Background and Objectives: Multilevel Modeling (MLM) is an increasingly popular statistical tool which extends the ordinary regression analysis to the situation of clustered as well as hierarchically structured data. Application of conventional regression methods by ignoring the structure of the data leads to the violation of the assumption of independence of errors which results in biased estimates of standard errors and spurious inference of the results. The objective of the study includes to reveal the importance of MLM over standard regression techniques when hierarchical or nested data structures are present in the datasets which is quite prevalent in educational and health sciences research.
The prevalence of depression is found to increase during adolescence especially from 15 to 18 years but often unrecognized. While adolescence is considered critical for the onset of most of the mental illnesses, depression figures out to be the most prevalent mental health problem. Depression has substantial antagonistic impact on school performance, family relations, socialization, and increases vulnerability to future depressive episodes, substance abuse, suicide, psychosocial impairment and antisocial behaviours (Frojd et. al., 2008). The current study also intends to identify the factors associated with symptoms of depression among urban adolescents in India using Multilevel modeling (MLM) approach.

Methodology: The data obtained from a cross-sectional school-based study has been used to identify the risk factors for depression among adolescents using MLM approach. The dataset consisted of youth studying in 8^{th} to 12^{th} standards from public and private institutions. The total sample comprised of 1428 school and college students aged 13 to 19 years. The students were nested in classes that were in turn nested within schools/colleges.
The demographic details of the participants were obtained using a socio-demographic data sheet. Beck depression inventory-II (BDI-II) was used to assess depression scores in individuals aged 13 and over. It is composed of items relating to depression symptoms, cognitions as well as physical symptoms. The scores on BDI range from 0 to 63 (Beck et al., 1996). A checklist for stress, coping and suicidal behaviours was used to assess the areas of stress, coping, and suicidal behaviours among the youth. Suicide probability scale (SPS) was used as a screening instrument to assess suicide risk in individuals aged 14 and older. Items of the SPS assess four areas: hopelessness, suicidal ideation, negative self-evaluation, and hostility (Cull and Gill, 1988).

Statistical analysis were done using R software version 3.5.1 (R Core Team, 2018). Bivariate analysis incorporating the nested structure of the data was done to identify the variables to retain in the final model. From the group of different multilevel models (such as two-level as well as three-level random intercept models and two-level random coefficient models) the relevant model should be chosen based on the necessity of parsimony in a model. The log-likelihoods of the two models were compared to select the model that best fits the data, if one model is a special case of the other model. The positive difference of 2 times log-likelihood has a χ^2 distribution with degrees of freedom obtained from the difference of the number of parameters to be estimated in the two models. The non-significance of the statistic suggests that a more complicated model may not be necessary. Different multilevel general linear models for the data were fitted and the models were compared using Likelihood Ratio (LR) tests, AIC, BIC along with estimates and standard error (SE)s. Intra-class correlation coefficient (ICC) values were also estimated at both school and class levels models. The final model was the random coefficient model which allowed suicidal probability score, suicide

ideation, solving family problems, academic stress to vary across schools/colleges. Bayesian multilevel models were also fitted and the results were compared.

Results: Variables found to be highly significant factors of the depression among adolescents were included in the final model and hence the best-fitted model was identified. The intraclass correlation coefficient was 0.118 specified the extent of shared variance among students in a school which indicated that multilevel modeling has to be used to incorporate the hierarchical structure of the data. Figure (1) indicated that the depression scores among students coming from different school/colleges are different and hence the school variable should be considered as a random effect in the model. The depression scores among students belong to different classes is displayed using box plot in Figure (2). The table (1) describes the importance of including the hierarchical structure of the data as random effects. The results indicated that the two-level model allowing for school variable as random effect is the best model compared to other models including three-level model.
The results from Table (2) indicated that the final random coefficients model allowing suicidal probability score, suicide ideation, solving family problems, academic stress to vary across schools/colleges best fitted the data and which outperformed the conventional as well as random intercept methods with respect to the estimated coefficients and fit indices. The results showed that the individual level variables such as suicidal probability score (β=0.447, $p<0.001$), academic stress (β=2.514, $p<0.001$), hostility (β=0.264, $p<0.001$), living with people other than parent (β=2.731, $p=0.008$), investing in close friends (β=0.267, $p=0.010$), problem in relationship with family members (β=3.136, $p<0.001$) were positively related with the depression whereas solving family problems (β=-0.228, $p<0.001$), ventilating feelings (β=-0.148, $p<0.036$) were negatively related. Also, the individual level variables such as suicide ideation, suicidal attempt, both parents alive, living with single parent, avoiding problems, problem in relationship with friends and problem in relationship with boy/girl friend were included in the final model but were not statistically significant at 5% level of significance.

