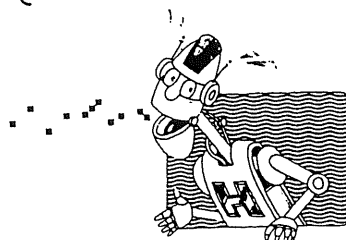


# HALBAUのバージョンアップについて



高木廣文 (統計数理研究所)

## 1. はじめに

HALBAU(はるぼう)は、柳井・高木他(1986)による「多変量解析ハンドブック」に掲載されたサブプログラムをもとに開発された。しかし、単に多変量解析の基本ルーチンの寄せ集めではなく、データ入力・編集、基礎的な統計学の手法のためのプログラムを大幅に加えて、統合的なパッケージとして作成されたものである。これは実際の分析において、いきなり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たからである。また、分析のためには手持ちのデータをコンピュータで使えるようにする必要がある。そのため、データ入力用のシステムも提供し、スクリーンエディタによりデータの入力ができるようになっている。さらに、データの変容(データを足したり、対数をとったりすること)も、ある程度は簡易型の変容システムでできるようになっている。

1986年以来、バグの修正や方法の追加などを行ってきたが、大幅な改訂は1989年のバージョン3の開発以後は行っていなかった。昨年のBASIC数学誌上で述べたように、この2年間ほどバージョン4の開発を行っていたが、新しい機能や手法を加えると必ずバグやエラーがあり、思いのほか時間を費やしてしまった。目立つようなバグもなくなった(と思う)ので、その旧バージョンとの相

違など、その概要について簡単に報告する。

## 2. HALBAU バージョン4について

HALBAUの基本的な構成はバージョン3とそれほど変わってはいない。表1, 2, 3に示したように、「データおよびファイルの操作」、「基礎統計学パッケージ」、「多変量解析パッケージ」のように区分されている。実際には、全ての方法で改訂がなされており、機能の追加や操作性の向上が図られている。

新しい機能をつけたり、方法を増やしたために、今まではフロッピーディスク1枚に納まっていたが、新バージョンからは、「基礎統計学編」と「多変量解析編」の2枚組になった。ただし、「データおよびファイルの操作」のためのプログラムは、両方ともに納められているので、実際には全体で約1.5Mバイト程度の大きさである。従来通り、フロッピーディスクしか使用できない機種でも動作は可能であるが、操作性からは後述する理由からも、ハードディスク上にコピーして使用す

表1 データおよびファイルの操作

| 名 称               | 主 な 機 能 説 明                                                             |
|-------------------|-------------------------------------------------------------------------|
| 1. データファイルの作成と編集  | 分析用データ入力と編集、2つのファイルのチェーンとマージ、HALBAU用ファイルの消書出力、ファイル消去など。                 |
| 2. データファイルの変換     | HALBAU用データファイルと固定長/MULTIPLANのSYLK形式/dBASE II/NAP形式(カンマ区切り)の各ファイルとの相互変換。 |
| 3. データ簡易変容とファイル分割 | 変数の重み付・べき乗線形和の計算、変数の再カテゴリ化、対数変換。ファイルの分割など。特定の変数によるソート。                  |
| 4. グラフの編集・作成      | HALBAUでファイルに保存したグラフィックスデータの編集、もしくは簡単な作図を行う。                             |
| 5. ドライブ・ファイル等の設定  | HALBAUで使用する入力用のドライブ/ファイル、出力用のドライブ/ファイル・プリンタ、作業用ドライブの設定。                 |
| (データ変容)*          | データの変容プログラム作成用エディタと、その実行。                                               |

\*メニュー画面には表示されない。

ることが望まれる。

手法によっては計算のアルゴリズムを変えたりしたものの多いが、主な変更点は以下の通りである。

- (1) コンパイラ版とインタプリタ版の区別をなくした。

HALBAU は N88BASIC でプログラムされているのだが、インタプリタではスピードが遅いため、できるだけコンパイルして用いる方がよい。両者のプログラムを若干変えていたのだが、その必要がないので、その区別をやめた。したがって、当初はインタプリタ上で HALBAU を使っていても、コンパイラさえ入手すれば、コンパイラ版を購入しなくとも、その日からすぐに数倍以上の処理速度の向上が望めるようになった。

