Package 'mmm'

February 20, 2015

Type Package

Title an R package for analyzing multivariate longitudinal data with
multivariate marginal models

Version 1.4

Date 2014-01-01

Author Ozgur Asar, Ozlem Ilk

Depends gee

Maintainer Ozgur Asar <o.asar@lancaster.ac.uk>

Description fits multivariate marginal models for multivariate longitudinal data for both continuous and discrete responses

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2014-01-01 05:19:53

R topics documented:

---

mmm-package	*mmm: an R package for analyzing multivariate longitudinal data with multivariate marginal models*

Description

fits multivariate marginal models for multivariate longitudinal data for both continuous and discrete responses

Asar, O., Ilk, O. (2013). mmm: an R package for analyzing multivariate longitudinal data with multivariate marginal models, *Computer Methods and Programs in Biomedicine*, **112**, 649-654.

Details

Package:	mmm
Type:	Package
Version:	1.4
Date:	2013-01-01
License:	GPL (>=2)

---

mmm	*an R function to fit the multivariate marginal models to analyze multivariate longitudinal data*

Description

fits multivariate marginal models to analyze multivariate longitudinal data, for both continuous and discrete responses

Usage

```
mmm(formula, id, data = NULL, correlation = NULL, initEstim = NULL, tol = 0.001,
maxiter = 25, family = "gaussian", corStruct = "independence", Mv = 1,
silent = TRUE, scale.fix = FALSE, scale.value = 1)
```

Arguments

formula	a formula expression, see the examples given below.
id	a vector for identification of the clusters.
data	an optional data frame.

correlation	a user specified square matrix for the working correlation matrix, appropriate when `corstr="fixed"`.
initEstim	user specified initials for the parameter estimates.
tol	the tolerance which specifies the convergency of the algorithm.
maxiter	the maximum number of iterations to be consumed by the algorithm.
family	an object which defines the link and variance function. The possible choices are same with the ones in the `"gee"` package. For details see the [gee](#) documentation. Note that `family=binomial` handles multivariate longitudinal binary data, `family=poisson` handles multivariate longitudinal count data, `family=gaussian` handles multivariate longitudinal (normal type) continous data and `family=gamma` handles multivariate longitudinal (gamma type) continous data.
corStruct	a character string which defines the structure of the working correlation matrix. For details see the [gee](#) documentation.
Mv	specifies the lag value, e.g. specification of `"corstr=AR-M"` and `"Mv=1"` indicates AR(1).
silent	a logial variable which decides the print of the iterations.
scale.fix	a logical variable for fixing the scale parameter to a user specified value.
scale.value	a user specified scale parameter value, appropriate when `scale.fix=TRUE`.

Value

Returns an onject of the results. See the examples given below.

Note

Version 1.1.

Author(s)

Ozgur Asar, Ozlem Ilk

References

Asar, O. (2012). *On multivariate longitudinal binary data models and their applications in forecasting*. MS Thesis, Middle East Technical University. Available online at http://www.lancaster.ac.uk/pg/asar/thesis-Ozgur-Asar

Asar, O., Ilk, O. (2013). mmm: an R package for analyzing multivariate longitudinal data with multivariate marginal models, *Computer Methods and Programs in Biomedicine*, **112**, 649-654.

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

Shelton, B. J., Gilbert, G. H., Liu, B., Fisher, M. (2004). A SAS macro for the analysis of multivariate longitudinal binary outcomes. *Computer Methods and Programs in Biomedicine*, **76**, 163-175.

