DOWNLOAD PDF SURVIVAL ANALYSIS WITH R

Chapter 1 : CRAN Task View: Survival Analysis

Kaplan Meier Analysis. The first thing to do is to use Surv() to build the standard survival object. The variable t1 records the time to death or the censored time; d1 indicates that the patient died (d1 = 1) or that the patient survived until the end of the study (d1 = 0).

Share Tweet With roots dating back to at least when John Graunt, a London merchant, published an extensive set of inferences based on mortality records, Survival Analysis is one of the oldest subfields of Statistics [1]. Basic life-table methods, including techniques for dealing with censored data, were known before [2]. In the early eighteenth century, the old masters, de Moivre working on annuities and Daniel Bernoulli studying competing risks for his work on smallpox inoculation, developed the foundations of time-to-event modeling [2]. Today, survival analysis models are important in Engineering, Insurance, Marketing and Medicine and many more application areas. So, it is not surprising that the R Task View on Survival Analysis, a curated, organized and annotated list of relevant R packages and functions, is formidable. Looking at the Task View on a small screen is a bit like standing too close to a brick wall â€" left-right, up-down, bricks all around. It is a fantastic edifice that gives some idea of the significant contributions R developers have made both to the theory and practice of Survival Analysis. As well-organized as it is, however, I imagine that even survival analysis experts need some time to find their way around this task view. I would be remiss not to mention that we all owe a great deal of gratitude to Arthur Allignol and Aurielien Latouche, the task view maintainers. Newcomers, people either new to R or new to survival analysis or both, must find it overwhelming. So, it is with newcomers in mind that I offer the following slim trajectory through the task view that relies on just a few packages: Not only is the package itself rich in features, but the object created by the Surv function, which contains failure time and censoring information, is the basic survival analysis data structure in R. My main reason for selecting the OIsurv package was to draw attention the very helpful guide to Survival Analysis in R , produced by the folks at OpenIntro. The ranger function is well-known for being a fast implementation of the Random Forests algorithm for building ensembles of classification and regression trees. But it was news to me that ranger also builds survival models. Benchmarks indicate that ranger is suitable for building time-to-event models with the large, high dimensional data sets important to internet marketing applications. I chose this because has number of covariates and no missing values. The only data preparation is to make the appropriate variables into factors. These are large-sample, simultaneous estimates and are plotted in red. They contrast with the point-wise confidence bands rendered as black dashed lines. Note, however, that this model does achieve an R2 value of 0. Note that ranger builds a model for each observation in the data set. The next block of code builds the model using the same variables used the the Cox model above, and each of the survival curves computed for the bmt data set, along with a curve of average values. Note that ta, tc, and dc are the same top three variables flagged in the Cox model. This four-package excursion only hints at the Survival Analysis tools that are available in R, but it does illustrate some of the richness of the R platform which has been under continuous development and improvement for nearly twenty years. The use of the Surv function shows how open source code allows generations of developers to build on the work of their predecessors. References For convenience, I have collected the references used throughout the post here. Cambridge University Press, 2nd ed. Wiley, p [3] Kaplan, E. Non-parametric estimation from incomplete observations, J American Stats Assn. Regression models and life-tables with discussion , Journal of the Royal Statistical Society B 34, â€" Survival Analysis in R.

Chapter 2 : Survival Analysis â€“ Part I | DataScience+

The survival package is the cornerstone of the entire R survival analysis edifice. Not only is the package itself rich in features, but the object created by the Surv() function, which contains failure time and censoring information, is the basic survival analysis data structure in R. Dr. Terry Therneau, the package author, began working on the.

We currently use R 2. You may want to make sure that packages on your local machine are up to date. You can perform update in R using update. We use the conf. We will use lifetab function presented in package KMsurv. In order to be able to use function lifetab, we need to create a couple of variables, mainly the number of censored at each time point and the number of events at each time point. Also notice that the time intervals have been grouped. The first step is to create grouped data. We use function gsummary from package nlme here to create grouped data. Based on the grouped data, we will create a couple of new variables for lifetab. Function lifetab requires that the length of the time variable is 1 greater than other variables, such as the variable of number of events, or the variable of number of censored. The confidence intervals in the book are calculated based on the standard errors. We write a function called stci for this calculation. Here is the definition of stci: The mean of the survivorship function, p. We will use survdiff for tests. Function survdiff is a family of tests parameterized by parameter rho. The following description is from R Documentation on survdiff: We will create a categorical age variable, agecat first. The easiest way to get Nelson-Aalen estimator is via cox regression using coxph function.

