

統計数理研究所
研究レポート51

クラスター分析プログラム・パッケージ

MINTS - 80

(Mini - Numerical Taxonomy System)

— 利用の手引 —

1980年5月

統計数理研究所

クラスター分析プログラム・パッケージ

MINTS - 80

(Mini - Numerical Taxonomy System)

－ 利用の手引 －

統 計 数 理 研 究 所

東京都港区南麻布 4 - 6 - 7

電話 (0 3) 4 4 6 - 1 5 0 1

この研究レポートは、「クラスター分析プログラム・パッケージ」の研究
成果報告書として作成されたものである。

当研究所では、

Annals of the Institute of Statistical Mathematics

統計数理研究所彙報

Computer Science Monographs

を発行している。

研究レポートは主として研究調査の成果の発表を目的とし、必要に応じて発行する。

(★) 本レポート内に掲載の内容を、ご利用希望の方は、事前に書面にて第6研究部までご連絡下さい。

ALL RIGHTS RESERVED.

NO PART OF THIS PUBLICATION MAY BE REPRODUCED OR TRANSMITTED IN ANY FORM OR BY ANY MEANS, ELECTRONIC OR MECHANICAL, INCLUDING PHOTOCOPY, RECORDING, OR ANY INFORMATION STORAGE AND RETRIEVAL SYSTEM, WITHOUT PERMISSION IN WRITING FROM THE INSTITUTE.

クラスター分析プログラム・パッケージ

MINTS

(Mini-Numerical Taxonomy System)

—利用の手引—

1980年5月

統計数理研究所

大 隅 昇

〔 目 次 〕

1. MINTS とは	1
2. MINTS の利用法	1
2.1 MINTS 命令文, MINTS ジョブ, MINTS ファイル	1
2.2 MINTS 命令文の構造	2
2.3 記号・記法の一般規則について	2
3. MINTS ジョブ制御命令文について	4
4. MINTS ジョブの標準的な利用法	8
HIERARCHICAL / 標準的な階層的クラスター分析	10
CLUSTERING FOR CATEGORICAL DATA / カテゴリカル・データ の階層的分類	13
FUZZY CLUSTERING / ファジィ・クラスタリング	16
KMEANS / 分割型クラスター分析	17
I SODATA / 分割反復最適化型クラスター分析	19
ASSOCIATION ANALYSIS / 二値型データの分枝型分類法	22
MAID / 多変量データ逐次二分法	23
PCA / 主成分分析法	25
5. 例題集, 付表等	27

1. MINTSとは

MINTS (MIni Numerical Taxonomy System) は、主にクラスター分析の解析処理を行うために開発したプログラム・パッケージである。その特徴として、次の事項を挙げることができる。

- 1) 独自の簡略化された命令文を持ち、利用者はこれを指定するだけで解析を行うことができる。
- 2) 多数組のデータを複数の課題（手法）により一括処理することが可能である。
- 3) 命令文の方式は、ほとんど自由書式であり、必要なキーワード、数値キーを簡単な規則に従って指定するだけでよい。また、それらの前後に任意の文章（英文）を挿入してもかまわない。
- 4) 標準値（default value）が設定されているので、これによる実行の場合には、命令文の指定はきわめて簡略である。
- 5) 一方、豊富なオプション機能も持つので、これを利用した、多様な分析も可能である。
- 6) 入力データの形式が多様である。すなわち、（サンプル）×（変数）の多変量構造データに限らず、主成分得点データ、距離行列、類似度行列などのデータまで多くの入力形式データが利用できる。
- 7) データの性格診断ルーチンをオプション機能として含む。クラスター分析は、手法によって、データの入力形式に様々な制約が起こる場合があるという性質を持っている。そこで、利用者が事前にデータの形式（多変量構造データ、距離行列、主成分得点データ、等）と、種類（カテゴリーカル、区間尺度、バイナリー型データなど）とを与えると、これと処理指定課題との相性を診断し、不適切の場合には、利用者に警告メッセージを呈示する。

2. MINTSの利用法

2.1 MINTS 命令文, MINTS ジョブ, MINTS ファイル

分析処理内容の指示は、すべてMINTS 命令文 により行う。命令文は原則としてカードで与えるが、このカード上には、処理の内容を指示する命令語とその内容明細を与える明細指示語が必要である。この両者は、必ず命令語、明細指示語の順に与える。

命令文の種類には大別して次の3種類がある。

- 1) MINTS ジョブ制御用
- 2) 分析課題名（手法名）指定用
- 3) その他

命令文はカードの集まりとしてセットされるが、これをMINTS 命令文カード・ブロックと呼び、処理の1単位（1ジョブの流れ）を MINTS ジョブと名づける。

1) はMINTS ジョブの実行に必要なデータの入力指示と作業ファイル上へのデータの格納、検証、課題ごとのパラメータ指定、オプション指定、ジョブ終了指示などを行うための命令文である。2) は、クラスター分析の各手法の課題名を指定する命令文である。3) は、1)、2) に含まれない命令文で、タイトル・カードと入力書式指定カードの2つがある。

なお、MINTS では、一旦入力用の命令文により読み込んだデータを作業ファイル（磁気ディスク）上に格納し、これを解析処理の指示順に、必要部分だけを読み取り処理を行うという解釈方式をとっている。この入力命令文により定義する原データのファイルをMINTS ファイルと呼ぶ。

2.2 MINTS 命令文の構造

MINTS 命令文の特徴は命令語と明細指示語を与えるだけですむという点にある。これをMINTS 内の解読ルーチンにより解読の上、誤まりがなければその内容に相当する副プログラムを呼び出した上で指定のデータの処理を行う。このとき、2つの解読方式を利用している。その1つは、キーワード解読方式であり、他の1つは、数値キー解読方式である。

キーワード解読方式とは、命令文カード上に与えた処理内容を指示する英文字を読みとり、その中からキーワードとして意味のある文字群をひろい出しこれを照合テーブルに照らしてその正誤を確認するという方法である。また、数値キー解読方式は、数値により指定した処理内容を判定表に照らして解読の上その適否を検証し処理を行うものである。このとき、処理の内容区分を与える機能キーとその中での具体的な条件を指定する条件キーとがある。

命令語はすべてキーワード解読方式であり、明細指示語は、両方式を用いている。前述のように、明細指示語は必ず命令語のあとに指定することが必要だが、キーワードと数値キーの指定の順(ならびの順)にはとくに指定の順位はないという点が特徴である。したがって、それらの間に任意の英文字、特殊文字(のほとんど)を置いてもよいという自由性を持つ。すなわち、必要なキーワード、数値キーを含んで、ある程度自由形式の文章化された(勿論英文のいみで)命令文を与えることができる。

なお、命令文の基本的な部分を一覧表として表1に整理してある。これらの表と、次に述べる命令文の基本的な指定規則を知れば、MINTS を容易に利用できる。

2.3 記号・記法の一般規則について

このマニュアルを通じて使用する記号と記法の概略をまとめておく。

- 1) 命令文中の下線部は必ず指定すべき部分で、解読のキーとなる。
- 2) [] は省略可能であることを示す。
- 3) { } または命令文に対して与えられる □ により指定内容を表わす。
- 4) 小文字のアルファベットにより数値を表わす。
 a, b, c, d 等で実数値を表わし、 n, nn, m, mm, l, k で整数値を表わす。
- 5) 条件キーをかこう □ は、その中の条件キーが標準値であることを示す。
(→各課題のオプション機能一覧表を参照)
- 6) □ は少くとも1個以上の空白の指定が必要であることを示す。

表 1. MINTS 命令文の一覧表

種類	指 定 方 法
シ ョ ブ 制 御 用	<p><u>READ</u></p> <p><u>INPUT</u> □ <u>TAPE</u> { <u>nn</u> }</p> <p><u>INPUT</u> □ <u>FILE</u> { <u>nn</u> }</p> <p>[<u>CHECK</u> □ <u>FORM</u> = [<u>RAW</u>, <u>SCORE</u>, <u>SIMILARITY</u>, <u>DISTANCE</u>, <u>RELATION</u>] <u>TYPE</u> = [<u>CATEGORY</u>, <u>BINARY</u>, <u>QUANTITATIVE</u>, <u>MIXED</u>]]</p> <p><u>PARAMETER</u> <u>N</u> = <u>n</u>, <u>M</u> = <u>m</u> …… , [キーワード]</p> <p>(*) ここで, <u>n</u> = サンプル数, <u>m</u> = 次元数 である。</p> <p>[<u>OPTION</u> <u>n</u>₁ = <u>m</u>₁ , <u>n</u>₂ = <u>m</u>₂ , …… , [キーワード]]</p> <p>(*) ここで, <u>n</u>₁ , <u>n</u>₂ …… は機能キー, <u>m</u>₁ , <u>m</u>₂ …… は条件キーである。</p> <p><u>STOP</u></p>
分 析 課 題 用	<p><u>ASSOCIATION ANALYSIS</u></p> <p><u>CLUSTERING FOR CATEGORICAL DATA</u></p> <p><u>FUZZY CLUSTERING</u></p> <p><u>HIERARCHICAL ANALYSIS</u></p> <p><u>ISODATA</u></p> <p><u>KMEANS METHOD</u></p> <p><u>MAID</u> (MULTIVARIATE - AID)</p> <p><u>PCA</u> (PRINCIPAL COMPONENTS ANALYSIS)</p>
そ の 他	<p>タイトル・カード</p> <p>入力データの書式指定</p>

3. MINTS ジョブ制御命令文について

MINTSの特徴の一つに制御用命令文によって、データの入力、処理条件、データの診断などの処理の指定が可能であるという点があげられる。

MINTS ジョブ制御用の命令文として次の表 2 にあげた 6 種類がある。

表 2 ジョブ制御用命令語一覧表

命 令 語	機 能
READ	カード・データの入力指示
INPUT	ファイル・データの入力指示
CHECK	データの診断
PARAMETER	パラメータの指示
OPTION	オプション機能の指定
STOP	MINTS ジョブの終了を示す

READ 文, INPUT 文

[機能]

分析データの入力機器および機番の指定を行う。

入力機器あるいはデータの入力媒体として、磁気テープ、磁気ディスク、およびカード・データの 3 種類利用できる。

[指定方法]

入力媒体の種類に対応して次の 4 通りの指示方式がある。

READ

INPUT_TAPE [nn]

INPUT_FILE [nn]

INPUT_SCORE

上から順にカード、テープ、ディスク、主成分得点データを入力する指定書式に相当する。磁気テープまたは磁気ディスクからデータを入力する場合には、入力機番 (nn) を指定する。この指定は省略可能である。

