Package ‘lawstat’

April 1, 2019

Type Package
Title Tools for Biostatistics, Public Policy, and Law
Version 3.3
Date 2019-03-31
Depends R (>= 2.6.0)
Imports Kendall, mvtnorm, Rdpack, VGAM
Suggests fBasics
License GPL (>= 2)
Description Statistical tests widely utilized in biostatistics, public policy, and law. Along with the well-known tests for equality of means and variances, randomness, and measures of relative variability, the package contains new robust tests of symmetry, omnibus and directional tests of normality, and their graphical counterparts such as robust QQ plot, robust trend tests for variances, etc. All implemented tests and methods are illustrated by simulations and real-life examples from legal statistics, economics, and biostatistics.

RdMacros Rdpack
RoxygenNote 6.1.1
Encoding UTF-8

Author Joseph L. Gastwirth [aut],
Yulia R. Gel [aut],
W. L. Wallace Hui [aut],
Vyacheslav Lyubchich [aut, cre],
Weiwen Miao [aut],
Kimihiro Noguchi [aut]

Maintainer Vyacheslav Lyubchich <lyubchic@umces.edu>
Repository CRAN

Date/Publication 2019-04-01 04:40:16 UTC
**Description**

Bartels (1982) test for randomness that is based on the ranked version of von Neumann’s ratio (RVN). Users can choose whether to test against two-sided, negative, or positive correlation. NAs from the data are omitted.

**Usage**

```r
bartels.test(y, alternative = c("two.sided", "positive.correlated", "negative.correlated"))
```
Arguments

- **y**: a numeric vector of data values.
- **alternative**: a character string specifying the alternative hypothesis, must be one of "two.sided" (default), "negative.correlated", or "positive.correlated".

Value

A list of class "htest" with the following components:

- **statistic**: the value of the standardized Bartels statistic.
- **parameter**: RVN ratio.
- **p.value**: the p-value for the test.
- **data.name**: a character string giving the names of the data.
- **alternative**: a character string describing the alternative hypothesis.

Author(s)

Kimihiro Noguchi, Wallace Hui, Yulia R. Gel, Joseph L. Gastwirth, Weiwen Miao

References


See Also

- `runs.test`

Examples

```r
## Simulate 100 observations from an autoregressive model of
## the first order AR(1)
y = arima.sim(n = 100, list(ar = c(0.5)))

## Test y for randomness
bartels.test(y)

## Sample Output
##
## Bartels Test - Two sided
## data: y
## Standardized Bartels Statistic -4.4929, RVN Ratio =
## 1.101, p-value = 7.024e-06
```
**bias**  
*Prediction Errors ("Biases") of Surface Temperature Forecasts*

**Description**
Prediction errors of 48-hour ahead MM5 forecasts of surface temperature measured at 96 locations in the US Pacific Northwest on 3-January-2000. The prediction error, or "bias", is the difference between the forecasted and observed surface temperature. (MM5 is the fifth-generation Pennsylvania State University – National Center for Atmospheric Research Mesoscale Model.)

**Usage**
data(bias)

**Format**
A numeric vector of length 96.

**Source**
The data were kindly provided by the research group of Professor Clifford Mass in the Department of Atmospheric Sciences at the University of Washington. Detailed information about the Pacific Northwest prediction effort and the associated data archive can be found online at [www.atmos.washington.edu/mm5rt/info.html](http://www.atmos.washington.edu/mm5rt/info.html) and [www.atmos.washington.edu/marka/pnw.html](http://www.atmos.washington.edu/marka/pnw.html), respectively.

---

**blackhire**  
*Hiring Data for Eight Professions and Two Races*

**Description**
Number of black and white candidates (hired or rejected) for eight professions (Gastwirth 1984).

**Usage**
data(blackhire)

**Format**
An array with 2 rows by 2 columns by 8 levels.

**References**
**Description**

The Brunner–Munzel test for stochastic equality of two samples, which is also known as the Generalized Wilcoxon Test. NAs from the data are omitted.

**Usage**

```r
brunner.munzel.test(x, y, alternative = c("two.sided", "greater", "less"), alpha = 0.05)
```

**Arguments**

- `x` the numeric vector of data values from the sample 1.
- `y` the numeric vector of data values from the sample 2.
- `alternative` a character string specifying the alternative hypothesis, must be one of "two.sided" (default), "greater" or "less". User can specify just the initial letter.
- `alpha` significance level, default is 0.05 for 95% confidence interval.

**Details**

There exist discrepancies with Brunner and Munzel (2000) because there is a typo in the paper. The corrected version is in Neubert and Brunner (2007) (e.g., compare the estimates for the case study on pain scores). The current function follows Neubert and Brunner (2007).

**Value**

A list of class "htest" with the following components:

- `statistic` the Brunner–Munzel test statistic.
- `parameter` the degrees of freedom.
- `conf.int` the confidence interval.
- `p.value` the p-value of the test.
- `data.name` a character string giving the name of the data.
- `estimate` an estimate of the effect size, i.e., \( P(X < Y) + 0.5 \times P(X = Y) \).

**Author(s)**

Wallace Hui, Yulia R. Gel, Joseph L. Gastwirth, Weiwen Miao. This function was updated with the help of Dr. Ian Fellows.
References


See Also

wilcox.test, pwilcox

Examples

```r
## Pain score on the third day after surgery for 14 patients under
## the treatment Y and 11 patients under the treatment N
## (see Brunner and Munzel, 2000; Neubert and Brunner, 2007).

Y <- c(1, 2, 1, 1, 1, 1, 1, 1, 1, 2, 4, 1, 1)
N <- c(3, 3, 4, 3, 1, 2, 3, 1, 1, 5, 4)

cd(Y, N)
```

Description

Coefficient of Dispersion – a Measure of Relative Variability

Measure of relative inequality (or relative variation) of the data. Coefficient of dispersion (CD) is the ratio of the mean absolute deviation from the median (MAAD) to the median of the data. NAs from the data are omitted. See Gastwirth (1988) and Bonett and Seier (2006).

Usage

cd(x)

Arguments

- `x` a numeric vector of data values.
The coefficient of dispersion.

