Generalization of Random Intercept 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@yahoo.com

Abstract

The concept of random intercept models has been generalized for k-levels in multilevel models. The random variation in intercepts at individual level is marginally split into components by incorporating higher levels of hierarchy in the single level model. This showed a control in the random variation in intercepts by incorporating the higher levels of hierarchy in the model.

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

1. Introduction

A random intercept model is one in which only the intercept of the model vary across the different levels of the model and extracts the information that how the values of response variable vary across the diverse nature of higher levels units in a hierarchical population. In such models, the random variation due to all the hierarchies is taking into account as we split overall random variation into variations due to all the levels. So, sometimes these models are also named as "Variance components Models" (Longford, 1986), "Hierarchical linear Models" (Raudenbaush and Bryk, 1986, 2002), and "Random Coefficient Models" (de Leeuw and Kreft, 1986, Moulton, 1986, Longford, 1992, Goldstein, 1995, Snijders and Bosker, 1999).

2. Random Intercept Model for K-Levels

The concept of random intercept models in a multilevel model developed by Goldstein (1986), Khan and Kamal (2012a,b, 2013) has been extended for k-levels. Suppose y_{i_1} is a response variable measured at individuals' level, where the subscript i_1 refers to Level-1 m_1 units. A simple model relating the response variable with the intercept term may be defined as,

$$y_{i_1} = \tau_0 + \varepsilon_{i_1} \tag{2.1}$$

where, τ_0 is an intercept term representing the average value of the response variable y_{i_1} and is fixed. The term ε_{i_1} is a residual term representing the random departure of values from the fixed intercept τ_0. Also $\varepsilon_{i_1} : NII(0, \sigma_1^2)$.

Again, suppose y_{i_1,i_2} is a response variable measured at Level-1 and the subscript i_2 refers to Level-2 m_2 units. Again, a two-levels random intercept model may be defined as,

$$\left.\begin{aligned} y_{i_1,i_2} &= \tau_{0i_2} + \varepsilon_{i_1,i_2} \\ \tau_{0i_2} &= \tau_0 + \varepsilon_{i_2} \\ y_{i_1,i_2} &= \tau_0 + \varepsilon_{i_2} + \varepsilon_{i_1,i_2} \end{aligned}\right\} \tag{2.2}$$

where, τ_0 is an intercept term representing the average value of all the intercepts. The term ε_{i_2} is a Level-2 residual term representing the variation of intercept terms across the second level units with $\varepsilon_{i_2} : NII(0,\sigma_2^2)$ and ε_{i_1,i_2} represents the random variation in intercepts due to Level-1 units. Also $\varepsilon_{i_1,i_2} : NII(0,\sigma_1^2)$. Model (2.2) is a random intercept multilevel model as it includes the variation of both the levels. Since partitions the individual variation into two components (one is for Level-1 units and the other is due to Level-2 units) so, it is also called random coefficient two-level model.

Similarly, y_{i_1,i_2,i_3} is $(i_1i_2i_3)^{th}$ value of a response variable y measured at Level-1 where the subscript i_3 refers to Level-3 m_3 units. Again, a three-level random intercept model may be defined as,

$$\left.\begin{aligned} y_{i_1,i_2,i_3} &= \tau_{0i_2,i_3} + \varepsilon_{i_1,i_2,i_3} \\ \tau_{0i_2,i_3} &= \tau_0 + \varepsilon_{i_2,i_3} + \varepsilon_{i_3} \\ y_{i_1,i_2,i_3} &= \tau_0 + \varepsilon_{i_3} + \varepsilon_{i_2,i_3} + \varepsilon_{i_1,i_2,i_3} \end{aligned}\right\} \tag{2.3}$$

where, τ_0 is an intercept term representing the average value of all the intercepts across Level-2 and Level-3 units. The term ε_{i_3} is a Level-3 residual term representing the variation of intercept terms across the third level units with $\varepsilon_{i_3} : NII(0,\sigma_3^2)$, ε_{i_1,i_2} represents the random variation in intercepts due to Level-2 units and $\varepsilon_{i_1,i_2,i_3}$ represents the random variation in intercepts due to primary units. Also $\varepsilon_{i_1,i_2} : NII(0,\sigma_2^2)$ and $\varepsilon_{i_1,i_2,i_3} : NII(0,\sigma_1^2)$. Model (2.3) is a random intercept three-level multilevel model as it includes the variation due to all three levels of hierarchy in the population.

