Package 'deming'

July 22, 2025

Title Deming, Theil-Sen, Passing-Bablock and Total Least Squares Regression

Maintainer Terry Therneau <therneau.terry@mayo.edu>

Description
Generalized Deming regression, Theil-Sen regression and Passing-Bablock regression functions.

Imports boot

Priority optional

Version 1.4-1

LazyData Yes

Author Terry Therneau

License LGPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2024-06-26 15:30:02 UTC

Contents

arsenate *Comparison of two assays for arsenate.*

Description

Arsenate(V) ion in natural river waters, as determined by two assay methods.

Usage

```
data(arsenate)
```

Format

A data frame with 30 observations on the following 4 variables.

`aas` micrograms/liter, by continuous selective reduction and atomic absorbtion spectectromitry

`se.aas` estimated standard error of the result

`aes` micrograms/liter, by non-selective reduction, cold trapping, and atomic emission spectroscopy

`se.aes` estimated standard error of the result

Source

The data is found in BD Ripley and M Thompson, Regression techniques for the detection of analytical bias, Analyst 112:377-383, 1987.

deming *Fit a generalized Deming regression*

Description

Find the MLE line relating x and y when both are measured with error. When the variances of x and y are constant and equal, this is the special case of Deming regression.

Usage

```
deming(formula, data, subset, weights, na.action, cv=FALSE,
       xstd, ystd, stdpat, conf=.95, jackknife=TRUE, dfbeta=FALSE,
       id, x=FALSE, y=FALSE, model=TRUE)
```

Arguments

formula	a model formula with a single continuous response on the left and a single continuous predictor on the right.
data	an optional data frame, list or environment containing the variables in the model.
subset	an optional vector specifying a subset of observations to be used in the fitting process.
weights	an optional vector of weights to be used in the fitting process. Should be NULL or a numeric vector.
I	
na.action	a function which indicates what should happen when the data contain NAs. The default is set by the na.action setting of options. The 'factory fresh' default is na.omit, the na.exclude option is often useful.
xstd	optional, the variable name of a vector that contains explicit error values for each of the predictor values. This data overrides the cv option if both are present.
ystd	optional, the variable name of a vector that contains explicit error values for each of the response values. This data overrides the cv option if both are present.
cv	constant coefficient of variation? The default of false corresponds to ordinary Deming regression, i.e., an assumption of constant error. A value of cv=TRUE corresponds to the assumption of constant coefficient of variation.
stdpat	pattern for the standard deviation, see comments below. If this is missing the default is based on the cv option.
conf	confidence level for the confidence interval
jackknife	compute a jackknife estimate of variance.
dfbeta	return the dfbeta matrix from the jackknife computation.
id	grouping values for the grouped jackknife
x, y, model	logicals. If TRUE the corresponding components of the fit (the model frame, the model matrix, or the response) is returned.

Details

Ordinary least squares regression minimizes the sum of distances between the y values and the regression line, Deming regression minimizes the sum of distances in both the x and y direction. As such it is often appropriate when both x and y are measured with error. A common use is in comparing two assays, each of which is designed to quantify the same compound.

The standard deviation of the x variate variate will often be of the form $\sigma(c + dx)$ for c and d some constants, where σ is the overal scale factor; similarly for y with constants e and f. Ordinary Deming regression corresponds to c=1 and d=0, i.e., constant variation over the range of the data. A more realistic assumption for many laboratory measurments is c=0 and d=1, i.e., constant coefficient of variation. Laboratory tests are often assumed to have constant coefficient of variation rather than constant variance.

There are 3 ways to specify the variation. The first is to directly set the pattern of (c,d,e,f) for the x and y standard deviations. If this is omitted, a default of (0,1,0,1) or (1,0,1,0) is chosen, based on whether the cv option is TRUE or FALSE, respectively. As a third option, the user can

specifiy xstd and ystd directly as vectors of data values. In this last case any values for the stdpat or ccs options are ignored. Note that the two calls deming(y ~ x, cv=TRUE) and deming(y ~ x, xstd=x, ystd=y) are subtly different. In the second the standard deviation values are based on the data, and in the first they will be based on the fitted values. The two outcomes will often be nearly identical.

Although a cv option of TRUE is often much better justified than an assumption of constant variance, assumpting a perfectly constant CV can also be questionable. Most actual biologic assays will have both a constant and a proportional component of error, with the former becoming dominant for values near zero and the latter dominant elsewhere. If all of the results are far from zero, however, the constant part may be ignored.