Chapter 3 : R Survival Analysis

Survival Analysis in R June David M Diez OpenIntro theinnatdunvilla.com This document is intended to assist individuals who are theinnatdunvilla.comdgable about the basics of survival analysis.

This is reinforced by the three significant tests of equality. Tests of equality of the survival function In the output we find three Chi-square based tests of the equality of the survival function over strata, which support our suspicion that survival differs between genders. In a nutshell, these statistics sum the weighted differences between the observed number of failures and the expected number of failures for each stratum at each timepoint, assuming the same survival function of each stratum. In other words, if all strata have the same survival function, then we expect the same proportion to die in each interval. Nonparametric estimation of the hazard function Standard nonparametric techniques do not typically estimate the hazard function directly. However, we can still get an idea of the hazard rate using a graph of the kernel-smoothed estimate. We generally expect the hazard rate to change smoothly if it changes over time, rather than jump around haphazardly. To accomplish this smoothing, the hazard function estimate at any time interval is a weighted average of differences within a window of time that includes many differences, known as the bandwidth. Widening the bandwidth smooths the function by averaging more differences together. However, widening will also mask changes in the hazard function as local changes in the hazard function are drowned out by the larger number of values that are being averaged together. Below is an example of obtaining a kernel-smoothed estimate of the hazard function across BMI strata with a bandwidth of days: From the plot we can see that the hazard function indeed appears higher at the beginning of follow-up time and then decreases until it levels off at around days and stays low and mostly constant. The hazard function is also generally higher for the two lowest BMI categories. The sudden upticks at the end of follow-up time are not to be trusted, as they are likely due to the few number of subjects at risk at the end. The red curve representing the lowest BMI category is truncated on the right because the last person in that group died long before the end of followup time. The Cox proportional hazards regression model 4. The hazard function for a particular time interval gives the probability that the subject will fail in that interval, given that the subject has not failed up to that point in time. In regression models for survival analysis, we attempt to estimate parameters which describe the relationship between our predictors and the hazard rate. A common way to address both issues is to parameterize the hazard function as: For such studies, a semi-parametric model, in which we estimate regression parameters as covariate effects but ignore leave unspecified the dependence on time, is appropriate. The exponential function is also equal to 1 when its argument is equal to 0. This parameterization forms the Cox proportional hazards model. It is called the proportional hazards model because the ratio of hazard rates between two groups with fixed covariates will stay constant over time in this model. Because of this parameterization, covariate effects are multiplicative rather than additive and are expressed as hazard ratios, rather than hazard differences. Instead, we need only assume that whatever the baseline hazard function is, covariate effects multiplicatively shift the hazard function and these multiplicative shifts are constant over time. Cox models are typically fitted by maximum likelihood methods, which estimate the regression parameters that maximize the probability of observing the given set of survival times. Cox proportional hazards regression in SAS using proc phreg 5. Previously, we graphed the survival functions of males in females in the WHAS dataset and suspected that the survival experience after heart attack may be different between the two genders. Perhaps you also suspect that the hazard rate changes with age as well. Below we demonstrate a simple model in proc phreg, where we determine the effects of a categorical predictor, gender, and a continuous predictor, age on the hazard rate: To specify that gender is a categorical predictor, we enter it on the class statement. On the model statement, on the left side of the equation, we provide the follow up time variable, lenfol, and the censoring variable, fstat, with all censoring values listed in parentheses. On the right side of the equation we list all the predictors.

Chapter 4 : Introduction to Survival Analysis in SAS

Survival analysis in R The core survival analysis functions are in the survival package. The survival package is one of the few "core" packages that comes bundled with your basic R installation, so you probably didn't need to theinnatdunvilla.comes() it.

