# Package ‘MultNonParam’

June 8, 2018

**Type**  Package  
**Title**  Multivariate Nonparametric Methods  
**Version**  1.2.6  
**Date**  2018-6-5  

**Description**  A collection of multivariate nonparametric methods, selected in part to support an MS level course in nonparametric statistical methods. Methods include adjustments for multiple comparisons, implementation of multivariate Mann-Whitney-Wilcoxon testing, inversion of these tests to produce a confidence region, some permutation tests for linear models, and some algorithms for calculating exact probabilities associated with one- and two-stage testing involving Mann-Whitney-Wilcoxon statistics.

**License**  GPL-2  
**RoxygenNote**  6.0.1  
**LazyData**  true  
**NeedsCompilation**  yes  
**Author**  John E. Kolassa [cre],  
            Stephane Jankowski [aut]  
**Maintainer**  John E. Kolassa <kolassa@stat.rutgers.edu>  
**Repository**  CRAN  
**Date/Publication**  2018-06-08 07:42:55 UTC  

## R topics documented:

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>MultNonParam-package</td>
<td>2</td>
</tr>
<tr>
<td>aov.P</td>
<td>2</td>
</tr>
<tr>
<td>betatest</td>
<td>3</td>
</tr>
<tr>
<td>dmannwhitney</td>
<td>3</td>
</tr>
<tr>
<td>higgins.fisher.kruskal.test</td>
<td>4</td>
</tr>
<tr>
<td>mood.median.test</td>
<td>5</td>
</tr>
<tr>
<td>page.test.unbalanced</td>
<td>5</td>
</tr>
<tr>
<td>probest</td>
<td>6</td>
</tr>
<tr>
<td>prostate</td>
<td>7</td>
</tr>
<tr>
<td>sensitivity.plot</td>
<td>8</td>
</tr>
</tbody>
</table>
MultNonParam-package  MultNonParam

Description
A collection of nonparametric methods.

Author(s)
Maintainer: John E. Kolassa <kolassa@stat.rutgers.edu>
Authors:
• Stephane Jankowski

aov.P  One-way ANOVA using permutation tests

Description
aov.P uses permutation tests instead of classic theory tests to run a one-way or two-way ANOVA.

Usage
aov.P(dattab, permi = NULL, be = NULL)

Arguments
dattab The table on which the ANOVA has to be done, or a vector of responses.
permi If dattab is a table, ignored. If dattab is a vector, a vector of treatment labels.
be If dattab is a table, ignored. If dattab is a vector, a vector of end points of blocks. In this case, blocks must form contiguous subvectors of dattab. If null, no blocking.

Details
The function calls a Fortran code to perform the permutation tests and the ANOVA. The function has to be applied directly on a cross-table of two variables.
**betatest**

**Description**

Permutation test of association

Calculate the p-value for the test of association between two variables using the permutation method.

**Usage**

```
betatest(x, y)
```

**Arguments**

- `x`: First vector to be associated.
- `y`: First vector to be associated.

**Value**

p-value

**Examples**

```
# Example using data from plant Qn1 from the CO2 data set.
betatest(CO2$Plant=="Qn1", CO2$Plant=="Qn1")
```

---

**dmannwhitney**

Mann Whitney Probability Mass function

Description

Calculates the Mann Whitney Probability Mass function recursively.

**Usage**

```
dmannwhitney(u, m, n)
```

**Arguments**

- `u`: Statistic value
- `m`: Group 1 size
- `n`: Group 2 size
Value

Probability that the Mann-Whitney statistic takes the value $u$ under H0

---

**higgins.fisher.kruskal.test**

*Fisher's LSD method applied to the Kruskal-Wallis test*

---

**Description**

This function applies a rank-based method for controlling experiment-wise error. Two hypothesis have to be respected: normality of the distribution and no ties in the data. The aim is to be able to detect, among $k$ treatments, those who lead to significant differences in the values for a variable of interest.

**Usage**

```
higgins.fisher.kruskal.test(resp, grp, alpha = 0.05)
```

**Arguments**

- `resp` vector containing the values for the variable of interest.
- `grp` vector specifying in which group is each observation.
- `alpha` level of the test.

**Details**

First, the Kruskal-Wallis test is used to test the equality of the distributions of each treatment. If the test is significant at the level $\alpha$, the method can be applied.

**Value**

A matrix with two columns. Each row indicates a combination of two groups that have significant different distributions.

