Package 'mma'

December 14, 2016

Type Package

Title Multiple Mediation Analysis

Version 4.0-0

Date 2016-12-12

Author Qingzhao Yu and Bin Li

Maintainer Qingzhao Yu <qyu@lsuhsc.edu>

Depends R (>= 2.14.1), gbm, car, gplots, splines, survival

Description Used for general multiple mediation analysis.
The analysis method is described in Yu et al. (2014) <doi:10.4172/2155-6180.1000189> ``General Multiple Mediation Analysis With an Application to Explore Racial Disparity in Breast Cancer Survival'', published on Journal of Biometrics & Biostatistics, 5(2):189.

License GPL (>= 2)

URL https://www.r-project.org,
https://publichealth.lsuhsc.edu/Faculty_pages/qyu/index.html

NeedsCompilation no

Repository CRAN

Date/Publication 2016-12-14 08:13:19

R topics documented:

---

mma-package   *Mediation Analysis Package*

Description

This package is used to identify mediators and for general mediation analysis. Mediation effect refers to the effect conveyed by intervening variables to an observed relationship between an exposure and a response variable (outcome). In this package, the exposure is called the predictor, the intervening variables are called mediators. The mediation effects include the total effect, direct effect, and indirect effect, which are defined and the statistical inferences described in Yu et al.(2014).

Details

Package:	mma
Type:	Package
Version:	4.0
Date:	2016-08-18
License:	GPL (>= 2)

"data.org" is used to identify potential mediators. *"med"* , and *"boot.med"* are used for statistical inferences on the mediation effects when the predictor is binary or continuous. *"mma"* is a combined function that automatically identify potential mediators, based on which to make statistical inference on the mediation effects.

Author(s)

Qingzhao Yu <qyu@lsuhsc.edu>

Maintainer: Qingzhao Yu <qyu@lsuhsc.edu>

References

Baron, R.M., and Kenny, D.A. (1986) <doi:10.1037/0022-3514.51.6.1173>. "The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations," J. Pers Soc Psychol, 51(6), 1173-1182.

J.H. Friedman, T. Hastie, R. Tibshirani (2000) <doi:10.1214/aos/1016120463>. "Additive Logistic Regression: a Statistical View of Boosting," Annals of Statistics 28(2):337-374.

J.H. Friedman (2001) <doi: 10.1214/aos/1013203451>. "Greedy Function Approximation: A Gradient Boosting Machine," Annals of Statistics 29(5):1189-1232.

Yu, Q., Fan, Y., and Wu, X. (2014) <doi:10.4172/2155-6180.1000189>. "General Multiple Mediation Analysis With an Application to Explore Racial Disparity in Breast Cancer Survival," Journal of Biometrics & Biostatistics,5(2): 189.

Examples

```
data("weight_behavior")
 #binary x
 #binary y
 x=weight_behavior[,2:14]
 y=weight_behavior[,15]
 temp.b.b.glm<-mma(x,y,pred=2,contmed=c(8:10,12:13),binmed=c(7,11),
   binref=c(1,1),catmed=6,catref=1,predref="M",alpha=0.4,alpha2=0.4,
   jointm=NULL,margin=1, n=2,seed=sample(1:1000,1),nonlinear=FALSE,nu=0.001,
   D=3,n2=2)

 temp.b.b.mart<-mma(x,y,pred=2,contmed=c(8:10,12:13),binmed=c(7,11),
   binref=c(1,1),catmed=6,catref=1,predref="M",alpha=0.4,alpha2=0.4,
   jointm=NULL,margin=1, n=2,seed=sample(1:1000,1),nonlinear=TRUE,nu=0.05,
   D=3,n2=5)

 #continuous y
 x=weight_behavior[,2:14]
 y=weight_behavior[,1]
 temp.b.c.glm<-mma(x,y,pred=2,contmed=c(8:10,12:13),binmed=c(7,11),
   binref=c(1,1),catmed=6,catref=1,jointm=list(n=1,j1=8:10),predref="M",
   alpha=0.4,alpha2=0.4,n=2,seed=1,nonlinear=FALSE,nu=0.05,D=3,n2=2)
 temp.b.c.mart<-mma(x,y,pred=2,contmed=c(8:10,12:13),binmed=c(7,11),
   binref=c(1,1),catmed=6,catref=1,jointm=list(n=1,j1=8:10, j2=12:13),
   predref="M",alpha=0.4,alpha2=0.4, margin=1, n=2,seed=1,nonlinear=TRUE,
   nu=0.05,n2=2)
```

---

boot.med *Statistical Inference on Mediation Analysis with Continuous or Binary Predictor*

Description

To make inferences on the mediation effects when the predictor is continuous or binary.

Usage

```
boot.med(data,x=data$x, y=data$y,dirx=data$dirx,binm=data$binm,
         contm=data$contm,catm=data$catm,jointm=data$jointm,margin=1,
         n=20,seed=sample(1:1000,1),nonlinear=F,df=1,nu=0.001,D=3,
         distn=NULL,family1=NULL,n2=50,weight=rep(1,nrow(x)),
         refy=NULL,x.new=x,binpred=data$binpred,type=NULL)
```

Arguments

data   the list of result from data.org that organize the covariates, mediators, predictor and outcome. If data is FALSE, then need to set x, y, dirx, contm, catm, and jointm.

