

改訂増補版： 統計検定を理解せずに使っている人のためにII

池田郁男

東北大学未来科学技術共同研究センター

改訂増補にあたって

この総説は、「統計検定を理解せずに使っている人のためにII」の改訂増補版であり、「改訂増補版：統計検定を理解せずに使っている人のためにI」の続きである。改訂増補に当たったの詳細は、「改訂増補版：I」の冒頭をお読みいただきたい。この改訂増補版では、「II」にあった誤りを修正した。また、理解しにくい部分について、さらにわかりやすい説明に努めた。

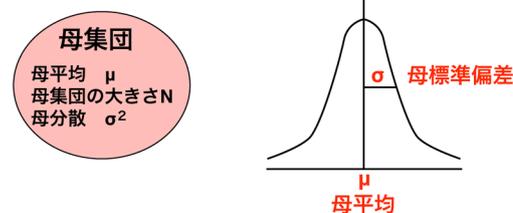
「改訂増補版：I」では、母集団、標本、母分散、母標準偏差、標本分散、標本標準偏差、不偏分散、不偏標準偏差、パラメトリック検定とノンパラメトリック検定の違い、正規性の検定について主に記述した。これらの理解が曖昧な場合は、再度「改訂増補版：I」をお読みいただきたい。今回は、標準誤差、2群のパラメトリック検定の基本、有意差の意味、2群のノンパラメトリック検定の基本をおもに、記述した。なお、図番号は、前回「改訂増補版：I」からの通し番号である。まず、図の内容をざっと眺めてから文章をお読みいただきたい。

パラメトリック検定 (図16)

パラメトリック検定は母集団が正規分布していると仮定する (図16)。平均値や分散などのパラメータを使うことからこの名称がある。パラメトリック検定の例とし

パラメトリック検定

母集団が正規分布していると仮定。
平均値や分散などのパラメータを使って検定。



例：t検定や分散分析

図16 ■ パラメトリック検定

ては、Studentのt検定や分散分析などがある。前回と同様、母集団の大きさは N 、母平均は μ 、母分散は σ^2 と定義する。母平均 μ 、母分散 σ^2 の正規分布を $N(\mu, \sigma^2)$ と記述することとする。

パラメトリック検定の基本：母集団から取り出した n 個の標本の平均値 (標本平均) の分布を考える (図17)

かなり込み入った話になるので、注意して読んでほしい。じっくり読まないといけないが、ここを理解できないと標準誤差SEとt検定は理解できない。ここでは、母

◆◆◆ コラム ◆◆◆

学生を含めて生命科学系研究者の多くは統計検定を利用してはいる。しかし、多くの論文、学会発表において統計検定が誤用されているため、研究の信頼性が揺らいでいる。世界中の研究論文の30~50%は統計に何らかの問題があるという調査結果があり、日本だけの問題ではない。統計の誤用が多い理由は、研究者が統計を理解していないからにほかならない。標準偏差(SD)や標準誤差(SE)が何を意味するのか理解していない研究者は少なくない。研究者は論文や学会発表で不備のない検定法を知りたいが、多くの統計書はその期待に答えてくれず、間違っただけの統計が横行する原因となっている。

また、統計は有意差を出すためのツールとしか認識していない研究者が多いように見受けられる。一度の実験で都合のよいデータが得られると、疑いもせず論文化している研究者が多いと考えられる。その結果、世界的に再現性のないデータが氾濫している。2016年のNatureの科学者を対象とした調査によると60~80%の研究者が他の研究者の実験は再現性が得られないと回答し、さらに驚くべきことに50

~60%の研究者は自分の実験に再現性がないと回答している。バイテク企業Amgenの研究者が、53のがん研究を再試験したところ、再現できたのは6研究しかなかったと報告している(僅か11%)。

この調査結果から、多くの研究が再現性を調べずに論文化されていること、および、かなりの研究者はそのことを認識していることがわかる。統計を理解していれば、標本データを用いた研究結果は必ずしも真実を語っておらず再現性を調べることは必須であるとわかるはずである。再現性をみていない研究がこれだけ多く世に出回っていることは、単純に統計を理解しておらず問題はないと考えているのか、あるいは、再現性がない可能性もあるが、業績を増やすためにとりあえず論文化しているかのどちらかであろう。また、上述の調査では、「都合のよい報告だけを行っている」と回答した研究者が70%弱も存在した。このような研究者の論文を真剣に読む価値はあるのだろうか? もはや、科学研究は危機的状況にあると言わざるをえない。このセミナーを読んだ皆さんが、統計を正しく理解し、倫理観をもって科学研究に貢献することを切に期待する。

集団から取り出した n 個の標本の平均値(標本平均) \bar{X}_i の分布を考える。個々の標本データの分布ではないので注意してほしい。

図17上のように、正規分布している母集団 $N(\mu, \sigma^2)$ から n 個の標本を取り出し標本平均 \bar{X}_1 を計算する。これを何度も繰り返すとそれぞれ標本平均 $\bar{X}_1, \bar{X}_2, \bar{X}_3, \dots$ が得られる。このたくさん得られた標本平均を一つの別の母集団と考えて分布を調べると、正規分布し、その平均値は母平均 μ に近づくことが知られている*1。しかし、最初の母集団の正規分布 $N(\mu, \sigma^2)$ とはバラツキ方(母分散) σ^2 が異なり、 $1/n$ だけ小さいバラツキ方 σ^2/n になることがすでにわかっている。つまり、 $N(\mu, \sigma^2/n)$ の正規分布となる(図17下グラフ)*1。すなわち、標本平均の母標準偏差 n は $\sqrt{\sigma^2/n} = \sigma/\sqrt{n}$ となり、 σ よりも $1/\sqrt{n}$ だけ小さくなる*2。図17では例として、ある一つの標本平均 \bar{X}_1 の位置を正規分布グラフに赤字で示している。

*1これはたいへん重要なポイントで、中心極限定理と呼ばれる。標本の大きさ n が大きいほど、標本平均 \bar{X}_i の平均値は母平均 μ に近づき、その分散は母分散 σ^2 の $1/n$ に近づく。 n が大きいほど σ/\sqrt{n} は小さくなるので、母平均の範囲が絞られてくることを意味する(図17下グラフ)。なお、 n が大きい場合、母集団が正規分布に限らず、正規分布から外れていても、その母集団からとった標本平均の分布は正規分布するというおもしろい性質がある。

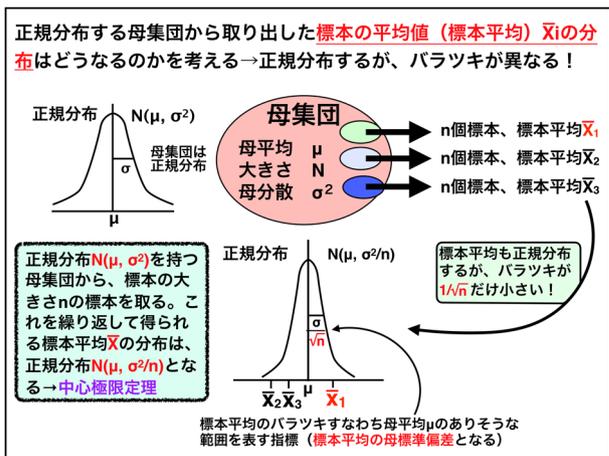


図17 ■ パラメトリック検定の基本

正規分布を標準正規分布に変換する(図18)

正規分布は μ と σ^2 が違えば分布が異なるので(図18左上と左下のグラフのように)、いろいろな母集団でい

*2実験動物の母集団を例にとれば、たとえばその中から6匹を標本として取り出して、測定したパラメータの標本平均を計算する。この操作を何度も繰り返して得られる多くの標本平均のバラツキは、個々に外れたデータがあっても平均化されるため、母集団のデータのバラツキよりも小さくなることは、予想できるであろう。

ちいち異なる正規分布を用いて考えるのは面倒である。そこで、母平均 μ が0(ゼロ)で、 σ^2 が1²の正規分布を標準正規分布と定義し、それぞれの異なる正規分布を標準正規分布に変換して考えることに決められている。そうすれば、どのような μ と σ^2 の母集団でも標準正規分布で考えることができる。図18の右のグラフが標準正規分布である*³。そこで、先ほどの標本平均 \bar{X}_1 (図17下グラフと図18左上と左下グラフ)を標準正規分布の値に変換することとする(これを標準化と呼ぶ)。まず、標準正規分布では母平均はゼロであるから、母平均 μ をゼロに移動するためには、 μ で引いてやればよい。すなわち、 $(\bar{X}_1 - \mu)$ である。一方、標準正規分布では母標準偏差は1であるから、 σ/\sqrt{n} を1に変換するためには σ/\sqrt{n} で割ってやればよい。すなわち、標本平均 \bar{X}_1 を標準化し、変換した値を Z_1 とすると、以下の式で表すことができる(図18中央の式)。

$$Z_1 = \frac{(\bar{X}_1 - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

そうすると、 \bar{X}_1 は図18右の標準正規分布グラフ中の Z_1 に変換される*⁴。このような標準化を行えば、異なる母集団の正規分布(図18の左上および下グラフ)の標本平均でも、同様に標準正規分布にデータ変換することができる。

このデータ変換はたいへん重要であるが少しわかりにくい。そこで、数式ではわかりにくいので、簡単な例を示す。図19を見て欲しい。今第1の母集団として、 $\mu = 10$ 、 $\sigma/\sqrt{n} = 2$ の正規分布があり、 $\bar{X}_1 = 12$ の標本平均が得られたとする(図19左上グラフ)。次に、第2の母集団として $\mu = 20$ 、 $\sigma/\sqrt{n} = 4$ の正規分布があり、 $\bar{X}_1 = 24$ が得られたとする(図19左下グラフ)。これらの数字を上式に代入すると、どちらも $Z_1 = 1$ が得られる(図19中央式と右グラフ)。これらの2つの異なる正規分布では、母平均 μ 、母標準偏差 σ/\sqrt{n} 、標本平均 \bar{X}_1 の値はそれぞれ異なるが、標準正規分布にデータ変換すると同じ値となり、標準正規分布での Z_1 の位置は、両母集団で同じであることがわかる。標本平均 \bar{X}_1 が標準正規分布中の Z_1 に標準化されたことを理解して欲しい。

