Package: CNPS (via r-universe)

August 23, 2024

Type Package

Title Nonparametric Statistics

Version 1.0.0

Author JiaSheng Zhang [aut,cre] (<zhangjiasheng0509@outlook.com>),
SiWei Deng [aut], Feng Yu [aut], YangYang Zhang [aut]

Maintainer JiaSheng Zhang <zhangjiasheng0509@outlook.com>

Description We unify various nonparametric hypothesis testing problems
in a framework of permutation testing, enabling hypothesis
testing on multi-sample, multidimensional data and contingency
tables. Most of the functions available in the R environment to
implement permutation tests are single functions constructed
for specific test problems; to facilitate the use of the
package, the package encapsulates similar tests in a
categorized manner, greatly improving ease of use. We will all
provide functions for self-selected permutation scoring methods
and self-selected p-value calculation methods (asymptotic,
exact, and sampling). For two-sample tests, we will provide
mean tests and estimate drift sizes; we will provide tests on
variance; we will provide paired-sample tests; we will provide
correlation coefficient tests under three measures. For
multi-sample problems, we will provide both ordinary and
ordered alternative test problems. For multidimensional data,
we will implement multivariate means (including ordered
alternatives) and multivariate pairwise tests based on four
statistics; the components with significant differences are
also calculated. For contingency tables, we will perform
permutation chi-square test or ordered alternative.

License GPL-2

Encoding UTF-8

Suggests e1071

NeedsCompilation no

Date/Publication 2021-05-25 07:00:06 UTC

Repository https://zhangmaotai.r-universe.dev

RemoteUrl https://github.com/cran/CNPS

RemoteRef HEAD

RemoteSha 6a04af3ac85ead79ea0854da3959bb8e38947205

Contents

---

cip *confidence interval for percentiles in the one-sample case*

Description

Finding confidence interval for (100p)th percentile in the one-sample case.

Usage

```
cip(x, conf.level = 0.95, p = 0.5)
```

Arguments

x	numeric vector of data values
conf.level	confidence level for the returned confidence interval
p	an arbitrary value from 0 to 1 which indicates the percentile

Details

Usually we take the set with the shortest interval. But if the upper bound is greater than the maximum value(or the lower bound is less than the minimum value), we will choose the maximum value as the upper bound(the minimum value as the lower bound).

Value

A list with following components

Lower.rank	rank of the lower bound of the confidence interval in the order statistic
Upper.rank	rank of the upper bound of the confidence interval in the order statistic
Lower	the lower bound of the confidence interval
Upper	the upper bound of the confidence interval

Author(s)

Jiasheng Zhang, Feng Yu, Yangyang Zhang, Siwei Deng. Tutored by YuKun Liu and Dongdong Xiang.

References

Higgins, J. J. (2004). An introduction to modern nonparametric statistics. Pacific Grove, CA: Brooks/Cole.

Examples

```
x <- c(72.1, 72.8, 72.9, 73.3, 76.1, 76.5, 78.8, 78.9, 79.7, 80.3, 80.5, 81.0)
cip(x)
cip(x, conf.level =0.9, p=0.7)
```

---

corr_test	*Correlation test*

Description

Test the correlation coefficient of the sample.

Usage

```
corr_test(x, y, alternative = "greater", measure = "pearson",
method_p = "sampling", samplenum = 1000, conf.level.sample = 0.95)
```

Arguments

x	numeric vectors of data values and should have the same length
y	numeric vectors of data values and should have the same length
alternative	a character string specifying the alternative hypothesis, must be one of "two.sided", "greater"(default) or "less"
measure	the way to measure the correlation coefficient and must be one of "pearson", "spearman" or "kendall"
method_p	a string indicating what method to use for p-value. "sampling" represents sampling; "asymptotic" represents using large sample approximations

samplenum   the number of SRS samples

conf.level.sample

                   p-value confidence level for SRS sampling

Details

All procedures and methods of the correlation coefficient test based on the Spearman Correlation Coefficient are the same as for the Pearson Correlation Coefficient. But pay attention to that the correlation coefficient test based on Kendall Correlation Coefficient is a little different from the above two due to its definition.

Value

A list with following components

method   the test uesd

score   the score which is used

stat   the statistic of the data under the given scoring system

conf.int   the confidence interval for p-value(only if method_p = "sampling")

pval   p-value for the test

null.value   a character string describing the alternative hypothesis

Author(s)

Jiasheng Zhang, Feng Yu, Yangyang Zhang, Siwei Deng. Tutored by YuKun Liu and Dongdong Xiang.

