PACIFIC STATISTICAL CONGRESS
I.S. Francis, B.F.J. Manly and F.C. Lam (Editors)

MINITAB AS A TEACHING AID FOR GENERALIZED LINEAR MODELS

David R Fox

Department of Mathematics and Statistics

Western Australian Institute of Technology

ABSTRACT: This paper describes how the statistical computing package, MINITAB may be utilized as an aid in the development of the statistical theory underlying the use of Generalized Linear Models.

KEY WORD: Generalized Linear Models; MINITAB; GLIM; Statistical Modelling; teaching aids.

1. INTRODUCTION

The theory of Generalized Linear Models has been well documented by Nelder & Wedderburn in their 1972 paper and since then in numerous texts. Although fairly straightforward, difficulties can arise when one attempts to introduce this theory to statistics undergraduates and/or the casual user of the GLIM software. The terminology adopted by Nelder & Wedderburn and used extensively throughout the GLIM documentation can at times obscure the thrust of the statistical explanation.

It has been this author's experience that the students and GLIM users referred to above have difficulty in coming to grips with the GLIM 'jargon'. The unfortunate consequence of this is that the theory underlying the iterative process used to obtain parameter estimates is less readily absorbed and understood.

The GLIM software is a powerful and flexible tool which embodies the theory of Generalized Linear Models. As with any statistical computing package its mere application to a given data set gives us little or no insight into the theory on which it is based.

On the other hand, an attempt to fit a GLM by a pencil and paper approach would be equally uninformative since the model-fitting process is iterative and relies heavily on matrix algebra. The tedium of inverting, transposing and multiplying matrices would clearly outweigh any learning opportunities afforded by this approach.

This paper describes an approach which avoids the tedium alluded to above, yet retains the advantages associated with the calculation by hand method.

The matrix handling facilities and stored macro capabilities of the MINITAB package have been exploited to assist the understanding of the mechanics and theory of the GLIM software

We begin with an overview of Generalized Linear Models and a brief description of the terminology and considerations.
An example of the approach is provided in §5. Additional MINITAB 'programs' for a variety of GLM's are provided in Fox (Technical Report No 3 - 1985).

2. GENERALIZED LINEAR MODELS

The theoretical development of GLM's will not be considered here. Instead we shall utilize some standard results and provide an outline of the iterative model-fitting procedure. The interested reader requiring more detail is advised to consult Fox (1984), Dobson (1983), or McCullagh & Nelder (1983).

2.1 THE STRUCTURE OF GLM'S

Many 'classical' linear statistical models may be written as :

$$Y = X\beta + \varepsilon \tag{2.1}$$

Where Y is an $N \times 1$ column vector of responses, X is an $N \times k$ design matrix, β is a $k \times 1$ column vector of parameters and ε is an $N \times 1$ column vector of error terms (assumed normally distributed with zero mean and constant variance σ^2_ε).

A more comprehensive class of structures may be accommodated by allowing models of the form :

$$Y = f(X\beta) + \varepsilon \tag{2.2}$$

for some function $f(\cdot)$ and error term ε . The only restriction placed on ε (and hence Y) is that its p.d.f. belong to the exponential family of distributions. Thus the p.d.f. of Y may be expressed as :

$$f_Y(y;\theta) = exp[a(y)b(\theta) + c(\theta) + d(y)] \tag{2.3}$$

for some functions $a(\cdot)$, $b(\cdot)$, $c(\cdot)$, and $d(\cdot)$. If $a(y) = y$ (2.3) is said to have the *cannonical form* and $b(\Theta)$ is called the *natural parameter* of the distribution. From (2.2) we have that

$$E[Y] = \mu = f(X\beta)$$

$$\Rightarrow g(\mu) = X\beta$$

where $g(\cdot)$ is the inverse of $f(\cdot)$. The function g is called the *link function* and the systematic component of our model, $X\beta$ is often replaced by the symbol η and is known as the *linear predictor.*

3. PARAMETER ESTIMATION FOR GLM'S

The method of parameter estimation for a GLM is equivalent to an iterative weighted least-squares procedure. The procedure is described fully in Nelder & Wedderburn (1972), although a more readable account is provided by Dobson (1983), and it is the latter we shall adopt here.
The parameter estimation procedure relies on obtaining successive approximations to the system of equations generated by setting

