Package ‘htestClust’

February 13, 2022

Title  Reweighted Marginal Hypothesis Tests for Clustered Data
Version 0.2.1
Description A collection of reweighted marginal hypothesis tests for clustered data, based
0420.00005>. The tests in this collection are clustered analogs to well-known hypothesis tests
in the classical setting, and are appropriate for data with cluster- and/or group-size
informativeness. The syntax and output of functions are modeled after common, recognizable functions native to R. Methods used in the package refer to

Depends R (>= 3.5.0)
License MIT + file LICENSE
Encoding UTF-8
LazyData true
RoxygenNote 7.1.0
Imports bootstrap, graphics, MASS, stats
NeedsCompilation no
Author Mary Gregg [aut, cre] (<https://orcid.org/0000-0003-2991-6939>),
Somnath Datta [aut] (<https://orcid.org/0000-0003-4381-1842>),
Doug Lorenz [aut] (<https://orcid.org/0000-0001-8114-0926>)
Maintainer Mary Gregg <megregg07@gmail.com>
Repository CRAN
Date/Publication 2022-02-13 13:00:09 UTC

R topics documented:

chisqtestClust .......................... 2
Chi-squared Test for Clustered Count Data

Description

chisqtestClust performs chi-squared contingency table tests and goodness-of-fit tests for clustered data with potentially informative cluster size.

Usage

chisqtestClust(x, y = NULL, id, p = NULL,
    variance = c("MoM", "sand.null", "sand.est", "emp"))

Arguments

x a numeric vector or factor. Can also be a table or data frame.
y a numeric vector or factor of the same length as x. Ignored if x is a table or data frame.
id a numeric vector or factor which identifies the clusters; ignored if x is a table or data frame. The length of id must be the same as the length of x.
p a vector of probabilities with length equal to the number of unique categories of x if x is a vector, or equal to the number of columns of x if x is a table or data frame.
variance character string specifying the method of variance estimation. Must be one of "sand.null", "sand.est", "emp", or "MoM".

Details

If x is 2-dimensional table or data frame, or if x is a vector or factor and y is not given, then the cluster-weighted goodness-of-fit test is performed. When x is a table or data frame, the rows of x must give the aggregate category counts across the clusters. In this case, the hypothesis tested is whether the marginal population probabilities equal those in p, or are all equal if p is not given.
When \(x\), \(y\), and \(id\) are all given as vectors or factors, the cluster-weighted *chi-squared test of independence* is performed. The lengths of \(x\), \(y\), and \(id\) must be equal. In this case, the hypothesis tested is that the joint probabilities of \(x\) and \(y\) are equal to the product of the marginal probabilities.

**Value**

A list with class "\(htest\)" containing the following components:

- `statistic`: the value of the test statistic.
- `parameter`: the degrees of freedom of the approximate chi-squared distribution of the test statistic.
- `p.value`: the p-value of the test.
- `method`: a character string indicating the test performed, and which variance estimation method was used.
- `data.name`: a character string giving the name(s) of the data and the total number of clusters.
- `M`: the number of clusters.
- `observed`: the observed reweighted proportions.
- `expected`: the expected proportions under the null hypothesis.

**References**


**Examples**

```r
data(screen8)
## is the marginal extracurricular activity participation evenly distributed across categories?
## Goodness of Fit test using vectors.
chisqtestClust(x=screen8$activity, id=screen8$sch.id)

## Goodness of Fit test using table.
act.table <- table(screen8$sch.id, screen8$activity)
chisqtestClust(act.table)

## test if extracurricular activity participation and gender are independent
chisqtestClust(screen8$gender, screen8$activity, screen8$sch.id)
```

**Description**

Test for marginal association between paired samples in clustered data with potentially informative cluster size.
Usage

cortestClust(x, ...)  