- (2) 分析方法／ファイル／変数などの指定での操作性を向上した。

各種の指定は、以前はメニュー画面に表示されている処理の番号を、キーボードから入力していたが、カーソルキーで簡単に指定できるよ

うになった。また、以前は分析用のデータ・ファイル名やドライブの指定を、分析方法を変更する度にキーボードから設定し直していたが、同一のドライブおよびファイルであれば、リターンキーを押すだけでよかった。これは、入力ファイルだけでなく、結果の出力ファイル、作業用のファイルなども、一度設定すれば、変更する必要が起るまで何もする必要がなくなった。

- (3) 分かりにくい指定画面でのヘルプ機能の設定。

バージョン 3 では、各種の指定についてヘルプ機能がある場合とない場合があった。また、多変

表 2 基礎統計学パッケージ

| 分 析 方 法                       | 主 な 機 能 説 明                                                         |
|-------------------------------|---------------------------------------------------------------------|
| 1. 基本統計量の計算                   | 標本数, 平均値, (標本) 分散, 標準偏差, 変動係数, 総和, 最小値, 最大値, 歪度, 尖度。                |
| 2. 項目別カテゴリの算出                 | 質的変数の各カテゴリの度数と百分率。                                                  |
| 3. 度数分布表とヒストグラム               | 適当な階級区分で度数, 累積度数, 相対度数, 累積相対度数, ヒストグラム, 折れ線グラフなどを作成。                |
| 4. 幹葉表示の作成                    | 適当な階級区分で幹葉表示を作成。                                                    |
| 5. クロス表の作成と検定                 | クロス集計と検定 (カイ 2 乗検定, フィッシャーの正確な確率, mid-p による正確な確率) など。               |
| 6. 四分表の併合                     | 層別四分表の共通オッズ比, リスク比のマンテル・ハイツェル推定量などの算出。                              |
| 7. 相関係数の計算の検定                 | ピアソンの積率相関係数 / スピアマンの順位相関係数 / ケンドールの順位相関係数などの算出と無相関の検定。              |
| 8. 相関図の作成                     | ピアソンの積率相関係数, 無相関の検定, 単回帰式など。(層別) 相関図の作成, 領域図の作成など。                  |
| 9. 3 変数以上の同時相関図               | 3 変数以上の相関係数行列のかわりに同時相関図を作成。単相関係数, 単回帰式など。                           |
| 10. 2 群の母平均値の差の検定             | 独立 2 標本の場合, および対応のある場合について母平均値の差の検定を行う。                             |
| 11. 一元配置分散分析 + $\alpha$       | 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, F 検定。各 2 群についての母平均値の差の検定。       |
| 12. 順位和検定など                   | ウィルコクソンの順位和検定 (U 検定), 対応のある場合には符号検定, ウィルコクソンの符号付順位和検定など。            |
| 13. Kruskal-Wallis + $\alpha$ | クラスカル・ウォリス検定と対比較 (ボンフェロニの方法, シュフェの方法)。                              |
| 14. 二元配置分散分析 + $\alpha$ *     | 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・不偏分散, F 検定など。          |
| 15. 5 数要約値と箱ヒゲ図               | 最小値, 第一四分位, 中央値, 第三四分位, 最大値, 四分位偏差, 箱ヒゲ図の作成など。                      |
| 16. 生命表解析                     | カプラン・マイヤ法による生命表解析, マンテル・ハイツェル検定, ログランク検定, 一般化ウィルコクソン検定, 累積生存率の作図など。 |

\* 層別変数は指定できない。

量解析ではほとんどヘルプ機能は設定されていなかった。バージョン 4 から、「分析方法の指定」、「入力ファイル名指定」、「分析用変数指定」、「層別変数指定」、「制御変数指定」、「欠損値指定」の各画面では、全ての分析方法でヘルプ機能が設定された。ヘルプ画面の内容は、現在は簡単なものであるが、今後もより分かりやすいように改訂していく予定である。

- (4) 分析結果の出力画面の停止, 逆スクロールなどが可能となった。

分析結果の画面出力は、従来は表示する量が多いと結果が一面に入らないため、見にくい場合

もあった。この点を改良し、画面上の表示がスクロール・逆スクロールできるようにした。

- (5) グラフィックス画面をファイルに保存できる。さらに、グラフィックス画面の編集・作成プログラムを追加した。

これまで、通常の分析結果はファイルに保存できるものの、ヒストグラム、相関図などのグラフィックス画面はプリンタ出力のみが可能で、ファイルに保存できなかった。グラフィックス画面をファイルに保存し、必要に応じてある程度の編集ができるように、グラフの編集・作図のためのプログラムを追加した。

