Indian Journal of Fundamental and Applied Life Sciences ISSN: 2231– 6345 (Online)
An Open Access, Online International Journal Available at www.cibtech.org/sp.ed/jls/2015/02/jls.htm
2015 Vol. 5 (S2), pp. 1432-1448/Vajargah and Nikbakht
Research Article

APPLICATION REMLMODEL AND DETERMINING CUT OFF OF ICC BY MULTI-LEVEL MODEL BASED ON MARKOV CHAINS SIMULATION IN HEALTH

***Kianoush Fathi Vajargah and Masoomehnikbakht**
Department of Statistics Islamic Azad University, North branch Tehran, Iran
**Author for Correspondence*

ABSTRACT
Multi-level analysis is an efficient method for analysis of the data in more than one levels and is the extended state of generalized linear models in which besides varied modeling, regression coefficients response is also modeled. Regarding continuous data in which dependence of dependent variables to independence variables is observed, to define which regression models or multi-levels are used, we can consider ICC (Intra-class correlation coefficient). In this study, we can determine ICC cut off point to use multi-level model of the data in which leveling nature is seen or REML method.

Keywords: *Markov Chain, Multi-level Models, REML Method, ICC*

INTRODUCTION
Many kinds of data, including observational data collected in the human and biological sciences, have a hierarchical or clustered structure. For example, animal and human studies of inheritance deal with a natural hierarchy where offspring are grouped within families. Offspring from the same parents tend to be more alike in their physical and mental characteristics than individuals chosen at random from the population at large. For instance, children from the same family may all tend to be small, perhaps because their parents are small or because of a common impoverished environment (Goldstein, 2010).
Such data are called nested data or hierarchy structure data. Although statistical analysis of such data is possible by common linear regression method, as one of the basic assumptions of simple regression is statistical independence of all observations, the problem of this assumption causes that standard error of model parameters estimation is estimated less potentially (Cohen, 1998).
Based on the above example, if independence assumption is not established between the observations, using regression common models has problems (Pinherio and Bates, 2000).
The suitable model of the analysis of above example data is multi-level model.
Linear mixed-effects models with nested grouping factors, generally called multilevel models (Goldstein, 1995) or hierarchical linear models (Bryk and Raudenbush, 1992.)
This model is called by other equal names in statistical articles and books as random effects, panel models, nested models and cluster models (Look, 2004).
The main feature of multi-level data is their leveling feature. Normally, the studied groups are selected as random and besides the error of measuring the intra-observation in each group, another error of sampling the groups is involved in multi-level data analysis. The traditional methods of regression models ignore this second error. In addition, we can refer to the lack of generalizing the results of leveling to total group and the lack of detecting variability of the group as other disadvantages of common regression models. Multi-level analysis models can eliminate these problems. By multi-level analysis models, we can estimate many parameters and reduce measurement error. The important point is that the parameter estimation method in multi-level model is different from estimation method in simple regression. Many data including het data of some of observations of human and biological sciences have cluster or hierarchy structure. The offspring of a family are much similar compared to other people in society form physical and mental features (e.g. the children are considered as level 1 and families as level 2). From statistical aspect, people similarity to each other indicates the lack of independence of data. Two people belonging to a father and mother are similar to each other from many conditions (Goldstain, 1987).

Research Article

As another example, if the child of a family has special disease as hepatitis, his brothers and sisters can have this problem or in the studies of the repeated measures on effective variables on blood pressure, blood pressure of samples can be measured in various times. In the blood pressure of each person, due to some conditions, some changes are made over time. In this example, the set of measures in various times is for each of the members of first level units as defined for each unit of second level. Regarding the justification of the fact as why in repetitive sizes, it is considered as level 1 in longitudinal studies, we can consider the correlation of repetitive observations as specific belonging to a person. These correlations are observed in cross section plans in nested structure (Hox, 2002).

Intra-class correlation coefficient (ICC) is a type of correlation coefficient expressing the agreement between some sets of data.

In multi-level modeling, this coefficient indicates the ratio of intergroup variance to total variance and the higher this ratio, we can say multilevel and grouping between the data are significant and using multi-level modeling is preferred to simple regression (Luke, 2004).

The comparison of estimators of multi-level model and equivalent estimator in regression model in various ICCs can show from which level of ICC, the estimator of multi-level model is preferred to regression.

This study attempts to show that by data simulation with pre-determined ICC and definite parameters by REML method in regression and multi-level case (random intercept) fitting is done for data and by comparison of AIC and Loglik mean and the mean of absolute value of models coefficients with pre-defined coefficients, we can define after which ICC , multi-level model acts better than regression model and we can achieve acceptable approximation of ICC for cutoff point of using multi-level model.

The history of multi-level models.

In the mid 1980, some researchers presented systematic views for modeling and analysis of the data with multi-level structure.

The first works of Aitkin *et al.,* (1980) on the relevant data of learning method and then Longford common works (1968) founded the basis of some progresses in this regard and the result of these progresses showed that in the early 1990, the main core of such models was formed. Goldstein (1987) wrote a book “Multilevel Models in Educational and Social Research”. Brky and Raudenbush (1992) developed linear models 2, 3 levels and their application on repetitive measures. Later, Longford (1993) extended multi-level models theory for factor analysis model of batch or classified responses and multi-variable models.

These models are used widely in analysis of various academic, educational, epidemiology, children growth, household investigations and the data with hierarchy structure. Leonardo and Carla (2005) introduced ICC=0.05 as an idea for threshold in using Multilevel factor models for ordinal variables with random intercept for discrete ordinal variables.

Explanation of Multi-level Modeling

For accurate analysis of the data needing multi-level modeling with nested structure, we need multi-level modeling. Multi-level analysis (nested) is extended case of generalized linear models in which besides modeling, regression coefficients response is also modeled.

The aim of multi-level models is modeling dependent variable based on a function of predictive variables (independent) in more than one levels. Multi-level analysis is also recognized by other names as Hierarchical linear models, mixed coefficients models and Random coefficients models.

Suppose, a researcher attempts to observe and test the relationship between the score of literature with two variables of the number of student studying hours (variable level 1) and class size (level 2 variable). He applies two-level model for modeling the data. Equation (1) indicates the modeling of response variable by him.

$$\text{level 1}: y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij} \quad (1)$$

$$\text{level 2}: \beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$$

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

$$U=\begin{bmatrix}u_{0j}\\u_{1j}\end{bmatrix}\approx N(o,G),\ G=\begin{bmatrix}\sigma_{u0}^{2} & .\\ \sigma_{01} & \sigma_{u1}^{2}\end{bmatrix} \qquad e_{ij}\approx N(0,R) \qquad (2)$$

Cov(U,e)=0

These equations not only show the predictive and dependent variables, but also they draw the multi-level nature of the model. In equation (1) level 1 is like an ordinary regression model but index j shows that level 1 for each of levels j of classroom (level 2 of study) is a different level 1 model as estimated and it means that each class has different intercepts (β_{0j}s) and slope of lines and the different effects of study time on score (β_{1j}s) are different.

