National Cheng Kung University

Institute of International Business

GENERALIZED LINEAR MODELS

Final Report - Spring 2022

Multilevel Linear Regression

Professor: Martin Tshishimbi Wa Lukusa

Student: Celia – R66107043

1.INTRODUCTION

The nature of the data is one of the primary assumptions in which observations are not correlated. However, in practice, it has witnessed many cases where the data structure is complex or hierarchical. For instance, a researcher wants to investigate factors that impact on GPA of students through six consecutive semesters. Hence, the six measurement occasions are represented by separate variables, or times of the measurements vary from subject to subject; however, GPA in the previous semester could be correlated with GPA in the following semester. Similarly, a human resource manager wants to examine employees' job satisfaction. Employees are nested in departments within a company. People within a particular department may share certain properties, including socialization patterns, traditions, attitudes, and goals. It would be reasonable to assume that the company department itself would impact the staff's satisfaction, resulting from the interactions between individuals and their groups. Otherwise, requiring the regression coefficients in all departments to be the same was generally seen as much too restrictive because there were many reasons why regressions within the department could be different. Various criteria could evaluate the job satisfaction of employees. A particular criterion is relatively important in some departments but less important in others. Intuitionally, if we assume the regression coefficient to be constant, the result could be biased and not meaningful.

Given that context, statisticians developed a mixed model, which contains both fixed and random effects. The term mixed model implies the existence of data in which individual observations on an outcome are distributed (or vary at random) across identifiable groups. Thus, the mixed model is differentiated from standard regression analysis through its capability of examining correlated data and unequal variances. There are generally three types of correlated data that the mixed model resolves: Multilevel data, Longitudinal data, and Clustered data. Multilevel data is data nested in groups within a data hierarchy (such as pupils within schools, individuals within families, companies within countries, etc.). Longitudinal data, often called repeated measurements in medicine and panel data in the social sciences, arise when units provide responses on multiple occasions. Longitudinal data provide clusters of observations on the same individual; meanwhile, cluster sample designs provide groups of observations within clusters.

The first example is an illustration of longitudinal data. Such data can be considered clustered or two-level data with occasions i at level 1 and units j at level 2. One feature distinguishing longitudinal data from other types of clustered data is the chronological ordering of the responses, implying that level-1 units cannot be viewed as exchangeable. Another feature of longitudinal data is that they often consist of many small clusters. The clustered data, where the observations are grouped, generally do not account for the chronological ordering or hierarchy of the responses. For example, when a pregnant woman comes to the hospital for a health check, the data on the mother would be clustered by one group, while the data on the baby would be clustered by the other.

Multilevel data sets are distinguished from single-level data sets by the nesting of individual observations within higher-level groups. In single-level data sets, participants are typically selected through simple random sampling. Each individual is assumed to have an equal chance of selection, and, at least in theory, the participants do not belong to any groups that might influence their responses. The simplest form of multilevel data is two-level data. In two-level data, a set of i observations (i referring to the total number of observations) at level 1 are nested within j participants at level 2. This would mean that there would be a maximum of i * j responses to analyze. The multilevel model anticipates both differences between the higher-level units and correlations within those units. It concerns components of variation of residuals at different levels. The correlations within units are expected because their members are assumed to be affected by the same aggregate effects, such as the job satisfaction of employees nested in the same department is likely to be affected by such department. The Generalized Linear Model is an expansion of the General Linear Model. In other words. General Linear Model is a special case where the link function is the identical link. Combining Multilevel Model and Generalized Linear Model, we can have Multilevel Logistic Model and Multilevel Linear Model. This report focuses on the Multilevel Linear Model in which the response variable is continuous and hierarchical, and the relationship between the response variable and independent variables is linear in parameters.

2.METHODOLOGY

Multilevel procedures also facilitate the investigation of variability in regression coef- ficients (slopes) across higher-order units in the study. Cross-level interaction involves the effects of explanatory variables at a higher level of the data hierarchy on a relationship at a lower level.

