Package ‘mgee2’

December 3, 2023

Title  Marginal Analysis of Misclassified Longitudinal Ordinal Data
Version  0.5
Description  Three estimating equation methods are provided in this package for marginal analysis of longitudinal ordinal data with misclassified responses and covariates.
The naive analysis which is solely based on the observed data without adjustment may lead to bias.
The corrected generalized estimating equations (GEE2) method which is unbiased requires the misclassification parameters to be known beforehand.
The corrected generalized estimating equations (GEE2) with validation subsample method estimates the misclassification parameters based on a given validation set. This package is an implementation of Chen (2013) <doi:10.1002/bimj.201200195>.

Depends  R (>= 3.6.0)
License  GPL (>= 2)
Encoding  UTF-8
Imports  MASS, ggplot2
LazyData  true
RoxygenNote  7.1.2
Suggests  rmarkdown, knitr
NeedsCompilation  yes
Author  Yuliang Xu [aut, cre],
        Zhijian Chen [aut],
        Shuo Shuo Liu [aut],
        Grace Yi [aut]
Maintainer  Yuliang Xu <yuliangx@umich.edu>
Repository  CRAN
Date/Publication  2023-12-03 04:00:02 UTC

R topics documented:

  heart ......................................................... 2
  mgee2k ....................................................... 4
heart: preprocessed Framingham Heart Study Teaching data

Description

heart: preprocessed Framingham Heart Study Teaching data

Usage

heart

Format

a dataframe with 1830 rows and 42 variables, a total of 915 participants.

**RANDID** individual id number

**HBP** a factor variable derived from SYSBP. HBP=0 indicates SBP below 140 mmHg, HBP=1 indicates SBP between 140 mmHg and 159 mmHg, and HBP=2 indicates SBP larger than 160 mmHg

**chol** a factor variable derived from TOTCHOL. 0=normal (less than 200 mg/dL), 1=borderline high (200-239mg/dL), 2=hypercholesterolemia (greater than 240 mg/dL)

**exam3** a factor variable. 1 if the observation belongs to exam 3, 0 otherwise.

For all other variables, please refer to https://biolincc.nhlbi.nih.gov/media/teachingstudies/FHS_Teaching_Longitudinal_Data_Documentation.pdf?link_time=2021-03-17_16:09:25.977880, The full teaching data set can be requested from https://biolincc.nhlbi.nih.gov/teaching/

Details

The authors thank Boston University and the National Heart, Lung, and Blood Institute (NHLBI) for providing the data set from the Framingham Heart Study (No. N01-HC-25195) in the illustration. The Framingham Heart Study is conducted and supported by the NHLBI in collaboration with Boston University. This package was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University, or NHLBI.
heart

References


Examples

{
  data(heart)
  #descriptive plots:
  if(0){
    library(mgee2)
    library(ggplot2)
    # covariates
    heart$chol = as.factor(heart$chol)
    heart$CURSMOKE = as.factor(heart$CURSMOKE)
    heart$exam3 = as.factor(heart$exam3)
    levels(heart$exam3) = c("exam2","exam3")
    ggplot(heart, aes(x=AGE, y=SYSBP)) +
    geom_line(aes(group=RANDID), alpha=0.5) +
    geom_smooth(se=FALSE, size=2) +
    ylab("SBP") +
    facet_grid(chol~CURSMOKE, labeller = label_both)
    # trend
    ggplot(heart, aes(x=AGE, y=SYSBP,
      colour = chol,linetype = CURSMOKE)) +
    geom_smooth(method="lm", se=FALSE) +
    ylab("SBP") + facet_wrap(~exam3)+
    scale_color_brewer(palette = "Dark2")
  }

  #Example 1:
  heart$chol = as.factor(heart$chol)
  heart$exam3 = as.factor(heart$exam3)
  ## set misclassification parameters to be known.
  varphiMat <- gamMat <- log( cbind(0.04/0.95, 0.01/0.95,
                          0.95/0.03, 0.02/0.03,
                          0.04/0.01, 0.95/0.01) )
  mgee2k.fit = mgee2k(formula = HBP~chol+AGE+CURSMOKE+exam3, id = "RANDID",
                      data = heart,
                      corstr = "exchangeable", misvariable = "chol",
                      gamMat = gamMat,
                      varphiMat = varphiMat)
  summary(mgee2k.fit)

  #Example 2:
  naigee.fit = ordGEE2(formula = HBP~chol+AGE+CURSMOKE+exam3, id = "RANDID",
                      data = heart, corstr = "exchangeable")
  summary(naigee.fit)
}
Description
Corrected GEE2 for ordinal data. This method yields unbiased estimators, but the misclassification parameters are required to be known.

