Open Journal of Statistics, 2024, 14, 689-720
https://www.scirp.org/journal/ojs
ISSN Online: 2161-7198
ISSN Print: 2161-718X

Multilevel Modeling Approach for Hierarchical Data an Empirical Investigation

Hiba Faisal*, Muhammad Zakria, Nazakat Ali*, Nimra Khalid

Department of Mathematics and Statistics, University of Agriculture, Faisalabad, Pakistan
Email: *hibafaisal016@gmail.com, *nazakatalibandesha@gmail.com

How to cite this paper: Faisal, H., Zakria, M., Ali, N. and Khalid, N. (2024) Multilevel Modeling Approach for Hierarchical Data an Empirical Investigation. *Open Journal of Statistics*, **14**, 689-720.
https://doi.org/10.4236/ojs.2024.146031

Received: October 25, 2024
Accepted: November 26, 2024
Published: November 29, 2024

Abstract

Multilevel modeling (MLM) has emerged as a powerful statistical framework for analyzing complex data structures with nested relationships. With its hierarchical modeling approach, MLM enables researchers to account for dependencies and variations within and between different levels of a hierarchy. By explicitly modeling these relationships, MLM provides a robust and accurate analysis of data. It has become increasingly popular in the field of education. MLM enables the investigation of various research issues, the evaluation of individual and group-level indicators, and the calculation of both fixed and random effects. Overall, MLM revolutionizes data analysis by uncovering patterns, understanding contextual effects, and making more precise statistical inferences in complex datasets. For fitting multilevel models in R, use lmer function provided by lme4 package. Through this examination, the use of a multilevel model is expected to increase and revolutionize data analysis and decision-making. The Constrained Intermediate Model (CIM) and Augmented Intermediate model (AIM) deviation are compared using the Likelihood-ratio (LR) test and the ANOVA function. This study analyzes student results from the University of Agriculture Faisalabad, collected via stratified random sampling. A linear mixed-effect model under multilevel modeling estimates the impact on CGPA, considering department, gender, intermediate marks, and entry test scores. These results indicate that Entry test is a significant predictor of CGPA, but the effect of department identifier CMC on CGPA is not statistically significant.

Keywords

Multilevel Modeling, Hierarchical Modeling, Intraclass Correlation Coefficient, Design Effect, Augmented Intermediate Model

1. Introduction

In recent years, multilevel modeling has developed into a potent technique for analyzing layered data structures and exploring complicated phenomena. In this method, study variables and parameters are specified at two or more levels, or stages, for example, pupils within schools or patients within clinics.

The method is sometimes referred to as hierarchical modeling or regression since these levels are organized in a hierarchy. Multilevel models were first introduced by Peter Diggle and Harvey Goldstein in 1970; a large portion of the early research was carried out by British statisticians.

Since it enables researchers to take into consideration hierarchical dependencies and nested linkages within their datasets, multilevel modeling has become a strong statistical framework for studying complicated data structures [1].

1.1. Background and Motivation for Studying Multilevel Models

Multilevel modeling (MLM), also referred to as hierarchical or nested modeling, originated in the 1950s. It gained prominence as psychologists and sociologists began to understand the need to consider the layered structure of social phenomena. Its use increased significantly in the 1990s with advancements in software and computing power, leading to applications in education, public health, and organizational research. MLM analyzes hierarchical data structures, such as students within schools or patients within hospitals, by modeling dependencies and variances across levels. It helps avoid biased estimates and incorrect statistical inferences, allowing researchers to study relationships within and between levels. MLM is particularly valuable for exploring the impact of individual and group-level predictors, contextual factors, and interactions across different levels. By incorporating both fixed and random effects, MLM offers a comprehensive approach to understanding complex data, making it a powerful tool for studies in sociology, psychology, and education.

1.2. Hierarchical Data

The development of specific statistical approaches is required due to the widespread use of hierarchical data structures in many different domains. When observations are arranged into nested groups or levels, hierarchical data are produced. To conduct effective statistical modeling, it is essential to comprehend the structure and features of hierarchical data.

At many levels, hierarchical data show linkages and variability. According to studies on schooling, for instance, students from the same school may be more similar than students from other schools due to overlapping environmental influences.

Additionally, they enable the study of cross-level interactions, allowing researchers to investigate relationships between higher levels and individual-level traits. Researchers can better understand the fundamental processes and mechanisms in hierarchical data by utilizing the flexibility and robustness of multilevel models.

1.3. Need for Multilevel Models

Before analyzing a nested dataset, it's important to determine if multilevel modeling is necessary. Not all nested datasets require this approach. If the response variable scores show no variation between Level-2 units (such as schools), ordinary least squares (OLS) multiple regression can be used for analysis. What follows is the new question, "How much change is there in the response variable in Level 2". Calculating the ICC and design effect statistics find the answer to this puzzle. Using a sample of student achievement scores from a set of NELS data, researchers can predict that student achievement scores will vary across educational institutions due to individual differences in ability and motivation, etc.

However, by averaging the science achievement scores of all students within each school, the science performance score for each school may be determined [2].

Multilevel modeling is necessary to distinguish the variance present both among students and between schools, especially when differences exist in the average scientific performance levels across schools. Traditional multiple regression methods are designed to account for variation in the response variable at only one level of analysis, such as students or schools, but not both simultaneously. Ordinary least squares regression and other conventional statistical methods make the assumption that the observations are independent.

1.4. An Empirical Investigation

This study conducts an empirical analysis to demonstrate the benefits of multilevel models in hierarchical data analysis, highlighting limitations of conventional techniques. By exploring various sample sizes, nesting levels, and distributional assumptions, we assess multilevel model performance using metrics like prediction accuracy, design effects, ICC, and goodness-of-fit indices. Insights aim to guide researchers on optimal conditions for multilevel modeling, supported by thorough comparisons with empirical results.

1.5. Intraclass Correlation Coefficient

The intra-class correlation coefficient (ICC) measures the degree of similarity within clusters, guiding the choice of using multilevel models. A high ICC indicates strong clustering, meaning observations within groups are more similar to each other than to those in other groups, while a low ICC suggests minimal clustering, making traditional statistical methods suitable. ICC values range from 0 to 1; values above 0.8 imply high reliability, while those below 0.5 indicate poor reliability. It is commonly used in quantitative assessments where consistency among grouped observations is essential, such as evaluating relatedness among siblings or consistency among different observers.

1.6. Design Effect

Multilevel modeling core idea of "design effect" takes interdependence or clustering into consideration when making statistical inferences. "Deff" stands for design effect. Because observations in hierarchical data are nested within groups or levels, there may be correlations or similarities between groupings.

The assumption of independence that is frequently made in conventional statistical analysis techniques is broken by this clustering.

The design effect assesses how the sample size required for precise estimates rises or reduces in comparison to independent data. It quantifies the inflation or deflation of the variance owing to clustering.

Design effect (deff) estimation for various sampling designs is described by Kish (1965). Design effect expressed as a ratio of operating sampling variance to simple randomization of variance.

A factor must be multiplied by the simple random sampling variance to obtain the real operating sampling variance. Defined as (1 + rho (n − 1)) in a basic cluster sampling scenario with equal cluster sizes, rho is the intraclass correlation, and n is the typical cluster size.

The objectives of this research are:

1) To evaluate the effectiveness of multilevel modeling (MLM) in analyzing nested educational data from the University of Agriculture Faisalabad, capturing variations within and between hierarchical levels such as departments and student demographics.

2) To examine the impact of predictors, including department, gender, intermediate marks, and entry test scores, on students' CGPA through a linear mixed-effects model using R's lme4 package.

3) To compare CIM and AIM models via likelihood ratio (LR) tests and ANOVA, assessing the predictive power of entry test scores on CGPA, while examining the non-significant effect of department identifiers.

2. Reviews of Literature

[3] found that emphasis of this essay had been on using SPSS 8. 0 output to interpret measurement error that exists in data. Before taking into account individual heterogeneity, error in measurement size (RMSE), and the variance between multiple measured averages, the r coefficient could not be properly evaluated. Saw Morrow and Jackson (1993) for details on how to appropriately present reliable results. [4] examined how multiple measurements that were not time-based. In situation of missing data, a single instance shows how to properly assess independent variables for categorical, continuous, or semi-continuous experimental stimuli or subjects.

[5] found in recent years, researchers in the social and psychological sciences have become more and more interested in using growth mixture and latent class modelling techniques. This has been made possible in part by technical improvements and the availability of computer software designed with this goal in mind. Two latent growth modelling techniques, growth mixture modelling (gmm) and latent class growth analysis (lcga), have grown in prominence due to their capacity to recognise significant groupings or classes of individuals as well as homogenous

subpopulations within larger heterogeneous populations. [6] examined how longitudinal models are often used in the behavioural sciences due to their main advantages, which include better power, more comprehensive evaluation, and the establishment of time precedence. The capacity to discern between inter-person and within-person influences during the regression of a result on a time-varying covariate was one especially notable benefit provided by longitudinal data. It's interesting to note that not all applications based on social science research have fully utilised the ability to divide apart these consequences. [7] The degree of bias depends on the informativeness of any ignored clustering in the sampling design and stratification. While some multilevel software programs support the inclusion of first-stage sampling design data in two-level models, not all do. Dependent variables are chosen based on previous research using these data sets. The implications of ignoring the sample design in two-level models both unconditional and conditional—were demonstrated using five examples from publicly available data sets.

