J.Natn.Sci.Foundation Sri Lanka 2014 42 (1): 23-36
DOI: http://dx.doi.org/10.4038/jnsfsr.v42i1.6676

RESEARCH ARTICLE

A multilevel Bayesian analysis of university entrance eligibility for selected districts in Sri Lanka: methods and application to educational data

Nirodha I. Jayawardena and M.R. Sooriyarachchi*
Department of Statistics, Faculty of Science, University of Colombo, Colombo 03.

Revised: 25 June 2013; Accepted: 19 July 2013

Abstract: Multilevel data structures are becoming a commonly encountered phenomenon in educational research. This type of data generates a number of statistical problems, of which clustering is particularly important. To solve the problems inherent in these, special statistical techniques are required. This study aimed to determine the factors affecting the university entrance eligibility of students from some selected districts in Sri Lanka, whilst capturing the layered structure of this educational data into pupil and school levels and determining how these layers interact and impact the dependent variable of interest. This study used university entrance eligibility of General Certificate of Education: Advanced Level (G.C.E) (A/L) student records in 3 districts of Sri Lanka. The response variable is university entrance eligibility of students, which is a binary variable. Thus a two level binary logistic model was fitted using the Bayesian Markov Chain Monte Carlo (MCMC) method as this method has some advantages over other classical statistical methods.

When determining the eligibility for university entrance, GCE A/L students find Science subjects more competitive than Arts and Commerce subjects. Students with a higher IQ level (as given with the data) and students with higher English ability stand a better chance. The chance is higher for students from national schools compared to provincial and private schools, and girls show more potential than boys. Students studying in English medium have a higher chance while those studying in Tamil medium have a lower chance compared to the students studying in Sinhala medium.

Keywords: Bayesian methods, binary responses, correlated data, cross level interaction, multilevel models, statistics education research.

INTRODUCTION

Multilevel data are structures that consist of multiple units of analysis, one nested within the other. The existence of hierarchies are found in many subject areas, hence the flexibility of multilevel models is reflected in a variety of applications, namely in the fields of medical, biological and social sciences such as educational and political sciences. The goal of multilevel analysis is to account for variance in a dependent variable that is measured at the lowest level of analysis by considering the information from all levels of analysis. Statistical modelling of multilevel data has been in discussion for many years and various developments have been reported on this aspect (Aitkin & Longford, 1986; Goldstein, 1986; Hedeker & Gibbons, 2003). However, most of the early developments were concentrated in the area of continuous response variables. Hence the field of multilevel modelling for discrete categorical responses is a relatively new approach (Goldstein, 1992; Rashbash *et al*., 2004; Fielding & Yang, 2005). The interest in multilevel models for binary outcome variables is a relatively recent development in sociological analysis. Many sociologists are often interested in explaining and predicting phenomena that can be characterized by a binary variable. This paper is based on the modelling of a binary response in the presence of a multilevel data structure.

The Sri Lankan education system is becoming increasingly competitive, and as such the Advanced

* Corresponding author (roshinis@hotmail.com)

Level (G.C.E) (A/L) examination, which is considered as a prerequisite to enter into a state university has also become very competitive. The prime objective of this study was to determine the factors affecting the university entrance eligibility of students from some selected districts in Sri Lanka, whilst capturing the layered structure of this educational data into pupil and school levels and determining how these layers interact and impact on the dependent variable of interest. Multilevel modelling techniques allow assessing the variation in a dependent variable at several levels simultaneously and addresses how much the university entrance eligibility varies between schools compared with the extent of variation in university entrance eligibility for pupils within schools. Hence it eliminates the possibility of obtaining misleading results due to biased estimates and large standard errors, which can occur by ignoring the clustered structure of the population. Bayesian methods are used in the fitting of the multilevel model due to advantages, which will be explained later.