- (6) 基礎統計学では、層別変数による分析が可能となった。

実際のデータ解析では、平均値を求める場合でも、性別

や学年別などのように、特定の変数の値別に集計が必要なものも多い。これまでは、そのような場合、制御変数を用いて繰り返し集計を行う必要があった。層別変数として複数の変数を指定することで、一度に性別・学年別などの基本的な解析や集計が行えるようになった。

- (7) 基礎統計学に「生命表解析」、多変量解析に「尺度構成・信頼性係数」を追加。

分析手法のメニュー画面上では、新たに加わったのはこの2つの手法であり、従来の「領域図の作成」は「相関図の作成」に併合された。また、「重回帰分析」と「数量化理論1類」を分離し、同様に「判別分析」と「数量化理論2類」も分離した。また、「非線形モデル」は「リスク関数などによる解析」とし、「条件付きロジスティック・モデル」を含めた。

各分析手法について個別に説明できないが、手法や機能を追加している。

表3 多変量解析パッケージ

| 分 析 方 法         | 主 な 機 能 説 明                                                                                                                               |
|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| 1. 重回帰分析        | 偏回帰係数・標準偏回帰係数・F値・有意確率、重回帰係数・多重決定係数とF検定など。変数選択は、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。てこ比の算出、残差プロットの作成等。                                           |
| 2. 数量化理論1類      | 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。                                                                                |
| 3. 主成分分析        | 主成分負荷量、固有値とその検定、寄与率、累積寄与率、主成分得点などの算出。主成分負荷量による変数の平面配置図の作成。                                                                                |
| 4. 正準相関分析       | 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。                                                                                                |
| 5. 数量化理論3類      | 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。各カテゴリの数量による平面配置図の作成。                                                                                           |
| 6. 判別分析         | 2群および多群の判別。各変数の判別係数、パートレットのカイ2乗検定、ウィルコクスの $\Lambda$ 、ボックスのM検定などの算出。各群の判別得点の平均値、誤判別率など。                                                    |
| 7. 数量化理論2類      | 2群および多群の判別。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、各群の判別得点の平均値、誤判別率など。                                                                                |
| 8. リスク関数などによる解析 | 多重ロジスティックモデル、多重ワイブルモデル、比例ハザードモデル、条件付きロジスティックモデルによる解析。各手法に関して各モデルに含まれる説明変数の係数、F検定、各変数のオッズ比と95%信頼区間の算出。生存解析については、期間別生存率・累積生存率など。累積生存率を作画する。 |
| 9. 因子分析         | 主因子法、最尤法。共通性の初期値指定、反復推定指定可能。因子の回転は、直交回転・斜交回転を行う。因子得点の算出、各変数について因子負荷量による平面配置図の作成など。                                                        |
| 10. クラスター分析     | 項目分類、個体分類を階層的手法により行い、デンドログラムを作成する。                                                                                                        |
| 11. 尺度構成・信頼性係数  | クロンバッハの $\alpha$ 信頼性係数、 $\theta$ 信頼性係数などを算出。各項目の第1主成分負荷量、他の変数によるSMCなどを算出する。                                                               |

- (8) 作業用ファイルを使用することにより、一部に制限のあった手法で、よりケース数の大きな解析が可能となった。

バージョン4から、分析の途中結果や出力を作業用のディスクに書き出すことで、メモリーの節約や、各種の操作を容易に行えるようになった。BASICの制限があるものの、より実用的なものとなった。ただし、作業用ディスクを用いるので、RAMディスクやハードディスクを用いる場合はよいが、フロッピーディスクだけの場合には、大量データでは若干時間がかかる場合もある。

- (9) マニュアルを整備した。

今までは、簡単なマニュアルしかなく、高木ら(1989)による解説書ぐらいしか参考になる文献はなかった。この点を考慮して、基本的な操作の説明のためのマニュアル(これには、「データおよびファイルの操作」のパッケージの完全対応の解