Usage
mgee2k(
  formula, id, data,
  corstr = "exchangeable", misvariable,
  gamMat, varphiMat,
  maxit = 50, tol = 0.001
)

Arguments
- **formula**: a formula object which specifies the relationship between the response and covariates for the observed data.
- **id**: a character object which records individual id in the data.
- **data**: a dataframe or matrix object for the observed data set.
- **corstr**: a character object. The default value is "exchangeable", corresponding to the structure where the association between two paired responses is considered to be a constant. The other option is "log-linear" which indicates the log-linear association between two paired responses.
- **misvariable**: a character object which names the error-prone covariate W.
- **gamMat**: a matrix object which records the misclassification parameter gamma for response Y.
- **varphiMat**: a matrix object which records the misclassification parameter phi for covariate X.
- **maxit**: an integer which specifies the maximum number of iterations. The default is 50.
- **tol**: a numeric object which indicates the tolerance threshold. The default is 1e-3.

Details
*mgee2k* implements the misclassification adjustment method outlined in Chen et al.(2014) where the misclassification parameters are known. In this case, validation data are not required, and only the observed data of the outcome and covariates are needed for the implementation.
**mgee2k**

**Value**

A list with component

- **beta**: the coefficients in the order as those specified in the formula for the response and covariates.
- **alpha**: the coefficients for paired responses global odds ratios. The number of alpha coefficients corresponds to the paired responses odds ratio structure selected in corstr. When corstr="exchangeable", only one baseline alpha is fitted. When corstr="log-linear", baseline, first order, second order (interaction) terms are fitted.
- **variance**: variance-covariance matrix of the estimator of all parameters.
- **convergence**: a logical variable; TRUE if the model converges.
- **iteration**: the number of iterations for the estimates of the model parameters to converge.
- **differ**: a list of difference of estimation for convergence.
- **call**: Function called.

**References**


**Examples**

```r
if(0){
data(obs1)
obs1$visit <- as.factor(obs1$visit)
obs1$treatment <- as.factor(obs1$treatment)
obs1$S <- as.factor(obs1$S)
obs1$W <- as.factor(obs1$W)
## set misclassification parameters to be known.
varphiMat <- gamMat <- log( cbind(0.04/0.95, 0.01/0.95, 0.95/0.03, 0.02/0.03, 0.04/0.01, 0.95/0.01) )
mgee2k.fit = mgee2k(formula = S~W+treatment+visit, id = "ID", data = obs1, corstr = "exchangeable", misvariable = "W", gamMat = gamMat, varphiMat = varphiMat)
}
```
mgee2v

Description

Corrected GEE2 for ordinal data, with validation subsample

Usage

mgee2v(
  formula,
  id,
  data,
  corstr = "exchangeable",
  misvariable = "W",
  valid.sample.ind = "delta",
  y.mcformula,
  x.mcformula,
  maxit = 50,
  tol = 0.001
)

Arguments

formula  a formula object which specifies the relationship between the response and co-
           variates for the observed data.

id       a character object which records individual id in the data.

data     a dataframe or matrix object for the observed data set.

corstr   a character object. The default value is "exchangeable", corresponding to the
           structure where the association between two paired responses is considered to
           be a constant. The other option is "log-linear" which indicates the log-linear
           association between two paired responses.

misvariable a character object which names the error-prone covariate W.

valid.sample.ind a string object which names the indicator variable delta. When a data point
                  belongs to the validation set, delta = 1; otherwise 0.

y.mcformula a string object which indicates the misclassification formula between true re-
              sponse Y and surrogate(observed) response S.

x.mcformula a string object which indicates the misclassification formula between true error-
              prone covariate X and surrogate W.

maxit     an integer which specifies the maximum number of iterations. The default is 50.

tol       a numeric object which indicates the tolerance threshold. The default is 1e-3.
mgee2v

Details

The function mgee2v does not require the misclassification parameters to be known, but require the availability of validation data. Similar to mgee2k, the function mgee2v needs the data set to be structured by individual id, i=1,...,n, and visit time, j_i=1,...,m_i. The data set should contain the observed response and covariates S and W. To indicate whether or not a subject is in the validation set, an indicator variable delta should be added in the data set, and we use a column named valid.sample.ind for this purpose. The column name of the error-prone covariate W should also be specified in misvariable.