Many times an assay will be done in duplicate, in which case the paired results can have correlated errors due to sample handling or manipulation that preceeds splitting it into separate aliquots for assay, and the ordinary variance will be too small (as it also is when the duplicate values are averaged together before fitting the regression line.) A correct grouped jackknife estimate of variance is obtained in this case by setting id to a vector of sample identifiers.

Value

a object of class 'deming' containing the components:

coefficient	the coefficient vector, containing the intercept and slope.
variance	The jackknife or bootstrap estimate of variance
ci	bootstrap confidence intervals, if nboot >0
dfbeta	pptionally, the dfbeta residuals. A 2 column matrix, each row is the change in the coefficient vector if that observation is removed from the data.

Author(s)

Terry Therneau

References

BD Ripley and M Thompson, Regression techniques for the detection of analytical bias, Analyst 112:377-383, 1987.

K Linnet, Estimation of the linear relationship between the measurements of two methods with proportional errors. Statistics in Medicine 9:1463-1473, 1990.

Examples

```
# Data from Ripley and Thompson
fit <- deming(aes ~ aas, data=arsenate, xstd=se.aas, ystd=se.aes)
print(fit)
## Not run:
            Coef se(coef) lower 0.95 upper 0.95
Intercept 0.1064   0.2477    -0.3790     0.5919
Slope     0.9730   0.1430     0.6928     1.2532

   Scale= 1.358
```

```
## End(Not run)
plot(1:30, fit$dfbeta[,2]) #subject 22 has a large effect on the slope

# Constant proportional error fit (constant CV)
fit2 <- deming(new.lot ~ old.lot, ferritin, cv=TRUE,
               subset=(period==3))
```

`ferritin` *Validation of a ferritin assay*

Description

For each of seven periods in which there was a new batch of reagent, a small set of patient samples was assayed for ferritin content using both the old and new batches.

Usage

```
data(ferritin)
```

Format

A data frame with 162 observations on the following 4 variables.

`sample` sample identifier

`period` the transition number, 1 to 7

`old.lot` assay result using the old lot of the reagent

`new.lot` assay result using the new lot

Details

The samples from each period are distinct. In the second data set `ferritin2` outliers have been added to the data for period 2, excess noise added to one lot in period 4, and deterministic laboratory error to period 6.

Source

Blinded data from a clinical laboratory.

Examples

```
data(ferritin)
temp <- ferritin[ferritin$period <4,]
plot(temp$old.lot, temp$new.lot, type='n', log='xy',
     xlab="Old lot", ylab="New Lot")
text(temp$old.lot, temp$new.lot, temp$period,
         col=temp$period)
```

pbreg *Passing-Bablock regressin*

Description

Passing-Bablock regression is a robust regression method for two variables that is symmetric in x and y.

Usage

```
pbreg(formula, data, subset, weights, na.action, conf=.95,
     nboot = 0, method=1, eps=sqrt(.Machine$double.eps),
     x = FALSE, y = FALSE, model = TRUE)
```

Arguments

formula	a model formula with a single continuous response on the left and a single continuous predictor on the right.
data	an optional data frame, list or environment containing the variables in the model.
subset	an optional vector specifying a subset of observations to be used in the fitting process.
weights	an optional vector of weights to be used in the fitting process. Should be NULL or a numeric vector.
I	
na.action	a function which indicates what should happen when the data contain NAs. The default is set by the na.action setting of options. The 'factory fresh' default for R is na.omit, the na.exclude option is often useful.
conf	the width of the computed confidence limit
nboot	number of bootstrap samples used to compute standard errors and/or confidence limits.
method	which of 3 related methods to use for the computation
eps	the tolerance used to detect tied values in x and y
x, y, model	logicals. If TRUE the corresponding components of the fit (the model frame, the model matrix, or the response) is returned.

Details

There are 3 related estimators under this heading. Method 1 is the original Passing-Bablock (1983) method, which is equal to a Theil-Sen estimate symmetric about the y=x line. Method 2 is the first extended method of the 1988 paper, designed to be scale invariant. Method 3 is the second extended method from the 1985 paper, the "scissors" estimate which is symmetric about both the x and y axes, and is also scale invariant.

The default confidence interval estimate is based on that derived by Sen, which is in turn based on the relationship to Kendall's tau. A theoretical justification of this approach for methods 2 and 3 is lacking, and we recommend a bootstrap based confidence interval based on 500-1000 replications.

