Package 'SROCt'

June 21, 2011

Type Package

Title Summary time-dependent ROC curves based on aggregated survival data

Version 1.0

Date 2011-06-20

Author Y. Foucher <Yohann.Foucher@univ-nantes.fr> and C. Combescure <Christophe.Combescure@hcuge.ch>

Maintainer Y. Foucher <Yohann.Foucher@univ-nantes.fr>

Description This package contains functions to assess the time-dependant ROC curve for a prognostic marker from aggregated data (survival probabilities in strata of the marker) and from several studies. The hazard function associated with the time-to-event is modeled as a piecewise constant function, and the package proposes 2 functions corresponding respectively to a 4-piece and a 5-piece constant functions. The area under the summary time-dependant ROC (SROCt) curve is assessed by using the trapezoidal rules.

License GPL (>=2)

LazyLoad yes

Depends splines, date, survival, nlme, statmod

Imports splines, date, survival, nlme, statmod

URL www.r-project.org, www.divat.fr

R topics documented:

SROCt-package *Summary ROC curves based on aggregated survival data*

Description

Computes the SROCt curve based on aggregate data of several studies.

Details

Package:	SROCt
Type:	Package
Version:	1.0
Date:	2011-06-15
License:	GPL (>=2)
LazyLoad:	yes

We propose an adaptation of the time-dependant ROC curve from published survival curves. The aggregate data used in this package are the survival probabilities in strata of the marker collected at various time-points. The cut-offs defining the strata of the marker are allowed to be different across the studies and need to be specified. The marker is assumed to be normally distributed. The survival function given the marker level is modeled with a piecewise-constant hazard function and the association of the marker is allowed to be different in each interval of time. Random effects are introduced in the models to take the inter-study variability into account. The regression coefficients are assessed by using the R package nlme. Illustrative data are provided: data from the cohort DIVAT, and data from a published meta-analysis (de Azambuja et al. 2007).

sroc4	Compute Summary ROC curve. The hazard function associated with the time-to-event was defined as a 4-piece constant function with a specific association of the marker at each interval.
sroc5	Compute Summary ROC curve. The hazard function associated with the time-to-event was defined as a 5-piece constant function with a specific association of the marker at each interval.
auc	Compute the area under the ROC curve from vectors of sensitivities and specificities using trapezoidal rule.

Author(s)

Y. Foucher <Yohann.Foucher@univ-nantes.fr>

References

Heagerty, P.J., Lumley, T., Pepe, M. S. (2000) Time-dependent ROC Curves for Censored Survival Data and a Diagnostic Marker.

See Also

URL: http://www.divat.fr

`aggregateDIVAT` *The kidney graft survival for different groups of recipients defined using the 1-year serum creatinine*

Description

We considered a subpopulation of 4195 adult patients and who had received a first kidney graft between January 1996 and Jun 2008 with a measure of the 1-year creatinine. Five centers participated. A total of 511 graft failures were observed (346 returns to dialysis and 165 deaths with a functional kidney). Based on this database, we constructed an aggregated dataset to perform a meta-analysis on 5 published monocentric studies.

Usage

```
data(aggregateDIVAT)
```

Format

A data frame with 106 observations (rows) with the 8 following variables (columns).

`classe` This numeric vector represents the groups of recipients defined using the 1-year serum creatinine. 1 is the first group with the lowest values of 1-year serum creatinine.

`year` This numeric vector represents the post transplant time (in years).

`surv` This numeric vector represents the survival probabilities at each year (obtained using the Kaplan and Meier estimator from the individual data).

`n` This numeric vector represents the number of recipients at the baseline (date of the transplantation) in each group.

`proba` This numeric vector represents the proportion of the patients in a center which belong to the corresponding group.

`marker.min` This numeric vector represents the minimum value of the interval of the 1-year serum creatinine (in $\mu mol/l$).

`marker.max` This numeric vector represents the maximum value of the interval of the 1-year serum creatinine (in $\mu mol/l$).

`centre.num` This numeric vector represents the centers.

Details

The immunology and nephrology department of the Nantes University hospital constituted a data bank with the monitoring of medical records for kidney and/or pancreas transplant recipients. Here, we considered a subpopulation of 4195 adult patients and who had received a first kidney graft between January 1996 and Jun 2008. Five centers participated. A total of 511 graft failures were observed (346 returns to dialysis and 165 deaths with a functional kidney). Based on this database, we constructed an aggregated dataset to perform a meta-analysis on 5 published monocentric studies. The medical objective was to evaluate whether 1-year serum creatinine (Cr) is a good predictive marker of graft failure. Cr is a breakdown product and is removed from the body by the kidneys. If kidney function is abnormal, blood Cr levels increase.

