Package ‘multgee’

February 15, 2013

Type Package

Title GEE Solver for Correlated Nominal or Ordinal Multinomial Responses

Version 1.1

Date 2012-11-12

Author Anestis Touloumis <anestis@ebi.ac.uk>

Maintainer Anestis Touloumis <anestis@ebi.ac.uk>

Depends R (>= 2.15.0), gnm, VGAM

Description GEE solver for correlated nominal or ordinal multinomial responses using a local odds ratios parameterization.

License GPL-3

LazyLoad yes

Repository CRAN

Date/Publication 2012-11-12 16:25:42

NeedsCompilation no

R topics documented:

multgee-package ................................................... 2
arthritis ............................................................... 3
housing ............................................................... 4
intrinsic.pars ....................................................... 5
ipfp.control ......................................................... 6
LORgee.control ..................................................... 7
matrixLOR ........................................................... 8
nomLORgee .......................................................... 9
ordLORgee .......................................................... 12

Index 15
multgee-package

A GEE Solver For Correlated Nominal Or Ordinal Multinomial Responses

Description

A generalized estimating equations (GEE) solver for fitting marginal regression models with correlated nominal or ordinal multinomial responses using a local odds ratios parameterization for the association structure.

Details

The package contains two core functions; ordLORgee for an ordinal response scale, and nomLORgee for a nominal response scale.

The following arguments are required in both functions: (i) a data frame (data), (ii) a model formula (formula), (iii) a subject identifier variable (id) and (iv) a vector that identifies the order of observations within each subject (repeated). Note that data must contain id, repeated and any variable present in formula.

Options for the marginal model in ordLORgee include cumulative link models or an adjacent category logit model, determined by the argument link. A marginal baseline category logit model is offered in nomLORgee. See the help files of nomLORgee and ordLORgee for the form of the linear predictor.

The association structure among the correlated multinomial responses is expressed via local odds ratios (Touloumis, Agresti and Kateri, 2012). The estimating procedure for the local odds ratios can be summarized as follows: For each level pair of the repeated variable, the responses are aggregated across subjects to form a square marginalized contingency table. Treating these tables as independent, the argument LORem indicates whether a model for counts is fitted either separately to each marginalized contingency table ("2way") or simultaneously to the full marginalized contingency table ("3way"). The model applied to the counts is the RC-G model proposed by Becker and Clogg (1989) and it belongs to the family of association models developed by Goodman (1981). The argument LORstr determines the form of the local odds ratios structure.

If the underlying association pattern does not change dramatically across the level pairs of repeated then a simple form for the local odds ratios should be a sufficient approximation for the association structure. To assess this, one might compare the intrinsic parameters of an RC-G model. See the utility function intrinsic.pars for more details.

Alternatively a user-defined local odds ratios structure can be provided if LORstr is "fixed". In this case, the utility function matrixLOR is useful in constructing the required LORterm.

Author(s)

Anestis Touloumis Maintainer: Anestis Touloumis <anestis@ebi.ac.uk>


References


See Also

`nomLORgee` and `ordLORgee`.

Examples

data(arthritis)
data <- arthritis
fitord <- ordLORgee(y~factor(time)+factor(trt)+factor(baseline),
                    id="id",repeated="time",data=data)
summary(fitord)

data(housing)
data <- housing
fitnom <- nomLORgee(y~factor(time)*sec, id="id",repeated="time",data=data)
summary(fitnom)

---

**arthritis**  
*Rheumatoid Arthritis Clinical Trial*

Description

Rheumatoid self-assessment scores for 302 patients, measured on a five-level ordinal response scale at three follow-up times.

Usage

data(arthritis)

Format

A data frame with 906 observations on the following 7 variables:

id  Patient identifier variable.
y  Self-assessment score of rheumatoid arthritis measured on a five-level ordinal response scale.
sex  Coded as (1) for female and (2) for male.
age  Recorded at the baseline.
trt  Treatment group variable, coded as (1) for the placebo group and (2) for the drug group.
baseline Self-assessment score of rheumatoid arthritis at the baseline.
time  Measurement occasions variable, coded as (1) at one month, (2) at three months and (3) at five months.

