Iraqi Journal of Statistical Sciences, Vol. 19, No. 2,2022 , Pp. (78-90)

Iraqi Journal of Statistical Sciences

www.stats.mosuljournals.com

Constructing a Multilevel Modeling to High-Resolution CT (HRCT) Lung in Patients with COVID-19 Infection

Didar.A.Wafa and Mohammad.M.Fage

Department of Mathematics, College of Science, University of Sulamania, Sulamania, Iraq

Article information

Article history:
Received July 4, 2022
Accepted August 10, 2022
Available online December 1, 2022

Keywords Multilevel Modeling, Fixed effect, Random Effect, Interclass correlation

Correspondence:
Didar.A.Wafa
Didar.rashid@univsul.edu.iq

Abstract

The coronavirus disease, also called COVID-19 is caused by the SARS-CoV-2 virus. Most the people contaminated with the virus will experience mild to moderate symptoms of respiratory diseases. The aim of this paper is constructing a model by multilevel modeling for these patients who sufferers by coronaviruses, we got seven hospitals which totals (636) patients in private and public that 27% from Erbil, 26% from Sulaimani, 23% from Duhok and 24% from Halabja from the period (September 1th, 2019 to February 1th, 2022). In these modelling of multilevel restricted maximum likelihood estimation (RMLE) and full maximum likelihood (FML) acclimate estimate the parameters of multilevel models (fixed and random). The application was on the HRCT lungs of patients, seven hospitals were selected randomly among the county in Kurdistan region of Iraq. The result shows that all three variables are significant at the hospital level, but in the two final models add level-2 predictor (Doctor Experience) that interaction with level-1 predictor (smoker), which is far from significant. However, there is a significant relationship between being a diabetic and having a CT scan, but the relationship between smoking and having a CT scan is not significant.

DOI: 10.33899/IQJOSS.2022.176224,

1. Introduction

Multilevel modeling is a common method used to study hierarchical data construction in various disciplines, such as education, behavioral health, and social sciences. Multilevel data analysis is a technique used to investigate data structures that cannot be adequately investigated using single-level analytical methods such as multiple regression, path analysis, and structural modeling. [8]

Various data, counting observational data collected from the human and biological sciences, often have a clustered structure, hierarchical or nested. For example, animal and human genetic studies deal with the natural hierarchy in which offspring are grouped inside a family. Descendants the same parent tends to have more similar physical and cerebral traits than randomly selected single from the general population. For example, in the same family youngster may often be small, perhaps because parents are little or because of a typical needy habitat. [4] Compared to regression as a classical form, multilevel modeling is commonly an improvement, but to varying degrees; multilevel modeling may be important for prediction, for data reduction it may be useful, and for causal inference it may be useful [1]. There are two main types of multilevel models: hierarchical and network. One category is multilevel regression models (often abbreviated as (MLM) for multilevel modeling), which are typically selected to explain variance in a single cross-sectional or longitudinal outcome with emphasis on the direct effects of specific predictors at two or more hierarchical outcome levels. Multilevel modeling is an extension of the single-level multiple regression model for calculating an existence of a hierarchical data structure. [4]

2: Literature review

Bowen (2016) discussed synthesizes existing research on HIV risk in the US and compares it to a model that takes into account the ecological factors that influence the spread of the virus. Social-ecological human development is a model that applies theories to identify factors shaping the HIV risk context for drug users who are unstable at global, societal, neighborhood, household and individual influence levels. [6].

Park et al. (2017) Examine eight factors associated with individual accidents and six additional high-level factors organized into two non-nested groups (company-level and regional-level). Also grew a single-level ordinary ordered logit model, two conventional multilevel ordered logit models, and a cross-classified multilevel ordered logit model (CCMM). The CCMM outperforms the opposing models in two main ways, first, the CCMM avoids the Type one error that contribute to happen when analyzing nested data with single-level models, and second, the CCMM can analyze two non-nested groups at the same time. Statistically significant factors are the type of vehicle ownership by the taxi companies and the size of the transport infrastructure budget of the municipalities [5].

Hair and Fávero (2019) Discuss longitudinal data of multilevel modeling and clarify the situation in which it is used. The methodology estimates of three-level repeated measures models that provide conditions for their correct explanation. And this is possible to detect the fixed with random effects on the Y variable, to recognize the variance decay of random effects at multiple levels, and to test alternative covariance construction to account for heteroscedasticity, and to compute and explain the within-class correlations of all levels of analysis. Conclusion comprehension of how hierarchical data structures and repeated measures data work allows researchers and managers to define several types of build from which MLM can be apply [3].

Prague et al. (2020) developed a multilevel model of the French COVID-19 outbreak at the regional level. Rely on a globally extended Susceptible Exposed-Infectious-Recovered (SEIR) reductionist model as a clarified representation of the average epidemic trajectory with the addition of region-specific random effects. Some French public datasets on the initial dynamics of the epidemic estimate region-specific key parameters depending on this atomistic model by SAEM (Stochastic Approximation Expectation Maximization) optimization with the Monolix software. The results shows the low immunity of the population, the power impact of the lockdown on the dynamics of epidemics, and therefore the need for further interventions in the lifting of the lockdown to prevent the scourge from breaking out again [4].

