

Package ‘multgee’

July 11, 2017

Type Package

Title GEE Solver for Correlated Nominal or Ordinal Multinomial Responses

Version 1.6.0

Author Anestis Touloumis <A.Touloumis@brighton.ac.uk>

Maintainer Anestis Touloumis <A.Touloumis@brighton.ac.uk>

Depends R (>= 2.15.0), gnm, VGAM

Imports utils, stats

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

License GPL-2 | GPL-3

LazyLoad yes

URL <http://github.com/AnestisTouloumis/multgee>

BugReports <http://github.com/AnestisTouloumis/multgee/issues>

NeedsCompilation no

Repository CRAN

Date/Publication 2017-07-11 21:45:42 UTC

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
waldts	15

Index	17
--------------	-----------

multgee-package	<i>A GEE Solver For Correlated Nominal Or Ordinal Multinomial Responses</i>
-----------------	---

Description

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

Details

The package contains two functions that fit GEE models for correlated multinomial responses; [ordLORgee](#) for an ordinal response scale and [nomLORgee](#) for a nominal response scale.

The main arguments in both functions are: (i) an optional data frame (`data`), (ii) a model formula (`formula`), (iii) a cluster identifier variable (`id`) and (iv) an optional vector that identifies the order of the observations within each cluster (`repeated`).

Options for the marginal model in the function [ordLORgee](#) include cumulative link models or an adjacent categories logit model. A marginal baseline category logit model is offered in the function [nomLORgee](#). For the form of the linear predictor in these models, see the Details sections in [nomLORgee](#) and [ordLORgee](#).

The association structure among the correlated multinomial responses is expressed via marginalized local odds ratios (*Touloumis et al., 2013*). The estimating procedure for the local odds ratios can be summarized as follows: For each level pair of the repeated variable, the available responses are aggregated across clusters to form a square marginalized contingency table. Treating these tables as independent, an RC-G(1) type model (*Becker and Clogg, 1989*) is fitted in order to estimate the marginalized local odds ratios. The `LORstr` argument determines the form of the marginalized local odds ratios structure. Since the general RC-G(1) model is closely related to the family of association models (*Goodman, 1985*), one can instead fit an association model to each of the marginalized contingency tables by setting `LORem="2way"`.

If the underlying association pattern does not change dramatically across the level pairs of repeated then parsimonious marginalized local odds ratios should sufficiently approximate the true underlying association structure. To assess the underlying association structure, one might use the utility function [intrinsic.pars](#).

Instead of estimating the local odds ratios structure, a user-defined structure can be provided by setting `LORstr="fixed"`. In this case, the utility function [matrixLOR](#) is useful in constructing the required `LORterm` argument.

The function [waldts](#) provides a goodness-of-fit test between two nested GEE models based on a Wald test statistic.

Author(s)

Anestis Touloumis Maintainer: Anestis Touloumis <A.Touloumis@brighton.ac.uk>

References

Becker, M. and Clogg, C. (1989) Analysis of sets of two-way contingency tables using association models. *Journal of the American Statistical Association*, **84**, 142-151.

Goodman, L. (1985) The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetry models for contingency tables with or without missing entries. *The Annals of Statistics*, **13**, 10-69.

Touloumis, A., Agresti, A. and Kateri, M. (2013) GEE for multinomial responses using a local odds ratios parameterization. *Biometrics*, **69**, 633-640.

Touloumis, A. (2015) R Package multgee: A Generalized Estimating Equations Solver for Multinomial Responses. *Journal of Statistical Software*, **64**, 1-14.

See Also

For a nominal response scale use the function [nomLORgee](#). For an ordinal response scale use the function [ordLORgee](#).

Examples

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

```
data(housing)
fitnom <- nomLORgee(y~factor(time)*sec, data=housing, id=id, repeated=time)
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 Follow-up time recorded in months.

Source

Lipsitz, S.R. and Kim, K. and Zhao, L. (1994) Analysis of repeated categorical data using generalized estimating equations. *Statistics in Medicine*, **13**, 1149–1163.

Examples

```
data(arthritis)
str(arthritis)
```

housing

Homeless Data

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 (1) for street living, (2) for community living and (3) for independent housing.

time Time recorded in months.

sec Section 8 rent certificate indicator.

Source

Hulburt M.S., Wood, P.A. and Hough, R.L. (1996) Providing independent housing for the homeless mentally ill: a novel approach to evaluating longitudinal housing patterns. *Journal of Community Psychology*, **24**, 291–310.

Examples