Wallace Hui, Yulia R. Gel, Joseph L. Gastwirth, Weiwen Miao


The Cochran–Mantel–Haenszel (CMH) procedure tests homogeneity of population proportions after taking into account other factors. This procedure is widely used in law cases, for example, on equal employment and discrimination, and in biological and pharmaceutical studies.

$\text{cmh.test}(x)$

$x$ a numeric $2 \times 2 \times k$ array of data values.
Details

The test is based on the CMH procedure discussed by Gastwirth (1984). The data should be input in an array of 2 rows \( \times \) 2 columns \( \times k \) levels. The output includes the Mantel–Haenszel Estimate, the pooled Odd Ratio, and the Odd Ratio between the rows and columns at each level. The Chi-square test of significance tests if there is an interaction or association between rows and columns.

The null hypothesis is that the pooled Odd Ratio is equal to 1, i.e., there is no interaction between rows and columns. For more details see Gastwirth (1984).

The \texttt{cmh.test} can be viewed as a subset of \texttt{mantelhaen.test}, in the sense that \texttt{cmh.test} is for a 2 by 2 by \( k \) table without continuity correction, whereas \texttt{mantelhaen.test} allows for a larger table, and for a 2 by 2 by \( k \) table, it has an option of performing continuity correction. However, in view of Gastwirth (1984), continuity correction is not recommended as it tends to overestimate the \( p \)-value.

Value

A list of class "htest" containing the following components:

- \texttt{MH.estimate} the value of the Cochran–Mantel–Haenszel estimate.
- \texttt{or} pooled Odd Ratio of the data.
- \texttt{ork} vector of Odd Ratio of each level.
- \texttt{cmh} the test statistic.
- \texttt{df} degrees of freedom.
- \texttt{p.value} the \( p \)-value of the test.
- \texttt{method} type of the performed test.
- \texttt{data.name} a character string giving the name of the data.

Author(s)

Min Qin, Wallace W. Hui, Yulia R. Gel, Joseph L. Gastwirth

References


See Also

\texttt{mantelhaen.test}

Examples

```r
## Sample Salary Data
data(blackhire)
cmh.test(blackhire)
```
**data1963**  
*Population Size and Number of Senators and Representatives in 1963*

**Description**
Number of senators and representatives and population size in 23 districts in the United States of America in 1963 (Gastwirth 1972).

**Usage**
```r
data(data1963)
```

**Format**
A data frame with 23 observations on the following 3 variables:
- `pop1963` population in 1963;
- `sen1963` number of senators in the district in 1963;
- `rep1963` number of representatives in the district in 1963.

**Source**
Gastwirth (1972).

**References**

---

**gini.index**  
*Measures of Relative Variability – Gini Index*

**Description**
Gini index for measuring relative inequality (or relative variation) of the data (Gini 1912). NAs from the data are omitted.

**Usage**
```r
gini.index(x)
```

**Arguments**
- `x` the input data.
Details

See also Gastwirth (1988).

Value

A list with the following components:

- statistic: the Gini index.
- parameter: the mean difference of the set of numbers.
- data.name: a character string giving the name of the data.

Author(s)

Wallace Hui, Yulia R. Gel, Joseph L. Gastwirth, Weiwen Miao

References


See Also

cd, j.maad, lorenz.curve

Examples

```r
## The Baker v. Carr Case: one-person-one-vote decision.
## Measure of Relative Inequality of Population data in 33 districts
## of the Tennessee Legislature in 1900 and 1972. See
## popdata (see Gastwirth (1988)).
data(popdata)
gini.index(popdata[,"pop1900")
gini.index(popdata[,"pop1972")
```

---

**j.maad**  
*MAAD Robust Standard Deviation*

Description

Compute average absolute deviation from the sample median, which is a consistent robust estimate of the population standard deviation for normally distribution data (Gastwirth 1982). NA's from the data are omitted.
Usage

j.maad(x)

Arguments

x a numeric vector of data values.

Value

Robust standard deviation.

Author(s)

Wallace Hui, Yulia R. Gel, Joseph L. Gastwirth, Weiwen Miao

References


See Also

cd, gini.index, rqq, rjb.test, sj.test

Examples

```r
# Sample 100 observations from the standard normal distribution
x = rnorm(100)
j.maad(x)
```

---

### laplace.test

Goodness-of-fit Test Statistics for the Laplace Distribution

Description

Goodness-of-fit test statistics A2 (Anderson–Darling), W2 (Cramer–von Mises), U2 (Watson), D (Kolmogorov–Smirnov), and V (Kuiper). By default, NAs are omitted. For the tables of critical values, see Stephens (1986) and Puig and Stephens (2000).

Usage

```
laplace.test(y)
```

Arguments

y a numeric vector of data values.
Value

A list with the following numeric components:

- $A_2$ the Anderson–Darling statistic.
- $W_2$ the Cramer–von Mises statistic.
- $U_2$ the Watson statistic.
- $D$ the Kolmogorov–Smirnov statistic.
- $V$ the Kuiper statistic.

Author(s)

Kimihiro Noguchi, Yulia R. Gel

References


See Also

plaplace

Examples

```r
## Differences in flood levels example taken from Puig and Stephens (2000)
y <- c(1.96,1.97,3.60,3.80,4.79,5.66,5.76,5.78,6.27,6.30,6.76,7.65,7.84,7.99,8.51,9.18,
    10.13,10.24,10.25,10.43,11.45,11.48,11.75,11.81,12.33,12.78,13.06,13.29,13.98,14.18,
    14.40,16.22,17.80)
laplace.test(y)$D
## [1] 0.9177726
## The critical value at the 0.05 significance level is approximately 0.906.
## Thus, the null hypothesis should be rejected at the 0.05 level.
```

---

**levene.test**

Levene’s Test of Equality of Variances

Description

Tests equality of the $k$ population variances.
Usage

levene.test(y, group, location = c("median", "mean", "trim.mean"),
    trim.alpha = 0.25, bootstrap = FALSE, num.bootstrap = 1000,
    kruskal.test = FALSE, correction.method = c("none",
        "correction.factor", "zero.removal", "zero.correction"))

