Package 'extractKM'

July 23, 2020

Type Package

Title Functions that Extract Data from the Kaplan Meier (KM) Curve

Version 0.1.0

Author Hyunsoo Hwang

Maintainer Hyunsoo Hwang <ilsanjoa@gmail.com>

Description

The algorithm by Guyot et al.(2012) derives data from KM curve that are close to the original individual patient time-to-event data (IPD). The functions in this package map from digitised curves back to KM data using the algorithm in an easier and more convenient way. There are 3 functions which require different information types or conditions.

Imports Rdpack, survival, graphics

RdMacros Rdpack

License GPL-3

Encoding UTF-8

LazyData TRUE

RoxygenNote 7.1.1

R topics documented:

---

ex.event   *Example Digitized Data for* km.event *Function*

Description

Data were extracted from Figure 1A (Boddu, 2018) digitised by the developer. There are infomration for two groups (MRD- & MRD+), where (x1,y1) and (x2,y2) contain the pairs of time and survival rate points for the 1st and 2nd group, respectively.

Usage

```
ex.event
```

Format

Data frame with x1,y1,x2,and y2 columns for time (x) and survival rate (y) points for two groups

References

Boddu P, Jorgensen J, Kantarjian H, Borthakur G, Kadia T, Daver N, Alvarado Y, Pemmaraju N, Bose P, Naqvi K, others (2018). "Achievement of a negative minimal residual disease state after hypomethylating agent therapy in older patients with AML reduces the risk of relapse." *Leukemia*, **32**(1), 241–244.

---

ex.tab.event	*Example Digitized Data for* km.tab.event *Function*

Description

Data were extracted from Figure 3A (Onecha, 2019) digitised by the developer. There are infomration for two groups (MRD- & MRD+), where (x1,y1) and (x2,y2) contain the pairs of time and survival rate points for the 1st and 2nd group, respectively.

Usage

```
ex.tab.event
```

Format

Data frame with x1,y1,x2,and y2 columns for time (x) and survival rate (y) points for two groups

References

Onecha E, Linares M, Rapado I, Ruiz-Heredia Y, Martinez-Sanchez P, Cedena T, Pratcorona M, Oteyza JP, Herrera P, Barragan E, others (2019). "A novel deep targeted sequencing method for minimal residual disease monitoring in acute myeloid leukemia." *haematologica*, **104**(2), 288–296.

---

ex.table	*Example Digitized Data for* km.table *Function*

Description

Data were extracted from Figure 1C (Malmberg, 2019) digitised by the developer. There are infomration for two groups (MRD- & MRD+), where (x1,y1) and (x2,y2) contain the pairs of time and survival rate points for the 1st and 2nd group, respectively.

Usage

```
ex.table
```

Format

Data frame with x1,y1,x2,and y2 columns for time (x) and survival rate (y) points for two groups

References

Delsing EM, Johansson SA, Nicklasson M, Lazarevic V, Stahlman S, Samuelsson T, Lenhoff S, Asp J, Ehinger M, Palmqvist L, others (2019). "Minimal residual disease assessed with deep sequencing of NPM1 mutations predicts relapse after allogeneic stem cell transplant in AML." *Leukemia \& lymphoma*, **60**(2), 409–417.

---

km.event	*Extract Data from KM Curve When Number of Events Information Is Available*

Description

This function derives or extracts IPD when only total number of events information are available

Usage

```
km.event(ssize, t.event = "NA", group.name = "group", data)
```

Arguments

ssize	vector of patients' number at risk at time points from the table
t.event	numeric value; total number of events
group.name	name of treatment group
data	data frame with 2 columns; 1st column has digitised time information and 2nd column has digitised survival rate information

Value

data frame with time to event, censoring, and group name information

References

Guyot P, Ades A, Ouwens MJ, Welton NJ (2012). "Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves." *BMC medical research methodology*, **12**(1), 9.

Examples

```
df <- ex.event[,1:2]
df <- df[!is.na(df$x1) & !is.na(df$y1),]  # only valid data
res1 <- km.event(ssize=16, t.event=13, group.name="group1", data=df); res1

df <- ex.event[,3:4]
df <- df[!is.na(df$x2) & !is.na(df$y2),]
res2 <- km.event(ssize=39, t.event=28, group.name="group2", data=df); res2

library(survival)
result <- rbind(res1,res2);
```

```
km <- survfit(Surv(Surv.Time, Censor) ~ Group, data=result);km
plot(km)
```

---

km.tab.event	*Extract Data from KM Curve When Both Number at Risk Table and Number of Events Are Available*

Description

This function derives or extracts IPD when both number at risk table and total number of events information are available

Usage

```
km.tab.event(t.risk.T, n.risk.T, t.event = "NA", group.name = "group", data)
```

Arguments

t.risk.T	vector of time points from the number at risk table
n.risk.T	vector of patients' number at risk at time points from the table
t.event	numeric value; total number of events
group.name	name of treatment group
data	data frame with 2 columns; 1st column has digitised time information and 2nd column has digitised survival rate information

Value

data frame with time to event, censoring, and group name information

References

Guyot P, Ades A, Ouwens MJ, Welton NJ (2012). "Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves." *BMC medical research methodology*, **12**(1), 9.

Examples

```
df <- ex.tab.event[,1:2]
df <- df[!is.na(df$x1) & !is.na(df$y1),]  # only valid data
res1 <- km.tab.event(t.risk.T = c(0,24,48,72,96,120),
                     n.risk.T = c(23,11,8,4,2,0),
                     t.event=5,
                     group.name="group1",
                     data=df);res1

df <- ex.tab.event[,3:4]
df <- df[!is.na(df$x2) & !is.na(df$y2),]
res2 <- km.tab.event(t.risk.T = c(0,24,48,72,96,120),
                     n.risk.T = c(12,3,3,2,1,1),
                     t.event=8,
                     group.name="group2",
                     data=df);res2
```

```
library(survival)
result <- rbind(res1,res2);
km <- survfit(Surv(Surv.Time, Censor) ~ Group, data=result);km
plot(km)
```

---

km.table	*Extract Data from KM Curve When Number at Risk Table Is Available*

Description

This function derives or extracts IPD when only number at risk table information are available

Usage

```
km.table(t.risk.T, n.risk.T, group.name = "group", data)
```

Arguments

t.risk.T	vector of time points from the number at risk table
n.risk.T	vector of patients' number at risk at time points from the table
group.name	name of treatment group
data	data frame with 2 columns; 1st column has digitised time information and 2nd column has digitised survival rate information

Value

data frame with time to event, censoring, and group name information

References

Guyot P, Ades A, Ouwens MJ, Welton NJ (2012). "Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves." *BMC medical research methodology*, **12**(1), 9.

Examples

```
df <- ex.table[,1:2]
df <- df[!is.na(df$x1) & !is.na(df$y1),]  # only valid data
res1 <- km.table(t.risk.T = c(0,25,50,75,100,125),
                 n.risk.T = c(22,18,10,7,5,2),
                 group.name="group1",
                 data=df);res1

df <- ex.event[,3:4]
df <- df[!is.na(df$x2) & !is.na(df$y2),]
res2 <- km.table(t.risk.T = c(0,25,50,75),
                 n.risk.T = c(5,1,1,0),
                 group.name="group2",
                 data=df);res2

library(survival)
result <- rbind(res1,res2);
```

```
km <- survfit(Surv(Surv.Time, Censor) ~ Group, data=result);km
plot(km)
```

Index