

Package ‘JPSurv’

July 31, 2019

Version 2.1

Date 2019/7/31

Title Methods for population-based cancer survival analysis

Author Yongwu Shao <ywshao@gmail.com>, Jun Luo <luoj@imsweb.com>

Maintainer Dennis Buck-

man <buckmand@imsweb.com>, Daniel Miller <millerd@imsweb.com>, Fanni Zhang <zhangf@imsweb.com>

Description Functions, methods, and datasets for cancer survival analysis,
including the proportional hazard relative survival model, the join point
relative survival model.

License GPL version 2 or newer

Imports ggplot2,ggrepel

R topics documented:

aapc	1
aapc.multiints	2
CoxFit	3
CoxModel_Year	3
deleteLastInterval	4
dictionary.overview	5
download.data	5
joinpoint	7
joinpoint.seerdata	9
plot.dying.year.annotate	10
plot.surv.int.multiyears	11
plot.surv.year.annotate	12

aapc

Trend summary measures for joint point relative survival model

Description

Get the trend summary measures for joint point relative survival model. Measures include annual percentage changes of hazard, annual percentage changes of cumulative relative survival, annual changes of cumulative relative survival.

Usage

```
aapc(fit.result, type="AbsChgSur")
```

Arguments

<code>fit</code>	Joinpoint object with predicted values and joinpoint model selections.
<code>type</code>	Type of trend summary measure. Supported measures are: RelChgHaz - Hazard of cancer death, AbsChgSur - Cumulative Relative Survival Average Annual-Absolute Change (in percentage points), RelChgSur - Cumulative Relative Survival Average Annual Percentage Change. The default is AbsChgSur.

Value

The estimates, standard errors, and confidence interval of the trend summary measure.

Examples

```
data(breast.example);
subsetStr="Year_of_diagnosis_1975 >= 1975 & Age_groups == \"00-49\" & Breast_stage == \"Local\"";
# Fit the survival join point model with zero join points,
# i.e., fit the proportional hazard relative survival model.
fit1 = joinpoint(data=breast.example,
                 subset = subsetStr,
                 year="Year_of_diagnosis_1975",
                 observedrelsurv="Relative_Survival_Cum",
                 model.form = NULL,
                 maxnum.jp = 0);
# Get the estimate, standard error, and confidence interval of the annual changes of cumulative hazard
haz_ac = aapc(fit1, type="AbsChgSur");
```

aapc.multiints

Trend summary measures for multiple selected intervals

Description

A function that returns the trend summary measures including Absolute Change in Survival, Relative Change in Survival and Relative Change in Hazard for multiple selected intervals.

Usage

```
aapc.multiints(fit, type = "AbsChgSur", int.select = NULL)
```

Arguments

<code>fit</code>	Joinpoint object with predicted values and joinpoint model selections.
<code>type</code>	Type of trend summary measure. Supported measures are: RelChgHaz - Hazard of cancer death, AbsChgSur - Cumulative Relative Survival Average Annual-Absolute Change (in percentage points), RelChgSur - Cumulative Relative Survival Average Annual Percentage Change. The default is AbsChgSur.
<code>int.select</code>	The single or multiple interval values selected for the trend measures calculation. The default is NULL.

Value

The estimates, standard errors, and confidence interval of the trend summary measure.

Author(s)

Fanni Zhang <zhangf@imsweb.com>

See Also

aapc

Examples

```
data("breast.example")
subsetStr<-"Age_groups=='45-65s' & Breast_stage=='Localized'"
fit <- joinpoint(breast.example, subsetStr,
                year="Year_of_diagnosis_1975", observedrelsurv="Relative_Survival_Cum",
                model.form = NULL, maxnum.jp = 3, proj.year.num = 5)
trend.rch<-aapc.multiints(fit, type="RelChgHaz",int.select=c(1,3,5))
```

CoxFit

Fitting a proportional hazard relative survival model

Description

Fitting a proportional hazard relative survival model

Usage

```
CoxFit(X, nAlive, nDied, nLost, expSurv)
```

Arguments

X	The input design matrix.
nAlive	Number of people at risk.
nDied	Number of people who died.
nLost	Number of people who are censored.
expSurv	The expected survival rate.

Value

A CoxFit class object.

CoxModel_Year

*Fitting a proportional hazard relative survival model***Description**

Fitting a proportional hazard relative survival model

Usage

```
CoxModel_Year(formula, data, subset, ...)
```

Arguments

formula	an object of class "formula": a symbolic description of the model to be fitted. If the input data is from SEER*Stat, the formula can be ~Year. Otherwise, the full formula should be specified as: ~Alive_at_Start + Died + Lost_to_Followup + Expected_Survival_Interval + Interval + Year
data	an optional data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model. If not found in data, the variables are taken from environment(formula).
subset	an optional vector specifying a subset of observations to be used in the fitting process.
...	additional arguments to be passed to the low level regression fitting functions.

