Original Research

Statistical Graphical User Interface Plug-In for Survival Analysis in R Statistical and Graphics Language and Environment

Daniel C. LEUCUŢA*, Andrei ACHIMAŞ CADARIU

„Iuliu Haţieganu" University of Medicine and Pharmacy Cluj-Napoca, Department of Medical Informatics and Biostatistics, 6 Louis Pasteur, 400349 Cluj-Napoca, Romania.
E-mail: dleucuta@umfcluj.ro

* Author to whom correspondence should be addressed; Tel.: +4-0264-431697; Fax: +4-0264-593847.

Abstract: *Introduction:* R is a statistical and graphics language and environment. Although it is extensively used in command line, graphical user interfaces exist to ease the accommodation with it for new users. Rcmdr is an R package providing a basic-statistics graphical user interface to R. Survival analysis interface is not provided by Rcmdr. The AIM of this paper was to create a plug-in for Rcmdr to provide survival analysis user interface for some basic R survival analysis functions. *Materials and Methods:* The Rcmdr plug-in code was written in Tinn-R. The plug-in package was tested and built with Rtools. The plug-in was installed and tested in R with Rcmdr package on a Windows XP workstation with the "aml" and "kidney" data sets from survival R package. *Results*: The Rcmdr survival analysis plug-in was successfully built and it provides the functionality it was designed to offer: interface for Kaplan Meier and log log survival graph, interface for the log-rank test, interface to create a Cox proportional hazard regression model, interface commands to test and assess graphically the proportional hazard assumption, and influence observations. *Conclusion:* Rcmdr and R though their flexible and well planed structure, offer an easy way to expand their functionality that was used here to make the statistical environment more user friendly in respect with survival analysis.

Keywords: Rcmdr library; Rcmdr plug-in; R language and environment; Survival analysis.

Introduction

R is a language and a fully planned coherent and flexible environment for statistical computing and graphics. It offers a very large spectrum of statistical and high quality graphical techniques: statistical tests, linear and nonlinear modeling, classification, clustering, survival analysis, time series analysis. In addition to these it provides facilities for data manipulation, calculations (including vector and matrix operators). R also provides a fully developed, robust object oriented programming language (named S), that handles user defined functions (simple or recursive), input and output facilities, grouping expressions, control statements like conditional execution (if statements), repetitive execution (for loops, repeat and while) [1].

The R functionality can be extended by creation of packages.

Beside this wealth of functionalities it is also free to use and modify, being open source software, with a GPL 2 license (GNU Public License).

R is used extensively in command line. Although this is a powerful way to work with it, this makes it much less appealing to new users, who might benefit from using it. There are several graphical user interfaces (GUI) to R. Some of the most known GUIs are Rcmdr (a R package that provides basic-statistics GUI for R with script, output windows, menus, buttons and information fields [2]), SciViews-R (a series of packages providing a GUI application programming interface for R [3]), JGR (Java Gui for R - a universal and unified Graphical User Interface for R [4]), RKWard

(a transparent front-end to R providing a convenient user-interface and integration with an office suite [5]).

Rcmdr GUI provides basic-statistics tools. Through plug-in packages it can be extended to offer user interface to more advanced statistical tools.

Survival analysis interface is not yet implemented in Rcmdr. Survival analysis is a set of statistical techniques designed to deal with survival data. Survival data is a special type of data that represents the time between a subject entering into a study and the occurrence of a predefined event. This type of data is very important in medicine, due to the necessity to study ways to lengthen the life span, or to lengthen the life span without signs or symptoms. Survival analysis is also used in many other research fields from technical ones to life sciences.

The AIM of this paper was to create a plug-in for Rcmdr to provide survival analysis user interface for some basic R survival analysis functions.

Material and Method

The code for the survival analysis Rcmdr plug-in package was written in Tinn-R (a free simple replacement for the basic editor provided by the R command line user interface [6]). The code was based on other pieces of code from Rcmdr code, like twoSampleWilcoxonTest, or linearModel functions. In addition, pieces of code from other papers were used [7].

The check for correctness and the building of the Rcmdr plug-in package was made by the use of Rtools – a collection of resources (Perl, command line tools, MinGW compilers) for building packages for R under Microsoft Windows [8].

The testing was made in R [9], with the Rcmdr [10] package loaded on a Windows XP workstation. The data from "aml" [11] and "kidney" [12] data sets was used for testing. Both datasets can be found in the survival [13] package.