[標準値]

- 1) 入力機番 nn を省略すると磁気テープの場合には nn = 09, 磁気ディスクの場合には nn = 10 が標準値として与えられる。

[制限事項]

- 1) カード・データを入力の場合、そのカードは READ 文に続いておく。
- 2) これらの入力指定命令文の省略は許されない。

CHECK 文

〔機能〕

各課題に対して、入力しようとする分析データの取扱いが可能か、すなわち課題とデータ形式の適合性をデータの形式 (FORM)、および種類 (TYPE) について吟味する。なお、この命令文はオプション機能であり省略が許される。

〔指定方法〕

分析データの形式および種類に対応して種々の指示方式の組み合わせが可能である (→表1参照)。なお次の例にみるようにキーワードFORMとTYPEの指定順に優先順はない。

〔例題〕

```
CHECK FORM = RAW, TYPE = QUANT
CHECK TYPE = BINARY, FORM = RAW
CHECK FORM = SCORE
```

判定の結果、指定した課題に対して入力データが適合しない場合には課題の処理を中止し、次の課題へ移行する。

〔制限事項〕

FORMとTYPEの両者を指定して、初めに指定したキーワードで不適切と判定されたとき、第2に指定のキーワードの処理は行わない。

PARAMETER 文

〔機能〕

この命令文は次の処理機能を持つ。

- 1) サンプル数の指定 (n)
- 2) 次元数の指定 (m)
- 3) その他のパラメータ (数値キー、キーワード)

〔指定方法〕

各課題において多小の違いがあるが、全体で7つの型がある。

表3 パラメータ文の標準書式

<型>	<指定書式>	<利用できる課題名>
I	PARAMETER n m	KMEAN, MAID以外で利用可能
II	PARAMETER n m l	FUZZY, PCA以外で利用可能
III	PARAMETER n m l k	ASSOCIATION, HIERARCHICAL, KMEANS のみで利用可能
IV	PARAMETER n (キーワード)	CLUSTER, FUZZYのみで利用可能
V	PARAMETER n l (キーワード)	CLUSTERのみ利用可能
VI	PARAMETER n n (キーワード)	FUZZYのみ利用可能
VII	PARAMETER n n l (キーワード)	CLUSTERのみ利用可能

ここで、 n はサンプル数、 m は次元数、 l および k はその他の数値キーである。

また、ここで(キーワード)は次の3種のデータ入力形式の識別コードである。

- 1) SIMILARITY
- 2) DISTANCE
- 3) RELATION

これらのキーワードの指定位置は、命令語のあとであればどの位置に挿入してもよい。

OPTION文

[機能]

各課題が備えているオプション機能を指定する命令文である。

[指定方法]

命令語OPTIONにつづいて、明細指示語を与える。OPTION文の明細指示語は次の構造をもっている。

- 機能キーと条件キー
- キーワード

ここで機能キーと条件キーは1対で1機能を表わし、必ず、この対の形式で指定する。

表4 OPTION文の標準書式

<型>	<指定書式>	<利用できる課題名>
I	OPTION $n_1=m_1$, $n_2=m_2$, ……	ASSOCIATIONを除くすべての課題
II	OPTION $n_1=m_1$, ……(キーワード)	KMEANSの場合に限る

上の表で、 $n_1, n_2, ……$ は機能キーを表わし、 $m_1, m_2, ……$ は条件キーを表わしている。

ここで、 (n_1, m_1) , (n_2, m_2) , ……という対と、その並びが処理条件の内容指定として意味がある。

(n_1, m_1) とは、各課題別に与えられたオプション機能一覧表の中のある処理内容 n_1 の実行の指定であり、そのときの具体的な処理条件を m_1 で指示する。

たとえば、

OPTION (1, 3) , (3, 1) , (2, 1) , ……

と与えると、まず初めに、機能キー1でその条件キー3の処理、次に、機能キー3で条件キー1の処理、……と

順に処理が進行する。

なお、OPTION文には、上の表にみるように、2つの型（Ⅰ型，Ⅱ型）があるが、Ⅱ型はKMEANSに限って利用できる。このとき、キーワードの位置は命令語OPTIONの後の機能キーと条件キーの対（ n_i, m_i ）の間のどこに位置してもよい。

なお、命令語、各数値キー、キーワードの間は、少なくとも1つ以上の、英文字で仕切られていなければならない。

〔標準値〕

課題別に用意した、オプション一覧表に挙げてある。

〔制限事項〕

- 1) 機能キーの指定は必ず整数値である。
- 2) 条件キーには整数値，実数値の2通りの指定が可能である。

STOP文

〔機能と指定方法〕

MINTSジョブの終了を指示する。必ずMINTSジョブの最後におく。

4. MINTS ジョブの標準的な利用法

MINTS ジョブの標準的な利用法

< 指定の優先順 >	< 処 理 の 内 容 >	
< 1 >	データ入力機器・機番の指定 (READ, INPUT)	
< 2 >	入力データ (CARD, TAPE, SCORE)	
< 3 >	分析課題名の指定 (課題 A)	} 第 1 課題 A の処理
< 4 >	[課題 A に対するデータ診断の指定 (CHECK)]	
< 5 >	課題 A のパラメータ・カード (PARAMETER)	
< 6 >	課題 A のタイトル・カード	
< 7 >	課題 A の入力データの書式指定	
< 8 >	[オプション機能の指定 (OPTION)]	} 第 2 課題 B の処理
< 9 >	分析課題名の指定 (課題 B)	
< 10 >	[課題 B に対するデータ診断の指定]	
⋮	⋮	} 第 3 課題 C の処理 (以下, 上の書式を反復指定する)
< k >	MINTS ジョブ終了の指示 (STOP)	

ここで、次の事項を守ることが必要である。

- 1) 印の命令文は省略できない (必ず指定する)。
- 2) [] 印の命令文は省略可能である。CHECK 文を指定しないときには、この機能は働かない。また、OPTION 文を指定しない場合には、各課題内に用意された標準値が与えられる。(→ 課題別のオプション一覧表を参照)。
- 3) 入力データがカード以外 (磁気テープ、ディスク上にある) のときには、データ入力機器、機番の指定の後に、第 1 課題の処理を指定する命令文カードを続けてよい。
- 4) STOP 文が現われるまで、MINTS ジョブは、指定課題の処理を順次続ける。なお、入力データの書式指定 (とくに、データの型)、オプションの条件はそれぞれ課題において異なるので、各課題の説明を参照されたい。
- 5) 上の書式は、1 組のデータセットを複数の課題で処理する形式を示している。
多数組のデータ・セットを複数 (あるいは 1 つ) の課題で処理する場合には、上の書式から STOP 文を除く、

ブロックを反復すればよい。

次に標準的なMINTSジョブの利用例をあげておく。

〔例題〕

① READ INPUT DATA

②

```
  +-----+
  |         |
  | DATA CARD |
  |         |
  +-----+
```

 [データ・セット(1)]

③

```
  [ HIERARCHICAL CLUSTERING BY COMBINATORIAL SCHEME.
    PARAMETER NUMBER OF SAMPLES=48, NUMBER OF DIMENSIONS=15
    < KENDALL'S DATA > ( THIS IS TITLE CARD )
    ( 3X,15F3.0 )
```

 データ・セット(1)
の課題指定

④ READ CARD

⑤

```
  +-----+
  |         |
  | DATA CARD |
  |         |
  +-----+
```

 [データ・セット(2)]

⑥

```
  [ KMEANS ANALYSIS BY MINIMIZING WITHIN CLUSTER SUM OF SQUARES.
    PARAMETER CASES ARE 150, VARIABLES ARE 4, AND CLUSTERS GENERATED 2 TO 3.
    FISHER'S IRIS DATA ( THIS IS TITLE CARD FOR FIRST TRIAL )
    ( 5X,4F5.2 )
    OPTION 1=0-FILE(6)=20, REPEAT ( THIS IS OPTION CARD )
```

⑦

```
  [ ISODATA METHOD
    PARAMETER SAMPLES 150 DIMENSIONS 4
    FISHER'S IRIS DATA ( SECOND TRIAL )
    ( 5X,4F5.2 )
```

⑧

```
  [ PCA ( PRINCIPAL COMPONENTS ANALYSIS )
    PARAMETER N=150, M=4
    *** FISHER'S IRIS DATA *** ( THIRD TRIAL )
    ( 5X,4F5.2 )
```

⑨

```
  [ INPUT SCORE DATA GENERATED FROM PCA.
    KMEANS PROCEDURE
    PARAMETER CASES 150, DIMENSIONS 2, INITIAL CLUSTER=2
    --- CLUSTERING INDIVIDUAL SCORES DERIVED FROM PCA OF FISHER'S IRIS DATA ---
    ( 2F8.4 )
```

⑩ STOP

[各ブロックの説明]

- ①：データ・セット(1)の入力指示
- ②：データ・セット(1)
- ③：データ・セット(1)の課題指定
- ④：データ・セット(2)の入力指示
- ⑤：データ・セット(2)
- ⑥：データ・セット(2)による第1課題の指定
- 以下⑦, ⑧, ……で第2, 第3, …… 課題の指定
- ⑩：MINTSジョブ終了の指示

[処理内容の概略]

- 1) データ・セット(1)を入力し, (①, ②), 階層的(HIERARCHICAL)手法を実行する。処理条件はすべて標準値を用いる(③)。
- 2) 次にデータ・セット(2)を入力する(④, ⑤)(このときデータ・セット(1)は消去されデータ(2)が格納される)。まずk-means法(KMEANS)を行う(⑥)。初期クラスターを2個から最大クラスターを5個と指定して分類を行う(PARAMETER)。オプション機能を用いてクラスタリングを反復試行する(OPTION)クラスタリングの結果を機番20のディスク上に出力する(I-O-FILE(6)=20で指示)。
- 3) 次にデータ・セット(2)の第2課題としてISODATAを実行する。このときすべて標準値を用いる(⑦)。
- 4) さらに, データ・セット(2)の第3課題として主成分分析(PCA)を行い(⑧), その結果えられる主成分得点データの第2主成分までを入力データとしてKMEANSを実行する(⑨)。この場合もすべて標準値を用いる。

HIERARCHICAL [標準的な階層的クラスター分析]

[標準書式]

```
HIERARCHICAL
PARAMETER N = n , M = m [ MAXCLUSTER = l , METHOD = k ]
タイトル・カード
入力データの書式指定
```