Source

URL: http://www.divat.fr.

References

M. Ladriere, Y. Foucher, C. Legendre, N. Kamar, V. Garrigue, E. Morelon, M. Kessler, J.P. Soulillou and M. Giral. The Western Europe Cohort Of Kidney Transplanted Recipients - The DIVAT Network. Clinical Transplant. 2011.

Examples

```
data(aggregateDIVAT)


# Kaplan Meier estimations of the graft survival in the first center
plot(aggregateDIVAT$year[aggregateDIVAT$centre.num==1],
 aggregateDIVAT$surv[aggregateDIVAT$centre.num==1],
 xlab="Post transplantation time (years)", ylab="Graft survival",
 ylim=c(0.7,1), xlim=c(0, 9), type="n")

# Goup 1
lines(c(0, aggregateDIVAT$year[aggregateDIVAT$centre.num==1 &
 aggregateDIVAT$classe==1]),
 c(1, aggregateDIVAT$surv[aggregateDIVAT$centre.num==1 &
 aggregateDIVAT$classe==1]),
 type="b", col=1, lty=1, lwd=2)

# Goup 2
lines(c(0, aggregateDIVAT$year[aggregateDIVAT$centre.num==1 &
 aggregateDIVAT$classe==2]),
 c(1, aggregateDIVAT$surv[aggregateDIVAT$centre.num==1 &
 aggregateDIVAT$classe==2]),
 type="b", col=2, lty=2, lwd=2)

# legend
legend("bottomleft", c("group #1 (1-year Cr<4.57)",
 "group #2 (1-year Cr>4.57)"), col=c(1, 2),
 lty=c(1, 2), lwd=c(2, 2))
```

---

auc *Area under the ROC curve from sensitivities and specificities*

Description

This function computes the area under the ROC curve using the trapezoidal rule.

Usage

```
auc(se, sp)
```

Arguments

se	A numeric vector with the sensitivities
sp	A numeric vector with the sensitivities

Details

This function computes the area under the ROC curve using the trapezoidal rule. The value of the area is directly returned.

Author(s)

Y. Foucher <Yohann.Foucher@univ-nantes.fr>

Examples

```
se.temp <- c(0, 0.5, 0.5, 1)
sp.temp <- c(1, 0.5, 0.5, 0)
auc(se.temp, sp.temp)
```

`dataKi67` *The data extracted from the Kaplan and Meier curves published in the meta-analysis by de Azambuja et al. (2007).*

Description

The available aggregated data are exhaustively presented. It concerns the meta-analysis already published by de Azambuja et al., the aim being to evaluate whether KI-67 can be considered as a good prognostic marker for breast cancer survival. The KI-67 is assumed log-normal distributed.

Usage

```
data(dataKi67)
```

Format

A data frame with 406 observations (rows) with the 10 following variables (columns).

`classe` This numeric vector represents the groups of patients defined using KI-67. 1 is the first group which is defined by the lowest KI-67 values.

`year` This numeric vector represents the survival time (in years).

`surv` This numeric vector represents the survival probabilities at each year (obtained using the Kaplan and Meier estimator from the published papers).

`n` This numeric vector represents the number of recipients at the baseline (date of KI-67 collection) in each group.

`proba` This numeric vector represents the proportion of the patients for a given paper which belong to the corresponding group.

`log.marker.min` This numeric vector represents the logarithm of the minimum value of the KI-67 interval.

`log.marker.max` This numeric vector represents the logarithm of the maximum value of the KI-67 interval.

`study.num` This numeric vector identifies the studies.

`author` This character vector identifies the first author of the paper.

`year.paper` This numeric vector identifies the year of publication.

Details

KI-67 is a marker of the proliferative activity of breast cancer, but its prognostic capacity is still unclear. In their recent meta-analysis, de Azambuja et al. concluded that KI-67 positivity conferred a worse survival. This work focused on the 35 evaluable studies of the relationship between KI-67 and the overall survival. 23 studies described survival curves according to the level of KI-67. Survival probabilities were measured every year.

