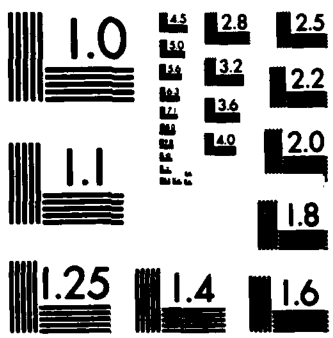
AD-A184 687 ANNOTATED COMPUTER OUTPUT FOR ILLUSTRATIVE EXAMPLES OF 1/1
CLUSTERING USING T (U) CORNELL UNIV ITHACA NY
MATHEMATICAL SCIENCES INST K E BASFORD ET AL
UNCLASSIFIED 26 JUN 87 ARO-23306 91-MA DAAG29-85-C-0018 F/G 12/3 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

DTIC FILE COPY

ARO 23306.91-MA

(2)

CORNELL UNIVERSITY

MATHEMATICAL SCIENCES INSTITUTE

ANNOTATED COMPUTER OUTPUT

'87 - 1

ANNOTATED COMPUTER OUTPUT FOR ILLUSTRATIVE EXAMPLES OF CLUSTERING USING THE MIXTURE METHOD AND TWO COMPARABLE METHODS FROM SAS

K.E. BASFORD, N.J. MILES-MCDERMOTT, AND W.T. FEDERER

February 1987

294 Caldwell Hall ■ Ithaca, New York 14853-2602

(607) 255-8005

UNCLASSIFIED MASTER COPY - FOR REPRODUCTION PURPOSES

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARO 23306.91-MA
6a. NAME OF PERFORMING ORGANIZATION Mathematical Sciences Inst. 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION U. S. Army Research Office
6c. ADDRESS (City, State, and ZIP Code) 294 Caldwell Hall, Cornell University Ithaca, New York 14853	7b. ADDRESS (City, State, and ZIP Code) P. O. Box 12211 Research Triangle Park, NC 27709-2211
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U. S. Army Research Office 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER DAAG29-85-C-0018
8c. ADDRESS (City, State, and ZIP Code) P. O. Box 12211 Research Triangle Park, NC 27709-2211	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification) ANNOTATED COMPUTER OUTPUT FOR ILLUSTRATIVE EXAMPLES OF CLUSTERING USING THE MIXTURE METHOD AND TWO COMPARABLE METHODS FROM SAS

12. PERSONAL AUTHOR(S) K.E. Basford, N.J. Miles-McDermott, and W.T. Federer

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Interim Technical	FROM ______ TO ______	June 26, 1987	45

16. SUPPLEMENTARY NOTATION The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	mixture model, control language, clustering, Habbema, variab analysis, observations, PLENGTH, PWIDTH, SWIDTH, SLENGTH,

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This is the annotated computer output for the three clustering methods considered in the associated technical report ~~BU-921-M~~ and '87-5, Illustrative Examples of Clustering using the Mixture Method and Two Comparable Methods from SAS, by ~~K.E.~~ Basford, ~~W.T.~~ Federer and ~~N.J.~~ Miles-McDermott. The computer output for the normal mixture model method is generated from a fortran program, KMM, written by ~~K.E.~~ Basford. Two other clustering methods are considered and are from SAS/CLUSTER, Version 5. These are Ward's method and the EML method. Two real data sets are processed.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code) / 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

ANNOTATED COMPUTER OUTPUT FOR ILLUSTRATIVE EXAMPLES OF CLUSTERING USING THE MIXTURE METHOD AND TWO COMPARABLE METHODS FROM SAS

K.E. BASFORD, N.J. MILES-MCDERMOTT, AND W.T. FEDERER

February 1987

* Partially supported by Mathematical Sciences Institute and by the Australian-American Education Foundation.

ABSTRACT

This is the annotated computer output for the three clustering methods considered in the associated technical report, BU-921-M and '87-5, Illustrative Examples of Clustering using the Mixture Method and Two Comparable Methods from SAS, by K.E. Basford, W.T. Federer, and N.J. Miles-McDermott. The computer output for the normal mixture model method is generated from a fortran program, KMM, written by K.E. Basford. Two other clustering methods are considered and are from SAS/CLUSTER, Version 5. These are Ward's method and the EML method. Two real data sets are processed.

COMMENTS

The annotated output should be read in sequence because explanations made on earlier pages are not necessarily repeated subsequently. Some pages may be composites of more than one output page, and some output pages are omitted because they are generally not useful for the purpose at hand. A general description of the mixture model approach to clustering is explained in detail and discussed in relation to other clustering methods in Basford (1986). SAS program documentation is in SAS User's Guide (1985a and b). Program documentation for KMM is available from K.E.

Basford and will appear in a forthcoming book by McLachlan and Basford (1987).

The data are presented below. Following on pages 9-10 the KMM and SAS control language for each example is presented. Control language is given in capital type with accompanying descriptions and notes given in boldface type. Program output follows on pages 12-44 with annotations in boldface and lower case type that describe output values in some detail.

DATA SETS

Two data sets are used for each of the three clustering methods presented. The first data set was taken from Habbema, Hermans, and van den Broek (1974). These examples are labeled CL-1-Habbema through CL-3-Habbema on the output pages. The second data set is the well known Iris data published by Fisher (1936). These examples are labeled CL-1-Fisher through CL-3-Fisher on the output pages. For each data set, the first example CL-1- illustrates the normal mixture model method of clustering using the KMM program. CL-2- illustrates Ward's method using SAS and CL-3- illustrates the EML method also using SAS.

The data taken from Habbema et al., consists of 45 observations on known haemophilia A carriers and 35 observations on known noncarriers. These data are shown in Table 1 and contain three variables. GROUP indicates whether the individual is a carrier (coded 2) or noncarrier (coded 1). The two other variables are used to discriminate between the normal individuals and the carriers in the clustering programs and are $\log_{10}$(AHF activity) and $\log_{10}$(AHF-like antigen). These variables were named ACTIVITY and ANTIGEN, respectively.

TABLE 1: Habbema et al., Haemophilia Data

GROUP	ACTIVITY	ANTIGEN
1	-0.00559	-0.16571
1	-0.16980	-0.15852
1	-0.34689	-0.18791
1	-0.08944	0.00642
1	-0.16791	0.07129
1	-0.08362	0.01059
1	-0.19789	-0.00054
1	-0.07621	0.03919
1	-0.19129	-0.21229
1	-0.10919	-0.11904
1	-0.52677	-0.47734
1	-0.08419	0.02482
1	-0.02252	-0.05805
1	0.00841	0.07821
1	-0.18266	-0.11384
1	0.12366	0.21397
1	-0.47022	-0.30989
1	-0.15191	-0.06864
1	0.00061	-0.11531
1	-0.20154	-0.04976
1	-0.19318	-0.22933
1	0.15069	0.09331
1	-0.12591	-0.06686
1	-0.15508	-0.12321
1	-0.19515	-0.10067
1	0.02908	0.04419

1	-0.22282	-0.17099
1	-0.09971	-0.07333
1	-0.19724	-0.06074
1	-0.08670	-0.05597
2	-0.49859	-0.08602
2	-0.50145	-0.29844
2	-0.13259	0.00970
2	-0.34787	-0.17209
2	-0.37553	-0.18652
2	-0.24466	-0.04067
2	-0.22047	0.00455
2	-0.21539	-0.02191
2	-0.25404	-0.05729
2	-0.37780	-0.26816
2	-0.06391	0.15694
2	-0.33510	-0.13676
2	-0.01493	0.15392
2	-0.03124	0.14001
2	-0.17402	-0.07764
2	-0.09636	0.05307
2	-0.02344	0.08038
2	-0.40546	-0.24184
2	-0.34776	0.11506
2	-0.36180	-0.20082
2	-0.69112	-0.33899
2	-0.36083	0.12372
2	-0.45348	-0.16817
2	-0.35388	0.07219
2	-0.47186	-0.10786
2	-0.36097	-0.03994
2	-0.32261	0.16697
2	-0.43193	-0.06869
2	-0.27342	-0.00203
2	-0.55728	0.05480
2	-0.49503	-0.01529
2	-0.51066	-0.24825
2	-0.16516	0.21321
2	-0.42318	-0.09981
2	-0.23746	0.28763
2	-0.34470	0.00969
2	-0.40465	-0.11618
2	-0.14158	0.16416
2	-0.15082	0.11372
2	-0.26421	0.08669
2	-0.33525	0.08753
2	-0.18782	0.25096
2	-0.17443	0.18924
2	-0.24443	0.16137
2	-0.47837	0.02821

The Fisher Iris data is shown in Table 2 and consists of four measurements on 50 plants from each of three species of Iris: *Iris setosa*, *Iris versicolor*, and *Iris virginica*, These species were coded 1, 2, and 3, respectively, with a variable name of GROUP. The four measurement variables input into the clustering programs were sepal length (SLENGTH), sepal width (SWIDTH), petal length (PLENGTH), and petal width (PWIDTH).

