

マイクロ・コンピューターによる 統計処理プログラム

吉 田 道 雄

Statistical Program Package by Microcomputer

Michio YOSHIDA

1. マイコンによる統計処理プログラムの存在意義

文科系・理科系をとわず、実験や調査によって得られたデータを、統計的に処理する必要性が増大している。その際、コンピューターを利用するのが最も効率のよい方法であることは、いまや多くのものが認めるところである。しかし、大は電算機センターなどに設置されている大型のコンピューターから、小はこどものゲームに使用されているマイコンに至るまで、ある程度自由に使いこなすためには、プログラムに関する知識が必要である。いかに性能が向上したとはいえ、処理したい課題を頭におもえがいているだけでは、コンピューターは少しも動かないのである。

そこで、コンピューターを利用しようとするれば、まずプログラムを学習しなければならない。ところが、近年きわめて簡単なプログラムが開発されてきてはいるが、それでも知っておかなければならない規則や命令は必ずしも少ないとはいえない。これらを学習し、一応満足のゆく処理が行なえるようになるまでには、ある程度の時間が必要である。また、得られた結果さえ正確であればそれで満足だというのであれば問題は無いが、資料として保存するので表としてみやすくプリントさせたい、などといった欲求をもちはじめると、さらにもう一つつこんだ学習と多くの試行錯誤を経験することになる。個人的にコンピューターに関心をもっているものや、いつもデータ処理に追われているものであれば、ある程度の時間を費やしても問題はないだろう。しかし、統計処理のために多くの時間をとる余裕のないものや、データを処理することが年に何回もないというものにとっては、プログラムの知識がなくても希望する統計処理が行なえる方法があればきわめて有用であろう。

そこで筆者は、一般に利用されることが多いと思われる統計について、最低限の知識で必要とする処理が行なえるマイコン用のプログラムを作成してゆくことにした。本稿では、現在すでに利用可能なプログラムについての解説を行なう。なお使用言語はすべてN-BASICである。

ところで、主として大型コンピューターを利用する統計パッケージがいくつか作成されている。その一つであるSPSS(Statistical Package for Social Sciences)は、約束に従ってパンチした少数のカードをデータカードと共に入力すると、多様な処理を行なうことができる。SPSSはその名からも察しがつくように、当初は社会科学者を対象にして開発されたものであるが、利用者は理科系にもわたっており、広く活用されている。このような大がかりなものと比較すれば、マイコンによる統計処理はきわめて制限の多い小規模なものにならざるをえない。しかしながら、SPSSは大型コンピューターあるいはその端末が設置されていなければ利用することができない(中型機用のミニ版もあるが、中型といってもマイコンに比べればはるかに大きい)。また約束に

従ってカードをパンチすればよいといっても、相当の事前学習が必要である(利用解説書は2分冊になっており、合計すると500ページをこえる)。要するにSPSSはある程度コンピューターを使いたないものために用意されているのであって、コンピューターに関する知識をもたないものがいきなり利用できるわけではない。これに対してマイコンは電卓とまではゆかないまでも、価格的にも入手しやすく、操作方法もきわめて容易なため、現在急速に普及しつつある。その意味でマイコンは大型コンピューターに比べれば、はるかに身近な存在である。もちろん記憶容量や演算スピードなどは大型コンピューターの比ではないが、大型コンピューターは多くのものと共同で利用することが多いため、入力するまでに待時間があったり、ミスの訂正に意外に時間をとられることがある。その点マイコンはミスなどの修正がきわめて速く簡単で、またディスクを使用すれば思ったよりも多量の情報を記録することが可能である。従って、データの入力からミスなどの修正を経て目的の結果を得るまでのトータルな時間は、大型コンピューターを利用したときとかわらない場合すらありうるのである。このような点を考慮すると、マイコンを利用する統計処理プログラム作成の意義は充分にあるといえる。

2. マイコン・システムの構成

作成されたプログラムの実行にあたっては、現在熊本大学教育学部附属教育工学センターに設置されているPC-8000シリーズ(日本電気㈱)を使用する。以下システムを構成する装置の概略を述べる。

(1) 本体(PC-8001)

8ビットのマイクロ・コンピューターで、マスクROM24Kバイト、ダイナミックRAM32Kバイトを実装している。

(2) カラーディスプレイ(PC-8043)

12インチで7色の表示ができる。画面は最大で1行あたり80字、25行まで。

(3) ドットインパクトプリンター(PC-8021)

1行最大80字で印字速度は125字/秒である。

この他に1行132字まで印字できるプリンター(EPSON MP-80)も1台設置されているが、作成されたプログラムはどちらにも利用できる。

(4) ミニディスクユニット(PC-8031)

ミニフロッピーディスク用のユニットで、1ドライブあたり143Kバイトで2台実装されている。この装置を用いることにより、ミニフロッピーディスク1枚あたり約120,000文字のデータを記録できる(これは80欄のパンチカード1,500枚に相当する)。

(5) ミニディスク拡張ユニット(PC-8032)

ミニディスクユニットの拡張のためのユニットで、これを接続することにより最大4ドライブまで使用できる。ただし今回作成したプログラムでは一部の例外を除いて原則として使用しない。

3. プログラムの概要

現在すでに利用可能なプログラムのリストを表1に示す。互いに関連した処理を行なうプログラムは一括してグループにまとめている。プログラムを利用する際には、原則としてプログラム名のみをコンピューターに指定すればよい。その後はコンピューターが指定された統計処理を実

