Preface

This book is the English adaptation of the second edition of the book *Statistiques avec R* which was published in 2008 and was a great success in the French-speaking world. In this version, a number of worked examples have been supplemented and new examples have been added. We hope that readers will enjoy using this book for reference when working with R.

This book is aimed at statisticians in the widest sense, that is to say, all those working with datasets: science students, biologists, economists, etc. All statistical studies depend on vast quantities of information, and computerised tools are therefore becoming more and more essential. There are currently a wide variety of software packages which meet these requirements. Here we have opted for R, which has the triple advantage of being free, comprehensive, and its use is booming. However, no prior experience of the software is required. This work aims to be accessible and useful for both novices and experts alike.

This book is organised into two main parts: Part I focuses on the R software and the way it works, and Part II on the implementation of traditional statistical methods with R. In order to render them as independent as possible, a brief chapter offers extra help getting started (Chapter 5, "A Quick Start with R") and acts as a transition: It will help those readers who are more interested in statistics than in software to be operational more quickly.

In the first chapter we present the basic elements of the software. The second chapter deals with data processing (reading data from file, merging factor levels, etc.), that is to say, common operations in statistics. As any statistical report depends on clear, concise visualisation of results, in Chapter 3 we detail the variety of possibilities available in this domain with R. We first present the construction of simple graphs with the various available options, and then detail the use of more complex graphs. Some programming basics are then presented in Chapter 4: We explain how to construct your own functions but also present some of the useful pre-defined functions to conduct repetitive analyses automatically. Focusing on the R software itself, this first main part enables the reader to understand the commands used in subsequent statistical studies.

The second part of the book offers solutions for dealing with a wide

range of traditional statistical data processing techniques: hypothesis tests (Chapter 6), regression methods (Chapter 7), analyses of variance and co-variance (Chapter 8), classification methods (Chapter 9), exploratory multi-variate analysis methods (Chapter 10), and clustering methods (Chapter 11). Each technique is preceded by a concrete example and is dealt with indepen-dently in its own specific worked example. Following a short presentation of the method, the R command lines are explicitly detailed and the results commented. Readers are able to download the datasets from the address `http://www.agrocampus-ouest.fr/math/RforStat`, where they can easily find all the results described in this work. R is a command line interface, but another easy-to-use graphical interface is available: At the end of each worked example there is a section called "Rcmdr Corner" which can be used to process the example using this second interface. This section also helps readers get to know the R commands better as, with each click on the drop-down menu, the lines of code corresponding to the chosen action are generated.