

Package ‘clustcurv’

March 25, 2019

Type Package

Title Determining Groups in Multiples Curves

URL <https://github.com/noramvillanueva/clustcurv>

BugReports <http://github.com/noramvillanueva/clustcurv/issues>

Version 1.0.0

Date 2019-03-15

Maintainer Nora M. Villanueva <nmvillanueva@uvigo.es>

Description A method for determining groups in multiple survival curves with an automatic selection of their number based on k-means or k-medians algorithms. The selection of the optimal number is provided by bootstrap methods.
Implemented methods are:
Grouping multiple survival curves described by Villanueva et al. (2018) <doi:10.1002/sim.8016>.

Depends R (>= 3.5.0)

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Imports doParallel, foreach, ggplot2, ggfortify, doRNG, Gmedian, survival, wesanderson

Suggests testthat, usethis, condSURV

RoxygenNote 6.1.1

NeedsCompilation no

Author Nora M. Villanueva [aut, cre] (<<https://orcid.org/0000-0001-8085-2745>>),
Marta Sestelo [aut]

Repository CRAN

Date/Publication 2019-03-25 09:23:23 UTC

R topics documented:

autoplot.clustcurv_surv	2
clustcurv	3
clustcurv_surv	4
kgroups_surv	6

Index	8
--------------	----------

autoplot.clustcurv_surv

Visualization of clustcurv_surv objects with ggplot2 graphics

Description

Useful for drawing the estimated survival functions grouped by color and the centroids (mean curve of the curves pertaining to the same group).

Usage

```
## S3 method for class 'clustcurv_surv'
autoplot(object = object,
  groups_by_colour = TRUE, centers = FALSE, conf.int = FALSE,
  censor = FALSE, xlab = "Time", ylab = "Survival", ...)
```

Arguments

object	Object of clustcurv_surv class.
groups_by_colour	A specification for the plotting groups by color.
centers	Draw the centroids (mean of the curves pertaining to the same group) into the plot. By default it is FALSE.
conf.int	Logical flag indicating whether to plot confidence intervals.
censor	Logical flag indicating whether to plot censors.
xlab	A title for the x axis.
ylab	A title for the y axis.
...	Other options.

Details

See help page of the function [autoplot.survfit](#).

Value

A ggplot object, so you can use common features from ggplot2 package to manipulate the plot.

Author(s)

Nora M. Villanueva and Marta Sestelo.

Examples

```
library(survival)
library(clustcurv)
library(condSURV)
library(ggplot2)
library(ggfortify)
data(veteran)
data(colonCS)

c12 <- kgroups_surv(time = veteran$time, status = veteran$status,
  fac = veteran$celltype, k = 2, algorithm = "kmeans")

autoplot(c12)
autoplot(c12, groups_by_colour = FALSE)
autoplot(c12, centers = TRUE)

colonCSm <- data.frame(time = colonCS$time, status = colonCS$event,
  nodes = colonCS$nodes)

table(colonCSm$nodes)
colonCSm$nodes[colonCSm$nodes == 0] <- NA
colonCSm <- na.omit(colonCSm)
colonCSm$nodes[colonCSm$nodes >= 10] <- 10
table(colonCSm$nodes) # ten levels

res <- clustcurv_surv(colonCSm$time, status = colonCSm$status,
  fac = colonCSm$nodes, algorithm = "kmeans", nboot = 20)

autoplot(res)
autoplot(res, groups_by_colour = FALSE)
autoplot(res, centers = TRUE)
```

clustcurv

clustcurv: *Determining Groups in Multiple Curves.*

Description

This package provides a method for determining groups in multiple survival curves with an automatic selection of their number based on k-means or k-medians algorithms. The selection of the optimal number is provided by bootstrap methods.

Details

Package: clustcurv
Type: Package
License: MIT + file LICENSE

clustcurv is designed along lines similar to those of other R packages. This software helps the user determine groups in multiple curves (survival curves in the current version). In addition, it enables both numerical and graphical outputs to be displayed (by means of ggplot2). The package provides the kgroups_surv() function that groups the curves given a number k and the clustcurv_surv() function that selects the optimal number of groups automatically through a bootstrap-based test. The autoplot() function let the user draw the resulted estimated curves coloured by groups.

For a listing of all routines in the clustcurv package type: library(help="clustcurv").

Author(s)

Nora M. Villanueva and Marta Sestelo

References

Villanueva, N. M., Sestelo, M., and Meira-Machado, J. (2019). A method for determining groups in multiple survival curves. *Statistics in Medicine*, 8(5):866-877

See Also

Useful links:

- <https://github.com/noramvillanueva/clustcurv>
- Report bugs at <http://github.com/noramvillanueva/clustcurv/issues>

clustcurv_surv

Clustering survival curves

Description

Function for grouping survival curves based on the k-means or k-medians algorithm. It returns the number of groups and the assignment.

