Package ‘wgeesel’

July 23, 2017

Type Package
Title Weighted Generalized Estimating Equations and Model Selection
Version 1.4
Date 2017-07-22
Encoding UTF-8
Author Cong Xu <congxu17@gmail.com>, Zheng Li <zheng.li@outlook.com>, Ming Wang <mwang@phs.psu.edu>
Maintainer Zheng Li <zheng.li@outlook.com>
Description Weighted generalized estimating equations (WGEE) is an extension of generalized linear models to longitudinal clustered data by incorporating the correlation within-cluster when data is missing at random (MAR). The parameters in mean, scale correlation structures are estimated based on quasi-likelihood. Multiple model selection criterion are provided for selection of mean model and working correlation structure based on WGEE/GEE.
License GPL-2
Depends R (>= 3.0.0)
Imports MASS, stats, geepack, bindata, PoisNor
NeedsCompilation no
Repository CRAN
Date/Publication 2017-07-23 08:34:55 UTC

R topics documented:

  wgeesel-package ........................................... 2
  data_sim ................................................... 3
  imps .......................................................... 4
  MLIC.gee .................................................... 5
  MQIC.gee ..................................................... 6
  ohio .......................................................... 8
  QIC.gee ...................................................... 8
  QICW.gee .................................................... 10
  RJ.gee ....................................................... 11
Description

Weighted Generalized estimating equations (WGEE) is an extension of generalized linear models to longitudinal or clustered data by incorporating the correlation within-cluster when data is missing at random (MAR). The parameters in mean, scale, correlation structures are estimate based on quasi-likelihood. The package \texttt{wgeesel} also contains model selection criteria for variable selection in the mean model and for the selection of a working correlation structure in longitudinal data with dropout or monotone missingness using WGEE.

Details

The collection of functions includes:

- \texttt{wgee} estimates parameters based on WGEE in mean, scale, and correlation structures, through mean link, scale link, and correlation link.
- \texttt{QIC.gee}, \texttt{MQIC.gee}, \texttt{RJ.gee} calculate the QIC (QIC$_{u}$), MQIC (MQIC$_{u}$), Rotnitzky-Jewell criteria for variable selection in the mean model and/or selection of a working correlation structure in GEE (unbalanced data is allowed).
- \texttt{MLIC.gee}, \texttt{QICW.gee} calculate the MLIC (MLICC) and QICW (QICW$_{u}$) for variable selection in the mean model and the selection of a working correlation structure in WGEE, which can accommodate dropout missing at random (MAR).
- \texttt{data_sim} can simulate longitudinal response data in different distribution (gaussian, binomial, poisson) with drop missingness.

For a complete list of functions, use \texttt{library(help = "wgeesel")}.

Author(s)

Cong Xu <congxu17@gmail.com>, Zheng Li <zheng.li@outlook.com>, Ming Wang <mwang@phs.psu.edu>

Maintainer: Zheng Li <zheng.li@outlook.com>
References


See Also

GEE methods exist for `geeglm` (`geepack`)

Examples

```r
data(imps)

fit <- wgee(Y ~ Drug+Sex+Time,data=imps,id=imps$id,family="binomial",
corstr="exchangeable",scale=NA,mismodel= R ~ Drug+Time)
```

---

**data_sim**

*Simulate longitudinal data*

Description

The function generate correlated normal, Bernoulli or Poisson longitudinal data.

Usage

```r
data_sim(id, rho, phi, x, beta, x_mis, para, corstr, family, lag_level)
```

Arguments

- **id**: subject id
- **rho**: within cluster correlation.
- **phi**: scale parameter in the variance covariance matrix.
- **x**: covariate associated with the response.
- **beta**: coefficients associated with x.
- **x_mis**: covariates associated with missing model.
Inpatient Multidimensional Psychiatric Scale (IMPS)

The `imps` data frame has 1544 rows and 8 columns. The data is from National Institute of the Mental Health Schizophrenia Collaborative Study, where the effect of chlorpromazine, fluphenazine, or thioridazine treatment on the overall severity of the schizophrenia disorder is of interest.

