

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

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 I」の改訂増補版である。今後、「II」および「III」も改訂増補する。これら三部作は2013年に出版されたが、最近でもJ-STAGEでの「化学と生物」の論文アクセスランキングで常に上位を占めている。しかし、これら三部作には正しくない記載や、言葉足らずでわかりにくい内容があった。そこで、誤りを修正し、内容をさらに充実しわかりやすくするため、改訂することとなった。間違いのあった理由は、筆者の統計に対する理解不足にほかならない。筆者は統計学が専門ではない。本来このような学術雑誌には専門家が執筆すべきと考えるが、専門家による総説や専門書は、往々にして生命科学系研究者（学生を含めて）が知りたい基本や核心については書かれていない場合が多い。この三部作が出版後6年も経つにもかかわらず一貫してアクセス数が多いことは、統計を理解できず救いの手を必要とする研究者がいかに多いかを物語っている。本総説が多くの研究者の研究の一助となることを願う。

序

このセミナーは、2012年5月に開催された日本栄養・食糧学会の教育講演「栄養学・食糧学のための実践統計

検定法」がもとになっている。この講演は好評を得て、その後、現在までに多くの講演や講義の依頼を受け、また、本も執筆した^(1,2)。統計を理解したいという研究者や学生の欲求がいかに強いかを実感している。私は統計の専門家ではないので、本セミナーの内容は基本的に多くの統計書を総合して生命科学系研究者にわかりやすいように解説したものである。実験研究に携わる研究者や学生にとって多くの統計書はわかりにくく、読むことには多大の苦痛を伴い、挫折することもしばしばである。そこで、本セミナーは、研究者の視点から、研究者が知りたいポイントに重点を置いてわかりやすく解説したつもりである。

高度な研究を行っている研究者が、用いている統計検定の基本をわかっていなくてよいはずがない。統計を利用している研究者は必ずしも難しい統計理論を必要としているわけではなく、ある程度の基本を理解し、学会や学術論文で文句を言われぬ適切な検定法を選ぶことがまずは重要である。統計が理解できていない場合、自らの研究と似かよった研究論文の統計検定を真似て検定を行っている場合もあるかもしれない。しかし、出版されている論文が正しい統計検定法を用いているかということ必ずしもそうではなく、不適切な検定法の論文も少なからず存在する。ある調査によれば、学術論文の約30~50%が統計に何らかの問題があることが指摘されている⁽³⁾。Impact factorの高いジャーナルの論文でも、これによ

◇◇◇ コラム ◇◇◇

筆者が参加するどの学会でも、長年にわたって統計検定に問題のある発表が多い印象があった。また、筆者の読む文献でも不適切な統計検定が多々見受けられ、日本だけの問題ではないと常々感じていた。他の多くの研究者もこのような懸念はもっており、この状況がよいことは思っていなかったはずである。多くの大学で統計学の講義はあるが、ほとんどは基礎的内容であり、研究者が望んでいるような教育になっていないため、直接的に研究の役に立つわけではない。したがって、研究に必要な統計は研究室の指導教員が指導すべきである。しかし、学会や論文の現状をみると、指導が十分なされていない場合が多いように見受けられる。統計ソフトの使い方を教えるだけの研究室も多いのではないだろうか？ 普通の学問であれば自力で勉強しても何とかなるものである。しかし、統計書を読み解くのはなかなか厄介で、筆者の場合何冊統計書を読んでも、いま一つ「ピンと来ない」状況で統計を利用して来た。しかも、恥ずかしながらそのような状況で統計の講義を行わざるを得なくなった。筆者が所属した農学部の学科

の場合には、統計を専門とする教員がいないため、統計をよく使っている教員が講義を任せられたのである。しかし、当時は十分な理解のないまま講義していた記憶がある。

そんなときに舞い込んだのが、2012年の日本栄養・食糧学会での教育講演である。「ピンと来ない」状況で学会講演などできるわけがない。このことが統計学を腰を据えて勉強し直すきっかけとなった。統計をわかりやすく説明するためには、まずは自らが理解する必要があるため、たくさんの統計書をわたり歩いて理解に努めた。完全に会得することは無理であったが、自分なりに咀嚼して講演を行った。不安ばかりの講演となったが、聴衆の皆さんの琴線に触れたようで、その後、講演、講義、総説や本の執筆依頼へとつながった。しかし、執筆した総説や本には素人的な間違いがあり、厳しい指摘を受けたが、それでも、多くの方々にはわかりやすいと好意的に受け止められたようである。特に、若手の研究者や学生の皆さんは統計を理解したいという欲求が強いようで、直接の講演依頼がいくつもあった。今回の改訂は、間違いを正し、講演や講義で指摘を受けた、わかりにくさを改善するためである。読者の統計への理解が進むことを祈るばかりである。

いのかと思われるような統計検定法が見受けられる。論文編集委員や査読者に統計の知識が乏しいと、不適切な統計検定法のまま雑誌に掲載される。その検定法を真似ると、さらに間違った統計検定法が世界中で横行することとなる。

統計検定は有意差があるかどうかを判断するための一手段であり、有意差の有無さえわかればよいと考えている研究者もいるかもしれない。しかし、この考え方だけでは研究者としては適切とは思えない。生命科学系研究者にとって、統計的考え方を理解することは真実を追究するうえで習得しておくべき重要な基本事項と筆者は考える。したがって、有意差検定の手法だけでなく、統計的考え方を理解し、研究に役立てて欲しい。

研究で必要とされる統計検定法は多様であり、研究者がそれらすべてを修得することは容易ではない。だからといって、統計書で自分が必要とする統計検定法の部分のみを読んでも、理解できない場合が多い。そもそも、どの検定法を選べばよいのかさえわからないことしばしばである。仕方なく、基礎から学んでみるが途中で挫折し、結局、研究室で従来から行われている統計検定を意味もわからず踏襲することになる。というのが、筆者が何度もたどった道であるが、このような研究者は多いのではないだろうか？ この最大の原因は、統計学の基

本中の基本を理解できていないことではないかと筆者は考える。その典型的な例は、標準偏差 (SD) と標準誤差 (SE) の違いを理解していないことである。たくさんの統計検定の手法は、それぞれ皆違うようにみえるが、根底に流れている考え方の基本は比較的共通している場合が多い。共通の考え方を習得するために必要な第一の基本はSDとSEの理解である。