Part 2 of equation (1) (level 2) indicates the relationship of level 1 parameters with level 2 variables. Instead of using such equations to define multi-level models, we can replace some parts of level 2 of model inside equation level 1. After placement and ordering the terms, we achieve equation 3.

$$y_{ij}=[\gamma_{00}+\gamma_{10}X_{ij}+\gamma_{01}w_j+\gamma_{11}w_jX_{ij}]+[u_{0j}+u_{1j}X_{ij}+e_{ij}] \qquad (3)$$

Fixed part | Random part

$$\Gamma=(\gamma_{00}+\gamma_{01}+\gamma_{10}+\gamma_{11})$$

$$u=(u_{0j},u_{1j})$$

Table 1: Different multi-level models

Class name	Multi-level equations system	Random equations system
Unconstrained	L1: $y_{ij}=\beta_{0j}+e_{ij}$	$\gamma_{ij}=\gamma_{00}+u_{0j}+e_{ij}$
	L2: $\beta_{0j}=\gamma_{00}+u_{0j}$	
Random intercepts	L1: $y_{ij}=\beta_{.j}+e_{ij}$	$y_{ij}=\gamma_{00}+\gamma_{01}w_j+u_{0j}+e_{ij}$
	L2: $\beta_{.j}=\gamma_{..}+\gamma_{.1}w_j+u_j$	
	L1: $y_{ij}=\beta_{0j}+\beta_{1j}X_{ij}+e_{ij}$	$y_{ij}=\gamma_{00}+\gamma_{10}X_{ij}+u_{0j}+e_{ij}$
	L2: $\beta_{0j}=\gamma_{00}+u_{0j}$	
	$\beta_{1j}=\gamma_{10}$	
random intercepts & slopes	L1: $y_{ij}=\beta_{0j}+\beta_{1j}X_{ij}+e_{ij}$	$y_{ij}=\gamma_{00}+\gamma_{10}X_{ij}+u_{0j}+u_{ij}X_{ij}+e_{ij}$
	L2: $\beta_{0j}=\gamma_{00}+u_{0j}$	
	$\beta_{1j}=\gamma_{10}+u_{1j}$	
	L1: $y_{ij}=\beta_{0j}+\beta_{1j}X_{ij}+e_{ij}$	$y_{ij}=\gamma_{00}+\gamma_{01}w_j+\gamma_{10}X_{ij}\gamma_{11}w_jX_{ij}+u_{0j}+u_{1j}X_{ij}+e_{ij}$
	L1: $\beta_{0j}=\gamma_{00}+\gamma_{01}w_j+u_{0j}$	
	$\beta_{1j}=\gamma_{10}+\gamma_{11}w_j+u_{1j}$	

The first parenthesis is called constant and the second parenthesis is called random part of model.

Now we can rewrite equation (3) as equation (4).

$$y=X\alpha+Z\beta+e \qquad (4)$$

Where, X is independent observations matrix for fixed effects Z as matrix of independent observations for random effects, α is fixed coefficients vector, β is random coefficients vector of model and finally e is model error.

Equation (3), (4) can be written as equation 5.

$$y_{ij} = \begin{bmatrix}1 & X_{ij} & W_{ij} & X_{ij}W_j\end{bmatrix} \times \begin{bmatrix}\gamma_{00}\\ \gamma_{10}\\ \gamma_{01}\\ \gamma_{11}\end{bmatrix} + \begin{bmatrix}1 & X_{ij}\end{bmatrix} \times \begin{bmatrix}u_{0j}\\ u_{1j}\end{bmatrix} + e_{ij} \quad (5)$$

$$X = \begin{bmatrix}1 & X_{ij} & W_{ij} & X_{ij}W_j\end{bmatrix}$$

$$Z = \begin{bmatrix}1 & X_{ij}\end{bmatrix}$$

$$\alpha = \begin{bmatrix}\gamma_{00}\\ \gamma_{10}\\ \gamma_{01}\\ \gamma_{11}\end{bmatrix} \quad , \quad \beta = \begin{bmatrix}u_{0j}\\ u_{1j}\end{bmatrix}$$

The major advantage of displaying multi-level equations as equation 5 is as this form and by normal distribution statistic features; we can estimate fixed, random coefficients and model variance components.
Another important issue is that multi-level models are classified into three general classes and these three classes in (Table 1) are presented for two-level models (Luke, 2004).

Model Fitting Methods

There are various estimation methods to estimate model coefficients in multi-level modeling. One of these methods is Maximum Likelihood Estimator (MLE), Residual Maximum Likelihood Estimation (REMLE) and it is called constrained liklihood method in some books. Indeed, it is MLE method in which

Likelihood function is modified. We can refer to Iterative Generalized Least Squares (IGLS), Restricted Iterative Generalized Least Squares (RIGLS), Bayesian methods, Monte Carlo methods (MCMC) (McCulloh, 2001; Brown 2006).

Maximum Likelihood Estimator Method (MLE)

It is a method to estimate the parameters and it was developed for the first time by Fisher (1922) and it is used for many years.
This method is used as basis method for big cases of sample of hypotheses test and confidence interval in analysis of time series data.
This estimation method is used in various grounds as survival analysis, regression analysis, spatial data analysis, variance component and etc.
In multi-level models, we are interested in estimation of three types of parameters, fixed coefficients of model, matrix G and error variance matrix of model R.
Based on equation 4, the feature of normal distribution and expected value we have:

$$E(y) = E(X\alpha + Z\beta + e) = E(X\alpha) + E(Z\beta) + E(e) = X\alpha \quad (6)$$

$$V(y) = V(X\alpha + Z\beta + e) = V(X\alpha) + V(Z\beta) + V(e) = ZV(\beta)Z' + V(e) \Longrightarrow$$

$$V(y) = ZGZ' + \Sigma \quad (7)$$

Thus, Y is defined as:

$$Y_i \sim N\left(X\alpha, ZGZ' + \Sigma\right) \quad (8)$$

Let γ is the vector of covariance variance parameters as found in $ZGZ' + \Sigma$. In other words, γ includes q(q+1)/2 different elements in matrix G and all Σ parameters.
By placement of values achieved in equations 6, 7 and its placing in normal distribution we have:

$$L_{ML}(\gamma, \alpha) = \frac{\exp\left[-\frac{1}{2}(Y - X\alpha)' V^{-1} (Y - X\alpha)\right]}{(2\pi)^{\left(\frac{1}{2}\right)n} |V|^{\frac{1}{2}}}. \quad (9)$$

By logarithm and derivation of equation $L_{ML}(\gamma, \alpha)$, by repetitive methods or *Newton–Raphson* method, we can use in solution of maximum likelihood equations in which most of statistical software are calculated. Thus, $\hat{\alpha}$ is achieved as: (Wulu, 1999; Myers, 2002; Verbeke, 2000).

$$\hat{\alpha}(\gamma) = \left(\sum_{i=1}^{N} X_i' V_i^{-1} X_i\right)^{-1} \sum_{i=1}^{N} X_i' V_i^{-1} y_i \quad (10)$$

Residual Maximum Likelihood (REML) Method

This method was raised for the first time by Thomposon (1962) and was explained by Peterson and Thompson (1974). By explanation of this method, Y should have a normal multi-variate distribution (Jiang, 2007).

Estimation of Residual Variance in Linear Regression

As a second example, we now consider the estimation of the residual variance σ^2in linear regretion model $Y = X\beta + \varepsilon$ where Y is an N dimensional vector, and with X a $(N \times p)$matrix with known covariate value. It is assumed that all elements in ε are independently normally distributed with mean zero and variance σ^2. The MLE for σ^2 equals

$$\widehat{\sigma^2} = \left(Y - X(X'X)^{-1}X'Y\right)'(Y - X(X'X)^{-1}X'Y)/N \quad (11)$$

Which can easily be shown to be biased downward by a factor (N-p)/N.

σ^2can be estimated using a set of error contrasts $U = A'y$ where A is now any $N \times (N-p)$ matrix with $N - p$ Linearly independent columns orthogonal to the columns of the design matrix X. We then have that U follows a normal distribution with mean vector 0 and covariance matrix $\sigma^2 A'A$, in which σ^2 is again the only unknown parameter. Maximizing the corresponding likelihood with respect to σ^2 yields

$$\widehat{\sigma^2} = \left(Y - X(X'X)^{-1}X'Y\right)'(Y - X(X'X)^{-1}X'Y)/(N-p) \quad (12)$$

Which is the mean squared error, unbiased for σ^2, and classically used as estimator for the residual variance in linear regression analysis (see, for example, Neter *et al.*, 1990; Seber, 1977) we again have that any matrix A satisfying the specified conditions leads to the same estimator for the residual variance, which is again called the REML estimator for σ^2 (Verbeke, 2000).

