

Package ‘crossmatch’

July 2, 2014

Type Package

Title The Cross-match Test

Version 1.3-1

Date 2012-06-16

Author Ruth Heller <ruheller@post.tau.ac.il>, Dylan Small
<dsmall@wharton.upenn.edu>, Paul Rosenbaum <rosenbap@wharton.upenn.edu>

Maintainer Ruth Heller <ruheller@gmail.com>

Description This package performs a test for comparing two multivariate distributions by using the distance between observations. The input is a distance matrix and the labels of the two groups to be compared, the output is the number of cross-matches and a p-value.

Depends survival,nbpMatching

Suggests MASS

License GPL-2

LazyLoad yes

Repository CRAN

Date/Publication 2012-06-17 00:13:00

NeedsCompilation no

R topics documented:

| | |
|------------------------------|---|
| crossmatch-package | 2 |
| crossmatchdist | 3 |
| crossmatchtest | 4 |

| | |
|--------------|----------|
| Index | 6 |
|--------------|----------|

crossmatch-package *The Cross-Match Test For Comparing Two Multivariate Distributions.*

Description

The cross-match test is an exact, distribution free test of equality of 2 high dimensional multivariate distributions.

Details

Package: crossmatch
 Type: Package
 Version: 1.3-1
 Date: 2012-06-16
 License: GPL-2
 LazyLoad: yes

For the cross-match test, use the function [crossmatchtest](#).

Author(s)

Ruth Heller, Paul Rosenbaum, Dylan Small. Maintainer: Ruth Heller <ruheller@post.tau.ac.il>

References

Rosenbaum, P.R. (2005), An exact distribution-free test comparing two multivariate distributions based on adjacency, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **67**, 4, 515-530.

Examples

```
## The example in Section 2 of the article (see References)

#The data consists of 2 outcomes measured on 9 treated cases and 9 controls:
dat <- rbind(c(0.47,0.39,0.47,0.78,1,1,0.54,1,0.38,1,0.27,0.63,0.22,0,-1,-0.42,-1,-1),
            c(0.03,0.11,0.16,-0.1,-0.05,0.16,0.12,0.4,0.04,0.71,0.01,0.21,-0.18,-0.08
            , -0.35,0.26,-0.6
z <- c(rep(0,9),rep(1,9))
X <- t(dat)

## Rank based Mahalanobis distance between each pair:
X <- as.matrix(X)
n <- dim(X)[1]
k <- dim(X)[2]
for (j in 1:k) X[,j] <- rank(X[,j])
cv <- cov(X)
vuntied <- var(1:n)
```

```

rat <- sqrt(vuntied/diag(cv))
cv <- diag(rat)%*%cv%*%diag(rat)
out <- matrix(NA,n,n)

library(MASS)

icov <- ginv(cv)
for (i in 1:n) out[i,] <- mahalnobis(X,X[i,],icov,inverted=TRUE)

dis <- out

## The cross-match test:

crossmatchtest(z,dis)

```

| | |
|----------------|--|
| crossmatchdist | <i>The Exact Null Distribution Of The Cross-match Statistic Under The Null</i> |
|----------------|--|

Description

The exact null distribution of the number of crossmatches for $\text{bigN} \geq 4$ cases, $n \geq 2$ from one type and $N-n \geq 2$ from another type.

Usage

```
crossmatchdist(bigN, n)
```

Arguments

| | |
|------|-----------------------------------|
| bigN | The total number of observations |
| n | The number of cases from one type |

Details

bigN is even. Let a_1 be the number of cross-matches pairs. Then $a_2 = (n - a_1) / 2$ and $a_0 = \text{bigN} / 2 - (n + a_1) / 2$ are the number of pairs both of one type and the other type respectively.

Value

| | |
|------|---|
| dist | A matrix with rows a_0 , a_1 , a_2 , $\Pr(A_1 = a_1)$ and $\Pr(A_1 \leq a_1)$. |
|------|---|

Author(s)

Ruth Heller

References

Rosenbaum, P.R. (2005), An exact distribution-free test comparing two multivariate distributions based on adjacency, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **67**, 4, 515-530.

Examples

```
crossmatchdist(18,9)
```

| | |
|----------------|-----------------------------|
| crossmatchtest | <i>The Cross-Match Test</i> |
|----------------|-----------------------------|

Description

A test for comparing two multivariate distributions by using the distance between the observations.

Usage

```
crossmatchtest(z, D)
```

Arguments

| | |
|---|---|
| z | A binary vector corresponding to observations class labels. |
| D | A distance matrix of dimensions NxN, where N is the total number of observations. |

Details

Observations are divided into pairs to minimize the total distance within pairs, using a polynomial time algorithm made available in R by Lu, B., Greevy, R., Xu, X., and Beck, C in the R package "nbpMatching". The cross-match test takes as the test statistic the number of times a subject from one group was paired with a subject from another group, rejecting the hypothesis of equal distribution for small values of the statistic; see Rosenbaum (2005) for details.

Value

A list with the following

| | |
|------------|--|
| a1 | The number of cross-matches |
| Ea1 | The expected number of cross-matches under the null |
| Va1 | The variance of number of cross-matches under the null |
| dev | The observed difference from expectation under null in SE units |
| pval | The p-value based on exact null distribution (NA for datasets with 340 observations or more) |
| approxpval | The approximate p-value based on normal approximation |

Author(s)

Ruth Heller

References

Rosenbaum, P.R. (2005), An exact distribution-free test comparing two multivariate distributions based on adjacency, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **67**, 4, 515-530.

Examples

```
## The example in Section 2 of the article (see References)

#The data consists of 2 outcomes measured on 9 treated cases and 9 controls:
dat <- rbind(c(0.47,0.39,0.47,0.78,1,1,0.54,1,0.38,1,0.27,0.63,0.22,0,-1,-0.42,-1,-1),
            c(0.03,0.11,0.16,-0.1,-0.05,0.16,0.12,0.4,0.04,0.71,0.01,0.21,-0.18,-0.08
            , -0.35,0.26,-0.6

z <- c(rep(0,9),rep(1,9))
X <- t(dat)

## Rank based Mahalanobis distance between each pair:
X <- as.matrix(X)
n <- dim(X)[1]
k <- dim(X)[2]
for (j in 1:k) X[,j] <- rank(X[,j])
cv <- cov(X)
vuntied <- var(1:n)
rat <- sqrt(vuntied/diag(cv))
cv <- diag(rat)%*%cv%*%diag(rat)
out <- matrix(NA,n,n)

library(MASS)

icov <- ginv(cv)
for (i in 1:n) out[i,] <- mahalanobis(X,X[i,],icov,inverted=TRUE)

dis <- out

## The cross-match test:

crossmatchtest(z,dis)
```

Index

*Topic **multivariate**

crossmatch-package, 2

*Topic **nonparametric**

crossmatch-package, 2

crossmatch (crossmatch-package), 2

crossmatch-package, 2

crossmatchdist, 3

crossmatchtest, 2, 4