## Default S3 method:  
cortestClust(  
x,  
y,  
id,  
method = c("pearson", "kendall", "spearman"),  
alternative = c("two.sided", "less", "greater"),  
conf.level = 0.95,  
...  
)  

## S3 method for class 'formula'  
cortestClust(formula, id, data, subset, na.action, ...)  

Arguments

x, y numeric vectors of data values.  
... further arguments to be passed to or from methods.  
id a vector or factor object which identifies the clusters. The length of id should be the same as the number of observations.  
method a character string indicating which correlation coefficient is to be used for the test. One of "pearson", "kendall", or "spearman". Can be abbreviated.  
alternative indicates the alternative hypothesis and must be one of "two.sided", "greater", or "less". You can specify just the initial letter. "greater" corresponds to positive association, "less" to negative association.  
conf.level confidence level for the returned confidence interval.  
formula a formula of the form ~ u + v, where each of u and v are numeric variables giving the data values for one sample. The samples must be of the sample length.  
data an optional matrix or data frame containing variables in the formula formula. By default the variables are taken from environment(formula).  
subset an optional vector specifying a subset of observations to be used.  
na.action a function which indicates what should happen when data contain NAs. Defaults togetOption("na.action").

Details

The three methods each estimate the marginal association between paired observations from clustered data and compute a test of the value being zero.

If method is "pearson" ("kendall"), the test statistic is based on the Pearson product-moment (Kendall concordance coefficient) analog of Lorenz et al. (2011).

If method is "spearman", the test statistic is based on the Spearman coefficient analog of Lorenz et al. (2018) modified for paired data.
Value
A list with class "htest" containing the following components:

- **statistic**: the value of the test statistic.
- **p.value**: the p-value of the test.
- **estimate**: the estimated measure of marginal association, with name "cluster-weighted cor", "cluster-weighted tau", or "cluster-weighted rho" corresponding to the method employed.
- **null.value**: the value of the association measure under the null hypothesis, always 0.
- **conf.int**: a confidence interval for the measure of association.
- **alternative**: a character string describing the alternative hypothesis.
- **method**: a character string indicating how the association was measured.
- **data.name**: a character string giving the name(s) of the data and the total number of clusters.
- **M**: the number of clusters.

References

Examples
```r
data(screen8)
## test if math and reading scores are marginally correlated using vectors
cortestClust(screen8$read, screen8$math, screen8$sch.id)

## formula interface
cortestClust(~ math + read, sch.id, data=screen8, method="kendall")
```

icsPlot

*icsPlot*

**Test of Marginal Proportion for Clustered Data**

Description
Function to visualize informative cluster size. Plots within-cluster summary statistic from quantitative variables against the size of each cluster. For categorical variables, a barplot of category proportions for quantiles of cluster size is produced.
Usage

icsPlot(
  x,
  id,
  FUN = c("mean", "median", "var", "sd", "range", "IQR", "prop"),
  breaks,
  xlab = NULL,
  ylab = NULL,
  legend = c(TRUE, FALSE),
  ...)

Arguments

x vector of data values. Alternatively a two-dimensional table or matrix.
id a vector which identifies the clusters, with length equal to length of x; ignored if x is a matrix or table.
FUN the name of the function that produces the desired intra-cluster summary statistic.
breaks a single number giving the number of desired quantiles for the barplot of categorical variables with >2 categories.
xlab a label for the x axis, defaults to "cluster size".
ylab a label for the y axis, defaults to a description of FUN of x.
legend a logical indicating whether a legend should be included in a barplot.
... further arguments to be passed to or from methods.

Details

If x is a matrix or table and x has exactly two columns, the first column should contain the cluster sizes and the second column the respective intra-cluster summary statistic (e.g., mean, variance) that will be plotted against cluster size.

If x has more than two columns, the first column is assumed to contain the cluster size and the subsequent columns the counts of intra-cluster observations belonging to the different categorical variable levels. If there are exactly two categorical levels (e.g., x has exactly three columns), a scatterplot of the proportion of intracluster observations belonging to the first category will be plotted against the cluster size. If the number of categories is >2, a barplot of category proportions against quantiles of cluster size is produced.