Arguments

y a numeric vector of data values.
group factor of the data.
location the default option is "median" corresponding to the robust Brown–Forsythe Levene-type procedure (Brown and Forsythe 1974); "mean" corresponds to the classical Levene's procedure (Levene 1960), and "trim.mean" corresponds to the robust Levene-type procedure using the group trimmed means.
trim.alpha the fraction (0 to 0.5) of observations to be trimmed from each end of x before the mean is computed.
bootstrap a logical value identifying whether to implement bootstrap. The default is FALSE, i.e., no bootstrap; if set to TRUE, the bootstrap method described in Lim and Loh (1996) for Levene's test is applied.
num.bootstrap number of bootstrap samples to be drawn when the bootstrap argument is set to TRUE. The default value is 1000.
kruskal.test logical value identifying whether to use the Kruskal–Wallis statistic. The default option is FALSE, i.e., the usual ANOVA statistic is used.
correction.method procedures to make the test more robust; the default option is "none"; "correction.factor" applies the correction factor described by O'Brien (1978) and Keyes and Levy (1997); "zero.removal" performs the structural zero removal method by Hines and Hines (2000); "zero.correction" performs a combination of the O'Brien's correction factor and the Hines–Hines structural zero removal method (Noguchi and Gel 2010). Note that the options "zero.removal" and "zero.correction" are only applicable when the location is set to "median", otherwise, "none" is applied.

Details

The test statistic is based on the classical Levene’s procedure (using the group means), the modified Brown–Forsythe Levene-type procedure (using the group medians), or the modified Levene-type procedure (using the group trimmed means). More robust versions of the test using the correction factor or structural zero removal method are also available. Two options for calculating critical values, namely, approximated and bootstrapped, are available. By default, NAs are omitted from the data.

Value

A list of class "htest" with the following components:

statistic the value of the test statistic.
levene.test

- **p.value**: the \( p \)-value of the test.
- **method**: type of test performed.
- **data.name**: a character string giving the name of the data.
- **non.bootstrap.p.value**: the \( p \)-value of the test without bootstrap method; i.e. the \( p \)-value using the approximated critical value.

**Note**

Instead of the ANOVA statistic suggested by Levene, the Kruskal–Wallis ANOVA may also be applied using this function (see the parameter `kruskal.test`).

Modified from a response posted by Brian Ripley to the R-help e-mail list.

**Author(s)**

Kimihiro Noguchi, W. Wallace Hui, Yulia R. Gel, Joseph L. Gastwirth, Weiwen Miao

**References**


**See Also**

`neuhauser.hothorn.test`, `lnested.test`, `ltrend.test`, `mma.test`, `robust.mmm.test`

**Examples**

```r
data(pot)
levene.test(pot[,"obs"], pot[,"type"],
  location = "median", correction.method = "zero.correction")
```

## Bootstrap version of the test. The calculation may take up a few minutes
## Description

The test statistic is based on the finite intersection approach.

## Usage

```r
lnested.test(y, group, location = c("median", "mean", "trim.mean"),
    tail = c("right", "left", "both"), trim.alpha = 0.25,
    bootstrap = FALSE, num.bootstrap = 1000,
    correction.method = c("none", "correction.factor", "zero.removal",
    "zero.correction"), correlation.method = c("pearson", "kendall",
    "spearman"))
```

## Arguments

- **y**: a numeric vector of data values.
- **group**: factor of the data.
- **location**: the default option is "median" corresponding to the robust Brown–Forsythe Levene-type procedure (Brown and Forsythe 1974); "mean" corresponds to the classical Levene’s procedure (Levene 1960), and "trim.mean" corresponds to the robust Levene-type procedure using the group trimmed means.
- **tail**: the default option is "right", corresponding to an increasing trend in variances as the one-sided alternative; "left" corresponds to a decreasing trend in variances, and "both" corresponds to any (increasing or decreasing) monotonic trend in variances as the two-sided alternative.
- **trim.alpha**: the fraction (0 to 0.5) of observations to be trimmed from each end of x before the mean is computed.
- **bootstrap**: a logical value identifying whether to implement bootstrap. The default is FALSE, i.e., no bootstrap; if set to TRUE, the bootstrap method described in Lim and Loh (1996) for Levene’s test is applied.
- **num.bootstrap**: number of bootstrap samples to be drawn when the bootstrap argument is set to TRUE. The default value is 1000.
- **correction.method**: procedures to make the test more robust; the default option is "none"; "correction.factor" applies the correction factor described by O’Brien (1978) and Keyes and Levy (1997); "zero.removal" performs the structural zero removal method by Hines and Hines (2000); "zero.correction" performs a combination of the O’Brien’s...
correction factor and the Hines–Hines structural zero removal method (Noguchi
and Gel 2010). Note that the options "zero.removal" and "zero.correction"
are only applicable when the location is set to "median", otherwise, "none" is
applied.

correlation.method
measures of correlation; the default option is "pearson", the linear correlation
coefficient that is equivalent to the t-test; nonparametric measures of correlation
such as "kendall" (Kendall’s tau) or "spearman" (Spearman’s rho) may also
be chosen.

Details
The test statistic is based on the classical Levene’s procedure (using the group means), the modified
Brown–Forsythe Levene-type procedure (using the group medians), or the modified Levene-type
procedure (using the group trimmed means). More robust versions of the test using the correction
factor or structural zero removal method are also available. Two options for calculating critical
values, namely, approximated and bootstrapped, are available. By default, NAs are omitted from the
data.

Value
A list with the following elements:

- **T** the statistic and \( p \)-value of the test based on the Tippett \( p \)-value combination.
- **F** the statistic and \( p \)-value of the test based on the Fisher \( p \)-value combination.
- **N** the statistic and \( p \)-value of the test based on the Liptak \( p \)-value combination.
- **L** the statistic and \( p \)-value of the test based on the Mudholkar-George \( p \)-value com-

Each of the list elements is a list of class "htest" with the following elements:

- **statistic** the value of the test statistic expressed in terms of correlation (Pearson, Kendall,
or Spearman).
- **p.value** the \( p \)-value of the test.
- **method** type of test performed.
- **data.name** a character string giving the name of the data.
- **non.bootstrap.statistic** the statistic of the test without bootstrap method.
- **non.bootstrap.p.value** the \( p \)-value of the test without bootstrap method.

Author(s)
Kimihiro Noguchi, W. Wallace Hui, Yulia R. Gel, Joseph L. Gastwirth, Weiwen Miao
**lorenz.curve**

**Lorenz Curve**

**Description**

Plots the Lorenz curve that is a graphical representation of the cumulative distribution function. The user can choose between the Lorenz curve with single (default) or multiple weighting of data, for example, taking into account for single or multiple legislature representatives (Gastwirth 1972).

