

統合型統計解析プログラムパッケージ HALBAU の開発

高 木 廣 文*
柳 井 晴 夫**

I はじめに

最近 10 年間の技術的進歩により、現在のパーソナルコンピュータは、10 年前の中型電子計算機に匹敵するか、もしくは上回る性能を有するようになった。そのため、以前では考えられもしなかったが、高度な統計計算が手軽にパソコン上で行えるようになった。しかし、現実には実際のデータ解析に有用であるとは、必ずしも断言できないソフトが少なくないことも事実である。

統計解析プログラムパッケージ HALBAU (High quality Analysis Libraries for Business and Academic Users) (愛称「ハル坊」) は、現代数学社発行の「多変量解析ハンドブック¹⁾」に収録された BASIC (NEC-N88 日本語 BASIC (86) (MS-DOS 版)) による多変量解析のサブプログラムを主要計算ルーチンとして開発されたものである。ただし、たんに同書に掲載されているプログラムのコピーではなく、スクリーンエディタによるデータ入力・編集プログラム、および各分析結果の出力プログラムを加え、システムとしてより完全なものとなるように設計されたものである。また、実際のデータ解析では分析当初から多変量解析の手法を用いることは稀であり、始めは単純な分析により、データのチェックやその特性を探索するのが一般的手順と考えられるため、基礎統計学の基本的な手法も加えられている。さらに、分析のためにデータを適当に加工する必要があるれば、HALBAU 内にある変容ルーチンを用いて、比較的簡単に行えるように設計されている。

このように、HALBAU はデータ入力から分析法の指定、実際の分析、結果のプリント出力・ファイル出力に至るまでを比較的容易に行えるように作成された、統合型の統計学プログラムパッケージである。

現在、聖路加看護大学では大学院生用のパソコンに、インストールしてあり、比較的良好に用いられている。

今回は、HALBAU 開発の過程と留意点や、パッケージに含れる分析可能な手法について、また HALBAU の長所・短所、さらに今後の開発の展望などについて、以下に述べていくことにする。

II HALBAU 開発の過程と開発の留意事項

1983 年の夏頃、多変量解析関係の解説書の出版についての問合わせが現代数学社より柳井 (大学入試センター、当時千葉大学) にあった。これを受けて、市川雅教 (東京工業大学、当時東京大学大学院博士課程)、服部芳明 (日本原子力事業(株)、佐藤俊哉 (東京大学医学部、当時同大学院博士課程)、丸井英二 (東京大学医学部) および柳井、高木の 6 名の分担執筆により、出版することとなった。これには出版社側から条件があり、パソコン用に BASIC による各手法のプログラムを各章に掲載して欲しいというものであった。このため、本文の編集は主に柳井が、そしてプログラムの編集は高木が担当することになった。当初は、単にプログラムを掲載するだけのつもりであっが、入出力関係を掲載するとプログラムが膨大になるため、主要計算部分だけを本文中に示し、完全な形のプログラムは別に作成することになった。

BASIC 言語としては、16 ビットマシンでもっともシェアの高い NEC-PC9800 シリーズ用として、N88-日本語 BASIC (MS-DOS 版) を採用することにした。

各分担執筆者は、担当する各章の分析方法の記述とともに、その分析用プログラムを作成することになり、それが集まった段階で高木がパッケージとして統合することになった。

1984 年の夏頃には一部を除いて本文、プログラムともに集まったので、パッケージシステムの作成にかかった。各種の仕様に関しては、高木に一任し、完成したものを各方面で使用してもらい、プログラム中のバグの発見や、パッケージに関する要望などを収集することにした。実際のパッケージの仕様は一任されていたため、かなり個人的な好み優先される結果となっが、以下の点を特に考慮して作成することにした。

* 聖路加看護大学、統計学

** 大学入試センター、研究開発部進学適性部門

① データ入力・編集が容易であり、ファイルの保存が簡単にできること：

この条件は、パッケージシステムの必須条件と考えられた。このために、スクリーンエディタを作成した。また、HALBAU 内のどの分析にも共通して使用できるデータファイル構成にすることが要求された。

