FORTRAN II PROGRAM FOR
MULTIVARIATE DISCRIMINANT
ANALYSIS USING AN IBM 1620
COMPUTER

By
JOHN C. DAVIS
and
ROBERT J. SAMPSON
Idaho State University

COMPUTER CONTRIBUTION 4
State Geological Survey
The University of Kansas, Lawrence
1966
Editor's Remarks

Many aids are available to the geologist who wishes to enter the computer world. Most computer manufacturers have literature available explaining different aspects of computers such as "You and the computer, a students' guide," available without charge from General Electric, "What is a computer?" available from NCR Electronic Data Processing, and "General information manual, Introduction to IBM data processing systems" distributed by IBM, Data Processing Division. There are also many excellent books on computers available commercially.

Glossaries of computer vocabulary include: "Automatic data processing... glossary" from Superintendent of Documents, U.S. Government Printing Office, Washington 25, D. C. ($0.40); "Reference manual glossary for information processing" from IBM Technical Publications Department, 112 East Park Road, White Plains, New York (available to users free-of-charge); "Glossary of process computer terms" from General Electric, Process Computer Section, Industry Control Department, Phoenix, Arizona (available on request). A list of data processing service centers is published by ADAPSO, 947 Old York Road, Abingdon, Pa. 19001 ($1.00).

Information on computer programs of interest to geologists can be obtained from several sources, some of which were mentioned in the editorial remarks of Kansas Geological Survey COMPUTER CONTRIBUTION 2. Organizations which have lists of available computer programs of interest to geologists include the Kansas Geological Survey (available on request). National Oceanographic Data Center, Data Processing Division, U.S. Naval Oceanographic Office, Washington, D. C. 20390 ($0.60), and U.S. Geological Survey, Chief, Branch of Computation, Room 1451, Interior Building, 18th and C Streets, N.W., Washington, D.C. 20242 (available on request).

...The problem of assembling a library of computer programs especially adapted to geological data is an important facet of data analysis. ...

W. C. Krumbein, June 29, 1962

Some of the problems of the availability of computer programs of interest to workers in the earth sciences would be lessened with the establishment of an organization such as GEOCOMP. GEOCOMP was organized by J. W. Horaugh for gathering and disseminating computer programs and information about computer programs applicable to the earth sciences. For the most part programs are now being distributed on an individual basis between "workers in the field," and with the noted exceptions are not generally available.

The Kansas Survey is the only geological organization known to be actively distributing computer program decks as well as data decks. The programs are sold for a limited time at a nominal cost. The programs are for Burroughs B5500, Elliott 803C, IBM 1620, 7040, and 7090/1401 computer systems. A list of available decks is given below.

<table>
<thead>
<tr>
<th>ALGOL</th>
<th>BALGOL</th>
<th>FORTRAN II</th>
<th>FORTRAN IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marine Simulation (CC 1)</td>
<td>$20.00</td>
<td>$10.00</td>
<td>$10.00</td>
</tr>
<tr>
<td>2D Regression (CC 2)</td>
<td>$10.00</td>
<td></td>
<td>$25.00</td>
</tr>
<tr>
<td>Trend-6 (CC 3)</td>
<td></td>
<td>$5.00</td>
<td></td>
</tr>
<tr>
<td>Discrim (CC 4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trend-3 (SDP 3)</td>
<td></td>
<td>$10.00</td>
<td></td>
</tr>
<tr>
<td>Match-Coeff (SDP 4)</td>
<td></td>
<td>$2.00</td>
<td></td>
</tr>
<tr>
<td>Correlation of distance Coeff (SDP 9)</td>
<td></td>
<td>$5.00 (each)</td>
<td></td>
</tr>
<tr>
<td>Time-trend (SDP 12)</td>
<td></td>
<td>$5.00</td>
<td>$5.00</td>
</tr>
<tr>
<td>Covap (SDP 13)</td>
<td></td>
<td>$15.00</td>
<td></td>
</tr>
<tr>
<td>Trend-3 (SDP 14)</td>
<td></td>
<td>$25.00</td>
<td>$25.00</td>
</tr>
<tr>
<td>Cross-Association (SDP 23)</td>
<td></td>
<td></td>
<td>$10.00</td>
</tr>
<tr>
<td>Single and doube Fourier (SDP 24)</td>
<td></td>
<td>$5.00</td>
<td>$15.00</td>
</tr>
<tr>
<td>Precambrian wells (SDP 25)</td>
<td></td>
<td>$50.00</td>
<td></td>
</tr>
<tr>
<td>List of about 2,600 Precambrian wells</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trend-4 (SDP 26)</td>
<td></td>
<td></td>
<td>$7.50</td>
</tr>
<tr>
<td>Sediment analysis (SDP 28)</td>
<td></td>
<td></td>
<td>$10.00</td>
</tr>
<tr>
<td>4D Trend (KGS 8171)</td>
<td></td>
<td></td>
<td>$10.00</td>
</tr>
<tr>
<td>Conversion of T&amp;R to Cartesian coordinates (KGS 8170-3)</td>
<td></td>
<td>$5.00</td>
<td>$5.00</td>
</tr>
</tbody>
</table>