**References**

mood.median.test

Mood’s Median test, extended to odd sample sizes.

Description

Test whether two samples come from the same distribution. This version of Mood’s median test is presented for pedagogical purposes only. Many authors successfully argue that it is not very powerful. The name "median test" is a misnomer, in that the null hypothesis is equality of distributions, and not just equality of median. Exact calculations are not optimal for the odd sample size case.

Usage

mood.median.test(x, y, exact = FALSE)

Arguments

x  
First data set.

y  
Second data set.

exact  
Indicator for whether the test should be done exactly or approximately.

Details

The exact case reduces to Fisher’s exact test.

Value

The two-sided p-value.

page.test.unbalanced

Perform Page test for unbalanced two-way design

Description

Perform Page test for unbalanced two-way design

Usage

page.test.unbalanced(x, trt, blk, sides = 2)

Arguments

x  
A vector of responses

trt  
A vector of consecutive integers starting at 1 indicating treatment

blk  
A vector of consecutive integers starting at 1 indicating block

sides  
A single integer indicating sides. Defaults to 2.
Value

P-value for Page test.

Examples

```r
page.test.unbalanced(rnorm(15), rep(1:3, 5), rep(1:5, rep(3, 5)))
```

---

**probest**

*Stratified Multivariate Kawaguchi Koch Wang Estimators*

Description

Function that return the estimators and their variance-covariance matrix calculated with the Kawaguchi - Koch - Wang method.

Usage

```r
probest(ds, resp, grp, str = NULL, covs = NULL, delta = NA, correct = FALSE)
```

Arguments

- **ds**: The data frame to be used.
- **resp**: The vector of the response manifest variable. There can be more than one variable. It has to be the name of the variable as a character string.
- **grp**: The vector of the variable that divides the population into groups. It has to be the name of the variable as a character string.
- **str**: The vector of the variable used for the strata. It has to be the name of the variable as a character string.
- **covs**: The covariates to be used in the model. It has to be the name of the variable as a character string.
- **delta**: Offset for covariates.
- **correct**: Should the variance estimator be corrected as in Chen and Kolassa?

Details

The function calls a Fortran code to calculate the estimators \( b \) and their variance-covariance matrix \( Vb \)

Value

A list with components \( b \), the vector of adjusted estimates from the method, and \( Vb \), the corresponding estimated covariance matrix.
References


Examples

```r
# Breast cancer data from the MultNonParam package.
data(sotiriou)
attach(sotiriou)
# First simple plot of the data
plot(AGE,TUMOR_SIZE,pch=record+1),main="Age and Tumor Size",
sub="Breast Cancer Recurrence Data",xlab="Age (years)",
ylab="Tumor Size",col=c("blue","darkolivegreen"))
legend(31,8,legend=c("Not Recurrent","Recurrent"),
pch=1:2,col=c("blue","darkolivegreen"))
# AGE and TUMOR_SIZE are the response variables, recur is used for the groups,
# TAMOXIFEN_TREATMENT for the stratum and ELSTON.ELLIS GRADE is a covariate.
poc<probest(sotiriou,c("AGE","TUMOR_SIZE"),"recur",
"TAMOXIFEN_TREATMENT","ELSTON.ELLIS GRADE")
```

Description

221 prostate cancer patients are collected in this data set.

Usage

data(prostate)

Format

- hosp: Hospital in which the patient is hospitalized.
- stage: stage of the cancer.
- gleason score: used to help evaluate the prognosis of the cancer.
- psa: prostate-specific antigen.
- age: age of the patient.
- advanced: boolean. TRUE if the cancer is advanced.
References


Examples

```r
data(prostate)
attach(prostate)
plot(age,psa,main="Age and PSA",sub="Prostate Cancer Data",
     xlab="Age (years)",ylab="PSA")
```

---

```
sensitivity.plot # Compare the sensitivity of different statistics.
```

Description

Compare the sensitivity of different statistics.

Usage

```r
sensitivity.plot(y, sub, stats)
```

Arguments

- `y` vector of the data.
- `sub` subtitle for the plot.
- `stats` vector of functions to be plotted.

Details

To compare the sensitivity, outliers are added to the original data. The shift of each statistics due to the new value is measured and plotted.
Description

187 breast cancer patients are collected in this data set.

