Random Intercept and Random Slope 2-Level Multilevel Models

Rehan Ahmad Khan
College of Statistical and Actuarial Sciences
University of the Punjab Lahore, Pakistan
stat4u@gmail.com

Shahid Kamal
College of Statistical and Actuarial Sciences
University of the Punjab Lahore, Pakistan
kamal_shahid@hotmail.com

Abstract

Random intercept model and random intercept & random slope model carrying two-levels of hierarchy in the population are presented and compared with the traditional regression approach. The impact of students' satisfaction on their grade point average (GPA) was explored with and without controlling teachers influence. The variation at level-1 can be controlled by introducing the higher levels of hierarchy in the model. The fanny movement of the fitted lines proves variation of student grades around teachers.

Keywords: Random Intercept, Random Slope, Multilevel Models, Iterative Generalized Least Square.

1. Introduction

The contextual or group effects are common in social, educational and health sciences, for example, drug users use drugs mostly due to social imbalance in their lives, means factors at community or society level are influencing them to take drugs. The depression is greatly developed by social and environmental stressors. Early childhood development is strongly affected by many environmental conditions (nature of diet, impurity in the environment, care given by mother, amount of stimulation in the environment etc). The likelihood of teenagers in risky behavior is associated with frequently accompanying the adults company. On many occasions, people avoid divorce in our society due to social or religious constraints.

The common phenomena in all the examples stated above is the influence of group or upper level characteristics on the individual or lower level traits. So there exist a natural hierarchy in all the said problems (multilevel problems) and for the proper exploration, specialized analytical tools are required. Multilevel analytical tools provide the proper estimation of such type of problems.

Multilevel Regression Models are also known as "Variance components Models" (Aitkin and Longford, 1986), "Hierarchical linear Models" (Raudenbaush and Bryk 1986, 1992, 2002), and "Random Coefficient Models" (de Leeuw and Kreft 1986, Longford 1993). Over the last few decades, the development of multilevel regression models (Goldstein 1995, 2003, Bryk&Raudenbush 1992, Longford 1993, Snijders&Bosker 1999) and their applications on quantitative and qualitative research remain in interest for the researchers (Smith 2011, Gelman and Jennifer 2007, Skrondal and Hesketh 2004, Reise and Duan, 2003). Multilevel regression models are used due to the natural hierarchy of the problematic data set. As we study and collect the data at different natural stages of the population, we should use techniques, methods, tools that indulge the variation at each stage of the hierarchy i.e., use multilevel techniques.

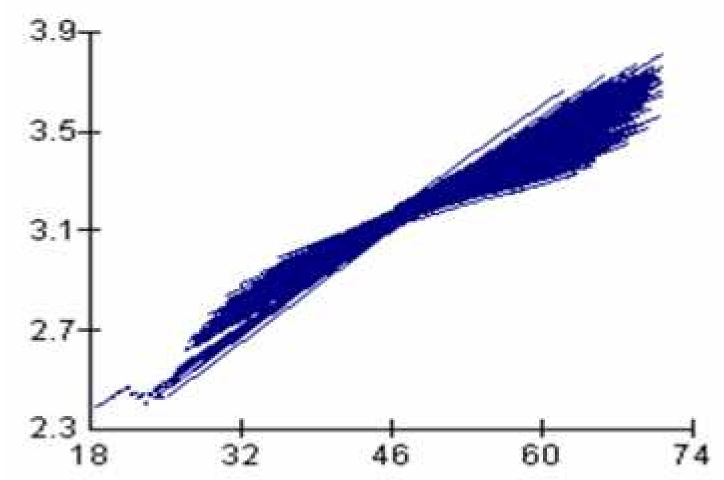
A substantial advancement has been made both in methodological and applied divisions of the multilevel models due to its wide range of applicability in every field of science. Earlier development of methodology of multilevel model is based on Junior School Project (JSP) data (Goldstein, 1986, 1989, 1995, 1997, Longford, 1987, 1993 Mortimore et al., 1988, Woodhouse, 1995). Webster et al. (1996) identify school and teacher effects on student's performance by using hierarchical linear models. Residential neighborhood has an effect on education (Benabou, 1993, Durlauf, 1996, Fernandez and Rogerson, 1997, Akerlof, 1997, Anselin, 2002). High school tracking characteristics such as selectivity, electivity, inclusiveness and scope on students have an effect on student performance (Gamoran, 1992). Extensive literature is available to study the application of multilevel models in education. Interested reader may see, Aitkin and Longford (1986), Nuttall et al. (1989), Willms (1992), Entwisle and Marton 1994), Gray et al. (1995), Goldstein and Spiegelhalter (1996), Roscigno (1998), Fielding (1999, 2004), Fraine et al. (2007, 2005). In this study, random intercept and random intercept & random slope two-level models are presented and compared with the traditional regression approach. The impact of students' satisfaction on their grade point average (GPA) was explored with and without controlling teachers influence.

