



## Fieger, Kastner, Heumann: WinMAREG Quick Start

Sonderforschungsbereich 386, Paper 108 (1998)

Online unter: <http://epub.ub.uni-muenchen.de/>

## Projektpartner



# WinMAREG Quick Start

A. Fieger, C. Kastner, C. Heumann

Institut für Statistik  
Akademiestr. 1  
80799 München

April 1998

## Abstract

This paper is a short introduction into the usage of WinMAREG. Two examples are used to illustrate the most common options and features of the software.

**Keywords:** User guide, MAREG, WinMAREG.

## 1 Introduction

This quick start will give the user a short introduction into the handling of WinMAREG. We present the analysis of two data sets, which are provided with the software. These are the ‘Respiratory Disorder Data’ described in Miller, Davis and Landis (1993) and the ‘Ohio Children Data’ described e. g. in Fitzmaurice and Laird (1993).

Most of the features and options are presented. A detailed description can be found in the (context sensitive) online help or in Fieger, Heumann and Kastner (1996).

Starting the program displays the main window of WinMAREG presented in figure 1. The menubar consists of the usual file menu (opening and closing data files), the options menu, the two statistics menues (ML and GEE1) and a help menu. The options menu provides entrys that determine the appearance of the main window (toolbar, statusbar, etc.) and a general options section, where the user can determine options which are to be used by MAREG, the computing routines (see figure 2).

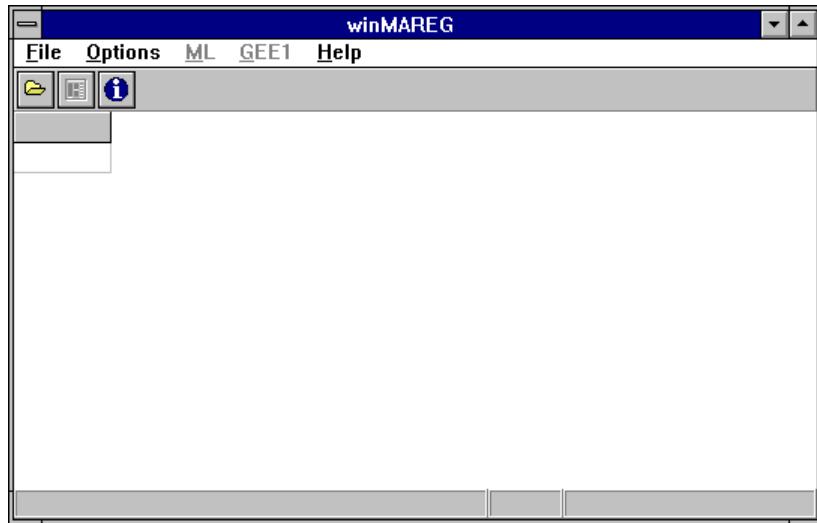


Figure 1: Main window of WinMAREG

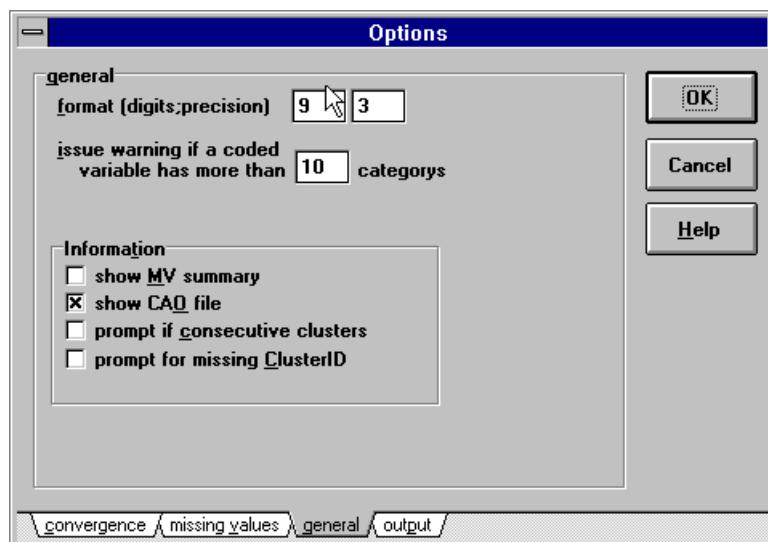


Figure 2: General options dialog of WinMAREG

Figure 3 shows the entries of the help menu, where information about the selected variables of the current model can also be found.

