入門 統計学　第6章

仮説検定と検出力

『入門 統計学 第2版　ー検定から多変量解析・実験計画法・ベイズ統計学までー』(オーム社)

※注:本書を購入された方へのサービスですので，教科書指定(参考図書は不可)していない授業での使用はお控えください。

（本章と次章で）仮説検定を学ぶと何を検証できる？

病気かどうかの判断

薬の効き目の有無

機械導入効果の有無

特定の値と標本平均を比較したり，処理前後の標本平均を比較することで，その背後にある母集団の平にも差があるか否かを判定できる

注：平均だけでなく分散なども比較できる

仮説検定の長所

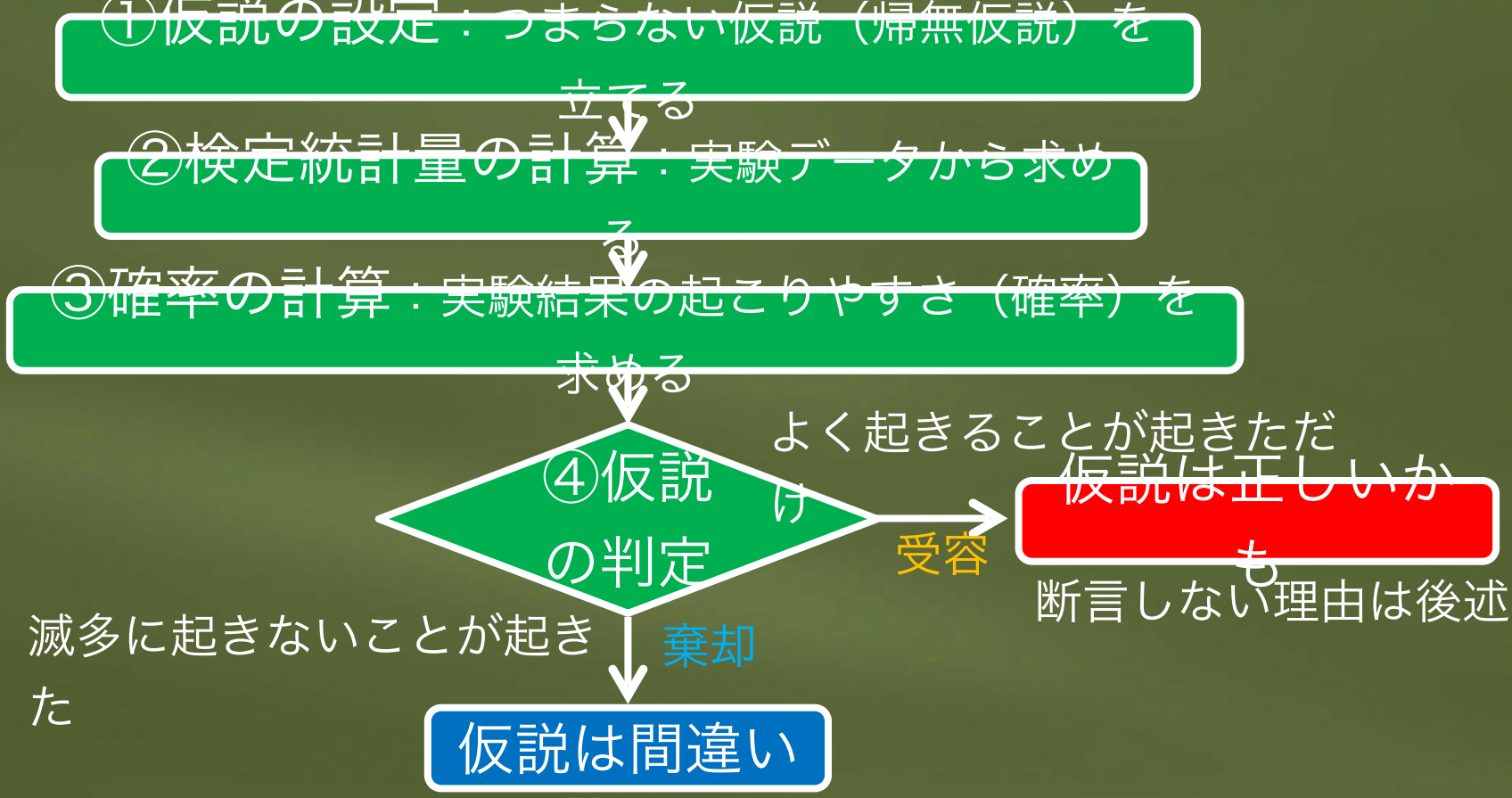
- 点推定や区間推定は，あくまで推定量や誤差を数値で表すだけ　→　統計学を学んでない人にはわかりにくい

- 仮説検定ならば…「処理の効果がある！」とはっきりいえるため，誰にでもわかりやすく伝えられる

仮説検定と検出力

- 仮説検定：母集団に対して主張したくない仮説（処理の効果がない等）を立て，その仮説が間違いであることを，実験結果の起き難ささから判断する。
- 検出力：母集団の特性に差があるときに，ちゃんと“差がある”と判断できる能力。検定の性能を表す指標で近年重視されている。

6.1　検定の概要


①仮説の設定：つまらない仮説（帰無仮説）を立てる
②検定統計量の計算：実験データから求める
③確率の計算：実験結果の起こりやすさ（確率）を求める
④仮説の判定
よく起きることが起きただけ
受容
仮説は正しいかも
断言しない理由は後述
滅多に起きないことが起きた
棄却
仮説は間違い

仮説検定の大まかな手順

手順①：仮説の設定

母集団における未知の事実を説明するため，とりあえず正しいものと仮定しておく

道筋を立てる

手順②：検定統計量の計算

観測されたデータのままでは検定できることは少ないので，目的に沿った検定のための統計量を求める（ただし，本章ではこの手順をスキップできる基本的な検定から解説する）

仮説検定の大まかな手順　続き

手順③：確率の計算

仮説が正しい下での実験結果の起こりやすさ（確率）を計算。実際には，確率計算は難しいので，仮説の分布のどれぐらい端の方に位置しているのかで仮説の是非を判定

手順④：仮説の判定

滅多に起きないことが起きたといえるぐらい確率が低いならば，そもそも仮説が間違いだったと判断。そうでなければ判定を保留する（仮説が正しかったとはいわない）

6.2　仮説の設定

（手順①の説明から）

- 母集団に立てる仮説には，帰無仮説（H_0）と対立仮説（H_1）の2つがある
- 基本となるのは帰無仮説で，帰無仮説が棄却されたときに採択する仮説が対立仮説
- 帰無仮説は「主張したい内容とは逆の，つまらない内容」とする

  例：処理効果が無い（2群の母平均に差は無い）

  →なぜ主張したい内容としないのだろうか？

帰無仮説の内容

例：殺虫剤の効果を検定する場合

主張したいのは「使用効果がある（使用前後で害虫数に差がある）」という内容だが…

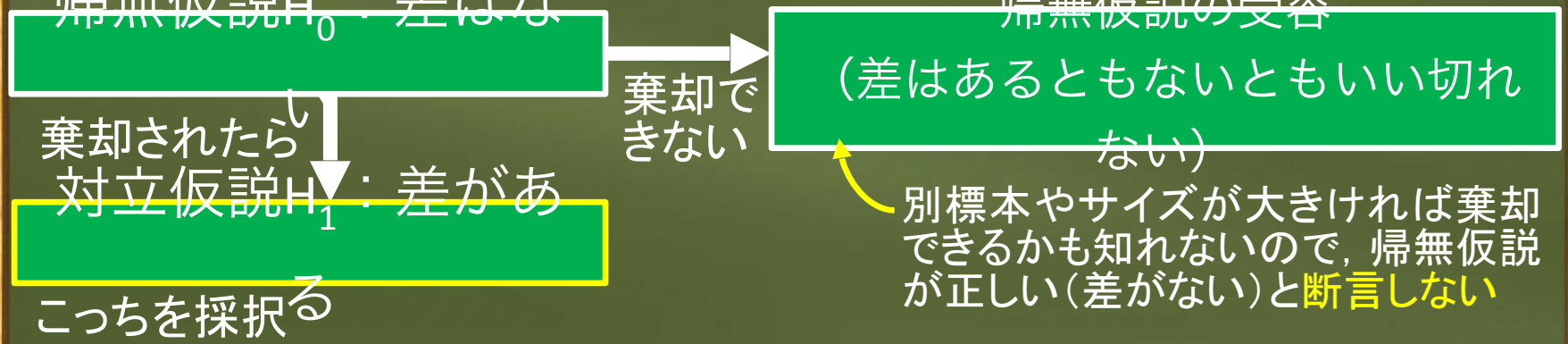
→それを検証するためには「＊匹減る」と具体的にわかってなければならない（無限に設定できる）

