Package 'ELCIC'

February 13, 2023

Type Package

Title The Empirical Likelihood-Based Consistent Information Criterion

Version 0.2.1

Description We developed a consistent and robust information criterion to conduct model selection for semiparametric models. It is free of distribution specification and powerful to locate the true model given large sample size. This package provides several usage of ELCIC with applications in generalized linear model (GLM), generalized estimating equation (GEE) for longitudinal data, and weighted GEE (WGEE) for missing longitudinal data under the mechanism of missing at random and drop-out. Chixiang Chen, Ming Wang, Rongling Wu, Runze Li (2020) <doi:10.5705/ss.202020.0254>.

License Artistic-2.0

Encoding UTF-8

LazyData true

RoxygenNote 7.2.1

URL https://github.com/chencxxy28/ELCIC

BugReports https://github.com/chencxxy28/ELCIC/issues

Suggests knitr, rmarkdown, markdown, testthat (>= 3.0.0)

VignetteBuilder knitr

Depends R (>= 3.5.0)

Imports MASS, mvtnorm, PoisNor,bindata,geepack,wgeesel

Config/testthat/edition 3

NeedsCompilation no

Author Chixiang Chen [cre],
Biyi Shen [aut],
Ming Wang [aut]

Maintainer Chixiang Chen <chencxxy@hotmail.com>

Repository CRAN

Date/Publication 2023-02-13 16:40:09 UTC

R topics documented:

---

cond.prob	*Calculate conditional probabilities for observing records at each time point*

Description

A function calculates conditional probabilities for longitudinal missing data. The observing probability is at observation-level.

Usage

```
cond.prob(x_mis,gamma,id,time)
```

Arguments

x_mis	A matrix containing covariates for the missing data model. The first column should be all ones corresponding to the intercept.
gamma	coefficients calculated from missing data model
id	A vector indicating subject id.
time	The number of observations in total for each subject

Value

a vector containing conditional probabilities.

Examples

```
## tests
# load data
data(wgeesimdata)
library(wgeesel)
data_wgee<-data.frame(do.call(cbind,wgeesimdata))
corstr<-"exchangeable"
dist<-"binomial"
id<-data_wgee$id
# obtain the estimates
fit<-wgee(y~x1+x2+x3,data_wgee,id,family=dist,corstr =corstr,scale = NULL,
        mismodel =obs_ind~x_mis1)
beta<-as.vector(summary(fit)$beta)
rho<-summary(fit)$corr
phi<-summary(fit)$phi
#calculate observing probabilies for all observations
gamma<-as.vector(summary(fit$mis_fit)$coefficients[,1])
x_mis<-wgeesimdata$x_mis
pi<-cond.prob(x_mis,gamma,id,time=3)
```

---

ee.gee	*Estimating equation for GEE without missingness or with data missing completely at random.*

Description

Calculate estimating equation from GEE in ELCIC without missingness or missing completely at random. This estimating equation is used for joint selection of marginal mean and working correlation structure.

Usage

```
ee.gee(y,x,r,id,beta,rho,phi,dist,corstr)
```

Arguments

y	A vector containing outcomes.
x	A matrix containing covariates. The first column should be all ones the represents the intercept.
r	A vector indicating the observation of outcomes: 1 for observed records, and 0 for unobserved records. The default setup is that all data are observed. See more in details section.
id	A vector indicating subject id.
beta	A plug-in estimator solved by an external estimation procedure.
rho	A correlation coefficients obtained from an external estimation procedure, such as GEE.
phi	An over-dispersion parameter obtained from an external estimation procedure, such as GEE.
dist	A specified distribution. It can be "gaussian", "poisson",and "binomial".
corstr	A candidate correlation structure. It can be "independence","exchangeable", and "ar1".

Details

If the element in argument "r" equals zero, the corresponding rows of "x" and "y" should be all zeros.

Value

A matrix containing values of calculated estimating equations.

Examples

```
## tests
# load data
data(geesimdata)
x<-geesimdata$x
y<-geesimdata$y
id<-geesimdata$id
corstr<-"exchangeable"
dist<-"poisson"
# obtain the estimates
library(geepack)
# x matrix already include the intercept column.
fit<-geeglm(y~x-1,data=geesimdata,family =dist,id=id,corstr = "ar1")
beta<-fit$coefficients
rho<-unlist(summary(fit)$corr[1])
phi<-unlist(summary(fit)$dispersion[1])
r<-rep(1,nrow(x))
```

```
ee.matrix<-ee.gee(y,x,r,id,beta,rho,phi,dist,corstr)
apply(ee.matrix,1,mean)
```

---

ee.gee.mean	*Estimating equation of marginal mean for GEE without missingness or missing completely at random*

Description

Calculate estimating equation from GEE in ELCIC. This estimating equation is used for marginal mean selection.

Usage

```
ee.gee.mean(y,x,r,id,beta,rho,phi,dist,corstr)
```

Arguments

y	A vector containing outcomes.
x	A matrix containing covariates. The first column should be all ones corresponding to the intercept.
r	A vector indicating the observation of outcomes: 1 for observed records, and 0 for unobserved records. The default setup is that all data are observed. See more in details section.
id	A vector indicating subject id.
beta	A plug-in estimator solved by an external estimation procedure, such as GEE.
rho	A correlation coefficients obtained from an external estimation procedure, such as GEE.
phi	An over-dispersion parameter obtained from an external estimation procedure, such as GEE.
dist	A specified distribution. It can be "gaussian", "poisson",and "binomial".
corstr	A candidate correlation structure. It can be "independence","exchangeable", and "ar1".

Details

If the element in argument "r" equals zero, the corresponding rows of "x" and "y" should be all zeros.

Value

A matrix containing values of calculated estimating equations.

Note

corstr should be prespecified.

Examples

```
## tests
# load data
data(geesimdata)
x<-geesimdata$x
y<-geesimdata$y
id<-geesimdata$id
corstr<-"exchangeable"
dist<-"poisson"
# obtain the estimates
library(geepack)
fit<-geeglm(y~x-1,data=geesimdata,family =dist,id=id,corstr = corstr)
beta<-fit$coefficients
rho<-unlist(summary(fit)$corr[1])
phi<-unlist(summary(fit)$dispersion[1])
r<-rep(1,nrow(x))
ee.matrix<-ee.gee.mean(y,x,r,id,beta,rho,phi,dist,corstr)
apply(ee.matrix,1,mean)
```

---

ee.glm   *Estimating equation for ELCIC under GLM*

Description

A specified estimating equation for ELCIC under GLM. This estimating equation is used for marginal mean selection.

Usage

```
ee.glm(x, y, beta, dist)
```

Arguments

x	A matrix containing covariates. The first column should be all ones corresponding to the intercept. See more details in
y	A vector containing outcomes.
beta	A plug-in estimator solved by an external estimating procedure.
dist	A specified distribution. It can be "gaussian", "poisson",and "binomial".

Value

A matrix containing values of calculated estimating equations.

Note

"x" and "y" should be all observed.

Examples

```
## tests
# load data
data(glmsimdata)
x<-glmsimdata$x
y<-glmsimdata$y
# obtain the estimates. Note that x matrix already contains intercept.
fit<-glm(y~x-1,family="poisson")
beta<-fit$coefficients
ee.matrix<-ee.glm(x, y, beta, dist="poisson")
apply(ee.matrix,1,mean)
```

| ee.wgee | *Estimating equation for weighted GEE (WGEE) for missing longitudinal data under the mechanism of missing at random and drop-out* |

Description

Calculate estimating equation from WGEE for missing longitudinal data under the mechanism of missing at random and drop-out. This estimating equation is used for joint selection of marginal mean and "working" correlation structure.