Shin et al. (2021) discussed a total of 1,796 raw data points extracted from 114 cases in 35 single case studies. Appling three-level multi-level modeling, both immediate effects and trends during the intervention phase were analyzed as moderation effects in connection with student characteristics (case level as the first level) and intervention characteristics (study level as the second level). The pair means immediate effect and the trend throughout the intervention were statistically significant. The overall effect of the prompt varied significantly by pupil grade, disability type, contriver, device, type of virtual manipulation, and the visual model implant in the virtual manipulation. Neither more student characteristics nor moderators related to disturbance characteristics significantly influenced the typical trend when using virtual manipulatives [9].

3: Methodology

3.1Modeling:

Modeling is a technique, as well as a science and, is directed toward finding a good approach model

$$Model = Pattern + Error \qquad (1)$$

Model represents a response variable then patterns are some explanatory variables displaying the behaviours of the relationship for them of response, the error also is a random variable explains the difference between the real values of the response against estimated or (predicted) value of the response. Statistical models impart powerful device to analyst in a wide array of disciplines. Such models allow for the assessment of connection among multiple variables, which can lead to a better concerning of the world [8].

3.2 Multilevel Modeling:

Multilevel modelling (MLM) also structural equation modelling (SEM) is commonly used in social and behavioral science. Multilevel Modelling is a statistical model that usually models the connection between response and explanatory data when there is a correlation between observations. Nearby, the individual observations are nested in different groups. The observations within each group are correlated [10]

3.2.1 Types of Multilevel Data:

1. Hierarchical data:

Multilevel modelling applies when data are organized hierarchically. With hierarchical data, there may be variables at each level. Only categorical variables above level 1 may be the clustering variables. Multilevel modelling will show how clustering variables and other variables at higher hierarchical levels affect the dependent variable at level 1.

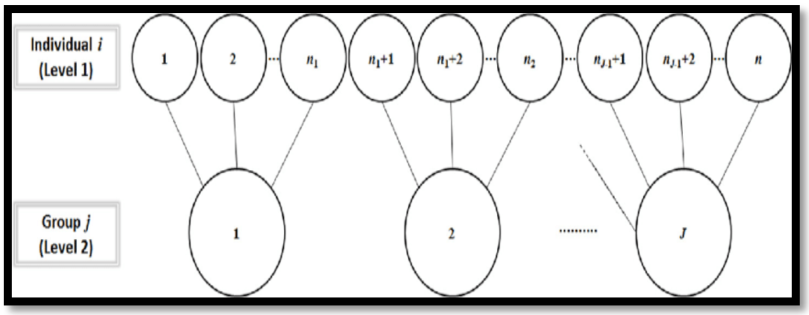

Figure3.1 Nested structure of clustered data for two-level [12]

2. Repeated measures data:

Repeated measures are conceivably seen as a special case of hierarchical data. The repeated measures go will with level 1 while the unit of analysis becomes level 2. Level 1 data not need to be repeated measures.

3. Random-effects data:

Random effects data are another special case of hierarchical data. An example is a marketing study in which consumer attitudes at level 1 are nested inside product brands at level 2, where the researcher is interested in whether the brand effect is significant.

4. Cross-classified data:

In some cases, data are not nested in a strict hierarchy. An example is where students are nested hierarchically within neighbourhoods (any student is in just one neighbourhood) but neighbourhoods are cross-classified within schools (a given neighbourhood may send students to multiple schools; a given school may draw from multiple neighbourhoods.

5. Multiple outcome data:

Multilevel modelling can also support models with more than one response variable at level-1. Again, this must be declared by the researcher who must use software supporting multivariate multilevel modeling (MMLM), in different circumstance multivariate linear mixed modeling (MLMM) or hierarchical multivariate linear modeling (HMLM).

3.2.2 Assumptions of Multilevel Modeling:

The assumption of multilevel level model has the same assumption of the general linear models as another major, but several of the assumptions are improved for the hierarchical nature of the design (hierarchical data).

1. **Linearity:** Linearity status presumption is a straight-line relationship between factors. Nevertheless, model can be extended to nonlinear relationships. In particular, if a middle part of the level 1 regression equation is restored with nonlinear parametric function, then such a model substructure is extensively referred to as a mixed effects nonlinear model [7].
2. **Normality and Multi-collinearity:** The normality assumption that the error terms are normally distributed at all model levels. Nonetheless, nearly all statistical software allows you to specify different distributions for the variance terms, such as poisson distribution, binomial and logistics. The multilevel modelling would be applied to all forms of generalized linear models [8].
3. **Homoscedasticity**: Homoscedasticity's assumption, also called as variance homogeneity, suppose that the population variances are equal [9]. Nevertheless, dissimiler variance-correlation matrices can be specified to account for this, and the variance heterogeneity can be modeled as such.

