

Package ‘mispitools’

March 29, 2022

Title Missing Person Identification Tools

Version 0.2.1

Description Open-source software for computing Likelihood ratios thresholds and error rates in DNA kinship testing. Marsico FL. et al (2021) <[doi:10.1016/j.fsigen.2021.102519](https://doi.org/10.1016/j.fsigen.2021.102519)>.

License GPL (>= 3)

Encoding UTF-8

LazyData true

Imports forrel, pedtools, plotly, dplyr, highcharter, tidyr, tidyverse, DirichletReg, stats, purrr

RoxygenNote 7.1.2

URL <https://github.com/MarsicoFL/mispitools>

BugReports <https://github.com/MarsicoFL/mispitools/issues>

Depends R (>= 2.10)

NeedsCompilation no

Author Franco Marsico [aut, cre] (<<https://orcid.org/0000-0002-0740-5516>>)

Maintainer Franco Marsico <franco.lmarsico@gmail.com>

Repository CRAN

Date/Publication 2022-03-29 06:40:02 UTC

R topics documented:

Argentina	2
Asia	2
Austria	3
BosniaHerz	3
China	3
combLR	4
deplot	4
DeT	5
Europe	6

getfreqs	6
Japan	7
LRdist	7
LRprelim	8
makeMPprelim	9
makePOIgen	10
makePOIprelim	11
simLRgen	12
simLRprelim	13
Trates	14
USA	15

Index	16
--------------	-----------

Argentina	<i>STRs allelic frequencies from specified country.</i>
-----------	---

Description

STRs allelic frequencies from specified country.

Usage

Argentina

Format

A data frame allele frequencies

Asia	<i>STRs allelic frequencies from specified country.</i>
------	---

Description

A dataset of allele frequencies.

Usage

Asia

Format

A data frame allele frequencies

Austria	<i>STRs allelic frequencies from specified country.</i>
---------	---

Description

STRs allelic frequencies from specified country.

Usage

Austria

Format

A data frame allele frequencies

BosniaHerz	<i>STRs allelic frequencies from specified country.</i>
------------	---

Description

STRs allelic frequencies from specified country.

Usage

BosniaHerz

Format

A data frame allele frequencies

China	<i>STRs allelic frequencies from specified country.</i>
-------	---

Description

STRs allelic frequencies from specified country.

Usage

China

Format

A data frame allele frequencies

combLR	<i>Combine LRs: a function for combining LRs obtained from simulations.</i>
--------	---

Description

Combine LRs: a function for combining LRs obtained from simulations.

Usage

```
combLR(LRdatasim1, LRdatasim2)
```

Arguments

LRdatasim1	A data frame object with the results of simulations. Outputs from simLRgen or simLRprelim functions.
LRdatasim2	A second data frame object with the results of simulations. Outputs from simLRgen or simLRprelim functions.

Value

An object of class data.frame combining the LRs obtained from simulations (the function multiplies the LRs).

Examples

```
library(mispitools)
library(forrel)
x = linearPed(2)
x = setMarkers(x, locusAttributes = NorwegianFrequencies[1:5])
x = profileSim(x, N = 1, ids = 2)[[1]]
LRdatasim1 = simLRgen(x, missing = 5, 10, 123)
LRdatasim2 = simLRprelim("sex")
combLR(LRdatasim1, LRdatasim2)
```

deplot	<i>Decision making plot: a function for plotting false positive and false negative rates for each LR threshold.</i>
--------	---

Description

Decision making plot: a function for plotting false positive and false negative rates for each LR threshold.

Usage

```
deplot(datasim)
```

Arguments

`datasim` Input dataframe containing expected LR_s for related and unrelated POIs. It should be the output from `makeLRsims` function.

Value

A plot showing false positive and false negative rates for each likelihood ratio threshold.

Examples

```
library(forrel)
library(plotly)
x = linearPed(2)
x = setMarkers(x, locusAttributes = NorwegianFrequencies[1:5])
x = profileSim(x, N = 1, ids = 2)[[1]]
datasim = simLRgen(x, missing = 5, 10, 123)
deplot(datasim)
```

DeT *Decision Threshold: a function for computing likelihood ratio decision threshold.*

Description

Decision Threshold: a function for computing likelihood ratio decision threshold.

Usage

```
DeT(datasim, weight)
```

Arguments

`datasim` Input dataframe containing expected LR_s for related and unrelated POIs. It should be the output from `makeLRsims` function.