Usage

```
ee.wgee(y,x,r,pi,id,time,beta,rho,phi,dist,corstr)
```

Arguments

y	A vector containing outcomes. use NA to indicate missing outcomes.
x	A matrix containing covariates. The first column should be all ones corresponding to the intercept.
r	A vector indicating the observation of outcomes: 1 for observed records, and 0 for unobserved records.
pi	A vector containing observing probabilities across all observations.
id	A vector indicating subject id.
time	The number of observations for each subject.
beta	A plug-in estimator solved by an external estimation procedure, such as WGEE.
rho	A correlation coefficients obtained from an external estimation procedure, such as WGEE.
phi	An over-dispersion parameter obtained from an external estimation procedure, such as GEE.

dist   A specified distribution. It can be "gaussian", "poisson",and "binomial".

corstr   A candidate correlation structure. It can be "independence","exchangeable", and
                        "ar1".

Value

A matrix containing values of calculated estimating equations.

Examples

```
## tests
# load data
data(wgeesimdata)
library(wgeesel)
data_wgee<-data.frame(do.call(cbind,wgeesimdata))
corstr<-"exchangeable"
dist<-"binomial"
id<-data_wgee$id
# obtain the estimates.
# Note that "obs_ind" is an indicator of observations in the missing data model.
fit<-wgee(y~x1+x2+x3,data_wgee,id,family=dist,corstr =corstr,
     scale = NULL,mismodel =obs_ind~x_mis1)
beta<-as.vector(summary(fit)$beta)
rho<-summary(fit)$corr
phi<-summary(fit)$phi
#calculate observing probabilies for all observations
gamma<-as.vector(summary(fit$mis_fit)$coefficients[,1])
x_mis<-wgeesimdata$x_mis
pi<-cond.prob(x_mis,gamma,id,time=3)
wgee.matrix<-ee.wgee(y=wgeesimdata$y,x=wgeesimdata$x,r=wgeesimdata$obs_ind,
pi=pi,id=wgeesimdata$id,time=3,beta=beta,rho=rho,phi=phi,dist=dist,corstr=corstr)
apply(wgee.matrix,1,mean)
```

---

ee.wgee.mean   *Estimating equation for marginal mean under WGEE for missing lon-*
                            gitudinal data under the mechanism of missing at random and drop-
                            out

Description

Calculate estimating function from WGEE. This estimating function is used for marginal mean
selection.

Usage

```
ee.wgee.mean(y,x,r,pi,id,time,beta,rho,phi,dist,corstr)
```

Arguments

y	A vector containing outcomes. use NA to indicate missing outcomes.
x	A matrix containing covariates. The first column should be all ones corresponding to the intercept.
r	A vector indicating the observation of outcomes: 1 for observed records, and 0 for unobserved records.
pi	A vector containing observing probabilities across all observations.
id	A vector indicating subject id.
time	The number of observations for each subject
beta	A plug-in estimator solved by an external estimation procedure, such as WGEE.
rho	A correlation coefficients obtained from an external estimation procedure, such as WGEE.
phi	An over-dispersion parameter obtained from an external estimation procedure, such as GEE.
dist	A specified distribution. It can be "gaussian", "poisson",and "binomial".
corstr	A candidate correlation structure. It can be "independence","exchangeable", and "ar1".

Value

A matrix containing values of calculated estimating equations.

Note

corstr should be prespecified.

Examples

```
## tests
# load data
data(wgeesimdata)
library(wgeesel)
data_wgee<-data.frame(do.call(cbind,wgeesimdata))
corstr<-"exchangeable"
dist<-"binomial"
id<-data_wgee$id
# obtain the estimates.
# Note that "obs_ind" is an indicator of observations in the missing data model.
fit<-wgee(y~x1+x2+x3,data_wgee,id,family=dist,corstr =corstr,
     scale = NULL,mismodel =obs_ind~x_mis1)
beta<-as.vector(summary(fit)$beta)
rho<-summary(fit)$corr
phi<-summary(fit)$phi
#calculate observing probabilies for all observations
gamma<-as.vector(summary(fit$mis_fit)$coefficients[,1])
x_mis<-wgeesimdata$x_mis
pi<-cond.prob(x_mis,gamma,id,time=3)
```

```
wgee.matrix<-ee.wgee.mean(y=wgeesimdata$y,x=wgeesimdata$x,r=wgeesimdata$obs_ind,
pi=pi,id=wgeesimdata$id,time=3,beta=beta,rho=rho,phi=phi,dist=dist,corstr=corstr)
apply(wgee.matrix,1,mean)
```

ELCIC.gee	*The whole procedure for joint selection of mean structure and corre-*
	lation structure in longitudinal data without missingness or missing
	completely at random

Description

The function ELCIC.gee provides the overall procedure for joint selection of mean structure and correlation structure in longitudinal data without missingness or missing completely at random.

Usage

```
ELCIC.gee(x,y,r,id,time,candidate.sets=NULL,name.var.sets=NULL,dist,
      candidate.cor.sets=c("independence","exchangeable", "ar1"), joints=TRUE)
```

Arguments

x	A matrix containing covariates. The first column should be all ones corresponding to the intercept if the intercept is expected in the marginal mean. Covariate matrix should be complete. NA values will be replaced by 0 if missingness is detected in x.
y	A vector containing outcomes. If y is missing completely at random, use NA to indicate missing outcomes and specify argument "r".
r	A vector indicating the observation of data: 1 for observed records (both outcome and covariates are observed for a given subject), and 0 for unobserved records. The default setup is that all data are observed.
id	A vector indicating subject id.
time	The number of observations in total for each subject
candidate.sets	A list containing index corresponding to candidate covariates. See more in details section.
name.var.sets	A list containing names of candidate covariates. The names should be subset of column names of x matrix. See more in details section.
dist	A specified distribution. It can be "gaussian", "poisson",and "binomial".
candidate.cor.sets	
	A vector containing candidate correlation structures. When joints=TRUE, it can be any subset of c("independence","exchangeable", "ar1"). The default is c("independence","exchangeable", "ar1"). When joints=FALSE, it should be either of "independence","exchangeable", "ar1". See more in details section.
joints	A logic value for joint selection of marginal mean and working correlation structure. The default is TRUE. See more in details section.

Details

Either arguments "candidate.sets" or "name.var.sets" is used to identify the set of candidate mean model. If both arguments are provided, only the argument "name.var.sets" will be used.

When joints=TRUE, the argument "candidate.cor.sets" can contain multiple correlation structures; however, when joints=FALSE, it should contain either of "independence","exchangeable", "ar1". If multiple correlation structures are provided, only the first one will be used.

Value

A matrix with each element containing ELCIC value for each candidate model.

Examples

```
## tests
# load data
data(geesimdata)
x<-geesimdata$x
y<-geesimdata$y
id<-geesimdata$id
r<-rep(1,nrow(x))
time<-3
candidate.sets<-list(c(1,2,3))
candidate.cor.sets<-c("exchangeable")
dist<-"poisson"
criterion.elcic<-ELCIC.gee(x=x,y=y,r=r,id=id,time=time,candidate.sets=candidate.sets,
                name.var.sets=NULL,dist=dist,candidate.cor.sets=candidate.cor.sets)
criterion.elcic
```

ELCIC.gee.single   *Calculate ELCIC value for a given candidate model under GEE framework with complete longitudinal data or data missing completely at random.*

Description

The function ELCIC.gee.single calculates ELCIC value for a given marginal mean candidate model with a specified working correlation structure. It is able to simultaneously evaluate mean model and working correlation structure.