REML Estimation for the Linear Mixed Model

The drawback of MLE method to estimate covariance components is that this method doesn't involve a part of degree of freedom lost due to the estimation of coefficients vector in estimation of covariance components and biased estimations are generated mostly.

The true example is the estimation of variance of multi-variate normal distribution. Let $y \sim MVN(\beta, \sigma^2 I)$, thus, ML estimation of parameter σ^2 is $\widehat{\sigma^2} = RSS/N$ in which RSS is the sum of squares of error. The above estimator is biased one. However, non-biased estimator σ^2 is equal to $\widehat{\sigma^2} = RSS/N - P$ in which P is the number of elements of β vector. This estimator is defined as ML estimator as a linear combination of observations ($U = A'y$). The distribution of converted observations is not dependent upon parameter α. A is not unique and any matrix true in $A^T X = 0$ can be used for this conversion (Jiang, 2007).

Estimation for the Linear Mixed Model

In practice, linear mixed models often contain many fixed effects. In such cases, it may be important to estimate the variance components, explicitly taking into account the loss of the degrees of freedom involved in estimating the fixed effects. In contrast to the simple cases, an unbiased estimator for the vector α of variance components cannot be obtained from appropriately transforming the ML estimator as suggested from the analytic calculation of its bias. However, the error contrasts approach can still be applied as follow. We first combine all N subject-specific regression models to one model:

$$y = X\alpha + Z\beta + \varepsilon \quad (13)$$

where the vectors Y,β and ε, and the matrix X are obtained from stacking the vectors Y_i, underneath each other, and where Z is the block-diagonal matrix with block $Z_i$on the main diagonal and zeroes elsewhere. The dimension of Y equals $\sum_{i=1}^{N} n_i$ and will be denoted by n.

The marginal distribution for Y is normal with mean vector $X\alpha$ and with covariance matrix $V(\gamma)$ equal to the block-diagonal matrix with blocks V_i on the main diagonal and zeros elsewhere. The REML estimator

Research Article

for the variance components γ is now obtained from maximizing the likelihood function of a set of error contrasts $U = A'Y$ where A is any (n$\times$ $(n-p)$) full-rank matrix with columns orthogonal to the columns of the X matrix. The vector U then follows a normal distribution with mean vector zero and covariance matrix $A'V(\gamma)A$, which is not dependent on α any longer. Further, Harville (1974) has shown that the likelihood function of the error contrast can be written as

$$L(\alpha) = (2\pi)^{-(n-p)/2}\left|\sum_{i=1}^{N} X_i' X_i\right|^{1/2} \tag{14}$$

$$\times \left|\sum_{i=1}^{N} X_i' V_i^{-1} X_i\right|^{-1/2} \prod_{i=1}^{N} |V_i|^{-1/2}$$

$$\times exp\left\{-\frac{1}{2}\sum_{i=1}^{N}(Y_i - X_i\hat{\alpha})' V_i^{-1}(Y_i - X_i\hat{\alpha})\right\}$$

Where $\hat{\alpha}$ is given by (10). Hence, REML estimator $\hat{\alpha}$ does not depend on error constant (i.e. the choise of A).

Note that the maximum likelihood estimator for the mean of a univariate normal population and for the vector of regression parameters in linear regression model are independent of the residual variance σ^2.

Finally note that the likelihood function in equals

$$L(\gamma) = C\left|\sum_{i=1}^{N} X_i' V_i^{-1} X_i\right|^{-1/2} L_{ML}(\gamma, \alpha) \tag{15}$$

Where C is constant not dependent on γ.

Because $\left|\sum_{i=1}^{N} X_i' V_i^{-1} X_i\right|$ does not depend on α it follows that the REML estimator for α and γ can also be found by maximizing the so-called REML likelihood function (Verbeke, 2000)

$$L_{REML}(\gamma, \alpha) = C\left|\sum_{i=1}^{N} X_i' V_i^{-1} X_i\right|^{-1/2} L_{ML}(\gamma, \alpha). \tag{16}$$

Parameters Test Methods and Model Adequacy Indices

Likelihood Ratio Test

Log-likelihood statistics is a method to test the significance H_0: β=0 regarding regression model parameters and this statistics is calculated by dividing likelihood function of H0 (l0) by likelihood function in H1 (I1) as:

$$\Lambda = -2\mathrm{Log}\left(\frac{l_0}{l_1}\right) = -2[\mathrm{Log}(l_0) - \mathrm{Log}(l_1)] = -2(l_0 - l_1) \tag{17}$$

Where $l_0$andl_1 show likelihood functions logarithm. This statistics in H0 has chi-square distribution with a degree of freedom (McCulloh, 2001).

-2log(L) value is used to compare two models. The model in which Λ is more and it is suitable model.

Akaike Information Criterion (AIC)

It is a criterion to evaluate goodness of fit. This criterion is based on entropy concept and it shows that using a statistical model can lose information. In other words, this criterion can establish a balance between accuracy of model and its complexity. This criterion is prposed by HirotesogoAkaike to select the best statistical model (Akaike, 1974).

Based on the data, some competing models can be ranked based on AIC value and the model has the lowest AIC as best.

The output of the summary function includes the values of the Akaike Information Criterion (AIC) (Sakamoto *et al.,* 1986) and the Bayesian Information Criterion (BIC) (Schwarz, 1978), which is also sometimes called Schwarz's Bayesian Criterion (SBC). These are model comparison criteria evaluated as

AIC = − 2 log Lik + 2n par , (18)

BIC = − 2 log Lik + n par log(N), (19)

where n par denotes the number of parameters in the model and N the total number of observations used to fit the model. Under these definitions, "smaller is better." That is, if we are using AIC to compare two or more models for the same data, we prefer the model with the lowest AIC.
Similarly, when using BIC we prefer the model with the lowest BIC (Pinherio and Bates, 2000)

Intra Class Correlation (ICC)

ICC is one of the best adequacy indices and model selection in multi-level data.
ICC is an index of correlation of intra cluster data and it is the intercluster variance ratio to total existing variance

$$\mathrm{ICC} = \frac{\sigma_{u0}^2}{\sigma_{u0}^2+\sigma_e^2} \tag{20}$$

This index reports correlation degree of data. Although numerical value of this index is close to one, it means that multi-level modeling is a suitable model to analyze the existing data and multi-level analysis can present better results compared to simple regression (Luke, 2004).
ICC value close to zero indicates that using simple regression method and multi-level analysis method has similar efficiency.
To prove that why ICC is called intra-cluster correlation coefficient, simple two –level model with equation 16 is considered.

$$y_{ij} = \gamma_{00} + U_{0j} + \varepsilon_{ij} \tag{21}$$

The variance of ith in cluster jth based on variance formula is as:

$$V(y_{ij}) = V(U_{0j}) + V(\varepsilon_{ij}) = \sigma_{u0}^2 + \sigma_e^2 \tag{22}$$

The covariance of ith and i' in cluster jth based on covariance formula is as

$$\mathrm{Cov}\left(y_{ij}, y_{i'j}\right) = \mathrm{Var}(U_{0j}) = \sigma_{u0}^2 \tag{23}$$

The correlation value of ith in cluster jth and based on correlation coefficient formula is equal to [6]:

$$\rho = \frac{\mathrm{Cov}\left(y_{ij}, y_{i'j}\right)}{\sqrt{V(y_{ij})}.\sqrt{V(y_{i'j})}} = \frac{\sigma_{u0}^2}{\sigma_{u0}^2+\sigma_e^2} \tag{24}$$

Simulation Algorithm

Simulation process for random intercept model with N=80 (number of samples), J=10 (the number of repetitions in each sample), b0=1.7 (γ_{00} in multilevel model as intercept of model) and b1=4 (γ_{10} value in multi-level model as slope of line in model) can be considered.
Sample n=80, J=10 is optimum sample volume in two-level model (Akhgari, 2013).
1- Based on relation σ_u^2 ،σ_e^2 and ICC by determining ICC and one of the variances, another value is determined.