表1. プログラム・リスト

グループ名	プログラム名	処 理 内 容
B A S I C	RWDISK	ディスクにデータを書きこむ(ランダムファイル)
	SRCONV	シーケンシャルファイルをランダムファイルに変換する
	PRDATA	データをプリントする(ランダムファイル)
	BASC	平均, SD, 歪度, 尖度, 最大値, 最小値, レンジ, メディアンを計算する
	DBASC	BASC にデータのプリントを追加する
	BASCZT	平均, SD, 歪度, 尖度, 最大値, 最小値, レンジ, メディアン, z-得点, T-得点を計算する。
	BASALL	BASCZT にデータのプリントを追加する
C O R R M T	WDDISK	ディスクにデータを書きこむ(シーケンシャルファイル)
	RSCONV	ランダムファイルをシーケンシャルファイルに変換する
	PDATA	データをプリントする(シーケンシャルファイル)
	MEAN	平均, SD を計算する
	DMEAN	MEAN にデータのプリントを追加する
	COR	相関係数を計算する
	DCOR	COR にデータのプリントを追加する
	MCOR	平均, SD, 相関係数を計算する
	DMCOR	MCOR にデータのプリントを追加する
	MT	平均, SD, t 値を計算する
	MTT	平均, SD, 2 個の t 値(独立・対応の場合)を計算する
	MCTT	平均, SD, 2 個の t 値(独立・対応の場合), 相関係数を計算する
	CORALL	MCTT にデータのプリントを追加する
F A C T O R	WDDISK	} CORR と同じ
	RSCONV	
	PDATA	
	FPRIN	主因子分析を行なう
	FVMAX	主因子分析の負荷量を使用してバリマックス回転を行なう
	FPVMAX	主因子分析, バリマックス回転を連続して行なう
C A T E G O R Y	CWDISK	ディスクに CATEGORY 用のデータを書きこむ(ランダムファイル)
	SRCONV	} BASIC と同じ
	PRDATA	
	HIST	変数のヒストグラムをプリントし, 平均, SD を計算する
	DHIST	HIST にデータのプリントを追加する
	CROSS	指定した2変数のクロス表をプリントし, χ^2 値, ピアソンの C 係数を計算する
	DCROSS	CROSS にデータのプリントを追加する
	SPEAR	スピアマンの順位相関係数 (ρ) を計算する(直接データを入力する)
KENDAL	ケンドールの一致係数 (W) を計算する(直接データを入力する)	
S U B F I L E	SUBDISK	特定の変数をキーにして, サブファイルを作りディスクに書きこむ
	DSUBDISK	SUBDISK にデータのプリントを追加する
	SUBMEAN	特定の変数をキーにして, サブファイルを作り, 各ファイルの平均, SD を計算する

行する。もちろん、プログラム名の入力以前に簡単な操作が必要であるが、その詳細については「マイコンによる統計処理プログラム利用の手引き」(熊本大学教育学部附属教育工学センター1980)を参照されたい。

以下、表1の順に各プログラムの内容について説明する。実行例を提示するために、表2にあげたサンプルデータを使用する。

表2 サンプルデータ

	国語	数学	理科	社会	英語	技術	美術	保健体育
1	78	95	73	60	75	69	78	73
2	75	79	75	84	81	69	85	95
3	76	73	74	71	83	90	77	88
4	53	56	76	68	69	79	67	60
5	75	68	78	85	75	72	80	79
6	73	65	80	74	75	66	73	76
7	79	76	77	69	74	50	90	80
8	68	62	58	45	58	60	74	70
9	68	60	58	52	54	64	52	52
10	52	64	56	66	54	52	80	45
11	64	52	45	53	60	59	70	76
12	82	51	62	80	50	58	52	50
13	62	75	67	75	60	75	70	80
14	65	75	75	80	70	75	70	80
15	62	70	70	80	75	80	75	80
16	70	75	67	75	72	80	65	75
17	75	75	65	75	70	85	75	80
18	60	75	65	70	60	80	70	80
19	60	75	65	70	60	80	70	80
20	65	75	90	80	85	85	80	95
21	75	75	70	80	69	70	80	75
22	77	72	75	72	75	65	70	70
23	60	80	75	70	72	80	67	70
24	67	85	70	80	64	75	65	75
25	65	80	60	75	75	80	75	80

(1) BASICグループ

このグループに属するプログラムは平均、SD、歪度、尖度、最大値、最小値、レンジ、メジアンといった基本的統計処理を行なう。

① RWDISK

作成されたプログラムはディスクにあるデータファイルからデータを読みとって、目的の統計処理を実行する。従って、それらのプログラムを実行させる前にディスクにデータを記録しておかなければならない。この作業を行なうのがRWDISKである。その点で後述するWDDISK、CWDISKも同様な目的をもったプログラムで、他の統計処理プログラムとは根本的に異なっている。RWDISKはディスクにランダムファイルを開いてデータを記録する。これはBASICグループに属するプログラムがランダムファイルを使用するからである。RWDISKはDATA文に書かれたデータをREAD文で読みこむ形式になっている。従って利用者がDATA文の部分を作成しRWDISKに付け加えなければならない。そのためにMERGEコマンドを用いる。このコマンドによって、本体にあるプログラムとディスク上のプログラムを結合することができる。具体的にいうと、RWDISKは行番号が34までのプログラムであるから、まず35以上の行番号をふってDATA文を作成する。そしてすべてのデータを入力し終えたらMERGEコマンドでディスクからRWDISKを呼び出す。これでRWDISKとDATA文の結合が行なわれ、実行が可能になる。そこでRUNさせると、

```
INPUT N OF SS (MAX=470) & VARIABLES (MAX=60) ?
```

と問いあわせてくる。キーボードから個体数と変数の数を入力すれば、ディスクにデータが記録される。なお個体数は最大470に、変数の数は最大60に制限している。これはあまりもデータ数が多くなると、マイコンの能力上多くの時間を要するので、一応の目安を定めているにすぎない。従って、利用目的に応じて変更することは可能である。他のグループについても同じことがいえる。

さて、データの書きこみが終ると、

INPUT VARIABLE NAMES ? YES=1, NO=2 ?

