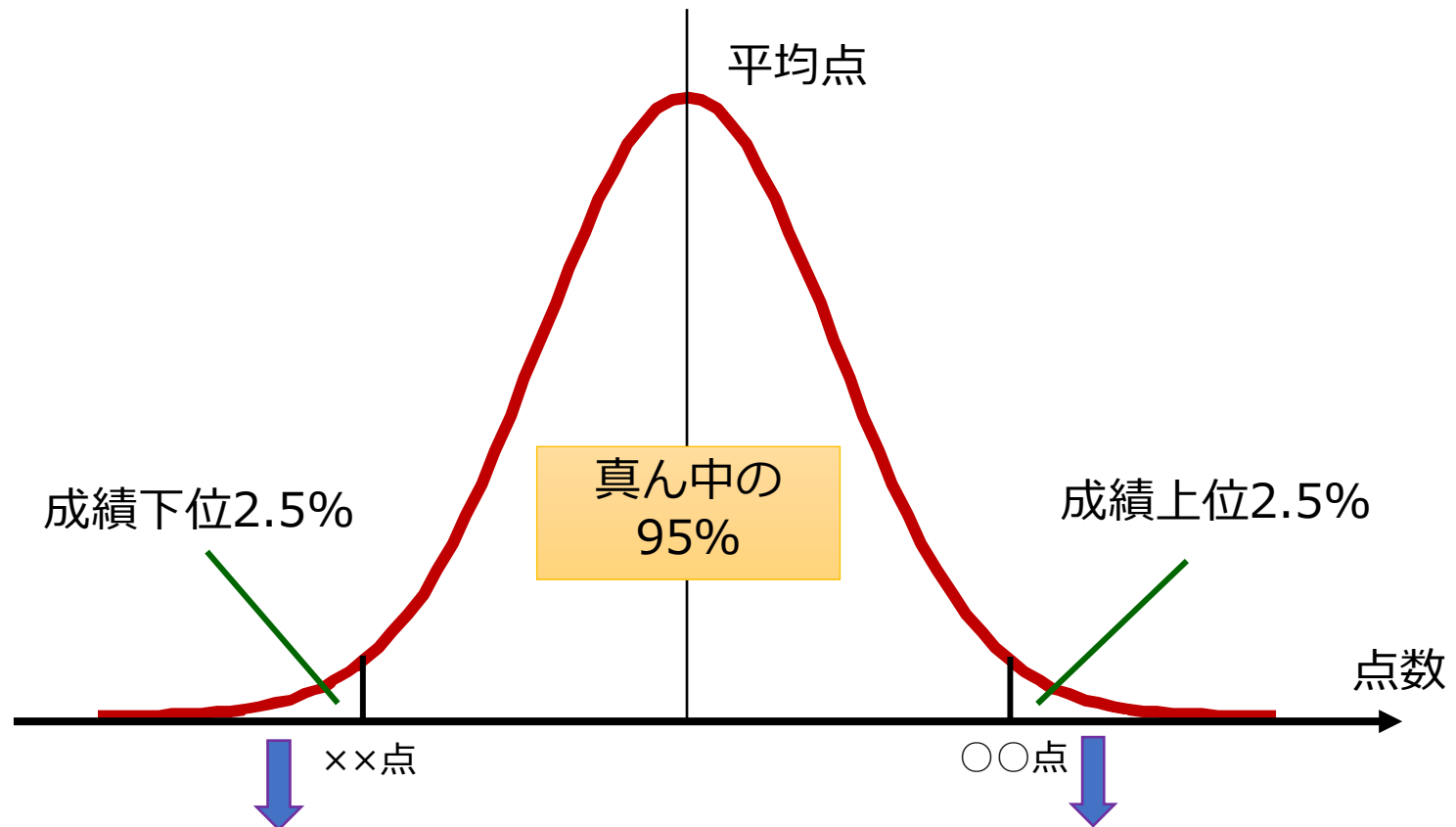
確率が5%以上

→帰無仮説を採用(偶然)

確率が5%未満

→対立仮説を採用(偶然ではない)

