みんぱくリポジトリ
国立民族学博物館学術情報リポジトリ National Museum of Ethnolo

A Versatile Software System for Ethnological Data Analysis

メタデータ	言語: jpn 出版者: 公開日: 2010-02-16 キーワード (Ja): キーワード (En): 作成者: 八村, 広三郎 メールアドレス: 所属:
URL	https://doi.org/10.15021/00004526

民族学データの統計的解析のための汎用ソフトウェアシステム

——民具データへの応用を中心として——

八　村　廣三郎*

A Versatile Software System for Ethnological Data Analysis

Kozaburo HACHIMURA

This short note describes a computer system for analyzing data sets acquired in the course of ethnological studies. The system accepts, as an input, any set of data whose individual constituents consist of several attribute items. This type of data set is frequently encountered in studies of material culture. Several metric and/or nominal attributes are defined and measured for all samples under investigation.

The major function of the system for this kind of data set is to depict frequency distribution of attributes. Measurement of the frequency distribution may be conditioned: The measurement may be made on only the samples that meet certain conditions imposed on some attribute items. By examining the attribute frequency distribution conditioned by other attributes, we can estimate inter-relationships among these attributes. Other principal functions of the system are linear regression analysis, correlation analysis and retrieval from the data set. All processing proceeds in man-machine interactive fashion and results are displayed graphically.

The functions and use of the system are illustrated by application to a data collection on Japanese basketry.

Ⅰ．はじめに
Ⅱ．取り扱うデータの形式とシステムの機能
Ⅲ．システムの構成と使用法
Ⅳ．あとがき

* 国立民族学博物館第5研究部

I. はじめに

国立民族学博物館では，中規模の研究用電子計算機システムを設置し，昭和54年1月より稼動を開始している。自然科学系でない，いわゆる文科系の研究分野での電子計算機の利用ということは，もはや目新しい話題とはいえないが，この中で，民族学研究での本格的な利用に関していえば，必ずしもそのスタートが早かった訳ではないようである。

そのひとつの理由には，民族学で取扱う情報には，数値化あるいは記号化のなされていない，生の，アナログ的性格の情報がきわめて多く，電子計算機の解析に直接のせるのが困難であったという事情があると考えられる。ところが，近年では，電子計算機の利用がさまざまな分野で盛んになり，それにともなって，計算機で取り扱うのが困難とされていた情報（たとえば，図形，画像などのパターン情報）の解析も次第に試みられるようになっている。民博でも，これらのアナログ情報，たとえば画像，音声，音楽などの情報を処理し解析するための特殊な装置が設置されており，その効用が期待されている。しかし，このような情報の処理技術は現在なお研究途上にあるものが多く，これらの研究成果をただちに民族学での情報の処理に広く応用し得るとは必ずしもいえない面が多い。この分野では，人間の感覚ではごく単純に見えても，機械による自動処理はきわめて困難，ということは案外多く残されているのである。

ところがこれとは逆に，情報処理サイドから見たときには，日常的でありふれた道具だてや手法であっても，使い方によっては，民族学研究への応用として大いに威力を発揮する，といった性質のものも多い。この小文で報告しようとしているのは，まさにそのようなものの実例であって，いわば民族学研究におけるソフトウェア・トゥールである。これは，機能的にはごく単純な仕事をするだけの計算機プログラムであるが，民族学研究の過程で生じる，ある程度のまとまった量のデータを解析する際に，それほど大げさになることなく簡便に利用でき，しかも多くの問題に広く適用できる。

民族学に限らず，一般に研究過程においては，研究の対象としている「もの」の集合について，個々の「もの」の持っている属性を調査し，得られた一群のデータの集まりから，この「もの」の集合を統轄する一般的法則を導いたり，逆に相異なる性質を持つ部分集合に分割したりすることは，きわめてひんぱんに要求される。ここで，対象となる「もの」としては，たとえば人間，標本資料などの具体的存在だけでなく，行政区画や言語などの抽象的な存在も考えられる。また，個々の「もの」に付随する

属性としては，たとえば，その名称，色，形状，大きさ，人口数，特徴などがある。

これらについて整理されたデータの集合に関してまず第一に行うべきことは，各属性の実現値（実測値）の生起する頻度の分布を調べることであろう。また，ある属性の実現値について特定の条件を設定したときの，他の属性の実現値の頻度分布を調査することにより，これらの属性間の相互関係を知ることもできる。数値で表現されるタイプの属性については，二つの属性間で相関係数を求めたり，回帰分析を行ったりすることも基本的で重要な仕事である。

このようなとき，頻度分布は数値で表として表現するよりも，グラフで図として表現する方が，分布の様相を直観的に把握しやすく，より大きな効果が得られる。また，数値属性間の相互関連性も，散布図と回帰直線を図示してみることによって，より直観的な把握が可能になる。すなわち，データをパターン化し，視覚化して表示することが，結果の一覧性を高めるうえできわめて効果的である。

この際，取り扱う標本の数が多く，また属性の種類も多いと，頻度分布の図を描くのも手仕事では割に面倒な仕事となるが，データの入力が手軽で，操作の容易な計算機ソフトウェアがあれば，試行錯誤的に繰り返して分布を求めたり，各種の多様な条件のもとでの分布を求めたりすることが容易に行える。これは単に，従来手作業で行っていた作業を自動化するという省力的な意味だけでなく，むしろデータ解析の速時性と結果の一覧性が高まることによって，より多角的な解析が可能になるのであって，解析の質的な向上が期待できるのである。

本報告は，上述したような見地に立って構成した，民族学データのための汎用データ解析システムについて述べるものである。

以下では，説明の便宜上，本館所蔵の日本のタケカゴの標本資料について，本館第4研究部中村俊亀智教授の整理作成になるデータを例としてとりあげ，この例にそってシステムの機能，データの形式や利用法などについて述べることにする。

Ⅱ．取り扱うデータの形式とシステムの機能

解析の対象となる「もの」に付随する属性として，ここでは，以下の4種のものを考える。

1) 数値属性
2) 記述属性
3) コード属性
4) パターン属性

第一の数値属性は，何らかの計測によって得られる計量的な情報であって，何桁かの数字で表現されるものである。タケカゴについてのデータの場合，カゴの縦，横の寸法，重さなどの情報がこれにあたる。

記述属性は，たとえば人や物の名前のように，有限個の文字で記述される名称的な情報を指す。タケカゴでは，その呼び名，採集者名，所用地名などをこれにあてる。

三番目のコード属性は，属性の実現値（実測値）をある一定のカテゴリー単位にカテゴライズし，各々のカテゴリーに特定のコード（符号）を割り当てて表現したものである。これは，元来は数値属性あるいは記述属性としても表現し得るものではあるが，少数のカテゴリーに区分して取り扱う方がデータが見やすくなって便利な場合に利用する。コードは，ここではたかだか3桁程度の整数で表わされるものとする。タケカゴの例では，たとえばその形状や編み方などの作成法をコード化している。

最後のパターン属性は，以上の3種のものとはデータの表現形式のうえでやや性格が異なり，システム上での取り扱いも他の属性とは異なったものとなるのであるが，これは，対象の「もの」そのものの外形写真や図面，あるいは付随する地図など，2