The program presented here, "FORTRAN II program for multivariate discriminant analysis using an IBM 1620 computer," by J. C. Davis and R. J. Sampson should be of interest to those who wish to assign "samples to previously defined populations on the basis of a number of variables considered simultaneously." This method will obviously be of use to workers in numerical taxonomy; the authors point out how this method has been used to solve problems in stratigraphy and sedimentology. Additional papers of interest on this subject include: "The palaeoecology of the Foraminifera of the Chalk Marl," by T. F. Burnaby (Palaeontology, v. 4, pt. 4, 1962, p. 599-608); "Multiple discriminant functions," by Emilio Casetti (ONR Tech. Rept. 11, Northwestern Univ., 1964, 63 p.); "On distinguishing basaltic lavas of circumsocoeanic and oceanic-island type by means of discriminant function," by F. Chayes and D. Velez (Am. Jour. Sci., v. 263, 1965, p. 205-222); "An example of the quantitative study of echinoid morphology," by J. C. Kelly (Contr. to Geol., v. 4, no. 1, Wyo. Univ., 1965, p. 15-20); and "Fourier series characterization of cyclic sediments for stratigraphic correlation," by F. W. Preston and J. H. Henderson (Kansas Geol. Survey Bull., 169, 1964, p. 415-425).

The Survey will make available for a limited time the card deck (in FORTRAN II) pertaining to the discriminant analysis program at a cost of $5.00.

Comments and suggestions concerning the COMPUTER CONTRIBUTION series are welcome and should be addressed to the editor.
FORTRAN II PROGRAM FOR MULTIVARIATE DISCRIMINANT ANALYSIS

USING AN IBM 1620 COMPUTER

By

JOHN C. DAVIS and ROBERT J. SAMPSON

INTRODUCTION

Multivariate discriminant analysis is a statistical method of assigning samples to previously defined populations on the basis of a number of variables considered simultaneously. If two sets of samples can be assigned to different populations on the basis of prior knowledge, they can be used to establish criteria for the classification of additional samples. In addition, the validity of the original classification can be tested. Discriminant analysis has been used to distinguish marine from fresh-water shales (Potter, Shimp, and Witters, 1963), to establish tectonic settings of sandstones (Middleton, 1962) and volcanics (Chayes, 1964), to classify depositional environments of carbonates (Krumbein and Graybill, 1964), and to distinguish ore-bearing from barren sediments (Griffiths, 1957). Melton (1964) has used the method to determine parameters controlling cement distribution in sandstone. The technique was originally developed by Fisher (1936) and has been widely used in biometrics and paleobiometrics (for applications, see Miller and Kahn, 1962, p. 277).

A population described by \( K \) variables may be pictured as a cluster of sample points in \( K \) dimensional space. A second population, described by the same \( K \) variables, consists of a second cluster of points. Discriminant analysis is the computation of a \( K \) dimensional plane that most efficiently separates the two clusters. An unknown sample is classified as belonging to one group or the other, depending upon which side of the plane it falls. In addition, the "locations" of the two populations are described by the \( K \) dimensional coordinates of their multivariate means. The degree of distinctness of the two groups is measured by the "distance" between these means.

