

限界値は「(比較対象となる特定の値)  $\pm$  (分布表から読み取った値)  $\times$  母標準誤差」となります。今回、比較対象は 5 mm,  $z$  の値は分布表で上側確率が 0.025 となる 1.96 (ステップ④を復習), 母標準誤差は 2 mm ですので、限界値は「 $5 \pm 1.96 \times 2$ 」となります。よって、標本平均である 11 mm が 1.08 mm よりも小さいか, 8.92 mm よりも大きければ帰無仮説を棄却し、そうでなければ受容します。

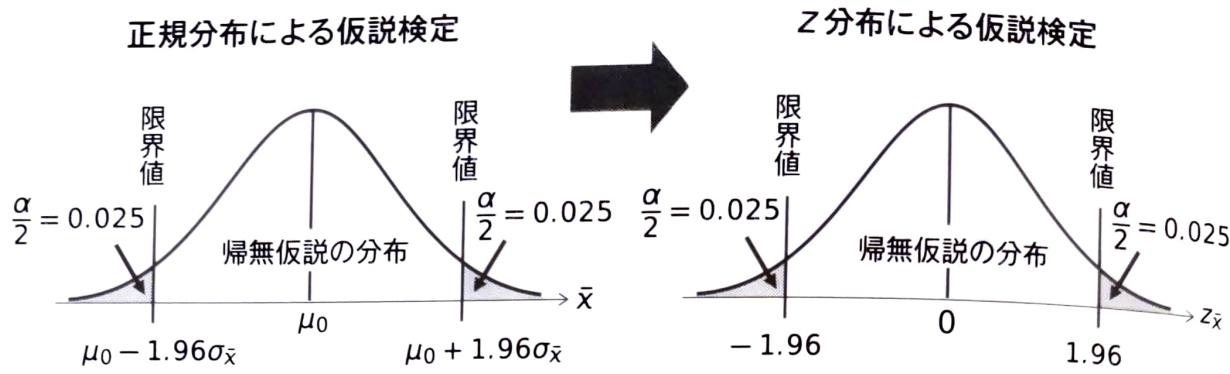
A君の見つけたテントウムシの平均体長 11 mm は、限界値（上限値）の 8.92 mm よりも大きいので、「今回の 6 mm という体長差は、偶然ではなかなか現れないほど大きい差といえるので、両者は異なる種である」と考え、帰無仮説を棄却して、対立仮説を採択します。つまり、A君の見つけたテントウムシは両側 5 % の有意水準で統計的に新種であるといえます（もちろん実際はこんな単純ではありません。章末のトピックス⑤をご覧ください）。

## 6.4 標準正規 ( $z$ ) 分布や $t$ 分布の利用

### ● 標準正規分布による検定 ( $z$ 検定)

本節では、同じ母平均の検定ではありますが、標本平均  $\bar{x}$  そのままではなく、それを標準化した  $z_{\bar{x}}$  や、母分散が不要の  $t_{\bar{x}}$  を検定統計量とした場合を考え、次章からの本格的な検定（母平均の検定も立派な検定ですが……）への準備をしたいと思います。

さて、前節の母平均の検定の方法では、標準化していないオリジナルの値のままの正規分布を使っていたため、観測された値（標本平均という検定統計量）が棄却域に入っているかどうかを判断するためには、比較する特定の値や母標準誤差を用いた、面倒な限界値の計算が必要でした。そして、検定対象が

図 6.10  $z$  分布を使うと限界値が単純になる

異なる度に、当然ながらその限界値の値は（同じ有意水準でも）変わってしまいます。

しかし、図 6.10 のように、標準化された正規分布である  $z$  分布を使った検定（ **$z$  検定** :  $z$ -test）ならば、母平均  $\mu$  はゼロ、母標準誤差は 1 に揃えられているので、限界値も  $\pm 1.96$  と単純になります。つまり、観測された標本平均を標準化した  $z_{\bar{x}}$  が  $1.96$  より大きい、もしくは  $-1.96$  よりも小さければ帰無仮説が棄却できるのです。そして、この限界値はどのような検定対象でも（有意水準が同じならば）変わらないという、わかりやすさもメリットです。

ただし、検定前に次の式で統計量  $\bar{x}$  を標準化しておくという手間がかかりますので、決して手続きが楽になるというわけではありません。

$$z \text{ 検定の統計量 (母平均の検定)} \quad z_{\bar{x}} = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

↑ 特定の値  
↑ 特定の値

以上のように、母平均の  $z$  検定では、観測された標本平均  $\bar{x}$  の標準化変量  $z_{\bar{x}}$  が検定統計量になります。それを特定の値  $\mu_0$  と比較するのですが、この比較対象  $\mu_0$  は帰無仮説の下では  $\bar{x}$  の真の値（母数）なので、 $z$  分布の平均（中心）である 0 になります。よって、 $z$  検定では、帰無仮説  $H_0 : z_{\bar{x}} = 0$ 、対立仮説  $H_1 : z_{\bar{x}} \neq 0$  と書くとわかりやすいでしょう。ただし、仮説は本来母集団について立てるものなので、やはり正式には前節と同じように  $H_0 : \mu = \mu_0$ 、 $H_1 : \mu \neq \mu_0$  となります。

## ● $t$ 分布による検定 ( $t$ 検定)

このように、よりシンプルになる  $z$  検定ですが、母分散が既知でないとできないという大きな欠点があります。母分散がわからなければ母標準誤差  $\sigma_{\bar{x}}$  も