The basic simple linear regression model describing the relation of response variable y and the explanatory variable x (both measured on level 1 of hierarchy) is defined as

$$y = \tau_0 + \tau_1 x + \varepsilon \qquad (1.1)$$

where, τ_0 represents intercept and τ_1 is the slope of the line. Also the random term ε is assumed to follow a normal distribution with mean 0 and constant variance σ^2. If the same variables y and x are measured for k groups and for each group regression line is fitted then we have k regression lines. Figure 1 is a display of such regression lines.

Figure 1. Regression Lines for k groups.



The fanny movement of the fitted lines suggests a variation of level-1 units across level-2 units. This means the characteristics measured at level-1 of hierarchy differ from level-2 unit to unit and if our attention is to consider the variation of all the variables measured at level 2 in a broader spectrum (population) then we need to build a single model accounting both level variables.

2. Random Intercept and Random Slope Model.

Let y_{ij} be the ij^{th} observation of a response variable y measured at level 1 and the corresponding explanatory variable x observation x_{ij} (j refers to level 2 unit and i represents level 1 unit), then the basic simple linear regression model describing a simple linear relation between the response variable y and the explanatory variable x is,

$$y_{ij} = \tau_{0j} + \tau_{1j} x_{ij} + \varepsilon_{ij} \qquad i = 1,2,...,n \quad j = 1,2,...,k \tag{2.1}$$

by fitting such k regression models we estimate $2k+1$ parameters, namely $(\tau_{0j}, \tau_{1j})\; j = 1,2,...,k \;\&\sigma_e^2$. Let τ_{0j} and τ_{1j} are the random variables as their magnitude varies in k linear regression lines and assume that

$$\tau_{0j} = \tau_0 + u_{0j} \tag{2.2}$$

$$\tau_{1j} = \tau_1 + u_{1j} \tag{2.3}$$

where, u_{0j} and u_{1j} are the unexplained parts of (τ_{0j}, τ_{1j}) with parameters

$$E(u_{0j}) = E(u_{1j}) = 0, \mathrm{var}(u_{0j}) = \sigma_{u0}^2, \mathrm{var}(u_{1j}) = \sigma_{u1}^2, and\, \mathrm{cov}(u_{0j}, u_{1j}) = \sigma_{u01}^2 .$$

The terms $\tau_0 \, and \, \tau_1$ are the average intercept and average slope of the k regression lines respectively and the random variables $u_{0j} \,\&\, u_{1j}$ referred to as residuals are the random departure of level-2 units from $\tau_0 \, and \, \tau_1$ respectively. In addition, $u_{0j} \sim N(0, \sigma_{u0}^2)$ and $u_{1j} \sim N(0, \sigma_{u1}^2)$. Also the residual term ε_{ij} introduced at level-1 represents a random variation within level-1 units.

By considering (2.2) and (2.3), model (2.1) can be written as

$$y_{ij} = \tau_0 + u_{0j} + \tau_1 x_{ij} + u_{1j} x_{ij} + \varepsilon_{ij}$$

$$or$$

$$y_{ij} = \underbrace{\tau_0 + \tau_1 x_{ij}}_{Fixed\ Part} + \underbrace{u_{0j} + u_{1j} x_{ij} + \varepsilon_{ij}}_{Random\ Part} \tag{2.4}$$

