

# Package ‘HDNRA’

February 27, 2024

**Type** Package

**Title** High-Dimensional Location Testing with Normal-Reference Approaches

**Version** 1.0.0

**Author** Pengfei Wang [aut, cre],  
Shuqi Luo [aut],  
Tianming Zhu [aut],  
Bu Zhou [aut]

**Maintainer** Pengfei Wang <nie23.wp8738@e.ntu.edu.sg>

**Description** We provide a collection of various classical tests and latest normal-reference tests for comparing high-dimensional mean vectors including two-sample and general linear hypothesis testing (GLHT) problem. Some existing tests for two-sample problem [see Bai, Zhidong, and Hewa Saranadasa.(1996) <<https://www.jstor.org/stable/24306018>>; Chen, Song Xi, and Ying-Li Qin.(2010) <[doi:10.1214/09-aos716](https://doi.org/10.1214/09-aos716)>; Srivastava, Muni S., and Meng Du.(2008) <[doi:10.1016/j.jmva.2006.11.002](https://doi.org/10.1016/j.jmva.2006.11.002)>; Srivastava, Muni S., Shota Katayama, and Yutaka Kano.(2013)<[doi:10.1016/j.jmva.2012.08.014](https://doi.org/10.1016/j.jmva.2012.08.014)>]. Normal-reference tests for two-sample problem [see Zhang, Jin-Ting, Jia Guo, Bu Zhou, and Ming-Yen Cheng.(2020) <[doi:10.1080/01621459.2019.1604366](https://doi.org/10.1080/01621459.2019.1604366)>; Zhang, Jin-Ting, Bu Zhou, Jia Guo, and Tianming Zhu.(2021) <[doi:10.1016/j.jspi.2020.11.008](https://doi.org/10.1016/j.jspi.2020.11.008)>; Zhang, Liang, Tianming Zhu, and Jin-Ting Zhang.(2020) <[doi:10.1016/j.ecosta.2019.12.002](https://doi.org/10.1016/j.ecosta.2019.12.002)>; Zhang, Liang, Tianming Zhu, and Jin-Ting Zhang.(2023) <[doi:10.1080/02664763.2020.1834516](https://doi.org/10.1080/02664763.2020.1834516)>; Zhang, Jin-Ting, and Tianming Zhu.(2022) <[doi:10.1080/10485252.2021.2015768](https://doi.org/10.1080/10485252.2021.2015768)>; Zhang, Jin-Ting, and Tianming Zhu.(2022) <[doi:10.1007/s42519-021-00232-w](https://doi.org/10.1007/s42519-021-00232-w)>; Zhu, Tianming, Pengfei Wang, and Jin-Ting Zhang.(2023) <[doi:10.1007/s00180-023-01433-6](https://doi.org/10.1007/s00180-023-01433-6)>]. Some existing tests for GLHT problem [see Fujikoshi, Yasunori, Tetsuto Himeno, and Hirofumi Wakaki.(2004) <[doi:10.14490/jjss.34.19](https://doi.org/10.14490/jjss.34.19)>; Srivastava, Muni S., and Yasunori Fujikoshi.(2006) <[doi:10.1016/j.jmva.2005.08.010](https://doi.org/10.1016/j.jmva.2005.08.010)>; Yamada, Takayuki, and Muni S. Srivastava.(2012) <[doi:10.1080/03610926.2011.581786](https://doi.org/10.1080/03610926.2011.581786)>; Schott, James R.(2007) <[doi:10.1016/j.jmva.2006.11.007](https://doi.org/10.1016/j.jmva.2006.11.007)>; Zhou, Bu, Ting Zhang.(2017) <[doi:10.1016/j.jspi.2017.03.005](https://doi.org/10.1016/j.jspi.2017.03.005)>]. Normal-reference tests for GLHT problem [see Zhang, Jin-Ting, Jia Guo, and Bu Zhou.(2017) <[doi:10.1016/j.jmva.2017.01.002](https://doi.org/10.1016/j.jmva.2017.01.002)>; Zhang, Jin-Ting, Bu Zhou, and Jia Guo.(2022) <[doi:10.1016/j.jmva.2021.104816](https://doi.org/10.1016/j.jmva.2021.104816)>; Zhu, Tian-

ming, Liang Zhang, and Jin-Ting Zhang.(2022) <doi:10.5705/ss.202020.0362>; Zhu, Tianming, and Jin-Ting Zhang.(2022) <doi:10.1007/s00180-021-01110-6>; Zhang, Jin-Ting, and Tianming Zhu.(2022) <doi:10.1016/j.csda.2021.107385>].

**License** GPL (>= 3)

**URL** <https://nie23wp8738.github.io/HDNRA/>

**BugReports** <https://github.com/nie23wp8738/HDNRA/issues>

**Encoding** UTF-8

**RoxygenNote** 7.3.1

**LinkingTo** Rcpp, RcppArmadillo

**Imports** expm, Rcpp, Rdpack, readr, stats

**Suggests** devtools, dplyr, knitr, rmarkdown, spelling, testthat (>= 3.0.0), tidyrr

**RdMacros** Rdpack

**Depends** R (>= 2.10)

**LazyData** true

**Language** en-US

**Config/testthat/edition** 3

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2024-02-27 17:40:02 UTC

## R topics documented:

corneal . . . . .	3
COVID19 . . . . .	4
glhtbf_zgz2017 . . . . .	5
glhtbf_zz2022 . . . . .	6
glhtbf_zzg2022 . . . . .	8
glht_fhw2004 . . . . .	10
glht_sf2006 . . . . .	11
glht_ys2012 . . . . .	13
glht_zgz2017 . . . . .	14
glht_zz2022 . . . . .	16
glht_zzz2022 . . . . .	18
ks_s2007 . . . . .	19
tsbf_cq2010 . . . . .	21
tsbf_skk2013 . . . . .	22
tsbf_zwz2023 . . . . .	24
tsbf_zz2022 . . . . .	25
tsbf_zzg2021 . . . . .	27
tsbf_zzz2023 . . . . .	29
ts_bs1996 . . . . .	30

<i>corneal</i>	3
ts_sd2008 . . . . .	32
ts_zgzc2020 . . . . .	33
ts_zz2022 . . . . .	35
ts_zzz2020 . . . . .	36
<b>Index</b>	<b>38</b>

corneal *HDNRA\_data corneal*

### Description

This dataset was acquired during a keratoconus study, a collaborative project involving Ms.Nancy Tripoli and Dr.Kenneth L.Cohen of Department of Ophthalmology at the University of North Carolina, Chapel Hill. The fitted feature vectors for the complete corneal surface dataset collectively into a feature matrix with dimensions of  $150 \times 2000$ .

### Usage

```
data(corneal)
```

### Format

#### 'corneal':

A data frame with 150 observations on the following 4 groups.

**normal group1** row 1 to row 43 in total 43 rows of the feature matrix correspond to observations from the normal group

**unilateral suspect group2** row 44 to row 57 in total 14 rows of the feature matrix correspond to observations from the unilateral suspect group

**suspect map group3** row 58 to row 78 in total 21 of the feature matrix correspond to observations from the suspect map group

**clinical keratoconus group4** row 79 to row 150 in total 72 of the feature matrix correspond to observations from the clinical keratoconus group

### References

Smaga Ł, Zhang J (2019). "Linear hypothesis testing with functional data." *Technometrics*, **61**(1), 99–110. doi:10.1080/00401706.2018.1456976.

### Examples

```
library(HDNRA)
data(corneal)
dim(corneal)
group1 <- as.matrix(corneal[1:43, ])
dim(group1)
group2 <- as.matrix(corneal[44:57, ])
dim(group2)
```

```
group3 <- as.matrix(corneal[58:78, ])  
dim(group3)  
group4 <- as.matrix(corneal[79:150, ])  
dim(group4)
```

---

COVID19

*HDNRA\_data COVID19*

---

### Description

A COVID19 data set from NCBI with ID GSE152641. The data set profiled peripheral blood from 24 healthy controls and 62 prospectively enrolled patients with community-acquired lower respiratory tract infection by SARS-COV-2 within the first 24 hours of hospital admission using RNA sequencing.

### Usage

```
data(COVID19)
```

### Format

**'COVID19':**

A data frame with 86 observations on the following 2 groups.

**healthy group1** row 2 to row 19, and row 82 to 87, in total 24 healthy controls

**patients group2** row 20 to 81, in total 62 prospectively enrolled patients

### References

Thair SA, He YD, Hasin-Brumshtein Y, Sakaram S, Pandya R, Toh J, Rawling D, Rimmel M, Coyle S, Dalekos GN, others (2021). "Transcriptomic similarities and differences in host response between SARS-CoV-2 and other viral infections." *Iscience*, **24**(1). doi:10.1016/j.isci.2020.101947.

### Examples

```
library(HDNRA)  
data(COVID19)  
dim(COVID19)  
group1 <- as.matrix(COVID19[c(2:19, 82:87), ])  
dim(group1)  
group2 <- as.matrix(COVID19[-c(1:19, 82:87), ])  
dim(group2)
```

glhtbf\_zgz2017

*Test proposed by Zhou et al. (2017)***Description**

Zhou et al. (2017)'s test for general linear hypothesis testing (GLHT) problem for high-dimensional data under heteroscedasticity.

**Usage**

```
glhtbf_zgz2017(Y,G,n,p)
```

**Arguments**

Y	A list of $k$ data matrices. The $i$ th element represents the data matrix ( $p \times n_i$ ) from the $i$ th population with each column representing a $p$ -dimensional observation.
G	A known full-rank coefficient matrix ( $q \times k$ ) with $\text{rank}(\mathbf{G}) < k$ .
n	A vector of $k$ sample sizes. The $i$ th element represents the sample size of group $i$ , $n_i$ .
p	The dimension of data.

**Details**

Suppose we have the following  $k$  independent high-dimensional samples:

$$\mathbf{y}_{i1}, \dots, \mathbf{y}_{in_i}, \text{ are i.i.d. with } E(\mathbf{y}_{i1}) = \boldsymbol{\mu}_i, \text{ Cov}(\mathbf{y}_{i1}) = \boldsymbol{\Sigma}_i, i = 1, \dots, k.$$

It is of interest to test the following GLHT problem:

$$H_0 : \mathbf{G}\mathbf{M} = \mathbf{0}, \quad \text{vs. } H_1 : \mathbf{G}\mathbf{M} \neq \mathbf{0},$$

where  $\mathbf{M} = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k)^\top$  is a  $k \times p$  matrix collecting  $k$  mean vectors and  $\mathbf{G} : q \times k$  is a known full-rank coefficient matrix with  $\text{rank}(\mathbf{G}) < k$ .