TABLE 2: Fisher Iris Data

GROUP	SLENGTH	SWIDTH	PLENGTH	PWIDTH
1	5.1	3.5	1.4	0.3
1	4.4	3.2	1.3	0.2
1	4.4	3.0	1.3	0.2
1	5.0	3.5	1.6	0.6
1	5.1	3.8	1.6	0.2
1	4.9	3.1	1.5	0.2
1	5.0	3.2	1.2	0.2
1	4.6	3.2	1.4	0.2
1	5.0	3.3	1.4	0.2
1	4.8	3.4	1.9	0.2
1	4.8	3.0	1.4	0.1
1	5.0	3.5	1.3	0.3
1	5.1	3.3	1.7	0.5
1	5.0	3.4	1.5	0.2
1	5.1	3.8	1.9	0.4
1	4.9	3.0	1.4	0.2
1	5.3	3.7	1.5	0.2
1	4.3	3.0	1.1	0.1
1	5.5	3.5	1.3	0.2
1	4.8	3.4	1.6	0.2
1	5.2	3.4	1.4	0.2
1	4.8	3.1	1.6	0.2
1	4.9	3.6	1.4	0.1
1	4.6	3.1	1.5	0.2
1	5.7	4.4	1.5	0.4
1	5.7	3.8	1.7	0.3
1	4.8	3.0	1.4	0.3
1	5.2	4.1	1.5	0.1
1	4.7	3.2	1.6	0.2
1	4.5	2.3	1.3	0.3
1	5.4	3.4	1.7	0.2
1	5.0	3.0	1.6	0.2
1	4.6	3.4	1.4	0.3
1	5.4	3.9	1.3	0.4

1	5.0	3.6	1.4	0.2
1	5.4	3.9	1.7	0.4
1	4.6	3.6	1.0	0.2
1	5.1	3.8	1.5	0.3
1	5.8	4.0	1.2	0.2
1	5.4	3.7	1.5	0.2
1	5.0	3.4	1.6	0.4
1	5.4	3.4	1.5	0.4
1	5.1	3.7	1.5	0.4
1	4.4	2.9	1.4	0.2
1	5.5	4.2	1.4	0.2
1	5.1	3.4	1.5	0.2
1	4.7	3.2	1.3	0.2
1	4.9	3.1	1.5	0.1
1	5.2	3.5	1.5	0.2
1	5.1	3.5	1.4	0.2
2	6.4	3.2	4.5	1.5
2	5.5	2.4	3.8	1.1
2	5.7	2.9	4.2	1.3
2	5.7	3.0	4.2	1.2
2	5.6	2.9	3.6	1.3
2	7.0	3.2	4.7	1.4
2	6.8	2.8	4.8	1.4
2	6.1	2.8	4.7	1.2
2	4.9	2.4	3.3	1.0
2	5.8	2.7	3.9	1.2
2	5.8	2.6	4.0	1.2
2	5.5	2.4	3.7	1.0
2	6.7	3.0	5.0	1.7
2	5.7	2.8	4.1	1.3
2	6.7	3.1	4.4	1.4
2	5.5	2.3	4.0	1.3
2	5.1	2.5	3.0	1.1
2	6.6	2.9	4.6	1.3
2	5.0	2.3	3.3	1.0
2	6.9	3.1	4.9	1.5
2	5.0	2.0	3.5	1.0
2	5.6	3.0	4.5	1.5
2	5.6	3.0	4.1	1.3
2	5.8	2.7	4.1	1.0
2	6.3	2.3	4.4	1.3
2	6.1	3.0	4.6	1.4
2	5.9	3.0	4.2	1.5
2	6.0	2.7	5.1	1.6
2	5.6	2.5	3.9	1.1
2	6.7	3.1	4.7	1.5
2	6.2	2.2	4.5	1.5
2	5.9	3.2	4.8	1.8
2	6.3	2.5	4.9	1.5
2	6.0	2.9	4.5	1.5
2	5.6	2.7	4.2	1.3
2	6.2	2.9	4.3	1.3

2	6.0	3.4	4.5	1.6
2	6.5	2.8	4.6	1.5
2	5.7	2.8	4.5	1.3
2	6.1	2.9	4.7	1.4
2	5.5	2.5	4.0	1.3
2	5.5	2.6	4.4	1.2
2	5.4	3.0	4.5	1.5
2	6.3	3.3	4.7	1.6
2	5.2	2.7	3.9	1.4
2	6.4	2.9	4.3	1.3
2	6.6	3.0	4.4	1.4
2	5.7	2.6	3.5	1.0
2	6.1	2.8	4.0	1.3
2	6.0	2.2	4.0	1.0
3	6.3	3.3	6.0	2.5
3	6.7	3.3	5.7	2.1
3	7.2	3.6	6.1	2.5
3	7.7	3.8	6.7	2.2
3	7.2	3.0	5.8	1.6
3	7.4	2.8	6.1	1.9
3	7.6	3.0	6.6	2.1
3	7.7	2.8	6.7	2.0
3	6.2	3.4	5.4	2.3
3	7.7	3.0	6.1	2.3
3	6.8	3.0	5.5	2.1
3	6.4	2.7	5.3	1.9
3	5.7	2.5	5.0	2.0
3	6.9	3.1	5.1	2.3
3	5.9	3.0	5.1	1.8
3	6.3	3.4	5.6	2.4
3	5.8	2.7	5.1	1.9
3	6.3	2.7	4.9	1.8
3	6.0	3.0	4.8	1.8
3	7.2	3.2	6.0	1.8
3	6.2	2.8	4.8	1.8
3	6.9	3.1	5.4	2.1
3	6.7	3.1	5.6	2.4
3	6.4	3.1	5.5	1.8
3	5.8	2.7	5.1	1.9
3	6.1	3.0	4.9	1.8
3	6.0	2.2	5.0	1.5
3	6.4	3.2	5.3	2.3
3	5.8	2.8	5.1	2.4
3	6.9	3.2	5.7	2.3
3	6.7	3.0	5.2	2.3
3	7.7	2.6	6.9	2.3
3	6.3	2.8	5.1	1.5
3	6.5	3.0	5.2	2.0
3	7.9	3.8	6.4	2.0
3	6.1	2.6	5.6	1.4
3	6.4	2.8	5.6	2.1
3	6.3	2.5	5.0	1.9

3	4.9	2.5	4.5	1.7
3	6.8	3.2	5.9	2.3
3	7.1	3.0	5.9	2.1
3	6.7	3.3	5.7	2.5
3	6.3	2.9	5.6	1.8
3	6.5	3.0	5.5	1.8
3	6.5	3.0	5.8	2.2
3	7.3	2.9	6.3	1.8
3	6.7	2.5	5.8	1.8
3	5.6	2.8	4.9	2.0
3	6.4	2.8	5.6	2.2
3	6.5	3.2	5.1	2.0

Control Language

CL-1-Habbema (Mixture method from KMM)

```
75 2                        ⇒ 75 is the number of observations and 2 is
                               the number of variables
-0.005595 -0.165712
-0.169805 -0.158521         }  INPUT DATA:  ACTIVITY and ANTIGEN
    :
    :
-0.478366  0.028215
2                  ⇒ number of clusters to be formed
2                  ⇒ instructs KMM to assume unequal covariance matrices
1                  ⇒ signals KMM that initial grouping estimates follow
1 1 2 1 1 1 1 1 1 1
2 1 1 1 1 1 1 2 1 1
1 1 1 1 1 1 1 1 1 1
2 2 1 2 2 1 1 1 1 2         }  Initial groupings of observations
1 2 1 1 1 1 1 2 1 2            (results of Ward's method were used)
2 1 1 1 2 1 1 2 1 2
2 1 2 2 1 1 2 1 1 1
1 2 1 1 2
```

CL-2-Habbema (Ward's method from SAS)

```
DATA GJ;
INPUT ACTIVITY ANTIGEN;        ⇒ Input variables
IF _N_ LE 30 THEN GROUP=1;     }  Defines the GROUP variable
ELSE GROUP=2;
CARDS;                         ⇒ Signals SAS that the data follow
-0.005595 -0.165712
-0.169805 -0.158521
    :
    :
-0.478366  0.028215
PROC CLUSTER OUTTREE=TREE METHOD=WARD;  ⇒ Requests CLUSTER analysis
                          using Ward's method on ACTIVITY and ANTIGEN
VAR ACTIVITY ANTIGEN;
COPY GROUP;
PROC TREE SORT HEIGHT=N;  ⇒ Requests the Cluster Tree from 1 to n
                             (75) clusters
ID GROUP;
PROC TREE NCL=2 OUT=OUT NOPRINT;
ID GROUP;                            }  Causes SAS to produce 2×2
PROC FREQ;                              table showing misclassifications
TABLE CLUSTER*GROUP;
```

CL-3-Habbema (EML method from SAS)

Same control language as for 2) above except substitute EML for WARD on PROC CLUSTER line.

CL-1-Fisher (Mixture method from KMM)

```
150 4                 ⇒ 150 is number of observations and 4 is the
                        number of variables
5.1 3.5 1.4 0.3
4.4 3.2 1.3 0.2                   Input data
:
6.5 3.2 5.1 2.0
3                    ⇒ Number of clusters to be formed
1                    ⇒ Instructs KMM to assume equal covariance matrices
1                    ⇒ Signals KMM that initial grouping estimates follow
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1               Initial grouping of observations
:                                 (results of Ward's method were used)
3 3 3 3 3 3 3 2 3 3
```

CL-2-Fisher (Ward's method from SAS)

```
DATA ONE;
INPUT SLENGTH SWIDTH PLENGTH PWIDTH; ⇒ Input variables
IF _N_ LE 50 THEN GROUP=1;
ELSE IF _N_ LE 100 THEN GROUP=2;       Defines the GROUP variable
ELSE GROUP=3;
CARDS;                              ⇒ Signals SAS that the data follow
5.1 3.5 1.4 0.3
4.4 3.2 1.3 0.2
:
6.5 3.2 5.1 2.0
PROC CLUSTER OUTTREE=TREE METHOD=WARD;  Requests the Cluster anlysis
VAR SLENGTH SWIDTH PLENGTH PWIDTH;      using Ward's method on the 4
COPY GROUP;                             variables SLENGTH, SWIDTH,
                                        PLENGTH, AND PWIDTH
PROC TREE DATA=TREE  SORT HEIGHT=N;      Requests cluster
ID GROUP;                                      tree
PROC TREE DATA=TREE NCL=3 OUT=OUT NOPRINT;
ID GROUP;
COPY SLENGTH SWIDTH PLENGTH PWIDTH;      Requests the 2×2 table
PROC FREQ;                               showing misclassifications
TABLE CLUSTER*GROUP;
PROC CANDISC NOPRINT OUT=CAN;            This series of commands is
CLASS CLUSTER;                           used to display cluster
VAR SLENGTH SWIDTH PLENGTH PWIDTH;       results.  The CANDISC pro-
PROC PLOT;                               cedure is run to produce
PLOT CAN2*CAN1=CLUSTER;                  canonical variables for the
PROC PLOT;                               cluster groups.  The first
PLOT CAN2*CAN1=GROUP;                    2 canonical variables are
                                         then plotted to show cluster
                                         membership
```

CL-3-Fisher (EML method from SAS)

Same control language as for 2) above except substitute EML for WARD on PROC CLUSTER line.

