Package: icensBKL (via r-universe)

September 16, 2024

Version 1.5

Date 2023-12-19

Title Accompanion to the Book on Interval Censoring by Bogaerts, Komarek, and Lesaffre

Depends R (>= 3.0.0), survival, MASS, Icens

Imports TeachingDemos, mvtnorm, smoothSurv, gtools, methods

Suggests interval, logspline, mixAK, bayesSurv, FHtest, icenReg, frailtypack, MLEcens, cubature, runjags, coda, dynsurv

Description Contains datasets and several smaller functions suitable for analysis of interval-censored data. The package complements the book Bogaerts, Komárek and Lesaffre (2017, ISBN: 978-1-4200-7747-6) ``Survival Analysis with Interval-Censored Data: A Practical Approach'' <https://www.routledge.com/Survival-Analysis-with-Interval-Censored-Data-A-Practical-Approach-with/Bogaerts-Komarek-Lesaffre/p/book/9781420077476>. Full R code related to the examples presented in the book can be found at <https://ibiostat.be/online-resources/icbook/supplemental>. Packages mentioned in the ``Suggests'' section are used in those examples.

Encoding UTF-8

License GPL (>= 2)

URL https://ibiostat.be/online-resources/icbook/supplemental/

NeedsCompilation no

Author Arnošt Komárek [aut, cre] (<https://orcid.org/0000-0001-8778-3762>), Kris Bogaerts [ctb] (<https://orcid.org/0000-0003-0188-8665>), Emmanuel Lesaffre [ctb]

Maintainer Arnošt Komárek <arnost.komarek@mff.cuni.cz>

Repository CRAN

Date/Publication 2023-12-20 09:50:17 UTC

Contents

`aidsCohort` *AIDS Cohort Study on Patients with Hemophilia*

Description

The study population for this example consists of 257 individuals with Type A or B hemophilia who had been treated at two French hospitals since 1978. These patients were at risk for infection by the human immunodeficiency virus (HIV) through contaminated blood factor received for their treatment. By the time of analysis, 188 patients were found to be infected with the virus, 41 of whom subsequently progressed to the acquired immunodeficiency syndrome (AIDS) or other related clinical symptoms. For reasons of simplicity, we refer to all of these events as AIDS. The primary goal of the analysis was to assess the effects of level of treatment received for hemophilia, and age, on the risk of developing AIDS-related symptoms.

The time scale was obtained by dividing the real time axis into 6-month intervals, with $T = 1$ denoting the time period from January 1, 1978 to June 30, 1978.

Usage

```
data(aidsCohort)
```

Format

a data frame with 257 rows and the following variables

L.Y left (lower) limit of interval containing the infection time.

R.Y right (upper) limit of interval containing the infection time. It is equal to NA for those with right-censored infection time at L.Y.

L.Z left (lower) limit of interval containing the time of the first clinical symptoms. It is equal to NA for those who were not infected by the end of the study period.

R.Z right (upper) limit of interval containing the time of the first clinical symptoms. It is equal to NA for those who were not infected by the end of the study period or for those whose time of the first clinical symptoms is right-censored at L.Z.

age the age covariate which is equal to 1 if the estimated age at time infection is < 20 years, and 2 otherwise.

group treatment group. It is 0 for lightly treated group and 1 for heavily treated group.

fage factor variable created from age.

fgroup factor variable created from group.

Source

Kim, M. Y., De Gruttola, V. G., and Lagakos, S. W. (1993). Analyzing doubly censored data with covariates, with application to AIDS. *Biometrics*, **49**, 13-22.

Examples

```
data("aidsCohort", package="icensBKL")
summary(aidsCohort)
```

---

aidsCT *AIDS Clinical Trial ACTG 181*

Description

AIDS Clinical Trials Group protocol ACTG 181 was a natural history substudy of a comparative trial of three anti-pneumocystis drugs. Patients were followed for shedding of cytomegalovirus (CMV) in urine and blood samples, and for colonization of mycobacterium avium complex (MAC) in the sputum and stool. Subjects were screened for CMV in the urine every 4 weeks and in the blood every 12 weeks. Subjects were screened for MAC every 12 weeks. Many patients missed several of the prescheduled clinic visits, and returned with new laboratory indications for CMV or MAC. Thus, their times until first CMV shedding or MAC colonization are censored into intervals of time when the missed clinic visits occurred. However, for the current analysis, visit times were rounded to the closest quarter because they are of practical interest to physicians in that they correspond to standard clinic schedules. Only 204 of the 232 subjects in the study who were tested for CMV shedding and MAC colonization at least once during the trial, and did not have a prior CMV or MAC diagnosis are included in the data base.

Usage

```
data(aidsCT)
```

Format

a data frame with 204 rows and the following variables

L.CMV left (lower) limit of interval (L.CMV, R.CMV] that contains time of CMV shedding (months). It is NA if the time of CMV shedding is left-censored at R.CMV.

R.CMV right (upper) limit of interval (L.CMV, R.CMV] that contains time of CMV shedding (months). It is NA if the time of CMV shedding is right-censored at L.CMV.

L.MAC left (lower) limit of interval (L.MAC, R.MAC] that contains time of MAC colonization (months). It is NA if the time of MAC colonization is left-censored at R.MAC.

R.MAC right (upper) limit of interval (L.MAC, R.MAC] that contains time of MAC colonization (months). It is NA if the time of MAC colonization is right-censored at L.MAC.

Source

Betensky, R. A. and Finkelstein, D. M. (1999). A non-parametric maximum likelihood estimator for bivariate interval censored data. *Statistics in Medicine*, **18**, 3089-3100.

Examples

```
data("aidsCT", package="icensBKL")
summary(aidsCT)
```

---

breastCancer *Cosmetic Data on Breast Cancer Patients*

Description

Breadle et al. (1984a, 1984b) report a retrospective study carried out to compare the cosmetic effects of radiotherapy alone versus radiotherapy and adjuvant chemotherapy on women with early breast cancer.

Patients were observed initially every 4 to 6 months, but as their recovery progressed, the interval betwenn visits lengthened. At each visit, the clinician recorded a measure of breast retraction on a 3-point scale (none, moderate, severe). Event of interest was the time to first appearance of moderate or severe breast retraction.

The subjects in this data were patients who had been treated at the Joint Center for Radiation Therapy in Boston between 1976 and 1980.

Usage

```
data(breastCancer)
```

Format

a data frame with 94 rows and the following variables

low lower limit of interval (lower,upper] that contains the event of interest (months). It is NA for left-censored observations.

upp upper limit of interval (lower,upper] that contains the event of interest (months). It is NA for right-censored observations.

treat treatment regimen

1 = `radio only` **=** radiotherapy only

2 = `radio+chemo` **=** radiotherapy + chemotherapy

Source

Finkelstein, D. M. and Wolfe, R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics*, **41**, 933-945. Table 4.

References

Beadle, G. F., Harris, J. R., Silver, B., Botnick, L., and Hellman, S. (1984a). Cosmetic results following primary radiation therapy for early breast cancer. *Cancer*, **54**, 2911-2918.

Beadle, G. F., Harris, J. R., Come, S., Henderson, C., Silver, B., and Hellman, S. (1984b). The effect of adjuvant chemotherapy on the cosmetic results after primary radiation treatment for early stage breast cancer: A preliminary analysis. *International Journal of Radiation Oncology, Biology and Physics*, **10**, 2131-2137.

