本書を読み始める皆さんへ

(1) 統計学の学習にRが使える時代

世の中には日常，たくさんのデータがあふれている．統計学の知識をもたずに，手元のデータから何らかの傾向を推測したり，グループごとの平均値間を比較することは，無謀極まりない．高校理科や大学の自然科学や社会科学の科目は，調査や実験のデザイン，その結果として得られるデータの分析など，すべて統計学に基づく．しかし，それをわかりやすく教える教科書や教材が，これまでは実に驚くほど少ない時代がずっと続いていた．なぜなら，統計学は実際のデータを解析してナンボの学問であるのに，その解析方法の原理を簡単に試し，理解する手段を提供してこなかったからである．関数電卓を使って検算したり，数値計算のコンピューター言語を使って面倒くさいプログラミングを実施し，結果を確認するしかなかった．

しかし今，統計学の教育に革命が起こっている．それは統計学とグラフィクス専門の“R”という名称の無料のソフトウェアが世界中で広く使われるようになったからである．Rは世界中の高度な統計の技術をもった専門家が，さまざまな関数やライブラリを開発して，一つのシステムとして，ウェブサイトCRANが世界中の国々で使えるようにしている．その恩恵を受けて，高校生や大学1年生の初心者であっても，あるいは統計学を専門とするわけではない異分野の利用者であっても，データ分析や統計解析にRを役立たせる時代となった．国内外でわかりやすいRのガイド本はいくつも出版されており，またインターネットでも，Rの関数の使い方などは懇切丁寧に解説されたサイトがいくつも見つかる．統計学を学ぶにはとてもよい時代である．

(2) Rを使うことと統計学を理解してRを使えることとは別の問題

多様で莫大なデータが利用可能な現代においては，データから有用な情報を得るための方法論として，データ解析やモデリングの手法はますます重要になってきている．そもそも統計学は，データをもとに現象を記述し，現象のモデルを構築し，知識を獲得するための方法論である．日本学術会議 数理科学委員会 統計学分野の参照基準検討分科会では，統計学の学習におい

ては，統計学の手法の知識と対象の理解とを両輪として進めることが重要であると主張している（2015 年 12 月 17 日分科会報告）．R が簡単に使える便利な時代とはいえ，対象への理解や検定法の背景にある理論，モデル構築と選択の理解も同時に学ぶことは，とても重要なことだ．

R が使えるようになっても，統計学の理論は初心者が独学で学ぶにはハードルが高い．たとえば，初歩の事例として出てくるベルヌーイ試行（コイントスの表と裏の出方，袋から取出す赤玉と白玉の出方）は確率が絡むので，よほど確率論が得意な人でない限りは，初心者はひと苦労するだろう．ましてや，確率分布としての正規分布などは見るからに複雑な数式だし，確率密度関数や累積分布関数などは体系立てて教わらないと，独学では修得の効率が非常に悪く，統計学嫌いをたくさん生み出す結果となるだろう．便利な R を使うことと，統計学を理解して R を使えるように修得することとは，おそらく別の次元の問題なのだと思われる．

しかし，この両方を初学者向けにわかりやすく解説した教科書があれば，統計学の初歩を体系立てて修得することができ，最初の基礎を固めれば，後は，使う用途によって，いろいろな関数をインターネットで探してきたり，より高度な統計学へと歩みを進めることができる．本書“R で学ぶ統計学入門”は，高校でのクラブ活動やスーパーサイエンスハイスクール（SSH）で学会のポスター発表する高校生とそれを指導する理科教師，大学教養課程や専門学科で統計学を学ぶ学生を対象としている．統計学専門の研究者を養成することは考えていないので，そういう読者には，巻末の少しレベルの高い文献を勧める．

(3) 日本の高校生や大学生は統計学を体系だって学ぶ機会が乏しい

筆者は，東京大学教養学部の広域科学科・学際科学科で 2〜3 年生を対象に，“統計学”の講義と R を利用した“統計学実習”を 15 年近く教えてきた．第二著者の阿部真人はその受講生であり，嶋田の研究室に進んだ後は，動物の移動や個体間の相互作用と因果性に関わるデータ解析で博士の学位を取得した．現在は国立情報学研究所のポスドク（博士研究員）である．嶋田と阿部は，国際生物オリンピックの国内第二次選抜を合格した 15 名を対象に，5

年ほど前から生物統計を特訓してきた．その中で痛感したことがある．それは，日本の高校では，統計学を学ぶチャンスが“皆無に等しい”ということだ．

中等教育での現学習指導要領では，統計の単元は，高校の数学Ⅰで“データの分布”として，散らばり（ばらつき）や散布図，箱ひげ図，2変量の相関などの概念を学ぶ．また，数学Bでは，“簡単な確率分布”として二項分布と正規分布を学び，そこから“母集団と標本”の関係を学ぶ．最後は，標本の標準偏差などを使って，母集団のばらつきを推定するところで終わる．

しかし，これは本書を手に取ってもらえればすぐにわかるだろうが，この内容は第2章“母集団と標本”までの抜粋でしかない．これでは，統計学という城の手前に立って，おずおずと高い城門を仰ぎ見ている状態でしかない．本書のレベル程度を体系立てて学ばないと，“統計を理解してRを使える”という段階に達しているとはいえないだろう．