Several methods of discriminant analysis have been devised. However, with one exception, all techniques described in the geological literature involve generation, inversion, and solution of a matrix. Matrix inversion is so laborious that it is essentially limited to computer operations, and in the case of large matrices, to large computers. Because the IBM 1620 has limited storage capacity, any approach of this type is restricted to small numbers of variables. Matrix inversion is avoided by using a modification of a procedure suggested by Rao (1952, p. 262-264) in which samples are operated upon one at a time. The method used by Miller and Kahn (1962, p. 276-283) is similar, but results are not comparable to those produced by matrix inversion. An excellent introduction to the procedure used here is given by Li (1964, p. 403-415).

MATHEMATICAL DEVELOPMENT

Suppose a set of \( n_1 \) samples consisting of measurements of variables \( A_1, B_1, C_1, \ldots, K_1 \) has been taken from population 1. A second set of \( n_2 \) samples consisting of measurements of the same \( K \) variables has been taken from population 2. The sums of variables, sums of squares of the variables, and sums of cross-products for each population sample are accumulated. These are used in the following series of equations to produce the linear discriminant function and related terms.

\[
\bar{A}_1 = \frac{\sum A_1}{n_1} \quad \bar{A}_2 = \frac{\sum A_2}{n_2} \\
\bar{K}_1 = \frac{\sum K_1}{n_1} \quad \bar{K}_2 = \frac{\sum K_2}{n_2}
\]

\( \Delta \bar{A} = \bar{A}_1 - \bar{A}_2 \)

\( \Delta \bar{K} = \bar{K}_1 - \bar{K}_2 \)

\[
SS_A = \left( \frac{\sum A_1^2}{n_1} + \frac{\sum A_2^2}{n_2} \right) - \left( \frac{\sum A_1}{n_1} \right)^2 - \left( \frac{\sum A_2}{n_2} \right)^2 \\
\vdots \\
SS_K = \left( \frac{\sum K_1^2}{n_1} + \frac{\sum K_2^2}{n_2} \right) - \left( \frac{\sum K_1}{n_1} \right)^2 - \left( \frac{\sum K_2}{n_2} \right)^2
\]
SS_{A} \ldots SS_{K} are unbiased estimates of the pooled variances of the K variables.

\[
SS_{AB} = \frac{\Sigma A_{1}B_{1} + \Sigma A_{2}B_{2}}{n_{1}} - \frac{\Sigma A_{1}\Sigma B_{1}}{n_{1}} + \frac{\Sigma A_{2}\Sigma B_{2}}{n_{2}}
\]

\[
SS_{AC} = \frac{\Sigma A_{1}C_{1} + \Sigma A_{2}C_{2}}{n_{1}} - \frac{\Sigma A_{1}\Sigma C_{1}}{n_{1}} + \frac{\Sigma A_{2}\Sigma C_{2}}{n_{2}}
\]

\[
\vdots
\]

\[
SS_{AK} = \frac{\Sigma A_{1}K_{1} + \Sigma A_{2}K_{2}}{n_{1}} - \frac{\Sigma A_{1}\Sigma K_{1}}{n_{1}} + \frac{\Sigma A_{2}\Sigma K_{2}}{n_{2}}
\]

\[
SS_{BK} = \frac{\Sigma B_{1}K_{1} + \Sigma B_{2}K_{2}}{n_{1}} - \frac{\Sigma B_{1}\Sigma K_{1}}{n_{1}} + \frac{\Sigma B_{2}\Sigma K_{2}}{n_{2}}
\]

\[
SS_{(K-1)K} = \frac{\Sigma (K-1)_{1}(K-1)_{2}}{n_{1}} - \frac{\Sigma (K-1)_{1}(K-1)_{1}}{n_{1}} + \frac{\Sigma (K-1)_{2}(K-1)_{2}}{n_{2}}
\]

All possible combinations of the K variables are represented by \( AB, AC, \ldots, (K-1)K \). The number of combinations is equal to \( \binom{K}{2} \), or in the case of 20 variables, \( \frac{20!}{2!} = 190 \). Unbiased estimates of the covariation of the K variables are \( SS_{AB}, SS_{AC}, \ldots, SS_{(K-1)K} \).