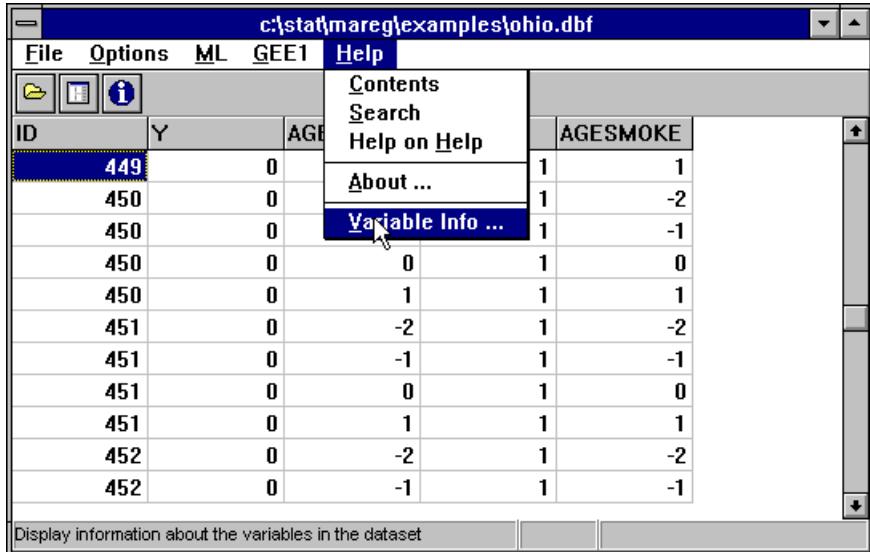


Figure 3: WinMAREG's help menu

## 2 Generalized Estimating Equations

This example illustrates the GEE features of MAREG. The data used are the 'Respiratory Disorder Data', which are a randomized clinical trial of a new treatment of respiratory disorder.

The response has three categories. Therefore, as shown in figure 4, a cumulative logit model is chosen. This opens the model selection dialog (figure 5). Selecting 'RESP' as dependent variable and 'TREAT' as independent variable by clicking the respective buttons specifies the model for the mean structure.

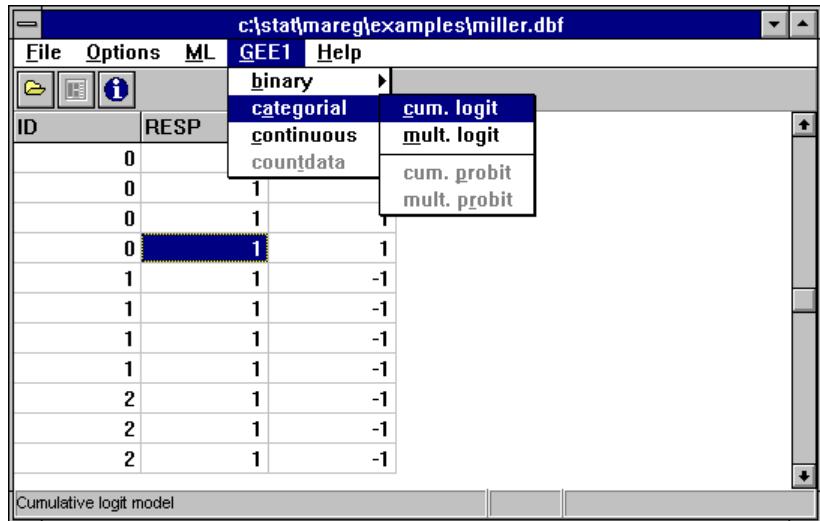


Figure 4: Selecting the GEE1 cumulative logit model

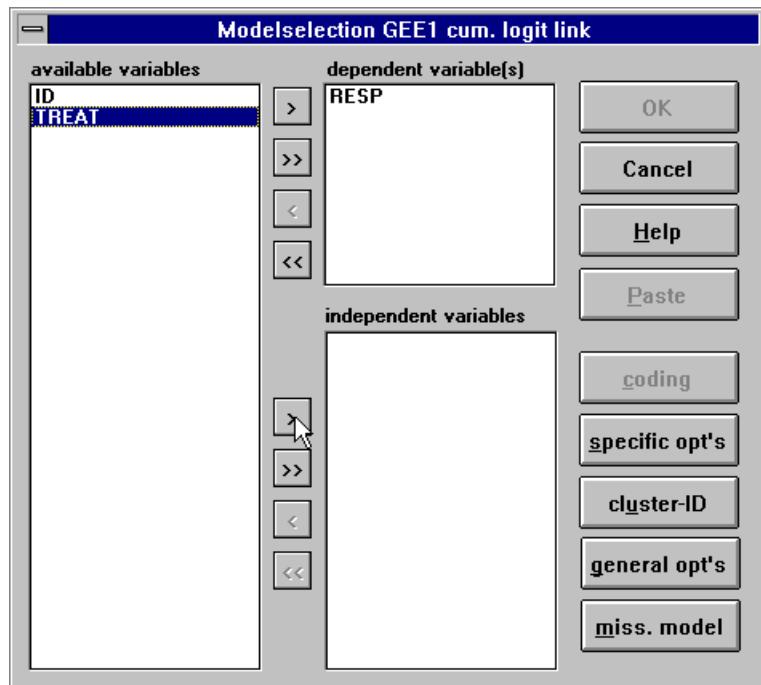


Figure 5: Model selection dialog (GEE1, cumulative logit link)

Dummy coding of the response variable (which has three categorys) can be obtained by right clicking the variable name or by clicking the coding button (see figure 6). As a response variable is coded, effect coding can not be selected. The treatment variable does not have to be coded as the existing codes are  $-1$  and  $1$ , which is already effect coding.