Standard graphical parameters can be passed to icsPlot through the ... argument.

Examples

data(screen8)
## VECTOR INPUT
## plot average math score by cluster size
icsPlot(x = screen8$math, id = screen8$sch.id, pch = 20)
## TABLE INPUT

### Plot proportion of females by cluster size
icsPlot(screen8$gender, screen8$sch.id, pch = 20, main = "Female proportion by cluster size")

### Barchart of activity proportion by quartile of cluster size
icsPlot(x = screen8$activity, id = screen8$sch.id)

### TABLE INPUT

### Plot intra-cluster variance of math score by cluster size
cl.size <- as.numeric(table(screen8$sch.id))
tab1 <- cbind(cl.size, aggregate(screen8$math, list(screen8$sch.id), var)[,2])
colnames(tab1) <- c("cl.size", "variance")
icsPlot(x = tab1, pch = 17, main = "Math score variance by cluster size")

### Barchart of activity proportion across five quantiles of cluster size
tab2 <- cbind(cl.size, table(screen8$sch.id, screen8$activity))
icsPlot(tab2, breaks = 5)

---

### icstestClust

#### Test for Informative Cluster Size

**Description**

Performs a test for informative cluster size.

**Usage**

icstestClust(x, id, test.method = c("TF", "TCM"), B = 1000, print.it = TRUE)

**Arguments**

- **x**
  - a vector of numeric responses. Can also be a data frame.
- **id**
  - a vector or factor object which identifies the clusters; ignored if x is a data frame.
  - The length of id must be the same as the length of x.
- **test.method**
  - character string specifying the method of construction for the test statistic. Must be one of "TF" or "TCM".
- **B**
  - the number of bootstrap iterations.
- **print.it**
  - a logical indicating whether to print the progression of bootstrap iterations.

**Details**

The null is that the marginal distributions of the responses are independent of the cluster sizes. A small p-value is evidence for the presence of informative cluster size.

When test.method = "TF", the test statistic is constructed based on differences between the null and alternative distribution functions. "TF" is the suggested method when there are a large number of unique cluster sizes and the number of clusters of each size is small. When test.method = "TCM",
the test statistic is a multisample Cramer von Mises-based test. This method is recommended when there are a small number of possible cluster sizes. See Nevalainen et al. (2017) for more details.

When x is a data frame, the first column should contain values denoting cluster membership and the second column the responses.

This test is computationally intensive and can take significant time to execute. print.it defaults to TRUE to identify the bootstrap progression.

Value

A list with class "htest" containing the following components:

- statistic: the value of the test statistic.
- p.value: the p-value of the test.
- method: a character string indicating the test performed and the method of construction.
- data.name: a character string giving the name(s) of the data.

References


Examples

```r
data(screen8)
## using vectors
## test if cluster size is related to math scores
icstestClust(screen8$math, screen8$sch.id, B=100)

## same test, but using a data frame and supressing iterations
tdat <- data.frame(screen8$sch.id, screen8$math)
icstestClust(tdat, B=100, print.it = FALSE)
```

levenetestClust

Reweighted Levene's Test for Homogeneity of Variance in Clustered Data

Description

Performs a reweighted test for homogeneity of marginal variances across intra-cluster groups in clustered data. Appropriate for clustered data with cluster- or group size informativeness.
Usage

levenetestClust(y, ...)  

## Default S3 method:  
levenetestClust(y, group, id, center = c("median", "mean"), trim = NA, ...)  

## S3 method for class 'formula'  
levenetestClust(formula, id, data, subset, na.action, ...)  

Arguments

y           vector of numeric responses.  
...          further arguments to be passed to or from methods.  
group       vector or factor object defining groups.  
id          vector or factor object denoting cluster membership for y responses.  
center      The name of a function to compute the center of each group. If mean, the reweighted group means will be used. The default median is the suggested measure of center, as it provides a more robust test.  
trim        optional numeric argument taking values [0, 0.5] to specify the percentage trimmed mean. Ignored if center = median.  
formula     a formula of the form lhs ~ rhs where lhs is a numeric variable giving the data values and rhs a factor with two or more levels giving the corresponding groups.  
data        an optional matrix or data frame containing variables in the formula formula and id. By default the variables are taken from environment(formula).  
subset      an optional vector specifying a subset of observations to be used.  
na.action   a function which indicates what should happen when data contain NAs. Defaults togetOption("na.action").

