

Preface to Second Edition

Analysis of variance and regression analysis have for many years been the mainstay of statistical modelling. These techniques usually have as a basic assumption that the residual or error terms are independently and identically distributed. Mixed models are an important new approach to modelling which allows us to relax the independence assumption and take into account more complicated data structures in a flexible way. Sometimes this interdependence of observations is modelled directly in a mixed model. For example, if a number of repeated measurements are made on a patient, mixed models allow us to specify a pattern for the correlation between these measurements. In other contexts, such as the cross-over clinical trial, specifying that patient effects are normally distributed, rather than fixed as in the classical approach, induces observations on the same patient to be correlated.

There are many benefits to be gained from using mixed models. In some situations the benefit will be an increase in the precision of our estimates. In others, we will be able to make wider inferences. We will sometimes be able to use a more appropriate model which will give us greater insight into what underpins the structure of the data. However, it is only the relatively recent availability of software in versatile packages such as SAS[®] which has made these techniques widely accessible. It is now important that suitable information on their use becomes available so that they may be applied confidently on a routine basis.

Our intention in this book is to put all types of mixed model into a general framework and to consider the practical implications of their use. We aim to do this at a level that can be understood by applied statisticians and numerate scientists. Greatest emphasis is placed on skills required for the application of mixed models and interpretation of the results. An in-depth understanding of the mathematical theory underlying mixed models is not essential to gain these, but an awareness of the practical consequences of fitting different types of mixed models is necessary. While many publications are available on various aspects of mixed models, these generally relate to specific types of model and often differ in their use of terminology. Such publications are not always readily comprehensible

to the applied statisticians who will be the most frequent users of the methods. An objective of this book is to help overcome this deficit.

Examples given will primarily relate to the medical field. However, the general concepts of mixed models apply equally to many other areas of application, e.g. social sciences, agriculture, official statistics. (In the social sciences mixed models are often referred to as 'multi-level' models.) Data are becoming easier to collect with the consequence that datasets are now often large and complex. We believe that mixed models provide useful tools for modelling the complex structures that occur in such data.

The second edition of this book retains the same structure as the first edition, but there are substantial changes to reflect the advances in software. All of the examples from the first edition have been reworked in Version 9 of SAS using, where relevant, the newly available PROC GLIMMIX for fitting generalised linear mixed models and categorical mixed models. There are also new sections on bioequivalence studies and cluster randomised trials, and a complicated laboratory experiment has been added to the examples. There has been a general updating throughout to reflect recent literature and software changes.

Chapter 1 provides an introduction to the capabilities of mixed models, defines general concepts and gives their basic statistical properties. Chapter 2 defines models and fitting methods for normally distributed data. Chapter 3 first introduces generalised linear models which can be used for the analysis of data which are binomial or Poisson or from any other member of the exponential family of distributions. These methods are then extended to incorporate mixed models concepts under the heading of generalised linear mixed models. The fourth chapter then examines how mixed models can be applied when the variable to be analysed is categorical. The main emphasis in these chapters, and indeed in the whole book, is on classical statistical approaches to inference, based on significance tests and confidence intervals. However, the Bayesian approach is also introduced, since it has several potential advantages and its use is becoming more widespread. Although the overall emphasis of the book is on the application of mixed models techniques, these chapters can also be used as a reference guide to the underlying theory of mixed models.

Chapters 5–7 consider the practical implications of using mixed models for particular designs. Each design illustrates a different feature of mixed models.

Multi-centre trials and meta-analyses are considered in Chapter 5. These are examples of hierarchical data structures and the use of a mixed model allows for any additional variation in treatment effects occurring between centres (or trials) and hence makes results more generalisable. The methods shown can be applied equally to any type of hierarchical data.

In Chapter 6 the uses of covariance pattern models and random coefficients models are described using the repeated measures design. These approaches take into account the correlated nature of the repeated observations and give more appropriate treatment effect estimates and standard errors. The material in this

chapter will apply equally to any situation where repeated observations are made on the same units.

Chapter 7 considers cross-over designs where each patient may receive several treatments. In this design more accurate treatment estimates are often achieved by fitting patient effects as random. This improvement in efficiency can occur for any dataset where a fixed effect is 'crossed' with a random effect.

In Chapter 8 a variety of other designs and data structures are considered. These either incorporate several of the design aspects covered in Chapters 5–7 or have structures that have arisen in a more unplanned fashion. They help to illustrate the broad scope of application of mixed models. This chapter includes the new section on bioequivalence studies and cluster randomised trials.

Chapter 9 gives information on software available for fitting mixed models. Most of the analyses in the book are carried out using PROC MIXED in SAS, supplemented by PROC GENMOD and PROC GLIMMIX, and this chapter introduces the basic syntax for these procedures. This information should be sufficient for fitting most of the analyses described but the full SAS documentation should be referenced for those who wish to use more complex features. The SAS code used for most of the examples is supplied within the text. Additionally, the example datasets and SAS code may be obtained electronically from www.chs.med.ac.uk/phs/mixed.

This book has been written to provide the reader with a thorough understanding of the concepts of mixed models and we trust it will serve well for this purpose. However, readers wishing to take a shortcut to the fitting of normal mixed models should read Chapter 1 for an introduction, Section 2.4 for practical details, and the chapter relevant to their design. To fit non-normal or categorical mixed models, Section 3.3 or Section 4.4 should be read in addition to Section 2.4. In an attempt to make this book easier to use, we have presented at the beginning of the text a summary of the notation we have used, while at the end we list some key definitions in a glossary.

Our writing of this book has been aided in many ways. The first edition evolved from a constantly changing set of course notes that accompanied a three-day course on the subject, run regularly over the previous six years. The second edition has been helped by many individuals who were kind enough to comment on the first edition, including the identification of some errors that had slipped in, and by further participants at our courses who have contributed to discussions and have thereby helped to shape our views. We are also grateful to many other colleagues who have read and commented on various sections of the manuscript and to our colleagues who have allowed us to use their data. We hope that readers will find the resulting book a useful reference in an interesting and expanding area of statistics.

Helen Brown
Robin Prescott
Edinburgh