[8] demonstrated that multilevel contingency and interactionism theories were based on cross-level interaction effects. Investigators had frequently bemoaned how challenging it was to detect hypothesized cross-level relationships, and there was currently no way to assess the statistical power of such experiments. We created such a strategy, validated its correctness, and presented evidence about the relative significance of elements that impact power to identify interactions of cross-level through the results of large-scale simulation research. Our findings suggested that magnitude interaction of cross-level, the standard deviation of lowered leveled slopes, lowered or upper leveled sample sizes were the main factors that affect the ability of statistical models to detect cross-level interactions. We offered a Monte Carlo tool to researchers so they could a priori design.

[9] discovered that to provide information on fieldwork organisation and methodology, sample design, weighting, and concerns for the use of design-based vs model-based estimates, this paper outlines the key aspects of the NCS-R design and field operations. Empirical data were given on the non-response bias, the design effect, and the trade-off between bias and efficiency-response bias. [10] studied the multilevel modelling approach, which offered several potential applications in personality and social psychology. This article offered an introduction to multilevel modelling with a focus on some of its applications in social and personality psychology in order to enable the author to fulfil his promise. Multilevel modelling was defined in this lecture, along with arguments in favour of it and an overview of its use in social and personality psychology research. Along with some of the subtleties of setting up multilevel studies and interpreting data, software solutions were explained.

[11] found that the intraclass correlation coefficient (ICC) from a one-way random effects model was widely used to dependability mean judgements in the overall behavioural, educational, and psychological investigations. Despite its evident use, the feature of ICC (2) as a focused estimator of average score intraclass correlation coefficient was hardly studied [12]. The study empirically assesses the

consequences of neglecting sampling strategy in two-level analyses weighted sets from the National Center Education Statistics (NCES). Currently, researchers have the option to ignore sampling designs above the levels they model, which could lead to biased standard error estimates that lead to incorrect inferences about hypotheses.

[13] examined the intraclass correlation coefficient, an indicator of consistency that measures both the degree of correlation and agreement between measurements. When a continuous data set met the criteria for applying parametric approaches, the ICC was frequently employed in orthodontic research. The ICC, however, had a total of ten distinct versions that were sometimes overlooked by researchers and may produce varied results. [14] has been discovered that physical multilevel models of deformation that were rigid with explanations for the changing nature of the structure of the content had the potential to lead to the creation of useful materials. In this research, we offered an enhanced statistical multilevel model for understanding thermomechanical process of polycrystals, which included a description of the dynamic recrystallization process, modified by analysing the mutual arrangement of crystallites [15]. The proposed authoritative voiced paper provides a thorough defence of multilevel modelling. In addition to encouraging used of multilevel modelling in company information systems or serving as inspiration for further study, it intends to conduct a comprehensive assessment of its prospects. Comparisons between instruments, general-purpose modelling languages, property modelling languages, and multilevel modelling were used to create the assessment.

3. Materials and Methods

3.1. Methodology

According to [16], the logic of scientific technique is called methodology. The systematic process of conducting research involves looking for fresh information in order to confirm existing information and the natural laws that underlie it. The scientific method is a set of precise guidelines and methods that serve as the foundation for research and as a yardstick by which to measure assertions of truth. The description of the study region, definitions of the materials used, and key components of the current investigation are all explained in this section of the study.

3.2. Data

For the above-mentioned purpose, the data of Multilevel model (MLM) Characterized by CGPA, Entry test marks, inter marks and collected data from Department of Mathematics & Statistics, Biochemistry, Chemistry, Computer Science, Botany, Zoology, Wildlife & Fisheries and Physics, University of agricultural Faisalabad.

3.3. Multilevel Modeling

Hierarchically organised data are common in research contexts, including educational, clinical, and other types. Just as students live in classrooms or teachers live in schools, teachers also live in schools. As an alternative, local public servant organisations may, in turn, be nested inside service receivers and house social workers who deliver services. Studies at any of these levels will yield misleading results if the specific level (students) or content level (school) is not included.

In order to appropriately account for the hierarchical (correlated) nesting of data, multilevel models have been created [17].

A statistical model known as multilevel modelling is used to simulate the relationship between the dependent variable and the independent variable when there is a correlation between the data.

These models are also known

- Hierarchical models
- Mixed effect models
- Nested data models
- Random coefficient models

Multilevel modeling analyzes relationships between lower-level and higher-level variables, known as cross-level interactions, and distinguishes individuals from contextual effects. It is essential to assess whether multilevel modeling is necessary before analyzing hierarchical datasets. If there is no variability in the response variable across Level 2 units (such as schools), using OLS regression might be appropriate. To determine if multilevel modeling is needed, calculate the ICC and design effect. These models explore different research questions, including analyzing how socioeconomic status (SES) relates to reading achievement at both the individual and school levels, accounting for influences from both perspectives [18].

The purpose of multilevel model is to determine constant and unpredictable effects at both and levels by taking into account the dependency of values in each group within (Level 1) and then between (Level 2).

Table 1. A few illustrations of mega and micro units

Mega-level	Micro-level
Schools	Teachers
Classes	Pupils
Neighbourhoods	Families
Districts	Voters
Departments	Employees
Families	Children
Doctors	Patients
Interviewers	Respondents
Firms	Departments

This results in more accurate estimates of the relationships between variables, especially when the within-group dependence is substantial. A variety of study topics are explored using multilevel models, such as understanding how collaborating impact outcomes, and estimating the variability of outcomes at different levels (Table 1).

Examining the relationship between variables at multiple levels of analysis. The ICC or design effect are important considerations in the design and analysis of multilevel models.

3.4. Types of Multilevel Modeling

Hierarchical Data

Different "levels" of the data (and analyses) must be discussed by researchers due to the nature of the data in hierarchical models.

Utilizing numbers is one method of doing this. The lowest level is often referred to as "Level-1," the next level above is "Level-2," and so on. In the illustration, "Level-1" denotes the inhabitants, and "Level-2" denotes the communities (Figure 1).

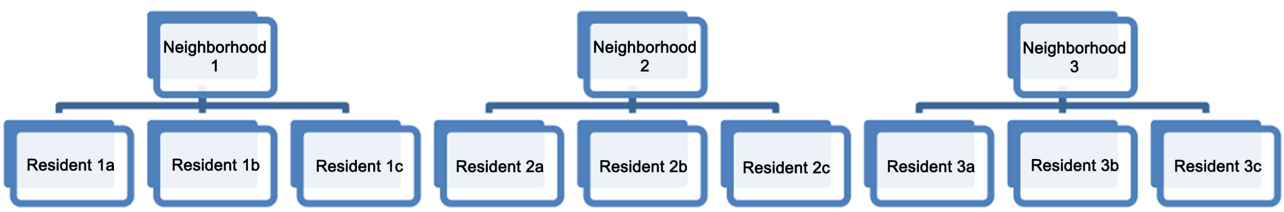

Figure 1. Two-level hierarchical data structure.

Using descriptive language is another method for expressing the stages. In the context of public health, this could refer to "resident level" as opposed to "neighbourhood level." Nested data frequently have more than two levels.

Using the public health scenario as an example, residents can be nested within towns and cities, and those 8 neighbourhoods may be nested within cities, creating a three-level nesting structure in Figure 2.

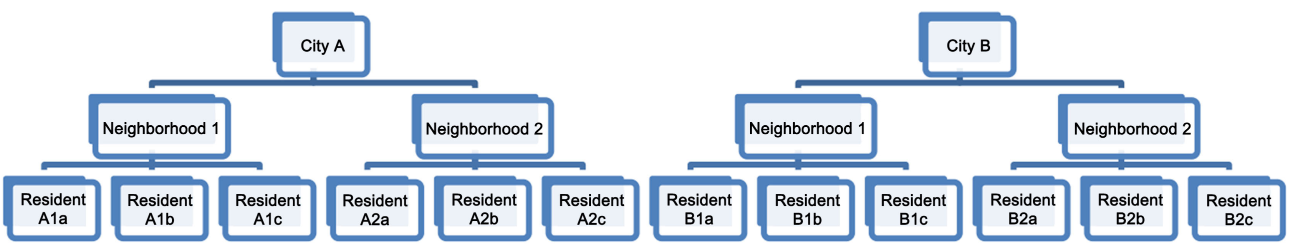

Figure 2. Three-level hierarchical data structure.

Longitudinal studies can also be seen as having a nested data building, where repeated measurements are collected over time for the same individuals. Additionally, individuals may be grouped within higher-level clusters. Figure 3 illustrates this model. In these models, it is typical to center the Level-1 time covariate on a fixed value rather than the mean. As a result, this dissertation will emphasize cross-sectional models.

Multilevel models provide techniques for studying hierarchical data and improving individual effect estimation [19]. For instance, a person might only be concerned with Level-1 (the resident level) impacts. It could possibly be inclined to perform a regression analysis that takes each resident's score as independent and disregards the fact that the data are nested.

