Package 'BenfordTests'

April 16, 2013

Type Package

Title Statistical Tests for Evaluating Conformity to Benford's Law

Version 0.5

Date 2013-04-09

Depends R (>= 2.15.1)

Author Dieter William Joenssen, with contributions from Thomas Muellerleile

Maintainer Dieter William Joenssen <Dieter.Joenssen@TU-Ilmenau.de>

Description This package contains seven specialized statistical tests and support functions for determining if numerical data could conform to Benford's law.

License GPL-3

R topics documented:

`BenfordTests-package` *Statistical Tests for Benford's Law.*

Description

This package contains seven specialized statistical tests and support functions for determining if numerical data could conform to Benford's law.

Details

Package:	BenfordTests
Type:	Package
Version:	0.5
Date:	2013-04-09
License:	GPL-3

`BenfordTests` is the implementation of the seven most commonly used goodness-of-fit (GOF) tests to assess if data conforms to Benford's law.
Tests include:
Pearson *chi-square* statistic (Pearson (1900))
Kolmogorov-Smirnov *D* statistic (Kolmogorov (1933))
Freedman's modification of Watson's *U-square* statistic (Freedman (1981), Watson (1961))
Chebyshev distance *m* statistic (Leemis (2000))
Euclidean distance *d* statistic (Cho and Gaines (2007))
Judge-Schechter mean deviation *a-star* statistic (Judge and Schechter (2009))
Joenssen's *JP-square* statistic, a Shapiro-Francia type correlation test (Shapiro and Francia (1972))

All tests may be performed using more than one leading digit. All tests simulate the specific p-values required for statistical inference, while p-values for the *chi-square* and *D* statistics may also be determined using their asymptotic distributions. The package version numbers are indicative of performance and breadth of functions.

Author(s)

Dieter William Joenssen

Maintainer: Dieter William Joenssen <Dieter.Joenssen@TU-Ilmenau.de>

References

Benford F. The law of anomalous numbers. Proceedings of the American Philosophical Society. 1938;78:551-572.

Cho WKT, Gaines BJ. Breaking the (Benford) law: Statistical fraud detection in campaign finance. The American Statistician. 2007;61(4):218-223.

Freedman LS. Watson's Un2 statistic for a discrete distribution. Biometrika. 1981;68(3):708-711.

Judge G, Schechter L. Detecting problems in survey data using Benford's law. Journal of Human Resources. 2009;44:1-24.

Kolmogorov AN. Sulla determinazione empirica di una legge di distibuzione. Giornale dell'Istituto Italiano degli Attuari. 1933;4:83-91.

Leemis LM, Schmeiser BW, Evans DL. Survival distributions satisfying Benford's law. The American Statistician. 2000;54(4):236-241.

Newcomb S. Note on the frequency of use of the different digits in natural numbers. American Journal of Mathematics. 1881;4(1):39-40.

Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine Series 5. 1900;50(302):157-175.

Shapiro SS, Francia RS. An approximate analysis of variance test for normality. Journal of the American Statistical Association. 1972;67:215-216.

Watson GS. Goodness-of-fit tests on a circle. Biometrika. 1961;48:109-114.

Examples

```
#Create a sample satisfying Benford's law
X<-rbenf(n=20)
X
#Look at the first digits of the sample
leading_digits(X)

#Perform a Chi-squared Test on the sample's first digits
chi_square_benford(X)
```

`Chebyshev_dist_benford`

Chebyshev Distance Test for Benford's Law

Description

`Chebyshev_dist_benford` takes any numerical vector reduces the sample to the specified number of significant digits and performs a goodness-of-fit test based on the Chebyshev distance between the first digits' distribution and Benford's distribution to assert if the data conforms to Benford's law.

Usage

```
Chebyshev_dist_benford(x = NULL, first_digits = 1, pvalmethod = "simulate", pvalsims = 10000)
```

Arguments

x	A numeric vector.
first_digits	An integer determining the number of first digits to use for testing, i.e. 1 for only the first, 2 for the first two etc.
pvalmethod	Method used for calculating the p-value. Currently only "simulate" is available.
pvalsims	An integer specifying the number of replicates used if pvalmethod = "simulate".

Details

A statistical test is performed utilizing the Chebyshev distance between leading_digits(x,first_digits) and pbenf(first_digits). x is a numeric vector of arbitrary length. Values of x should be continuous, as dictated by theory, but may also be integers. first_digits should be chosen so that leading_digits(x,first_digits) is not influenced by previous rounding.

