統計にだまされるな

川瀬　雅也

「この結果を統計処理すると」あるいは「有意差がある」などの文章を目にしたことがない人はいないと思う.「統計処理して有意差があった」と言われると「差があるのだな」と納得してしまうのではないだろうか．しかし，よく考えてみると「そもそも有意差とは何だろうか」「差があるとは何だろうか」という問にすぐに答えることができる人は少ないのではないか.

何故，答えることができないのか．これは，統計処理はできても統計学が分かっていないからである. 今回は，初心に戻って統計学の本質を学んでみたい.

統計学は正しいのか

よく耳にする疑問だと思う.「実験を行っていると違いがあると思われるのに，統計処理をしてみたら違いがないと出た」ということはよくあると思う．このとき，統計の答えは正しいのかという疑問を持つこともよくあることである.

そもそも統計学とは「調査により得られたバラツキのあるデータを確率論的に解析する方法を研究する学問，あるいは，データの要約もしくは解釈の根拠を与える学問」であり，データ処理の方法ではない．生物工学会の諸兄が使われる統計処理は「統計学の研究成果を用いて，データを要約する作業」といえばお分かりいただけるのではないかと思う.

では，表題の質問「統計学は正しいのか」の答えは「統計学が要求する条件が満たされている状況下では統計学は正しい」となる．つまり，データの統計処理を行う場合，根拠となる統計学が成り立つ条件に合致しているのか，また，どの程度ずれているのかを評価した上で処理を行えば妥当な結果を得ることができるわけである．これだけでは，何を言っているのか分からないと，おっしゃる方も多いと思う．平均値を例にして具体的に説明しよう.

平均値とは

よく使う平均は「データの和をデータ数で割った値」つまり算術平均である．計算式は，

$$\bar{x}=\frac{1}{n}\sum_i x_i$$

ここで，$\bar{x}$は平均，nはデータ数，x_iはデータである.

非常に簡単な式であり誰も間違うことはないはずである．しかし，この平均値が何を意味しているのかと問われると，答えることができるだろうか.

質問を変えて「何のために平均値を求めるのか」としたらどうだろうか．平均とは統計学的には「2個以上のデータから，データ全体の性質を記述するための指標の一つ」である．我々の使い方に近い言い方に言い換えれ

連載にあたって

「生物工学基礎講座バイオよもやま話」はちょうど2年前から連載を開始しました．実験の原理や技術の歴史を述べるだけでなく，裏側にある苦労話も紹介し，楽しみながら基礎知識を学んでいただきたいと思いました．内容によってはすでに引退されている先生方にお願いした原稿もありました．和文誌がWEB上で誰でもダウンロードできるようになったこともあり，多くの方にキーワード検索され，参考書代わりに活用されてきました．学会誌はゴミ箱に捨てられていても，ダウンロードされた「よもやま話」のファイルはiPadなどで読まれており，複雑な気持ちで，時代を感じてしまいます．もともと期間限定の企画でしたが，継続を望む声も多く，このたび「続・生物工学基礎講座―バイオよもやま話―」として続けることになりました．これまで同様に会員皆様からのご意見や質問を募集致します．会員の関心が高い内容は今後取り上げて参りたいと思います．（和文誌編集委員会）