**Usage**

```r
lorenz.curve(data, weight = NULL, mul = FALSE, plot.it = TRUE,
             main = NULL, xlab = NULL, ylab = NULL, xlim = c(0, 1),
             ylim = c(0, 1), ...)```

**References**


**See Also**

`levene.test`, `ltrend.test`, `mma.test`, `neuhauser.hothorn.test`, `robust.mmm.test`
Arguments

`data`    input data. If the argument is an array, a matrix, a data.frame, or a list with two or more columns, then the first column will be treated as a data vector, and the second column to be treated as a weight vector. A separate weight vector is then ignored and not required. If the argument is a single column vector, then a user must enter a separate single-column weight vector. NAs or character are not allowed.

`weight`  one-column vector contains factors of single or multiple weights. Ignored if included in the `data` argument. NAs or character are not allowed.

`mul`     logical value indicates whether the Lorenz curve with multiple weight is to be plotted. Default is `FALSE`, i.e., single.

`plot.it` logical value indicates whether the Lorenz curve should be plotted. Default is `TRUE`, i.e., to plot.

`main`    title of Lorenz curve. Only required if user wants to override the default value.

`xlab`    label of x-axis. Only required if user wants to override the default value.

`ylab`    label of y-axis. Only required if user wants to override the default value.

`xlim`    plotting range of x-axis. Only required if user wants to override the default value.

`ylim`    plotting range of y-axis. Only required if user wants to override the default value.

...       other graphical parameters to be passed to the `plot` function.

Details

The input data should be a data frame with 2 columns. The first column will be treated as data vector, and the second column to be treated as a weight vector. Alternatively, data and weights can be entered as separate one-column vectors.

Value

A Lorenz curve plot with x-axis being the cumulative fraction of the data argument, and y-axis being the cumulative fraction of the weight argument. In the legend to the plot, the following values are reported:

`RMD`     relative mean deviation of the input data.

`GI`      the Gini index of the input data.

`L(1/2)`  median of the cumulative fraction sum of the data.

Author(s)

Man Jin, Wallace W. Hui, Yulia R. Gel, Joseph L. Gastwirth

References

ltrend.test

Test for a Linear Trend in Variances

Description
Test for a linear trend in variances.

Usage
ltrend.test(y, group, score = NULL, location = c("median", "mean", "trim.mean"), tail = c("right", "left", "both"), trim.alpha = 0.25, bootstrap = FALSE, num.bootstrap = 1000, correction.method = c("none", "correction.factor", "zero.removal", "zero.correction"), correlation.method = c("pearson", "kendall", "spearman"))

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>a numeric vector of data values.</td>
</tr>
<tr>
<td>group</td>
<td>factor of the data.</td>
</tr>
<tr>
<td>score</td>
<td>weights to be used in testing an increasing/decreasing trend in group variances, score coincides by default with group; it can be chosen as a linear, quadratic or any other monotone function.</td>
</tr>
<tr>
<td>location</td>
<td>the default option is &quot;median&quot; corresponding to the robust Brown–Forsythe Levene-type procedure (Brown and Forsythe 1974); &quot;mean&quot; corresponds to the classical Levene’s procedure (Levene 1960), and &quot;trim.mean&quot; corresponds to the robust Levene-type procedure using the group trimmed means.</td>
</tr>
</tbody>
</table>

See Also
gini.index

Examples

```r
## Data on: number of senators (second column) and
## representatives (third column) relative to population size (first column) in 1963
## First column is treated as the data argument.
data(data1963)

## Single weight Lorenz Curve using number of senators as weight argument.
lorenz.curve(data1963)

## Multiple weight Lorenz Curve using number of senators as weight argument.
lorenz.curve(data1963, mul = TRUE)

## Multiple weight Lorenz Curve using number of representatives
## as weight argument.
lorenz.curve(data1963[, "pop1963"], data1963[, "rep1963"], mul = TRUE)
```
tail

the default option is "right", corresponding to an increasing trend in variances as the one-sided alternative; "left" corresponds to a decreasing trend in variances, and "both" corresponds to any (increasing or decreasing) monotonic trend in variances as the two-sided alternative.

trim.alpha

the fraction (0 to 0.5) of observations to be trimmed from each end of x before the mean is computed.

bootstrap

a logical value identifying whether to implement bootstrap. The default is FALSE, i.e., no bootstrap; if set to TRUE, the bootstrap method described in Lim and Loh (1996) for Levene’s test is applied.

num.bootstrap

number of bootstrap samples to be drawn when the bootstrap argument is set to TRUE. The default value is 1000.

correction.method

procedures to make the test more robust; the default option is "none"; "correction.factor" applies the correction factor described by O’Brien (1978) and Keyes and Levy (1997); "zero.removal" performs the structural zero removal method by Hines and Hines (2000); "zero.correction" performs a combination of the O’Brien’s correction factor and the Hines–Hines structural zero removal method (Noguchi and Gel 2010). Note that the options "zero.removal" and "zero.correction" are only applicable when the location is set to "median", otherwise, "none" is applied.

correlation.method

measures of correlation; the default option is "pearson", the linear correlation coefficient that is equivalent to the t-test; nonparametric measures of correlation such as "kendall" (Kendall’s tau) or "spearman" (Spearman’s rho) may also be chosen.

Details

The test statistic is based on the classical Levene’s procedure (using the group means), the modified Brown–Forsythe Levene-type procedure (using the group medians), or the modified Levene-type procedure (using the group trimmed means). More robust versions of the test using the correction factor or structural zero removal method are also available. Two options for calculating critical values, namely, approximated and bootstrapped, are available. By default, NAs are omitted from the data.

Value

A list of class "htest" containing the following components:

statistic the value of the test statistic expressed in terms of correlation (Pearson, Kendall, or Spearman).

p.value the p-value of the test.

method type of test performed.

data.name a character string giving the name of the data.

t.statistic the value of the test statistic from Student’s t-test.

non.bootstrap.p.value the p-value of the test without bootstrap method.
**Author(s)**
Kimihiro Noguchi, W. Wallace Hui, Yulia R. Gel, Joseph L. Gastwirth, Weiwen Miao

**References**


**See Also**

`neuhauser.hothorn.test`, `levene.test`, `lnested.test`, `mma.test`, `robust.mmm.test`

**Examples**