ここで n = サンプル数
 m = 次元数
 l = 最大クラスター数
 k = 手法の指定 (→オプション機能参照)

である。

この課題により, 標準的な, 階層的クラスター分析によるクラスタリングを行なうことができる。

特徴的な機能として次の事項が挙げられる。

- 1) クラスターごとの各種統計量(平均ベクトル, クラスター内分散, メンバシップ・リストなど)の豊富

- な出力内容をもつ。また、これらの出力の水準をオプションにより制御できる。
- 2) クラスタ数を目安を与えるための“クラスタ数評価基準”をいくつか算出する。これによりクラスタ数をいくつかとするか見当をつけることができる。
 - 3) 乱数を利用して、手法の“感度分析”を行うことができる(シミュレーション機能)。

すなわち与えられたデータ布置に対し、各サンプルにノイズを付加して、この布置を攪乱し、それに対してクラスタの個数がどのように変化するかその挙動を追跡することができる。これにより、まとまりのよいクラスタの存在の有無の手掛りを得るとともに、各手法の与えられたデータに対する適合性を評価することも可能である。
 - 4) 8つの手法を任意に指定することができる(1種類または多種類のいずれも可能である)。

[制限事項]

- 1) l と k の関係について

手法キー k だけを指定することができるが最大クラスタ数 l だけの指定は許さない。 l の指定は k と対にして指定することが必要である。

なお、 l および k を指定しないとき、標準値はそれぞれ次の値が与えられる。

$$l = \min \{ \sqrt{n}, 20 \}$$

$$k = 1 \text{ (single linkage)}$$
- 2) シミュレーション機能はサンプル数が少ない場合に利用することが望ましい($n \leq 100$ 程度)。 n が大であると計算時間を費すので、注意を要する。
- 3) 手法の指定に関する注意

1つの手法だけを利用する場合には、 k を指定すればよい。このとき k の数値は、オプション機能一覧表の機能4にあげた条件キーと同じ数値キーを利用する。

複数個の手法を指定したい場合には、まず k で第1番目に利用する手法のキーを指定し、次にオプション機能の機能キー-4を使って最終の手法のキーを指示する。したがって、つねに、 $k \leq$ (オプション機能キー-4で指定の条件キーの値 k^*)でなければならない。

このように与えると第 k 番目の手法から k^* までの手法をすべて実行する。

[例題]

```

HIERAR
PARAMETER N = 100 , M = 3 , METHOD = 2
DEMO-DATA (1)
( 3 F 2 . 0 )
OPTION ( 4 , 8 )

HIERAR
PARAMETER N = 150 , M = 10
DEMO-DATA (2)
( 10 F 8 . 2 )
OPTION ( 4 , 5 )

```

第1の例は、手法2から手法8まで実行する。

第2の例は、 k を省略したため、 k には標準値1が与えられる。したがって、手法1から手法5まで実行する。

[オプション機能]

処理の内容	機能キー	条件キー	説明
出力条件の指定	1	-1 0 1 2 3 4 5	シミュレーション時の評価基準統計量のみ出力 最終情報(クラスター連結情報, クラスター数評価基準の一覧など) 0+(メンバーシップ・リスト) 1+(デンドログラム) 2+(各分割レベルでの全クラスター級内, 級間分散・共分散行列およびそのトレース) 3+(各分散レベルでの全クラスターおよび各クラスターの級内, 級間分散・共分散行列およびそのトレース) 4+(各分割レベルでのクラスター数評価基準の出力)
データの正規化の指定	2	0 1	正規化しない 正規化する
順位相関係数の算出指定	3	0 1	計算しない 計算する
手法の指定	4	1 2 3 4 5 6 7 8	Nearest Neighbor(Single Linkage) Furthest Neighbor(Complete Linkage) WPG(Weighted Pair-Group) Median Group Average Centroid Ward's Method Minimum Spanning Tree
シミュレーション回数の指定	5	0 n_s	シミュレーションは実行しない n_s 回シミュレーションを実行する
クラスター数評価基準量の算出	6	0 1	計算しない 計算する
付加する乱数の大きさの指定	7	0.0 a	$N(0, cI)$, ここで $c = \sigma / a^2$ またここで $\sigma = \max(\text{データの分散})$

CLUSTERING FOR CATEGORICAL DATA (カテゴリカル・データの階層的分類)

(標準書式)

CLUSTER
PARAMETER $N=n$, $M=m$, [**METHOD= ℓ**] (利用する親縁性の種類の指定キーワード)
 タイトル・カード
 入力データの書式指定

ここで n = サンプル数
 m = 次元数
 ℓ = 手法の指定

である。

- (1) この課題は主としてカテゴリカル・データの階層的クラスタリングに適した課題であり、つぎの特徴を持っている。
- a) 入力データの型は**実数型**、**整数型**のいずれも扱うことができる。
 - b) 入力データの型の識別(実数型か、整数型か)は、入力書式カードの指定内容の解釈により自動的に設定される。すなわち、実数型書式を与えると実数型のデータと判定する。整数型についても同様である。
 - c) 距離行列や類似度行列などの親縁性データを扱うことができる。
 - d) 19通りの類似度や非類似度の指定が可能である(→オプション機能一覧表参照)。
- (2) 手法としては、次の表の7種の指定が可能である。

手法指定の数値キ- ℓ	手 法 名
1	Nearest Neighbor (Single Linkage)
2	Furthest Neighbor (Complete Linkage)
3	Group Average
4	WPG (Weighted Pair-Group)
5	Centroid
6	Median
7	Farris' Method

なお、手法の指定にあたって次の点に留意することが必要である。

- i) $\ell = 7$ は実数型データで距離を利用するデータに対してのみ有効である。
- ii) $\ell = 1$ および 2 については、入力データの形式指定との関係で次のような組み合わせが考えられる。

<親縁性の種類の指定キーワード>

	DISTANCE	SIMILARITY, RELATION
$\ell = 1$	Single-Linkage	Complete-Linkage
$\ell = 2$	Complete-Linkage	Single-Linkage

すなわち、データ形式により、手法の切り換えが行なわれる。

- (3) 距離行列型データの入力時には、正方行列 ($n = m$) となる。このとき n (または m) の値だけ指定すればよい。

(サンプル) × (変数) の多変量型データで、サンプル数と次元数が等しい場合にも $n = m$ となるが、このときには n, m も指定が不可決である。

〔オプション機能〕

処理の内容	機能キー	条件キー	説 明
類似度・非類似度の指定	1	1	Distance for binary data 〔dissimilarity (0, ∞)〕(*)
		2	Simple matching coefficient 〔similarity (0, 1)〕
		3	Jaccard coefficient 〔similarity (0, 1)〕
		4	Czekanowski-dice coefficient 〔similarity (0, 1)〕
		5	unnamed coefficient (Sokal-Sneath) 〔similarity (0, 1)〕
		6	unnamed coefficient (Sokal-Sneath) 〔similarity (0, 1)〕
		7	Rogers and Tanimoto coefficient 〔similarity (0, 1)〕
		8	Kulczynski coefficient 〔similarity (0, 1)〕
		9	Sokal-Sneath coefficient 〔similarity (0, ∞)〕
		10	Haman coefficient 〔similarity (-1, +1)〕
		11	dot product (Russell-Rao coefficient) 〔similarity (0, 1)〕
		12	Kulczynski coefficient 〔similarity (0, 1)〕
		13	unnamed coefficient (Sokal-Sneath) 〔similarity (0, 1)〕
		14	Ochiai coefficient (cosine) 〔similarity (0, 1)〕
		15	unnamed coefficient (Sokal-Sneath) 〔similarity (0, 1)〕
		16	correlation 〔relational (-1, +1)〕
		17	Yule coefficient 〔relational (-1, +1)〕
		18	distance for binary data 〔dissimilarity (-∞, ∞)〕

		19	dispersion (relation (-1, +1))
データの正規化	2	0 1	正規化する 正規化しない
距離関数に対する 重みの指定	3	1 2 3 4	重みなし 次元数 m の 2 乗の逆数を重みとする 各変数の標準偏差の逆数を変数ごとの重みとする サンプル数 n の $1/\sqrt{n}$ を重みとする
距離計算式の指定	4	1 2 3	ミンコフスキー型距離 (1 - 相関係数) / 2 Canberra metric
ミンコフスキー距 離のべきの指定	5	2.0 a	ユークリッド距離 ミンコフスキー距離 (a は正の実数)

<注>

(*) これは、非類似度 (dissimilarity) で、その値が区間 (0, ∞) に分布することを示す。他の () 内も、同様に解釈する。

FUZZY CLUSTERING [ファジイ・クラスタリング]

ファジイ関係を利用して、階層構造を作り出すクラスタリングの方法である。
一般書式は次のとおりである。

[標準書式]

<p><u>FUZZY</u> <u>PARAMETER</u> N=<u>n</u>, M=<u>m</u>, [親縁性の種類の指定キーワード] タイトル・カード 入力データの書式指定</p>
--

ここで n = サンプル数
 m = 次元数

である。

この課題には次の機能がある。

1) 入力データの形式として次のいずれかを指定する。

- a) (サンプル) × (変数) (通常の変量型データ行列)
 - b) 距離行列, 類似度行列, 関連行列
- b) の場合は, 上の書式で親縁性の種類の指定キーワードを次のように与える。

距離行列に対して, DISTANCE

類似度行列に対して, SIMILARITY

関連行列に対して, RELATIONAL

2) b) の場合対称行列, 非対称行列のいずれに対しても処理が可能である。しかし, 非対称行列を入力する場合は, 次のオプション機能の指定が必要である。

[オプション機能]

処理の内容	機能キー	条件キー	説 明
親縁性の種類の指定	1	0	対 称 行 列 の 入 力
		1	非 対 称 行 列 の 入 力 (類 似 度 , 距 離 , 関 連 行 列)

KMEANS [分割型クラスター分析 / k -means 法]

クラスター内平方和の最小化を行なう典型的な分割型の手法の一つである。

初期値の与え方やセントロイドの更新方法により解が一意に定まらない(局所最適である)ので、それを出来るだけ回避するための手当てが必要である。この課題の特徴として、

- 1) 初期分割の与え方を選択できる。
- 2) クラスタリングの反復試行機能を持つ。すなわち、初期条件を変えて利用者の必要とする回数だけ分類試行をくり返すことが出来る。(REPEAT機能)
- 3) クラスタリングの結果として与えられるメンバーシップ・リスト(各サンプルのクラスターへの所属のリスト)の情報をカードおよび一時作業ファイル上(磁気ディスク)に出力することができる。
- 4) クラスター分割の判定基準を目安に、過剰な分割を制御する機能を持つ。
- 5) クラスター数評価基準(アフィニティー基準)によりクラスター数の大まかな見積りが可能である。

