Package ‘Hotelling’

September 9, 2021

Version 1.0-8
Date 2021-09-09
Title Hotelling’s T^2 Test and Variants
Maintainer James Curran <j.curran@auckland.ac.nz>
Description A set of R functions which implements Hotelling’s T^2 test and some variants of it. Functions are also included for Aitchison’s additive log ratio and centred log ratio transformations.
Depends corpcor
License GPL (>= 2)
URL https://github.com/jmcurran/Hotelling
BugReports https://github.com/jmcurran/Hotelling/issues
Encoding UTF-8
RoxygenNote 7.1.1
NeedsCompilation no
Author James Curran [aut, cre],
Taylor Hersh [aut]
Repository CRAN
Date/Publication 2021-09-09 04:30:05 UTC

R topics documented:

Hotelling-package ........................................... 2
alr .............................................................. 2
bottle.df ..................................................... 3
clr ............................................................ 3
container.df .................................................. 4
hotelling.stat ................................................. 5
hotelling.test ............................................... 6
manova1.df ................................................... 9
plot.hotelling.test ......................................... 9
print.hotelling.test ....................................... 10
summarise .................................................... 11
Description

A set of R functions and data sets which implements Hotelling's T^2 test, and some variants of it. Functions are also included for Aitchison's additive log ratio and centred log ratio transformations.

Author(s)

James Curran Maintainer: James Curran <j.curran@auckland.ac.nz> ~~ The author and/or maintainer of the package ~~

Usage

```
alr(form, data, group = NULL)
```

Arguments

- **form**: a formula which specifies the denominator variable as the response
- **data**: a data frame in which the data is stored
- **group**: if not NULL then a character string specifying the name of the grouping variable

Details

This function will give a warning if zeros are present because the transformed data will have -lnfs.

Value

A data frame with the ALR transformation applied to data. Each row in the data frame is standardized with respect to a specific variable by dividing by that variable. The logarithms of the resulting ratios are returned. If a grouping variable is specified, then this is preserved.

Author(s)

James M. Curran
References


See Also

clr

Examples

data(bottle.df)

## transform with respect to manganese
alr(Mn~., bottle.df, "Number")

## transform the data with respect to barium, but removing the
## grouping in column 1
alr(Ba~., bottle.df[, -1])

Description

This data contains the elemental concentration of five different elements (Manganese, Barium, Strontium, Zirconium, and Titanium) in samples of glass taken from six different Heineken beer bottles. 20 measurements were taken from each bottle.

References


clr

Description

Aitchison’s centered log ratio tranformation for compositional data

Usage

clr(data, group = NULL)
Arguments

- **data**: a data frame in which the data is stored
- **group**: if not NULL then a character string specifying the name of the grouping variable

Details

This function will give a warning if zeros are present because the transformed data will have -Inf.

Value

A data frame with the CLR transformation applied to data. Each row in the data frame is standardized by dividing by the geometric mean of that row. The logarithms of the resulting ratios are returned. If a grouping variable is specified, then this is preserved.

Author(s)

James M. Curran

References


See Also

- alr

Examples

```r
data(bottle.df)

## transform preserving grouping
clr(bottle.df, "Number")

## transform the data but remove the
## grouping in column 1
clr(bottle.df[, -1])
```

---

**container.df**

**Container data**

Description

This data contains the elemental concentration of nine different elements (Titanium, Aluminium, Iron, Manganese, Magnesium, Calcium, Barium, Strontium, and Zirconium) in specimens of glass taken from two different containers. Ten measurements were taken from each container.
hotelling.stat

References


hotelling.stat  Calculate Hotelling’s two sample T-squared test statistic

Description

Calculate Hotelling’s T-squared test statistic for the difference in two multivariate means.

Usage

hotelling.stat(x, y, shrinkage = FALSE, var.equal = TRUE)

Arguments

x  a nx by p matrix containing the data points from sample 1 or a list containing elements mean, cov, and n where mean is a mean vector of length p, cov is a variance-covariance matrix of dimension p by p, and n is the sample size

y  a ny by p matrix containing the data points from sample 2 or a list containing elements mean, cov, and n where mean is a mean vector of length p, cov is a variance-covariance matrix of dimension p by p, and n is the sample size

shrinkage  set to TRUE if the covariance matrices are to be estimated using Schaefer and Strimmer’s James-Stein shrinkage estimator. Note this only works when raw data is supplied, and will not work if summary statistics are supplied.

var.equal  set to TRUE if the covariance matrices are (assumed to be) equal

Details

Note, the sample size requirements are that nx + ny - 1 > p. The procedure will stop if this is not met and the shrinkage estimator is not being used. The shrinkage estimator has not been rigorously tested for this application (small p, smaller n).

