ATHENS UNIVERSITY
OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

MULTILEVEL MODELS ANALYSIS IN HIERARCHICAL DATA STRUCTURE: AN APPLICATION TO EDUCATIONAL DATA

By

Andreas Moursellas

A THESIS
Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
April 2013

ABSTRACT

Andreas Moursellas

MULTILEVEL MODELS ANALYSIS IN HIERARCHICAL DATA STRUCTURE: AN APPLICATION TO EDUCATIONAL DATA

April 2013

Hierarchically structured datasets are very common in real life situations. These types of datasets appear in various areas of statistical analysis, such as Educational Statistics, Spatial Statistics, Health Statistics, Repeated Measures and Survey Research.

In this thesis we focus on Multilevel Models techniques and methods of analysis of hierarchical data, since it is proved that they provide more accurate results compared to more classical approaches. We also apply multilevel statistical analysis to educational data obtained from the Greek Ministry of Education, Lifelong Learning and Religious Affairs referring to the General Admission Grades of students in the National Exams for access to Universities and Technical Institutions for the years 2006 up to 2009. This is the first time multilevel analysis is carried out for data in the particular educational system of access. The results of the analysis show that multilevel models are more effective compared to simple ones, since a 3-level model (students nested in schools and schools in prefectures) is a significant improvement compared to 1-level models. Also, the factors that are detected to have a significant effect on the performance of students are consistent with the results of previous studies. According to the analysis, female students perform better than male students. Also, students examined for the Exact Sciences orientation have the highest performance, while students examined for the Human Sciences orientation, in general, perform slightly better than those of Technical Sciences orientation. However, males for the Technical Sciences orientation perform better than those for the Human Sciences orientation. The best year for students' performance is 2009, while 2006 -first year of application of the new educational system- is by far the worst. Finally, the performance of students from private schools in general is much better than those from public schools. In public schools the performance of students for the Exact Sciences orientation is higher than the "usual pattern".

ΠΕΡΙΛΗΨΗ

Ανδρέας Μουρσελλάς

ΑΝΑΛΥΣΗ ΠΟΛΥΕΠΙΠΕΔΩΝ ΜΟΝΤΕΛΩΝ ΣΕ ΙΕΡΑΡΧΙΚΗ ΔΟΜΗ ΔΕΔΟΜΕΝΩΝ: ΜΙΑ ΕΦΑΡΜΟΓΗ ΣΕ ΕΚΠΑΙΔΕΥΤΙΚΑ ΔΕΔΟΜΕΝΑ

Απρίλιος 2013

Τα ιεραρχικά δομημένα σύνολα δεδομένων είναι πολύ συνηθισμένα σε πραγματικές καταστάσεις. Τέτοιων ειδών σύνολα δεδομένων εμφανίζονται σε διάφορους τομείς της στατιστικής ανάλυσης, όπως Εκπαιδευτική Στατιστική, Χωρική Στατιστική, Στατιστική Υγείας, Επαναλαμβανόμενες Μετρήσεις και Μελέτη Ερευνών.

Σε αυτήν την διατριβή εστιάζουμε σε τεχνικές και μεθόδους Πολυεπίπεδων Μοντέλων ανάλυσης ιεραρχικών δεδομένων, δεδομένου ότι αποδεικνύεται ότι παρέχουν ακριβέστερα αποτελέσματα σε σύγκριση με περισσότερο κλασσικές προσεγγίσεις. Εφαρμόζουμε επίσης πολυεπίπεδη στατιστική ανάλυση σε εκπαιδευτικά δεδομένα που λαμβάνονται από το Ελληνικό Υπουργείο Παιδείας, Δια Βίου Μάθησης και Θρησκευμάτων και αναφέρονται στους Βαθμούς Γενικής Πρόσβασης των μαθητών στις Εθνικές Εξετάσεις για πρόσβασή σε Πανεπιστήμια και Τεχνολογικά Ιδρύματα για τα έτη 2006 μέχρι 2009. Είναι η πρώτη φορά που πολυεπίπεδη ανάλυση εφαρμόζεται για δεδομένα στο συγκεκριμένο εκπαιδευτικό σύστημα πρόσβασης. Τα αποτελέσματα της ανάλυσης δείχνουν ότι τα πολυεπίπεδα μοντέλα είναι πιο αποτελεσματικά σε σύγκριση με τα απλά, καθώς ένα 3-επίπεδο μοντέλο (μαθητές να βρίσκονται μέσα σε σχολεία και σχολεία σε νομαρχιακά διαμερίσματα) είναι μια σημαντική βελτίωση σε σύγκριση με 1-επίπεδα μοντέλα. Επίσης, οι παράγοντες που διακρίνονται να έχουν μια σημαντική επίδραση στην απόδοση των μαθητών είναι σύμφωνοι με τα αποτελέσματα προηγούμενων ερευνών. Σύμφωνα με την ανάλυση, οι μαθήτριες αποδίδουν καλύτερα από τους μαθητές. Επίσης οι μαθητές που εξετάζονται για τη Θετική κατεύθυνση έχουν την υψηλότερη απόδοση, ενώ οι μαθητές που εξετάζονται για τη Θεωρητική κατεύθυνση, γενικά, αποδίδουν ελαφρά καλύτερα από αυτούς της Τεχνολογικής κατεύθυνσης. Ωστόσο, οι μαθητές (αγόρια) για την Τεχνολογική κατεύθυνση αποδίδουν καλύτερα από αυτούς

για τη Θεωρητική κατεύθυνση. Το καλύτερο έτος για την απόδοση των μαθητών είναι το 2009 ενώ το 2006 –πρώτο έτος εφαρμογής του νέου εκπαιδευτικού συστήματος- είναι μακράν το χειρότερο. Τέλος, η απόδοση των μαθητών από ιδιωτικά σχολεία είναι γενικά πολύ καλύτερη από αυτούς από δημόσια σχολεία. Στα δημόσια σχολεία, η απόδοση των μαθητών για τη Θετική κατεύθυνση είναι ψηλότερη από το «συνηθισμένο μοτίβο».

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

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.

CHAPTER 2

2 INITIAL CONCEPTS OF HIERARCHICAL DATA STRUCTURE

In this Chapter we introduce all the basic concepts of Hierarchical Data Structure analysis, the research areas where hierarchical structure is met, as well as all the possible statistical approaches to analyze hierarchical data structures in these various research areas.

The scope of this chapter is, starting from simple to most sophisticated models, to describe the limitations of the simplest models for analyzing hierarchical data and, therefore, to justify the need of using Multilevel Models for an effective analysis.

2.1 Types of Variables in Hierarchical Data Structure

In traditional data analysis techniques, where all data are measured at the same level, we simply refer to the "response" (dependent) variable, which is the variable to be examined, by using a number of "explanatory" (dependent) variables. In Multilevel Research, however, the situation is quite more complex, since variables can be defined at any level of hierarchy. In order to make clear to which level the measurements properly belong, we present a brief description of each type of variables (based on Hox, 1995) as well as a table with the relations between different types of variables, defined at different levels.

Some of the variables may be measured directly at their natural level; for example, at the school level we may measure school size and denomination, and at the students' level intelligence and school success. In addition, we may move variables from one level to another by aggregation or disaggregation. Aggregation means that the variables at the lower level are moved to a higher level, for instance by computing the school mean of the students' intelligence scores. Disaggregation means moving variables to a lower level, for instance by assigning to all pupils a variable that reflects the denomination of the school they belong to.

Table 2.1: Types of variables in Hierarchical Structure Data Models

Level:	1		2		3	etc
Variable Type:	Absolute	⇒	Analytical			
	Relational	⇒	Structural			
	Contextual	⇐	Global	⇒	Analytical	
			Relational	⇒	Structural	
			Contextual	⇐	Global	⇒
					Relational	⇒
					contextual	⇐

As shown in Table 2.1, at each level we have several types of variables, however some of them are related to each other. *Global* and *Absolute* variables refer only to the level at which they are defined, without reference to any other units or levels ('absolute variables' is simply the term used for global variables defined at the lowest level). A pupil's intelligence would be a global or absolute variable. *Relational* variables also refer to one single level; they describe the relationships of a unit to the other units at the same level. Many sociometric indices are relational variables. *Analytical* and *Structural* variables are measured by referring to the subunits at a lower level. Analytical variables refer to the distribution of an absolute or a global variable at a lower level, for instance to the mean of a global variable from a lower level. Structural variables refer to the distribution of relational variables at the lower level; many social network indices are of this type. Constructing an analytical or relational variable from the lower level data involves *aggregation* (indicated by ⇒): data on lower level units are aggregated into data on a smaller number of higher level units. *Contextual* variables, on the other hand, refer to the superunits; all units at the lower level receive the value of a variable for the superunit to which they belong at the higher level. This is called *disaggregation* (indicated by ⇐): data on higher level units are disaggregated into data on a larger number of lower level units. The resulting variable is called a *contextual* variable, because it refers to the higher level context of the units we are investigating.

2.2 Areas with Hierarchical Data Structure

Multilevel Analysis can be applied in situations where the existence of data hierarchies is neither accidental nor ignorable (Goldstein, 1995), in other words the hierarchical data structure is straightforward, or in cases where the applications of multilevel research are not so obvious. This chapter is just an introductory description of some research areas where Multilevel Analysis is commonly used. More detailed examples, when necessary, are presented in Chapter 4 of the thesis, both from educational and from other areas of interest:

Organizational Research: The case where individuals are nested within their organizations. An illustrative example was performed by Kreft et al. (1995), when data were collected on workers in 12 different industries. The hierarchical structure of such model is straightforward, since the workers (first-level units) are nested in industries (second-level groups). The response variable of interest was the income of the workers, measured at the first level of hierarchy and explanatory variables were measured at both levels. Also aggregated measures were used in the analysis. We won't go further to the techniques as well as the results of this particular study since the method used are "overlapped" in following examples.

Clinical Therapy and health research: Clinical psychology is another area where multilevel techniques can be a rational approach, especially in the evaluation of group therapy research. In such studies the patients (first-level objects) are very often gathered in particular therapy groups (second-level objects). In group therapy the type of therapy is an effect under the control of the researcher, but the group dynamics is not. Therapy groups are, at the outset, as much alike as chance can make them by randomly assigning clients to therapy groups, but they change over time. The interactions within each group depend on the dynamics of the group, which develops over time in unpredictable directions.

If the two types of group therapy administered are directive intervention and non-directive interventions, groups within the same treatment can become different, especially under the non-directive intervention treatment. The behaviour of each client starts to reflect the type of therapy as well as the specific dynamics that develops in the client's therapy groups. The interaction between group members makes clients in the same group more alike than clients in different groups. Consequently, the observations of group members can no longer be considered statistically independent.

Application of traditional statistical methods, such as ANCOVA, will fail to analyze correctly this kind of data, in the sense that it will ignore the intra-class correlation that develops over time, leading to an underestimation of the error variance of the estimated coefficients. The group dynamics cannot be modeled in a traditional ANCOVA model, nor can characteristics of the therapist. A multilevel model, on the contrary, will take care of the dependency of observations within groups, and also will model differences between groups by means of macro-level characteristics, such as different approaches by therapists and different group dynamics.

Health economics is another area of health research where multilevel techniques can be applied. We will focus more on this area in Chapter 4 of the thesis.

Twin Studies: A very special case of Clinical Research in which multilevel models can apply is the Twin Studies. This case is indeed special because we may have a large number of groups (2^{nd} level units) but all these groups are of size two. This is an uncommon case in hierarchical or clustered data where a large number of individuals (level-one units) are nested within a small number of common groups. On the other hand, within these small samples of two, we expect to have a high intra-class correlation. Since it is difficult to estimate a model within each group with only two observations, the statistical stability has to come from the number of groups. A Multilevel approach provides the advantage to introduce separate variables for the individuals in the pairs (1^{st} level variables) and variables which the members of the pairs have in common (2^{nd} level variables).

Repeated Measures and Growth curve analysis: Repeated-measures and growth curve analysis is a wide area of research in Statistical Analysis. Although it is not straightforward, longitudinal studies can be considered as an issue where Multilevel Analysis techniques can be performed. However, they form a special case in Multilevel Analysis, since the individuals are the 'macro level' instead of the 'micro level' as in common cases. In other word, here the occasions (measurements) are clustered within individuals that represent the level-two units with measurement occasions the level-one units. Such structures are typically strong hierarchies because there is much more variation between individuals in general, than between occasions within individuals who are, naturally, correlated. Intra-class correlation, in this case measures the degree to which behaviour of the same person is more similar to his/her own previous behaviour in comparison to behaviour of other people (Kreft, 1998). In the case of child growth, for example, once we have adjusted for the overall trend with age, the variance between successive measurements on the same individual is generally no more than 5% of the variation in height between children (Goldstein, 1995).

The major advantage of performing Multilevel Techniques in repeated measures analysis is that they can handily deal with unbalanced data structure with missing values. There are cases in practice where individuals are measured irregularly, some of them a great number of times and some perhaps only once. By considering such data as a general 2-level structure (measurements-the 1st level and individuals-the 2nd level) we can apply the standard set of multilevel modeling techniques while providing statistically efficient parameter estimation and at the same time presenting a simpler conceptual understanding of the data.
Later in the thesis we will focus more on application of multilevel analysis in the area of repeated measures, by presenting a rather special example of autobiographical memories.

Geographical information systems: Spatial Statistics is also an area where multilevel techniques can be easily adapted to analyze the measured data. Such cases are census data, election data, demographical studies and so on. It is obvious that in all these cases sites or individuals are nested within geographic regions, and thus the intra-class correlation comes from spatial autocorrelation. The measured variables can refer to, either geographical characteristics and information about the region (2nd level units), or characteristics of the individuals themselves (1st level units).

An illustrative example was presented by Courgeau and Baccaini (1998). In their work the migration flows of the 19 Norwegian regions is being examined, by using multilevel logit techniques. We will focus on this example with more details later on in the thesis.

Survey Research: Survey research is a wide area where multilevel techniques can be and are already applied in order to examine the, more or less, obvious hierarchical structure of the data measured. This concerns both the methological aspect of sampling design and the practical aspect of data collection through interviews.

The standard literature on surveys, reflected in survey practice, recognizes the importance of taking account of the clustering in complex sample designs. Thus, in a household survey, the first-stage sampling unit will often be a well-defined geographical unit. From those geographical areas, which are randomly chosen, further stages of random selection are carried out until the final households (the final sample) are selected. This is an obvious case of hierarchical data structure where respondents nested in the same geographical area will be more similar to each other than respondents from different areas. However in the analysis insofar the population structure, as it is mirrored in the sampling design, is seen as a 'nuisance factor'. Ignoring hierarchy in the analysis will cause estimates for standard errors that are too small, and 'spurious' significant results – the so-called 'design effect' in survey research. The most usual correction for design effects, taking into account the clustering of individuals within groups, is to compute the standard errors by ordinary analysis methods, estimate the intra-class correlation between respondents within clusters (geographical regions) and to employ a correction formula to the standard errors (Kish,1987).

Although these procedures developed to produce valid statistical inferences can be quite powerful (Skinner et al, 1989), they still do not allow for simultaneous analysis from variables taken from different levels, using a statistical model that includes the various dependencies. In other words, the multilevel modeling approach views the population structure as of potential interest in itself, so that a sample designed to reflect that structure is not merely a matter of saving costs as in traditional survey design, but can be used to collect and analyze data about the higher level units in the population. The subsequent modeling can then incorporate this information and obviate the need to carry out special adjustment procedures, which are built into the analysis model directly.

In chapter 4 we will focus more on an example of survey research taking into account the hierarchy of individuals nested within households nested within geographical areas.

Another aspect of survey research where multilevel techniques are applied is in the so-called 'Interviewer Effect" on the results a study. Whenever the sampling method of the survey is telephone or personal interviews, the interview is carried out at the respondents from a smaller number of interviewers. Even though it is not so obvious, this kind of survey forms a hierarchical structure where respondents (the 1^{st} level units) are "nested" within their interviewers (the second-level of the analysis). It is logical, even though it is not desirable, that in many cases the interviewer characteristics affect the results of the survey, or in other words the respondents being interviewed by the same interviewer tend to have more similar answers. Multilevel analysis can detect such interviewers effect by introducing variables measured in both levels, as well as interactions between interviewers and respondents (cross-level interactions). An illustrative example performed by Hox (1994) will be presented later on in the thesis (Chapter 4).

Meta-Analysis: Meta-analysis or integrative analysis, as it is often called, is a quantitative approach to reviewing the research literature. The term Meta Analysis (Hedges & Olkin, 1985) refers to the pooling of results of separate studies, all of which are concerned with the same research hypothesis. The aim is to achieve greater accuracy than that obtainable from a single study and also to allow the investigation of factors responsible for between-study variation. In other words, the primary goal of meta-analysis is to generalize from a set of studies about a specific substantive issue, by statistically combining quantitative study outcomes from existing research on a particular question. The basic idea is to apply formal statistical methods to the results of a specific set of studies. The statistical approach is the one of the main characteristics that distinguishes meta-analysis from the more traditional narrative literature review (Bangert-Drowns, 1986). Each study typically provides an estimate for an 'effect', for example a group difference, for a 'common' response and the original data are unavailable for analysis. In general, the response measure used will vary, and care is needed in interpreting them as meaning the same thing. Furthermore, the scales of measurement will differ, so that the effect is usually standardized using a suitable within-study estimate of between-unit standard deviation.

Clearly, in a meta-analysis the most important preliminary question is, whether the results differ more from each other than corresponds to the random sampling variation that is expected given the studies' sample size. If the results do not differ more than is expected given the pure sampling error, they are called homogeneous, meaning that they come from a single population. In the next analysis step we would want to estimate the common value of the population parameter of interest. If the results differ more than expected given the pure sampling variation, they are called heterogeneous, meaning that they come from different populations. In this case, estimating the 'average' result is not the primary goal; instead, our goal becomes to analyze the excess variation as a function of the known study characteristics such as the age or sex composition of the sample, or methological characteristics such as the methological quality of the study.

There are various methods to analyze and combine separate study results, one of them, even not so profound, is multilevel analysis. The problem of combining the varying results from different studies has some similarity to the multilevel problem of combining the varying micro-models from different groups or contexts. If we had access to the original data of all the studies, we could analyze them using the hierarchical regression model. But in meta-analysis we generally do not have access to the row data. Still, the statistical problem looks familiar. In multilevel modeling we have a number of regression models computed in different contexts, and we want to estimate the expectation and the variability of the various regression coefficients, and draw conclusions based on all available information. In meta-analysis we have a number of statistics computed in different contexts, and we want to assess their average value and their variability, and again draw conclusions based on all available information. According to Raudenbush & Bryk (1992) meta-analysis may be viewed as a special case of the two-level hierarchical linear model. In each study, a within study model is estimated, and a second level or between study model is added to explain the variation in the within study parameters as a function of differences between the studies. The variability within the studies is considered to be sampling variability, which is known if the relevant sampling distribution and sample size are known. The variability between the studies reflects both sampling variance and systematic differences between the results of different studies. If the study level variance is significant, the studies' results are assumed to be heterogeneous, meaning that there are indeed systematic differences between the studies. If the study level variance is not significant, they are assumed to be homogeneous, meaning that the apparent differences between the studies are just sampling variance.

School Effectiveness Research: Schooling and more generally educational systems is an area where the hierarchical data structure is profound, since students (the 1st level units) are nested or clustered within schools/universities (the 2nd level units), which themselves may be clustered within education authorities or boards or geographical regions (the 3rd level units) and so on. In special cases, students form the 2nd level of hierarchy when repeated measures (the 1st level) are performed within the same student.

In one or another way, school and other institutions effectiveness has been widely examined by educational researchers who have been interested in comparing schools and other educational institutions, most often in terms of the achievements of their pupils. In other examples, the teachers effectives on students achievement is of primary interest, see Bennet (1976) and the analysis of different teaching styles, which was reconsidered by Aitkin et al. (1981) using more advanced statistical methodology. In other cases we are interested in studying the extent to which schools differ for different kinds of students, for example to see whether the variation between schools is greater for initially high scoring students than for initially low scoring students (Goldstein et al, 1993) and whether some factors are better at accounting for or 'explaining' the variation for the former students than for the latter. Moreover, there is often considerable interest in the relative ranking of individual schools, using the performances of their students after adjusting for intake achievements. The response variable in most of the cases discussed above is the students' performance, which is measured, for example, by an examination test.

Most of the traditional approaches for the analysis of such data have been carried out by researchers, such as regression analysis, sometimes by fitting a separate regression line within each group, or ANCOVA models that treat schools as fixed factor. All of them however, either ignore the hierarchy of the data structure at all, or fail to correct for the intra-class correlation within each group. Multilevel techniques stand for a straightforward solution, since multilevel analysis models introduce variables in all levels of hierarchy simultaneously, as well as interactions of the characteristics between levels. They can answer, therefore, to all the theoretical questions as stated in the previous paragraph, and, at the same time obtain statistically efficient estimates of regression coefficients, correct standard errors, confidence intervals and significance tests.

In Greece, although the school effectiveness and students' performance issue has been discussed widely, statistical approaches are very poor in literature and, when they exist, they are mainly constrained in descriptive presentation of the results (Centre of Development of Educational Policy, General Confederation of

Greek Workers (GSEE) 2009). Marouga (2004) used more sophisticated factor analysis and cluster analysis techniques in order to examine students' performance and preferences according to the Greek National Exams. Moreover, Kosmopoulou's dissertation (1998) is the only statistical project where multilevel techniques were performed in Greek educational data in order to assess school effectiveness and students' performance in the National Exams of 1990 and 1991. However, both projects referred before analyze educational data from a national educational system of student's Access in the National Universities and Technical Institutions that does not longer exist.

A more thorough examination of multilevel techniques approach in educational data and school effectiveness research will be performed later on in the thesis, by reviewing, presenting or applying representative examples.

2.3 Possible Approaches for Hierarchical Data Structure - Traditional Models to Random Coefficient Models

In this Chapter we present a number of variations on the ordinary linear model and on OLS regression that have been suggested to deal with hierarchically nested data. They vary from total or pooled regression, which completely ignores the between-group variation, to aggregate regression, which completely ignores the within-group variation. And, on another dimension, they vary from separate regressions for each group, with separate sets of regression parameters, to a single regression with only one set of parameters.

In many cases, however, it makes sense to take the group structure into account more explicitly. Forms of regression analysis, in which both individual and group level variables are used, are known as contextual analyses. In contextual analysis group membership is not neglected. The units of observation are treated as members of certain groups, because the research interest is in individuals as well as in their contexts. Traditional contextual models, the Cronbach model, the ANCOVA model and the various multilevel models decompose the variation in the data into a within and a between part, but each in their one way.

2.3.1 Models and formulae

In contextual analysis techniques the free parameters of the linear model are estimated based on the following model, where $\underline{y}$ is the response variable, x the explanatory variable at the individual level and z is the explanatory variable at the context level. The subscript i is for individual, and j is for context. The model is:

$$\underline{y}_{ij} = a_j + bx_{ij} + cz_j + \underline{\varepsilon}_{ij}. \qquad (2.1)$$

The $\underline{\varepsilon}_{ij}$ are disturbances, which are centered, homoscedastic and independent. This means they have expectation zero and constant variance σ^2. Generally, of course, there may be more than one explanatory variable on both levels. We will discuss more on model structure in the next Chapter by using even more appropriate formulas, since in this Chapter the main interest is to explain the differences between the various models.

Model (2.1) can be expressed in a slightly different way, that more clearly shows its structure. We write:

$$\underline{y}_{ij} = a_j + bx_{ij} + \underline{\varepsilon}_{ij} \qquad (2.2a)$$

$$a_j = a + cz_j \qquad (2.2b)$$

Equation (2.2b) shows that the contextual models of equation (2.1) are varying intercept models, i.e. regression models for each group which are linked because they have the same slope b and the same error variance σ^2. They differ, however, in their intercepts. The different contextual models we discuss in this chapter specify the relationship between the varying intercepts and the group-level variables in different ways.

In hierarchically nested data with two levels the variances and covariances of the observed variables can be divided into a between-group and a within-group matrix. This distinction of between and within variation of variables is not straightforward and differs from technique to technique. To explain the definition of regression coefficients in different models we make use of the notion of the correlation ratio. The correlation ratio is the percentage group variance of a variable, which can be explained as follows. Variables such as x as an explanatory variable and y as response variable can be divided into a between- and a within-group part. This induces a corresponding decomposition of the variances as

$$V_T(x) = V_B(x) + V_W(x) \qquad (2.3a)$$

and equally

$$V_T(y) = V_B(y) + V_W(y) \qquad (2.3b)$$

where the indices T, B and W denote total variance, between-group variance and within-group variance, respectively. Moreover, the total covariance between variables x and y can be divided in the same way into a within and a between part,

$$C_T(x,y) = C_B(x,y) + C_W(x,y) \qquad (2.3c)$$

where C denotes covariance.

The coefficients for regressions over the total sample b_T, between groups b_B and within groups b_W, can be defined by the variances within or between groups, compared to the total variance, as follows:

$$b_T = \frac{C_T(x,y)}{V_T(x)}, \qquad (2.4a)$$

$$b_B = \frac{C_B(x,y)}{V_B(x)}, \qquad (2.4b)$$

$$b_W = \frac{C_W(x,y)}{V_W(x)}, \qquad (2.4c)$$

These coefficients can be related to the correlation ratio η^2, defined for x and y in the following way:

$$\eta^2(x) = \frac{V_B(x)}{V_T(x)}, \qquad (2.5a)$$

$$\eta^2(y) = \frac{V_B(y)}{V_T(y)}. \qquad (2.5b)$$

The equations show group variation in the response variable as the percentage of the total variance in y declared between groups. This is at the same time the definition of the intra-class correlation. We will return to this measure later on in the thesis, since it is of great importance in Multilevel Analysis.

Also,

$$1 - \eta^2(x) = \frac{V_W(x)}{V_T(x)}, \quad (2.6a)$$

$$1 - \eta^2(y) = \frac{V_W(y)}{V_T(y)}, \quad (2.6b)$$

The proportion of variance within groups is equal to $1-\eta^2(x)$, and equal to the ratio of the within variance and the total variance.

We know from classical regression theory that the "best" estimate of b for the regression over the total sample, irrespective of group membership, is b_T. It can be shown that the estimate of b_T is a weighted composite of the between-group regression b_B and the within-group regression b_W, as we can see in the following equation:

$$b_T = \eta^2(x)b_B + (1-\eta^2(x))b_W. \qquad (2.7)$$

2.3.2 Total or Pooled regression

The first technique we discuss is a simple one. It is not a multilevel analysis, and in most cases not even a contextual analysis. We analyze the effect of the explanatory variable of the individual level on the response variable in a single regression for the total sample. No context variable is used; the fact that some individuals are in the same group and others are in different is not reflected in the model.

Executing a regression analysis over the total sample of individuals, ignoring group membership, is the same as ignoring the subscript j in equation (2.1). The model becomes

$$\underline{y}_{ij} = a + bx_{ij} + \underline{\varepsilon}_{ij}, \qquad (2.8)$$

where the $\underline{\varepsilon}_{ij}$ are independent, with mean zero and constant variance σ^2. For completeness, and for later comparisons, we also fit the corresponding null model, with only the intercept a and no explanatory variable. This null model is:

$$\underline{y}_{ij} = a + \underline{\varepsilon}_{ij}. \qquad (2.9)$$

A regression analyzing individual observations over the total group is called a total regression. The individual is the unit of analysis, the unit of sampling and the unit of decision-making. Using this analysis means that no systematic influence of groups on the response variable is expected, and all influences of the groups are incorporated in the error term of the model. The fact that the observations are nested within groups is disregarded, and assumed to be of no importance for the research question. In terms of the contextual model (2.2a), in the total regression the intercepts a_j are assumed to be equal for all groups j.

2.3.3 Aggregate regression

One rather crude way to take the grouping of the individuals into account is to do a regression over the group means, a so-called aggregated analysis. There is a priori no real reason to expect that regression coefficients from a total regression analysis and those from an aggregate regression analysis will be similar. In fact, it is easy to construct examples in which the differences between the two techniques will be very large.

For the analysis we form the means for the explanatory variable $x_{\bullet j}$, the means for the response variable $y_{\bullet j}$, and we fit the model

$$\underline{y}_{\bullet j} = a + bx_{\bullet j} + \underline{\varepsilon}_j, \quad (2.10)$$

where the bullet replaces the index for individuals i to indicate that the x and y are summed over individuals. As usual, it is assumed that $\underline{\varepsilon}_j$ has a mean of zero. The variance of $\underline{\varepsilon}_j$ is now, compared to the total model, $n_j^{-1}\sigma^2$, because it is a mean of n_j disturbances, each with variance σ^2. In this analysis we fit a weighted regression, with weights equal to n_j. The regression is heteroscedastic.

Clearly aggregate regression ignores all within-groups variation, and thus throws away a large amount of possibly important variance. At the same time, the standard errors of the regression coefficients normally become much larger, because they are based on only n_j observations. Aggregate regression equations must be interpreted carefully. From the prediction point of view, we can state predictions and merely draw conclusions for individuals, and actually making such statements on the basis of aggregated results. This is known as the "ecological fallacy" (Robinson, 1950).

2.3.4 The contextual model

The contextual model has been used widely in the past in research interested in the effect of group membership on individual behaviour. Typically in this type of analysis the group mean of an individual-level variable is used as a contextual variable. Together with the individual level characteristic x_{ij}, a characteristic of groups is defined as the average value of the groups' members $x_{\bullet j}$. The same measurement is used twice in the same regression, once as the original individual measurement, and once as the mean for each group. In other words, the characteristic is aggregated from individual to group level. The model is thus written as follows:

$$\underline{y}_{ij} = a_j + bx_{ij} + \underline{\varepsilon}_{ij} \quad \text{(2.11a)}$$

$$a_j = a + cx_{\bullet j} \quad \text{(2.11b)}$$

Substitution gives us the following equation for the contextual model:

$$\underline{y}_{ij} = a + bx_{ij} + cx_{\bullet j} + \underline{\varepsilon}_{ij}\,. \quad \text{(2.12)}$$

It turns out that the best estimate of b in equation (2.12) is b_W, while the best estimate of c is $b_B - b_W$. It can be shown (Duncan et al., 1966) that the within regression (b_W) is confounded with the between regression (b_B) in the estimation of the context effect.

Some more technical problems are present in this contextual model, one related to multicollinearity and one to the level of analysis. Multicollinearity is introduced in this analysis by the correlation of the individual variable and the group mean for this variable. The level of analysis is the individual, because the response variable is defined at the individual level. Performing a regression analysis at one level ignores the true hierarchically nested structure of the data, and treats the aggregated variable as if it was still measured at the first level. The contextual effect in this contextual model is merely the difference between b_B and b_W. It is clear that the individual and group effects are confounded in c, and as a result interesting and significant relationships can be distorted by this procedure.