$$\mathrm{ICC} = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2}$$

As it can be said

$$\sigma_{u0}^2 = \frac{ICC}{1 - \mathrm{ICC}}\sigma_e^2$$

By considering values for ICC and $\sigma_e^2 = 5$, σ_u^2 values are determined.
2-Simulation of u_{0j} values: The data are simulated as n times of distribution N(0,σ_{u0}^2) and each time, for J times, these data are repeated (for each set of data, u_{0j} value is considered and values u_{0j} are similar for the data of a set of values).
3- For simulation X_{ij} as σ_x^2 is 16 and data of X_{ij} as $N = n \times j$ of distribution N(0,σ_x^2) can be simulated.
4- For simulation e_{ij}, the data as $N = n \times j$ are simulated as N(0,σ_e^2) distribution.
5- Values Y_{ij} are simulated as followings

$Y_{ij} = b_0 + b_1.X_{ij} + u_{0j} + e_{ij}$

In classic case in 12 values of ICC ranging from 0.01 to 0.9 and in each ICC as 1000 times data are generated and each time, two regression and multi-level models by REML method are fitted on data and coefficients of model $b_0$andb_1 and AIC and Loglik values and R2 value of regression model are registered and in each fitting time , absolute value of coefficients deviation of real value of two models with assumption of model coefficients or biased absolute value of effect size can be computed.
By considering three criteria comparison models, we can compare two models in various ICCs.
The mean values of AIC and Loglik and biased absolute value of effective size of two models are compared with each other. Also, average R^2 in regression model is computed as a criterion to show the efficiency of regression model in each model. To compare the biasedness of effect size of two models, percent of growth coefficient of effect size in model change of regression to multi-level model in each ICC can be computed and compared. When multi-level model is better than regression, this value is bigger than zero and the growth of this value shows the superiority of multi-level model to regression model.
The formula of effect size growth coefficient percentage in this study is as:

$$\left(\frac{\text{Absolute vlue of effect size of multi-level model}}{\text{Absolute value of regression model effect size}} - 1\right) \times 100 \quad (25)$$

In the results of Bayesian approach, the results of simulation by Gibbs algorithm is for random intercept model. The number of simulations based on 100000 times sampling with burn-in phase 50000 times. Then, of 50000 residual Gibbs samples, of each 50, one is selected. Thus, to estimate model parameters, 1000 samples are available.

The Investigation of the Results in Classic Case

Table 4-1 shows 12 values of various ICCs ranging 0.01-0.9 with average intercept (b0) and slope (b1) and two criteria AIC and log likelihood in two regression and multi-level models each with 1000 times simulations.

Table 2: The general results in classic case

ICC	Regression model b_0	Regression model b_1	Regression model AIC	Regression model Loglik	R^2	Multi-level model b_0	Multi-level model b_1	Multi-level model AIC	Multi-level model Loglik
0.01	1.713487	4.000034	3570.369	-1782.25	0.981	1.713468	3.999893	3571.529	-1781.76
0.03	1.697784	4.00227	3603.919	-1798.86	0.980	1.697829	4.002233	3601.964	-1796.98
0.05	1.676696	3.999154	3608.99	-1801.09	0.980	1.676746	3.998981	3603.415	-1797.72
0.055	1.737523	3.998667	3611.946	-1802.97	0.980	1.737539	3.998872	3604.535	-1798.27
0.06	1.696821	4.003567	3621.294	-1807.66	0.979	1.696757	4.003719	3612.89	-1801.95
0.065	1.704008	3.998016	3622.76	-1809.57	0.980	1.703984	3.998372	3609.651	-1801.82
0.1	1.686862	4.001383	3657.287	-1824.74	0.979	1.686848	4.001221	3632.272	-1812.14
0.2	1.725452	3.989761	3753.08	-1874.54	0.976	1.725355	3.989578	3677.012	-1834.5
0.3	1.669395	3.998506	3856.562	-1925.78	0.973	1.669386	3.998595	3710.697	-1851.37
0.5	1.681176	3.999091	4109.402	-2051.1	0.963	1.681264	4.000779	3752.868	-1872.42
0.7	1.76127	4.000235	4534.57	-2264.28	0.938	1.76113	3.996224	3824.399	-1908.2
0.9	1.664998	3.998298	5385.125	-2679.56	0.839	1.664029	3.995434	3821.926	-1959.32

Table 4-2 shows the biased mean of effect size of regression model, multi-level model and biased growth coefficient of effect size of regression to multi-level in 12 levels of ICC.

Indian Journal of Fundamental and Applied Life Sciences ISSN: 2231– 6345 (Online)
An Open Access, Online International Journal Available at www.cibtech.org/sp.ed/jls/2015/02/jls.htm
2015 Vol. 5 (S2), pp. 1432-1448/Vajargah and Nikbakht
Research Article

Table 3: The results in classic case 2

ICC	Biasedness of effect size of regression model coefficients:		Biased effect size of multi-level model coefficients:		Biased growth coefficient of effect size of regression to multi-level	
	B0	B1	B0	B1	B0	B1
0.01	0.06727812	0.01679103	0.06728266	0.016875903	-0.007%	-0.503%
0.03	0.07450571	0.0164156	0.0745417	0.0164955	-0.048%	-0.484%
0.05	0.07668905	0.01744417	0.0766934	0.01752309	-0.006%	-0.450%
0.055	0.1007628	0.01869448	0.10079871	0.01879788	-0.036%	-0.550%
0.06	0.0841757	0.01635698	0.08417861	0.01642354	-0.003%	-0.405%
0.065	0.0861131	0.01542158	0.086116747	0.01501944	-0.004%	2.677%
0.1	0.08097108	0.01546579	0.080899377	0.01507619	0.089%	2.584%
0.2	0.08452881	0.02517319	0.08441917	0.02400584	0.130%	4.863%
0.3	0.14568395	0.01956844	0.145584353	0.0163644	0.068%	19.579%
0.5	0.21035854	0.02498493	0.20975297	0.01803338	0.289%	38.548%
0.7	0.30382888	0.02594441	0.30325008	0.016225647	0.191%	59.898%
0.9	0.62194652	0.05217747	0.62146553	0.01470385	0.077%	254.856%

Likelihood Ratio Test (Loglik)

The changes of this parameter in various ICCs are shown in the following Table and chart.