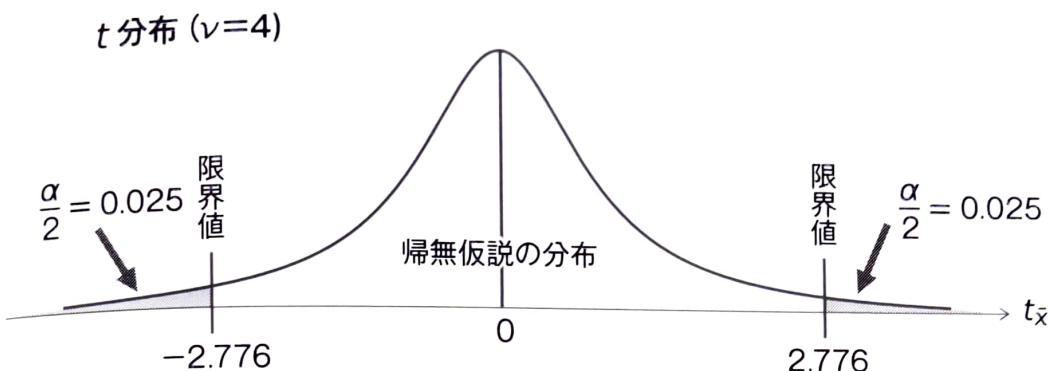


図 6.11 t分布を使った検定（自由度が4の場合）

わからぬいため、標準化できないからです。そこで、母分散が未知のときにz分布の代わりとするt分布を使った検定（**t検定**；*t-test*）を実施します。

*t*分布とは、観測データから推定できる不偏標準誤差  $\hat{\sigma}_{\bar{x}}$  を使って準標準化した統計量  $t_{\bar{x}}$  が従う確率分布でした。

しかし、この*t*分布を使って検定する場合、面倒なことがあります。それは、分布の形を決める母数として、**自由度ν** ( $= n - 1$ ) が関わってくるということです。つまり、*t*検定で帰無仮説の是非を判断する基準である限界値は、（有意水準だけでなく）自由度によっても変化するということです（*z*分布に自由度は関係ないので、*z*検定では  $\alpha = 0.05$  ならば、いつでも  $\pm 1.96$  です）。

図6.11は、例として自由度が4（つまり標本サイズ  $n$  が5）の場合の*t*検定（有意水準  $\alpha$  は両側 5 %）を描いたものです。

図6.10（の右図）の*z*検定と比べると、同じ有意水準なのに限界値の絶対値が大きくなっている ( $\pm 1.96 \rightarrow \pm 2.776$ ) ことがわかります（自由度が小さくなるほど限界値の絶対値は大きくなります）。*t*分布では、分布のバラツキである母標準誤差がわからぬいため、その代わりに不偏標準誤差としてやや大きめに見積もります。そのため、*t*検定における帰無仮説の棄却域は（*z*検定よりも）やや狭くなり、判定も保守的になる（棄却が難しくなる）のです。このあたりは*t*分布を使った区間推定における解説と同じですね。

（1標本の）母平均の*t*検定の検定統計量は次のようにになります。

$$\text{t検定の統計量 (母平均の検定)} \quad t_{\bar{x}} = \frac{\bar{x} - \mu_0}{\hat{\sigma}_{\bar{x}}} = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\hat{\sigma}^2}{n}}}$$

↑ 特定の値

このように、母平均の*t*検定では、観測された標本平均  $\bar{x}$  の準標準化変量  $t_{\bar{x}}$  が検定統計量になります。それを特定の値  $\mu_0$  と比較するのですが、帰無仮説

の下では  $\mu_0$  は標本平均の真の値なので、 $z$  検定のときと同じように、 $t$  分布の平均（中心）である 0 となります。よって、 $t$  検定でも、帰無仮説  $H_0 : t_{\bar{x}} = 0$ 、対立仮説  $H_1 : t_{\bar{x}} \neq 0$  とするとわかりやすいのですが、やはり仮説は母集団について立てるものなので、 $H_0 : \mu = \mu_0$ 、 $H_1 : \mu \neq \mu_0$  と書きましょう。

### 例題

痛風の疑いのある 5 名に血液検査を実施したところ、下記のような尿酸値を観測しました（低い順、単位は mg/dL）。

7, 10, 12, 13, 13

このグループ（個人ではありません）は痛風に罹患しているといえるでしょうか？ なお、尿酸値の正常値は上限 7.0 mg/dL であることがわかっています。

解：

痛風の疑いのある 5 名の標本（血中尿酸値）から平均を計算すると 11 mg/dL です。正常値 7 mg/dL と比較すると、標本は 4 mg/dL 高いことがわかります。特定の値との差が偶然の範囲といえるのかどうかは、「母平均の検定」で確かめることができます。今回は母分散がわからないので  $t$  分布を使いましょう（母分散の  $t$  検定）。また、本事例は上限値なので片側検定でも良いのですが、一般的な両側 5% 有意水準で検定しましょう。

### 手順①：仮説の設定

$$\left\{ \begin{array}{l} \text{帰無仮説 } H_0 : \mu(\bar{x} = 11 \text{ mg/dL}) = \mu_0(7 \text{ mg/dL}) \rightarrow \text{正常である (痛風ではない)} \\ \text{対立仮説 } H_1 : \mu(\bar{x} = 11 \text{ mg/dL}) \neq \mu_0(7 \text{ mg/dL}) \rightarrow \text{正常ではない (痛風である)} \end{array} \right.$$

### 手順②：検定統計量の計算

$t$  検定の統計量の式に、標本平均 ( $\bar{x} = 11$ )、特定の値 ( $\mu_0 = 7$ )、不偏標準誤差 ( $\hat{\sigma}_{\bar{x}} = 1.14$ ) を代入します。なお、不偏標準誤差  $\hat{\sigma}_{\bar{x}}$  は、不偏分散 ( $\hat{\sigma}^2 = 6.51$ ) を標本サイズ ( $n = 5$ ) で割って平方根を取ったものです。すると、検定統計量  $t$  は “3.51” となります。

$$t_{\bar{x}} = \frac{\bar{x} - \mu_0}{\sqrt{\hat{\sigma}_{\bar{x}}^2}} = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\hat{\sigma}^2}{n}}} = \frac{11 - 7}{\sqrt{\frac{6.51}{5}}} = 3.51$$

