A STATISTICAL TECHNIQUE
FOR COMPUTER IDENTIFICATION
OF OUTLIERS IN MULTIVARIATE DATA

by

Ram Swaroop
Computing and Software, Inc.
Field Team at Flight Research Center

and

William R. Winter
Flight Research Center
Edwards, Calif. 93523
16. Abstract

A statistical technique and the necessary computer program for editing multivariate data are presented. The technique is particularly useful when large quantities of data are collected and the editing must be performed by automatic means. One task in the editing process is the identification of outliers, or observations which deviate markedly from the rest of the sample. A statistical technique, and the related computer program, for identifying the outliers in univariate data was presented in NASA TN D-6275. The current report is a multivariate analog which considers the statistical linear relationship between the variables in identifying the outliers. The program requires as inputs the number of variables, the data set, and the level of significance at which outliers are to be identified. It is assumed that the data are from a multivariate normal population and the sample size is at least two greater than the number of variables.

Although the technique has been used primarily in editing biodata, the method is applicable to any multivariate data encountered in engineering and the physical sciences.

An example is presented to illustrate the technique.
A STATISTICAL TECHNIQUE FOR COMPUTER IDENTIFICATION OF
OUTLIERS IN MULTIVARIATE DATA

Ram Swaroop
Computing and Software, Inc.
Field Team at Flight Research Center

and

William R. Winter
Flight Research Center

INTRODUCTION

The NASA Flight Research Center is engaged in an extensive biomedical research and development program. Objectives of this program include advancing the state of the art in the medical monitoring of humans in flight (ref. 1); predicting and extending the limit of man's operational capacity in the flight environment; and developing improved protection, restraint, and life support systems. As a result of this program, large quantities of biomedical information are collected in flight, necessitating dependence on the Flight Research Center's capacity for collecting, reducing, and analyzing these data by automatic means.

Experience has shown that no matter how sophisticated the monitoring, collection, and reduction systems, some editing of the biodata is required before they can be analyzed statistically. The reduced biodata may contain observations that deviate markedly from the rest of the sample. Such observations may be due to errors other than the usual random fluctuations characterizing the population to which the data belong, or may merely occur too infrequently to be considered in a particular analysis. If, upon examination, an observation falls outside a standardized region, it is usually identified as an outlier. Outliers often provide useful information. Their identification is important not only for improving the analysis but also for indicating anomalies which may require further investigation.

A statistical technique, and the related computer program, for identifying the outliers in univariate data was presented in reference 2. A method for identifying outliers in multivariate data is derived and demonstrated in this report. This method was chosen because of its simplicity and applicability in editing biodata. A program for automatic editing was written in FORTRAN IV. Inputs to this program are the number of variables, the data set, and the selected level of significance. An example is presented to illustrate the use of the method, and a scatter plot of the data is shown. The program source listing, user instructions, and a sample output are also presented.

The program computes and prints the means and standard deviations of all the variables before and after the outliers are identified and deleted. A list of the data with outliers identified by asterisks is also printed.
The authors would like to acknowledge the assistance of M. C. Nesel in writing the computer program.

**SYMBOLS**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>nonsingular matrix</td>
</tr>
<tr>
<td>$F_{\alpha, p, n-p-1}$</td>
<td>$\alpha$-level value of $F$-distribution with $p$ and $(n - p - 1)$ degrees of freedom</td>
</tr>
<tr>
<td>$G$</td>
<td>normal component of acceleration as experienced by the subject, $g$</td>
</tr>
<tr>
<td>$H/R$</td>
<td>heart rate, beats per minute</td>
</tr>
<tr>
<td>$I$</td>
<td>identity matrix</td>
</tr>
<tr>
<td>$k$</td>
<td>summation starting from $i$ through $k$, where $i$ and $k$ are integers between 1 and $n$, and $i$ is less than $k$</td>
</tr>
<tr>
<td>$N_p(\mu, \Sigma)$</td>
<td>$p$-variate normal distribution with mean, $\mu$, and covariance matrix, $\Sigma$</td>
</tr>
<tr>
<td>$n$</td>
<td>sample size</td>
</tr>
<tr>
<td>$p$</td>
<td>number of variables</td>
</tr>
<tr>
<td>$S$</td>
<td>$(p \times p)$ matrix of sums and cross-products of deviations of observations from $X$ divided by $(n - 1)$</td>
</tr>
<tr>
<td>S.D.</td>
<td>standard deviation</td>
</tr>
<tr>
<td>$T^2$</td>
<td>Hotelling's $T^2$ statistic</td>
</tr>
<tr>
<td>$u, v$</td>
<td>column vectors of $p$ dimensions</td>
</tr>
<tr>
<td>$X_i, X_j$</td>
<td>$i$th or $j$th observation vector of $p$ dimensions, where $i$ or $j$ ranges from 1 to $n$</td>
</tr>
<tr>
<td>$\bar{X}$</td>
<td>mean vector computed from $n$ observation vectors</td>
</tr>
<tr>
<td>$Z_i$</td>
<td>$i$th vector obtained by orthogonal transformation of vector $X_i$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>level of significance</td>
</tr>
<tr>
<td>$\Delta_i$</td>
<td>positive real number corresponding to observation vector $X_i$, where $i$ ranges from 1 to $n$</td>
</tr>
<tr>
<td>$\Delta_*$</td>
<td>positive real number computed from $F_{\alpha, p, n-p-1}$ for the data set, to compare with $\Delta_i$</td>
</tr>
</tbody>
</table>
\( r^2 \) random variable related to \( T^2 \)

Superscript:

\( T \) transpose

**BRIEF DESCRIPTION OF TECHNIQUE**

Outliers are identified by computing, at the given level of significance, the critical value, \( \Delta_* \), for the data set and \( \Delta_i \) for each observation vector, \( X_i \). If \( \Delta_i \) is larger than \( \Delta_* \), observation \( X_i \) is identified as an outlier. The quantity \( \Delta_* \) is a function of total sample size, \( n \), number of variables, \( p \), and the \( F \)-value for the given level of significance, whereas each \( \Delta_i \) is a function of the observation \( X_i \) and the estimated mean and covariance matrix from all the observations. It is assumed that all the observations constitute a random sample from a \( p \)-variate normal distribution.

**DERIVATION OF TECHNIQUE**

Let \( X_1, X_2, \ldots, X_n \) be a random sample of size \( n \) from a \( p \)-dimensional normal distribution, \( N_p(\mu, \Gamma) \). The observations will be considered as \( n \) greater than \( p + 2 \) column vectors in a \( p \)-dimensional vector space. Consider any \((n \times n)\) orthogonal matrix, with first two rows as shown,

\[
\begin{bmatrix}
\frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\
\frac{1}{\sqrt{n}} & -\frac{1}{\sqrt{n(n-1)}} & \cdots & -\frac{1}{\sqrt{n(n-1)}} \\
\frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & \frac{1}{\sqrt{n(n-1)}} \\
\frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & \frac{1}{\sqrt{n(n-1)}} \\
\end{bmatrix}
\]

representing the rotation of \( n \)-dimensional space so that the observations \( X_1, X_2, \ldots, X_n \) are transformed into vectors \( Z_1, Z_2, \ldots, Z_n \), where

\[
Z_1 = \sum_{i=1}^{n} \frac{1}{\sqrt{n}} X_i = \sqrt{n} \bar{X}
\]

\[
Z_2 = \sqrt{\frac{n-1}{n}} X_1 - \sum_{i=2}^{n} \frac{1}{\sqrt{n(n-1)}} X_i = \sqrt{\frac{n}{n-1}} (X_1 - \bar{X})
\]
It may be noted that $Z_1$ is distributed as $N_p(\sqrt{n}\mu, \Gamma)$, $Z_2$, $Z_3$, ..., $Z_n$ are all distributed as $N_p(0, \Gamma)$, and all are stochastically independent of one another (ref. 3, pp. 50-52). Let $S$ denote the estimate of the covariance matrix $\Gamma$. Then the following relation holds:

$$(n - 1)S = \sum_{1}^{n} (X_1 - \bar{X})(X_1 - \bar{X})^T = \sum_{2}^{n} Z_iZ_i^T$$

Define a $(p \times p)$ matrix, $S_0$, such that

$$(n - 2)S_0 = \sum_{3}^{n} Z_iZ_i^T$$

Independence of $Z_2$ and set $(Z_3, \ldots, Z_n)$ implies that $Z_2$ is independent of $S_0$ and the

$$T^2 = Z_2^TS_0^{-1}Z_2$$  \hspace{1cm} (1)

statistic is distributed as Hotelling's $T^2$. From the relationship between $T^2$ and $F$ (ref. 3, pp. 106-107), it follows that

$$T^2 = Z_2^TS_0^{-1}Z_2$$

is distributed as

$$\frac{p(n - 2)}{(n - p - 1)}F_{p, n - p - 1}$$

Because $Z_2$ is not independent of $S$, the preceding distribution does not hold for

$$\tau^2 = Z_2^TS^{-1}Z_2$$  \hspace{1cm} (2)

and the distribution of $\tau^2$ must be derived.

