Issue No. 7 1981 March

The GENSTAT Newsletter

EDITORS: N. G. ALVEY
ROTHAMSTED EXPERIMENTAL STATION
HARPENDEN
HERTFORDSHIRE
AL5 2JQ

M. G. RICHARDSON
NAG CENTRAL OFFICE
7 BANBURY ROAD
OXFORD
OX2 6NN

NAg

CONTENTS

Published twice yearly by Rothamsted Experimental Station Statistics Department and the Numerical Algorithms Group Limited.

Printed at NAG Central Office

EDITORIAL

This is the first issue of the GENSTAT Newsletter to be issued in its new format. Future issues will be published twice yearly, in February and August. All paid-up GENSTAT sites licensed through NAG will receive one copy of each issue free of charge; additional copies may be purchased using the order form at the back of the Newsletter.

It is intended that the Newsletter should, in future, serve as more of a GENSTAT users' forum; we therefore strongly urge users to submit material (letters to the editor, short notes, full length articles or any other form which seems appropriate). No bounds have been drawn on the content of submissions, other than those of propriety and of being in some way concerned with GENSTAT; topics which potential authors might consider include macros for specific analyses, unusual applications of GENSTAT, suggestions concerning the future development of the package and using GENSTAT in the teaching of statistics. Material may be submitted to either of the editors whose names appear on the front cover and should be received by the end of June or the end of December for inclusion in the following issue. The editors reserve the right to rephrase, amend or abridge any material submitted for inclusion in the Newsletter; they will, of course, always endeavour to preserve the authors' intentions.

The contents of any article in the Newsletter should not be considered to reflect the opinions or policy of Rothamsted Experimental Station, NAG or the editors.

STATUS REPORT

Genstat 4.03 was released to users at Rothamsted in July 1980. By the end of the February versions were available for eleven different computer ranges with eleven more in preparation. The status of Genstat implementations at the end of February is shown below.

Computer Range	Converter	Version available
Burroughs B6700	Dr. B.G. Cox, Otago	4.01
CDC (NOS etc.)	Dr. P.A. Baghurst, Adelaide	4.03
CDC (SCOPE 2)	J.D. Lloyd-Jones, Manchester	4.03
CII Iris 80	Mlle. A. Bouvier, Jouy-en-Josas	4.03
DEC System-10	J.C. Byrne, York	4.03
System-20	J.D. Griffiths, Nottingham	4.01
VAX 11	A. Nairn, Stirling	4.03
Harris	S. Morris, Chelsea	-
Hewlett Packard 3000	B.J. Martin, Ulverston	-
Honeywell (GCOS)	J.S. Lemon, Aberdeen	4.03
(MULTICS)	Dr. S.P. Evans, Bristol	-

Computer Range	Converter	Version available
IBM 360/370 & similar (OS)	N.M. McLaren, Cambridge	4.03
IBM 360/370 & similar (VM/CMS)	P. Nicholson, Leeds	4.03
ICL 1900	Dr. A. Sumner, Reading	4.01
2900	B.J. Fletcher, Edinburgh	4.03
System 4	Statistics, Rothamsted	4.03
Modcomp Classic	Dr. W. Haase, Cologne	-
Prime 400+	H.R. Simpson, Rothamsted	4.03
Siemens	Dr. J. v.d. Vooren, Naaldwijk	4.01
Telefunken TR440	A. Schutt, Osnabrück	-
Univac 1100	J. Wasniewski, Copenhagen	4.01
Xerox Sigma 6	G. West, Milton Keynes	-

We are very grateful to the converters for the large part they play in making Genstat widely available.

For the benefit of sites which may wish to make contact with others in their area, below are lists of all known GENSTAT sites, as of the end of February, classified by country and machine range. Anyone wishing to make such contact should first approach the GENSTAT coordinator at NAG Central Office.

GENSTAT SITES LISTED BY COUNTRY

AUSTRALIA

AUSTRALIAN DEP HEALTH CANBERRA
CSIRO AUSTRALIA (COMP RESEARCH)
S AUSTRALIA AGRIC ADELAIDE
S AUSTRALIA INS TECH (COMP CENT) POORAKA
U ADELAIDE (WAITE INS) S. AUSTRALIA
U AUSTRALIA NAT (COM CEN) CANBERRA
U JAMES COOK (COM CEN) N. QUEENSLAND AUS
U SYDNEY NSW (FISHER LIB)
U TASMANIA (COMP CEN) HOBART AUSTRALIA
U W AUSTRALIA (RAINE MED STATS) NEDLANDS
VICTORIA DEP OF AGRIC MELBOURNE AUSTRALIA
W AUSTRALIA DEPT AGRICULTURE S.PERTH

AUSTRIA

U GRAZ TECH (EDV) AUSTRIA
U SALZBURG (EDV) AUSTRIA

BELGIUM

U GHENT STATE (CEN DIG COM CEN) BELGIUM

BOTSWANA

BOTSWANA CENT STATISTICS OFF GABORONE

BRASIL

EMBRAPA BRASIL

CANADA

ONTARIO INST STUDIES EDUC (CANADA)
U TORONTO (FAC DENTISTRY - STATS) CANADA

DENMARK

KOEBENHAVN SC ECON BUS ADMIN DENMARK
NEUCC LYNGBY (EDB-CENT) DENMARK
NOVO RES INST BAGSVAERD DENMARK
RECKU KOEBENHAVN DENMARK
U AARHUS (RECAU)
U ODENSE (DATACENTER) DENMARK

FRANCE

CENTRE AIR ST-CYR FRANCE
CEN NAT RES ZOOT JOUY-EN-JOSAS FRANCE
U PARIS SUD (MATH) FRANCE

WEST GERMANY

INST NUM STAT KOELN W GERMANY
U BERLIN FREE W. GER.
U DUSSELDORF (RZ) W. GERMANY
U HOHENHEIM STUTTGART W GERMANY
U KIEL (RZ) WEST GERMANY
U OSNABRUECK (RZ) W GERMANY

HONG KONG

U & POLY COMP CENT HONG KONG

ICELAND

U ICELAND (COMP SER) REYKJAVIK

INDIA

ICRISAT PATANCHERU AP INDIA

ITALY

ISTIT APPL CALCOLO ROMA ITALY
U NAPOLI (STAT & DEMOG)
U ROMA (CENT DI CALC INTERFAC) ITALY

MEXICO

S.A.R.H.-I.N.I.A. MEXICO

THE NETHERLANDS

INSTITUUT TNO THE HAGUE NETHERLANDS
NETHERLANDS GOVT SER LAND RECONSTRUCTION UTRECHT
PROEFSTATION GROENTEN EN FRUITTEELT NETH
SIEMENS DATA THE HAGUE NL
U UTRECHT (SOC WET CEN DAT ANAL) NETHERLANDS
U WAGENINGEN AGRIC (COMP CEN) NL

NEW ZEALAND

LEVIN HORT RES STN N.Z.
N.Z. FOREST PRODUCTS TOKOROA NEW ZEALAND
N.Z. STATE SERVICES COMMISSION (CSD)
U CANTERBURY (LINCOLN COLL) N ZEALAND
U MASSEY (COMP CEN) PALMERSTON NORTH N.Z.
U OTAGO (COMP CEN) DUNEDIN NZ
U WELLINGTON (COMP CEN) N ZEALAND