$\frac{\partial l}{\partial \beta} = 0$ where l is the log-likelihood function

for Y and β the vector of model parameters.
Using the Newton-Raphson method we thus have :

$$b^{(m)} = b^{(m-1)} - \left[\frac{\partial^2 l}{\partial \beta_j \partial \beta_k}\right]^{-1} U^{(m-1)} \qquad (3.1)$$

where $b^{(m)}$ and $b^{(m-1)}$ are the current parameter estimates at the m^{th} and $(m-1)^{th}$ iterations respectively, and $U_j = \frac{\partial l}{\partial \beta_j}$ is the score with respect to β_j. Equation (3.1) essentially describes the iterative provedure employed by the GLIM software with the exception that the matrix of second derivatives is replaced by the matrix of expected values

$$E\left[\frac{\partial^2 l}{\partial \beta_j \partial \beta_k}\right].$$

This modified procedure is often referred to as Fisher's *method of scoring.*
In matrix notation, the procedure may be defined as :

$$X'WXB^{(m)} = X'WZ \qquad (3.2)$$

where X is an N x k design matrix
W is an N x N diagonal matrix having elements :

$$W_{ii} = \frac{1}{Var[Y_i]}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2$$

where μ is a vector of means, η the linear predictor and Z is an N x 1 column vector with elements :

$$Z_i = \sum_k x_{ik} b_k^{(m-1)} + (y_i - \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right).$$

4. MEASURE OF GOODNESS OF FIT

Nelder and Wedderburn (1972) proposed the *deviance statistic* D as a measure of discrepancy between the current model relative to the full model (which has as many parameters as there are observations).
The discrepancy of a fit is proportional to twice the difference between the maximum log-liklihood achievable and that achieved by the model under investigation. The forms taken by the deviance statistic for some common members of the exponential family of distributors may be found in the GLIM manual or McCullagh & Nelder.

5. AN EXAMPLE OF THE GLIM PROCEDURE USING MINITAB

To illustrate the calculations performed by the GLIM software we consider the following bioassay data reported by Bliss (1935). Data are given for the number of beetles killed after 5 hours exposure to gaseous carbon disulphide at various concentrations.

Dose, x_i (log CS_2 mg l^{-1})	Number of insects, n_i	Number Killed r_i
1.6907	59	6
1.7242	60	13
1.7552	62	18
1.7842	56	28
1.8113	63	52
1.8369	59	53
1.8610	62	61
1.8839	60	60

Let R_i denote the number of insects killed (out of n_i) at dose level x_i.
Assume R_i has the binomial distribution with parameters (n_i, Θ_i) where Θ_i is the true proportion killed at dose level x_i.

$$\text{i.e.} \quad P[R_i = r_i] = \binom{n_i}{r_i} \Theta_i^{r_i} (1-\Theta_i)^{n_i - r_i}$$

Define a new random variable Y_i as the proportion of insects killed at dose level x_i.

$$\text{i.e.} \quad Y_i = \frac{R_i}{n_i}$$

$$\Rightarrow P[Y_i = y_i] = \binom{n_i}{n_i y_i} \Theta_i^{n_i y_i} (1-\Theta_i)^{n_i - n_i y_i}$$

We take our Model as the logistic function

$$Y_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} + \varepsilon_i$$

with

$$E[Y_i] = \mu_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} = \frac{e^{\eta_i}}{1 + e^{\eta_i}}$$

and

$$\ln\left(\frac{\mu_i}{1 - \mu_i}\right) = \eta_i$$

giving :

$$\frac{\partial \mu_i}{\partial \eta_i} = \frac{e^{\eta_i}}{(1+e^{\eta_i})^2} \quad \text{and} \quad \frac{\partial \eta_i}{\partial \mu_i} = \frac{1}{\mu_i(1-\mu_i)}$$

Also,

$$E[Y_i] = \theta_i$$

$$\text{and} \quad \mathrm{Var}[Y_i] = \frac{\theta_i(1-\theta_i)}{n_i}$$

We will now utilize these results and solve (3.2) iteratively using MINITAB.
Note that before executing the stored set of instructions the user must have previously defined the following three matrices;

- M1 - design matrix X
- M2 - vector of observations Y
- M3 - initial parameter estimates $\hat{\beta}^{(0)}$

This is achieved in MINITAB as follows :

```
MTB > set c11
DATA> 8(1)
DATA> set c12
DATA> 1.6907,1.7242,1.7552,1.7842
DATA> 1.8113,1.8369,1.8610,1.8839
DATA> copy c11 c12 m1
MTB > set c13
DATA> 59,60,62,56
DATA> 63,59,62,60.5
DATA> set c14
DATA> 6,13,18,28
DATA> 52,53,61,60
DATA> divide c14 c13 c15
MTB > copy c15 m2
MTB > set c16
DATA> 0,0
DATA> copy c16 m3
MTB >
```

```
MTB > print m1-m3
MATRIX M1                  MATRIX M2      MATRIX M3

  1.0000   1.6907          0.101695         0
  1.0000   1.7242          0.216667         0
  1.0000   1.7552          0.290323
  1.0000   1.7842          0.500000
  1.0000   1.8113          0.825397
  1.0000   1.8369          0.983871
  1.0000   1.8610          0.991736
  1.0000   1.8839
```