表1　タケカゴデータの内容と構成

項目番号	属　　性	種　類	書　　式*	単位，記法など
1	横の寸法	数　値	×××.×	cm
2	縦の寸法	〃	×××.×	cm
3	高　　さ	〃	×××.×	cm
4	重　　さ	〃	××.×	kg
5	立竹の幅	〃	××.×	cm
6	廻し竹の幅	〃	××.×	cm
7	巻竹の幅	〃	××.×	cm
8	縁竹の幅	〃	××.×	cm
9	形　　状	コード	××	
10	底の作り方	〃	××	
11	胴の作り方	〃	××	
12	縁の作り方	〃	××	
13	採集年次	記　述	××××	西　暦
14	用途など	〃	20文字	訓令式ローマ字綴
15	備　　考	〃	3文字	
16	収集番号	〃	5文字	アルファベット，数字
17	呼び名	〃	20文字	訓令式ローマ字綴
18	所用地	〃	25文字	〃　（二名法）
19	採集者氏名	〃	20文字	〃

*：×は0～9の数字

表2　コードの定義

項目 / コード	9：形状	10：底 11：胴の作り方	12：縁の作り方
1	球形	網代（市松）	共縁
2	円錐形	網代編	巻口仕上げ
3	円筒形	笊目編	蛇腹巻き
4	三角錐形	四つ目編	野田口仕上げ
5	四角錐形	四つ目変形	重ね縁
6	直方体，立方体形	六つ目編	その他
7	円口方底形	六つ目変形	—
8	半球形	八つ目編	—
9	円錐台形	双子編，巻き籠	—
10	その他	藤箕編	—
11	—	筒，曲物	—
12	—	板（折りまげ）	—

次元平面のうえで表現されるパターン情報を指すものとする。カゴデータの場合はタケカゴの外形写真を用いている。

さて，対象となる「もの」の集合について，上に挙げたような4種類の属性が，それぞれ任意の個数，任意の組み合せで定義されており，個々の標本（サンプル）についてこれらの属性の実現値（実測値）が得られているものとする。もちろん，対象となるものの本質的な性質上，あるいは解析の目的などの理由により，上述の4種の属性のうちある種のものはデータの中に含まれていないこともある。

このようなデータの具体例であるタケカゴデータの内容と構成を表1に示す。ここでは，項目番号1～8を数値属性項目，9～12をコード属性項目，13～19を記述属性項目としており，さらに，各々のカゴの外形写真をパターン属性として定義している。また，コード属性項目で用いるコードの定義を表2に掲げる。表3はデータの内容の一部を示している。（実際には，全部で396個のタケカゴについてこれらの19種の属性とパターン属性が得られている。）

このようなデータの集合に対してシステムの行う機能は次のようなものである。タケカゴのデータに対して行った処理例を挙げながらそれぞれの機能について述べる。

1)　頻度分布の表示：特定の属性項目について，その実現値の頻度の度数分布図（ヒストグラム）を求め図示する。数値属性項目については，分布の平均値，標準偏差などの統計量の値も同時に表示する。記述属性項目の分布は出現頻度順にソートして並べた表として示す。

表3 タケカゴデータの一部 (a)

項目番号 / 通し番号	1	2	3	4	5	6	7	8	9	10	11	12	13	14
00001	42.0	28.5	46.5		0.8	0.3	0.6	2.0	6	5	3	4	1939	YASAIIRE
00002	45.0	43.5	55.5	3.1	0.7	0.3	1.0	2.5	3	7	7	2	1934	
00003	45.0	42.0	49.0	2.5	0.7	0.3	0.7		3	5	3	1	1931	
00004	43.3	29.0	56.0	1.8	1.5	2.0	0.7	2.3	6	7	7	2	1937	
00005	51.0	49.5	53.0	2.8	1.5	1.3			6	7	7	2		
00006	50.0	48.0	47.0	2.5	1.5	2.5	1.5	2.5	3	7	7	2	1935	
00007	49.5	46.5	52.0	2.6	1.2	0.4			3	5	3	2	1935	
00008	42.0	41.0	47.0	1.9	1.5	1.5			3	7	7			
00009	40.0	40.0	40.0	1.5	1.5	0.6			7	5	3			
00010	38.0	33.0	46.5	2.8	1.5	1.5	1.5		6	2	2	1	1934	
00011	47.0	45.5	56.0	3.4	1.3	0.3	1.0	1.0	7	2	3	2		
00012	43.5	26.5	37.0	0.9	0.7	0.3	0.5	0.3	10	7	7	5	1950	
00013	51.0	36.0	44.0	1.4	1.0	1.1	1.0	0.5	10	7	7	5	1934	
00014	40.5	39.5	54.5	1.7	1.1	0.7			3	4	3	5	1950	
00015	46.0	45.0	43.5	1.6	1.0	0.3	1.5	2.0	7	5	3	2		
00016	57.0	57.0	49.0	1.6	0.8	1.0	1.0	2.3	3	6	6	2		
00017	19.0	18.5	20.0	0.3	0.5	0.2			7	2	3	3		TABEMONOIRE
00018	32.0	30.0	32.0	0.9	0.7	0.3			7	2	3	3		
00019	21.0	21.0	22.3	0.5	0.7	0.2			7	2	3	3	1928	
00020	34.0	33.0	28.5	0.9	0.9	0.9			7	3	3	2	1927	

表3　タケカゴデータの一部 (b)

項目番号 / 通し番号	15	16	17	18	19
00001	A 8	8637	HA-KEGO	YAMAGATAKEN-MINAMIHARA	KUNUGI KINAI
00002	A 8	3683	SYOI-KAGO	TOTIGIKEN-KORI	AM DOOZIN
00003	A 8	2854	SYOI-KAGO	TIBAKEN-HUNAGATA	SAZI YUUKITI
00004	A 8	7119	YAMA-KAGO	SAITAMAKEN-MINAMIKOMA	MURAKAMI SEIBUN
00005	A 8	Z 67			
00006	A10	5079	SEI-KAGO	SIZUOKAKEN-SIZUOKASI	UTIDA TAKESI
00007	A10	5796	SEI-PUKU	SIZUOKAKEN-MIWA	UTIDA TAKESI
00008	A10	Z 68		YAMANASIKEN	
00009	A10	Z 69		NAGANOKEN-SUWA	
00010	A10	3625	HIRA-KOOZA	GIHUKEN-TAKAYAMA	TAKAHASI BUNTAROO
00011	A10	Z 70			
00012	A12	23457		MIYAZAKIKEN-TAKATIHO	YAHATA ITIROO
00013	A12	3746	TAKEGARAI	MIYAZAKIKEN-TAKATIHO	HAYAKAWA KOOTAROO
00014	A12	21089	MEGO-TEBO	NAGASAKIKEN-TOYOSAKI	MIYAMOTO KEITAROO
00015	A12	Z 71			
00016	A12	16394			
00017	A14	6435	TABIRROKO	TOOKYOTO-NIIZIMA	HUZIKI KIKUMARO
00018	A14	6436	KOME-KAGO	TOOKYOTO-NIIZIMA	HUZIKI KIKUMARO
00019	A14	2045	BIKU	TOOKYOTO-KOOZUSIMA	HUZIKI KIKUMARO
00020	A14	2040		TOOKYOTO-MIYAKEZIMA	HUZIKI KIKUMARO

