

標本調査支援システム ; Three-S System

木村, 宏
岡山理科大学理学部応用数学科

一村, 稔
岡山理科大学理学部応用数学科

浅野, 長一郎
九州大学大学院総合理工学研究科情報システム学専攻

<https://doi.org/10.15017/17164>

出版情報 : 九州大学大学院総合理工学報告. 11 (3), pp.349-345, 1989-12-01. 九州大学大学院総合理工学研究科
バージョン :
権利関係 :

標本調査支援システム：Three-S System

木 村 宏*・一 村 稔*・浅 野 長一郎**

(平成元年 8 月31日 受理)

Sampling Survey Support System : Three-S System

Hiroshi KIMURA, Minoru ICHIMURA and Chooichiro ASANO

A sampling survey support system, Three-S System, that works on personal computers has been developed. As recent word processor softwares provide features to insert comments into a document file, a designer of questionnaire sheets can enter attributes of each question in a comment record when he prepares his sheets with such a word processor. This system extracts these comments from a document file and generates an attribute file to use it on the input phase of questionnaire data and calculation of some basic statistics for the resultant data file. The detailed syntax to define attributes of questionnaire have been described with some examples.

1. 緒 言

最近では、パーソナルコンピュータの性能向上にともない、中・小規模の世論調査やアンケート調査等の標本調査に関する集計作業、特にその初期段階の作業にパーソナルコンピュータを用いる機会が増えてきた。ところが実際のデータ入力から単純集計作業（クロス集計程度）まではデータハンドリングに伴う単純作業が中心であり、多くの人力を必要とする。また、集計結果の出力に要するプログラムは、ほぼ要求される集計表の個数だけ必要となり、しかも調査毎に異なる書式の入出力を取り扱わなければならない、結構手間取るものである。その中で一番わずらわしいのがデータ入力である。調査票設計者や入力プログラム作成者が直接入力作業を担当すれば、それなりに正確なデータ入力が期待できるが、現実にはアルバイト要員やキー入力に不慣れな者が作業に加わる事が多く、入力データファイルの確認だけでも大変な手間がかかる。ここで提案するシステムは、これらの作業に関する労力を極力軽減し、信頼性を高めることを目的とし、調査票の作成から単純集計作業まで新規プログラミングなしで遂行する手段を提案する。原始データファイル作成

後の統計処理は Micro-NISAN 等の既存の統計処理パッケージを適用する。ただ、公表されているアンケート調査や世論調査には様々な設問様式が採用されており、あらゆるタイプの調査票に適應できる汎用システムの実現は、データベースシステムのデータ定義とデータ入力部分の構築に担当する大がかりなシステムとなる。したがって、本システムでは調査票の各設問項目の性質を記述する構文規則を定義し、実用上問題とならない程度の制限を設けることで、パーソナルコンピュータに適した調査データ集計支援システムを構成する。

2. システム設計概要

本システムの目的は調査票集計作業を支援する汎用システムの研究開発である。したがって、まず調査票を用いた標本調査作業の処理手順について検討を行う一般的な標本調査処理の概略ステップは以下のようなものである。

- (1) 調査目的の設定
- (2) 設問項目の設計
- (3) 調査票の作成
- (4) 調査票の配布と回収
- (5) データファイルの作成
- (6) 各種統計処理

*岡山理科大学理学部応用数学科

**情報システム学専攻

この中で本システムが直接関与する作業は(3), (5), (6)である。(2)はデータ処理プログラムの複雑さを決定する重要なステップであり, ここでは本システムの条件に合致した設計がなされることを前提として開発を進めた。

まず, 調査票の作成段階ではパーソナルコンピュータの高性能化とワードプロセッサ・ソフトウェアの機能充実により, 調査票の原稿をこれらで作成することが可能となった。その場合, 調査票の原稿ファイル中に各設問項目の性質を記録しておき, この情報をもとにしてデータ入力から各種統計処理まで, 個別プログラムの作成無しで一貫して処理する方法が考えられる。事実, 国内でよく利用されている日本語ワードプロセッサ“一太郎”等には文書内へのコメント挿入機能が用意されており, コメントが各設問属性(設問形式, 選択肢の個数や範囲など)の記述域として利用できる。同様の機能は英文処理系“Wordstar”にも備わっており, 英文アンケートの集計支援も同様に行える。ただし, この方式はワードプロセッサの改版に応じて, コメント部分の抽出方法を再確立しなければならない点が問題である。また, ページプリンタなどの高品位プリンタを利用すれば, 調査票あるいはその印刷原稿を直接出力できる場合もある。

