Package 'LTASR'

April 13, 2022

Title Functions to Replicate the Center for Disease Control and Prevention's 'LTAS' Software in R

Version 0.0.1

Description A suite of functions for reading in a rate file in XML format, stratify a cohort, and calculate 'SMRs' from the stratified cohort and rate file.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.1.2

Imports dplyr, knitr, lubridate, magrittr, purrr, readr, rlang, stringr, tidyr, XML, zoo

Suggests rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

Depends R (>= 2.10)

LazyData true

Config/testthat/edition 3

NeedsCompilation no

Author Stephen Bertke [aut, cre]

Maintainer Stephen Bertke <nioshltas@cdc.gov>

Repository CRAN

Date/Publication 2022-04-13 09:02:29 UTC

R topics documented:

---

checkPerson *Check person dataframe*

Description

checkPerson checks that the person dataframe is formatted correctly

Usage

```
checkPerson(person)
```

Arguments

person	person dataframe

Value

No return value, called for side effects

Examples

```
library(LTASR)
library(dplyr)

#Import example peron file
person <- person_example %>%
  mutate(dob = as.Date(dob, format='%m/%d/%Y'),
         pybegin = as.Date(pybegin, format='%m/%d/%Y'),
         dlo = as.Date(dlo, format='%m/%d/%Y'))

#Returns nothing
checkPerson(person)
```

checkStrata *Checks all strata in py_table are contained in rate file*

Description

Checks all strata in py_table are contained in rate file

Usage

```
checkStrata(py_table, rateobj)
```

Arguments

py_table	A stratified cohort created by get_table
rateobj	A rate object created by parseRate

Value

A list containing:

1. The py_table with strata removed not found in rateobj
2. The observations from py_table that were removed

Examples

```
library(LTASR)
library(dplyr)

#Import example person file
person <- person_example %>%
  mutate(dob = as.Date(dob, format='%m/%d/%Y'),
         pybegin = as.Date(pybegin, format='%m/%d/%Y'),
         dlo = as.Date(dlo, format='%m/%d/%Y'))

#Import default rate object
rateobj <- us_119ucod_19602020

#Stratify person table
py_table <- get_table(person, rateobj)

#Check Strata are in rate file
checkStrata(py_table, rateobj)
```

expand_dates *Expand a data.frame to include all dates between a start and end value defined by parameters x and y*

Description

Expand a data.frame to include all dates between a start and end value defined by parameters x and y

Usage

```
expand_dates(df, start, end)
```

Arguments

df	Input data.frame
start	start date
end	end date

Value

A data.frame/tibble containing all variables of the input data.frame as well as a new variable, date, with repeated rows for each date between start and end.

Examples

```
library(LTASR)
data <- data.frame(start = as.Date('3/1/2015', format='%m/%d/%Y'),
                   end = as.Date('3/15/2015', format='%m/%d/%Y'))
expand_dates(data, start, end)
```

get_table *Stratify Person Table*

Description

get_table reads in a data.frame/tibble containing basic demographic information for each person of the cohort and stratifies the person-time and deaths into 5-year age, 5-year calendar period, race, and sex strata. See Details for information on how the person file must be formatted.

Usage

```
get_table(persondf, rateobj, batch_size = 25, strata = dplyr::vars())
```

Arguments

persondf	data.frame like object containing one row per person with the required demographic information
rateobj	a rate object created by the parseRate function
batch_size	a number specifying how many persons to stratify at a time. Default is 25.
strata	any additional variables contained in persondf on which to stratify. Must be wrapped in a vars() call from dplyr.

Details

The person file must contain the variables:

- id,
- gender (numeric: 'M'/'F'),
- race (numeric: 'W'/'N'),
- dob (date),
- pybegin (date),
- dlo (date),
- rev (numeric: 5-10),
- code (character: ICD code)

Value

A data.frame with a row for each strata containing the number of observed deaths within each of the defined minors/outcomes (_o1-_oxxx) and the number of person days.

Examples

```
library(LTASR)
library(dplyr)

#Import example person file
person <- person_example %>%
  mutate(dob = as.Date(dob, format='%m/%d/%Y'),
         pybegin = as.Date(pybegin, format='%m/%d/%Y'),
         dlo = as.Date(dlo, format='%m/%d/%Y'))

#Import default rate object
rateobj <- us_119ucod_19602020

#Stratify person table
py_table <- get_table(person, rateobj)
```

`get_table_history` *Stratify Person Table with Time Varying Co-variate*

Description

`get_table_history` reads in a data.frame/tibble containing basic demographic information for each person of the cohort as well as a data.frame/tibble containing time varying exposure information and stratifies the person-time and deaths into 5-year age, 5-year calendar period, race, sex and exposure categories. See `Details` for information on how the person file and history file must be formatted.

