Package 'snht'

September 23, 2014

Type Package

Title Standard Normal Homogeneity Test

Version 1.0

Date 2014-09-23

Depends Rcpp, RcppArmadillo, inline, MASS, zoo, plyr

Author Josh Browning

Maintainer Josh Browning <jbrownin@mines.edu>

Description Robust and non-robust SNHT tests for changepoint detection.

License What license is it under?

R topics documented:

`snht-package` *Robust and Non-Robust Standard Normal Homogeneity Test*

Description

Computes test statistics for the SNHT and robust SNHT test. Additionally, users may supply a custom function for estimating the mean and standard deviation, and this function will be used for computing the test statistic.

Details

Package:	snht
Type:	Package
Version:	1.0
Date:	2014-09-23
License:	What license is it under?

The main function is snht, which then calls the other functions in this package. However, users may also wish to call robustSNHT which allows for a custom estimator function.

Author(s)

Josh Browning

Maintainer: Josh Browning <jbrownin@mines.edu>

References

L. Haimberger. Homogenization of radiosonde temperature time series using innovation statistics. Journal of Climate, 20(7): 1377-1403, 2007.

`snht` *Standard Normal Homogeneity Test*

Description

This function performs a standard normal homogeneity test on the data supplied. This test searches the data for potential changepoints.

Usage

```
snht(data, period, robust = F, useC = F, time = NULL, ...)
robustSNHT(data, period, estimator = function(x, minObs = 5) {
    x = x[!is.na(x)]
    if (length(x) < minObs)
        return(c(NA, NA))
    if (max(table(x)) > length(x)/2)
        return(c(NA, NA))
    fit = MASS::huber(x)
    return(c(fit[[1]], fit[[2]]))
})
robustSNHTunequal(data, period, time, estimator = NULL)
```

Arguments

data	The data to be analyzed for changepoints.
period	The SNHT works by calculating the mean of the data on the previous period observations and the following period observations. Thus, this argument controls the window size for the test statistics.
robust	Flag indicating whether or not robust estimators should be used. If T, then Huber's robust estimator for the mean and variance will be used (see ?MASS::huber).
useC	Should the C++ implementation of the SNHT be used? This algorithm is much faster than the R equivalent, but is currently only implemented in cases where the data are equally spaced and the non-robust mean and sd are used.
time	Numeric vector specifying times for the observations. If not supplied, it is assumed that each observation occurs on one time period. If supplied, then the algorithm will create a new dataset with the same number of observations for each time unit by adding missing values.
estimator	A custom function may be supplied to this function which computes estimates for the mean and standard deviation. The function should only take one argument (a numeric vector of data) and should return a vector of length two: the estimated center and spread. The huber function from MASS is implemented for the robust SNHT by default (along with some data quality checks).
...	Additional arguments to pass to the robustSNHT function.

Details

The SNHT works by calculating the mean of the data on the previous period and on the following period. The test statistic at each observation is then computed as described in Haimberger (2007). Essentially, though, it just compares the means of these two periods and normalizes by the standard deviation.

Note: if there are not enough observations both before and after the current observation, no test is performed.

Large values of the test statistic suggests the presence of a changepoint. Haimberger (see references) suggests values larger than 100 should be considered changepoints.

Value

Returns a data.frame, with columns Statistic, leftMean, and rightMean. Statistic is the SNHT test statistic described above, and leftMean (rightMean) are the means to the left (right) of the current observation.

Additionally, if time is supplied, then time is returned on the output data.frame. Note that new (missing) observations were introduced to the dataset to ensure the same number of observations occur per day.

Author(s)

Josh Browning (jbrownin@mines.edu)

References

L. Haimberger. Homogenization of radiosonde temperature time series using innovation statistics. Journal of Climate, 20(7): 1377-1403, 2007.

See Also

huber, robustSNHT

Examples

```
data = rnorm(1000)
brk = sample(1000, size=1)
data[1:brk] = data[1:brk]-2
out = snht( data, period=50, robust=FALSE, useC=FALSE )
summary(out)

data = rnorm(1000)
time = 1:1000 + rnorm(1000)
brk = sample(1000, size=1)
data[1:brk] = data[1:brk]-2
out = snht( data, period=50, time=time, robust=FALSE, useC=FALSE )
summary(out)
```

Index