当代科学前沿论丛

NEW FRONTIERS OF SCIENCES

Multilevel Models: Applications Using SAS

多层统计分析模型：SAS与应用

$$\beta_{1j} = \gamma_{10} + \sum_{m=1}^{M} \gamma_{1m} w_{mj} + u_{1j}$$

$$\cdots$$

$$\beta_{Qj} = \gamma_{Q0} + \sum_{m=1}^{M} \gamma_{Qm} w_{mj} + u_{Qj}$$

$$y_{ij} = \gamma_{00} + \sum_{m=1}^{M} \gamma_{0m} w_{mj} + \sum_{p=1}^{P} \alpha_p X_{pij} + \sum_{q=1}^{Q} \gamma_{q0} z_{qij} + \sum_{q=1}^{Q} \sum_{m=1}^{M} \gamma_{qm} w_{mj} z_{qij} + \left(u_{0j} + \sum_{q=1}^{Q}$$

Jichuan Wang · Haiyi Xie · James Henry Fisher

高等教育出版社 HIGHER EDUCATION PRESS

当代科学前沿论丛

Multilevel Models:

Applications Using SAS

多层统计分析模型：SAS与应用

Jichuan Wang · Haiyi Xie · James Henry Fisher

高等教育出版社

图书在版编目（CIP）数据

多层统计分析模型：SAS与应用=Multilevel Models: Applications Using SAS: 英文 / 王济川，谢海义，（美）费舍余（Fisher, J.）著. —北京：高等教育出版社，2009. 6

ISBN 978-7-04-027568-1

Ⅰ. 多… Ⅱ. ①王…②谢…③费… Ⅲ. 统计分析-应用软件，SAS-英文 Ⅳ. C812

中国版本图书馆CIP数据核字 (2009) 第 071199 号

策划编辑 王丽萍 **责任编辑** 王丽萍 **封面设计** 于 涛

责任校对 杨凤玲 **责任印制** 宋克学

出版发行	高等教育出版社	**购书热线**	010 - 58581118
社 址	北京市西城区德外大街 4 号	**免费咨询**	400 - 810 - 0598
邮政编码	100120	**网 址**	http://www.hep.edu.cn
总 机	010 - 58581000		http://www.hep.com.cn
		网上订购	http://www.landraco.com
经 销	蓝色畅想图书发行有限公司		http://www.landraco.com.cn
印 刷	北京新华印刷厂	**畅想教育**	http://www.widedu.com
开 本	787 × 1092 1/16	**版 次**	2009 年 6 月第 1 版
印 张	17.25	**印 次**	2009 年 6 月第 1 次印刷
字 数	430 000	**定 价**	58.00 元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换。

物料号 27568 - 00

《当代科学前沿论丛》专家委员会

(按姓氏笔画为序)

(国内部分)

王　夔	冯　端	师昌绪	曲钦岳	朱清时
孙　枢	李三立	李大潜	李国杰	杨芙清
吴建屏	邹承鲁	张尧庭	陈　竺	陈佳洱
陈希孺	陈宜瑜	周秀骥	姜伯驹	袁亚湘
钱　易	徐光宪	徐端夫	徐冠华	翟中和
戴立信	戴汝为			

(海外部分)

王中林	文小刚	邓兴旺	田　刚	丛京生
刘　钧	汤　超	许　田	危　岩	严晓海
李　凯	李　明	邱子强	余振苏	范剑青
周午纵	郑元芳	宫　鹏	俞陆平	袁钧瑛
徐希平	鄂维南	程正迪		

Preface

Interest in multilevel statistical models for social science and public health studies has been aroused dramatically since the mid-1980s. New multilevel modeling techniques are giving researchers tools for analyzing data that have a hierarchical or clustered structure. Multilevel models are now applied to a wide range of studies in sociology, population studies, education studies, psychology, economics, epidemiology, and public health.