図1は，タケカゴの「重さ」の分布，図2は「高さ」の頻度分布を求めた結果である。また，図3はコード属性である「形」の分布を示している。図3では，右上部の枠内に，用いたコードの定義が示されている。図4は記述属性に対する例として，「所用地」の分布を求めたもので，出現頻度の多い方から40番目までを示している。

2)　線形回帰分析：任意の二つの数値属性項目間で線形回帰分析を行い，散布図と回帰直線を図示する。

図5はタケカゴの「縦」の寸法と「横」の寸法について行った回帰分析の結果である。図の上部には分布の平均値，標準偏差，相関係数などの統計量と回帰直線の方程式も印字している。

3)　相関：すべての数値属性項目間で相関係数を計算し印字する。

図6にタケカゴデータについての結果を示す。各項目番号の表す属性は表1に記しているとおりである。

4)　条件付き分析：以上の 1)～3) の処理に際して条件付きの処理を行う。すなわち，直接の対象としている属性とは別の任意の複数の属性について，その実現値にあ

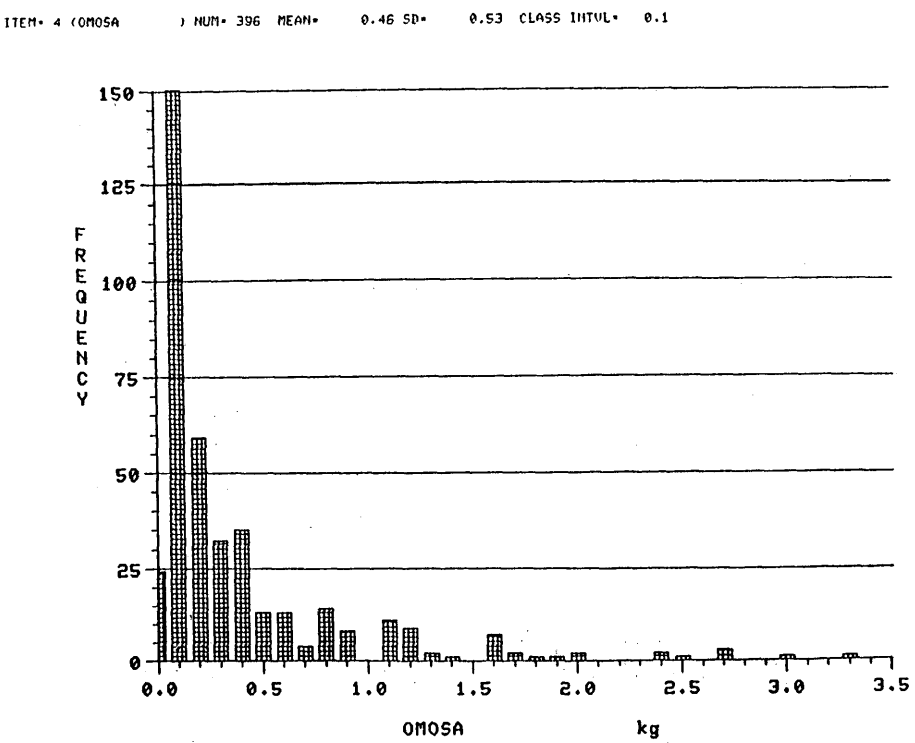

図1　数値属性の頻度分布の表示 (1)：カゴの「重さ」の分布

ITEM= 3 (TAKASA　　) NUM= 396 MEAN= 20.30 SD= 12.63 CLASS INTVL= 1.0

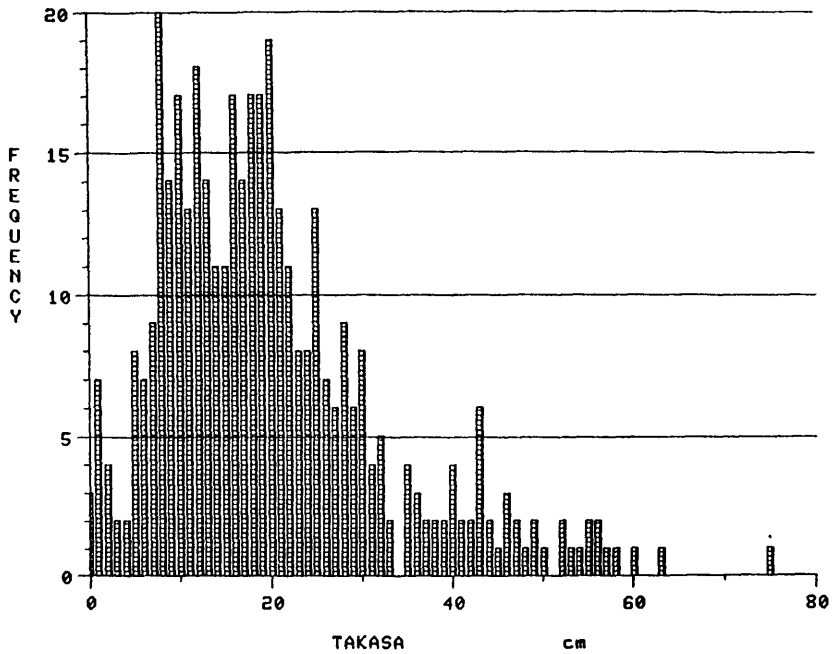

図2　数値属性の頻度分布の表示 (2)：カゴの「高さ」の分布

ITEM= 9 (KATATI　　) NUM= 395

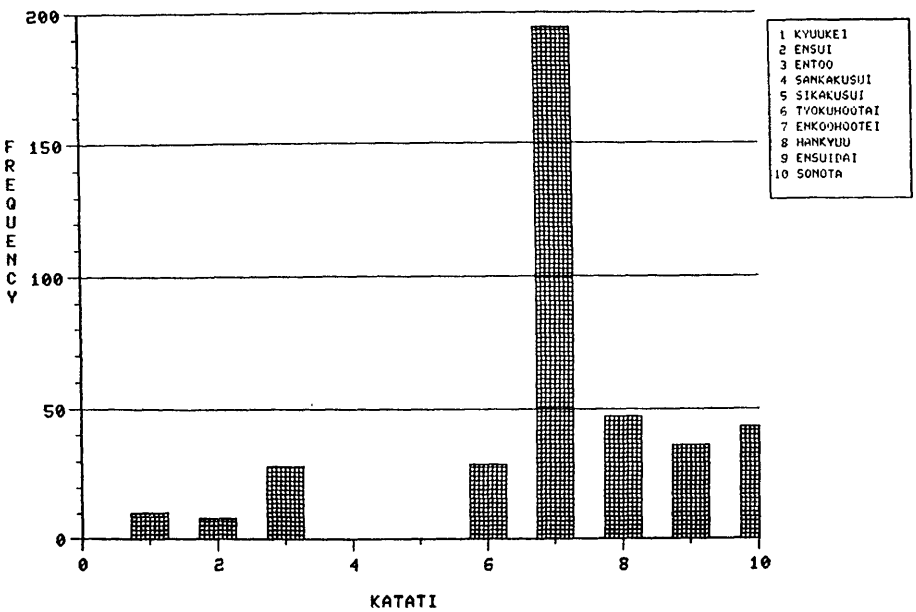

図3　コード属性の頻度分布の表示：カゴの「形」の分布

```
 129 CATEGORIES FOUND
ENTER NUMBER OF CATEGORIES YOU WANT
?
.40

ITEM-18 (SYOYOO TI    )      NUM- 308
```

