Issue No. 11

/ R.W. Payne

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

This issue of the Genstat Newsletter appears, experimentally, in typeset form. In the past, a major factor slowing the production of the Newsletter has been the difficulty of generating multi-font output on word processors: we hope that the new approach may overcome this source of delay. Readers' comments on the format of this issue would be welcomed, to help determine our future strategy.

It is the editors' intention to devote a future issue of the Newsletter largely to the Third Genstat Conference, which takes place in Orsay, on the outskirts of Paris, in October (for details, see p.73). It would help considerably if authors of conference papers could prepare these in a form suitable for publication.

We look forward to meeting as many Genstat users as possible at the conference.

STATUS REPORT

Genstat 4.04 was released to ARC System 4 users in February. A large number of errors was soon revealed and this version was replaced by a much improved 4.04A, six weeks later. This version has been transferred to the new VAX 780 at Rothamsted and a few more errors discovered.

The source code of 4.04 will be sent out to convertors in July, with distribution of the VAX version at about the same time. Meanwhile, the Prime and IBM conversions are in progress at Rothamsted.

The number of permanently licenced sites continues to grow at about five per month; by the end of June the total number of such sites was 220. In addition to this, there is a growing trend for existing sites to order further copies of Genstat.

Lists of all known 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
- APPLIED COMPUTING NEDLANDS W AUSTRALIA
- AUSTRALIAN DEP HEALTH CANBERRA
- CSIRO AUSTRALIA (COMP RESEARCH)
- CSIRO (COTTON RES UT) NARRABRI NSW AUSTRALIA
- CSR LIMITED SYDNEY AUSTRALIA
- INST FAMILY STUDIES MELBOURNE AUSTRALIA
- MITCHELL COLL/AGRISEARCH BATHURST NSW AUSTRALIA
- NSW DEPT AGRIC HAYMARKET AUSTRALIA
- S AUSTRALIA DEPT AGRIC ADELAIDE
- S AUSTRALIA INS TECH (COM CEN) POURAKA
- SEQEB BRISBANE AUSTRALIA
- TASMANIAN TREASURY
- UNCLE BEN'S OF AUSTRALIA WODONGA VICTORIA
- U ADELAIDE (WAITE INS) S. AUSTRALIA
- U ADELAIDE S AUSTRALIA
- U AUSTRALIA NAT (COM SER CEN) CANBERRA
- U JAMES COOK (COM CEN) N. QUEENSLAND AUS
- U MACQUARIE (COMP CENT) NORTH RYDE AUSTRALIA
- U NEW ENGLAND (COM CEN) ARMIDALE AUS
- U QUEENSLAND (PRENTICE COM CEN) ST. LUCIA 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 BRASILIA BRAZIL

CANADA
- AGRICULTURE CANADA CE FARM OTTAWA
- ONTARIO INST STUDIES EDUC (CANADA)
- PACIFIC BIOLOGICAL STN NANAIMO
- U MCGILL (MATHS)
- U TORONTO (FAC DENTISTRY) CANADA

CHILE
- U CHILE (AGRON) SANTIAGO DE CHILE

COLOMBIA
- CIAT CALI COLUMBIA

DENMARK
- DANMARKS GEOLOGISKE UNDERSOEGELSE KOEBENHAVN
- ELSAM FREDERICIA DENMARK
- KOEBENHAVN SC ECON BUS ADMIN DENMARK
- NEUCC LYNGBY (EDB-CENT) DENMARK
- NOVO RES INST BAGSVAERD DENMARK
- RECKU KOEBENHAVN DENMARK
- RIGSHOSPITALET KOEBENHAVN DENMARK
- U AALBORG (COMP CENT) DENMARK
- U AARHUS (RECAU)
- U ODENSE (DATACENTER) DENMARK

EIRE
- AGRICULTURAL INST DUBLIN
- GUINNESS DUBLIN EIRE
- U COLL CORK (COMP BUREAU) EIRE

FRANCE
- CENT AIR ST-CYR FRANCE

CENT INTERNAT RECH DERMATOL VALBONNE FRANCE
CENT NAT RES ZOOTEC JOUY-EN-JOSAS FRANCE
CESA JOUY-EN-JOSAS FRANCE
INSERM VILLEJUIF FRANCE
U PARIS SUD (MATH) FRANCE
U PAUL SABATIER TOULOUSE
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 DUISBURG (HSRZ) W. GERMANY
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
CINECA BOLOGNA ITALY
IATA-CNR FIRENZE
ISTIT APPL CALCOLO ROMA ITALY
U FIRENZE (CENT DI CALC) ITALY
U GENOVA (INST DI MAT) ITALY
U NAPOLI (CENTRO DI CALC) ITALY
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

MALAYSIA
U PERTANIAN MALAYSIA, SERDANG

MEXICO
U MEXICO NACIONAL AUTONOMA

THE NETHERLANDS
CIVO-INST TNO ZEIST NETHERLANDS
EUROPEAN SPACE AGENCY
GIST-BROCADES N.V. DELFT NETHERLANDS
INST TNO THE HAGUE NETHERLANDS
ITAL WAGENINGEN NL
LANDINRICHTINGSDIENST UTRECHT NETHERLANDS
PROEFSTATION TUINBOUW ONDER 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 AUCKLAND (COM CEN) NEW ZEALAND
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
IITA IBADAN NIGERIA
U AHMADU BELLO (COMP CEN) NIGERIA

NORWAY
NORSK REGNESENTRAL OSLO
U TROMSO (EDB-SENTRET) NORWAY

PAPUA NEW GUINEA
PAPUA NEW GUINEA NAT COMP CEN WAIGANI

PORTUGAL
LAB NAC ENG & TEC IND. LISBOA PORTUGAL

SENEGAL
INST SENEGALAIS RES AG (OCEANOG) DAKAR

SOUTH AFRICA
PREMIER GROUP JOHANNESBURG S. AFRICA
S AFRICAN DEPT AGRIC (BIOMET & DATAMET) PRETORIA
U FORT HARE 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
C.E.R.S. SYRIA C/O SYSTIME LEEDS
ICARDA ALEPPO SYRIA

TAIWAN R.O.C.
PIG RESEARCH INST TAIWAN

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

UNITED KINGDOM
A. & A.E.E. BOSCOMBE DOWN
AERE (COMP SCI) HARWELL
ASSOCIATED OCTEL
BP INTERNATIONAL LTD LONDON
BRITISH - AMERICAN TOBACCO SOUTHAMPTON
BRITISH SHIP RES ASSN WALLSEND
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
GREATER LONDON COUNCIL (DCTR-GNL'S DEPT)
HMGCC MILTON KEYNES
HOME OFFICE (SRDB) LONDON
HUNTINGDON RESEARCH CENTRE
IMPERIAL CANCER RESEARCH FUND LONDON
IMPERIAL TOBACCO (ITL) BRISTOL
INST NAVAL MEDICINE GOSPORT
INST OCCUPATIONAL MEDICINE EDINBURGH
LEATHERHEAD FOOD RES ASS
LIFE SCIENCE RESEARCH OCCOLD SUFFOLK
MAFF GUILDFORD
MEAT & LIVESTOCK COMMISSION, BLETCHLEY
METAL BOX READING
MET OFFICE BRACKNELL
MIN AG FISH FOOD (ADAS) 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
NATURE CONSERVANCY COUNCIL HUNTINGDON
NAT RADIOLOGICAL PROTECTION BOARD DIDCOT
NEGAS (COMP SERV) LEEDS
NERC COMP SERV BIDSTON BIRKENHEAD
NORTHERN REG HEALTH AUTH NEWCASTLE
NORTH STAFFS HOSPS (COMP DEP) HARTSHILL
NORTH WEST REG HEALTH A MANCHESTER
NORTH WEST WATER (COMP CENT) WARRINGTON
NORTHERN REGIONAL HEALTH AUTH NEWCASTLE
PERKINS ENGINES PETERBOROUGH
PFIZER CENTRAL RESEARCH SANDWICH
POLY BRIGHTON (COMP CENT)
POLY CENT LONDON (COMP UNIT)
POLY CITY LONDON (COM CEN)
POLY COVENTRY (LANCHESTER) (COM CEN)
POLY HUDDERSFIELD
POLY KINGSTON (COMP UNIT)
POLY LIVERPOOL (COMP SER)
POLY NORTH EAST LONDON (COMP CEN)
POLY PLYMOUTH (COMP CENT)
POLY PORTSMOUTH (COMP CENT)
POLY SOUTH BANK (COMP CENTRE) LONDON
POLY THAMES (COMP SERV) LONDON
POLY TRENT (DEP COM SER) NOTTS
RANK XEROX ENG GROUP WELWYN
R.I.B.A. LONDON
ROTHAMSTED ARC COMP CENT HARPENDEN
R.S.P.B. SANDY BEDFORDSHIRE
SCHLUMBERGER CAMBRIDGE RESEARCH
SCOTTISH OFFICE (COMP SERV) EDINBURGH
SEVERN-TRENT WATER AUTHORITY
SHELL RESEARCH CENTRE SITTINGBOURNE
SHELL THORNTON RESEARCH CENTRE CHESTER
SHELL UK LTD LONDON
SMITH KLINE & FRENCH WELWYN GARDEN CITY
THAMES WATER D.O.P. (COMPUTING) READING
UNILEVER RES LAB PORT SUNLIGHT
U.K.A.E.A. WINFRITH DORSET
U ABERDEEN (COMP CEN)
U ASTON (COMP SER)
U BATH (SWRCC)
U BELFAST QUEEN'S (COMP CEN)
U BRADFORD (COMP LAB)

U BRISTOL (SWUCN)
U BRISTOL (COMP CEN)
U CAMBRIDGE (COMP LAB)
U COLL CARDIFF (COMP CEN)
U COLL N WALES (COMP LAB) BANGOR
U EDINBURGH (PLU)
U ESSEX (COMP CEN)
U HULL (COMP CEN)
U LANCASTER (COMP LAB)
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 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 (COMP SERVICES)
U STIRLING (COMP UNIT)
U ST ANDREWS (COMP LAB)
U SURREY (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
WEST MIDLANDS COUNTY COUNCIL (CSIRU)
WEST MIDLANDS GAS

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 CALIFORNIA-RIVERSIDE 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
NSW DEPT AGRIC HAYMARKET AUSTRALIA
S AFRICAN DEPT AGRIC (BIOMET & DATAMET) PRETORIA
U CALIFORNIA-DAVIS (COMP CENT) U.S.A.
U DUISBURG (HSRZ) W. GERMANY
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
CINECA BOLOGNA ITALY
CSIRO AUSTRALIA (COMP RESEARCH)
DOE (TRRL) CROWTHORNE
INST TNO THE HAGUE NETHERLANDS
LANDINRICHTINGSDIENST UTRECHT NETHERLANDS
RIGSHOSPITALET KOEBENHAVN DENMARK
S AUSTRALIA DEPT AGRIC ADELAIDE
S AUSTRALIA INS TECH (COM CEN) POURAKA
U AALBORG (COMP CENT) DENMARK
U AARHUS (RECAU)
U ADELAIDE (WAITE INS) S. AUSTRALIA
U AHMADU BELLO (COMP CEN) NIGERIA
U BERLIN FREE W. GER.
U BOLOGNA ITALY
U BRADFORD (COMP LAB)
U COLORADO STATE FORT COLLINS U.S.A.
ULCC (ADVISORY SERV)
U LONDON IMP COLL (COMP CEN)
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) NLANDS
U W AUSTRALIA (RAINE MED STATS) NEDLANDS

CII IRIS

INSERM VILLEJUIF 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 CITY LONDON (COM CEN)
POLY KINGSTON (COMP UNIT)
POLY LIVERPOOL (COMP SER)
POLY SOUTH BANK (COMP CENTRE) LONDON
POLY TRENT (DEP COM SER) NOTTS
U AUSTRALIA NAT (COM SER CEN) CANBERRA
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 QUEENSLAND (PRENTICE COM CEN) ST. LUCIA AUS
U WAGENINGEN AGRIC (COMP CEN) NL
U YORK (COMP SCI DEPT)

DEC VAX-11

AGRICULTURAL INST DUBLIN
AGRICULTURE CANADA CE FARM OTTAWA
APPLIED COMPUTING NEDLANDS W AUSTRALIA
CAMBRIDGE COLL ARTS & TECH (COM CEN)
C.E.R.S. SYRIA C/O SYSTIME LEEDS
CIVO-INST TNO ZEIST NETHERLANDS
CSIRO (COTTON RES UT) NARRABRI NSW AUSTRALIA
CSR LIMITED SYDNEY AUSTRALIA
DANMARKS GEOLOGISKE UNDERSOEGELSE KOEBENHAVN
DSIR (PHYS & ENG) LOWER HUTT N. ZEALAND
EUROPEAN SPACE AGENCY
GKN TECHNOLOGY LTD WOLVERHAMPTON
HOME OFFICE (SRDB) LONDON
IATA-CNR FIRENZE
ICARDA ALEPPO SYRIA
ICRISAT PATANCHERU AP INDIA
IITA IBADAN NIGERIA
ITAL WAGENINGEN NL
NAT RADIOLOGICAL PROTECTION BOARD DIDCOT
NORSK REGNESENTRAL OSLO
N.Y. STATE DEPT. HEALTH ALBANY N.Y. USA
ONTARIO INST STUDIES EDUC (CANADA)
PACIFIC BIOLOGICAL STN NANAIMO
PFIZER CENTRAL RESEARCH SANDWICH
POLY BRIGHTON (COMP CENT)
RANK XEROX ENG GROUP WELWYN
ROTHAMSTED ARC COMP CENT HARPENDEN
SCHLUMBERGER CAMBRIDGE RESEARCH
SHELL RESEARCH CENTRE SITTINGBOURNE
UNCLE BEN'S OF AUSTRALIA WODONGA VICTORIA
U ADELAIDE S AUSTRALIA
U BELFAST QUEEN'S (COMP CEN)
U CALIFORNIA-DAVIS (COMP CENT) U.S.A.
U CALIFORNIA-RIVERSIDE U.S.A.
U CANTERBURY (LINCOLN COLL) N ZEALAND
U COLL CORK (COMP BUREAU) EIRE
U ICELAND (COMP SER) REYKJAVIK
U LANCASTER (COMP LAB)
U MACQUARIE (COMP CENT) NORTH RYDE AUSTRALIA
U OXFORD (OUCS)
U STIRLING (COMP UNIT)
U ST ANDREWS (COMP LAB)
U SUSSEX (COMP CEN)
UTAC LINAS-MONTLHERY FRANCE
W SUSSEX C.C. (COUNTY SURVEYOR) CHICHESTER
WATER RESEARCH CENTRE MEDMENHAM
WEST MIDLANDS GAS

HARRIS

POLY COVENTRY (LANCHESTER) (COM CEN)
UNILEVER RES LAB PORT SUNLIGHT
U ASTON (COMP SER)
U LONDON CHELSEA COLL (COMP CEN)

HEWLETT PACKARD 3000

GLAXO OPERATIONS UK ULVERSTON

HONEYWELL SERIES 60

ASSOCIATED OCTEL
CENT NAT RES ZOOTEC JOUY-EN-JOSAS FRANCE

NERC COMP SERV BIDSTON BIRKENHEAD
U ABERDEEN (COMP CEN)
U BATH (SWRCC)
U BRISTOL (COMP CEN)
U COLL CARDIFF (COMP CEN)
U LOUGHBOROUGH (COMP CEN)
U PAUL SABATIER TOULOUSE
U PAVIA (CENT CALC)

HONEYWELL SIGMA 6

A. & A.E.E. BOSCOMBE DOWN
HMGCC MILTON KEYNES
MOD(A&AEE) BOSCOMBE DOWN

IBM 360 & SIMILAR

AERE (COMP SCI) HARWELL
AUSTRALIAN DEP HEALTH CANBERRA
BRITISH SHIP RES ASSN WALLSEND
CADBURY SCHWEPPES LTD BIRMINGHAM
CENT AIR ST-CYR FRANCE
CENT INTERNAT RECH DERMATOL VALBONNE FRANCE
CIAT CALI COLUMBIA
EMBRAPA BRASILIA BRAZIL
FUYO TOKYO JAPAN
GREATER LONDON COUNCIL (DCTR-GNL'S DEPT)
IMPERIAL TOBACCO (ITL) BRISTOL
INST SENEGALAIS RES AG (OCEANOG) DAKAR
ISTIT APPL CALCOLO ROMA ITALY
MET OFFICE BRACKNELL
MITCHELL COLL/AGRISEARCH BATHURST NSW AUSTRALIA
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)
PERKINS ENGINES PETERBOROUGH
PIG RESEARCH INST TAIWAN
PREMIER GROUP JOHANNESBURG S. AFRICA
SCOTTISH OFFICE (COMP SERV) EDINBURGH
SEQEB BRISBANE AUSTRALIA
SHELL RESEARCH CENTRE SITTINGBOURNE
SHELL THORNTON RESEARCH CENTRE CHESTER
SMITH KLINE & FRENCH WELWYN GARDEN CITY
U AUCKLAND (COM CEN) NEW ZEALAND
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) CANADA
U UPPSALA (DATACENTER) SWEDEN
U WELLINGTON (COMP CEN) N ZEALAND

ICL SYSTEM 4

ROTHAMSTED ARC COMP CENT HARPENDEN
U BRISTOL (SWUCN)

ICL 1900 & SIMILAR

DES MOWDEN HALL DARLINGTON
DOE BUILDING RES STN GARSTON
HUNTINGDON RESEARCH CENTRE
INST NAVAL MEDICINE GOSPORT
METAL BOX READING
MIN AG FISH FOOD (ADAS) WEYBRIDGE
NORTHERN REG HEALTH AUTH NEWCASTLE
NORTHERN REGIONAL HEALTH AUTH NEWCASTLE
THAMES WATER D.O.P. (COMPUTING) READING
U HULL (COMP CEN)
U READING (COMP UNIT)
U SALFORD (COMP LAB)
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 (COMP CENT) WARRINGTON
NOVO RES INST BAGSVAERD DENMARK
N.Z. STATE SERVICES COMMISSION (CSD)
PAPUA NEW GUINEA NAT COMP CEN WAIGANI
POLY HUDDERSFIELD
POLY PORTSMOUTH (COMP CENT)
SEVERN-TRENT WATER AUTHORITY
U.K.A.E.A. WINFRITH DORSET
U BATH (SWRCC)
U EDINBURGH (PLU)
U HOHENHEIM STUTTGART W GERMANY
U LANCASTER (COMP LAB)
U LONDON QMC (COMP CEN)
U NOTTINGHAM (COMP CEN)
U OXFORD (OUCS)
U SOUTHAMPTON (COMP SERVICES)
WEST MIDLANDS COUNTY COUNCIL (CSIRU)

MODCOMP

INST NUM STAT KOELN W. GERMANY

PERKIN ELMER

LIFE SCIENCE RESEARCH OCCOLD SUFFOLK
ZIMBABWE BIOMETRICS BUREAU, CAUSEWAY
ZIMBABWE SCIENTIFIC COMP CENT HARARE

PR1ME

CESA JOUY-EN-JOSAS FRANCE
CSR LIMITED SYDNEY AUSTRALIA
DOE BUILDING RES STN GARSTON
FORESTRY COMMISSION FARNHAM
GLAXO OPERATIONS UK GREENFORD
GLOUCESTERSHIRE C.A.T. CHELTENHAM
INST FAMILY STUDIES MELBOURNE AUSTRALIA
INST OCCUPATIONAL MEDICINE EDINBURGH
KOEBENHAVN SC ECON BUS ADMIN DENMARK
LEATHERHEAD FOOD RES ASS
LEVIN HORTIC.RES.STN.LEVIN N. ZEALAND
MAFF GUILDFORD
NAPIER COLL COMM & TECH (COM CEN) EDINBURGH
NATURE CONSERVANCY COUNCIL HUNTINGDON
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 ARC COMP CENT HARPENDEN
R.I.B.A. LONDON
R.S.P.B. SANDY BEDFORDSHIRE
SHELL RESEARCH CENTRE SITTINGBOURNE
U FORT HARE SOUTH AFRICA
U MASSEY (COMP CEN) PALMERSTON NORTH N.Z.
U SHEFFIELD (COMP SER)
U SURREY (COMP UNIT)
W AUSTRALIA DEPT AGRICULTURE S. PERTH

SIEMENS BS2000

PROEFSTATION TUINBOUW ONDER 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
ELSAM FREDERICIA DENMARK
RECKU KOEBENHAVN DENMARK
SHELL UK LTD LONDON
TASMANIAN TREASURY
U AUSTRALIA NAT (COM SER CEN) CANBERRA
U KUWAIT (COM SER) ADELIYAH
U LUND (COMP CEN) SWEDEN
U NAPOLI (CENTRO DI CALC) ITALY
U NATAL SA (COMP CEN)
U ODENSE (DATACENTER) DENMARK
U PARIS SUD (MATH) FRANCE
U PERTANIAN MALAYSIA, SERDANG
U ROMA (CENT DI CALC INTERFAC) ITALY
U ZAGREB (COMP CENT) YUGOSLAVIA

OTHER MACHINES

LAB NAC ENG & TEC IND, LISBOA PORTUGAL
MEAT & LIVESTOCK COMMISSION, BLETCHLEY

M.G. Richardson
NAG Central Office
Mayfield House
256 Banbury Road
Oxford OX2 7DE
United Kingdom

H.R. Simpson
Rothamsted Experimental Station
Harpenden
Herfordshire AL5 2JQ
United Kingdom

THE GENSTAT MACRO LIBRARY

My thanks to those who sent comments about existing macros, we hope we have sorted out any problems.

The first issue of the new Library will be made during the lifetime of Genstat 4.04. Macros submitted will be included in this issue if they can pass through the refereeing process by 1 October 1983. However, as we plan to add to the Macro Library at regular intervals, probably once a year, please keep sending macros to me!

Guidelines for submitting and writing macros follow below.

Writing Macros for the Genstat Macro Library

We want the Macro Library to be both useful and used. Not only will the macros be offering a wealth of new and different techniques but they will be saving users' time and energies. For this they will need to be reliable and convenient to use.

We need a consistent programming style for macros within the Library. After some discussion the following is proposed (comments welcome). It incorporates most of the needs for clarity, conciseness and efficiency. Those of you at the Genstat Conference 1981 in Holland will have also heard Ron Baxter propound some of these principles. I recommend Kernighan and Plauger's, 'The Elements of Programming Style', Second Edition, (McGraw–Hill, 1978); in particular their Summary of Rules is well worth looking at.

Programming Style for Genstat Macros

1. declare all local structures (even variates);
2. devalue local structures (except single valued ones);
3. remove restrictions on input-only globals imposed within the macro;
4. don't clutter the working part of the macro with declarations or devalues;
5. minimise the number of statements;
6. use other library macros where appropriate;
7. make input easy: e.g. few input structures, SETs, optional input;
8. if options are not set, or are unrecognisable, do something sensible;
9. have acceptable ouput: e.g. optional printed output; consider optional output structures;
10. be aware of special structures that are preserved or restored on exit depending on the option used at invocation [II 2;6.1].
11. comment usefully but not excessively.

Submitting Macros for the Genstat Macro Library

The following notes are based on the recommendations for contributions to the Algorithms Section of 'Applied Statistics'.

Submission

Macros should be sent to the Editor of the Macro Library and should consist of:

1. the description (for the Macro Manual);
2. the macro code;
3. a driver program for the use of the referee: this should be well commented and should call the macro and produce checkable output;
4. an example program to be included in the Genstat Example Files as an illustration of the use of the macro;
5. any additional information to aid referees.

Items 1 to 4 should be in machine readable form with a corresponding computer listing. 'Machine readable' for ARCUS includes punched cards, paper tape, unlabelled magnetic tape (preferred = 1600 bpi, record length 80, blocksize 4000), but not cassettes or floppy discs (although this may change). All magnetic media will be returned, a contribution to postage would be appreciated.

Refereeing

Adequate refereeing of a macro can entail much computer testing and authors should attempt to made the referees' task as straightforward as possible.

The Editor would welcome volunteers to join the pool of referees.

Description

For the convenience of authors, referees and users we are introducing standard headings, which should appear if applicable.

Name
Macro name and date.

Author
Author's name and affiliation.

Date
Of description.

Outline
A very brief summary, which will also be included in an information retrieval macro `INFO` in the Macro Library.

Keywords
In the conventional form. A separate list of abbreviated keywords will be used in the information retrieval macro.

Description or **Purpose**
This should enable the potential user to decide whether the macro is likely to serve the purpose. Reference to published work should be made when this will shorten the description. Subheadings are preferred, the following may be helpful: Purpose; Notation; Theory; Numerical Method.

Method of Use
With example.

Structure

Global Identifiers
For each, its *type*, its *use* or classification as input and/or output and a brief *description* of its purpose. If a parameter is listed as input only, the macro must not under any circumstances change its values or any of its attributes, e.g. its restriction.

Local Structures
The numbers of named and unnamed structures (NIDs and NUNNs).

