Package ‘VirtualPop’

October 12, 2022

Type Package

Title Simulation of Populations by Sampling Waiting-Time Distributions

Version 1.0.2

Imports msm,HMDHFDplus

Suggests knitr, rmarkdown,ggplot2,foreign,lubridate,xml2,remotes

BuildResaveData best

VignetteBuilder knitr

LazyData true

Date 2022-06-23

Maintainer Frans Willekens `<willekens@nidi.nl>`

Description

Generates lifespans and fertility histories in continuous time using individual-level state transition (multi-state) models and data from the Human Mortality Database and the Human Fertility Database. To facilitate virtual population analysis, data on virtual individuals are stored in a data structure commonly used in sample surveys. Life histories are generated for multiple generations. The genealogies that result facilitate the study of family ties.

License GPL-2

NeedsCompilation no

Depends R (>= 3.5.0)

Encoding UTF-8

BugReports `https://github.com/willekens/VirtualPop/issues`

RoxygenNote 7.2.0

Author Frans Willekens [aut, cre] (<`https://orcid.org/0000-0001-6125-0212`>),
Tim Riffe [ctb] (<`https://orcid.org/0000-0002-2673-4622`>)

Repository CRAN

Date/Publication 2022-06-23 18:10:15 UTC

R topics documented:

`Children` *Generates Individual Fertility Histories, Using Function Sim_bio.*

Description

Individual fertility histories

Usage

```
Children(dat0, rates)
```

Arguments

`dat0`	Data frame with base individual data on members of virtual population
`rates`	Mortality and fertility rates. The object 'rates' is produced by Getrates_refyear.R

Value

List object with two objects: (a) data frame with individual info and fertility history of egos and (b) children data frame

Author(s)

Frans Willekens

Examples

```
utils::data(dataLH)
utils::data(rates)
dat0 <- dataLH[1:10,]
out <- Children(dat0=dat0,rates=rates)
```

`dataLH` *dataLH data*

Description

simulated population of four generations

Format

A data frame with data on 29954 individuals (10000 in initial cohort).

ID Identification number

gen Generation

sex Sex. A factor with levels Males and Females

bdated Date of birth (decimal date

ddated Date of death (decimal date

x_D Age at death (decimal number

IDpartner ID of partner

IDmother ID of mother

IDfather ID of father

jch Child's line number in the household

nch Number of children ever born

id.1 ID of first child

id.2 ID of 2nd child

id.3 ID of 3rd child

id.4 ID of 4th child

id.5 ID of 5th child

id.6 ID of 6th child

id.7 ID of 7th child

id.8 ID of 8th child

id.9 ID of 9th child

age.1 Age of mother at birth of first child

age.2 Age of mother at birth of 2nd child

age.3 Age of mother at birth of 3rd child

age.4 Age of mother at birth of 4th child

age.5 Age of mother at birth of 5th child

age.6 Age of mother at birth of 6th child

age.7 Age of mother at birth of 7th child

age.8 Age of mother at birth of 8th child

age.9 Age of mother at birth of 9th child

Source

Simulation uses period mortality rates and fertility rates by birth order from the United States 2019. The data are downloaded from the Human Mortality Database (HMD) and the Human Fertility Database (HFD).

`dpopus` *dpopus data Population of the United States in 2019 reported in the HMD (Population.txt file)*

Description

dpopus data

Population of the United States in 2019 reported in the HMD (Population.txt file)

Format

A data frame with 111 age groups (single years of age).

