# Package 'washeR'

July 12, 2018

**Type** Package

**Title** Time Series Outlier Detection

**Version** 0.1.2

**Author** Andrea Venturini

**Maintainer** Andrea Venturini <andrea.venturini@bancaditalia.it>

**Description** Time series outlier detection by mean of non parametric test. Outlier detection regarding two methodologies: single time series variability (a vector) and grouped similar time series (a data frame). Andrea Venturini(2011) Statistica-Universita' Bologna, Vol.71, pp.329-344.

**License** GPL (>= 2)

**Encoding** UTF-8

**LazyData** true

**Imports** gplots,grDevices,graphics,stats,utils

**RoxygenNote** 6.0.1

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2018-07-12 16:00:03 UTC

## R topics documented:

## dati *Data frame of meteorological data*

### Description

A dataset containing values to test outlier detection on meteorological number completely random

### Usage

```
dati
```

### Format

A data frame with 800 rows and 4 variables:

**phen** Temperature, Rain

**time** ordered numbers for time

**zone** label classification for group data

**value** numbers for values

## ts *Time series*

### Description

A dataset containing a random time series with increasing trend and some variability

### Usage

```
ts
```

### Format

A data frame with 35 rows and 1 variable:

**dati** pseudo random numers

wash.out *Time Series Outlier Detection (washer)*

## Description

This function provides a new outlier detection methodology (washer): efficient for timesaving elaboration and implementation procedures, adaptable for general assumptions and for needing very short time series, reliable and effective as involving robust non parametric test. You can input a vector or a data frame with ordered information (as showed below). See: Andrea Venturini, "Time Series Outlier Detection: A New Non Parametric Methodology (washer)" Statistica — Universita' di Bologna, Vol. 71, 2011, pp. 329-344.

## Usage

```
wash.out(dati, graph = FALSE, linear_analysis = FALSE, val_test_limit = 5,
  save_out = FALSE, out_out = "out.csv", pdf_out = "out.pdf", r_out = 3,
  c_out = 2, first_line = 1, pace_line = 6)
```

## Arguments

| | |
|---|---|
| dati | A data frame (grouped time series: phenomenon+date+group+values) or a vector (single time series) |
| graph | A logical for graphical analysis (default=FALSE) |
| linear_analysis | A logical for linear analysis (default=FALSE) |
| val_test_limit | A number for testing outlier (default=5) |
| save_out | A logical for saving outliers (default=FALSE) |
| out_out | A character file name for saving outliers in csv form ";" delimited and ',' for decimal (default="out.csv") |
| pdf_out | A character file name for saving graphic analysis in pdf form (default="out.pdf") |
| r_out | A number of rows in graphic analysis (default=3) |
| c_out | A number of cols in graphic analysis (default=2) |
| first_line | A number for first dotted line in graphic analysis (default=1) |
| pace_line | A number for pace in dotted line in graphic analysis (default=6) |

## Value

A data frame of possible outliers in a triad. Output record: rows /time.2/series/y1/y2/y3/test(AV)/AV/n /median(AV)/mad(AV)/madindex(AV). Where time.2 is the center of the triad y1, y2, y3; test(AV) is the number that over 5 detect outlier; n is the number of observations of the group ....

## Examples

```
## data without outliers but structured with co-movement between groups
data("dati")
## first column for phenomenon
## 2° col for time written in ordered number or strings
## 3° col for group classification variable
## 4° col for values
str(dati)
## 1° -> data.frame + no outlier
out=wash.out(dati)
out
length(out[,1])
## add two outliers
####  time=3 temperature value=0
dati[99,4]=  0
## ... and after for "rain" phenomenon!
####  time=3 rain value=37
dati[118,4]=  37
##  2° -> data.frame + 2 outliers
out=wash.out(dati)
##  all "three terms" time series
## let's take a look at anomalous time series
out
## ... the same but save results in a specified file....
## If you don't specify the file the name is by default "out.csv"
out=wash.out(dati,save_out=TRUE,out_out="tabel_out.csv")
out
## put the limit from 5 to 10 to capture only the worst outliers
out=wash.out(dati, val_test_limit = 10  )
out
## save plots and outliers in a pdf file "out.pdf" as a default
out=wash.out(dati, val_test_limit = 10,graph=TRUE)
out
## make the usual analysis for groups but also that for every single time series
## (linear_analysis): two files for saved outliers ("out.csv" and "linout.csv")
##  and for graph display in two pdf files ("out.pdf" and "linout.pdf")
out=wash.out(dati, val_test_limit = 5, save_out=TRUE,
          linear_analysis=TRUE  ,graph=TRUE)
out
## out return only the second linear analysis...
############################################################
##  single time series analysis
############################################################
data(ts)
str(ts)
dati= ts$dati
plot(dati,type="b",pch=20,col="red")
## a time series with a variability and an increasing trend
## dati is a vector and linear analysis is a default
out=wash.out(dati)
out
## no outlier
```

```
out=wash.out(dati, val_test_limit = 5,linear_analysis=TRUE   ,graph=TRUE)
out
## no outlier
## add an outlier with limited amount
dati[5]=dati[5]*2
plot(dati,type="b",pch=20,col="red")
out=wash.out(dati, val_test_limit = 5)
out
## test is over 5 for a bit
out=wash.out(dati, val_test_limit = 5,save_out=TRUE ,graph=TRUE)
out
data(ts)
dati= ts$dati
dati[5]=dati[5]*3
## try a greater outlier
plot(dati,type="b",pch=20,col="blue")
out=wash.out(dati, val_test_limit = 5,save_out=TRUE ,graph=TRUE)
out
## washer identify three triads of possible outliers
```

---

washeR *washeR.*

---

## Description

Time series outlier detection using non parametric test. This methodology is based on three points tendence to fall or not to fall into line. In grouped time series we use the grouped tendence to align or not with a very small time range. If we use the linear analysis we use the general variance inside the time series to test if a single variation is too heavy or not. For other informations see: A. Venturini (2011) (https://rivista-statistica.unibo.it/article/download/3617/2968) but also (https://sites.google.com/site/andreaventurini65/home/outlier-detection) or (https://www.r-bloggers.com/time-series-outlier-detection-a-simple-r-function/) for an informal explanation.

## References

Andrea Venturini "Time Series Outlier Detection: A New Non Parametric Methodology (washer)" Statistica — Universita' di Bologna, Vol. 71, 2011, pp. 329-344.

# Index