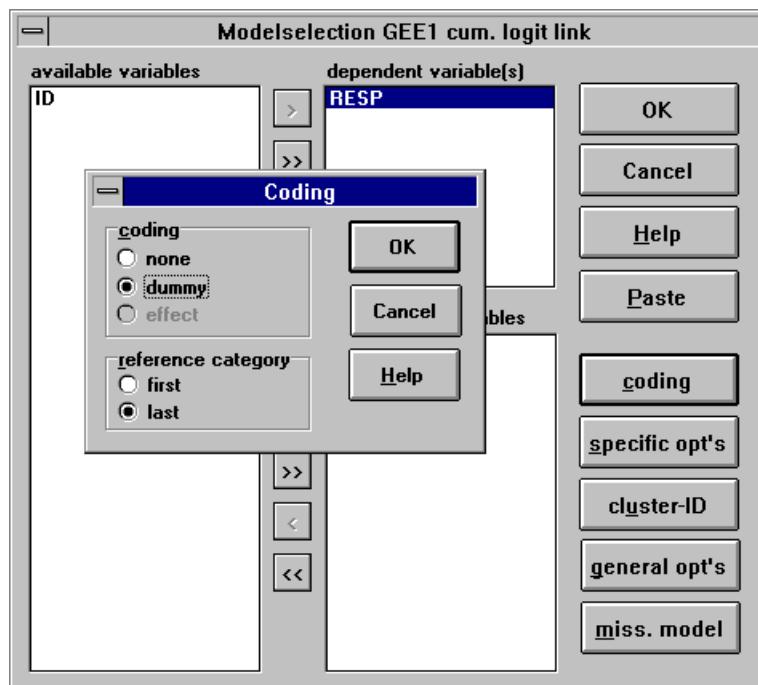


Figure 6: Coding of the response variable

The longitudinal structure of the data provides four observations per patient. Clicking the cluster-ID button opens the dialog shown in figure 7, where the variable 'ID' can be selected to be the cluster-ID. If no cluster-ID is selected each row in the data table would be treated as a single cluster.

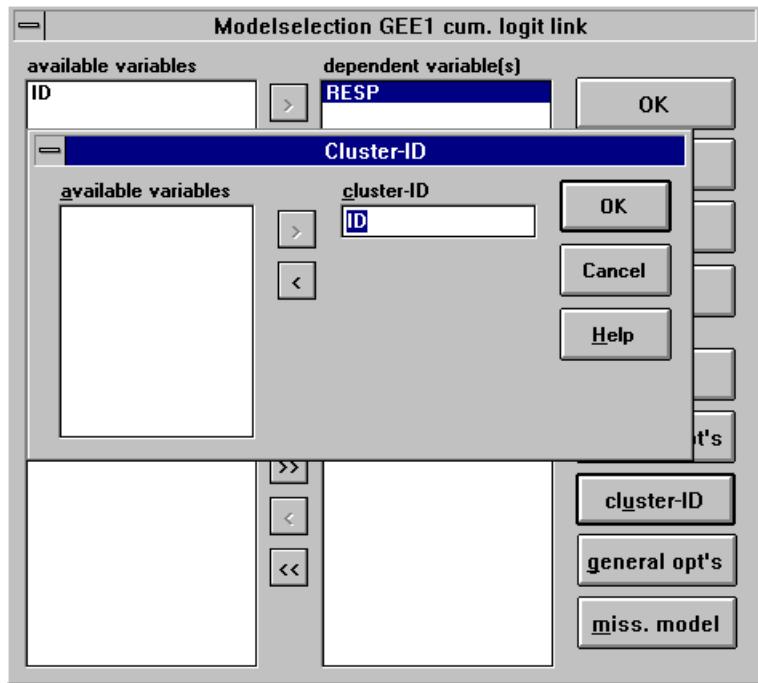


Figure 7: Selecting a cluster-ID

By clicking on the specific options button the dialog in figure 8 shows up. Here one of the methods that are appropriate for the current data level can be selected. Selecting Method of Prentice and clicking OK opens the next dialog (see figure 9), where two different ways of specifying the covariance matrix of the association model can be chosen.