→使用効果がない（全く減らない）としておけば，それと矛盾した結果が観測されれば，効果があるといえる（背理法）

対立仮説の内容

- 対立仮説は，帰無仮説の逆の（主張したい）内容とし，帰無仮説が棄却されたときに採用する
- 対立仮説は何のために立てる？

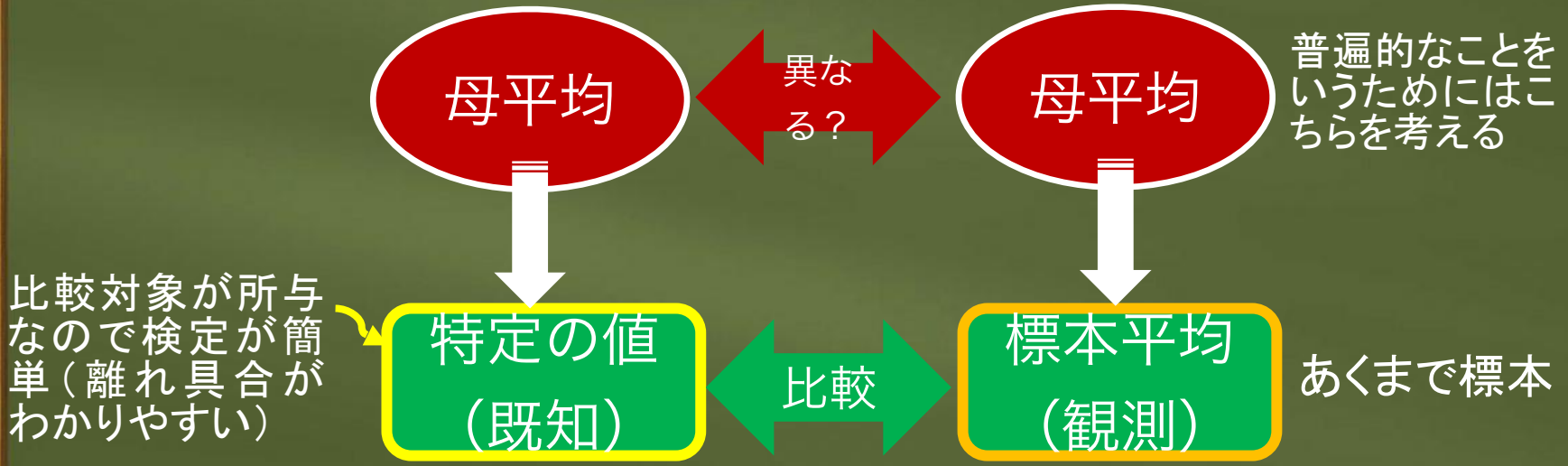
→検出力や第二種の過誤確率β（後述）のため



6.3　（1標本の）母平均の検定

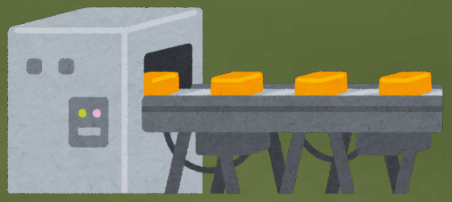
検定統計量が不要（容易）な検定から学びます

- 特定の既知の値と観測された標本平均とを比較し，その差が偶然（誤差）の範囲内なのか，母集団においても差があるといえるほど大きい差なのかを判定する



母平均の検定の適用例

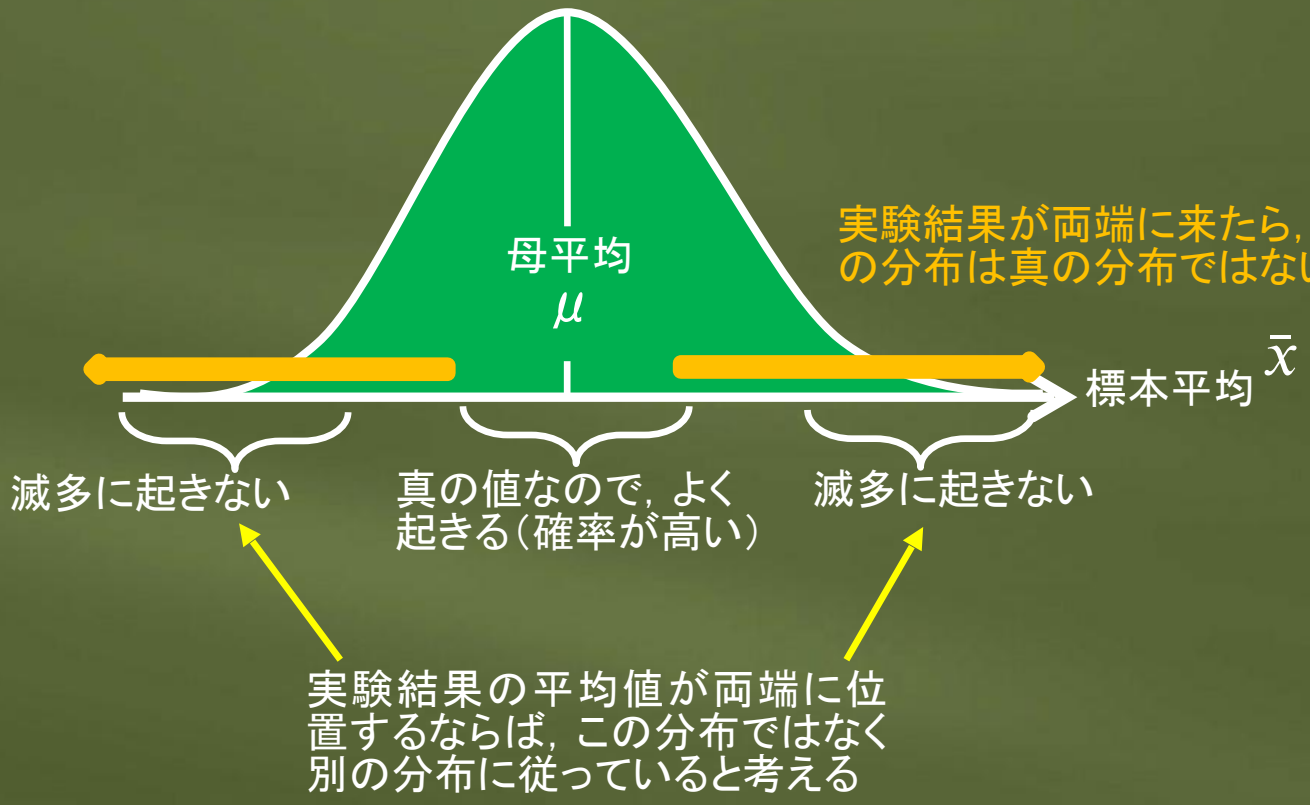
病気の疑いのあるグループの平均と正常値とを比較して，病気か健康かを判定



製造している商品の容量が袋に表記されている基準値と異なっていないかを確認

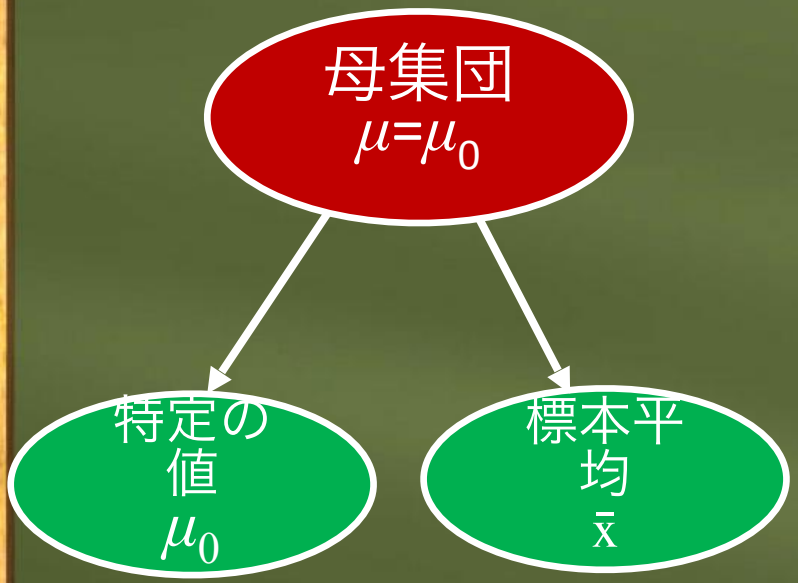
調査した農家の平均年齢が真の値からズレていないかを検証

検定の原理


母平均
μ
実験結果が両端に来たら，この分布は真の分布ではない
標本平均 $\bar{x}$
滅多に起きない
真の値なので，よく起きる（確率が高い）
滅多に起きない
実験結果の平均値が両端に位置するならば，この分布ではなく別の分布に従っていると考える

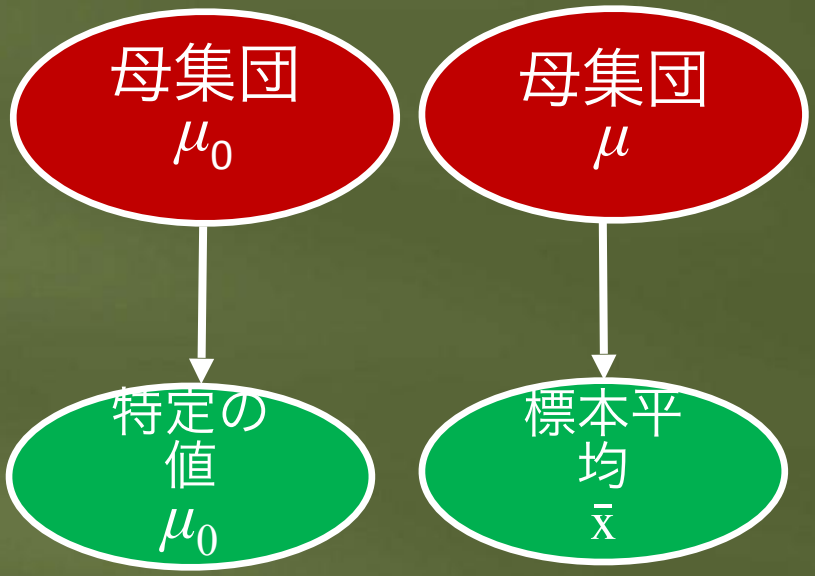
ステップ①：２つの仮説を考える

帰無仮説 $H_0: \mu = \mu_0$



同じ母集団から抽出
（μ_0は$\bar{x}$の真の値）

対立仮説 $H_1: \mu \neq \mu_0$



異なる母集団から抽出
（μ_0は$\bar{x}$の真の値ではない）

対立仮説の内容は3種類

パターン①：$\mu \neq \mu_0$ →どちらが大きくなるかは不明

（とにかく差があることに興味がある）

← 両側検定

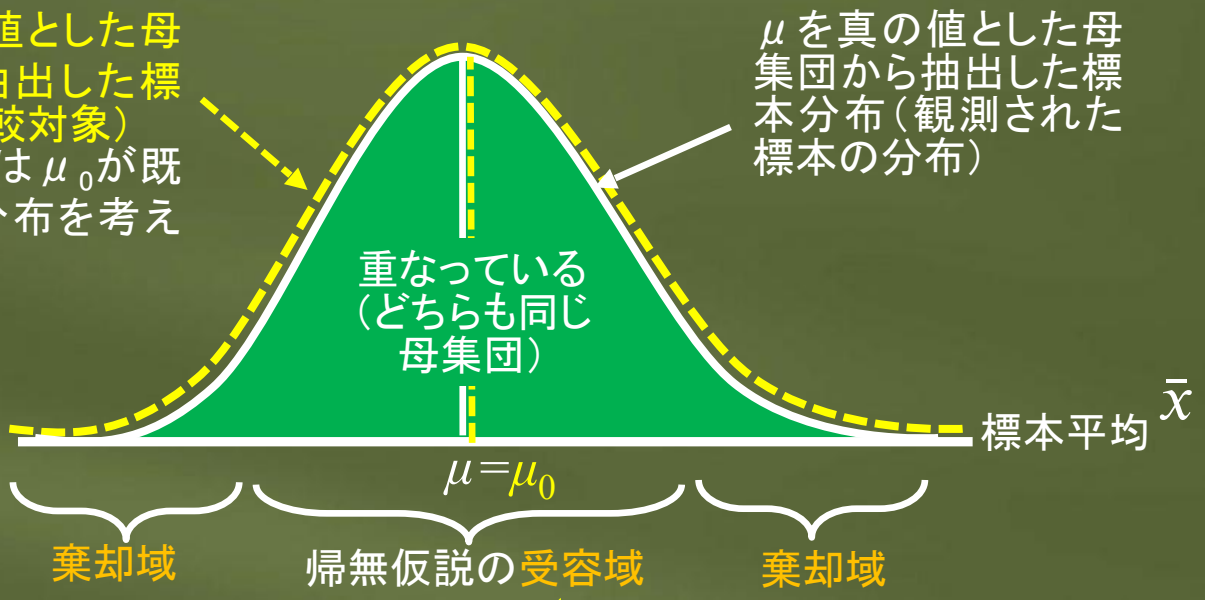
パターン②：$\mu < \mu_0$ →比較値μ_0が標本の母平均μよりも大きいと仮定できる（あるいは比較値の方が小さいことには興味がない）

