Package ‘NPIstats’

October 12, 2022

Type Package
Title Nonparametric Predictive Inference
Version 0.1.0
Description An implementation of the Nonparametric Predictive Inference approach in R. It provides tools for quantifying uncertainty via lower and upper probabilities. It includes useful functions for pairwise and multiple comparisons: comparing two groups with and without terminated tails, selecting the best group, selecting the subset of best groups, selecting the subset including the best group.
License GPL-3
Depends R (>= 3.5.0)
Imports dplyr (>= 1.0.0)
Suggests knitr, rmarkdown, testthat (>= 3.0.0)
VignetteBuilder knitr
Encoding UTF-8
Language en-US
LazyData true
RoxygenNote 7.1.1
Config/testthat/edition 3
NeedsCompilation no
Author Tahani Coolen-Maturi [cre, aut]
Maintainer Tahani Coolen-Maturi <tahani.maturi@durham.ac.uk>
Repository CRAN
Date/Publication 2021-02-16 09:50:05 UTC

R topics documented:

best.pair ................................................. 2
best.pair.tt ............................................ 3
BirthWeights .......................................... 4
BreakdownTimes ...................................... 4
**best.pair**

### Description

NPI lower and upper probabilities for the event that the next future observation from group Y is larger than the next future observation from group X.

### Usage

`best.pair(X, Y)`

### Arguments

- **X**: numeric vector of data values
- **Y**: numeric vector of data values, to check if it is the best group

### Value

NPI lower and upper probabilities for the event that the next future observation from group Y is larger than the next future observation from group X.

### References


### Examples

```r
data(BreakdownTimes)
data2<-split(BreakdownTimes$times, BreakdownTimes$group)
# No terminated tails, complete data
best.pair(data2$X, data2$Y)
```
**Description**

NPI lower and upper probabilities for the event that the next future observation from group Y is larger than the next future observation from group X. The information available consists of precise measurements of real-valued data only within a specific range, between the cut points, where the numbers of observations to the left and to the right of this range available.

**Usage**

```r
best.pair.tt(X, Y, Lx = -Inf, Ux = Inf, Ly = -Inf, Uy = Inf)
```

**Arguments**

- `X`: numeric vector of data values
- `Y`: numeric vector of data values, to check if it is the best group
- `Lx`: numeric value, lower cut point for group X, default set to -Inf
- `Ux`: numeric value, upper cut point for group X, default set to Inf
- `Ly`: numeric value, lower cut point for group Y, default set to -Inf
- `Uy`: numeric value, lower cut point for group Y, default set to Inf

**Value**

NPI lower and upper probabilities for the event that the next future observation from group Y is larger than the next future observation from group X.

**References**


**Examples**

```r
data(BreakdownTimes)
data2<-split(BreakdownTimes$times, BreakdownTimes$group)
# No terminated tails, complete data
best.pair(tt(data2$X, data2$Y))
# terminated tails with Ly = 0.5, Uy = 4 and Ux = 10, but as Lx is not given then Lx=-Inf
best.pair.tt(data2$X, data2$Y, Ux = 10, Ly = 0.5, Uy = 4)
```
BirthWeights

**BirthWeights data set**

**Description**

**Usage**

data(BirthWeights)

**Format**
An object of class "data.frame"

- **group** male or female
- **weights** BirthWeights for 12 male and 12 female babies

**References**

**Examples**

data(BirthWeights)
head(BirthWeights)

BreakdownTimes

**Breakdown times of units from two groups**

**Description**
We consider a data set used by Nelson (1982, p.462), which gives the breakdown times of units from 6 different groups. In this data set, only the first two groups are used to illustrate the NPI method for pairwise comparison with tails termination. Both groups consist of 10 observations. The first unit of group X has a reported breakdown time of 0.00, we interpret this as a very small but positive breakdown time.

**Usage**

data(BreakdownTimes)
ChemicalReaction

Format
An object of class "data.frame"

- **group**: group X or Y
- **times**: Breakdown times

References

Examples
```r
data(BreakdownTimes)
head(BreakdownTimes)
```

---

**ChemicalReaction**  
*Chemical reaction of two methods*

Description
This data set is presented by Box et al. (1978, p. 159), where a chemical reaction was studied by making 10 runs with a standard method X, and 10 runs with a new, supposedly improved method Y.