Failure Indications
Types of failure should be listed and described.

Auxiliary Macros
Reference to any Library macros used.

Restrictions
Any restrictions on the use of the macro, e.g. size of problem, inability to deal with missing values or restricted structures.

Accuracy
If relevant.

Related Macros
Where relevant, the possibilities of combining the macro with other Library macros should be commented on.

Additional Comments
Comments on the best way to use the macro may be included here.

References
In the conventional form.

Layout

The appearance of a macro will depend largely upon the author's initial presentation of the code. We rely on authors to conform to a standard layout for code:

use white space for readability (it's generally cheap), indent continuation lines, use spaces to clarify long expressions and sequences;

use full length system words of 5 or fewer letters, otherwise use 4 (or 5) letter contractions;

use keywords in nameable lists and option-words in option lists;

where possible comments should be in upper and lower case; all identifiers should use upper case only.

See also 'Writing Macros for the Genstat Macro Library', above, and the code in Edwards (1981).

Comments

Every macro must have an opening comment which identifies the macro and the author. Other comments should be included where necessary but should be kept concise.

Machine Dependence

All Genstat implementations should provide all the features of the language; however, it would be wise to avoid:

unusual characters within text strings;
the use of more than 2 input channels or 3 files;

tolerance levels, upper bounds such as 10E40.

There will be more information about this in a later Newsletter.

Error Diagnostics

Where possible, macros should detect errors and be protected against misuse. However, where the author recognises the possibility of an error but considers that the normal Genstat fault trapping mechanism is sufficient for the user, this should be explained in the Macro Description. The author should consider using a global scalar FAULT to indicate whether or not the macro has been successful.

Disclaimer

Authors will be given full credit for the macros they have written. If work utilising the macros is published, reference should be made to the author and to Genstat.

Whilst every effort will be made to ensure the accuracy and efficiency of macros in the Macro Library, no liability is assumed by any contributor, the Macro Editor, The Statistics Department at Rothamsted Experimental Station or NAG.

References

Edwards (1981) The Anderson–Darling Test for Normality. Genstat Newsletter, 7, 22–23.

Working Party on Statistical Computing (1979) The Construction and Description of Algorithms. Applied Statistics, **28**, No.3, 311–318.

Jane Bryan–Jones
Editor, Genstat Macro Library
ARC Unit of Statistics
University of Edinburgh
James Clerk Maxwell Building
The King's Buildings
Edinburgh EH9 3JZ

THE ANALYSIS OF EXPERIMENTS WITH REPEATED MEASURES

Experiments in which the same plots or subjects are observed more than once occur in many fields of application. The problems which they pose are more concerned with the statistical assumptions for the analysis than with the specification of the necessary Genstat instructions. In this article, we sketch out some of the problems, giving references for further information, and present listings of Genstat input to analyse examples of designs which can occur.

In deciding upon the appropriate analysis, it is necessary to distinguish between two very different situations which can be illustrated as follows:

(1) In an animal experiment, each animal receives just one diet, whose effect is investigated by observing the animal at regular intervals throughout the experiment, or, in a drug trial, the effect of a stimulus is measured several times following its application.

(2) Several drivers are given different cars to drive on successive evenings, each time using an identical route (with which they are very familiar), in order to assess the petrol consumption of each car.

Superficially, it might seem that both of these situations could be analysed by an analysis of variance with 'BLOCKS'

```
(1) ANIMALS/TIME
```

and (2) DRIVERS/EVENING

(or perhaps, (2) DRIVERS*EVENINGS, if the design has been suitably chosen). However, the assumptions for these analyses require that all contrasts of differences between observations from an individual (animal or driver) have equal variances. Inequality of contrast variances can be due to unequal correlations between observations. The assumption of equal correlations between observations from an individual is perhaps reasonable in (2) but, in (1), there will almost certainly be a greater correlation between the observations at time 1 and time 2 than between those at time 1 and time 3, and so on.

One approach in situation (1) is to analyse the data separately for each time. One has to be careful not to regard the appearance of similar patterns of treatment effects at different times as providing independent confirmation but, otherwise, the analyses will be valid, although unadventurous. They fail to give any information about how the effects vary with the time or how this variation differs from treatment to treatment. An alternative, perhaps complementary, approach (Rowell and Walters (1976); Yates (1982)) is to analyse contrasts of the measurements of each subject over time. Often these will be polynomial contrasts (i.e. the linear increase or decrease with time, the quadratic etc.) but these are not the only contrasts which one might consider. Janson, Keen and Thissen (1981) discuss the use of Principal Components to determine suitable linear contrasts and Yates (1982) discusses the use of coefficients from sigmoid-type curves, like the Gompertz curve. If coefficients from non-linear curves are to be analysed, the form of curve to be used should be carefully chosen to try to decrease the correlation between the fitted parameters. (This also aids the fitting process – see the article by Ross on page 58).

The calculated contrasts of Rowell and Walters do not, of course, represent the best estimates of the polynomial trends (for which a regression program catering for correlated errors would be required). However, in practice, this should not be important. It will certainly not invalidate the subsequent Analysis of Variance. For non-linear curves this aspect should, perhaps, be given more attention. For equally spaced observations, fitting can be done using the FIT MODEL command of the program MLP, also developed at Rothamsted.

Examples JANSEN(1,2) below show the analysis of the data given by Jansen et al.(1981) at the 1981 Genstat Conference, including the first example of Rowell and Walters (1976) (JANSEN(1)). We are grateful to Ir. J. Jansen for supplying details. Examples ROWELL(3,4) show 2 other examples from Rowell and Walters (1976): we thank J.G. Rowell for providing the necessary data.

Situation (2) covers the type of experiment in which one may assume that the observations are sufficiently separated in time to avoid any undue correlation between observations at adjacent times, compared to other observations on the same subject. Provided this assumption is valid (see below) the specification of these experiments poses few problems beyond those faced in ordinary statified experiments, for example, field experiments (see Examples WINER(582,267,525,635,646,731,803,806) below). One source of extra complexity may be the need to check whether there are 'residual treatment effects', i.e. whether the effect of a treatment on a particular individual (animal, plot, etc.) may persist into the subsequent period (during which some other treatment is applied). Designs are available which enable such effects to be estimated; Example AVCCOX(135), from Cochran and Cox (1957) p. 135, is reproduced below. A further problem is that the variances of the observations may not be constant at different times, thus necessitating a weighted analysis.

The analysis of contrasts over time, suggested for situation (1), is feasible only if each subject has received only a single treatment (either the same one at each time, or an initial treatment whose effect is left to develop). If a second treatment had been allocated (at random) to the time periods of each subject, this analysis would no longer be appropriate. Our advice is that, where different

treatments are applied at different times, care should be taken to try to avoid undue correlations between adjacent times, for example by allowing rest periods between experimental periods, so that situation (2) may be assumed to apply.

Of course, the decision about whether or not the extra correlation between adjacent times can be ignored may not always be clear cut. The assumption that is being made is that all constrasts of differences have the same variance, which is true when (but not only when) the covariance matrix of observations at different times has compound symmetry (i.e. it has a value σ^2 on the diagonal and $\rho\sigma^2$ off the diagonal). Box (1950) has suggested a test procedure for checking whether the covariance matrices are homogeneous between groups and whether the pooled covariance matrix has compound symmetry (see also Winer (1971) pp. 594–599). This has been programmed in a macro REPMEAS, (described in the following article) together with the estimation of the Greenhouse and Geiser ε (see below).

If it seems that the matrix does not have compound symmetry, a possible approach is to take the analysis of variance table from the data, using the split-plot model, and use alternative critical values when testing within-subject effects. Greenhouse and Geisser (1959) define ε as a measure of the departure of the contrast variances from equality. ε takes values on the interval (1/(t-1), 1] (where t is the number of occasions on which measurements were taken) and can be used as a multiplicative factor for both degrees of freedom of the critical F–values. Using the lower bound of the range produces a conservative test (see Gill and Hafs (1971)). An estimate of ε based on the sample covariance matrix may be preferred (see Winer (1971) pp. 523–524).

Examples

```
'REFE' JANSEN(1)
''
POLYNOMIAL TREND FITTED TO REPEATED MEASURES WITH COVARIATES
(JANSEN,KEEN AND THISSEN - DOG DATA)
```

Four treatments were applied to 16 dogs in a completely randomised design (4 dogs per treatment). Measurements were taken before the application of treatments and 1, 3 and 5 minutes after application. The measurement taken before treatment was used as a covariate.

```
''
'UNIT' $ 4
'HEAD' HX=''TIME''
 :     H1=''INDIVIDUAL CURVES FOR TREATMENT''
 :     H2=''MEAN CURVES PER TREATMENT''
 :     PLOTTYPE=''L''
 :     SYMBOLS=''1234567890''
'VARI' S(1)=1...4
 :     S(2)=5...8
 :     S(3)=9...12
 :     S(4)=13...16
 :     TIME=0,1,3,5
'READ/S,NUN=Q,PRIN=DE' INDVAR(1...16)
'MACRO' INDCURV $
'GRAPH/ATX=HX,ATY=HY' INDVAR(SS) ; TIME $PLOTTYPE;SYMBOLS
'CALC' TRMEAN=VMEAN(INDVAR(SS))
'ENDM'
'FOR' SS=S(1...4)
;TREATNR=1...4 ; TRMEAN=TRMEAN(1...4)
'JOIN' HY=H1,TREATNR
'USE/R' INDCURV $
'REPE'
'GRAPH/ATX=HX,ATY=H2' TRMEAN(1...4) ; TIME $PLOTTYPE;SYMBOLS
'RUN'
```

```
0.04 0.20 0.10 0.08 0.02 0.06 0.02 0.02 0.07 1.40 0.48 0.24
0.17 0.57 0.35 0.24 0.10 0.09 0.13 0.14 0.12 0.11 0.10 *
0.07 0.07 0.07 0.07 0.05 0.07 0.06 0.07 0.03 0.62 0.31 0.22
0.03 1.05 0.73 0.60 0.07 0.83 1.07 0.80 0.09 3.13 2.06 1.23
0.10 0.09 0.09 0.08 0.08 0.09 0.09 0.10 0.13 0.10 0.12 0.12
0.06 0.05 0.05 0.05
'UNIT' $ 16
'MATR' DATAMAT $ 16,3
'EQUA' DATAMAT=INDVAR(1...16) $ X,3
 :     COVARIATE=INDVAR(1...16) $ 1,3X
'VARI' CONTR(0)=3(1)
 :     CONTR(1)=-1,0,1
 :     CONTR(2)=1,0,1
'CALC' CONTR(1,2)=CONTR(1,2)/SQRT(SUM(CONTR(1,2)**2))
 :     CONTR(0)=CONTR(0)/3
 :     AVERAGE,LINEAR,QUADRATIC=PDT(DATAMAT;CONTR(0,1,2))
'NAME' NDRUG=MORPHINE,TRIMETHA
 :     NHIS=INTACT,DEPLETED
'FACTOR' DRUG $ NDRUG=8(1,2)
 :       HISLEVEL $ NHIS=4(1,2)2
'TREAT' DRUG*HISLEVEL
'COVAR' COVARIATE
'ANOVA/PRX=0,PRYU=0' AVERAGE,LINEAR,QUADRATIC
'RUN'
'CLOSE'


'REFE' JANSEN(2)
''
```

POLYNOMIAL TREND FITTED TO REPEATED MEASURES
(JANSEN,KEEN AND THISSEN - RAT DATA)

Three treatments were applied to 27 rats with unequal replication. Measurements were taken at five equal intervals after application of treatment.

```
''
'UNIT' $ 5
'HEAD' HX=''TIME''
 :     H1=''INDIVIDUAL CURVES FOR TREATMENT''
 :     H2=''MEAN CURVES PER TREATMENT''
 :     PLOTTYPE=''L''
 :     SYMBOLS=''0123456789''
 :     TRNAM(1)=''CONTROL''
 :     TRNAM(2)=''THYROXIN''
 :     TRNAM(3)=''THIORACIL''
'VARI' S(1)=1...10
 :     S(2)=11...17
 :     S(3)=18...27
 :    TIME=1...5
'READ/S,NUN=Q,PRIN=DE' INDVAR(1...27)
'MACRO' INDCURV $
'GRAPH/ATX=HX,ATY=HY' INDVAR(SS) ; TIME $PLOTTYPE;SYMBOLS
'CALC' TRMEAN=VMEAN(INDVAR(SS))
'ENDM'
```

```
'FOR' SS=S(1...3)
;TRNAM=TRNAM(1...3) ; TRMEAN=TRMEAN(1...3)
'JOIN' HY=H1,TRNAM
'USE/R' INDCURV $
'REPE'
'GRAPH/ATX=HX,ATY=H2' TRMEAN(1...3) ; TIME $PLOTTYPE;SYMBOLS
'RUN'
57 86 114 139 172 60 93 123 146 177 52 77 111 144 185 49 67 100 129 164
56 81 104 121 151 46 70 102 131 153 51 71 94 110 141 63 91 112 130 154
49 67 90 112 140 57 82 110 139 169 59 85 121 146 181 54 71 90 110 138
56 75 108 151 189 59 85 116 148 177 57 72 97 120 144 52 73 97 116 140
52 70 105 138 171 61 86 109 120 129 59 80 101 111 122 53 79 100 106 133
59 88 100 111 122 51 75 101 123 140 51 75 92 100 119 56 78 95 103 108
58 69 93 116 140 46 61 78 90 107 53 72 89 104 122
'UNIT' $ 27
'MATR' DATAMAT $ 27,5
'EQUA' DATAMAT=INDVAR(1...27)
'VARI' CONTR(1)=-2,-1,0,1,2
 :     CONTR(2)=2,-1,-2,-1,2
 :     CONTR(3)=-1,2,0,-2,1
'CALC' CONTR(1,2,3)=CONTR(1,2,3)/SQRT(SUM(CONTR(1,2,3)**2))
 :     LINEAR,QUADRATIC,CUBIC=PDT(DATAMAT;CONTR(1,2,3))
'NAME' NDRUG=CONTROL,THYROXIN,THIORACIL
'FACTOR' DRUG $ NDRUG=10(1),7(2),10(3)
'TREAT' DRUG
'ANOVA' LINEAR,QUADRATIC,CUBIC
'RUN'
'CLOSE'


'REFE' ROWELL(3)
''
POLYNOMIAL TREND FITTED TO REPEATED MEASURES ON A RANDOMISED BLOCK DESIGN
(ROWELL AND WALTERS - EXPERIMENT 3)
```

Five treatments were applied to 20 plots of 20 chickens arranged in 4 blocks. Weights of the chickens were assessed at five weekly intervals.

```
''
'UNIT' $ 100
'FACTOR' TIME $ 5=20(1...5)
 :      TREAT $ 5=4(1...5)5
 :       BLOCK $ 4=(1...4)25
 :       CHICKEN $ 20=(1...20)5
'READ/PRIN=DE' WEIGHT
'TREAT' BLOCK+TREAT+CHICKEN.BLOCK.TREAT+POL(TIME,4)/(TREAT+BLOCK)
      +CHICKEN.BLOCK.TREAT.POL(TIME,4)
'ANOVA/LIMA=999,SE=N' WEIGHT
'RUN'
```

```
50.55 50.45 50.35 50.35 50.70 50.45 50.45 50.75 50.55 50.55
50.60 50.50 50.55 50.30 50.45 50.40 50.40 50.35 50.35 50.45
102.55 104.75 104.80 107.10 110.00 111.10 110.30 112.55 102.60 103.95
103.10 106.05 102.05 95.40 105.25 105.80 100.75 100.35 101.90 101.35
181.4 185.0 182.5 191.1 191.3 195.0 193.8 198.7 168.9 172.7 168.6
176.7 167.1 158.8 174.6 167.1 173.0 170.2 173.3 170.1
290.7 295.6 287.7 310.2 300.1 298.8 303.5 312.7 267.7 273.7
262.7 282.2 264.1 251.8 271.6 260.1 276.2 267.5 277.7 266.7
433.1 434.1 422.8 461.0 437.8 423.9 428.5 458.5 392.3 406.2
389.3 422.9 390.7 392.7 393.5 391.1 414.6 397.0 416.7 401.7
'EOD'
'CLOSE'


'REFE' ROWELL(4)
''
POLYNOMIAL TREND FITTED TO UNEQUALLY SPACED REPEATED MEASURES
(ROWELL AND WALTERS - EXPERIMENT 4)
```

Five treatments were applied to four blocks of 20 plots, each plot consisting of 250 pineapple plants. The number of flowers per fruit were assessed 4, 9, 12, 14 and 21 days after application of the treatment.

```
''
'UNIT' $ 100
'FACTOR' TIME $ 5=20(1...5)
 :       TREAT $ 5=4(1...5)5
 :        BLOCK $ 4=(1...4)25
 :        PLANT $ 20=(1...20)5
'VARI' DAYS=4,9,12,14,21
'READ/PRIN=DE' FLOWERS
'TREAT' BLOCK+TREAT+PLANT.BLOCK.TREAT+POL(TIME,4,DAYS)/(TREAT+BLOCK)
      +PLANT.BLOCK.TREAT.POL(TIME,4,DAYS)
'ANOVA/LIMA=999,SE=N' FLOWERS
'RUN'
0.803 0.247 0.844 0.719 0.456 0.342 0.514 0.331 0.397 0.304 0.406 0.319
0.401 0.198 0.081 0.246 0.088 0.067 0.229 0.152 5.388 2.871 3.911 3.222
3.548 2.685 3.283 2.714 2.985 3.500 2.465 3.412 2.807 2.653 2.153 2.383
1.318 1.876 2.442 1.925 6.356 4.185 5.577 5.528 5.300 4.883 4.849 4.576
4.934 4.993 4.970 4.817 4.695 4.090 3.224 3.943 1.915 3.746 3.264 3.049
5.877 5.525 5.955 5.668 5.296 5.701 5.155 4.714 5.397 5.601 5.504 4.728
5.411 5.941 4.704 5.023 2.833 5.246 5.729 3.561 4.143 4.355 3.844 5.006
3.440 4.552 4.339 3.647 4.000 4.601 3.774 3.599 2.852 3.513 4.612 3.957
2.857 4.716 3.793 2.466
'EOD'
'CLOSE'
```

```
'REFERENCE' WINER(582)
''
ESTIMATION OF LINEAR,QUADRATIC AND CUBIC TRENDS IN A REPEATED MEASURES DESIGN
(WINER - PAGE 582)
```

This is a learning trial in which factors A and B are randomised onto subjects and factor C represents 4 equally spaced trials. We analyse the linear, quadratic and cubic contrasts of C. Two complementary methods of analysis are shown.

The first method of analysis gives a full factorial breakdown of the sums of squares, ignoring the block model.

```
''
'UNIT' $ 48
'FACTOR' SUBJECTS $ 12=4(1...12)
 :       A $ 2=24(1,2)
 :       B $ 2=12(1,2)2
 :       C $ 4=(1...4)12
'READ/PRIN=DE' DATA
'PAGE'
'CAPT' ''
VARIANCE RATIOS IN THE ANALYSIS OF VARIANCE TABLE NEED TO BE CALCULATED BY
DIVIDING THE MEAN SQUARE ERROR FOR THE TREATMENT TERM BY THE MEAN SQUARE ERROR
FOR THE NEXT INTERACTION WITH SUBJECTS ( OR THE CORRESPONDING CONTRAST OF
THAT INTERACTION)
EG FOR LINEAR TREND IN C. V.R. = 570.4/2.575 = 221.5
''
'TREATMENTS' A*B+A.B.SUBJECTS+POL(C,3)/(A*B)+A.B.SUBJECTS.POL(C,3)
'ANOVA/SE=N,LIMA=499,PR=613' DATA
'RUN'
1 6 5 7 0 6 7 9 3 8 8 9 2 7 12 15
1 6 8 9  3 7 10 11 1 2 7 12 1 1 4 10
1 1 4 8  2 2 8 12 3 2 10 15 2 2 7 13
'EOD'
''
```

The second method of analysis generates variates corresponding to the orthogonal polynomials and analyses each variate separately. This does not give the between-subjects analysis, nor the overall linear, quadratic and cubic effects of C. Each analysis gives the interactions between A, B and A.B and a particular contrast (LIN, QUAD, CUBIC) of C.

```
''
'UNIT' $ 12
'FACTOR' SUBJECTS $ 12=(1...12)
 :       A $ 2=6(1,2)
 :       B $ 2=3(1,2)2
'VARI' TIME=1...4
'MATR' DATAMAT $ 12,4
'EQUA' DATAMAT=DATA
'MACRO' ORTHPOL $
''MACRO FOR CALCULATING ORTHOGONAL POLYNOMIALS''
'LOCAL' SO,S1,PO,P1,P2,I,LIN ,ONE
'SCAL' SO,S1  :ONE=1
  'FOR' P1=XVAR 'VARI' PSET$P1 'REPE'
  'ASSI' LIN=PSET$ONE : PO= ONE$ONE : P1=LIN$ONE
  'CALC' LIN=XVAR-MEAN(XVAR) : SO=NVAL(XVAR) : S1=SUM(P1*P1)
  'FOR' I=2...DEGREE
    'ASSI' P2=PSET$I
```

```
    'CALC' P2=LIN*P1
    : P2=P2-(SUM(P1*P2)/S1)*P1 -(S1/S0)*P0
    : S0=S1 : S1=SUM(P2*P2)
    'ASSI' P0=P1$ONE : P1=P2$ONE
  'REPE'
'ENDM'
'VARI' CONTR(1...3) $ 4
'SCAL' DEGREE=3
'SET' PSET=CONTR(1...3)
 :    XVAR=TIME
'USE' ORTHPOL $
'CALC' CONTR(1...3)=CONTR(1...3)/SQRT(SUM(CONTR(1...3)**2))
'CALC' LINEAR,QUADRATIC,CUBIC=PDT(DATAMAT ; CONTR(1...3))
'BLOCK' SUBJECTS
'TREAT' A*B
'ANOVA' LINEAR,QUADRATIC,CUBIC
'RUN'
'CLOSE'

'REFERENCE' WINER(267)
''
```

ONE FACTOR REPEATED MEASURES DESIGN
(WINER - PAGE 267)

The aim of the experiment was to study the effects of four drugs on the reaction time to a series of standardised tasks. There were five subjects (PEOPLE) and four drugs. Each subject was observed under each drug, the order of application of the drugs being randomised independently within each subject and a sufficient time was allowed between the administration of each drug to avoid any carry-over effects.