# 確率変数は試験点数と同じ



- ・勉強できない環境的理由？
- ・経済的理由？

- ・人より多い努力量？
- ・先天的に優れた脳機能？

偶然では説明しがたい「理由」があるとみなす

# 試験の結果

# 統計学的検定

受験者の解答内容

平均、個体数、分散

採点

確率変数の計算

点数の決定

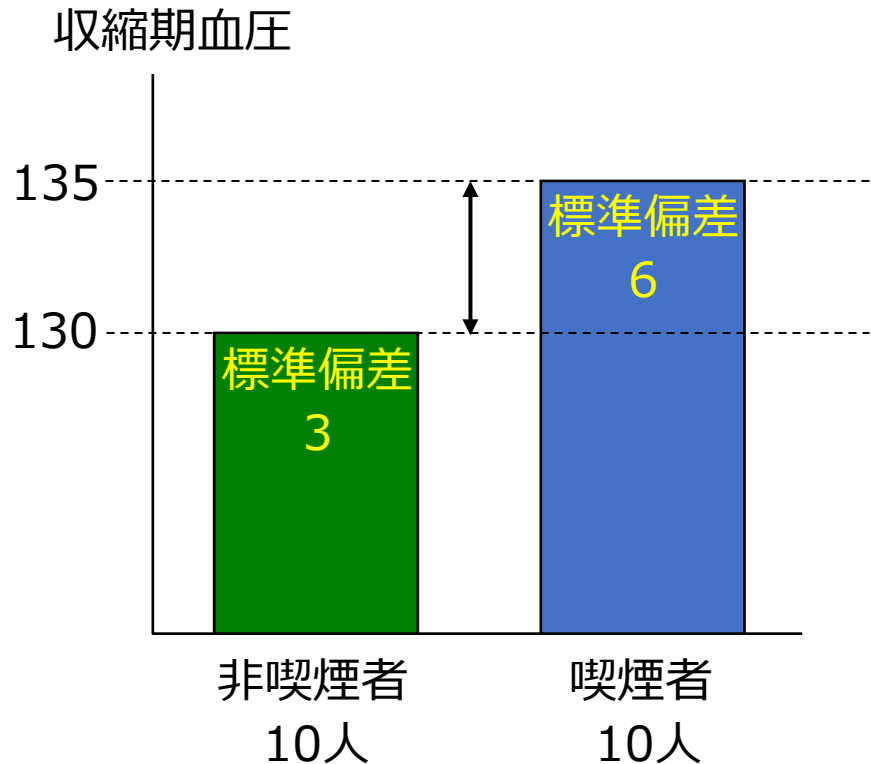
確率変数の決定

分布の端5%なら…  
偶然ではない(天才? or  
何かしら問題?)  
真ん中95%なら…  
偶然(天才とはいえず)

分布の端5%なら…  
偶然ではない(新薬の効果  
か?)  
真ん中95%なら…  
偶然(効果ありとはいえず)