Table 1 : Description of the data and its abbreviations

Variable	Notation	Description	Levels	Code
*Gender	Gender	Gender of the student	Male (M)	0
			Female (F)	1
*English grade	English code	Grade obtained for the General English test	Marks (75 - 100)	A
			Marks (65 - 74)	B
			Marks (55 - 64)	C
			Marks (35 - 54)	S
			Fail (0 - 34)	F
			Absent for the exam	E
*IQ Code	IQ	The IQ marks were categorized using 20th percentile	Marks (75 - 100)	5
			Marks (65 - 74)	4
			Marks (55 - 64)	3
			Marks (35 - 54)	2
			Marks (0 - 34)/ absent for the exam	1
*Subject stream	Stream	The subject stream which the students followed	Bio Science	1
			Combined Mathematics	2
			Commerce	3
			Arts	4
*Medium of study	Medium	The medium which the students followed	Sinhala	1
			English	2
			Tamil	3
*Year	Year	Academic year which the students did the examination	2006	1
			2007	0
**School sector	School sec	The categorization of the type of the school	National schools	1
			Provincial schools	2
			Private schools	3
**School class sector	Class sec	The class formulation	Mixed schools	1
			Girls' schools	2
			Boys' schools	3
University eligibility	Qual	University eligibility of a particular student	Yes	1
			No	0

** school level (1st level) variables
* student level (2nd level) variables

The data for this study was provided by the University Grants Commission of Sri Lanka. All the students who sat for the GCE A/L in the districts of Colombo, Gampaha and Moneragala in the years 2006 and 2007 were considered. However, to begin with the study, possible modification to the original dataset was required to deal with the modelling difficulties and the convergence problems encountered with the MLwiN 2.19 software used in this study. After a heavy data manipulation process, the initial dataset was reduced to 41997 student records within 90 schools. Table 1 gives the variables and their respective categories for each level in this study.

REVIEW OF LITERATURE

Hierarchical data modelling sparked an enthusiasm in the mid 1980's when the influential works of Aitkin *et al*. (1981) showed that the aggregation over individual observations may lead to misleading results because of violating the heavy assumption of independence. This provoked many researchers to develop systematic approaches to deal with this type of data. While the schooling system presents an obvious example of hierarchical structure, it is very important to account for the multilevel structure of data when measuring the educational performance, to explore the hierarchy of pupils, classes, schools and sometimes also from local education authorities. The existence of such hierarchies cannot be ignored since when the groupings are established, even if at random, the groups tend to differentiate (Steedman, 1983). To ignore this relationship, risks overlooking the importance of group effects may even render invalid many of the traditional statistical methods used to study school effectiveness and the quality of school systems (Goldstein, 1995).

Tinkalin (2005) presented the relationships among the social background, gender, attainment and a range of other factors in their secondary analysis of the Scottish School Leavers Survey. Multilevel modelling was used as this allows an assessment of the relationships between different factors and high attainment, as well as an assessment of the variation between schools in their ability to produce high attainers when other potential affecting factors are held constant.

Furthermore, initial multilevel techniques were mainly confined to continuous responses, however, the early 1990's showed the extension of multilevel theory and its implementation in software to capably handle different types of outcomes such as binary, nominal scale multi-categorical and ordered categorical. Some striking methods among these were an improved second order approximation proposed by Goldstein (1995), Gauss-Hermite quadrature approximations to maximum likelihood proposed by Hedeker and Gibbons (1994) and Pinheiro and Bates (1995) and a higher order Laplace transformation proposed by Raudenbush *et al*. (2000). Browne (1998) discussed the varied approaches for fitting multilevel models. With the availability of powerful, high speed computers, Bayesian methods have become computationally feasible with the development of Markov Chain Monte Carlo (MCMC) methods, especially Gibbs sampling (Zeger & Karim, 1991). The advantage of this method of estimation is that in small samples, it takes account of the uncertainty associated with the estimates of the random parameters and can provide exact measures of uncertainty. The maximum likelihood methods tend to overestimate precision because they ignore this uncertainty. In small samples this will be important especially when obtaining 'posterior' estimates for residuals (Goldstein, 2003). These improvements together with the development of more sophisticated software such as MLwiN and STATA have brought multilevel modelling to a new phase of application.

METHODS AND MATERIALS

Univariate analysis-using Zhang and Boos test

Prior to fitting any statistical model, it is always important to test the nature and the strength of the relationships between the explanatory variables and the response. Several different methods for assessing these relationships exist and these methods vary according to the nature of the variables in question. However, as noted previously, most traditional techniques for assessing the relationship among categorical variables are not flexible enough to handle stratified data in a multilevel situation. Therefore, traditional univariate techniques such as chi-squared tests will not suffice in the present scenario. Hence, in order to address the above fact it was decided to identify a suitable technique for assessing the nature of the relationship among the variables. Generalized Cochran-Mantel-Haenszel test statistics for correlated categorical data (Zhang & Boos, 1997) provided three different test statistics, which could be used in place of the traditional chi-squared test for testing associations between stratified categorical variables. The work presented a detailed discussion about the suitability of each of the statistics (T_p, T_u and T_{EL}). According to the simulation study conducted by Zhang and Boos (1997), T_p was found to be the most suitable statistic for unequal strata size as is the case in this study.

