

Package: SequentialDesign (via r-universe)

October 11, 2024

Title Observational Database Study Planning using Exact Sequential Analysis for Poisson and Binomial Data

Version 1.0

Author Judith Maro [aut, cre], Laura Hou [aut]

Maintainer Judith Maro <judy_maro@harvardpilgrim.org>

Description Functions to be used in conjunction with the 'Sequential' package that allows for planning of observational database studies that will be analyzed with exact sequential analysis. This package supports Poisson- and binomial-based data. The primary function, seq_wrapper(...), accepts parameters for simulation of a simple exposure pattern and for the 'Sequential' package setup and analysis functions. The exposure matrix is used to simulate the true and false positive and negative populations (Green (1983) <doi:10.1093/oxfordjournals.aje.a113521>, Brenner (1993) <doi:10.1093/oxfordjournals.aje.a116805>). Functions are then run from the 'Sequential' package on these populations, which allows for the exploration of outcome misclassification in data.

Depends R (>= 3.3.1)

Imports Sequential, stats

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2017-11-16 09:52:56 UTC

Contents

create.exposure	2
---------------------------	---

fun.decum	2
fun.exposure	3
initialize.data	3
SCRI.seq	5
seq_wrapper	5
sim.exposure	8

Index	9
--------------	----------

create.exposure	<i>Create Exposure Data Matrix</i>
-----------------	------------------------------------

Description

This function create.exposure is a sub-function used in conjunction with the initialize.data function and creates an exposure matrix. The columns represent strata for the observational data and the rows represent new exposures in unit time. It take cumulative data and segregates it by time period. Do not run create.exposure as a stand-alone function.

Usage

```
create.exposure(params)
```

Arguments

params	This is a set of parameters from the initialize.data function that allows for simulation of a sequence of sequential exposures in unit time.
--------	--

Examples

```
paramtest <- initialize.data(seed=8768, N=1, t0=0, tf=2, NStrata=2,
strataRatio=c(0.2, 0.3, 0.3, 0.2), EventRate=c(0.4, 0.5), sensitivity=0.9, PPVest=0.9, RR=3.0,
MatchRatio=1, maxSampleSize=200, maxTest=1, totalAlpha=0.05, minEvents=5, AlphaSpendType="Wald",
AlphaParameter=0.5, address=getwd(), rate=20, offset=30)
create.exposure(paramtest)
```

fun.decum	<i>Decumulate Values in a Matrix</i>
-----------	--------------------------------------

Description

The function takes a matrix of values that have accumulated over rows and returns a matrix of the incremental increase between each row. Do not run fun.decum as a stand-alone function

Usage

```
fun.decum(matrix)
```

Arguments

`matrix` This is a matrix of values to be decumulated where the cumulation occurs over rows within the same column.

Examples

```
testarray <- array(NA, dim=c(5,2,2))
testarray[,,1] <- cbind(c(1:5), c(1:5)*2)
testarray[,,2] <- cbind(c(1:5)*1.5, c(1:5)*3)
fun.decum(testarray)
```

`fun.exposure` *Define Exposure Accumulation Function*

Description

This function creates a dataset simulating the accumulation of individuals exposed to treatment under the self-controlled risk interval design.

Usage

```
fun.exposure(rate, offset, t)
```

Arguments

`rate` Rate of accumulation.
`offset` Initial exposed population.
`t` Time at which individuals are exposed.

Examples

```
fun.exposure(rate=100, offset=20, t=20)
```

`initialize.data` *Create Simulated Sequential Data Parameter Data Frame*

Description

The function creates a data frame with all the needed parameters for simulation and initializes the simulation problem. Do not run `initialize.data` as a stand-alone function.

Usage

```
initialize.data(seed, N, t0, tf, NStrata, strataRatio, EventRate, sensitivity,
  PPVest, RR, MatchRatio, maxSampleSize, maxTest, totalAlpha, minEvents,
  AlphaSpendType, AlphaParameter, address, rate, offset)
```