Usage

```
ELCIC.gee.single(x, y, r, id, time, index.var=NULL, name.var = NULL,
                    dist, corstr, joints=TRUE)
```

Arguments

x	A matrix containing covariates. The first column should be all ones corresponding to the intercept. If y and x are missing completely at random, use NA to indicate missingness and specify argument "r".
y	A vector containing outcomes. If y and x are missing completely at random, use NA to indicate missing outcomes and specify argument "r".
r	A vector indicating the observation of data: 1 for observed records (both outcome and covariates are observed for a given subject), and 0 for unobserved records. The default setup is that all data are observed.
id	A vector indicating subject id.
time	The number of observations in total for each subject
index.var	A vector containing index corresponding to candidate covariates. See more in details section.
name.var	A vector containing names of candidate covariates. The names should be subset of column names of x matrix. See more in details section.
dist	A specified distribution. It can be "gaussian", "poisson",and "binomial".
corstr	A candidate correlation structure. It can be either of "independence","exchangeable", and "ar1".
joints	A logic value for joint selection of marginal mean and working correlation structure. The default is TRUE. See more in details section.

Details

Either arguments "index.var" or "name.var" is used to identify the candidate mean model. If both arguments are provided, only the argument "name.var" will be used.

When the argument "joints" is TRUE, `ELCIC.gee.single` will calculate ELCIC value based on the function `lambda.find.gee` and `ee.gee`, which involve estimating equations for both marginal mean and correlation coefficient. When the argument "joints" is FALSE, `ELCIC.gee.single` will calculate ELCIC value based on the function `lambda.find.gee.mean` and `ee.gee.mean`, which only involve estimating equations for marginal mean.

Value

A ELCIC value for a given candidate model.

Examples

```
## tests
# load data
data(geesimdata)
x<-geesimdata$x
y<-geesimdata$y
id<-geesimdata$id
r<-rep(1,nrow(x))
time<-3
corstr<-"exchangeable"
```

```
dist<-"poisson"
criteria<-ELCIC.gee.single(x=x,y=y,r=r,id=id,time=time,index.var=c(1,2,3),
           name.var=NULL,dist=dist,corstr=corstr)
criteria
```

ELCIC.glm   *The whole variable selection procedure for mean structure in GLM*

Description

The function ELCIC.glm provides the overall procedure for variable selection in GLM.

Usage

```
ELCIC.glm(x,y,candidate.sets,name.var.sets=NULL,dist)
```

Arguments

x	A matrix containing covariates. The first column should contain all ones corresponding to the intercept if the intercept is expected in the mean structure.
y	A vector containing outcomes.
candidate.sets	A list containing index corresponding to candidate covariates in each candidate model. See more in details section.
name.var.sets	A list containing names of candidate covariates corresponding to each candidate model. The names should be subset of column names of the x matrix. See more in details section.
dist	A specified distribution. It can be "gaussian", "poisson",and "binomial".

Details

"x" and "y" should be all observed. The corresponding individual data will be deleted if any missingness is detected.

Either arguments "candidate.sets" or "name.var.sets" is used to identify the set of candidate mean model. If both arguments are provided, only the argument "name.var.sets" will be used.

Value

A matrix with each element containing ELCIC value for each candidate model (in columns) and (in rows)

Examples

```
## tests
# load data
data(glmsimdata)
x<-glmsimdata$x
y<-glmsimdata$y
#candidate model index
candidate.sets<-list(c(1,2),c(1,2,3),c(1,2,3,4))
criteria<-ELCIC.glm(x, y, candidate.sets, name.var.sets = NULL, dist="poisson")
criteria
```

ELCIC.glm.single   *Variable selection in generalized linear models (GLM)*

Description

The function `ELCIC.glm.single` provides values of several model selection criteria including AIC, BIC, GIC, and ELCIC, given a candidate mean model.

Usage

```
ELCIC.glm.single(x, y, index.var=NULL, name.var = NULL, dist)
```

Arguments

x	A matrix containing covariates. The first column should contain all ones corresponding to the intercept if intercept is considered in your mean model.
y	A vector containing outcomes.
index.var	A vector containing index corresponding to candidate covariates (including the intercept). See more in details section.
name.var	A vector containing names of candidate covariates. The names should be subset of column names of x matrix. See more in details section.
dist	A specified distribution. It can be "gaussian", "poisson",and "binomial".

Details

"x" and "y" should be all observed. The corresponding individual data will be omitted in analysis if any missingness is detected.

Either arguments "index.var" or "name.var" is used to identify the candidate mean model. If both arguments are provided, only the argument "name.var" will be used.

Value

A vector containing information criteria including ELCIC, AIC, BIC, and GIC.

Examples

```
## tests
# load data
data(glmsimdata)
x<-glmsimdata$x
y<-glmsimdata$y
#candidate model index
name.var<-c("intercept","x1","x2")
index.var<-c(1,2,3)
criteria<-ELCIC.glm.single(x, y, index.var =index.var, name.var = NULL, dist="poisson")
criteria
```

ELCIC.wgee   *The whole procedure for joint selection of mean structure and correlation structure for missing longitudinal data under the mechanism of missing at random and drop-out*

Description

The function `ELCIC.wgee` provides the overall procedure for joint selection of mean structure and correlation structure in longitudinal data under missing at random. It is also able to implement marginal mean structure selection given a prespecified working correlation structure. The data is dropout missing and missing at random.

Usage

```
ELCIC.wgee(x,y,x_mis,r,id,time,candidate.sets=NULL,name.var.sets=NULL,
    dist,candidate.cor.sets=c("independence","exchangeable", "ar1"), joints=TRUE,lag=1)
```

Arguments

x	A matrix containing covariates. The first column should be all ones corresponding to the intercept if the intercept is considered in the marginal mean. Covariate matrix should be complete.
y	A vector containing outcomes. Use NA to indicate missing outcomes.
x_mis	A matrix containing covariates for the missing data model. The first column should be all ones corresponding to the intercept. This covariate matrix should be complete and all observed. See more in details section.
r	A vector indicating the observation of outcomes: 1 for observed records, and 0 for unobserved records.
id	A vector indicating subject id.
time	The number of observations in total for each subject
candidate.sets	A list containing index corresponding to candidate covariates. See more in details section.

name.var.sets   A list containing names of candidate covariates. The names should be subset of
                column names of x matrix. See more in details section.

dist   A specified distribution. It can be "gaussian", "poisson",and "binomial".

candidate.cor.sets

                A vector containing candidate correlation structures. When joints=TRUE, it
                can be any subset of c("independence","exchangeable", "ar1"). The default is
                c("independence","exchangeable", "ar1"). When joints=FALSE, it should be
                either of "independence","exchangeable", "ar1". See more in details section.

joints   A logic value for joint selection of marginal mean and working correlation struc-
                ture. The default is TRUE. See more in details section.

lag   A numeric value indicating lag-response involved in the missing data model. It
                can be either 1 or 2. The default is 1.

Details

Covariate matrix "x" should be complete. If missing data are present in "x", the elements in covari-
ate vector will be replaced by zeros for individuals who have missing covariates.

The argument "x_mis" includes all covariates to fit the missing data model. It does not contains a
lag variable based on the outcome y. The argument "lag" in this function will automatically add
lag-response variables to indicate the data missing at random.

Either arguments "candidate.sets" or "name.var.sets" is used to identify the set of candidate mean
model. If both arguments are provided, only the argument "name.var.sets" will be used.