NIGERIA

U AHMADU BELLO (COMP CEN) NIGERIA

NORWAY

U TROMSO (EDB-SENTRET) NORWAY

PAPUA NEW GUINEA

PAPUA NEW GUINEA NAT COMP CEN WAIGANI

SENEGAL

INS SENEGALAIS RES AG (OCEANOG) DAKAR

SOUTH AFRICA

U NATAL SA (COMP CEN)
U SOUTH AFRICA PRETORIA

SPAIN

U MADRID (DEP MORFOLOGIA) SPAIN

SWEDEN

U LUND (COMP CEN) SWEDEN
U STOCKHOLM (QZ DATA CENTER) SWEDEN
U UPPSALA (DATACENTER) SWEDEN

UNITED KINGDOM

BRITISH - AMERICAN TOBACCO SOUTHAMPTON
BUSINESS STATISTICS OFFICE NEWPORT
CEGB NW REGION STOCKPORT
CHEMICAL DEFENSE EST PORTON DOWN
CLINICAL RES CEN HARROW
CONTROL DATA LONDON
DOE BUILDING RES STN GARSTON
DOE (TRRL) CROWTHORNE
FORESTRY COMMISSION FARNHAM
GLAXO OPERATIONS UK GREENFORD
GLAXO OPERATIONS UK ULVERSTON
HMGCC MILTON KEYNES
IMPERIAL CANCER RESEARCH FUND LONDON
IMPERIAL TOBACCO BRISTOL
INS OCCUPATIONAL MEDICINE EDINBURGH
INST NAVAL MEDICINE GOSPORT
LUCS LTD LONDON
MET OFFICE BRACKNELL
MIN AG FISH FOOD (CENT VET LAB) WEYBRIDGE
MOD WHITEHALL LONDON
NAPIER COLL (COMP SERV) EDINBURGH
NAT FOUNDATION ED RESEARCH SLOUGH
NERC COMP SERV BIDSTON BIRKENHEAD
NERC COMP SER WALLINGFORD
NORTH WEST REG HEALTH A MANCHESTER
NORTH WEST WATER WARRINGTON
PFIZER CENTRAL RES SANDWICH
PLESSEY COMPUTERLINE ADDLESTONE
POLY CENT LONDON (COMP UNIT)
POLY COVENTRY (LANCHESTER) (COM CEN)
POLY HUDDERSFIELD
POLY NORTH EAST LONDON (COMP CEN)
POLY PLYMOUTH (COMP CENT)
POLY SOUTH BANK (COMP CENTRE) LONDON
POLY THAMES (COMP SERV) LONDON
POLY TRENT (DEP COM SER) NOTTS
RANK XEROX ENG GROUP WELWYN
RECKITT & COLMAN HULL
ROTHAMSTED EXP STN HARPENDEN
RUTHERFORD LAB - ATLAS (PRIME)
RUTHERFORD LAB (IBM) COMP &AUTO DIV
SEVERN-TRENT WATER AUTHORITY
SHELL RESEARCH LAB SITTINGBOURNE
SHELL UK LTD LONDON
SMITH KLINE & FRENCH WELWYN GARDEN CITY
TYMSHARE UK LONDON
UCC (GB) LTD LONDON
U.K.A.E.A.
U ABERDEEN (COMP CEN)
U BATH (SWRCC)
U BRADFORD (COMP LAB)
U BRISTOL (COMP CEN)
U BRISTOL (SWUCN)
U CAMBRIDGE (COMP LAB)
U COLL N WALES (COMP LAB) BANGOR
U DUNDEE (COMP CENT)
U EDINBURGH (PLU)
U ESSEX (COMP CEN)
U GLASGOW (COMP SERVICE)
U HULL (COMP CEN)
ULCC (ADVISORY SERV)

U LEEDS (COMP CEN)
U LONDON CHELSEA COLL (COMP CEN)
U LONDON IMP COLL (ZOOLOGY)
U LONDON QMC (COMP CEN)
U LOUGHBOROUGH (COMP CEN)
U MANCHESTER (RCC)
U NEWCASTLE (NUMAC)
U NOTTINGHAM (COMP CEN)
U OXFORD (OUCS)
U READING (COMP UNIT)
U SALFORD (COMP LAB)
U SHEFFIELD (COMP SER)
U SOUTHAMPTON (DEPT MATHS)
U STIRLING (COMP UNIT)
U WARWICK (COMP UNIT)
U YORK (COMP SCI DEPT)
WATER RESEARCH CENTRE MEDMENHAM
WESSEX REGIONAL HEALTH AUTH (COMP CENT) WINCHESTER

UNITED STATES OF AMERICA

N.Y. STATE AGRIC EXP STN GENEVA NY USA
U CINCINNATI (STATS) USA
U CORNELL NEW YORK
U MICHIGAN STATE (COMP CEN) E. LANSING USA
U PENNSYLVANIA (RES MED) PHILADELPHIA USA
U ROCHESTER MED CEN (BIOSTAT) NY USA

YUGOSLAVIA

U ZAGREB COMP CEN YUGOSLAVIA

GENSTAT SITES LISTED BY MACHINE RANGE

BURROUGHS

U CANTERBURY (LINCOLN COLL) N ZEALAND
U MASSEY (COMP CEN) PALMERSTON NORTH N.Z.
U OTAGO (COMP CEN) DUNEDIN NZ
U SOUTH AFRICA PRETORIA
U TASMANIA (COMP CEN) HOBART AUSTRALIA
U WARWICK (COMP UNIT)
U WELLINGTON (COMP CEN) N ZEALAND
VICTORIA DEP OF AGRIC MELBOURNE AUSTRALIA

CDC 6000/7600/CYBER

BRITISH - AMERICAN TOBACCO SOUTHAMPTON
CONTROL DATA LONDON
CSIRO AUSTRALIA (COMP RESEARCH)
DOE (TRRL) CROWTHORNE
INSTITUUT TNO THE HAGUE NETHERLANDS
ISTIT APPL CALCOLO ROMA ITALY
LUCS LTD LONDON
NETHERLANDS GOVT SER LAND RECONSTRUCTION UTRECHT
S.A.R.H.-I.N.I.A. MEXICO
S AUSTRALIA AGRIC ADELAIDE
S AUSTRALIA INS TECH (COMP CENT) POORAKA
U AARHUS (RECAU)
U ADELAIDE (WAITE INS) S. AUSTRALIA
U AHMADU BELLO (COMP CEN) NIGERIA
U BERLIN FREE W. GER.
ULCC (ADVISORY SERV)
U LONDON IMP COLL (ZOOLOGY)
U MANCHESTER (RCC)
U MICHIGAN STATE (COMP CEN) E. LANSING USA
U SYDNEY NSW (FISHER LIB)
U TROMSO (EDB-SENTRET) NORWAY
U UTRECHT (SOC WET CEN DAT ANAL) NETHERLANDS
U W AUSTRALIA (RAINE MED STATS) NEDLANDS

