Package 'mvhtests'

January 9, 2025

Type Package

Title Multivariate Hypothesis Tests

Version 1.1

Date 2025-01-08

Author Michail Tsagris [aut, cre]

Maintainer Michail Tsagris <mtsagris@uoc.gr>

Depends R (>= 4.0)

Imports doParallel, emplik, foreach, parallel, Rfast, Rfast2, stats

Suggests highd2means

Description Hypothesis tests for multivariate data. Tests for one and two mean vectors, multivariate analysis of variance, tests for one, two or more covariance matrices. References include: Mardia K.V., Kent J.T. and Bibby J.M. (1979). Multivariate Analysis. ISBN: 978-0124712522. London: Academic Press.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2025-01-09 09:20:01 UTC

Contents

---

mvhtests-package   *Multivariate Hypothesis Tests*

Description

Multivariate Hypothesis Tests.

Details

Package:	mvhtests
Type:	Package
Version:	1.1
Date:	2025-01-08
License:	GPL-2

Maintainers

Michail Tsagris <mtsagris@uoc.gr>.

Author(s)

Michail Tsagris <mtsagris@uoc.gr>.

References

Aitchison J. (1986). The statistical analysis of compositional data. Chapman & Hall.

Amaral G.J.A., Dryden I.L. and Wood A.T.A. (2007). Pivotal bootstrap methods for k-sample problems in directional statistics and shape analysis. Journal of the American Statistical Association, 102(478): 695–707.

Efron B. (1981) Nonparametric standard errors and confidence intervals. Canadian Journal of Statistics, 9(2): 139–158.

Emerson S. (2009). Small sample performance and calibration of the Empirical Likelihood method. PhD thesis, Stanford university.

Everitt B. (2005). An R and S-Plus Companion to Multivariate Analysis. Springer.

James G.S. (1954). Tests of Linear Hypotheses in Univariate and Multivariate Analysis when the Ratios of the Population Variances are Unknown. Biometrika, 41(1/2): 19–43.

Jing B.Y. and Wood A.T.A. (1996). Exponential empirical likelihood is not Bartlett correctable. Annals of Statistics, 24(1): 365–369.

Jing B.Y. and Robinson J. (1997). Two-sample nonparametric tilting method. Australian Journal of Statistics, 39(1): 25–34.

Johnson R.A. and Wichern D.W. (2007, 6th Edition). Applied Multivariate Statistical Analysis.

Krishnamoorthy K. and Yu J. (2004). Modified Nel and Van der Merwe test for the multivariate Behrens-Fisher problem. Statistics & Probability Letters, 66(2): 161–169.

Krishnamoorthy K. and Yanping X. (2006). On Selecting Tests for Equality of Two Normal Mean Vectors. Multivariate Behavioral Research, 41(4): 533–548.

Mardia K.V., Kent J.T. and Bibby J.M. (1979). Multivariate Analysis. London: Academic Press.

Owen A.B. (1988). Empirical likelihood ratio confidence intervals for a single functional. Biometrika, 75(2): 237–249.

Owen A. (1990). Empirical likelihood ratio confidence regions. Annals of Statistics, 18(1): 90–120.

Owen A. B. (2001). Empirical likelihood. Chapman and Hall/CRC Press.

Preston S.P. and Wood A.T.A. (2010). Two-Sample Bootstrap Hypothesis Tests for Three-Dimensional Labelled Landmark Data. Scandinavian Journal of Statistics 37(4): 568–587.

Todorov V. and Filzmoser P. (2010). Robust Statistic for the One-way MANOVA. Computational Statistics & Data Analysis 54(1): 37–48.

Box's M test for equality of two or more covariance matrices

Box's M test for equality of two or more covariance matrices

Description

Box's M test for equality of two or more covariance matrices.

Usage

```
Mtest.cov(x, ina, a = 0.05)
```

Arguments

x	A matrix containing Euclidean data.
ina	A vector denoting the groups of the data.
a	The significance level, set to 0.05 by default.

Details

According to Mardia, Kent and Bibby (1979, pg. 140), it may be argued that if n_i is small, then the log-likelihood ratio test (function `likel.cov`) gives too much weight to the contribution of $\mathbf{S}$. This consideration led Box (1949) to propose another test statistic in place of that seen in `likel.cov` . Box's M is given by

$$M = \gamma \sum_{i=1}^{k} (n_i - 1) \log \left| \mathbf{S}_i^{-1} \mathbf{S}_p \right|,$$

where $\gamma = 1 - \frac{2p^2+3p-1}{6(p+1)(k-1)} \left(\sum_{i=1}^{k} \frac{1}{n_i-1} - \frac{1}{n-k} \right)$ and $\mathbf{S}_i$ and $\mathbf{S}_p$ are the i-th unbiased covariance estimator and the pooled covariance matrix, respectively with $\mathbf{S}_p = \frac{\sum_{i=1}^{k}(n_i-1)\mathbf{S}_i}{n-k}$. Box's M also has an asymptotic χ^2 distribution with $\frac{1}{2}(p+1)(k-1)$ degrees of freedom. Box's approximation seems to be good if each n_i exceeds 20 and if k and p do not exceed 5 (Bibby and Kent (1979) pg. 140).

Value

A vector with the the test statistic, the p-value, the degrees of freedom and the critical value of the test.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Mardia K.V., Kent J.T. and Bibby J.M. (1979). Multivariate Analysis. London: Academic Press.

See Also

equal.cov, likel.cov

Examples

```
x <- as.matrix( iris[, 1:4] )
ina <- iris[, 5]
Mtest.cov(x, ina)
```

Empirical likelihood for a one sample mean vector hypothesis testing
Empirical likelihood for a one sample mean vector hypothesis testing

---

Description

Empirical likelihood for a one sample mean vector hypothesis testing.

Usage

```
el.test1(x, mu, R = 1, ncores = 1, graph = FALSE)
```

Arguments

x	A matrix containing Euclidean data.
mu	The hypothesized mean vector.
R	If R is 1 no bootstrap calibration is performed and the classical p-value via the χ^2 distribution is returned. If R is greater than 1, the bootstrap p-value is returned.
ncores	The number of cores to use, set to 1 by default.
graph	A boolean variable which is taken into consideration only when bootstrap calibration is performed. IF TRUE the histogram of the bootstrap test statistic values is plotted.