Usage

data(sotiriou)

Format

A data set with the following variables

- AGE : Age of the patient
- TUMOR_SIZE : The size of the tumor, numeric variable
- recur : 1 if the patient has a recurrent breast cancer, 0 if it is not recurrent.
- ELSTON.ELLIS_GRADE : Elston Ellis grading system in order to classify the breast cancers. It can be a low, intermediate or high grade (high being the worst prognosis)
- TAMOXIFEN_TREATMENT : boolean. TRUE if the patient is treated with the Tamoxifen treatment.

Source

https://gdoc.georgetown.edu/gdoc/

References


Examples

data(sotiriou)
attach(sotiriou)
plot(AGE,TUMOR_SIZE,pch=(recur+1),
main="Age and Tumor Size",
sub="Breast Cancer Recurrence Data",
 xlab="Age (years)",ylab="Tumor Size",
col=c("blue","darkolivegreen"))
**Description**

This function returns either exact or asymptotic p-values for score tests of the null hypothesis of univariate symmetry about 0.

**Usage**

`symscorestat(y, scores = NULL, exact = F, sides = 1)`

**Arguments**

- **y**: Vector of data on which test will be run.
- **scores**: Scores to be used for the test. Defaults to integers 1:length(y).
- **exact**: Logical variable indicating whether the exact p-value should be calculated. Default is false.
- **sides**: Integer; 1 for one sided test rejecting for large values of the statistic, and 2 for the two-sided test. Defaults to 1.

**Details**

The statistic considered here is the sum of scores corresponding to those entries in y that are positive. If exact=T, the function calls a Fortran code to cycle through all permutations. If exact=F, the expectation of the statistic is calculated as half the sum of the scores, the variance is calculated as one quarter the sum of squares of scores about their mean, and the statistic is compared to its approximating normal distribution.

**Value**

A list with components `pv`, the p-value obtained with the permutation tests, and `tot`, the total number of rearrangements of the data considered in calculating the p-value.

**References**


**Examples**

`symscorestat(y=c(1,-2,3,-4,5), exact=TRUE)`
Description

Diagnostic tool that verifies the normality of the estimates of the probabilities b with the Kawaguchi - Koch - Wang method. The diagnostic method is based on a Monte Carlo method.

Usage

testve(n, m, k, nsamp = 100, delta = 0, beta = 0, disc = 0)

Arguments

n  number of observations in the first group.
m  number of observations in the second group.
k  number of strata.
nsamp  The number of estimates that will be calculated. Must be enough to be sure that the results are interpretable.
delta  Offset that depends on group.
beta  Correlation between x and y.
disc  The Mann Whitney test is designed to handle continuous data, but this method applies to discretized data; disc adjusts the discreteness.

Details

This functions serves as a diagnosis to prove that the Kawaguchi - Koch - Wang method gives Gaussian estimates for b. It generates random data sets, to which the Mann Whitney test gets applied. y is the generated response variable and x the generated covariable related to y through a regression model.

Value

Nothing is returned. A QQ plot is drawn.

References


Examples

testve(10,15,3,100,0.4)
tukey.kruskal.test \hspace{1em} Tukey HSD procedure

**Description**

Rank-based method for controlling experiment-wise error. Assume normality of the distribution for the variable of interest.

**Usage**

tukey.kruskal.test(resp, grp, alpha = 0.05)

**Arguments**

- `resp` vector containing the values for the variable of interest.
- `grp` vector specifying in which group is each observation.
- `alpha` level of the test.

**Details**

The original Tuckey HSD procedure is supposed to be applied for equal sample sizes. However, the `tukey.kruskal.test` function performs the Tukey-Kramer procedure that works for unequal sample sizes.

**Value**

A logical vector for every combination of two groups. `TRUE` if the distribution in one group is significantly different from the distribution in the other group.

**References**


util.jplot \hspace{1em} Plot a curve, skipping bits where there is a large jump.

**Description**

Plot a curve, skipping bits where there is a large jump.

**Usage**

util.jplot(x, y, ...)

Arguments

- x: Ordinates to be plotted.
- y: Abcissas to be plotted.
- . . .: Arguments passed directly to plot.
Index

aov.P, 2
betatest, 3
dmannwhitney, 3
higgins.fisher.kruskal.test, 4
mood.median.test, 5
MultNonParam (MultNonParam-package), 2
MultNonParam-package, 2
page.test.unbalanced, 5
probest, 6
prostate, 7
sensitivity.plot, 8
sotiriou, 9
symscorestat, 10
testve, 11
tukey.kruskal.test, 12
util.jplot, 12