2.3.5 The Cronbach model

The Cronbach Model (Cronbach & Webb, 1975) provides a clearer picture of the individual effect together with the group mean effect on the response variable. The individual variables are first centered around their respective group means, as in the following equation:

$$\underline{y}_{ij} = a + b_1(x_{ij} - x_{\bullet j}) + b_2(x_{\bullet j} - x_{\bullet\bullet}) + \underline{\varepsilon}_{ij}. \quad (2.13)$$

In equation (2.13) the centered individual scores $x_{ij} - x_{\bullet j}$ form a variable that is orthogonal to the variable formed by the centered group-level scores $x_{\bullet j} - x_{\bullet\bullet}$. Raw scores are thus transformed into deviation scores from the group mean. Centering explanatory variables in this model provides a convenient way of avoiding the problem of correlation between the two variables that are measurements for a characteristic at the two different levels. The two predictors in the Cronbach model are a centered individual characteristic and the centered group mean for this characteristic analyzed again with regression. Because the two predictors are orthogonal, the best estimate of b_1 is equal to b_W, and thus also to the estimate in the contextual model discussed previously. The difference compared to the contextual model is in the estimate for the contextual effect, where b_2 is now equal to b_B and thus equal to the effect of b_B in the aggregate model. Within and between effects are no longer confounded in the Cronbach model.

Although the collinearity problem of the correlation between the individual variable and its aggregated counterpart is solved in the Cronbach model, the significance tests are just as suspect as they are in the contextual model. In both contextual models discussed so far, the analysis is executed at the lower level. As a result the standard error for the coefficient of the group mean is underestimated. The result is an increase in the alpha level of the test of significance. The group mean has only as many independent observations as the number of groups. Since we have say k groups with nk observations each, the total number of observations on which the standard error is based is $k \times n_k$, instead of the correct number k. Another threat to the validity of the standard errors in the above contextual model is intra-class correlation in the sense that when intra-class correlation is present, the alpha level enhances.

2.3.6 Analysis of Covariance (ANCOVA)

Analysis of covariance is another traditional way of analyzing group data. Both levels are included in the model but not in equal roles. Individual-level explanatory variables are involved, as in regression models, but at the same time groups are allowed to differ in the intercepts. The ANCOVA model incorporates both quantitative and qualitative variables and therefore has a mixed character. It is a regression model, with dummy variables to code group membership. While the regression model enables us to access the effect of quantitative factors, ANCOVA enables us to model qualitative factors.

ANCOVA is a technique with a somewhat different purpose from contextual analyses. It evaluates the effect of groups, correcting for pre-existing differences among these groups. With this technique we can study if the groups are equal in the response variable, corrected for the differences in the amount of a first- level variable. Such an analysis would tell us if groups differ in average response, and which group "scores", on average, the best. In ANCOVA the individual effects are neglected, or considered as noise, and the emphasis is on the group effect.

The individual variable(s) functions as covariate (s), while the grouping is used as the important factor in the design. Because the model was originally developed for designed experiments, groups in ANCOVA are considered to be different treatment categories. The equation for the analysis of Covariance is

$$\underline{y}_{ij} = a_j + bx_{ij} + \underline{\varepsilon}_{ij}. \quad (2.14)$$

Different values for a_j mean that some groups have higher "starting values" for the response variable than others. The assumption in ANCOVA, that all groups have the same slope (the b in the model), means that we assume that the relation between the first-level explanatory variable and response variable is the same for all groups. We can see that equation (2.14) is the same as (2.2a) and that (2.2b) is missing. There is no additional structure imposed on the a_j; they can take all possible values.

Since ANCOVA expresses the differences between k groups using all $k-1$ degrees of freedom, the model provides an upper limit on the amount of variance potentially attributable to overall differences in contexts. In contrast to the traditional contextual model in equation (2.1), ANCOVA cannot tell us which characteristics of the context explain the differences between them. The only thing it shows is how large the overall group effect is, by giving a measure of the explained between-group variation of the intercepts.

The chief advantage of ANCOVA is that it has greater predictive power than the traditional contextual models, as in equations (2.2a) and (2.2b). ANCOVA accounts for all variability between the context means, and not only for variability related to a context-specific explanatory variable, as in contextual models. At the same time the specificity is strength of the contextual model, because it identifies important group characteristics. Most researchers consider the analysis of (co)variance useful as an estimate of the composite group effect preliminary to contextual analysis. It is true that where the a_j in a covariance analysis adds little explained variance, we know from the outset that none of the context characteristics can explain much additional variance of the response variable in subsequent models. But that is only true if variation among contexts is studied in relation to the intercepts, the main effects. But more and more research is dedicated to studying differences among contexts in relationships between explanatory variables and response variable, the b-coefficients in model (2.14). The assumption of ANCOVA that each of the k explanatory variables, or covariates, has the same relation with the response variable over all groups is unrealistic. Each group may need its own unique solution, and its own unique relation between the response variable and the explanatory variable.

2.3.7 Moving from one single-level to multilevel-model techniques

By now, we have discussed some of the traditional ways of analyzing grouped data that consist of two levels, an individual one and a contextual one. The data analysis in these models is always executed at one single level, which can be either the individual or the context level. Analyses executed at the individual level can still be different in the way they handle the between variation. As a result, different regression estimates for the contextual effect are observed among models. From the discussion of the models and the different results, we see that we are in need of a more general model. We need a model that treats the data at the level they are measured, and can answer research questions about the influence of all explanatory variables on the response variable, irrespective of the level in the hierarchy at which they are measured, or to which they are aggregated. Such models are multilevel models, i.e. the Varying Coefficients Models and their modern version, the Random Coefficient (RC) models. Varying coefficient models are also known as the "slopes-as-outcomes" approach.

Varying coefficients or "slopes-as-outcomes"

Traditional strategies for analyzing group data are several forms of regression analysis. The basic equation defining this linear model is

$$\underline{y}_{ij} = a_j + b_j x_{ij} + \underline{\varepsilon}_{ij}, \qquad (2.15)$$

which is similar to equation (1.2a) discussed previously. In equation (2.15) x is again the individual explanatory variable, and y the response variable. The a_j are intercepts and the b_j are slopes. We use the plural form, since instead of the usual single intercept and single slope, separate ones are estimated for each context. To indicate that fact in the formula the subscript j is added to the coefficients a and b. Thus subscript j refers to contexts and subscript i to individuals. The $\underline{\varepsilon}_{ij}$ is the usual individual error term, with an expectation (mean) of zero and a variance of σ^2. In equation (2.15) only $\underline{\varepsilon}_{ij}$ and $\underline{y}_{ij}$ are random variables. Later, when we move away from varying to random coefficient models, the a_j and b_j will be underlined too, implying that a_j and b_j are also random variables.

Within the traditional fixed effects linear framework the ‘slopes-as-outcomes’ approach can be considered a multilevel analysis approach. This approach is the first step toward modern multilevel modeling. A linear model with individual-level explanatory variables and an individual-level response variable estimates separate parameters within each group, allowing each context to have its own micro model. Three hypothetical situations of varying coefficients analysis are used for comparison:

- A situation with varying intercepts only;
- A situation with varying slopes only; and
- A situation with varying intercepts and varying slopes.

In the first case, the groups’ regression lines are parallel, meaning that the slope of the regression of y on x is equal for each group. But the lines start at different points, showing that the overall mean level for y is different from context to context. Unequal intercepts mean that some groups “score” better than others on the response variable, after the amount of the explanatory variable is taken into account. This situation resembles an ANCOVA solution already discussed previously, where unequal intercepts but equal relationships (or parallel lines) between x and y are assumed.

The second case represents a situation where all regression lines for groups start at the same point, thus having the same intercept. But the regression of y on x

is stronger in some groups, resulting in different slopes. The steeper the slope, the stronger the relationship between x and y.

The third case exhibits a more realistic situation where both intercepts and slopes in the regression model differ. This is an example of where the 'slopes-as-outcomes' approach is most valuable. Each group is allowed to have its own unique solution, which may be a more realistic situation than forcing groups to have some or all features in common.

All three situations show that different intercepts and/or slopes are estimated for each context, representing the first step in the slopes-as-outcomes approach. In subsequent steps parameter estimates for intercepts and slopes are used as response variables in macro-level regressions together with macro-level explanatory variables.

Another name sometimes used for this type of analysis is 'two-step-analysis', because in a first step the individual, or micro-level, parameters are estimated within each context and used in a second step as response variables, predicted by macro-level variable(s). In both steps Ordinary Least Squares (OLS) is the estimation method. The following equations show the second step, which is at the macro level

$$a_j = c_0 + c_1 z_j , \qquad (2.16a)$$

$$b_j = d_0 + d_1 z_j , \qquad (2.16b)$$

where a_j and b_j are the regression coefficients for intercept and slope respectively. The number of observations in each step can be different. In the micro analyses of the first step the number of observations varies for each group. In the macro analyses, with either intercepts or slopes as response variable, the number of observations is equal to the number of groups. m groups produce m different slopes b_j and m different intercepts a_j. The macro equations produce macro intercepts and slopes which are c_0 and d_0 and c_1 and d_1 respectively in (2.16a) and (2.16b). The same equations show the group-level variable z is used to explain the variation among intercepts and slopes. z can either be a global variable or an aggregated variable, as were determined in previous chapter.

The 'slopes-as-outcomes' approach is promising and a potentially good way to find interesting features in the data, features that were previously ignored. But the approach has a practical disadvantage; it requires a separate analysis for each context. Separate analyses for each group may be the best way to represent its group in its uniqueness, but with a large number of groups, this method is hardly feasible, not parsimonious and ignores the fact that groups also may have things in common. The model of the 'slopes-as-outcomes' approach has also some other drawbacks. First, the error structure is not specified correctly, which makes the p-values for the parameter estimates questionable. Secondly, the regression coefficients obtained in the first step are not equally efficient: some may have large standard errors and some small ones. This is not accounted for in the second step. Each coefficient is weighted equally.

An alternative way is the extension of this approach to RC models, which will be discussed in the next section. This approach combines the conceptually interesting features of the 'slopes-as-outcomes' approach with the statistical advantage of parsimony, and the practical advantage of taking into account not only the uniqueness of each group but also what they have in common.

The Random Coefficient (RC) Model

The RC model is conceptually based on the 'slopes-as-outcomes' model. One difference between the two models is that the RC model does not estimate coefficients for each context separately, although each context is allowed to differ from the other contexts in intercept, in slope(s), or in both. A single model is estimated from which the groups are allowed to deviate. From a "graphical" point of view, in the RC approach, a single (solid) regression line is calculated with two other (dashed) lines on either side of it. These two lines capture the variation of the groups from the average line, corresponding with the variance in the 'fixed but varying coefficient' cases in the previous section.

The three situations discussed in the previous chapter are compared with similar ones referring to RC models, in order to show similarities and differences between models. These three situations are again:

- A situation with varying intercepts only;
- A situation with varying slopes only; and
- A situation with varying intercepts and varying slopes.

In the first case where intercepts vary but slopes are the same, this is reflected in a variance around the regression line, which is regular and equal for all values of x.

In the second case, where intercepts do not vary but slopes do, the space around the regression line is not equal for all values of x. This is to be expected in

RC models, because variation in slopes is related to values of x, the explanatory variable. The higher the value of x, the larger the spread around the mean (regression) line.

Finally, in the third case where both slopes and intercepts are different for the groups, the variation around the regression line is produced by the combination of the variance of the slope, the variance of the intercept and the covariance between the two. The variation in slopes is related to values of x, as is the covariance between the variances of intercept and slope. The total variance around the line is the sum of all three (co)variances. As a result, the pattern of the variation of the groups around the average line is irregular, with a minimum and a maximum at certain values of x. If the variation around the average (regression) line, as indicated by the values for the variances, is large, we say that the single line does not represent all groups equally well. Since the regression line is an 'average', we know by the value of the dispersion or variance of the coefficients that some groups are above the line, while others are below it. If, on the other hand, the variances of the intercept and slope are small, the line is close to equal for all groups. A single-level regression analysis would then represent the relationship in this data equally well. The groups can differ either in intercept, in slope, or in both. In RC models each coefficient has its own variance, allowing groups to be unique. Uniqueness for each context is translated into the extent of the deviation of a group from the overall regression line. This deviation (or error) can be used to calculate the posterior means, which are separate values for intercepts and slope(s) for separate contexts, very similar to the 'slope-as-outcomes' approach.

In the previous paragraphs we have described the principles of RC modeling, as well as the differences between RC models and 'slopes-as-outcomes' models. Next we formalize the same principles in equation form. We have shown that the coefficient estimates for separate contexts are represented as varying around the overall regression line. As a result coefficients in RC models consist of two parts: a mean or fixed part, and a variance or random part. The random part is represented by a macro variance, showing the deviation from the overall solution. This variance is referred to as macro-level variance, because the coefficients differ from each other at the macro or context level. The equation of the random model starts with the familiar regression equation, where we underline random variables as before:

$$\underline{y}_{ij} = \underline{a}_j + \underline{b}_j x_{ij} + \underline{\varepsilon}_{ij} . \qquad (2.17)$$

Index i is again used for individuals and index j for groups. $\underline{y}_{ij}$ is the score on the response variable of an observation i within a context j., while x_{ij} is the individual level explanatory variable of the same observation. The variable $\underline{a}_j$ is the random intercept, $\underline{b}_j$ is the random slope, and $\underline{\varepsilon}_{ij}$ is the disturbance term. We assume that $\underline{\varepsilon}_{ij}$ has expectation zero. All $\underline{\varepsilon}_{ij}$ are independent of each other. The variance of $\underline{\varepsilon}_{ij}$ is equal to σ^2.

Note that the underlining of a and b in equation (2.17) is a new feature signifying random coefficients. Observe that this underlying is the only difference between this equation and equation (2.15) for the 'slopes-as-outcomes' model.

The models discussed so far have fixed coefficients. In RC models coefficients can be either fixed or random. The choice between random and fixed coefficients can be made separately for each coefficient in an analysis based on an RC model. Coefficients in RC models are estimated as a main effect with a variance around it. This variance represents the deviation of contexts from that overall or main effect. To specify the properties of the random coefficients, we define them as fixed components plus disturbances. These disturbances are at the group level. They have expectation zero, as usual, and they are independent of the individual-level disturbances $\underline{\varepsilon}_{ij}$. The macro-level equations express the properties of the random slope and intercept in terms of overall population values plus error, as specified in the following macro equations:

$$\underline{a}_j = \gamma_{00} + \underline{u}_{0j}, \qquad (2.18a)$$

$$\underline{b}_j = \gamma_{10} + \underline{u}_{1j}. \qquad (2.18b)$$

The macro-level errors $\underline{u}_{0j}$ and $\underline{u}_{1j}$ in (2.18a) and (2.18b) indicate that both the intercept γ_{00} and slope γ_{10} vary over contexts. The grand mean effect in (2.18a) is γ_{00} while $\underline{u}_{0j}$, the macro-error term, measures the deviation of each context from this overall or grand mean. In the same manner the grand slope estimate across all contexts is γ_{10}, while $\underline{u}_{1j}$ represents the deviation of the slope within each context from the overall slope, as in equation (2.18b). For the gammas the subscript is defined as follows: the fist index is the number of the variable at the micro level, the second represents the number of the variable at the macro level. Hence, γ_{st} is the effect of the macro variable t on the regression coefficient of micro variables s. Zero signifies the intercept, that is to say, the variable with all values equal to +1, either at the micro level or at the macro level. For instance, γ_{00} is the effect of the macro-level intercept on the micro-level coefficient of the intercept. Note that equations (2.18a) and (2.18b) display the model coefficients $\underline{a}_j$ and $\underline{b}_j$ as a function of two components: a fixed component γ_{00} and γ_{10} respectively, and a random component $\underline{u}_{0j}$ and $\underline{u}_{1j}$ respectively, where $\underline{u}_{0j}$ has variance τ_{00}, $\underline{u}_{1j}$ has variance τ_{11}, while $\underline{u}_{0j}$ and $\underline{u}_{1j}$ have covariance τ_{01}.

The elements in the matrix T in equation (2.19) summarize the variance components of an RC model with a random intercept and one random slope and indicate the extra parameters that are estimated in RC models. The τ parameters show the degree to which the groups differ from the overall line.

$$T = \begin{array}{cc} & \begin{array}{cc} \underline{u}_{0j} & \underline{u}_{1j} \end{array} \\ \begin{array}{c} \underline{u}_{0j} \\ \underline{u}_{1j} \end{array} & \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix} \end{array} \qquad (2.19)$$

To show that the separate equations are not really separate, but part of the model, we substitute the separate equations (2.18a) and (2.18b) into equation (2.17) resulting in

$$\underline{y}_{ij} = (\gamma_{00} + \underline{u}_{0j}) + (\gamma_{10} + \underline{u}_{1j})x_{ij} + \underline{\varepsilon}_{ij}. \qquad (2.20)$$

Expanding and rearranging terms yields

$$\underline{y}_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + (\underline{u}_{0j} + \underline{u}_{1j}x_{ij} + \underline{\varepsilon}_{ij}). \qquad (2.21)$$

The rearranging of the terms yields an equation that looks a bit more organized. The fixed effects (gammas) are together and the micro error $\underline{\varepsilon}_{ij}$ and the two macro errors $\underline{u}_{0j}$ and $\underline{u}_{1j}x_{ij}$ are also collected together (in parentheses). The result is a single equation that resembles a traditional regression equation, except for the error terms in parentheses. It is already mentioned that the macro-level variance of the slope (the variance of $\underline{u}_{1j}$) is related to the values of x, as in equation (2.21), the error term, in parentheses, depends on the variable x.

The uniqueness of each context is expressed in these macro errors (the $\underline{u}$ s) which are the deviances from the overall solution. Solutions based on this model no longer produce unique regression lines for each context, such as in the 'slopes-as-outcomes' approach. The result of the RC analysis is a single regression line as an overall solution. Groups fluctuate around this average line. The parameters of the line are the gammas in the above equation, also called the fixed effects. The random effects or macro variances are $\underline{u}_{0j}$ and $\underline{u}_{1j}x_{ij}$. If these variances are significantly different from zero we say that context effects are present.

The next step in the RC approach is to add a new second-level variable to the analysis model, in order to 'explain' the variation in the coefficients for slope and intercept. By adding a macro-level variable z, the variation among groups in general (in the intercepts) or in particular (in the slopes) may disappear. If that works we say that the macro-level variable 'explains' the variation among groups.

As in the 'slopes-as-outcomes' approach, we can choose to model the intercept variance or the slope variance. What we could not do in the 'slopes-as-outcomes' approach was model both variances in the same step. We will show how the model can be extended by fitting macro variances together. All parameters are estimated in a single model, instead of fitting two different macro models as in the 'slopes-as-outcomes' approach.

Out task is to add an explanatory macro-level variable z_j that can account for the explanation of the intercept variance as well as the slope variation among groups. We relate the macro-level variable to the intercept and slope by changing the equations (2.18a) and (2.18b) respectively to:

$$\underline{a}_j = \gamma_{00} + \gamma_{01}z_j + \underline{u}_{0j}, \qquad (2.22a)$$

$$\underline{b}_j = \gamma_{10} + \gamma_{11}z_j + \underline{u}_{1j}. \qquad (2.22b)$$

By fitting this model we assume that intercepts vary as a function of the macro-level explanatory variable z_j plus a random fluctuation, which is represented in the macro-error term $\underline{u}_{0j}$ in (2.22a) and, at the same time, we create and introduce an interaction of the micro-level variable x_{ij} with the macro-level variable z_j. Substituting the new macro equations for the slope (2.22b) and for the intercept (2.22a) into the basic equation (2.17), we produce the single equation:

$$\underline{y}_{ij} = \gamma_{00} + \gamma_{01}z_j + \gamma_{10}x_{ij} + \gamma_{11}x_{ij}z_j + (\underline{u}_{0j} + \underline{u}_{1j}x_{ij} + \underline{\varepsilon}_{ij}). \qquad (2.23)$$

The difference between model (2.23) and (2.21) is in the estimation of two more parameter coefficients, γ_{01} and γ_{11}, while the rest stay the same.

We will discuss in more extent about the estimation methods for all the parameters of the RC model (fixed and random) in the next Chapter of the thesis.

2.3.8 Assumptions and Differences for the Linear Models – A Brief Summary

Table 2.2 summarizes the differences between the models discussed in this Chapter, two traditional linear models, regression and ANCOVA, and two multilevel linear models, 'slopes-as-outcomes' and random coefficients. Most models in Table 2.2 are fixed effects linear models, while the RC model is the only random effects linear model. Within the fixed models the choice is to allow intercepts to be equal (2.24a) or different (2.24b):

$$a_1 = a_2 = \cdots = a_m, \qquad (2.24a)$$

$$a_1 \neq a_2 \neq \cdots \neq a_m. \qquad (2.24b)$$

Equation (2.24a) applies to the total regression model, where group membership is ignored, and all contexts are assumed to have the same effect on individuals. ANCOVA models assume unequal intercepts over contexts, as in equation (2.24b).
Linear models can also differ in what they assume concerning slope coefficients. Slopes can also be assumed to be equal or unequal over contexts. Equal slopes are assumed in the analysis of variance model, where a pooled within slope is estimated, as in equation (2.25a):

$$b_1 = b_2 = \cdots = b_m, \qquad (2.25a)$$

$$b_1 \neq b_2 \neq \cdots \neq b_m. \qquad (2.25b)$$

Random and varying coefficient models allow slopes to differ, as in equation (2.25b).
RC and 'slopes-as-outcomes' model allow researchers to assume that coefficients within contexts vary systematically as a function of the context. Different intercepts together with different slopes can be fitted.

ANCOVA and regression are based on a more restrictive model than the two multilevel models. Multilevel models are more general, because some restrictions are lifted and more parameters are estimated. While more general models allow more freedom than restricted models, they are at the same time less parsimonious.

The equations of RC model presented previously show that this model is an intermediate solution between a totally restricted one, such as standard regression that ignores the context, and a totally unrestricted one, such as the 'slopes-as-outcomes' approach that takes the context too literally. In the 'slopes-as-outcomes' approach all contexts are treated as separate entities as if they have nothing in common, while in the total regression approach contexts are treated as if they are the same and interchangeable. The RC model is also statistically in between the two extremes. The RC model estimates fewer fixed parameters than the 'slopes-as-outcomes' approach, but RC models estimate more parameters than are estimated in the total regression model.

Table 2.2: Assumptions of Traditional Linear Models and Multilevel Models

Model	Intercepts	Slopes
Traditional linear regression	Equal	Equal
ANCOVA	Unequal	Equal
'Slopes-as-outcomes'	Unequal	Unequal
Random Coefficients (RC)	Unequal	Either equal or unequal

2.4 Conclusions of the Chapter

As we can conclude from the thorough discussion of the Chapter, despite the attempts to introduce more sophisticated models in order to analyse hierarchical data (Total or Pooled Regression, Aggregated Regression, Contextual Model, Cronbach Model, ANCOVA, "Slopes-as-Outcomes" Model), all these attempts fail to describe effectively the hierarchy of the data. On the other hand, Multilevel Models seem to be more effective and, therefore, the theoretical aspects of these models will be elaborated more in the following Chapter.

CHAPTER 3

3 THE BASIC MULTILEVEL MODEL AND EXTENSIONS

In the previous Chapter we introduced a number of models and we cleared out the advantages of Multilevel Models in the analysis of hierarchically nested data. First of all, these models "respect" the hierarchy of the data and analyze data simultaneously in all levels. They allow for variables entry in all levels as well as cross-level interactions (interactions of variables measured in different levels). In Random Coefficient Models the lowest level regression coefficient are treated as random variables at the higher level, which explains further the variability of the model.

In this Chapter we first elaborate more on the development of a basic 2-level model. We reconsider alternative ways and notations of setting up and motivating the model and introduce procedures for estimating parameters, forming and testing functions of the parameters and constructing confidence intervals. Then we extend to the natural extensions of the basic 2-level model by introducing higher-level structure, as well as special cases. These are the cross-classified models, the generalized multilevel models for proportion as outcome and the multivariate multilevel model.

The scope of this Chapter is, therefore, to present in extent all theoretical aspects and advantages of a Multilevel Model and to show how this kind of analysis can be effective both in simple hierarchical data problems, as well as in even more complex theoretical statistical data structures.

3.1 The Basic Two-Level Model - The Formulas

3.1.1 The 2-level model and basic notation

We first consider a simple model for one group, relating the response variable to one simple explanatory variable. We write:

$$y_i = \alpha + \beta x_i + e_i \qquad (3.1)$$

where standard interpretations can be given to the intercept (α), slope (β) and residual (e_i). We follow the normal convention of using Greek letters for the regression coefficients and place a circumflex over any coefficient (parameter) which is a sample estimate. This is the formal model and describes a single-level relationship. To describe simultaneously the relationships for several groups we write, for group j

$$y_{ij} = \alpha_j + \beta_j x_{ij} + e_{ij} \qquad (3.2)$$

This is now the formal model where j refers to the level 2 unit and i to the level 1 unit. As it stands, (3.2) is still essentially a single level model, albeit describing a separate relationship for each group. In some situations, for example where there are few groups and interest centres on just those groups in the sample, we may analyze (3.2) by fitting all the $2n+1$ parameters, namely

$$(\alpha_j, \beta_j) \quad j = 1,...,n \text{ and } \sigma_e^2$$

assuming a common 'within-group' residual variance and separate lines for each group.

If we wish to focus not just on these groups, but on a wider 'population' of groups then we need to regard the chosen groups as giving us information about the characteristics of all the groups in the population. Just as we choose random samples of individuals to provide estimates of population means etc., so a randomly chosen sample of groups can provide information about the characteristics of the population of groups. In particular, such a sample can provide estimates of the variation and covariation between groups in the slope and intercept parameters and will allow us to compare groups with different characteristics.

An important class of situations arises when we wish primarily to have information about each individual group in a sample, but where we have a large number of groups so that (3.2) would involve estimating a very large number of parameters. Furthermore, some groups may have rather small numbers of observations and application of (3.2) would result in imprecise estimates. In such cases, if we regard the groups as members of a population and then use our population estimates of the mean and between-group variation, we can utilize this information to obtain more precise estimates for each individual group. This will be discussed later in the section dealing with 'residuals'.

To make (3.2) into a genuine 2-level model we let α_j, β_j become random variables. For consistency of notation replace α_j by β_{0j} and β_j by β_{1j} and assume that

$$\beta_{0j} = \beta_0 + u_{0j} \qquad \text{(3.3a)}$$

$$\beta_{1j} = \beta_1 + u_{1j} \qquad \text{(3.3b)}$$

where u_{0j}, u_{1j} are random variables with parameters

$$E(u_{0j}) = E(u_{1j}) = 0 \qquad \text{(3.4a)}$$

$$\text{var}(u_{0j}) = \sigma_{u0}^2, \ \text{var}(u_{1j}) = \sigma_{u1}^2, \ \text{cov}(u_{0j}, u_{1j}) = \sigma_{u01} \qquad \text{(3.4b)}$$

We can now write (3.2) in the form

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + (u_{0j} + u_{1j} x_{ij} + e_{0ij}) \qquad \text{(3.5)}$$

where

$$\text{var}(e_{0ij}) = \sigma_{e0}^2 \qquad \text{(3.6)}$$

We shall require the extra suffix in the level 1 residual term for models with more complex residual term.

We have expressed the response variable y_{ij} as the sum of a fixed part and a random part within the brackets. We shall also generally write the fixed part of (3.5) in the matrix form

$$E(Y) = X\beta \text{ with } Y = \{y_{ij}\} \qquad \text{(3.7)}$$

$$E(y_{ij}) = X_{ij}\beta = (X\beta)_{ij}, \ X = \{X_{ij}\} \qquad \text{(3.8)}$$

where $\{\}$ denotes a matrix, X is the design matrix for the explanatory variables and X_{ij} is the *ij*-th row of X. For model (3.5) we have $X = \{1 \ \ x_{ij}\}$. Note the alternative representation for the *i*-th row of the fixed part of the model.

The random variables are referred to as 'residuals' and in the case of a single level model the level 1 residual e_{0ij} becomes the usual linear model residual term. To make the model symmetrical so that each coefficient has an associated explanatory variable, we can define a further explanatory variable for the intercept

β_0 and its associated residual, u_{0j}, namely x_{0ij}, which takes the value 1.0. For simplicity this variable may often be omitted.

The feature of (3.5) which distinguishes it from standard linear models of the regression or analysis of variance type is the presence of more than one residual term and this implies that special procedures are required to obtain satisfactory parameter estimates. Note that it is the structure of the random part of the model, which is the key factor. In the fixed part the variables can be measured at any level. We can also include so called 'compositional' variables such as the average value of an explanatory variable for all individuals in each group. The presence of such variables does not alter the estimation procedure, although results will require careful interpretation. We will elaborate more on estimation procedures in the following section.

3.1.2 Parameter estimation for the variance components model

Equation (2.5) requires the estimation of two fixed coefficients, β_0, β_1, and four other parameters, $\sigma^2_{u0}, \sigma^2_{u1}, \sigma_{u01}$ and σ^2_{e0}. We refer to such variances and covariances as *random parameters*. We start, however, by considering the simplest 2-level model, which includes only the random parameters $\sigma^2_{u0}, \sigma^2_{e0}$. It is termed a variance components model because the variance of the response, about the fixed component, the *fixed predictor*, is

$$\mathrm{var}(y_{ij} | \beta_0, \beta_1, x_{ij}) = \mathrm{var}(u_0 + e_{0ij}) = \sigma^2_{u0} + \sigma^2_{e0} \qquad (3.9)$$

that is, the sum of a level 1 and a level 2 variance. This model implies that the total variance for each individual is constant and that the covariance between two individuals (denoted by i_1, i_2) in the same group is given by

$$\mathrm{cov}(u_{0j} + e_{0i_1j}, u_{0j} + e_{i_2j}) = \mathrm{cov}(u_{0j}, u_{0j}) = \sigma^2_{u0} \qquad (3.10)$$

since the level 1 residuals are assumed to be independent. The correlation between two such individuals is therefore

$$\rho = \frac{\sigma^2_{u0}}{(\sigma^2_{u0} + \sigma^2_{e0})} \qquad (3.11)$$

which is referred to as the 'intra-level-2-unit correlation' or the 'intra-class' correlation. This correlation measures the proportion of the total variance which is between-groups. In a model with 3 levels, we will have two such correlations; the 'intra-level-3-unit correlation' and the intra-level-2-unit correlation', and so on.

The existence of a non-zero intra-unit correlation, resulting from the presence of more than one residual term in the model, means that traditional estimation procedures such as 'ordinary least squares' (OLS) which are used for example in multiple regression, are inapplicable. A later section illustrates how the application of OLS techniques leads to incorrect inferences. We now look in more detail at the structure of a 2-level data set, focusing on the covariance structure typified by Figure 3.1.

Figure 3.1: Covariance matrix of three first-level units in a single 2-level context for a variance components model

$$\begin{pmatrix} \sigma_{u0}^2 + \sigma_{e0}^2 & \sigma_{u0}^2 & \sigma_{u0}^2 \\ \sigma_{u0}^2 & \sigma_{u0}^2 + \sigma_{e0}^2 & \sigma_{u0}^2 \\ \sigma_{u0}^2 & \sigma_{u0}^2 & \sigma_{u0}^2 + \sigma_{e0}^2 \end{pmatrix}$$

The matrix in figure 3.1 is the (3 x 3) covariance matrix for the scores of three individuals in a single group, derived from the above expressions. For two groups, one with three individuals and one with two, the overall covariance matrix is shown in Figure 3.2. This 'block-diagonal' structure reflects the fact that the covariance between individuals in different groups is zero, and clearly extends to any number of level 2 units.