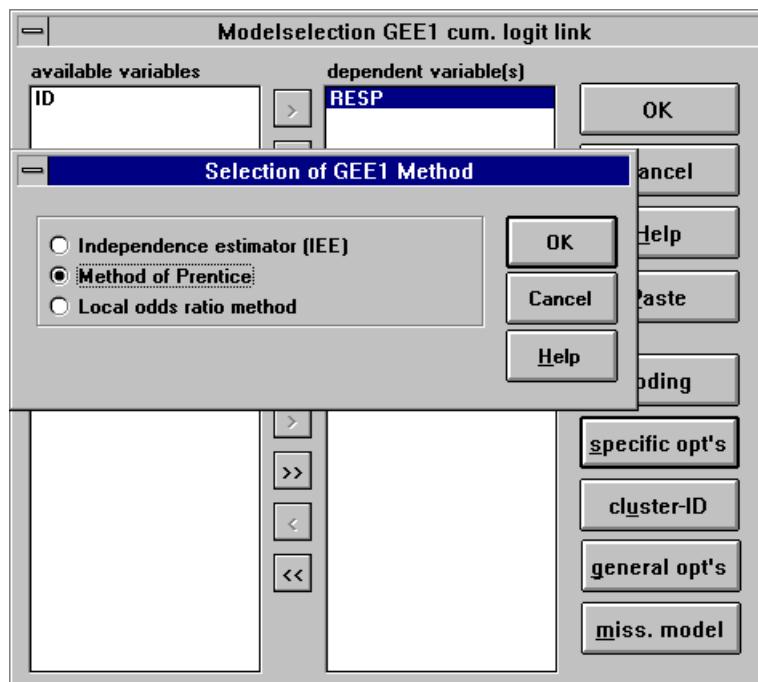


Figure 8: Selecting a GEE1 method

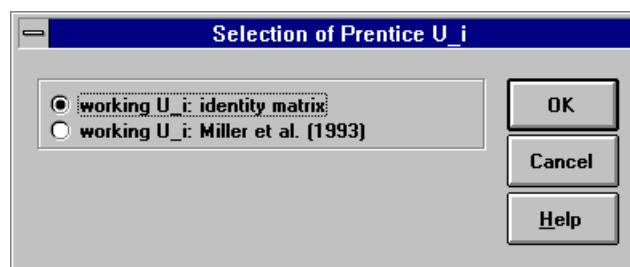


Figure 9: Options for Method of Prentice

Again, clicking OK opens the dialog where the design (matrix) of the mean and association structure can be specified (see figure 10). Clicking OK brings us back to the model selection dialog (figure 5). Clicking OK or Paste opens the file dialog shown in figure 11. OK immediately starts MAREG as shown in figure 12, whereas clicking Paste only generates the data and information files needed by MAREG, which can be run later.

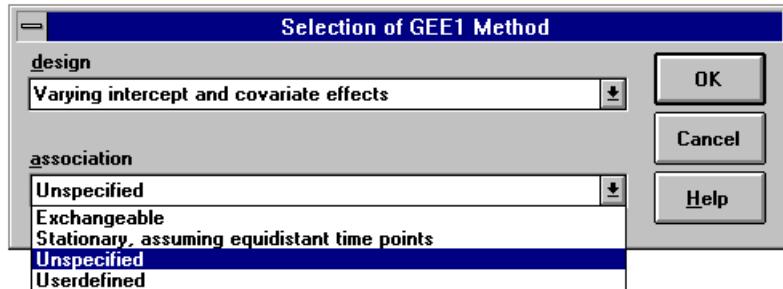


Figure 10: Specifying the design of the mean and association structure

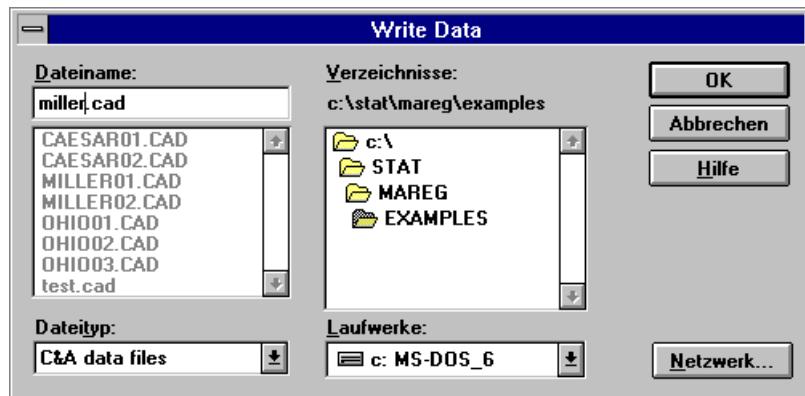


Figure 11: Saving the data and information files for MAREG

```

c:\stat\mareg\mareg.exe
parsing infile <c:\stat\mareg\examples\miller.cai> ...
MAREG version 0.0.7b. (c)25.07.1997 SFB386-C3. All rights reserved.
start reading data...
finished reading data...
--> Compute initial values for parameters beta by OLS Method...
--> Initial estimates of beta (OLS)
  0.12577
  0.562212
-0.0131579
  0.178186
  0.475403
-0.0767544
  0.166848
  0.446111
-0.0499513
  0.21511
  0.422317
-0.0409357
--> Compute IEE Estimator (scoring)...
--> IEE Estimator : 1

```

Figure 12: Running MAREG. The level of information shown can be selected in the general options dialog. The information is also written to the log file (extension \*.cal)

By default the results of the computations will be displayed after MAREG finishes (figure 13).