と問いあわせてくる。変数名を入れるならば1を入力し、順に名前をキーボードから入れてゆく。名前は10文字以内であれば、英数字、カタカナ、記号などなんでもよい。こうしてRWDISKの実行が終る。

② SRCONV

このプログラムはシーケンシャルファイルとして記録されているデータをランダムファイルに変換するものである。先にも述べたように、BASICグループではランダムファイルを使用するが、他のグループにはシーケンシャルファイルを用いるものがある。それらを実行するために作られたシーケンシャルファイルはそのままではBASICグループのプログラムに使用できない。

VARIABLE LIST

1)コゴ*	2)スガ*?	3)カ	4)シヤカ
5)エゴ*	6)キ*ジ*ユ?	7)ヒ*ジ*ユ?	8)キ*ジ*ユ*ユ?

DATA TABLE (N= 25 ,N OF VARIABLES= 8)

1)	78	95	73	60	75	69	78	73
2)	75	79	75	84	81	69	85	95
3)	76	73	74	71	83	90	77	88
4)	53	56	76	68	69	79	67	60
5)	75	68	78	85	75	72	80	79
6)	73	65	80	74	75	66	73	76
7)	79	76	77	69	74	50	90	80
8)	68	62	58	45	58	60	74	70
9)	68	60	58	52	54	64	52	52
10)	52	64	56	66	54	52	80	45
11)	64	52	45	53	60	59	70	76
12)	82	51	62	80	50	58	52	50
13)	62	75	67	75	60	75	70	80
14)	65	75	75	80	70	75	70	80
15)	62	70	70	80	75	80	75	80
16)	70	75	67	75	72	80	65	75
17)	75	75	65	75	70	85	75	80
18)	60	75	65	70	60	80	70	80
19)	60	75	65	70	60	80	70	80
20)	65	75	90	80	85	85	80	95
21)	75	75	70	80	69	70	80	75
22)	77	72	75	72	75	65	70	70
23)	60	80	75	70	72	80	67	70
24)	67	85	70	80	64	75	65	75
25)	65	80	60	75	75	80	75	80

END OF *PRDATA* (TIME=00:00:50)

図1 PRDATAの出力例

RWDISKでデータを記録しなおせばよいが、内容は同じデータなのだから二度手間になる。そこでこのSRCONVを使うと、自動的にランダムファイルへの変換が行なわれる。ちなみにサンプルデータ(データ数 $25 \times 8 = 200$, 変数名8)で実行させたところ、28秒で変換が終了した。

なお、変換はドライブ2に入っているディスクのシーケンシャルファイルを読みとり、その同じディスクにランダムファイルを作ることによって行なわれる。もしデータ数が多くて1枚のディスクに追加できない場合(一応個体数で200を目安にしている)には、自動的にドライブ3を使用し、まったく新しいディスクにランダムファイルを作る。このときにはミニディスク拡張ユニット(PC-8032)が用いられる。

③ PRDATA

このプログラムも実質的な統計処理は行なわない。RWDISKによってディスクに書きこまれたランダムファイルのデータを単純にプリントアウトする。図1はサンプルデータを使用した例

VARIABLE LIST							
1)コク*	2)スカ*?	3)リカ	4)シカ+	5)イク*	6)キ*シ*?	7)ヒ*シ*?	8)キ*カ*?
VARIABLE 1 =コク*							
MEAN	68.240	SD	7.936	SKEWNESS	-0.232	KURTOSIS	2.195
MAXIMUM	82.000	MINIMUM	52.000	RANGE	30.000	MEDIAN	68.000
VARIABLE 2 =スカ*?							
MEAN	71.520	SD	9.864	SKEWNESS	-0.229	KURTOSIS	3.226
MAXIMUM	95.000	MINIMUM	51.000	RANGE	44.000	MEDIAN	75.000
VARIABLE 3 =リカ							
MEAN	69.040	SD	9.202	SKEWNESS	-0.359	KURTOSIS	3.507
MAXIMUM	90.000	MINIMUM	45.000	RANGE	45.000	MEDIAN	70.000
VARIABLE 4 =シカ+							
MEAN	71.560	SD	9.880	SKEWNESS	-1.089	KURTOSIS	3.650
MAXIMUM	85.000	MINIMUM	45.000	RANGE	40.000	MEDIAN	74.000
VARIABLE 5 =イク*							
MEAN	68.600	SD	9.239	SKEWNESS	-0.273	KURTOSIS	2.172
MAXIMUM	85.000	MINIMUM	50.000	RANGE	35.000	MEDIAN	70.000
VARIABLE 6 =キ*シ*?							
MEAN	71.920	SD	10.365	SKEWNESS	-0.434	KURTOSIS	2.340
MAXIMUM	90.000	MINIMUM	50.000	RANGE	40.000	MEDIAN	75.000
VARIABLE 7 =ヒ*シ*?							
MEAN	72.400	SD	8.509	SKEWNESS	-0.562	KURTOSIS	3.742
MAXIMUM	90.000	MINIMUM	52.000	RANGE	38.000	MEDIAN	73.000
VARIABLE 8 =キ*カ*?							
MEAN	74.560	SD	11.900	SKEWNESS	-0.822	KURTOSIS	3.634
MAXIMUM	95.000	MINIMUM	45.000	RANGE	50.000	MEDIAN	76.000

