

Vol. 22 No. 6
December 2010
日本知能情報ファジィ学会誌

知能と情報

Journal of Japan Society for Fuzzy Theory
and Intelligent Informatics



日本知能情報ファジィ学会

Japan Society for Fuzzy Theory and Intelligent Informatics

平成22年12月15日発行第22巻第6号（通巻123号）年6回発行（2, 4, 6, 8, 10, 12月）
平成 5 年12月20日 学術刊行物認可 ISSN 1347-7986

書評

中村 永友（著）、金 明哲（編）

「多次元データ解析法 (Rで学ぶデータサイエンス2)」

- ・出版社名：共立出版
- ・発行日：2009年8月25日
- ・B5判、264頁 3,500円
- ・ISBN：978-4-320-01922-5

清水 信夫(評)
統計数理研究所データ科学研究系

昨今の情報化社会の進展は、データの形態や収集方法の複雑化や多様化、およびデータの平均サイズの増大をもたらしている。その一方で、計算機は高速化・高性能化・低価格化が進み、また計算機上で統計計算やシミュレーションを行うための専用ソフトウェアも多数開発されている。それらの結果として、データを扱うための科学として統計学の役割の重要性が広く認知されてきている。

これらのデータを扱う上では、収集や加工などデータの流れの上流から、データの解析およびマイニングなど下流に至るまでの幅広い理解が必要であり、それらを包含する概念として「データサイエンス」という用語がよく用いられているが、その中核を成すのはやはり解析であろう。そのために開発されている統計解析ソフトウェアは多種多様であり、それらの多くは年々高速化、高性能化してはいるものの、入手が容易で手軽に利用でき、かつ大量のデータの解析にも適しているという両方の条件を同時に満たすものは極めて少ないのが問題であった。

そのような背景の下で、1990年代から開発が始められた統計的データ解析ツールであるRはオープンソース方式で現在も開発が続けられており、大量かつ様々なタイプのデータの解析にも対応できる統計計算環境や専用パッケージが多く用意されている。またRはフリーのソフトウェアとして、インターネットに接続された環境であれば誰でもいつでも何処からでも自由にダウンロードが可能であり、インストールおよび使用も容易である。

Rの使用による統計学やデータサイエンス関連の本はここ数年で急速に増えているが、多くはRによる初級統計学やRのマニュアルという形態であり、データサイエンスに関する数理的・統計的な理論と実践を両立させたものは少なかった。そのような状況の中で、本書はその両者を様々な統計的データ解析手法において本格的に両立させている。

本書は11の章から構成されている。まず第1章「統計学の基礎的事項」では多次元データの解析法の分類およびデータの各種記述方法、さらに統計学でよく利用される各種手法についての紹介を行っている。第2章「Rの基礎的コマンド」では基本的なデータの取り扱い方や基礎的な統計量の計算に必要なRのコマンドについて簡潔に紹介している。

第3章以降の各章では、各種統計解析手法におけるRの利用方法について詳述している。各章のタイトルは、第3章「線形解析モデル」、第4章「判別分析法」、第5章「ロジスティック回帰モデル」、第6章「主成分分析法」、第7章「対応分析法」、第8章「因子分析法」、第9章「正準相関分析法」、第10章「多次元尺度法」、第11章「クラスター分析法」である。これらの各章ではいずれも各種手法の統計学的な理論を系統的に説明した上で、実データを用いた簡単な解析例についてもわかりやすく丁寧に記述している。またRのコマンドを使用した場合の解析方法と結果についてもソースコードをそのまま示しており、初心者にも使いやすい形でまとめている。各章の最後には、更に深く学びたい人向けの参考文献を紹介している。

なお、多次元データ解析を行う上でよく用いられる行列やベクトルによる各種演算についても「行列の基本演算」に付録としてまとめており、巻末の索引も統計用語およびRのコマンドに関して充実している。

本書を含む「Rで学ぶデータサイエンス」シリーズの各巻は、それぞれが対象とする分野において使用されている数理的理論とRによる実践を共に理解できるよう編集されているが、その中でも本書はデータサイエンスの基礎を扱っている巻として、様々な分野における解析手法の理解を助ける上でも有用と言える。