

本シリーズの編集にあたって

R言語が急速に普及しているその主な理由としては、オープンソースであることや膨大なパッケージが自由に利用できること、世界最高峰の統計専門家やデータサイエンティストなどがもっぱらRを使うようになってきていることなどがあげられる。RはSPSSやSASを王座から引きずり下ろす勢いで、統計分析のスタンダードなツールになりつつあるといわれている。このようなニーズを見据えて2009年から「Rで学ぶデータサイエンス」全20巻の刊行を開始した。その後、朝倉書店からはシリーズ「統計科学のプラクティス」が刊行された。このようなシリーズを含めRに関する和書・訳書は100冊を超えている。多くの執筆者や出版業界の努力により、Rが統計科学とデータサイエンスの教育、研究および応用に浸透しつつある。

昨今、ビッグデータという言葉をよく耳にするようになった。また、西内啓著「統計学が最強の学問である」(ダイヤモンド社)をはじめ統計科学やデータサイエンスに関する啓蒙的な書物のおかげで、統計学とデータサイエンスが今までにないちょっとしたブームになっている。大量のデータをすばやく解析するためには、時代に適したツールが必要である。

近年、ビッグデータ処理に関するシステムが続々公開されているが、その多くのシステムの内部の統計的データ処理や機械学習はRを用いている。たとえば、データベースシステムの大手メーカーオラクルはOracle R Enterpriseを公開している。Oracle R EnterpriseはRを使ってオラクルのデータベース内のビッグデータをRの機能を用いて統計処理を行う環境である。また、ヨーロッパ最大級のデータベース会社SAPは、インメモリーデータベース製品「SAP HANA」の新バージョンSP3にRICEを用いてビッグデータ解析の環境を提供し、自社のデータウェアハウス構築ソフトSAP BWにもRICEを利用可能にすることを目指している。RICEはHANAが提供するインメモリーDB技術を生かして、ビッグデータを扱うアプリケーションをRで開発・実行できるようにしている。なお、PCサーバを多数つなげた大規模のデータを分散処理するHadoopでもデータの統計処理はRに託している。したがって、Rを利用しているユーザは、なじみのインタフェースを使い、大規模のデータの統計解析を行うことができる。

ii 本シリーズの編集にあたって

このようなことから、Rはビッグデータ時代にますます利用が増えると予測される。Rは奥が深く、さまざまな分野での応用が予想される。R言語の構造、Rを用いた情報の発信、Rを用いたツールの作成のようなRに関する内容やデータ分析プロセスから、金融データ解析やトランスクリプトーム解析のようなデータ解析に関する広範囲の内容を柔軟に預かるため、シリーズ「Rで学ぶデータサイエンス」の姉妹編として、シリーズ「Useful R」を刊行する。

本シリーズの刊行が統計科学やデータサイエンスにかかわる諸分野の振興に少しでも寄与できれば幸いである。

編 者 金 明 哲