CII IRIS

CEN NAT RES ZOOT JOUY-EN-JOSAS FRANCE

DEC SYSTEMS 10 & 20

CLINICAL RES CEN HARROW
IMPERIAL CANCER RESEARCH FUND LONDON
ONTARIO INST STUDIES EDUC (CANADA)
POLY CENT LONDON (COMP UNIT)
POLY SOUTH BANK (COMP CENTRE) LONDON
POLY TRENT (DEP COM SER) NOTTS
U COLL N WALES (COMP LAB) BANGOR
U DUNDEE (COMP CENT)
U ESSEX (COMP CEN)
U JAMES COOK (COM CEN) N. QUEENSLAND AUS
U KIEL (RZ) WEST GERMANY
U PENNSYLVANIA (RES MED) PHILADELPHIA USA
U ROCHESTER MED CEN (BIOSTAT) NY USA
U WAGENINGEN AGRIC (COMP CEN) NL
U YORK (COMP SCI DEPT)

DEC VAX-11

ICRISAT PATANCHERU AP INDIA
PFIZER CENTRAL RES SANDWICH
RANK XEROX ENG GROUP WELWYN
U ICELAND (COMP SER) REYKJAVIK
U STIRLING (COMP UNIT)
WATER RESEARCH CENTRE MEDMENHAM

HARRIS/4

U LONDON CHELSEA COLL (COMP CEN)

HEWLETT PACKARD 3000

GLAXO OPERATIONS UK ULVERSTON

HONEYWELL SERIES 60

NERC COMP SERV BIDSTON BIRKENHEAD
U ABERDEEN (COMP CEN)
U BRISTOL (COMP CEN)

IBM 360/370 & SIMILAR

AUSTRALIAN DEP HEALTH CANBERRA
CENTRE AIR ST-CYR FRANCE
EMBRAPA BRASIL
INS OCCUPATIONAL MEDICINE EDINBURGH
INS SENEGALAIS RES AG (OCEANOG) DAKAR
ISTIT APPL CALCOLO ROMA ITALY
MET OFFICE BRACKNELL
NAT FOUNDATION ED RESEARCH SLOUGH
NEUCC LYNGBY (EDB-CENT) DENMARK
N.Z. FOREST PRODUCTS TOKOROA NEW ZEALAND
N.Z. STATE SERVICES COMMISSION (CSD)
RECKITT & COLMAN HULL
RUTHERFORD LAB (IBM) COMP &AUTO DIV
TYMSHARE UK LONDON
U.K.A.E.A.
U CAMBRIDGE (COMP LAB)
U CINCINNATI (STATS) USA

U CORNELL NEW YORK
U LEEDS (COMP CEN)
U NEWCASTLE (NUMAC)
U PARIS SUD (MATH) FRANCE
U & POLY COMP CENT HONG KONG
U SALZBURG (EDV) AUSTRIA
U STOCKHOLM (QZ DATA CENTER) SWEDEN
U TORONTO (FAC DENTISTRY - STATS) CANADA
U UPPSALA (DATACENTER) SWEDEN

ICL SYSTEM 4

DOE (TRRL) CROWTHORNE
ROTHAMSTED EXP STN HARPENDEN
U BRISTOL (SWUCN)

ICL 1900 & SIMILAR

BOTSWANA CENT STATISTICS OFF GABORONE
BUSINESS STATISTICS OFFICE NEWPORT
DOE BUILDING RES STN GARSTON
IMPERIAL TOBACCO BRISTOL
INST NAVAL MEDICINE GOSPORT
MIN AG FISH FOOD (CENT VET LAB) WEYBRIDGE
NORTH WEST REG HEALTH A MANCHESTER
NOVO RES INST BAGSVAERD DENMARK
PAPUA NEW GUINEA NAT COMP CEN WAIGANI
PLESSEY COMPUTERLINE ADDLESTONE
POLY COVENTRY (LANCHESTER) (COM CEN)
U BRADFORD (COMP LAB)
U HULL (COMP CEN)
U LOUGHBOROUGH (COMP CEN)
U NOTTINGHAM (COMP CEN)
U READING (COMP UNIT)
U SALFORD (COMP LAB)
U SHEFFIELD (COMP SER)
WESSEX REGIONAL HEALTH AUTH (COMP CENT) WINCHESTER

ICL UPPER 2900

CEGB NW REGION STOCKPORT
MOD WHITEHALL LONDON
NORTH WEST WATER WARRINGTON
N.Z. STATE SERVICES COMMISSION (CSD)
POLY HUDDERSFIELD
SEVERN-TRENT WATER AUTHORITY
U BATH (SWRCC).
U EDINBURGH (PLU)
U GLASGOW (COMP SERVICE)
U HOHENHEIM STUTTGART W GERMANY
U LONDON QMC (COMP CEN)
U OXFORD (OUCS)
U SOUTHAMPTON (DEPT MATHS)

MODCOMP

INST NUM STAT KOELN W GERMANY

PRIME

FORESTRY COMMISSION FARNHAM
GLAXO OPERATIONS UK GREENFORD
KOEBENHAVN SC ECON BUS ADMIN DENMARK
LEVIN HORT RES STN N.Z.
NAPIER COLL (COMP SERV) EDINBURGH
N.Y. STATE AGRIC EXP STN GENEVA NY USA
POLY NORTH EAST LONDON (COMP CEN)
POLY PLYMOUTH (COMP CENT)
POLY THAMES (COMP SERV) LONDON
ROTHAMSTED EXP STN HARPENDEN
RUTHERFORD LAB - ATLAS (PRIME)
SHELL RESEARCH LAB SITTINGBOURNE
W AUSTRALIA DEPT AGRICULTURE S. PERTH

SEL

CHEMICAL DEFENSE EST PORTON DOWN

SIEMENS 4004.

PROEFSTATION GROENTEN EN FRUITTEELT NETH
SIEMENS DATA THE HAGUE NL
U DUSSELDORF (RZ) W. GERMANY
U GHENT STATE (CEN DIG COM CEN) BELGIUM

TELEFUNKEN TR440

U OSNABRUECK (RZ) W GERMANY

UNIVAC 1100

NERC COMP SER WALLINGFORD
RECKU KOEBENHAVN DENMARK
SHELL UK LTD LONDON
SMITH KLINE & FRENCH WELWYN GARDEN CITY
UCC (GB) LTD LONDON
U AUSTRALIA NAT (COM CEN) CANBERRA
U GRAZ TECH (EDV) AUSTRIA
U LUND (COMP CEN) SWEDEN
U MADRID (DEP MORFOLOGIA) SPAIN
U NAPOLI (STAT & DEMOG)
U NATAL SA (COMP CEN)
U ODENSE (DATACENTER) DENMARK
U PARIS SUD (MATH) FRANCE
U ROMA (CENT DI CALC INTERFAC) ITALY
U ZAGREB COMP CEN YUGOSLAVIA

XEROX SIGMA 6

HMGCC MILTON KEYNES

M.G. Richardson
NAG Central Office
7 Banbury Road
Oxford OX2 6NN
U.K.

H.R. Simpson
Rothamsted Experimental Station
Harpenden
Hertfordshire AL5 2JQ
U.K.

PRINCIPAL COORDINATE ANALYSIS OF GROUPED DATA

Given a matrix of distances or similarities between n objects grouped into g groups it is possible to obtain the matrix of distances between the g group centroids. Also the $n \times g$ matrix of distances of the n objects from the g centroids can be obtained. The distance matrix of order g can be used as input to the *PCO* directive to locate the group centroids in k ($k < g$) dimensions. Subsequently, the $n \times g$ matrix can be used by the *ADPT* directive to add points for the n original objects to the group centroid locations.

