

訳者からの序

この本は、たとえば大学で研究室配属になって「ああ……これから統計解析の勉強をしなければ……やっぱパソコンとか使わないといけないよね……」と憂鬱な気分になっている学生が、とりあえず何も考えずに買って大丈夫な本です。この本にしたがって統計解析の実習をやることで、解析作業の基礎的なステップが身に付きます。使うのは、パソコンとRStudioです。世の中ではパソコンと言えばWindowsが多い一方、生命科学分野の研究室ではMacが増えていると思いますが、RStudioの使い方はどちらでも同じです。さらにRStudioはパソコンがなくても、タブレットでも使えます（RStudio Cloudというのがブラウザから使えます）。つまりいろんな人が、この本1冊で勉強できるわけです。大学の研究室や高校の部活（学生だけでなく先生たちにも）、またデータサイエンスに興味を持つ企業の研究チームなどにも最適です。Rはバージョンアップがよくありますが、この本にはR version 3.5.1/RStudio 1.1.463をWindows 10での結果を載せています。Rのバージョンなどによっては実行結果が少し違ったりするかもしれませんが、大きな問題になることはないだろうと思います。

この本には、統計そのものの原理や検定の理論の解説はほとんどありません。統計解析の知識が少しある人に向けて、堅実な結論を出すための作業手順をケーススタディで練習する内容です。なので、最初のページから順に進んでいくように書いてあります（知りたいところだけを必要な時に抜き出して読む、というような書き方にはなっていません）。このような作業手順を身に付けるのは、間違った結論を出して、そうとは知らずに堂々と主張する、といった悲劇を避けるためです。多くの統計検定には、データの分布やモデルの当てはまり方などについて前提条件があります。また、データを観測した時の実験計画の意図をよく理解していないと、適切なモデルを選べません。これらの点について繰り返し実習して、どんなデータ形式にどんな検定法を使うのかを体験しつつ、作業手順を体で覚えるのがこの本の目的です。

近年、人工知能やビッグデータといったキーワードが注目されていますが、この本はそのどちらも扱っていません。ただ、これらはデータサイエンスの一分野です。そしてこの本はデータサイエンスのもっとも重要な基礎を身に付けるためのものです。生命科学分野でビッグデータ解析に興味を持っている人も、基礎的な解析作業を身に付けておいて損はありません。むしろ、身に付けておくべきだろうと思います。これが、自分の扱おうとしているデータがビッグデータかどうかを判断する力の基礎にもなるでしょう。

生命科学分野では、遺伝子発現データを始めとして大規模な数値データがよく見られますが、これらは必ずしもビッグデータではありません。遺伝子数が多くてもサンプル数が少なければ、各遺伝子についてはむしろ小サンプルです。小サンプルデータの解析にはしばしば落とし穴があって、出版されている論文でも怪しい統計解析がときどきあります。その原因の一部は、モデルが適切でなかったり、データの分布が前提を満たしていなかったりすることです。これにはやはり、適切で堅実な作業手順を身に付けることが重要です。

この本はまさにそのための本です。「基礎的で堅実な作業方法」が主題なので、これ一冊あれば生命科学分野の統計解析法は一通りわかるといった網羅性は（全然）ありません。実際の研究現場では、この本に載ってない解析法をたくさん調べて使う必要があります（たとえば、フィッシャーの正確確率検定とかウィルコクソンの順位和検定とか、またいろんな次元縮約も重要です）。RStudio（が内部で使っているR）には必要な機能はほぼすべて揃っていますが、それを使うためには、自分でやり方を調べる必要があります。と言っても臆することはありません。この本で基本的な考え方と作業の流れを理解したなら、応用的なことを理解する力も備わっているとと言えるでしょう。この本の中では「Rのヘルプ画面が役に立つ」と書いてありますが、全部英語なので、敷居が高いと思ったら調べものはネット検索か、または自分にとってわかりやすい解説書を探すとうまいと思います。

この本は、楽しい本です。また、難しい統計の理論は（ほとんど）ないので、わかりやすいと思います。ちゃんと解析して、堅実な結論を出して、きれいな図を描くと、一流の研究者がやっているのと同じ統計解析が自分にもできた、この結論は自信を持って発表できるのだ、という喜び、うれしさが湧いてきます。「自信が持てる結論と発表のための解析手順を身に付けて、楽しく、みんなの役に立つ研究ができる」のが幸せな研究生活だろうと思います。みなさんがそうできるよう、心より祈っています。この本の原著者らは楽しくやってそうですしね。

2019年2月

富永大介