The plug-in was made to provide user interface to some basic R survival analysis functions found in the survival package: interface to draw a Kaplan-Meier graph, interface for a log-rank test, interface for creating a Cox proportional hazard regression, interfaces for Cox regression diagnostics (test and graphical assessments of proportional hazard assumption, and assessment of outliers).

Results

P The survival analysis Rcmdr plug-in that was built was named RcmdrPlugin.SurvivalT. After a prior installation of the package in R by calling Rcmdr library from R, Rcmdr GUI starts. In the Rcmdr Tools menu the survival analysis plug-in, can be installed. After installation a new menu called SurvivalT is added to the standard Rcmdr menu.

In order to analyze survival data, a dataset containing survival data should be set as an active data set through Rcmdr Data menu. If the data set contains numeric and factor variables then the commands in the SurvivalT menu becomes active (Fig. 1).

By selecting Kaplan Meyer graph command from the Survival T menu a dialog windows shows up as can be seen in Fig. 2, along with the resulted graph, for the aml data set from the survival package.

The -LogLogSurvival graph command provides a similar dialog window with the Kaplan Meier graph one.

In order to test for differences in survival time between two or more groups the Log rank test command from SurvivalT menu should be selected. The dialog window along with its result is shown in Fig. 3.

To do a Cox proportional hazard regression the Cox Model command should be selected from the SurvivalT menu. The dialog window is shown in Fig. 4.
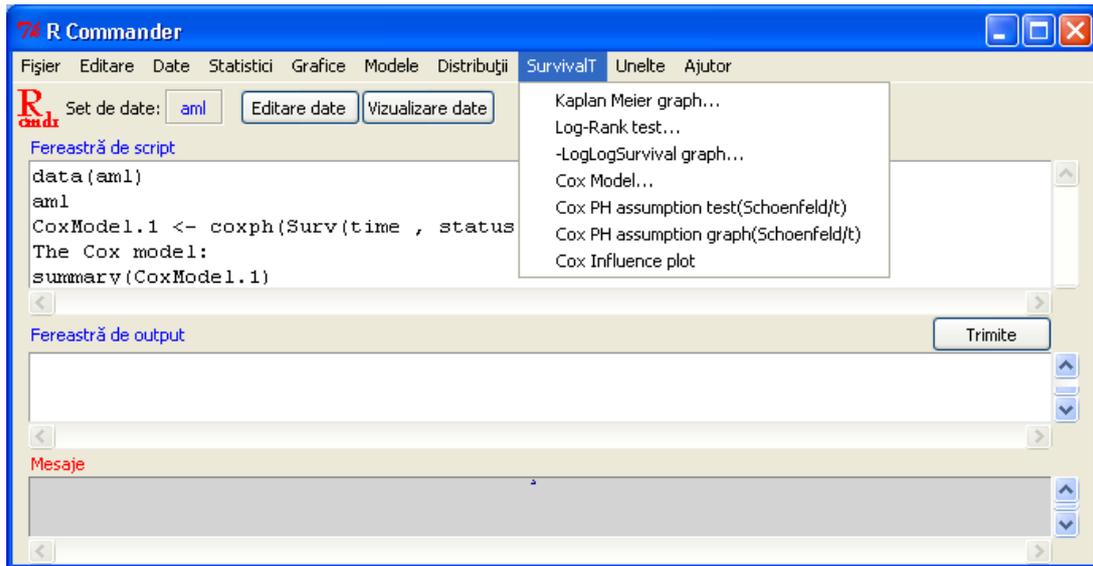
Figure 1. The SurvivalT menu that shows up by installing the Rcmdr survival analysis plug-in
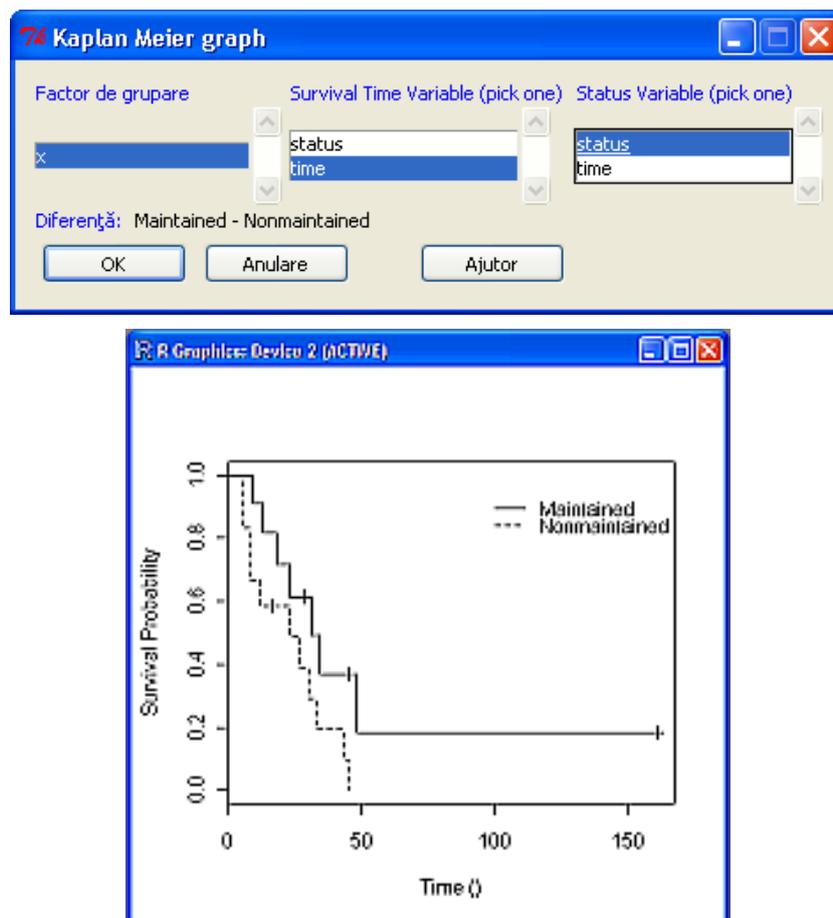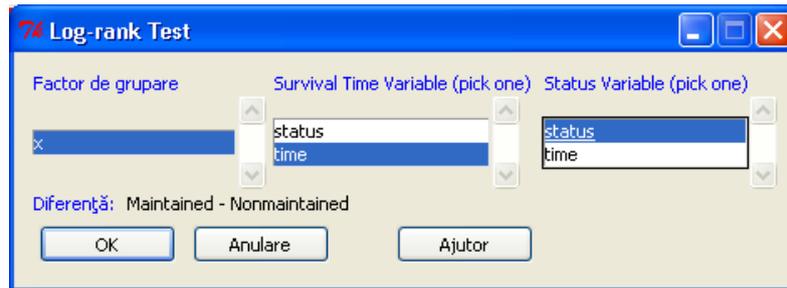