We generalize the random intercept multilevel model for k-levels by method of induction. Suppose $y_{i_1,\ldots,i_k}$ be the $i_1 \ldots i_k{}^{th}$ observation of a response variable y measured at Level-1 where the subscript i_1 refers to Level-1 m_1 units, i_2 refers to Level-2 m_2 units, i_3 refers to Level-3 m_3 units and so on the subscript i_k represents m_k units measured across Level-k. If we assume the variation of response variable across Level-1,...,Level-k only in the intercept term then a k-levels model may be define as,

$$y_{i_1,\ldots,i_k} = \tau_{i_1,\ldots,i_k} + \varepsilon_{i_1,\ldots,i_k} \tag{2.4}$$

where,

$$\tau_{i_1,\dots,i_k} = \tau_0 + \varepsilon_{i_2,\dots,i_k} + \varepsilon_{i_3,\dots,i_k} + \dots + \varepsilon_{i_k} \tag{2.5}$$

If we substitute (2.5) in (2.4), we have

$$y_{i_1,\dots,i_k} = \tau_0 + \varepsilon_{i_1,\dots,i_k} + \varepsilon_{i_2,\dots,i_k} + \varepsilon_{i_3,\dots,i_k} + \dots + \varepsilon_{i_k} \tag{2.6}$$

where, τ_0 is the average intercept and the terms $\varepsilon_{i_1,\dots,i_k}, \varepsilon_{i_2,\dots,i_k}, \varepsilon_{i_3,\dots,i_k}, \dots, \varepsilon_{i_k}$ are the residuals due to random variations of Level-1, Level-2, Level-3,..., Level-k units in the model (2.4). Also $\varepsilon_{i_1,\dots,i_k} : NII(0,\sigma_1^2)$, $\varepsilon_{i_2,\dots,i_k} : NII(0,\sigma_2^2), \dots,$ $\varepsilon_{i_k} : NII(0,\sigma_k^2)$. Since we measure no predictor at any level of the model and for estimation purpose we assume a dummy explanatory variable $x_{i_1,\dots,i_k}$ at Level-1 with all the value zero, so the model (2.6) now takes the form,

$$y_{i_1,\dots,i_k} = \tau_0 + \tau_1 x_{i_1,\dots,i_k} + \varepsilon_{i_1,\dots,i_k} + \varepsilon_{i_2,\dots,i_k} + \varepsilon_{i_3,\dots,i_k} + \dots + \varepsilon_{i_k} \tag{2.7}$$

The model (2.7) can be written in matrices form as,

$$\mathbf{Y} = \boldsymbol{\tau}\mathbf{X} + \mathbf{R} \tag{2.8}$$

where,

$$\mathbf{Y} = \left[y_{i_1,\dots,i_k} \right]_{m_1\times 1}, \mathbf{X} = \left[\mathbf{1} \quad \mathbf{x_{i_1,\dots,i_k}} \right]_{m_1\times 2}, \boldsymbol{\tau} = \begin{bmatrix} \tau_0 \\ \tau_1 \end{bmatrix} \tag{2.9}$$

