Package: highd2means (via r-universe)

October 21, 2024

Type Package

Title High-Dimensional Tests for Two Population Mean Vectors

Version 1.0

Date 2024-08-17

Author Michail Tsagris [aut, cre], Manos Papadakis [aut]

Maintainer Michail Tsagris <mtsagris@uoc.gr>

Depends R (>= 4.0)

Imports doParallel, foreach, parallel, stats, Rfast

Suggests highmean, highDmean, PEtests

Description Tests for two high-dimensional population mean vectors. The user has the option to compute the asymptotic, the permutation or the bootstrap based p-value of the test. Some references are: Chen S.X. and Qin Y.L. (2010). <doi:10.1214/09-AOS716>, Cai T.T., Liu W., and Xia Y. (2014) <doi:10.1111/rssb.12034> and Yu X., Li D., Xue L. and Li, R. (2023) <doi:10.1080/01621459.2022.2061354>.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2024-08-21 15:00:02 UTC

Contents

highd2means-package   *High-Dimensional Tests for two Population Mean Vectors*

Description

The package contains a few tests for the case of two high-dimensional population mean vectors. The user has the option to compute the asymptotic, the permutation or the bootstrap based p-value of the test.The tests are based on three other packages, namely the **highmean**, **highDmean** and the **PEtests**. We took the codes, modified them and made them more efficient.

Details

Package:	highd2means
Type:	Package
Version:	1.0
Date:	2024-08-17
License:	GPL-2

Maintainers

Michail Tsagris <mtsagris@uoc.gr>.

Author(s)

Michail Tsagris <mtsagris@uoc.gr> and Manos Papadakis <papadakm95@gmail.com>.

References

Bai Z.D. and Saranadasa H. (1996). Effect of high dimension: by an example of a two sample problem. Statistica Sinica, 6(2): 311–329.

Cai T.T., Liu W., and Xia Y. (2014). Two-sample test of high dimensional means under dependence. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76(2): 349–372.

Chen S.X. and Qin Y.L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. The Annals of Statistics, 38(2) 808–835.

Chen S.X., Li J., and Zhong P.S. (2014). Two-Sample Tests for High Dimensional Means with Thresholding and Data Transformation. arXiv preprint arXiv:1410.2848.

Srivastava M.S. and Du M. (2008). A test for the mean vector with fewer observations than the dimension. Journal of Multivariate Analysis, 99(3): 386–402.

Srivastava, M.S., Katayama, S., and Kano, Y. (2013). A two sample test in high dimensional data. Journal of Multivariate Analysis, 114: 349–358.

Yu X., Li D., Xue L. and Li, R. (2023). Power-enhanced simultaneous test of high-dimensional mean vectors and covariance matrices with application to gene-set testing. Journal of the American Statistical Association, 118(544): 2548–2561.

High-dimensional tests for 2 population mean vectors

High-dimensional tests for 2 population mean vectors

Description

High-dimensional tests for 2 population mean vectors.

Usage

```
means.test(x1, x2, test = "baisara", cov.equal = TRUE, Rp = 1, Rb = 1, ncores = 1)
```

Arguments

x1	A numerical matrix with data, where the number of columns (variables) is greater than the number of rows (observations).
x2	A numerical matrix with data, where the number of columns (variables) is greater than the number of rows (observations).
test	The test to run. "baisara" corresponds to the test proposed by Bai and Saranadasa (1996), "cailiuxia" to the test proposed by Cai, Liu and Xia (2014), "chenlizho" to the test proposed by Chen, Li and Zhong (2014), "chenqin" to the test proposed by Chen and Qin (2010), and "sridu" to the test proposed by Srivastava and Du (2008). Finally, the "skk" corresponds to the test proposed by Srivastava, Katayama, and Kano(2013).
cov.equal	This argument applies to the "cailiuxia", "chenlizho" and "chenqin" tests only. If this is true, it assumes that the two population covariance matrices are equal.
Rp	The number of permutations to perform.
Rb	The number of bootstrap samples to draw.
ncores	The number of cores to use. By default is 1, so no parallel computations take place. You are advised to load the necessary libraries prior to using this argument.

Details

These methods are designed for high-dimensional data. If the data dimension p is smaller than 30 this may result in an inflated Type-I error rate.

Note that these tests were taken from the **highmean** package, except for the "skk" which was taken from the **highDmean** package. The "cailiaxiu" and "chenqin" tests here should be the same as in the function pe.test ("clx" and "cq", respectively), but the authors of that package and of the package **highmean** give different formulas.