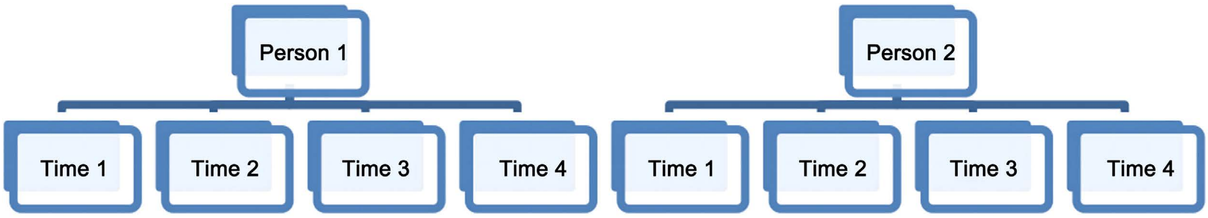

Figure 3. Example of longitudinal data represented in two-level model.

The problem with this strategy is that people who live in the same neighbourhood are probably more alike than people who live in separate neighbourhoods.

The regression analysis's independence assumption is broken as a result. Models that do not take this violation into account will have inaccurate estimations of the standard error and ineffective parameter values.

Researchers can evaluate data using multilevel models, which improve parameter and standard error estimation and take into account the hierarchical nature of the data.

To estimate regression equations for samples that are too small to do so on their own (such as modest numbers of minority students across 25 colleges).

[19] points out that multilevel models can draw strength from a wider pool of data. The modeling of distinct random errors 9 for each Level-2 unit improves standard error estimates in multilevel models by addressing the dependence between individuals belonging to the same Level-2 group.

3.5. Repeated Measure Data

In repeated measure data, the observations within each individual or group are not independent, as they are likely to be similar to each other observations from other persons or groups. Multilevel modeling takes into account this hierarchical structure and allows for the analysis of both within-individual/group variability and between-individual/group variability.

The basic idea behind multilevel modeling is to model the dependent variable at different levels. In the case of repeated measure data, the lowest level (Level 1) represents individual observations, and higher level (Level 2) represents individuals or groups from which observations are nested.

By incorporating random effects at different levels, multilevel models can estimate the variability within and between levels. This allows for the examination of individual differences in the dependent variable, as well as the identification of contextual factors that may explain the variability between individuals or groups.

Multilevel models can handle various types of data, including continuous, binary, and categorical outcomes. They also accommodate unbalanced designs, missing data, and unequal spacing between measurements.

3.6. Random Effect Data

In statistical analysis, random effects refer to sources of variability that are considered random or unobserved in the population being studied. These sources of variability are assumed to be randomly selected from a larger population or distribution. Random effects are used to model and account for the variation that exists between different groups or clusters within the data.

For example, consider a study examining the test scores of students from different schools. The schools would be considered as random effects because we are not interested in making specific inferences about individual schools but rather want to generalize the findings to a larger population of schools (Figure 4).

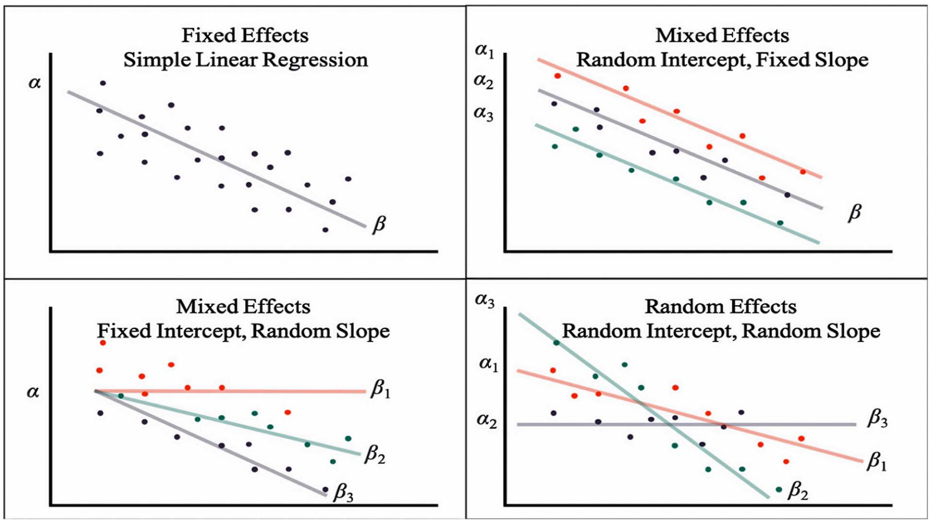

Figure 4. Random effects and fixed effects slope.

3.7. Multiple Outcome Data

Multiple outcome data refers to a situation in which multiple outcomes or dependent variables are measured for each unit or observation in a study. Instead of focusing on a single outcome variable, researchers may collect data on multiple outcomes to gain a more comprehensive understanding of the phenomenon under investigation [20].

Using multilevel modelling, more than one dependent variable (DV) may be contained in a model. Once more, the researcher is required to make this clear and use software that enables multivariate multilevel modelling (MMLM), also known as multivariate linear mixed modelling (MLMM) or hierarchical multivariate linear modelling (HMLM).

3.8. Multivariate Analysis

Multivariate analysis techniques, such as multivariate regression or multivariate analysis of variance (MANOVA), allow for the simultaneous analysis of multiple outcome variables while accounting for their interrelationships.

3.8.1. Structural Equation Modeling (SEM)

Statistical method SEM enables complex relationships examination among multiple variables, including multiple outcomes. It allows for the testing of direct and indirect effects among variables and can provide insights into causal relationships.

3.8.2. Latent Class Analysis (LCA)

LCA is a statistical method that uses patterns of various outcomes to find latent (unobserved) subgroups within a population. It can help uncover distinct profiles or classes of individuals based on their response patterns.

3.9. Nested Data and Cluster Sampling Designs

Multilevel modeling is the presumption that the various observations within a sample will have independently distributed error terms. This presumption basically states that once the independent factors in the analysis are taken into account, there are no correlations between the people in the sample for the dependent variable.

Multilevel modeling gives a direct mechanism to add indicators for clusters at all levels in complex hierarchical data where the data are hierarchical and hence clustered or acquired from clustered sampling.

Multilevel analysis offers proper standard errors by exploiting the cluster information, which are typically more conservative than those produced by ignoring the presence of clustering.

3.10. REML Estimation

REML estimation is used to provide estimates for maximum likelihood of the variance parameters. These values are then used to estimate the constant effects using extended least squares methods.

As opposed to ML estimates, REML estimates the variance parameters while taking into account uncertainty in the fixed effects. One would assume that given the fixed effects' uncertainty is more pronounced with lower sample sizes, the discrepancies between different methodologies will normally be larger when sample sizes are less.

Several empirical studies have noted differences between ML and REML estimates across various scenarios, though these studies do not consistently favor one approach over the other. As sample sizes increase, the sampling distributions of these estimates tend to approach a normal distribution. This leads to asymptotic efficiency in the estimates of fixed effects and variance parameters, and asymptotically unbiased estimates for both. Thus, ML and REML are often recommended for larger sample sizes. However, with smaller sample sizes, particularly when data are irregular, the reliability of ML and REML diminishes, prompting researchers to explore alternative methods. Moreover, assumptions of error-free variance estimates when inferring fixed effects become increasingly tenuous with smaller sample sizes.

3.11. Hypothesis Testing Likelihood Ratio Testing

An essential part of interpreting any model is testing hypotheses. Knowing whether a parameter is important or not is essential. Depending on the parameter being observed, different statistical tests will be used. For our fixed effect parameters, we can employ the conventional z-tests and t-tests. However, likelihood ratio testing will be required to look for random effects.

Making use of likelihood ratios in reality, likelihood ratio tests are rather simple to interpret. Consider a situation where the model's intercept is random. We will fit the model with and without a random intercept and compute the log-likelihood of each model in order to do an LRT.

3.12. Intraclass Correlation

When people are grouped inside an advanced level unit (for example, a classroom, school, or school district), the intraclass correlation (ICC, abbreviated _I in the population) can be used to quantify the correlation between individual scores inside a cluster or nested structure [21].

The ρ_I is a measurement of how much of the outcome variable's variance occurs between groups compared to all other variations. Alternatively, it can be thought of as the correlation between measurements of two randomly selected individuals from the same group. Its range is 0 (no variation among clusters) to 1 (variance among clusters but no within-cluster variance). It can be said to be

$$\text{ICC} = \frac{\text{Between} - \text{Cluster variance}}{\text{Total variance}} \tag{3.1}$$

$$\rho_I = \frac{\tau^2}{\tau^2 + \sigma^2} \tag{3.2}$$

where τ^2 denotes clusters between population variance and σ^2 indicates variations in population within clusters. Higher values of I show a stronger correlation between introductions to Multilevel Data Structure 25 scores of two cluster members, indicating that a larger proportion of the outcome measure's overall variation is related to cluster membership.