Share Tweet With roots dating back to at least when John Graunt, a London merchant, published an extensive set of inferences based on mortality records, survival analysis is one of the oldest subfields of Statistics [1]. Basic life-table methods, including techniques for dealing with censored data, were discovered before [2], and in the early eighteenth century, the old masters â€" de Moivre working on annuities, and Daniel Bernoulli studying competing risks for the analysis of smallpox inoculation â€" developed the modern foundations of the field [2]. Today, survival analysis models are important in Engineering, Insurance, Marketing, Medicine, and many more application areas. So, it is not surprising that R should be rich in survival analysis functions. We all owe a great deal of gratitude to Arthur Allignol and Aurielien Latouche, the task view maintainers. Looking at the Task View on a small screen, however, is a bit like standing too close to a brick wall â€" left-right, up-down, bricks all around. It is a fantastic edifice that gives some idea of the significant contributions R developers have made both to the theory and practice of Survival Analysis. As well-organized as it is, however, I imagine that even survival analysis experts need some time to find their way around this task view. Newcomers â€" people either new to R or new to survival analysis or both â€" must find it overwhelming. So, it is with newcomers in mind that I offer the following narrow trajectory through the task view that relies on just a few packages: Not only is the package itself rich in features, but the object created by the Surv function, which contains failure time and censoring information, is the basic survival analysis data structure in R. Terry Therneau, the package author, began working on the survival package in The first public release, in late , used the Statlib service hosted by Carnegie Mellon University. Thereafter, the package was incorporated directly into Splus , and subsequently into R. The ranger function is well-known for being a fast implementation of the Random Forests algorithm for building ensembles of classification and regression trees. But ranger also works with survival data. Benchmarks indicate that ranger is suitable for building time-to-event models with the large, high-dimensional data sets important to internet marketing applications. Load the data This first block of code loads the required packages, along with the veteran dataset from the survival package that contains data from a two-treatment, randomized trial for lung cancer. The times parameter of the summary function gives some control over which times to print. Here, it is set to print the estimates for 1, 30, 60 and 90 days, and then every 90 days thereafter. This is the simplest possible model. It only takes three lines of R code to fit it, and produce numerical and graphical summaries. First, I create a new data frame with a categorical variable AG that has values LT60 and GT60, which respectively describe veterans younger and older than sixty. While I am at it, I make trt and prior into factor variables. But note, survfit and npsurv worked just fine without this refinement. However, some caution needs to be exercised in interpreting these results. For example, the Cox model assumes that the covariates do not vary with time. In a vignette [12] that accompanies the survival package Therneau, Crowson and Atkinson demonstrate that the Karnofsky score karno is, in fact, time-dependent so the assumptions for the Cox model are not met. The vignette authors go on to present a strategy for dealing with time dependent covariates. Data scientists who are accustomed to computing ROC curves to assess model performance should be interested in the Concordance statistic. Notice the steep slope and then abrupt change in slope of karno. Note however, that there is nothing new about building tree models of survival data. The next block of code builds the model using the same variables used in the Cox model above, and plots twenty random curves, along with a curve that represents the global average for all of the patients. Note that I am using plain old base R graphics here. Also note that the importance results just give variable names and not level names. This is because ranger and other tree models do not usually create dummy variables. This is a generalization of the ROC curve, which reduces to the Wilcoxon-Mann-Whitney statistic for binary variables, which in turn, is equivalent to computing the area under the ROC curve. This apparently is a challenge. In a paper [16], Hamad observes: However, in the

context of survival trees, a further difficulty arises when timeâ€”varying effects are included. Hence, we feel that the interpretation of covariate effects with tree ensembles in general is still mainly unsolved and should attract future research. I believe that the major use for tree-based models for survival data will be to deal with very large data sets. The following code pulls out the survival data from the three model objects and puts them into a data frame for ggplot. I suspect that there are neither enough observations nor enough explanatory variables for the ranger model to do better. This four-package excursion only hints at the Survival Analysis tools that are available in R, but it does illustrate some of the richness of the R platform, which has been under continuous development and improvement for nearly twenty years. The ranger package, which suggests the survival package, and ggfortify, which depends on ggplot2 and also suggests the survival package, illustrate how open-source code allows developers to build on the work of their predecessors. For an elementary treatment of evaluating the proportional hazards assumption that uses the veterans data set, see the text by Kleinbaum and Klein [13]. See the paper [15] by Intrator and Kooperberg for an early review of using classification and regression trees to study survival data. References For convenience, I have collected the references used throughout the post here. Cambridge University Press, 2nd ed. Non-parametric estimation from incomplete observations, J American Stats Assn. Regression models and life-tables with discussion , Journal of the Royal Statistical Society B 34, pp.