Source


Examples

data(arthritis)
str(arthritis)

data(housing)
str(housing)

Description

Housing status for 362 severely mentally ill homeless subjects measured at baseline and at three follow-up times.

Usage

data(housing)

Format

A data frame with 1448 observations on the following 4 variables:

id  Subject identifier variable.
y  Housing status response, coded as (0) for street living, (1) for community living and (2) for independent housing.
time Measurement occasions variable, coded as (0) at baseline, (1) at 6 months, (2) at twelve months and (3) at twenty four months.
sec  Section 8 rent certificate indicator.

Source


Examples

data(housing)
str(housing)
Intrinsic Parameters Estimation

Description

Utility function to assess the underlying association pattern.

Usage

`intrinsic.pars(response, id, repeated, ncategories, rscale = "ordinal")`

Arguments

- `response`: a vector that indicates the response variables.
- `id`: a vector that identifies the subjects.
- `repeated`: a vector that identifies the order of the observations within each subject.
- `ncategories`: a positive constant that indicates the maximum number of the observed response categories.
- `rscale`: a character string that indicates the nature of the response scale. Options include "ordinal" or "nominal".

Details

Simulation studies in Touloumis, Agresti and Kateri (2012) suggested that if the range of the intrinsic parameter estimates is small then simple local odds ratios structures should adequately approximate the association pattern.

The intrinsic parameters are estimated under the heterogeneous uniform association model for ordinal response categories and under the RC-G model with homogeneous score parameters for nominal response categories.

Value

Returns a vector with the estimated intrinsic parameters.

Author(s)

Anestis Touloumis

References


See Also

`nomLORgee` and `ordLORgee`
Examples

data(arthritis)
intrinsic.pars(arthritis$y, arthritis$id, arthritis$time, 5)
## The intrinsic parameters do not differ much. The 'uniform' local odds ratios
## structure might be a good approximation for the association pattern.

set.seed(1)
data(housing)
intrinsic.pars(housing$y, housing$id, housing$time, 3, rscale="nominal")
## The intrinsic parameters seem to vary. The 'RC' local odds ratios structure
## might be a good approximation for the association pattern

ipfp.control

Description

Control function for the Iterative Proportion Fitting Procedure function ipfp.

Usage

ipfp.control(tol = 1e-06, maxit = 200)

Arguments

  tol  positive convergence tolerance. The algorithm converges when the absolute dif-
        ference between the observed and the given row or column totals is less than or
        equal to tol.

  maxit positive integer that indicates the maximum number of iterations.

Note

Currently ipfp function is internal.

Author(s)

Anestis Touloumis

See Also

nomLORgee and ordLORgee.
Control function for the GEE solver in the \texttt{nomLORgee} and \texttt{ordLORgee} functions.

**Usage**

\texttt{LORgee.control(tolerance = 0.001, maxiter = 15, verbose = FALSE, TRACE = FALSE)}

**Arguments**

- **tolerance**: positive convergence tolerance. The algorithm converges when the pairwise maximum of the absolute relative difference in parameter estimates is less than or equal to tolerance.
- **maxiter**: positive integer that indicates the maximum number of iterations in the Fisher-scoring iterative procedure.
- **verbose**: logical that indicates if output should be printed for each iteration.
- **TRACE**: logical that indicates if the parameter estimates and the convergence criterion for each iteration should be saved.

**Author(s)**

Anestis Touloumis

**See Also**

\texttt{nomLORgee} and \texttt{ordLORgee}.

**Examples**

\begin{verbatim}
data(arthritis)
fitmod <- ordLORgee(y~factor(trt)+factor(baseline)+factor(time), id = "id",
                      repeated = "time", data = arthritis)
## A one-step GEE estimator
fitmod1 <- update(fitmod, control = LORgee.control(maxiter=1))
coef(fitmod)
coef(fitmod1)
\end{verbatim}
Creating A Probability Matrix With Specified Local Odds Ratios

Description
Utility function to create a square probability matrix with a specified local odds ratios structure.

Usage
matrixLOR(x)

Arguments
x
a square matrix with positive entries that describes the desired local odds ratios matrix.