Another way to think about this problem is that people who belong to the same cluster (say a school) are more similar to one another than people who belong to different groups on the measured variable.

$$0 \le \text{ICC} \le 1 \tag{3.3}$$

The ICC is a measure of the reliability or consistency ratings of a set or measurements different judges by made. In the context of multilevel models (MLMs). The ICC can vary from 0 to 1, where 0 denotes total disagreement and 1 denotes complete agreement among evaluators. An ICC value greater than 0.75 is generally considered to indicate good agreement.

In a multilevel model, ICC can be used to quantify the degree of clustering in the data; this may be used to choose a right number of random variables for model.

Additionally, group-level predictors' dependability is evaluated using it and to quantify the extent to which variability in response variables can be attributed to differences between higher-level units.

The ICC is an essential metric in multilevel modeling, as it highlights the influence that a hierarchical data structure can have on the outcome variable. Higher ICC values indicate that clustering has a greater effect. As a result, as the ICC value increases, we must exercise greater caution while employing multilevel modelling approaches for data analysis.

Before moving on to strategies for dealing with it directly, we will first talk about the issues of disregarding this layered structure in the following section.

3.13. Design Effect

The Design Effect is more informative to decide whether multilevel modelling is required [22]. When examining intricate surveys or clustered data, the field of education and other applicable disciplines frequently refer to the design impact.

Design effect demonstrates the cluster-related inflection in the estimate's variability (Level 2).

$$\text{DEFF} = 1 + (n-1) \times \rho \tag{3.4}$$

$$\text{DEFF} = n - 1 + \rho \tag{3.5}$$

n = expected number of measurements made each cluster;

ρ = ICC for that variable;

If DEFF > 1, account for clustering and consider LMM for modeling.

- One may simply disregard the hierarchical nature of their data and perform conventional regression when the DEFF is less than 1.5.
- The confidence intervals must be three times as wide as they would be for a simple random sample, for instance, if the DEFF is 3.04.

3.14. Multilevel Linear Models

The fundamental principles of MLMs. Our intention is to acquaint readers with terminology that appears repeatedly throughout the book and explain them in a way that is largely nontechnical. Prior to discussing the fundamentals of parameter estimation, we should first distinguish between random and fixed effects.

Next, they should address the two widely used methods: maximum likelihood estimation (MLE) and restricted maximum likelihood (REML). Finally, we should review the underlying presuppositions of MLMs and provide an overview of their most common applications, along with examples.

We shall discuss the topic of centering in this part as well as the significance of this idea in MLM. The reader will have enough technical knowledge of MLMs after finishing this chapter to start utilizing the R software package for fitting MLMs.

3.15. Random Intercept

Basic simple linear regression model

$$y = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (3.6)$$

- A function of an independent variable, x, is used to express the dependent variable y;
- A slope coefficient of β_1 multiplied;
- Intercept β_0;
- Unpredictable fluctuation between subjects ε.

We defined the intercept by using the conditional mean of y when the value of x is zero. One intercept is shared by every member of the target population in a single-level regression model like this one.

However, when individuals are grouped in some way (such as students in classes and schools, or organizational groups in the firm), it is possible that each group has a different intercept, that is, specific means by which the dependent variables can exist for $x = 0$ in groups. The single intercept model, as shown in Equation (3.6), works well when there is no clustering effect. Determining whether differences in means exist among clusters is an empirical question. It's important to note that this discussion focuses specifically on scenarios where the intercept varies across clusters. It is also possible for β_1 to change depending on the group or possibly other model coefficients from more complex models.

$$y_{ij} = \beta_{0j} + \beta_{1j} x + \varepsilon_{ij} \quad (3.7)$$

where the ith person in the jth group is represented by the subscript ij. We will begin analyzing MLM signatures and models by working with the simplest multilevel model, which predicts only the outcome of the effect that we allow to vary for each group.

$$y_{ij} = \beta_{0j} + \varepsilon_{ij} \quad (3.8)$$

This leads to the random intercept that we express as

$$\beta_{0j} = \gamma_{00} + U_{0j} \quad (3.9)$$

In this framework,

γ_{00} represents an average or general intercept value that holds across clusters as a fixed effect because it remains constant across all clusters.

Whereas U_{0j} is a group-specific effect on the intercept and random effect because it varies from cluster to cluster.

$$y = \gamma_{00} + U_{0j} + \beta_{1x} + \varepsilon \quad (3.10)$$

In multilevel modeling (MLM), our analysis of a dataset often begins with a basic random intercept model, commonly referred to as the null model, represented in the following form

$$y_{ij} = \gamma_{00} + U_{0j} + \varepsilon \quad (3.11)$$

The null model serves as a foundational reference point for constructing and comparing other models, as will be discussed in subsequent sections.

3.16. Centering

First, consider how to centre predictor(s). estimate various effects depending on the type of centring use, especially with regard to Level-1 predictor(s).

When it comes to Level-1 predictors, there are essentially only two centring methods: grand-mean centring and cluster-mean centring. Be cautious since certain software programs that aren't included in this primer could centre the variables.

Regression analysis "centring" a predictor variable (X). Variables that have departed from a specific value are said to be centred.

An example is a raw score variable (X_{ij}) that deviates from the overall mean ($\bar{X}_{..}$) of the sample:

$$X_{CGM} = X_{ij} - \bar{X}_{..} \tag{3.12}$$

The sample mean is not the only value that can be utilized. Centring is occasionally used in regression studies on data that are not hierarchically nested to generate a meaningful zero point for a measure that lacks one.

The X measuring a person's height or weight is a prime illustration of this. In this situation, the predicted value of Y when X is zero would be the intercept's interpretation. This is not very helpful as neither height nor weight can ever be zero.

But when X is centered on the sample mean, the intercept is interpreted as the predicted Y for a person whose weight or height is at the sample mean of X.

There are two options for mean-centering the Level-1 predictor variable in two-level models.

The first choice is to focus on the sample as a whole's mean. The term "grand mean centering" or "centering grand mean" (CGM) is used to describe this.

This translates to taking the mean of X across all residents, regardless of where they dwell (neighbourhood or city), and deducting that value from the observed X for each resident in the public health example above:

$$X_{CGM} = X_{ij} - \bar{X}_{..} \tag{3.13}$$

The second alternative is to place the Level 1 predictor variable's centre of gravity on the Level-2 group's mean of that variable.

The term "Centering within context" (CWC) is used to describe this. In the public health example, every neighbourhood has a unique mean for X, which is the total mean of X for the local population.

The X that is centered within context is the X for a particular resident less the mean for that resident's neighbourhood:

$$X_{CWC} = X_{ij} - \bar{X}_{.j} \tag{3.14}$$

The Centering options increase in the three-level model. The Level-1 X variable can also be centered on the Level-3 mean in addition to the CWC.

$$X_{CWC\text{-}L3} = X_{ijk} - \overline{X}_{..k} \quad (3.15)$$

To indicate that it is being centered on the Level 3 mean, I designate this type of Centering within context as CWC-L3.

Although it is a Centering alternative, there isn't a compelling substantive argument for why this Centering approach would be beneficial in a three-level model. As a result, this method of Centering is not further discussed in this dissertation.

3.17. Assumptions Underlying MLMs

This section explains these presumptions and discusses how they affect MLM-using researchers. In later chapters, we go over how to verify that these presumptions are true for particular sets of data.

At Level 2, it is initially assumed that the residuals are independent across clusters. Additionally, the random intercept and slope(s) at this level are considered to be independent of each other when comparing across different clusters. Second, it is assumed that the Level 2 intercepts and coefficients are independent of the Level 1 residual, meaning that errors in the estimates at the cluster and individual levels are unconnected.

Finally, the Level 1 residuals exhibit constant variance and follow a normal distribution, similar to the assumptions made for residuals in traditional linear regression models. Additionally, the Level 2 intercept and slope(s) are assumed to follow a multivariate normal distribution with a consistent covariance matrix.

3.18. Overview of Two-Level MLMs

Specific requirements for Multilevel Modeling (MLM) include Residuals with Random Effects at Level 1 and Level 2. This section will include examples of two- and three-level MLMs in addition to using MLMs with information that goes beyond implementation research. By the end of this section, readers will have a solid foundation for using R to evaluate MLMs in the following chapters. We will begin by examining the random slope model and the two-level MLM.

$$y_{ij} = \gamma_{00} + \gamma_{00} x_{ij} + U_{0j} + U_{1j} x_{ij} + \varepsilon_{ij} \quad (3.16)$$

The autonomous variable x_{ij} and random errors at both the student and school levels were influenced by the dependent variable y_{ij} (reading achievement).

The two parts of this model are expressed as

$$\text{Level 1}: y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_{ij} \quad (3.17)$$

$$\text{Level 2}: \beta_{hj} = \gamma_{h0} + \gamma_{h1} z_j + U_{hj} \quad (3.18)$$

$\gamma_{h1} z_j$, which represents the slope for (γ_{h1}), and value of the average vocabulary score for the school (z_j). a single equation for the two-level MLM.

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + \gamma_{01} z_j + \gamma_{1001} x_{ij} z_j + U_{0j} + U_{1j} x_{ij} + \varepsilon_{ij} \tag{3.19}$$

The term "cross-level interaction" refers to the interaction between predictors at Level 1 and Level 2.