4. **Independence of observations**: Independence is an assumption of general linear models that states that instance are random samples from the population and those results on the response variable are mutually explaining [9]. One of the main aims of multilevel models is to settle cases where the assumption of independence is contravene; however, multilevel models conclude that first, the residuals at levels one and two are uncorrelated, and second, the errors (as measured by the residuals) are uncorrelated at the highest level. This assumption was checked by the Durbin-Watson test

3.2.3 Types of Multilevel Model:

There are many models possible with multilevel modeling and, unfortunately, an even larger number of labels for these models. In this part, we briefly describe the most common model types. For these types, we assume only two levels and a maximum of one predictor variable at either level (not counting the level variable), but in some research there may well be more predictor variables at either level and there may be more hierarchical levels or cross-classification of levels[2][11].

1. **The Null or Unconditional Random Intercept Model:**

In multilevel modelling, the null model is not one with just the intercept (constant) of the response variable, without any predictor variables, as in OLS regression. Instead it is the model with only the grouping (clustering, level) variable as a determinant of the intercept of the response variable "unconditional" models in that there are no other predictor variables to condition the estimates. It is a "random intercept" model since it is predicting the level 1 intercept of the outcome variable and is not predicting any (b coefficients) at lower levels (there aren't any).

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

where we will refer to β_{0j} as the mean of productivity for the jth group, and (ε_{ij}) is the residual component for individual i in organization j (i.e., ε_{ij} represents the deviation from the level-2 unit mean for individual i). The level-1 residual is supposed to have a mean of 0 and a constant variance σ^2. In a two-level model, fixed effects of level-1 are commonly communicated as unstandardized β coefficients and level-2 is not desirable. Unstandardized means the coefficients are in their original metrics.

2. **The Conditional Random Intercept Model**

The random intercept model, also called a conditional random intercept model, is random because it incorporates the random effect of the clustering variable. It is an intercept model because only the intercept of the outcome variable is adjusted for the random effect. It is conditional because predictor variables are present in addition to the clustering variable(s) which define level 2 or higher. The intercept, represents the mean of the dependent variable. The slopes (b coefficients) of any level 1 predictor variables are not modelled as random effects.

a. **Adding Level-1 Predictors to Explain Variability in Intercepts**

The level-1 predictors as X variables. For all individual i in organization j, this model summarizing the effect of explanatory variable motivation on response can be expressed as

$$\mathrm{Y_{ij}} = \beta_{0j} + \beta_{1j}\mathrm{X_{1ij}} + \varepsilon_{ij} \qquad (2)$$

Where $\mathrm{Y_{ij}}$is the observation for the ith individual in level-2 unit j, $\beta_{0j}$is the level-1 intercept within unit j, β_{1j} is a level-1 slope for predictor $\mathrm{X_{1ij}}$ and ε_{ij} is error for individual i in organization$j$ .Within all level-2 unit ε_{ij} is assumed to have a mean of (0) and constant variance over all levels of $\mathrm{X_{1ij}}$.

If sufficient variation exists within and between the level-2 units, this model can yield a dissimilar set of estimates of β_{0j} for each level-2 unit. Variation in level 1 intercepts can be described by an organization-level intercept (γ_{00}), or grand mean, and a random parameter catching variation in individual organization means (u_{0j}) from the grand mean

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad (3)$$

$$\beta_{1j} = \gamma_{10} \qquad (4)$$

Equation (4) implies variability in level-1 slopes can be reported by an organizational-level average slope coefficient (γ_{10}), or grand mean, which initially we usually specify as fixed across level-2 units.

Through substituting Equations (4) and (3) into Equation (2), and with some rearranging of fixed and variance terms, it obtains the combined model

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

b. **Adding Level-2 Predictors to Explain Variability in Intercepts**

This step is often to specify one or more group-level projection that can describe variability in the randomly varying intercepts. Assuming the level-1 model remains the same, the level-2 models would appear as follows:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_{1j} + u_{0j} \qquad (6)$$

$$\beta_{1j} = \gamma_{10} \qquad (7)$$

Where the level-1 intercept is β_{0j} in level-2 unit j; γ_{00} is the mean of the level-1 outcome, runing for the level-2 predictor, (W_j ; γ_{01}) is the slope for the level-2 variable (W_j ; u_{0j}) is the random variability for the organization (j ; β_{1j}) is the level-1 slope in level-2 unit *j* and γ_{10} is its mean value at the group level. Because it is no random effect (γ_{10}) in Equation (7), the slope coefficient is again fixed to one value for the sample. In difference to level-1 outcomes, which are based on *N* individual-level observations, the level-2 estimates specified in Equations (6) and (7) are based on *j* unit-level observations. The unite model with a level-1 predictor and level-2 predictor is as follows:

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{1ij} + \gamma_{10} W_{1j} + u_{0j} + \varepsilon_{ij} \qquad (8)$$

3. The Conditional Random Coefficients Model

The coefficients term in random coefficients model should not obscure the fact that a random coefficients model estimates the intercept (mean) and the slope (regression coefficient) at level 1. Synonyms for the random coefficients model are random coefficient model and random coefficients regression model.