END OF *BASC*(TIME=00:04:10)

図2 BASICの出力例

である。50秒かかってプリントアウトしていることがわかる。

④ BASC

Z-SCORE (N= 25 ,N OF VARIABLES= 8)

1)	1.230	2.380	0.430	-1.170	0.693	-0.282	0.658	-0.131
2)	0.852	0.758	0.648	1.259	1.342	-0.282	1.481	1.718
3)	0.978	0.150	0.539	-0.057	1.559	1.744	0.541	1.129
4)	-1.920	-1.573	0.756	-0.360	0.043	0.683	-0.635	-1.224
5)	0.852	-0.357	0.974	1.360	0.693	0.008	0.893	0.373
6)	0.600	-0.661	1.191	0.247	0.693	-0.571	0.071	0.121
7)	1.356	0.454	0.865	-0.259	0.584	-2.115	2.068	0.457
8)	-0.030	-0.965	-1.200	-2.688	-1.147	-1.150	0.188	-0.383
9)	-0.030	-1.168	-1.200	-1.980	-1.580	-0.764	-2.398	-1.896
10)	-2.046	-0.762	-1.417	-0.563	-1.580	-1.922	0.893	-2.484
11)	-0.534	-1.979	-2.612	-1.879	-0.931	-1.246	-0.282	0.121
12)	1.734	-2.080	-0.765	0.854	-2.013	-1.343	-2.398	-2.064
13)	-0.786	0.353	-0.222	0.348	-0.931	0.297	-0.282	0.457
14)	-0.408	0.353	0.648	0.854	0.152	0.297	-0.282	0.457
15)	-0.786	-0.154	0.104	0.854	0.693	0.780	0.306	0.457
16)	0.222	0.353	-0.222	0.348	0.368	0.780	-0.870	0.037
17)	0.852	0.353	-0.439	0.348	0.152	1.262	0.306	0.457
18)	-1.038	0.353	-0.439	-0.158	-0.931	0.780	-0.282	0.457
19)	-1.038	0.353	-0.439	-0.158	-0.931	0.780	-0.282	0.457
20)	-0.408	0.353	2.278	0.854	1.775	1.262	0.893	1.718
21)	0.852	0.353	0.104	0.854	0.043	-0.185	0.893	0.037
22)	1.104	0.049	0.648	0.045	0.693	-0.668	-0.282	-0.383
23)	-1.038	0.860	0.648	-0.158	0.368	0.780	-0.635	-0.383
24)	-0.156	1.367	0.104	0.854	-0.498	0.297	-0.870	0.037
25)	-0.408	0.860	-0.982	0.348	0.693	0.780	0.306	0.457

図3 BASCZTによるz-得点の出力例

T-SCORE (N= 25 ,N OF VARIABLES= 8)

1)	62.298	73.804	54.303	38.299	56.927	47.183	56.581	48.689
2)	58.518	57.583	56.477	62.591	63.421	47.183	64.808	67.177
3)	59.778	51.500	55.390	49.433	65.586	67.443	55.406	61.294
4)	30.797	34.266	57.563	46.397	50.433	56.831	43.654	37.765
5)	58.518	46.431	59.737	63.603	56.927	50.077	58.932	53.731
6)	55.998	43.390	61.910	52.470	56.927	44.289	50.705	51.210
7)	63.558	54.542	58.650	47.409	55.845	28.852	70.685	54.572
8)	49.698	40.348	38.003	23.117	38.527	38.500	51.880	46.168
9)	49.698	38.321	38.003	30.202	34.198	42.359	26.025	31.042
10)	29.537	42.376	35.830	44.372	34.198	30.782	58.932	25.159
11)	44.658	30.210	23.876	31.214	40.692	37.535	47.179	51.210
12)	67.338	29.196	42.350	58.543	29.868	36.570	26.025	29.361
13)	42.138	53.528	47.783	53.482	40.692	52.971	47.179	54.572
14)	45.918	53.528	56.477	58.543	51.515	52.971	47.179	54.572
15)	42.138	48.459	51.043	58.543	56.927	57.795	53.056	54.572
16)	52.218	53.528	47.783	53.482	53.680	57.795	41.303	50.370
17)	58.518	53.528	45.610	53.482	51.515	62.619	53.056	54.572
18)	39.617	53.528	45.610	48.421	40.692	57.795	47.179	54.572
19)	39.617	53.528	45.610	48.421	40.692	57.795	47.179	54.572
20)	45.918	53.528	72.777	58.543	67.751	62.619	58.932	67.177
21)	58.518	53.528	51.043	58.543	50.433	48.148	58.932	50.370
22)	61.038	50.487	56.477	50.445	56.927	43.324	47.179	46.168
23)	39.617	58.597	56.477	48.421	53.680	57.795	43.654	46.168
24)	48.438	63.666	51.043	58.543	45.021	52.971	41.303	50.370
25)	45.918	58.597	40.176	53.482	56.927	57.795	53.056	54.572

図4 BASCZTによるT-得点の出力例

このプログラムから、ディスクに入れられたデータを用いて何らかの統計処理を行なう。BASCを実行して得られる結果は、平均、SD、歪度、尖度、最大値、最小値、レンジ、メディアンである。サンプルデータを使用した例を図2に示す。所要時間は4分10秒である。

⑤ DBASC

BASCで得られる結果にデータのプリントが追加される。サンプルデータで5分2秒かかった。

⑥ BASCZT

BASCの結果に加えて、z-得点、T-得点を算出する。正確には正規分布でない場合T-得点はZ-得点になる。それぞれの得点の出力例を図3、図4に示す。サンプルデータで6分22秒を要した。