**Usage**

`data(imps)`
Format

This data frame contains the following columns:

- **ID**  patient ID
- **IMPS79**  the severity of the schizophrenia disorder (ranges from 0 to 7)
- **Week**  the fixed visit time
- **Drug**  the indicator of treatment (1: chlorpromazine, fluphenazine, or thioridazine treatment; 0: placebo)
- **Sex**  the indicator of sex of the patients (1: male; 0: female)
- **R**  an indicator of the missingness (1: observed; 0: missing)
- **Time**  square root of the Week covariate
- **Y**  an indicator if IMPS >= 4

Source


References


---

**MLIC.gee**  
*MLIC and MLICC for Weighted GEE*

Description

Calculate the MLIC (missing longitudinal information criterion) for selection of mean model, and the MLICC (missing longitudinal information correlation criterion) for selection of working correlation structure, based on the expected quadratic loss and the WGEE.

Usage

```r
MLIC.gee(object, object_full)
```

Arguments

- `object`  a fitted model object of class "wgee".
- `object_full`  a fitted model object of class "wgee": the largest candidate model under consideration to be fitted.

Value

Return a data frame of MLIC, MLICC and Wquad_loss.
Note
MLIC and MLICC model selection criterion for longitudinal data criterion with dropouts or monotone missingness under the assumption of MAR.

Author(s)
Cong Xu, Zheng Li and Ming Wang

References


See Also
wgee

Examples
data(imps)
fit1 <- wgee(Y ~ Drug+Sex+Time,data=imps,id=imps$ID,family="binomial", corstr="exchangeable",scale=NULL,mismodel= R ~ Drug+Time)

fit_f <- wgee(Y ~ Drug+Sex+Time*Sex+Time*Drug,data=imps,id=imps$ID, family="binomial", corstr="exchangeable",scale=NULL,mismodel= R ~ Drug+Time)

###not run####
###MLIC.ggee(fit1,fit_f)

---

MQIC.ggee

**MQIC and MQICu for GEE**

Description
Calculate MQIC (Modified QIC) and MQICu. MQIC is an asymptotic unbiased estimator of the risk function based on the independent quasi-likelihood. MQIC and original QIC may have non-negligible effect for model selection, especially when the true correlation structure completely different from the working correlation structure.

Usage
MQIC.ggee(object)
**Arguments**

object a fitted model object of class "wgee".

**Value**

Return a list of MQIC, MQIC_u and Quasi-likelihood.

**Author(s)**

Cong Xu, Zheng Li and Ming Wang

**References**


**See Also**

geeglm (geepack)

**Examples**

```r
data(imps)

fit <- wgee(IMPS79 ~ Drug+Sex+Time, data=imps, id=imps$ID,family="gaussian",
corstr="independence")

MQIC.gee(fit)

data(ohio)

fit2 <- wgee(resp ~ age + smoke+age:smoke, data=ohio, id=ohio$id,
family="binomial", corstr="independence")

MQIC.gee(fit2)
```
Ohio Children Wheeze Status

Description
The ohio data frame has 2148 rows and 4 columns. The dataset is a subset of the six-city study, a longitudinal study of the health effects of air pollution.

Usage
data(ohio)

Format
This data frame contains the following columns:
- **resp**: an indicator of wheeze status (1=yes, 0=no)
- **id**: a numeric vector for subject id
- **age**: a numeric vector of age, 0 to 9 years old
- **smoke**: an indicator of maternal smoking at the first year of the study (1=yes, 0=no)

References

See Also
- ohio (geepack)

QIC.gee

QIC and QICu for GEE

Description
Calculate quasi-likelihood under the independence model criterion (QIC) and QIC_u based on GEE.

Usage
QIC.gee(object)

Arguments
- **object**: a fitted model object of class "wgee".
**Value**

Return a vector of QIC, QIC\_u and Quasi-likelihood.

**Note**

QIC can be used to select the best correlation structure and the best fitting model in GEE analyses. The GEE is fitted by \texttt{geeglm (geepack)}. QIC\_u is a simplified version of QIC, which can not be applied to select the optimal working correlation structure. \texttt{geeglm (geepack)} only works for complete data. Thus if there are NA’s in data, the missing values are automatically removed by \texttt{na.omit}.

**Author(s)**

Cong Xu, Zheng Li and Ming Wang

**References**


**See Also**

\texttt{geeglm (geepack)}. \texttt{MuMIn} also provides QIC value.