x	a data frame contains the predictor, all mediators and covariates. Need to set up only when data is FALSE.
y	the vector of outcome variable. Need to set up only when data is FALSE.
dirx	the predictor name or the column number of x that locates the predictor. Need to set up only when data is FALSE.
binm	the variable names or the column number of x that locates the binary mediators. Need to set up only when data is FALSE.
contm	the variable names or the column numbers of x that locate the potential continuous mediators. Need to set up only when data is FALSE.
catm	categorical mediators should be binarized and be presented as a list, where the first item is the number of categorical variables and the following items are the names or the column numbers of each binarized categorical variable in x. data.org organizes the categorical mediators in this format after they pass the mediator tests. Need to set up only when data is FALSE.
jointm	a list where the first item is the number of groups of joint mediators to be considered, and each of the following items identifies the names or the column numbers of the mediators in x for each group of joint mediators. Need to set up only when data is FALSE.
margin	the change in predictor when calculating the mediation effects, see Yu et al. (2014).
n	the time of resampling in calculating the indirect effects, default is n=20, see Yu et al. (2014).
seed	set seed to make the calculation repeatable. The default value of seed is from sample(1:1000,1).
nonlinear	if TURE, Multiple Additive Regression Trees (MART) will be used to fit the final full model in estimating the outcome. The default value of nonlinear is FALSE, in which case, a generalized linear model will be used to fit the final full model.
df	if nonlinear is TURE, natural cubic spline will be used to fit the relationship between the predictor and each mediator. The df is the degree of freedom in the ns() function, the default is 1.
nu	set the parameter "interaction.depth" in gbm function if MART is to be used, by default, nu=0.001. See also help(gbm.fit).
D	set the parameter "shrinkage" in gbm function if MART is to be used, by default, D=3. See also help(gbm.fit).
distn	the assumed distribution of the outcome if MART is used for final full model. The default value of distn is "gaussian" for coninuous y, and "bernoulli" for binary y.
family1	define the conditional distribution of y given x, and the linkage function that links the mean of y with the system component if generalized linear model is used as the final full model. The default value of family1 is gaussian(link="identity") for continuous y, and binomial(link = "logit") for binary y.
n2	the number of times of bootstrap resampling. The default value is n2=50.

weight	the weight for observations.
refy	if y is binary, the reference group of y.
x.new	of the same format as x, with a new set of predictor and corresponding covariates on which to calculate the mediation effects.
binpred	if TRUE, the predictor is binary.
type	the type of prediction when y is class Surv. Is "risk" if not specified.

Details

The mediators are not tested in this function. data.org should be used for the tests and data organizing, and then the resulting list from data.org can be used directly to define the arguments in this function. boot.med considers all variables in x as mediators or covariates in the final model and all variables identified by contm, binm, catm, or jointm as mediators.

Value

Returns an mma object.

estimation	list the estimation of ie (indirect effect), te (total effect), and de (direct effect from the predictor) separately.
bootsresults	a list where the first item, ie, is a matrix of n2 rows where each column gives the estimated indirect effect from the corresponding mediator (identified by the column name) from the n2 bootstrap samples; the second item, te, is a vector of estimated total effects from the bootstrap sample; and the 3rd item, de, is a vector of estimated direct effect of the predictor from the bootstrap sample.
model	a list where the first item, MART, is T if mart is fitted for the final model; the second item, Survival, is T if a survival model is fitted; the third item, type, is the type of prediction when a survival model is fitted; the fourth item, model, is the fitted final full model where y is the outcome and all predictor, covariates, and mediators are the explanatory variables; and the fourth item, best.iter is the number of best iterations if MART is used to fit the final model.
data	a list that contains all the used data: x=x, y=y, dirx=dirx, binm=binm, contm=contm, catm=catm, jointm=jointm, binpred=F.
boot.detail	a list that contains the mediation effects on each row of new.x: new.x=new.x, te1, de1, ie1.

Author(s)

Qingzhao Yu <qyu@lsuhsc.edu>

References

Yu, Q., Fan, Y., and Wu, X. (2014) <doi:10.4172/2155-6180.1000189>. "General Multiple Mediation Analysis With an Application to Explore Racial Disparity in Breast Cancer Survival," Journal of Biometrics & Biostatistics,5(2): 189.

See Also

"[med](#)" just estimate the mediation effects.

Examples

```
data("weight_behavior")
##binary x
#binary y
x=weight_behavior[,2:14]
y=weight_behavior[,15]
data.bin<-data.org(x,y,pred=2,contmed=c(8:10,12:13),binmed=c(7,11),
          binref=c(1,1),catmed=6,catref=1,predref="M",alpha=0.4,alpha2=0.4)
temp1<-boot.med(data=data.bin,n=2,n2=4,seed=1)
temp2<-boot.med(data=data.bin,n=2,n2=4,nu=0.05,seed=1,nonlinear=TRUE)

#continuous y
x=weight_behavior[,2:14]
y=weight_behavior[,1]
data.cont<-data.org(x,y,pred=2,contmed=c(8:10,12:13),binmed=c(7,11),
                     binref=c(1,1),catmed=6,catref=1, predref="M",
                     family1=gaussian(link="identity"),alpha=0.4,alpha2=0.4)

temp3<-boot.med(data=data.cont,n=2,n2=4,seed=1)
temp4<-boot.med(data=data.cont,n=2,n2=4,nu=0.05,seed=1, nonlinear=TRUE)

##continuous x
#binary y
x=weight_behavior[,2:14]
y=weight_behavior[,15]
data.contx<-data.org(x,y,pred=1,contmed=c(8:10,12:13),binmed=c(7,11),
                      binref=c(1,1),catmed=6,catref=1,alpha=0.4,alpha2=0.4)

temp5<-boot.med(data=data.contx,seed=1,n=1,n2=2)
plot(temp5,vari="exercises",xlim=c(0,30))
temp6<-boot.med(data=data.contx,seed=1,n=1,refy=0,nonlinear=T,n2=2)

#continuous y
x=weight_behavior[,2:14]
y=weight_behavior[,1]
data.contx<-data.org(x,y,pred=1,contmed=c(8:10,12:13),binmed=c(7,11),
                      binref=c(1,1),catmed=6,catref=1,alpha=0.4,alpha2=0.4)

temp7<-boot.med(data=data.contx,seed=1,n=1,n2=2)
temp8<-boot.med(data=data.contx,nonlinear=T,seed=1,n=1,n2=2)

##Surv class outcome (survival analysis)
data(cgd0)        #a dataset in the survival package
cgd1<-cgd0
x=cgd1[,c(4:12)]
status<-ifelse(is.na(cgd1$etime1),0,1)
y=Surv(cgd1$futime,status)
#for continuous predictor
```

```
data.surv.contx<-data.org(x,y,pred=3,mediator=(1:ncol(x))[-3],
                          alpha=0.5,alpha2=0.5)
temp9.contx<-boot.med(data=data.surv.contx,seed=1,n=1,n2=2, type="lp")
summary(temp9.contx)

temp10.contx<-boot.med(data=data.surv.contx,nonlinear=TRUE,seed=1,n=1,n2=2)
summary(temp10.contx)

#for binary predictor
data.surv.binx<-data.org(x,y,pred=1,mediator=(1:ncol(x))[-1],
                         alpha=0.4,alpha2=0.4)
temp9.binx<-boot.med(data=data.surv.binx,seed=1,n=1,n2=2, type="lp")
summary(temp9.binx)

temp10.binx<-boot.med(data=data.surv.binx,nonlinear=TRUE,seed=1,n=1,n2=2)
summary(temp10.binx)
```