CL-1-Habbema

Initial partition as specified by input

1	1	2	1	1	1	1	1	1	1
2	1	1	1	1	1	1	2	1	1
1	1	1	1	1	1	1	1	1	1
2	2	1	2	2	1	1	1	1	2
1	2	1	1	1	1	1	2	1	2
2	1	1	1	2	1	1	2	1	2
2	1	2	2	1	1	2	1	1	1
1	2	1	1	2					

Initial group allocation for each observation. The entry for row 1 column 1 refers to observation 1, row 1 column 2 refers to observation 2, and so on

Estimated mean (as a row vector) for each group

ACTIVITY	ANTIGEN	
-0.221538	-0.032402	= GROUP 1
-0.282643	-0.040757	= GROUP 2

Group means for each variable based on initial group allocation above

Estimated covariance matrix for group 1 = $S_{ij(\text{group 1})}$

$0.031661 = S_1^2$

$0.010517 = S_{12}$ $0.019972 = S_2^2$

Estimated covariance matrix for group 2 = $S_{ij(\text{group 2})}$

$0.022859 = S_1^2$

$0.016834 = S_{12}$ $0.030533 = S_2^2$

Covariance matrices for each group based on initial group allocation

Proportion from each group as specified by input

0.720 0.280

= Number initially assigned to group i/total number of observations

In loop 55 log likelihood is 77.035

77.035 is the solution of the likelihood equation based on 55 iterations of the EM algorithm

Estimate of mixing proportion for each group

0.508 0.492

Estimate of the final proportion for each group under the normal mixture model

Entity: Final estimates of posterior probabilities of group membership

OBSERVATION	GROUP=1	GROUP=2
1	0.999	0.001
2	0.971	0.029
3	0.245	0.755
4	0.985	0.015
5	0.710	0.290
6	0.986	0.014
7	0.773	0.227
8	0.983	0.017
9	0.958	0.042
10	0.992	0.008
11	0.001	0.999
12	0.983	0.017
13	0.999	0.001
14	0.997	0.003
15	0.948	0.052

These estimates indicate the degree of certainty with which each observation belongs to one of the two groups

For example, observation 1 has a probability of .999 of belonging to group 1 and .001 of belonging to group 2

CL-1-Habbema

16	0.998	0.002
17	0.014	0.986
18	0.966	0.034
19	0.999	0.001
20	0.855	0.145
21	0.957	0.043
22	1.000	0.000
23	0.982	0.018
24	0.976	0.024
25	0.922	0.078
26	0.999	0.001
27	0.899	0.101
28	0.992	0.008
29	0.882	0.118
30	0.993	0.007
31	0.003	0.997
32	0.185	0.815
33	0.001	0.999
34	0.006	0.994
35	0.949	0.051
36	0.000	1.000
37	0.002	0.998
38	0.012	0.988
39	0.223	0.777
40	0.008	0.992
41	0.004	0.996
42	0.045	0.955
43	0.002	0.998
44	0.008	0.992
45	0.274	0.726
46	0.000	1.000

47	0.125	0.875
48	0.000	1.000
49	0.004	0.996
50	0.092	0.908
51	0.606	0.394
52	0.015	0.985
53	0.001	0.999
54	0.620	0.380
55	0.736	0.264
56	0.034	0.966
57	0.591	0.409
58	0.152	0.848
59	0.032	0.968
60	0.899	0.101
61	0.241	0.759
62	0.975	0.025
63	0.970	0.030
64	0.944	0.056
65	0.426	0.574
66	0.636	0.364
67	0.963	0.037
68	0.089	0.911
69	0.992	0.008
70	0.010	0.990
71	0.016	0.984
72	0.126	0.874
73	0.073	0.927
74	0.031	0.969
75	0.000	1.000

Resulting partition of the entities into NG groups

1	1	2	1	1	1	1	1	1	1
2	1	1	1	1	1	2	1	1	1
1	1	1	1	1	1	1	1	1	1
2	2	2	2	1	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2
1	2	2	1	1	2	1	2	2	1
2	1	1	1	2	1	1	2	1	2
2	2	2	2	2					

Final group allocations after 55 iterations of clustering algorithm

Number assigned to each group
39 36

Estimates of correct allocation rates for each group
0.934 0.908

Overall estimate of degree of certainty with which observations are allocated to each group

Estimate of overall correct allocation rate 0.921 Weighted average of estimates of correct allocation rates for each group

Estimated mean (as a row vector) for each group

ACTIVITY	ANTIGEN		
-0.115406	-0.024497	= GROUP 1	Group means for each variable based on final estimates of posterior probability of group membership
-0.365950	-0.045323	= GROUP 2	

Estimated covariance matrix for group 1 = $S_{ij(\text{group 1})}$

0.011245		Based on final estimates of posterior probability of group membership
0.006548	0.012367	

Estimated covariance matrix for group 2 = $S_{ij(\text{group 2})}$

0.015898	
0.015029	0.032278

CL-1-FISHER

Initial partition as specified by input

1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2
2	2	3	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3
3	3	2	3	2	3	2	2	2	3
2	3	3	3	2	2	2	3	2	3
3	3	2	3	3	2	3	2	2	3
3	3	3	3	3	3	3	2	3	3

Initial group allocation for each observation. The entry for row 1 column 1 refers to observation 1, row 1 column 2 refers to observation 2, and so on

Estimated mean (as a row vector) for each group

SLENGTH	SWIDTH	PLENGTH	PWIDTH		
5.005994	3.427995	1.461996	0.246000	GROUP 1	Group means for each variable based on initial group allocation above
5.920269	2.751557	4.420300	1.434370	GROUP 2	
6.869439	3.086106	5.769438	2.105549	GROUP 3	

Estimated covariance matrix for group 1 = $S_{ij(\text{group 1})}$

0.124213			
0.099176	0.143674		
0.016347	0.011713	0.030165	
0.010327	0.009296	0.006070	0.011106

Estimated covariance matrix for group 2 = $S_{ij(\text{group 2})}$

0.227175			
0.066786	0.087267		
0.141501	0.053037	0.277231	
0.034401	0.028532	0.117393	0.085792

Estimated covariance matrix for group 3 = $S_{ij(\text{group 3})}$

0.241609			
0.016371	0.082387		
0.185024	0.011265	0.230741	
-0.008398	0.027246	0.009312	0.059419

Covariance matrices for each group based on initial group allocation

Estimated common covariance matrix

0.196290			
0.065579	0.104907		
0.110146	0.029316	0.183807	
0.016186	0.021814	0.054552	0.054618

In this run we specified that KMM assume equal covariance matrices for each group

This is the pooled estimate of that matrix based on the weighted average of the individual estimated covariance matrices

CL-1-FISHER

Proportion from each group as specified by input
0.333 0.427 0.240

.333 = 50/150 = Number initially assigned to group 1/total number of observations

In loop 30 log likelihood is -256.354 = **Solution to the likelihood equation based on 30 iterations of the EM algorithm**

Estimate of mixing proportion for each group
0.333 0.330 0.337

Estimate of the final proportion for each group under the normal mixture model

Entity: Final estimates of posterior probabilities of group membership

OBSERVATION	GROUP 1	GROUP 2	GROUP 3
1	1.000	0.000	0.000
2	1.000	0.000	0.000
3	1.000	0.000	0.000
4	1.000	0.000	0.000
5	1.000	0.000	0.000
6	1.000	0.000	0.000
7	1.000	0.000	0.000
8	1.000	0.000	0.000
9	1.000	0.000	0.000
10	1.000	0.000	0.000
11	1.000	0.000	0.000
12	1.000	0.000	0.000
13	1.000	0.000	0.000
14	1.000	0.000	0.000
15	1.000	0.000	0.000
16	1.000	0.000	0.000
17	1.000	0.000	0.000
18	1.000	0.000	0.000
19	1.000	0.000	0.000
20	1.000	0.000	0.000
21	1.000	0.000	0.000
22	1.000	0.000	0.000
23	1.000	0.000	0.000
24	1.000	0.000	0.000
25	1.000	0.000	0.000
26	1.000	0.000	0.000
27	1.000	0.000	0.000
28	1.000	0.000	0.000
29	1.000	0.000	0.000
30	1.000	0.000	0.000
31	1.000	0.000	0.000
32	1.000	0.000	0.000
33	1.000	0.000	0.000
34	1.000	0.000	0.000
35	1.000	0.000	0.000
36	1.000	0.000	0.000
37	1.000	0.000	0.000
38	1.000	0.000	0.000
39	1.000	0.000	0.000
40	1.000	0.000	0.000
41	1.000	0.000	0.000
42	1.000	0.000	0.000
43	1.000	0.000	0.000
44	1.000	0.000	0.000
45	1.000	0.000	0.000
46	1.000	0.000	0.000
47	1.000	0.000	0.000
48	1.000	0.000	0.000

These estimates indicate the degree of certainty with which each observation belongs to one of the three groups. Observation 1 has a probability of 1.0 of belonging to group 1 and 0 of belonging to the other two groups

49	1.000	0.000	0.000
50	1.000	0.000	0.000
51	0.000	0.999	0.001
52	0.000	1.000	0.000
53	0.000	1.000	0.000
54	0.000	1.000	0.000
55	0.000	1.000	0.000
56	0.000	1.000	0.000
57	0.000	0.999	0.001
58	0.000	1.000	0.000
59	0.000	1.000	0.000
60	0.000	1.000	0.000
61	0.000	1.000	0.000
62	0.000	1.000	0.000
63	0.000	0.704	0.296
64	0.000	1.000	0.000
65	0.000	1.000	0.000
66	0.000	1.000	0.000
67	0.000	1.000	0.000
68	0.000	1.000	0.000
69	0.000	1.000	0.000
70	0.000	0.997	0.003
71	0.000	1.000	0.000
72	0.000	0.967	0.033
73	0.000	1.000	0.000
74	0.000	1.000	0.000
75	0.000	1.000	0.000
76	0.000	0.998	0.002
77	0.000	0.999	0.001
78	0.000	0.127	0.873
79	0.000	1.000	0.000
80	0.000	0.999	0.001
81	0.000	0.979	0.021
82	0.000	0.133	0.867
83	0.000	0.868	0.132
84	0.000	0.991	0.009
85	0.000	1.000	0.000
86	0.000	1.000	0.000
87	0.000	0.988	0.012
88	0.000	0.997	0.003
89	0.000	0.998	0.002
90	0.000	0.994	0.006
91	0.000	1.000	0.000
92	0.000	0.999	0.001
93	0.000	0.929	0.071
94	0.000	0.979	0.021
95	0.000	0.999	0.001
96	0.000	1.000	0.000
97	0.000	1.000	0.000
98	0.000	1.000	0.000
99	0.000	1.000	0.000

