

# Package ‘HDLSSkST’

August 7, 2020

**Type** Package

**Title** Distribution-Free Exact High Dimensional Low Sample Size  
k-Sample Tests

**Version** 1.0.1

**Date** 2020-08-05

**Maintainer** Biplab Paul <biplab.paul@niser.ac.in>

**Description** We construct four new exact level (size) alpha tests for testing the equality of k distributions, which can be conveniently used in high dimensional low sample size setup based on clustering. These tests are easy to implement and distribution-free. Under mild conditions, we have proved the consistency of these tests as the dimension d of each observation grows to infinity, whereas the sample size remains fixed. We also apply step-down-procedure (1979) for multiple testing. Details are in Biplab Paul, Shyamal K De and Anil K Ghosh (2020); Soham Sarkar and Anil K Ghosh (2019) <doi:10.1109/TPAMI.2019.2912599>; William M Rand (1971) <doi:10.1080/01621459.1971.10482356>; Cyrus R Mehta and Nitin R Patel (1983) <doi:10.2307/2288652>; Joseph C Dunn (1973) <doi:10.1080/01969727308546046>; Sture Holm (1979) <doi:10.2307/4615733>.

**License** GPL (>= 2)

**Imports** Rcpp (>= 1.0.3), stats, utils

**LinkingTo** Rcpp

**Author** Biplab Paul [aut, cre],  
Shyamal K. De [aut],  
Anil K. Ghosh [aut]

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2020-08-07 09:42:15 UTC

## R topics documented:

HDLSSkST-package . . . . .	2
AFEIttest . . . . .	3
ARIttest . . . . .	5
FEIttest . . . . .	7

gMADD . . . . .	9
gMADD_DI . . . . .	10
pmf . . . . .	12
randfun . . . . .	13
rctab . . . . .	14
RItest . . . . .	15
<b>Index</b>	<b>18</b>

---

HDLSSkST-package	<i>Distribution-Free Exact High Dimensional Low Sample Size k-Sample Tests</i>
------------------	--

---

**Description**

This package performs some k-sample tests in high dimensional low sample size (HDLSS) regime based on clustering. This package includes four tests, namely (i) RI test, (ii) FEI test, (iii) ARI test, and (iv) AFEI test. These tests are distribution free, and exact level (size)  $\alpha$  test, also consistent as dimension  $d$  goes to  $\infty$ . We also apply Holm’s step-down-procedure (1979) for multiple testing.

**Author(s)**

Biplab Paul, Shyamal K. De and Anil K. Ghosh  
Maintainer: Biplab Paul<biplab.paul@niser.ac.in>

**References**

Biplab Paul, Shyamal K De and Anil K Ghosh (2020). Distribution-free Exact k-sample Tests for High Dimensional Low Sample Size data Based on Clustering.

Soham Sarkar and Anil K Ghosh (2019). On perfect clustering of high dimension, low sample size data, *IEEE transactions on pattern analysis and machine intelligence*, doi:10.1109/TPAMI.2019.2912599.

William M Rand (1971). Objective criteria for the evaluation of clustering methods, *Journal of the American Statistical association*, 66(336):846-850, doi:10.1080/01621459.1971.10482356.

Cyrus R Mehta and Nitin R Patel (1983). A network algorithm for performing Fisher’s exact test in rxc contingency tables, *Journal of the American Statistical Association*, 78(382):427-434, doi:10.2307/2288652.

Joseph C Dunn (1973). A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters, doi:10.1080/01969727308546046.

Sture Holm (1979). A simple sequentially rejective multiple test procedure, *Scandinavian journal of statistics*, 65-70, doi:10.2307/4615733.

AFEItest

*k*-Sample AFEI Test of Equal Distributions**Description**

Performs the distribution free exact  $k$ -sample test for equality of multivariate distributions in the HDLSS regime. This an aggregate test of the two sample versions of the FEI test over  $\frac{k(k-1)}{2}$  numbers of two-sample comparisons, and the test statistic is the minimum of these two sample FEI test statistics. Holm's step-down-procedure (1979) is applied for multiple testing.