	CATEGORY	OCCURRENCE
1	KAGOSIMAKEN-KIKAIGASIMA	19
2	KAGOSIMAKEN-AMAMIOOSIMA	15
3	KOOTIKEN-NAKAMURA	13
4	SIZUOKAKEN-SIZUOKASI	11
5	MIYAZAKIKEN-TAKATIHO	11
6	NIIGATAKEN-SADO	10
7	KAGOSIMAKEN	7
8	NAGANOKEN-SUWA	6
9	GIHUKEN-TAKAYAMA	6
10	TOOKYOTO-NIIZIMA	5
11	KAGOSIMAKEN-IOOZIMA	5
12	NAGANOKEN-UEDA	5
13	NAGASAKIKEN-TUSIMA	5
14	KAGOSIMAKEN-KAGOSIMASI	5
15	YAMAGATAKEN-MINAMIHARA	4
16	SIZUOKAKEN-MIWA	4
17	KAGOSIMAKEN-NAKANOSIMA	4
18	OKINAWAKEN-HONTOO	4
19	OKINAWAKEN-ISIGAKIZIMA	4
20	NIIGATAKEN-MURAKAMI	4
21	AITIKEN-HONGOO	4
22	TOTIGIKEN-KORI	3
23	SAITAMAKEN-MINAMIKOMA	3
24	TOOKYOTO-HATIZYOZIMA	3
25	TOOKYOTO	3
26	KAGOSIMAKEN-KUTINOERABU	3
27	KAGOSIMAKEN-TAKESIMA	3
28	OKINAWAKEN	3
29	SIZUOKAKEN-TAGO	3
30	AOMORIKEN	3
31	TOOKYOTO-AOGASIMA	3
32	YAMAGATAKEN	3
33	AKITAKEN-KUMOSAWA	3
34	WAKAYAMAKEN	3
35	TOOKYOTO-MIYAKEZIMA	2
36	KAGOSIMAKEN-KUTINOSIMA	2
37	KAGOSIMAKEN-KOTAKARAZIMA	2
38	AOMORIKEN-HATINOHE	2
39	KAGOSIMAKEN-SAKURAZIMA	2
40	TOOKYOTO-OOMORI	2

図4　記述属性の頻度分布の表示：カゴの「所用地」の分布

る条件を設定し，この条件を満たす標本だけについて上の 1)～3) で述べた処理を行う。

たとえば，図7はカゴの「形」が「円口方底形」であるものについて，その「高さ」の分布を見たものであって，コード属性の実現値に条件付けを行って数値属性の頻度分布を調べている。結果の図の上部には設定した条件を印字している。

また，図8は複数の条件付けを行って分布を調べた例であって，「横」の寸法が 10

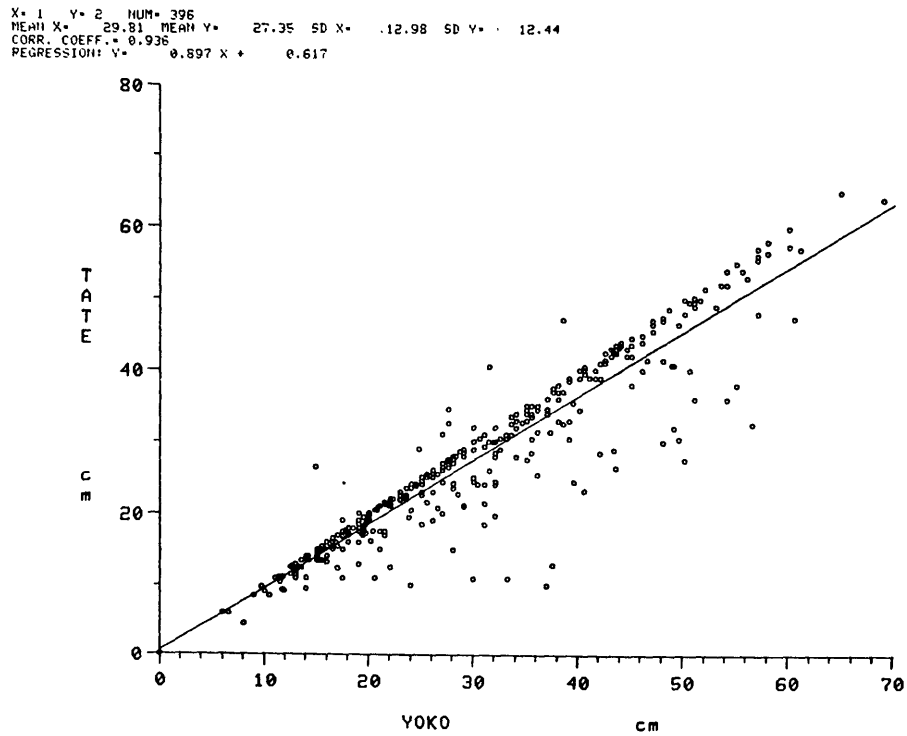

図5　線形回帰分析：カゴの「横」の寸法と「縦」の寸法

CORRELATION COEFFICIENTS

ITEM	1	2	3	4	5	6	7	8
1	1.000							
2	0.936	1.000						
3	0.248	0.210	1.000					
4	0.618	0.597	0.729	1.000				
5	0.254	0.248	0.184	0.306	1.000			
6	0.315	0.281	0.244	0.332	0.584	1.000		
7	0.306	0.297	0.206	0.299	0.105	0.083	1.000	
8	0.347	0.383	0.026	0.192	0.176	0.265	0.121	1.000

図6　相関係数の表示

cm から 50 cm の範囲内で，「高さ」が 30 cm 以下であって，しかも「所用地」が「鹿児島県」でないカゴについて，その「形」の分布を求めている。このように，複数の異なった種類の属性項目で条件付けが可能で，さらに，この条件のなかにはネガティブな条件も設定できるようにしている。

図9は回帰分析の際に（ネガティブな）条件付けを行った例であって，「形」が「直方体」でも「その他」でもないものについて，「横」の寸法と「縦」の寸法の関係を求めたものである。