Covariance and variance estimates are set into linear equations which are equal to the values \( \Delta A, \Delta B, \ldots, \Delta K \) times the constant \( (n_{1}+n_{2}-2) \).

\[
SS_{A} = SS_{AB} + SS_{AC} + \ldots + SS_{AK}
\]

\[
SS_{B} = SS_{AB} + SS_{BC} + SS_{AC} + \ldots + SS_{BK}
\]

\[
SS_{K} = SS_{BK} + SS_{CK} + \ldots + SS_{(K-1)K}
\]

The \( K \) simultaneous equations are then solved to produce values of \( \lambda_{A}, \lambda_{B}, \ldots, \lambda_{K} \).

The linear discriminant function, \( R \), takes the form

\[
R = \lambda_{A} A + \lambda_{B} B + \lambda_{C} C + \ldots + \lambda_{K} K
\]

The discriminant index, \( R_{0} \), is found by using the means of the combined population samples in the discriminant function

\[
R_{0} = \frac{\Sigma A_{1} + \Sigma A_{2}}{n_{1}+n_{2}} + \frac{\Sigma B_{1} + \Sigma B_{2}}{n_{1}+n_{2}} + \ldots + \frac{\Sigma K_{1} + \Sigma K_{2}}{n_{1}+n_{2}}
\]

By substituting the means for each set of data separately into equation (6), discriminant values for each population sample are determined. An unknown sample is assigned to one of the two populations by substituting the appropriate values for the \( K \) variables into the discriminant function. If the discriminant value is on the \( R_{1} \) side of \( R_{0} \), it is assigned to population 1. If it is on the \( R_{2} \) side of \( R_{0} \), it is assigned to population 2. Probabilities of misclassifying samples from either population 1 or 2 are equal if \( R_{0} \) is used as the discriminant index. Other values may be chosen which will decrease the probability of misclassification in one population while increasing the probability of misclassification in the other.

Discriminant analysis is based on the assumption that the two samples are drawn from different populations. It is desirable to test this assumption, as a discriminant analysis will be useless if the multivariate means of the two groups are not statistically distinct. A test for significance may be derived from Mahalanobis' Generalized Distance (\( D^{2} \)), which is a measure of the "distance" between the multivariate means of the two sample clusters.

Middleton (1962, p. 121) derives \( D^{2} \) by substituting the differences between the variable means into the discriminant function.

\[
d^{2} = \lambda_{A}\Delta A + \lambda_{B}\Delta B + \lambda_{C}\Delta C + \ldots + \lambda_{K}\Delta K
\]

The significance of the difference between the multivariate means is tested by

\[
F_{K,n_{1}+n_{2}-K-1}(\frac{n_{1}n_{2}}{n_{1}+n_{2}-2}) \frac{R_{0}^{2}}{K-1} \geq \frac{R^{2}}{K-1} D^{2} \]

The contribution that each variable makes to the total distance between the multivariate means is determined by the equations

\[
% \text{contributed by } A = \frac{\lambda_{A}\Delta A}{D^{2}} \times 100
\]

\[
% \text{contributed by } B = \frac{\lambda_{B}\Delta B}{D^{2}} \times 100
\]

\[
% \text{contributed by } K = \frac{\lambda_{K}\Delta K}{D^{2}} \times 100
\]

In this manner, variables making insignificant contributions to \( D^{2} \) can be detected and eliminated from future considerations. This criterion tests only for the contribution made directly by the variable being considered. It does not consider contributions made by correlations between variables. It two or more of the variables are not truly independent, their interactions contribute to \( D^{2} \) to a greater extent than this test suggests. Since the discriminant function is based on the assumption of independence of variables, known sets of dependent variables should not purposely be included.

Variables with low or negative contributions to \( D^{2} \) should be deleted on successive runs until the F-value for the test of significance of the difference between the multivariate means (equation 9) is reduced below the value of the assigned significance level. All variables contribute to the function, but it is obviously more economical to confine considerations to those necessary for the desired degree of discrimination. This condition reduces labor involved in subsequent data gathering. Schultz and Goggins (1961) describe a method for identifying the absolutely
most potent discriminants. Their method may yield
different variables than those chosen by this method.
However, their technique is not compatible with
computational procedures used to establish the dis-
graminant function in this program, because they use
a stepwise modification of the Doolittle procedure to
solve the simultaneous equations.