Details

The H_0 is that $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ and the constraint imposed by EL is

$$\frac{1}{n} \sum_{i=1}^{n} \left\{ \left[1 + \boldsymbol{\lambda}^T \left(\mathbf{x}_i - \boldsymbol{\mu}_0 \right) \right]^{-1} \left(\mathbf{x}_i - \boldsymbol{\mu}_0 \right) \right\} = \mathbf{0},$$

where the $\boldsymbol{\lambda}$ is the Lagrangian parameter introduced to maximize the above expression. Note that the maximization of is with respect to the $\boldsymbol{\lambda}$. The probabilities have the following form

$$p_i = \frac{1}{n} \left[1 + \boldsymbol{\lambda}^T \left(\mathbf{x}_i - \boldsymbol{\mu}_0 \right) \right]^{-1}.$$

The log-likelihood ratio test statistic can be written as

$$\Lambda = \sum_{i=1}^{n} \log n p_i.$$

where d denotes the number of variables. Under H_0 $\Lambda \sim \chi_d^2$, asymptotically. Alternatively the bootstrap p-value may be computed.

Value

A list with the outcome of the function **el.test**() from the package **emplik** which includes the -2 log-likelihood ratio, the observed P-value by chi-square approximation, the final value of Lagrange multiplier λ, the gradient at the maximum, the Hessian matrix, the weights on the observations (probabilities multiplied by the sample size) and the number of iteration performed. In addition the runtime of the procedure is reported. In the case of bootstrap, the bootstrap p-value is also returned.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Jing B.Y. and Wood A.T.A. (1996). Exponential empirical likelihood is not Bartlett correctable. Annals of Statistics, 24(1): 365–369.

Owen A. (1990). Empirical likelihood ratio confidence regions. Annals of Statistics, 18(1): 90–120.

Owen A.B. (2001). Empirical likelihood. Chapman and Hall/CRC Press.

See Also

eel.test1, hotel1T2, james, hotel2T2, maov, el.test2

Examples

```
x <- as.matrix(iris[, 1:4])
el.test1(x, mu = numeric(4) )
eel.test1(x, mu = numeric(4) )
```

Empirical likelihood hypothesis testing for two mean vectors

Empirical likelihood hypothesis testing for two mean vectors

Description

Empirical likelihood hypothesis testing for two mean vectors.

Usage

```
el.test2(y1, y2, R = 0, ncores = 1, graph = FALSE)
```

Arguments

y1	A matrix containing the Euclidean data of the first group.
y2	A matrix containing the Euclidean data of the second group.
R	If R is 0, the classical chi-square distribution is used, if R = 1, the corrected chi-square distribution (James, 1954) is used and if R = 2, the modified F distribution (Krishnamoorthy and Yanping, 2006) is used. If R is greater than 3 bootstrap calibration is performed.
ncores	How many to cores to use.
graph	A boolean variable which is taken into consideration only when bootstrap calibration is performed. IF TRUE the histogram of the bootstrap test statistic values is plotted.

Details

The H_0 is that $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ and the two constraints imposed by EL are

$$\frac{1}{n_j} \sum_{i=1}^{n_j} \left\{ \left[1 + \boldsymbol{\lambda}_j^T \left(\mathbf{x}_{ji} - \boldsymbol{\mu} \right) \right]^{-1} \left(\mathbf{x}_{ij} - \boldsymbol{\mu} \right) \right\} = \mathbf{0},$$

where $j = 1, 2$ and the $\boldsymbol{\lambda}_j s$ are Lagrangian parameters introduced to maximize the above expression. Note that the maximization of is with respect to the $\boldsymbol{\lambda}_j s$. The probabilities of the j-th sample have the following form

$$p_{ji} = \frac{1}{n_j} \left[1 + \boldsymbol{\lambda}_j^T \left(\mathbf{x}_{ji} - \boldsymbol{\mu} \right) \right]^{-1}$$

. The log-likelihood ratio test statistic can be written as

$$\Lambda = \sum_{j=1}^{2} \sum_{i=1}^{n_j} \log n_j p_{ij}.$$

The test is implemented by searching for the mean vector that minimizes the sum of the two one sample EL test statistics. See `el.test1` for the test statistic in the one-sample case.

Value

A list including:

`test`	The empirical likelihood test statistic value.
`modif.test`	The modified test statistic, either via the chi-square or the F distribution.
`dof`	Thre degrees of freedom of the chi-square or the F distribution.
`pvalue`	The asymptotic or the bootstrap p-value.
`mu`	The estimated common mean vector.
`runtime`	The runtime of the bootstrap calibration.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Amaral G.J.A., Dryden I.L. and Wood A.T.A. (2007). Pivotal bootstrap methods for k-sample problems in directional statistics and shape analysis. Journal of the American Statistical Association, 102(478): 695–707.

Owen A. B. (2001). Empirical likelihood. Chapman and Hall/CRC Press.

Owen A.B. (1988). Empirical likelihood ratio confidence intervals for a single functional. Biometrika, 75(2): 237–249.

Preston S.P. and Wood A.T.A. (2010). Two-Sample Bootstrap Hypothesis Tests for Three-Dimensional Labelled Landmark Data. Scandinavian Journal of Statistics, 37(4): 568–587.

See Also

eel.test2, maovjames, maov, hotel2T2, james

Examples

```
el.test2( y1 = as.matrix(iris[1:25, 1:4]), y2 = as.matrix(iris[26:50, 1:4]), R = 0 )
el.test2( y1 = as.matrix(iris[1:25, 1:4]), y2 = as.matrix(iris[26:50, 1:4]), R = 1 )
el.test2( y1 =as.matrix(iris[1:25, 1:4]), y2 = as.matrix(iris[26:50, 1:4]), R = 2 )
```

Exponential empirical likelihood for a one sample mean vector
hypothesis testing

Exponential empirical likelihood for a one sample mean vector hypothesis testing

---

Description

Exponential empirical likelihood for a one sample mean vector hypothesis testing.