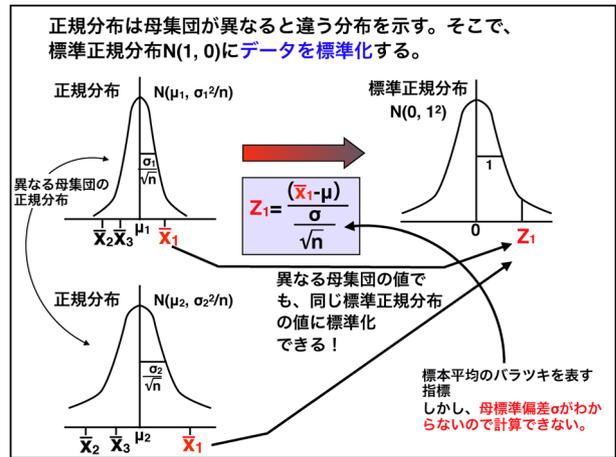


図18 ■ 標準正規分布へのデータ変換

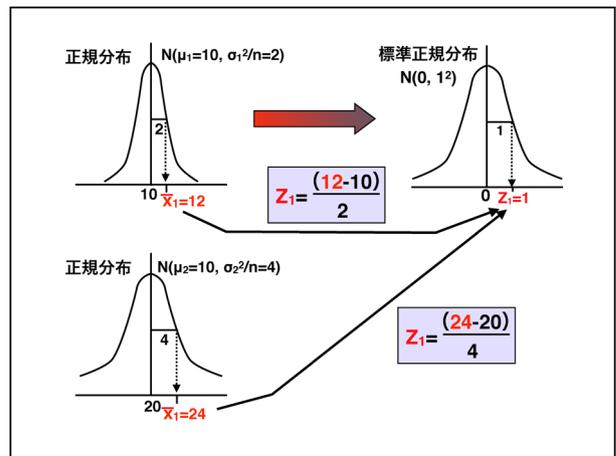


図19 ■ 標準正規分布への標準化の実例

母標準偏差 σ はわからないので、不偏標準偏差 u に置き換える(図20)

ここまで、正規分布している母集団から取り出した標本データの平均値すなわち標本平均の分布は正規分布することを述べてきたが(図20左上グラフ、これを $N(\mu, \sigma^2/n)$ とした)、バラツキ(母標準偏差)は σ/\sqrt{n} である。ここで、もともとの母集団の母標準偏差 σ は、生命科学系の研究では通常知ることができない値であり、このままでは、 σ/\sqrt{n} を計算することができない。では、どうすればよいのか?

ここで、前回「改訂増補版：I」で登場した不偏標準偏差 u を思い出して欲しい(前回の図12および13)*⁵。

*³前回でも述べたが、本セミナーのすべての図は筆者が模式的に作成したものであり正確なものではない。

*⁴この式でなぜ標準正規分布の値に変換されるのか不思議に思うかもしれないが、数学的な証明はなされている。

*⁵本総説では、不偏分散 u^2 をルートした値 u を不偏標準偏差と名付け、母標準偏差を推定する値として論じている。しかし、 u は真の不偏標準偏差ではない(「改訂増補版：I」の*¹³参照)。正確な不偏標準偏差とはズレがあり、特に、 n が10以下で、小さいほどズレが大きい。

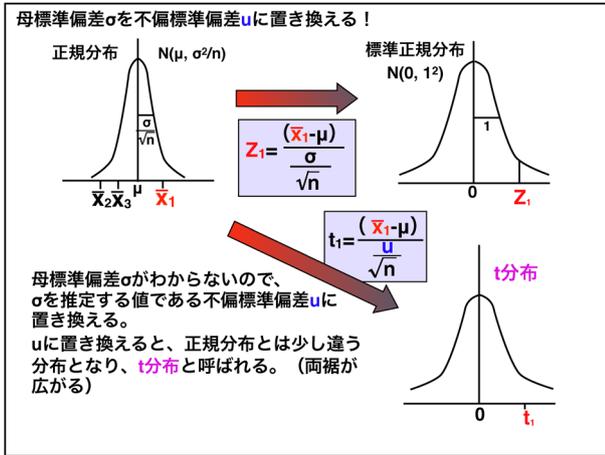
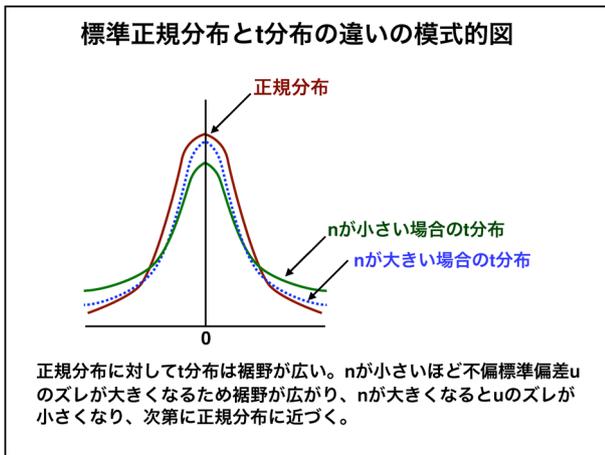
図20 ■ 母標準偏差 σ の不偏標準偏差 u への置き換え

図21 ■ 標準正規分布とt分布の違い

標本データから母標準偏差 σ を推定する値として計算できるのが不偏標準偏差 u である。そこで、母標準偏差 σ を不偏標準偏差 u に置き換えることとする。この計算値を t_1 とする。

$$t_1 = \frac{(\bar{X}_1 - \mu)}{\frac{u}{\sqrt{n}}}$$

そこで、何度も標本を取って標本平均 $\bar{X}_1, \bar{X}_2, \bar{X}_3 \dots$ および $u_1, u_2, u_3 \dots$ を計算し、さらに $t_1, t_2, t_3 \dots$ を計算してt値の分布を調べる。 σ を用いた場合、分布は正規分布するので標準正規分布に変換されたが(図20左上グラフから右上グラフへ)、 u に置き換えると、正規分布とは少し異なる分布になる(図20左上グラフから右下グラフへ)*⁶。この分布はt分布と名付けられた(図20右下グラフおよび図21のグラフ)。t分布のグラフは、正規分布よりも両裾が広がっていることに注意して欲しい。

不偏標準偏差 u は母標準偏差 σ を推定している。しか

し、 u は真の不偏標準偏差ではないため、標本の大きさ n の影響を受ける*⁶。 n が小さいと大きな影響を受け u は σ からのズレが大きく、値としては小さくなる(改訂増補版: Iの*¹³参照)。一方、 n を大きくすると、 u はあまり影響を受けず σ に近い値となる*⁵。 t_1 の式と図21のt分布グラフを見て欲しいが、 n が小さいと u は σ より小さな値となるため、計算されるt値はt分布の中心のゼロから離れた値となる場合が相対的に多くなる。そうすると、正規分布に比べてt分布の裾野が相対的に広がることとなる。一方、 n が大きくなると、 u は σ に近づくので、t値の分布は正規分布に近づいていく(図21のグラフ)。つまり、t分布は n が違っているとグラフの形が異なることを理解して欲しい*⁷。

ここでは標本平均の分布やバラツキを考えてきたことから、上の t_1 式の分母である u/\sqrt{n} は標本平均がどのようなバラツキ方をしているかを示しており、これが多くの研究者が用いる標準誤差(standard error; SE)である。 σ/\sqrt{n} は標本平均の母標準偏差であったので(図17)、標準誤差は標本平均の母標準偏差を推定する値となる。つまり、SEは母平均 μ がどのあたりにあるのか、すなわち、母平均 μ のありそうな範囲を表している。

SEを標本データのバラツキの一種と考えている読者がいるかもしれないが、それは間違いであり、SEには標本データのバラツキの意味はない。 t_1 の式の分母がSEであるから、上の式は以下のようにも書ける。

*⁶この点が、 u は“不偏”でないことを示している。“不偏”とは偏らないことで、偏りなく母集団を推定できることを意味する。もし、 u が“不偏”であれば、母標準偏差 σ を推定する値となるので、 σ を u に置き換えても正規分布となるはずである。ところが、 u は σ とはズレがあるために正規分布とはすこしずれたt分布となる。 n が小さいほどズレが大きくなるため、t分布は正規分布と大きくずれ、 n が大きいほど正規分布に近い分布となる(図21のグラフ参照)。

*⁷t分布は、実際は n ではなく自由度 $n-1$ に依存して形が変化する。自由度は何度か登場した。不偏分散の計算では、平方和を n で割るのではなく、自由度 $n-1$ で割ると母分散を推定できる値となることを述べた。今後は $n-1$ 以外の自由度も登場するが、どのような場合でも、自由度を用いることで母集団の情報を推定できると考えればよい。“バイオサイエンスの統計学”では、自由度とは、データのバラツキや偏りを予測する際に(つまり、分散や標準偏差を計算する際に)、他と独立して扱えるデータ数のことと述べている⁽¹⁾。たいへんわかりづらいが、たとえば、ある母集団から、標本を6個採取した場合、それぞれが関連のない独立した標本であれば、自由度は6である。しかし、すでに述べてきたように不偏分散を計算する際に、計算式に標本平均が入っている。式に標本平均があると、 n が6の場合、5個のデータがあれば、6個目のデータは標本平均 $\times 6$ から(5個の標本データの合計)を引き算すれば求まる。つまり、独立して扱えるデータ数は5であり、6個目は自動的に決まってしまう自由に動けない。したがって、自由度は1減ってしまい、5となる。つまり、標本の大きさを n とすると自由度は $n-1$ となるのである。

$$t_1 = \frac{(\bar{X}_1 - \mu)}{SE}$$

つまり、この式は、母平均 μ のありそうな範囲である標準誤差SEを基準（分母）にして、標本平均 \bar{X}_1 が母平均 μ からどれ位離れているか（分子）を計算しているのである。たとえば、 $\bar{X}_1 - \mu > 0$ と仮定して、分母のSEが小さければ、分子の $(\bar{X}_1 - \mu)$ は相対的に大きくなるので t_1 値はゼロから離れる。逆に、SEが大きければ、 $(\bar{X}_1 - \mu)$ は相対的に小さくなるので t_1 値はゼロに近づく。この感覚を理解して欲しい。

