GLIM - GENERALISED LINEAR INTERACTIVE MODELLING

M. A. Toleman.

1. INTRODUCTION

GLIM provides an excellent means for fitting generalised linear models to data. It also doubles as a general statistical programming language.

The program was written in FORTRAN by statisticians from Rothamsted Experimental Station. It is maintained by the Numerical Algorithms Group (NAG) and is available on the CYBER 76, supported by the CSIRO Division of Mathematics and Statistics. GLIM was written for an interactive computing environment but this is masked on the CYBER 76; hence, GLIM runs as batch only.

The GLIM manual is well presented having four sections:

(a) Theory - of generalised linear models

(b) User's Guide - to GLIM

(c) Reference Section

(d) Appendices - summary of notation, examples.

2. GLIM SYNTAX

A GLIM program consists of a sequence of statements in the form of directives. A directive consists of a directive-name followed by a set of items.

Although many other directives are available the following eleven directives could form the basis for fitting almost any general linear model.

(a) $C comment terminated by the next occurrence of the dollar symbol.

eg. $C GLIM EXAMPLE

(b) $UNITS number of cases

eg. $UNITS 8

(c) $FACTOR factor name 1 factor levels 1 factor name 2 etc.

eg. $FACTOR T 2 L 2 S 2

Factor names can have up to four characters

(d) $DATA variable 1 variable 2 etc.

declares variables to be read by next $READ command

eg. $DATA A B C

Variable names can have up to four characters

(e) $CALC calculate directive used to create new variables

eg. $CALC T = %GL(2,4) : L = %GL(2,2) : S = %GL(2,1) generates values

T	L	S
1	1	1
1	1	2
1	2	1
1	2	2
2	1	1
2	1	2
2	2	1
2	2	2

(f) $READ read data by vectors as specified in last $DATA command

(g) $YVAR variable

y - data for next $FIT directive

eg. $YVAR Y

(h) $ERROR distribution for error

B = binomial

G = gamma

N = normal

P = poisson

eg. $ERROR P

(i) $FIT model

used to specify terms to be included in a model

notation allowed includes + - . * /

eg. $FIT T+L+T.L implies that the factors T and L and their interaction T.L are to be included in this fit. An equivalent model formula is

$FIT T*L

A subsequent directive could be

$FIT -T.L which implies that the factors to be fitted are T and L and that their interaction is to be excluded.

Nested models may also be developed.

eg. $FIT T/L implies that the factor T and the interaction T.L are to be included in this fit.

(j) $DISPLAY C D E R

after a $FIT, display the output as per the mnemonics

C = correlation matrix

D = deviance of residual and degrees of freedom

E = estimates and standard errors

R = observed and fitted values and standardized residuals

eg. $DISPLAY D E R

(If no $DISPLAY command follows a $FIT this is equivalent to $DISPLAY D)

(k) $STOP last command in a deck.

3. CSIRO CYBER 76 VERSION 3.01 OF GLIM

Access to GLIM is made via the following control cards.

```
JOB.
GETSET(CMS8736)
ATTACH(GLIM3,GLIM3,ID=CMSXXX,SN=CMS8736)
GLIM3.
```

If data is on file, say TAPE1 then this is attached before the GLIM3. command and the GLIM command $DINPUT 1 is used instead of the $READ command.

4. EXAMPLES OF THE USE OF GLIM

Appendix D of the GLIM manual contains six examples of the use of GLIM. To complement these examples two more are given here.

The first example (Table 1) represents hypothetical data for 510 cows on the relationship between three variables, (1) breed of cow, (2) age of cow and (3) calf loss. There are two breeds; Brahman and Sahiwal, three age groups; 2 year old, 3 year old and 4 year old, and, two categories of calf loss; loss and no loss.

Table 1 Frequency data on breed x age x calf loss

Breed	Age	Calf Loss Yes	Calf Loss No
	2	55	67
Brahman	3	16	44
	4	8	45
	2	48	66
Sahiwal	3	20	52
	4	18	71

This first example could be coded for GLIM as follows:

```
$C     EXAMPLE     LOG-LINEAR ANALYSIS OF A THREE DIMENSIONAL TABLE.
$C
       A = BREED           B = AGE           C = CALF LOSS
       A1= BRAHMAN         B1= 2 YEARS       C1= CALF LOSS
       A2= SAHIWAL         B2= 3 YEARS       C2= NO LOSS
                           B3= 4 YEARS
$C
       DECLARE NUMBER OF UNITS AND FACTOR LEVELS
```

```
$C
$UNITS 12
$FACTOR A 2 B 3 C 2
$C
      GENERATE FACTOR LEVELS USING %GL FUNCTION
$C
$CALC A=%GL(2,6) : B=%GL(3,2) : C=%GL(2,1)
$C
      DECLARE AND READ DATA (FREE FORMAT)
$C
$DATA Y
$READ
55 67 16 44 8 45 48 66 20 52 18 71
$C
      DECLARE DEPENDENT VARIABLE AND ERROR STRUCTURE
$C
$YVAR Y $ERROR P
$C
      FIT VARIOUS MODELS DISPLAYING THE SCALED DEVIANCE (G SQUARED)
$C
$FIT A + B + C $
$FIT A + B + C + B.C $
$FIT A + B + C + A.B + B.C $
$FIT A + B + C + A.B + A.C + B.C $
$C
      A POTENTIAL MODEL IS A + B + C + B.C SO GET ESTIMATES AND RESIDUALS
$C
$FIT A + B + C + B.C $DISPLAY E R $
$STOP
```

The second example (Table 2) represents hypothetical data for 564 steers categorized by four variables, (1) breed, (2) horn length pre-dehorning, (3) dehorning instrument and (4) regrowth at six months post-dehorning. There are two breeds; Brahman and Sahiwal, three categories of horn length; less than 2.5 cm, 2.5-3.5 cm and greater than 3.5 cm, two dehorning instruments; scoop dehorner and large hodge pattern calf dehorner, and two categories of regrowth; nil and some regrowth.

