Rw Payne

*Issue No. 9* *1982 March*

# The GENSTAT Newsletter

*EDITORS:*

*R.W. PAYNE*
*ROTHAMSTED EXPERIMENTAL STATION*
*HARPENDEN*
*HERTFORDSHIRE*
*AL5 2JQ*

*M.G. RICHARDSON*
*NAG CENTRAL OFFICE*
*MAYFIELD HOUSE*
*256 BANBURY ROAD*
*OXFORD*
*OX2 7DE*

NAg

CONTENTS



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

## EDITORIAL

Much of this issue is taken up with reports of the Second Genstat Conference and the editors wish to thank speakers for taking the trouble to summarise their talks for the Newsletter.

As always, we wish to stress that the Newsletter is meant largely as a users' forum and to encourage Genstat users to contribute letters, notes, articles, announcements or anything else which they feel is of interest.

Occasionally, subscribers write or telephone with subscription queries. In the former case, it would be very helpful if the address label from the Newsletter could be enclosed; in the latter, if the reference at the top right hand corner of this label could be quoted. An explanation of this reference may be of interest. For licenced Genstat sites' free copies, it is simply the licence agreement number. For other subscribers, it consists of two numbers separated by a solidus (/). The first of these numbers indicates the final issue number of the subscription, the second, the number of copies ordered.

## STATUS REPORT

Versions of Genstat 4.03 for the Burroughs B6700 and the Univac 1100 are now available. The ICL 1900 version has been completed by Roger Sammons of the Computer Centre at Reading University and will be distributed when the Installation Notes are ready. Only the Siemens BS 2000 version of 4.03 remains to be completed.

New orders for Genstat have averaged three per month over the last six months, which, although not as high as might be hoped, is nevertheless encouraging in the present economic climate.

Lists of all Genstat sites, classified by country and machine range, are given below. Anyone wishing to contact another site should first approach the Genstat Co-ordinator at NAG Central Office.

GENSTAT SITES LISTED BY COUNTRY
================================

AUSTRALIA

AUSTRALIAN DEP HEALTH CANBERRA
CSIRO AUSTRALIA (COMP RESEARCH)
CSR LIMITED (MGMT SER GR) SYDNEY AUSTRALIA
NEW S. WALES DEPT. AGRIC HAYMARKET AUSTRALIA
S AUSTRALIA DEPT AGRIC ADELAIDE
S AUSTRALIA INS TECH (COM CEN) POURAKA
SEQEB BRISBANE AUSTRALIA
U ADELAIDE (WAITE INS) S. AUSTRALIA
U ADELAIDE S AUSTRALIA
U AUSTRALIA NAT (COM CEN) CANBERRA
U JAMES COOK (COM CEN) N. QUEENSLAND AUS
U MACQUARIE (COMP CENT) NORTH RYDE AUSTRALIA
U MELBOURNE (COMP CENT) PARKVILLE
U NEW ENGLAND (COM CEN) ARMIDALE 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 SALZBURG (EDV) AUSTRIA

BELGIUM

U GHENT STATE (CEN DIG COM CEN) BELGIUM

BRAZIL

EMBRAPA BRASIL

CANADA

PACIFIC BIOLOGICAL STN NANAIMO B.C.
U MCGILL (MATHS)
U TORONTO (FAC DENTISTRY - STATS) CANADA

CHILE

U CHILE (AGRON) SANTIAGO DE CHILE

COLOMBIA

CIAT CALI COLUMBIA

DENMARK

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

EIRE

GUINNESS DUBLIN EIRE

FRANCE

CEN NAT RES ZOOTEC JOUY-EN-JOSAS FRANCE
CENTRE AIR ST-CYR FRANCE
U PARIS SUD (MATH) FRANCE
UTAC LINAS-MONTLHERY FRANCE

WEST GERMANY

INST NUM STAT KOELN W. GERMANY
U BERLIN FREE W. GER.
U BERLIN TECH (INST STAT OEK, & OR)
U DUSSELDORF (RZ) W. GERMANY
U HOHENHEIM STUTTGART W 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 FIRENZE (CENT DI CALC) ITALY
U GENOVA (INST DI MAT) ITALY
U NAPOLI (STAT & DEMOG)
U PADOVA (CENT CALC)
U PAVIA (CENT CALC)
U ROMA (CENT DI CALC INTERFAC) ITALY

JAPAN

FUYO TOKYO JAPAN

KUWAIT

U KUWAIT (COM SER) ADELIYAH

MEXICO

S.A.R.H.-I.N.I.A. MEXICO
U MEXICO NACIONAL AUTONOMA

THE NETHERLANDS

GIST-BROCADES N.V. DELFT NETHERLANDS
INST TNO THE HAGUE NETHERLANDS
LANDINRICHTINGSDIENST UTRECHT NETHERLANDS
PROEFSTATION TUINBOUW UNDER GLAS NETH
U UTRECHT (SOC WET CEN DAT ANAL) NLANDS
U WAGENINGEN AGRIC (COMP CEN) NL

NEW ZEALAND

DSIR (PHYS & ENG) LOWER HUTT N. ZEALAND
LEVIN HORTIC.RES.STN.LEVIN N. ZEALAND
N.Z. FOREST PRODUCTS TOKOROA NEW ZEALAND
N.Z. MIN WORKS & DEVELOP WELLINGTON NORTH N.Z.
N.Z. STATE SERVICES COMMISSION (CSD)
U AUKLAND
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

INST SENEGALAIS RES AG (OCEANOG) DAKAR

SOUTH AFRICA

U NATAL SA (COMP CEN)
U SOUTH AFRICA PRETORIA

SWEDEN

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

SYRIA

ICARDA ALEPPO SYRIA

TURKEY

U EGE (E.H.B.ENS.) IZMIR TURKEY

UNITED KINGDOM

BP INTERNATIONAL LTD LONDON
BRITISH - AMERICAN TOBACCO SOUTHAMPTON
BUSINESS STATISTICS OFFICE NEWPORT (VIA CCA)
CADBURY SCHWEPPES LTD BIRMINGHAM
CAMBRIDGE COLL ARTS & TECH (COM CEN)
CEGB NW REGION STOCKPORT
CLINICAL RES CENT HARROW
DES MOWDEN HALL DARLINGTON
DOE BUILDING RES STN GARSTON
DOE (TRRL) CROWTHORNE
FORESTRY COMMISSION FARNHAM
GKN TECHNOLOGY LTD WOLVERHAMPTON
GLAXO OPERATIONS UK GREENFORD
GLAXO OPERATIONS UK ULVERSTON
HMGCC MILTON KEYNES
HUNTINGDON RESEARCH CENTRE
IMPERIAL CANCER RESEARCH FUND LONDON
INST NAVAL MEDICINE GOSPORT
INST OCCUPATIONAL MEDICINE EDINBURGH
LIFE SCIENCE RESEARCH OCCOLD SUFFOLK
LUCS LTD LONDON
MET OFFICE BRACKNELL
MIDLAND BANK COMP OPS SHEFFIELD
MIN AG FISH FOOD (CENT VET LAB) WEYBRIDGE
MOD (A&AEE) BOSCOMBE DOWN
MRC NAT INST MED RES MILL HILL
NAPIER COLL COMM & TECH (COM CEN) EDINBURGH
NAT FOUNDATION ED RESEARCH SLOUGH
NEGAS (COMP SERV) LEEDS
NERC COMP SERV BIDSTON BIRKENHEAD
NORTH STAFFS HOSPS (COMP DEP) HARTSHILL
NORTH WEST REG HEALTH A MANCHESTER
NORTH WEST WATER WARRINGTON
PFIZER CENTRAL RESEARCH SANDWICH
POLY BRIGHTON (COMP CENT)
POLY CENT LONDON (COMP UNIT)
POLY HUDDERSFIELD
POLY KINGSTON (COMP UNIT)
POLY LIVERPOOL (COMP SER)
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 EXPERIMENTAL STN (COMP DEP)
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
U.K.A.E.A. HARWELL
U.K.A.E.A. WINFRITH DORSET
U ABERDEEN (COMP CEN)
U BATH (SWRCC)
U BRADFORD (COMP LAB)
U BRISTOL (SWUCN)
U BRISTOL (COMP CEN)
U CAMBRIDGE (COMP LAB)
U COLL N WALES (COMP LAB) BANGOR
U EDINBURGH (PLU)
U ESSEX (COMP CEN)
U HULL (COMP CEN)
ULCC (ADVISORY SERV)
U LEEDS (COMP STUDIES)
U LIVERPOOL (COMP LAB)
U LONDON CHELSEA COLL (COMP CEN)
U LONDON IMP COLL (COMP CEN)
U LONDON QMC (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 (COMP SERVICES)
U STIRLING (COMP UNIT)
U SUSSEX (COMP CEN)
U WARWICK (COMP UNIT)
U YORK (COMP SCI DEPT)
W SUSSEX C.C. (COUNTY SURVEYOR) CHICHESTER
WATER RESEARCH CENTRE MEDMENHAM
WELSH WATER AUTHORITY BRECON POWYS
WESSEX REGIONAL HEALTH AUTH (COMP CENT) WINCHESTER

UNITED STATES OF AMERICA

N.Y. STATE AGRIC EXP STN GENEVA NY USA
N.Y.STATE DEPT. HEALTH ALBANY N.Y. USA
U CALIFORNIA-DAVIS (COMP CENT) U.S.A.
U CINCINNATI (STATS) USA
U COLORADO STATE FORT COLLINS U.S.A.
U CORNELL NEW YORK
U MICHIGAN STATE (COMP CEN) E. LANSING USA

YUGOSLAVIA

U ZAGREB (COMP CENT) YUGOSLAVIA

GENSTAT SITES LISTED BY MACHINE RANGE
=====================================

BURROUGHS

GIST-BROCADES N.V. DELFT NETHERLANDS
MIDLAND BANK COMP OPS SHEFFIELD
NEW S. WALES DEPT. AGRIC HAYMARKET AUSTRALIA
U CALIFORNIA-DAVIS (COMP CENT) U.S.A.
U GENOVA (INST DI MAT) ITALY
U MEXICO NACIONAL AUTONOMA
U OTAGO (COMP CEN) DUNEDIN NZ
U SOUTH AFRICA PRETORIA
U TASMANIA (COMP CEN) HOBART AUSTRALIA
U WARWICK (COMP UNIT)
VICTORIA DEP OF AGRIC MELBOURNE AUSTRALIA