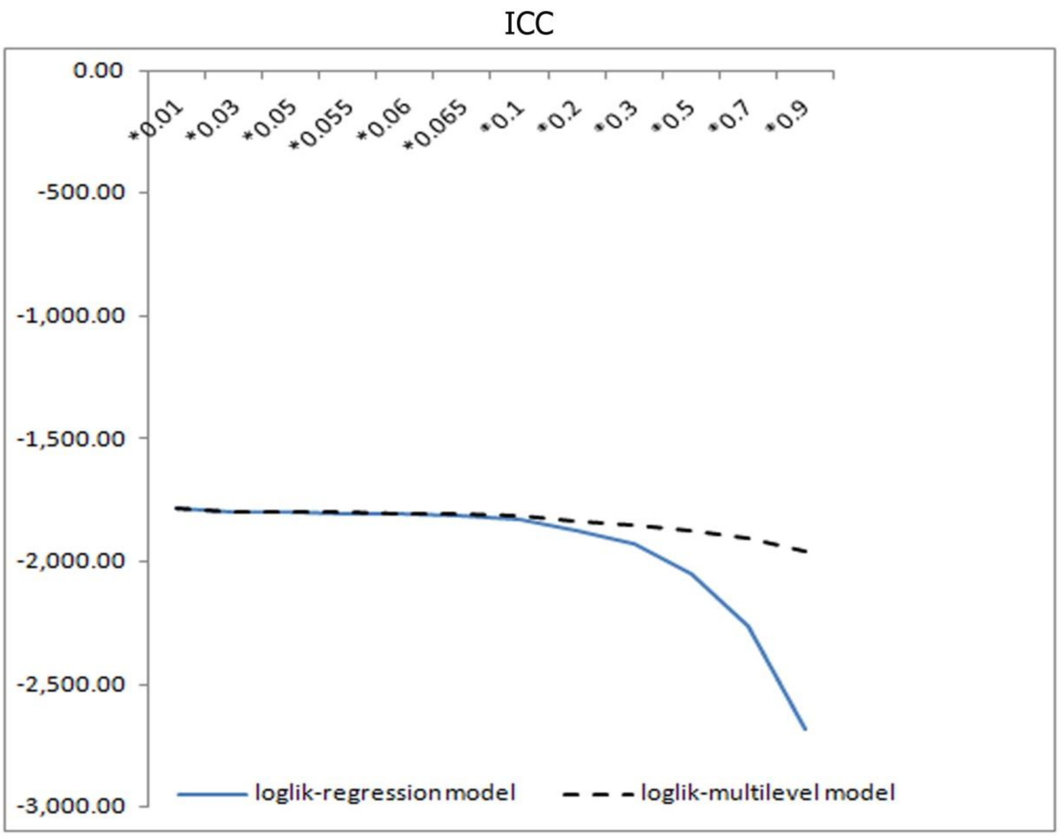

Image 1: The chart of Loglik changes in classic case-comparison with regression model

Table 4: Loglik difference in multi-level model with regression model

ICC	0.01	0.03	0.05	0.055	0.06	0.065	0.1	0.2	0.3	0.5	0.7	0.9
Difference of Loglik in multi-level model with regression model	0.49	1.88	3.37	4.71	5.72	7.75	12.61	40.04	74.41	178.69	356.08	720.24

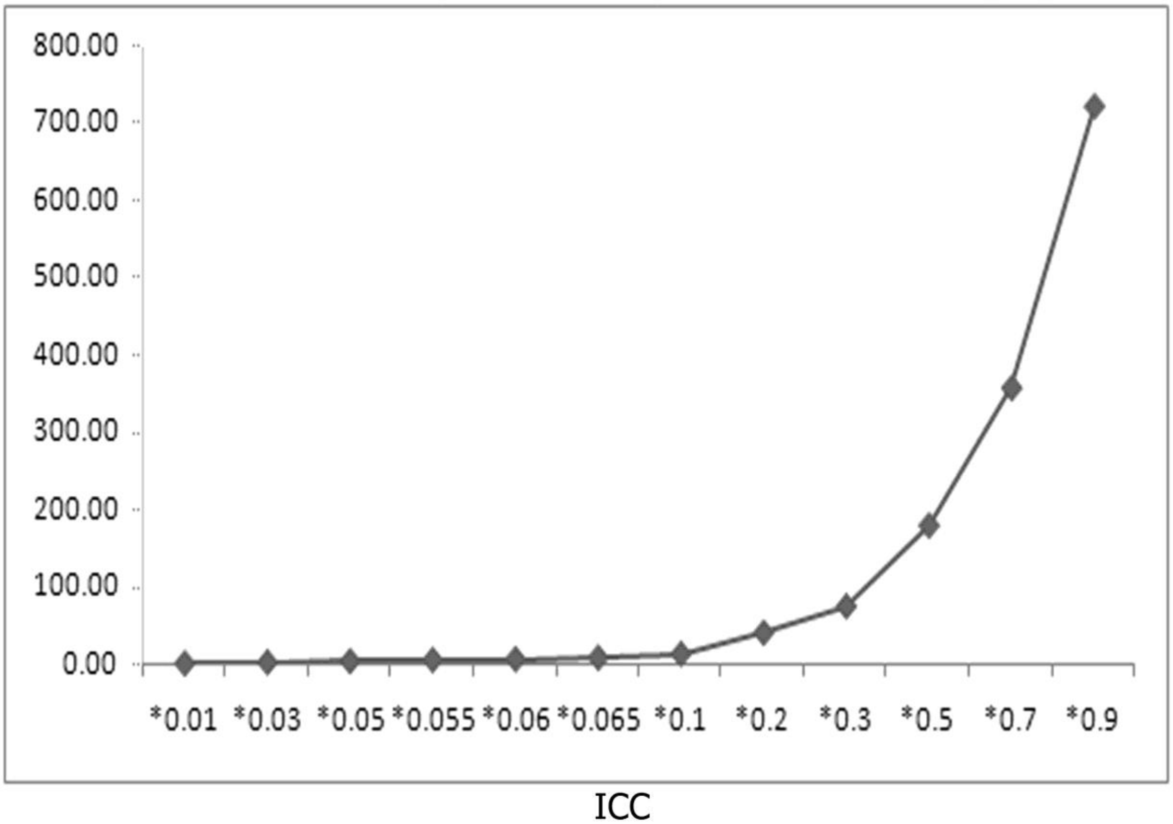

Image 2: The chart of Loglik difference in multi-level model compared to regression model

As it was said before, Loglik parameter is a criterion for determining model superiority. Compared to the two models, Loglik model of big value has better fitness for data. Chart 4-2 shows the comparison of the value of this parameter in various ICCs and Table 4-3 shows loglik difference in multi-level model with regression model (Loglik multilevel-Loglik regression). Based on the results, it seems that the difference of this parameter between two regression models and multilevel models to ICC=0.065 is very little and this difference is higher later and to ICC=0.1, this difference is ignored and after ICC=0.1, this difference is higher. IN addition, it shows that the higher ICC, Loglik parameter is increased with mild slope to ICC=0.5 and then the slope gets higher and multi-level model has high efficiency compared to regression model.

Akaike Information Criterion

The changes of this parameter in various ICCs are observed in the following Table and chart.

Research Article

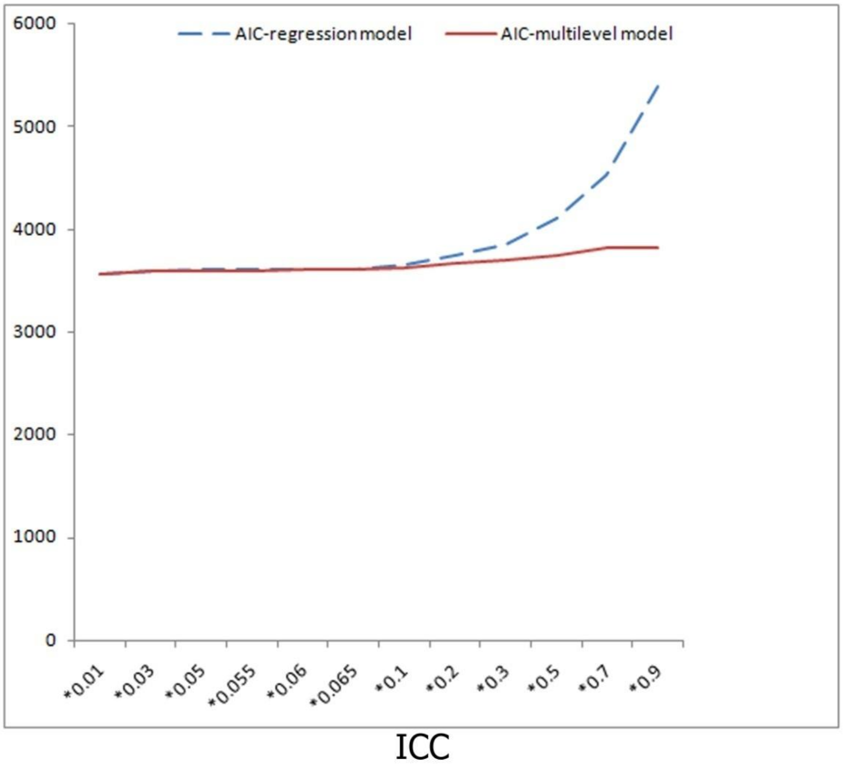

Image 3: The chart of AIC changes in classic case-comparison with regression model