Value

A list with attributes:

coefficients	a named vector of coefficients and standard errors
converged	convergence status
predicted	the fitted relative survival rates
xbeta	the linear predictor
ll	log likelihood
aic	AIC
bic	BIC

deleteLastInterval *delete records of last intervals of all years*

Description

delete records of last intervals of all years in a data.

Usage

```
deleteLastInterval(data, byvarnames, yearcol, intervalcol)
```

Arguments

data	an data frame, containing byvars, the names of which are given by byvarnames, year column, the name of which is given by yearcol, and interval column, the name of which is given by intervalcol.
byvarnames	a vector of characters that gives names of byvars.
yearcol	a character variable that gives the name of year variable.
intervalcol	a character variable that gives the name of interval variable.

Value

res	a data.frame object that contains all from data but records of last intervals of all years for cohorts defined by byvarnames.
-----	---

Examples

```
#define a data set
mydata=data.frame(
  cohort1=c("A", "A", "A", "B", "B", "B"),
  cohort2=c("C", "C", "C", "D", "D", "D"),
  year=c(1990,1990,1990,1991,1991,1991),
  interval=c(1,2,3,1,2,4)
);
x=deleteLastInterval(mydata,c("cohort1","cohort2"),"year","interval");
print(x);
```

dictionary.overview

SEERStat Dictionary Overview

Description

Reads the dictionary file and returns a parsed overview.

Usage

```
dictionary.overview(dictionary)
```

Arguments

dictionary	File name/path for dictionary (.dic) file corresponding to SEERStat data file.
------------	--

Value

A list which contains the parsed information stored in the dictionary file.

download.data	<i>Combine inputs and outputs</i>
---------------	-----------------------------------

Description

A function that returns a merged data set including the selected cohort input data and the accompanying data for plot.surv.year/plot.dying.year for the download feature in the JPSurv web app.

Usage

```
download.data(input, fit, nJP, yearvar, downloadtype, subset = NULL, interval = "Interval")
```

Arguments

input	The input dataset read in by function joinpoint.seerdata.
fit	The joinpoint object containing the model output.
nJP	The number of joinpoints in the model.
yearvar	The variable name for year of diagnosis used in argument 'year' of the function joinpoint.
downloadtype	Either "graph" for graph data or "full" for full data.
subset	An optional string specifying a subset of observations used in the fitting process.
interval	The variable name for year since diagnosis. The default is 'Interval'.
int.select	The interval values selected for the plot if the downloadtype="graph". The default is NULL.

Value

A data frame containing the selected cohort input data and the below predicted columns.

Predicted_Survival_Int	The predicted interval survival.
Predicted_ProbDeath_Int	The predicted probability of Dying of Cancer.
Predicted_Survival_Cum	The predicted cumulative survival.
Predicted_Survival_Int_SE	The standard error of the predicted interval survival.
Predicted_Survival_Cum_SE	The standard error of the predicted probability of Dying of Cancer.
Predicted_ProbDeath_Int_SE	The standard error of the predicted cumulative survival.

Author(s)

Fanni Zhang <zhangf@imsweb.com>

References

The JPSurv web app <https://analysistools.nci.nih.gov/jpsurv/>.

See Also

`plot.dying.year.annotate, plot.surv.year.annotate, plot.surv.int.multipleyears.`

Examples

```
data("breast.example")
subsetStr<-"Age_groups=='45-65s' & Breast_stage=='Localized'"
fit <- joinpoint(breast.example, subsetStr,
                year="Year_of_diagnosis_1975", observedrelsurv="Relative_Survival_Cum",
                model.form = NULL, maxnum.jp = 3, proj.year.num = 5)
nJP<-2
data.graph135<-download.data(breast.example,fit,nJP,"Year_of_diagnosis_1975","graph",subsetStr)
data.full<-download.data(breast.example,fit,nJP,"Year_of_diagnosis_1975","full",subsetStr)
```

joinpoint

Fitting a join point relative survival model

Description

Fitting a joinpoint relative survival model

Usage

```
joinpoint(data, subset=NULL, na.action = na.fail,
          year="Year", interval="Interval",
          number.event="Died", number.alive="Alive_at_Start", number.loss="Lost_to_Followup",
          expected.rate="Expected_Survival_Interval", observedrelsurv = NULL,
          model.form = NULL, maxnum.jp = 0,
          op=list(),
          delLastIntvl=F)
```