さらに、多くの研究者は有意差があるか、ないかで一喜一憂する。しかし、有意差があるとは、必ずしも本当に差があることを意味しない。われわれ研究者が知りたいのはあくまでも真実であるが、有意差がある、なしの判断が真実を示すとは限らない。真実を明らかにする研究を行うためにも、統計の理解は必須である。また、統計を理解することは、実験結果の処理のためだけでなく、実験計画を立てるうえでも重要である。本来、統計を考えつつ実験計画を立てるのが基本であり、実験を行ってから統計検定法を考えるべきではない。そのためにも統計を理解していないと実験計画は立てられないのである。

筆者は、食品機能学および栄養化学、脂質生化学を専門としており、培養細胞試験や試験管内試験、および、動物飼育試験での2群および3群以上の群間比較を主に用いている。統計検定法は多様であり、それらを網羅し

てしまうと、焦点がぼけてしまう可能性がある。そこで、本セミナーでは筆者自身が研究において主に利用し、最低限必要と考える検定法を詳述する。より細かい検定法は、拙著を参考にされたい^(1,2)。

本セミナー4~6では講演のスライド形式で図を挿入する。本文を読むにあたって、まず、その項目の図の内容をざっと眺めることをお勧めする。この改訂増補版では、「統計検定を理解せずに使っている人のために I」「II」「III」の図表を、最近の講演で用いている図表に差し替え、また、追加もしている。

本セミナーでは、主に以下の項目について言及する。

- ・母集団と標本の違いを理解して研究しているか？
- ・標準偏差 (SD) と標準誤差 (SE) の意味は何か？何を表現しようとしているのか？
- ・正規性の検定、等分散性の検定は必要か？
- ・帰無仮説、有意水準 (危険率) の意味を理解しているか？
- ・パラメトリック検定 (t 検定) とノンパラメトリック検定 (Wilcoxon 順位和検定) の原理を理解しよう！
- ・パラメトリック検定とノンパラメトリック検定はどちらを選ぶべきか？
- ・片側検定と両側検定のどちらを用いるべきか？
- ・多重比較と多重性の問題とは何か？
- ・一元配置分散分析と多重比較の原理と使い分けを理解しよう！
- ・二元配置分散分析と群間比較を理解しよう！ (多くの論文で間違っている！)
- ・統計検定法を理解して、実験計画を立てよう！

母集団や標本のデータのタイプ

研究は多様で、データのタイプもさまざまであるが、主に3つのタイプがある。

1. 測定データ (たとえば、ネズミの体重や細胞内のある物質の濃度などの測定値)
2. 順位データ (たとえば、魚の大きさを大, 中, 小と分類する場合やネズミの体重を軽い順に順位に置き換えた場合など)
3. カテゴリーデータ (たとえば、ある品種の花の色, 赤, 青, 黄がどのような割合で出現するのかなど)

このセミナーでは、主に **1. 測定データ**を取り扱うが、ノンパラメトリック検定では **2. 順位データ**を取り扱う。

母集団と標本の違いを理解する (図1)

基礎統計学を学習すれば、母集団と標本は最初に登場し、それほど難しい話ではない。しかし、読者は自らの研究が母集団なのか標本なのかを意識して研究しているだろうか？ 研究者が母集団と標本の違いを無意識にでも意識していなければ、統計を何もわかっていないと言われても仕方がない。

典型的な母集団と標本の例は以下のようなものがある。

例1) S市民の血漿コレステロール濃度の平均値や分布を知りたい。しかし、全員を調べることはできないので、一部の市民を抽出して調べる。S市民全員は**母集団**であり、抽出した一部の市民は**標本**である。この例はわかりやすい。しかし、生命科学系研究者には今一つピンとこない。

そこで次の例2)として、ある系統のマウス6匹に「ある特殊な成分X」を摂食させて影響をみたところ、画期的な発見があった。この発見は世界初であり6匹以外に成分Xを摂取したマウスはいない。さて、この実験は、母集団の実験か、標本の実験か？ 多くの生命科学系の実験は、ほかに誰も行っていない世界初の実験が多いのではないだろうか？

例1)の研究の目的を考えると、研究者は実験した標本の結果だけを知りたいわけではなく、母集団のS市民全体 (母集団) の情報を知りたいはずである。では、例2)はどうであろうか？ 研究者であれば、この試験結果が6匹だけで起こるのでは満足せず、世界中の同じ系統のマウスを用いて同じ試験を行っても、同様の結果が得られることを期待するのではないだろうか？^{*1} つま

*1 実験動物を用いる試験の場合、理想的には母集団は世界中の同系統のマウスと考えれば世界的な研究として位置づけられる。しかし、現実はそのようなわけではない。同じ系統のマウスでも、動物を納入するブリーダーごとに遺伝子は全く同じというわけではないので、ブリーダーが違えば実験に対する応答は異なることはよくある。したがって、購入したブリーダーのマウス全体を母集団と考えたほうがよい。筆者には、ある系統のラットで、研究対象の遺伝子がブリーダー A では欠損し、別のブリーダー B では欠損していなかったため、研究が大混乱に陥った経験がある。同じ系統ならどのブリーダーの動物でも同じように応答すると思っている研究者が多いかもしれないが、必ずしもそうではないので注意が必要である。このように、動物実験は世界的な研究と位置づけたいが、実はローカルな実験を行っているため、類似の試験でもほかの研究者が行うと研究結果が異なることは起こりうると思っておいたほうがよい。このような状況は、植物、魚や微生物などの研究でも同様と考えられる。類似の実験でも研究者によって異なる結果が出現するので、研究者を悩ます一因となるが、多様な結果に惑わされないことである。生物を用いた試験とはそのようなものであり、普遍的な真実を見つけることは容易ではない。