⑦ BASALL

BASICグループで最も多くの出力を行なうプログラムで、データのプリントアウトからT得点の算出まで、これまで述べたすべてのものを処理する。サンプルデータでの実行時間は7分1秒であった。

(2)CORRグループ

CORRグループでは主として相関係数やt値の算出を行なう。データはシーケンシャルファイルとして記録されているものを使用する。

① WDDISK

さきに述べたRWDISKに対応するプログラムで、シーケンシャルファイルをオープンし、データをディスクに書きこむ。

② RSCONV

ランダムファイルとして記録されているデータをシーケンシャルファイルに変換する。ちょうどSRCONVと逆の処理を行なう。

③ PDATA

シーケンシャルファイルとして記録されているデータをプリントアウトする。PRDATAと同じ働きをするプログラムである。

WDDISK, RSCONV, PDATAの処理時間はRWDISK, SRCONV, PRDATAとほとんどかわらない。

④ MEAN

平均値、SDのみを計算するプログラムである。サンプルデータによる実行結果を図5に示す。出力までに29秒かかっている。

⑤ DMEAN

MEANにデータのプリントが追加される。サンプルデータで1分11秒を要した。

⑥ COR

相関係数のみを計算する。1分18秒でサンプルデータの処理が終了した(図6)。

⑦ DCOR

CORにデータのプリントが追加される。サンプルデータの所要時間は1分59秒であった。

⑧ MCOR

MEANとCORの計算を一度に行なうプログラムである。サンプルデータで1分37秒を要した。

⑨ DMCOR

MCORにデータのプリントが追加される。2分20秒でサンプルデータの処理を終えた。

VARIABLE	MEAN	S D	(N= 25)
1)コゴ*	68.240	7.936	
2)スガ*?	71.520	9.864	
3)リカ	69.040	9.202	
4)ツカカ	71.560	9.880	
5)エイ*	68.600	9.239	
6)キ*ジ*ユ?	71.920	10.365	
7)ヒ*ジ*ユ?	72.400	8.509	
8)キ*ジ*ユ?	74.560	11.900	

END OF *MEAN*(TIME=00:00:29)

図5 MEANの出力例

CORRELATION TABLE (N= 25)

1)コゴ*	1.000								
2)スガ*?	0.153	1.000							
3)リカ	0.233	0.415	1.000						
4)ツカカ	0.189	0.304	0.570	1.000					
5)エイ*	0.293	0.505	0.743	0.443	1.000				
6)キ*ジ*ユ?	-0.200	0.396	0.358	0.388	0.470	1.000			
7)ヒ*ジ*ユ?	0.127	0.413	0.365	0.200	0.596	-0.042	1.000		
8)キ*ジ*ユ?	0.197	0.523	0.471	0.400	0.723	0.535	0.562	1.000	

END OF *COR*(TIME=00:01:18)

図6 CORの出力例

T-VALUES (N= 25 ,DF= 24)

1)コゴ*	0.000								
2)スガ*?	1.269	0.000							
3)リカ	0.323	-0.901	0.000						
4)ツカカ	1.283	0.014	0.914	0.000					
5)エイ*	0.145	-1.058	-0.165	-1.072	0.000				
6)キ*ジ*ユ?	1.381	0.137	1.018	0.123	1.171	0.000			
7)ヒ*ジ*ユ?	1.752	0.331	1.313	0.316	1.482	0.175	0.000		
8)キ*ジ*ユ?	2.165	0.964	1.798	0.950	1.938	0.820	0.723	0.000	

END OF *MT*(TIME=00:00:53)

図7 MTのうちt値の出力例

⑩ MT

平均値, SD, および独立した平均値の差の検定を行なう場合に利用される t 値の計算を行なう。図7はサンプルデータを使用したときの出力のうち、 t 値の部分を示している。所要時間は53秒である。

⑪ MTT

MTで計算されるものの他に、対応のある場合の t 値をも計算する。追加される t 値の部分の出力例を図8に示す。2分12秒で処理している。

```
T-VALUES WITH CORRELATION (N= 25 ,DF= 24 )

1)コゴ*          0.000
2)ヌカ*?        0.161  0.000
3)リカ          0.040 -0.121  0.000
4)ツツカI       0.163  0.002  0.123  0.000
5)イイ*         0.018 -0.143 -0.022 -0.145  0.000
6)チ*シ*ユ*     0.180  0.019  0.140  0.017  0.162  0.000
7)ヒ*ツ*ユ*     0.204  0.042  0.163  0.040  0.185  0.023  0.000
8)チ*ツ*ユ*     0.304  0.143  0.264  0.141  0.286  0.123  0.101  0.000

END OF *MTT*(TIME=00:02:12)
```

図8 MTTの出力例(対応のある場合の t 値のみ)

⑫ MCTT

MTTに相関係数の処理を追加したプログラムである。サンプルデータを処理するのに2分36秒を要した。

⑬ CORALL

MCTTにデータのプリントアウトを追加する。要するにこれまで述べたCORグループのすべての処理を行なう。サンプルデータの処理に3分26秒かかった。

(3)FACTORグループ

このグループのプログラムは主因子分析、およびその結果にもとづいたバリマックス解を求める。表1のはじめから3つのプログラム(WDDISK, RSCONV, PDATA)はCORRグループと同一のものである。

① FPRIN

主因子分析を行なう。平均, SD, 相関係数もプリントアウトする。RUNさせると、
INPUT N OF FACTORS (MAX=10)?