&

$$\mathbf{R} = \begin{bmatrix} \boldsymbol{\varepsilon}_{\mathbf{i_1\dots 1k}} + \boldsymbol{\varepsilon}_{\mathbf{i_2\dots 1k}} + \dots + \varepsilon_{1k} \\ \boldsymbol{\varepsilon}_{\mathbf{i_1\dots 2k}} + \boldsymbol{\varepsilon}_{\mathbf{i_2\dots 2k}} + \dots + \varepsilon_{2k} \\ \vdots \\ \boldsymbol{\varepsilon}_{\mathbf{i_1\dots m_k k}} + \boldsymbol{\varepsilon}_{\mathbf{i_2\dots m_k k}} + \dots + \varepsilon_{m_k k} \end{bmatrix} = e^{(1)}_{i_1\dots i_k} + e^{(2)}_{i_2\dots i_k} + \dots + e^{(k)}_{i_k} \tag{2.10}$$

$$e^{(1)}_{i_1\dots i_k} = R_1 = \varepsilon_{i_1\dots i_k}, e^{(2)}_{i_2\dots i_k} = R_2 = \varepsilon_{i_2\dots i_k}, \dots, e^{(k)}_{i_k} = R_k = \varepsilon_{i_k} \tag{2.11}$$

Now, $\operatorname{var}(y_{i_1,\dots,i_k}) = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_k^2 = \sum_{j=1}^{k} \sigma_j^2$

where, $\mathbf{Y}$ is a vector of m_1 units of response variable, $\mathbf{1}$ is a vector of ones, $\mathbf{x_{i_1,\dots,i_k}}$ is a vector of m_1 units, The terms $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$ are the variances across Level-1, Level-2,...,Level-k units respectively. We assume that the covariances between the residuals are zero. Also the matrices of residuals have the following assumptions:

$$\left.\begin{aligned} & E(\mathbf{R_h}) = 0, \qquad \text{for h=1,2,...,k} \\ & E(\mathbf{R_1 R_1^{\mathit{T}}}) = \mathbf{V_{k(1)}}, E(\mathbf{R_2 R_2^{\mathit{T}}}) = \mathbf{V_{k(2)}}, \dots, E(\mathbf{R_k R_k^{\mathit{T}}}) = \mathbf{V_{k(k)}} \\ & E(\mathbf{R_f R_h^{\mathit{T}}}) = \mathbf{0}, \quad \text{for } f \neq h, \text{ where f,g=1,2,...,k} \\ & \mathbf{V} = Cov(\mathbf{R}) = \mathbf{V_{k(1)}} + \mathbf{V_{k(2)}} + \dots + \mathbf{V_{k(k)}} \end{aligned}\right\} \tag{2.12}$$

If $\mathbf{V}$ is known then the Generalized Least Square (GLS) estimates of $\boldsymbol{\tau}$ can be obtained by using the relation,

$$\hat{\boldsymbol{\tau}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y} \quad \text{with} \quad \text{cov}(\hat{\boldsymbol{\tau}}) = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1} \tag{2.13}$$

If $\boldsymbol{\tau}$ is known and $\mathbf{V}$ is unknown then the estimates of $\boldsymbol{\tau}$ can be obtained again by using GLS estimator,

$$\hat{\boldsymbol{\tau}} = (\mathbf{X}'^T\mathbf{V}'^{-1}\mathbf{X}')^{-1}\mathbf{X}'^T\mathbf{V}'^{-1}\mathbf{Y}' \tag{2.14}$$