Multilevel model for a binary response variable

This section is concerned about the theory behind multilevel models that have a binary response. In many situations the response variable is not continuous but is instead binary. For example, the interest may be in whether or not a student is university eligible and would have a response variable coded *1=eligible, 0=not eligible*. Similarly one could be interested in the variation in university eligibility in terms of the school. For example, university eligibility may be associated with a student's own characteristics and or by the characteristics of the school they attended.

Structure of the model

A general (one level) logistic model for binary response data is given by Agresti (1990).

Consider a two-level logistic regression model (Goldstein, 2003) with binary response Y_{ij}, which equals 1 if student i, in school j was university eligible and 0 if not. Let π_{ij} be the probability of student i in school j being eligible for university entry. That is $\pi_{ij} = \Pr(Y_{ij} = 1)$.

We begin with a random intercept or variance components model that allows the overall probability of university eligibility to vary across schools. If we have a single explanatory variable, x_{ij}, measured at the student level, then extended two-level random intercept model is as follows:

$$logit(\pi_{ij}) = \beta_{0j} + \beta_1 x_{ij} \text{ where } \beta_{0j} = \beta_0 + u_{0j} \text{ and } u_{0j} \sim N(0,\sigma_u^2) \quad ...(1)$$

For a random intercept model for a binary response, the intercept consists of two terms: a fixed component β_0 and a school specific component, the random effect u_{0j}. Here it is assumed that the u_{0j} follow a normal distribution with mean zero and variance σ_{uo}^2 (Rashbash *et al*., 2004).

Fitting the model

In the fitting of model (1) the model parameters were first estimated using the iterative generalized least squares (IGLS) method followed by Markov Chain Monte Carlo (MCMC) estimation method. As far as the IGLS method is concerned it uses marginal quasi -likelihood (MQL) *versus* penalized quasi-likelihood (PQL) approximations to fit binary response models. However the PQL method produces more accurate estimates than the MQL method (Rashbash *et al*., 2004), therefore PQL (II), which denotes PQL method with second order Taylor series approximation was used as the estimation procedure. A model is estimated from MCMC by setting its monitoring chain length and thinning to required amounts (Gelman & Rubin, 1992).

The Bayesian approach to statistics can be thought of as a sequential learning approach where the prior beliefs/ ideas are combined with the data collected to produce posterior beliefs/ideas to a problem. Then the previous posterior ideas act as prior knowledge and combined with data simulated from the joint posterior distribution using a sampler such as Gibbs sampler. For example, for an unknown parameter θ when the prior beliefs are condensed to a prior distribution $p(\theta)$ and when the collected data y (with the distributional assumption) produces a likelihood function $L(y|\theta)$, which is used as the function that maximum likelihood methods maximize. Then the prior distribution and the likelihood function are combined to produce the posterior distribution for θ, $p(\theta|y) \propto p(\theta)L(y|\theta)$ where this equation can be used to reach inferences about θ. However in this approach calculating the proportionality constant is much of an issue. MCMC methods circumvent this problem as it does not calculate the exact form of the posterior distribution but instead produce simulated draws from it. MCMC methods are more general in that they can be used to fit many statistical models. It is a simulation based procedure so that rather than simply producing point estimates, the methods are run for many iterations and at each iteration an estimate for each unknown parameter is produced. These estimates will not be independent, as at each iteration the estimates from the last iteration are used to produce new estimates. However, to choose the starting values it is important to run IGLS or RIGLS before running the MCMC estimation and then the process is simply repeated many times using the previously generated set of parameter values to generate the next set. The chain of values generated by this sampling procedure is known as a Markov chain, as every new value generated for a parameter only depends on its previous values through the last value generated. The field of MCMC convergence diagnostics is concerned with calculating when a chain has converged to its equilibrium distribution. In MLwiN by default the burn-in period is set as 500 iterations (Browne, 2012).