Var.	beta	std.	r. std.	Z	r. Z	p	r. p
resp(1)_1	-1.965	0.283	0.283	-6.942	-6.931	0.000	0.000
resp(2)_1	0.834	0.206	0.210	4.055	3.970	0.000	0.000
treat_1	-0.237	0.187	0.186	-1.266	-1.274	0.206	0.203
resp(1)_2	-1.270	0.265	0.262	-6.463	-6.533	0.000	0.000

Figure 13: Results of a MAREG run

## A short description of the output file (\*.cao)

This section briefly describes how to read the output file produced by MAREG.

The output file begins by giving some information on the files used by MAREG and the estimation method for the mean structure specified with WinMAREG along with some information on the convergence of the used algorithm.

```
MAREG version 0.0.7b. (c)25.07.1997 SFB386-C3. All rights reserved.
----- GEE Prentice estimator :
Inifile: miller.cai
Out file: miller.cao
Log file: miller.cal
Data file: miller.cad
Orig. sample size, #clusters: 444, 111
Act. sample size, #clusters: 444, 111
Estimation method: GEE
Special estimator: Method of Prentice
Link: cumulative logit link
Variance function: Multinomial variance function
Design: Varying intercept and covariate effects
User given epsilon: 1e-10
User given maxiter: 100
Tolerance beta reached: 4.08424e-12
Iterations needed: 14
Estimated overdispersion: no overdispersion was estimated
```

The next section displays the parameter estimates, their standard deviations and the resulting test statistics and p-values. The parameter estimates for the four time points are indicated by `xxx_1` to `xxx_4`, where intercept variables are indicated by the name of the response variable. A three categorical response variable leads to two thresholds (intercepts) indicated by (1) and (2). This way `resp(2)_3` is the second threshold for the third time point.

Var.	beta	std.	r. std.	Z	r. Z	p	r. p
resp(1)_1	-1.965	0.283	0.283	-6.942	-6.931	0.000	0.000
resp(2)_1	0.834	0.206	0.210	4.055	3.970	0.000	0.000
treat_1	-0.237	0.187	0.186	-1.266	-1.274	0.206	0.203
resp(1)_2	-1.710	0.265	0.262	-6.463	-6.533	0.000	0.000
resp(2)_2	0.723	0.212	0.215	3.411	3.364	0.001	0.001
treat_2	-0.744	0.193	0.193	-3.854	-3.849	0.000	0.000
resp(1)_3	-1.501	0.246	0.246	-6.111	-6.097	0.000	0.000
resp(2)_3	0.597	0.201	0.212	2.976	2.822	0.003	0.005
treat_3	-0.526	0.182	0.181	-2.885	-2.903	0.004	0.004
resp(1)_4	-1.300	0.227	0.245	-5.735	-5.317	0.000	0.000
resp(2)_4	0.589	0.195	0.200	3.015	2.948	0.003	0.003
treat_4	-0.312	0.178	0.178	-1.750	-1.753	0.080	0.080

As with the mean model, a short description of the association structure and the convergence of the used algorithm is given. The following table states the estimates of the association parameters. As the data in the analysed examples supply three categories and four time points we have  $(3-1)^2 \cdot 4 \cdot (4-1)/2 = 24$  parameters. They are displayed as associations between

time point 1/category 1 and time point 2/category 1,  
 time point 1/category 1 and time point 2/category 2,  
 ...  
 time point 1/category 1 and time point 4/category 2,  
 time point 1/category 2 and time point 2/category 1,  
 ...  
 time point 3/category 2 and time point 4/category 2,