Figure 3.2: The block-diagonal covariance matrix for the response vector Y for a 2-level variance components model with two level 2 units

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = \begin{bmatrix} \begin{pmatrix} \sigma_{u0}^2 + \sigma_{e0}^2 & \sigma_{u0}^2 & \sigma_{u0}^2 \\ \sigma_{u0}^2 & \sigma_{u0}^2 + \sigma_{e0}^2 & \sigma_{u0}^2 \\ \sigma_{u0}^2 & \sigma_{u0}^2 & \sigma_{u0}^2 + \sigma_{e0}^2 \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} \sigma_{u0}^2 + \sigma_{e0}^2 & \sigma_{u0}^2 \\ \sigma_{u0}^2 & \sigma_{u0}^2 + \sigma_{e0}^2 \end{pmatrix} \end{bmatrix}$$

A more compact way of presenting this matrix, which we shall use, again is given in figure 3.3.

Figure 3.3: Block-diagonal covariance matrix using general notation

$$V_2 = \begin{bmatrix} \sigma^2_{u0} J_{(3)} + \sigma^2_{e0} I_{(3)} & 0 \\ 0 & \sigma^2_{u0} J_{(2)} + \sigma^2_{e0} I_{(2)} \end{bmatrix}$$

where $I_{(n)}$ is the (n x n) identity matrix and $J_{(n)}$ is the (n x n) matrix of ones. The subscript 2 for V indicates a 2-level model. In single-level OLS models σ^2_{u0} is zero and this covariance matrix then reduces to the standard form $\sigma^2 I$ where σ^2 is the (single) residual variance.

3.1.3 The general 2-level model including random coefficients

We now extend (3.5) in the standard way to include further fixed explanatory variables

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \sum_{h=2}^{p} \beta_h x_{hij} + (u_{0j} + u_{1j} x_{1ij} + e_{0ij}) \qquad (3.12)$$

and more compactly as

$$y_{ij} = X_{ij}\beta + \sum_{h=0}^{1} u_{hj} z_{hij} + e_{0ij} z_{0ij} \qquad (3.13)$$

where we use new explanatory variables for the random part of the model and write these more generally as

$$Z = \{Z_0 \; Z_1\} \qquad (3.14)$$

where $Z_0 = \{1\}$ i.e a vector of 1s and $Z_1 = \{x_{1ij}\}$.

The explanatory variables for the random part of the model are often a subset of those in the fixed part, as here, but this is not necessary. Also, any of the explanatory variables may be measured at any of the levels; for example we may have individual characteristics at level 1 or group characteristics at level 2.

This model (3.13), with the coefficient of X_1 random at level 2, gives rise to the following typical block structure, for a level-two block with two level-one units. The matrix Ω_2 is the covariance matrix of the random intercept and slope at level 2. Note that we need to distinguish carefully between the covariance matrix of the responses given in the following structure and the covariance matrix of the random coefficients. We also refer to the intercept as a random coefficient. The matrix Ω_1 is the covariance matrix for the set of level-one random coefficients; in this case there is just a single variance term at level one. We also write $\Omega = \{\Omega_i\}$ for the set of these covariance matrices. More explicitly:

Figure 3.4: Response covariance matrix for a level 2 unit with two level 1 units for a 2-level model with a random intercept and random regression coefficient at level-2

$$\begin{pmatrix} A & B \\ B & C \end{pmatrix} = \begin{pmatrix} \sigma_{u0}^2 + 2\sigma_{u01}x_{1j} + \sigma_{u1}^2 x_{1j}^2 + \sigma_{e0}^2 & \sigma_{u0}^2 + \sigma_{u01}(x_{1j} + x_{2j}) + \sigma_{u1}^2 x_{1j} x_{2j} \\ \sigma_{u0}^2 + \sigma_{u01}(x_{1j} + x_{2j}) + \sigma_{u1}^2 x_{1j} x_{2j} & \sigma_{u0}^2 + 2\sigma_{u01}x_{2j} + \sigma_{u1}^2 x_{2j}^2 + \sigma_{e0}^2 \end{pmatrix}$$

$$A = \sigma_{u0}^2 + 2\sigma_{u01}x_{1j} + \sigma_{u1}^2 x_{1j}^2 + \sigma_{e0}^2$$

$$B = \sigma_{u0}^2 + \sigma_{u01}(x_{1j} + x_{2j}) + \sigma_{u1}^2 x_{1j} x_{2j}$$

$$C = \sigma_{u0}^2 + 2\sigma_{u01}x_{2j} + \sigma_{u1}^2 x_{2j}^2 + \sigma_{e0}^2$$

giving

$$\begin{pmatrix} A & B \\ B & C \end{pmatrix} = X_j \Omega_2 X_j^T + \begin{pmatrix} \Omega_1 & \\ & \Omega_1 \end{pmatrix}$$

where

$$X_j = \begin{pmatrix} 1 & x_{1j} \\ 1 & x_{2j} \end{pmatrix},$$

$$\Omega_2 = \begin{pmatrix} \sigma_{u0}^2 & \sigma_{u01} \\ \sigma_{u01} & \sigma_{u1}^2 \end{pmatrix},$$

$$\Omega_1 = \sigma_{e0}^2$$

We also see here the general pattern for constructing the response covariance matrix which generalizes both to higher order models and to complex variation at level 1.

3.1.4 Parameter Estimates - Possible Approaches – Algorithms

It is obvious even from the previous discussion that the parameters estimation for both the fixed and the random part of the model is a crucial issue in Multilevel analysis, especially due to the large number of parameters that have to be estimated. For a model of P predictors for the lowest level and Q predictors for the highest level the number of estimates is shown in the following Table (taken by Hox (1995)):

Table 3.1: Number of parameters to be estimated in a "full" Multilevel model

Parameters	Number of Estimates
Intercept	1
Lowest level error variance	1
Slopes for the lowest level predictors	P
Highest level error variances for these slopes	P
Highest level covariances of the intercept with all slopes	P
Highest level covariances between all slopes	P(P-1)/2
Slopes for the highest level predictors	Q
Slopes for cross level interactions	P x Q

Several techniques and principles and their corresponding algorithms have been proposed in order to reach reliable estimates for both fixed and random part. As far as Maximum Likelihood technique is concerned, two different varieties of Maximum Likelihood estimation are used for multilevel regression analysis. One is called Full Maximum Likelihood (FML); in this method both the regression coefficients and the variance components are included in the likelihood function. The other method is called Restricted Maximum Likelihood (REML), here only the variance components are included in the likelihood function. The difference is that FML treats the estimates for the regression coefficients as known quantities when the variance components are estimated, while REML treats them as estimates that carry some amount of uncertainty (Bryk & Raudenbush, 1992). Since REML is more realistic, it should, in theory, lead to better estimates, especially when the number of groups is small (Bryk & Raudenbush, 1992). However, in practice, the differences between the two methods are not very important.

Computing the Maximum Likelihood estimates requires an iterative procedure. The most commons of the algorithms (The Iterative Generalized Least Square Method and the EM algorithm) are discussed in this chapter, as well as other techniques and procedures.

The Iterative Generalized Least Square (IGLS) Method

We now give an overview of the Iterative Generalized Least Squares (IGLS) method which also forms the basis for many of the developments in more complex analysis.

We consider the simple 2-level variance components model

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + e_{0ij} \qquad (3.15)$$

Suppose that we knew the values of the variances, and so could construct immediately the block-diagonal matrix V_2, which we will refer to simply as V. We can then apply immediately the usual Generalized Least Squares (GLS) estimation procedure to obtain the estimator for the fixed coefficients

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y \qquad (3.16)$$

where in this case

$$X = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{n_m m} \end{pmatrix} \qquad Y = \begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n_m m} \end{pmatrix} \qquad (3.17)$$

with m level 2 units and n_j level 1 units in the *j*-th level 2 unit. When the residuals have Normal distributions (3.16) also yields maximum likelihood estimates.

Our estimation procedure is iterative. We would usually start from 'reasonable' estimates of the fixed parameters. Typically these will be those from an initial OLS fit (that is assuming $\sigma_{u0}^2 = 0$), to give the OLS estimates of the fixed coefficients $\hat{\beta}_{(0)}$. From these we form the 'raw' residuals

$$\tilde{y}_{ij} = y_{ij} - \hat{\beta}_0 - \hat{\beta}_1 x_{ij} \qquad (3.18)$$

The vector of raw residuals is written

$$\tilde{Y} = \{\tilde{y}_{ij}\} \qquad (3.19)$$

If we form the cross-product matrix $\tilde{Y}\tilde{Y}^T$ we see that the expected value of this is simply *V*. We can rearrange this cross product matrix as a vector by stacking the columns one on top of the other which is written as $vec(\tilde{Y}\tilde{Y}^T)$ and similarly we can construct the vector $vec(V)$. For the structure given in figure 3.2, these both have $3^2 + 2^2 = 13$ elements. The relationship between these vectors can be expressed as the following linear model

$$\begin{pmatrix} \tilde{y}_{11}^2 \\ \tilde{y}_{21}\tilde{y}_{11} \\ \vdots \\ \tilde{y}_{22}^2 \end{pmatrix} = \begin{pmatrix} \sigma_{u0}^2 + \sigma_{e0}^2 \\ \sigma_{u0}^2 \\ \vdots \\ \sigma_{u0}^2 + \sigma_{e0}^2 \end{pmatrix} + R = \sigma_{u0}^2 \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \sigma_{e0}^2 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 1 \end{pmatrix} + R \qquad (3.20)$$

where R is a residual vector. The left hand side of (3.20) is the response vector in the linear model and the right hand side contains two explanatory variables, with coefficients σ_{u0}^2, σ_{e0}^2 which are to be estimated. The estimation involves an application of GLS using the estimated covariance matrix of $vec(\tilde{Y}\tilde{Y}^T)$, assuming Normality, namely $2(V^{-1} \otimes V^{-1})$ where $\otimes$ is the Kronecker product. The Normality assumption allows us to express this covariance matrix as a function of the random parameters. Even if the Normality assumption fails to hold, the resulting estimates are still consistent, although not fully efficient, but standard errors, estimated using the Normality assumption and, for example confidence intervals will generally not be consistent. For certain variance component models alternative distributional assumptions have been studied, especially for discrete response models of the kind discussed later in the thesis and maximum likelihood estimates obtained. For more general models, however, with several random coefficients, the assumption of multivariate Normality is a flexible one, which allows a convenient parameterization for complex covariance structures at several levels.

With the estimates obtained from applying GLS to (3.20) we return to (3.16) to obtain new estimates of the fixed effects and so alternate between the random and fixed parameter estimation until the procedure converges, that is the estimates for all the parameters do not change from one cycle to the next. At convergence, assuming multivariate Normality, the estimates are maximum likelihood. Essentially the same procedure can be used for the more complicated models discussed later on in the thesis. The maximum likelihood procedure produces biased estimates of the random parameters because it takes no account of the sampling variation of the fixed parameters. This may be important in small samples. Goldstein (1989a) shows how a simple modification leads to restricted iterative generalized least squares (RIGLS) or restricted maximum likelihood (REML) estimates which are unbiased. The IGLS algorithm is readily modified to produce these restricted estimates (RIGLS)

Full details of efficient computational procedures for carrying out all these calculations are given by Goldstein & Rasbash (1992).

The EM algorithm

To illustrate the procedure, consider the 2-level variance components model

$$y_{ij} = (X\beta)_{ij} + u_j + e_{ij}, \qquad \mathrm{var}(e_{ij}) = \sigma^2_{\mathrm{e}}, \quad \mathrm{var}(\mathrm{u}_{\mathrm{j}}) = \sigma^2_u \qquad (3.21)$$

The vector of level 2 residuals is treated as missing data and the 'complete' data therefore consists of the observed vector Y and the u_j treated as observations. The joint distribution of these, assuming Normality, and using our standard notation is

$$\begin{bmatrix} Y \\ u \end{bmatrix} = N\left\{ \begin{bmatrix} X\beta \\ 0 \end{bmatrix}, \begin{bmatrix} V & J^T\sigma^2_u \\ \sigma^2_u J & \sigma^2_u I \end{bmatrix} \right\} \qquad (3.22)$$

This generalizes readily to the case where there are several random coefficients. If we denote these by β_j we note that some of them may have zero variances. We can now derive the distribution of $\beta_j|Y$, and we can also write down the Normal log likelihood function for (3.22) with a general set of random coefficients, namely

$$\log(L) \propto -N\log(\sigma_e{}^2) - J\log|\Omega| - \sigma_e{}^{-2}\sum_{ij} e_{ij}^2 - \sum_j \beta_j^T \Omega_u{}^{-1}\beta_j \quad (3.23)$$

Maximizing the latter for the random parameters we obtain

$$\hat{\sigma}_e^2 = N^{-1}\sum_{ij} e_{ij}^2 \quad (3.24)$$

$$\hat{\Omega}_u = m^{-1}\sum_j \beta_j \beta_j^T \quad (3.25)$$

where m is the number of level 2 units. We do not know the values of the individual random variables. We require the expected values, conditional on the Y and the current parameters, of the terms under the summation signs, these being the sufficient statistics. We then substitute these expected values in (3.24) and (3.25) for the updated random parameters. These conditional values are based upon the 'shrunken' predicted values and their (conditional) covariance matrix. With these updated values of the random parameters we can form V and hence obtain the updated estimates for the fixed parameters using generalized least squares. We note that the expected values of the sufficient statistics can be obtained using the general result for a random parameter vector θ.

$$E(\theta\theta^T) = \mathrm{cov}(\theta) + [E(\theta)][E(\theta)]^T \quad (3.26)$$

The prediction is known as the E (expectation) step of the algorithm and the computations in (3.25) and (3.26) the M (maximization) step. Given starting values, based upon OLS, these computations are iterated until convergence is obtained. Convenient computational formulae for computing these quantities at each iteration can be found in Bryk & Raudenbush (1992).

Markov Chain Monte Carlo estimation – The Gibbs Sampling

Markov Chain Monte Carlo algorithms exploit the properties of Markov chains where the probability of an event is conditionally dependent on a previous state. The procedure is iterative and at each stage from the full multivariate distribution the distribution of each component conditional on the remaining components is computed and used to generate a random variable. The components may be variates, regression coefficients, covariance matrices etc. After a suitable number of iterations, we obtain a sample of values from the distribution of any component, which we can then use to derive any desired characteristic such as the mean, covariance matrix, etc. The most common procedure is that of Gibbs Sampling and Gilks et al. (1993) provide a comprehensive discussion with applications and an application to a 2-level logit model is given by Zeger & Karim (1991). It allows the fitting of Bayesian models where prior distributions for the parameters are specified.

We outline a Gibbs Sampling procedure for a 2-level model. We write:

$$Y = X\beta + Z^{(2)}u + Z^{(1)}e \qquad (3.27)$$

We first consider the distribution $\beta|u^{(k)},Y$ where k refers to the *k*-th iteration.

Given $u^{(k)}$, $Z^{(2)}u$ is just an offset so that we can regress y_{ij} on x_{ij} to estimate $\hat{\beta}^{(k)}$ and $\mathrm{var}(\hat{\beta}^{(k)})$.

We can then select a random vector from this distribution, assumed to be multivariate normal $(\hat{\beta}^{(k)}, \mathrm{var}(\hat{\beta}^{(k)}))$.

We now consider the distribution of $\Omega_2|u^{(k)}$. We have (with a non-informative prior) that the (posterior) distribution of Ω_2^{-1} is a Wishart distribution with parameter (i.e. covariance) matrix

$$S^{(k)} = \sum_{j=1}^{J} u_j^{(k)} u_j^{(k)^T} \quad with \;\; d = J - q + 1 \text{ d.f.} \qquad (3.28)$$

where J is the number of level 2 units and q is the number of random coefficients.

A simple way of generating such a Wishart distribution is to generate *d* multivariate normal vectors from $N(0, S^{(k)})$ and form their SSP matrix. This provides $\hat{\Omega}_2^{(k)}$.

Finally we consider the distribution $u_j|\beta,\Omega_2,Y$. These are the usual level 2 residuals, for which we have standard expressions for their expected values and covariance matrix. We note that for a 2-level model (but not within a three level model) these are block-independent. Assuming Normality we can now generate a set of $u_j^{(k)}$ and this completes an iterative cycle.

There are some particular computational details to be noted. For example 'rejection sampling' at each cycle can be used and we can do several cycles for Ω_2, u_j for each β since the former tend to have higher autocorrelations across cycles.

The procedure can be applied to any existing models, e.g. logit models, where the conditional distributional assumptions are explicit. Gibbs Sampling tends to be computationally demanding, with hundreds if not thousands of iterations required and this can be particularly burdensome when several different models are being explored for their fit to the data. On the other hand, this approach has the advantage, in small samples, that it takes account of the uncertainty associated with the estimates of the random parameters and can provide exact measures of uncertainty. The maximum likelihood methods tend to overestimate precision because they ignore this uncertainty. In small samples this will be important especially when obtaining 'posterior' estimates for residuals, which will be discussed in the following section. Gibbs sampling approach is therefore useful for small and moderate sized samples and when used in conjunction with likelihood based EM or IGLS algorithms.

Other estimation procedures

A variation on IGLS is Expected Generalized Least Squares (EGLS) or the “Gauss-Newton method” as it is mentioned by other authors (Kreft & Leeuw, 1998). This focuses interest on the fixed part parameters and uses the estimate of V obtained after the first iteration merely to obtain a consistent estimator of the fixed part coefficients without further iterations. A variant of this separates the level 1 variance from V as a parameter to be estimated iteratively along with the fixed part coefficients.

Longford (1987) developed a procedure based upon a 'Fisher scoring' algorithm which can be seen that it is formally equivalent to IGLS. This algorithm can also incorporate certain extensions, for example to handle discrete response data.

We have already mentioned the full Bayesian approach, which has become computationally feasible with the development of 'Markov Chain Monte Carlo' (MCMC) methods, especially Gibbs Sampling (Zeger & Karim, 1991). An alternative to the full Bayes estimation, known as 'Empirical Bayes’, ignores the prior distributions of the random parameters, treating them as known for purposes of inference. When Normality is assumed, these estimates are the same as IGLS or RIGLS.

Another approach, which parallels all that was mentioned so far, is that of Generalized Estimating Equations (GEE) introduced by Liang & Zeger (1986). The principal difference is that GEE obtains the estimate of V using simple regression or 'moment' procedures based upon functions of the actual calculated raw residuals. It is concerned principally with modeling the fixed coefficients rather than exploring the structure of the random component of the model. While the resulting coefficient estimates are consistent they are not fully efficient. In some circumstances, however, GEE coefficient estimates may be preferable, since they will usually be quicker to obtain and they make weaker assumptions about the structure of *V*. The GEE procedure can be extended to handle most of the models dealt with more complex cases.

3.1.5 Estimating the residuals

In a single level model such as (3.1) the usual estimate of the single residual term e_i is just $\tilde{y}_i$ the raw residual. In a multilevel model, however, we shall generally have several residuals at different levels. In this chapter we consider estimating the individual residuals in all levels.

Given the parameter estimates, consider predicting a specific residual, say u_{0j} in a 2-level variance components model. Specifically we require for each level 2 unit

$$\hat{u}_{0j} = E(u_{0j}|Y,\hat{\beta},\hat{\Omega}) \qquad (3.29)$$

We shall refer to these as estimated or predicted residuals or, using Bayesian terminology, as posterior residual estimates. If we ignore the sampling variation attached to the parameter estimates in (3.29) we have

$$\text{cov}(\tilde{y}_{ij}, u_{0j}) = \text{var}(u_{0j}) = \sigma_{u0}^2 \qquad (3.30a)$$

$$\text{cov}(\tilde{y}_{ij}, e_{0ij}) = \sigma_{e0}^2 \qquad (3.30b)$$

$$\text{var}(\tilde{y}_{ij}) = \sigma_{u0}^2 + \sigma_{e0}^2 \qquad (3.30c)$$

We regard (3.29) as a linear regression of u_{0j} on the set of $\{\tilde{y}_{ij}\}$ for the *j*-th level 2 unit and (3.13) defines the quantities required to estimate the regression coefficients and hence $\hat{u}_{0j}$. For the variance components model we obtain

$$\hat{u}_{0j} = \frac{n_j\sigma_u^2}{(n_j\sigma_u^2 + \sigma_{e0}^2)}\tilde{y}_j \qquad (3.31a)$$

$$\tilde{e}_{0ij} = \tilde{y}_{ij} - \hat{u}_{0j} \qquad (3.31b)$$

$$\tilde{y}_j = (\sum_i \tilde{y}_{ij})/n_j \qquad (3.31c)$$

where n_j is the number of level 1 units in the *j*-th level 2 unit. The residual estimates are not, unconditionally, unbiased but they are consistent. The factor multiplying the mean ($\tilde{y}_j$) of the raw residuals for the *j*-th unit is often referred to as a 'shrinkage factor' since it is always less than or equal to one. As n_j increases this factor tends to one, and as the number of level 1 units in a level 2 unit decreases the 'shrinkage estimator' of u_{0j} becomes closer to zero. In many applications the higher level residuals are of interest in their own right and the increased shrinkage for a small level 2 unit can be regarded as expressing the relative lack of information in the unit so that the best estimate places the predicted residual close to the overall population value as given by the fixed part.

These residuals therefore can have two roles. Their basic interpretation is as random variables with a distribution whose parameter values tell us about the variation among the level 2 units, and which provide efficient estimates for the fixed coefficients. A second interpretation is as individual estimates for each level 2 unit where we use the assumption that they belong to a population of units to predict their values. In particular, for units which have only a few level 1 units, we can obtain more precise estimates than if we were to ignore the population membership assumption and use only the information from those units. This becomes especially important for estimates of residuals for random coefficients, where in the extreme case of only one level-one unit in a level-two unit we lack information to form an independent estimate.

As in single level models we can use the estimated residuals to help check on the assumptions of the model. The two particular assumptions that can be studied readily are the assumption of Normality and that the variances in the model are constant. Because the variances of the residual estimates depends in general on the values of the fixed coefficients it is common to standardize the residuals by dividing by the appropriate standard errors, which are referred as 'diagnostic' or 'unconditional' standard errors (Goldstein, 1995).

When the residuals at higher levels are of interest in their own right, we need to be able to provide interval estimates and significance tests as well as point estimates for them or functions of them. For these purposes we require estimates of the standard errors (the so-called 'conditional' or 'comparative' standard errors) of the estimated residuals, where the sample estimate is viewed as a random realization from repeated sampling of the same higher-level units whose unknown true values are of interest.

The level 1 residuals are generally not of interest in their own right but are used rather for model checking, having first been standardized using the diagnostic standard errors. Checking the model assumptions in a multilevel model are used in an exactly analogous way as in simple regression models. In other words, we use plot of the standardized level 1 residuals against the fixed part predicted value to check the assumption of a constant level 1 variance ('homoscedasticity') and Normal score plots for level-one (and level-two) residuals to check the assumption of Normality.

3.1.6 Hypothesis testing and confidence intervals

In this section we deal with large sample procedures for constructing interval estimates for parameters or linear functions of parameters and for hypothesis testing. Hypothesis tests are used sparingly in multilevel analysis since the usual form of a null hypothesis, that a parameter value or a function of parameter values is zero, is usually implausible and also relatively uninteresting. Moreover, with large enough samples a null hypothesis will almost certainly be rejected. The exception to this is where we are interested in whether a difference is positive or negative, and this is discussed in the section on residuals below. Confidence intervals emphasize the uncertainty surrounding the parameter estimates and the importance of their substantive significance.

Fixed parameters

We have already presented parameter estimates techniques for the fixed part parameters together with their standard errors. These are adequate for hypothesis testing or confidence interval construction separately for each parameter. In many cases, however, we are interested in combinations of parameters. For hypothesis testing, this most often arises for grouped or categorized explanatory variables where n group effects are defined in terms of $n-1$ dummy variable contrasts and we wish simultaneously to test whether these contrasts are zero. We may also be interested in providing a pair of confidence intervals for the parameter estimates. We proceed as follows:

Define a *(r x p)* contrast matrix ***C***. This is used to form linearly independent functions of the *p* fixed parameters in the model of the form $f = C\beta$, so that each row of ***C*** defines a particular linear function. Parameters that are not involved have the corresponding elements set to zero. Suppose we wish to test the hypothesis that the coefficients of two variables each having two categories are jointly zero. We define

$$C = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad f = \begin{pmatrix} \beta_2 \\ \beta_3 \end{pmatrix}$$

and the general null hypothesis is

$$H_0: f = k, \quad k = \{0\} \text{ here}$$

We form

$$R = (\hat{f} - k)^T [C(X^T \hat{V}^{-1} X)^{-1} C^T]^{-1} (\hat{f} - k) \qquad (3.32a)$$

$$\hat{f} = C\hat{\beta} \qquad (3.32b)$$

If the null hypothesis is true this is distributed as approximately χ^2 with *r* degrees of freedom. Note that the term $(X^T \hat{V}^{-1} X)^{-1}$ is the estimated covariance matrix of the fixed coefficients.

If we find a statistically significant result we may wish to explore which particular linear combinations of the coefficients involved are significantly different from zero. The common instance of this is where we find that *n* groups differ and we wish to carry out all possible pairwise comparisons. A simultaneous comparisons procedure which maintains the overall type I error at the specified level involves carrying out the above procedure with either a subset of the rows of *C* or a set of (less than *r*) linearly independent contrasts. The value of *R* obtained is then judged against the critical values of the chi-squared distribution with *r* degrees of freedom.

We can also obtain an α% confidence region for the parameters by setting $\hat{R}$ equal to the α% tail region of the χ^2 distribution with *r* degrees of freedom in the expression

$$\hat{R} = (f - \hat{f})^T [C(X^T \hat{V}^{-1} X)^{-1} C^T]^{-1} (f - \hat{f}) \qquad (3.33)$$

This yields a quadratic function of the estimated coefficients, giving an *r*-dimensional ellipsoidal region.

In some situations we may be interested in separate confidence intervals for all possible linear functions involving a subset of ***q*** parameters or ***q*** linearly independent functions of the parameters, while maintaining a fixed probability that all the intervals include the population value of these functions of the parameters. As before, this may arise when we have an explanatory variable with several categories and we are interested in intervals for sets of contrasts. For a $(1-\alpha)\%$ interval write C_i for the *i*-th row of *C*, then a simultaneous $(1-\alpha)\%$ interval for $C_i\beta$, for all C_i is given by

$$(C_i\hat{\beta} - d_i, C_i\hat{\beta} + d_i) \qquad (3.34)$$

where

$$d_i = [C_i (X^T \hat{V}^{-1} X)^{-1} C_i^T \chi^2_{q,(\alpha)}]^{0.5} \qquad (3.35)$$

where $\chi^2_{q,(\alpha)}$ is the $\alpha\%$ point of the χ^2_q distribution.

We can also use the likelihood ratio test criterion for testing hypotheses about the fixed parameters, although generally the results will be similar. The difference arises because the random parameter estimates used in (3.32a) and (3.32b) are those obtained for the full model rather than those under the null hypothesis assumption, although this modification can easily be made. We shall discuss the likelihood ratio test in the next section dealing with the random parameters.

Random parameters

In very large samples it is possible to use the same procedures for hypothesis testing and confidence intervals as for the fixed parameters. Generally, however, procedures based upon the likelihood statistic are preferable. To test a null hypothesis H_0 against an alternative H_1 involving the fitting of additional parameters we form the log likelihood ratio or deviance statistic

$$D_{01} = -2\log_e(\lambda_0 / \lambda_1) \qquad (3.36)$$

where λ_0, λ_1 are the likelihoods for the null and alternative hypotheses and this is referred to tables of the chi-squared distribution with degrees of freedom equal to the difference (q) in the number of parameters fitted under the two models.

We can also use (3.36) as the basis for constructing a $(1-\alpha)\%$ confidence region for the additional parameters. If D_{01} is set to the value of the $\alpha\%$ point of the chi squared distribution with q degrees of freedom, then a region is constructed to satisfy (3.36), using a suitable search procedure. This is a computationally intensive task, however, since all the parameter estimates are recomputed for each search point.

An alternative is to use the 'profile likelihood' (McCullagh & Nelder, 1989). In this case the likelihood is computed for a suitable region containing values of the random parameters of interest, for fixed values of the remaining random parameters. Interval estimates can be provided also by bootstrap simulations.

Residuals

In studies of institutions (e.g. schools, hospitals etc) effectiveness (Goldstein & Spiegelhalter, 1996), one requirement is sometimes to try to identify institutions with residuals which are substantially different. From a significance testing standpoint, we will often be interested in the null hypothesis that institution (group) A has a smaller residual than institution (group) B against the alternative that the residual for institution (group) A is larger than that for institution (group) B (ignoring the vanishingly small probability that they are equal). In the case when a standard significance test accepts the alternative hypothesis (at a chosen level) of some difference against the null hypothesis of no difference, this is equivalent to accepting one of the alternatives (A > B, A < B) at the same level of significance and we shall use this interpretation.

Where we can identify two particular institutions (groups) then it is straightforward to construct a confidence interval for their difference or carry out a significance test. Often, however, the results are made available to a number of individuals, each of whom are interested in comparing their own institutions (e.g. schools) of interest. This may occur, for example where policy makers wish to select a few schools within a small geographical area for comparison, out of a much larger study. In the following discussion, we suppose that individuals wish to compare only

pairs of institutions, although the procedure can be extended to multiple comparisons of three or more residuals. Further details are given by Goldstein & Healy (1994).
When the sample size of a study is fairly large, we can assume that the estimated residuals together with their comparative standard errors estimates are uncorrelated.

First, we order the residuals from smallest to largest. We construct an interval about each residual so that the criterion for judging statistical significance at the $(1-\alpha)\%$ level for any pair of residuals is whether their confidence intervals overlap. For example, if we consider a pair of residuals with a common standard error (se) and assuming Normality, the confidence interval width for judging a difference significant at the 5% level are given by $\pm 1.39(\text{se})$. The general procedure defines a set of confidence intervals for each residual *i* as

$$\hat{u}_i \pm c(\text{se})_i \qquad (3.37)$$

For each possible pair of intervals, (3.37) there is a significance level associated with the overlap criterion, and the value ***c*** is determined so that the average, over all possible pairs is $(1-\alpha)\%$. A search procedure can be devised to determine ***c***. When the ratios of the standard errors do not vary appreciably, say by not more than 2:1, the value 1.4 can be used for ***c***. As this ratio increases so does the value of ***c***.

These kinds of residual analyses are useful for conveying the inherent uncertainty associated with estimates for individual level 2 (or higher) units, where the number of level 1 units per higher-level unit is not large. This uncertainty in turn places inherent limitations upon such comparisons.

3.2 Extensions of the 2-Level Linear Model

What we have discussed so far refers to notations, techniques and estimations for the two-level linear model, which is the most common case in the multilevel analysis theory. However, in order to examine more demanded applications presented in the next chapter, we need to present the logical extensions of the two-level linear model. In all cases discussed here, the extensions are straightforward and stem either from the hierarchy of the subjects or from the nature of the data that are being measured. The extensions discussed here are:

- The 3-Level linear Model
- Cross-Classification Models
- Models for Discrete response data – The Proportions as responses case

- Multivariate Multilevel Models – The basic 2-level Multivariate model
- Multilevel Structural Equation Models – Multilevel Factor Analysis case

3.2.1 The Three-Level Linear Model

The most profound, maybe, extension of a 2-Level linear model comes when we add more levels of hierarchy in the model. We focus on the 3-level model since higher-level cases are rarely of importance in practice. Some examples of 3-level hierarchical structures are students (Level-1) nested within schools (Level-2) nested within prefectures (Level-3). Or in another point of view repeated visits (Level-1) of patients (Level-2) in heath provider units (Level-3).