[標準書式]

<p>KMEANS PARAMETER $N = \underline{n}$, $M = \underline{m}$, INITIAL-CLUSTER = $\underline{\ell}$, [MAX-CLUSTER = \underline{k}] タイトル・カード 入力データの書式指定</p>
--

ここで n = サンプル数
 m = 次元数
 ℓ = 初期クラスター数
 k = 最大クラスター数

である。

なお、この課題では、PARAMETER文として、(n , m , ℓ)の3つの数値キーを指定することが必要である。また最大クラスター数 k を省略した場合、 k の標準値として $k^* = \min\{\sqrt{n}, 20\}$ が与えられる。

[オプション機能]

処理の内容	機能キー	条件キー	説 明
メンバーシップの間リストの出力	1	0	出力しない
		1	出力する
入力データリストの出力	2	0	出力しない
		1	出力する
分割の判定条件	3	0	分割条件を与えない
		1	$w_i \leq \epsilon$ (w_i : 第 i クラスターの分散共分散行列式の値)
		2	$d_i \leq \epsilon$ (d_i : 第 i クラスターのクラスター内分散)
		3	$c_i \leq \epsilon$ (c_i : 第 i クラスターのクラスター・サイズ)

機能キー 3 で与えた条件に対応する処理	4	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3	<p>強制的に処理を続行</p> <p>処理を中断する</p> <p>第 i クラスターを削除して処理を続行</p> <p>第 i クラスター内のデータをそのクラスターの最近隣クラスターに合併して処理を続行</p>
クラスタリングの反復試行回数の指定	5	<input type="checkbox"/> 0 n_r	<p>反復しない</p> <p>n_r 回反復する (→REPEAT を指定すること)</p>
分類反復結果の分類情報 (メンバーシップ・リスト) の出力指定	6	<input type="checkbox"/> 11 nn	<p>機番 11 のファイルに出力する</p> <p>機番 nn のファイルに出力する。 $15 \leq nn \leq$ (計算機で使える機番の限界)。</p>
クラスタリング反復結果のカード出力	7	<input type="checkbox"/> -1 <input type="checkbox"/> 0 <input type="checkbox"/> 1	<p>(サンプル) × (反復回数) の行列形式で出力する</p> <p>出力しない</p> <p>反復ごとのメンバーシップ・リストをそのまま行ベクトル形式としてカードに出力する</p>
閾値の指定	8	<input type="checkbox"/> 0.0 ϵ	<p>機能キー 3 の閾値 ϵ を 0.0 とする</p> <p>機能キー 3 の閾値として ϵ を与える</p>

処理の内容	キーワード	説明
初期クラスターの配置の与え方の指定	<input type="checkbox"/> 与えない	与えたクラスター数 k までの整数を系統的に付与する。 (すなわち, 1, 2, 3, ..., k , 1, 2, 3, ..., k , 1, 2, ... と与える)
	GIVEN	初期クラスターの配置を利用者がカードで与える。このとき (16G 5.0) で突孔した初期クラスターカード群を OPTION カードの後に付ける必要がある。
	RANDOM	自動生成した乱数により, ランダムに与える。
クラスタリングの反復試行の指定	REPEAT	初期条件を変えて反復分類試行する。このとき最大クラスター数 k に対してだけ反復が行なわれる (これは RANDOM を指定のときのみ有効)。

ISODATA [分割反復最適化型クラスター分析 / ISODATA 法]

この課題は、クラスターの等質性の基準としてクラスター内の平方和を利用する典型的な方法の一つである。基本原理はまず全データを、適当に代表点を与えて複数の群に初期分割し、この群落に対して個々のデータの再配置とセントロイドの更新を適当な反復回数を満たすまでくり返すというものである。

この他、ISODATAの特徴として、次の諸点がある。

- 1) クラスター・サイズの制御機能 (クラスター・サイズの小さいクラスターを一時的に除去する機能をもつ)。
- 2) 各クラスターについて、次元方向別にデータのちらばりの大きさを調べて分割の有無を、次元別かつクラスター単位に吟味する (locally splitting)。
- 3) クラスター間距離を調べて基準に照らして近いクラスター同士はなるべく合併させる (lumping 機能)。
- 4) クラスターの個数の決定がある程度自動的に行なわれる。すなわち、事前に与えたクラスター数を固定せずに、ある範囲で変化させる。また、パラメータをいろいろ変えながら interactive に処理できる。
- 5) クラスターの等質性の基準としてクラスター内分散または平方和、クラスター間平均距離などを使うのでよく分離したクラスターが存在し、しかも各クラスターのサイズが比較的同程度のクラスターの検出、生成に適している。
- 6) 与えた典型 (事前に与えた試みの分類情報) と、えられたクラスターとの間のクロス表を出力表示する。
- 7) クラスターごとの初期平均ベクトル (初期セントロイド) を利用者が与えると、それを核とするクラスター生成が可能である。
- 8) OPTION 文と初期セントロイドカード群を 1 単位とするカードの組を必要な組数だけ与えると同一データについて初期セントロイドを任意にかえて、何組もの異なった処理を連続的に行うことができる。
- 9) 入力データが主成分得点データの場合には初期セントロイドの自動付与機能が働く。これはデータの入力機番が主成分得点用ファイルであることによって識別される。

[標準書式]

ISODATA
PARAMETER N=n, M=m, [CLASS=ℓ]
タイトル・カード
入力データの書式指定

ここで n = サンプル数

m = 次元数

$ℓ$ = 典型の入力指定

$ℓ = 0$ (典型の入力を行わない)
 (標準値)
 $ℓ = 1$ (典型を入力する)

である。

$ℓ$ を指定した場合には、データ中に典型データが与えられねばならない。このとき次の 2 つの形式が許される。

a) NTYPE, (X(i), i = 1, m)

b) (X(i), i = 1, m), NTYPE

ここで、NTYPE は典型データ、(X(i) (i = 1, 2, …, m)) は 1 サンプルに対する m 次元観測データ・ベクトルである。

このどちらの型であるかの識別は、上の a) または b) のデータの入力データ書式命令文のいずれかを指示すると、これを解釈して自動的に判別する。なお、NTYPE は整数型、X(i) は実数型データとして入力することが必要である。

〔オプション機能〕

処理の内容	機能キー	条件キー	説明
初期セントロイド数の指定	1	<input type="checkbox"/> 0 n_i	初期セントロイドとして原点を与える n_i 個の初期セントロイドを読み込む。このとき、OPTION文の後に初期セントロイド・ベクトルを与えるカードが必要である
希望するクラスター数の数	2	<input type="checkbox"/> 0 k	$k * = \min \{ \sqrt{n}, 20 \}$ を目安に分割する k 個を目安に分割する
一時的に除去したクラスターのクラスターサイズの上限	3	<input type="checkbox"/> 2 n_s	クラスター・サイズが2のクラスターは計算から一時除去する クラスター・サイズが n_s のクラスターは計算から一時除去する
合併時のクラスターセントロイド間の距離の対の取り上げる個数	4	<input type="checkbox"/> 0 n_e	取り上げない 対の組み合わせのうち、値の小さいものから n_e 個まで取り上げる（リンク・リストの生成と指定）
最大反復回数の指定	5	<input type="checkbox"/> 5 n_r	5回以内反復する n_r 以内反復する
セントロイド更新時のクラスターごとの最大反復回数	6	<input type="checkbox"/> 10 n_c	クラスターごとの、クラスター内反復は10回以内である 反復は n_c 回以内である
反復ごとの処理情報の出力	7	<input type="checkbox"/> -1 0 1 n_u	最終回にのみ出力 中間情報は出力しない 各回ごとおよび最終回に出力する n_u 回反復すること、および最終回に出力する
メンバーシップリストの出力指定	8	<input type="checkbox"/> -1 0 1 n_m	最終回に出力 出力しない 1回ごとおよび最終回に出力する n_m 回ごとおよび最終回に出力する
メンバーシップ・リストの一時作業ファイルへの出力	9	<input type="checkbox"/> 0 nn	出力せず 機番 nn 上に出力する
クラスターのセントロイド更新時のセントロイドの移動変化量の最大許容値	10	<input type="checkbox"/> 0.0 a	最大許容値 0.0 最大許容値を a とする ($a > 0$)
クラスター内データの各次元方向のバラツキの最大許容量 σ_{max} (標準偏差) の指定	11	<input type="checkbox"/> 1.0 b - c	許容量は 1.0 である 許容量は b である ($b > 0$) 全データを包含する超立方体の最大対角線の c ($c > 0$) 倍を許容量とする

合併 (lumping) パ ラメータの指定	12	$\boxed{1.0}$ d	クラスター・セントロイド間距離が 1.0 以下のクラスター は合併する クラスター・セントロイド間距離が d ($d \geq 0$) 以下のク ラスターは合併する
分割 (splitting) パ ラメータの指定	13	$\boxed{0.5}$ e	機能パラメータ 11 で与えた許容量の 0.5 倍 (あるいは e 倍) をクラスター・セントロイドの移動量として与える ($0.5 \sigma_{max}$ または, $e \cdot \sigma_{max}$)

ASSOCIATION ANALYSIS [二値型データの分枝型分類法]

[標準書式]

ASSOCIATION ANALYSIS
 PARAMETER N=n, M=m, [MAXCLUSTER=ℓ, [MINCLUSTER=k]]
 タイトル・カード
 入力データの書式指定

ここで、 n = サンプル数
 m = 次元数
 $ℓ$ = クラスタ数の最大許容量
 k = クラスタ・サイズの最小許容サンプル数

である。

[制限事項]

- 1) $ℓ$ の指定を省略して k だけを指定することはできない。
- 2) オプションとして入力データの項目 (変数) に対して 8 文字以内の項目名を与えることができる。項目名を指定するカード群は入力データ書式指定のあとに続けて置く。
- 3) 計数多値データ (カテゴリーカル・データ) を、2 値型のバイナリー・データに変換した上で、分類を行なうことができる (→オプション機能参照)。