＜要望・質問の宛先＞E-mail: info@sbj.or.jp

***著者紹介**　長浜バイオ大学　バイオサイエンス学部（教授）　E-mail: m_kawase@nagahama-i-bio.ac.jp

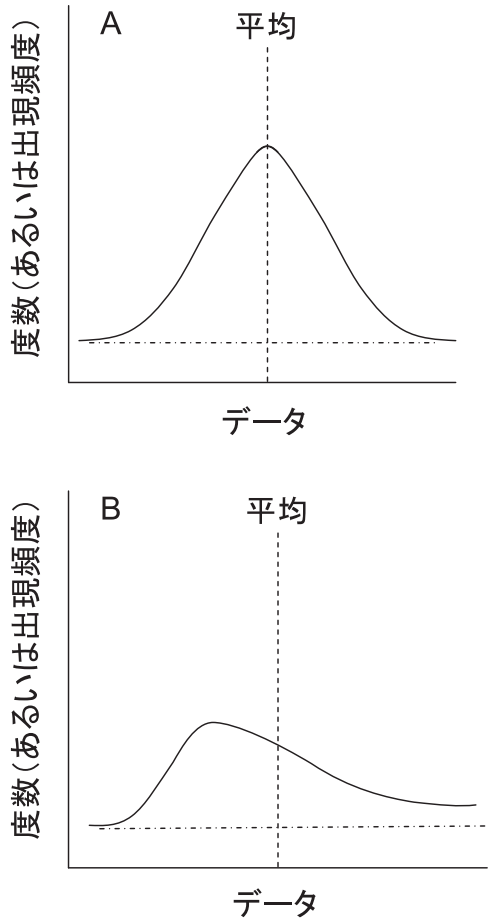

図1．典型的な2種類のデータ分布．（A）左右対称の分布，（B）偏りのある分布．

ば，「平均値とデータの差から，そのデータの得られる確率（あるいは，出現頻度）を簡単に推定できるような指標」となるはずである．確かに図1(A)のように平均値を中心に左右対称に近い分布をデータがとる場合は，統計学的な意味を持つ平均値が得られる．しかし，どのような場合でも平均値がこのような指標となるとは限らない．たとえば，図1(B)のようにデータに偏りがある場合やいくつも分布に山がある場合などは，単に平均値とデータの差から，そのデータの出現頻度を簡単に推定することはできないので，このような場合の平均値は単なる計算値以上の意味をもってこない．また，どのデータも同じような頻度で現れる一様分布の場合は平均値を計算することに意味があるのかどうかも考える必要がある．

多くの場合，機械的に平均値を計算し議論に使っているが，その値が本当に統計学的に意味のある値なのか考えてみたことはあるだろうか．多くの場合は，考えずに使っているのではないだろうか．

単純な平均であっても，平均が統計学的な意味を持つには，幾つかの条件が整う必要がある．「平均値だけで議論しても，何の意味もない」と指摘されることがあるとおもうが，図1(A)から外れた分布のデータについては的を射ている．しかし，単にデータの分布が広い（分散が大きい）との理由の場合は，この指摘はデータの見方という面からは確かな意見であるが，統計学的には間違った指摘である．平均といえども，その奥が深いことを分かって頂けたであろうか．

ここまでの話で気がついたのではないかと思うが，なぜ，仮説検定でデータの分布が正規分布なのかどうかをうるさく言うかの理由がここにある．仮説検定は，平均値について議論する統計的手法であるので，統計学的に意味のない平均値を取り扱うことができない（正確には，取り扱っても，その結果には何の意味もない）からである．

有意差とは

仮説検定の話が出たので，次に有意差について考えてみたい．「検定の結果，両者に有意差が認められた」という議論を置く耳にしたり，目にしたりすると思う．このように聞くと，「両者の間に違いがあるのだ」と納得する（納得してしまう）のではないだろうか．まず，仮説検定の際に設定する有意水準の正体を解き明かしてみよう．

仮説検定の手順を思い出してほしい．最初に帰無仮説を設定する．帰無仮設は両者に差がないとする．この理由は，“差がない”の否定（帰無仮説の棄却という）は反例を一つ出せばできるからである．この逆は非常に難しく，普通は帰無仮設とはしない．有意水準とは「帰無仮設を棄却する基準」と定義されている．また，同時に「帰無仮説が成り立っているにもかかわらず，棄却してしまう確率（第1種の過誤）」とも等しい．

仮説検定では，データより検定統計量という値を計算し，データの分布（多くの場合はt-分布）関数より，その値が生じる確率を計算する．その確率が有意水準を下回ったときに，帰無仮説を棄却し有意差があるという．

言い換えれば，「有意差があるとは，積極的に違いがあることを認めているのではなく，有意水準を下回るような確率でしか生じない事象は，偶然生じたとは考えられないので，帰無仮説が成り立っているとは考えにくい」という意味である．つまり，有意差があるといっても，違いのあることを完全に保障しているのではなく「違いはないと積極的に言うことは少し難しいのではないか」という気持ちが強い消極的否定の意味合いが強いことを理解していただきたい．有意とは統計学的に「偶然に起こったとは確率論的に考えにくく，何かの意味がそこに