The level-2 slope can be specified as randomly varying as follows

$$\beta_{1j} = \gamma_{10} + u_{1j} \qquad (9)$$

More generally, for the intercept and each level-1 slope, the implied between-unit model is

$$\beta_{qj} = \gamma_{q0} + u_{qj} \;\; \text{For } q = 0, 1, 2, \dots, Q \qquad (10)$$

Where β_{qj} represents the number (Q) of within-unit (*j*) regression parameters from the level-1 model in equation (4) and γ_{q0} is the mean value for each of the within-unit parameters. Across all level-2 units, then, each β_{qj} has a mean of 0 and some variance of distribution with. Should there exist significant variance in any level-1 coefficient between level-2 units β_{qj}. If only the intercept is randomly varying (u_{0j}), we can specify an identity covariance matrix at level-2

$$\text{Var}\left[u_{0j}\right] : (0, \sigma^2_{u0}) \qquad (11)$$

Which suggests there is simply one level-2 variance component (σ^2_{u0}), the variance in intercepts for level-2 units, whatever is assumed to be normally distributed with a mean of 0 and constant variance. Note also the covariance between u_{oj} and ϵ_{ij} assumed to be 0 in two-level models. For a given group j, it can specify an unstructured covariance matrix of random effects at level 2 to lodge the covariance between the random intercept and slope as follows:

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{u0} & \sigma_{u01} \\ \sigma_{u10} & \sigma^2_{u1} \end{bmatrix}\right) \qquad (12)$$

As Equation (12) indicates, specify the level-2 variances in the intercept and slope in the diagonal of the matrix and the covariance is the off-diagonal element, because the covariance matrix is a square, symmetrical matrix, only requires either the upper or lower covariance coefficient. If we do not include the covariance term between the intercept and slope, it will specify a diagonal covariance matrix at level 2, which has the intercept and slope variances as the respective diagonal elements and the covariance is fixed at 0:

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{u0} & 0 \\ 0 & \sigma^2_{u1} \end{bmatrix}\right) \qquad (13)$$

4. The Random Intercept Regression Model

Random Intercept Models with Level 2 Variable, Describe Individual and Mean Differences in Outcome, this model is the random intercept model with level 2 predictors or the means as outcomes regression model.

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + \epsilon_{ij} \qquad (14)$$

Where:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j}$$

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

Then

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}W_j + u_j + \epsilon_{ij} \qquad (15)$$

5. The Random Intercept Analysis of Covariance (ANCOVA) Model

A random intercept ANCOVA model is a random intercept regression model to which one or more level-1 regressors (predictor variables) have been added, the random intercept ANCOVA model combines the monolevel regression model with the random intercept regression model. Synonyms are the random intercept model with level-1 and level-2 variables or the means as an outcomes ANCOVA model

$$Y_{ij}= \gamma_{00}+\gamma_{10}\, x_{i1}+\gamma_{01}\, x_{1j}+ \gamma_{02}x_{2j} + \cdots + u_{1j}x_{ij} + u_{2j}x_{ij} + \cdots + u_{oj}+ e_{ij} \qquad (16)$$

Where

$$y_{ij}= \gamma_{00} +\gamma_{01}\, x_j+ u_{oj}+\gamma_{10}+\ u_{1j}\, x_{ij}+\gamma_{20} +\ u_{2j}\, x_{ij}+..+e_{ij}$$

6. The Random Coefficients ANCOVA Model

The Fully Specified Multilevel Model, Use the Slope at Level 1 as an Outcome at Level 2 and "Cross-Level Interaction"

Full Random Coefficient Model

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \epsilon_{ij} \qquad (17)$$

Where

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \ u_{1j}$$

Cross-Level Interaction Model

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \epsilon_{ij} \qquad (18)$$

where

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + \ u_{1j}$$

W_j Is Add on to the Model as a Predictor of β_{1j}

3.3 Interclass correlation (ICC):

Inter-class correlation is a measures of reliability scores or measures for clustered data collected or variety into groups. A related term is an interclass correlation, which is normally another name for Pearson's correlation (more statistics can be used, such as Cohen's kappa, but this is rare). Pearson's is typically used for inter-rater reliability when you only have one or two meaningful pairs from one or two raters. Such as most correlation coefficients, the ICC ranges from 0 to 1.

$$\rho = \frac{\sigma_b^2}{(\sigma_b^2+\sigma_w^2)} \qquad (19)$$

the ICC is main because it changes the error variance in one-stage regression analyses. while clusters and non-trivial ICCs are present, independent errors resulting from simple random sampling are probable to violate the OLS regression assumption. [3]

4: Description and Analysis of data

4.1 Data Collection:

In this section, the accumulated data will be hand over and argued in detail. The multilevel modelling have been used to analyze the data that we used from 636 patients in Erbil, Sulaimani, Duhok and Halabja along in each county proceeds 2 Hospitals, one of them is private hospital and the other is a public hospital between (September 1th,2019 to February 1th, 2022)

Analysis with multilevel modeling algorithms would be conducted to score estimable result. The patients we collected data from were identified with coronavirus and the PCR test is positive after that the patient was radiated by CT scan to find out how much the virus had on his lungs. The objective of this search is to correctly classify the patients based on their diseases by considering 3 explanatory variables 2 of variables from level-1 and the other from level-2 variable. The captured data has been translated to a form which can be gather it and used by the JAMOVI program. Converted data has been imported into JAMOVI program to be analyzed.