大学のカリキュラムでも同様である．理系の学生は，本来はデータ分析を経験するはずだから，全員必修にしてもしかるべきだと思われるが，一般教養の課程では少数の学生しか履修しない．入学したばかりの1年生にはデータ分析の重要性を理解できないだろう．しかも，多くは数理統計学専門の教員が教えるので，確率分布や確率変数などの説明に多くの時間を費やし，証明問題などが多い．その結果，統計学を受講しても，分散分析の計算法（“平均値間の差を調べるのに，なぜ分散を使うのか？”という根本的な理解）すら頭に定着していない学生が多い．専門学科での学生実験で初めてデータ分析の統計学を学ぶが，それとて，その実験テーマに特有の統計分析しか学ばない．――体系立てたデータ分析の統計学入門をどこで学べばよいのだろう？

(4) 本書の特徴：Rを使ってデータ分析の入門を体系立て学ぶ

本書の特徴は，一言でいえば，“Rを使ってデータ分析の入門を体系立てて学ぶ点”である．本書は数理統計学の専門家を養成するための教科書ではない．あくまでも，統計学の技法を使ってデータ分析する高校生や高校の理科教師，大学でのエンドユーザを育てるのが目的である．

よって，国内で広く高校や大学の授業で使ってもらえる教科書として，確率分布や確率変数を説明する数式はできるだけ少なく抑え，数学的な証明は脚注で引用文献を紹介するにとどめた．最初に記述統計学，つまり標本のデータの平均値とばらつき（分散）から説き明かして，徐々に検定と推定に関わる理論を学び，いくつかの検定法（*t* 検定，分散分析，回帰分析と相関）を修得する．そして，R の普及とともに最近，頻繁に使われるようになった一般化線形モデル（GLM）を学び，最後にノンパラメトリック法のいくつかを学ぶ．本書が示すこれらの内容を体系立てて学ぶことで，統計学の基本は必ず身に付くはずである．さらに，本書を卒業した読者が次に進むべき教材として，少し上のレベルの文献やサイトを巻末にあげておいた．

本書は，高校や大学での統計学初心者からデータ分析や論文を執筆する大学院生などの中級レベル利用者まで幅広い読者層を対象にしている．そこで，中級の読者層向けの章や節には目次に ★ 印をつけて区別した．星なしは初心者レベル，★ は学部 4 年生～修士レベル，★★ は博士院生～研究者レベル（統計学以外の分野）に相当する．★ 印をつけた章や節を理解した読者は，レベルとしては日本統計学会公式認定 統計検定 2 級の内容も理解できるだろう．

また，R で演習する本書の事例はすべて小標本（標本サイズが数十～百程度）を対象としている．その理由は，一つは高校での科学部の実験や調査，大学での学生実験や卒業論文，修士課程で学会発表を初めて試みるレベルだと，標本サイズは小標本であることが多い点，もう一つは分野としては生態学，分類学，生理学，進化学，動物や人の行動学，認知心理学，薬学などの研究分野では，小標本が多い点である．もちろん，R ではもう少し大きなサイズの組込みデータセットも標準で用意されており，R コンソールで **`data()`** の一覧を出して，適当なデータセットの名称を入力すれば自由に中身が出力されて使える．だが，基本的には本書で培った小標本の扱いでそのまま対応できるので心配ない．

本書のもう一つの特徴は，ごく少数の模式図などを除いて，ほとんどの図版を R で描画している点である．これは阿部がすべて清書図版の描画を担当した（阿部は R スクリプトの確認も担当）．R はグラフの描画ソフトとし

ても優れており，投稿論文用の清書図版もすべてRで描ける．各章の図版を作成するためのスクリプトを東京化学同人ホームページ（http://www.tkd-pbl.com/）上に載せておいた．コピーペーストして自由に使えるので，ぜひ有効活用してもらいたい．もちろん，描画関数の各引数を適当にはしょって，それらをデフォルトにすると，図の見栄えは劣るが，それでも似たような図は得られる．両者を比較することで，各引数の意味が試行錯誤でわかるだろう．

なお，Rのガイド本としては，“The R Tips：データ解析環境Rの基本技・グラフィクス活用法（第2版）”（舟尾暢男 著，オーム社）を薦める．ただし，これは統計学入門の本ではないので注意してほしい．

宮下 直氏（東京大学大学院農学生命科学研究科）には第10章と第11章の原稿を，また粕谷英一氏（九州大学大学院理学研究院）には第10章～第13章の原稿を下読みしていただき，多くの有効なコメントを頂戴した．深く感謝している．特に，粕谷氏からはGLMやノンパラメトリック法について，私たちの理解の至らない内容や文言，事例について詳細な指摘をいただき，大きく改善に役立った．もちろん，誤りの箇所や不適切な文言がまだ残っていれば，それは嶋田と阿部の責任である．

最後に，嶋田はけっこうな遅筆で，申し訳ないことに，東京化学同人の住田六連編集部長が毎月原稿を受取りついでに発破をかけに来られた．これがペースメーカーとなってようやく脱稿し，たいへん感謝している．また，編集部の井野未央子さんには，これまでの東京化学同人から刊行したいくつかの教科書同様に，厳しくも慈悲深い見守りと本編集のすべてのプロセスでお世話になった．ここに厚く御礼申し上げたい．

2017年1月

嶋 田 正 和