The presence of two residuals in the model separates it from the simple linear regression model (1) and for the estimation of the model we need to estimate two fixed parameters $\tau_0 \, and \, \tau_1$, and four random parameters $\sigma_{u0}^2, \sigma_{u1}^2, \sigma_{u01} \, and\, , \sigma_{e0}^2$ where $\sigma_{e0}^2 = \mathrm{var}(\varepsilon_{ij})$. The model (2.4) can also be described in matrices forms as:

$$\mathbf{Y} = \mathbf{X\tau} + \mathbf{R} \tag{2.5}$$

where

$$\mathbf{Y}=\begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n_k k} \end{bmatrix}, \quad \mathbf{X}=\begin{bmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{n_k k} \end{bmatrix}, \boldsymbol{\tau}=\begin{bmatrix} \tau_0 \\ \tau_1 \end{bmatrix} \tag{2.6}$$

$$\mathbf{R}=\begin{bmatrix} e_{11}+u_1 & e_{12}+u_2 & \dots & e_{1j}+u_j & \dots & e_{1k}+u_k \\ e_{21}+u_1 & e_{22}+u_2 & \dots & e_{2j}+u_j & \dots & e_{2k}+u_k \\ \vdots & \vdots & & \vdots & & \vdots \\ e_{i1}+u_1 & e_{i2}+u_2 & \dots & e_{ij}+u_j & \dots & e_{ik}+u_k \\ \vdots & \vdots & & \vdots & & \vdots \\ e_{n1}+u_1 & e_{n2}+u_2 & \dots & e_{nj}+u_j & \dots & e_{nk}+u_k \end{bmatrix} \tag{2.7}$$

$$\mathbf{R}=\begin{bmatrix} e_{11} & e_{12} & \dots & e_{1j} & \dots & e_{1k} \\ e_{21} & e_{22} & \dots & e_{2j} & \dots & e_{2k} \\ \vdots & \vdots & & \vdots & & \vdots \\ e_{i1} & e_{i2} & \dots & e_{ij} & \dots & e_{ik} \\ \vdots & \vdots & & \vdots & & \vdots \\ e_{n1} & e_{n2} & \dots & e_{nj} & \dots & e_{nk} \end{bmatrix}+\begin{bmatrix} u_1 & u_2 & \dots & u_j & \dots & u_k \\ u_1 & u_2 & \dots & u_j & \dots & u_k \\ \vdots & \vdots & & \vdots & & \vdots \\ u_1 & u_2 & \dots & u_j & \dots & u_k \\ \vdots & \vdots & & \vdots & & \vdots \\ u_1 & u_2 & \dots & u_j & \dots & u_k \end{bmatrix} \tag{2.8}$$

$$\mathbf{R}=\mathbf{R}_1+\mathbf{R_2} \tag{2.9}$$

Where,

$$\mathbf{R}_1=e_{ij}^{(1)}=e_{ij} \;\&\; \mathbf{R_2}=e_j^{(2)}=u_j \tag{2.10}$$

also, the matrices of residuals have the following assumptions:

$$E(\mathbf{R_1})=E(\mathbf{R_2})=\mathbf{0} \tag{2.11}$$

$$E(\mathbf{R_1}\mathbf{R_1}^T)=\mathbf{V_{2(1)}}, E(\mathbf{R_2}\mathbf{R_2}^T)=\mathbf{V_{2(2)}} \tag{2.12}$$

$$E(\mathbf{R_1}\mathbf{R_2}^T)=0, \mathbf{V_2}=\mathbf{V_{2(1)}}+\mathbf{V_{2(2)}} \tag{2.13}$$

It is also assumed that the residuals at level-1 are independent to each other, so $\mathbf{V_{2(1)}}$ is a diagonal with ij^{th} element. Thus,

$$\mathbf{V_{2(1)}}=\text{var}(e_{ij})=\sigma_{e_{ij}}^2=\mathbf{X_j}^T\Omega_1\mathbf{X_j},\ \Omega_1=\sigma_{e0}^2 \tag{2.14}$$

Similarly by assuming independence of residuals at level- 2, we gain $\mathbf{V_{2(2)}}$ block diagonal with j^{th} elements

$$\mathbf{V_{2(2)}}=\text{var}(u_j)=\mathbf{X_j}^T\mathbf{\Omega_2}\mathbf{X_j},\ \mathbf{\Omega_2}=\begin{bmatrix} \sigma_{u0}^2 & \sigma_{u01} \\ \sigma_{u01} & \sigma_{u1}^2 \end{bmatrix} \tag{2.15}$$