標本平均±SEは何を意味しているのか？（図22）

標本データの平均値（標本平均）は母平均 μ を推定する値と位置づけられる。そして、SEは母平均 μ のありそうな範囲を示している（図22）。SEの式を見ると、SEは標本の大きさ n が大きくなればなるほど、小さくなるのがわかる。これは、 n が増えれば増えるほど、母平均 μ の存在する範囲が絞り込まれることを意味する。 n が増えて、母集団の大きさ N に近くなればなるほど、 μ が絞り込まれていくことは感覚的に理解できるのではないだろうか*8。

標本平均±SDと標本平均±SEは何が違うのか？（図23）

標本平均±SDについては、前回の図15でも記述したように、研究者は母平均 μ と母標準偏差 σ の両方に関心があるはずである（図23）。それでは、標本平均±SEではどうであろうか？ SEは母平均 μ のありそうな範囲を示していることから、母平均 μ に関心が集中しているのであり、母標準偏差 σ には原則的には関心がないことになる。母平均 μ をSEによりどこまで絞り込んだかが焦点である（この点は『らくらく生物統計学』⁽²⁾に詳しく述べられている）。

実例を図23中段に示している。ある機能性成分が母平均 μ を変化させるかどうかに関心があるのであれば、SEを使えばよい。図23下段に示すような例では、SDが適している。しかし、学会発表や論文を見るかぎり、SEとSDを上記の観点から使い分けている研究者は少ないように思われる。SDを使うか、SEを使うかは研

*8なお、学会発表において、スライドに標本の大きさ n や標本平均±SEなのか±SDなのか表記していない発表が数多く見受けられる。これらは重要な情報であり、その表記は研究者としての基本である。書き忘れたで済むものではない。

標本平均±SEが表しているのは、

標本平均 \bar{X} で母平均 μ を推定。SEは母平均 μ のありそうな範囲を表す。

$$SE = \frac{u}{\sqrt{n}}$$

n を増やせば増やすほど、SEは小さくなる。（母平均 μ が絞り込まれる）

図22 ■ 標本平均±SEの意味

標本平均±SD

研究者は母平均 μ と母標準偏差 σ の両方に関心を持っている。

標本平均±SE

母平均 μ に関心が集中しており、母標準偏差 σ には関心がない。母平均 μ を標本平均±SEでどこまで絞り込んだかに関心が集中。

ある食品成分の有効性を調べる場合

その食品成分が、ある測定パラメータの母平均 μ を変化させるかどうか最大の関心であるなら、SEが適している。

以下の場合、SDも重要な情報となる。

ある食品成分は、あるパラメータの母平均 μ を変化させるだけでなく、母標準偏差すなわちバラツキも変化させる。その原因を知りたい。

らくらく生物統計学：中山書店

図23 ■ 標本平均±SDと標本平均±SEの違い

研究室毎に決まっているのではなく、あくまでも何が知りたいのかで決まるのである。このセミナーの読者は、SDとするのかSEとするのか、よく考えて使い分けて欲しい。

2群の差の検定（Studentのt検定）（図24、25）

ここからいよいよ群間比較、群間の有意差検定に入り、「有意差」について説明していく。生命科学系の研究では、2群あるいは3群以上を設定して、群間比較を行う実験が多いのではないだろうか？ その基本となるのが、Studentのt検定である。t検定は2群のパラメトリック検定であり、t分布を用いることからその名がある。しかし、統計書を読んでも、計算方法はわかるが、その原理は簡単には理解できない場合が多い。かなり複雑であるが、t検定の原理をマスターすれば、そのほかの多様な検定も、実は基本的な手順は同じであることか

ら、理解しやすいはずである。

詳しくは後述するが、 t 検定には主に対応のない独立2群の検定 (unpaired t test) と対応のある関連2群の検定 (paired t test) がある。ここでは、unpaired t test を例として詳述する。

図24の右上図に、2群の標本平均 \bar{X}_1 と \bar{X}_2 が示されている。この2群の間に「差がある」かどうかは、基準がないと決められない。たとえば、 \bar{X}_1 と \bar{X}_2 間に10%以上差があれば差があることにしようという基準である。しかし、「改訂増補版：I」の図4で説明したように、差があるかどうかは標本平均の差だけでなくデータのバラツキ方によっても判断が変わるので、基準を決めることは困難である。そこで統計学では、「差がない」という仮説から考えていくことに決まっている。この「差がない」という仮説を帰無仮説と呼ぶ。これに対して「差がある」という仮説を対立仮説と呼ぶ。「差がない」ことから考えていく考え方は、群間比較を行う検定においては基本的考え方であり、共通しているので、記憶して欲しい。

そこで、考え方として、同一母集団から標本をとると考える (図24下図)。たとえば、今一つの母集団から、それぞれ n_1 および n_2 個の標本を取り2群とする。この場合、2群のそれぞれ個々の標本には対応 (関連) はなく独立しているので、対応のない独立2群の検定 (unpaired t test) となる。2群それぞれの標本平均を \bar{X}_1 と \bar{X}_2 とする (図24)。このように標本を取れば、もともと同一母集団であるから、本来は差はないはずであるが、標本なので標本平均 \bar{X}_1 と \bar{X}_2 は必ずしも一致はしないという状況となる。

ここで、便宜的に、2つの標本 \bar{X}_1 と \bar{X}_2 の母集団の母平均をそれぞれ μ_1 と μ_2 とし、母分散を σ_1^2 と σ_2^2 とする。(同じ母集団からの標本であるから、実際は $\mu_1 = \mu_2$ および $\sigma_1^2 = \sigma_2^2$ であるが)。ここで、帰無仮説は「差がない」とするので $\mu_1 = \mu_2$ となる。これに対して対立仮説は、「差がある」とするので、 $\mu_1 \neq \mu_2$ となる。

母集団から n 個の標本を取って標本平均 \bar{X}_i を得たとして、これを無限に繰り返したとき標本平均は正規分布し、母分散は σ^2/n 、母標準偏差は σ/\sqrt{n} となることはすでに述べた (図17下グラフ)。それでは、ある正規分布する母集団から標本データを n_1 および n_2 個取り出し、標本平均 \bar{X}_1 と \bar{X}_2 を求める。そして、標本平均の差 ($\bar{X}_1 - \bar{X}_2$) を求めることとして、この操作を何度も繰り返して ($\bar{X}_1 - \bar{X}_2$) の分布を求めると、どうなるであろうか？ 実はこの場合も正規分布するのである (図25左上グラフ)。($\bar{X}_1 - \bar{X}_2$) は同じ母集団からの標本平均の差

統計検定での重要なポイント

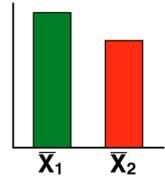
統計検定では、差があるという考え方はせずに、差がないということから考えていく。(差があるという判断基準がない！)

従って、各群の標本データは、同じ母集団から得られたと仮定する

→本来、2群間には差がないと考えられる。

正規分布する母集団

母平均 μ
母集団の大きさ N
母分散 σ^2



n_1 個標本、標本平均 \bar{X}_1
 n_2 個標本、標本平均 \bar{X}_2

図24 ■ 統計検定の重要なポイント

同じ母集団から、 n_1 個と n_2 個の標本を2回とり、その標本平均 \bar{X}_1 と \bar{X}_2 の差をとる。すると、 $\bar{X}_1 - \bar{X}_2$ もまた正規分布する！ (左図)

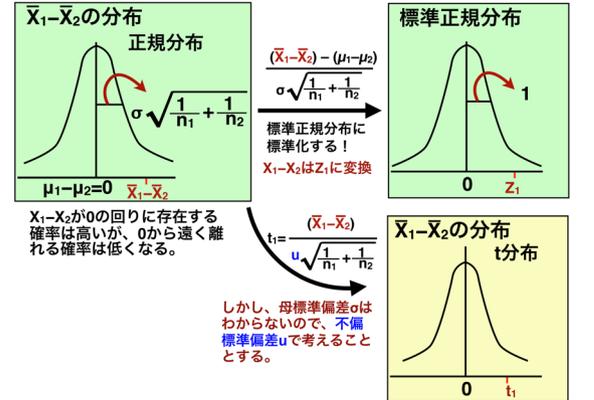


図25 ■ 2群の標本平均の差も正規分布する！

なので、ゼロを中心にはばらつくと予想される。すなわち、正規分布の中心は $\mu_1 - \mu_2 = 0$ である。($\bar{X}_1 - \bar{X}_2$) を一つの母集団としたときの母分散は、それぞれの群の母分散 σ_1^2/n_1 および σ_2^2/n_2 を足した値となる。すなわち、

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

これをルートすることで、母標準偏差が計算できる。

すなわち、母標準偏差は、

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

となるが、もともと同一母集団なので $\sigma_1 = \sigma_2$ であるから、これらを σ とすると、以下の式となる (図25左グラフ)。

$$\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

そこで、標本平均の差 ($\bar{X}_1 - \bar{X}_2$) がこの正規分布のどこにあるかを考えるが、ここで、図18で説明した、標準正規分布へのデータ変換 (標準化) を思い出して欲しい。

($\bar{X}_1 - \bar{X}_2$) を標準正規分布上の Z_1 に変換すると以下の式が出現する (図25中央上の式)。

$$Z_1 = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

この式が出現することがピンとこないようであれば、「正規分布を標準正規分布に変換する (図18)」の項を読み返して欲しい。

ここで、 $\mu_1 - \mu_2 = 0$ であるから、結局、次の式が得られ、標準正規分布の値に変換される (図25右上グラフ)。

$$Z_1 = \frac{\bar{X}_1 - \bar{X}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

ここで、 σ はもともとの母集団の母標準偏差であり、研究者は知ることができない。そこで、図20で説明したことと同様の理由で、母標準偏差を推定する値である不偏標準偏差 u を用いることとする。このときの計算値を t_1 とすると、次式となる (図25中央下式)。

$$t_1 = \frac{\bar{X}_1 - \bar{X}_2}{u \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

さて、ここで不偏標準偏差 u はどのように計算するのであろうか？ 不偏標準偏差 u の計算方法は前回の図12でも説明したが、ここでは図26上の式を見て欲しい。簡単には、平方和を自由度 $n-1$ で割り算し不偏分散を求めルートした。しかし、ここでは2群あるので、2群両方のバラツキを考慮すべきである。そこで、2群の平方和を合算し、合算した自由度で割り算して、ルートすればよい (図26最下の式)。