はある」という意味を持つ用語であることからも理解できると思う．

また，帰無仮説が棄却できない場合も同様である．仮説検定で帰無仮説が棄却されなかった場合「両者の間には差がない」と考えている方も多いのではないだろうか．これは大きな誤解である．帰無仮説が棄却されないということは「両者の間に生じている違いは偶然生じたもので，違いがあるとは考えにくい」という意味であり，どこかに「違いはあるのかもしれないが，あるとはハッキリと言うことはできない」という意味合いを含んでいる．

最初に「実験を行っていると違いがあると思われるのに，統計処理をしてみたら違いがないと出た」という例を紹介したが，仮説検定の結果の意味するところをしっかりと理解すれば，この結果は「違いがないと出た」のではなく「違いがあるとは言えない違いである」もしくは「違うということができるほどの違いではない」という結果であり落胆する必要はないと思う．

標準偏差と分散とは

平均値と同様に基本統計量として算出される量に標準偏差と分散がある．ここで，この両者の意味について考えてみたい．

まず，思い出していただきたいのが「標本」と「母集団」という言葉である．両者の関係については後でもう一度触れることにするので，今は「母集団＝研究対象全範囲」「標本＝データ」と理解して頂ければ結構である．母集団を直接調べることは不可能な場合が多いので，その中から標本を選び調べているのが実験であり，調査である．この二つの用語は，統計学の最初に出てくる言葉であり，統計学の理解には欠かせない言葉であるので，再度，確認いただきたい．

確率論では分散とは「確率変数（データのことだと思っていただいて結構）がその分布の中心（確率的には期待値，平均値と思っていただいて結構）からどの程度散らばっているかを表す量」とされている．統計学では平均値からの散らばりの度合いと理解している．

標本から計算される分散を標本分散（s^2）とよび

$$s^2 = \frac{1}{n}\sum_i \left(x_i - \bar{x}\right)^2$$

で求められる．標本分散は母分散（σ^2）とは異なることが知られており，母分散の推定量（不偏推定量という）として不偏分散が用いられる．不偏分散は

$$V = \frac{1}{n-1}\sum_i \left(x_i - \bar{x}\right)^2$$

として計算される．分散について言えば母集団の分散を推定することができるわけである．

では，標準偏差はというと標本に関する標準偏差は標本分散の平方根として求めることができる．

$$s = \sqrt{\frac{\sum_i \left(x_i - \bar{x}\right)^2}{n}}$$

いろいろな統計処理では標準偏差として不偏分散の平方根を用いている．この値は次の式で求められる．

$$\sigma = \sqrt{\frac{\sum_i \left(x_i - \bar{x}\right)^2}{n-1}}$$

しかし，この値は母集団の標準偏差の不偏推定量ではない．母集団の標準偏差は母集団が正規分布に従う場合，

$$\sigma_t = \sqrt{\frac{n-1}{2}}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\sqrt{\frac{\sum_i \left(x_i - \bar{x}\right)^2}{n-1}}$$

が不偏推定量となる．$\Gamma(x)$ は Γ 関数である．

統計処理では単に不偏分散の平方根を用いるという意味で使っていたのが，いつの間にか不偏分散の平方根が母集団の標準偏差の推定量であると勘違いされてしまったわけである．先に説明した有意差の意味の勘違いもそうである．

不変分散は母集団の分散の推定量であるが，その平方根は違うという点はお分かりいただけたと思う．では，不変分散の平方根（σ）にはどのような役目があるのだろうか．図2を見ていただきたい．