PROGRAM DESCRIPTION

The program described here computes the
discriminant function for two sample sets, using up
to 20 variables, and having any number of samples
in either or both sets. Sums, sums of squares, and
sums of cross-products are accumulated as the sample
data are read into the computer. From these, un-
biased estimates of the sample variances and covari-
ances (see equations 3 and 4) are computed and
become terms in a series of $K$ simultaneous equations.
These equations are solved for $\lambda_k$ by a mod-
ification of the Gauss-Jordan method (Golden, 1965,
p. 93). The values of $\lambda_k$ are constants in the
linear discriminant function.

Output from the program consists of the constant terms ($\lambda_k$) of the linear discriminant
function, the discriminant index ($R_0$), the discrimi-
nant values for each of the two groups ($R_1$ and $R_2$),
Mahalanobis’ Generalized Distance ($D^2$), and the
percent contributed by each variable to $D^2$. An $F$-
value for determining the significance of the dif-
ference between the multivariate means is punched,
accompanied by the first ($K$) and second $(n_1+n_2-K-1)$
degrees of freedom. The machine then will accept
data from individual samples and classify them as
belonging to population 1 or 2 by computing their multivariate “location.”

This program is designed for use on an IBM
1620 computer with 20K bits storage, automatic
divide, and indirect addressing. A 1622 Card Read
Punch or equivalent, a 407 accounting machine or
similar line printer, PDQ FORTRAN Processor C2
without reread version 1 modification 0, and PDQ
FORTRAN fixed format subroutines without reread
version 1 modification 0 (IBM User’s Group Program
2.0.031) are also required.

If ten or fewer variables are used, the data
for each sample are punched in seven-column fields
on a single card. If 11 to 20 variables are used,
the additional data are punched on a second card
in seven-column fields. Columns 71-80 on the
first card are used for sample numbers. After load-
ing the program object deck and the PDQ fixed
format subroutines, data are fed into the computer
in the following manner.

Card 1. This card contains a two-digit
number specifying the number of variables to be con-
sidered. The number must be right-justified in
columns 1 and 2.

Card 2. This is the first of any number of
cards containing the $K$ variables of samples in set
1. If Card 1 specified any number between or
including 11 to 20, the data cards will be read in
pairs.

Data cards from set 1 are separated from those
of set 2 by a card containing the number 9.0E 48
punched in columns 1 through 7. When this number
is read, data following it will be accumulated in
locations reserved for set 2. The second set of data
is followed by a card with 9.0E 48 punched in
columns 1 through 7, which directs the computer to
begin computations and punch output. If individual
samples are to be classified following completion of
the discriminant function, SENSE SWITCH 1 should
be off. If it is on, the computer will return to
manual mode. If SENSE SWITCH 1 is off, the computer
will accept individual sample cards one at a time (or
pair at a time if there are more than ten variables)
and will compute and punch out the discriminant value
of each sample.

The program has been tested using data pub-
lished by Miller and Kahn (1962, p. 250), giving
measurements on four characters in a series of indi-
viduals of Merychoidodon culbertsoni and Prodesmat-
choerus meekii, and on four-variable data on car-
bonate environments published by Krumbein and Gray-
bill (1965, p. 363). The program was also tested with
seven variables, using the trace element composition
of marine versus fresh-water argillaceous sediments
reported by Potter, Shimp, and Witters (1963, p. 685).

A more complete test of the program was made using
data on 19 measurements made on specimens of the
bryozoan Constellaria (Upper Ordovician) collected
from central Kentucky. Unpublished data were sup-
plied by John Cutler, Idaho State University. In all
cases, the discriminant program performed satisfac-
torily. The length of time required to complete the
program depends upon the number of samples and
variables used. Approximately 5 minutes per vari-
able are required to process input consisting of 20
samples per set.