Examples

```
data("breastCancer", package="icensBKL")
summary(breastCancer)
```

---

clayton.copula *Cumulative density function of the Clayton copula*

Description

Cumulative density function of the Clayton copula evaluated at points (u, v) with given parameter exp(beta %*%cov)

Usage

```
clayton.copula(u, v, beta, cov)
```

Arguments

u	vector of points in [0,1] representing the first coordinate where the Clayton copula must be evaluated
v	vector of points in [0,1] representing the second coordinate where the Clayton copula must be evaluated
beta	vector of coefficients to be multiplied with the covariates in order to determine the parameter of the Clayton copula
cov	vector of covariates to be multipleid with the coefficients in order to determine the parameter of the Clayton copula

Value

Cumulative density function of the Clayton copula evaluated at points (u,v) with given parameter exp(beta %*%cov)

Note

This is not to be called by the user.

Author(s)

Kris Bogaerts <kris.bogaerts@kuleuven.be>

References

Clayton, D. G. (1978). A model for association in bivariate life-tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65, 141-151.

---

contourProb *Contour Probability*

Description

Calculation of a contour probability (possible Bayesian counterpart of a p-value) based on the MCMC posterior sample for a univariate parameter corresponding to the equal-tail credible interval.

For details, see Bogaerts, Komárek and Lesaffre (201X, Sec. 9.1.4.2).

Usage

```
contourProb(sample, theta0 = 0)
```

Arguments

sample	a numeric vector with the MCMC sample from the posterior distribution of a univariate parameter.
theta0	a value of the parameter to which the contour probability is to be related.

Value

A value of the contour probability.

Author(s)

Arnošt Komárek <arnost.komarek@mff.cuni.cz>

References

Bogaerts, K., Komárek, A. and Lesaffre, E. (201X). *Survival Analysis with Interval-Censored Data: A Practical Approach*. Boca Raton: Chapman and Hall/CRC.

See Also

PGM, ictest.

Examples

```
set.seed(20170127)
sample <- rnorm(1000, mean = 2, sd = 1)
contourProb(sample)
contourProb(sample, theta0 = 2)
```

---

fit.copula *Function to fit a survival copula*

Description

The function fits a survival copula (Clayton, Gaussian or Plackett) to interval censored data using a two-stage procedure. The marginal dsitributions are fitted using an acceleated failure time model with a smoothed error distribution as implemented in the smoothSurv package. The copula parameter may depend on covariates as well.

Usage

```
fit.copula(data, copula = "normal", init.param = NULL, cov = ~1,
           marginal1 = formula(data), logscale1 = ~1, lambda1 = exp(3:(-3)),
           marginal2 = formula(data), logscale2 = ~1, lambda2 = exp(3:(-3)),
           bootstrap = FALSE, nboot = 1000,
           control1 = smoothSurvReg.control(info = FALSE),
           control2 = smoothSurvReg.control(info = FALSE),
           seed = 12345)
```

Arguments

data	Data frame in which to interpret the variables occurring in the formula.
copula	A character string specifying the copula used to fit the model. Valid choices are "normal", "clayton" or "plackett".
init.param	Optional vector of the initial values of the regression parameter(s) of the copula.
cov	A formula expression to determine a possible dependence of the copula parameter. For the Clayton and Plackett copula, the dependence will be modelled on the log-scale. For the normal copula, the dependence will be modelled modulo a Fisher transformation.
marginal1	A formula expression as for other regression models to be used in a smoothSurvReg fit for the first marginal. Use Surv on the left hand side of the formula.
logscale1	A formula expression to determine a possible dependence of the log-scale in the first marginal on covariates. It is used in a smoothSurvReg fit for the first marginal.
lambda1	A vector of values of the tuning parameter λ for the model of the first marginal. It is used in a smoothSurvReg fit for the first marginal.
marginal2	A formula expression as for other regression models to be used in a smoothSurvReg fit for the second marginal. Use Surv on the left hand side of the formula.
logscale2	A formula expression to determine a possible dependence of the log-scale in the second marginal on covariates. It is used in a smoothSurvReg fit for the second marginal.
lambda2	A vector of values of the tuning parameter λ for the model of the second marginal. It is used in a smoothSurvReg fit for the second marginal.
bootstrap	If TRUE, a bootstrap is applied in order to determine the standard erros of the copula parameter(s).
nboot	The number of bootstrap samples to be used in case the bootstrap argument is TRUE.
control1	A smoothSurvReg.control object which determines the settings for the smoothSurv fit of the first marginal.
control2	A smoothSurvReg.control object which determines the settings for the smoothSurv fit of the first marginal.
seed	seed for random numbers generator.

Value

A list with elements fit, variance, BScoefficients, BSresults.

Author(s)

Kris Bogaerts <kris.bogaerts@kuleuven.be>

Examples

```
### Signal Tandmobiel study
### Plackett copula fitted to emergence times
### of teeth 14 and 24, covariate = gender
data(tandmob, package = "icensBKL")
tand1424 <- subset(tandmob,
  select = c("GENDER", "fGENDER", "L14", "R14", "L24", "R24"))
summary(tand1424)

T1424.plackett <- fit.copula(tand1424,
   copula = "plackett", init.param = NULL, cov = ~GENDER,
   marginal1 = Surv(L14, R14, type = "interval2") ~ GENDER,
   logscale1 = ~GENDER, lambda1 = exp((-3):3),
   marginal2 = Surv(L24, R24, type = "interval2") ~ GENDER,
   logscale2 = ~GENDER, lambda2 = exp((-3):3),
   bootstrap = FALSE)
print(T1424.plackett)
```

`graft` *Homograft's Survival Times*

Description

SOMETHING WILL COME HERE.

Usage

```
data(graft)
```

Format

a data frame with 301 rows (272 patients; 243 patients contributing one observation, 29 patients contributiong two observations) and the variables that can be divided into the following conceptual groups:

Identification of the patient and the homograft

idnr identification number of a patient

Hsecond 0/1, equal to 1 if this is already the second homograft (replacement) for a given patient

Time variable and censoring indicators

timeFU follow-up time after the operation (years)

homo.repl.biod 0/1, value of 1 indicates that the homograft failed after the time `timeFU` because of biodegeneration, that is for other reasons than infection. Value of 0 indicates right censoring when considering the failure of the homograft because of biodegeneration as event

homo.failure 0/1, value of 1 indicates that the homograft failed after the time `timeFU` for various reasons. Value of 0 indicates right censoring when considering the failure of the homograft as event

death 0/1, value of 1 indicates death `timeFU` years after the operation

Basic characteristics of the patients

age age at operation (years)

gender gender of the patient

0 = `male=` male

1 = `female=` female

Description of diagnosis

clamp.time a simple way how to reflect the diagnosis, cross-clamp time (min)

position codes between grafts which are anatomically correctly (concordant) and which are not correctly (discordant) placed. It has a close relationship to the diagnosis

0 = `CONC` = concordant

1 = `DISC` = discordant

type further distinguishes *discordant* grafts according to the *Truncus* status

0 = `CONC` = concordant

1 = `DISC` = discordant, not Truncus

2 = `TRUNCUS` = discordant, Truncus

Truncus further distinguishes *discordant* grafts according to the *Truncus* status. This variable is equal to 0 for all *concordant* grafts and for all discordant, not Truncus grafts

0 = `no` = no

1 = `yes` = yes

Ross further distinguishes *concordant* grafts according to the *Ross* status. This variable is also equal to 0 for all *discordant* grafts.

