CHAPTER 7

The Ditto Algorithm

7.1 INTRODUCTION

The data in Table 7.1 record the evaluation of various wines from 1961 to 1970. The evaluations are given in terms of five categories—disastrous, poor, average, good, excellent. In many statistical analyses, it is customary to quickly shuffle category variables out of sight by replacing them with interval scale variables—for example, 1, 2, 3, 4, 5 would be reasonable here because the categories are ordered. For clustering, category variables are more natural than interval variables and should be cherished and treated as they lie. A category variable partitions the set of cases. For example, the 1961 evaluations divide the wines into classes—excellent, good, and average. An overall partition of the data is thus a combination of the individual partitions. There is no more reason for a single partition to be adequate to explain the data than for the first principal component to be adequate for interval scale data.

In the ditto algorithm, the center for each cluster is equal to the mode of each variable over cases in the cluster. A measure of partition error is defined similar to that in the K-means algorithm.

In the K-means algorithm, it is necessary to fix the number of clusters in the partition; otherwise the partition of M clusters is best. In this algorithm, the error function does not necessarily decrease as the number of clusters increases, and so the partition size varies during the course of the algorithm.

7.2 DITTO ALGORITHM

Preliminaries. There are M cases, N variables, and a value $A(I, J)$ of the Ith case for the Jth variable. Let $P(M, K)$ denote a partition of the cases. The error of the partition $P(M, K)$ is

$$e[P(M, K)] = M + \sum \{1 \leq J \leq N\} \sum \{1 \leq L \leq K\}\, \rho[B(0, J), B(L, J)] + \sum \{1 \leq J \leq N\} \sum \{1 \leq I \leq M\}\, \rho[A(I, J), B(L(I), J)],$$

where $L(I)$ is the cluster containing I, $\rho(A, B) = 1$ if $A \neq B$, and $\rho(A, B) = 0$ if $A = B$. The vector $\{B(L, J), 1 \leq J \leq N\}$ for the Lth cluster is the center of the cluster. It will be chosen to minimize the expression $e[P(M, K)]$; this requirement does not uniquely determine $B(L, J)$, but it is always possible to find $B(L, J)$ minimizing e and equal to a mode of $A(I, J)$ over cases I in the cluster L. The vector $\{B(0, J), 1 \leq J \leq N\}$ is the "grand mode" of the cluster centers.

Table 7.1 Evaluation of Wines 1961–1970

D disastrous
P poor
A average
G good
E excellent

		1961	62	63	64	65	66	67	68	69	70
	Red Bordeaux										
MG	Medoc and Graves	E	G	P	G	D	G	G	P	A	G
EP	Saint Emilion and Pomerol	E	A	P	G	P	G	G	P	A	G
	White Bordeaux										
SS	Sauternes	G	G	D	D	D	A	G	P	G	G
GS	Graves	G	G	D	G	D	G	G	P	G	G
RB	Red Burgundy	E	G	A	G	P	G	A	D	G	G
	White Burgundy										
CB	Cote de Beaune	E	G	A	G	A	G	G	A	G	G
CS	Chablis	E	G	A	G	P	G	G	A	G	G
BS	Beaujolais	E	G	P	G	D	G	A	P	G	G
	Red Rhone										
RN	North	E	G	P	G	A	G	G	A	G	G
RS	South	G	A	P	G	A	A	G	A	G	G
WL	White Loire	A	P	P	G	P	A	G	G	G	G
AE	Alsace	G	A	P	E	P	G	G	P	G	G
RE	Rhine	G	A	P	G	P	G	G	P	G	G
ME	Moselle	G	A	P	E	P	G	G	P	G	G
CA	California	-	-	-	P	G	A	G	G	A	G

From *Gourmet Magazine* (August, 1971), pp. 30–33.