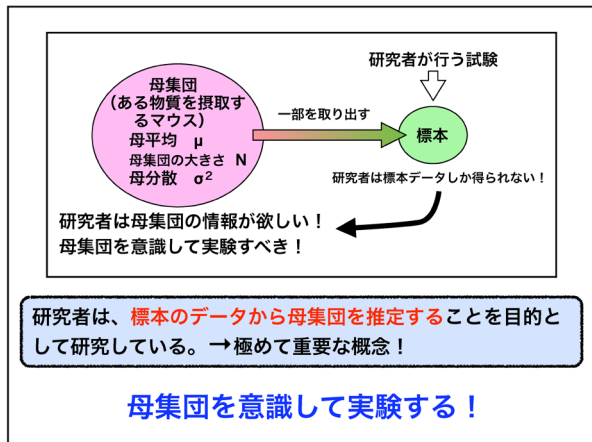


図1 ■ 標本で試験し母集団の情報を推定する

り、いくら世界初であっても6匹の試験は標本の試験ととらえ、より大きな母集団を想定し（そのような母集団は実在しないが）、その母集団でも同じ結果が得られることを期待すべきである。すなわち、研究者は標本のデータしか得ることはできないが、母集団の情報を知りたい（推定したい）ために研究を行っているのである（図1）。これは極めて重要な概念である。なぜ、母集団の情報を知りたいのかというと、母集団の情報はすなわち真実の情報だからである。研究者は真実を探求するために研究を行っているはずである。読者、特に学生の皆さんは、標本から母集団を推定するという概念を意識しているであろうか？ もしこの概念がなく自分の得たデータを標本ではなく母集団の情報と考え、間違いのない真実ととらえてしまうと、大きな誤りを犯すことになりかねない。

ここで、例2)のように実際には存在しない母集団を無限母集団と呼ぶ。これに対して、例1)のS市民全体といった具体的な母集団が存在する場合は有限母集団と呼ぶ。

もちろん、無限母集団を想定する必要のない研究もあり得る。購入したあるいは繁殖で生まれたマウス10匹の情報のみが欲しいといった研究である。その場合はこの10匹を母集団と捉えることができる。すなわち、母集団の想定は研究の目的で変化する。研究を行う際には、まず、研究の目的をよく考え、母集団は何かを想定すべきである。主にマウスの例で説明してきたが、植物や魚などでの試験、あるいは、培養細胞試験などでも考え方は全く同じであり、母集団と標本の概念はもたなければならない。

生物の場合、個々の集団はあるバラツキをもっている。たとえば、ある母集団のマウスの体重は、ある範囲

でばらついている。その母集団から標本をとると当然バラツキがあり、平均値もバラツキも必ずしも母集団とは一致しない。そのような条件下で、研究者は母集団の情報を推定しようとしている。しかし、標本から母集団を推測することは容易ではない。特に、標本の大きさ n^{*2} が小さい場合は、データはばらつく可能性が高くなり、母集団の推定はぶれる。また、そもそもバラツキの大きいパラメータの場合も同様である^{*3}。バラツキが大きいと、全く同じ試験をもう一度行くと、異なる結果となる可能性が高くなる。すなわち、再現性がとれない確率が高いことは認識しておいた方がよい。

研究者は真実を知るために研究する。しかし、生物を用いる研究では n が小さい標本で研究している限り、真実を突き止めることは容易ではない。したがって、真実を知るためには、必ず複数の実験を行い再現性を調べるべきである^{*4}。2016年のNature誌の調査では、多くの研究者が再現性を調べずに論文化していることを指摘している⁽⁴⁾。調査に応じた研究者の70%以上がほかの研究者の研究に再現性がないと答えている。再現性のない論文が多数出回るとは、混乱を招く要因となる。日本でも博士論文や修士論文に間に合わせるためや業績を増やすために、再現性を調べずに論文化している例は多いのではないだろうか？

ここで、後々出てくるので母集団を定義する。母集団（たとえば、世界中のマウス）の総数（「母集団の大きさ」と呼ばれる）を N 、あるパラメータの総平均値は母平均と呼び μ 、そのバラツキ方を表す母分散を σ^2 とす

^{*2}統計学では、母集団から抽出した標本の数を「標本の大きさ」あるいは「標本サイズ」「サンプルサイズ」と呼び、英語ではsample sizeである。標本数とは呼ばない。標本数は、英語でthe number of samplesでありサンプル数とも呼び、簡単に言えば群数のことである。たいへん紛らわしいので注意してほしい。

^{*3}このバラツキは測定誤差によるバラツキのことではなく（次項参照）、生物本来のバラツキである。 n をどれくらいとればよいのかは、それぞれの実験に依存する。主に、データのバラツキ方つまり分散が関係する。バラツキが大きければ、 n を大きくすべきである。しかし、一般にバラツキは実験してみないとわからない場合が多く、また、測定パラメータによりバラツキは異なるので、実験開始時に n を予測することは困難である。論文や過去の実験結果などから大きなバラツキが予測される場合は、 n を大きくすることを考えるべきである。

^{*4}実験遂行上の問題や測定誤差などの理由でデータに自信がない場合は、全く同じ試験を行って結果を確認することは有用である。*in vitro*試験では比較的容易にやり直せる。しかし、動物試験や長期の観察期間が必要な試料の場合はやり直しが困難な場合が多いので、試験を確実に遂行することが重要となる。通常の実験において再現性を確認する場合は、「改訂増補版：III」で述べる“多重性の問題”を考慮すべきで、全く同じ試験を繰り返すよりも、別の角度の実験を複数回行って、同様の結果が得られるかを確認したほうが、信頼性の高い試験と認識される。

る。 σ^2 をルートした σ は母標準偏差と呼ばれる(σ^2 と σ については後述する)なお、すでに登場したが、母集団から取り出した標本の数は、「標本の大きさ」*1と呼び、 n と表記する。

統計検定は測定誤差を考慮してくれない！(図2)

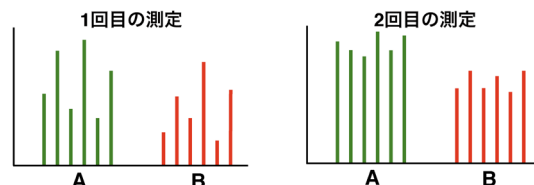
このセミナーでは、主に測定データ(定量値)の統計処理を扱う。測定データは研究者や学生が測定した数値である。そこで問題になるのが測定誤差であり、主題に入る前に注意しておきたい。個々の生物にはバラツキがあるが、それに加えて、測定者の測定誤差が上乘せされる。測定誤差が大きいとデータの信頼度は低くなり、統計検定に値しない。当たり前である。筆者の経験では、卒業論文研究で研究室に入る学生諸君の定量分析能力はかなり低く、信用できない場合が多い。しかも、学生は自分の定量能力の低さを理解していない。どれだけ長く学生実験をやっているか、定量性を高めようという意識と訓練がなければ、定量能力は高くない。したがって、学生の定量能力を把握し、測定誤差を極力小さくする訓練を行うことが、指導者には求められる。同じ試料を2回測定させたとき、全く異なるデータとなることは珍しくない(図2)。それぞれの測定値で相関をとってみると一目瞭然である(図3)。どちらが正しいデータであろうか？どちらも正しくないこともありうる。