Usage

```
eel.test1(x, mu, tol = 1e-06, R = 1)
```

Arguments

x	A matrix containing Euclidean data.
mu	The hypothesized mean vector.
tol	The tolerance value used to stop the Newton-Raphson algorithm.
R	The number of bootstrap samples used to calculate the p-value. If R = 1 (default value), no bootstrap calibration is performed

Details

Exponential empirical likelihood or exponential tilting was first introduced by Efron (1981) as a way to perform a "tilted" version of the bootstrap for the one sample mean hypothesis testing. Similarly to the empirical likelihood, positive weights p_i, which sum to one, are allocated to the observations, such that the weighted sample mean $\bar{\mathbf{x}}$ is equal to some population mean $\boldsymbol{\mu}_0$, under the H_0. Under H_1 the weights are equal to $\frac{1}{n}$, where n is the sample size. Following Efron (1981), the choice of $p_i s$ will minimize the Kullback-Leibler distance from H_0 to H_1

$$D\left(L_0, L_1\right) = \sum_{i=1}^{n} p_i \log\left(np_i\right),$$

subject to the constraint $\sum_{i=1}^{n} p_i \mathbf{x}_i = \boldsymbol{\mu}_0$. The probabilities take the form

$$p_i = \frac{e^{\boldsymbol{\lambda}^T \mathbf{x}_i}}{\sum_{j=1}^{n} e^{\boldsymbol{\lambda}^T \mathbf{x}_j}}$$

and the constraint becomes

$$\frac{\sum_{i=1}^{n} e^{\boldsymbol{\lambda}^T \mathbf{x}_i} \left(\mathbf{x}_i - \boldsymbol{\mu}_0 \right)}{\sum_{j=1}^{n} e^{\boldsymbol{\lambda}^T \mathbf{x}_j}} = 0 \Rightarrow \frac{\sum_{i=1}^{n} \mathbf{x}_i e^{\boldsymbol{\lambda}^T \mathbf{x}_i}}{\sum_{j=1}^{n} e^{\boldsymbol{\lambda}^T \mathbf{x}_j}} - \boldsymbol{\mu}_0 = \mathbf{0}.$$

A numerical search over $\boldsymbol{\lambda}$ is required. Under H_0 $\Lambda \sim \chi_d^2$, where d denotes the number of variables. Alternatively the bootstrap p-value may be computed.

Value

A list including:

p	The estimated probabilities.
lambda	The value of the Lagrangian parameter λ.
iter	The number of iterations required by the newton-Raphson algorithm.
info	The value of the log-likelihood ratio test statistic along with its corresponding p-value.
runtime	The runtime of the process.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Efron B. (1981) Nonparametric standard errors and confidence intervals. Canadian Journal of Statistics, 9(2): 139–158.

Jing B.Y. and Wood A.T.A. (1996). Exponential empirical likelihood is not Bartlett correctable. Annals of Statistics, 24(1): 365–369.

Owen A. B. (2001). Empirical likelihood. Chapman and Hall/CRC Press.

See Also

el.test1, hotel1T2, james, hotel2T2, maov, el.test2

Examples

```
x <- as.matrix( iris[, 1:4] )
eel.test1(x, numeric(4) )
el.test1(x, numeric(4) )
```

Exponential empirical likelihood hypothesis testing for two mean
vectors

Exponential empirical likelihood hypothesis testing for two mean vectors

---

Description

Exponential empirical likelihood hypothesis testing for two mean vectors.

Usage

```
eel.test2(y1, y2, tol = 1e-07, R = 0, graph = FALSE)
```

Arguments

y1	A matrix containing the Euclidean data of the first group.
y2	A matrix containing the Euclidean data of the second group.
tol	The tolerance level used to terminate the Newton-Raphson algorithm.
R	If R is 0, the classical chi-square distribution is used, if R = 1, the corrected chi-square distribution (James, 1954) is used and if R = 2, the modified F distribution (Krishnamoorthy and Yanping, 2006) is used. If R is greater than 3 bootstrap calibration is performed.
graph	A boolean variable which is taken into consideration only when bootstrap calibration is performed. IF TRUE the histogram of the bootstrap test statistic values is plotted.

Details

Exponential empirical likelihood or exponential tilting was first introduced by Efron (1981) as a way to perform a "tilted" version of the bootstrap for the one sample mean hypothesis testing. Similarly to the empirical likelihood, positive weights p_i, which sum to one, are allocated to the observations, such that the weighted sample mean $\bar{\mathbf{x}}$ is equal to some population mean $\boldsymbol{\mu}$, under the H_0. Under H_1 the weights are equal to $\frac{1}{n}$, where n is the sample size. Following Efron (1981), the choice of $p_i s$ will minimize the Kullback-Leibler distance from H_0 to H_1

$$D\left(L_0, L_1\right) = \sum_{i=1}^{n} p_i \log\left(n p_i\right),$$

subject to the constraint $\sum_{i=1}^{n} p_i \mathbf{x}_i = \boldsymbol{\mu}$. The probabilities take the form

$$p_i = \frac{e^{\boldsymbol{\lambda}^T \mathbf{x}_i}}{\sum_{j=1}^{n} e^{\boldsymbol{\lambda}^T \mathbf{x}_j}}$$

and the constraint becomes

$$\frac{\sum_{i=1}^{n} e^{\boldsymbol{\lambda}^T \mathbf{x}_i}\left(\mathbf{x}_i - \boldsymbol{\mu}\right)}{\sum_{j=1}^{n} e^{\boldsymbol{\lambda}^T \mathbf{x}_j}} = 0 \Rightarrow \frac{\sum_{i=1}^{n} \mathbf{x}_i e^{\boldsymbol{\lambda}^T \mathbf{x}_i}}{\sum_{j=1}^{n} e^{\boldsymbol{\lambda}^T \mathbf{x}_j}} - \boldsymbol{\mu} = 0.$$

Similarly to empirical likelihood a numerical search over $\boldsymbol{\lambda}$ is required.

We can derive the asymptotic form of the test statistic in the two sample means case but in a simpler form, generalizing the approach of Jing and Robinson (1997) to the multivariate case as follows. The three constraints are

$$
\begin{aligned}
\left(\sum_{j=1}^{n_1} e^{\boldsymbol{\lambda}_1^T \mathbf{x}_j}\right)^{-1} \left(\sum_{i=1}^{n_1} \mathbf{x}_i e^{\boldsymbol{\lambda}_1^T \mathbf{x}_i}\right) - \boldsymbol{\mu} &= \mathbf{0} \\
\left(\sum_{j=1}^{n_2} e^{\boldsymbol{\lambda}_2^T \mathbf{y}_j}\right)^{-1} \left(\sum_{i=1}^{n_2} \mathbf{y}_i e^{\boldsymbol{\lambda}_2^T \mathbf{y}_i}\right) - \boldsymbol{\mu} &= \mathbf{0} \\
n_1 \boldsymbol{\lambda}_1 + n_2 \boldsymbol{\lambda}_2 &= \mathbf{0}.
\end{aligned}
$$

