Package ‘Peirce’

September 7, 2012

Type  Package

Title  Functions for removing outliers, with illustrations

Version  0.5

Date  2012-08-26

Author  Christopher Dardis

Maintainer  <christopherdardis@gmail.com>

Description  Peirces and Chauvenets functions for removing outliers with examples and illustrations.

Depends  R (>= 2.13.0), NORMT3

Imports  ggplot2, gridExtra, Hmisc, outliers, pracma, rgl,RColorBrewer, scatterplot3d

Suggests  emdbook, outliers, TeachingDemos, compositions

License  GPL (>= 2)

LazyLoad  yes

R topics documented:

Peirce-package ......................................................... 2
Chauvenet ............................................................... 2
libor ................................................................. 3
liborUSD3M ............................................................ 4
NIST ................................................................. 5
NlogQ ............................................................... 5
NlogQLimits .......................................................... 6
Peirce ............................................................... 6
Peirce1852 ............................................................ 7
PeirceGould .......................................................... 8
PeirceLimits .......................................................... 9
PeirceVsChauvenet .................................................... 10
PeirceVsLibor ......................................................... 10
Ross ............................................................... 11

Index  12
Description

Peirce and Chauvenet functions for removal of outliers. Examples and illustrations from the original papers.

Details

<table>
<thead>
<tr>
<th>Package:</th>
<th>Peirce</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type:</td>
<td>Package</td>
</tr>
<tr>
<td>Version:</td>
<td>0.5</td>
</tr>
<tr>
<td>Date:</td>
<td>2012-08-26</td>
</tr>
<tr>
<td>License:</td>
<td>GPL (&gt;= 2)</td>
</tr>
<tr>
<td>LazyLoad:</td>
<td>yes</td>
</tr>
</tbody>
</table>

Functions for removing outliers: Peirce, Chauvenet. Other functions are provided as examples and illustrations.

Author(s)

Christopher Dardis Maintainer: christopherdardis@gmail.com

References


Examples

```r
v1 <- c(101.2, 90.0, 99.0, 102.0, 103.0, 100.2, 89.0, 98.1, 101.5, 102.0)
Peirce(v1) # 2 outliers removed
Chauvenet(v1) # 2 outliers removed
```

Description

Remove outliers from a vector based on Chauvenets criterion.

Usage

```r
Chauvenet(v, loop = FALSE)
```
Arguments

v  v vector of numeric values
loop  loop logical value; if TRUE, process will be repeated until no more values can be removed

Details

Calculates z value for each value in the vector (from mean and standard deviation). Then removes elements where the \text{erfc} of the z value is >0.5. (\text{erfc} = \text{complementary error function}). If loop = TRUE, this process will be repeated until no more values can be removed.

Value

A numeric vector, with outliers removed; preserves order of original vector. The vector will contain at least two values.

References


Examples

v1 <- c(1/0.0, 9/0.0, 99.0, 1/0.2, 1/0.3, 1/0.2, 89.0, 98.1, 1/0.15, 1/0.2)
Chauvenet(v1) # 2 outliers removed
Chauvenet(v1, loop=TRUE) # 8 outliers removed (the maximum)

libor  \textit{bbalibor function to remove outliers then return mean of those remaining.}

Description

Remove outliers from a vector based on bbalibor criterion, then calculate mean of those remaining. Number removed is based on length of vector.

• If 15-18, 4 highest and lowest removed.
• If 11-14, 3 highest and lowest removed.
• If 8-10, 2 highest and lowest removed.
• If 6-7, 1 highest and lowest removed.

Usage

libor(v)

Arguments

v  v vector of numeric values, length 6-18

Value

Numeric. The mean of the vector with outliers removed.
References

bbalibor explained

See Also

PeirceVsLibor

Examples

libor(runif(n=16, min=0, max=1))

---

liborUSD3M  LIBOR rate in US Dollars for 3 month loans

Description

LIBOR rate submissions for 16 banks during a 3 month period in 2008. Each submission reflects the rate of interest the bank would expect to pay on a 3 month loan in US Dollars.