MLwiN 2.19 uses the Metropolis Gibbs hybrid estimation method with univariate updates (Browne, 1998). This is briefly described here. By extending the model defined in equation (3) to include several explanatory variables, the form of the model for the i^{th} individual in the j^{th} cluster can be expressed as given in equation (4). Here l refers to the number of the explanatory variable.

$$logit\,(\pi_{ij}) = \Sigma_{t=0}^{p} \beta_t X_{ijl} + u_{oj} \quad ...(2)$$

Where $Y_i \sim \text{Bernoulli}(\pi_i)$ and $u_{oj} \sim N(0,\sigma_u^2)$

(a) Fixed effects, β_t's

Browne (1998) has used the following general prior distribution for the fixed effects, β_t (t = 0, ...,p) where t refers to the t[th] iteration.

$$\beta_t^{(t)} = \begin{cases} \beta_t^* = \beta_t^{(t-1)} + \lambda_t \text{ with probability } \min\left(1, \dfrac{p(\beta_t^* \mid y...)}{p(\beta_t^{(t-1)} \mid y..)}\right) \\ \text{and } \lambda_t \sim N(0, \sigma_t^2) \\ \beta_t^{(t-1)} \text{otherwise} \end{cases}$$

The joint posterior distribution of the β_t's is then given by :

$\{P(\beta)^*\Pi_{ij}[\pi_{ij}^{yij}(1-\pi_{ij})^{(1-yij)}]\}$, where the first part of the product is the joint prior of the β_t's and the second part of the product is the likelihood function of the data. Here π_{ij} can be substituted from model. This is proportional to the marginal posterior distribution of β_t conditional on the data.

(b) Second level residuals, u_{oj}

$u_{oj} \sim N(0, \sigma^2_u)$ and let the variable containing u_{oj} for j=1,…, k be denoted by u_o. Here k is the number of 2[nd] level units (schools).

The marginal posterior distribution of u_{oj} is then proportional to :

$$\{\{\Pi_{ij}[\pi_{ij}^{yij}(1-\pi_{ij})^{(1-yij)}]]^*[(2\pi\sigma^2_u)^{-1/2}\exp(-\tfrac{1}{2}(\frac{u_0^2}{\sigma^2_u}))]\}$$

(c) Second level variance, σ^2_u

As Browne (1998) considers a model with several random effect terms resulting in a matrix V consisting of many second level variance terms, he takes a general inverse Wishart prior for V, that is V ~ 1W. However our model consists of only one random effect, u_{oj} resulting in a single second level variance term, (σ^2_u), the corresponding prior distribution of $(\sigma^2_u)^{-1}$ is taken to be $\chi^2_{(v)} \mid S = \sum_j (u_{oj}^2)$ where $\chi^2_{(v)}$ is the chi-square distribution with ν degrees of freedom and ν =k-2. Here k is the number of 2[nd] level units (schools).

So $(\sigma^2_u)^1 \sim x^2(v)|S$. Suppose the prior distribution of $(\sigma^2_u)^{-1}$ is denoted by $P((\sigma^2_u)^{-1})$ and the conditional distribution of u_{oj} given σ^2_u is denoted by $P(u_{oj}|\sigma^2_u) \sim N(0, \sigma^2_u)$ then the posterior distribution of $(\sigma^2_u)^{-1}$ is proportional to the product of $P(u_{oj}|\sigma^2_u)$ and $P((\sigma^2_u)^{-1})$.

Variable Selection

Although many techniques are available for selecting variables for the model in this study, the Bayesian variable selection method was used together with the Wald statistic (Agresti, 1990; Polit, 1996). While Bayesian approaches have been known in ecology and the environmental sciences for some time, using Bayesian approaches were virtually impossible until recently. But the advent of cheap computing has fostered the development of algorithms that provide precise numerical approximations for most problems, making the routine application of Bayes theorem a practical option. In order to determine the best model, a forward selection procedure was adopted. At each stage the Wald statistic was calculated together with the deviance information criterion (DIC) value to observe the significance of the added factors and to evaluate the fit of the model, respectively. Since the estimation technique used here is not maximum likelihood, the well-known likelihood ratio statistic is not applicable in this scenario. Several criteria have been proposed for use in model comparison and selection. Many proposed criteria have a component that quantifies the goodness of model fit, along with a component that penalizes model complexity; namely, the Akaike information criterion (AIC), the Bayesian information criteria (BIC), and the deviance information criterion (DIC). In the context of a Bayesian hierarchical model, the number of independent parameters included in the model is difficult to determine, which makes the use of AIC or BIC problematic. DIC has been proposed for model comparison in this context. As with all the other criteria, a lower value of DIC is preferred over a higher value. Spiegelhalter *et al.* (2002), have offered guidelines for using DIC to compare competing models.