A macro, and its associated macros, for this will be put in the macro library in due course. For anyone who wishes to use them meanwhile, a brief description and listings are given below.

The macro *BGDIST* take a symmetric matrix of original distances, *D*, together with a grouping factor, *G*, and produces the symmetric matrix of inter-centroid distances, *CD*, and the object-to-centroid distance matrix, *DCD*. Note that the centroids indexing these matrices are in the same order as the levels of the factor; however the objects indexing *DCD* are reordered so that all the objects in the first group precede all the objects in the second group, and so on: the order of groups is as for the centroids and within group order is preserved. The macro *BGDIST* should be used at run-time.

```
'MACR' BGDIST $
'LOCA' N,K,KK
'SCAL' N,K,KK
'CALC' N = NVAL(G)
   :   KK = (K = MAX(FLOAT(G)))*K
'USE/R' BGDISTB $
'ENDM'

'MACR' BGDISTB $
'LOCA' R,RC,PR,PC,I,J,IJ,LI,LJ,RS,CS,RI,CI,IJR,IJC,SUB(1...KK),NI,NJ,
       DCDT,DD,DDV,NONE
'POIN' R = R(1...N)
'SYMM' D $ R = D
'VARI' RC $ N
'SYMM' CD $ K
'INTE' PR,PC
'SCAL' I,J,IJ,IJR,IJC,NI,NJ
'CALC' I,IJ = 0
'LABE' LI
'CALC' I = I+1
   :   J = 0
'LABE' LJ
'CALC' J,IJ = J,IJ+1
   :   IJR,IJC = I,J+(J,I-1)*K
```

```
'REST' RC $ G = I; PR
   :   RC $ G = J; PC
'CALC' NI,NJ = NVAL(PR,PC)
'USE/R' BGDISTL $
'DEVA' PR,PC
'GOTO' LJ * (J.LT.I)
   :   LI * (I.LT.K)
'MATR' DCD $ N,K
   :   DCDT $ K,N
'EQUA' DCDT = SUB(1...KK)
'DIAG' DD $ K
'MATR' DDV $ 1,K
   :   NONE $ N,1 = (1)N
'CALC' DD = CD
'EQUA' DDV = DD
'CALC' DCD = PDT(NONE;DDV)-2*TRANS(DCDT)
'DEVA' R,SUB(1...KK),DS,RS,CS,DCDT,DD,DDV,NONE
'ENDM'

'MACR' BGDISTL $
'MATR' DS $ 1,1
'POIN' RS = R(PR)
   :   CS = R(PC)
'MATR' DS $ RS,CS
   :   RI,SUB(IJR) $ NI,1
   :   CI,SUB(IJC) $ NJ,1
'EQUA' RI,CI   1
'CALC' DS = SUBMAT(D)
   :   ELEM(CD; IJ) = MEAN(DS)
   :   SUB(IJR) = PDT(DS;CI)/NJ
   :   SUB(IJC) = TPDT(DS;RI)/NI
'DEVA' RI,CI
'ENDM'
```

P.G.N. Digby
Rothamsted Experimental Station
Harpenden
Hertfordshire AL5 2JQ
U.K.

HIDDEN DISCONNECTEDNESS AND ANALYSIS OF VARIANCE

Table 1, representing an example of incomplete data in a three-way classification, has been taken from L.C.A. Corsten (1957).

	a_1			a_2			a_3			a_4			a_5			a_6		
	c_1	c_2	c_3	c_1	c_2	c_3	c_1	c_2	c_3	c_1	c_2	c_3	c_1	c_2	c_3	c_1	c_2	c_3
b_1	21.6	20.1	16.4	20.2			18.9		20.1	20.2	30.9		19.8	20.1	26.5	32.2		
b_2		26.7	17.3					14.3				22.8			15.9		11.0	
b_3					22.6	25.1			21.9		20.9						21.7	28.8
b_4					18.8	12.2		24.6				17.3		20.1				10.9

Table 1

Not only are there many missing plots but also some confounding occurs as may be seen in table 2, showing the product classification $(B * A)*(C * A)$ $(B * A)$ and $(C * A)$.

C	1	2	3	1	2	3	1	3	2	1	2	3	1	2	3	1	2	3
A	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5	6	6	6
B A																		
1 1	X	X	X															
2 1		X	X															
1 2				X														
3 2					X	X												
4 2					X	X												
1 3							X	X										
2 3									X									
3 3								X										
4 3									X									
1 4										X	X							
2 4												X						
3 4											X							
4 4												X						
1 5													X	X	X			
2 5															X			
4 5														X				
1 6																X		
2 6																	X	
3 6																	X	X
4 6																		X

Table 2

This table shows that disconnectedness (i.e. subclasses with no connected values) between *C* and *B* occurs in the classes A_2, A_3, A_4 and A_6. This implies confounding of some *BA* and *CA* effects. Also the interaction *ABC* has zero dimension (d.f.) indicated by the absence of cubes $2 * 2 * 2$.

We tried to analyse these data using several packages but, in 1978, only Genstat gave the analysis that would satisfy a statistician. Others gave erroneous results or the analysis was halted by difficulty with the matrix inversion.

The Genstat statements for carrying out the analysis and the results are shown. Note that the user is warned about singularities (linear dependencies) so that corresponding corrections of dimension (d.f.) can be made. Each row of 'modifications to the model' has been annotated to show all the terms and total d.f. fitted up to that point. Table 3 lists some components that are useful in carrying through F-tests.

```
'REFE' CORSTEN
'UNIT' $ 30
'FACT' A $ 6=5(1...6)
   :  B $ 4=3(1),2(2),1,2(3,4),(2(1),2,3,4)2,3(1),2,4,1,2,2(3),4
   :  C $ 3=(1,(2,3)2)2,1,3,2,3,2,1,2,3,2,3,1,2,3,3,2,1,2,2,3,3
'VARI' Y=21.6,20.1,16.4,26.7,17.3,20.2,22.6,25.1,18.8,12.2,18.9,
         20.1,14.3,21.9,24.6,20.2,30.9,22.8,20.9,17.3,19.8,20.1,
         26.5,15.9,20.1,32.2,11.0,21.7,28.8,10.9
'PRIN/P' A,B,C,Y $ 3(4),7.1
'TERM' Y+A*B*C
'Y' Y
'FIT/ANDEV=I,PRIN=Z' A*B*C
'DROP/PRIN=Z' A.C+A.B.C
'SWIT/PRIN=Z' A.B+A.C
'DROP/PRIN=Z' B.C
'SWIT/PRIN=Z' C+A.B+A.C
'SWIT/PRIN=Z' B+C+A.B+A.C
'SWIT/PRIN=Z' A+B+A.C+B.C
'DROP/PRIN=Z' C+B.C
'SWIT/ANDEV=T,PRIN=C' B+C
'RUN'
```

A	B	C	Y
1	1	1	21.6
1	1	2	20.1
1	1	3	16.4
1	2	2	26.7
1	2	3	17.3
2	1	1	20.2
2	3	2	22.6
2	3	3	25.1
2	4	2	18.8
2	4	3	12.2
3	1	1	18.9
3	1	3	20.1
3	2	2	14.3
3	3	3	21.9
3	4	2	24.6
4	1	1	20.2
4	1	2	30.9
4	2	3	22.8
4	-3	2	20.9
4	4	3	17.3
5	1	1	19.8
5	1	2	20.1
5	1	3	26.5
5	2	3	15.9
5	4	2	20.1
6	1	1	32.2
6	2	2	11.0
6	3	2	21.7
6	3	3	28.8
6	4	3	10.9

12...

