Package 'LCRFsim'

July 27, 2019

Title Lung Cancer Risk Factor Simulator (LCRFsim)

Version 0.9.7.00

Description

This package was developed to generate risk factors for lung cancer, conditioned on smoking history. These risk factors include BMI, personal history, family history, and COPD. The smoking histories can be simulated using the Smoking History Generator (SHG), a program maintained by CISNET. This package contains a wrapper to run the SHG, and functions to simulate the set of correlated risk factors for lung cancer from the SHG cohort, given gender, and the year of birth of the cohort. Updated to include models for the 1950 and 1960 cohorts.

Depends R (>= 3.5.0)

License GPL

Encoding UTF-8

LazyData TRUE

Imports splines

RoxygenNote 6.1.1

R topics documented:

`processSHG` *Process raw SHG output*

Description

The SHG creates an out_1950.out file which is the raw output. The following function runs python code that decomposes that output into a few other files that will create an R-readable table. A SHG file is saved to "smoking_histories.csv" in the working directory. The function also returns the smoking history as a data frame object in R. If you have a output_XXXX.out file from your own SHG simulations, you can process it using this function to load an R object that can be used for simulating the risk factors.

Usage

```
processSHG(file, birth_cohort)
```

Arguments

file	the location of the output_1950.out file (ie: "shg/output_1950.out")
birth_cohort	the year of birth for the cohort (ie: 1950)

Examples

```
smoking_history <- processSHG(file = "~/SHG3.6.4/output_1950.out", birth_cohort = 1950)
```

---

riskFactorSimulator *Risk Factor Simulator*

Description

For a given SHG cohort with a specific birth year, for a given gender ("F" or "M"), for a range of ages (ie: 45:90), this function will return a simulated set of risk factors for lung cancer (BMI, family history, personal history, COPD) for each individual in the SHG cohort, conditioned on their smoking histories provided by runSHG(). Use this function after you have created a SHG cohort using runSHG(). You can also load a sample processed dataset of the SHG by typing data(smkhist). smkhist is a processed SHG data frame of 10000 men born in 1950 simulated using SHG3.6.4. This function requires a processed SHG object (SHG_out) that is created by processsSHG(). If you have your own simulated SHG cohort (as an output_19X0.out file), you can process that using processSHG() to create a data.frame that can be inputted into this function.

Usage

```
riskFactorSimulator(gender, birth_cohort, SHG_out, ages, seed)
```

Arguments

gender	the gender (ie: "M" or "F")
birth_cohort	the birth year of the cohort (ie: 1950)
SHG_out	the name of the object that is the output from the SHG
ages	the range of longitudinal ages to be passed to reformatSHG (ie: 45:90)
seed	the randomization seed

Examples

```
library(LCRFsim)
library(splines)
runSHG("~/SHG3.6.4", 10000, "M", 1950, 1) # runs SHG to create output_1950.out in SHG directory
smoking_history <- processSHG(file = "~/SHG3.6.4/output_1950.out", birth_cohort = 1950)
data(smkhist) # or you can load a sample processed SHG output of 10K men born in 1950
ANS_m <- riskFactorSimulator(gender="M", birth_cohort = 1950, SHG_out = smkhist, ages = 45:90, seed = 1618)
summary(ANS_m$outputOnly[,1:5]) # look at results of simulation
```

runSHG *Runs Smoking History Generator*

Description

The SHG is written in Python and is maintained seperately from this package by CISNET. This function calls the SHG from within R, creates the SHG output_19X0.out file, and also returns the SHG cohort. The SHG needs to be installed properly prior to use. To use runSHG(), you pass it the directory where the SHG is installed. Tested using SHG6.3.4 and Python v2.7.10 on MacOSX. SHG6.3.4.zip was obtained from cisnet.flexkb.net, see their readme.md for setup instructions. In python, you'll likely want these dependencies: six, ipdb, enum, ipython_genutils, ipython. To install the SHG, run install.sh. To run the SHG from R, run: (1) python run_tests.py directly, or use this R wrapper function to create a custom 'run_custom_tests.py'. No object is returned by this wrapper function. You will be prompted to look for the appropriate output_19X0.out, input_19X0.txt, and errors_19X0.txt in the SHG directory. If you do not wish to use the SHG, you can load a processed smoking history dataset by typing: data(smkhist) that is included with this package. The smkhist dataset is for a 1950 birth cohort of 10000 men from SHG3.6.4.

Usage

```
runSHG(path, n, gender, birth_cohort, seed)
```

Arguments

path	a string indicating the directory that the SHG is installed in (ie: "~/SHG6.3.4")
n	the size of the cohort (ie: 10000)
gender	the gender (ie: "M" == 0 or "F" == 1)
birth_cohort	the year of birth for the cohort (ie: 1950 or 1960)
seed	the seed for random number generator

Examples

```
library(LCRFsim)
runSHG("~/SHG3.6.4", 10000, "F", 1950, 1) # runs SHG to create output_1950.out in SHG directory
smoking_history <- processSHG(file = "~/SHG3.6.4/output_1950.out", birth_cohort = 1950)
data(smkhist) # or you can load a sample processed SHG output
```

smkhist *Sample SHG data*

Description

This is a sample processed SHG dataset of 10,000 men born in 1950. The code for generating is shown below. The cohort was generated using SHG 6.3.4

Examples

```
library(LCRFsim)
library(splines)
runSHG("~/SHG3.6.4", 10000, "M", 1950, 1) # runs SHG to create output_1950.out in SHG directory
smkhist <- processSHG(file = "~/SHG3.6.4/output_1950.out", birth_cohort = 1950)
OUT_m <- riskFactorSimulator(gender="M", birth_cohort = 1950, SHG_out = smkhist, ages = 45:90, seed = 1618)
summary(OUT_m$outputOnly[,1:5]) # look at results of simulation
```

Index