Where $\mathbf{Y}'$ is the vector of upper triangle elements of $(\mathbf{Y}-\mathbf{X\tau})(\mathbf{Y}-\mathbf{X\tau})^T$, $\mathbf{V}'$ is the covariance matrix of $\mathbf{Y}'$ and $\mathbf{X}'$ is the design matrix linking $\mathbf{Y}'$ to $\mathbf{V}'$ in the regression of $\mathbf{Y}'$ on $\mathbf{X}'$. If both $\boldsymbol{\tau}$ and $\mathbf{V}$ are unknown then the estimates of $\boldsymbol{\tau}$ can be obtained using iterative generalized least square procedure. The procedure starts by initially assuming an estimate of $\mathbf{V}$ and by using this estimate of $\mathbf{V}$ one can obtain the values of $\boldsymbol{\tau}$ and then get the better estimate of $\mathbf{V}$ by using $\boldsymbol{\tau}$. This process continues simultaneously and iteratively until both the estimates converge. A useful initial start of the process is by assuming $\mathbf{V} = \sigma^2 I$, where I_{m_1} is an identity matrix. The initial estimates of $\boldsymbol{\tau}$ estimated by assuming $\mathbf{V} = \sigma^2 I$ are asymptotically consistent. If we assume the initial estimate of $\mathbf{V_1} = \mathbf{V_{k(1)}} = \sigma^2 I_{m_1}$, then the initial consistent estimator of $\boldsymbol{\tau}$ is $\hat{\boldsymbol{\tau}} = (\mathbf{X}^T\mathbf{V}_1^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}_1^{-1}\mathbf{Y}$. A more useful estimate $\mathbf{V}$ of can be obtained by incorporating the variation due to all levels as,

$$\mathbf{V}_2 = \mathbf{V_{k(2)}} = \bigoplus_{i_2}^{m_2}\left\{\sigma_1^2 I_{m_2} + \sigma_2^2 I_{m_2}\right\} = \bigoplus_{i_2}^{m_2}\left\{\sigma_1^2 I_{m_2} + \sigma_2^2 J_{m_2\times 1} I_{(1)} J_{1\times m_2}\right\} \tag{2.15}$$

$$\mathbf{V}_3 = \mathbf{V_{k(3)}} = \mathbf{V}_2 + \sigma_3^2 J_{m_3\times 1} I_{(1)} J_{1\times m_3} \tag{2.16}$$

$$\mathbf{V}_4 = \mathbf{V_{k(4)}} = \mathbf{V}_3 + \sigma_4^2 J_{m_4\times 1} I_{(1)} J_{1\times m_4} \tag{2.17}$$

and so on,

$$\mathbf{V}_k = \mathbf{V_{k(\mathit{k})}} = \mathbf{V}_{k-1} + \sigma_k^2 J_{m_k\times 1} I_{(1)} J_{1\times m_k} \tag{2.18}$$

The inverse of $(2.15), (2.16), (2.17)$ and (2.18) are obtained by using the relation (2.19):

$$\left(X + YZY^T\right)^{-1} = X^{-1} - X^{-1}YZ\left(I + Y^T X^{-1}YZ\right)^{-1} Y^T X^{-1} \tag{2.19}$$

$$\mathbf{V_2^{-1}} = \bigoplus_{i_2}^{m_2}\sigma^{-2}\left\{I_{m_2} - \sigma_2^2(m_2\sigma_2^2 + \sigma_1^2)^{-1} J_{m_2}\right\} \tag{2.20}$$

$$\mathbf{V}_3^{-1} = \bigoplus_{i_3}^{m_3}\left[\mathbf{V}_2^{-1} - \mathbf{V}_2^{-1} J_{m_3\times 1}\left\{\sigma_3^{-2} + J_{1\times m_3}\mathbf{V}_2^{-1} J_{m_3\times 1}\right\}^{-1}\right] J_{1\times m_3}\mathbf{V}_2^{-1} \tag{2.21}$$

$$\mathbf{V}_4^{-1} = \bigoplus_{i_4}^{m_4}\left[\mathbf{V}_3^{-1} - \mathbf{V}_3^{-1}J_{m_4\times 1}\left\{\sigma_4^{-2} + J_{1\times m_4}\mathbf{V}_3^{-1}J_{m_4\times 1}\right\}^{-1}\right]J_{1\times m_4}\mathbf{V}_3^{-1} \tag{2.22}$$