`weight` The differential weight between false positives and false negatives. A value of 10 is suggested.

Value

A value of Likelihood ratio suggested as threshold based on false positive-false negative trade-off.

Examples

```
library(forrel)
x = linearPed(2)
x = setMarkers(x, locusAttributes = NorwegianFrequencies[1:5])
x = profileSim(x, N = 1, ids = 2)[[1]]
datasim = simLRgen(x, missing = 5, 10, 123)
DeT(datasim, 10)
```

Europe	<i>STRs allelic frequencies from specified country.</i>
--------	---

Description

STRs allelic frequencies from specified country.

Usage

Europe

Format

A data frame allele frequencies

getfreqs	<i>Function for getting STR allele frequencies from different world populations.</i>
----------	--

Description

Function for getting STR allele frequencies from different world populations.

Usage

```
getfreqs(region)
```

Arguments

region	select the place of the allele frequency database. Possible values are listed: "Argentina", "Asia", "Europe", "USA", "Austria", "BosniaHerz", "China" and "Japan".
--------	--

Value

An allele frequency database adapted compatible with pedtools format.

Source

<https://doi.org/10.1016/j.fsigss.2009.08.178>; <https://doi.org/10.1016/j.fsigen.2016.06.008>; <https://doi.org/10.1016/j.fsigen.2016.06.008>

Japan	<i>STRs allelic frequencies from specified country.</i>
-------	---

Description

STRs allelic frequencies from specified country.

Usage

Japan

Format

A data frame allele frequencies

LRdist	<i>Likelihood ratio distribution: a function for plotting expected log10(LR) distributions under relatedness and unrelatedness.</i>
--------	---

Description

Likelihood ratio distribution: a function for plotting expected log10(LR) distributions under relatedness and unrelatedness.

Usage

```
LRdist(datasim, type = 1)
```

Arguments

datasim	Input dataframe containing expected LR _s for related and unrelated POIs. It should be the output from makeLRsims function.
type	Select between a density plot (type = 1, default) or a violin plot (type = 2).

Value

A plot showing likelihood ratio distributions under relatedness and unrelatedness hypothesis.

Examples

```
library(forrel)
x = linearPed(2)
x = setMarkers(x, locusAttributes = NorwegianFrequencies[1:5])
x = profileSim(x, N = 1, ids = 2)[[1]]
datasim = simLRgen(x, missing = 5, 10, 123)
LRdist(datasim)
```

LRprelim

Likelihood ratio for preliminary investigation data: a function for computing likelihood ratio based on preliminary investigation data.

Description

Likelihood ratio for preliminary investigation data: a function for computing likelihood ratio based on preliminary investigation data.

Usage

```
LRprelim(
  ABD = "1976-05-31",
  DBD = "1976-07-15",
  PrelimData,
  alpha = c(1, 4, 60, 11, 6, 4, 4),
  cuts = c(-120, -30, 30, 120, 240, 360),
  draw = 500,
  type = 1,
  seed = 123
)
```

Arguments

ABD	Actual birth date of the missing person.
DBD	Declared birth date of the person of interest.
PrelimData	Used when type = 2, is the dataframe with the DBD of the persons of interest in the database.
alpha	A vector containing the alpha values for the dirichlet. It should contain the number of categories of differences between DBD and ABD.
cuts	Value of differences between DBD and ABD used for category definition.
draw	Number of simulations for Dirichlet distribution computation.
type	Type of scenario, type 1 is an "open search", where it is unknown if the missing person is in the database. Type 2 refers to a scenario where the missing person is in the database.
seed	Seed for simulations.

Value

A value of Likelihood ratio based on preliminary investigation data. In this case, birth date.

Examples

```
library(DirichletReg)
LRprelim(ABD = "1976-05-31", DBD = "1976-07-15",
PrelimData, alpha = c(1, 4, 60, 11, 6, 4, 4),
cuts = c(-120, -30, 30, 120, 240, 360),
type = 1, seed = 123)
```

makeMPprelim	<i>Make preliminary investigation MP data simulations: a function for obtaining a database of preliminary investigation data for a missing person search.</i>
--------------	---

Description

Make preliminary investigation MP data simulations: a function for obtaining a database of preliminary investigation data for a missing person search.