Similarly to EL the sum of a linear combination of the $\boldsymbol{\lambda}s$ is set to zero. We can equate the first two constraints of

$$
\left(\sum_{j=1}^{n_1} e^{\boldsymbol{\lambda}_1^T \mathbf{x}_j}\right)^{-1} \left(\sum_{i=1}^{n_1} \mathbf{x}_i e^{\boldsymbol{\lambda}_1^T \mathbf{x}_i}\right) = \left(\sum_{j=1}^{n_2} e^{\boldsymbol{\lambda}_2^T \mathbf{y}_j}\right)^{-1} \left(\sum_{i=1}^{n_2} \mathbf{y}_i e^{\boldsymbol{\lambda}_2^T \mathbf{y}_i}\right).
$$

Also, we can write the third constraint of as $\boldsymbol{\lambda}_2 = -\frac{n_1}{n_2}\boldsymbol{\lambda}_1$ and thus rewrite the first two constraints as

$$
\left(\sum_{j=1}^{n_1} e^{\boldsymbol{\lambda}^T \mathbf{x}_j}\right)^{-1} \left(\sum_{i=1}^{n_1} \mathbf{x}_i e^{\boldsymbol{\lambda}^T \mathbf{x}_i}\right) = \left(\sum_{j=1}^{n_2} e^{-\frac{n_1}{n_2}\boldsymbol{\lambda}^T \mathbf{y}_j}\right)^{-1} \left(\sum_{i=1}^{n_2} \mathbf{y}_i e^{-\frac{n_1}{n_2}\boldsymbol{\lambda}^T \mathbf{y}_i}\right).
$$

This trick allows us to avoid the estimation of the common mean. It is not possible though to do this in the empirical likelihood method. Instead of minimisation of the sum of the one-sample test statistics from the common mean, we can define the probabilities by searching for the $\boldsymbol{\lambda}$ which makes the last equation hold true. The third constraint of is a convenient constraint, but Jing and Robinson (1997) mention that even though as a constraint is simple it does not lead to second-order accurate confidence intervals unless the two sample sizes are equal. Asymptotically, the test statistic follows a χ_d^2 under the null hypothesis.

Value

A list including:

test	The empirical likelihood test statistic value.
modif.test	The modified test statistic, either via the chi-square or the F distribution.
dof	The degrees of freedom of the chi-square or the F distribution.
pvalue	The asymptotic or the bootstrap p-value.
mu	The estimated common mean vector.
runtime	The runtime of the bootstrap calibration.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Efron B. (1981) Nonparametric standard errors and confidence intervals. Canadian Journal of Statistics, 9(2): 139–158.

Jing B.Y. and Wood A.T.A. (1996). Exponential empirical likelihood is not Bartlett correctable. Annals of Statistics, 24(1): 365–369.

Jing B.Y. and Robinson J. (1997). Two-sample nonparametric tilting method. Australian Journal of Statistics, 39(1): 25–34.

Owen A.B. (2001). Empirical likelihood. Chapman and Hall/CRC Press.

Preston S.P. and Wood A.T.A. (2010). Two-Sample Bootstrap Hypothesis Tests for Three-Dimensional Labelled Landmark Data. Scandinavian Journal of Statistics 37(4): 568–587.

Tsagris M., Preston S. and Wood A.T.A. (2017). Nonparametric hypothesis testing for equality of means on the simplex. Journal of Statistical Computation and Simulation, 87(2): 406–422.

See Also

el.test2, maovjames, maov, hotel2T2, james

Examples

```
y1 = as.matrix(iris[1:25, 1:4])
y2 = as.matrix(iris[26:50, 1:4])
eel.test2(y1, y2)
eel.test2(y1, y2 )
eel.test2( y1, y2 )
```

Hotelling's multivariate version of the 1 sample t-test for Euclidean
data

*Hotelling's multivariate version of the 1 sample t-test for Euclidean
data*

Description

Hotelling's test for testing one Euclidean population mean vector.

Usage

```
hotel1T2(x, M, a = 0.05, R = 999, graph = FALSE)
```

Arguments

x	A matrix containing Euclidean data.
a	The significance level, set to 0.05 by default.
M	The hypothesized mean vector.

R	If R is 1 no bootstrap calibration is performed and the classical p-value via the F distribution is returned. If R is greater than 1, the bootstrap p-value is returned.
graph	A boolean variable which is taken into consideration only when bootstrap calibration is performed. IF TRUE the histogram of the bootstrap test statistic values is plotted.

Details

The hypothesis test is that a mean vector is equal to some specified vector $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$. We assume that $\boldsymbol{\Sigma}$ is unknown. The first approach to this hypothesis test is parametrically, using the Hotelling's T^2 test Mardia, Bibby and Kent (1979, pg. 125-126). The test statistic is given by

$$T^2 = \frac{(n-p)\,n}{(n-1)\,p} \left(\bar{\mathbf{X}} - \boldsymbol{\mu}\right)^T \mathbf{S}^{-1} \left(\bar{\mathbf{X}} - \boldsymbol{\mu}\right).$$

Under the null hypothesis, the above test statistic follows the $F_{p,n-p}$ distribution. The bootstrap version of the one-sample multivariate generalization of the simple t-test is also included in the function. An extra argument (R) indicates whether bootstrap calibration should be used or not. If R=1, then the asymptotic theory applies, if R>1, then the bootstrap p-value will be applied and the number of re-samples is equal to R.

Value

A list including:

m	The sample mean vector.
info	The test statistic, the p-value, the critical value and the degrees of freedom of the F distribution (numerator and denominator). This is given if no bootstrap calibration is employed.
pvalue	The bootstrap p-value is bootstrap is employed.
runtime	The runtime of the bootstrap calibration.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Mardia K.V., Kent J.T. and Bibby J.M. (1979). Multivariate analysis. London: Academic Press.

See Also

eel.test1, el.test1, james, hotel2T2, maov, el.test2

Examples

```
x <- matrix( rnorm( 100 * 4), ncol = 4)
hotel1T2(x, numeric(4), R = 1)
hotel1T2(x, numeric(4), R = 999, graph = TRUE)
```

Hotelling's multivariate version of the 2 sample t-test for Euclidean
data

Description

Hotelling's test for testing the equality of two Euclidean population mean vectors.

Usage

```
hotel2T2(x1, x2, a = 0.05, R = 999, graph = FALSE)
```

Arguments

x1	A matrix containing the Euclidean data of the first group.
x2	A matrix containing the Euclidean data of the second group.
a	The significance level, set to 0.05 by default.
R	If R is 1 no bootstrap calibration is performed and the classical p-value via the F distribution is returned. If R is greater than 1, the bootstrap p-value is returned.
graph	A boolean variable which is taken into consideration only when bootstrap calibration is performed. IF TRUE the histogram of the bootstrap test statistic values is plotted.