Association model: Unspecified correlation model, method of Prentice  
 e  
 Tolerance alpha reached: 6.21491e-11  
 Score equation for alpha uses: Identity matrix

alpha	std.	r. std.	Z	r. Z	p	r. p
1.055	0.248	0.330	4.259	3.194	0.000	0.001
-0.299	0.194	0.183	-1.539	-1.633	0.124	0.102
-0.329	0.195	0.191	-1.688	-1.723	0.091	0.085
0.780	0.220	0.204	3.544	3.829	0.000	0.000
0.682	0.213	0.267	3.204	2.552	0.001	0.011
-0.148	0.191	0.177	-0.775	-0.836	0.439	0.403
-0.029	0.190	0.184	-0.150	-0.156	0.880	0.876
0.640	0.210	0.199	3.047	3.215	0.002	0.001
0.797	0.222	0.269	3.595	2.961	0.000	0.003
-0.244	0.193	0.166	-1.268	-1.473	0.205	0.141
-0.126	0.191	0.191	-0.662	-0.659	0.508	0.510
0.477	0.201	0.194	2.374	2.461	0.018	0.014
0.831	0.225	0.251	3.702	3.318	0.000	0.001
-0.217	0.192	0.186	-1.128	-1.162	0.259	0.245
-0.109	0.190	0.188	-0.574	-0.583	0.566	0.560
0.916	0.233	0.216	3.940	4.235	0.000	0.000
0.913	0.232	0.270	3.932	3.384	0.000	0.001
-0.126	0.191	0.188	-0.659	-0.668	0.510	0.504
-0.005	0.190	0.192	-0.028	-0.027	0.978	0.978
0.553	0.205	0.198	2.700	2.796	0.007	0.005
1.238	0.272	0.287	4.545	4.313	0.000	0.000
-0.298	0.194	0.179	-1.533	-1.659	0.125	0.097
-0.173	0.191	0.185	-0.903	-0.934	0.366	0.351
0.699	0.214	0.204	3.266	3.419	0.001	0.001

### 3 Likelihood Method

The following example illustrates the use of the ML method implemented in MAREG/WinMAREG. The data used are the 'Ohio Children Data'. This is a subset of the six-cities study, a longitudinal study of the health effects of air pollution. The response is the presence or absence of respiratory infection, observed for the age of 7, 8, 9 and 10 of the children in the study. The age of the child and the smoking status of the child's mother are observed covariates.

Opening the data file and selecting the logistic regression procedure from the ML-menu as in figure 14 opens the Model selection dialog shown in figure 15.

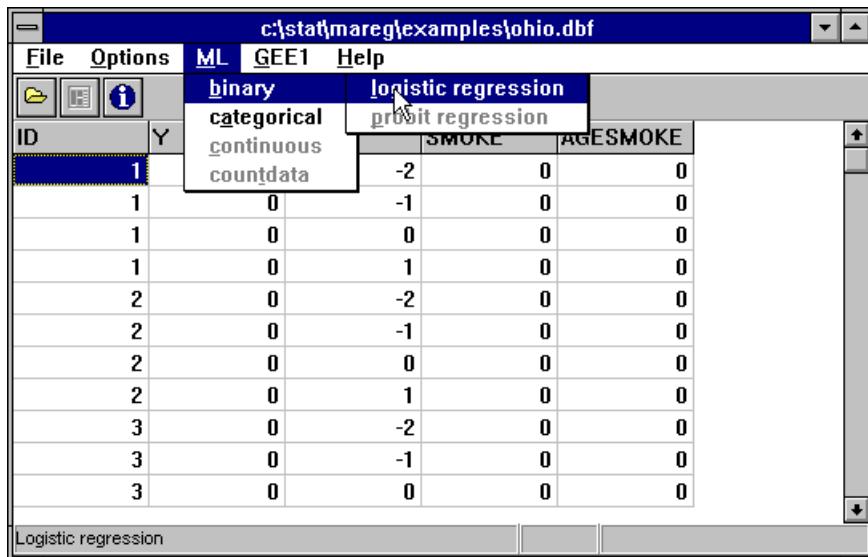


Figure 14: Selecting the logistic regression procedure from the ML-menu

The interaction 'age\*smoke' can't currently be created directly by WinMAREG (this is planned as a future feature). In this example the interaction is represented by the variable AGESMOKE, already provided in the data file.

The covariates are coded as follows: AGE is coded as the age in years since the child's 9th birthday, the smoking status SMOKE is coded as 0: mother does not smoke, 1: mother smokes. This means that the variables are already coded in the data file, which requires no further coding by WinMAREG. Figure 15 shows this selection for all three variables simultaneously.