次のデータ入力の段階では, 上で定義した調査票の設問項目の属性を取り込み, 出力データファイルの構造を自動的に決定して, ファイル化する。調査データ入力時にこのファイル構造に関する情報を参照すれば, システム主導型の会話形式の入力が可能となり, 字類検査や範囲検査が常時行えるので, 誤りの少ないデータ入力が保証される。当然, この属性ファイルはデータ訂正作業などにも広く応用できる。

各種の統計処理としては, 簡単な集計処理までを行うこととし, 高度な処理は統計パッケージを適用する方が有利である。したがって適当なファイル変換フィルタを用意すれば充分である。

3. システム構成

本システムは日本語 MS-DOS が動作するパーソナルコンピュータ上(開発は NEC PC-9801)での標本調査支援システムであり, そのデータフローを Fig. 1 に示す。ここで, 流線に添付した名前はプロセスを表している。開発言語は画面制御の容易さなどから C 言語を用いた。

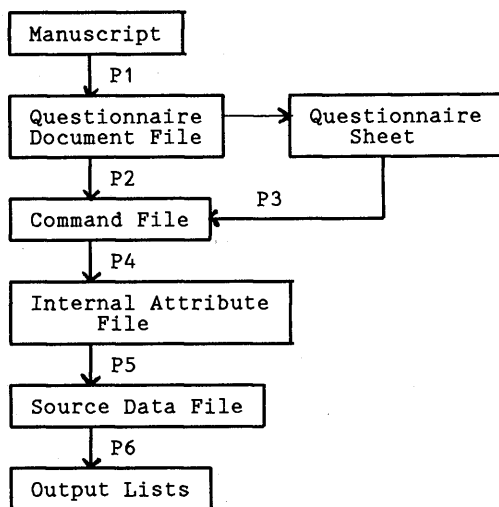


Fig. 1 System construction

P 1 は原稿から属性定義を含む調査票文書ファイルを作成するプロセスで, 市販のワードプロセッサ(現版では“一太郎”と“Wordstar”)を示す。具体的な文書ファイルの例は Fig. 4 に示した。

P 2 は文書ファイルから属性を定義したコメント行を抽出しファイル化する新規プログラム群で, 同じ結果を得る別手段が P 3 (市販のエディタ)によるコマンドファイルの直接作成である。

プロセス P 4 は, 以後の処理の高速化を図るため, コマンドファイルの正当性の検査しながら各コマンドを内部形式のコマンドファイル・レコードに変換する。コマンドエラーを検出すれば的確なメッセージを表示するので原ファイルの修正が容易に行える。

P 5 が内部属性ファイルを参照して Fig. 6 のようなデータ入力画面を作成し, 調査データを 1 レコード毎打ち込ませて原始データファイルを構築する入力プログラムである。各項目の入力エリアは枠組みで保護され, 右詰め固定長方式のデータ入力が容易に行えるようになっている。当然, ファイル作成の継続処理もサポートしている。

P 6 が単純集計を行うプロセスで, 内部属性ファイルを参照しながら, 項目毎の件数, 欠測データ件数, 選択肢の出現比率のグラフ表示, 最大値, 最小値, 平均値, 標準偏差などを表示する。(Fig. 8 参照)

4. 調査票および設問属性の記述

調査票の設問項目が要求する回答形式には多くのパ

タンが採用されているが、その構成規則は比較的単純である。例えば、各設問の回答にはせいぜい数桁の数字か文字が出現するだけで、長い文字列などが現れることはない。もちろん、自由記入の回答欄を設けた調査票もあるが、これは機械集計の対象には適さないものでここでは除外する。本システムでは調査票および設問項目の性質を記述するために以下の属性定義コマンドを設定した。

4.1 コマンドの一般形式

属性定義コマンドの一般形式は

#<コマンド名> /<属性>|

とする。ここで、<・・・>は構文要素、[・・・]はオプション、{|・・・|}は0回以上の反復を表わす超記号である。このように属性定義コマンドは、第一欄が#で始まる ANK 文字列で構成する。コマンド名が調査票の性質および各設問を識別し、後に性質を規定する属性指定が続く。

属性指定の一般形式は

<属性名>(<属性毎のパラメータリスト>)

