The Society for Bioscience and Bioengineering, Japan

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

NII-Electronic Library Service