There are four stages in the ditto algorithm:

(i) selecting the initial partition;

(ii) moving cases from one cluster to another, or creating new clusters;

(iii) updating the modal values $B(L, J)$ within clusters, and updating the grand mode $B(0, J)$;

(iv) printing out the original data, recording the value $A(I, J)$ as a dot if it agrees with $B[L(I), J]$, and recording $B(L, J)$ as a dot if it agrees with $B(0, J)$. The number of nondot values is $e[P(M, K)]$.

STEP 1. Initially define a different cluster center $\{B(L, J), J = 1, \ldots, N\}$ for each value taken by a variable. The total number of clusters K will be the number of different values taken by each variable, summed over variables. For the Lth cluster, suppose a variable $J(L)$ takes a value $V(L)$. Then $B(L, J)$ is a mode of values $A(I, J)$ for which $A[I, J(L)] = V(L)$.

Finally, $B(0, J)$ is a mode of $B(L, J)$ over all clusters L $(1 \leq L \leq K)$.

STEP 2. For each case I $(1 \leq I \leq M)$ allocate I to the first cluster L $(0 \leq L \leq K)$ for which $\sum \{1 \leq J \leq N\}\, \rho[A(I, J), B(L, J)]$ is a minimum.

STEP 3. Delete clusters containing no objects. Delete clusters containing one object, and allocate the object to the 0 cluster.

STEP 4. For each cluster L, replace $B(L, J)$ $(1 \leq J \leq N)$ by a mode of values $A(I, J)$ $(I \in L)$ and the value $B(0, J)$. If $B(0, J)$ is a possible value of $B(L, J)$, set $B(L, J) = B(0, J)$. For the cluster 0, replace $B(0, J)$ by the mode of $B(L, J)$ over all clusters and over $A(I, J)$ for cases allocated to 0.

STEP 5. If any change occurs in Steps 2–4, return to Step 2. Otherwise, replace $A(I, J)$ by a dot, if $I \in L$, $A(I, J) = B(L, J)$ $(1 \leq I \leq M,\ 1 \leq J \leq N)$. Replace $B(L, J)$ by a dot if $B(L, J) = B(0, J)$ $(1 \leq L \leq K)$.

7.3 APPLICATION OF DITTO ALGORITHM TO WINES

STEP 1. There are 10 variables, taking 28 different values, so initially there are 28 clusters. The first cluster center is computed by using variable $J(1) = 1$ and value $V(1) = E$. The cases I such that $A(I, 1) = E$ are MG EP RB CB CS BS RN. The modal values $B(1, J)$ are

$B(1, 1) = E$, the mode of E E E E E E E
$B(1, 2) = G$, the mode of G A G G G G G
$B(1, 3) = P$, the mode of P P A A A P P

and so on. The next cluster center uses $J(2) = 1$ and $V(2) = G$, the second value taken by the first variable. The matching cases are SS GS RS AE RE ME. The complete set of initial cluster values appears in Table 7.2. Some clusters, such as 20, 26, and 28, are identical. The redundant ones will be eliminated in later steps.

STEP 2. Each case is now allocated to a cluster—the cluster whose center it best matches. For example, case 1, MG, differs from the center of cluster 4 only in the year 1969, so MG is allocated to cluster 4. The complete allocation is given in Table 7.3. There are many single clusters that will be allocated to the 0 cluster in the next step.

STEP 3. Delete the many clusters to which no cases are allocated. For example cluster 10, having an identical center to 0, will receive no cases and be eliminated. Delete clusters such as cluster 7 that contain a single object.

STEP 4. Recompute cluster centers. For example, cluster 4 contains MG and BS. To compute $B(4, 1)$, use the data values E, E and the value $B(0, 1) = E$. The mode of these three values is $B(4, 1) = E$. To consider a less straightforward case, consider $B(9, 4)$. The cluster 9 contains cases SS and GS taking values D and G in variable 4, 1964. The overall mode is $B(0, 4) = G$. Thus the mode of D, G, G is $B(9, 4) = G$.

Finally, the grand mode is recomputed by using cluster centers and cases allocated to the grand mode.

