Kastner, Fieger, Heumann:

New features in MAREG 0.2.0

Sonderforschungsbereich 386, Paper 156 (1999)

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

New features in MAREG 0.2.0

C. Kastner* C. Heumann† A. Fieger‡

Institut für Statistik
Ludwig-Maximilians-Universität München

June 7, 1999

Abstract

This paper describes changes in the software tool MAREG from version 0.0.7 (July 1997) to version 0.2.0 (June 1999). As these new features are not implemented in WinMAREG yet, they can only be used via editing the ini-files (`*.cai`). Handling and features of Version 0.0.7 are described in Fieger, Heumann and Kastner (1996), Fieger, Kastner and Heumann (1998) and Kastner, Fieger and Heumann (1997).

1 Changes

1.1 Test statistics

There is no unique solution to testing hypothesis in generalized linear models. Instead, there exist several test statistics based on t or χ^2 distributions. In

*kchris@stat.uni-muenchen.de
†chris@stat.uni-muenchen.de
‡andreas@stat.uni-muenchen.de

version 0.0.7 the Z statistic, which was defined as

$$Z = \frac{\beta_j}{\sqrt{\mathrm{var}(\beta_j)}} \sim t\,,$$

was used in regression tables. To be consistent with the global Wald test, available now, the Wald statistic

$$\mathrm{Wald} = \frac{\beta_j^2}{\mathrm{var}(\beta_j)} \sim \chi^2\,.$$

is used instead of the Z statistic.

1.2 GEE output

The association parameters used in the method of Prentice (1988) are the correlations transformed by Fishers z. In version 0.2.0 these correlations (`rho`) are also displayed. Note, that the standard deviations, tests statistics and p-values which are listed correspond to the association parameter (not to the correlation).

2 New general options

This options described below are not depending on the estimation procedure.

2.1 Expected values of the response variable

It is now possible to get the expected values of the response variable given a specific stratum of the covariates. Additionally, the number of observations in each stratum, the sum of responses (y-sum) and the mean of the responses (y-mean) are given. Note, that the covariate strata are build using the design matrix X created by MAREG instead of the raw data.

With binary response the expectation $E(y|x)$ is the estimated probability of the event coded by '1', y-sum is the number of events in each stratum and y-mean is the relative response frequency (observed probability).

When the response is categorical (with k categories), we can use a cumulative logit model. In this case, MAREG creates a design matrix were each covariate stratum is given $k-1$ times. Therefore the number of observations per stratum is also given $k-1$ times.

2.2 Global Wald tests

It is now possible to specify a general linear hypothesis of the form

$$H_0 : C\theta = d \qquad \text{versus} \qquad H_1 : C\theta \neq d$$

using the Wald test. First the number of Wald tests has to be specified. This has to be done for the β and α parameters separately (`[options]`-section). Then a name for each test has to be specified. This is done in the sections `[waldnamesbeta]` and `[waldnamesalpha]`, respectively. At last, the matrices C and the vectors d have to be generated. A special syntax in the section `[waldtestsbeta]` and `[waldtestsalpha]` is used:

- The matrix C and the vector d are specified for each test in following form: Each row in the matrix C corresponds to a contrast of the form

$$c_s\theta_{(k)} = -c_t\theta_{(l)} \qquad \text{or} \qquad c_s\theta_{(k)} + c_t\theta_{(l)} = d$$

  where $k, l = 0, \ldots, p-1$ indicate the index in the original vectors β or α and $s, t = 0, \ldots, q-1$ correspond to the q parameters of θ.

- Each contrast is given by five values: the coefficients c_s and c_t, the indices k and l and the value d in the form: c_s, k, c_t, l, d

- Rows of the contrast matrix and the corresponding component of d are separated by ';'

- This results in one test being specified by a syntax like `waldbeta0=1,1,-1,2,0;1,2,-1,3,0;0,1,1,3,0`.
  The next test would be specified on the line beginning with `waldbeta1` and so on.

Example 1: Assume a categorical variable with four categories. In the model dummy-coding with the last category as reference category was used. Now you want to test the overall hypothesis

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$$

which corresponds to the test that the effects of all categories are equal. Note, that the reference category (β_4) is a priori set to zero. Then one set of contrasts may be $\beta_1 = \beta_2$, $\beta_2 = \beta_3$ and $\beta_3 = 0$. This corresponds to

$$C = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \theta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}, \quad d = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Usually the whole parameter vector β consists of more than these three parameters to be tested. Therefore, the position of these three parameters within the whole vector has to be specified (where the first position has the index 0). If the parameter vector contains only one more parameter (the intercept) at position 0, the syntax for this test would be:

```
[options]
waldtestsbeta=1
[waldtestsbeta]
waldbeta0=1,1,-1,2,0;1,2,-1,3,0;0,1,1,3,0
[waldnamesbeta]
namebeta0=overall-test
```