The j^{th} block of V_2 is therefore given by,

$$\mathbf{V_{2j}} = \mathbf{V} = \begin{bmatrix} \sigma_{u0}^2 \mathbf{J_{(n)}} + \sigma_{e0}^2 \mathbf{I_{(n)}} & 0 \\ 0 & \sigma_{u0}^2 \mathbf{J_{(n-1)}} + \sigma_{e0}^2 \mathbf{I_{(n-1)}} \end{bmatrix} = \oplus_i \sigma_{e_{ij}}^2 + V_{2(2)j} \tag{2.16}$$

where, $\mathbf{I_{(n)}}$ is an $(n \times n)$ identity matrix and $\mathbf{J_{(n)}}$ is a $(n \times n)$ matrix of ones. Furthermore, $\oplus$ is a direct sum operator.

The Generalized Least Square (GLS) estimates of $\boldsymbol{\tau}$ can be obtained by using the relation:

$$\hat{\boldsymbol{\tau}} = \left(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y} \tag{2.17}$$

with variance-covariance matrix $\left(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}\right)^{-1}$. For known $\boldsymbol{\tau}$ one can estimate the residuals as $\tilde{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\boldsymbol{\tau} = \mathbf{E}_1 + \mathbf{E}_2$ with covariance matrix $\mathbf{Y}^* = \tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^T$ and $E(\mathbf{Y}^*) = \mathbf{V}$.

This is an iterative procedure by starting some reasonable estimates of fixed parameters and setting $\sigma_{u0}^2 = 0$. The estimates obtained from (2.17) are known as iterative generalized least square estimates as the procedure continues until the estimates converge (Goldstein, 1995).

Model (1.1) and (2.1) are estimated by using a real educational data collected by the researcher for PhD work. The data was collected from 40000 university students nested within 1000 university teachers. Students were considered as level-1 units and teachers as level-2. The response variable y was recorded as students grade point average score (GPA) and the explanatory variable x was student satisfaction with the university. Suppose we want to investigate whether the student grades vary from teacher to teacher and the impact of student satisfaction on their grades. We treat student satisfaction variable as a random variable across the teachers i.e., the coefficient of student satisfaction will vary across the teachers. We now assume a model which includes the possibility that the teachers have different slopes. This implies that the coefficient of explanatory variable will vary from teacher to teacher. Model (2.1) may be re-write by relating grade point average and student satisfaction (Stu_Sat) as,

$$GPA_{ij} = \tau_{0j} + \tau_{1j} Stu\_Sat_{ij} + e_{ij} \tag{2.18}$$

$$\tau_{0j} = \tau_0 + u_{oj} \tag{2.19}$$

$$\tau_{1j} = \tau_1 + u_{1j} \tag{2.20}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N\left(0, \Omega_u\right) \quad : \quad \mathbf{\Omega_u} = \begin{bmatrix} \sigma_{u0}^2 & \\ \sigma_{u01} & \sigma_{u1}^2 \end{bmatrix} \tag{2.21}$$

$$e_{ij} \sim N(0, \sigma_e^2) \tag{2.22}$$

Now both the intercept and the slope vary randomly across teachers. Hence both the parameters τ_{0j} and τ_{1j} have a subscript j. Equation (2.19) describe that the intercept for the jth teacher (τ_{0j}) is given by τ_0, the average intercept across all the teachers, plus a random departure u_{0j}. In the same way, equation (2.20) states that the slope for the jth teacher (τ_{1j}) is given by τ_1, the average intercept across all the teachers, plus a random departure u_{1j}. The parameters τ_0 and τ_1 are the fixed intercept and slope of (2.18) and jointly give the average line across all students nested in all teachers. The term u_{0j} and u_{1j} represent the random departures from τ_0 and τ_1, or residuals at the teachers level i.e., they allow the jth teacher summary line to differ from the average line in both its slope and its intercept. The term u_{0j} and u_{1j} follows a bivariate Normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{\Omega_u}$. Here the covariance matrix $\mathbf{\Omega_u}$ is a 2×2 matrix having $\sigma^2_{u0}, \sigma^2_{u1}$ and σ_{u01} its elements. The term σ^2_{u0} represents the variation in the intercepts across the teachers summary line, σ^2_{u1} represent the variation in the slopes across the teachers summary line and the term σ_{u01} shows the covariance between the teachers intercepts and slopes. Finally, student's grades depart from their teachers summary line by an amount e_{ij}, which is assumed to be normally distributed with mean 0 and variance σ^2_e. Least square estimates of parameters under model 1.1 are available in table 1 and the iterative generalized least square estimates under model (2.18) are available in table 2.