**Usage**

```
AFEItest(M, sizes, randomization = TRUE, clust_alg = "knwClustNo",
s_fn = 1, lb = 1, n_sts = 1000, alpha = 0.05)
```

**Arguments**

M	$n \times d$ observations matrix of pooled sample, the observations should be grouped by their respective classes
sizes	vector of sample sizes
randomization	logical; if TRUE (default), randomization test and FALSE, non-randomization test
clust_alg	"knwClustNo"(default) or "estclustNo"; modified K-means algorithm used for clustering
s_fn	function required for clustering, 1 (default) for $1 - \exp(-t)$ , 2 for $\log(1 + t)$ , 3 for $t$
lb	each observation is partitioned into some numbers of smaller vectors of same length $lb$ , default: 1
n_sts	number of simulation of the test statistic, default: 1000
alpha	numeric, confidence level $\alpha$ , default: 0.05

**Value**

AFEItest returns a list containing the following items:

AFEIStat	value of the observed test statistic
AFCutoff	cut-off of the test
randomGamma	randomized coefficient of the test
decisionAFEI	if returns 1, reject the null hypothesis and if returns 0, fails to reject the null hypothesis
multipleTest	indicates where two populations are different

**Author(s)**

Biplab Paul, Shyamal K. De and Anil K. Ghosh

Maintainer: Biplab Paul<biplab.paul@niser.ac.in>

**References**

Biplab Paul, Shyamal K De and Anil K Ghosh (2020). Distribution-free Exact k-sample Tests for High Dimensional Low Sample Size data Based on Clustering.

Cyrus R Mehta and Nitin R Patel (1983). A network algorithm for performing Fisher's exact test in rxc contingency tables, *Journal of the American Statistical Association*, 78(382):427-434, doi:10.2307/2288652.

Sture Holm (1979). A simple sequentially rejective multiple test procedure, *Scandinavian journal of statistics*, 65-70, doi:10.2307/4615733.

**Examples**

```
# multivariate normal distribution:
# generate data with dimension d = 500
set.seed(151)
n1=n2=n3=n4=10
d = 500
I1 <- matrix(rnorm(n1*d,mean=0,sd=1),n1,d)
I2 <- matrix(rnorm(n2*d,mean=0.5,sd=1),n2,d)
I3 <- matrix(rnorm(n3*d,mean=1,sd=1),n3,d)
I4 <- matrix(rnorm(n4*d,mean=1.5,sd=1),n4,d)
X <- as.matrix(rbind(I1,I2,I3,I4))
#AFEI test:
results <- AFEItest(M=X, sizes = c(n1,n2,n3,n4))

## outputs:
results$AFEIStat
#[1] 5.412544e-06

results$AFCutoff
#[1] 0.0109604

results$randomGamma
#[1] 0

results$decisionAFEI
#[1] 1

results$multipleTest
# Population.1 Population.2 pvalues criticalValues
#1          1          2      0    0.008333333
#2          1          3      0    0.010000000
#3          1          4      0    0.012500000
#4          2          3      0    0.016666667
#5          2          4      0    0.025000000
#6          3          4      0    0.050000000
```

ARItest

*k*-Sample ARI Test of Equal Distributions**Description**

Performs the distribution free exact  $k$ -sample test for equality of multivariate distributions in the HDLSS regime. This an aggregate test of the two sample versions of the RI test over  $\frac{k(k-1)}{2}$  numbers of two-sample comparisons, and the test statistic is the minimum of these two sample RI test statistics. Holm's step-down-procedure (1979) is applied for multiple testing.

