

統計プログラムパッケージのあり方と Star/B

川崎医科大学 数学教室

林 篤 裕

(昭和62年8月31日受理)

Some Remarks of Statistical Program Packages and Star/B

Atsuhiro HAYASHI

*Department of Mathematics,
Kawasaki Medical School,
Kurashiki, 701-01, Japan*

(Received on Aug. 31, 1987)

概 要

計算機が容易に利用できるようになってきた現在、統計プログラムパッケージはデータ解析を行う場合に無くてはならないソフトウェアの1つであるといえる。一方、近年のパーソナルコンピュータの発展と普及には目を見張るものがあり、一世代前の大型計算機並の作業を個人レベルで利用することも夢ではなくなってきた。こうした状況から、パーソナルコンピュータ上で動作する統計プログラムパッケージが望まれるようになり、また技術的にも可能となってきた。

本論文では、まず統計プログラムパッケージの利用に際して要求される項目を考察した。続いて、これら要求を考慮し、かつパーソナルコンピュータの動作特性を生かして開発した統計プログラムパッケージ Star/B の特徴を述べた。このシステムは多くの解析手法を含んでおり、グラフィック機能をサポートしているばかりではなく、会話的に対象データを編集・加工・抽出できるデータマネジメント機能を有しているため広範な分野での利用が可能である。

Abstract

The development and popularization of personal computers have had a strong influence on the field of statistics during these ten years. We can analyze our data by using personal computers easily, just as we used them on main frame computers before.

We developed a new statistical package that made good use of personal computers named Star/B. It is designed not only for beginners, but also for senior statisticians. Users need to have smallest knowledge of computers. It includes many statistical methods, from introductory to advanced, with graphical outputs and powerful utilities for data management.

In this paper, we discuss some important items of statistical program packages for analysts and developers. And we mention the structures and the features of Star/B.

1. はじめに

今日の情報化社会における統計解析理論の研究は、計算機の発達と普及にともない、一層重要な役割を期待されている。データ解析理論は、元来、自然や人文社会の諸科学に協力支援し、データを収集・整理・分析・要約して、現象の説明や情報の把握を行うことを目的とする。そして、このような意図を直接的に支援・実行する統計プログラムパッケージ (Statistical Program Package, 以下 S.P.P. と略) の研究・開発は、社会と諸科学に大きく寄与するものと位置づけられる。

S.P.P. は、当初、独立した統計プログラムとして利用されていたものの蓄積が、全体に統一的しかも有機的に利用できるように総轄されたものであった。また、利用形態の変化から言えば、パンチカードを用いたバッチ処理形式から、各人が所有するターミナルやワークステーションを用いた TSS 形式へと発展してきた。個人で容易に利用できるようになってきたことから会話型の S.P.P. が望まれるようになり、最近では R. Becker, J. Chambers らの『S』¹⁾ に代表されるような関数型言語が注目されるようになってきた。ハードウェア的には、パーソナルコンピュータ (パソコン) と呼ばれる小規模な計算機の発展が目ざましく、周辺機器をはじめ処理速度およびメモリーサイズにおいて一世代前のメインフレームを凌ぐ勢いである。こうした状況から、パソコンの動作環境を生かした統計プログラムの開発が望まれるようになってきた。

そこで我々は、多くのパソコンで利用できるように BASIC 言語で記述した S.P.P. として Star/B を開発した。これは、これまでパソコン上に独立して開発・蓄積してきた統計プログラム群²⁾ を体系的にまとめ、またデータの編集・加工・抽出を会話的に行うことのできるデータマネージメント機能を新たに加え、パッケージとして再編成したものである。

本論文では、まず S.P.P. 利用に際して要求される項目を明らかにし、つづいて、これらを踏まえて開発された Star/B の特徴を述べる。

2. 統計プログラムパッケージのあり方

統計プログラムパッケージは、当然電算機の発展に基づくが、当初、独立した統計プログラムの蓄積として利用されていたものが、全体を統一的に、しかも有機的に利用できるように総轄されたものである。現在、国内的³⁾にも国際的⁴⁾にも多くの S.P.P. が研究・開発・利用されているが、それらは個々に独自の特徴を持っており、一概に優劣を付ける事はできない。つまり、S.P.P. の存在性は、その利用環境と全機能的評価に大きく依存する。