Figure 2. The Kaplan Meier graph dialog window, along with its result in Rcmdr using the survival analysis plug-in, for the aml data set from the survival package

Figure 3. The log-rank dialog window, along with its result in Rcmdr using the survival analysis plug-in, for the aml data set from the survival package
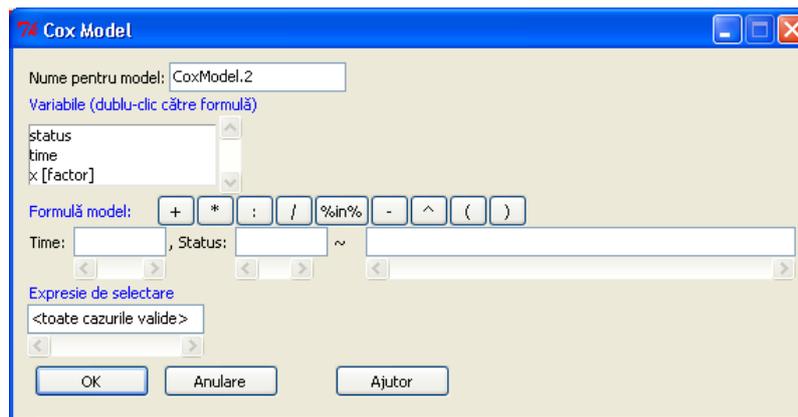

Figure 4. The Cox Model dialog window, in Rcmdr using the survival analysis plug-in, for the aml data set from the survival package

Having a Cox Model created different diagnostic tools become available in the SurvivalT menu. This way the Cox proportional hazard assumption test based on scaled Schoenfeld residuals correlated with a Kaplan Meier transformation of time is shown by using the command Cox PH assumption test (Schoenfeld/t) in the survival T menu having as result what is shown in Fig. 5.