Chapter 5 : Applied Survival Analysis, Chapter 2 | R Textbook Examples

Or copy & paste this link into an email or IM.

Definitions of common terms in survival analysis[edit] The following terms are commonly used in survival analyses: Death, disease occurrence, disease recurrence, recovery, or other experience of interest Time: The time from the beginning of an observation period such as surgery or beginning treatment to i an event, or ii end of the study, or iii loss of contact or withdrawal from the study. If a subject does not have an event during the observation time, they are described as censored. The subject is censored in the sense that nothing is observed or known about that subject after the time of censoring. A censored subject may or may not have an event after the end of observation time. Survival function S t: The probability that a subject survives longer than time t. Acute myelogenous leukemia survival data[edit] This example uses the Acute Myelogenous Leukemia survival data set "aml" from the "survival" package in R. The aml data set sorted by survival time is shown in the box. Censoring indicates that the patient did not have an event no recurrence of aml cancer. This subject was only in the study for 13 weeks, and the aml cancer did not recur during those 13 weeks. It is possible that this patient was enrolled near the end of the study, so that they could only be observed for 13 weeks. It is also possible that the patient was enrolled early in the study, but was lost to follow up or withdrew from the study. The remaining subjects all experienced events recurrence of aml cancer while in the study. The question of interest is whether recurrence occurs later in maintained patients than in non-maintained patients. Kaplan-Meier plot for the aml data[edit] The Survival function S t , is the probability that a subject survives longer than time t. S t is theoretically a smooth curve, but it is usually estimated using the Kaplan-Meier KM curve. The graph shows the KM plot for the aml data and can be interpreted as follows: The x axis is time, from zero when observation began to the last observed time point. The y axis is the proportion of subjects surviving. The solid line similar to a staircase shows the progression of event occurrences. A vertical drop indicates an event. In the aml table shown above, two subjects had events at 5 weeks, two had events at 8 weeks, one had an event at 9 weeks, and so on. These events at 5 weeks, 8 weeks and so on are indicated by the vertical drops in the KM plot at those time points. At the far right end of the KM plot there is a tick mark at weeks. The vertical tick mark indicates that a patient was censored at this time. In the aml data table five subjects were censored, at 13, 16, 28, 45 and weeks. There are five tick marks in the KM plot, corresponding to these censored observations. Life table for the aml data[edit] A life table summarizes survival data in terms of the number of events and the proportion surviving at each event time point. The life table for the aml data, created using the R software, is shown. Life table for the aml data The life table summarizes the events and the proportion surviving at each event time point. The columns in the life table have the following interpretation: Being "at risk" means that the subject has not had an event before time t, and is not censored before or at time t. Testing for differences in survival in the aml data[edit] The log-rank test compares the survival times of two or more groups. This example uses a log-rank test for a difference in survival in the maintained versus non-maintained treatment groups in the aml data. The graph shows KM plots for the aml data broken out by treatment group, which is indicated by the variable "x" in the data. Kaplan-Meier graph by treatment group in aml The null hypothesis for a log-rank test is that the groups have the same survival. The expected number of subjects surviving at each time point in each is adjusted for the number of subjects at risk in the groups at each event time. The log-rank test determines if the observed number of events in each group is significantly different from the expected number. The formal test is based on a chi-squared statistic. When the log-rank statistic is large, it is evidence for a difference in the survival times between the groups. The log-rank statistic approximately has a chi-squared distribution with one degree of freedom, and the p-value is calculated using the chi-squared distribution. The sample size of 23 subjects is modest, so there is little power to detect differences between the treatment groups. The chi-squared test is based on asymptotic approximation, so the p-value should be regarded with caution for small sample sizes. Cox proportional hazards PH regression analysis [edit] Kaplan-Meier curves and log-rank tests are most useful when the predictor variable is categorical e. For quantitative predictor variables, an alternative method is Cox