### 手順③：確率の計算

帰無仮説（このグループは正常である）が正しいとした下で  $4 \text{ mg/dL}$  以上の差が観測される確率を計算するのは難しいので※，その代わりに帰無仮説の分布のどのあたりに位置するのかで，結果の起こりやすさ／起こりにくさを判定します。次の手順で求める限界値よりも極端な値を取っているようならば，今回観測された差は偶然とはいえないぐらい大きいものであったと考えます。

### 手順④：仮説の判定

両側 5% の有意水準に対応する限界値を  $t$  分布表（付録Ⅱ）から読み取ります。すると，自由度  $\nu = n - 1 = 4$  の行と上側確率  $p = 0.025$  の列がクロスする “2.776” が限界値であることがわかります（下側の −2.776 も限界値になります）。検定統計量 ( $t = 3.51$ ) は限界値 (2.776) よりも大きいので，「有意差があった」と判定できます。いいかえれば，このグループは，正常な母集団とは異なる母集団から抽出された (= このグループは痛風である) といえます。

※手計算では手順③の確率の計算はできませんが，Excel 関数の  $T.DIST.2T(3.51,4)$  で，両側を合わせた  $p$  値 = 0.0247 を求めることができます。有意水準 0.05 と比べて  $p$  値の方が小さいので，「有意差あり」と判定できます ( $p < 0.05$ )。

## 6.5 検出力分析

### ● 統計的過誤

本章の後半は，仮説検定の能力（性能）を表す“検出力”について解説します。一昔前の統計学の入門書ならば扱わなかった内容ですが，近年は卒論レベルでも，実施した検定の「検出力がどのくらいだったのか」を表記することが求められるようになりました。また，目標の検出力を設定すれば，事前にどのくらいの大きさの標本を確保すれば良いのか（データをいくつ集めれば良いのか）を計算することもできます。

検出力を学ぶには，仮説検定における 2 種類の間違い（統計的過誤：statistical error）について理解しておく必要がありますので，まずはそちらから説明します。

仮説検定は，あくまで母集団の一部である標本を用いるため，間違った判断を下してしまうことも当然あるでしょう。それは次の 2 種類です。

1 つ目は，第一種の過誤（Type I error）と呼ばれる過ちで，「差（処理効果）がないのに，あると判断してしまう間違い」です。この過ちを犯す確率は危険

率とも呼ばれ、 $\alpha$ で表記します。

2つ目は、第一種の過誤の逆で「差があるのに、ないと判断してしまう間違い」です。こちらは**第二種の過誤** (Type II error) と呼ばれ、この過ちを犯す確率は $\beta$ で表記されます。

これらは重要なので整理しておきましょう（「真」とは仮説の内容が正しいこと、「偽」とは間違っていることです）。

**第一種の過誤**：帰無仮説が真なのに棄却してしまう過ち（確率は $\alpha$ ）

**第二種の過誤**：帰無仮説が偽なのに棄却しない過ち（確率は $\beta$ ）

## ● 第一種の過誤

2種類の統計的過誤と、それを犯す確率 $\alpha$ 、 $\beta$ について、仮説の分布を使って確認しておきましょう。第一種の過誤は、帰無仮説が正しいのに、たまたま偶然、実験結果（検定統計量）が帰無仮説の分布の端の値を取ってしまったために、帰無仮説を棄却して対立仮説を採択してしまったという状況です。ですから本検定で、それを犯す確率（危険率） $\alpha$ は、図6.12のように帰無仮説の分布の両裾を合わせた部分となります。

$\alpha$ という記号が出てきた時点で気がついたかもしれません、仮説検定で事前に設定する有意水準こそが、この第一種の過誤を犯す確率のことです。つまり、前々節のステップ③で設定した $\alpha$ は、検定における第一種の過誤を「この程度に抑えておきたいな」という、許容できる危険率の上限だったのです。

例えば、製薬会社が効き目のない癌治療薬を誤って発売したら大変です（会社は傾くでしょう）。ですから、仮説検定では、この第一種の過誤ができるだ

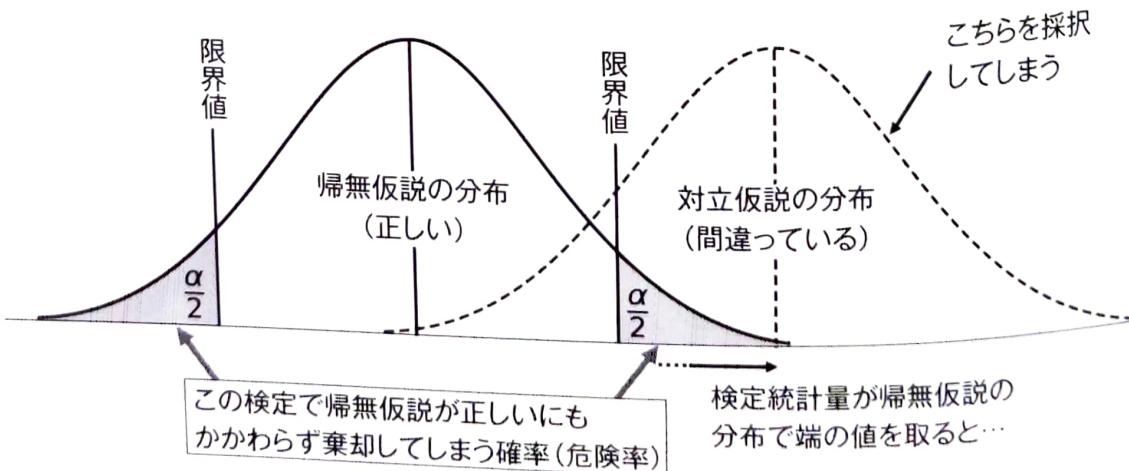


図 6.12 第一種の過誤を犯す確率 $\alpha$ （危険率）

け犯さないように厳しく設定するのです。その厳しさとして、仮に実験と検定を20回実施しても1回(5%)までしか過誤を許さない、いいかえれば19回(95%)は過誤を犯さないという基準が一般的に用いられるのです。