By the preceding definitions

$$(n - 1)S = Z_2Z_2^T + \sum_{3}^{n} Z_iZ_i^T = Z_2Z_2^T + (n - 2)S_0$$
or
\[(n - 2)S_0 = (n - 1)S - Z_2Z_2^T\quad (3)\]

To express the relation between \(T^2\) and \(\tau^2\), the following lemma is used:

**Lemma:** Let \(A\) be a \((p \times p)\) nonsingular matrix and \(u, v\) be \(p\)-dimensional vectors. Then
\[
(A - uv^T)^{-1} = A^{-1} + \frac{(A^{-1}u)(v^TA^{-1})}{1 - v^TA^{-1}u}
\]

**Proof:** The proof of the lemma is presented in appendix A.

Applying the result (eq. (4)) of the lemma to equation (3),
\[
S_o^{-1} = \frac{n - 2}{n - 1}\left[ S^{-1} + \frac{S^{-1}Z_2Z_2^TS^{-1}}{(n - 1) - Z_2^TS^{-1}Z_2} \right]
\]

Substituting this expression for \(S_o^{-1}\) in equation (1) and applying equation (2),
\[
T^2 = Z_2^TS_o^{-1}Z_2 = \frac{n - 2}{n - 1}Z_2^T\left[ S^{-1} + \frac{S^{-1}Z_2Z_2^TS^{-1}}{(n - 1) - Z_2^TS^{-1}Z_2} \right]Z_2
\]
\[
= \frac{n - 2}{n - 1}\left[ \tau^2 + \frac{\tau^4}{(n - 1) - \tau^2} \right]
\]
\[
= \frac{(n - 2)\tau^2}{(n - 1) - \tau^2}
\]

This relation provides the distribution of \(\tau^2\), and appropriate probability statements can be made.
Define  
\[ \Delta_i = (X_i - \bar{X})^T S^{-1} (X_i - \bar{X}) \text{ for } i = 1, 2, \ldots, n \]

With no loss in generality, \( \Delta_1 \) is used. From equation (2)

\[ \tau^2 = Z_2^T S^{-1} Z_2 = \frac{n}{n - 1} \Delta_1 \]

and from equation (5)

\[ T^2 = \frac{(n - 2) \Delta_1}{(n - 1) - \frac{n}{n - 1} \Delta_1} = \frac{n(n - 2) \Delta_1}{(n - 1)^2 - n \Delta_1} \]

is distributed as

\[ \frac{p(n - 2)}{(n - p - 1)} F_{p, n-p-1} \]

From the distribution of \( T^2 \), the statement

\[ \text{Probability} \left[ \frac{n(n - 2) \Delta_1}{(n - 1)^2 - n \Delta_1} \geq \frac{p(n - 2)}{(n - p - 1)} F_{p, n-p-1} \right] = \alpha \]

provides criteria for identifying the \( Z_2 \) (or \( X_1 \)) as an outlier at the assigned level of significance, \( \alpha \). This statement is equivalent to

\[ \text{Probability} \left[ \Delta_1 \geq \frac{p(n - 1)^2 F_{p, n-p-1}}{n(n - p - 1) + np F_{p, n-p-1}} \right] = \alpha \]

For significance level \( \alpha \), denote

\[ \Delta_* = \frac{p(n - 1)^2 F_{p, n-p-1}}{n(n - p - 1) + np F_{p, n-p-1}} \]

then \( X_1 \) will be identified as an outlier at \( \alpha \) level if

\[ \Delta_1 > \Delta_* \]
The quantity $X_1$ (or $Z_2$) was chosen for convenience of the preceding derivation and the derivation holds for all $X_i$. Thus $X_i$ will be identified as an outlier at $\alpha$ level of significance if $\Delta_i > \Delta_*$. 

**PROGRAM APPLICATION**

Given a sample of data vectors $X_1$, $X_2$, ..., $X_n$, the mean vector

$$\bar{X} = \frac{1}{n} \sum_{1}^{n} X_i$$

and the estimate of the covariance matrix

$$S = \frac{1}{n - 1} \sum_{1}^{n} (X_i - \bar{X})(X_i - \bar{X})^T$$

is computed. Then, for assigned $\alpha$, the critical value

$$\Delta_* = \frac{p(n - 1)^2 F_{\alpha p, n-p-1}}{n(n - p - 1) + np F_{\alpha p, n-p-1}}$$

for the data set is computed. Corresponding to each observation vector, $X_i$,

$$\Delta_i = (X_i - \bar{X})^T S^{-1} (X_i - \bar{X})$$

is computed. If $\Delta_i > \Delta_*$, observation vector $X_i$ is identified as an outlier at level $\alpha$.

The program (appendix B) follows this technique. The output (appendix C) of the program contains the data set, $\Delta_i$, $\Delta_*$, outliers marked by asterisks (*), the number of outliers identified, and the level of significance. The output also shows the means and standard deviations of the variables before and after the deletion of outliers. The required input parameters are: (1) format of the data to be read, (2) number of variables, (3) significance level, $\alpha$, and (4) the data set, formatted as specified. Program options allow the user to select either a 5 percent or a 1 percent level of significance and to print the names of the variables, if desired. This program is designed so that it can be used as a subroutine in other than biodata applications, in engineering and the physical sciences, for example.

The program is particularly useful when large quantities of data are collected and the editing must be performed by automatic means.
Heart rate, H/R, and normal component of acceleration, G, data from a 66-minute flight by a student pilot at the Aerospace Research Pilot School, Edwards Air Force Base, Calif., are used to demonstrate the described technique of computer editing of biodata. These data were chosen because centrifuge studies (ref. 4) have shown that H/R and G are linearly related. The program was used to identify the outliers at a 1 percent level of significance considering H/R and G separately as univariate data and together as bivariate data. The computer output for these cases is shown in appendix C.

The results of the two univariate analyses and the one bivariate analysis of the same data are presented in figure 1. The point labeled H is identified as an outlier on the basis of H/R analysis alone; the point labeled G is identified as an outlier on the basis of G alone; and points labeled B are identified as outliers on the basis of bivariate analysis of H/R and G.

![Figure 1. Minute heart rate and acceleration data for a 66-minute flight of a student pilot from the Aerospace Research Pilot School showing outliers identified by the automatic multivariate outlier technique.](image)

Bivariate analysis is based on the fact that high H/R is associated with high G, whereas univariate analysis cannot take this information into account. For this reason the point labeled H was not identified by the bivariate analysis, but was identified as an outlier on the basis of univariate H/R analysis. Also, both points labeled B appear not to follow a statistical linear relationship and are identified by the bivariate
analysis; however, only one of these points was identified by one of the univariate analyses (G alone). This example thus focuses on the fact that the multivariate technique, which utilizes the statistical linear relationships between the variables, is preferable in identifying outliers in multivariate data.

CONCLUDING REMARKS

A statistical technique to identify outliers, or observations which deviate markedly from the rest of the sample, in multivariate data at a given level of significance was derived. The use of the technique was illustrated by a biodata example. The example also demonstrated that the results obtained when each variable was considered separately could be different from the results obtained when the variables were considered jointly. The latter technique takes into account the statistical linear relationship between the variables and is the preferred method.

Although this method of detecting and identifying outliers is being used for biodata editing at the NASA Flight Research Center, it is also applicable to multivariate data encountered in other disciplines, such as engineering and the physical sciences. This technique is particularly useful when large quantities of data are collected and the editing must be performed by automatic means.

The program can be used as a subroutine in multivariate analyses.

Flight Research Center,
National Aeronautics and Space Administration,
APPENDIX A

PROOF OF THE LEMMA

**Lemma:** If \( A \) is a \((p \times p)\) nonsingular matrix, and \( u, v \) are \( p \)-dimensional vectors, then

\[
(A - uv^T)^{-1} = A^{-1} + \frac{(A^{-1}u)(v^TA^{-1})}{1 - v^TA^{-1}u}
\]

**Proof:** The result is obtained by showing that

\[
(A - uv^T) \left[ A^{-1} + \frac{(A^{-1}u)(v^TA^{-1})}{1 - v^TA^{-1}u} \right] = I
\]

Simplification of the left-hand expression gives

\[
AA^{-1} + \frac{AA^{-1}u(v^TA^{-1})}{1 - v^TA^{-1}u} = uv^TA^{-1} - \frac{uv^TA^{-1}uv^TA^{-1}}{1 - v^TA^{-1}u}
\]

or

\[
I + \frac{1}{1 - v^TA^{-1}u} \left[ uv^TA^{-1} - uv^TA^{-1} + (v^TA^{-1}u)(uv^TA^{-1}) - uv^TA^{-1}uv^TA^{-1} \right]
\]

Because \( v^TA^{-1}u \) is a scalar, the expression becomes

\[
I + \frac{1}{1 - v^TA^{-1}u} \left[ (v^TA^{-1}u)(uv^TA^{-1}) - (v^TA^{-1}u)(uv^TA^{-1}) \right]
\]

or

\[
I
\]

which is the same as the right-hand side.
APPENDIX B

PROGRAM SOURCE LISTING

MTVOUT

PURPOSE
IDENTIFY OUTLIERS IN MULTIVARIATE DATA

METHOD
LET VECTORS X(1) THRU X(N) BE OBSERVATIONS FROM A NP-VARIATE
NORMAL DISTRIBUTION

THEN
XBAR AND COVARIANCE MATRIX S ARE OBTAINED BY EQUATIONS
XBAR = (1/N) * SUM X(J)
S * (N-1) = SUM (X(J)-XBAR) * (X(J)-XBAR) TRANSPOSE