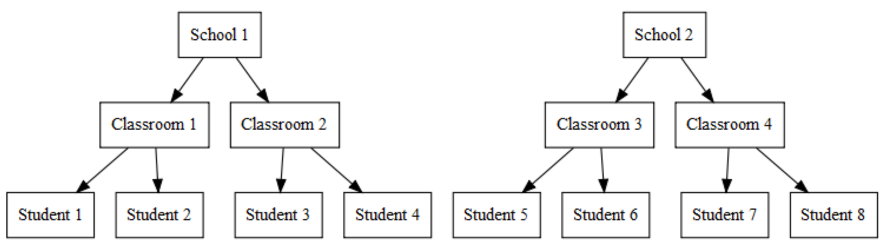


Figure 1: Illustration of 3-level Data

Null model predicting the outcome from only an intercept that is allowed to vary randomly for each group. Consider the sample has n observations and p groups. The null model for individual i in organization j can be represented as:

$$Y_{ij} = \beta_{0j} + \varepsilon_{ij} \qquad (1)$$

Where,

β_{0j}: is the mean of productivity (intercept) for the j^{th} group,

ε_{ij}: the errors in estimating individual productivity within groups

Y_{ij}: outcome variable of individual i in organization j

i is from 1 to n (n is the number of observations), j is from 1 to p (p is the number of groups).

The subscript j indicates that the intercept represents the mean outcome for groups. Individual-level error is also considered a random component.

Between groups, variation in intercepts (β_{0j}) can be represented as:

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{2}$$

In this framework, γ_{00} represents an average or general intercept value that holds across clusters (Level 2 fixed effect), whereas u_{0j} is a group-specific effect on the intercept (Level 2 random effect). We can think of γ_{00} as a fixed effect because it remains constant across all clusters, and u_{0j} is a random effect because it varies from cluster to cluster.

Through substituting Equation 2 into Equation (1) , we arrive at the single-equation model, which can be written as

$$Y_{ij} = \gamma_{00} + u_{0j} + \varepsilon_{ij} \tag{3}$$

As group were randomly selected and the group effects are not of primary interest, the terms u_{0j} can be defined as independent, identically distributed (i.i.d) random variables with $u_{0j} \sim N(0, \sigma_u^2)$.

Similarly, the terms ε_{ij} are (i.i.d) random variables $\varepsilon_{ij} \sim N(0, \sigma_e^2)$ and u_{0j} and ε_{ij} are independent.

In this case,

$E(Y_{ij}) = \gamma_{00}$,

$$Var(Y_{ij}) = E[(Y_{ij} - \gamma_{00})^2] = E[(u_{0j} + \varepsilon_{ij})^2] = (\sigma_u^2 + \sigma_e^2)$$

where σ_u^2 denotes population variance between clusters and σ_e^2 indicates population variance within clusters.

For individual in the same organization (group),

$$cov(Y_{kj}, Y_{mj}) = E[((u_{0j} + \varepsilon_{kj})(u_{0j} + \varepsilon_{mj})] = \sigma_u^2$$

$k \neq m$; k and m are from 1 to n; j is from 1 to p

and for individual in different group,

$$cov(Y_{ik}, Y_{lm}) = E[((u_{0k} + \varepsilon_{ik})(u_{0m} + \varepsilon_{lm})] = 0$$

$i \neq l$; $k \neq m$; i and l are from 1 to n; k and m are from 1 to p

If y_j is the vector of responses for individuals in group j then the variance - covariance matrix for y_j is

$$V_j = \begin{pmatrix} \sigma_u^2 + \sigma_e^2 & \sigma_u^2 & \dots\dots\dots & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 + \sigma_e^2 & \dots\dots\dots & \sigma_u^2 \\ \dots & \dots & \dots & \dots \\ \sigma_u^2 & \sigma_u^2 & \dots\dots\dots & \sigma_u^2 + \sigma_e^2 \end{pmatrix}$$

Then,

$$V_j = (\sigma_u^2 + \sigma_e^2) \begin{pmatrix} 1 & \rho & \dots\dots & \rho \\ \rho & 1 & \dots\dots & \rho \\ \dots & \dots & \dots & \dots \\ \rho & \rho & \dots\dots & 1 \end{pmatrix}$$

