Study Guide

Survival Analysis STAT827 / SVA

2008

Prepared by: A/Prof Gillian Heller Department of Statistics Macquarie University
Edited by: Dr Petra Graham Department of Statistics Macquarie University

Contents

Welcome to Survival Analysis. This unit is a component of Macquarie University's postgraduate Applied Statistics program, and the program of the Biostatistics Collaboration of Australia (BCA). The Macquarie University unit code is STAT827, and the BCA's code is SVA. All aspects of the unit, including assessment, will be identical for BCA and MU students.

1 Contacts

The **unit coordinator** is
Dr Petra Graham
Phone: (02) 9850 6138
E-mail: pgraham@efs.mq.edu.au

General questions about the unit as well as questions about exercises and assignments should be addressed to Petra. Petra works Monday to Thursday and will also be overseeing the Blackboard CE6 (formerly known as WebCT) discussions.

The **postgraduate administrator** is
Lesley Mooney
Phone: (02) 9850 8550
e-mail: lmooney@efs.mq.edu.au

Enquiries about the receipt of course materials and assessments should be addressed to Lesley. Her working days are Tuesday, Wednesday and Thursday.

2 Overview

Survival analysis is concerned with the analysis of responses which are times to an event of interest. Such events can be, for example, recurrence of disease (such as a particular type of cancer); death; cessation of smoking; or weaning of breastfed infants. We seek to relate the time-to-event to covariates. In principle, we utilize the same tools that are applicable to other statistical analyses, such as multiple regression, but we have to take account of two issues that are common to time-to-event responses, viz. non-normality and censoring. The models that we study accommodate these features, and enable us to assess the effect of covariates in the usual statistical framework of estimation and hypothesis testing. We will be using the statistical software Stata throughout the unit.

3 Unit objectives

At the end of this unit you should be able to:

a. Summarise and display survival data using nonparametric methods;

b. Analyse survival data using the Cox proportional hazards model, including:

   i. hyopthesis testing;

   ii. diagnostic testing;

iii. the use of stratification and time-dependent variables, where appropriate;

c. Analyse survival data using parametric (accelerated failure-time) models;

d. Formulate and implement multiple-event models;

e. Determine sample size for a simple survival analysis;

f. Produce appropriate displays for publication.

The unit is divided into seven modules, each taking two weeks, except for the last one:

Module	Weeks	Content
1	1,2	The nature of survival data, including censoring; the survival (or survivorship) function: definition and estimation via the Kaplan-Meier curve; the stset command in Stata.
2	3,4	Kaplan-Meier estimate of the survival (or survivorship) function: confidence intervals and hypothesis testing; the Nelson-Aalen estimate of the cumulative hazard function; the density, survival, hazard and cumulative hazard functions.
3	5,6	Definition of the proportional hazards model; construction of the partial likelihood for the Cox model; the treatment of tied failure times; hypothesis testing on the coefficients, using Wald and partial likelihood ratio tests.
4	7,8	For the Cox PH model: hypothesis testing on the coefficients, contd; estimation of the baseline functions $S_0(t)$ and $H_0(t)$, and their adjustment for covariate values; the effect of a change in scale and origin of units of measurement of covariates.
5	9,10	Model diagnostics for the Cox PH model; the stratified Cox model
6	11,12	Time-dependent covariates in the Cox model; parametric survival time models, in particular the accelerated failure time model, with an exponential and Weibull distrubution; multiple-event models, including competing-risk models
7	13	Sample size determination for comparing two response rates and two survival distributions; good practice for the display of survival analysis results in scientific publications

Module objectives are given in Section A.

4 Assessment

Assessment for the unit consists of: three assignments, tutorial exercises, participation in online discussions and a written examination. The weighting is as follows:

Three assignments (22% each)	66%
Online participation	8%
Examination	26%

4.1 Assignments

Rules for assignment submission are given in the document BCA Assessment Guide, which is included with these notes and is also on Blackboard. While this has been written for BCA students, it applies equally to MU students. *Please read this document carefully.*

Plagiarism Please read the section on plagiarism in the BCA Assessment Guide.

Late submission Requests for an extension of the due date for an assignment must be made in advance of the due date for that assessment. These requests must be made directly to the unit coordinator by email. The unit coordinator will reply by email with the decision as to whether an extension has been granted and the new due date.
The penalty for late submission, where permission has not been granted, is as follows: 5% will be deducted for each day that an assignment is late, up to a maximum of 50%.

4.2 Exercises

There are exercises in the notes. While submission of the exercises is not required, these may provide a useful starting point for online Discussions. Participation in online discussions, as described below, is assessed.