(This is analagous to a randomised block design, with PEOPLE forming blocks.)

```
'UNIT' $ 20
'FACTOR' DRUGS $ 4=(1...4)5
 :       PEOPLE $ 5=4(1...5)
'READ/PRIN=DE' DATA
'BLOCK' PEOPLE
'TREATMENT' DRUGS
'ANOVA' DATA
'RUN'
30 28 16 34
14 18 10 22
24 20 18 30
38 34 20 44
26 28 14 30
'EOD'
'CLOSE'

'REFERENCE' WINER(525)
''
```

TWO FACTOR DESIGN WITH REPEATED MEASURES ON ONE FACTOR.
(WINER - PAGE 525)

In an experiment to study the effects of two methods of calibration and four shapes of dial on the accuracy of recording three subjects were allocated (at random) to dials calibrated by each of the two methods. Each subject is observed with all four shapes of dial, the order of observations being randomised independently for each subject.

```
''
'UNITS' $ 24
'FACTOR' SUBJECTS $ 6=4(1...6)
 :       CALIBRATION $ 2=12(1,2)
 :       SHAPE $ 4=(1...4)6
'READ/PRIN=DE' ACCURACY
'BLOCK' SUBJECTS/SHAPE
'TREATMENTS' CALIBRATION*SHAPE
'ANOVA' ACCURACY
'RUN'
0 0 5 3
3 1 5 4
4 3 6 2
4 2 7 8
5 4 6 6
7 5 8 9
'EOD'
'CLOSE'
'REFE' WINER(635)
''
```

CONFOUNDED 2X2X2 FACTORIAL DESIGN
(WINER - PAGE 636)

This experiment evaluates preference for advertisements of different sizes, style of typeface, and colour (all factors at 2 levels). Each of the subjects examined four out of the eight treatment combinations, the SIZE.TYPEFACE.COLOUR interaction being confounded between subjects.

```
''
'UNIT'  $ 24
'FACTOR' SIZE $ 2=(2,1,1,2)3,(1,2,2,1)3
 :       TYPEFACE $ 2=(1,2)12
 :       COLOUR $ 2=2(1,2)6
 :       SUBJECTS $ 6=4(1...6)
'READ/PRIN=DE' DATA
'BLOCK' SUBJECTS
'TREAT' SIZE*TYPEFACE*COLOUR
'ANOVA' DATA
'RUN'
16 8 2 8
10 4 3 7
9 3 0 5
10 12 8 3
11 16 10 5
4 7 7 2
'EOD'
'CLOSE'

'REFE' WINER(646)
''
```

PARTIALLY CONFOUNDED 3X3 FACTORIAL DESIGN
(WINER - PAGE 646)

12 subjects (divided into 6 blocks of 2) were each observed under 3 treatment combinations. There were 2 treatment factors, A and B, each with 3 levels. In blocks 1–3 the interaction contrasts A.(B**1) are confounded, while in 4–6 A.(N**2) is confounded.

```
''
'UNIT' $ 36
'FACTOR' A $ 3=(1,2,3)12
 :       B $ 3=(1,3,2)2,(2,1,3)2,(3,2,1)2,(1,2,3)2,(3,1,2)2,(2,3,1)2
 :       BLOCK $ 6=6(1...6)
 :       SUBJECT $ 12=3(1...12)
'READ/PRIN=DE' DATA
'BLOCK' BLOCK/SUBJECT
'TREAT' A*B
'ANOVA' DATA
'RUN'
14 7 15 6 3 5 3 4 10 7 6 30
5 15 7 5 5 3 10 10 15 15 20 25
3 5 12 7 10 18 6 7 5 4 3 5
'EOD'
'CLOSE'

'REFE' WINER(731)
''
```

CONFOUNDED 3X3X3 FACTORIAL ANALYSED USING PSEUDO FACTORS
(WINER - PAGE 731)

This experiment studies the effectiveness of type of material (factor A), style of printing (factor B), and colour (factor C) on the effectiveness of the design of a package (A, B, C all at 3 levels). There are 18 subjects divided into 9 groups of two, each of which evaluates 3 treatment combinations. Factor C and half of the contrasts of the A.B and A.B.C interactions are confounded with the groups of subjects.

```
''
'UNIT' $ 54
'FACTOR' A $ 3=(1,2,3)18
 :       B $ 3=(2,3,1)6,(1,2,3)6,(3,1,2)6
 :       C$ 3=6(1,2,3)3
 :       GROUPS $ 9=6(1,4,7,2,5,8,3,6,9)
 :       SUBJECT $ 18=3(1...18)
 :       PFAB $ 3
'GENERATE' PFAB ; A.B ; C ; GROUPS
'READ/PRIN=DE' DATA
'BLOCK' SUBJECT.GROUPS
'TREAT' ((A*B)//PFAB)*C
'ANOVA' DATA
'RUN'
2 2 3 1 1 7 5 8 1 9 12 7
5 4 6 7 5 9 5 4 7 8 5 8
8 10 4 10 14 6 10 10 8 7 3 9
3 2 5 6 4 9 8 9 6 10 10 5
12 6 10 8 8 2
'EOD'
'CLOSE'

'REFE' WINER(803)
''
```

2X2 FACTORIAL WITH COVARIATES MEASURED BEFORE EXPERIMENT
(WINER - PAGE 803)

Factor A was applied to subjects, factor B was applied to observations within subjects. Covariate X was measured on each subject before the experiment.

```
''
'UNIT' $ 16
'FACT' A $ 2=8(1,2)
 :     B $ 2=(1,2)8
 :     SUBJECT $ 8=2(1...8)
'READ/P,PRIN=DE' X,Y
'BLOCK' SUBJECT
'TREAT' A*B
'COVA' X
'ANOVA' Y
'RUN'
3 10 3 8
5 15 5 12
8 20 8 14
2 12 2 6
1 15 1 10
8 25 8 20
10 20 10 15
2 15 2 10
'EOD'
'CLOSE'
```

```
'REFE' WINER(806)
''
2X2 FACTORIAL WITH COVARIATES MEASURED BEFORE EACH OBSERVATION
(WINER - PAGE 806)
```

Factor A (3 levels) was applied to the subjects, factor B (2 levels) was applied to observations within subjects. Covariate X was measured before each observation on a subject.

```
''
'UNIT' $ 18
'FACT' A $ 3=6(1...3)
 :     B $ 2=(1,2)9
 :     SUBJECT $ 9=2(1...9)
'READ/P,PRIN=DE' X,Y
'BLOCK' SUBJECT
'TREAT' A*B
'COVA' X
'ANOVA' Y
'RUN'
3 8 4 14 5 11 9 18 11 16 14 22 2 6 1 8 8 12 9 14
10 9 9 107 10 4 10 8 14 10 18 9 15 12 22
'EOD'
'CLOSE'
```

```
'REFERENCE'      AVCCOX(135)
''
```

LATIN SQUARE WITH RESIDUAL EFFECTS
(COCHRAN AND COX - PAGE 135)

The treatments (3 levels) were applied to subjects in a Latin Square design. The analysis tests for the existence of residual effects (factor NORES) and also between the residual effects of the three treatments (RESTREAT).

```
''
'UNITS' $ 18
'FACTORS' SQUARE,NORES $ 2
        : PERIOD,SEQUENCE,TREAT $ 3
        : RESTREAT $ 4
'INTEGER' Z=1,2,-3,-4
'GENERATE' SQUARE,SEQUENCE,PERIOD
'VALUES' TREAT=1,2,3, 2,3,1, 3,1,2, 1,3,2, 2,1,3, 3,2,1
  : RESTREAT=4,1,2, 4,2,3, 4,3,1, 4,1,3, 4,2,1, 4,3,2
'GROUPS' NORES=GROUP(RESTREAT;Z)
'VALUES' Y=38,25,15,109,86,39,124,72,27,86,76,46,75,35,34,101,63,1
'BLOCKS' SQUARE/(SEQUENCE*PERIOD)
'TREATMENTS' TREAT+NORES/RESTREAT
'ANOVA' Y
'TREATMENTS' NORES/RESTREAT+TREAT
'ANOVA' Y
'RUN'
'CLOSE'
```

References

Box, G.E.P. (1950) Problems in the Analysis of Growth and Wear Curves. Biometrics, **6**, 362–389.

Cochran, W.G. and Cox, G.M. (1957) Experimental Design (2nd Edition). Wiley, New York.

Gill, J.L. and Hafs, H.D. (1971) Analysis of Repeated Measurements of Animals. Journal of Animal Science, **33**, 331–336.

Greenhouse, S.W. and Geisser, S. (1959) On Methods in the Analysis of Profile Data. Psychometrika, **24**, 95–112.

Jansen, J., Keen, A. and Thissen, J.T.M.N. (1981) Using Genstat in the Analysis of Designed Experiments with Observations Repeated in Time. Instituut TNO voor Wiskunde, Informatieverwerking en Statistick, Wageningen.

Ross, G.J.S. (1983) Fitting General models with OPTIMIZE. Genstat Newsletter **11**, 56–58.

Rowell, J.G. and Walters, D.E. (1976) Analysing Data with Repeated Observations on each Experimental Unit. J.Agric.Sci.Camb., **87**, 423–432.

Winer, B.J. (1971) Statistical Principles in Experimental Design (2nd Edition). McGraw–Hill, New York.

Yates, F. (1982) Reader Reaction: Regression Models for Repeated Measurements. Biometrics, **38**, 850–853.

R.W. Payne
T.J. Dixon
Statistics Department
Rothamsted Experimental Station
Harpenden
Hertfordshire AL5 2JQ
United Kingdom

MACRO REPMEAS

Purpose

Testing Homogeneity and compound symmetry of variance-covariance matrices from Repeated Measures designs.

Background and Description

A repeated-measures design is one in which the subjects (animals, people, plots, etc.) are observed on several different occasions. Usually there is one or more treatment factor randomly allocated to the subjects (i.e. each subject receives a single combination of these factors throughout the experiment); there may also be factors randomised onto the occasions within each subject.

The design may thus seem analogous to a split-plot design, with subjects corresponding to the whole plots (receiving the first set of treatments) and the occasions of observation to the subplots. However, for a split-plot analysis to be valid it must be assumed that the variances of all contrasts of differences between observations on a whole plot are equal. This will be true if the covariance matrices of observations at different times (or residual effects after removing effects randomised onto the occasions) are equal and have compound symmetry (a common value σ^2 on the diagonal, $\rho\sigma^2$ off the diagonal). This macro constructs the covariance matrices of observations (or residuals), checks that they are homogeneous, forms a pooled covariance matrix and tests whether this has compound symmetry using the method of Box (1950) (see also Winer (1971) pp. 594–599). In addition, the correction factor ε suggested by Greenhouse and Geisser (1959) (see also Winer (1971) pp. 523–524) is estimated, for use when compound symmetry is not present.

Method of Use

The user must declare three scalars: NSUBJ, the number of subjects in the experiment; NTIME, the number of occasions on which each subject is observed; and NGRP the number of groups of subjects receiving different treatment combinations. In addition, three factors need to be declared: SUBJECT, with NSUBJ levels, classifying the subjects; TIME, with NTIME levels, classifying the occasions (both of length NTIME*NSUBJ)); and SUBGROUP, with NGRP levels and length NSUBJ, classifying the subjects into groups. A table RESTAB either of resulting observations (which need to be read or tabulated) or of residuals (which need to be extracted from the analysis of variance) must be supplied. The classifying factors of RESTAB are SUBJECT and TIME.

REPMEAS has 28+NTIME local named structures and 18+NGRP unnamed structures.

Global identifiers

All of these are input parameters.

SUBJECT	FACTOR	factor classifying the subjects
TIME	FACTOR	factor classifying the occasions
NSUBJ	SCALAR	number of subjects
NTIME	SCALAR	number of occasions
RESTAB	TABLE	SUBJECT,TIME table of data
SUBGROUP	FACTOR	factor, length NSUBJ, indicating the group to which each subject belongs.
NGRP	SCALAR	number of groups

Note

The values of all the scalars must be known at the time when the macro is compiled.

Examples of both methods of use, taken from Winer (1971), are given below.

The Macro and Examples

```
'REFERENCE/NUNN=100' EXAMPLES

'MACRO' REPMEAS $
'LOCAL' M,C,SBARJJ,SBAR,DF1,DF2,CHISQ1,CHISQ2,SBARJJ,SBAR,E,DFPOOL,
          UNITY,TWO,THREE,FOUR,SIX,ONE,SBARJ,SSPV,SSV,SO,SSPOOL,DGM,INT,
         V(1...NTIME),VSET,TAB
'SCALAR' M,C,SBARJJ,SBAR,DF1,DF2,CHISQ1,CHISQ2,SBARJJ,SBAR,E,DFPOOL
 : UNITY=1 : TWO=2 : THREE=3 : FOUR=4 : SIX=6
'VARI' ONE,SBARJ $ NTIME
 :     V(1...NTIME) $ NSUBJ
'SET' VSET=V(1...NTIME)
'TABL' TAB $ TIME,SUBJECT
'CALC' TAB=RESTAB
'EQUA' VSET=TAB
'INTEGER' INT
'DSSP' SSPV $ VSET
'SYMMAT' SSV,SO,SSPOOL $ NTIME
'DIAGMAT' DGM $ NTIME
'CALC' SSPOOL,DFPOOL,M,C=0
'FOR' I=1...NGRP
'REST' VSET,SUBGROUP $ SUBGROUP=I ; INT
'SSP' SSPV
'EQUATE' SSV=SSPV
'CALC' SSV=SSV/(NVAL(SUBGROUP)-UNITY)
```

```
'CALC' M=M-(NVAL(SUBGROUP)-UNITY)*LOG(DET(SSV))
 :     C=C+UNITY/(NVAL(SUBGROUP)-UNITY)
 :     SSPOOL=SSPOOL+(NVAL(SUBGROUP)-UNITY)*SSV
 :     DFPOOL=DFPOOL+(NVAL(SUBGROUP)-UNITY)
'REPE'
'REST' VSET,SUBGROUP
'CALC' SSV=SSPOOL/DFPOOL
'CAPT' ''
    ** POOLED COVARIANCE MATRIX **''
'PRINT' SSV $ 12.4
'CALC' M=M+DFPOOL*LOG(DET(SSV))
 :     C=C-UNITY/DFPOOL
 :     C=C*(TWO*NTIME*NTIME+THREE*NTIME-UNITY)/(SIX*(NTIME+UNITY)*(NGRP-UNITY))
 :     DF1=NTIME*(NTIME+UNITY)*(NGRP-UNITY)/TWO
 :     CHISQ1=(UNITY-C)*M
 :     M=(SUM(SSV)-TRACE(SSV))/(NTIME*(NTIME-UNITY)/TWO)
 :     C=TRACE(SSV)/NTIME-M
 :     DGM=C
 :     SO=M
 :     SO=SO+DGM
'CAPT' ''
    ** COMPOUND SYMMETRIC COVARIANCE MATRIX **''
'PRINT' SO $ 12.4
'CALC' M=-(NVAL(SUBGROUP)-NGRP)*LOG(DET(SSV)/DET(SO))
 :     C=NTIME*(NTIME+UNITY)*(NTIME+UNITY)*(TWO*NTIME-THREE)
 :     C=C/(SIX*(NVAL(SUBGROUP)-NGRP)*(NTIME-UNITY)*(NTIME*NTIME+NTIME-FOUR))
 :     DF2=(NTIME*NTIME+NTIME-FOUR)/TWO
 :     CHISQ2=(UNITY-C)*M
'CAPT' ''
    ** BOX'S TEST OF EQUALITY OF COVARIANCE MATRICES **''
'PRINT' DF1,CHISQ1 $ 5.0,10.4

'CAPT' ''
    ** BOX'S TEST OF SYMMETRY OF COVARIANCE MATRICES **''
'PRINT' DF2,CHISQ2 $ 5.0,10.4
'CALC'  ONE=UNITY
 :      SBARJJ=TRACE(SSV)/NTIME
 :      SBARJ=PDT(SSV;ONE)/NTIME
 :      SBAR=MEAN(SBARJ)
 :      E=NTIME*NTIME*(SBARJJ-SBAR)*(SBARJJ-SBAR)
 :      E=E/((NTIME-UNITY)*(SUM(PDT(SSV*SSV;ONE))-TWO*NTIME*SUM(SBARJ*SBARJ)
          +NTIME*NTIME*SBAR*SBAR))
'CAPT' ''
    ** GREENHOUSE-GEISSER EPSILON **''
'PRINT' E $ 10.4
'ENDMACRO'
```

```
''
```

TESTING EQUALITY AND SYMMETRY OF COVARIANCE MATRICES USING MACRO REPMEAS (WINER - PAGE 594)

The assumptions made in analysing a repeated measures design as a split-plot design include the assumption that the covariance matrices of the response at different times are equal for all treatments and have compound symmetry (i.e. all covariances are equal and all variances are equal).

In this experiment measurements were made at three times on ten subjects, each receiving a single treatment for the whole experiment. There were five subjects for each of the two treatments.

```
''
'SCAL' NSUBJ=10
 :     NTIME=3
 :     NGRP=2
'UNIT' $ 30
'FACTOR' SUBGROUP $ 2=5(1,2)
 :       SUBJECT $ NSUBJ=3(1...10)
 :       TIME $ NTIME=(1...3)10
'READ/PRIN=DE' DATA
'RUN'
4 7 2 3 5 1 7 9 6 6 6 2 5 5 1
8 2 5 4 1 1 6 3 4 9 5 2 7 1 1
'EOD'
''
```

The first method of using REPMEAS is to tabulate the data and use this as input to the macro.

```
''
'TABL' RESTAB $ SUBJECT,TIME
'TABU' DATA ; RESTAB
'USE' REPMEAS $
'DEVALUE' RESTAB
'RUN'
''
```

The second method of using REPMEAS is to perform an analysis of variance and to extract the residual effects from the SUBJECTS and SUBJECTS.TIME strata and add them together.

This is the only method of use when factors have been randomised onto the different occasions, in which case their effects have to be removed before examining the covariance structure.

```
''
'BLOCK' SUBJECT/TIME
'TREAT'
'ANOVA/PR=0' DATA ; OUT=OV
'EXTRACT' OV ; SUBJECT/TIME $ EFF=RESTAB1,RESTAB
'CALC' RESTAB=RESTAB+RESTAB1
'USE' REPMEAS $
'RUN'
'CLOSE'
```

References

Box, G.E.P.	(1950)	Problems in the Analysis of Growth and Wear Curves. Biometrics, **6**, 362–389.
Greenhouse, S.W. and Geisser, S.	(1959)	On Methods in the Analysis of Profile Data. Psychometrika, **24**, 95–112.
Winer, B.J.	(1971)	Statistical Principles in Experimental Design (2nd Edition). McGraw–Hill, New York.

T.J. Dixon
R.W. Payne
Statistics Department
Rothamsted Experimental Station
Harpenden
Hertfordshire AL5 2JQ
United Kingdom

REGRESSION DIAGNOSTICS IN GENSTAT

Introduction

The following three macros have been written to implement some of the diagnostic procedures available for linear regression. The techniques are used to highlight those units which have considerable influence on the regression and to elucidate the nature of that influence. The procedures presented here are essentially based on those of Atkinson (1982).

Description of the Functions of Each Macro

DIAGNOSE This macro fits a regression equation to the user's data. (An intercept is automatically fitted, there is no convenient way of omitting this term.) The diagonal elements of the projection matrix (leverages) are calculated and printed in serial order.

The conventional, standardised and jack-knife residuals, as well as the modified Cook statistic (see Atkinson), are calculated and saved. As it is likely that the user will want to inspect those units with large values for the various statistics, the units with the ten largest values for jack-knife residuals and for Cook statistics are printed. Finally, depending on the value of a flag, half-normal plots of the jack-knife residuals and Cook statistics are produced.

CALENV One of the innovations in Atkinson's paper is the use of simulated samples to aid in the discrimination between points which are correctly influential and those whose influence may be due to an error in the covariate vector. This macro generates the envelopes described in Atkinson's paper, based on NSIM simulated samples, and draws the half-normal plots of the envelopes together with the actual curve derived from the data.

COJACK This macro calculates the residual mean square and regression coefficients resulting from omitting a particular point, without refitting the regression. The formulae used can be found in Cook and Weisberg (page 33) and Atkinson.

Using the Macros

Although, of course, each macro can be used 'standing alone' (and sufficient information is given below to allow this) it is intended that DIAGNOSE be used first. This will ensure that quantities such as the conventional residuals and leverages will be saved in structures whose names are compatible with their use in the other two macros.

Structures Required and Created by the Macros

DIAGNOSE

Input

SCALARS	N	Number of units
	P	Number of regression coefficients (this must include the intercept)
	IGRAPH	Flag controlling plotting of the graphs, 1 for plotting and 0 otherwise
VARIATES (of length N)	Y	Response variable
	X	List of (P-1) VARIATES, each of length N, defined using a SET directive

OUTPUT

VARIATES (of length N)	LEVER	Leverages in serial order
	RESIDS	Conventional residuals
	STRES	Standardised residuals
	JACKRES	Jack-knife residuals
	COOK	Modified Cook statistic
	ABJACK	Ordered absolute values of JACKRES
	ABCOOK	Ordered absolute values of COOK
	SQJ,LVJ	Unit numbers and leverages in order of increasing jack-knife residual
	SQC,LVC	As for SQJ, LVJ, but using modified Cook statistic
	HALFN	Half-normal order statistics
VARIATES (of length P)	FCOF	Fitted regression coefficients
SYMMAT (P by P)	INVM	Inverse matrix

CALENV

INPUT

	N,P,X	As for DIAGNOSE
	LEVER	As from DIAGNOSE
	ABCOOK,ABJACK	As from DIAGNOSE
SCALARS	SEED	Starting value for random number generator
	NSIM	Number of samples to be simulated

OUTPUT

VARIATES (of length N)	CHI,CLOW,JHI,	Envelope of simulated order statistics of

	JLOW	absolute value of Cook and jack-knife residuals

COJACK

INPUT

	N,P,X	As for DIAGNOSE
	RESIDS,LEVER,INVM	As from DIAGNOSE
	FCOF	As from DIAGNOSE
SCALAR	IDROP	Unit number of case to be dropped

OUTPUT

VARIATES (of length P)	XDROP	Covariate vector to be dropped
	JACKCO	Jack-knife regression coefficients
SCALAR	MSS	Jack-knife residual mean square

Macros

DIAGNOSE

```
''
      THIS MACRO IMPLEMENTS SOME OF THE REGRESSION
      DIAGNOSTICS DESCRIBED IN ATKINSON(1982) J.R.S.S.(B) PP1-36
           THE USER SUPPLIES THE FOLLOWING STRUCTURES:-
              X - A LIST OF COVARIATES
              Y - THE RESPONSE VARIATE
            N,P - SCALARS HOLDING THE NO. OF CASES AND THE NO.
                  OF REGRESSION COEFFICIENTS FITTED RESPECTIVELY
                  (NOTE - THE INTERCEPT MUST BE COUNTED IN P)
         IGRAPH - A SCALAR TO CONTROL PRINTING OF GRAPHS
                   0  NO GRAPHS        1  GRAPHS
```

```
''
'LOCAL' L1,SEQ,ONE,LP.MSS,MXT,HX,HJ,HC,HU,J1,J2,J3,C1,C2,C3
'VARI' SEQ,SQJ,SQC,ONE,COOK,HALFN,JACKRES,LEVER,STRES $ N
'VARI' ABJACK,ABCOOK,LVJ,LVC $ N
'VARI' J1,J2,J3,C1,C2,C3 $ 10
'SCAL' MSS,N1   'MATR' MXT $ P,N   'DIAG' LP $ N
'CALC' ONE=1    'CALC' SEQ,SQJ,SQC=CUM(ONE)
'CALC' HALFN=CUM(ONE)   'CALC' HALFN=0.5*(HALFN+1+N)/(1+N)
'CALC' HALFN=NED(HALFN)
'HEAD' HX=''HALF NORMAL ORDER STATISTICS''
'HEAD' HJ=''JACK KNIFE RESIDUALS''
'HEAD' HC='' MODIFIED COOK STATISTICS''
'HEAD' HU=''UNIT NUMBER''
''
```

THE REGRESSION EQUATION IS FITTED AND THE SO CALLED LEVERAGES OF THE POINTS ARE CALCULATED AND PRINTED

```
''
'EQUA' MXT=ONE,X
'TERM' Y,X
'Y' Y
'FIT' X;INV=INVM;RES=RESIDS;COEF=FCOF
'CALC' LP=PDT(TPDT(MXT;INVM);MXT)
'LINE' 10
'CAPT' '' THE LEVERAGES OF THE POINTS IN SERIAL ORDER''
'LINE' 4
'PRIN' LP $ 8.3
'PAGE'
'EQUA' LEVER=LP   'DEVA' LP,MXT
''
```

VARIOUS RESIDUAL STATISTICS ARE COMPUTED.