and so on,

$$\mathbf{V}_k^{-1} = \bigoplus_{i_k}^{m_k}\left[\mathbf{V}_{k-1}^{-1} - \mathbf{V}_{k-1}^{-1}J_{m_k\times 1}\left\{\sigma_k^{-2} + J_{1\times m_k}\mathbf{V}_{k-1}^{-1}J_{m_k\times 1}\right\}^{-1}\right]J_{1\times m_k}\mathbf{V}_{k-1}^{-1} \tag{2.23}$$

where $I_{(.)}$ is an identity matrix and $J_{(.)}$ is a matrix of ones. Furthermore, $\oplus$ is a direct sum operator. Thus consistent estimates of $\boldsymbol{\tau}$ can be obtained by using $\mathbf{V}$ in $(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$.

We estimate the random intercept model for single level, 2-levels, 3-levels, 4-levels and 5-levels by using a higher education data collected from five universities in Pakistan. The educational structure in universities in Pakistan is hierarchical where students are nested within teachers, teachers nested within directors/chairpersons, directors/chairpersons nested within deans, deans nested within universities and the universities. The response variable is student grade point average (GPA) score collected from 40000 students registered in five universities from Pakistan. Random intercept models defined across Level-1, Level-2, Level-3, Level-4 & Level-5 are as follows:

$$\left. GPA_i = \tau_0 + \varepsilon_i \right\} \tag{2.24}$$

$$\left.\begin{aligned} GPA_{ij} &= \tau_{0j} + \varepsilon_{ij} \\ \tau_{0j} &= \tau_0 + u_{0j} \end{aligned}\right\} \tag{2.25}$$

$$\left.\begin{aligned} GPA_{ijk} &= \tau_{0jk} + \varepsilon_{ijk} \\ \tau_{0jk} &= \tau_0 + u_{0jk} + v_k \end{aligned}\right\} \tag{2.26}$$

$$\left.\begin{aligned} GPA_{ijkl} &= \tau_{0jkl} + \varepsilon_{ijkl} \\ \tau_{0jkl} &= \tau_0 + u_{0jkl} + v_{kl} + w_l \end{aligned}\right\} \tag{2.27}$$

and,

$$\left.\begin{aligned} GPA_{ijklm} &= \tau_{0jklm} + \varepsilon_{ijklm} \\ \tau_{0jklm} &= \tau_0 + u_{0jklm} + v_{klm} + w_{lm} + g_m \end{aligned}\right\} \tag{2.28}$$

where, where, τ_0 is the average intercept and the terms $\varepsilon_{ijklm}, u_{jklm}, v_{klm}, w_{lm}$ & g_m are the residuals due to random variations of Level-1,..., Level-5 units respectively under model (2.24), (2.25), (2.26), (2.27) and (2.28) and represent the random departure from the average intercept τ_0. Also $\varepsilon_{ijklm} : NII(0,\sigma_{e0}^2)$, $u_{jklm} : NII(0,\sigma_{u0}^2)$, $v_{klm} : NII(0,\sigma_{v0}^2)$, $w_{lm} : NII(0,\sigma_{w0}^2)$, and $g_m : NII(0,\sigma_{g0}^2)$. The iterative generalized least square estimates of the parameters under (2.24), (2.25), (2.26), (2.27) and (2.28) are presented in table 1.

Table 1: Estimates of Parameters under model $(2.24), (2.25), (2.26), (2.27)$ and (2.28)

Parameters	1-Level Model	2-Levels Model	3-Levels Model	4-Levels Model	5-Levels Model
τ_0 : Average Intercept.	3.049 * (0.001)	3.049 * (0.006)	3.049 * (0.012)	3.049 * (0.027)	3.049 * (0.072)
σ^2_{g0} : Intercept Variation at Level-5	-	-	-	-	0.025 * (0.012)
σ^2_{w0} : Intercept Variation at Level-4	-	-	-	0.027 * (0.007)	0.004 * (0.001)
σ^2_{v0} : Intercept Variation at Level-3	-	-	0.029 * (0.003)	0.002 * (0.00052)	0.00097 * (0.00013)
σ^2_{u0} : Intercept Variation at Level-2	-	0.030 * (0.001)	0.001 * (0.0005)	0.001 * (0.00047)	0.001 * (0.00054)
σ^2_{e0} : Variation at Level-1.	0.073 * (0.001)	0.043 * (0.00028)	0.043 * (0.00028)	0.043 * (0.00029)	0.043 * (0.00028)
-2*loglikelihood	-8705.8	-9282.9	-11272.5	-11899.9	-11959.9