4.2 Data Description:

The dataset used in this search has been collected from patients in in Erbil, Sulaimani, Duhok and Halabja, in each county precede 2 Hospitals, between September 1th,2019 to January 2th, 2022. This dataset consists of total 636 patient records.

Table: 4.1 Data Description

N0	Variable	Attribute Name	Attribute Description	Values
1	PID	Patient ID	Identity document of patient	No particular range
2	HID	Hospital ID	Identity document of hospital	Lalav=1,Zheen=2,Shahi dhemn=3 Smart=4,Azady=5Vazhe n=6,Anab=7
3	CT	HRCT	HRCT of lungs patient	Min=15 Max=100
4	Diabetic	Diabetic	Diabetic of the patient	No particular range
5	Smoker	Is Smoker?	Is the patient is smoker or not?	No = 1 Past smoker = 2 Smoker=3 Passive smoker =4 High smoker = 5
6	Dexp	Doctor experience	The experience of the doctor	Min=16, Max=23

Table: 4.2 Counties name with Hospital name in each counties and Number of the patient that take it in Each Hospital

N0	Counties name	Hospital name in each counties	Number of the patient that take it
1	Erbil	LALAV	100 PATIENT
		ZHEEN	70 PATIENT
2	Sulaimany	SHAHID HEMN	100 PATEINT
		SMART	65 PATIENT
3	Duhok	AZADY	90 patient
		VAZHIN	70 PATIENT
4	Halabja	ANAB	141 PATIENT

4.3 Building Multilevel Modeling:

After verifying the assumption of the model, we can build the form of the model that includes the explanatory variable which effected to CT s patient. It would be useful to conduct ANOVA at level-2 of the model, the result of the ANOVA showed the effect of both a patient and hospital on the response variable. The hospital ID that we cluster the data is significance as shown from table 4.3 that it is <0.001 then we can build the multilevel model. The result of ANOVA as follows:

Table: 4.3 Represent the ANOVA table

	DF	Sum Sq	Mean Sq	F value	P
PID	140	50242	358.87	0.778	0.963
HID	6	19503	3250.5	7.047	<.001
Residual	489	225545	461.23		
Total	635	293900	462.83		

4.4 Results and Discussion of Multilevel Modeling:

The analysis has been conducted with the use of Multilevel modeling with six models. JAMOVI version 2.2.5.0 is used to analyze data, that the CT dataset consist of patient from different hospitals and because each patient in each counties be in to one unique hospital, it is nested design.

1. **Intercept-only Model (Unconditional Model):**
   The unconditional mixed model statements is similar to a one-way ANOVA with u_{0j}as the overall mean and u_{0j} as the hospital effect. However, we consider u_{0j} as being a random effect (a normally distributed variable with a mean of zero) rather than a fixed factor effect as in ANOVA

Table: 4.4 Result of Intercept-only Model

Fixed Effects Parameter Estimates							
			99% Confidence Interval				
	Name	Estimate	SE	Lower	Upper	df	t	p
Intercept		34.3	2.4	28.0	40.5	6	14.2	<.001

Random Components					
		99% Confidence Interval			
Groups	Name	SD	Variance	ICC	p
Hospitals (u_{0j})	Intercept	5.97	35.6	0.08	<.001
Residual (e_{ij})		20.94	438.6		<.001

$$\mathrm{CT}_{ij} = 34.3 + 35.6 + 438.6 \qquad (20)$$

We can conclude that mean CT score among hospitals is 34.3, and that there is more variation within the hospitals (442.8) than among the different hospitals (35.3), because the p-value is <0.001 then the intercept is significant. Then the inter-class correlation is

$$\rho = \frac{\sigma^2_{u_{0j}}}{\sigma^2_{u_{0j}} + \sigma^2_{e_{ij}}} = \frac{35.6}{35.6+438.6} = 0.08 \quad ,$$

ICC = %8

1. Random Intercept with one Fixed level-1 Factor (Non-Random Slop):

This model added a fixed patient-level factor, diabetics. The mixed model looks like an ANCOVA-based hospital with the covariate diabetic, but remember we're still accounting for thatu_{oj}, to be a random effect, not a fixed effect. Thus, the estimate for γ_{10} differs from what would be established by an ANCOVA procedure.