Value

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

statistic	the value of the Chebyshev distance test statistic
p.value	the p-value for the test
method	a character string indicating the type of test performed

Author(s)

Dieter William Joenssen <Dieter.Joenssen@TU-Ilmenau.de>

References

Benford F. The law of anomalous numbers. Proceedings of the American Philosophical Society. 1938;78:551-572.

Leemis LM, Schmeiser BW, Evans DL. Survival distributions satisfying Benford's law. The American Statistician. 2000;54(4):236-241.

Morrow J. Benford's law, families of distributions and a test basis. 2010. http://www.johnmorrow.info/projects/benford/benfordMain.pdf.

See Also

pbenf

chi_square_benford *Pearsons's Chi-squared Goodness-of-Fit Test for Benford's Law*

Description

chi_square_benford takes any numerical vector reduces the sample to the specified number of significant digits and performs Pearson's chi-square goodness-of-fit test to assert if the data conforms to Benford's law.

Usage

```
chi_square_benford(x = NULL, first_digits = 1, pvalmethod = "asymptotic", pvalsims = 10000)
```

Arguments

x	A numeric vector.
first_digits	An integer determining the number of first digits to use for testing, i.e. 1 for only the first, 2 for the first two etc.
pvalmethod	Method used for calculating the p-value. Either "asymptotic" or "simulate".
pvalsims	An integer specifying the number of replicates to use if pvalmethod = "simulate".

Details

A chi-square goodness-of fit test is performed on leading_digits(x,first_digits) versus pbenf(first_digits). x is a numeric vector of arbitrary length. Values of x should be continuous, as dictated by theory, but may also be integers. first_digits should be chosen so that leading_digits(x,first_digits) is not influenced by previous rounding.

Value

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

statistic	the value of the chi-squared test statistic
p.value	the p-value for the test
method	a character string indicating the type of test performed

Author(s)

Dieter William Joenssen <Dieter.Joenssen@TU-Ilmenau.de>

References

Benford F. The law of anomalous numbers. Proceedings of the American Philosophical Society. 1938;78:551-572.

Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine Series 5. 1900;50(302):157-175.

See Also

pbenf

Euclidean_dist_benford

Euclidean Distance Test for Benford's Law

Description

Euclidean_dist_benford takes any numerical vector reduces the sample to the specified number of significant digits and performs a goodness-of-fit test based on the Euclidean distance between the first digits' distribution and Benford's distribution to assert if the data conforms to Benford's law.

Usage

```
Euclidean_dist_benford(x = NULL, first_digits = 1, pvalmethod = "simulate", pvalsims = 10000)
```

Arguments

x	A numeric vector.
first_digits	An integer determining the number of first digits to use for testing, i.e. 1 for only the first, 2 for the first two etc.
pvalmethod	Method used for calculating the p-value. Currently only "simulate" is available.
pvalsims	An integer specifying the number of replicates used if pvalmethod = "simulate".

Details

A statistical test is performed utilizing the Euclidean distance between leading_digits(x,first_digits) and pbenf(first_digits). x is a numeric vector of arbitrary length. Values of x should be continuous, as dictated by theory, but may also be integers. first_digits should be chosen so that leading_digits(x,first_digits) is not influenced by previous rounding.

Value

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

statistic	the value of the Euclidean distance test statistic
p.value	the p-value for the test
method	a character string indicating the type of test performed

Author(s)

Dieter William Joenssen <Dieter.Joenssen@TU-Ilmenau.de>

References

Benford F. The law of anomalous numbers. Proceedings of the American Philosophical Society. 1938;78:551-572.

Cho WKT, Gaines BJ. Breaking the (Benford) law: Statistical fraud detection in campaign finance. The American Statistician. 2007;61(4):218-223.

Morrow J. Benford's law, families of distributions and a test basis. 2010. http://www.johnmorrow.info/projects/benford/benfordMain.pdf.

See Also

pbenf

`Freedman_Watson_Usquare_benford` *Freedman-Watson U-squared Test for Benford's Law*

Description

`Freedman_Watson_Usquare_benford` takes any numerical vector reduces the sample to the specified number of significant digits and performs the Freedman-Watson test for discreet distributions between the first digits' distribution and Benford's distribution to assert if the data conforms to Benford's law.

Usage

```
Freedman_Watson_Usquare_benford(x = NULL, first_digits = 1, pvalmethod = "simulate", pvalsims =
```

Arguments

`x`	A numeric vector.
`first_digits`	An integer determining the number of first digits to use for testing, i.e. 1 for only the first, 2 for the first two etc.
`pvalmethod`	Method used for calculating the p-value. Currently only `"simulate"` is available.
`pvalsims`	An integer specifying the number of replicates used if `pvalmethod = "simulate"`.