Usage

```
get_table_history(
  persondf,
  rateobj,
  batch_size = 25,
  strata = dplyr::vars(),
  historydf,
  exp,
  cutpt,
  lag = 0
)
```

Arguments

`persondf`	data.frame like object containing one row per person with the required demographic information
`rateobj`	a rate object created by the `parseRate` function
`batch_size`	a number specifying how many persons to stratify at a time. Default is 25.
`strata`	any additional variables contained in `persondf` on which to stratify. Must be wrapped in a `vars()` call from `dplyr`.
`historydf`	data.frame like object containing temporal exposure data. Each row represents a period of time during which the values f the temporal stratifiers remain constant. Multiple rows per `id` are typical.
`exp`	the name of the exposure variable found within `historydf`
`cutpt`	a numeric vector containing the cut-points to use to stratify the calculated cumulative exposure for variable `exp`. Should include min and max values (typically -Inf and Inf).
`lag`	An optional numeric variable indicating a lag, in years, to be applied to exposure variables. Specified in years. Default is 0 yrs (un-lagged).

Details

The person file must contain the variables:

- id,
- gender (numeric: 'M'/'F'),
- race (numeric: 'W'/'N'),
- dob (date),
- pybegin (date),
- dlo (date),
- rev (numeric: 5-10),
- code (character: ICD code)

Additionally, a history file is read in containing daily exposures levels and the begin and end dates of that level. The history file must contain the following variables:

- id,
- begin_dt (date),
- end_dt (date),
- <daily exposure levels>

Value

A data.frame with a row for each strata containing the number of observed deaths within each of the defined minors/outcomes (_o1-_oxxx) and the number of person days.

Examples

```
library(LTASR)
library(dplyr)
library(lubridate)

#Import example person file
person <- person_example %>%
  mutate(dob = as.Date(dob, format='%m/%d/%Y'),
         pybegin = as.Date(pybegin, format='%m/%d/%Y'),
         dlo = as.Date(dlo, format='%m/%d/%Y'))

#Import example history file
history <- history_example %>%
  mutate(begin_dt = as.Date(begin_dt, format='%m/%d/%Y'),
         end_dt = as.Date(end_dt, format='%m/%d/%Y'))

#Import default rate object
rateobj <- us_119ucod_19602020

#Stratify cohort employed variable. The employed variable indicates (0/1) periods of employment
#and will be summed each day of each exposure period. Therefore, this calculates
#duration of employment in days. The cut-points used below will stratify
```

```
#by person-time with less than and greater than a year of employment (365 days of employment).
py_table <- get_table_history(persondf = person,
                              rateobj = rateobj,
                              historydf = history,
                              exp = employed,
                              cutpt = c(-Inf, 365, Inf))

#Investigate cumulative exposure values for the exposure_level variable.
#This aids in determining appropriate cut-points for stratification.
history_example %>%
  mutate(begin_dt = as.Date(begin_dt, format='%m/%d/%Y'),
         end_dt = as.Date(end_dt, format='%m/%d/%Y')) %>%
  mutate(dur = difftime(end_dt, begin_dt, units = 'days') %>%
           as.numeric() %>%
           `+`(1),
         cum = dur*exposure_level) %>%
  group_by(id) %>%
  summarize(cum = sum(cum))

#Stratify cohort with exposure variable. This will stratify by person-time with
#with less than and greater than 15000 unit-days of cumulative exposure.
py_table <- get_table_history(persondf = person,
                              rateobj = rateobj,
                              historydf = history,
                              exp = exposure_level,
                              cutpt = c(-Inf, 15000, Inf),
                              lag = 10)
```

`history_example` *Example History File for Testing*

Description

A tibble containing example history file data to be used for testing and demonstration of the package

Usage

```
history_example
```

Format

A data frame with 4 rows and 5 variables:

id unique identifier; numeric

begin_dt beginning date of an exposure period; character

end_dt beginning date of an exposure period; character

employed a hypothetical variable indicating employment during the given exposure period; numeric (0/1)

exposure_level a hypothetical variable identifying daily exposure levels to be summed to calculate a cumulative exposure; numeric

...

Source

Internally Generated

`mapDeaths` *Map ICD codes to grouped minors*

Description

Map ICD codes to grouped minors

Usage

```
mapDeaths(persondf, rateobj)
```

Arguments

`persondf`	Person data.frame
`rateobj`	A rate object created from `parseRate`

Value

A data.frame for each death observed in the person file with the following variables: id, code, rev: from the `persondf` minor: the minor/outcome from the rate file that the death was mapped to

Examples

```
library(LTASR)

#Import example person file
person <- person_example

#Import default rate object
rateobj <- us_119ucod_19602020

#Check mapping of deaths to minors/outcomes
mapDeaths(person, rateobj)
```

parseRate *Parses LTAS rate file in .xml format*

Description

Parses LTAS rate file in .xml format

Usage

```
parseRate(xmlpath)
```

Arguments

xmlpath	path of LTAS rate file

Value

returns a list containing:

1. $residual: the minor number where all unknown deaths will be assigned
2. $MinorDesc: a dataframe/tibble giving descriptions of minor numbers as well as how minors are mapped to majors
3. $mapping: a dataframe/tibble listing how each icd-code and revision will be mapped to each minor number

person_example *Example Person File for Testing*

Description

A tibble containing example person file data to be used for testing and demonstration of the package

Usage

```
person_example
```

Format

A tibble with 3 observations and 9 variables:

id unique identifier; character

gender Gender/Sex; character 'M' or 'F'

race Race; character 'W' or 'N'

dob Date of Birth; character to be converted to date

pybegin date to begin follow-up/at-risk accumulation, character to be converted to date

dlo Date last observed; character to be converted to date

vs indicator identifying the vital status of the cohort. A value of 'D' indicates an observed death; character

rev ICD revision of the ICD code; numeric

code ICD-code for the cause of death; character ...