パターン③：$\mu > \mu_0$ →比較値μ_0が標本の母平均μよりも小さいと仮定できる（あるいは比較値の方が大きいこ

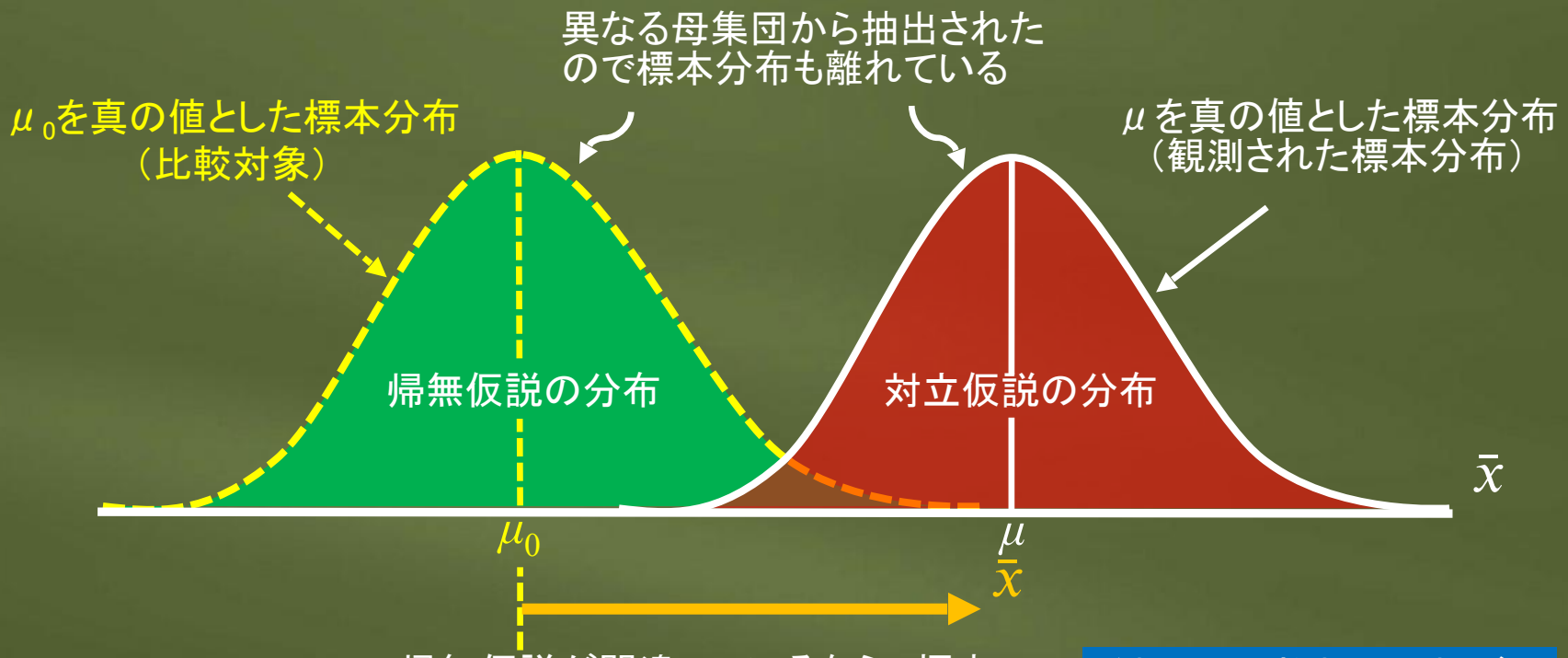
片側検定

片側検定の方が差があることをいいやすいが，普通はどちらが大きいかはわからないため，両側検定を基本とする

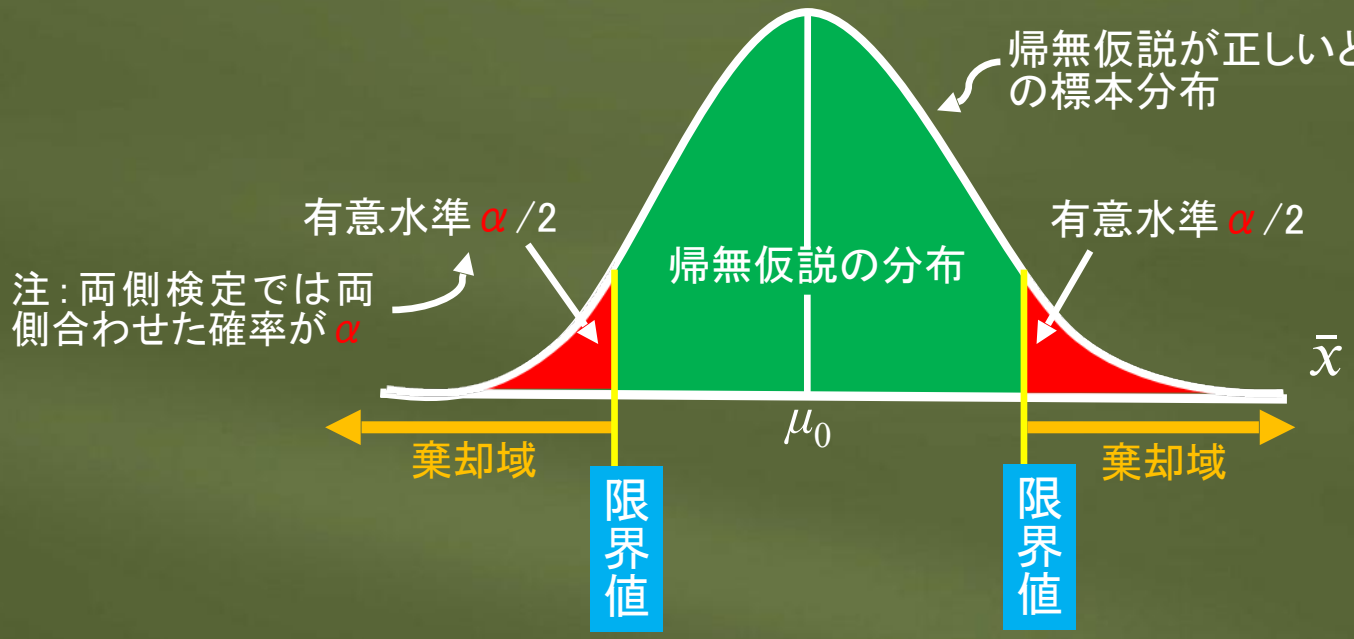
ステップ②：帰無仮説が正しい場合の標本の位置は？



ステップ②の続き：
帰無仮説が間違っている場合は？

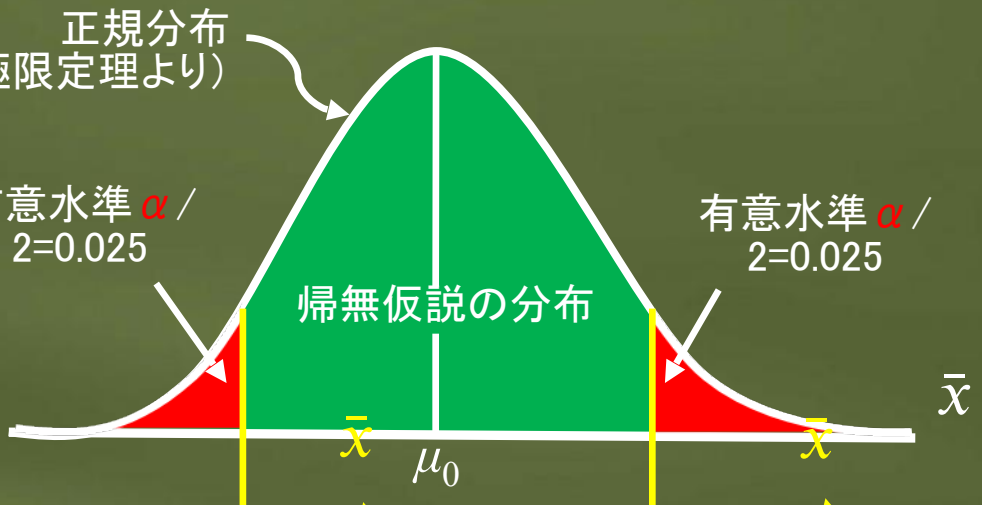


ステップ③：どこから帰無仮説を棄却する？



任意の有意水準（一般的に $\alpha = 5\%$ とする）に対応する値を基準（限界値）として，帰無仮説の是非を判定
（片側検定では片方だけで5%とするので棄却域が広がる）