説が含まれている)を作成した。さらに、「基礎統計学パッケージ」の内容についても、方法の説明とともに、画面上の操作方法を完全に対応させたマニュアルを作成した。ただし、多変量解析については、「多変量解析ハンドブック」があるので、いまのところマニュアルの作成は行っていない。

### 3. 実行例

#### (1) メニュー画面の例

実際に HALBAU を始動すると、図 1 のような画面が表示される。

メニュー画面で方法を指定するためには、カーソルキー [↑] と [↓] で縦方向の分析方法を指定し、左右のメニューへの移動は [←] と [→]、またはリターンキーで行う。画面上で黄色の帯になっている箇所が、カーソルのある位置である。

左側のメニューでリターンキーを押すと、右側のメニューにカーソルが移動し、黄色の帯も移動する。右側のメニューでリターンキーを押すと、そのプログラムを実行する。

#### (2) ヘルプ画面の例

各種のメニュー画面を表示中に [HELP] キーを押すと、ヘルプ画面が表示され、簡単な説明が得られる。図 2 は層別変数のヘルプ画面である。

#### (3) 結果の出力画面の例

分析結果の出力を表示すると、図 3 のような画面表示になる。

| HALBAU-4 メニュー    | データとファイルの操作関係メニュー |
|------------------|-------------------|
| 1. データおよびファイルの操作 | 1. データファイルの作成と編集  |
| 2. 基礎統計学パッケージ    | 2. データファイルの変換     |
| 3. 多変量解析パッケージ    | 3. データ簡易変容とファイル分割 |
| 4. 終了            | 4. グラフの編集・作成      |
|                  | 5. ドライブ・ファイル等の設定  |

移動: [→] [←] [↑] [↓]  
確定: [↵] リターンキー

図 1 HALBAU バージョン 4 のメニュー画面

移動  
確定

[ 層別変数について ]

分析を、ある変数の特定の数値ごとに行いたい場合があります。例えば、「男」、「女」の別に集計をしたい、または、学年別に集計をしたい、といった場合がそれにあたります。

そのような場合、「性別」や「学年」という変数が「層別変数」ということになります。

当然ですが、層別変数はあらかじめ変数として、各ケースにそれぞれデータが入力されていなければなりません。

◎層別変数は 1 つ以上指定できます。ただし、あまり多数指定すると、層別変数の各組合わせにより、有効な標本数が極端に少なくなり、分析に支障をきたすことがあります。

どれかキーを押して下さい。

図 2 層別変数についてのヘルプ画面

| [ 分析結果の画面出力 ]                                 |     |                  |         |        |        |  |
|-----------------------------------------------|-----|------------------|---------|--------|--------|--|
| [基本統計量解析結果]    File = hal.hal    94/03/03     |     |                  |         |        |        |  |
| 層別変数名                                         |     | ----- 層の範囲 ----- |         |        |        |  |
| 1)性                                           |     | 層 1              | 1.000   | ～      | 1.000  |  |
|                                               |     | 層 2              | 2.000   | ～      | 2.000  |  |
| [全体]                                          |     |                  |         |        |        |  |
| 変数名                                           | 標本数 | 平均値              | 分散      | 標準偏差   | 変動係数   |  |
| 1)年齢                                          | 50  | 57.700           | 80.970  | 8.998  | 15.595 |  |
| 2)身長                                          | 49  | 155.912          | 44.373  | 6.661  | 4.272  |  |
| 3)体重                                          | 50  | 57.012           | 80.687  | 8.983  | 15.756 |  |
| 4)最高血圧                                        | 50  | 127.460          | 216.008 | 14.697 | 11.531 |  |
| 5)最低血圧                                        | 50  | 80.320           | 93.618  | 9.676  | 12.046 |  |
| 総 和                                           |     | 最小値              | 最大値     | 歪度     | 尖度     |  |
| [↑]上 [↓]下 [ROLL DOWN]前画面 [ROLL UP]次画面 [ESC]終了 |     |                  |         |        |        |  |