## ● 第二種の過誤と効果量

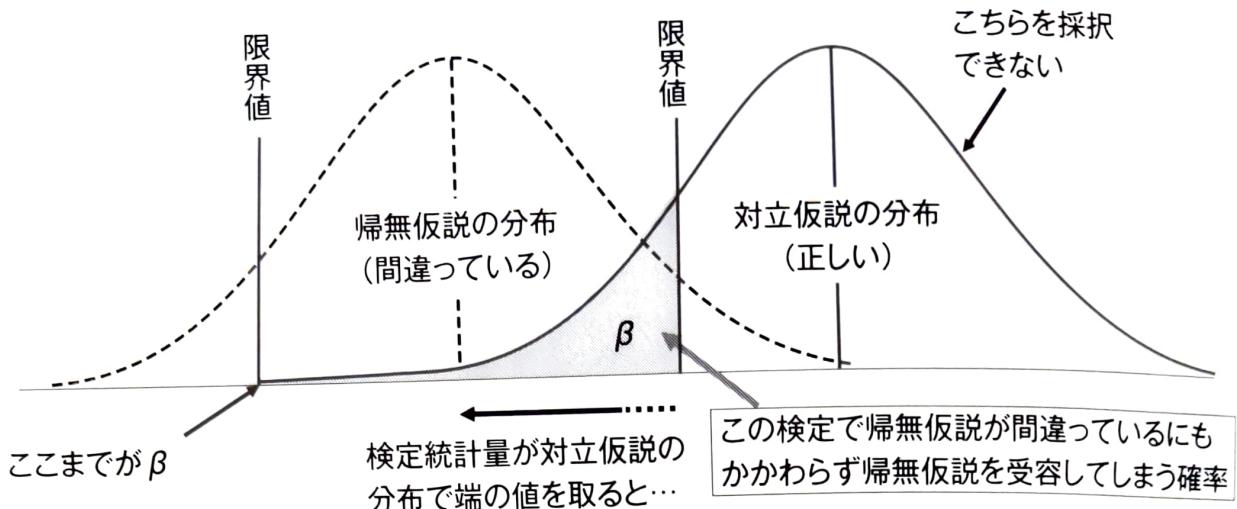
一方、第二種の過誤は、対立仮説が正しいのに、たまたま偶然、検定統計量が対立仮説の分布の（帰無仮説寄りの）端の値を取ってしまったために、帰無仮説を受容してしまったという状況です。いいかえれば、対立仮説が正しかったのに、それを見逃して採択できなかったということです。検定における第二種の過誤を犯す確率 $\beta$ は、図6.13のグレー部分のように、対立仮説の分布の確率となります。 $\beta$ は帰無仮説が棄却されたときに採択される対立仮説の立場から考えるので、帰無仮説側だけの確率でOKです。

さて、こちらの過誤、製薬会社の事例だったら、良い薬をせっかく開発しても、それを発売するチャンスを逃してしまうのですから、やはり低く抑えたい過ちですね。でも、検定のステップでは、( $\alpha$ のように)事前に $\beta = ○○$ として許容できる上限を設定しませんでした。というより、次のような理由から「したくてもできなかった」のです。

対立仮説の分布の母平均は、図6.13のように帰無仮説のそれとは離れています。この両分布の離れ具合を**非心度**(noncentrality)と呼ぶのですが、その大きさによって対立仮説の分布の形が変わるので(z分布のように自由度が関係ない場合は変わりません)。そして、やっかいなことに、非心度は事前にはわからぬのです。というのも、非心度は標本サイズと効果量で決まるためです(z検定やt検定のように、帰無仮説の真値が0になる場合は検定統計量が非心度です)。

**効果量**(effect size)とは、母集団が持つ処理効果そのものの大きさ(を母標準偏差で割った値)のことです。例えば、降圧剤ならば、実際にその薬を飲んでどのくらい血圧が下がるかという、薬が持つ効き目そのもののことなので、実験や検定を実施する前から決まっています。

効果量の計算式は検定によって異なりますが、母平均の検定ならば、「母平均の差 $(\mu - \mu_0) \div$ 母標準偏差 $\sigma$ 」となります(母平均も母標準偏差も未知ですし、標本サイズは関係ありません)。ただし、事後ならば、観測データから計算した検定統計量を非心度に見立て、それと標本サイズから逆算して推定できます。近年は、こうした事後に推定した効果量を、検出力(後述)と一緒に論

図 6.13 第二種の過誤を犯す確率  $\beta$ 

文に記載することが求められるようになってきました。ちなみに母平均の  $t$  検定の効果量  $d$  の推定式は  $t/\sqrt{n}$  です（同じ検定でもいろいろあり、この式はその1つです）。

なお、図 6.12 と図 6.13 を見比べていただくとわかると思いますが、標本サイズと効果量が同じ、つまり両分布の離れ具合（非心度）が同じ場合、 $\alpha$  と  $\beta$  はトレードオフの関係にあります。ですから、仮に  $\beta$  を小さくしようとしても  $\alpha$  が大きくなってしまうのです。

というわけで、第一種の過誤  $\alpha$  も第二種の過誤  $\beta$  も抑えたいのはやまやまなのですが、 $\beta$  を事前には計算できること、またコンプライアンスが重視される現代において、会社に大きなインパクトを与えるのは第一種の過誤である（効かない薬を発売してしまった場合と、効く薬を発売し損ねた場合の社会的影響を想像してみればわかりますね）ことから、仮説検定では有意水準  $\alpha$  を優先して厳しく（小さめに）設定されてきたのです。

## ● 標本サイズで $\beta$ を抑える

一般には第一種の過誤を犯さないことが重要だと述べましたが、場合によっては第二種の過誤を犯さないことが重要になることもあります。例えば、環境保護政策の効果の検定 ( $H_0$ : 効果なし,  $H_1$ : 効果あり) において、本来効果のある政策を採択し損なった場合を考えてみてください。自然環境は破壊されたら最後、もう二度と復活しないでしょう。また、食品添加物の毒性を検定 ( $H_0$ : 毒性なし,  $H_1$ : 毒性あり) する場合も、毒性を見過ごして、それを添加した食品を売り出したら大変なことになりますね。このような場合には、第二種の過誤を犯す確率  $\beta$  を低く抑えることが重要になるでしょう。しかし、先ほ