この指定法は他の課題と若干異なるので、次に説明しておく。

$$\begin{array}{l}
 \textcircled{1} \left[\text{OPTION } n_1, n_2, \dots, n_i \right. \\
 \left. \begin{array}{l}
 m_1^{(1)}, m_2^{(1)}, \dots, m_{n_1}^{(1)} \\
 m_1^{(2)}, m_2^{(2)}, \dots, m_{n_2}^{(2)} \\
 \vdots \\
 m_1^{(i)}, m_2^{(i)}, \dots, m_{n_i}^{(i)}
 \end{array} \right. \\
 \textcircled{2} \left[\begin{array}{l}
 \vdots \\
 \vdots \\
 \vdots \\
 \vdots \\
 \vdots
 \end{array} \right.
 \end{array}$$

まず①で i 個の項目に対して、各項目ごとにそれぞれ何個のカテゴリーを 0 値に変換したいかその個数 n_i を指定する。つまり n_i は第 i 項目に対しては n_i 個のカテゴリーを 0 値に変換することを意味する。

つづいて②は①で指定した各項目内でどのカテゴリーを 0 にかえたいか、そのもとの数値コードを列挙する。このときの書式は固定で (20 I 4) であることが必要である。

たとえば第 1 の項目は、「 n_1 個を 0 値に変換するが、そのコードは $m_1^{(1)}, m_2^{(1)}, \dots, m_{n_1}^{(1)}$ である」と与える。他の項目についても同様である。

また指定したカテゴリー以外の数値コードはすべて "1" となり結果として、多値データが (0, 1) 型データに変換されたことになる。

- 4) データ入力書式は可変であるが、入力形式に若干の制約がある。

すなわち、NSAMPLE, (IX(i), $i=1, m$) の形式であることが必要である。ここで NSAMPLE (サンプル識別番号) は文字型変数 (Aタイプ)、分類データ IX(.) は整数型でなければならない。

MAID [多変量データ逐次二分除法]

[標準書式]

<u>MAID</u> <u>PARAMETER</u> N= <u>n</u> , PREDICTOR= <u>m</u> , DEPENDENT= <u>ℓ</u> <small>タイトル・カード</small> <small>入力データの書式指定</small>

ここで n = サンプル数
 m = 説明変数の次元数
 $ℓ$ = 外的基準の次元数

である。

AID法 (Automatic Interaction Detector method) は、カテゴリカル・データである説明変数群 X の情報を利用して、外的基準 (従属変数) Y として与えられた量的データを逐次二分する方法である。

標準的な AID 法では、Y として 1 次元データを与えるが、ここでは、さらに一般化して多次元の外的基準の場合も扱うことができる。なお入力データ形式として、次の 4 つの型がある。

- a) (Y(j), j = 1, ℓ), (I X(i), i = 1, m)
- b) (I X(i), i = 1, m), (Y(j), j = 1, ℓ)
- c) NSAMPLE, (Y(j), j = 1, ℓ), (I X(i), i = 1, m)
- d) NSAMPLE, (I X(i), i = 1, m), (Y(j), j = 1, ℓ)

ここで c), d) は a), b) の変形でサンプル番号 (NSAMPLE) まで入力する場合に相当する。

- 1) I X(·) は説明変数ベクトル, Y(·) が外的基準データである。ここで I X(·) は整数型, Y(·) は実数型で与える。
- 2) a) が標準形式でありその他の型のデータを入力する場合には、オプション機能の機能キー 1 および機能キー 7 を組み合わせて条件を指定することができる (→ オプション機能)。
- 3) 説明変数に対する名称付与の機能。

説明変数の名項目に対して名称を付与することができる。この名称は出力リストの中で利用される。名称を付与したい場合には項目の数(m)だけの枚数のカードを用意し、1 変数名に対し 1 枚のカードを使って 8 文字以内の名称を与える。このとき、その 8 文字がカード上の 80 クラム内のどこに位置していてもよい。

なおこのカード群は必ず、入力データ書式指定命令文につづいて置く。

この名称付与カードを与えない場合には、標準値が与えられる。

[オプション機能]

処理の内容	機能キー	条件キー	説明
データ入力形式の指定	1	<input type="checkbox"/> 0 1	a) 型または c) 型データのとき b) 型または d) 型データのとき
クラスター数の最大数	2	<input type="checkbox"/> 0 n_{max}	\sqrt{n} (サンプル数の平方根) n_{max} 個まで分割する
クラスター・サイズの最小数	3	<input type="checkbox"/> 5 n_{min}	クラスター・サイズが 5 以下のクラスターが現われるまで分割する クラスター・サイズが n_{min} になるクラスターが現われるまで分割する

クラスタリングの終了判定基準の指定	4	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4	機能キー 2, 3, 8, 9, のすべてを照合する 機能キー 2のみ照合する 機能キー 3のみ照合する 機能キー 8のみ照合する 機能キー 9のみ照合する
入力データの出力の有無	5	<input type="checkbox"/> 0 <input type="checkbox"/> 1	出力する 出力しない
説明変数の分割の条件の指定	6	<input type="checkbox"/> 0 <input type="checkbox"/> 1	各カテゴリごとに求めた外的基準の平均で順序づけし、順序尺度化する。 カテゴリのすべての2分割の組み合わせを調べる
3に付与したサンプル番号の入力指示	7	<input type="checkbox"/> 0 <input type="checkbox"/> 1	入力しない 入力する
TSSI/TSSTの値。全データの平方和(TSST)とクラスター内平方和(TSSI)の比の限界値の指定	8	<input type="checkbox"/> 0.08 <input type="checkbox"/> a	$(TSSI/TSST) \leq 0.08$ のとき計算終了 利用者が与えた $a (\geq 0)$ のとき計算終了
BSSI/TSSTの値。総平方和(TSST)と、ある2分割が行なわれたときのクラスター間平方和(BSSI)の比の限界値の指定	9	<input type="checkbox"/> 0.05 <input type="checkbox"/> b	$(BSSI/TSST) \leq 0.05$ のとき計算終了 利用者が与えた $b (\geq 0)$ のとき計算終了

PCA〔主成分分析〕

この課題指定により標準的な主成分分析法の処理が可能である。

なお、主成分分析の結果えられた主成分得点データを他の課題の入力データとしてクラスター化することが可能である。

〔標準書式〕

<p>PCA(PRINCIPAL COMPONENTS ANALYSIS)</p> <p>PARAMETER N=<u>n</u>, M=<u>m</u></p> <p>タイトル・カード</p> <p>入力データの書式指定</p>

ここで n = サンプル数

m = 次元数

である。

この課題を利用するときには次の点に注意しなければならない。

- 1) 変数が70次元以上 ($m > 70$) では利用できない。
- 2) 求めた主成分得点データを一時作業ファイル(磁気ディスク, 機番 $nn=04$) 上に出力する。このとき出力書式は(10 F 8.4) である。したがって得点データを他の課題で利用する場合の入力書式はこれを指定する。
- 3) 主成分の次元数は $\min \{ m, 10 \}$ で与えられる(すなわち最大で10成分まで)。

〔オプション機能〕

処 理 の 内 容	機能キー	条件キー	説 明
χ^2 検定結果の出力	1	<input type="checkbox"/> 0 1	出力しない 出力する
入力データ・リストの出力指定	2	<input type="checkbox"/> 0 1	出力しない 出力する
分散・共分散行列の出力指定	3	<input type="checkbox"/> 0 1	出力しない 出力する
正規化データ・リストの出力	4	<input type="checkbox"/> 0 1	出力しない 出力する
分散・共分散行列の行列式の値	5	<input type="checkbox"/> 0 1	出力しない 出力する
固有ベクトル行列の出力	6	<input type="checkbox"/> 0 1	出力しない 出力する
相関行列, 因子負荷量行列, 固有値, 寄与率の一覧表の出力	7	0 <input type="checkbox"/> 1	出力しない 出力する
因子負荷量の散布図の出力	8	0 <input type="checkbox"/> 1	出力しない 出力する(第1, 2, 3成分に対して, 3通りの散布図)

付 録

- 例 題 ， 付 表 -

TEST SCHEDULE (1) (ASSOCIATION ANALYSIS)

READ INPUT BINARY DATA

```
-----  
|          |  
| INSERT HERE DATA SET (A) |  
|          |  
-----
```

ASSOCIATION

PARAMETER N=50, M=5

ASSOCIATION ANALYSIS/ THIS IS THE FIRST TRIAL

(1X,A4,5X,5I1)

ASSOCIATION

PARAMETER CASES 50 DIMENSIONS 5

THIS IS THE SECOND TRIAL/ THE CASE WITH THE ADDITIONAL ITEM NAMES

(1X,A4,5X,5I1)

ITEM-A

ITEM-B

ITEM-C

ITEM-D

ITEM-E

INPUT DATA IS REGISTERED ON DISK

ASSOCIATION ANALYSIS

PARAMETER SAMPLES=48,DIMENSIONS=6

DEMO SCHEDULE BY KENDALL'S DATA DEFINED ON DISK

(A3,6I3)

OPTION 3,5,4,3,4,2

7	8	9		
1	2	3	4	5
0	7	8	10	
8	9	10		
3	4	7	8	
9	10			

STOP

TEST SCHEDULE (2) (CLUSTERING FOR CATEGORICAL DATA)

READ CARDS

```

+-----+
|       |
| INSERT HERE DATA SET (B) |
|       |
+-----+
  
```

CLUSTERING FOR BINARY DATA
 PARAMETER N=32, M=6
 THE RESULT OF THE EXAMINATION
 (2X,6I2)

CLUSCATEGORICAL DATA
 PARAMETER CASES ARE 32 ; VARIABLES ARE 6
 --- TITLE IS " THE RESULT OF TEST "
 (2X,6I2)
 OPTION SIMILARITY IS USED, AND CONDITION IS (1,6)

CLUSCATEGORY
 PARAMETER SAMPLES=32,VARIABLES=6
 --- SCHOOL DATA ---
 (2X,6F2.0)

READ DATA CARDS

1	HORI	2	1	2	2	2	3	1	2	2	2	1	2	1	2
2	SATOU	2	1	2	2	2	1	3	2	2	2	1	2	1	3
3	URANO	1	1	1	3	3	2	2	3	2	1	2	2	2	2
4	SHIINA	2	2	1	3	2	1	3	3	2	3	2	1	2	1
5	KAMODA	1	2	2	3	1	3	1	2	3	2	3	1	3	2
6	HAMANO	2	3	1	2	1	1	3	3	2	1	1	2	3	2
7	KURIYAMA	2	2	1	2	3	3	1	2	1	3	3	1	3	1
8	TANBA	1	1	3	2	3	1	3	1	2	1	2	2	3	1
9	HUNADA	1	3	1	3	2	1	3	3	1	3	2	1	2	3
10	NAGAE	1	3	1	2	3	1	3	1	1	2	3	3	3	3