Table 1: Estimates of Parameters of Simple Linear Regression Model

Parameters	Estimates	S.E	Statistic	P-Value
τ_0 : Intercept	2.054	0.003	684.667	0.000***
τ_1 : Student Satisfaction	0.024	0.00013	184.615	0.000***
σ^2_e : Variation at Student Level	0.017	0.00023	73.913	0.000***
-2*loglikelihood = -48566.671				

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

Table 2: Estimates of Parameters of Random Intercept and Random Slope Model

Parameters	Estimates	S.E	Statistic	P-Value
τ_0 : Average Intercept	2.091	0.007	298.714	0.000***
τ_1 : Student Satisfaction	0.023	0.00031	74.194	0.000***
σ^2_{u0} : Intercept Variation at Teachers Level	0.030	0.002	15.000	0.000***
σ^2_{u1} : Slope Variation at Teachers Level	0.00014	0.000025	5.600	0.000***
σ_{u01} : Variation b/w Intercept and Slope	-.001	0.00038	-2.632	0.00432**
σ^2_e : Variation at Student Level	0.015	0.00029	51.724	0.000***
-2*loglikelihood = -52291.468				

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

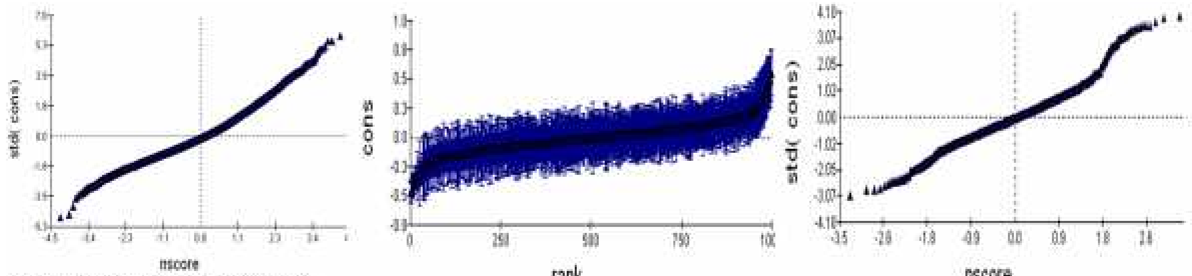
The estimated coefficient of τ_1 is close to the estimate obtained from the model with a simple slope. However, the individual teacher slopes vary about this mean with a estimated variance $0.00014(s.e=0.000025)$. The intercepts of the individual teacher lines also differ. Their mean is $2.091(s.e=0.007)$ and their variance is $0.030(s.e=0.002)$. In addition, there is a negative covariance between intercepts and slopes estimated as $-.001(s.e=0.00038)$ suggesting that teachers with lower intercepts tend to have steeper slopes. This can also be confirmed with the correlation between intercepts and slopes across teachers estimated as -0.488. This negative correlation will lead to a fanning pattern of teachers lines. The variance for level-1 residuals (e_{ij}) is 0.015 with standard error 0.00029. The goodness of fit of a model may be explored through graphs of residuals and predicted values. The residuals estimated at any level can be used to test the assumption of normality i.e., at each level of hierarchy, it is assumed that the residuals should follow a normal distribution and this assumption may be checked by using a Normal probability plot, in which the ranked residuals are plotted against corresponding points on a Normal distribution curve and a straight line of points on a Normal plot indicate normality in residuals.

Graph 2(a) is a normal probability plot of residuals at level-1 and graph 2(c) is a normal probability plot of residuals at level-2. Linear pattern of points in these graphs proves the normality assumption. Graph 2(b) is a caterpillar plot showing the residuals of 1000 level-2 units (teachers) with the 95% confidence intervals around them. Clearly, there is no considerable overlap of intervals indicating teachers have significant different means of student's grades. In last, graph 2(d) is a normal probability plot of standardized residuals under model 1.1 depicting normality.