S.P.P. を論議する場合、どの様な観点から S.P.P. がとらえられ、運用されているかを知ることにより、その意図の形態を把握することができる。ここでは、利用環境と S.P.P. の評価の二点から、形態を考察する。

(I) 利用環境

利用環境は、利用者の要求や対象とする母集団の性質に密接に関係しており S.P.P. を設計・開発する上で重要な項目となる。

① 利用目的 (利用分野)

標本調査・社会福祉・計量経済分析・医療分析など、特徴的な適用分野をもつ S.P.P. と、一般的な汎用のものがある。当然、これらの違いにより、包含されるデータ・ハンドリングや解析の方法論が異なる。

② 計算機設備

所属の設備・利用計算機の規模・機種とシステム構成、また使用言語や OS などソフトウェア言語と体系に依存して S.P.P. の様式や利用法が異なる。

③ データの量と質

量については、対象分野・現象・資料の収集法・出力の目的・経済効率等に関連する。質については、その精度・信頼性等が問われる。資料の形式は文字型か数値か、その数値が名義・順序・区間・比率の各尺度をどれ程に含むか、にも関係する。

④ 利用者のデータ解析や統計数学・実質科学・計算機に関する知識水準

利用者層の経験・熟練度・技術知識・好みに依存し多様である。

(II) 統計プログラムパッケージの評価

S.P.P. の評価は、(I)の各項目とも密接に関連し、とくに統計的データ解析面の実行者の観点で多彩かつ独自の改善意見が盛んである。また、統計学の専門家は、S.P.P. 開発の非常な努力とは別に、適用方法論や出力の精度等の上で一般利用者の誤用を憂える論議が多い。S.P.P. の評価は機能評価と有用性の二面の観点から整理できる。

i) 機能評価

統計的データ解析には、データを解釈し表現する数学模型について、探索的・発見的立場からのデータ解析や検証的立場の統計解析が、S.P.P. によって自在に実行できる必要がある。

① 入出力機能

検証的または探索的データ解析においては、種々の統計モデルを試行錯誤しながら、S.P.P. に適用する。したがって、バッチ処理形式より、会話型に入出力を行える TSS 形式の方が有利である。また、入力操作が動的に実施できるためには、入力誤りの検出・修復・標準省略・障害時のデータ保護などに配慮が必要である。

② データマネジメント機能

収集された素データの形式は、多種多様で、必ずしも直ちに入力情報として処理や解析に適した形になっているとは限らない。したがって、解析前にデータの再編集や

数値変更, 加工を行ったり計算結果からモデル式を再構成したりする事も頻繁に必要となる。

③ 解析機能

広範な解法を高品質で含むべきことは当然であるが, とくに計算アルゴリズム上の計算精度と速度に配慮される必要がある。また, データ集合が病的状態にある時や数学的構造模型に合致していない時などには, 不適當解など出力結果の信頼性に問題が生じるので, 利用者に対して警告を発する必要がある。

④ 新研究成果の追加機能

システム機能は固定的・閉鎖的であってはならない。統計的解析や情報処理の研究は年々進展し新成果が生成される。そこで, これらの新成果を, 多くの努力を投じて研究開発された S.P.P. 上に容易に増強・追加できれば, 新旧の機能を不断に蓄積・改善・削除し成長させることができるため, 非常に有益である。

ii) 有用性

実際に役立つ・使い易いと言う観点から S.P.P. を評価すると, S.P.P. の有用性は, 優れた機能を持つ良い S.P.P. の評価とは必ずしも一致せず, 前述の利用環境中の各項目に密接に関連し, 利用者の主体的評価となる。そこでは, 利用者が種々の特性に付与する重みづけに依存して多様である。次に, これらの要因を挙げておく。

① S.P.P. と利用者間のインターフェース

S.P.P. の各種ドキュメンテーションや手引書類の完備性, 入出力の形式, 制御の方式, 教育的配慮, 誤操作や問い合わせに対する応答機能, 融通性, 拡張性。