ステップ④：限界値を求めて検定統計量（観測された標本平均）と比較する



母標準誤差（標本平均$\bar{x}$の母標準偏差）

$$\mu_0 \pm 1.96\sigma_{\bar{x}}$$

z分布表から上側確率が0.025になるz値を読み取る（区間推定と同じ）

受容

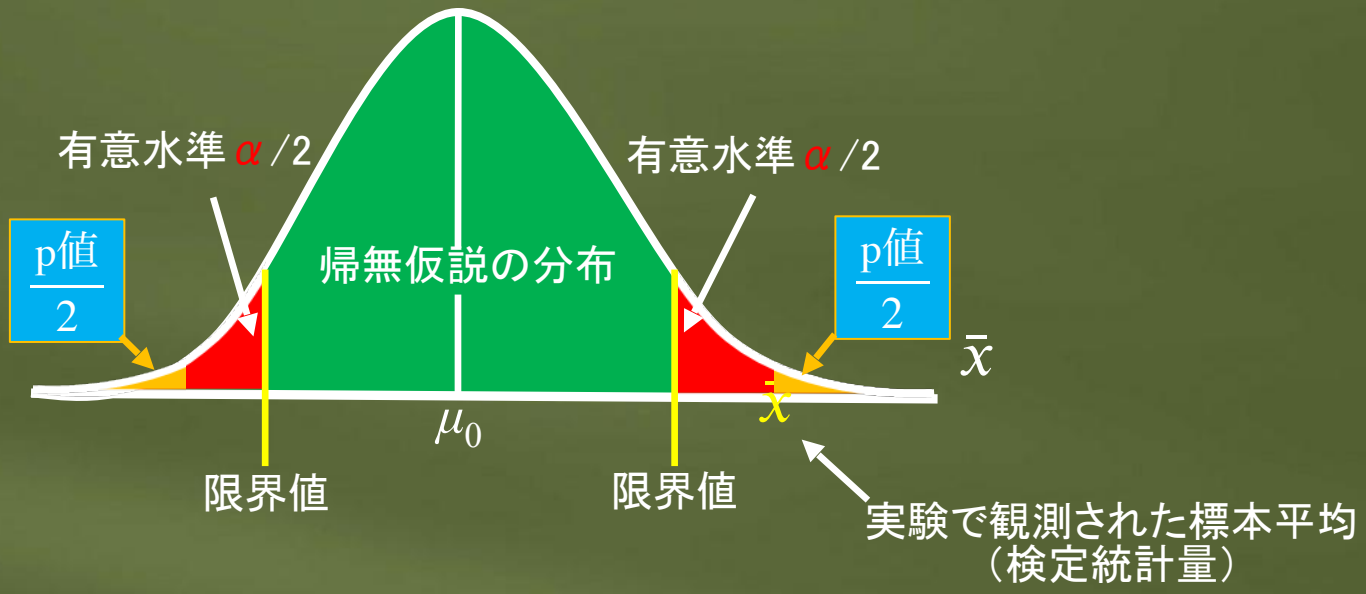
これぐらいの差ならば偶然に観測されることがある（帰無仮説は間違いとはいえない）

棄却

滅多に起きないぐらいの大きな差なので，別の分布に従う（帰無仮説は間違い）→有意差がある

ステップ④の補足：p値を使った検定法

（ソフトウェアが必要）



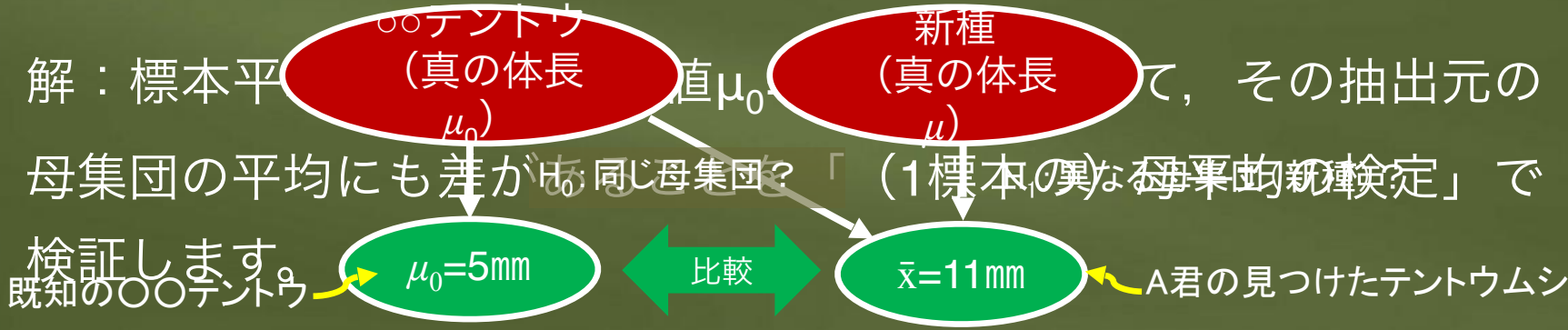
p値：帰無仮説の下で，実験結果（検定統計量）よりも極端な値が観測される確率

→p値を有意水準 α と比較して“p＜α”ならば帰無仮説を棄却（有意）

例　題

A君が大きなテントウムシ（平均体長$\bar{x}$=11㎜）を数匹採集した。このテントウムシは新種だろうか？それとも既知の種（○○テントウ）だろうか？ただし，既知の種の真の体長，つまり母平均μ_0=5㎜，母標準誤差$\sigma_{\bar{x}}$=2㎜とし，種は体長で決まるとする。

解：標本平均値$\bar{x}$と母平均値μ_0について，その抽出元の母集団の平均にも差がH_0：同じ母集団？「（1標本の）母平均の検定」で検証します。



例題　続き

手順①：仮説の設定

H_0：μ（$\bar{x}$=11mm）＝μ_0（5mm）→新種でない（同じ母集団から抽出）

H_1：μ（$\bar{x}$=11mm）≠μ_0（5mm）→新種である（異なる母集団から抽出）

手順②の検定統計量の計算は不要

手順③の平均体長が11mm以上の〇〇テントウが見つかる確率の計算も難しいので，その珍しさを帰無仮説の分布の位置で判断する（手順④）

手順④：仮説の判定

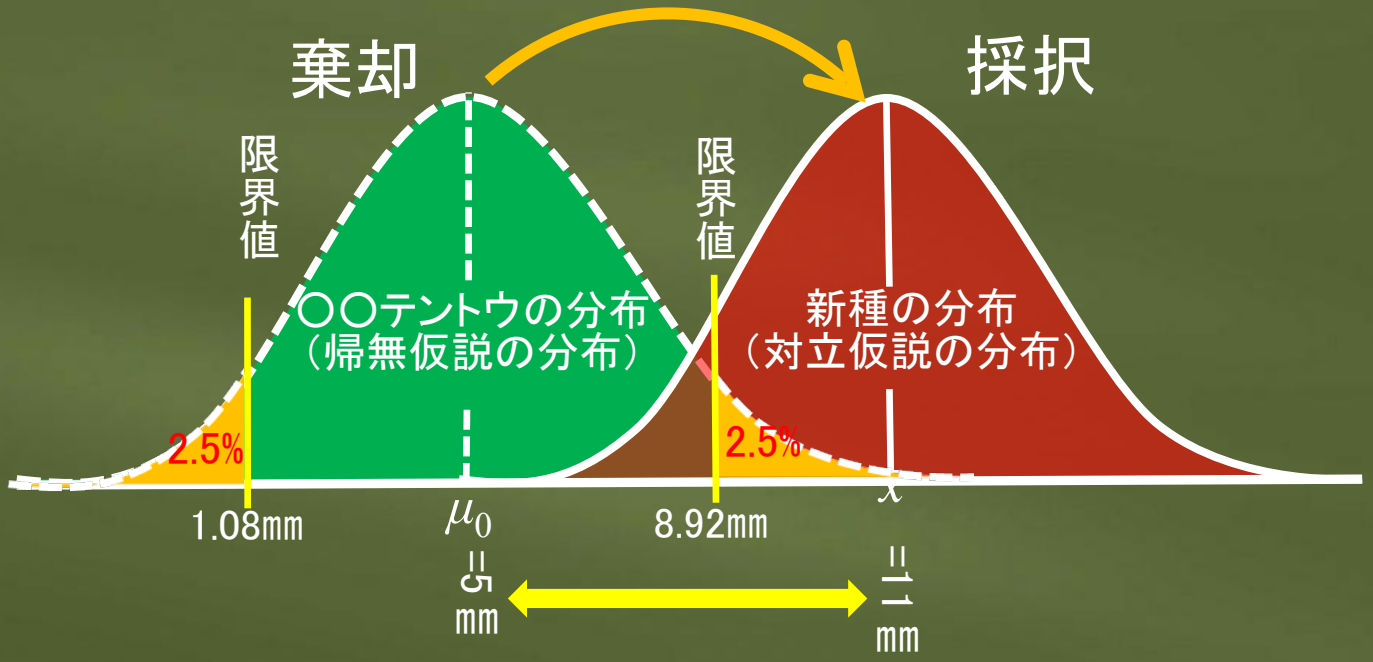
設定した有意水準（今回は両側5%とする）に対応する限界値を求め，それと標本平均11mmを比較し，帰無仮説の是非を判定する。

限界値は，（比較対象値）±（有意水準に対応するz値）×母標準誤差なので，5±1.96×2から，下限値が1.08mm，上限値が8.92mmとなる。

∴　標本平均＞上限値なので，帰無仮説は棄却してよいだろう（次に図掲）。

例題　図解

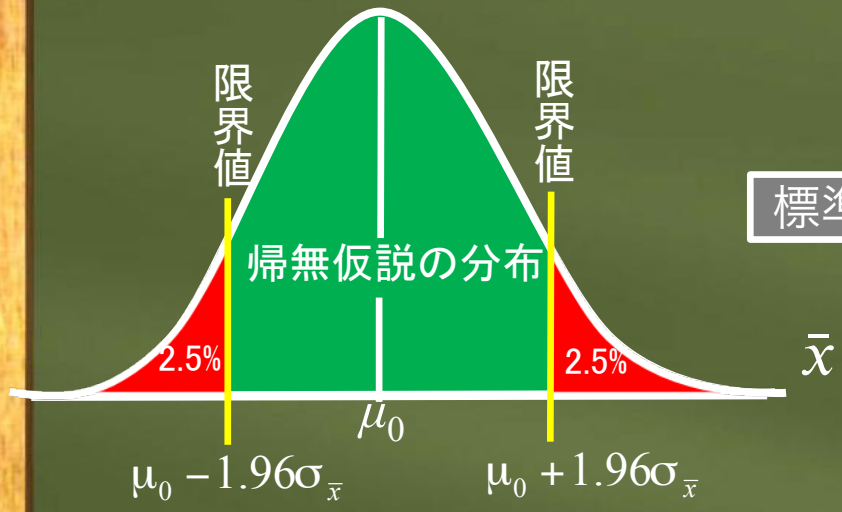
A君の採集したテントウは新種の分布に従っていると考えた方が合理的



6mmという差は偶然とはいえない大きさ

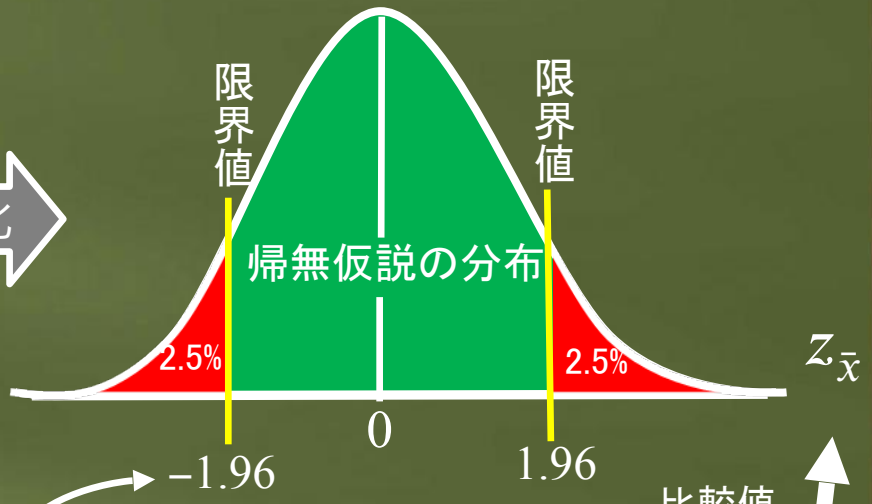
6.4　標準正規分布による母平均の検定（z検定）

正規分布による仮説検定



標準化

標準正規(z)分布による仮説検定



z検定では，限界値が単純になり，どんな対象でも同じ値(統計量を計算するので，検定が容易になるわけではない)

