

訳者あとがき

本書は、Michael A. Bailey (2020) *Real Stats: Using Econometrics for Political Science and Public Policy*, 2nd edition (Oxford University Press) の全訳である¹。著者であるマイケル・ベイリーは、アメリカ政治を中心に関心に世論調査方法論や国際関係に至るまで幅広い研究業績がある、ジョージタウン大学所属の著名な政治学者である。

本書の内容は、一般的な統計学の教科書とは一線を画している。大きくいうと、本書が目指すのは「ときめかない」統計学から脱却し、社会科学を学ぶ人が日々の研究で実際に「使える」統計知識を教えることである。政治学の教育方法論を議論する学術誌に2019年に発表された論文で、著者自身がその狙いを具体的に説明している²。「統計学の教え方——恐ろしくて、退屈で、役に立たないものから、そう、何かもっと良いものへ」と題する本論文によれば、これから（社会科学の学生・研究者のための）統計学教育に必要なことは、4つにまとめられる。

ときめく内容だけを教えること

1つ目は、「お片づけ！」である。著者ベイリーは、片づけコンサルタントの近藤麻理恵さんの言葉を借りながら、「ときめく」内容だけを残すことが重要だと説く（そう、「こんまりメソッド」が政治学の学術誌で引用されているのである！）。すなわち、研究で実際に使うことがイメージできるような、ワクワクする内容だけを教えよう、ということである。

1 姉妹書として、同じ著者による Michael A. Bailey (2019) *Real Econometrics: The Right Tools to Answer Important Questions*, 2nd edition (Oxford University Press, 2nd Edition) がある。Real Stats が政治や公共政策に関心がある人向けに書かれているのに対して、Real Econometrics は、経済や経済政策をメインテーマとしたテキストになっている。

2 Bailey, Michael A. (2019). “Teaching Statistics: Going from Scary, Boring, and Useless to, Well, Something Better,” *PS: Political Science & Politics* 52 (2): 367-370.

教科書の冒頭で、さまざまな確率分布の名前を列挙して、細々とその性質について説明することを考えてみよう。そういう辞書的な知識は、実際には社会科学の分析にすぐには使えない。そんな教科書を読み始めたら、役に立たないものであふれた汚部屋に踏み込んだ気分になるだろう。省略変数をモデルに含めたら、係数値が0.001変わった、という例もいらない。ときめかないとある。 R^2 がモデルの当てはまりの良さを示すことを長々と説明した後に、 R^2 の高さは、分析がうまくいっているかどうかとほぼ無関係である（洗練された実験の分析モデルの R^2 は得てして低いが、さまざまな問題を抱えるラグ従属変数モデルの R^2 は必然的に高くなる）と指摘するのはどうだろうか。ならば、どうして延々と R^2 の説明を読まされなくてはいけなかったのか。これも、全然ときめかない。

本書は、そんなときめかない内容を「片づけ」て、社会科学を学ぶ人たちが日々の研究で実際に「使える」統計学の知識を厳選して扱う。よって、本書の内容は、統計分析の数学的・技術的な背景を正確に把握したい人には、詳細が不足していたり、表現が厳密には正しくないと感じたりする部分があるかもしれない。しかし、社会科学の研究にどんな分析手法を使えばいいのか迷っている人にとっては、統計分析を研究で適切かつ有意義に「使う」ために必要十分な知識だけが詰まっている本書は、統計を研究で使うハードルを確実に低くしてくれるだろう。

| リアルなデータに早く触れること

著者が重視している2つ目の点は、「リアルなデータに、できるだけ早く触れること」である。リアルな例を使って統計学を学ぶことは、実際の研究で統計がどのように使えるかをイメージする近道である。統計学におけるさまざまな概念を、難しい専門用語を使うことなく現実社会の例を通して表現できれば、（数学が嫌いな人を含めて）幅広い人が統計的な分析思考を身につけるきっかけになるだろう。

本書の本文にある図表のほとんどは、現実に存在する政治や経済、社会に関するデータ（もしくは現実社会を模して直感的なラベル付けがしてある仮想データ）に基づいて作成されている。また、演習問題のセクションで扱う問題には、さまざまな政治・社会問題に関わる実際のデータが付属している。したがって、本書は、研究関心と直接関連するような現実のデータを使って、統計学で用いられる概念の本質をつかむ機会を、読者に提供している。

| 理論は必要な時にだけ、ジャスト・イン・タイムで説明すること

3つ目の点は、統計の概念・理論の導入に関わるものである。ここで著者は「理論をジャスト・イン・タイムで教える」べきであると主張する。すなわち、統計理論は、社会科学の研究が実際に直面している課題と紐づけて、必要な時に、必要な分だけ説明されるべきだということである。内生性の概念を理解し、その課題に実験や OLS、操作変数、回帰不連続デザインなどの手法がどう対処しているかを理論的に学ぶことは、データ分析を適切に行うために必要だ。しかし、条件付き確率分布の名称と定義をひとつずつ押さえることや、中心極限定理にどんなバリエーションがあるかを知ることは、現実の課題を考えるうえでは、ちょっとマニアックすぎる知識といえるだろう。