Table: 4.5 Result of Random Intercept with one Fixed level-1 factor Model

Fixed Effects Parameter Estimates							
			99% Confidence Interval				
Name	Estimate	SE	Lower	Upper	df	t	p
Intercept	34.09	1.13	31.18	37.016	6	30.1	<.001
Diabetic	0.327	0.006	0.311	0.342	630	54.4	<.001
Random Components							
			99% Confidence Interval				
Groups	Name	SD	Variance	ICC	p		
Hospitals (u_{0j})	Intercept	2.84	8.08	0.095	<.001		
Residual (e_{ij})		8.80	77.44		<.001		

We now have an estimate to the fixed effect of Diabetic. For one and all unit increase in a patient's reported Diabetic score, there is a 0.33 increase in their CT score, the p-value showed that it is significant.

$$CT_{ij} = 34.09 + 0.33\ Diabetic_{ij} + 8.08 + 77.4 \qquad (21)$$

The ICC for this model is equal to

$$\rho = \frac{\sigma^2_{u_{0j}}}{\sigma^2_{u_{0j}} + \sigma^2_{e_{ij}}} = \frac{8.08}{8.08+77.4} = 0.1 \qquad , \qquad ICC = \%10$$

With one patient-level fixed factor, nearly one-half of the total variation in CT can be accounted for by both the hospital of the patient and the patient-level fixed factor Diabetic.

2. Random Intercept and Slop for One Level-1 Factor:

This model holds a random slope for Diabetic, which means that we allow the slope of the regression equation to vary by hospital. This model is more attitude than the previous model for the variables acceptance used since it is intuitive to assume that Diabetic varies from hospital to hospital.

Table: 4.6 Result of Random Intercept and Slope for one level-1 factor Model

Fixed Effects Parameter Estimates							
			99% Confidence Interval				
Name	Estimate	SE	Lower	Upper	df	t	p
Intercept	33.9	1.029	31.269	36.554	6	33.1	<.001
Diabetic	0.325	0.0138	0.289	0.360	630	23.5	<.001
Random Components							

99% Confidence Interval					
Groups	Name	SD	Variance	ICC	p
Hospitals (u_{0j})	Intercept	2.55	6.5	0.08	<.001
	Diabetic	0.033	0.001		<.001
Residual (e_{ij})		8.65	74.9		<.001

$$CT_{ij} = 33.9 + 0.33\ Diabetic_{ij} + 0.033 Diabetic_{ij} + 8.65 + 74.9 \qquad (22)$$

The estimate for random Diabetics slope is significant (p-value<0.001), and therefore we would say that the Patient Diabetic do vary by hospital.

The ICC for this model is

$$\rho = \frac{\sigma^2_{u_{0j}}}{\sigma^2_{u_{0j}} + \sigma^2_{e_{ij}}} = \frac{6.5}{6.5+74.9} = 0.08 \qquad , \qquad ICC = \%\ 8$$

3. Random Intercept and Random Slop for Two level-1 Factors:

For this model, we are counting a second student-level variable, Smoker, which also has a random slope u_{3j} . This means that we are accounting for the smoker of the patient and their Diabetic score, and we are allowing the effects of these factors to vary by hospital.

Table: 4.7 Result of Random Intercept and Random Slop for two level-1 factor Model

Fixed Effects Parameter Estimates							
99% Confidence Interval							
Name	Estimate	SE	Lower	Upper	df	t	p
Intercept	33.93	1.069	31.18	36.69	6	31.72	<.001
Diabetic	0.325	0.012	0.292	0.357	6	25.69	<.001
Smoker	-0.0868	0.25	-1.513	-0.223	74	-3.46	<.001

Random Components					
99% Confidence Interval					
Groups	Name	SD	Variance	ICC	p
Hospitals (u_{0j})	Intercept	2.67	7.165	0.09	<.001
	Diabetic	0.03	0.001		<.001
	Smoker	0.11	0.01		<.001
Residual (e_{ij})		8.58	73.63		<.001

$$CT_{ij} = 33.9 + 0.33\ Diabetic_{ij} - 0.087 Smoker_{ij} + 0.001 Diabetic_{ij} + 0.01 Smoker_{ij} + 7.16 + 73.63 \quad (23)$$

In the output, we see that Smoker does have a significant effect on patient's CT (p-value<0.001). The fixed estimate for Smoker, means that No Smoker (NO=1) has a CT of -0.0868, holding Diabetic constant.

The ICC for this model is equal to

$$\rho = \frac{\sigma^2_{u_{0j}}}{\sigma^2_{u_{0j}} + \sigma^2_{e_{ij}}} = \frac{7.165}{7.165+73.63} = 0.09 \quad , \quad ICC = \%\ 9$$

4. One level-2 Factor and Two Random Level-1 Factors (No Interactions):

The first model that we have seen has a level-2 (patient-level) variable: Doctor's experience in years (Dexp), in the hierarchical arrangement, you can see that it has a fixed slope coefficient γ_{01}, and is unique for all hospital j. This model does not have any interaction between Doctor's experience and the Patient-level variables. We use this model when we had reason to accept that Dexp does not mitigate the effects of smoking and diabetics on CT, meaning that the slopes for our patient-level variables are the same whether the patient is a new or new doctor may or may not have years of experience

Table: 4.8 Result of One level-2 factor and Two Random Level-1 Factors (No Interactions) Model