# 確率変数 t の計算

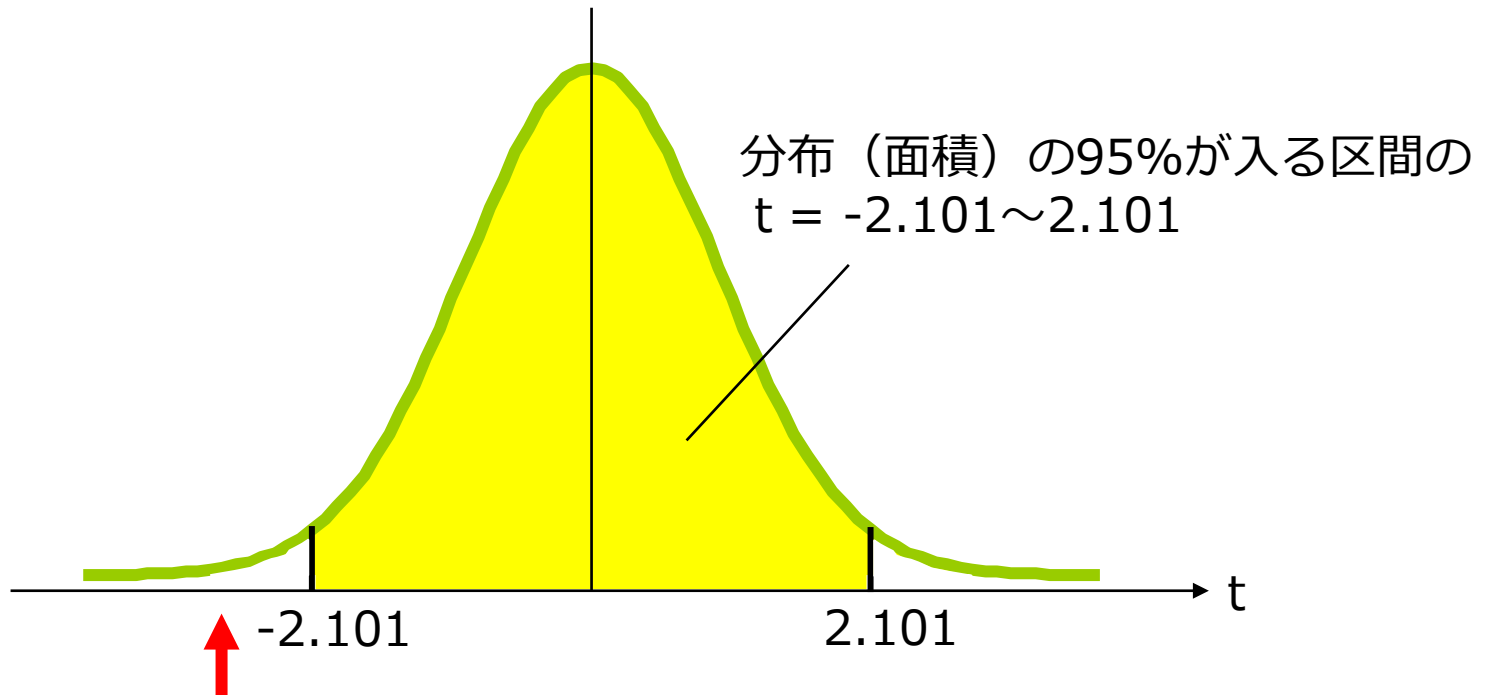


算出する変数： t

$$t = \frac{\text{標本Aの平均} - \text{標本Bの平均}}{\sqrt{\frac{\text{標本Aの分散}}{Aの個体数} + \frac{\text{標本Bの分散}}{Bの個体数}}}$$
$$= \frac{130 - 135}{\sqrt{\frac{9}{10} + \frac{36}{10}}} = \frac{-5}{2.121} = \mathbf{-2.357}$$

# 変数 $t$ の分布

自由度18の  $t$  分布



$$t = -2.357 < -2.101$$

$t$  分布の95%分布の区間内に入らない（5%の区間内に入る）

→ **非喫煙者と喫煙者の血圧差は統計学的に有意な差とみなす  
（喫煙者の血圧値が有意に高い）**

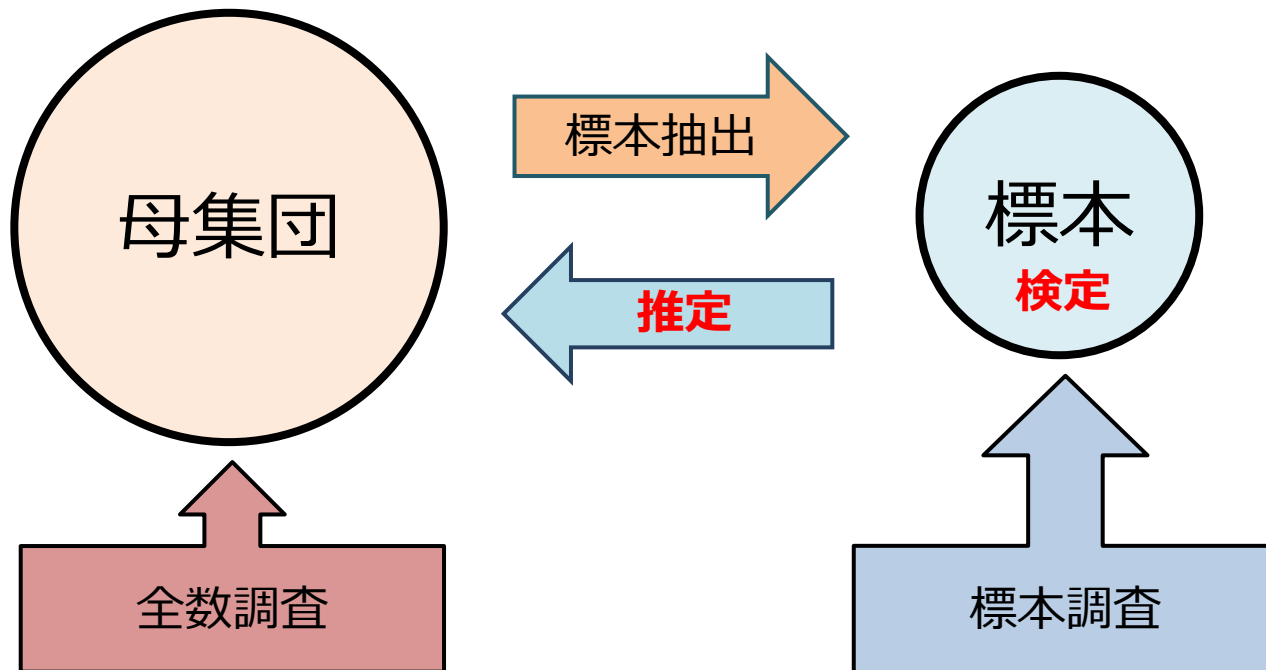
# 「検定」と「推定」

検定：標本を用い、母集団の特性に関する仮説を統計学的に検証すること

母平均がわかっていて本当に正しいかを判定

推定：標本が属する母集団の特性を、統計学的に推測すること

母平均が未知であり、どの範囲にあるかを知る

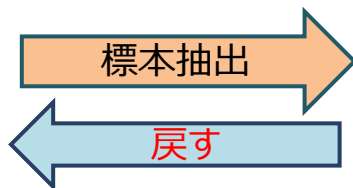
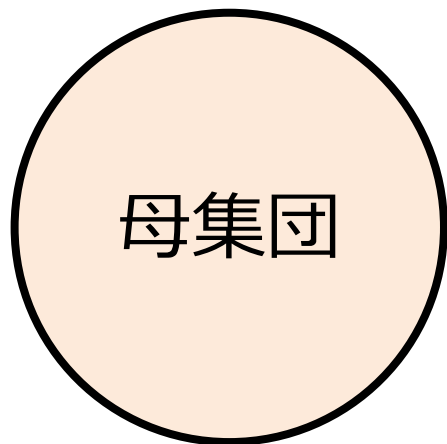