WHERE
N = SAMPLE SIZE (N.LE.500)
NP = NUMBER OF VARIABLES (NP.LE.10)
ALPH = SIGNIFICANCE LEVEL
FALPH = F-VALUE FOR ALPHA AT NP AND N-NP-1
DEGREES OF FREEDOM
X(J,1) = THE J'TH ELEMENT OF THE J'TH VECTOR,
WHERE J=1,2,...,NP AND J=1,2,...,N

CALCULATE
DELS = ((N-1)**2*NP*FALPH) / (N*(N-NP-1)+NP*FALPH)

CALCULATE FOR EACH OBSERVATION VECTOR X(J)
RR = (X(J)-XBAR) TRANSPOSE * (S) INVERSE * (X(J)-XBAR)

IF RR > DELS, THEN X(J) IS IDENTIFIED AS AN OUTLIER

REFERENCE
1. AN INTRODUCTION TO MULTIVARIATE STATISTICAL ANALYSIS,
ANDERSON, 1965
2. A SIMPLE TECHNIQUE FOR AUTOMATIC COMPUTER EDITING
OF BIODATA, NASA TN D-5275
3. SYS/360 SCIENTIFIC SUBROUTINE PACKAGE (360A-CM-03X)
PROGRAMMERS MANUAL, IBM INC., 1968

SUBROUTINES

F TABLE
M PRO
LOC
DS I NV
DMF SD
APPENDIX B

INPUT

CARD 1  FORMAT OF X-ARRAY CARDS TO BE READ IN  (20A4)
CARD 2
  COL 1-2  NP  (NUMBER OF VARIABLES)  NP.LE.10  (12)
  COL 3-6  BLANK
  COL 7-9  ALPH  (SIGNIFICANCE LEVEL) .05 OR .01 (F3.2)
  COL 10-12  BLANK
  COL 13-15  VARIABLE NAME CARD INDICATOR
              1 - NAME CARD FOLLOWS
              BLANK - NO NAME CARD

CARD 3  (OPTIONAL)
  TEN FIELDS OF EIGHT CHARACTERS EACH (10A4), WHICH MAY
  BE USED TO ASSIGN MEANINGFUL NAMES TO THE NP VARIABLES.
  IF COL 10 OF THE PREVIOUS CARD IS PITCHED, NAMES MUST
  BE ASSIGNED FOR ALL NP VARIABLES.  DEFAULT NAMES ARE
  'X1', 'X2', ..., 'X(NP)'.

CARDS 4-  DATA FORMATTED AS PRESCRIBED IN CARD 1

MULTIPLE RUNS ARE PERMITTED, AS LONG AS EACH DATA DECK IS
PRECEDED BY APPROPRIATE CONTROL CARDS (CARDS 1, 2 AND
3 ABOVE), AND IS FOLLOWED BY A CARD WITH **** PUNCHED
IN COLUMNS 1 THRU 4.

OUTPUT

1  LIST OF VECTORS WITH OUTLIERS IDENTIFIED BY ASTERISKS
2  MEAN (ORIGINAL DATA)
3  STANDARD DEVIATION (ORIGINAL DATA)
4  MEAN (OUTLIERS DELETED)
5  STANDARD DEVIATION (OUTLIERS DELETED)

REAL*8  SUMX(10), XBAR(10), SDI(10), XBARI(10), SDI(10)
REAL*8  A(55), B(10), S(10,10)
REAL*8  RR, DELSTR, R(1C), FMT(10), BLANK/
       '  ', VAR(10)
REAL*8  DEF(10)/  X1  '  ',  X2  '  ',  X3  '  ',  X4  '  ',  X5
               1  '  ',  X6  '  ',  X7  '  ',  X8  '  ',  X9  '  ',  X10  '  /
REAL  X(501,10), Y(501,10)
INTEGER  STAR/****/, BLANK/' ', FLAG, FLGTOT, DSWI
INTEGER  PREFERENCES(2)/'(A4,','T1,','GPT(2)'/YES',' NC'/
INTEGER*2  IFLAG(500)
LOGICAL*1  FORMAT(89), RPA(1)'/'
EQUIVALENCE (PREFERENCES,FMT(1)),(FMT,FFORMAT(9)),(RPA,RFORMAT(89))
1  DO 2  I=1,10
   2  VAR(I) = DEF(I)
2  DO 5  I=1,500
   5  FMT(I) = BLNK
   5  DO 5  I=1,500
   5  IFLAG(I) = 0
   5  FLGTOT = 0
   5  LINES = 1
   5  NSWI = 0
   5  DSWI = 0
   5  OH = 0.01
APPENDIX B

ID = 1
IZ = 0
MM = 1
N = 1
READ(1,1000,EVD=999) FMT
READ(1,1010) NP, ALPH, ISW1
C
C TEST IF NAMES HAVE BEEN ASSIGNED TO THE VARIABLES
IF(ISW1.GT.0) GO TO 10
MM = 2
GO TO 15
10 READ(1,1050) (VAR(I), I=1,NP)
C
C READ IN THE DATA AS X-ARRAY
15 READ(1,FORMAT,END=995) IXX, (X(N,I), I=1,NP)
16 IF(IXX.EQ.STAR) GO TO 19
N = N + 1
IF(N.GT.501) GO TO 990
GO TO 15
19 N = N - 1
C
C SELECT APPROPRIATE F-VALUE
CALL FTABLE (V,NP,ALPH,FALPH)
C
C WRITE THE INPUT CONTROL INFORMATION
WRITE(3,5002) NP, ALPH, FALPH, OPT(MM)
DELSR = ((N-1)**2 + NP * FALPH) / (N * ((N-NP-1) + NP * FALPH))
20 DO 86 I=1,10
SUMX(I)=0.0
21 DO 86 J=1,10
86 S(I,J)=0.0
DO 30 J=1,N
C
C TEST FOR FLAGGED VECTORS IDENTIFIED AS OUTLIERS
IF(IFLAG(J).GT.0) GO TO 30
DO 25 I=1,NP
25 SUMX(I) = SUMX(I) + X(J,I)
30 CONTINUE
C
C FIND MEAN OF EACH VARIABLE
DO 40 I=1,NP
XBAR(I) = SUMX(I) / (N-FLGTOT)
40 DO J=1,N
Y(J,I) = X(J,I) - XBAR(I)
41 IF(IFLAG(J).GT.0) Y(J,I) = 0.D0
JJ = 0
42 DO 70 I=1,NP
43 DO 70 K=1,I
44 DO 60 J=1,N
60 S(I,K) = Y(J,I) * Y(J,K) + S(I,K)
C
C FIND STANDARD DEVIATION OF EACH VARIABLE
SD(I) = DSQRT(S(I,1) / (N-1-FLGTOT))
JJ = JJ + 1
70 A(JJ) = S(I,K)
C
C IF COMPUTATIONS ARE COMPLETE, BRANCH TO
C PRINT TABLE OF MEANS AND S.D.'S
IF(DSW1.GT.0) GO TO 200
CALL DSINV(A,NP,OH,IER)
APPENDIX B

IF (IER) 991, 80, 991
C
WRITE THE LIST HEADING
30 WRITE (3,1015) (VAR(I), I=1, NP)
C
100 J=1, N
C
90 K=1, NP
B(K) = Y(J,K)
CALL MPRD (B,A,R,IO,NP,IZ,IC,NP)
CALL MPRD (R,B,RR,IO,NP,IZ,IZ,IO)
RR = RR*(N-1)
FLAG = BLANK
C
FLAG THIS VECTOR WITH AN ASTRISK IF
C
IT IS IDENTIFIED AS AN OUTLIER
IF (RR.GT.DELESTR) FLAG = STAR
IF (FLAG.NE.STAR) GO TO 95
IF (FLAG(J) = 1)
FLGTOT = FLGTOT + 1
C
WRITE THE DATA VECTOR AND IF IDENTIFIED AS
C
AN OUTLIER, LABEL WITH AN ASTRISK
95 WRITE (3,1020) J, RR, FLAG, (X(J,K), K=1, NP)
C
LINES = LINES + 1
IF (LINES.LE.55) GO TO 100
WRITE (3,1015) (VAR(I), I=1, NP)
C
LINES = 1
100 CONTINUE
WRITE (3,1025) ALPH
WRITE (3,1030) N
WRITE (3,1035) FLGTOT
WRITE (3,1040) DELSTR
C
SAVE MEANS AND S.D.'S, THEN LCCP BACK AND
C
COMPUTE NEW MEANS AND S.D.'S AFTER DELETING OUTLIERS
110 DO 110 I=1, NP
XBAR(I) = XBAR(I)
SD(I) = SD(I)
DSWI = 1
GO TO 20
C
WRITE TABLE OF MEANS AND S.D.'S BEFORE AND
C
AFTER DELETION OF OUTLIERS
200 WRITE (3,2000)
WRITE (3,2005)
WRITE (3,2010)
DO 210 I=1, NP
210 WRITE (3,2015) VAR(I), XBAR(I), SD(I), XBAR(I), SD(I)
IF (NSWI.EQ.1) GO TO 999
GO TO 1
C
990 WRITE (3,5000)
GO TO 999
C
991 WRITE (3,5001)
GO TO 999
C
995 NSWI = 1
GO TO 19
C
999 WRITE (3,5009)
APPENDIX B