```
data(housing)
str(housing)
```

intrinsic.pars	<i>Intrinsic Parameters Estimation</i>
----------------	--

Description

Utility function to assess the underlying association pattern.

Usage

```
intrinsic.pars(y = y, data, id = id, repeated = NULL, rscale="ordinal")
```

Arguments

y	a vector that identifies the response vector of the desired marginal model.
data	an optional data frame containing the variables provided in y, id and repeated.
id	a vector that identifies the clusters.
repeated	an optional vector that identifies the order of observations within each cluster.
rscale	a character string that indicates the nature of the response scale. Options include "ordinal" or "nominal".

Details

Simulation studies in *Touloumis et al. (2013)* suggested that if the range of the intrinsic parameter estimates is small then simple local odds ratios structures should adequately approximate the association pattern. Otherwise more complicated structures should be employed.

The intrinsic parameters are estimated under the heterogeneous linear-by-linear association model (*Agresti, 2013*) for ordinal response categories and under the RC-G(1) model (*Becker and Clogg, 1989*) with homogeneous score parameters for nominal response categories.

A detailed description of the arguments `id` and `repeated` can be found in the Details section of [nomLORgee](#) or [ordLORgee](#).

Value

Returns a numerical vector with the estimated intrinsic parameters.

Author(s)

Anestis Touloumis

References

Agresti, A. (2013) *Categorical Data Analysis*. New York: John Wiley and Sons, Inc., 3rd Edition.

Becker, M. and Clogg, C. (1989) Analysis of sets of two-way contingency tables using association models. *Journal of the American Statistical Association*, **84**, 142-151.

Touloumis, A., Agresti, A. and Kateri, M. (2013) GEE for multinomial responses using a local odds ratios parameterization. *Biometrics*, **69**, 633-640.

See Also

[nomLORgee](#) and [ordLORgee](#).

Examples

```
data(arthritis)
intrinsic.pars(y,arthritis,id,time,rscale="ordinal")
## 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(y,housing,id,time,rscale="nominal")
## The intrinsic parameters vary. The 'RC' local odds ratios structure
## might be a good approximation for the association pattern.
```

ipfp.control

IPFP Control

Description

Control variables for the Iterative Proportion Fitting Procedure function `ipfp`.

Usage

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

Arguments

<code>tol</code>	positive convergence tolerance. The algorithm converges when the absolute difference between the observed and the given row or column totals is less than or equal to <code>tol</code> .
<code>maxit</code>	positive integer that indicates the maximum number of iterations.

Note

Currently the function `ipfp` is internal.

Author(s)

Anestis Touloumis

See Also

[nomLORgee](#) and [ordLORgee](#).

LORgee.control *Control For The GEE Solver*

Description

Control variables for the GEE solver in the functions [nomLORgee](#) and [ordLORgee](#).

Usage

```
LORgee.control(tolerance = 0.001, maxiter = 15, verbose = FALSE, TRACE = FALSE)
```

Arguments

tolerance	positive convergence tolerance. The algorithm converges when the 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 algorithm.
verbose	logical that indicates if output should be printed at each iteration.
TRACE	logical that indicates if the parameter estimates and the convergence criterion at each iteration should be saved.

Author(s)

Anestis Touloumis

See Also

[nomLORgee](#) and [ordLORgee](#).

Examples

```
data(arthritis)
fitmod <- ordLORgee(y~factor(trt)+factor(baseline)+factor(time),
                  data = arthritis, id = id, repeated = time)
## A one-step GEE estimator
fitmod1 <- update(fitmod, control = LORgee.control(maxiter=1))
coef(fitmod)
coef(fitmod1)
```

Description

Utility function to create a square probability matrix that satisfies the 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 the `LORterm` argument in the functions [nomLORgee](#) and [ordLORgee](#).

Value

Returns a square probability matrix that satisfies the local odds ratios structure defined 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 ratios 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)

## Fit the marginal model.
data(arthritis)
fitmod.fixed <- ordLORgee(y~factor(trt)+factor(time)+factor(baseline),data=arthritis,
  id=id,repeated=time,LORstr="fixed",LORterm=lorterm)

fitmod.fixed

