Preface

This is a book about linear models in Statistics. A linear model describes a quantitative response in terms of a linear combination of predictors. You can use a linear model to make predictions or explain the relationship between the response and the predictors. Linear models are very flexible and widely used in applications in physical science, engineering, social science and business. Linear models are part of the core of Statistics and understanding them well is crucial to a broader competence in the practice of statistics.

This is not an introductory textbook. You will need some basic prior knowledge of statistics as might be obtained in one or two courses at the university level. You will need to be familiar with essential ideas such as hypothesis testing, confidence intervals, likelihood and parameter estimation. You will also need to be competent in the mathematical methods of calculus and linear algebra. This is not a particularly theoretical book as I have preferred intuition over rigorous proof. Nevertheless, successful statistics requires an appreciation of the principles and it is my hope that the reader will absorb these through the many examples I present.

This book is written in three languages: English, Mathematics and R. My motivation in writing the first edition of this text was to combine these three seamlessly to allow coherent exposition of the practice of linear modeling. This was uncommon ten years ago but now this has become popular in the now large number of statistics books that integrate R. Clearly it is a method that works but it does require the reader to become somewhat fluent in R. This is not a book about learning R but like any foreign language, one becomes proficient by practicing it rather than by memorizing the dictionary. The reader is advised to look elsewhere for a basic introduction to R but should not hesitate to dive into this book and pick it up as you go. I shall try to help. See the appendix to get started.

The book website can be found at:

http://people.bath.ac.uk/jjf23/LMR

Ten years have passed since the first edition and the world of R has expanded enormously. Here I am taking the opportunity to fix some errors and make some improvements. This new edition makes several revisions:

1. I have reorganized the material on interpreting linear models to distinguish the main applications of prediction and explanation. Although these two share a common basis in theory, the modeling approach diverges, depending on the purpose of the analysis. In particular, I have introduced some elementary notions of causality into the chapter on explanation.

2. Several new topics have been added including the QR decomposition, splines, additive models, lasso, multiple imputation and false discovery rates. The chapter on statistical strategy from the first edition has gone but its contents have been redistributed throughout the revised edition.

3. I have made extensive use of the `ggplot2` graphics package in addition to base graphics. As with the first edition, I assume that the reader has a basic knowledge of R. Some of the R commands are quite complex. Sometimes I have explained the more difficult points but the reader will need to make some effort to fully understand these commands. This effort is a necessary part of the learning process.

My thanks to many past students and readers of the first edition whose comments and questions have helped me make many improvements to this edition. Thanks to the builders of R (R Core Team (2013)) who made all this possible.