Readme for GHotelling(Data, Nboot)

PURPOSE

This program computes a generalized Hotelling T-squared test for differences between two groups, A and B. A unique feature is that an individual can contribute data for both groups, as for example if an individual provides data for a skin sample (group A) and for a saliva sample (group B). Correlations from such samples need to be taken into account, which is done by a bootstrap procedure to compute P-values.

ARGUMENTS

Nboot is the number of bootstrap replicates used for calculating p-values. Default Nboot=10,000.

Data is a data frame with *n* rows that correspond to individuals who contribute data. For a given row, columns 1 and 2 of Data contain dA,stool and dA,nasal respectively, which are the mean distances from the group A sample to the Human Microbiome Project (HMP) reference sets of 92 stool samples and 74 nasal samples, respectively. Columns 3 and 4 contain dB,stool and dB,nasal respectively, which are the mean distances from the group B sample to the HMP stool and nasal reference sets. Missing data are indicated by NA. Here are some examples with comments.

(1,2,3,4) This individual contributed two distances from group A and two distances from group B. This individual will contribute to three analyses. (1) One (T2\_1)is a generalized squared t-test to see if mean distances to the HMP reference stool samples differ between groups A and B. (2) Another (T2\_2) is a generalized squared t-test to see if mean distances to the HMP reference nasal samples differ between groups A and B. (3) The third (T2\_12) is a generalized Hotelling’s T-squared test to see if the means of the vectors (dA,stool , dA,nasal) differ from the means of vectors (dB,stool , dB,nasal).

Only the following missingness patterns are permitted. The following missingness patterns are associated with analyses to which they can contribute:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Pattern | | | | Analyses to which pattern contributes |
| 1 | 2 | 3 | 4 | 1(ignore cols 2,4),2(ignore cols 1,3),3 |
| 1 | 2 | NA | NA | 1(ignore cols 2,3,4),2(ignore cols 1,3,4),3(ignore cols 3,4) |
| NA | NA | 3 | 4 | 1(ignore cols 1,2,4),2(ignore cols 1,2,3)3(ignore cols 1,2) |

OUTPUT

Original Hotelling’s like Statistics for tests 1,2, and 3

Corresponding p-values (proportion of bootstraps in which the original statistic was less than or equal to the (null) bootstrap statistic.

Nboot

Full statistics for the analysis of the original data

EXAMPLES

First compile or load the functions GHotelling\_function=function(data,nBoot), T2forAB=function(DA,DB,DAB,T2only), and T2withnoAB=function(DA,DB,T2only). These functions are included.

Then run the code

load("test-mx.rdata") #This is test data  
data=test.mx #These data (90x 4) include some individuals who contribute to groups A and B  
nBoot=10000  
dim(data)  
GHotelling\_function(data,nBoot)

> load("test-mx.rdata")

> data=test.mx

> nBoot=10000

> dim(data)

[1] 90 4

> GHotelling\_function(data,nBoot)

Output  
[[1]]

[1] 0 #phat\_12

[[2]]

[1] 0.0395 #phat\_1

[[3]]

[1] 0 #phat\_2

[[4]]

[1] 72.54674 #T2\_12 original Hotelling-like statistic

[[5]]

[1] 4.44436 #T2\_1

[[6]]

[1] 72.33653 #T2\_2

[[7]]

[1] 10000 #nBoot

[[8]]

[[8]][[1]] #output from out1, a list

[1] 0.7362909 0.4758886

[[8]][[2]]

[1] 49

[[8]][[3]]

[1] 0.6843044 0.6902029

[[8]][[4]]

[1] 26

[[8]][[5]]

[1] 0.7270853 0.4975778 0.6980028 0.7074237

[[8]][[6]]

[1] 15

[[8]][[7]]

[1] 0.7341333 0.4809720

[[8]][[8]]

[1] 0.6893160 0.6965032

[[8]][[9]]

[,1] [,2] [,3] [,4]

[1,] 2.287032e-02 -0.0078611567 7.085304e-05 0.0001203015

[2,] -7.861157e-03 0.0341147880 3.117209e-04 0.0003633302

[3,] 7.085304e-05 0.0003117209 3.911512e-03 0.0008310508

[4,] 1.203015e-04 0.0003633302 8.310508e-04 0.0046452507

[[8]][[10]]

[,1] [,2] [,3] [,4]