Parameter interpretation for a 2 level model with cross level interaction terms

Cross level interaction refers to the interaction between higher level and lower level variables that is, for modification of the effects of lower level variables by characteristics of the higher level units to which the lower level units belong (or *vice versa*). For example, if the relation between a particular student's level of IQ and English proficiency differs by school characteristics (that is, school and individual level variables interact), there is said to be a cross level interaction. In multilevel models, whenever group specific estimates of the effect of a lower level variable are modelled as a function of higher level variables, a cross level interaction appears in

the final model. However, when the numbers of variables at different levels are large, there are vast number of possible cross-level interactions. The odds ratios associated with cross level interactions can be calculated as any other odds ratio (Agresti, 1990).

Residual analysis and model adequacy

Similar to any modelling procedure, multilevel modelling also requires a comprehensive residual analysis in order to validate model assumptions as well as the fit of the model. Even though the model specifications are different for different types of response variables, the definition and the analysis of multilevel residuals is common to all models. It is also important to note that the area of residual analysis and diagnostic testing of multilevel models have not been addressed thoroughly by researches up to date. Hence only the available analytical techniques are used for the purpose of validating model assumptions and the fit of the model. Since the residual analyses do not differ for different multilevel models, the theory associated with it will be presented for the most basic multilevel model with a continuous response (Rashbash *et al*., 2004).

RESULTS

As explained previously, the dataset in this study takes a hierarchical form with respect to students being clustered within schools. The 'Schools' can be considered as the stratification factor according to which the students are clustered. The response variable termed as 'Eligibility' refers to a binary variable to check whether a particular student is eligible or not. The dataset contains six explanatory variables at the student level, namely, 'Gender', 'Subject Stream', 'IQ Score', 'English Grade', 'Year' and 'Medium' followed by 'School Class Setting' and 'School Sector' being the respective school level variables. All eight variables are categorical variables with

Table 2 : Generalized Cochran-Mantel-Haenszel test (T_p) results on student level factors with the response

Variable	Levels	T_p	Degrees of freedom	p value
Subject stream	Bio	4678.164	3	0.00*
	Combined Mathematics			
	Commerce			
	Arts			
English Grade	A	1407.252	5	3.692566e-302*
	B			
	C			
	F			
	S			
	Absent/E			
Medium	Sinhala	7.77939	2	0.02045158*
	English			
	Tamil			
Year	2006	5.758711	1	0.01640711*
	2007			
IQ Score	Marks (0 - 34)/ Absent	1734.326	4	0.00*
	Marks (35 - 54)			
	Marks (55 - 64)			
	Marks (65 - 74)			
	Marks (75 - 100)			
Gender	Male	832.5471	1	4.527937e-183*
	Female			

* Significant at 5 % level

the English grade and IQ score being ordinal categorical. Although originally there were 74755 students in the dataset, after removing observations with missing values it was reduced to 41997.

Univariate analysis for identifying student level factor impact on the response

The univariate analysis was carried out for student level variables with the intention of identifying the effect of explanatory variables on the response. Since the dataset concerned in this study takes a hierarchical form, the generalized Cochran-Mantel-Haenszel tests for correlated categorical data was used in place of the traditional chi-squared techniques, using schools as the respective stratification factor. The notion of doing a univariate analysis prior to the model fitting is to gain important information about the variables to be included in the model. This adjusted univariate technique used is based on the T_p statistic proposed by Zhang and Boos (1997). The required calculations were carried out using the R-macro proposed by De Silva and Sooriyarachchi (2012). Table 2 demonstrates the results of the adjusted univariate test carried out to check the significance of student level covariates in the presence of school as the respective stratification factor.

According to the results in Table 2, it is clear that all the student level variables significantly affect the response variable at 5 % level of significance. Of these the highest significance is observed for the factors *Stream* and *IQ Score*.