# 区間推定と点推定

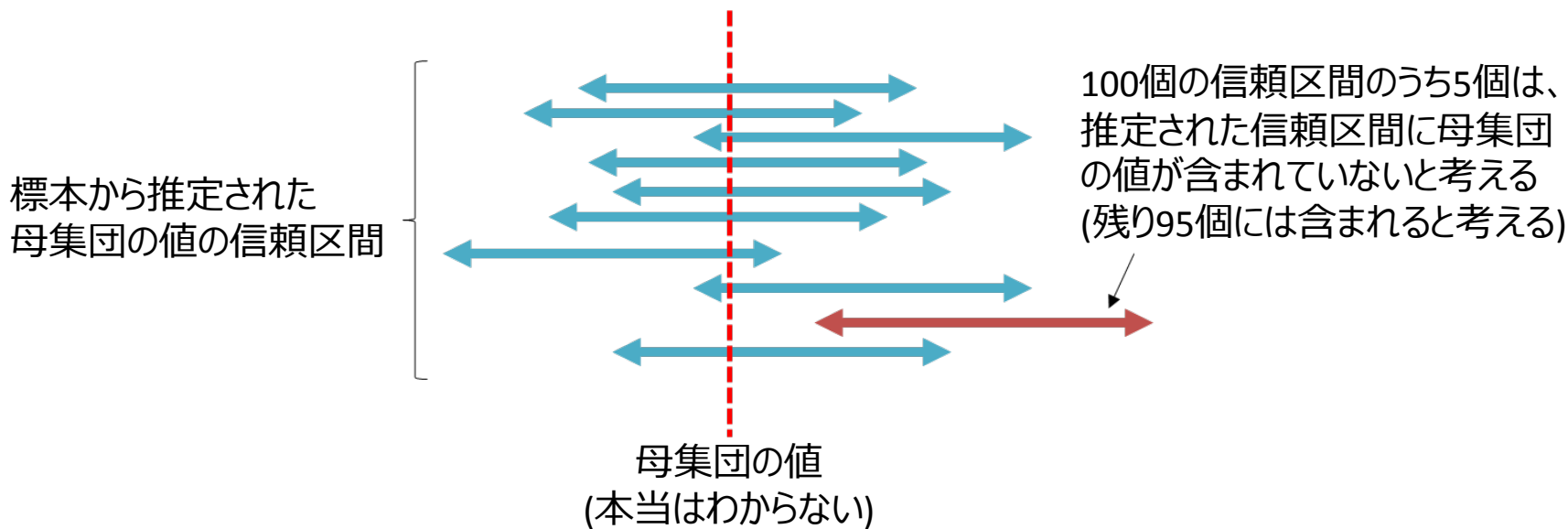
- 推定 → 通常は区間推定を指す  
(点推定は、検定の結果を母集団に外挿する)
- 区間推定
  - 母集団の特性について、ある区間で推定すること
  - 通常は、95%信頼区間を用いる

# 95%信頼区間

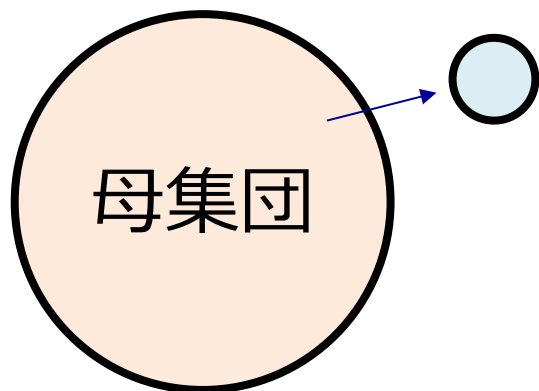
母集団から標本を取ってきて信頼区間を求める、  
という作業を100回やったときに、95回はその中に  
母集団の値が含まれると考える区間



- 標本1から信頼区間を算出
- 標本2から信頼区間を算出
- ⋮
- 標本100から信頼区間を算出



# 母集団と標本集団との関係



標本サンプル数10	平均 $\bar{x}_1$	分散 $\sigma^2_1$
標本サンプル数20	平均 $\bar{x}_2$	分散 $\sigma^2_2$
標本サンプル数30	平均 $\bar{x}_3$	分散 $\sigma^2_3$
	⋮	
標本サンプル数 $n$	平均 $\bar{x}_n$	分散 $\sigma^2_n$