---

data.org	*Data Organization and Identify Potential Mediators*

Description

Do a preliminary data analysis to identify potential mediators and covariates. Each variable listed in jointm is forced in the final estimation model as a mediator. Also organize the data into a format that can be directly used for the mediation analysis functions.

Usage

```
data.org(x,y,pred,mediator=NULL,contmed=NULL,binmed=NULL,binref=NULL,catmed=NULL,
                  catref=NULL,jointm=NULL,refy=NULL, family1=NULL,
                  predref=NULL,alpha=0.1,alpha2=0.1,testtype=1)
```

Arguments

x	a data frame contains the predictor, all potential mediators and covariates
y	the vector of outcome variable.The outcome can be binary, continuous, or of "Surv" class (see survival package for help).
pred	the column number that locates the predictor in x or the predictor name.
mediator	the list of mediators (column numbers in x or by variable names). The mediators to be checked can be identified by "contmed", "binmed" and "catmed", or by this argument, "mediator", where binary and categorical mediators in x are identified as factors or characters, the reference group is the first level of the factor or factorized character. if a mediator has only two unique values, the mediator is identified as binary. If the reference groups need to be changed, the binary or categorical mediators can be listed in binmed or catmed, and the corresponding reference group in binref or catref.
contmed	a vector of variable names or column numbers that locate the potential continuous mediators in x.

binmed	a vector of column numbers that locate the potential binary mediators in x.
binref	the defined reference groups of the binary potential mediators in binmed. The first levels of the mediators if is null.
catmed	a vector of variable names or column numbers that locate the potential categorical mediators in x. The first levels of the mediators if is null.
catref	the defined reference groups of the categorical potential mediators in catmed.
jointm	a list that identifies the mediators that need to be considered jointly, where the first item indicates the number of groups of mediators to be considered jointly, and each of the following items identifies the variable names or column numbers of the mediators in x for each group of joint mediators.
refy	if y is binary, the reference group of y. The default is the first level of as.factor(y). y for the reference group is assigned as 0.
family1	define the conditional distribution of y given x, and the linkage function that links the mean of y with the system component in generalized linear model. The default value of family1 is binomial(link = "logit") for binary y, and gaussian(link="identity") for continuous y.
predref	if the predictor is binary, identify the reference group of the binary predictor. The default is the first level of the predictor. The value of the predictor is 0 for the reference grop.
alpha	the significance level at which to test if the potential mediators (identified by contmed, binmed, and catmed) can be used as a covariate or mediator in estimating y when all variables in x are included in the model. The default value is alpha=0.1
alpha2	the significant level at which to test if a potential mediator is related with the predictor. The default value is alpha2=0.1.
testtype	if the testtype is 1 (by default), covariates/mediators are identified using full model; if the testtype is 2, covariates/mediators are tested one by one in models with the predictor only.

Value

data.org returns a list with the organized data and identifiers of the potential mediators in the organized data set.

x	the organized data frame that include all potential mediators and covariates that should be used to estimate the outcome.
dirx	the column number of x that locates the predictor.
contm	the column numbers of x that locate the potential continuous mediators.
binm	when the predictor is continuous, binm gives the column numbers of x that locate the potential binary mediators.
catm	when the predictor is binary, catm gives the column numbers of x that locate the potential binary or categorical mediators; when the predictor is continuous, catm gives a list where the first item is the number of potential categorical mediators, and the following items give the column numbers of each binarized categorical mediator in x.

jointm	a list where the first item is the number of groups of joint mediators, and each of the following items identifies the column numbers of the mediators in the newly organized x for each group of joint mediators.
y	the vector of outcome.
fullmodel	the full linear model fitted with all potential mediators and covariates.
rela	p-values of tests on the realtionship between the predictor and each potiential mediator.
surv	True if y is of class Surv.
P1	If testtype=2, P1 gives the p-value of the corresponding variables in predicting the outcome when only the variable and predictor are covariates in the model.

Note

All other variables in x but not identified by mediator, contmed, binmed, or catmed are forced in the final model as covariates. Compared with data.org, joint mediators are considered in this function. Every variable in the jointm should be listed in contmed, binmed, or catmed, and these variables are forced to be included as mediators for further mediation analysis. A variables can be included in more than one groups of joint mediators in jointm.

Author(s)

Qingzhao Yu <qyu@lsuhsc.edu>

References

Baron, R.M., and Kenny, D.A. (1986) <doi:10.1037/0022-3514.51.6.1173>. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J. Pers Soc Psychol, 51(6), 1173-1182.

See Also

"data.org" that does not consider joint mediators, which can be added freely in the mediation analysis functions later.