Usage

data(liborUSD3M)

Format

A data frame with 496 observations on the following 3 variables.

- Bank  a character vector
- Date  a character vector
- X3M  a character vector

Source

Data on google docs spreadsheet

References

Guardian newspaper 2012

Examples

data(liborUSD3M)
str(liborUSD3M)

## Not run:
liborUSD3M <- transform(liborUSD3M, Date = as.Date(Date, format="%m/%d/%Y")) # convert to date format
liborUSD3M <- transform(liborUSD3M, X3M = as.numeric(X3M)) # convert to numeric
liborUSD3M <- transform(liborUSD3M, Bank = as.factor(Bank)) # convert to factor

## End(Not run)
**NIST outliers dataset**

**Description**
Outliers dataset from NIST. 90 ordered observations.

**Usage**
data(NIST)

**Format**
The format is: num [1:90] 30 171 184 201 212 250 265 270 272 289 ...

**Source**
NIST outliers example

**Examples**
data(NIST)
str(NIST)
hist(NIST)

---

**NlogQ**

**Description**
Calculate the value of NlogQ.

**Usage**
NlogQ(N, k)

**Arguments**
N Number of observations in dataset.
k Number of observations proposed to be rejected in dataset.

**Details**
Calculates NlogQ, as given in Table III of Goulds paper below. This value is used in the calculation of R, the maximum allowable absolute error for a value in a vector of observations.

**Value**
Numeric.
References

Examples
NlogQ(N=10, k=1) # = 8.58818...

Description
Produces a 3d rotatable plot using rgl which gives values of NlogQ for various values of N and k.

Usage
NlogQlimits()

Details
Produces a 3d rotatable plot using rgl which gives values of NlogQ for various values of N and k.
N = number of observations, k = numer of outliers proposed to be rejected.

Value
A 3d plot with rgl.

See Also
See Also NlogQ

Examples
## Not run: NlogQlimits()

Peirce
Peirces function to remove outliers

Description
Remove outliers from a vector based on Peirce criterion.

Usage
Peirce(v, m = 1)

Arguments
v v vector of numeric values
m m number of unknown quantities
Details

Calculates $R$, the maximum allowable absolute error for a value in the vector, and removes those values exceeding this value. $m$, the 'number of unknown quantities', may be treated as the number of independent processes giving rise to errors. For practical purposes $m$ should generally be restricted to 1.

Value

A numeric vector, with outliers removed; ordered by value of absolute error.

See Also

See Also PeirceGould, PeirceVsChauvenet

Examples

```r
## Not run:
v1 <- c(101.2, 90.0, 99.0, 102.0, 103.0, 100.2, 89.0, 98.1, 101.5, 102.0)
Peirce(v1) # 2 outliers removed
## End(Not run)
```

Description

Produce plots of functions from Peirces 1852 paper. Initial two plots are with Hmisc::labcurve. Subsequent plots are with ggplot2. Produces plots corresponding to his functions for $\phi$ and $\psi$ and compares the latter with erfc, the complementary error function.

Usage

```
Peirce1852(width = 1366, height = 768)
```

Arguments

- **width**: width width of graphical display (screen) in pixels
- **height**: height height of graphical display (screen) in pixels

Value

Plots two curves with Hmisc::labcurve and returns a list of 4 plots from ggplot2.

Note

The tail end of function contains attempts to generate probability functions from Peirces paper. These values are not returned.

References

PeirceGould

See Also

Peirce

Examples

## Not run: Peirce1852()

---

**Description**

Remove outliers from a vector based on Peirce's criterion. Method follows the original technique in the paper by Gould below.

**Usage**

PeirceGould(v, m = 1)

**Arguments**

- `v`: vector of numeric values
- `m`: number of unknown quantities

**Details**

Calculates $R$, the maximum allowable absolute error for a value in the vector, and removes those values exceeding this value. $m$, the 'number of unknown quantities', may be treated as degrees of freedom of the system of observations, or the number of independent processes giving rise to errors. For practical purposes $m$ should generally be restricted to 1. For efficiency, `Peirce` is preferred.

