Package ‘DIFplus’

March 20, 2020

Title Multilevel Mantel-Haenszel Statistics for Differential Item Functioning Detection

Version 1.1

Author Shenghai Dai [aut, cre], Brian F. French [aut], W. Holmes Finch [aut], Andrew Iverson [aut]

Maintainer Shenghai Dai <s.dai@wsu.edu>

Description

Depends R (>= 3.5.0), stats
Imports TestDataImputation, plyr
NeedsCompilation no
Encoding UTF-8
LazyData true
License GPL (>= 2)
RoxygenNote 7.0.2
Repository CRAN
Date/Publication 2020-03-20 17:10:06 UTC

R topics documented:

<table>
<thead>
<tr>
<th>R topics</th>
<th>R topics documented:</th>
</tr>
</thead>
<tbody>
<tr>
<td>ContingencyTables</td>
<td>ContingencyTables</td>
</tr>
<tr>
<td>data.adult</td>
<td>data.adult</td>
</tr>
<tr>
<td>data.adult.revised</td>
<td>data.adult.revised</td>
</tr>
<tr>
<td>data.ordinal</td>
<td>data.ordinal</td>
</tr>
<tr>
<td>ML.DIF</td>
<td>ML.DIF</td>
</tr>
</tbody>
</table>

2 3 4 5 6
ContigencyTables

Description

This function creates contingency tables by strata for each item. Both dichotomous and polytomous item responses are allowed. It also handles missing responses and returns a cleaned data set with no missing data.

Usage

ContigencyTables (Response.data, Response.code=c(0,1),
Group, group.names=NULL, Stratum=NULL, Cluster=NULL,
missing.code="NA", missing.impute="LW", print.information=TRUE)

Arguments

Response.data A scored item responses matrix in the form of matrix or data frame. This matrix should not include any other variables (group, stratum, cluster, etc.).
Response.code A numerical vector of all possible item responses. By default, Response.code=c(0,1).
Group The variable of group membership (e.g., gender). Its length should be equal to the sample size of the item response matrix.
group.names Names for each defined group (e.g., c('Male','Female')). This argument is optional. By default, group.names=NULL. If not provided, group names of "Group.1, Group.2, etc." will be automatically generated.
Stratum The matching variable. By default, Stratum=NULL. If not provided, the observed total score will be used.
Cluster The cluster variable. Its length should be equal to the sample size of the item response matrix. By default, Cluster=NULL. This variable will not be used to generate contingency tables. It will be included in the returned data set for DIF analysis.
missing.code Indication of how missing values were defined in the data. By default, missing.code="NA".
missing.impute The approach selected to handle missing item responses. By default, missing.impute="LW", indicating the list-wise deletion will be used. Other options include: "PM" (person mean or row mean imputation), "IM" (item mean or column mean imputation), "TW" (two-way imputation), "LR" (logistic regression imputation), and EM (EM imputation). Check the package "TestDataImputation" (https://cran.r-project.org/package=TestDataImputation) for more details.
Note. If any missing data are detected on group, cluster, or stratum variables, listwise deletion will be used before handling missing item responses.
print.information Indicator of whether function running information is printed on screen. By default, print.information=TRUE.
Details

This function creates contingency tables.

Value

A list of strata statistics, contingency tables, etc.

Strata.stats Summary statistics for each item: n.valid.strata, n.valid.category, and also sample sizes for each stratum across items.

c.table.list.all A list that contains all contingency tables across items and strata.

c.table.list.valid A list that contains only valid contingency tables across items and strata. Strata that have missing item response categories or zero marginal means are removed.

data.out A cleaned data set with variables "Group", "Group.factor","Cluster", "Stratum", and all item responses (with missing data handled).