1000 FORMAT (20A4)
1010 FORMAT (12,1X,F3.2,3X,I1)
1015 FORMAT (1H1,'LIST OF VECTORS WITH OUTLIERS IDENTIFIED BY ASTERISKS'
1/1H0,' J',7X,'DELTA',7X,'C(2X,A8)'/)
1020 FORMAT (1H1,3F12.4,2X,A1,6X,1CF10.2)
1025 FORMAT (1H1,'///1H0,12X,* OUTLIER IDENTIFIED AT ',F4.2,' SIGNIFICANCE'
1/1LEVEL')
1030 FORMAT (1H1,12X,*SAMPLE SIZE IS ',I3)
1035 FORMAT (1H1,12X,*NO OF OUTLIERS IS ',I3)
1040 FORMAT (1H1,12X,*DELSTAR = ',F10.4)
1050 FORMAT (10A8)
2000 FORMAT (1H1,20X,'MEAN AND STANDARD DEVIATION OF THE VARIABLES')
2005 FORMAT (1H0,14X,'DATA BEFORE IDENTIFICATION',9X,'DATA AFTER DELETION'/
1H1,22X,'OF OUTLIERS',21X,'OF OUTLIERS')
2010 FORMAT (1H0,'VARIABLES',8X,'MEAN',13X,'S.D.',11X,'MEAN',14X,'S.D.')
2015 FORMAT (1H0,A8,F15.4,F17.4,F15.4,F18.4)
5000 FORMAT (1H1,'SAMPLE SIZE EXCEEDS 500 - PROGRAM TERMINATED')
5001 FORMAT (1H1,'ERROR IN THE MATRIX INVERSION PROCESS - PROGRAM TERMINATED')
5002 FORMAT (1H1, '**INPUT CONTROL INFORMATION**//1H0,'**NUMBER OF VARIABLES IS',14/1H1,'**SIGNIFICANCE LEVEL IS',F5.2/1H1,'**F-VALUE IS',
2F16.4/1H1,'**VARIABLE NAME CARD:',A4)
5009 FORMAT (1H1,'END OF JOB')
9999 STOP
END
### APPENDIX B

**SUBROUTINE FTABLE (N,M,SL,F)**

```
*************** FTAB0000  FTAB0010  FTAB0020  FTAB0030  FTAB0040  FTAB0050  FTAB0060  FTAB0070  FTAB0080  FTAB0090  FTAB0100  FTAB0110  FTAB0120  FTAB0130  FTAB0140  FTAB0150  FTAB0160  FTAB0170  FTAB0180  FTAB0190  FTAB0200  FTAB0210  FTAB0220  FTAB0230  FTAB0240  FTAB0250  FTAB0260  FTAB0270  FTAB0280  FTAB0290  FTAB0300  FTAB0310  FTAB0320  FTAB0330  FTAB0340  FTAB0350  FTAB0360  FTAB0370  FTAB0380  FTAB0390  FTAB0400  FTAB0410  FTAB0420  FTAB0430  FTAB0440  FTAB0450  FTAB0460  FTAB0470  FTAB0480  FTAB0490  FTAB0500
```

**CALLING PARAMETERS**

- **N = SAMPLE SIZE**
- **M = NUMBER OF VARIABLES**
- **SL = SIGNIFICANCE LEVEL**
- **F = SELECTED F-VALUE**

**DIMENSION TABLE(71/31,40,60,120,20C,44C,1000/1, DF**

```
REAL T(21/0,5,01/)
```

**McG.**
APPENDIX B

5.238, 3.43, 4.35, 8.10, 3.45, 9.58, 3.10, 4.39, 2.39, 2.46, 2.35, 3.37

63.87, 2.51, 3.70, 2.45, 3.56, 2.39, 2.46, 2.35, 3.37

1.40, 0.257, 3.81, 2.49, 3.04, 2.42, 2.51, 2.37, 3.04, 2.32, 3.14, 3.07, 0.95

23.44, 5.72, 3.05, 4.82, 2.82, 4.31, 2.66, 3.09, 2.55, 3.76, 2.46, 3.59, 2.40

3.34, 2.34, 3.35, 2.30, 3.26, 4.28, 7.89, 3.42, 2.56, 3.30, 3.4, 2.76, 2.46

2.64, 3.94, 2.53, 3.71, 2.44, 3.54, 2.37, 3.41, 2.32, 3.30, 2.27, 3.21, 4.26

57.82, 3.40, 5.61, 3.01, 4.72, 1.24, 2.78, 4.22, 2.62, 2.39, 2.0, 2.51, 3.67, 2.42, 3.50


74.18, 2.60, 3.05, 2.49, 3.03, 2.42, 3.46, 2.34, 3.28, 2.28, 2.22, 2.4, 3.13

94.23, 7.72, 3.37, 5.52, 2.98, 4.62, 2.74, 4.14, 2.59, 3.02, 2.47, 3.59, 2.39

93.42, 2.32, 3.29, 2.27, 3.18, 2.22, 3.04, 4.21, 7.68, 3.35, 5.49, 2.96, 4.60

12.3, 4.1, 2.57, 3.78, 2.46, 3.56, 2.27, 3.39, 2.31, 2.62, 2.29, 3.15, 2.20

23.06, 4.20, 7.6, 3.34, 5.45, 2.95, 4.57, 2.71, 4.07, 5.6, 3.75, 2.45, 3.53

32.36, 3.36, 2.29, 3.23, 2.24, 3.12, 2.19, 3.04, 4.18, 7.60, 3.33, 5.42, 2.93

44.54, 2.70, 4.04, 2.55, 3.73, 2.43, 3.50, 2.35, 3.33, 2.28, 3.20, 2.22, 3.09

52.18, 3.00, 4.17, 7.56, 3.32, 5.35, 2.92, 4.51, 2.89, 4.02, 2.53, 3.70, 2.42

63.47, 2.33, 3.30, 2.27, 3.17, 2.21, 3.07, 2.16, 2.98

REAL TABL (140) / 0.48, 7.31, 3.23, 5.18, 2.84, 4.31, 2.61, 3.83, 2.45

13.51, 2.23, 3.29, 2.25, 3.12, 2.18, 2.99, 2.12, 3.12, 2.89, 3.20, 2.94, 4.30, 7.08

23.15, 4.98, 2.8, 76, 4.13, 2.53, 3.65, 2.37, 3.34, 2.25, 3.12, 2.17, 2.95, 2.10

32.82, 2.04, 2.72, 1.99, 2.63, 3.92, 2.52, 6.38, 3.69, 2.95, 2.45, 3.48

42.29, 3.17, 2.17, 2.98, 2.09, 2.79, 2.02, .97, 3.41, 2.96, 2.76, 1.91, 2.47, 3.89

56.70, 3.04, 4.71, 2.65, 3.08, 2.41, 3.41, 2.26, 3.11, 2.14, 2.90, 2.05, 2.73

61.98, 2.60, 1.52, 2.30, 1.87, 2.41, 3.66, 2.66, 7.03, 3.92, 4.66, 2.62, 2.02, 3.9

73.36, 2.23, 3.06, 2.12, 2.85, 2.07, 2.69, 1.96, 2.55, 1.90, 2.96, 1.85, 2.37

83.85, 4.64, 3.00, 4.62, 2.61, 3.38, 2.38, 3.34, 2.22, 2.34, 2.04, 2.10, 2.82, 2.02

92.65, 1.95, 2.53, 1.89, 2.43, 1.84, 2.34, 3.84, 4.63, 3.86, 4.6, 2.51, 1.88, 2.41, 1.83

22.32/

EQUIVALENCE (TABLE(1,1), TABLE(1,1)) ; (TABLE(1,1), TABLE(2,1))

EQUIVALENCE (TABLE(1,1), TABLE(2,1)) ; (TABLE(1,1), TABLE(4,1))

IF(MG,GT,10) GO TO 590

DF = M-N-1

IF(SL,GE,SL(1)) GO TO 10

IF(SL,GE,SL(T)) GO TO 19

GO TO 591

10 L = 1

GO TO 16

15 L = 2

IF(DF,LE,30) GO TO 40

DD 20 I = 2,7

IF(DF,D2D(1)) 50,3C,20

20 CONTINUE

F = TABLE(L,M,36) + (1,/,N,/,0,0) * (TABLE(L,M,36) - TABLE(L,M,37))

RETURN

30 DF = I + 29

RETURN

GO TO 590

50 F = TABLE(L,M,129) + (1,/,DF,1,/,XDF(1)) / (1,/,XDF(1)-1,/,XDF)

1(I,J) * (TABLE(L,M,128) - TABLE(L,M,129))