**Value**

A numeric vector, with outliers removed; ordered by value of absolute error.

**References**


**See Also**

Peirce

**Examples**

```r
v1 <- c(101.2, 90.0, 99.0, 102.0, 103.0, 100.2, 89.0, 98.1, 101.5, 102.0)
Peirce(v1) # 2x outliers removed
```
PeirceLimits

Plot limits of Peirce’s function

Description

Produces 3d plots using rgl (rotatable) and scatterplot3d which gives values of \( R \) for various values of \( N \) and \( m \). \( R \) is the maximum allowable absolute error for a value in a vector of observations. \( N \) is the number of observations (length of vector). \( m \), the ‘number of unknown quantities’, may be treated as the number of independent processes giving rise to errors in the vector.

Usage

PeirceLimits(N=1000L, plots=TRUE, noPoints=100)

Arguments

- **N** is the number of observations (length of vector)
- **plots** if TRUE then will display 3d plots
- **noPoints** no of intervals into which to divide \( k \) and \( n \). Defaults to 100, corresponding to 100% of \( N \). Larger values will give a more detailed resolution for the plotted image. Set to the same value as \( N \) to generate results for every combination of \( k \) and \( m \) up to \( N \)
- **asPercent** if TRUE then will display returned values of \( k \) and \( m \) as percentages of \( N \)

Value

3d plots with rgl and scatterplot3d. A dataframe with value of \( R \) for a given combination of \( k \) and \( m \).

Note

This can take some time to run for large values of \( N \) e.g. >10000.

See Also

Peirce

Examples

```r
## Not run: PeirceLimits()
```
PeirceVsChauvenet

Description

Compares Peirce's function with Chauvenet's for four datasets. Plots these datasets using base graphics and \texttt{ggplot2}. Returns a matrix giving number of observations removed by each method.

Usage

\texttt{PeirceVsChauvenet(width = 1366, height = 768)}

Arguments

- \texttt{width}: width width of graphical display (screen) in pixels
- \texttt{height}: height height of graphical display (screen) in pixels

Value

A matrix giving number of observations removed by each method.

See Also

\texttt{Peirce}

Examples

```r
## Not run: PeirceVsChauvenet()
```

PeirceVsLibor

Description

Compares Peirce's function with bbLibor for an example dataset. Peirce and libor function are used to remove outliers, then the mean is calculated from the remaining values. Plots the number of observations removed by each method for each date with base graphics. Plots the dataset, including averaged values by each method, using \texttt{ggplot2}. Returns a data frame giving number of observations removed by Peirce's method for each date.

Usage

\texttt{PeirceVsLibor(width = 1366, height = 768)}

Arguments

- \texttt{width}: width width of graphical display (screen) in pixels
- \texttt{height}: height height of graphical display (screen) in pixels
Ross

Value

A data frame giving number of observations removed by Peirces method for each date.

Author(s)

Christopher Dardis

See Also

liborUSD3M

Examples

```r
## Not run: PeirceVsLibor()
```

---

Ross

Outliers dataset from Ross paper

Description

Outliers dataset from Ross paper. 10 observations from a pressure gage, taken in an experiment at one setting.

Usage

data(Ross)

Format

The format is: num [1:10] 101 90 99 102 103 ...

Source


Examples

```r
data(Ross)
hist(Ross)
```
Index

*Topic **datasets**
  - liborUSD3M, 4
  - NIST, 5
  - Ross, 11
*Topic **package**
  - Peirce-package, 2

BBALibor (libor), 3
BBAlibor (libor), 3

Chauvenet, 2

LIBOR (libor), 3
Libor (libor), 3
libor, 3
liborUSD3M, 4, 11

NIST, 5
NlogQ, 5, 6
NlogQLimits, 6

Peirce, 6, 8–10
Peirce-package, 2
Peirce1852, 7
PeirceGould, 7, 8
PeirceLimits, 9
PeirceVsChauvenet, 7, 10
PeirceVsLibor, 4, 10

Ross, 11