② 統計機能的有用性

データの点検, データマネージメント機能, 欠測データ処理, 図式表現, 収容解法, 精度, モデルビルディング機能。

③ 実行性

機種独立性, 移植と稼働の容易さ, 経費, 保守の容易さ, 入出力操作の多様性と容易さ, 処理速度, 会話機能。

3. ソフトウェア設計

前章で示した項目を十分考慮しつつ, より広範な解説手法を包括し, また探索的および検証的データ解析に適應する S.P.P. とするために, Star/B の開発にあたって次の 5 点を考慮してソフトウェアを構築した。

(1) 応答方式による入力

解析に必要なパラメータの指定方法には, キーワード方式と応答方式の 2 通りがある。前者はキーワードと共にパラメータを入力する方法であり, 後者はシステムからの入力内容の問い合わせに対して解析者が逐次的に応答する方法である。マニュアルが手元になくても解

析を進めていくことができるという点で、我々は応答方式を採用した。

(2) グラフィック機能のサポート

解析結果をグラフ表現することは、結果を視覚的にとらえることができ、その理解を助けることができる。一方、近年多くのパソコンにグラフィック機能が搭載され、BASIC 言語を用いて容易に操作することができる。そこで、グラフィック機能を用いたグラフ解析法をサポートした。特に散布図関係では、散布図行列やマルチ SCAT 等も用意し、ケースラベルの有無や軸のタイプも指定できるようにオプションを豊富に準備した。また、図形出力用周辺機器として X-Y プロッターも普及してきたので、これへの出力も少量のプログラム修正で対応できるように考慮した。

(3) 共通モジュールによる構成

解析手法はそれぞれ固有の目的のために開発・利用されているが、それらを処理内容別に細分化してみると基本的な機能が多くの解析手法で共通に利用されていることがわかる。そこで我々は、この基本機能、例えば行列演算、確率計算、データ入出力、グラフィック機能等を共通モジュールとして開発し、個々の解析手法プログラムは、これらのうち必要なものを選択して利用するようにした。この手続きにより、プログラム管理が整然と行え、また保守性も向上した。

(4) 移植性

BASIC 言語が利用可能なパソコンは市中に多く出回っており、中でもマイクロソフト社の BASIC 言語を搭載したものが大半を占めている。この BASIC 言語の中で機種依存性の少ない命令群を用いてプログラムを作成したので、移植性の高いシステムとすることができた。また前項で述べたように、共通モジュールによる構成となっているので、機種に依存するモジュールのみを修正することで移植が容易に行えるように配慮した。

(5) データマネージメント機能

解析を進めていく過程において、対象データを逐次的に編集・加工する必要が生じてくる。この要求に対して会話的に操作ができるデータマネージメント機能をサポートした。加えて、変数変換や抽出に関しては簡単な言語体系を持った命令群を用いてプログラムを記述することにより一括して処理が行え、また、そのプログラム自身も保存できるようにしたので、典型的な作業にも能率的に対処できる。一方、パソコンには多くのアプリケーションプログラムが流通しているので、MULTIPLAN, dBASE, LOTUS 1-2-3, JUSE-QCAS といった代表的なソフトウェアのシステムファイルやテキスト形式ファイルを Star/B に取り込んで利用できるようにした。

4. ハードウェア構成

前章で述べた機能を十分発揮するために、図1に示したハードウェア構成によるパソコンを対象とした。これらの条件は、高性能・低価格化する16ビットパソコンの標準的な基本構成であり、機種を限定するほどの制約とはならない。また、グラフィック機能は、高解像度ディスプレイのみならずX-Yプロッターでも利用可能なように配慮した。

5. Star/Bの構造

Star/Bは、スーパーバイザー部と機能部の2つの部分から構成される。スーパーバイザー部は、メニュー方式による機能の選択や指定された機能の呼び出しといった処理行程を管理するプログラムである。機能部は統計解析手法の計算処理や対象データの編集・加工といった種々の処理を行うプログラムの集合体である。

スーパーバイザー部と機能部とを交互に呼び出し、各々の処理を行うように設計してあるので、利用者は会話的に目的の処理を行うことができる。この様子を模式的に図2に示す。