母集団から n 個の標本を取って標本平均 \bar{x}_i を得たとして、これを無限に繰り返したときの標本平均の分布は、不偏標準偏差 u を用いると t 分布することはすでに述べた (図20右下グラフ)。このことと同様に、上式で計算した t_1 は、この計算を無限に繰り返すと t 分布するので、 t 分布中の値に変換される (図25右下グラフ)。 t

$\bar{X}_1 - \bar{X}_2$ の不偏標準偏差 u の計算方法

第1群のデータを x_1 から x_5
第2群のデータを x_6 から x_{10} とする。

$$u_1^2 = \frac{(x_1 - \bar{X}_1)^2 + (x_2 - \bar{X}_1)^2 + (x_3 - \bar{X}_1)^2 + (x_4 - \bar{X}_1)^2 + (x_5 - \bar{X}_1)^2}{n_1 - 1} \rightarrow \text{平方和1とする}$$

$$u_2^2 = \frac{(x_6 - \bar{X}_2)^2 + (x_7 - \bar{X}_2)^2 + (x_8 - \bar{X}_2)^2 + (x_9 - \bar{X}_2)^2 + (x_{10} - \bar{X}_2)^2}{n_2 - 1} \rightarrow \text{平方和2とする}$$

合算の不偏分散

$$u^2 = \frac{\text{平方和1} + \text{平方和2}}{(n_1 - 1) + (n_2 - 1)} \quad \text{t検定では2群の平方和を合算！} \\ \text{自由度も合算！}$$

合算の不偏標準偏差

$$u = \sqrt{\frac{\text{平方和1} + \text{平方和2}}{(n_1 - 1) + (n_2 - 1)}}$$

図26 ■ ($\bar{X}_1 - \bar{X}_2$) の不偏標準偏差 u の計算

分布については、図21の説明で詳しく説明した。

この t_1 値が t 分布のどこに位置するかで、有意差を決めていくことになるのである。ここまでの内容をしっかり理解したうえで、先に進んで欲しい。

ここまで、2つの標本平均の差 ($\bar{X}_1 - \bar{X}_2$) の分布は不偏標準偏差 u を用いると t 分布することを述べてきた。そして、ある2群の実験を行って得られたデータから、 t_1 を計算した。この t_1 が t 分布のどの位置にあるのかを調べる (図25右下グラフの t_1 の位置を調べる)。 t 分布はゼロを中心に左右にばらついている。 \bar{X}_1 と \bar{X}_2 に大きな差がなければ、($\bar{X}_1 - \bar{X}_2$) はゼロあたりにあるが、たまたま大きな差があれば、ゼロから離れる。しかし、 t 分布のグラフからわかるように、ゼロから大きく離れる確率は低く、滅多には起こらない。そこで、 t_1 が t 分布のかなり外れたあたりにくると、それは滅多に起こらないまれなことが起こったのであるから、差があることにしてしまおうというのが、「有意差がある」という決め方になっている。つまり、同じ母集団からの2つの標本平均の差 ($\bar{X}_1 - \bar{X}_2$) を、「差がない」という帰無仮説 $\mu_1 = \mu_2$ で考えてきたが、「差がない」と考えるには、あまりにも外れたあたりにあるので、帰無仮説 $\mu_1 = \mu_2$ を棄て (棄却し)、差がある $\mu_1 \neq \mu_2$ と考えることにしてしまおうという考え方である (対立仮説を採用することにする)。この「差がない」ことから考えてきたが、あまりにも差が大きいので「差がある」ことにしてしまおうというのが、群間比較の検定の基本的な考え方であるので、記憶して欲しい。

ここで重要なポイントが見える。つまり、2群間に「有意差がある」は、「真に差がある」ことを意味している訳ではないことである。多くの研究者は標本で研究している。標本で研究しているかぎり、たとえ有意差が

あっても、差があるとは断定できないのである。したがって、標本を用いた1回の試験だけでは真実かどうかはわからない。研究者はいろいろな角度から研究し真実を追究すべきである。

なお、 t 分布は自由度 $n-1$ によって分布の形が変化することはすでに述べた(図21と*7)。ここでは2群あるので、自由度は図26で述べたように合算して $(n_1-1)+(n_2-1)$ となり、 n_1+n_2-2 である。したがって、自由度 n_1+n_2-2 の t 分布で考えることとなる。

減多に起こらないまれな確率の決め方(図27)

それでは、減多に起こらないまれな確率をどの程度にとればよいのであろうか? 図27の上にも書いているように、 t 分布全体の面積の両端それぞれ2.5%または0.5%の合計(あるいは片端、これについては後述する)の5%または1%とすることに決められている(母集団全体を1とすれば、0.05あるいは0.01)*9。この判定基準を有意水準と呼ぶ。標本平均 \bar{X}_1 と \bar{X}_2 はどちらが大きな値になるのかわからない場合は、 $(\bar{X}_1-\bar{X}_2)$ はプラスになるかマイナスになるかわからないので、 t 分布の左右両端が設定され、それぞれ2.5%ずつあるいは0.5%ずつに t 値が入ってくると、まれなことが起こったと判定するのである。しかし、減多に起こらないまれなことといっても、もともと「差がない」ことから考えてきたのであり、「差がある」と断言するのは危険である。そこで、有意水準は危険率とも呼ばれる。なお、帰無仮説は正しい、つまり差がないのに、帰無仮説を棄却する、つまり差があるとしてしまう誤りを犯すことはありうる。この誤りを第一種の過誤と呼ぶ*10。

最近では、 t 値がわかればパソコンで t 値よりも両端の面積を計算してくれる。 t 分布の面積を1としたときのこの両端の面積を p 値と呼ぶ(図28)。したがって、有意水準(危険率)は p 値が0.05(5%)あるいは0.01(1%)である。 p はprobabilityの頭文字であり、 p 値とは観察された差が偶然生じる可能性を示す尺度ということになる。たとえば、 $p=0.005$ とは、観察された差が偶然起こるのは0.5%、つまり200回に1回であることを示す。偶然に起こるのが200回に1回であれば、偶然に起こったのではなく、ある意味をもって(有意に)起こ

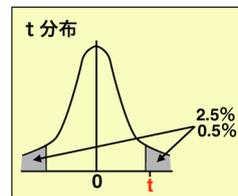
*9 この5% (0.05) を決めたのは、統計学では有名なRonald Fisherと言われている。実は、0.05に科学的根拠はないのである。Fisherが0.07と決めれば、そう決まったかもしれない。

*10 第一種の過誤に対して、差があるのに対立仮説を棄却して、差がないとしてしまう誤りを、第二種の過誤と呼ぶ。これらは統計書によく登場し、わかりにくい単語である。

減多に起こらない稀な確率をどの程度にとるか?

一般的に、 t 値が t 分布の両端または片端の5または1%とすることに決められている。(分布の面積を100とする)
 t 分布の面積を1とすると、0.05または0.01と表記する。
この判定基準を有意水準と呼ぶ。

t 値が t 分布の両端2.5%または0.5%に入ってくると、稀であると判定する。
($\bar{X}_1-\bar{X}_2$ はプラスになるかマイナスになるかわからないので両側をみる)
しかし、もともと同じ母集団から考えてきたので、稀でも起こりうることであるから、帰無仮説($\mu_1=\mu_2$)は正しいのかもしれない(正しい危険性がある)。従って、有意水準を危険率とも呼ぶ。



有意水準(危険率)5%(1%)以下で有意差ありと記述する。

帰無仮説は正しいのに、帰無仮説を棄却する誤りを第一種の過誤と呼ぶ。(本当は有意差がないのに、有意差があると結論してしまう誤り)

図27 ■ 有意水準(危険率)とは?

t 値がわかれば、 t 値よりも両端の面積を計算できる。

t 分布の面積を1としたときの、この面積を p 値と呼ぶ。

p (probability)値とは、観察された差が偶然生じる可能性を示す尺度

例えば、 $p=0.005$ とは、観察された差が偶然起こるのは、0.5%、つまり200回に1回であることを示す。偶然に起こるのが200回に1回であれば、偶然に起こったのではなく、ある意味を持って(有意に)起こった可能性がかなり高いと判断する。

また角度を変えると、 p 値は間違った結論を出してしまう確率でもある。

有意水準5%は、20回に1回は有意差がないのに有意となる可能性のある確率。
1回目に起こるかもしれない!
20回目に起こるかもしれない!

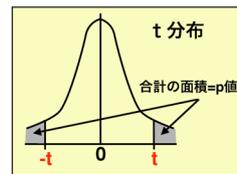


図28 ■ p 値とは?

た可能性が高いと判断する。これを「有意差がある」と表現する。

有意水準5%で考えると(図28左下枠内)、有意水準5%は同じ実験を20回行くと、1回程度は有意差がないのに有意となる可能性のある確率となる。それは1回目に起こるかもしれない! 20回目に起こるかもしれない! 実験は通常1回しか行わない。もし、1回目に起これば、本当は有意差がないのに有意差ありと判定されることになる(第一種の過誤)。研究者はこのことを念頭に置いておかねばならない。1回の実験だけで結果を論文化する危うさはここにある。このことは先程述べたように、真実を明らかにするためには角度を変えた研究を行い確認を取ることが重要である。

t_1 の計算式の分母の式をみると、

$$u \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

n が大きくなればなるほど、この値が小さくなり、 $\bar{X}_1 > \bar{X}_2$ と仮定すると、 t_1 値は大きくなるのがわかる。 t 値が大きくなると、 t 分布の端のほうに寄るので、有意となる可能性が高まる。したがって、 t 検定では、標本データの大きさ n を増やせば有意差は出やすい。微妙な差しかない、あるいは、バラツキが大きいことが最初からわかっているパラメータの場合に、 n を増やすことは有益な手段である。

$p < 0.05$ をどのように表現するのか？

有意差をどのように表現するのかは図29に記載したとおりである。有意差があると、「差があった!」と表現することが多いが、これまで述べてきたように、「差がある」という表現は適切ではない。正確には、「有意水準5%未満で統計的に有意差がある」が正しい。せめて、「(統計的に)有意差がある」という表現を使うようにしたい。また、 $p > 0.05$ の場合では、「(統計的に)有意差はない」が妥当な表現である。

母集団の情報わかっている場合の検定で統計的有意差の意味を再度理解する!