STEP 5. Steps 2–4 are repeated until there is no change in the clusters, which occurs after two allocations. The array is now prepared for display in a dot diagram. Since MG is allocated to cluster 4, $A(1, 1) = B(4, 1) = E$ and $A(1, 1)$ is replaced by a dot.

Table 7.2 Initial Cluster Centers Applying Ditto Algorithm to Wines

CLUSTER	VARIABLE	VALUE	CLUSTER CENTRE 1961	62	63	64	65	66	67	68	69	70
1	61	E	E	G	P	G	P	G	G	A	G	G
2	61	G	G	A	P	G	P	G	G	P	G	G
3	61	A	A	P	P	G	P	A	G	G	G	G
4	62	G	E	G	P	G	D	G	G	P	G	G
5	62	A	G	A	P	G	P	G	G	P	G	G
6	62	P	G	P	P	G	P	A	G	G	G	G
7	63	P	E	A	P	G	P	G	G	P	G	G
8	63	A	E	G	A	G	P	G	G	A	G	G
9	63	D	G	G	D	G	D	G	G	P	G	G
10	64	G	E	G	P	G	P	G	G	P	G	G
11	64	E	G	A	P	E	P	G	G	P	G	G
12	64	D	G	G	D	D	D	A	G	P	G	G
13	64	P	-	-	-	P	G	A	G	G	A	G
14	65	P	E	A	P	G	P	G	G	P	G	G
15	65	A	G	A	P	G	A	A	G	A	G	G
16	65	D	E	G	P	G	D	G	G	P	G	G
17	65	G	-	-	-	P	G	A	G	G	A	G
18	66	G	E	G	P	G	P	G	G	P	G	G
19	66	A	G	G	P	G	P	A	G	G	G	G
20	67	G	G	G	P	G	P	G	G	P	G	G
21	67	A	E	G	P	G	P	G	G	P	G	G
22	68	P	E	G	P	G	P	G	G	P	G	G
23	68	A	E	G	A	G	P	G	G	A	G	G
24	68	G	A	P	P	G	P	A	G	G	G	G
25	68	D	E	G	A	G	P	G	A	D	G	G
26	69	G	G	G	P	G	P	G	G	P	G	G
27	69	A	E	G	P	G	P	G	G	P	A	G
28	70	G	G	G	P	G	P	G	G	P	G	G
0			E	G	P	G	P	G	G	P	G	G

Similarly, since $B(4, 1) = B(0, 1) = E$, $B(4, 1)$ is replaced by a dot. The final dotted array appears in Table 7.4.

The total number of symbols necessary to represent the data is 41, which should be compared to the 150 data values and to the 28 different values taken by all variables. The story told by the clustering is as follows. The usual grading of wine is "good." Overall there were three poor years, 1963, 1965, and 1968, and one excellent year, 1961. There is a gallimaufry of four wines, St. Emilion and Pomerol, Red Rhone North, White Loire, and California, which vary (differently) from this overall

Table 7.3 Successive Passes of Ditto Algorithm on Wines

		CLUSTER MODE									
CLUSTER	ALLOCATION 1	1961	62	63	64	65	66	67	68	69	70
4	MG BS	E	G	P	G	D	G	G	P	G	G
7	EP	E	G	P	G	P	G	G	P	G	G
9	SS GS	G	G	D	G	D	G	G	P	G	G
25	RB	E	G	P	G	P	G	G	P	G	G
8	CB CS	E	G	A	G	P	G	G	A	G	G
1	RN	E	G	P	G	P	G	G	P	G	G
15	RS	E	G	P	G	P	G	G	P	G	G
24	WL	E	G	P	G	P	G	G	P	G	G
2	AE RE ME	G	A	P	G	P	G	G	P	G	G
13	CA	E	G	P	G	P	G	G	P	G	G
0		E	G	P	G	P	G	G	P	G	G
CLUSTER	ALLOCATION 2										
4	MG BS	E	G	P	G	D	G	G	P	G	G
9	SS GS	G	G	D	G	D	G	G	P	G	G
8	RB CB CS	E	G	A	G	P	G	G	A	G	G
2	RS AE RE ME	G	A	P	G	P	G	G	P	G	G
0	EP RN WL CA	E	G	P	G	P	G	G	P	G	G
CLUSTER	ALLOCATION 3	No change.									
4	MG BS										
9	SS GS										
8	RB CB CS										
2	RS AE RE ME										
0	EP RN WL CA										