Details
This function is designed to ease the construction of a fixed local odds ratios structure in the functions nomLORgee or ordLORgee. See also the example.

Value
Returns a square probability matrix that satisfies the local odds ratios structure described by x.

Warning
Caution is needed for local odds ratios close to zero.

Author(s)
Anestis Touloumis

See Also
nomLORgee and ordLORgee.

Examples
## Illustrating the construction of a "fixed" local odds ratio structure
## using the arthritis dataset. Here, we assume a uniform local odds ratios
## structure equal to 2 for each time pair.

## Create the uniform local odds ratios structure.
lorterm <- matrixLOR(matrix(2,4,4))

## Create the LORterm argument.
lorterm <- c(lorterm)
lorterm <- matrix(c(lorterm),3,25,TRUE)
data(arthritis)
fitmod.fixed <- ordLORgee(y~factor(trt)+factor(time)+factor(baseline), id="id",
    repeated="time", data=arthritis, LORstr="fixed", LORterm=lorterm)
fitmod.fixed

---

**Marginal Models For Correlated Nominal Multinomial Responses**

**Description**

Solving the generalized estimating equations for correlated nominal multinomial responses assuming a baseline category logit model for the marginal probabilities.

**Usage**

```r
nomLORgee(formula = formula, data = data, id = id, repeated = repeated,
    bstart = NULL, LORstr = "time.exch", LORem = "3way", LORterm = NULL,
    add = 0, homogeneous = TRUE, control = LORgee.control(),
    ipfp.ctrl = ipfp.control(), IM = "solve")
```

**Arguments**

- **formula**: a formula expression as for other regression models for multinomial responses. An intercept term must be included.
- **data**: a mandatory data frame that should include the variables provided in the formula, id and repeated arguments.
- **id**: a vector that identifies the subjects.
- **repeated**: a vector that identifies the order of the observations within each subject.
- **bstart**: a vector that includes an initial estimate for the marginal regression parameter vector.
- **LORstr**: a character string that indicates the local odds ratios structure. Options include "independence", "time.exch", "RC" or "fixed".
- **LORem**: a character string that indicates if the marginalized local odds ratios structure is estimated simultaneously ("3way") or seperately at each level pair of repeated ("2way").
- **LORterm**: a matrix that contains the desired local odds ratios structure. It should be used when LORstr is "fixed".
- **add**: a positive constant to be added at each cell of the full marginalized contingency table in the presence of zero observed counts.
- **homogeneous**: a logical that indicates homogeneous score parameters when LORstr is "time.exch" or "RC".
- **control**: a vector that specifies the control variables for the GEE solver.
- **ipfp.ctrl**: a vector that specifies the control variables for the ipfp function.
- **IM**: a character string that indicates the method used for inverting a matrix. Options include "solve", "qr.solve" or "cholesky".
Details

The data must be provided in a subject level or equivalently in ‘long’ format. See details about the ‘long’ format in the \texttt{reshape} function.

A term of the form \texttt{offset(expression)} is allowed in the \texttt{formula}.

The \texttt{id} and the \texttt{repeated} do not need to be pre-sorted. Instead the function reshapes data in an ascending order of \texttt{id} and \texttt{repeated}.

The default set for the response categories is 1, \ldots, I, where I \(> 2\) is the maximum observed response category. If otherwise, the function recodes the observed response categories onto this set.

The default set for the levels of \texttt{repeated} is 1, \ldots, T, where T is the number of observed levels. If otherwise, the function recodes the observed levels onto this set.

The I\textsuperscript{th} response category is treated as baseline.

The linear predictor is of the form

\[
\beta_{0j} + \beta_{j}'x_{it}
\]

where \(\beta_{0j}\) is the \(j\)-th intercept, \(\beta_{j}\) is the \(j\)-th response category specific parameter vector and \(x_{it}\) is the covariate vector for the \(i\)-th subject at the \(t\)-th level of \texttt{repeated}.

The \texttt{LORterm} argument must be an \(L \times I^2\) matrix, where \(L\) is the number of level pairs of \texttt{repeated}. These are ordered as \((1, 2), (1, 3), \ldots, (1, T), (2, 3), \ldots, (T - 1, T)\) and the rows of \texttt{LORterm} are supposed to preserve this order. Each row is assumed to contain the vectorized form of a probability table that satisfies the desired local odds ratios structure.