4.3 Online participation

You will receive a mark for your Discussion contributions for each module. If you are not very forthcoming in the online discussions, you can make up for that by showing in your submitted work that you have been "listening" to the discussions, e.g. by providing critical comment or synthesis of points that have been made by others. Marks for online contributions will be awarded for quantity (i.e. "effort" e.g. making lots of online contributions), but mainly for quality, i.e. for succinct, focused presentation of points. When assessing the submitted exercises we will take off marks for work that appears to be simply borrowed from online contributions without proper attribution and discussion in your own words.

Instructors will generally let Discussions flow between the students in each group, except where key points seem to need resolution. The instructors will not usually make comments on the exercises required for submission, at least until after the deadline.

Any general Discussion items, in particular on the module notes, can be posted to the other Discussion areas, e.g. Stata, module 1, module 2, etc.

4.4 Examination

The examination will be released on Blackboard at 9am on Friday 13th June, and is due on Monday 16th June at 5pm. You should allocate about two free days to complete the examination. You should consider yourself under examination conditions, i.e. there is to be no communication between yourself and other students or any other person, in connection with the examination.

5 Timetable

We will be following the Macquarie University semester, with dates as follows:

Week	Starting date	Module	Assessment due
1	3 March	1	
2	10 March	1	
3	17 March	2	
4	25 March	2	Assignment 1
5	31 March	3	
6	7 April	3	
Mid semester break			
7	21 April	4	
8	28 April	4	Assignment 2
9	5 May	5	
10	12 May	5	
11	19 May	6	
12	26 May	6	Assignment 3
13	2 June	7	

Assessment is due on the Monday of the indicated weeks. The examination (see Section 4.4) will be held from 13 to 16 June.

6 Method of delivery and communication

The unit is offered in distance mode. Our means of communication will be via printed notes which will be mailed out, e-mail, and Blackboard.

The unit relies heavily on the prescribed text Hosmer and Lemeshow (see below). The study notes provide a guide to readings in this text, as well as sometimes to other readings, which will be provided. They also provide additional explanation where this is needed. In the study notes for each module, tutorial exercises are given, mostly referring to exercises in Hosmer and Lemeshow. Some of these are to be handed in for assessment (see above).

Study notes will be mailed to you, usually in blocks of two modules at a time. These will also be posted on the Blackboard site. You will be alerted by email when material has been mailed out, so if you do not receive these within a few days then please contact Lesley Mooney.

Blackboard

We will be using Blackboard Campus Edition 6 (CE6) for online discussions, posting of course notes, assignments, solutions and data sets, and submission of assignments. The link is
`http://learn.mq.edu.au/`
which you should bookmark.

BCA students: Note that we are using the Macquarie University Blackboard CE6 site, not the BCA E-learning site which you are using in other BCA units.

As Blackboard is the primary medium for communication in this unit, we expect that students will access the site at least every second day during semester.

- *Macquarie students:* Your Blackboard login is the username and password which were supplied to you by the University on enrolment. (This includes BCA students who are enrolled through Macquarie University.)
- *BCA students:* Your Blackboard login name is your first initial, your surname (all lowercase) and the digit 1. For example, David Hosmer would have login name dhosmer1. Your password is your BCA userid.

Student homepages

In order to introduce ourselves to each other, we will create homepages on Blackboard, with profiles of the students, unit coordinator and teaching assistants. The Homepages page is accessed from the main page of the unit in Blackboard. You will see that you have access to edit your own home page. You are encouraged to include at least the following:

- a photo (add this as a banner image, and upload the picture file from your computer);
- your preferred name;
- the degree program you are studying;
- your current job role, and the organisation for which you work;
- what you hope to gain from the unit;

plus any other information which you think would be interesting to your fellow students. You may include your e-mail address if you want to be contactable directly by other students. Please would you do this by the end of the first week of semester.

7 Textbooks

The prescribed text is

Hosmer DW and Lemeshow S (1999). *Applied Survival Analysis*, Wiley Interscience.

There are numerous texts on survival analysis which you may wish to consult, but the following may be particularly helpful because of its use of Stata:

Cleves MA, Gould WW and Gutierrez RG (2004). *An Introduction to Survival Analysis using Stata*, Revised Edition, Stata Press.

Other useful texts are:

- Klein JP and Moeschberger ML (1997). *Survival analysis : techniques for censored and truncated data*, Springer.
- Kleinbaum DG (1995). *Survival analysis : a self-learning text*, Springer-Verlag.

