Package 'SubCultCon'

February 19, 2015

Type Package

Title Maximum-Likelihood Cultural Consensus Analysis with Sub-Cultures

Version 1.0

Date 2013-09-23

Author Mary C Meyer, Jeffrey G Snodgrass, Michael Lacy

Maintainer Mary C Meyer <meyer@stat.colostate.edu>

Description The three functions in the package compute the maximum likelihood estimates of the informants' competence scores, tests for two answer keys with known groups, and finds ``best'' split of the informants into sub-culture groups.

License GPL-2 | GPL-3

Depends graphics, grDevices, stats, utils

NeedsCompilation no

Repository CRAN

Date/Publication 2013-09-25 18:38:35

R topics documented:

1

SubCultCon–package   *Subcultural Consensus Modeling*

Description

The three functions in the package compute the maximum likelihood estimates of the informants' competence scores, test for two answer keys with known groups, and find best split of the informants into sub-culture groups.

Details

Package:	SubCultCon
Type:	Package
Version:	1.0
Date:	2013-09-25
License:	GPL-2 or GPL-3

Author(s)

Mary C Meyer, Jeffrey G Snodgrass, Michael Lacy

Maintainer: Mary C Meyer <meyer@stat.colostate.edu>

References

Kimball A. Romney, Susan C. Weller, and William H. Batchelder (1986). Culture as Consensus: A Theory of Culture and Informant Accuracy. *American Anthropologist*, **88(2)**, 313-338.

See Also

ccgrouptest, GAmaxcomp, ccmle, contagion, , village

---

ccgrouptest	*Function to test the null hypothesis that the two groups have the same answer key, versus the alternative hypothesis that the two groups have answer keys that differ for at least one question.*

Description

Given group assignments and an answer matrix, the maximum likelihood estimates of cultural consensus are found for the null and alternative hypothesis. The test statistic is the difference in the sums of the competence scores. The null distribution is simulated, and the approximate p-value is the proportion of the simulated distribution that is larger than the observed statistic.

Usage

```
ccgrouptest(answermat, group)
```

Arguments

answermat
: A matrix of informant answers to a fixed set of questions. The matrix should have n rows and m columns, where n is the number of informants and m is the number of questions. The element in row i and column j should be the answer provided to question j, by informant i. All questions should have the same number of possible responses.

group
: A vector of length n containing 1s and 2s, indicating the group for the ith informant (according to the rows of answermat)

Details

First, the function `ccmle` is called using the entire answer matrix, to get the one-answer-key solution. Next, the function is called for the two groups separately, to get the two-answer-key solution. If the answer keys for the two groups are identical, then the null hypothesis is accepted and the p-value is set to 1. Otherwise, let CSUM1 be the sum of the competence scores for the one-answer-key solution, and let CSUM2 be the sum of the competence scores for the two-answer-key solution. The idea is that if CSUM2 is considerably larger than CSUM1, this is evidence that the two-answer-key solution is "better." The test statistic is CDIFF=CSUM2-CSUM1. To obtain the simulated null distribution, we use the answer key and competence scores from the one-answer-key solution, and simulate 1000 answer matrices under these assumptions. For each, we compute a test statistic SDIFF(1),....SDIFF(1000). The p-value is the fraction of SDIFF values that are larger than CDIFF.

Value

pval
: The p-value for the test

key1
: The answer key for the first group, for the two-answer-keys solution

key2
: The answer key for the second group, for the two-answer-keys solution

comp1
: The competence scores for the one-answer-key solution

comp2
: The competence scores for the two-answer-keys solution

diff
: The test statistic – sum(comp2)-sum(comp1)

simdist
: The 1000 values of the simulated distribution for the test statistic

Author(s)

Mary C Meyer

See Also

[ccmle](), [GAmaxcomp]()