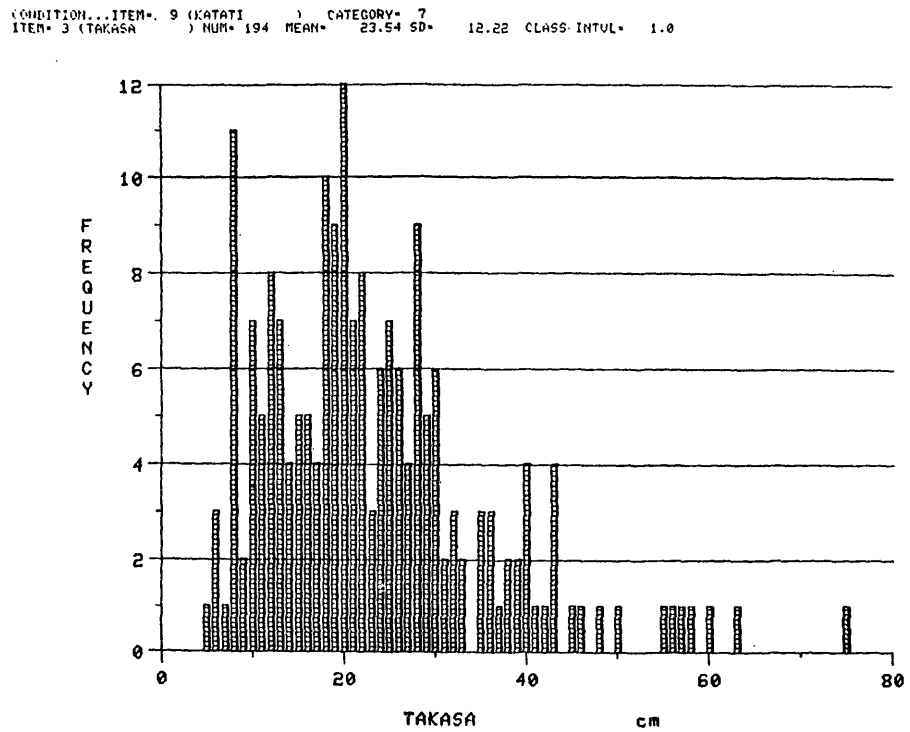

図7　条件つき頻度分布(1)

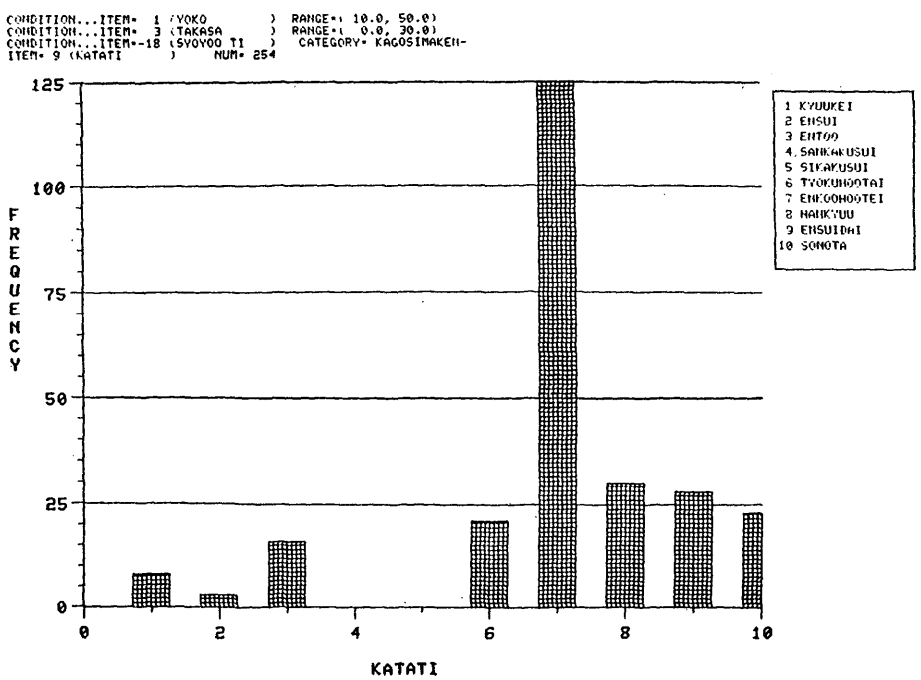

図8　条件つき頻度分布(2)

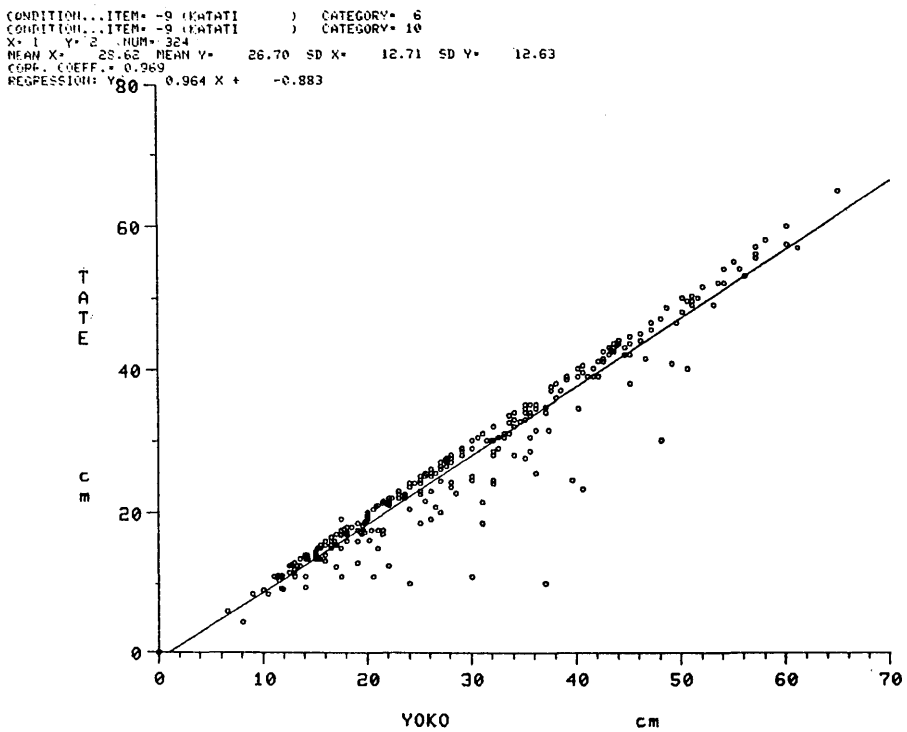

図9　条件つき回帰分析

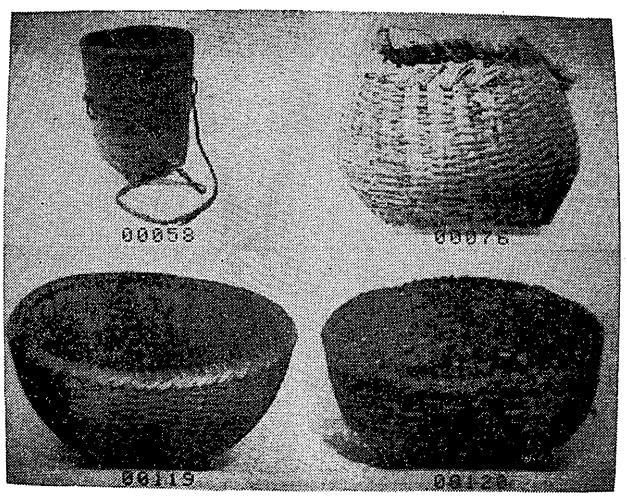

写真1　パターン属性の表示例：図10の検索で得られたカゴの外形写真

```
QUERIES ENTERED:
ITEM  CATEGORY OR RANGE
 18 = OKINAWAKEN-
  9     7
-10     5
  3 = (    10.0,    40.0)

DATA FOUND:
RECORD NO.
   58
   76
  119
  120

  4 RECORDS FOUND

DISPLAY ? (REC NO, -1: ALL, 0: SKIP)
?
```