標本の平均



母集団の平均に近づく

標本の分散 /  $n$

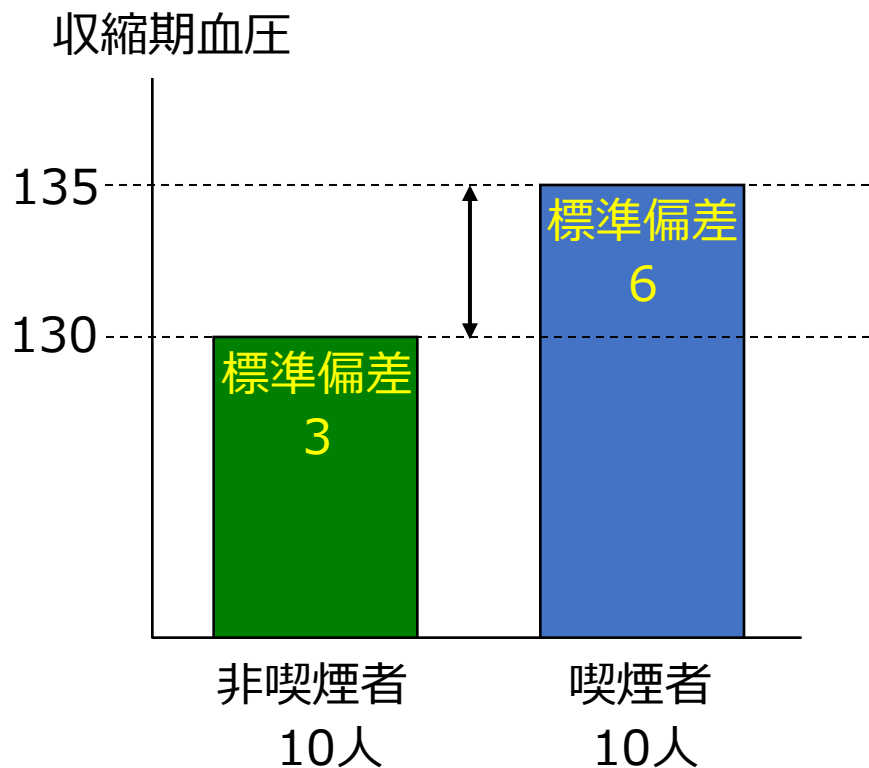


母集団の標準偏差に近づく

- 母集団の平均 = 標本の平均
- 母集団の標準偏差 = 標本の分散 /  $n$

= 標準偏差 /  $\sqrt{n}$   
→ 標準誤差(SE)

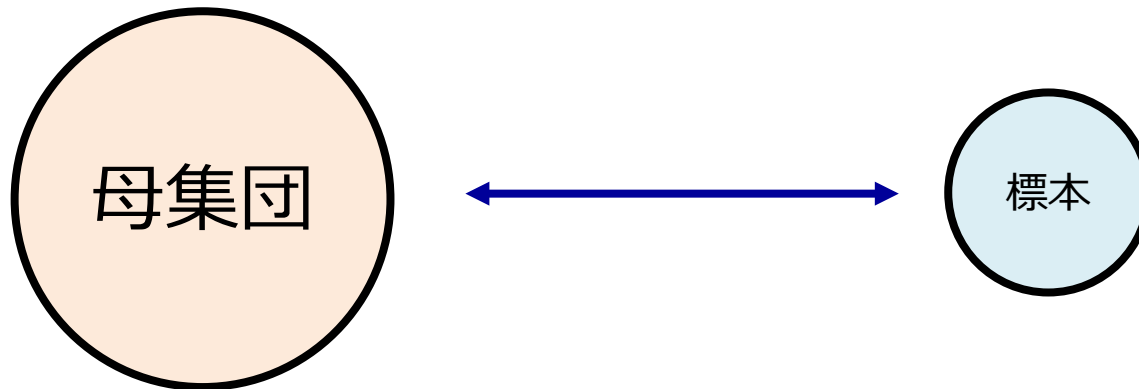
# 検定：標本を用い、母集団の特性に関する仮説を統計学的に検証すること



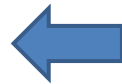
- 帰無仮説  
非喫煙者と喫煙者の血圧平均値の差5 mmHgは、差があるとはいえない(偶然)
- 対立仮説  
非喫煙者と喫煙者の血圧平均値の差5 mmHgは、差がある(偶然でない)
- この差が偶然に出現する確率  
= 0.030 < 5%
- 統計学的に有意な差がある