② 方法の選択が簡単にできること：

このためには、メニュー画面方式を採用した。

③ 分析のための変数指定が容易であること：

変数指定画面において、各変数名の一覧を表示し、変数の番号指定に誤りが少なくなるようにした。

④ 欠損値の処理ができること：

調査データでは、データに欠損のある場合が普通であり、その処理が自動的にできなければならない。ペア単位の除去とリスト単位の除去ができるようにプログラムを作成した（分析方法によってはリスト単位の除去のみのものもある）。

⑤ 各メニュー、指定画面、結果の表示は全て日本語であること：

データ解析の素人が、各種の指定や分析結果を間違えないようにするためには、英語などではなく、可能な限り日本語で表示することが絶対に必要であった。

⑥ エラー処理、ストップキーの処理などがある程度可能なこと：

分析中に誤ってキーを押したり、エラーが起こった場合などに、日本語による簡単なメッセージが画面に出力され、それに対応できるようにした。

⑦ 実際の調査データの分析に用いて実用的であること：

実際のデータ解析への応用が十分できなければ、開発の意味がないので、中途半端なものを作成しないことにした。

⑧ デフォルト処理がある程度可能なこと：

分析を素人が行う場合、分析手法によっては細かい指定が必要となるが、たんにリターンキーを押すだけで設定が行えるようにした。この点は、中途半端な方法の理解の仕方、解析を行うという可能性を高くするため、デフォルト処理をわざと行わない部分を残してある。

⑨ MS-DOS 上で作動すること：

従来 DISK-BASIC で作成されている統計パッケージが多かったが、ファイル操作や将来の開発に便利のように、MS-DOS 上で動くことは必須であると考えられた。また、インタープリタ上ではなく、コンパイルしてより高速に作動するように考慮した。

⑩ 分析できるケース数に制限がないこと：

分析できる変数の個数は、BASIC の配列の大きさの制限があるが、扱えるケース数に制限がないことを基

本とするように設計した（ただし、最尤解や順位を求めるような分析には制限がある）。

⑪ 主要計算はすべて倍精度で行うこと：

これは数値計算の基本である。

⑫ 分析手法を簡単にパッケージに組み込めること：

将来のユーザーの要望に応じて分析手法を HALBAU に付け加える場合、複雑な変更を必要としないこと（現在でも、分析手法の要望は受付けている。ただし、システムに付け加えるかどうかは個人的な見解に左右される）。

⑬ 「統計学パッケージ」というものの最低備えていなければならない操作性、機能などの水準を示すこと：

今後作成されるであろう各種の統計学パッケージの最低水準を示し、質の悪いパッケージでユーザーが多大な損失を被らないように、HALBAU を設計することにした。

以上のような点を考慮して、実際の HALBAU のシステム作成を行った。

1985 年の春頃には、多変量解析とデータ入力部分がほぼ完成した。これに従来から聖路加看護大学で使用していた高木による基礎統計解析プログラム（NEC PC-8801 用、日本語入出力）を大幅に手直しして組み込み、同年の夏頃には HALBAU の原型が完成した。これを関係者の知人等に配り、また東京大学医学部保健学科の統計情報処理実習などでも使用してもらい、プログラム中のバグや、操作上の問題点などについて意見をきき、改良に務めた。