永年学生を指導した経験からは、オートピペット*5、ガラスピペット、いずれでもピペッティングの誤差が大きい。測定の際にピペット操作が多いと誤差が誤差を生む。特に、微量サンプリング、たとえば、数 μL から20 μL のサンプリングは大学院生でも誤差が大きく、再現性のある測定値が得られないことがある(mRNA発現量の測定など)。サンプリング能力を高めておかない

*5オートピペットは、目盛りどおりの容量が測り取れると信じている学生がいるが、必ずしもそうではない。しかも、目盛りと測り取れる容量は、ずれている。また、熟練度が低いとバラツキが大きくなる。さらに、オートピペット自体しばしば狂いを生じる。その点を理解している研究者であれば、使用するすべてのオートピペットを定期的に検定するか、あるいは、使用前に検定しているはずである。ちなみに私の研究室では、使用前に、水を用いて測り取る容量を精密天秤で繰り返し重量測定し、正しく測り取れているかを検定してから使用している。特に、複数の研究者がオートピペットを共用して使用している場合は、信用してはいけない。使用前に検定することをお薦めする。この検定を行うと、不具合があればすぐに気づくし、オートピペットの繰り返し精度がどれくらいなのかを認識でき、また、熟練度が低いとバラツキが大きくなることを理解できる。オートピペットを検定して使用することは研究者としての基本である。実は、ガラスピペットも基本ができていない学生の秤量誤差は大きい。

データには必ず測定誤差が含まれる。
しかし、統計は測定誤差を配慮してくれない！
研究結果に大きく影響する。

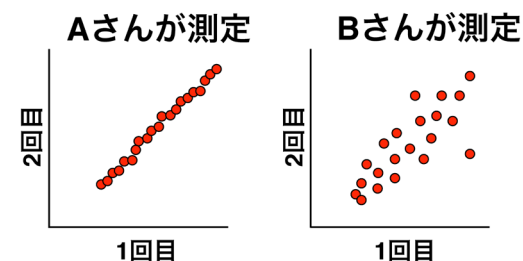
どちらが正しい？ こんなことは普通に起こる！



測定誤差は極力少なく！ 個々の分析能力を高めるべき！

図2 ■ 統計検定は測定誤差を考慮してくれない！

自分の測定誤差を知っていますか？



バラツキのあるデータを2回測ってみると
分析能力がわかる！

図3 ■ 自分の測定誤差を知って分析能力を磨こう！

と折角の実験が台なしになる。学生諸君も実験技術が未熟であることを自覚し、測定誤差を小さくする訓練を行うべきである。正しい測定値が得られてこそ、正しい研究結果や統計検定結果が得られる。

標本の抽出は無作為抽出が基本

すでに述べたが、生命科学系の研究では多くの場合、標本で研究する。母集団からの標本の抽出は、基本的に無作為抽出(ランダムサンプリング)で行うべきである。無作為抽出とは、簡単に言えば、母集団を反映する標本を作為なく取り出すことであり、決してデタラメに取り出すことではない。これは極めて重要で、必ず頭に置いて欲しい。母集団を反映する標本が取り出せれば正しい情報が得られるが、もし、標本が偏っていると統計的推定は間違った情報となる。

無作為抽出は簡単ではない。すでに述べた例1)でS市民全体の情報を知るために、S市民から標本を50名抽

出しなさいと言われると、途方に暮れるのではないだろうか？ 標本の抽出方法は、個々の研究により多様であり一概に言えないが、よく考えて行わないと偏ることになる。たとえば、ある水域から魚を抽出する場合や畑の作物から何本か抽出する場合など、何気なく選びそうであるが、専門分野に適した無作為抽出法を勉強すべきである*6。なお、実験動物を用いる場合は、ブリーダーから納入される動物を母集団から無作為抽出された標本と考えるしかないが、実際は、同じ母集団から無作為抽出されたとは思えないような、応答の異なる動物が納入されることはありうる。

データの特徴をつかむ—差があるかどうかをどう判断するのか？ (図4)

研究者にとって、データの特徴をつかむことは重要である。たとえば、図4左の棒グラフは2群の平均値である(各群6匹)。A群とB群間で差があると言えるかどうかは、個人個人で判断が分かれるはずである。差があるかどうかの判断は平均値の差だけでは決められない。そこで、図4右側上下の2つの棒グラフを見てほしい。A、B群の個々のデータを棒グラフにしている。右上のグラフを見ると、データはよくそろっておりA群とB群間に差があるように見える。ところが、右下のグラフはデータがバラバラであり、差があるようには見えない。しかし、いずれのグラフも平均値を取ると左のグラフとなる。

ここで重要なポイントが見えてくる。差の有無の判断は、平均値の差だけでなく、バラツキ方で変わるということである。すなわち、差の有無は平均値の差とバラツキの兼ね合いで判断できそうである。この考え方こそが、後に出てくるパラメトリック検定の原理である。

もう一つの重要なポイントは、個々のデータを図示するとデータの特徴がつかめることである。平均値と標準誤差や標準偏差(これらは後述する)だけを図示するのではなく、図4右側にあるように、個々のデータも図示することをお勧めする。

パラメトリック検定とノンパラメトリック検定 (図5)

統計検定には大きく分けてパラメトリック検定とノンパラメトリック検定の2つがある(図5)。多くの研究者はパラメトリック検定しか利用していないが、ノンパラ

*6 無作為抽出法は研究の種類により一概に言えないので、このセミナーでは述べない。統計書や標本抽出の専門書を参考に、それぞれの研究に適した方法を考えるべきである。