0 = `no` = no

1 = `yes` = yes

Features of the homografts

Hsize size of the homograft (mm)

Hgraft donor graft

0 = `PH` = pulmonary donor graft

1 = `AH` = aortic donor graft

Immunological factors

BG.compatible is the blood group compatible between recipient and donor?

0 = `no` = no

1 = `yes` = yes

BG.recip blood group of recipient

0 = `0` = 0

1 = `A` = A

1 = `B` = B

2 = `AB` = AB

BG.donor blood group of donor

0 = `0` = 0

1 = `A` = A

1 = `B` = B

2 = `AB` = AB

Rh.recip Rhesus factor of recipient

0 = `n` = negative

1 = `p` = positive

Rh.donor Rhesus factor of donor

0 = `n` = negative

1 = `p` = positive

Iwarm warm ischemia (ischemic time) (hours)

Icold cold ischemia (ischemic time) (days)

Source

Department of Cardiac Surgery, University Hospital Gasthuisberg, Herestraat 49, 3000 Leuven, Belgium

References

Meyns, B., Jashari, R., Gewillig, M., Mertens, L., Komárek, A., Lesaffre, E., Budts, W., and Daenen, W. (2005). Factors influencing the survival of cryopreserved homografts. The second homograft performs as well as the first. *European Journal of Cardio-thoracic Surgery*, **28**, 211-216.

Examples

```
data("graft", package="icensBKL")
summary(graft)
```

`hiv` *HIV Data on Hemophilia Patients*

Description

The data arises from a 16-center prospective study in the 1980s on people with hemophilia for the purpose of investigating the risk of HIV-1 infection on these people. The event of interest is the HIV-1 infection. Patients received either no or low-dose factor VIII concentrate. Times are recorded in quarters and 0 represents January 1, 1978, the start of the epidemic and the time at which all patients are considered to be negative.

Usage

```
data(hiv)
```

Format

a data frame with 368 rows and the following variables

id patient identification number

low lower limit of interval (lower,upper] that contains the event of interest (quarters).

upp upper limit of interval (lower,upper] that contains the event of interest (quarters). It is NA for right-censored observations.

treat treatment regimen

0 = no **=** no factor VIII concentrate

1 = low dose **=** low-dose factor VIII concentrate

Source

Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data.* New York: Springer. ISBN 978-0387-32905-5. Table A.2.

References

Goedert, J. J., Kessler, C. M., Aledort, L. M., Biggar, R. J., Andes, W. A., White, G. C., Drummond, J. E., Vaidya, K., Mann, D. L., Eyster, M. E. and et al. (1989). A prospective study of human immunodeficiency virus type 1 infection and the development of AIDS in subjects with hemophilia. *The New England Journal of Medicine*, **321**, 1141-1148.

Kroner, B. L., Rosenberg, P. S., Aledort, L. M., Alvord, W. G., and Goedert, J. J. (1994) HIV-1 infection incidence among persons with hemophilia in the United States and western Europe, 1978-1990. Multicenter Hemophilia Cohort Study. *Journal of Acquired Immune Deficiency Syndromes*, **7**, 279-286.

Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data.* New York: Springer. ISBN 978-0387-32905-5. Section 3.4.

Examples

```
data("hiv", package="icensBKL")
summary(hiv)
```

`icbiplot` *Interval-censored biplot*

Description

Principal Component Analysis for interval-censored data as described in Cecere, Groenen and Lesaffre (2013).

Usage

```
icbiplot(L, R, p = 2, MaxIter = 10000, tol = 1e-06, plotit = TRUE, seed = NULL, ...)
```

Arguments

`L`	Matrix of dimension number of individuals/samples by number of variables with left endpoints of observed intervals.
`R`	Matrix of dimension number of individuals/samples by number of variables with right endpoints of observed intervals.
`p`	Dimension of the solution. Default value is $p = 2$.
`MaxIter`	Maximum number of iterations in the iterative minimazation algorithm
`tol`	Tolerance when convergence is declared
`plotit`	Logical value. Default equals TRUE. A biplot in dimension 2 is plotted.
`seed`	The seed for the random number generator. If NULL, current R system seed is used.
`...`	further arguments to be passed.

Value

Returns a list with the following components

`X`	matrix of number of individuals times 2 (p) with coordinates representing the individuals
`Y`	matrix of number of variables times 2 (p) with coordinates representing the variables
`H`	matrix of number of individuals times number of variables with approximated events
`DAF`	Disperssion accounted for (DAF) index
`FpV`	matrix showing the fit per variable
`iter`	number of iterations performed

Author(s)

Silvia Cecere, port into icensBKL by Arnošt Komárek <arnost.komarek@mff.cuni.cz>

References

Cecere, S., Groenen, P. J. F., and Lesaffre, E. (2013). The interval-censored biplot. *Journal of Computational and Graphical Statistics*, **22**(1), 123-134.

Examples

```
data("tandmob", package = "icensBKL")
Boys <- subset(tandmob, fGENDER=="boy")
L <- cbind(Boys$L14, Boys$L24, Boys$L34, Boys$L44)
R <- cbind(Boys$R14, Boys$R24, Boys$R34, Boys$R44)
L[is.na(L)] <- 0
R[is.na(R)] <- 20    ## 20 = infinity in this case

icb <- icbiplot(L, R, p = 2, MaxIter = 10000, tol = 1e-6,
                plotit = TRUE, seed = 12345)
```

---

icsurv2cdf *Conversion of icsurv objects in a data.frame*

Description

Convert an object of class `icsurv` (created by several functions of the package `Icens` to a two-column `data.frame` that can be easily used to plot the fitted distribution function.

Usage

```
icsurv2cdf(fit)
```

Arguments

`fit`	an object of class `icsurv`, typically obtained by one of the functions EM, PGM, VEM, ISDM, EMICM of the package `Icens`.

Value

A `data.frame` with columns labeled `time` and `cdf`.

Author(s)

Arnošt Komárek <arnost.komarek@mff.cuni.cz>

See Also

EM, PGM, VEM, ISDM, EMICM.

Examples

```
### Distribution function of the emergence of tooth 44 on boys
### (this example: only a subset of boys)
data("tandmob", package="icensBKL")
Boys <- subset(tandmob, fGENDER=="boy")
Sboy <- Surv(Boys$L44, Boys$R44, type="interval2")

Aboy <- subset(Boys, select=c("L44", "R44"))
Aboy$L44[is.na(Aboy$L44)] <- 0
Aboy$R44[is.na(Aboy$R44)] <- 20    ## 20 = infinity in this case
fitB.NPMLE <- EMICM(Aboy)
print(fitB.NPMLE)
plot(fitB.NPMLE)

fitB.NPMLE <- icsurv2cdf(fitB.NPMLE)
print(fitB.NPMLE)
plot(fitB.NPMLE$time, fitB.NPMLE$cdf, type="l", xlim=c(6, 13), ylim=c(0, 1),
     xlab="Age (years)", ylab="Proportion emerged")
```

`kSampleIcens` *Non-parametric comparison of k survival curves*

Description

Weighted log-rank tests for non-parametric comparison of k survival curves observed as interval-censored data. It implements an interval-censored analog to well known $G^{\varrho,\gamma}$ class of right-censored k-sample tests of Fleming and Harrington (1991, Chapter 7) proposed by Gómez and Oller (2008) and described also in Gómez et al. (2009, Sec. 3).