THESE ARE : -

STRES - STANDARDISED RESIDUALS

JACKRES - JACK KNIFE RESIDUALS

COOK - MODIFIED COOK STATISTICS

```
''
'CALC' STRES=RESIDS**2 'CALC' MSS=SUM(STRES)
'CALC' MSS=MSS/(N-P)   'CALC' MSS=SQRT(MSS)
'CALC' STRES=RESIDS/SQRT(1-LEVER) 'CALC' STRES=STRES/MSS
'CALC' JACKRES=((N-P)-(STRES**2))/(N-P-1)
'CALC' JACKRES=STRES/SQRT(JACKRES)
'CALC' COOK=LEVER/(1-LEVER) 'CALC'COOK=SQRT((N-P)*COOK/P)
'CALC' COOK=COOK*JACKRES
'CALC' ABCOOK,ABJACK=ABS(COOK,JACKRES)
'CALC' SQJ=ORDER(SQJ;ABJACK)   'CALC' SQC=ORDER(SQC;ABCOOK)
'CALC' LVJ=ORDER(LEVER;ABJACK) 'CALC' LVC=ORDER(LEVER;ABCOOK)
'CALC' ABCOOK,ABJACK=ORDER(ABCOOK,ABJACK)
''
```

FOR EACH OF JACK KNIFE RESIDUALS AND COOK STATISTICS THE UNIT NUMBERS OF THE TEN LARGEST MODULI ARE FOUND.

```
''
'CALC' N1=N-10
'FOR' IC=1...10
'CALC' ELEM(C1,C2,C3;IC)=ELEM(SQC,LVC,ABCOOK;N1+IC)
'CALC' ELEM(J1,J2,J3;IC)=ELEM(SQJ,LVJ,ABJACK;N1+IC)
'REPE'
''

    DEPENDING ON THE VALUE OF IGRAPH SERIAL AND

    HALF NORMAL PLOTS ARE PRODUCED
''
'JUMP' L1*(IGRAPH.LT.0.5)
'GRAP/ATX=HU,ATY=HJ' JACKRES;SEQ
'GRAP/ATX=HX,ATY=HJ' ABJACK;HALFN
'GRAP/ATX=HU,ATY=HC' COOK;SEQ
'GRAP/ATX=HX,ATY=HC' ABCOOK;HALFN
'LABEL' L1
'PAGE'
'CAPT' ''
      THE TEN LARGEST VALUES OF ABSOLUTE JACK KNIFE RESIDUALS

      AND THEIR UNIT NUMBER'S AND LEVERAGE'S

        J1-  UNIT NUMBER

        J2-  LEVERAGE

        J3- ABSOLUTE VALUE OF JACK KNIFE RESIDUAL
''
'PRIN/P' J1,J2,J3 $ 3.0,7.3,8.3
'LINE' 3
'CAPT' ''
      THE TEN LARGEST VALUES OF ABSOLUTE MODIFIED COOK STATISTIC

      AND THEIR UNIT NUMBER'S AND LEVERAGE'S

        C1-  UNIT NUMBER

        C2-  LEVERAGE

        C3- ABSOLUTE VALUE OF MODIFIED COOK STATISTIC
''
'PRIN/P' C1,C2,C3 $ 3.0,7.3,8.3
'ENDMACRO'
```

CALENV

```
''
      THIS MACRO COMPUTES THE SIMULATED ENVELOPES OF BOTH
      THE JACK KNIFE RESIDUALS AND COOK STATISTICS AND DRAWS
      THE HALF NORMAL PLOTS.
      IF THE USER CALLS DIAGNOSE BEFORE CALENV THEN THE ONLY
      INPUT STRUCTURES THAT HE NEEDS TO SUPPLY, IN ADDITION
      TO THOSE REQUIRED BY DIAGNOSE ARE:-
          SEED,NSIM - SCALARS HOLDING AN ARBITRARY STARTING
                      VALUE (.GT.1) FOR THE RANDOM NUMBER GENERATOR
                      AND THE NUMBER OF SIMULATED SAMPLES REQUIRED
                      RESPECTIVELY.
      THE VARIATES CONTAINING THE ENVELOPES ARE PASSED TO THE
      USER IN VARIATES XLOW,XHI.
        X=J JACK KNIFE RESIDUALS ;   X=C  MODIFIED COOK STATISTIC
''
'LOCAL' G1,G2,ERR,MSS,STUR,JRES,ECOOK,YY,HH,HJR,HCS,HN,HALFN
'HEAD' HH=''--*''
'HEAD' HJR=''JACK KNIFE RESIDUALS''
'HEAD' HCS=''MODIFIED COOK STATISTIC''
'HEAD' HN=''HALF NORMAL ORDER STATISTICS''
'VARI' STUR,JRES,ECOOK,YY $ N
'VARI' HALFN=1...N 'CALC' HALFN=0.5*(HALFN+1+N)/(1+N)
'CALC' HALFN=NED(HALFN)
'SCAL' MSS
'CALC' MSS=RANDU(SEED)
'FOR' CNT=1...NSIM
'CALC' YY=RANDU(0)    'CALC' YY=NED(YY)
'TERM' YY,X
'Y' YY
'FIT/PRIN=Z' X ;RES=ERR;INV=INVM
'CALC' STUR=ERR**2 'CALC' MSS=SUM(STUR)
'CALC' MSS=MSS/(N-P) 'CALC' MSS=SQRT(MSS)
'CALC' STUR=ERR/SQRT(1-LEVER)
'CALC' STUR=STUR/MSS
'CALC' JRES=((N-P)-(STUR**2))/(N-P-1)
'CALC' JRES=STUR/SQRT(JRES)
'CALC' JRES=ABS(JRES)
'CALC' ECOOK=LEVER/(1-LEVER)
'CALC' ECOOK=SQRT((N-P)*ECOOK/P)
'CALC' ECOOK=ECOOK*JRES
'CALC' ECOOK,JRES=ORDER(ECOOK,JRES)
'JUMP' G1*(CNT.EQ.1)
'CALC' CLOW=VMIN(CLOW,ECOOK) 'CALC' CHI=VMAX(CHI,ECOOK)
'CALC' JLOW=VMIN(JLOW,JRES) 'CALC' JHI=VMAX(JHI,JRES)
'JUMP' G2
'LABEL' G1
'CALC' CLOW,CHI=ECOOK   'CALC' JLOW,JHI=JRES
'LABEL' G2
'REPE'
'GRAP/ATY=HJR,ATX=HN' JHI,JLOW,ABJACK;HALFN $ ;HH
'GRAP/ATY=HCS,ATX=HN' CHI,CLOW,ABCOOK;HALFN $ ;HH
'PAGE'
'ENDMACRO'
```

COJACK

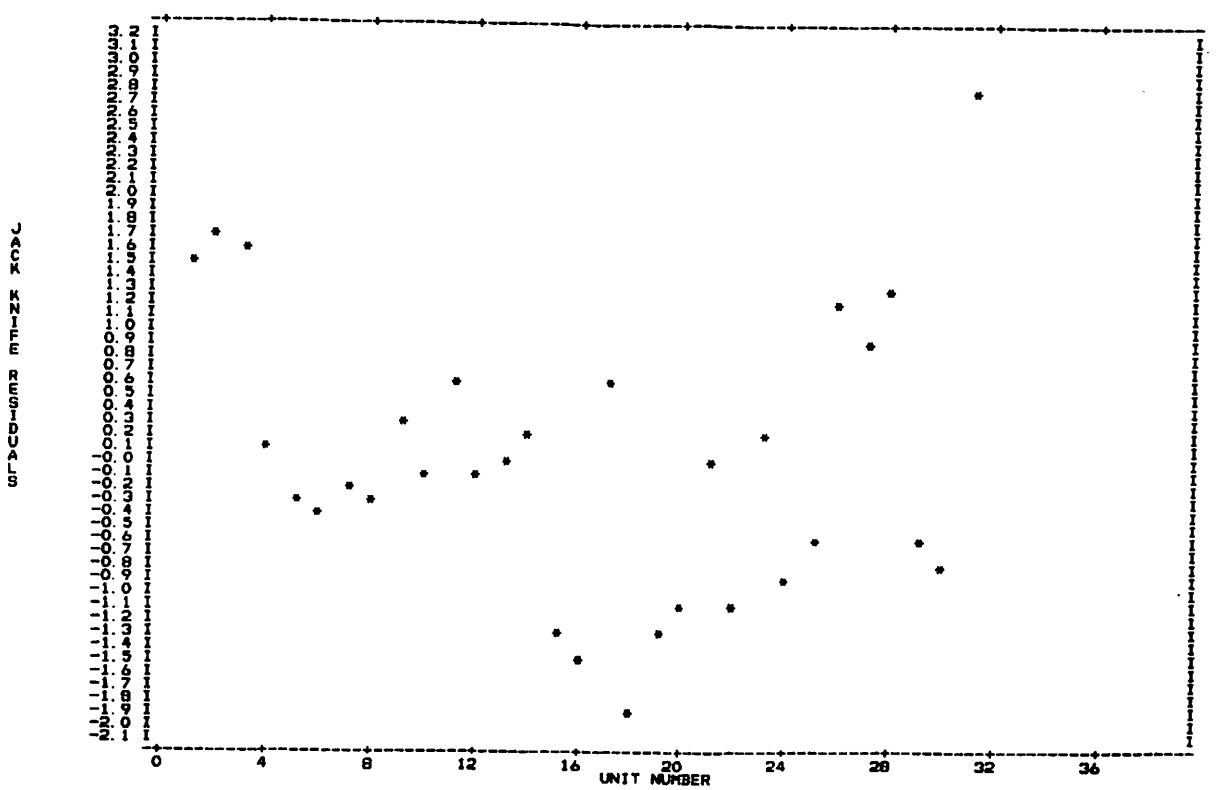
```
''
  THIS MACRO DETERMINES THE CHANGE IN THE REGRESSION
  COEFFICIENTS DUE TO OMITTING A POINT FROM THE FITTING
  WITHOUT REPEATING THE FITTING PROCEDURE. THE JACK KNIFE RESIDUAL
  MEAN SQUARE IS ALSO CALCULATED.
        IF THE USER CALLS DIAGNOSE FIRST THEN THE INPUT
        STRUCTURES REQUIRED IN ADDITION TO THOSE FOR DIAGNOSE
        IS:-
             IDROP - SCALAR HOLDING UNIT NUMBER OF CASE TO BE
                     DROPPED
''
'LOCAL' SPR,ONE,ACO,MXT,CHOO,MSS,JMS
'VARI' ONE,CHOO,SPR $ N   'VARI' XDROP,JACKCO $ P
'MATR' MXT $ P,N   'SCAL' ACO,MSS,JMS
'CALC' ONE=1  'CALC' SPR=RESIDS/(1-LEVER)
'EQUA' MXT=ONE,X  'CALC' CHOO=0
'CALC' ELEM(CHOO;IDROP)=1
'CALC' XDROP=PDT(MXT;CHOO)  'DEVA' MXT
'CALC' JACKCO=PDT(INVM;XDROP)
'CALC' CHOO=CHOO*SPR  'CALC' ACO=SUM(CHOO)
'CALC' ONE=CHOO*RESIDS   'CALC' JMS=SUM(ONE)
'CALC' ONE=RESIDS**2   'CALC' MSS=SUM(ONE)
'CALC' JMS=(MSS-JMS)/(N-1-P)
'CALC' JACKCO=FCOF-ACO*JACKCO
'LINE' 10
'CAPT' ''**** JACK KNIFING COEFFICIENTS ****''
'LINE' 4
'CAPT' ''----UNIT DROPPED WAS ''
'PRIN' IDROP $ 3.0
'LINE' 2
'CAPT' ''----THIS UNIT HAD COVARIATES XDROP''
'PRIN' XDROP $ -15.4
'LINE' 6
'CAPT' ''FULL FIT COEFFICIENTS ARE IN FCOF''
'LINE' 1
'CAPT' ''JACK KNIFE COEFFICIENTS ARE IN JACKCO''
'LINE' 2
'PRIN/P' FCOF,JACKCO $ -15.4
'LINE' 3
'CAPT' ''  THE JACK KNIFE RESIDUAL MEAN SQUARE IS IN JMS''
'LINE' 1
'PRIN' JMS $ -15.4
'ENDMACRO'
```

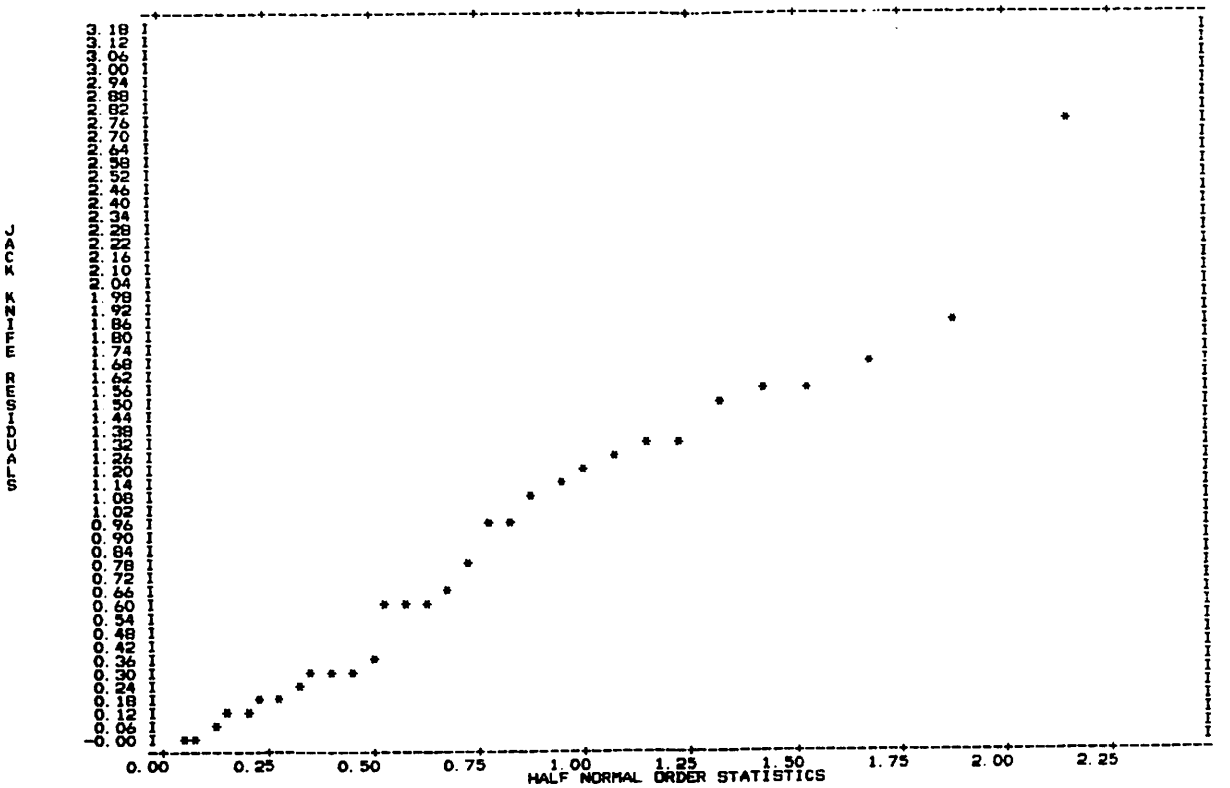
Example Output

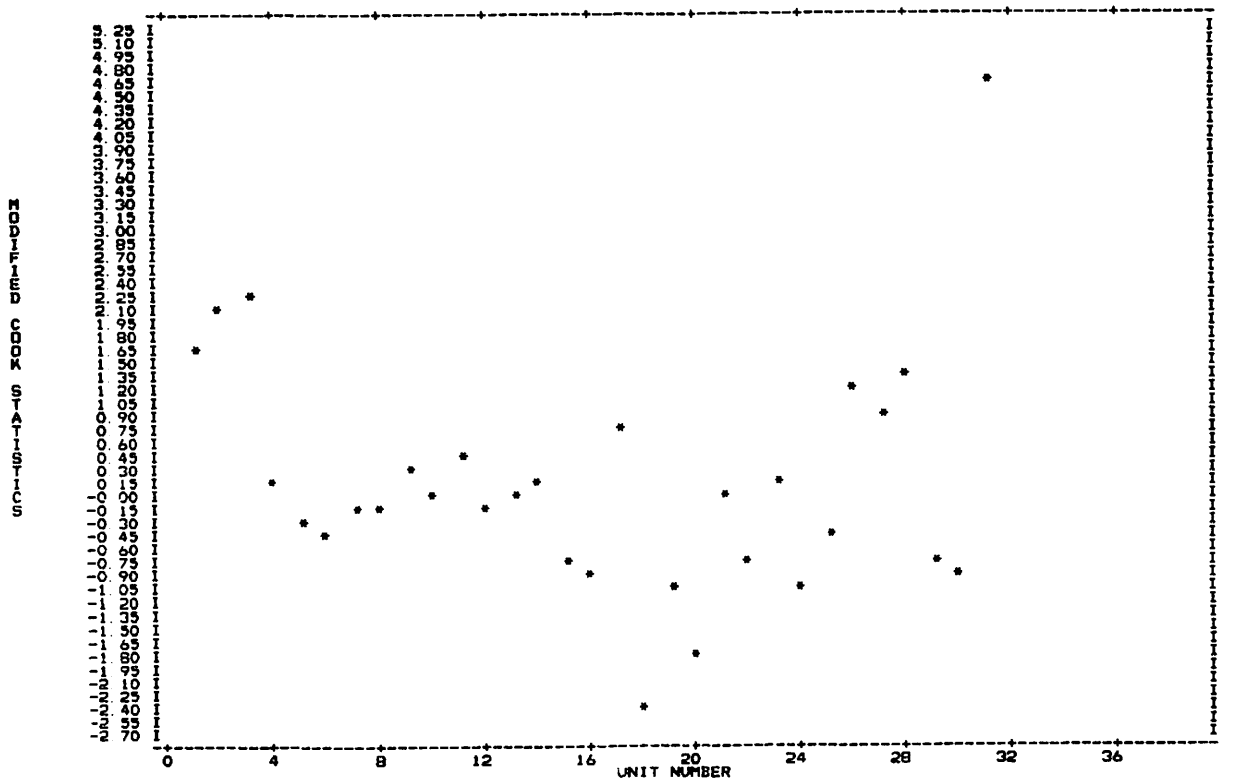
The example uses the MINITAB tree data analysed by Atkinson in his paper and taken from Ryan et al. (page 278).

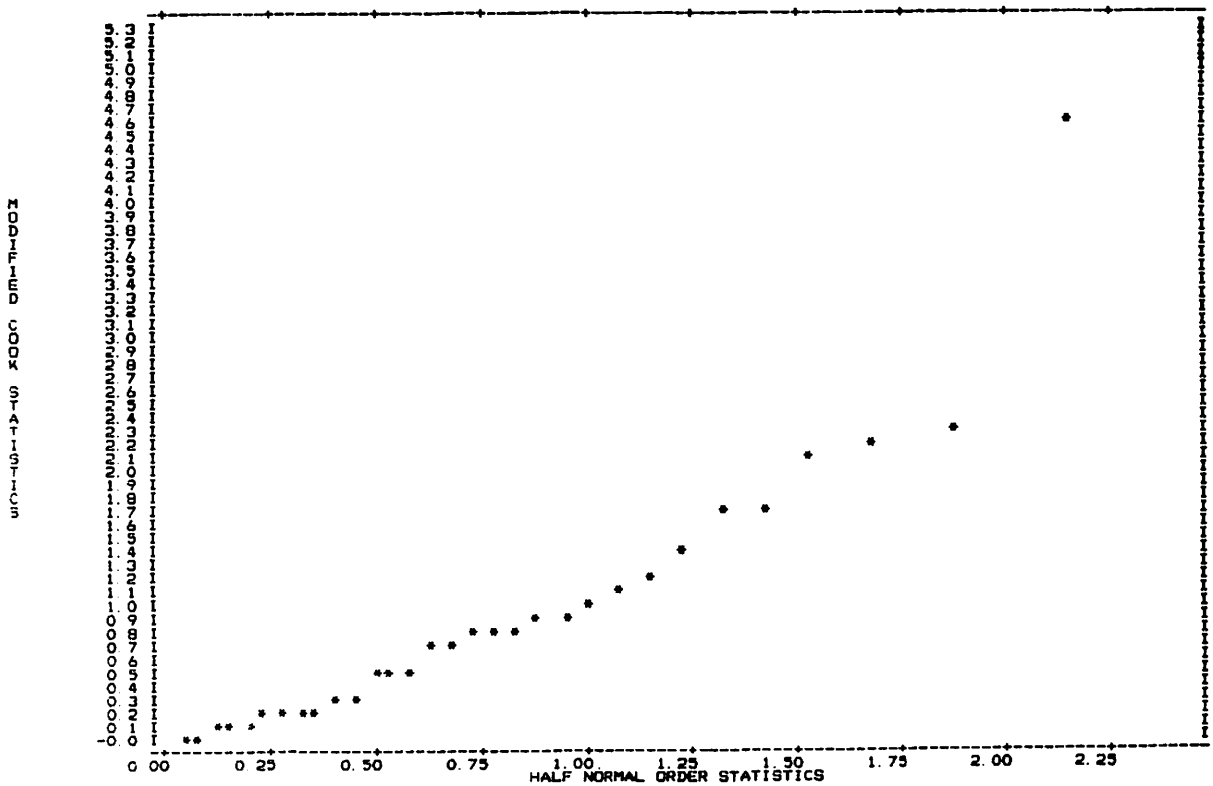
Output from DIAGNOSE

```
***** REGRESSION ANALYSIS *****

*** REGRESSION COEFFICIENTS ***
Y-VARIATE: Y
                ESTIMATE          S.E.            T
CONSTANT       -57.98769       8.63824        -6.71
XX(1)            4.70816       0.26427        17.82
XX(2)            0.33925       0.13015         2.61
*** ANALYSIS OF VARIANCE ***
                DF            SS            MS
REGRESSN         2        7684.2        3842.08
RESIDUAL        28         421.9          15.07
   TOTAL        30        8106.1         270.20
  CHANGE        -2       -7684.2        3842.08
PERCENTAGE VARIANCE ACCOUNTED FOR 94.4

  THE LEVERAGES OF THE POINTS IN SERIAL ORDER

     LP
      1
  0.116      2      3      4      5      6      7      8      9     10     11     12     13     14
          0.147  0.177  0.059  0.121  0.156  0.115  0.051  0.092  0.048  0.074  0.048  0.048  0.073
     15     16     17     18     19     20     21     22     23     24     25     26     27     28
  0.038  0.036  0.131  0.143  0.067  0.211  0.036  0.045  0.050  0.111  0.069  0.088  0.096  0.106
     29     30     31
  0.110  0.110  0.227
```




JACK KNIFE RESIDUALS
3.18
3.12
3.06
3.00
2.94
2.88
2.82
2.76
2.70
2.64
2.58
2.52
2.46
2.40
2.34
2.28
2.22
2.16
2.10
2.04
1.98
1.92
1.86
1.80
1.74
1.68
1.62
1.56
1.50
1.44
1.38
1.32
1.26
1.20
1.14
1.08
1.02
0.96
0.90
0.84
0.78
0.72
0.66
0.60
0.54
0.48
0.42
0.36
0.30
0.24
0.18
0.12
0.06
-0.00
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
HALF NORMAL ORDER STATISTICS


MODIFIED COOK STATISTICS
5.25
5.10
4.95
4.80
4.65
4.50
4.35
4.20
4.05
3.90
3.75
3.60
3.45
3.30
3.15
3.00
2.85
2.70
2.55
2.40
2.25
2.10
1.95
1.80
1.65
1.50
1.35
1.20
1.05
0.90
0.75
0.60
0.45
0.30
0.15
-0.00
-0.15
-0.30
-0.45
-0.60
-0.75
-0.90
-1.05
-1.20
-1.35
-1.50
-1.65
-1.80
-1.95
-2.10
-2.25
-2.40
-2.55
-2.70
0
4
8
12
16
20
24
28
32
36
UNIT NUMBER



THE TEN LARGEST VALUES OF ABSOLUTE JACK KNIFE RESIDUALS
AND THEIR UNIT NUMBER'S AND LEVERAGE'S

J1- UNIT NUMBER
J2- LEVERAGE
J3- ABSOLUTE VALUE OF JACK KNIFE RESIDUAL

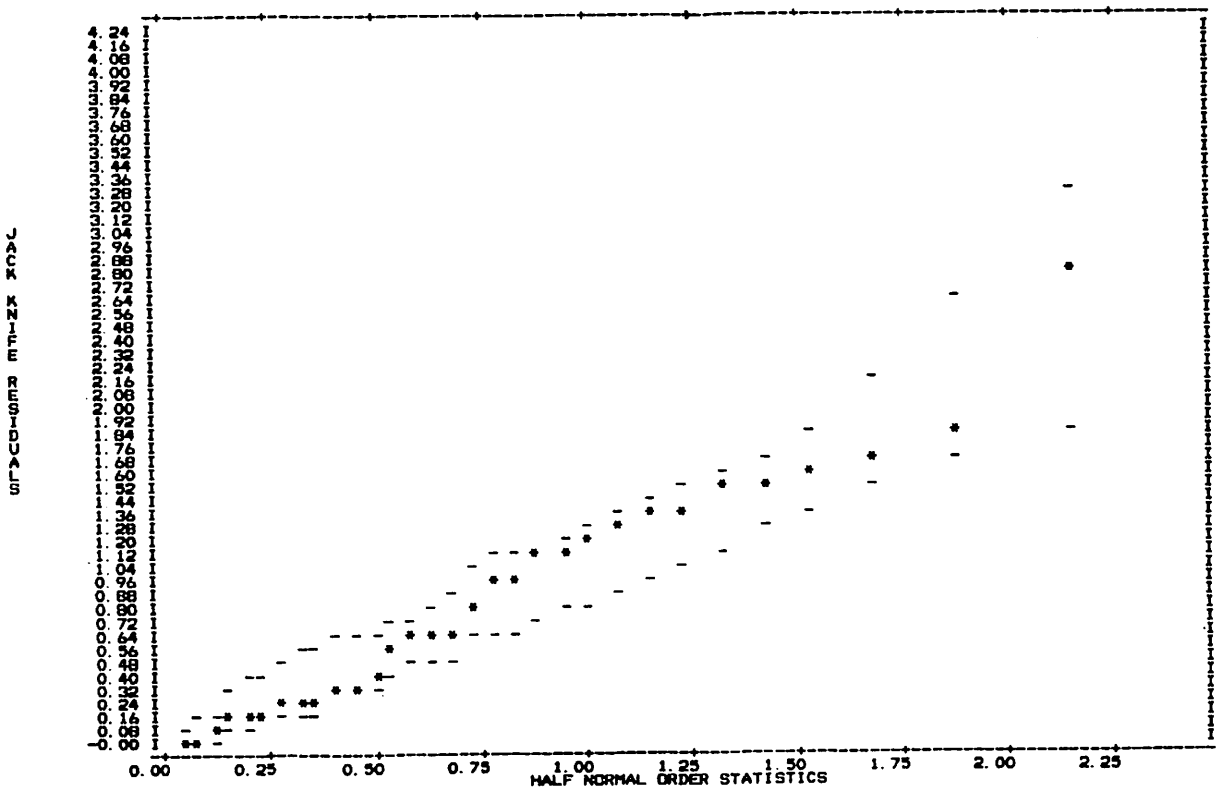
J1	J2	J3
26	0.088	1.213
15	0.038	1.290
19	0.067	1.324
28	0.106	1.347
16	0.036	1.517
1	0.116	1.532
3	0.177	1.568
2	0.147	1.652
18	0.143	1.860
31	0.227	2.766

