Package ‘parmsurvfit’

October 14, 2022

**Title**  Parametric Models for Survival Data

**Version**  0.1.0

**Description**  Executes simple parametric models for right-censored survival data. Functionality emulates capabilities in 'Minitab', including fitting right-censored data, assessing fit, plotting survival functions, and summary statistics and probabilities.

**Depends**  R (>= 3.5.0)

**License**  GPL-2

**URL**  https://github.com/apjacobson/parmsurvfit

**BugReports**  https://github.com/apjacobson/parmsurvfit/issues

**Encoding**  UTF-8

**LazyData**  true

**RoxygenNote**  6.1.1

**Imports**  fitdistrplus, flexsurv, ggplot2, stats, graphics

**Suggests**  survival, utils, knitr, rmarkdown

**VignetteBuilder**  utils, knitr

**NeedsCompilation**  no

**Author**  Ashley Jacobson [aut, cre],
            Victor Wilson [aut],
            Shannon Pileggi [aut]

**Maintainer**  Ashley Jacobson <ashleypjacobson@gmail.com>

**Repository**  CRAN

**Date/Publication**  2018-12-07 05:30:03 UTC

**R topics documented:**

aggressive ................................................................. 2
compute_AD .............................................................. 3
firstdrink ............................................................... 3
Description

Diekmann et al. (1996) investigated the association between driver characteristics and social status of cars to aggressive driver responses by measuring the time that elapsed between the being blocked and honking the horn. Researchers intentionally blocked 57 motorists at a green light by a Volkswagen Jetta, and recorded the time it took for motorists to show signs of aggression. Signs of aggression included honking their horn or beaming the headlights at the Jetta.

Usage

aggressive

Format

A data frame with 57 rows and 2 variables:

- **seconds** Number of seconds until showing signs of aggression
- **censor** censoring status indicator variable (0 = censored event time, 1 = complete event time)

Source

https://stats.idre.ucla.edu/other/examples/alda/
compute_AD

Anderson-Darling goodness of fit test statistic

Description
Computes Anderson-Darling goodness of fit test statistic given that the data follows a specified parametric distribution.

Usage
compute_AD(data, dist, time = "time", censor = "censor")

Arguments
data
A dataframe containing a time column and a censor column.
dist
A string name for a distribution that has a corresponding density function and a distribution function. Examples include "norm", "lnorm", "exp", "weibull", "logis", "ilogis", "gompertz", etc.
time
The string name of the time column of the dataframe. Defaults to "time".
censor
The string name of the censor column of the dataframe. Defaults to "censor". The censor column must be a numeric indicator variable where complete times correspond to a value of 1 and incomplete times correspond to 0.

Examples
data("rearrest")
compute_AD(rearrest, "lnorm", time = "months")
compute_AD(rearrest, "weibull", time = "months")

firstdrink
Data on age at first drink of alcohol.

Description
Data on age at first drink of alcohol.

Usage
firstdrink

Format
A data frame with 1000 rows and 3 variables:

- **age** the age at which the survey respondent had their first drink of alcohol
- **censor** censoring status indicator variable (0 = censored event time, 1 = complete event time)
- **gender** a dichotomous variable identifying gender (1 = male, 2 = female)
fit_data  
Fitting right censored survival data to distribution

Description

Fits right censored data to a distribution using maximum likelihood estimates.

Usage

fit_data(data, dist, time = "time", censor = "censor", by = "")

Arguments

data  A dataframe containing a time column and a censor column.
dist  A string name for a distribution that has a corresponding density function and a distribution function. Examples include "norm", "lnorm", "exp", "weibull", "logis", "llogis", "gompertz", etc.
time  The string name of the time column of the dataframe. Defaults to "time".
censor  The string name of the censor column of the dataframe. Defaults to "censor". The censor column must be a numeric indicator variable where complete times correspond to a value of 1 and incomplete times correspond to 0.
by  The string name of a grouping variable. If specified, the function returns a list. The list will be in alphabetical order of the values in the by column. Variable can contain logical, string, character, or numeric data.

See Also

fitdistcens

Examples

data("rearrest")
fit_data(rearrest, "lnorm", time = "months")
fit_data(rearrest, "weibull", time = "months", by = "personal")

Source

"National Comorbidity Survey (1990-1992)"
**graduate**

*Data on time until graduation for 1000 college students.*

**Description**
A dataset that contains the time (in years) that 1000 students (472 males and 528 females) took to graduate (obtain a bachelor’s degree) from college (measured from the time they entered a post-secondary institution, i.e. either a junior college or four year degree granting institution). The Gender column contains the gender of each student (1 = male, 2 = female), and Censor contains the values of the censoring status variable.