推定：標本が属する母集団の特性を、統計学的に推測すること → 区間推定

集団の特性を示す指標は・・・**平均と標準偏差**



- 血圧平均値の差 = 5 mmHg
- 平均値の差の標準偏差 = 2.121



- 血圧平均値の差 = 5 mmHg

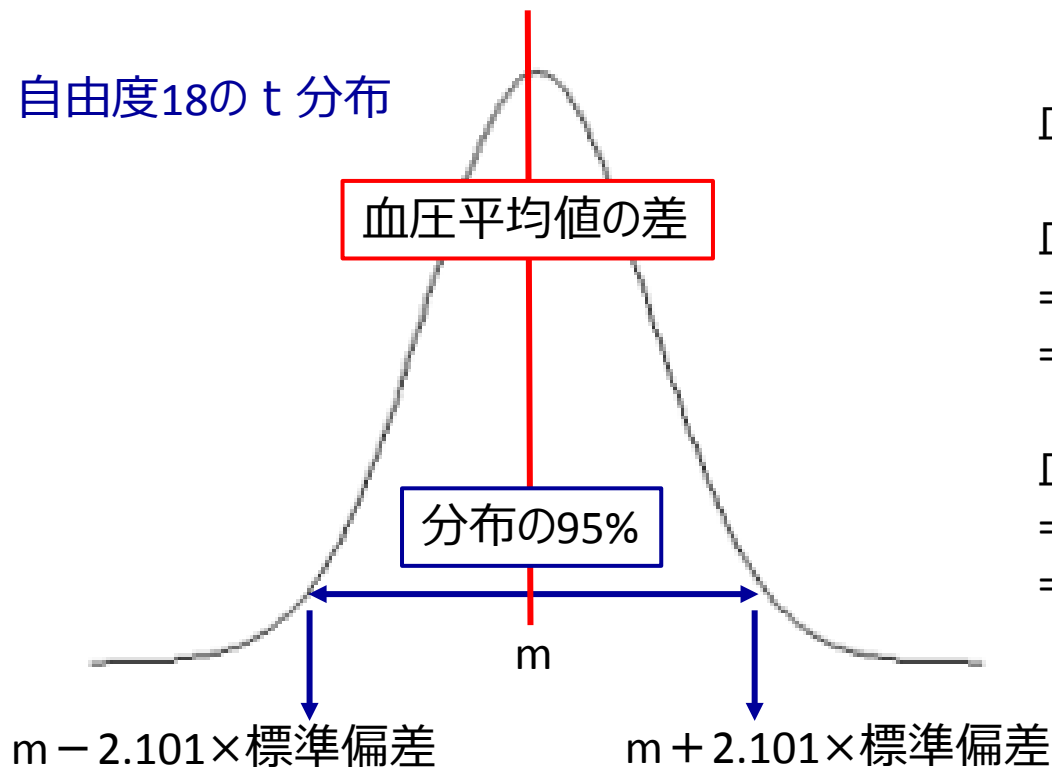
- 血圧平均値の差の標準誤差

$$= \sqrt{\frac{9}{10} + \frac{36}{10}} = \sqrt{\frac{45}{10}} = 2.121$$



# 母集団における非喫煙者と喫煙者の血圧平均値の差の区間推定

自由度18の t 分布



血圧平均値の差の95%信頼区間

$$\begin{aligned} & \text{血圧平均値の差} - 2.101 \times \text{標準偏差} \\ &= 5 - 2.101 \times 2.121 \\ &= 0.543 \end{aligned}$$

$$\begin{aligned} & \text{血圧平均値の差} + 2.101 \times \text{標準偏差} \\ &= 5 + 2.101 \times 2.121 \\ &= 9.457 \end{aligned}$$