と問いあわせてくるので因子数を入力する。ここでわかるように因子数は最大で10となっている。その出力例を図9に示す。サンプルデータは変数の数に比べて人数が少ないので、さらに15名分のデータを追加して処理を行なった結果である。処理時間は因子数やデータの性質に大きく影響を受けるので、データ数だけで一般的にどのくらいかかるかはわからない。このことは以下に述べるFVMAX, FPMAXについても同様である。

なお、このプログラムを作成するにあたって芝(1972)を参考にした。

② FVMAX

FPRINで得られた負荷量を使用して、バリマックス解を求める。FPRINを実行させると、

PRINCIPAL FACTOR SOLUTION (BY POWER METHOD)

N OF SS= 40, N OF VARIABLES= 8 (REP=500, EPS=0.0001)

VARIABLES	FACTOR LOAD. (FACTOR=3)		
1) コワコ	0.408	-0.698	0.118
2) スカカ	0.556	0.354	0.623
3) リカ	0.036	0.594	0.139
4) シカカ	-0.196	-0.726	-0.407
5) イカ	0.195	0.231	0.199
6) キシユ	0.614	0.441	0.659
7) ヒシユ	0.692	0.101	0.661
8) カカカカ	-0.105	0.079	-0.129
ITERATION	2	8	3
CONTRIBUT	1.421	0.910	0.204

図9 FPRIN の出力例

VARIMAX FACTOR SOLUTION

N OF SS= 40, N OF VARIABLES= 8 (REP=500, EPS=0.0001)

VARIABLES	FACTOR LOAD. (FACTOR=3)		
1) コワコ	0.076	-0.814	-0.032
2) スカカ	0.905	0.017	0.058
3) リカ	0.343	0.505	-0.030
4) シカカ	-0.671	-0.527	-0.048
5) イカ	0.347	0.105	-0.015
6) キシユ	0.998	0.067	0.032
7) ヒシユ	0.922	-0.273	0.050
8) カカカカ	-0.122	0.129	-0.047
CONTRIBUT	3.378	1.302	0.014

図10 FVMAX の出力例

FVMAXが必要とする情報がディスクに記録される。FVMAXはこの情報を読みこんで処理を行なう。図10はその出力例である。作成にあたっては塗師(1979)を参考にした。

③ FVMAX

主因子分析を行ない、そのまま引き続いてバリマックス解をも算出する。従って、最初に因子数を指定するだけで、自動的にバリマックス解まで出力する。このプログラムがあればFPRIN, FVMAXは不要だと思われるかもしれない。しかし、実際にやってみるとわかるが、主因子分析だけで相当な時間を要することがある。そのような場合には、一担FPRINのみを行ない、あらためてFVMAXを実行させればよい。そうすれば、例えば時間の余裕がないようなときなど、2日にわたって分析するといったことが可能になる。また当然のことながら、バリマックス解が不要なときにはFPRINだけですませればよい。

(4) CATEGORY グループ

このグループでは、いわゆるカテゴリーデータを分析の対象とし、ヒストグラムやクロス表を作成したり、 χ^2 値の算出などを行なう。

① CWDISK

基本的にはRWDISKと同じもので、ディスクにランダムファイルをオープンし、データを書きこむプログラムである。ただこのグループの分析には各変数のカテゴリー数がわかっていなければならないので、その情報もディスクに記録しておく必要がある。そこでCWDISKをRUNさせると、データの記録後に、

N OF CATEGORIES IS SAME THROUGH ALL VARIABLES? YES = 1, NO = 0? と問いあわせてくる。すべての変数のカテゴリー数が同じなら1を入力し、つづいてそのカテゴリー数を入れる。もし各変数のカテゴリー数に違いがあれば0を入力し、その後それぞれのカテゴリー数を入れてゆく。ちなみにサンプルデータ(表2)をもとにして、65以下を1、65以上75以下を2、それ以上を3に分類してカテゴリーデータを作ってみた。その結果を表3に示す。以下このデータをもちいて実行例を示す。なお表1でCWDISKにつづくSRCONV, PRDATAはすでに説明したBASICグループのものと同じのプログラムである。

表3 カテゴリー化したサンプルデータ

	国語	数学	理科	社会	英語	技術	美術	保健体育
1	3	3	2	1	3	2	3	2
2	3	3	3	3	3	2	3	3
3	3	2	2	2	3	3	3	3
4	1	1	3	2	2	3	2	1
5	3	2	3	3	3	2	3	3
6	2	2	3	2	3	2	2	3
7	3	3	3	2	2	1	3	3
8	2	1	1	1	1	1	2	2
9	2	1	1	1	1	1	1	1
10	1	1	1	2	1	1	3	1
11	1	1	1	1	1	1	2	3
12	3	1	1	3	1	1	1	1
13	1	3	2	3	1	3	2	3
14	2	3	3	3	2	3	2	3
15	1	2	2	3	3	3	3	3
16	2	3	2	3	2	3	2	3
17	3	3	2	3	2	3	3	3
18	1	3	2	2	1	3	2	3
19	1	3	2	2	1	3	2	3
20	2	3	3	3	3	3	3	3
21	3	3	2	3	2	2	3	3
22	3	3	3	2	3	2	2	2
23	1	3	3	2	2	3	2	2
24	2	3	2	3	1	3	2	3
25	2	3	1	3	3	3	3	3

② HIST

各変数のヒストグラムをプリントアウトする。同時に各カテゴリーごとの個体数、累計パーセントイジなども表示し、最後に平均、SDを計算して処理が終了する。サンプルデータをもちいて実行した結果の一部を図11に示す。3分1秒で処理をした。

③ DHIST

HISTにデータのプリントアウトを追加する。サンプルデータで3分53秒かかった。

④ CROSS

VARIABLE LIST

1) コゴ	2) スカ	3) リカ	4) シカ	5) イゴ	6) キシ	7) ビシ	8) ナシ
-------	-------	-------	-------	-------	-------	-------	-------