When joints=TRUE, the argument "candidate.cor.sets" can contain multiple correlation structures;
however, when joints=FALSE, it should contain either of "independence","exchangeable", "ar1". If
multiple correlation structures are provided, only the first one will be used.

Value

A matrix with each element containing ELCIC value for each candidate model.

Examples

```
## tests
# load data
data(wgeesimdata)
dist<-"binomial"
x<-wgeesimdata$x
y<-wgeesimdata$y
x_mis<-wgeesimdata$x_mis
r<-wgeesimdata$obs_ind
id<-wgeesimdata$id
time<-3
candidate.sets<-list(c(1,2,3))
candidate.cor.sets<-c("exchangeable")
criterion.elcic<-ELCIC.wgee(x,y,x_mis,r,id,time,candidate.sets,name.var.sets=NULL,
                                dist,candidate.cor.sets,joints=TRUE)
criterion.elcic
```

ELCIC.wgee.single	*Calculate ELCIC value for a given candidate model under WGEE framework for missing longitudinal data under the mechanism of missing at random and drop-out*

Description

The function `ELCIC.wgee.single` to calculate ELCIC value for a given candidate mean model with specified working correlation structure. It is able to simultaneously evaluate mean model and working correlation structure. The data is dropout missing and missing at random.

Usage

```
ELCIC.wgee.single(x,y,x_mis,r,id,time,index.var=NULL,
     name.var=NULL,dist,corstr,joints=TRUE,lag=1)
```

Arguments

x	A matrix containing covariates. The first column should be all ones corresponding to the intercept if the intercept is considered in the marginal mean. Covariate matrix should be complete. See more in details section.
y	A vector containing outcomes. Use NA to indicate missing outcomes.
x_mis	A matrix containing covariates for the missing data model. The first column should be all ones corresponding to the intercept. This covariate matrix should be complete and all observed. See more in details section.
r	A vector indicating the observation of outcomes: 1 for observed records, and 0 for unobserved records.
id	A vector indicating subject id.
time	The number of observations in total for each subject
index.var	A vector containing index corresponding to candidate covariates. See more in details section.
name.var	A vector containing names of candidate covariates. The names should be subset of column names of x matrix. See more in details section.
dist	A specified distribution. It can be "gaussian", "poisson",and "binomial".
corstr	A candidate correlation structure. It can be "independence","exchangeable", and "ar1".
joints	A logic value for joint selection of marginal mean and working correlation structure. The default is TRUE. See more in details section.
lag	A numeric value indicating lag-response involved in the missing data model. It can be one of 0, 1, and 2. The default is 1.

Details

Covariate matrix "x" should be complete. If missing data are present in "x", the elements in covariate vector will be replaced by zeros for individuals who have missing covariates.

The argument "x_mis" includes all covariates to fit the missing data model. It does not contains a lag variable based on the outcome y. The argument "lag" in this function will automatically add lag-response variables to indicate the data missing at random.

Either arguments "index.var" or "name.var" is used to identify the candidate mean model. If both arguments are provided, only the argument "name.var" will be used.

When the argument "joints" is TRUE, `ELCIC.wgee.single` will calculate ELCIC value based on the function `lambda.find.wgee` and `ee.wgee`, which involve estimating equations for both marginal mean and correlation coefficient. When the argument "joints" is FALSE, `ELCIC.wgee.single` will calculate ELCIC value based on the function `lambda.find.wgee.mean` and `ee.wgee.mean`, which only involve estimating equations for marginal mean.

Value

A matrix containing values of calculated estimating equations.

Examples

```
## tests
# load data
data(wgeesimdata)
corstr<-"exchangeable"
dist<-"binomial"
x<-wgeesimdata$x
y<-wgeesimdata$y
x_mis<-wgeesimdata$x_mis
r<-wgeesimdata$obs_ind
id<-wgeesimdata$id
time<-3
index.var<-c(1,2,3)
ELCIC_value<-ELCIC.wgee.single(x,y,x_mis,r,id,time,index.var,name.var=NULL,
                    dist,corstr,joints=TRUE)
ELCIC_value
```

ELCICgee	*Model selection based on ELCIC under the syntax of GEE (Main function).*

Description

The function `ELCICgee` provides the model selection under the syntax of the geepack package.

Usage

```
ELCICgee(models, candidate.cor.sets,data, family,r,id,time)
```

Arguments

`models`	A list of formulas. See the corresponding documentation to geeglm.
`candidate.cor.sets`	
	A vector containing candidate correlation structures. It can be any subset of c("independence","exchangeable", "ar1").
`data`	A data frame containing the variables in the model.
`family`	A description of the error distribution and link function to be used in the model. The details are given under "Details".
`r`	A vector indicating the observation of data: 1 for observed records (both outcome and covariates are observed for a given subject), and 0 for unobserved records. The default setup is that all data are observed.
`id`	A vector indicating subject id.
`time`	The number of observations in total for each subject.

Details

Three commonly used distributions are considered: "gaussian", "poisson", "binomial". For the current package, the identity link is considered for a "gaussian" distribution; the log link is considered for a "poisson" distribution; the logit link is considered for a "binomial" distribution;

Value

A list with two items: model selection result based on ELCIC; An object of "geeglm" based on the selected model.

Examples

```
## tests
# load data
data(geesimdata)
id<-geesimdata$id
r<-rep(1,length(id))
time<-3
dat <- data.frame(y=geesimdata$y, geesimdata$x,id=id)
models <- list(y~x1+x2)
candidate.cor.sets<-c("exchangeable")
family<-poisson()
output<-ELCICgee(models, candidate.cor.sets,data=dat,family,r,id,time)
output$model.selection
output$gee.output
```

ELCICglm   *Variable selection based on ELCIC under the syntax of GLM (Main function).*

Description

The function `ELCICglm` provides the variable selection under the syntax of the GLM package.

Usage

```
ELCICglm(models, data, family)
```

Arguments

models	A list of formulas. See the corresponding documentation to glm.
data	A data frame containing the variables in the model.
family	A description of the error distribution and link function to be used in the model. The details are given under "Details".

Details

Three commonly used distributions are considered: "gaussian", "poisson", "binomial". For the current package, the identity link is considered for a "gaussian" distribution; the log link is considered for a "poisson" distribution; the logit link is considered for a "binomial" distribution;

Value

A list with two items: model selection results based on ELCIC, AIC, BIC, and GIC; An object of "glm" based on the selected model.

Examples

```
## tests
# load data
data(glmsimdata)
dat <- data.frame(y=glmsimdata$y, glmsimdata$x) ####x is a covariate matrix.
models <- list(y~x1, y~x1+x2, y~x1+x2+x3)
output<-ELCICglm(models, dat, poisson())
output$model.selection
output$glm.output
```

| ELCICwgee | *Model selection based on ELCIC under the syntax of WGEE (Main function).* |

Description

The function `ELCICwgee` provides the model selection under the syntax of the wgeesel package.

Usage

```
ELCICwgee(models, candidate.cor.sets, data, model_mis, family,r,id,time)
```

Arguments

models	A list of formulas. See corresponding documentation to wgeesel.
candidate.cor.sets	
	A vector containing candidate correlation structures. It can be any subset of c("independence","exchangeable", "ar1").
data	A data frame containing the variables in both the main model and the missing model.
model_mis	A formula used in the missing data model.
family	A description of the error distribution and link function to be used in the model. The details are given under "Details".
r	A vector indicating the observation of data: 1 for observed records (both outcome and covariates are observed for a given subject), and 0 for unobserved records. The default setup is that all data are observed.
id	A vector indicating subject id.
time	The number of observations in total for each subject.