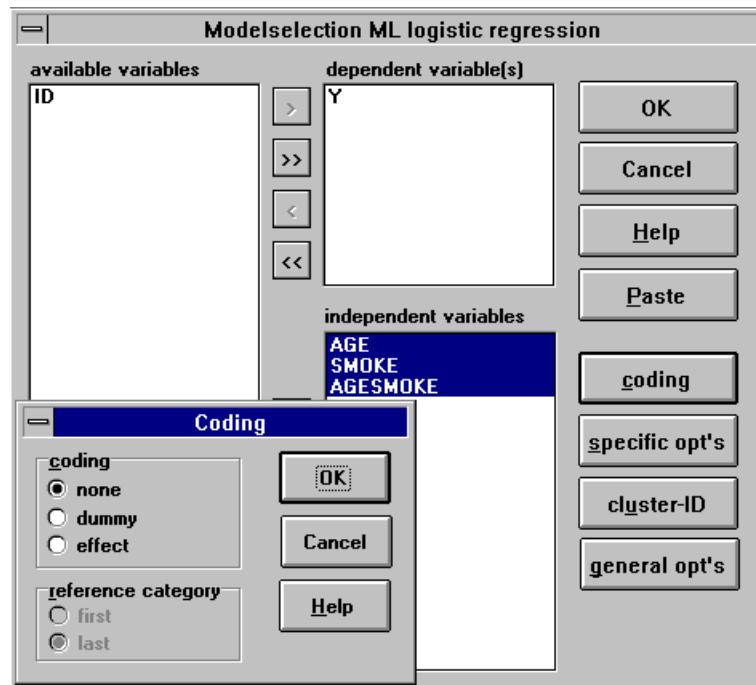


Figure 15: Model selection for ML

Clicking the specific options button opens the dialog shown in figure 16, where the design (matrix) of the mean and association structure can be specified. If you are unsure about your choice, clicking the help button opens the context sensitive help window, where you can find further information (see figure 17).

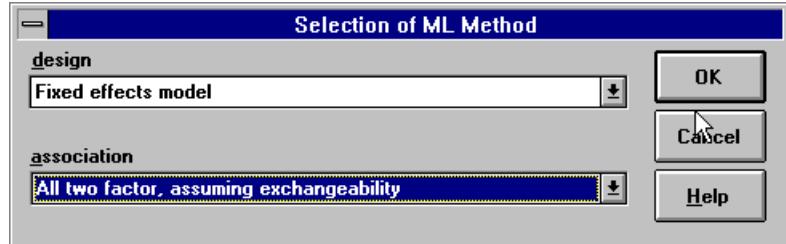


Figure 16: Specific options for ML

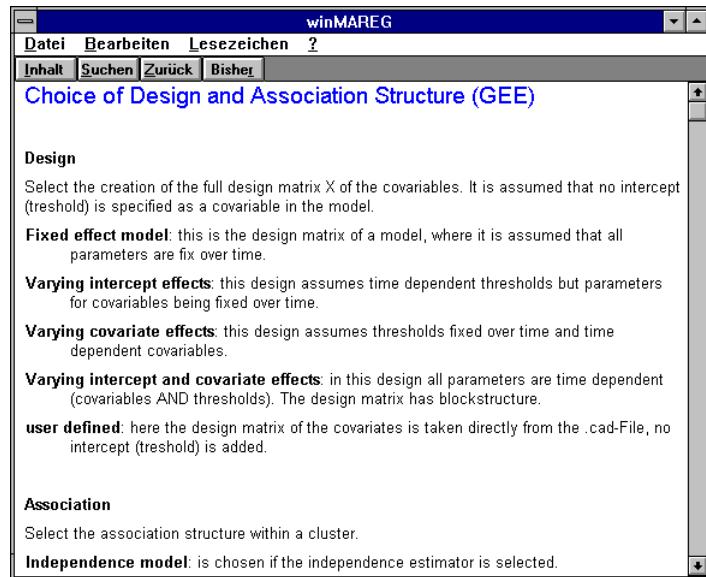


Figure 17: WinMAREG's online help

Clicking OK in the options dialog opens a further dialog (see figure 18) where we specify grouping of data. We select ‘grouping and table output’ which groups the data with respect to the single covariate strata (two in this example, as age is no stratifying variable) for computing. In addition this feature produces further output as we will see later.

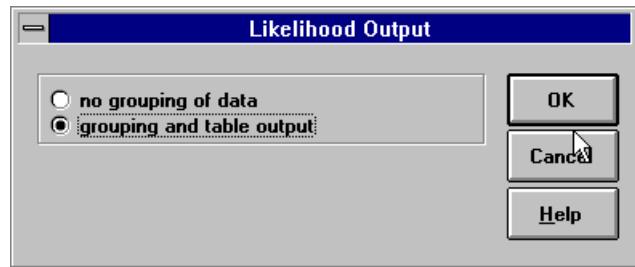


Figure 18: ML grouping dialog

Clicking OK brings us back to the model selection dialog (figure 15), where we proceed as in the first (GEE) example.

## Description of the output file

This section briefly describes how to read the output file produced by MAREG. The output file begins by giving some information on the files used by MAREG and the estimation method for the mean structure specified with WinMAREG along with some information on the convergence of the used algorithm.

