

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