Let  $\bar{\mathbf{y}}_i, i = 1, \dots, k$  be the sample mean vectors and  $\hat{\boldsymbol{\Sigma}}_i, i = 1, \dots, k$  be the sample covariance matrices.

Zhou et al. (2017) proposed the following U-statistic based test statistic:

$$T_{ZGZ} = \|\mathbf{C}\hat{\boldsymbol{\mu}}\|^2 - \sum_{i=1}^k h_{ii} \text{tr}(\hat{\boldsymbol{\Sigma}}_i)/n_i,$$

where  $\mathbf{C} = [(\mathbf{G}\mathbf{D}\mathbf{G}^\top)^{-1/2}\mathbf{G}] \otimes \mathbf{I}_p$ ,  $\mathbf{D} = \text{diag}(1/n_1, \dots, 1/n_k)$ , and  $h_{ij}$  is the  $(i, j)$ th entry of the  $k \times k$  matrix  $\mathbf{H} = \mathbf{G}^\top (\mathbf{G}\mathbf{D}\mathbf{G}^\top)^{-1} \mathbf{G}$ .

They showed that under the null hypothesis,  $T_{ZGZ}$  is asymptotically normally distributed.

**Value**

A (list) object of S3 class `htest` containing the following elements:

**statistic** the test statistic proposed by Zhou et al. (2017).

**p.value** the  $p$ -value of the test proposed by Zhou et al. (2017).

**References**

Zhou B, Guo J, Zhang J (2017). “High-dimensional general linear hypothesis testing under heteroscedasticity.” *Journal of Statistical Planning and Inference*, **188**, 36–54. doi:10.1016/j.jspi.2017.03.005.

**Examples**

```
set.seed(1234)
k <- 3
p <- 50
n <- c(25, 30, 40)
rho <- 0.1
M <- matrix(rep(0, k * p), nrow = k, ncol = p)
avec <- seq(1, k)
Y <- list()
for (g in 1:k) {
  a <- avec[g]
  y <- (-2 * sqrt(a * (1 - rho)) + sqrt(4 * a * (1 - rho) + 4 * p * a * rho)) / (2 * p)
  x <- y + sqrt(a * (1 - rho))
  Gamma <- matrix(rep(y, p * p), nrow = p)
  diag(Gamma) <- rep(x, p)
  Z <- matrix(rnorm(n[g] * p, mean = 0, sd = 1), p, n[g])
  Y[[g]] <- Gamma %*% Z + t(t(M[g, ])) %*% (rep(1, n[g]))
}
G <- cbind(diag(k - 1), rep(-1, k - 1))
glhtbf_zgz2017(Y, G, n, p)
```

---

glhtbf\_zz2022

*Test proposed by Zhang and Zhu (2022)*


---

**Description**

Zhang and Zhu (2022)’s test for general linear hypothesis testing (GLHT) problem for high-dimensional data under heteroscedasticity.

**Usage**

```
glhtbf_zz2022(Y, G, n, p)
```

**Arguments**

Y	A list of $k$ data matrices. The $i$ th element represents the data matrix ( $p \times n_i$ ) from the $i$ th population with each column representing a $p$ -dimensional observation.
G	A known full-rank coefficient matrix ( $q \times k$ ) with $\text{rank}(\mathbf{G}) < k$ .
n	A vector of $k$ sample sizes. The $i$ th element represents the sample size of group $i$ , $n_i$ .
p	The dimension of data.

**Details**

Suppose we have the following  $k$  independent high-dimensional samples:

$$\mathbf{y}_{i1}, \dots, \mathbf{y}_{in_i}, \text{ are i.i.d. with } E(\mathbf{y}_{i1}) = \boldsymbol{\mu}_i, \text{ Cov}(\mathbf{y}_{i1}) = \boldsymbol{\Sigma}_i, i = 1, \dots, k.$$

It is of interest to test the following GLHT problem:

$$H_0 : \mathbf{G}\mathbf{M} = \mathbf{0}, \quad \text{vs. } H_1 : \mathbf{G}\mathbf{M} \neq \mathbf{0},$$

where  $\mathbf{M} = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k)^\top$  is a  $k \times p$  matrix collecting  $k$  mean vectors and  $\mathbf{G} : q \times k$  is a known full-rank coefficient matrix with  $\text{rank}(\mathbf{G}) < k$ .

Let  $\bar{\mathbf{y}}_i, i = 1, \dots, k$  be the sample mean vectors and  $\hat{\boldsymbol{\Sigma}}_i, i = 1, \dots, k$  be the sample covariance matrices.

Zhang and Zhu (2022) proposed the following U-statistic based test statistic:

$$T_{ZZ} = \|\mathbf{C}\hat{\boldsymbol{\mu}}\|^2 - \sum_{i=1}^k h_{ii} \text{tr}(\hat{\boldsymbol{\Sigma}}_i)/n_i,$$

where  $\mathbf{C} = [(\mathbf{G}\mathbf{D}\mathbf{G}^\top)^{-1/2}\mathbf{G}] \otimes \mathbf{I}_p$ ,  $\mathbf{D} = \text{diag}(1/n_1, \dots, 1/n_k)$ , and  $h_{ij}$  is the  $(i, j)$ th entry of the  $k \times k$  matrix  $\mathbf{H} = \mathbf{G}^\top (\mathbf{G}\mathbf{D}\mathbf{G}^\top)^{-1} \mathbf{G}$ .

**Value**

A (list) object of S3 class `htest` containing the following elements:

**p.value** the  $p$ -value of the test proposed by Zhang and Zhu (2022).

**statistic** the test statistic proposed by Zhang and Zhu (2022).

**beta0** the parameter used in Zhang and Zhu (2022)'s test.

**beta1** the parameter used in Zhang and Zhu (2022)'s test.

**df** estimated approximate degrees of freedom of Zhang and Zhu (2022)'s test.

**References**

Zhang J, Zhu T (2022). "A new normal reference test for linear hypothesis testing in high-dimensional heteroscedastic one-way MANOVA." *Computational Statistics & Data Analysis*, **168**, 107385. [doi:10.1016/j.csda.2021.107385](https://doi.org/10.1016/j.csda.2021.107385).

**Examples**

```

set.seed(1234)
k <- 3
p <- 50
n <- c(25, 30, 40)
rho <- 0.1
M <- matrix(rep(0, k * p), nrow = k, ncol = p)
avec <- seq(1, k)
Y <- list()
for (g in 1:k) {
  a <- avec[g]
  y <- (-2 * sqrt(a * (1 - rho)) + sqrt(4 * a * (1 - rho) + 4 * p * a * rho)) / (2 * p)
  x <- y + sqrt(a * (1 - rho))
  Gamma <- matrix(rep(y, p * p), nrow = p)
  diag(Gamma) <- rep(x, p)
  Z <- matrix(rnorm(n[g] * p, mean = 0, sd = 1), p, n[g])
  Y[[g]] <- Gamma %*% Z + t(t(M[g, ])) %*% (rep(1, n[g]))
}
G <- cbind(diag(k - 1), rep(-1, k - 1))
glhtbf_zz2022(Y, G, n, p)

```

---

glhtbf\_zzg2022

*Test proposed by Zhang et al. (2022)*


---

**Description**

Zhang et al. (2022)'s test for general linear hypothesis testing (GLHT) problem for high-dimensional data under heteroscedasticity.

**Usage**

```
glhtbf_zzg2022(Y, G, n, p)
```

**Arguments**

- |   |  |
|---|--|
| Y | A list of $k$ data matrices. The $i$ th element represents the data matrix ( $p \times n_i$ ) from the $i$ th population with each column representing a $p$ -dimensional observation. |
| G | A known full-rank coefficient matrix ( $q \times k$ ) with $\text{rank}(\mathbf{G}) < k$ .   |
| n | A vector of $k$ sample sizes. The $i$ th element represents the sample size of group $i$ , $n_i$ .   |
| p | The dimension of data.   |



## Details

Suppose we have the following  $k$  independent high-dimensional samples:

$$\mathbf{y}_{i1}, \dots, \mathbf{y}_{in_i}, \text{ are i. i. d. with } E(\mathbf{y}_{i1}) = \boldsymbol{\mu}_i, \text{ Cov}(\mathbf{y}_{i1}) = \boldsymbol{\Sigma}_i, i = 1, \dots, k.$$

It is of interest to test the following GLHT problem:

$$H_0 : \mathbf{GM} = \mathbf{0}, \quad \text{vs. } H_1 : \mathbf{GM} \neq \mathbf{0},$$

where  $\mathbf{M} = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k)^\top$  is a  $k \times p$  matrix collecting  $k$  mean vectors and  $\mathbf{G} : q \times k$  is a known full-rank coefficient matrix with  $\text{rank}(\mathbf{G}) < k$ .

Zhang et al. (2022) proposed the following test statistic:

$$T_{ZZG} = \|\mathbf{C}\hat{\boldsymbol{\mu}}\|^2,$$

where  $\mathbf{C} = [(\mathbf{GDG}^\top)^{-1/2}\mathbf{G}] \otimes \mathbf{I}_p$  with  $\mathbf{D} = \text{diag}(1/n_1, \dots, 1/n_k)$ , and  $\hat{\boldsymbol{\mu}} = (\bar{\mathbf{y}}_1^\top, \dots, \bar{\mathbf{y}}_k^\top)^\top$  with  $\bar{\mathbf{y}}_i, i = 1, \dots, k$  being the sample mean vectors.

They showed that under the null hypothesis,  $T_{ZZG}$  and a chi-squared-type mixture have the same normal or non-normal limiting distribution.

## Value

A (list) object of S3 class `htest` containing the following elements:

**p.value** the  $p$ -value of the test proposed by Zhang et al. (2022)

**statistic** the test statistic proposed by Zhang et al. (2022).

**beta** the parameters used in Zhang et al. (2022)'s test.

**df** estimated approximate degrees of freedom of Zhang et al. (2022)'s test.

## References

Zhang J, Zhou B, Guo J (2022). "Linear hypothesis testing in high-dimensional heteroscedastic one-way MANOVA: A normal reference  $L^2$ -norm based test." *Journal of Multivariate Analysis*, **187**, 104816. doi:10.1016/j.jmva.2021.104816.

