The Computer Program LOGREG

The computer program LOGREG can be used to fit a variety of logistic regression models for proportion data, allowing for the effects of both quantitative variables (such as age in years) and factors (such as species of plant). If factors are considered then the user of LOGREG must set up appropriate 0-1 dummy variables. Fitted models always include a constant term.

According to the logistic assumption, the probability of a 'success' for cases in the ith of several groups is

$$p_i = \frac{\exp(ß_0 + ß_1 x_{i1} + ... + ß_p x_{ip})}{1 + \exp(ß_0 + ß_1 x_{i1} + ... + ß_p x_{ip})}.$$

This probability can then be estimated by knowing the number of 'successes' that occur with n trials for the ith group.

Input

Two types of input to the program are allowed:

(a) Input can be directly from the keyboard. In that case the user is first asked for the number of groups (proportions) and the number of predictor variables (X variables). Next the numbers of "successes" and the number of trials (Y and n) are requested for the cases in turn. Finally, the values of predictor variables are requested, in the order X_1, X_2, ..., X_p. The analysis is then conducted, with output as shown in the example below.

(b) Input can be from a text file. In this case the first line of the file must contain a title and the second line must contain two numbers, separated by one or more spaces. These are the number of groups and the number of predictor variables. The rest of the file contains the number of 'successes' (Y), the number of trials (n), and the values of predictor variables (X_1, X_2, ..., X_p), one group at a time. The ASCII file can be set up in a text editor, a word processor, or in a spreadsheet. The data should be set up as needed and then printed to an ASCII file.

Output

In addition to appearing on the screen, output it is sent to the file LOGREG.OUT for later printing if desired.

Example

This example is concerned with the site preference of two species of lizard, as shown in the following table of frequencies. The logistic regression model fits the proportion of species A present under various conditions. Only the main effects of the factors perch height, perch diameter, sun and time of day are allowed for.

Perch height	Perch diameter	Sun or shade	Time of day	Species A	Total lizards
low	low	sun	early	20	22
high	low	sun	early	13	13
low	high	sun	early	8	11
high	high	sun	early	6	6
low	low	shade	early	34	45
high	low	shade	early	31	36
low	high	shade	early	17	32
high	high	shade	early	12	13
low	low	sun	mid-day	8	9
high	low	sun	mid-day	8	8
low	high	sun	mid-day	4	5
low	low	shade	mid-day	69	89

high	low	shade	mid-day	55	59
low	high	shade	mid-day	60	92
high	high	shade	mid-day	21	26
low	low	sun	late	4	4
high	low	sun	late	12	12
low	high	sun	late	5	8
high	high	sun	late	1	2
low	low	shade	late	18	28
high	low	shade	late	13	16
low	high	shade	late	8	16
high	high	shade	late	4	8

Note that there were no lizards in the sun on large diameter high perches, at mid-day.

The input file for this example is as shown below, with 0-1 variables set up to account for the four factors, but no interactions.

```
HABITAT CHOICE OF LIZARDS
23 5
20 22  0  0  0  0  0
13 13  1  0  0  0  0
 8 11  0  1  0  0  0
 6  6  1  1  0  0  0
34 45  0  0  1  0  0
31 36  1  0  1  0  0
17 32  0  1  1  0  0
12 13  1  1  1  0  0
 8  9  0  0  0  1  0
 8  8  1  0  0  1  0
 4  5  0  1  0  1  0
69 89  0  0  1  1  0
55 59  1  0  1  1  0
60 92  0  1  1  1  0
21 26  1  1  1  1  0
 4  8  0  0  0  0  1
12 12  1  0  0  0  1
 5  8  0  1  0  0  1
 1  2  1  1  0  0  1
18 28  0  0  1  0  1
13 16  1  0  1  0  1
 8 16  0  1  1  0  1
 4  8  1  1  1  0  1
```

The output follows, with comments added in italics.

```
               *****  LOGISTIC REGRESSION *****
================================================================

Problem title: HABITAT CHOICE OF LIZARDS

Number of X variables =  5  Number of groups =  23

Initial approximations for parameters
   b0  1.176      b1   .000      b2   .000      b3   .000      b4   .000
   b5   .000
```

The 5 0-1 variables represent the perch height (X_1), the perch diameter (X_2), sun or shade (X_3), and the three times of day (X_4 and X_5).

```
Chi-squared values for null model with zero coefficients for
X variables:
Pearson =     61.35,  Log-likelihood =     70.10,  with   22 df.

Iteration:  1  Initial Deviance =  70.102
Fraction of corrections made = 1.00000  New Deviance =    16.822

Iteration:  2  Initial Deviance =  16.822
Fraction of corrections made = 1.00000  New Deviance =    14.228

Iteration:  3 = Initial Deviance =  14.228
Fraction of corrections made = 1.00000  New Deviance =    14.205
```

```
Iteration:  4  Initial Deviance =  14.205
Fraction of corrections made = 1.00000  New Deviance =    14.205

Iteration:  5  Initial Deviance =  14.205
Fraction of corrections made = 1.00000  New Deviance =    14.205
```

Five iterations are required.

Parameter	Final estimate	Standard error	Ratio	
b0	.1945E+01	.3415E+00	5.69	*Constant*
b1	.1130E+01	.2571E+00	4.40	*Perch height parameter*
b2	-.7626E+00	.2113E+00	-3.61	*Perch diameter parameter*
b3	-.8473E+00	.3224E+00	-2.63	*Shade (not sun) parameter*
b4	.2271E+00	.2502E+00	.91	*Mid-day parameter*
b5	-.7368E+00	.2990E+00	-2.46	*Late parameter*

Group	Y	N	E(Y)	Chi-sq	Group	Y	N	E(Y)	Chi-sq
1	20	22	19.2	.24	2	13	13	12.4	.60
3	8	11	8.4	.09	4	6	6	5.5	.59
5	34	45	33.7	.01	6	31	36	32.5	.71
7	17	32	18.7	.35	8	12	13	10.6	1.05
9	8	9	8.1	.01	10	8	8	7.7	.29
11	4	5	4.0	.00	12	69	89	70.3	.12
13	55	59	54.3	.10	14	60	92	58.6	.09
15	21	26	22.0	.27	16	4	8	6.2	3.29
17	12	12	10.9	1.16	18	5	8	4.9	.01
19	1	2	1.7	1.52	20	18	28	16.5	.33
21	13	16	13.1	.00	22	8	16	6.4	.66
23	4	8	5.4	1.11					

```
Goodness of fit: Pearson =    12.59, Log-likelihood =    14.20 , df =  17
```

The model is a good fit.

```
COVARIANCE MATRIX FOR PARAMETERS
   b0   .117E+00
   b1  -.157E-01  .661E-01
   b2  -.216E-01  .447E-02  .446E-01
   b3  -.825E-01 -.201E-02  .208E-03  .104E+00
   b4  -.227E-01  .356E-02 -.432E-02 -.169E-01  .626E-01
   b5  -.405E-01 -.303E-02  .220E-02  .361E-02  .363E-01  .894E-01

CORRELATION MATRIX FOR PARAMETERS
   b0   1.000
   b1   -.179  1.000
   b2   -.299   .082  1.000
   b3   -.750  -.024   .003  1.000
   b4   -.266   .055  -.082  -.209  1.000
   b5   -.397  -.039   .035   .037   .486  1.000
```

The model seems to fit, with no large residuals. All the parameters seem important except that the proportion of species A at mid-day does not seem much different from the proportion in the early part of the day.