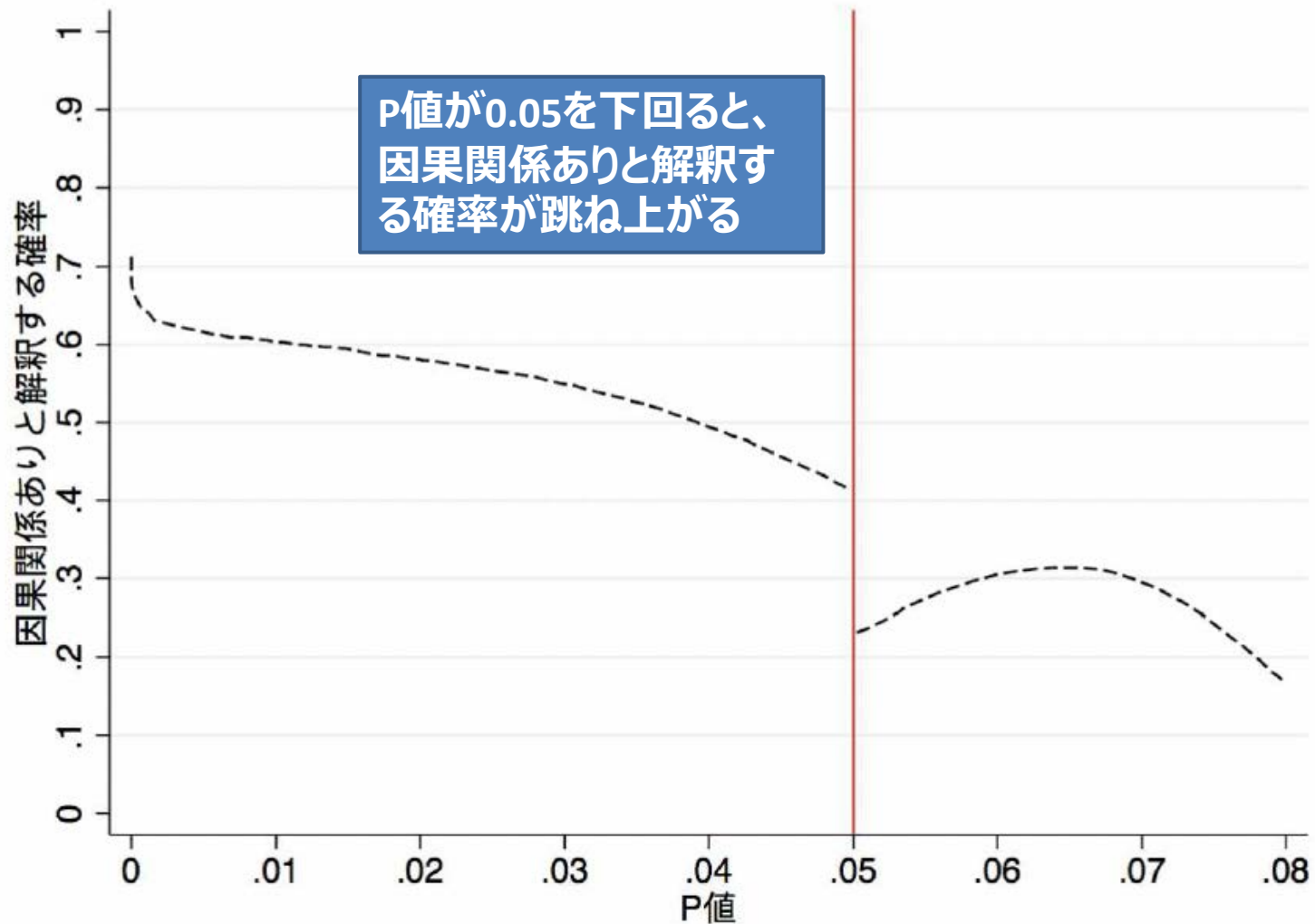
結論：

血圧平均値の差の95%信頼区間は 0.543 ~ 9.457 mmHgである  
(ゼロを含まない → 統計的有意)

**「神は  $p < 0.06$  を  $p < 0.05$  と等しく、  
そして同じくらい強く愛してくださる」**

Rosnow and Rosenthal (1989)

# P値と「因果関係あり」の解釈



# p hacking

## p hackingの具体例

1. 行った条件や測定した変数の一部しか報告しない
2. 参加者を少しずつ足しながら分析を行い、有意差に至ったところで止める
3. 様々な共変量を用いて分析を行い、有意になった組み合わせのみを報告する

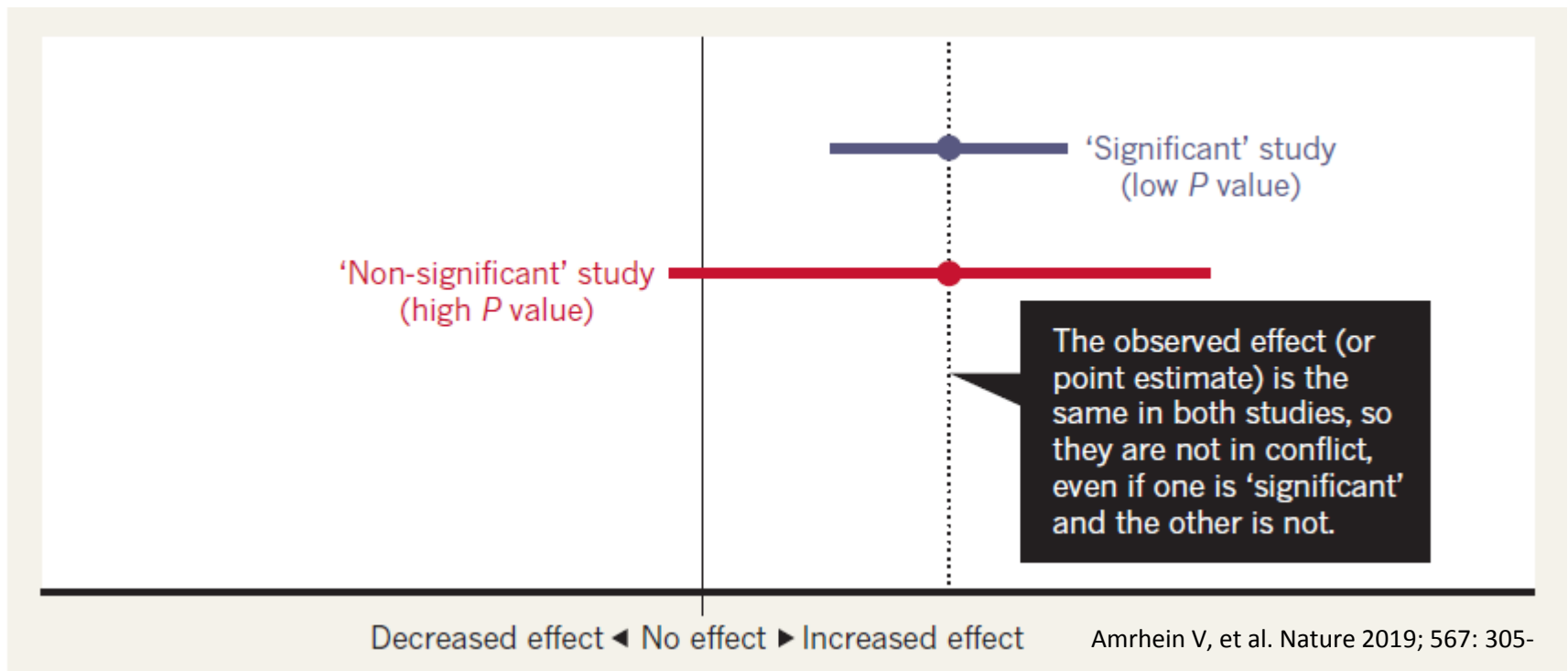
これら「研究者の自由度」を組み合わせることにより、少なくとも何らかの分析で有意差が見つかる可能性は...