## Examples

```
set.seed(1234)
k <- 3
p <- 50
n <- c(25, 30, 40)
rho <- 0.1
M <- matrix(rep(0, k * p), nrow = k, ncol = p)
avec <- seq(1, k)
Y <- list()
for (g in 1:k) {
  a <- avec[g]
  y <- (-2 * sqrt(a * (1 - rho)) + sqrt(4 * a * (1 - rho) + 4 * p * a * rho)) / (2 * p)
  x <- y + sqrt(a * (1 - rho))
  Gamma <- matrix(rep(y, p * p), nrow = p)
```

```

diag(Gamma) <- rep(x, p)
Z <- matrix(rnorm(n[g] * p, mean = 0, sd = 1), p, n[g])
Y[[g]] <- Gamma %*% Z + t(t(M[g, ])) %*% (rep(1, n[g]))
}
G <- cbind(diag(k - 1), rep(-1, k - 1))
glhtbf_zzg2022(Y, G, n, p)

```

---

glht\_fhw2004

*Test proposed by Fujikoshi et al. (2004)*


---

### Description

Fujikoshi et al. (2004)'s test for general linear hypothesis testing (GLHT) problem for high-dimensional data with assuming that underlying covariance matrices are the same.

### Usage

```
glht_fhw2004(Y, X, C)
```

### Arguments

Y	An $n \times p$ response matrix obtained by independently observing a $p$ -dimensional response variable for $n$ subjects.
X	A known $n \times k$ full-rank design matrix with $\text{rank}(\mathbf{G}) = k < n$ .
C	A known matrix of size $q \times k$ with $\text{rank}(\mathbf{C}) = q < k$ .

### Details

A high-dimensional linear regression model can be expressed as

$$\mathbf{Y} = \mathbf{X}\Theta + \epsilon,$$

where  $\Theta$  is a  $k \times p$  unknown parameter matrix and  $\epsilon$  is an  $n \times p$  error matrix.

It is of interest to test the following GLHT problem

$$H_0 : \mathbf{C}\Theta = \mathbf{0}, \quad \text{vs.} \quad H_1 : \mathbf{C}\Theta \neq \mathbf{0}.$$

Fujikoshi et al. (2004) proposed the following test statistic:

$$T_{FHW} = \sqrt{p} \left[ (n - k) \frac{\text{tr}(\mathbf{S}_h)}{\text{tr}(\mathbf{S}_e)} - q \right],$$

where  $\mathbf{S}_h$  and  $\mathbf{S}_e$  are the matrices of sums of squares and products due to the hypothesis and the error, respectively.

They showed that under the null hypothesis,  $T_{FHW}$  is asymptotically normally distributed.

**Value**

A (list) object of S3 class `htest` containing the following elements:

**statistic** the test statistic proposed by Fujikoshi et al. (2004).

**p.value** the  $p$ -value of the test proposed by Fujikoshi et al. (2004).

**References**

Fujikoshi Y, Himeno T, Wakaki H (2004). "Asymptotic results of a high dimensional MANOVA test and power comparison when the dimension is large compared to the sample size." *Journal of the Japan Statistical Society*, **34**(1), 19–26. doi:10.14490/jjss.34.19.

**Examples**

```
set.seed(1234)
k <- 3
q <- k-1
p <- 50
n <- c(25,30,40)
rho <- 0.01
Theta <- matrix(rep(0,k*p),nrow=k)
X <- matrix(c(rep(1,n[1]),rep(0,sum(n)),rep(1,n[2]),rep(0,sum(n)),rep(1,n[3])),ncol=k,nrow=sum(n))
y <- (-2*sqrt(1-rho)+sqrt(4*(1-rho)+4*p*rho))/(2*p)
x <- y+sqrt((1-rho))
Gamma <- matrix(rep(y,p*p),nrow=p)
diag(Gamma) <- rep(x,p)
U <- matrix(ncol = sum(n),nrow=p)
for(i in 1:sum(n)){
  U[,i] <- rnorm(p,0,1)
}
Y <- X%%Theta+t(U)%%Gamma
C <- cbind(diag(q),-rep(1,q))
glht_fhw2004(Y,X,C)
```

---

glht\_sf2006

*Test proposed by Srivastava and Fujikoshi (2006)*


---

**Description**

Srivastava and Fujikoshi (2006)'s test for general linear hypothesis testing (GLHT) problem for high-dimensional data with assuming that underlying covariance matrices are the same.

**Usage**

```
glht_sf2006(Y,X,C)
```

**Arguments**

<b>Y</b>	An $n \times p$ response matrix obtained by independently observing a $p$ -dimensional response variable for $n$ subjects.
<b>X</b>	A known $n \times k$ full-rank design matrix with $\text{rank}(\mathbf{G}) = k < n$ .
<b>C</b>	A known matrix of size $q \times k$ with $\text{rank}(\mathbf{C}) = q < k$ .

**Details**

A high-dimensional linear regression model can be expressed as

$$\mathbf{Y} = \mathbf{X}\Theta + \epsilon,$$

where  $\Theta$  is a  $k \times p$  unknown parameter matrix and  $\epsilon$  is an  $n \times p$  error matrix.

It is of interest to test the following GLHT problem

$$H_0 : \mathbf{C}\Theta = \mathbf{0}, \quad \text{vs.} \quad H_1 : \mathbf{C}\Theta \neq \mathbf{0}.$$

Srivastava and Fujikoshi (2006) proposed the following test statistic:

$$T_{SF} = [2q\hat{a}_2(1 + (n - k)^{-1}q)]^{-1/2} \left[ \frac{\text{tr}(\mathbf{B})}{\sqrt{p}} - \frac{q}{\sqrt{n - k}} \frac{\text{tr}(\mathbf{W})}{\sqrt{(n - k)p}} \right].$$

where  $\mathbf{W}$  and  $\mathbf{B}$  are the matrix of sum of squares and products due to error and the error, respectively, and  $\hat{a}_2 = [\text{tr}(\mathbf{W}^2) - \text{tr}^2(\mathbf{W})/(n - k)]/[(n - k - 1)(n - k + 2)p]$ . They showed that under the null hypothesis,  $T_{SF}$  is asymptotically normally distributed.

**Value**

A (list) object of S3 class `htest` containing the following elements:

**statistic** the test statistic proposed by Srivastava and Fujikoshi (2006).

**p.value** the  $p$ -value of the test proposed by Srivastava and Fujikoshi (2006).

**References**

Srivastava MS, Fujikoshi Y (2006). "Multivariate analysis of variance with fewer observations than the dimension." *Journal of Multivariate Analysis*, **97**(9), 1927–1940. doi:10.1016/j.jmva.2005.08.010.

**Examples**

```
set.seed(1234)
k <- 3
q <- k-1
p <- 50
n <- c(25, 30, 40)
rho <- 0.01
Theta <- matrix(rep(0, k*p), nrow=k)
X <- matrix(c(rep(1, n[1]), rep(0, sum(n)), rep(1, n[2]), rep(0, sum(n)), rep(1, n[3])), ncol=k, nrow=sum(n))
y <- (-2*sqrt(1-rho)+sqrt(4*(1-rho)+4*p*rho))/(2*p)
```

```

x <- y+sqrt((1-rho))
Gamma <- matrix(rep(y,p*p),nrow=p)
diag(Gamma) <- rep(x,p)
U <- matrix(ncol = sum(n),nrow=p)
for(i in 1:sum(n)){
U[,i] <- rnorm(p,0,1)
}
Y <- X%%Theta+t(U)%%Gamma
C <- cbind(diag(q),-rep(1,q))
glht_sf2006(Y,X,C)

```

glht\_ys2012

*Test proposed by Yamada and Srivastava (2012)***Description**

Yamada and Srivastava (2012)'test for general linear hypothesis testing (GLHT) problem for high-dimensional data with assuming that underlying covariance matrices are the same.

**Usage**

```
glht_ys2012(Y,X,C)
```

**Arguments**

Y	An $n \times p$ response matrix obtained by independently observing a $p$ -dimensional response variable for $n$ subjects.
X	A known $n \times k$ full-rank design matrix with $\text{rank}(\mathbf{G}) = k < n$ .
C	A known matrix of size $q \times k$ with $\text{rank}(\mathbf{C}) = q < k$ .

**Details**

A high-dimensional linear regression model can be expressed as

$$\mathbf{Y} = \mathbf{X}\Theta + \epsilon,$$

where  $\Theta$  is a  $k \times p$  unknown parameter matrix and  $\epsilon$  is an  $n \times p$  error matrix.

It is of interest to test the following GLHT problem

$$H_0 : \mathbf{C}\Theta = \mathbf{0}, \quad \text{vs. } H_1 : \mathbf{C}\Theta \neq \mathbf{0}.$$

Yamada and Srivastava (2012) proposed the following test statistic:

$$T_{YS} = \frac{(n-k) \text{tr}(\mathbf{S}_h \mathbf{D}_{\mathbf{S}_e}^{-1}) - (n-k)pq/(n-k-2)}{\sqrt{2q[\text{tr}(\mathbf{R}^2) - p^2/(n-k)]c_{p,n}}},$$

where  $\mathbf{S}_h$  and  $\mathbf{S}_e$  are the variation matrices due to the hypothesis and error, respectively, and  $\mathbf{D}_{\mathbf{S}_e}$  and  $\mathbf{R}$  are diagonal matrix with the diagonal elements of  $\mathbf{S}_e$  and the sample correlation matrix, respectively.  $c_{p,n}$  is the adjustment coefficient proposed by Yamada and Srivastava (2012). They showed that under the null hypothesis,  $T_{YS}$  is asymptotically normally distributed.

**Value**

A (list) object of S3 class `htest` containing the following elements:

**statistic** the test statistic proposed by Yamada and Srivastava (2012).

**p.value** the  $p$ -value of the test proposed by Yamada and Srivastava (2012).

**References**

Yamada T, Srivastava MS (2012). “A test for multivariate analysis of variance in high dimension.” *Communications in Statistics-Theory and Methods*, **41**(13-14), 2602–2615. doi:[10.1080/03610926.2011.581786](https://doi.org/10.1080/03610926.2011.581786).