100	0.000	1.000	0.000
101	0.000	0.000	1.000
102	0.000	0.000	1.000
103	0.000	0.000	1.000
104	0.000	0.000	1.000
105	0.000	0.148	0.852
106	0.000	0.000	1.000
107	0.000	0.000	1.000
108	0.000	0.000	1.000
109	0.000	0.000	1.000
110	0.000	0.000	1.000
111	0.000	0.000	1.000
112	0.000	0.002	0.998
113	0.000	0.000	1.000
114	0.000	0.000	1.000
115	0.000	0.009	0.991
116	0.000	0.000	1.000
117	0.000	0.001	0.999
118	0.000	0.094	0.906
119	0.000	0.123	0.877
120	0.000	0.003	0.997
121	0.000	0.162	0.838
122	0.000	0.001	0.999
123	0.000	0.000	1.000
124	0.000	0.004	0.996
125	0.000	0.001	0.999
126	0.000	0.089	0.911
127	0.000	0.302	0.698
128	0.000	0.000	1.000
129	0.000	0.000	1.000
130	0.000	0.000	1.000
131	0.000	0.000	1.000
132	0.000	0.000	1.000
133	0.000	0.746	0.254
134	0.000	0.002	0.998
135	0.000	0.000	1.000
136	0.000	0.073	0.927
137	0.000	0.000	1.000
138	0.000	0.006	0.994
139	0.000	0.022	0.978
140	0.000	0.000	1.000
141	0.000	0.000	1.000
142	0.000	0.000	1.000
143	0.000	0.001	0.999
144	0.000	0.005	0.995
145	0.000	0.000	1.000
146	0.000	0.000	1.000
147	0.000	0.000	1.000
148	0.000	0.000	1.000
149	0.000	0.000	1.000
150	0.000	0.008	0.992

CL-1-FISHER

Resulting partition of the entities into NG groups

1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	3	2	2
2	3	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3
3	3	3	3	3	3	3	3	3	3
3	3	3	3	3	3	3	3	3	3
3	3	2	3	3	3	3	3	3	3
3	3	3	3	3	3	3	3	3	3

Final group allocations after 30 iterations

Number assigned to each group
50 49 51

Estimates of correct allocation rates for each group
1.000 0.973 0.983

Overall estimate of degree of certainty with which observations are allocated to each group

Estimate of overall correct allocation rate 0.985 = Weighted average of estimates of correct allocation rates for each group

Estimated mean (as a row vector) for each group

SLENGTH	SWIDTH	PLENGTH	PWIDTH
5.005994	3.427995	1.461996	0.246000
5.942309	2.760773	4.258801	1.319220
6.574652	2.980818	5.539058	2.024963

Group means for each variable based on estimates of posterior probability of group membership

Estimated common covariance matrix

0.263932			
0.089847	0.111946		
0.169658	0.051118	0.186544	
0.039336	0.029976	0.041973	0.039709

This pooled estimate of the common covariance matrix is based on the final estimates of posterior probability of group membership

WARD'S MINIMUM VARIANCE CLUSTER ANALYSIS

EIGENVALUES OF THE COVARIANCE MATRIX

	EIGENVALUE λ_i	DIFFERENCE $(\lambda_i - \lambda_{i+1})$	PROPORTION $\lambda_i/\Sigma\lambda_i$	CUMULATIVE
1	0.038779	0.025371	0.743072	0.74307
2	0.013409	.	0.256928	1.00000

ROOT-MEAN-SQUARE TOTAL-SAMPLE STANDARD DEVIATION = 0.161536
ROOT-MEAN-SQUARE DISTANCE BETWEEN OBSERVATIONS = 0.323072 → squared Eucludean distances

NUMBER = number of clusters formed at each step
FREQUENCY = number of observations in the newly formed cluster

Clusters joined identifies the two clusters being joined at a particular step. Clusters of one observation being joined is identified as OBn, where n is the observation number. Clusters of more than one observation are identified as CLn, where n is the number of clusters existing after the cluster is formed

NUMBER OF CLUSTERS	CLUSTERS JOINED		FREQUENCY OF NEW CLUSTER	SEMIPARTIAL R-SQUARED	R-SQUARED
74	OB4	OB6	2	0.000007	0.999993
73	OB20	OB29	2	0.000018	0.999975
72	OB49	OB52	2	0.000032	0.999944
71	OB3	OB34	2	0.000033	0.999911
70	OB8	OB12	2	0.000035	0.999876
69	OB9	OB21	2	0.000038	0.999838
68	OB15	OB25	2	0.000043	0.999795
67	OB36	OB39	2	0.000047	0.999748
66	OB35	OB50	2	0.000051	0.999697
65	OB43	OB44	2	0.000059	0.999638
64	OB28	OB30	2	0.000061	0.999577
63	OB7	OB37	2	0.000069	0.999507
62	OB18	OB45	2	0.000074	0.999434
61	OB54	OB71	2	0.000075	0.999358
60	OB64	OB67	2	0.000079	0.999279
59	OB63	OB73	2	0.000085	0.999194
58	CL63	OB38	3	0.000105	0.999088

Semipartial R^2 is the decrease in the proportion of variance accounted for resulting from joining two clusters.

.000007 is the decrease in the proportion of variances accounted for resulting from joining observations 4 and 6 into one cluster.

R^2 is the squared multiple correlation and is the proportion of variances accounted for by the clusters at a particular step.

i.e. .997537 = R^2 for 49 clusters

57	CL70	OB46	3	0.000122	0.998967
56	OB14	OB47	2	0.000132	0.998835
55	OB17	OB32	2	0.000143	0.998691
54	OB31	OB55	2	0.000154	0.998537
53	CL71	CL66	4	0.000166	0.998371
52	OB23	CL64	3	0.000185	0.998186
51	OB40	OB48	2	0.000189	0.997997
50	OB2	OB24	2	0.000190	0.997808
49	OB5	OB69	2	0.000271	0.997537
48	OB61	OB75	2	0.000281	0.997256
47	CL74	CL57	5	0.000290	0.996966
46	OB41	CL65	3	0.000305	0.996661
45	OB58	CL60	3	0.000323	0.996338
44	OB1	OB19	2	0.000334	0.996005
43	OB56	OB66	2	0.000353	0.995652
42	CL59	OB68	3	0.000375	0.995277
41	CL62	CL73	4	0.000426	0.994851
40	CL72	CL61	4	0.000429	0.994421
39	CL56	OB26	3	0.000444	0.993977
38	CL50	CL68	4	0.000474	0.993504
37	CL67	OB59	3	0.000481	0.993023
36	OB65	OB72	2	0.000493	0.992530
35	CL47	OB33	6	0.000532	0.991998
34	OB10	CL52	4	0.000564	0.991434
33	CL69	OB27	3	0.000590	0.990844
32	CL53	OB42	5	0.000628	0.990215
31	CL55	OB62	3	0.000646	0.989569
30	OB70	OB74	2	0.000773	0.988797
29	CL54	OB53	3	0.001050	0.987747
28	CL40	OB57	5	0.001088	0.986658
27	CL58	CL37	6	0.001117	0.985541
26	CL44	OB13	3	0.001243	0.984298
25	OB60	CL48	3	0.001263	0.983035
24	CL29	CL45	6	0.001421	0.981613
23	CL38	CL41	8	0.001872	0.979741
22	OB16	OB22	2	0.001980	0.977762
21	CL49	CL30	4	0.002592	0.975170

20	CL32	CL51	7	0.002801	0.972369
19	CL42	CL36	5	0.002858	0.969511
18	CL39	CL46	6	0.003321	0.966190
17	CL23	CL34	12	0.003847	0.962343
16	CL27	CL43	8	0.005464	0.956879
15	OB11	OB51	2	0.005975	0.950904
14	CL21	CL19	9	0.007694	0.943210
13	CL15	CL31	5	0.008771	0.934439
12	CL17	CL33	15	0.009699	0.924741
11	CL18	CL22	8	0.009889	0.914851
10	CL24	CL25	9	0.010854	0.903998
9	CL26	CL35	9	0.013422	0.890575
8	CL16	CL28	13	0.018916	0.871660
7	CL9	CL12	24	0.025938	0.845722
6	CL20	CL10	16	0.029542	0.816180
5	CL14	CL8	22	0.040895	0.775285
4	CL6	CL13	21	0.057587	0.717698
3	CL7	CL11	32	0.095677	0.622021
2	CL3	CL5	54	0.136697	0.485324
1	CL2	CL4	75	0.485324	0.000000

SAS

TABLE OF CLUSTER BY GROUP

CLUSTER / GROUP FREQUENCY PERCENT ROW PCT COL PCT	1	2	TOTAL	
1	27 36.00 50.00 90.00	27 36.00 50.00 60.00	54 72.00	= row total = row percent = 54/75 = .72
2	3 4.00 14.29 10.00	18 24.00 85.71 40.00	21 28.00	
TOTAL	30 40.00	45 60.00	75 100.00	

Cell frequency = 27
Cell percent = 27/75 = .36
Row percent = 27/54 = .50
column percent = 27/30 = .90

column total = 30
column percent = 30/75 = .4

This 2x2 table shows misclassifications. GROUP contains the true group allocation of each observation and CLUSTER contains the group allocation based on the Ward's clustering. 27 observations were correctly assigned to cluster 1 and 18 to cluster 2. 27 observations from group 2 were incorrectly assigned to cluster 1 while 3 from group 1 were incorrectly assigned to cluster 2

WARD'S MINIMUM VARIANCE CLUSTER ANALYSIS

EIGENVALUES OF THE COVARIANCE MATRIX

	EIGENVALUE λ_i	DIFFERENCE $(\lambda_i - \lambda_{i+1})$	PROPORTION $\lambda_i/\Sigma\lambda_i$	CUMULATIVE
1	4.22824	3.98557	0.924619	0.92462
2	0.24267	0.16446	0.053066	0.97769
3	0.07821	0.05437	0.017103	0.99479
4	0.02384	.	0.005212	1.00000