Usage

```
makeMPprelim(
  casetype = "children",
  dateinit = "1975/01/01",
  scenario = 1,
  femaleprop = 0.5,
  ext = 100,
  numsims = 10000,
  seed = 123,
  region = c("North America", "South America", "Africa", "Asia", "Europe", "Oceania"),
  regionprob = c(0.2, 0.2, 0.2, 0.1, 0.2, 0.1)
)
```

Arguments

casetype	Type of missing person search case. Two options are available: "migrants" or "children".
dateinit	Minimum birth date of simulated missing person. Casetype: Children.
scenario	Birth date distribution scenarios: (1) non-uniform, (2) uniform. Casetype: Children.
femaleprop	Proportion of females. Casetype: All.
ext	Time extension for minimum birth date, range in scenario 1 and days in scenario 2. Casetype: Children.
numsims	Number of simulated MPs. Casetype: All.
seed	Select a seed for simulations. If it is defined, results will be reproducible. Casetype: All.
region	Birth region or place in missing children case or place of place of the last seen in missing migrant case. Casetype: All.
regionprob	Region proportions. Casetype: All.

Value

An object of class `data.frame` with preliminary investigation data.

Examples

```
makeMPprelim()
```

makePOIgen	<i>Make POIs gen: a function for obtaining a database with genetic information from simulated POIs or UHRs.</i>
------------	---

Description

Make POIs gen: a function for obtaining a database with genetic information from simulated POIs or UHRs.

Usage

```
makePOIgen(numsims = 100, reference, seed = 123)
```

Arguments

numsims	Number of simulations performed (number of POIs or UHRs).
reference	Indicate the reference STRs/SNPs frequency database used for simulations.
seed	Select a seed for simulations. If it is defined, results will be reproducible. Suggested, seed = 123

Value

An object of class `data.frame` with genetic information from POIs (randomly sampled from the frequency database).

Examples

```
library(forrel)
freqdata <- getfreqs(Argentina)
makePOIgen(numsims = 100, reference = freqdata, seed = 123)
```

makePOIprelim	<i>Make preliminary investigation POI/UHR data simulations: a function for obtaining a database of preliminary investigation data for a missing person search.</i>
---------------	--

Description

Make preliminary investigation POI/UHR data simulations: a function for obtaining a database of preliminary investigation data for a missing person search.

Usage

```
makePOIprelim(
  casetype = "children",
  dateinit = "1975/01/01",
  scenario = 1,
  femaleprop = 0.5,
  ext = 100,
  numsims = 10000,
  seed = 123,
  birthprob = c(0.09, 0.9, 0.01),
  region = c("North America", "South America", "Africa", "Asia", "Europe", "Oceania"),
  regionprob = c(0.2, 0.2, 0.2, 0.1, 0.2, 0.1)
)
```

Arguments

casetype	Type of missing person search case. Two options are available: "migrants" or "children".
dateinit	Minimum birth date of simulated persons of interest. Casetype: Children.
scenario	Birth date distribution scenarios: (1) non-uniform, (2) uniform. Casetype: Children.
femaleprop	Proportion of females. Casetype: All.
ext	Time extension for minimum birth date, range in scenario 1 and days in scenario 2. Casetype: Children.
numsims	Number of simulated POIs/UHRs. Casetype: All.
seed	Select a seed for simulations. If it is defined, results will be reproducible. Casetype: All.
birthprob	Birth type probabilities: home birth, hospital birth and unknown-adoption. Casetype: Children.
region	Birth region or place in missing children case or place of discovery of the human remain in missing migrant case. Casetype: All.
regionprob	Region proportions. Casetype: All.

Value

An object of class data.frame with preliminary investigation data.

Examples

```
makePOIprelim(
  dateinit = "1975/01/01",
  scenario = 1,
  femaleprop = 0.5,
  ext = 100,
  numsims = 10000,
  seed = 123,
  birthprob = c(0.09, 0.9, 0.01),
  region = c("North America", "South America", "Africa", "Asia", "Europe", "Oceania"),
  regionprob = c(0.2, 0.2, 0.2, 0.1, 0.2, 0.1))
```

simLRgen	<i>Simulate likelihoods ratio (LRs) based on genetic data: a function for obtaining expected LRs under relatedness and unrelatedness kinship hypothesis.</i>
----------	--

Description

Simulate likelihoods ratio (LRs) based on genetic data: a function for obtaining expected LRs under relatedness and unrelatedness kinship hypothesis.