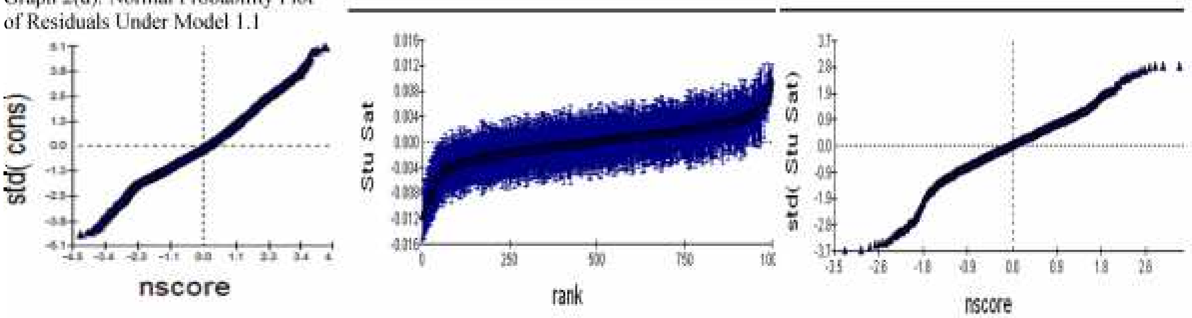
Graph 2(a). Normal Probability Plot of Residuals at Level-1

Graph 2(b). Caterpillar Plot of Residuals at Level-2

Graph 2(c). Normal Probability Plots of Residuals at Level-2



Graph 2(d). Normal Probability Plot of Residuals Under Model 1.1



3. Random Intercept Model

Suppose we wish to examine the relationship between grade point average (response variable) and the teachers. The population is considered to have a two-level hierarchical structure with student's grade point average (GPA) "y_{ij}" at level-1 and teachers at level-2. The random effect model with no explanatory variable can be described as,

$$y_{ij} = \tau_{0j} + e_{ij} \qquad e_{ij} \sim N(0, \sigma_e^2) \tag{3.1}$$

$$\tau_{0j} = \tau_0 + u_{oj} \qquad u_{0j} \sim N(0, \sigma_{u0}^2) \tag{3.2}$$

from (3.2), (3.1) becomes,

$$y_{ij} = \tau_0 + e_{ij} + u_{oj} \tag{3.3}$$

Where e_{ij} and u_{0j} are level-1 and level-2 residuals respectively. In this model u_{0j}, the teacher's effect is assumed to be random variable having a normal distribution with variance σ_{u0}^2.

Model (3.1) is also known as variance components model because it partitions the residual variance into two components, level-2 variance (Between groups variance σ_{u0}^2) and level-1 variance (Within a group variance σ_e^2). The iterative generalized least square estimates of the parameters of model are given in table 3.

Table 3: Estimated values of parameters of Random Intercept Model

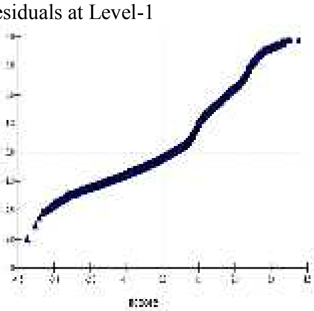
Parameters	Estimates	S.E	Statistic	P-Value
τ_{0j}	3.049	0.006	508.167	0.000***
σ^2_{u0}	0.030	0.001	30.0	0.000***
σ^2_e	0.043	0.00023	186.96	0.000***
-2*loglikelihood = -9282.89				

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

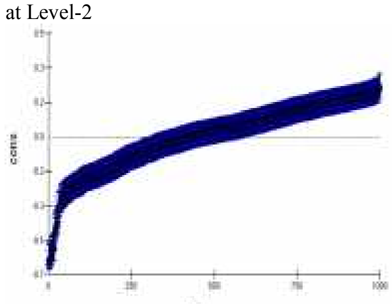
The overall mean of GPA is estimated as $\hat{\tau}_{0j} = 3.049$. The means for the different teachers are distributed about their overall mean with an estimated variance of 0.030. The variance among students within teachers is estimated as $\sigma^2_e = 0.043$ and among teachers variance is estimated as $\sigma^2_{u0} = 0.030$. In order to test $H_0 : \sigma^2_{u0} = 0$ (analogues to testing $H_0 : \tau_1 = \tau_2 = \tau_3 ... = \tau_{1000} = 0$ in the fixed effect model), this variance appears significantly different from zero $(z = 30.0, p < 0.001)$.

