

Package: Domean (via r-universe)

March 5, 2025

Title Distributed Online Mean Tests

Date 2025-02-28

Version 0.1

Description Distributed Online Mean Tests is a powerful tool designed to efficiently process and analyze distributed datasets. It enables users to perform mean tests in an online, distributed manner, making it highly suitable for large-scale data analysis. By leveraging advanced computational techniques, 'Domean' ensures robust and scalable solutions for statistical analysis, particularly in scenarios where data is dispersed across multiple nodes or sources. This package is ideal for researchers and practitioners working with high-dimensional data, providing a flexible and efficient framework for mean testing. The philosophy of 'Domean' is described in Guo G.(2025) <[doi:10.1016/j.physa.2024.130308](https://doi.org/10.1016/j.physa.2024.130308)>.

License MIT + file LICENSE

RoxygenNote 7.3.2

Imports stats, MASS

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Guangbao Guo [aut, cre]
(<<https://orcid.org/0000-0002-4115-6218>>), Qianwen Liu [aut]

Maintainer Guangbao Guo <ggb11111111@163.com>

Depends R (>= 3.5.0)

Date/Publication 2025-03-04 10:10:09 UTC

Repository <https://guangbaog.r-universe.dev>

RemoteUrl <https://github.com/cran/Domean>

RemoteRef HEAD

RemoteSha fd5283f54cd43d20ed8dab3549a916f21f8aee1b

Contents

CLX	2
covclx	3
CQ2	4
CZZZ	5
SKK	6
zwl	7
Index	9

 CLX

Two-Sample CLX Test for High-Dimensional Data

Description

Performs a two-sample CLX test to compare the means of two high-dimensional samples. This test is suitable for situations where the number of variables (p) is large relative to the sample sizes.

Usage

CLX(X, Y, alpha)

Arguments

X	A numeric matrix representing the first sample, where rows are variables and columns are observations.
Y	A numeric matrix representing the second sample, where rows are variables and columns are observations.
alpha	The significance level for the test (e.g., 0.05).

Details

The CLX test is designed to handle high-dimensional data by estimating the covariance matrix, applying thresholding to reduce noise, and transforming the data to white noise. The test statistic is calculated based on the maximum squared difference between the mean vectors, weighted by the inverse of the variances.

Value

A list containing the following components:

statistics	The test statistic.
p.value	The p-value of the test.
alternative	The alternative hypothesis ("two.sided").
method	The method used ("Two-Sample CLX test").

See Also

[eigen](#): Used for eigen-decomposition of the covariance matrix. [solve](#): Used to compute the inverse of the covariance matrix.

Examples

```
# Example usage:
set.seed(123)
p <- 100 # Number of variables
n1 <- 20 # Sample size for X
n2 <- 20 # Sample size for Y
X <- matrix(rnorm(n1 * p), nrow = p, ncol = n1)
Y <- matrix(rnorm(n2 * p, mean = 0.5), nrow = p, ncol = n2)
result <- CLX(X, Y, alpha = 0.05)
print(result)
```

 covclx

Two-Sample Covariance Test for High-Dimensional Data

Description

Performs a test to compare the covariance matrices of two high-dimensional samples. This test is designed for situations where the number of variables (p) is large relative to the sample sizes (n_1) and (n_2) .

Usage

```
covclx(X, Y)
```

Arguments

X	A numeric matrix representing the first sample, where rows are observations and columns are variables.
Y	A numeric matrix representing the second sample, where rows are observations and columns are variables.

Details

This function tests the null hypothesis that the covariance matrices of two samples are equal:

$$H_0 : \Sigma_1 = \Sigma_2$$

against the alternative hypothesis that they are not equal.

The test statistic is based on the maximum normalized squared difference between the two sample covariance matrices. The p-value is computed using an extreme value distribution.

Value

A list containing the following components:

stat	The test statistic.
pval	The p-value of the test.

See Also

[cov](#): Used for calculating sample covariance matrices.

Examples

```
# Example usage:
set.seed(123)
n1 <- 20
n2 <- 30
p <- 50
X <- matrix(rnorm(n1 * p), nrow = n1, ncol = p)
Y <- matrix(rnorm(n2 * p), nrow = n2, ncol = p)
result <- covclx(X, Y)
print(result)
```

CQ2

Two-Sample CQ Test for High-Dimensional Covariance Matrices

Description

Performs a two-sample test to compare the covariance matrices of two high-dimensional samples. This test is designed for situations where the number of variables (p) is large relative to the sample sizes (n_1) and (n_2) .

Usage

```
CQ2(X, Y)
```

Arguments

X	A numeric matrix representing the first sample, where rows are variables and columns are observations.
Y	A numeric matrix representing the second sample, where rows are variables and columns are observations.

Details

The test statistic is based on the difference between the sample covariance matrices, normalized by their variances. The p-value is computed using a normal approximation.