Examples

#Specify the item responses matrix
data(data.adult)
Response.data<-data.adult[,2:13]
#Run the function with specifications
c.table.out<-ContigencyTables(Response.data, Response.code=c(0,1),
                                  Group=data.adult$Group, group.names=NULL,
                                  Stratum=NULL, Cluster=NULL, missing.code="NA",
                                  missing.impute= "LW",print.information = TRUE)

#Obtain results
c.tables.all<-c.table.out$c.table.list.all
c.tables.valid<-c.table.out$c.table.list.valid
c.table.out$Strata.stats
data.use<-c.table.out$data.out

Description

This data example contains binary (0/1) responses of 684 participants to 12 items. Participants were classified into 34 clusters and 2 groups.

Usage

data("data.adult")
Format

A data frame with 684 observations on the following 14 variables.

Cluster  The cluster variable
I1  Item 1
I2  Item 2
I3  Item 3
I4  Item 4
I5  Item 5
I6  Item 6
I7  Item 7
I8  Item 8
I9  Item 9
I10  Item 10
I11  Item 11
I12  Item 12

Group  Binary group membership variable

Details

A data set with 14 variables: (1) binary (0/1) responses of 684 participants to 12 items; (2) a cluster indicator variable; and (3) a group indicator variable.

Examples

```r
data(data.adult)
## maybe str(data.adult) ; plot(data.adult) ...
```

```r
data.adult.revised  Modified data.adult by removing all strata with zero marginal means.
```

Description

This data example contains binary (0/1) responses of 684 participants to 12 items. Participants were classified into 10 clusters, 2 groups, and 3 strata.

Usage

```r
data("data.adult.revised")
```
**Format**

A data frame with 684 observations on the following 15 variables.

Cluster  The cluster variable
I1  Item 1
I2  Item 2
I3  Item 3
I4  Item 4
I5  Item 5
I6  Item 6
I7  Item 7
I8  Item 8
I9  Item 9
I10  Item 10
I11  Item 11
I12  Item 12

Group  Binary group membership variable
Stratum  A prespecified matching variable with three levels

**Details**

A data set with 15 variables: (1) binary (0/1) responses of 684 participants to 12 items; (2) a cluster indicator variable with ten levels; (3) a group indicator variable with two levels; and (4) a stratum variable with three levels.

**Examples**

```r
data(data.adult.revised)
## maybe str(data.adult.revised) ; plot(data.adult.revised) ...
```

---

**Description**

This data example contains ordinal (1/2/3/4) responses of 300 participants to 5 items. Participants were classified into 6 clusters and 2 groups.

**Usage**

data("data.ordinal")
Format

A data frame with 300 observations on the following 7 variables.

Group  Group membership
Cluster Cluster membership
I1  Item 1
I2  Item 2
I3  Item 3
I4  Item 4
I5  Item 5

Details

A data set with 7 variables: (1) ordinal (1/2/3/4) responses of 300 participants to 5 items; (2) a cluster indicator variable with six levels; and (3) a group indicator variable with two levels.

Examples

data(data.ordinal)
## maybe str(data.ordinal) ; plot(data.ordinal) ...

ML.DIF  Main function to compute adjusted Mantel-Haenszel statistics

Description


Usage

ML.DIF (Response.data, Response.code=c(0,1), Cluster, Group, group.names=NULL, Stratum=NULL, correct.factor=0.85, missing.code="NA", missing.impute="LW", anchor.items=NULL, purification=FALSE, max.iter=10, alpha = .05)

Arguments

Response.data  A scored item responses matrix in the form of matrix or data frame. This matrix should not include any other variables (group, stratum, cluster, etc.).
Response.code  A numerical vector of all possible item responses. By default, Response.code=c(0,1).
Cluster  The cluster variable. Its length should be equal to the sample size of the item response matrix.
**Group**  
The variable of group membership (e.g., gender). Its length should be equal to the sample size of the item response matrix.

**group.names**  
Names for each defined group (e.g., c('Male','Female')). This argument is optional. By default, group.names=NULL. If not provided, group names of "Group.1, Group.2, etc." will be automatically generated.