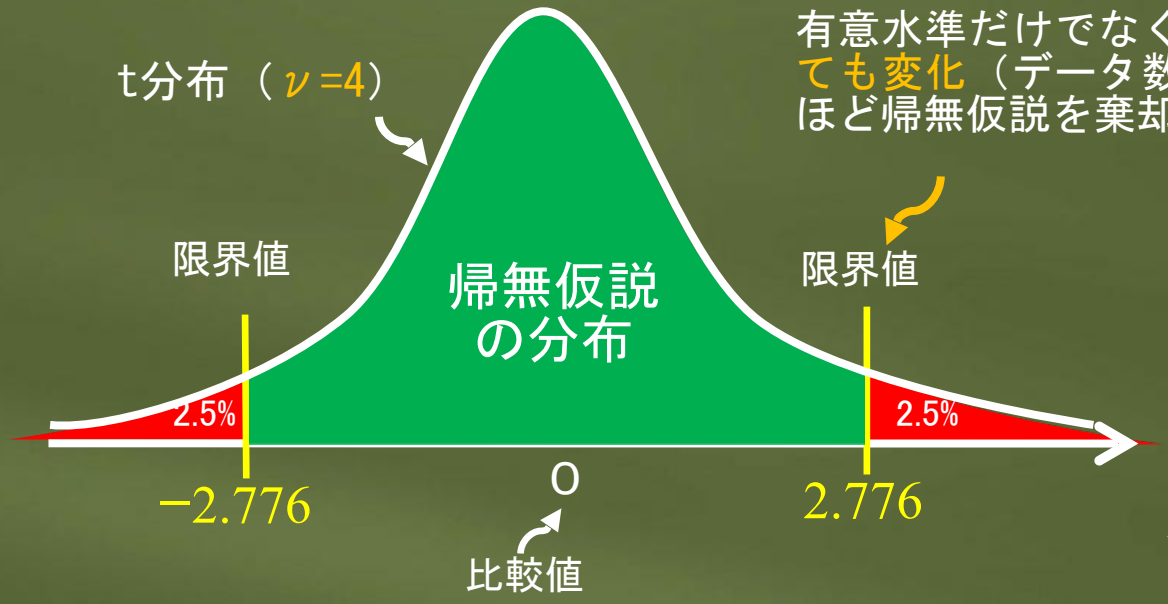
z検定の統計量 $z_{\bar{x}} = \dfrac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \dfrac{\bar{x} - \mu_0}{\sqrt{\dfrac{\sigma^2}{n}}}$

（母平均の検定）

比較値

母標準誤差

t 分布による母平均の検定（t 検定）



正規分布による検定やz検定は母分散が既知の必要があったが，t検定は標本データのみでOK

t 検定の統計量（母平均の検定）

$$t_{\bar{x}} = \frac{\bar{x} - \mu_0}{\hat{\sigma}_{\bar{x}}} = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\hat{\sigma}^2}{n}}}$$

不偏標準誤差

t 検定の例題

痛風の疑いのある5名に血液検査をした結果，7，10，12，13，13という尿酸値（mg/dL）を観測した。このグループは痛風に罹患しているといえるか？　ただし，正常値の上限は7.0とする。

解：標本平均$\bar{x}$は11mg/dLなので，正常値μ_0に比べて4mg/dL高い。これが偶然の範囲内か否かを母平均の検定で確かめる。ただし，母分散は不明なのでt分布を使うことになる。（痛風は尿酸値が高い場合なので上片方側で良いが，基本的な両側検定を実施しておく）

手順①：仮説の設定

同じ母集団から抽出

- 帰無仮説H_0：μ（$\bar{x}$=11mg）＝μ_0（7mg）→正常である（痛風ではない）
- 対立仮説H_1：μ（$\bar{x}$=11mg）≠μ_0（7mg）→正常ではない（痛風である）

事例の続き

手順②：検定統計量の計算　$t_{\bar{x}} = \dfrac{\bar{x} - \mu_0}{\hat{\sigma}_{\bar{x}}} = \dfrac{\bar{x} - \mu_0}{\sqrt{\dfrac{\hat{\sigma}^2}{n}}} = \dfrac{11 - 7}{\sqrt{\dfrac{6.51}{5}}} = 3.51$

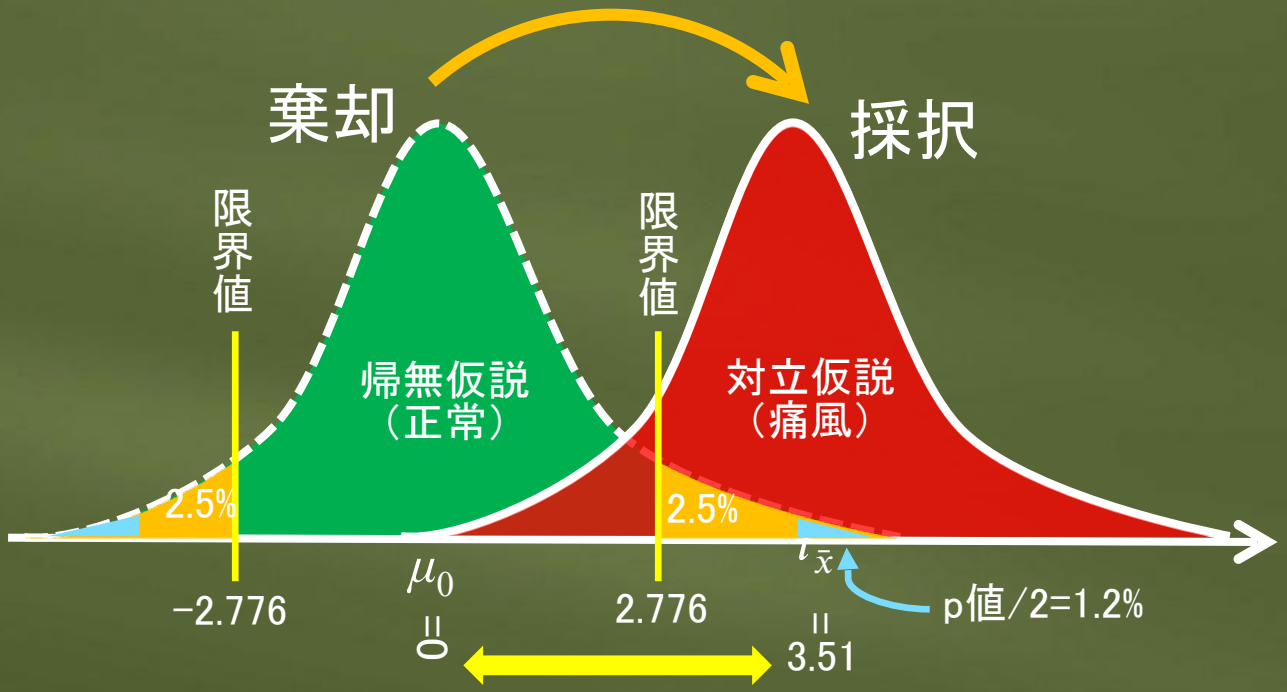
不偏分散

手順③：確率の計算（ソフトがなければ手順④にスキップ）
帰無仮説が正しいとした下で4mg以上の差が観測される確率（p値）はExcel関数を使えばT.DIST.2T(3.51, 4)で0.0247と計算できる。p値は5%よりも小さいので，正常という帰無仮説は棄却され，痛風と判定できる。

手順④：仮説の判定（限界値と比較する方法）
ソフトが使えない場合には，t分布表から読み取った限界値（上側2.5%で自由度4のt値は“2.776”）と検定統計量とを比較する。t値は限界値よりも大きいので5%有意水準で帰無仮説は棄却されることがわかる（次掲）。

事例の図解

このグループは痛風の分布に従っていると考えた方が合理的



7mgという差は偶然とはいえない大きさ

6.5　検出力分析
（近年重要になりつつある分野）

①実施した検定の能力（検出力）はどのぐらいだったのかを示すのが重要になってきている

②処理効果（2群の差）そのものの大きさ（効果量）を推定することも重要になってきている

→こうした検出力や効果量を求めて検定の良し悪しを評価したり，望ましい標本サイズを事前に決める分野を検出力分析と総称

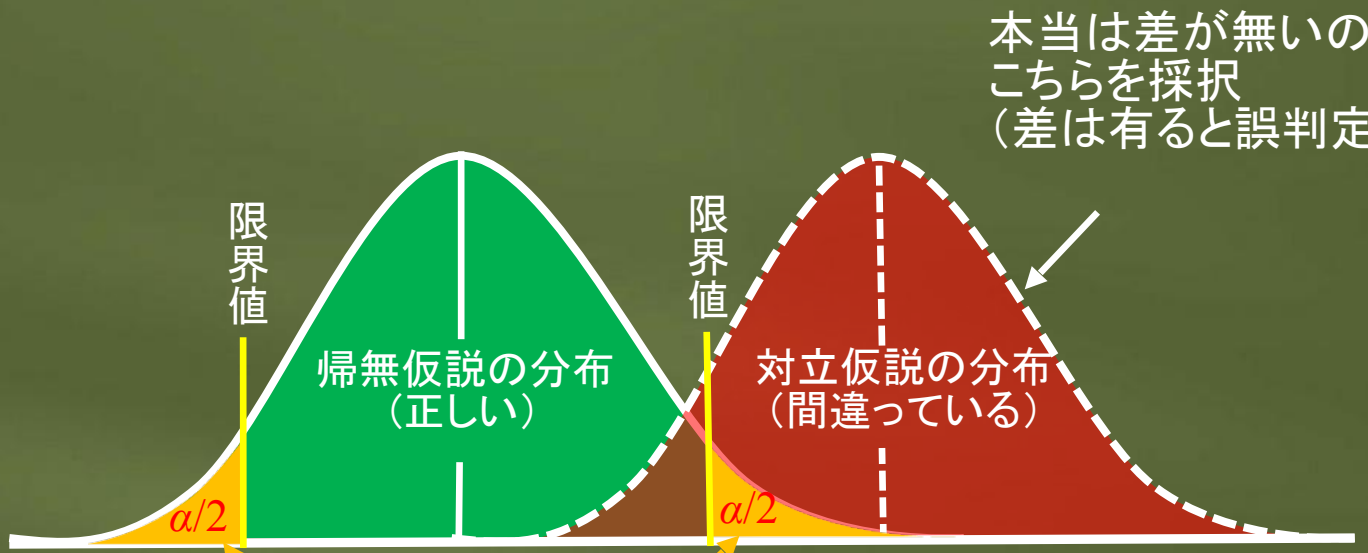
→検出力分析の基本は，検定における2種類の間違い（統計的過誤）なので，まずはこれから解説