**Examples**

```r
data(imps)
fit <- wgee(y ~ drug+sex+time, data=imps, id=imps$ID, family="binomial",
            corstr="exchangeable")
QIC.ggee(fit)

data(seizure)
# reshape the seizure data to "long" format
seiz.long <- reshape(seizure, varying=list(c("base","y1", "y2", "y3", "y4")),
                      v.names="y", times=8:4, direction="long")
seiz.long <- seiz.long[order(seiz.long$id, seiz.long$time),]
fit <- wgee(y ~ age + trt + time, data=seiz.long, id=seiz.long$id,
            family="poisson", corstr="independence")
QIC.ggee(fit)
```
QICW.gee

QICWr and QICWp for WGEE

Description

Calculate the QICW_r and QICW_p (an information criterion based on the weighted quasi-likelihood function) for selection of mean model and correlation structure based on the WGEE.

Usage

QICW.gee(object)

Arguments

object a fitted model object of class "wgee".

Value

Return a data frame of QICW_r, QICW_p and Wquasi_lik.

Note

QICW_r can be used for variable selection and for selecting the correlation structure in WGEE analyses. QICW_p is a simplified version of QICW_r, which can not be applied to select the optimal working correlation structure in WGEE.

Author(s)

Cong Xu, Zheng Li and Ming Wang

References


See Also

wgee
### RJ.gee

**RJC for GEE**

**Description**

Calculate RJC (Rotnitzky-Jewell information criterion) based on GEE.

**Usage**

```r
RJ.gee(object)
```

**Arguments**

- `object`: a fitted model object of class "wgee".

**Details**

Rotnitzky-Jewell information criterion (RJC) is usually used for working correlation structure selection.

**Value**

Return the value of the Rotnitzky-Jewell information criterion (RJ).

**Author(s)**

Cong Xu, Zheng Li and Ming Wang

**References**


**See Also**

- `geeglm` (geepack)
**Examples**

```r
data(ohio)

fit <- wgee(resp ~ age + smoke+age:smoke, data=ohio, id=ohio$id, 
            family="binomial", corstr="exchangeable")

RJ2.gee(fit)
```

---

**Description**

Calculate corrected RJC (Rotnitzky-Jewell information criterion) based on GEE with a modified robust variance estimator.

**Usage**

```r
RJ2.gee(object)
```

**Arguments**

- **object**
  
  a fitted model object of class "wgee".

**Value**

Return the value of the corrected Rotnitzky-Jewell information criterion (RJC).

**Note**

`RJ2.gee` can only handle balanced data (data with dropout missingness). Two assumptions should be satisfied. (A1) The conditional variance of $Y_{ij}$ given $X_{ij}$ is correctly specified; (A2) A common correlation structure, $R_c$, exists across all subjects. If there is missingness, one can group the subjects by the cluster size of the response variable. And, calculate the modified robust variance in each group to get the pooled estimate of the variance.

**Author(s)**

Cong Xu, Zheng Li and Ming Wang

**References**


See Also

ggee (geepack), RJ. gee

Examples

data(ohio)

fit <- wgee(resp ~ age + smoke+age:smoke, data=ohio, id=ohio$id, family="binomial",corstr="exchangeable")
RJ2.gee(fit)

seizure | Epiliptic Seizures
---------|------------------------

Description

The dataset has the number of epiliptic seizures in each of four two-week intervals, and in a baseline eight-week interval, for treatment and control groups with a total of 59 individuals.

Usage

data(seizure)

Format

This data frame contains the following columns:

- **y1**: the number of epiliptic seizures in the 1st 2-week interval
- **y2**: the number of epiliptic seizures in the 2nd 2-week interval
- **y3**: the number of epiliptic seizures in the 3rd 2-week interval
- **y4**: the number of epiliptic seizures in the 4th 2-week interval
- **trt**: an indicator of treatment
- **base**: the number of epilitic seizures in a baseline 8-week interval
- **age**: a numeric vector of subject age

Source


References


See Also

seizure (geepack)
wgee

Fit Weighted Generalized Estimating Equations (WGEE)

Description

wgee fits weighted generalized estimating equations (WGEE) with Newton Raphson algorithm. wgee has a syntax similar to glm and returns an object similar to a glm object.

Usage

wgee(model, data, id, family, corstr, scale = NULL, mismodel = NULL)

Arguments

model an object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted.
data a data frame containing the variables in the model.
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.
family a description of the error distribution and link function to be used in the model. This is a character string naming a family function. The following are permitted: "gaussian", "binomial", "poisson".
corstr a character string specifies the working correlation structure. The following are permitted: "independence", "exchangeable", "ar1".
scale a numeric variable giving the value to which the scale parameter should be fixed; if NA, the scale parameter is not fixed.
mismodel an object of class "formula" (or one that can be coerced to that class): a symbolic description of the missingness model to be fitted.