This program was written by the authors at
the Idaho State University Computer Center as a
part of ISU Geology Department Research Project 51.
Dr. Robert S. Cochran of the Statistics Laboratory,
University of Wyoming, kindly reviewed the analysis
of the discriminant function. Machine time was
donated by the ISU Computer Center on equipment
furnished in part under NSF Grant GP-2275.
PROGRAM LISTING

C DISCRIMINATE ANALYSIS OF TWO POPULATIONS, USING UP TO TWENTY
C VARIABLES AND AN UNLIMITED NUMBER OF SAMPLES IN EITHER OR BOTH
C GROUPS. COMPUTES DISCRIMINATE EQUATIONS AND AN F-VALUE FOR A
C TEST OF EQUALITY OF MULTIVARIATE MEANS OF THE TWO POPULATIONS.
C
ISU GEOLOGY DEPARTMENT RESEARCH PROJECT 51 2-19-66

BEGIN TRACE
DIMENSION N(2) , A(2,20) , A2(2,20) , C(20,1) , X(20)
90 READ 1000, K
   K1=(K-1)/10+1
   N(1)=0
   N(2)=0
   DO 100 I=1,2
   DO 100 J=1,20
      A(I,J)=0.0
   100    A2(I,J)=0.0
   DO 101 I=1,20
   DO 101 J=1,21
   101   C(I,J)=0.0
   DO 102 I=1,2
      READ 1001 , X(I,1), X(I,2), X(I,3), X(I,4), X(I,5), X(I,6), X(I,7), X(I,8), X(I,9), X(I,10)
      IF (X(I,1) -9.0E48) 2,102,2
   2      GO TO (3, 4), K1
   4      READ 1001 , X(I,11), X(I,12), X(I,13), X(I,14), X(I,15), X(I,16), X(I,17), X(I,18),
           X(I,19), X(I,20)
   3      N(1)=N(1)+1
   DO 103 J=1,K
      A(I,J)=A(I,J)+X(J)
   103    A2(I,J)=A2(I,J)+X(J)**2
   DO 104 J=2,K
      M=J-1
   104    DO 105 L=1,M
   105   C(L,J)=C(L,J)+X(L)*X(J)
   CONTINUE
   GO TO 1
   102 CONTINUE
   AN1=N(1)
   AN2=N(2)
   AN3=AN1+AN2-2,0
   DO 106 I=1,K
      C(I,K+1)=(A(I,1)/AN1-A(I,2)/AN2)*AN3
   106    C(I,I)=A2(I,1)+A2(I,2)-(A(I,1)**2)/AN1-(A(I,2)**2)/AN2
   DO 120 I=2,K
      J=I-1
   120    DO 121 L=I,J
   121   C(L,I)=C(L,I)-(A(I,1)*A(I,1)/AN1+A(I,2)*A(I,2)/AN2)
   120 CONTINUE
   DO 109 I=2,K
      L=I-1
   109    DO 110 J=1,L
   110   C(I,J)=C(J,I)
   CONTINUE
   J1=K+1
   DO 112 I=1,K
      DO 112 J=1,K
         IF (I=J) 113,112,113
   113   F=(-C(J,J)/C(I,I))
DO 114 L=I,J1
114 C(J,L)=C(J,L)+(F*C(I,L))
112 CONTINUE
DO 116 I=1,K
116 C(I,1)=C(I,J1)/C(I,I)
RZERO=0.0
R1=0.0
R2=0.0
R2=0.0
DO 111 I=1,K
RZERO=RZERO+C(I,21)*(A(1,I)+A(2,I))/(AN1+AN2)
R1=R1+C(I,21)*(A(1,I)/AN1)
R2=R2+C(I,21)*(A(2,I)/AN2)
111 D2=D2+C(I,21)*((A(1,I)/AN1-A(2,I)/AN2)
AK=K
AK1=AN1+AN2-AK-1.0
F=(AN1*AN2*AK1/(AK*(AN1+AN2)*AN3))*D2
NAK1=AK1
PUNCH 1002,F,K,NAK1
PUNCH 1004,D2
PUNCH 1006,R1,RZERO,R2
PUNCH 1003
PUNCH 1008
DO 115 I=1,K
B=(C(I,21)*((A(1,I)/AN1-A(2,I)/AN2)/D2)*100.0
115 PUNCH 1005,I,C(I,21),B
IF (SENSE SWITCH 1) 92,91
91 READ 1001,X(1),X(2),X(3),X(4),X(5),X(6),X(7),X(8),X(9),X(10)
IF (9.00E+08-X(1)) 93,92,93
93 GO TO (80,81),K1
81 READ 1001,X(1),X(2),X(3),X(4),X(5),X(6),X(7),X(8),X(9),X(10),
1X(20)
80 D=0.0
DO 216 I=1,K
216 D=D+C(I,21)*X(I)
PUNCH 1007,D
GO TO 91
92 PAUSE
GO TO 90
1000 FORMAT (12)
1001 FORMAT (10F7.0)
1002 FORMAT (3HF =,F12.4,5H WITH,13,4H AND,14,19H DEGREES OF FREEDOM)
1003 FORMAT (/)
1004 FORMAT (16HMAHALANOBIS D2 =,F12.5)
1005 FORMAT (5X,I4,2F20.4)
1006 FORMAT (/4HR1 =,F15.4,7HNRZERO =,F15.4,4HR2 =,F15.4)
1007 FORMAT (F20.5)
1008 FORMAT (4X,8Hvariable,10X,8Hconstant,9X,11Hprct. added)
END TRACE
END
CLEAR, SHALLOW WATER (U)  
ABUNDANT ALGAE (V)  

