Package ‘wcox’

May 16, 2023

Title Weights to Correct for Outcome Dependent Sampling in Time to Event Data

Version 1.0

Description A new inverse probability of selection weighted Cox model to deal with outcome-dependent sampling in survival analysis.

License GPL-3

Encoding UTF-8

RoxygenNote 7.2.3

LazyData true

Imports dplyr, tidyr, survival, stats

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

NeedsCompilation no

Author Vera Arntzen [aut, cre] (<https://orcid.org/0000-0002-2642-9898>),
Mar Rodriguez-Girondo [aut] (<https://orcid.org/0000-0003-3126-1003>)

Maintainer Vera Arntzen <v.h.arntzen@math.leidenuniv.nl>

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2023-05-16 08:30:02 UTC

R topics documented:

Calculate_weights ................................................. 2
fam_dat .......................................................... 2
Prepare_data ....................................................... 3

Index 4
Calculate_weights  Calculate inverse probability of selection weights.

Description

This function calculates weights to correct for ascertainment bias in time-to-event data where clusters are outcome-dependently sampled, for example high-risk families in genetic epidemiological studies in cancer research.

Usage

Calculate_weights(dat)

Arguments

dat  Data.frame with one row per individual with columns d non-censoring indicator; k interval of (age) group; S_k population interval-based proportion of individuals experiencing the event in intervals later than k; S_k. sample proportion of individuals experiencing the event in intervals later than k.

Details

Weights are based on a comparison between the survival between sample and population. Therefore, besides the sample data, the population incidence rate (per 100 000) is needed as input, as well as the cut-offs of the (age/time-to-event) groups for which this is available. The function provides two options for the latter: cut-offs can be provided manually or using the standard 5- or 10-years (age) categories (0-4, 5-9, ... or 0-9, 10-14, ...). Note that resulting intervals are of the form [xx, xx).

Value

Vector with weights.

fam_dat  Simulated time-to-event data.

Description

This toy data set is simulated for educational purposes explaining the package. It concerns of families in which at least two individuals experienced the event of interest during following up. The covariate of interest is risk modifier ‘x’. This data set is inspired by data that is often seen in genetic epidemiological studies in cancer research.

Usage

data(fam_dat)
Prepare_data

Format
Data.frame.

References

Prepare_data
Prepare data for weight calculation using the population incidence.

Description
This function prepares the sample data for weight calculation using external information, i.e. the incidence in the population.

Usage
Prepare_data(dat, population_incidence, breaks)

Arguments
- dat: Data.frame with one row per individual which at least includes a column d with event indicator (1 for event, 0 for censored), a column y with event/censoring time.
- population_incidence: A vector (in combination with breaks) or a data.frame (columns 1) ‘start age group’, 2) ‘end age group’, 3) ‘S_pop’) with population incidence per 100,000 per interval k.
- breaks: Cut-points for the (age/time) groups. Only needed when population_incidence is a vector.

Details
Weights are based on a comparison between the survival between sample and population. Therefore, besides the sample data, the population incidence rate (per 100 000) is needed as input, as well as the cut-offs of the (age/time-to-event) groups for which this is available, unless this is provided in a data.frame.

Value
Data.frame ready for weight calculation using function ‘Calculate_weights()’: one row per individual and a.o. columns id unique ID; d non-censoring indicator; k interval of (age) group; S_k population interval-based proportion of individuals experiencing the event in intervals later than k; S_k. sample proportion of individuals experiencing the event in intervals later than k.
Index

* datasets
  fam_dat, 2

Calculate_weights, 2
fam_dat, 2
Prepare_data, 3