In the simplest case the basic linear 3-level model can be written as follows:

$$y_{ijk} = \beta x_{ijk} + (v_k + u_{jk} + e_{ijk}) \qquad (3.38)$$

where x_{ijk} is a vector of covariates and $\boldsymbol{\beta}$ a corresponding vector of parameter estimates. The vector of covariates includes a constant together with explanatory variables measured at any of the three levels. The error terms v_k, u_{jk} and e_{ijk} are considered are considered as random variables with mean zero and variances

$$\text{var}(v_k) = \sigma_v^2 \quad (3.39a)$$

$$\text{var}(u_{jk}) = \sigma_u^2 \quad (3.39b)$$

$$\text{var}(e_{ijk}) = \sigma_e^2 \quad (3.39c)$$

If we now introduce Z explanatory variables in the random part of the model, in any of the three levels, we obtain the more general form of the 3-level model, as follows:

$$y_{ijk} = X_{ijk}\beta + \sum_{h=0}^{q_3} v_{hk} z_{hk}^{(3)} + \sum_{h=0}^{q_2} u_{hjk} z_{hjk}^{(2)} + \sum_{h=0}^{q_1} e_{hijk} z_{hijk}^{(1)} \qquad (3.40)$$

where x_{ijk} is again the vector of covariates, $\boldsymbol{\beta}$ the corresponding vector of parameter estimates, and $z_{hk}^{(3)}, z_{hjk}^{(2)}$ and $z_{hijk}^{(1)}$ the explanatory variables of the random part of the 3rd, 2nd and 1st level of hierarchy, respectively.

Although such models seem more complicated and demanding than the two-level models, the computations, estimation techniques and algorithms are totally analogous to the methods described before for the two-levels case.

3.2.2 Cross-Classified Models

So far we have considered only data where the units have a purely hierarchical or nested structure. In many cases, however, a unit may be classified along more than one dimension. An example is students classified both by the school they attend and by the neighbourhood where they live. This is diagrammatically represented as follows for three schools and four neighbourhoods with between one and six students per school/neighbourhood cell. The cross classification is at level 2 with students at level 1.

Table 3.2: A random cross-classification at level 2

	School 1	School 2	School 3
Neighbourhood 1	x x x x	x x	x
Neighbourhood 2	x	x x x x x x	x x x
Neighbourhood 3	x x	x	x x x x
Neighbourhood 4	x x x	x x	x x

Another example is in a repeated measures study where children are measured by different raters at different occasions. If each child has its own set of raters not shared with other children then the cross classification is at level 1, occasions by raters, nested within children at level 2. We note that, by definition, a level 1 cross classification has only one unit per cell.

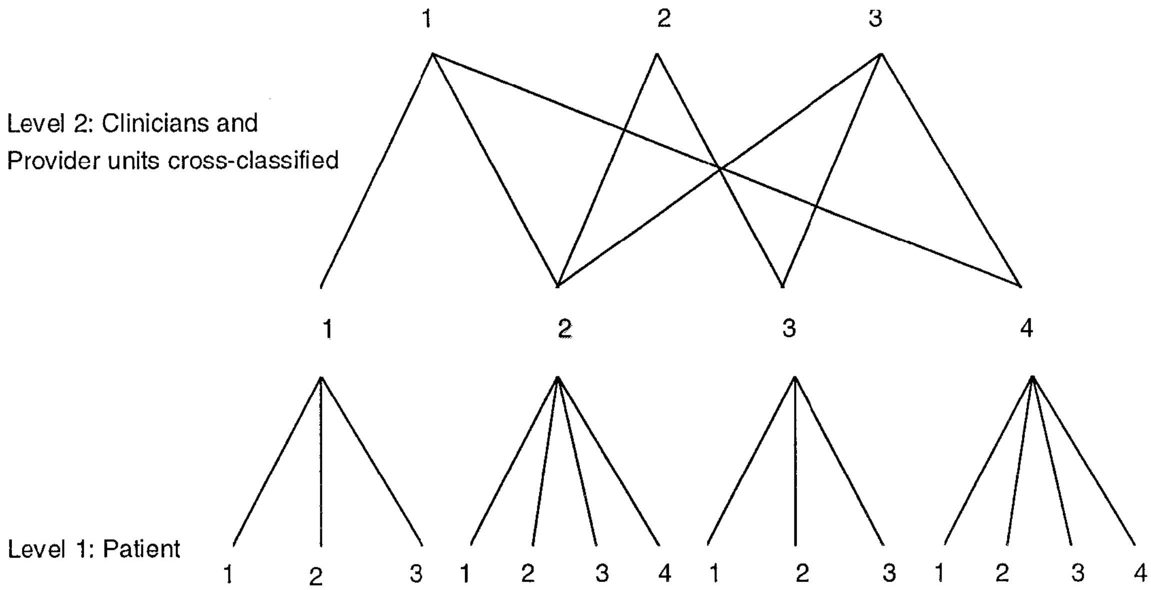
These basic cross-classifications occur commonly when a simple hierarchical structure breaks down in practice. Consider, for example, a repeated measures design, which follows a sample of students over time, say once a year, within a set of classes for a single school. If students change classes during the course, that is a cross classification at level 2 for classes by students. If we now include schools these will be classified as level 3 units, but if students also change schools during the course of the study then we obtain a level 3 cross classification of students by schools with classes nested at level 2 within schools and occasions as the level 1 units. The students have moved from being crossed with classes to being crossed with schools. Note that since students are crossed at level 3 with schools they are also automatically crossed with any units nested within schools and we do not need separately to specify the crossing of classes with students.

Such designs will occur also in panel or longitudinal studies of individuals who move from one locality to another, or workers who change their place of employment.

Other examples of such designs occur in panel studies of households where, over time, some households split up and form new households. The total set of all households is crossed with individual at level 2 with occasion at level 1. The households, which remain intact for more than one occasion, provide the information for estimating level 1 variation.

In health studies cross-classification occurs naturally in many cases. Consider for example the case where patients may be classified both by the hospitals they visit and by the clinicians the frequent, so that individuals within one hospital cluster are not grouped in the same way under clinicians. This type of cross-classification does not occur when clinicians operate within a single medical care, but this is not always the case. This kind of cross-classification is illustrated diagramatically in the following figure (Rice & Jones, 1997):

Figure 3.5: Patients within Cross-classified clinicians and Provider Units



In another example, patients may receive care from more that one medical centre during the year. This arrangement forms a multiple or cross-unit membership model, a special case of cross-classification (Carey, 2000).

We now set out the structure of the basic models described above and then go on to consider extensions and special cases of interest.

A basic cross-classified model

We consider first the simple model of Table 3.2 with variance components at level 2 and a single variance term at level 1.
We shall refer to the two classifications at level 2 using the subscripts j_1, j_2 and in general parentheses will group classifications at the same level. We write the model as

$$y_{i(j_1 j_2)} = X_{i(j_1 j_2)}\beta + u_{j_1} + u_{j_2} + e_{i(j_1 j_2)} \qquad (3.41)$$

The covariance structure at level 2 can be written in the following form

$$\operatorname{cov}(y_{i(j_1 j_2)} y_{i'(j_1 j'_2)}) = \sigma^2_{u_1} \qquad (3.42a)$$

$$\operatorname{cov}(y_{i(j_1 j_2)} y_{i'(j'_1 j_2)}) = \sigma^2_{u_2} \qquad (3.42b)$$

$$\operatorname{var}(y_{i(j_1 j_2)}) = \operatorname{cov}(y_{i(j_1 j_2)} y_{i'(j_1 j_2)}) = \sigma^2_{u_1} + \sigma^2_{u_2} \quad (3.42c)$$

Note that if there is no more than one unit per cell, then model (3.41) is still valid and can be used to specify a level 1 cross classification.

Thus the level-2 variance is the sum of the separate classification variances, the covariance for two level 1 units in the same classification is equal to the variance for that classification and the covariance for two level 1 units, which do not share either classification, is zero. If we have a model where random coefficients are included for either or both classifications, then analogous structures are obtained. We can also add further ways of classification with obvious extensions to the covariance structure.

We can now show how cross-classified models can be specified and estimated efficiently using a purely hierarchical formulation, including random cross-classified structures.

We illustrate the procedure using a 2-level model with crossing at level 2. The 2-level cross-classified model, using the same notations as in previous chapters for the basic model, can be written

$$y_{i(j_1 j_2)} = X_{i(j_1 j_2)}\beta + \sum_{h=1}^{q_1} z_{1hij_1} u_{1hj_1} + \sum_{h=1}^{q_2} z_{2hij_2} u_{2hj_2} + e_{i(j_1 j_2)} \qquad (3.43)$$

Parentheses group the ways of classification at each level. We have two sets of explanatory variables, type 1 and type 2, for the random components defined by the columns of $Z_1(n \times p_1 q_1)$, $Z_2(n \times p_2 q_2)$ where p_1, p_2 are respectively the number of categories of each classification, i.e.

$$Z_1 = \{z_{1hij_1}\} \qquad (3.44a)$$

where $z_{1hij_1} = z_{1him} \quad if \quad j_1 = m, \; for \;\; m - th\, type\, 1\, level\, 2\, unit, 0\, otherwise$

and

$$Z_2 = \{z_{2hij_2}\} \qquad (3.44b)$$

where $z_{2hij_2} = z_{2him} \quad if \quad j_2 = m, \; for \;\; m - th\, type\, 2\, level\, 2\, unit, 0\, otherwise$

These variables are dummy variables where for each level 2 unit of type 1 we have q_1 random coefficients with covariance matrix $\Omega_{(1)2}$ and likewise for the type 2 units. To simplify the exposition we restrict ourselves to the variance component case where we have

$$\Omega_{(1)2} = \sigma^2_{(1)2}, \;\; \Omega_{(2)2} = \sigma^2_{(2)2} \qquad (3.45a)$$

$$E(\tilde{Y}\tilde{Y}^T) = V_1 + Z_1(\sigma^2_{(1)2} I_{(p_1)}) Z_1^T + Z_2(\sigma^2_{(2)2} I_{(p_2)}) Z_2^T \quad (3.45b)$$

It is clear that the second term in (3.45b) can be written as

$$Z_1(\sigma^2_{(1)2} I_{(p_1)}) Z_1^T = J \sigma^2_{(1)2} J^T \qquad (3.46)$$

where J is a (n x 1) vector of ones. The third term is of the general form $Z_3 \Omega_3 Z_3^T$, namely a level 3 contribution where in this case there is only a single level 3 unit and with no covariances between the random coefficients of the Z_{2h} and with the variance terms constrained to be equal to a single value, $\sigma^2_{(2)2}$.

More generally we can specify a level 2 cross classified variance components model by modeling one of the classifications as a standard hierarchical component and the second as a set of dummy explanatory variables, one for each category, with the random coefficients uncorrelated and with variances constrained to be equal. We can summarize the procedure using the simple model of (3.41). We specify one of the classifications, most efficiently the one with the larger number of units, as a standard hierarchical level 2 classification. For the other classification we define a dummy (0,1) variable for each unit, which is one if the observation belongs to that unit and zero if not. Then we specify that each of these dummy variables has a coefficient random at level 3 and in addition constrain the resulting set of level 3 variances to be equal. The variance estimate obtained is that required for this classification and the level-2 variance for the other classification is the one we require for that.

To extend this to further ways of classification we add levels. Thus, for a three way cross classification at level 2 we choose one classification, typically that with the largest number of categories, to model in standard hierarchical fashion at level 2, the second to model with coefficients random at level 3 as above and the third to model in a similar fashion with coefficients random at level 4. So we can obtain the third variance by defining a similar set of dummy variables with coefficients varying at level 4 and variances constrained to be equal. This procedure generalizes straightforwardly to sets of several random coefficients for each classification, with dummy variables defined as the products of the basic (0,1) dummy variables used in the variance components case and with corresponding variances and covariances constrained to be equal within classifications. In general a p-way cross classification at any level can be modeled by inserting sets of random variables at the next p-1 higher levels. Thus in a 2-level model with two crossed classifications at level 1 we would obtain a three level model with the original level 2 at level 3 and the level 1 cross classifications occupying levels 1 and 2.

Interactions in cross-classifications

If the second (type-2) classification has further explanatory variables with random coefficients as in (3.43) then we form extended dummy variable 'interactions' as the product of the basic dummy variables and the further explanatory variables with random coefficients, so that these coefficients have variances and covariances within the same type-2 level-2 unit but not across units. In addition the corresponding variances and covariances are constrained to be equal. We illustrate this case using the simple model with variance components at level-2 and a single variance term at level-1 (3.41). Consider the following extension of equation (3.41)

$$y_{i(j_1 j_2)} = X_{i(j_1 j_2)}\beta + u_{j_1} + u_{j_2} + u_{(j_1 j_2)} + e_{i(j_1 j_2)} \quad (3.47)$$

We have now added an 'interaction' term to the model which was previously an additive one for the two variances. The usual specification for such a random interaction term is that it has simple variance $\sigma^2_{u_{(12)}}$ across all the level 2 cells (Searle et al, 1992). To fit such a model we would define each cell of the cross classification as a level 2 unit with a between cell variance $\sigma^2_{u_{(12)}}$, a single level 3 unit with a variance σ^2_{u1} and a single level 4 unit with a variance σ^2_{u2}. The adequacy of such a model can be tested against an additive model using a likelihood ratio test criterion.

Extensions to this model are possible by adding random coefficients for the interaction component, just as random coefficients can be added to the additive components.

Level 1 cross classifications

Some interesting models occur when units are basically cross-classified at level 1. By definition we have a design with only one unit per cell, and we can also have a level 2 cross classification which is formally equivalent to a level 1 cross-classification where there is just one unit per cell. This case should be distinguished from the case where a level 2 cross classification happens to produce no more than 1 level 1 unit in a cell as a result of sampling, so that the confounding occurs by chance rather than by design.

A 2-level variance components model with a cross classification at level 1 can be written as

$$y_{(i_1 i_2)j} = X_{(i_1 i_2)j}\beta + u_j + e_{i_1 j} + e_{i_2 j} + e_{(i_1 i_2)j} \qquad (3.48)$$

where for level 1 we use a straightforward extension of the notation for a level 2 cross-classification. The term $e_{(i_1,i_2)j}$ is analogous to the interaction term in (3.47). To specify this model we would define the u_j as random at level 4, the $e_{i_1 j},\ e_{i_2 j}$ as random at levels 3 and 2, each with a single unit and the interaction term random across the cells of the cross classification at level 1, within the original level 2 units.

Suppose now that we were able to extend the design by replicating measurements for each cell of the level 1 cross-classification. Then (3.48) would refer to a 3-level model with replications as level 1 units, and which could be written as follows where the subscript *h* denotes replications

$$y_{h(i_1 i_2)j} = X_{h(i_1 i_2)j}\beta + u_j + e_{i_1 j} + e_{i_2 j} + e_{h(i_1 i_2)j} \qquad (3.49)$$

Since (3.48) is just model (3.49) with one unit per cell, we could interpret the 'interaction' variance in (3.48) as an estimate of the extent to which the additive variances of the cross-classification fail to account for the total level 1 variance.
So called 'generalisability theory' models (Cronbach & Webb, 1975) can be formulated as level-1 cross-classifications. The basic model is one where a test or other instrument consisting of a set of items, for example ratings or questions, is administered to a sample of individuals. The individuals are therefore cross-classified by the items at level 1 and may be further nested within schools etc. at higher levels. In educational test settings the item responses are often binary so that we would apply the methods discussed in next chapter (3.5.3.) to the present procedures in a straightforward way. Since each individual can only respond once to each item this an example of a genuine level 1 cross classification.

Another extension to what we have discussed so far is to allow simultaneous crossing at more than one level. However, the approach of such structures is totally analogous. Thus for example, if there is a 2-way cross classification at level 1 and a 3-way cross classification at level 2, we will require five levels, the first two describing the level 1 cross classification and the next three describing the level 2 cross classification.

Cross-unit membership models

In some circumstances units can be members of more than one higher-level unit at the same time. An example is friendship patterns where at any time individuals can be members of more than one friendship group. Another example is where children belong to more than one 'extended' family, which includes aunts and uncles as well as parents. In an educational system students may attend more than one institution. In all such cases we shall assume that for each higher level unit to which a lower level unit belongs there is a known weight, summing to 1.0 for each lower level unit, which represents, for example, the amount of time spent in that unit. We may also have data where, although there is no cross-unit membership, there is some uncertainty about which higher-level unit some lower level units belong to. For example, in a survey of students information about their neighbourhood of residence may only be available for a few students for larger geographical units. For these cases it may be possible to assign a weight for each of the constituent neighbourhoods, which is in effect a probability of belonging to each based upon available information.

Such a structure can be analyzed formally as a cross-unit membership model with most students having a single weight of 1.0 and the remainder zero.

Consider the 2-level variance components model (3.41) with each level 1 unit belonging to at most two level-2 units where the j_1, j_2 subscripts now refer to the same type of unit:

$$y_{i(j_1j_2)} = X_{i(j_1j_2)}\beta + w_{1ij_1}u_{j_1} + w_{2ij_2}u_{j_2} + e_{i(j_1j_2)} \quad (3.50)$$

where $w_{1ij_1} + w_{2ij_2} = 1$.

The overall contribution at level 2 is therefore the weighted sum over the level 2 units to which each level 1 unit belongs. This leads to the following covariance structure

$$\mathrm{var}(y_{i(j_1j_2)}) = (w_{1ij_1}^2 + w_{2ij_2}^2)\sigma_u^2 + \sigma_e^2 \quad (3.51a)$$

$$\mathrm{cov}(y_{i(j_1j_2)}y_{i'(j_1j_2)}) = (w_{1ij_1}w_{1i'j_1} + w_{2ij_2}w_{2i'j_2})\sigma_u^2 \quad (3.51b)$$

$$\mathrm{cov}(y_{i(j_1j_2)}y_{i'(j_1'j_2)}) = w_{2ij_2}w_{2i'j_2}\sigma_u^2 \quad (3.51c)$$

This has the structure of a standard 2-level cross-classified model with the additional constraint $\sigma_{u1}^2 = \sigma_{u2}^2 = \sigma_u^2$ and where the explanatory indicator variables Z_1, Z_2 described in (3.44a) and (3.44b) have the value 1 replaced by the relevant weights for each level 1 unit. As with the standard cross-classification this model can be extended to include random coefficients and general p-unit membership. In this case we need only in fact specify a single level 2 unit with explanatory variable design matrix *Z*, containing dummy weight vectors, and Ω_u as diagonal of order equal to the number of level 2 units, and elements equal to σ_u^2.

In the next Chapter we introduce examples where the cross-classified structure has to be taken seriously into account.

3.2.3 Models for Discrete response data – The Proportions as responses case

All the models of previous chapters have assumed that the response variable is continuously distributed. We now look at data where the response is essentially a count of events. This count may be the number of times an event occurs out of a fixed number of 'trials' in which case we usually deal with the resulting proportion as response: an example is the proportion of deaths in a population, classified by age.

We may have a vector of counts representing the numbers of events of different kinds which occur out of a total number of events: an example is the number of responses to each, ordered, category of a question on abortion attitudes. In all these cases the assumption of normality for the response variable is clearly violated as well as the assumption of homoscedastic error terms.

The approach to the problem of non-normally distributed variables is to include the necessary transformation and the choice of the appropriate error distribution (not necessarily a normal distribution) explicitly in the statistical model. Statistical models for such data are referred to as 'generalized linear models' (McCullagh & Nelder, 1989). A 2-level model can be written in the general form

$$\pi_{ij} = f(X_{ij}\beta_j) \quad (3.52)$$

where π_{ij} is the expected value of the response for the *ij*-th level 1 unit and *f* is a nonlinear function (called the 'link function) of the 'linear predictor' $X_{ij}\beta_j$. Note that we allow random coefficients at level 2. The model is completed by specifying a distribution for the *observed* response $y_{ij}|\pi_{ij}$. Where the response is a proportion this is typically taken to be binomial and where the response is a count taken to be Poisson. Equation (3.52) is a special case of the nonlinear models and the estimation methods for fixed and random part are extensions to those of linear models, using the appropriate transformations (e.g. the Taylor series expansion). However, what remains is to specify the nonlinear 'link' function *f*. Table 3.3 lists some of the standard choices, with logarithms chosen to base *e*. In addition to these we can also have the 'identity' function $f^{-1}(\pi) = \pi$, but this can create difficulties since it allows, in principle, predicted counts or proportions which are respectively less than zero or outside the range (0,1). Nevertheless, in many cases, using the identity function produces acceptable results, which may differ little from those obtained with the nonlinear functions. For the purpose of the thesis we consider only the type of model where responses are proportions in dichotomous variables.

Table 3.3: Some nonlinear link functions

Response	$f^{-1}(\pi)$	**Name**
Proportion	$\log\{(\pi)/(1-\pi)\}$	logit
Proportion	$\log\{-\log(1-\pi)\}$	complementary log log

Vector of proportions	$\log(\pi_s / \pi_t) \;\; (s = 1,\ldots,t-1)$	multivariate logit
Count	$\log(\pi)$	log

Proportions as responses

Consider the 2-level variance components model with a single explanatory variable where the expected proportion is modeled using a logit link function

$$\pi_{ij} = \{1 + \exp(-[\beta_0 + \beta_1 x_{1ij} + u_{0j}])\}^{-1} \qquad (3.53)$$

The observed responses y_{ij} are proportions with the standard assumption that they are binomially distributed

$$y_{ij} \sim Bin(\pi_{ij}, n_{ij})$$

where n_{ij} is the denominator for the proportion. We also have

$$\mathrm{var}(y_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / n_{ij} \quad (3.54)$$

We now write the model in the standard way including the level 1 variation as

$$y_{ij} = \pi_{ij} + e_{ij} z_{ij}, \quad z_{ij} = \sqrt{\pi_{ij}\left(1 - \pi_{ij}\right) / n_{ij}}, \quad \sigma_e^2 = 1 \quad (3.55)$$

Using this explanatory variable Z and constraining the level 1 variance associated with this to be one we obtain the required binomial variance in equation (3.54). When fitting a model we can also allow the level-1 variance to be estimated and by comparing the estimated variance with the value 1.0 obtain a test for 'extra binomial' variation. Such variation may arise in a number of ways.

Estimation methods for both fixed and random parameters for a proportion as response model is a demanding procedure and therefore will not be discussed further. Goldstein (1995) refers to the distinction between 'predictive quasilikelihood' (PQL) and 'marginal quasilikelihood' (MQL) estimation procedures (Breslow & Clayton, 1993). In many applications the MQL procedure will tend to underestimate the values of both the fixed and random parameters, especially where n_{ij} is small. When the sample size is small the unbiased (RIGLS, REML) procedure should be used.

In the next Chapter we focus more on examples where proportions are used as the response variables and generalized linear models in the form discussed here are used.

3.2.4 Multivariate Multilevel Models – The basic 2-level Multivariate model

Multivariate Multilevel models

In the models discussed so far we have considered only a single response variable, either normal or not, measured at the first level of hierarchy. We now look at models where we wish simultaneously to model several responses as functions of explanatory variables. As we shall see, the ability to do this provides us with tools for tackling a very wide range of problems. These problems include missing data, rotation or matrix designs for surveys and prediction models.

We develop the model considering the case of two response variables measured at the individual level while explanatory variables are measured at all levels of hierarchy.

The basic 2-level multivariate model

To define a multivariate, in this case a 2-variate, model we treat the individuals as a level 2 unit and the 'within-individuals' measurements as level 1 units. Each level 1 measurement 'record' has a response, which is either the first or the second variable. The basic explanatory variables are a set of dummy variables that indicate which response variable is present. Further explanatory variables are defined by multiplying these dummy variables by individual level explanatory variables, for example a dichotomous variable with values 1 and 0. The model is written as

$$y_{ij} = \beta_{01} z_{1ij} + \beta_{02} z_{2ij} + \beta_{11} z_{1ij} x_j + \beta_{12} z_{2ij} x_j + u_{01j} + u_{02j} \quad (3.56)$$

where

$$z_{1ij} = \begin{Bmatrix} 1 \text{ if 1st variable is present} \\ 0 \text{ if 2nd variable is present} \end{Bmatrix}, \quad z_{2ij} = 1 - z_{1ij}, \quad x_j = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix} \quad (3.57)$$

and

$$\text{var}(u_{01j}) = \sigma_{01}^2 \quad (3.58a)$$

$$\text{var}(u_{02j}) = \sigma_{02}^2 \quad (3.58b)$$

$$\text{cov}(u_{01j} u_{02j}) = \sigma_{012} \quad (3.58c)$$

There are several features of this model. There is no level 1 variation specified because level 1 exists solely to define the multivariate structure. The level 2 variances and covariance are the (residual) between-individuals variances. In the case where only the intercept dummy variables are fitted, and since every individual has scores for both the response variables, the model estimates of these parameters become the usual between-subjects estimates of the variances and covariance. The multilevel estimates are statistically efficient even where some responses are missing, and in the case where the measurements have a multivariate Normal distribution they are maximum likelihood. Thus the formulation as a 2-level model allows for the efficient estimation of a covariance matrix with missing responses.

We can further allow the individuals to be grouped within level 3 units and therefore add more variability terms in the 3-rd level of the model. These can be variances and a covariance for the two components added at level 3 as well as additional variance terms for the second level explanatory variables.

Designs with subsets of responses – Missing cases

We have already seen that fully balanced multivariate designs are unnecessary and randomly missing responses are handled automatically. The basic 2-level formulation does not formally recognize that a response is missing, since we only record those present. We now look at designs where responses are effectively missing by design and we see how this can be useful in a number of circumstances.

In many kinds of surveys the amount of information required from respondents is so large that it is too onerous to expect each one to respond to all the questions or items. In education we may require achievement information covering a large number of areas, in surveys of businesses we may wish to have a large amount of detailed information, and in household questionnaires we may wish to obtain information on a wide range of topics. We consider only measurements that are used as responses in a model. If we denote the total set of responses as $\{N\}$ then we choose *p* subsets $\{N_i, i = 1, \ldots p\}$ each of which is suitable for administering to a subject (level 1).

When choosing these subsets we can only estimate subject-level covariances between those responses that appear together in a subtest. It is therefore common in such designs to ensure that every possible pair of responses is present. If we wish to estimate covariances for higher-level units such as schools it is necessary only to ensure that the relevant pair of responses are assigned to the some schools - a large enough number to provide efficient estimates. The subjects are assigned at random to subtest and higher-level units are also assigned randomly, possibly with stratification. Each subset is viewed formally as a multivariate response vector with randomly missing values, although the missing observations are produced by design. We can then fit a multivariate response model for such data and obtain efficient estimates for the fixed part coefficients and covariance structures at any level. In this formulation, the variables to be used as explanatory variables should be measured for each level 1 unit.

Multivariate cross-classified models

We know consider the multivariate case of a type of models discussed previously – the cross-classified models. For multivariate models the responses may have different structures. Thus in a bivariate model one response may have a 2-level hierarchical structure and the other may have a cross classification at level 2. Suppose, for example that we measure the height and the mathematics attainment of a sample of students from a sample of schools. The mathematics attainment is assessed by a different set of teachers in each school and the heights are measured by a single anthropometrist. For the mathematics scores there is a level-1 cross-classification of students within each school whereas for height there is a 2-level hierarchy with students nested within schools. Height and mathematics attainment will be correlated at both the student and the school level and we can write a model for this structure as follows

$$y_{h(i_1i_2)j} = \delta_{1h}(X_{1(i_1i_2)j}\beta_1 + u_{1j} + e_{1i_1j} + e_{1i_2j}) + \delta_{2h}(X_{2i_1j}\beta_2 + u_{2j} + e_{2i_1j}) \quad (3.59)$$

where

$$\delta_{1h} = 1 \; if \;\; mathematics, \; 0 \; if \;\; height, \quad \delta_{2h} = 1 - \delta_{1h} \qquad (3.60)$$

and

$$\operatorname{cov}(u_{1j}u_{2j}) = \sigma_{u12} \qquad (3.61a)$$

$$\operatorname{cov}(e_{1i_1j}e_{2i_1j}) = \sigma_{e12} \qquad (3.61b)$$

and all other covariances are zero. This will therefore be specified as a 4-level model with the bivariate structure as level 1 and level 2 units being individual students. There will be a single level 3 unit with the coefficients of the dummy variables for teachers having variances random at this level, with level 4 being that of the school.

3.3 Conclusions of the Chapter

The conclusions drawn by the discussion of this Chapter can verify the theoretical advantage of Multilevel Models compared to other techniques. First of all, in simple situations they respect totally the hierarchy of the data and they end up to more precise estimations both for the model parameters and for the model variability. Secondly, all known statistical techniques and procedures (such as Maximum Likelihood, EM Algorithm, MCMC estimations etc) for statistical inferences (parameter estimates, testing functions, confidence intervals) can be easily used, making Multilevel Methods theoretically understandable and easily applicable for statisticians and researchers. Finally, the simple multilevel models can be readily extended to more sophisticated theoretical concepts, such as multivariate or generalized models, and therefore can be applied effectively to more complex situations. In the following Chapter we will discuss if the theoretical advantages of the Multilevel Techniques are also present in practice, or, in other words, why and how Multilevel Analysis is useful and effective in practical situations.

CHAPTER 4

4 REVIEW OF APPLICATIONS

In this Chapter we introduce a number of practical examples and applications of hierarchical data structures that were previously described within the thesis. These examples cover the whole application area of multilevel modeling, from educational data analysis to health analysis, from social statistics to survey research, as well as meta-analysis, repeated measures analysis and so on. It is profound that the applications reviewed here are only a minor sample of the thousands printed articles dealing with applications and data analysis by using, among others, multilevel techniques. However, the main reason they were chosen is that they have formed the basis for theoretical discussions as well for other applications in the same or other areas of interest, from the same or other authors.

In this Chapter we will focus more on the description of the example, the data description and analysis, the statistical methods and techniques and the discussion of the results rather than the results themselves. We will also mention other relative to the examples, reviews and articles from the same or other authors, which were based or "inspired" from the particular articles. We split our review into examples taken from the area of education and applications in all other areas of interest.

The scope of this Chapter is, therefore, to show if Multilevel Analysis is handly applicable in practice and if the theoretical advantages of the Multilevel Techniques, as discussed in the previous Chapter, are also present in practical situations derived from various research areas.

4.1 Applications in Education

Analysis of educational data, students and organizations performance is the area were multilevel model techniques were firstly introduced and have been used to a great extent, due to the profound hierarchical structure of the data measured in this area. We start our review from an introductory multilevel analysis of school examination results (Goldstein et al, 1993), which, however, form the basis to almost all the analysis in these areas.