1986 年の 4 月に「多変量解析ハンドブック」の発行とともに市販することになり、その後いくつかの修正を加え現在に至っている。

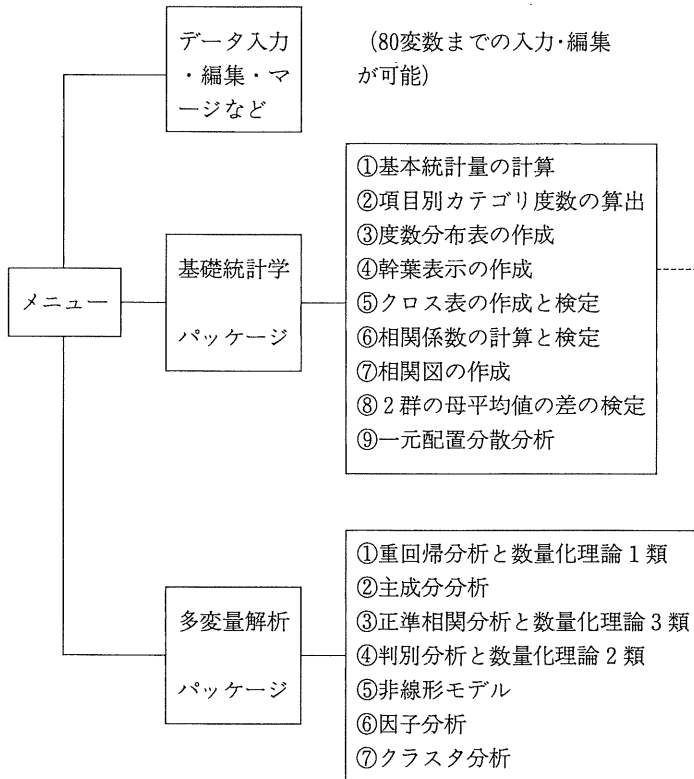
次に、HALBAU のシステム構成について説明しよう。

III HALBAU のシステム構成について

図 1 に HALBAU のバージョン 1 からのシステム構成を簡単に示した。また、各分析方法の計算機能は付録に示したので、詳細を知りたい場合には参照して欲しい。なお、現在市販中のバージョンは 2 であるが、すでにバージョン 3 の基本設計は終了し、開発途上バージョン 2.61（2 月 15 日付）として各方面に配布し、モニタリングを行っている。なお、図には示していないが、「因子分析」、「数量化理論 3 類」および「正準相関分析」については、計算に時間がかかるということから、FORTRAN 言語によるプログラムがすでに作成されている。

図 1 に示したように、HALBAU を構成するプログラムは大きく 3 つに分類できる。それらは(1)データ入

〈バージョン 1 の構成〉



〈バージョン 2 での変更点〉

○分析結果のファイル出力を追加
(クラスタ分析とグラフィック画面を除く)

〈バージョン 3 では変数の個数5000まで〉

- ⑩四分表の併合
- ⑪ 3 変数以上の同時相関図の作成
- ⑫順位和検定, 符号付順位和検定
符号検定
- ⑬Kruskal-Wallis検定
を追加, 一元配置分散分析に対比較を追加.

〈バージョン 3 での追加予定〉

- ⑭二元配置分散分析
- ⑮ 5 数要約と箱ヒゲ図

④が「判別分析」と「数量化理論 2 類」の 2 つに分割された.

〈バージョン 3 での追加予定〉

- ⑤に生命表解析のために「一般化ウイルクソン検定」を追加.

図 1 HALBAU のシステム構成

力・編集関係のプログラムシステム, (2)データの記述・探索のための基本統計関係のプログラムシステム, そして(3)多変量解析関係のプログラムシステムである (ただし, 分散分析などは便宜上(2)に含ませているが, どちらかというと(3)に近い)。

(1)のデータ入力・編集システムは, データ解析をほとんど行ったことのない素人でも, 簡単に手持ちの調査票などからデータを入力し, ディスク上にデータファイルとして保存できるように意図して作成されたものである。さらに, データの誤入力などを訂正し, 編集するのも比較的簡単に行えるように設計されている。また, 地域別や学年別などに作成したファイルを統合するために, ファイルのチェーン用プログラムも用意した。同様に, いくつかの変数ごとにファイルが作成されている場合には, それをマージするためのプログラムも組み込まれている。