Details

The fist case scenario is when we assume equality of the two covariance matrices. This is called the two-sample Hotelling's T^2 test (Mardia, Kent and Bibby, 1979, pg. 131-140) and Everitt (2005, pg. 139). The test statistic is defined as

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} \left(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2 \right)^T \mathbf{S}^{-1} \left(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2 \right),$$

where $\mathbf{S}$ is the pooled covariance matrix calculated under the assumption of equal covariance matrices $\mathbf{S} = \frac{(n_1-1)\mathbf{S}_1 + (n_2-1)\mathbf{S}_2}{n_1+n_2-2}$. Under H_0 the statistic F given by

$$F = \frac{(n_1 + n_2 - p - 1) T^2}{(n_1 + n_2 - 2) p}$$

follows the F distribution with p and $n_1 + n_2 - p - 1$ degrees of freedom. Similar to the one-sample test, an extra argument (R) indicates whether bootstrap calibration should be used or not. If R=1, then the asymptotic theory applies, if R>1, then the bootstrap p-value will be applied and the number of re-samples is equal to R. The estimate of the common mean used in the bootstrap to transform the data under the null hypothesis the mean vector of the combined sample, of all the observations.

The built-in command manova does the same thing exactly. Try it, the asymptotic F test is what you have to see. In addition, this command allows for more mean vector hypothesis testing for more than two groups. I noticed this command after I had written my function and nevertheless as I mention in the introduction this document has an educational character as well.

Value

A list including:

`mesoi`	The two mean vectors.
`info`	The test statistic, the p-value, the critical value and the degrees of freedom of the F distribution (numerator and denominator). This is given if no bootstrap calibration is employed.
`pvalue`	The bootstrap p-value is bootstrap is employed.
`note`	A message informing the user that bootstrap calibration has been employed.
`runtime`	The runtime of the bootstrap calibration.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Everitt B. (2005). An R and S-Plus Companion to Multivariate Analysis. Springer.

Mardia K.V., Kent J.T. and Bibby J.M. (1979). Multivariate Analysis. London: Academic Press.

Tsagris M., Preston S. and Wood A.T.A. (2017). Nonparametric hypothesis testing for equality of means on the simplex. Journal of Statistical Computation and Simulation, 87(2): 406–422.

See Also

[james](#), [maov](#), [el.test2](#), [eel.test2](#)

Examples

```
hotel2T2( as.matrix(iris[1:25, 1:4]), as.matrix(iris[26:50, 1:4]) )
hotel2T2( as.matrix(iris[1:25, 1:4]), as.matrix(iris[26:50, 1:4]), R = 1 )
```

Hypothesis test for two high-dimensional mean vectors

Hypothesis test for two high-dimensional mean vectors

Description

Hypothesis test for two high-dimensional mean vectors.

Usage

```
sarabai(x1, x2)
```

Arguments

x1   A matrix containing the Euclidean data of the first group.

x2   A matrix containing the Euclidean data of the second group.

Details

High dimensional data are the multivariate data which have many variables (p) and usually a small number of observations (n). It also happens that $p > n$ and this is the case here in this Section. We will see a simple test for the case of $p > n$. In this case, the covariance matrix is not invertible and in addition it can have a lot of zero eigenvalues.

The test we will see was proposed by Bai and Saranadasa (1996). Ever since, there have been some more suggestions but I chose this one for its simplicity. There are two datasets, $\mathbf{X}_1$ and $\mathbf{X}_2$ of sample sizes n_1 and n_2, respectively. Their corresponding sample mean vectors and covariance matrices are $\bar{\mathbf{X}}_1$, $\bar{\mathbf{X}}_2$ and $\mathbf{S}_1$, $\mathbf{S}_2$ respectively. The assumption here is the same as that of the Hotelling's test we saw before.

Let us define the pooled covariance matrix at first, calculated under the assumption of equal covariance matrices $\mathbf{S}_n = \frac{(n_1-1)\mathbf{S}_1 + (n_2-1)\mathbf{S}_2}{n}$, where $n = n_1 + n_2$. Then define $B_n = \sqrt{\frac{n^2}{(n+2)(n-1)} \left\{ \text{tr}\left(\mathbf{S}_n^2\right) - \frac{1}{n} \left[\text{tr}\left(\mathbf{S}_n\right)\right]^2 \right\}}$. The test statistic is

$$Z = \frac{\frac{n_1 n_2}{n_1 + n_2} \left(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2\right)^T \left(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2\right) - \text{tr}\left(\mathbf{S}_n\right)}{\sqrt{\frac{2(n+1)}{n}} B_n}.$$

Under the null hypothesis (equality of the two mean vectors) the test statistic follows the standard normal distribution. Bai and Saranadasa (1996) established the asymptotic normality of the test statistics and showed that it has attractive power property when $p/n \to c < \infty$ and under some restriction on the maximum eigenvalue of the common population covariance matrix. However, the requirement of p and n being of the same order is too restrictive to be used in the "large p small n" situation.

Value

A vector with the test statistic and the p-value.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

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.

See Also

hotel2T2, maov, el.test2, eel.test2

Examples

```
x1 <- matrix( rnorm(40 * 100), ncol = 100 )
x2 <- matrix( rnorm(50 * 100), ncol = 100 )
sarabai(x1, x2)
```

```
James multivariate version of the t-test
```
James multivariate version of the t-test

Description

James test for testing the equality of two population mean vectors without assuming equality of the covariance matrices.

Usage

```
james(y1, y2, a = 0.05, R = 999, graph = FALSE)
```

Arguments

y1	A matrix containing the Euclidean data of the first group.
y2	A matrix containing the Euclidean data of the second group.
a	The significance level, set to 0.05 by default.
R	If R is 1 no bootstrap calibration is performed and the classical p-value via the F distribution is returned. If R is greater than 1, the bootstrap p-value is returned.
graph	A boolean variable which is taken into consideration only when bootstrap calibration is performed. If TRUE the histogram of the bootstrap test statistic values is plotted.

Details

Here we show the modified version of the two-sample T^2 test (function [hotel2T2](#)) in the case where the two covariances matrices cannot be assumed to be equal.

James (1954) proposed a test for linear hypotheses of the population means when the variances (or the covariance matrices) are not known. Its form for two p-dimensional samples is:

$$T_u^2 = \left(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2 \right)^T \tilde{\mathbf{S}}^{-1} \left(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2 \right),$$

where $\tilde{\mathbf{S}} = \tilde{\mathbf{S}}_1 + \tilde{\mathbf{S}}_2 = \frac{\mathbf{S}_1}{n_1} + \frac{\mathbf{S}_2}{n_2}$.