HISTGRAM N OF SS= 25

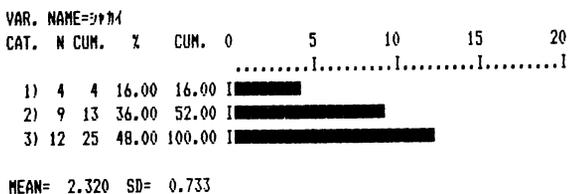
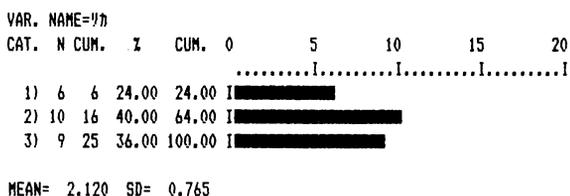
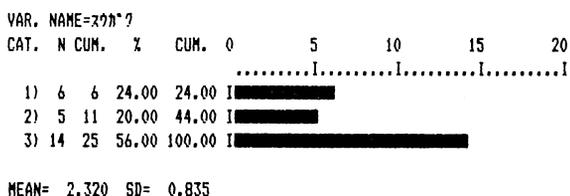
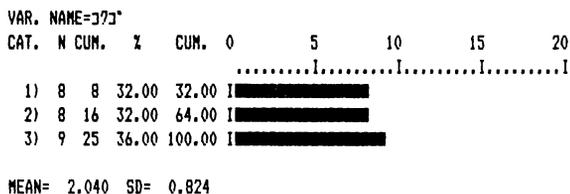


図11 HIST の出力例

2 変数を指定すれば、クロス表を作成し、 χ^2 値およびピアソンの C 係数を求める。RUN させると、

HOW MANY CROSS TABLES DO YOU WANT ? (MAX=200) ?

と問いあわせてくる。そこで作成したいクロス表の数だけ入力する。表は200まで作成可能である。それが終るとつづいて

INPUT NO. OF ROW VAR. & COLUMN VAR. ?

VARIABLE LIST

1)コウゴ	2)スウカク	3)リカ	4)シヨク
5)イイゴ	6)キョウゴ	7)ヒョウゴ	8)キョウゴ

CROSS TABLE (TOTAL N= 25)

コウゴ	(ROW) * スウカク	(COLUMN)			
	1	2	3	SUM	
1	3	1	4	8	
2	2	1	5	8	
3	1	3	5	9	
SUM	6	5	14		

CHI SQUARE= 2.596 (DF= 4)
PEARSON C= 0.307

CROSS TABLE (TOTAL N= 25)

コウゴ	(ROW) * リカ	(COLUMN)			
	1	2	3	SUM	
1	2	4	2	8	
2	3	2	3	8	
3	1	4	4	9	
SUM	6	10	9		

CHI SQUARE= 2.380 (DF= 4)
PEARSON C= 0.295

END OF *CROSS*(TIME=00:00:55)

図12 CROSSの出力例

と表示される。実際にクロスさせるべき変数の組合わせを入力するように要求しているのである。表の行および列に入れるべき変数の番号を入力する。作成すべきクロス表の数だけ問いあわせが繰り返される。サンプルデータを使って国語と数学、理科の組合せを指定して得られた結果が図12である。これはあくまでサンプルであるが、いずれかのセルが5以下の場合には χ^2 値の修正が必要になるので注意を要する。処理時間は55秒になっているが、作成すべき表の数やカテゴリー数によって大きく変動する。

⑤ DCROSS

CROSSにデータのプリントアウトを追加するプログラムで、手続きはCROSSとまったく同じである。

⑥ SPEAR

スピアマンの順位相関係数(ρ)を求める。ディスクからデータを読みこむのではなく、直接キーボードから入力する。個体数の上限は500である。終了を指示するまでは、別のデータを何度入力しても実行を繰り返す。図13はその出力例である。

```

INPUT DATA (N= 12)

1)
 2 6 5 1 10 9 8 3 4 12 7 11

2)
 3 4 2 1 8 11 10 6 7 12 5 9

RHO= 0.818

END OF *SPEAR*(TIME=00:00:34)

```

図13 SPEARの出力例

⑦ KENDAL

ケンドールの一致係数(W)を計算する。SPEARと同様キーボードからデータを直接入力する。最大20×200のデータを処理することができる。このプログラムも終了の指示があるまでは処理を繰り返して行なえる。出力例を図14に示す。

```

INPUT DATA (N= 6)

1) 5 3 4 2 6 1

2) 4 1 5 3 6 2

3) 4 2 5 1 6 3

4) 4 3 5 2 6 1

KENDALL W= 0.871

END OF *KENDAL*(TIME=00:00:33)

```

図14 KENDALの出力例

(5)SUBFILEグループ

一組のデータを特定の変数によって分類する必要にせまられることがある。例えば、男女こみで入っているデータを性別に分けるといった場合である。SUBFILEグループのプログラムはこの分類作業を行なう。

① SUBDISK

特定の変数をキーにして指定した名前のサブファイルを作り、ディスクに書きこむ。RUNさせると、

HOW MANY SUBFILES DO YOU MAKE? (MAX=20) ?

と問いあわせてくるので、ディスクに作成しようとするサブファイルの数を入力する。つづいて、
INPUT NAME OF EACH FILE (WITHIN 6 CHARACTERS) ?

と各ファイル名の入力を要求してくる。6文字以内の名前を入力する。指定したサブファイルの数だけ名前を問いあわせてくる。次に、

INPUT NO. OF KEY VARIABLE ?