また, 実際の調査では複数個の項目の反応から, 新しい尺度を構成することがよく行われるが, この目的のためにデータ変容ルーチンを用意した。ただし, デー

タ変容を行うためには, 若干の BASIC に関する知識が要求され, 全くの初心者には使用は困難である。

データ入力・編集システムに関しては, 今後他のデータベース作成のファイルを, HALBAU 用に相互に変換可能となるように, 現在システムを改訂中である。ファイル変換を計画中的ものは, dBASE-II, MULTIPLAN および固定長データファイル (FORTRAN プログラムなどでよく用いられている) などである。また, HALBAU 用データファイルの清書出力プログラムも現在作成中であり, 1988 年の春頃には完成する予定である。

(2)の基本統計プログラムシステムは, データの特性を記述し, 探索するために付け加えたものである。当初は多変量解析のみのパッケージを作成することが目的であった。しかし, 多変量解析のみでは, データ解析の立場から好ましくないという意見が多数を占めた。そこで, 「医学・保健学の例題による統計学²⁾」, 「ナースのための統計学³⁾」, 「探索的データ解析の方法⁴⁾」などを参考にして, 高木がこのプログラムシステ

ムを作成することになった。

基本統計解析プログラムシステムには、分布の特性値である平均値、分散、尖度、歪度などを求めたり、度数分布表や幹葉表示、もしくは箱ヒゲ図（バージョン3から）などを作成するプログラムなどが含まれている。また、相関係数、散布図、クロス集計などのルーチンもある。さらに、ノンパラメトリックな手法を含む各種の検定方法が組み込まれているが、詳しくは付録の機能説明を参照して欲しい。

(3)の多変量解析プログラムシステムは、より複雑な事象間の関係を分析するために作成されたものである。

「予測」のために、重回帰分析と数量化理論1類があり、「識別問題」のためには判別分析と数量化理論2類がある。また、「内的構造分析」のためには因子分析と数量化理論3類があり、「情報圧縮」のためには主成分分析がある。また、変数群の関係を分析するためには正準相関分析と双対尺度法（数量化理論3類）がある。「個体分類」や「項目分類」のためにはクラスタ分析があり、「生命表解析」のためには各種の非線形モデルおよびカプラン・マイヤ法、一般化ウィルコクソン検定（バージョン3から）などが組み込まれている。

上記のように、HALBAUには極めて多数の分析方法が含まれており、ある程度は各種の方法についての知識がなければ、どれを用いればよいのか判断できない場合もありえる。しかし、これはある意味では望ましいことでもある。よく理解できるものには、分析方法の多様さは、解析の幅を広げることになる。逆に、素人は理解の程度を深めながら、慎重に分析を行うかもしれないからである（ただし、これは制作者側の希望でもある）。

各方法の詳細をここで説明するのは困難であり、付録と成書¹⁻⁴⁾を参考にして欲しい。また、今後はパソコンも32ビットマシンが主流になると考えられるので、そのような変化にも対応させていくつもりである。

IV HALBAUの特徴と問題点について

HALBAUは当初の目的に沿って、データ解析に不慣れなものでも、かなり高度な分析が可能のように、すべてメニュー画面方式を採用している。従って、使い始めに簡単な使用説明書を一読するか指導を受ければ、その後は独力でかなりの水準まで使用できるようになる。

また、一度作成したデータファイルは、HALBAUに含まれる全ての分析プログラムでそのまま使用できるため、入力に時間がかかってもその後の処理が簡単に行える。

さらに、データ変容も可能なため、データ入力後に

必要に応じて変数の和を計算したり、対数変換したりして新しい変数に加えることも可能である。また、デフォルト処理を組み込んであるので、分析手法にそれほど詳しくなくとも、一通りの分析結果を得ることができる。