This R implementation considerably exploited the example code shown in Gómez et al. (2009, Sec. 3.3).

Usage

```
kSampleIcens(A, group, icsurv, rho=0, gamma=0)
```

Arguments

`A`	two column matrix or `data.frame` with lower and upper limits of observed intervals in a pooled sample. It is passed to function `PGM` from the `Icens` package which calculates the NPMLE of the cdf function based on a pooled sample.
`group`	a vector of group indicators. Its length must be the same as number of rows in `A` or as number of columns in `icsurv$clmat`.
`icsurv`	estimated cdf of based on a pooled sample. It must be an object of class `icsurv` obtained by using the function `PGM` with `A` matrix. It does not have to be supplied. Nevertheless, if supplied by the user, it is not re-calculated inside the function call which spares some computational time, especially if the test is to be run with different ϱ and γ values.

rho — parameter of the weighted log-rank (denoted as ϱ in Bogaerts, Komárek and Lesaffre (2017)).

gamma — parameter of the weighted log-rank (denoted as γ in Bogaerts, Komárek and Lesaffre (2017))

Value

An object of class `htest`.

Author(s)

Arnošt Komárek <arnost.komarek@mff.cuni.cz>

References

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.

Gómez, G. and Oller Pique, R. (2008). *A new class of rank tests for interval-censored data.* Harvard University Biostatistics Working Paper Series, Working Paper 93. https://biostats.bepress.com/harvardbiostat/paper93/

Gómez, G., Calle, M. L., Oller, R., Langohr, K. (2009). Tutorial on methods for interval-censored data and their implementation in R. *Statistical Modelling*, **9**, 259-297.

Bogaerts, K., Komárek, A. and Lesaffre, E. (2017). *Survival Analysis with Interval-Censored Data: A Practical Approach*. Boca Raton: Chapman and Hall/CRC.

See Also

PGM, ictest.

Examples

```
### Comparison of emergence distributions
##  of tooth 44 on boys and girls
data("tandmob", package="icensBKL")

  ## take only first 50 children here
  ## to decrease the CPU time
  ## of the example
tandmob50 <- tandmob[1:50,]

  ## only needed variables
Acompare <- subset(tandmob50, select=c("fGENDER", "L44", "R44"))

  ## left-censored observations:
  ##  change lower limit denoted by NA to 0
Acompare$L44[is.na(Acompare$L44)] <- 0

  ## right-censored observations:
  ##  change upper limit denoted by NA to 20
  ##  20 = infinity in this case
```

```
Acompare$R44[is.na(Acompare$R44)] <- 20

  ## inputs for kSampleIcens function
Amat <- Acompare[, c("L44", "R44")]
Group <- Acompare$fGENDER

  ## two-sample test
  ## (interval-censored version of classical Mantel's log-rank)
kSampleIcens(A=Amat, group=Group, rho=0, gamma=0)

  ## some other choices of rho and gamma,
  ## pooled CDF is supplied to kSampleIcens function
  ## to speed-up the calculation
  ## and also to set maxiter to higher value than above
  ## to ensure convergence
poolcdf <- PGM(A=Amat, maxiter=10000)

  ## IC version of classical Mantel's log-rank again
kSampleIcens(A=Amat, group=Group, icsurv=poolcdf, rho=0, gamma=0)

  ## IC version of Peto-Prentice generalization of
  ## the Wilcoxon test
kSampleIcens(A=Amat, group=Group, icsurv=poolcdf, rho=1, gamma=0)

kSampleIcens(A=Amat, group=Group, icsurv=poolcdf, rho=0, gamma=1)
kSampleIcens(A=Amat, group=Group, icsurv=poolcdf, rho=1, gamma=1)
```

`loglogistic` *Log-logistic distribution*

Description

Density, distribution function, quantile function and random generation for the log-logistic distribution.

Usage

```
dllogis(x, shape, scale=1, log=FALSE)

pllogis(q, shape, scale=1, lower.tail=TRUE, log.p=FALSE)

qllogis(p, shape, scale=1, lower.tail=TRUE, log.p=FALSE)

rllogis(n, shape, scale=1)
```

Arguments

x, q	vector of quantiles.
p	vector of probabilities.

n	number of observations.
shape	the shape parameter γ.
scale	the scale parameter α^{-1}.
log, log.p	logical; if TRUE, probabilities p are given as log(p).
lower.tail	logical; if TRUE (default), probabilities are $\mathrm{P}(X \leq x)$, otherwise, $\mathrm{P}(X > x)$.

Details

Log-logistic distribution $\mathrm{LL}(\alpha,\, \gamma)$ has a density

$$f(x) = \frac{\alpha\gamma(\alpha\, x)^{\gamma-1}}{\left\{1 + (\alpha\, x)^{\gamma}\right\}^2}, \quad x > 0,$$

and a distribution function

$$F(x) = 1 - \frac{1}{(1 + (\alpha\, x)^{\gamma})}, x > 0,$$

where α and γ are positive parameters (α is the inverse of the scale parameter and γ is the shape parameter). The mean and the variance are given by

$$\begin{aligned} \mathrm{E}X &= \frac{1}{\alpha}\frac{\pi}{\gamma \sin\left(\frac{\pi}{\gamma}\right)}, && \gamma > 1, \\ \mathrm{var}X &= \frac{1}{\alpha^2}\left\{\frac{2\pi}{\gamma \sin\left(\frac{2\pi}{\gamma}\right)} - \frac{\pi^2}{\gamma^2 \sin^2\left(\frac{\pi}{\gamma}\right)}\right\}, && \gamma > 2, \end{aligned}$$

Value

dllogis gives the density, pllogis gives the distribution function, qllogis gives the quantile function, and rllogis generates random deviates.

Author(s)

Arnošt Komárek <arnost.komarek@mff.cuni.cz>

See Also

Logistic.

Examples

```
set.seed(1977)
print(x <- rllogis(10, shape=3, scale=5))
print(d <- dllogis(x, shape=3, scale=5))
print(p <- pllogis(x, shape=3, scale=5))
qllogis(p, shape=3, scale=5)
```

mastitis *Data on Mastitis in Dairy Cattle*

Description

Mastitis in dairy cattle is the inflammation of the udder and the most important disease in the dairy sector of the western world. Mastitis reduces the milk production and the quality of the milk. For this mastitis study, 100 cows were included into the study from the time of parturition (assumed to be free of infection). They were screened monthly at the udder-quarter level for bacterial infections. Since the udder quarters are separated, one quarter might be infected while other quarters remain free of infection. The cows were followed up until the end of the lactation period, which lasted approximately 300 to 350 days. Some cows were lost to follow-up, due to, e.g. culling. Because of the approximately monthly follow-up (except during July/August for which only one visit was planned due to lack of personnel), data are interval censored. Right censored data are present when no infection occurred before the end of the lactation period or the lost to follow up time. As visits were planned independently of infection times, independent noninformative censoring is a valid assumption.