と表示される。キー変数の番号を入力すればよい。さらに、

DO YOU WRITE ALL VARIABLES TO DISK ? YES=1, NO=0 ?

と問い返してくる。キー変数も含めてすべての変数を分類してディスクに書きこむのであれば1を、一部の変数のみでよい場合には0を入力する。0と入れたときは分類すべき変数を入力する

FILE NAME	N
1)コワ*1	8
2)コワ*2	8
3)コワ*3	9

END OF *SUBDISK*(TIME=00:01:36)

図 15 SUBDISKによる出力例

SUBFILE=コワ*1

1)	1	1	3	2	2	3	2	1
2)	1	1	1	2	1	1	3	1
3)	1	1	1	1	1	1	2	3
4)	1	3	2	3	1	3	2	3
5)	1	2	2	3	3	3	3	3
6)	1	3	2	2	1	3	2	3
7)	1	3	2	2	1	3	2	3
8)	1	3	3	2	2	3	2	2

SUBFILE=コワ*2

1)	2	2	3	2	3	2	2	3
2)	2	1	1	1	1	1	2	2
3)	2	1	1	1	1	1	1	1
4)	2	3	3	3	2	3	2	3
5)	2	3	2	3	2	3	2	3
6)	2	3	3	3	3	3	3	3
7)	2	3	2	3	1	3	2	3
8)	2	3	1	3	3	3	3	3

SUBFILE=コワ*3

1)	3	3	2	1	3	2	3	2
2)	3	3	3	3	3	2	3	3
3)	3	2	2	2	3	3	3	3
4)	3	2	3	3	3	2	3	3
5)	3	3	3	2	2	1	3	3
6)	3	1	1	3	1	1	1	1
7)	3	3	2	3	2	3	3	3
8)	3	3	2	3	2	2	3	3
9)	3	2	3	2	3	2	2	2

図16 各サブファイルの内容

ように要求してくる。図15はその出力例である。サンプルデータの変数1(国語)をキーにして、コクゴ1, 2, 3という名前をそれぞれつけてサブファイルを作成した。ディスクに記録されたサブファイル名と各サブファイルの個体数がプリントアウトされる。念のため作成された各サブファイルの内容をプリントアウトさせた結果が図16である。国語のカテゴリー1, 2, 3によって、正確に分類されていることがわかる。

② DSUBDISK

これまでのプログラムと同様、SUBDISKにデータのプリントアウトを追加する。

③ SUBMEAN

このプログラムはキー変数によって他の変数を分類し、その後に各サブファイルごとに、分類した変数の平均, SDを求める。ただし、ディスクに対する書きこみは行なわない。サンプルデータの理科をキー変数にして、技術, 美術, 保健体育の分類を指定して処理した結果が図17である。

(SUB リカ1) N= 6		
	MEAN	SD
1) *リカ1	1.333	0.745
2) *リカ2	2.000	0.816
3) *リカ3	1.833	0.898

(SUB リカ2) N= 10		
	MEAN	SD
1) *リカ1	2.800	0.400
2) *リカ2	2.500	0.500
3) *リカ3	2.900	0.300

(SUB リカ3) N= 9		
	MEAN	SD
1) *リカ1	2.333	0.667
2) *リカ2	2.444	0.497
3) *リカ3	2.556	0.685

END OF *SUBMEAN*(TIME=00:00:49)

図17 SUBMEANの出力例

以上でプログラムの概要についての説明を終わる。CORRグループまでは参考のために、比較的詳細に処理時間をあげたが、FACTOR以降はそれをほとんど省略した。それは、先にも述べたようにデータの性質に影響を受けたり、あるいは各種の問あわせに対する回答の入力に個人差があったりなどして、処理時間が一定していないと考えられたからである。

4. ステップ・モニター

最近のマイコンは、その名前から想像される以上の能力をもっている。しかし、実際にデータを入れてプログラムを実行させてみるとわかるが、データの量が多くなると、やはりそれなりの時間を要する。その点では大型コンピューターなどとは比較にならない。理論的にはディスクを取りかえるなどの工夫をすれば、無限大のデータでも取り扱えるはずであるが、それこそ何日、何カ月かかるかわからない。RWDISKのところでも説明したように、データ数を最大470×60にしているのは、実用的な時間内で処理が終るようにと考えたからである。しかし、それでも因子分析などのように繰り返しの多いステップを含んでいるプログラムの実行にあたっては、2時間、3時間かかることもまれではない。コンピューターを外からみていただけでは何の変化もなく、一体どこまで処理が進んでいるのかまったくわからない。結果が出るまでただ漫然と待っているわけにもゆかない。そこで、現在何人までのデータがすでに処理されているとか、何因子まで結果がみいだされている、といった処理中の情報が提供されると便利である。1人あたりのデータ処理時間から全体の所要時間の予測が可能になるし、あるいは途中で様子をみて、すでにどこまで処理が終っているのかを確認することもできる。本稿で紹介したプログラムは、ほとんどの場合この種の情報をディスプレイ上に表示する。これをステップ・モニターと呼んでいる。この情報を有効に活用すれば、結果がでるまでじっとコンピューターの前で待つといったことをしなくてすむのである。

参考文献

- 三宅一郎・山本嘉一郎(1976)：SPPSS統計パッケージ I基礎編 東洋経済新報社。
三宅一郎・中野嘉弘・水野欽司・山本嘉一郎(1977)：SPSS統計パッケージ II解析編 東洋経済新報社。
塗師斌(1979)：コンピュータープログラム 新曜社。
芝祐順(1972)：因子分析法 東京大学出版会。

(1981年5月20日 受理)