これらの点は、もともとデータ解析の素人にも統計学の手法を、手軽に使用して貰うために考えたのであり、所期の目的は達成されていると考えてよいだろう。

分析結果の表示方法に関しては、すべて一度画面に出力し、これを出力用のデバイス（プリンタかファイル）に出力するかを問うようになっている（分析方法によっては若干異なるが）。このため、必要に応じて画面上で検討するだけでも、またプリンタに出力したり、ファイルに保存したりすることも極めて容易にできる。とくに、ファイルに出力した分析結果を、「一太郎」などの日本語ワープロで編集し、報告書や論文などの表としてそのまま使用することができるので便利である。

上記の点は、HALBAUの長所と考えられるが、問題点も下記のように存在する。

HALBAUは調査データの分析を主眼として設計されているため、データが負の値を取るような場合に、分析を思うように行うことができない場合がある。この理由は、欠損値として-99の値をデータとしているため、これ以下のデータはすべて欠損扱いになるからである。この点に関しては、欠損値の自由指定などのオプションを付けるなどが考えられるが、操作が煩雑になりHALBAUのシステム構成に馴染まないため、よほど優れた方法を開発しない限り、現在のところ改訂の意志はない。

データ入力に関しては、バージョン2までは一度に80変数までしか扱えなかったため、それ以上の変数を扱う場合、マージする必要があった。また、各分析の変数指定画面も80変数を基準に考えられていたため、それ以上の変数を扱うと画面が乱れるという問題があった（分析には差障りはないが）。バージョン3ではこれらの問題はすべて解決されている。ただし、変数の個数が1000以上の場合、画面が若干乱れるが大きな問題にはならないと考えられる。

HALBAUの欠点で最も強調すべきことは、統計的解析の誤用を助長するかもしれないという点にあるだろう。パッケージとして含まれる手法は極めて多数であり、必ずしも最適な方法を用いてくれる保証はない。さらに、結果の解釈が正しく行えるか疑問もあり、今後どのようにその対策をプログラムに組み込むかが当面の最大問題である。ただし、上記の問題点は、一般にどのような統計学パッケージでも問題となることである。この解決にはユーザー自身が、データ解析の各

手法を正確に理解し、幅広い一般教養および専門的知識を身に付け、さらに直観力、知恵、知力（脳力）を育成する以外には方法がないのかもしれない。

V まとめ

データ解析のための統合型統計解析パッケージ HALBAU に関して、開発過程やシステム構成などについて、簡単な説明を行った。

HALBAU は、統計学やデータ解析に不慣れなものにも、簡単に高度な解析が行えるように設計されたものである。HALBAU を構成するプログラムシステムは、(1)データ入力・編集関係のプログラムシステム、(2)データの説述・探索および各種検定のための基本統計解析関係のプログラムシステム、および(3)予測・判別・分類など複雑な事象間の関係を分析するための多変量解析関係のプログラムシステム、と大きく3分類

できる。

HALBAU の大きな特徴は、データ入力から分析方法の指定、分析用の変数の指定、分析の実行、結果の出力まで、全て日本語によるメニュー画面方式を採用している点があげられる。このため、各種の設定を行うのに、特別な知識や訓練を多くは必要としない。

しかし、この操作性の容易さと組み込まれている分析手法の多様さが、各方法に対する知識の欠如と重なった場合、統計学の誤用および結果の解釈を誤るという可能性を、極めて大きくしているとも考えられる。これらの問題は今後のシステム開発における主要な課題であり、データ解析を指向する統計学パッケージにおいては、何等かの対策を構じる必要があるものと考えられる。

(昭和 62 年 12 月 30 日受理)

参考文献

- 1) 柳井晴夫, 高木廣文 編著: 多変量解析ハンドブック, 現代数学社, 1986
- 2) 豊川裕之, 柳井晴夫 編著: 医学・保健学の例題によ

る統計学, 現代数学社, 1982

- 3) 高木廣文: ナースのための統計学, 医学書院, 1984
- 4) Hartwig, F. and Dearing, D.: Exploratory Data Analysis, Sage Pub., 1979; 柳井晴夫, 高木廣文訳, 探索的データ解析の方法, 朝倉書店, 1981