RETURN

590 WRITE(3,990)

RETURN

WRITE(3,991) SL

RETURN

FORMAT (1H1, * , NUM VARIABLES > 10, F HAS BEEN SET TO A DUM M)

FORMAT (1H1, * , SIGNIFICANCE LEVEL 'F6.3', IS NOT ACCEPTABLE. LEV)

END
APPENDIX B

SUBROUTINE MPRD

PURPOSE
MULTIPLY TWO MATRICES TO FORM A RESULTANT MATRIX

USAGE
CALL MPRD(A,B,R,N,M,MSA,MSB,L)

DESCRIPTION OF PARAMETERS
A - NAME OF FIRST INPUT MATRIX
B - NAME OF SECOND INPUT MATRIX
R - NAME OF OUTPUT MATRIX
N - NUMBER OF ROWS IN A AND R
M - NUMBER OF COLUMNS IN A AND ROWS IN B
MSA - ONE DIGIT NUMBER FOR STORAGE MODE OF MATRIX A
  0 - GENERAL
  1 - SYMMETRIC
  2 - DIAGONAL
MSB - SAME AS MSA EXCEPT FOR MATRIX B
L - NUMBER OF COLUMNS IN B AND R

REMARKS
MATRIX R CANNOT BE IN THE SAME LOCATION AS MATRICES A OR B
NUMBER OF COLUMNS OF MATRIX A MUST BE EQUAL TO NUMBER OF ROWS OF MATRIX B

SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
LOC

METHOD
THE M BY L MATRIX B IS PREMULTIPLIED BY THE N BY M MATRIX A
AND THE RESULT IS STORED IN THE N BY L MATRIX R. THIS IS A
ROW INTO COLUMN PRODUCT.
THE FOLLOWING TABLE SHOWS THE STORAGE MODE OF THE OUTPUT
MATRIX FOR ALL COMBINATIONS OF INPUT MATRICES

A       B       R
GENERAL  GENERAL  GENERAL
GENERAL  SYMMETRIC  GENERAL
GENERAL  DIAGONAL  GENERAL
SYMMETRIC  GENERAL  GENERAL
SYMMETRIC  SYMMETRIC  GENERAL
SYMMETRIC  DIAGONAL  GENERAL
DIAGONAL  GENERAL  GENERAL
DIAGONAL  SYMMETRIC  GENERAL
DIAGONAL  DIAGONAL  DIAGONAL

SUBROUTINE MPRD(A,B,R,N,M,MSA,MSB,L)
APPENDIX B

DOUBLE PRECISION A, B, R
DIMENSION A(I), B(I), R(I)

C
C SPECIAL CASE FOR DIAGONAL BY DIAGONAL
C
MS=MSA*10+MSB
IF(MS-22) 30,1C,30
10 DO 20 I=1,N
20 R(I)=A(I)*B(I)
RETURN
C
C ALL OTHER CASES
C
30 IR=1
DO 90 K=1,L
DO 90 J=1,N
R(IR)=0
DO 80 I=1,M
IF(MS) 40,60,4C
40 CALL LOC(J,1,IA,N,M,MSA)
CALL LOC(I,K,IB,M,L,MSB)
IF(IA) 50,8C,5C
50 IF(IB) 70,8C,7C
60 IA=N*(I-1)+J
IB=M*(K-1)+I
70 R(IR)=R(IR)+A(IA)*B(IB)
90 CONTINUE
90 IR=IR+1
RETURN
END
SUBROUTINE LOC

PURPOSE

COMPUTE A VECTOR SUBSCRIPT FOR AN ELEMENT IN A MATRIX OF
SPECIFIED STORAGE MODE

USAGE

CALL LOC %I,J,IR,N,M,MS<

DESCRIPTION OF PARAMETERS

I - ROW NUMBER OF ELEMENT
J - COLUMN NUMBER OF ELEMENT
IR - RESULTANT VECTOR SUBSCRIPT
N - NUMBER OF ROWS IN MATRIX
M - NUMBER OF COLUMNS IN MATRIX
MS - ONE DIGIT NUMBER FOR STORAGE MODE OF MATRIX
  0 - GENERAL
  1 - SYMMETRIC
  2 - DIAGONAL

REMARKS

NONE

SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED

NONE

METHOD

MS#0 SUBSCRIPT IS COMPUTED FOR A MATRIX WITH N*M ELEMENTS
IN STORAGE <GENERAL MATRIX>

MS#1 SUBSCRIPT IS COMPUTED FOR A MATRIX WITH N*(N+1)/2 ELEMENTS
IN STORAGE <UPPER TRAPEZIUM OF SYMMETRIC MATRIX>, IF ELEMENT IS IN LOWER TRAPEZIUM, SUBSCRIPT IS
CORRESPONDING ELEMENT IN UPPER TRAPEZIUM.

MS#2 SUBSCRIPT IS COMPUTED FOR A MATRIX WITH N ELEMENTS
IN STORAGE <DIAGONAL ELEMENTS OF DIAGONAL MATRIX>, IF ELEMENT IS NOT ON DIAGONAL AND THEREFORE NOT IN
STORAGE, IR IS SET TO ZERO.

SUBROUTINE LOC(I,J,IR,N,M,MS)

IX=I
JX=J
IF(MS=1) 10,20,30
10 IRX=N*(JX-1)+IX
GO TO 36
20 IF(MS-JX) 22,24,24
22 IRX=IX+(IX*JX-JX)/2
GO TO 36
24 IRX=JX+(IX*IX-IX)/2
GO TO 36
30 IRX=0
IF(MS-JX) 36,32,36
32 IRX=IX
36 IR=IRX
RETURN
END
APPENDIX B

SUBROUTINE DSINV

PURPOSE
INVERT A GIVEN SYMMETRIC POSITIVE DEFINITE MATRIX

USAGE
CALL DSINV(A,N,EPS,IER)

DESCRIPTION OF PARAMETERS
A - DOUBLE PRECISION UPPER TRIANGULAR PART OF GIVEN
SYMMETRIC POSITIVE DEFINITE N BY N COEFFICIENT
MATRIX.
ON RETURN A CONTAINS THE RESULTANT UPPER
TRIANGULAR MATRIX IN DOUBLE PRECISION.
N - THE NUMBER OF ROWS (COLUMNS) IN GIVEN MATRIX.
EPS - SINGLE PRECISION INPUT CONSTANT WHICH IS USED
AS RELATIVE TOLERANCE FOR TEST ON LOSS OF
SIGNIFICANCE.
IER - RESULTING ERROR PARAMETER CODED AS FOLLOWS
IER = 0 - NO ERROR
IER = -1 - NO RESULT BECAUSE OF WRONG INPUT PARAME-
TER N OR BECAUSE SOME RADICAND IS NON-
POSITIVE (MATRIX A IS NOT POSITIVE
DEFINITE, POSSIBLY DUE TO LOSS OF SIGNI-
FICANCE)
IER = K - WARNING WHICH INDICATES LOSS OF SIGNIFI-
CANCE. THE RADICAND FORMED AT FACTORIZA-
TION STEP K+1 WAS STILL POSITIVE BUT NO
LONGER GREATER THAN ABS(EPS*AIK+1,K+1)).

REMARKS
THE UPPER TRIANGULAR PART OF GIVEN MATRIX IS ASSUMED TO BE
STORED COLUMNWISE IN K*(K+1)/2 SUCCESSIVE STORAGE LOCATIONS.
IN THE SAME STORAGE LOCATIONS THE RESULTING UPPER TRIANGU-
LAR MATRIX IS STORED COLUMNWISE.
THE PROCEDURE GIVES RESULTS IF N IS GREATER THAN 0 AND ALL
CALCULATED RADICANDS ARE POSITIVE.

SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
DMFSD

METHOD
SOLUTION IS DONE USING FACTORIZATION BY SUBROUTINE DMFSC.

