

Package ‘OutlierDC’

July 2, 2014

Title Outlier Detection using quantile regression for Censored Data

Date 2014-03-23

Version 0.3-0

Description

This package provides three algorithms to detect outlying observations for censored survival data.

Depends R (>= 3.0.0), methods, survival, quantreg, Formula

Maintainer Soo-Heang Eo <eo.sooheang@gmail.com>

License GPL-3

Author Soo-Heang Eo [aut, cre],Huixia Judy Wang [ctb],Lan Wang [ctb],HyungJun Cho [aut]

NeedsCompilation no

Repository CRAN

Date/Publication 2014-03-24 13:18:13

R topics documented:

OutlierDC-package	2
coef	3
ebd	4
odc	4
OutlierDC-class	7
plot	8
show	9
update	9

Index	11
--------------	-----------

OutlierDC-package *Functions for detecting outlying observations for censored data using quantile regression*

Description

This package offers three outlier detection algorithms for censored data using quantile regression.

Details

Package: OutlierDC
Type: Package
Version: 0.3-0
Date: 2014-03-23
License: GPL version 3
LazyLoad: no

Note

We would like to thank Huxia Judy Wang and Lan Wang for permission to use their LCRQ functions.

Author(s)

Soo-Heang Eo and HyungJun Cho Maintainer: Soo-Heang Eo <eo.sooheang@gmail.com>

References

Eo S-H, Hong S-M Hong, Cho H (2014). Identification of outlying observations with quantile regression for censored data, *Submitted*.

Wang HJ, Wang L (2009) Locally Weighted Censored Quantile Regression. *JASA* 104:1117–1128. doi: 10.1198/jasa.2009.tm08230

See Also

[odc](#), [plot](#), [coef](#), [show](#), [quantreg](#)

Examples

```
## Not run:  
require(OutlierDC)  
# Toy example  
data(ebd)  
# The data consists of 402 observations with 6 variables.
```

```

dim(ebd)
# To show the first six observations of the dataset,
head(ebd)

#scoring algorithm
fit <- odc(Surv(log(time), status) ~ meta, data = ebd)
fit
coef(fit)
plot(fit)

# Add upper bound for the selection of outliers
fit1 <- update(fit, k_s = 4)
fit1
plot(fit1)

# residual-based algorithm
fit2 <- odc(Surv(log(time), status) ~ meta, data = ebd, method = "residual", k_r = 1.5)
fit2
plot(fit2)

# To display all of outlying observations in the fitted object
fit2@outlier.data

# boxplot algorithm
fit3 <- odc(Surv(log(time), status) ~ meta, data = ebd, method = "boxplot", k_b = 1.5)
fit3
plot(fit3, ylab = "log survival times", xlab = "metastasis lymph nodes")

## End(Not run)

```

coef

a coef method for "OutlierDC".

Description

coef is a generic function which extracts model coefficient matrix including the 10th, 25th, 50th, 75th, 90th quantile estimates.

Usage

```
## S4 method for signature 'OutlierDC'
coef(object)
```

Arguments

object an object with class `OutlierDC`.

Details

This function is a generic function `coef` for the S4 class `OutlierDC`. It can be invoked by calling `print` for an object of the appropriate class, or directly by calling `coef` regardless of the class of the object.

See Also

[odc](#) and `OutlierDC` class

ebd	<i>Extrahepatic Cholangiocarcinoma Data</i>
-----	---

Description

The extrahepatic cholangiocarcinoma data comes from the US Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute.

Usage

```
data(ebd)
```

Source

Hankey B., Ries L., and Edwards B. (1999). The surveillance, epidemiology, and end results program a national resource. *Cancer Epidemiology Biomarkers and Prevention*, 12:1117-1121.

Examples

```
data(ebd)
```

odc	<i>Outlier detection using quantile regression for censored data</i>
-----	--

Description

outlier detection algorithms using quantile regression for censored data

Usage

```
odc(formula, data,
     method = c("score", "boxplot", "residual"),
     rq.model = c("Wang", "PengHuang", "Portnoy"),
     k_r = 1.5, k_b = 1.5, h = .05)
```