Discussion: Since the conventional methods does not consider the nested or hierarchical structure of the subjects within the same cluster, the use of such models results in the underestimation of the standard errors which leads to inflation in type-I errors as well as biased inferences. Multilevel models handles this situation by including random components of cluster effects in the statistical model allowing the subjects to vary among groups. Modeling the data using multilevel approach helps to understand the variability of random effects across clusters. Hence the importance of the clusters could be evaluated by dividing the total variance in the dependent variable into between-cluster and within-cluster parts. In MLM modeling both individual-level as well as cluster-level covariates can be included in the model. Also, the aggregation bias results because of the clustering can be eliminated using MLM approach.

Conclusion: Hierarchical or nested data structures are quite prevalent in educational and health research. Ignoring the hierarchical or nested data structure of the data can lead to biased estimates of SEs which would highly influence the statistical significance. The analysis should account for the sampling technique adopted during data collection as well as the structure of the data, wherever possible. The research could identified the factors responsible for depression among adolescents using random coeffiients model so that relevant strategies can be adopted to reduce the risk among adolescents.

Keywords: Depression, Adolescents, Hierarchical structure, Multilevel models, Random coefficients model, Mixed-effects linear regression, Intraclass correlation coefficient.

Figure 1. Depression (BDI) scores of adolescents among schools/colleges

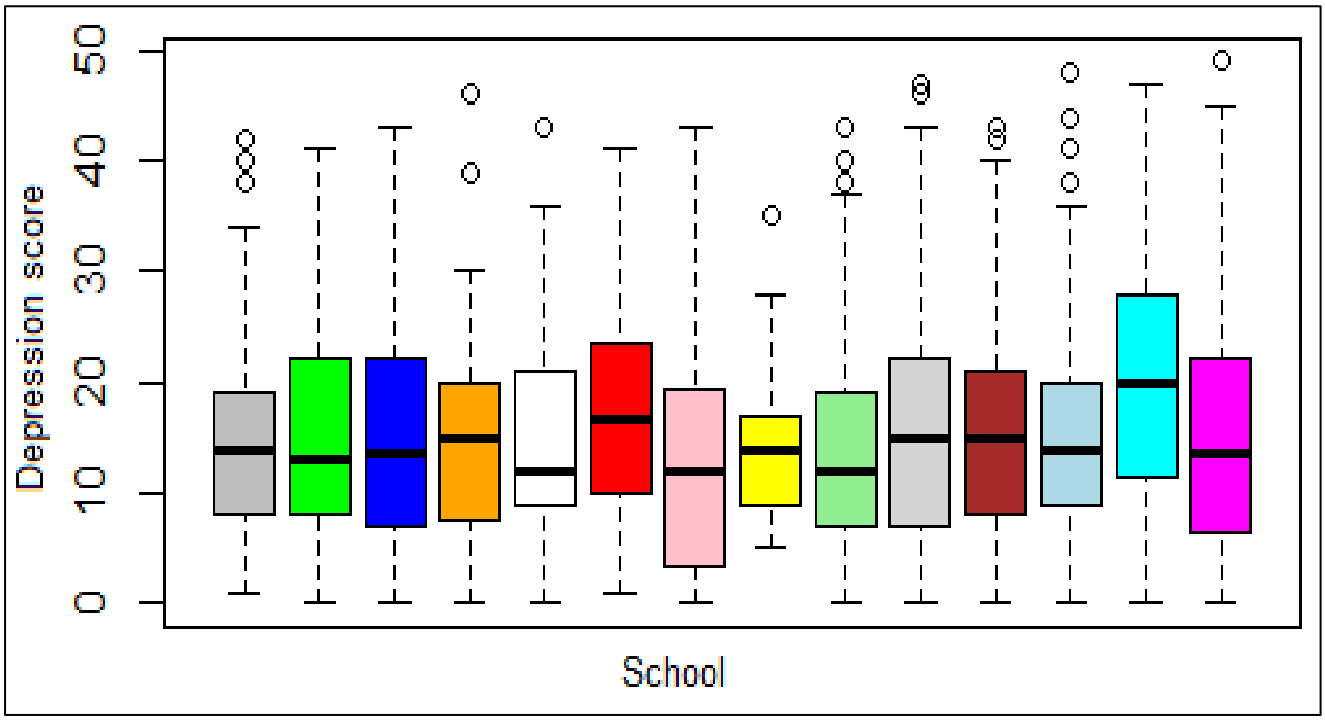

Figure 2. Depression (BDI) scores of adolescents among classes

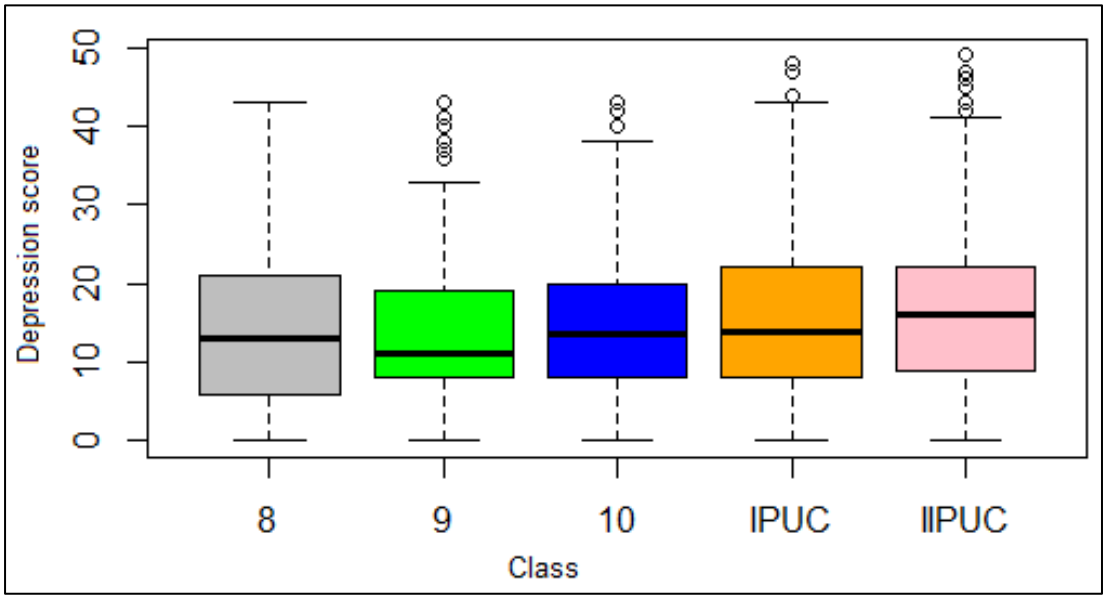

Table 1. Depression (BDI) scores of adolescents among schools/colleges

Intercept only model	Random intercept	AIC	BIC	RE*_intercept (variance)	RE*_Residual (variance)	ICC#
1	School	10505.61	10521.40	1.2785	9.518001	0.118