**Examples**

```
set.seed(1234)
k <- 3
q <- k-1
p <- 50
n <- c(25,30,40)
rho <- 0.01
Theta <- matrix(rep(0,k*p),nrow=k)
X <- matrix(c(rep(1,n[1]),rep(0,sum(n)),rep(1,n[2]),rep(0,sum(n)),rep(1,n[3])),ncol=k,nrow=sum(n))
y <- (-2*sqrt(1-rho)+sqrt(4*(1-rho)+4*p*rho))/(2*p)
x <- y+sqrt((1-rho))
Gamma <- matrix(rep(y,p*p),nrow=p)
diag(Gamma) <- rep(x,p)
U <- matrix(ncol = sum(n),nrow=p)
for(i in 1:sum(n)){
  U[,i] <- rnorm(p,0,1)
}
Y <- X%%Theta+t(U)%%Gamma
C <- cbind(diag(q),-rep(1,q))
glht_ys2012(Y,X,C)
```

---

glht\_zgz2017

*Test proposed by Zhang et al. (2017)*


---

**Description**

Zhang et al. (2017)’s test for general linear hypothesis testing (GLHT) problem for high-dimensional data with assuming that underlying covariance matrices are the same.

**Usage**

```
glht_zgz2017(Y,G,n,p)
```

### Arguments

Y	A list of $k$ data matrices. The $i$ th element represents the data matrix ( $p \times n_i$ ) from the $i$ th population with each column representing a $p$ -dimensional observation.
G	A known full-rank coefficient matrix ( $q \times k$ ) with $\text{rank}(\mathbf{G}) < k$ .
n	A vector of $k$ sample sizes. The $i$ th element represents the sample size of group $i$ , $n_i$ .
p	The dimension of data.

### Details

Suppose we have the following  $k$  independent high-dimensional samples:

$$\mathbf{y}_{i1}, \dots, \mathbf{y}_{in_i}, \text{ are i. i. d. with } E(\mathbf{y}_{i1}) = \boldsymbol{\mu}_i, \text{ Cov}(\mathbf{y}_{i1}) = \boldsymbol{\Sigma}, i = 1, \dots, k.$$

It is of interest to test the following GLHT problem:

$$H_0 : \mathbf{GM} = \mathbf{0}, \quad \text{vs.} \quad H_1 : \mathbf{GM} \neq \mathbf{0},$$

where  $\mathbf{M} = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k)^\top$  is a  $k \times p$  matrix collecting  $k$  mean vectors and  $\mathbf{G} : q \times k$  is a known full-rank coefficient matrix with  $\text{rank}(\mathbf{G}) < k$ .

Zhang et al. (2017) proposed the following test statistic:

$$T_{ZGZ} = \|\mathbf{C}\hat{\boldsymbol{\mu}}\|^2,$$

where  $\mathbf{C} = [(\mathbf{GDG}^\top)^{-1/2}\mathbf{G}] \otimes \mathbf{I}_p$ , and  $\hat{\boldsymbol{\mu}} = (\bar{\mathbf{y}}_1^\top, \dots, \bar{\mathbf{y}}_k^\top)^\top$ , with  $\bar{\mathbf{y}}_i, i = 1, \dots, k$  being the sample mean vectors and  $\mathbf{D} = \text{diag}(1/n_1, \dots, 1/n_k)$ .

They showed that under the null hypothesis,  $T_{ZGZ}$  and a chi-squared-type mixture have the same normal or non-normal limiting distribution.

### Value

A (list) object of S3 class `htest` containing the following elements:

**statistic** the test statistic proposed by Zhang et al. (2017)

**p.value** the  $p$ -value of the test proposed by Zhang et al. (2017).

**beta** the parameters used in Zhang et al. (2017)'s test.

**df** estimated approximate degrees of freedom of Zhang et al.(2017)'s test.

### References

Zhang J, Guo J, Zhou B (2017). "Linear hypothesis testing in high-dimensional one-way MANOVA." *Journal of Multivariate Analysis*, **155**, 200–216. doi:10.1016/j.jmva.2017.01.002.

**Examples**

```

set.seed(1234)
k <- 3
p <- 50
n <- c(25, 30, 40)
rho <- 0.1
M <- matrix(rep(0, k * p), nrow = k, ncol = p)
y <- (-2 * sqrt(1 - rho) + sqrt(4 * (1 - rho) + 4 * p * rho)) / (2 * p)
x <- y + sqrt((1 - rho))
Gamma <- matrix(rep(y, p * p), nrow = p)
diag(Gamma) <- rep(x, p)
Y <- list()
for (g in 1:k) {
  Z <- matrix(rnorm(n[g] * p, mean = 0, sd = 1), p, n[g])
  Y[[g]] <- Gamma %*% Z + t(t(M[g, ])) %*% (rep(1, n[g]))
}
G <- cbind(diag(k - 1), rep(-1, k - 1))
glht_zgz2017(Y, G, n, p)

```

glht\_zz2022

*Test proposed by Zhu and Zhang (2022)***Description**

Zhu and Zhang (2022)'s test for general linear hypothesis testing (GLHT) problem for high-dimensional data with assuming that underlying covariance matrices are the same.

**Usage**

```
glht_zz2022(Y, G, n, p)
```

**Arguments**

Y	A list of $k$ data matrices. The $i$ th element represents the data matrix ( $p \times n_i$ ) from the $i$ th population with each column representing a $p$ -dimensional observation.
G	A known full-rank coefficient matrix ( $q \times k$ ) with $\text{rank}(\mathbf{G}) < k$ .
n	A vector of $k$ sample sizes. The $i$ th element represents the sample size of group $i$ , $n_i$ .
p	The dimension of data.

**Details**

Suppose we have the following  $k$  independent high-dimensional samples:

$$\mathbf{y}_{i1}, \dots, \mathbf{y}_{in_i}, \text{ are i. i. d. with } E(\mathbf{y}_{i1}) = \boldsymbol{\mu}_i, \text{ Cov}(\mathbf{y}_{i1}) = \boldsymbol{\Sigma}, i = 1, \dots, k.$$

It is of interest to test the following GLHT problem:

$$H_0 : \mathbf{GM} = \mathbf{0}, \quad \text{vs.} \quad H_1 : \mathbf{GM} \neq \mathbf{0},$$



where  $M = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k)^\top$  is a  $k \times p$  matrix collecting  $k$  mean vectors and  $G : q \times k$  is a known full-rank coefficient matrix with  $\text{rank}(G) < k$ .

Zhu and Zhang (2022) proposed the following test statistic:

$$T_{ZZ} = \|C\hat{\boldsymbol{\mu}}\|^2 - q \text{tr}(\hat{\boldsymbol{\Sigma}}),$$

where  $C = [(GDG^\top)^{-1/2}G] \otimes I_p$ , and  $\hat{\boldsymbol{\mu}} = (\bar{\mathbf{y}}_1^\top, \dots, \bar{\mathbf{y}}_k^\top)^\top$ , with  $\bar{\mathbf{y}}_i, i = 1, \dots, k$  being the sample mean vectors and  $\hat{\boldsymbol{\Sigma}}$  being the usual pooled sample covariance matrix of the  $k$  samples.

They showed that under the null hypothesis,  $T_{ZZ}$  and a chi-squared-type mixture have the same normal or non-normal limiting distribution.

## Value

A (list) object of S3 class `htest` containing the following elements:

**p.value** the  $p$ -value of the test proposed by Zhu and Zhang (2022).

**statistic** the test statistic proposed by Zhu and Zhang (2022).

**beta0** the parameter used in Zhu and Zhang (2022)'s test.

**beta1** the parameter used in Zhu and Zhang (2022)'s test.

**df** estimated approximate degrees of freedom of Zhu and Zhang (2022)'s test.

## References

Zhu T, Zhang J (2022). "Linear hypothesis testing in high-dimensional one-way MANOVA: a new normal reference approach." *Computational Statistics*, **37**(1), 1–27. doi:10.1007/s0018002101110-6.

## Examples

```
set.seed(1234)
k <- 3
p <- 50
n <- c(25, 30, 40)
rho <- 0.1
M <- matrix(rep(0, k * p), nrow = k, ncol = p)
y <- (-2 * sqrt(1 - rho) + sqrt(4 * (1 - rho) + 4 * p * rho)) / (2 * p)
x <- y + sqrt((1 - rho))
Gamma <- matrix(rep(y, p * p), nrow = p)
diag(Gamma) <- rep(x, p)
Y <- list()
for (g in 1:k) {
  Z <- matrix(rnorm(n[g] * p, mean = 0, sd = 1), p, n[g])
  Y[[g]] <- Gamma %*% Z + t(t(M[g, ])) %*% (rep(1, n[g]))
}
G <- cbind(diag(k - 1), rep(-1, k - 1))
glht_zz2022(Y, G, n, p)
```

glht\_zzz2022

*Test proposed by Zhu et al. (2022)***Description**

Zhu et al. (2022)'s test for general linear hypothesis testing (GLHT) problem for high-dimensional data with assuming that underlying covariance matrices are the same.

**Usage**

```
glht_zzz2022(Y, X, C)
```

**Arguments**

<b>Y</b>	An $n \times p$ response matrix obtained by independently observing a $p$ -dimensional response variable for $n$ subjects.
<b>X</b>	A known $n \times k$ full-rank design matrix with $\text{rank}(\mathbf{G}) = k < n - 2$ .
<b>C</b>	A known matrix of size $q \times k$ with $\text{rank}(\mathbf{C}) = q < k$ .

**Details**

A high-dimensional linear regression model can be expressed as

$$\mathbf{Y} = \mathbf{X}\Theta + \epsilon,$$

where  $\Theta$  is a  $k \times p$  unknown parameter matrix and  $\epsilon$  is an  $n \times p$  error matrix.

It is of interest to test the following GLHT problem

$$H_0 : \mathbf{C}\Theta = \mathbf{0}, \quad \text{vs. } H_1 : \mathbf{C}\Theta \neq \mathbf{0}.$$

Zhu et al. (2022) proposed the following test statistic:

$$T_{ZZZ} = \frac{(n - k - 2)}{(n - k)pq} \text{tr}(\mathbf{S}_h \mathbf{D}^{-1}),$$

where  $\mathbf{S}_h$  and  $\mathbf{S}_e$  are the variation matrices due to the hypothesis and error, respectively, and  $\mathbf{D}$  is the diagonal matrix with the diagonal elements of  $\mathbf{S}_e / (n - k)$ . They showed that under the null hypothesis,  $T_{ZZZ}$  and a chi-squared-type mixture have the same limiting distribution.

**Value**

A (list) object of S3 class `htest` containing the following elements:

**p.value** the  $p$ -value of the test proposed by Zhu et al. (2022)

**statistic** the test statistic proposed by Zhu et al. (2022).

**df** estimated approximate degrees of freedom of Zhu et al. (2022)'s test.

## References

Zhu T, Zhang L, Zhang J (2023). “Hypothesis Testing in High-Dimensional Linear Regression: A Normal Reference Scale-Invariant Test.” *Statistica Sinica*. doi:10.5705/ss.202020.0362.