Graph 3(a) & 3(c) are normal probability plots of residuals at level-1 and level-2 respectively. Both graphs showed a fairly linear pattern of points indicating no worry about violation of normality assumption. In addition, points on graph 3(a) are linear as our response variable is normally distributed. Graph 3(c) is a caterpillar plot showing the residuals of 1000 level-2 units (teachers) with the 95% confidence intervals around them. Clearly, there is no considerable overlap of intervals indicating teachers have significant different means of student's grades. These residuals also represent the departure of level-2 units (teachers) from the overall average predicted by the fixed parameter, this means that these are the teachers that differ significantly from the average at 5% level.

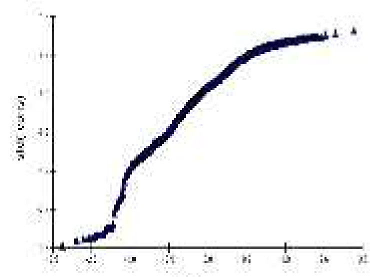
Graph 3(a). Normal Probability Plot of Residuals at Level-1



Graph 3(b). Caterpillar Plot of Residuals at Level-2



Graph 3(c). Normal Probability Plots of Residuals at Level-2



Discussion

The value of average intercept has increased from the fixed intercept while the value of random slope slightly decline with the value of fixed slope. The variation at level-1 also reduced due to introducing the level-2 variation suggesting that the individual level variation may be controlled by introducing higher levels of hierarchy in the model. The reduction in the students' grades variation under two-level model from the one-level model confirms the teachers influence over students grades.

The goodness of fit of model (3.1) may be tested through a compact test named likelihood ratio test. In a likelihood ratio test of $H_0 : \sigma^2_{u0} = 0$, we compare the model (3.1) with a model where σ^2_{u0} is constrained to equal to zero, i.e., the single level model with only an intercept term $(y = \tau_0 + e)$. The value of the likelihood ratio statistic is the difference between the likelihood ratio of model (3.1) and the likelihood of the single level model $(y = \tau_0 + e)$ which is compared to a chi-squared distribution with 1 degree of freedom i.e.,

$$\chi^2 = likelihood\ ratio\ of\ single\ level\ model\ -\ likelihood\ ratio\ of\ level-2\ model$$
$$= 8705.819 - (-9282.89)$$
$$= 17988.709$$

We conclude that there is a significant variation among teachers $(\chi^2 = 17988.709, p < 0.001)$. The variance partition coefficient

$$VPC = \frac{\sigma^2_{u0}}{\sigma^2_{u0} + \sigma^2_e} = \frac{0.030}{0.030 + 0.043} = 0.411$$

is showing 41% of the total variance in students grade point average is due to the differences among the teachers.

Similarly, model (1.1) is compared with a random slope model (2.1) and noticed that the value of $-2 * \log likelihood$ has decreased from $-48566.671 to -52291.468$, a difference of 3724.797. Since the model (2.1) involves two additional parameters, the variance of slope residuals σ^2_{u1}, and the covariance between intercepts and slopes σ_{u01}, so this difference follows a $\chi^2 - distribution$ with $2 d.f$. Under the null hypothesis that the extra parameters have population values of zero the change is highly significant, confirming the better fit of the model. In addition, about $65\% (VPC = 0.653)$ variation in students' grades is due to the variation among teachers. Finally, it is concluded that a model may be predicted more efficiently by considering higher levels of hierarchy in a hierarchical population.

References

1. Aitkin, M. and Longford, N. (1986). Statistical Modelling in School Effectiveness studies (with discussion). *Journal of the Royal Statistical Society*, A149, 1-43.
2. Akerlof, G.A. (1997). Social Distance and Social Decisions. *Econometrica*, Econometric Society, 65, 5,1005-1028.
3. Anselin, L. (2002). Under the hood: Issues in the specification and interpretation of spatial regression models.*Agricultural Economics*, Blackwell, 27, 3, 247-267.