Usage
```r
data(ChemicalReaction)
```

Format
An object of class "data.frame"

- **method**: Method X or Y
- **value**: Chemical reaction values

References

Examples
```r
data(ChemicalReaction)
head(ChemicalReaction)
```
### FourSources

**Description**

This data set is used by Coolen and van der Laan (2001) to introduce NPI for multiple comparisons.

**Usage**

```r
data(FourSources)
```

**Format**

An object of class "data.frame"

- **source** Source A, B, C or D
- **value** values given for these sources

**References**


**Examples**

```r
data(FourSources)
head(FourSources)
```

---

### NPIstats

**Description**

An implementation of the Nonparametric Predictive Inference approach in R. It provides tools for quantifying uncertainty via lower and upper probabilities. It includes useful functions for pairwise and multiple comparisons: comparing two groups with and without terminated tails, selecting the best group, selecting the subset of best groups, selecting the subset including the best group.
Nonparametric Predictive Inference (NPI) is a statistical method which uses few modelling assumptions, enabled by the use of lower and upper probabilities to quantify uncertainty. NPI has been presented for many problems in Statistics, Risk and Reliability and Operations Research. NPI approach is based on Hill’s assumption A(n), which gives a direct conditional probability for a future observable random quantity, conditional on observed values of related random quantities. Inferences based on A(n) are predictive and nonparametric, and can be considered suitable if there is hardly any knowledge about the random quantity of interest, other than the n observations, or if one does not want to use such information, e.g. to study effects of additional assumptions underlying other statistical methods. A(n) is not sufficient to derive precise probabilities for many events of interest, but it provides optimal bounds for probabilities for all events of interest involving the next future observation. These bounds are lower and upper probabilities in the theories of imprecise probability and interval probability, and as such they have strong consistency properties. NPI is a framework of statistical theory and methods that use these A(n)-based lower and upper probabilities, and also considers several variations of A(n) which are suitable for different inferences. For more info, visit NPI webpage.

References


select.best.groups

NPI for selecting the subset of best groups

Description

NPI lower and upper probabilities for the event that the next future observations from groups S are greater than all future observations from the other groups.

Usage

select.best.groups(data, S)
select.include.best

Arguments

data a list of numeric data vectors
S a vector of indices of the groups in the data list to be considered as the subset of best groups

Value

NPI lower and upper probabilities for the event that the next future observations from groups S are greater than all future observations from the other groups.

Examples

# NPI lower and upper probabilities for the event that
# the next future observations from groups 1 and 3 are greater than
# all future observations from the other groups.

data(FourSources)
# Convert the dataframe to a list of groups
data2<-split(FourSources$value,FourSources$source)
select.best.groups(data2,S=c(1,3))

select.include.best NPI for selecting the subset including the best group

Description

NPI lower and upper probabilities for the event that at least one of the next future observations from groups S is greater than all future observations from the other groups.

Usage

select.include.best(data, S)

Arguments

data a list of numeric data vectors
S a vector of indices of the groups in the data list to be considered as the subset of groups that includes the best group.

Value

NPI lower and upper probabilities for the event that at least one of the next future observations from groups S is greater than all future observations from the other groups.
**Examples**

NPI lower and upper probabilities for the event that the next future observation from one (the Sth) group is greater than all future observations from the other groups.

```r
data(FourSources)
data2<-split(FourSources$value,FourSources$source)
select.include.best(data2,S=c(1,3))
```

**Description**

NPI lower and upper probabilities for the event that the next future observation from one (the Sth) group is greater than all future observations from the other groups.

**Usage**

```r
select.the.best(data, S)
```

**Arguments**

- **data**: a list of numeric data vectors
- **S**: an index of the group in the data list to be considered as the best group

**Value**

NPI lower and upper probabilities for the event that the next future observation from the Sth group is greater than all future observations from the other groups.

**Examples**

```r
data(FourSources)
data2<-split(FourSources$value,FourSources$source)
select.the.best(data2, 2)
```
Index

* datasets
  BirthWeights, 4
  BreakdownTimes, 4
  ChemicalReaction, 5
  FourSources, 6

best.pair, 2
best.pair.tt, 3
BirthWeights, 4
BreakdownTimes, 4

ChemicalReaction, 5
FourSources, 6

NPIstats, 6

select.best.groups, 7
select.include.best, 8
select.the.best, 9