ROOT-MEAN-SQUARE TOTAL-SAMPLE STANDARD DEVIATION = 1.06922
ROOT-MEAN-SQUARE DISTANCE BETWEEN OBSERVATIONS = 3.02422

NUMBER OF CLUSTERS	CLUSTERS JOINED		FREQUENCY OF NEW CLUSTER	SEMIPARTIAL R-SQUARED	R-SQUARED
149	OB117	OB125	2	0.000000	1.000000
148	OB14	OB46	2	0.000007	0.999993
147	OB6	OB48	2	0.000007	0.999985
146	OB1	OB50	2	0.000007	0.999978
145	OB137	OB149	2	0.000007	0.999971
144	OB17	OB40	2	0.000007	0.999963
143	OB52	OB62	2	0.000015	0.999949
142	OB5	OB38	2	0.000015	0.999934
141	OB124	OB144	2	0.000015	0.999919
140	OB119	OB126	2	0.000015	0.999905
139	OB65	OB97	2	0.000015	0.999890
138	OB53	OB54	2	0.000015	0.999875
137	OB22	OB29	2	0.000015	0.999861
136	OB23	OB35	2	0.000015	0.999846
135	OB11	OB16	2	0.000015	0.999831
134	OB76	OB90	2	0.000015	0.999817
133	OB59	OB69	2	0.000015	0.999802

.000007 is the decrease in the proportion of variance explained resulting from joining observations 137 and 149. The associated R^2 is .999971

132	OB60	OB61	2	0.000015	0.999787
131	OB21	OB49	2	0.000015	0.999773
130	OB8	OB47	2	0.000015	0.999758
129	OB3	OB44	2	0.000015	0.999743
128	OB118	OB121	2	0.000022	0.999721
127	CL146	OB12	3	0.000022	0.999699
126	OB111	OB122	2	0.000022	0.999677
125	OB64	OB85	2	0.000022	0.999655
124	OB9	CL148	3	0.000022	0.999633
123	CL135	OB27	3	0.000024	0.999609
122	CL138	OB73	3	0.000024	0.999584
121	OB86	OB96	2	0.000029	0.999555
120	OB72	OB93	2	0.000029	0.999525
119	OB13	OB41	2	0.000029	0.999496
118	OB66	OB91	2	0.000029	0.999467
117	CL147	OB32	3	0.000032	0.999435
116	CL142	OB43	3	0.000034	0.999401
115	CL137	OB24	3	0.000034	0.999366
114	OB134	OB150	2	0.000037	0.999330
113	OB130	OB140	2	0.000037	0.999293
112	CL134	OB84	3	0.000044	0.999249
111	OB114	OB131	2	0.000044	0.999205
110	OB109	OB116	2	0.000044	0.999161
109	CL143	OB79	3	0.000044	0.999117
108	OB123	OB142	2	0.000044	0.999073
107	OB68	OB88	2	0.000044	0.999029
106	OB4	CL119	3	0.000049	0.998980
105	OB51	OB94	2	0.000051	0.998929
104	OB106	OB146	2	0.000051	0.998877
103	OB75	OB81	2	0.000051	0.998826
102	OB56	OB70	2	0.000051	0.998775
101	OB107	OB108	2	0.000051	0.998723
100	CL130	OB33	3	0.000054	0.998669
99	CL141	OB143	3	0.000054	0.998616
98	CL127	CL131	5	0.000056	0.998560
97	OB31	OB42	2	0.000059	0.998501
96	CL132	OB74	3	0.000064	0.998437
95	OB82	CL140	3	0.000064	0.998374

94	OB2	CL129	3	0.000064	0.998310
93	OB10	OB20	2	0.000066	0.998244
92	OB113	CL149	3	0.000068	0.998176
91	CL98	CL124	8	0.000073	0.998103
90	OB89	OB92	2	0.000073	0.998030
89	CL117	CL123	6	0.000073	0.997956
88	CL102	OB80	3	0.000076	0.997880
87	OB78	OB133	2	0.000081	0.997800
86	OB102	CL113	3	0.000081	0.997719
85	OB26	OB36	2	0.000081	0.997638
84	CL94	OB18	4	0.000083	0.997555
83	CL122	CL125	5	0.000086	0.997470
82	OB105	OB120	2	0.000088	0.997382
81	OB28	OB45	2	0.000088	0.997293
80	CL145	OB145	3	0.000091	0.997203
79	OB58	CL112	4	0.000092	0.997111
78	CL95	OB115	4	0.000094	0.997017
77	CL128	OB138	3	0.000095	0.996922
76	CL92	OB148	4	0.000097	0.996825
75	CL144	OB19	3	0.000100	0.996725
74	OB57	CL107	3	0.000103	0.996622
73	CL110	OB128	3	0.000103	0.996519
72	OB77	OB99	2	0.000117	0.996402
71	CL91	CL136	10	0.000121	0.996281
70	CL87	OB83	3	0.000125	0.996156
69	OB104	OB135	2	0.000125	0.996031
68	CL86	CL108	5	0.000126	0.995905
67	CL100	CL115	6	0.000139	0.995766
66	CL133	OB67	3	0.000142	0.995624
65	CL75	OB34	4	0.000142	0.995482
64	CL89	OB7	7	0.000142	0.995340
63	OB55	OB98	2	0.000147	0.995193
62	CL105	OB87	3	0.000154	0.995039
61	OB63	CL114	3	0.000159	0.994880
60	CL116	OB15	4	0.000160	0.994720
59	OB112	CL99	4	0.000170	0.994550
58	CL109	CL118	5	0.000170	0.994380
57	CL82	OB141	3	0.000176	0.994204

56	CL126	CL111	4	0.000194	0.994009
55	CL72	CL121	4	0.000205	0.993804
54	CL74	CL139	5	0.000208	0.993595
53	CL101	OB132	3	0.000213	0.993382
52	OB25	OB39	2	0.000220	0.993162
51	CL120	CL90	4	0.000242	0.992920
50	CL70	CL77	6	0.000247	0.992673
49	CL65	CL97	6	0.000259	0.992414
48	CL76	OB129	5	0.000269	0.992145
47	CL66	OB71	4	0.000276	0.991868
46	CL106	CL93	5	0.000278	0.991590
45	CL52	CL81	4	0.000286	0.991304
44	CL58	OB95	6	0.000292	0.991011
43	CL57	CL104	5	0.000301	0.990711
42	CL96	OB100	4	0.000303	0.990407
41	CL49	CL85	8	0.000322	0.990085
40	CL64	CL67	13	0.000366	0.989719
39	CL103	OB127	3	0.000369	0.989349
38	OB101	CL73	4	0.000385	0.988964
37	CL80	OB147	4	0.000394	0.988570
36	CL83	CL51	9	0.000395	0.988175
35	CL88	CL54	8	0.000396	0.987778
34	CL59	CL37	8	0.000427	0.987351
33	OB103	OB110	2	0.000477	0.986874
32	CL61	CL56	7	0.000496	0.986378
31	CL44	CL63	8	0.000501	0.985878
30	CL84	OB37	5	0.000537	0.985340
29	CL60	CL41	12	0.000540	0.984800
28	CL79	CL55	8	0.000549	0.984252
27	CL71	CL46	15	0.000567	0.983685
26	CL31	CL42	12	0.000633	0.983052
25	CL50	OB136	7	0.000634	0.982419
24	CL62	CL78	7	0.000742	0.981677
23	CL38	CL68	9	0.000805	0.980873
22	CL30	OB30	6	0.000896	0.979977
21	CL33	CL69	4	0.000976	0.979001
20	CL36	OB139	10	0.001087	0.977914
19	CL22	CL40	19	0.001141	0.976773

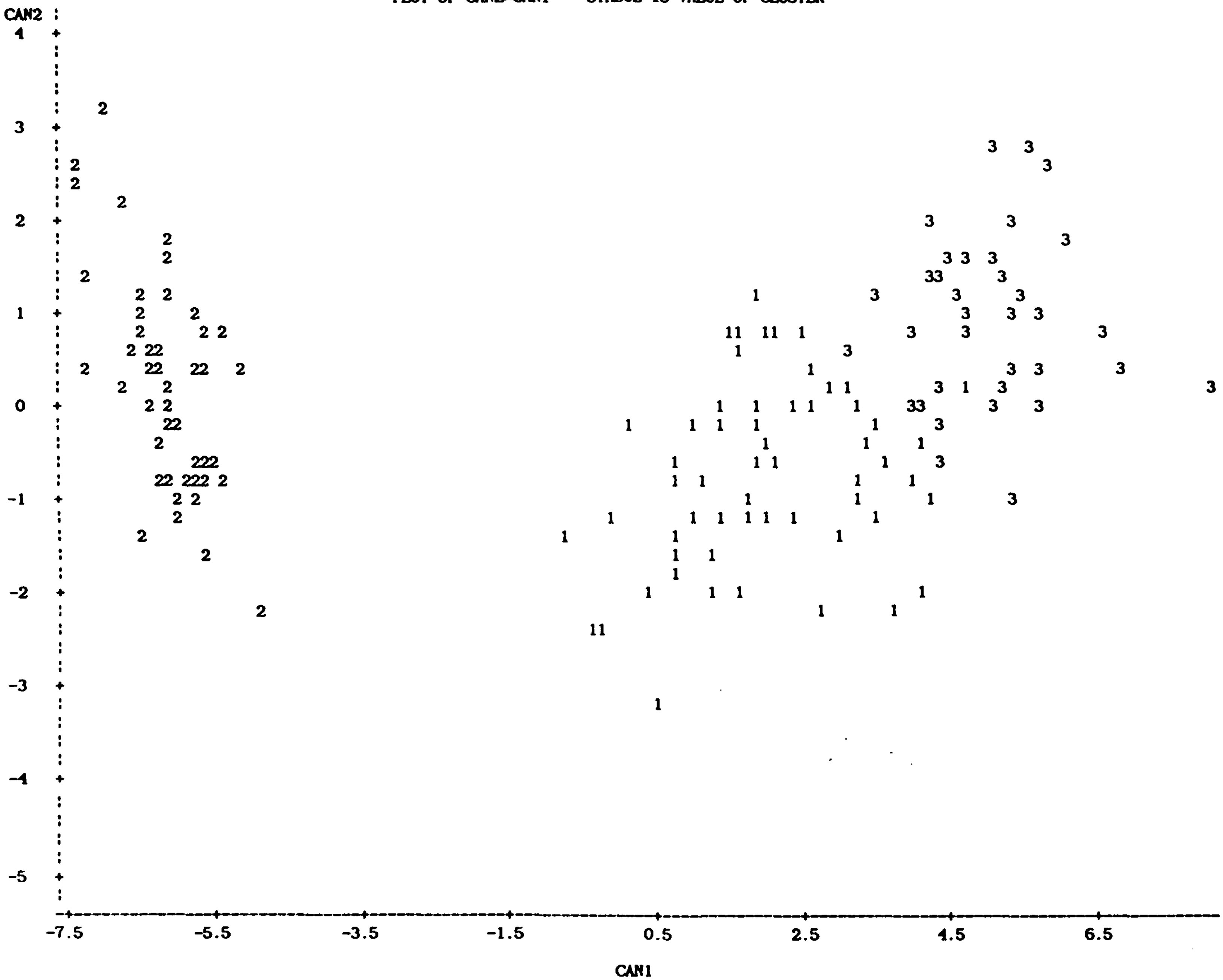
18	CL39	CL25	10	0.001249	0.975524
17	CL29	CL45	16	0.001351	0.974172
16	CL32	CL34	15	0.001462	0.972710
15	CL24	CL28	15	0.001641	0.971069
14	CL21	CL53	7	0.001873	0.969196
13	CL18	CL48	15	0.002271	0.966925
12	CL16	CL23	24	0.002274	0.964651
11	CL14	CL43	12	0.002500	0.962151
10	CL26	CL20	22	0.002694	0.959457
9	CL27	CL17	31	0.003060	0.956397
8	CL15	CL35	23	0.003095	0.953302
7	CL10	CL47	26	0.005811	0.947491
6	CL8	CL13	38	0.006042	0.941449
5	CL9	CL19	50	0.010532	0.930917
4	CL12	CL11	36	0.017245	0.913673
3	CL6	CL7	64	0.030051	0.883621
2	CL3	CL4	100	0.111026	0.772595
1	CL5	CL2	150	0.772595	0.000000