Examples

```
## example with simulated data for 9 informants answering 7 questions
## there are 4 possible answers per question
## there are two subgroups with answer keys that differ in 3 questions
## five informants in group 1 and four in group 2
n=9
m=7
nl=4
group=c(1,1,1,1,1,2,2,2,2)
key=matrix(nrow=m,ncol=2)
key[,1]=trunc(runif(m)*nl+1)
key[,2]=key[,1];key[5:7,2]=5-key[5:7,1]
answermat=matrix(nrow=n,ncol=m)
comp=round(rbeta(n,3,1),4)
for(i in 1:n){for(j in 1:m){
if(runif(1)<comp[i]){
answermat[i,j]=key[j,group[i]]
}else{answermat[i,j]=trunc(runif(1)*nl)+1}
}}
ans=ccgrouptest(answermat,group)
rng=c(min(ans$simdist),max(max(ans$simdist),ans$diff))
par(mar=c(3,4,1,1))
 hist(ans$simdist,br=0:20/20*(rng[2]-rng[1])+rng[1],main="Simulated Null Distribution")
points(ans$diff,0,pch="X",cex=1.3,col=2)
ans$pval  ## here is the p-value
###for a longer-running example, un-comment
#data(village)
#ans=ccgrouptest(village$answermat,village$group) ## takes a few minutes to simulate distribution
#par(mar=c(3,4,3,1))
#hist(ans$simdist,br=0:50/50*(ans$diff-min(ans$simdist))+min(ans$simdist),
#main="simulated distribution of test statistic
#observed value is X")
#points(ans$diff,0,pch="X",cex=1.2,col=2)
#ans$pval  # the computed p-value is zero because the observed test statistic
#                 #  is larger than all simulated values
```

---

ccmle   *Maximum Likelihood Estimation of Cultural Competence Scores*

Description

Given a matrix of answers, the function returns the maximum-likelihood estimates of the informants' competence scores, and an answer key estimated using the matrix of answers, weighted by the competence scores.

Usage

```
ccmle(answermat)
```

Arguments

answermat   A matrix of informant answers to a fixed set of questions. The matrix should have n rows and m columns, where n is the number of informants and m is the number of questions. The element in row i and column j should be the answer provided to question j, by informant i. All questions should have the same number of possible responses.

Details

Each question in `answermat` should have the same number of possible responses.

Value

comp   Maximum-likelihood estimates of the competence scores

key   Estimate of the answer key

conv   Indication of convergence – a value of 0 means convergence; anything else means failure

val   The value of the negative log likelihood function at the solution

Author(s)

Mary C Meyer, Professor, Statistics Department, Colorado State University.

References

Kimball A. Romney, Susan C. Weller, and William H. Batchelder (1986). Culture as Consensus: A Theory of Culture and Informant Accuract. *American Anthropologist* **88(2)**

See Also

[ccgrouptest](), [GAmaxcomp]()

Examples

```
data(contagion)
## Compute the competence scores and the answer key for the contagion data
mlescores=ccmle(contagion$answermat)
## make a histogram of the competence scores
hist(mlescores$comp,main="Competency Scores for the Informants",br=28:50/50)
## look at the answer key
mlescores$key
```

---

contagion	*Guatemalen Women's Perceptions of Contagion in Disease*

Description

This data set was collected by Weller (1983,1984) and used in Romney, et al. (1986). Twenty-four urban Guatemalen women were asked about the role of contagion in 27 diseases.

Usage

```
data(contagion)
```

Format

There are two objects:

questions: list of diseases

answermat: 24 by 27 matrix of responses

Source

Susan C. Weller (1983) New Data on Intracultural Variability: The Hot-cold Concept of Medicine and Illness. *Human Organization* **42(3)**: 249-257.

References

Kimball A. Romney, Susan C. Weller, and William H. Batchelder (1986). Culture as Consensus: A Theory of Culture and Informant Accuract. *American Anthropologist* **88(2)**

Susan C. Weller (1984) Cross Cultural Concepts of Illness: Variation and Validation. *American Anthropologist* **86(2)**:341-351.

Examples

```
data(contagion)
contagion$questions
contagion$answermat
```

GAmaxcomp	*Determines best assignment of informants into subcultures based on responses to cultural queries.*

Description

Given an answer matrix and a number of subcultural groups, a genetic algorithm is employed to determine which assignment of informants to subgroups maximizes the overall competence score. That is, the maximum-likelihood estimates of competence scores are summed over all informants; this is the criterion for best assignment.