図 3 結果の出力画面

# 多変量解析ハンドブック

柳井晴夫 / 高木広文 編著 定価4,944円

- 1 これまでに開発され、また、比較的最近までの発展を含めた多変量解析法（非計量の多次元尺度解析法、対数線形モデルを除く）について、それぞれの手法の基本原理をなるべく平易に解説することを試みた。
- 2 読者自身がプログラムを作成できるように、各種技法の計算手順をなるべく詳しく解説した。（各章の巻末に Basic 言語によるプログラムを付けてある）
- 3 各章に、実際の生データ（架空データではない）を数値として与え、それに基づく、一通りの解析結果を示した。（これにより、読者の作成したプログラムのチェックを行うことができる。さらに、計算の過程を follow することによって各種技法の基本理論に対する読者の理解度が深まる）
- 4 多変量解析に関して歴史的に著名な文献（多変量分布論に関するものは除く）および著書はなるべく広く収集し、それらを巻末にまとめて示した。（文献の数は約300）また、これらの文献を通して、多変量解析のそれぞれの手法が歴史的にみて、どのような人によって開発されたかについて、なるべく、詳しく解説した。
- 5 多変量解析に対するここ数年の新しい考え方（重回帰分析における残差分析：非線形モデルにおける比例ハザードモデル、因果分析：正準相関分析における冗長性係数）を解説した。
- 6 これまで、同一の手法が異なる名称で呼ばれるものがあつたが（例えば、多群の判別分析）これらについて、著者の見解を示した。
- 7 第10章において、多変量解析の適用上の問題点（データの信頼性：質的データの取り扱い：多変量外れ値の取り扱い：多変量解析の各種技法によって得られる、種々の係数の解釈についての問題点（多重共線性の取り扱いを含む））を解説し、今後への傾向として、探索的多変量解析法が重視されていくようになることを指摘した。また、10章の後半部分では、多変量解析の方法がどのような分野で広く適用されているかについて、領域別に解説した。
- 8 巻末に、本書を読むのに最低限必要とされる数学的知識（線形数学の基礎：確率分布の基本概念：確率分布とその近似計算法等）について解説してある。
- 9 巻末に、世界的にもっとも広まっている多変量解析パッケージ SAS, BMDP, SPSS について簡単に解説した。

以上の特色にみられるように、本書は、今日、up to dateな方法として注目されている多変量解析（より限定された意味では、多変量データ解析）に関するなるべく多くの情報を読者に提供することを試みたものである。

## 〔プログラムパッケージ：HALBAU〕別売

- データの入力の方法が EDIT 機能の完備により、これまで市販されているプログラムパッケージの中では、もっとも充実している。
- 変数の指定が容易である。
- 欠損値の取り扱いが工夫されている。
- 多変量解析だけでなく、一変量の解析、二変量の解析：クロス集計と独立性の検定が含まれている。etc.,--  
(適用機種 PC 9800 シリーズ 28,840円)

現代数学社 〒606京都市左京区鹿ヶ谷西寺ノ前町1  
電話 (075) 751-0727 振替京都 11144

図3は基本統計量の出力画面を表示したものであるが、カーソルキー[↑][↓],および[ROLL DOWN]・[ROLL UP]キーで画面表示をスクロール・逆スクロールさせてみる事ができる。画面の最下段に表示されているように、終了させたい場合には、[ESC]キーを押せばよい。その後、分析結果を出力用のファイル、プリンタなどに出力するかをきいてくる。この手順は従来と同様である。

## 4. HALBAU の今後について

極めて簡単に HALBAU のバージョン 4 について説明したが、今後さらにバグの訂正、手法の追加などもあるかもしれない。

気の早い話かもしれないが、簡単に今後の開発計画を述べておく。現在、HALBAU は N88 BASIC によってプログラムされているため、NEC の 98 シリーズ上でしか動作しない。しかし、IBM コンパチ機やマッキントッシュ上でも動かしたいという希望もある。当面、Quick-BASIC への移植を行い、時間のかかる箇所を FORTRAN などで書き換えてリンクするつもりである。実際には、数年前から、HALBAU のうち時間のかかる分析方法については、MS-FORTRAN によりプログラムされたものが存在するのであるが、一般にはリリースされていない。また、Visual BASIC などへの移植も考慮しているが、どのような手順で行っていくかは未定である。1年後には何らかの形で、NEC(と EPSON)以外の機種でも、HALBAU が動くようにしたいと考えている。

## 参考文献

- 柳井晴夫、高木広文 編著 (1986)、多変量解析ハンドブック、現代数学社 (京都)。  
高木広文、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)。  
高木広文 (1992)、HALBAU による研究の実際、BASIC 数学、50-55、現代数学社 (京都)。

(たかぎ ひろふみ)