γ_{00} intercept or grand mean for the model;

γ_{10} fixed effect of variable x on the outcome;

U_{0j} represents random variation for the intercept across groups;

U_{1j} represents random variation for the slope across groups;

γ_{01} represents fixed effect of Level 2 variable z (average vocabulary) on the outcome;

γ_{11} represents slope for and value of the average vocabulary score for the school.

The degree of vocabulary performance at a student's school may have a considerable impact on the association between their vocabulary test score and overall reading achievement, according to a significant value for this coefficient (Table 2, Table 3).

Table 2. Model building process for 2-level linear models.

Model 1	No predictors, just random effect for the intercept	Output used to calculate ICC provides information on how much variation in the outcome exists between Level-2 and Level-3 units
Model 2	Model1 + Level-1 fixed effects	The association between Level-1 predictors and the outcome is shown by the results.
Model 3	Model 2 + random slopes for Level-1 predictors	The same information as Model 2 is provided by fixed effect results; random slope findings show if there are differences in the associations between Level-1 predictors and the outcome between Level-2 units and Level-3 units.
Model 4	Model 3 + Level-2 fixed effects	Results with Level-2 fixed effects show how Level-2 predictors and the result are related. The remaining findings offer the same details as those given for Model 3

Table 3. Model building process for 3-level linear models.

Model 1	No predictors, just random effect for the intercept	The output used to compute the ICC reveals how much difference there is between Level-2 and Level-3 units in the result.
Model 2	Model1 + Level-1 fixed effects	The association between Level-1 predictors and the outcome is shown by the results.
Model 3	Model 2 + random slopes for Level-1 predictors	The same information as Model 2 is provided by fixed effect results; random slope findings show if there are differences in the associations between Level-1 predictors and the outcome between Level-2 units and Level-3 units.
Model 4	Model 3 + Level-2 fixed effects	Results with Level-2 fixed effects show how Level-2 predictors and the result are related. The information is the same for the remaining findings.

Continued

Model 5	Model 4 + random slopes for Level-2 predictors	The Model 4 findings are still relevant. If there are different correlations between Level-2 predictors and the outcome, Level-2 variable random slope findings show this.
Model 6	Model 5 + Level-3 fixed effects	Results with Level 3 fixed effects show how Level 3 predictors and the result are related. The same kind of information is provided by all the other Level-1 and Level-2 outcomes.

We'll use one as an illustration. A group of academics from New Zealand conducted a study on 700 cats from 200 houses in the early 2000s (*i.e.*, an average of 3.5 cats per household [23].

The researchers separated the impacts of Level-1 cat characteristics, such as whether the cat has long legs, from Level-2 household variables, such as if there is a dog in the home, in order to predict cat obesity.

The researchers considered cats to be Level-1 units that were nested in homes, which were Level-2 units. They found that when short-legged cats reside in homes without dogs, they frequently weigh more.

One method for representing data using the "MODEL" object is regression. The extent to which a model fails to properly describe the data is referred to as "RESIDUALS".

Clearly, we are just social scientists, and we can only expect our models to explain so much of the actual world. To put it another way, our models can never be totally accurate.

$$\text{DATA} = \text{MODEL} + \text{RESIDUALS} \tag{3.20}$$

$$Y_i = B_0 + e_i \tag{3.21}$$

Data can be described using the simplest regression equation possible, which is a regression with no predictor and the mean as the constant.

Depending on the nature of the independence issue, using classic regression while ignoring this issue will almost surely lead to biased standard errors and false-positive or false-negative conclusions.

Must consequently utilize two-level linear regression in this circumstance. The objective of two-level regression, like traditional regression, is to characterize data using an object called "MODEL." The extent to which such a model fails to accurately represent the data is referred to, similarly to traditional regression, as "RESIDUALS."

There are, however, two distinct residual types this time around:

1) "Level-2 RESIDUALS" refers to the degree to which the model fails to accurately capture between-cluster variances.

2) "Level-1 RESIDUALS" are the degree to which the model falls short of accurately capturing within-cluster variability.

$$\text{DATA} = \text{MODEL} + \text{LEVEL-2 RESIDUALS} + \text{LEVEL-1 RESIDUALS} \tag{3.22}$$

$$Y_{ij} = B_{00} + u_{0j} + e_{ij} \tag{3.23}$$

It's important to note that all equations for two-level regression follow the same structure. A regression with no predictor and the overall mean as the constant is the most straightforward two-level linear regression equation. Both conventional regression and multilevel modeling employ two distinct estimation techniques.

Multilevel modelling typically uses the maximum likelihood (ML) estimator (the coefficients and variance terms are jointly estimated by maximising the likelihood of the predicted values given the data).

3.19. To Estimate the (Co)variance Terms, Intermediate Models Are Built

Construct the two intermediate models to assess if it is necessary to predict this anticipated variation:

1) A restricted intermediate model that ignores the Level-1 effect's between-cluster variation

2) An enhanced intermediate model that takes into account this variance. After that, we must contrast the two intermediate models [24].

3.20. Constrained Intermediate Model

The cross-level interactions are not included in this model. The objective is to estimate the crude slope residuals and they are likely to account for some of the residual variation

$$Y_{ij} = B_{00} + B_{10} \times x_{ij}^{cmc} + B_{01} \times X_j + u_{0j} + e_{ij} \tag{3.24}$$

In the above constrained intermediate model equation

1) Coefficient estimate B_{10} (the fixed slope) corresponds to the overall effect of Level-1 predictor;

2) x_{ij}^{cmc} (cluster-mean cantered hotness);

3) Coefficient estimate B_{01} corresponds to the effect of Level-2 predictor;

4) X_j (period of success).

3.21. Augmented Intermediate Model

Then, the two intermediate models will be compared (Aguinis *et al.*, 2013).

$$Y_{ij} = B_{00} + \left(B_{10} + u_{1j}\right) \times x_{ij}^{cmc} + B_{01} \times X_j + u_{0j} + e_{ij} \tag{3.25}$$

3.22. Data of Student's Result

It is expected that universities will generate highly qualified graduates who achieve all requirements. From the time that students enrolled in the university until the time that they graduate, a variety of factors influence how well they perform academically. The most common measurement used to assess students' academic success is their cumulative grade point average (CGPA).

During the study time period, many factors have their influence on the performance of student's performance on their cumulative grade point average (CGPA) such as intelligence level, hours of study, teacher's conveying method.

Many researches have been done to check the influence the influence of different factors on cumulative grade point average (CGPA) just like [25] has examined the current review's objective, which was to determine the impact of online recordings on students' difficult results. In this study, straight blended impact models fitted to the data were used to differentiate the effects of video access from the effects of many other factors, such as orientation, academic year, course section, and students' aggregate grade point average (CGPA).

They actually made a very big admission and came to the conclusion that online recordings often work best for students with lower CGPAs.

However, in this investigation, the main objective was to examine the impact of previous teaching methods on current teaching interventions. For this, CGPA was used as a response variable, while Intermediate marks and Entry test marks were utilized as explanatory variables.

A dataset with N = 406 student grades uses regression analysis. Regression may be considered a technique for describing data using the "MODEL" object. In other words, our models can never accurately reflect the facts, and the degree to which a model falls short of doing so is referred to as "RESIDUALS".

3.23. lmer

The lmer () function ("Linear Mixed Effects in R") from the lme4 package is used to execute a multilevel linear model. Independent variables at all data levels were also treated for null fit, random intercept, and the random slope model and comparative methods of model fit were optimized.

4. Results and Discussion

The University of Agriculture Faisalabad provided the data set for this study, which was collected using a stratified random sampling technique based on students' result data. This linear mixed-effect model is fitted under multilevel modeling process and used to jointly estimate the effect on CGPA and the set of other variables, including department identifier, gender, intermediate marks and entry test marks.

In this portion, the experiment is accomplished by multilevel modeling approaches for empirical investigation.

- "Inter" represents a numerical variable with values ranging from 498.0 to 1078.0.
- "Gender" is a binary variable with values 0 and 1, representing the gender of the students.
- "CGPA" represents the response variable (Cumulative Grade Point Average), with values ranging from 1.94 to 3.90.
- "Entry. Test" is a numerical variable indicating the entry test score.
- "Department" is a character variable indicating the department of the students, a total of 406 students.

4.1. Entering Variables

In **Table 4** and **Table 5**, the summary provides information about the distribution and characteristics of the CGPA, Entry Test and Inter variables. As Inter is the cluster, CGPA is the outcome obtained of eight different departments that are dependent on the entry test and inter marks of the students.

Table 4. Summary of statistics department identifier CMC.

	Department Identifier GMC	Department identifier CMC
Min.	−3.4901	−3.833
1st Qu.	−2.4901	−1.500
Median	0.0099	0.000
Mean	0.0000	0.000
3rd Qu.	1.5098	1.500
Max.	3.5099	4.000

Table 5. Summary of CGPA, entry test marks, inter marks.