Individuals and social contexts (e.g., communities, schools, organizations, or geographic locations) to which individuals belong are conceptualized as a hierarchical system, in which individuals are micro units and contexts are macro units. Research interest often centers on whether and how individual outcome varies across contexts, and how the variation is explained by contextual factors; what and how the relationships between the outcome measures and individual characteristics vary across contexts, and how the relationships are influenced or moderated by contextual factors. To address these questions, studies often employ data collected from more than one level of observation units, i.e., observations are collected at both an individual level (e.g., students) and one or more contextual levels (e.g., schools, cities). As a result, the data are characterized by a hierarchical structure in which individuals are nested within units at the higher levels. This kind of data is called hierarchically structured data or multilevel data. The conventional single-level statistical methods, such as ordinary least square(OLS) regression are inappropriate for analysis of multilevel data because observations are nonindependent and the contextual effects cannot be addressed appropriately in such models. Multilevel modeling not only takes into account observation dependence in the multilevel data, but also provides a more meaningful conceptual framework by allowing assessment of both individual and contextual effects, as well as cross-level interaction effects.

This book covers a broad range of topics about multilevel modeling. Our goal is to help students and researchers who are interested in analysis of multilevel data to understand the basic concepts, theoretical frameworks and application methods of multilevel modeling. This book is written in non-mathematical terms, focusing on the methods and application of various multilevel models, using the internationally widely used statistical software, the *Statistics Analysis System* (SAS). Examples are drawn from analysis of real-world research data. We focus on two-level models in this book because it is most frequently encountered situation in real research. These models can be readily expanded to models with three or more levels when applicable. A wide range of linear and non-linear multilevel models are introduced and demonstrated.

There are six chapters in this book.

Chapter 1 presents a brief introduction and overview of multilevel modeling. In this chapter, we discuss the problems inherent in applying traditional analytical methods to hierarchically structured or multilevel data; we explain why multilevel models are needed for analyzing such data; and we discuss the conceptual framework, its advantages, and limitations of multilevel modeling. Chapter 1 concludes with a brief overview of computer software for multilevel modeling.

Chapter 2 summarizes basic concepts of multilevel models, including intra-class correlation (ICC), model formulation, statistical assumptions, model estimation, model fit and model comparison, explained micro and macro level variances, and strategies of model building. Expansion of the 2-level model to 3-level models is also discussed.

Chapter 3 demonstrates linear multilevel models, also known as hierarchical linear models (HLM) using cross-sectional data. This chapter presents detailed model building strategies and illustrates model development and statistical testing procedures step by step.

Chapter 4 extends multilevel models to longitudinal data. The chapter covers both linear and curvilinear growth models. Some complex modeling strategies such as orthogonal polynomial modeling and piecewise modeling techniques are also presented.

Chapter 5 discusses advanced multilevel models for discrete outcome measures, such as binary, ordinal, nominal and count outcomes. The chapter starts with introduction of the generalized linear models.Then we present the model formulation for each type of discrete outcomes: multilevel logistic regression for binary outcome, multilevel cumulative logistic regression for ordinal outcomes, multilevel multinomial models for nominal outcome, and multilevel Poisson model, as well as multilevel negative binomial model, for count data. Alternative SAS procedures are used to analyze different types of discrete outcomes, and detailed count of model specifications and interpretations of model results are presented.

Chapter 6 discusses some special issues that are often encountered in multilevel modeling, including approaches for modeling count data with extra zeros, semi-continuous outcome measures, and multilevel data with a small number of groups (i.e., level-2 units). We demonstrate multilevel or random effects zero-inflated Poisson (RE-ZIP) models, random effect zero-inflated negative binomial models (RE-ZINB), mixed-effect mixed-distribution models, bootstrapping multilevel models using SAS procedures. In addition, group-based models are introduced to assess growth trajectories of various outcome measures using longitudinal data. A special SAS procedure, SAS *PROC TRAJ* is used to demonstrate group-based logit models, group-based ZIP models, group-based Poisson models, group-based censored normal models, and group-based normal models. Finally, missing values and sample size/statistical power estimation for multilevel modeling are discussed.

While many computer programs are available for multilevel modeling, we have chosen the internationally distributed statistics package *Statistics Analysis System* (SAS) to demonstrate multilevel models in this book. SAS is a suitable package for many analysts because of its powerful data manipulation and modeling capabilities. The models demonstrated in this book are intended to show readers, step by step, how to build multilevel models using SAS for both cross-sectional and longitudinal data. SAS syntax for all of the models covered in the book are provided in each corresponding chapter of the book. The data used, as well as SAS syntax for all examples, can be downloaded from the website of the China Higher Education Press (academic.hep.com.cn). Although data used for these examples are drawn from public health studies, the methods and analytical techniques are applicable to other fields of social sciences.