SUBROUTINE DSINV(A,N,EPS,IER)
APPENDIX B

DIMENSION A(11)
DOUBLE PRECISION A, DIN, WORK
C
C FACTORIZE GIVEN MATRIX BY MEANS OF SUBROUTINE DMFSD
C A = TRANSPOSE(T) * T
CALL DMFSD(A, V, EPS, IER)
IF(IER) 9, 1, 1
C
C INVERT UPPER TRIANGULAR MATRIX T
C PREPARE INVERSION-LOOP
1 IPIV=N*(N+1)/2
IND=IPIV
C
C INITIALIZE INVERSION-LOOP
DO 6 I=1,N
DIN=1.0D0/A(IPIV)
A(IPIV)=DIN
MIN=N
KEND=I-1
LANF=N-KEND
IF(KEND) 5, 5, 2
2 J=IND
C
C INITIALIZE ROW-LOOP
DO 4 K=1,KEND
WORK=0.0D0
MIN=MIN-1
LHOR=IPIV
LVER=J
C
C START INNER LOOP
DO 3 L=LANF,MIN
LVER=LVER+1
LHOR=LHOR+L
3 WORK=WORK+A(LVER)*A(LHOR)
C
END OF INNER LOOP
C
A(J)=-WORK*DIN
4 J=J-MIN
C
END OF ROW-LOOP
C
5 IPIV=IPIV-MIN
6 IND=IND-1
C
END OF INVERSION-LOOP
C
C CALCULATE INVERSE(A) BY MEANS OF INVERSE(T)
C INVERSE(A) = INVERSE(T) * TRANSPOSE(INVERSE(T))
C INITIALIZE MULTIPLICATION-LOOP
DO 8 I=1,N
IPIV=IPIV+1
J=IPIV
C
SINV 510
SINV 530
SINV 550
SINV 570
SINV 580
SINV 590
SINV 600
SINV 610
SINV 620
SINV 630
SINV 640
SINV 650
SINV 660
SINV 670
SINV 680
SINV 690
SINV 700
SINV 710
SINV 720
SINV 730
SINV 740
SINV 750
SINV 760
SINV 770
SINV 780
SINV 790
SINV 800
SINV 810
SINV 820
SINV 830
SINV 840
SINV 860
SINV 870
SINV 880
SINV 890
SINV 900
SINV 910
SINV 920
SINV 930
SINV 940
SINV 950
SINV 960
SINV 970
SINV 980
SINV 990
SINV 1000
SINV 1010
SINV 1020
APPENDIX B

C INITIALIZE ROW-LOOP
DO 8 K=1,N
 WORK=0. DO
 LHOR=J
C C START INNER LOOP
DO 7 L=K,N
 LVER=LHOR+K-I
 WORK=WORK+A(LHOR)*A(LVER)
 7 LHOR=LHOR+L
C END OF INNER LOOP
C A(IJ)=WORK
 8 J=J+K
C END OF ROW- AND MULTIPLICATION-LOOP
C
 9 RETURN
END
APPENDIX B

SUBROUTINE DMFSD

PURPOSE
FACTOR A GIVEN SYMMETRIC POSITIVE DEFINITE MATRIX

USAGE
CALL DMFSD(A,N, EPS,IER)

DESCRIPTION OF PARAMETERS
A - DOUBLE PRECISION UPPER TRIANGULAR PART OF GIVEN
SYMMETRIC POSITIVE DEFINITE N BY N COEFFICIENT
MATRIX.
ON RETURN A CONTAINS THE RESULTANT UPPER
TRIANGULAR MATRIX IN DOUBLE PRECISION.
N - THE NUMBER OF ROWS (COLUMNS) IN GIVEN MATRIX.
EPS - SINGLE PRECISION INPUT CONSTANT WHICH IS USED
AS RELATIVE TOLERANCE FOR TEST ON LOSS OF
SIGNIFICANCE.
IER - RESULTING ERROR PARAMETER CODED AS FOLLOWS
IER=0 - NO ERROR
IER=1 - NO RESULT BECAUSE OF WRONG INPUT PARAME-
TER N OR BECAUSE SOME RADICAND IS NON-
POSITIVE (MATRIX A IS NOT POSITIVE
DEFINITE, POSSIBLY DUE TO LOSS OF SIGNI-
FICANCE)
IER=K - WARNING WHICH INDICATES LOSS OF SIGNIFI-
CANCE, THE RADICAND FORMED AT FACTORIZATION
STEP K+1 WAS STILL POSITIVE BUT NO
LONGER GREATER THAN ABS(EPS*A(K+1,K+1)).

REMARKS
THE UPPER TRIANGULAR PART OF GIVEN MATRIX IS ASSUMED TO BE
STORED COLUMNWISE IN N*(N+1)/2 SUCCESSIVE STORAGE LOCATIONS.
IN THE SAME STORAGE LOCATIONS THE RESULTING UPPER TRIANGU-
LAR MATRIX IS STORED COLUMNWISE too.
THE PROCEDURE GIVES RESULTS IF N IS GREATER THAN 0 AND ALL
CALCULATED RADICANDS ARE POSITIVE.
THE PRODUCT OF RETURNED DIAGONAL TERMS IS EQUAL TO THE
SQUARE-ROOT OF THE DETERMINANT OF THE GIVEN MATRIX.

SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
NONE

METHOD
SOLUTION IS DONE USING THE SQUARE-ROOT METHOD OF CHOLESKY.
THE GIVEN MATRIX IS REPRESENTED AS PRODUCT OF TWO TRIANGULAR
MATRICES, WHERE THE LEFT HAND FACTOR IS THE TRANSPOSE OF
THE RETURNED RIGHT HAND FACTOR.
APPENDIX B

C SUBROUTINE DMFS0(A,N,EPS,IER)
C
C DIMENSION A(1)
C DOUBLE PRECISION DPIV,DSUM,A
C
C TEST ON WRONG INPUT PARAMETER N
C IF(N<1) 12,1,1
C 1 IER=0
C
C INITIALIZE DIAGONAL-LOOP
C KPIV=0
C DO 11 K=1,N
C KPIV=KPIV+K
C IND=KPIV
C LEND=K-1
C
C CALCULATE TOLERANCE
C TOL=ABS(EPS*SVGLI(A(KPIV)))
C
C START FACTORIZATION-LOOP OVER K-TH ROW
C DO 11 I=K,N
C DSUM=0.DO
C IF(LEND) 2,4,2
C
C START INNER LOOP
C 2 DO 3 L=1,LEND
C LANF=KPIV-L
C LIND=IND-L
C 3 DSUM=DSUM+A(LANF)*A(LIND)
C END OF INNER LOOP
C
C TRANSFORM ELEMENT A(IND)
C DSUM=A(IND)-DSUM
C IF(I<K) 10,5,1C
C
C TEST FOR NEGATIVE PIVOT ELEMENT AND FOR LOSS OF SIGNIFICANCE
C 5 IF(SNFL(DSUM)-TOL) 6,6,9
C 6 IF(DSUM) 12,12,7
C 7 IF(IER) 8,8,9
C 8 IER=K-1
C
C COMPUTE PIVOT ELEMENT
C DPIV=DSQRT(DSUM)
C A(KPIV)=DPIV
C DPIV=1.DO/DPIV
C GO TO 11
C
C CALCULATE TERMS IN ROW
C 10 A(IND)=DSUM*DPIV
C 11 IND=IND+I
C
C END OF DIAGONAL-LOOP
C RETURN
C 12 IER=-1
C RETURN
C END
APPENDIX C