Examples

```
data("weight_behavior")
#binary x
 #binary y
 x=weight_behavior[,2:14]
 y=weight_behavior[,15]
 data.b.b.2.1<-data.org(x,y,mediator=6:13,jointm=list(n=1,j1=c(6,8,10)),
                        pred=2,predref="M", alpha=0.4,alpha2=0.4)
 summary(data.b.b.2.1)
 #Or you can specify the potential mediators and change the reference
 #group for binary or categorical mediators. In the following code,
 #potential continuous mediators are columns 8,9,10,12, and 13 of x,
 #binary mediators are columns 7 and 11, and categorical mediator is
 #column 6 of x with 1 to be the reference group for all categorical
```

```
#and binary mediators.
 data.b.b.2<-data.org(x,y,pred=2,contmed=c(8:10,12:13),binmed=c(7,11),
  binref=c(1,1),catmed=6,catref=1,jointm=list(n=1,j1=c(6,8,10)),
  predref="M",alpha=0.4,alpha2=0.4)
 summary(data.b.b.2)
#use the mediator argument instead of contmet, binmed and catmed

#continuous y
x=weight_behavior[,2:14]
y=weight_behavior[,1]
data.b.c.2<-data.org(x,y,pred=2,mediator=6:13,jointm=list(n=1,j1=8:10),
  predref="M",alpha=0.4,alpha2=0.4)
summary(data.b.c.2)

#continuous x
#binary y
x=weight_behavior[,2:14]
y=weight_behavior[,15]
data.c.b.2<-data.org(x,y,pred=1,mediator=6:13,catref=1,
       jointm=list(n=2,j1=8:10,j2=c(6,8)),alpha=0.4,alpha2=0.4)
summary(data.c.b.2)

#continuous y
x=weight_behavior[,2:14]
y=weight_behavior[,1]
data.c.c.2<-data.org(x,y,pred=1,contmed=c(8:10,12:13),binmed=c(7,11),
  binref=c(1,1),catmed=6,catref=1,jointm=list(n=2,j1=8:10,j2=c(6,8)),
  alpha=0.4,alpha2=0.4)
summary(data.c.c.2)

#Surv class outcome (survival analysis)
data(cgd0)       #a dataset in the survival package
cgd1<-cgd0
x=cgd1[,c(4:12)]
status<-ifelse(is.na(cgd1$etime1),0,1)
y=Surv(cgd1$futime,status)
#for continuous predictor
#all other variables are considered as potential mediator
data.surv.contx<-data.org(x,y,pred=3,mediator=(1:ncol(x))[-3],
                  alpha=0.5,alpha2=0.5)
summary(data.surv.contx)

#for binary predictor
data.surv.binx<-data.org(x,y,pred=1,mediator=(1:ncol(x))[-1],
                  alpha=0.4,alpha2=0.4)
summary(data.surv.binx)
```

med   *Mediation Analysis with Binary or Continuous Predictor*

Description

To estimate the mediation effects when the predictor is binary or continuous.

Usage

```
med(data, x=data$x, y=data$y, dirx=data$dirx, binm=data$binm,
    contm = data$contm,catm = data$catm, jointm = data$jointm,
    allm = c(contm, catm), margin=1,n=20,seed=sample(1:1000,1),
    nonlinear=F, df=1, nu=0.001,D=3,distn=NULL,family1=NULL,
    refy=0,binpred=data$binpred,x.new=x,type=NULL)
```

Arguments

data	the list of result from data.org that organize the covariates, mediators, predictor and outcome. If data is FALSE, then need to set x1, y1, dirx, contm, catm, and jointm.
x	a data frame contains the predictor, all mediators and covariates. Need to set up only when data is FALSE.
y	the vector of outcome variable. Need to set up only when data is FALSE.
dirx	the predictor name or the column number of x that locates the predictor. Need to set up only when data is FALSE.
binm	the variable names or the column number of x that locates the binary mediators. Need to set up only when data is FALSE.
contm	the variable names or the column numbers of x that locate the potential continuous mediators. Need to set up only when data is FALSE.
catm	categorical mediators should be binarized and be presented as a list, where the first item is the number of categorical variables and the following items are the names or the column numbers of each binarized categorical variable in x. data.org organizes the categorical mediators in this format after they pass the mediator tests. Need to set up only when data is FALSE.
jointm	a list where the first item is the number of groups of joint mediators to be considered, and each of the following items identifies the names or the column numbers of the mediators in x for each group of joint mediators. Need to set up only when data is FALSE.
allm	the column numbers of all mediators. Need to set up only when data is FALSE. The default value of allm is c(contm,catm).
margin	the change in predictor when calculating the mediation effects, see Yu et al. (2014).
n	the time of resampling in calculating the indirect effects, default is n=20, see Yu et al. (2014).
seed	set seed to make the calculation repeatable. The default value of seed is from sample(1:1000,1).
nonlinear	if TURE, Multiple Additive Regression Trees (MART) will be used to fit the final full model in estimating the outcome. The default value of nonlinear is FALSE, in which case, a generalized linear model will be used to fit the final full model.

df	if nonlinear is TURE, natural cubic spline will be used to fit the relationship between the predictor and each mediator. The df is the degree of freedom in the ns() function, the default is 1.
nu	set the parameter "interaction.depth" in gbm function if MART is to be used, by default, nu=0.001. See also help(gbm.fit).
D	set the parameter "shrinkage" in gbm function if MART is to be used, by default, D=3. See also help(gbm.fit).
distn	the assumed distribution of the outcome if MART is used for final full model. The default value of distn is "gaussian" for continuous y and "bernoulli" for binary y.
family1	define the conditional distribution of y given x, and the linkage function that links the mean of y with the system component if generalized linear model is used as the final full model. The default value of family1 is gaussian(link="identity") for contiuous y, and binomial(link = "logit") for binary y.
refy	if y is binary, the reference group of y.
binpred	if TRUE, the predict variable is binary.
x.new	of the same format as x, with a new set of predictor and corresponding covariates on which to calculate the mediation effects. For continuous predictor only.
type	the type of prediction when y is class Surv. By default, type is "risk".

Details

The mediators are not tested in this function. data.org should be used first for the tests and data organizing, and then the resulting list from data.org can be used directly to define the arguments in this function. med considers all variables in x as mediators or covariates in the final model and all variables identified by contm, binm, catm, or jointm as mediators.