pattern. California particularly is quite different. There are four clusters of wines: the Beaujolais cluster (disastrous in 1965), the Sauternes cluster (disastrous in 1963 and 1965), the Chablis cluster (average in 1963 and 1968 when others were poor), the Moselle cluster (not as good as others in 1961 and 1962).

7.4 THINGS TO DO

7.4.1 Running the Ditto Algorithm

This algorithm is especially appropriate for category data. Continuous variables may be converted to this form by division into classes such as low, middle, and high. Variables with large numbers of categories will have little effect on the final classification but will increase the expense of computation, so such variables should be avoided. An unusual feature of the algorithm is the presence of a gallimaufry of objects, each forming a singleton cluster.

Table 7.4 Ditto Diagram of Wines

	1961	62	63	64	65	66	67	68	69	70
CLUSTER 0	E	G	P	G	P	G	G	P	G	G
St. Emilion and Pomerol	•	A	•	•	•	•	•	•	A	•
Red Rhone North	•	•	•	•	A	•	•	A	•	•
White Loire	A	P	•	•	•	A	•	•	•	•
California	-	-	-	P	G	A	•	G	•	•
CLUSTER 4	•	•	•	•	D	•	•	•	•	•
Medoc and Graves	•	•	•	•	•	•	•	•	A	•
Beaujolais	•	•	•	•	•	•	A	•	•	•
CLUSTER 9	G	•	D	•	D	•	•	•	•	•
Sauternes	•	•	•	D	•	A	•	•	•	•
Graves	•	•	•	•	•	•	•	•	•	•
CLUSTER 8	•	•	A	•	•	•	•	A	•	•
Red Burgundy	•	•	•	•	•	•	A	D	•	•
Cote de Beaune	•	•	•	•	A	•	•	•	•	•
Chablis	•	•	•	•	•	•	•	•	•	•
CLUSTER 2	G	A	•	•	•	•	•	•	•	•
Red Rhone South	•	•	•	•	A	A	•	A	•	•
Alsace	•	•	•		•	•	•	•	•	•
Rhine	•	•	•	•	•	•	•	•	•	•
Moselle	•	•	•	E	•	•	•	•	•	•

Number of symbols = 41

Number without partitions = 150

Fill in cluster centers from cluster 0. Fill in data values from cluster centers.

A good data set for this algorithm is the sleeping pattern of seventeen monkeys in a vervet troop observed by Struhsaker (Table 7.5). Also see the metamorphosis sequences of British butterflies (Table 7.6).

PROGRAMS

SCALE converts continuous variables to category variables.

DITTO computes partition of category data to maximize matching between cases in a cluster and the cluster mode.

DITOUT using output from ditto, this program prints out dot matrix, where each dot represents an identity between a value in the matrix and the corresponding cluster mode.

Table 7.5 Vervet Sleeping Groups

I	- Adult male	IX	- Young juvenile female
II	- Older adult male	X	- Juvenile female
III	- Adult male	XI	- Subadult female
IV	- Adult female	XII	- Adult female
V	- Juvenile male	XIII	- Two young indist. juv. males
VI	- Adult female	XIV	- Infant male, son of IV
VII	- Young juvenile female	XV	- Infant female, from XII
VIII	- Young juvenile female	XVI	- Infant male, from VI