As fixed effects were selected, there is only one set of parameters, as opposed to the GEE example above. The binary response corresponds to one intercept only (**y**). The exchangeable association structure results in one association parameter alpha. This information is completed with some measures for goodness of fit.

```
MAREG version 0.0.7b. (c)25.07.1997 SFB386-C3. All rights reserved.
----- LIKELIHOOD ESTIMATION WITH TABLE OUTPUT :
Inifile:          ohio.cai
Out file:         ohio.cao
Log file:         ohio.cal
Data file:        ohio.cad
Orig. sample size, #clusters: 2148, 537
Act. sample size, #clusters: 2148, 537
Estimation method: Maximum Likelihood
Special estimator: Conditional log odds ratios
Link:             cumulative logit link
Variance function: Binomial variance function
Design:           Fixed effects model
User given epsilon: 1e-10
User given maxiter: 100
Tolerance beta reached: 1.09741e-11
Iterations needed: 6
Estimated overdispersion: no overdispersion was estimated

      Var.     beta     std.   r. std.      Z     r. Z      p     r. p
        y    -1.901    0.117    0.119  -16.193  -15.962    0.000    0.000
        age   -0.141    0.057    0.058   -2.480   -2.424    0.013    0.015
       smoke   0.314    0.188    0.188   1.667   1.671    0.095    0.095
      agesmoke   0.071    0.089    0.088   0.799   0.801    0.424    0.423

Association model: All two factor association model assuming exchangeability
Tolerance alpha reached: 2.84473e-11

      alpha     std.   r. std.      Z     r. Z      p     r. p
        1.266    0.073    0.073  17.406  17.440    0.000    0.000

Likelihood: -797.391
Deviance: 16.387
Chisquare: 17.843
Cressie-Read (lambda=2/3): 17.215
```

### 3.1 Description of the log-file (\*.cal)

The log file contains the information printed to the screen during the MAREG run. As table output was chosen above, additional information is given here: the observed values, the expected values and the residuals of the different response patterns are presented for each stratum. They are ordered by the rule 'last index varying fastest', which means from '1,1,1,1', '1,1,1,0', ... to '0,0,0,0' (taking into account that '0' is the reference category).

```
[...]
--> Strata: 0
      age   smoke  agesmoke
      -2      0       0
      -1      0       0
      0      0       0
      1      0       0
  observed  expected      resid  std. res
  11.0000  10.0013  0.9987  0.3158
  5.0000   4.5488  0.4512  0.2116
  2.0000   3.6757 -1.6757 -0.8740
  6.0000   5.9277  0.0723  0.0297
  2.0000   2.9685 -0.9685 -0.5621
  3.0000   4.7871 -1.7871 -0.8168
  3.0000   3.8683 -0.8683 -0.4415
  24.0000  22.1189  1.8811  0.4000
  3.0000   2.3966  0.6034  0.3898
  7.0000   3.8649  3.1351  1.5947
  2.0000   3.1231 -1.1231 -0.6355
  16.0000  17.8575 -1.8575 -0.4396
  4.0000   2.5222  1.4778  0.9305
  15.0000  14.4216  0.5784  0.1523
  10.0000  11.6536 -1.6536 -0.4844
  237.0000 236.2643  0.7357  0.0479
  deviance, chisquare: 6.4079 / 6.6656

--> Strata: 1
      age   smoke  agesmoke
      -2.000  1.000  -2.000
      -1.000  1.000  -1.000
      0.000   1.000  0.000
      1.000   1.000  1.000
  observed  expected      resid  std. res
  7.0000   8.5384 -1.5384 -0.5265
  4.0000   2.9174  1.0826  0.6339
  2.0000   2.6031 -0.6031 -0.3738
  4.0000   3.1536  0.8464  0.4766
  1.0000   2.3226 -1.3226 -0.8678
  3.0000   2.8137  0.1863  0.1111
  3.0000   2.5106  0.4894  0.3089
  7.0000  10.7842 -3.7842 -1.1523
  4.0000   2.0723  1.9277  1.3391
  6.0000   2.5105  3.4895  2.2024
  1.0000   2.2400 -1.2400 -0.8285
  11.0000  9.6220  1.3780  0.4442
  2.0000   1.9986  0.0014  0.0010
  8.0000   8.5850 -0.5850 -0.1997
  6.0000   7.6602 -1.6602 -0.5999
  118.0000 116.6678  1.3322  0.1233
  deviance, chisquare: 9.9788 / 11.1770
[...]
```

## Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft.

## References

Fieger, A., Heumann, C. and Kastner, C. (1996). MAREG and winMAREG, *SFB386 – Discussion Paper 45*, Ludwig-Maximilians-Universität München.

Fitzmaurice, G. M. and Laird, N. M. (1993). A likelihood-based method for analysing longitudinal binary responses, *Biometrika* **80**: 141–151.

Miller, M. E., Davis, C. S. and Landis, R. J. (1993). The analysis of longitudinal polytomous data: Generalized estimating equations and connections with weighted least squares, *Biometrics* **49**: 1033–1044.