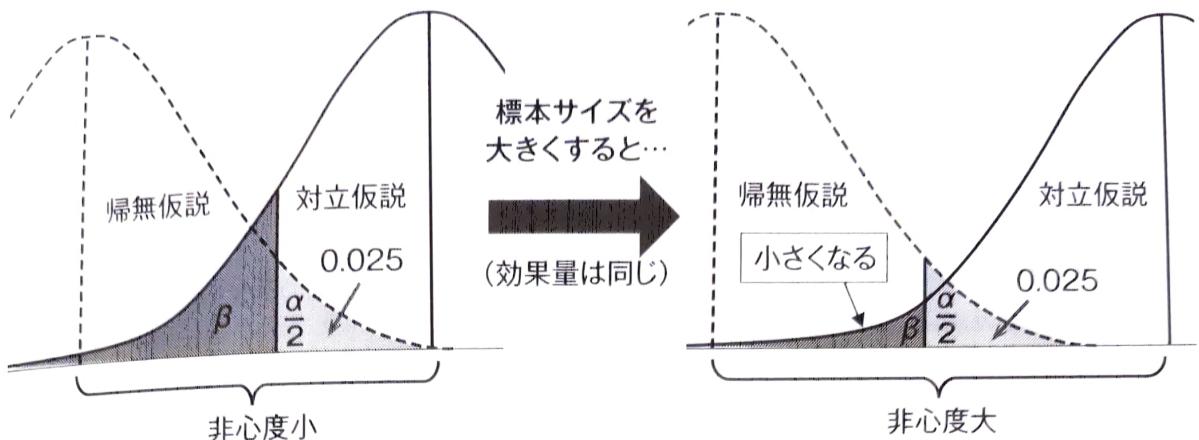


図 6.14 非心度が大きくなると  $\beta$  が小さくなる（有意水準  $\alpha$  は同じ）

と説明したように、事前に  $\beta$  は知り得ませんので、上限を設定することはできません……。さて、どうすれば良いでしょうか。

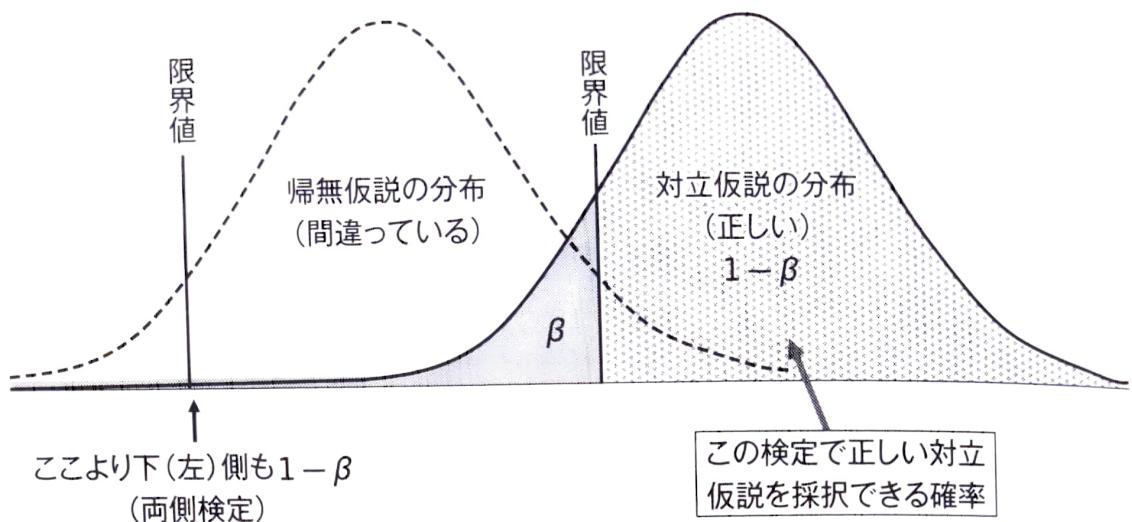
ヒントは、非心度（両分布の離れ具合）を決める標本サイズと効果量にあります。母集団の特性である効果量については、自分ではどうしようもない（検定と関係なしに決まっているもの）ですが、標本サイズは実験の計画段階で調整することができます。

つまり、図 6.14 のように、独立した実験を繰り返してデータをたくさん観測（標本を大きく）すれば非心度が大きくなり、（有意水準  $\alpha$  をそのままに）  $\beta$  を小さくすることができるのです。ただし、効果量は未知ですので、「 $\beta = \text{○○}$  とするために標本サイズを  $\text{○○}$  とする」とはっきり決めることはできません。あくまで、予備実験や類似の既往研究などから、効果量はこのぐらいだろうと想定した値を使って、（任意の  $\alpha$  の下で） 望ましい  $\beta$  になるような「おおよその標本サイズを逆算」して実験を計画するのです。

## ● 検出力

以上のように、（所与の  $\alpha$  の下で） 望ましい  $\beta$  の小ささで検定するためのおおよその標本サイズは、効果量と有意水準から計算することになるのですが、直接  $\beta$  の大きさを目標とすることはあまりしません。というのも、どうせ対立仮説の分布で考えるならば、その“補数”的な方がわかりやすいからです。

正しい対立仮説を採択し損なってしまう確率が  $\beta$  なので、その補数  $(1 - \beta)$  は、間違っている帰無仮説をちゃんと棄却して、正しい対立仮説を採択できる確率ということになります。いいかえれば「本当に差があるときに、差があると正しく判定できる確率」です（こちらの方がしっくりくるでしょ？）。このように、確率  $(1 - \beta)$  は、本来の有意差を検出できる“検定の能力”を表して

