Jayant V. Deshpande
Sudha G. Purohit

Second Edition

Series on Quality, Reliability and Engineering Statistics Vol. 16

Lifetime Data: Statistical Models and Methods

World Scientific

Series on Quality, Reliability and Engineering Statistics Vol. 16

Lifetime Data: Statistical Models and Methods

Second Edition

Jayant V. Deshpande & Sudha G. Purohit
University of Pune, India

World Scientific

NEW JERSEY · LONDON · SINGAPORE · BEIJING · SHANGHAI · HONG KONG · TAIPEI · CHENNAI · TOKYO

Published by

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data
Names: Deshpande, J. V., author. | Purohit, Sudha G., author.
Title: Lifetime data : statistical models and methods /
by Jayant V. Deshpande, Sudha G. Purohit.
Description: 2nd edition. | New Jersey : World Scientific, 2016. |
Series: Series on quality, reliability, and engineering statistics ; vol. 16 |
Includes bibliographical references and index.
Identifiers: LCCN 2015031961 | ISBN 9789814730662 (hardcover : alk. paper)
Subjects: LCSH: Failure time data analysis. | Survival analysis (Biometry) |
Reliability (Engineering)--Statistical methods.
Classification: LCC QA276 .D446 2016 | DDC 519.5/46--dc23
LC record available at http://lccn.loc.gov/2015031961

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Printed in Singapore

Lifetime Data: Statistical Models and Methods

SERIES ON QUALITY, RELIABILITY AND ENGINEERING STATISTICS

Published

Vol. 3: Contributions to Hardware and Software Reliability
P. K. Kapur, R. B. Garg and S. Kumar

Vol. 4: Frontiers in Reliability
A. P. Basu, S. K. Basu and S. Mukhopadhyay

Vol. 5: System and Bayesian Reliability
Y. Hayakawa, T. Irony and M. Xie

Vol. 6: Multi-State System Reliability
Assessment, Optimization and Applications
A. Lisnianski and G. Levitin

Vol. 7: Mathematical and Statistical Methods in Reliability
B. H. Lindqvist and K. A. Doksum

Vol. 8: Response Modeling Methodology: Empirical Modeling for Engineering and Science
H. Shore

Vol. 9: Reliability Modeling, Analysis and Optimization
H. Pham

Vol. 10: Modern Statistical and Mathematical Methods in Reliability
A. Wilson, S. Keller-McNulty, Y. Armijo and N. Limnios

Vol. 11: Life-Time Data: Statistical Models and Methods
J. V. Deshpande and S. G. Purohit

Vol. 12: Encyclopedia and Handbook of Process Capability Indices:
A Comprehensive Exposition of Quality Control Measures
W. L. Pearn and S. Kotz

Vol. 13: An Introduction to the Basics of Reliability and Risk Analysis
E. Zio

Vol. 14: Computational Methods for Reliability and Risk Analysis
E. Zio

Vol. 15: Basics of Reliability and Risk Analysis: Worked Out Problems and Solutions
E. Zio, P. Baraldi and F. Cadini

Vol. 16: Lifetime Data: Statistical Models and Methods (2nd Edition)
J. V. Deshpande and S. G. Purohit

The complete list of the published volumes in the series can be found at
http://www.worldscientific.com/series/sqres

Preface to the Second Edition

The first edition of the book appeared in 2005. The copies are not easily available now. The authors and the publisher are happy to bring out a revised edition.

We have taken this opportunity to correct several typographical errors and also expand the material in various chapters. In particular, Chapter 8, which discussed only the proportional hazards model, now also includes the 'Frailty Model' as well as the 'Accelerated Life Time Model'. It has now been renamed.

This book has been used as a textbook at the Master's level at many institutions. The teachers have communicated typographical errors which they encountered. We are grateful to all of them and particularly to Prof. Isha Dewan of the Indian Statistical Institute for this curtsey.

Preface to the First Edition

The last 50 years have seen a surge in the development of statistical models and methodology for data consisting of lifetimes. This book presents a selection from this area in a coherent form suitable for teaching postgraduate students. In particular, the background and needs of students in India have been kept in mind.