References

Higgins, J. J. (2004). An introduction to modern nonparametric statistics. Pacific Grove, CA: Brooks/Cole.

Examples

```
x=c(68,70,71,72)
y=c(153,155,140,180)
corr_test(x , y , measure = "kendall" , method = "asymptotic")
corr_test(x , y , measure = "kendall" , method = "sampling")
```

---

emcdf   *Estimating the population cdf*

Description

Finding confidence interval for the population cdf.

Usage

```
emcdf(x, conf.level = 0.05)
```

Arguments

x	numeric vector of data values
conf.level	confidence level for the returned confidence interval

Details

This "emcdf" constructs the approximation interval according to the central limit theorem. And use "plot(emcdf(data))" will help us draw a plot conveniently.

Value

A list with following components

sample	the given vector
empirical.cdf	the value of the empirical cdf
Lower	the lower bound of the confidence interval of the empirical cdf
Upper	the upper bound of the confidence interval of the empirical cdf

Author(s)

Jiasheng Zhang, Feng Yu, Yangyang Zhang, Siwei Deng. Tutored by YuKun Liu and Dongdong Xiang.

References

Higgins, J. J. (2004). An introduction to modern nonparametric statistics. Pacific Grove, CA: Brooks/Cole.

Examples

```
x <- c(7,11,15, 16, 20, 22, 24, 25, 29, 33, 34, 37, 41, 42, 49, 57, 66, 71, 84, 90)
em <- emcdf(x)
plot(em)
```

---

ksample_test   *Multiple sample permutation test*

Description

Test whether there is a difference among k treatments.

Usage

```
ksample_test(x, group, score = "kruskal", method_p = "sampling", type = "normal",
samplenum = 1000, conf.level.sample = 0.95)
```

Arguments

x
: numeric vector of data values

group
: factor that determines the grouping of elements in x

score
: a discrete value indicating the type of score. There are "original", "Wilcoxon", "van" and "exp" to be selected

method_p
: a string indicating what method to use for p-value. "sampling" represents sampling; "asymptotic" represents using large sample approximations

type
: "normal" refers to ordinary test, "JT" refers to ordered alternative hypothesis

samplenum
: the number of SRS samples

conf.level.sample
: p-value confidence level for SRS sampling

Details

Use a permutation sample based on the F-statistic or use a large sample approximation to determine if there is a difference between the populations (treatments). If the sample is ordered, you can also use the JT test. The argument "score" allows you to choose different scoring system to do the test. If you want to use a special scoring system defined by yourself, just transform the data first and then choose score="original" in the function.

Value

A list with following components

method
: the test uesd

stat
: the statistic of the data under the given scoring system

conf.int
: the confidence interval for p-value(only if method_p = "sampling")

pval
: p-value for the test

alternative
: a character string describing the alternative hypothesis

Author(s)

Jiasheng Zhang, Feng Yu, Yangyang Zhang, Siwei Deng. Tutored by YuKun Liu and Dongdong Xiang.

References

Higgins, J. J. (2004). An introduction to modern nonparametric statistics. Pacific Grove, CA: Brooks/Cole.

Examples

```
x1=c( 13.0, 24.1, 11.7, 16.3, 15.5, 24.5)
x2=c( 42.0, 18.0, 14.0, 36.0, 11.6, 19.0)
x3=c( 15.6, 23.8, 24.4, 24.0, 21.0, 21.1)
x4=c( 35.3, 22.5, 16.9, 25.0, 23.1, 26.0)
x <- c(x1, x2, x3, x4)
ind=c(rep(1,length(x1)), rep(2, length(x2)), rep(3, length(x3)), rep(4, length(x4)))
group=as.factor(ind)
ksample_test(x , group , type = "JT" , samplenum = 4000)
```

MultiDimen_test	*Multivariate Permutation Test and Paired Comparisons*

Description

Performs multivariate permutation tests, including paired tests.

Usage

```
MultiDimen_test (data , stat = "HT",pair=FALSE, method_p = "sampling",rank = FALSE,
diff = FALSE , samplenum = 1000)
```

Arguments

data	a matrix or data frame of data values.
stat	a character string specifying the statistic, must be one of "HT" (default), "tmax", "tmaxabs", "wsum", "zmax", "zmaxabs".
pair	a logical indicating whether you want a paired test.
method_p	a character string specifying the method of calculating p-value, must be one of "sampling" (default), " asymptotic", "exact".
rank	a logical indicating whether you want Wilcoxon test.
diff	a logical indicating whether you want to present which variables are different.
samplenum	a number specifying the number of sampling.

