

The mblm Package

February 16, 2008

Type Package

Title Median-Based Linear Models

Version 0.11

Date 2007-07-23

Author Lukasz Komsta <luke@novum.am.lublin.pl>

Maintainer Lukasz Komsta <luke@novum.am.lublin.pl>

Description This package provides linear models based on Theil-Sen single median and Siegel repeated medians. They are very robust (29 or 50 percent breakdown point, respectively), and if no outliers are present, the estimators are very similar to OLS.

License GPL version 2 or newer. <http://www.gnu.org/copyleft/gpl.html>

URL <http://www.r-project.org>, <http://www.komsta.net/>

R topics documented:

confint.mblm	1
mblm	3
summary.mblm	4

Index	7
--------------	----------

confint.mblm	<i>Confidence Intervals for 'mblm' Model</i>
--------------	--

Description

Computes confidence intervals for one or more parameters in a fitted model of 'mblm' class.

Usage

```
confint.mblm(object, parm, level = 0.95, ...)
```

Arguments

object	a fitted model object
parm	a specification of which parameters are to be given confidence intervals, either a vector of numbers or a vector of names. Not yet implemented for 'mblm'
level	the confidence level required
...	additional arguments

Details

This function computes confidence intervals for slope and intercept in linear model based on single median or repeated medians. The confidence intervals are computed in simplest way, as confidence interval for the median of all slopes or intercepts found during fitting.

Value

A matrix (or vector) with columns giving lower and upper confidence limits for each parameter.

Note

The recommended method of calculating confidence intervals, given by Sen and based on Kendall's tau, not Wilcoxon test, is not implemented at this time and is considered to be implemented in next version of this package.

Author(s)

Lukasz Komsta

References

Sen, P.K. (1968). Estimates of Regression Coefficient Based on Kendall's tau. J. Am. Stat. Ass. 63, 324, 1379-1389.

See Also

[mblm](#), [summary.mblm](#)

Examples

```
set.seed(1234)
x <- 1:100+rnorm(100)
y <- x+rnorm(100)
y[100] <- 200
fit <- mblm(y~x)
fit
summary(fit)
confint(fit)
```

Description

This function is used to fit linear models based on Theil-Sen single median, or Siegel repeated medians.

Usage

```
mblm(formula, dataframe, repeated = TRUE)
```

Arguments

<code>formula</code>	A formula of type $y \sim x$ (only linear models are accepted)
<code>dataframe</code>	Optional dataframe
<code>repeated</code>	If set to true, model is computed using repeated medians. If false, a single median estimators are calculated

Details

Theil-Sen single median method computes slopes of lines crossing all possible pairs of points, when x coordinates differ. After calculating these $n(n-1)/2$ slopes (these value are true only if x is distinct), the median of them is taken as slope estimator. Next, the intercepts of n lines, crossing each point and having calculated slope are calculated. The median from them is intercept estimator.

Siegel repeated medians is more complicated. For each point, the slopes between it and the others are calculated (resulting n-1 slopes) and the median is taken. This results in n medians and median from this medians is slope estimator. Intercept is calculated in similar way, for more information please take a look in function source.

The breakdown point of Theil-Sen method is about 29%, Siegel extended it to 50%, so these regression methods are very robust. Additionally, if the errors are normally distributed and no outliers are present, the estimators are very similar to classic least squares.

Value

An object of class `c("mblm", "lm")`, containing minimal set of data to perform basic operations, such as in case of `lm` model. Additionally, the return value contains 2 fields:

<code>slopes</code>	The slopes (in single median), or medians of slopes (in repeated medians) between tested point pairs
<code>intercepts</code>	The intercepts calculated

Note

This function should have compatibility with all 'lm' methods, but it is not guaranteed that they will work or have any cognitive value (this method is nonparametric). The compatibility was only introduced to use some basic methods from 'lm' without programming new functions.

Author(s)

Lukasz Komsta, some fixes by Sven Garbade

References

Theil, H. (1950) A rank invariant method for linear and polynomial regression analysis. *Nederl. Akad. Wetensch. Proc. Ser. A* 53, 386-392 (Part I), 521-525 (Part II), 1397-1412 (Part III).

Sen, P.K. (1968). Estimates of Regression Coefficient Based on Kendall's tau. *J. Am. Stat. Ass.* 63, 324, 1379-1389.

Siegel, A.F. (1982). Robust Regression Using Repeated Medians. *Biometrika*, 69, 1, 242-244.

See Also

[lm](#), [summary.mblm](#), [confint.mblm](#)

Examples

```
set.seed(1234)
x <- 1:100+rnorm(100)
y <- x+rnorm(100)
y[100] <- 200
fit <- mblm(y~x)
fit
summary(fit)
fit2 <- lm(y~x)
plot(x,y)
abline(fit)
abline(fit2,lty=2)
plot(fit)
residuals(fit)
fitted(fit)
plot(density(fit$slopes))
plot(density(fit$intercepts))
anova(fit)
anova(fit2)
anova(fit,fit2)
confint(fit)
AIC(fit,fit2)
```

summary.mblm

Summarizing median-based linear models

Description

'summary' method for class 'mblm'

Usage

```
summary.mblm(object, ...)
```

Arguments

object	an object of class 'mblm', usually, a result of a call to 'mblm'.
...	additional arguments

Details

This function is based on `summary.lm` code, and the base difference is use of nonparametric `wilcox.test` to obtain significance and `mad` instead of standard error of estimates. Of course you can force standard `lm` behavior by calling `summary.lm`, but values received in such way has low cognitive value.

Value

For the return value, see `summary.lm`. Summary of 'mblm' class does not contain R-squared values and F-test result.

Note

The significance of estimators can be computed more "lege artis" based on Kendall's tau, as suggested by Sen, but today such feature is not yet implemented.

Author(s)

Lukasz Komsta

References

Theil, H. (1950) A rank invariant method for linear and polynomial regression analysis. *Nederl. Akad. Wetensch. Proc. Ser. A* 53, 386-392 (Part I), 521-525 (Part II), 1397-1412 (Part III).

Sen, P.K. (1968). Estimates of Regression Coefficient Based on Kendall's tau. *J. Am. Stat. Ass.* 63, 324, 1379-1389.

Siegel, A.F. (1982). Robust Regression Using Repeated Medians. *Biometrika*, 69, 1, 242-244.

See Also

[summary.lm](#), [mblm](#), [confint.mblm](#)

Examples

```
set.seed(1234)
x <- 1:10+rnorm(10)
y <- x+rnorm(10)
y[10] <- 20
fit <- mblm(y~x)
```

```
summary(fit)
```

Index

*Topic **models**

confint.mblm, 1
mblm, 3
summary.mblm, 4

confint.mblm, 1, 4, 5

lm, 4

mblm, 2, 3, 5

summary.lm, 5

summary.mblm, 2, 4, 4