○

(付録) 各手法の機能説明

以下に HALBAU に含れている各プログラムの計算機能などを示す。なお、プログラム名にはイクステンション部は省かれている。コンパイル版では、このプログラム名を入力することで、各手法を直接起動させることができる（コンパイル後）。

[1] データファイル作成および編集 (INP)

HALBAU 用のデータファイルの作成と編集のためのプログラム。具体的な構成内容は以下のようになっている。

① データファイルの作成と編集

HALBAU 用のデータファイルのデータ入力・編集のためのスクリーンエディター。変数名入力ルーチン、データ入力ルーチン、変数の削除・追加ルーチンからなる。

② チェーンマージ

2つの HALBAU 用ファイルを1つのファイルに統合する。

③ データの変容

HALBAU 用ファイルの変数を変換するため、プログラム作成用スクリーンエディターと、その実行ルーチンを提供。実行はインタープリタ上で行

う。

④ プリンター出力

①で作成したデータファイルをチェックするために、プリンターに出力する（バージョン3では、清書用ルーチンを追加）。

⑤ ファイルの消去

不必要なファイルを消去する。

[2] 基礎統計学パッケージ

(1) 基本統計量の計算 (BAS)

量的変数の標本数、平均値、分散、標準偏差、変動係数、最小値、最大値、範囲、歪度、尖度を求める。

(2) 項目別カテゴリ度数の算出 (FREQ)

質的変数の各カテゴリの度数と百分率を求める。

(3) 度数分布表の作成 (DOSU)

量的変数について、適当な階級区分（5, 10, 20 など）により、度数、累積度数、相対度数（百分率）、累積相対度数（百分率）を求める。また、ヒストグラム、折れ線グラフなどを作成する（バージョン3から）。

(4) 幹葉表示の作成 (STEM)

量的変数（非負の整数値）について、適当な階級区分（5, 10, 20 など）により、幹葉表示を作成

する。

(5) クロス表の作成と検定 (CROS)

2つの質的変数に関するクロス表の作成とカイ2乗検定を行う。

四分表の場合、連続性の補正を行った場合と行わない場合のカイ2乗値、それぞれの有意確率、フィッシャーの直接確率、関連係数、オッズ比・リスク比とその信頼区間などを求める。一般のクロス表の場合、カイ2乗値、自由度、有意確率、関連係数を求める。

(6) 四分表の併合 (FOUR)

2つ以上の四分表を併合し、Mantel-Haenszel 推定量を求める。コホート研究の場合、リスク比、リスク差、Cochran-Mantel-Haenszel 検定 (カイ2乗値、有意確率)、Shore-Posternack-Curnen 検定 (同左)、共通リスク比の Mantel-Haenszel 推定量とその95%信頼区間、共通リスク差の Mantel-Haenszel 推定量とその95%信頼区間を求める。ケース・コントロール研究の場合、オッズ比、Cochran-Mantel-Haenszel 検定 (カイ2乗値、有意確率)、共通オッズ比の Mantel-Haenszel 推定量とその95%信頼区間を求める。断面研究の場合、Yule の Q、ファイ係数、Cochran-Mantel-Haenszel 検定 (カイ2乗値、有意確率)、共通 Yule の Q の Mantel-Haenszel 推定量とその95%信頼区間、共通ファイ係数の Mantel-Haenszel 推定量とその95%信頼区間を求める。

(7) 相関係数の計算と検定 (COR)

量的変数について、各変数の平均値、標準偏差、変数間のピアソンの積率相関係数、無相関の検定 (t 値、自由度、有意確率) を求める。

(8) 相関図の作成 (SCAT)

2つの量的変数について、平均値、標準偏差、変数間のピアソンの積率相関係数、無相関の検定 (t 値、自由度、有意確率)、単回帰式、および相関図 (散布図) を求める。

(9) 3変数以上の同時相関図の作成 (MULSCAT)