Value

A list containing the following components:

statistics	The test statistic (Q_n) .
p.value	The p-value of the test.
alternative	The alternative hypothesis ("two.sided").
method	The method used ("Two-Sample CQ test").

Examples

```
# Example usage:
set.seed(123)
p <- 50
n1 <- 30
n2 <- 30
X <- matrix(rnorm(n1 * p), nrow = p, ncol = n1)
Y <- matrix(rnorm(n2 * p), nrow = p, ncol = n2)
result <- CQ2(X, Y)
print(result)
```

CZZZ

High-Dimensional Two-Sample Mean Test

Description

Conducts a high-dimensional two-sample mean test with optional variable filtering. This function performs both non-studentized and studentized tests to determine whether the means of two groups are significantly different.

Usage

```
CZZZ(X, Y, m = 2500, filter = TRUE, alpha = 0.05)
```

Arguments

X	Matrix representing the first group of data (variables in rows, observations in columns).
Y	Matrix representing the second group of data (variables in rows, observations in columns).
m	Number of bootstrap samples used for the test (default is 2500).
filter	Logical parameter indicating whether to filter variables based on mean differences (default is TRUE).
alpha	Significance level for the test (default is 0.05).

Details

This function performs a high-dimensional two-sample mean test, which is useful when the number of variables (p) is much larger than the number of observations (n). The function includes an optional filtering step to reduce the number of variables based on the difference in means between the two groups.

Value

A list containing the results of the non-studentized and studentized tests. Each result includes:

<code>statistics</code>	The test statistic.
<code>p.value</code>	The p-value of the test.
<code>alternative</code>	The alternative hypothesis (two-sided).
<code>method</code>	The method description.

Examples

```
# Example usage:
library(MASS)
set.seed(123)
X <- matrix(rnorm(1000), nrow = 100, ncol = 10) # 100 variables, 10 observations
Y <- matrix(rnorm(1000, mean = 0.5), nrow = 100, ncol = 10) # Different mean
result <- CZZZ(X, Y, m = 1000, filter = TRUE, alpha = 0.05)
print(result)
```

SKK

*High-Dimensional Two-Sample Mean Test***Description**

Conducts a high-dimensional two-sample mean test using a modified Hotelling's T-squared statistic. This test is suitable for cases where the number of variables (p) is larger than the sample size (n).

Usage

```
SKK(X, Y)
```

Arguments

<code>X</code>	Matrix representing the first sample (rows are observations, columns are variables).
<code>Y</code>	Matrix representing the second sample (rows are observations, columns are variables).

Details

This function implements a high-dimensional two-sample mean test by adjusting the Hotelling's T-squared statistic. It uses diagonal matrices and a correction factor to handle high-dimensional data.

Value

A list containing:

TSvalue	The test statistic value.
pvalue	The p-value of the test.

Examples

```
# Example usage:
set.seed(123)
X <- matrix(rnorm(200), nrow = 10, ncol = 20) # 10 samples, 20 variables
Y <- matrix(rnorm(200, mean = 0.5), nrow = 10, ncol = 20) # Different mean
result <- SKK(X, Y)
print(result)
# Output:
# TSvalue: The test statistic value
# pvalue: The p-value indicating the significance of the test
```

zwl

High-Dimensional Two-Sample Mean Test with Centering Adjustment

Description

Conducts a high-dimensional two-sample mean test with centering adjustment. This function is designed for cases where the number of variables (p) is larger than the sample sizes (n) and (m).

Usage

```
zwl(X, Y, order = 0)
```

Arguments

X	Matrix representing the first sample (rows are observations, columns are variables).
Y	Matrix representing the second sample (rows are observations, columns are variables).
order	Integer specifying the order of centering adjustment (default is 0).

Details

This function performs a high-dimensional two-sample mean test by adjusting the test statistic for centering. It uses a modified t-statistic and estimates the variance to handle high-dimensional data. The function also includes a custom centering adjustment based on the specified order.

Value

A list containing:

<code>statistic</code>	The test statistic value.
<code>pvalue</code>	The p-value of the test.
<code>Tn</code>	The adjusted test statistic before centering.
<code>var</code>	The estimated variance.

Examples

```
# Example usage:
set.seed(123)
X <- matrix(rnorm(200), nrow = 10, ncol = 20) # 10 samples, 20 variables
Y <- matrix(rnorm(200, mean = 0.5), nrow = 10, ncol = 20) # Different mean
result <- zwl(X, Y, order = 0)
print(result)

# Output:
# $statistic: The test statistic value
# $pvalue: The p-value indicating the significance of the test
# $Tn: The adjusted test statistic before centering
# $var: The estimated variance
```


Index

CLX, [2](#)
cov, [4](#)
covclx, [3](#)
CQ2, [4](#)
CZZZ, [5](#)

eigen, [3](#)

SKK, [6](#)
solve, [3](#)

zw1, [7](#)