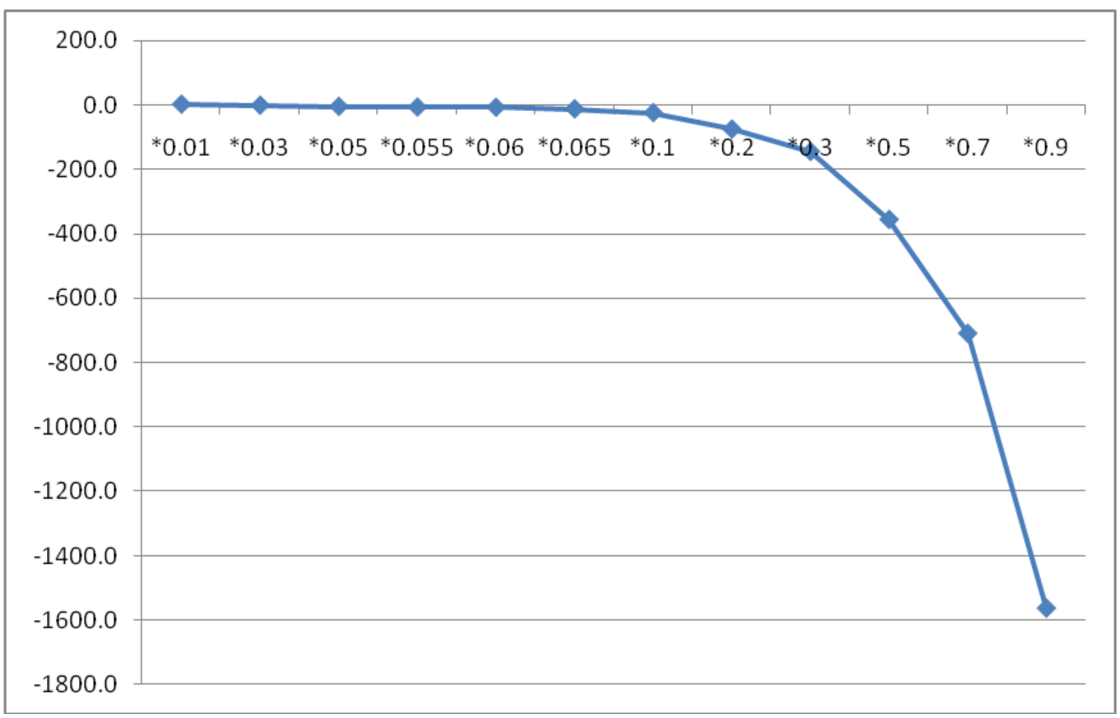

Image 4: The chart of AIC difference in multi-level model compared to regression model-in classed case

As it was said in 2-5-2, one of the goodness of fit criteria is Akaike information criterion or AIC and the model in which this feature is smaller is a good model.

Table 5: AIC difference in multi-level model with regression model-in classic case

ICC	**0.01**	**0.03**	**0.05**	**0.055**	**0.06**	**0.065**	**0.1**	**0.2**	**0.3**	**0.5**	**0.7**	**0.9**
AIC difference of regression and multi-level models	1.2	-2.0	-5.6	-7.4	-8.4	-13.1	-25.0	-76.1	-145.9	-356.5	-710.2	-1563.2

Chart 4-4 shows this parameter in various ICCs and Table 4-4 shows AIC difference in multi-level model with regression model (Loglik multilevel-Loglik regression). Based on the results, it seems that the difference of this parameter between two models to ICC=0.065 is ignored and later this difference is more and after ICC=0.1, this difference is considerable. Also, it shows that the higher ICC, AIC parameter in regression model is increased with mild slope to ICC=0.5 and the ascending slope is increased and regression model has low efficiency compared to multi-level model.

Coefficient of Determination (R^2)

As it was said before, coefficient of determination is a criterion for suitability of regression model and the bigger this criterion, the higher the efficiency of regression in determining the dependent variable based on independent variable or variables. The changes of coefficient of determination in various ICCs are observed in the following chart:

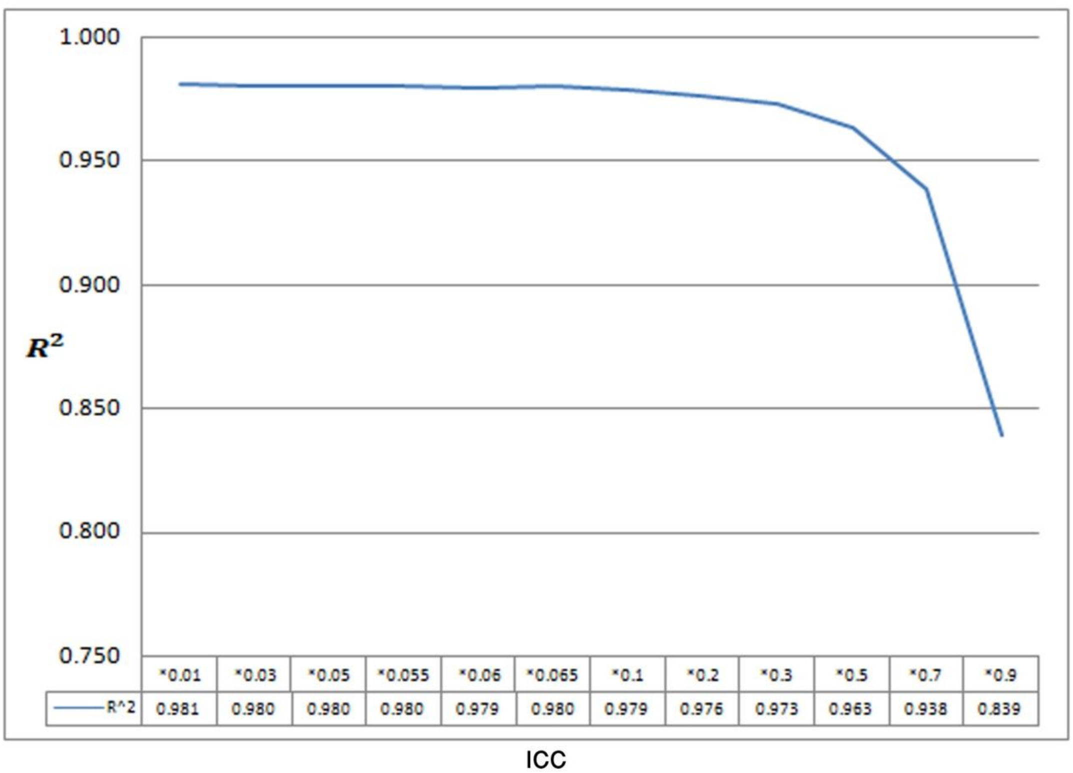

Image 5: The chart of coefficient of determination changes in various ICCs in classic case

As it is observed in ICCs above 0.1, coefficient of determination is reduced gradually and by increasing ICC, efficiency of regression model is reduced.

Indian Journal of Fundamental and Applied Life Sciences ISSN: 2231– 6345 (Online)
An Open Access, Online International Journal Available at www.cibtech.org/sp.ed/jls/2015/02/jls.htm
2015 Vol. 5 (S2), pp. 1432-1448/Vajargah and Nikbakht
Research Article

The Growth Coefficient of the Mean Effect Size of Real Coefficients

To compute this growth coefficient (G), if we consider the second model as regression model and the first model as multi-level model, the mean absolute value of deviation from initial value is model coefficient or effect size of equation 25.

The positive values of this growth coefficient show the increase of regression model error compared to multi-level model and negative values show the reduction of error of regression model compared to multi-level model to estimate B1 value or model slope. Thus, in each value of ICC, this growth coefficient is calculated.