図10a　データの検索：検索結果

```
RECORD NO=    58
 1  YOKO                25.5 cm
 2  TATE                25.0 cm
 3  TAKASA              30.5 cm
 4  OMOSA                0.5 kg
 5  TATI DAKE            0.5 cm
 6  MAWASI DAKE          0.5 cm
 7  MAKI DAKE            0.5 cm
 8  HUTI DAKE            0.0 cm
 9  KATATI          7
10  SOKO            4
11  DOO             3
12  HUTI            1
13  SAISYU NENZI   1935
14  YOOTO NADO
15  SONOTA         A26
16  BANGOO          5667
17  YOBINA         KUIZIKI
18  SYOYOO TI      OKINAWAKEN-HONTOO
19  SAISYUU SYA    HAYAKAWA KOOTAROO
--- NO.  1 OF    4

NEED PICTURE?
?
.1
```

図10b　データの検索：属性の表示

5)　データの検索：複数の属性項目について実現値に条件を設定し，この条件を満たす標本を探し出して得られた標本の通し番号を印字する。さらに，見つかったすべての標本，あるいはこのうちの指定した一部の標本について，付随する属性データを表示する。またこのデータにパターン属性が定義されておればこれを画像ディスプレイ上に表示する。

図10aは，「所用地」が「沖縄県」，「形」が「円口方底形」で「高さ」が10 cmから40 cmの範囲内にあって，しかも「胴」の編み方が「四つ目変形」でないカゴを検索したときの例で，四つの標本が見つかっている。図10bはこの結果のうち，標本の通し番号が58のものについて実際の属性データを表示したものである。また，見つかった四つのカゴのパターン属性（外形写真）を画像ディスプレイに表示したのが写真1である。

6)　属性の表示：標本の通し番号を指定しその標本のすべての属性を表示する。たとえば，図10bで示したような属性データの表示と写真1のような画像の表示が得られる。

Ⅲ．システムの構成と使用法

本システムは図11に示すようなハードウェア構成のもとで利用する。主計算機のIBM 370/138は対話型モニタシステム（CMS）のもとで動作させ，利用者はストレ

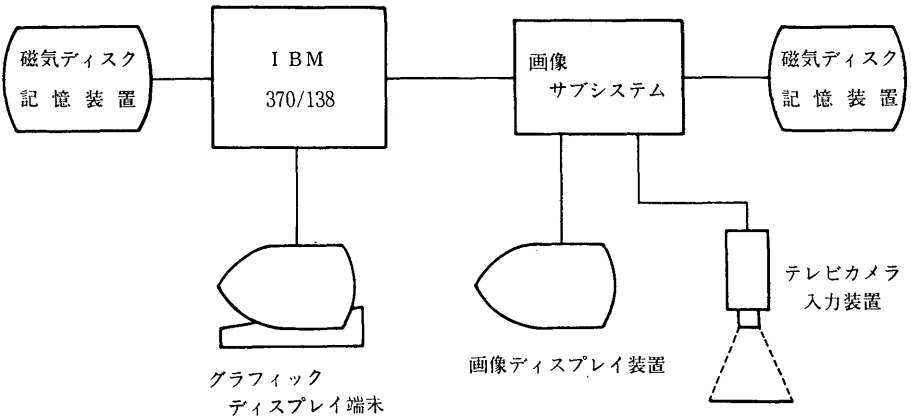

図11 システム構成図

ージタイプ・グラフィックディスプレイ端末からシステムを利用する。グラフなどの計算機からの情報はこのグラフィックディスプレイ端末に表示される。

解析の対象となるデータのうちパターン属性以外のデータは，CMS ファイルとして IBM システムの磁気ディスク装置内に記憶させておく。また，写真等のパターン属性データは，IBM システムに接続された画像サブシステムで取り扱う。すなわち，写真や図面などの画像情報は，画像サブシステムの画像入力装置（たとえばテレビカメラ入力装置）を利用してディジタル化し，画像サブシステムの磁気ディスクに記憶させておく。システムからパターン属性の表示が要求されると，対応する画像データを磁気ディスク装置内から捜し画像ディスプレイ装置に表示する。しかし，対象としているデータにパターン属性が定義されていないとき，あるいはこれを表示する必要のないときは画像サブシステムを利用する必要はなく，IBM システムとグラフィックディスプレイ端末だけがあればよい。

さて，一般に計算機でデータを取り扱う場合には，データの形式（たとえば，属性項目の構成や数値の桁数，小数点位置など）に関して，実際に入力されるデータと計算機プログラムの側で認識するデータの形式との間で厳密な対応が保たれていなければならない。このためには，利用者が準備したデータの形式に合致するように，プログラム側のデータ形式を定めるのが最も簡単な方法である。しかし，これでは，取り扱うデータが変るたびにシステムを作成しなおしたり，入力するデータを既存のシステムが受けつける形に無理に直したりする必要があり，システムは汎用性にとぼしいものとなる。

本報告のシステムでは以下のような方法で汎用性を維持している。まず，利用者の

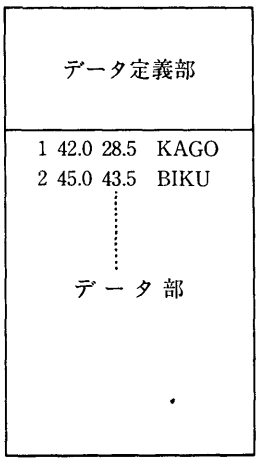

図12
データファイルの構成

側でデータの構成と書式を定め，それに従ってデータを整理する。計算機へのデータの入力に際しては，データそのものだけでなく，データの構成と書式に関する情報も，「データ定義情報」として，データとともに入力しファイルに記憶させる（図12）。システムはこのデータファイルから，データ定義情報を記したデータ定義部をまず読み込み，これに従って実際のデータを読み込んでゆく。こうすることによって，システムと利用者の作成したデータとの対応が保たれ，常に利用者の意図したとおりのデータがシステム内に再構成されることになる。対象とする「もの」やデータの構成が変っても，それに応じてデータ定義部を書きかえておけば，システムのプログラム側には全く変更を加えることなく，新しいデータを取り扱うことができる。

実際にデータを作成する場合には次のようなことに留意すればよい。まず，各属性項目の種類と書式を定める。属性の種類は属性の本質的な性質により自然に決まるものが多いが，書式にはいろいろな可能性が考えられるので，効率が良くしかも無理のない方法に決める。すなわち，数値属性項目では表現する数値の桁数と小数以下の桁数，コード属性では用いるコードの桁数，記述属性では使用する文字桁数を決める。これらの情報とともに，計測単位や記法などの補助的な情報をまとめて表1のような形に整理する。このとき，属性項目の順序は全く任意に決めればよく，異なる種類の属性が混在していても構わない。項目は先頭の項目から順に項目番号をつける。今後，処理の際には，各属性はこの項目番号で指定されることになる。