proportional hazards regression analysis. Cox proportional hazards regression analysis for melanoma[edit] This example uses the melanoma data set from Dalgaard Chapter The Cox proportional hazards regression using R gives the results shown in the box. Cox proportional hazards regression output for melanoma data. Predictor variable is sex 1: The Cox regression results are interpreted as follows. Sex is encoded as a numeric vector 1: The R summary for the Cox model gives the hazard ratio HR for the second group relative to the first group, that is, male versus female. The summary for the Cox model gives the hazard ratio for the second group relative to the first group, that is, male versus female. The estimated hazard ratio of 1. Dividing the coef by its standard error gives the z score. Finally, the output gives p-values for three alternative tests for overall significance of the model: For large enough N, they will give similar results. For small N, they may differ somewhat. The Likelihood ratio test has better behavior for small sample sizes, so it is generally preferred. Cox model using a covariate in the melanoma data[edit] The Cox model extends the log-rank test by allowing the inclusion of additional covariates. Regression models, including the Cox model, generally give more reliable results with normally-distributed variables. For this example use a log transform. The log of the thickness of the tumor looks to be more normally distributed, so the Cox models will use log thickness. The Cox PH analysis gives the results in the box. Cox PH output for melanoma data set with covariate log tumor thickness The p-value for all three overall tests likelihood, Wald, and score are significant, indicating that the model is significant. The p-value for log thick is 6. Because the confidence interval for HR includes 1, these results indicate that sex makes a smaller contribution to the difference in the HR after controlling for the thickness of the tumor, and only trend toward significance. Examination of graphs of log thickness by sex and a t-test of log thickness by sex both indicate that there is a significant difference between men and women in the thickness of the tumor when they first see the clinician. The Cox model assumes that the hazards are proportional. The proportional hazard assumption may be tested using the R function cox. A p-value is less than 0. Additional tests and graphs for examining a Cox model are described in the textbooks cited. Extensions to Cox models[edit] Cox models can be extended to deal with variations on the simple analysis. The subjects can be divided into strata, where subjects within a stratum are expected to be relatively more similar to each other than to randomly chosen subjects from other strata. The regression parameters are assumed to be the same across the strata, but a different baseline hazard may exist for each stratum. Stratification is useful for analyses using matched subjects, for dealing with patient subsets, such as different clinics, and for dealing with violations of the proportional hazard assumption. Some variables, such as gender and treatment group, generally stay the same in a clinical trial. Other clinical variables, such as serum protein levels or dose of concomitant medications may change over the course of a study. Cox models may be extended for such time-varying covariates. Tree-structured survival models[edit] The Cox PH regression model is a linear model. It is similar to linear regression and logistic regression. Specifically, these methods assume that a single line, curve, plane, or surface is sufficient to separate groups alive, dead or to estimate a quantitative response survival time. In some cases alternative partitions give more accurate classification or quantitative estimates. One set of alternative methods are tree-structured survival models, including survival random forests. Tree-structured survival models may give more accurate predictions than Cox models. Examining both types of models for a given data set is a reasonable strategy. Example survival tree analysis[edit] This example of a survival tree analysis uses the R package "rpart". The example is based on stage C prostate cancer patients in the data set stagec in rpart. The variables in stagec are: Survival tree for prostate cancer data set Each branch in the tree indicates a split on the value of a variable. The terminal nodes indicate the number of subjects in the node, the number of subjects who have events, and the relative event rate compared to the root. Survival random forests[edit] An alternative to building a single survival tree is to build many survival trees, where each tree is constructed using a sample of the data, and average the trees to predict survival. This is the method underlying the survival random forest models. The randomForestSRC package includes an example survival random forest analysis using the data set pbc. In the example, the random forest survival model gives more accurate predictions of survival than the Cox PH model. The prediction errors are estimated by bootstrap re-sampling.

Chapter 6 : Survival Analysis in R Programming - Steps To Perform Analysis - DataFlair

Implementation of a Survival Analysis in R. With these concepts at hand, you can now start to analyze an actual dataset and try to answer some of the questions above.