Value

The result is an med object with:

denm	a matrix where each column gives the estimated direct effect not from the corresponding mediator (identified by the column name), see Yu et al. (2014) for the definition, and each row corresponding to the results from one resampling for binary predictor or the results on a row of x.new for continuous predictor.
ie	a matrix where each column gives the estimated indirect effect from the corresponding mediator (identified by the column name) and each row corresponding to the results from one resampling for binary predictor or the results on a row of x.new for continuous predictor.
te	a vector of the estimated total effect on x.new.
model	a list, where the first item, MART, is TRUE if a mart is fitted as the final model; the second item, Survial is T if a survival model is fit; the third item, type, is the type of prediction when a survival model is fitted; the fourth item, full.model, is the fitted final full model where y is the outcome and all predictor, covariates, and mediators are the explanatory variables; and the fifth item, best.iter, is the number of best iterations if MART is used to fit the final model, is NULL if the final model is a generalized linear model.

Author(s)

Qingzhao Yu <qyu@lsuhsc.edu>

References

J.H. Friedman, T. Hastie, R. Tibshirani (2000) <doi:10.1214/aos/1016120463>. "Additive Logistic Regression: a Statistical View of Boosting," Annals of Statistics 28(2):337-374.

J.H. Friedman (2001) <doi:10.1214/aos/1013203451>. "Greedy Function Approximation: A Gradient Boosting Machine," Annals of Statistics 29(5):1189-1232.

Yu, Q., Fan, Y., and Wu, X. (2014) <doi:10.4172/2155-6180.1000189>. "General Multiple Mediation Analysis With an Application to Explore Racial Disparity in Breast Cancer Survival," Journal of Biometrics & Biostatistics,5(2): 189.

See Also

"`boot.med`*"* to make inferences on the estimated mediation effects using bootstrap method.

Examples

```
data("weight_behavior")
##binary x
#binary y
x=weight_behavior[,2:14]
y=weight_behavior[,15]
data.bin<-data.org(x,y,pred=2,contmed=c(8:10,12:13),binmed=c(7,11),
  binref=c(1,1),catmed=6,catref=1,predref="M",alpha=0.4,alpha2=0.4)
temp1<-med(data=data.bin,n=2,seed=1)
temp2<-med(data=data.bin,n=2,seed=1,nonlinear=TRUE)

#continuous y
x=weight_behavior[,2:14]
y=weight_behavior[,1]
data.cont<-data.org(x,y,pred=2,contmed=c(8:10,12:13),binmed=c(7,11),
                    binref=c(1,1),catmed=6,catref=1,
                    predref="M",alpha=0.4,alpha2=0.4)
temp3<-med(data=data.cont,n=2,seed=1)
temp4<-med(data=data.cont,n=2,seed=1,nonlinear=TRUE)

##continuous x
#binary y
x=weight_behavior[,2:14]
y=weight_behavior[,15]
data.contx<-data.org(x,y,pred=1,mediator=5:11,alpha=0.4,alpha2=0.4)
temp5<-med(data=data.contx,seed=1,n=2)
temp6<-med(data=data.contx,seed=1,n=2,nonlinear=TRUE,nu=0.05)

#continuous y
x=weight_behavior[,2:14]
y=weight_behavior[,1]
data.contx<-data.org(x,y,pred=1,contmed=c(12:13),binmed=c(7,11),
                     binref=c(1,1),catmed=6,catref=1,
```

```
                          alpha=0.4,alpha2=0.4)

temp7<-med(data=data.contx,seed=1,n=2)
temp8<-med(data=data.contx,seed=1,n=2,nonlinear=TRUE,nu=0.05)

##Surv class outcome (survival analysis)
data(cgd0)        #a dataset in the survival package
cgd1<-cgd0
x=cgd1[,c(4:12)]
status<-ifelse(is.na(cgd1$etime1),0,1)
y=Surv(cgd1$futime,status)
#for continuous predictor
data.surv.contx<-data.org(x,y,pred=3,mediator=(1:ncol(x))[-3],
                          alpha=0.5,alpha2=0.5)
temp9.contx<-med(data=data.surv.contx,seed=1,n=2,type="lp")
#close to mart results when use type="lp"
temp9.contx
temp10.contx<-med(data=data.surv.contx,n=2,seed=1,nonlinear=TRUE)
#results in the linear part unit
temp10.contx

#for binary predictor

data.surv.binx<-data.org(x,y,pred=1,mediator=(1:ncol(x))[-1],
                     alpha=0.4,alpha2=0.4)
temp9.binx<-med(data=data.surv.binx,seed=1,n=2,type="lp")
temp9.binx
temp10.binx<-med(data=data.surv.binx,n=2,seed=1,nonlinear=TRUE)
temp10.binx
```

---

mma   *Multiple Mediation Analysis*

Description

Test for mediators and do statistical inferences on the identified mediation effects.

Usage

```
mma(x,y,pred,mediator=NULL, contmed=NULL,binmed=NULL,binref=NULL,
    catmed=NULL,catref=NULL,jointm=NULL,refy=NULL,predref=NULL,
    alpha=0.1,alpha2=0.1, margin=1, n=20,seed=sample(1:1000,1),
    nonlinear=F,df=1,nu=0.001,D=3,distn=NULL,family1=NULL,n2=50,
    weight=rep(1,nrow(x)),x.new=NULL,type=NULL)
```

Arguments

x	a data frame contains the predictor, all potential mediators and covariates.
y	the vector of outcome variable.

