Package 'bfsl'

December 16, 2018

Title Best-Fit Straight Line

Version 0.1.0

Description Provides the solution from York (1968) <doi:10.1016/S0012-821X(68)80059-7> for fitting a straight line to bivariate data with errors in both coordinates. It gives unbiased estimates of the intercept, slope and standard errors of the best-fit straight line to independent points with (possibly correlated) normally distributed errors in both x and y. Other commonly used errors-in-variables methods, such as orthogonal distance regression, geometric mean regression or Deming regression are special cases of York's solution.

Depends R (>= 3.5.0)

License GPL-3

URL https://github.com/pasturm/bfsl

BugReports https://github.com/pasturm/bfsl/issues

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

Suggests testthat

NeedsCompilation no

Author Patrick Sturm [aut, cre]

Maintainer Patrick Sturm <sturm@tofwerk.com>

Repository CRAN

Date/Publication 2018-12-16 14:50:03 UTC

R topics documented:

bfsl *Calculates the Best-fit Straight Line*

Description

bfsl calculates the best-fit straight line to independent points with (possibly correlated) normally distributed errors in both coordinates.

Usage

```
bfsl(x, y = NULL, sd_x = 0, sd_y = 1, r = 0, control = bfsl_control())
```

Arguments

x	A vector of *x* observations or a data frame (or an object coercible by as.data.frame to a data frame) containing the named vectors *x*, *y*, and optionally *sd_x*, *sd_y* and *r*. If weights *w_x* and *w_y* are given, then *sd_x* and *sd_y* are calculated from *sd_x = 1/sqrt(w_x)* and *sd_y = 1/sqrt(w_y)*. Specifying y, sd_x, sd_y or r directly as function arguments overwrites these variables in the data structure.
y	A vector of *y* observations.
sd_x	A vector of *x* measurement error standard deviations. If it is of length one, all data points are assumed to have the same *x* standard deviation.
sd_y	A vector of *y* measurement error standard deviations. If it is of length one, all data points are assumed to have the same *y* standard deviation.
r	A vector of correlation coefficients between errors in *x* and *y*. If it is of length one, all data points are assumed to have the same correlation coefficient.
control	A list of control settings. See bfsl_control for the names of the settable control values and their effect.

Details

bfsl provides the general least-squares estimation solution to the problem of fitting a straight line to independent data with (possibly correlated) normally distributed errors in both x and y.

With sd_x = 0 the (weighted) ordinary least squares solution is obtained. The calculated standard errors of the slope and intercept multiplied with sqrt(chisq) correspond to the ordinary least squares standard errors.

With sd_x = c, sd_y = d, where c and d are positive numbers, and r = 0 the Deming regression solution is obtained. If additionally c = d, the orthogonal distance regression solution, also known as major axis regression, is obtained.

Setting sd_x = sd(x), sd_y = sd(y) and r = 0 leads to the geometric mean regression solution, also known as reduced major axis regression or standardised major axis regression.

The goodness of fit metric chisq is a weighted reduced chi-squared statistic. It compares the deviations of the points from the fit line to the assigned measurement error standard deviations. If x and y are indeed related by a straight line, and if the assigned measurement errors are correct

(and normally distributed), then chisq will equal 1. A chisq > 1 indicates underfitting: the fit does not fully capture the data or the measurement errors have been underestimated. A chisq < 1 indicates overfitting: either the model is improperly fitting noise, or the measurement errors have been overestimated.

Value

An object of class "bfsl", which is a list containing the following components:

coefficients	A 2x2 matrix with columns of the fitted coefficients (intercept and slope) and their standard errors.
chisq	The goodness of fit (see Details).
control	The control list used, see the control argument.
convInfo	A list with convergence information.
call	The matched call.
data	A list containing x, y, sd_x, sd_y and r.

References

York, D. (1968). Least squares fitting of a straight line with correlated errors. *Earth and Planetary Science Letters*, 5, 320–324, https://doi.org/10.1016/S0012-821X(68)80059-7

Examples

```
x = pearson_york$x
y = pearson_york$y
sd_x = 1/sqrt(pearson_york$w_x)
sd_y = 1/sqrt(pearson_york$w_y)
bfsl(x, y, sd_x, sd_y)

fit = bfsl(pearson_york)
plot(fit)
```

---

bfsl_control *Controls the Iterations in the bfsl Algorithm*

Description

bfsl_control allows the user to set some characteristics of the bfsl best-fit straight line algorithm.

Usage

```
bfsl_control(tol = 1e-10, maxit = 100)
```

Arguments

`tol`	A positive numeric value specifying the tolerance level for the convergence criterion
`maxit`	A positive integer specifying the maximum number of iterations allowed.

Value

A `list` with two components named as the arguments.

See Also

`bfsl`

Examples

```
bfsl_control(tol = 1e-8, maxit = 1000)
```

`pearson_york` *Example data*

Description

Example data set of Pearson (1901) with weights suggested by York (1966).

Usage

```
pearson_york
```

Format

A data frame with 10 rows and 4 variables:

x *x* observations

w_x weights of *x*

y *y* observations

w_y weights of *y*

References

Pearson K. (1901), On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 59-572, https://doi.org/10.1080/1478644010946

York, D. (1966). Least-squares fitting of a straight line. *Canadian Journal of Physics*, 44(5), 1079–1086, https://doi.org/10.1139/p66-090

Examples

```
bfsl(pearson_york)
```

plot.bfsl *Plot Method for bfsl Results*

Description

Plot method for objects of class "bfsl".

Usage

```
## S3 method for class 'bfsl'
plot(x, grid = TRUE, ...)
```

Arguments

x	An object of class "bfsl".
grid	If TRUE (default) grid lines are plotted.
...	Further parameters to be passed to the plotting routines.

Details

plot.bfsl plots the data points with error bars and the calculated best-fit straight line.

print.bfsl *Print Method for bfsl Results*

Description

print method for class "bfsl".

Usage

```
## S3 method for class 'bfsl'
print(x, digits = max(3L, getOption("digits") - 3L), ...)
```

Arguments

x	An object of class "bfsl".
digits	The number of significant digits to use when printing.
...	Further arguments passed to print.default.

Index