Contents

Chapter 1 Introduction

Over the past two decades *multilevel models* (Mason, Wong *et. al.* 1983; Bryk & Raudenbush, 1992; Raudenbush & Bryk, 2002; Goldstein, 1987, 1995) have gained popularity in various research fields including education, psychology, sociology, economics, and public health. Multilevel models extend ordinary least square (OLS) regression to analyze multiple level data or hierarchical data that involve both micro and macro observation information. Multilevel models also appear under different names in the literature, including *hierarchical linear models* (Bryk & Raudenbush, 1992; Raudenbush & Bryk, 2002), *random-effect models* (Laird & Ware, 1982), *random coefficient models* (DeLeeuw & Kreft, 1986), *variance component models* (Dempster, Rubin, & Tsutakawa, 1981), *mixed models* (Longford, 1987), and *empirical Bayes models* (Strenio, Weisberg, & Bryk, 1983).

Prior to the development of formal statistical methodology for multilevel models, sociologists engaged in contextual or multilevel analysis of hierarchically structured data. In the late 1950s and early 1960s Lazarsfeld (1961) and Merton (1957) at Columbia University began to assess contextual effects on individual behavior. The 1970s witnessed a significant jump in analysis of multilevel data in education (Barr & Dreeben, 1977; Block & Burns, 1976; Bronfenbrenner, 1976; Burstein, 1980; Cronbach, 1976; Herriot & Muse, 1973; Pedhazur, 1975; Snow, 1976; Spady, 1973; Walberg, 1976). In a systematic study of contextual analysis, Boyd and Iversen (1979) discussed how to model multilevel data with micro-macro models, i.e., to formulate within group regression model at individual level, then relate the within group regression coefficients to contextual variables that describe the groups. Although multilevel observations are discussed in their models, their estimation was conducted using ordinary least square (OLS) techniques that were inappropriate for multilevel analysis.

Statistical theories of multilevel models and corresponding computer programs were developed in early 1980s by sociologists and demographers. Models were applied to analyze the large scale multilevel data of the United Nation's World Fertility Survey (WFS) (Hermalin and Mason, 1980; Mason, Wong *et al*, 1983). Further methodological and substantive work in educational studies and the user-friendly windows-based computer programs by Bryk & Raudenbush (1992) and Goldstein (1987, 1995) have popularized the multilevel models. Multilevel models are now applied in a wide range of studies in the social sciences.

1.1 Conceptual framework of multilevel modeling

A key concept in social sciences is that a society can be described in hierarchical structures. By hierarchy, we mean that units at a lower level are nested within or grouped into units at a higher level. People cannot be treated as isolated individuals but as social beings. Individuals are members of many different types of groups and are embedded in different social contexts. For example, individuals belong to families, neighborhoods, organizations and communities. Awareness has been mounting that individual behaviors and outcomes are affected not only by individual

characteristics, but also shaped by the social contexts in which they are imbedded (Lazarsfeld, 1961; Merton, 1957; Bronfenbrenner, 1976; Blalock, 1984; Iversen, 1991). Davis' so-called "frog-pond" theory proposes that individual students evaluate personal ability relative to in-groups and pay little attention to out-groups (Davis, 1966). A moderately intelligent student (a medium size "frog") in a highly intelligent school (a large "pond") may become discouraged and thus become an under-achiever, while the same student in a considerably less intelligent school (a small "pond") may gain confidence and become an over-achiever. On the contrary, a moderately intelligent student might be motivated to study harder in a highly intelligent school and become more successful. The effect of an individual student's intelligence on his/her achievement may be influenced by specific features in the school he/she attends. In addition to composite measures (e.g., average intelligence level), student academic achievement may also be influenced by a variety of school level variables such as student/teacher ratio, teachers' work experience, school facilities, budget, etc.

The relationships between academic achievement and individual level variables vary across schools. For example, differences in academic achievement among ethnic groups may be larger in some schools and smaller in others. In such cases the extent of the effect of ethnicity on academic achievement may relate to identifiable school-level characteristics. On the other hand, the school's effect on student academic achievement may also vary among individuals. For example, while students usually benefit from smaller student/teacher ratios, these ratios and other school features are unlikely to influence all types of students equally. Cross-level interactions in multilevel modeling enable us to assess the degree to which relationships between individual explanatory and outcome variables are moderated by group level variables.