2	Class	10511.54	10527.33	0.95189	9.560777	0.091
3	School/ Class	10505.82	10526.87	0.92914	9.488461	

* Random effects, # Intra-class correlation coefficient

Table 2. Parameter estimates and SEs in a conventional Linear Regression model and the final (Random coefficient) model

Predictors	Conventional linear regression model		Final (Random coefficient) model	
	Coefficient	SE	Coefficient	SE
Fixed part				
Intercept	14.81***	1.178	14.85***	1.206
Probability score	0.33***	0.060	0.45***	0.088
Suicide Ideation	0.14*	0.064	0.07	0.092
Solving Family Problems	-0.20***	0.048	-0.23***	0.059
Academic stress	2.78***	0.463	2.51**	0.805
Hostility	0.31***	0.076	0.26***	0.078
Ventilating Feelings	-0.14*	0.072	-0.15*	0.071
Suicidal attempt	2.03	1.404	1.36	1.430
Both parents alive	-1.85	1.147	-1.64	1.130
Living with single parent	-0.41	1.039	-0.39	1.025
Living with others	3.20**	1.043	2.73**	1.027
Avoiding Problems	0.08	0.068	0.08	0.067
Investing in Close Friends	0.27**	0.105	0.27*	0.105
Problem in relationship with family	3.60***	0.844	3.14***	0.852
Problem in relationship with friends	1.06	0.874	1.44	0.872
Problem in relationship with boy/girl friend	1.11	1.045	1.14	1.039
Random part				
σ_e^2 (Variance at the student level)		8.127		7.930
σ_u^2 (Variance of intercept at the school level)				1.029
σ_{v1}^2 (Variance of the coefficient of Suicidal probability)				0.207
σ_{v2}^2 (Variance of the coefficient of Suicide ideation)				0.222
σ_{v3}^2 (Variance of the coefficient of Solving family problems)				0.119
σ_{v4}^2 (Variance of the coefficient of Academic stress)				2.292

***** p<0.001; ** p<0.01; * p<0.05**

References:

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck Depression Inventory-II (BDI-II) (2nd edition). San Antonio, TX: The Psychological Corporation.

Cull J, Gill W. Suicide Probability Scale (SPS) Manual. Los Angeles, Calif: Western Psychological Services, 1988.

Frojd, S. A., Nissinen, E. S., Pelkonen, M. U., Marttunen, M. J., Koivisto, A. M. and Kaltiala-Heino, R. (2008). Depression and school performance in middle adolescent boys and girls. Journal of Adolescence, 31, 485–498.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.