3変数以上の量的変数 (10変数程度まで) に関して、相関係数行列のかわりに同時相関図を作成する。

(10) 2群の母平均値の差の検定 (TWO)

質的変数の値によりケースを2群にわけ、量的変数の母平均値に差があるかを検定する。独立2標本の場合、各群の平均値、分散、分散比 (F 値)・自由度・有意確率 (等分散の検定)、等分散の場合の2群の母平均値の差の検定 (t 値・自由度・有意確率)、および不等分散の場合の Welch 検定 (t 値・自由度・有意確率) を求める。

対応のある場合の母平均値の差の検定 (各群の平均値・分散、差の平均値・分散、t 値・自由度・有意確率) を求める。

(11) 一元配置分散分析 + α (MUL)

質的変数の値によりケースを指定したグループ数にわけ、量的変数の母平均値に差があるかを分散分析により検定する。各群の平均値、分散、級間変動・級内変動・全変動およびその不偏分散、分散比 (F 値)・自由度・有意確率を求める。各2群の組合せについて、対比較を行う (Bonferroni の方法、Scheffe の方法による)。

(12) 順位和検定、符号付順位和検定、符号検定 (WIL)

質的変数の値によりケースを2群にわけ、量的変数の分布の位置に差があるかを、ノンパラメトリックな手法である Wilcoxon の順位和検定 (U 検定) を用いて行う。2群に対応のある場合には、符号検定、Wilcoxon の符号付順位和検定を行う。

(13) Kruskal-Wallis 検定 + α (KW)

質的変数の値によりケースを指定したグループ数にわけ、量的変数の分布の位置に差があるかを、ノンパラメトリックな手法である Wilcoxon の順位和検定 (U 検定) を3群以上に拡張した手法である Kruskal-Wallis 検定を用いて行う。各2群の組合せについて、対比較を行う (Bonferroni の方法、Schéffe の方法による)。

(14) 二元配置分散分析 (2WAY) (バージョン 3 から)

量的変数の変動に関して、2要因の主効果、交互作用の有無を分析する。反復数が1の場合、反復数が2以上で等しいか比例関係にある場合、反復数が等しくない場合の3通りについて分析が可能。

各要因の水準ごとの平均値、(不偏) 標準偏差・各要因変動・交互作用・残差変動・全変動およびその不偏分散、分散比 (F 値)・自由度・有意確率を求める。また、交互作用を残差に含めて再計算も可能。

各要因ごとの各2水準間の平均値の差について、対比較を行う (Bonferroni の方法、Schéffe の方法による)。

(15) 5数要約と箱ヒゲ図 (BOX) (バージョン 3 から)

量的変数の最小値、第一四分位、中央値、第3分位、最大値、および四分位偏差を求める。また、これをもとに箱ヒゲ図を作成する。

[3] 多変量解析パッケージ

(1) 重回帰分析と数量化理論 1 類 (REG)

説明変数が多数あり、説明変数が量的データである場合に重回帰分析を、質的データである場合に数量化理論1類を行うことができる。重回帰分析の場合、各変数の平均値、標準偏差、各変数間の単相関係数、説明変数の偏回帰係数・標準偏回帰係数・F値・有意確率、重相関係数・多重決定係数とその検定（F値・自由度・有意確率）、および自由度調整済重相関係数、自由度再調整済重相関係数を求める。

説明変数の選択方法は、一括投入法、変数増加法・変数減少法・変数増減法・変数減増法があり、F値を基準とする場合とAIC（赤池の情報量規準）による場合を選ぶことができる。また残差プロット、標準化残差のHALBAU用ファイル出力が可能。

数量化理論1類の場合、説明変数間のクロス表・カテゴリごとの基準変数の平均値と標準偏差、標準化数量、基準変数との偏相関係数、重相関係数などを求める。

(2) 主成分分析 (PRI)

各変数の平均値・分散・標準偏差、変数間の相関係数、各変数の主成分負荷量、固有値とその検定、寄与率、累積寄与率、主成分得点などを求める。主成分得点はHALBAU用ファイル出力が可能。