統計的過誤は2種類

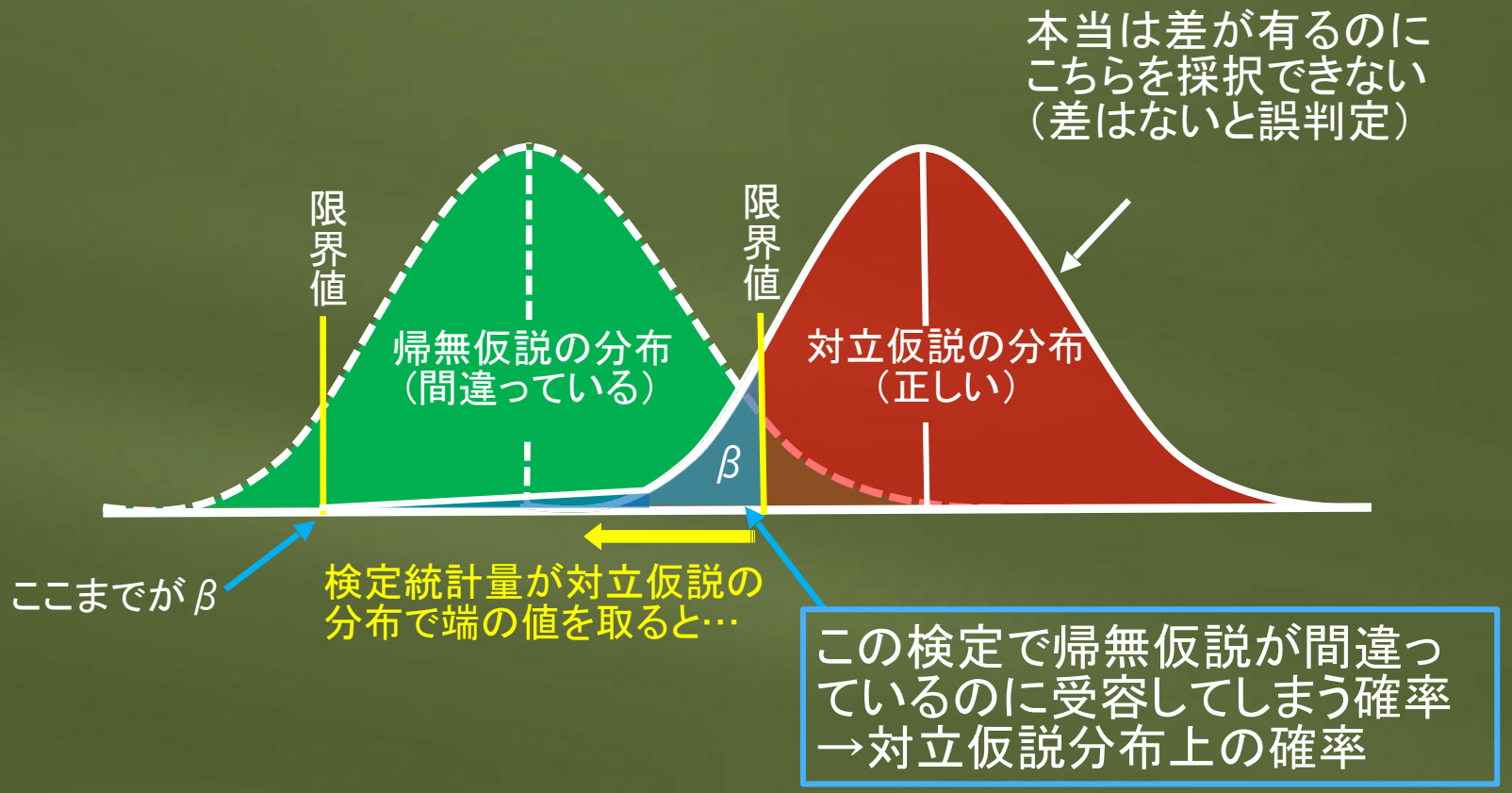
- 標本を用いている以上，仮説検定で間違った判断（統計的過誤）を下してしまうこともある
- 第一種の過誤：処理効果（群間で差）が無いのに有ると判断してしまう間違い→帰無仮説が真なのに棄却してしまう過ち（確率はαで表す）
- 第二種の過誤：処理効果（差）が有るのに無いと判断してしまう間違い→帰無仮説が偽なのに棄却しない（真の対立仮説を採択し損なう）過ち（β）

それぞれを図解すると…

第一種の過誤（Type Ⅰ error）



第二種の過誤（Type Ⅱ error）



第一種と第二種の過誤では どちらが重大（致命的）？

- 例えば，効果が無い薬を売り出した場合（第一種過誤）と，効果が有る薬を開発したのに売り出せない場合（第二種過誤）では，どちらが製薬会社にとって致命的なダメージとなるか？

  →第一種の過誤の方が致命的な事が多い

  →その検定で許容できる第一種の過誤を犯す確率α（有意水準）を小さく設定するのが一般的

  →α=0.05は，実験と検定を100回実施しても5回までしか過誤を許さないということ

（あまりないですが…）

第二種の過誤の方が致命的な事例

- 一旦，効果（差）を見逃してしまったら，取り返しの付かない場合

→下例で正しい対立仮説を採択し損ねたら？

事例①：環境保護政策の効果の検定（H_0：効果無し，H_1：効果有り）→環境は一度破壊されたら二度と戻らない

事例②：食品添加物の毒性の検定（H_0：毒性無し，H_1：毒性有り）→毒の入っ

第二種の過誤を抑えた検定はどうやる？

- 有意水準αのように，その検定で許容できる第二種の過誤確率βを事前に設定すれば良い？

×できない

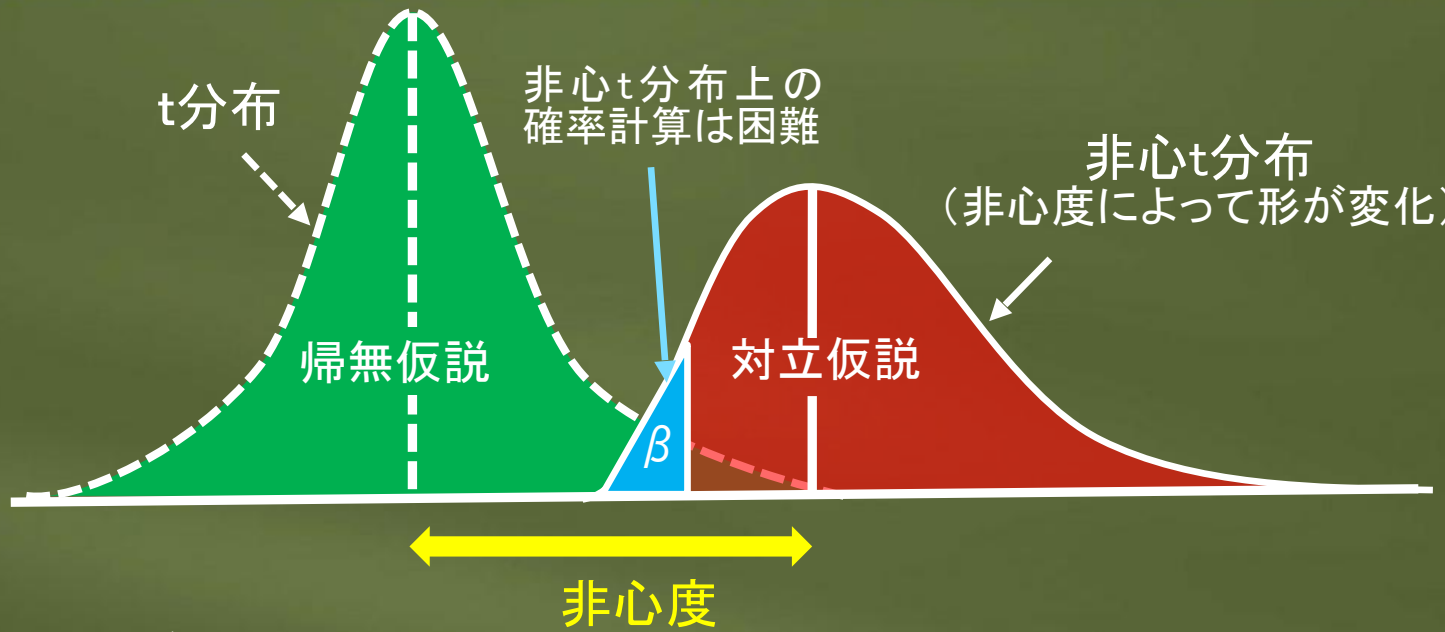
理由：対立仮説の分布の確率計算が難しいから

→帰無仮説分布からの離れ具合（非心度）によって対立仮説分布の形状が変わる（図を次掲）

→非心度は実質的な差（効果量）と標本サイズから決まるため，事前（検定前）に計算できない

→対立仮説を判定するためのβに対応する限界値は不明（有意水準の限界値を便宜的に共有する）

対立仮説の分布（非心分布）



（標本サイズと効果量によって決まり，zやt検定では検定統計量そのもの）

実質的な差の大きさなので事前には不明

注：自由度が関係ない（標準）正規分布の場合，対立仮説の分布でも形は不変

効果量

（注：検定によって計算式は色々ある）

- 母集団が持つ処理効果そのもの（実質的な差）の大きさ

母平均の検定の効果量

$$d = \frac{\mu - \mu_0}{\sigma}$$

（σ：未知）

← 標本の母平均 μ と比較値 μ_0 の差が母標準偏差 σ（標準誤差ではないので標本サイズとは無関係）の何倍なのかを示す

→観測された標本から検定後に推定できる

母平均の検定の推定効果量

$$\hat{d} = \frac{\bar{x} - \mu_0}{\hat{\sigma}} \quad or \quad \frac{t}{\sqrt{n}}$$