CDC

BRITISH - AMERICAN TOBACCO SOUTHAMPTON
CSIRO AUSTRALIA (COMP RESEARCH)
DOE (TRRL) CROWTHORNE
INST TNO THE HAGUE NETHERLANDS
LANDINRICHTINGSDIENST UTRECHT NETHERLANDS
LUCS LTD LONDON
RIGSHOSPITALET KOEBNHAVN DENMARK
S.A.R.H.-I.N.I.A. MEXICO
S AUSTRALIA DEPT AGRIC ADELAIDE
S AUSTRALIA INS TECH (COM CEN) POURAKA
U AARHUS (RECAU)
U ADELAIDE (WAITE INS) S. AUSTRALIA
U AHMADU BELLO (COMP CEN) NIGERIA
U BERLIN FREE W. GER.
U BOLOGNA ITALY
U COLORADO STATE FORT COLLINS U.S.A.
ULCC (ADVISORY SERV)
U LONDON IMP COLL (COMP CEN)
U MANCHESTER (RCC)
U MELBOURNE (COMP CENT) PARKVILLE
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) NLANDS
U W AUSTRALIA (RAINE MED STATS) NEDLANDS

CII IRIS

CEN NAT RES ZOOTEC JOUY-EN-JOSAS FRANCE
U FIRENZE (CENT DI CALC) ITALY

DEC SYSTEMS 10 & 20

CLINICAL RES CENT HARROW
IMPERIAL CANCER RESEARCH FUND LONDON
MRC NAT INST MED RES MILL HILL
POLY CENT LONDON (COMP UNIT)
POLY KINGSTON (COMP UNIT)
POLY LIVERPOOL (COMP SER)
POLY SOUTH BANK (COMP CENTRE) LONDON
POLY TRENT (DEP COM SER) NOTTS
U COLL N WALES (COMP LAB) BANGOR
U ESSEX (COMP CEN)
U JAMES COOK (COM CEN) N. QUEENSLAND AUS
U NEW ENGLAND (COM CEN) ARMIDALE AUS
U WAGENINGEN AGRIC (COMP CEN) NL
U YORK (COMP SCI DEPT)

DEC VAX-11

CAMBRIDGE COLL ARTS & TECH (COM CEN)
DSIR (PHYS & ENG) LOWER HUTT N. ZEALAND
GKN TECHNOLOGY LTD WOLVERHAMPTON
ICARDA ALEPPO SYRIA
ICRISAT PATANCHERU AP INDIA
N.Y.STATE DEPT. HEALTH ALBANY N.Y. USA
PACIFIC BIOLOGICAL STN NANAIMO B.C.
PFIZER CENTRAL RESEARCH SANDWICH
POLY BRIGHTON (COMP CENT)
RANK XEROX ENG GROUP WELWYN
U ADELAIDE S AUSTRALIA
U CANTERBURY (LINCOLN COLL) N ZEALAND
U ICELAND (COMP SER) REYKJAVIK
U MACQUARIE (COMP CENT) NORTH RYDE AUSTRALIA
U STIRLING (COMP UNIT)
U SUSSEX (COMP CEN)
UTAC LINAS-MONTLHERY FRANCE
W SUSSEX C.C. (COUNTY SURVEYOR) CHICHESTER
WATER RESEARCH CENTRE MEDMENHAM

HONEYWELL SERIES 60

NERC COMP SERV BIDSTON BIRKENHEAD
U ABERDEEN (COMP CEN)
U BRISTOL (COMP CEN)
U PAVIA (CENT CALC)

HONEYWELL SIGMA 6

HMGCC MILTON KEYNES
MOD (A&AEE) BOSCOMBE DOWN

IBM 360 & SIMILAR

AUSTRALIAN DEP HEALTH CANBERRA
CADBURY SCHWEPPES LTD BIRMINGHAM
CENTRE AIR ST-CYR FRANCE
CIAT CALI COLUMBIA
EMBRAPA BRASIL
FUYO TOKYO JAPAN
INST OCCUPATIONAL MEDICINE EDINBURGH
INST 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. MIN WORKS & DEVELOP WELLINGTON NORTH N.Z.
N.Z. STATE SERVICES COMMISSION (CSD)
RECKITT & COLMAN HULL
RUTHERFORD LAB (IBM) COMP &AUTO DIV
SEQEB BRISBANE AUSTRALIA
SHELL RESEARCH LAB SITTINGBOURNE
SHELL THORNTON RESEARCH CENTRE CHESTER
SMITH KLINE & FRENCH WELWYN GARDEN CITY
TYMSHARE UK LONDON
U.K.A.E.A. HARWELL
U AUKLAND
U BERLIN TECH (INST STAT OEK, & OR)
U CAMBRIDGE (COMP LAB)
U CHILE (AGRON) SANTIAGO DE CHILE
U CINCINNATI (STATS) USA
U CORNELL NEW YORK
U EGE (E.H.B.ENS.) IZMIR TURKEY
ULCC (ADVISORY SERV)
U LEEDS (COMP STUDIES)
U LIVERPOOL (COMP LAB)
U MCGILL (MATHS)
U NEWCASTLE (NUMAC)
U PADOVA (CENT CALC)
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
U WELLINGTON (COMP CEN) N ZEALAND

ICL SYSTEM 4

ROTHAMSTED EXPERIMENTAL STN (COMP DEP)
U BRISTOL (SWUCN)

ICL 1900 & SIMILAR

BUSINESS STATISTICS OFFICE NEWPORT (VIA CCA)
DES MOWDEN HALL DARLINGTON
DOE BUILDING RES STN GARSTON
HUNTINGDON RESEARCH CENTRE
INST NAVAL MEDICINE GOSPORT
MIN AG FISH FOOD (CENT VET LAB) WEYBRIDGE
PAPUA NEW GUINEA NAT COMP CEN WAIGANI
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)
WELSH WATER AUTHORITY BRECON POWYS
WESSEX REGIONAL HEALTH AUTH (COMP CENT) WINCHESTER

ICL UPPER 2900

BUSINESS STATISTICS OFFICE NEWPORT (VIA CCA)
CEGB NW REGION STOCKPORT
GUINNESS DUBLIN EIRE
NEGAS (COMP SERV) LEEDS
NORTH STAFFS HOSPS (COMP DEP) HARTSHILL
NORTH WEST WATER WARRINGTON
NOVO RES INST BAGSVAERD DENMARK
N.Z. STATE SERVICES COMMISSION (CSD)
POLY HUDDERSFIELD
SEVERN-TRENT WATER AUTHORITY
U.K.A.E.A. WINFRITH DORSET
U BATH (SWRCC)
U EDINBURGH (PLU)
U HOHENHEIM STUTTGART W GERMANY
U LONDON QMC (COMP CEN)
U OXFORD (OUCS)
U SOUTHAMPTON (COMP SERVICES)

PR1ME

CSR LIMITED (MGMT SER GR) SYDNEY AUSTRALIA
DOE BUILDING RES STN GARSTON
FORESTRY COMMISSION FARNHAM
GLAXO OPERATIONS UK GREENFORD
KOEBENHAVN SC ECON BUS ADMIN DENMARK
LEVIN HORTIC.RES.STN.LEVIN N. ZEALAND
NAPIER COLL COMM & TECH (COM CEN) 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 EXPERIMENTAL STN (COMP DEP)
SHELL RESEARCH LAB SITTINGBOURNE
U MASSEY (COMP CEN) PALMERSTON NORTH N.Z.
W AUSTRALIA DEPT AGRICULTURE S. PERTH

SIEMENS BS2000

PROEFSTATION TUINBOUW UNDER GLAS NETH
U DUSSELDORF (RZ) W. GERMANY
U GHENT STATE (CEN DIG COM CEN) BELGIUM

TELEFUNKEN TR400

U OSNABRUECK (RZ) W GERMANY

UNIVAC 1100

BP INTERNATIONAL LTD LONDON
RECKU KOEBENHAVN DENMARK
SHELL UK LTD LONDON
U AUSTRALIA NAT (COM CEN) CANBERRA
U KUWAIT (COM SER) ADELIYAH
U LUND (COMP CEN) SWEDEN
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 CENT) YUGOSLAVIA

OTHER MACHINES

GLAXO OPERATIONS UK ULVERSTON
INST NUM STAT KOELN W. GERMANY
LIFE SCIENCE RESEARCH OCCOLD SUFFOLK
POLY COVENTRY (LANCHESTER) (COM CEN)
U LONDON CHELSEA COLL (COMP CEN)

*M.G. Richardson*
*NAG Central Office*
*Mayfield House*
*256 Banbury Road*
*Oxford OX2 7DE*
*U.K.*

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

## GENSTAT MACRO LIBRARY

A new-style Macro Library will be distributed with Version 4.04 of Genstat.

In the past RES and CSIRO have generously made their own Macro libraries available to all users. However, there has been much pressure from users for a Macro Library to which all can contribute. The enthusiasm at the Wageningen Genstat conference has led to the planning of a different library system which will contain refereed macros and will be, to some extent, a "supported" collection of software. The Library will be distributed by NAG and edited by Jane Bryan-Jones of the ARC Unit of Statistics and the University of Edinburgh.

Guidelines for submission of macros for the Library will be printed in the next Newsletter - draft copies are available from me and I would be glad of comments from a wide range of users.

The Library itself will contain brief documentation for each macro and a key-word classification of macros; a "help" macro will offer information such as:

> a complete list of macros, key-words or macros and their associated key-words,
>
> a list of macros associated with a subset of key-words provided by the user,
>
> a brief outline of macros, perhaps with the names of the required input parameters.

A "Macro Manual" will contain the detailed documentation of the macros and a standard set of headings will be used for each macro description.

At present we have no plans for commissioning macros - but if you identify a need for a macro, and for some reason are not able to provide the code, please let me know.

The first issue of the Macro Library will contain many of the existing macros, so please send me any comments you may have about these. If you have any macro(s) you would like to be included in this issue please write to me soon, before July if possible, for further information.

Finally if you wish to volunteer to referee for the Library please send me a note of your particular areas of experience and interest.

*Jane Bryan-Jones*
*ARC Unit of Statistics*
*University of Edinburgh*
*James Clerk Maxwell Building*
*The King's Buildings*
*Mayfield Road*
*Edinburgh EH9 3JZ*
*U.K.*

## THE SECOND GENSTAT CONFERENCE, WAGENINGEN 1981 OCTOBER 7-9

A total of 85 participants from nine countries were welcomed to the IAC, Wageningen by Ir J. Remmelts, director of the Institute TNO for Mathematics, Information Processing and Statistics. Dr J.A. Nelder (Rothamsted) opened the proceedings by outlining the changes that had occurred since the first Genstat conference in Cambridge in April 1979. The number of sites using Genstat had risen from 80 in 14 countries to 174 in 31 countries. Release 4.03, issued in July 1980, is presently available at most sites. The main developments had been the addition of time series analysis, full implementation of optimise, the addition of histograms and smoothed lines, additional calculate functions and an extended macro library, including in addition the CSIRO macro library.