*** LINEAR DEPENDENCE DETECTED WHILE FITTING TERM A.B
4 PARAMETER(S) OF THIS TERM ARE ALIASED.
*** LINEAR DEPENDENCE DETECTED WHILE FITTING TERM A.C
4 PARAMETER(S) OF THIS TERM ARE ALIASED.
*** LINEAR DEPENDENCE DETECTED WHILE FITTING TERM B.C
4 PARAMETER(S) OF THIS TERM ARE ALIASED.
*** LINEAR DEPENDENCE DETECTED WHILE FITTING TERM A.B.C
ALL PARAMETER(S) OF THIS TERM ARE ALIASED.

13...

*** WHILE DROPPING TERM A.C
R FITTED 1 PARAMETER(S) OF TERM B.C

16...

*** LINEAR DEPENDENCE DETECTED WHILE FITTING TERM A.B
4 PARAMETER(S) OF THIS TERM ARE ALIASED.

18...

*** LINEAR DEPENDENCE DETECTED WHILE FITTING TERM B.C
3 PARAMETER(S) OF THIS TERM ARE ALIASED.

20...

***** REGRESSION ANALYSIS *****

*** REGRESSION COEFFICIENTS ***

Y-VARIATE: Y

	ESTIMATE	S.E.	T
CONSTANT	19.6000	1.5385	12.74
C 1	2.5500	2.6648	0.96
C 2	1.3833	2.1758	0.64

*** ANALYSIS OF DEVIANCE ***

Y-VARIATE: Y

TERMS	RESIDUAL DF	RESIDUAL DEVIANCE	CHANGE DF	CHANGE DEVIANCE	MEAN CHANGE	MN DEV. RATIO	Authors' annotation: Terms fitted	Total d.f. used
INITIAL MODEL								
CONSTANT	29	794.97	*	*				
MODIFICATIONS TO MODEL								
+A	24	772.37	5	22.60	4.52	0.31	A	6
+B	21	583.09	3	189.28	63.09	4.38	A+B	9
+C	19	571.61	2	11.48	5.74	0.40	A+B+C	11
+A.B	8	174.05	11	397.56	36.14	2.51	A+B+C+A.B	22
+A.C	2	28.82	6	145.22	24.20	1.68	A+B+C+A.B+A.C	28

** DENOMINATOR OF RATIO IS (RES. DEV./RES. DF) FROM LINE ABOVE, = 14.41

TERMS	RESIDUAL DF	RESIDUAL DEVIANCE	CHANGE DF	CHANGE DEVIANCE	MEAN CHANGE	MN DEV. RATIO	Authors' annotation: Terms fitted	Total d.f. used
+B.C	0	0.00	2	28.82	14.41	1.00	A+B+C+A.B+A.C+B.C	30
+A.B.C	0	0.00	0	0.00	*		A+B+C+A.B+A.C+B.C+A.B.C	30
-A.B.C	0	0.00	0	0.03	*			
-A.C	5	89.70	-5	-89.70	17.94	1.24	A+B+C+A.B +B.C	25
-A.B	16	415.75	-11	-326.05	29.64	2.06	A+B+C +B.C	14
+A.C	6	195.19	10	220.57	22.06	1.53	A+B+C +A.C+B.C	24
-B.C	9	338.07	-3	-142.88	47.63	3.30	A+B+C +A.C	21
-A.C	19	571.61	-10	-233.54	23.35	1.62	A+B+C	11
-C	21	583.09	-2	-11.48	5.74	0.40	A+B	9
+A.B	10	195.23	11	387.87	35.26	2.45	A+B +A.B	20
-A.B	21	583.09	-11	-387.87	35.26	2.45	A+B	9
-B	24	772.37	-3	-189.28	63.09	4.38	A	6
+C	22	744.31	2	28.06	14.03	0.97	A +C	8
+A.C	12	506.03	10	238.28	23.83	1.65	A +C +A.C	18
-A.C	22	744.31	-10	-238.28	23.83	1.65	A +C	8
-A	27	766.91	-5	-22.60	4.52	0.31	C	3
+B	24	595.12	3	171.79	57.26	3.97	B+C	6
+B.C	21	485.33	3	109.78	36.59	2.54	B+C +B.C	9
-B.C	24	595.12	-3	-109.78	36.59	2.54	B+C	6
-C	26	606.72	-2	-11.60	5.80	0.40	B	4
-B	29	794.97	-3	-188.25	62.75	4.35		1
+C	27	766.91	2	28.06	14.03	0.97	C	3

22 'CLOSE'
******* END OF CORSTEN AT LINE 10 USED 6656 LEFT 58579

	d.f.	S.S.		d.f.	S.S.
A	5	22.60	A∿B†	5	23.63
B	3	188.25	B∿C	3	171.79
C	2	28.06	C∿A	2	28.06
	d.f.	S.S.		d.f.	S.S.
A∿B	5	22.60	A∿B,C	5	23.51
B∿A	3	189.28	B∿A,C	3	172.70
C∿B	2	11.60	C∿A,B	2	11.48
	d.f.	S.S.		d.f.	S.S.
A∿B,C,B.C	5	69.59	A.B∿A,B	11	387.87
B∿A,C,A.C	3	167.96	A.C∿A,C	10	238.28
C∿A,B,A.B	2	21.18	B.C∿B,C	3	109.78
	d.f.	S.S.		d.f.	S.S.
A.B∿A,B,C	11	397.56	A.B∿A,B,C,A.C	7*	309.24
A.C∿A,B,C	10	233.54	A.C∿A,B,C,B.C	10	220.57
B.C∿A,B,C	3	155.86	B.C∿A,B,C,A.B	3	84.35
	d.f.	S.S.		d.f.	S.S.
A.B∿A,B,C,B.C	11	326.05	A.B∿A,B,C,A.C,B.C	6*	195.19
A.C∿A,B,C,A.B	6*	145.22	A.C∿A,B,C,A.B,B.C	5*	89.70
B.C∿A,B,C,A.C	3	142.88	B.C∿A,B,C,A.B,A.C	2*	28.83

† A∿B indicates "A eliminating B"

* Differences in degrees of freedom caused by confounding

Table 3

Note that the output was produced by release 4.01, the version current in 1978. In later versions the regression coefficients are printed using a different convention.

REFERENCE

Corsten, L.C.A. (1957) Vectors, a tool in statistical regression theory (Thesis). Mededeling Landbouwhogeschool Wagningen 58 (1), 1958. Mededeling van het Instituut voor Rassenonderzoek van Landbouwgewassen te Wageningen, no. 35.