The students are expected to have adequate mastery over calculus and introductory probability theory, including the classical laws of large numbers and central limit theorems. They are also expected to have undergone a basic course in statistical inference. Certain specialized concepts and results such as U-statistics limit theorems are explained in this book itself. Further concepts and results, e.g., weak convergence of processes and martingale central limit theorem, are alluded to and exploited at a few places, but are not considered in depth.

We illustrate the use of many of these methods through the commands of software R. The choice of R was made because it is in public domain and also because the successive commands bring out the stages in the statistical computations. It is hoped that users of statistics will be able to choose methods appropriate for their needs, based on the discussions in this book, and will be able to apply them to real problems and data with the help of the R-commands.

We have taught courses based on this material at the University of Pune and elsewhere. It is our experience that most of this material can be taught in a one-semester course (about 45–50 one-hour lectures over 15/16 weeks). Lecture notes prepared by the authors for this course have been in circulation at Pune and elsewhere for several years. Inputs from colleagues and successive batches of students have been useful in finalizing this book. We are grateful to all of them. We also record our appreciation of the support received from our families, friends and all the members of the Department of Statistics, University of Pune.

Contents

Chapter 1

Introduction

It is universally recognized that lifetimes of individuals, components, systems, etc. are unpredictable and random, and hence amenable only to probabilistic and statistical laws. The development of models and methods to deal with such random variables took place in the second half of the 20-th century, although certain explicit and implicit results are from earlier times as well. The development proceeded in two main intermingling streams. The reliability theory stream is concerned with models for lifetimes of components and systems, in the engineering and industrial fields. The survival analysis stream mainly drew inspiration from medical and similar biological phenomena. In this book we bring the two streams together. Our aim is to emphasise the basic unity of the subject and yet to develop it in its diversity.

In all the diverse application, the random variable of interest is the time upto the occurrence of the specified event often called "death", "failure", "break down" etc. It is called the lifetime of the concerned unit. However, there are situations where the technical term "time" does not represent time in the literal sense. For example, it could be the number of operations a component performs before it breaks down. It could even be the amount that a health insurance company pays in a particular case.

Examples of failure or lifetime situations:

(1) A mechanical engineer conducts a fatigue test to determine the expected life of rods made of steel by subjecting n specimens to an axial load that causes a specified stress. The number of cycles are recorded at the time of failure of every specimen.

(2) A manufacturer of end mill cutters introduces a new ceramic cutter material. In order to estimate the expected life of a cutter, the manu-

facturer places n units under test and monitors the tool wear. A failure of the cutter occurs when the wear-out exceeds a predetermined value. Because of the budgeting constraints, the manufacturer runs the test for a month.

(3) A 72-hr. test was carried out on 25 gizmos, resulting in r_1 failure times (in hrs.). Of the remaining working gizmos on test r_2 were removed before the end of test duration (72 hrs.) to satisfy customer demands. The rest were still working at the end of the 72-hr. test.

(4) *Leukemia patients*: Leukemia is the cancer of blood and as in any other type of cancer, there are remission periods. In a remission period, the patient though not free of disease is free of symptoms. The length of the remission period is a variable of interest in this study. The patients in the state of remission are followed over time to see how long they stay in remission.

(5) *A prospective study of heart condition.* A disease-free cohort of individuals is followed over several years to see who develops heart disease and when does it happen.

(6) *Recidivism study*: A recidivist is a person who relapses into crime. In this study, newly released parolees are followed in time to see whether and when they get rearrested for another crime.

(7) *Spring testing*: Springs are tested under cycles of repeated loading and failure time is the number of cycles leading to failure. Samples of springs are allocated to different stress levels to study the relationship between the lifetimes at different stress levels. At the lower stress levels failure times could be longer than at higher stress levels.

Measurement of Survival Time (or Failure Time): The following points should be kept in mind while measuring the survival time. The time origin should be precisely defined for each individual. The individuals under study should be as similar as possible at their time origin. The time origin need *not* be and usually is *not* the same calender time for each individual. Most clinical trials have staggered entries, so that patients enter the study over a period of time. The survival time of a patient is measured from his/her own date of entry. Figures 1.1 and 1.2 show staggered entries and how these are aligned to have a common origin.