Females Female population

Males Male population

Source

The data are downloaded from the Human Mortality Database (HMD). Country: USA. Year: 2019

`GetData` *Reads data from the HMD and HFD*

Description

Reads data from the HMD and HFD

Usage

```
GetData(country, user, pw_HMD, pw_HFD)
```

Arguments

`country`	country
`user`	Name of the user, used at registration with the HMD and HFD. It is assumed that the same name is used for both HMD and HFD.
`pw_HMD`	Password to access HMD, provided at registration
`pw_HFD`	Password to access HFD, provided at registration

Value

`data_raw`	5 objects: country,life tables females,life tables males,fertility rates,female population (from HFD): exposures

Author(s)

Frans Willekens

Examples

```
## Not run: dataLH <- GetData(country="USA",user,pw_HMD,pw_HFD)
```

`GetGenerations` *Creates Database 'dataLH' from Mortality Rates and Fertility Rates*

Description

Creates database 'dataLH' from mortality rates by age and sex, and fertility rates by age of mother and birth order

Usage

```
GetGenerations(
  rates,
  ncohort,
  ngen,
  age_end_perc = NULL,
  iages = NULL,
  ID1 = NULL
)
```

Arguments

`rates`	List object with death rates (ASDR) and birth rates (ASFR)
`ncohort`	Size of hypothetical birth cohort
`ngen`	Number of generations to be simulated
`age_end_perc`	If age_end_perc is not missing (NULL), then the simulated ages at death are replaced by the age distribution given by age_end_perc. The age distribution is a matrix with 2 dolumns, one for females (column 1) and one for males (column 2). The distribution is given by single years of age.
`iages`	If iages is not missing, the vector of simulated ages at death is replaced by the vector of individual ages at censoring
`ID1`	Identification number of first person in virtual population being created (optional)

Details

age_end_prec or iages are used to simulate ages at censoring. For instance, to compare the virtual population with a real population for which information is collected retorspectively in a cross-sectional survey, the simulation window must be equal to the observation window. In other words, the virtual population and the real population must have the same censoring.

Value

`dataLH`	The database of simulated individual lifespans and fertility histories. The object 'dataLH' has two attributes: (a) the calendar year of period rates and (b) the country

Author(s)

Frans Willekens

Examples

```
# The object rates is produced by the function GetRates.
utils::data(rates)
dLH <- GetGenerations (rates=rates,ncohort=100,ngen=4)
```

`GetRates` *Retrieves Data from HMD and HFD for a Selected Country (All Years)*

Description

(a) Retrieves rates, the period life tables and the period fertility tables. (b) Computes death rates by age and sex, and birth rates by age and birth order.

Usage

```
GetRates(data, refyear)
```

Arguments

`data`	data
`refyear`	Reference year, which is the year of period data

Details

The user needs to register as a new user before data can be downloaded. To register with HMD, go to https://www.mortality.org. To register with HFD, go to https://www.humanfertility.org/cgi-bin/main.php.

Value

`ASDR`	Age-specific death rates, by sex (for reference year or all years)
`ASFR`	Age-specific birth rates by birth order (for reference year or all years)
`e0`	REMOVE

Note

To access the HMD and HFD, the function used HMDHFDplus written by Tim Riffe and other at the Max Planck Institute for Demographic Research, Rostock, Germany

Author(s)

Frans Willekens

Examples

```
## Not run: ratesR <- GetRates(data,refyear)
```

`H_pw` *Computes cumulative hazard at duration t.*

Description

Computes cumulative hazard at duration t from age-specific demographic rates.

Usage

```
H_pw(t, breakpoints, rates)
```

Arguments

`t`	Duration at which cumulative hazard is required.
`breakpoints`	Breakpoints: values of x at which piecewise-constant rates change.
`rates`	Piecewise-constant rates

Value

Cumulative hazard at duration t

Author(s)

Frans Willekens

See Also

Function H_pw called by pw_root, which is called by r_pw_exp.

Examples

```
breakpoints <- c(0, 10, 20, 30, 60)
rates <- c(0.01,0.02,0.04,0.15)
z <- H_pw(t=0:40, breakpoints=breakpoints, rates=rates)


utils::data(rates)
ages <- as.numeric(rownames(rates$ASDR))
breakpoints <- c(ages,120)
zz <- H_pw(t=ages, breakpoints=breakpoints, rates=rates$ASDR[,1])
```

`Lifespan` *Generates Individual Lifespan(s)*

Description

Simulate length of life using age-specific death rates.Generate date of death and age at death. The function uses the rpexp function from the package msm and uniroot of base R

Usage

```
Lifespan(data, ASDR)
```

Arguments

`data`	Data frame with individual data
`ASDR`	Age-specific death rates

Value

data: data frame 'dataLH' with date of death and age of death completed.