Details

The test can be used for multivariate permutation test and multivariate paired comparisons.

When doing multivariate paired comparisons, that is pair = TRUE, the statistic wsum is not suitable. Meanwhile, asymptotic method can only be used when statistic is HT. Besides, the second last column of the data must only contain two unique numbers to represent the two samples; the last column represents different pairs.

When doing multivariate permutation test, that is pair = FALSE, the statistic zmax and zmaxabs are not suitable. Meanwhile, the last column of the data must only contain 0 and 1 to represent the two samples. Besides, asymptotic method can not be used when statistic is tmax or tmaxabs.

Value

method	the test which is used.
score	a character string describing the score used for test.
stat	the test statistic.
pval	p-value for the test.
alternative	a character string describing the alternative hypothesis.
addition	a character string describing which variable is different in two samples.(presents only if pair = FALSE)

Author(s)

Jiasheng Zhang, Feng Yu, Yangyang Zhang, Siwei Deng. Tutored by YuKun Liu and Dongdong Xiang.

References

Higgins, J. J. (2004). An introduction to modern nonparametric statistics. Pacific Grove, CA: Brooks/Cole.

Examples

```
## Multivariate permutation test
data = matrix(c(6.81, 6.16, 5.92, 5.86, 5.80, 5.39,
                6.68, 6.30, 6.12, 5.71, 6.09, 5.28,
                6.34, 6.22, 5.90, 5.38, 5.20, 5.46,
                6.68, 5.24, 5.83, 5.49, 5.37, 5.43,
                6.79, 6.28, 6.23, 5.85, 5.56, 5.38,
                6.85, 6.51, 5.95, 6.06, 6.31, 5.39,
                6.64, 5.91, 5.59, 5.41, 5.24, 5.23,
                6.57, 5.89, 5.32, 5.41, 5.32, 5.30,
                6.84, 6.01, 5.34, 5.31, 5.38, 5.45,
                6.71, 5.60, 5.29, 5.37, 5.26, 5.41,
                6.58, 5.63, 5.38, 5.44, 5.17, 6.62,
                6.68, 6.04, 5.62, 5.31, 5.41, 5.44),
                nrow = 12,ncol = 6,byrow = TRUE
)
data=as.matrix(data)
```

```
index=c(rep(0,6),rep(1,6))
data = cbind(data,index)
x = MultiDimen_test(data ,  rank = FALSE ,  method_p = "sampling", samplenum = 100
, stat = "HT",diff = TRUE )
y = MultiDimen_test(data ,  rank = FALSE ,  method_p = "sampling", samplenum = 100
, stat = "tmax",diff = TRUE)
z = MultiDimen_test(data ,  rank = TRUE , method_p = "sampling"  , stat = "HT"
, samplenum = 100,diff = TRUE)

## Multivaraite paired comparisons
data = matrix(c(82, 60,  72, 62,
                75, 71,  70, 68,
                85, 59,  87, 64,
                90, 77,  87, 78),
             nrow = 4,ncol = 4,byrow = TRUE
)
x = data[,c(1,2)]
y = data[,c(3,4)]
data = cbind(rbind(x,y) , c(0,0,1,1) , c(1,2,1,2))
MultiDimen_test(data , method_p = "exact" , pair = TRUE)
```

---

pairwise_test   *Paired Comparisons*

Description

Detects differences between two related samples.

Usage

```
pairwise_test(x, y, alternative = "greater", score = "wilcoxon", method_p = "asymptotic",
method_asymptotic = "norm", method_wilcoxon = "type1", samplenum = 1000,
conf.level.sample = 0.95, samplemethod = "R")
```

Arguments

x	numeric vectors of data values and should have the same length
y	numeric vectors of data values and should have the same length
alternative	a character string specifying the alternative hypothesis, must be one of "two.sided", "greater"(default) or "less"
score	determines scoring systems and must be one of "original", "wilcoxon" or "sign"
method_p	a string indicating what method to use for p-value. "sampling" represents sampling; "asymptotic" represents using large sample approximations; "exact" represents Iterate through all combinations
method_asymptotic	
	determines the asymptotic distribution and should be one of "norm" or "binomial"(only for method_p="sign")

method_wilcoxon

> indicates the way to compute wilcoxon ranks when the ties are 0 and could be one of "type1" or "type2"

samplenum   the number of SRS samples

conf.level.sample

> p-value confidence level for SRS sampling

samplemethod   a discrete value indicating the method of sampling. "S" represents sample function sampling; "R" represents Put-back sampling

Details

If score="sign", then method_p must be "asymptotic". Three scoring systems can use the normal approximation but only "sign" can use binomial approximation. Namely, the argument method_asymptotic can be selected as "binomial" only if method_p="sign". And method_wilcoxon indicates the method to deal with ties. "type1" means ranking with zeros and "type2" means ranking without zeros.