*<反復回数>

とする。ここで、<属性名>はシステムで定めたキーワードで、各属性固有のパラメータリストと属性の反復回数が続く。

属性定義コマンドの出現順序は、調査票上の設問の順序に一致させる。調査票上では無回答の設問が現れてもよいが、連結順序にループ構造が現れないように編成しなければならない。

4.2 タイトル指定コマンド

T /<タイトル文字列>

データ入力画面や集計結果リストに表示するタイトル文字列を指定する。文字列が空白を含む場合は“ ”で囲んで空白を明示する。このコマンドはオプションであり、省略時には既定のタイトルが使用される。

4.3 欠測データ定義コマンド

M /<欠測データコード>

欠測データの入力に用いるコードを一文字指定する。デフォルト値は*である。本システムでは、必ずしも全てのデータ項目に欠測データ入力が必要とされるようになっていない。例えば、分岐質問に対し欠測データを入力すると、後続の質問が選択できないので許していない。このような状況が予測される場合には、「回答なし」の選択肢を準備する等の工夫が必要である。

4.4 ID 定義コマンド

ID /RNG(<始値>, <終値>)

/CNT(<最大レコード数>)

各調査票を識別する番号の範囲と、最大レコード数を指定する。本システムではデータレコードの最初のフィールドを ID 欄と限定している。現在の版では最大レコード数を積極的に利用していないが、今後の改訂によりファイル容量の推定などに用いる予定である。

4.5 設問の分類とコマンド名

調査票の設問を下記の通り分類し、各々に対応したコマンド名(設問名)の識別記号と意味を示す。実際のコマンド名はこの識別記号の後に一意名を付加して構成する。一意名は番号または適当な文字列で構成し、コマンド名全体で8文字以内とする。当然、全てのコマンド名は一意的でなければならない。例えば、

Q 1, BQA 2, SF 3

などである。

(1) 設問部

これは調査票の各設問に対する回答欄を定義する部分であり、設問の形式に応じて次の3種類に分ける。

共通設問: Q

全ての回答者が回答すべき必須部分を表わす。

分岐設問: BQ

この設問に対する回答状況に応じて、以降の適当な設問が選択される分岐設問を表わす。

副設問: SQ

分岐設問に従って一部の回答者だけが回答する設問を表わす。

(2) フェースシート

これは回答者の性質を規定する部分である。設問部と同様に、分岐の発生を許すので次のように分類する。

フェースシート共通設問: F

全ての回答者が回答する設問を示す。

フェースシート分岐設問: BF

回答者の性質を細分するために分岐を設けた回答欄である。

フェースシート副設問: SF

BFによって細分された一部の回答者だけが答える欄を示す。

4.6 共通設問定義コマンド

Q<一意名> /<回答形式指定>

全ての被調査者が回答すべき共通質問を定義する。<回答形式指定>は設問に対する回答の形式を定義する部分であり、4.13節で詳細に述べる。

4.7 分岐設問定義コマンド

```
#BQ<一意名> /<回答形式指定>
/SEL(<値の範囲>=<設問名>
|, <値の範囲>=<設問名>|)
```

分岐設問は回答内容により、次に答えるべき設問を特定する。したがって、分岐設問に対する回答は単一回答形式しか認めない。／SELの部分には全ての回答値に対して選択される接続先（設問名）を記述する。これは次節の／LNK定義と共に用いて、調査票の回答順序を定義する。設問の構造が定義できたので、データの入力時に理論上回答のない欄を自動的にスキップでき、入力作業の省力化が可能になる。

左辺の<値の範囲>は回答が取りうる値を示し
<値>

または

<値の範囲の下限>:<値の範囲の上限>

の構文に従うものとする。右辺の<設問名>は各設問を識別するコマンド名

```
Q<一意名>, BQ<一意名>, SQ<一意名>,
F<一意名>, BF<一意名>, SF<一意名>
```

であり、必ずこの設問より後に出現する設問を指定しなければならない。等式の組により、回答値が取り得る全ての値と各々の場合に後続する設問名を重複無く指定しなければならない。

4.8 副設問定義コマンド

```
#SQ<一意名> /<回答形式指定>
/LNK(<設問名>)
```

分岐質問により場合分けされた回答経路を構成する設問名を記述する。ただし、この設問名は当該質問より後方に存在する設問名でなければならない。

4.9 フェースシート共通設問定義コマンド

```
#F<一意名> /<回答形式指定>
```

性別、年齢など全ての回答者に関する共通設問を定義する。エラーチェックの対象とはならないが、原則的にフェースシートは調査票の最後に存在することを想定している。

4.10 フェースシート分岐設問定義コマンド

```
#BF<一意名> /<回答形式指定>
/SEL(<値の範囲>=<設問名>
|, <値の範囲>=<設問名>|)
```