## Examples

```
set.seed(1234)
k <- 3
q <- k - 1
p <- 50
n <- c(25, 30, 40)
rho <- 0.01
Theta <- matrix(rep(0, k * p), nrow = k)
X <- matrix(c(rep(1, n[1]), rep(0, sum(n)), rep(1, n[2]), rep(0, sum(n)), rep(1, n[3])),
  ncol = k, nrow = sum(n)
)
y <- (-2 * sqrt(1 - rho) + sqrt(4 * (1 - rho) + 4 * p * rho)) / (2 * p)
x <- y + sqrt((1 - rho))
Gamma <- matrix(rep(y, p * p), nrow = p)
diag(Gamma) <- rep(x, p)
U <- matrix(ncol = sum(n), nrow = p)
for (i in 1:sum(n)) {
  U[, i] <- rnorm(p, 0, 1)
}
Y <- X %*% Theta + t(U) %*% Gamma
C <- cbind(diag(q), -rep(1, q))
glht_zzz2022(Y, X, C)
```

---

ks\_s2007

*Test proposed by Schott (2007)*


---

## Description

Schott, J. R. (2007)’s test for one-way MANOVA problem for high-dimensional data with assuming that underlying covariance matrices are the same.

## Usage

```
ks_s2007(Y, n, p)
```

## Arguments

Y	A list of $k$ data matrices. The $i$ th element represents the data matrix ( $p \times n_i$ ) from the $i$ th population with each column representing a $p$ -dimensional observation.
n	A vector of $k$ sample sizes. The $i$ th element represents the sample size of group $i$ , $n_i$ .
p	The dimension of data.

## Details

Suppose we have the following  $k$  independent high-dimensional samples:

$$\mathbf{y}_{i1}, \dots, \mathbf{y}_{in_i}, \text{ are i.i.d. with } E(\mathbf{y}_{i1}) = \boldsymbol{\mu}_i, \text{ Cov}(\mathbf{y}_{i1}) = \boldsymbol{\Sigma}, i = 1, \dots, k.$$

It is of interest to test the following one-way MANOVA problem:

$$H_0 : \boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_k, \quad \text{vs. } H_1 : H_0 \text{ is not true.}$$

Schott (2007) proposed the following test statistic:

$$T_S = [\text{tr}(\mathbf{H})/h - \text{tr}(\mathbf{E})/e]/\sqrt{N-1},$$

where  $\mathbf{H} = \sum_{i=1}^k n_i (\bar{\mathbf{y}}_i - \bar{\mathbf{y}})(\bar{\mathbf{y}}_i - \bar{\mathbf{y}})^\top$ ,  $\mathbf{E} = \sum_{i=1}^k \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)^\top$ ,  $h = k - 1$ , and  $e = N - k$ , with  $N = n_1 + \dots + n_k$ . They showed that under the null hypothesis,  $T_S$  is asymptotically normally distributed.

## Value

A (list) object of S3 class `htest` containing the following elements:

**statistic** the test statistic proposed by Schott (2007).

**p.value** the  $p$ -value of the test proposed by Schott (2007).

## References

Schott JR (2007). "Some high-dimensional tests for a one-way MANOVA." *Journal of Multivariate Analysis*, **98**(9), 1825–1839. doi:10.1016/j.jmva.2006.11.007.

## Examples

```
set.seed(1234)
k <- 3
p <- 50
n <- c(25, 30, 40)
rho <- 0.1
M <- matrix(rep(0, k * p), nrow = k, ncol = p)
y <- (-2 * sqrt(1 - rho) + sqrt(4 * (1 - rho) + 4 * p * rho)) / (2 * p)
x <- y + sqrt((1 - rho))
Gamma <- matrix(rep(y, p * p), nrow = p)
diag(Gamma) <- rep(x, p)
Y <- list()
for (g in 1:k) {
  Z <- matrix(rnorm(n[g] * p, mean = 0, sd = 1), p, n[g])
  Y[[g]] <- Gamma %*% Z + t(t(M[g, ])) %*% (rep(1, n[g]))
}
ks_s2007(Y, n, p)
```

tsbf\_cq2010

*Test proposed by Chen and Qin (2010)***Description**

Chen and Qin (2010)'s test for testing equality of two-sample high-dimensional mean vectors without assuming that two covariance matrices are the same.

**Usage**

```
tsbf_cq2010(y1, y2)
```

**Arguments**

**y1** The data matrix (p by n1) from the first population. Each column represents a p-dimensional observation.

**y2** The data matrix (p by n2) from the first population. Each column represents a p-dimensional observation.

**Details**

Suppose we have two independent high-dimensional samples:

$$\mathbf{y}_{i1}, \dots, \mathbf{y}_{in_i}, \text{ are i. i. d. with } E(\mathbf{y}_{i1}) = \boldsymbol{\mu}_i, \text{ Cov}(\mathbf{y}_{i1}) = \boldsymbol{\Sigma}_i, i = 1, 2.$$

The primary object is to test

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ versus } H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2.$$

Chen and Qin (2010) proposed the following test statistic:

$$T_{CQ} = \frac{\sum_{i \neq j}^{n_1} \mathbf{y}_{1i}^\top \mathbf{y}_{1j}}{n_1(n_1 - 1)} + \frac{\sum_{i \neq j}^{n_2} \mathbf{y}_{2i}^\top \mathbf{y}_{2j}}{n_2(n_2 - 1)} - 2 \frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \mathbf{y}_{1i}^\top \mathbf{y}_{2j}}{n_1 n_2}.$$

They showed that under the null hypothesis,  $T_{CQ}$  is asymptotically normally distributed.

**Value**

A (list) object of S3 class `htest` containing the following elements:

**statistic** the test statistic proposed by Chen and Qin (2010)

**p.value** the p-value of the test proposed by Chen and Qin (2010).

**References**

Chen SX, Qin Y (2010). "A two-sample test for high-dimensional data with applications to gene-set testing." *The Annals of Statistics*, **38**(2). doi:10.1214/09a0s716.

**Examples**

```

set.seed(1234)
n1 <- 20
n2 <- 30
p <- 50
mu1 <- t(t(rep(0, p)))
mu2 <- mu1
rho1 <- 0.1
rho2 <- 0.2
a1 <- 1
a2 <- 2
w1 <- (-2 * sqrt(a1 * (1 - rho1)) + sqrt(4 * a1 * (1 - rho1) + 4 * p * a1 * rho1)) / (2 * p)
x1 <- w1 + sqrt(a1 * (1 - rho1))
Gamma1 <- matrix(rep(w1, p * p), nrow = p)
diag(Gamma1) <- rep(x1, p)
w2 <- (-2 * sqrt(a2 * (1 - rho2)) + sqrt(4 * a2 * (1 - rho2) + 4 * p * a2 * rho2)) / (2 * p)
x2 <- w2 + sqrt(a2 * (1 - rho2))
Gamma2 <- matrix(rep(w2, p * p), nrow = p)
diag(Gamma2) <- rep(x2, p)
Z1 <- matrix(rnorm(n1*p,mean = 0,sd = 1), p, n1)
Z2 <- matrix(rnorm(n2*p,mean = 0,sd = 1), p, n2)
y1 <- Gamma1 %*% Z1 + mu1%*(rep(1,n1))
y2 <- Gamma2 %*% Z2 + mu2%*(rep(1,n2))
tsbf_cq2010(y1, y2)

```

---

tsbf\_skk2013

*Test proposed by Srivastava et al. (2013)*


---

**Description**

Srivastava et al. (2013)'s test for testing equality of two-sample high-dimensional mean vectors without assuming that two covariance matrices are the same.

**Usage**

```
tsbf_skk2013(y1, y2)
```

**Arguments**

y1	The data matrix (p by n1) from the first population. Each column represents a p-dimensional observation.
y2	The data matrix (p by n2) from the first population. Each column represents a p-dimensional observation.

## Details

Suppose we have two independent high-dimensional samples:

$$\mathbf{y}_{i1}, \dots, \mathbf{y}_{in_i}, \text{ are i. i. d. with } E(\mathbf{y}_{i1}) = \boldsymbol{\mu}_i, \text{ Cov}(\mathbf{y}_{i1}) = \boldsymbol{\Sigma}_i, i = 1, 2.$$

The primary object is to test

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ versus } H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2.$$

Srivastava et al. (2013) proposed the following test statistic:

$$T_{SKK} = \frac{(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^\top \hat{\mathbf{D}}^{-1} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2) - p}{\sqrt{2\widehat{\text{Var}}(\hat{q}_n)c_{p,n}}},$$

where  $\bar{\mathbf{y}}_i, i = 1, 2$  are the sample mean vectors,  $\hat{\mathbf{D}} = \hat{\mathbf{D}}_1/n_1 + \hat{\mathbf{D}}_2/n_2$  with  $\hat{\mathbf{D}}_i, i = 1, 2$  being the diagonal matrices consisting of only the diagonal elements of the sample covariance matrices.  $\widehat{\text{Var}}(\hat{q}_n)$  is given by equation (1.18) in Srivastava et al. (2013), and  $c_{p,n}$  is the adjustment coefficient proposed by Srivastava et al. (2013). They showed that under the null hypothesis,  $T_{SKK}$  is asymptotically normally distributed.

## Value

A (list) object of S3 class `htest` containing the following elements:

**statistic** the test statistic proposed by Srivastava et al. (2013)

**p.value** the  $p$ -value of the test proposed by Srivastava et al. (2013)

**cpn** the adjustment coefficient proposed by Srivastava et al. (2013)

## References

Srivastava MS, Katayama S, Kano Y (2013). "A two sample test in high dimensional data." *Journal of Multivariate Analysis*, **114**, 349–358. doi:10.1016/j.jmva.2012.08.014.