Details

Three commonly used distributions are considered: "gaussian", "poisson", "binomial". For the current package, the identity link is considered for a "gaussian" distribution; the log link is considered for a "poisson" distribution; the logit link is considered for a "binomial" distribution;n;

Value

A list with two items: model selection result based on ELCIC; An object of "wgee" based on the selected model.

Examples

```
## tests
# load data
data(wgeesimdata)
family<-binomial()
r<-wgeesimdata$obs_ind
id<-wgeesimdata$id
time<-3
dat <- data.frame(y=wgeesimdata$y, wgeesimdata$x,x_mis1=wgeesimdata$x_mis[,2])
models <- list(y~x1+x2)
model_mis<-r~x_mis1
candidate.cor.sets<-c("exchangeable")
output<-ELCICwgee(models, candidate.cor.sets,data=dat,model_mis,family,r,id,time)
output$model.selection
output$wgee.output
```

---

gee.generator   *Generate longitudinal data without missingness*

Description

A function for generating longitudinal data without missingness. All covariates (except for intercept) are normally distributed.

Usage

```
gee.generator(beta,samplesize,time,num.time.dep,num.time.indep,
      rho,x.rho,dist,cor.str,x.cor.str)
```

Arguments

beta	A vector containing underlying true coefficients for each covariate in the model (including the intercept).
samplesize	The sample size.
time	The number of observations per subject.
num.time.dep	The number of time-dependent covariates.
num.time.indep	The number of time-independent covariates (not include intercept).
rho	The correlation coefficient for residuals across time.
x.rho	The correlation coefficient for time-dependent covariates across time.
dist	A specified distribution. It can be "gaussian", "poisson",and "binomial".
cor.str	The correlation structure for residuals across time. It can be "independence","exchangeable", and "ar1".
x.cor.str	The correlation structure for time-dependent covariates across time. It can be "independence","exchangeable", and "ar1".

Value

x: a matrix containing continuous covariates. The first column should contain all ones corresponding to the intercept.

y: a vector containing outcomes.

id: a vector indicating subject id.

Examples

```
beta<-c(-1,1,0.5,0)
samplesize<-100
geesimdata<-gee.generator(beta=beta,samplesize=samplesize,time=3,num.time.dep=2,
num.time.indep=1,rho=0.4,x.rho=0.2,dist="poisson",cor.str="exchangeable",
x.cor.str="exchangeable")
geesimdata$y
```

---

geesimdata	*Data simulated for model selection under GEE framework without missingness*

Description

Data simulated for model selection under GEE framework without missingness

Usage

```
data(geesimdata)
```

Format

An object of class `"list"`

y The outcomes generated from Poisson distribution with three repeated measurements from each subject

x A covariate matrix, of which the first column are all ones and rest columns contain normally distributed. Two are time-dependent variables, and one is time-independent variable.

References

This data set was artificially created for the ELCIC package.

Examples

```
data(geesimdata)
geesimdata$y
```

---

glm.generator   *Cross-sectional data generation under GLM*

Description

A function provides simulated outcomes as well as covariates under the framework of GLM. All covariates (except for intercept) are normally distributed.

Usage

```
glm.generator(beta, samplesize, rho = 0, dist, sd.gaussian = NULL, ov = NULL)
```

Arguments

beta	The underlying true coefficient for each covariates in the model (including the intercept).
samplesize	The sample size.
rho	The correlation coefficient among covariates.
dist	A specified distribution. It can be "gaussian", "poisson",and "binomial".
sd.gaussian	The standard deviation for the outcome from Gaussian distribution.
ov	The dispersion parameter for the outcome from Negative Binomial distribution.

Value

x: a matrix containing continuous covariates. The first column should contain all ones corresponding to the intercept.

y: a vector containing outcomes.

Examples

```
beta<-c(0.5,0.5,0.5,0)
samplesize<-100
data<-glm.generator(beta=beta,samplesize=samplesize,rho=0.5,dist="poisson")
```

---

glmsimdata *Data simulated for variable selection under GLM framework*

Description

Data simulated for variable selection under GLM framework

Usage

```
data(glmsimdata)
```

Format

An object of class "list"

y The outcome generated from Negative Binomial distribution with the dispersion parameter parameter=2

x A covariate matrix, of which the first column are all ones and rest columns contain normally distributed values

References

This data set was artificially created for the ELCIC package.

Examples

```
data(glmsimdata)
glmsimdata$y
```

---

impsdata *Inpatient Multidimensional Psychiatric Scale (IMPS)*

Description

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(impsdata)
```

Format

An object of class `"list"`

y The binary outcomes indicating whether IMPS >=4, which is longitudinal dropout and missing at random. IMPS describes severity of the schizophrenia disorder (ranges from 0 to 7)

x A full covariate matrix. It contains intercept, sex (1:male,0:female), drug (1: chlorphromazine, fluphenazine, or thioridazine treatment; 0: placebo), time: square root of the week covariate, and their two-way interactions.

x_mis A covariate matrix for missing data model. It contains intercept, drug, time, and sex.

id Patient ID

r An indicator of the missingness (1: observed; 0: missing).

References

add here

Examples

```
data(impsdata)
id<-impsdata$id
r<-impsdata$r
data.real <- data.frame(id=id,y=impsdata$y, impsdata$x)
head(data.real,n=10)
# each participant has three records
time<-4
# the outcome is binary
family=binomial()
models <- list(y~Time, y~Drug,y~Time+Drug,
               y~Time*Drug,y~Time+Sex+Drug,
               y~Time+Sex+Drug+Time:Sex+Sex:Drug+Drug:Time)

model_mis<-r~Drug+Time+Sex

candidate.cor.sets<-c("exchangeable","independence","ar1")
#not run
#output_ELCIC<-ELCICwgee(models, candidate.cor.sets,data=data.real,model_mis,family,r,id,time)
#output_MLIC<-MLICwgee(models, candidate.cor.sets,data=data.real,model_mis,family,r,id,time)
#output_QICW<-QICWwgee(models, candidate.cor.sets,data=data.real,model_mis,family,r,id,time)
```

---

lambda.find.gee   *Calculate the tuning parameters involved in ELCIC under GEE*

Description

This function provides an efficient algorithm to calculate the tuning parameters involved in ELCIC under GEE.

Usage

```
lambda.find.gee(x, y, id, beta, r, dist, rho, phi, corstr)
```

Arguments

x	A matrix containing covariates. The first column should be all ones corresponding to the intercept.
y	A vector containing outcomes.
id	A vector indicating subject id.
beta	A plug-in estimator solved by an external estimation procedure, such as GEE.
r	A vector indicating the observation of outcomes: 1 for observed records, and 0 for unobserved records. The default setup is that all data are observed. See more in details section.
dist	A specified distribution. It can be "gaussian", "poisson",and "binomial".
rho	A correlation coefficients obtained from an external estimation procedure, such as GEE.
phi	An over-dispersion parameter obtained from an external estimation procedure, such as GEE.
corstr	A candidate correlation structure. It can be "independence","exchangeable", and "ar1".

Details

If the element in argument "r" equals zero, the corresponding rows of "x" and "y" should be all zeros.

Value

Tuning parameter values.

Examples

```
## tests
# load data
data(geesimdata)
x<-geesimdata$x
y<-geesimdata$y
id<-geesimdata$id
corstr<-"exchangeable"
dist<-"poisson"
# obtain the estimates
library(geepack)
fit<-geeglm(y~x-1,data=geesimdata,family =dist,id=id,corstr = corstr)
beta<-fit$coefficients
rho<-unlist(summary(fit)$corr[1])
phi<-unlist(summary(fit)$dispersion[1])
r=rep(1,nrow(x))
lambda<-lambda.find.gee(x,y,id,beta,r,dist,rho,phi,corstr)
```

    lambda

---

lambda.find.gee.mean   *Calculate the tuning parameters under marginal mean selection in*
                        GEE

Description

This function provides an efficient algorithm to calculate the tuning parameters involved in marginal mean selection in GEE.

Usage