* p<0.01

The value of average intercept $\tau_0 = 3.049$ remains unchanged under $(2.21), (2.22), (2.23)$, (2.24) and (2.25) but its significance reduces by considering higher levels of hierarchy in the model (2.21). The value of random variation σ^2_{e0} at higher levels is marginally reduced to $0.043(se = 0.00028, p < 0.001)$ from individual level variation $0.043(se = 0.00028,\ p < 0.001)$. The random variation at individual level is split into two portions when considering second level of hierarchy in the model. One is at individual level $\sigma^2_{e0} = 0.043$ and the other one is due to second level $\sigma^2_{u0} = 0.030$ .This means 41.1% of the random variation in students' grades is due to the variation among teachers. Similarly, the variation at second level $\sigma^2_{u0} = 0.030$ is further subdivided into two parts, one is due to second level $\sigma^2_{u0} = 0.001$ and the other one is by taking third level $\sigma^2_{v0} = 0.029$ of hierarchy in the model. Again, 96.7% of the variation in Level-2 units is due to incorporating third level in the model. This suggests the contribution of department level units in the random variation of grades is more when compared to variation due to teachers. Again when we included the fourth level in the model (2.23) , the variation at part of third level $\sigma^2_{v0} = 0.029$ is subdivided into the variation due to third level $\sigma^2_{v0} = 0.002$ and the variation due to fourth level $\sigma^2_{w0} = 0.027$. This information tells the fact that students grades vary faculty (Dean) wise and their impact is comparatively more highlighted as compare to variation due to department. Finally, a major contribution in random variation at Level-4 is due to Level-5 units and students grades vary from university to university.

Conclusions

The random variation in intercepts at individual level is marginally divided into higher levels when consider higher levels of hierarchy in the single level model. So, one can control the random variation in intercepts by incorporating the higher levels in the model. The decreasing trend of -2*loglikelihood confirms the better fit of the more nested model.

References

1. De Leeuw, J. and Kreft, Ita G.G. (1986). Random Coefficient Models. Journal of Educational Statistics, 11, 1, 55-85.
2. Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalised least squares. *Biometrika*, 73, 43-56.
3. Goldstein, H. (1995). *Multilevel Statistical Models*, 2nd edition. London, Edward Arnold: New York, Wiley.
4. Khan, R. A. and Kamal, S. (2012). Random Intercept and Random Slope 2-Levels Multilevel Models. *PJSOR*, 8(4), pp. 777-788.
5. Khan, R. A. and Kamal, S. (2012b). Random Intercept and Random Slope 2-Levels Multivariate Multilevel Models. *Science Series Data Record*, 4(10), pp. 1-11.
6. Khan, R. A. and Kamal, S. (2013). Random Intercept and Random Slope 3-Levels Multilevel Models. *CJASR*, Accepted for publication.
7. Longford, N.T. (1992). *Random Coefficient Models*. Oxford: Clarendon Press.
8. Moulton, B. R. (1986). Random Group Effects and the Precision of Regression Estimates. *Journal of Econometrics*, 32, 3, 385-397.
9. Raudenbush, S.W. and Bryk, A.S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 12, 241-269.
10. Raudenbush, S.W. and Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd ed. Newbury Park, CA: Sage.
11. Snijders, T and Bosker, R. (1999). *Multilevel Analysis*. London, Sage.