```r
data(pot)
ltrtrend.test(pot[, "obs"], pot[, "type"], location = "median", tail = "left", correction.method = "zero.correction")

## Bootstrap version of the test. The calculation may take up a few minutes
## depending on the number of bootstrap samples.
ltrtrend.test(pot[, "obs"], pot[, "type"], location = "median", tail = "left", correction.method = "zero.correction",
             bootstrap = TRUE, num.bootstrap = 500)
```
michigan  

*Dioxin Levels for Counties in the Upper Peninsula of Michigan*

**Description**

Data contains 16 observations of dioxin levels for counties in the Upper Peninsula of Michigan.

**Usage**

```r
data(michigan)
```

**Format**

A numeric vector of length 16.

**Source**

The Environmental Protection Agency (EPA) of the State of Michigan.

mma.test  

*Mudholkar–McDermott–Aumont Test for Ordered Variances for Normal Samples*

**Description**

Test for a monotonic trend in variances for normal samples. The test statistic is based on a combination of the finite intersection approach and the classical $F$ (variance ratio) test (Mudholkar et al. 1993). By default, NAs are omitted.

**Usage**

```r
mma.test(y, group, tail = c("right", "left", "both"))
```

**Arguments**

- `y`  
  a numeric vector of data values.
- `group`  
  factor of the data.
- `tail`  
  the default option is "right", corresponding to an increasing trend in variances as the one-sided alternative; "left" corresponds to a decreasing trend in variances, and "both" corresponds to any (increasing or decreasing) monotonic trend in variances as the two-sided alternative.
Value

A list with the following components:

- T: the statistic and p-value of the test based on the Tippett p-value combination.
- F: the statistic and p-value of the test based on the Fisher p-value combination.
- N: the statistic and p-value of the test based on the Liptak p-value combination.
- L: the statistic and p-value of the test based on the Mudholkar–George p-value combination.

Each of the list elements is a list of class "htest" with the following elements:

- statistic: the value of the test statistic.
- p.value: the p-value of the test.
- method: type of test performed.
- data.name: a character string giving the name of the data.

Author(s)

Kimihiro Noguchi, Yulia R. Gel

References


See Also

neuhauser.hothorn.test, levene.test, lnested.test, ltrend.test, robust.mmm.test

Examples

data(pot)
mma.test(pot[, "obs"], pot[, "type"], tail = "left")$N

neuhauser.hothorn.test

Neuhauser-Hothorn Double Contrast Test for a Monotonic Trend in Variances

Description

The test statistic suggested by Neuhauser and Hothorn (2000).
Usage

neuhauser.hothorn.test(y, group, location = c("median", "mean", "trim.mean"), tail = c("right", "left", "both"), trim.alpha = 0.25, bootstrap = FALSE, num.bootstrap = 1000, correction.method = c("none", "correction.factor", "zero.removal", "zero.correction"))

Arguments

y  a numeric vector of data values.
group  factor of the data.
location  the default option is "median" corresponding to the robust Brown–Forsythe Levene-type procedure (Brown and Forsythe 1974); "mean" corresponds to the classical Levene's procedure (Levene 1960), and "trim.mean" corresponds to the robust Levene-type procedure using the group trimmed means.
tail  the default option is "right", corresponding to an increasing trend in variances as the one-sided alternative; "left" corresponds to a decreasing trend in variances, and "both" corresponds to any (increasing or decreasing) monotonic trend in variances as the two-sided alternative.
trim.alpha  the fraction (0 to 0.5) of observations to be trimmed from each end of x before the mean is computed.
bootstrap  a logical value identifying whether to implement bootstrap. The default is FALSE, i.e., no bootstrap; if set to TRUE, the bootstrap method described in Lim and Loh (1996) for Levene’s test is applied.
num.bootstrap  number of bootstrap samples to be drawn when the bootstrap argument is set to TRUE. The default value is 1000.
correction.method  procedures to make the test more robust; the default option is "none"; "correction.factor" applies the correction factor described by O’Brien (1978) and Keyes and Levy (1997); "zero.removal" performs the structural zero removal method by Hines and Hines (2000); "zero.correction" performs a combination of the O’Brien’s correction factor and the Hines–Hines structural zero removal method (Noguchi and Gel 2010). Note that the options "zero.removal" and "zero.correction" are only applicable when the location is set to "median", otherwise, "none" is applied.

Details

The test statistic is based on the classical Levene’s procedure (using the group means), the modified Brown–Forsythe Levene-type procedure (using the group medians), or the modified Levene-type procedure (using the group trimmed means). More robust versions of the test using the correction factor or structural zero removal method are also available. Two options for calculating critical values, namely, approximated and bootstrapped, are available. By default, NAs are omitted from the data.
Value

A list of class "htest" with the following components:

- **statistic**: the value of the test statistic.
- **p.value**: the p-value of the test.
- **method**: type of test performed.
- **data.name**: a character string giving the name of the data.
- **non.bootstrap.p.value**: the p-value of the test without bootstrap method.

Author(s)

Kimihiro Noguchi, Yulia R. Gel

References


See Also

- `levene.test`, `lnested.test`, `ltrend.test`, `mma.test`, `robust.mmm.test`

Examples

```r
data(pot)
neuhauser.hothorn.test(pot[, "obs"], pot[, "type"], location = "median",
tail = "left", correction.method = "zero.correction")```
nig.parameter

Generate Parameters for the Normal Inverse Gaussian (NIG) Distribution

Description

Produce four parameters, alpha (tail heavyness), beta (asymmetry), delta (scale), and mu (location) from the four variables: mean, variance, kurtosis, and skewness.

Usage

nig.parameter(mean = mean, variance = variance, kurtosis = kurtosis, skewness = skewness)

Arguments

mean mean of the NIG distribution.
variance variance of the NIG distribution.
kurtosis excess kurtosis of the NIG distribution.
skewness skewness of the NIG distribution.

Details

The parameters are generated with three conditions: 1) \(3 \times \text{kurtosis} > 5 \times \text{skewness}^2\); 2) \(\text{skewness} > 0\), and 3) \(\text{variance} > 0\). See Atkinson (1982), Barndorff-Nielsen and Blaesild (1983), and Noguchi and Gel (2010).

