

はしがき

本書は、医学、薬学、農学など、情報科学や統計科学を専門としない生物系の学科を履修した学生、大学院生で、すぐにデータ分析を行いたい状況にある人に向けて書かれた統計を扱った書籍です。近年、次世代シーケンサーやマイクロアレイなど、あるいは、多くの検査項目を含む臨床検査データなどを集計して、統計解析を行わなければならない機会が増えてきました。このようなビッグデータの時代において、情報科学や統計科学の重要性はますます高まっています。データサイエンスの重要性の高まりとともに、関連の書籍も多く出版されるようになってきました。統計解析ソフトRやプログラミング言語Pythonなどを用いたデータ分析も徐々に普及しています。しかし、関連の書籍は、ややもするとプログラミングや統計解析を前提としていたり、数式や数学用語が多かったり、なかなかとつきにくいものも多いです。一方で、全くの初心者向けに抽象的でわかったような気にさせるような「なんちゃって統計本」も出回っています。

生物系の学生・研究者が本当に押さえるべき「統計の基礎」とはどのようなものでしょうか。実際のデータ分析の場面ではすぐにでも結果を出したい、すぐに結果を評価して意思決定を行いたいという需要も少なくありません。そのなかで、実際の統計ツールをどのように使いこなし、どのようにデータ分析の結果を解釈すべきかという視点に立った書籍は多くないように思います。ステューデントの t 検定やマン・ホイットニーの U 検定にはどのような利点、注意点があり、どのような場合に使うのか。そのような視点から書かれた本はこれからすぐにデータ分析にとりかかりたい人には切実でしょう。

筆者は毎日のように、次世代シーケンサーやマイクロアレイなどのデータ分析をこなしています。そのなかで、データ分析のノウハウ、コツのようなものがあり、そのような視点に立った書籍が必要ではと考えていました。また、そのためには数式を省いた「なんちゃって統計本」ではなく、理解を助ける最低限の数式などの記述は必要であると考えていました。

本書は、統計や数学などの前提を必要としない生物系向けと謳いながらも、数式が多めであることも特徴かと思います。数式の細かい記述はともかく、直感的理解に必要な最小限の数式を掲載しました。最低限これだけの内容を踏まえれば、次世代シーケンサーやマイクロアレイなどのデータ分析の世界に入っていける。そのような内容の書籍としました。

実践的に使えることを優先したために、理論的なことを置き去りにしている感はありませんが、既存の統計学の教科書にあるような数学的な厳密さよりとにかくデータ分析がすぐにできるようになることを優先しています。一方で、入門者向けの「なんちゃって統計本」にあるような、冗長な周りくどい説明も意図的に避け、手を動かして実感できるようなものにするのを心がけました。本書は、その点、他の統計の教科書にあるような記述も省き、使い方やその注意点を優先して記述することに徹しました。その意味では異端の本になっていると思います。

ビッグデータ時代にあって、生物学は、ケミカルをベースにしたものから数学や情報科学をベースにしたものに大きく変貌しつつあります。今後、数学や情報科学は生物学にとってますます重要なものになっていくでしょう。コーディングとアルゴリズムは生物学にとって不可欠なスキルになると思います。このようななかで、最初の一步を踏み出すきっかけとして本書を活用していただければと思います。

2017年2月

東京農工大学特任教授
石井一夫