Table 2 Frequency data on breed x horn length pre-dehorning x dehorning instrument x regrowth at six months post-dehorning.

Breed	Horn length pre-dehorning	Dehorning Instrument: Scoop		Hodge	
		Regrowth			
		Nil	Some	Nil	Some
Brahman	≤ 2.5 cm	54	44	25	18
	2.5-3.5 cm	77	63	64	21
	≥ 3.5 cm	22	25	13	5
Sahiwal	≤ 2.5 cm	13	5	18	1
	2.5-3.5 cm	21	4	33	5
	≥ 3.5 cm	14	3	11	5

The GLIM code for this second example illustrates the model selection procedure "the method of standardized parameter estimates".

```
$C      EXAMPLE   LOG-LINEAR ANALYSIS OF A FOUR DIMENSIONAL TABLE
                  METHOD OF STANDARDIZED PARAMETER ESTIMATES
$C      A = BREED                          B = HORN LENGTH PRE-DEHORNING
        A1= BRAHMAN                        B1= < 2.5 CM
        A2= SAHIWAL                        B2= 2.5 - 3.5 CM
                                           B3= > 3.5 CM
$C      C = DEHORNING INSTRUMENT           D = REGROWTH
        C1= SCOOP DEHORNER                 D1= NIL REGROWTH
        C2= HODGE DEHORNER                 D2= SOME REGROWTH
$C
        DECLARE NUMBER OF UNITS AND FACTOR LEVELS
        FOR THIS METHOD FACTOR LEVELS MUST SUM TO ZERO
        LINEAR CONTRASTS FOR FACTORS A, C AND D
        LINEAR AND QUADRATIC CONTRASTS FOR FACTOR B
$C
$UNITS 24
$DATA A BL BQ C D Y
$READ
```

```
 1  -1   1   1   1  54
 1  -1   1   1  -1  44
 1  -1   1  -1   1  25
 1  -1   1  -1  -1  18
 1   0  -2   1   1  77
 1   0  -2   1  -1  63
 1   0  -2  -1   1  64
 1   0  -2  -1  -1  21
 1   1   1   1   1  22
 1   1   1   1  -1  25
 1   1   1  -1   1  13
 1   1   1  -1  -1   5
-1  -1   1   1   1  13
-1  -1   1   1  -1   5
-1  -1   1  -1   1  18
-1  -1   1  -1  -1   1
-1   0  -2   1   1  21
-1   0  -2   1  -1   4
-1   0  -2  -1   1  33
-1   0  -2  -1  -1   5
-1   1   1   1   1  14
-1   1   1   1  -1   3
-1   1   1  -1   1  11
-1   1   1  -1  -1   5
$C
       SET UP DESIGN MATRIX FOR FULL MODEL
$C
$CALC ABL=A*BL : ABQ=A*BQ : AC=A*C : AD=A*D
$CALC BLC=BL*C : BQC=BQ*C : BLD=BL*D : BQD=BQ*D
$CALC CD=C*D
$CALC ABLC=A*BL*C : ABQC=A*BQ*C
$CALC ABLD=A*BL*D : ABQD=A*BQ*D
$CALC ACD=A*C*D
$CALC BLCD=BL*C*D : BQCD=BQ*C*D
$CALC INT1=A*BL*C*D : INT2=A*BQ*C*D
$C
       CALCULATE AVERAGE STANDARD DEVIATION TO USE AS SCALE PARAMETER
$C
$CALC X=1/Y : %A=0 : %A=%A+X : %A=%CU(X) : %A=%SQRT(%A)/%NU
```

```
$C
       TAKE LOGARITHMS OF Y
$C
$CALC Y=%LOG(Y)
$C
       DECLARE DEPENDENT VARIABLE , ERROR STRUCTURE AND SCALE PARAMETER
$C
$YVAR Y
$ERROR N
$SCALE %A
$C
       FIT FULL MODEL AND DISPLAY ESTIMATES
$C
$FIT A+BL+BQ+C+D+ABL+ABQ+AC+AD+BLC+BQC+BLD+BQD+CD+ABLC+ABQC+ABLD+ABQD+ACD+
BLCD+BQCD+INT1+INT2
$DISPLAY E $
$STOP
```

Output for these examples is attached as Appendix A.

5. GLIM - ADVANTAGES AND DISADVANTAGES

GLIM has few advantages or disadvantages over its rival GENSTAT when it comes to fitting general linear models. It is very simple to use (something not often said of GENSTAT) with an excellent reference manual. GENSTAT has one facility GLIM does not have; the ability to construct an analysis of deviance table from a series of FIT's. This is not seen as a serious disadvantage.

GLIM is essentially a program for fitting general linear models. As such it does not have the features of a general statistical program like GENSTAT for doing analysis of variance or principal components analysis, etc. It can however be used, as a calculator for doing arithmetic on vectors and scalars, for producing scattergrams and summary statistics and for data exploration.