Two covariates were recorded. The first is the number of calvings, i.e., parity. This is a categorical cow-level covariate with the following categories: (1) one calving, (2) 2 to 4 calvings and (3) more than 4 calvings and is also represented by two dummy variables (representing classes 2 and 3). The second covariate is the position of the udder quarter (front or rear). Both variables have been suggested in the literature to impact the incidence of mastitis (Weller et al., 1992; Adkinson et al., 1993).

Date are adopted from Goethal et al. (2009).

Usage

```
data(mastitis)
```

Format

a data frame with 400 rows and the following variables

cow identification number of the cow.

frear `factor` indication of the location of the udder (`front`/`rear`.)

rear indication of the location of the udder, 0 = `front`, 1 = `rear`.

fpar `factor` variable indicating parity (`1`/`2-4`/`>4`.)

par1 dummy variable for a parity of 1.

par24 dummy variable for a parity of 2 to 4.

par56 dummy variable for a parity of >4.

ll lower limit of interval $(\text{lower}, \text{upper}]$ that contains time of infection.

ul upper limit of interval $(\text{lower}, \text{upper}]$ that contains time of infection.

censor censor indicator, 0 = right-censored, 1 = interval-censored. Left-censored observations have a missing value in the lower limit (`ll`). Right-censored observations have a missing value in the upper limit (`ul`).

Source

Goethals, K., Ampe, B., Berkvens, D., Laevens, H., Janssen, P. and Duchateau, L. (2009). Modeling interval-censored, clustered cow udder quarter infection times through the shared gamma frailty model. *Journal of Agricultural, Biological, and Environmental Statistics*, **14**(1), 1-14.

References

Adkinson, R. W., Ingawa, K. H., Blouin, D. C., and Nickerson, S. C. (1993). Distribution of clinical mastitis among quarters of the bovine udder. *Journal of Dairy Science*, **76**(11), 3453-3459.

Goethals, K., Ampe, B., Berkvens, D., Laevens, H., Janssen, P., and Duchateau, L. (2009). Modeling interval-censored, clustered cow udder quarter infection times through the shared gamma frailty model. *Journal of Agricultural, Biological, and Environmental Statistics*, **14**(1), 1-14.

Weller, J. I., Saran, A., and Zeliger, Y. (1992). Genetic and environmental relationships among somatic cell count, bacterial infection, and clinical mastitis. *Journal of Dairy Science*, **75**(9), 2532-2540.

Examples

```
data("mastitis", package="icensBKL")
summary(mastitis)
```

`mobile` *Survey on Mobile Phone Purchases*

Description

In February 2013 a survey on mobile phone purchases was held among 15 to 79 years old owners of a mobile phone in Finland. The participants were randomly sampled from a publicly available phone number directory by setting quotas in the gender, age and region of the respondents. A total of 536 completed interviews were recorded using a computer-assisted telephone interview (CATI) system. The amount of female owners but also 15-24 years old owners were underrepresented in the data while male and 65-79 years old owners were overrepresented in the study compared to the 2012 Finnish official statistics. The respondents answered several questions about the purchase of their current and previous mobile phone and reported also some family characteristics. More details about the survey may be found in Karvanen et al. (2014).

The purchase times are interval censored because only the month and not the day of purchase was asked. In addition, many respondents could not recall the time of purchase. For their current phone, 310 respondents were able to report the month and year of the purchase, an additional 115 were able to provide the season and year and 37 were not able to recall even the year. Out of 517 respondents who answered the questions about their previous phone, 117 were able to report the purchase month and year, an additional 91 were able to report the season and year, 146 provided only the year and 163 were not able to recall even the year. A maximum purchase interval of 200 months is assumed when the purchase year is missing. Three respondents who reported their previous phone to have bought after their current phone, 30 respondents for whom the purchase intervals of their previous and current phone are completely overlapping and 6 respondents who did not report their household size were excluded from the analysis.

The dataset shown here includes 478 respondents who correctly reported to have had a previous phone and have non-overlapping intervals for the purchase times of the previous and current phone.

Usage

```
data(mobile)
```

Format

a data frame with 478 rows and the following variables

dPLL lower limit purchase date of the previous phone.

dPUL upper limit purchase date of the previous phone.

dCLL lower limit purchase date of the current phone.

dCUL upper limit purchase date of the current phone.

PLL numeric lower limit purchase date of the previous phone (0 = January 1, 1992)

PUL numeric upper limit purchase date of the previous phone.

CLL numeric lower limit purchase date of the current phone.

CUL numeric upper limit purchase date of the current phone.

gender gender (0 = *male*, 1 = *female*).

fgender `factor` derived from `gender`.

agegrp age group (1 = 15-24 years, 2 = 25-34 years, 3 = 35-44 years, 4 = 45-54 years, 5 = 55-64 years, 6 = 65-79 years).

fagegrp `factor` derived from `agegrp`.

hhsize size of household (1 = 1 person, 2 = 2 persons, 3 = 3 persons, 4 = 4 persons, 5 = 5 persons or more).

fhhsize `factor` derived from `hhsize`.

income household income before taxes (1 = 30 000 EUR or less, 2 = 30 001-50 000 EUR, 3 = 50 001-70 000 EUR, 4 = more than 70 000 EUR, 5 = No answer).

fincome `factor` derived from `income`.

Source

https://jyx.jyu.fi/handle/123456789/77334/

References

Karvanen, J., Rantanen, A., and Luoma, L. (2014). Survey data and Bayesian analysis: A cost-efficient way to estimate customer equity. *Quantitative Marketing and Economics*, **12**(3), 305-329.

Examples

```
data("mobile", package="icensBKL")
summary(mobile)
```

`normal.copula` *Cumulative density function of the normal copula*

Description

Cumulative density function of the normal copula evaluated at points (u,v) with given parameter exp(beta %*%cov)

Usage

```
normal.copula(u, v, beta, cov)
```

Arguments

`u`	vector of points in [0,1] representing the first coordinate where the normal copula must be evaluated
`v`	vector of points in [0,1] representing the second coordinate where the normal copula must be evaluated
`beta`	vector of coefficients to be multiplied with the covariates in order to determine the parameter of the normal copula
`cov`	vector of covariates to be multipleid with the coefficients in order to determine the parameter of the normal copula

Value

Cumulative density function of the normal copula evaluated at points (u,v) with given parameter exp(beta %*%cov)

Note

This is not to be called by the user.

Author(s)

Kris Bogaerts <kris.bogaerts@kuleuven.be>

References

Nelsen, R. B. (1998). *An Introduction to Copulas*. Lecture Notes in Statistics 139. Springer-Verlag, New-York.