CLUSTERING
 PARAMETER SAMPLES 10 / VARIABLES 14 / METHOD 3 /
 --- AN EXAMPLE OF VOTING DATA ---
 (1 7 X , 1 4 I 2)

READ / INPUT DATA IS A RELATIONAL-TYPE

100	80	80	40	20	10
80	100	85	70	40	30
80	85	100	40	30	25
40	70	40	100	75	65
20	40	30	75	100	70
10	30	25	65	70	100

CLUS CATEGORICAL ANARISYS
 PARAMETER CASES(6),DIMENSIONS(6) : SIMILARITY-TYPE
 TEST DATA OF VISUAL RESPONSE
 (6F4.2)

OPTION VALUES DIST(4,1),POWER=2.5,AND WEIGHT=3

READ A DISTANCE MATRIX

1	0.0	0.7	0.6	0.9
2	0.7	0.0	0.9	0.5
3	0.6	0.9	0.0	0.7
4	0.9	0.5	0.7	0.0

CLUSCATEGORY DISTANCE TEST
 PARA VARIABLES,CASES=4,TYPE IS DISTANCE
 CLUSTER ANALYSIS
 (5X,4F5.1)

OPTION NORMALIZE YES(2,1),POWER(5) IS 1.8

READ A CORRELATION MATRIX

1	1.0	-0.3	-0.4	0.2
2	-0.3	1.0	0.1	-0.4
3	-0.4	0.1	1.0	0.3
4	0.2	-0.4	0.3	1.0

CLUSCATEGORY : HIERARCHICAL CLUSTERING METHOD FOR CATEGORICAL AND BINARY DATA
 PARAMETER CASES 8 VARIABLES = 4 TYPE=RELATIONAL
 CLUSTER GENERATION FOR CORRELATION MATRIX
 (5 X , 4 F 5 . 1)
 OPTIONS ARE SPECIFIED THE FOLLOWING/(2,1),(4,3),(5,1.5)/
 STOP

TEST SCHEDULE (3) (FUZZY CLUSTERING)

READ INPUT DATA

1	11	15
2	20	20
3	31	32
4	34	35
5	40	42
6	45	46
7	51	42
8	56	34
9	25	25
10	30	28

FUZZY

PARAMETER N=10, M=2

EXAMPLE (A), CASES BY VARIABLES

(5X,2F5.1)

READ INPUT CARD

100	79	65	58	47	31	34	13
79	100	76	81	66	56	60	33
65	76	100	75	56	43	64	46
58	81	75	100	79	67	76	60
47	66	56	79	100	86	68	54
31	56	43	67	86	100	60	49
34	60	64	76	68	60	100	75
13	33	46	60	54	49	75	100

FUZZY CLUSTERING

PARAMETERS N OF CASES 8, N OF VARIABLES 8, SIMILARITY-TYPE

EXAMPLE (B), SIMILARITY MATRIX

(8F4.2)

READ DATA

1	0.0	0.3	0.4	0.2
2	0.3	0.0	0.1	0.4
3	0.4	0.1	0.0	0.3
4	0.2	0.4	0.3	0.0

FUZZY CLUSTERING

PARAMETER CASES, VARIABLES = 4, DISTANCE

EXAMPLE (C), DISTANCE TYPE

(5X, 4F5.1)

READ INPUT MEDIUM IS CARD IMAGE

1	1.0	-0.6	-0.4	0.8
2	0.3	1.0	0.1	0.4
3	0.4	0.1	1.0	0.3
4	0.2	-0.4	0.3	1.0

FUZZY TRANSITIVE CLOSURE

CHECK FORM=RELATIONAL

PARAMETER SAMPLES(N,N)=4, RELATION

RELATIONAL AND ASSYMMETRY DATA

(5X, 4 F 5.1)

OPTION PROXIMITY IS ASSYMMETRY, SO SPECIFY FUNC(1)=1

STOP

TEST SCHEDULE (4) (HIERARCHICAL CLUSTERING)

READ KENDALL'S DATA

```
-----  
| INSERT HERE KENDALL'S DATA |  
-----
```

HIERARCHICAL ANALYSIS

PARAMETER CASES 48, DIMENSIONS 15

*** KENDALL'S DATA ***

(3X, 15F3.0)

OPTIONS OUTPUT LEVEL (1,2) RANKCOR(3)=1 (CALCULATE), RANKCOR(2) = 1 (YES)

HIERARCHICAL

PARAMETER CASES BY VARIABLES, 48 X 15

--- KENDALL'S DATA ---

(15F3.0)

OPTION RANDOM-ADD(7)=0.1, OUTPUT LEVEL IS (1,3)

HIERARCHICAL CLUSTERING

PARAM SAMPLES 48/ VARIABLES 15/ METHOD=2

... KENDALL'S DATA

(3 X , 1 5 F 3 . 0)

OPTION METHOD(4)=6 (SPECIFIED METHODS ARE FROM TWO TO SIX)

INPUT FILE

HIERARCHICAL CLUSTERING METHOD BY COMBINATORIAL SCHEME

PARAM SAMPLES ARE 48, VARS ARE 15 AND METHOD 1 IS SPECIFIED

/ KENDALL'S DATA /

(3X,15F3.0)

OPTIONS METHOD IS CARRIED OUT UNDER THE SPECIFIED KEY(4,8)

STOP

TEST SCHEDULE (5) (ISODATA AND KMEANS)

```
-----  
| INSERT HERE FISHER'S IRIS DATA |  
-----
```

ISODATA ANALYSIS

PARAM 150/ 4/ CLASS=1/

DEMO SCHEDULE BY FISHER'S IRIS DATA

(15,4F5.2)

OPTIONS (7,1),(8,1),(1,4),(10,0.01),(12,0.9)

1.	0.	2.	4.
5.	2.	4.	3.
4.	3.	4.	5.
3.	5.	4.	5.

KMEANS METHOD

PARAM N=150 M=4 : INITIAL-CLUSTER=2,MAX-CLUSTER=10

FISHER'S DATA (IRIS-SETOSA,IRIS-VERSICOLAR,IRIS-VIRGINICA)

(5X,4F5.2)

OPTION REPEAT, MEMBERSHIP FILE(6)=15, MEMBERSHIP LIST(YES) (1,1)

INPUT FILE NO.15 (CLUSTERING-REPEAT DATA)

ISODATA METHOD

PARAMETER SAMPLES=150, DIMENSIONS=10

FISHER'S DATA : IRIS-SETOSA,IRIS-VERSICOLAR,IRIS-VIRGINICA

(10F2.0)

OPTIONS MERGE-LINK(4,1),MAX-ITERATIONS(12) IN EACH CLUSTER ARE 15

KMEANS ANALYSIS

PARA 150,4,INITIAL CLUSTER=6,MAXCLUSTER=10/

FISHER'S DATA

(5 X, 4 F 4 . 2)

OPTIONS ITER. NUMBERS(5) ARE 5, THRESHOLD(8) IS 0.05

INPUT FILE 11 RECLUSTERING THE RESULT OBTAINED BY REPEAT-TRIALS OF CLUSTERING

ISODATA

PARAM (N,M) = (150,5)

./ FISHER'S IRIS DATA

(5 F 2 . 0)

STOP

TEST SCHEDULE (6) (MULTIVARIATE-AID)

READ

```
-----  
| INSERT HERE AID TEST DATA (1) |  
-----
```

MAID

PARAMETERS N=24, PREDICTORS=2, M=2

< AID TEST DATA >

(F5.0,F5.2,2I5)

MAID PROCEDURE

PARAMETER CASES,24 / PREDICTOR,2 / DIMENSIONS,2 /

/ AID TEST DATA /

(5X,F5.0,F5.2,2I5)

OPTIONS MAXGROUPS (2,8), STOPPING-RULE IS (4,1), DATA-LIST IS YES(5,1)

READ DATA

```
-----  
| INSERT HERE AID TEST DATA (2) |  
-----
```

MAID ANALYSIS

PARA SAMPLES=50, PREDICTORS(4),VARIABLES(3)

DEMO DATA GENERATED FROM RANDOM NUMBERS

(15, 4I2, 3F7.2)

OPTIONS INPUT FORM(1)=1, SEGNUMBER(7)=1

MAID/MULTIVARIATE AUTOMATIC INTERACTION DETECTOR

PARA : N OF CASES = 50 ; PRED= 4 VARIABLES= 3

(RANDOM DATA)

(5X,4I2,3F7.2)

OPTIONS TSSI/TSST(8)=0.07, BSSI/TSSI(9)=0.06, INPUT-FORM IS (1,1)

READ CARD

```
-----  
| INSERT HERE AID TEST DATA (3) |  
-----
```

MAID : TEST DATA

PARAMETER = (CASES 50, PREDICTOR 13, VARIABLES 5)

<<< TEST DATA >>>

(1 5 , F 5 . 0 , F 5 . 1 , 3 F 5 . 0 , 5 X , 1 3 I 1)

* TEST-A 3

* TEST-B 4

* TEST-C 4

* TEST-D 5

* TEST-E 3

* TEST-F 3

* TEST-G 4

* TEST-H 3

* TEST-I 3

* TEST-J 4

* TEST-K 3

* TEST-L 2

* TEST-M 2

OPTIONS VALUE (2,10;7,1)

STOP

TEST SCHEDULE (7) (PRINCIPAL COMPONENTS ANALYSIS)

INPUT FILE

PCA

PARAMETER N=48, M=15

< KENDALL'S DATA >

(3X,15F3.0)

INPUT FILE ; < KENDALL'S DATA >

PCA PRINCIPAL COMPONENTS ANALYSIS

PARAMETER N=48, M=15

< KENDALL'S DATA >

(3X, 15F3.0)

OPTION EIGEN VECTORS=(6,1), CHISQ=(1,0)

INPUT TAPE NO.9

PCA IS PROCEDURE

PARAMETER SAMPLES (2500) DIMENSIONS (3)

TEST BY NORMAL-MIXTURE DATA

(5X,2X,F12.5,F12.5,F12.5/)

OPTION VALUES = SUMMARY(7,0)(YES), V-COV. MATRIX(3,1)(YES), NORM. DATA(4,1)(YES)

READ CARD

```
+-----+
|       |
| INSERT HERE FISHER'S IRIS DATA |
|       |
+-----+
```

PCA

PARAM CASES ARE 150, VARIABLES ARE 4

.... FISHER'S DATA

(5X,4F5.2)