Value

pbreg returns an object of `class` "pbreg". The generic accessor functions `coef`, `fitted` and `residuals` extract the relevant components.

Author(s)

Terry Therneau

References

Passing, H. and Bablock, W. (1983). A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in Clinical Chemistry, Part I. J. Clin. Chem. Clin. Biochem. 21:709-720.

Passing, H. and Bablock, W. (1984). Comparison of several regression procedures for method comparison studies and determination of sample size. Application of linear regression procedures for method comparison studies in Clinical Chemistry, Part II. J. Clin. Chem. Clin. Biochem. 22:431-435.

Bablock, W., Passing, H., Bender, R. and Schneider, B. (1988). A general regression procedure for method transformations. Application of linear regression procedures for method comparison studies in Clinical Chemistry, Part III. J. Clin. Chem. Clin. Biochem. 26:783-790.

See Also

deming

Examples

```
afit1 <- pbreg(aes ~ aas, data= arsenate)
afit2 <- pbreg(aas ~ aes, data= arsenate)
rbind(coef(afit1), coef(afit2))  # symmetric results
1/coef(afit1)[2]
```

---

theilsen *Theil-Sen regression*

Description

Thiel-Sen regression is a robust regression method for two variables. The symmetric option gives a variant that is symmentric in x and y.

Usage

```
theilsen(formula, data, subset, weights, na.action, conf=.95,
     nboot = 0, symmetric=FALSE, eps=sqrt(.Machine$double.eps),
     x = FALSE, y = FALSE, model = TRUE)
```

Arguments

formula	a model formula with a single continuous response on the left and a single continuous predictor on the right.
data	an optional data frame, list or environment containing the variables in the model.
subset	an optional vector specifying a subset of observations to be used in the fitting process.
weights	an optional vector of weights to be used in the fitting process.
na.action	a function which indicates what should happen when the data contain NAs. The default is set by the na.action setting of options.
conf	the width of the computed confidence limit.
nboot	number of bootstrap samples used to compute standard errors and/or confidence limits. If this is 0 or missing then an asypmtotic formula is used.
symmetric	compute an estimate whose slope is symmetric in x and y.
eps	the tolerance used to detect tied values in x and y
x, y, model	logicals. If TRUE the corresponding components of the fit (the model frame, the model matrix, or the response) is returned.

Details

One way to characterize the slope of an ordinary least squares line is that $\rho(x, r)$ =0, where where ρ is the correlation coefficient and r is the vector of residuals from the fitted line. Thiel-Sen regression replaces ρ with Kendall's τ, a non-parametric alternative. It it resistant to outliers while retaining good statistical efficiency.

The symmetric form of the estimate is based on solving the inverse equation: find that rotation of the original data such that $\tau(x, y) = 0$ for the rotated data. (In a similar fashion,the rotation such the least squares slope is zero yields Deming regression.) In this case it is possible to have multiple solutions, i.e., slopes that yeild a 0 correlation, although this is rare unless the deviations from the fitted line are large.

The default confidence interval estimate is based on the result of Sen, which is in turn based on the relationship to Kendall's tau and is essentially an inversion of the confidence interval for tau. The argument does not extend to the symmetric case, for which we recommend using a bootstrap confidence interval based on 500-1000 replications.

Value

theilsen returns an object of class "theilsen" with components

coefficients	the intercept and slope
residuals	residuals from the fitted line
angle	if the symmetric option is chosen, this contains all of the solutions for the angle of the regression line
n	number of data points
model, x, y	optional componets as specified by the x, y, and model arguments
terms	the terms object corresponding to the formula

na.action	na.action information, if applicable
call	a copy of the call to the function

The generic accessor functions coef, residuals, and terms extract the relevant components.

Author(s)

Terry Therneau

References

Thiel, H. (1950), A rank-invariant method of linear and polynomial regression analysis. I, II, III, Nederl. Akad. Wetensch., Proc. 53: 386-392, 521-525, 1397-1412.

Sen, P.B. (1968), Estimates of the regression coefficient based on Kendall's tau, Journal of the American Statistical Association 63: 1379-1389.

See Also

deming, pbreg

Examples

```
afit1 <- theilsen(aes ~ aas, symmetric=TRUE, data= arsenate)
afit2 <- theilsen(aas ~ aes, symmetric=TRUE, data= arsenate)
rbind(coef(afit1), coef(afit2))  # symmetric results
1/coef(afit1)[2]
```

Index