pred	the column number that locates the predictor in x.
mediator	the list of mediators (column numbers in x or by variable names). The mediators to be checked can be identified by "contmed", "binmed" and "catmed", or by this argument, "mediator", where binary and categorical mediators in x are identified by factors, the reference group is the first level of the factor.
contmed	a vector of column numbers that locate the potential continuous mediators in x.
binmed	a vector of column numbers that locate the potential binary mediators in x.
binref	the defined reference groups of the binary potential mediators in binmed.
catmed	a vector of column numbers that locate the potential categorical mediators in x.
catref	the defined reference groups of the categorical potential mediators in catmed.
jointm	a list that identifies the mediators that need to be considered jointly, where the first item indicates the number of groups of mediators to be considered jointly, and each of the following items identifies the column numbers of the mediators in x for each group of joint mediators.
refy	if y is binary, the reference group of y.By default, the reference group will be the first level of as.factor(y).
predref	if predictor is binary, identify the reference group of the binary predictor. By default, the reference group will be the first level of the predictor.
alpha	the significance level at which to test if the potential mediators (identified by contmed, binmed, and catmed) can be used as a covariate or mediator in estimating y when all variables in x are included in the model. The default value is alpha=0.1
alpha2	the significant level at which to test if a potential mediator is related with the predictor. The default value is alpha2=0.1.
margin	if binpred is FALSE, define the change in predictor when calculating the mediation effects, see Yu et al. (2014).
n	the time of resampling in calculating the indirect effects, default is n=20, see Yu et al. (2014).
seed	set seed to make the calculation repeatable. The default value of seed is from sample(1:1000,1).
nonlinear	if TURE, Multiple Additive Regression Trees (MART) will be used to fit the final full model in estimating the outcome. The default value of nonlinear is FALSE, in which case, a generalized linear model will be used to fit the final full model.
df	if nonlinear is TURE, natural cubic spline will be used to fit the relationship between the countinuous predictor and each mediator. The df is the degree of freedom in the ns() function, the default is 1.
nu	set the parameter "interaction.depth" in gbm function if MART is to be used, by default, nu=0.001. See also help(gbm.fit).
D	set the parameter "shrinkage" in gbm function if MART is to be used, by default, D=3. See also help(gbm.fit).
distn	the assumed distribution of the outcome if MART is used for final full model. The default value of distn is "bernoulli" for binary y, and "gaussian" for continuous y.

family1	define the conditional distribution of y given x, and the linkage function that links the mean of y with the system component if generalized linear model is used as the final full model. The default value of family1 is binomial(link = "logit") for binary y, gaussian(link="identity") for continuous y.
n2	the number of times of bootstrap resampling. The default value is n2=50.
weight	the weight for each observation.
x.new	for continuous predictor only. of the same format as x (after data.org), with a new set of predictor and corresponding covariates on which to calculate the mediation effects. If is NULL, the mediation effects will be calculated based on the original data set.
type	the type of prediction when y is class Surv. Is "risk" if not specified.

Details

mma first tests if the potential mediators defined by binm, contm, and catm should be treated as mediators or covariates (if none, the variable will be deleted from further analysis). All variables identified by jointm are treated as mediators. All other variables in x that are not tested are treated as covariates. Then mma does the mediation effects estimation and inference on the selected variables.

Value

Returns an mma object.

estimation	list the estimation of ie (indirect effect), te (total effect), and de (direct effect from the predictor) separately.
bootsresults	a list where the first item, ie, is a matrix of n2 rows where each column gives the estimated indirect effect from the corresponding mediator (identified by the column name) from the n2 bootstrap samples; the second item, te, is a vector of estimated total effects from the bootstrap sample; and the 3rd item, de, is a vector of estimated direct effect of the predictor from the bootstrap sample.
model	a list where the first item, MART, is T if mart is fitted for the final model; the second item, Survival, is T if a survival model is fitted; the third item, type, is the type of prediction when a survival model is fitted; the fourth item, model, is the fitted final full model where y is the outcome and all predictor, covariates, and mediators are the explanatory variables; and the fourth item, best.iter is the number of best iterations if MART is used to fit the final model.
data	a list that contains all the used data: x=x, y=y, dirx=dirx, binm=binm, contm=contm, catm=catm, jointm=jointm, binpred=F.
boot.detail	for continuous predictor only: a list that contains the mediation effects on each row of new.x: new.x=new.x, te1, de1, ie1.

Author(s)

Qingzhao Yu <qyu@lsuhsc.edu>

References

Baron, R.M., and Kenny, D.A. (1986) <doi:10.1037/0022-3514.51.6.1173>. "The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations," J. Pers Soc Psychol, 51(6), 1173-1182.

J.H. Friedman, T. Hastie, R. Tibshirani (2000) <doi:10.1214/aos/1016120463>. "Additive Logistic Regression: a Statistical View of Boosting," Annals of Statistics 28(2):337-374.

J.H. Friedman (2001) <doi:10.1214/aos/1013203451>. "Greedy Function Approximation: A Gradient Boosting Machine," Annals of Statistics 29(5):1189-1232.

Yu, Q., Fan, Y., and Wu, X. (2014) <doi:10.4172/2155-6180.1000189>. "General Multiple Mediation Analysis With an Application to Explore Racial Disparity in Breast Cancer Survival," Journal of Biometrics & Biostatistics,5(2): 189.

See Also

"data.org" is for mediator tests, and *"med"* , and *"boot.med"* for mediation analysis and inferences.

