1 Introduction

1.1 Introduction

The popularity of applying mixed model analysis has increased rapidly over since around 2005. A very small, non-systematic literature search showed that in 2005, 705 papers were published in which mixed model analysis was applied. In 2010 this number increased to 1292, while in 2016 the number of papers in which mixed model analysis was applied rose to over 2200. Figure 1.1 shows the development from 2005 to 2016 in the number of published papers in which mixed model analysis was applied.

1.2 Background of Mixed Model Analysis

Mixed model analysis was first developed for educational research (Goldstein, 1987, 1992; Goldstein and Cuttance, 1988; Nuttall et al., 1989; Woodhouse and Goldstein, 1989; Plewis, 1991). When analysing the performance of students, the researchers realised that the observations of students in the same class were not independent of each other. Because standard statistical methods assume independent observations, it is not appropriate to use these methods to analyse the performance of students. The structure of such a study can be described as a sort of hierarchy; students are clustered within classes (see Figure 1.2). Because of this hierarchy, mixed model analysis is also known as hierarchical linear modelling.

This situation is known as a two-level data structure, with the first level being the students and the second level being the classes. Because of the different levels, mixed model analysis is also known as multilevel analysis.

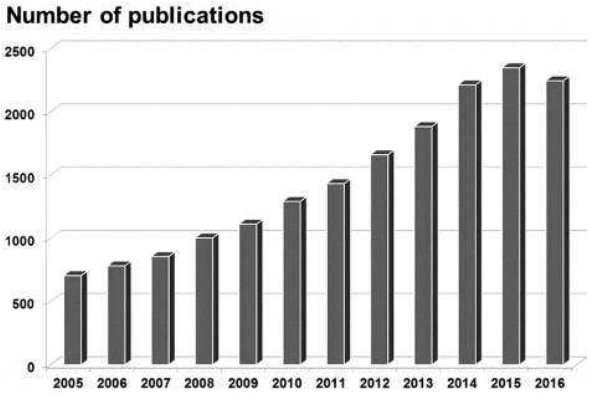

Figure 1.1 Development from 2005 to 2016 in the number of published papers in which mixed model analysis was applied.

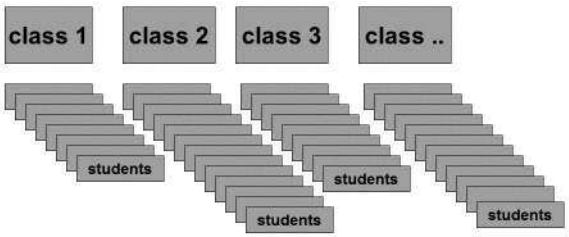

Figure 1.2 Illustration of a two-level hierarchical data structure. Observations of students are clustered within classes.

The general idea of mixed model analysis is that this hierarchy is taken into account in the analysis or, in other words, the analysis takes into account the dependency of the observations. Within the educational setting there can be another (i.e. higher) level of clustering, because not only are the students clustered within classes, but the classes are also clustered within schools (see Figure 1.3). This situation is referred to as a three-level data structure, with the students being level 1, the classes being level 2 and the schools being level 3. Again, the general idea of mixed model analysis in this situation is that it takes into account the dependency of observations, not only within classes, but also within schools.

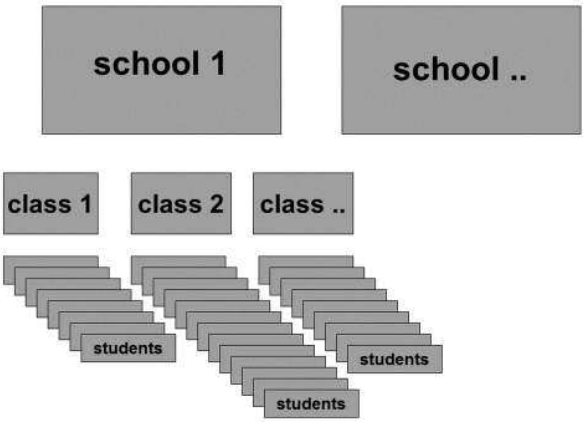

Figure 1.3 Illustration of a three-level hierarchical data structure. Observations of students are clustered within classes and the observations of classes are clustered within schools.

1.3 General Approach

Although there is a considerable amount of basic literature on mixed model analysis, most of it is characterised by a mathematical approach (Bryk and Raudenbush, 1992; Goldstein, 1995, 2003; Kreft and De Leeuw, 1998; Snijders and Bosker, 1999; Little et al., 2000; McCullagh and Searle, 2001; Hox, 2002; Raudenbush and Bryk, 2002; Reise and Duan, 2003; Jiang, 2007, 2016; Demidenko, 2013; Galecki and Burzykowski, 2013; West et al., 2015).

Only a few papers have tried to follow a more practical approach (see, for instance, Korff et al., 1992; Rice and Leyland, 1996; Greenland, 2000a, 2000b; Livert et al., 2001; Diez Roux, 2002; Merlo, 2003; Leyland and Groenewegen, 2003). This book will also follow a more practical approach, which will make it easier to read and more understandable for non-mathematical readers. The emphasis of this book lies on the interpretation of the results of mixed model analysis, on the research questions that can be answered with mixed model analysis, and on the differences between mixed model analysis and the so-called naive approaches that do not take into account the dependency of observations. Therefore, in each chapter, the (mathematically difficult) statistical analyses will be explained by using relatively simple examples, accompanied by computer output.

1.4 Prior Knowledge

In this book an attempt has been made to keep the description of the mixed model analyses as simple as possible. The basis of the explanations will be the underlying research question and the interpretation of the results of the mixed model analyses. However, it will be assumed that the reader has some prior knowledge about standard statistical regression techniques, such as linear regression analysis, logistic regression analysis, multinomial logistic regression analysis, Poisson regression analysis and survival analysis. This is necessary because mixed model analysis can be seen as an extension of the standard regression techniques. So, mixed model analysis with a continuous outcome variable is an extension of linear regression analysis, mixed model analysis with a dichotomous outcome variable is an extension of logistic regression analysis, and so on.

1.5 Example Datasets

All datasets that will be used in the examples will be available from the internet and can be reanalysed by the reader. This will certainly improve understanding of the general theories underlying mixed model analysis.

1.6 Software

All the analyses in the first part of the book are performed with STATA (version 14). In Chapter 13, the use of mixed model analysis in other software packages, such as SPSS, R and SAS, will be discussed. The data used in the examples will be reanalysed with other software and the results will be compared. Both syntax and output will accompany the discussion of the different software packages.