References

de Azambuja E, Cardoso F, de Castro G, Colozza M, Mano M, Durbecq V, Sotiriou C, Larsimont D, Piccart-Gebhart M, Paesmans M. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12 155 patients. British Journal of Cancer 2007; 96:1504-1513.

Examples

```
data(dataKi67)

# Kaplan Meier estimations of graft survivals in Wintzer et al. (1991)
plot(dataKi67$year[dataKi67$study.num==1],
 dataKi67$surv[dataKi67$study.num==1],
 xlab="Post transplantation time (years)",
 ylab="Graft survival", ylim=c(0.6,1), xlim=c(0, 4), type="n")

# Goup 1
lines(c(0, dataKi67$year[dataKi67$study.num==1 & dataKi67$classe==1]),
 c(1, dataKi67$surv[dataKi67$study.num==1 & dataKi67$classe==1]),
 type="b", col=1, lty=1, lwd=2)

# Goup 2
lines(c(0, dataKi67$year[dataKi67$study.num==1 & dataKi67$classe==2]),
 c(1, dataKi67$surv[dataKi67$study.num==1 & dataKi67$classe==2]),
 type="b", col=2, lty=2, lwd=2)

# legend
legend("bottomleft", c("group #1 (log Ki67 < 2.49)",
 "group #2 (log Ki67 > 2.49)"), col=c(1, 2), lty=c(1, 2), lwd=c(2, 2))
```

---

sroc4 *Summary ROC curve with 4-piece hazard function*

Description

This function computes Summary ROC curve. The hazard function associated with the time-to-event was defined as a 4-piece constant function with a specific association of the marker at each interval.

Usage

```
sroc4(study.num, classe, year, surv, n, proba, marker.min,
 marker.max, init.nlme1, precision, pro.time, time.cutoff)
```

Arguments

study.num	A numeric vector (1,2,3,...) with the study identification.
classe	A numeric vector with integers (1,2,3,...) for identifying the groups defined using the studied marker. 1 is the first group with the lowest values of the marker.
year	A numeric vector with the survival times.
surv	A numeric vector with the survival probabilities corresponding to the previous times (often obtained graphically using the published survival curves).
n	A numeric vector with the number of subjects at the baseline (date of marker collection).
proba	This numeric vector represents the proportion of the patients in a center which belong to the corresponding group.
marker.min	A numeric vector with the minimum values of the marker interval corresponding to the previous class.
marker.max	A numeric vector with the maximum values of the marker interval corresponding to the previous class.
init.nlme1	A numeric vector with the initiate values (mean, sd) of the maker distribution which is assumed to be Gaussian.
precision	A numeric vector with the initiate values (mean, sd) of the maker distribution which is assumed to be Gaussian.
pro.time	The value of prognostic time is the maximum delay for which the capacity of the variable is evaluated. The same unit than the one used in the argument time.
time.cutoff	The value of internal threasholds for the definition of the piecewise hazard function (3 values).

Details

This function computes Summary ROC curve. The hazard function associated with the time-to-event was defined as a 4-piece constant function with a specific association of the marker at each interval. The maker distribution is assumed to be Gaussian distributed.

Value

nlme1	An object of class nlme representing the nonlinear mixed-effects model of the marker distribution. The marker is assumed Gaussian distributed. mu and sigma represent the mean and the standard deviation. The inter-study variability is modeled with a random effect on the mean. See nlmeObject for the components of the fit.
nlme2	An object of class nlme representing the nonlinear mixed-effects model of the time distribution. The hazard function is a stepwise function with 4 intervals. exp(beta0.1) and exp(beta0.2) represent the baseline hazard and the hazard ratio in the first interval. exp(beta1.1) and exp(beta1.2) represent the corrections of these parameters for the second interval... The inter-study variability is modeled with a random effect on the baseline parameter beta0.1. See nlmeObject for the components of the fit.
table	This data frame presents the sensitivities (se) and specificities (sp) associated with the cut-off values (cut.off).
auc	The area under the SROC curve for a prognostic up to prognostic time.

Author(s)

Y. Foucher <Yohann.Foucher@univ-nantes.fr>

References

C. Combescure, JP. Daures and Y. Foucher. A literature-based approach to evaluate the predictive capacity of a marker using time-dependent Summary Receiver Operating Characteristics (SROCt).