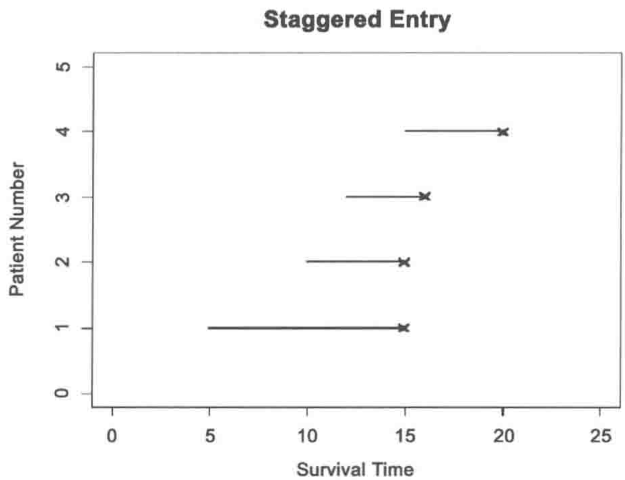

Figure 1.1

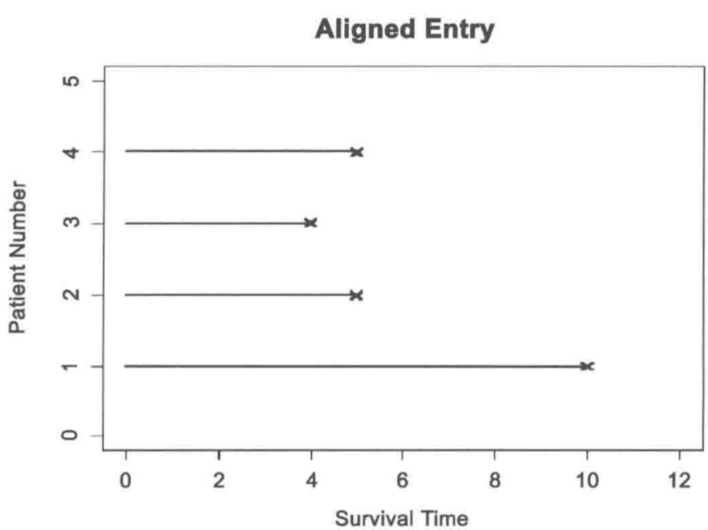

Figure 1.2

The concept of the point event of failure should be defined precisely. If a light bulb, for example, is operating continuously, then the number of hours for which it burned should be used as the lifetime. If the light bulb is turned on and off, as most are, the meaning of number of hours burned will be different as the shocks of lighting and putting off decreases the light bulb's life. This example indicates that there may be more to defining a lifetime than just the amount of time spent under operation.

Censoring: The techniques for reducing experimental time are known as censoring. In survival analysis the observations are lifetimes which can be indefinitely long. So quite often the experiment is so designed that the time required for collecting the data is reduced to manageable levels.

Two types of censoring are built into the design of the experiment to reduce the time taken for completing the study.

Type I (Time Censoring): A number (say n) of identical items are simultaneously put into operation. However, the study is discontinued at a predetermined time t_0. Suppose n_u items have failed by this time and the remaining $n_c = n - n_u$ items remain operative. These are called the censored items. Therefore the data consists of the lifetimes of the n_u failed items and the censoring time t_0 for the remaining n_c items (see Figure 1.3).

Example of type I censoring

Power supplies are major units for most electronic products. Suppose a manufacturer conducts a reliability test in which 15 power supplies are operated over the same duration. The manufacturer decides to terminate the test after 80000 hrs. Suppose 10 power supplies fail during the fixed time interval. Then the remaining five are type I censored.

Type II (Order Censoring): Again a number (say n) of identical components are simultaneously put into operation. The study is discontinued when a predetermined number $k(< n)$ of the items fail. Hence the failure times of the k failed items are available. These are the k smallest order statistics of the complete random sample. For the remaining items the censoring time $x_{(k)}$, which is the failure time of the item failing last, is available (see Figure 1.4).