James (1954) suggested that the test statistic is compared with $2h\left(\alpha\right)$, a corrected χ^2 distribution whose form is

$$2h\left(\alpha\right) = \chi^2 \left(A + B\chi^2 \right),$$

where $A = 1 + \frac{1}{2p} \sum_{i=1}^2 \frac{\left(tr\tilde{\mathbf{S}}^{-1}\tilde{\mathbf{S}}_i \right)^2}{n_i - 1}$ and $B = \frac{1}{p(p+2)} \left[\sum_{i=1}^2 \frac{tr\left(\tilde{\mathbf{S}}^{-1}\tilde{\mathbf{S}}_i\right)^2}{n_i - 1} + \frac{1}{2} \sum_{i=1}^2 \frac{\left(tr\tilde{\mathbf{S}}^{-1}\tilde{\mathbf{S}}_i \right)^2}{n_i - 1} \right]$.

If you want to do bootstrap to get the p-value, then you must transform the data under the null hypothesis. The estimate of the common mean is given by Aitchison (1986)

$$
\hat{\boldsymbol{\mu}}_c = \left(n_1 \mathbf{S}_1^{-1} + n_2 \mathbf{S}_2^{-1} \right)^{-1} \left(n_1 \mathbf{S}_1^{-1} \bar{\mathbf{X}}_1 + n_2 \mathbf{S}_2^{-1} \bar{\mathbf{X}}_2 \right) = \left(\tilde{\mathbf{S}}_1^{-1} + \tilde{\mathbf{S}}_2^{-1} \right)^{-1} \left(\tilde{\mathbf{S}}_1^{-1} \bar{\mathbf{X}}_1 + \tilde{\mathbf{S}}_2^{-1} \bar{\mathbf{X}}_2 \right).
$$

The modified Nel and van der Merwe (1986) test is based on the same quadratic form as that of James (1954) but the distribution used to compare the value of the test statistic is different. It is shown in Krishnamoorthy and Yanping (2006) that $T_u^2 \sim \frac{\nu p}{\nu - p + 1} F_{p, \nu - p + 1}$ approximately, where $\nu = \frac{p + p^2}{\frac{1}{n_1} \left\{ \mathrm{tr} \left[\left(\mathbf{S}_1 \tilde{\mathbf{S}} \right)^2 \right] + \mathrm{tr} \left[\left(\mathbf{S}_1 \tilde{\mathbf{S}} \right) \right]^2 \right\} + \frac{1}{n_2} \left\{ \mathrm{tr} \left[\left(\mathbf{S}_2 \tilde{\mathbf{S}} \right)^2 \right] + \mathrm{tr} \left[\left(\mathbf{S}_2 \tilde{\mathbf{S}} \right) \right]^2 \right\}}$.

The algorithm is taken by Krishnamoorthy and Yu (2004).

Value

A list including:

note	A message informing the user about the test used.
mesoi	The two mean vectors.
info	The test statistic, the p-value, the correction factor and the corrected critical value of the chi-square distribution if the James test has been used or, the test statistic, the p-value, the critical value and the degrees of freedom (numerator and denominator) of the F distribution if the modified James test has been used.
pvalue	The bootstrap p-value if bootstrap is employed.
runtime	The runtime of the bootstrap calibration.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Aitchison J. (1986). The statistical analysis of compositional data. Chapman & Hall.

James G.S. (1954). Tests of Linear Hypothese in Univariate and Multivariate Analysis when the Ratios of the Population Variances are Unknown. Biometrika, 41(1/2): 19–43.

Krishnamoorthy K. and Yu J. (2004). Modified Nel and Van der Merwe test for the multivariate Behrens-Fisher problem. Statistics & Probability Letters, 66(2): 161–169.

Krishnamoorthy K. and Yanping Xia (2006). On Selecting Tests for Equality of Two Normal Mean Vectors. Multivariate Behavioral Research, 41(4): 533–548.

Tsagris M., Preston S. and Wood A.T.A. (2017). Nonparametric hypothesis testing for equality of means on the simplex. Journal of Statistical Computation and Simulation, 87(2): 406–422.

See Also

hotel2T2, maovjames, el.test2, eel.test2

Examples

```
james( as.matrix(iris[1:25, 1:4]), as.matrix(iris[26:50, 1:4]), R = 1 )
james( as.matrix(iris[1:25, 1:4]), as.matrix(iris[26:50, 1:4]), R = 2 )
james( as.matrix(iris[1:25, 1:4]), as.matrix(iris[26:50, 1:4]) )
```

Log-likelihood ratio test for equality of one covariance matrix
Log-likelihood ratio test for equality of one covariance matrix

---

Description

Log-likelihood ratio test for equality of one covariance matrix.

Usage

```
equal.cov(x, Sigma, a = 0.05)
```

Arguments

x	A matrix containing Euclidean data.
Sigma	The hypothesis covariance matrix.
a	The significance level, set to 0.05 by default.

Details

The hypothesis test is that the the sample covariance is equal to some specified covariance matrix: $H_0 : \mathbf{\Sigma} = \mathbf{\Sigma}_0$, with $\mathbf{\mu}$ unknown. The algorithm for this test is taken from Mardia, Bibby and Kent (1979, pg. 126-127). The test is based upon the log-likelihood ratio test. The form of the test is

$$-2 \log \lambda = n \text{tr} \left\{ \mathbf{\Sigma}_0^{-1} \mathbf{S} \right\} - n \log \left| \mathbf{\Sigma}_0^{-1} \mathbf{S} \right| - np,$$

where n is the sample size, $\mathbf{\Sigma}_0$ is the specified covariance matrix under the null hypothesis, $\mathbf{S}$ is the sample covariance matrix and p is the dimensionality of the data (or the number of variables). Let α and g denote the arithmetic mean and the geometric mean respectively of the eigenvalues of $\mathbf{\Sigma}_0^{-1} \mathbf{S}$, so that $tr \left\{ \mathbf{\Sigma}_0^{-1} \mathbf{S} \right\} = p\alpha$ and $\left| \mathbf{\Sigma}_0^{-1} \mathbf{S} \right| = g^p$, then the test statistic becomes

$$-2 \log \lambda = np \left(\alpha - log(g) - 1 \right).$$

The degrees of freedom of the χ^2 distribution are $\frac{1}{2} p \left(p + 1 \right)$.

Value

A vector with the the test statistic, the p-value, the degrees of freedom and the critical value of the test.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Mardia K.V., Kent J.T. and Bibby J.M. (1979). Multivariate Analysis. London: Academic Press.