**Usage**

graduate

**Format**
A data frame with 1000 rows and 3 variables:

- **years**: years until graduation
- **censor**: censoring status indicator variable (0 = censored event time, 1 = complete event time)
- **gender**: a dichotomous variable identifying gender (1 = male, 2 = female)

**Source**
National Educational Longitudinal Survey (NELS) from 1988-2002

---

**oscars**

*Data on time until actors receive their first Academy Award nomination*

**Description**
The dataset contains data for the top 128 grossing actors up to 2017 as listed on Box Office Mojo. The data for the first film appearance and for the first Oscar nomination was taken from IMDb. It should be noted that of the 128 observations in the data set, 48 were right-censored. Right-censored observations represent actors who have not received an Oscar nomination by the year 2017 or actors that died before 2017 without ever receiving an Oscar nomination. For the censor variable "1" represents complete observations, actors who received an Oscar nomination by the year 2017, and "0" represents right-censored observations.

**Usage**
oscars
Format

A data frame with 128 rows and 12 variables:

- **obs**  observation number
- **name**  name of actor
- **adj_gross**  actor’s total adjusted gross earnings (in millions)
- **num_movies**  number of movies actor received credit for
- **avg_gross**  actor’s average gross earnings per movie
- **top_movie**  title of actor’s movie with the top gross earnings
- **top_gross**  actor’s top gross earnings from a single movie
- **gender**  actor’s gender
- **years_until_nom**  number of years between actor’s first full film appearance and first Oscar nomination
- **censor**  censoring status indicator variable (0 = censored event time, 1 = complete event time)
- **first_film_appearance**  year of actor’s first full film appearance
- **first_oscar_nom**  year of actor’s first Oscar nomination

Source

https://github.com/shannonpileggi/SP--Pablo--RProgramming

---

parmsurvfit  

*parmsurvfit: Fitting right censored data to parametric distributions.*

Description

Executes parametric survival analysis techniques similar to those in 'Minitab'. Fits right censored data to a given parametric distribution, produces summary statistics of the fitted distribution, and plots parametric survival, hazard, and cumulative hazard plots. Produces Anderson-Darling test statistic and probability plots to assess goodness of fit of right censored data to a distribution.

Details

Functions

- **fit_data**
- **surv_summary**
- **surv_prob**
- **plot_surv**
- **plot_haz**
- **plot_cumhaz**
- **plot_density**
plot_cumhaz

- plot_ppsurv
- compute_AD

Datasets
- aggressive
- firstdrink
- graduate
- oscars
- rearrest

---

### plot_cumhaz

**Plotting parametric cumulative hazard curves**

**Description**

Plots cumulative hazard curve of right censored data given that it follows a specified parametric distribution.

**Usage**

```r
plot_cumhaz(data, dist, time = "time", censor = "censor", by = "")
```

**Arguments**

- `data`: A dataframe containing a time column and a censor column.
- `dist`: A string name for a distribution that has a corresponding density function and distribution function. Examples include "norm", "lnorm", "exp", "weibull", "logis", "llogis", "gompertz", etc.
- `time`: The string name of the time column of the dataframe. Defaults to "time".
- `censor`: The string name of the censor column of the dataframe. Defaults to "censor". The censor column must be a numeric indicator variable where complete times correspond to a value of 1 and incomplete times correspond to 0.
- `by`: The string name of a grouping variable. If specified, multiple lines will be plotted. Variable can contain logical, string, character, or numeric data.

**Examples**

```r
data("rearrest")
plot_cumhaz(rearrest, "lnorm", time = "months")
plot_cumhaz(rearrest, "weibull", time = "months", by = "personal")
```
plot_density

Plotting density function overlayed on top of a histogram of data

Description

Creates histogram of right censored data with the density function of a fitted parametric distribution overlayed.

Usage

plot_density(data, dist, time = "time", censor = "censor", by = "")

Arguments

data A dataframe containing a time column and a censor column.
dist A string name for a distribution that has a corresponding density function and distribution function. Examples include "norm", "lnorm", "exp", "weibull", "logis", "llogis", "gompertz", etc.
time The string name of the time column of the dataframe. Defaults to "time".
censor The string name of the censor column of the dataframe. Defaults to "censor". The censor column must be a numeric indicator variable where complete times correspond to a value of 1 and incomplete times correspond to 0.
by The string name of a grouping variable. If specified, the function plots each group individually along with the plot for all groups together. Variable can contain logical, string, character, or numeric data.

Examples

data("rearrest")
plot_density(rearrest, "exp", time = "months")
plot_density(rearrest, "weibull", time = "months", by = "personal")

plot_haz

Plotting parametric hazard curves

Description

Plots hazard curve of right censored data given that it follows a specified parametric distribution.

Usage

plot_haz(data, dist, time = "time", censor = "censor", by = "")
### Arguments

- **data**: A dataframe containing a time column and a censor column.

- **dist**: A string name for a distribution that has a corresponding density function and distribution function. Examples include "norm", "lnorm", "exp", "weibull", "logis", "llogis", "gompertz", etc.

- **time**: The string name of the time column of the dataframe. Defaults to "time".

- **censor**: The string name of the censor column of the dataframe. Defaults to "censor". The censor column must be a numeric indicator variable where complete times correspond to a value of 1 and incomplete times correspond to 0.