Arguments

seed	Seed used for randomization.
N	Number of simulations to be created. Because adverse event assignment is stochastic, this number is usually at least 10,000.
t0	Initial time point, a number in units of either days, weeks, months, or years. It is important to be consistent.
tf	Final time point, a number in units of either days, weeks, months, or years.
NStrata	Number of strata in the observational study design, where a "stratum" can be defined by age categories, sex, and any other defining characteristics. Event rate of the adverse event of interest is also segregated by strata and database population size is also segregated by strata. For example, a single strata might 0-17 year old females.
strataRatio	Ratio of individuals within a single strata for exposed and unexposed individuals. The number of elements in this list should be 2*NStrata.
EventRate	Rate of event accrual given in events /person-time where the time constant is the same constant being used throughout the study. Additionally, the number of elements in the EventRate matrix should be equal to the NStrata.
sensitivity	True sensitivity of the outcome of interest. $\text{sensitivity} = (\text{true positive case}) / (\text{true positive case} + \text{false negative case})$.
PPVest	True positive predictive value of outcome in the unexposed group. $\text{PPV} = (\text{true positive case}) / (\text{true positive case} + \text{false positive case})$.
RR	Intended relative risk to detect (and therefore to simulate) in the dataset.
MatchRatio	Single numeric value. In a self-controlled risk interval design, it is the ratio of the length of the control window to the length of the risk window.
maxSampleSize	Maximum number of events before sequential analysis is ended or the upper limit on sample size expressed in terms of total number of events. This is the same variable as N from R Sequential.
maxTest	Number of tests to perform on simulation data.
totalAlpha	Total amount of alpha available to spend.
minEvents	Minimum number of events needed before the null hypothesis can be rejected. Represented as M in R Sequential.
AlphaSpendType	Method of alpha expenditure. Available values are "Wald" or "power-type". This is the same as AlphaSpending R Sequential.
AlphaParameter	Rho parameter for power-type alpha spending function. This is the same as rho in R Sequential.
address	File directory where data for sequential analysis is stored for future tests.
rate	Rate of exposure/cohort accrual.
offset	Offset for exposure/cohort accrual.

Examples

```
initialize.data(seed=8768, N=1, t0=0, tf=2, NStrata=2, strataRatio=c(0.2, 0.3, 0.3, 0.2),
EventRate=c(0.4, 0.5), sensitivity=0.9, PPVest=0.9, RR=3.0, MatchRatio=1, maxSampleSize=200,
maxTest=1, totalAlpha=0.05, minEvents=5, AlphaSpendType="Wald", AlphaParameter=0.5, address=getwd(),
rate=20, offset=30)
```

SCRI.seq	<i>Perform sequential analysis on true and misclassified binomial data</i>
----------	--

Description

This function performs a prespecified number of binomial sequential analyses on real and misclassified binomial data that is designed to simulate a self-controlled risk interval design. Do not run SCRI.seq as a stand-alone function.

Usage

```
SCRI.seq(data, params)
```

Arguments

data	Output from sim.exposure that contains real and observed data.
params	Output from initialize.data function.

Examples

```
#paramtest <- initialize.data(seed=8768, N=1, t0=0, tf=2, NStrata=2,
#strataRatio=c(0.2, 0.3, 0.3, 0.2), EventRate=c(0.4, 0.5), sensitivity=0.9, PPVest=0.9, RR=3.0,
#MatchRatio=1, maxSampleSize=200, maxTest=1, totalAlpha=0.05, minEvents=5, AlphaSpendType="Wald",
#AlphaParameter=0.5, address=getwd(), rate=20, offset=30)
#exposed1 <- create.exposure(paramtest)
#exposed2 <- sim.exposure(exposed1, paramtest)
#SCRI.seq(exposed2, paramtest)
```

seq_wrapper	<i>Execute Simulated Exact Sequential Analysis in Multi-Site Observational Database Studies</i>
-------------	---

Description

SequentialDesign is designed for planning observational database studies that use exact sequential analysis. It is designed to be used in conjunction with the R package Sequential. This package is appropriate to use when one is performing a multi-site observational database study (i.e., an epidemiologic study) and planning to use sequential statistical analysis. This package supports two types of observational study designs:

- a self-controlled risk interval design which creates binomial data, and
- a current v. historical design which creates Poisson data.

The goal of this package is to allow the investigator to plan for the optimal study.

Usage

```
seq_wrapper(seed, N, t0, tf, NStrata, strataRatio, EventRate, sensitivity,
  PPVest, RR, MatchRatio, maxSampleSize, maxTest, totalAlpha, minEvents,
  AlphaSpendType, AlphaParameter, rate, offset, address, ...)
```

Arguments

seed	Seed used for randomization
N	Number of simulations to be created. Because adverse event assignment is stochastic, this number is usually at least 10,000.
t0	Initial time point, a number in units of either days, weeks, months, or years. It is important to be consistent.
tf	Final time point, a number in units of either days, weeks, months, or years.
NStrata	Number of strata in the observational study design, where a "stratum" can be defined by age categories, sex, and any other defining characteristics. Event rate of the adverse event of interest is also segregated by strata and database population size is also segregated by strata. For example, a single strata might 0-17 year old females.
strataRatio	Ratio of individuals within a single strata for exposed and unexposed individuals. The number of elements in this list should be 2*NStrata.
EventRate	Rate of event accrual given in events /person-time where the time constant is the same constant being used throughout the study. Additionally, the number of elements in the EventRate matrix should be equal to the NStrata.
sensitivity	True sensitivity of the outcome of interest. $\text{sensitivity} = (\text{true positive case}) / (\text{true positive case} + \text{false negative case})$.
PPVest	True positive predictive value of outcome in the unexposed group. $\text{PPV} = (\text{true positive case}) / (\text{true positive case} + \text{false positive case})$.
RR	Intended relative risk to detect (and therefore to simulate) in the dataset.
MatchRatio	Single numeric value. In a self-controlled risk interval design, it is the ratio of the length of the control window to the length of the risk window.
maxSampleSize	Maximum number of events before sequential analysis is ended or the upper limit on sample size expressed in terms of total number of events. This is the same variable as N from R Sequential.
maxTest	Number of tests to perform on simulation data.
totalAlpha	Total amount of alpha available to spend.
minEvents	Minimum number of events needed before the null hypothesis can be rejected. Represented as M in R Sequential.
AlphaSpendType	Method of alpha spenditure. Available values are "Wald" or "power-type". This is the same as AlphaSpending R Sequential.
AlphaParameter	Rho parameter for power-type alpha spending function. This is the same as rho in R Sequential.
rate	Rate of exposure/cohort accrual.
offset	Offset for exposure/cohort accrual.