Arguments

<code>data</code>	an optional data frame, list or environment (or object coercible by <code>as.data.frame</code> to a data frame) containing the variables in the model. If not found in data, the variables are taken from <code>environment(formula)</code> .
<code>subset</code>	an optional vector specifying a subset of observations to be used in the fitting process.
<code>na.action</code>	how NAs are treated. The default is first, any <code>na.action</code> attribute of data, second a <code>na.action</code> setting of options, and third <code>na.fail</code> if that is unset. The <code><al><ae>factory-fresh<al><af></code> default is <code>na.omit</code> . Another possible value is <code>NULL</code> . The default is <code>na.fail</code>
<code>year</code>	It is about the year or time values and could be a vector of numeric or a character string giving a column name of the argument 'data'. The default is the string 'Year'.
<code>interval</code>	It is about the time elapsed from start to the event and could be a vector of numeric or a character string giving a column name of the argument 'data'. The default is the string 'Interval'.
<code>number.event</code>	It is the number of events or died and could be a vector of numeric or a character string giving a column name of the argument 'data'. The default is the string 'Died'.

<code>number.alive</code>	It is about the number of alive and could be a vector of numeric or a character string giving a column name of the argument 'data'. The default is the string 'Alive_at_Start'.
<code>number.loss</code>	It is about the number of Lost_to_Followup and could be a vector of numeric or a character string giving a column name of the argument 'data'. The default is the string 'Lost_to_Followup'.
<code>expected.rate</code>	It is about the interval expected survival and could be a vector of numeric or a character string giving a column name of the argument 'data'. The default is the string 'Expected_Survival_Interval'.
<code>observedrelsurv</code>	It is about the observed cumulative relative survival and could be a vector of numeric or a character string giving a column name of the argument 'data'. If NULL, then no observed values. The default is NULL.
<code>model.form</code>	an object of class "formula": a symbolic description of covariates. Example: <code>~-1+age+as.factor(stage)</code>
<code>proj.year.num</code>	The number of projection years for use in the prediction step. Default value is 10 years, with a valid range of 0 to 30 years.
<code>maxnum.jp</code>	The maximum number of join points allowed. The default is zero, which is equivalent to a proportional hazard relative survival model.
<code>op</code>	More options. Details —
	<code>numbetwn</code> : integer value, number of skipped obs between joinpoints exclusive (not count for the joinpoints). Default is 2.
	<code>numfromstart</code> : integer value, number of skipped obs from the first obs to joinpoints exclusive (not count for the joinpoint). Default is 3.
	<code>numtoend</code> : integer value, number of skipped obs from the first obs to joinpoints exclusive (not count for the joinpoint). Default is 4.
<code>delLastIntvl</code>	an logical value indicating whether or not deleting records of last intervals of all years. The default is false.

Value

An object of class "joinpoint" will be returned with attributes:

<code>coefficients</code>	a named vector of coefficients and standard errors
<code>jp</code>	the estimates of the join points
<code>converged</code>	convergence status
<code>predicted</code>	the fitted relative survival rates
<code>fullpredicted</code>	the full output matrix, with all year/interval combinations and projections
<code>xbeta</code>	the linear predictor
<code>ll</code>	log likelihood
<code>aic</code>	AIC
<code>bic</code>	BIC
<code>FitList</code>	a list that contains fitting results for the number of joinpoints = 0,1,...,numJPoints respectively.

References

Yu, B., Huang, L., Tiwari, R. C., Feuer, E. J. and Johnson, K. A. (2009), Modeling population-based cancer survival trends by using join point models for grouped survival data. *Journal of the Royal Statistical Society: Series A*, 172, 405-425.

Examples

```
#Load the provided SEER 18 breast cancer example data.
data(breast.example);
subsetStr="Year_of_diagnosis_1975 >= 1975 & Age_groups == \"00-49\" & Breast_stage == \"Local\"";
# Fit the survival join point model with zero join points,
# i.e., fit the proportional hazard relative survival model.
fit1 = joinpoint(data=breast.example,
                 subset = subsetStr,
                 year="Year_of_diagnosis_1975",
                 observedrelsurv="Relative_Survival_Cum",
                 model.form = NULL,
                 maxnum.jp = 0);
# The fit1 object can now be passed to either plot.surv.year or plot.surv.int, or combined with
# input data to create an input/output data table with output.overview.
```

joinpoint.seerdata *Fitting a join point relative survival model*

Description

Read in and format SEER*Stat data in a single step.