See Also

likel.cov, Mtest.cov

Examples

```
x <- as.matrix( iris[, 1:4] )
s <- cov(x) * 1.5
equal.cov(x, s)
```

Log-likelihood ratio test for equality of two or more covariance
matrices

Log-likelihood ratio test for equality of two or more covariance matrices

---

Description

Log-likelihood ratio test for equality of two or more covariance matrices.

Usage

```
likel.cov(x, ina, a = 0.05)
```

Arguments

x	A matrix containing Euclidean data.
ina	A vector denoting the groups of the data.
a	The significance level, set to 0.05 by default.

Details

Tthe hypothesis test is that of the equality of at least two covariance matrices: $H_0 : \mathbf{\Sigma}_1 = \ldots = \mathbf{\Sigma}_k$. The algorithm is taken from Mardia, Bibby and Kent (1979, pg. 140). The log-likelihood ratio test is the multivariate generalization of Bartlett's test of homogeneity of variances. The test statistic takes the following form

$$-2log\lambda = n \log |\mathbf{S}| - \sum_{i=1}^{k} n_i \log |\mathbf{S_i}| = \sum_{i=1}^{k} n_i \log \left|\mathbf{S}_i^{-1}\mathbf{S}\right|,$$

where $\mathbf{S}_i$ is the i-th sample biased covariance matrix and $\mathbf{S} = n^{-1} \sum_{i=1}^{k} n_i \mathbf{S}_i$ is the maximum likelihood estimate of the common covariance matrix (under the null hypothesis) with $n = \sum_{i=1}^{k} n_i$. The degrees of freedom of the asymptotic chi-square distribution are $\frac{1}{2}(p+1)(k-1)$.

Value

A vector with the the test statistic, the p-value, the degrees of freedom and the critical value of the test.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Mardia K.V., Kent J.T. and Bibby J.M. (1979). Multivariate Analysis. London: Academic Press.

See Also

equal.cov, Mtest.cov

Examples

```
x <- as.matrix( iris[, 1:4] )
ina <- iris[, 5]
likel.cov(x, ina)
```

Multivariate analysis of variance (James test)

Multivariate analysis of variance (James test)

Description

Multivariate analysis of variance without assuming equality of the covariance matrices.

Usage

```
maovjames(x, ina, a = 0.05)
```

Arguments

x	A matrix containing Euclidean data.
ina	A numerical or factor variable indicating the groups of the data.
a	The significance level, set to 0.005 by default.

Details

James (1954) also proposed an alternative to MANOVA when the covariance matrices are not assumed equal. The test statistic for k samples is

$$J = \sum_{i=1}^{k} \left(\bar{\mathbf{x}}_i - \bar{\mathbf{X}} \right)^T \mathbf{W}_i \left(\bar{\mathbf{x}}_i - \bar{\mathbf{X}} \right),$$

where $\bar{x}_i$ and n_i are the sample mean vector and sample size of the i-th sample respectively and $\mathbf{W}_i = \left(\frac{\mathbf{S}_i}{n_i} \right)^{-1}$, where $\mathbf{S}_i$ is the covariance matrix of the i-sample mean vector and $\bar{\mathbf{X}}$ is the estimate of the common mean $\bar{\mathbf{X}} = \left(\sum_{i=1}^{k} \mathbf{W}_i \right)^{-1} \sum_{i=1}^{k} \mathbf{W}_i \bar{\mathbf{x}}_i$.

Normally one would compare the test statistic with a $\chi^2_{r,1-\alpha}$, where $r = p(k-1)$ are the degrees of freedom with k denoting the number of groups and p the dimensionality of the data. There are r constraints (how many univariate means must be equal, so that the null hypothesis, that all the mean vectors are equal, holds true), that is where these degrees of freedom come from. James (1954) compared the test statistic with a corrected χ^2 distribution instead. Let A and B be $A = 1 + \frac{1}{2r} \sum_{i=1}^{k} \frac{\left[\text{tr} \left(\mathbf{I}_p - \mathbf{W}^{-1} \mathbf{W}_i \right) \right]^2}{n_i - 1}$ and $B = \frac{1}{r(r+2)} \sum_{i=1}^{k} \left\{ \frac{\text{tr} \left[\left(\mathbf{I}_p - \mathbf{W}^{-1} \mathbf{W}_i \right)^2 \right]}{n_i - 1} + \frac{\left[\text{tr} \left(\mathbf{I}_p - \mathbf{W}^{-1} \mathbf{W}_i \right) \right]^2}{2(n_i - 1)} \right\}$.

The corrected quantile of the χ^2 distribution is given as before by $2h(\alpha) = \chi^2 \left(A + B\chi^2 \right)$.

Value

A vector with the next 4 elements:

`test`	The test statistic.
`correction`	The value of the correction factor.
`corr.critical`	The corrected critical value of the chi-square distribution.
`p-value`	The p-value of the corrected test statistic.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

James G.S. (1954). Tests of Linear Hypotheses in Univariate and Multivariate Analysis when the Ratios of the Population Variances are Unknown. Biometrika, 41(1/2): 19–43.

See Also

maov, hotel2T2, james

Examples

```
maov( as.matrix(iris[,1:4]), iris[,5] )
maovjames( as.matrix(iris[,1:4]), iris[,5] )
```

Multivariate analysis of variance assuming equality of the
covariance matrices

Multivariate analysis of variance assuming equality of the covariance matrices

---

Description

Multivariate analysis of variance assuming equality of the covariance matrices.

Usage

```
maov(x, ina)
```

Arguments

x	A matrix containing Euclidean data.
ina	A numerical or factor variable indicating the groups of the data.

Details

Multivariate analysis of variance assuming equality of the covariance matrices.

Value

A list including:

note	A message stating whether the F or the chi^2 approximation has been used.
result	The test statistic and the p-value.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Johnson R.A. and Wichern D.W. (2007, 6th Edition). Applied Multivariate Statistical Analysis, pg. 302–303.

Todorov V. and Filzmoser P. (2010). Robust Statistic for the One-way MANOVA. Computational Statistics & Data Analysis, 54(1): 37–48.

See Also

maovjames, hotel2T2, james

Examples

```
maov( as.matrix(iris[,1:4]), iris[,5] )
maovjames( as.matrix(iris[,1:4]), iris[,5] )
```

Relationship between Hotelling's T2 test and James' MANOVA
            Relationship between Hotelling's $T\hat{\ }2$ test and James' MANOVA

Description

Relationship between Hotelling's T^2 test and James' MANOVA.