Examples

```
data("weight_behavior")
 #binary x
 #binary y
 x=weight_behavior[,2:14]
 y=weight_behavior[,15]
 temp.b.b.glm<-mma(x,y,pred=2,mediator=6:13,predref="M",alpha=0.4,
                   alpha2=0.4,n=2,n2=2)


 temp.b.b.mart<-mma(x,y,pred=2,mediator=6:13,predref="M",alpha=0.4,
                     alpha2=0.4,nonlinear=TRUE,n=2,n2=5)
 #continuous y
 x=weight_behavior[,2:14]
 y=weight_behavior[,1]
 temp.b.c.glm<-mma(x,y,pred=2,mediator=c(8:10,12:13,7,11,6),
   jointm=list(n=1,j1=8:10), seed=1, n=2, predref="M",alpha=0.4,
   alpha2=0.4,n2=20)
 temp.b.c.mart<-mma(x,y,pred=2,mediator=6:13,jointm=list(n=1,j1=8:10,
                    j2=12:13), predref="M",alpha=0.4,alpha2=0.4,
                    n=2,seed=1,nonlinear=TRUE,n2=20)

##Surv class outcome (survival analysis)
data(cgd0)      #a dataset in the survival package
cgd1<-cgd0
x=cgd1[,c(4:12)]
status<-ifelse(is.na(cgd1$etime1),0,1)
y=Surv(cgd1$futime,status)
#for continuous predictor
temp.cox.contx<-mma(x,y,pred=3,mediator=(1:ncol(x))[-3],
             alpha=0.5,alpha2=0.5,seed=1,n=1,n2=2,type="lp")
summary(temp.cox.contx)
temp.surv.mart.contx<-mma(x,y,pred=3,mediator=(1:ncol(x))[-3], n=1,
```

```
                           alpha=0.5,alpha2=0.5,nonlinear=TRUE,seed=1,n2=2)
summary(temp.surv.mart.contx)
plot(temp.surv.mart.contx,vari="steroids")
plot(temp.cox.contx,vari="steroids")

#for binary predictor
temp.cox.binx<-mma(x,y,pred=1,mediator=(1:ncol(x))[-1],
                    alpha=0.4,alpha2=0.4,seed=1,n=1,n2=2,type="lp")
summary(temp.cox.binx)
temp.surv.mart.binx<-mma(x,y,pred=1,mediator=(1:ncol(x))[-1], seed=1,
                          alpha=0.4,alpha2=0.4,nonlinear=TRUE,n=1,n2=2)
summary(temp.surv.mart.binx)
plot(temp.surv.mart.binx,vari="hos.cat")
plot(temp.cox.binx,vari="hos.cat")
```

---

plot.mma   *Relative effects plot of the fitted mma object*

Description

Plot the marginal effect of the selected variable on the outcome, and the marginal effect of the predictor on the selected variable.

Usage

```
## S3 method for class 'mma'
plot(x,...,vari,xlim=range(x$data$x[,vari],na.rm=T),alpha=0.95,
quantile=F)
```

Arguments

x	a mma object created initially call to mma, boot.met.binx, or boot.met.contx.
vari	an indices or the name of the variable to plot.
xlim	the range of the variable to be plotted.
alpha	for continuous predictor only, to draw the alpha confidence interval of the indirect effect.
quantile	for continuous predictor only, if true to draw the alpha confidence interval of the indirect effect based on quantile, otherwise, based on the normal approximation.
...	other arguments passed to the plot function.

Details

plot.mma plots the marginal effect of the selected variable on the outcome, and the marginal effect of the predictor on the selected variable. If the predictor is binary, draw a histogram or boxplot of the marginal density of the variable at each different value of the predictor.

Author(s)

Qingzhao Yu <qyu@lsuhsc.edu>

References

Yu, Q., Fan, Y., and Wu, X. (2014) <doi:10.4172/2155-6180.1000189>. "General Multiple Mediation Analysis With an Application to Explore Racial Disparity in Breast Cancer Survival," Journal of Biometrics & Biostatistics,5(2): 189.

See Also

"mma",*"boot.med"*

Examples

```
data("weight_behavior")
x=weight_behavior[,2:14]
y=weight_behavior[,15]
temp.b.b.glm<-mma(x,y,pred=2,contmed=c(8:10,12:13),binmed=c(7,11),
   binref=c(1,1),catmed=6,catref=1,predref="M",alpha=0.4,alpha2=0.4,
   jointm=NULL,margin=1, n=2,seed=1,nonlinear=FALSE,nu=0.001,D=3,n2=2)
plot(temp.b.b.glm,vari="exercises",xlim=c(0,50))
plot(temp.b.b.glm,vari="sports")
```

---

print.med	*Print an med object*

Description

Print the estimation of mediation effects from an med object: from functions med.

Usage

```
## S3 method for class 'med'
print(x,...,digit=4)
```

Arguments

x	a med object created initially call to med.
...	other arguments passed to the print function.
digit	the number of digits to keep at printing.

Author(s)

Qingzhao Yu <qyu@lsuhsc.edu>

References

Yu, Q., Fan, Y., and Wu, X. (2014) <doi:10.4172/2155-6180.1000189>. "General Multiple Media-
tion Analysis With an Application to Explore Racial Disparity in Breast Cancer Survival," Journal
of Biometrics & Biostatistics,5(2): 189.

See Also

["med"](#)

Examples

```
data("weight_behavior")
##binary x
#binary y
x=weight_behavior[,2:14]
y=weight_behavior[,15]
data.bin<-data.org(x,y,pred=2,contmed=c(8:10,12:13),binmed=c(7,11),
  binref=c(1,1),catmed=6,catref=1,predref="M",alpha=0.4,alpha2=0.4)
temp1<-med(data=data.bin,n=2,seed=1)
temp2<-med(data=data.bin,n=2,seed=1,nonlinear=TRUE)
temp1
temp2
```

---

  print.mma   *Print a mma object*

Description

Print the estimation of mediation effects from an mma object.

Usage

```
## S3 method for class 'mma'
print(x,...)
```

Arguments

x	a mma object created initially call to mma, boot.med.binx, or boot.med.contx.
...	other arguments passed to the print function.

Value

Return a LIST

ie	estimation of the indirect effects from the mma object.
te	estimation of the total effect from the mma object.
de	estimation of the direct effect from the mma object.

Author(s)

Qingzhao Yu <qyu@lsuhsc.edu>

References

Yu, Q., Fan, Y., and Wu, X. (2014) <doi:10.4172/2155-6180.1000189>. "General Multiple Mediation Analysis With an Application to Explore Racial Disparity in Breast Cancer Survival," Journal of Biometrics & Biostatistics,5(2): 189.

See Also

"mma",*"boot.med"*

Examples

```
data("weight_behavior")
 x=weight_behavior[,2:14]
 y=weight_behavior[,15]
 temp.b.b.glm<-mma(x,y,pred=2,contmed=c(8:10,12:13),binmed=c(7,11),
   binref=c(1,1),catmed=6,catref=1,predref="M",alpha=0.4,alpha2=0.4,
   jointm=NULL,margin=1, n=2,seed=sample(1:1000,1),nonlinear=FALSE,
   nu=0.001,D=3,n2=2)
 print(temp.b.b.glm)
```

---

summary.med_iden	*Summary method for class "med_iden".*

Description

Compute the estimations, standard deviations and confidence intervals of the mediation effects.