よく御存じと思うが平均からσ離れた点が関数の変曲点にあたる．つまり，σには正規分布の形を決めるとい

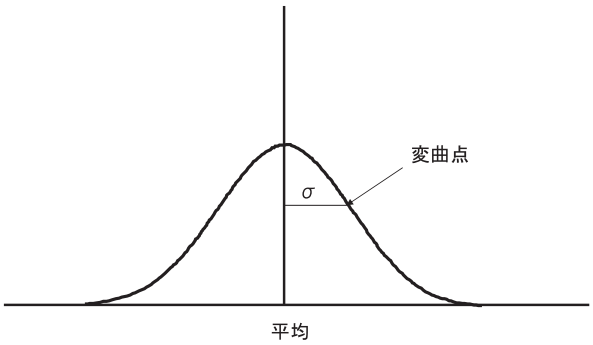

図2．正規分布

う重要な役割があり，さらに，

μ ± σの範囲に構成要素の68.26%

μ ± 2σの範囲に構成要素の95.44%

μ ± 3σの範囲に構成要素の99.73%

という正規分布の性質を表す量でもあり，母集団の不変推定量でないので大した役割はないと思ってはいけない．

分散と標準偏差の意味をしっかりと押さえずに使うと思わぬ落とし穴にはまることもあるので注意されたい．

母集団と標本とは

統計学を勉強するとき，最初に理解してほしい用語が母集団と標本である．先にも触れたが母集団＝研究対象全範囲」「標本＝データ」という関係にある．これだけでは，よくわからないという諸君のために，もう少し詳しく見て行こう．まず，母集団であるが「研究対象全範囲」とは何か．よく使う例であるが「風邪薬の効き目を調べる」という調査を行うとする．この場合，母集団は「調査の時点で風邪をひいている人すべて」となる．日本国内に限っても，今，風邪をひいている人は何人いるかなど分かるはずもない．しかも，その人すべてに，この薬を飲んでもらい効き目を調べるなどできるはずがないことは，すぐに，分かってもらえると思う．このため何人かの人に効き目を調べるために風邪薬を飲んでもらい，そのデータを使って効き目を評価するわけである．この風邪薬を飲んだ人が標本にあたる．別の例として，酵素活性の測定を考えてみたい．微生物中の酵素活性を調べる場合，培養を行い微生物から酵素を取り出して調べる方法をとったとする．微生物の培養の状況や取り出す際の条件により酵素活性は影響を受ける．

図3を見ていただきたい．本来の酵素活性の値（実際には，どの値がそうなのかは簡単に決めることが難しい）と測定値との間の差（誤差）の分布を示している．誤差の分布は中心極限定理により正規分布になることが知られている．何度か実験を繰り返すと平均値が本来の活性値に近くなってくる．多数回（母集団にあたる），実験を繰り返すと図3のような分布曲線を描くことができ，本来の酵素活性値を知ることができるが，酵素活性ばかり測定を行うことができないので，標本数を幾つとるか（酵素活性の測定を何回行うか）が重要となる．1回や2回の測定では図3のAやBに示すような極端に平均から離れた値を測定する可能性もあり，統計学的には6回以上の測定は必要になる（t-分布の成り立つ最低標本数が6である）．この2例を見ても分かるように，母集団のすべてを調べることは非常に難しい場合が多い．母集団の一部を無作為に抽出し標本として調べ，その結果が母集団の調査結果と考えるわけである．

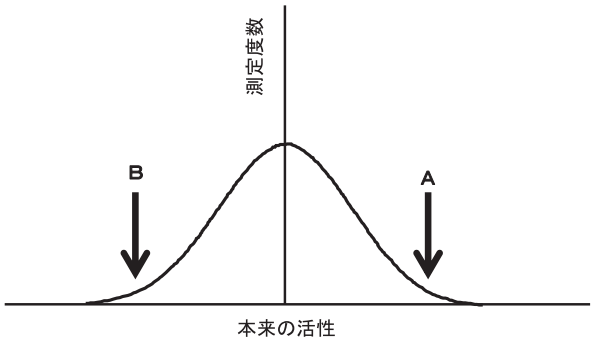

図3．平均と誤差の関係

ここで問題となるのが無作為というキーワードである．無作為とは「意図的に手を加えることなく，偶然にまかせること」と広辞苑にある．統計学の立場から言えば「偏りなく」という言葉も入る．実験の場合を考えると，ある条件だけを変えて，他は同じであるので，変更した条件の影響を見ているとしても，条件変更が他の条件に影響と与えていない保証ができるかと言えば，これは非常に難しいことである．このようなことを考えると実験なんかできないとおっしゃると思う．確かにそうである．ここで何が言いたいかというと，統計学の要求を満たすような実験は不可能に近い場合が多いということである．言い換えれば，実験結果を統計処理した結果を妄信的に信用して議論することは危険であるということである．最初に，もう一度戻るが「実験を行っていると違いがあると思われるのに，統計処理をしてみたら違いがないと出た」場合，標本の取り方に問題がある場合もあり，もう一度，実験系を見直すということを忘れてはいけない．