```

nomLORgee

*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

```

nomLORgee(formula, data, id = id, repeated = NULL,
  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	an optional data frame containing the variables provided in formula, id and repeated.
id	a vector that identifies the clusters.
repeated	an optional vector that identifies the order of the observations within each cluster.
bstart	a vector that includes an initial estimate for the marginal regression parameter vector.
LORstr	a character string that indicates the marginalized 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 independently at each level pair of repeated ("2way").
LORterm	a matrix that satisfies the user-defined local odds ratios structure. It is ignored unless LORstr="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="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 function ipfp.
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 case level or equivalently in 'long' format. See details about the 'long' format in the function [reshape](#).

A term of the form `offset(expression)` is allowed in the right hand side of `formula`.

The default set for the response categories is $\{1, \dots, J\}$, where $J > 2$ is the maximum observed response category. If otherwise, the function recodes the observed response categories onto this set.

The J -th response category is treated as baseline.

The default set for the id labels is $\{1, \dots, N\}$, where N is the sample size. If otherwise, the function recodes the given labels onto this set.

The argument `repeated` can be ignored only when data is written in such a way that the t -th observation in each cluster is recorded at the t -th measurement occasion. If this is not the case, then the user must provide `repeated`. The suggested set for the levels of `repeated` is $\{1, \dots, T\}$, where T is the number of observed levels. If otherwise, the function recodes the given levels onto this set.

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

The fitted marginal baseline category logit model is

$$\log \frac{\Pr(Y_{it} = j | x_{it})}{\Pr(Y_{it} = J | x_{it})} = \beta_{j0} + \beta'_j x_{it}$$

where Y_{it} is the t -th multinomial response for cluster i , x_{it} is the associated covariates vector, β_{j0} is the j -th response category specific intercept and β_j is the j -th response category specific parameter vector.

The formula is easier to read from either the Vignette or the Reference Manual (both available [here](#)).

The `LORterm` argument must be an $L \times J^2$ matrix, where L is the number of level pairs of `repeated`. These are ordered as $(1, 2), (1, 3), \dots, (1, T), (2, 3), \dots, (T - 1, T)$ 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:

call	the matched call.
title	title for the GEE model.
version	the current version of the GEE solver.
link	the marginal link function.
local.odds.ratios	the marginalized local odds ratios structure variables.

terms	the terms structure describing the marginal model.
contrasts	the contrasts used for the factors.
nobs	the number of observations.
convergence	the values of the convergence variables.
coefficients	the estimated regression parameter vector of the marginal model.
linear.pred	the estimated linear predictor of the marginal regression model. The j -th column corresponds to the j -th response category.
fitted.values	the estimated fitted values of the marginal regression model. The j -th column corresponds 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.
max.id	the number of clusters.
clusz	the number of observations within each cluster.
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.
LORgee.control	the control values for the GEE solver.
ipfp.control	the control values for the function ipfp.
inverse.method	the method used for inverting matrices.
adding.constant	the value used for add.
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 $H_0 : \beta_1 = \dots = \beta_{J-1} = 0$ based on the Wald test statistic.

Author(s)

Anestis Touloumis

References

- Touloumis, A. (2011) *GEE for multinomial responses*. PhD dissertation, University of Florida.
- Touloumis, A., Agresti, A. and Kateri, M. (2013) GEE for multinomial responses using a local odds ratios parameterization. *Biometrics*, **69**, 633-640.
- Touloumis, A. (2015) R Package multgee: A Generalized Estimating Equations Solver for Multinomial Responses. *Journal of Statistical Software*, **64**, 1-14.

See Also

For an ordinal response scale use the function [ordLORgee](#).

Examples

```
## See the interpretation in Touloumis (2011).
data(housing)
fitmod <- nomLORgee(y~factor(time)*sec,data=housing,id=id, repeated=time)
summary(fitmod)
```

ordLORgee

Marginal Models For Correlated Ordinal Multinomial Responses

Description

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

Usage

```
ordLORgee(formula, data, id = id, repeated = NULL,
           link = "logit", 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	an optional data frame containing the variables provided in formula, id and repeated.
id	a vector that identifies the clusters.
repeated	an optional vector that identifies the order of observations within each cluster.
link	a character string that specifies the link function. Options include "logit", "probit", "cauchit", "cloglog" or "acl".
bstart	a vector that includes an initial estimate for the marginal regression parameter vector.
LORstr	a character string that indicates the marginalized 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 independently at each level pair of repeated ("2way").

LORterm	a matrix that satisfies the user-defined local odds ratios structure. It is ignored unless LORstr="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="time.exch" or "RC".
restricted	a logical that indicates monotone score parameters when LORstr="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 function ipfp.
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 case level or equivalently in 'long' format. See details about the 'long' format in the function [reshape](#).