回答者の性質を細分する場合に適用され、基本的に分岐設問と同様である。

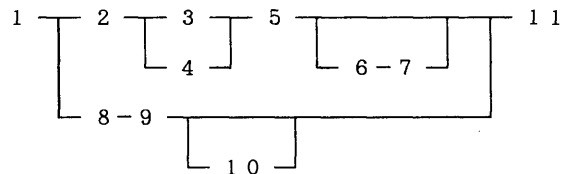
4.11 フェースシート副設問定義コマンド

```
#SF<一意名> /<回答形式指定>
/LNK(<設問名>)
```

BFにより場合分けされたフェースシート設問を定義し、SQと同様に扱われる。

4.12 分岐設問コマンドと副設問コマンドの関係

誤りの無い調査データを入力するとき、先頭の設問から最後の設問まで常に後戻りなく入力させるには、各設問を有向グラフのノードとみて、これらのノードの集合に半順序関係を定義できればよい。調査票がトポロジカルソートされたグラフの一例であるかどうかを判定するために、分岐設問と副設問に接続関係を表す属性／SELと／LNKを設けた。現実の調査票ではあまり複雑な分岐構造は出現しないが、例として2重の分岐を含む場合を Fig. 2 に示す。



```
#BQ 1 /SEL (··=BQ 2, ··=SQ 8)
#BQ 2 /SEL (··=SQ 3, ··=SQ 4)
#SQ 3 /LNK (BQ 5)
#SQ 4 /LNK (BQ 5)
#BQ 5 /SEL (··=SQ 6, ··=Q 11)
#SQ 6 /LNK (SQ 7)
#SQ 7 /LNK (Q 11)
#SQ 8 /LNK (BQ 9)
#BQ 9 /SEL (··=Q 11, ··=SQ 10)
#SQ 10 /LNK (Q 11)
#Q 11 /
```

Fig. 2 Link structure of questions

4.13 回答形式指定

回答形式は単一回答と複数回答に分けられる。以下にそれぞれの構造と意味を定義する。

4.13.1 単一回答形式

```
SA(<回答値の種類>, <回答値の範囲指定>)
[*<反復数>]
```

単一回答形式とは1個の値だけを回答する方式で、<回答値の種類>には以下の6種類が指定できる。それぞれの場合の()内の記述形式を示す。

(1) 番号:D, [<始値>:<終値>]

整数値による番号を選択する形式であり、最も多く

使用される。〈始値〉が省略されると1を仮定する。入力時の桁数は、〈始値〉または〈終値〉から必要最小限の桁数が自動的に割り出される。

(2) カナ：K, 〈選択肢の個数〉

これはアイウエオ順の仮名文字を1桁だけ回答させる形式である。最大45個までの選択肢が1桁で指定できる。ただし、先頭の値は常に'A'とする。

(3) 英字：A, 〈選択肢の個数〉

これは仮名文字の代わりに英字を用いる形式で、最大26個の選択肢が指定できる。本システムでは大文字と小文字の区別はしない。先頭は'A'である。

(4) 整数：I, 〈下限〉, 〈上限〉[, 〈桁数〉]

具体的に整数値を回答させる場合に使用する。オプションの〈桁数〉は、整数値の入力時に常に復帰改行キーで打ち込みを終わらせるなどリズムミカルな入力操作に対処できるよう、必要桁数より余分に入力桁数を指定させる目的で設けた。

(5) 実数：R, 〈下限〉, 〈上限〉, 〈桁数〉

実数値を回答させる設問に適用する。実数の場合、入力欄の桁数は必須である。

4.13.2 複数回答形式

MA (〈回答値の種類〉, 〈回答値の範囲指定〉, 〈最小回答個数〉, 〈最大回答個数〉)
[*〈反復数〉]

複数回答形式とは、限定された複数個あるいは任意個の回答を要求する設問を規定する形式である。この場合、〈回答値の種類〉と〈回答値の範囲指定〉は次の3通りしか取れない。これらは単一回答形式と同一である。