**Usage**

```
ARItest(M, sizes, randomization = TRUE, clust_alg = "knwClustNo",
s_fn = 1, lb = 1, n_sts = 1000, alpha = 0.05)
```

**Arguments**

M	$n \times d$ observations matrix of pooled sample, the observations should be grouped by their respective classes
sizes	vector of sample sizes
randomization	logical; if TRUE (default), randomization test and FALSE, non-randomization test
clust_alg	"knwClustNo"(default) or "estclustNo"; modified K-means algorithm used for clustering
s_fn	function required for clustering, 1 (default) for $1 - \exp(-t)$ , 2 for $\log(1 + t)$ , 3 for $t$
lb	each observation is partitioned into some numbers of smaller vectors of same length $lb$ , default: 1
n_sts	number of simulation of the test statistic, default: 1000
alpha	numeric, confidence level $\alpha$ , default: 0.05

**Value**

ARItest returns a list containing the following items:

ARISat	value of the observed test statistic
Cutoff	cut-off of the test
randomGamma	randomized coefficient of the test
decisionARI	if returns 1, reject the null hypothesis and if returns 0, fails to reject the null hypothesis
multipleTest	indicates where two populations are different

**Author(s)**

Biplab Paul, Shyamal K. De and Anil K. Ghosh

Maintainer: Biplab Paul<biplab.paul@niser.ac.in>

**References**

Biplab Paul, Shyamal K De and Anil K Ghosh (2020). Distribution-free Exact k-sample Tests for High Dimensional Low Sample Size data Based on Clustering.

William M Rand (1971). Objective criteria for the evaluation of clustering methods, *Journal of the American Statistical association*, 66(336):846-850, doi:10.1080/01621459.1971.10482356.

Sture Holm (1979). A simple sequentially rejective multiple test procedure, *Scandinavian journal of statistics*, 65-70, doi:10.2307/4615733.

**Examples**

```
# multivariate normal distribution:
# generate data with dimension d = 500
set.seed(151)
n1=n2=n3=n4=10
d = 500
I1 <- matrix(rnorm(n1*d,mean=0,sd=1),n1,d)
I2 <- matrix(rnorm(n2*d,mean=0.5,sd=1),n2,d)
I3 <- matrix(rnorm(n3*d,mean=1,sd=1),n3,d)
I4 <- matrix(rnorm(n4*d,mean=1.5,sd=1),n4,d)
X <- as.matrix(rbind(I1,I2,I3,I4))
#ARI test:
results <- ARItest(M=X, sizes = c(n1,n2,n3,n4))

## outputs:
results$ARISat
#[1] 0

results$ARICutoff
#[1] 0.3368421

results$randomGamma
#[1] 0

results$decisionARI
#[1] 1

results$multipleTest
# Population.1 Population.2 pvalues criticalValues
#1          1          2      0    0.008333333
#2          1          3      0    0.010000000
#3          1          4      0    0.012500000
#4          2          3      0    0.016666667
#5          2          4      0    0.025000000
#6          3          4      0    0.050000000
```

FEItest

*k*-Sample FEI Test of Equal Distributions**Description**

Performs the distribution free exact k-sample test for equality of multivariate distributions in the HDLSS regime.

**Usage**

```
FEItest(M, labels, sizes, randomization = TRUE,
        clust_alg = "knwClustNo", s_fn = 1, lb = 1, n_sts = 1000,
        alpha = 0.05)
```

**Arguments**

<b>M</b>	$n \times d$ observations matrix of pooled sample, the observations should be grouped by their respective classes
<b>labels</b>	length $n$ vector of membership index of observations
<b>sizes</b>	vector of sample sizes
<b>randomization</b>	logical; if TRUE (default), randomization test and FALSE, non-randomization test
<b>clust_alg</b>	"knwClustNo"(default) or "estclustNo"; modified K-means algorithm used for clustering
<b>s_fn</b>	function required for clustering, 1 (default) for $1 - \exp(-t)$ , 2 for $\log(1 + t)$ , 3 for $t$
<b>lb</b>	each observation is partitioned into some numbers of smaller vectors of same length $lb$ , default: 1
<b>n_sts</b>	number of simulation of the test statistic, default: 1000
<b>alpha</b>	numeric, confidence level $\alpha$ , default: 0.05