Details

The null hypothesis is that all levels of group have equal marginal variances.

Value

A list with class "htest" containing the following components:

statistic    the value of the test statistic.  
p.value      the p-value of the test.  
parameter    the degrees of freedom of the chi square distribution.  
method       a character string indicating the test performed.  
data.name    a character string giving the name of the data and the total number of clusters.  
M            the number of clusters.
References


Examples

data(screen8)

## Do boys and girls have the same variability in math scores?
## Test using vectors
levenetestClust(y=screen8$math, group=screen8$gender, id=screen8$sch.id)

## Test using formula method
levenetestClust(math~gender, id=sch.id, data=screen8)

## Using 10% trimmed mean
levenetestClust(math~gender, id=sch.id, data=screen8, center="mean", trim=.1)

$\texttt{mcnemartestClust}$  $\textit{Test of Marginal Homogeneity for Clustered Data}$

Description

Performs a test of marginal homogeneity of paired clustered data with potentially informative cluster size.

Usage

$\texttt{mcnemartestClust}(x, y, id, \text{variance = c("MoM", "emp")})$

Arguments

$x, y$ vector or factor objects of equal length.

$id$ a vector or factor object which identifies the clusters. The length of $id$ must be the same as the length of $x$.

$\text{variance}$ character string specifying the method of variance estimation. Must be one of "MoM" or "emp".

Details

The null is that the marginal probabilities of being classified into cells $[i, j]$ and $[j, i]$ are equal. Arguments $x, y$, and $id$ must be vectors or factors of the same length. Incomplete cases are removed.

When $\text{variance}$ is $\text{MoM}$, a method of moments variance estimate evaluated under the null is used. This is equivalent to the test by Durkalski et al. (2003). When $\text{variance}$ is $\text{emp}$, an empirical variance estimate is used. See Gregg (2020) for details.
onewaytestClust

Value
A list with class "htest" containing the following components:

- **statistic**: the value of the test statistic.
- **parameter**: the degrees of freedom of the approximate chi-squared distribution of the test statistic.
- **p.value**: the p-value of the test.
- **method**: a character string indicating the test performed and which variance estimation method was used.
- **data.name**: a character string giving the name(s) of the data and the total number of clusters.
- **M**: the number of clusters.

References

Examples
data(screen8)
## Is marginal proportion of students in lowest fitness category
## at the end of year equal to the beginning of year?
screen8$low.start <- 1*(screen8$qfit.s=='Q1')
screen8$low.end <- 1*(screen8$qfit=='Q1')
mcnemartestClust(screen8$low.start, screen8$low.end, screen8$sch.id)

onewaytestClust

Test for Equal Marginal Means in Clustered Data

Description
Test whether two or more intra-cluster groups have the same marginal means in clustered data. Reweighted to correct for potential cluster- or group size informativeness.

Usage
onewaytestClust(x, ...)

## Default S3 method:
onewaytestClust(x, ...)

## S3 method for class 'formula'
onewaytestClust(formula, id, data, subset, ...)
Arguments

- **x**: a two-dimensional matrix or data frame containing the within-cluster group means, where rows are the clusters and columns are the group means.
- **...**: further arguments to be passed to or from methods.
- **formula**: a formula of the form `lhs ~ rhs`, where `lhs` is a numeric variable giving the data values and `rhs` a numeric or factor with two or more levels giving the corresponding groups.
- **id**: a vector or factor object denoting cluster membership.
- **data**: an optional matrix or data frame containing variables in the formula `formula` and `id`. By default the variables are taken from `environment(formula)`.
- **subset**: an optional vector specifying a subset of observations to be used.