OUTPUT FOR THE EXAMPLE

LIST OF VECTORS WITH OUTLIERS IDENTIFIED BY ASTERISKS

<table>
<thead>
<tr>
<th>J</th>
<th>Δelta</th>
<th>H/R</th>
<th>J</th>
<th>Δelta</th>
<th>H/R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.6921</td>
<td>91.12</td>
<td>34</td>
<td>1.1316</td>
<td>85.42</td>
</tr>
<tr>
<td>2</td>
<td>0.0240</td>
<td>96.05</td>
<td>35</td>
<td>0.6436</td>
<td>90.48</td>
</tr>
<tr>
<td>3</td>
<td>0.0139</td>
<td>96.32</td>
<td>36</td>
<td>2.7647</td>
<td>85.05</td>
</tr>
<tr>
<td>4</td>
<td>0.4924</td>
<td>102.30</td>
<td>37</td>
<td>0.0249</td>
<td>96.03</td>
</tr>
<tr>
<td>5</td>
<td>0.3994</td>
<td>92.37</td>
<td>38</td>
<td>0.0203</td>
<td>86.22</td>
</tr>
<tr>
<td>6</td>
<td>0.0446</td>
<td>95.64</td>
<td>39</td>
<td>0.0452</td>
<td>95.03</td>
</tr>
<tr>
<td>7</td>
<td>0.6132</td>
<td>91.44</td>
<td>40</td>
<td>1.5831</td>
<td>106.36</td>
</tr>
<tr>
<td>8</td>
<td>0.3098</td>
<td>93.12</td>
<td>41</td>
<td>0.0170</td>
<td>96.23</td>
</tr>
<tr>
<td>9</td>
<td>0.1806</td>
<td>94.08</td>
<td>42</td>
<td>0.7866</td>
<td>95.71</td>
</tr>
<tr>
<td>10</td>
<td>0.5299</td>
<td>91.90</td>
<td>43</td>
<td>0.1973</td>
<td>93.94</td>
</tr>
<tr>
<td>11</td>
<td>0.0063</td>
<td>96.60</td>
<td>44</td>
<td>0.2601</td>
<td>93.46</td>
</tr>
<tr>
<td>12</td>
<td>0.0173</td>
<td>98.14</td>
<td>45</td>
<td>0.2643</td>
<td>93.43</td>
</tr>
<tr>
<td>13</td>
<td>0.3266</td>
<td>101.35</td>
<td>46</td>
<td>0.0440</td>
<td>95.65</td>
</tr>
<tr>
<td>14</td>
<td>0.1059</td>
<td>99.55</td>
<td>47</td>
<td>0.0087</td>
<td>96.50</td>
</tr>
<tr>
<td>15</td>
<td>0.0390</td>
<td>95.74</td>
<td>48</td>
<td>0.4256</td>
<td>101.94</td>
</tr>
<tr>
<td>16</td>
<td>0.0244</td>
<td>98.32</td>
<td>49</td>
<td>2.1912</td>
<td>107.98</td>
</tr>
<tr>
<td>17</td>
<td>0.1427</td>
<td>99.94</td>
<td>50</td>
<td>0.0968</td>
<td>65.45</td>
</tr>
<tr>
<td>18</td>
<td>0.6820</td>
<td>103.21</td>
<td>51</td>
<td>3.7721</td>
<td>111.35</td>
</tr>
<tr>
<td>19</td>
<td>0.5956</td>
<td>102.83</td>
<td>52</td>
<td>0.9550</td>
<td>104.31</td>
</tr>
<tr>
<td>20</td>
<td>0.1279</td>
<td>99.74</td>
<td>53</td>
<td>0.7815</td>
<td>103.63</td>
</tr>
<tr>
<td>21</td>
<td>1.1196</td>
<td>104.90</td>
<td>54</td>
<td>0.3489</td>
<td>101.49</td>
</tr>
<tr>
<td>22</td>
<td>0.0388</td>
<td>98.98</td>
<td>55</td>
<td>3.8612</td>
<td>111.52</td>
</tr>
<tr>
<td>23</td>
<td>0.0896</td>
<td>103.24</td>
<td>56</td>
<td>7.8021*</td>
<td>117.56</td>
</tr>
<tr>
<td>24</td>
<td>0.9047</td>
<td>104.12</td>
<td>57</td>
<td>3.2624</td>
<td>110.36</td>
</tr>
<tr>
<td>25</td>
<td>0.2784</td>
<td>101.03</td>
<td>58</td>
<td>0.0253</td>
<td>58.34</td>
</tr>
<tr>
<td>26</td>
<td>0.8715</td>
<td>90.37</td>
<td>59</td>
<td>0.0146</td>
<td>96.30</td>
</tr>
<tr>
<td>27</td>
<td>0.7036</td>
<td>103.30</td>
<td>60</td>
<td>4.2446</td>
<td>82.15</td>
</tr>
<tr>
<td>28</td>
<td>1.8903</td>
<td>87.15</td>
<td>61</td>
<td>3.9288</td>
<td>82.72</td>
</tr>
<tr>
<td>29</td>
<td>2.6336</td>
<td>109.02</td>
<td>62</td>
<td>5.2739</td>
<td>83.98</td>
</tr>
<tr>
<td>30</td>
<td>0.0016</td>
<td>96.09</td>
<td>63</td>
<td>0.7131</td>
<td>91.02</td>
</tr>
<tr>
<td>31</td>
<td>0.5893</td>
<td>91.38</td>
<td>64</td>
<td>0.1841</td>
<td>94.05</td>
</tr>
<tr>
<td>32</td>
<td>2.7238</td>
<td>85.14</td>
<td>65</td>
<td>0.9473</td>
<td>90.08</td>
</tr>
<tr>
<td>33</td>
<td>1.9017</td>
<td>87.12</td>
<td>66</td>
<td>0.4621</td>
<td>102.14</td>
</tr>
</tbody>
</table>

* OUTLIER IDENTIFIED AT 0.01 SIGNIFICANCE LEVEL
SAMPLE SIZE IS 66
NC OF OUTLIERS IS 1
DELTAR = 6.3530

DATA BEFORE IDENTIFICATION OF OUTLIERS
DATA AFTER DELETION OF OUTLIERS

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>MEAN</th>
<th>S.D.</th>
<th>MEAN</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>H/R</td>
<td>97.1806</td>
<td>7.2556</td>
<td>96.8671</td>
<td>6.8897</td>
</tr>
</tbody>
</table>
APPENDIX C

LIST OF VECTORS WITH OUTLIERS IDENTIFIED BY ASTERISKS

<table>
<thead>
<tr>
<th>J</th>
<th>DELTA</th>
<th>G</th>
<th>J</th>
<th>DELTA</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.3790</td>
<td>0.70</td>
<td>34</td>
<td>0.8358</td>
<td>0.85</td>
</tr>
<tr>
<td>2</td>
<td>1.1829</td>
<td>0.75</td>
<td>35</td>
<td>0.6848</td>
<td>0.90</td>
</tr>
<tr>
<td>3</td>
<td>1.0019</td>
<td>0.80</td>
<td>36</td>
<td>0.5488</td>
<td>0.95</td>
</tr>
<tr>
<td>4</td>
<td>0.8358</td>
<td>0.85</td>
<td>37</td>
<td>0.4279</td>
<td>1.00</td>
</tr>
<tr>
<td>5</td>
<td>0.6848</td>
<td>0.90</td>
<td>38</td>
<td>0.3222</td>
<td>1.05</td>
</tr>
<tr>
<td>6</td>
<td>0.5488</td>
<td>0.95</td>
<td>39</td>
<td>0.2311</td>
<td>1.10</td>
</tr>
<tr>
<td>7</td>
<td>0.4279</td>
<td>1.00</td>
<td>40</td>
<td>0.1553</td>
<td>1.15</td>
</tr>
<tr>
<td>8</td>
<td>0.3220</td>
<td>1.05</td>
<td>41</td>
<td>0.0945</td>
<td>1.20</td>
</tr>
<tr>
<td>9</td>
<td>0.2311</td>
<td>1.10</td>
<td>42</td>
<td>0.0487</td>
<td>1.25</td>
</tr>
<tr>
<td>10</td>
<td>0.1553</td>
<td>1.15</td>
<td>43</td>
<td>0.0180</td>
<td>1.30</td>
</tr>
<tr>
<td>11</td>
<td>0.0945</td>
<td>1.20</td>
<td>44</td>
<td>0.0022</td>
<td>1.35</td>
</tr>
<tr>
<td>12</td>
<td>0.0487</td>
<td>1.25</td>
<td>45</td>
<td>0.0016</td>
<td>1.40</td>
</tr>
<tr>
<td>13</td>
<td>0.0180</td>
<td>1.30</td>
<td>46</td>
<td>0.0159</td>
<td>1.45</td>
</tr>
<tr>
<td>14</td>
<td>0.0016</td>
<td>1.35</td>
<td>47</td>
<td>0.0097</td>
<td>1.55</td>
</tr>
<tr>
<td>15</td>
<td>1.1459</td>
<td>1.40</td>
<td>48</td>
<td>0.1491</td>
<td>1.60</td>
</tr>
<tr>
<td>16</td>
<td>0.0453</td>
<td>1.45</td>
<td>49</td>
<td>0.2236</td>
<td>1.65</td>
</tr>
<tr>
<td>17</td>
<td>0.0897</td>
<td>1.50</td>
<td>50</td>
<td>0.3131</td>
<td>1.70</td>
</tr>
<tr>
<td>18</td>
<td>0.1491</td>
<td>1.55</td>
<td>51</td>
<td>0.4177</td>
<td>1.75</td>
</tr>
<tr>
<td>19</td>
<td>0.3131</td>
<td>1.60</td>
<td>52</td>
<td>0.5372</td>
<td>1.80</td>
</tr>
<tr>
<td>20</td>
<td>0.5372</td>
<td>1.70</td>
<td>53</td>
<td>0.9861</td>
<td>1.95</td>
</tr>
<tr>
<td>21</td>
<td>0.6181</td>
<td>1.80</td>
<td>54</td>
<td>1.7951</td>
<td>2.15</td>
</tr>
<tr>
<td>22</td>
<td>0.8215</td>
<td>1.85</td>
<td>55</td>
<td>2.0349</td>
<td>2.20</td>
</tr>
<tr>
<td>23</td>
<td>0.8661</td>
<td>1.90</td>
<td>56</td>
<td>4.1346</td>
<td>2.55</td>
</tr>
<tr>
<td>24</td>
<td>1.5703</td>
<td>1.95</td>
<td>57</td>
<td>4.4947</td>
<td>2.60</td>
</tr>
<tr>
<td>25</td>
<td>1.7951</td>
<td>2.10</td>
<td>58</td>
<td>14.1922 *</td>
<td>3.55</td>
</tr>
<tr>
<td>26</td>
<td>2.0349</td>
<td>2.15</td>
<td>59</td>
<td>1.0018</td>
<td>0.80</td>
</tr>
<tr>
<td>27</td>
<td>2.3137</td>
<td>2.20</td>
<td>60</td>
<td>0.8358</td>
<td>0.85</td>
</tr>
<tr>
<td>28</td>
<td>2.0575</td>
<td>2.50</td>
<td>61</td>
<td>0.6848</td>
<td>0.90</td>
</tr>
<tr>
<td>29</td>
<td>1.5901</td>
<td>0.55</td>
<td>62</td>
<td>0.5488</td>
<td>0.95</td>
</tr>
<tr>
<td>30</td>
<td>1.3790</td>
<td>0.65</td>
<td>63</td>
<td>0.4279</td>
<td>1.00</td>
</tr>
<tr>
<td>31</td>
<td>1.1829</td>
<td>0.70</td>
<td>64</td>
<td>0.3222</td>
<td>1.05</td>
</tr>
<tr>
<td>32</td>
<td>1.0019</td>
<td>0.75</td>
<td>65</td>
<td>0.0897</td>
<td>1.55</td>
</tr>
<tr>
<td>33</td>
<td>0.8661</td>
<td>0.80</td>
<td>66</td>
<td>2.8446</td>
<td>2.35</td>
</tr>
</tbody>
</table>