Value

A list with the following numeric components:

alpha tail-heavyness parameter of the NIG distribution.
beta asymmetry parameter of the NIG distribution.
delta scale parameter of the NIG distribution.
mu location parameter of the NIG distribution.

Author(s)

Kimihiro Noguchi, Yulia R. Gel
References


See Also

rnig

Examples

```r
library(fBasics)

# Testing with alpha = 0, beta = 1, and delta = 1
mu = test$mu

# Test with alpha = 1, beta = 1, and delta = 1
mu = test$mu

test <- nиг.parameter(0, 2, 5, 1)
random <- rnиг(1000000, alpha = test$alpha, beta = test$beta, mu = test$mu, delta = test$delta)

mean(random)
var(random)
kurtosis(random)
skewness(random)
```

---

**popdata**

Population Size of 33 Districts of the Tennessee Legislature in 1900, 1960, and 1972

Description


Usage

```r
data(popdata)
```

Format

A data frame with 33 observations on the following 3 numeric variables:

- pop1900 population data in 1900
- pop1960 population data in 1960
- pop1972 population data in 1972
Source


References


---

### Description

The apertures of the chupa pots from three Philippine locations: Dalupa (Apdl), Dangtalan (Apdg), and Paradijon (ApP).

### Usage
data(pot)

### Format

A data frame with 343 observations of 2 variables: obs (integer values of observed apertures) and locations (factor with 3 levels).

### Details

Archaeologists are concerned with the effect that increasing economic activity had on older civilizations. Economic growth and its related economic specialization led to the "standardization hypothesis", i.e., increased production of an item would lead to its becoming more uniform. Kvamme et al. (1996) focused on earthenware, chupa-pots from three Philippine communities that differ in the way they organize ceramic production. In Dangtalan, pottery is primarily made for household use; in Dalupa there is a non-market barter economy where potters exchange their works. In the village of Paradijon, near the provincial capital, full-time pottery specialists sell their output to shopkeepers for sale to the general public.

### Source

The data are kindly provided by Professor Kvamme (Kvamme et al. 1996).

### References

Description

The robust and classical Jarque–Bera tests of normality.

Usage

```r
rjb.test(x, option = c("RJB", "JB"),
         crit.values = c("chisq.approximation", "empirical"), N = 0)
```

Arguments

- `x`: a numeric vector of data values.
- `option`: the choice of whether to perform the robust test, "RJB" (default) or classic test, "JB".
- `crit.values`: a character string specifying how the critical values should be obtained: approximated by the Chi-square distribution (default) or empirically.
- `N`: number of Monte Carlo simulations for the empirical critical values.

Details

The test is based on a joint statistic using skewness and kurtosis coefficients. The Robust Jarque–Bera (RJB) is the robust version of the Jarque–Bera (JB) test of normality. The RJB (default option) utilizes the robust standard deviation (specifically, the Average Absolute Deviation from the Median; MAAD) to estimate sample kurtosis and skewness. For more details, see Gel and Gastwirth (2008). Users can also choose to perform the classical Jarque–Bera test (Jarque and Bera 1980).

Value

A list of class "htest" with the following components:

- `statistic`: the value of the test statistic.
- `parameter`: the degrees of freedom.
- `p.value`: the p-value of the test.
- `method`: type of test was performed.
- `data.name`: a character string giving the name of the data.

Note

Modified from `jarque.bera.test` (tseries package).

Author(s)

W. Wallace Hui, Yulia R. Gel, Joseph L. Gastwirth, Weiwen Miao
References


See Also

`sj.test, rqq, jarque.bera.test`

Examples

```r
## Normally distributed data
x <- rnorm(100)
rjb.test(x)

## Using zuni data
data(zuni)
rjb.test(zuni[, "Revenue"])
```

---

**rlm.test**

*Robust L1 Moment-Based (RLM) Goodness-of-Fit Test for the Laplace Distribution*

**Description**

Robust test for the Laplace distribution. Two options for calculating critical values, namely, approximated with Chi-square distribution and empirical, are available.

**Usage**

`rlm.test(x, crit.values = c("chisq.approximation", "empirical"), N = 0)`

**Arguments**

- `x` a numeric vector of data values.
- `crit.values` a character string specifying how the critical values should be obtained: approximated by the Chi-square distribution (default) or empirically.
- `N` number of Monte Carlo simulations for the empirical critical values.

**Details**

The test is based on a joint statistic using skewness and kurtosis coefficients. In particular, RLM uses the Average Absolute Deviation from the Median (MAAD), a robust estimate of standard deviation. See Gel (2010).
**Value**

A list of class "htest" with the following components:

- **statistic**: the value of the test statistic.
- **parameter**: the degrees of freedom.
- **p.value**: the p-value of the test.
- **method**: type of test was performed.
- **data.name**: a character string giving the name of the data.

**Author(s)**

Kimihiro Noguchi, W. Wallace Hui, Yulia R. Gel

**References**


**See Also**

`sj.test`, `rjb.test`, `rqq`, `jarque.bera.test`

**Examples**

```r
## Laplace distributed data
x = rexp(100) - rexp(100)
rlm.test(x)
```

---

**Description**

A test for a monotonic trend in variances (Mudholkar et al. 1995). The test statistic is based on a combination of the finite intersection approach and the two-sample t-test using Miller’s transformation. By default, NAs are omitted.

**Usage**

```r
robust.mmm.test(y, group, tail = c("right", "left", "both"))
```
Arguments

y a numeric vector of data values.
group factor of the data.
tail the default option is "right", corresponding to an increasing trend in variances as the one-sided alternative; "left" corresponds to a decreasing trend in variances, and "both" corresponds to any (increasing or decreasing) monotonic trend in variances as the two-sided alternative.

Value

A list with the following elements:

T the statistic and p-value of the test based on the Tippett p-value combination.
F the statistic and p-value of the test based on the Fisher p-value combination.
N the statistic and p-value of the test based on the Liptak p-value combination.
L the statistic and p-value of the test based on the Mudholkar-George p-value combination.

Each of the list elements is a list of class "htest" with the following elements:

statistic the value of the test statistic.
p.value the p-value of the test.
method type of test performed.
data.name a character string giving the name of the data.