Value

Returns an object of the class “\texttt{LORgee}”. This has components:

- \texttt{title}: title for the GEE model.
- \texttt{version}: the current version of the GEE solver.
- \texttt{link}: the marginal link function.
- \texttt{odds.ratio}: the ‘working’ local odds ratios structure name along with the estimates.
- \texttt{terms}: the terms structure describing the model.
- \texttt{contrasts}: the contrasts used for the factors.
- \texttt{nobs}: the number of observations after recoding the response variables into equivalent binary responses.
- \texttt{convergence}: the values of convergence variables.
- \texttt{coefficients}: the estimated regression parameter vector of the marginal model.
- \texttt{linear.pred}: the estimated linear predictor of the marginal regression model with \(j\)-th column corresponding to the \(j\)-th response category.
- \texttt{fitted.values}: the estimated fitted values of the marginal regression model with \(j\)-th column corresponding to the \(j\)-th response category.
- \texttt{residuals}: the residuals of the marginal regression model based on the binary responses. The \(j\)-th column corresponds to the \(j\)-th response category.
- \texttt{y}: the multinomial response variables.
- \texttt{id}: the \texttt{id} variable based on the binary responses.
nomLORgee

max.id the number of patients.
clusz the number of observations within each patient.
robust.variance the estimated "robust" covariance matrix.
naive.variance the estimated "naive" or model-based covariance matrix.
xnames the regression coefficients' symbolic names.
categories the number of observed response categories.
occasions the levels of the repeated variable.
gee.control the control values for the GEE solver.
ipfp.control the control values for the ipfp function.
inverse.method the method used for inverting matrices.
adding.constant the value used for add.
call the matched call.
pvalue the p-value based on a Wald test that no covariates are statistically significant.

Generic coef, summary, print, fitted and residuals methods are available. The pvalue of the Null model corresponds to the hypothesis $H_0: \beta_1 = \ldots = \beta_{I-1} = 0$ based on the Wald test statistic.

Author(s)

Anestis Touloumis

References


See Also

For an ordinal response scale use ordLORgee.

Examples

data(housing)
data <- housing
fitmod <- nomLORgee(y~factor(time)*sec, id="id",repeated="time",data=data)
summary(fitmod)
coef(fitmod)
Description

Solving the generalized estimating equations for correlated ordinal multinomial responses assuming a cumulative link model or an adjacent category logit model for the marginal probabilities.

Usage

```r
ordLORgee(formula = formula, data = data, id = id, repeated = repeated,
  link = "logistic", bstart = NULL, LORstr = "category.exch",
  LORem = "3way", LORterm = NULL, add = 0, homogeneous = TRUE,
  restricted = FALSE, control = LORgee.control(),
  ipfp.ctrl = ipfp.control(), IM = "solve")
```

Arguments

- `formula`: a formula expression as for other regression models for multinomial responses. An intercept term must be included.
- `data`: a mandatory data frame that should include the variables provided in the `formula`, `id` and `repeated` arguments.
- `id`: a vector that identifies the subjects.
- `repeated`: a vector that identifies the order of the observations within each subject.
- `link`: a character string that specifies the link function. Options include "logistic", "probit", "cauchit", "cloglog" and "acl".
- `bstart`: a vector that includes an initial estimate for the marginal regression parameter vector.
- `LORstr`: a character string that indicates the local odds ratios structure. Options include "independence", "uniform", "category.exch", "time.exch", "RC" or "fixed".
- `LORem`: a character string that indicates if the marginalized local odds ratios structure is estimated simultaneously ("3way") or seperately at each level pair of `repeated` ("2way").
- `LORterm`: a matrix that contains the desired local odds ratios structure. It should be used when `LORstr` is "fixed".
- `add`: a positive constant to be added at each cell of the full marginalized contingency table in the presence of zero observed counts.
- `homogeneous`: a logical that indicates homogeneous score parameters when `LORstr` is "time.exch" or "RC".
- `restricted`: a logical that indicates monotone score parameters when `LORstr` is "time.exch" or "RC".
- `control`: a vector that specifies the control variables for the GEE solver.
ipfp.ctrl  a vector that specifies the control variables for the ipfp function.
IM       a character string that indicates the method used for inverting a matrix. Options include "solve", "qr.solve" or "cholesky".

