CHAPTER 1

1 INTRODUCTION

1.1 The Scope of the Thesis

Statistical Analysis and the corresponding Statistical Models have been developed in order to describe real life situations and relationships. One of the most common situations, especially in social and behavioural sciences are individuals nested in groups. The general concept is that individuals interact with the social contexts to which they belong, meaning that individual persons are influenced by the social groups to which they belong, and that the properties of those groups are in turn influenced by the individuals who make up that group (Hox, 1995). These kinds of relationships straightforwardly have a hierarchical or clustered structure and the corresponding data stem from different levels of hierarchy. The classical statistical analysis only concerns with the existence of such hierarchies but not their provenance. The proper recognition, however, of these natural hierarchies leads to satisfactory answers to major theoretical and practical questions.

The scope of this project is, therefore, to describe in detail the most appropriate statistical models in analyzing hierarchical data structures, to introduce Multilevel Models Analysis and, specifically, to discuss the comparative advantages of Multilevel Analysis in relation to other methods of such research enquiries, both in theory and in practice. In order to answer this main research question of the project, which is why and how Multilevel Analysis is more useful and effective, we will argue both on theoretical aspects, as well as practical situations where the method can apply. Especially, we will focus on the answer of the main research question by including an application to Greek educational data referring to the General Admission Grade of students in the National Exams.

This type of analysis, where data are measured in different levels, is known as "Multilevel Analysis" and the corresponding models as "Multilevel Models" (Aitkin et al, 1981). Examples and applications will be described and carried out on data taken by a wide area of research, under the presuppositions that data have a hierarchical structure. Such areas are educational performance of students and

organizations, where hierarchical structure of students (the 1st Level units) nested within organizations (schools/universities) (the 2nd Level of Multilevel Analysis) is obvious, as well as other research areas where hierarchy applies (survey research, health statistics and so on).

1.2 The Structure of the Thesis

Chapter 2 introduces the reader to the nature and the basic concepts of Hierarchical Data Structure analysis and discusses the main differences in comparison to analysis of more general and classical types of data. We first introduce some basic notations of variables appeared when hierarchical structure is considered, which are not familiar in a single level regression analysis. We then give a brief idea of the research areas where hierarchical structure is met, in a more or less obvious way. We also comment on the usual techniques of statistical analysis in these areas and the extent to which multilevel techniques have also been applied. We then give an extended description of the possible statistical approaches to analyze hierarchical data structures. We start with the basic model where no hierarchy is considered and we end up to more sophisticated models such as the random coefficient and multilevel models for hierarchical data. In every step of the description we clarify the advantages of the latter compared to more restrictive techniques and we introduce the idea of elaborating more on multilevel models throughout the thesis.

Chapter 3 is an extended description of all the concepts that arise in multilevel modeling. We start from the basic 2-level model and we present all the necessary notations and formulas, the parameter estimations for all the parts of the models as well as the most usual estimation procedures and algorithms. We additionally present hypothesis tests, confidence intervals and models comparison techniques. In the second part of Chapter 3 we introduce, to some extend, the natural extensions of the 2-level basic model, which are necessary to cover the practical needs of multilevel analysis of more complex data that lie beyond the basic 2-level form.

In Chapter 4 we introduce the most important applications, examples and articles that have given rise to discussions about the practical use of multilevel techniques. Some examples are reviewed more extensively, presenting a thorough description of data analysis, methods and results, while for others a simple reference is presented. The applications were chosen so as to cover a wide range of research

areas where multilevel models apply, as well as a wide area of the theoretical considerations of multilevel models, including the extensions of the basic model. On the other hand, they were chosen because their results and comments have formed, or can form, the basis for further discussion in the areas of interest. Although applications were chosen from a wide range of interest areas, educational research was analyzed separately due to the fact that it is a major area of applications of multilevel techniques.

In Chapter 5 we apply techniques and methods and we perform multilevel analysis to real educational data made available by the Greek Ministry of Education, Lifelong Learning and Religious Affairs referring to the years 2006 up to 2009. The aim of the analysis is to detect the factors that affect students' performance in the National Exams for access to the National Universities and Technical Institutions, as measured by their General Admission Grade.

Chapter 6 presents the overall conclusions detected from the whole thesis. Firstly, we refer to all theoretical approaches, techniques and reviews described within the thesis. Moreover, we present the main results drawn by the practical application of multilevel techniques in the real dataset and we compare them to the results of other relative reports. Finally, according to the results of the analysis and the advantages and disadvantages of the techniques described in the thesis, we introduce areas of further research, both in theoretical and in practical basis.