Introduction to the “Survival” Library for R

I have used R to create Survival Analysis outputs during the presentations for these resources, and I have also used R to develop exercises. In this document I will provide a short overview of R libraries and commands that I have used.

# The “Survival” Library

The main library is “survival”, which you can install and load using the following commands:

install.package(“survival”)

library(survival)

You can check this repository for more information:

<https://cran.r-project.org/web/packages/survival/index.html>

# Create Survival Objects

Once installed, you can use the package to create a *survival object*, i.e. ensure that the dataset you are using is read as a dataset for survival analysis.

For example, using the dataset employed as an exemplar in Chapter 10 of Singer and Willett’s (2003) book “*Applied Longitudinal Data Analysis”*, you can get the “teachers” dataset and prepare it by using the Surv command from the *survival* library:

library(survival)

*# Preliminary steps to create Table 10.1 on page 327.*

teachers<-read.table("https://stats.idre.ucla.edu/stat/examples/alda/teachers.csv", sep=",", header=T)

*# Create a variable to indicate whether the event took place or not, based on censoring:*

teachers$event <- ifelse(teachers$censor == 1, 0, 1)

*# check the new variable is correctly specified*

table(teachers$event, teachers$censor)

*#create a survival object:*

**survobj<-Surv(teachers$t, teachers$event)**

The final line creates a survival object I called “survobj” using the command Surv : the first argument specifies the time variable in the survival analysis dataset, while the second argument specifies the variable indicating the event.

See the UCLA Advanced Research Computing web-pages for a slightly different example:

<https://stats.oarc.ucla.edu/r/examples/alda/r-applied-longitudinal-data-analysis-ch-10/>

If you want to check what the survival object looks like you can use:

survobj[1:15]

which will show this output:

[1] 1 2 1 1 12+ 1 12+ 1 2 2 7 12+ 1 12+ 12+

The output lists the first 15 cases in the dataset. The variable reported is the timing of the event of interest. In some cases, the timing is followed by the “+” sign: this indicates the case is censored.

# Estimate Key Survival Analysis Statistics

Once you have created a survival object, you can use this as an argument to create survival curves and collate other relevant information. To this end, the “survival” library offers the survfit function. Following from the previous commands, you can use the survival object created as follows:

ts <- survfit(survobj ~ 1, data = teachers)

The line above creates the “ts” object. The ~ 1 indicates that the survival curve should be estimated for the entire dataset as a whole, without stratifying by any covariates or groups. (Covariates can be included by substituting 1 with the covariate(s) of interest.)

You will see that the object “ts” contains a set of information about the model:

A screenshot of a computer program

Description automatically generated

For example, the number of cases in the dataset (ts$n), the time intervals (ts$time), the risk set within each time interval (ts$n.risk), the number of events observed in each time interval (ts$n.event), the number of censored cases within each time interval (ts$n.censor), the survival function for each time interval (ts$surv) and other information (e.g., the 95% confidence intervals of the survival function based on the Kaplan-Meier formula).

When we want to create a life table or plot the survival function, we can extract the information obtained through the survfit function.

However, the survfit function does not estimate the hazard function, but just the cumulative hazard function. To calculate the hazard function we can resort to its formula, where for each time interval *j*, the hazard function is estimated as:

Therefore, we can simply calculate the hazard function using the information in the survival function estimate created above:

h<-ts$n.event/ts$n.risk

Once we estimate the hazard function, we can include it in a dataset to report it in the life tables or to plot it.

Further examples of life tables and plotting the estimated functions based on Singer and Willatt’s (2003) book are available here:

<https://stats.oarc.ucla.edu/r/examples/alda/r-applied-longitudinal-data-analysis-ch-10/>

and here:

<https://stats.oarc.ucla.edu/other/examples/alda/>