Good examples of this class of multilevel studies can be drawn from population studies. It is well-known that fertility levels vary among countries. In general, fertility is low in developed countries and high in developing countries. Fertility has multi-level determinants. Individual fertility behavior is determined not only by individual characteristics such as a couple's preference for children, ethnicity, education, and income at the micro level. Features of the social contexts or social environments where the individuals live, such as culture or subculture, GDP, average education level, and in particular, the intensity/efficacy of the family planning programs (FPP) at the macro-level can also produce measurable effects. Assessing cross-level interactions is very important in fertility studies. FPP analysts and officers are interested in knowing: What individual characteristics influence individual fertility behaviors? Do family planning programs work? How do differences in program implementation among various locations or macro-level units affect individual fertility behaviors? And, for what classes of people are family planning programs most effective? Multilevel modeling helps us to gauge how family planning programs interact with individual characteristics to affect fertility behavior.

Public health studies indicate that individual health behaviors and outcomes are jointly determined by individual and environmental factors (Von Korff *et al.*, 1992; Duncan *et al.*, 1996; Diez-Roux, 1998; Wang *et al.*, 1998). For example, initiation of smoking among adolescents may be associated with gender, ethnicity, school achievement and family background, as well as the social setting in which the individual is imbedded, such as geographic location, prevalence of smoking, and restrictions on smoking in public areas.

From these examples we can see that research interest in social science studies often centers on questions like: 1) what and how explanatory variables measured at the individual level affect the individual-level dependent variable, 2) what and how variables measured at the context or group level affect the individual-level dependent variables, 3) how the relationships between the individual-level explanatory and dependent variables vary across contexts or groups, and 4) what and how group-level variables moderate the effects of individual level variables on the individual-level dependent variable.

To answer these questions, both micro and macro data are needed. A common challenge in multilevel data is within-group observation dependence. That is, individuals in the same group tend to be alike and share similar attitudes and behaviors relative to individuals from other groups. For example, people living in the same neighborhood may share similarities with each other because they are influenced by the same neighborhood socio-economic characteristics. This may be true even for groupings that are only recently established. For example, students who are in the same school may not be associated with each other before they get into the same school. Once students enter a school, they become members of the same group. Once groupings are established, individuals in the group will tend to share traits that differentiate them from members of other groups. In statistical terms we say there exist within-group homogeneity and between-group heterogeneity in the hierarchically structured data.

Traditional analytical methods such as Ordinary Lease Squares (OLS) Regression assume that observations are independently, identically distributed (IID). The same assumption is required for generalized linear models. Violation of this assumption will result in incorrect inference in statistical analysis. Chapter 2 demonstrates how observation dependence can be measured using an *Intraclass Correlation* coefficient (ICC). Studies show that even a small ICC can lead to substantial Type-I errors in statistical testing, thus falsely rejecting a true null hypothesis. Dealing with ICC has been a challenge in statistical analysis of multilevel data for many years.

Multilevel models provide an appropriate analytical framework to deal with observation dependence in multilevel data. More importantly, multilevel models permit us to explore the nature and extent of the relationships at both micro and macro levels, as well as across levels.

1.2 Hierarchically structured data

Hierarchical social structures naturally give rise to hierarchical or multilevel data in which the lower level units are nested or grouped in the next higher level units. Such hierarchically structured data exist in many real life situations. The simplest and the most often used multilevel data are collected at two levels (i.e., one micro level and one macro level). For example, a study on student academic achievement may collect information at the student level and at the school level for multilevel modeling.

Multilevel designs can be readily extended to more than two levels. For example, students are nested in classes, and classes nested in schools; thus observation units lie at three levels of a hierarchy: the level-1 units are students; the level-2 units are classes; and the level-3 units are schools. The lowest level units (e.g., students) are the micro-level units or individual units, while the higher level units are the macro level units or context/group units.

Hierarchically structured data may arise in a variety of forms and from a variety of situations, either observed or by design. Survey data obtained from a complex sampling design are hierarchically structured. Multi-stage or cluster sampling is conducted to take full advantage of information from a hierarchy of study units. The first stage or "*Primary Sampling Unit*" (PSU) is often a well-defined geographic unit (e.g., county in a state). Once the PSUs are randomly selected, further stages of random selection are carried out within the PSUs (e.g., districts in a county) until the final units (e.g., households or individuals) are selected (Kalton, 1983). As a result, the survey data collected from cluster sampling design have a hierarchical structure in which individuals are nested within higher level sampling units.