Interactive Genstat is another new development and Mr H.R. Simpson (Rothamsted) discussed the progress being made. Although in its early stages, the important principle had been established that there should be little difference between batch and interactive mode Genstat. The general feeling from those who had used Genstat interactively was that it had been found particularly useful for developing programs; some did, however, question whether such a version was in fact necessary. Mr P.G.N. Digby (Rothamsted) then talked about macros, gave hints on how to write them and gave general encouragement to participants to 'have a go'. Macros likely to be of general interest could be submitted, after thorough testing, to the Newsletter.

After the break, Dr E.R. Williams (CSIRO) launched himself into the question of which designs were analysable by Genstat. He felt that there were a number of different definitions of balance and general balance in the literature and that the definition in the manual was none too clear. He stressed the importance of the sequential nature of model fitting and was again critical of the lack of emphasis of this point in the manual. Dr R.I. Baxter (CSIRO) then gave a paper, prepared with Dr D. Ratcliff (CSIRO), on a Genstat approach to statistical analysis. He stressed the need to write a program in segments and clearly set out, with continual testing at each stage to ensure it was working properly.

Mr R.W. Payne (Rothamsted) began the afternoon session by enlightening the audience as to the meaning of the covariance efficiency factor, a concept not discussed in the statistical textbooks. It indicates the extent to which covariates and treatments are non-orthogonal and a low value, indicating a high correlation, could mean that covariance analysis was being misused. Dr R. Thompson (Edinburgh) then considered the problem of combining information from the different strata of an experiment and a macro using the REGRESS directive was described.

Mr D.G. Edwards (Copenhagen) compared a macro he had developed for producing high quality graphical displays with Genstat code developed at Rothamsted. In the discussion it was pointed out that eventually it was intended to link the Genstat code with the NAG graphical subroutines, which would connect with a wide range of graphical facilities. The first day ended with a beautifully prepared and presented talk by Mrs M.E. Van den Bol (TNO, The Hague) on TNO's experiences over the past three years in learning and using Genstat. She pleaded for an introductory manual since it was difficult to learn Genstat without expert assistance readily available. Part of the difficulty was due to the manual being in English rather than Dutch; although Dr E.R. Williams, still in defiant mood, thought the language used had been 'Double Dutch'.

Dr G. Granath (Uppsala) opened the second day by describing the use of Genstat to analyse till and rock samples in an attempt to detect new areas of ore. Quadratic discriminant analysis was being used with generalised distances based on robust methods of calculating the dispersion matrix. Mr S.A. Watts (Shell), in a talk prepared with Mr G. Paterson, described the way that certain basic analyses are carried out in an automatic way at Sittingbourne and how Genstat was used for the repeated use of preset analyses. Mr P.W. Lane (Rothamsted) then described a new directive PREDICT proposed for the next release of Genstat. It is aimed at making regression output using model formulae more readable and interpretable. There was also a need to provide standard errors for predictions but this facility may not be available in time for the next release.

Dr G. Tunnicliffe Wilson (Lancaster) spoke on the new time series facilities in Genstat. He presented examples to illustrate the use of the seven directives available. We were assured that the facilities were not difficult to use and were again encouraged to 'have a go'. Mr A. Keen (TNO, Wageningen) wound up the morning session by considering various approaches to the problem of analysing data from designed experiments with measurements repeated in time.

The use of Genstat as an educational tool occupied the first session in the afternoon. M R. Astier and Mme E. Lesquoy (Paris) described their experience in teaching Genstat to postgraduate students. They found many advantages in using Genstat but again experienced difficulties, as the manual was only in the English language. Mr J. Wilkin (Coventry) then explained how Genstat was used in the modular science degree at Coventry Polytechnic. After the tea break Miss J. Bryan-Jones (Edinburgh) gave a talk on how she is trying to enable the non-statistical, non-computing person to use Genstat through a conversational version of the language. At the moment, this potentially useful work is at the exploratory stage.

The Genstat manual had been the subject of some criticism during the first two days. Mr N.G. Alvey (Rothamsted) likened it to Principia Mathematica or the Kama Sutra - worth reading if you could put in the effort. The need for a gentle lead in to the manual had been realised and a book was soon to be published which was aimed at achieving this. We were also informed that manuals or books were now available or under preparation in Danish, Dutch, French, German and Italian. Other aids, such as the Standard Analysis Forms, for making Genstat more accessible to users were also brought to our attention. NAG had taken over the distribution of Genstat at the beginning of 1980 and Mr M.G. Richardson (NAG) provided some background on the facilities available. A microfiche version of the manual was available from the NAG office (at a cost of £5) and Newsletters 1-6 were also available in this form. The second day was concluded by Mr J. Wasniewski (RECKU, Copenhagen) who spoke on the UNIVAC 1100 implementation of Genstat.

Also in progress during the second day was filming for the third Genstat audio-visual tape. A number of participants at the conference were interviewed on their use of Genstat and on how useful they had found it in their work. At least one person found the whole procedure nerve racking!

The final session of the conference was devoted to a far ranging discussion on the future of Genstat. Dr J.A. Nelder (Rothamsted) outlined the major new facilities to be available in the next version (4.04) of Genstat. The discussion then covered the type of facilities users would like to see in the future and various general current topics were aired. The conference ended after a little crystal gazing into possible new developments in both hardware and software computing facilities.

*J.A. John*
*Department of Mathematics*
*University of Southampton*
*Southampton SO9 5NH*
*U.K.*

A selection of summaries of talks from the conference follows. Authors whose conference contributions do not appear here are cordiallly invited to submit summaries for the next edition of the Newsletter.

*The Editors*

## COMBINATION OF INFORMATION FROM DIFFERENT STRATA

### SUMMARY

The information to be combined might be on treatment effects or variance parameters. In some circumstances combination is a simple modification of regression calculations. Macros to do these calculations have been written and their implementation is discussed.

### INTRODUCTION

As an example of the type of problem we are considering we take a balanced incomplete block design (BIB). For example, with three treatments, three blocks and two treatments per block the layout

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| I | A | B | II | B | C | III | C | A |

represents a BIB. Then there is usually information from a BIB on

I) Treatments, from contrasts (i) within blocks,
(ii) between blocks.

II) Variances, from contrasts (i) within blocks,
(ii) between blocks (although not in the design above),
(iii) comparison of treatment estimates in the two strata.

In the next section we concentrate on combining information on treatments and in the following one combining information on variances from balanced designs.

COMBINATION OF INFORMATION ON TREATMENTS

In many cases the data is 'balanced' and an *ANOVA* gives an analysis of variance and treatment estimates in each stratum. Efficient estimation is then merely a matter of weighting the treatment estimates in the different strata appropriately. The weighting depends on the variance parameters and their estimation fits into the framework of the next section.

An example of the type of model we are interested in is

$$y_{ij} = t_i + b_j e_{ij}$$

where $t_i$ represents the effect of treatment $i$,

$b_j$ represents the block effect, assumed to be normally distributed with variance $\sigma_b^2$ and

$e_{ij}$ represents a residual with variance $\sigma^2$.

Alternatively this model can be written as

$$E(y_{ij}) = t_i \, , \quad V(y_{ij}) = \sigma^2 + \sigma_b^2 \, , \quad Cov(y_{ij}, y_{kj}) = \sigma_b^2 \, , \quad Cov(y_{ij}, y_{km}) = 0 \quad (j \neq m)$$

indicating that plots in the same block have the same covariance. This formulation parallels the GENSTAT specification formulation where *TREATMENTS* (and *COVARIATES*) relate to the expectation of the variate and *BLOCKS* relates to the variance structure (and null analysis of variance). The effects $t_i$ and $b_j$ are often called fixed and random effects respectively.

If not all combinations of treatments and blocks are observed, the design may be unbalanced and then *REGRESS* could be used to fit the model with sums of squares due to fitting blocks and treatments in turn. However, treatment estimates will only be efficient if $\sigma_b^2 = 0$. If the ratio of $\sigma_b^2/\sigma^2$ is known, a simple modification to the least squares equations (Henderson et al, 1959), essentially adding $\sigma^2/\sigma_b^2$ to some of the diagonal terms in the sums of squares and cross products matrix (*SSP*), gives efficient estimation of the fixed effects. Further, the estimation of the variance components $\sigma^2$, $\sigma_b^2$ by maximum likelihood can easily be formulated (Patterson and Thompson, 1971; Harville, 1977) in terms of these modified least square equations.

A macro is available from the authors for fitting the general model

$$E(y) = \underset{\sim}{X}\underset{\sim}{B}, \quad V(y) = \sum_{i=1}^{N} \underset{\sim}{Z}_i \underset{\sim}{Z}_i' \theta_i + I\theta_{N+1} ,$$

where $\underset{\sim}{B}$ can include qualitative factor and quantitative (covariate) effects and $\underset{\sim}{Z}_i$ represents a blocking factor. The user has to specify initial values for the variance components. The variate to be analysed, fixed effects and random effects have to be specified twice: firstly, as lists of factors and variates (no doubt the macro could be adapted to use model formulae) secondly, as pointers in order to access the sums of squares and products matrix.

The space used and time taken naturally depend on the number of effects in *SSP* and can be a limitation in some cases. Other approaches take more advantage of the structure. For example, several of the submatrices of *SSP* are diagonal (Thompson 1977). We found it impossible to implement such procedures generally.

Unfortunately, when the diagonal terms are added to *SSP* it is reduced from double to single precision; however, the next version of Genstat will allow *SSP* to be kept in double precision. Another difficulty is that an extra level is needed for each of the random effects, because the natural aliasing between levels of a factor is lost when additions are made to the diagonal of *SSP*.

## COMBINATION OF INFORMATION ON VARIANCES

The previous section considered a procedure which allowed estimation of treatment effects and variance components in parallel. There are cases when combination of information on variances can be carried out separately.