OPTIONS ARE SPECIFIED AS FOLLOWS; (2,1), (5,1), (8,0)

INPUT SCORE

HIERARCHICAL CLUSTERING BY SCORE-DATA

PARAMETER CASES: 150 VARIABLES: 4 METHOD: 2

SCORE DATA DERIVED FROM FISHER'S IRIS DATA

(4F8.4)

OPTION IS GIVEN TO THE OUTPUT OF DENDROGRAM, (1,2)

STOP

DATA SET (A), ASSOCIATION ANALYSIS

1	00101	26	10110
2	00001	27	00010
3	10110	28	00111
4	11000	29	11000
5	01011	30	01100
6	01001	31	10000
7	10010	32	10000
8	10110	33	10000
9	00000	34	01101
10	10111	35	11101
11	00100	36	10011
12	01011	37	10000
13	11010	38	10001
14	10101	39	10100
15	10100	40	00000
16	11001	41	01100
17	00101	42	00000
18	11000	43	10110
19	00010	44	00011
20	10001	45	00111
21	10111	46	10011
22	01100	47	10000
23	10110	48	01110
24	10101	49	00101
25	11001	50	11100

DATA SET (B), CLUSTERING FOR CATEGORICAL DATA

1	1	1	1	1	0	0
2	1	1	1	0	0	1
3	1	1	1	1	0	0
4	0	1	0	0	0	0
5	0	0	1	0	1	0
6	0	1	0	0	0	0
7	0	1	0	0	0	0
8	1	1	1	1	0	1
9	0	0	0	1	0	0
10	0	0	0	1	0	0
11	0	1	0	1	0	0
12	0	0	0	0	0	1
13	0	0	1	0	0	1
14	1	0	1	0	0	1
15	0	0	0	1	1	1
16	0	1	1	1	0	0
17	0	1	1	1	0	1
18	1	1	1	0	1	0
19	1	0	0	0	1	0
20	1	0	0	0	0	0
21	1	0	0	0	0	0
22	0	1	0	0	0	0
23	0	0	0	1	0	0
24	0	0	0	0	1	0
25	0	1	0	1	1	0
26	1	1	1	1	1	1
27	0	1	0	0	0	0
28	0	0	0	1	0	0
29	1	0	1	1	0	0
30	0	0	0	1	0	0
31	1	0	0	0	1	0
32	0	0	0	1	0	0

DATA SET AID TEST DATA (1)

1	79.	2.56	5	3
2	20.	1.45	2	1
3	80.	1.98	6	2
4	15.	2.27	1	1
5	29.	2.24	3	2
6	78.	1.76	5	2
7	25.	4.63	2	1
8	27.	1.80	3	2
9	51.	1.49	4	3
10	48.	2.33	3	2
11	5.	1.01	1	1
12	26.	1.48	2	2
13	42.	2.53	3	2
14	73.	3.00	5	3
15	62.	2.72	5	3
16	84.	1.48	6	3
17	10.	1.71	1	1
18	9.	1.13	1	1
19	95.	3.14	6	3
20	61.	2.31	5	2
21	81.	3.01	6	3
22	0.	0.98	1	1
23	7.	1.12	1	2
24	55.	2.61	4	3

DATA SET AID TEST DATA (2)

1	0	1	0	3	2.75	3.12	2.49
2	1	1	1	0	5.16	4.72	5.50
3	2	1	0	1	6.95	6.28	7.37
4	0	0	0	0	0.70	-0.85	0.56
5	3	0	2	3	41.88	41.67	41.34
6	3	1	1	0	8.69	9.66	8.76
7	1	1	0	1	5.56	5.69	6.36
8	1	1	1	1	10.66	9.94	10.72
9	0	0	0	3	-0.41	-0.67	-1.15
10	0	0	1	0	0.23	1.88	1.02
11	0	0	0	2	-0.25	-0.69	-1.06
12	3	0	1	1	11.49	11.59	10.90
13	3	1	0	3	8.52	9.05	8.27
14	3	1	2	1	20.99	20.76	19.81
15	3	1	2	1	21.18	21.40	21.91
16	0	1	0	3	3.22	3.94	3.50
17	2	1	1	0	7.17	7.26	7.51
18	0	0	0	1	-0.04	-0.83	-0.16
19	3	0	0	1	5.86	6.17	5.93
20	0	0	1	2	12.74	12.51	13.26
21	3	0	2	0	5.75	5.61	5.89
22	1	0	1	3	19.99	19.21	19.46
23	2	0	2	1	16.34	17.48	18.03
24	1	1	2	0	4.94	4.96	4.27
25	1	0	1	0	2.12	3.15	3.06
26	1	1	2	2	29.00	28.08	29.16
27	2	0	1	0	4.11	3.88	3.80
28	2	1	2	1	18.59	18.85	17.57
29	2	0	0	1	3.99	4.28	4.95
30	2	1	0	3	7.08	7.14	7.01
31	3	0	2	2	30.12	29.19	30.48
32	1	1	2	0	5.43	6.70	6.33
33	0	0	0	1	0.04	-0.88	-0.60
34	2	0	1	1	9.88	10.67	10.64
35	0	0	0	2	0.70	0.43	1.33
36	3	1	2	1	21.28	20.84	21.40
37	0	1	2	3	37.99	37.75	36.29
38	1	0	1	1	8.12	8.27	8.40
39	3	1	2	1	20.42	19.07	19.52
40	0	1	1	0	2.00	2.84	1.52
41	1	1	0	1	5.09	4.96	5.08
42	1	1	1	0	5.06	5.21	4.55
43	0	0	1	3	17.38	17.37	16.63
44	1	1	2	1	17.09	17.07	16.75
45	2	0	1	1	9.34	9.67	9.21
46	3	1	1	1	14.83	16.18	14.64
47	3	0	0	0	6.20	6.35	5.51
48	2	0	0	3	3.93	4.97	4.19
49	3	0	1	3	23.92	23.61	24.36
50	1	0	1	2	13.55	13.83	13.07

DATA SET AID TEST DATA (3)

00001	33.	7.0	7.	45.	13.	2221134122221
00002	34.	7.0	7.	45.	13.	2221134122222
00003	67.	6.0	5.	48.	10.	2442322334221
00004	68.	6.0	5.	48.	10.	2432322334222
00005	99.	5.5	5.	45.	6.	3241314132211
00006	100.	5.5	5.	45.	6.	3241314132222
00007	26.	6.0	5.	48.	5.	1431321114121
00008	29.	6.0	5.	48.	5.	2221321112122
00009	51.	5.5	5.	45.	3.	3441322334221
00010	52.	5.5	5.	45.	3.	3441322334222
00011	103.	6.0	5.	45.	6.	3241334132221
00012	104.	6.0	5.	45.	6.	3241334132222
00013	71.	6.0	5.	48.	10.	3342121233321
00014	72.	6.0	5.	48.	10.	3222121232212
00015	3.	5.5	5.	56.	4.	2221324212311
00016	5.	5.5	5.	56.	4.	3341324213222
00017	55.	5.5	5.	45.	3.	3343324133221
00018	56.	5.5	5.	45.	3.	3443324132222
00019	107.	6.0	5.	45.	6.	2223122232211
00020	108.	6.0	5.	45.	6.	3243122232322
00021	37.	7.0	7.	45.	13.	3343122133321
00022	38.	7.0	7.	45.	13.	2343122133222
00023	79.	5.5	5.	48.	10.	2443324234221
00024	80.	5.5	5.	48.	10.	3443324234322
00025	112.	5.5	5.	45.	6.	3143312332211
00026	111.	5.5	5.	45.	6.	3333312333222
00027	32.	7.0	5.	48.	5.	3334121223211
00028	31.	7.0	5.	48.	5.	2224121222212
00029	59.	5.5	5.	48.	3.	2333111132211
00030	60.	5.5	5.	45.	3.	2433111134222
00031	115.	5.5	5.	45.	6.	2433323234221
00032	116.	5.5	5.	45.	6.	2333323234222
00033	83.	6.0	5.	48.	10.	2433322334221
00034	84.	6.0	5.	48.	10.	1223322332212
00035	79.	7.0	7.	45.	13.	3443323134321
00036	80.	7.0	7.	45.	13.	3243323132322
00037	87.	6.0	5.	48.	10.	3443122134321
00038	88.	6.0	5.	48.	10.	3343122134322
00039	119.	5.5	5.	45.	6.	3444314134321
00040	120.	5.5	5.	45.	6.	3334314134312
00041	22.	7.5	5.	48.	5.	2434122334221
00042	30.	7.5	5.	48.	5.	2434122334212
00043	42.	7.0	7.	45.	13.	2434322234221
00044	41.	7.0	7.	45.	13.	2434322234222
00045	91.	6.0	5.	48.	10.	1444324134321
00046	92.	6.0	5.	48.	10.	2434224134222
00047	123.	6.0	5.	45.	6.	1425121134211
00048	124.	6.0	5.	45.	6.	2435121134312
00049	1.	8.0	5.	48.	5.	2435332234311
00050	2.	8.0	5.	48.	5.	2435332234312

DATA SET (FISHER'S IRIS DATA WITH CLASS-CODE)

