Package 'GeneralOaxaca'

January 20, 2025

Type Package

Title Blinder-Oaxaca Decomposition for Generalized Linear Model

Version 1.0

Date 2015-08-12

Author Aurelien Nicosia and Simon Baillargeon-Ladouceur

Maintainer Aurelien Nicosia <nicosia.aurelien@gmail.com>

Description Perform the Blinder-Oaxaca decomposition for generalized linear model with bootstrapped standard errors. The twofold and threefold decomposition are given, even the generalized linear model output in each group.

License GPL (>= 2)

Imports boot, stats

Suggests MASS

Collate 'statBO.R' 'GeneralOaxaca.R'

NeedsCompilation no

Repository CRAN

Date/Publication 2015-08-17 00:33:44

Contents

GeneralOaxaca-package *Blinder-Oaxaca Decomposition for Generalized Linear Model*

Description

Perform the Blinder-Oaxaca decomposition for generalized linear model with bootstrapped standard errors. The twofold and threefold decomposition are given, even the generalized linear model output in each group.

Details

The DESCRIPTION file:

Package:	GeneralOaxaca
Type:	Package
Title:	Blinder-Oaxaca Decomposition for Generalized Linear Model
Version:	1.0
Date:	2015-08-12
Author:	Aurelien Nicosia and Simon Baillargeon-Ladouceur
Maintainer:	Aurelien Nicosia <nicosia.aurelien@gmail.com>
Description:	Perform the Blinder-Oaxaca decomposition for generalized linear model with bootstrapped standard errors. Th
License:	GPL (>= 2)
Imports:	boot, stats
Suggests:	MASS
Collate:	'statBO.R' 'GeneralOaxaca.R'

Index of help topics:

Author(s)

Aurelien Nicosia and Simon Baillargeon-Ladouceur

Maintainer: Aurelien Nicosia <nicosia.aurelien@gmail.com>

References

T. Bauer and M. Sinning. An extension of the Blinder-Oaxaca decomposition to nonlinear models (2008). Advances in Statistical Analysis, Springer-Verlag.

Examples

```
data("chicago")
formula=ln.real.wage~ age + female + LTHS + some.college + college + advanced.degree
BO_A <- GeneralOaxaca(formula,  family= Gamma, data=chicago, groupInd=chicago$foreign.born,B=100)
BO_A$twofold
```

---

chicago *Labor market and demographic data for employed Hispanic workers in metropolitan Chicago*

Description

Data from a 2013 sample of employed Hispanic workers in metropolitan Chicago. It is a subset of the 2013 Current Population Survey (CPS) Outgoing Rotation Groups (ORG) data set provided by the Center for Economic and Policy Research in Washington, DC (CEPR, 2014).

Usage

```
data("chicago")
```

Format

A data frame with 712 observations on the following 9 variables. The 9 variables contain labor market and demographic information on a sample of employed Hispanic workers in the Chicago metropolitan area.

`age` the worker's age, expressed in years

`female` an indicator for female gender

`foreign.born` an indicator for foreign-born status

`LTHS` an indicator for having completed less than a high school (LTHS) education

`high.school` an indicator for having completed a high school education

`some.college` an indicator for having completed some college education

`college` an indicator for having completed a college education

`advanced.degree` an indicator for having completed an advanced degree

`ln.real.wage` the natural logarithm of the worker's real wage (in 2013 U.S. dollars)

Source

Center for Economic and Policy Research (CEPR). 2014. CPS ORG Uniform Extracts, Version 1.9 . Washington, DC.

Examples

```
data("chicago")
summary(chicago)
```

GeneralOaxaca *General Blinder-Oaxaca Decomposition*

Description

Blinder-Oaxaca decomposition for generalized linear model. It provide the twofold and threefold decomposition describe in Bauer and Sinning (2008), as the characteristic, coefficient and interaction part of the observed difference on the dependent variable between the two groups. Bootstrapped standard errors are calculated (e.g., Efron, 1979).

Usage

```
GeneralOaxaca(formula, family = stats::gaussian, data,
groupInd, groupRef = "A", B = 1000, control = list())
```

Arguments

formula	an object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted.
family	a description of the error distribution and link function to be used in the model. (See family for details of family functions.)
data	an optional data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model. If not found in data, the variables are taken from environment(formula).
groupInd	is an indicator variable that is TRUE (or equal to 1) when an observation belongs to Group A, and FALSE (or equal to 0) when it belongs to Group B
groupRef	Group of reference for the decomposition, by default Group A.
B	number of bootstrap replications for the calculation of standard errors
control	a list of parameters for controlling the fitting process.

Details

The twofold and threefold decomposition contains the characteristic and coefficient part (also the interaction for the threefold) of the decomposition, with their proportion with respect to the observed difference between groups. It also give the z value, p value and 95% confidence intervals computed using the bootstrapped standard errors. The regoutput are the results of the generalized linear model applied to data in each group (A and B). See glm for more details about the outputs.

Value

GeneralOaxaca returns the following results:

regoutput	List of two elements (names GroupA and GroupB) with the standard generalized linear model output in each group.
twofold	the twofold decomposition with the respect groupInd.

threefold	the threefold decomposition with the respect groupInd.
n	the size of each respective group.
summaryStat	descriptive statistic of the independent variable in each group.

Note

The function can perform the Blinder-Oaxaca decomposition for any generalized linear model supported by the glm function.

Author(s)

Aurelien Nicosia and Simon Baillargeon-Ladouceur

References

T. Bauer and M. Sinning. An extension of the Blinder-Oaxaca decomposition to nonlinear models (2008). Advances in Statistical Analysis, Springer-Verlag.

B. Efron. Bootstrap Methods: Another Look at the Jackknife (1979). Annals of Statistics, 7(1), 1-26.

Examples

```
data("chicago")
chicago$real.wage=exp(chicago$ln.real.wage)
formula=ln.real.wage~ age + female + LTHS + some.college + college +
advanced.degree

# exemple with gamma distribution
BO_A <- GeneralOaxaca(formula,  family= Gamma, data=chicago,
groupInd=chicago$foreign.born,B=100)
BO_A$twofold
BO_A$regoutput$GroupA
BO_A$threefold
```

---

statBO *statBO*

Description

statBO performs a Blinder-Oaxaca decomposition for generalized linear regression models for the individual with indice in the data. statBO is used for the bootstrap estimation of the standard errors

Usage

```
statBO(data, formula, family, groupRef, groupInd, indice)
```

Arguments

data	an optional data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model. If not found in data, the variables are taken from environment(formula).
formula	an object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted.
family	a description of the error distribution and link function to be used in the model. (See family for details of family functions.)
groupRef	Group of reference for the decomposition, by default Group A.
groupInd	is an indicator variable that is TRUE (or equal to 1) when an observation belongs to Group A, and FALSE (or equal to 0) when it belongs to Group B
indice	identification of each individual in data. Useful for the call of the boot function.

Value

char	characteristic part in the twofold decomposition
coef	coefficient part in the twofold decomposition
diff	diff of the dependent variable between group
char3	characteristic part in the threefold decomposition
coef3	coefficient part in the threefold decomposition
int	interaction part in the threefold decomposition

Author(s)

Aurelien Nicosia and Simon Baillargeon-Ladouceur

See Also

See Also as GeneralOaxaca

Index