MINITAB Parameter Estimation Routine

```
MTB > exec 'blparm' 10            Peform 10 iterations
MTB > multiply m1 m3 m4           Matrix M4 = η
MTB > copy m4 c1                  Store ηi in C1
MTB > let c2=exp(c1)              C2 = e^ηi
MTB > let c3=c2/(1+c2)            C3 = μi
MTB > let c4=c3/(1+c2)            C4 = ∂μi/∂ηi
MTB > let c5=c3*(1-c3)/c13        C5 = Var[Yi]
MTB > let c6=1/(c3*(1-c3))        C6 = ∂ηi/∂μi
MTB > let c7=(1/c5)*(c4**2)       C7 = ωi
MTB > diag c7 m8                  Matrix M8 = W
MTB > copy m2 c8                  Store yi in C8
MTB > subtract c3 c8 c9           C9 = yi - μi
MTB > multiply c9 c6 c10          C10 = (yi - μi) ∂ηi/∂μi
MTB > add c10 c1 c11              C11 = C10 + ηi
MTB > copy c1 m7                  Matrix M7 = Z
MTB > trans m1 m9                 Matrix M9 = X'
MTB > multiply m9 m8 m10          Matrix M10 = X'W
MTB > multiply m10 m1 m11         Matrix M11 = X'WX
MTB > invert m11 m11              Matrix M11 = (X'WX)^-1
MTB > note . . . covariance matrix . . .
MTB > print m11
MATRIX M11

  6.72953  -3.74660
 -3.74660   2.08846


MTB > multiply m10 m7 m12         Matrix M12 = X'WZ
MTB > multiply m11 m12 m13        Matrix M13 =
MTB > print m13                   (X'WX)^-1 X'WZ = β^
MATRIX M13

 -37.679
  21.236


MTB > copy m13 m3                 Store revised
MTB > end                         ESTIMATES IN M3.
MTB > multiply m1 m3 m4    }
MTB > copy m4 c1           }      more iterations
      :                    }
```

MINITAB: Calculation of Deviance

```
MTB > exec 'bldvnce'
MTB > let cl7=1-c3            c17= 1-μi
MTB > let cl8=1-cl5           c18= 1-yi
MTB > let cl9=loge(c3/cl7*cl8/cl5)
MTB > let cl9=cl3*cl5*cl9
MTB > let c2Ø=cl3*loge(cl7/cl8)
MTB > add cl9 c2Ø c21
MTB > multiply c21 -2 c21
MTB > sum c21 kl
   SUM     =     9.2798
MTB > note . . . Deviance = . . .
MTB > print kl
Kl      9.27976
MTB > end
MTB > stop
```

REFERENCES:

[1] Armstrong, W. A Users' Review of GLIM 3, The Statistician 29 (1980) 57-63.

[2] Baker, R.J. & Nelder, J.A. The GLIM System - Release 3 (NAG, 1978).

[3] Barndorff-Neilsen, O. Information and Exponential Families in Statistical Theory (Wiley, New York, 1978).

[4] Bliss, C.I. The Calculation of the dosage-mortality curve, Annals of Applied Biology 22 (1935) 134-167.

[5] Dobson, A.J. An Introduction to Statistical Modelling (Chapman & Hall, London, 1983).

[6] Fox, D.R. A Survey of Generalized Linear Models and the GLIM Software, Technical Report No. 9 (School of Mathematics and Computing, W.A.I.T. 1984).

[7] Fox, D.R. MINITAB as a Teaching Aid for Generalized Linear Models, Technical Report No. 3 (School of Mathematics and Computing, W.A.I.T. 1985).

[8] Gilchrist, R. (ed.) GLIM 82: Proceedings of the International Conference on Generalized Linear Models (Springer-Verlag, New York, 1982).

[9] McCullagh, P. & Nelder, J.A. Generalized Linear Models (Chapman & Hall, London, 1983).

[10] Nelder, J.A. & Wedderburn, R.W.M. Generalized Linear Models, J.R. Statist. Soc. (A) 135 (1972) 370-384.