Arguments

formula	a type of Formula object with a survival object on the left-hand side of the ~ operator and covariate terms on the right-hand side. The survival object with survival time and its censoring status is constructed by the <code>Surv</code> function in survival package.
data	a data frame with variables used in the formula. It needs at least three variables, including survival time, censoring status, and covariates.
method	the outlier detection method to be used. The options "socre", "boxplot", and "residual" conduct the scoring, boxplot, and residual-based algorithm, respectively. The default algorithm is "score".
rq.model	the type of censored quantile regression to be used for fitting. The options "Wang", "Portnoy", and "PengHuang" conduct Wang and Wang's, Portnoy's, and Peng and Huang's censored quantile regression approaches, respectively. The default is "Wang".
k_r	a value to control the tightness of cut-offs for the residual algorithm with a default value of 1.5.
k_b	a value to control the tightness of cut-offs for the boxplot algorithm with a default value of 1.5.
h	bandwidth for locally weighted censored quantile regression with a default value of 0.05.

Details

The `odc` function conducts three outlier detection algorithms on the basis of censored quantile regression. Three outlier detection algorithms were implemented: residual-based, boxplot, and scoring algorithms. The residual-based algorithm detects outlying observations using constant scale estimates; however, it does not account for the heterogeneity of variability. When the data is extremely heterogeneous, the boxplot algorithm with censored quantile regression is more effective. The residual-based and boxplot algorithms produce cut-offs to determine whether observations are outliers. In contrast, the scoring algorithm provides the outlying magnitude or deviation of each point from the distribution of observations. Outlier detection is achieved by visualising the scores.

Value

an object of the S4 class "OutlierDC" with the following slots:

call: evaluated function call

formula: formula to be used

raw.data: data to be used for model fitting

refined.data: the data set after removing outliers

refined.data: the data set containing outliers

coefficients: the estimated censored quantile regression coefficient matrix consisting of 10th, 25th, 50th, 75th, and 90th quantiles

fitted.mat: the censored quantile regression fitted value matrix consisting of 10th, 25th, 50th, 75th, and 90th quantiles

score: outlying scores (scoring algorithm) or residuals (residual-based algorithm)

cutoff: estimated scale parameter for the residual-based algorithm

lower: lower fence vector used for the boxplot and scoring algorithms

upper: upper fence vector used for the boxplot and scoring algorithms
 outliers: logical vector to determine which observations are outliers
 n.outliers: number of outliers detected
 method: outlier detection method to be used
 rq.model: censored quantile regression to be used
 k_r: a value to be used for the tightness of cut-offs in the residual algorithm
 k_b: a value to be used for the tightness of cut-offs in the boxplot algorithm
 k_s: a value to be used for the tightness of upper fence cut-offs used for the scoring algorithm with the [update](#) function

Source

Eo S-H, Hong S-M Hong, Cho H (2014). Identification of outlying observations with quantile regression for censored data, *Submitted*.

Wang HJ, Wang L (2009) Locally Weighted Censored Quantile Regression. *JASA* 104:1117–1128. doi: 10.1198/jasa.2009.tm08230

See Also

[OutlierDC-package](#)
[coef](#), [plot](#), [show](#), [update](#)

Examples

```
## Not run:
require(OutlierDC)
# Toy example
data(ebd)
# The data consists of 402 observations with 6 variables.
dim(ebd)
# To show the first six observations of the dataset,
head(ebd)

#scoring algorithm
fit <- odc(Surv(log(time), status) ~ meta, data = ebd)
fit
coef(fit)
plot(fit)

# Add upper bound for the selection of outliers
fit1 <- update(fit, k_s = 4)
fit1
plot(fit1)

# residual-based algorithm
fit2 <- odc(Surv(log(time), status) ~ meta, data = ebd, method = "residual", k_r = 1.5)
fit2
plot(fit2)

# To display all of outlying observations in the fitted object
```

```

fit2@outlier.data

# boxplot algorithm
fit3 <- odc(Surv(log(time), status) ~ meta, data = ebd, method = "boxplot", k_b = 1.5)
fit3
plot(fit3, ylab = "log survival times", xlab = "metastasis lymph nodes")

## End(Not run)

```

OutlierDC-class	<i>"OutlierDC" class</i>
-----------------	--------------------------

Description

The OutlierDC class presents outlier detection algorithms for censored data.

Objects from the Class

Objects can be created by calls of the form `new("OutlierDC")`.