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII
4Jan64	A	D	A	B	A	E	E	E	C	C	C	A	B	A	B		
30Jan64	B	B	B	B	B	A	A	A	A	A	B	B	B	B	B	B	
5Feb64	B	A	C	A	A	D	D	D	D	D	A	A	A	A	A	A	D
10Feb64	A	B	B	D	B	D	D	D	C	C	A	D	A	A	D	D	D
24Feb64	B	B	B	B	B	B	B	B	B	B	A	A	B	B	B	A	B
25Feb64	A	B	A	B	B	B	B	B	B	B	A	B	B	B	B	B	B
10Mar64	C	C	B	C	C	C	C	C	C	A	B	B	B	C	C	B	C
20Mar64	B	B	A	B	B	B	B	B	B	B	A	A	B	B	B	A	B
1Apr64	D	C	A	D	D	D	D	D	B	A	B	C	A	D	D	C	D
2Apr64	D	A	D	C	D	D	D	D	B	D	B	A	B	C	C	A	D
5Apr64	A	A	C	E	D	B	E	E	E	C	E	E	E	E	E	E	B
7Apr64	C	C	B	A	C	B	C	B	A	B	A	B	A	A	A	B	B
18Apr64	C	D	B	A	C	C	C	C	A	A	A	B	A	A	A	B	C
28Apr64	A		A	D	A	D	D	D	A	A	C	B	A	D	D	B	D
1May64	B		D	A	C	C	C	C	A	A	B	B	C	A	A	B	C
5May64	D		C	B	F	A	E	E	E	E	D	C	E	B	B	C	A
10May64	B		B	E	B	D	D	D	A	F	A	C	A	B	E	C	D
12May64	D		B	D	D	E	E	E	B	B	C	A	B	D	D	A	E
20May64	A		B	D	C	C	C	C	A	C	A	C	D	C	D	C	C
22May64	E		B	F	F	A	A	A	C	F	C	D	F	F	F	D	A
2Jun64	C		D	C	C	E	E	E	C	C	B	A	C	E	C	A	E
5Jun64	A		D	C	A	D	D	A	D	A	B	A	C	A	C	A	D

From Struhsaker, T. T. (1967). "Behaviour of vervet monkeys and other cercopithecines." *Science* **156,** 1197–1203.

Table 7.6 Times of Appearance of British Butterflies

O (ova); L (larva); P (pupa); I (imago)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Adonis Blue	L	L	L	L	PI	LI	P	PI	LI	L	L	L
Bath White	P	P	P	P	LI	L	P	PI	LP	P	P	P
Black Hairstreak	O	O	O	O	L	LPI	I	O	O	O	O	O
Black-veined White	L	L	L	L	LP	PI	LI	L	L	L	L	L
Brinstone	I	I	I	I	LI	LI	PI	PI	I	I	I	I
Brown Argus	L	L	L	L	L	P	PI	I	L	L	L	L
Brown Hairstreak	O	O	O	O	L	LP	PI	I	I	O	O	O
Camberwell Beauty	I	I	I	I	I	L	LP	I	I	I	I	I
Chalk Hill Blue	L	L	L	L	LPI	LP	PI	I	L	L	L	L
Chequered Skipper	L	L	L	P	P	I	O	O	L	L	L	L
Clouded Yellow	-	-	-	-	I	LI	LP	I	LI	LPI	I	-
Comina	I	I	I	I	LI	LP	PI	LI	PI	I	I	I
Common Blue	L	L	L	LP	I	IL	LI	I	I	L	L	L
Dark Green Fritillary	L	L	L	L	L	LP	PI	LI	L	L	L	L
Dingy Skipper	L	L	L	LP	I	LI	LP	I	L	L	L	L
Duke of Burgundy Fritillary	P	P	P	P	PI	IOL	L	L	LP	P	P	P
Essex Skipper	O	O	O	OL	L	P	I	I	O	O	O	O
Gatekeeper	L	L	L	L	L	LP	I	I	L	L	L	L
Glanville Fritillary	L	L	L	LP	LPI	LPI	L	L	L	L	L	L
Grayling	L	L	L	L	L	LPI	PI	LI	LI	L	L	L
Green Hairstreak	P	P	P	P	PI	LI	LP	P	P	P	P	P
Green-veined White	P	P	P	PI	LI	LPI	LPI	LPI	LP	P	P	P
Grizzled Skipper	L	L	L	LP	I	LI	LP	I	L	L	L	L
Heath Fritillary	L	L	L	L	L	LPI	LI	L	L	L	L	L
High Brown Fritillary	L	L	L	L	L	LP	PI	LI	L	L	L	L
Holly Blue	P	P	P	PI	I	LP	PI	LI	LPI	LP	P	P
Large Blue	L	L	L	L	L	LI	I	L	L	L	L	L