Hierarchical data also frequently arise from experimental designs. For example, clinical trails may be carried-out in randomly selected clinics or medical centers, thus creating data that

have a hierarchical structure. However, in practice clinics and medical centers are often not randomly selected. This is also true in many multi-site research projects. For example, a national multi-site research project on public health is often conducted with many project sites located in different regions, states, or cities. Very often, rather than being randomly selected, project sites are selected based on the quality of the grant proposals, the level of seriousness of the health problems under study, or the feasibility of conducting a successful study in a specific site. Although the distribution of the project sites may be carefully taken into consideration, they are not randomly selected, thus they are not representative of the corresponding higher level units in the targeted population. As a result, inferences based on the multilevel analysis for non-randomly selected study sites should be interpreted with caution.

Hierarchical data structures are not confined to cross-sectional settings with multiple units. Individuals may also be higher level units. For example, in longitudinal or panel studies individuals are followed up over time. Data are collected repeatedly from the same individuals. Such longitudinal data can be considered hierarchically structured. The repeated measures for each individual at different times are level-1 observation units, and individuals become the level-2 units. A third level can be introduced into the data structure, if the higher level units (e.g., clinics) in which individuals are nested are available, to create a multilevel data with more than two levels.

Depending on the situation, some individuals may be considered as level-1 units while other individuals are higher level units. For example, patients and doctors can form a multilevel data structure. As a doctor treats multiple patients, the doctor may be considered as a level-2 unit, while patients are level-1 units. Similar situations include teachers and students, coaches and athletes, as well as interviewers and interviewees.

Finally, a special type of hierarchical data arise from meta-analysis in which results or findings from a series of related studies are summarized quantitatively to assess consistency or inconsistency of study results (Glass, 1976). In meta-analysis data, individuals are nested within specific studies. However, it is usually difficult or impossible for researchers to obtain the raw data from all the studies of interest. As such, a special approach is required for multilevel modeling of meta-analysis data. Detailed examples of formulating multilevel models for meta-analysis are available in many studies (Goldstein *et al.*, 2000; Raudenbush and Bryk, 2002).

1.3 Variables in multilevel data

In addition to the format of multilevel data, choosing the variables that describe the features of the distinct levels of the hierarchical structure is another important consideration. For multilevel analysis, the dependent variable is measured at the individual level and explanatory variables are measured at both individual and group level or at both micro and macro levels. As in regular statistical analysis, individual explanatory variables usually include socio-demographic characteristics (e.g., gender, ethnicity, education, age) and other measures such as psychological status and behaviors, depending upon the analyst's conceptual model.

Contextual variables are group level measures. They can be aggregate measures, such as mean values of some individual measures (e.g., average family income) or proportion of individuals for a particular characteristic within a particular context (e.g., percentage of minority population). These contextual variables represent the collective social characteristics of contexts/groups. They can be derived from either the sample or obtained separately from other sources such as census or government statistics.

Many contextual variables are not aggregations of individual information. Some characteristics are unique to contexts/groups and can't be captured at the individual level. For example, in studies of student school performance, contextual variables could include aggregate measures such as student gender ratio or average enrollment test scores, and school feature measures such as school ranking, student-teacher ratio or teacher's level of experience. The former is an aggregation of student data and can be generated from the sample; the latter represents contextual aspects of the schools that must be collected from other sources at the school level.

Contextual variables can also be categorical measures. For example, in a multilevel study on childhood obesity in which children are level-1 units and neighborhoods are the level-2 units. The researcher may include a dummy variable (1-yes; 0-no) to indicate whether there are fast-food restaurants in the neighborhood because easy access to fast food may have a significant impact on children's diet, thus on their obesity level.

Conceptually, one might use $J-1$ dummy variables to represent all the contextual features of the J groups. This approach, however, is not feasible even with a moderately large number of groups because too many dummy variables would be needed to represent the groups.

The following tables illustrate a fictional two-level data structure. Table 1.3.1 shows the individual level outcome variable y_{ij} and independent variable x_{ij} for the i^{th} individual in the j^{th} group. There are a total of n_j individuals in the j^{th} group, and individuals in all the groups sum up to the total sample size $\sum n_j = N$. z_j is a contextual variable describing the group. The values of the variable z_j for specific groups ($j = 1, 2, \ldots, J$ groups) are shown in Table 1.3.2.

