

MDSA：多次元統計解析システム

大 隅 昇

多次元統計解析システム (MDSA システム: Multivariate Descriptive Statistical Analysis System) は、日本科学技術研修所 (JUSE) と共同で開発を進めた統計ソフトウェアであり、現在多くのユーザにより利用されている。比較的順調にユーザに受け入れられた理由として、設計にあたり開発委員会方式を採用し、従来の類似システムの問題点や行き届かない部分に目を配り、できるだけ現場のデータ解析に適した利用環境の実現を目標としてきたことがある。設計の基本指針として、データ (とくにその種類、属性) と手法の関連性を意識したこと、手法相互間のデータ授受機能の重視、コンピュータ利用環境に応じたアプリケーション・サイズの調整機能 (add-on/remove module 方式による必要モジュールの移植やアプリケーション・サイズの調整機能) などがある。さらに、① できるだけ操作性のよいインターフェースを備えること、② ある程度、統計的データ解析についての知識を持つユーザを対象とすること、③ 比較的規模の大きいデータセットの処理が可能であること、④ 利用度の高い多変量解析手法に的を絞って、オプション機能の充実を重視したこと、⑤ 他のファイル形式のインポート機能があること、⑥ 手法の基礎数理、解析例、サンプルデータなどを活用ガイドブックにより提供することなどがある。システムは「基本解析編」と「応用解析編」とから構成される。基本解析編では、データセットの入出力処理、データ編集加工などの一般的な基本処理を初め「データの特徴抽出」の多数の独自の機能を含む。たとえば、データの属性に応じた各種多変量連関図、多重クロス表の生成、それらに関連する統計量の算出とグラフィカル表示等の多様な機能が含まれる。グラフィカル機能と算出統計量を同時に表示・観察できるメニューと画面設計にも特徴がある。応用解析編には、現在のところ主成分分析型手法 (主成分分析、対応分析)、予測型手法 (重回帰分析、数量化I類)、分類型手法 (k -平均法、ハイブリッド法) などが含まれる。主成分分析型手法には共通して、追加処理機能 (データの一時除去と再配置)、算出スコアの保存・再利用の機能などが用意されている。とくに対応分析には、通常の2元クロス表型データ (あるいはそれに同等のデータ表) を初め、多重クロス表 (パート表) の対応分析、それをういた大量データの処理が可能である。出発行列として多様なデータ表形式に対応できるように設計されており、この点で他の統計ソフトウェアに見られない特色がある。また、複数のデータ表の併合で得られるデータの解析にも対処できる。重回帰分析は質的・量的データの混在を許すもので、変数選択もこうしたデータに対して適用できる。また、回帰診断、残差診断の豊富なオプション機能 (トレランス、テコ比など)、AIC などを使った逐次的モデル選択機能なども含む。分類手法は、主成分分析型、重回帰分析型手法の事前・事後にリンクさせて利用する機会が多いので、これらの手法とのデータ相互授受や結果のグラフィカル表示に重点をおいた機能が多く含まれる。モニター制を活用しながら現場のデータアナリストが必要としているデータ解析支援ツールの機能とは何かを配慮しつつ、今後も他手法の移植や既存機能の改善強化を続ける。

HALBAU の開発について

高 木 廣 文

人間集団を対象として調査を行う場合、質問項目は極めて多数となるのが普通である。そのような調査データのコンピュータへの入力、保存、編集そしてデータ解析を効率よく行うシス

テムは必須である。HALBAU (はるぼう) は、そのような目的で開発されてきた。

HALBAU の開発に当たって、パソコンの機種は NEC-98 シリーズとし、システム開発用のプログラム言語は N88BASIC を用い、コンパイルして MS-DOS 上で使用するものとした。また、コンピュータやデータ解析について、初心者でも使用できるように、方法の選択等にはメニュー画面方式を採用した。指定などが分かりにくいと思われる箇所には、ヘルプ機能を設定することにした。解析方法は基本的なものから、高度な研究が行えるものまで、広範に取り入れることにした。これらの経緯については、柳井・高木 (1986)、高木 他 (1989) に簡単に述べられている。

今回のバージョン 4 での改訂でも、従来と同様に、HALBAU は 3 つのサブ・システムからなるような概念構成で構築された。すなわち、1) データの入力・編集などに関するシステム、2) 基礎的なデータ解析のためのシステム、3) 多変量解析や特殊な解析方法のためのシステム、の 3 つである。具体的には、1) には調査データを入力し、分析用のデータ・ファイルに保存するためのプログラムの他に、他の形式 (固定長データ、カンマ区切りデータ等) への変換、対数や合計の計算、HALBAU で作成したグラフの編集、などのためのプログラムが含まれる。2) には平均値や分散などの基本統計量の計算、クロス集計、相関係数の計算、各種の検定方法などの基本的な手法と散布図、箱ヒゲ図、生命表解析などが含まれる。3) には重回帰分析、判別分析、主成分分析、数量化理論 1, 2, 3 類、リスク評価のためのロジスティック・モデルや比例ハザードモデルによる解析、さらに質問紙などによる尺度構成のための信頼性係数などの算出も可能になった。また、マニュアルも整備され (高木 (1994a, 1994b)), これまで以上に各種の解析が容易に、かつ統合的に行えるシステムとなったものと考えられる。

参 考 文 献

- 柳井晴夫, 高木廣文 編著 (1986). 『多変量解析ハンドブック』, 現代数学社, 京都.
 高木廣文, 佐伯圭一郎, 中井里史 (1989). 『HALBAU によるデータ解析入門』, 現代数学社, 京都.
 高木廣文 (1994a). 『HALBAU-4 マニュアル, I. 基本編』, 現代数学社, 京都.
 高木廣文 (1994b). 『HALBAU-4 マニュアル, II. 基礎統計学』, 現代数学社, 京都.

サンプリング方式と尤度関数

(客員) 東京大学 教養学部 岸 野 洋 久

ほとんどの場合、データ収集に当たっては効率のよい調査を組んで、層別多段サンプリングを実施したりする。そこから得られたデータの解析に際しては不等確率抽出に起因するデータ間の重みの違い、単純ランダムに比して特性値の分散が大きくなる可能性がある、といったことを考慮に入れる必要がある。多次元解析や多変量解析において調査データ個々の重みをどのように考えるか、あるいは最尤法によるモデル分析において、上述のようなサンプリング方式によって抽出されたサンプルの中の相関構造をどのようにして解析に取り組むか、などという問題である。

この方面の関心もこのところ出始めており、たとえば最尤法に関連しては quasi likelihood によるアプローチが開発されてきている (McCullagh and Nelder (1989), Skinner et al. (1989)). ここでは母平均の推定 (Hansen and Hurwitz (1943), Horvitz and Thompson (1952)) を参考に、各データに抽出率に逆比例した重みをつけた対数尤度を考え、サンプリング方式を勘案した精度評価、および情報量規準を提案した。ただし重みづけに関しては、サンプリング