Where,

$$\rho = \sigma_u^2 \; / \; (\sigma_u^2 + \sigma_e^2)$$

is the intra-class correlation (ICC) coefficient which describes the proportion of the total variance due to within-cluster variance. Higher values of ρ indicate that a greater share of the total variation in the outcome measure is associated with cluster membership. When the responses within a cluster are much more alike than responses from different clusters, then σ_e^2 is much smaller than σ_u^2

Now, if we add a single predictor (x_{ij}) at the individual level (Level 1) to the model, we obtain:

$$Y_{ij} = \gamma_{00} + \gamma_{10}\, x_{ij} + u_{0j} + \varepsilon_{ij} \tag{4}$$

Where,

γ_{00}: an average or general intercept value that holds across clusters (Level 2 fixed effect)

γ_{10}: the average relationship of x with y across clusters

u_{0j}: a group-specific effect on the intercept (Level 2 random effect)

ε_{ij}: the errors in estimating individual productivity within groups

Y_{ij}: outcome variable of individual i in organization j

i is from 1 to n (n is the number of observations), j is from 1 to p (p is the number of groups)

This model can also be expressed in two separate levels:

Level 1: $Y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + u_{0j}$

$\beta_{1j} = \gamma_{10}$

β_{0j}: intercept; β_{1j}: the impact of x on Y

The model now includes the predictor and the slope relating it to the dependent variable γ_{10}, which we acknowledge as being at Level 1 by the subscript 10. We interpret γ_{10} in the same way as β_1 in the linear regression model, i.e., as a measure of the impact on y of a one-unit change in x. In this model, both γ_{10} and γ_{00} are fixed effects, while σ_u^2 and σ_e^2 remain random.

One implication of the model in Equation (4) is that the dependent variable is impacted by variations among individuals (σ_e^2), variations among clusters (σ_u^2), an overall mean common to all clusters (γ_{00}), and the impact of the independent variable as measured by γ_{10}, which is also common to all clusters.

In practice, however, there is no reason that the impact of x on y must be common for all clusters. In other words, it is entirely possible that rather than having a single γ_{10} common to all clusters, there is actually a unique effect for the cluster of $\gamma_{10} + u_{1j}$, where γ_{10} is the average relationship of x with y across clusters, and u_{1j} is the cluster-specific variation of the relationship between the two variables.

Level 1: $Y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + u_{0j}$

$\beta_{1j} = \gamma_{10} + u_{1j}$

β_{0j}: intercept; β_{1j}: the impact of x on Y

i is from 1 to n; j is from 1 to p

The combined model (random slopes model) is

$$Y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + u_{0j} + u_{1j} x_{ij} + \varepsilon_{ij} \quad (5)$$

Written in this way, we have separated the model into its fixed ($\gamma_{00} + \gamma_{10} x_{ij}$) and random ($u_{0j} + u_{1j} x_{ij} + \varepsilon_{ij}$) components. The Equation (5) model simply indicates an interaction between cluster and x, such that the relationship of x and y is not constant across clusters.

Parameters that are proposed to vary randomly across units are referred to as random effects or random coefficients from various statistical perspectives. In experimental research, for example, a random effect describes a situation where the levels of a treatment are assumed to represent a sample drawn from a universe of treatments or treatment levels. Because the effect is considered as randomly varying across a universe of treatment levels, the intent is to make inferences beyond the specific treatment levels included in the study. The effects, therefore, are not assumed to be constant. In contrast, a fixed effect describes the situation where all possible treatments are present in the experiment. In this latter case, inferences can only be made about the specific treatments used. The effects are considered to be constant and measured without error because all possible cases are included.