統計理論について一通りカバーしてから実践例に進む、というのは古典的な統計学の教科書にありがちな構成である。統計学という学問自体に関心がある、新しい統計手法を開発したいと思っている人には、理論を網羅的に教えられることは有意義かもしれない。しかし、とくに社会科学の学生・研究者を対象とする場合、理論だけを先に集中して教える構成は、学習者の統計に対する意欲を削ぎかねないし、ときめかない内容を「片づける」ことを難しくすると著者は指摘する。本書は、そのような著者の主張に沿って、統計理論は社会科学のデータ分析において生じる具体的な課題に応答する形で、ジャスト・イン・タイムに導入される構成になっている。

| 分析の主役となる OLS にさっさと触れること

著者が統計学教育に必要だと考える第4の点は、「OLS にさっさと触れる」ことである。本書はその通り、第3章にはすでに OLS（最小二乗法）が紹介され、第4章以降はほぼすべて OLS の枠組みをベースに展開していく。

一般的な統計分析入門の教科書では、まずは変数の種類や頻度分布表、中央値・平均値などの記述統計の示し方をカバーし、その後に平均値の比較などの単純な二項分析について丁寧に説明をした後に、OLS は巻末で「ラスボス」として登場することが多い。対照的に、本書は記述統計に関する説明は第2章の数ページでサッと終わらせてしまい、平均値の比較については OLS の枠組みの中で説明が行われている。

著者は、このような構成をとる理由として、記述統計の知識は得てして「散らかりやすい」こと、大学より前の数学教育ですでにカバーされている内容が多いこと、そして、因果関係や内生性の問題に対処するほとんどの手法は

OLS の延長線上にあることを挙げる。社会科学における問い合わせに対して意味のある統計分析をするには、結局 OLS を使う必要があるのだから、早めに学んだほうがよい、ということである。

記述統計に関する知識はもちろん重要だ。しかし、ほとんどの場合、それだけでは「 X は Y を引き起こしているか」という、本当に検証したい問い合わせには答えられない。OLS という、因果関係に関する問い合わせを検証する際に主役となる手法をさっさと紹介することで、社会科学の学生・研究者に、統計学が「ときめく」学問だと感じてほしい、というのが本書における最後の狙いである。

*

上で紹介した通り、本書には、政治学・公共政策学を含む社会科学全般に関心がある学生・研究者が、統計分析を研究手法として魅力的に感じられるようになる工夫が随所にちりばめられている。

訳者として、最後に語っておきたいのは、本書全体を通して登場する、カジュアルな例示と表現についてである。たとえば、本書で最初に紹介されるデータ事例は、ザ・シンプソンズ（アメリカの超有名なテレビアニメ）に登場するキャラクターを対象とするドーナツの消費量と体重の関係であるし、第 12 章のケーススタディでは、犬を飼っているかどうかとオバマ大統領支持の関係が分析される。それ以外にも「一から再現できない分析は、役には立たない。いっそのこと、忘れて寝てしまったほうがマシかもしない」（第 2 章、2.2 節）、「信頼区間は真値を含むか、含まないかのどちらかであるので（スターウォーズのヨーダが言うように、「とりあえず試してみるなどない」のである）……」（第 6 章、注 5）など、随所でくだけた表現が登場する。

一瞬、著者は悪ふざけをしているのではないかと思ってしまう（し、訳者としては翻訳に苦労した）が、この教科書が書かれた意図を知れば、これらも、統計学へのハードルを下げるための著者の戦略だということがわかる。これらの例示や表現は、実は重要な理論や概念を語るときに登場しており、（難解な用語を使わなくても）内容が頭に残るように工夫されているのである。読者にとっては、いつの間にか統計的手法・理論の要点が頭に焼きついている、という実にありがたい構成になっているので、なんだかバカみたいな例だなと思うても、クスっと笑いながら、飛ばさずにしっかり読み進めてもらえればと思う。

なお、各章末のプログラミング・コーナーにある分析コードは、使いやすい

ように日本語化した.do ファイルと.R ファイルを作成してgithub.com/gentok/JpnRealStats にアップロードしているので、分析の実践にぜひ活用してほしい。

本書の翻訳プロジェクトが始動したのは、2020 年の末にさかのばる。それからすでに 4 年半以上の月日が経過し、著者のベイリー先生には「もうすぐ終わります」と終わる終わる詐欺（?）を重ねて、長らくお待たせしてしまった。また、担当編集者の勁草書房の上原正信さんにも、辛抱を重ねてお待ちいただいたことに、深くお詫びと感謝を申し上げる。関係各方面に多大なご心労をおかけしたが、ついに出版にこぎつけることができ、訳者一同ほっとしているところである。また、2023 年にはベイリー先生を早稲田大学にお招きして研究発表会を開催した。その際にご尽力くださった日野愛郎教授をはじめ、早稲田大学関係者の方々にも、この場を借りて感謝を申し上げたい。

本書は、統計分析に初めて触れる人のみならず、統計学に今まで関心が持てなかつたり、やりかけたものの挫折したりしたことがある人にも手にとってもらいたい本である。本書を通じて、統計学を用いたデータ分析のおもしろさや有用性に目覚めて、統計分析を自分の研究ツールボックスに追加してくれる人が増えることを、訳者一同、心から願ってやまない。

2025 年 7 月吉日
訳者代表 加藤 言人