A term of the form `offset(expression)` is allowed in the right hand side of `formula`.

The default set for the response categories is $\{1, \dots, J\}$, where $J > 2$ is the maximum observed response category. If otherwise, the function recodes the observed response categories onto this set.

The J -th response category is omitted.

The default set for the id labels is $\{1, \dots, N\}$, where N is the sample size. If otherwise, the function recodes the given labels onto this set.

The argument `repeated` can be ignored only when data is written in such a way that the t -th observation in each cluster is recorded at the t -th measurement occasion. If this is not the case, then the user must provide `repeated`. The suggested set for the levels of `repeated` is $\{1, \dots, T\}$, where T is the number of observed levels. If otherwise, the function recodes the given levels onto this set.

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

The fitted marginal cumulative link model is

$$Pr(Y_{it} \leq j | x_{it}) = F(\beta_{j0} + \beta' x_{it})$$

where Y_{it} is the t -th multinomial response for cluster i , x_{it} is the associated covariates vector, F is the cumulative distribution function determined by `link`, β_{j0} is the j -th response category specific intercept and β is the marginal regression parameter vector excluding intercepts.

The marginal adjacent categories logit model

$$\log \frac{Pr(Y_{it} = j | x_{it})}{Pr(Y_{it} = j + 1 | x_{it})} = \beta_{j0} + \beta' x_{it}$$

is fitted if and only if `link="acl"`. In contrast to a marginal cumulative link model, here the intercepts do not need to be monotone increasing.

The formulae are easier to read from either the Vignette or the Reference Manual (both available [here](#)).

The LORterm argument must be an $L \times J^2$ matrix, where L is the number of level pairs of repeated. These are ordered as $(1, 2), (1, 3), \dots, (1, T), (2, 3), \dots, (T - 1, T)$ 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:

call	the matched call.
title	title for the GEE model.
version	the current version of the GEE solver.
link	the marginal link function.
local.odds.ratios	the marginalized local odds ratios structure variables.
terms	the terms structure describing the model.
contrasts	the contrasts used for the factors.
nobs	the number of observations.
convergence	the values of the convergence variables.
coefficients	the estimated regression parameter vector of the marginal model.
linear.pred	the estimated linear predictor of the marginal regression model. The j -th column corresponds to the j -th response category.
fitted.values	the estimated fitted values of the marginal regression model. The j -th column corresponds to the j -th response category.
residuals	the residuals of the marginal regression model. The j -th column corresponds to the j -th response category.
y	the multinomial response variables.
id	the id variable.
max.id	the number of clusters.
clusz	the number of observations within each cluster.
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.
LORgee.control	the control values for the GEE solver.
ipfp.control	the control values for the function ipfp.
inverse.method	the method used for inverting matrices.
adding.constant	the value used for add.
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 $H_0 : \beta = 0$ based on the Wald test statistic.

Author(s)

Anestis Touloumis

References

Touloumis, A., Agresti, A. and Kateri, M. (2013) GEE for multinomial responses using a local odds ratios parameterization. *Biometrics*, **69**, 633-640.

Touloumis, A. (2015) R Package multgee: A Generalized Estimating Equations Solver for Multinomial Responses. *Journal of Statistical Software*, **64**, 1-14.

See Also

For a nominal response scale use the function [nomLORgee](#).

Examples

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

 waldts

Wald Test of Nested GEE Models

Description

Comparing two nested GEE models by carrying out a Wald test.

Usage

```
waldts(object0, object1)
```

Arguments

`object0` A GEE model of the class "LORgee".
`object1` A GEE model of the class "LORgee".

Details

The two GEE models implied by `object0` and `object1` must be nested.

Author(s)

Anestis Touloumis

Examples

```
data(housing)
set.seed(1)
fitmod1 <- nomLORgee(y~factor(time)*sec,data=housing,id=id,repeated=time)
set.seed(1)
fitmod0 <- update(fitmod1,formula=y~factor(time)+sec)
waldts(fitmod0,fitmod1)
```


Index

*Topic **datasets**

arthritis, [3](#)

housing, [4](#)

*Topic **package**

multgee-package, [2](#)

arthritis, [3](#)

coef, [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](#), [15](#)

ordLORgee, [2](#), [3](#), [5–8](#), [12](#), [12](#)

print, [11](#), [14](#)

reshape, [10](#), [13](#)

residuals, [11](#), [14](#)

summary, [11](#), [14](#)

waldts, [2](#), [15](#)