[Ford, T. L. E. (1963). *Practical Entomology*, Warne, London, p. 181.] (A subset has been selected from the full list.)

```
      SUBROUTINE SCALE(A,M,N,KL,KK)
C...............................................................20 MAY 1973
C.... SCALES ARRAY TO TAKE INTEGER OR ALPHAMERIC VALUES 1,2,3,... KL
C     MINIMUM AND MAXIMUM VALUES ARE COMPUTED FOR EACH VARIABLE, AND EACH VALUE
C     IS THEN CLASSIFIED INTO ONE OF KL INTERVALS OF EQUAL LENGTH BETWEEN THE
C     MINIMUM AND MAXIMUM.
C.... M = NUMBER OF ROWS
C.... N = NUMBER OF COLUMNS
C.... A = M BY N BORDERED ARRAY
C.... KL = NUMBER OF LEVELS
C.... KK = LEVELLING OPTION
C            KK = 1 UNIFORM OVER ALL DATA VALUES
C            KK = 2 UNIFORM WITHIN VARIABLES
C            KK = 3 SAME AS OPTION 1, CONVERTED TO ALPHAMERIC
C            KK = 4 SAME AS OPTION 2, CONVERTED TO ALPHAMERIC
C.........................................................................
      DIMENSION A(M,N)
      DIMENSION CC(9)
      DATA CC/1H1,1H2,1H3,1H4,1H5,1H6,1H7,1H8,1H9/
           IF(KK.GE.3.AND.KL.GT.9) WRITE(6,1) KL
    1 FORMAT(I5,27H TOO MANY ALPHAMERIC LEVELS
      IF(KK.EQ.1.OR.KK.EQ.3) GO TO 20
C.... COMPUTE MINIMUM AND MAXIMUM
      DO 10 J=2,N
      XMIN=A(2,J)
      XMAX=A(2,J)
      DO 11 I=2,M
      IF(A(I,J).GT.XMAX) XMAX=A(I,J)
      IF(A(I,J).LT.XMIN) XMIN=A(I,J)
   11 CONTINUE
C.... CHANGE A VALUES TO INTEGER OR ALPHAMERIC
      IF(XMIN.EQ.XMAX) XMAX=XMIN+.000001
      ZZ=KL/(XMAX-XMIN)
      DO 12 I=2,M
      K=(A(I,J)-XMIN)*ZZ+1
      IF(KK.EQ.4) A(I,J)=CC(K)
      IF(KK.EQ.2) A(I,J)=K
   12 CONTINUE
   10 CONTINUE
       RETURN
   20 CONTINUE
C.... MINIMUM AND MAXIMUM
      XMIN=A(2,2)
      XMAX=A(2,2)
      DO 21 I=2,M
      DO 21 J=2,N
      IF(A(I,J).LT.XMIN) XMIN=A(I,J)
   21 IF(A(I,J).GT.XMAX) XMAX=A(I,J)
C.... CHANGE A VALUES
      IF(XMIN.EQ.XMAX) XMAX=XMIN+.000001
      ZZ=KL/(XMAX-XMIN)
      DO 22 I=2,M
      DO 22 J=2,N
      K=(A(I,J)-XMIN)*ZZ+1.
      IF(KK.EQ.3) A(I,J)=CC(K)
      IF(KK.EQ.1) A(I,J)=K
   22 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE DITTO(M,N,KL,KC,A,X,LC,LK,Y,Z)
C.............................................................20 MAY 1973
C.... COMPUTES PARTITION OF CATEGORY DATA TO MAXIMIZE MATCHING BETWEEN CASES IN