## Examples

```
set.seed(1234)
n1 <- 20
n2 <- 30
p <- 50
mu1 <- t(rep(0, p))
mu2 <- mu1
rho1 <- 0.1
rho2 <- 0.2
a1 <- 1
a2 <- 2
w1 <- (-2 * sqrt(a1 * (1 - rho1)) + sqrt(4 * a1 * (1 - rho1) + 4 * p * a1 * rho1)) / (2 * p)
x1 <- w1 + sqrt(a1 * (1 - rho1))
Gamma1 <- matrix(rep(w1, p * p), nrow = p)
diag(Gamma1) <- rep(x1, p)
w2 <- (-2 * sqrt(a2 * (1 - rho2)) + sqrt(4 * a2 * (1 - rho2) + 4 * p * a2 * rho2)) / (2 * p)
```

```

x2 <- w2 + sqrt(a2 * (1 - rho2))
Gamma2 <- matrix(rep(w2, p * p), nrow = p)
diag(Gamma2) <- rep(x2, p)
Z1 <- matrix(rnorm(n1*p,mean = 0,sd = 1), p, n1)
Z2 <- matrix(rnorm(n2*p,mean = 0,sd = 1), p, n2)
y1 <- Gamma1 %*% Z1 + mu1%*(rep(1,n1))
y2 <- Gamma2 %*% Z2 + mu2%*(rep(1,n2))
tsbf_skk2013(y1, y2)

```

---

tsbf\_zwz2023

*Test proposed by Zhu et al. (2023)*


---

### Description

Zhu et al. (2023)'s test for testing equality of two-sample high-dimensional mean vectors without assuming that two covariance matrices are the same.

### Usage

```
tsbf_zwz2023(y1, y2)
```

### Arguments

- |    |   |
|----|---|
| y1 | The data matrix (p by n1) from the first population. Each column represents a $p$ -dimensional observation. |
| y2 | The data matrix (p by n2) from the first population. Each column represents a $p$ -dimensional observation. |

### Details

Suppose we have two independent high-dimensional samples:

$$\mathbf{y}_{i1}, \dots, \mathbf{y}_{in_i}, \text{ are i. i. d. with } E(\mathbf{y}_{i1}) = \boldsymbol{\mu}_i, \text{ Cov}(\mathbf{y}_{i1}) = \boldsymbol{\Sigma}_i, i = 1, 2.$$

The primary object is to test

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ versus } H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2.$$

Zhu et al. (2023) proposed the following test statistic:

$$T_{ZWZ} = \frac{n_1 n_2 n^{-1} \|\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2\|^2}{\text{tr}(\hat{\boldsymbol{\Omega}}_n)},$$

where  $\bar{\mathbf{y}}_i, i = 1, 2$  are the sample mean vectors and  $\hat{\boldsymbol{\Omega}}_n$  is the estimator of  $\text{Cov}[(n_1 n_2 / n)^{1/2} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)]$ . They showed that under the null hypothesis,  $T_{ZWZ}$  and an F-type mixture have the same normal or non-normal limiting distribution.



**Value**

A (list) object of S3 class `htest` containing the following elements:

**p.value** the p-value of the test proposed by Zhu et al. (2023).

**statistic** the test statistic proposed by Zhu et al. (2023).

**df1** estimated approximate degrees of freedom  $d_1$  of Zhu et al. (2023)'s test.

**df2** estimated approximate degrees of freedom  $d_2$  of Zhu et al. (2023)'s test.

**References**

Zhu T, Wang P, Zhang J (2023). “Two-sample Behrens–Fisher problems for high-dimensional data: a normal reference F-type test.” *Computational Statistics*, 1–24. doi:10.1007/s00180023014336.

**Examples**

```
set.seed(1234)
n1 <- 20
n2 <- 30
p <- 50
mu1 <- t(t(rep(0, p)))
mu2 <- mu1
rho1 <- 0.1
rho2 <- 0.2
a1 <- 1
a2 <- 2
w1 <- (-2 * sqrt(a1 * (1 - rho1)) + sqrt(4 * a1 * (1 - rho1) + 4 * p * a1 * rho1)) / (2 * p)
x1 <- w1 + sqrt(a1 * (1 - rho1))
Gamma1 <- matrix(rep(w1, p * p), nrow = p)
diag(Gamma1) <- rep(x1, p)
w2 <- (-2 * sqrt(a2 * (1 - rho2)) + sqrt(4 * a2 * (1 - rho2) + 4 * p * a2 * rho2)) / (2 * p)
x2 <- w2 + sqrt(a2 * (1 - rho2))
Gamma2 <- matrix(rep(w2, p * p), nrow = p)
diag(Gamma2) <- rep(x2, p)
Z1 <- matrix(rnorm(n1*p, mean = 0, sd = 1), p, n1)
Z2 <- matrix(rnorm(n2*p, mean = 0, sd = 1), p, n2)
y1 <- Gamma1 %%% Z1 + mu1%%(rep(1, n1))
y2 <- Gamma2 %%% Z2 + mu2%%(rep(1, n2))
tsbf_zwz2023(y1, y2)
```

**Description**

Zhang and Zhu (2022)'s test for testing equality of two-sample high-dimensional mean vectors without assuming that two covariance matrices are the same.

**Usage**

```
tsbf_zz2022(y1, y2)
```

**Arguments**

- y1** The data matrix (p by n1) from the first population. Each column represents a p-dimensional observation.
- y2** The data matrix (p by n2) from the first population. Each column represents a p-dimensional observation.

**Details**

Suppose we have two independent high-dimensional samples:

$$\mathbf{y}_{i1}, \dots, \mathbf{y}_{in_i}, \text{ are i. i. d. with } E(\mathbf{y}_{i1}) = \boldsymbol{\mu}_i, \text{ Cov}(\mathbf{y}_{i1}) = \boldsymbol{\Sigma}_i, i = 1, 2.$$

The primary object is to test

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ versus } H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2.$$

Zhang and Zhu (2022) proposed the following test statistic:

$$T_{ZZ} = \|\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2\|^2 - \text{tr}(\hat{\boldsymbol{\Omega}}_n),$$

where  $\bar{\mathbf{y}}_i, i = 1, 2$  are the sample mean vectors and  $\hat{\boldsymbol{\Omega}}_n$  is the estimator of  $\text{Cov}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)$ . They showed that under the null hypothesis,  $T_{ZZ}$  and a chi-squared-type mixture have the same normal or non-normal limiting distribution.

**Value**

A (list) object of S3 class `htest` containing the following elements:

**p.value** the p-value of the test proposed by Zhang and Zhu (2022).

**statistic** the test statistic proposed by Zhang and Zhu (2022).

**beta0** parameter used in Zhang and Zhu (2022)'s test.

**beta1** parameter used in Zhang and Zhu (2022)'s test.

**df** estimated approximate degrees of freedom of Zhang and Zhu (2022)'s test.

**References**

Zhang J, Zhu T (2022). "A further study on Chen-Qin's test for two-sample Behrens-Fisher problems for high-dimensional data." *Journal of Statistical Theory and Practice*, **16**(1), 1. doi:10.1007/s4251902100232w.

**Examples**

```

set.seed(1234)
n1 <- 20
n2 <- 30
p <- 50
mu1 <- t(t(rep(0, p)))
mu2 <- mu1
rho1 <- 0.1
rho2 <- 0.2
a1 <- 1
a2 <- 2
w1 <- (-2 * sqrt(a1 * (1 - rho1)) + sqrt(4 * a1 * (1 - rho1) + 4 * p * a1 * rho1)) / (2 * p)
x1 <- w1 + sqrt(a1 * (1 - rho1))
Gamma1 <- matrix(rep(w1, p * p), nrow = p)
diag(Gamma1) <- rep(x1, p)
w2 <- (-2 * sqrt(a2 * (1 - rho2)) + sqrt(4 * a2 * (1 - rho2) + 4 * p * a2 * rho2)) / (2 * p)
x2 <- w2 + sqrt(a2 * (1 - rho2))
Gamma2 <- matrix(rep(w2, p * p), nrow = p)
diag(Gamma2) <- rep(x2, p)
Z1 <- matrix(rnorm(n1 * p, mean = 0, sd = 1), p, n1)
Z2 <- matrix(rnorm(n2 * p, mean = 0, sd = 1), p, n2)
y1 <- Gamma1 %*% Z1 + mu1 %*% (rep(1, n1))
y2 <- Gamma2 %*% Z2 + mu2 %*% (rep(1, n2))
tsbf_zz2022(y1, y2)

```

---

tsbf\_zzgz2021

*Test proposed by Zhang et al. (2021)*


---

**Description**

Zhang et al. (2021)'s test for testing equality of two-sample high-dimensional mean vectors without assuming that two covariance matrices are the same.

**Usage**

```
tsbf_zzgz2021(y1, y2)
```

**Arguments**

y1	The data matrix (p by n1) from the first population. Each column represents a $p$ -dimensional observation.
y2	The data matrix (p by n2) from the first population. Each column represents a $p$ -dimensional observation.

## Details

Suppose we have two independent high-dimensional samples:

$$\mathbf{y}_{i1}, \dots, \mathbf{y}_{in_i}, \text{ are i. i. d. with } E(\mathbf{y}_{i1}) = \boldsymbol{\mu}_i, \text{ Cov}(\mathbf{y}_{i1}) = \boldsymbol{\Sigma}_i, i = 1, 2.$$

The primary object is to test

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ versus } H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2.$$

Zhang et al.(2021) proposed the following test statistic:

$$T_{ZZGZ} = \frac{n_1 n_2}{n} \|\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2\|^2,$$

where  $\bar{\mathbf{y}}_i, i = 1, 2$  are the sample mean vectors. They showed that under the null hypothesis,  $T_{ZZGZ}$  and a chi-squared-type mixture have the same normal or non-normal limiting distribution.

## Value

A (list) object of S3 class `hstest` containing the following elements:

**p.value** the p-value of the test proposed by Zhang et al. (2021).

**statistic** the test statistic proposed by Zhang et al. (2021).

**beta** parameter used in Zhang et al. (2021)'s test.

**df** estimated approximate degrees of freedom of Zhang et al. (2021)'s test.

## References

Zhang J, Zhou B, Guo J, Zhu T (2021). "Two-sample Behrens-Fisher problems for high-dimensional data: A normal reference approach." *Journal of Statistical Planning and Inference*, **213**, 142–161. [doi:10.1016/j.jspi.2020.11.008](https://doi.org/10.1016/j.jspi.2020.11.008).