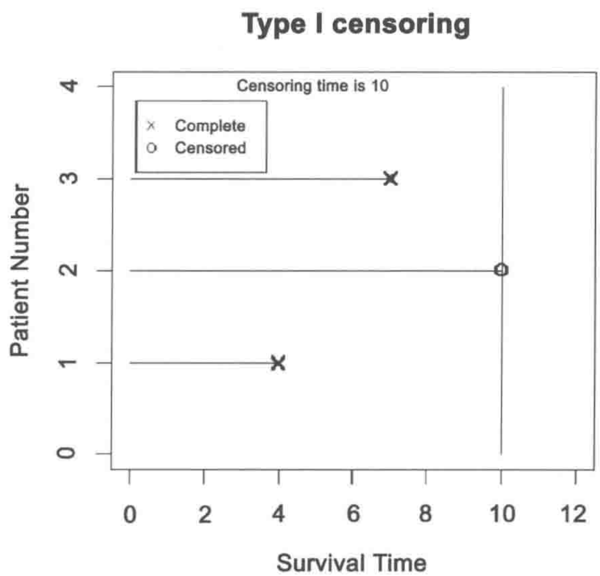

Figure 1.3

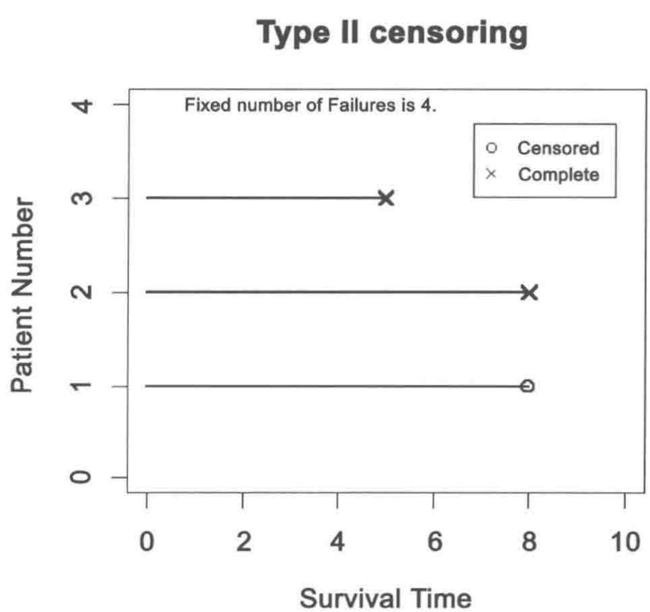

Figure 1.4

Example of Type II censoring

Twelve ceramic capacitors are subjected to a life test. In order to reduce the test time, the test is terminated after eight capacitors fail. The remaining are type II censored.

The above types of censoring are more prevalent in reliability studies (of engineering systems). In survival studies (of biomedical items) censoring is more a part of the experimental situation rather than a matter of deliberate design.

Undesigned censoring occurs when some information about individual survival time is available but exact survival time is *not* known. As a simple example of such undesigned censoring, consider leukemia patients who are followed until they go out of remission. If for a given patient, the study ends while the patient is still in remission (that is the event defining failure does not occur), then the patient's survival time is considered as censored. For this person, it is known that the survival time is not less than the period for which the person was observed. However, the complete survival time is *not* known.

The most frequent type of censoring is known as *right random censoring*. It occurs when the complete lifetimes are not observed for reasons which are beyond the control of the experimenter. For example, it may occur in any of the following situations: (i) loss to follow-up; the patient may decide to move elsewhere and therefore the experimenter may not see him/her again, (ii) withdrawal from the study; the therapy may have bad side effects so it may become necessary to discontinue the treatment or the patient may become non-cooperative, (iii) termination of the study; a person does not experience the event before the study ends, (iv) the value yielded by the unit under study may be outside the range of the measuring instrument, etc. Figure 1.5 illustrates a possible trial in which random censoring occurs. In this figure, patient 1 entered the study at $t = 0$ and died at $T = 5$, giving an uncensored observation. Patient 2 also entered the study at $t = 0$ and was still alive by the end of study, thus, giving a censored observation. Patient 3 has entered the study at $t = 0$ but did not follow up before the end of study to give another censored observation.

Example of Random (right) Censoring

A mining company owns a 1,400 car fleet of 80-ton high-side, rotary-dump gondolas. A car will accumulate about 100,000 miles per year. In their travels from mines to a power plant, the cars are subjected to vibrations due to track input in addition to the dynamic effects of the longitudinal shocks coming through the couplers. As a consequence, the cou-