Author(s)

Kimihiro Noguchi, Yulia R. Gel

References


See Also

neuhauser.hothorn.test, levene.test, lnested.test, ltrend.test, mma.test

Examples

data(pot)
robust.mmm.test(pot[, "obs"], pot[, "type"], tail = "left")$N
Test of Normality Using RQQ Plots

Description

Produce robust quantile-quantile (RQQ) and classical quantile-quantile (QQ) plots for graphical assessment of normality and optionally add a line, a QQ line, to the produced plot. The QQ line may be chosen to be a 45-degree line or to pass through the first and third quartiles of the data. NAs from the data are omitted.

Usage

\[ \text{rqq}(y, \text{plot.it} = \text{TRUE}, \text{square.it} = \text{TRUE}, \text{scale} = \text{c("MAD", "J", "classical")}, \text{location} = \text{c("median", "mean")}, \text{line.it} = \text{FALSE}, \text{line.type} = \text{c("45 degrees", "QQ")}, \text{col.line} = 1, \text{lwd} = 1, \text{outliers} = \text{FALSE}, \alpha = 0.05, \ldots) \]

Arguments

- \( y \)  
  the input data.
- \( \text{plot.it} \)  
  logical. Should the result be plotted?
- \( \text{square.it} \)  
  logical. Should the plot scales be square? The default is \text{TRUE}.
- \( \text{scale} \)  
  the choice of a scale estimator, i.e., the classical or robust estimate of the standard deviation.
- \( \text{location} \)  
  the choice of a location estimator, i.e., the mean or median.
- \( \text{line.it} \)  
  logical. Should the line be plotted? No line is the default.
- \( \text{line.type} \)  
  If \( \text{line.it} = \text{TRUE} \), the choice of a line to be plotted, i.e., the 45-degree line or the line passing through the first and third quartiles of the data.
- \( \text{col.line} \)  
  the color of the line (if plotted).
- \( \text{lwd} \)  
  the line width (if plotted).
- \( \text{outliers} \)  
  logical. Should the outliers be listed in the output?
- \( \alpha \)  
  significance level of outliers. If \( \text{outliers} = \text{TRUE} \), then all observations that are less than the \( 100\times\alpha \)-th standard normal percentile or greater than the \( 100\times(1-\alpha) \)-th standard normal percentile will be listed in the output.
- \( \ldots \)  
  other parameters passed to the \text{plot} function.

Details

An RQQ plot is a modified QQ plot where data are robustly standardized by the median and robust measure of spread (rather than mean and classical standard deviation as in the basic QQ plots) and then are plotted against the expected standard normal order statistics (Gel et al. 2005; Weisberg 2005). Under normality, the plot of the standardized observations should follow the 45-degree line, or QQ line. Both the median and robust standard deviation are significantly less sensitive to outliers than mean and classical standard deviation and therefore are more preferable in many practical
situations to assess graphically deviations from normality (if any). We choose median and MAD as a robust measure of location and spread for our RQQ plots since this standardization typically provides a clearer graphical diagnostics of normality. In particular, deviations from the QQ line are usually more noticeable in RQQ plots in the case of outliers and heavy tails. Users can also choose to plot the 45-degree line or the 1st-3rd quartile line (see the argument line.type). No line is the default.

Value

A list with the following numeric components:

- x: the x-coordinates of the points that were/would be plotted.
- y: the original data vector, i.e., the corresponding y-coordinates, including NAs (if any).

Author(s)

W. Wallace Hui, Yulia R. Gel, Joseph L. Gastwirth, Weiwen Miao

References


See Also

- `rjb.test`, `sj.test`, `qqnorm`, `qqplot`, `qline`

Examples

```r
## Simulate 100 observations from standard normal distribution:
y = rnorm(100)
rrq(y)

## Using Michigan data
data(michigan)
rrq(michigan)
```

---

**Runs Test for Randomness**

**Description**

Performs the runs test for randomness (Mendenhall and Reinmuth 1982). Users can choose whether to plot the correlation graph or not, and whether to test against two-sided, negative, or positive correlation. NAs from the data are omitted.
Usage

```r
runs.test(y, plot.it = FALSE, alternative = c("two.sided",
    "positive.correlated", "negative.correlated"))
```

Arguments

- **y**: a numeric vector of data values.
- **plot.it**: logical. If TRUE, then the graph will be plotted. If FALSE (default), then it is not plotted.
- **alternative**: a character string specifying the alternative hypothesis, must be one of "two.sided" (default), "negative.correlated", or "positive.correlated".

Details

On the graph, observations that are less than the sample median are represented by red letters "A", and observations that are greater or equal to the sample median are represented by blue letters "B".

Value

A list of class "htest" with the following components:

- **statistic**: the value of the standardized runs statistic.
- **p.value**: the p-value for the test.
- **data.name**: a character string giving the names of the data.
- **alternative**: a character string describing the alternative hypothesis.

Author(s)

Wallace Hui, Yulia R. Gel, Joseph L. Gastwirth, Weiwen Miao

References


See Also

- `bartels.test`

Examples

```r
# Simulate 100 observations from an autoregressive model
# of the first order (AR(1))
y = arima.sim(n = 100, list(ar = c(0.5)))

# Test y for randomness
runs.test(y)
```
Test of Normality – SJ Test

Description

Perform the robust directed test of normality, which is based on the ratio of the classical standard deviation \( S \) to the robust standard deviation \( J \) (Average Absolute Deviation from the Median, MAAD) of the sample data. See Gel et al. (2007).

Usage

\[
\texttt{sj.test}(x, \text{crit.values = c("t.approximation", "empirical"), N = 0})
\]

Arguments

- \( x \) a numeric vector of data values.
- \( \text{crit.values} \) a character string specifying how the critical values should be obtained, i.e., approximated by the \( t \)-distribution (default) or empirically.
- \( N \) number of Monte Carlo simulations for the empirical critical values.

Value

A list of class "htest" with the following components:

- \( \text{statistic} \) the standardized test statistic.
- \( \text{p.value} \) the \( p \)-value.
- \( \text{parameter} \) the ratio of the classical standard deviation \( S \) to the robust standard deviation \( J \).
- \( \text{data.name} \) a character string giving the name of the data.