Unlike single-level (ordinary least squares [OLS]) regression, where random errors are assumed to be independent, to be normally distributed, and to have constant variance, in multilevel models the error structures are more complex. The individual-level errors are dependent within each unit because they are common to every individual within that unit. Errors do not have constant variance because the residual components describing intercepts and slopes may also vary across units. The estimation of these unknown random parameters associated with intercepts or slopes may also depend on characteristics of the data (e.g., sample size, degree of imbalance in sample sizes of higher level units, and degree of similarity among individuals within groups), the type of analysis conducted, and the measurement scale of the dependent variables. Because the model's random parameters must be estimated with group samples containing differing numbers of individuals, iterative estimation procedures must be used to obtain efficient estimates.

3. DATA ANALYSIS

3.1. Sample and Data Collection

The empirical study was conducted on a dataset collected in 2004 by The General Statistics Office of Vietnam. Information of 2287 students (grade-12) from 131 random high schools was collected. The empirical study aims to examine relationship of Math scores in national graduation exam and other factors such as the time on studying, and characteristics of nested school – school size. School size that measured by the number of classes in schools can reflect the property of a school since a larger school with more classes likely to go along with its reputation. A hypothesis is given that the quality of education can varies across school and students enjoy different education can have different result.

The research question focuses on

+ Whether math score varies across high school

+ The relationship between math score and time on studying
+ Whether the effects of individual time on studying tend to compound at the school level to influence student math achievement
+ Whether features of schools' contexts (school-size) affect the relationship between individual student time on studying and math achievement.

The study, therefore, provide an illustration of building a two-level model to investigate (a) a randomly varying intercept (Math score level) and, subsequently, (b) a randomly varying slope (i.e., the individual time on studying–math achievement relationship).

3.2. Descriptive statistics

3.2.1 Variable description

There are totally 2287 observations and 5 variables in the data. The type and the levels/range of each variable is summarized as follows.

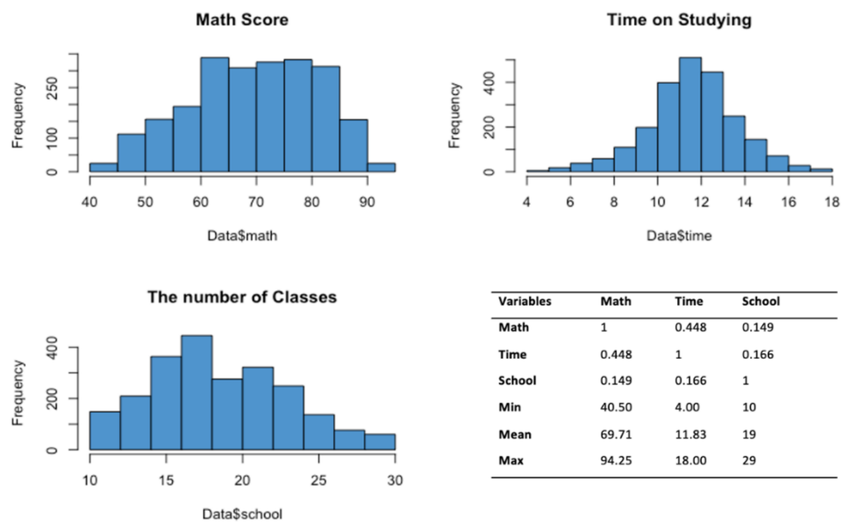
Variable	Level	Description	Type
Id	Individual – Level 1	Student identifier	Continuous
Math	Individual – Level 1	Math score of national graduation exam	Continuous
Time	Individual – Level 1	Time on study per week (Math only)	Continuous
Pro_code	School – Level 2	school identifier	
School	School – Level 2	The size of school (measured by the number of school's classes).	Continuous

Table 1: Variables description

The number of observations in each school is different with ranges from 4 to 35 observation, and an average of 17.46.

3.2.2 Explore variables

For continuous variables, important statistics such as histogram, min, mean, max are given is as follows



Variables	Math	Time	School
Math	1	0.448	0.149
Time	0.448	1	0.166
School	0.149	0.166	1
Min	40.50	4.00	10
Mean	69.71	11.83	19
Max	94.25	18.00	29

Figure 1: Histogram and statistics values

From the figure 1, we can see that time on studying seems to be normal distributed while other variables have no normal distribution which is a good signal for data analysis. Additionally, we can see that the number of classes are very different among schools.