(3) 正準相関分析と数量化理論3類 (CAN)

正準相関分析では各変数の平均値・標準偏差、変数間の相関係数、各変数の重みベクトル・構造ベクトル、各成分の固有値の検定・正準相関係数・冗長性などを求める。

数量化理論3類では、各カテゴリおよびケースの重み数量、各成分の固有値などを求める。

(4) 判別分析 (DIS)

説明変数が量的データの場合、2群および多群の判別を行う。グループ別に各変数の平均値・標準偏差・変数間の相関係数、各変数の判別係数、バートレットのカイ2乗検定、自由度・有意確率、ボックスのM検定（カイ2乗値・自由度・有意確率）、各群の判別得点の平均値、誤判別率、判別得点のHALBAU用ファイルの出力、説明変数の選択は一括投入及び変数増加法が可能。

(5) 数量化理論2類 (HQ2)

説明変数が質的データの場合、2群および多群の判別を行う。各群ごとに説明変数間の各カテゴリのクロス表、カテゴリ数量、各群ごとの判別得点の平均値などを求める。

(6) 非線形モデル (NON)

多重ロジスティックモデル、多重ワイブルモデル、Coxの生命表解析（比例ハザードモデル）、カプラン・マイヤの生命表解析、一般化ウィルコクソン検定（バージョン3から）が含まれている。

各手法に関して各モデルに含まれる説明変数の係数、F値、有意確率、AICなどを求める。3つの生命表解析の手法に関しては、期間別生存率・累積生存率などを求める。

(7) 因子分析 (FAC)

因子分析の手法として主因子法、最尤法の2通りの選択ができる。

共通性としては、すべて1、SMC、相関係数最大、自由指定、などが選択でき、共通性の反復推定も可能・因子数も指定できる。

因子の回転は、直交回転（バリマックス法、バイコーティマックス法、コーティマックス法）および斜交回転（コバリミン法、バイコーティミン法、コーティミン法）を選択できる。

回転前および回転後の因子負荷量、斜交回転の場合には回転後の構造行列・因子間の相関係数行列を求める。

因子得点の指定には、残差平方最小、独自性を除外、Thomsonの方法、Bartlettの方法を選択することができる。また、因子得点のHALBAU用ファイル出力が可能。

(8) クラスタ分析 (CLUS)

個体分類、または項目分類を行う。個体分類の場合、用いる変数のデータを標準化するか、そのままか、もしくはマハラノビスの汎距離かを指定できる。項目分類の場合、相関係数を求め、距離に変換する。分析手法は階層的手法（最近隣法、最遠隣法、メディアン法、群平均法、重心法、ウォード法）であり、デンドログラムを出力する。

DEVELOPING THE INTEGRATED STATISTICAL ANALYSIS PROGRAM PACKAGE 'HALBAU'

by Hirofumi TAKAGI and Haruo YANAI

We have developed the integrated statistical analysis program package 'HALBAU' (High quality Analysis Libraries for Business and Academic Users) mainly for the nonexperimental studies.

HALBAU consists of the three major subsystems ; i.e., (1) the data edition program system for the data input and correction, the file chain and merge, and so on, (2) the fundamental statistical analysis program system for the data description, the exploratory usage, and the statistical test of the hypothesis, and (3) the multivariate analysis program system for the prediction, the discrimination, and so on by elucidating the relationships among the complicated phenomena.

For nonprofessionals in the field of the data analysis, the menu selection method in Japanese has been adopted in all cases where we must chose the method of use, the variables, the output device for the outcomes, and so on. Therefore, no one needs special training except collecting own research data.

However, even though HALBAU has the operative easiness and includes various statistical methods, one may fall into the statistical misapplication of the method and/or the misunderstanding of the results without the accurate knowledge of the method of use and the experience of the data analysis. Since these unsolved problems will be the main subject about the development of the statistical packages for the data analysis, we are urged to take new measures against them.