Author(s)

Frans Willekens

Examples

```
utils::data(dataLH)
utils::data(rates)
z <- Lifespan (dataLH[1:5,],ASDR=rates$ASDR)
```

`Partnership` *Allocates Partners to Members of Virtual Population.*

Description

Randomly allocates partners to egos

Usage

```
Partnership(dLH)
```

Arguments

`dLH`	Database

Value

Updated version of database (dLH), which includes the IDs of partners.

Author(s)

Frans Willekens

Examples

```
utils::data(dataLH)
dLH=dataLH[1:10,]
# Remove current partner
dLH$IDpartner <- NA
d <- Partnership(dLH=dLH)
# NOTE: partners are randomly selected from the individuals documented in dLH.
```

`pw_root` *Equation for which root must be determined.*

Description

Equation: cumulative hazard functionn + log(uu) = 0

Usage

```
pw_root(t, breakpoints, rates, uu)
```

Arguments

`t`	Vector of durations to be considered in determining root.
`breakpoints`	Breakpoints
`rates`	Piecewise-constant rates
`uu`	Random draw from standard uniform distribution.

Details

The function is called by function uniroot (base R), which is called by r.pw_exp

Value

Vector of differences between cumulative hazard and -log(uu) for different values of t.

Author(s)

Frans Willekens

See Also

Functions H_pw and r.pw_exp

Examples

```
breakpoints <- c(0, 10, 20, 30, 60)
rates <- c(0.01,0.02,0.04,0.15)
z <- pw_root (t= c(10,18.3,23.6,54.7),breakpoints,rates,uu=0.43)
```

`r.pw_exp` *Sample from a piecewise-constant exponential distribution.*

Description

Takes n random draws from a piecewise-constant exponential distribution.

Usage

```
r.pw_exp(n, breakpoints, rates)
```

Arguments

`n`	Number of random draws required
`breakpoints`	Breakpoints in piecewise-constant exponential distribution
`rates`	Piecewise-constant rates

Value

Vector of waiting times, drawn from piecewise-exponential survival function.

Author(s)

Frans Willekens

Examples

```
breakpoints <- c(0, 10, 20, 30, 60)
rates <- c(0.01,0.02,0.04,0.15)
pw_sample <- r.pw_exp (n=10, breakpoints, rates=rates)
```

`rates` *rates data*

Description

Mortality rates by age and sex: fertility rates by age and birth order

Format

A list of three objects.

ASDR Mortality rates

ASFR Fertility rates

ratesM Multistate transition rates

Source

The data are downloaded from the Human Mortality Database (HMD) and the Human Fertility Database (HFD). Country: USA. Year: 2019

`Sim_bio` *Generic Function to Generate Single Life History*

Description

The function is called from the function Children. It uses the rpexp function of the msm package.

Usage

```
Sim_bio(datsim, ratesM)
```

Arguments

`datsim`	Data frame with individual data
`ratesM`	Multistate transition rates in standard (multistate) format

Value

`age_startSim`	Age at start of simulation
`age_endSim`	Age at end of simulation
`nstates`	Number of states
`path`	path: sequence of states occupied
`ages_trans`	Ages at transition

Author(s)

Frans Willekens

Examples

```
# Generates single fertility history from mortality rates by age
# and fertility rates by age and parity
# Fertily history is simulated from starting age to ending age
# Individual starts in state "par0"
# ratesM is an object with the rates in the proper format for multistate analysis
utils::data(rates)
popsim <- data.frame(ID=1,born=2000.450,start=0,end=80,st_start="par0")
ch <- Sim_bio (datsim=popsim,ratesM=rates$ratesM)
```

Index