Details

The null hypothesis is that all levels of `group` have equal marginal means.

If `x` is a matrix or data frame, the dimension of `x` should be MxK, where M is the number of clusters and K is the number of groups. Each row of `x` corresponds to a cluster, where the column values contain the respective group means from that cluster. Clusters which do not contain observations in a particular group should have `NA` in the corresponding column.

Value

A list with class "htest" containing the following components:

- **statistic**: the value of the test statistic.
- **p.value**: the p-value of the test.
- **estimate**: the estimated marginal group means.
- **parameter**: the degrees of freedom of the chi square distribution.
- **method**: a character string indicating the test performed.
- **data.name**: a character string giving the name of the data and the total number of clusters.
- **M**: the number of clusters.

References


Examples

```r
data(screen8)
## do average reading scores differ across after-school activities?
## test using a table
read.tab <- tapply(screen8$read, list(screen8$sch.id, screen8$activity), mean)
onewaytestClust(read.tab)

## test using formula method
onewaytestClust(read~activity, id=sch.id, data=screen8)
```
Description

proptestClust can be used for testing the null that the marginal proportion (probability of success) is equal to certain given values in clustered data with potentially informative cluster size.

Usage

proptestClust(x, id, p = NULL, alternative = c("two.sided", "less", "greater"), variance = c("sand.null", "sand.est", "emp", "MoM"), conf.level = 0.95)

Arguments

x a vector of binary indicators denoting success/failure of each observation, or a two-dimensional table (or matrix) with 2 columns giving the aggregate counts of failures and successes (respectively) across clusters.

id a vector which identifies the clusters; ignored if x is a matrix or table. The length of id must be the same as the length of x.

p the null hypothesized value of the marginal proportion. Must be a single number greater than 0 and less than 1.

alternative a character string specifying the alternative hypothesis. Must be one of "two.sided", "greater", or "less". You can specify just the initial letter.

variance character string specifying the method of variance estimation. Must be one of "sand.null", "sand.est", "emp", or "MoM".

conf.level confidence level of the returned confidence interval. Must be a single number between 0 and 1.

Details

If p is not given, the null tested is that the underlying marginal probability of success is .5.

The variance argument allows the user to specify the method of variance estimation, selecting from the sandwich estimate evaluated at the null hypothesis (sand.null), the sandwich estimate evaluated at the cluster-weighted proportion (sand.est), the empirical estimate (emp), or the method of moments estimate (MoM).

Value

A list with class "htest" containing the following components:

statistic the value of the test statistic.
p.value the p-value of the test.
estimate the estimated marginal proportion.
null.value the value of p under the null hypothesis.
conf.int a confidence interval for the true marginal proportion.
alternative a character string describing the alternative hypothesis.
method a character string indicating the test performed and method of variance estimation.
data.name a character string giving the name of the data and the total number of clusters.
M the number of clusters.

References

Examples
data(screen8)
## using vectors
## suppose math proficiency is determined by score >= 65
## is the marginal proportion of students proficient in math at least 75%?
screen8$math.p <- 1*(screen8$math>=65)
proptestClust(screen8$math.p, screen8$sch.id, p = .75, alternative = "great")

## using table and empirical variance; two-sided CI
## (note that "failure" counts are the first column and "success" counts are the second column)
mathp.tab <- table(screen8$sch.id, screen8$math.p)
proptestClust(mathp.tab, variance="emp", p=.75)

---

screen8 Example data for informative cluster size

Description
Simulated hypothetical clustered data created for illustration of functions in the htestClust package.

Usage
data(screen8)

Format
A data frame with 2224 rows and 12 columns:
sch.id identification variable for school (clusters).
stud.id identification variable for students within schools (observations within clusters).
age student age in years.
gender binary student gender.
**height**  student height in inches.