8 Software

We will be using Stata (version 9). [1] The software is available in the computer laboratory E4B 202 on Macquarie University campus. If you need your own copy of Stata, you will need to purchase it directly from the suppliers. You can place your order via the Survey Design website at:
`www.survey-design.com.au`

Prices

- Advanced GradPlan Intercooled Stata 9 with *perpetual licence* (consisting of CD for Windows, Mac or Linux, and a Getting Started Manual) = AU$231 (+ postage)
- Basic GradPlan Intercooled Stata 9 - as above but with a one-year licence = AU$143 (+ postage)
- There is also a Small Stata option; however, this is limited to around 1,000 observations which may not be sufficient.

Arrangements

Go to the Survey Design website listed above and download a copy of the order form (LH column 'Order Form' link). You should order a **GradPlan Intercooled** package. On the order form you should note that you are enrolled in a BCA course (BCA students) or STAT827 (MU students), your student ID number, the university in which you are enrolled, and the operating system of the computer that you will be using.

A note about versions 7, 8 and 9

BCA study and bridging notes are written in standard Stata code that is common to version 7 and later, with the exception of methods for producing graphs. A major change from version 8 onwards is that all commands are available via drop-down menus.

[1] If you have version 7 or 8, there is no need to upgrade for the purposes of this unit.

Note from suppliers:

Commands and output obtained through the GUI in Stata 8 or 9 are identical to those obtained from a command line in (BCA) study notes for version 7. There is still a Do editor, and we would always advocate the use of a sequence of command lines for serious analysis.

9 Complaints policy

BCA students: please see the BCA complaints policy in the Assessment Guide.

A Module objectives

A.1 Module 1

To understand:

a. the nature of survival data, including censoring;

b. the survival (or survivorship) function: definition and estimation.

To be able to compute and interpret:

a. the `stset` command in Stata;

b. the Kaplan-Meier curve.

A.2 Module 2

To understand:

a. the Kaplan-Meier estimate of the survival (or survivorship) function: confidence intervals and hypothesis testing;

b. the Nelson-Aalen estimate of the cumulative hazard function;

c. the relationship between the density, survival, hazard and cumulative hazard functions, and their derivation.

To be able to compute and interpret:

a. the Kaplan-Meier curve; logrank and Wilcoxon tests;

b. the Nelson-Aalen estimate of the cumulative hazard function;

c. the density, survival, hazard and cumulative hazard functions over a grid of parameter values.

A.3 Module 3

To understand:

a. the definition of the proportional hazards model;

b. construction of the partial likelihood for the Cox model;

c. the treatment of tied failure times;

d. hypothesis testing on the coefficients, using Wald and partial likelihood ratio tests.

To be able to compute and interpret:

a. the Cox regression model with both categorical and continuous predictors;

b. Estimates using the different methods for treating tied times;

c. Hypothesis tests on the βs, individually and jointly, using both the Wald and partial likelihood ratio tests.

A.4 Module 4

To understand, for the Cox proportional hazards model:

a. hypothesis testing on the coefficients, contd;

b. estimation of the baseline functions $S_0(t)$ and $H_0(t)$, and their adjustment for covariate values;

c. the effect of a change in scale and origin of units of measurement of covariates.

To be able to compute and interpret:

a. Hypothesis tests on the βs, individually and jointly, using both the Wald and partial likelihood ratio tests;

b. $\widehat{S}_0(t)$ and $\widehat{H}_0(t)$, adjusted for covariate values, plotted against time;

c. hazard ratios and confidence intervals for scaled covariates.

A.5 Module 5

To understand:

a. model diagnostics for the Cox PH model;

b. the stratified Cox model.

To be able to compute and interpret:

a. martingale residuals for the correct functional form of a covariate;

b. categorical forms of continuous covariates;

c. log-log survival plots;

d. Schoenfeld residuals;

e. Cox-Snell residuals for the overall goodness of fit;

f. dfbeta residuals for identification of influential observations;

g. the stratified Cox model.

A.6 Module 6

To understand:

a. Time-dependent covariates in the Cox model;

b. Parametric survival time models, in particular the accelerated failure time model, with an exponential and Weibull distribution;

c. Multiple-event models, including competing-risk models.

To be able to compute and interpret:

a. Cox regression with time-dependent covariates;

b. Coefficients and hazard ratios from exponential and Weibull regressions;

c. Models for data with multiple outcomes.

A.7 Module 7

To understand:

a. Sample size determination for comparing two response rates and two survival distributions;

b. Good practice for the display of survival analysis results in scientific publications.

To be able to compute and interpret:

a. Sample size calculations;

b. Survival plots in a variety of forms.