For example, consider a balanced design with information on treatments in three strata. The residual mean squares in the three strata provide information on the strata variances, say $\theta_1$ , $\theta_2$ and $\theta_3$ . There is also information from comparisons of treatment effects in the three strata. This can be summarised as a (3-1) × (3-1) matrix:-

$$\begin{pmatrix} \sum_i (t_{i1} - t_{i3})^2 & \sum_i (t_{i1} - t_{i3})(t_{i2} - t_{i3}) \\ \sum_i (t_{i1} - t_{i3})(t_{i2} - t_{i3}) & \sum_i (t_{i2} - t_{i3})^2 \end{pmatrix}$$

where $t_{ij}$ is the estimate of treatment effect $i$ in stratum $j$ . The expectation of this matrix is a function of the strata variances and the efficiencies in the different strata (Thompson, 1980). For example,

$$E(\sum_i(t_{i1}-t_{i3})^2) = d(\theta_1/rE_1+\theta_3/rE_3)$$

where $d$ is degrees of freedom associated with the treatments,

$E_i$ is the efficiency factor in the i-th stratum and

$r$ is the replication factor.

Data of this type fit into a framework where we have a series of matrices of mean squares and products, $\underset{\sim}{y}_h$ , and their expectations are linear functions of known matrices $\underset{\sim}{X}_{hi}$ , i.e.

$$E(\underset{\sim}{y}_h) = \underset{\sim}{\mu}h = \sum_{j=1}^{P} \underset{\sim}{X}_{hj}\,\theta_j \ .$$

This is rather like linear regression with matrices $\underset{\sim}{X}_{hi}$ and $\underset{\sim}{y}_h$ replacing independent variables and dependent variates. Anderson (1973) suggests an iterative maximum likelihood scheme resembling weighted least squares with the weight depending on $\underset{\sim}{\mu}_h$.

A macro is available from the authors to fit the above model. The user has to specify the number of $\underset{\sim}{y}_h$ matrices, dimensions and degrees of freedom associated with the $\underset{\sim}{y}_h$ matrices, $p$ , the number of parameters fitted and the $\underset{\sim}{y}_h$ and $\underset{\sim}{X}_{hi}$ matrices.

## REFERENCES


Anderson, T.W. (1973) Asymptotically efficient estimation of covariance matrices with linear structure. Ann.Statist., 1, 135-141.

Harville, D.A. (1977) Maximum likelihood approaches to variance components estimation and to related problems. J.Amer.Statist. Assoc., 72, 320-338.

Henderson, C.R., Kempthorne, O., Searle, S.R. and Von Krosigk, C.N. (1959) Estimation of environmental and genetic trends from records subject to culling. Biometrics, 13, 192-218.

Patterson, H.D. and Thompson, R. (1971) Recovery of inter-block information when block sizes are unequal. Biometrika, 58, 545-554.

Thompson, R. (1977) The estimation of heritability with unbalanced data. I. Observations available on parents and offspring. Biometrics, 33, 485-495.

Thompson, R. (1980) Maximum Likelihood Estimation of Variance Components. Math. Operationsforsch.Statist., 4, 545-561.

*R. Thompson*
*P. Phillips*
*A.R.C. Unit of Statistics*
*University of Edinburgh*
*King's Buildings*
*Edinburgh EH9 3JZ*
*U.K.*

# THE USE OF GENSTAT AT IWIS-TNO

## TNO AND IWIS-TNO

"TNO", in Dutch, is short for "Nederlandsche Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek", which translated means: "Netherlands Organization for Applied Scientific Research". The Organization consists of about 35 institutes, which operate in the fields of industrial and food research, national defence and health research. IWIS is one of the institutes of the "Central Organization", the co-ordinating body of TNO. The initials "IWIS" stand for "Instituut TNO voor Wiskunde, Informatieverwerking en Statistiek", which translated means: "TNO Institute for Mathematics, Information Processing and Statistics". One of the departments of IWIS is the department of Statistics, which consists of 18 statisticians and 7 programmers. The group of Genstat users consists mainly of these people.

## EXPERIENCE DURING THE LEARNING PHASE

Our people learned to use Genstat through a three-day course given by two experts from Rothamsted. A few started to use Genstat immediately after the course, others went back to their work and only used it later. From our experience, it was a pity that we did not have an expert to consult while we were learning to use Genstat, because:

1. It is sometimes difficult to understand the text of the manual. This is caused not only by the choice of words, or by the fact that the language used is English rather than Dutch, but is also due to our bad reading habits.

2. The diagnostics are not always too clear, which interferes with their interpretation.

I think we lost a great deal of time in running jobs whilst not understanding the text of the manual, which, I suppose, would not have happened, if there had been an expert available. Less frequent users and people who are learning to use the package find the manual too voluminous; perhaps the introductory guide will lower the threshold for using Genstat.

## EXPERIENCE OF USING GENSTAT OVER A THREE YEAR PERIOD

Using Genstat has changed our work methods. Formerly, the statisticians did not know anything about computers and computer packages; all the computing was done by the programmers. However, most programmers are familiar only with the basic ideas of statistics, so the statisticians had to explain in complete detail what had to be calculated. Genstat is too complicated to work in this way and, as a result, most statisticians now do all the work themselves, only a few of them still instructing programmers to carry out the calculations.

## RESULTS OF A SURVEY INTO SOME ASPECTS OF GENSTAT

### *SATISFACTION WITH GENSTAT*

The satisfied people are also the frequent users and the group of less satisfied people coincides with the group of less frequent users. This implies that Genstat is a rather difficult package and that one needs a lot of practice to be able to use all of its features.

As for the facilities of Genstat, a feeling of satisfaction prevails. People are, in particular, quite happy with the several facilities for reading data, with the ease of using arithmetic operations on scalars, variates and matrices and with the options for saving results of procedures in structures for further use.

### *COMPARISON OF THE FACILITIES OF GENSTAT WITH OTHER PACKAGES*

Genstat is not used very often for descriptive statistical work: it is too expensive for that purpose and the data files are often too large. For more mathematical statistical use, people find Genstat superior because of the many facilities available. An experienced user of Genstat can do much more with it than with other packages.

## CONCLUSIONS

Genstat is too complicated a package for occasional use. Optimal use of all of its facilities can be achieved only if one is willing to spend a lot of time on learning the Genstat language and to persevere with its use.

*M.E. van den Bol*
*IWIS-TNO*
*Postbus 297*
*2501 BD Den Haag*
*The Netherlands*

## MAKING PREDICTIONS FROM A REGRESSION MODEL

When a regression model includes the effects of factors and perhaps also interactions between factors, the list of estimated regression coefficients is not a convenient summary of the fit. A tabular summary of fitted values, indexed by the levels of factors in the model, is more informative but can involve complex programming in Genstat.

A new directive called *PREDICT* will be available in release 4.04. This can provide tables of mean values and standard errors, classified by the factors in the current regression model. For example

*'Y' Y 'FIT' A*B*

*'PREDICT' A,B*

will print a table of means of variate *Y* classified by factors *A* and *B*. Standard errors are calculated and displayed with the means if the *PRINT* option is set to *S*, as in

*'PREDICT/PRINT=S'A,B*

When there are several y-variates, predictions are calculated for each of them.

If there are variates in the model, the means calculated by *PREDICT* are adjusted for their effects: by default all means are calculated at the mean value of each variate. This corresponds to the calculation of adjusted means by the *ANOVA* directive in an analysis of covariance for a balanced design. The adjusted means may be regarded as "predictions" of the average value the y-variate would have taken if the covariates took their mean value in all units analysed.

Predictions can be calculated for another, or for several other, values of a covariate by supplying the values in a nameable list of the directive. Similarly predictions may be formed for subsets of factor levels.

The tables formed by *PREDICT* do not have margins. To calculate marginal predictions, a *PREDICT* statement must be given which includes a subset of the model factors. Marginal predictions are then calculated from the full predictions by averaging over the levels of the other factors. By default, each level of a factor is weighted according to its overall representation in the data: this weighting is constant for all levels of the other factors. The resulting values are therefore adjusted, or standardised, for the effects of some factors. They may be regarded as predictions of the average values the y-variate would have taken if the standardising factors were proportionately replicated in the data. For example,

*'FIT' A*B 'PREDICT' A*

gives predictions for factor *A*, standardised for the effects of factor *B*.

If the marginal weights are not appropriate, equal weights may be specified by setting the option *WEIGHT=EQUAL*. Means are then adjusted for equal representation of each level of standardising factors. Any other weighting policy can be specified by supplying weights in further nameable lists.

The predictions may be formed for any generalised linear model, as well as for the linear regression model. Standard errors for prediction from non-linear models are approximations and should be treated with due caution, especially for predictions near extremes of distributions and when there is high variability.

I hope this new directive will make it easier to summarise regression analysis in Genstat. Any comments or experiences with the new directive (once it is available at your site) will be welcome.

*P.W. Lane*
*Rothamsted Experimental Station*
*Harpenden*
*Hertfordshire AL5 2JQ*
*U.K.*

## POST-GRADUATE USE OF GENSTAT

Teaching of Genstat at the University of Paris XI takes place during the first year of post-graduate statistics, at the same time as courses on Linear Models, Time Series Analysis and Multivariate Analysis. During the second post-graduate year, each student works on a statistical subject coming from such fields as Biometry, Medicine, Economics and Theoretical Statistics.

There are 15 sessions on Genstat, each 2½ hours long, beginning with teaching of the language; teaching is then oriented to the statistical interest of directives. Most of the students have no computing experience but they quickly learn how to run a simple Genstat program in batch-mode.

We choose Genstat from the many packages available because all the subjects of the post graduate course are more or less included in it and because its qualities allow a good introduction to statistics. We have found the following features attractive:

- the possibility of options in directives, with a clever choice of the default case, so that it is a very easy language for beginners to use;

- very short programs with attractive output from most of the statistical directives;

- a large choice of mathematical and statistical functions and of algebraic operations;

- the possibility of obtaining all the statistical results from every procedure.

We have found some faults in Genstat, especially in messages and diagnostics resulting from misuse.

As a pedagogical tool, we can mention:

*1. Performance of the necessary calculations for statistical work:*

For instance, when dealing with regression you can use the associated directives just as a black box, which delivers the results as output given some structures as input; this is not good pedagogically. A better way of doing it is to perform the calculations (solving the normal equations and, most of the time, inverting a matrix, which Genstat happily allows you to do) and, after various cases have been studied, the directives associated with regression can be used to provide results quickly (without ignoring the various steps along the way).

*2. Outline of a statistical problem*

For instance, to obtain results in a regression context the items needed may be summarised as follows:

| | |
|---|---|
| 1. Knowing the various variates? | 'REGRESS' |
| 2. Knowing which is the dependant variate? | 'Y' |
| 3. Knowing which are the independant variates? | 'FIT' |

To solve the problem in Genstat, you only have to answer the same questions. So, when you show the way Genstat solves a problem, you also get an outline of the items involved in that problem. Similarly for some calculations in regression, Genstat allows you *to give the variates* involved in the regression, *by the S.S.P. structure* built from these variates; with that remark you may point out that, to some extent, the information necessary for regression, obtained from the data, is contained in the S.S.P. or covariance matrix. So teaching Genstat can lead to the review of some theoretical knowledge.