The following chart shows the changes of this growth coefficient for line slope in two models (B1).

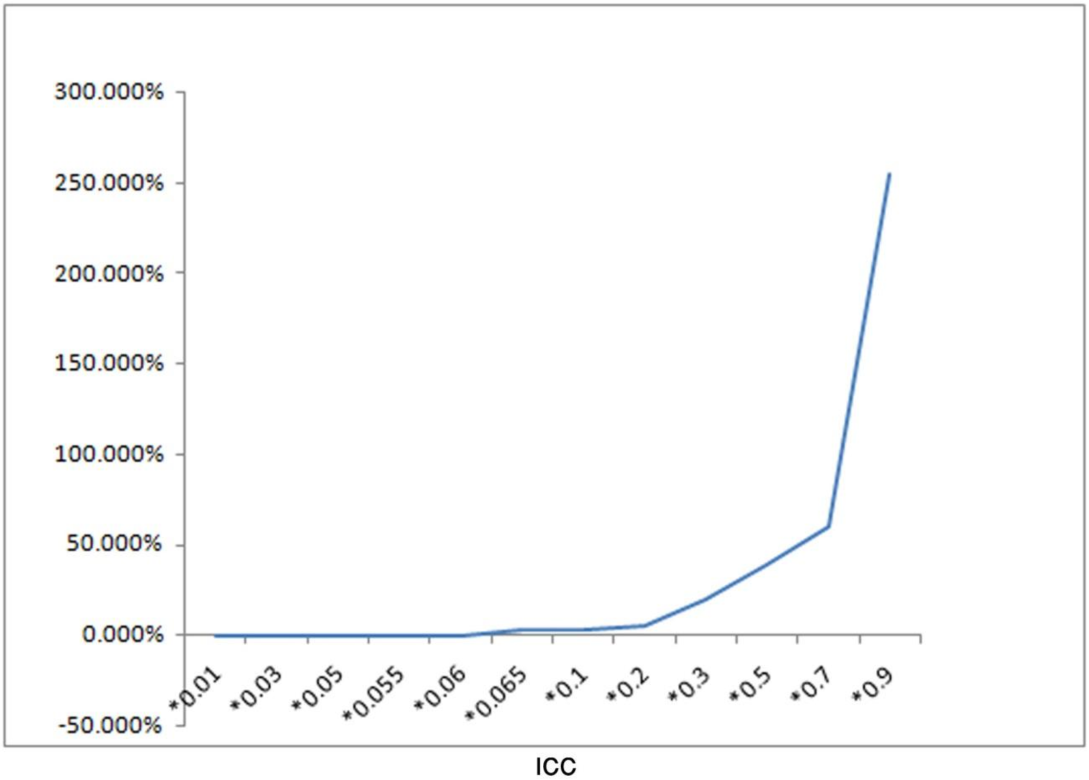

Image 6: The chart of mean growth coefficient of effect size B1- in classic case

Table 6: The growth coefficient of the mean absolute value of deviation from real coefficients B1 in classic case

ICC	**0.01**	**0.03**	**0.05**	**0.055**	**0.06**	**0.065**	**0.1**	**0.2**	**0.3**	**0.5**	**0.7**	**0.9**
The coefficient of mean growth of absolute value of deviation from real coefficients B1	-0.50%	-0.48%	-0.45%	-0.55%	-0.41%	2.68%	2.58%	4.86%	19.58%	38.55%	59.90%	254.86%

In this chart, it is seen that in ICC higher than 0.065, average absolute value of coefficients error in regression model is more than multi-level model and this increase to ICC=0.1 is mild and then continues

with ascending slope as in high ICC values, the superiority of multi-level model in reduction of the error of model coefficients compared to regression model can be observed well.
This coefficient for b0 in various ICCs is about to zero and we ignore to investigate it. (The values are in Table 4-1).
The summary of comparison with the estimation of
By considering the above items, we can say cut off point to be used in multi-level model in random intercept is with ignoring ICC=0.1

Applied Example

In this applied example, a sample including 81 files of children is selected of health center of south of Tehran as randomly. Of each file at seven different ages, the child below 2 years, ages and weights are registered. The aim is the investigation of the relationship between weights and age of children.

Table 7: The summary of data

	Age	Weight
Min	0	1.16
Max	24	16.1
Mean	7.76	7.908

Growth chart of weight index for children at various ages below two years

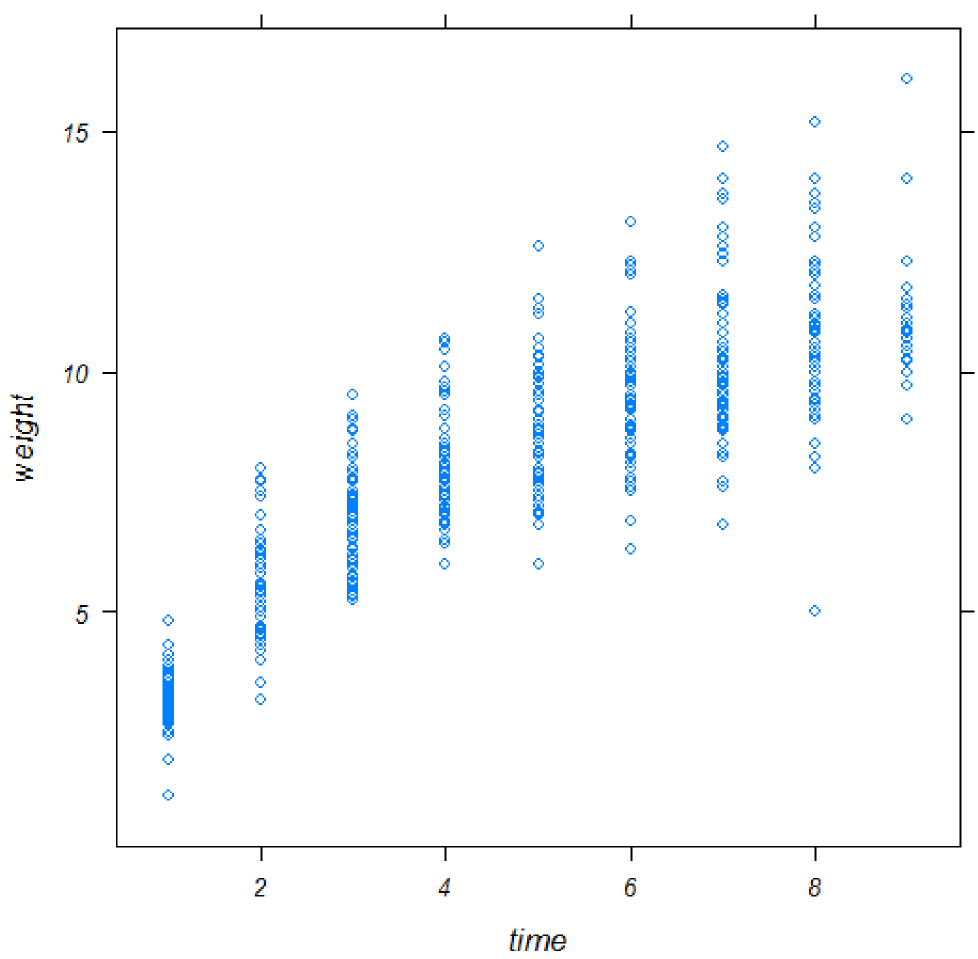

Image 7: The weight chart to growth time (age)