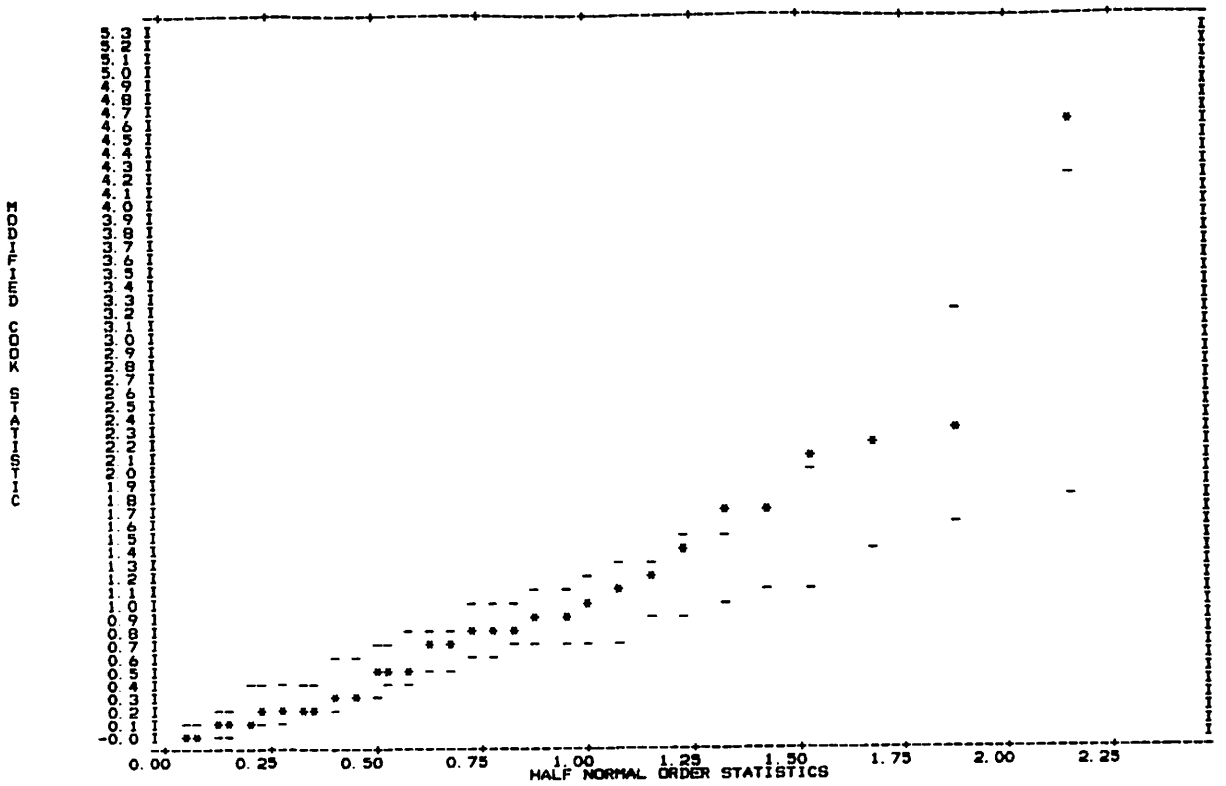
THE TEN LARGEST VALUES OF ABSOLUTE MODIFIED COOK STATISTIC
AND THEIR UNIT NUMBER'S AND LEVERAGE'S

C1- UNIT NUMBER
C2- LEVERAGE
C3- ABSOLUTE VALUE OF MODIFIED COOK STATISTIC

C1	C2	C3
24	0.111	1.024
19	0.067	1.081
26	0.088	1.154
28	0.106	1.420
1	0.116	1.694
20	0.211	1.748
2	0.147	2.096
3	0.177	2.220
18	0.143	2.325
31	0.227	4.579

Output from CALENV





Output from COJACK

```
**** JACK KNIFING COEFFICIENTS ****


----UNIT DROPPED WAS
        IDROP  2

-----THIS UNIT HAD COVARIATES XDROP

          XDROP
     1.0000E  0
     8.6000E  0
     6.5000E  1



FULL FIT COEFFICIENTS ARE IN FCOF
JACK KNIFE COEFFICIENTS ARE IN JACKCO


          FCOF          JACKCO
    -5.7988E  1    -6.2475E  1
     4.7082E  0     4.7655E  0
     3.3925E -1     3.8545E -1

 THE JACK KNIFE RESIDUAL MEAN SQUARE IS IN JMS

          JMS     1.4193E  1

 53  'PAGE'
 54  'PRIN' DIAGNOSE
 55  'LINE' 10
 56  'PRIN' CALENV
 57  'LINE' 10
 58  'PRIN' COJACK
 59  'RUN'
```

Acknowledgement

I would like to thank Tom Marshall for helpful discussions when writing these macros.

References

Atkinson, A.C. (1982) Regression Diagnostics, Transformations and Constructed Variables. J.R.S.S. (B), 44, 1–36.

Cook, R.D. and Weisberg, S. (1982) Residuals and Influence in Regression. (Chapman & Hall).

Ryan, T.A., Joiner, B.L. and Ryan, B.F. (1976) Minitab Student Handbook. (Duxbury Press).

J.N.S. Matthews
Department of Medical Statistics
London School of Hygiene
& Tropical Medicine
Keppel Street
London WC1E 7HT
United Kingdom

BOX-COX TRANSFORMATION

Introduction

The two parameter power transformation of Box and Cox, (1964) is well known. The transformed scale, Z, from data, Y, involves a location parameter, λ_2 and a power parameter, λ_1, via the relation,

$$Z = \frac{(Y+\lambda_2)^{\lambda_1} - 1}{\lambda_1} \qquad \lambda_1 \neq 0$$
$$Z = ln(Y+\lambda_2) \qquad \lambda_1 = 0 \qquad 1.1$$

The values of λ_1 and λ_2 are chosen to give, as far as possible, a transformed scale which is additive with respect to some specified factor effects and possesses a Normal distribution with constant variance over all factor combinations.

The equivalent normalised transformation is

$$Z = \frac{(Y+\lambda_2)^{\lambda_1} - 1}{\lambda_1 \, g^{(\lambda_1 - 1)}} \qquad \lambda_1 \neq 0$$
$$Z = g ln(Y+\lambda_2) \qquad \lambda_1 = 0 \qquad 1.2$$

where g is the geometric mean of the $Y + \lambda_2$ values.

The log-likelihood of Z (from 1.2) maximised over the desired additive model is $L(\lambda_1,\lambda_2)$ where,

$$L(\lambda_1,\lambda_2) = \frac{n}{2} \ln\left(\frac{\text{RESS}}{n}\right) \qquad 1.3$$

The number of observations in Y is n and the residual sum of squares in the ANOVA of Y is RESS.

The optimum values of λ_1 and λ_2 are those which maximise $L(\lambda_1,\lambda_2)$ (equivalently minimise RESS) and can be found iteratively by some maximisation routine, or more simply, by careful examination of $L(\lambda_1,\lambda_2)$ over a grid of λ_1 and λ_2 values.

A macro, called BOX, written in Genstat has been developed to produce the optimum transformation. This was in response to demand for data transformation facilities and is run on the Polytechnic's DEC–20.

The program picks out the largest $L(\lambda_1,\lambda_2)$ over a user specified grid of λ_1,λ_2 values. The transformation of Y to Z for computing purposes is given by

$$Z = \frac{g}{|\lambda_1|}\left(\frac{Y+\lambda_2}{g}\right)^{\lambda_1} \qquad \lambda_1 \neq 0$$
$$Z = g \, ln(Y+\lambda_2) \qquad \lambda_1 = 0 \qquad 1.4$$

where g is the geometric mean of $Y + \lambda_2$ values and where $\lambda_2 > -MIN(Y_1,....,Y_n)$.

Transformation 1.4 is equivalent to 1.2 but is more numerically reliable than 1.2. It has been tried on a number of examples, including the two in Box and Cox, (1964) and the 'Plankton Haul' data in Snedecor and Cochran, (1973). The Snedecor and Cochran data range from *387* to *43300*. Transformation 1.4 in the program agreed with those optimal transformations found by the authors in these three examples and has also given sensible optimal transformations for other unpublished data sets.

The output from the program includes optimum values of λ_1 and λ_2 and also λ_1 when λ_2 is constrained to be zero. An approximate *95%* confidence region for λ_1 and λ_2 is given and an approximate *95%* confidence interval for λ_1 is given when λ_2 is constrained to be zero.

Full details of user supplied data and output of results are given in the following sections.

Input of Data

The data, Y, are assumed to be structured in a balanced factorial experiment. The number of factors, their levels and the grid values of λ_1 and λ_2 have to be specified by the user in a data file. The data file should consist of, firstly, five lines of control information, secondly, the data values, Y, and, lastly, the end of data terminator 'EOD'

The file should contain:

''HEADING''	HEADING is text which the user may wish to describe the data, Y. This heading should be contained in two pairs of single quote symbols and is printed in the output.
NF RE	Two SCALARs in free format. NF is the number of factors and RE is the number of replicate observations for each combination of factor levels. The program assumes $1 \leq$ NF ≤ 8 and RE ≥ 1
M LAM1 S1 N LAM2 S2	Six SCALARs in free format (M, N, are INTEGERs). The grid of λ_1 and λ_2 values is given by M values of λ_1 starting at LAM1 and increasing in steps of S1 Hence LAM1 $\leq \lambda_1 \leq$ LAM1 + (M-1)*S1. Similarly there are N values of λ_2 starting at LAM2 and increasing in steps of S2. Hence LAM2 $\leq \lambda_2 \leq$ LAM2 + (N-1)*S2
LSEN PRIV	Two SCALARs in free format. LSEN is a 'sensible' value of λ_1 in terms of the experiment. This value is used by the program if LSEN is inside the *95%* confidence interval of λ_1. If no such sensible value exists, put a value of LSEN that is not in the range of λ_1 values specified by M, LAM1, S1. PRIV controls the output of the transformed data. If PRIV ≥ 1, the transformed data are output. If PRIV < 1, no output of data is given.
B(1) B(2) ... B(NF)	There must now be NF INTEGERs in free format giving the number of levels of the NF factors.
Y	The VARIATE Y of data value in free format. There will be NU values where

$$NU = RE * \prod_{I=1}^{NF} B(I)$$

The order of the factors specified by the list of levels B(1), B(2),..., B(NF) [B(1...NF)] specifies the arrangement of the NU values of *Y*. The order of the *Y* values, in fact, depends upon the list of factors in a 'GENERATE' directive.

The first factor has B(1) levels the second factor has B(2) levels and so on. If these levels index the data (*Y* values) then the *Y* values are arranged such that the levels of the first factor are *changing most slowly*, the levels of the second factor are changing *less slowly* (*2nd slowest*) and so on until the levels of the last factor are changing the fastest except for any replication 'Levels' which must change faster than any factor levels.

'EOD' — The directive 'EOD'
The end of data terminator EOD in single quotes.

Output of Results

Some program output will contain default Genstat output, though the printing of the Genstat instructions is suppressed. The output is summarised in the following list.

1. (i) The heading given in the first line of input is printed as a title.
   (ii) The mean, minimum and maximum of the *Y*–values are output as Genstat default.
2. ANOVA table for the ANOVA of *Y* using a *full* model with interaction terms.
3. (i) ANOVA table for the ANOVA of *Y* using *additive* terms in the model formula only.
   (ii) Plot of residuals against fitted values from the additive model.
   (iii) Plot of Expected Normal quantiles against residuals from the additive model.
   (iv) Histogram of residuals.
4. (i) Plot of log-likelihood of transformed data, maximised under the additive model against the user's specified values of λ_1. [λ_2 is constrained to be zero]. Hence a plot of $L(\lambda_1, \lambda_2 = 0)$ against λ_1 is output.
   (ii) The largest maximised log-likelihood over the M λ_1 values (when λ_2 is constrained to be zero) and the value of λ_1 to give this.
   (iii) The largest maximised log-likelihood over the M*N grid of λ_1 and λ_2 values and the values of λ_1 and λ_2 which achieve this.
   (iv) The (approximate) Chi-squared statistic to test the hypothesis 'λ_2 is zero' on one degree of freedom. The statistic is twice the difference between the largest maximised log-likelihoods found in (ii) and (iii). In addition, text is printed stating whether or not this statistic is significant at the *5%* level. It must be remembered that the optimisation of the additive model log-likelihood is over user supplied values for λ_1 and λ_2 and hence such values of λ_1 and λ_2 must be chosen with care.
5. (i) A contour plot of $L(\lambda_1, \lambda_2)$ over the grid of λ_1 and λ_2 values. The contours are cut off to provide an approximate *95%* confidence region for λ_1 and λ_2.

(ii) A list of λ_1 values in the user's grid which would be inside a *95%* confidence interval of λ_1 when λ_2 is constrained to be zero.

6. The program records whether or not the Chi-squared statistic of 4(iv) is significant. If the statistic is signficant part (ii) is output, otherwise part (i) is output.

(i) $\lambda_2 = 0$ *not rejected*

The ANOVA table of the transformed data, Z ,when λ_2 is zero and when using the optimal value of λ_1 or the user supplied value (LSEN) of λ_1 if LSEN is inside the *95%* confidence interval of 5(ii).

The ANOVA table is supplied for both additive and full models when RE is at least *2*. When the value of RE is *1* the ANOVA table for the additive model only is output. For the additive model 3(ii), (iii) and (iv) are repeated to enable the user to examine the effectiveness of the transformation used.

(ii) $\lambda_2 = 0$ *rejected*

The ANOVA tables for both additive and full models of the transformed data, Z, are output when the transformation is determined by the optimal values of λ_1 and λ_2; only the additive model table is output if RE is *1*. Then 3(ii), (iii) and (iv) are repeated for the additive model, to enable the user to examine the effectiveness of the transformation used.

7. The transformed data (either from 6(i) or 6(ii)) are output if the user's value of PRIV is at least *1*.

The data are given a field width of 8 characters including 3 decimal places.

The Program

```
'REFE/NIT=500,NUNN=700,DUMP=Y,PR=N' INCTRAN
'SCAL' RE:NU:LAM1:LAM2:NT:S1:S2:I:LM:PT:RT:M:N
'SCAL' R(1...9)=1:FNJ:NF
'SCAL' IP1:IM:PL1:PL2:NCL:LCL:RPT:LSEN:LZC=0:PRIV:PSIG
''                                                                    ''
''PROGRAM READS FROM THE FIRST LINE OF DATA FILE A TITLE(HNAM)''
''                                                                    ''
''AND THEN READS THE NEXT THREE LINES CONTAINING THE NUMBER OF ''
''                                                                    ''
''FACTORS,NUMBER OF REPLICATIONS OF EACH COMBINATION OF FACTORS,''
''                                                                    ''
''THE SIX SCALARS SETTING UP USER'S GRID OF PARAMETER VALUES,''
''                                                                    ''
''A SENSIBLE VALUE OF POWER PARAMETER AND PRIV(DATA PRINTING SCALAR)''
'READ' HNAM
'INPUT' 2
'READ' NAME
'READ' NF,RE
'READ' M,LAM1,S1,,,LAM2,S2
'READ' LSEN,PRIV
'INPUT' 1
'READ'
'CALC' M=12+(M-12)*(M.LT.12)
'CALC' NF=R*(NF-R)*(NF.LT.6)
'LIST' 2
'PRINT' HNAM
'LIST' 3
'READ'
''
```

```
''PROGRAM READS THE NUMBER OF LEVELS OF THE NF FACTORS''
''                                                                    ''
'INPUT' 2
'READ' B(1...NF)
'INPUT' 1
'RUN'
'CALC' NU=NE*B(1)*B(2)*B(3)*B(4)*B(5)*B(6)
'CALC' NU=NU*B(7)*B(8)*B(9)
'RUN'
'CALC' N1=M**N
'RUN'
''                                                                    ''
'' THE PROGRAM SETS UP THE VARIATE OF DATA Y AND READS IN Y  ''
''                                                                    ''
'UNITS' $ NU
'VARI' Y:R:F:RESSNT
'INPUT' 2
'READ' Y
'INPUT' 1
'RUN'
''                                                          ''
''THE N VALUES OF LAMDA2 ARE ADDED TO Y TO FORM N VARIATES TY(1...N)''
''AND THEIR GEOMETRIC MEANS ARE CALCULATED''
''                                                                    ''
'SCAL' GM(1...N):P=1:DG:MP=0:LAMB1(1...M):L2:J:MY:RG
'CALC' MY=-MIN(Y) :RG=0.5*(MAX(Y)+MY)/NU
'CALC' LAM2=LAM2+(MY+RG-LAM2)*(LAM2.LE.MY)
'CALC' S1=ABS(S1) :S2=ABS(S2)
'RUN'
'FOR' LDUM=LAMB1(1...M)
'CALC' LDUM=LAM1+MP*S1:MP=MP+1
'REPEAT'
'VARI' TY(1...N):DY:LAMB2SN:EY:MSN1
'FOR' G=GM(1...N) ; TV=TY(1...N)
'CALC' ELEM(LAMB2;P)=LAM2+(P-1)*S2
:TV=T+ELEM(LAMB2;P)
:G=MEAN(LOG(TV)) : G=EXP(G) : P=P+1
'REPEAT'
'SCAL' PM:10
'RUN'
''                                                          ''
''THE NF FACTORS A(1...NF) ARE SET UP ALONG WITH A REPLICATION FACTOR''
''(REP) AND A DUMMY FACTOR CALLED DF.THE VALUES OF THE FACTORS ARE SET''
''UP IN A GENERATE DIRECTIVE.''
''                                                          ''
''                                                                    ''
'FACT' EXUNIT 5 NU=1...NU
'FACT' REPSRE :DFS1
'FACT' A(1)SB(1) :A(2)SB(2) :A(3)SB(3) :A(4)SB(4)
'FACT' A(5)SB(5) :A(6)SB(6) :A(7)SB(7) :A(8)SB(8) :A(9)SB(9)
'GENE' DF,A(1...NF),REP
'RUN'
''                                                                    ''
''ADDITIVE AND FULL FACTORIAL MODELS FOR SUBSEQUENT ANOVA'S''
''ARE SET UP USING MACRO MODSETS (THE ADDITIVE MODEL ONLY IF ''
''THE NUMBER OF FACTORS OR NUMBER OF REPLICATIONS IS ONE ''
''                                                          ''
'SET/*' LIF=A(1) :LFF=A(1)
'MACRO' MODSETS
'GOTO' LBV*((NF.EQ.1).OR.(FNJ.EQ.1))
'SET/*' LIF=LIF+A(FNJ)
'SET/*' LFF=LFF*A(FNJ)
'LABEL' LBV
'END*'
'RUN'
'FOR' FSUM=1...NF
'CALC' FNJ=FSUM
```

```
'USE/R' MODSETS
'REPEAT'
'RUN'
''                                                    ''
'' THE ANOVA OF RAW DATA Y WITH VARIOUS GRAPHS OF RESIDUALS ENABLING''
''THE USER TO CONSIDER THE ANOVA ASSUMPTIONS''
''                                                          ''
'GOTO' L28*((RE.EQ.1).OR.(NF.EQ.1))
'TREAT' LPF
'CAPTION'
''ANOVA OF DATA VALUES : FULL MODEL''
'ANOV' Y
'LABEL' L28
'BLOCK' EXUNIT
'TREAT' LTF
'VARI' TTY(1...M):LM3SM:LB1SM
'CAPTION'
''ANOVA OF DATA VALUES : ADDITIVE MODEL''
'ANOV' Y;RES=R;FVAL=F
'RUN'
'RES1' R,FSLF=1
'RUN'
'MACRO' RYLS
 'LOCAL' XTLE,YTLE,Y2TLE,Y1TLE,X2TLE,X3TLE,NVON,A,B,K,P ,YB,CH,
        EX,AA,FSQ,SQRES,LSA,L(1...4),SEL,QNORM,NRES,GON
'VARI' GNORM,EX,FSQ,SQRES $ K
'RESTRICT' GNORM,EX,FSQ,SQRES $ P
'READING'  XTLE=''  FITTED VALUES  ''
         :  YTLE=''  RESIDUALS FROM FITTED MODEL  ''
         : Y1TLE=''  EXPECTED NORMAL QUANTILES  ''
         : Y2TLE=''  FREQUENCY  ''
         : X2TLE=''  HISTOGRAM OF STANDARDISED RESIDUALS  ''
         : X3TLE=''  RESIDUALS FROM ANALYSIS OF SQUARED FITTED VALUES  ''
'SCALAR' NVON,NRES,A,B,K,P,YB,LSA,L(1...4)
'VARI' SEL=1,1,1,M
'CONF' LSA*(NVAL(SELECT).NE.4)
'EQUATE' SEL=SELECT
'LABEL' LSA
'INTEGER' CH $ 1=2
'CALCULATE' YB=MEAN(F)
'JUMP' L(1)*(ELEM(SEL ; 1).NE.1)
'GRAPH/AIX=XTLE,AIY=YTLE' R ; F
'LABEL' L(1)
'CALCULATE' GNORM=R : NVON,NRES=NVAL(GNORM,P)
'USE/R' GNORM $
'JUMP' L(2)*(ELEM(SEL ; 2).NE.1)
'CALCULATE' EX=QNORM*SQRT(VAR(P))
'READING' AA=''LP''
'GRAPH/AIX=YTLE,AIY=Y1TLE,XCH1=CH' GNORM ; EX,R $ AA
'LABEL' L(2)
'JUMP' L(3)*(ELEM(SEL ; 3).NE.1)
'USE/R' GSQRT $
'LABEL' L(3)
'JUMP' L(4)*(ELEM(SEL ; 4).NE.1)
'CALCULATE' FSQ=F*F
'COVAR'
'ANOVA/FR=-1' FSQ ; RES=SQRES
'GRAPH/AIX=X3TLE,AIY=YTLE' R ; SQRES
'CALCULATE' B=SUM(R*SQRES) : A=SUM(SQRES*SQRES) : K=B/A :  P=1-2*K*YB
'PRINT' K,P $ 10.4
'CAPTION'
''REF. ANSCOMBE AND TUKEY ; TECHNOMETRICS VOL 5,NO 2, MAY 1963
     NOTE: THIS GRAPH ONLY APPLIES TO LINEAR MODELS FITTED VIA
           THE 'ANOVA' STATEMENT  ''
'LABEL' L(4)
'ENDMACRO'
'RUN'
'RUN'
```

```
'MACRO' GGPLOTS
'LOCAL' SHGGN,RHIST,BOUNDS,L,IRV
'FACTOR' RHISTS30
'INTEGER' IRV=1...NRES
'VARI' SHGGN $NRES
'VARIATE' BOUNDS $ 29=-2.8,-2.6...2.8
'JUMP' L*(NRES.EG.NVGN)
'REST' GGNS R=0;IRV
'LABEL' L
'COPY' SHGGN=GGN $ IRV
'GROUPS' RHIST=LIMITS(SHGGN;BOUNDS)
'GRAPH/ATX=X2TLE,ATY=Y2TLE,NRF=18,NCF=60' RHIST;BOUNDS
'ENDM'
'RUN'
'RUN'
'MACRO' GGNAS
'LOCAL' G,GG,MM,MK
'SCAL' MM:VMM:MK=-1.0E20
'CALC' MM=NMV(GNORM)
:VMM=NVGN-MM
'VARI' G,GG,GGN $ GNORM
'REST' G,GG,GGN $ GNORM
'EGUA' GG=1,2...NVGN
'CALC' GGN=(GNORM-MEAN(GNORM))/SQRT(VAR(GNORM))
:GNORM=GG
:GGN=REPMV(MK)
:G=ORDER(GNORM;GGN)
:GNORM=ORDER(GG;G)
:GNORM=GNORM+0/(GGN.GT.MK)
:GGN=GGN+0/(GGN.GT.MK)
:GNORM=GNORM-MM
:GNORM=GNORM+0/(GNORM.GT.0)
:GNORM=(GNORM-0.363)/(VMM-2.0*0.363+1.0)
:GNORM=NEG(GNORM)
'ENDM'
'RUN'
'RUN'
'USE' KYLS
'CAPTION'
''**************************************************************************''
'RUN'
''                                                                        ''
''THE MACROS PASSINS AND POWERTS AND THE NEXT 14 LINES CALCULATES''
''THE VARIATE RES WHICH CONTAINS THE RESIDUAL SUMS OF SQUARES OF ''
''THE M*N ANOVA'S OF THE RAW DATA TRANSFORMED OVER THE GRID OF   ''
''POWER AND LOCATION PARAMETERS                                  ''
''                                                               ''
'MACRO' PASSIN $
'LOCAL' D1,D2
'ASSI' D1=TY(1...N)$PM:D2=GM(1...N)$PM
'CALC' DY=D1/D2:DG=D2
'ENDM'
'RUN'
'MACRO' POWERT $
'UNITS' $ NU
'LOCAL' I,DY,EY,LJ,LI,LH,LM,DG,D1,D2
'USE/R' PASSIN $
'FOR' EY=TY(1...M);LM=LAMB1(1...M)
'GOTO' LI*(LM.EG.0)
'CALC' EY=DG*(DY**LM)/LM
'GOTO' LH
'LABEL' LI 'CALC' EY=DG*LOG(TY(PM))
'LABEL' LH
'REPEAT'
'DEVA' DY,DG
'ENDM'
'RUN'
```

```
'FOR' DPP=1...N
'CALC' PM=DPP
'USE/R' POWERT $
'FOR' IDUM=1...M;TW=TTY(1...M)
'CALC' ID=IDUM
'ANOV/PR=******' TW;OUT=FOUT
'EXTRACT' FOUT;EXUNIT+LTF $ SS=RS
'CALC' PI=(PM-1)*M+ID
:ELEM(RES;PI)=RS
'DEVA' FOUT,RS,PI
'REPEAT'
'DEVA' TTY(1...M),PM
'REPEAT'
'RUN'
''                                                                    ''
'' RES WILL NOW CONTAIN THE LOG-LIKELIHOOD OF TRANSFORMED DATA        ''
''                                                                    ''
'CALC' RES=LOG(RES) : RES=0.5*NU*(LOG(NU)-RES)
'RUN'
''                                                                    ''
''IF ZERO IS CONTAINED IN THE USER'S RANGE FOR THE LOCATION PARAMETER''
''THEN THE LOG-LIKELIHOOD IS SELECTED FROM RES,OTHERWISE SUCH        ''
''LOG-LIKELIHOODS ARE CALCULATED.                                    ''
''                                                                    ''
'FOR' DP=1...N
'CALC' J=DP
'CALC' L2=ELEM(LAMB2;J)
'CALC' I=0
'LABEL' LK 'CALC' I=I+1
'CALC' PJ=(LP-1)*M+I
:PT=ELEM(RES;PJ)
:LM=LAM1+(I-1)*S1
'GOTO' LP*(L2.NE.0)
'CALC' ELEM(LM3;I)=RT
'CALC' LZC=LZC+1
'LABEL' LP
'GOTO' LK*(I.LT.M)
'REPEAT'
'DEVA' GM(1...N)
'RUN'
'CALC' GM(1)=MEAN(LOG(Y)) :GM(1)=EXP(GM(1))
'RUN'
'DEVA' TTY(1...M),PM,TY(1)
'CALC' PM=1:TY(1)=Y
'RUN'
'USE/R' POWERTS
'RUN'
'GOTO' LC7*(M.LE.LZC)
'SCAL' IZ2
'FOR' IZD=1...M;TZ=TTY(1...M)
'CALC' IZ2=IZD
'ANOV/PR=******' TZ;OUT=FOUT
'EXTR' FOUT;EXUNIT+LTF SSS=RS4
'CALC' ELEM(LM3;IZ2)=RS4
'DEVA' FOUT,RS4
'REPEAT'
'LABEL' LC7
'RUN'
'CAPTION'
''LIST OF GRIDVALUES OF LOCATION PARAMETER(LAMDA2) ''
'PRINT/LABC=1' LAMB2 S8.2
'EQUATE' LB1=LAMP1(1...M)
'RUN'
'CAPTION'
''LIST OF GRID VALUES OF POWER PARAMETER(LAMBDA1)''
'LINE' 2
'PRINT/LABC=1' LB1 S7.3
'RUN'
'GOTO' L1K*(M.LE.LZC)
```

```
'CALC' LM3=LOG(LM3):LM3=M.5*NU*(LOG(NU)-LM3)
'LABEL' L1M
'READ' HY='''MAX-LOG-LIKELIHOOD'':HX='''LAMBDA1(POWER PARAMETER)'''
:H*='''LARGEST NORMAL ADDITIVE LOG-LIKELIHOOD OVER''
:H*2='' TWO TRANSFORMATION PARAMETERS ''
:H*3='' TOGETHER WITH ITS ELEMENT NUMBER IN GRID OF PARAMETERS''
:HZ='''LARGEST NORMAL ADDITIVE LOG-LIKELIHOOD OVER''
:HZ2='' POWER PARAMETER VALUES ONLY WITH ITS ELEMENT NUMBER''
:HZ3='' IN LIST OF POWER PARAMETER VALUES''
:H*='''CHI-SQUARE STATISTIC FOR LAMBDA2 TO BE ZERO ON ONE DF IS ''
:HL*='''BEST VALUE OF POWER PARAMETER WITH ZERO LOCATION PARAMETER IS''
:HLC='''LIST OF POWER PARAMETER VALUES IN 95% CONFIDENCE INTERVAL''
:HLC2='''WHEN LOCATION PARAMETER IS ZERO''
:HL*2='''BEST VALUES OF POWER & LOCATION PARAMETERS''
:HL1='''POWER PARAMETER IS ''
:HL12='' LOCATION PARAMETER IS ''
:HG5='''PLOT OF MAXIMISED LOG-LIKELIHOOD AGAINST POWER PARAMETER''
:HG6='''   WHEN LOCATION PARAMETER IS ZERO''
:HG7='''**********************************************************''
'RUN'
''                                                                    ''
'''THE MAXIMUM LOG-LIKELIHOODS ARE FOUND ALONG WITH THE OPTIMUM     ''
'''VALUES OF THE POWER AND LOCATION PARAMETERS.                      ''
''                                                                   ''
'GRAP/ATY=HY,ATX=HX,NRF=18,NCF=60' LM3;LM1
'PRINT' HG5
'PRINT' HG6
'LINE' 2
'PRINT' HG7
'SCAL' C:TU:TL:MP1:ML:H2:H1
'VAR1' HCS4:X=1...NT:XL=1...M:W2SM
'CALC' W=ORDER(X;RES):MPT=ELEM(W;N1)
:H2=ELEM(RES;MPT)
'CALC' RP1=MPT/M :IPT=INTPT(RP1) :IM=MPT-M*IPT
'RUN'
'GOTO' L20*(IM.G1.0)
'CALC' IPT=IPT-1
:IM=M
'LABEL' L20
'RUN'
'CALC' BL1=LAMB1(IM) :IPT=IPT+1
:BL2=ELEM(LAMB2;IPT)
'PRINT' HW
'PRINT' HW2
'PRINT' HW3
'PRINT/FORM=C,LABR=1' H2,MPT $10.3,3X,3.0
'PRINT' HLM2
'PRINT/FORM=C,LABR=1' HLT,BL1,HLT2,BL2 $2X,0,6.2,0,8.2
'CALC' W2=ORDER(XL;LM3):ML=ELEM(W2;M)
:H1=ELEM(LM3;ML)
'PRINT' HZ
'PRINT' HZ2
'PRINT' HZ3
'PRINT/FORM=C,LABR=1' H1,ML $10.3,3X,3.0
'RUN'
''                                                                    ''
'''THE NECESSARY VALUES FOR THE OPTIONS IN A CONTOUR PLOT OF GRID  ''
'''OF LOG-LIKELIHOODS ARE CALCULATED. THE CHI-SQUARE STATISTIC TO  ''
'''TEST THE HYPOTHESIS THAT THE LOCATION PARAMETER IS ZERO IS ALSO''
'''CALCULATED.                                                      ''
''                                                                   ''
'PRINT/FORM=C,LABR=1' HLM,LAMB1(ML) $0,6.2
'CALC' ELEM(BC;3)=LAM2:ELEM(BC;4)=ELEM(LAMB2;N)
:ELEM(BC;1)=LAM1:ELEM(BC;2)=LAMB1(M)
:C=(H2-MIN(RES))/NT
:TL=(H2-2.990)/C
:TU=(H2+100)/C
'RUN'
```