-261.  7.56  8.82  1.3  
 110.  4.44  2.31  9.4  
  83.  4.3   2.51  5.6  
 -45.  4.28  2.14  7.9  
-214.  6.56  2.41 10.  
   0.  7.08  1.13  1.57 
-158.  5.53  2.38  1.01  
-107.  5.86  1.93  1.13  
-264.  7.22  1.9   1.2  
  43.  6.29  1.91  1.21  
 74.  5.65  1.78  1.41  
 34.  8.36  8.88  1.23  
-200.  4.86  1.93  1.55  
-158.  5.19  1.72  1.67  

TEST DATA - KRUMBEIN AND GRAYBILL  
RESULTS OF ANALYSIS  

CLEAR, SHALLOW WATER (U)  
ABUNDANT ALGAE (V)  

-9.03485  
-5.45156  
-4.69602  
-5.72241  
-6.3819  
-7.64124  
-7.66771  
-7.58854  
-9.28619  
-7.19183  
-6.34887  
-6.36624  
-7.72307  
-8.62563  
-8.74144  

-7.53561  
-8.25118  
-9.91152  
-9.00858  
-8.43920  
-8.76173  
-12.48141  
-12.76228  
-13.36184  
-13.23099  
-11.17719  
-11.89867  
-8.84938  
-12.33243  
-15.02300  

F = 6.1053 WITH 4 AND 25 DEGREES OF FREEDOM  
MAHALANOBIS D2 = 3.64694  

R1 = -7.2213  
RZERO = -9.0448  
R2 = -10.8683  

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>CONSTANT</th>
<th>PRCT. ADDED</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.053</td>
<td>15.5656</td>
</tr>
<tr>
<td>2</td>
<td>-0.5131</td>
<td>1.1913</td>
</tr>
<tr>
<td>3</td>
<td>-0.6110</td>
<td>14.2642</td>
</tr>
<tr>
<td>4</td>
<td>-2.5022</td>
<td>68.9788</td>
</tr>
</tbody>
</table>
REFERENCES

Chayes, Felix, 1964, A petrographic distinction between Cenozoic volcanics in and around the open ocean: Jour. Geophysical Res., v. 69, p. 1573-1588.


PROGRAM ABSTRACT

Title (If subroutine state in title):

FORTRAN II program for multivariate discriminant analysis using an IBM 1620 computer

Computer: IBM 1620 Date: February 19, 1966

Programming language: FORTRAN II

Author, organization: J. C. Davis and R. J. Sampson, Department of Geology, Idaho State University,
Pocatello, Idaho.

Direct inquiries to: Authors or

Name: D. F. Merriam Address: Kansas Geological Survey

The University of Kansas, Lawrence

Purpose/description: Computes the multivariate discriminant function and Mahalanobis' generalized distance for two groups, analyzes the significance of the difference between the groups, and classifies individual samples.

Mathematical method: Uses pooled estimates of variance and covariance in a series of simultaneous equations, which are solved by the Gauss-Jordan method of approximations.

Restrictions, range: 20 variables may be considered simultaneously. The two sample groups may contain any number of samples.