	CGPA	Entry test	Inter
Minimum	1.94	37.00	498.0
1st Quartile	3.01	54.00	799.0
Median	3.36	63.00	856.0
Mean	3.31	63.59	849.7
3rd Quartile	3.60	71.75	908.0
Maximum	3.90	88.00	1078.0

The department is the cluster mean Centering of the Level 1 predictor.

Inter represents a numerical variable with values ranging from 498.0 to 1078.0. Gender is a binary variable with values 0 and 1, representing the gender of the students. CGPA represents the response variable (Cumulative Grade Point Average), with values ranging from 1.94 to 3.90. Entry test is a numerical variable indicating the entry test score.

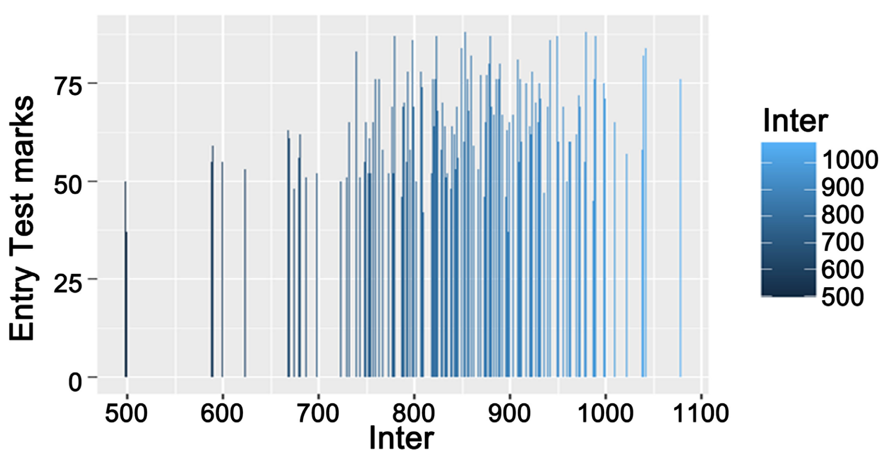

Figure 5. Bar chart of entry test marks by inter.

Figure 5 compares the Entry Test marks for each level of "Inter" using the height of the bars.

Figure 6 shows a histogram that represents the distribution of CGPA values in the dataset. Each bar in the histogram represents a range of CGPA values, and the height of the bar indicates the frequency or count of observations falling within that range.

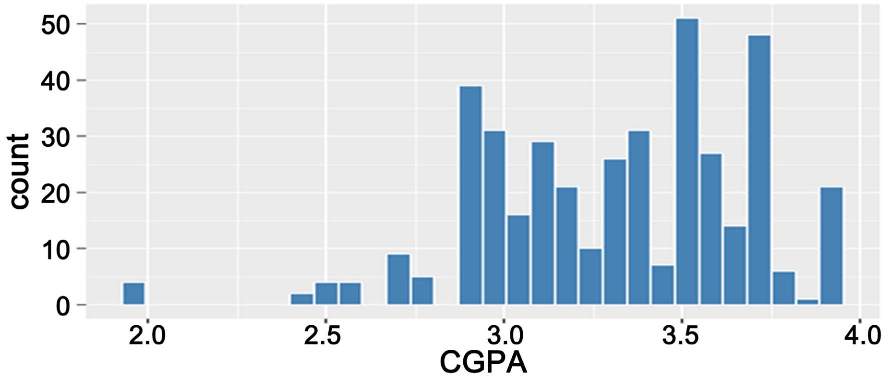

Figure 6. Histogram of CGPA.

In Figure 7, scatter plot visualizes the relationship between Entry Test Scores and CGPA. Each point represents a data point. The x-axis represents Entry Test Scores, while the y-axis represents CGPA. By examining the plot, we can observe the distribution and potential patterns or trends between Entry Test Scores and CGPA, as well as any variation based on the "Inter" variable.

Figure 7. Scatter plot of CGPA vs entry test scores.

4.2. Building an Empty Model to Calculate the ICC/DEFF

It performs a linear mixed-effects regression model (multilevel model) with "CGPA" as the dependent variable, a fixed intercept term, and a random intercept for the variable "Inter". This section presents a summary of the linear mixed model results, covering the random effects, fixed effects, and model fit statistics. Provides a summary of the linear mixed model results, including random effects, fixed effects and model fit statistics.

4.2.1. Random Effects

In Table 6, random effects component of the model estimates the variance of the random intercepts for the "Inter" variable. The estimated standard deviation of the random intercepts is 0.2773, indicating the variability in "CGPA" across different levels of "Inter".

Table 6. Summary of random effects.

Groups	Name	Variance	Std. Dev.
Inter. Marks	Intercept	0.0769	0.2773
Residuals		0.0647	0.2544

4.2.2. Fixed Effects

In Table 7, fixed intercept estimate is 3.284 with a standard error of 0.0291. This means that, on average, the expected value of "CGPA" is 3.284 when all other predictors are held constant.

Table 7. Summary of fixed effects.

	Estimate	Standard Error	T-value
Intercept	3.2838	0.0291	113

4.2.3. Model Fit Statistics

In Table 8, restricted maximum likelihood estimation (REML) criterion at convergence is 221.9. This value represents the goodness of fit of the linear mixed model to the data.

Table 8. REML criterion.

REML criterion at convergence	221.9

A lower REML criterion suggests a better fit, indicating that the model explains a larger proportion of the variation in the response variable (CGPA) based on the fixed and random effects.

In conclusion, the results indicate that the intercept term is significantly associated with the CGPA scores. The random intercepts suggest that there is variability in CGPA between different levels of the "Inter" grouping variable. The overall model fit, as indicated by the REML criterion, suggests a reasonable fit to the data.

4.2.4. ICC and Design Effect (DEFF)

In Table 9, the ICC value of adjusted and unadjusted 0.543 indicates a moderate level of clustering or dependency of CGPA scores within the levels of the "Inter" grouping variable.

Table 9. ICC and DEFF.

	Adjusted ICC	Unadjusted ICC	Design Effect
ICC	0.543	0.543	
DEFF			2.194

Approximately 54.3% of the total variability in CGPA scores can be attributed to between-group differences. The DEFF value of 2.194 implies the need to adjust the standard errors of the fixed effects estimates. These errors should be multiplied by approximately 2.194 to account for the clustering effect. This adjustment reflects the increased uncertainty in estimates caused by dependency among observations within the same group.

The DEFF value of 2.194 implies that the standard errors of the fixed effects estimates should be adjusted or multiplied by approximately 2.194 to account for the clustering effect. This adjustment accounts for the increased uncertainty in the estimates due to the dependency among observations within the same group.

Overall, these results indicate the presence of clustering in the data and highlight the need to account for the group-level variation when analyzing or designing studies.

4.3. Building Intermediate Models to Estimate (Co)variance Terms

4.3.1. Constrained Intermediate Model (CIM)

Table 10 summarizes the goodness-of-fit measures for the CIM. These measures provide information about the model's fit to the data and can be used for model comparison purposes.

Table 10. Summary of constrained intermediate model.

Model	AIC	BIC	Log Likelihood	Deviance
CIM	166.93	186.94	−78.463	156.93

In this case, the CIM has an AIC (Akaike Information Criterion) value of 166.93 and a BIC (Bayesian Information Criterion) value of 186.94. Lower AIC and BIC values indicate a better fit, so the CIM with these values suggests a relatively good fit to the data.

The log likelihood value of −78.463 and the deviance value of 156.93 represent the goodness-of-fit statistics for the CIM. These values indicate how well the model predicts the observed data. A lower deviance value suggests a better fit of the model to the data.

1) Random Effects

In Table 11, inter group, which represents the random intercept for the clusters, the estimated variance is 0.0546, indicating the variability in the outcome variable across different clusters.

The corresponding standard deviation is 0.2337. Residual row represents the residual variance, which captures the unexplained variability in the outcome variable after accounting for the fixed effects and random intercept. The estimated residual variance is 0.05881, with a corresponding standard deviation of 0.2425.

Table 11. Summary of random effects.

Group	Name	Variance	Std Dev.
Inter	(Intercept)	0.0546	0.2337
Residual		0.0588	0.2425

2) Fixed Effects

Table 12 shows fixed effects, the result of estimates, standard errors, and t-values for each predictor variable. The intercept has an estimated value of 2.5200, which represents the expected value of the outcome when all predictors are at zero.

Table 12. Summary of fixed effects.

Variable	Estimate	Std. Error	t-value
(Intercept)	2.5200	0.0979	25.734
Department Identifier CMC	−0.0015	0.0060	−0.248
Entry Test	0.0122	0.0015	8.088

The department identifier CMC variable has an estimated coefficient of −0.001493, indicating a negligible effect on the outcome. The Entry test variable has an estimated coefficient of 0.012180, indicating a statistically significant positive effect on the outcome.

Table 13 are showing Constrained Intermediate Model suggests that the Entry test variable has a statistically significant impact on CGPA, while the department identifier CMC variable does not show a significant relationship.

Table 13. Confidence intervals for CIM.