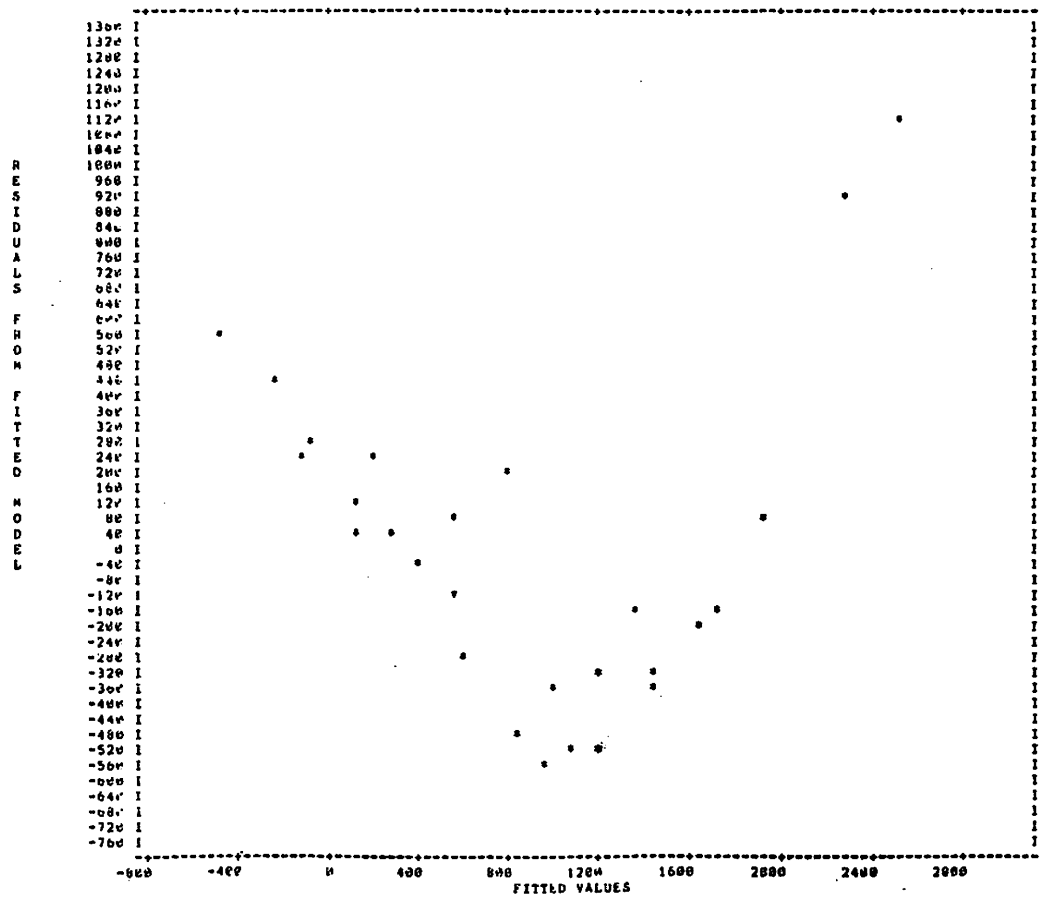
```
'CALC' M2=2.0*(M2-M1)
'PRINT/FORM=C,LABR=1' HV,M2 S 2X,0,10.3
'CALC' PSIG=(M2.GE.3.841)
'RUN'
'GOTO' L17*PSIG
'CAPTION'
''THE CHI-SQUARE STATISTIC IS NOT SIGNIFICANT AT 5% LEVEL''
'GOTO' L18
'LABEL' L17
'CAPTION'
''THE CHI-SQUARE STATISTIC IS SIGNIFICANT AT 5% LEVEL''
'LABEL' L18
'RUN'
'PAGE'
'CONTOUR/CI=C,UCO=TU,LCO=TL,NR=N,NC=M,PV=MC' RES
'RUN'
'CAPTION'
''THE CONTOUR PLOT VERTICAL AXIS IS LAMDA2''
'LINE' 1
'CAPTION'
''THE HORIZONTAL AXIS IS LAMDA1''
''                                                                    ''
''THE POWER PARAMETER GRID VALUES THAT ARE INSIDE THE                 ''
'' 95% CONFIDENCE INTERVAL(WHEN LOCATION PARAMETER IS ZERO)           ''
''ARE CALCULATED AND OUTPUT.                                          ''
''                                                                    ''
'DEVA' K:F:TTY(1...M)
'SCAL' CM:MCRIT:CCM=M:ULAM
'CALC' MCRIT=M1-1.9205
'VARI' CLAMSM :YL
'RUN'
'FOR' CDUM=1...M
'CALC' CM=CDUM
'GOTO' LC*(ELEM(LM3;CM).LT.MCRIT)
'CALC' CCM=CCM+1
:ELEM(CLAM;CCM)=ELEM(LB1;CM)
'LABEL' LC
'REPEAT'
'RUN'
'VARI' CLA SCCM
'EQUATE' CLA=CLAM
'PRINT' MLC
'PRINT' MLC2
'PRINT/LABC=1' CLA $6.2
'LINE' 2
'SCAL'  MICL:MACL
'CALC' MICL=MIN(CLA) :MACL=MAX(CLA)
'CALC' MACL=MACL*(ABS(MACL).GT.0.001)
'CALC' MICL=MICL*(ABS(MICL).GT.0.001)
'CALC' ULAM=(MICL.LE.LSEN).AND.(LSEN.LE.MACL)
:ULAM=LSEN*ULAM+(1-ULAM)*LAMB1(ML)
:TTY(1)=Y/GM(1)
''                                                                    ''
''THE MACROS ONEPARS AND TWOPARS PRODUCE THE ANOVA'S FOR THE          ''
''DATA TRANSFORMED BY ONE AND TWO PARAMETERS RESPECTIVELY             ''
''THE MACRO TWOPARS IS EXECUTED ONLY IF THE CHI-SQUARE STATISTIC      ''
''IS SIGNIFICANT AT THE 5% LEVEL(DETERMINED BY PROGRAM).              ''
''                                                                    ''
'MACRO' ONEPARS
'LOCAL' LC3,LC4,LC7
'CAPTION'
''POWER PARAMETER USED IN TRANSFORM OF OBSERVED VALUES IF
 TRANSFORMATION INVOLVES POWER PARAMETER ONLY.THIS WILL BE WHEN
 THE CHI-SQUARE STATISTIC IS NOT SIGNIFICANT''
'LINE' 2
'PRINT/FORM=C,LABR=1' MLT,ULAM $2X,0,7.3
'GOTO' LC3*(ULAM.EQ.0)
'CALC' YL=GM(1)*(TTY(1)**ULAM)/ABS(ULAM)
'CALC' TTY(2)=Y**ULAM
'GOTO' LC4
```

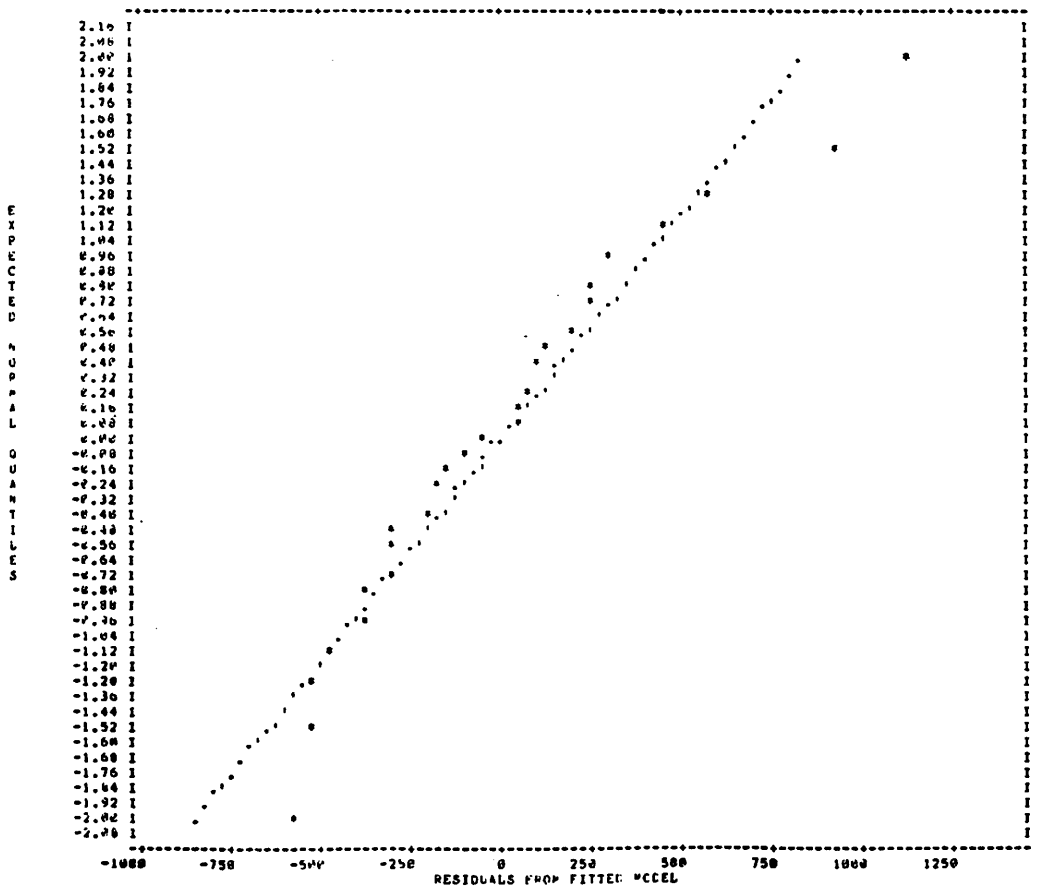
```
'LABEL' LC3 'CALC' YL=LOG(Y)
'CALC' TTY(2)=LOG(Y)
'LABEL' LC4
'CAPTION'
''*****************************************************************''
'LINE' 2
'CAPTION'
''ANOVA OF VALUES TRANSFORMED BY POWER PARAMETER ONLY''
'LINE' 1
'CAPTION'
''    ADDITIVE MODEL    ''
'ANOV' YL;RES=R;FVAL=F
'USE' RYLS
'CAPTION'
''*****************************************************************''
'GOTO' LC7*(RE.EQ.1)
'TREAT' LPF
'LINE' 3
'CAPTION'
''    MODEL WITH INTERACTION TERM(S)  ''
'ANOV' YL
'LABEL' LC7
'ENDM'
'RUN'
'MACRO' TWOPARS
'LOCAL' LC5,LC6,LC8
'CALC' TTY(1)=Y+EL2
:TTY(2)=TTY(1)
:GM(1)=MEAN(LOG(TTY(1))) :GM(1)=EXP(GM(1))
:TTY(1)=TTY(1)/GM(1)
'GOTO' LC5*(BL1.EQ.0)
'CALC' YL=GM(1)**(TTY(1)**BL1)/ABS(BL1)
'CALC' TTY(2)=TTY(2)**BL1
'GOTO' LC6
'LABEL' LC5
'CALC' YL=LOG(TTY(1))
'CALC' TTY(2)=LOG(TTY(2))
'LABEL' LC6
'LINE' 3
'CAPTION'
''ANOVA OF VALUES TRANSFORMED BY BOTH PARAMETERS''
'TREAT' LTF

'CAPTION'
''      ADDITIVE MODEL    ''
'ANOV' YL;RES=R;FVAL=F
'USE' RYLS
'LINE' 2
'CAPTION'
''*****************************************************************''
'GOTO' LC8*((RE.EQ.1).OR.(NF.EQ.1))
'TREAT' LPF
'CAPTION'
''    MODEL WITH INTERACTION TERM(S)  ''
'ANOV' YL
'LABEL' LC8
'ENDM'
'RUN'
'MACRO'
```

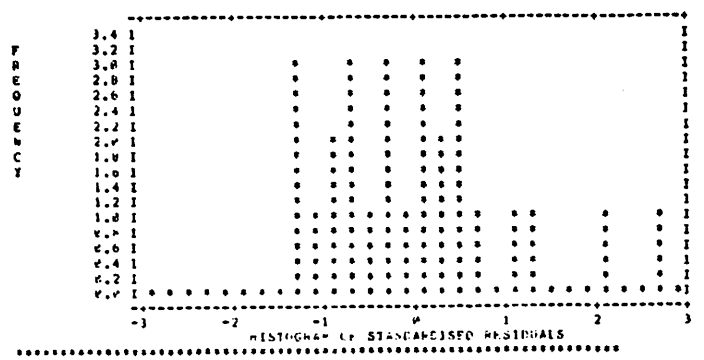
```
'GOTO' L15*PSIG
'USE/R' CHEPARS
'GOTO' L16
'LABEL' L15
'USE/R' TWOPARS
'LABEL' L16
''                                                                    ''
''THE PROGRAM DECIDES FROM THE VALUE OF PRIV WHETHER OR NOT THE       ''
''USER WISHES THE DATA,TRANSFORMED BY THE OPTIMAL TRANSFORMATION,''
''TO BE OUTPUT.MACRO DPRINS PRINTS THE TRANSFORMED DATA IF REQUIRED''
''                                                                    ''
'MACRO' DPRINS
'LOCAL' SYT(1...NU)
'SCAL' SYT(1...NU)
'EQUATE' SYT(1...NU)=TTY(2)
'CAPTION'
''LIST OF SCALED TRANSFORMED DATA VALUES AS READ FROM DATA FILE.THE
  TRANSFORM USES BOTH PARAMETERS ONLY IF THE CHISQUARE STATISTIC
  FOR LAMBDA2 IS SIGNIFICANT.''
'LINE' 4
'PRIN1/FORM=P,LABR=1' SYT(1...NU) $8.3
'ENDM'
'RUN'
'GOTO' L27*(PRIV.LT.1)
'USE/P' DPRINS
'LABEL' L27
'RUN'
'CLOS'
'STOP'
```

Specimen Output

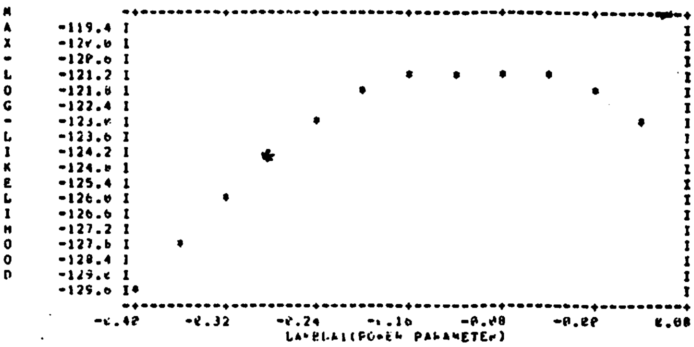
```
GENSTAT V     MARK 4.03
(C)1980 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STATION)



BOX&COX 3 FACTOR TEXTILE EXAMPLE



      Y      MINIMAX       861.3333       90.0000       3636.0000    27  VALUES     0 MISSING  **



ANOVA OF DATA VALUES : ADDITIVE MODEL



***** ANALYSIS OF VARIANCE *****

VARIATE: Y

SOURCE OF VARIATION                  DF          SS        SS%           MS         VR

EXUNIT STRATUM
   A(1)                               2     8182571      40.45      4091285     17.524
   A(2)                               2     5625137      27.81      2812566     12.047
   A(3)                               2     1752508       8.66       876254      3.753
  RESIDUAL                           20     4669241      23.08       233462
TOTAL                                26    20229456     100.00       778056

GRAND TOTAL                          26    20229456     100.00

GRAND MEAN                                861
TOTAL NUMBER OF OBSERVATIONS               27

***** TABLES OF MEANS *****

VARIATE: Y

    GRAND MEAN       861

          A(1)         1          2          3
                     261        742       1581

          A(2)         1          2          3
                    1489        679        417

          A(3)         1          2          3
                    1156        894        534

***** STANDARD ERRORS OF DIFFERENCES OF MEANS *****

TABLE                      A(1)           A(2)           A(3)
-----------------------------------------------------------------
REP                           9              9              9
SED                       227.8          227.8          227.8

***** STRATUM STANDARD ERRORS AND COEFFICIENTS OF VARIATION *****

STRATUM                    DF            SE          CV%

EXUNIT                     20         483.2         56.1
```


RESIDUALS FROM FITTED MODEL
FITTED VALUES


EXPECTED NORMAL QUANTILES
RESIDUALS FROM FITTED ACCEL


FREQUENCY
HISTOGRAM OF STANDARDISED RESIDUALS

```
LIST OF GRIDVALUES OF LOCATION PARAMETER(LAMDA2)

  -80.00
  -70.00
  -60.00
  -50.00
  -40.00
  -30.00
  -20.00
  -10.00
    0.00
   10.00
   20.00
   30.00
   40.00
   50.00


LIST OF GRID VALUES OF POWER PARAMETER(LAMBDA1)


 -0.400
 -0.360
 -0.320
 -0.280
 -0.240
 -0.200
 -0.160
 -0.120
 -0.080
 -0.040
  0.000
  0.040
```