**weight**  student weight in lbs.

**math**  score from standardized math test.

**read**  score from standardized reading test.

**phq2**  ordinal (0-6) score from a mental health screening; higher scores correspond to higher levels of depression.

**qfit**  age-adjusted fitness quartile from physical health assessment at end of school year.

**qfit.s**  age-adjusted fitness quartile from physical health assessment at beginning of school year.

**activity**  student’s primary after-school activity.

**Details**

Hypothetical data simulated for the following scenario. An urban school district has collected demographic, biometric, and academic performance data from graduating 8th grade students. `screen8` contains a sample of this data from 2224 students across 73 schools. Student-level observations are clustered within schools. The school district has implemented an incentive program in which schools with higher participation rates are prioritized for classroom and technology upgrades. Cluster size could be informative in this data, as resource-poor schools might have higher participation rates (larger cluster size), but also tend to have worse health metrics and lower standardized test scores.

**Author(s)**

Mary Gregg

**Examples**

```r
data(screen8)
head(screen8)

## plot average math scores by cluster size
c.l.size <- as.numeric(table(screen8$sch.id))
ave.math <- tapply(screen8$math, list(screen8$sch.id), mean)
plot(cl.size, ave.math)
```

---

**ttestClust**  
*Test of Marginal Means in Clustered Data*

**Description**

Performs one and two sample tests of marginal means in clustered data, reweighted to correct for potential cluster- or group-size informativeness.
Usage

ttestClust(x, ...)

## Default S3 method:
ttestClust(
  x,
  y = NULL,
  idx,
  idy = NULL,
  alternative = c("two.sided", "less", "greater"),
  mu = 0,
  paired = FALSE,
  conf.level = 0.95,
  ...
)

## S3 method for class 'formula'
ttestClust(formula, id, data, subset, na.action, ...)

Arguments

x, y numeric vectors of data values.
...
... further arguments to be passed to or from methods.
idx vector or factor object denoting cluster membership for x observations (or cluster membership for paired observations when paired is TRUE). Length must be equal to length of x.
idy vector or factor object denoting cluster membership for y observations. Length must be equal to length of y
alternative indicates the alternative hypothesis and must be one of "two.sided", "greater", or "less". You can specify just the initial letter.
mu a number specifying an optional parameter used to form the null hypothesis.
paired a logical indicating whether x and y are paired.
conf.level confidence level of the interval.
formula a formula of the form lhs ~ rhs where lhs is a numeric variable giving the data values and rhs a factor with two levels giving the corresponding groups.
id a vector or factor giving the corresponding cluster membership.
data an optional matrix or data frame containing variables in the formula formula and id. By default the variables are taken from environment(formula).
subset an optional vector specifying a subset of observations to be used.
na.action a function which indicates what should happen when data contain NAs. Defaults togetOption("na.action").

Details

The formula interface is only applicable for the 2-sample tests.
If paired is TRUE then x, y, and idx must be given and be of the same length. idy is ignored.
ttestClust

Value

A list with class "htest" containing the following components:

- **statistic**: the value of the test statistic.
- **p.value**: the p-value of the test.
- **conf.int**: a confidence interval for the mean appropriate to the specified alternative hypothesis.
- **estimate**: the estimated mean or difference in means, depending on whether it was a one-sample or two-sample test.
- **null.value**: the specified hypothesized value of the mean or mean difference.
- **alternative**: a character string describing the alternative hypothesis.
- **method**: a character string indicating what type of reweighted test of means was performed.
- **data.name**: a character string giving the name of the data and the total number of clusters.
- **M**: the number of clusters.