Value

A list containing the following components:

statistic  Hotelling’s (unscaled) T-squared statistic

m  The scaling factor - this can be used by by multiplying it with the test statistic, or dividing the critical F value

df  a vector of length containing the numerator and denominator degrees of freedom

nx  The sample size of sample 1

ny  The sample size of sample 2

p  The number of variables to be used in the comparison
Author(s)
James M. Curran
Taylor Hersh

References

Examples

data(container.df)
split.data = split(container.df[,,-1],container.df$gp)
x = split.data[[1]]
y = split.data[[2]]
hotelling.stat(x, y)
hotelling.stat(x, y, TRUE)

hotelling.test

Two-sample Hotelling’s T-squared test

Description
Performs a two-sample Hotelling’s T-squared test for the difference in two multivariate means

Usage
hotelling.test(x, ...)

# Default S3 method:
hotelling.test(
  x,
  y,
  shrinkage = FALSE,
  var.equal = TRUE,
  perm = FALSE,
  B = 10000,
  progBar = (perm && TRUE),
)
hotelling.test

... )

## S3 method for class 'formula'
hotelling.test(x, data = NULL, pair = c(1, 2), ...)

Arguments

x
- a matrix containing the data points from sample 1, or a formula specifying the
elements to be used as a response and the grouping variable as a predictor, or a
list containing elements mean, cov, and n where mean is a mean vector of length
p, cov is a variance-covariance matrix of dimension p by p, and n is the sample
size

... any additional arguments. This is useful to pass the optional arguments for the
default call from the formula version

y
- a matrix containing the data points from sample 2, or a list containing elements
mean, cov, and n where mean is a mean vector of length p, cov is a variance-
covariance matrix of dimension p by p, and n is the sample size

shrinkage
- if TRUE then Shaefer and Strimmer’s James-Stein shrinkage estimator is used to
calculate the sample covariance matrices

var.equal
- set to TRUE if the covariance matrices are (assumed to be) equal

perm
- if TRUE then permutation testing is used to estimate the non-parametric P-value
  for the hypothesis test

B
- if perm is TRUE, then B is the number of permutations to perform

progBar
- if TRUE and perm is TRUE then a progress bar will be displayed whilst the per-
  mutation procedure is carried out

data
- a data frame needs to be specified if a formula is to be used to perform the test

pair
- a vector of length two which can be used when the grouping factor has more
  than two levels to select different pairs of groups. For example for a 3-level
  factor, pairs could be set to c(1, 3) to perform Hotelling’s test between groups
  1 an an 3

Value

A list (which is also of class 'hotelling.test') with the following elements:

stats
- a list containing all of the output from hotelling.stat

pval
- the P-value from the test

results
- if perm == TRUE, then all of the permutation test statistics are stored in results

Methods (by class)

- default: Two-sample Hotelling’s T-squared test
- formula: Two-sample Hotelling’s T-squared test
Author(s)

James M. Curran

References


See Also

hotelling.stat

Examples

data(container.df)
fit = hotelling.test(.~gp, data = container.df)
fit

subs.df = container.df[1:10,]
subs.df$gp = rep(1:2, c(5,5))
fitPerm = hotelling.test(Al+Fe~gp, data = subs.df, perm = TRUE)
fitPerm
plot(fitPerm)

data(bottle.df)
fit12 = hotelling.test(.~Number, data = bottle.df)
fit12

fit23 = hotelling.test(.~Number, data = bottle.df, pair = c(2,3))
fit23

data(manova1.df)
fit = hotelling.test(wratr+wrata~treatment, data = manova1.df, var.equal = FALSE)
fit

x = list(mean = c(7.81, 108.77, 44.92),
  cov = matrix(c(0.461, 1.18, 4.49,
                1.18, 3776.4, -17.35,
                4.49, -17.35, 147.24), nc = 3, byrow = TRUE),
  n = 13)
y = list(mean = c(5.89, 41.9, 20.8),
  cov = matrix(c(0.148, -0.679, 0.209,
                 -0.679, 96.10, 20.20,
                 0.209, 20.20, 147.24), nc = 3, byrow = TRUE),
  n = 13)
The data contains example data for testing the unequal variance option in the package. The dataset has four variables, wratr, wrata, treatment, and disability. treatment is the grouping variable and wratr and wrata are the responses. disability is not used.