読者は統計的有意差の意味が理解できたであろうか? かなり複雑な内容であったため、いま一つピンと来ていないかもしれない。そこで、ここでは、生命科学研究者が用いることはほとんどない母集団の情報(母平均 μ と母分散 σ^2)がわかっている場合の検定について説明し、有意差検定の理解を深めたい。すでに有意差の意味を理解できた読者は、おそらく簡単に理解できるはずである。

筆者が作った架空の母集団であるが、日本人男性の身長分布を正規分布する母集団と考え、身長の平均(母平均)が $\mu = 170\text{cm}$ で、母標準偏差 $\sigma = 10$ であることがわかっていると仮定して考える(図30上)(通常生命科学系研究ではこの部分が不明であり、それを知らたいがために研究している)。日本人男性は6,000万人いるとする。分布は正規分布であるから、170cm前後の人が多く、身長がかなり高いおよび低い人の人数は減っていくというベル型である(図30左上グラフ)。ここで、170cm前後のヒトはたくさんいるので、「日本人男性と同等と判断する」。一方で、身長が正規分布の両側のかなり外れたあたりにあると、170cmからかなり外れて

$p < 0.05$ の表現

「有意水準(危険率)5%未満で統計的に有意差がある」と表現する。

もともと、同じ母集団からの標本データで考えてきたので、本来、差はないのである。しかし、差がないと考えるにはあまりにも外れているから、差があることにしようという考え方である。従って、「差がある」という表現はよくない。

$p > 0.05$ の表現

「統計的に有意差はない」という表現が妥当。

図29 ■ $p < 0.05$ の表現は？

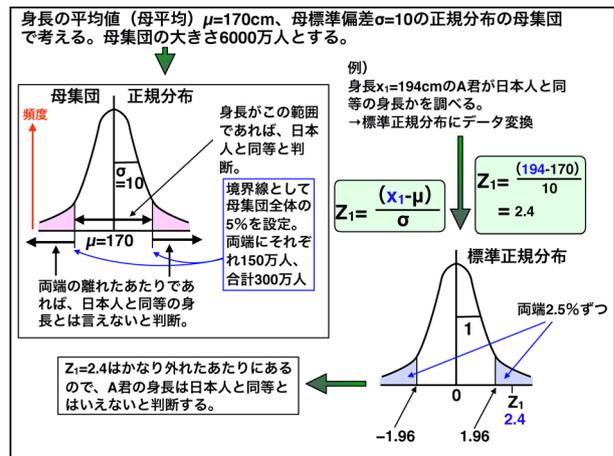


図30 ■ 母集団の情報わかっている場合の統計検定の考え方(1)

おり、そのような身長の日本人は減多にいないので、「日本人と同等とは言えない」と判断する。これが有意差がない、あるいは、あることを意味する統計検定の原理である。

これまで、統計検定とはもともと差がないところから考えていく。しかし、差がないというにはあまりにも差が大きいため、差があることにしてしまおうと考えると説明してきた。ここでもこの考え方は同様である。もともと日本人である。しかし、日本人と考えるにはあまりにも背が高い(低い)ので、日本人とは言えないことにしてしまおうという考え方である。

ここで、日本人と同等か、同等とはいえないかを判断するための境界線として考えられたのが、有意水準の5%あるいは1%である。6,000万人の5%は300万人である。両端それぞれで2.5%は150万人に相当する。かなり背が高い(低い)人で、この両端150万人に入ってい

れば、あまりにも背が高い（低い）ので、日本人とは同等の身長ではないことにしてしまおうと考える。これが統計検定での「有意差がある」ことを意味する。

さらに実例をあげると、今ここに、身長 $x_1=194$ cmのA君がいる。A君が日本人と同等の身長かどうかを調べたい（A君は日本人かどうか不明とする）。そのために、194cmがこの正規分布のどのあたりにあるのかを調べることとする。図30左の正規分布で、194cmがどのあたりにあるかを調べることもできるが、図18で述べたことと同様に、 $\mu=0$, $\sigma=1$ の標準正規分布に当てはめ、データを標準化することに決まっている（図30右下）。平均値をゼロへ座標移動するために母平均 μ を引き、 σ を1とするために、 σ で割り算する。すると、以下の式ができる（図30右中央式）。

$$Z_1 = \frac{(x_1 - \mu)}{\sigma}$$

$\mu=170$ cm, $\sigma=10$ であるから、

$$Z_1 = \frac{(194 - 170)}{10} = 2.4$$

となる。

標準正規分布で、両端2.5%に相当するZ値は、1.96（あるいは-1.96）と決まっている。計算値2.4は、1.96よりもさらにゼロより遠い位置にあり（図30右下グラフ）、身長194cmのA君はかなり外れた身長であることがわかる。したがって、A君は日本人の身長としてはかなり外れているので、日本人と同等の身長とは言えないと判断する（図30左下）。ここで、「A君の身長は日本人と同等である」は帰無仮説となり、「A君の身長は日本人と同等とは言えない」は対立仮説となる。A君の場合は、帰無仮説を捨て、対立仮説を採用する。これが有意差検定である。

このように、A君は日本人にしてはあまりにも身長が高いので、日本人とは同等の身長ではないことにしてしまおうと考えたのであって、決して日本人と同等ではないと断定したわけではない。日本人でもこの身長の人存在する。実はA君が日本人であることはありうる。A君が日本人ではないとは、決して断定できないのである。

A君の場合は、標本として $n=1$ であったが、標本が複数の場合はどうなるのであろうか？たとえば、日本全体の大学生の平均身長が、 $\mu=170$ cmで、母標準偏差 $\sigma=10$ であることがわかっていると仮定する（母集団）（図31）。Y大学の無作為抽出した学生 $n=25$ の平均身長

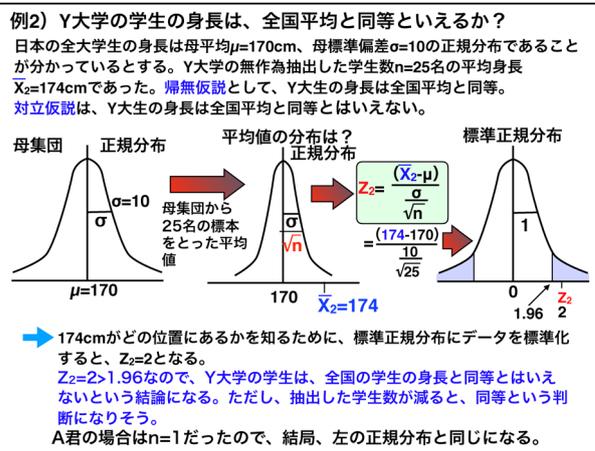


図31 ■ 母集団の情報わかっている場合の統計検定の考え方(2)

は、 $\bar{X}_2=174$ cmであった。Y大学の学生の身長は、全国平均よりも大きいと言えるかという設問があったとする。ここで、帰無仮説は全国平均と同等である。対立仮説は全国平均と同等とは言えないと設定する。

ここでは、 $n=25$ の標本平均174cmが焦点である。これは、日本人学生の母集団から25名を標本としてとった平均値（標本平均）である。そうすると、25名の標本を取ることを繰り返して得られた標本平均の分布で考える必要がある。これは、図17の設定と同じなので思い出して欲しい。標本平均の分布は正規分布し、母平均 μ を中心にバラつく。そのバラツキ（母標準偏差）は

$$\frac{\sigma}{\sqrt{n}}$$

であった（図17および31中央グラフ）。この正規分布で、 $\bar{X}_2=174$ cmがどこにあるかを考えるが、何度も説明してきたように、ここでも標準正規分布に標準化する（図18および31の式）。そうすると、変換式は以下となる。得られた値を Z_2 とすると、

$$Z_2 = \frac{\bar{X}_2 - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{174 - 170}{\frac{10}{\sqrt{25}}} = 2$$

となる（図31右グラフ）。

そこで、 $Z_2=2$ が標準正規分布のどの位置にあるかを調べる。先程記述したように、両端2.5%のときのZの値は、1.96であるから、2は1.96よりも僅かに大きいので、Y大学の学生の身長は、ぎりぎり全国平均と同等とは言えず、有意に背が高いという結論となる。統計学ではこのような考え方をすることを再度認識して欲しい。

ここでもし標本の大きさ n が25名よりも小さくと、

上式の分母 $\frac{\sigma}{\sqrt{n}}$ は大きくなるため、 Z_2 の計算値は 1.96 よりも小さくなる。そうすると統計的有意差はなくなり、Y大学の学生は全国平均と同等と言えるという結果になる。つまり、標本の大きさ n が小さくなると結論が変わる。 n が小さくなると、情報の正確さに欠けてくるので、全国平均と同等と言わざるをえなくなってくるのである。

ここで、先程の A 君 $n=1$ の例に戻るが、 Z の変換式は以下であった。

$$Z_1 = \frac{(x_1 - \mu)}{\sigma}$$

一方、Y大学の学生 25名の例では、

$$Z_2 = \frac{x_2 - \mu}{\frac{\sigma}{\sqrt{n}}}$$

であった。この Z_2 の式に $n=1$ を入れれば、 Z_1 と同じ式であることに気づく。すなわち、図 31 の中央グラフは、 $n=1$ の場合は、左グラフと一致する。

ここでは、母集団の情報があるので、標準正規分布で検定を行ったことを理解して欲しい。すでに述べた t 検定では、母標準偏差 σ がわからないために、仕方なく不偏標準偏差 u を用いざるをえず、そのために、 t 分布が登場した (図 25)。違いはこの部分のみである。

対立仮説の立て方で検定結果は異なる！ (図 32)

これまで 2 群の実験で、同一母集団からの標本で 2 群とした。2 群それぞれの母平均は便宜的に μ_1 と μ_2 とした (同一母集団であるから $\mu_1 = \mu_2$ である)。帰無仮説として母平均 μ_1 と μ_2 は差がない、すなわち、 $\mu_1 = \mu_2$ とした。 μ_1 と μ_2 はどちらが大きな値になるかは、通常はわからない。そこで、対立仮説は $\mu_1 \neq \mu_2$ (すなわち $\mu_1 - \mu_2 \neq 0$) とした。標本平均 \bar{X}_1 と \bar{X}_2 で考えると、 $(\bar{X}_1 - \bar{X}_2)$ はプラスかマイナスかわからない。したがって、 t 分布の両端の 2.5% ずつの範囲に入れば、有意水準 5% で有意差ありとする。このような検定を両側検定と呼ぶ (図 32)。未知の機能性成分の影響を調べる場合では、通常 \bar{X}_1 と \bar{X}_2 はどちらが大きくなるかわからないので、両側検定とすべきである。しかし、前もって μ_1 と μ_2 のどちらかが大きいという十分な情報がある場合は、 $\mu_1 > \mu_2$ あるいは $\mu_1 < \mu_2$ と対立仮説を立てることも可能である。具体的な例を図 32 左下に記述している。もし、 $\mu_1 < \mu_2$ と対立仮説を立てることができれば、 $(\bar{X}_1 - \bar{X}_2)$ はマイ