## Examples

```
set.seed(1234)
n1 <- 20
n2 <- 30
p <- 50
mu1 <- t(rep(0, p))
mu2 <- mu1
rho1 <- 0.1
rho2 <- 0.2
a1 <- 1
a2 <- 2
w1 <- (-2 * sqrt(a1 * (1 - rho1)) + sqrt(4 * a1 * (1 - rho1) + 4 * p * a1 * rho1)) / (2 * p)
x1 <- w1 + sqrt(a1 * (1 - rho1))
Gamma1 <- matrix(rep(w1, p * p), nrow = p)
diag(Gamma1) <- rep(x1, p)
w2 <- (-2 * sqrt(a2 * (1 - rho2)) + sqrt(4 * a2 * (1 - rho2) + 4 * p * a2 * rho2)) / (2 * p)
x2 <- w2 + sqrt(a2 * (1 - rho2))
Gamma2 <- matrix(rep(w2, p * p), nrow = p)
```

```
diag(Gamma2) <- rep(x2, p)
Z1 <- matrix(rnorm(n1*p,mean = 0,sd = 1), p, n1)
Z2 <- matrix(rnorm(n2*p,mean = 0,sd = 1), p, n2)
y1 <- Gamma1 %*% Z1 + mu1%*(rep(1,n1))
y2 <- Gamma2 %*% Z2 + mu2%*(rep(1,n2))
tsbf_zzg2021(y1, y2)
```

tsbf\_zzz2023

*Test proposed by Zhang et al. (2023)***Description**

Zhang et al. (2023)'s test for testing equality of two-sample high-dimensional mean vectors without assuming that two covariance matrices are the same.

**Usage**

```
tsbf_zzz2023(y1, y2, cutoff)
```

**Arguments**

y1	The data matrix (p by n1) from the first population. Each column represents a $p$ -dimensional observation.
y2	The data matrix (p by n2) from the first population. Each column represents a $p$ -dimensional observation.
cutoff	An empirical criterion for applying the adjustment coefficient

**Details**

Suppose we have two independent high-dimensional samples:

$$\mathbf{y}_{i1}, \dots, \mathbf{y}_{in_i}, \text{ are i. i. d. with } E(\mathbf{y}_{i1}) = \boldsymbol{\mu}_i, \text{ Cov}(\mathbf{y}_{i1}) = \boldsymbol{\Sigma}_i, i = 1, 2.$$

The primary object is to test

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ versus } H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2.$$

Zhang et al.(2023) proposed the following test statistic:

$$T_{ZZZ} = \frac{n_1 n_2}{np} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^\top \hat{\mathbf{D}}_n^{-1} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2),$$

where  $\bar{\mathbf{y}}_i, i = 1, 2$  are the sample mean vectors, and  $\hat{\mathbf{D}}_n = \text{diag}(\hat{\boldsymbol{\Sigma}}_1/n + \hat{\boldsymbol{\Sigma}}_2/n)$  with  $n = n_1 + n_2$ . They showed that under the null hypothesis,  $T_{ZZZ}$  and a chi-squared-type mixture have the same limiting distribution.

**Value**

A (list) object of S3 class `htest` containing the following elements:

**p.value** the p-value of the test proposed by Zhang et al. (2023)'s test.

**statistic** the test statistic proposed by Zhang et al. (2023)'s test.

**df** estimated approximate degrees of freedom of Zhang et al. (2023)'s test.

**cpn** the adjustment coefficient used in Zhang et al. (2023)'s test.

**References**

Zhang L, Zhu T, Zhang J (2023). “Two-sample Behrens–Fisher problems for high-dimensional data: a normal reference scale-invariant test.” *Journal of Applied Statistics*, **50**(3), 456–476. doi:10.1080/02664763.2020.1834516.

**Examples**

```
set.seed(1234)
n1 <- 20
n2 <- 30
p <- 50
mu1 <- t(rep(0, p))
mu2 <- mu1
rho1 <- 0.1
rho2 <- 0.2
a1 <- 1
a2 <- 2
w1 <- (-2 * sqrt(a1 * (1 - rho1)) + sqrt(4 * a1 * (1 - rho1) + 4 * p * a1 * rho1)) / (2 * p)
x1 <- w1 + sqrt(a1 * (1 - rho1))
Gamma1 <- matrix(rep(w1, p * p), nrow = p)
diag(Gamma1) <- rep(x1, p)
w2 <- (-2 * sqrt(a2 * (1 - rho2)) + sqrt(4 * a2 * (1 - rho2) + 4 * p * a2 * rho2)) / (2 * p)
x2 <- w2 + sqrt(a2 * (1 - rho2))
Gamma2 <- matrix(rep(w2, p * p), nrow = p)
diag(Gamma2) <- rep(x2, p)
Z1 <- matrix(rnorm(n1*p, mean = 0, sd = 1), p, n1)
Z2 <- matrix(rnorm(n2*p, mean = 0, sd = 1), p, n2)
y1 <- Gamma1 %>% Z1 + mu1 %>% (rep(1, n1))
y2 <- Gamma2 %>% Z2 + mu2 %>% (rep(1, n2))
tsbf_zzz2023(y1, y2, cutoff=1.2)
```

---

ts\_bs1996

*Test proposed by Bai and Saranadasa (1996)*


---

**Description**

Bai and Saranadasa (1996)'s test for testing equality of two-sample high-dimensional mean vectors with assuming that two covariance matrices are the same.

**Usage**

```
ts_bs1996(y1, y2)
```

**Arguments**

**y1** The data matrix (p by n1) from the first population. Each column represents a  $p$ -dimensional observation.

**y2** The data matrix (p by n2) from the first population. Each column represents a  $p$ -dimensional observation.

**Details**

Suppose we have two independent high-dimensional samples:

$$\mathbf{y}_{i1}, \dots, \mathbf{y}_{in_i}, \text{ are i.i.d. with } E(\mathbf{y}_{i1}) = \boldsymbol{\mu}_i, \text{ Cov}(\mathbf{y}_{i1}) = \boldsymbol{\Sigma}, i = 1, 2.$$

The primary object is to test

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ versus } H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2.$$

Bai and Saranadasa (1996) proposed the following centralised  $L^2$ -norm-based test statistic:

$$T_{BS} = \frac{n_1 n_2}{n} \|\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2\|^2 - \text{tr}(\hat{\boldsymbol{\Sigma}}),$$

where  $\bar{\mathbf{y}}_i, i = 1, 2$  are the sample mean vectors and  $\hat{\boldsymbol{\Sigma}}$  is the pooled sample covariance matrix. They showed that under the null hypothesis,  $T_{BS}$  is asymptotically normally distributed.

**Value**

A (list) object of S3 class `htest` containing the following elements:

**statistic** the test statistic proposed by Bai and Saranadasa (1996)

**p.value** the  $p$ -value of the test proposed by Bai and Saranadasa (1996).

**References**

Bai Z, Saranadasa H (1996). "Effect of high dimension: by an example of a two sample problem." *Statistica Sinica*, 311–329. <https://www.jstor.org/stable/24306018>.

**Examples**

```
set.seed(1234)
n1 <- 20
n2 <- 30
p <- 50
mu1 <- t(rep(0, p))
mu2 <- mu1
rho <- 0.1
y <- (-2 * sqrt(1 - rho) + sqrt(4 * (1 - rho) + 4 * p * rho)) / (2 * p)
x <- y + sqrt((1 - rho))
```

```

Gamma <- matrix(rep(y, p * p), nrow = p)
diag(Gamma) <- rep(x, p)
Z1 <- matrix(rnorm(n1 * p, mean = 0, sd = 1), p, n1)
Z2 <- matrix(rnorm(n2 * p, mean = 0, sd = 1), p, n2)
y1 <- Gamma %% Z1 + mu1 %% (rep(1, n1))
y2 <- Gamma %% Z2 + mu2 %% (rep(1, n2))
ts_bs1996(y1, y2)

```

ts\_sd2008

*Test proposed by Srivastava and Du (2008)***Description**

Srivastava and Du (2008)'s test for testing equality of two-sample high-dimensional mean vectors with assuming that two covariance matrices are the same.

**Usage**

```
ts_sd2008(y1, y2)
```

**Arguments**

**y1** The data matrix (p by n1) from the first population. Each column represents a p-dimensional observation.

**y2** The data matrix (p by n2) from the first population. Each column represents a p-dimensional observation.

**Details**

Suppose we have two independent high-dimensional samples:

$$\mathbf{y}_{i1}, \dots, \mathbf{y}_{in_i}, \text{ are i. i. d. with } E(\mathbf{y}_{i1}) = \boldsymbol{\mu}_i, \text{ Cov}(\mathbf{y}_{i1}) = \boldsymbol{\Sigma}, i = 1, 2.$$

The primary object is to test

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ versus } H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2.$$

Srivastava and Du (2008) proposed the following test statistic:

$$T_{SD} = \frac{n^{-1}n_1n_2(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^\top \mathbf{D}_S^{-1}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2) - \frac{(n-2)p}{n-4}}{\sqrt{2 \left[ \text{tr}(\mathbf{R}^2) - \frac{p^2}{n-2} \right] c_{p,n}}},$$

where  $\bar{\mathbf{y}}_i, i = 1, 2$  are the sample mean vectors,  $\mathbf{D}_S$  is the diagonal matrix of sample variance,  $\mathbf{R}$  is the sample correlation matrix and  $c_{p,n}$  is the adjustment coefficient proposed by Srivastava and Du (2008). They showed that under the null hypothesis,  $T_{SD}$  is asymptotically normally distributed.



**Value**

A (list) object of S3 class `htest` containing the following elements:

**statistic** the test statistic proposed by Srivastava and Du (2008).

**p.value** the  $p$ -value of the test proposed by Srivastava and Du (2008).

**cpn** the adjustment coefficient proposed by Srivastava and Du (2008).

**References**

Srivastava MS, Du M (2008). "A test for the mean vector with fewer observations than the dimension." *Journal of Multivariate Analysis*, **99**(3), 386–402. doi:10.1016/j.jmva.2006.11.002.

**Examples**

```
set.seed(1234)
n1 <- 20
n2 <- 30
p <- 50
mu1 <- t(rep(0, p))
mu2 <- mu1
rho <- 0.1
y <- (-2 * sqrt(1 - rho) + sqrt(4 * (1 - rho) + 4 * p * rho)) / (2 * p)
x <- y + sqrt((1 - rho))
Gamma <- matrix(rep(y, p * p), nrow = p)
diag(Gamma) <- rep(x, p)
Z1 <- matrix(rnorm(n1 * p, mean = 0, sd = 1), p, n1)
Z2 <- matrix(rnorm(n2 * p, mean = 0, sd = 1), p, n2)
y1 <- Gamma %% Z1 + mu1 %% (rep(1, n1))
y2 <- Gamma %% Z2 + mu2 %% (rep(1, n2))
ts_sd2008(y1, y2)
```

---

 ts\_zgzc2020

---

*Test proposed by Zhang et al. (2020)*


---

**Description**

Zhang et al. (2020)'s test for testing equality of two-sample high-dimensional mean vectors with assuming that two covariance matrices are the same.

**Usage**

```
ts_zgzc2020(y1, y2)
```

**Arguments**

`y1` The data matrix ( $p$  by  $n_1$ ) from the first population. Each column represents a  $p$ -dimensional observation.