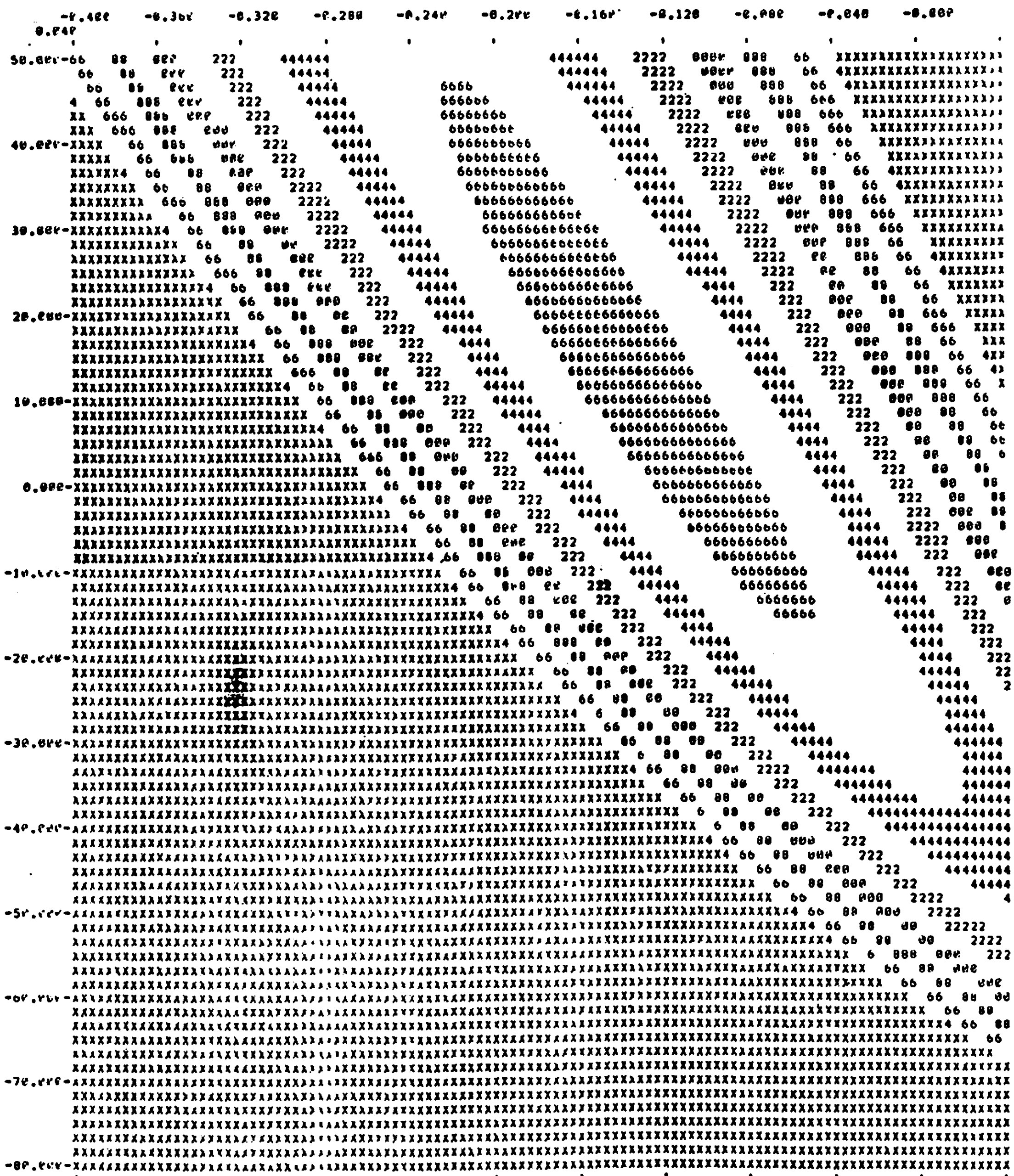



```
PLOT OF MAXIMISED LOG-LIKELIHOOD AGAINST POWER PARAMETER

   WHEN LOCATION PARAMETER IS ZERO


************************************************************


LARGEST NORMAL ADDITIVE LOG-LIKELIHOOD OVER

 TWO TRANSFORMATION PARAMETERS

 TOGETHER WITH ITS ELEMENT NUMBER IN GRID OF PARAMETERS
  -120.661     116

BEST VALUES OF POWER & LOCATION PARAMETERS
  POWER PARAMETER IS -0.12 LOCATION PARAMETER IS   10.00

LARGEST NORMAL ADDITIVE LOG-LIKELIHOOD OVER

 POWER PARAMETER VALUES ONLY WITH ITS ELEMENT NUMBER

 IN LIST OF POWER PARAMETER VALUES
  -120.915      8


BEST VALUE OF POWER PARAMETER WITH ZERO LOCATION PARAMETER IS -0.12


  CHI-SQUARE STATISTIC FOR LAMBDA2 TO BE ZERO ON ONE DF IS      0.110


THE CHI-SQUARE STATISTIC IS NOT SIGNIFICANT AT 5% LEVEL
```

****CONTOUR PLOT OF RES AT INTERVALS OF 0.2610****

SCALED VALUES AT GRID POINTS

-473.6316	-469.0886	-466.9740	-465.0106	-464.1071	-464.2905	-465.5539	-467.0299	-471.0080	-474.9552	-479.5261	-484.5842
-476.7438	-472.3410	-468.7138	-466.0152	-464.3733	-463.8710	-464.5309	-466.3111	-469.1131	-472.7999	-477.2162	-482.2072
-480.5023	-475.4576	-471.1352	-467.6486	-465.1999	-463.9150	-463.8607	-465.0323	-467.3537	-470.6936	-474.8878	-479.7637
-485.1635	-479.4460	-474.3433	-470.0352	-466.7152	-464.5521	-463.6646	-464.0977	-465.6120	-468.6935	-472.5765	-477.2718
-490.0310	-484.2694	-478.4807	-473.3059	-469.0058	-465.9379	-464.0955	-463.6455	-464.6014	-466.8640	-470.3379	-474.7628
-497.0011	-490.1270	-483.9901	-477.0041	-472.4172	-468.2553	-465.3434	-463.0556	-463.6780	-465.3874	-468.2575	-472.2087
-504.5720	-497.0072	-489.0700	-483.0500	-476.5558	-471.7178	-467.6300	-464.9605	-463.0552	-464.3003	-466.4606	-469.9354
-513.3255	-505.3303	-497.0017	-489.5499	-482.0952	-476.5697	-471.2576	-467.2551	-464.6208	-464.1160	-465.1655	-467.0425
-523.5495	-515.0009	-506.0600	-496.4195	-490.4937	-483.1071	-476.5380	-471.1112	-467.1597	-464.9598	-464.6027	-466.2400
-535.0057	-526.6835	-517.7334	-508.0313	-500.0959	-491.7032	-483.8989	-477.0018	-471.3652	-467.4258	-465.4218	-465.5043
-550.1069	-540.7207	-531.2430	-521.7200	-512.2295	-502.0575	-493.9166	-485.5624	-478.1992	-472.2580	-468.1712	-466.2672
-566.2298	-556.2999	-546.2445	-536.0005	-527.6473	-517.6196	-507.5253	-497.7680	-488.0521	-480.5844	-474.0717	-469.6213
-592.6164	-581.9247	-571.1029	-560.1463	-549.0025	-537.8796	-526.6604	-515.5225	-504.6605	-494.4002	-485.2032	-477.6342
-631.6455	-619.5012	-607.2705	-594.9352	-582.4794	-569.0944	-557.1067	-544.3903	-531.5071	-518.9367	-506.7300	-495.4177

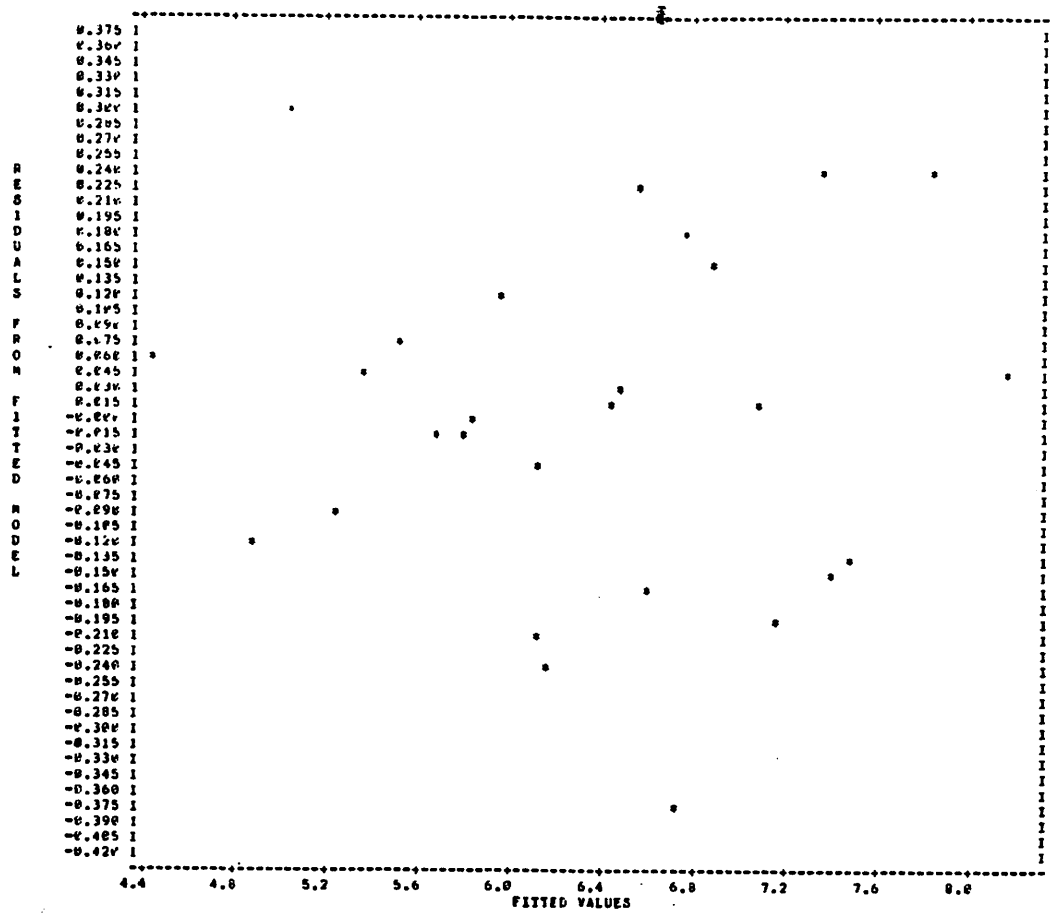


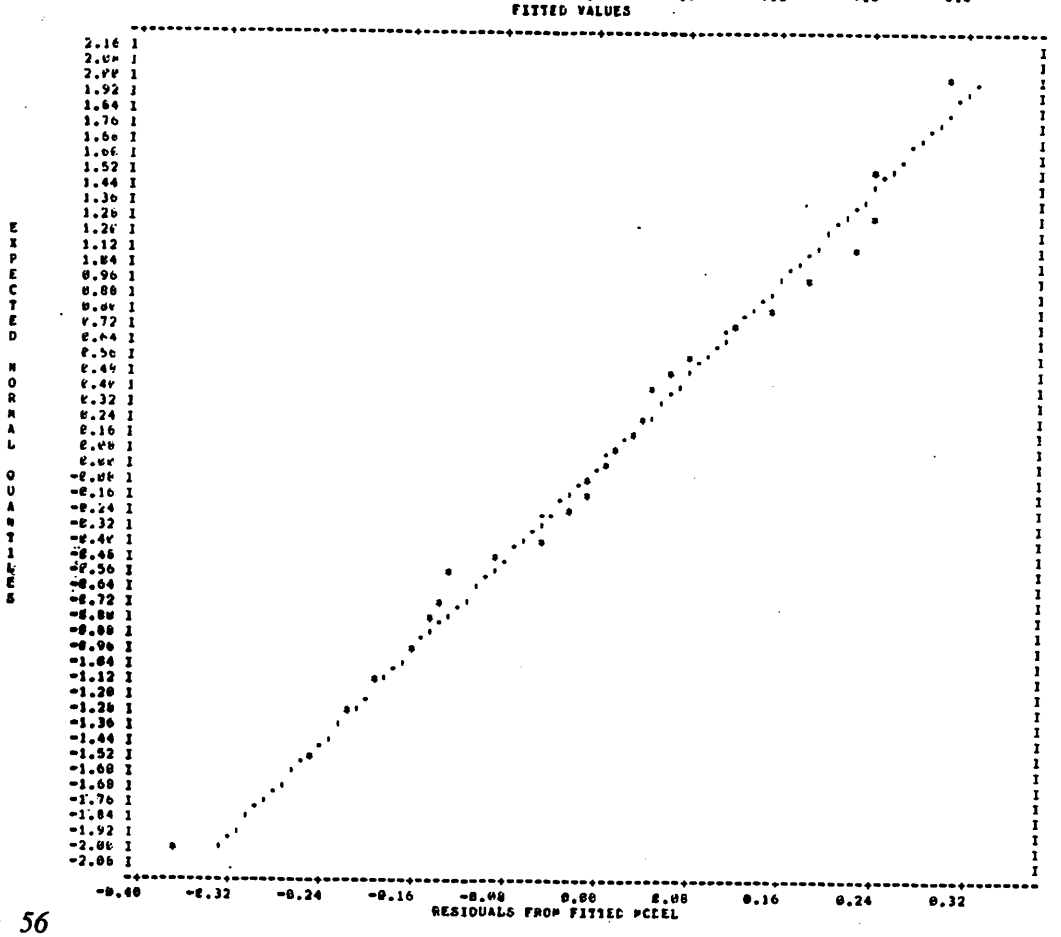
THE CONTOUR PLOT VERTICAL AXIS IS LAMDA2

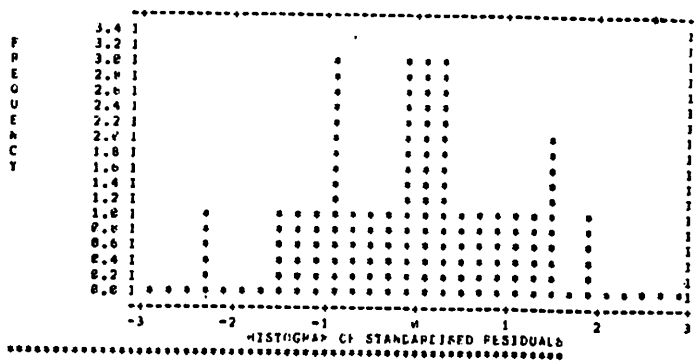
THE HORIZONTAL AXIS IS LAMDA1

```
 LIST OF POWER PARAMETER VALUES IN 95% CONFIDENCE INTERVAL


WHEN LOCATION PARAMETER IS ZERO

 -0.20
 -0.16
 -0.12
 -0.08
 -0.04
  0.00
```

```
POWER PARAMETER USED IN TRANSFORM OF OBSERVED VALUES IF
 TRANSFORMATION INVOLVES POWER PARAMETER ONLY.THIS WILL BE WHEN
 THE CHI-SQUARE STATISTIC IS NOT SIGNIFICANT

  POWER PARAMETER IS  0.000
**************************************************************

ANOVA OF VALUES TRANSFORMED BY POWER PARAMETER ONLY

   ADDITIVE MODEL


***** ANALYSIS OF VARIANCE *****

VARIATE: YL

SOURCE OF VARIATION                DF          SS        SS%          MS          VR

EXUNIT STRATUM
   A(1)                             2    12.51561      53.94     6.25780     174.401
   A(2)                             2     7.17023      30.90     3.58511      99.915
   A(3)                             2     2.80023      12.07     1.40011      39.020
  RESIDUAL                         20     0.71703       3.09     0.03585
TOTAL                              26    23.20309     100.00     0.89242

GRAND TOTAL                        26    23.20309     100.00

GRAND MEAN                              6.335
TOTAL NUMBER OF OBSERVATIONS              27

***** TABLES OF MEANS *****

VARIATE: YL

    GRAND MEAN    6.335

         A(1)         1         2         3
                  5.474     6.392     7.138

         A(2)         1         2         3
                  6.974     6.319     5.712

         A(3)         1         2         3
                  6.765     6.388     5.920

***** STANDARD ERRORS OF DIFFERENCES OF MEANS *****

TABLE                    A(1)        A(2)        A(3)
-------------------------------------------------------
REP                         9           9           9
SED                    0.0893      0.0893      0.0893

***** STRATUM STANDARD ERRORS AND COEFFICIENTS OF VARIATION *****

STRATUM                  DF          SE         CV%

EXUNIT                   20      0.1894         3.0
```


RESIDUALS FROM FITTED MODEL
4.4
4.8
5.2
5.6
6.0
6.4
6.8
7.2
7.6
8.0
FITTED VALUES


EXPECTED NORMAL QUANTILES
-0.40
-0.32
-0.24
-0.16
-0.08
0.00
0.08
0.16
0.24
0.32
RESIDUALS FROM FITTED MODEL




```
LIST OF SCALED TRANSFORMED DATA VALUES AS READ FROM DATA FILE.THE
  TRANSFORM USES BOTH PARAMETERS ONLY IF THE CHISQUARE STATISTIC
  FOR LAMBDA2 IS SIGNIFICANT.

  6.513   5.914   5.677   5.823   5.563   5.347   5.136   4.771   4.582   7.254   7.088   6.452   6.938   6.439   6.682
  6.891   5.805   5.394   8.199   8.066   7.021   7.358   6.975   6.335   7.239   6.784   5.886

$24  'CLOS'
```

References

Box, G.E.P. and Cox, D.R. (1964) An Analysis of Transformations. JRSS (B), **26**, 211–252.

Snedecor, G.W. and Cochran, W.G. (1973) Statistical Methods, 6th Edition, (section 11.17). Ames Iowa Press.

G.V. Stevens
Department of Mathematics
Liverpool Polytechnic
Byrom Street
Liverpool L3 3AF
United Kingdom

FITTING GENERAL MODELS WITH OPTIMIZE

The existence of OPTIMIZE, which allows specification of a very wide range of non-linear models and error distributions, does not absolve users from thinking about the models, how they should be specified and whether they are appropriate for the data. The 'black-box' approach may be sufficient for the single occasion, the user's time being more valuable than that of the computer, but in the long run this approach can be wasteful and frustrating.

As originators of OPTIMIZE and also of the Maximum Likelihood Program (MLP), we are often asked to explain why the programs do not seem to be working properly. At times, these queries have uncovered genuine deficiencies in the programs but quite often the problem is in the model itself. In an ideal program, diagnostic tests might be devised for all the examples to be discussed here. Meanwhile, potential users may find them instructive.

Valid Models

The rules of model specification are very simple; MODEL specifies calculations leading to a variate E(Y) of predicted means given the parameter values (and usually some independent variates) and the options of OPTIMIZE specify the error distribution of the observation variate Y about these means, with associated weights where appropriate. The commonest application is non-linear regression (curve-fitting) with independent Normal errors. What, then, is meant by an 'invalid' model if the Genstat syntax is adhered to?

Assuming that E(Y) is correctly described both in MODEL and OPTIMIZE and the parameters listed in OPTIMIZE occur on the right-hand side in the MODEL, they may still fail to produce a variate that can be matched with Y. The following problems tend to occur:

Inactive parameters

One or more of the parameters may not be mentioned in the model, or may produce no effect on E(Y), for the data being analysed. For example, the expression Z**P for parameter P has no effect if Z only takes the values 0 or 1. Models using logical variables, (X.LT.P) for example, may be unaffected by P if the condition is always satisfied (P greater than all X).

Redundant parameters

One or more parameters can be cancelled from the model. This often occurs when the user writes an expression from a physical or chemical textbook, say, which is correct for evaluating the model given all the parameters, but not appropriate for fitting a model. For example

```
E(Y) = (P1+P2*X)/(P3+P4*X)
```

has only three *independent* parameters; one parameter must be held constant, (e.g. P3 = 1). A recent example which caused unnecessary trouble was something like

```
E(Y) = P1*EXP(P2/P3) + P2*(P3**P4)*X
```

Of course this was really just a straight line in X, with only two parameters, and the program played about with endless combinations of four parameters always getting the same fit but never converging, because there was no single set of optimum parameter values.

Inactive data

Some data values contribute virtually nothing extra to the log likelihood or sum of squares. The cause may be (a) a weighting variate which gives a wide range of weights, so that the smallest are virtually ignored, (b) replicated data where there are too few

different x-values (e.g. an attempt to fit a quadratic to replicate data for only 2 x's), (c) formulae where certain terms cancel for some data values.

Incorrect ranges of values

When data are essentially *non-negative* or perhaps in range (0,1), not only should the *distribution* be appropriate (Poisson, Gamma, Binomial, etc.) but the values of E(Y) should also lie in the range. It may be necessary to control the parameter bounds so that E(Y) always lies within range. The program may proceed, by treating invalid means as missing observations, but the ultimate responsibility is with the user. Strictly speaking, a model is invalid if it is possible to simulate data out of range. Positive variables should not use Normally distributed errors, because the combination of small mean and large negative error could lead to impossible values. In practice, this advice is usually ignored and the results are acceptable if data values are not close to the limit of the range.

Overflow and underflow

These occur in models with exponential or power terms. For example, the logistic function

```
E(Y) = P1/(1+EXP(-P2*(X-P3)))**P4
```

can experience overflow for quite modest values of X and the parameters. Try it on your pocket calculator!

Parameter Transformations

The same numerical results can be obtained in a number of ways. For example the models

```
E(Y) = P1*EXP(-P2*X)
```

and `E(Y) = Q1*EXP(-Q2*(X-XA))`

are the same, although the parameters are different. It can be shown that `Q2 = P2` and `P1 = Q1*EXP(Q2*XA)`. All we have done is to take off a working mean, XA from the variate X. The parameter Q1 is the height of the curve at X=XA, whereas P1 is the height at X=0. Now if `X = 0` is outside the range of data X but XA is close to the mean value or mid-range, it will be found that it is much easier to fit the Q parameters, which are nearly independent, than the P parameters which are correlated in a non-linear manner. An even better transformation, based on the values at two interior points, XA and XB, is

```
E(Y) = R1*EXP((X-XA)*LOG(R2/R1)/(XB-XA))
```

This looks very complicated algebraically but is very satisfactory numerically. If parameters P1 and P2 are needed, they may be evaluated when the model has been fitted.

The principle is the same as that used in fitting polynomials to data. Algebraic polynomials are easy to write down but orthogonal polynomials are easier to fit.

Most models are more complicated than the simple one given here but a few moments thought can often save a wasted computer run. Any transformation which reduces the dependence between parameters is likely to be useful.

Linear parameters

When a model is expressed as a linear combination of non-linear terms and errors are Normal distributed, only the non-linear parts of the model need be written. This means that some of the parameter fitting may be done by linear regression within the optimisation routine. This not only saves time and improves accuracy but also avoids the need for the user to suggest suitable values for the linear parameters. A well known example is the exponential curve

```
E(Y) = B1 + B2*EXP(-P1*X)
```

where, if P1 is known, the model is linear with parameters B1 and B2. The difference in performance between the one parameter fit and the three parameter fit is very remarkable.

G.J.S. Ross
Statistics Department
Rothamsted Experimental Station
Harpenden
Hertfordshire AL5 2JQ
United Kingdom

PROCESSING THE RESULTS OF MULTIVARIATE ANALYSES

[This is intended to be read in conjunction with the next article.]

Problems arise when the results of multivariate analyses are being processed, because of the two methods of representing a data matrix: a matrix with n rows and p columns, or a set of p variates of length n (here the term 'data matrix' is used generally to include any units × variates situation, for example the matrix of principle component scores). The EQUATE statement which copies from one form to the other needs a format that is (at best) messy; sometimes the use of the format is realistically unavoidable, however it is possible to sidestep it in some situations.

The basic idea, which allows us to omit the awkward formats, is to work in terms of transposed matrices where possible. A single example, based on canonical variate analysis (CVA), should be sufficient to show this device; however the applicability is wider. There are one or two extra points that I leave to the end.

Suppose that we have a data matrix of n rows and p columns, the n units are grouped into g classes and we will restrict our interest to the first k dimensions of the results from the analysis. IN Genstat we have

```
'SCALARS' N,P,G,K
```

and we will need subsequently a scalar with value $(k-1)$, e.g.

```
'SCALAR'  KM1
```

The data matrix is held as a set of variates and the grouping into g classes is held as a factor

```
'SET'       XV = XV(1...P)
'VARIATES'  XV $ N
'FACTOR'    F $ G,N
```

and it is assumed that these have been given values. For the algebra that occurs below the other form of the data matrix is denoted by the matrix X $(n \times p)$.