address Output folder where Sequential TXT files are to be stored. These should be preserved between runs, as detailed within the Sequential package.

... additional arguments to be passed to or from methods.

Details

The simulation has the following steps:

1. Sample Size Calculations for the study using the R Sequential package
2. Given these sample size calculations and an exposure uptake function, calculate new exposure accrual in calendar time for the exposures of interest.
3. Given the simulated exposure information, generate adverse events of interest according to a pre-specified effect size.
4. Perform sequential analysis on these simulated data.
5. Generate calendar time descriptive statistics with respect to stopping points.

These steps will be discussed in more detail.

First, the investigator should work with the R package Sequential in order to calculate design parameters for their study. These are the statistical parameters that govern stopping points in statistical analysis. The relevant ones required for this analysis are: `maxSampleSize`, `totalAlpha`, `minEvents`, `AlphaSpendType`, `AlphaParameter`. If binomial data is being used for sequential analysis of a self-controlled risk interval design, then `MatchRatio` is also needed.

Second, this function will generate incident exposure to a simulated study population based on the parameters of an exposure accrual function.

Third, with incremental exposure accrual information, new adverse events will be assigned based on user-specified characteristics. This function also allows for outcome misclassification so true positive adverse events, false positive adverse events, and false negative adverse events are all simulated. Fourth, Sequential analysis is implemented on these simulated data using function in R Sequential. Fifth, the investigator is able to generate descriptive statistics in calendar time to enable the investigator to plan for their analysis.

Simulating sequential analysis in observational data requires many parameter inputs about

- the parameters that control the epidemiologic study design,
- the parameters that describe the characteristics of the databases, and
- the parameters of the simulation.

In addition to the parameter inputs, there are many sub-functions that are needed to perform different steps in the simulation. These sub-functions are not intended to be run as stand-alone functions but rather always in the sequence specified in this function.

Examples

```
#paramtest <- initialize.data(seed=8768, N=1, t0=0, tf=2, NStrata=2,
#strataRatio=c(0.2, 0.3, 0.3, 0.2), EventRate=c(0.4, 0.5), sensitivity=0.9, PPVest=0.9, RR=3.0,
#MatchRatio=1, maxSampleSize=200, maxTest=1, totalAlpha=0.05, minEvents=5, AlphaSpendType="Wald",
#AlphaParameter=0.5, address=getwd(), rate=20, offset=30)
#exposed1 <- create.exposure(paramtest)
#exposed2 <- sim.exposure(exposed1, paramtest)
#SCRI.seq(exposed2, paramtest)
```

`sim.exposure`*Create Simulated Exposure Matrix for Real and Observed Data*

Description

This function creates an exposure matrix with real and observed data after taking into account true positive, false negative, and false positive rates. The columns represent strata for the observational data #' and the rows represent new events in unit time. Do not run `sim.exposure` as a stand-alone function.

Usage

```
sim.exposure(exposed.matrix, params)
```

Arguments

`exposed.matrix` Output exposure matrix from `create.exposure` function.

`params` Output from `initialize.data` function.

Examples

```
paramtest <- initialize.data(seed=8768, N=1, t0=0, tf=2, NStrata=2,
strataRatio=c(0.2, 0.3, 0.3, 0.2), EventRate=c(0.4, 0.5), sensitivity=0.9, PPVest=0.9, RR=3.0,
MatchRatio=1, maxSampleSize=200, maxTest=1, totalAlpha=0.05, minEvents=5, AlphaSpendType="Wald",
AlphaParameter=0.5, address=getwd(), rate=20, offset=30)
exposed1 <- create.exposure(paramtest)
sim.exposure(exposed1, paramtest)
```


Index

- * **exposure**
 - fun.exposure, 3
- * **sequential**
 - create.exposure, 2
 - fun.decum, 2
 - fun.exposure, 3
 - initialize.data, 3
 - SCRI.seq, 5
 - seq_wrapper, 5
 - sim.exposure, 8
- * **simulation**
 - fun.exposure, 3

create.exposure, 2

fun.decum, 2

fun.exposure, 3

initialize.data, 3

SCRI.seq, 5

seq_wrapper, 5

sim.exposure, 8