(1) 番号：D, [〈始値〉:]〈終値〉

(2) カナ：K, 〈選択肢の個数〉

(3) 英字：A, 〈選択肢の個数〉

回答個数の設定欄にはその最小個数と最大個数を指定する。例えば、選択肢がAからJまであり、この中から3個まで選んで回答する場合には

MA (A, 10, 0:3)

と指定する。

単一回答および複数回答の両者に共通した*〈反復回数〉は、同一形式の部分設問を集めて1個の設問にまとめた場合の、部分設問の個数を指定する。例えば

(a) 回答1 (1.良い 2.普通 3.悪い)

(b) 回答2 (1.良い 2.普通 3.悪い)

のように1つの設問中に同じSA形式の部分設問を2

<input type="text"/>	Single answer, no iteration
<input type="text"/> <input type="text"/> <input type="text"/>	Single answer, 3 iterations
<input type="text"/> <input type="text"/> <input type="text"/>	Multiple answer (Max 3 answers), no iteration
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	Multiple answer (Max 2 answers), 2 iterations

Fig. 3 Examples of input fields

個含む場合

SA (D, 3) * 2

と指定する。多くの場合、部分設問は1個であるから省略できる。

入力時の画面上で各設問形式は Fig. 3 のように表現されるので、入力者の視認性が向上する。

4.14 コメント

##〈文字列〉

これによりコマンド列の中に注釈を挿入できる。コメント行はシステムの処理に何ら影響を与えない。

5. 調査票の諸特性

本システムの処理条件を決定する際の資料として、既に公表されているいくつかの世論調査やアンケート調査票の性質を調査した。その結果を Table 1 に示す。調査対象は総理府発行の「世論調査」誌の調査票および6節で取り上げる例の29件である。

Table 1 Properties of questionnaires

Items	Mean	Min	Max	Ex. 1	Ex. 2
# of questions	20.9	10	44	44	20
# of face shts.	6.3	0	17	13	5
# of BQ cmnds.	3.6	0	11	11	4
# of BF cmnds.	0.4	0	3	3	0
depth of BQ/BF	1.2	0	2	2	1
max selectors	12.9	6	22	18	6
max iterations	3.7	1	9	6	4
# of mult. ans.	7.1	0	20	20	4
record length				270	43

Number of samples : 29

Ex. 1: 「世論調査」 Vol. 20, no. 2, 1988

“Food life and role of rural district”

Ex. 2: “Political reformation”, The Yomiuri

Table 1 の例 1 は調査対象の内で最も規模の大きかった例であるが、画面構成の制限を越えるため本システム（現版）の処理能力をオーバーしている。ただ、この規模まで大きくなると作表などの面でもパーソナルコンピュータには適さないケースだと言えよう。

6. 具 体 例

ここでは実際のアンケート調査票とその属性定義ファイルおよびデータ入力画面を示す。

6.1 世論調査の例

この例は読売新聞本社が実施した政治改革に関する全国世論調査（1989年5月3日付朝刊に掲載）の質問票を本システムで表現したものである。この例は規模が本システムにとって典型的であり、また多くの回答形式も使用されているので選択した。Fig. 4 が設問属

```
## 読売新聞社の世論調査
#T/ "政治改革に関する世論調査 (1989/4)"
#ID/ RNG(1, 3000)/ CNT(3000)
(1) あなたは、いまの国の政治に満足していますか
満足していませんか
①非常に満足している ②多少は満足している
③やや不満だ ④非常に不満だ 回答欄 [ ]
#Q1/ SA(D, 4)
(5) 国会議員や地方議員が冠婚葬祭や地域の行事に、
ご祝儀や香典、花輪などの金品を出すことについて、
あなたは、出すのが当然だと思いますか、場合によ
ってはやむを得ないと思いますか、それとも出す必
要はないと思いますか
①当然だ =>質問 (6)へ
②場合によってはやむを得ない=>質問 (6)へ
③必要ない=>質問 (7)へ 回答欄 [ ]
#BQ5/ SA(D, 3)/ SEL(1=SQ6, 2=SQ6, 3=SQ7)
(6) 【前問で「当然だ」「場合によってはやむを得
ない」と答えた人だけ】 そう思う理由を、次の中
から、いくつでもあげて下さい
① [ ] 投票に対する政治家からの見返りだから
② [ ] 政治家と親しい関係を持てるから
③ [ ] 政治家は金持ちだから
④ [ ] 世の中の慣習だから
⑤ [ ] その他
#SQ6/ MA(D, 5, 1:5)/ LNK(Q8)
(9) 政治家への献金のあり方について、次のあげる
意見に、①賛成か、②反対か、順にお答え下さい
1. 企業や労働組合などからの献金を一切禁止すべ
きだ 回答欄 [ ]
2. 献金を受けるために政治家が設けている政治団
体の数を限定すべきだ 回答欄 [ ]
3. 法律で許されている献金の金額の枠を経済の成
長にあわせて拡大すべきだ 回答欄 [ ]
4. 公開（報告）を義務づけられている献金額の基
準を引き下げるべきだ 回答欄 [ ]
#Q9/ SA(D, 2)*4
【フ ェ ー ス シ ー ト】
(2) 年 齢
1. 20歳代 2. 30歳代 3. 40歳代 4. 50歳代
5. 60歳代 6. 70歳代以上 回答欄 [ ]
#F2/ SA(D, 6)
```