Usage

```
clustcurv_surv(time, status, fac, kvector = NULL, kbin = 50,  
nboot = 100, algorithm = "kmeans", alpha = 0.05, cluster = FALSE,  
ncores = NULL, seed = NULL, multiple = FALSE,  
multiple.method = "holm")
```

Arguments

time	Survival time.
status	Censoring indicator of the survival time of the process; 0 if the total time is censored and 1 otherwise.
fac	Categorical variable indicating the population to which the subject belongs
kvector	A vector specifying the number of groups of curves to be checking.
kbin	Size of the grid over which the survival functions are to be estimated.
nboot	Number of bootstrap repeats.
algorithm	A character string specifying which clustering algorithm is used, i.e., k-means("kmeans") or k-medians ("kmedians").
alpha	Seed to be used in the procedure.
cluster	A logical value. If TRUE (default), the testing procedure is parallelized. Note that there are cases (e.g., a low number of bootstrap repetitions) that R will gain in performance through serial computation. R takes time to distribute tasks across the processors also it will need time for binding them all together later on. Therefore, if the time for distributing and gathering pieces together is greater than the time need for single-thread computing, it does not worth parallelize.
ncores	An integer value specifying the number of cores to be used in the parallelized procedure. If NULL (default), the number of cores to be used is equal to the number of cores of the machine - 1.
seed	Seed to be used in the procedure.
multiple	A logical value. If TRUE (not default), the resulted pvalues are adjunted by using one of several methods for multiple comparisons.
multiple.method	Correction method. See Details.

Details

The adjustment methods include the Bonferroni correction ("bonferroni") in which the p-values are multiplied by the number of comparisons. Less conservative corrections are also included by Holm (1979) ("holm"), Hochberg (1988) ("hochberg"), Hommel (1988) ("hommel"), Benjamini & Hochberg (1995) ("BH" or its alias "fdr"), and Benjamini & Yekutieli (2001) ("BY"), respectively. A pass-through option ("none") is also included.

Value

A list containing the following items:

table	A data frame containing the null hypothesis tested, the values of the test statistics and the obtained pvalues.
levels	Original levels of the variable fac.
cluster	A vector of integers (from 1:k) indicating the cluster to which each curve is allocated.
centers	An object of class <code>survfit</code> containing the centroids (mean of the curves pertaining to the same group).
curves	An object of class <code>survfit</code> containing the survival curves for each population.

Author(s)

Marta Sestelo, Nora M. Villanueva.

Examples

```
library(clustcurv)
library(survival)
library(condSURV)
data(veteran)
data(colonCS)

res <- clustcurv_surv(time = veteran$time, status = veteran$status,
  fac = veteran$celltype, algorithm = "kmeans")

#res <- clustcurv_surv(colonCS$time, status = colonCS$status, fac = colonCS$nodes, nboot = 20)
```

kgroups_surv

k-groups of survival curves

Description

Function for grouping survival curves, given a number k , based on the k -means or k -medians algorithm.

Usage

```
kgroups_surv(time, status, fac, k, kbin = 50, algorithm = "kmeans",
  seed = NULL)
```

Arguments

time	Survival time.
status	Censoring indicator of the survival time of the process; 0 if the total time is censored and 1 otherwise.
fac	Categorical variable indicating the population to which the subject belongs
k	An integer specifying the number of groups of curves to be performed.
kbin	Size of the grid over which the survival functions are to be estimated.
algorithm	A character string specifying which clustering algorithm is used, i.e., k -means("kmeans") or k -medians("kmedians").
seed	Seed to be used in the procedure.

Value

A list containing the following items:

measure	A measure of...
levels	Original levels of the variable fac.
cluster	A vector of integers (from 1:k) indicating the cluster to which each curve is allocated.
centers	An object of class survfit containing the centroids (mean of the curves pertaining to the same group).
curves	An object of class survfit containing the survival curves for each population.

Author(s)

Marta Sestelo, Nora M. Villanueva.

Examples

```
library(clustcurv)
library(survival)
data(veteran)

# 2 groups k-means
cl2 <- kgroups_surv(time = veteran$time, status = veteran$status,
  fac = veteran$celltype, k = 2, algorithm = "kmeans")

data.frame(level = cl2$level, cluster = cl2$cluster)

# 2 groups k-medians
cl2 <- kgroups_surv(time = veteran$time, status = veteran$status,
  fac = veteran$celltype, k = 2, algorithm = "kmedians")

data.frame(level = cl2$level, cluster = cl2$cluster)

# 3 groups
cl3 <- kgroups_surv(time = veteran$time, status = veteran$status,
  fac = veteran$celltype, k = 3, algorithm = "kmeans")

data.frame(level = cl3$level, cluster = cl3$cluster)
```

Index

[autoplot.clustcurv_surv](#), 2
[autoplot.survfit](#), 2

[clustcurv](#), 3
[clustcurv-package \(clustcurv\)](#), 3
[clustcurv_surv](#), 4

[kgroups_surv](#), 6