

# はじめに

本書は、手っ取り早く R を使って確率統計学を学びたい読者のために、『R で学ぶ確率統計学（一変量統計編）』と『R で学ぶ確率統計学（多変量統計編）』の 2 冊を再編集したものである。

R は、オークランド大の Ross Ihaka と Robert Clifford Gentleman（二人の R）によって 1993 年に開発された統計解析ソフトウェアであり、R 言語とも呼ばれている。フリーで提供されており、Windows, Mac, UNIX 上で動作する。R は R Studio と一緒に利用することにより、スムーズな対話的統計解析を行うことができる。データサイエンスにおいては、R とならんで Python も多く利用されているため、どちらを学ばよいかと訊かれることがある。対話的な統計処理については圧倒的に R の方が手軽で便利である。R は、統計処理に限れば、ほとんどどんな処理も可能で、多くの場合、数行の入力で済む。Python で同様のことをするには、若干余計な手間がかかる。一方、人工知能なども含め、より汎用性を求めるのであれば、Python に分があるかもしれない。筆者（神永）は、数式処理、暗号研究では主に Python を使っているが、実データの統計処理やモンテカルロシミュレーションには R を使っている。暗号のプログラムには Python の方が適切だし、統計処理に関しても、まとまった量のデータの細かいハンドリングが必要な場面では Python の方が便利だ。R と Python のいずれにも利点があるので、両方使えるようにして、用途に応じて使い分けるのが賢い方法だと思う。

本書の内容は、一変量統計編、多変量統計編に含まれているが、再編集にあたり、数学的な説明は大幅に簡略化し、R の操作、可視化の方法についての記述を拡充した。数学的な説明の簡略化により、数学に苦手意識がある学生にも近づきやすいものになったと期待している。

本書のゴールは、読者が一通りの標準的な統計学の知識を得たのち、R を用いて実データの分析ができるようになることである。確率統計学を学ぶために必要な基礎的概念から始まり、R を使って実際に手を動かして学ぶ段階を経て理解を深めていく。はじめにデータの可視化、平均、分散、標準偏差、相関などの基本統計量の扱いを学び、これらを基礎として区間推定と検定のような一変量推測統計を理解していく。さらに、より進んだ統計解析として、分割表の分析、単回帰分析、重回帰分析、モデル選択のために必要な AIC（赤池情報量基準）、一般化線形モデルとして、ロジスティック回帰、プロビット回帰のほか、計数データの分析を学ぶ。今日では、いずれも統計学の常識に属するものであるが、一般化線形モデルまで習得すれば、大抵の統計解析は可能となるだろう。

本書でベースになる数学の予備知識は、おおむね高等学校 2 年までの学習内容だが、特に、指数、対数の基本的性質と総和記号  $\sum$  の使い方、多項式の微積分である。高 3 以降の知識、特に微積分の知識はほぼ仮定していないが、一部で必要に見える場合もある。しかし、その場合でも、積分と面積の対応が理解できていれば大きな問題はなく、仮にそのような箇所を飛ばしても本質の理解は損なわれない作りになっている。一部で使われている線形代数の知識は、初歩的な内容のみである。シミュレーションを用いた説明もあるので、数式が理解しづらいときには、シミュレーション結果を見てとりあえず先に進むのもよい勉強法だと思う。演習に使うデータは、Excel 形式 (xlsx) のものと CSV 形式のものがあり、R スクリプトとともに出版社のサイト ([http://www.rokakuho.co.jp/data/04\\_support.html](http://www.rokakuho.co.jp/data/04_support.html)) からダウンロードできる（動作確認は、R version 4.2.2(64bit), R Studio 2022.12.0 Build353(64bit)で行っている）。

本書を通して読者の統計学の学習が少しでも楽になれば、筆者にとって望外の喜びである。

2023 年 8 月

神永正博・木下 勉