Storage requirements: Program requires approximately 20K bits storage

Equipment specifications:

Memory 20K X 40K 60K K

Automatic divide: Yes X No Indirect addressing: Yes X No

Other special features required PDQ FORTRAN compiler C2 without reread version 1 modification 0

Additional remarks (include at author's discretion: fixed/float, relocatability; optional: running time, approximate number of times run successfully, programming hours)
Mathematical simulation of marine sedimentation with IBM 7090/7094 computers, by J. W. Harbaugh, 1966 .......................... $1.00
2. A generalized two-dimensional regression procedure, by J. R. Dempsey, 1966 .............................................................. $0.50
3. FORTRAN IV and MAP program for computation and plotting of trend surfaces for degrees 1 through 6, by Mont O'Leary, R. H. Lippert, and O. T. Spitz, 1966 .......................... $0.75
4. FORTRAN II program for multivariate discriminant analysis using an IBM 1620 computer, by J. C. Davis and R. J. Sampson, 1966 ...................................................... $0.50

Special Distribution Publication

3. BALGOL program for trend-surface mapping using an IBM 7090 computer, by J. W. Harbaugh, 1963 ............................................. $0.50
4. FORTRAN II program for coefficient of association (Match-Coeff) using an IBM 1620 computer, by R. L. Kaesler, F. W. Preston, and D. I. Good, 1963 .................................................. $0.25
9. BALGOL programs for calculation of distance coefficients and correlation coefficients using an IBM 7090 computer, by J. W. Harbaugh, 1964 ...................................................... $0.75
11. Trend-surface analysis of regional and residual components of geologic structure in Kansas, by D. F. Merriam and J. W. Harbaugh, 1964 ............................................................... $0.75
12. FORTRAN and FAP program for calculating and plotting time-trend curves using an IBM 7090 or 7094/1401 computer system, by W. T. Fox, 1964 .......................................................... $0.75
13. FORTRAN program for factor and vector analysis of geologic data using an IBM 7090 or 7094/1401 computer system, by Vincent Manso and John Imbrie, 1964 ............................................. $0.75
14. FORTRAN II trend-surface program for the IBM 1620, by D. I. Good, 1964 ................................................................. $1.00
15. Application of factor analysis to petrologic variations of Americas Limestone (Lower Permian), Kansas and Oklahoma, by J. W. Harbaugh and Ferruh Demirman, 1964 ............................................................... $1.00
23. ALGOL program for cross-association of nonnumeric sequences using a medium-size computer, by M. J. Sackin, P. H. A. Sneath, and D. F. Merriam, 1965 ........................................... $0.75
24. BALGOL program and geologic application for single and double Fourier series using IBM 7090/7094 computers, by F. W. Preston and J. W. Harbaugh, 1965 ........................................... $1.00
25. Final report of the Kansas Geological Society Basement Rock Committee and list of Kansas wells drilled into Precambrian rocks, by V. B. Cole, D. F. Merriam, and W. W. Hambleton, 1965 .......................... $0.75
26. FORTRAN II trend-surface program with unrestricted input for the IBM 1620 computer, by R. J. Sampson and J. C. Davis, 1966 ................................................................. $0.50
27. Application of factor analysis to a facies study of the Lovenworth Limestone (Pennsylvanian-Virginian) of Kansas and environs, by D. F. Toomey, 1966 ........................................... $0.75
28. FORTRAN II program for standard-size analysis of unconsolidated sediments, by J. W. Pierce and D. I. Good, 1966 .......................... $0.75

Reprints (available for limited time)

Mathematical conversion of section, township, and range notation to Cartesian Coordinates, by D. I. Good (Kansas Geological Survey Bulletin 170, pt. 3, 1964) ...................................................... $0.50

A computer method for four-variable trend analysis illustrated by a study of oil-gravity variations in southeastern Kansas, by J. W. Harbaugh (Kansas Geological Survey Bulletin 171, 1964) ............................................................... $1.00


Geology and the computer, by D. F. Merriam (reprinted from New Scientist, v. 26, no. 444, 1965) ............................................................... no charge


Generation of orthogonal polynomials for trend surfacing with a digital computer, by O. T. Spitz (reprinted from Computers and operations research in mineral industries, Pennsylvania State University, 6th Annual Symposium, 1966) ............................................................... no charge

The use of statistical communication theory for characterization of porous media, by F. W. Preston, D. W. Green, and W. D. Aldenderfer (reprinted from Computers and operations research in mineral industries, Pennsylvania State University, 6th Annual Symposium, 1966) ............................................................... no charge