図 6.15 検出力 ( $1 - \beta$ )

いるので**検出力**（検定力；power）と呼ばれます。

検出力を図で表しますと、図6.15のように対立仮説の分布における $\beta$ の残り部分となります。

さて、第二種の過誤を犯さない確率である検出力ですが、一体、どれぐらいが望ましいのでしょうか？もちろん、強ければ強い方が良いでしょうが、 $\alpha$ を大きくせずに検出力を強くするためには標本を大きくしなければならないので、実験をする側にしてみれば負担になってしまいます。

実は、コーベンという統計学者が「検出力は0.8ぐらいで良いだろう」と指針を示してくれています。分野によっても異なるでしょうが、入門者は、この0.8という検出力を目標として逆算した標本サイズを計画すればよいでしょう。検出力が0.8ということは、実験と検定を10回実施した場合、8回は正しい対立仮説をきちんと採択できるということです。また、検出力が0.8ということは $\beta$ が0.2ということですから、コーベンは $\alpha$ （こちらは0.05とするのが一般的です）と $\beta$ の比を1:4とするのが妥当であるといっているわけですね。もちろん、先に例を示した、環境保護政策の効果や食品添加物の毒性など、 $\beta$ が重大な過誤と考えられる場合には、強い検出力を得るために、より大きな標本の実験を計画したり、場合によっては大きな $\alpha$ を設定することになります。

## ● ソフトを使った検出力の計算

検出力も $\beta$ 同様、対立仮説の分布における確率なので、事前に求めることはできません。しかし、事後ならば、観測データを使って、実施した検定の検出力を推定できます。近年では、検出力の記載が求められることが多いので、可能ならば計算しておくと良いでしょう。ただし、対立仮説の分布は極めて複雑

(式は省略)なので、ソフトウェアが必要となります。

いろいろなソフトウェアがありますが、使いにくいものが多く、現在のところ、ドイツの大学の先生方が開発した無料ソフトG\*Powerの一択といつて良いでしょう。2021年4月現在、ハインリッヒ・ハイネ大学のWebページ (<https://www.psychologie.hhu.de/>) でバージョン3.1.9が公開されており、WindowsとMacのOSで動作します。正式に公開されているマニュアルは英語だけですが、日本語で使い方を説明しているWebページも数多くあるので、困ることはないと思います。

さて、G\*Powerをインストールして起動すると、図6.16のような画面が出てきます。今回は、母平均のt検定の例題（痛風）の検出力を計算してみましょう。（1標本の）母平均の検定は [t tests] の中の [Means: Difference from constant (one sample case)] になります。そして、[Type of power analysis] で [Post hoc] を選べば事後分析、つまり検出力の計算画面になります。