Details

The data must be provided in a subject level or equivalently in ‘long’ format. See details about the ‘long’ format in the reshape function.

A term of the form offset(expression) is allowed in the formula.

The id and the repeated do not need to be pre-sorted. Instead the function reshapes data in an ascending order of id and repeated.

The default set for the response categories is 1, ..., I, where I > 2 is the maximum observed response category. If otherwise, the function recodes the observed response categories onto this set.

The default set for the levels of repeated is 1, ..., T, where T is the number of observed levels. If otherwise, the function recodes the observed levels onto this set.

The I-th response category is omitted.

An adjacent category logit model is fitted if and only if link is "acl". Otherwise a cumulative link model is fitted.

The linear predictor is of the form

$$\beta_{0j} + \beta \cdot x_{it}$$

where $$\beta_{0j}$$ is the j-th intercept and $$x_{it}$$ is the covariate vector for the i-th subject at the t-th level of repeated.

The LORterm argument must be an L x I^2 matrix, where L is the number of level pairs of repeated. These are ordered as (1, 2), (1, 3), ..., (1, T), (2, 3), ..., (T - 1, T) and the rows of LORterm are supposed to preserve this order. Each row is assumed to contain the vectorized form of a probability table that satisfies the desired local odds ratios structure.

Value

Returns an object of the class "LORgee". This has components:

title       title for the GEE model.
version     the current version of the GEE solver.
link        the marginal link function.
odds.ratio  the ‘working’ local odds ratios structure name along with the estimates.
terms       the terms structure describing the model.
contrasts    the contrasts used for the factors.
nobs        the number of observations after recoding the response variables into equivalent binary responses.
convergence the values of convergence variables.
coefficients the estimated regression parameter vector of the marginal model.
linear.pred the estimated linear predictor of the marginal regression model with j-th column corresponding to the j-th response category.
fitted.values  the estimated fitted values of the marginal regression model with j-th column corresponding to the j-th response category.
residuals    the residuals of the marginal regression model based on the binary responses. The j-th column corresponds to the j-th response category.
y            the multinomial response variables.
id            the id variable based on the binary responses.
max.id        the number of patients.
clusz         the number of observations within each patient.
robust.variance the estimated "robust" covariance matrix.
naive.variance the estimated "naive" or model-based covariance matrix.
xnames        the regression coefficients' symbolic names.
categories    the number of observed response categories.
ocasions      the levels of the repeated variable.
gee.control   the control values for the GEE solver.
ipfp.control  the control values for the ipfp function.
inverse.method the method used for inverting matrices.
adding.constant the value used for add.
call          the matched call.
pvalue        the p-value based on a Wald test that no covariates are statistically significant.

Generic coef, summary, print, fitted and residuals methods are available. The pvalue of the Null model corresponds to the hypothesis $H_0: \beta_1 = ... = \beta_{I-1} = 0$ based on the Wald test statistic.

Author(s)
Anestis Touloumis

References

See Also
For a nominal response scale use nomLORgee.

Examples
```
data(arthritis)
intrinsic.pars(arthritis$y, arthritis$id, arthritis$time, 5)
fitmod <- ordLORgee(y~factor(time)+factor(trt)+factor(baseline),id="id", repeated="time",data=arthritis, LORstr="uniform")
summary(fitmod)
```
Index

*Topic datasets
  arthritis, 3
  housing, 4
*Topic package
  multgee-package, 2

arthritis, 3
coeff, 11, 14
fitted, 11, 14
housing, 4
intrinsic.pars, 2, 5
ipfp.control, 6
LORgee.control, 7
matrixLOR, 2, 8
multgee (multgee-package), 2
multgee-package, 2
nomLORgee, 2, 3, 5–8, 9, 14
ordLORgee, 2, 3, 5–8, 11, 12
print, 11, 14
reshape, 10, 13
residuals, 11, 14
summary, 11, 14