対立仮説の立て方は重要

$\mu_1 - \mu_2 \neq 0$ (この場合どちらが大きいかかわからないので、両側検定) ($\mu_1 \neq \mu_2$) となる。

$\mu_1 > \mu_2$ (どちらかが大きい (小さい) と立てると片側検定となる) ($\mu_1 < \mu_2$)

未知の成分の機能性を調べる場合、両側検定が妥当。

前もって、 $\mu_1 > \mu_2$ あるいは $\mu_1 < \mu_2$ であるという情報がある場合は、片側検定もありうる。

例えば、植物ステロール含有機能性食品を開発する場合

植物ステロールの血漿コレステロール濃度低下作用は十分立証されているので、「植物ステロール摂取群で血漿コレステロール濃度は低い」と対立仮説を立てることが可能 → 片側検定となる！片側検定のほうが、有意差は出やすい。しかし、余程の確証がない限り、片側検定は用いるべきではない。

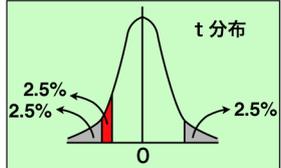


図 32 ■ 対立仮説の立て方

ナスとなることが最初から期待できることから、図 32 の t 分布グラフのマイナス側だけを考えればよいので、マイナス側に 5% を設定できる。このような検定を片側検定と呼ぶ。図 32 右下図を見るとわかるように、片側検定では両側検定よりもマイナス側の t 分布の面積が 2 倍となるので、有意となりやすくなることがわかる。

片側検定を使えるのは十分な証明のある場合である。図 32 に片側検定の一例を示した。しかし、一般的な生命科学系の研究では十分な立証がないから研究を行っている場合が多いので、生命科学系の研究において片側検定を用いることはほとんどない。なお、ある試験物質で 1, 2 回実験を行ってみて両側検定で有意な影響が得られたからといって、それ以降の少し条件を変えた実験において、片側検定を用いて有意差を得やすくするといった判断は許されることではない。あくまでも、 μ_1 と μ_2 のどちらかが大きいという十分な情報がある場合に限られる。原著論文において、一般的には両側検定か片側検定かの記述はないが、常識的には両側検定を行っているはずである。統計ソフトでは、片側か両側かをチェックして検定に進む場合が多いので、間違えないようにしなければならない。もし、チェックせずに進む統計ソフトがある場合は普通は両側検定が行われる。

2 群のパラメトリック検定の流れ

図 33 に 2 群のパラメトリック検定の流れを記載している。これまで述べてきたのは、対応のない独立 2 群の検定 unpaired t test であり (図 33 左端の上から下の流れ)、同一母集団からの標本なので、母分散は等しい、すなわち $\sigma_1^2 = \sigma_2^2$ と仮定した (2 群の差の検定 (Student の t 検定の項参照))。しかし、標本として 2 群をとって

いるので、たまたま、分散が等しいとはいえない場合も起こりうる。分散が等しい、すなわち、等分散かどうかは、等分散性の検定を行い、等分散と判定されれば、unpaired t testを行い、等分散と判定されなければ、Welchの検定を行うというのが定番で、多くの統計書にそのように書かれている（図33左上から下への流れ）。しかし、Welchの検定は、等分散でない場合だけでなく、等分散かどうか分からない場合でも正確な検定ができることから最近では推奨されている。一方、unpaired t testは、等分散から外れると検定が不正確となる。つまり、Welchの検定を用いるのであれば、等分散性の検定を行う必要がない。unpaired t testは2群の検定の定番のように言われてきたが、 t 検定を指定すると、実際はWelchの検定が実行される統計ソフトも現れている。まず、等分散性の検定について説明する。

等分散性の検定（F検定）の原理（図34）

すでに述べてきたが、unpaired t testは、同一母集団から2群の標本を取り、それぞれの母分散を便宜的に σ_1^2 と σ_2^2 としたとき、同一母集団であれば当然 $\sigma_1^2 = \sigma_2^2$ であるから、2群の分散は等しい（等分散）として検定が考えられている。しかし、実際には、得られるのは標本データであり、母分散はわからないので、2群の不偏分散 u_1^2 と u_2^2 で考えざるをえない。標本データであるから、 u_1^2 と u_2^2 が必ずしも近い値になるわけではなく、かけ離れてしまうこともありうる。そうすると、unpaired t testではうまく検定できない。そこで、 u_1^2 と u_2^2 がかけ離れているかどうかを検定する方法として等分散性の検定がある。この検定は、**F検定**とも呼ばれる。

図34上図にあるように、同じ母集団から2群を取り出し、それぞれの不偏分散 u_1^2 と u_2^2 を計算する。等分散であれば、 $u_1^2 = u_2^2$ を考えればよい、 $u_1^2 = u_2^2$ であれば、 u_1^2/u_2^2 は1になる（この値をFとする）。 u_1^2 と u_2^2 がかなり離れた値なら、 u_1^2/u_2^2 は1から外れた値になる。これを利用する。そこで、母集団から2群を取って、F値を計算することを何度も繰り返してプロットすると、ある分布が出現する（図34右グラフ）。これを**F分布**と呼ぶ^{*11}。 u_1^2/u_2^2 の分布であるから、マイナスはありえない。Fは1前後になる可能性が最も高いので、1をピークにして、1から離れるほど少なくなる分布となる。そ

*11 F分布は自由度により形が変動する。ここでは2群あるので、第1群目を分子とし、2群目を分母とすると、自由度 $n_1 - 1$ と $n_2 - 1$ のF分布となる。なお、図34のF分布の図は筆者が適当に描いたもので正しい図ではない。

2群のパラメトリック検定

母集団の情報がない場合-----標本2群の場合の統計検定

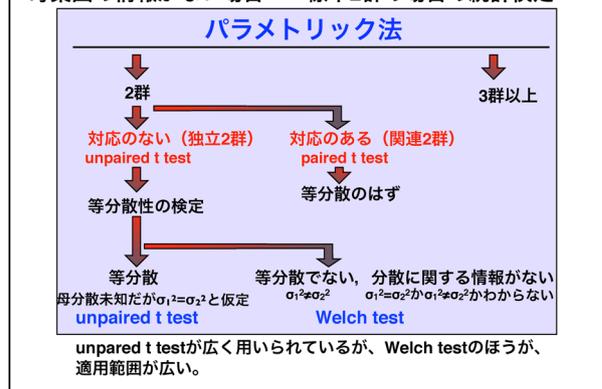


図33 ■ 2群のパラメトリック検定の流れ

等分散性の検定（F検定）

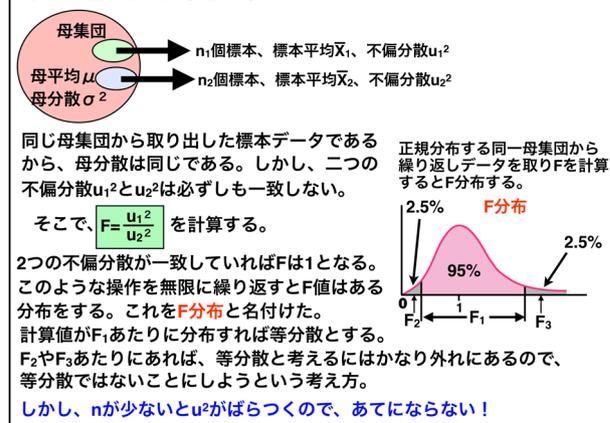


図34 ■ 等分散性の検定

ここで、Fが1あたりにあると等分散と考え（図34中 F_1 前後）、1から大きく離れた場合（図中 F_2 や F_3 前後）、等分散と考えるにはあまりにも離れており、無理があると判定する。その判定基準であるが、 t 検定の場合と同様に考える。すなわち、全体の5%を基準に考える。 u_1^2 と u_2^2 はどちらが大きい値になるかはわからないので、F値は、1より大きくなるか小さくなるかわからない。そこで、両端の2.5%ずつを基準とし、そこにF値が入った場合、等分散とは言えないと判定することにしようという考え方である。これが、等分散性の検定の原理である^{*12}。

しかしながら、標本の大きさ n が小さいと、不偏分散が大きくばらついてくる恐れがあるので、 u_1^2/u_2^2 も大きく変動し、等分散かどうかの判断が難しくなる。 n が

*12 t 検定でも述べたように、同じ母集団から考えてきたことから、等分散ではないと断定することはできないことはおわかりいただけるはずである。

小さい場合、等分散性の検定では等分散と判断されることが多い。これは等分散ではないと判断するにはばらつきが大きすぎるため、等分散と判定しているにすぎない。等分散かどうかの判断は、一般的に $n=30$ 以上必要とされる。これらのことから、標本の大きさ n が小さい場合は等分散性の検定はあてにならないのである。

そうすると、 n が小さい場合に等分散と判定されても、unpaired t testを行うのは危険かもしれない。生命科学系の研究では、 n が30よりも小さい実験が多いのではないだろうか？ このような場合、Welchの検定を選択したほうが妥当と考えられる（Welchの検定は次項で説明する）。もちろん、 n が大きい場合でもWelchの検定は使えるので、結局、標本の大きさ n にかかわらず、等分散性の検定は行わず、Welchの検定を行えばよい。なお、この検定の流れは比較的最近推奨されてきているが、まだ広く認められたわけではなく、等分散性の検定→unpaired t testという考え方で書かれた本がほとんどである。この議論に関してネット上にまとめられている⁽³⁾。検定を利用するだけの研究者にとって、この論争は迷惑であり、どちらかに決めてほしいところである。筆者が調べたかぎりでは、等分散性の検定は行わず、Welchの検定でよいと考えられる。この考え方で、図33を修正したのが、図35である。