Usage

```
GAmaxcomp(xmat, ng, npop, ngen)
```

Arguments

xmat	A matrix of informant answers to a fixed set of questions. The matrix should have n rows and m columns, where n is the number of informants and m is the number of questions. The element in row i and column j should be the answer provided to question j, by informant i. All questions should have the same number of possible responses.
ng	The number of subgroups. Should be at least two. The code gets slower for larger groups, perhaps four or five is a maximum.
npop	The size of the "population of phenotypes" for the genetic algorithm. This should be in the hundreds or thousands depending on the size of the data set. About 2*n*ng is reasonable.
ngen	The maximum number of "generations" for the genetic algorithm. ngen=100 should be enough unless your data set is quite large

Details

There is no way to guarantee or check convergence of the genetic algorithm. Typically if the code is started several times with different random seeds, and the same solution is obtained, this is a good indication that the optimal solution is reached. If there are different solution, try a larger value of npop. The algorithm can take a long time for large data sets, so the "best" solution for each generation is printed, enabling the user to track the progress. Also printed are the current maximum total competence score (TCS) and the 75th percentile. The algorithm stops when these two values are quite close to each other.

Value

bgrp	The assignment that maximizes the sum of the competence scores
compsum	The sum of the competence scores for the best assignment
comp	The individual competence scores for the best assignment

keymat	The matrix of answer keys, where the ng columns contain the keys for the sub-groups
numgen	The number of generations before the stopping criterion was reached. If this is ngen, the stopping criterion may not have been met. Choose a larger npop and ngen.

Author(s)

Mary C Meyer, Professor, Statistics Department, Colorado State University

See Also

ccgrouptest, ccmle

Examples

```
## example with simulated data for 9 informants answering 7 questions
## there are 4 possible answers per question
## there are two subgroups with answer keys that differ in 3 questions
## five informants in group 1 and four in group 2
n=9
m=7
nl=4
group=c(1,1,1,1,1,2,2,2,2)
key=matrix(nrow=m,ncol=2)
key[,1]=trunc(runif(m)*nl+1)
key[,2]=key[,1];key[5:7,2]=5-key[5:7,1]
answermat=matrix(nrow=n,ncol=m)
comp=round(rbeta(n,3,1),4)
for(i in 1:n){for(j in 1:m){
if(runif(1)<comp[i]){
answermat[i,j]=key[j,group[i]]
}else{answermat[i,j]=trunc(runif(1)*nl)+1}
}}
ans=GAmaxcomp(answermat,2,100,100)
ans$keymat
##########
## for an example that takes longer to run:  (uncomment)
#data(contagion)
#ans=GAmaxcomp(contagion$answermat,2,400,100) ## get best two-group solution
#ans$keymat    ## these are the answer keys
#ans$bgrp         ## this is the best assignment
##########
```

---

village   *Sahariya Village Quality*

Description

Informants from Sahariya villages in the Indian state of Madhya Pradesh responded to fourteen queries as to their perceptions of social and material well-being and village quality.

Usage

```
data(village)
```

Format

There are three objects:

group: indicator for which village the informant comes from

questions: topics for the queries

answermat: 157 by 14 matrix of answers

Details

Eighty of the informants came from a village which remains in its traditional location, adjacent to a relatively rich forest. Seventy-seven came a similar village which was relocated.

Source

Collected by Jeffrey G. Snodgrass and his research team in Madhya Pradesh, India Spring 2011, in a project funded by the National Science Foundation, Environmental Displacement and Human Resilience: New Explanations Using Data from Central India. PI is Jeffrey G. Snodgrass, Co-PI is Sammy Zahran. 2011-14.

Examples

```
data(village)
village$questions
##
# uncomment to do an hypothesis test of H_0: same answer key for both villages, versus
#                                        H_1: villages have different answer keys
#ans=ccgrouptest(village$answermat,village$group) ## takes a few minutes to simulate distribution
#par(mar=c(3,4,3,1))
#hist(ans$simdist,br=0:50/50*(ans$diff-min(ans$simdist))+min(ans$simdist),
#main="simulated distribution of test statistic
#observed value is X")
#points(ans$diff,0,pch="X",cex=1.2,col=2)
#ans$pval  # the computed p-value is zero because the observed test statistic
#                 #  is larger than all simulated values
```

Index