C            A CLUSTER AND THE CLUSTER MODE.
C.... M = NUMBER OF ROWS
C.... N = NUMBER OF COLUMNS
C.... KL = MAXIMUM NUMBER OF DIFFERENT VALUES TAKEN BY A VARIABLE PLUS ONE.
C.... KC = KL BY N +1
C.... A = M BY N BORDERED DATA ARRAY
C.... X = KC BY N ARRAY OF CLUSTER MODES
C.... LC = 1 BY M ARRAY ASSIGNING CASES TO CLUSTERS
C.... LK = 1 BY KC ARRAY COUNTING CASES IN CLUSTERS
C.... Y = KL BY N ARRAY COUNTING FREQUENCIES IN CLUSTERS
C.... Z = KL BY N ARRAY SPECIFYING DIFFERENT VALUES OF VARIABLES
C...............................................................................
      DIMENSION A(M,N),X(KC,N),LC(M),LK(KC),Y(KL,N),Z(KL,N)
      DIMENSION CC(10)
      DATA CC/1H1,1H2,1H3,1H4,1H5,1H6,1H7,1H8,1H9,2H10/
      DATA RNGE,XMOD/4HRNGE,4HMODE/
C.... PUT LABELS IN VARIOUS ARRAYS
      DO 80 K=2,KL
      KK=(K-1)-((K-2)/10)*10
   80 Z(K,1)=CC(KK)
      DO 81 J=2,N
   81 Z(1,J)=A(1,J)
      Z(1,1)=RNGE
      DO 82 K=2,KC
      KK=(K-1)-((K-2)/10)*10
   82 X(K,1)=CC(KK)
      DO 83 J=2,N
   83 X(1,J)=A(1,J)
      X(1,1)=XMOD
C.... FIND DIFFERENT VALUES TAKEN BY VARIABLES
      DO 9 I=2,M
    9 LC(I)=0
      DO 8 K=1,KC
    8 LK(K)=1
      NC=-1
      DO 10 J=2,N
      DO 10 K=2,KL
   10 Z(K,J)=0
      DO 11 J=2,N
      DO 12 I=2,M
      DO 13 K=2,KL
      IF(Z(K,J).EQ.0) Z(K,J)=A(I,J)
      IF(Z(K,J).EQ.A(I,J)) GO TO 12
   13 CONTINUE
   12 CONTINUE
   11 CONTINUE
C.... COMPUTE MODES
   70 NC=NC+1
      DO 20 J=2,N
      DO 20 K=2,KL
      DO 21 JJ=2,N
      DO 21 KK=2,KL
   21 Y(KK,JJ)=0
      KT=K+1+(J-2)*(KL-1)
      IF(LK(KT).EQ.0) GO TO 20
      DO 23 I=2,M
      IF(NC.EQ.0.AND.A(I,J).NE.Z(K,J)) GO TO 23
      IF(NC.NE.0.AND.LC(I).NE.KT) GO TO 23
      DO 25 JJ=2,N
      DO 25 KK=2,KL
   25 IF(A(I,JJ).EQ.Z(KK,JJ)) Y(KK,JJ)=Y(KK,JJ)+1
   23 CONTINUE
      DO 26 JJ=2,N
      YM=Y(2,JJ)
      KM=2
      DO 27 KK=2,KL
      IF(Y(KK,JJ).LE.YM) GO TO 27
      YM=Y(KK,JJ)
      KM=KK
   27 CONTINUE
   26 X(KT,JJ)=Z(KM,JJ)
   20 CONTINUE
C.... COMPUTE GRAND MODE
      DO 55 J=2,N
      DO 50 K=2,KL
```

```
   50 Y(K,J)=0
      DO 51 K=3,KT
      IF(LK(K).EQ.0) GO TO 51
      DO 52 KK=2,KL
   52 IF(X(K,J).EQ.Z(KK,J)) Y(KK,J)=Y(KK,J)+1
   51 CONTINUE
      DO 53 KK=2,KL
      YM=Y(2,J)
      X(2,J)=Z(2,J)