Analysis based on modelling

Fitting a two level random intercept model

Before applying the modelling techniques some modifications were carried out by re-categorizing the variables in order to deal with the sparseness of data, which gave rise to zero observations. In the modelling phase all factors were considered as these were all significant in the univariate phase. The MLwiN software version of 2.19 was used in the model fitting and it sets the level coded with the lower value as the base by default. Also it assigns zero for the coefficient of the base category. Each student level factor was fitted separately and the model was first estimated from the IGLS; PQL (II) method followed by the MCMC method to obtain the Wald statistic and DIC values, respectively. The Wald statistic was calculated for each parameter coefficient and the p values of the statistic were then compared with the 5 % significance level in order to assess the significance of the coefficient. Thus the starting variable was decided based on both Wald statistic and DIC value. However since the factors have an unequal number of categories, using only Wald statistic in factor selection is difficult. Hence some other additional tool is needed to select the most significant factor at each stage. This procedure is continued by adding the second student level term to the factor selected above until it attains the lowest DIC value. However it should be noted that according to the forward selection framework, once a variable is selected to the model, it will not be removed throughout the process. After selecting the student level variables that should be included in the model, the next step is to focus on the school level factors that should be added to the above selected model. A similar procedure as above is carried out in selecting school level factors as well. The model, which results in a lower DIC value together with satisfactory Wald statistic measures is considered to be selected as the best model. According to the significance levels of the coefficients and by considering the DIC values it can be shown that the addition of school level variables to the model with student level variables has resulted in slight increment in the DIC value. Also from the Wald test several levels of those school level variables were found to be insignificant at 5 % level of significance. Then for the selected model, cross level interaction terms were added separately using the forward selection criteria. According to the significance levels of the coefficients and by considering the DIC values, it can be shown that the addition of cross level interaction terms has resulted in a decreased DIC value. The final model includes all the student level factors, namely, stream, IQ code, English grade, medium, gender and year and cross level interactions between Stream and class setting, IQ and school sector and IQ and class setting. Table 3 gives the parameter estimates, standard errors of the estimates and p values associated with the parameters for the fitted model. Also given are the DIC and the parameter estimates and standard errors of the estimates of the fixed and random coefficients.

School level variance component analysis - for the model with interactions

In order to justify whether fitting a multilevel model is sensible, it is advisable to first look at the significance of the school level variance component. This can be checked by the following hypothesis.

H_0: School level residual variance is zero ($\sigma_u^2 = 0$)

H_1: School level residual variance is not zero ($\sigma_u^2 \neq 0$)

According to the 2.5^{th} percentile value of 0.087 and 97.5^{th} percentile value of 0.162 for the school level variance, it

Table 3 : Final model parameters with interactions

Model	Factor	Level	Coefficient	SE	p value
		Fixed part			
Main effects	Stream	2	- 0.056	0.079	0.478411
		3	2.322	0.072	3.5E-228*
		4	2.115	0.087	1.5E-130*
	IQ code	2	0.672	0.100	1.82E-11*
		3	1.144	0.119	7.02E-22*
		4	2.154	0.128	1.52E-63*
		5	- 1.839	0.509	0.000303*
	English	S	0.481	0.033	4.01E-48*
		E	0.645	0.083	7.78E-15*
		C	0.868	0.037	1.1E-121*
		B	1.166	0.048	2.4E-130*
		A	1.632	0.056	1E-186*
	Medium	3	0.132	0.065	0.042278*
		4	- 0.556	0.078	1.02E-12*
	Gender	1	0.794	0.050	8.72E-57*
	Year	2007	0.086	0.023	0.000185*
	School sector	SchlSec2	- 0.079	0.098	0.420172
		SchlSec3	- 0.287	0.119	0.015876*
	Class sector	Claset2	- 0.254	0.157	0.105698
		Claset3	0.937	0.160	4.73E-09*
	Stream* class setting	Stream2*Claset2	- 0.316	0.095	0.00088*
		Stream3*Claset2	0.083	0.098	0.397029
		Stream4*Claset2	- 0.264	0.116	0.022854*
		Stream2*Claset3	- 0.523	0.098	9.46E-08*
		Stream3*Claset3	- 1.099	0.092	6.84E-33*
		Stream4*Claset3	- 1.627	0.120	7.07E-42*
	IQ* school sector	IQ2* SchlSec2	- 0.102	0.088	0.246419
		IQ3* SchlSec2	- 0.199	0.099	0.04442*
		IQ4* SchlSec2	- 0.572	0.096	2.55E-09*
		IQ5* SchlSec2	- 0.270	0.409	0.50916
		IQ2* SchlSec3	0.097	0.095	0.30723
		IQ3* SchlSec3	0.084	0.104	0.419268
		IQ4* SchlSec3	0.016	0.102	0.875353
		IQ5* SchlSec3	-0.803	0.614	0.190935
	IQ* class setting	IQ2* Claset2	-0.048	0.082	0.558302
		IQ3* Claset2	0.316	0.108	0.003434*
		IQ4* Claset2	0.320	0.128	0.012419*
		IQ5* Claset2	0.257	0.431	0.550983
		IQ2* Claset3	- 0.060	0.099	0.544475
		IQ3* Claset3	- 0.307	0.503	0.541638
		IQ4* Claset3	- 0.572	0.096	2.55E-09*
		IQ5* Claset3	- 0.803	0.614	0.190935
		Random part			
Ω_u	0.115	0.021			
DIC (Deviance Information Criteria: 46504.33					