```
lambda.find.gee.mean(x, y, id, beta, r, dist, rho, phi, corstr)
```

Arguments

x	A matrix containing covariates. The first column should be all ones corresponding to the intercept.
y	A vector containing outcomes.
id	A vector indicating subject id.
beta	A plug-in estimator solved by an external estimation procedure, such as GEE.
r	A vector indicating the observation of outcomes: 1 for observed records, and 0 for unobserved records. The default setup is that all data are observed. See more in details section.
dist	A specified distribution. It can be "gaussian", "poisson",and "binomial".
rho	A correlation coefficients obtained from an external estimation procedure, such as GEE.
phi	An over-dispersion parameter obtained from an external estimation procedure, such as GEE.
corstr	A candidate correlation structure. It can be "independence","exchangeable", and "ar1".

Details

If the element in argument "r" equals zero, the corresponding rows of "x" and "y" should be all zeros.

Value

Tuning parameter values.

Note

corstr should be prespecified.

Examples

```
## tests
# load data
data(geesimdata)
x<-geesimdata$x
y<-geesimdata$y
id<-geesimdata$id
corstr<-"exchangeable"
dist<-"poisson"
# obtain the estimates
library(geepack)
fit<-geeglm(y~x-1,data=geesimdata,family =dist,id=id,corstr = corstr)
beta<-fit$coefficients
rho<-unlist(summary(fit)$corr[1])
phi<-unlist(summary(fit)$dispersion[1])
r<-rep(1,nrow(x))
lambda<-lambda.find.gee.mean(x,y,id,beta,r,dist,rho,phi,corstr)
lambda
```

---

lambda.find.glm *To calculate tuning parameter involved in ELCIC under GLM*

Description

This function aims to efficiently calculate the tuning parameter lambda in ELCIC.

Usage

```
lambda.find.glm(x, y, beta, dist)
```

Arguments

x	A matrix containing covariates. The first column should contain all ones corresponding to the intercept.
y	A vector containing outcomes.
beta	A plug-in estimator solved by an external estimating procedure.
dist	A specified distribution. It can be "gaussian", "poisson",and "binomial".

Value

A value of lambda (tuning parameter) vector involved in the empirical likelihood.

Note

All "x" and "y" should be observed.

Examples

```
## tests
# load data
data(glmsimdata)
x<-glmsimdata$x
y<-glmsimdata$y
# obtain the estimates
fit<-glm(y~x-1,family="poisson")
beta<-fit$coefficients
lambda<-lambda.find.glm(x, y, beta, dist="poisson")
lambda
```

---

lambda.find.wgee   *Calculate the tuning parameters involved in ELCIC under WGEE with*
                         data missing at random

Description

This function provides an efficient algorithm to calculate the tuning parameters involved in ELCIC under WGEE with data missing at random.

Usage

```
lambda.find.wgee(y,x,r,pi,id,time,beta,rho,phi,dist,corstr)
```

Arguments

y	A vector containing outcomes. use NA to indicate missing outcomes.
x	A matrix containing covariates. The first column should be all ones corresponding to the intercept.
r	A vector indicating the observation of outcomes: 1 for observed records, and 0 for unobserved records.
pi	A vector containing observing probabilities across all observations.
id	A vector indicating subject id.
time	The number of observations for each subject
beta	A plug-in estimator solved by an external estimation procedure, such as WGEE.
rho	A correlation coefficients obtained from an external estimation procedure, such as WGEE.
phi	An over-dispersion parameter obtained from an external estimation procedure, such as GEE.
dist	A specified distribution. It can be "gaussian", "poisson",and "binomial".
corstr	A candidate correlation structure. It can be "independence","exchangeable", and "ar1".

Value

Tuning parameter values.

Examples

```
## tests
# load data
data(wgeesimdata)
library(wgeesel)
data_wgee<-data.frame(do.call(cbind,wgeesimdata))
corstr<-"exchangeable"
dist<-"binomial"
id<-data_wgee$id
# obtain the estimates
fit<-wgee(y~x1+x2+x3,data_wgee,id,family=dist,corstr =corstr,scale = NULL,
        mismodel =obs_ind~x_mis1)
beta<-as.vector(summary(fit)$beta)
rho<-summary(fit)$corr
phi<-summary(fit)$phi
#calculate observing probabilies for all observations
gamma<-as.vector(summary(fit$mis_fit)$coefficients[,1])
x_mis<-wgeesimdata$x_mis
pi<-cond.prob(x_mis,gamma,id,time=3)
lambda<-lambda.find.wgee(y=wgeesimdata$y,x=wgeesimdata$x,r=wgeesimdata$obs_ind,
pi=pi,id=wgeesimdata$id,time=3,beta=beta,rho=rho,phi=phi,dist=dist,corstr=corstr)
lambda
```

---

lambda.find.wgee.mean	*Calculate the tuning parameters involved in marginal mean selection under WGEE with data missing at random*

Description

This function provides an efficient algorithm to calculate the tuning parameters involved in marginal mean selection under WGEE with data missing at random.

Usage

```
lambda.find.wgee.mean(y,x,r,pi,id,time,beta,rho,phi,dist,corstr)
```

Arguments

y	A vector containing outcomes. use NA to indicate missing outcomes.
x	A matrix containing covariates. The first column should be all ones corresponding to the intercept.
r	A vector indicating the observation of outcomes: 1 for observed records, and 0 for unobserved records.

pi	A vector containing observing probabilities across all observations.
id	A vector indicating subject id.
time	The number of observations for each subject
beta	A plug-in estimator solved by an external estimation procedure, such as WGEE.
rho	A correlation coefficients obtained from an external estimation procedure, such as WGEE.
phi	An over-dispersion parameter obtained from an external estimation procedure, such as GEE.
dist	A specified distribution. It can be "gaussian", "poisson",and "binomial".
corstr	A candidate correlation structure. It can be "independence","exchangeable", and "ar1".

Value

Tuning parameter values.

Note

corstr should be prespecified.

Examples

```
## tests
# load data
data(wgeesimdata)
library(wgeesel)
data_wgee<-data.frame(do.call(cbind,wgeesimdata))
corstr<-"exchangeable"
dist<-"binomial"
id<-data_wgee$id
# obtain the estimates
fit<-wgee(y~x1+x2+x3,data_wgee,id,family=dist,corstr =corstr,scale = NULL,
        mismodel =obs_ind~x_mis1)
beta<-as.vector(summary(fit)$beta)
rho<-summary(fit)$corr
phi<-summary(fit)$phi
#calculate observing probabilies for all observations
gamma<-as.vector(summary(fit$mis_fit)$coefficients[,1])
x_mis<-wgeesimdata$x_mis
pi<-cond.prob(x_mis,gamma,id,time=3)
lambda<-lambda.find.wgee.mean(y=wgeesimdata$y,x=wgeesimdata$x,r=wgeesimdata$obs_ind,
pi=pi,id=wgeesimdata$id,time=3,beta=beta,rho=rho,phi=phi,dist=dist,corstr=corstr)
lambda
```

---

marg.prob	*Calculate the inverse of marginal probability for observing records at each time point*

Description

A function calculates the inverse of joint probabilities for weight calculation involved in WGEE.

Usage