However, this failure time may not be observed within the relevant time period, producing so-called censored observations. This task view aims at presenting the useful R packages for the analysis of time to event data. Please let the maintainers know if something is inaccurate or missing. The Task View is also on github. Feel free to open an issue or submit a pull request. The prodlim package implements a fast algorithm and some features not included in survival. Various confidence intervals and confidence bands for the Kaplan-Meier estimator are implemented in the km. Surv of package eha plots the Kaplan-Meier estimator. The kaplan-meier function in spatstat computes the Kaplan-Meier estimator from histogram data. The KM function in package rhosp plots the survival function using a variant of the Kaplan-Meier estimator in a hospitalisation risk context. The survPresmooth package computes presmoothed estimates of the main quantities used for right-censored data, i. The asbio package permits to compute the Kaplan-Meier estimator following Pollock et al. The bpcp package provides several functions for computing confidence intervals of the survival distribution e. The lbiassurv package offers various length-bias corrections to survival curve estimation. Non-Parametric confidance bands for the Kaplan-Meier estimator can be computed using the kmconfband package. The landest package allows landmark estimation and testing of survival probabilities. The jackknifeKME package computes the original and modified jackknife estimates of Kaplan-Meier estimators. The condSURV package provides methods for estimating the conditional survival function for ordered multivariate failure time data. The gte package implements the generalised Turnbull estimator proposed by Dehghan and Duchesne for estimating the conditional survival function with interval-censored data. The Icens package provides several ways to compute the NPMLE of the survival distribution for various censoring and truncation schemes. The DTDA package implements several algorithms permitting to analyse possibly doubly truncated survival data. The fitdistrplus package permits to fit an univariate distribution by maximum likelihood. Data can be interval censored. The vitality package provides routines for fitting models in the vitality family of mortality models. Hazard Estimation The muhaz package permits to estimate the hazard function through kernel methods for right-censored data. The ICE package aims at estimating the hazard function for interval censored data. The bshazard package provides non-parametric smoothing of the hazard through B-splines. Testing The survdiff function in survival compares survival curves using the Fleming-Harrington G-rho family of test. NADA implements this class of tests for left-censored data. The exactRankTests implements the shift-algorithm by Streitberg and Roehmel for computing exact conditional p-values and quantiles, possibly for censored data. SurvTest in the coin package implements the logrank test reformulated as a linear rank test. The maxstat package performs tests using maximally selected rank statistics. The interval package implements logrank and Wilcoxon type tests for interval-censored data. Three generalised logrank tests and a score test for interval-censored data are implemented in the glrt package. The Survgini package proposes to test the equality of two survival distributions based on the Gini index. The FHtest package offers several tests based on the Fleming-Harrington class for comparing surival curves with right- and interval-censored data. The LogrankA package provides a logrank test for which aggregated data can be used as input. The short term and long term hazard ratio model for two samples survival data can be found in the YPmodel package. The controlTest implements a nonparametric two-sample procedure for comparing the median survival time. The survRM2 package performs two-sample comparison of the restricted mean survival time The emplik2 package permits to compare two samples with censored data using empirical likelihood ratio tests. Regression Modelling Cox model: The coxph function in the survival package fits the Cox model. An implementation of weighted estimation in Cox regression can be found in coxphw. The coxrobust package proposes a robust implementation of the Cox model. The mfp package permits to fit Cox models with multiple fractional polynomial. The NestedCohort fits Cox models for covariates with missing data. A Cox model model can be fitted to data from complex survey design using the svycoxph function in survey. The multipleNCC package