* Significant at 5 % level

can be shown that the above interval does not include the value zero thus we reject H_0 at 5 % level of significance and conclude that the school-level residual variance component is not equal to zero. Thus, the use of a two level model taking school as the respective second level is justifiable.

Residual analysis and classification power of the final model

In order to evaluate the fit of the model, the school level residuals obtained for the final fitted model was analyzed using two graphical tools, namely, the Caterpillar plot and the Normal plot.

Graphical techniques in checking model adequacy: According to the Caterpillar plot illustrated in Figure 1 most school level residuals contain zero within their 95 % confidence interval. This implies that these schools do not show significant differences in university eligibility from the overall mean predicted by the fixed part of the model. However there are some schools whose residual shows the highest positive deviations and whose residual shows the highest negative deviations. Therefore it can be concluded that these schools have the largest effect on the response. The normal probability plot given in Figure 2 illustrates an approximate straight line indicating that the residuals are approximately normal.

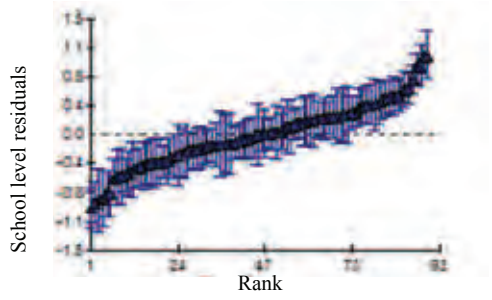

Figure 1. Estimated school-level residuals for the final full model

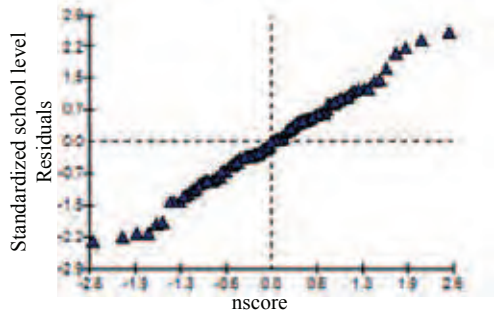

Figure 2. Normal Plot for school-level residuals for the final full model

Distributional Assumptions:

Anderson Darling normality test results

In order to confirm the results of Figure 2, an Anderson Darling normality test was carried out.

The Anderson Darling test for the estimated school-level residuals tests the hypothesis,

H_0: The data is distributed normally

H_1: The data is not distributed normally

If the p value obtained for the Anderson Darling statistic is less than 0.05, we reject H_0 at 5 % level of significance to conclude that the data is not distributed normally. However, when the Anderson Darling test was carried out, a p value of 0.702 was obtained. Hence, we do not reject H_0 at 5 % level of significance, and thus conclude that the school-level residuals follow a normal distribution in the fitted model with interactions.

Testing binomial assumption of the data: When the response is the number of times an event occurs out of a fixed number of 'trials', the distribution is typically taken to be binomial. Thus in this study the logistic model was used. The assumption of the binomial distribution can be evaluated using the 'extra binomial' variation (Goldstein, 2003). As a rule of thumb if the extra binomial parameter (EBP) is close to 1, there exhibits no over or under dispersion in the model. This model gives a EBP value of 0.907, which is close to 1 and thus it can be concluded that the binomial assumption is valid.