Usage

```
joinpoint.seerdata(seerfilename,newvarnames=character(0),UseVarLabelsInData=FALSE, ...)
```

Arguments

seerfilename The file name of the seer files. Here one needs two files: the seer directory file with name extension 'dic' and the seer data file with the same name of the directory file but different name extension, i.e. 'txt'.

newvarnames A list of key words used for referencing the variable in the seer directory file and data file.

UseVarLabelsInData a logic value or a character variables giving variable names. If true or the variable names, then variable labels read from the dic file will replace associated numeric values in the data.frame object, which stores data from the associated data file, for all vars in dic or those (names of which are given by UseVarLabelsInData). If false, then data read from the associated data file won't be changed.

Examples

```
# For this example we will be referencig the SEER*Stat session that was used to create the b
# data included with teh package.
# If the "breast.example.txt" is the output data file from SEER*Stat, then joinpoint.seerdat
# to input the data quickly, while taking into account the relevant information in the accom
# Input data is stored in breast.example
#breast.data = joinpoint.seerdata(seerfilename="breast.example",
#                                newvarnames=c("Age_groups", "Breast_Stage", "Year_of_diagnos
#                                UseVarLabelsInData=FALSE
#);
# The breast.data object can immediately be used as input to teh joinpoint modeling function
```

```
plot.dying.year.annotate
```

*Figure - Percent Change in the Annual Probability of Dying of Cancer
by Diagnosis Year*

Description

A function that returns a plot (or a list including the Relative Change in Hazard trend measures and the plot) for percent change in the annual probability of dying of cancer by diagnosis year using ggplot. The annotation feature is available for $nJP \leq 3$ and the number of multiple intervals selected ≤ 3 .

Usage

```
plot.dying.year.annotate(plotdata, fit, nJP, yearvar, obsintvar = "Relative_Survival")
```

Arguments

plotdata	The graph data returned by function download.data with downloadtype="graph".
fit	The joinpoint object containing the model output.
nJP	The number of joinpoints in the model.
yearvar	The variable name for year of diagnosis used in argument 'year' of the function joinpoint.
obsintvar	The variable name for observed interval survival. The default is "Relative_Survival_Interval" for relative survival data. For cause-specific data, it needs to be changed accordingly.
predintvar	The variable name for predicted interval survival. The default is "Predicted_ProbDeath_Int".
interval	The variable name for year since diagnosis. The default is 'Interval'.
annotation	The indicator for the annotation feature. The default is 0 (no annotation on the plot). Two plots with and without annotations will be returned in a list when annotation=1.
topanno	The indicator for showing the top curve annotation. The default is 1 (annotation for the top curve).
trend	The indicator for returning the Relative Change in Hazard trend measures . The default is 0 (no trend tables returned).

Value

trend=0, annotation=0
A plot with no annotation for percent change in the annual probability of dying of cancer by diagnosis year.

trend=0, annotation=1
A list containing two plots with and without annotations.

trend=1, annotation=0
A list containing a table for the relative change in hazard trend measures and a plot without annotations.

trend=1, annotation=1
A list containing a table for the relative change in hazard trend measures and two plots with and without annotations.

Author(s)

Fanni Zhang <zhangf@imsweb.com>

See Also

download.data, plot.surv.year.annotate, plot.surv.int.mutiyears, aapc.

Examples

```
data("breast.example")
subsetStr<-"Age_groups=='45-65s' & Breast_stage=='Localized'"
yearvar<-"Year_of_diagnosis_1975"
obsintvar<-"Relative_Survival_Interval"
predintvar<-"Predicted_ProbDeath_Int"
interval<-"Interval"
fit <- joinpoint(breast.example, subsetStr,
                 year=yearvar, observedrelsurv="Relative_Survival_Cum",
                 model.form = NULL, maxnum.jp = 3, proj.year.num = 5)
nJP<-3
data.graph<-download.data(breast.example, fit, nJP, yearvar, "graph", subsetStr, interval, int.sele
out.anno<-plot.dying.year.annotate(data.graph, fit, nJP, yearvar, obsintvar, predintvar, interval,
out<-plot.dying.year.annotate(data.graph, fit, nJP, yearvar, obsintvar, predintvar, interval, annot
trend.rch<-out.anno[[1]]
plot.rch.anno<-out.anno[[2]]
plot.rch<-out.anno[[3]] ### same as out[[2]]
```

plot.surv.int.mutiyears

Figure - Cumulative Survival by Interval

Description

A function that returns a plot for Cumulative Survival by Interval supporting multiple selected years.