M. Keuls, Department of Mathematics
W.J. Koops, Department of Animal Production
Agricultural University
Wageningen
The Netherlands

J.T.N.M. Thissen
IWIS - TNO
Staringgebouw
Marijkeweg 11
Wageningen
The Netherlands

AVOIDING VA 18

In Newsletter 6 Howard Simpson described a new consistency check which the compiler makes when a structure, *P* say, is redefined: if any structures currently depend on *P* a warning message (VA 18) is produced. For instance, the warning occurs if the following statements are executed:

```
'POINTER' P = R(1...4)
'SYMMAT' S $ P
'MACRO' M $
'POINTER' PSUB = DUM
'SYMMAT' SSUB $ PSUB
'CALC' SSUB = SUBMAT(S)
'ENDMACRO'
'FOR' DUM = R(1...4)
'USE/R' M $
'REPEAT'
```

The warning is generated on the second, third and fourth passes through the loop when *PSUB* is redefined since *SSUB* depends on *PSUB*.

There is nothing 'wrong' with the statements above; in a slightly more general form they give the only way of selecting a series of arbitrary sub-matrices of the matrix *S*. Here we are well aware of the redefinition of *PSUB* and do not need the warning messages to remind us. To suppress them, we must ensure that *SSUB* does not depend on *PSUB* when *PSUB* is redefined. This can be done by declaring *SSUB*, independent of *PSUB*, before *PSUB* is redefined and then redeclaring *SSUB*, dependent on *PSUB*, after *PSUB* has been redefined. So the macro is now:

```
'MACRO' M $
'SYMMAT' SSUB $ 1
'POINTER' PSUB = DUM
'SYMMAT' SSUB $ PSUB
'CALC' SSUB = SUBMAT(S)
'ENDMACRO'
```

Similar solutions can be used to avoid other situations in which VA 18 error messages are anticipated.

P.G.N. Digby
Rothamsted Experimental Station
Harpenden
Hertfordshire AL5 2JQ
U.K.

HISTOGRAMS

The new *'HISTOGRAM'* directive has proved very useful for the preliminary inspection of data. However, one problem which has occurred relates to the case when the data are supplied as counts, which must be held in an *INTEGER* vector. Often the counts are not available directly but have to be obtained using other directives (for instance *'CALCULATE'* or *'TABULATE'*); consequently the counts are held in real valued structures. Another case where this arises is when the cells of the histogram represents intervals of unequal width: for the heights of the histogram to be directly comparable they should be divided by these widths before plotting. Again the counts will be held in real valued structures (and in this case they will probably be non-integral).

In both of these cases the problem is to produce an INTEGER vector containing the corresponding counts for use with *'HISTOGRAM'*. The following run time macro will produce an *INTEGER* vector *IHT* from a *VARIATE* vector *VHT*, which is assumed to hold the counts, and then plot the histogram.

```
'MACRO' HISTOG $
'INTEGER' IHT=VHT
'HISTOGRAM' IHT
'ENDMACRO'
```

Once values have been assigned to *VHT* the histogram is obtained by the statement *'USE/R' HISTOG $*.

One point to be remembered is that the values of *VHT* will be rounded to the nearest integer before being assigned to *IHT* (see the GENSTAT Manual Part II § 4.7). If it is felt that the values should be truncated, rather than rounded, then the *'USE/R'* statement should be preceded by the statement *'CALCULATE' VHT=VHT-0.5*

G.E.L. Morris
National Vegetable Research Station
Wellesbourne
Warwickshire CV35 9EF
U.K.

WHEN READ FAILS ...

A *READ* instruction will normally attempt to read the whole of the data set defined by its list of structures. The only exception to this is when an internal inconsistency is found, such as structures of different lengths in a parallel read.

By default, reports are given on all suspect elements found as READ proceeds but this can be modified in two ways:

(1) By use of the *PRIN* option: if *E* is not set such error reports are suppressed;

(2) By use of the *ERCT* option: if this is set to a negative integer $-m$ then reports on only the first m errors are output.

When the data have been read, subsequent action depends on the settings of two options *ERCT* and *QUIT*. If *ERCT* is set to $\pm m$ and k errors have been found, then if $k \leq m$ the read is deemed successful and the job proceeds normally, i.e. the user can specify the number of data errors which he can tolerate.

If the read is unsuccessful - because of too many data errors or for more serious reasons - the rest of the current job is skipped and the *REFERENCE* command beginning the next job is sought (if there is no subsequent job a *STOP* should bring the run to a tidy conclusion).

During this search the input stream is reflected to the output. This action is not unreasonable when the data set is small, but can be tiresome if it is large.

To overcome this problem, a new option *QUIT* has been provided in GENSTAT 4.03. This allows the user to specify an input channel which is to be switched on if *READ* fails. Thus, if the data set is large, it is advisable to place it on a secondary channel and set *QUIT = 1*. If the read fails, the rest of the data file is completely ignored and subsequent jobs sought with a minimum of fuss. Care is needed if a later job is to read from the faulty data file - if input is switched back to that file it will continue from where is left off.

If more than two input channels are available (most versions of GENSTAT allow four) a further refinement is possible. If an input failure is deemed serious enough to abandon, not only the current job, but also all subsequent jobs then *QUIT = 3*, say, will switch input to another file of instructions which might contain only *'STOP'*!

H.R. Simpson
Rothamsted Experimental Station
Harpenden
Hertfordshire AL5 2JQ
U.K.

SOME NOTES ON THE PRINTING OF DATA

While the printing of most types of analysis undertaken in GENSTAT is achieved through print options of the analysis directives, the *'PRINT'* directive is needed for printing the data themselves. You have only to look at long term trials, where it may be necessary to refer to an output file months, if not years, after the original analysis, to see how important it is to print the input data and any derived variates in the output file. Hence a few comments on *'PRINT'*.

The default option of *'PRINT'* is *FORM = S*, commonly used for headings but not usually appropriate for groups of factors and variates. Here *'PRIN/P'* is better, the factors and variates appearing in parallel columns. If the number of structures to be printed is small (say less than 6), and in particular if the number of units is large, then the use of *VAR = 1* which causes parallel printing across the page makes better use of the paper and provides an easily readable output.

e.g. *'PRIN/P, VAR = 1' BLOCK,TREAT,V(1...3) $ 11.1*

results in

```
BLOCK        .              .                .
TREAT        .              .                .
V(1)       Unit 1        Unit 2     ...   Unit 10
V(2)         .              .                .
V(3)         .              .                .

BLOCK        .
TREAT        .
V(1)       Unit 11       etc.
V(2)         .
V(3)         .
```

The first 13 positions are reserved for factor and variate headings, thus leaving 119 for the data, on 132 character printers. Hence by careful setting of the format statement it can be arranged to have a specific number of units falling on the same line. This can be further simplified by the use of the print option *RHM* = n, where n is an integer less than 132, which restricts the field width to the first n positions from the left hand margin (leaving n-13 for the data).

When symbolic variate names (e.g. *V(1...3)*) are used it is particularly advisable to print headings to describe the data. As an alternative to

```
'PRIN/S' H(1...3)
```

a useful option here is *'PRIN/C'* as follows:

```
'HEAD'  H(1) = '' YIELD (KG) ''
     :  H(2) = '' FRUIT SET ''
     :  H(3) = '' MEAN FRUIT WT (G) ''
     :  HB = ''V(''
     :  HE = '')=''
     :  HG = '' VARIATES ARE:- ''
'PRIN/S' HG
'FOR' I = 1,2,3; HD = H(1...3)
'PRIN/C' HB,I,HE,HD $ 1
'REPE'
'PRIN/P' BLOCK,TREAT,V(1...3) $ 2(10),3(10.1)
```

with output

```
 VARIATES ARE:-
V( 1) = YIELD (KG)
V( 2) = FRUIT SET
V( 3) = MEAN FRUIT WT (G)

BLOCK     TREAT     V(1)     V(2)     V(3)

  .         .         .        .        .
  .         .         .        .        .
  .         .         .        .        .
```

The format for *'PRIN/C'* given overleaf is arbitrary but its presence ensures that the integer is not printed in E format.