* OUTLIER IDENTIFIED AT 0.01 SIGNIFICANCE LEVEL
SAMPLE SIZE IS 66
NC OF OUTLIERS IS 1
DELSTAR = 6.3530

DATA BEFORE IDENTIFICATION OF OUTLIERS

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>MEAN</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>1.3773</td>
<td>0.5767</td>
</tr>
</tbody>
</table>

DATA AFTER DELETION OF OUTLIERS

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>MEAN</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>1.3438</td>
<td>0.5129</td>
</tr>
</tbody>
</table>
APPENDIX C

LIST OF VECTORS WITH OUTLIERS IDENTIFIED BY ASTERISKS

<table>
<thead>
<tr>
<th>J</th>
<th>DELTA</th>
<th>H/R</th>
<th>G</th>
<th>J</th>
<th>DELTA</th>
<th>H/R</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.4132</td>
<td>91.12</td>
<td>0.70</td>
<td>34</td>
<td>1.2653</td>
<td>89.42</td>
<td>0.85</td>
</tr>
<tr>
<td>2</td>
<td>1.5222</td>
<td>96.05</td>
<td>0.75</td>
<td>35</td>
<td>0.9751</td>
<td>90.48</td>
<td>0.90</td>
</tr>
<tr>
<td>3</td>
<td>1.3226</td>
<td>96.32</td>
<td>0.80</td>
<td>36</td>
<td>2.8389</td>
<td>85.05</td>
<td>0.95</td>
</tr>
<tr>
<td>4</td>
<td>3.1165</td>
<td>102.30</td>
<td>0.85</td>
<td>37</td>
<td>0.5015</td>
<td>96.03</td>
<td>1.00</td>
</tr>
<tr>
<td>5</td>
<td>0.7199</td>
<td>92.57</td>
<td>0.90</td>
<td>38</td>
<td>0.6559</td>
<td>98.22</td>
<td>1.05</td>
</tr>
<tr>
<td>6</td>
<td>0.6203</td>
<td>95.64</td>
<td>0.95</td>
<td>39</td>
<td>0.2377</td>
<td>95.63</td>
<td>1.10</td>
</tr>
<tr>
<td>7</td>
<td>0.6784</td>
<td>91.44</td>
<td>1.00</td>
<td>40</td>
<td>3.4797</td>
<td>196.36</td>
<td>1.15</td>
</tr>
<tr>
<td>8</td>
<td>0.4002</td>
<td>93.12</td>
<td>1.05</td>
<td>41</td>
<td>0.0979</td>
<td>96.23</td>
<td>1.20</td>
</tr>
<tr>
<td>9</td>
<td>0.2635</td>
<td>94.08</td>
<td>1.10</td>
<td>42</td>
<td>0.9158</td>
<td>90.71</td>
<td>1.25</td>
</tr>
<tr>
<td>10</td>
<td>0.5249</td>
<td>91.90</td>
<td>1.15</td>
<td>43</td>
<td>0.2202</td>
<td>93.94</td>
<td>1.30</td>
</tr>
<tr>
<td>11</td>
<td>0.1091</td>
<td>96.60</td>
<td>1.20</td>
<td>44</td>
<td>0.3527</td>
<td>93.46</td>
<td>1.35</td>
</tr>
<tr>
<td>12</td>
<td>0.1499</td>
<td>98.14</td>
<td>1.25</td>
<td>45</td>
<td>0.4353</td>
<td>93.43</td>
<td>1.40</td>
</tr>
<tr>
<td>13</td>
<td>0.6519</td>
<td>101.35</td>
<td>1.30</td>
<td>46</td>
<td>0.1363</td>
<td>95.65</td>
<td>1.45</td>
</tr>
<tr>
<td>14</td>
<td>0.1888</td>
<td>99.55</td>
<td>1.35</td>
<td>47</td>
<td>0.1967</td>
<td>96.50</td>
<td>1.55</td>
</tr>
<tr>
<td>15</td>
<td>0.0746</td>
<td>95.74</td>
<td>1.40</td>
<td>48</td>
<td>0.4257</td>
<td>101.94</td>
<td>1.60</td>
</tr>
<tr>
<td>16</td>
<td>0.0263</td>
<td>98.32</td>
<td>1.45</td>
<td>49</td>
<td>2.4135</td>
<td>107.98</td>
<td>1.65</td>
</tr>
<tr>
<td>17</td>
<td>0.1421</td>
<td>99.93</td>
<td>1.50</td>
<td>50</td>
<td>0.3134</td>
<td>99.45</td>
<td>1.70</td>
</tr>
<tr>
<td>18</td>
<td>0.7313</td>
<td>103.21</td>
<td>1.55</td>
<td>51</td>
<td>4.1166</td>
<td>111.35</td>
<td>1.75</td>
</tr>
<tr>
<td>19</td>
<td>0.6055</td>
<td>102.83</td>
<td>1.60</td>
<td>52</td>
<td>0.9969</td>
<td>104.31</td>
<td>1.80</td>
</tr>
<tr>
<td>20</td>
<td>0.3148</td>
<td>99.79</td>
<td>1.70</td>
<td>53</td>
<td>1.1297</td>
<td>103.63</td>
<td>1.95</td>
</tr>
<tr>
<td>21</td>
<td>1.1413</td>
<td>104.90</td>
<td>1.80</td>
<td>54</td>
<td>1.8467</td>
<td>101.49</td>
<td>2.15</td>
</tr>
<tr>
<td>22</td>
<td>0.7922</td>
<td>98.58</td>
<td>1.85</td>
<td>55</td>
<td>3.9882</td>
<td>111.52</td>
<td>2.20</td>
</tr>
<tr>
<td>23</td>
<td>0.0621</td>
<td>103.24</td>
<td>1.90</td>
<td>56</td>
<td>8.0630</td>
<td>117.56</td>
<td>2.55</td>
</tr>
<tr>
<td>24</td>
<td>1.1990</td>
<td>104.12</td>
<td>1.95</td>
<td>57</td>
<td>4.9986</td>
<td>110.36</td>
<td>2.60</td>
</tr>
<tr>
<td>25</td>
<td>1.6294</td>
<td>101.03</td>
<td>2.10</td>
<td>58</td>
<td>20.3487 *</td>
<td>98.34</td>
<td>3.55</td>
</tr>
<tr>
<td>26</td>
<td>6.1920</td>
<td>90.37</td>
<td>2.15</td>
<td>59</td>
<td>3.1388</td>
<td>96.30</td>
<td>0.80</td>
</tr>
<tr>
<td>27</td>
<td>2.6352</td>
<td>103.30</td>
<td>2.20</td>
<td>60</td>
<td>4.3616</td>
<td>82.15</td>
<td>0.85</td>
</tr>
<tr>
<td>28</td>
<td>2.6809</td>
<td>87.15</td>
<td>0.50</td>
<td>61</td>
<td>4.0832</td>
<td>82.72</td>
<td>0.90</td>
</tr>
<tr>
<td>29</td>
<td>11.1150 *</td>
<td>169.02</td>
<td>0.55</td>
<td>62</td>
<td>3.4158</td>
<td>83.98</td>
<td>0.95</td>
</tr>
<tr>
<td>30</td>
<td>2.3074</td>
<td>96.89</td>
<td>0.65</td>
<td>63</td>
<td>0.7541</td>
<td>91.02</td>
<td>1.00</td>
</tr>
<tr>
<td>31</td>
<td>1.3906</td>
<td>91.58</td>
<td>0.70</td>
<td>64</td>
<td>0.3372</td>
<td>94.05</td>
<td>1.05</td>
</tr>
<tr>
<td>32</td>
<td>2.7509</td>
<td>95.14</td>
<td>0.75</td>
<td>65</td>
<td>2.0682</td>
<td>90.08</td>
<td>1.55</td>
</tr>
<tr>
<td>33</td>
<td>1.9632</td>
<td>87.12</td>
<td>0.80</td>
<td>66</td>
<td>2.9773</td>
<td>102.14</td>
<td>2.35</td>
</tr>
</tbody>
</table>

* OUTLIER IDENTIFIED AT 0.01 SIGNIFICANCE LEVEL
SAMPLE SIZE IS 66
NO OF OUTLIERS IS 2
DELSTAR = 8.7115

DATA BEFORE IDENTIFICATION OF OUTLiERS
<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>MEAN</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>H/R</td>
<td>97.1805</td>
<td>7.2556</td>
</tr>
<tr>
<td>G</td>
<td>1.3773</td>
<td>0.5767</td>
</tr>
</tbody>
</table>

DATA AFTER DELETION OF OUTLIERS
<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>MEAN</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>H/R</td>
<td>96.9775</td>
<td>7.2544</td>
</tr>
<tr>
<td>G</td>
<td>1.3562</td>
<td>0.5069</td>
</tr>
</tbody>
</table>
REFERENCES