Table 1.3.1 Individual level data

Unit		Variable	
Group	Individual	y_{ij}	x_{ij}
1	1	5	11
1	2	3	8
⋮	⋮	⋮	⋮
1	n_1	2	7
2	1	6	12
2	2	9	10
⋮	⋮	⋮	⋮
2	n_2	10	15
3	1	11	15
3	2	15	18
⋮	⋮	⋮	⋮
3	n_3	16	20
⋮	⋮	⋮	⋮
J	1	4	7
J	2	5	9
⋮	⋮	⋮	⋮
J	n_J	6	8

Note:

n_1, n_2, and n_J — Number of individuals in the first, second, and the j^{th} groups, respectively. $\sum n_j = N$.

y_{ij} and x_{ij}—Individual level outcome and independent variable, respectively.

Individual level data and group level data shown in Tables 1.3.1 and 1.3.2 are integrated into a mixed data set and shown in Table 1.3.3. When merging the data sets, both individual ID and group ID must be matched for every individual. As such, the same value of the contextual vari-

able z_j of group j is assigned to each individual in this group. Consequently, the value of z_j does not vary across individuals within the same group (see Table 1.3.3).

Table 1.3.2 Group level data

Group	z_j
1	8.7
2	12.3
3	17.6
⋮	⋮
J	8.0

Note:
z_j — Contextual variable at the group level.

Table 1.3.3 Individual and group level mixed data

Unit		Variable		
Group	Individual	y_{ij}	x_{ij}	z_j^*
1	1	5	11	8.7
1	2	3	8	8.7
⋮	⋮	⋮	⋮	⋮
1	n_1	2	7	8.7
2	1	6	12	12.3
2	2	9	10	12.3
⋮	⋮	⋮	⋮	⋮
2	n_2	10	15	12.3
3	1	11	15	17.6
3	2	15	18	17.6
⋮	⋮	⋮	⋮	⋮
3	n_3	16	20	17.6
⋮	⋮	⋮	⋮	⋮
J	1	4	7	8.0
J	2	5	9	8.0
⋮	⋮	⋮	⋮	⋮
J	n_J	6	8	8.0

Note:
*—The same value of contextual variable z_j of the j^{th} group is assigned to each individual in the group.

The data format for multilevel modeling varies slightly by computer programs. Some programs require separate individual and groups data sets while others work with mixed data formats like the one shown in Table 1.3.3.

1.4 Analytical problems with multilevel data

Prior to the availability of multilevel analytical techniques and computer programs, multilevel data were analyzed separately at a single level, either the individual level or the group level[1]:

[1] For the purpose of simplicity, only one independent variable is included in each model.

Individual level model:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij} \tag{1.4.1}$$

Group level model:

$$\bar{y}_j = \gamma_0 + \gamma_1 \bar{x}_j + \varepsilon_j \tag{1.4.2}$$

Equation 1.4.1 is a model at the individual level in which both dependent and explanatory variables are measured at the individual level. Equation 1.4.2 is a model at the aggregate or group level in which both dependent and explanatory variables are measured as the mean values of the corresponding individual level variables. The underlying problem encountered in such an approach is that it ignores the multilevel structure of the data. Model 1.4.1 ignores group membership and focuses exclusively on individual-level characteristics and inter-individual variation and thus ignores the potential importance of group-level features in influencing individual-level outcomes. Another serious problem with this model is that it assumes the independence of observations. As discussed in Section 1.1, generally individuals within each group are more alike compared with those in other groups. Thus the within-group observations are unlikely to be independent.

Model 1.4.1 cannot control the intraclass correlation coefficient (ICC), it ignores the within-group observation dependence, and thus violates the basic assumption underlying traditional regression models. As a result, standard errors of parameter estimates would be biased downwards, resulting in a large Type I error — falsely rejecting a true null hypothesis in statistical significance testing (De Leeuw and Kreft, 1986; Snijders and Bosker, 1999; Hox, 1998, 2002). Even a small ICC can lead to Type I errors that are much larger than the nominal alpha level. (Hox, 1998; Barcikowski, 1981). Consequently, analyzing multilevel data with traditional regression models can produce misleading conclusions.