4. Bénabou, R. (1993). Workings of a City: Location, Education and Production. *Quavlerly Journal of Economics*, 108, 619-652.
5. Bryk, A.S. and Raudenbush, S.W. (1992). *Hierarchical Linear Models*. Newbury Park, CA: Sage.
6. De Fraine, B., Van Damme, J., &Onghena, P. (2007). A Longitudinal analysis of gender differences in academic self-concept and language achievement: A multivariate latent growth curve approach. *Contemporary Educational Psychology*, 32, 132-150.
7. De Fraine, B., Van Landeghem, G., Van Damme, J., &Onghena, P. (2005). An analysis of well-being in secondary school with multilevel growth curve models and multilevel multivariate models. *Quality and Quantity*, 39, 297-316.
8. De Leeuw, J. and Kreft, Ita G.G. (1986). Random Coefficient Models. Journal of Educational Statistics, 11, 1, 55-85.
9. Durlauf, S.N. (1996). Statistical Mechanics Approaches to Socioeconomic Behavior.NBER Technical Working Papers 0203, *National Bureau of Economic Research*, Inc.
10. Entwistle, N.J. and Marton, F. (1994). Knowledge objects: understandings constituted through intensive academic study. *British Journal of Educational Psychology*, 64,161-78.
11. Fielding, M. (1999). Radical collegiality: Affirming teaching as an inclusive professional practice.*Australian Educational Researcher*, 26, 2, 1–34.
12. Fielding, M. (2004). Transformative approaches to student voice: Theoretical underpinnings, recalcitrant realities. *British Educational Research Journal*, 30, 2, 295–311.
13. Gelman, A. and Jennifer, H. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.
14. Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalised least squares. *Biometrika*, 73, 43-56.
15. Goldstein, H. (1989a). Restricted unbiased iterative generalised least squares estimation. *Biometrika*, 76, 622-623.
16. Goldstein, H. (1995). *Multilevel Statistical Models*, 2nd edition. London, Edward Arnold: New York, Wiley.
17. Goldstein, H. (2003). *Multilevel Statistical Models*, 3rd edition. London, Edward Arnold: New York, Wiley.
18. Goldstein, H. (1997). Methods in school effectiveness research. *School Effectiveness and School Improvement*, 8, 369-395.
19. Goldstein, H. and Spiegelhalter, D.J. (1996). League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society*, A159, 505-513.
20. Gray, J., Jesson, D., Goldstein, H., Hedger, K. and Rasbash, J. (1995). A multi-level analysis of school improvement: Changes in school performance over time. *School Effectiveness and School Improvement*, 6, 2, 97-114.

21. Longford, N.T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, 74, 817-827.
22. Longford, N.T. (1993). *Random Coefficient Models*. Oxford: Clarendon Press.
23. Mortimore, P., Sammons, P., Stoll, L., Lewis, D. and Ecob, R. (1988). School Matters. Wells, Open Books.
24. Nuttall, D.L., Goldstein, H., Prosser, R. and Rasbash, J. (1989). Differential school effectiveness. *International Journal of Educational Research*, 13, 769-776.
25. Raudenbush, S.W. and Bryk, A.S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 12, 241-269.
26. Raudenbush, S.W. and Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd ed. Newbury Park, CA: Sage.
27. Reise, S.P. and Duan, N. (2003). *Multilevel Modeling: Methodological Advances, Issues, and Applications*. Lawrence Erlbaum Associate, Inc.
28. Roscigno, V.J. (1998). Race and Reproduction of Educational Disadvantages. *Social Forces,* 76, 1033-1060.
29. Skrondal, A. and Hesketh, S.R. (2004). *Generalized Latent Variable Models: Multilevel, longitudinal, and structural equation models*. Boca Raton: Chapman & Hall.
30. Smith, R.B. (2011). *Multilevel Modeling of Social Problems: A Causal Perspective*. New York, Springer.
31. Snijders, T and Bosker, R. (1999). *Multilevel Analysis*. London, Sage.
32. Webster, W., Mendro, R., Orsak, T. and Weerasinghe, D. (1998). An application of hierarchical linear modeling to the estimation of school and teacher effect. *Paper presented at the Annual Meeting of the American Educational Research Association*, San Diego, CA, April 13-17.
33. Willms, J.D. (1992). *Monitoring school performance: A guide for educators*. Lewes: Falmer Press.
34. Woodhouse, G. (1995) (Ed.). *A Guide to MLn for New Users*. London: Multilevel Models Project, University of London.