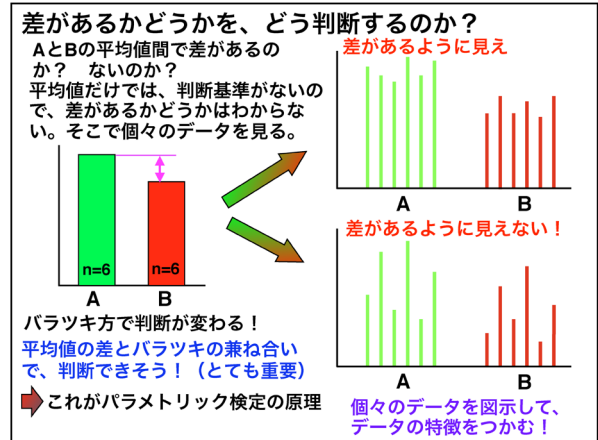


図4 ■ データの特徴をつかむ

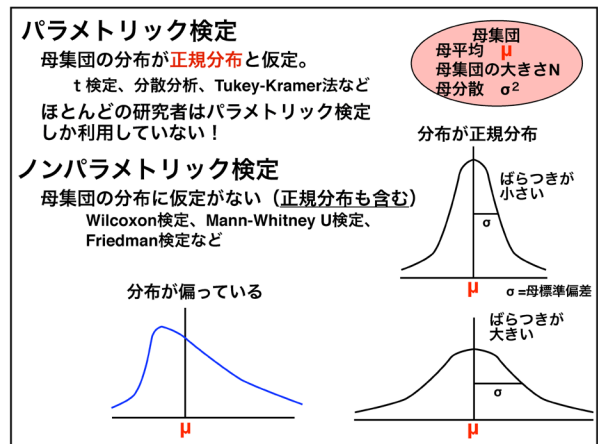


図5 ■ パラメトリック検定とノンパラメトリック検定

メトリック検定も考慮に値する検定法であり、その原理を知っておくことはデータの処理に際して役に立つ。

あるパラメータの母集団の分布は、いろいろな分布が考えられる。すでに述べたように、生命科学系研究者は標本で研究しており、母集団の情報がわからない場合が多い。しかし、母集団の情報を想定したほうが、標本から母集団の情報を推定しやすい。そこで、母集団が正規分布することを仮定して構築されたのが、パラメトリック検定である(図5の右の2つのグラフ*7)。一方、ノンパラメトリック検定は母集団の分布に仮定がなく分布が偏っていてもよい(図5のすべてのグラフ)。注意してほしいのは、ノンパラメトリック検定には母集団が正規分布している場合も含まれることである。ノンパラメトリック検定については次回「改訂増補版：II」に記述する。正規分布の詳細はここでは書かない。統計書にはい

*7 本セミナーのすべての図は筆者が模式的に描いたもので必ずしも正確なものではない。

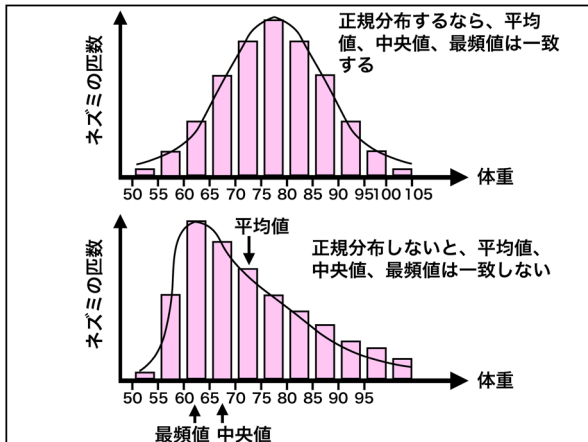


図6 ■ 正規分布とは？

くからでも登場するので参照して欲しい。

生物のかかわる多くのパラメータは正規分布に従うことが知られる。たとえば、母集団のネズミの体重を測定し分布を調べると、図6上グラフのようになったとする。横軸は体重の分布を、縦軸は頻度を示し、ここではネズミの匹数である。平均体重付近で最もネズミの数が多く、平均体重から離れるほどネズミの数は減少するベル型である。重要なパラメータは、母平均 μ とバラツキの指標である母標準偏差 σ である（パラメータの説明は後述）。正規分布の図は、 σ が小さければ裾野が狭く、大きければ裾野が広がる（図5右グラフ上と下を比較）。

もちろん、正規分布しない偏った分布のパラメータも多く存在する（図5左グラフは一例、図6下グラフ）。正規分布しない場合、平均値はあまり意味をなさない（図5左グラフの μ や図6下グラフの平均値の位置に注意）。偏った分布の場合は、中央値や最頻値が用いられるが、生命科学系の基礎研究の論文ではほとんどお目にかからない。中央値とは、データを大きさの順に並べたときの真ん中の値であり、最頻値とは、最もたびたび現れる値である（図6下グラフ）。なお、正規分布であれば、平均値、中央値、最頻値はほぼ一致する（図6上グラフ）。

次回「改訂増補版：II」に記述するが、ノンパラメトリック検定は中央値を用いた検定法である。平均値や後述の標準偏差などのパラメータは母集団が正規分布することが前提であり、正規分布していないとあまり意味がない。しかし、ノンパラメトリック検定を用いた論文でもデータ表記は平均値や標準偏差などを用いている論文が多い。これは、考えてみるとおかしなことである。

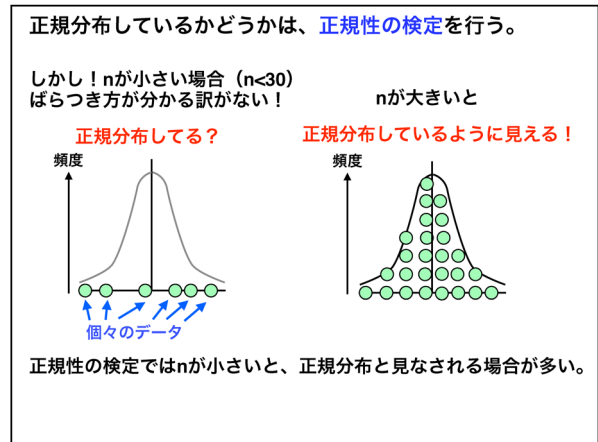


図7 ■ 正規性の検定

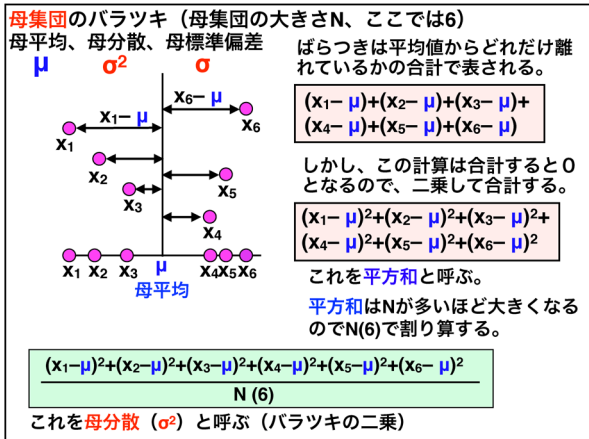
正規性の検定（図7）

パラメトリック検定は母集団が正規分布することが前提である。しかし、生命科学系研究者が測定するパラメータの母集団が正規分布するかどうかはわからない場合がほとんどである。母集団が正規分布しているかどうかは正規性の検定^{*8}を行えば知ることができ、正規分布していれば、パラメトリック検定を行うと統計書には記述されている。読者は正規性の検定を行っているであろうか？