```
marg.prob(pi)
```

Arguments

pi   A matrix containing covariates for the missing data model. The first column should be all ones corresponding to the intercept.

Value

a vector containing the inverse of joint probabilities.

Examples

```
## tests
# load data
data(wgeesimdata)
library(wgeesel)
data_wgee<-data.frame(do.call(cbind,wgeesimdata))
corstr<-"exchangeable"
dist<-"binomial"
id<-data_wgee$id
# obtain the estimates
fit<-wgee(y~x1+x2+x3,data_wgee,id,family=dist,corstr =corstr,scale = NULL,
        mismodel =obs_ind~x_mis1)
beta<-as.vector(summary(fit)$beta)
rho<-summary(fit)$corr
phi<-summary(fit)$phi
#calculate observing probabilies for all observations
gamma<-as.vector(summary(fit$mis_fit)$coefficients[,1])
x_mis<-wgeesimdata$x_mis
pi<-cond.prob(x_mis,gamma,id,time=3)
joint_prob<-marg.prob(pi)
```

MLIC.wgee	*The whole MLIC procedure for joint selection of mean structure and correlation structure for missing longitudinal data under the mechanism of missing at random and drop-out*

Description

This function provides the overall MLIC procedure for joint selection of mean structure and correlation structure in longitudinal data missing at random. It is also able to implement marginal mean structure selection given a prespecified working correlation structure. The data is dropout missing and missing at random.

Usage

```
MLIC.wgee(x,y,x_mis,r,id,time,candidate.sets=NULL, name.var.sets=NULL,dist,
    candidate.cor.sets=c("independence","exchangeable", "ar1"), joints=TRUE,lag=1)
```

Arguments

x	A matrix containing covariates. The first column should be all ones corresponding to the intercept if the intercept is expected in the marginal mean model. Covariate matrix should be complete.
y	A vector containing outcomes. Use NA to indicate missing outcomes.
x_mis	A matrix containing covariates for the missing data model. The first column should be all ones corresponding to the intercept. This covariate matrix should be complete and all observed. See more in details section.
r	A vector indicating the observation of outcomes: 1 for observed records, and 0 for unobserved records.
id	A vector indicating subject id.
time	The number of observations in total for each subject.
candidate.sets	A list containing index corresponding to candidate covariates. See more in details section.
name.var.sets	A list containing names of candidate covariates. The names should be subset of column names of x matrix. See more in details section.
dist	A specified distribution. It can be "gaussian", "poisson",and "binomial".
candidate.cor.sets	
	A vector containing candidate correlation structures. When joints=TRUE, it is c("independence","exchangeable", "ar1") as default. When joints=FALSE, it should be either of "independence","exchangeable", "ar1". See more in details section.
joints	A logic value for joint selection of marginal mean and working correlation structure. The default is TRUE. See more in details section.
lag	A numeric value indicating lag-response involved in the missing data model. It can be either 1 or 2. The default is 1.

Details

Covariate matrix "x" should be complete. If missing data are present in "x", the elements in covariate vector will be replaced by zeros for individuals who have missing covariates.

The argument "x_mis" includes all covariates to fit the missing data model. It does not contains a lag variable based on the outcome y. The argument "lag" in this function will automatically add lag-response variables to indicate the data missing at random.

Either arguments "candidate.sets" or "name.var.sets" is used to identify the set of candidate mean model. If both arguments are provided, only the argument "name.var.sets" will be used.

When joints=TRUE, the argument "candidate.cor.sets" can contain multiple correlation structures; however, when joints=FALSE, it should contain either of "independence","exchangeable", "ar1". If multiple correlation structures are provided, only the first one will be used.

Value

A vector with each element containing MLIC value for each candidate model. The row name of this vector is the selected correlation structure.

Examples

```
## tests
# load data
data(wgeesimdata)
dist<-"binomial"
x<-wgeesimdata$x
y<-wgeesimdata$y
x_mis<-wgeesimdata$x_mis
r<-wgeesimdata$obs_ind
id<-wgeesimdata$id
time=3
candidate.sets<-list(c(1,2))
candidate.cor.sets<-c("independence")
criterion.mlic<-MLIC.wgee(x,y,x_mis,r,id,time,candidate.sets,
            name.var.sets=NULL,dist,candidate.cor.sets,joints=FALSE)
criterion.mlic
```

| MLICwgee | *Model selection based on MLIC under the syntax of WGEE (Main function).* |

Description

The function [MLICwgee](#) provides the model selection under the syntax of the wgeesel package.

Usage

```
MLICwgee(models, candidate.cor.sets, data, model_mis, family,r,id,time)
```

Arguments

`models`	A list of formulas. See corresponding documentation to wgeesel.
`candidate.cor.sets`	
	A vector containing candidate correlation structures. It can be any subset of c("independence","exchangeable", "ar1").
`data`	A data frame containing the variables in both the main model and the missing model.
`model_mis`	A formula used in the missing data model.
`family`	A description of the error distribution and link function to be used in the model. The details are given under "Details".
`r`	A vector indicating the observation of data: 1 for observed records (both outcome and covariates are observed for a given subject), and 0 for unobserved records. The default setup is that all data are observed.
`id`	A vector indicating subject id.
`time`	The number of observations in total for each subject.

Details

Three commonly used distributions are considered: "gaussian", "poisson", "binomial". For the current package, the identity link is considered for a "gaussian" distribution; the log link is considered for a "poisson" distribution; the logit link is considered for a "binomial" distribution;n;

Value

A list with two items: model selection result based on ELCIC; An object of "wgee" based on the selected model.

Examples

```
## tests
# load data
data(wgeesimdata)
family<-binomial()
r<-wgeesimdata$obs_ind
id<-wgeesimdata$id
time<-3
dat <- data.frame(y=wgeesimdata$y, wgeesimdata$x,x_mis1=wgeesimdata$x_mis[,2])
models <- list(y~x1+x2)
model_mis<-r~x_mis1
candidate.cor.sets<-c("exchangeable")
##not run
#output<-MLICwgee(models, candidate.cor.sets,data=dat,model_mis,family,r,id,time)
#output$model.selection
#output$wgee.output
```

QICc.gee	*Joint selection procedure of marginal mean and correlation structures in longitudinal data based on QIC*

Description

This function provides the Joint selection of marginal mean and correlation structures in longitudinal data based on QIC.

Usage

```
QICc.gee(x,y,id,dist,candidate.sets=NULL, name.var.sets=NULL,
    candidate.cor.sets=c("independence","exchangeable", "ar1"), joints=TRUE)
```

Arguments

x	A matrix containing covariates. The first column should be all ones corresponding to the intercept. Covariate matrix should be complete.
y	A vector containing outcomes.
id	A vector indicating subject id.
dist	A specified distribution. It can be "gaussian", "poisson",and "binomial".
candidate.sets	A list containing index corresponding to candidate covariates.
name.var.sets	A list containing names of candidate covariates. The names should be subset of column names of x matrix.
candidate.cor.sets	
	A vector containing candidate correlation structures. When joints=TRUE, it can be any subset of c("independence","exchangeable", "ar1"). The default is c("independence","exchangeable", "ar1"). When joints=FALSE, it should be either of "independence","exchangeable", "ar1". See more in details section.
joints	A logic value for joint selection of marginal mean and working correlation structure. The default is TRUE.

Details

Either arguments "index.var" or "name.var" is used to identify the candidate mean model. If both arguments are provided, only the argument "name.var" will be used.

When joints=TRUE, the argument "candidate.cor.sets" can contain multiple correlation structures; however, when joints=FALSE, it should contain either of "independence","exchangeable", "ar1". If multiple correlation structures are provided, only the first one will be used.

Value

A vector with each element containing QIC value for each candidate model. The row name of this vector is the selected correlation structure.

Examples