計算といっても、左側の空白になっている3カ所に、検出力を決める3つの要素（パラメータ）を入力するだけです。なお、統計学におけるパラメータと

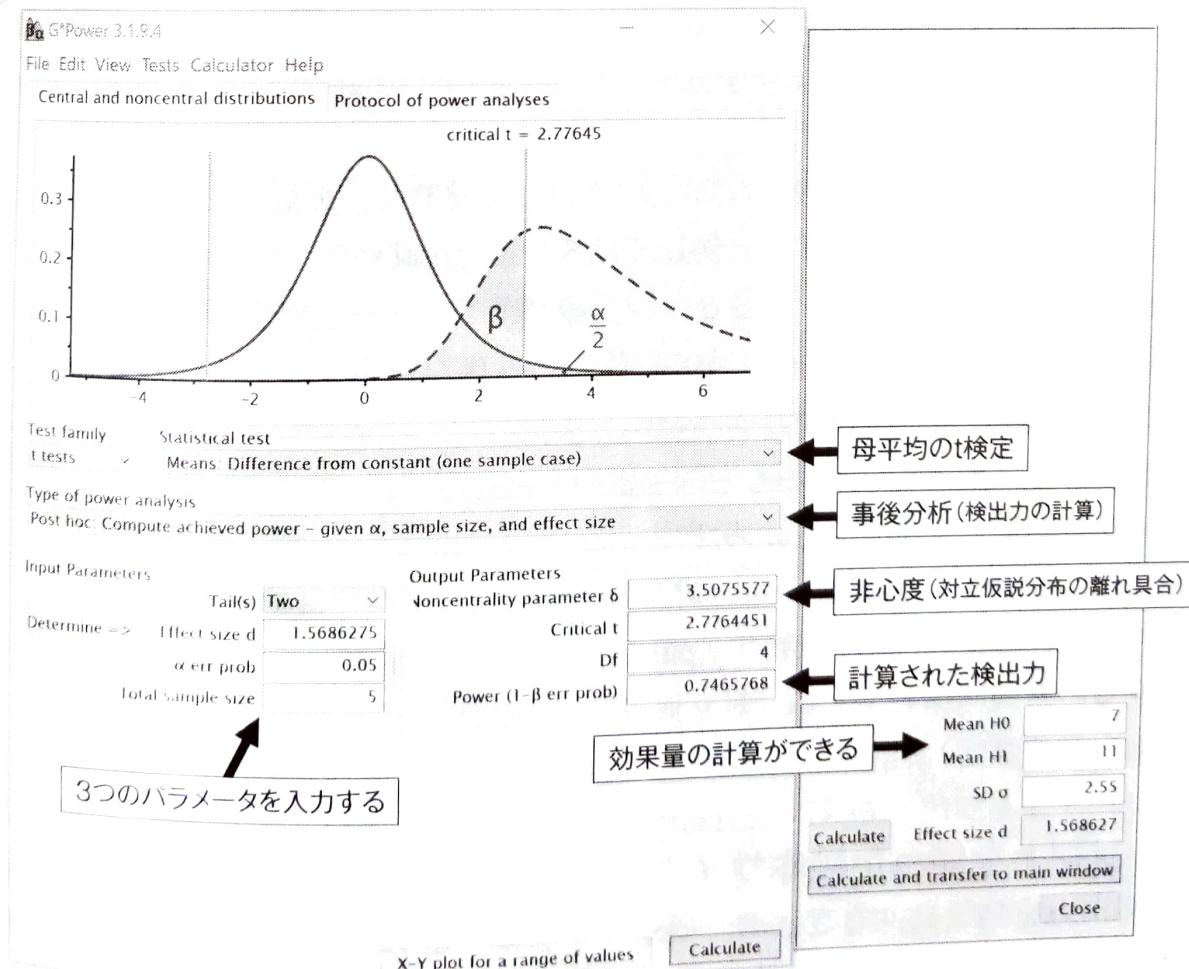


図 6.16 G\*Power の起動画面（事後分析）

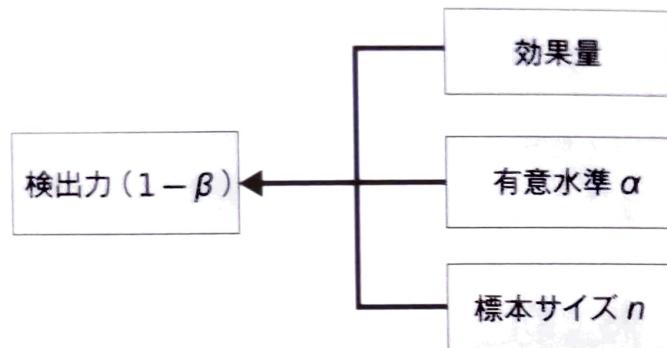


図 6.17 検出力を決める3つの要素

は、母平均など母数を意味するのですが、工学分野では電子機器の動作を設定する値を指すこともあります（ここでは後者の意味で使います）。

検出力を決める3要素とは、 $\beta$ を決める3要素と同じですので、非心度に影響する「効果量」と「標本サイズ $n$ 」、そして $\beta$ とトレードオフの関係にある「有意水準 $\alpha$ 」ということになります（図6.17）。

まず効果量ですが、[Determine] というボタンを押すと右側に計算するためのウインドウが出てきますので、比較対象の特定の値（例題では尿酸正常値の7）を [Mean H0] に、標本平均（例題では11）を [Mean H1] に、不偏標準偏差（例題では2.55）を [SD δ] に入力して [Calculate and transfer to main window] を押します。すると自動的に左の [Effect size d]（「d」は数種類ある効果量の1つ）に計算値（例題では1.57）が入ります。有意水準 [α err prob] には、一般的な0.05（両側なのでTailsで「Two」を選択）を入れておきましょう。最後に標本サイズ（例題では5）を [total sample size] に入力して [Calculate] を押します。すると、例題のパラメータ値（両側有意水準=0.05、効果量=1.57、標本サイズ=5）から、検出力は0.747（74.7%）と計算されました。この例題で、標本が小さかったにも関わらず、理想（0.8）に近い検出力となったのは、推定した効果量（平均差／標準偏差）が大きかった（平均差が大きいか標準偏差が小さい）からです。

また、G\*powerの画面から、対立仮説の $t$ 分布（右側破線）が3.5を平均として（=非心度 [Noncentrality parameter δ]）左右非対称となっているのがわかるのも興味深いですね（対立仮説の $t$ 分布を、中心がズレているので非心 $t$ 分布と呼びます）。

### ● ソフトを使った標本サイズの計算

検出力を効果量、有意水準、標本サイズから計算できるということは、標本サイズを検出力、有意水準、効果量の3要素から計算できるということです

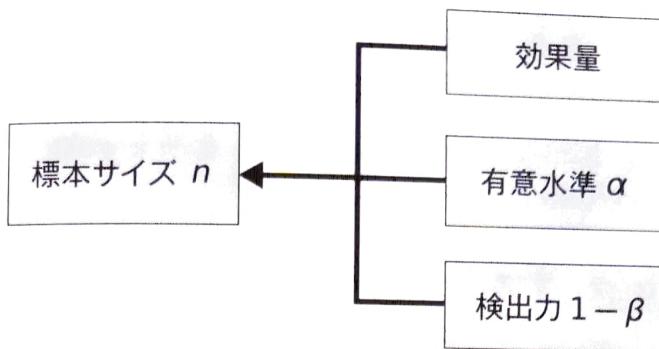


図 6.18 標本サイズを決める3つの要素

(図6.18)。事前に標本サイズを決める作業は、本来、第10章の実験計画法に含まれる内容ですが、せっかくここまで来たので試してみましょう。

ただし、この計算は事前に行うので（事後に標本サイズを計算しても仕方がないません）、効果量を推定できないという問題が発生します。検出力が推定できたのは、事後で観測データが手元にあったからです。そこで、予備実験や類似の既往研究などから、効果量の推定に必要な値を想定することになります。もし、予備実験を実施できず、類似の既往研究もない場合には、仕方がないので適当な効果量を入力するしかありません。ただし、検定ごとに、大きな効果量は○○ぐらい、中程度の効果量は○○ぐらい、小さな効果量は○○ぐらいと、目安となる効果量の値は提案されています。

さて、標本サイズの計算も G\*power を使えば大変簡単にできます。[Type of power analysis] のところで事前分析を意味する [A prior] を選択すると、図 6.19 のような画面が現れます。あとは、事後分析（検出力）のときのように、3要素（効果量、有意水準、検出力）を入力するだけです。事例では、有意水準と検出力はもっとも一般的な値 ( $\alpha = 0.05$ , 検出力 = 0.8) としておきましょう。効果量は、事後分析のときのように別ウィンドウから計算することができますが、予備実験や既往研究など、ヒントにできるデータがなにもない場合は、カーソルを [Effect size d] の右の入力部分に近づけると、図 6.19 左画面のように効果量の目安（小 : 0.2, 中 : 0.5, 大 : 0.8）が表示されますので、それを参考に入力します（事例では大きな効果量が得られたことを想定して 0.8 とします）。

