The Computer Program LOGLIN

The computer program LOGLIN can be used to fit a log-linear model of the form

$$Y = B\exp(ß_0 + ß_1x_1 + ... + ß_px_p) + e,$$

where Y is an observed count with a Poisson distribution, B is an optional base rate, and e is a random error. The program allows for up to 500 data frequencies and up to 30 parameters.

Data Input

The data input for LOGLIN can either be directly through the keyboard (as directed by the program) or from a file set up previously. If keyboard input is chosen then the data are saved in a file for later use. Data files can either be produced in this way or by using a text editor or word processor that outputs ASCII text files. If the user chooses to set up a file themselves then the data format must be as shown below:

Line 1 contains a title for data (up to 50 characters).

Line 2 contains NFREQ (number of data frequencies), NX (number of X variables), and IB (1 if there are base rates, otherwise 0).

Line 3 contains the first data frequency, the base rate for this frequency (if IB=1), and X(1) to X(NX) for this frequency.

.

.

Line NFREQ+2 contains the last data frequency, the base rate (if IB=1), and X(1) to X(NX) for this frequency.

Once the data are input, the user is asked to identify the X variables to be used in the analysis.

Output

The output from LOGLIN is shown in the example that follows. An iterative process is used to improve initial estimates, and as this proceed a summary is output to the screen, including the gradually increasing log-likelihood function. When it is decided that estimates cannot be improved the following are output: (i) final estimates with standard errors; (ii) observed and expected frequencies; (iv) Pearson and log-likelihood goodness of fit statistics, with the degrees of freedom, (v) the covariance matrix for estimated parameters, and (vi) the correlation matrix for estimated parameters. After fitting the requested model it is possible to choose either to change the X variables being used and fit another model, or to terminate the program.

The output from the program is also sent to the file LOGLIN.OUT for later printing if desired. If the program is started twice then the second output overwrites what is already in LOGLIN.OUT.

Example

The following Table 1 shows the level of church attendance classified by age and religion for a sample of 1591 individuals obtained from a survey. In this example a loglinear model for the counts is fitted on the assumption that the three classifications are independent. Table 2 shows the data in a form suitable for input to LOGLIN, with 0-1 variables set up to allow the counts to depend on the three classification factors but with these having independent effects. Table 3 shows the output from the program, with comments.

Table 2 Church attendance related to religion and age from a sample survey carried out in 1972.

Religion	Age	Church attendance Low	Medium	High	Total
Non-Catholic	Young	322	122	141	585
Non-Catholic	Old	250	152	194	596
Catholic	Young	88	45	106	239
Catholic	Old	28	24	119	171
		688	343	560	1591

Table 3 Data in a form suitable for input to LOGLIN, assuming independent effects of religion, age and church attendance.

```
CHURCH ATTENDANCE DATA
12 4 0
322  0  0  1  0
250  0  1  1  0
 88  1  0  1  0
 28  1  1  1  0
124  0  0  0  1
152  0  1  0  1
 45  1  0  0  1
 24  1  1  0  1
141  0  0  0  0
194  0  1  0  0
106  1  0  0  0
119  1  1  0  0
```

Table 4 Output from LOGLIN.

```
               *****  LOG-LINEAR FITTING *****
=============================================================

Problem title: CHURCH ATTENDANCE DATA

Number of X variables =  4  Number of frequencies =  12

Initial approximations for parameters
   b0  4.888      b1   .000      b2   .000      b3   .000      b4   .000

Chi-squared values for null model with zero coefficients for
X variables:
Pearson =    657.15,  Log-likelihood =    662.89,  with   11 df.
```

This shows that the null model, with all counts expected to be the same, gives a very poor fit to the data.

```
Iteration:  1  Initial Deviance = 662.894
Fraction of corrections made = 1.00000  New Deviance =   191.543

Iteration:  2  Initial Deviance = 191.543
Fraction of corrections made = 1.00000  New Deviance =   151.155

Iteration:  3  Initial Deviance = 151.155
Fraction of corrections made = 1.00000  New Deviance =   150.863

Iteration:  4  Initial Deviance = 150.863
Fraction of corrections made = 1.00000  New Deviance =   150.863

Iteration:  5  Initial Deviance = 150.863
```

```
Fraction of corrections made =  .50000  New Deviance =   150.863
```

Five iterations were needed to fit the model with 4 X variables.

Parameter	Final estimate	Standard error	Ratio
b0	.5374E+01	.5085E-01	105.67
b1	-.1060E+01	.5731E-01	-18.49
b2	-.7411E-01	.5014E-01	-1.48
b3	.2059E+00	.5691E-01	3.62
b4	-.4844E+00	.6844E-01	-7.08

Class	Obs freq	Exp freq	Chi-sq	Class	Obs freq	Exp freq	Chi-sq
1	322	264.9	12.30	2	250	246.00	.07
3	88	91.8	.16	4	28	85.26	38.45
5	124	132.8	.59	6	152	123.36	6.65
7	45	46.0	.02	8	24	42.75	8.23
9	141	215.6	25.83	10	194	200.23	.19
11	106	74.7	13.08	12	119	69.40	35.46

```
Goodness of fit: Pearson =    141.03, Log-likelihood =    150.86 , df =   7
```

This model does not fit much better than the null model.

COVARIANCE MATRIX FOR PARAMETERS

b0	.259E-02				
b1	-.845E-03	.328E-02			
b2	-.121E-02	-.365E-10	.251E-02		
b3	-.179E-02	.611E-10	.561E-10	.324E-02	
b4	-.179E-02	.135E-10	-.225E-10	.179E-02	.468E-02

CORRELATION MATRIX FOR PARAMETERS

b0	1.000				
b1	-.290	1.000			
b2	-.475	.000	1.000		
b3	-.617	.000	.000	1.000	
b4	-.513	.000	.000	.458	1.000

High correlations (close to +1 or -1) indicate difficulty in separating the effects of different X variables. This problem is not apparent here.