Value

A list with component

beta the coefficients in the order of 1) all non-baseline levels for response, 2) covariates - same order as specified in the formula

alpha the coefficients for paired responses global odds ratios. Number of alpha coefficients corresponds to the paired responses odds ratio structure selected in "corstr"; when corstr="exchangeable", only one baseline alpha is fitted.

variance variance-covariance matrix of all fitted parameters

convergence a logical variable, TRUE if the model converges

iteration number of iterations for the model to converge

call Function called

References


Examples

if(0){
data(obs1)
obs1$Y <- as.factor(obs1$Y)
obs1$X <- as.factor(obs1$X)
obs1$visit <- as.factor(obs1$visit)
obs1$treatment <- as.factor(obs1$treatment)
obs1$S <- as.factor(obs1$S)
obs1$W <- as.factor(obs1$W)
mgee2v.fit = mgee2v(formula = S~W+treatment+visit, id = "ID", data = obs1,
y.mcformula = "S-1", x.mcformula = "W-1", misvariable = "W",
valid.sample.ind = "delta",
corstr = "exchangeable")
}

mgee2_pkg

A list of external packages and functions used in mgee2

Description

A list of external packages and functions used in mgee2

obs1

obs1: simulated observed data

Description

obs1: simulated observed data

Usage

obs1

Format

a dataframe with 3000 rows and 8 variables

ID  individual id number
Y   true response, factor variable
X   true error-prone covariate, factor variable
treatment  error-free covariate
visit  serial number of each visit
S   observed response, same as Y when in the validation set (delta=1)
W   observed error-prone covariate, same as X when in the validation set (delta=1)
delta  indicator variable, 1 if in the validation set, 0 if not.
Description

This function provides a naive approach to estimate the data without any correction or misclassification parameters. This may lead to biased estimation for response parameters.

Usage

ordGEE2(formula, id, data, corstr = "exchangeable", maxit = 50, tol = 0.001)

Arguments

- **formula**: a formula object: a symbolic description of the model with error-prone response, error-prone covariates and other covariates.
- **id**: a character object which records individual id in the data.
- **data**: a dataframe or matrix of the observed data, including id, error-prone ordinal response error-prone ordinal covaritaeas, other covariates.
- **corstr**: a character object. The default value is "exchangeable", corresponding to the structure where the association between two paired responses is considered to be a constant. The other option is "log-linear" which indicates the log-linear association between two paired responses.
- **maxit**: an integer which specifies the maximum number of iterations. The default is 50.
- **tol**: a numeric object which indicates the tolerance threshold. The default is 1e-3.

Details

In addition to developing the package *mgee2* to implement the methods of Chen et al.(2014) which accommodate misclassification effects in inferential procedures, we also implement the naive method of ignoring the feature of misclassification, and call the resulting function *ordGEE2*. This function can be used together with the precedingly described *mgee2k* or *mgee2v* to evaluate the impact of not addressing misclassification effects.

Value

A list with component

- **beta**: the coefficients in the order of 1) all non-baseline levels for response, 2) covariates - same order as specified in the formula.
- **alpha**: the coefficients for paired responses global odds ratios. Number of alpha coefficients corresponds to the paired responses odds ratio structure selected in "corstr"; when corstr="exchangeable", only one baseline alpha is fitted.
- **variance**: variance-covariance matrix of all fitted parameters.
- **convergence**: a logical variable, TRUE if the model converges.
iteration number of iterations for the model to converge
differ a list of difference of estimation for convergence

##
call Function called

References


Examples

data(obs1)
obs1$Y <- as.factor(obs1$Y)
obs1$X <- as.factor(obs1$X)
obs1$visit <- as.factor(obs1$visit)
obs1$treatment <- as.factor(obs1$treatment)
obs1$S <- as.factor(obs1$S)
obs1$W <- as.factor(obs1$W)

naigee.fit = ordGEE2(formula = S~W+treatment+visit, id = "ID",
data = obs1, corstr = "exchangeable")

plot_model

Description

This function gives plot of the odds ratio or shows the iteration for convergence.

Usage

plot_model(x, conv = FALSE)

Arguments

x results from the fitted model.
conv defulated for odds ratio plot, otherwise show the iteration plot.

Value

plot odds ratio with CIs or plot of the iterations.
Examples

```r
beta=c(0.1,0.2,0.3)
alpha=c(0.4,0.5)
variance=c(0.8,0.5,0.7,0.3,0.4)
x=list(beta,alpha,variance)
names(x)=c("beta","alpha","variance")
plot_model(x)
```

Description

print.summary.mgee2

Usage

```r
## S3 method for class 'summary.mgee2'
print(x, ...)
```

Arguments

- `x` the summary results
- `...` Other parameters

Value

a table of summary statistics

Description

summary.mgee2

Usage

```r
## S3 method for class 'mgee2'
summary(object, ...)
```

Arguments

- `object` The fitted model
- `...` Other parameters summary function for mgee2 method output
Index

* datasets
  heart, 2
  obs1, 8

heart, 2
mgee2_pkg, 8
mgee2k, 4
mgee2v, 6
obs1, 8
ordGEE2, 9

plot_model, 10
print.summary.mgee2, 11

summary.mgee2, 11