Value

A list including:

stat	The test statistic value.
pvalue	The asymptotic p-value of the test.
perm.pvalue	The permutation based p-value will appear only if the argument Rp is greater than 1.
boot.pvalue	The bootstrap based p-value will appear only if the argument Rp is greater than 1.

Author(s)

Michail Tsagris and Manos Papadakis.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr> and Manos Papadakis <papadakm95@gmail.com>.

References

Bai Z.D. and Saranadasa H. (1996). Effect of high dimension: by an example of a two sample problem. Statistica Sinica, 6(2): 311–329.

Cai T.T., Liu W., and Xia Y. (2014). Two-sample test of high dimensional means under dependence. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76(2): 349–372.

Chen S.X. and Qin Y.L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. The Annals of Statistics, 38(2) 808–835.

Chen S.X., Li J., and Zhong P.S. (2014). Two-Sample Tests for High Dimensional Means with Thresholding and Data Transformation. arXiv preprint arXiv:1410.2848.

Srivastava M.S. and Du M. (2008). A test for the mean vector with fewer observations than the dimension. Journal of Multivariate Analysis, 99(3): 386–402.

Srivastava, M.S., Katayama, S., and Kano, Y. (2013). A two sample test in high dimensional data. Journal of Multivariate Analysis, 114: 349–358.

See Also

pe.test

Examples

```
n1 <- 50  ; n2 <- 100 ; p <- 500
x1 <- matrix( rnorm(n1 * p), nrow = n1, ncol = p)
x2 <- matrix( rnorm(n2 * p), nrow = n2, ncol = p)
means.test(x1, x2)
```

Power-enhanced high-dimensional tests for 2 population mean vectors
Power-enhanced high-dimensional tests for 2 population mean vectors

---

Description

Power-enhanced high-dimensional tests for 2 population mean vectors.

Usage

```
pe.test(x1, x2, test = "clx", Rp = 1, Rb = 1, ncores = 1)
```

Arguments

x1	A numerical matrix with data, where the number of columns (variables) is greater than the number of rows (observations).
x2	A numerical matrix with data, where the number of columns (variables) is greater than the number of rows (observations).
test	The test to run, either "clx" corresponding to the test proposed by Cai, Liu and Xia (2014), "cq" corresponding to the test proposed by Chen and Qin (2010). The "pecf" combines the two previous methods using the Cauchy and the Fisher approaches. The "pecomp" updates the "cq" test via the power-enhanced component.
Rp	The number of permutations to perform. This is applicable only for the "clx" and "cq" tests.
Rb	The number of bootstrap samples to draw. This is applicable only for the "clx" and "cq" tests.
ncores	The number of cores to use. By default is 1, so no parallel computations take place. You are advised to load the necessary libraries prior to using this argument.

Details

These methods are designed for high-dimensional data. If the data dimension p is smaller than 30 this may result in an inflated Type-I error rate.

Note that these tests were taken from the **PEtests** package. The "clx" and "cq" tests here should be the same as in the function [means.test](#) ("cailiaxiu" and "chenqin", respectively), but the authors of that package and of the package **highmean** give different formulas.

Value

If the test is "clx", "cq" or "pecomp" a list including:

stat	The test statistic value.
pvalue	The asymptotic p-value of the test.

| perm.pvalue | The permutation based p-value will appear only if the argument Rp is greater than 1. |
| boot.pvalue | The bootstrap based p-value will appear only if the argument Rp is greater than 1. |

If the test is "pecf" a list including:

cq	A list with the results of the "cq" test, the stat and the pvalue.
clx	A list with the results of the "clx" test, the stat and the pvalue.
pecauchy	A list with the results of the Cauchy combination, the stat and the pvalue.
pefisher	A list with the results of the Fisher combination, the stat and the pvalue.

Author(s)

Michail Tsagris and Manos Papadakis.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr> and Manos Papadakis <papadakm95@gmail.com>.

References

Cai T.T., Liu W., and Xia Y. (2014). Two-sample test of high dimensional means under dependence. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76(2): 349–372.

Chen S.X. and Qin Y.L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. The Annals of Statistics, 38(2) 808–835.

Yu X., Li D., Xue L. and Li, R. (2023). Power-enhanced simultaneous test of high-dimensional mean vectors and covariance matrices with application to gene-set testing. Journal of the American Statistical Association, 118(544): 2548–2561.

See Also

means.test

Examples

```
n1 <- 50  ; n2 <- 100 ; p <- 500
x1 <- matrix( rnorm(n1 * p), nrow = n1, ncol = p)
x2 <- matrix( rnorm(n2 * p), nrow = n2, ncol = p)
pe.test(x1, x2, test = "pecf")
```

Index