`y2` The data matrix ( $p$  by  $n_2$ ) from the first population. Each column represents a  $p$ -dimensional observation.

## Details

Suppose we have two independent high-dimensional samples:

$$\mathbf{y}_{i1}, \dots, \mathbf{y}_{in_i}, \text{ are i. i. d. with } E(\mathbf{y}_{i1}) = \boldsymbol{\mu}_i, \text{ Cov}(\mathbf{y}_{i1}) = \boldsymbol{\Sigma}, i = 1, 2.$$

The primary object is to test

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ versus } H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2.$$

Zhang et al.(2020) proposed the following test statistic:

$$T_{ZGZC} = \frac{n_1 n_2}{n} \|\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2\|^2,$$

where  $\bar{\mathbf{y}}_i, i = 1, 2$  are the sample mean vectors. They showed that under the null hypothesis,  $T_{ZGZC}$  and a chi-squared-type mixture have the same normal or non-normal limiting distribution.

## Value

A (list) object of S3 class `htest` containing the following elements:

**p.value** the p-value of the test proposed by Zhang et al. (2020).

**statistic** the test statistic proposed by Zhang et al. (2020).

**beta** parameter used in Zhang et al. (2020)'s test.

**df** estimated approximate degrees of freedom of Zhang et al. (2020)'s test.

## References

Zhang J, Guo J, Zhou B, Cheng M (2020). "A simple two-sample test in high dimensions based on L 2-norm." *Journal of the American Statistical Association*, **115**(530), 1011–1027. doi:10.1080/01621459.2019.1604366.

## Examples

```
set.seed(1234)
n1 <- 20
n2 <- 30
p <- 50
mu1 <- t(rep(0, p))
mu2 <- mu1
rho <- 0.1
y <- (-2 * sqrt(1 - rho) + sqrt(4 * (1 - rho) + 4 * p * rho)) / (2 * p)
x <- y + sqrt((1 - rho))
Gamma <- matrix(rep(y, p * p), nrow = p)
diag(Gamma) <- rep(x, p)
Z1 <- matrix(rnorm(n1 * p, mean = 0, sd = 1), p, n1)
Z2 <- matrix(rnorm(n2 * p, mean = 0, sd = 1), p, n2)
y1 <- Gamma %*% Z1 + mu1 %*% (rep(1, n1))
y2 <- Gamma %*% Z2 + mu2 %*% (rep(1, n2))
ts_zgzc2020(y1, y2)
```

ts\_zz2022

*Test proposed by Zhang and Zhu (2022)***Description**

Zhang and Zhu (2022)'s test for testing equality of two-sample high-dimensional mean vectors with assuming that two covariance matrices are the same.

**Usage**

```
ts_zz2022(y1, y2)
```

**Arguments**

**y1** The data matrix ( $p$  by  $n_1$ ) from the first population. Each column represents a  $p$ -dimensional observation.

**y2** The data matrix ( $p$  by  $n_2$ ) from the first population. Each column represents a  $p$ -dimensional observation.

**Details**

Suppose we have two independent high-dimensional samples:

$$\mathbf{y}_{i1}, \dots, \mathbf{y}_{in_i}, \text{ are i. i. d. with } E(\mathbf{y}_{i1}) = \boldsymbol{\mu}_i, \text{ Cov}(\mathbf{y}_{i1}) = \boldsymbol{\Sigma}, i = 1, 2.$$

The primary object is to test

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ versus } H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2.$$

Zhang et al.(2022) proposed the following test statistic:

$$T_{ZZ} = \frac{n_1 n_2}{n} \|\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2\|^2 - \text{tr}(\hat{\boldsymbol{\Sigma}}),$$

where  $\bar{\mathbf{y}}_i, i = 1, 2$  are the sample mean vectors and  $\hat{\boldsymbol{\Sigma}}$  is the pooled sample covariance matrix. They showed that under the null hypothesis,  $T_{ZZ}$  and a chi-squared-type mixture have the same normal or non-normal limiting distribution.

**Value**

A (list) object of S3 class `htest` containing the following elements:

**p.value** the p-value of the test proposed by Zhang and Zhu (2022).

**statistic** the test statistic proposed by Zhang and Zhu (2022).

**beta0** parameter used in Zhang and Zhu (2022)'s test

**beta1** parameter used in Zhang and Zhu (2022)'s test

**df** estimated approximate degrees of freedom of Zhang and Zhu (2022)'s test.

## References

Zhang J, Zhu T (2022). “A revisit to Bai–Saranadasa’s two-sample test.” *Journal of Nonparametric Statistics*, **34**(1), 58–76. doi:10.1080/10485252.2021.2015768.

## Examples

```
set.seed(1234)
n1 <- 20
n2 <- 30
p <- 50
mu1 <- t(t(rep(0, p)))
mu2 <- mu1
rho <- 0.1
y <- (-2 * sqrt(1 - rho) + sqrt(4 * (1 - rho) + 4 * p * rho)) / (2 * p)
x <- y + sqrt((1 - rho))
Gamma <- matrix(rep(y, p * p), nrow = p)
diag(Gamma) <- rep(x, p)
Z1 <- matrix(rnorm(n1 * p, mean = 0, sd = 1), p, n1)
Z2 <- matrix(rnorm(n2 * p, mean = 0, sd = 1), p, n2)
y1 <- Gamma %>% Z1 + mu1 %>% (rep(1, n1))
y2 <- Gamma %>% Z2 + mu2 %>% (rep(1, n2))
ts_zz2022(y1, y2)
```

---

ts\_zzz2020

*Test proposed by Zhang et al. (2020)*

---

## Description

Zhang et al. (2020)’s test for testing equality of two-sample high-dimensional mean vectors with assuming that two covariance matrices are the same.

## Usage

```
ts_zzz2020(y1, y2)
```

## Arguments

- y1            The data matrix ( $p$  by  $n_1$ ) from the first population. Each column represents a  $p$ -dimensional observation.
- y2            The data matrix ( $p$  by  $n_2$ ) from the first population. Each column represents a  $p$ -dimensional observation.

## Details

Suppose we have two independent high-dimensional samples:

$$\mathbf{y}_{i1}, \dots, \mathbf{y}_{in_i}, \text{ are i. i. d. with } E(\mathbf{y}_{i1}) = \boldsymbol{\mu}_i, \text{ Cov}(\mathbf{y}_{i1}) = \boldsymbol{\Sigma}, i = 1, 2.$$

The primary object is to test

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ versus } H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2.$$

Zhang et al.(2020) proposed the following test statistic:

$$T_{ZZZ} = \frac{n_1 n_2}{np} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^\top \hat{\mathbf{D}}^{-1} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2),$$

where  $\bar{\mathbf{y}}_i, i = 1, 2$  are the sample mean vectors,  $\hat{\mathbf{D}}$  is the diagonal matrix of sample covariance matrix. They showed that under the null hypothesis,  $T_{ZZZ}$  and a chi-squared-type mixture have the same limiting distribution.

## Value

A (list) object of S3 class `htest` containing the following elements:

**p.value** the p-value of the test proposed by Zhang et al. (2020).

**statistic** the test statistic proposed by Zhang et al. (2020).

**df** estimated approximate degrees of freedom of Zhang et al. (2020)'s test.

## References

Zhang L, Zhu T, Zhang J (2020). "A simple scale-invariant two-sample test for high-dimensional data." *Econometrics and Statistics*, **14**, 131–144. doi:10.1016/j.ecosta.2019.12.002.

## Examples

```
set.seed(1234)
n1 <- 20
n2 <- 30
p <- 50
mu1 <- t(t(rep(0, p)))
mu2 <- mu1
rho <- 0.1
y <- (-2 * sqrt(1 - rho) + sqrt(4 * (1 - rho) + 4 * p * rho)) / (2 * p)
x <- y + sqrt((1 - rho))
Gamma <- matrix(rep(y, p * p), nrow = p)
diag(Gamma) <- rep(x, p)
Z1 <- matrix(rnorm(n1 * p, mean = 0, sd = 1), p, n1)
Z2 <- matrix(rnorm(n2 * p, mean = 0, sd = 1), p, n2)
y1 <- Gamma %*% Z1 + mu1 %*% (rep(1, n1))
y2 <- Gamma %*% Z2 + mu2 %*% (rep(1, n2))
ts_zzz2020(y1, y2)
```

# Index

- \* **datasets**
    - corneal, [3](#)
    - COVID19, [4](#)
  - \* **data**
    - corneal, [3](#)
    - COVID19, [4](#)
  - \* **glht**
    - glht\_fhw2004, [10](#)
    - glht\_sf2006, [11](#)
    - glht\_ys2012, [13](#)
    - glhtbf\_zgz2017, [5](#)
    - ks\_s2007, [19](#)
  - \* **nraglht**
    - glht\_zgz2017, [14](#)
    - glht\_zz2022, [16](#)
    - glht\_zzz2022, [18](#)
    - glhtbf\_zz2022, [6](#)
    - glhtbf\_zzg2022, [8](#)
  - \* **nrats**
    - ts\_zgzc2020, [33](#)
    - ts\_zz2022, [35](#)
    - ts\_zzz2020, [36](#)
    - tsbf\_zwz2023, [24](#)
    - tsbf\_zz2022, [25](#)
    - tsbf\_zzgz2021, [27](#)
    - tsbf\_zzz2023, [29](#)
  - \* **ts**
    - ts\_bs1996, [30](#)
    - ts\_sd2008, [32](#)
    - tsbf\_cq2010, [21](#)
    - tsbf\_skk2013, [22](#)
- corneal, [3](#)  
COVID19, [4](#)
- glht\_fhw2004, [10](#)  
glht\_sf2006, [11](#)  
glht\_ys2012, [13](#)  
glht\_zgz2017, [14](#)  
glht\_zz2022, [16](#)
- glht\_zzz2022, [18](#)  
glhtbf\_zgz2017, [5](#)  
glhtbf\_zz2022, [6](#)  
glhtbf\_zzg2022, [8](#)
- ks\_s2007, [19](#)
- ts\_bs1996, [30](#)  
ts\_sd2008, [32](#)  
ts\_zgzc2020, [33](#)  
ts\_zz2022, [35](#)  
ts\_zzz2020, [36](#)  
tsbf\_cq2010, [21](#)  
tsbf\_skk2013, [22](#)  
tsbf\_zwz2023, [24](#)  
tsbf\_zz2022, [25](#)  
tsbf\_zzgz2021, [27](#)  
tsbf\_zzz2023, [29](#)