また、この方法を採用したことにより、新しい機能を追加する場合にもスーパーバイザーに登録するだけで利用可能となり容易に新機能の追加が行える。

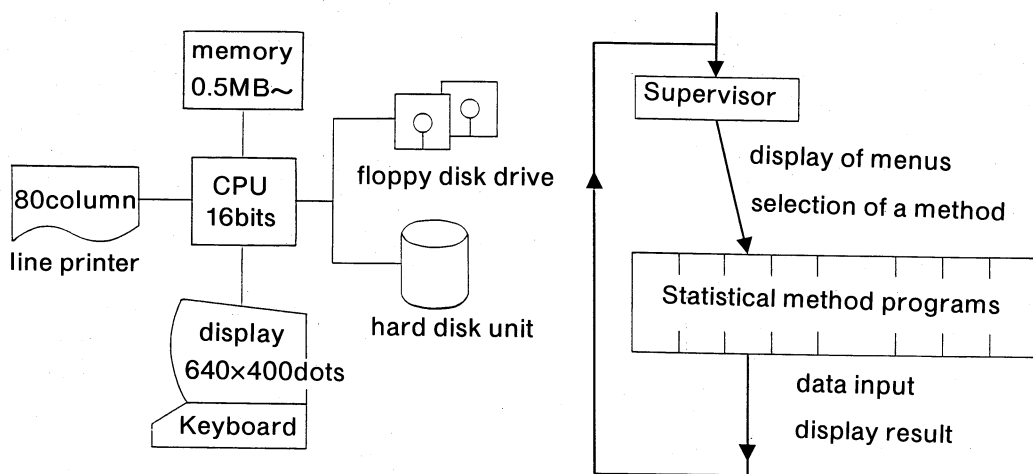


図1. ハードウェア構成

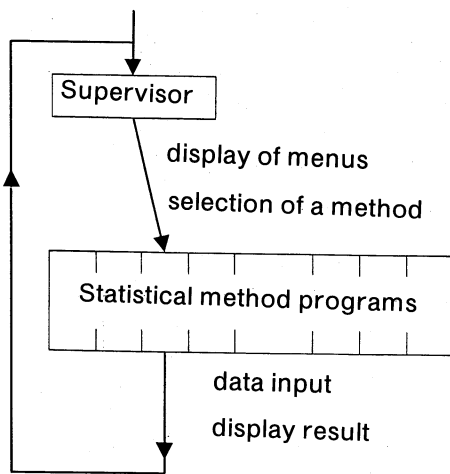


図2. 制御の流れ

6. Star/Bの機能部

前章で述べたようにStar/Bの機能部はスーパーバイザーから見ると種々の機能プログラムの集まりとして捉えることができるが、その作業内容からは解析手法群とデータマネジメント機能群に分けることができる。解析手法群は以下の(1)~(5)に示すように、入門的なものから応用的なものまで広範な統計解析手法を含んでいるので、多くの解析現場での利用が可能であ

る。また、(6)のデータマネージメント機能群は、従来の S.P.P. にみられるような対象データの編集・加工といった機能以外に、第3章で述べたようなパソコンならではのデータマネージメント機能も含んでいる。

Star/B の機能部は総計で56本、約60,000行のプログラムである。具体的な手法については、付表Aを参照されたい。

- | | |
|-----------------|----------------------|
| (1)記述統計 [6本] | (4)ノンパラメトリック法 [14本] |
| (2)グラフ解析法 [10本] | (5)多変量解析 [15本] |
| (3)推定・検定 [4本] | (6)データマネージメント機能 [7本] |

7. ま と め

我々は、機種独立性が高くパソコンの動作特性を生かした統計プログラムパッケージ Star/B を開発した。このシステムは多くの解析手法を含んでおり、グラフィック機能をサポートしているばかりでなく、会話的に対象データを編集・加工・抽出できるデータマネージメント機能を有しているため広範な分野での利用が可能である。

より多くの方々の参考にしていただきたいので、Star/B においても『パソコン統計解析ハンドブック』と同様にプログラムリストを公開する。また、プログラム中には注意を払っていてもミス（バグ）が残るので、利用者に迅速かつ正確にバグの存在と修正方法を伝えることが望ましいと考える。我々は、この目的のためにニュースレターを発行するだけでなく、BBS等のネットワークを利用する事を考慮中である。