TABLE OF CLUSTER BY GROUP

CLUSTER / GROUP FREQUENCY PERCENT ROW PCT COL PCT	1	2	3	TOTAL
1	0 0.00 0.00 0.00	49 32.67 76.56 98.00	15 10.00 23.44 30.00	64 42.67
2	50 33.33 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	50 33.33
3	0 0.00 0.00 0.00	1 0.67 2.78 2.00	35 23.33 97.22 70.00	36 24.00
TOTAL	50 33.33	50 33.33	50 33.33	150 100.00

2x2 table displaying misclassifications. The true group allocation is labelled GROUP and is displayed as columns. The rows show the group allocation based on Ward's Clustering. One observation from GROUP 2 was misclassified and 15 from GROUP 3

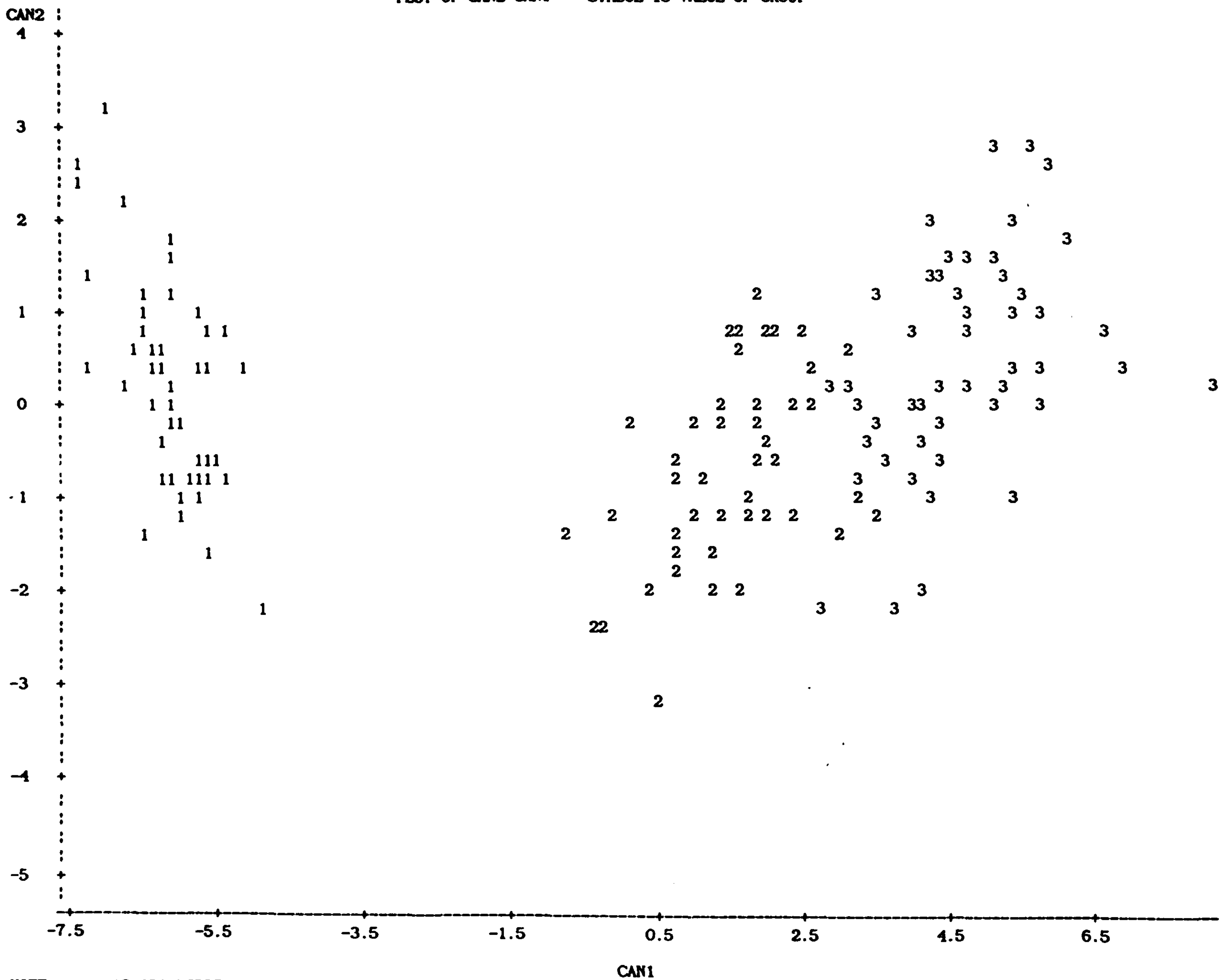
PLOT OF CAN2*CAN1 SYMBOL IS VALUE OF CLUSTER



NOTE: 13 OBS HIDDEN

This plot displays the results when three clusters are formed. The first two canonical variables (CAN1 and CAN2) for discriminating among the three clusters were computed and plotted to show cluster membership. The symbol plotted is the value of CLUSTER.

PLOT OF CAN2*CAN1 SYMBOL IS VALUE OF GROUP



NOTE: 13 OBS HIDDEN

This plot is exactly the same as the one on the previous page except the symbol plotted is the value of GROUP.

EQUAL VARIANCE MAXIMUM LIKELIHOOD METHOD

EIGENVALUES OF THE COVARIANCE MATRIX

	EIGENVALUE	DIFFERENCE	PROPORTION	CUMULATIVE
1	0.038779	0.025371	0.743072	0.74307
2	0.013409	.	0.256928	1.00000

ROOT-MEAN-SQUARE TOTAL-SAMPLE STANDARD DEVIATION = 0.161536
ROOT-MEAN-SQUARE DISTANCE BETWEEN OBSERVATIONS = 0.323072

NUMBER OF CLUSTERS	CLUSTERS JOINED		FREQUENCY OF NEW CLUSTER	LOG LIKELIHOOD RATIO	LOG LIKELIHOOD
74	OB4	OB6	2	.	1692.3
73	OB20	OB29	2	193.8	1498.5
72	OB49	OB52	2	121.7	1376.9
71	OB3	OB34	2	65.4659	1311.4
70	OB8	OB12	2	46.9378	1264.4
69	OB9	OB21	2	37.3953	1227.1
68	OB15	OB25	2	32.2544	1194.8
67	OB36	OB39	2	28.3152	1166.5
66	OB35	OB50	2	24.8173	1141.7
65	OB43	OB44	2	24.1433	1117.5
64	OB28	OB30	2	20.5569	1097.0
63	OB7	OB37	2	20.0080	1077.0
62	OB18	OB45	2	18.1699	1058.8
61	OB54	OB71	2	15.9815	1042.8
60	OB64	OB67	2	14.6690	1028.1
59	OB63	OB73	2	14.0356	1014.1
58	CL63	OB38	3	14.6115	999.5
57	CL70	OB46	3	14.9649	984.5
56	OB14	OB47	2	15.2553	969.3
55	CL71	CL66	4	14.4026	954.9

54	OB17	OB32	2	12.5620	942.3
53	OB31	OB55	2	12.1489	930.2
52	OB23	CL64	3	12.3538	917.8
51	OB40	OB48	2	12.0734	905.7
50	OB2	OB24	2	10.7887	894.9
49	CL74	CL57	5	11.8875	883.1
48	OB5	OB69	2	12.7666	870.3
47	OB61	OB75	2	11.8045	858.5
46	OB41	CL65	3	10.5419	847.9
45	OB58	CL60	3	10.0196	837.9
44	OB1	OB19	2	10.3168	827.6
43	CL59	OB68	3	9.6231	818.0
42	CL62	CL73	4	8.4188	809.6
41	CL72	CL61	4	7.3100	802.3
40	OB56	OB66	2	7.0361	795.2
39	CL50	CL68	4	6.6825	788.5
38	CL56	OB26	3	6.7983	781.7
37	CL49	OB33	6	6.4054	775.3
36	CL67	OB59	3	6.1010	769.2
35	OB10	CL52	4	6.3651	762.9
34	OB65	OB72	2	6.1221	756.7
33	CL55	OB42	5	5.6129	751.1
32	CL69	OB27	3	5.5075	745.6
31	CL54	OB62	3	5.7742	739.8
30	CL58	CL36	6	6.9427	732.9
29	OB70	OB74	2	6.9427	726.0
28	CL41	OB57	5	7.6944	718.3
27	CL53	CL45	5	7.4183	710.8
26	CL27	OB53	6	5.8181	705.0
25	CL39	CL42	8	5.6233	699.4
24	CL44	OB13	3	6.3334	693.1
23	OB60	CL47	3	5.8405	687.2
22	CL25	CL35	12	10.8017	676.4
21	CL33	CL51	7	8.1111	668.3
20	CL28	CL29	7	7.5983	660.7
19	OB16	OB22	2	6.8347	653.9
18	CL43	CL34	5	6.1380	647.7
17	CL38	CL46	6	5.3666	642.4