生命科学系研究の場合、標本の大きさ n は10以下の場合が多い。図7の左側のグラフにあるように、たとえば n が6と小さいと母集団が正規分布しているかどうかは見た目ではわからない。正規分布しているかどうかは、 n が30以上は必要と言われている（図7の左右のグラフを比較）。ある統計ソフトでは、 $n < 12$ では、正規性の検定はあてにならないと記されている。したがって、 n が小さい場合には正規性の検定を行う意味はあまりないように思われる。一方で、 n が大きいのであれば、正規性の検定を行って判断すればよい。

n が小さい場合、正規性の検定を行うと正規分布と判断される場合が多い（これは、正規性の検定の数式に当てはめると正規分布と判定されるだけであって、本当に正規分布していることを意味している訳ではない）。この判断を適用してパラメトリック検定を行うのが一般的である。しかし、 n が小さい場合は正規分布しているかどうか判断できないので、正規分布してもしなくても用いることができるノンパラメトリック検定は利用価値がある。しかし、利用するためには、その原理を理解する必要がある。なお、パラメトリック検定は母集団が正規分布する場合に用いるが、この点はあまり厳密に考える

*8 正規性の検定の原理はこのセミナーでは説明しない。

図8 ■ 母分散 σ^2 の計算法

必要はなく、正規分布から多少外れていても用いることができる」とされている。

母集団のパラツキ方 (母分散) はどのように計算するのか?

パラメトリック検定では、母集団は正規分布であり、平均値とバラツキというパラメータが重要である。数値のパラツキ方を表す言葉に分散がある。分散の計算は理解しているという方は多いと思うが、重要なポイントを含むので再確認と思って読んでほしい。

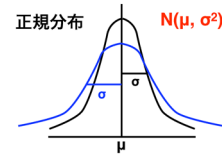
まずは、母集団での分散 (母分散 σ^2) である。標本ではないので注意して欲しい。母集団の大きさNは相当地に大きい数であるが、わかりやすくするため図8では $N=6$ としている。個々のデータは x_1 から x_6 で表している。その平均値は母平均 μ である。図8左のグラフを見て欲しい。データのバラツキは、母平均 μ からどれくらい離れているかで表すことができるので、たとえば、データ x_1 のバラツキは $(x_1 - \mu)$ で表すことができる。6つのデータすべてでこれを計算し合計すれば、バラツキの総和を求めることができる (図8右上の計算式)。しかし、個々のデータは μ よりも大きい場合と小さい場合があるので、 μ を引くとプラスとマイナスが生じ、この総和はゼロになる。ゼロではバラツキは求められないので、2乗 (平方) して足すことに決まっている (図8右中段)。平方して足すので平方和と呼ばれる。平方和は母集団の大きさNが大きくなればなるほど大きくなるので、N (ここでは6) で割り算して、平均の分散を求める (図8下の計算式)。これを母分散と呼び σ^2 で表す。母分散は母集団のデータのバラツキ方を表す指標であるが、データから母平均 μ を引いた後、平方して足しているの、言うなれば (バラツキ) 2 を表している。

母標準偏差 (σ) とは

$\sigma = \sqrt{\sigma^2}$ (平均値と比較するには次元 (単位) を揃える必要があるので、 $\sqrt{\quad}$)

$$\sigma = \sqrt{\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2 + (x_4 - \mu)^2 + (x_5 - \mu)^2 + (x_6 - \mu)^2}{N(6)}}$$

母集団のNは通常、極めて大きな値と仮定。



ばらついているほど σ が大きくなり裾野が広がる

図9 ■ 母標準偏差 σ

母標準偏差 σ とは何か? (図9)

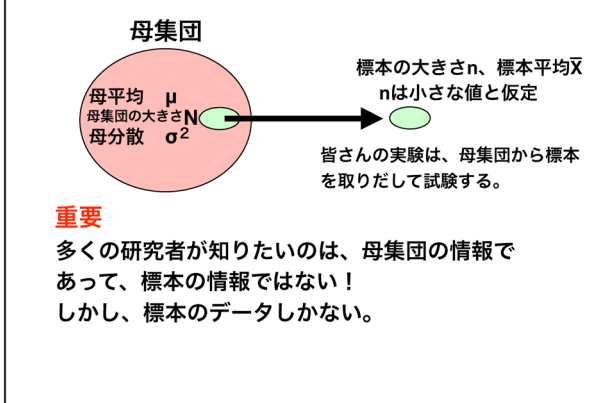
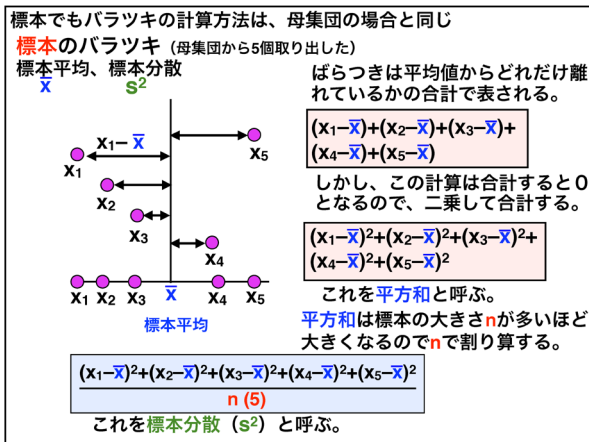
母分散 σ^2 の計算式からわかるように、 σ^2 は母平均とは次元が異なり、平均値と直接比較ができない。そこで、次元をそろえるために母分散 σ^2 をルートし σ とする (図9上の式)。この σ を母標準偏差と呼ぶ。先に、 σ^2 を (バラツキ) 2 と表したが、これをルートすることで得られた σ が (バラツキ) を表すこととなり、母平均 μ と次元が同じになる。母標準偏差は母集団のデータのありそうな範囲を示している。多くの統計書では、単に標準偏差と書かれていることがあり、後で出てくる標本標準偏差や不偏標準偏差と混乱する可能性があることから、本セミナーでは、母標準偏差、標本標準偏差、不偏標準偏差を明確に区別することとする。