Slots

call: evaluated function call
formula: formula to be used with the type of "Formula"
raw.data: data to be used with the type of "data.frame"
refined.data: the data set after removing outliers
outlier.data: the data set containing outliers
coefficients: the estimated censored quantile regression coefficient matrix
fitted.mat: the censored quantile regression fitted value matrix with the type of "matrix"
score: outlying scores (scoring algorithm) or residuals (residual-based algorithm)
cutoff: estimated scale parameter for the residual-based algorithm
lower: lower fence vector used for the boxplot and scoring algorithms with the type of "vector"
upper: upper fence vector used for the boxplot and scoring algorithms with the type of "vector"
outliers: logical vector to determine which observations are outliers
n.outliers: number of outliers to be used. The object of class "integer".
method: outlier detection method to be used
rq.model: censored quantile regression to be fitted
k_r: a value to be used for the tightness of cut-offs in the residual-based algorithm
k_b: a value to be used for the tightness of cut-offs in the boxplot algorithm
bound: type of fence to be used in the model fitting
k_s: a value to be used for the tightness of upper fence cut-offs used for the scoring algorithm with update function

Methods

coef signature(object = "OutlierDC"): Print the coefficient matrix of censored quantile regression to be used. See [coef](#).

plot signature(x = "OutlierDC", y = "missing"): See [plot](#).

show signature(object = "OutlierDC"): See [show](#).

update signature(object = "OutlierDC"): Update the fitted object to find outliers in scoring algorithm. See [update](#).

See Also

[OutlierDC-package](#)
[coef](#), [plot](#), [show](#), [update](#)

Examples

```
showClass("OutlierDC")
```

plot	<i>a plot-method for a "OutlierDC" object</i>
------	---

Description

This function provides three different results. If the algorithm is `score`, it draws a normal quantile-quantile plot of the outlying scores. If the algorithm is `boxplot`, the scatter plot of log survival times against the covariate used is given. Lastly, if the algorithm is `residual`, it offers a residual plot.

Usage

```
## S4 method for signature 'OutlierDC'
plot(x, y = NA, ...)
```

Arguments

x	fitted model object of class OutlierDC .
y	missing value
...	plot.default arguments

See Also

[odc](#) and [OutlierDC](#) class

show	<i>a show method for OutlierDC</i>
------	------------------------------------

Description

This function provides a summary for the OutlierDC class.

Usage

```
## S4 method for signature 'OutlierDC'  
show(object)
```

Arguments

object fitted model object of class [OutlierDC](#).

Details

This function is a method for the generic function show for the S4 class OutlierDC. It can be invoked by calling print for an object of the appropriate class, or directly by calling show regardless of the class of the object.

See Also

[odc](#) and [OutlierDC](#)

update	<i>Update a scoring algorithm.</i>
--------	------------------------------------

Description

This function updates a scoring algorithm using upper and/or lower fences. Using the call stored in the object, the update function declares outlying observations based on the QQ plot. k_s is used to set the upper cut-off bound.

Usage

```
## S4 method for signature 'OutlierDC'  
update(object, k_s = NA, LB = NA)
```

Arguments

object fitted model object of class [OutlierDC](#).
k_s cut-off value for the upper fence
LB cut-off value for the lower fence

Details

This function is a generic function called `update` for the S4 class `OutlierDC`. Cut-off bounds are added to find outliers on the normal QQ plot.

See Also

[odc](#) and `OutlierDC` class

Index

- *Topic **class**
 - OutlierDC-class, [7](#)
- *Topic **datasets**
 - ebd, [4](#)
- *Topic **methods**
 - coef, [3](#)
 - plot, [8](#)
 - show, [9](#)
 - update, [9](#)
- *Topic **models**
 - odc, [4](#)
- *Topic **package**
 - OutlierDC-package, [2](#)

coef, [2](#), [3](#), [6](#), [8](#)
coef, OutlierDC-class
 (OutlierDC-class), [7](#)
coef, OutlierDC-method (coef), [3](#)

ebd, [4](#)

odc, [2](#), [4](#), [4](#), [8–10](#)
OutlierDC, [3](#), [4](#), [8–10](#)
OutlierDC (OutlierDC-package), [2](#)
OutlierDC-class, [7](#)
OutlierDC-package, [2](#)

plot, [2](#), [6](#), [8](#), [8](#)
plot, OutlierDC-method (plot), [8](#)
plot.default, [8](#)

quantreg, [2](#)

show, [2](#), [6](#), [8](#), [9](#)
show, OutlierDC-method (show), [9](#)
Surv, [5](#)

update, [6](#), [8](#), [9](#)
update, OutlierDC-method (update), [9](#)