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

See Also

gee

Examples

```
###########################
## Binary data example ##
###########################

data(motherStress)
fit1<-mmm(formula=cbind(stress,illness)~married+education+
employed+chlth+mhlth+race+csex+housize+bstress+billness+
week,id=motherStress$id,data=motherStress,family=binomial,corStruct="exchangeable")
summary(fit1)

#########################
## Count data example ##
#########################

## First we illustrate how the data set is simulated
## Then the R script to analyze the data set by mmm is given
## Note: no need to run the script to generate the data set, unless of interest

## Not run:
### Generating the data by the help of 'corcounts' package

# loading the package 'corcounts'
library("corcounts")

# setting the seed to 12
set.seed(12)

# number of subjects in the study
n1 <- 500

# defining the response and covariate families (Poi indicates Poisson distribution)
margins <- c("Poi","Poi","Poi","Poi","Poi","Poi","Poi","Poi","Poi")

# the means of the responses and covariate. while 5 and 8 are the means of the responses
# 20 is the mean of the time independent covariate
mu <- c(5, 8, 20, 5, 8, 5, 8, 5, 8)

# the correlation structure which 'corcounts' uses to generate correlated count data
# (unstr indicates unstructured correlation structure)
corstr <- "unstr"

# the correlation matrix corcounts assumes the correlated count data have
corpar<-matrix(c(1,0.4,0.6,0.9,0.37,0.8,0.34,0.7,0.31,
0.4,1,0.6,0.37,0.9,0.34,0.8,0.31,0.7,
0.6,0.6,1,0.6,0.6,0.6,0.6,0.6,0.6,
0.9,0.37,0.6,1,0.4,0.9,0.37,0.8,0.34,
0.37,0.9,0.6,0.4,1,0.37,0.9,0.34,0.8,
```

```
0.8,0.34,0.6,0.9,0.37,1,0.4,0.9,0.37,
0.34,0.8,0.6,0.37,0.9,0.4,1,0.37,0.9,
0.7,0.31,0.6,0.8,0.34,0.9,0.37,1,0.4,
0.31,0.7,0.6,0.34,0.8,0.37,0.9,0.4,1),ncol=9,byrow=T)

# generating the correlated count data by 'rcounts' function avaiable in 'corcounts'
data1 <- rcounts(N=n1,margins=margins,mu=mu,corstr=corstr,corpar=corpar)

### The reconstruction of the generated correlated count data to
### the longitudinal data (long) format

# seperating the bivariate responses measured at the first time
# point and the time independent covariate
time11<-data1[,1:3]

# seperating the bivariate responses measured at the second time
# point and combining them with the time independent covariate
time12<-cbind(data1[,4:5],data1[,3])

# seperating the bivariate responses measured at the third time
# point and combining them with the time independent covariate
time13<-cbind(data1[,6:7],data1[,3])

# seperating the bivariate responses measured at the fourth time
# point and combining them with the time independent covariate
time14<-cbind(data1[,8:9],data1[,3])

# combining the data for all the time points
data12<-rbind(time11,time12,time13,time14)

# constructing the time variable
time1<-matrix(rep(seq(1:4),each=n1))

# constructing the id variable
id1<-matrix(rep(seq(1:n1),4))

# combining the id of the subjects, the simulated data and the time variable
data13<-cbind(id1,data12,time1)

# reconstructing the data subject by subject which 'mmm' expects it has
data14<-NULL
for (i in 1:n1) data14<-rbind(data14,data13[data13[,1]==i,])

### Data manipulations on the covariates

# taking natural logarithm of the time independent covariate
data14[,4]<-log(data14[,4])

# standardizing time variable
data14[,5]<-scale(data14[,5])

# adding the interaction of the time independent covariate
# and time as a new covariate
```

```
multiLongCount<-as.data.frame(cbind(data14,data14[,4]*data14[,5]))
names(multiLongCount)<-c("ID","resp1","resp2","X","time","X.time")