Examples

```
# import and attach the data example
data(aggregateDIVAT)

# Compute the SROC curve for a prognostic up to 9 years
roc9y<-sroc4(aggregateDIVAT$centre.num, aggregateDIVAT$classe,
 aggregateDIVAT$year, aggregateDIVAT$surv, aggregateDIVAT$n,
 aggregateDIVAT$proba, aggregateDIVAT$marker.min,
 aggregateDIVAT$marker.max, init.nlme1=c(4.86, -1.19),
 precision=50, pro.time=9, time.cutoff=c(2, 4, 6))

# Compute the SROC curve for a prognostic up to 3 years
roc3y<-sroc4(aggregateDIVAT$centre.num, aggregateDIVAT$classe,
 aggregateDIVAT$year, aggregateDIVAT$surv, aggregateDIVAT$n,
 aggregateDIVAT$proba, aggregateDIVAT$marker.min,
 aggregateDIVAT$marker.max, init.nlme1=c(4.86, -1.19),
 precision=50, pro.time=3, time.cutoff=c(2, 4, 6))

# The ROC graph associated to these to SROC curves
plot(c(1,1-roc9y$table$sp,0), c(1,roc9y$table$se,0),
 xlab="1-specificity", ylab="sensitivity", type="l",
 col=1, lty=1, lwd=2)

lines(c(1,1-roc3y$table$sp,0), c(1,roc3y$table$se,0), type="l",
 col=2, lty=2, lwd=2)

legend("bottomright", c(paste("Prognostic up to 9 years (AUC=",
 round(roc9y$auc,2), ")", sep=""), paste(
 "Prognostic up to 3 years (AUC=",  round(roc3y$auc,2), ")",
 sep="")), col=c(1, 2), lty=c(1, 2), lwd=c(2, 2))

# Check of the goodness-of-fit: the observed proportions of
# patients in the $g$th interval of the study $k$ versus the
# fitted proportions (equation 3).
plot(roc9y$data.marker$proba, roc9y$data.marker$fitted,
 xlab="Observed probabilities", ylab="Fitted probabilities",
 ylim=c(0,1), xlim=c(0,1))
abline(0,1)

# Check of the goodness-of-fit: the observed bivariate
# probabilities versus the fitted bivariate
# probabilities (equation 4).
plot(roc9y$data.surv$p.joint, roc9y$data.surv$fitted,
 xlab="Observed probabilities", ylab="Fitted probabilities",
 ylim=c(0,1), xlim=c(0,1))
abline(0,1)
```

```
# Check of the goodness-of-fit: the residuals of the bivariate
# probabilities (equation 4) versus the times.
plot(roc9y$data.surv$year, roc9y$data.surv$resid,
 xlab="Survival time (years)", ylab="Residuals")
lines(lowess(roc9y$data.surv$year,
 I(roc9y$data.surv$resid), iter=0))
```

`sroc5` *Summary ROC curve with 5-piece hazard function*

Description

This function computes Summary ROC curve. The hazard function associated with the time-to-event was defined as a 5-piece constant function with a specific association of the marker at each interval.

Usage

```
sroc5(study.num, classe, year, surv, n, proba, marker.min,
 marker.max, init.nlme1, precision, pro.time, time.cutoff)
```

Arguments

`study.num`	A numeric vector (1,2,3,...) with the study identification.
`classe`	A numeric vector with integers (1,2,3,...) for identifying the groups defined using the studied marker. 1 is the first group with the lowest values of the marker.
`year`	A numeric vector with the survival times.
`surv`	A numeric vector with the survival probabilities corresponding to the previous times (often obtained graphically using the published survival curves).
`n`	A numeric vector with the number of subjects at the baseline (date of marker collection).
`proba`	This numeric vector represents the proportion of the patients in a center which belong to the corresponding group.
`marker.min`	A numeric vector with the minimum values of the marker interval corresponding to the previous class.
`marker.max`	A numeric vector with the maximum values of the marker interval corresponding to the previous class.
`init.nlme1`	A numeric vector with the initiate values (mean, sd) of the maker distribution which is assumed to be Gaussian.
`precision`	A numeric vector with the initiate values (mean, sd) of the maker distribution which is assumed to be Gaussian.
`pro.time`	The value of prognostic time is the maximum delay for which the capacity of the variable is evaluated. The same unit than the one used in the argument `time`.
`time.cutoff`	The value of internal threasholds for the definition of the piecewise hazard function (4 values).