*3. Genstat does not decide for you*

Very few statistical tests are calculated and, when they exist, do not indicate whether the result of the test is acceptable. On the other hand, one can easily obtain various tests, even in multivariate analysis thanks to:

- the Genstat operations on matrices (determinant, latent roots and vectors)

- the facilities for getting *all the results* from ANOVA or regression.

so that the dispersion matrices involved in multivariate analysis are readily calculated. Thus Genstat forces users to know the appropriate statistical tests if they wish to know the significance of results.

After 8 hours' learning, students are able to use a simple Genstat program; after that, they easily learn how to do ANOVA and regression and are very soon able to solve such problems by themselves. The time spent at the beginning is not lost, because it allows the language to be learnt so that in dealing with new directives all the interest is focused on the statistical background. We do not use all the facilities of Genstat with the students but we know if they need a feature which has not been taught, they are able to find it for themselves.

As well as being helpful in teaching, Genstat is a useful tool for post-graduate statistics students.

*R. Astier*
*E.R.A. 532*
*Bâtiment 425 Mathêmatique*
*Universitê Paris-Sud*
*91405 Orsay*
*France*

## GENSTAT IN A CNAA DEGREE

### THE DEGREE

About 200 students enter the first year of the Polytechnic Modular Degree in Science. A range of modules is offered from each of seven main subject areas, allowing students to qualify for a Degree or Honours Degree in either a single subject or combination of two main subjects. Single subject degrees are normally four year sandwich courses in which the third year is a professional training period. Combined science programmes can sometimes include a professional training year but are usually three year full time polytechnic based courses.

### *FIRST YEAR*

The first year is designed as a broad based study so that students can defer their commitment to single subject or combined science programmes until the beginning of the second year. All students take three main subject modules in the first year. An introductory statistics half module of 3 hours per week for 30 weeks is offered to all students in the first year. Of the 100 or so first year students who choose this option, about half have no prior knowledge of calculus. These students take another half module in Foundations of Mathematics. This is common choice for students whose main interest is in biology, geography or economics and who wish to consider combining statistics with another subject in subsequent years.

Students who enter the first year with a mathematics degree in mind take 6 hours per week in the main subject mathematics module plus a half module in numerical methods, together with the introductory statistics half module.

*SECOND YEAR*

In the second year students choose 6 modules from one or more subject areas. Four statistics modules are available, each occupying 4 hours per week for 30 weeks.

BS1, 2, 3 are for students taking statistics as a major subject in their degrees. They may be aiming for mathematics degrees or for combined science degrees in statistics with another subject.

BS4 is a second course in statistics for non-specialist statisticians.

BS1 is the core module which must be taken by all students following the statistics option. BS2 is available only to students who have studied the main subject mathematics module in the first year. BS3 can be taken by mathematicians and combined science statisticians entering with less mathematical background. BS4 is primarily for single subject biologists and chemists wanting a second course in statistics.

BS1 is half operations research and half statistics, containing the necessary distribution theory, design and analysis of experiments, and regression to enable progression to final year statistics modules. BS3 provides a full module in which statistical methods can be developed in more detail with realistic examples. Genstat is used as an integral part of this module.

Some use is made of Genstat in BS4 although, given the weak mathematical background of some students and the fact that this may be their first contact with a computer, there are problems.

*FINAL YEAR*

At final year, level 2 statistics modules are offered. CS1 "Design and Multivariate" is a module in which coursework problems often demand the use of a computer statistical package. Those who are familiar with it usually opt for Genstat.

USE OF GENSTAT

*INTRODUCING GENSTAT*

Unfortunately it is rare for combined science students to have a positive interest in programming. For many students, particularly those who are not primarily mathematicians, their first attempts to use a large mainframe computer, through an introductory course in BASIC or FORTRAN, are traumatic. Therefore, Genstat is introduced gradually as the need for the computer is generated by the statistical techniques used and by examples of realistically large data sets.

Logistic difficulties in submitting a card deck which satisfies computer centre rules and receiving output within a reasonable time interval are frequently major barriers to be overcome before students can make effective use of Genstat. I aim to get through this initial period as quickly as possible, by minimising the potential sources of error which might cause a job to abort and produce pages of incomprehensible output. To this end, examples of small programs are given to the students: these examples are already punched and can be run without further alteration, when the student has punched the necessary personal identification on the first card.

Writing effective programs in Genstat at this level does not need great skill in the organisation of algorithms, flowcharts and formatting of input and output which may be essential for the competent FORTRAN programmer. Understanding and manipulating the various data structures in Genstat is the key and has a much more direct relationship to the practical problems of data analysis.

About 4 hours in the first two weeks of the course are spent giving the students a brief introduction to the syntax of Genstat and showing them how to use card punches and run a job. They are given a list of the most common directives, with examples of their most straightforward use.

Throughout the course graded exercises are set for homework. Initially, these involve only small changes to the punched examples provided and later progress to the writing of small programs from scratch. Typically, there are about 10 homework assignments which require the use of Genstat to a greater or lesser extent. These are spread over the academic year from October until May, followed by an examination in June. Examples and exercises are chosen from application areas in biology, economics and geography which are the most common subjects to be combined with statistics.

### *TEACHING STATISTICS WITH GENSTAT*

From easy exercises involving use of the calculate directive, tabular and graphical output for small data sets, I progress very rapidly to an example in which a range of physiological measurements is given for a sample of 200 critically ill patients. Raw data held in computer filestore is read by a Genstat program with the necessary data matrix declarations and read format provided for the students in a further computer file. This generates the need for a certain amount of detailed information about connecting programs to data files and the use of several input channels in one job. However, although a handful of more experienced mathematics students may do some file editing at a terminal, a deck of 8 cards is enough to obtain a first tabulation from this data set.

Students are asked to investigate such questions as the relationship between sex, weight, height, survival and some of the physiological variables. Class presentation is relatively straightforward, looking at the construction of one or two way tables of counts, means and percentages. This is dovetailed with the discussion of statistical ideas in examining two way tables of counts, frequency distributions and tables of means. Examples are the use of chi-squared and its limitations for contigency tables, t tests paired and unpaired for means, and fitting theoretical distributions to frequency curves. Histograms are examined for departures from normality.

Emphasis is always on using the computer where necessary, for example, to access a large data set or in multiple regression to perform tasks which would be tedious or impossible to carry out by hand. Hand calculation and graph plots may be used even where a few more lines in the Genstat program might provide a more rounded piece of computer output. A mixture of hand and computer work is used throughout. Sometimes duplicated output from Genstat programs is provided for discussion and interpretive exercises.

*REGRESSION*

Regression is probably the central topic in the applied statistics module. Some theory and a small amount of practical application is taught in the core statistics module over four 3 hour sessions at roughly the level of the first two chapters in the book by Draper and Smith. The applied course takes a more leisurely approach, mixing more theory and practical examples. Simple linear regressions, scatter plots of raw data and standardised residuals are produced by the students. Model violations are discussed in detail. Transformations to linearity and normality are considered with students trying various possibilities on different data sets.

The pattern of Genstat directives for simple linear regression is easily extended into multiple regression. Residual plots can be produced both in illustrative examples and exercises, so that the various kinds of problems which arise can be shown and different solutions tried. Automatic selection of predictor variables is introduced together with the various criteria for assessing model fits.

Use is made of the facility for saving the variance-covariance matrix of parameter estimates to calculate confidence intervals for predicted values. The need for separate regressions, with male and female patients for example, is examined by introducing factors into the regression model.

*ANOVA AND MULTIVARIATE ANALYSIS*

The Genstat directives for the analysis of variance have not been used, because students seem to need extensive familiarity with hand calculations in ANOVA before they can understand the estimation and testing involved. The powerful directives for specifying factorial models with blocking factors seems to offer more than can be readily assimilated at this level. Some use is made of the ANOVA facilities in the final year design and analysis of experiments module. Principal components analysis is introduced in second year applied statistics as a practical procedure for summarising multivariate data. In the final year multivariate course, some students choose to approach computer based coursework through the use of the Genstat multivariate directives.

EXPERIENCE

Experience with three sets of students, so far, suggests that the element of discovery in exporing a large data set held in computer file store (and therefore not physically available) results in a more positive attitude to computer use than is sometimes the case with introductory programming courses. Delight and satisfaction at the appearance of an interesting histogram or scatter plot or a successful regression model can be a spur to further interest in both statistics and computing. A measure of this interest is the fact that combined science students frequently use Genstat in the analysis of data under consideration in biology, geography or economics modules.

Although combined science students can complete their degree in three years many are choosing to take an extra year for professional training in industry, agricultural research or government establishments. Since Genstat is increasingly used at such places, their polytechnic experience with the program can be very useful. Prospects for employment may even be enhanced when the iob specification says that familiarity with Genstat would be an advantage.

REFERENCES

Draper, N.R., and Smith, H. (2nd edition, 1980) Applied Regression Analysis. Wiley

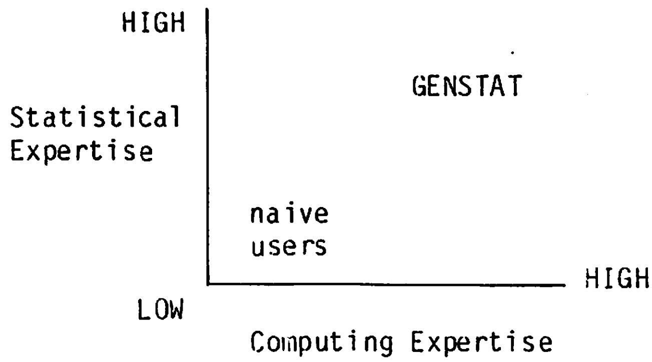
*J. Wilkin*
*Department of Statistics & Operational Research*
*Coventry (Lanchester) Polytechnic*
*Priory Street*
*Coventry CV1 5FB*
*U.K.*

## A CONVERSATIONAL APPROACH TO USING GENSTAT

### INTRODUCTION

In my talk at the Genstat Conference, I considered why a conversational approach to Genstat is needed, how it might be provided and some alternatives. I used the term "naive user" of a statistical package for a user who is inexperienced in both statistical analysis and computing, as shown in the "User Profile" (Figure 1, based on a diagram by Payne (1981)). On the Profile I have shown a gap between the naive user and Genstat; however, nowadays there are many systems that cope well with the naive computer user. In what follows, the main consideration is to broaden the scope of Genstat over the range of computing expertise and I assume that the naive user works within his statistical competence and consults a statistician for advice about interpretation and inference.