16	CL37	CL48	8	4.1815	638.2
15	CL30	CL40	8	9.5200	628.7
14	CL22	CL32	15	13.1016	615.6
13	CL16	CL17	14	9.5519	606.0
12	CL21	CL26	13	7.5668	598.5
11	OB11	OB51	2	7.8476	590.6
10	CL11	CL31	5	7.6117	583.0
9	CL14	CL15	23	5.4097	577.6
8	CL20	CL18	12	9.1725	568.4
7	CL24	CL9	26	5.7861	562.6
6	CL13	CL19	16	6.6930	555.9
5	CL12	CL23	16	4.7377	551.2
4	CL7	CL6	42	6.9678	544.2
3	CL4	CL8	54	-10.1224	554.4
2	CL5	CL10	21	-5.2539	559.6
1	CL3	CL2	75	10.6897	548.9

TABLE OF CLUSTER BY GROUP

CLUSTER / GROUP FREQUENCY PERCENT ROW PCT COL PCT	1	2	TOTAL
1	27 36.00 50.00 90.00	27 36.00 50.00 60.00	54 72.00
2	3 4.00 14.29 10.00	18 24.00 85.71 40.00	21 28.00
TOTAL	30 40.00	45 60.00	75 100.00

EQUAL VARIANCE MAXIMUM LIKELIHOOD METHOD

EIGENVALUES OF THE COVARIANCE MATRIX

	EIGENVALUE	DIFFERENCE	PROPORTION	CUMULATIVE
1	4.22824	3.98557	0.924619	0.92462
2	0.24267	0.16446	0.053066	0.97769
3	0.07821	0.05437	0.017103	0.99479
4	0.02384	.	0.005212	1.00000

ROOT-MEAN-SQUARE TOTAL-SAMPLE STANDARD DEVIATION = 1.06922
ROOT-MEAN-SQUARE DISTANCE BETWEEN OBSERVATIONS = 3.02422

NUMBER OF CLUSTERS	CLUSTERS JOINED		FREQUENCY OF NEW CLUSTER	LOG LIKELIHOOD RATIO	LOG LIKELIHOOD
149	OB117	OB125	2	.	.
148	OB14	OB46	2	.	4016.3
147	OB6	OB48	2	413.1	3603.2
146	OB1	OB50	2	240.5	3362.7
145	OB137	OB149	2	169.8	3192.9
144	OB17	OB40	2	131.1	3061.7
143	OB52	OB62	2	199.1	2862.6
142	OB5	OB38	2	148.0	2714.6
141	OB124	OB144	2	117.6	2597.0
140	OB119	OB126	2	97.4599	2499.5
139	OB65	OB97	2	83.0879	2416.4
138	OB53	OB54	2	72.3253	2344.1
137	OB23	OB35	2	63.9628	2280.1
136	OB22	OB29	2	57.2775	2222.9
135	OB11	OB16	2	51.8105	2171.1
134	OB76	OB90	2	47.2564	2123.8
133	OB59	OB69	2	43.4040	2080.4
132	OB60	OB61	2	40.1028	2040.3

131	OB21	OB49	2	37.2422	2003.1
130	OB8	OB47	2	34.7396	1968.3
129	OB3	OB44	2	32.5317	1935.8
128	OB9	CL148	3	45.5238	1890.3
127	CL146	OB12	3	41.7725	1848.5
126	OB111	OB122	2	39.5980	1808.9
125	OB118	OB121	2	36.8022	1772.1
124	OB64	OB85	2	34.3527	1737.7
123	CL135	OB27	3	34.9040	1702.8
122	CL138	OB73	3	32.5557	1670.3
121	OB86	OB96	2	38.1524	1632.1
120	OB72	OB93	2	35.5383	1596.6
119	OB66	OB91	2	33.2382	1563.3
118	OB13	OB41	2	31.1987	1532.1
117	CL147	OB32	3	30.9345	1501.2
116	CL142	OB43	3	31.4852	1469.7
115	CL136	OB24	3	29.5228	1440.2
114	OB134	OB150	2	31.0074	1409.2
113	OB130	OB140	2	29.2066	1380.0
112	CL134	OB84	3	32.4336	1347.6
111	CL143	OB79	3	30.3674	1317.2
110	OB123	OB142	2	29.5706	1287.6
109	OB114	OB131	2	27.9159	1259.7
108	OB109	OB116	2	26.4224	1233.3
107	OB68	OB88	2	25.0675	1208.2
106	OB4	CL118	3	25.6709	1182.5
105	CL127	CL131	5	25.2102	1157.3
104	OB106	OB146	2	25.2138	1132.1
103	OB107	OB108	2	23.9664	1108.2
102	OB51	OB94	2	22.8254	1085.3
101	OB75	OB81	2	21.7779	1063.5
100	OB56	OB70	2	20.8127	1042.7
99	CL141	OB143	3	19.9336	1022.8
98	CL130	OB33	3	19.0290	1003.8
97	CL105	CL128	8	18.9414	984.8
96	OB31	OB42	2	20.0692	964.8
95	CL117	CL123	6	19.0634	945.7
94	OB82	CL140	3	18.9408	926.8

93	CL132	OB74	3	18.1089	908.6
92	OB2	CL129	3	17.3356	891.3
91	OB113	CL149	3	18.1585	873.2
90	OB10	OB20	2	17.6839	855.5
89	CL122	CL124	5	18.7901	836.7
88	CL100	OB80	3	17.9123	818.8
87	OB89	OB92	2	17.5339	801.2
86	CL92	OB18	4	17.7140	783.5
85	OB102	CL113	3	16.9813	766.5
84	OB78	OB133	2	17.3308	749.2
83	OB26	OB36	2	16.6790	732.5
82	OB58	CL112	4	16.8660	715.7
81	CL145	OB145	3	16.5405	699.1
80	CL94	OB115	4	15.9775	683.1
79	CL97	CL137	10	15.3316	667.8
78	CL91	OB148	4	14.9811	652.8
77	OB105	OB120	2	14.4466	638.4
76	OB28	OB45	2	13.9662	624.4
75	CL125	OB138	3	13.8027	610.6
74	CL144	OB19	3	14.1647	596.5
73	CL108	OB128	3	14.0609	582.4
72	OB57	CL107	3	13.5435	568.8
71	CL85	CL110	5	13.9353	554.9
70	CL95	CL115	9	13.4751	541.4
69	OB77	OB99	2	15.0862	526.4
68	CL84	OB83	3	14.5910	511.8
67	OB104	OB135	2	15.0893	496.7
66	CL74	OB34	4	15.1889	481.5
65	CL133	OB67	3	15.2430	466.2
64	CL79	OB7	11	14.9585	451.3
63	CL111	CL119	5	14.6028	436.7
62	CL102	OB87	3	14.8579	421.8
61	CL116	OB15	4	14.3220	407.5
60	OB55	OB98	2	13.9645	393.5
59	OB63	CL114	3	13.8010	379.7
58	OB112	CL99	4	13.7858	365.9
57	CL77	OB141	3	14.5532	351.4
56	CL126	CL109	4	14.1081	337.3