**Value**

FEItest returns a list containing the following items:

<b>estClustLabel</b>	a vector of length $n$ of estimated class membership index of all observations
<b>obsCtyTab</b>	observed contingency table
<b>ObservedProb</b>	value of the observed test statistic
<b>FCutoff</b>	cut-off of the test
<b>randomGamma</b>	randomized coefficient of the test
<b>estPvalue</b>	estimated p-value of the test
<b>decisionF</b>	if returns 1, reject the null hypothesis and if returns 0, fails to reject the null hypothesis
<b>estClustNo</b>	total number of the estimated classes

**Author(s)**

Biplab Paul, Shyamal K. De and Anil K. Ghosh

Maintainer: Biplab Paul<biplab.paul@niser.ac.in>

**References**

Biplab Paul, Shyamal K De and Anil K Ghosh (2020). Distribution-free Exact k-sample Tests for High Dimensional Low Sample Size data Based on Clustering.

Cyrus R Mehta and Nitin R Patel (1983). A network algorithm for performing Fisher's exact test in rxc contingency tables, *Journal of the American Statistical Association*, 78(382):427-434, doi:10.2307/2288652.

**Examples**

```
# multivariate normal distribution:
# generate data with dimension d = 500
set.seed(151)
n1=n2=n3=n4=10
d = 500
I1 <- matrix(rnorm(n1*d,mean=0,sd=1),n1,d)
I2 <- matrix(rnorm(n2*d,mean=0.5,sd=1),n2,d)
I3 <- matrix(rnorm(n3*d,mean=1,sd=1),n3,d)
I4 <- matrix(rnorm(n4*d,mean=1.5,sd=1),n4,d)
levels <- c(rep(0,n1), rep(1,n2), rep(2,n3), rep(3,n4))
X <- as.matrix(rbind(I1,I2,I3,I4))
#FEI test:
results <- FEItest(M=X, labels=levels, sizes = c(n1,n2,n3,n4))

## outputs:
results$estClustLabel
#[1] 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3

results$obsCtyTab
#      [,1] [,2] [,3] [,4]
#[1,]  10    0    0    0
#[2,]   0   10    0    0
#[3,]   0    0   10    0
#[4,]   0    0    0   10

results$ObservedProb
#[1] 2.125236e-22

results$FCutoff
#[1] 1.115958e-07

results$randomGamma
#[1] 0

results$estPvalue
#[1] 0
```



```
results$decisionF
#[1] 1
```

---

gMADD	<i>Modified K-Means Algorithm by Using a New Dissimilarity Measure, MADD</i>
-------	--

---

### Description

Performs modified K-means algorithm by using a new dissimilarity measure, called MADD, and provides estimated cluster (class) labels or memberships of observations.

### Usage

```
gMADD(s_fn, n_clust, lb, M)
```

### Arguments

s_fn	function required for clustering, 1 for $1 - \exp(-t)$ , 2 for $\log(1 + t)$ , 3 for $t$
n_clust	total number of the classes in the whole observations
lb	each observation is partitioned into some numbers of smaller vectors of same length $lb$
M	$n \times d$ observations matrix of pooled sample, the observations should be grouped by their respective classes

### Value

a vector of length n of estimated cluster (class) labels of observations

### Author(s)

Biplab Paul, Shyamal K. De and Anil K. Ghosh

Maintainer: Biaplal Paul<biplab.paul@niser.ac.in>

### References

Biplab Paul, Shyamal K De and Anil K Ghosh (2020). Distribution-free Exact k-sample Tests for High Dimensional Low Sample Size data Based on Clustering.