Figure 1: "USER PROFILE" for Statistical Computing.



One such naive user is Jim, a scientist who wants to use simple data summaries and statistical analysis in his research; he is neither a statistician nor a computer user but time and economic pressures force him to do as much as he can for himself. He recognises that he should use a computer and a statistical package for the computational work, and that there is scope for him to build on his experience and use the statistical package for other computations.

## CHOOSING A STATISTICAL PACKAGE

It is actually very difficult for someone like Jim to decide which statistical package to use. He doesn't really know what he should be looking for, nor how to assess what he finds out, and the literature on this subject is much more likely to confuse than help. It is easy to say that this is an area where the statistician can and should advise but most of our experience is with packages that suit us and we cannot assess packages from a naive user's point of view. (Francis (1980) discusses the collection and organisation of information about statistical programs.)

At the 1981 Inter-University Software Committee (IUSC) Workshop on Interactive Statistics Software it was generally agreed that if you want to do any "real" statistics you should use Genstat or Glim; it was also clear that Genstat is considered a statisticians' package, not a statistical package. However, the most frequently used SPSS techniques are those for data summary, tabulation and simple regression (Byrne 1981): work which, except for very large data sets, Genstat can do just as well as, if not better than, packages like SPSS and MINITAB.

## THE NAIVE USER AND GENSTAT

How can we create an environment in which the naive user can easily make Genstat do what he sees as simple calculations? If he wants to "plan and write Genstat programs" he needs to know the syntax and semantics of the language; this calls for a large investment of his time even if he learns with Alvey, Galway and Lane (1982), a very welcome introductory text. But why should "using Genstat" mean "planning and writing Genstat programs"? Indeed, there are several ways to postpone, or even remove, the need to learn the Genstat language: I will describe three possibilities external to Genstat, and then discuss in more detail what can be done within Genstat itself.

a) A *Genstat Consultant* could provide a complete Genstat program, with good clear instructions, but this is inflexible and there could be problems when Jim wants to do something different next time.

b) A *Genstat tutorial program* could be provided, so that Jim writes his own Genstat code with tutorial help and guidance. Associated with the tutorial program there would be a friendly compiler to compile Jim's program and make suggestions (hopefully sensible ones) for correcting any mistakes and errors. This solution looks promising and it is particularly relevant for preparing code on a microprocessor and sending a complete job to another computer.

c) A *Genstat Code Generator* is a program that questions the user in order to complete a prepared piece of Genstat code, in the manner of Alvey's "Standard Analysis Forms". Some code generators exist but were written in local or machine-dependent languages and, as far as I know, do not incorporate logical checks for the resulting Genstat program or syntax checks of any user-supplied Genstat code: in Edinburgh we have ARCUSGEN written in IMP (though now being re-designed); in Kent, Shell Research Centre have used the Univac command interpreter RSG; Schektman (1981) described the LAPUMS system, which could be used with Genstat. The skeleton Genstat code may be embedded in the question/answer program (Shell, LAPUMS) or may be contained in a computer file (ARCUS). A Code Generator may require the user to actually respond in the Genstat language, though this has obvious disadvantages for the naive user.

An extension of the Code Generator would be to run the conversational program and Genstat *apparently simultaneously*. This has similarities with some of the external solutions but provides more flexibility, because we need not limit the possible tasks the user sets out to complete, nor the order in which he tackles these tasks (except where one task is dependent on the successful completion of another). One possibility is to call Genstat from a program written in another language more suited to conversations: some machine-dependent work has been done on this at the Forestry Commission at Farnham. Another possibility, which we are investigating at ARCUS, is to use the Genstat language to provide both the conversations and the computations *within Genstat:* this is discussed in the next section.

## CONVERSATIONS WITHIN GENSTAT

Our aim is to provide a series of software tools which let the user work with his data without having to know the Genstat language. Two sets of macros are being designed and written: one set questions the user to discover what he wants to do next and what parameters describe his data and requirements; other macros are then called to actually manipulate the data. Obviously, the conversational macros are designed for interactive work; the computational ones could be equally useful for both batch and interactive work. A naive user would probably use both sets of macros and his work would be *menu-driven;* when he had more experience, he could forgo the conversational macros, write Genstat code to set parameter values and then call the computational macros: a *user-driven* mode of working. Note that a small site/machine-dependent program will be needed to assign file channels, start interactive Genstat and begin the conversation.

Although I claim that this introduces flexibility, it also introduces yet another language. We plan to involve some researchers in the design of the interface, so that the terminology of the language is familiar to a scientist using data analysis and statistics.

Often "conversational" systems are faulted because users gain experience and then find a lot of redundancy in the conversation. I am proposing three levels of communication: sentence, short prompts and no conversation at all.

## THE SUITABILITY OF THE GENSTAT LANGUAGE FOR "CONVERSATIONS"

Some useful features of the Genstat language, in addition to *MACRO*, are: *POINTER* for referring to structures without knowing their names; variables in a *FORMAT* for sensible printing; *POSITION* to permit alternative responses from the user,

e.g. *'POSITION' I = RESPONSE $ YES,yes,OK,oui.*

There are some limitations, for example *string handling* can only be done at a fairly crude level and thus, although a reasonably intelligent data entry environment can be provided, natural language conversations cannot!

A serious difficulty is that of recognising potential Genstat faults before the program intervenes with its error diagnostic. For example, there are some problems (*IO 8, IO 11, SX 8,* a *dump*) if the user responds "*HELP*" when the program expects a scalar structure. It would be safer, though slower, to read all the user's responses as text and convert to real values when appropriate, though problems may still arise if the user types non-name-characters.

## MACHINES VERSUS PEOPLE

Performance is a key issue these days. People are now more expensive than computers and "efficient" use of computing time is becoming less important than "efficient" use of people. Also, personal computers are improving at a remarkably fast rate and machines such as the PERQ, the SYMBOLICS "LISP machine" and descendants of the ALTO are almost with us. We need to think in terms of using computing resources to make the person's job easier. It is unfortunate that a lot of the computers we are using today do not cope well with many interactive users using large programs but, from the naive user's point of view, even a slow responding computer environment is vastly superior to the other options he could consider.

## SUMMARY

The Genstat input/output interface is not attractive to a naive user, though this is not particularly surprising, as the provision of a friendly environment for naive users was not a design aim of Genstat. Several things can be done to make Genstat more attractive and more usable: much can be done within Genstat itself but we can only go so far, for the language is cumbersome for this sort of manipulation and inevitably the resulting system cannot give the fast interaction we can expect nowadays. The main advantage of "doing it within Genstat" is *portability*: if Genstat works for you, then this system will work.

There will be two versions of the next ARCUSGEN, one "within Genstat", the other an external but portable Code Generator. We will offer the naive user a wide range of facilities for working with his data; we hope that he will gain confidence and become more adventurous, perhaps starting to write Genstat code for some aspects of his work.

REFERENCES

Alvey, N.G. (1980) Genstat Standard Analysis Forms The Statistics Department, Rothamsted Experimental Station.

Alvey, N.G., Galwey, N.W. and Lane, P.W. (1982) Introduction to Genstat. Academic Press, London.

Byrne, J. (1981) Interactive statistics packages at York. IUSC Workshop on Interactive Statistics Software, July, 1981, (unpublished).

Francis, I. (1981) A taxonomy of statistical software. COMPSTAT 1980: Proceedings in computational statistics. Physica-Verlag.

Payne, C.D. (1981) PSTAT - The thinking man's SPSS/SCSS. IUSC Workshop on Interactive Statistics Software, July, 1981, (unpublished).

Schektman, Y. (1981) Setting up a library using the GLIM and LAPUMS Systems. GLIM Newsletter, 1981 June, 33-38

*Jane Bryan-Jones*
*ARC Unit of Statistics*
*University of Edinburgh*
*Edinburgh EH9 3JZ*
*U.K.*

- o 0 o -

## ABSORPTION - A METHOD FOR FITTING MODELS WITH MANY PARAMETERS

The *ANOVA* directive in Genstat can analyse models which involve a large number of linear parameters. This is done efficiently in terms of computer storage since the sweeping algorithm involves storage of structures only the same size as the variates being analysed. In balanced experiments the length of the variates is not usually such as to cause storage problems.

When data are unbalanced, as is usually the case in surveys, for example, the regression directives must be used instead of *ANOVA*. The regression algorithm requires storage of a sums-of-squares-and-products (SSP) matrix, which has one row and column for each parameter in the model. Very large matrices are therefore produced for large models. Consider the analysis of a hypothetical survey of 200 herds of cattle, where a model might include a "herd" effect together with a treatment effect with, say, 10 levels and a covariate with separate effects for each treatment. There are thus rows for 219 parameters (constant, 1 variate, 199 parameters for herd, 9 for treatment, 9 for non-parallelism) plus one row for the y-variate, so the SSP matrix has 24310 values.

The number of parameters can be effectively reduced when the size of the model is due to one large factor, as in the example. The parameters corresponding to the factor may be excluded from the SSP by "absorbing" with respect to the factor. This method involves partial solution of the normal equations during the formation of the SSP matrix. The absorbed SSP matrix contains the same values as would an unabsorbed matrix after the effect of the factor had been "swept out", except that no values corresponding to that factor are in the absorbed matrix. (This means that no parameter estimates are available for the absorbed factor, so fitted values, for example, cannot be formed.

Absorbing with respect to herd in the above example leaves only 21 rows and columns and so only 231 values in the SSP matrix.

The method of absorption is otherwise known as "within groups analysis". This can be done in Genstat by defining a within groups SSP structure e.g.

```
'READ' Y, HERD, X, TREAT

'DSSP' W $ Y+X*TREAT; HERD

'TERMS' W

'Y' Y

'FIT' X*TREAT
```

At present Genstat allows only one absorbing factor, i.e. analysis within only one grouping of the data. There is also a problem (in releases up to 4.03) if a more complex model is fitted. Sometimes an interaction is needed between a factor in the model and the grouping factor. If such an interaction is specified, Genstat will interpret it as the effect of the grouping factor within the model factor, since the grouping factor is not present in the model formula in the *DSSP* statement. This fault will be corrected in release 4.04.

I am grateful to Professor Clyde Anderson of Michigan State University for bringing this method to my attention.

*P.W. Lane*
*Rothamsted Experimental Station*
*Harpenden*
*Hertfordshire AL5 2JQ*
*U.K.*

# DISPLAYING RESIDUALS

## OUTLINE

Analysis of experiments in which the units are laid out sequentially, either in space or time, often includes examination of residuals. It is convenient to have some form of graphical display of these as well as a listing of the values. Our display has similarities with that of Andrews and Tukey (1973), except that it is printed down the page and we have divided the standardised residuals into levels of width one third (1/3), allowing 16 columns in all.