NPbayesSurv *Bayesian non-parametric estimation of a survival curve with right-censored data*

Description

Bayesian non-parametric estimation of a survival curve for right-censored data as proposed by Susarla and Van Ryzin (1976, 1978)

Usage

```
NPbayesSurv(time, censor,
  choice = c("exp", "weibull", "lnorm"), c = 1, parm,
  xlab = "Time", ylab = "Survival Probability", maintitle = "",
  cex.lab = 1.2, cex.axis = 1.0, cex.main = 1.5, cex.text = 1.2, lwd = 2)
```

Arguments

`time`, `censor`	numeric vectors with (right-censored) survival times and 0/1 censoring indicators (1 for event, 0 for censored)
`choice`	a character string indicating the initial guess (S^*) of the survival distribution
`c`	parameter of the Dirichlet process prior
`parm`	a numeric vector of parameters for the initial guess: `rate` parameter for exponential (see also `Exponential`), a two-element vector with `shape` and `scale` parameters for weibull (see also `Weibull`), a two-element vector with `meanlog` and `sdlog` parameters for log-normal (see also `Lognormal`). If not given, parameters for the initial guess are taken from the ML fit
`xlab`, `ylab`	labels for axes of the plot
`maintitle`	text for the main title
`cex.lab`, `cex.axis`, `cex.main`, `cex.text`, `lwd`	graphical parameters

Value

A vector corresponding to the `parm` argument

Author(s)

Emmanuel Lesaffre <emmanuel.lesaffre@kuleuven.be>, Arnošt Komárek <arnost.komarek@mff.cuni.cz>

References

Susarla, V. and Van Ryzin, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *Journal of the American Statistical Association*, **71**(356), 897-902.

Susarla, V. and Van Ryzin, J. (1978). Large sample theory for a Bayesian nonparametric survival curve estimator based on censored samples. *The Annals of Statistics*, **6**(4), 755-768.

Examples

```
## Nonparametric Bayesian estimation of a survival curve
## Homograft study, aortic homograft patients
data("graft", package = "icensBKL")

graft.AH <- subset(graft, Hgraft == "AH") # aortic homograft patients
time <- graft$timeFU[graft$Hgraft == "AH"]
censor <- graft$homo.failure[graft$Hgraft == "AH"]

  ## Initial guess: Weibull, c = 0.1 and 100
oldpar <- par(mfrow = c(1, 2))
NPbayesSurv(time, censor, "weibull", c = 100,
   xlab = "Follow-up time since the operation (years)", maintitle = "c = 100")
NPbayesSurv(time, censor, "weibull", c = 100,
   xlab = "Follow-up time since the operation (years)", maintitle = "c = 100")
par(oldpar)

  ## Initial guess: Exponential, c = 100
oldpar <- par(mfrow = c(1, 1))
NPbayesSurv(time, censor, "exp", c = 100,
   xlab = "Follow-up time since the operation (years)", maintitle = "Exp: c = 100")

  ## Initial guess: Log-normal, c = 100
NPbayesSurv(time, censor, "lnorm", c = 100,
   xlab = "Follow-up time since the operation (years)", maintitle = "Log-Normal: c = 100")
par(oldpar)
```

`NPICbayesSurv` *Bayesian non-parametric estimation of a survival curve with interval-censored data*

Description

Bayesian non-parametric estimation of a survival curve for right-censored data as proposed by Susarla and Van Ryzin (1976, 1978)

Usage

```
NPICbayesSurv(low, upp,
  choice = c("exp", "weibull", "lnorm"), cc, parm,
  n.sample = 5000, n.burn = 5000, cred.level = 0.95)
```

Arguments

`low`	lower limits of observed intervals with `NA` if left-censored
`upp`	upper limits of observed intervals with `NA` if right-censored
`choice`	a character string indicating the initial guess (S^*) of the survival distribution
`cc`	parameter of the Dirichlet process prior

`parm`	a numeric vector of parameters for the initial guess: `rate` parameter for exponential (see also `Exponential`), a two-element vector with `shape` and `scale` parameters for weibull (see also `Weibull`), a two-element vector with `meanlog` and `sdlog` parameters for log-normal (see also `Lognormal`). If not given, parameters for the initial guess are taken from the ML fit
`n.sample`	number of iterations of the Gibbs sampler after the burn-in
`n.burn`	length of the burn-in
`cred.level`	credibility level of calculated pointwise credible intervals for values of the survival function

Value

A `list` with the following components

S a `data.frame` with columns: `t` (time points), `S` (posterior mean of the value of the survival function at `t`), `Lower`, `Upper` (lower and upper bound of the pointwise credible interval for the value of the survival function)

t grid of time points (excluding an "infinity" value)

w a matrix with sampled weights

n a matrix with sampled values of n

Ssample a matrix with sampled values of the survival function

c parameter of the Dirichlet process prior which was used

choice character indicating the initial guess

parm parameters of the initial guess

n.burn length of the burn-in

n.sample number of sampled values

Author(s)

Emmanuel Lesaffre <emmanuel.lesaffre@kuleuven.be>, Arnošt Komárek <arnost.komarek@mff.cuni.cz>

References

Calle, M. L. and Gómez, G. (2001). Nonparametric Bayesian estimation from interval-censored data using Monte Carlo methods. *Journal of Statistical Planning and Inference*, **98**(1-2), 73-87.

Examples

```
## Breast Cancer study (radiotherapy only group)
## Dirichlet process approach to estimate nonparametrically
## the survival distribution with interval-censored data
data("breastCancer", package = "icensBKL")
breastR <- subset(breastCancer, treat == "radio only", select = c("low", "upp"))

### Lower and upper interval limit to be used here
low <- breastR[, "low"]
upp <- breastR[, "upp"]
```

```
### Common parameters for sampling
### (quite low, only for testing)
n.sample <- 100
n.burn <- 100

### Gibbs sampler
set.seed(19680821)
Samp <- NPICbayesSurv(low, upp, choice = "weibull", n.sample = n.sample, n.burn = n.burn)

print(ncol(Samp$w))         ## number of supporting intervals
print(nrow(Samp$S))         ## number of grid points (without "infinity")
print(Samp$S[, "t"])        ## grid points (without "infinity")
print(Samp$t)               ## grid points (without "infinity")
print(Samp$S)               ## posterior mean and pointwise credible intervals

print(Samp$w[1:10,])         ## sampled weights (the first 10 iterations)
print(Samp$n[1:10,])         ## sampled latend vectors (the first 10 iterations)
print(Samp$Ssample[1:10,])   ## sampled S (the first 10 iterations)

print(Samp$parm)     ## parameters of the guess

### Fitted survival function including pointwise credible bands
ngrid <- nrow(Samp$S)

plot(Samp$S[1:(ngrid-1), "t"], Samp$S[1:(ngrid-1), "Mean"], type = "l",
     xlim = c(0, 50), ylim = c(0, 1), xlab = "Time", ylab = expression(hat(S)(t)))
polygon(c(Samp$S[1:(ngrid-1), "t"], Samp$S[(ngrid-1):1, "t"]),
        c(Samp$S[1:(ngrid-1), "Lower"], Samp$S[(ngrid-1):1, "Upper"]),
        col = "grey95", border = NA)
lines(Samp$S[1:(ngrid - 1), "t"], Samp$S[1:(ngrid - 1), "Lower"], col = "grey", lwd = 2)
lines(Samp$S[1:(ngrid - 1), "t"], Samp$S[1:(ngrid - 1), "Upper"], col = "grey", lwd = 2)
lines(Samp$S[1:(ngrid - 1), "t"], Samp$S[1:(ngrid - 1), "Mean"], col = "black", lwd = 3)
```

`plackett.copula` *Cumulative density function of the Plackett copula*

Description

Cumulative density function of the Plackett copula evaluated at points (u,v) with given parameter exp(beta %*%cov)

Usage

```
plackett.copula(u, v, beta, cov)
```

Arguments

u	vector of points in [0,1] representing the first coordinate where the Plackett copula must be evaluated

v	vector of points in [0,1] representing the second coordinate where the Plackett copula must be evaluated
beta	vector of coefficients to be multiplied with the covariates in order to determine the parameter of the Plackett copula
cov	vector of covariates to be multipleid with the coefficients in order to determine the parameter of the Plackett copula

Value

Cumulative density function of the Plackett copula evaluated at points (u,v) with given parameter exp(beta %*%cov)

Note

This is not to be called by the user.