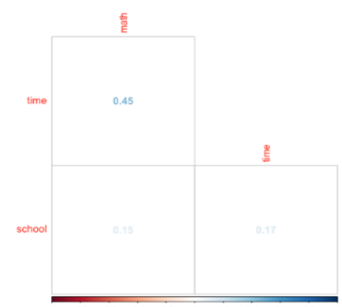


Figure 2: The correlation plot without school split

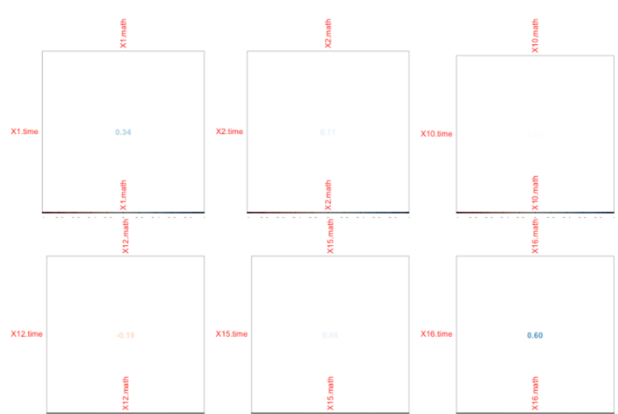


Figure 3: The correlation plots of math score and time studying for the first six school

From the figure 3, we can see that the correlation of math sore and time studying varies across schools so if we do not account for the effect of school when building model, our model can have bias results.

3.3. Results

In a single-level regression model, the impact of a variable like time on studying is assumed to be fixed across all individuals in the sample. However, the nature of data shown a hierarchical structure and the average math scores students receive may vary across schools in the sample. Given that reasons, we conduct a multi-level model. In this model, intercepts (math score) and slopes (the strength of relationship between time on study and math scores) vary across schools.

Typically, there are three distinct steps in developing the multilevel model:

Step 1: Specification of the null, or no predictors model.
Step 2: Specification of the Level 1 model
Step 3: Specification of the Level 2 model.
Now we come to the first step.

3.3.1. Variance Components Model (Null/No Predictors Model

The first step in a multilevel analysis usually is to develop a null (or no predictors) model to partition the variance in the outcome into its within- and between-groups components. This will help determine how much of the variance in math achievement lies between the schools in the sample.

The null model for student (i) in school (j) can be expressed as:

$$Y_{ij} = \beta_{0j} + \varepsilon_{ij}$$

i is from 1 to n (the number of students); j is from 1 to p (the number of schools)

where β_{0j} is the intercept and ε_{ij} represents variation in estimating individual achievement within groups. Between groups, variation in intercepts can be represented as

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

Through substitution, the null model can be written as

$$Y_{ij} = \gamma_{00} + u_{0j} + \varepsilon_{ij}$$

The null model therefore provides an estimated mean achievement score for all schools. It also provides a partitioning of the variance between Level 1 (ε_{ij}) and Level 2 (u_{0j}). Altogether there are three effects to estimate: the intercept, the between-school variation in intercepts (u_{0j}), and the variation in individual scores within schools (ε_{ij}).

		Number of levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
Random Effects	Intercept	1	Variance Components	1	Pro_code
Residual				1	

Table 2. Model Dimension

```
     AIC      BIC   logLik deviance df.resid
 17663.5  17680.7  -8828.8  17657.5     2284
```

Figure 4. Information Criteria with dependent variable is Math score

Fixed Effect			
Parameter	Estimate	Std. Error	T-Value
Intercept	69.3601	0.4172	166.2
Random Effect			
Group	Name	Variance	Std. Dev
Pro_code	Intercept	14.73	3.837
Residual		124.09	11.140

Table 3: Estimated Result with dependent variable is Math score

$$\text{ICC: } \rho = \sigma_u^2 \;/\; (\sigma_u^2 + \sigma_e^2)$$

The result reports intra-class correlation (ICC) of 0.1060825 so (~10.6%) of the variation in math scores can be attributed to the schools.