## METHOD OF USE

In order for the display to be meaningful, data must be in the order in which units occur, for instance by layout in the field or chronologically. The user must supply a set *PSET* of structures to be printed (for example, unit identifier, block and treatment factors and variate identifier); a variate *VFORM* containing the format elements required for printing *PSET*; a variate *VRES* containing residuals saved from an analysis of variance or regression analysis of the variate; scalars *SSRES* and *DFRES* containing respectively the residual sum of squares for the error stratum or deviance sum of squares and the degrees of freedom appropriate to this sum of squares.

The macro may be used with restricted variates.

## OUTPUT

A listing of the members of *PSET* is printed in parallel according to the format of *VFORM* with a listing and display of the normalised residuals printed alongside. A variate *NORM_RES*, of the length of *VRES*, containing the normalised residuals, is returned from the macro. Missing values in this variate are set to zero. The displayed residuals are stored in 2 factors *NEGATIVE* and *POSITIVE*, with level names *NN* and *NP*. These names could obviously be changed if combinations of other symbols are preferred to '_' and '*X*'. The factors *NEGATIVE* and *POSITIVE* could be returned from the macro and manipulated, if a two-way layout were required.

THE MACRO

```
'MACRO'      RDISPLAY   $
'LOCAL'      NN,NP,RM,RP,NEGATIVE,POSITIVE,VF,THREE,ZERO
'NAME'       NN $ 9=XXXXXXXX,_XXXXXXX,__XXXXXX,___XXXXX,____XXXX,
                    _____XXX,______XX,_______X,________
   :         NP $ 9=________,X_______,XX______,XXX_____,XXXX____,
                    XXXXX___,XXXXXX__,XXXXXXX_,XXXXXXXX
'FACTOR'     NEGATIVE   $ NN , VRES
   :         POSITIVE   $ NP , VRES
'VARIATE'    RM   $ 8 = -7,-6,-5,-4,-3,-2,-1,-.00001
   :         RP   $ 8 = 0.00001,1,2,3,4,5,6,7
   :         NORM_RES $ VRES
   :         VF=2(12.4),12,8
'SCALAR'     THREE=3
   :         ZERO=0
'CALCULATE'  RM,RP = RM,RP/THREE
'CALCULATE'  NORM_RES = VRES/SQRT(SSRES/DFRES)
   :         NORM_RES = REPMV(ZERO)
'GROUPS'     NEGATIVE,POSITIVE = LIMITS(NORM_RES;RM,RP)
'PAGE'
'PRINT/FORM=P' PSET,VRES,NORM_RES,NEGATIVE,POSITIVE $ VFORM,VF
'ENDMACRO'
```

*GLOBAL IDENTIFIERS*

| | | | |
|---|---|---|---|
| *PSET* | set | input: | list of structures to be printed |
| *VFORM* | variate | input: | format for printing *PSET* |
| *VRES* | variate | input: | containing residuals |
| *SSRES* | scalar | input: | containing residual sum of squares |
| *DFRES* | scalar | input: | containing residual d.f. |
| *NORM_RES* | variate | output: | containing normalised residuals |

*STORAGE*

The numbers of named and unnamed structures are 10 and 3 respectively.

*RDISPLAY* is a run-time macro

## EXAMPLE

```
'SET'          PSET = UNIT, BLOCK, VARIETY, YIELD
'VARIATE'      VFORM = 3(12), 12.2
'BLOCKS'       BLOCK/PLOT
'TREATMENT'    VARIETY
'ANOVA'        YIELD;  RES = VRES;  OUT = OUTV
'EXTRACT'      OUTV;  BLOCK . PLOT $ SS = SSRES;  DF = DFRES
'USE/R'        RDISPLAY $
```

*EXAMPLE OUTPUT*

| UNIT | BLOCK | VARIETY | YIELD | VRES | NORM RES | NEGATIVEPOSITIVE |
|---|---|---|---|---|---|---|
| 1 | 1 | A | 1.94 | -0.0653 | -0.9726 | XXX |
| 2 | 1 | B | 1.97 | -0.0771 | -1.1483 | XXXX |
| 3 | 1 | C | 2.16 | 0.0981 | 1.4604 | XXXXX |
| 4 | 1 | D | 1.93 | -0.0171 | -0.2542 | X |
| 5 | 1 | E | 1.96 | 0.0309 | 0.4606 | XX |
| 6 | 1 | F | 2.03 | 0.0305 | 0.4541 | XX |
| 7 | 2 | B | 1.07 | 0.0824 | 1.2267 | XXXX |
| 8 | 2 | C | 1.04 | 0.0358 | 0.5333 | XX |
| 9 | 2 | D | 0.85 | -0.0464 | -0.6909 | XXX |
| 10 | 2 | E | * | -0.0000 | -0.0000 | |
| 11 | 2 | F | 0.90 | -0.0449 | -0.6685 | XXX |
| 12 | 2 | A | 0.93 | -0.0269 | -0.4006 | XX |
| 13 | 3 | C | 3.05 | -0.0571 | -0.8497 | XXX |
| 14 | 3 | D | 3.07 | 0.0676 | 1.0073 | XXXX |
| 15 | 3 | E | 2.96 | -0.0222 | -0.3308 | X |
| 16 | 3 | F | 3.00 | -0.0436 | -0.6498 | XX |
| 17 | 3 | A | 3.14 | 0.0843 | 1.2551 | XXXX |
| 18 | 3 | B | 3.06 | -0.0290 | -0.4322 | XX |
| 19 | 4 | D | 4.00 | -0.0042 | -0.0622 | X |
| 20 | 4 | E | 3.98 | -0.0087 | -0.1298 | X |
| 21 | 4 | F | 4.11 | 0.0580 | 0.8643 | XXX |
| 22 | 4 | A | 4.07 | 0.0079 | 0.1180 | X |
| 23 | 4 | B | 4.12 | 0.0238 | 0.3537 | XX |
| 24 | 4 | C | 4.04 | -0.0768 | -1.1440 | XXXX |

## REFERENCE

Andrews, D.F. and Tukey, J.W. (1973) Teletypewriter Plots for Data Analysis can be Fast: 6-line Plots, Including Probability Plots. J.R.S.S.(C), 22, No. 2, 192-202.

*J. Dickson*
*E.A. Hunter*
*ARC Unit of Statistics*
*University of Edinburgh*
*Edinburgh EH9 3JZ*
*U.K.*

## HOW TO PRINT STRUCTURE IDENTIFIERS WITHOUT THEIR VALUES

When writing general programs or macros, it can be useful to print structure identifiers. A structure may be defined outside the macro, and its identifier may be required as a label for some output (other than the structure values, which are automatically labelled with the identifier by *PRINT*).

The following mechanism produces the required label:

*'POINTER' DUMMY = <structure identifier>*

*'PRINT/LABC=1' DUMMY*

This will actually print two blank lines followed by the identifier right-justified in a field of 12 characters. The field width may be altered by a format. The two blank lines may be suppressed by setting the option *FORM=C* in *PRINT* but this produces a blank line after the identifier. No blank lines are produced by:

*'PRINT/FORM=C,VAR=1,LABR=1' DUMMY*

Several identifiers may be printed at once by printing several pointers, each with one value, or by printing one pointer with several values.

Identifiers may be stored in heading structures by the same mechanism, but using *JOIN* rather than *PRINT*. For example,

*'JOIN/VAR=1,LABR=1' H = DUMMY $ 8*

sets up *H* as a heading, containing the identifier right-justified in a field of 8 characters.

*P.W. Lane*
*Rothamsted Experimental Station*
*Harpenden*
*Hertfordshire AL5 2JQ*
*U.K.*