Value

A list with following components

method   the test uesd

score   the score which is used

stat   the statistic of the data under the given scoring system

conf.int   the confidence interval for p-value(only if method_p = "sampling")

pval   p-value for the test

null.value   a character string describing the alternative hypothesis

Author(s)

Jiasheng Zhang, Feng Yu, Yangyang Zhang, Siwei Deng. Tutored by YuKun Liu and Dongdong Xiang.

References

Higgins, J. J. (2004). An introduction to modern nonparametric statistics. Pacific Grove, CA: Brooks/Cole.

Examples

```
x1=c(1530, 2130,2940,1960,2270)
x2=c(1290, 2250,2430,1900,2120)
pairwise_test(x1 , x2)
pairwise_test(x1 , x2 , method_p = "sampling" , samplenum = 4000)
pairwise_test(x1 , x2 , method_p = "asymptotic" , method_asymptotic = "norm")
```

---

permu_table	*Permutation Tests for Contingency Tables*

Description

Performs permutation tests on contingency tables, including tables with ordered or disordered categories.

Usage

```
permu_table(data , permu = "row" , row = NULL , col = NULL , fix = "row" ,
samplenum = 1000)
```

Arguments

data	a matrix or data frame of data values.
permu	a character string specifying the method of generating permutation samples, must be one of "row" (default), "col".
row	a numeric vector of the order of row categories. "row =NULL" indicates the categories are disordered.
col	a numeric vector of the order of colnum categories."col =NULL" indicates the categories are disordered.
fix	a character string specifying the group characteristic, must be one of "row" (default), "col". This argument is used for JT test when both "row" and "col" arguments are not "NULL".
samplenum	a number specifying the number of sampling.

Details

The test can deal with the contingency tables with or without ordered categories. if both `row` and `col` are `null`, the data is viewed as normal contingency table with disordered categories. The test will use chi-square statistic.

If one of `row` and `col` is not `null`, the data is viewed as contingency table with one ordered characteristic. The vector you put in represents the order. For another characteristic, the Wilcoxon test can be performed if it has two classes; if the characteristic has several different values, the Kruskal-Wallis test can be used.

if both `row` and `col` are not `null`, that is both characteristics are ordered, the different values of one of the characteristic can be treated as observations and the JT test used for the other characteristic.

Value

method	a character string describing the type of test.
stat	the value of the test statistic with a name describing it.
pval	the p-value for the test.

alternative	a character string describing the alternative hypothesis.
conf.int	95% confidence interval of p-value(presents if either "row" and "col" arguments are not "NULL".)

Author(s)

Jiasheng Zhang, Feng Yu, Yangyang Zhang, Siwei Deng. Tutored by YuKun Liu and Dongdong Xiang.

References

Higgins, J. J. (2004). An introduction to modern nonparametric statistics. Pacific Grove, CA: Brooks/Cole.

Examples

```
## generate a contingency table
x1=c(10,12, 17, 30)
x2=c( 9, 9, 11, 35)
x3=c( 7, 8, 12, 43)
data = rbind(x1,x2,x3)

## without ordered categories
permu_table(data)

## with ordered column categories
permu_table(data , col = c(1,2,3,4) )

## with ordered row categories
permu_table(data , row = c(1,2,3))

## with ordered row and column categories
permu_table(data , col = c(1,2,3,4),row = c(1,2,3),fix = ”row”)
```

RMD_test   *RMD Test*

Description

Perform two-sample RMD test on vectors of data.

Usage

```
RMD_test(x , y , alternative = ”greater” , mu1=median(x) , mu2=median(y),
method_p=”exact” , samplenum = 2000 , samplemethod = ”R” , conf.level.sample = 0.95 )
```

Arguments

x	numeric vector of data values.
y	numeric vector of data values.
alternative	a character string specifying the alternative hypothesis, must be one of "two.sided" (default), "greater" or "less".
mu1	mean value of x, or median if not given
mu2	mean value of y, or median if not given
method_p	a character string specifying the method of calculating p-value, must be one of "exact" (default), "sampling".
samplenum	The number of samples
samplemethod	a discrete value indicating the method of sampling. "S" represents sample function sampling; "W" represents Cistern Sampling; "R" represents Put-back sampling.
conf.level.sample	
	p-value confidence level for SRS sampling

Details

The test is to test the difference of deviance of two samples, which has different sample means.