Figure 5. The Cox proportional hazard assumption test based on scaled Schoenfeld residuals correlated with a Kaplan Meier transformation of time in Rcmdr using the survival analysis plug-in, for the kidney data set from the survival package

Other diagnostics for the Cox model are provided by the SurvivalT menu commands: Cox PH assumption test (Schoenfeld/t) – to assess graphically the Cox proportional hazard assumption, and Cox influence plot – to identify influential observations in the data set. The results of these two commands for a Cox model for the kidney data set from the survival package is shown in Figure 6.
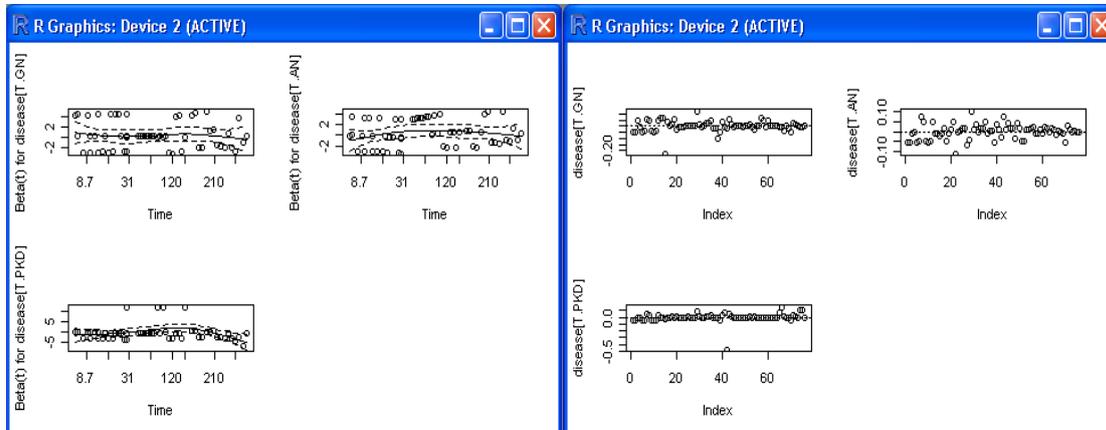

Figure 6. The Cox proportional hazard assumption graphical assessment based on scaled Schoenfeld residuals correlated with a Kaplan Meier transformation of time, and the influential observation graphical assessment results in Rcmdr using the survival analysis plug-in, for the kidney data set from the survival package

Discussion

The Rcmdr survival analysis plug-in was successfully built and it provides the functionality it was designed to offer: interface for Kaplan Meier and log log survival graph, interface for the log-rank test, interface to create a Cox proportional hazard regression model, interface commands to test and assess graphically the proportional hazard assumption, and influence observations.

This version provides basic survival analysis options. Future versions may add new functionalities, like interface for accelerated failure times models, robust survival analysis, handling truncation for survival data.

Conclusion

Rcmdr and R though their flexible and well planed structure, offer an easy way to expand their functionality that was used here to make the statistical environment more user friendly in respect with survival analysis.

References

1. Venables WN, Smith DM, et al. An Introduction to R: Notes on R: A Programming Environment for Data Analysis and Graphics. R Development Core Team 2006.
2. Fox J. The R Commander: A Basic-Statistics GUI for R [online]. [cited 2008 November]. Available from: URL: http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/
3. Sciviews-R [online], [cited 2008 November]. Available from: URL: http://www.sciviews.org/SciViews-R/index.html
4. JGR – Java GUI for R [online] [cited 2008 November]. Available from: URL: http://jgr.markushelbig.org/JGR.html
5. RKWard – An overview [online] [cited 2008 November]. Available from: URL: http://rkward.sourceforge.net/?content=overview

6. Tinn-R [online] [cited 2008 November]. Available from: URL: http://sourceforge.net/projects/tinn-r

7. Fox J. Cox Proportional-Hazards Regression for Survival Data. Appendix to An R and S-PLUS Compaion to Applied Regression. 2003.

8. Building R for Windows R [online] [cited 2008 November]. Available from: URL: http://www.murdoch-sutherland.com/Rtools/

9. R Development Core Team (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org

10. Fox J, with contributions from Ash M, Boye T, Calza S, Chang A, Grosjean P, et al. Rcmdr: R Commander. R package version 1.4-5. [online] 2007 [cited 2008 November] URL: http://socserv.socsci.mcmaster.ca/jfox/Misc/Rcmdr/

11. Miller RG. Survival Analysis. John Wiley & Sons 1997.

12. McGilchrist CA, Aisbett CW. Regression with frailty in survival analysis. Biometrics 1991;47:461–6.

13. S original by Therneau T and ported by Lumley T. Survival: Survival analysis, including penalised likelihood. R package version 2.34.