## End(Not run)

### R script to analyze the count data set
### It is already stored in mmm pacakge

data(multiLongCount)
fit2<-mmm(formula=cbind(resp1,resp2)~X+time+X.time,
id=multiLongCount$ID,data=multiLongCount,family=poisson,corStruct="unstructured")
summary(fit2)

#############################
## Continuous data example ##
#############################

## First we illustrate how the data set is simulated
## Then the R script to analyze the data set by mmm is given
## Note: no need to run the script to generate the data set, unless of interest

## Not run:
### Generating the data by the help of mvtnorm package

# loading package 'mvtnorm'
library("mvtnorm")

# number of subjects in the study
n2<-500

# setting the seed to 12
set.seed(12)

# specifying the correlation matrix which 'mvtnorm' assumes the correlated data have
cormat<-matrix(c(1,0.4,0.6,0.9,0.37,0.8,0.34,0.7,0.31,
0.4,1,0.6,0.37,0.9,0.34,0.8,0.31,0.7,
0.6,0.6,1,0.6,0.6,0.6,0.6,0.6,0.6,
0.9,0.37,0.6,1,0.4,0.9,0.37,0.8,0.34,
0.37,0.9,0.6,0.4,1,0.37,0.9,0.34,0.8,
0.8,0.34,0.6,0.9,0.37,1,0.4,0.9,0.37,
0.34,0.8,0.6,0.37,0.9,0.4,1,0.37,0.9,
0.7,0.31,0.6,0.8,0.34,0.9,0.37,1,0.4,
0.31,0.7,0.6,0.34,0.8,0.37,0.9,0.4,1),ncol=9,byrow=T)

# variances of the responses and time independent covariate
# while 0.97 and 1.1 correspond to the variances of the bivariate responses
# 4 corresponds to the variance of the time independent covariate
variance<-c(0.97,1.1,4,0.97,1.1,0.97,1.1,0.97,1.1)

# constructing the (diaonal) standard deviation matrix
std<-diag(sqrt(variance),9)

# constructing the variance covariance matrix, sigma
```

```
sigma<-std

# generating the correlated continuous data utilizing 'rmvnorm' function available
# in 'mvtnorm'; method="svd" indicates use of 'singular value decomposition method
data2<-rmvnorm(n2,mean = rep(0,nrow(sigma)),sigma=sigma,method="svd")


### The reconstruction of the generated correlated continuous data to the
### longitudinal data (long) format

# seperating the bivariate responses measured at first time point
# and the time independent covariate
time21<-data2[,1:3]

# seperating the bivariate responses measured at second time point
# and combining them with the time independent covariate
time22<-cbind(data2[,4:5],data2[,3])

# seperating the bivariate responses measured at third time point
# and combining them with the time independent covariate
time23<-cbind(data2[,6:7],data2[,3])

# seperating the bivariate responses measured at fourth time point
# and combining them with the time independent covariate
time24<-cbind(data2[,8:9],data2[,3])

# combining the data for all the time points
data22<-rbind(time21,time22,time23,time24)

# constructing the time variable
time2<-matrix(rep(seq(1:4),each=n2))

# constructing the id variable
id2<-matrix(rep(seq(1:n2),4))

# combining the id of the subjects, the generated data and the time variable
data23<-cbind(id2,data22,time2)

# reconstructing the data subject by subject which 'mmm' expects it has
data24<-NULL
for (i in 1:n2) data24<-rbind(data24,data23[data23[,1]==i,])

### Data manipulations on the covariates

# standardizing the time variable
data24[,5]<-scale(data24[,5])

# adding the interaction of the time independent covariate
# and time as a new covariate
multiLongGaussian<-as.data.frame(cbind(data24,data24[,4]*data24[,5]))
names(multiLongGaussian)<-c("ID","resp1","resp2","X","time","X.time")

## End(Not run)
```

```
### R script to analyze the continuous data set
### It is already stored in mmm pacakge

data(multiLongGaussian)
fit3<-mmm(formula=cbind(resp1,resp2)~X+time+X.time,
id=multiLongGaussian$ID,data=multiLongGaussian,family=gaussian,corStruct="unstructured")
summary(fit3)
```