Usage

```
simLRgen(reference, missing, numsims, seed)
```

Arguments

reference	Reference pedigree. It could be an input from read_fam() function or a pedigree built with pedtools.
missing	Missing person ID/label indicated in the pedigree.
numsims	Number of simulations performed.
seed	Select a seed for simulations. If it is defined, results will be reproducible. Suggested, seed = 123

Value

An object of class data.frame with LRs obtained for both hypothesis, Unrelated where POI is not MP or Related where POI is MP.

Examples

```

library(forre1)
x = linearPed(2)
x = setMarkers(x, locusAttributes = NorwegianFrequencies[1:5])
x = profileSim(x, N = 1, ids = 2)[[1]]
plot(x)
datasim = simLRgen(x, missing = 5, 10, 123)

```

simLRprelim	<i>Simulate likelihoods ratio (LRs) based on preliminary investigation data: a function for obtaining expected LRs under relatedness and unrelatedness kinship hypothesis.</i>
-------------	--

Description

Simulate likelihoods ratio (LRs) based on preliminary investigation data: a function for obtaining expected LRs under relatedness and unrelatedness kinship hypothesis.

Usage

```

simLRprelim(
  vartype,
  numsims = 1000,
  seed = 123,
  int = 5,
  ErrorRate = 0.05,
  alphaBdate = c(1, 4, 60, 11, 6, 4, 4),
  numReg = 6,
  MP = NULL,
  database,
  cuts = c(-120, -30, 30, 120, 240, 360)
)

```

Arguments

vartype	Indicates type of preliminary investigation variable. Options are: sex, region, age, birthDate and height.
numsims	Number of simulations performed.
seed	Seed for simulations.
int	Interval parameter, used for height and age vartypes. It defines the estimation range, for example, if MP age is 55, and int is 10, the estimated age range will be between 45 and 65.
ErrorRate	Error rate for sex, region, age and Height LR calculations.
alphaBdate	Vector containing alpha parameters for Dirichlet distribution. Usually they are the frequencies of the solved cases in each category.

numReg	Number of regions present in the case.
MP	Introduce the preliminary data of the selected variable (vartype) of the MP. If it is null, open search is carried out. If it is not NULL, close search LR is computed. Variables values must be named as those presented in makePOIprelim function.
database	It is used when the close search (MP not NULL), is carried out. It could be the output from makePOIprelim or a database with the same structure.
cuts	Value of differences between DBD and ABD used for category definition. They must be the same as the ones selected for alphaBdate vector.

Value

An object of class data.frame with LRs obtained for both hypothesis, Unrelated where POI/UHR is not MP or Related where POI/UHR is MP.

Examples

```
library(mispitools)
simLRprelim("sex")
```

Trates	<i>Threshold rates: a function for computing error rates and Matthews correlation coefficient of a specific LR threshold.</i>
--------	---

Description

Threshold rates: a function for computing error rates and Matthews correlation coefficient of a specific LR threshold.

Usage

```
Trates(datasim, threshold)
```

Arguments

datasim	Input dataframe containing expected LRs for related and unrelated POIs. It should be the output from makeLRsims function.
threshold	Likelihood ratio threshold selected for error rates calculation.

Value

Values of false positive and false negative rates and MCC for a specific LR threshold.

Examples

```
library(forre1)
x = linearPed(2)
x = setMarkers(x, locusAttributes = NorwegianFrequencies[1:5])
x = profileSim(x, N = 1, ids = 2)[[1]]
datasim = simLRgen(x, missing = 5, 10, 123)
Trates(datasim, 10)
```

USA

STRs allelic frequencies from specified country.

Description

STRs allelic frequencies from specified country.

Usage

USA

Format

A data frame allele frequencies

Index

* datasets

Argentina, [2](#)
Asia, [2](#)
Austria, [3](#)
BosniaHerz, [3](#)
China, [3](#)
Europe, [6](#)
Japan, [7](#)
USA, [15](#)

Argentina, [2](#)
Asia, [2](#)
Austria, [3](#)

BosniaHerz, [3](#)

China, [3](#)
combLR, [4](#)

deplot, [4](#)
DeT, [5](#)

Europe, [6](#)

getfreqs, [6](#)

Japan, [7](#)

LRdist, [7](#)
LRprelim, [8](#)

makeMPprelim, [9](#)
makePOIgen, [10](#)
makePOIprelim, [11](#)

simLRgen, [12](#)
simLRprelim, [13](#)

Trates, [14](#)

USA, [15](#)