今後の課題としては、以下に示す2項目が挙げられる。

① 手法の増強・追加

Star/B では容易に新機能を登録できるので、現在利用可能な解析手法を機能強化するのはもちろんのこと、実験計画法等の新しい手法を追加していく。

② 多国語仕様

パソコンの所有台数から判断するとアメリカ合衆国をはじめ英語圏が圧倒的に多いため、当面 Star/B も英語版とした。しかし、将来的には日本語版のみならず多国語仕様としたいので、個々のメッセージに番号を付け、メッセージを番号で管理する方向に持っていく。

参 考 文 献

- 1) R.A.Becker, J.M.Chambers, S - An Interactive Environment for Data Analysis and Graphics -, Wadsworth, 1984, (邦訳：渋谷, 柴田訳, Sシステム, 共立出版, 1987)
- 2) 脇本和昌, 垂水共之, 田中豊, パソコン統計解析ハンドブック I 基礎統計編, II 多変量解析編, 共立出版, 1984.
- 3) 浅野長一郎ほか, 新・会話型統計解析 (NISAN) システムの機能拡張と実用化の研究開発, 昭和57年度科学研究費補助金試験研究(1)研究成果報告書, 1983.
- 4) 竹内啓監修, 市川伸一, 大橋靖雄著, SAS によるデータ解析入門, 東京大学出版会, 1987.

付 表 A.

The Functions and contents of Star/B

(0) Main menu

Statistical package for personal computer written in BASIC (Star/B)

- Simple Statistics
- Graphical Methods
- Test and Estimation
- Nonparametric Test
- Multivariate Analysis
- Utilities
- MS-DOS Command
- End (return to BASIC system)

(1) Simple statistics

BSTAT	Basic statistics
BOX	Box and whisker chart
SREG	Scattergram and regression line
RCOR	Spearman's and Kendall's rank correlations
SMOOTH	Smoothing
CROSS	Crosstabulation

(2) Graphical methods

SCAT	Scatter plot
SCATMAT	Scatter plot matrix
MSCAT	Multiple Scatter plot
RADAR	Radar chart
ANDRE	Andrews' plot
CONSTE	Constellation graph
MFACE	Face graph
LVECT	Linked vector graph
PQnq	Probability plot (Normal)
PQmul	Probability plot (Multivariate normal)

(3) Estimation and testing methods

ESTEST	Estimation and testing for mean and variance
TMEAND	Test for equality between two means
TFIT	Test of fit
TINDEP	Test of independence in a contingency table

(4) Nonparametric test

WILCOX	Wilcoxon test
MANNWH	Mann-Whitney test
MEDIAN	Median test
WAERDE	van der Waerden test
SIGN	Sign test
SWILCO	Wilcoxon signed rank test
SPEARM	Spearman's rank correlation test
KENDAL	Kendall's rank correlation test
KRUSK	Kruskal-Wallis test
JONCKH	Jonckheele test
FRIEDM	Friedman test
PAGE	Page test
CONCOR	Concordance test
K-MULT	k-Multiple chart test

(5) Multivariate statistical analysis

MREG1	Multiple regression analysis
MREG	Multiple regression analysis with variable selection and regression diagnostics
DISC12	Linear discriminant analysis & quadratic discriminant analysis (Two groups)
DISC34	Linear discriminant analysis with variable selection & canonical discriminant analysis (Several groups)
PCA	Principal component analysis
CANCOR	Canonical correlation analysis
FACTOR	Factor analysis
CLUST	Cluster analysis
ASSOC	Association measures of contingency table
QUANT1	Hayashi's first method of quantification
QUANT2	Hayashi's second method of quantification
QUANT3	Hayashi's third method of quantification
QUANT4	Hayashi's fourth method of quantification
PCOORD	Principal co-ordinates analysis
BIPLOT	Biplot

(6) Utilities

Input/Edit	Data entry and editing
Merge	Case & variable merge of system file
Transpose	Transpose system file
Transformation/Selection	Data transformation and selection
Convert 1	Converter to system file from data statement
Convert 2	Converter to system file from text file
Import	Converter to system file from other system file