Usage

```
plot.surv.int.mutiyears(plotdata, fit, nJP, yearvar, obscumvar = "Relative_Surviva
```

Arguments

plotdata	Either of the following data set would work. 1) the graph data returned by function download.data with downloadtype="graph" and all intervals needed for int.col. 2) the full data returned by function download.data with downloadtype="full".
fit	The joinpoint object containing the model output.
nJP	The number of joinpoints in the model.
yearvar	The variable name for year of diagnosis used in argument 'year' of the function joinpoint.
obscumvar	The variable name for observed relative cumulative survival. The default is "Relative_Survival_Cum" for relative survival data. For cause-specific data, it needs to be changed accordingly.
predcumvar	The variable name for predicted cumulative survival. The default is "Predicted_Survival_Cum".
interval	The variable name for year since diagnosis. The default is 'Interval'.
year.select	The year values selected for the plot. The default is NULL.

Author(s)

Fanni Zhang <zhangf@imsweb.com>

See Also

download.data, plot.dying.year.annotate, plot.surv.year.annotate.

Examples

```
data("breast.example")
subsetStr<-"Age_groups=='00-49' & Breast_stage=='Regional'"
fit <- joinpoint(breast.example, subsetStr,
  year="Year_of_diagnosis_1975", observedrelsurv="Relative_Survival_Cum",
  model.form = NULL, maxnum.jp = 3, proj.year.num = 5)

yearvar<-"Year_of_diagnosis_1975"
obscumvar<-"Relative_Survival_Cum"
predcumvar<-"Predicted_Survival_Cum"
interval<-"Interval"
nJP<-2
data.full<-download.data(breast.example, fit, nJP, yearvar, "full", subsetStr, interval="Interval")
plot<-plot.surv.int.multipleyears(data.full, fit, nJP, yearvar, obscumvar, predcumvar, interval, year.
```

plot.surv.year.annotate

Figure - Average Change in Cumulative Survival by Diagnosis Year

Description

A function that returns a plot (or a list including the Absolute Change in Survival trend measures and the plot) for average change in cumulative survival by diagnosis year using ggplot. The annotation feature is available for nJP<=3 and the number of multiple intervals selected <=3.

Usage

```
plot.surv.year.annotate(plotdata, fit, nJP, yearvar, obscumvar = "Relative_Survival_Cum", predcumvar = "Predicted_Survival_Cum", interval = "Interval", annotation = 0, trend = 0)
```

Arguments

plotdata	The graph data returned by function download.data with downloadtype="graph".
fit	The joinpoint object containing the model output.
nJP	The number of joinpoints in the model.
yearvar	The variable name for year of diagnosis used in argument 'year' of the function joinpoint.
obscumvar	The variable name for observed relative cumulative survival. The default is "Relative_Survival_Cum" for relative survival data. For cause-specific data, it needs to be changed accordingly.
predcumvar	The variable name for predicted cumulative survival. The default is "Predicted_Survival_Cum".
interval	The variable name for year since diagnosis. The default is 'Interval'.
annotation	The indicator for the annotation feature. The default is 0 (no annotation on the plot). Two plots with and without annotations will be returned in a list when annotation=1.
trend	The indicator for returning the Absolute Change in Survival trend measure tables. The default is 0 (no trend tables returned).

Value

trend=0, annotation=0	A plot with no annotation for average change in cumulative survival by diagnosis year.
trend=0, annotation=1	A list containing two plots with and without annotations.
trend=1, annotation=0	A list containing a table for the absolute change in survival trend measures and a plot without annotations.
trend=1, annotation=1	A list containing a table for the absolute change in survival trend measures and two plots with and without annotations.

Author(s)

Fanni Zhang <zhangf@imsweb.com>

See Also

download.data, plot.dying.year.annotate, plot.surv.int.multiyears, aapc.multiints.

Examples

```
data("breast.example")
subsetStr<-"Age_groups=='00-49' & Breast_stage=='Regional'"
yearvar<-"Year_of_diagnosis_1975"
obscumvar<-"Relative_Survival_Cum"
predcumvar<-"Predicted_Survival_Cum"
interval<-"Interval"
```

```
fit <- joinpoint(breast.example, subsetStr,
               year=yearvar, observedreldsurv=obscumvar,
               model.form = NULL, maxnum.jp = 3, proj.year.num = 5)
nJP<-2
data.graph<-download.data(breast.example,fit,nJP,yearvar,"graph",subsetStr,interval,int.sele
out.anno<-plot.surv.year.annotate(data.graph,fit,nJP,yearvar,obscumvar,predcumvar,interval,a
out<-plot.surv.year.annotate(data.graph,fit,nJP,yearvar,obscumvar,predcumvar,interval,annota
trend.acs<-out.anno[[1]]
plot.acs.anno<-out.anno[[2]]
plot.acs<-out
```