The statements needed to perform the CVA and save the coordinates of the group means in the canonical variate space (M), the latent vectors in L, roots in R and trace in T, are as follows:

```
'DSSP'      W $ XV; F
'MATRIX'    M $ G,K
   :        L $ P,K
'DIAGMAT'   R $ K
'SCALAR'    T
'SSP'       W
'CVA'       W; L,R,T; SCORES=M
```

To plot the group means, we require a set of k variates of length g which must be copied from the matrix M; here a format for EQUATE is used.

```
'SET'       MV = MV(1...K)
'VARIATES'  MV $ G
'EQUATE'    MV = M $ (1,(X)KM1)G,X
```

It is normally useful to plot the positions of the individual units in the canonical variate space (superimposing these on the graphs of the variates MV). The matrix of these canonical variate scores, S $(n \times k)$, is formed in two steps: first we obtain the rotation of axes S = XL; then S is 'centered', i.e. its column means are subtracted from their respective columns. Of course, the columns of S will need to be put into a set of variates for use in plotting, so cumbersome formats will be required both to do this and to get X from the set of variates XV.

Instead of working with X and S we use their transposes, X' $(p \times n)$ and S' $(k \times n)$. We have S' = L ' X' (from S = XL) which translates easily into the following Genstat statements.

```
'MATRIX'    XT  $ P,N
   :        ST  $ K,N
'EQUATE'    XT  = XV
'CALC'      ST  = TPDT(L;XT)
```

To centre S we need to subtract row means from S', however this is easier to do after the rows of S' (columns of S) have been put into the variates needed for plotting.

```
'SET'       SV = SV(1...K)
'VARIATES'  SV $ N
'EQUATE'    SV = ST
'CALC'      SV = SV - MEAN(SV)
```

Now a typical GRAPH statement might be

```
'GRAPH/EQXY=Y,37,61' MV(2),SV(2); MV(1),SV(1); *,F
```

which will plot the coordinates of the group means, as asterisks, and the individuals, using the grouping factor, in the first and second dimensions of the canonical variate space.

I conclude with a few general comments. The device illustrated is to transpose the matrix algebra; it is particularly useful for multi-stage analyses, i.e. when several analyses are being strung together. In this case you may need to switch quite often between the two forms of data matrix: a set of variates, for plotting (and printing), and a matrix, for the algebra.

You must be careful to centre matrices at the right time, for example with principal components analysis it is X which must be centred, not S. Centering is easiest when the data matrix is held as a set of variates, as in 'double-centering', where row means and then column means (or vice versa) are subtracted.

```
'CALC'  XV  =  XV-VMEAN(XV)
   :    XV  =  SV-MEAN(XV)
```

I hope that some users of the multivariate directives will benefit from these comments.

P.G.N. Digby
Statistics Department
Rothamsted Experimental Station
Harpenden
Hertfordshire AL5 2JQ
United Kingdom

PLOTTING VARIABLES IN PAIRS

The main purpose of this article is to pass on some hints for use when plotting the results of a multivariate analysis. However some points have a wider application, hence the more general title.

Most results from the multivariate directives are returned in matrices, for plotting purposes we need a set of variates containing the columns of the matrix, e.g. if the matrix has N rows and R columns we need (up to) R variates of length N. Most people have seen something like:

```
'EQUATE'    VSET = MAT $ (1,(X)RM1)N,X
```

where RM1 equals (R-1) and must be known at compile time. This requirement is tiresome, especially if it occurs within a macro since an inner block is required to calculate RM1. Dynamic formatting is allowed in Genstat 4.04 (see Lane, Payne and Simpson) which eases the problem. Alternatively, we can transpose the matrix into an R×N matrix and equate directly from the transpose. In general, I suspect that the use of a dynamic format is best.

Having obtained the set of R variates there is now the problem of getting all the R(R-1)/2 graphs of each variate with each other variate. If you are not too bothered about the order in which the graphs are produced you can use the device suggested by Gower, the gist of which is:

```
'SCALAR'    JGM,JGP,JGQ,JGR
'CALC'      JGQ=3-(JGP=R-2*(JGR=R-1-(JGM=INTPT(R/2))))
'RUN'
'FOR'       YD=V(JGP...R);XD=V((JGQ...R)JGR,(1)JGM)
'GRAPH'     YD;XD
'REPEAT'
```

The four scalars correspond to scalars M, P, Q and R in the original article; I have condensed their calculation into one statement (note that there is a typographic error in the first CALC statement in the original). I dislike the rather odd order of the graphs from this method so I routinely uses a less subtle approach based on the following:

```
'FOR'       YD = V(1...R)
  :         XD = V(1...R)
'GOTO'      FORGETIT * (XD.IS.YD)
'GRAPH'     YD;XD
'REPEAT'
'LABEL'     FORGETIT
'REPEAT'
```

Note the use of the .IS. operator to test the equality of the identifiers pointed to by the pair of dummies. As can easily be confirmed, this produces graphs in the order V(2,3,3,4,4,4) against V(1,1,2,1,2,3) etc.

The final paragraphs deal with the GRAPH statement itself. Even with the latter form of generating the graphs it is easy to lose track of which variates are included in each graph. This is easiest to remedy by putting axis titles on the graphs. By declaring headings HDIM(1...R), with e.g. HDIM(2)=''DIMENSION TWO'', parallel lists can be put on both loops, e.g. 'FOR' YD=V(1...R);YHD=HDIM(1...R), and the ATY=YDH and ATX=XHD options of GRAPH can be set. Within macros, it is best to declare a large number of such headings to cater for differing settings of R. I normally declare HDIM(1...6) which should be adequate (note that 15 graphs are produced if one goes up to 6 dimensions).

More care must be taken when plotting graphs of the dimensions of the results of a multivariate analysis than when a set of observed variables is plotted. Interpretation of multivariate results varies according to the type of analysis. However, many methods, e.g. PCP, PCO, CVA and ROTATE produce results where the relationship between two points depends on the (mathematical) distance between the points. Thus it is important that the scales of the axes, e.g. 3 units/inch, are the same, so that the same physical distance in any direction represents the same mathematical distance. This is arranged by making the graph square and setting the EQXY=Y option. Most printers give 10 characters per inch (across) and 6 lines per inch (down), but this may vary (beware). To get graphs that are t inches square set the NRF and NCF options to $(6t+1)$ and $(10t+1)$, respectively. I nearly always use $t=6$ or $t=10$.

To cater for these two possibilities I extend the use of the scalar (normally called GRAPHOPT) which indicates the number of dimensions I want to plot: now its absolute value gives the dimensionality and its sign indicates the graph size. Using two scalars, NRF and NCF, we have

```
'CALC' NRF,NCF=37,61+24,40*(GRAPHOPT.GT.0)
```

and the option settings are NRF=NRF and NCF=NCF.

The other way to ensure that mathematical and physical (paper) distances correspond correctly is to work out minima and maxima and use the BV option. This is more complicated and usually results in awkward scales on the axes; however, it is the only way to keep the same scales on all graphs – they tend to differ with the previous method.

It is useful to label the points uniquely; if Genstat realises that this has been done it will list coincident points below the graph, rather than printing a colon on the graph. By declaring 'FACTOR' GFAC $ N=1...N we can use GFAC as the final list on the graph statement to specify this.

By this stage the GRAPH statement begins to look horrendous. Careful use of the rules of Genstat syntax, and the knowledge that the five options given above (excluding BV which is the third option) are first (ATY), second (ATX), seventh (EQXY), eighth (NRF) and ninth (NCF), allows the form

```
'GRAPH/YHD,XHD,EQXY=Y,NRF,NCF' YD;XD;GFAC
```

which is not too daunting.

I hope that these comments will be received in the spirit in which they were written, rather than the response that I might expect from (some) computer managers, 'Oh no, not more ways of generating scrap paper'!

References

Gower, J. (1980) Counting over Triangular Arrays. Genstat Newsletter, **6**, 9–10.

Lane, P.W., Payne R.W. and Simpson, H.R. (1982) Genstat 4.04. Genstat Newsletter, **10**, 7–12

P.G.N. Digby
Statistics Department
Rothamsted Experimental Station
Harpenden
Hertfordshire AL5 2JQ
United Kingdom

GENSTAT ERROR NOTICE S4010

The following changes should be made to Error Notice S3010 (Newsletter No. 10):

Page 48 Line −23 should read `DO  20  I=1,NVZ`

Page 49 Line 10 should read `KSYM(1) = KCH(56)`
Lines 30 and 34 should be flagged with '*'.

ADDITIONS TO THE NOTICE BOARD

***** ERROR ***** 23/03/83

'FIT' ETC. RESIDUALS FROM A GENERALISED LINEAR MODEL MAY BE WRONG. WHEN THE LINK FUNCTION IS THE INVERSE, OR A POWER FUNCTION WITH NEGATIVE EXPONENT, THE SIGN OF THE RESIDUALS IS REVERSED.

(SEE ERROR RG7)

***** ERROR ***** 23/03/83

'FIT' ETC. SOME RESULTS ARE NOT SAVED CORRECTLY IF PRINTING IS SUPPRESSED AT THE SAME TIME AS REORDERING COEFFICIENTS. IF THE OPTIONS PRINT=Z AND ORDER=MAX ARE SET, COVARIANCE MATRICES AND COEFFICIENTS SAVED WITH KEYWORDS VCOV AND COEF ARE WRONG.

(SEE ERROR RG8)

***** ERROR ***** 23/03/83

'OPTIMISE' IF THE LIKELIHOOD IS MULTINOMIAL (LIK=7) THE WRONG RESIDUAL DEGREES OF FREEDOM ARE PRINTED: THE NUMBER PRINTED SHOULD BE REDUCED BY 1.
IF THERE ARE NO RESIDUAL DEGREES OF FREEDOM AND A LIKELIHOOD WITH NO SCALE PARAMETER IS SPECIFIED (LIK=4,5,6,7), THEN NO STANDARD ERRORS ARE PRINTED EVEN THOUGH THEY ARE ESTIMABLE.
IF THERE ARE MISSING VALUES IN A VARIATE OF UPPER BOUNDS, SPECIFIED WITH KEYWORD UPPER, THE STATEMENT WILL FAIL WRONGLY WITH A DIAGNOSTIC OP 11 (MEANING INITIAL PARAMETER VALUE IS GREATER THAN UPPER LIMIT).

(SEE ERROR OP3)

***** ERROR ***** 23/03/83

'OPTIMISE' IF 'RESTRICT' IS USED WITH 'OPTIMISE' THEN VARIATES OF RESIDUALS OR FITTED VALUES SAVED WITH KEYWORDS RES OR FVAL ARE ASSIGNED VALUES IN THE WRONG ORDER.

(SEE ERROR OP4)

***** ERROR ***** 23/03/83

'DERIVE' IF PREDICTION ERROR VARIANCES ARE REQUESTED USING FUNCTION PACF, THE LAST VALUE (HIGHEST LAG) IS NOT CALCULATED BUT GIVEN AS A MISSING VALUE.

(SEE ERROR TS4)

***** ERROR ***** 23/03/83

'DERIVE' FUNCTION THACF CAUSES PROBLEMS IF THE OPTIONAL ARGUMENT (TO SAVE THE TIME SERIES VARIANCE) IS NOT SET. THIS RESULTS IN THE TRANSFORMATION PARAMETER OF THE MODEL BEING WRONGLY CHANGED.

(SEE ERROR TS5)

***** ERROR ***** 23/03/83

'ESTIMATE' IF THE TRANSFORMATION PARAMETER OF A MODEL IS UNSET, THIS DIRECTIVE WILL FAIL. IF NO TRANSFORMATION IS REQUIRED ,SET THE PARAMETER EXPLICITLY TO 1.0 IN THE VARIATE OF PARAMETERS OR USE THE 'PRELIM' DIRECTIVE.

(SEE ERROR TS6)

***** ERROR ***** 23/03/83

'PRELIM' THIS DIRECTIVE WRONGLY RESETS THE OBLIGATORY PARAMETERS (CONSTANT, TRANSFORMATION AND VARIANCE) OF A MODEL IF NO VALUES ARE ASSIGNED FOR THEM. THUS, IF THESE PARAMETERS ARE GIVEN EXPLICITLY IN THE PARAMETERS VECTOR AND 'PRELIM' IS USED TO FORM PRELIMINARY ESTIMATES OF THE OTHER PARAMETERS, YOU MUST REDEFINE THE OBLIGATORY PARAMETERS USING KEYWORDS CONST, TRANS AND VAR.

(SEE ERROR TS7)

***** ERROR ***** 23/03/83

'ESTIMATE' IF THE FIX OPTION IS SET, THE REFERENCE NUMBERS OF PARAMETERS ARE NOT CORRECTLY USED TO LABEL THE MATRIX OF PARAMETER CORRELATIONS.

(SEE ERROR TS8)

```
           ***** ERROR *****                                    23/03/83
'FORECAST' THE FUTEREX OPTION FAILS WITH AN INTEGER VECTOR.
           THIS OPTION WILL ONLY WORK FOR FORECASTING WITH
           A SINGLE INPUT VECTOR SERIES: GIVE THE REQUIRED
           CODE FOR THE OPTION IN A SCALAR RATHER THAN IN
           AN INTEGER VECTOR WITH ONE VALUE.

                          (SEE ERROR TS9)


           ***** ERROR *****                                    23/03/83
'ESTIMATE' IF A MODEL CONTAINS NO NON-SEASONAL PARAMETERS,
           THIS DIRECTIVE WILL FAIL.  REDEFINE THE MODEL WITH
           A NON-SEASONAL PART, USING EXPLICIT LAGS IF
           NECESSARY.

                          (SEE ERROR TS10)


           ***** ERROR *****                                    23/03/83
'FORECAST' IF MODELS INVOLVE DIFFERENCING OF ORDER 2 OR MORE,
           FORECASTS WILL NOT BE PRODUCED.

                          (SEE ERROR TS11)



***** GENSTAT 4.03  ERROR NOTICE NO.12   22.3.83
**
**
**
***** ERROR OP3
***** MODULE OP  SUBPROGRAM EXOPT
**
***** (1) ERROR IF THERE ARE MISSING UPPER BOUNDS
**
      J = KPAR(I)
      IF(RDATA(J).NE.RMV) PARINI(I) = RDATA(J)
*     IF(PARMAX(I).EQ.RMV) PARMAX(I) = RLARGE
*     IF(PARMIN(I).EQ.RMV) PARMIN(I) = -RLARGE
      IF(KGRID.GE.2) PARINI(I) = PARMIN(1)
      RDATA(J) = PARINI(I)
      IF(PARINI(I).LT.PARMIN(I).OR.PARINI(I).GT.PARMAX(I)) GO TO 1261
  350 CONTINUE
**
***** (2) WRONG DF PRINTED FOR MULTINOMIAL, LIK=7
**
      ISDATA(KREND) = I
  450 CONTINUE
*     IF(LIK.EQ.7) NDF = NDF-1
      IF(KREND.LT.KRBEG) GO TO 1199
C
**
```

```
***** (3) NO S.E.S PRINTED IF DF=0 BUT LIK=4,5,6,7 OR 8
**
  510 RDEV = FVALUE
      IF(LIK.GT.3) RDEV = RDEV+RDEV
      RMS = RMV
      RR = 0.0
      IF(LIK.EQ.1) RR = 1.0
*     IF(NDF.LT.0.OR.NDF.EQ.IMV) GO TO 520
*     IF(NDF.GT.0) RMS = RDEV/NDF
*     IF(NDF.GT.0) RR = RMS+RMS
      IF(LIK.GT.3.AND.LIK.NE.9) RR = 1.0
C
**
***** END OF ERROR OP3
**
**
***** ERROR OP4
***** MODULE OP  SUBPROGRAM OPFRES
**
***** WRONG RESIDUALS AND FITTED VALUES ARE SAVED
***** WHEN 'RESTRICT' IS USED WITH 'OPTIMIZE'
**
      GO TO 30
   25 Z = RMV
*  30 K = KFIT+I
*     IF(FITTED) RDATA(K) = Z
      IF(Z.EQ.RMV.OR.Y.EQ.RMV.OR.W.EQ.RMV.OR.T.EQ.RMV) GO TO 40
      R = 0.0
**
      GO TO 37
   40 R = RMV
*  45 K = KRES+I
*     IF(RESID) RDATA(K) = R
      IF(.NOT.PRINT) GO TO 50
      CALL NEWLIN(1)
**
***** END OF ERROR OP4
**
**
***** ERROR RG7
***** MODULE RGA SUBPROGRAM FRES
**
***** WRONG RESIDUALS FOR NEGATIVE POWER LINK FUNCTIONS
**
      IF(SCALE) WT = WT/RSCALE
      RDATA(IRES) = (RDATA(IWV)-RDATA(ILP))*SQRT(WT)
C       CHANGE THE SIGN FOR NEGATIVE POWER TRANSFORMATION
*     IF(RPOWER.LT.0.0) RDATA(IRES) = -RDATA(IRES)
  250 CONTINUE
C
  998 FRES = .TRUE.
**
***** END OF ERROR RG7
**
```

```
**
***** ERROR RG8
***** MODULE RGB SUBPROGRAM SETLST
**
***** WRONG COEFFS AND COVARS SAVED IF ORDER=MAX, PRINT=Z
**
   50 CONTINUE
C
C        IF ORDER IS TAKEN FROM MAX. MODEL, GO THROUGH ALL TERMS.
C
*     IF(NORDER) GO TO 80
      NX = NNRC
*     IF(NOXSET) GO TO 80
C        CODE UP TO 70 IS AS CODE IN PRTINT UP TO 96.
      DO 70 ITERM=1,NT
      IBP = IORBP+ITERM
**
***** END OF ERROR RG8
**
**
***** ERROR TS4
***** MODULE TSC SUBPROGRAM TSPACF
**
***** FUNCTION PACF DOESNT SET LAST VALUE OF PEV VECTOR
**
   30 S=S*(1.-A*A)
      IF(S.LE.0)GOTO 1285
      N=P7+K1
*     IF(P6.GE.K1) RDATA(N)=S
      N=W2+K1
      RDATA(N)=A
**
***** END OF ERROR TS4
**
**
***** ERROR TS5
***** MODULE TSA SUBPROGRAM DERIVE
**
***** FUNCTION THACF OVERWRITES TRANSFORMATION PARAMETER
**
  220 CONTINUE
      IF(KODE.GT.4) GO TO 250
*     KVOR = 0
      IF(NUMBER.EQ.2.OR.L(3).EQ.0) GO TO 230
      IF(K3.EQ.K(3)+L(3)) K3 = K(3)
**
***** END OF ERROR TS5
**
```

```
**
***** ERROR TS6
***** MODULE TSC SUBPROGRAM TSTRAN
**
***** UNSET TRANSFORMATION PARAMETER IN 'ESTIMATE' CAUSES FAILURE
**
      J= IDATA(J)+1
      LAMBDA=RDATA(J)
* 2 LINES DELETED
      J = BC+I
      J = IDATA(J)
*     IF(J.EQ.0.AND.LAMBDA.EQ.RMV) LAMBDA = 1.0
*     I2=1
*     IF(LAMBDA.EQ.1.0)I2=0
      IF(J.EQ.1)NBCX=NBCX+1
      IF(J.EQ.0.AND.I2.EQ.1)I1=1
      IF(I1.EQ.0)GOTO 612
      IF(GWSP(P2,2).NE.0) GO TO 1000
      M=SER+NCS*P2
      NCS=NCS+1
  612 J=SRXO+I
**
***** END OF ERROR TS6
**
**
***** ERROR TS7
***** MODULE TSA SUBPROGRAM PRELIM
**
***** 'PRELIM' RESETS FIRST 3 PARAMS EVEN IF CORRESPONDING LISTS UNSET
**
      L3=IDATA(P2+5)
      V3=IDATA(P2+6)
*     X=RDATA(V2+1)
      IF(P5.NE.0)X=RDATA(P5+1)
      IF(X.EQ.RMV)X=1.0
      RDATA(V2+1)=X
*     X=RDATA(V2+2)
      IF(P6.NE.0)X=RDATA(P6+1)
      IF(X.EQ.RMV)X=0.0
      RDATA(V2+2)=X
*     R=RDATA(V2+3)
      IF(P7.NE.0)R=RDATA(P7+1)
      IF(R.EQ.RMV)R=1.0
      V2=V2+3
      B=0
      IF(P1.EQ.0)GOTO 10
      B=ISDATA(V1+1)
      V1=V1+1
      V2=V2-1
   10 SK=1
**
***** END OF ERROR TS7
**
```

```
**
***** ERROR TS8
***** MODULE TSC SUBPROGRAM TSPRIN
**
***** WRONG LABELS FOR CORRELATION MATRIX IF FIX OPTION SET
**
      LABCOL = 3
      KREF = 0
*     IF(KFIX.EQ.0) GO TO 150
C        LABEL ROW AND COLUMN BY INDEX IF PRESENT
*     LABROW = 0
*     LABCOL = 0
*     KREF = DEX
* 150 CALL PRTSYM(0,NPAR,KREF,3,LENLAB,K2INV,2,LENIT)
  200 CALL NEWLIN(1)
      RETURN
**
***** END OF ERROR TS8
**
**
***** ERROR TS9
***** MODULE TSA SUBPROGRAM FORCAS
**
***** FUTUREX OPTION OF 'FORECAST' REJECTS INTEGER VECTOR
**
      IVOR = 0
      IF(IVEC.EQ.0) GO TO (10,20,30,40),I
*     IF(I.EQ.4) ITYPE = 1
      IF(VECSET(IVEC,ITYPE,IMODE,IVAL,IVOR,-1,3).NE.0) GO TO 1000
      IF(I.EQ.4) GO TO 40
      R = RDATA(IVOR+1)
      IF(R.LT.0.0) GO TO 1173
      GO TO (10,20,30),I
   10 NORIGN = R
**
***** 'FORECAST' FAILS IF NO STRUCTURES ARE SAVED
**
      KC = KLISTS+ISDATA(IJ)
      N = ISDATA(KC)
*     IF(N.EQ.0) GO TO 190
      DO 180 J=1,N
      IC = KC+J
      IVEC = ISDATA(IC)
      IF(GETATT(1,IVEC).NE.0) GO TO 1000
      MVPTR(1) = CNMV(1)
      IF(PUTATT(1,IVEC).NE.0) GO TO 1000
  180 CONTINUE
**
***** END OF ERROR TS9
**
```

```
**
***** ERROR TS10
***** MODULE TSC SUBPROGRAM TSMVAL
**
***** MODEL WITH NO NON-SEASONAL PART REJECTED BY 'ESTIMATE'
**
   25 M=4
   30 IF(L1.LT.(K+M)) GOTO 100
*     IF(P4.EQ.0)GOTO 40
      S=ISDATA(V1+4)
      IF(S.LE.0)GOTO 100
**
***** END OF ERROR TS10
**
**
***** ERROR TS11
***** MODULE TSC SUBPROGRAM TSINVO
**
***** FAILURE IN 'FORECAST' IF MODEL TAKES 2 OR MORE DIFFERENCES
**
   20 IF(N.LT.0)GOTO 50
      IF(P5.EQ.0)GOTO 40
      IF(M.EQ.0)GOTO 40
*     ID = 1
*     IF(P.EQ.0) ID = N
*     M = M/ID
*     DO 35 IJ=1,ID
      I5=I5-M
      DO 30 I=1,M
      U=I2+I
      V=I5+I
   30 RDATA(V)=RDATA(U)
      I2=I2+M
*  35 CONTINUE
   40 IF(P6.EQ.0)GOTO 50
      I3=I3+1
**
***** END OF ERROR TS11
**
**
**
***** END OF ERROR NOTICE NO. 12
```

H.R. Simpson
Statistics Department
Rothamsted Experimental Station
Harpenden
Herfordshire AL5 2JQ
United Kingdom

NOTICES

10th and 11th International Time Series Meetings

Special Topics Conference:

on Hydrological, Geophysical and Spatial Time Series:
Toronto, Canada, 1984 August 10–14.

General Interest Conference:

Toronto, Canada, 1984 August 18–21.

Information on these conferences may be obtained from the Convenor (see below). Further abstracts are welcome.

O.D. Anderson, ITSM Convenor
9 Ingham Grove
Lenton Gardens
Nottingham NG7 2LQ
United Kingdom

Genstat Conference 1983

The third International Conference of users of the Genstat statistical package will take place at Orsay (near Paris) France, 1983 October 10–12.

Sessions of selected contributed papers are planned on:

* new facilities and future developments
* macros
* comparisons of Genstat with other packages
* interesting applications of Genstat
* the use of Genstat in education

Participants will be able to view the Genstat video tapes, see demonstrations of the latest Genstat release and consult expert users of the program.

If you would like to receive further information about the Conference, please complete and return the form below.

To R.W. Payne Statistics Department, Rothamsted Experimental Station, Harpenden, Hertfordshire AL5 2JQ, U.K.

I am interested in attending the third Genstat Conference. Please send further information.

Name: _____

Organisation: _____

Address: _____

_____

_____

_____

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
United Kingdom

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

Please note that each subscription to the Newsletter costs £ 5.00 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 (24×). The first contains issues 1 – 6 and each subsequent fiche contains two issues – 7/8, 9/10 etc. Each fiche costs the same as a year's subscription to the Newletter (£5.00).

To: NAG Ltd., Mayfield House, 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 ..... .

(Minimum subscription period: 2 years)

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 to be made payable to the Numerical Algorithms Group Ltd.

** delete one alternative*

 # tick one box