## GENSTAT ERROR NOTICE S2010 AND MANUAL AMENDMENT

The following document contains an amendment to the Genstat notice board and source corrections for an error discovered since Error Notice S1020 was published in the previous Newsletter. Site representatives should bring these amendments to the attention of their computing centres.

```
                ADDITION TO THE NOTICE BOARD
                ----------------------------

R8         ****ERROR****                               16/3/82
'GRAPH'    IN SOME CIRCUMSTANCES COINCIDENCES ARE
           INCORRECTLY INDICATED, PARTICULARLY WHEN :
           IS THE INDICATOR . ALSO, IF PLOTTING SYMBOLS 1...8
           HAVE BEEN SPECIFIED USING A HEADING THEY
           WILL BE MISTAKEN FOR INDICATORS OF COINCIDENCES
           IN PLOTS USING THE DEFAULT SYMBOL AND
           INCREMENTED BY 1

                  (SEE ERRORS DG7 - DG9)


***** GENSTAT ERROR NOTICE NO. 10          16.3.82
**
**
***** ERROR DG7
***** MODULE DGA SUBROUTINE GRAFF
**
***** 24 LINES BEFORE LABEL 1
**
C     ---------------------------------------------------------------
C
*     IF(NCR.LT.22+NBV(3))  GO TO 1009
      N=PCI
      NOPT=TABLE(N)
**
*****  8 LINES BEFORE LABEL 9
**
      IF(NFM.EQ.0)  GO TO 9
      IJK=TABLE(NFOR+1)
*     IF(IJK.NE.ISMV)  GO TO 4
*     NFM=0
*     CALL DGM(17)
*     GO TO 9
*   4 IF(GETATT(3,IJK).NE.0)  GO TO 1000
      IF(TYPE(3).NE.5)  GO TO 1181
      IF(VALOR(3).EQ.0)  GO TO 1014
**
***** 1 LINE BELOW LABEL 49
```

```
**
   26 CONTINUE
   49 CALL SCFUNC(IAD,NOR,KWOR,67,KVAL,MVP(40))
*     IF(RDATA(IAD).EQ.RMV)  GO TO 91
*     IF(INTERP(J).EQ.0) RDATA(IAD)=RDATA(IAD)-RDATA(IAD)/50.0
      IF(RDATA(IAD).LT.ZMIN)  ZMIN=RDATA(IAD)
*  91 CALL SCFUNC(IAD,NOR,KWOR,68,KVAL,MVP(40))
*     IF(INTERP(J).EQ.0) RDATA(IAD)=RDATA(IAD)+RDATA(IAD)/50.0
      IF(RDATA(IAD).GT.ZMAX)  ZMAX=RDATA(IAD)
      IF(NVZ.EQ.1)  GO TO 42
**
***** 1 LINE BELOW LABEL 60

**
C
   60 DO 71  I=1,NVZ
*     ICPNT(1)=1
      IXO=I+NVZ
      IVLRX=IWOR
**
***** 5 LINES BELOW LABEL 60
**
      IVLRY=IWOR
      IVLFAC=IWOR
*     IF(IVAL(I).EQ.IVAL(IXO))  GO TO 86
*     IF(GETATT(3,VCN(I)).NE.0)  GO TO 1000
*     GO TO 1183
*  86 IF (LFORM(1).EQ.2)  GO TO 73
      IF(STV(I).NE.0)  GO TO 72
      STV(I)=STV(IXO)
**
***** 4 LINES ABOVE LABEL 68
**
      CALL PBYT(IDATA(ISAD),K3,KCH(47))
      CALL PBYT(IDATA(ISAD),K4,KCH(41))
*     IF (LFORM(I).EQ.3)  GO TO 88
*     ICPNT(1)=2
*     GO TO 61
*  88 CALL PBYT(IDATA(ISAD),K4,KCH(55))
      LFORM(I)=4
      GO TO 69
**
***** 1 LINE BELOW LABEL 77
**
      IVLFAC=IVLRY
   77 IF((IVAL(I).GT.JVAL(I)).AND.(LFORM(I).NE.3))  GO TO 1183
*     IF((IABS(LFORM(I)+2).EQ.1).AND.(IVAL(I).GT.360))  CPLIST=.FALSE.
      IF(GWSP(KVAL,3).NE.0)  GO TO 1000
      ISAD=WSPOR+1
**
***** 4 LINES ABOVE LABEL 79
**
      IF(ISDATA(M).NE.ISMV)  GO TO 79
      CALL PBYT(IDATA(WSPOR),K4,KCH(41))
*     IF(LFORM(I).EQ.3)  GO TO 87
*     ICPNT(1)=2
*     GO TO 67
*  87 CALL PBYT(IDATA(WSPOR),K4,KCH(55))
      LFORM(I)=4
      GO TO 67
```

```
**
***** 3 LINES BELOW LABEL 18
**
      CALL LPOUT(SCL(1),IDATA(NN),ISDATA(IHX),IDATA(IHY),ISDATA(IHH),
     1  NHX,NHY,NHH,KDP,MAXN,PFORM,DEV)
*     XORIG=XORIG+NCL+30
*     IF(NLP+6.GT.YMAX)  YMAX=NLP+6
   17 CONTINUE
      IF(.NOT.CPLIST)  GO TO 100
**
***** 1 LINE ABOVE LABEL 1000
**
 1181 DIAG=181

      GO TO 1000
C        SP-9 INADEQUATE PAGE WIDTH
*1009 DIAG=9
 1000 CALL DIAGUP(SPNAME)
      RETURN
**
**
***** END OF ERROR DG7
**
**
***** ERROR DG8
***** MODULE DGB SUBROUTINE LPOUT
**
***** 1 LINE BELOW LABEL 4
**
      XAX(M)=SCL(INK+3)
    4 CONTINUE
*     NCLL=NCL+COLADD
*     NCHR=NCLL/NBV(3)
*     NCHL=NACR*NCHR
*     IF(COLADD.GE.6)  NCLL=NCLL-COLADD+5+IDBLGD
      KA=(NLP-NHY)/2
      KLM=NHY/2
**
**
***** END OF ERROR DG8
**
**
***** ERROR DG9
***** MODULE DGB SUBROUTINE LPPLOT
**
***** 11 LINES BELOW LABEL 80
**
      CALL GBYT(ISYMB(KNT),K3,KSYM(2)
      IF(KSYM(2).NE.KCH(47))  ICNT=2
*     IF(ICPNT(1).EQ.2)  GO TO 190
      IF(IFORM.LT.-2)  GO TO 190
*     IF((IFORM.GT.1).OR.(KA.EQ.KCH(47)))  GO TO 190
      IF(KB.NE.KCH(47))  GO TO 90
      KB=KA
**
***** 2 LINES BELOW LABEL 90
```

```
**
   90 CALL PBYT(ISYM,K3,KA)
      CALL PBYT(ISYM,K4,KB)
*     IF(ICPNT(1).NE.1)  GO TO 120
*      IF(ISYM.EQ.KOLON)  GO TO 300
*     KSYM(1)=KCH(56)
*     IF(.NOT.PRIME)  GO TO 190
      KSYM(1)=KCH(47)
      KSYM(2)=KCH(56)
**
***** 3 LINES BELOW LABEL 200
**
      IX=IX-1
      IPOINT=(IY+2)*NCHL+IX+INC+1
*     IF(ICPNT(1).NE.2)  GO TO 230
      IF(KA.EQ.KCH(47))  GO TO 230
      IF(KA.NE.KCH(10)  GO TO 210
**

**
***** END OF ERROR DG9
**
**
**
***** END OF ERROR NOTICE NO. 10
```

MANUAL AMENDMENT FOR GENSTAT 4.03

In Genstat release 4.03, the *RECYCLE* option of the X-set directives should be set to the identifier of a variate of fitted values.

Please amend Pt II, Section 7.7 of the manual accordingly.

(A formal update will be issued with release 4.04.)

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

# - NOTICES -

## GENSTAT COURSES

The following Genstat courses will be held over the next six months.

| | | |
|---|---|---|
| May 17 - 21 | Rome University | Introductory Course, plus non-linear models and multivariate analysis |
| June 21 - 24 | Copenhagen University | Time Series Analysis |
| July (date open) | New York State Agricultural Experiment Station | Introductory Course |

These courses may have space available for outside participants. Enquires should be made to:

*P.W. Lane*
*Rothamsted Experimental Station*
*Harpenden*
*Hertfordshire AL5 2JQ*
*U.K.*

## THIRD AMERICAN TIME SERIES MEETING (EIGHTH ITSM): 1982

This relatively low-cost conference to be held in Cincinnati, Ohio from the 19th to the 21st of August, 1982 will feature both invited and contributed papers, the objects being to discuss recent developments in the theory and practice of Time Series Analysis and Forecasting and to bring together practitioners from diverse parent disciplines, work environments and geographical locations.

The conference will convene directly after the 1982 ASA annual summer Meetings and use the same headquarters hotel, Stouffer's Cincinnati Towers, at very competitive rates. Thus, interested participants will be able to attend both events conveniently and economically in a single trip, without changing accommodation. Major time series themes will be: Statistical Methodology; Applications to Economics and in Econometrics; Government, Business and Industrial Examples; the Hydrosciences, such as Limnology, Hydrology, Water Quality Regulation and Control and the Modelling of Marine Environments; the Geosciences, especially such areas as Oil Exploration and Seismology; Civil Engineering and allied disciplines; Spatial and Space-Time Processes - their theory and application - especially in Geography and related areas such as city planning or energy demand forecasting; Biology and Ecology; Medical Applications and Biomedical Engineering; Irregularly Spaced Data (including Outliers and Missing Observations); Robust Methods; Seasonal Modelling and Adjustment, Calendar Effects; Causality; Bayesian Approaches; Box-Jenkins Univariate ARIMA, Transfer-function, Intervention and Multivariate Modelling;

State Space; Nonlinear Modelling; Estimation; Diagnostic Checking; Signal Extraction; Comparative Studies; Spectral Analysis, especially for the Physical Sciences; Business Cycle and Expectations Data; Data Revisions; Computer Software and Numerical Analysis; Forecasting and, no doubt, many other areas of the subject.

A fine programme of invited speakers is planned for this event; 150-word abstracts and offers to act as session chairmen should be sent to Oliver Anderson as soon as possible. As usual, the Proceedings will be published. Further information from:

*O.D. Anderson*
*ITSM Cincinnati*
*9 Ingham Grove*
*Lenton Gardens*
*Nottingham NG7 2LQ*
*U.K.*

*Telephone:* *National* *0602 708085*
*International* *+44602 708085*

## 1983 NORTH AMERICAN TIME SERIES MEETINGS IN TORONTO: INITIAL NOTICE

(1) 11-14 August, Special Topics ITSM on Hydrological, Geophysical and Spatial Time Series.

(2) 18-21 August, General Interest ITSM.

This pair of International Time Series Meetings is being planned as satellites to the 15-18 August, ASA (American Statistical Association) joint annual summer Meetings - all three events being held at the exciting Sheraton Centre in downtown Toronto.

The ITSMs are organised independently of ASA by the undersigned, from whom details may be obtained.

*O.D. Anderson*
*9 Ingham Grove*
*Lenton Gardens*
*Nottingham NG7 2LQ*
*U.K.*

## NAG USERS ASSOCIATION

### 1982 MEETING

This year's meeting of the NAG Users Association will be held in Oxford from Wednesday, 29th September to Friday, 1st October. Accommodation will again be in St. John's College and the technical sessions in the Mathematical Institute.

A programme has been arranged to cover technical interests and services offered by NAG to users. This will include:

- Library Development
- User experiences with NAG products
- GLIM, GENSTAT and TSA
- Finite Element Library
- Workshops dealing in depth with particular parts of the NAG Library
- Open Forum

There will also be a general meeting of the Users Association, which will consider the constitution and hold elections to the committee.

The conference fee of £90 (£105 for non-members) includes accommodation in a single room in St. John's College for three nights (28th, 29th, 30th September) and all meals including a conference dinner.

Invitations to the meeting have been sent to all sites which hold a licence for the NAG Library, GLIM or GENSTAT. If you are interested in attending the meeting and have not received an invitation please contact Janet Bentley or Christine Macdougall at NAG Central Office.

The closing date for applications is Friday, 3rd September 1982.

*Janet Bentley*
*NAG Central Office*
*Mayfield House*
*256 Banbury Road*
*OXFORD OX2 7DE*
*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
Mayfield House
256 Banbury Road
OXFORD
OX2 7DE
U.K.

(Each Genstat site representative *automatically* receives one copy, free of charge.)

Please note that each subscription to the Newsletter costs £2.50 per annum (2 issues). This price includes 2nd class/surface postage. Postage at other rates will be charged at cost.

Back issues of the Newsletter are available on microfiche (24X). The first contains issues 1 - 6 and and each subsequent fiche contains two issues - 7/8, 9/10 etc. Each fiche costs the same as a year's subscription to the Newsletter (£2.50).

---

To: NAG Ltd., 256 Banbury Road, OXFORD OX2 7DE, U.K.

Please supply me with .... copies of each future issue of the Genstat Newsletter †{for .... years / until further notice} beginning with issue number .... .

Please supply me with .... microfiche of each of the following issues .......... .

* ☐ 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.

† *delete one alternative*
* *tick one box*

*NP140/136:Mar82*