G.H.Maude
East Malling Research Station
Maidstone
Kent ME19 6BJ
U.K.

LINE SPACING WITH PRINT

When a number of vector structures is printed in parallel the output is in the form of a two-way array with one column per structure and one row per element. The rows are single spaced on the output. Occasionally wider spacings are desirable. At present there is no *PRINT* option to control line spacing but some methods of obtaining this effect are described below. For illustration, let us declare three vector structures of length 5:

```
'NAME' COUNTRY = COLOMBIA,BRAZIL,PERU,ECUADOR,CHILE
'VARIATE' AREA = 440,3287,496,105,286
  :        POPN = 14447,66302,10857,4317,7802
'SET' OUTSET   = COUNTRY,AREA,POPN
```

Ordinarily we would use:

```
'PRINT/P' OUTSET $ 8,7,7
```

which results in single line spacing.

One possibility, for small arrays, is to use concatenated printing (see the GENSTAT Manual, part II § 5.2.3) which allows more format control; unfortunately it operates like serial printing and so the user must explicitly refer to every element of every vector. However we could get double line spacing by:

```
'NAME' T(1,4...13) $ 1
'SCALAR' T((2,5...14),(3,6...15))
'EQUATE' T(1,4...13) = COUNTRY
  :       T((2,5...14),(3,6...15)) = AREA,POPN
'PRINT'C,LABC = 1' T(1...15) $ 8,7,7,2/
```

Unfortunately this creates a lot of directory entries for any but the smallest arrays.

An alternative is to use a separate *PRINT* for each row of the table:

```
'NAME' C $ 1
'SCALAR' A,P, I = 0
'CAPTION' '' COUNTRY   AREA   POPN''
'FOR' D = (I)5
'CALC' I = I+1
'COPY' C,A,P, = OUTSET $ I
'PRINT/P,LABC = 1' C,A,P $ 8,7,7
'REPEAT'
```

This gives triple line spacing and quadruple (and greater) line spacing can be obtained by putting a *LINE* directive before the *PRINT*. (Note the use of the *LABC* option in *PRINT* - we do not want column labels on every row of the array.

In the above code, the use of *'FOR' D = (I)5*, with *I* incremented in a *CALCULATE* statement, results in the use of a single structure, *I*. The alternative *'FOR' I = 1...n* (with *n = 5* in this example) sets up *n* unnamed structures and so is very wasteful of directory space when *n* is large. See Lane (1979).

A better method is to replace the *FOR* statement by *'LABEL' TOP* and replace the *'REPEAT'* statement by *'GOTO' TOP*(I.LT.5)*. Unfortunately, such loops seem more difficult to follow.

To get double line spacing we need to fool Genstat twice: at the beginning of the *PRINT* it must think that it has enough columns on the output channel to accommodate the array without splitting it into parts; during the printing of each line it must find that one of the data structures overflows its format specification so much that a second line needs to be used to accommodate all the structures. So that none of the data structures of interest overflows onto a second line an extra names structure may be used:

```
'NAME'  FILLER $ 5
  :     FILLERIN = --------
'EQUATE' FILLER = FILLERIN
```

Since we will not want a heading for this extra structure we must label the array ourselves:

```
'CAPTION' '' COUNTRY   AREA   POPN''
```

and remember to use *LABC = 1* in *PRINT*. This extra structure occupies *8* columns, so it will overflow its format width, *f* say, if *f* < *8*. Of course this will not automatically cause a new line to be used; this will only happen if the right-hand margin setting (*rhm*) is so small that the overflow from an assumed field width of *f* to an actual field width of *8* means that *rhm* is exceeded. However we must ensure that the assumed total line width is not greater than *rhm* as, in that case, the whole array would be printed in two parts. The final constraint to note is that on machines with four bytes per word (for instance ICL System 4, IBM 360/370) any setting of *rhm* is rounded down by Genstat to an integer multiple of *4*; on other machines a similar thing may happen. A little thought should yield appropriate settings for *f* and *rhm*: in the example we can set *f = 2* to give an overall format of *8,7,7,2* and thus *rhm = 24*(or *28*). *rhm* can be set by the *RHM* option in *PRINT* or by redefining the number of characters per record using the *OUTPUT* directive. (In the latter case *NCR* must be reset before any further printing.) Thus, the instructions are:

```
'OUTPUT' 1 $ 24
'PRINT/P LABC = 1' OUTSET FILLER $ 8,7,7,2
```

The resulting output is shown below. Although the intervening lines are not completely blank, the result is visually similar to double spacing.

```
'REFERENCE' NEWSLETTER
'NAME' COUNTRY = COLOMBIA,BRAZIL,PERU,ECUADOR,CHILE
'VARIATE' AREA = 440,3287,496,105,286
    :     POPN = 14447,66302,10857,4317,7802
'SET' OUTSET = COUNTRY,AREA,POPN
'PRINT/P' OUTSET $ 8,7,7
'RUN'

 COUNTRY   AREA   POPN
COLOMBIA    440  14447
  BRAZIL   3287  66302
    PERU    496  10857
 ECUADOR    105   4317
   CHILE    286   7802

'NAME' T(1,4...13) $ 1
'SCALAR' T((2,5...14),(3,6...15))
'EQUATE' T(1,4...13) = COUNTRY
    :     T((2,5...14),(3,6...15)) = AREA,POPN
'PRINT/C,LABC=1,LABR=1' T(1...15) $ 8,7,7,2/
'RUN'

COLOMBIA
            440  14447
  BRAZIL
           3287  66302
    PERU
            496  10857
 ECUADOR
            105   4317
   CHILE
            286   7802

'NAME' C $ 1
'SCALAR' A,P,I = 0
'CAPTION' '' COUNTRY   AREA   POPN''
'FOR' D = (I)5
'CALC' I = I+1
'COPY' C,A,P = COUNTRY,AREA,POPN $ I
'PRINT/P,LABC=1' C,A,P $ 8,7,7
'REPEAT'
'RUN'
```

```
  COUNTRY   AREA   POPN


 COLOMODIA   440  14447


   BRAZIL   3287  66302


     PERU    496  10857


  ECUADOR    105   4317


    CHILE    286   7802

   23  'NAME' FILLER $ 5
   24     :   FILLERIN = --------
   25  'EQUATE' FILLER = FILLERIN
   26  'CAPTION' '' COUNTRY   AREA   POPN''
   27  'OUTPUT' 1 $ 24
   28  'PRINT/P,LABC=1' OUTSET,FILLER $ 8,7,7,2
   29  'RUN'

  COUNTRY   AREA   POPN

 COLOMBIA    440  14447
---------
   BRAZIL   3287  66302
---------
     PERU    496  10857
---------
  ECUADOR    105   4317
---------
    CHILE    286   7802
---------


   30  'CLOSE'
  ******* END OF
     NEWSLETT
 AT LINE    28     USED 481
     LEFT 976
```

REFERENCE

Lane, P. (1979) Using the 'FOR' directive.
GENSTAT Newsletter 5, 5.