‘A Multilevel Analysis of School Examination Results’

a. INTRODUCTION

Goldstein et al. (1993) examined data on examination results from inner London schools in relation to intake achievement, pupil gender and school type. The examination achievement, averaged over subjects, was studied as well as achievement in the separate subjects of mathematics and English. Multilevel models were fitted, so that the variation between schools could be studied. Specifically he focused on two measures of intake achievement for each school in the study, and examined the interpretational issue by studying the dimensionality of school differences.

b. DATA

The data were examination results from 5748 students in 66 schools in six Inner London Education Authorities. These students had data on their General Certificate of Secondary Examination (GCSE) grades in Mathematics and English, together with a total score for all the subjects taken in that examination. For mathematics and English, a scale ranging from 0 (no grade awarded) to 7 (grade A) was used in the analysis and, for the total score, the scale ranged from 0 to 70. These students also had scores on a common reading test taken when they were 11 years old-the London Reading Test (LRT) (Levy & Goldstein, 1984) and were graded also into three categories on the basis of a verbal reasoning (VR) test at 11 years (Nuttall et al., 1989). All three scores were scaled to have mean zero and standard deviation 1. The pattern was similar for all three response variables.

The original number of students on whom some examination data had been obtained was 8857 in 74 schools. Students were omitted from the analysis if they did not have both intake measures. Where students did not take an examination they were given a score of 0, the same as if they obtained an ungraded result. The exclusion of these students resulted in a sample with a higher total examination score, 23.7 as opposed to 20.0. As the author pointed out this differential loss of students with lower examination achievements needs to be born in mind when interpreting the results, and is a persistent problem with data of this kind.

Two separate models had been fitted to the data. The first analyses the total examination score and the second is a bivariate analysis of the English and mathematics scores. All the response variables had been transformed using normal scoring to conform as closely as possible to multivariate normality.

c. TOTAL EXAMINATION SCORE

The explanatory variables used in this analysis were as follows:

- standardized London reading test (LRT);
- verbal reasoning category;
- gender;
- school gender (mixed, girls, boys);
- school religious denomination (State, Church of England, Roman Catholic, other).

Formally, the model was written as follows:

$$y_{ij} = \sum_{h=0}^{5} \beta_h x_{hij} + \sum_{h=6}^{10} \beta_h x_{hj} + \sum_{h=0}^{2} u_{hj} x_{hij} + \sum_{h=0}^{1} e_{hij} x_{hij} \qquad (4.1)$$

where i refers to student and j refers to school. Throughout this equation the subscript 0 refers to the constant term (= l), the subscript 1 to LRT and 2 to the dummy variable for VR group 1. Subscripts 3-5 refer to the square of LRT, the dummy variable for verbal reasoning group 2, and the dummy variable for gender. The subscripts 6-10 refer to the five school-level defined i.e. the 'fixed part'. These are:

- Girls-mixed school
- Boys-mixed school
- CE-State school
- RC-State school
- Other-State school

The first summation refers to the explanatory variables defined at the student level, the second to those defined at the school level, the third to the random part of the model defining variation at the school level, that is level 2, and the fourth summation defines the random variation at the student level, that is level 1. We also have, at level 2,

$$\text{var}(u_{hj}) = \sigma_h^2 \quad (4.2a)$$

$$\text{var}(e_{hij}) = \sigma_{eh}^h \quad (4.2b)$$

The level 1 contribution to the variance is:

$$\sigma_{e0}^2 + 2\sigma_{e01} + \sigma_{e1}^2 x^2 \quad (4.3.)$$

That is, a quadratic function of x and the individual level variances and covariance in this expression do not have separate interpretations. On the other hand, since the x_{hij} are defined at level 1, the level 2 variances and covariances are interpreted directly as between-school variances and covariances for the relevant coefficients. Several exploratory models were fitted and the main results for the model found to give the most satisfactory fit are the following:

- The level 1 (between students) variance is a linear function of LRT score given by: $Variance = 0.55 + 0.092LRT$
- Likelihood ratio test statistics for:
  - (a) Level 1, LRT (covariance): $X_1^2 = 66.0, P < 0.001$
  - (b) Level 2, VRl (VR2,VR3) variances and covariance: $X_3^2 = 11.0, P = 0.012$
  - (c) Level 2, LRT variance: $X_3^2 = 24.7, P < 0.001$.
- The effect of school gender is small and the differences are about the same order of magnitude as the estimated standard errors.
- There seems to be a small advantage for those attending Roman Catholic schools.
- Girls do better than boys
- There are large differences between those in the different verbal reasoning categories and there is a strong quadratic relationship with LRT.
- The relationship between examination score and LRT varies, as does the difference between verbal reasoning categories 1 and 3, with high positive correlations.
- At the student level the variance increases with increasing LRT score.

After having fitted the model it was possible to estimate the level-2 residuals. These are obtained by estimating the regression model with the (unknown) residuals as responses. The resulting estimates are often known as 'shrunken' estimates since, like all regression predictions they have smaller variances than that of the true values. To illustrate the implications of the model, the authors formed particular extreme combinations of the school residuals. So for each school they calculated, the two combinations $u_{0j} - 2u_{1j}$ and $u_{0j} + 2u_{1j} + u_{2j}$ that is, first the estimated school 'effect' for a student with an LRT score of - 2, the approximate lower 2.5th percentile, and in verbal reasoning group 2 or 3, and second the estimated 'effect' for a student at the approximate upper 2.5th percentile and in verbal reasoning group 1.

It was concluded that there is a positive correlation between the school 'effects' for the low and high achievers on intake. Nevertheless, there were some schools with below average values for the low achievers which had above average values for the high achievers, and vice versa. This emphasizes the point that schools appear to be differentially effective for different kinds of students.

Then, approximate 95% confidence intervals for estimate of the intercept residual were constructed and plotted. That is the school 'effect' estimated at the mean LRT score for those in verbal reasoning groups 2 and 3. It was noted that these intervals were calculated separately for each residual, and were based upon the estimated standard error, which in general are an underestimate of the true standard error. For comparing any two particular schools, the usual significance test and confidence interval procedures were used. It was seen that there is a very considerable overlap of intervals, which suggests that it is not possible statistically to discriminate easily between schools. In particular, there are no natural division points in the sequence of estimates, which would allow the authors to classify schools into homogeneous subgroups.

d. JOINT ANALYSIS OF ENGLISH AND MATHEMATICS SCORE

The next step was the analysis of the English and mathematics examination scores. These were chosen because, in principle, these examinations are taken by all students. It would be possible to carry out a joint analysis of these two scores together with the total score on the other subjects, but for simplicity of interpretation the authors restricted to just the two. Also for simplicity, they used only the student level variables as explanatory variables, and at the between-school level they used only the intercept and LRT coefficient as random variables.

A multivariate model was fitted by treating the multiple variates within each student as the level 1 classification. In this case, therefore, there were two level 1 units within each student (level 2) with schools at level 3. It was concluded that:

- In the fixed part of this model the average difference between girls and boys is 0.1 units in favor of the boys for mathematics and 0.23 units in favor of the girls for English.

- For LRT and verbal reasoning categories, there is little difference between the relationships for maths and English.
- In the random part of the model the LRT coefficients for English and maths do not vary greatly. The standard deviation for maths is 0.006 units while that for English is only 0.003.
- While schools differ in terms of overall maths and English performance, at least for English, there is little differential effect according to LRT at intake. The intercepts for maths and English have a small correlation (0.09), and there is only a moderate correlation for the intercepts and LRT coefficients for both maths and English.
- At the student level the between-students variation decreases from VR1 to VR3 category students. This is similar to the finding in the analysis of total score, where the lower achieving intake students (based on LRT) had smaller variance
- If a model is fitted with just a variance term for mathematics and English and a covariance term at level 1, the effect is to increase the standard errors for both the fixed part of the model and the level 2 random parameters by up to 20%. The estimates of the coefficients and parameters themselves do not change appreciably, but the decrease in precision emphasizes the importance of accurate level 1 modeling.

To illustrate the relationship among the school level residuals or 'effects', the residual estimates for the two intercepts, that is at the mean LRT score were plotted. It was shown that there is little relationship between English and maths performance. The school with the greatest English residual was only average for maths and one of the schools with high maths residual had a low value for English. This relationship is for those students with average LRT scores. Since the LRT coefficients vary across schools, the relationship between maths and English residuals will also vary with the LRT score.

e. DISCUSSION

The analysis leaded to some important comments mentioned by the authors:

- There is an association between the between-student variance and the intake achievement score, with increasing variation as the intake achievement increases. This is of substantive interest and it is also important to incorporate it in the model since it helps to ensure that the overall model is correctly specified and will generally improve the precision of the remaining parameters.

- The authors criticize on the reliability of the explanatory variables used in the analysis (LRT score or the VR band allocation). Also, whether the use of particular variables (verbal reasoning and reading achievement measures) are entirely satisfactory to adjust for intake. They comment that, ideally, when total examination score is used as the response, initial achievement measures should cover the full range of school subjects taken in the examination
- They also criticize 'league tables' (published Tables with the average General Certificate of Secondary Education examination results in England and Wales, used by parents in order to choose schools and colleges) whether or not these are adjusted for intake achievement and whether or not multilevel modeling has been used. The ordering of school effects depends on the intake achievements of students as well as the curriculum subject being examined. Also, a study of residuals differentiated by intake achievement and by subject, can suffice as a screening device and as feedback to individual schools about potential problems Even then however, the ability of comparison between schools based on residuals is still controversial. Fine distinctions and detailed rank orderings are statistically invalid.

Harvey Goldstein as the "leader" and a large group of other scientists have introduced the "Institute of Education" in the University of London. This organization with a large number of publications has always given rise to various multilevel issues and has inspired other scientists from the area of education research or other areas. They have also established particular software (MLwiN) for multilevel analysis, which is frequently updated to catch up with the new considerations. Their articles cover a wide area of issues in multilevel techniques.

(a) "League Tables", their limitations and, school performance and comparisons are a major issue of importance in many of their studies. They introduce the use of "Value Added Information" i.e. how many units of improvement an institution has added to its students' achievement and performance. Such studies are performed by Goldstein & Spiegelhalter (1996), O'Donoghue, Thomas, Goldstein & Knight (1996), Goldstein (1997), Goldstein (1998), Goldstein & Woodhouse (2000), Goldstein, Huiqi, Rath & Hill (2000).

(b) Another issue of interest analyzed in details from the particular group of scientists are applications in the extensions of the basic multilevel models, such as multivariate multilevel models, cross-classification and multilevel repeated measures models. We have already discussed the multivariate case of joint analysis of English and Mathematics Scores (Goldstein et al, 1993). Yang et al. (2002) have elaborated significantly on this issue. Cross-classification in educational data has been analyzed, among others by Hill & Goldstein (1998). Another example of cross-classification of schools in examination results (Goldstein, 1995) is presented here:

The data consist of scores on school leaving examinations obtained by 3435 students who attended 19 secondary schools cross-classified by 148 primary schools in Fife, Scotland (Paterson, 1991). Before their transfer to secondary school at the age of 12 each student obtained a score on a verbal reasoning test, measured about the population mean of 100 and with a population standard deviation of 15. The model is as follows:

$$y_{i(j_1 j_2)} = \beta_0 + \beta_1 x_{1i(j_1 j_2)} + u_{j_1} + u_{j_2} + e_{i(j_1 j_2)} \qquad (4.4)$$

and a number of alternative models were fitted. Random coefficients for verbal reasoning were estimated as zero.

The results of the study are as follows:

- Ignoring the verbal reasoning score, the between-primary school variance is estimated to be more than three times that between secondary schools. The principal reason for this is that the secondary schools are on average far larger than primary schools, so that within a secondary school, primary school differences are averaged. Such an effect will often be observed where one classification has far fewer units than another, for example where a small number of schools is crossed with a large number of small neighbourhoods or a small number of teachers is crossed with a large number of students at level 1 within schools. In such circumstances we need to be careful about our interpretation of the relative sizes of the variances.
- When the verbal reasoning score is added to the fixed part of the model the between secondary school variance becomes very small, the between primary school variance is also considerably reduced and the level 1 variance also.
- When cross-classification is removed by primary school, the between secondary school variance is only a little smaller than in analysis without

verbal reasoning score. Using this kind of analysis, which is typically the case with school effectiveness studies, which control for initial achievement, there were important differences between the progress made in secondary schools. However, most of this is explained by the primary schools attended.

Of course, the verbal reasoning score is only one measure of initial achievement, but these results illustrate that adjusting for achievement at a single previous time may not be adequate.

(c) Although the estimation techniques and algorithms for both the fixed and the random part of a multilevel model were of major importance from the beginning of the adaptation of such models, the rapid development of IT has given rise to this subject. More demanding algorithms and techniques, such as Bootstrap methods, can now be easily applicable in the multilevel software. The "institute of Education" team of scientists has thoroughly dealt with estimation issues. We refer, among others, to the recent works of Browne et al. (2002) and Carpenter, Goldstein & Rasbash (2003).

Multilevel analysis for educational data was introduced by Aitkin (1981). Since then, a great number of articles have been published for this issue, apart from those already mentioned from the 'Institute of Education'. We can refer to Kreft (1993) and McArdle & Hamagami (1994) for simple applications of basic multilevel and logit multilevel models in educational performance. Kreft & Leuew (1998) use basic models multilevel techniques to analyze a large set of data from the National Education Longitudinal Study of 1988 (NELS-88) taken by the National Center for Education Statistics of the US Department of Education. They follow the basic Two-Level approach, however their work has a number of practical advantages. Firstly, they follow a "step-by-step model building" approach which explains perfectly the practical use of Multilevel Analysis. Their critical comments throughout the analysis, stretch some of the major drawbacks in Multilevel Analysis and set the basis for practical considerations. Such problems are estimation techniques and considerations, the effect of "centering" on data in multilevel approach models, 'multicolinearity' issues and so on. Since they compare the results of different subsets of different sample sizes, they illustrate the "sample size selection" issue. Finally, in each step of their analysis, they introduce the basic commands of Mln software (the previous version of MLwiN) as well as references to other known software programs. Afshartous & DeLeeuw (2002) elaborate more on this data set trying to set best techniques for parameter estimation.

In Greece, although the school effectiveness and students' performance issue is of great interest, the only serious attempt to perform multilevel techniques in educational data was taken by Kosmopoulou (1998). In her dissertation project, the author performed multilevel models analysis, and more specifically fitted a 3-level model, in Greek educational data in order to assess school effectiveness and students' performance in the National Entrance Exams of 1990 and 1991 for the student's access in the National Universities and Technical Institutions. The response variable of interest was the mean score of students in the National Entrance Exams. We should notice that according to the educational system being in use at the period of the study, students in the 3rd of Lyceum were divided into 4 scientific orientations ("desmes") and each student according to the scientific orientation should be examined in four subjects in the National Entrance Exams. The explanatory variables were the mean score of students in the third class of Lyceum, the type of school (public/private), the gender of student, the scientific orientation (4 categories) and the year of examination (1990/1991). In the year 1990 the number of level-3 units (prefectures) was 51, the number of level-2 units (schools) was 961 and the number of level-1 units (students) was 52041 while in the year 1991 the corresponding numbers were 51, 978 and 54200, respectively. The author performed a 3-level model and the main final results of the analysis were the following:

- The 3-level model is a significant improvement compared to the simpler 2-level and 1-level models.
- When unadjusted examination results were used (the explanatory variable "mean score of students in the third class of Lyceum" was omitted), girls concluded to perform much better than boys, the type of school was not statistically significant, students of the 1st, 2nd and especially 3rd scientific orientation performed better than those in the 4th and finally, students in 1990 performed better than students in 1991.
- When adjustment was made, it was concluded that boys do better than girls, public schools perform better than the private ones, students who chose the 4th scientific orientation performed better than those in the 1st and finally, again students in 1990 performed better than students in 1991.
- According to the descriptive statistics, the prefecture with the highest mean score in both years was Chios and the prefectures with the lowest mean score were Evros for 1990 and Evritania for 1991.

Although the educational system of National Exams described above is totally different than the system that will be described in our application in the following Chapter, the results of Kosmopoulou (1998) will be the most reliable reference of comparison, due to the lack of other relevant studies in this area of research.

We should also refer to some other projects and results from the Greek literature that will be used for comparison reasons, although their statistical approach is not multilevel modeling but is mainly constrained in descriptive statistics. Some of them are:

- The scaling of the General Admission Grade of the candidate students for the access to the National Universities and Technical Institutions by scientific orientation, published every year by the Greek Ministry of Education, Lifelong Learning and Religious Affairs (www.ypepth.gr).
- The dissertation of Marouga (2004) for the examination of students' performance and preferences according to the Greek National Exams.
- The project of the Centre of Development of Educational Policy of the General Confederation of Greek Workers (GSEE) (2009) about the system of access in the third degree education (National Exams 2004, 2005 and 2006) (www.kanep-gsee.gr/index.php?download=keimeno%206-4-09.doc).
- The report of the Centre of Research and Documentation of the Greek Federation of State School Teachers of Secondary Education (OLME) (2004) expressing the views of the organization for the suppression of the National Exams and the autonomy of Lyceum (http://www.smarinis.gr/aei1.pdf).

4.2 Applications in various areas

In this Chapter we introduce particular examples where various multilevel model techniques are performed in various areas of interests. Such areas have already been described in previous Chapters according to the hierarchical structure of the data measured, and stem from spatial statistics, health research, survey research, repeated measures and meta-analysis.

The scope of this review is to detect in which ways multilevel analysis has introduced in practice in these areas and present the basis of more extended use when multilevel models are proved to be appropriate.

4.2.1 Spatial Statistics

The first article is taken from the area of spatial statistics (Courgeau & Baccaini, 1998).

“Multilevel Analysis in the Social Sciences”

a. INTRODUCTION

The authors used the multilevel approach to study human behaviour taking into account not only individual characteristics but also the fact that these individuals belong to larger geographical units such as communes and regions. This article gives a detailed critical presentation of the models used already and the models to be introduced, as well as the aims and formulations of these models. Attention ranges from the most basic models, which introduce the many different levels in the form of individual and aggregated characteristics, to more complex models, which operate with the random characteristics specific to each level.

Demography has for long favored analysis at aggregated levels. Attention focuses on identifying the relations which exist between the classic demographic rates, corresponding to the phenomenon being studied in each sub-population, and the average values of the characteristics, also calculated for each subpopulation. An analysis of the emigration rates of different regions, for example, would try to link them to the unemployment rates, average incomes, percentage of dependants, etc., found in these regions. The authors state the danger of mistaken inference when we try to infer individual behaviour from aggregated measures (‘ecological fallacy) or when we examine individuals separately (‘atomical fallacy’) and focus on the need to introduce multilevel analysis when referring to emigration models given rise to the idea of working simultaneously on different levels of aggregation, with the aim of explaining a behaviour which is always treated as individual, rather than aggregated as before. Then the authors set up the basic multilevel, as well as the logit multilevel models in a way that has already been discussed in previous chapters.

b. DATA

The example the authors chose to work is that of regional migration in Norway. Data were taken from the Norwegian Population Register to demonstrate the importance of aggregation effects. Various types of model can be used according to the type of data i.e. exponential regression to model the emigration rates of the regions, logit and event history models to explain the individual risks of migrating in terms of the characteristics of the zones being considered and of the individuals themselves.

Norway has local population registers in which are recorded the demographic events of the individuals living in the country, and in particular their internal migrations (changes of municipal districts). The file used contains the 54814 individuals born in 1958, who lived in Norway in 1991 and who had not migrated abroad. For each of these individuals the successive changes of region (Norway is divided into 19 regions, was known. Only the regional emigration flows, observed over a short period of two years, 1980 and 1981 were considered when the individuals were aged 22-23. A census was conducted in 1980, so various characteristics of the individuals at this date were also known, and there was also the ability to establish how long the individual had been living in the region of residence at the start of 1980. At the individual level eight characteristics had been selected as having a possible effect on the chances of moving out of the region: marital status (married Vs unmarried), being economically active (active Vs non-active), type of occupation (farmer Vs non-farmer), educational level (more than 12 years in full-time education Vs less than 13 years in full-time education), having children (at least one child Vs no children), and the level of income (high income; low income; no income). The authors were then able to reconstitute the aggregated characteristics for the 19 regions (percentage of individuals having left the region in 1980- 1981, percentage of individuals married, percentage of farmers, etc.).

c. ANALYSIS OF INDIVIDUAL AND AGGREGATED CHARACTERISTICS

In the first step, all the analysis was performed separately, either in individual or in regions level (using aggregated data). Although the results were highlighted by the authors we mention that the authors conclude their discussion by: “These early analyses, which need to be further developed, show how understanding of migratory processes can be increased by the simultaneous inclusion in the models of the characteristics of the individuals and of the regions of origin and destination”. They also illustrate the care needed when interpreting the results.

d. ANALYSIS WITH MULTILEVEL RANDOM VARIABLES – DISCRETE RESPONSE DATA

Many demographic characteristics are observed in the form of dichotomous or polytomous variables: an individual is married or not, an individual can migrate between n regions, for example. The authors examined the dichotomous case, the migration flows of the 19 Norwegian regions, for individuals born in 1958 and who migrated in 1980-81. To explain the behaviour they also used the 8 individual and aggregated characteristics that have already defined. Because individual and aggregate characteristics have a specific effect on the probabilities of emigrating from the regions, they introduced them first for each type of characteristic in a simple logit model and in a multilevel logit model. After fitting a number of alternative models for men, the main conclusions were the following:

- The non-random parameters estimated with a multilevel model are in general very close to those we obtain with a simple logit model. But when the effect of the random terms related to the characteristics is not zero at the regional level, a large increase in the dispersion of these parameters is observed, with an approximate doubling of their standard deviation. Despite this, most of the effects that are significant at the 5% threshold in the simple logit model are also significant in the multilevel model. The only exceptions to this rule are two effects of aggregated characteristics: the positive influence that living in a low-income region had on the chances of migrating becomes non-significant in the multilevel model; on the other hand, regions with a high educational level are found to have a significant effect of reducing the chances of migrating when the multilevel model is used, whereas this was not apparent at all with the simple logit model.
- Farmers have a much lower probability of migrating than the other categories, but the higher the percentage of farmers in a particular region, the higher the probability of migrating for all categories. This result highlights the danger of inferring individual results from results obtained at a more aggregated level: the presence of a large number of farmers in a region results in a higher probability of emigrating for all the categories of population, doubtless due to the greater scarcity of non-agricultural employment in such regions. But this does not mean that farmers have a higher probability of emigrating than the other categories, since it is the exact opposite that is observed.

- Individuals with at least one child are found to have a probability of migrating that is always lower than those without children, whether or not we introduce the percentage with at least one child. It is also verified that the variance between regions of the logits of the probability of migrating of those with at least one child is three times higher than that of those with no children when the percentages with at least one child are not introduced. When this percentage is introduced it has a highly significant effect and, above all, it reduces the between-region variances and covariances by half, so that they lose their significant effect: We can therefore say that it does indeed explain a part of these random effects.
- In the case of individuals with more than 12 years of full-time education, they have a higher probability of migrating than the others.
- In a model in which all the characteristics considered are included which have an effect on the regional probability of migrating, the results of a simple logit model compared with a multilevel model show that only the characteristics of educational level are considered to be random between regions.
- The effects of the non-random characteristics are very similar whether the first or second model is used. The case of the farmers now becomes fully significant: the fact of being a farmer always reduces the probability of migrating. By contrast, the fact of having a high income becomes non-significant in a multilevel model.
- The random parameters at the level of the region are modified the most, in relation to the model in which only the fixed educational level characteristic is included. The between-region variance is reduced to a fifth of what it was by the introduction of the other characteristics. On the other hand, the between-region variance remains close to what it was. The correlation between the regional hazards of individuals with more or less than 12 years of full-time education is highly negative whereas it was almost null in the earlier model.
- The final conclusion by the authors is that the introduction of a model using the multilevel random variables does not alter the essential of the conclusions obtained with a simple logit model taking into account the characteristics measured at different levels of aggregation. On the other hand, these random variables provide some valuable information about the relationships between

probabilities of emigrating from the different regions of individuals who have or do not have a given characteristic.

e. CONCLUSIONS

When drawing the final conclusions, the authors state a number of questions to be discussed when a multilevel model is introduced. These are, in brief, the appropriateness and the accuracy of a higher lever measures for human behaviour, the number of levels to be introduced, the need of examination of the behaviours that are specific to each level before we perform a jointed analysis and so on. In any case, however, they recognize the 'rich potential of multilevel models, but also the need to situate them in a coherent theoretical framework'.

4.2.2 Health Statistics

Health statistics is another area where multilevel models can easily be applied due to the clear hierarchy of the data measured in this area (patients are clearly nested within health care providers). Since the similarities with the educational area are obvious, it is not surprise that all the formulas, techniques, algorithms and inferences are borrowed by the educational literature. We have, therefore performance indicators for heath care units analogous to those for educational institutions (Goldstein & Spiegelhalter, 1996) and in general authors from one area to draw conclusions for the other (Goldstein, Browne & Rasbash, 2002). In this chapter we focus on an example taken from the health literature (Carey, 2000).

'A multilevel modelling approach to analysis of patient costs under managed care'

a. INTRODUCTION

The paper is interested in the performance of health care providers and more specifically to analyse the effects of managed care penetration on patient level costs for a sample of 24 medical centres operated by the Veterans Health Administration (VHA) using multilevel modelling techniques. The appropriateness of a two level approach to this problem over ordinary least squares (OLS) is demonstrated. To date, the majority of studies of the determinants of costs have been performed at the hospital level. However, a fundamental practice of managed care associations is moving patient services out of the hospital inpatient setting to ambulatory sites, medical offices, nursing homes, or patients' own homes. A statistical technique that is well-suited to examining the effects of provider institutions on patients' costs in a managed care environment is multilevel modelling. This framework is used to analyse data that fall naturally into hierarchical structures consisting of multiple 'micro' units nested within 'macro' units. The author clearly emphasises the advantages of a multilevel technique compared to the classical OLS approach, which have been discussed in depth in previous chapters. Also provides the appropriate techniques and formulations for basic multilevel models, as well as for cross-classified models, mentioned but not used in the analysis.

This paper uses a multilevel modelling approach to explain variation in patient costs in facilities operated by the (VHA). The analytic framework developed here explores cost variation occurring at the VHA medical facility level and its interpretation, after controlling for differences in individual patients. The multilevel approach allows for drawing insights regarding the relative performance and efficiency of these public institutions operating in the current environment of managed care.

b. DATA

During 1997, US Veterans Health Administration (VHA) operated 148 medical centres providing a continuum of services under an organized model of managed care in which veterans were gradually being assigned to a primary care physician or team. These institutions provide care for military service connected injuries as well as a broader spectrum of care for poorer veterans. VHA was organized into offering veterans a continuum of treatment settings that in addition to inpatient care include outpatient clinical and surgical services, nursing home care, and home health care. The medical centres were grouped into 22 geographically-based Veterans Integrated Service Networks (VISNs), or groups of medical facilities providing a structural and organizational foundation for integrating services and planning. The structural and functional changes can be summarily characterized as a movement away from hospital based care toward managed care within integrated delivery systems. The data in this study includes 24 medical centres operating in three randomly chosen VISNs and servicing 526117 individual patients in fiscal year 1997.

The primary data sources used in this study come from VHA administrative files. The two level model nests patients within provider units for the fiscal year 1997 (the 1-year period commencing on 1 October 1996).

The response variable is the total direct cost of an individual accrued by VHA during the fiscal year. These costs are adaptations of patient level costs derived from the Cost Distribution Report (CDR), a standard costing procedure in which VHA allocates costs from central accounting sources to specific patient care programmes. The cost of hospital inpatient care, outpatient care, and long-term care are included in the response variable, also including physician costs. The dependent variable used in the analysis is the log of this patient cost.

A critical factor in determining health care cost variation is the clinical state of the patients served. An advantage of multilevel modelling is that it can integrate micro data stemming from a growing volume of patient care information systems. Individual patient health status is measured using the Diagnostic Cost Group (DCG) methodology. DCG models use demographics and diagnoses to predict individuals' relative resource. These may be interpreted as measures of relative health status based on expected expenditure differences in comparison to a benchmark population normalized around a mean of 1.00c. DCGs form risk groups of persons based on diagnoses that are similar. DCG risk scores are well-equipped to control for the clinical status of individual patients in this model of patient annualized costs. Patient level data also includes age and sex, drawn from VHA's Patient Treatment File (PTF) and Outpatient Care File (OPC).

At the medical centre level, the effects of the managed care effort on costs may be related to certain practices that are observable. In order to capture the magnitude of that relationship, a managed care penetration rate variable is included. This is the percentage of ambulatory patients treated at that facility who were assigned to primary care persons or team. As the managed care model is being implemented over time across the agency, the level of primary care oversight achieved at a facility is a means of capturing variation in managed care effort in this cross-sectional analysis. Additional controls at the provider level include size, measured by the number of operating beds (CDR data source), and an indicator for teaching hospitals, gauged by membership in the Council of Teaching Hospitals. Dummy variables for VISN are added in order to control for regional differences in cost such as wage differentials.

Although the author mentions that patients may be classified both by the hospitals they visit and by the physicians they frequent so that individuals within one hospital cluster are not grouped in the same way under physicians (cross-classification), in the VHA this type of cross-classification does not occur, since, in general, physicians and other clinicians operate within a single medical centre. However, patients may receive care from more than one medical centre during the year. This arrangement forms a multiple or cross-unit membership model, a special case of cross-classification.

c. THE RESULTS

Both OLS results and Multilevel Model results were presented for matter of comparisons. The main conclusions drawn from the multilevel models fitted are:

- From the unconditional (null) model, it can be seen that most of the variation in costs occurs at the patient level.
- Adding level-one, the DCG measure is the most powerful predictor of patient costs. Controlling for health status, age is also very highly significant as is sex, with men being more expensive. All of the random effects are also significant, indicating that the relationships between health status and cost and between age and cost differ by medical centre, and adding to the observed strength of the group.
- From the full model results, where the level two predictors are added, these have not ‘explained away’ the variance at level two: both the intercept and slope variances are still significantly different from zero. Also, the value of the variance component for the age slope coefficient has risen.
- None of the fixed main effects of the level two predictors are significantly different from zero. The managed care penetration rate shows some effect in its interaction with DCG. It is possible that the effect of the managed care variable is operating through its interaction with DCG such that costs rise with DCG risk scores but less so depending upon the extent to which the primary care model has been applied in the facility.
- The significance of the random terms indicates the appropriateness of applying a multilevel model. As anticipated, the standard errors in the multilevel model are much higher than in OLS. A number of coefficients that were significant in OLS are no longer significant in the final mixed model.

- Using the predicted residuals for the intercept and the corresponding C.I., in order to draw comparisons between institutions, the author concluded that there is some difference ib cost containment performance among institutions.
- Judging from the magnitudes of the fixed and random effects, there appears to be greater volatility in cost among institutions across Age than across DCG.
- Using the model coefficients estimates in order to make inferences on patient costs, the author calculates that the quantitative effect of managed care effect on patients cost.

d. CONCLUSIONS

The final conclusions and discussion by the author from the analysis conducted can be summarized into the following:

(a). The author mentions that multilevel statistical modeling provides an important complement to existing econometric techniques in analyzing health care provider costs and efficiency, compared to other techniques such as OLS analysis and panel data models

(b) The author uses the results with caution to perform comparisons between institutions and concludes that 'some deviations were identified and such departures from the average may well serve as useful diagnostics regarding performance evaluation that can inform both local and central management'.