σ^2 および σ はデータが大きくばらつければ、当然大きな値となる。正規分布グラフは σ が大きければ、裾野が大きく広がることになる (図9下の正規分布グラフ)。なお、 σ は母平均 μ から、分布グラフの変曲点までの距離である。

母集団から標本n個を取って実験を行う場合 (図10)

生命科学系の基礎研究では母集団で実験することはほとんどない。そこで、母集団から標本n個を取り出して実験することとなる。動物実験ではブリーダーから購入して、ある試験用の食餌を与えたマウスが標本に相当する。その平均値 (標本平均) を \bar{X} と表す (図10)。重要なので再度述べるが、生命科学系研究において、われわれが知りたいのは、多くの場合、母集団の情報であって標本の情報ではない。しかし、得られるのは標本の情報のみである! このポイントをもう一度頭に入れて欲しい (図1)。

母集団からn個の標本をとった場合

図10 ■ 母集団から標本 n 個を取って実験を行う場合図11 ■ 標本分散 s^2 の計算法

標本データのバラツキ方 (標本分散) はどのように計算するのか? (図11)

図11にあるように、母集団から取り出した標本の大きさを $n=5$ とする。(バラツキ)²の計算方法は母集団の場合(図8)と全く同じであり、 $N=6$ が $n=5$ に変わり、母平均 μ が標本平均 \bar{x} に変わったのみである。すなわち、平方和を n で割ると計算できる(説明は割愛するが、図11を見てほしい)。母集団の場合は母分散であったが、今回は標本データであるから、標本分散と呼ぶこととし、 s^2 で表す。

標本標準偏差 s 、不偏標準偏差 u とは? (図12, 13)

当然のことながら、標本分散 s^2 は標本の (バラツキ)² である。標本のバラツキ方を知りたいのであればこの計算でよい^{*9}。標本の (バラツキ) を知りたければ、標本分散 s^2 をルートして、 s を求めればよい (図12の式)。

標本分散をルートすると標本標準偏差 s が求まる。

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + (x_4 - \bar{x})^2 + (x_5 - \bar{x})^2}{n (5)}}$$

S は標本のバラツキ。しかし、多くの研究者が知りたいのは、母集団のバラツキ (母分散 σ^2 と母標準偏差 σ) である。 σ^2 や σ は、 n ではなく $n-1$ で割ると推定できる。

$n-1$ で割ったとき

$$\text{不偏分散 (} u^2 \text{)} \quad u^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + (x_4 - \bar{x})^2 + (x_5 - \bar{x})^2}{n-1 (4)}$$

不偏標準偏差 (u)

$$u = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + (x_4 - \bar{x})^2 + (x_5 - \bar{x})^2}{n-1 (4)}}$$

n は小さいほど不偏標準偏差 (母標準偏差の推定値) と標本標準偏差の違いが大きい!
 n が大きくなると、それほど大きな違いがなくなる。

図12 ■ 標本標準偏差と標準偏差

この s を標本標準偏差と呼ぶこととする。しかし、群間比較研究を行う多くの研究者が知りたいのは、標本分散や標本標準偏差ではないはずである。研究者は母集団の (バラツキ)²、すなわち母分散 σ^2 および母標準偏差 σ を知りたい (推定したい) のである。実は、標本分散は母分散と同じではなく、母分散よりも少し小さな値となることが理論的にすでに知られており、母分散を推定するためには n ではなく、 $n-1$ で割り算すると丁度よいことがわかっている^{*10}。ここで、 $n-1$ は自由度と呼ばれる^{*11}。すなわち、 n で割ると標本分散がわかり、自由度 $n-1$ で割ると母分散を推定できるのである。

$n-1$ で割った分散は不偏分散と呼び u^2 で表すこととし、標本分散 s^2 と区別する (図12の中段)。“不偏”とは偏らないという意味であり、母分散 σ^2 と等しくなるように標本分散を補正した値である^{*12}。統計書を読む場合、不偏分散と標本分散を明確に区別しておかないと混乱する。また、本によっては、不偏分散を標本分散と記述してある場合があり、注意が必要である。

^{*9}たとえば、あるブリーダーから購入したICRマウス6匹の血清コレステロール濃度がどれくらいバラついていてほしい場合は標本分散でよい。しかし、世界中のICRマウス(これを母集団と仮定)の血清コレステロール濃度のバラツキを推定したければ、不偏分散を計算すべきである。

^{*10}実際に正規分布する母集団から標本を取ることを何度も繰り返して、標本分散を計算してみると母分散よりも小さくなるのである。また、理論的に計算で求めることができるので、興味があれば『バイオサイエンスの統計学』(南江堂)を読んで欲しい⁽⁶⁾。

^{*11}自由度については、次回「改訂増補版:II」で解説する。

^{*12}もう少し詳しく述べると、不偏性があるとは、標本の大きさ n にかかわらず(依存せず)、母平均や母分散を偏りなく推定できることを意味する。もちろん、不偏分散 u^2 は標本から計算した値であるから、母分散 σ^2 と必ずしも一致する値にはならない。しかし、母集団から何度も標本をとって u^2 を計算することを繰り返して平均すると σ^2 に一致するのである。

標本分散 s^2 (標本の分散)・・・母分散と式は同じ。

$$\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+(x_4-\bar{x})^2+(x_5-\bar{x})^2}{n(5)}$$

標本標準偏差 s ・・・母標準偏差と式は同じ。

$$\sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+(x_4-\bar{x})^2+(x_5-\bar{x})^2}{n(5)}}$$

不偏分散 u^2 (母分散を推定)

$$\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+(x_4-\bar{x})^2+(x_5-\bar{x})^2}{n-1(4)}$$

不偏標準偏差 u

$$\sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+(x_4-\bar{x})^2+(x_5-\bar{x})^2}{n-1(4)}}$$

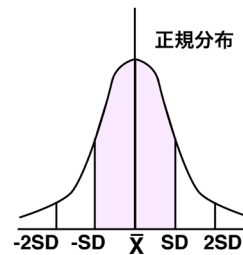
不偏標準偏差も標本標準偏差も標準偏差と呼ぶ場合があるので注意!