Variable	2.5%	97.5%
(Intercept)	2.3270	2.7136
Department Identifier CMC	−0.0133	0.0103
Entry Test	0.0092	0.0152

4.3.2. Augmented Intermediate Model (AIM)

Table 14 shows that AIM model has an AIC value of 170.7, indicating the model's goodness of fit. Lower AIC values indicate better-fitting models.

Table 14. Summary of augmented intermediate model.

AIC	BIC	Log Likelihood	Deviance
170.7	198.8	−78.4	156.7

The BIC value for the AIM model is 198.8. Similar to AIC, lower BIC values indicate better model fit, while considering the complexity of the model. The log-likelihood of the AIM model is −78.4. The deviance of the AIM model is 156.7, which represents the measure of model fit.

1) Random Effects

In Table 15, group inter estimated standard deviation for this random effect is 0.2337, indicating the average amount of variability in the intercept between different groups.

Table 15. Summary of random effect.

Groups	Name	Variance	Std. Dev.
Inter. Marks	Intercept	5.465	0.2337
Department Identifier CMC		9.909	0.0031
Residuals		5.876	0.2424

The correlation is not applicable in this case. Department identifier CMC predictor variable across different groups. The estimated standard deviation for this random effect is 0.0031, indicating the average amount of variability in the effect of "department identifier CMC" between different groups.

The correlation between this random effect and the intercept is 1.00, suggesting a positive correlation between the intercept and the effect of "department identifier CMC" across groups.

2) Fixed Effect

In Table 16, estimated intercept is 2.516269. This represents the expected average value of the response variable (CGPA) when all the predictor variables are held at zero.

Table 16. Summary of fixed effects.

	Estimate	Std. Error	t value
Intercept	2.5163	0.0979	25.706
Department Identifier CMC	−0.0018	0.0060	−0.304
Entry Test	0.0122	0.0015	8.133

The coefficient suggests that there is a small negative effect of "Department Identifier CMC "on CGPA, although the effect size is quite small. There is a positive effect of "Entry test" on CGPA, indicating that higher scores on the entry test are associated with higher CGPA values (Table 17).

Table 17. Confidence intervals for AIM.

Fixed Effect	2.5%	97.5%
(Intercept)	2.322656506	2.71054336
Department Identifier CMC	−0.013810866	0.01010980
Entry test	0.009250244	0.01522480

4.3.3. Comparing the Constrained and Augmented Intermediate Model

Table 18 represents the final result. The Chi-squared test compares the deviances of the CIM and AIM models. In this case, the p-value is 0.9118, which is greater than the commonly used significance level of 0.05. Therefore, there is insufficient evidence to reject the null hypothesis that the CIM model fits the data significantly better than the AIM model.

Table 18. Summary of model.

Model	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
CIM	5	168.10	188.13	−79.051	158.10			
AIM	7	171.92	199.96	−78.958	157.92	0.1848	2	0.9118

In Figure 8, the tile plot visualizes the relationship between Entry Test scores (x-axis), Inter values (y-axis), and the predicted CGPA values. Each tile represents a combination of an Entry Test score and an Inter value.

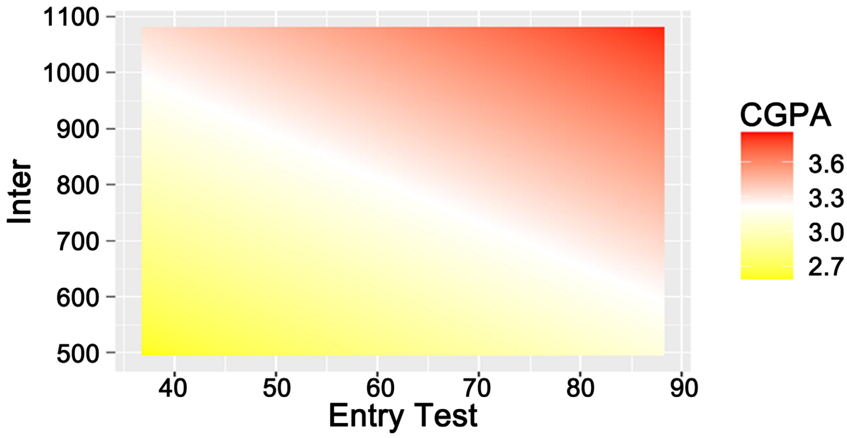

Figure 8. Predicted CGPA value.

This plot helps in understanding the predicted CGPA values based on different combinations of Entry Test scores and Inter values.

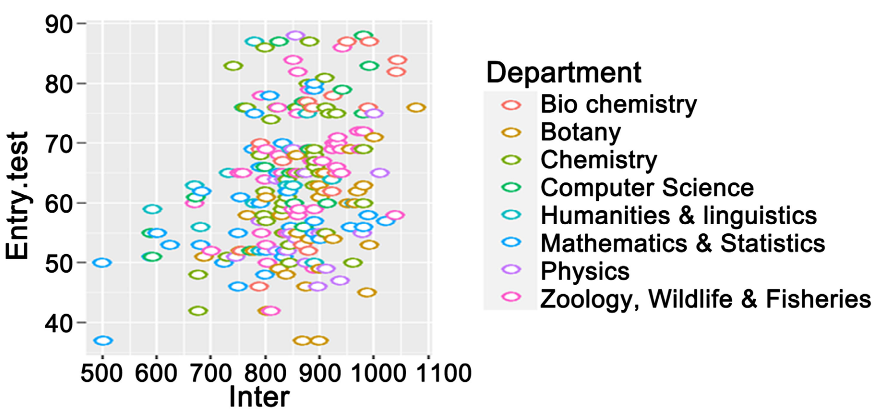

Figure 9. Observed CGPA values.

In **Figure 9**, plot shows the observed CGPA values corresponding to different combinations of Inter and Entry test scores. Each point represents a data point in the dataset.

In **Figure 10**, the graph displays the relationship between the Entry Test Score and CGPA (Cumulative Grade Point Average) based on different departments. Each department is represented by a different color.

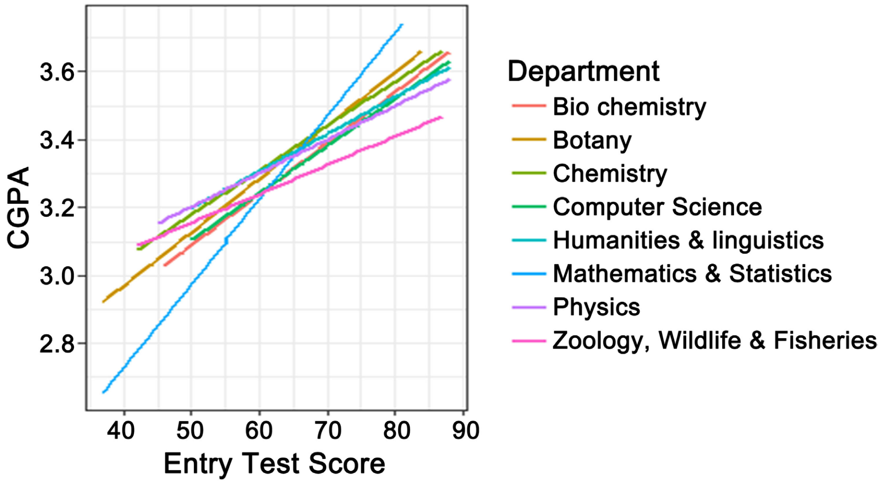

Figure 10. Relationship between entry test score and CGPA by department.

The graph shows a scatter plot of the data points, and a linear regression line is fitted for each department separately.

The x-axis represents the Entry Test Score, which is a measure of the students' performance on a standardized test. The y-axis represents the CGPA, which is a measure of academic performance.

Based on this graph, we conclude that there is a positive relationship between the Entry Test Score and CGPA across departments, indicating that higher scores on the Entry Test are associated with higher CGPA.

However, the strength of this relationship and the extent to which it varies across departments can be further examined by considering additional statistical analysis and interpreting the coefficients of the linear regression models.

5. Discussion

The study described the use of multilevel modeling (MLM) in educational research by Peugh (2010). It describes the seven key steps for performing a multilevel analysis: defining the research question, choosing the suitable parameter estimator, determining the necessity of MLM, constructing Level-1 and Level-2 models, offering multilevel effect sizes, and evaluating the likelihood ratio model. The article seeks to guide applied researchers in performing and assessing multilevel analyses and offers recommendations for effectively presenting the results. The study emphasizes the potential for larger bias in MLM results compared to traditional analyses that pool data from all clusters, due to the smaller sample sizes typically present at the lowest level of any multilevel model. It highlights the trade-off between the potential advantages of using MLM and the need for results to be easily comprehensible to policymakers and practitioners.

Investigate the application of linear mixed effects models in comparing dissolution profiles, with a focus on making the methodology more understandable to individuals working in dissolution laboratories. The study uses theoretical components and real data to demonstrate the use of linear mixed effects models in this context.

They propose a simple iterative method for estimating and selecting fixed and random effects in linear mixed models. They suggest using a data-oriented penalty function and demonstrate the consistency of their method's variable selection process through simulation experiments and real data analysis.