Fixed Effects Parameter Estimates							
			99% Confidence Interval				
Name	Estimate	SE	Lower	Upper	df	t	p
Intercept	33.9	0.52	32.46	35.16	3	64.49	<.001
Diabetic	0.327	0.012	0.296	0.358	6	27.27	<.001
Smoker	-0.088	0.25	-1.53	-0.243	244	-3.55	<.001
Dexp	0.777	0.199	0.264	1.29	17	3.6	<.001

Random Components					
		99% Confidence Interval			
Groups	Name	SD	Variance	ICC	p
Hospitals (u_{0j})	Intercept	1.044	1.09	0.02	<.001
	Diabetic	0.027	0.001		<.001
	Smoker	0.037	0.001		<.001
Residual (e_{ij})		8.6	73.7		<.001

$$CT_{ij} = 33.9 + 0.77\ Dexp_j + 0.33\ Diabetic_{ij} - 0.088 Smoker_{ij} + 0.001 Diabetic_{ij} + 0.001 Smoker_{ij} + 1.09 + 73.7 \quad (24)$$

See Dexp in the fixed effect table, with an estimate of 0.777 and a significant p-value (<0.001). This means that, holding the patient's diabetic and smoker score constant, for every additional year's experience the doctor has, that patient's CT score increases by 0.777.

The ICC for this model is equal to

$$\rho = \frac{\sigma^2_{u_{0j}}}{\sigma^2_{u_{0j}} + \sigma^2_{e_{ij}}} = \frac{1.1}{1.1+73.7} = 0.02 \quad \text{then} \quad ICC = \%\ 2$$

Notice that the ICC for this model has decreased from the previous model (ρ = 0.09)

5. One level-2 factor and Two Random Level-1 Factors with Interaction:

In the hierarchical layout, Dexp has a slope coefficient within the three β equations. This connected to the interaction terms in the mixed model for doctor's experience by diabetic, well as doctor's experience by Smoker.

Table: 4.9 Result of One level-2 factor and Two Random Level-1 Factors with Interaction Model

Fixed Effects Parameter Estimates							
			99% Confidence Interval				
Name	Estimate	SE	Lower	Upper	df	t	p
Intercept	33.72	0.477	32.49	34.95	6	28.157	<.001
Diabetic	0.322	0.008	0.3	0.347	5.39	37.45	<.001

Smoker	-0.85	0.25	-1.5	-0.20	66.15	-3.37	<.001
Dexp	1.065	0.22	0.48	1.64	7.20	4.75	<.001
Diabetic* Dexp	0.01	0.003	-0.014	0.002	6.65	-2.02	0.03
Smoker* Dexp	-0.059	0.126	-0.386	0.267	104.64	-0.469	0.64

Random Components					
		99% Confidence Interval			
Groups	Name	SD	Variance	ICC	p
Hospitals (u_{0j})	Intercept	0.86	0.74	0.02	<.001
Diabetic		0.024	0.002		
Smoker		0.059	0.003		
Residual (e_{ij})		8.58	73.7		<.001

$$\text{CT}_{ij}=33.7+1.065\ \text{Dexp}_j+0.32\ \text{Diabetic}_{ij}-0.85\ Smoker_{ij}+0.01\ \text{Dexp}_j*\ \text{Diabetic}_{ij}+0.002\text{Diabetic}_{ij}+0.003\ \text{Smoker}_{ij}+0.74+73.7 \qquad (25)$$

In the fixed effects table, there are two interaction terms, one of which (γ_{21})is far from significant, with a p-value of > 0.001. However, γ_{11} is significant, meaning that Doctor's experience moderates the relationship between Diabetic and CT, but not the relationship between Smoker and CT**.**

The ICC for this model is equal to

$$\rho = \frac{\sigma^2_{u_{0j}}}{\sigma^2_{u_{0j}}+\sigma^2_{e_{ij}}} = \frac{0.74}{0.74+73.7} = 0.01$$

ICC = % 1

The ICC is nearly exactly the same as with model 5, the explanation that the interaction terms did not change the proportion of variance accounted for by the hospital.

5: Conclusions and Recommendations

5.1 Conclusion:

In this Paper, we have discussed the constriction a multilevel Modeling to High-Resolution CT (HRCT) Lung in Patients with COVID-19. Used patients in Erbil, Sulaimani, Duhok and Halabja in each county precedes 2 Hospitals, between September 1th,2019 to January 2th, 2022.. Six different Model techniques have been described and their factors were calculated. According to the results achieved from this paper:

