Survival Analysis for Epidemiologic and Medical Research

A Practical Guide

This practical guide to the analysis of survival data written for readers with a minimal background in statistics explains why the analytic methods work and describes how to effectively analyze and interpret epidemiologic and medical survival data with the help of modern computer systems.

This text contains a variety of statistical methods that not only are key elements of survival analysis but also are central to statistical analysis in general. Techniques such as statistical tests, transformations, confidence intervals, and analytic modeling are discussed in the context of survival data but are, in fact, statistical tools that apply to many kinds of data. Similarly, discussions of such statistical concepts as bias, confounding, independence, and interaction are presented and also are basic to a broad range of applications. These topics make up essentially a second-year, one-semester biostatistics course in survival analysis concepts and techniques for nonstatisticians.

Steve Selvin is Professor of Biostatistics and Epidemiology at the University of California, Berkeley. He has taught on the Berkeley campus for 35 years and has authored or co-authored more than 200 scientific articles in the areas of applied statistics and epidemiology. He has received two university teaching awards and is a member of the ASPH/Pfizer Public Health Academy of Distinguished Teachers.

Practical Guides to Biostatistics and Epidemiology

This is a series of short and practical but authoritative books for biomedical researchers, clinical investigators, public health researchers, epidemiologists, and nonacademic and consulting biostatisticians who work with data from biomedical, epidemiologic, and genetic studies. Some books are explorations of a modern statistical method and its application, others focus on a particular disease or condition and the statistical techniques most commonly used in studying it.

This series is for people who use statistics to answer specific research questions. The books explain the application of techniques, specifically the use of computational tools, and emphasize the interpretation of results, not the underlying mathematical and statistical theory.

Published in the series:
Applied Multilevel Analysis, by **Jos W. R. Twisk**
Secondary Data Sources for Public Health, by **Sarah Boslaugh**

Survival Analysis for Epidemiologic and Medical Research

A Practical Guide

Steve Selvin
University of California, Berkeley

CAMBRIDGE UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment, a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9780521895194

First published 2008

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging-in-Publication data
Selvin, S.
Survival analysis for epidemiologic and medical research : a practical guide / Steve Selvin
p. ; cm. – (Practical guides to biostatistics and epidemiology)
Includes bibliographical references and index.
ISBN 978-0-521-89519-4 (hardback) – ISBN 978-0-521-71937-7 (pbk.)
1. Medicine – Research – Statistical methods. 2. Epidemiology – Research – Statistical methods. 3. Survival analysis (Biometry) I. Title. II. Series.
[DNLM: 1. Survival Analysis. 2. Models, Statistical. WA 950 S469S 2008]
R853.S7S453 2008
610.72′7 – dc22 2007037910

ISBN 978-0-521-89519-4 Hardback
ISBN 978-0-521-71937-7 Paperback

For Liz and David

Contents

Overview

The description of survival analysis techniques can be mathematically complex. The primary goal of the following description, however, is a sophisticated introduction to survival analysis theory and practice using only elementary mathematics, with an emphasis on examples and intuitive explanations. The mathematical level is completely accessible with knowledge of high school algebra, a tiny bit of calculus, and a one-year course in basic statistical methods (for example, *t*-tests, chi-square analysis, correlation, and some experience with linear regression models). With this minimal background, the reader will be able to appreciate why the analytic methods work and, with the help of modern computer systems, to effectively analyze and interpret much of epidemiologic and medical survival data.

A secondary goal is the introduction (perhaps the review) of a variety of statistical methods that are key elements of survival analysis but are also central to statistical data analysis in general. Such techniques as statistical tests, transformations, confidence intervals, analytic modeling, and likelihood methods are presented in the context of survival data but, in fact, are statistical tools that apply to many kinds of data. Similarly, discussions of such statistical concepts as bias, confounding, independence, and interaction are presented in the context of survival analysis but also are basic to a broad range of applications.

To achieve these two goals, the presented material is divided into nine topics:

Chapter 1: Rates and their properties
Chapter 2: Life tables
Chapter 3: Two especially useful estimation tools
Chapter 4: Product-limit estimation

Chapter 5: Exponential survival time probability distribution
Chapter 6: Weibull survival time probability distribution
Chapter 7: Analysis of two-sample survival data
Chapter 8: General hazards model: parametric
Chapter 9: General hazards model: nonparametric

These topics make up essentially a second-year, one-semester biostatistics course. In fact, this course has been taught at the University of California, Berkeley as part of the biostatistics/epidemiology master of public heath degree major, at the Graduate Summer Institute of Epidemiology and Biostatistics at Johns Hopkins Bloomberg School of Public Health, and at the Graduate Summer Session in Epidemiology at the University of Michigan.

All statistical methods are extensively illustrated with both analytic and graphical examples from the San Francisco Men's Health Study. This unique study was established in 1983 to conduct a population-based prospective investigation of the epidemiology and natural history of the newly emerging disease Acquired Immunodeficiency Syndrome (AIDS). The collected data are a source of valuable and comprehensive information about the AIDS epidemic in its earliest years. These data illustrate realistically the discussed statistical techniques. A "workbook" of noncomputer problems is included to further explore the practical side of survival analysis methods. Finally, a small amount of computer code gives a sense of survival analysis software. The statistical analysis system called "R" is chosen because it is extensive and fully documented and both the software and documentation can be obtained without cost (http://www.r-project.org).

Clearly many kinds of phenomena fail. Data collected to study the failure of equipment, machine components, numerous kinds of products, and the structural integrity of various materials are frequently analyzed with survival analysis techniques (sometimes called time-to-failure data and methods). For the following description of survival analysis, however, the terminology is by and large in terms of human mortality (survived/died). For example, rates are described in terms of mortality risk (risk of death). The language of human mortality is chosen strictly for simplicity. The theory and applications of the methods discussed are essentially the same regardless of the subject matter context. Using general terminology complicates explanations and is avoided to clearly focus on the statistical issues important in the analysis of epidemiologic and medical survival data.

It has been remarked (by Churchill Eisenhert) that the practical power of a statistical procedure is the statistical power multiplied by the probability that the procedure will be used. The material in this text has some of this same spirit. A number of analytic approaches are presented because they are simple rather than optimally efficient. For example, simple stratification procedures are suggested for estimation, exploring linearity of a variable, identifying the source of interactions, and assessing the proportionality of hazard functions. Also in the spirit of simplicity, all confidence intervals are set at the 95% level because other levels of significance are rarely used.

Steve Selvin, 2007