統計ソフトでは、等分散性の検定ののちにunpaired t testとWelchの検定の両方の検定結果が表示されるものがある。両方で検定結果が同じであればその結果を採用すればよいので何も問題がない。しかし、 n が小さい場合、もし、unpaired t testとWelchの検定の検定結果が異なると困ったことになる。等分散性の検定があてにならないと、どちらの検定結果を採用してよいか決められない。このような場合は、Welchの検定の検定結果を採

用してよいと考えられる。

Welchの検定の考え方（図36）

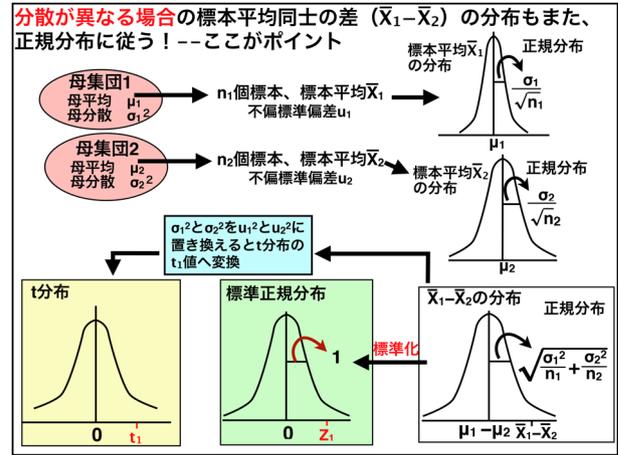


図36 ■ Welchの検定の考え方

Welchの検定の考え方は、unpaired t testの考え方とよく似ているが、分散が異なるかもしれない2群で考えるので、母分散の異なる2つの母集団からそれぞれ標本をとって2群としたと考えたほうが、考えやすい（図36左上図）。つまり、母平均 μ は同じであるが、母分散は異なる2つの母集団からの標本と考えるのである。便宜的に2つの母平均は μ_1 と μ_2 とするが、 $\mu_1 = \mu_2$ である。それぞれの標本平均を \bar{X}_1 と \bar{X}_2 とし、 $(\bar{X}_1 - \bar{X}_2)$ を計算する。これを何度も繰り返して分布を調べると、unpaired t testの場合と同様に正規分布する（図36右下グラフ）*1。このとき、母分散 σ_1^2 および σ_2^2 は異なるので、以下の式のように、別々に足すことで合算の母分散を求める。

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

これをルートした値が母標準偏差となる（図36右下グラフ）。

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

そこで、unpaired t testの場合と同様に、 $(\bar{X}_1 - \bar{X}_2)$ を標準正規分布の値に標準化する（図36右下グラフから中央下グラフへ）。

そうすると、以下の式が出現する（標準化の方法については図18参照）。これにより、 $(\bar{X}_1 - \bar{X}_2)$ は標準正規

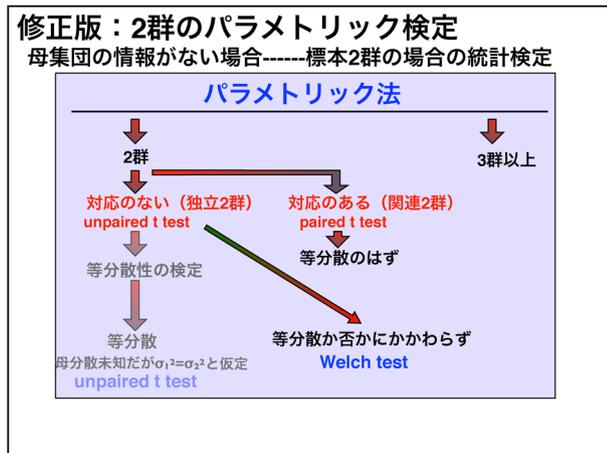


図35 ■ 修正版：2群のパラメトリック検定の流れ

分布の Z_1 へ変換される (図36中央下グラフ)。

$$Z_1 = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

unpaired t test の場合は、以下の式であった (図25)。分母の標準偏差の違いを認識して欲しい。

$$Z_1 = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

ここで、 $\mu_1 = \mu_2$ であるから、分子の $(\mu_1 - \mu_2)$ はゼロとなり消すことができる。また、 σ_1^2 および σ_2^2 はわからないので、それぞれの群の不偏分散 u_1^2 と u_2^2 を用いることとなる。

unpaired t test では、図26にあるように、合算の不偏分散を計算した。これは2群がそれぞれ等分散であると仮定したので、合算した。しかし、Welchの検定では、2群の母標準偏差が異なるので、それぞれの群の不偏分散 u_1^2 と u_2^2 をそのまま代入する。そうすると、標準化の式の子母は以下の t_1 の式となる。

$$t_1 = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{u_1^2}{n_1} + \frac{u_2^2}{n_2}}}$$

この t_1 値が、 t 分布のどこにあるかを調べる (図36左下のグラフ)。

このように、 $(\bar{X}_1 - \bar{X}_2)$ を標準正規分布に標準化するが、母標準偏差がわからないので不偏標準偏差に置き換えることで t 分布となり、 t_1 値が t 分布のどこにあるかを考えるという手順は unpaired t test の場合と同じであることがわかる。

ただし、unpaired t test では、自由度は2群の自由度を足した $(n_1 - 1) + (n_2 - 1)$ であった。つまり自由度 $(n_1 + n_2 - 2)$ の t 分布を用いる。しかし、Welchの検定では、自由度が異なる。自由度の計算式は以下のような複雑な式となっている。

$$\text{自由度} = \frac{\left(\frac{u_1^2}{n_1} + \frac{u_2^2}{n_2} \right)^2}{\left(\frac{u_1^2}{n_1} \right)^2 + \left(\frac{u_2^2}{n_2} \right)^2} \cdot \frac{1}{\frac{1}{n_1 - 1} + \frac{1}{n_2 - 1}}$$

自由度については、すでに説明した*7。しかし、この計算式がどのように誘導されたのか、また、なぜこの自

由度の t 分布を用いれば適正に検定できるのかは原報を読んでも理解できず、筆者の能力を超える⁽⁴⁾。この自由度は整数にはならないので、四捨五入して整数にして、その自由度の t 分布を用いることとなる。統計ソフトでは自動的に計算してくれ、その t 分布に当てはめてくれる。その後の手順は、unpaired t test と同様に検定する (図27)。

unpaired と paired t test はどう違うのか? (図37)

対応のない独立2群の差の検定 (unpaired t test)

例：独立した2群のネズミで、対照食と試験食を与えて比較する場合。

(対照食と試験食ではネズミが同じではない)

対応のある関連2群の差の検定 (paired t test)

例：同じネズミで、食品成分投与前と投与後の血圧を比較する場合。

(投与前と投与後は同じネズミ)

paired t test を行うべき場合に、unpaired t test を行うと、有意差が出にくい!

図37 ■ unpaired と paired t test の違い

これまでは、対応のない独立2群の差の検定 (unpaired t test) を述べてきたが、対応のある関連2群の差の検定 (paired t test) もある。わかりやすい例を図37に記載した。2群の別々のネズミで試験すると、ネズミは対応していないので対応のない独立2群の検定である。一方、同じネズミで投与前と投与後の比較をする場合は、対応があるので paired t test となる。paired t test の考え方は、これまで述べてきた unpaired t test を理解していれば容易に理解できる。

簡単に記すと、paired t test の場合には、同じネズミからのデータなので、投与後のデータ x_1 と投与前のデータ x_2 を個々のネズミで比較できる。そこで、仮に6匹のネズミがいるとして、それぞれのネズミで差 $d = x_1 - x_2$ を計算する。投与前と投与後で大きな変化がなければ d はゼロの前後に分布することになり、何らかの影響があれば、 d はゼロから離れる。パラメトリック検定であるからもとの母集団は正規分布と仮定しており、差 d もまた正規分布することが知られている。そこで、6匹の d の平均値 (標本平均) \bar{d} とその不偏標準偏差を計算する。その後は、 \bar{d} を標準正規分布に標準化

し、不偏標準偏差に置き換えて、 t 分布に当てはめるのは、unpaired t testと同様である。詳しくは拙著を参照されたい⁶⁾。

対応のある関連2群の実験は、同じネズミで比較し合うので個体差が消えるため、全く異なるネズミで比較する対応のない独立2群の実験よりも、実験計画によっては、信頼性の高いデータが得られる可能性がある^{*13}。特に、個体差の大きいことが予想される実験の場合は威力を発揮する。問題のない実験計画が立てられるのであれば、paired t testで検定できるような計画にするとよいが、そうもいかない実験は多い^{*14}。

なお、統計検定をpaired t testで行うべきところをunpairedで行うと、せっかくの有意差を見落とすことになりかねないので、間違えないようにしなければならない。

ノンパラメトリック検定 (図38)

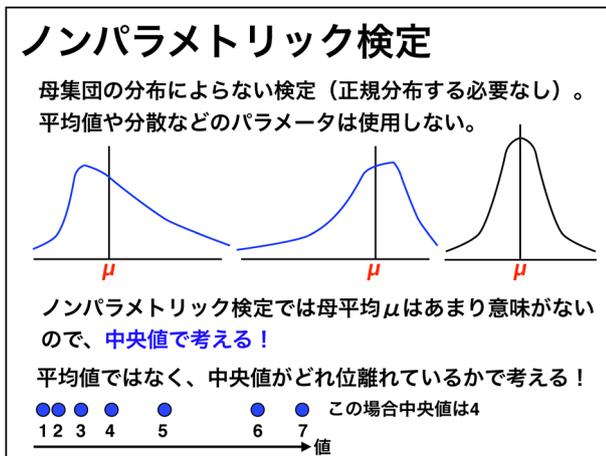


図38 ■ ノンパラメトリック検定

パラメトリック検定は、母集団のデータが正規分布していることが前提で組み立てられていることはすでに述べてきた。ノンパラメトリック検定は、母集団の分布は正規分布する必要がなく、平均値や分散などのパラメータは使用しない。ノンパラメトリック検定は基本的に平

*13 同じネズミからの前と後のデータなので、分散は同じと考え、等分散性の検定は行わない (図35)。

*14 ヒトは個体のバラツキが大きいため、対応のない独立2群の試験では、標本の大きさをかなり大きくする必要がある。しかし、被験者を集めるのはたいへんである。そこで、対応のある関連2群の試験を用いるケースは多い。たとえば、試験前後の血圧を比較するといった試験である。一方で、長期の試験になると、夏と冬で血圧が変動するといった季節変動などの問題が生じるという欠点もあり、目的をよく考えて実施する必要がある。