References


Examples

data(screen8)
## One sample test
## Test if marginal math scores are equal to 70
ttestClust(x=screen8$math, idx=screen8$sch.id, mu = 70)

## paired test
## Test equality of marginal means in math and reading scores
ttestClust(x=screen8$math, y=screen8$read, idx=screen8$sch.id, paired=TRUE)

## unpaired test
## Test if boys and girls have equal marginal math scores
boys <- subset(screen8, gender=='M')
girls <- subset(screen8, gender=='F')
ttestClust(x=boys$math, y=girls$math, idx=boys$sch.id, idy=girls$sch.id)

## unpaired test using formula method
ttestClust(math~gender, id=sch.id, data=screen8)
vartestClust  
\textit{Reweighted Test to Compare Two Variances in Clustered Data}

\textbf{Description}

 Performs a reweighted test to compare marginal variances of intra-cluster groups in clustered data. Appropriate for clustered data with cluster- or group-size informativeness.

\textbf{Usage}

\begin{verbatim}
vartestClust(x, ...)  
## Default S3 method:  
vartestClust(  
x,  
y,  
idx,  
idy,  
difference = 0,  
alternative = c("two.sided", "less", "greater"),  
conf.level = 0.95,  
...  
)  

## S3 method for class 'formula'  
vartestClust(formula, id, data, subset, na.action, ...)
\end{verbatim}

\textbf{Arguments}

\begin{itemize}
  \item \textbf{x, y} numeric vectors of data values.
  \item ... further vectors to be passed to or from methods.
  \item \textbf{idx} vector or factor object denoting cluster membership for \textit{x} observations. Length must be equal to length of \textit{x}.
  \item \textbf{idy} vector or factor object denoting cluster membership for \textit{y} observations. Length must be equal to length of \textit{y}.
  \item \textbf{difference} the hypothesized difference of the marginal population variances of \textit{x} and \textit{y}.
  \item \textbf{alternative} indicates the alternative hypothesis and must be one of "two.sided", "greater", or "less". You can specify just the initial letter.
  \item \textbf{conf.level} confidence level of the interval.
  \item \textbf{formula} a formula of the form \textit{lhs} \sim \textit{rhs} where \textit{lhs} is a numeric variable giving the data values and \textit{rhs} a factor with two levels giving the corresponding groups.
  \item \textbf{id} a vector or factor giving the corresponding cluster membership.
  \item \textbf{data} an optional matrix or data frame containing variables in the formula \textit{formula} and \textit{id}. By default the variables are taken from \textit{environment(formula)}.
\end{itemize}
subset an optional vector specifying a subset of observations to be used.

na.action a function which indicates what should happen when data contain NAs. Defaults
to getOption("na.action").

Details

The null hypothesis is that the difference of the marginal variances of the populations of intra-cluster
groups from which x and y were drawn is equal to difference.

Using the default method, difference is the difference of the reweighted sample variances of x
and y. When using the formula method, the order of the difference is determined by the order of
the factor levels of rhs.

Value

A list with class "htest" containing the following components:

statistic the value of the test statistic.
p.value the p-value of the test.
conf.int a confidence interval for the difference of the population marginal variances.
estimate the difference in reweighted sample variances of x and y.
null.value the difference of population marginal variances under the null.
alternative a character string describing the alternative hypothesis.
method a character string indicating the test performed.
data.name a character string giving the name of the data and the total number of clusters.
M the number of clusters.

References


Examples

data(screen8)
boys <- subset(screen8, gender=='M')
girls <- subset(screen8, gender=='F')

## Do boys and girls have the same variability in math scores?
## Test using vectors
vartestClust(x=boys$math, y=girls$math, idx=boys$sch.id, idy=girls$sch.id)

## Test using formula method.
vartestClust(math~gender, id=sch.id, data=screen8)

## Note that in this example, the sign of the estimate returned when using the formula
## method is opposite to that when the test was performed using vectors. This is due to
## the order of the gender factor levels
wilcoxttestClust  

Rank Sum and Signed Rank Tests for Clustered Data

Description

Performs a one-sample or paired cluster-weighted signed rank test, or a cluster- or group-weighted rank sum test. These tests are appropriate for clustered data with potentially informative cluster size.

Usage

wilcoxttestClust(x, ...)  