3.2 The Individual-Level (or Level 1)

3.2.1 Random Intercept Model

For each individual student i in school j, a proposed model that examines the effect of studying time on math scores) can be expressed as:

$$Y_{ij} = \beta_{0j} + \beta_{1j}Time_{ij} + \varepsilon_{ij}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{0j} = \gamma_{10}$$

Combined Model:

$$Y_{ij} = \gamma_{00} + \gamma_{10}Time_{ij} + u_{0j} + \varepsilon_{ij}$$

$\beta_{0j:}$ is the mean of math score (intercept) for the j^{th} group

β_{1j}: the impact of time on studying on math scores for the j^{th} group

i is from 1 to n (the number of students); j is from 1 to p (the number of schools)

γ_{00}: an average math scores value that holds across clusters

γ_{10}: the average relationship of time on studying with math scores across clusters

u_{0j}: a group-specific effect (school effect) on the average math score

ε_{ij}: error term

		Number of levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
	Time	1		1	
Random Effects	Intercept	1	Variance Components	1	Pro_code
Residual				1	

Table 4. Model Dimension

AIC	BIC	logLik	deviance	df.resid
17190.759	17213.699	-8591.379	17182.759	2283

Figure 5. Information Criteria with dependent variable is Math score

Fixed Effect		
Parameter	Estimate	
Intercept	69.534	
Time	2.462	
Random Effect		
Group	Name	Std.Dev
Pro_code	Intercept	2.999
Residual		10.094

Table 5: Estimated Result with dependent variable is Math score

The expected math score for a student with average time on studying averages 69.534 across all schools, but shows substantial variation from one school to another, with a standard deviation of 2.99. The common slope is estimated as a gain of 2.462 points in math score per one hour of studying.

3.2.2 Random Slope Model

For each individual student i in school j, a proposed model similar (summarizing the effect of studying time on math scores) can be expressed as:

$$Y_{ij} = \beta_{0j} + \beta_{1j}Time_{ij} + \varepsilon_{ij}$$

$$\beta_0 = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Combined Model:

$$Y_{ij} = \gamma_{00} + (\gamma_{10} + u_{1j})*Time_{ij} + u_{0j} + \varepsilon_{ij}$$

Where,

$\beta_{0j:}$ is the mean of math score (intercept) for the j^{th} group

β_{1j}: the impact of time on studying on math scores for the j^{th} group

i is from 1 to n; j is from 1 to p

γ_{00}: an average math scores value that holds across clusters

γ_{10}: the average relationship of time on studying with math scores across clusters

u_{0j}: a group-specific effect (school effect) on the average math score

u_{1j}: a group-specific effect (school effect) on the studying time - math score relationship

ε_{ij}: error term

AIC	BIC	logLik	deviance	df.resid
17191.98	17226.39	-8589.99	17179.98	2281

Figure 6. Information Criteria with dependent variable is Math score

Fixed Effect			
Parameter	Estimate		
Intercept	69.459		
Time	2.495		
Random Effect			
Group	Name	Std.Dev	Corr
Pro_code	Intercept	2.9944	
	Time	0.5356	0.13
Residual		10.0372	

Table 6: Estimated Result with dependent variable is Math score

The expected math score for a student with average studying time now averages 69.459 across schools, with a standard deviation of 2.9944. The expected gain in math score per unit (hour) of studying time 2.495, almost the same as before, with a standard deviation of 0.5356. The intercept and slope have a positive correlation of -0.13 across schools, so schools with higher math scores for a kid with average studying time tend to show bigger size.

3.3. The School-Level (or Level 2) Random Intercept Model

At this step, a school-level variable (school size) is added into the model to explain the variability in intercepts across schools. A hypothesis is proposed that school size will impact the remaining variability in achievement between schools. Additionally, the cross-level interactions, or school-level variables could moderate (enhance or diminish) the strong of the within-school time on studying–math score relationship. Because a level 2 predictor is added to the model, every individual in each unit has the same value on that variable. For example, every student in the first school will have the same value at school size.