均値ではなく中央値で考える (図38)。標本の大きさ n が大きい場合は、正規性の検定により、正規分布と判定されない場合に用いることができる (正規分布と判定されても用いることはできるが)。また、外れ値^{*15}がある場合でも利用することができる。前回図7で述べたが、 n が小さい場合は、母集団が正規分布しているかどうかは判然としないため、パラメトリック検定にこだわる必要はない。ノンパラメトリック検定は応用範囲が広い。したがって、筆者は悩ましいデータが得られた場合にパラメトリック検定とノンパラメトリック検定を併用して、結果を考えることにしている。

2群間のノンパラメトリック検定 (図39)

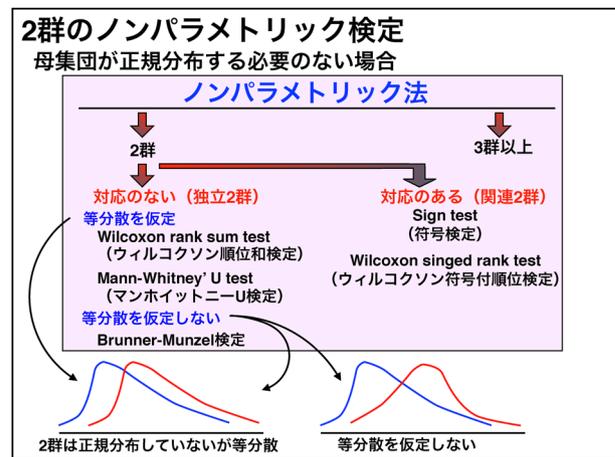


図39 ■ 2群のノンパラメトリック検定の流れ

ノンパラメトリック検定でも、2群間および3群以上の検定法が存在する。図39には2群間の検定法をいくつか挙げています。パラメトリック検定同様に、対応のない場合と対応のある場合の検定法がある。なお、よく用いられるのはウィルコクソンの順位和検定やマン-ホイットニー u 検定であるが、2群は等分散を仮定している (図39左下グラフ)。しかし、 n が少ないと等分散かどうかは判断しづらい。そこで、等分散を仮定せず、等分散でも等分散でなくても用いることができる Brunner-Munzel 検定^{*16}がある (図39左、右グラフ)。なお、ウィルコクソンの順位和検定とマン-ホイットニー u 検定

*15 ほかの値から大きく外れた値のことである。研究ではよくお目にかかり、どう処理するか悩む場合が多い。統計検定における外れ値の取り扱いについては、次回に記述する。

*16 筆者にはこの検定法の原理等の知識がないので、紹介に止める。

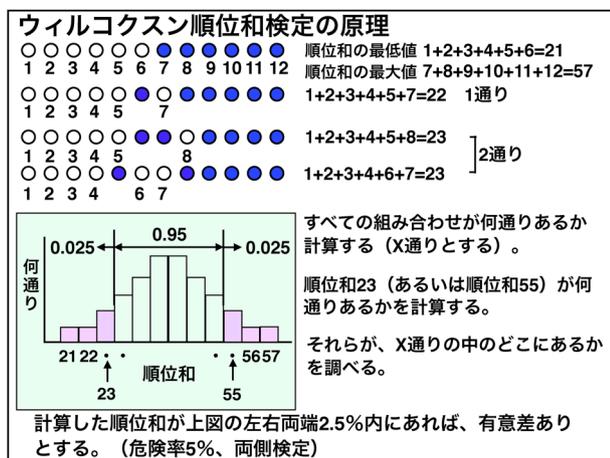


図42 ■ 順位和検定の考え方(3)

のような基準で判定するかを決めればよい。

そこで、順位和の組み合わせがどれくらいあるかを考える。まず、順位和の最小値は、たとえば白が1から6であり合計21、青は自動的に決まり7から12となり、合計57で最大値となる(図42上)。この組み合わせは1とおりにしかない。ここで、白が決まれば、青は自動的に決まるので、白だけを考えることとする。そこで次に、白の6が7に替わったとすると1, 2, 3, 4, 5, 7となり合計22であり、この組み合わせも1とおりにしかない。次に、白が1, 2, 3, 4, 5, 8となると合計23であるが、合計が23になる組み合わせは、1, 2, 3, 4, 6, 7もありうる。したがって、2とおりにあることになる(図42上)。このような考え方で、合計が24, 25…となる組み合わせが何とおりにあるかを最大値の57まで計算していく。それぞれの合計値に対して組み合わせが何通りあるかを棒グラフにすると、図42左下のようなグラフが得られる(このグラフは模式的な図で正確ではない)。次に、すべての組み合わせが何とおりにあるかを合計する(ここでは合計Xとおりにする)。そこで、図40の③で示した例である順位和23がこのグラフのどこにあるかを考える。Xとおりの全体を100として順位和23の2とおりが、たとえば、両端の2.5%以内に入った場合、有意差があると判定する(有意水準5%、両側検定の場合)(図42左下グラフ)。この判定は、青の場合の順位和55で考えても、グラフでの位置は左右逆側になるが、同じ結果を得ることができる(図42左下グラフ)。パラメトリック検定で説明したので詳しくは述べないが、対立仮説として $A > B$ が仮定できれば、片側検定もありうる。これが順位和検定で有意差を求める原理である*17。

なお、上述の順位和検定の例は、対応のない独立2群の検定であるウィルコクソンの順位和検定であり、t検

定ではunpaired t testに相当する。ノンパラメトリック検定でも、対応のある関連2群の検定(paired t testに相当)がある。代表的なものは、図39右に記載した。ウィルコクソンの順位和検定に対しては、ウィルコクソン符号付順位検定がある。ここでは詳細に述べないが、「対応のある」の基本的考え方は、パラメトリックのpaired t testと同様であり、対応するデータ x_1 と x_2 の差 d を順位に置き換えて考えることになる。詳しくは拙著⁽⁵⁾あるいは『すぐわかる統計解析』⁽⁶⁾を参照されたい。類似した名称であるので間違わないことである。なお、統計アプリによっては、「ウィルコクソン順位検定」という名称で、対応なしか対応ありにチェックするという手順となっている。これは上記の2つの検定法に対応する。

ノンパラメトリック検定の利点、欠点(図43)

ノンパラメトリック検定の利点、欠点

母集団の分布に神経質になる必要がない。
 母集団が正規分布の場合、検出力が落ちる。
 しかし、その場合でも検出力の低下は意外と少ない。
 外れ値がある場合は、有効性が高い可能性あり。
 (安易に、棄却検定をしないこと)
 nが小さいと検定が厳しくなる。

パラメトリック検定だけでなく、ノンパラメトリック検定も選択肢にいらてよい。

検出力が落ちるとは?→正確に検定できなくなるという意味

図43 ■ ノンパラメトリック検定の利点、欠点

ノンパラメトリック検定の原理を知ると、多くの方は大きくばらついたデータでも、そうでなくても検定結果が同じになることに対して、それでよいのだろうかと思うのではないだろうか? 私も最初はそうであった。図43に利点、欠点を示している。母集団が正規分布する場合は検出力が落ちるらしいが、それほど大きく落ちるわけではない⁽²⁾。また、外れ値があるとき、棄却検定で外すことが可能になる場合もあるが、安易に外すことに

*17 われわれは論文などにデータを表示する場合に、標本平均±標準誤差(SE)あるいは標準偏差(SD)を書く。これらはパラメトリック検定のためのパラメータであり、ノンパラメトリック検定を行った場合は、これらを表記することは意味がない。しかし、論文や学会発表では慣習的にこれらの表記が行われている場合がほとんどである。考えてみれば奇妙である。

は問題があるので、外さずにノンパラメトリック検定を行ってみる価値はある。ノンパラメトリック検定でも有意差が得られなければ、あきらめることができる。外れ値の取り扱いについては次回に述べるが、思い悩むデータがあるときは、まずは、パラメトリック検定とノンパラメトリック検定、さらには棄却検定の結果をみてから考えてはどうであろうか？

なお、ノンパラメトリック検定は標本の大きさ n が小さいと、パラメトリック検定よりも検定は厳しくなる。 t 検定では2群とも $n=3$ でも検定できる。しかし、有意水準5%で両側検定の場合、ウイルコクソンの順位和検定では、最低必要な n は2群共4である。この場合、図42の上から2行目のように、白丸と青丸が完全に分離している場合のみ有意となるのでかなり厳しい。また、ウイルコクソンの符号付順位検定の場合、両側検定で最低必要な n は6である。したがって、これらの検定では最低必要な n での検定は厳しいので、それよりもいくらか多い n 以上にするほうが無難である。

おわりに

次回最終回は、3群以上の場合の検定である、一元配置分散分析、多重比較、二元配置分散分析を中心に記述する。研究論文や学会発表において、3群以上ある試験で多重比較を用いず、不適切に2群の検定法、たとえば t 検定を繰り返している例は減少してきているが、まだ

完全にはなくなっていない。なぜ t 検定ではいけないのかを理解すべきである。また、二元配置分散分析時の群間比較は間違いのない検定法で行われているほうが少ないように思われ大きく混乱しているので、正しい理解が必要である。

文献

- 1) 市原清志：“バイオサイエンスの統計学”，南江堂，1990.
- 2) 足立堅一：“らくらく生物統計学”，中山書店，1998.
- 3) http://www2.vmas.kitasato-u.ac.jp/lecture0/statistics/stat_info03.pdf
- 4) B. L. Welch: *Biometrika*, **29**, 350 (1938).
- 5) 池田郁男：“実験で使うとこだけ生物統計2 キホンのホン 改訂版”，羊土社，2017.
- 6) 石村貞夫：“すぐわかる統計解析”，東京図書，1993.

プロフィール



池田 郁男 (Ikuo IKEDA)

<略歴>1975年九州大学農学部食糧化学工学科卒業/1980年同大学大学院農学研究科博士後期課程修了，農学博士/同年同大学農学部助手，助教授を経て2005年より東北大学大学院農学研究科教授/2018年より同大学未来科学技術共同研究センター教授，現在に至る<研究テーマと抱負>食品成分および栄養素の肥満，動脈硬化症予防作用に関する研究，ステロール吸収機構の解明<趣味>以前はバドミントンでしたが，今はドライブ&温泉巡り

Copyright © 2019 公益社団法人日本農芸化学会
DOI: 10.1271/kagakutoseibutsu.57.562