Details

A Freedman-Watson test for discreet distributions is performed between `leading_digits(x,first_digits)` and `pbenf(first_digits)`. `x` is a numeric vector of arbitrary length. Values of `x` should be continuous, as dictated by theory, but may also be integers. `first_digits` should be chosen so that `leading_digits(x,first_digits)` is not influenced by previous rounding.

Value

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

statistic	the value of the *U-square* test statistic
p.value	the p-value for the test
method	a character string indicating the type of test performed

Author(s)

Dieter William Joenssen <Dieter.Joenssen@TU-Ilmenau.de>

References

Benford F. The law of anomalous numbers. Proceedings of the American Philosophical Society. 1938;78:551-572.

Freedman LS. Watson's Un2 statistic for a discrete distribution. Biometrika. 1981;68(3):708-711.

Watson GS. Goodness-of-fit tests on a circle. Biometrika. 1961;48:109-114.

See Also

pbenf

J_stat_squ_benford *Joenssen's* JP-square *Test for Benford's Law*

Description

J_stat_squ_benford takes any numerical vector reduces the sample to the specified number of significant digits and performs a goodness-of-fit test based on the correlation between the first digits' distribution and Benford's distribution to assert if the data conforms to Benford's law.

Usage

```
J_stat_squ_benford(x = NULL, first_digits = 1, method = "pearson", pvalmethod = "simulate", pval
```

Arguments

x	A numeric vector.
first_digits	An integer determining the number of first digits to use for testing, i.e. 1 for only the first, 2 for the first two etc.
method	A character string indicating which correlation coefficient is to be computed. One of "pearson" (default), "kendall", or "spearman", can be abbreviated.
pvalmethod	Method used for calculating the p-value. Currently only "simulate" is available.
pvalsims	An integer specifying the number of replicates used if pvalmethod = "simulate".

Details

A statistical test is performed utilizing the sign-preserved squared correlation between `leading_digits(x,first_digit` and `pbenf(first_digits)`. `x` is a numeric vector of arbitrary length. Values of `x` should be continuous, as dictated by theory, but may also be integers. `first_digits` should be chosen so that `leading_digits(x,first_digits)` is not influenced by previous rounding.

Value

A list with class "`htest`" containing the following components:

`statistic`	the value of the *JP-square* test statistic
`p.value`	the p-value for the test
`method`	a character string indicating the type of test performed

Author(s)

Dieter William Joenssen <`Dieter.Joenssen@TU-Ilmenau.de`>

References

Benford F. The law of anomalous numbers. Proceedings of the American Philosophical Society. 1938;78:551-572.

Joenssen DW. A new test for Benford's distribution [abstract]. In: Abstract-proceedings of the 3rd joint Statistical Meeting DAGStat, March 18-22, 2013; Freiburg, Germany.

Shapiro SS, Francia RS. An approximate analysis of variance test for normality. Journal of the American Statistical Association. 1972;67:215-216.

See Also

`pbenf`

`J_S_avg_dev_benford` *Judge-Schechter Mean Deviation for Benford's Law*

Description

`J_S_avg_dev_benford` takes any numerical vector reduces the sample to the specified number of significant digits and performs a goodness-of-fit test based on the deviation in means of the first digits' distribution and Benford's distribution to assert if the data conforms to Benford's law.

Usage

```
J_S_avg_dev_benford(x = NULL, first_digits = 1, pvalmethod = "simulate", pvalsims = 10000)
```

Arguments

x	A numeric vector.
first_digits	An integer determining the number of first digits to use for testing, i.e. 1 for only the first, 2 for the first two etc.
pvalmethod	Method used for calculating the p-value. Currently only "simulate" is available.
pvalsims	An integer specifying the number of replicates used if pvalmethod = "simulate".

Details

A statistical test is performed utilizing the deviation beteen the mean digit of leading_digits(x,first_digits) and pbenf(first_digits). The resulting statistic is normalized to [0,1]. x is a numeric vector of arbitrary length. Values of x should be continuous, as dictated by theory, but may also be integers. first_digits should be chosen so that leading_digits(x,first_digits) is not influenced by previous rounding.

Value

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

statistic	the value of the *a-star* test statistic
p.value	the p-value for the test
method	a character string indicating the type of test performed

Author(s)

Dieter William Joenssen <Dieter.Joenssen@TU-Ilmenau.de>

References

Benford F. The law of anomalous numbers. Proceedings of the American Philosophical Society. 1938;78:551-572.

Judge G, Schechter L. Detecting problems in survey data using Benford's law. Journal of Human Resources. 2009;44:1-24.