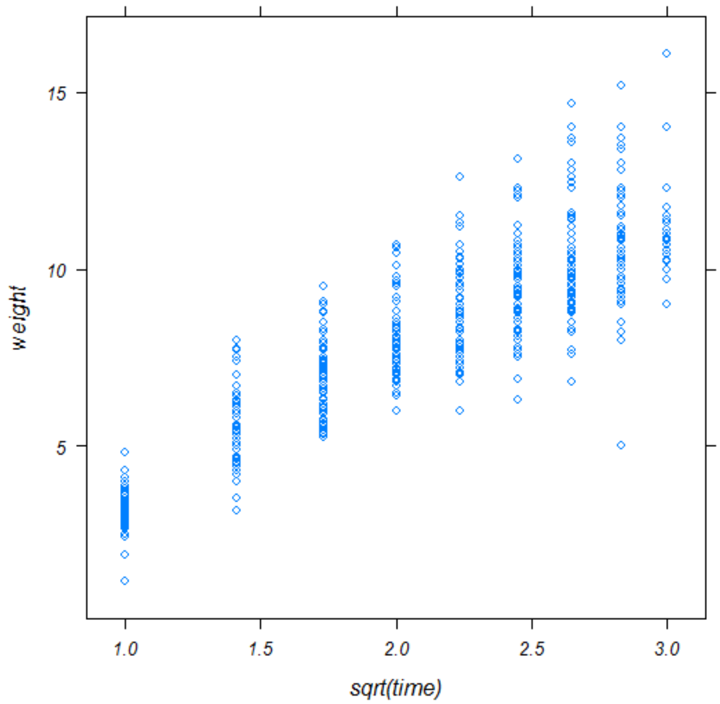

Image 8: Weight chart to square growth time (age)

Histogram and weight morality chart

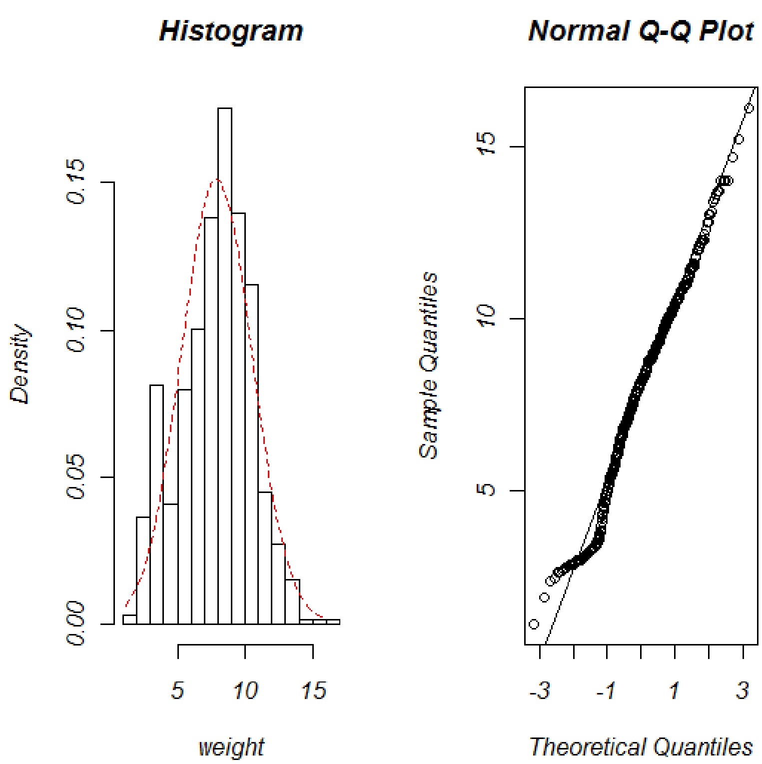

Image 9: Histogram and qq-plot to investigate normality of weight data

Based on these two charts, we can say the data distribution follows normal distribution a little.

Regression Model of Weight to Age

Table 8: The output of regression model fitting

model	**Weight~sqrt(age)**		
Method	**Generalized least squares fit by REML**		
Coefficients:			
	value	Std. Error	p-value
Intercept	3.107177	0.0891512	0
Sqrt(age)	1.931126	0.03195432	0

R^2	AIC	BIC	loglik
0.8	1955.589	1969.088	-974.7945

Regression model: $weight = 3.107177 + (1.9311260)(age)^2$

Fitting of random intercept model (model estimation method of REML method).

Table 9: Output of multi-level model fitting

σ_u	σ_e
0.8418146	0.6089281

model	**Weight~sqrt(age)**		
Method	**Linear mixed-effects model fit by REML**		
Coefficients:	**value**	**Std. Error**	**p-value**
Intercept	3.123677	0.10735478	0
Sqrt(age)	1.925886	0.01896518	0

AIC	BIC	Loglik
1474.142	1492.141	-733.0711

Multi-level model

$weight = 3.123677 + 1.925886(age)^2$

By comparison of AIC and Loglik we have:

Table 10: The comparison of AIC and Loglik of multi-level and regression models

	Regression model	Multi-level model
AIC	1955.589	1955.589
Loglik	-974.795	-733.071

σ_e^2 value is 0.370793 (0.608928 squared) and σ_u^2 is 0.708652 and of ICC, 0.785 is obtained.

Based on Table, AIC of multi-level is less and Loglik is higher than regression model showing that multi-level model acts better than regression model.

Based on simulation results, this is supported by observation of ICC value (ICC =0.785 is higher than 0.1).

Based on the results, multi-level model is suitable for these data.

CONCLUSION

Based on the results, we can say based on ICC=0.1, we can observe the improvement of multi-level performance compared to regression model in random intercept and we can consider ICC cut off point in using multi-level model in random intercept by REML method compared to regression model by ignoring 0.1.

REFERENCES

Aitkin M, Anderson D and Hinde J (1981). Statistical Modelling of Data on Teaching Styles (With Discussion). *Journal of Royal Statistical Society*.
Akhgari Omid and Golalizade Musa (2013). The investigation of the impact of sample size in multi-level models with the approach of subsampling. *Advanced Modeling Journal of Math* (1).
Akaike H (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**(6) 716–723.
Brown H and Prescott R (1999). *Applied Mixed Models in Medicine* (New York: John Wiley & Sons).
Bryk A and Raudenbush S (1992). *Hierarchical Linear Models for Social and Behavioral Research* (Sage) Newbury Park, CA.
Bryk AS and Raudenbush SW (1992). *Hierarchical Linear Models* (Newbury Park, Sage).
Cohen M (1998). Determining Sample Sizes for Surveys with Data Analyzed by Hierarchical Linear Models. *Journal of Official Statistics* **14** 267-275.
Goldstein H (1987). *Multilevel Models in Educational and Social Research* (C. Griffin and Co., London, Oxford University Press, New York).
Goldstein H (1995). *Multilevel Statistical Models* (Halstead Press) New York.
Goldstein H (2010). *Multilevel Statistical Models*, 4th edition (London: Amold Publishing) 12-13.
Grilli L and Rampichini C (2007). Multilevel factor models for ordinal variables. *Structural Equation Modeling* **14**(1) 1-25.
Hox JJ (2002). *Multilevel Analysis, Techniques and Application* (Mahwah/ N.J, Lawrence Erlbaum Associates).
Longford NT (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. Biometrika.
Luke DA (2004). Multilevel modeling, A sage university paper series. Quanitive application in social sciences.
McCulloh C and Searel S (2001). *Generalized Linear and Mixed Models* (New York: Wiley).
Myers RH, Montgomery DC and Vining GG (2002). *Generalized Linear Models with Application in Engineering and Sciences* (New York: John Wiley & Sons).
Pinherio JC and Bates DM (2000). *Mixed-Effect Models in R and S-plass* (New York: springer-verlag).
Sakamoto Y, Ishiguro M and Kitagawa G (1986). *Akaike Information Criterion Statistics*, Reidel, Dordrecht, Holland.
Schwarz G (1978). Estimating the dimension of a model. *Annals of Statistics* **6** 461–464.
Verbeke G and Molenberghs G (2000). *Linear Mixed Models for Longitudinal Data* (New York: Springer) 41-46.
Wulu JT (1999). Generalized Poisson Regression Models with applications, thesis submitted to Degree of doctor of philosophy in the Graduate Faculty of University of Alabama at Birmingham; USA.