Package 'gee'

November 7, 2019

Title Generalized Estimation Equation Solver

Version 4.13-20

Depends stats

Suggests MASS

Date 2019-11-07

DateNote Gee version 1998-01-27

Author Vincent J Carey. Ported to R by Thomas Lumley (versions 3.13 and 4.4)
and Brian Ripley <ripley@stats.ox.ac.uk> (version 4.13).
Files src/dgedi.f and src/dgefa.f are for LINPACK authored by
Cleve Moler.
Note that maintainers are not available to give advice on using a
package they did not author.

Maintainer Brian Ripley <ripley@stats.ox.ac.uk>

Description Generalized Estimation Equation solver.

License GPL-2

NeedsCompilation yes

Repository CRAN

Date/Publication 2019-11-07 13:12:54 UTC

R topics documented:

---

gee   *Function to solve a Generalized Estimation Equation Model*

Description

Produces an object of class "gee" which is a Generalized Estimation Equation fit of the data.

Usage

```
gee(formula, id,
    data, subset, na.action,
    R = NULL, b = NULL,
    tol = 0.001, maxiter = 25,
    family = gaussian, corstr = "independence",
    Mv = 1, silent = TRUE, contrasts = NULL,
    scale.fix = FALSE, scale.value = 1, v4.4compat = FALSE)
```

Arguments

formula	a formula expression as for other regression models, of the form response ~ predictors. See the documentation of [lm](#) and [formula](#) for details.
id	a vector which identifies the clusters. The length of id should be the same as the number of observations. Data are assumed to be sorted so that observations on a cluster are contiguous rows for all entities in the formula.
data	an optional data frame in which to interpret the variables occurring in the formula, along with the id and n variables.
subset	expression saying which subset of the rows of the data should be used in the fit. This can be a logical vector (which is replicated to have length equal to the number of observations), or a numeric vector indicating which observation numbers are to be included, or a character vector of the row names to be included. All observations are included by default.
na.action	a function to filter missing data. For gee only na.omit should be used here.
R	a square matrix of dimension maximum cluster size containing the user specified correlation. This is only appropriate if corstr = "fixed".
b	an initial estimate for the parameters.
tol	the tolerance used in the fitting algorithm.
maxiter	the maximum number of iterations.
family	a family object: a list of functions and expressions for defining link and variance functions. Families supported in gee are gaussian, binomial, poisson, Gamma, and quasi; see the [glm](#) and [family](#) documentation. Some links are not currently available: 1/mu^2 and sqrt have not been hard-coded in the 'cgee' engine at present. The inverse gaussian variance function is not available. All combinations of remaining functions can be obtained either by family selection or by the use of quasi.

corstr	a character string specifying the correlation structure. The following are permitted: `"independence"`, `"fixed"`, `"stat_M_dep"`, `"non_stat_M_dep"`, `"exchangeable"`, `"AR-M"` and `"unstructured"`
Mv	When `corstr` is `"stat_M_dep"`, `"non_stat_M_dep"`, or `"AR-M"` then Mv must be specified.
silent	a logical variable controlling whether parameter estimates at each iteration are printed.
contrasts	a list giving contrasts for some or all of the factors appearing in the model formula. The elements of the list should have the same name as the variable and should be either a contrast matrix (specifically, any full-rank matrix with as many rows as there are levels in the factor), or else a function to compute such a matrix given the number of levels.
scale.fix	a logical variable; if true, the scale parameter is fixed at the value of `scale.value`.
scale.value	numeric variable giving the value to which the scale parameter should be fixed; used only if `scale.fix == TRUE`.
v4.4compat	logical variable requesting compatibility of correlation parameter estimates with previous versions; the current version revises to be more faithful to the Liang and Zeger (1986) proposals (compatible with the Groemping SAS macro, version 2.03)

Details

Though input data need not be sorted by the variable named `"id"`, the program will interpret physically contiguous records possessing the same value of id as members of the same cluster. Thus it is possible to use the following vector as an id vector to discriminate 4 clusters of size 4: `c(0,0,0,0,1,1,1,1,0,0,0,0,1,1,1,1)`.

Value

An object of class `"gee"` representing the fit.

Side Effects

Offsets must be specified in the model formula, as in [glm](glm).

Note

This is version 4.8 of this user documentation file, revised 98/01/27. The assistance of Dr B Ripley is gratefully acknowledged.

References

Liang, K.Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73** 13–22.

Zeger, S.L. and Liang, K.Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42** 121–130.

See Also

glm, lm, formula.

Examples

```
data(warpbreaks)
## marginal analysis of random effects model for wool
summary(gee(breaks ~ tension, id=wool, data=warpbreaks, corstr="exchangeable"))
## test for serial correlation in blocks
summary(gee(breaks ~ tension, id=wool, data=warpbreaks, corstr="AR-M", Mv=1))

if(require(MASS)) {
data(OME)
## not fully appropriate link for these data.
(fm <- gee(cbind(Correct, Trials-Correct) ~ Loud + Age + OME, id = ID,
          data = OME, family = binomial, corstr = "exchangeable"))
summary(fm)
}
```

Index