P.G.N. Digby
Rothamsted Experimental Station
Harpenden
Hertfordshire AL5 2JQ
U.K.

THE ANDERSON - DARLING TEST FOR NORMALITY

OUTLINE

The macro listed below performs the Anderson - Darling test for normality on an arbitrary variate *X*. Approximate significance points are calculated by the method given by Pettitt (1977): these should be accurate to the third decimal place.

METHOD OF USE

The user must supply a variate X containing the observations. *X* may contain missing values, which will be ignored, and may be restricted before calls of the macro. The global identifiers are:

X	variate	The variate to be tested.

OUTPUT

The number of (non-missing) observations, the value of the Anderson - Darling statistic and the two adjacent significance points in the set 5%, 10%,..,95%, 97.5%, 99%, 99.5% are output. If the observed statistic is greater than the 99.5% point, or less than the 5% point, only the one adjacent point is output.

EXAMPLE

'UNIT'$ 50
'READ'X
'USE/R' NORMAL $

IDENTIFIERS

The numbers of named and unnamed structures are 22 and 9, respectively. NORMAL is a run-time macro and must be called by the method described in the GENSTAT manual, Part II § 2.2.1.

```
'MACRO'  NORMAL $
'LOCAL'  A,A1,AINF,AN,ANS,AN2,B0,B1,C,C1,END,FF,H1,H2,H3,II,N,NN,P,PS,XX,XY
'SCAL'   ANS,AN2,N,NN,PS  :  C1=1
'START'
'CALC'   N=NVAL(X)-NMV(X)
 $       NN=N+N-C1
'RUN'
'JUMP'   A*(N.GT.3)
'CAPT' ''NUMBER OF (NON-MISSING) VALUES MUST EXCEED 3''
'JUMP'   END
'LABEL'  A
```

```
'VARI'   AN $ 22  :  XX $ N  :  XY $ X
 :        C = 1,3...NN
 :        P = 5,10...95,97.5,99,99.5
 :       B0 = 0.512,0.552,0.608,0.643,0.707,0.735,0.772,0.770,
              0.778,0.779,0.803,0.818,0.818,0.801,0.800,0.756,
              0.749,0.750,0.795,0.881,1.013,1.063
 :       B1 = 2.10,1.25,1.07,0.93,1.03,1.02,1.04,0.90,0.80,0.67,
              0.70,0.58,0.42,0.12,-0.09,-0.39,-0.59,-0.80,-0.89,
              -0.94,-0.93,-1.34
 :     AINF = 0.1674,0.1938,0.2147,0.2333,0.2509,0.2681,0.2853,
              0.3030,0.3213,0.3405,0.3612,0.3836,0.4085,0.4367,
              0.4695,0.5091,0.5597,0.6305,0.7514,0.8728,1.0348,1.1578
'FACT'   FF
'INTE'   II
'HEAD'   H1 = ''NUMBER OF (NON-MISSING) VALUES IS''
 :       H2 = ''VALUE OF ANDERSON-DARLING STATISTIC IS''
 :       H3 = ''% PT OF NULL DISTRIBUTION IS''
'GROUP'  FF=INTPT(X)
'REST/C,MV=IN'  XY $ FF=*;II
'COPY'   XX=X $ II
'CALC'   XX = ORDER(NPI((XX-MEAN(XX))/SQRT(VAR(XX))))
 $      AN2 = -SUM(C*LOG(XX*(C1-REV(XX))))/N-N
 $       AN = AINF*(C1-B0/N+B1/(N*N))
 $       NN = SUM(AN.LE.AN2)
'PRIN/C,LABR=1'  H1,N,H2,AN2 $ X,0,10,/,X,0,10.4
'JUMP'   A1*(NN.EQ.0)
'CALC'   ANS=ELEM(AN;NN)  $  PS=ELEM(P;NN)
'PRIN/C,LABR=1'  PS,H3,ANS $ 3.1,0,16.4
'LABEL'  A1
'JUMP'   END*((NN=NN+C1).GT.22)
'CALC'   ANS=ELEM(AN;NN)  $  PS=ELEM(P;NN)
'PRIN/C,LABR=1'  PS,H3,ANS $ 3.1,0,16.4
'LABEL'  END
'ENDMACRO'
```

REFERENCE

Pettitt, A.N. (1977) Testing the normality of several independent samples using the Anderson-Darling statistic. Appl. Stat., 26, 156-161.

D.G. Edwards
RECKU
Vermundsgade 5
2100 København
Denmark

GENSTAT CONFERENCE

The second GENSTAT conference will be held at the International Agricultural Centre, Wageningen, The Netherlands, from 7th to 9th October 1981.

During the conference recent additions to the program and practical applications of these and other procedures will be discussed. Explanation of some of the more complex procedures will be given and there will be talks on the use of GENSTAT as an educational tool. Participants will be invited to discuss the future development of the program and the originators will be available to advise on problems encountered in using the program.

There will be three levels of conference fee:

1. £35 for those requiring lunch only.

2. £65 for those staying at the Centre. Here it may be necessary to share a room.

3. £100 for those staying in hotels (probably in Arnhem).

The last two rates cover overnight accommodation, breakfast and lunch for the three days. Fees should be paid by May 1st 1981.

Would those interested in attending please write to:

N.G. Alvey
Rothamsted Experimental Station
Harpenden
Hertfordshire AL5 2JQ
U.K.

GENSTAT VIDEO TAPES

A set of three colour video tapes introducing GENSTAT is being prepared by the London University Audio Visual Unit. Each tape will run for a little over half an hour. The first, "Getting Started", is now ready and the second, "Fitting Models", should be ready by the time this newsletter appears. The third tape, which will consist of discussions with GENSTAT users, is expected to be ready early in the autumn. The tapes are intended to give a "preview" of GENSTAT by showing examples of its use; they do not provide a self-instructional course in the package.

Each tape may be hired for £7.00 for ten days, or purchased outright for £25.00 plus the cost of the tape (about £7 to £10). Postal charges and V.A.T. are extra.

Further information may be obtained from:

Miss P. Gulliford
London University Audio Visual Unit
11 Bedford Square
LONDON WC1 B3RA
U.K.

GENSTAT NEWSLETTER ORDER FORM

To order future issues of the GENSTAT Newsletter, please complete the form below and return it to:

The GENSTAT Co-ordinator
NAG Central Office
7 Banbury Road
OXFORD
OX2 6NN
U.K.

Please note that each subscription to the newsletter costs £2.50 per annum (2 issues). If your payment, *in sterling*, is sent with the order, a special rate of £1.50 per annum applies, provided that no invoice is required. These prices include 2nd class/surface postage. Postage at other rates will be charged at cost.

To: NAG Ltd., 7 Banbury Road, OXFORD, OX2 6NN, U.K.

Please supply me with .... copies of each future issue of the GENSTAT newsletter for .... years, beginning with issue number .....

* ☐ Enclosed is my remittance for ....

* ☐ Please invoice me

Signature ....

Name and address for posting ....
(Please type or print)
....
....
....
....

Special mailing instructions ....

Cheques should be made payable to the Numerical Algorithms Group Ltd.

** tick one box*

NP128/0:Mar81