[1,] 3.573488e-04 -1.228306e-04 4.050288e-07 6.876989e-07

[2,] -1.228306e-04 5.330436e-04 1.781941e-06 2.076964e-06

[3,] 4.050288e-07 6.876989e-07 9.540274e-05 2.026953e-05

[4,] 1.781941e-06 2.076964e-06 2.026953e-05 1.132988e-04

[[8]][[11]]

[,1]

[1,] 0.0448173

[2,] -0.2155312

[[8]][[12]]

[,1]

[1,] 0.0448173

[[8]][[13]]

[,1]

[1,] -0.2155312

[[8]][[14]]

[,1] [,2] [,3] [,4]

[1,] 1 0 -1 0

[2,] 0 1 0 -1

[[8]][[15]]

[,1] [,2] [,3] [,4]

[1,] 1 0 -1 0

[[8]][[16]]

[,1] [,2] [,3] [,4]

[1,] 0 1 0 -1

[[8]][[17]]

[,1]

[1,] 72.54674

[[8]][[18]]

[,1]

[1,] 4.44436

[[8]][[19]]

[,1]

[1,] 72.33653

The second example is a reanalysis of the data, but with individuals who contribute to both

groups A and B excluded

load("test-mx.rdata") #This is test data  
data=test.mx #These data (90x 4) include some individuals who contribute to groups A and B  
data=data[-c(50:64),] #Exclude those who contribute to both A and B  
nBoot=10000  
dim(data)  
GHotelling\_function(data,nBoot)

Output

|  |
| --- |
| > load("test-mx.rdata")  > data=test.mx  > data=data[-c(50:64),]  > nBoot=10000  > dim(data)  [1] 75 4  > GHotelling\_function(data,nBoot)  [[1]]  [1] 0  [[2]]  [1] 0.0574  [[3]]  [1] 0  [[4]]  [1] 46.90417  [[5]]  [1] 3.633159  [[6]]  [1] 45.92472  [[7]]  [1] 10000  [[8]]  [[8]][[1]]  [1] 0.7362909 0.4758886  [[8]][[2]]  [1] 49  [[8]][[3]]  [1] 0.6843044 0.6902029  [[8]][[4]]  [1] 26  [[8]][[5]]  [1] 0.7362909 0.4758886  [[8]][[6]]  [1] 0.6843044 0.6902029  [[8]][[7]]  [,1] [,2] [,3] [,4]  [1,] 0.024837616 -0.008028944 0.000000000 0.000000000  [2,] -0.008028944 0.035360785 0.000000000 0.000000000  [3,] 0.000000000 0.000000000 0.006161458 0.001196945  [4,] 0.000000000 0.000000000 0.001196945 0.007240466  [[8]][[8]]  [,1] [,2] [,3] [,4]  [1,] 0.0005068901 -0.0001638560 0.000000e+00 0.000000e+00  [2,] -0.0001638560 0.0007216487 0.000000e+00 0.000000e+00  [3,] 0.0000000000 0.0000000000 2.369792e-04 4.603635e-05  [4,] 0.0000000000 0.0000000000 4.603635e-05 2.784795e-04  [[8]][[9]]  [,1]  [1,] 0.05198649  [2,] -0.21431427  [[8]][[10]]  [,1]  [1,] 0.05198649  [[8]][[11]]  [,1]  [1,] -0.2143143  [[8]][[12]]  [,1] [,2] [,3] [,4]  [1,] 1 0 -1 0  [2,] 0 1 0 -1  [[8]][[13]]  [,1] [,2] [,3] [,4]  [1,] 1 0 -1 0  [[8]][[14]]  [,1] [,2] [,3] [,4]  [1,] 0 1 0 -1  [[8]][[15]]  [,1]  [1,] 46.90417  [[8]][[16]]  [,1]  [1,] 3.633159  [[8]][[17]]  [,1]  [1,] 45.92472 |
|  |
| |  | | --- | |  | |

Listing of “data”