(c) Although the results of multilevel analysis are judged as "inconclusive", the author comments that 'facilities more heavily penetrated by the primary care model appear to be slightly more effective at controlling the costs of their sicker patients'.

(d) The necessity of extensions of the model provided by adding repeated observations on individual providers, as well as more facilities to the data is mentioned, in order to empower the ability of conclusion making by the model.

What can be added in the above conclusions is that the particular article can be used as the basis for research in terms of health care costs for patients, in a variety of health systems, except from the particular one discussed here.

Stemming from the same area of health research and health economics, we refer to an article from Rice & Jones (1997). This article uses all the concepts of the multilevel model already discussed in the thesis (especially from the educational research) 'transformed' into the heath terminology perspectives. The issues described in the article are the basic multilevel model for patients nested within provider units, parameter estimation for both fixed and random part, residual estimation, some extensions to the basic model, such as cross-classification of patients within provider units and clinicians, and finally, proposals for applications of multilevel techniques in the area of health economics. This review can be seen as an introductory reference for everyone who wants to perform a multilevel analysis in health research.

4.2.3 Repeated Measures

The next example comes from the area of repeated measurements models. Repeated measures are, undoubtedly, a wide area of interest with applications in most of research issues. From the multilevel point of view, the basic characteristic of such models is that individuals form the 2nd level of measurement, where responses/events for the same individual are nested within them, forming the 1st level of analysis. This is exactly the case to be reviewed here (Wright, 1998).

‘Modeling Clustered Data in Autobiographical Memory Research: The Multilevel Approach’

a. INTRODUCTION

The article focuses on the issue of autobiographical memory research which involves recording several autobiographical memories for each of several people. These memories are not independent of each other, an assumption of the statistical procedures used in many cognitive psychology papers. To explore autobiographical memory, researchers often ask people to remember several different events. These events will, in some way, be sampled from each person's life. The events are usually different for each person. Conceptually, these events are nested within each person, that is each event happens only for a single person. Therefore the author, compared to all traditional research, states the advantages of the multilevel modeling approach in such research, using examples and concepts from other areas. The four models that are introduced and compared are the simple model that ignores the hierarchy, the aggregated model, the ANCOVA model and finally the multilevel model. Also for the purpose of the analysis the logit multilevel model is introduced.

When studying autobiographical memory, the events and the memories are unique for the person and this is important for theories of autobiographical memory.

The events are nested within the person in a manner analogous to the basic hierarchical structure in multilevel modeling.

b. DATA

The example used here is an example of autobio- graphical memory research by Burt et al. (1995). At the beginning of the summer holiday, 27 volunteers were given several rolls of film with the instructions that they should take photographs as they normally would. Several thousand photographs were taken. Some were excluded because they were out of focus or one of several depicting the same event (for example, many pictures of a wedding ceremony). Others excluded for missing values. The number of photographs kept for analyses is 1342. The number per person ranged from 23 to 147. Photographs are nested within people. The photographs were coded by whether they were of an activity, of participants, of a location, or some combination of these.

After the summer holiday, these photographs were presented to subjects for 10 seconds on a tachistoscope. Several distractors, taken from the researchers' own collections, were also presented. Over 97% of them were correctly recognized as foils. Subjects were asked if they could remember the event and the reaction times were recorded. If they could remember the event, they made several ratings about it, including the importance of the event and the level of emotion provoked by the event at the time. If they could not remember the event, they were asked whether they thought it was a distractor, whether there were too many similar events to be sure, or whether the picture did not provide sufficient cues for them to decide if it was their Only three relationships were examined by the author since they cover three of the main statistical procedures used by cognitive psychologists. The first is comparing reaction times by whether the event was correctly recognized, or not, and if not recognized the reason they gave. The second is the relationship between importance and emotion. Both variables were measured on 7-point rating scales and the Pearson correlation was found in the original analyses. The final relationship is between the type of photograph (of activities, participants and/or locations) and whether the subject could remember the event.

c. REACTION TIMES

There was a missing value for one reaction time, making the n=1341 for these analyses. The reaction times were highly skewed. There were some that were longer than 10 seconds, which was the duration that the picture was presented on the tachistoscope. These cases were recoded as 10 seconds. The natural logarithms of these data were taken so these transformations removed most of the skewness and made the distribution appear roughly Normal. The response variable was called ‘lntime’. About 65% of the holiday photographs were remembered (n=879). This will serve as the baseline category, with dummy variables for distractors (`dist' n=40), those not recalled because there were others that were too similar (`simil' n =222) and those for which the cues were insufficient (`cues' n =200). Besides `remembered' being the most frequent category, and therefore providing more reliable parameter estimates, the most logical contrasts are between `remembered' and each of the others. The final approach for this model is multilevel modeling which allows the intercept, here the times for remembered photographs, to vary among subjects but in addition it assumes that these response times are Normally distributed.

The results showed that the reaction times are faster for the remembered events and moreover that the variance of the subject level residuals is significant. In other words, that it is important to take into account the variation among subjects and the assumption of independence is not valid. Also, by fitting random variables in the random part of the model, it was shown that the variance for the non-remembered photographs was higher than for the remembered photographs.

d. IMPORTANCE AND EMOTION

In this step the relationship between the importance of an event (import) and people's emotional reactions to the event (emot) is examined

Subjects only made ratings for the events they remembered. Of the 879 remembered photographs, there was one missing value for emotion reducing the number of photographs to 878. Both ratings were on 7-point scales and they were treated as continuous interval measures

The multilevel approach treats photographs as a random sample from each person's summer but also treats subjects as a random sample from some population. The author fits the model with both random intercept and random slope and the results show that emotion gets higher when importance gets higher, as well as, the variation among subjects is largest when importance scores are either high or low, since the relationship between importance and subject variance is quadratic.

e. PREDICTING REMEMBERING

In the last part of the analysis the author tried to determine the best cues for retrieving memories. In other words, to predict whether a person remembers the event, using three dummy variables for whether a participant, a location and/or an activity are depicted in a photograph.

In the multilevel approach a logistic multilevel model was used as follows:

$$\operatorname{logit}\pi_{ij} = \beta_0 + \beta_1 part_{ij} + \beta_2 act_{ij} + \beta_3 loc_{ij} + u_j + e_{ij} \qquad (4.5)$$

where $part_{ij}$, act_{ij} and loc_{ij} are the three dummy variables as defined before and π_{ij} is the probability of remembering. The model provided a relatively good fit and moreover the assumption of binomial variation in the residuals of photographs was not rejected.

f. CONCLUSIONS

The author concludes once again by mentioning the advantages of multilevel techniques when analyzing autobiographical memory data. Also the advantages in other areas of psychology, where events are nested into individuals. It is pointed out, however that, in contrast to other areas, where the hierarchy is profound, ‘it is perhaps less intuitive that memories are nested within a person’.

4.2.4 Survey Research

The next two examples to be reviewed come from the area of survey research which is a wide area for applications for multilevel modeling techniques. The first (Rice et al, 1998) examine a clear structure of hierarchy where individuals are nested within households, which are also nested within geographical areas of residence. In the second case the effects of interviewers and respondents on the data collected from a survey research are examined (Hox, 1994).

1. 'The Influence of Households on Drinking Behaviour: A Multilevel Analysis'

a. INTRODUCTION

The purpose of the analysis is to examine the influence of household membership and area of residence on individual drinking behaviour. Before any statistical analysis a thorough theoretical review points out, one the one hand, the dangers of long heavy drinking compared to the intermediate drinking, and on the other examines the factors that affect drinking consumption and drinking behaviour of an individual. Using a large number of examples and references, the authors point out that grouping of individuals affects their drinking behaviour. Such groups can be of many kinds (social, religious, geographical) and can affect either in a positive or negative way. Households in which individuals are naturally nested are said to be one of the fundamental factors that affect the individual's behaviour as well as the area of residence. It is also discussed that such groups tend to make individuals that belong to them more "homogeneous" that individuals nested in different groups. Due to this hierarchy, the authors introduce and suggest multilevel model for data analysis and also introduce the basic models and parameter estimations, already explained in previous chapters.

b. DATA

Data from the 1993 Health Survey for England (HSE) was used for the study. Collection was performed throughout 1993 and on into early 1994, and consists of 17687 interviews with adults (aged 16 or over) living in 9700 households in England. The sample was distributed relatively evenly across the 14 English Regional Health Authorities and was obtained by sampling households from 2 or 3 electoral wards of residence in each Authorities' area. The survey is unusual in seeking responses from all adults in each household, providing a rare opportunity to explore effects within households. The 1993 survey focused on cardiovascular disease and associated risk factors, including alcohol consumption, as well as general health and various long-standing illnesses. There are well-known problems with the measurement of lifestyles in household surveys, associated with under-reporting and low response rates of heavy drinkers (Warner, 1978). Nevertheless, such information sources remain the method by which governments monitor their success in reaching drinking targets and are the only datasets large enough to answer the sort of questions addressed in this study. Alcohol consumption data was incomplete for 1139 individuals and these were excluded from the analysis. A further 1119 individuals were also excluded due to missing responses to the various explanatory variables used. This resulted in a total of 15429 individuals within 8737 households within 495 enumeration districts presenting for analysis. One-person households were retained in the analysis as they contribute to the estimates of the covariates of alcohol consumption and to area variations in consumption. The vast majority of the sample live in two-person households. The following variables were included in this analysis:

Personal characteristics: Gender, age.

Social environment/support: No. of persons in household, whether single or have a partner, perceived social support.

Health: Perceived stress.

Health related activity: Physical activity level.

Educational: Educational attainment.

Socio-economic: Social class, car ownership, whether or not economically active, whether in receipt of income support.

Various other potential explanatory variables were available in the HSE, for example, smoking status and self-reported general health. However, they were excluded since the relationship between these variables and drinking status was likely to be simultaneously determined.

The dependent variable in this analysis was an estimate of the number of units of alcohol drunk in a 1-week period based on respondents' answers to questions relating to the frequency of consumption and the number of units consumed on a usual occasion (Bennett, 1993). The response data displayed strong skewness and transformation to a more symmetrical distribution was sought by taking the natural logarithm. The distribution of alcohol consumption also often contains a significant proportion of zero values. Because of the problem of zero observations, a constant of unity was added to all observations before taking natural logarithms.

Empirical analyses of individual behaviour incorporating a household effect relied on the use of an explanatory variable often in the form of a dummy variable indicating whether or not other household members drink (or drink heavily), or a continuous measure of the average units consumed per individual within the household

The approach adopted here was to model both area and household effects as random components within a multilevel framework. To investigate the effects of area of residence and household membership on individual drinking behaviour a multilevel model including random components for individual, household and geographical area was specified. All explanatory variables were entered as dummy (0,1) variables except age, which is continuous. To ensure variations at each of the three levels were estimated at typical values of age, age was centred about its mean of 46 before being placed in the models.

c. RESULTS

After fitting the appropriate models, the results can be summarized as follows:

- Drinking declines with age while levels of consumption amongst the younger age groups are moderately high. In the case of women drinkers aged between 16 and 34 the average alcohol consumption was moderate as well as for men of the same age, but with higher average values for men.
- Males generally consume more alcohol than females and there is a quadratic age effect indicating that older people drink less. There is also an indication that there is a differential age effect for males and females shown by the significance of male by age interaction terms. It appears that although males drink more than females, the difference decreases with age. Single people tended to consume more alcohol.
- Of the general health and activities characteristics, individuals who report moderate and vigorous activities generally consume more alcohol compared to their baseline category of inactive individuals. There is no evidence in these data to suggest that a lack of social support or increased stress has an effect on levels of alcohol consumption. Individuals who are unemployed, inactive or working part-time are less likely to drink heavily than those engage in full-time employment (base-line category). However, the effect observed for inactive respondents decreases dramatically when household size and car ownership are included. This suggests that in the ``individual model'' the economically inactive was inappropriately picking up, an effect that should be properly attributed to household-level variables
- As for household characteristics, there was a clear household size gradient indicating strongly that larger households are associated with decreased

individual alcohol consumption. Car ownership is likely to be a reflection of individual current and capital household wealth and ownership of multiple cars appears to be associated with increased alcohol consumption.

- In all the models presented, the estimated variation at each of the three levels appears as statistically significant.
- The majority of the variance is attributed to differences between individuals (56%). However, 42% of variation occurs at the household level indicating that household membership and composition is very influential in determining inhabitants' consumption levels. Very little of the variation is attributed to area effects (2%), and speculation of strong geographical contextual effects of drinking behaviour appears unfounded in these data.
- There is a large amount of unexplained variation in individual alcohol consumption, which can be attributed to household membership. Further, little variation is attributed to differences in geographical area influences. The influence of household membership is nearly as great as that due to differences between individual characteristics in determining consumption of alcohol.

d. CONCLUSION

Conclusions of the author are mainly focused on suggestions about policies to reduce drinking consumption taking into account, of course, the results of the analysis. It is pointed therefore that focus will be given on both individuals and households, since their ‘contribution’ in drinking behaviour is, more or less, equivalently important. Geographical contextual effects were found to be minimal, however the authors mention that the definition of ‘area’ used in the analysis might be inappropriate.

2. ‘Hierarchical Regression Models for Interviewer and Respondent Effect’

a. INTRODUCTION

The example, reported by Hox (1994) concerns the effect of interviewers and respondents on survey results. Both respondents and interviewers have been recognized as a potential source of error (observational error) in survey interview data. Interviewer and respondent characteristics can have an important effect on the survey results and quality, and much methological research has been spent on the question how much interviewer and respondent bias is present in social survey data. Since respondents are nested within interviewers, in methological terms, this is clearly a multilevel problem. The specific example investigates how much interviewer and respondent characteristics influence the speed of interviewing (i.e. how many questions have been asked and answered in a given time period) All the traditional methods are mentioned, however the analysis is performed by applying hierarchical regression model techniques.

b. DATA, DATA ANALYSIS & RESULTS

The example data stem from a controlled field experiment on mode effects (De Leeuw 1992). In this example, data are analyzed from 515 respondents, who were questioned by 20 interviewers. Three data collection methods are compared: 221 of the interviews were conducted face-to-face, 219 by telephone using a pencil-and-paper questionnaire, and 75 by telephone using Computer Assisted Telephone Interviewing (CATI), all three using the same interviewers. The respondents were randomly assigned to the different collection methods: in both telephone conditions, they were randomly assigned to interviewers. Due to financial constraints, in the face-to-face condition, random assignment of respondents to interviewers was used within four broad geographical regions.

The dependent variable in the analysis is the total time needed for an interview. Because time measures generally have a skewed distribution, an inverse transformation is used, which transforms the variable, time, into the variable speed. Thus the dependent variable Y_{ij} is the speed of the interview measured in number of questions completed per minute. The research problem is whether interviewers differ in the speed with which they complete an interview. In addition, the author wishes to analyze which interviewer and/or respondent characteristics influence the interviewing speed. The explanatory variables $X_{pij}^{(1)}$ at the respondent level include two dummy variables indicating the three data collection methods: one contrast variable, tel (coded +1, -1), that compares the two telephone conditions to the face-to-face condition, and one contrast variable, cati, that compares the CATI-condition with the pencil-and-paper telephone condition (cati). The other respondent variables are respondent age (r-age) and loneliness (lonely), a measured by a multi-item scale. The explanatory variables $X_{qj}^{(2)}$ at the interviewer level are amount of previous interviewing experience, interviewer age (i-age), interviewer preference for telephone interviewing (pref.tel) and the interviewer's score on five personality scales: extroversion (extro), friendly disposition (friendly), conscientiousness (cons.), social assurance (soc.ass.), and ability to terminate awkward situations (term.).

Because the design is not completely orthogonal, the first step in the analysis is to inspect the correlations between respondent and interviewer explanatory variables. The correlations between respondents and interviewers are generally low, indicating that the partial othogonalization was successful. But, because the respondent and interviewer effects to be investigated are generally also small, it is safer to take these correlations into account in the analysis by modeling the interviewer effects conditional on the respondent variables.

The starting point for the model construction is the intercept-only model, which is a model with no explanatory variables. It is given by the following equation:

$$y_{ij} = \beta_0 + (u_{0j} + e_{ij}) \qquad (4.6)$$

This model contains the fixed regression coefficient β_0 for the grand mean for y_{ij} and the two variance estimates, σ_e^2 for the residual variance at the respondent level and σ_{uo}^2 for the residual variance at the interviewer level. In the example, β_0 is estimated as 3.19, indicating an overall interviewing speed of slightly more than three questions per minute. The respondent level variance σ_e^2 is estimated as 0.68 and the interviewer level variance σ_{uo}^2 as 0.11; this model produces an estimate for the intra-interviewer correlation ρ_1 of 0.14.

The next analysis step analyzes explanatory variables at the lowest (respondent) level as fixed variables; that is, without the corresponding variance components for the regression slopes. Equation is now:

$$y_{ij} = \beta_0 + \beta_p X_{pij}^{(1)} + (u_{0j} + e_{ij}) \quad (4.7)$$

In the subsequent model, the regression coefficients of the respondent variables are assumed to be random; that is they are assumed to vary between interviewers. This is described by the following equation:

$$y_{ij} = \beta_0 + \beta_p X_{pij} + (u_{pj} Z_{pij} + u_{0j} + e_{ij}) \qquad (4.8)$$

In equation (4.8), each regression slope β_p, has a corresponding random error term $u_{pj}Z_{pij}$. In the example data, the effect of the CATI contrast turns out to be not significant (p=0.25). The other explanatory respondent variables are all significant. Only the regression slope for the telephone contrast has a significant variance component. The conclusion is that the model for the correspondent effects might be simplified by dropping the CATI contrast altogether and assuming a random slope only for the telephone contrast. In this model, the respondent level variance is 0.53, and the interviewer level intercept variance is 0.08.

The next analysis step adds the explanatory variables at the interviewer level, giving

$$y_{ij} = \beta_0 + \beta_p X_{pij}^{(1)} + \beta_q X_{qj}^{(2)} + (u_{pj}Z_{pij} + u_{0j} + e_{ij}) \quad (4.9)$$

In the example, only three of the nine interviewer variables are significant; interviewer training, preference for telephone and extroversion.

The between-interviewers variation of the regression slopes for the telephone condition can be modeled by including interactions between the telephone condition variable and explanatory variables at the interviewer level. This gives the full model, which can be formulated as

$$y_{ij} = \beta_0 + \beta_p X_{pij}^{(1)} + \beta_q X_{qj}^{(2)} + \beta_{pq} X_{qj}^{(2)} X_{pij}^{(1)} + (u_{pj}Z_{pij} + u_{0j} + e_{ij}) \quad (4.10)$$

The only significant interaction effect is the interaction of the telephone contrast with the interviewer variable, social assurance. Because the interpretation of interaction effects requires that the corresponding simple effects are also included in the model, the (nonsignificant) interviewer variable, social assurance, is again added to the model. To aid interpretation, social assurance is centered around its overall mean of 61.8, and the interaction term is computed using the centered variable. In model (4.6), the residual variance at the respondent level is 0.68, and the residual variance at the interviewer level is 0.11. In model (4.8) σ_e^2 is estimated as 0.53; this can be expressed by saying that the respondent variables reduce the residual variance at the respondent level by 23%. Similarly, the interviewer variables can be said to reduce the residual variance at the interviewer level by 22%. The explanatory interviewer variables added in model (4.9) reduce the intercept variance by a further 43%. Adding the (nonsignificant) interviewer variable, social assurance, and its interaction with the telephone contrast (model (4.10)) reduces the intercept variance by 3% and the variance of the regression slope for the telephone contrast by 22%. Interpreting these variance reductions as explained variance, a comparison of the explained variance across the different models fitted shows that both the respondent and the interviewer variables explain a significant portion of the initial variance in interview speed. The interaction that is added in model (4.10) does not appear to explain much variance but, in fact, does explain a considerable proportion of the slope variance that appears in the previous model (4.9).

Calculations, however, of the explained variance, by using simply the variance components and residuals in a model with random slopes should be performed with caution. This is mainly because the variance components are not generally invariant under admissible linear transformations of the explanatory variables.

Using the models' deviances for a chi-square test shows that, in all comparisons of consecutive models, the more complicated models have a significantly better fit. Most of the regression coefficients are stable between different models. Although interviewer and respondent variables are correlated, adding the interviewer variables to the model does not appreciably change the regression slopes for the respondent variables. Only the intercept changes. The interpretation of the regression slopes is straightforward. Interviews take longer with older and lonely respondents, previously trained and extrovert interviewers are faster, and interviewers that have expressed a preference for using the telephone are also faster. The regression contrast for the telephone condition is coded –1 for the face-to-face condition and +1 for both telephone conditions. Its slope coefficient of 0.30 means that the telephone condition is faster by (2 x 0.30=) 0.6 questions per minute; at an overall average of 3.19 questions per minute, this means that telephone interviews are 19% faster. However, because the variable, telephone condition, is involved in an interaction, the interaction effect and the corresponding simple effects cannot be interpreted in isolation. When an interaction between two explanatory variables is involved, the simple regression coefficients for either of these variables reflect a conditional relationship, which is the relationship that holds when the other explanatory variable has the value zero. Because social assurance is centered around its overall mean of 61.8, the regression slope for the telephone contrast reflects the effect of this explanatory variable for interviewers with an average social assurance. To interpret the interaction, the author concludes that over the observed range of values for social assurance, telephone interviewing is faster than face-to-face interviewing. In the telephone interview, there is no relationship between social assurance and interviewing time, but in the face-to-face interview, interviewers with a higher social assurance tend to use more time. For an explanation, it could be hypothesized that the more personal situation in the face-to-face interview leads the less socially assured interviewers to adopt a task-orientated role, whereas the more socially assured interviewers adopt a social role, which uses up more time. In the more businesslike situation of the telephone interview, this differential role assumption does not take place.

c. CONCLUSIONS - REMARKS

The author ends with some critical conclusions, which can have both theoretical and practical affect on the area of survey research.

- Instrument effects, such as the type of questionnaire, is suggested to be a new level of analysis in a multilevel model. To do so, different type of questions could be treated as repeated measures within respondents i.e. all types could be asked to all respondents.
- Even though the interviewer effect is not of practical important for a survey (not included in the results), it should be measured in the hierarchical analysis as a control factor. Even a small intra-class correlation can cause large bias in the parameter estimates and therefore lead to misleading results
- Ideally, respondents should be assigned randomly within interviewers, otherwise respondent and interviewer characteristics could be confounded. However, in survey research companies, where this is not a frequent situation, multilevel techniques could solve the problem of confounding since respondent variables would be controlled for interviewer effects and vice versa.

Although survey research appears to be an area where multilevel techniques could be readily used, both in theory and in practice, compared to other research areas such as educational research, these techniques are still rarely preferred to the classical techniques, especially in survey research companies.

4.3 Conclusions of the Chapter

The main conclusion which can be drawn from the discussion of this Chapter is that Multilevel Analysis Techniques are extensively applied from a great range of scientists and researchers, not only in the area of Educational Statistics, where the hierarchical structure of data is profound, but also in less profound research areas such as spatial statistics, health research, survey research, repeated measures and meta-analysis. Moreover, as discussed also by the authors, the use of multilevel instead of more classical statistical techniques has significant advantages both in the precision and the statistical accuracy of the results, as well as in the interpretation of the hierarchical structure of the data. However, in order to enhance more on the advantages of the multilevel techniques, in the following Chapter we will perform a practical application of Multilevel Analysis in a real Greek educational dataset referring to the General Admission Grade of students to the National Exams.

CHAPTER 5

5 AN APPLICATION TO GREEK EDUCATIONAL DATA

In this Chapter we apply a statistical analysis to a real dataset obtained 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 assess the effectiveness of Greek Lyceums and to detect the factors that affect students' performance in high school, especially according to the National Exams for their access in the National Universities and Technical Institutions.

The hierarchical structure of the data is profound. Students are nested in schools and schools are nested in prefectures, so the influences of this grouping must be taken into account. Thus, as described in previous Chapters, multilevel modeling is required for the analysis of such data, where students represent the 1st Level, schools the 2nd Level and prefectures the 3rd Level of analysis. It should be noticed that this is the first time multilevel analysis is carried out in the particular educational system of access which is described in the following paragraph.

The scope, therefore, of this Chapter is to apply Multilevel Techniques as described in previous chapters, in this particular situation, to discuss firstly on the results of the analysis themselves but, moreover to discuss on the advantages and the applicability of the use of Multilevel Analysis in this real example.

5.1 Description of the Educational System

In order to detect the variables of interest for the analysis, let us give a brief description of the Greek educational system in the Greek Lyceums (at the years of interest 2006-2009) as described in the website of the Greek Ministry of Education, Lifelong Learning and Religious Affairs (www.ypepth.gr). Studies in Lyceum are optional and last for three years. All students during their studies are asked to choose one of the three "Scientific Orientations" (Human Sciences, Exact Sciences, and Technical Sciences) which, more or less, "direct" their studies. Technical scientific orientation is further divided in two cycles of studies (Technology & Production and Informatics & Services). In the third class of Lyceum, students are examined at a

National level, in four subjects of orientation and two (or three-as will be described later) subjects of general education. According to their marks on these six (or seven) subjects students receive the so-called “General Admission Grade” (ranging from 0 to 20) which is the first basis for the calculation of the “Final Admission Grade (“moria”)” used for the access of students to the National Universities and Technical Institutions. The calculation of the “Final Admission Grade (“moria”)” is somehow complicated and depends on the “weight” given in each of the six (or seven) subjects according to the new Scientific Area (“Pedio”) chosen by the students prior to their potential access to a University or Technical Institution. There are five “scientific areas” and are chosen by the students who apply for the selection process for their access to a University or Technical Institution independently of their initial choice of “Scientific Orientation”. Students of the first four “scientific areas” are examined in six subjects, while students of the 5^{th} “scientific area” are examined in seven. Moreover, according to the National Exams Regulations, in order a student to be eligible to apply for the selection process for the access to a University or Technical Institution (and, therefore, in order the “Final Admission Grade (“moria”)” to be calculated) the “General Admission Grade” of the student has to be at least 10. These restrictions in the calculation of the “Final Admission Grade” (not calculated for all students and, when it is calculated, the formula is different for each student) should be taken into account in the analysis.

5.2 Variables

Having in mind all the above considerations, we can now refer to the variables that will be used in the analysis. The “General Admission Grade” will be the response variable, since it is calculated in a “common” basis for all the students, and the potential independent factors that might affect students’ performance and schools’ effectiveness are the Gender of the student, the Scientific Orientation of Studies, the Type of School and the Year of Examination in which the General Admission Grade corresponds. Also possible interactions between factors will be examined. We should notice that all potential explanatory variables are categorical and no appropriate continuous explanatory variables can be used. More specifically:

Response Variable

The General Admission Grade is the most appropriate measure to be used as the response variable in order to detect student's performance and schools' effectiveness, since it is calculated for the whole of students directly by the performance of the students in the National Exams in the six (or seven) pre-determined subjects. The two subjects of general education are "New Hellenic Grammar" (compulsory) and one of "History of Modern World", "Mathematics & Elements of Statistics" or "Biology & Physics". For the 5^{th} scientific area the 7^{th} subject is "Elements of Economical Theory". Also, according to the initial Scientific Orientation, the four subjects of orientation are:

- "Ancient Greek", "Latin", "New Hellenic Literature" and "History" for the Human Sciences Orientation.
- "Biology", "Mathematics", "Physics" and "Chemistry" for the Exact Sciences Orientation.
- "Electrology", "Mathematics", "Physics" and "Chemistry/Biochemistry" for the 1^{st} cycle (Technology & Production) of Technical Sciences Orientation.
- "Mathematics", "Physics", "Elements of Business Administration" and "Applications Development in Programming Environment" for the 2^{nd} cycle (Informatics & Services) of Technical Sciences Orientation.

The score of the General Admission Grade ranges from 0 to 20 and, according to initial goodness-of-fit analysis, the original scores were used, without the need of any transformation.

Explanatory Variables

As mentioned before, there is no continuous explanatory variable used in the analysis, since there were no such appropriate available variables. This should be taken into account, since as mentioned in almost all previous references (Goldstein, 1993 for instance), an explanatory variable is usually used as an "adjustment" for the existing achievements of the students. The categorical explanatory variables (factors) used in the analysis are the following:

1. The Type of School is a profound potential factor (2^{nd} Level variable) to be used in the analysis. There are two types of schools of interest, Public and

Private schools. The variable indicating the type of school is a dummy dichotomous variable coded 1 for public schools and 0 for private.

2. The Gender of students is also an important explanatory variable (1st Level variable). It is also a dummy variable coded 1 for male students and 0 for female.
3. The initial Scientific Orientation of Studies is also an interesting potential factor which might affects the response variable. It was preferred from the "Scientific Area (pedio)" as possible explanatory variable because it exists for all students, while the latter is present only for the "eligible" students. Also it should be mentioned that Technical Scences Orientation was examined as it is and was not separated into the two cycles of studies, due to the very small number of students examined in the first cycle (Technology & Production). The categorical variable "Scientific Orientation" has three categories, so we need 2 dummy variables in order to make the appropriate comparisons between the three orientations. More specifically, the 1st dummy variable is coded 1 for the 1st scientific orientation (Human Sciences) and 0 for all the others. The 2nd dummy variable is coded 1 for the 2nd scientific orientation (Exact Sciences) and 0 for all the others, while the 3rd scientific orientation (Technical Scences) is the base category.
4. Another important factor to be examined for its effect on the response variable is the Year of Examination. We should notice that, in order to avoid duplications, if the same student had taken the National Exams more than once within the time period of interest (2006-2009) only their first General Admission Grade was taken into account. For example, if a student had examined in the years 2006, 2007 and 2008, only the score in 2006 was used in the analysis and the two other scores were omitted. The categorical variable "Year of Examination" has four values (2006, 2007, 2008, 2009) so three dummy variables were needed in order to make the appropriate comparisons between years. The last year (2009) was used as the base category.

5.3 Descriptive Statistics

Before any further analysis we present some descriptive statistics for our data concerning the number of units in each level of analysis as well as the General Admission Grade according to the explanatory variables of the analysis.

As shown in the following table (Table 5.1) in the total dataset there are 325724 students (1st Level units), nested within 1387 schools (2nd Level units), nested within 54 prefectures (3rd Level units). In the year 2006 there are 98333 unique students, nested within 1345 schools nested within 54 prefectures. In 2007 77311 unique students nested within 1354 schools, in 2008 76917 unique students nested in 1360 schools and in 2009 73163 students nested within 1365 schools. As mentioned in the previous Chapter, all students participating in the analysis are unique, that is only the General Admission Grade from their first year of examination was taken into account. Of course, the vast majority of the school units and all 54 prefectures participated in the analysis every year.