これら3つのパラメータ（効果量=0.8, 両側有意水準=0.05, 検出力=0.8）を入力し、[Calculate] を押すと、[Total sample size] は 15 と出てきます。つまり、効果量が大きい 0.8 と想定できるとき、（有意水準が 0.05 の下で）0.8 の検出力を得る検定を成功させる（有意に差があるという結果を得る）ためには、15 という標本サイズ（データ数  $n$ ）を確保する実験を計画すれば良いのです。

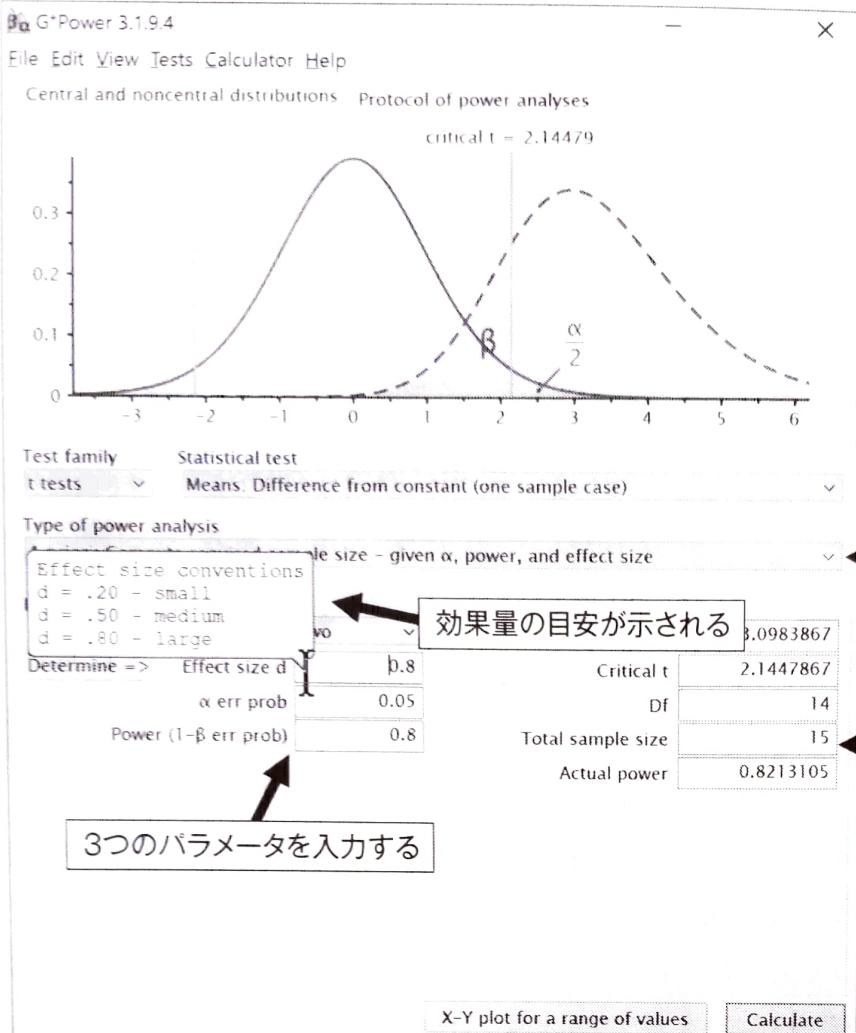


図 6.19 G\*Powerによる標本サイズの計算（事前分析）

このソフトウェアを使って、効果量や検出力を変化させると、必要な標本サイズはどれくらいになるか、いろいろな値を入れてシミュレーションしてみてください。このように、理想的な実験・検定を実施するために事前に標本サイズを求めたり、事後に検出力や効果量を求めて検定の良し悪しを評価したりすることを、総称して**検出力分析**（power analysis）といいます。

## トピックス⑤

### 新種の発見

例題では簡単に「新種（＝学名のない種）」などという言葉を使っていますが、もちろん簡単にみなさんが「新種を発見した！」などということはできません。

例えば、みなさんが見たことのない昆虫を発見したら、まずは、それが属するであろうグループに関する膨大な文献や標本群と比較し、どれにも該当しないことを確認しなければなりません。それはまさに大仕事で、脚や体の形や色、模様、剛毛の並び方を解剖切斷しながら比較していきます。次に、「完模式標本（ホロタイプ）」という、それぞれの種を明確に定義するために、世界にたった

1つだけ保存されている標本と比較して、ようやく新種であることを確信できるのです（ただし、この標本は大変貴重なものなので、それに準ずる副標本を使うか、あるいはこのプロセス自体がスキップされることも多いようです）。そして最終的には、『国際動物命名規約』という国際的な学名の用法規範（他にも植物に関する規約や細菌に関する規約があります）に沿って、学名を決め、ようやく論文を書くことができるのです。そして、その論文も、複数の匿名査読者から「これでよし」という評価を受けなければ学会誌に載せてもらえない。

——参考『Web TOKAI 虫を見る 第11回：新種を観る／杉本美華』

## 章末問題

**問1** A君は、卒論研究で地元の販売農家100人に対してアンケート調査を実施しました。その結果、調査した農家の平均年齢は65歳で、標準偏差は18歳でした。なお、全数調査（2005年農林業センサス）によって、日本の販売農家は196万人で、その年齢の平均は63歳、標準偏差は11歳で、正規分布に従うことがわかっています。また、販売農家とは年間の農産物販売金額が50万円以上の農家を指します。

このように、A君の調査した農家の平均年齢は日本全体の農家の平均年齢に比べると2歳ほど高い値となっているが、果たしてA君の調査した農家の年齢は真の値とは異なっているというべきであろうか？ 両側5%の有意水準で検定しなさい。

**問2** G\*powerなどのソフトウェアを使用して、問1の事後分析（検出力と効果量を推定すること）を実施しなさい。また、本来、どのくらいの標本サイズで検定すれば、（有意水準は5%のままで）検出力0.8の検定が実現できたのか（事前分析）を、やはりG\*powerなどを使って計算しなさい。