

はじめに

本書は、主として一変量の統計学と R の入門書である*¹。数学は好きだが R に馴染みがない人を念頭に置いて書いた。ソフトウェアの実用書と数理統計学の専門書は数多く出版されているが、前者には理論が、後者にはソフトが少ない。このような現状では、理屈は数理統計学の教科書で学び、実際に使う場合は実用書を読む、という以外の選択肢がないが、数理統計学の理論とソフトウェアの使い方を自力で対応付ける必要があるため非効率である。本書ではこの 2 つを同時に学習する。

本書のねらいは、R を統計学の理解の補助として使うことにより、読者が、統計学の理屈と R の使い方を同時にマスターすることである。現在はウェブを検索すれば様々な学習資料を見つけることができるが、本書の目標は、検索した情報を読み解く基礎学力を身に着けることである。標語的に言えば、「ググれば理解できる」段階が本書のゴールである。統計学は本質的に応用志向の学問であり、統計学の理論だけ追求しても得るものが少ない。かといってノウハウだけ学ぶのでは「こういうときはこうする」ということばかり覚えることになり、なぜそうするのがわからないので、使っていてもどこか頼りないと思う。本書を読めば、R に用意されている統計関数がどのような数学的原理に基づいて計算を行うかがわかるはずである。

R は、オークランド大学の Ross Ihaka と Robert Clifford Gentleman (2 人の R) により開発されたオープンソースの統計処理環境であり、R 言語と呼ばれることもある。比較的簡単な操作だけで高度な統計処理を行うことができるのが R のよいところである。R に煩雑な処理を任せ、一種の統計電卓、シミュレーションソフトとして使うことで、統計概念を効率よく修得することができる。R は、フリーで提供され、UNIX, Mac OS, Windows 上で動作する。洗練された数多くの機能を持ち、多くのユーザーに愛されている R は、今後数十年にわたって統計言語のスタンダードとして使われていくことになるだろう。本書で統計学と共に R を勉強しておくことは決して無駄にはならないと確信している。

本書は、東北学院大学工学部における講義用の教材に大幅に加筆したものである。理論的説明の分量が多いので講義で全て説明するのは難しいかもしれない。筆者は、講義で理論の細かいところは解説していないが、積率母関数の扱いから中心極限定理の証明に至る流れは省かずに説明している。予備知識としては、大学初年次の線形代数学および一変数・多変数の微積分を仮定した。本書は、自習書としても、教科書としても使うことができる。各章ごとにいくつかの練習問題があり、読者はこれらの問題を解きながら、R の操作を確認するとともに本文の内容の理解を深めることができる。演習用データはエクセル形式のファイルまたは CSV 形式のファイルであり、内田老鶴圃の書籍サポートページ*²からダウンロードできる。また、章末問題には、本文では紹介しきれなかった統計概念も盛り込んである。解答は一部を除いて完全なものを載せておいた。

最終章は、「べき分布」と `poweRlaw` パッケージについての説明である。べき分布は通常の統計学の教科書ではほとんど省略されてきた内容であり、耳慣れない話だと思うが、ぜひとも読んでいただきたい。正規分布万能の世界観を根底から揺るがす大きな問題だからである。

なお、本書は Windows マシンを使って操作することを想定して書かれているため、Mac OS や

*¹ 本書の統編「R で学ぶ確率統計学 (多変量統計編)」が予定されている。

*² http://www.rokakuho.co.jp/data/04_support.html

UNIX を用いる場合，操作上若干の違いがある．また，本書では 64 bit 版 R (ver3.5.1) と R Studio (ver1.1.456) を使い執筆を行った．この点にご容赦願いたい，違いはわずかなので，本書を読む際の大きな障害にはならないと信じている．重要なことは統計学の考え方を理解すること，統計ツールに慣れること，そして実際の仕事で使うことである．本書がその役に立てば望外の喜びである．

神永は統計学を学習する際に，多数の教科書，論文の他，多くのウェブサイトのお世話になった．とりわけ，青木繁伸先生のサイト*³，奥村晴彦先生のサイト*⁴，中澤港先生のサイト*⁵ (五十音順) から多くを学んだ．また，本書を出版する機会を与えて下さった内田老鶴園社長の内田学氏をはじめ，編集部の皆様には大変お世話になった．記して感謝したい．

本書の執筆は，神永の講義ノートを下木がチェックし，それを再び書き直す形で行われた．下木は主として数式と R スクリプトの動作のチェックを担当した．統計学的な問題については神永の責任でまとめた．もし，統計学的な誤りがあれば全面的に神永の責任である．

2019 年 2 月

著 者

*³ <http://aoki2.si.gunma-u.ac.jp/>

*⁴ <https://oku.edu.mie-u.ac.jp/~okumura/>

*⁵ <http://minato.sip21c.org/>