**Stratum**  
The matching variable. By default, Stratum=NULL. If not provided, the observed total score will be used.

**correct.factor**  
The value of adjustment applied to the adjusted MH statistic (i.e., f). The default value used here is .85. The adjusted MH statistic was found to exhibit low statistical power for DIF detection in some conditions. One solution to this is to reduce the magnitude of f through multiplying it by the correct factor (e.g., .85, .90, .95). The value of .85 is suggested by French & Finch (2013) <doi:10.1177/0013164412472341>.

**missing.code**  
Indication of how missing values were defined in the data. By default, missing.code="NA".

**missing.impute**  
The approach selected to handle missing item responses. By default, missing.impute="LW", indicating the list-wise deletion will be used. Other options include: "PM" (person mean or row mean imputation), "IM" (item mean or column mean imputation), "TW" (two-way imputation), "LR" (logistic regression imputation), and EM (EM imputation). Check the package "TestDataImputation" (<https://cran.r-project.org/package=TestDataImputation>) for more details.

**anchor.items**  
A scored item responses matrix of selected anchor items. This matrix should be a subset of the response data matrix specified above. By default, anchor.items=NULL.

**purification**  
True or false argument, indicating whether purification will be used. By default, purification=FALSE.
Note: Purification will not be applied if anchor items are specified and/or the matching variable is defined.

**max.iter**  
The maximum number of iterations for purification. The default value is 10.

**alpha**  
The alpha value used to decide on the DIF items. The default value is .05.

**Details**

This main function computes both unadjusted and adjusted Mantel-Haenszel statistics in the presence of multilevel data.

**Value**

A list of MH statistics, contingency tables, etc.

**MH.values**  
Summary of estimated MH statistics and corresponding p-values. Specifically,
* MH.unadj is the unadjusted MH test statistic.
* MH.score is the MH statistic based on working score test (Begg, 1999).
* MH.GMH is the MH test statistic based on Holland & Thayer's (1998) formula.
* MH.Yates is the MH.GMH statistic with Yates’ correction.
* MH.adj is the adjusted MH statistic for clustered data;
* f.adj is the adjustment value based on Begg (1999).
* f.adj.correct is the product of f and the correction factor (.85, etc.).
* DIF.Item (Yes) = 1 indicates the item is flagged as a DIF item;
* N.Valid, N.Strata, and N.Cluster refer to the sample size, number of valid stata
  and cluster that are used in the analysis.

Stratum.statistics
summary statistics for each item: n.valid.strata, n.valid.category, and also sam-
ple sizes for each stratum across items.

c.table.list.all
A list that contains all contingency tables across items and strata.

c.table.list.valid
A list that contains only valid contingency tables across items and strata. Strata
that have missign item response categories or zero marginal means are removed.

data.out A cleaned data set with variables "Group", "Group.factor","Cluster", "Stratum",
and all item responses (with missing data handled).

References
doi:10.1080/00949650108812115.
procedure." In H. Wainer & H. I. Braun (Eds.), Test validity (pp.129-145). Lawrence Erlbaum
Associates, Inc.

Examples
#Specify the item responses matrix
data(data.adult)
Response.data<-data.adult[,2:13]
#Run the function with specifications
ML.DIF.out<-ML.DIF (Response.data, Response.code=c(0,1),Cluster=data.adult$Cluster,
Group=data.adult$Group, group.names=c('Reference', 'Focal'),
Stratum=NULL, correct.factor=0.85,
missing.code="NA", missing.impute="LW",
anchor.items=NULL, purification=FALSE,
max.iter=10, alpha = .05)
#Obtain results
ML.DIF.out$MH.values
ML.DIF.out$Stratum.statistics
Index

*Topic datasets
  data.adult, 3
  data.adult.revised, 4
  data.ordinal, 5

ContigencyTables, 2

data.adult, 3
data.adult.revised, 4
data.ordinal, 5

ML.DIF, 6