Fig. 4 An example of questionnaire sheet

```
## 読売新聞社の世論調査
#T/ "政治改革に関する世論調査 (1989/4)"
#ID/ RNG(1, 3000)/ CNT(3000)
#Q1/ SA(D, 4)
#Q2/ SA(D, 2)
#Q3/ SA(D, 2)
#Q4/ SA(D, 2)
#BQ5/ SA(D, 3)/ SEL(1=SQ6, 2=SQ6, 3=SQ7)
#SQ6/ MA(D, 5, 1:5)/ LNK(Q8)
#SQ7/ MA(D, 5, 1:5)/ LNK(Q8)
#Q8/ SA(D, 3)
#Q9/ SA(D, 2)*4
#Q10/ SA(D, 3)
#Q11/ SA(D, 3)
#Q12/ SA(D, 3)
#Q13/ SA(D, 3)
#BQ14/ SA(D, 2)/ SEL(1=SQ15, 2=Q16)
#SQ15/ SA(D, 4)/ LNK(Q16)
#Q16/ SA(D, 3)
#BQ17/ SA(D, 2)/ SEL(1=SQ18, 2=BQ19)
#SQ18/ MA(D, 5, 1:2)/ LNK(BQ19)
#BQ19/ SA(D, 4)/ SEL(1:2=F1, 3:4=SQ20)
#SQ20/ MA(D, 5, 1:2)/ LNK(F1)
#F1/ SA(D, 2)
#F2/ SA(D, 6)
#F3/ SA(D, 6)
#F4/ SA(D, 12)
#F5/ SA(D, 5)
```

Fig. 5 An example of command file

性を含む調査票文書ファイルの一部を示している。こ
こから抽出された属性定義ファイルを Fig. 5 に示す。

この世論調査のデータを入力する時の画面構成を
Fig. 6 に示す。図の枠内が個々のデータを打ち込む
領域である。通常、データとして数値（字）が要求さ
れるので、基本的にはキーボード上のテンキーとカー
ソル移動キーのみで入力作業が進行するように設計さ
れている。また、範囲外のデータを入力したり、間違
ったキーを入力した場合、警告としてベルを鳴らし
カーソルが次の枠へ移動しない等の誤操作防止対策が
組み込まれている。

6.2 アンケート調査の例

この調査は本システムを用いて実施した具体例であ
る。調査件数が10,000件に近く、データ量のかかなり多
い調査であったが、データ入力段階で本システムが有
効に機能した例である。この調査は設問形式がすべて
単一回答形式であるので、属性定義ファイルの一部の
みを Fig. 7 に、単純集計結果の一部を Fig. 8 に示す。
ただし、この図はファイルに出力された印刷イメージ
の右端を削除してある。

7. 結 論

各種の世論調査やアンケート調査の調査作成から

Fig. 6 An example of data input screen

Fig. 7 Command file of exmple 5.2

Fig. 8 An example of output result

参考文献

- 1) H. Kimura, M. Ichimura, Ch. Asano: Sampling Survey Support Systm in a Microcomputer, COMPSAT 88, p.143 (1988).
- 2) 総理府広報室：世論調査，**20**，No. 1-12 (1988).
- 3) H. Kimura, Ch. Asano: The First Version of Micro-NISAN System, Proc. 3-rd Japan-China Symp. on Stat., p.298 (1989)
- 4) 林 知己夫監修：調査とサンプリング，p. 114，同文書院，(1985)。
- 5) 池田 央：統計調査のコンピュータ解析，p. 8，東洋経済新聞，(1971)。