さらに，コード属性項目では，使用するコードの定義を表2のようにまとめる。表1，表2のような約束に従って，データを表3のような形に整理する。この際，データを記入する用紙には，項目ごとに縦の欄を区切り，各項目の欄内を，さらに，表1に記した各項目で使用する文字数分だけ細かく欄に分けたものを使用すると，実際にカードパンチなどの方法で計算機に入力する際にそのまま使用でき便利である。

システムの操作は対話的に行われ，利用者は，システムが端末に印字するメッセージに従って必要な情報をキー入力してゆけばよい。

まず，グラフィック端末から本システムを起動すると，図13aに示したようなメッセージが端末の管面に表示される。この中から必要なファンクションを選択し，

```
CHOOSE ONE OF THE FOLLOWING FUNCTIONS

DISTRIBUTION ......... 1
CORRELATION ........... 2
REGRESSION ............ 3
SEARCH ................ 4
DISPLAY ............... 5
TERMINATE JOB ........ 0

?
.1
ENTER CONDITION ITEM NO.
( 0 : TERMINATE, 99 : EXPLAIN )
?
.0

ENTER ITEM NUMBER (99 : EXPLAIN)
?
.99

ITEM  ATTRIBUTE       KIND

 1    YOKO             NUM
 2    TATE             NUM
 3    TAKASA           NUM
 4    OMOSA            NUM
 5    TATI DAKE        NUM
 6    MAWASI DAKE      NUM
 7    MAKI DAKE        NUM
 8    HUTI DAKE        NUM
 9    KATATI           CODE
10    SOKO             CODE
11    DOO              CODE
12    HUTI             CODE
13    SAISYU NENZI     DESC
14    YOOTO NADO       DESC
15    SONOTA           DESC
16    BANGOO           DESC
17    YOBINA           DESC
18    SYOYOO TI        DESC
19    SAISYUU SYA      DESC


ENTER ITEM NUMBER (99 : EXPLAIN)
?
.17
```

図13a　システムの起動と“EXPLAIN”

その番号をキー入力する。システムは，指定したファンクションに応じて，それぞれの仕事に必要なパラメータ，たとえば対象とする項目番号，条件を与える項目やその値などを問い返してくるので，以後，システムの出す指示に従って，順次それに答えてゆけば所要の結果が得られる。また，処理にあたって項目番号を指示する際に，項目番号と実際の属性との対応が分らなくなることがしばしばある。このようなときには，“EXPLAIN”のファンクションを指示すると対応表を表示するので，利用者はこれを見て必要な項目番号をあらためてキー入力すればよい（図 13a)。

図13では，一連の対話的な処理の流れの例を示している。システム起動後まず，“EXPLAIN” によって項目番号と属性との対応を調べた後（図 13a），項目番号17の「呼び名」の分布を求めてみる（図 13b）。その結果，“TEGO” または，語尾に “TEGO” のつく呼び名が比較的多いことが分る。これについて，“TEGO” という呼び名の地域的特質を見るため，「呼び名」が “TEGO” で終るものだけについて，

```
 139 CATEGORIES FOUND
ENTER NUMBER OF CATEGORIES YOU WANT
?
.40

ITEM-17 (YOBINA        )      NUM- 277
```

	CATEGORY	OCCURRENCE
1	TEGO	14
2	HA-KEGO	11
3	BIKU	10
4	KAGO	5
5	SOOKE	5
6	KAKKOBE	5
7	SIO-KAGO	5
8	TERU	4
9	MISO-KOSI	4
10	IDOU-URABAKUU	3
11	IKAKI	3
12	SOOBINKAA	3
13	SAN-BARA	3
14	AZIKA	3
15	ME-KAGO	3
16	SIO-BIKU	3
17	SYOI-KAGO	2
18	URABAKUU	2
19	KOSITUKE-KAGO	2
20	ISUMAGUU	2
21	TIRUNAA	2
22	ARAI-KAGO	2
23	SADO-SOOKE	2
24	TESIMA	2
25	SA-DOOSI	2
26	TANE-BIKU	2
27	SAKANA-KAGO	2
28	ITAYA-KAKKOBE	2
29	MARU-BIKU	2
30	SEOI-BIKU	2
31	UBIKU	2
32	ISU-TIIRU	2
33	OBOKE	2
34	SIO-TEGO	2
35	MASYUU	2
36	NORITUMI-KAGO	2
37	YAMA-KAGO	1
38	SEI-KAGO	1
39	SEI-PUKU	1
40	HIRA-KOOZA	1

図13b 「呼び名」の頻度分布

その「所用地」の分布を調べている（図 13c）。この結果，「鹿児島県」に多いことが分る。次に，語尾 "TEGO" は主にどのような単語に接続するかを見るため，「呼び名」が "TEGO" で終るものについて，その「呼び名」自身の分布を調べた（図 13d）。さらに，「呼び名」における "TEGO" とカゴの形との関連を見るため，「呼び名」が "TEGO" で終るものについて，「形」の分布を求めた（図 13e）。最後に，"TEGO" で終る名を持つもので，しかも「形」が「円錐台形」の標本を検索した結果が図 13f

```
CONDITION...ITEM= 17 (YOBINA       )   CATEGORY= -TEGO
  14 CATEGORIES FOUND
ENTER NUMBER OF CATEGORIES YOU WANT
?
.14

ITEM=18 (SYOYOO TI    )     NUM=  23

      CATEGORY                          OCCURRENCE

  1   KAGOSIMAKEN-AMAMIOOSIMA              5
  2   KAGOSIMAKEN-NAKANOSIMA               3
  3   KAGOSIMAKEN-KUTINOSIMA               2
  4   KAGOSIMAKEN-KOTAKARAZIMA             2
  5   KAGOSIMAKEN-KAGOSIMASI               2
  6   KAGOSIMAKEN-TAKESIMA                 1
  7   KAGOSIMAKEN-IOOZIMA                  1
  8   KAGOSIMAKEN-SAKURAZIMA               1
  9   KAGOSIMAKEN-KUTINOERABU              1
 10   KAGOSIMAKEN-AKUSEKIZIMA              1
 11   KAGOSIMAKEN-IBUSUKI                  1
 12   AKITAKEN                             1
 13   NIIGATAKEN-KANAYA                    1
 14   YAMAGATAKEN-HIGASI                   1
.
```

図13 c 「呼び名」の語尾に "TEGO" の付くカゴの「所用地」

```
CONDITION...ITEM= 17 (YOBINA       )   CATEGORY= -TEGO
   9 CATEGORIES FOUND
ENTER NUMBER OF CATEGORIES YOU WANT
?
.9

ITEM=17 (YOBINA       )     NUM=  23

      CATEGORY                          OCCURRENCE

  1   TEGO                                14
  2   SIO-TEGO                             2
  3   TYUUHAN-TEGO                         1
  4   NABATORI-TEGO                        1
  5   OO-TEGO                              1
  6   KO-TEGO                              1
  7   MESI-TEGO                            1
  8   MATUDON-TEGO                         1
  9   AMI-TEGO                             1
.
```