Example 2: Assume longitudinal binary data with $t = 3$ time points. Using an unspecified association structure, there are 3 parameters: $\alpha_0 \equiv \alpha_{12}$, $\alpha_1 \equiv \alpha_{13}$ and $\alpha_2 \equiv \alpha_{23}$. Now you want to test the hypothesis $H_0 : \alpha_0 = \alpha_2$, i.e. the association between equi-distant time points is equal. In this situation we get

$$C = \begin{pmatrix} 1 & -1 \end{pmatrix}, \quad \theta = \begin{pmatrix} \alpha_0 \\ \alpha_2 \end{pmatrix}, \quad d = 0$$

The syntax for this test is

```
[options]
waldtestsalpha=1
[waldtestsalpha]
waldalpha0=1,0,-1,2,0
[waldnamesalpha]
namealpha0=equi-distant
```

2.3 Specifying start values

You can specify start values for the parameters. The syntax is given in section 4.

2.4 Sequential logit link

For binary data the sequential logit link is now implemented, too. For details see e.g. Fahrmeir and Tutz (1994).

3 New options for the GEE module

For the Independence Estimating Equation (IEE) and the method of Prentice in the GEE module several methods for estimating the covariance matrix of the parameters are now available. For the IEE method you can choose between the usual implementation for the robust covariance matrix or the approximation of the jackknife estimator as described, e. g., by Ziegler (1997). Using the method of Prentice there are three methods for estimating the covariance matrix of the parameters:

1. standard estimation
2. stabilized covariance estimation
3. jackknife estimation

In contrast to the standard estimation of the covariance of θ, the stabilized form sets the matrix B in Prentice (1988, eq. (15)) to zero. The jackknife implementation is the natural extension of the jackknife for the IEE, as described in Ziegler (1997).

4 New values for `.cai` files

This section explains valid values for the items of the sections in the `.cai` files for MAREG. Expressions in brackets '<>' explain the expected type of value; '//' starts a comment (comments are not allowed in `.cai` files that are actually used with MAREG, they are only used here for documentation).

```
[link]
link=<integer>          // 1=identity link
                        // 2=cumulative logit link
                        // 3=multinomial logit link
                        // 4=sequential logit link

[betastart]
beta0=<float>           // value of first beta parameter
beta1=<float>           // value of first beta parameter
...
```

```
[alphastart]
alpha0=<float>          // value of first alpha parameter
alpha1=<float>          // value of first alpha parameter
...

[options]
printexpectedvalues=<integer>              // 1=print expected values in log-file, 0=don't
prenticestabilizedvariance=<integer>       // 0=standard, 1=stabilized estimation
jackknifedvariance=<integer>               // 0=standard, 1=jackknife estimation
                                           // NOTE: only prenticestabilizedvariance
                                           // or jackknifedvariance can be used
waldtestsbeta=<integer>                    // number of tests for beta parameters
waldtestsalpha=<integer>                   // number of tests for alpha parameters

[waldtestsbeta]
waldbeta0=<float>,<integer>,<float>,<integer>,<float>;...
                                           //contrast matrix of first wald test for beta
waldbeta1=<float>,<integer>,<float>,<integer>,<float>;...
                                           //contrast matrix of second wald test for beta
...

[waldnamesbeta]
namebeta0=<string>                 // name of the first wald test for beta parameters
namebeta1=<string>                 // name of the second wald test for beta parameters
...

[waldtestsalpha]
waldalpha0=<float>,<integer>,<float>,<integer>,<float>;...
                                   //contrast matrix of first wald test for alpha
waldalpha1=<float>,<integer>,<float>,<integer>,<float>;...
                                   //contrast matrix of second wald test for alpha
...

[waldtestalpha]
namealpha0=<string>                // name of the first wald test for alpha parameters
namealpha1=<string>                // name of the second wald test for alpha parameters
...
```

5 Licensing agreement

If you use WinMAREG and/or MAREG, please acknowledge in presentations or publications.

6 Availability

The latest versions can be obtained via anonymous ftp from `ftp.stat.uni-muenchen.de`. It is located in the directory `/pub/sfb386/c3/mareg`. The following distributions are available:

- `w32mareg.zip`, Windows NT/Windows 95 version of MAREG plus WinMAREG.
- `Solaris_mareg.tar.gz`, Solaris 2.6 version of MAREG.
- `SolarisOpt_mareg.tar.gz`, (runtime) optimized Solaris 2.6 version of MAREG.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft.

References

Fahrmeir, L. and Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer, New York.

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

Fieger, A., Kastner, C. and Heumann, C. (1998). WinMAREG quick start, *SFB386 – Discussion paper 108*, Ludwig-Maximilians-Universität München.

Kastner, C., Fieger, A. and Heumann, C. (1997). MAREG and WinMAREG—a tool for marginal regression models, *Computational Statistics and Data Analysis* **24**: 235–241.

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation, *Biometrics* **44**: 1033–1048.

Ziegler, A. (1997). Practical considerations of the jackknife estimator of variance for generalized estimationg equations, *Statistical Papers* **38**: 363–369.