Author(s)

Wallace Hui, Yulia R. Gel, Joseph L. Gastwirth, Weiwen Miao

References


See Also

- rqq, rjb.test, jarque.bera.test

Examples

\[
\text{data(bias)}
\]
\[
\text{sj.test(bias)}
\]
**Description**

Perform test for symmetry about an unknown median. Users can choose among the Cabilio–Masaro test (Cabilio and Masaro 1996), the Mira test (Mira 1999), or the MGG test (Miao et al. 2006); and between using asymptotic distribution of the respective statistics or a distribution from m-out-of-n bootstrap (Lyubchich et al. 2016). Additionally to the general distribution asymmetry, the function allows to test for negative or positive skewness (see the argument side). NAs from the data are omitted.

**Usage**

```r
symmetry.test(x, option = c("MGG", "CM", "M"), side = c("both", "left", "right"), boot = TRUE, B = 1000, q = 8/9)
```

**Arguments**

- `x` : data to be tested for symmetry.
- `option` : test statistic to be applied. The options include statistic by Miao et al. (2006) (default), Cabilio and Masaro (1996), and Mira (1999).
- `side` : choice from the three possible alternative hypotheses: general distribution asymmetry (`side = "both"`, default), left skewness (`side = "left"`), or right skewness (`side = "right"`).
- `boot` : logical value indicates whether m-out-of-n bootstrap will be used to obtain critical values (default), or asymptotic distribution of the chosen statistic.
- `B` : number of bootstrap replications to perform (default is 1000).
- `q` : scalar from 0 to 1 to define a set of possible m for the m-out-of-n bootstrap. Default `q = 8/9`. Possible m are then set as the values `unique(round(n*(q^j)))` greater than 4, where `n = length(x)` and `j = c(0:20)`.

**Details**

If the bootstrap option is used (`boot = TRUE`), a bootstrap distribution is obtained for each candidate subsample size m. Then, a heuristic method (Bickel et al. 1997; Bickel and Sakov 2008) is used for the choice of optimal m. Specifically, we use the Wasserstein metric (Ruschendorf 2001) to calculate distances between different bootstrap distributions and select m, which corresponds to the minimal distance. See Lyubchich et al. (2016) for more details.

**Value**

A list of class "htest" with the following components:

- `method` : name of the method.
- `data.name` : name of the data.
statistic value of the test statistic.
p.value p-value of the test.
alternative alternative hypothesis.
estimate bootstrap optimal \( m \) (given in the output only if bootstrap was used, i.e., boot = TRUE).

Author(s)
Joseph L. Gastwirth, Yulia R. Gel, Wallace Hui, Vyacheslav Lyubchich, Weiwen Miao, Xingyu Wang (in alphabetical order)

References


Examples

data(zuni) #run ?zuni to see the data description
symmetry.test(zuni[,"Revenue"], boot = FALSE)

---

**zuni**

*The Zuni Data from the Law Case: Zuni Public School v. United States Department of Education*

Description

Number of students and available revenue per student in each school district in New Mexico.
Usage

data(zuni)

Format

A data frame with 89 observations on 3 variables: District, Revenue, and Mem (number of students).

Details

The Zuni data come from a law case "The Zuni Public School District No. 89, Gallup-McKinley County Public School District No. 1, Petitioners v. United States Department of Education" concerning whether the revenue per pupil satisfied the standard for "equal" expenditures per pupil in the state. This classification determines whether most of the federal money given to the state under the law goes to the state or to the local school districts.

Source


References

Index

*Topic data
  bias, 4
  blackhire, 4
  dataQYVS, 9
  michigan, 22
  popdata, 27
  pot, 28
  zuni, 38
*Topic distribution
  bartels.test, 2
  laplace.test, 11
  nig.parameter, 26
  rjb.test, 29
  rlm.test, 30
  rqq, 33
  runs.test, 34
  sj.test, 36
  symmetry.test, 37
*Topic homogeneity
  cd, 6
  cmh.test, 7
  gini.index, 9
  j.maad, 10
  levene.test, 12
*Topic htest
  bartels.test, 2
  brunner.munzel.test, 5
  cmh.test, 7
  levene.test, 12
  lnested.test, 15
  ltrend.test, 19
  mma.test, 22
  neuhauser.hothorn.test, 23
  rjb.test, 29
  rlm.test, 30
  robust.mmm.test, 31
  runs.test, 34
  sj.test, 36
  symmetry.test, 37

*Topic nonparametric
  brunner.munzel.test, 5
*Topic plot
  lorenz.curve, 17
*Topic robust
  j.maad, 10
  levene.test, 12
  lnested.test, 15
  ltrend.test, 19
  neuhauser.hothorn.test, 23
  rjb.test, 29
  rlm.test, 30
  robust.mmm.test, 31
  rqq, 33
  sj.test, 36
  symmetry.test, 37
*Topic variability
  cd, 6
  gini.index, 9
  j.maad, 10
  levene.test, 12
  lnested.test, 15
  ltrend.test, 19
  mma.test, 22
  neuhauser.hothorn.test, 23
  robust.mmm.test, 31
  bartels.test, 2, 35
  bias, 4
  blackhire, 4
  brunner.munzel.test, 5
  cd, 6, 10, 11
  cmh.test, 7
  data1963, 9
  gini.index, 7, 9, 11, 19
  j.maad, 7, 10, 10
  jarque.bera.test, 29–31, 36
INDEX

laplace.test, 11
levene.test, 12, 17, 21, 23, 25, 32
lnested.test, 14, 15, 21, 23, 25, 32
lorenz.curve, 10, 17
ltrend.test, 14, 17, 19, 23, 25, 32
mantelhaen.test, 8
michigan, 22
mma.test, 14, 17, 21, 22, 25, 32
neuhauser.hothorn.test, 14, 17, 21, 23, 23, 32
nig.parameter, 26
plaplace, 12
plot, 18, 33
popdata, 27
pot, 28
pwilcox, 6
qcline, 34
qqnorm, 34
qqplot, 34
rjb.test, 11, 29, 31, 34, 36
rlm.test, 30
rnig, 27
robust.mmm.test, 14, 17, 21, 23, 25, 31
rqq, 11, 30, 31, 33, 36
runs.test, 3, 34
sj.test, 11, 30, 31, 34, 36
symmetry.test, 37
wilcox.test, 6
zuni, 38