1. In our study we have 5 attributes that Patient ID, Hospital ID, HRCT of the patient, Diabetic of the patient and if the patient is a smoker (No Smoker or Past Smoker or Smoker or Passive Smoker or High Smoker), then the last one is Doctor Experience. With percentage descriptive of variable that %26.3 of the CT's patient is 15, the point is that 26.3 percent of patients with coronavirus have 15 percent of their lungs infected, and %0.6 is 100, for diabetic of the patient %0.2 is 72, %0.2 is 402. %64.5 of patients infected with coronavirus are non-smokers, %13.5 are past smoker and %10.4 are a passive smoker then %9.4 of the patients are a high smoker
2. The interclass correlation coefficient is a significant value that tells you how much variability there is between your clusters/groups. Also interpretable and useful for random cut models. The higher the correlation within the clusters, the lower the variability within the clusters and consequently the higher the variability between the clusters. Adding the level 1 predictor increased the ICC. Nevertheless, when we added a level 2 predictor, the ICC dropped dramatically to an even lower value than the unconditional model. This is due to a decrease in unexplained variation at level 2, the random intercept term u_{oj}, when a hospital-level predictor was added.
3. Finally, these three variables we received were analyzed by two-level of multilevel modeling showed that all three variables are significance at the hospital level. But in the two final models added level-2 predictor (Doctor Experience) that interaction with level-1 predictor smoker is not significant, with a p-value of > 0.001. However, diabetics are significant, meaning that Doctor's experience moderates the relationship between Diabetic and CT, but not the relationship between Smoker and CT**.**

Reference

1. Gelman, A. (2006). Multilevel (hierarchical) modeling: what it can and cannot do. Technometrics, 48(3), 432-435.
2. Goldstein, H. (2011). Multilevel statistical models. John Wiley & Sons.
3. Hair Jr, J. F., & Fávero, L. P. (2019). Multilevel modeling for longitudinal data: concepts and applications. RAUSP management journal, 54, 459-489.
4. Heck, R. H., & Thomas, S. L. (2020). An introduction to multilevel modeling techniques: MLM and SEM approaches. Routledge.
5. Maulana, R., Opdenakker, M. C., & Bosker, R. (2014). Teacher–student interpersonal relationships do change and affect academic motivation: A multilevel growth curve modelling. British journal of educational psychology, 84(3), 459-482.
6. Park, H. C., Kim, D. K., Kho, S. Y., & Park, P. Y. (2017). Cross-classified multilevel models for severity of commercial motor vehicle crashes considering heterogeneity among companies and regions. Accident Analysis & Prevention, 106, 305-314.
7. Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. Psychometrika, 69(2), 167-190.
8. Scott, M. A., Simonoff, J. S., & Marx, B. D. (Eds.). (2013). *The SAGE handbook of multilevel*
9. Shin, M., Park, J., Grimes, R., & Bryant, D. P. (2021). Effects of using virtual manipulatives for students with disabilities: Three-level multilevel modeling for single-case data. Exceptional Children, 87(4), 418-437.
10. Skrondal, A., & Rabe-Hesketh, S. (2007). Redundant overdispersion parameters in multilevel models for categorical responses.
11. Steenbergen, M. R., & Jones, B. S. (2002). Modeling multilevel data structures. american Journal of political Science, 218-237.
12. http://www.bristol.ac.uk/cmm/learning/multilevel-models/data-structures.html

إنشاء نمذجة متعددة المستويات للرئة عالية الدقة بالتصوير المقطعي المحوسب (HRCT) في المرضى المصابين بعدوى COVID-19

ديدار احمد وفا و محمد محمد الفقي

قسم الرياضيات، كلية العلوم ، جامعة السليمانية، السليمانية ، العراق.

الخلاصة

مرض الفيروس التاجي ، المعروف أيضًا باسم COVID-19 ناجم عن فيروس SARS-CoV-2. يعاني معظم الأشخاص المصابين بالفيروس من أعراض خفيفة إلى معتدلة لأمراض الجهاز التنفسي. الهدف من هذه الورقة هو بناء نموذج بنمذجة متعددة المستويات لهؤلاء المرضى الذين يعانون من فيروسات كورونا ، لدينا سبعة مستشفيات تضم (636) مريضاً في القطاعين الخاص والعام 27٪ من أربيل و 26٪ من السليمانية و 23٪ من دهوك و 24٪ من حلبجة من الفترة (1 أيلول 2019 إلى 1 شباط 2022). في هذه النمذجة لتقدير الاحتمال الأقصى المقيد متعدد المستويات (RMLE) والاحتمالية القصوى القصوى (FML) ، يقدر التأقلم معلمات النماذج متعددة المستويات (الثابتة والعشوائية). كان التطبيق على رئتي HRCT للمرضى ، وتم اختيار سبعة مستشفيات بشكل عشوائي من بين محافظة كردستان العراق. تظهر النتيجة أن المتغيرات الثلاثة جميعها مهمة على مستوى المستشفى ، ولكن في النموذجين النهائيين ، أضف متنبئًا من المستوى 2 (تجربة الطبيب) ذلك التفاعل مع متنبئ المستوى 1 (مدخن) ، وهو بعيد كل البعد عن الأهمية. ومع ذلك ، هناك علاقة مهمة بين الإصابة بمرض السكري وإجراء الأشعة المقطعية ، ولكن العلاقة بين التدخين وإجراء الأشعة المقطعية ليست مهمة.

الكلمات المفتاحية: النمذجة متعددة المستويات ، التأثير الثابت ، التأثير العشوائي ، الارتباط البيني.