図13 d 語尾に "TEGO" の付く「呼び名」

で，見つかった標本の属性データを表示したのが図 13g である。またこれらの標本のパターン属性を表示したものが写真2である。

このように，得られた結果を検討しながら対話的にデータの解析を進めてゆくことができるので，即時性を保ったまま，試行錯誤的に多角的な解析を行うことができる。

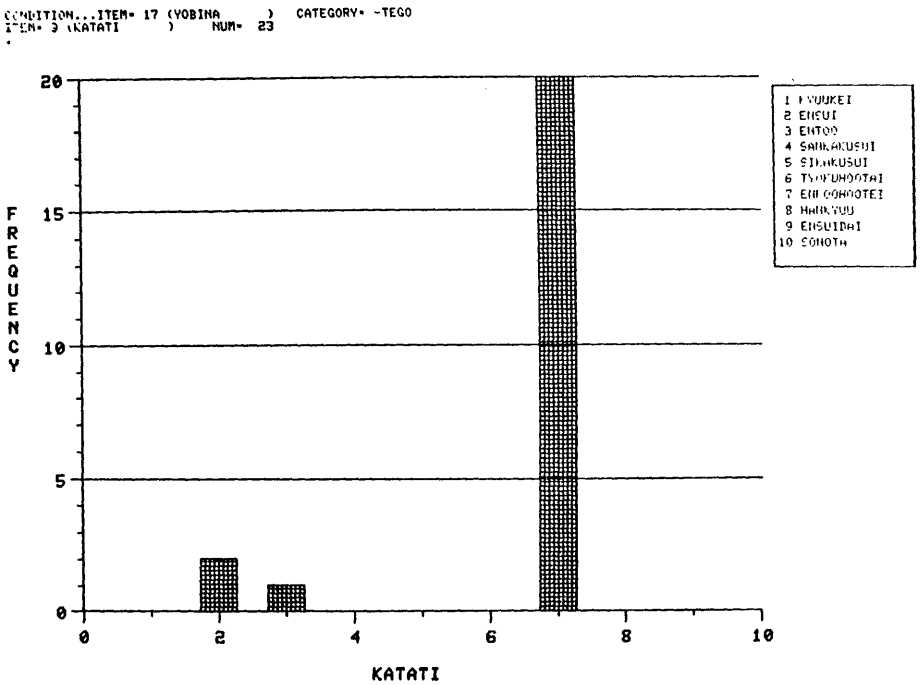

図13e 「呼び名」の語尾に“TEGO”の付くカゴの「形」の分布

```
QUERIES ENTERED:
ITEM  CATEGORY OR RANGE
 17 = -TEGO
  9     2

DATA FOUND:
RECORD NO.
  354
  355

  2 RECORDS FOUND

DISPLAY ? (REC NO, -1: ALL, 0: SKIP)
?
.
```

図13f 「呼び名」の語尾に“TEGO”を持ち「形」が「円錐形」のカゴの検索

```
RECORD NO=    354
 1  YOKO                    21.0 cm
 2  TATE                    21.0 cm
 3  TAKASA                  22.0 cm
 4  OMOSA                    0.1 kg
 5  TATI DAKE                1.5 cm
 6  MAWASI DAKE              1.5 cm
 7  MAKI DAKE                0.0 cm
 8  HUTI DAKE                2.5 cm
 9  KATATI            2
10  SOKO              0
11  DOO               2
12  HUTI              4
13  SAISYU NENZI    1932
14  YOOTO NADO
15  SONOTA          A12
16  BANGOO           3048
17  YOBINA          SIO-TEGO
18  SYOYOO TI       KAGOSIMAKEN-KAGOSIMASI
19  SAISYUU SYA     HAYAKAWA KOOTAROO
--- NO.  1 OF    2

NEED PICTURE?
?
.

RECORD NO=    355
 1  YOKO                    20.8 CM
 2  TATE                    20.8 CM
 3  TAKASA                  19.5 CM
 4  OMOSA                    0.1 KG
 5  TATI DAKE                1.1 CM
 6  MAWASI DAKE              1.1 CM
 7  MAKI DAKE                0.0 CM
 8  HUTI DAKE                1.5 CM
 9  KATATI            2
10  SOKO              0
11  DOO               2
12  HUTI              4
13  SAISYU NENZI    1934
14  YOOTO NADO
15  SONOTA          A12
16  BANGOO           4293
17  YOBINA          SIO-TEGO
18  SYOYOO TI       KAGOSIMAKEN-AMAMIOOSIMA
19  SAISYUU SYA     AM DOOZIN
--- NO.  2 OF    2

NEED PICTURE?
?
.1
```

図13g　図13fの検索で得られたカゴの属性データの表示

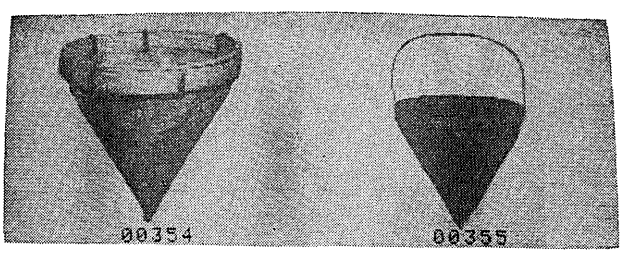

写真2　「呼び名」の語尾が“TEGO”で「形」が「円錐形」のカゴ

Ⅳ．あとがき

報告したデータ解析システムによって，手もとのデータについて，各種の条件のもとでの属性の頻度分布を，容易にグラフとして視覚化して表示することが可能になった。これは，ごくありふれた手段を用いたものであって，機能的にも特に目新しいものとはいえない。しかし，このような比較的単純なシステムであっても，その効用は意外に大きく，まとまったデータの集まりから，埋もれていた多くの情報を効率良く引き出すのに役立つことが分った。ある仮説にもとづいて分布を調べ，その結果を視覚的に把握して直ちに次の解析の方向を決めるといった対話的な使用方法が可能なので，グラフを描くための単なる省力化のための手段というより，むしろ本システムにより解析の質的な変化が期待できるのである。また，本報告では対象データの例としてタケカゴに関するデータをとりあげたが，属性が数値，文字，コード，パターンで構成されるものであればどのようなデータでも取り扱うことができるので，各方面のデータ解析に広く利用できるものと考える。今後，本システムが民族学研究，特に物質文化の研究に大いに利用されることを期待する。

末筆ながら，日頃ご指導たまわる京都大学長尾真教授，ならびに，本システム作成に際し多くのご教示をいただいた本館中村俊亀智教授に心から謝意を表する。

文　献

中村俊亀智
1977「中国地方タケカゴ細工の一側面」『国立民族学博物館研究報告』2(4)：806-827。
1979「国立民族学博物館所蔵の東南アジア島嶼部採集のカゴ細工について」『国立民族学博物館研究報告』4(1)：130-151。