2	7.00	3.20	4.70	1.40
1	4.80	3.10	1.60	0.20
1	5.40	3.40	1.50	0.40
1	5.20	4.10	1.50	0.10
2	5.50	2.40	3.80	1.10
2	6.70	3.10	4.70	1.50
1	4.60	3.40	1.40	0.30
1	5.00	3.40	1.50	0.20
1	4.40	2.90	1.40	0.20
2	5.20	2.70	3.90	1.40
2	5.90	3.00	4.20	1.50
3	6.10	3.00	4.90	1.80
3	6.40	2.80	5.60	2.10
3	7.20	3.00	5.80	1.60
1	4.80	3.40	1.60	0.20
1	4.80	3.00	1.40	0.10
1	4.30	3.00	1.10	0.10
2	6.20	2.90	4.30	1.30
3	7.40	2.80	6.10	1.90
2	6.00	2.20	4.00	1.00
2	6.30	2.30	4.40	1.30
3	7.90	3.80	6.40	2.00
3	6.40	2.80	5.60	2.20
1	5.40	3.40	1.70	0.20
2	5.60	3.00	4.10	1.30
1	5.10	3.70	1.50	0.40
2	5.50	2.50	4.00	1.30
1	5.00	3.50	1.30	0.30
1	4.90	3.10	1.50	0.10
3	6.30	2.80	5.10	1.50
1	4.50	2.30	1.30	0.30
2	5.60	2.50	3.90	1.10
3	6.10	2.60	5.60	1.40
1	5.40	3.70	1.50	0.20
1	4.40	3.20	1.30	0.20
1	5.00	3.50	1.60	0.60
1	5.10	3.80	1.90	0.40
3	7.70	3.00	6.10	2.30
3	6.30	3.40	5.60	2.40
1	4.80	3.00	1.40	0.30
1	5.10	3.80	1.60	0.20
2	5.50	2.60	4.40	1.20
3	6.40	3.10	5.50	1.80
3	6.00	3.00	4.80	1.80
1	4.60	3.20	1.40	0.20
2	6.50	2.80	4.60	1.50
2	6.10	3.00	4.60	1.40
1	5.30	3.70	1.50	0.20
1	5.80	4.00	1.20	0.20
1	5.70	4.40	1.50	0.40
2	5.10	2.50	3.00	1.10
3	6.90	3.10	5.40	2.10
2	5.70	2.80	4.50	1.30
2	6.10	2.90	4.70	1.40
2	5.60	2.90	3.60	1.30
3	6.70	3.10	5.60	2.40
1	5.40	3.90	1.30	0.40
3	6.90	3.10	5.10	2.30
1	5.10	3.50	1.40	0.30
2	6.70	3.10	4.40	1.40
3	5.80	2.70	5.10	1.90
3	6.80	3.20	5.90	2.30
1	5.70	3.80	1.70	0.30
1	5.10	3.80	1.50	0.30
3	6.70	3.30	5.70	2.50
3	6.70	3.00	5.20	2.30
2	6.30	3.30	4.70	1.60
2	5.60	3.00	4.50	1.50
2	5.80	2.70	4.10	1.00
1	5.00	3.20	1.20	0.20
1	5.50	3.50	1.30	0.20
1	4.60	3.60	1.00	0.20
1	5.50	4.20	1.40	0.20
1	5.10	3.30	1.70	0.50
1	4.90	3.10	1.50	0.20
3	5.70	2.80	4.10	1.30
3	6.30	3.30	6.00	2.50
3	5.80	2.70	5.10	1.90
3	7.10	3.00	5.90	2.10
1	5.00	3.40	1.60	0.40
1	5.10	3.50	1.40	0.20
1	4.90	3.00	1.40	0.20
1	4.70	3.20	1.30	0.20
2	5.50	2.40	3.70	1.00
2	5.80	2.70	3.90	1.20
2	6.00	2.70	5.10	1.60
2	4.90	2.40	3.30	1.00
3	6.30	2.90	5.60	1.80
3	6.30	2.50	5.00	1.90
3	6.50	3.00	5.20	2.00
2	5.80	2.60	4.00	1.20
2	6.40	3.20	4.50	1.50
1	4.60	3.10	1.50	0.20
2	6.60	2.90	4.60	1.30
2	5.00	2.30	3.30	1.00
2	5.60	2.70	4.20	1.30
3	6.50	3.00	5.80	2.20
3	7.60	3.00	6.60	2.10
3	4.90	2.50	4.50	1.70
3	7.30	2.90	6.30	1.80
3	5.80	2.80	5.10	2.40
3	6.40	3.20	5.30	2.30
1	4.90	3.60	1.40	0.10
1	4.40	3.00	1.30	0.20
3	6.50	3.00	5.50	1.80
3	7.70	3.80	6.70	2.20
3	7.70	2.60	6.90	2.30
2	6.10	2.80	4.00	1.30
3	6.00	2.20	5.00	1.50
3	6.70	2.50	5.80	1.80
2	6.30	2.50	4.90	1.50
2	6.10	2.80	4.70	1.20
3	6.90	3.20	5.70	2.30
3	5.60	2.80	4.90	2.00
1	4.80	3.40	1.90	0.20
3	7.70	2.80	6.70	2.00
3	6.30	2.70	4.90	1.80
1	5.00	3.00	1.60	0.20
1	5.20	3.50	1.50	0.20
3	6.70	3.30	5.70	2.10
2	6.40	2.80	4.30	1.30
2	5.00	2.00	3.50	1.00
3	7.20	3.60	6.10	2.50
3	6.20	3.40	5.40	2.30
3	6.50	3.20	5.10	2.00
3	6.40	2.70	5.30	1.90
3	6.80	3.00	5.50	2.10
3	5.70	2.50	5.00	2.00
1	5.90	3.00	5.10	1.80
2	5.00	3.30	1.40	0.20
3	7.20	3.20	6.00	1.80
1	5.00	3.60	1.40	0.20
2	6.60	3.00	4.40	1.40
2	6.80	2.80	4.80	1.40
2	6.90	3.10	4.90	1.50
1	5.20	3.40	1.40	0.20
1	4.70	3.20	1.60	0.20
2	5.50	2.30	4.00	1.30
1	5.40	3.90	1.70	0.40
2	6.70	3.00	5.00	1.70
2	6.00	2.90	4.50	1.50
2	5.70	2.60	3.50	1.00
2	5.70	3.00	4.20	1.20
2	6.20	2.20	4.50	1.50
1	5.10	3.40	1.50	0.20
2	5.70	2.90	4.20	1.30
3	6.20	2.80	4.80	1.80
2	5.40	3.00	4.50	1.50
2	6.00	3.40	4.50	1.60
2	5.90	3.20	4.80	1.80

DATA SET (KENDALL'S DATA)

```

1 6 7 2 5 8 7 8 8 3 8 9 7 5 7 10
2 9 10 5 8 10 9 9 10 5 9 9 8 8 8 10
3 7 8 3 6 9 8 9 7 4 9 9 8 6 8 10
4 5 6 8 5 6 5 9 2 8 4 5 8 7 6 5
5 6 8 8 8 4 4 9 2 8 5 5 8 8 7 7
6 7 7 7 6 8 7 10 5 9 6 5 8 6 6 6
7 9 9 8 8 8 8 8 10 8 10 8 9 8 10
8 9 9 9 8 9 9 8 8 10 9 10 9 9 9 10
9 9 9 7 8 8 8 8 5 9 8 9 8 8 8 10
10 4 7 10 2 10 10 7 10 3 10 10 10 9 3 10
11 4 7 10 0 10 8 3 9 5 9 10 8 10 2 5
12 4 7 10 4 10 10 7 8 2 8 8 10 10 3 7
13 6 9 8 10 5 4 9 4 4 4 5 4 7 6 8
14 8 9 8 9 6 3 8 2 5 2 6 6 7 5 6
15 4 8 8 7 5 4 10 2 7 5 3 6 6 4 6
16 6 9 6 7 8 9 8 9 8 8 7 6 8 6 10
17 8 7 7 7 9 5 8 6 6 7 8 6 6 7 8
18 6 8 8 4 8 8 6 4 3 3 6 7 2 6 4
19 6 7 8 4 7 8 5 4 4 2 6 8 3 5 4
20 4 8 7 8 8 9 10 5 2 6 7 9 8 8 9
21 3 8 6 8 8 8 8 10 5 3 6 7 8 8 5 8
22 9 8 7 8 9 10 10 10 3 10 8 10 8 10 8
23 7 10 7 9 9 9 10 10 3 9 9 10 9 10 8
24 9 8 7 10 8 10 10 10 2 9 7 9 9 10 8
25 6 9 7 7 4 5 9 3 2 4 4 4 4 5 4
26 7 8 7 8 5 4 8 2 3 4 5 6 5 5 6
27 2 10 7 9 8 9 10 5 3 5 6 7 6 4 5
28 6 3 5 3 5 3 5 0 0 3 3 0 0 5 0
29 4 3 4 3 3 0 0 0 0 4 4 0 0 5 0
30 4 6 5 6 9 4 10 3 1 3 3 2 2 7 3
31 5 5 4 7 8 4 10 3 2 5 5 3 4 8 3
32 3 3 5 7 7 9 10 3 2 5 3 7 5 5 2
33 2 3 5 7 7 9 10 3 2 2 3 6 4 5 2
34 3 4 6 4 3 3 8 1 1 3 3 3 2 5 2
35 6 7 4 3 3 0 9 0 1 0 2 3 1 5 3
36 9 8 5 5 6 6 8 2 2 2 4 5 6 6 3
37 4 9 6 4 10 8 8 9 1 3 9 7 5 3 2
38 4 9 6 6 9 9 7 9 1 2 10 8 5 5 2
39 10 6 9 10 9 10 10 10 10 10 8 10 10 10 10
40 10 6 9 10 9 10 10 10 10 10 10 10 10 10 10
41 10 7 8 0 2 1 2 0 10 2 0 3 0 0 10
42 10 3 8 0 1 1 0 0 10 0 0 0 0 0 10
43 3 4 9 8 2 4 5 3 6 2 1 3 3 3 8
44 7 7 7 6 9 8 8 6 8 8 10 8 8 6 5
45 9 6 10 9 7 7 10 2 1 5 5 7 8 4 5
46 9 8 10 10 7 9 10 3 1 5 7 9 9 4 4
47 0 7 10 3 5 0 10 0 0 2 2 0 0 0 0
48 0 6 10 1 5 0 10 0 0 2 2 0 0 0 0

```

〔各課題とCHECK命令文の関係〕

データ 課題	F o r m						T y p e			
	raw	score	regress	distance	similarity	relation	category	binary	quantitative	mixed
HIERARCHICAL	<input type="checkbox"/>	<input type="checkbox"/>							<input type="checkbox"/>	
CLUSTER	<input type="checkbox"/>			<input type="checkbox"/>						
MAID			<input type="checkbox"/>							<input type="checkbox"/>
ASSOCIATION	<input type="checkbox"/>						<input type="checkbox"/>	<input type="checkbox"/>		
ISODATA	<input type="checkbox"/>	<input type="checkbox"/>							<input type="checkbox"/>	
KMEANS	<input type="checkbox"/>	<input type="checkbox"/>							<input type="checkbox"/>	
FUZZY	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			<input type="checkbox"/>	
PCA	<input type="checkbox"/>								<input type="checkbox"/>	

(*) ○印のところが正しい組み合わせと判定される。

Research Report

General Series No. 51

**A USER'S MANUAL
TO
MINTS - 80**

— MINI-NUMERICAL TAXONOMY SYSTEM —

by

NOBORU OHSUMI

THE INSTITUTE OF STATISTICAL MATHEMATICS

MAY 1980

TÔKEI-SÛRI KENKYÛZYO

Institute of Statistical Mathematics

4-6-7 Minami-Azabu, Minato-ku,

Tôkyô 106, Japan