Data Simulation for Log-rank Test Power Calculation

rong.lu@utsw.edu

October 07, 2019

Description

The function **LRPowerSimulation()** in R package **LogrankPower** is used to generate 1 simulation dataset for power calculation of the Log-rank test, with user specified total sample size, effect size, sample size ratio between comparison groups, the death rate in the reference group, and the distribution of follow-up time (negative binomial).

Usage

LRPowerSimulation(total.sample.size, effect.size=0.6, group.sample.size.ratio=1, reference.group.incidence=0.5, time.distribution.para=c(100, 0.8))

Arguments

- **total.sample.size** Total sample size in this simulation dataset.

- **effect.size** A scalar parameter (between 0 and 1). effect.size=0.2 (small effect size); effect.size=0.5 (median effect size); effect.size=0.8 (large effect size). effect.size=0.6 by default.

- **group.sample.size.ratio** Ratio between samples of contrast group and the reference or control group. group.sample.size.ratio=1 by default.

- **reference.group.incidence** The incidence of death or event of interest in the reference patient group. reference.group.incidence=0.5 by default.

- **time.distribution.para** A numeric vector of length 2. This vector is used to specify the negative binomial distribution that will be used to simulate follow-up time. The 1st element in this vector is the target for number of successful trials, or dispersion parameter (the shape parameter of the gamma mixing distribution). Must be strictly positive, need not be integer. Then 2nd parameter is the probability of success in each trial. time.distribution.para=c(100, 0.8) by default.

Simulation Steps:

- determine sample size of the reference patient group (group 1) by calculating:

      round(total.sample.size**/**(group.sample.size.ratio**+1**))

- determine sample size of the contrast group (group 2) by calculating:

      total.sample.size **-** reference.group.n

- generate follow-up time for both the referencee and the contrast group using negative bionomial distribution with user specified parameters:

```
                nb.size <- time.distribution.para[1]
                nb.prob <- time.distribution.para[2]
                reference.group.time <- rnbinom(reference.group.n, nb.size, nb.prob)
```

- generate vital status indictor for the reference group by sampling from the bionomial distribution with the number of trials = 1 and the probability of success/death = reference.group.incidence:

```
                reference.group.event.indicator <- rbinom(reference.group.n,
                                                1,reference.group.incidence)
```

- generate vital status indictor for the contrast group by sampling from the bionomial distribution with the number of trials = 1 and the probability of success (death) = reference.group.incidence*(1-effect.size):

```
                contrast.group.event.indicator <- rbinom(reference.group.n,
                                                1,reference.group.incidence*(1-effect.size))
```

Complete Code:

```
LRPowerSimulation <- function(total.sample.size, effect.size=0.6,
                              group.sample.size.ratio=1,
                              reference.group.incidence=0.5,
                              time.distribution.para=c(100, 0.8)){

        reference.group.n <- round(total.sample.size/(group.sample.size.ratio+1))
        contrast.group.n <- total.sample.size - reference.group.n

        nb.size <- time.distribution.para[1]
        nb.prob <- time.distribution.para[2]
        reference.group.time <- rnbinom(reference.group.n, nb.size, nb.prob)
        reference.group.event.indicator <- rbinom(reference.group.n,
                                                1,reference.group.incidence)
        reference.data <- data.frame(time=reference.group.time,
                                event=reference.group.event.indicator, group="group 1")

        contrast.group.time <- rnbinom(reference.group.n, nb.size, nb.prob)
        contrast.group.event.indicator <- rbinom(reference.group.n,
                                                1,reference.group.incidence*(1-effect.size))
        contrast.data <- data.frame(time=contrast.group.time,
                                event=contrast.group.event.indicator, group="group 2")

        surv.data <- rbind(reference.data, contrast.data)
        surv.data$SurvObj <- with(surv.data, Surv(time, event))

   return(surv.data) ## return 1 simulation dataset based on the assume parameters
}
```

Example:

```
library(survival)
library(survminer)
```

```
## Loading required package: ggplot2

## Loading required package: ggpubr

## Loading required package: magrittr
```

```r
set.seed(1234)
surv.data <- LRPowerSimulation(total.sample.size=160, effect.size=0.5,
group.sample.size.ratio=1, reference.group.incidence=0.5)
ggsurvplot(survfit(SurvObj ~ group, surv.data), surv.median.line = "hv",
           conf.int = TRUE, pval=TRUE)
```