```
## tests
# load data
data(geesimdata)
x<-geesimdata$x
y<-geesimdata$y
id<-geesimdata$id
r<-rep(1,nrow(x))
time<-3
candidate.sets<-list(c(1,2),c(1,2,3))
candidate.cor.sets<-c("exchangeable")
dist="poisson"
criterion.qic<-QICc.gee(x=x,y=y,id=id,dist=dist,candidate.sets=candidate.sets,
                    name.var.sets=NULL,candidate.cor.sets=candidate.cor.sets)
criterion.qic
```

QICW.wgee   *The whole QICW procedure for joint selection of mean structure and correlation structure for missing longitudinal data under the mechanism of missing at random and drop-out*

Description

This function provides the overall QICW procedure for joint selection of mean structure and correlation structure in longitudinal data missing at random. It is also able to implement marginal mean structure selection given a prespecified working correlation structure. The data is dropout missing and missing at random.

Usage

```
QICW.wgee(x,y,x_mis,r,id,time,candidate.sets,name.var.sets=NULL,
    dist,candidate.cor.sets=c("independence","exchangeable", "ar1"), joints=TRUE,lag=1)
```

Arguments

x	A matrix containing covariates. The first column should be all ones corresponding to the intercept if the intercept is expected in the marginal mean model. Covariate matrix should be complete.
y	A vector containing outcomes. Use NA to indicate missing outcomes.
x_mis	A matrix containing covariates for the missing data model. The first column should be all ones corresponding to the intercept. This covariate matrix should be complete and all observed. See more in details section.
r	A vector indicating the observation of outcomes: 1 for observed records, and 0 for unobserved records.
id	A vector indicating subject id.

time	The number of observations in total for each subject.
candidate.sets	A list containing index corresponding to candidate covariates. See more in details section.
name.var.sets	A list containing names of candidate covariates. The names should be subset of column names of x matrix. See more in details section.
dist	A specified distribution. It can be "gaussian", "poisson",and "binomial".
candidate.cor.sets	
	A vector containing candidate correlation structures. When joints=TRUE, it is c("independence","exchangeable", "ar1") as default. When joints=FALSE, it should be either of "independence","exchangeable", "ar1". See more in details section.
joints	A logic value for joint selection of marginal mean and working correlation structure. The default is TRUE. See more in details section.
lag	A numeric value indicating lag-response involved in the missing data model. It can be either 1 or 2. The default is 1.

Details

Covariate matrix "x" should be complete. If missing data are present in "x", the elements in covariate vector will be replaced by zeros for individuals who have missing covariates.

The argument "x_mis" includes all covariates to fit the missing data model. It does not contains a lag variable based on the outcome y. The argument "lag" in this function will automatically add lag-response variables to indicate the data missing at random.

Either arguments "candidate.sets" or "name.var.sets" is used to identify the set of candidate mean model. If both arguments are provided, only the argument "name.var.sets" will be used.

When joints=TRUE, the argument "candidate.cor.sets" can contain multiple correlation structures; however, when joints=FALSE, it should contain either of "independence","exchangeable", "ar1". If multiple correlation structures are provided, only the first one will be used.

Value

A vector with each element containing QICW value for each candidate model. The row name of this vector is the selected correlation structure.

Examples

```
## tests
# load data
data(wgeesimdata)
dist="binomial"
x<-wgeesimdata$x
y<-wgeesimdata$y
x_mis<-wgeesimdata$x_mis
r<-wgeesimdata$obs_ind
id<-wgeesimdata$id
time=3
candidate.sets<-list(c(1,2))
```

```
candidate.cor.sets<-c("independence")
criterion.qicw<-QICW.wgee(x,y,x_mis,r,id,time,candidate.sets,
          name.var.sets=NULL,dist,candidate.cor.sets,joints=FALSE)
criterion.qicw
```

QICWwgee   *Model selection based on QICW under the syntax of WGEE (Main function).*

Description

The function `QICWwgee` provides the model selection under the syntax of the wgeesel package.

Usage

```
QICWwgee(models, candidate.cor.sets, data, model_mis, family,r,id,time)
```

Arguments

models	A list of formulas. See corresponding documentation to wgeesel.
candidate.cor.sets	
	A vector containing candidate correlation structures. It can be any subset of c("independence","exchangeable", "ar1").
data	A data frame containing the variables in both the main model and the missing model.
model_mis	A formula used in the missing data model.
family	A description of the error distribution and link function to be used in the model. The details are given under "Details".
r	A vector indicating the observation of data: 1 for observed records (both outcome and covariates are observed for a given subject), and 0 for unobserved records. The default setup is that all data are observed.
id	A vector indicating subject id.
time	The number of observations in total for each subject.

Details

Three commonly used distributions are considered: "gaussian", "poisson", "binomial". For the current package, the identity link is considered for a "gaussian" distribution; the log link is considered for a "poisson" distribution; the logit link is considered for a "binomial" distribution;n;

Value

A list with two items: model selection result based on ELCIC; An object of "wgee" based on the selected model.

Examples

```
## tests
# load data
data(wgeesimdata)
family<-binomial()
r<-wgeesimdata$obs_ind
id<-wgeesimdata$id
time<-3
dat <- data.frame(y=wgeesimdata$y, wgeesimdata$x,x_mis1=wgeesimdata$x_mis[,2])
models <- list(y~x1+x2)
model_mis<-r~x_mis1
candidate.cor.sets<-c("exchangeable")
##not run
#output<-QICWwgee(models, candidate.cor.sets,data=dat,model_mis,family,r,id,time)
#output$model.selection
#output$wgee.output
```

---

respiratorydata	*Data from a clinical trial comparing two treatments for a respiratory illness*

Description

The data are from a clinical trial of patients with respiratory illness, where 111 patients from two different clinics were randomized to receive either placebo or an active treatment. Patients were examined at baseline and at four visits during treatment.

Usage

```
data(respiratorydata)
```

Format

An object of class `"list"`

y Respiratory status at each visit, categorized as 1 = good, 0 = poor.

x A full covariate matrix. It contains intercept, center(1: center 2, 0: center 1), sex (1: male,0: female), treat (1: treatment, 0: placebo), visit: id of each of four visits, baseline: respiratory status at baseline, age: in years at baseline.

id Patient ID

r A vector indicating the observation of data (1: observed; 0: missing). All data are observed.

References

add here

wgeesimdata	*Data simulated for model selection under WGEE framework for miss-ing longitudinal data under the mechanism of missing at random and drop-out.*

Description

Data simulated for model selection under WGEE framework for missing longitudinal data under the mechanism of missing at random and drop-out.

Usage

```
data(wgeesimdata)
```

Format

An object of class `"list"`

y The outcomes generated from Bernoulli distribution with three repeated measurements from each subject. The data is longitudinal dropout and missing at random.

x A full covariate matrix. The first column corresponds to the intercept; the second column contains time-independent continuous-scale variable x1; the third column contains a doctor-visit variable x2; the third column contains time-independent binary-scale variable x3.

x_mis A covariate matrix for missing data model. The first column corresponds to the intercept; the second column contains continuous variable x_mis1.

id A vector indicating subject id.

obs_ind A vector indicating observations for each record: 1 for observed records, and 0 for unobserved records.

References

This data set was artificially created for the ELCIC package.

Examples

```
data(wgeesimdata)
wgeesimdata$y
```

Index