“統計処理で出た結果はすべて正しいと思い込んではいけない．”このことを筆者は伝えたいわけである．

再び平均値

ここまでの話で，少し統計が分かった気がすると思ってもらえれば，この原稿は成功であると思う．少し分かって来たときに陥りやすい間違いがあり，紹介しようと思う．微生物の増殖曲線を考えてみよう．微生物の増殖は図4のように培養初期には，ゆっくりと増殖（誘導期）し，その後に対数増殖期（指数関数的に菌体が増加）に入る．培養終期には菌体量は飽和に達する．培養を並行して何本か行うと誘導期は培養ごとに異なってくる．対数増殖期の比増殖速度は全培養で同じような値となる．このような場合，各時間での菌体量は正規分布に近くなる．こ

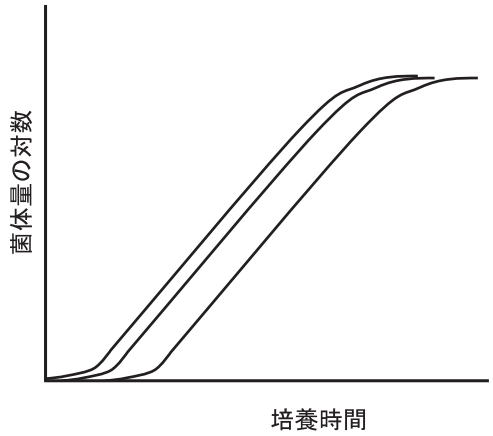

図4. 菌体の増殖曲線

うなれば，幾つもグラフを書くのは見難いので平均をとって1本のグラフにしてもいいと考えるだろう．しかし，実際に平均をとって描いたグラフから求められる比増殖速度は個々の測定で出した値と異なってくる．

何故かを考えてみよう．まず，上の例で皆さんが求める平均は

$$\bar{x}=\frac{1}{n}\sum_i x_i$$

による平均（算術平均）であると思う．小学校の時代から「平均＝算術平均」という指導がなされており，無理もないことだと思う．もう一度，図4を見ていただきたい．このグラフの縦軸は対数である．今，ある時点で測定した菌体量としてA，B，Cの3つの値を得たとする．この平均値は

$$\log\bar{x}=\frac{1}{3}(\log A+\log B+\log C)$$

となるはずである．対数というのがミソである．もう少し式を展開してみよう．

$$\log\bar{x}=\frac{1}{3}(\log A+\log B+\log C)$$
$$\log\bar{x}=\frac{1}{3}\log ABC$$
$$\log\bar{x}=\log\sqrt[3]{ABC}$$
$$\therefore\ \bar{x}=\sqrt[3]{ABC}$$

となる．この右辺は幾何平均とよばれる量である．統計的に平均をとるという判断は正しいが，どのような平均をとるかを判断しないと，とんでもない間違いを犯すことになる．統計学は数学の一分野であることも理解していただけたのではないだろうか．

統計ソフトは大丈夫？

統計処理を行う時，必ず統計ソフト（表計算ソフトも含む）を使うと思う．統計処理の結果は“**単なる計算結果**”であるので，ソフトの使い方を間違わなければ“**正しい計算結果**”を得ることができる．しかし，統計処理の原理（統計学的な基礎）とその手法の限界を知らなければ“**計算結果**”を“**解析結果**”にすることはできない．つまり，偶然見つけたフリーソフトは別として，ある程度，使われているソフトなら大丈夫と考えてよい．その結果を生かすも殺すも，皆さん次第である．

ソフトの使い方さえわかっていれば統計は簡単にできるというのは大きな間違いであり，くれぐれも“統計にだまされない”ように勉強していただきたい．

孫子曰く「**彼（てき）を知り，己を知れば，百戦危うからず**」である．

統計学の参考書は多数出版されている．どれがいいかという優劣はつけることができない．まず，手に取ってみて，これなら読んでみようかと思った本が，自分に合った本である．是非，自分に合った参考書を探していただきたい．このような理由で，あえて参考書は示さない．