図13 ■ 分散と標準偏差のまとめ

母分散を推定する不偏分散 u^2 のルートが u であり、不偏標準偏差と呼ぶこととする^{*13}。不偏標準偏差は母標準偏差を推定する値となり、母集団のデータのありそうな範囲を推定する値である。この不偏標準偏差 u が、母集団のバラツキを推定したい研究者が一般的に用いる標準偏差 (standard deviation; SD) である。もし、この u を標本データのバラツキと考えているならば、それは間違っている。標本データのバラツキは標本標準偏差 s である。間違えないで欲しい。ところで、実は標本標準偏差 s もまた標準偏差と呼ばれるため、混乱する原因となることから、統計書を読む際には注意を要する。母集団を推定する場合は、 $n-1$ で割った標準偏差 (不偏標準偏差) が正しい。Excel の関数では、不偏標準偏差は

*13 “不偏” についてはすでに*12で説明した。ここで u を不偏標準偏差と名付けた。不偏標準偏差は不偏の定義からすると母標準偏差 σ を推定する値のはずである。「統計検定を理解せずに使っている人のためにI」や拙著でもそのように書いた^(1,2)。しかし、不偏分散 u^2 をルートした u は不偏ではないという指摘を受けた。調べてみるとそのとおりであった⁽⁶⁾。ここで、母標準偏差 σ を推定する値を真の不偏標準偏差と名付けると、 u は真の不偏標準偏差とはズレがある。このズレは、*12の最後に述べた、何度も標本をとって計算した u^2 の平均値をルートした値 (これが真の不偏標準偏差) と、 u^2 を計算する毎にルートして u を計算し、その平均を求めた値とが一致しないことに起因する。我々は通常一回しか実験しないので、一回の試験で求めた u^2 をルートするしかなく、真の不偏標準偏差を知ることはできないのである。 n が大きいと (おおむね $n > 10$)、 u は真の不偏標準偏差に近いが、 n が小さいと ($n \leq 10$)、ズレが大きくなり補正が必要になる (n に依存して変化し n が小さいほどズレが大きくなる)。従って、真の不偏標準偏差を知るには u を1未満の係数で割って補正する必要がある。例えば、 $n=5$ では係数は0.94であり、 $u/0.94$ が真の不偏標準偏差となり σ を推定する値となる⁽⁶⁾。 $n \leq 10$ で実験している研究者は多いと思うが、データを標本平均 \bar{x} SD で表記し、SDとして u を用いる場合は、真の不偏標準偏差ではなく若干低めの値を表記していることになる。このセミナーでは、 u は真の不偏標準偏差ではないが、不偏標準偏差と呼ぶこととし、少しズレがあるが母標準偏差 σ を推定する値として話を進める。

標本平均 \pm SD (不偏標準偏差) は、



Excelの関数では、
stdev.s (またはstdev) → 不偏標準偏差
stdev.p (またはstdevp) → 母標準偏差、標本標準偏差

図14 ■ 標本平均 \pm SD の意味

stdev.s (またはstdev) であり、 n で割った標本標準偏差はstdev.p (またはstdevp) であるので、間違わないように注意しなければならない (母標準偏差も N で割るので、stdev.p で計算できる) (図14)。なお、標本の大きさ n が小さいほど、標準偏差 u と標本標準偏差 s の違いが大きくなり、 n が大きくなるとその違いは小さくなることは、式を見ればわかるはずである (図12)。

標本分散、標本標準偏差、不偏分散、不偏標準偏差をまとめて、図13に記載する。この区別をしっかりと認識し、統計書の混乱した記述に惑わされないことが肝要である。

標本平均 \pm SD は何を意味しているのか? (図14)

母平均 μ と母標準偏差 σ を推定したい場合、標本平均 \bar{x} は母平均 μ を推定する値と位置づけられる。また、すでに述べたように不偏標準偏差のSDは母標準偏差 σ を推定する。 $\bar{x} \pm$ SDの範囲は、図14にあるように、正規分布する母集団のデータの68%が存在することが期待される範囲を示している。また、 $\bar{x} \pm 2$ SDでは、正規分布する母集団のデータの96%が存在することが期待される範囲を示している。

研究者は標本平均 \pm SD で何を表現したいのか? (図15)

$\bar{x} \pm$ SDは標本平均 \bar{x} から母平均 μ を、SDから母標準偏差 σ を推定していることから、『らくらく生物統計学』(中山書店)によると、研究者は母平均 μ を推定したいのは当然として、さらに、母集団のデータのバラツキにも同様に關心があるはずであると述べている⁽⁷⁾ (図15)。

研究者はSD（不偏標準偏差 σ ）で何を表現したい？

研究者は母平均 μ と母標準偏差 σ に関心があるはず。

SDを用いている研究者の皆さん！
母標準偏差 σ に関心がありますか？

標本のバラツキを知りたい実験であれば、**標本標準偏差 s** すなわち、 n で割った値を用いる。

もし母平均 μ だけに関心がある場合は、**標準誤差（SE）**を用いたほうがよい。
標準誤差SEは次回で説明する。

図15 ■ 研究者は標本平均 \pm SDで何を表現したいのか？

SDを用いている研究者は、母集団のデータのバラツキに関心をもって標本平均 \pm SDを用いているであろうか？ もし、標本平均にしか興味がなければ、SDよりもSEを用いたほうがよい（SEは次回「改訂増補版：II」に登場する）。

おわりに

今回は、パラメトリック検定およびノンパラメトリック検定の基礎について述べる。パラメトリック検定では、標準誤差（SE）が登場し、標準偏差SDとの違いを明確にする。本セミナーを書くにあたっては多くの統計

書にお世話になった。特に、文献に挙げた本は、比較的わかりやすく書かれている⁽⁶⁻⁹⁾。

文献

- 1) 池田郁男：“実験で使うとこだけ生物統計1 キホンのキ”改訂版，羊土社，2017.
- 2) 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.
- 3) A. Reinhart: Statistics done wrong: The woefully complete guide, William Pollock, 2015.
- 4) M. Baker: *Nature*, **533**, 452 (2016).
- 5) 市原清志：“バイオサイエンスの統計学”，南江堂，1990.
- 6) 石居進：“生物統計学入門”，培風館，1975.
- 7) 足立堅一：“らくらく生物統計学”，中山書店，1998.
- 8) 石村貞夫：“すぐわかる統計処理”，東京図書，1994.
- 9) 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>前はバドミントンでしたが，今はドライブと温泉巡り

Copyright © 2019 公益社団法人日本農芸化学会
DOI: 10.1271/kagakutoseibutsu.57.492