Soham Sarkar and Anil K Ghosh (2019). On perfect clustering of high dimension, low sample size data, *IEEE transactions on pattern analysis and machine intelligence*, doi:10.1109/TPAMI.2019.2912599.

## Examples

```
# Modified K-means algorithm:
# multivariate normal distribution
# generate data with dimension d = 500
set.seed(151)
n1=n2=n3=n4=10
d = 500
I1 <- matrix(rnorm(n1*d,mean=0,sd=1),n1,d)
I2 <- matrix(rnorm(n2*d,mean=0.5,sd=1),n2,d)
I3 <- matrix(rnorm(n3*d,mean=1,sd=1),n3,d)
I4 <- matrix(rnorm(n4*d,mean=1.5,sd=1),n4,d)
n_cl <- 4
X <- as.matrix(rbind(I1,I2,I3,I4))
gMADD(1,n_cl,1,X)

## outputs:
#[1] 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3
```

---

gMADD\_DI

---

*Modified K-Means Algorithm by Using a New Dissimilarity Measure,  
MADD and DUNN Index*


---

## Description

Performs modified K-means algorithm by using a new dissimilarity measure, called MADD and DUNN index, and provides estimated cluster (class) labels or memberships and corresponding DUNN index of the observations.

## Usage

```
gMADD_DI(s_fn, kmax, lb, M)
```

## Arguments

s_fn	function required for clustering, 1 for $1 - \exp(-t)$ , 2 for $\log(1 + t)$ , 3 for $t$
kmax	maximum value of total number of clusters to estimate total number of the classes in the whole observations
lb	each observation is partitioned into some numbers of smaller vectors of same length $lb$
M	$n \times d$ observations matrix of pooled sample, the observations should be grouped by their respective classes

## Details

DUNN index is used for cluster validation, but here we use it to estimate total number of cluster  $k$  by  $\hat{k} = \operatorname{argmax}_{2 \leq k' \leq k^*} DI(k')$ . Here  $DI(k')$  represents the DUNN index and we use  $k^* = 2 * k$ .

**Value**

a  $kmax \times (n + 1)$  matrix of the estimated cluster (class) labels and corresponding DUNN indexes of observations

**Note**

The result of this gMADD\_DI function is a matrix. The 1st row of this matrix doesn't provide anything about estimated class labels or DUNN index of observations since the DUNN index is only defined for  $k \geq 2$ . The last column of this matrix represents the DUNN indexes. The estimated cluster labels of observations are calculated by finding out the corresponding row of maximum DUNN index.

**Author(s)**

Biplab Paul, Shyamal K. De and Anil K. Ghosh

Maintainer: Biplab Paul<biplab.paul@niser.ac.in>

**References**

Biplab Paul, Shyamal K De and Anil K Ghosh (2020). Distribution-free Exact k-sample Tests for High Dimensional Low Sample Size data Based on Clustering.

Soham Sarkar and Anil K Ghosh (2019). On perfect clustering of high dimension, low sample size data, *IEEE transactions on pattern analysis and machine intelligence*, doi:10.1109/TPAMI.2019.2912599.

Joseph C Dunn (1973). A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters, doi:10.1080/01969727308546046.

**Examples**

```
# Modified K-means algorithm:
# multivariate normal distribution
# generate data with dimension d = 500
set.seed(151)
n1=n2=n3=n4=10
d = 500
I1 <- matrix(rnorm(n1*d,mean=0,sd=1),n1,d)
I2 <- matrix(rnorm(n2*d,mean=0.5,sd=1),n2,d)
I3 <- matrix(rnorm(n3*d,mean=1,sd=1),n3,d)
I4 <- matrix(rnorm(n4*d,mean=1.5,sd=1),n4,d)
n_cl <- 4
N <- n1+n2+n3+n4
X <- as.matrix(rbind(I1,I2,I3,I4))
dvec_di_mat <- gMADD_DI(1,2*n_cl,1,X)
est_no_cl <- which.max(dvec_di_mat[, (N+1)])
dvec_di_mat[est_no_cl,1:N]

## outputs:
#[1] 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3
```

pmf

*Generalized Hypergeometric Probability***Description**

A function that provides the probability of observing an  $r \times c$  contingency table using generalized hypergeometric probability.