See Also

pbenf

K_S_benford *Kolmogorov-Smirnov Test for Benford's Law*

Description

K_S_benford takes any numerical vector reduces the sample to the specified number of significant digits and performs the Kolmogorov-Smirnov goodness-of-fit test to assert if the data conforms to Benford's law.

Usage

```
K_S_benford(x = NULL, first_digits = 1, pvalmethod = "simulate", pvalsims = 10000)
```

Arguments

x	A numeric vector.
first_digits	An integer determining the number of first digits to use for testing, i.e. 1 for only the first, 2 for the first two etc.
pvalmethod	Method used for calculating the p-value. Currently only "simulate" is available.
pvalsims	An integer specifying the number of replicates used if pvalmethod = "simulate".

Details

A Kolmogorov-Smirnov test is performed between leading_digits(x,first_digits) and pbenf(first_digits). x is a numeric vector of arbitrary length. Values of x should be continuous, as dictated by theory, but may also be integers. first_digits should be chosen so that leading_digits(x,first_digits) is not influenced by previous rounding.

Value

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

statistic	the value of the Kolmogorov-Smirnov *D* test statistic
p.value	the p-value for the test
method	a character string indicating the type of test performed

Author(s)

Dieter William Joenssen <Dieter.Joenssen@TU-Ilmenau.de>

References

Benford F. The law of anomalous numbers. Proceedings of the American Philosophical Society. 1938;78:551-572.

Kolmogorov AN. Sulla determinazione empirica di una legge di distibuzione. Giornale dell'Istituto Italiano degli Attuari. 1933;4:83-91.

See Also

pbenf

`leading_digits` *Leading Digits*

Description

Returns the specified number of significant digits for each element of a given vector.

Usage

```
leading_digits(x = NULL, number = 1)
```

Arguments

`x`	A numeric vector.
`number`	An integer determining the number of first digits to use for testing, i.e. 1 for only the first, 2 for the first two etc.

Details

`x` is a numeric vector of arbitrary length. Unlike other solutions, this function will work reliably with all real numbers.

Value

Returns a vector of integers the same length as the input vector `x`.

Author(s)

Dieter William Joenssen <Dieter.Joenssen@TU-Ilmenau.de>

See Also

chi_square_benford; K_S_benford; Freedman_Watson_Usquare_benford; Chebyshev_dist_benford; Euclidean_dist_benford; J_S_avg_dev_benford; J_stat_squ_benford

pbenf *Distribution Function for Benford's Distribution*

Description

Returns the complete Benford distribution function for a given number of first digits.

Usage

```
pbenf(digits = 1)
```

Arguments

digits	An integer determining the number of first digits for which the pdf is returned, i.e. 1 for 1:9, 2 for 10:99 etc.

Value

Returns an object of class "table" containing the expected density of Benford's distribution for the given number of digits.

Author(s)

Dieter William Joenssen <Dieter.Joenssen@TU-Ilmenau.de>

References

Benford F. The law of anomalous numbers. Proceedings of the American Philosophical Society. 1938;78:551-572.

See Also

qbenf; rbenf

qbenf *Quantile Function for Benford's Distribution*

Description

Returns the complete quantile function for Benford's distribution with a given number of first digits.

Usage

```
qbenf(digits = 1)
```

Arguments

digits	An integer determining the number of first digits for which the qdf is returned, i.e. 1 for 1:9, 2 for 10:99 etc.

Value

Returns an object of class "table" containing the expected quantile function of Benford's distribution with a given number of digits.

Author(s)

Dieter William Joenssen <Dieter.Joenssen@TU-Ilmenau.de>

References

Benford F. The law of anomalous numbers. Proceedings of the American Philosophical Society. 1938;78:551-572.

See Also

pbenf; rbenf

---

rbenf *Random Sample Satisfying Benford's Law*

Description

Returns a random sample with length n satisfying Benford's law.

Usage

```
rbenf(n)
```

Arguments

n	Number of observations.

Value

Returns a random sample with length n satisfying Benford's law.

Author(s)

Dieter William Joenssen <Dieter.Joenssen@TU-Ilmenau.de>

References

Benford F. The law of anomalous numbers. Proceedings of the American Philosophical Society. 1938;78:551-572.

See Also

qbenf; pbenf

sequence_leading *Sequence of Possible Leading Digits*

Description

Returns a vector containing all possible significant digits for a given number of places.

Usage

```
sequence_leading(number = 1)
```

Arguments

number	An integer determining the number of first digits to be returned, i.e. 1 for 1:9, 2 for 10:99 etc.

Value

Returns an integer vector.

Author(s)

Dieter William Joenssen <Dieter.Joenssen@TU-Ilmenau.de>

Index