---

motherStress   *Mother's Stress and Children's Morbidity Study*

Description

A data frame with 2004 observations on the following 14 variables. motherStress is a longitudinal dataset which includes daily information of the participants. There are 167 mothers and children enrolled in the study.

Usage

```
data(motherStress)
```

Format

The details of the columns of the data frame are given below.

id a vector for subject id

stress a vector for mother's stress at time t:1=presence, 0=absence

illness a vector for children's illness at time t: 1=presence, 0=absence

married a vector for marriage status of mother: 1=married, 0=other

education a vector for mother's education level: 0=high school or less, 1=high school graduate

employed a numeric vector for mother's employment status: 1=employed, 0=unemployed

chlth a vector for children's health status at baseline: 0=very poor/poor, 1=fair, 2=good, 3=very good

mhlth a vector for mother's health status at baseline: 0=very poor/poor, 1=fair, 2=good, 3=very good

race a vector for child's race: 1=non-white, 0=white

csex a vector for child's gender: 1=female, 0=male

housize a vector for the size of the household: 0=2-3 people, 1=more than 3 people

bstress a vector for the baseline stress for the period of day 1 to 16; calculated as the mean of the stress status of the subjects in the period of day 1 to 16

billness a vector for the baseline illness for the period of day 1 to 16; calculated as the mean of the illness status of the subjects in the period of day 1 to 16

week a numeric vector for time: (day-22)/7

Details

The original data contains the information of the mothers and children in the study for 28 days. Because of the weak serial correlation in the period of day 1 to 16, it is ignored. Only the period of day 17 to 28 is included here. To catch the specific characteristic of the mothers and children, the averages of the stress and illness status of them are added as new covariates; bstress and billness. While the covariates have no missing observation, responses have very low percentages of missing values, 0.97

Source

http://faculty.washington.edu/heagerty/Books/AnalysisLongitudinal/datasets.html

References

Alexander, C. S., Markowitz, R. (1986). Maternal employment and use of pediatric clinic services. *Medical Care*, **24(2)**, 134-147.

Diggle, P. J., Heagerty, P., Liang, K. Y., Zeger, S. L. (2002). *Analysis of Longitudinal Data*. New York: Oxford University Press.

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

Examples

```
data(motherStress)
head(motherStress,10)
require(graphics)
mosaicplot(~motherStress$employed+motherStress$housize+motherStress$stress,color=TRUE)
```

---

multiLongCount	*Multivariate Longitudinal Count Data*

Description

A data frame with 2000 observations on the following 6 variables. multiLongCount is a simulated bivariate longitudinal count dataset assuming there are 500 subjects in the study whose data are collected at 4 equally-spaced time points.

Usage

```
data(multiLongCount)
```

Format

A data frame with 2000 observations on the following 6 variables.

ID a numeric vector for subject ID

resp1 a numeric vector for the first longitudinal count response

resp2 a numeric vector for the second longitudinal count response

X a numeric vector for the covariate, X

time a numeric vector for the time point at which observations are collected

X.time a numeric vector for the interaction between X and time

Details

The covariates, X and time are the standardized values indeed. The related interaction is calculated by using these standardized values. X is a time-independent covariate. The R script to generate the data set is given in the Examples section of the mmm function.

References

Asar, O. (2012). *On multivariate longitudinal binary data models and their applications in forecasting*. MS Thesis, Middle East Technical University

Erhardt, V. (2009). corcounts: Generate Correlated Count Random Variable. R package version 1.4. URL http://CRAN.R-project.org/package=corcounts.

Examples

```
data(multiLongCount)
plot(multiLongCount$X,multiLongCount$resp1)
```

---

multiLongGaussian   *Multivariate Longitudinal Continuous (Gaussian) Data*

Description

A data frame with 2000 observations on the following 6 variables. multiLongGaussian is a simulated bivariate longitudinal continuous dataset assuming there are 500 subjects in the study whose data are collected at 4 equally-spaced time points.

Usage

```
data(multiLongGaussian)
```

Format

A data frame with 2000 observations on the following 6 variables.

ID a numeric vector for subject ID

resp1 a numeric vector for the first longitudinal count response

resp2 a numeric vector for the second longitudinal count response

X a numeric vector for the covariate, X

time a numeric vector for the time point at which observations are collected

X.time a numeric vector for the interaction between X and time

Details

The covariates, X and time are the standardized values indeed. The related interaction is calculated by using these standardized values. X is a time-independent covariate. The R script to generate the data set is given in the Examples section of the mmm function.

References

Asar, O. (2012). *On multivariate longitudinal binary data models and their applications in forecasting*. MS Thesis, Middle East Technical University

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., Hothorn, T. (2011). mvtnorm: Multivariate Normal and t Distributions. R package version 0.9-96. URL http://CRAN.R-project.org/package=mvtnorm.

Examples

```
data(multiLongGaussian)
plot(multiLongGaussian$X,multiLongGaussian$resp1)
```

Index