**Usage**

```
pmf(M)
```

**Arguments**

M  $r \times c$  contingency table

**Value**

a single value between 0 and 1

**Author(s)**

Biplab Paul, Shyamal K. De and Anil K. Ghosh

Maintainer: Bioplal Paul<bioplal.paul@niser.ac.in>

**References**

Biplab Paul, Shyamal K De and Anil K Ghosh (2020). Distribution-free Exact k-sample Tests for High Dimensional Low Sample Size data Based on Clustering.

Cyrus R Mehta and Nitin R Patel (1983). A network algorithm for performing Fisher's exact test in rxc contingency tables, *Journal of the American Statistical Association*, 78(382):427-434, doi:10.2307/2288652.

**Examples**

```
# Generalized hypergeometric probability of rxc Contingency Table:
mat <- matrix(1:20,5,4, byrow = TRUE)
pmf(mat)

## outputs:
#[1] 4.556478e-09
```

---

randfun

*Rand Index*


---

**Description**

Measures to compare the dissimilarity of exact cluster labels (memberships) and estimated cluster labels (memberships) of the observations.

**Usage**

```
randfun(lvel, dv)
```

**Arguments**

lvel	exact cluster labels of the observations
dv	estimated cluster labels of the observations

**Value**

a single value between 0 and 1

**Author(s)**

Biplab Paul, Shyamal K. De and Anil K. Ghosh

Maintainer: Bioplal Paul<bioplal.paul@niser.ac.in>

**References**

William M Rand (1971). Objective criteria for the evaluation of clustering methods, *Journal of the American Statistical association*, 66(336):846-850, doi:10.1080/01621459.1971.10482356.

**Examples**

```
# Measures of dissimilarity:
ex1 <- c(rep(0,5), rep(1,5), rep(2,5), rep(3,5))
el <- c(0,0,1,0,0,1,2,1,0,1,2,2,3,2,2,3,2,3,1,3)
randfun(ex1,el)

## outputs:
#[1] 0.2368421
```

---

rctab	<i>Generates an <math>r \times c</math> Contingency Table</i>
-------	---

---

**Description**

A function that generates an  $r \times c$  contingency table with the same marginal totals as given  $r \times c$  contingency table.

**Usage**

```
rctab(M)
```

**Arguments**

M  $r \times c$  contingency table

**Value**

generated  $r \times c$  contingency table

**Author(s)**

Biplab Paul, Shyamal K. De and Anil K. Ghosh

Maintainer: Biplab Paul<biplab.paul@niser.ac.in>

**References**

Cyrus R Mehta and Nitin R Patel (1983). A network algorithm for performing Fisher's exact test in rxc contingency tables, *Journal of the American Statistical Association*, 78(382):427-434, doi:10.2307/2288652.

**Examples**

```
# Generation of rxc Contingency Table:
set.seed(151)
mat <- matrix(1:20,5,4, byrow = TRUE)
rctab(mat)

## outputs:
#      [,1] [,2] [,3] [,4]
# [1,]    3    4    0    3
# [2,]    4    5   10    7
# [3,]    8    7   12   15
# [4,]   18   16   13   11
# [5,]   12   18   20   24
```

---

RIttest	<i>k</i> -Sample RI Test of Equal Distributions
---------	---

---

**Description**

Performs the distribution free exact  $k$ -sample test for equality of multivariate distributions in the HDLSS regime.