fits Cox models using a weighted partial likelihood for nested case-control studies. The dynsurv package fits time-varying coefficient models for interval censored and right censored survival data using a Bayesian Cox model, a spline based Cox model or a transformation model. The OrdFacReg package implements the Cox model using an active set algorithm for dummy variables of ordered factors. The survivalMPL package fits Cox models using maximum penalised likelihood and provide a non parametric smooth estimate of the baseline hazard function. A Cox model with piecewise constant hazards can be fitted using the pch package. The isoph allows nonparametric estimation of an isotonic covariate effect for proportional hazards model. The icenReg package implements several models for interval-censored data, e. A Cox type Self-Exciting Intensity model can be fitted to right-censored data using the coxsei package. The SurvLong contains methods for estimation of proportional hazards models with intermittently observed longitudinal covariates. The plac package provides routines to fit the Cox model with left-truncated data using augmented information from the marginal of the truncation times. The cumres function in gof computes goodness-of-fit methods for the Cox proportional hazards model. The proportionality assumption can be checked using the cox. The coxphQuantile in the latter package draws a quantile curve of the survival distribution as a function of covariates. The multcomp package computes simultaneous tests and confidence intervals for the Cox model and other parametric survival models. The lsmeans package permits to obtain least-squares means and contrasts thereof from linear models. In particular, it provides support for the coxph, survreg and coxme functions. The multtest package on Bioconductor proposes a resampling based multiple hypothesis testing that can be applied to the Cox model. Testing coefficients of Cox regression models using a Wald test with a sandwich estimator of variance can be done using the saws package. The rankhazard package permits to plot visualisation of the relative importance of covariates in a proportional hazards model. The smoothHR package provides hazard ratio curves that allows for nonlinear relationship between predictor and survival. The PHeval package proposes tools to check the proportional hazards assumption using a standardised score process. Parametric Proportional Hazards Model: The eha and mixPHM packages implement a proportional hazards model with a parametric baseline hazard. The pphsm in rms translates an AFT model to a proportional hazards form. The polspline package includes the hare function that fits a hazard regression model, using splines to model the baseline hazard. Hazards can be, but not necessarily, proportional. The flexsurv package implements the model of Royston and Parmar The model uses natural cubic splines for the baseline survival function, and proportional hazards, proportional odds or probit functions for regression. The SurvRegCensCov package allows estimation of a Weibull Regression for a right-censored endpoint, one interval-censored covariate, and an arbitrary number of non-censored covariates. The survreg function in package survival can fit an accelerated failure time model. A modified version of survreg is implemented in the rms package psm function. It permits to use some of the rms functionalities. The eha package also proposes an implementation of the AFT model function aftreg. An AFT model with an error distribution assumed to be a mixture of G-splines is implemented in the smoothSurv package. The NADA package proposes the front end of the survreg function for left-censored data. A least-square principled implementation of the AFT model can be found in the lss package. The simexaft package implements the Simulation-Extrapolation algorithm for the AFT model, that can be used when covariates are subject to measurement error. A robust version of the accelerated failure time model can be found in RobustAFT. An alternative weighting scheme for parameter estimation in the AFT model is proposed in the imputeYn package. Both survival and timereg fit the additive hazards model of Aalen in functions aareg and aalen, respectively. A version of the Cox-Aalen model for interval censored data is available in the coxinterval package. The uniah package fits shape-restricted additive hazards models. The addhazard package contains tools to fit additive hazards model to random sampling, two-phase sampling and two-phase sampling with auxiliary information. The bj function in rms and BJnoint in emplik compute the Buckley-James model, though the latter does it without an intercept term. The bujar package fits the Buckley-James model with high-dimensional covariates L2 boosting, regression trees and boosted MARS, elastic net. Functions like survreg can fit other types of models depending on the chosen distribution, e. The AER package provides the tobit function, which is a wrapper of survreg to fit the tobit model. An implementation of the tobit model for cross-sectional data and panel data can be found in the

censReg package. The timereg package provides implementation of the proportional odds model and of the proportional excess hazards model. The invGauss package fits the inverse Gaussian distribution to survival data. The model is based on describing time to event as the barrier hitting time of a Wiener process, where drift towards the barrier has been randomized with a Gaussian distribution. The pseudo package computes the pseudo-observation for modelling the survival function based on the Kaplan-Meier estimator and the restricted mean.

Chapter 7 : Survival analysis - Wikipedia

Survival analysis deals with predicting the time when a specific event is going to occur. It is also known as failure time analysis or analysis of time to death. For example predicting the number of days a person with cancer will survive or predicting the time when a mechanical system is going to.

Chapter 8 : RPubs - Introduction to Survival analysis in R

Survival Analysis in R David Diez This document is intended to assist an individual who has familiarity with R and who is taking a survival analysis course.

Chapter 9 : Survival Analysis with Plotly: R vs Python

Survival analysis is a set of statistical methods for analyzing the occurrence of events over time. It is also used to determine the relationship of co-variates to the time-to-events, and accurately compare time-to-event between two or more groups.