Usage

```
## S3 method for class 'med_iden'
summary(object,...)
## S3 method for class 'summary.med_iden'
print(x,...)
```

Arguments

object	a med_iden object created initially call to data.org.
x	a summary.med_iden object created initially call to summary.med_iden
...	other arguments passed to the print function.

Details

summary.med_iden gives a list of identified mediators, covariates and the test results.

Value

The function summary.med_iden return a list of covariates and mediators as identified by a series tests.

mediator	variable names of the identified mediators, either siginicant in both full model and in relate to the predictor, or being a member of the pre-identified joint mediators.
covariate	variable names of covariates: being significant in the full model but not significantly relate with the predictor.
tests	statistical test results.
results	the original object.

Author(s)

Qingzhao Yu <qyu@lsuhsc.edu>

References

Yu, Q., Fan, Y., and Wu, X. (2014) <doi:10.4172/2155-6180.1000189>. "General Multiple Mediation Analysis With an Application to Explore Racial Disparity in Breast Cancer Survival," Journal of Biometrics & Biostatistics,5(2): 189.

See Also

"mma",*"boot.med"*

Examples

```
data("weight_behavior")
x=weight_behavior[,2:14]
y=weight_behavior[,15]
data.b.b.2<-data.org(x,y,pred=2,contmed=c(8:10,12:13),binmed=c(7,11),
   binref=c(1,1),catmed=6,catref=1,jointm=list(n=1,j1=c(6,8,10)),
   predref="M",alpha=0.4,alpha2=0.4)
summary(data.b.b.2)
```

---

summary.mma	*Summary of an mma project*

Description

Compute the estimations, standard deviations and confidence intervals of the mediation effects.

Usage

```
## S3 method for class 'mma'
summary(object,..., alpha=0.05, plot=TRUE, RE=FALSE,quant=T)
## S3 method for class 'summary.mma'
print(x,...)
```

Arguments

`object`	a mma object created initially call to mma, boot.met.binx, or boot.met.contx.
`x`	a summary.mma object created initially call to summary.mma.
`...`	other arguments passed to the print function.
`alpha`	the alpha level for confidence interval.
`plot`	default is TRUE, if ture, draw a barplot of the relative mediation effects with confidence intervals.
`RE`	default is FALSE, if ture, show the inferences on relative mediation effects.
`quant`	default is TRUE, if ture, draw the confidence intervals of relative effects using quantile.

Details

summary.mma gives a list of the estimations and summary statistics based on the bootstrap results. If plot=T, draw a barplot of the relative effects of the direct effect of the predictor and indirect effects of the mediators. Relative effects is defined as the (in)direct effect divided by the total effect. The plot is arranged in order from the largest to the smallest relative effect.

Value

The function summary.mma return a list. The first item, results, is the list for mediation effects, and the second item, re, is the list for relative effects. Under them, results have the items ie, te and de; re has the items ie and de. In each of the items, est is the estimation of the corresponding (relative) mediation effects based on the whole data, mean is the average estimated (relative) effects from the bootstrap samples, sd is the standard deviation of the estimates from the bootstrap sample. upbd and lwbd are the upper and lower bound of the confidence interval of the estimation using parametric method from the bootstrap sample, upbd_q and lwbd_q are the corresponding quantiles of the estimation from the bootstrap sample.

`ie`	a matrix of statistics inference on the (relative) indirect effects from the mma object.
`te`	statistics inference on the total effects from the mma object.
`de`	statistics inference on the (relative) direct effects from the mma object.

If plot=T, draw a barplot of the relative mediation effects.

Author(s)

Qingzhao Yu <qyu@lsuhsc.edu>

References

Yu, Q., Fan, Y., and Wu, X. (2014) <doi:10.4172/2155-6180.1000189>. "General Multiple Mediation Analysis With an Application to Explore Racial Disparity in Breast Cancer Survival," Journal of Biometrics & Biostatistics,5(2): 189.

See Also

"mma","boot.med"

Examples

```
data("weight_behavior")
 x=weight_behavior[,2:14]
 y=weight_behavior[,15]
 temp.b.b.glm<-mma(x,y,pred=2,contmed=c(8:10,12:13),binmed=c(7,11),
   binref=c(1,1),catmed=6,catref=1,predref="M",alpha=0.4,alpha2=0.4,
   jointm=NULL,margin=1, n=2,seed=sample(1:1000,1),nonlinear=FALSE,
   nu=0.001,D=3,n2=2)
 summary(temp.b.b.glm)
```

---

weight_behavior   *Weight_Behavior Data Set*

Description

This database was obtained from the Louisiana State University Health Sciences Center, New Orleans, by Dr. Richard Scribner. He explored the relationship between BMI and kids behavior through a survey at children, teachers and parents in Grenada in 2014. This data set includes 691 observations and 15 variables.

Usage

```
data(weight_behavior)
```

Format

The data set contains the following variables:

bmi - body mass index, calculated by weight(kg)/height(cm)^2, numeric

age - children's age in years at the time of survey, numeric

sex - sex of the children, factor

race - race of the children, factor

numpeople - number of people in family, numeric

car - the number of cars in family, numeric

gotosch - the method used to go to school, factor

snack - eat snack or not in a day, binary

tvhours - number of hours watching TV per week, numeric

cmpthours - number of hours using computer per week, numeric

cellhours - number of hours playing with cell phones per week, numeric

sports - join in a sport team or not, 1: yes; and 2: no

exercises - number of hours of exercises per week, numeric

sweat - number of hours of sweating activities per week, numeric

overweigh - the child is overweighed or not, binary

Examples

```
data(weight_behavior)
names(weight_behavior)
```

Index