Usage

```
maovjames.hotel(x, ina)
```

Arguments

x	A matrix containing the Euclidean data of the first group.
ina	A numerical or factor variable indicating the groups of the data.

Details

The relationship for the James two sample test (see the function `james.hotel`) is true for the case of the MANOVA. The estimate of the common mean, $\boldsymbol{mu}_c$ (see the function `james` for the expression of $\boldsymbol{\mu}_c$), is in general, for g groups, each of sample size n_i, written as

$$\hat{\boldsymbol{\mu}}_c = \left(\sum_{i=1}^{g} n_i \mathbf{S}_i^{-1} \right)^{-1} \sum_{i=1}^{g} n_i \mathbf{S}_i^{-1} \bar{\mathbf{X}}_i.$$

The function is just a proof of the mathematics you will find in Emerson (2009, pg. 76–81) and is again intended for educational purposes.

Value

A list including:

test	The value of the test statistic, the sum of the two Hotelling's test statistic using the common mean.
mc	The common mean.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Emerson S. (2009). Small sample performance and calibration of the Empirical Likelihood method. PhD thesis, Stanford university.

James G.S. (1954). Tests of Linear Hypothese in Univariate and Multivariate Analysis when the Ratios of the Population Variances are Unknown. Biometrika, 41(1/2): 19–43.

See Also

hotel2T2, maovjames, el.test2, eel.test2

Examples

```
maovjames.hotel( as.matrix(iris[, 1:4]), iris[, 5] )
maovjames( as.matrix(iris[, 1:4]), iris[, 5] )
```

Relationship between the Hotelling's and James' tests
Relationship between the Hotelling's $T\hat{\;}2$ and James' test

---

Description

Relationship between the Hotelling's T^2 and James' test.

Usage

```
james.hotel(x1, x2)
```

Arguments

x1	A matrix containing the Euclidean data of the first group.
x2	A matrix containing the Euclidean data of the second group.

Details

Emerson (2009, pg. 76–81) mentioned a very nice result between the Hotelling's one sample T^2 and James test for two mean vectors

$$J\left(\boldsymbol{\mu}\right) = T_1^2\left(\boldsymbol{\mu}\right) + T_2^2\left(\boldsymbol{\mu}\right),$$

where $J\left(\boldsymbol{\mu}\right)$ is the James test statistic (James, 1954) and $T_1^2\left(\boldsymbol{\mu}\right)$ and $T_1^2\left(\boldsymbol{\mu}\right)$ are the two one sample Hotelling's T^2 test statistic values (see function hotel1T2) for each sample from their common mean vector $\boldsymbol{\mu}_c$ (see the help file of james). In fact, James test statistic is found from minimizing the right hand side of the above expression with respect to $\boldsymbol{\mu}$. The sum is mimized when $\boldsymbol{\mu}$ takes the form of the common mean vector $\boldsymbol{\mu}_c$. The same is true for the t-test in the univariate case.

I have created this function illustrating this result, so this one is for educational purposes. It calculates the James test statistic, the sum of the two T^2 test statistics, the common mean vector and the one found via numerical optimization. In the univariate case, the common mean vector is a weighted linear combination of the two sample means. So, if we take a segment connecting the two means, the common mean is somewhere on that segment.

Value

A list including:

tests   A vector with two values, the James test statistic value and the sum of the two
                Hotelling's test statistic using the common mean.

mathematics.mean

                The common mean computed the closed form expression seen in the help file of
                [james](#).

optimised.mean The common mean vector obtained from the minimisation process.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Emerson S. (2009). Small sample performance and calibration of the Empirical Likelihood method.
PhD thesis, Stanford university.

James G.S. (1954). Tests of Linear Hypothese in Univariate and Multivariate Analysis when the
Ratios of the Population Variances are Unknown. Biometrika, 41(1/2): 19–43.

See Also

[hotel2T2](#), [maovjames](#), [el.test2](#), [eel.test2](#)

Examples

```
james.hotel( as.matrix(iris[1:50, 1:4]), as.matrix(iris[51:100, 1:4]) )
james( as.matrix(iris[1:50, 1:4]), as.matrix(iris[51:100, 1:4]), R = 1 )
```

Repeated measures ANOVA (univariate data) using Hotelling's T2 test

*Repeated measures ANOVA (univariate data) using Hotelling's T^2
test*

Description

Repeated measures ANOVA (univariate data) using Hotelling's T^2 test.

Usage

```
rm.hotel(x, a = 0.05)
```

Arguments

x	A numerical matrix with the repeated measurements. Each column contains the values of the repeated measurements.
a	The level of significance, default value is equal to 0.05.

Details

We now show how can one use Hotelling's T^2 test to analyse univariate repeated measures. Univariate analysis of variance for repeated measures is the classical way, but we can use this multivariate test as well. In the repeated measures ANOVA case, we have many repeated observations from the same n subjects, usually at different time points and the interest is to see whether the means of the samples are equal or not $\mu_1 = \mu_2 = \ldots = \mu_k$ assuming k repeated measurements. We can of course change this null hypothesis and test many combinations of means. The idea in any case is to construct a matrix of contrasts. I will focus here in the first case only and in particular the null hypothesis and the matrix of contrasts $\mathbf{C}$ are

$$
\begin{pmatrix}
\mu_1 = \mu_2 \\
\mu_2 = \mu_3 \\
\vdots \\
\mu_{k-1} = \mu_k
\end{pmatrix}
=
\begin{pmatrix}
1 & -1 & 0 & \ldots & 0 \\
1 & 0 & -1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
1 & 0 & 0 & \ldots & -1
\end{pmatrix}
\boldsymbol{\mu} = \mathbf{C}\boldsymbol{\mu}.
$$

The contrast matrix $\mathbf{C}$ has $k-1$ independent rows and if there is no treatment effect, $\mathbf{C}\boldsymbol{\mu} = \mathbf{0}$.

The test statistic is

$$
T_r^2 = \frac{(n-k+1)}{(n-1)(k-1)} n \left(\mathbf{C}\bar{\mathbf{x}} \right)^T \left(\mathbf{C}\mathbf{S}\mathbf{C}^T \right)^{-1} \left(\mathbf{C}\bar{\mathbf{x}} \right) \sim F_{k-1,n-k+1}.
$$

Value

A list including:

m	The mean vector.
result	A vector with the test statistic value, it's associated p-value, the numerator and denominator degrees of freedom and the critical value.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

See Also

maov, hotel2T2, james

Examples

```
x <- as.matrix(iris[, 1:4]) ## assume they are repeated measurements
rm.hotel(x)
```

Index