**5% から約61% へ上昇**

# 統計的有意差なし ( $p \geq 0.05$ ) = データに差がない?

有意差なし... × データに差がなかった

○ データに差があることを示せなかった



同じ効果量だが上（青）は「有意差あり」、下（赤）は「有意差なし」

# 統計的有意性とp値に関するASA声明

ASA：米国統計学会

1. p値は「そのデータがある特定の統計モデルとどれくらい適合しないか」を示し得る
2. p値は調べている仮説が正しい確率や、データが偶然のみで得られた確率を測るものではない
3. 科学的結論及びビジネス・政策上の意思決定は「p値がある特定の値を超えたかどうか」のみに拠るべきではない
4. 適切な推論のためには、全てを報告する透明性が必要である
5. p値や統計的有意性は効果の大きさや結果の重要性を測るものではない
6. p値そのものだけではモデルや仮説に関するエビデンスの良い指標たり得ない

# どうすれば良いのか？

1. 効果量（平均値の差、オッズ比、ハザード比等）も必ず記載する。
2. P値（検定結果）のみならず、95%信頼区間（推定結果）を記載する。

✗ A群と比べ、B群にて有意な▲▲の増加を認めた（ $p=0.01$ ）

○ A群と比べ、B群にて有意な▲▲の増加を認めた（▲▲の両群差  $5.1 \pm 2.6$  mg/L、差の95%信頼区間  $3.2-9.0$ 、 $p=0.01$ ）

✗ A群を参照としたとき、B群にて有意なリスクの上昇を認めた（ $p=0.002$ ）

○ A群を参照としたとき、B群にて有意なリスクの上昇を認めた（オッズ比  $1.36$ 、95%信頼区間  $1.20-1.57$ 、 $p=0.002$ ）

# 後付け解析の「誤魔化し」

## 武田薬品 ARB・プロプレスの臨床研究CASE-Jで組織的関与 糖尿病の新規発症で有利な結果得る

公開日時 2014/06/23 03:52



共有



Twitter



印刷



ARB・プロプレス（一般名：カンデサルタン）の臨床研究「CASE-J」の臨床研究不正をめぐり、武田薬品は6月20日記者会見を開き、同剤の付加価値最大化を目的に、研究の企画段階から学会発表、論文作成まで、一貫して組織的かつ継続的な関与があったとの第三者機関の調査結果を公表した。奨学寄附金による総額37億5000万円の資金提供や、複数の社員による労務提供により、京都大学EBMセンターの運営を含めてサポートされており、「医師主導臨床試験である実質的なスポンサーであった」と結論付けた。副次評価項目である糖尿病の新規発症については、当初の解析方法では“有意差なし”だったが、同社が研究者に定義を変更するよう働きかけ、結果としてプロプレスに有利な結果を引き出していたこともわかった。試験結果の公平性、客観性も疑問視されるとこ

- 後付け（post-hoc）解析は悪くない（エビデンスは低いが）
- 後付け解析時に定義を変更し、有意差を見つけ、あたかも事前から計画された解析の様に見せる → 実質的な改ざん



# Take-home message

- 統計解析とは、統計学の手法を用いて、集団の特性について客観的な解釈を与えることである。
- 統計学的検定は「背理法」を用いる。
- 確率変数を求め、確率分布での位置を知ることにより検定を行う。
- 「推定」によって、母集団の特徴を推測する。
- $p < 0.05$ を過信してはいけない！