- **by**: The string name of a grouping variable. If specified, multiple lines will be plotted. Variable can contain logical, string, character, or numeric data.

### Examples

```r
data("rearrest")
plot_haz(rearrest, "logis", time = "months")
plot_haz(rearrest, "weibull", time = "months", by = "personal")
```

### Description

Creates percent-percent plot of right censored data given that it follows a specified parametric distribution.

### Usage

```r
plot_ppsurv(data, dist, time = "time", censor = "censor")
```
**plot_surv**  
*Plotting parametric survival curves*

**Description**

Plots survival curve of right censored data given that it follows a specified parametric distribution.

**Usage**

```r
plot_surv(data, dist, time = "time", censor = "censor", by = "")
```

**Arguments**

- **data**: A dataframe containing a time column and a censor column.
- **dist**: A string name for a distribution that has a corresponding density function and distribution function. Examples include "norm", "lnorm", "exp", "weibull", "logis", "llogis", "gompertz", etc.
- **time**: The string name of the time column of the dataframe. Defaults to "time".
- **censor**: The string name of the censor column of the dataframe. Defaults to "censor". The censor column must be a numeric indicator variable where complete times correspond to a value of 1 and incomplete times correspond to 0.
- **by**: The string name of a grouping variable. If specified, multiple lines will be plotted. Variable can contain logical, string, character, or numeric data.

**Examples**

```r
data("rearrest")
plot_surv(rearrest, "lnorm", time = "months")
plot_surv(rearrest, "weibull", time = "months", by = "personal")
```

**rearrest**  
*Data on time until re-incarceration for 194 inmates.*

**Description**

Henning and Frueh (1996) followed criminal activities of 194 inmates released from a medium security prison for 36 months. The data from this study can be used to investigate the time until the former inmates were re-arrested. If the former inmate had been re-arrested for a criminal act before 36 months (after initial prison release) had passed, then that former inmate’s event time was complete. If the former inmate had not been re-arrested for a criminal act after 36 months had passed, or had completely dropped out of the study, then that former inmate’s event time was right censored.
**Survival probability based on parametric distribution**

Computes probability of survival beyond time $t$ given that the data follows a specified parametric distribution.

**Usage**

```
surv_prob(data, dist, x, lower.tail = F, time = "time", censor = "censor", by = "")
```

**Arguments**

- **data**: A dataframe containing a time column and a censor column.
- **dist**: A string name for a distribution that has a corresponding density function and a distribution function. Examples include "norm", "lnorm", "exp", "weibull", "logis", "llogis", "gompertz", etc.
- **x**: A scalar quantity, time at which the probability of survival is computed.
- **lower.tail**: Logical; if `F` (default), probability is $P(T > x)$, otherwise, $P(T < x)$.
- **time**: The string name of the time column of the dataframe. Defaults to "time".
surv_summary

censor

The string name of the censor column of the dataframe. Defaults to "censor". The censor column must be a numeric indicator variable where complete times correspond to a value of 1 and incomplete times correspond to 0.

by

The string name of a grouping variable. If specified, the function prints probability for each group individually along with the overall probability. Variable can contain logical, string, character, or numeric data.

Examples

data("rearrest")
surv_prob(rearrest, "lnorm", 110, time = "months")
surv_prob(rearrest, "weibull", 90, time = "months", lower.tail = TRUE)

surv_summary

Summary statistics based on parametric distribution

Description

Estimates various statistics, including median, mean, standard deviation, and percentiles of survival time given that the data follows a specified parametric distribution.

Usage

surv_summary(data, dist, time = "time", censor = "censor", by = "")

Arguments

data

A dataframe containing a time column and a censor column.
dist

A string name for a distribution that has a corresponding density function and a distribution function. Examples include "norm", "lnorm", "exp", "weibull", "logis", "llogis", "gompertz", etc.
time

The string name of the time column of the dataframe. Defaults to "time".
censor

The string name of the censor column of the dataframe. Defaults to "censor". The censor column must be a numeric indicator variable where complete times correspond to a value of 1 and incomplete times correspond to 0.
by

The string name of a grouping variable. If specified, returns summary statistics for each group. Variable can contain logical, string, character, or numeric data.

Examples

data("rearrest")
surv_summary(rearrest, "lnorm", time = "months")
surv_summary(rearrest, "weibull", time = "months", by = "personal")
Index

* datasets
  aggressive, 2
  firstdrink, 3
  graduate, 5
  oscars, 5
  rearrest, 10

aggressive, 2, 7
compute_AD, 3, 7
firstdrink, 3, 7
fit_data, 4, 6
fitdistcens, 4

graduate, 5, 7

oscars, 5, 7

parmsurvfit, 6
parmsurvfit-package (parmsurvfit), 6
plot_cumhaz, 6, 7
plot_density, 6, 8
plot_haz, 6, 8
plot_ppsurv, 7, 9
plot_surv, 6, 10

rearrest, 7, 10

surv_prob, 6, 11
surv_summary, 6, 12