**Usage**

```
RIttest(M, labels, sizes, randomization = TRUE,
        clust_alg = "knwClustNo", s_fn = 1, lb = 1, n_sts = 1000,
        alpha = 0.05)
```

**Arguments**

M	$n \times d$ observations matrix of pooled sample, the observations should be grouped by their respective classes
labels	length $n$ vector of membership index of observations
sizes	vector of sample sizes
randomization	logical; if TRUE (default), randomization test and FALSE, non-randomization test
clust_alg	"knwClustNo"(default) or "estclustNo"; modified K-means algorithm used for clustering
s_fn	function required for clustering, 1 (default) for $1 - \exp(-t)$ , 2 for $\log(1 + t)$ , 3 for $t$
lb	each observation is partitioned into some numbers of smaller vectors of same length $lb$ , default: 1
n_sts	number of simulation of the test statistic, default: 1000
alpha	numeric, confidence level $\alpha$ , default: 0.05

**Value**

RIttest returns a list containing the following items:

estClustLabel	a vector of length $n$ of estimated class membership index of all observations
obsCtyTab	observed contingency table
ObservedRI	value of the observed test statistic
RICutoff	cut-off of the test
randomGamma	randomized coefficient of the test
estPvalue	estimated p-value of the test
decisionRI	if returns 1, reject the null hypothesis and if returns 0, fails to reject the null hypothesis
estClustNo	total number of the estimated classes

**Author(s)**

Biplab Paul, Shyamal K. De and Anil K. Ghosh

Maintainer: Bioplal Paul<bioplal.paul@niser.ac.in>

**References**

Biplab Paul, Shyamal K De and Anil K Ghosh (2020). Distribution-free Exact k-sample Tests for High Dimensional Low Sample Size data Based on Clustering.

William M Rand (1971). Objective criteria for the evaluation of clustering methods, *Journal of the American Statistical association*, 66(336):846-850, doi:10.1080/01621459.1971.10482356.

**Examples**

```
# multivariate normal distribution:
# generate data with dimension d = 500
set.seed(151)
n1=n2=n3=n4=10
d = 500
I1 <- matrix(rnorm(n1*d,mean=0,sd=1),n1,d)
I2 <- matrix(rnorm(n2*d,mean=0.5,sd=1),n2,d)
I3 <- matrix(rnorm(n3*d,mean=1,sd=1),n3,d)
I4 <- matrix(rnorm(n4*d,mean=1.5,sd=1),n4,d)
levels <- c(rep(0,n1), rep(1,n2), rep(2,n3), rep(3,n4))
X <- as.matrix(rbind(I1,I2,I3,I4))
# RI test:
results <- RItest(M=X, labels=levels, sizes = c(n1,n2,n3,n4))

## outputs:
results$estClustLabel
#[1] 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3

results$obsCtyTab
#      [,1] [,2] [,3] [,4]
#[1,]   10    0    0    0
#[2,]    0   10    0    0
#[3,]    0    0   10    0
#[4,]    0    0    0   10

results$ObservedRI
#[1] 0

results$RICutoff
#[1] 0.3307692

results$randomGamma
#[1] 0

results$estPvalue
#[1] 0

results$decisionRI
```



#[1] 1

# Index

- \* **AFEItest function**  
AFEItest, [3](#)
- \* **ARItest function**  
ARItest, [5](#)
- \* **FEItest function**  
FEItest, [7](#)
- \* **HDLSS**  
HDLSSkST-package, [2](#)
- \* **RItest function**  
RItest, [15](#)
- \* **gMADD function**  
gMADD, [9](#)
- \* **gMADD\_DI function**  
gMADD\_DI, [10](#)
- \* **package**  
HDLSSkST-package, [2](#)
- \* **pmf function**  
pmf, [12](#)
- \* **randfun function**  
randfun, [13](#)
- \* **rctab function**  
rctab, [14](#)

AFEItest, [3](#)  
ARItest, [5](#)

FEItest, [7](#)

gMADD, [9](#)  
gMADD\_DI, [10](#)

HDLSSkST-package, [2](#)

pmf, [12](#)

randfun, [13](#)  
rctab, [14](#)  
RItest, [15](#)