Model 1.4.2 focuses exclusively on the inter-group variation and on the data aggregated to the group level. The group-level model eliminates the observation dependence problem, but ignores the role of individual-level variables in shaping the outcome on one hand; and on the other hand, it substantially reduces statistical power by using a group level sample with a much smaller sample size.

Traditionally, researchers tended to use model results at one level to draw statistical inferences at another level. This has proven incorrect. The results from the two single level models frequently differ either in magnitude or in sign. The relationships found at the group level are not reliable predictors for relationships at the individual level, and vice versa. This phenomenon is known as the *ecological fallacy, aggregation bias,* or the *Robinson effect* (Robinson, 1950).

What causes *Robinsin effect*? Model 1.4.2 analyzes the variation in variable $\bar{y}_{ij}$ at the group level. *Aggregating* individual measures changes their meaning. If x_{ij} is a continuous measure (e.g., age), then $\bar{x}_j$ would be the average or mean value of the x_{ij} (e.g., mean age) in the j^{th} group of individuals. If x_{ij} is a dichotomous variable, denoting gender (e.g., 1-male; 0-female), then $\bar{x}_j$ would be the proportion of males in the j^{th} group. Clearly, x_{ij} and $\bar{x}_j$ are different measures, and we should not expect them to have the same effect in separate models based on either individual or group data.

A critical analytical problem with multilevel data is the *heterogeneity of relationships* of independent variables with the dependent variable. The relationship between individual level dependent and independent variables may vary across groups. For example, suppose we were studying academic performance of minority students in high schools. The average academic performance score for the minority students may vary across schools. The effect of minority status on the dependent variable may vary across schools for a variety of reasons. The proportion of minority students in a school, a "sample composition contextual variable" might partially

account for the variation in performance in addition to other contextual variables at the school level.

In the past, heterogeneity of micro level relationships was often studied using the following fixed-effect regression model:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 \bar{x}_j + \beta_3 x_{ij} \cdot \bar{x}_j + \varepsilon_{ij} \quad (1.4.3)$$

where y_{ij} denotes the performance score for the i^{th} student in the j^{th} school; x_{ij} is a dummy variable (1-yes; 0-no) indicating the minority status at the student level; and $\bar{x}_j$ denotes the proportion of minority students in the j^{th} school. In this model the macro-level (e.g., school level) variables were disaggregated to the micro-level (e.g., student level). In this example, students are assigned various school-level variables and all students in the same school are assigned the same value on the school-level variable (e.g., $\bar{x}_j$). The model is then run at the student level. Slope coefficients of β_1 and β_2 are the main effects of the individual level variable x_{ij} and school level variable $\bar{x}_j$ respectively. The slope coefficient β_3 is the interaction of these two variables. If the cross-level interaction is statistically significant, we conclude that the relation between student's minority status and the performance score is influenced or moderated by the proportion of minority students at the school level. This kind of model takes into account the effects of contextual variables on the relationships of individual explanatory variables and the dependent variable at individual level.

One serious problem with this model is that it treats observations as independent though they are not, thus leading to biased standard error estimates. In addition, in this fixed-effect model, the variation in the intercept and slope coefficients are assumed to be perfectly explained by group level variables without error, which is highly unlikely.

Van de Eeden (1988) and others have examined the *heterogeneity of relationships* problem using a two-step approach. In Step 1, they estimated the individual level regression models for each group separately. The assumption of invariance in the intercept and slope coefficients is tested by running multi-group regression models with and without equality restrictions on the coefficients across groups, using structural equation modeling software such as LISREL. If the coefficients show significant variance across groups, then the second step is to regress each of the regression coefficients on the contextual variables at the group level.

Although this approach enables analysts to test the significance of variations in the regression coefficients estimated in Step 1, it has several limitations. OLS models are used at both Steps 1 and 2, even though it is technically incorrect to use OLS to estimate the standard errors in the second step (De Leeuw & Kreft, 1986, p. 61). It is also impractical to run separate regression for each group when the number of groups is large, and particularly when the number of observations per group is small. This approach treats the groups as unrelated and ignores the likelihood that the groups are drawn from a larger population of groups that share common attributes.

Given the shortcomings of traditional methods, a new statistical method, called multilevel modeling is needed. Multilevel models are explicitly designed to analyze hierarchically structured data, modeling variables at both micro and macro levels simultaneously without aggregation or disaggregation. In the following section we will discuss the advantages as well as the limitations of multilevel models for multilevel data analysis.