Author(s)

Kris Bogaerts <kris.bogaerts@kuleuven.be>

References

R. L. Plackett (1965). A class of bivariate distributions. *Journal of the American Statistical Association*, **60**, 516-522.

---

survfitS.smoothSurvReg

Survivor function for Objects of Class 'smoothSurvReg'

Description

Compute survivor function at left and right endpoint based on the fitted model. The function is an adapted version of the survfit.smoothSurvReg function of the package smoothSurv.

Usage

```
survfitS.smoothSurvReg(formula, cov, logscale.cov)
```

Arguments

formula	Object of class smoothSurvReg.
cov	Vector or matrix with covariates values for which the survivor function is to be computed. It must be a matrix with as many columns as is the number of covariates (interactions included) or the vector of length equal to the number of covariates (interactions included). Intercept is not to be included in cov. If cov is missing a survivor curve for the value of a covariate vector equal to zero is calculated.

`logscale.cov` Vector or matrix with covariate values for the expression of log-scale (if this depended on covariates). It can be omitted in the case that log-scale was common for all observations.

Value

A `data.frame` with columns named `S1` and `S2` containing the value of survivor function at the left and right endpoints, respectively.

Note

This is not to be called by the user.

Author(s)

Kris Bogaerts <kris.bogaerts@kuleuven.be>

---

tandmob *Signal Tandmobiel Data, Subsample*

Description

This is a random sample of 500 children (256 boys and 244 girls) from the dataset resulting from a longitudinal prospective dental study performed in Flanders (North of Belgium) in 1996 – 2001. The cohort of 4 468 randomly sampled children who attended the first year of the basic school at the beginning of the study was annualy dental examined by one of 16 trained dentists. The original dataset consists thus of at most 6 dental observations for each child.

The dataset presented here contains mainly the information on the emergence and caries times summarized in the interval-censored observations. In addition to the interval censored observation of the emergence time of several teeth, a random visit was selected in order to create current status data of the emergence times for the same children. Also the time to caries of the four permanent first molars are included in the data set. Finally, the covariates gender, frequency of brushing, the presence of sealants and occlusal plaque on the first permanent molars collected in the first year were included as potential confounders in the data set.

For more detail on the design of the study see Vanobbergen et al. (2000).

Usage

```
data(tandmob)
```

Format

a data frame with 500 rows and the following variables

IDNR identification number of a child.

GENDER numeric gender; 0=boy, 1=girl.

fGENDER factor derived from a variable `GENDER`.

DMF_1 dmft-score at baseline around the age of 7 years.

OH16o, OH26o, OH36o, OH46o numeric occlusal plaque status of teeth 16, 26, 36, 46 (permanent first molars); 0=no plaque, 1=pits/fissures, 2=total.

fOH16o, fOH26o, fOH36o, fOH46o factors derived from variables `OH16o, OH26o, OH36o, OH46o`.

BRUSHING numeric brushing frequency at baseline; 0=less than daily, 1=at least once a day.

fBRUSHING factor derived from a variable `BRUSHING`.

SEAL16, SEAL26, SEAL36, SEAL46 numeric baseline presence of sealing on teeth 16, 26, 36, 46 (permanent first molars); 0=no, 1=yes.

fSEAL16, fSEAL26, fSEAL36, fSEAL46 factors derived from variables `SEAL16, SEAL26, SEAL36, SEAL46`.

DMF_54, DMF_64, DMF_74, DMF_84 numeric baseline dmft score of teeth 54, 64, 74, 84 (primary first molars); 0=sound, 1=caries experience.

DMF_55, DMF_65, DMF_75, DMF_85 numeric baseline dmft score of teeth 55, 65, 75, 85 (primary second molars); 0=sound, 1=caries experience.

fDMF_54, fDMF_64, fDMF_74, fDMF_84 factors derived from variables `DMF_54, DMF_64, DMF_74, DMF_84`.

fDMF_55, fDMF_65, fDMF_75, fDMF_85 factors derived from variables `DMF_55, DMF_65, DMF_75, DMF_85`.

L14, L24, L34, L44 left (lower) limit of observed emergence time of teeth 14, 24, 34, 44 (permanent first premolars), `NA` for left-censored observations.

R14, R24, R34, R44 right (upper) limit of observed emergence time of teeth 14, 24, 34, 44 (permanent first premolars), `NA` for right-censored observations.

L15, L25, L35, L45 left (lower) limit of observed emergence time of teeth 15, 25, 35, 45 (permanent second premolars), `NA` for left-censored observations.

R15, R25, R35, R45 right (upper) limit of observed emergence time of teeth 15, 25, 35, 45 (permanent second premolars), `NA` for right-censored observations.

L16, L26, L36, L46 left (lower) limit of observed emergence time of teeth 16, 26, 36, 46 (permanent first molars), `NA` for left-censored observations.

R16, R26, R36, R46 right (upper) limit of observed emergence time of teeth 16, 26, 36, 46 (permanent first molars), `NA` for right-censored observations.

CS_age age at visit selected to determine current status.

CS_14, CS_24, CS_34, CS_44 numeric current status of teeth 14, 24, 34, 44 (permanent first premolars) at age given by variable `CS_age`; 0=not emerged, 1=emerged.

fCS_14, fCS_24, fCS_34, fCS_44 factors derived from variables `CS_14, CS_24, CS_34, CS_44`.

CL16, CL26, CL36, CL46 left (lower) limit of observed time to caries of teeth 16, 26, 36, 46 (permanent first molars), `NA` for left-censored observations.

CR16, CR26, CR36, CR46 right (upper) limit of observed time to caries of teeth 16, 26, 36, 46 (permanent first molars), `NA` for right-censored observations.

Source

Leuven Biostatistics and Statistical Bioinformatics Centre (L-Biostat), Katholieke Universiteit Leuven, Kapucijnenvoer 35, 3000 Leuven, Belgium

URL: http://med.kuleuven.be/biostat/

Data collection was supported by Unilever, Belgium. The Signal Tandmobiel project comprises the following partners: D. Declerck (Dental School, Catholic University Leuven), L. Martens (Dental School, University Ghent), J. Vanobbergen (Oral Health Promotion and Prevention, Flemish Dental Association), P. Bottenberg (Dental School, University Brussels), E. Lesaffre (Biostatistical Centre, Catholic University Leuven), K. Hoppenbrouwers (Youth Health Department, Catholic University Leuven; Flemish Association for Youth Health Care).

References

Carvalho, J. C., Ekstrand, K. R., and Thylstrup, A. (1989). Dental plaque and caries on occlusal surfaces of first permanent molars in relation to stage of eruption. *Journal of Dental Research*, **68**, 773–779.