## Default S3 method:
wilcoxttestClust(  
  x,  
  y = NULL,  
  idx,  
  idy = NULL,  
  alternative = c("two.sided", "less", "greater"),  
  mu = 0,  
  paired = FALSE,  
  method = c("cluster", "group"),  
  ...  
)

## S3 method for class 'formula'
wilcoxttestClust(formula, id, data, subset, na.action, ...)

Arguments

x, y  
numeric vectors of data values.

...  
further arguments to be passed to or from methods.

idx  
vector or factor object denoting cluster membership for x observations (or cluster membership for paired observations when paired is TRUE). Length must be equal to length of x.

idy  
vector or factor object denoting cluster membership for y observations. Length must be equal to length of y

alternative  
indicates the alternative hypothesis and must be one of "two.sided", "greater", or "less". You can specify just the initial letter.

mu  
a number specifying an optional parameter used to form the null hypothesis. Ignored when performing a rank-sum test. See 'Details'.

paired  
a logical indicating whether x and y are paired. When TRUE, the cluster-weighted signed rank test is performed.
method: a character string specifying the method of rank sum test to be performed. See "Details".

formula: a formula of the form lhs ~ rhs, where lhs is a numeric variable giving the data values and rhs a numeric or factor with two levels giving the corresponding groups.

id: a vector or factor object denoting cluster membership.

data: an optional matrix or data frame containing variables in the formula formula and id. By default the variables are taken from environment(formula).

subset: an optional vector specifying a subset of observations to be used.

na.action: a function which indicates what should happen when data contain NA s. Defaults togetOption("na.action").

Details

The formula interface is only applicable for the 2-sample rank-sum tests.

If only x and idx are given, a cluster-weighted signed rank test of the null that the distribution of x is symmetric about mu is performed.

If x and y are given and paired is TRUE, only idx is necessary (idy is ignored). In this case, a cluster-weighted signed-rank test of the null that the distribution of x - y is symmetric about mu is performed.

When method is cluster, the cluster-weighted rank sum test of Datta and Satten (2005) is performed. The data must have complete intra-cluster group distribution (i.e., all clusters must contain observations belonging to both groups) for this test to be performed.

When method is group, the group-weighted rank-sum test of Dutta and Datta (2015) is performed. This test is appropriate for clustered data with potentially informative intra-cluster group size. Incomplete intra-cluster group distribution is permitted.

For the rank sum tests, the null is that the two groups follow the same marginal distribution. mu is ignored when performing these tests.

The tests performed by this function involve computation of reweighted empirical CDFs. This is computationally intensive and can result in lengthy execution time for large data sets.

Value

A list with class "htest" containing the following components:

statistic: the value of the test statistic.
p.value: the p-value of the test.
null.value: the location parameter mu. Always 0 for rank sum test.
data.name: a character string giving the name(s) of the data and the total number of clusters.
method: a character string indicating the test performed and method of construction.
alternative: a character string describing the alternative hypothesis.
M: the number of clusters.
References


Examples

```r
data(screen8)
## One-sample signed rank test
wilcoctestClust(x=screen8$math, idx=screen8$sch.id, mu=70)

## Paired signed rank test
wilcoctestClust(x=screen8$math, y=screen8$read, idx=screen8$sch.id, paired=TRUE, mu=10)

## Cluster-weighted rank sum test
wilcoctestClust(math~gender, id=sch.id, data=screen8)

## Group-weighted rank sum test
boys <- subset(screen8, gender=="M")
girls <- subset(screen8, gender=="F")
wilcoctestClust(x=boys$math, y=girls$math, idx=boys$sch.id, idy=girls$sch.id, method="group")

## Group-weighted rank sum test using formula method
wilcoctestClust(math~gender, id=sch.id, data=screen8, method="group")
```
Index

* datasets
  screen8, 14
chisqtestClust, 2
cortestClust, 3
icsPlot, 5
icstestClust, 7
levenetestClust, 8
mcnemartestClust, 10
onewaytestClust, 11
proptestClust, 13
screen8, 14
ttestClust, 15
vartestClust, 18
wilcoxtestClust, 20