|  |
| --- |
| > data  [,1] [,2] [,3] [,4]  [1,] 0.8891262 0.5393743 NA NA  [2,] 0.6086559 0.4926765 NA NA  [3,] 0.5576324 0.4230861 NA NA  [4,] 0.6202238 0.5801520 NA NA  [5,] 0.6716623 0.6808685 NA NA  [6,] 0.9085683 0.5960691 NA NA  [7,] 0.9169964 0.3599369 NA NA  [8,] 0.6318522 0.5994725 NA NA  [9,] 0.5891847 0.4468154 NA NA  [10,] 0.8355427 0.7454346 NA NA  [11,] 0.8468401 0.6815416 NA NA  [12,] 0.8048558 0.2573758 NA NA  [13,] 0.7041567 0.3238072 NA NA  [14,] 0.7435672 0.3123777 NA NA  [15,] 0.8090252 0.2613502 NA NA  [16,] 0.3434949 0.9820981 NA NA  [17,] 0.6419931 0.5487027 NA NA  [18,] 0.7988042 0.2947895 NA NA  [19,] 0.7454450 0.3470540 NA NA  [20,] 0.8602455 0.3025820 NA NA  [21,] 0.5106520 0.4373603 NA NA  [22,] 0.6999326 0.2767769 NA NA  [23,] 0.7062435 0.2566572 NA NA  [24,] 0.7491845 0.5446015 NA NA  [25,] 0.6817700 0.3880376 NA NA  [26,] 0.7533255 0.4715248 NA NA  [27,] 0.6771983 0.4374261 NA NA  [28,] 0.8954557 0.3227732 NA NA  [29,] 0.7109037 0.3040642 NA NA  [30,] 0.9730085 0.4005991 NA NA  [31,] 0.7293814 0.2519109 NA NA  [32,] 0.9160121 0.3333571 NA NA  [33,] 0.8383893 0.2861348 NA NA  [34,] 0.8422147 0.3048039 NA NA  [35,] 0.4913181 0.4111496 NA NA  [36,] 0.7052787 0.7173436 NA NA  [37,] 0.9551113 0.3846980 NA NA  [38,] 0.7052787 0.7173436 NA NA  [39,] 0.1913229 0.7549907 NA NA  [40,] 0.7052787 0.7173436 NA NA  [41,] 0.9969031 0.4266393 NA NA  [42,] 0.7052787 0.7173436 NA NA  [43,] 0.9992160 0.9984026 NA NA  [44,] 0.7052787 0.7173436 NA NA  [45,] 0.9372732 0.3517056 NA NA  [46,] 0.6815404 0.3957496 NA NA  [47,] 0.6760257 0.3953155 NA NA  [48,] 0.6866441 0.4429350 NA NA  [49,] 0.7249606 0.3786472 NA NA  [50,] 0.6400566 0.6511278 0.7029759 0.7136018  [51,] 0.8329836 0.7450835 0.7032013 0.7143167  [52,] 0.8267445 0.4231805 0.6972365 0.7072342  [53,] 0.7982652 0.5997181 0.6858586 0.6842543  [54,] 0.7613881 0.3981847 0.6935702 0.7019639  [55,] 0.6508773 0.5704893 0.6964355 0.7114471  [56,] 0.5885381 0.3855215 0.6826419 0.6759895  [57,] 0.7573890 0.3721890 0.7043901 0.7164105  [58,] 0.7075599 0.3133640 0.6919188 0.7052712  [59,] 0.7439767 0.2644019 0.7009846 0.7124965  [60,] 0.8257555 0.2812157 0.7024527 0.7146073  [61,] 0.8757406 0.3295845 0.6997325 0.7105801  [62,] 0.7962656 0.7168541 0.7025694 0.7131337  [63,] 0.3417755 0.7489067 0.7027402 0.7149955  [64,] 0.7589627 0.6638452 0.7033342 0.7150532  [65,] NA NA 0.7038382 0.7160411  [66,] NA NA 0.7032732 0.7154599  [67,] NA NA 0.6963844 0.7058613  [68,] NA NA 0.6807999 0.6775183  [69,] NA NA 0.6881731 0.6922054  [70,] NA NA 0.6918215 0.6970938  [71,] NA NA 0.6981244 0.7082174  [72,] NA NA 0.7052787 0.7173436  [73,] NA NA 0.7046314 0.7166264  [74,] NA NA 0.6966688 0.7083345  [75,] NA NA 0.7041139 0.7160756  [76,] NA NA 0.7046650 0.7167093  [77,] NA NA 0.7050647 0.7170528  [78,] NA NA 0.6889059 0.7027948  [79,] NA NA 0.6963772 0.7050173  [80,] NA NA 0.7046416 0.7170741  [81,] NA NA 0.7033849 0.7152708  [82,] NA NA 0.3008636 0.6185371  [83,] NA NA 0.7048675 0.7169501  [84,] NA NA 0.7052787 0.7173436  [85,] NA NA 0.7046184 0.7164899  [86,] NA NA 0.6894985 0.7002000  [87,] NA NA 0.7041585 0.7160084  [88,] NA NA 0.7001885 0.2851486  [89,] NA NA 0.7030232 0.7145864  [90,] NA NA 0.7032706 0.7153154 |
|  |
| |  | | --- | |  | |