Table 5.1: No of units (students/schools/prefectures) for each Year of Examination

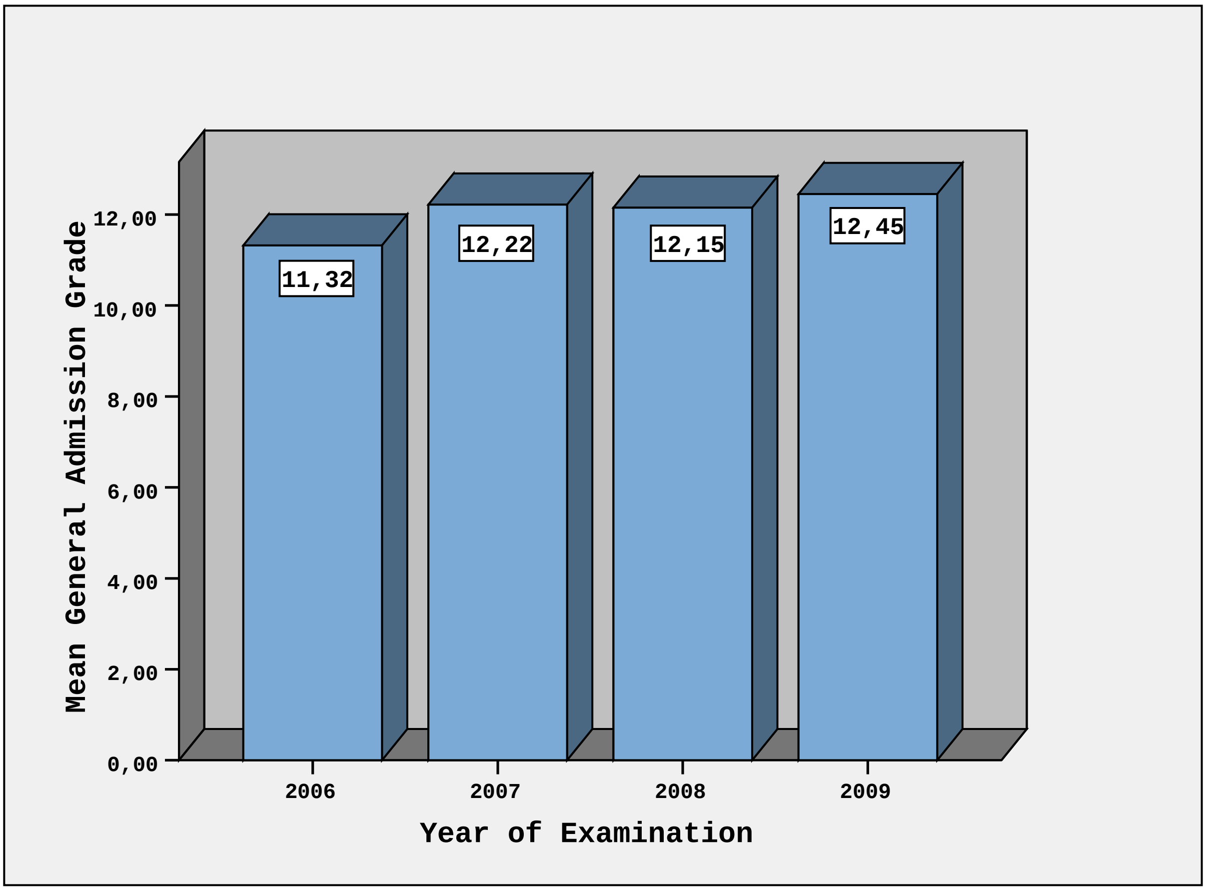
Year of Examination	No of Students	No of Schools	No of prefectures
2006	98333	1345	54
2007	77311	1354	54
2008	76917	1360	54
2009	73163	1365	54
Total	**325724**	**1387**	**54**

The following tables (Table 5.2-5.5) and the corresponding figures (Figure 5.1-5.4) refer to the descriptive statistics of the General Admission Grade according to the explanatory variables. We should notice that 15663 students were omitted from the analysis because of missing of their General Admission Grade and, therefore, the final analysis was based on 310061 students. The total mean for General Admission Grade for the whole period of interest (2006-2009) is 11.99 (±4.66).

Table 5.2: Descriptive Statistics for the General Admission Grade according to the Year of Examination

Variable	Year of Examination	Mean	Std. Dev.	Minimum	Maximum	N of cases
	2006	11.32	4.43	.02	19.93	92814
General	2007	12.22	4.56	.04	19.91	73142
Admission	2008	12.15	4.78	.03	19.88	73513
Grade	2009	12.45	4.84	.05	19.95	70592
	Total	**11.99**	**4.66**	**.02**	**19.95**	**310061**

Figure 5.1: Figure for the Mean General Admission Grade according to the Year of Examination



As shown in the above table (Table 5.2) and the corresponding figure (Figure 5.1) the highest mean score of the General Admission Grade (12.45) with respect to the year of examination was observed by students examined in 2009. In contrast, students examined in 2006 seem to have the worst performance (mean General Admission Grade=11.32). This result can be probably explained by the fact that 2006 was the first year of application of the particular education system of access, so students were not yet adapted to this new system.

Table 5.3: Descriptive Statistics for the General Admission Grade according to the Gender of Students

Variable	Gender	Mean	Std. Dev.	Minimum	Maximum	N of cases
General	Male	11.37	4.81	.03	19.90	140621
Admission	Female	12.50	4.47	.02	19.95	169440
Grade	**Total**	**11.99**	**4.66**	**.02**	**19.95**	**310061**

Figure 5.2: Figure for the Mean General Admission Grade according to the Gender of Students

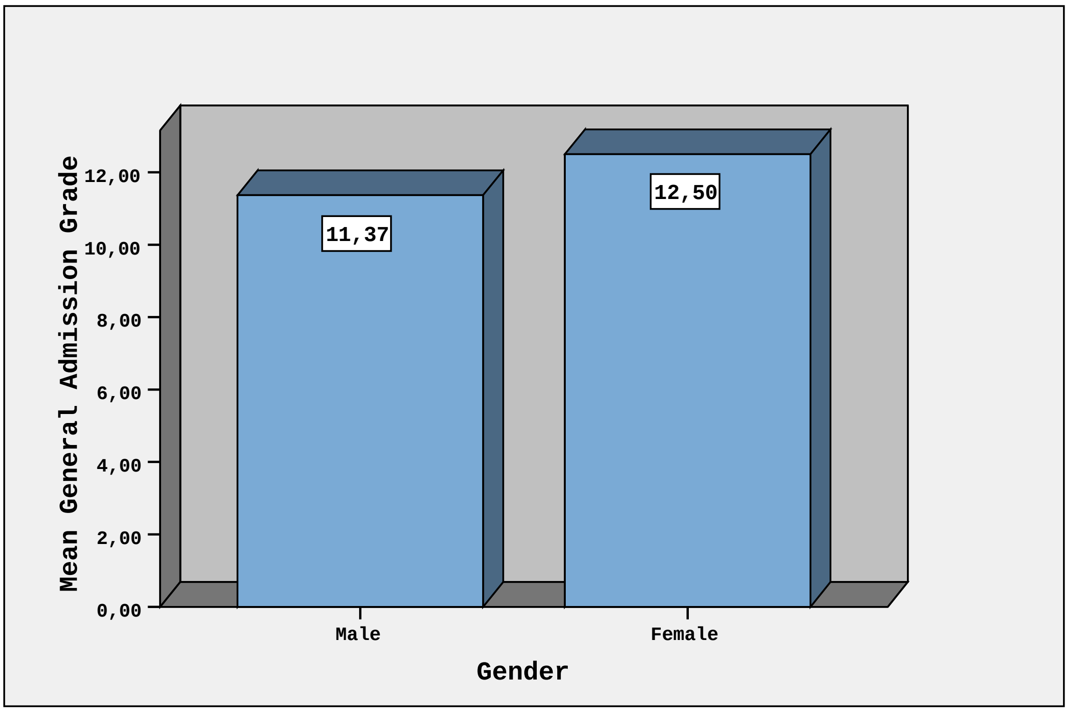
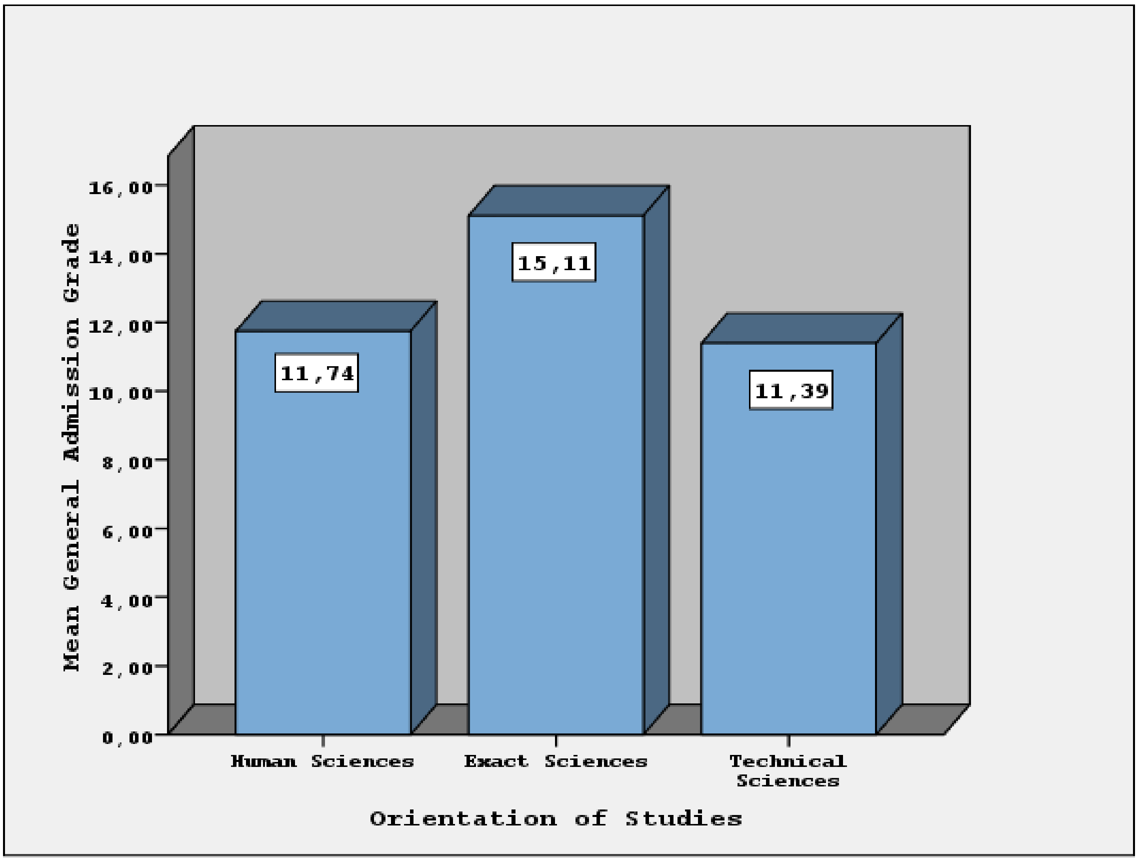


Table 5.3 and the corresponding figure (Figure 5.2) show that female students perform much better than male students, since the mean score for General Admission Grade for girls is 12.5 versus only 11.37 for boys. However, we will elaborate more on this result in the following Chapter of multilevel analysis, since this difference might be also related to the differences in scientific orientations chosen by girls and boys.

Table 5.4: Descriptive Statistics for the General Admission Grade according to the Scientific Orientation of Studies

Variable	Orientation of Studies	Mean	Std. Dev.	Minimum	Maximum	N of cases
General Admission Grade	Human Sciences	11.74	4.61	.03	19.93	121258
	Exact Sciences	15.11	3.92	.07	19.92	38117
	Technical Sciences	11.39	4.57	.02	19.95	150686
	Total	**11.99**	**4.66**	**.02**	**19.95**	**310061**

Figure 5.3: Figure for the Mean General Admission Grade according to the Scientific Orientation of Studies



From the above table (Table 5.4 and the corresponding figure (Figure 5.3) we can easily detect that the performance of the students who have chosen the "Exact Sciences" scientific orientation is much higher than those who have chosen the two other orientations ("Human Sciences" and "Technical Sciences"). The mean General Admission Grade for the Exact Sciences scientific orientation is 15.11 versus 11.74 and 11.39 for the Human Sciences and Technical Sciences orientation respectively. However, the highest score (19.95) was accomplished by a student of the Technical Sciences scientific orientation. In order to give a first explanation for the superiority of the students of Exact Sciences orientation, as regards their General Admission Grade performance, we should mention the relative small number of students who choose this scientific orientation, as well as the fact that this orientation is chosen by students focused mainly on health studies which demand very high entrance exams scores from the students.

Table 5.5: Descriptive Statistics for the General Admission Grade according to the Type of School

Variable	Type of School	Mean	Std. Dev.	Minimum	Maximum	N of cases
General	Public	11.82	4.64	.03	19.95	289287
Admission	Private	14.33	4.37	.02	19.90	20774
Grade	**Total**	**11.99**	**4.66**	**.02**	**19.95**	**310061**

Figure 5.4: Figure for the Mean General Admission Grade according to the Type of School

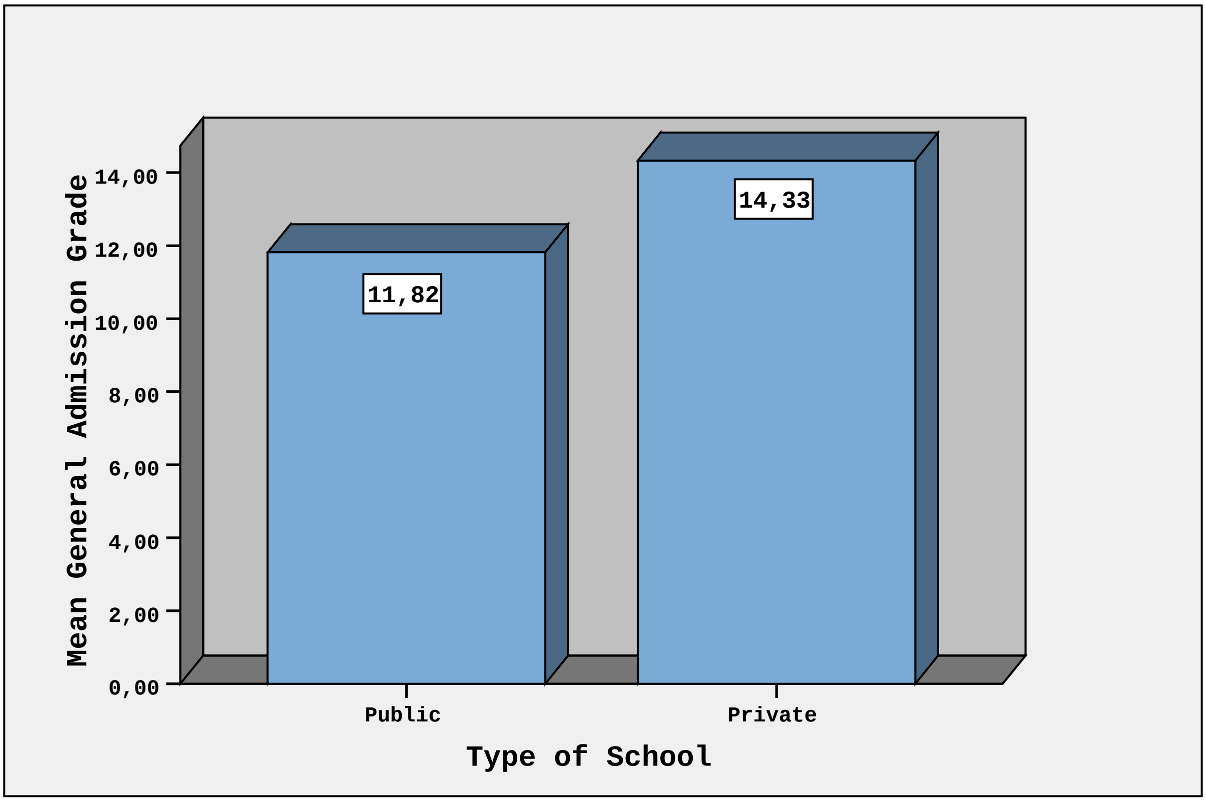


Table 5.5 and the corresponding figure (Figure 5.4) show the descriptive results for the General Admission Grade according to a 2nd Level explanatory variable (type of student). We can see from the results an obvious higher performance for students in private schools compared to students from public schools (mean scores are 14.33 versus 11.82 respectively). However, the highest score for General Admission Grade (19.95) was accomplished by a student in a public school. Again, we have to take into consideration the relatively very small number of students attending private schools. In the following Chapter of multilevel analysis we will also examine the possible cross-level interaction between the type of school and some 1st level variables (gender and scientific orientation).

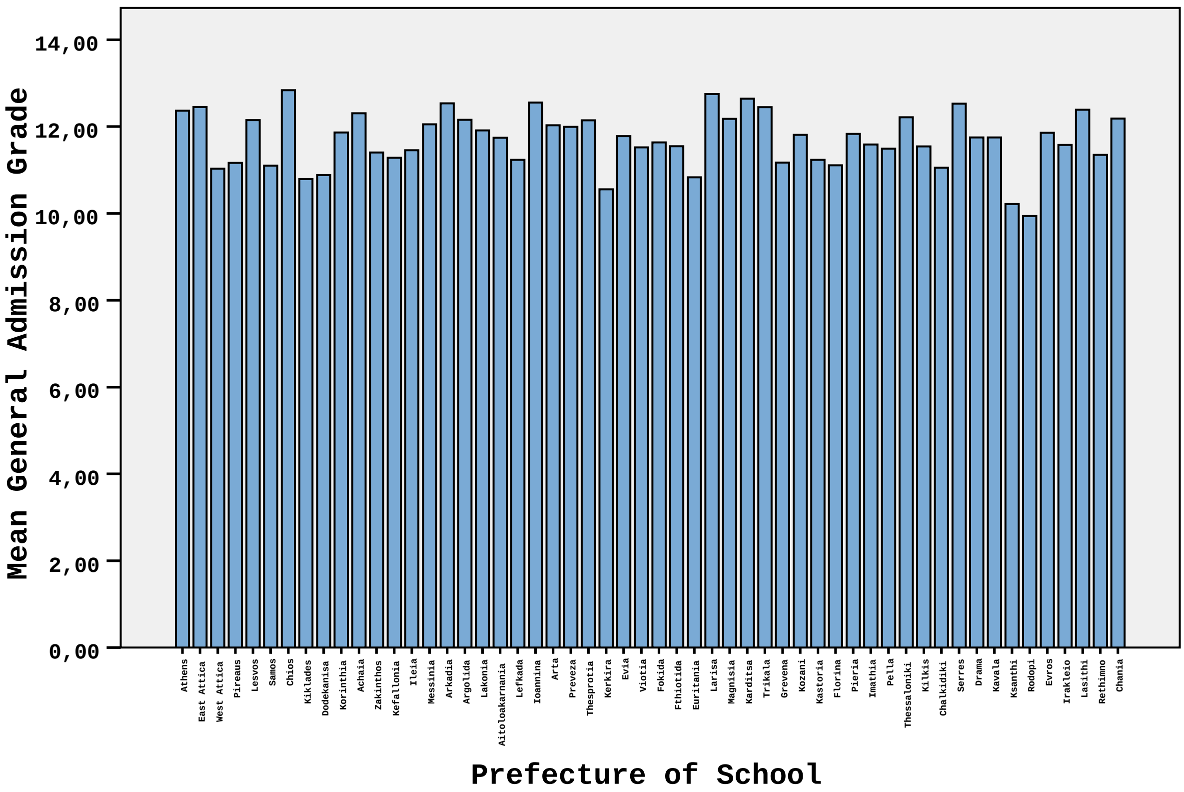
The following table (Table 5.6) and the corresponding figure (Figure 5.5) present the descriptive statistics for the General Admission Grade for all 54 prefectures in the analysis. As shown in the table, the prefecture with the best mean performance within the period of analysis is Chios (mean score for Admission Grade for the 1416 students of Chios is 12.84). The prefecture with the second highest mean score is Larisa with mean score 12.75 in a total of 8878 students. The highest score of Admission Grade for all these years (19.95) was accomplished by a student in Magnisia in the year 2009. On the contrast, the prefecture with the lowest mean score is Rodopi with mean score only 9.94 in a total of 2234 students.

Table 5.6: Descriptive Statistics for the General Admission Grade according to the Prefecture of School

Prefecture of School	Mean	Std. Dev.	Minimum	Maximum	N of cases
Athens	12.37	4.53	.03	19.90	77836
East Attica	12.45	4.62	.07	19.90	14120
West Attica	11.03	4.58	1.18	19.81	3710
Pireaus	11.17	4.54	.14	19.76	14999
Lesvos	12.15	4.63	1.37	19.85	2464
Samos	11.10	4.55	2.62	19.52	1027
Chios	12.84	4.44	.05	19.80	1416
Kiklades	10.79	4.55	.48	19.78	2651
Dodekanisa	10.89	4.82	.70	19.75	5233
Korinthia	11.87	4.78	.04	19.75	3940
Achaia	12.31	4.62	.03	19.83	9512
Zakinthos	11.40	4.93	1.82	19.65	1101
Kefallonia	11.28	4.79	.48	19.67	1124
Ileia	11.46	4.77	.53	19.78	3903
Messinia	12.05	4.74	1.07	19.81	4425
Arkadia	12.54	4.65	1.48	19.83	2582
Argolida	12.16	4.76	1.73	19.77	3044
Lakonia	11.91	4.85	1.45	19.90	2342
Aitoloakarnania	11.74	4.76	.63	19.75	6728
Lefkada	11.23	4.98	1.47	19.62	742
Ioannina	12.55	4.60	1.13	19.73	4644
Arta	12.03	4.70	1.33	19.77	1974
Preveza	12.00	4.64	1.76	19.88	1786
Thesprotia	12.15	4.58	.04	19.75	1186
Kerkira	10.56	4.63	.15	19.88	3098
Evia	11.78	4.69	.88	19.81	5909
Viotia	11.52	4.80	.03	19.85	3213
Fokida	11.64	4.79	1.18	19.91	940
Fthiotida	11.55	4.87	1.48	19.66	4687
Euritania	10.83	4.69	1.32	19.63	428
Larisa	12.75	4.60	.30	19.78	8878
Magnisia	12.18	4.78	.70	19.95	5978
Karditsa	12.64	4.66	.82	19.77	3421

Trikala	12.45	4.67	.68	19.92	4267
Grevena	11.17	4.78	1.01	19.63	782
Kozani	11.81	4.63	.83	19.77	5701
Kastoria	11.23	4.89	1.30	19.70	1876
Florina	11.11	4.87	1.12	19.60	1712
Pieria	11.83	4.89	.95	19.88	3814
Imathia	11.59	4.60	1.30	19.73	4212
Pella	11.49	4.64	.18	19.67	4152
Thessaloniki	12.21	4.53	.02	19.85	33122
Kilkis	11.54	4.60	.48	19.73	1887
Chalkidiki	11.06	4.62	.62	19.62	2257
Serres	12.53	4.52	.03	19.78	4592
Drama	11.75	4.67	.69	19.80	2862
Kavala	11.75	4.62	1.72	19.78	3870
Ksanthi	10.22	5.20	.40	19.73	2711
Rodopi	9.94	5.61	.28	19.93	2234
Evros	11.86	4.69	.97	19.76	3523
Irakleio	11.58	4.62	.50	19.87	8952
Lasithi	12.39	4.39	1.27	19.71	1910
Rethimno	11.35	4.59	1.37	19.65	2195
Chania	12.19	4.56	1.60	19.87	4389
Total	**11.99**	**4.66**	**.02**	**19.95**	**310061**

Figure 5.5: Figure for the Mean General Admission Grade according to the Prefecture of School



5.4 Multilevel Data Analysis

In the previous Chapter we have presented some descriptive results for our response variable which is the General Admission Grade according to the explanatory variables of interest (gender of student, scientific orientation of studies, year of examination and type of school). Although a brief idea of the differences in the performance of the students according to these factors was taken, we will now try to introduce statistical models in order to estimate and evaluate the effect of each factor on the score of the student. The hierarchical structure of our data is profound. Students are nested within schools and schools are nested within prefectures. Therefore, in our analysis we will try to perform Multilevel Statistical Models described in previous Chapters, in order to take advantage of this structure and present more accurate estimates. Of course, the statistical significance of the use of higher level models was tested.

The selection method of the models will be the forward method, meaning that we will start from the simple null model and we will add successively factors and test whether the more advanced model is a significant improvement of the previous one. To estimate the significance of the improvement we will carry out a likelihood-ratio test by comparing the deviances of the two models and test the difference in deviances referring to tables of the chi-square distribution. For the estimation of the parameters of the model we will use the REML (Restricted Maximum Likelihood) method. As a practical test for the significance of a parameter we will simply check if the parameter estimate is more than three times the estimate of its standard error.

Model 1

At the first stage of the analysis we simply fit the null 2-level model with the General Admission Grade as the response variable and no explanatory variables (except of course from the constant parameter). This first model (Model 1) is simply used as the basis of the analysis and the parameter values are displayed in table 5.7. The table only contains the constant estimate, the estimates for the level-1 variance (between students) and the level-2 variance (between schools), as well as the standard errors of the estimates. It also contains the deviance (-2*log(likelihood)) of the model in order to perform the appropriate likelihood-ratio tests.

Before any further discussion we should check whether this null 2-level (Model 1) is a significant improvement of the null 1-level model which can be produced if we omit the level-2 variance σ^2_{u0} (between schools). The deviance of the model that does not contain the level-2 variance is 1834534 and is compared to the deviance of Model 1 which is 1802536. The difference is referred to tables of the chi-square distribution with one degree of freedom and is found to be highly statistically significant. Therefore, this is the first evidence that the use of a 2-level model containing also a between-schools variance estimate is a significant improvement of the simple 1-level model which only analyzes the between-students differences. In other words, we have proven that by taking into account the hierarchical structure of the data (students nested in schools) we can conclude to more accurate results.

The null 2-level model can also be used in order to estimate the so-called 'intra-class correlation' which is given by the formula $\rho = \frac{\sigma_{u0}^2}{(\sigma_{u0}^2 + \sigma_{e0}^2)}$ and estimates the proportion of the total variance which is between-schools. In our case the intra-class correlation is $\rho = \frac{3.419}{(3.419 + 19.311)} = 0.15$ meaning that almost 15% of the total variance is attributable to school traits. This proportion is relatively high, giving another evidence that the effect of 2-level units (schools) on the total performance of students should be examined.

Table 5.7: Parameter estimates for Model 1

Parameter	Estimate	Std. Error
Fixed:		
Constant	11.335	0.051
Random:		
σ_{u0}^2 (between schools)	3.419	0.142
σ_{e0}^2 (between students)	19.311	0.049
-2*log(likelihood)	1802536	

Model 2

The first explanatory variable we add in the model is a 1st level variable, the gender of the student. This is a dummy variable coded 1 for male students and 0 for female, so female will be the base category. The estimates of the parameters for this new model (Model 2) are given in the following table (Table 5.8). The difference of the deviances of this model (Model 2) and the null model (Model 1) is 1802536-1797168 and is apparently significant referring to the tables of the chi-square distribution with one degree of freedom. So Model 2 is a significant improvement from the previous one. The parameter estimate for the male category is also significant since the estimate of the standard error of the parameter is less than a third of the parameter estimate. Also the value of the estimate is negative (-1.165) implying that the score of the male students in the General Admission Grade is significally lower than the score of the female students, so, as it was expected, the gender difference is in favor of girls. As far as the random parameters are concerned, we observe that both the level-1 (between students) and the level-2 (between schools) variances are slightly decreased with the inclusion of the gender in the model.

Table 5.8: Parameter estimates for Model 2

Parameter	Estimate	Std. Error
Fixed:		
Constant	11.869	0.051
Gender (Male)	-1.165	0.016
Random:		
σ^2_{u0} (between schools)	3.363	0.139
σ^2_{e0} (between students)	18.979	0.049
-2*log(likelihood)	1797168	

Model 3

The new explanatory variable we introduce in the model is another 1st level variable, namely the scientific orientation of the studies of the student. This factor has three categories and the Technical Sciences orientation is the base category. Therefore, two parameters are estimated in the new model (Model 3), one for the Human Sciences and one for the Exact Sciences orientation, and the new estimates are presented in table 5.9. By comparing the deviances of the new model (1779371) and the previous one (1797168) we conclude that the difference of the deviances is highly significant referring to the tables of the chi-square distribution with two degrees of freedom, so we keep Model 3 as a better approach. Also all the parameter estimates of the model are statistically significant since the estimates are more than three times of the estimates of the standard errors. As we observe from the parameters, students examined for Human Sciences scientific orientation have slightly better performance than students examined for Technical Sciences orientation. However, students of Exact Sciences orientation seem to have much higher score in the General Admission Grade compared to students of Technical Sciences orientation, since the parameter estimate for this category is relatively high (3.254). Also, by adding the new variable, the values of the parameter estimates for the constant and male category have slightly decreased, as well as the level-1 (between students) and the level-2 (between schools) variances.

Table 5.9: Parameter estimates for Model 3

Parameter	Estimate	Std. Error
Fixed:		
Constant	11.392	0.050
Gender (Male)	-1.019	0.017
Scientific Orientation (Human Sciences)	0.159	0.018
Scientific Orientation (Exact Sciences)	3.254	0.025
Random:		
σ_{u0}^2 (between schools)	3.023	0.126
σ_{e0}^2 (between students)	17.923	0.046
-2*log(likelihood)	1779371	

Model 4

In the new model (Model 4) we describe in the following table (Table 5.10) we have not added a new explanatory variable. Instead, we add the interaction between the gender of the student and the scientific orientation of their studies, in order to detect possible differences in the performance of males and females according to their scientific orientation. Since the first variable (gender) has two categories and the other (scientific orientation of studies) has three, only two parameters need to be estimated, one for male with Human Sciences orientation and one for male with Exact Sciences orientation. The deviance of the new model is 1778050 and by comparing it to the deviance of the previous model (1779371) the difference is significant referring to the tables of the chi-square distribution with two degrees of freedom. So we can consider Model 4 as a significant improvement for our analysis. Also all parameter estimates of the model are significant since the estimates of the parameters are more than three times of the estimates of their standard errors.

By examining the two new parameters for the interaction term we can detect an important result which alterates our previous conclusions. The parameter estimate for male students of Human Sciences orientation is negative (-0.864). Therefore, although for the total of students examined for Human Sciences orientation the performance is better than those examined for Technical Sciences orientation, for boys the score in the General Admission Grade seem to be worse if they have chosen Human Sciences scientific orientation instead of Technical Sciences. In other words male students seem to have better performance for the Technical Sciences orientation than for Human Sciences, while all students, independently of gender, perform better in the Exact Sciences scientific orientation than in the other two. As far as the random parts of the model are concerned, we observe a further slight decrease of both the variance estimates with the inclusion of the interaction term in the model

Table 5.10: Parameter estimates for Model 4

Parameter	Estimate	Std. Error
Fixed:		
Constant	11.310	0.050
Gender (Male)	-0.868	0.023
Scientific Orientation (Human Sciences)	0.424	0.023
Scientific Orientation (Exact Sciences)	2.896	0.033
Gender*Scientific Orientation (Male/Human Sciences)	-0.864	0.037
Gender*Scientific Orientation (Male/Exact Sciences)	0.993	0.050
Random:		
σ_{u0}^2 (between schools)	2.934	0.122
σ_{e0}^2 (between students)	17.849	0.045
-2*log(likelihood)	1778050	

Model 5

The explanatory variable we add in the new model (Model 5) is the year of examination of the student. We have already noted that, in order to avoid duplications and repeated measures for the students, each student is unique and measured only in the first year of their examination. The new variable is another 1st level variable with four categories and the base category is the year 2009. The estimations for the three new parameters, as well as all the other new estimations for the parameters introduced in previous steps are presented in table 5.11. The deviance of the new model is 1775015 and if we compare it to the deviance of the previous model (1778050) we conclude that the difference is highly significant referring to the tables of the chi-square distribution with three degrees of freedom. Therefore, Model 5 is a significant improvement compared to the previous model. Also, once again, all the parameter estimates of the model are highly statistically significant since the estimates are more than three times of the estimates of their standard errors. We can observe from the table that all the new parameter estimates for the three years are negative (-1.078, -0.247 and -0.328 for years 2006, 2007 and 2008 respectively) and therefore the performance of the students according to their score in the General Admission Grade is higher in 2009 than in all other years of examination. The worst year of all seems to be 2006, since the parameter estimate is relatively high (-1.078). This result seems rational considering the fact that 2006 was the first year of application of the particular educational system, so students were probably not yet adapted to this new system. The parameter estimates for all other fixed parts of the model have not altered dramatically, and so have not alterated the conclusions made in previous steps. For the random parts of the model we should mention another slight decrease in the level-1 (between students) and the level-2 (between schools) estimates of variances after adding the new explanatory variable in Model 5.