The arguments samplenum and samplemethod only work when method_p="sampling".

Value

method	the test uesd
stat	RMD of the original data.
pval	p-value for the test.
conf.int	the confidence interval for p-value(only if method_p = "sampling")
alternative	a character string describing the alternative hypothesis

Author(s)

Jiasheng Zhang, Feng Yu, Yangyang Zhang, Siwei Deng. Tutored by YuKun Liu and Dongdong Xiang.

References

Higgins, J. J. (2004). An introduction to modern nonparametric statistics. Pacific Grove, CA: Brooks/Cole.

Examples

```
## A simple example
x=c(16.55, 15.36, 15.94, 16.43, 16.01)
y=c(16.05, 15.98, 16.10, 15.88, 15.91)
RMD_test(x , y , alternative = "greater" )
```

siegel_tukey   *Siegel-Tukey Test*

Description

Performs two-sample Siegel-Tukey test on vectors of data.

Usage

```
siegel_tukey (x,y,adjust.median=FALSE,...)
```

Arguments

x	numeric vector of data values.
y	numeric vector of data values.
adjust.median	a logical indicating whether you want the adjusted median, which represents difference of each value with median.
...	The input from twosample_test function.

Details

The test is about the scale parameter, to test the deviance. The arguments you want to modify is almost same with twosample_test.

Value

method	the test used.
stat	the statistic of the original data.
conf.int	the confidence interval for p-value(only if method_p = "sampling")
pval	p-value for the test
alternative	a character string describing the alternative hypothesis.

Author(s)

Jiasheng Zhang, Feng Yu, Yangyang Zhang, Siwei Deng. Tutored by YuKun Liu and Dongdong Xiang.

References

Higgins, J. J. (2004). An introduction to modern nonparametric statistics. Pacific Grove, CA: Brooks/Cole.

Examples

```
## A simple example
x <- c(33, 62, 84, 85, 88, 93, 97)
y<-c(4, 16, 48, 51, 66, 98)
siegel_tukey(x,y,adjust.median=FALSE)
```

---

twosample_test   *Comprehensive two-sample permutation tests*

Description

Perform two-sample permutation test on vectors of data.

Usage

```
twosample_test (x , y , alternative = "greater" , score = "wilcoxon" ,
method_p = "sampling" , samplenum = 2000 ,samplemethod="R",
conf.level.sample = 0.95 , conf.diff = TRUE, conf.level.diff = 0.95)
```

Arguments

x	numeric vector of data values.
y	numeric vector of data values.
alternative	a character string specifying the alternative hypothesis, must be one of "two.sided", "greater"(default) or "less".
score	a discrete value indicating the type of score. There are "original", "Wilcoxon", "van" and "exp" to be selected.
method_p	a string indicating what method to use for p-value. "sampling" represents sampling; "asymptotic" represents using large sample approximations; "exact" represents Iterate through all combinations.
samplenum	The number of samples
samplemethod	a discrete value indicating the method of sampling. "S" represents sample function sampling; "W" represents Cistern Sampling; "R" represents Put-back sampling.
conf.level.sample	
	p-value confidence level for SRS sampling
conf.diff	a logical indicating whether to calculate the confidence interval of drift parameters.
conf.level.diff	
	the level of confidence of drift parameters.

Details

score has 4 options: "original", "Wilcoxon", "van" and "exp". When choosing "original", the test is based on the original data; if score = ″Wilcoxon″, the test is baesd on rank-sum; if score = ″van″, the test is based on Van der Waerden score; if score = ″exp″, the test is based on exponential score.

samplenum and samplemethod only work when method_p=″sampling″. Similarly, conf.level.diff only works when conf.diff =TRUE.

Value

method	the test used.
score	the score which is used.
stat	the statistic of the original data.
conf.int	the confidence interval for p-value(only if method_p = "sampling")
pval	p-value for the test
alternative	a character string describing the alternative hypothesis.
addition	a character string describing the Hodges-Lehmann estimate and the confidence interval of the drift parameter.

Author(s)

Jiasheng Zhang, Feng Yu, Yangyang Zhang, Siwei Deng. Tutored by YuKun Liu and Dongdong Xiang.

References

Higgins, J. J. (2004). An introduction to modern nonparametric statistics. Pacific Grove, CA: Brooks/Cole.

Examples

```
## A simple example
x = c(1,2,3,4,5)
y = c(2,3,4,5,6)
twosample_test(x,y,samplemethod = ″R″ )
```

Index