      IF(Y(KK,J).LE.YM) GO TO 53
      YM=Y(KK,J)
      X(2,J)=Z(KK,J)
   53 CONTINUE
   55 CONTINUE
C.... REASSIGN CASES
      DO 30 I=2,M
      DM=10.**10
      KM=1
      DO 32 K=2,KT
      IF(LK(K).EQ.0) GO TO 32
      IF(NC.NE.0.AND.LK(K).LE.1) GO TO 32
      DD=0
      DO 31 J=2,N
   31 IF(A(I,J).NE.X(K,J)) DD=DD+1
      IF(DD.GE.DM) GO TO 32
      KM=K
      DM=DD
   32 CONTINUE
   30 LC(I)=KM
C.... COUNT CASES IN CLUSTERS
      DO 35 K=3,KC
   35 LK(K)=0
      DO 36 I=2,M
      K=LC(I)
   36 LK(K)=LK(K)+1
      DO 37 K=3,KC
      IF(LK(K).NE.0) GO TO 37
      DO 38 J=2,N
   38 X(K,J)=0.
   37 CONTINUE
      IF(NC.LT.5) GO TO 70
      RETURN
      END
```

```
      SUBROUTINE DITOUT(A,M,N,X,KC,LC)
C...............................................................20 MAY 1973
C.... PRINTS OUT DATA MATRIX WITH CLUSTER VALUES DOTTED OUT
C.... DATA MATRIX A AND CLUSTER MATRIX X ARE ALPHABETIC
C.... USE AFTER PROGRAM DITTO
C.... M = NUMBER OF ROWS
C.... N = NUMBER OF COLUMNS
C.... KC = NUMBER OF CLUSTERS
C.... A = M BY N BORDERED ARRAY,DATA VALUES
C.... X = KC BY N BORDERED ARRAY, CLUSTER VALUES
C.... SECOND ROW OF X CONTAINS GRAND CLUSTER
C.... LC = 1 BY M ARRAY ASSIGNING CASES TO CLUSTERS
C..........................................................................
      DIMENSION A(M,N),X(KC,N),LC(M)
      DIMENSION AA(120)
      DATA DOT/1H./
      WRITE(6,1) A(1,1)
    1 FORMAT(14H1DOT MATRIX OF,A5)
      WRITE(6,5)
    5 FORMAT(
     *63H FIRST CLUSTER IS GRAND MODE.  ALL OTHER CLUSTER VALUES ARE
     *63HREPLACED BY  '.' IF THEY AGREE WITH VALUE OF GRAND MODE
     *63H   WITHIN A CLUSTER VALUES ARE REPLACED BY '.' IF THEY AGREE
     *63H WITH CLUSTER VALUE.
     *)
      DO 20 K=2,6
   20 WRITE(6,2)(A(1,J),J=K,N,5)
    2 FORMAT(10X,24A5)
C.... DATA
      DO 30 K=2,KC
      NC=0
      DO 40 I=2,M
      IF(LC(I).NE.K) GO TO 40
      NC=NC+1
      DO 60 J=2,N
   60 IF(K.GE.3.AND.X(K,J).EQ.X(2,J)) X(K,J)=DOT
      IF(NC.EQ.1) WRITE(6,3) K,(X(K,J),J=2,N)
    3 FORMAT(5H0CLUS,I3,3X,120A1)
      DO 50 J=2,N
      AA(J)=A(I,J)
      IF(X(K,J).EQ.DOT) X(K,J)=X(2,J)
   50 IF(A(I,J).EQ.X(K,J)) AA(J)=DOT
      WRITE(6,4) A(I,1),(AA(J),J=2,N)
    4 FORMAT(3X,A5,3X,120A1)
   40 CONTINUE
   30 CONTINUE
      RETURN
      END
```