Details

wgee analyzes longitudinal data with missing values by weighted genralized estimating equations (WGEE), proposed by Robins, Totnizky and Zhao (1995). WGEE can handle missing at random problem. The standard error of the estimates are calculated as described in (Fitzmaurice, Laird, and Ware, 2011) and Preisser, Lohman, and Rathouz (2002).

Value

beta covariate effect estimates
var variance covariances estimates for beta
w_r_square weighted R square for continuous data
mu_fit fitted values of response
scale scale estimates
rho estimates
weight The weight of response y
model WGEE model structure
x covariates in WGEE
y response in WGEE
mis_fit estimates of the missingness model
call the function to be called
id as input
data as input
family as input
corstr as input

Author(s)
Zheng Li, Cong Xu and Ming Wang

References

See Also
geeglm (geepack)

Examples

### Example 1
data(imps)

```r
fit <- wgee(IMPS79 ~ Drug+Sex+Time, data=imps, id=imps$ID, family="gaussian",
corstr="exchangeable", scale=NULL, mismodel= R ~ Drug+Time)
```

### Example 2
data(seizure)

```r
# reshape the seizure data to "long" format
```
Wsquare.ggee

Weighted R Square for WGEE

Description

Calculate the weighted $R^2$ (missing longitudinal information criterion) for selection of mean model.

Usage

Wsquare.ggee(object, weight_mean)

Arguments

object | a fitted model object of class "wgee".
weight_mean | logical; whether weighted mean of the response should be used for weighted R square.

Value

Return a list of weighted R square.
ylag

Author(s)
Zheng Li, Cong Xu and Ming Wang

References

See Also
wgee

Examples
data(imps)

fit <- wgee(Y ~ Drug+Sex+Time, data=imps, id=imps$ID, family="binomial", corstr="exchangeable", scale=NULL, mismodel= R ~ Drug+Time)
WRSquare.gee(fit, weight_mean=TRUE)

ylag

Generate subject-level lagged response

Description
Suppose you have longitudinal response y and it’s subject id. This function generates lagged y for each subject.

Usage
ylag(id,y,lag,na=FALSE)

Arguments
id subject id
y response
lag how many lags for y
na logical; whether remain NAs in the response when lag > 1. Default to FALSE and output 0s instead of NAs.

Value
return a vector of lagged y

Author(s)
Cong Xu, Zheng Li and Ming Wang
Examples

```r
id <- rep(c(1:20), each=3)
y <- rnorm(length(id))
ylag(id, y, 1) # lag=1
ylag(id, y, 2, na=FALSE) # lag=2
```
Index

*Topic MAR
  wgee, 14
  WRsquare.ggee, 16
*Topic MLIC
  MLIC.ggee, 5
*Topic MQIC
  MQIC.ggee, 6
*Topic QICW
  QICW.ggee, 10
*Topic QIC
  QIC.ggee, 8
*Topic RJC
  RJ.ggee, 11
  RJ2.ggee, 12
*Topic datasets
  imps, 4
  ohio, 8
  seizure, 13
*Topic gee
  data_sim, 3
  MQIC.ggee, 6
  QIC.ggee, 8
  RJ.ggee, 11
  RJ2.ggee, 12
  wgeesel-package, 2
  ylag, 17
*Topic lag
  ylag, 17
*Topic package
  wgeesel-package, 2
*Topic qic
  wgeesel-package, 2
*Topic simulate
  data_sim, 3
*Topic wgee
  MLIC.ggee, 5
  QICW.ggee, 10
  wgee, 14
  wgeesel-package, 2
  WRsquare.ggee, 16
data_sim, 2, 3
geeglm, 3, 7, 9, 11, 13, 15
glm, 14
imps, 4
MLIC.ggee, 2, 5
MQIC.ggee, 2, 6
na.omit, 9
ohio, 8, 8
QIC.ggee, 2, 8
QICW.ggee, 2, 10
RJ.ggee, 2, 11, 13
RJ2.ggee, 12, 12
seizure, 13, 13
wgee, 2, 6, 10, 14, 14, 17
wgeesel (wgeesel-package), 2
wgeesel-package, 2
WRsquare.ggee, 16
ylag, 17