Package ‘OutlierDC’

March 18, 2013

Type Package
Version 0.2.1
Date 2013-03-16
Title Outlier Detection using Quantile Regression for Censored Data
Author Soo-Heang Eo <hanansh@korea.ac.kr> and HyungJun Cho <hj4cho@korea.ac.kr>
Maintainer Soo-Heang Eo <hanansh@korea.ac.kr>
Depends R (>= 2.15.3), methods, quantreg, survival, Formula
Description This provides three algorithms to detect outliers based on censored quantile regression.
URL http://statlab.korea.ac.kr
License GPL (>= 3)
NeedsCompilation no
Repository CRAN
Date/Publication 2013-03-18 09:26:31

R topics documented:

OutlierDC-package ........................................... 2
coeff ......................................................... 2
ecb .......................................................... 3
odc ........................................................... 4
odcNews ...................................................... 6
OutlierDC-class .............................................. 7
plot ........................................................... 8
show .......................................................... 9
update ......................................................... 9

Index 11
OutlierDC-package  
*Functions for Detecting Outlying Observations using Quantile Regression in Censored Data*

**Description**

This package offers three outlier detection algorithms based on censored quantile regression.

**Details**

<table>
<thead>
<tr>
<th>Package</th>
<th>OutlierDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Package</td>
</tr>
<tr>
<td>Version</td>
<td>0.2-1</td>
</tr>
<tr>
<td>Date</td>
<td>2013-03-16</td>
</tr>
<tr>
<td>License</td>
<td>GPL version 2 or newer</td>
</tr>
<tr>
<td>LazyLoad</td>
<td>no</td>
</tr>
</tbody>
</table>

The package OutlierDC is the implementation of the outlier detection algorithms proposed by Eo, Hong and Cho (2013+). Three outlier detection algorithms using quantile regression are proposed, which are modified versions of existing algorithms for uncensored data. We illustrated with real data from the SEER database, which contains a number of data sets related to various cancers.

**Author(s)**

Soo-Heang Eo Seung-Mo Hong and HyungJun Cho Maintainer: Soo-Heang Eo <hanansh@korea.ac.kr>

**References**


**See Also**

`odc`, `plot`, `coef`, `show`, `quantreg`

---

**Description**

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

```r
## 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**

*Extrahepatic Cholangiocarcinoma Data*

Description

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

Usage

```r
data(ebd)
```

References

Description

outlier detection algorithms using quantile regression for censored data

Usage

```r
odc(formula, data, method = c("score", "boxplot", "residual"),
     rq.model = c("Wang", "PengHuang", "Portnoy"), k = 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 `Surv` function in the `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 "score", "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**: a constant to control the tightness of cut-offs for residual-based and boxplot algorithms 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 visualizing 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
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: constant value to be used for the tightness of cut-offs in the residual-based and boxplot algorithms
UB: upper fence used for the scoring algorithm with the update function
LB: lower fence used for the scoring algorithm with the update function

Author(s)

Soo-Heang Eo <hanansh@korea.ac.kr>
HyungJun Cho <hj4cho@korea.ac.kr>

References


See Also

OutlierDC-package
coef, plot, show, update

Examples

## Not run:
require(OutlierDC)
data(ebd)

#scoring algorithm
fit <- odc(Surv(log(time), status) ~ meta, data = ebd)
fit
coef(fit)
plot(fit)
```r
fit1 <- update(fit, UB = 4, LB = -4)
plot(fit1)

fit2 <- update(fit, UB = NA, LB = -4)
fit3 <- update(fit, UB = 4, LB = NA)

# residual-based algorithm
fit4 <- odc(Surv(log(time), status) ~ meta, data = ebd, method = "residual", k = 3)
plot(fit4)

# boxplot algorithm
fit5 <- odc(Surv(log(time), status) ~ meta, data = ebd, method = "boxplot", k = 1.5)
plot(fit5, ylab = "log survival times", xlab = "metastasis lymph nodes")

## End(Not run)
```

### odcNews

Show the NEWS file of the OutlierDC package.

#### Description

Show the NEWS file of the OutlierDC package.

#### Usage

```r
odcNews()
```

#### See Also

- OutlierDC-package
- odc
- coef.plot, show, update

#### Examples

```r
odcNews()
```
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 with the type of "language"
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 with the type of "data.frame"
coefficients: the estimated censored quantile regression coefficient matrix with the type of "data.frame"
fitted.mat: the censored quantile regression fitted value matrix with the type of "matrix"
score: outlying scores (scoring algorithm) or residuals (residual-based algorithm). The object of class "vector"
cutoff: estimated scale parameter for the residual-based algorithm with the type of "vector"
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. The object of class "vector"
n.outliers: number of outliers to be used. The object of class "integer"
method: outlier detection method to be used with the type of "character"
_rq.model: censored quantile regression to be fitted. The object of class "character"
k: constant value to be used for the tightness of cut-offs in the residual-based and boxplot algorithm with the type of "numeric"
UB: upper fence used for the scoring algorithm with update function. The Object of class "numeric"
LB: lower fence used for the scoring algorithm with update function. The object of class "numeric"

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.
Author(s)

Soo-Heang Eo <hanansh@korea.ac.kr>

See Also

OutlierDC-package
coeff, plot, show, update

Examples

showClass("OutlierDC")

plot

a plot method for OutlierDC

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**

*a show method for OutlierDC*

Description

This function provides a summary for the OutlierDC class.

Usage

```r
## 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**

*Update a scoring algorithm.*

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. UB is used to set the upper cut-off bound and LB is used to set the lower cut-off bound.

Usage

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

Arguments

- `object`: fitted model object of class `OutlierDC`.
- `UB`: 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 classes
  OutlierDC-class, 7

*Topic datasets
  ebd, 3

*Topic methods
  coef, 2
  odcNews, 6
  plot, 8
  show, 9
  update, 9

*Topic models
  odc, 4

*Topic package
  OutlierDC-package, 2

coef, 2, 5–8
coef, OutlierDC-method (coef), 2

ebd, 3

odc, 2, 3, 4, 6, 8–10
odcNews, 6
OutlierDC, 3, 8–10
OutlierDC-class, 7
OutlierDC-package, 2

plot, 2, 5–8, 8
plot, OutlierDC-method (plot), 8
plot.default, 8

quantreg, 2

show, 2, 5–8, 9
show, OutlierDC-method (show), 9
Surv, 4

update, 5–8, 9
update, OutlierDC-method (update), 9