Source

Internally Generated

`smr_custom` *Calculate SMRs for Custom minor groupings*

Description

`smr_major` will collapse minor outcomes into "major" groupings as defined in the rate object, `rateobj`.

Usage

```
smr_custom(smr_minor_table, minor_grouping)
```

Arguments

`smr_minor_table`	A data.frame/tibble as created by `smr_minor` containing observed and expected number of deaths for each minor outcome
`minor_grouping`	A numeric vector defining which minors to group together

Value

A data.frame/tibble containing the expected and observed number of deaths as well the SMR, lower CI and upper CI for the outcome by the user

Examples

```
library(LTASR)
library(dplyr)

#Import example person file
person <- person_example %>%
  mutate(dob = as.Date(dob, format='%m/%d/%Y'),
         pybegin = as.Date(pybegin, format='%m/%d/%Y'),
         dlo = as.Date(dlo, format='%m/%d/%Y'))

#Import default rate object
rateobj <- us_119ucod_19602020
```

```
#Stratify person table
py_table <- get_table(person, rateobj)

#Calculate SMRs for all minors
smr_minor_table <- smr_minor(py_table, rateobj)

#Calculate custom minor grouping for all deaths
smr_custom(smr_minor_table, 1:119)

#' #Calculate custom minor grouping for all deaths
smr_custom(smr_minor_table, 4:40)
```

`smr_major` *Calculate SMRs for Major groupings*

Description

`smr_major` will collapse minor outcomes into "major" groupings as defined in the rate object, `rateobj`.

Usage

```
smr_major(smr_minor_table, rateobj)
```

Arguments

`smr_minor_table`	A data.frame/tibble as created by `smr_minor` containing observed and expected number of deaths for each minor outcome
`rateobj`	A rate object created by `parseRate`

Value

A data.frame/tibble containing the expected and observed number of deaths as well as SMRs, lower CI and upper CI for each major as defined in the rate object `rateobj`

Examples

```
library(LTASR)
library(dplyr)

#Import example person file
person <- person_example %>%
  mutate(dob = as.Date(dob, format='%m/%d/%Y'),
         pybegin = as.Date(pybegin, format='%m/%d/%Y'),
         dlo = as.Date(dlo, format='%m/%d/%Y'))

#Import default rate object
```

```
rateobj <- us_119ucod_19602020

#Stratify person table
py_table <- get_table(person, rateobj)

#Calculate SMRs for all minors
smr_minor_table <- smr_minor(py_table, rateobj)

#Calculate SMRs major groupings found within rate file
smr_major(smr_minor_table, rateobj)
```

`smr_minor` *Calculate SMRs for Minors*

Description

`smr_minor` calculates SMRs for all minor groupings found within the rate object, `rateobj`, for the stratified cohort `py_table`

Usage

```
smr_minor(py_table, rateobj)
```

Arguments

`py_table`	A stratified cohort created by `get_table`
`rateobj`	A rate object created by `parseRate`

Value

A dataframe/tibble containing the expected and observed number of deaths as well as SMRs, lower CI and upper CI for each minor found in the rate object `rateobj`

Examples

```
library(LTASR)
library(dplyr)

#Import example person file
person <- person_example %>%
  mutate(dob = as.Date(dob, format='%m/%d/%Y'),
         pybegin = as.Date(pybegin, format='%m/%d/%Y'),
         dlo = as.Date(dlo, format='%m/%d/%Y'))

#Import default rate object
rateobj <- us_119ucod_19602020

#Stratify person table
```

```
py_table <- get_table(person, rateobj)

#Calculate SMRs for all minors
smr_minor(py_table, rateobj)
```

`us_119ucod_19602020` *119 UCOD U.S. Death Rate, 1960-2020*

Description

A list containing referent underlying cause of death (UCOD) rate information for the US population from 1960-2020 for the 119 minor/outcome LTAS groupings

Usage

```
us_119ucod_19602020
```

Format

A list with 4 elements:

residual the minor/outcome number to which unknown/uncategorized outcomes will be mapped to

MinorDesc a data.frame containing descriptions for each minor and major grouping

mapping a tibble detailing which minor number each icd-code and revision combination will be mapped to

rates the population referent rate for each minor for each gender/race/calendar period/age strata ...

Source

Available upon request from nioshltas@cdc.gov

Index