Details

This function computes Summary ROC curve. The hazard function associated with the time-to-event was defined as a 5-piece constant function with a specific association of the marker at each interval. The maker distribution is assumed to be Gaussian distributed.

Value

`nlme1`	An object of class `nlme` representing the nonlinear mixed-effects model of the marker distribution. The marker is assumed Gaussian distributed. `mu` and `sigma` represent the mean and the standard deviation. The inter-study variability is modeled with a random effect on the mean. See nlmeObject for the components of the fit.
`nlme2`	An object of class `nlme` representing the nonlinear mixed-effects model of the time distribution. The hazard function is a stepwise function with 5 intervals. `exp(beta0.1)` and `exp(beta0.2)` represent the baseline hazard and the hazard ratio in the first interval. `exp(beta1.1)` and `exp(beta1.2)` represent the corrections of these parameters for the second interval... The inter-study variability is modeled with a random effect on the baseline parameter `beta0.1`. See nlmeObject for the components of the fit.
`table`	This data frame presents the sensitivities (`se`) and specificities (`sp`) associated with the cut-off values (`cut.off`).
`auc`	The area under the SROC curve for a prognostic up to prognostic time.

Author(s)

Y. Foucher <Yohann.Foucher@univ-nantes.fr>

References

C. Combescure, JP. Daures and Y. Foucher. A literature-based approach to evaluate the predictive capacity of a marker using time-dependent Summary Receiver Operating Characteristics (SROCt).

de Azambuja E, Cardoso F, de Castro G, Colozza M, Mano M, Durbecq V, Sotiriou C, Larsimont D, Piccart-Gebhart M, Paesmans M. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12 155 patients. British Journal of Cancer 2007; 96:1504-1513.

Examples

```
# import and attach the data example
data(dataKi67)

# Compute the SROC curve for a prognostic up to 9 years
roc9y<-sroc5(dataKi67$study.num, dataKi67$classe, dataKi67$year,
 dataKi67$surv, dataKi67$n, dataKi67$proba, dataKi67$log.marker.min,
 dataKi67$log.marker.max, init.nlme1=c(2.55, -0.29),
 precision=50, pro.time=9, time.cutoff=c(2, 4, 6, 8))

# Compute the SROC curve for a prognostic up to 3 years
roc3y<-sroc5(dataKi67$study.num, dataKi67$classe, dataKi67$year,
 dataKi67$surv, dataKi67$n, dataKi67$proba, dataKi67$log.marker.min,
 dataKi67$log.marker.max, init.nlme1=c(2.55, -0.29),
 precision=50, pro.time=3, time.cutoff=c(2, 4, 6, 8))
```

```
# The ROC graph associated to these to SROC curves
plot(c(1,1-roc9y$table$sp,0), c(1,roc9y$table$se,0),
 xlab="1-specificity", ylab="sensitivity", type="l",
 col=1, lty=1, lwd=2)

lines(c(1,1-roc3y$table$sp,0), c(1,roc3y$table$se,0), type="l",
 col=2, lty=2, lwd=2)

legend("bottomright", c(paste("Prognostic up to 9 years (AUC=",
  round(roc9y$auc,2), ")", sep=""), paste(
  "Prognostic up to 3 years (AUC=",  round(roc3y$auc,2), ")",
  sep="")), col=c(1, 2), lty=c(1, 2), lwd=c(2, 2))

# Check of the goodness-of-fit: the observed proportions of
# patients in the $g$th interval of the study $k$ versus the
# fitted proportions (equation 3).
plot(roc9y$data.marker$proba, roc9y$data.marker$fitted,
 xlab="Observed probabilities", ylab="Fitted probabilities",
 ylim=c(0,1), xlim=c(0,1))
abline(0,1)

# Check of the goodness-of-fit: the observed bivariate
# probabilities versus the fitted bivariate
# probabilities (equation 4).
plot(roc9y$data.surv$p.joint, roc9y$data.surv$fitted,
 xlab="Observed probabilities", ylab="Fitted probabilities",
 ylim=c(0,1), xlim=c(0,1))
abline(0,1)

# Check of the goodness-of-fit: the residuals of the bivariate
# probabilities (equation 4) versus the times.
plot(roc9y$data.surv$year, roc9y$data.surv$resid,
 xlab="Survival time (years)", ylab="Residuals")
lines(lowess(roc9y$data.surv$year,
 I(roc9y$data.surv$resid), iter=0))
```

Index