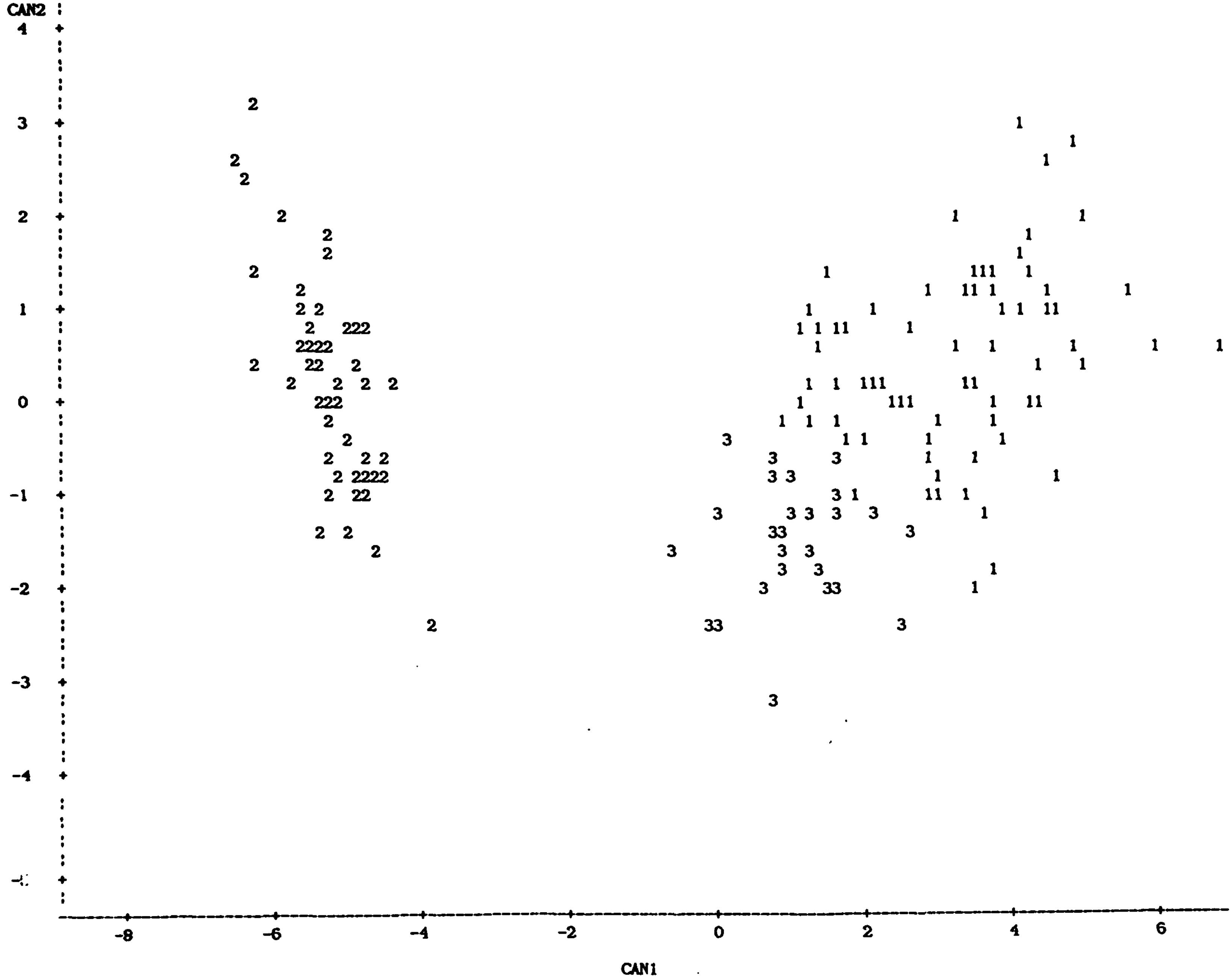
55	CL72	CL139	5	13.6412	323.6
54	CL69	CL121	4	13.8841	309.8
53	CL68	CL75	6	14.2400	295.5
52	CL89	CL87	7	13.9381	281.6
51	CL70	CL98	12	13.1141	268.5
50	CL66	CL96	6	13.4041	255.1
49	CL103	OB132	3	12.9165	242.1
48	OB25	OB39	2	14.0625	228.1
47	CL106	CL90	5	13.9022	214.2
46	CL50	CL83	8	14.0355	200.1
45	CL78	OB129	5	13.5881	186.6
44	CL57	CL104	5	13.3811	173.2
43	CL48	CL76	4	12.9790	160.2
42	CL52	CL93	10	12.5952	147.6
41	CL65	OB71	4	12.1714	135.4
40	CL63	OB95	6	11.7183	123.7
39	CL88	CL55	8	11.9019	111.8
38	CL58	CL81	7	12.1729	99.6391
37	CL64	CL47	16	12.5030	87.1362
36	CL61	CL46	12	11.5860	75.5502
35	CL53	OB127	7	12.7557	62.7945
34	OB101	CL71	6	13.2697	49.5248
33	CL59	CL56	7	12.6880	36.8369
32	CL38	OB147	8	12.3602	24.4767
31	CL82	CL54	8	11.9039	12.5727
30	CL40	CL60	8	11.2545	1.3182
29	CL101	OB100	3	12.5860	-11.2678
28	CL42	CL120	12	13.8733	-25.1411
27	CL44	OB110	6	13.8089	-38.9500
26	CL86	CL51	16	13.7295	-52.6795
25	CL26	OB37	17	12.6505	-65.3300
24	CL62	CL31	11	11.9355	-77.2655
23	CL35	OB136	8	12.8183	-90.0838
22	CL34	CL73	9	12.2185	-102.3
21	CL37	CL36	28	10.8425	-113.1
20	CL80	CL45	9	9.8004	-122.9
19	OB103	CL67	3	15.3968	-138.3
18	CL33	CL32	15	15.2276	-153.6

17	CL25	OB30	18	15.8760	-169.4
16	CL23	CL20	17	15.6063	-185.1
15	CL30	CL28	20	15.1236	-200.2
14	CL18	CL22	24	12.1927	-212.4
13	CL15	OB139	21	15.9966	-228.4
12	CL24	CL39	19	17.4915	-245.9
11	CL13	CL29	24	20.6397	-266.5
10	CL27	CL49	9	19.4921	-286.0
9	CL21	CL43	32	17.9355	-303.9
8	CL19	CL10	12	20.3710	-324.3
7	CL12	CL16	36	43.3135	-367.6
6	CL9	CL17	50	39.8843	-407.5
5	CL11	CL41	28	34.1956	-441.7
4	CL7	CL14	60	78.4355	-520.1
3	CL4	CL8	72	163.6	-683.7
2	CL3	CL5	100	198.1	-881.8
1	CL6	CL2	150	697.7	-1579.5

TABLE OF CLUSTER BY GROUP

CLUSTER (FREQUENCY / PERCENT / ROW PCT / COL PCT)	GROUP 1	GROUP 2	GROUP 3	TOTAL
1	0	23	49	72
	0.00	15.33	32.67	48.00
	0.00	31.94	68.06	
	0.00	46.00	98.00	
2	50	0	0	50
	33.33	0.00	0.00	33.33
	100.00	0.00	0.00	
	100.00	0.00	0.00	
3	0	27	1	28
	0.00	18.00	0.67	18.67
	0.00	96.43	3.57	
	0.00	54.00	2.00	
TOTAL	50	50	50	150
	33.33	33.33	33.33	100.00

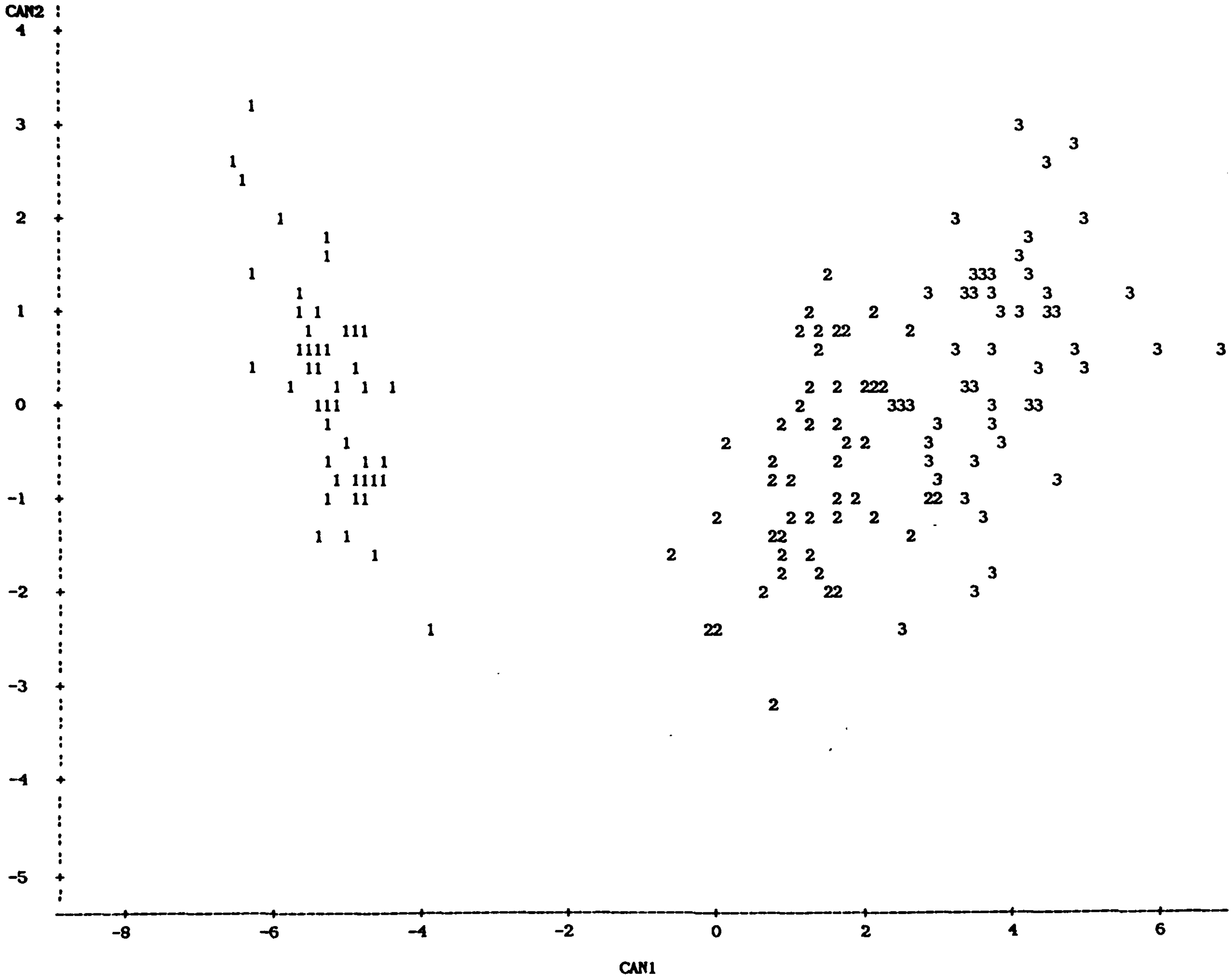
PLOT OF CAN2*CAN1 SYMBOL IS VALUE OF CLUSTER



NOTE: 7 OBS HIDDEN

SAS

PLOT OF CAN2*CAN1   SYMBOL IS VALUE OF GROUP



NOTE: 7 OBS HIDDEN

REFERENCES

Basford, K.E. (1986). An intoduction to the use of mixture models in clustering. Cornell University Biometrics Unit Technical Report BU-920-M and Mathematical Sciences Institute Technical Report '86-38, Cornell University.

Basford, K.E., Federer, W.T. and Miles-McDermott, N.J. (1987). Illustrative examples of clustering using the mixture method and two comparable methods from SAS. Cornell University Biometrics Unit Technical Report BU-921-M and Mathematical Sciences Institute Technical Report '87-5, Cornell University.

Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. Annals of Eugenics 7, 179-188.

Habbema, J.D.J., Hermans, J. and van den Broek, K. (1974). A stepwise discriminant analysis program using density estimation. Compstat 1974: Proceedings in Computational Statistics. Vienna: Physica-Verlag, 101-110.

McLachlan, G.J. and Basford, K.E. (1987). Mixture Models: Inference and Applications to Clustering. Marcel Dekker, New York (in the process of publication).

END

10-87

DTIC