（$\hat{\sigma}$：不偏標準偏差、t：検定統計量）

← 推定した効果量から非心度を推定すれば，対立仮説分布の確率 β も計算できる

効果量を検定結果に加えよう！

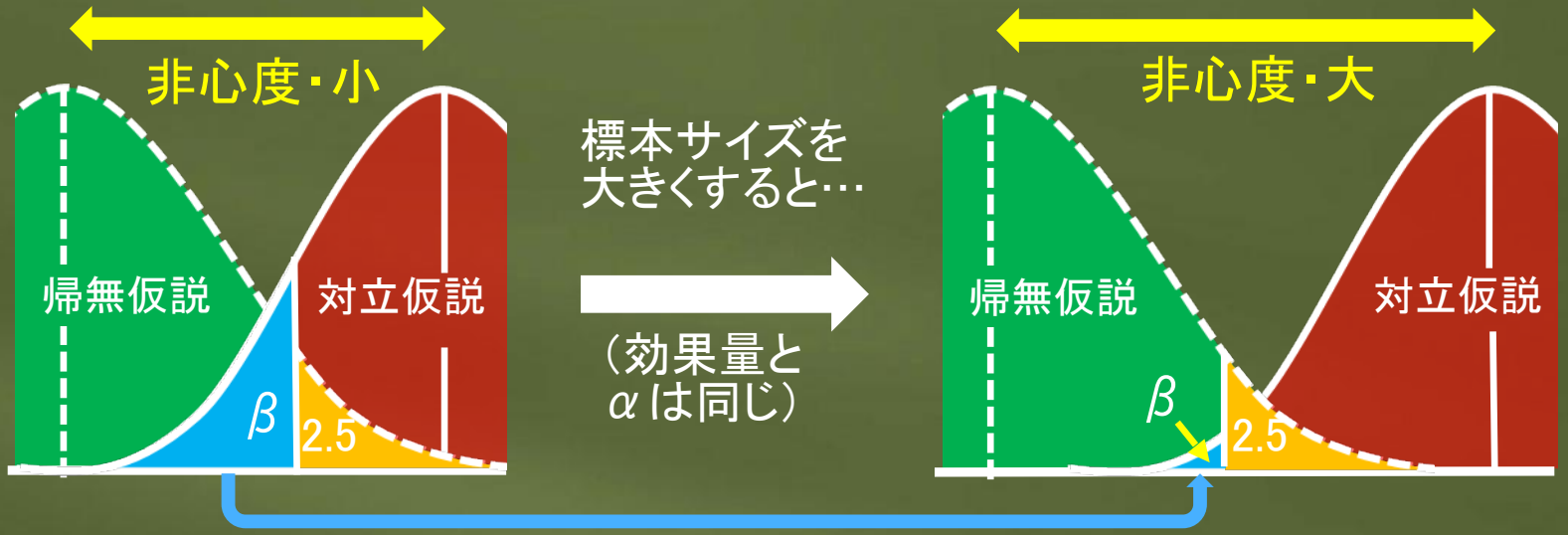
- 検定は標本サイズに影響を受けるが，効果量は標本サイズと関係ない

→検定結果が有意でなくても効果量が大きいこともある（逆に，有意でも効果量が小さいこともある）

→検定統計量を差（効果）の大きさと勘違いしたり，有意でなかったときに処理が全く無意味と思われないように，推定効果量も書いておく

第二種の過誤が致命的な場合の対処法に戻って…

大きな標本サイズでβを抑える

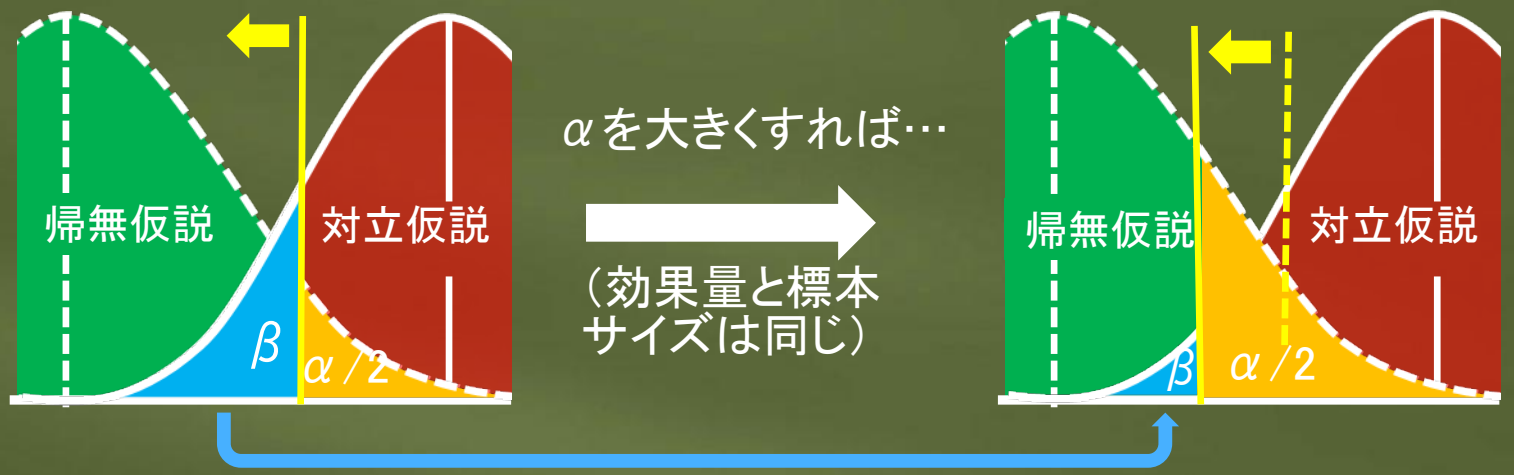


標本サイズの大きな実験を計画することで，β を小さくできる
（任意の α の下で望ましい β になるような，おおよその標本サイズを逆算して実験を計画すればよい→ソフトウェアを使った標本サイズの決め方を最後に解説）

第二種の過誤を抑えるもう一つの裏技

αを大きくしてβを小さくする

αとβで共用している限界値(の絶対値)を小さくする

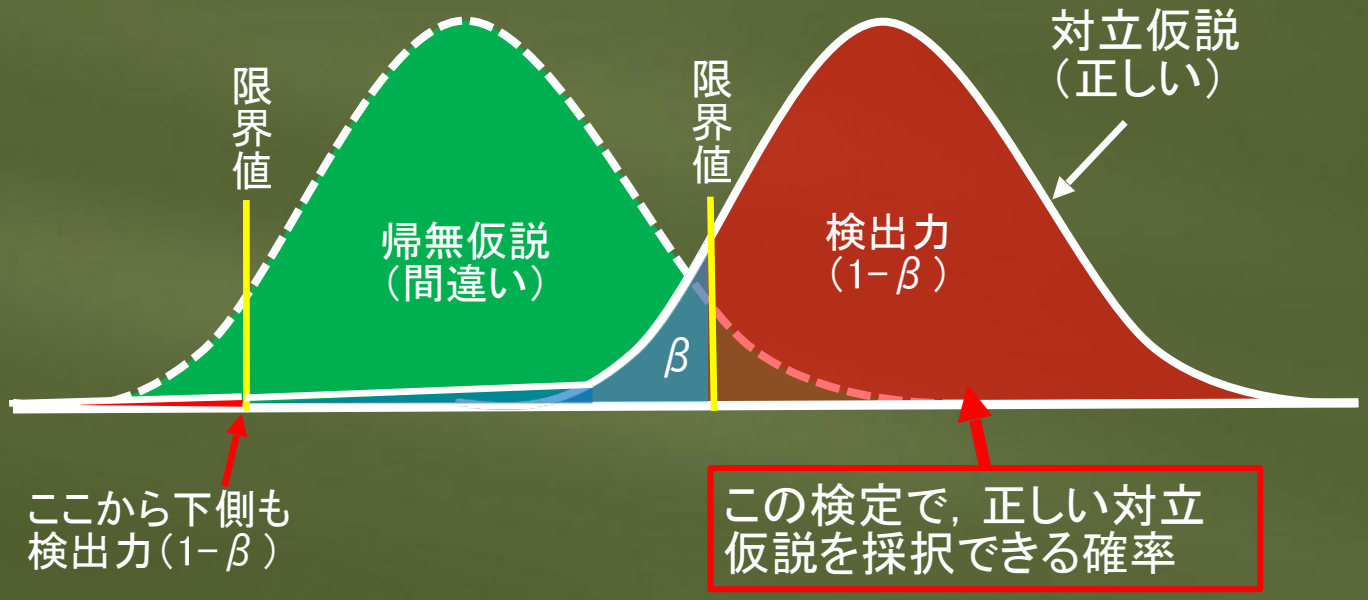


第一種の過誤が致命的でないなら，有意水準αの許容を大きくすることで，αとトレードオフ関係にあるβは小さくなる
(ただし，よほどデータを集めるのが困難な場合を除いて用いられない)

検出力

ここまでβを使ってきましたが…

第二種の過誤確率βよりも，その補数（1−β）である検出力を考える方がわかりやすい
→本当に処理効果（差）があるときに，「ある」といえる能力

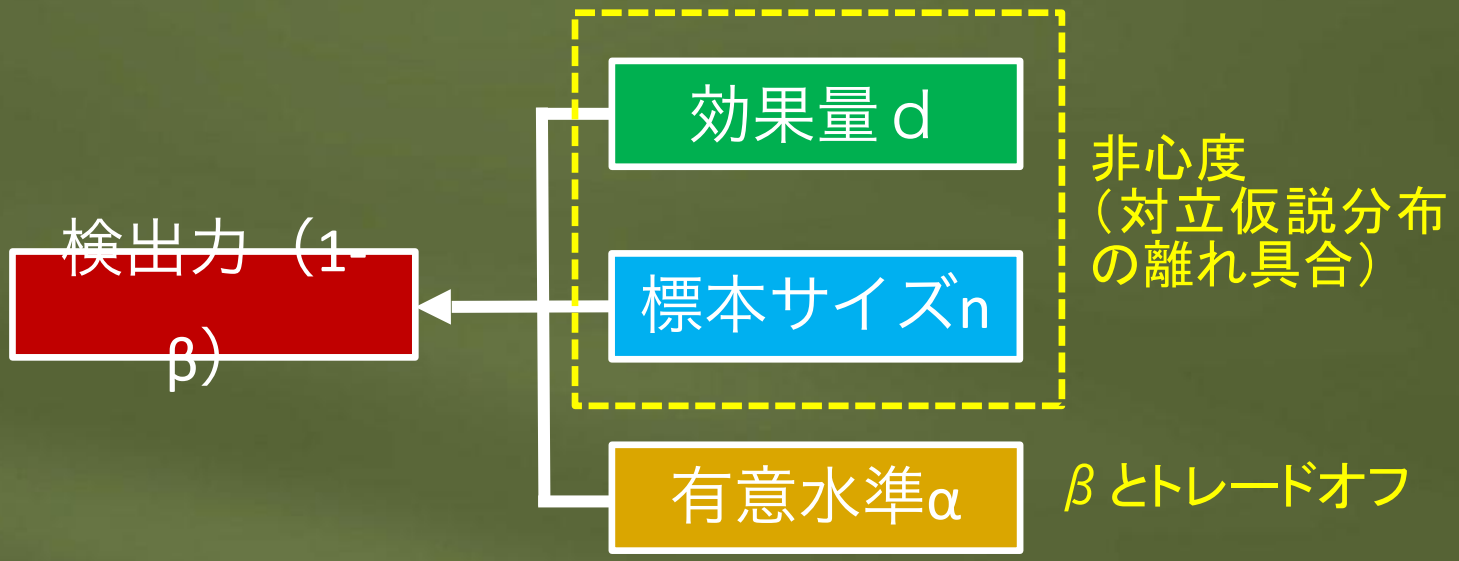


検出力も検定結果に加えよう！

- 検出力は，その検定の能力を表している
- ｐ値や有意水準のみでは，正しい対立仮説を高い確率で採択し損なっていてもわからない
- 事後（検定後）ならば計算できるので，検定結果に書いておくと良い