At the school level, the model is as follows.

$$Y_{ij} = \beta_{0j} + \beta_{1j}Time_{ij} + \varepsilon_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Size + \gamma_{02}Size*Time + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Combined Model:

$$Y_{ij} = \gamma_{00} + \gamma_{01}Size + \gamma_{02}Size*Time + u_{0j} + (\gamma_{10} + u_{1j})* Time_{ij} + \varepsilon_{ij}$$

$$Y_{ij} = \underbrace{\gamma_{00} + \gamma_{10}Time_{ij} + \gamma_{01}Size + \gamma_{02}Size*Time}_{\text{Fixed Effect}} + \underbrace{u_{0j} + u_{1j}\,Time_{ij}}_{\text{Random effect}} + \varepsilon_{ij}$$

Where,

i is from 1 to n; j is from 1 to p

γ_{00}: an average math scores value that holds across clusters

γ_{10}: the average relationship of time on studying with math scores across clusters

u_{0j}: a group-specific effect (school effect) on the average math score

u_{1j}: a group-specific effect (school effect) on the studying time - math score relationship

γ_{01}: the effect of school's size on the average math score

γ_{02}: the effect of school's size on the average math score under the moderating effect of time on studying (interaction term of school's size and studying time)

ε_{ij} : error term

Fixed Effect			
Parameter	Estimate		
Intercept	69.3452		
Time	2.4497		
Size	0.2338		
Time*Size	0.0406		
Random Effect			
Group	Name	Std.Dev	Corr
Pro_code	Intercept	2.8135	
	Time	0.5329	0.07
Residual		10.0329	

Table 7: Estimated Result with dependent variable is Math score

We find that the relationship between math scores and time on studying varies substantially from school to school, depending on the school's size and unobserved factors.

For the average school size, the mean math score for a student with average time on studying is 69.3452 and increases an average of 2.4497 points per unit (hour) of studying time.

The expected score for an average student increases 0.2338 points per unit of school's size (the number of classes), but this effect has a standard deviation across schools of 2.8135 points, so there remain substantial unobserved school effects.

The average gain in math scores per unit of studying time increases 0.0406 points per unit of school's size (the number of classes). This effect has a standard deviation of 0.5329 across schools.

Finally, we note that the random intercept and slope have a positive correlation of 0.07. This means that schools that tend to show higher math scores for average student also tend to show bigger size.

AIC	BIC	logLik	deviance	df.resid
17184.983	17230.863	-8584.492	17168.983	2279

Figure 7. Information Criteria with dependent variable is Math score

4. CONCLUSION

The empirical study conducts a multilevel linear regression on a set of continuous variables. Its limitations are the number of levels and the type of variables. The further studies should concern other types when building model such as multilevel logistic model with for binary dependent variable and encounter categorical independent variables in model. This study stops at 2 levels, however, in reality, the structure of data could show higher levels. For example, a 3-level data includes level 1- patients, level 2-hopital departments, level 3-hospitals. Although this study illustrates a simple multi-level model, we can see that the data structure affects the estimated result. Hence without encountering this effect, the model likely to be biased. Otherwise, the multi-level model seems to be more flexible than ordinary single-level model since it does not need strong assumptions, however, the interpretation of estimated effects is still challenge for researchers.

REFERENCE:

Finch, W.H., Bolin, J.E. and Kelley, K., 2019. Multilevel modeling using R. Crc Press.

Snijders, T.A. and Bosker, R.J., 2011. Multilevel analysis: An introduction to basic and advanced multilevel modeling. sage.

Understanding and analyzing multilevel data from real-time monitoring studies: An easily-accessible tutorial using R

De Leeuw, J., Meijer, E. and Goldstein, H., 2008. Handbook of multilevel analysis (p. 1). New York: Springer.

Heck, R.H., Thomas, S.L. and Tabata, L.N., 2013. Multilevel and longitudinal modeling with IBM SPSS. Routledge.