Jongmans (2021) examines the intra-class correlation coefficient (ICC) as a measure of differences within groups in hierarchical data structures. The study investigates the ICC Bayes factor test for various sample sizes and concludes that the ICC can effectively communicate intra-group differences and provide further context for mean scores in nested data.

In the study mentioned, the main goal was to evaluate different models within the linear mixed effects modeling structure using data collected from the University of Agriculture, Faisalabad.

The dataset included 406 observations on student's result data, and the lmer package in R (a statistical programming language) was used to estimate the Level 1 and Level 2 regression equations for the linear mixed effects model. The variables considered in the models included CGPA, intermediate marks, entry test marks, department identifier, and gender.

Overall, these studies contribute to the understanding and application of multilevel modeling and linear mixed effects models in educational research, providing insights into their methodology, interpretation, and potential challenges.

The LR test is performed using the ANOVA function to compare the deviance of the CIM and AIM models. The result of the LR test shows that the chi-square test statistic is 0.1848 with 2 degrees of freedom, resulting in a p-value of 0.9118. This indicates that the difference in deviance between the two models is not statistically significant.

6. Summary

This study demonstrates the importance of multilevel modeling (MLM) in behavioural sciences, highlighting its capability to analyze hierarchical data where traditional regression may fall short. Using linear mixed-effects models, an extension of MLM, this research explores how MLM accounts for clustered, cross-classified data structures that ordinary least squares (OLS) regression cannot fully address due to assumptions of data independence.

MLM is particularly valuable when outcome variables are clustered by categorical factors, and overlooking these clusters can lead to misinterpretations. A null model, which includes clustering variables modifying the intercept of the dependent variable, confirms the necessity of MLM by testing clustering significance. Multilevel models allow for complex analysis, accommodating nested or cross-classified structures and specifying different covariance structures across hierarchical levels.

This research utilized data from the University of Agriculture, Faisalabad, comprising 406 student test results gathered via stratified random sampling. Students were nested within classes, which were further nested within departments. Using the R lmer package, Level 1 and Level 2 equations in the linear mixed-effects model assessed variables such as CGPA, intermediate marks, entry test marks, department identifier, and gender.

The constrained intermediate model (CIM) and augmented intermediate model (AIM) both indicate that the department identifier has no significant effect on CGPA, while entry test marks positively correlate with CGPA. Deviance values between CIM and AIM were similar, suggesting the augmented model did not offer a significantly better fit. Results reveal that entry test scores significantly predict CGPA, but department identifiers show no statistical significance, underscoring MLM's effectiveness in capturing meaningful patterns in educational data.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Kreft, I. and de Leeuw, J. (1998) Introducing Multilevel Modeling. SAGE Publications Ltd. https://doi.org/10.4135/9781849209366

[2] Paccagnella, O. (2006) Centering or Not Centering in Multilevel Models? The Role of the Group Mean and the Assessment of Group Effects. *Evaluation Review*, **30**, 66-85. https://doi.org/10.1177/0193841x05275649

[3] Looney, M.A. (2000) When Is the Intraclass Correlation Coefficient Misleading? *Measurement in Physical Education and Exercise Science*, **4**, 73-78. https://doi.org/10.1207/s15327841mpee0402_3

[4] Hoffman, L. and Rovine, M.J. (2007) Multilevel Models for the Experimental Psychologist: Foundations and Illustrative Examples. *Behavior Research Methods*, **39**, 101-117. https://doi.org/10.3758/bf03192848

[5] Jung, T. and Wickrama, K.A.S. (2007) An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Social and Personality Psychology Compass*, **2**, 302-317. https://doi.org/10.1111/j.1751-9004.2007.00054.x

[6] Curran, P.J. and Bauer, D.J. (2011) The Disaggregation of Within-Person and Between-Person Effects in Longitudinal Models of Change. *Annual Review of Psychology*, **62**, 583-619. https://doi.org/10.1146/annurev.psych.093008.100356

[7] Musca, S.C., Kamiejski, R., Nugier, A., Méot, A., Er-Rafiy, A. and Brauer, M. (2011) Data with Hierarchical Structure: Impact of Intraclass Correlation and Sample Size on Type-I Error. *Frontiers in Psychology*, **2**, Article 74. https://doi.org/10.3389/fpsyg.2011.00074

[8] Mathieu, J.E., Aguinis, H., Culpepper, S.A. and Chen, G. (2012) Understanding and

Estimating the Power to Detect Cross-Level Interaction Effects in Multilevel Modeling. *Journal of Applied Psychology*, **97**, 951-966. https://doi.org/10.1037/a0028380

[9] Dong, N. and Maynard, R. (2013) *PowerUp!*: A Tool for Calculating Minimum Detectable Effect Sizes and Minimum Required Sample Sizes for Experimental and Quasi-Experimental Design Studies. *Journal of Research on Educational Effectiveness*, **6**, 24-67. https://doi.org/10.1080/19345747.2012.673143

[10] Nezlek, J. (2011) Multilevel Modeling for Social and Personality Psychology. SAGE Publications Ltd. https://doi.org/10.4135/9781446287996

[11] Shieh, G. (2015) Choosing the Best Index for the Average Score Intraclass Correlation Coefficient. *Behavior Research Methods*, **48**, 994-1003. https://doi.org/10.3758/s13428-015-0623-y

[12] Stapleton, L.M. and Kang, Y. (2016) Design Effects of Multilevel Estimates from National Probability Samples. *Sociological Methods & Research*, **47**, 430-457. https://doi.org/10.1177/0049124116630563

[13] Perinetti, G. (2018) StaTips Part IV: Selection, Interpretation and Reporting of the Intraclass Correlation Coefficient. *South European Journal of Orthodontics and Dentofacial Research*, **5**, 3-5. https://doi.org/10.5937/sejodr5-17434

[14] Trusov, P., Kondratev, N. and Podsedertsev, A. (2022) Description of Dynamic Recrystallization by Means of an Advanced Statistical Multilevel Model: Grain Structure Evolution Analysis. *Crystals*, **12**, Article 653. https://doi.org/10.3390/cryst12050653

[15] Frank, U. (2022) Multi-Level Modeling: Cornerstones of a Rationale: Comparative Evaluation, Integration with Programming Languages, and Dissemination Strategies. *Software and Systems Modeling*, **21**, 451-480. https://doi.org/10.1007/s10270-021-00955-1

[16] Loy, J. and Booth, D. (2004) Social Structure and Social Theory: The Intellectual Insights of Robert K. Merton. In: Giulianotti, R., Ed., *Sport and Modern Social Theorists*, Palgrave Macmillan UK, 33-47. https://doi.org/10.1057/9780230523180_3

[17] Hox, J. and Maas, C. (2004) Multilevel Structural Equation Models: The Limited Information Approach and the Multivariate Multilevel Approach. In: van Montfort, K., Oud, J. and Satorra, A., Eds., *Recent Developments on Structural Equation Models*, Springer, 135-149. https://doi.org/10.1007/978-1-4020-1958-6_8

[18] Koo, T.K. and Li, M.Y. (2016) A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, **15**, 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

[19] Rindskopf, D. (2010) Hierarchical Linear Models. In: Peterson, P., Baker, E. and McGaw, B., Eds. *International Encyclopedia of Education* (*Third Edition*), Elsevier, 210-215. https://doi.org/10.1016/b978-0-08-044894-7.01336-1

[20] Wicker, P., Hallmann, K. and Breuer, C. (2013) Analyzing the Impact of Sport Infrastructure on Sport Participation Using Geo-Coded Data: Evidence from Multi-Level Models. *Sport Management Review*, **16**, 54-67. https://doi.org/10.1016/j.smr.2012.05.001

[21] Bosker, R.J., Snijders, T.A.B. and Guldemond, H. (1999) PINT (Power IN Two-Level Designs): Estimating Standard Errors of Regression Coefficients in Hierarchical Linear Models for Power Calculations. User's Manual. RU Groningen/UT-NWO Project 500-278-203. https://research.utwente.nl/en/publications/pint-power-in-two-level-designs-estimating-standard-errors-of-reg

[22] Kish, L. (1965) Sampling Organizations and Groups of Unequal Sizes. *American Sociological Review*, **30**, 564*572. https://doi.org/10.2307/2091346

[23] Allan, F.J., Pfeiffer, D.U., Jones, B.R., Esslemont, D.H.B. and Wiseman, M.S. (2000) A Cross-Sectional Study of Risk Factors for Obesity in Cats in New Zealand. *Preventive Veterinary Medicine*, **46**, 183-196. https://doi.org/10.1016/s0167-5877(00)00147-1

[24] Aguinis, H., Gottfredson, R.K. and Culpepper, S.A. (2013) Best-practice Recommendations for Estimating Cross-Level Interaction Effects Using Multilevel Modeling. *Journal of Management*, **39**, 1490-1528. https://doi.org/10.1177/0149206313478188

[25] Dupuis, J., Coutu, J. and Laneuville, O. (2013) Application of Linear Mixed-Effect Models for the Analysis of Exam Scores: Online Video Associated with Higher Scores for Undergraduate Students with Lower Grades. *Computers & Education*, **66**, 64-73. https://doi.org/10.1016/j.compedu.2013.02.011