

Cambridge University Press

0521846285 - Introduction to Computer-Intensive Methods of Data Analysis in Biology

Derek A. Roff

Frontmatter

[More information](#)

Introduction to Computer-Intensive Methods of Data Analysis in Biology

This guide to the contemporary toolbox of methods for data analysis will serve graduate students and researchers across the biological sciences. Modern computational tools, such as Bootstrap, Monte Carlo and Bayesian methods, mean that data analysis no longer depends on elaborate assumptions designed to make analytical approaches tractable. These new ‘computer-intensive’ methods are currently not consistently available in statistical software packages and often require more detailed instructions. The purpose of this book therefore is to introduce some of the most common of these methods by providing a relatively simple description of the techniques. Examples of their application are provided throughout, using real data taken from a wide range of biological research. A series of software instructions for the statistical software package S-PLUS are provided along with problems and solutions for each chapter.

DEREK A. ROFF is a Professor in the Department of Biology at the University of California, Riverside.

Cambridge University Press

0521846285 - Introduction to Computer-Intensive Methods of Data Analysis in Biology

Derek A. Roff

Frontmatter

[More information](#)

Introduction to Computer- Intensive Methods of Data Analysis in Biology

DEREK A. ROFF

Department of Biology

University of California



Cambridge University Press
 0521846285 - Introduction to Computer-Intensive Methods of Data Analysis in Biology
 Derek A. Roff
 Frontmatter
[More information](#)

CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
 The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9780521846288

© Cambridge University Press 2006

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2006

Printed in the United Kingdom at the University Press, Cambridge

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

Library of Congress Cataloging in Publication data

Roff, Derek A., 1949–

Introduction to computer-intensive methods of data analysis in biology / Derek A. Roff.
 p. cm.

Includes bibliographical references.

ISBN-13: 978-0-521-84628-8 (hardback)

ISBN-13: 978-0-521-60865-7 (pbk.)

ISBN-10: 0-521-84628-5 (hardback)

ISBN-10: 0-521-60865-1 (pbk.)

1. Biology—Data processing. I. Title.

QH324.2 R62 2006

570.285—dc22

2006001857

ISBN-13: 978-0-521-84628-8 hardback

ISBN-10: 0-521-84628-5 hardback

ISBN-13: 978-0-521-60865-7 paperback

ISBN-10: 0-521-60865-1 paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

Preface vii

1 An Introduction to Computer-intensive Methods 1

2 Maximum Likelihood 9

3 The Jackknife 42

4 The Bootstrap 66

5 Randomization and Monte Carlo Methods 102

6 Regression Methods 157

7 Bayesian Methods 204

References 233

**Appendix A – An Overview of S-PLUS Methods Used in
this Book 242**

**Appendix B – Brief Description of S-PLUS Subroutines Used
in this Book 249**

Appendix C – S-PLUS Codes Cited in Text 253

Appendix D – Solutions to Exercises 316

Index 365

Preface

Easy access to computers has created a revolution in the analysis of biological data. Prior to this easy access even “simple” analyses, such as one-way analysis of variance, were very time-consuming. On the other hand, statistical theory became increasingly sophisticated and far outstripped the typical computational means available. The advent of computers, particularly the personal computer, and statistical software packages, changed this and made such approaches generally available.

Much of the development of statistical tools has been premised on a set of assumptions, designed to make the analytical approaches tractable (e.g., the assumption of normality, which underlies most parametric methods). We have now entered an era where we can, in many instances, dispense with such assumptions and use statistical approaches that are rigorous but largely freed from the straight-jacket imposed by the relative simplicity of analytical solution. Such techniques are generally termed “computer-intensive” methods, because they generally require extensive numerical approaches, practical only with a computer. At present, these methods are rather spottily available in statistical software packages and very frequently require more than simple “point and click” instructions. The purpose of the present book is to introduce some of the more common methods of computer-intensive methods by providing a relatively simple mathematical description of the techniques, examples from biology of their application, and a series of software instructions for one particular statistical software package (S-PLUS). I have assumed that the reader has at least an introductory course in statistics and is familiar with techniques such as analysis of variance, linear and multiple regression, and the χ^2 test. To relieve one of the task of typing in the coding provided in an appendix to this book, I have also made it available on the web at <http://www.biology.ucr.edu/people/faculty/Roff.html>.