Vanobbergen, J., Martens, L., Lesaffre, E., and Declerck, D. (2000). The Signal-Tandmobiel project – a longitudinal intervention health promotion study in Flanders (Belgium): baseline and first year results. *European Journal of Paediatric Dentistry*, **2**, 87–96.

Examples

```
data("tandmob", package="icensBKL")
summary(tandmob)
```

---

tandmobAll *Signal Tandmobiel Data*

Description

This is the dataset resulting from a longitudinal prospective dental study performed in Flanders (North of Belgium) in 1996 – 2001. The cohort of 4\,468 randomly sampled children who attended the first year of the basic school at the beginning of the study was annualy dental examined by one of 16 trained dentists. The original dataset consists thus of at most 6 dental observations for each child.

The dataset presented here contains mainly the information on the emergence and caries times summarized in the interval-censored observations. Some baseline covariates are also included here.

For more detail on the design of the study see Vanobbergen et al. (2000).

Usage

```
data(tandmobAll)
```

Format

a data frame with 4\,430 rows (38 sampled children did not come to any of the designed dental examinations) and the following variables

IDNR identification number of a child

GENDER character *boy* or *girl*

GENDERNum numeric, 0 = *boy*, 1 = *girl*

DOB character, date of birth in the format DDmmmYY

PROVINCE factor, code of the province with

- **0 =** `Ant` = Antwerpen
- **1 =** `VlB` = Vlaams Brabant
- **2 =** `Lim` = Limburg
- **3 =** `OVl` = Oost Vlaanderen
- **4 =** `WVl` = West Vlaanderen

EDUC factor, code of the educational system with

- **0 =** `Free` = Free school
- **1 =** `Community` = Community school
- **2 =** `Province/council` = Province/council school

STARTBR factor, code indicating the starting age of brushing the teeth (as reported by parents) with

- **1 =** `<=1` = [0, 1] years
- **2 =** `(1,2]` = (1, 2] years
- **3 =** `(2,3]` = (2, 3] years
- **4 =** `(3,4]` = (3, 4] years
- **5 =** `(4,5]` = (4, 5] years
- **6 =** `>5` = later than at the age of 5

EBEG.xx lower limit of the emergence (in years of age) of the permanent tooth xx. `NA` if the emergence was left-censored.

xx takes values 11, 21, 31, 41 (permanent incisors), 12, 22, 32, 42 (permanent central canines), 13, 23, 33, 43 (permanent lateral canines), 14, 24, 34, 44 (permanent first premolars), 15, 25, 35, 45 (permanent second premolars), 16, 26, 36, 46 (permanent first molars), 17, 27, 37, 47 (permanent second molars).

EEND.xx upper limit of the emergence (in years of age) of the permanent tooth xx. `NA` if the emergence was right-censored.

xx takes values as for the variable `EBEG.xx`.

FBEG.xx lower limit for the caries time (in years of age, 'F' stands for 'failure') of the permanent tooth xx. `NA` if the caries time was left-censored.

xx takes values as for the variable `EBEG.xx`.

FEND.xx upper limit for the caries time (in years of age, 'F' stands for 'failure') of the permanent tooth xx. `NA` if the caries time was right-censored.

xx takes values as for the variable `EBEG.xx`.

Unfortunately, for all teeth except 16, 26, 36 and 46 almost all the caries times are right-censored. For teeth 16, 26, 36, 46, the amount of right-censoring is only about 25%.

Txx.DMF indicator whether a deciduous tooth xx was ***d****ecayed* or ***m****issing due to caries* or ***f****illed* on at most the last examination before the first examination when the emergence of the permanent successor was recorded.

xx takes values 53, 63, 73, 83 (deciduous lateral incisors), 54, 64, 74, 84 (deciduous first molars), 55, 65, 75, 85 (deciduous second molars).

Txx.CAR indicator whether a~deciduous tooth xx was removed due to the orthodontical reasons or decayed on at most the last examination before the first examination when the emergence of the permanent successor was recorded.

Source

Biostatistical Centre, Katholieke Universiteit Leuven, Kapucijnenvoer 35, 3000 Leuven, Belgium

URL: http://med.kuleuven.be/biostat/

Data collection was supported by Unilever, Belgium. The Signal Tandmobiel project comprises the following partners: D. Declerck (Dental School, Catholic University Leuven), L. Martens (Dental School, University Ghent), J. Vanobbergen (Oral Health Promotion and Prevention, Flemish Dental Association), P. Bottenberg (Dental School, University Brussels), E. Lesaffre (Biostatistical Centre, Catholic University Leuven), K. Hoppenbrouwers (Youth Health Department, Catholic University Leuven; Flemish Association for Youth Health Care).

References

Vanobbergen, J., Martens, L., Lesaffre, E., and Declerck, D. (2000). The Signal-Tandmobiel project – a longitudinal intervention health promotion study in Flanders (Belgium): baseline and first year results. *European Journal of Paediatric Dentistry*, **2**, 87–96.

Examples

```
data("tandmobAll", package="icensBKL")
summary(tandmobAll)
```

---

yoghurt *Sensory Shelf Life of Yoghurt*

Description

For the current study, the objective was to determine the sensory shelf life (SSL) of whole-fat, stirred yoghurt with strawberry pulp. A reversed storage design was used in which yoghurt pots of 150 g were kept at 4°C, and some of them were stored in a 42°C oven for 0, 4, 8, 12, 24, 36 and 48 hours. These times were chosen because previous experiments showed that deterioration in flavor occurred quickly up to approximately 12 hours and then slowed down. After being stored at 42°C, the samples were refrigerated at 4°C until they were tasted. More details about the experiment can be found in Hough (2010).

Fifty adults between 18 and 30 years and 50 children between 10 and 12 years who consumed stirred yoghurt at least once a week were recruited from a town in Argentina. For each of the 7 samples presented in random order, the subject tasted the sample and answered the question: "Would you

normally consume this product? Yes or No?". If a subject would consume the samples up to 8 hours' storage but not the samples with 12 hours' storage or longer, it is known that SSL is somewhere between 8 and 12 hours storage. The data are thus interval-censored. Right-censored data occur when the subject accepts all samples and left-censored data if the sample with the first storage time is rejected. For subjects with inconsistent answers, several options to construct the interval are possible. Here, the widest uncertainty interval as to the storage time at which the subject rejects the yoghurt was applied. That is, from the first "yes" before a "no" until the last "no" which occurs after a "yes".

Several subjects were excluded from the analysis (4 adults and 3 children) because they preferred the stored product to the fresh product. The data are reproduced from Hough (2010).

Usage

```
data(yoghurt)
```

Format

a data frame with 93 rows and the following variables

left lower (left) limit of interval that contains an event of interest, set to `NA` for left-censored observations.

right upper (right) limit of interval that contains an event of interest, set to `NA` for right-censored observations.

adult a binary variable indicating whether respondent is child (0) or adult (1).

fadult `factor` derived from variable `adult`.

Source

Hough, G. (2010). *Sensory Shelf Life Estimation of Food Products*. CRC press. ISBN 9781420092912.

References

Hough, G. (2010). *Sensory Shelf Life Estimation of Food Products*. CRC press. ISBN 9781420092912.

Examples

```
data("yoghurt", package="icensBKL")
summary(yoghurt)
```

Index