Preface

Linear models are central to the practice of statistics. They are part of the core knowledge expected of any applied statistician. Linear models are the foundation of a broad range of statistical methodologies; this book is a survey of techniques that grow from a linear model. Our starting point is the regression model with response y and predictors $x_1, \ldots x_p$. The model takes the form:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon$$

where ε is normally distributed. This book presents three extensions to this framework. The first generalizes the y part; the second, the ε part; and the third, the x part of the linear model.

Generalized Linear Models: The standard linear model cannot handle nonnormal responses, y, such as counts or proportions. This motivates the development of generalized linear models that can represent categorical, binary and other response types.

Mixed Effect Models: Some data has a grouped, nested or hierarchical structure. Repeated measures, longitudinal and multilevel data consist of several observations taken on the same individual or group. This induces a correlation structure in the error, ε. Mixed effect models allow the modeling of such data.

Nonparametric Regression Models: In the linear model, the predictors, x, are combined in a linear way to model the effect on the response. Sometimes this linearity is insufficient to capture the structure of the data and more flexibility is required. Methods such as additive models, trees and neural networks allow a more flexible regression modeling of the response that combine the predictors in a nonparametric manner.

This book aims to provide the reader with a well-stocked toolbox of statistical methodologies. A practicing statistician needs to be aware of and familiar with the basic use of a broad range of ideas and techniques. This book will be a success if the reader is able to recognize and get started on a wide range of problems. However, the breadth comes at the expense of some depth. Fortunately, there are book-length treatments of topics discussed in every chapter of this book, so the reader will know where to go next if needed.

R is a free software environment for statistical computing and graphics. It runs on a wide variety of platforms including the Windows, Linux and Macintosh operating systems. Although there are several excellent statistical packages, only R is both free and possesses the power to perform the analyses demonstrated in this book. While it is possible in principle to learn statistical methods from purely theoretical expositions, I believe most readers learn best from the demonstrated interplay of

theory and practice. The data analysis of real examples is woven into this book and all the R commands necessary to reproduce the analyses are provided.

Prerequisites: Readers should possess some knowledge of linear models. The first chapter provides a review of these models. This book can be viewed as a sequel to *Linear Models with R*, Faraway (2004). Even so there are plenty of other good books on linear models such as Draper and Smith (1998) or Weisberg (2005), that would provide ample grounding. Some knowledge of likelihood theory is also very useful. An outline is provided in Appendix A, but this may be insufficient for those who have never seen it before. A general knowledge of statistical theory is also expected concerning such topics as hypothesis tests or confidence intervals. Even so, the emphasis in this text is on application, so readers without much statistical theory can still learn something here.

This is not a book about learning R, but the reader will inevitably pick up the language by reading through the example data analyses. Readers completely new to R will benefit from studying an introductory book such as Dalgaard (2002) or one of the many tutorials available for free at the R website. Even so, the book should be intelligible to a reader without prior knowledge of R just by reading the text and output. R skills can be further developed by modifying the examples in this book, trying the exercises and studying the help pages for each command as needed. There is a large amount of detailed help on the commands available within the software and there is no point in duplicating that here. Please refer to Appendix B for details on obtaining and installing R along with the necessary add-on packages and data necessary for running the examples in this text. S-plus derives from the same S language as R, so many of the commands in this book will work. However, there are some differences in the syntax and the availability of add-on packages, so not everything here will work in S-plus.

The website for this book is at `www.stat.lsa.umich.edu/~faraway/ELM` where data described in this book appears. Updates and errata will also appear there.

Thanks to the builders of R without whom this book would not have been possible.

Contents