Table 5.11: Parameter estimates for Model 5

Parameter	Estimate	Std. Error
Fixed:		
Constant	11.792	0.052
Gender (Male)	-0.877	0.023
Scientific Orientation (Human Sciences)	0.379	0.023

Scientific Orientation (Exact Sciences)	2.869	0.033
Gender*Scientific Orientation (Male/Human Sciences)	-0.864	0.036
Gender*Scientific Orientation (Male/Exact Sciences)	1.001	0.050
Year of Examination (2006)	-1.078	0.021
Year of Examination (2007)	-0.247	0.022
Year of Examination (2008)	-0.328	0.022
Random:		
σ^2_{u0} (between schools)	2.895	0.121
σ^2_{e0} (between students)	17.674	0.045
-2*log(likelihood)	1775015	

Model 6

So far, in all the previous steps we have introduced 1st level explanatory variables referring only to the student units (gender of student, scientific orientation of studies and year of examination), as well as the interaction term between some of them (gender*scientific orientation). However, we have seen by the null model and the calculation of the intra-class correlation that almost 15% of the total variation is attributable to school units. So, in the new model (Model 6) we will add a level-2 explanatory variable, the type of the school. This is a dummy variable coded 1 for public and 0 for private schools, so private schools will be the base category. The estimates of the parameters for this new model (Model 6) are given in table 5.12. The difference of the deviances of the new model (Model 6) and the previous one is 1775015-1774905 and is highly significant referring to the tables of the chi-square distribution with one degree of freedom. So, we can keep Model 6 and the inclusion of the new level-2 variable as a significant improvement from the previous model. Also, all the parameter estimates for the fixed and the random parts of the model are significant since the estimates of the standard error of the parameters are less than a third of the parameter estimates. We can observe that the value of the estimate of the new variable for the category “public” is negative (-1.844). From this we can conclude that the score of the students of public schools in their General Admission Grade is significally lower than the score of the students from private schools, so, as it was expected, the type of school difference is in favor of private schools. All the other parameter estimates for the fixed part of the model have no important alterations and therefore, all conclusions made in previous steps hold. As for the random parameters of the model, we observe that, as was expected, the estimate of the level-2 (between schools) variance has decreased after adding a new 2nd level explanatory variable in the model, while the estimate for the level-1 variance has remained almost the same.

Table 5.12: Parameter estimates for Model 6

Parameter	Estimate	Std. Error
Fixed:		
Constant	13.500	0.165
Gender (Male)	-0.878	0.023
Scientific Orientation (Human Sciences)	0.380	0.023
Scientific Orientation (Exact Sciences)	2.868	0.033
Gender*Scientific Orientation (Male/Human Sciences)	-0.865	0.036
Gender*Scientific Orientation (Male/Exact Sciences)	1.000	0.050
Year of Examination (2006)	-1.078	0.021
Year of Examination (2007)	-0.247	0.022
Year of Examination (2008)	-0.328	0.022
Type of School (Public)	-1.844	0.170
Random:		
σ_{u0}^2 (between schools)	2.634	0.111
σ_{e0}^2 (between students)	17.675	0.045
-2*log(likelihood)	1774905	

Model 7

One of the main advantages of a multilevel model is that it gives the opportunity to combine explanatory variables taken from different levels in order to examine more precisely the response variable, simply by specifying a cross-level interaction, that is the interaction between two variables from different levels. In the new model (Model 7) we introduce the interaction term between the type of school and the gender of the students, in order to detect possible differences in the performance of males and females according to the type of school they study. Since both variables have two categories, only one new parameter needs to be estimated and that is for male students in public schools. The new estimate, as well as all the alterations in the previous estimates, are presented in the following table (Table 5.13). The deviance of the new model is 1774898 and by comparing it to the deviance of the previous model (1774905) the difference is significant referring to the tables of the chi-square distribution with one degree of freedom, so we can keep Model 7 as an improvement from the previous model. Also, once again all parameter estimates of the model are significant since the estimates of the parameters are more than three times of the estimates of their standard errors and, moreover, the values of the estimates for both the fixed and the random part have not altered by the previous model. Only the estimates for the gender and the type of school factors have slightly altered which is logical since the new parameter estimate is the interaction term of these two variables. However these new parameter estimates as well as the estimate for the new interaction term have not changed our previous conclusions. In other words, as was mentioned before, the performance referring to the score in the General Admission Grade for male students and for public schools are relatively worse than female students and private schools.

Table 5.13: Parameter estimates for Model 7

Parameter	Estimate	Std. Error
Fixed:		
Constant	13.406	0.168
Gender (Male)	-0.689	0.062
Scientific Orientation (Human Sciences)	0.378	0.023
Scientific Orientation (Exact Sciences)	2.871	0.033

Gender*Scientific Orientation (Male/Human Sciences)	-0.863	0.036
Gender*Scientific Orientation (Male/Exact Sciences)	0.991	0.050
Year of Examination (2006)	-1.078	0.021
Year of Examination (2007)	-0.247	0.022
Year of Examination (2008)	-0.328	0.022
Type of School (Public)	-1.743	0.173
Type of School*Gender (Public/Male)	-0.202	0.062
Random:		
σ_{u0}^2 (between schools)	2.636	0.111
σ_{e0}^2 (between students)	17.674	0.045
-2*log(likelihood)	1774898	

Model 8

The new term we introduce to the new model (Model 8) is another cross-level interaction term, the interaction between the type of school and the scientific orientation of the studies. By adding this term we try to detect if the performance of students of different scientific orientations is related to the type of school the study. Since the first variable (type of school) has two categories and the second (scientific orientation) has three, only two new parameters will be estimated, one for students from public schools with Human Sciences orientation and one for students from public schools with Exact Sciences orientation. The new estimates, as well as the estimates of all the other parameters discussed previously are presented in table 5.14. The difference of the deviances of the new model (Model 8) and the previous one is 1774898-1774550 and is highly significant referring to the tables of the chi-square distribution with two degrees of freedom. So the new model (Model 8) is a significant improvement compared to the previous model. However, it is important to observe that the parameter estimates for the interaction terms "public school*male student" and "public school* Human Sciences orientation" are not significant since the estimates of the standard error of the parameters are more than a third of the parameter estimates for the particular terms. In other word, contrary to what we concluded in the previous step, we can now say that there is no significant interaction between the type of school and the gender of student or, in other word, both male and female students perform in the same "pattern" in public and private schools. Thinking in the same way, we can say that students of Human Sciences orientation and students of Technical Sciences orientation also perform in the same "pattern" in public and private schools. On the other hand, we should seriously pay attention to the high decrease of the parameter estimate for Exact Sciences scientific orientation compared to the previous model (1.533 vs. 2.871) and at the same time to the relatively high estimate for the new interaction term "public school* Exact Sciences orientation". If we combine these two observations, we can conclude that the performance of students examined for Exact Sciences orientation is generally higher than the other two orientations (a conclusion that we have already mentioned in previous steps), but especially in public schools, the performance of students of Exact Sciences orientation is relatively even higher. For all the other parameters for both mixed and random parts of the new model, the effect of the inclusion of the new interaction term is not so important.

Table 5.14: Parameter estimates for Model 8

Parameter	Estimate	Std. Error
Fixed:		
Constant	13.653	0.173
Gender (Male)	-0.715	0.065
Scientific Orientation (Human Sciences)	0.390	0.074
Scientific Orientation (Exact Sciences)	1.533	0.083
Gender*Scientific Orientation (Male/Human Sciences)	-0.864	0.036
Gender*Scientific Orientation (Male/Exact Sciences)	1.028	0.050
Year of Examination (2006)	-1.077	0.021
Year of Examination (2007)	-0.245	0.022
Year of Examination (2008)	-0.328	0.022
Type of School (Public)	-2.009	0.179
Type of School*Gender (Public/Male)	-0.174	0.066
Type of School*Scientific	-0.007	0.075

Orientation (Public/Human Sciences)		
Type of School*Scientific Orientation (Public/Exact Sciences)	1.471	0.084
Random:		
σ^2_{u0} (between schools)	2.661	0.112
σ^2_{e0} (between students)	17.653	0.045
-2*log(likelihood)	1774550	

Model 9

In the new model (Model 9), which is also the final model of the multilevel analysis, we do not introduce another explanatory variable. Instead, we add a parameter in the random part of the model which is the variance term between prefectures (σ^2_{v0}). In other words our model now becomes a 3-level model with students (1st level) nested within schools (2nd level) nested within prefectures (3rd level). The new estimate of the level-3 variance term, as well as the estimates of all the other parameters, are presented in table 5.15. First, we need to perform a likelihood-ratio test in order to check whether the inclusion of a third level random term is a significant improvement to the previous 2-level model. The deviance of the new model is 1774456 and by comparing it to the deviance of the previous model (1774550) the difference is highly significant referring to the tables of the chi-square distribution with one degree of freedom. So, Model 9 is a significant improvement of the previous model. Therefore, we can conclude that the use of a 3-level model containing also a between-prefecture variance estimate is a significant improvement of the 2-level model which only contains a between-schools and between-students random part. In other words, we have proven that taking into account the full hierarchical structure of the data (students nested in schools nested in prefectures) we can conclude to more accurate results. However, by comparing the parameter estimates of the fixed part of the new model to the respective parameters of the previous model we observe minor differences, with only exception the decrease of the estimate of the type of school (-1.747 vs. -2.009 in the previous model). As it is obvious, the part of the model which has been affected more by the inclusion of the third level variance is the random part. More specifically the parameter estimate for the level-2 variance (between schools) has decreased from 2.661 to 2.364, while the estimate for the level-1 variance (between students) has remained unaffected.

Table 5.15: Parameter estimates for Model 9

Parameter	Estimate	Std. Error
Fixed:		
Constant	13.177	0.193
Gender (Male)	-0.717	0.065
Scientific Orientation (Human Sciences)	0.388	0.074
Scientific Orientation (Exact Sciences)	1.536	0.083
Gender*Scientific Orientation (Male/Human Sciences)	-0.865	0.036
Gender*Scientific Orientation (Male/Exact Sciences)	1.029	0.050
Year of Examination (2006)	-1.078	0.021
Year of Examination (2007)	-0.244	0.022
Year of Examination (2008)	-0.328	0.022
Type of School (Public)	-1.747	0.174
Type of School*Gender (Public/Male)	-0.172	0.066
Type of School*Scientific Orientation (Public/Human Sciences)	-0.005	0.075
Type of School*Scientific Orientation (Public/Exact Sciences)	1.469	0.083
Random:		
σ^2_{v0} (between prefectures)	0.308	0.091
σ^2_{u0} (between schools)	2.364	0.102
σ^2_{e0} (between students)	17.653	0.045

-2*log(likelihood)	1774456

5.5 Conclusions of the Chapter

Apart from the usefulness of the results themselves, since we refer to real data from the Greek educational system, the main conclusion that can be drawn of the analysis in this Chapter, is that the use of Multilevel Analysis Techniques, and more specifically of a 3-level model, has significant advantages compared to simplest models, both concerning the precision of the estimates, as well as the interpretation of the hierarchical structure of the data.

CHAPTER 6

6 CONCLUSIONS – FURTHER RESEARCH

The purpose of the current thesis is to describe the most appropriate statistical models and methods in order to analyze hierarchical data structures. In the first part of the thesis (the "theoretical") we described all the possible known approaches for analyzing this type of data structures and we elaborated more on the "Multilevel Analysis" and the corresponding "Multilevel Models". We did so because this kind of approach was proved to be the most appropriate in order to analyze hierarchical structured datasets. According to the related literature, the main advantage of multilevel models is that they "respect" the hierarchical structure of the data. In other words, they allow variables from all levels of interest to be introduced in the model, they can also examine cross-level interactions between variables in different levels and, furthermore, they can examine in detail the random part of a model by examining the variation of data separately in all levels of interest. By doing so, multilevel models offer a more precise approach for the response variable of interest compared to the classical models. In the first Chapters we described in detail all the theoretical basis of these models, the formulas and notations, the estimation methods and testing methods for all the parameters of the model, both fixed and random, the algorithms and also the extensions of the simple multilevel model in more sophisticated cases. We presented cases where simple multilevel models can readily extent to more complex data structures (multivariate models, models with discrete response variable etc.). Moreover, in this part of the thesis we have chosen and described in detail a number of practical examples and applications of multilevel models in all areas of applications of multilevel modelling. These areas are educational statistics, spatial statistics, health statistics, repeated measures and survey research. By reviewing the methods and the results of these applications described by the authors, we have shown that multilevel techniques can be applicable and also very useful in many areas where hierarchical data structure exists. Also, the chosen examples were proved to be very representative and can form the basis for other applications in the same area of interest.

In the second part of the thesis (the "practical") we applied the multilevel techniques and models into real data made available from the Greek Ministry of

Education, Lifelong Learning and Religious Affairs referring to the students' score of the General Admission Grade in the National Exams in the years 2006-2009. Our results from the analysis confirmed that the use of multilevel techniques may provide more precise results and estimates for the students' performance referring to their General Admission Grade. More specifically the conclusions obtained from our analysis can be briefly summarized to the following:

- As concerns the prefectures, and utilizing the results of the descriptive statistics, Chios seems to be the prefecture with the best average performance through the years of interest (2006-2009). This result is consistent with the descriptive results given by Kosmopoulou (1998) for the years 1990 and 1991 indicating an ongoing stability of the students' performance of this particular prefecture during the years. The result is also consistent with the results of the project of the Centre of Development of Educational Policy of the General Confederation of Greek Workers (GSEE) (2009) (www.kanep-gsee.gr/index.php?download=keimeno%206-4-09.doc) concerning the indices of access in higher education. However, this conclusion should be drawn with caution since we refer only to descriptive and not confirmatory statistics.
- The use of level-2 and level-3 analysis which contains the effect of between-schools and between-prefectures variability is a significant improvement compared to models which contain only the between-students effect.
- According to the multilevel analysis, female students seem to perform significantly better than male students concerning their score in the General Admission Grade. This result seems rational, taking into account the general belief that girls perform better than boys in their school examinations.
- As concerns the scientific orientation of the studies, students examined for the Exact Sciences orientation seem to have highly better performance than students examined for the other two orientations (Human Sciences and Technical Sciences). This is also an expected result since Exact Sciences orientation is chosen by less (more "conscious") students mainly focused on entering Institutions related to health studies which demand very high entrance exams scores. As for the other two scientific orientations, students of Human Sciences orientation have a slightly higher performance than those of Technical Sciences orientation. The above results are also confirmed by the official results on the scaling of the General Admission Grade of the candidate

students for the access to the National Universities and Technical Institutions by scientific orientation, published every year by the Greek Ministry of Education, Lifelong Learning and Religious Affairs (www.ypepth.gr).

- Although for the whole of students of Human Sciences orientation the General Admission Grade is better than those of Technical Sciences orientation, for boys the performance in the Technical Sciences orientation seems to be better than in Human Sciences orientation. This conclusion also seems to be in agreement with the "common sense" that boys perform better in technological areas and girls in classical. Analogous results were drawn by Marouga (2004), although the educational system of access and the descriptions of the scientific orientations presented in the two projects are different.
- The year of examination also seems to have a significant effect on the performance of the students and their scores in the General Admission Grade. We can conclude from the analysis that 2009 is the best year for students' performance, while 2006 is by far the worst. This result is probably due to the fact that 2006 was the first year of application of the new educational system of access, so students needed several years in order to adapt to this new system. The most profound reason, in general, for the differences in the performance of students over the years is the fact that the level of difficulty of the National Exams is impossible to be equally predetermined among the years. It has been discussed by many authors, politicians, professors and sociologists (see for example the report of the Centre of Research and Documentation of the Greek Federation of State School Teachers of Secondary Education (OLME) (2004) (http://www.smarinis.gr/aei1.pdf)) that this fact is maybe one of the main disadvantages of the Greek National Exams system, since it does not offer equal opportunities for students in their attempt to access the Greek National Universities and Technical Institutions. From the statistical point of view, this factor (year of examination) produces serious problems to the interpretation of the results, since it cannot easily be controlled and therefore does dot allow the detection of the effect of more important factors.
- As far as the type of school is concerned, we can conclude from the model that the performance of students from private schools according to their General Admission Grade is much higher than the students from public schools. This is

also an expected result, however we should take into account the very small number of students in private schools compared to those of public schools.

- By examining the cross-level interactions between type of school and gender and type of school and scientific orientation we conclude that the performance of students in the two types of schools does not depend on the gender of the student. In other words, boys and girls perform "in the same pattern" both in private and public schools. On the other hand, it can be deduced by the analysis that, especially in public schools, the performance of students for the Exact Sciences orientation is even higher than the "usual pattern". With regard to this particular outcome, Marouga (2004) has also come to analogous results.

Apart from the usefulness of the results of the analysis themselves, which is straightforward since we refer to real data from the Greek educational system, we should also argue on the reflections of the practical application of this kind of Multilevel Analysis to this real situation. It is verified throughout the analysis that, from a statistical and methodological point of view, the use of Multilevel Analysis Techniques has significant advantages compared to simplest (non-multilevel models) concerning the precision of the estimates, the explanation of the variability of the model (especially in the random part), as well as the improvement in the interpretation of the hierarchical structure of the data.

So far we have pointed out the obvious advantages of the use of multilevel modeling techniques in our analysis. The statistical advantages (more precise estimates, elaboration on the random part of the model), the consistency of our results with the results of other reports and projects (not necessarily statistical), as well as the "harmony" of the results with our "common sense" and knowledge for the National Examination Results. On the other hand, we should point out some technical disadvantages. Firstly, it has been argued by many authors (see for example Goldstein, 1993) that the lack of an explanatory (continuous) variable referring to the previous achievements of the students and the use of raw unadjusted results may lead to some invalid comparisons and some misleading outcomes. For example, the results of our analysis strongly agree with the results of the previous analysis conducted by Kosmopoulou (1998) when raw data are used in both analyses, but somehow differ when an "adjuster" was used. Further research on the same subject should be focused on the addition of more explanatory variables in all levels of interest, especially an explanatory variable as an adjustment of the outcome and, also, more higher-level variables. Furthermore, the use of SPSS program to analyze the data offers applicability, easy handling of data transformation situations, multiple choices of syntax procedures and of course high quality of the produced results, however it is rather time-demanding compared to other more specialized statistical programs (MLwiN, HLM or even SAS).

From all the above we can conclude that the use of multilevel analysis techniques is a significant improvement for the examination of hierarchical structured datasets in many areas of interest -including education- however results of this type of analysis should be treated with caution. The application presented in the thesis, despite the constraints, is the first multilevel analysis approach concerning this particular educational system of access in the National Universities and Technical Institutions and the second (after Kosmopoulou, 1998) considering the use of multilevel techniques in Greek educational datasets in general. Since it is profound that the use of such models is a significant improvement for the analysis of hierarchical data structures, further research needs to be done, especially in the education area. More specifically:

- We have already pointed out the necessity of using more explanatory variables in all the levels of interest. Especially, in the 1st level (students) a continuous explanatory variable, such as the average grade of the students in previous high-school exams, could be used as an initial adjustment of the outcome. Moreover, more higher-level variables, such as the size of the school (2nd level variable) or the demographical structure of the prefecture (3rd level variable), would be useful in order to obtain more precise estimates for both the fixed and the random part of the model. However, such information is not easily available in real datasets and, if available, requires the combination of datasets from many sources.
- Multilevel Models outcomes have been used by many authors in order to rank Institutions (schools and/or Universities) according to their students' performance. A further application of such techniques regarding the ranking of the Greek Institutions would be interesting. However, as discussed by many authors (see Goldstein, 1993) the results of ranking are very sensitive to data alterations and constraints and, therefore, should be used with great caution.

- The rapid development of IT-Systems allows the use of more sophisticated and memory-demanding algorithms for the estimation of the fixed and random parameters of the model. Comparisons between the results of various estimation techniques, especially in small datasets, may reveal significant improvements regarding the robustness of the estimates.
- Extensions of the Linear Multilevel Model have already been discussed, from a theoretical perspective, within the thesis. Applications of such techniques in Greek datasets would have great practical importance. For instance, in practice, the performance of the students in the National Exams is usually measured by a dichotomous variable ("pass"/"fail") or a polynomial variable. Analysis of a dichotomous or polynomial response variable using multilevel approaches requires the introduction of a "Generalized Linear Multilevel Model'. However, the application of such models in practice has difficulties, due to the fact that it requires advanced algorithms and memory-demanding computational techniques, not available by the classical computer software.
- Finally, it is well-known that the Greek educational system of access in the National Universities and Technical Institutions has been altered various times through the last years. The lack of a stable system, does not allow an ongoing comparison of the performance of students in the National Exams for their access in the third-degree education. Therefore, it would be of great practical importance to be able to perform techniques of multilevel analysis (Meta-Analysis techniques for instance) in order to adjust the outcomes of various educational systems of access and make the appropriate comparisons.

CHAPTER 7

7 REFERENCES

- **Afshartous, D. & DeLeeuw, J. (2002).** An application of multilevel model prediction to NELS :88. National Institute for Statistical Sciences.
- **Aitkin, M., Anderson, D. & Hinde, J. (1981).** Statistical Modeling of Data on Teaching Styles. Journal of the Royal Statistical Society, A, 144, 419-461.
- **Albright, J.J & Marinova, M.D. (2010).** Estimating Multilevel Models using SPSS, Stata, SAS, and R. Stat/Math Center, University of Indiana.
- **Bangert-Drowns, R.L. (1986).** Review of developments in meta-analytic method. Psychological Bulletin, 99, 388-399.
- **Bennet, N. (1976).** Teaching Styles and Pupil Progress. London, Open Books.
- **Breslow, N.E. & Clayton, D.G. (1993).** Approximate inference in generalized linear mixed models. Journal of American Statistical Association, 88, 9-25.
- **Browne, W.J., Draper, D., Goldstein, H. & Rasbash, J.** Bayesian and likelihood methods for fitting multilevel models with complex level-1 variation. Computational Statistics and Data Analysis, 39, 203-225.
- **Bryk, A.S. & Raudenbush, S.W. (1992).** Hierarchical Linear Models. Newbury Park, Sage.
- **Burt, C.D.B., Mitchell, D.A., Raggatt, P.T.F., Jones, C.A. & Cowan, T.M. (1995).** A snapshot of autobiographical memory retrieval characteristics. Applied Cognitive Psychology, 9, 61-74.
- **Carey, K.** (2000). A multilevel modeling approach to analysis of patient costs under managed care. Health Economics, 9, 435-446.
- **Carpenter, J.R. Goldstein, H. & Rasbash, J. (2003).** A novel bootstrap procedure for assessing the relationship between class size and achievement. Journal of the Royal Statistical Society, C, 52, 431-443.
- **Courgeau, D. & Baccaini, B. (1998).** Multilevel Analysis in the Social Sciences. Population: An English Selection, special issue New Methological Approaches in the Social Sciences, 1998, 39-71.

- **Cronbach, L.J. & Webb, N. (1975).** Between class and within class effects in a reported aptitude x treatment interaction: A reanalysis of a study by G.L. Anderson. Journal of Educational Psychology, 67, 717-24.
- **DeLeeuw, J. (1992).** Data quality in mail, telephone, and face-to-face surveys. Amsterdam, TT-Publikaties.
- **Duncan, O.D., Curzort, R.P. & Duncan, R.P. (1966).** Statistical Geography: Problems in analyzing Areal Data. Free Press, Glencoe, IL.
- **Gilks, W.R., Clayton, D.G., Spiegelhalter, D.J., Best, N.G., McNiel, A.J., Sharples, L.D. & Kirby, A.J. (1993).** Modeling complexity: applications of Gibbs Sampling in medicine. (With discussion). Journal of the Royal Statistical Society, B, 55, 39-102.
- **Goldstein, H., (1989a).** Restricted unbiased iterative generalized least square estimation. Biometrica, 76, 622-23.
- **Goldstein, H. (1997).** Methods in School Effectiveness Research. School Effectiveness and School Improvement, 8, 369-95.
- **Goldstein, H. (1998).** Multilevel Models for Analyzing Social Data. Encyclopaedia of Social research Methods.
- **Goldstein, H. (1995).** Multilevel Statistical Models, (2nd Edition). London, Edward Arnold.
- **Goldstein, H., Browne, W.J. & Rasbash, J. (2002).** Multilevel modeling in Medical Data. Statistics in Medicine, 21, 3291-3315.
- **Goldstein, H., & Healy, M.J.R. (1994).** The graphical presentation of a collection of means. Journal of the Royal Statistical Society, A, 158, 175-7.
- **Goldstein, H., Huiqi, P, Rath, T. & Hill, N. (2000).** The use of value added information in judging school performance. Perspectives on Education Policy, 8.
- **Goldstein, H. & Rasbash, J. (1992).** Efficient computational procedures for the estimation of parameters in multilevel models based on iterative generalized least squares. Computational Statistics and Data Analysis, 13, 63-71.
- **Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttall, D. & Thomas, S. (1993).** A Multilevel Analysis of School Examination Results. Oxford Review of Education, 19, 425-433.

- **Goldstein, H. & Spiegelhalter, D. (1996).** League Tables and their Limitations: Statistical Issues in Comparisons of Institutional Performance. Journal of the Royal Statistical Society, A, 159, 385-443.
- **Goldstein, H. & Thomas, S. (1996).** Using Results as Indicators of School and College Performance. Journal of the Royal Statistical Society, A, 159, 149-164.
- **Goldstein, H. & Woodhouse, G. (2000).** School effectiveness research and Educational Policy. Oxford Review of Education.
- **Hedges, L.**V. & Olkin. I. (1985). Statistical Methods for Meta Analysis. Orlando, Academic Press.
- **Hill, P.W. & Goldstein, H. (1998).** Multilevel Modeling of Educational Data with Cross-Classification and Missing Identification for Units. Journal of Educational and Behavioural Statistics, 23, 117-128.
- **Hox, J.J. (2010).** Multilevel Analysis: Techniques and Applications, (2nd Edition). New York, Routledge Academic.
- **Hox, J.J. (1995).** Applied Multilevel Analysis. Amsterdam, TT-Publikaties.
- **Hox, J.J. (1994).** Hierarchical Regression Models for Interviewer and Respondent Effects. Sociological Methods and Research, 22, 300-318.
- **Kish, L. (1987).** Statistical Design for Research. New York, Wiley.
- **Kosmopoulou, A. (1998).** Assessment of School Effectiveness Using Multilevel Models. Dissertation for the MSc program of Department of Statistics, Athens University of Economics and Business.
- **Kreft, I.G.G. (1993).** Using multilevel analysis to assess school effectiveness: A study of Dutch secondary schools. Sociology of Education, 66, 2, 104-129.
- **Kreft, I. & DeLeeuw, J. (1998).** Introducing Multilevel Modeling. London, Sage
- **Kreft, I., DeLeeuw, J. & Aiken, L. (1995).** The effect of different forms of centering in hierarchical linear models. Multivariate Behavioural Research, 30, 1-22.
- **Levy, P. & Goldstein, H. (1984).** Tests in Education. London, Academic Press.
- **Leyland, H.A. (2004).** A review of multilevel modeling in SPSS. MRC Social and Public Health Sciences Unit, University of Glasgow.

- **Liang, K. & Zeger, S.L. (1986).** Longitudinal data analysis using generalized linear models. Biometrika, 73, 45-51.
- **Longford, N.T. (1987).** A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. Biometrika, 74, 817-27.
- **Marinis, S. (2004).** Documentation of the view of the organization for the suppression of the National Exams and the autonomy of Lyceum. Centre of Research and Documentation, Greek Federation of State School Teachers of Secondary Education.
- **Marouga, K. (2004).** Statistical Analysis of Marks and Preferences of the Candidates in Year 1993 with the System of General Exams. Dissertation for the MSc program of Department of Statistics, Athens University of Economics and Business.
- **McArdle, J.J. & Hamagami, F. (1994).** Logit and multilevel logit modeling of college graduation for 1984-1985 freshman student athletes. Journal of the American Statistical Association, 89, 427, 1107-1123.
- **McCullagh, P. & Nelder, J. (1989).** Generalized Linear Models (2nd edition). London, Chapman and Hall.
- **Nuttall, D.N., Goldstein, H., Prosser, R. & Rasbash, J. (1989).** Differential school effectiveness. Journal of Educational Research, 13, 769-776.
- **O'Donoghue, C.**, **Thomas, S., Goldstein, H. & Knight, T. (1997).** 1996 DfEE study of Value Added for 16-18 year olds in England. DfEE research series, March, London, DfEE.
- **Painter, J. (2003).** Designing Multilevel Models Using SPSS 11.5 Mixed Model. Jordan Institute for Families, School of Social Work, University of North Carolina.
- **Paizis, N. (2009).** Research of the system of access in the third degree education (National Exams 2004, 2005 and 2006). Centre of Development of Educational Policy, General Confederation of Greek Workers.
- **Paterson, L. (1991).** Socio-economic status and educational attainment: a multidimensional and multilevel study. Evaluation and Research in Education, 5, 97-121.

- **Rice, N., Carr-Hill, R., Dixon, P. & Sutton, M. (1998).** The influence of households on drinking behaviour: A multilevel analysis. Social Science & Medicine, 46, 8, 971-979.
- **Rice, N. & Jones, A. (1997).** Multilevel models and health economics. Health Economics, 6, 561-575.
- **Robinson, W.S. (1950).** Ecological correlations and the behaviour of individuals. American Sociological Review, 15, 351-357.
- **Searle, S.R., Casella, G. & McCulloch, C.E. (1992).** Variance Components. New York, Wiley.
- **Skinner, C.J., Holt, D. & Smith, T.M.F. (1989).** Analysis of Complex Surveys. New York, Wiley.
- **Wright, D. (1998).** Modeling clustered data in autobiographical memory research: The multilevel approach. Applied Cognitive Psychology, 12, 339-357.
- **Yang, M.**, Goldstein, H., Browne, W. & Woodhouse, G. Multivariate multilevel analyses on examination results. Journal of the Royal Statistical Society, A, 165, 137-153.
- **Zeger, S.L. & Karim, M.R. (1991).** Generalized linear models with random effects; a Gibbs Sampling approach. Journal of American Statistical Society, 86, 79-102.