



A First Guide to Statistical Computations in R

Torben Martinussen, Ib Michael Skovgaard and Helle Sørensen

Biofolia

A First Guide to Statistical Computations in R

Torben Martinussen
Ib M. Skovgaard
Helle Sørensen

Torben Martinussen, Ib Michael Skovgaard and Helle Sørensen

A First Guide to Statistical Computations in R

1. edition 2012, Biofolia

© Biofolia

Cover: Sofie Meedom

Typeset: Torben Martinussen, Ib Michael Skovgaard and Helle Sørensen

Print: Narayana Press, Gylling, Denmark

Trykt bog ISBN 978-87-913-1956-3

E-bog ISBN 978-87-913-1998-3

Published by:

Samfundslitteratur

Rosenoerns Allé 9 DK-1970 Frederiksberg C

Denmark

Tlf: + 45 38 15 38 80

Fax: + 45 35 35 78 22

slforlagene@samfundslitteratur.dk

www.biofolia.dk

All rights reserved.

No parts of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage or retrieval system, without permission in writing from the publisher.

Preface

R is a statistical computer program used and developed by statisticians around the world. It is probably the leading statistical program, at least among statisticians, and it is freely available [R Development Core Team, 2010]. You can read more about the program on the R home page www.r-project.org.

This guide to R is intended for the newcomer who wants to do statistical analysis with R, and needs a guide to get started and a reference for common data handling and statistical analysis. The guide is in no way complete, but you should be able to do quite a bit based on this guide alone. Once acquainted with R you will surely need information from the numerous other sources, many of which are freely available and easily found from the R pages. Notice that R can also be used as a general purpose programming language, but this is outside the scope of this guide.

The selection of material reflects what is needed for our teaching in statistics for students of various disciplines, mostly biological, at the University of Copenhagen.

The guide is divided into two parts, the first part on R basics, and the second part on statistical analyses using R. In the second part we use various datasets for illustration. They are all available in the R package `Guide1data`, see Appendix C for installation of the package. There is also a supporting web site

<http://www.r-guide.dk>

where you among other things find most of the R-code used in the book.

To get started you should not read from one end to the other, but first install R (see Appendix A), then read the first few sections in the first chapter while making sure that you can redo what is shown, and as soon as possible start working on your own data, reading sections as needed. We have tried to make the purpose of each chapter and section clear so that it is easier for you to select what to read. This is done explicitly in Part I with an introductory remark to each section or subsection (in *italic*), whereas it is more obvious in Part II.

The way you work with R does not depend much on the operating system, and we have chosen the few such instances to describe how it works in Windows in the hope that the analogue can be guessed by users of other platforms.

To improve the layout, the graphs in the book have been produced with changed graphical parameters, such as the size of plot symbols and labels, rather than with the default values. The exception to this is Section 5.2.1 where some of these options are described. Some graphs use colors, but as the book has been typeset in greyscale the colors are not shown in the book. The colored versions of these graphs can be found at the supporting web site. Moreover, the output from R has been edited a few places to make it more readable.