

# S-PLUS の利用環境と最新の S 言語

山本義郎 (多摩大学経営情報学部)

## 1 . S-PLUS の利用環境

PC とインターネットの普及にともない統計解析パッケージだけでなく、多くのシステムにおいて最適な環境へのシフトが行われている。統計解析環境 S-PLUS は、統計解析のための高級言語であり、データを眺めながら対話的にプログラミングをおこなうための環境である。その性質を有効に利用するため、UNIX 版と Windows 版では提供されているユーザインタフェースが異なる。さらに、データマイニング領域などでは、ダイナミックにデータベースからのデータを扱う必要もあり、クライアントサーバシステムでの利用の要求があり、それに対応するものとして StatServer が提供されている。

このように、利用環境は異なるものの、対話的にプログラミングを行ってデータ解析を行うという、本来の S-PLUS の特徴を有効に利用できるように構成された、オブジェクト指向に基づく「すべてのものはオブジェクトである」というスタイルが、S における処理を更に効率的なものとしている。また、S は統計計算を行うための関数群を有するため、多くの研究者の新しい研究を支援するにとどまらず、その成果を容易に再利用できる環境を提供している。そのため、多くの研究者が、その成果を公開しているだけでなく、そのような成果が

StatLib <http://lib.stat.cmu.edu/S/>

などにまとめられ公開されており、それらの産物を利用することにより最新の研究に生かされている。そのような成果の多くは現在の S-PLUS に取り込まれており、このようなサイクルが多くの実務家や研究者が好んで S-PLUS を利用している主な理由ではないだろうか。

S-PLUS の先進的な機能を利用するための参考書である「Modern Applied Statistics with S-PLUS (Third edition)」の訳本(「S-PLUS による統計解析」)も出版されたが、原著の Web サイト

<http://www.stats.ox.ac.uk/pub/MASS3/>

には、本の中で利用しているプログラムコードやデータが提供されているだけでなく、バグ情報やプログラムの更新や最新情報が提供され、そこで提供されている差分により2版、3版がこの種のものとしてはかなり早いペースで提供されている。このようなサイクルが実現できていることも、S が現在のインターネットを活用した研究・実践のスタイルに適応していることを示している。後述する新しいS言語の仕様においても、このようなコントリビュータが支えている S というシステムの利便性を向上するための機能が取り込まれている。

## 2 . R について

R は GNU S と呼ばれ、S と同等の環境で統計解析を行うために作られたフリーソフトである。そのホームページ

<http://stat.auckland.ac.nz/r/r.html>

には、S と R の違いについて書かれているが、アーキテクチャーや利用可能な関数には違いはあるものの、ほぼ S の特徴を取り込んでいる。これは、S 言語が 1998 年に ACM のソフトウェア賞を受賞していることからわかるように、言語仕様が非常に高い S をもとにすることの有効性を考慮し、更にコントリビュータを基点とするプログラムサイクルを見越したためであると考えられる。実際、Version が1になるまではかなりゆるい進展であったが、その後は急速にバージョンアップし、利便性が高まってきている。そのプログラムおよびドキュメントや情報は

The R Project for Statistical Computing <http://www.r-project.org/>

で管理されている。

この R の発展は、S にとって良い関係となるであろう。というのも、R を含んだ S ユーザ層が広がることは、R が S での成果を有効に取り込んで成功したように、R を利用した研究成果も S に取り込みやすいため、想像に難くない。利用者環境も、大学や職場では S-PLUS、家では R という選択肢を選ぶことが可能であることもフリーソフトとして独自に展開している R が存在することは、他のパッケージには無い利点である。

### 3 . 新しい S-PLUS (「Programming with Data」の紹介)

一般にはあまり知られていないことであるが、S 言語にはバージョンがあり、ベル研究所での S のバージョンを把握するには、S 言語の解説書と連動していると考えるとわかりやすい。つまり、「S システム」(共立出版)は S 言語 V1 の解説書であり、青本と呼ばれる「S 言語 I, II」(共立出版)は V2 の解説書であり、「S と統計モデル」(共立出版)では V3 が対象となっている。そして現在の V4 に関する解説書がこのたび訳本として日本にも紹介されることとなった「Programming with Data」、邦題「データによるプログラミング」(森北出版)である。S-PLUS のバージョン番号と S 言語のバージョン番号は、これらと連動しているわけではなく、S-PLUS V5 以降(現在 V6)が S 言語 V4 の仕様である。S-PLUS V5 は Windows では対応しておらず、S-PLUS for Windows version2000 は S-PLUS V5 と同時期に世に出たが言語仕様は S 言語 V3 であり、旧来のシステムに GUI を充実させたという印象であったが、S-PLUS V6 では UNIX, Windows ともに同じバージョン番号、同じ S 言語仕様へと正常化された。そのため、「データによるプログラミング」は S-PLUS を有効に利用するためのバイブルとなるでしょう。

「データによるプログラミング」という書名に代表されるように、ソフトウェアにデータを与えて解析するという概念ではなく、データがあり、それをオブジェクトとして扱い、データの表示・要約・分析をオブジェクトに対するメソッドとして実現するというオブジェクト指向の概念が全面的に取り入れられている。この概念のもとでデータ解析・分析を行うというパラダイムは、すべてがオブジェクトであるというオブジェクト指向環境で実現されている。

この新しい S 言語は、S 言語 V3 から取り入れられたオブジェクト化やモデル化の特徴をより明確な形で再構成した全く新しい S に衣替えした。主な特徴としては、オブジェクト指向の強化とそれに伴う関数・クラス・メソッドの取り扱い、プロジェクトごとに作業しやすい環境を与えるためのチャプタの利用、ドキュメントとドキュメントオブジェクト、旧バージョンからの移行のためのツールなどがあげられ、それらについての十分な解説が「Programming with Data」から得ることができる。

### 参考文献

- [1] John M. Chambers (1998) *Programming with Data*. (邦訳 垂水共之・越智義道・水田正弘・森 裕一・山本義郎 訳「データによるプログラミング - データ解析言語 S における新しいプログラミング -」, 森北出版, 近刊)
- [2] W.N. Venables, B.D. Ripely (2000) *Modern Applied Statistics with S-PLUS 3rd ed.*. (伊藤幹夫, 大津泰介, 戸瀬信之, 中東雅樹 訳「S-PLUS による統計解析」, シュプリンガー・フェアラーク東京株式会社, 2001 )
- [3] R.A. Beaker J.M. Chambers, and A. R. Wilks. *The New S Language*. Chapman and Hall, 1988. (邦訳: 渋谷政昭, 柴田里程 訳, 「S 言語 データ解析とグラフィックスのためのプログラミング環境 - 1, 2」, 共立出版, 1991)
- [4] John M. Chambers and Trevor Hastie (eds) (1993) *Statistical Models in S*. Chapman and Hall (邦訳: 柴田里程 訳, 「S と統計モデル」, 共立出版, 1994)