References


Description

Plots the results from a permutation based version of Hotelling’s T-squared test for the difference in two multivariate sample means

Usage

```r
## S3 method for class 'hotelling.test'
plot(x, ...)
```

Arguments

- `x` an object of type hotelling.test
- `...` any additional arguments to be passed to the hist command

Details

This function only works if you have performed a permutation test. It will return an error message if not. It could be programmed to draw the relevant F distribution in the standard case, but this seems rather pointless.

Author(s)

James M. Curran
Examples

```r
data(bottle.df)
bottle.df = subset(bottle.df, Number == 1)
bottle.df$Number = rep(1:2,c(10,10))
fit = hotelling.test(~Number, bottle.df, perm = TRUE)
plot(fit)
plot(fit, col = "lightblue")
```

print.hotelling.test

Prints the results from a Hotelling’s T-squared test for the difference in two multivariate sample means

Description

Prints the test statistic, degrees of freedom and P-value from Hotelling’s T-squared test for the difference in two multivariate sample means

Usage

```r
## S3 method for class 'hotelling.test'
print(x, ...)
```

Arguments

- `x`: an object of type hotelling.test
- `...`: any additional arguments to be passed to the hist command

Author(s)

James M. Curran

Examples

```r
data(bottle.df)
bottle.df = subset(bottle.df, Number == 1)
bottle.df$Number = rep(1:2,c(10,10))
fit = hotelling.test(~Number, bottle.df, perm = TRUE)
fit
fit = hotelling.test(~Number, bottle.df)
fit

## an explicit call
print(fit)
```
### summarise

**Summary statistics for grouped data**

**Description**

Easily get summary statistics for each group present in the data

**Usage**

```r
summarise(x, ...)
```

**Arguments**

- `x` a matrix of multivariate observations, a list of summary statistics from multivariate observations, or a data.frame of multivariate observations, OR a data.frame of multivariate observations, OR a formula with a multivariate response on the right hand side, and a grouping variable/factor on the left hand side.
- `y` a matrix of multivariate observations, a list of summary statistics from multivariate observations, OR a data.frame of multivariate observations
- `stats` a named list of summary statistics to compute on each variable in each group. Note 1: Quantiles are not supported yet because I can’t think of a good way to handle the extra arguments. Help welcome. Note 2: The names of the elements in the list are used to label the columns of the output. They probably should be unique.
- `data` a data.frame containing the variables used in a formula
- `...` other arguments such as another matrix of multivariate observations: see summarise.default, or a data to be used with a formula: see summarise.formula
Methods (by class)

- default: Summary statistics for grouped data
- formula: Summary statistics for grouped data
- data.frame: Summary statistics for grouped data

Examples

data(container.df)
split.data = split(container.df[, -1], container.df$gp)
x = split.data[[1]]
y = split.data[[2]]
summarise(x, y)

## Using the formula interface

data(container.df)
summarise(gp~, data = container.df)

summarise(gp~Al+Ti, data = container.df)
Index

* datasets
  - bottle.df, 3
  - container.df, 4
  - manova1.df, 9
* htest
  - hotelling.stat, 5
  - hotelling.test, 6
* package
  - Hotelling-package, 2
* plot
  - plot.hotelling.test, 9
* print
  - print.hotelling.test, 10
* transformation
  - alr, 2
  - clr, 3
  
  alr, 2

  bottle.df, 3

  clr, 3
  container.df, 4

  hotel.stat (hotelling.stat), 5
  hotel.test (hotelling.test), 6
  Hotelling (Hotelling-package), 2
  Hotelling-package, 2
  hotelling.stat, 5
  hotelling.test, 6

  manova1.df, 9

  plot.hotelling.test, 9
  print.hotelling.test, 10

  summarise, 11
  summarize (summarise), 11