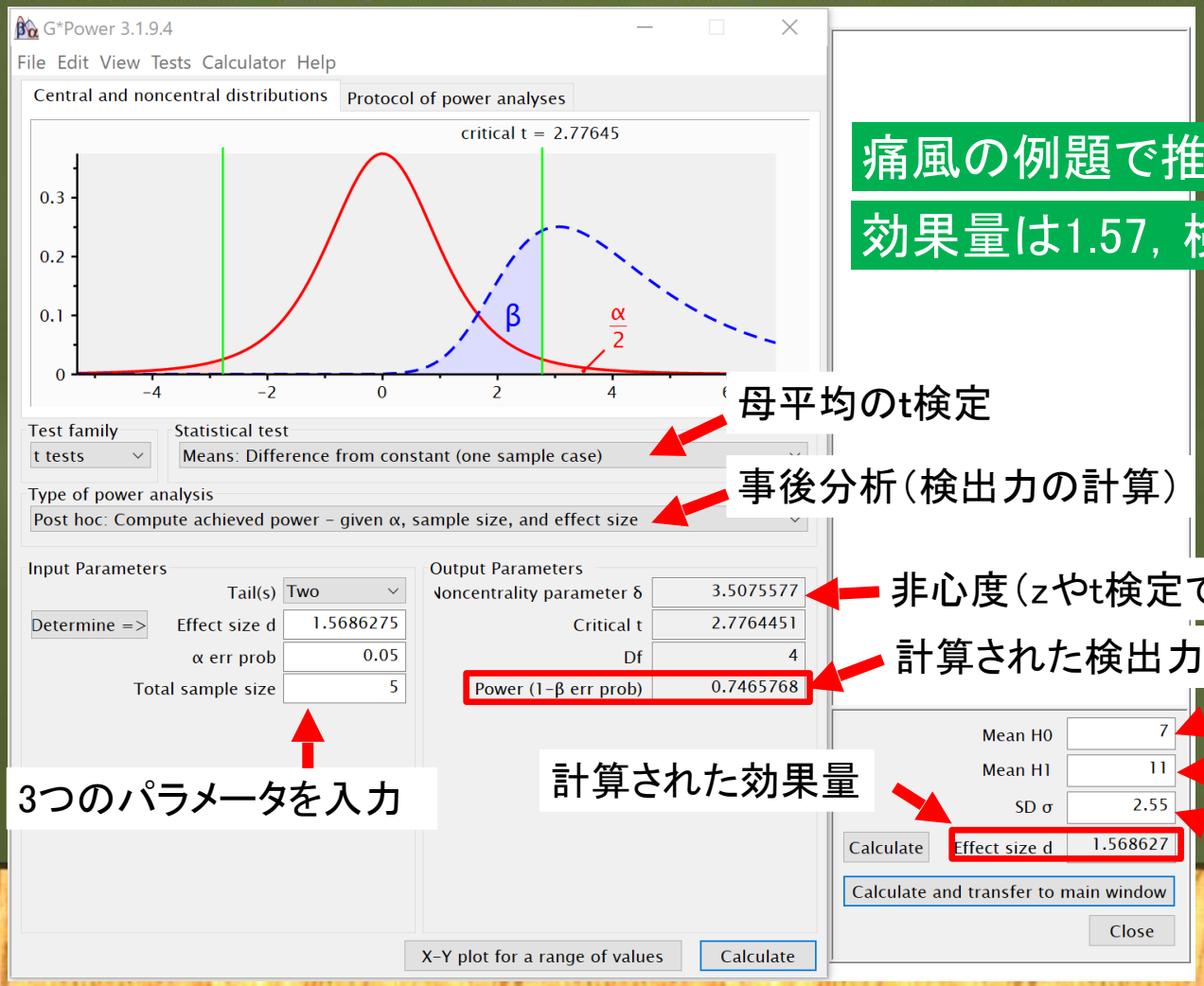
→非心分布の確率計算は難しいので，G*powerなどのソフトウェアが必要

検出力の計算（事後分析）



検出力は3つの要素から決まる（これら3要素をパラメータとして，ソフトウェアに入力すれば計算できる）

ソフト（G*power)を使った検出力計算



痛風の例題で推定してみると…

効果量は1.57，検出力は0.75

母平均のt検定

事後分析（検出力の計算）

非心度（zやt検定では検定統計量）

計算された検出力

比較値 μ_0

標本平均 $\bar{x}$

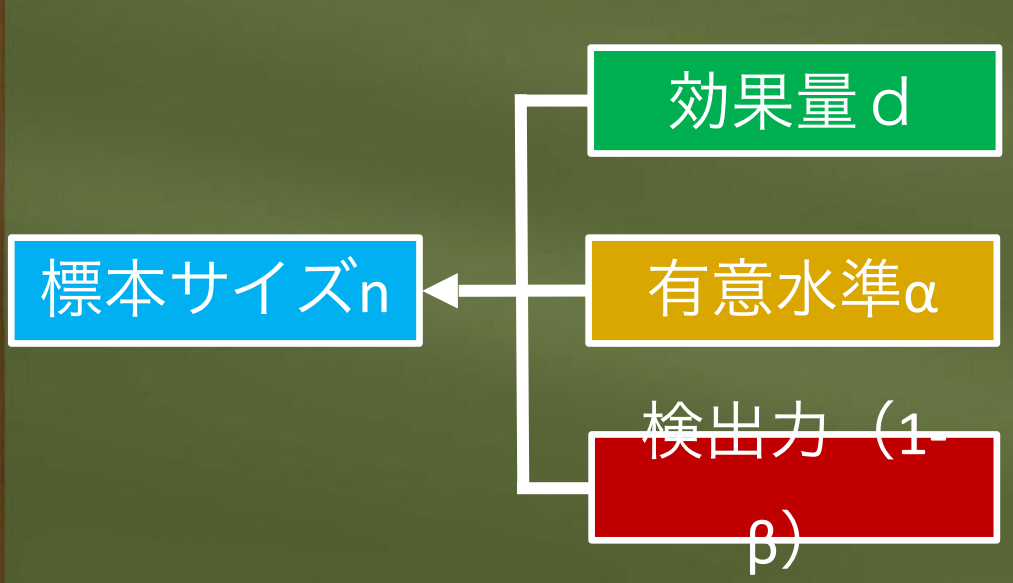
不偏標準偏差 $\hat{\sigma}$

計算された効果量

3つのパラメータを入力

標本サイズの計算（事前分析）

任意の有意水準の下で望ましい検出力になるような標本サイズを事前に求める（効果量は予想するしかない）

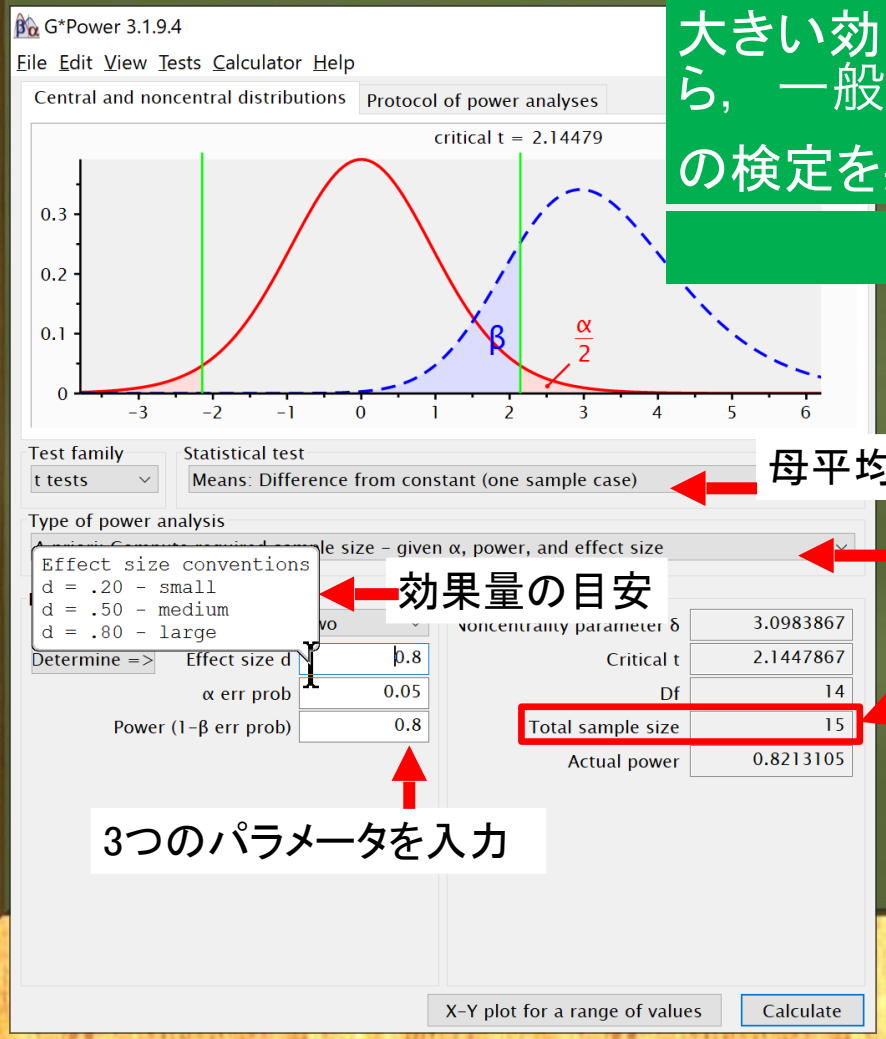


関連データがあれば推定し，なければ検定ごとに示されている目安を使う（母平均の検定で0.2～0.8）

0.05（5%）がもっとも一般的

0.8（80%）程度で良いだろうとコーエンが提唱

ソフト（G*power)を使った標本サイズの計算



大きい効果量（0.8）が得られると予想されるなら，一般的な有意水準（0.05）と検出力（0.8）の検定を実現するための標本サイズは…

n=15

母平均のt検定

事前分析（標本サイズの計算）

効果量の目安

計算された標本サイズ

3つのパラメータを入力

注：第二種の過誤が致命的な場合は，検出力か有意水準αを大きく設定する

以上で第６章は終了です。