Prediction accuracy analysis: Table 4 gives the prediction accuracy according to the fitted model based on classification.

Table 4: Classification table

		Actual outcome		Total
		Not eligible	Eligible	
Predicted outcome by the model	Not eligible	9624 (56 %)	4549 (18 %)	14173
	Eligible	7525 (44 %)	20299 (82%)	27824
Total		17149	24847	41997

Table 5: Odds ratios for each of the significant main effects and interaction terms

Interaction terms	Levels	Odds ratio	95 % confidence interval	
			Lower bound	Upper bound
Stream*class setting	Stream2[when Claset2]	0.689354	0.616943	0.770264
	Stream3[when Claset2]	11.07843	9.810142	12.51069
	Stream4[when Claset2]	6.366183	5.538839	7.317109
	Stream2[when Claset3]	0.560459	0.503005	0.624475
	Stream3[when Claset3]	3.397365	3.023009	3.81808
	Stream4[when Claset3]	1.629055	1.374592	1.930623
	Claset2[when Stream2]	0.565525	0.39956	0.800427
	Claset2[when Stream3]	0.842822*	0.657689	1.080068
	Claset2[when Stream4]	0.595711	0.467443	0.759176
	Claset3[when Stream2]	1.512857	1.062279	2.154553
	Claset3[when Stream3]	0.850441*	0.618184	1.169959
	Claset3[when Stream4]	0.501576	0.358536	0.701683
IQ*school sector	SchlSec2[when IQ code2]	0.834435*	0.676575	1.029127
	SchlSec2[when IQ code3]	0.757297	0.597676	0.959547
	SchlSec2[when IQ code4]	0.521524	0.413304	0.658081
	SchlSec2[when IQ code5]	0.705393*	0.327171	1.520854
	SchlSec3[when IQ code2]	0.826959	1.078069	1.924812
	SchlSec3[when IQ code3]	0.816278*	0.601707	1.107426
	SchlSec3[when IQ code4]	0.762616	0.618688	0.939988
	SchlSec3[when IQ code5]	0.336216	1.569701	1.991351
IQ*class setting	Claset2 [whenIQ 2]	0.953134*	0.797577	1.13903
	Claset2 [whenIQ 3]	1.37163	1.139071	1.65167
	Claset2 [whenIQ 4]	1.377128	1.138703	1.665474
	Claset2 [whenIQ 5]	1.293045*	0.354594	4.715151
	Claset3 [whenIQ 2]	2.403678	1.945166	2.97027
	Claset3 [whenIQ 3]	1.877611*	0.697863	5.051738
	Claset3 [whenIQ 4]	1.440514	1.182042	1.755504
	Claset3 [whenIQ 5]	1.143393*	0.249022	5.249938
Common interactions	IQ code2[when SchlSec2, Claset2]	1.813833	0.716559	0.924312
	IQ code3[when SchlSec2, Claset2]	3.528949	3.02112	4.12214
	IQ code4[when SchlSec2, Claset2]	6.69928	5.636818	7.962001
	IQ code5[when SchlSec2, Claset2]	0.156923*	0.085917	1.286613
	IQ code2[when SchlSec2, Claset3]	1.804125	0.02345	0.84345
	IQ code3when SchlSec2, Claset3]	1.892692*	0.712968	5.02446
	IQ code4[when SchlSec2, Claset3]	2.745601	2.572798	2.93001
	IQ code5[when SchlSec2, Claset3]	0.054367*	0.019351	1.152743
	IQ code2[when SchlSec3, Claset2]	0.993024*	0.851765	1.15771
	IQ code3[when SchlSec3, Claset2]	4.683286	3.926559	5.58585
	IQ code4[when SchlSec3, Claset2]	12.06128	9.946505	14.62568
	IQ code5[when SchlSec3, Claset2]	0.092089*	0.029245	1.28998
	IQ code2[when SchlSec3, Claset3]	1.208119	0.008234	0.3967
	IQ code3[when SchlSec3, Claset3]	2.511801*	0.945152	6.675268
	IQ code4[when SchlSec3, Claset3]	4.943136	4.531287	5.392418
	IQ code5[when SchlSec3, Claset3]	0.031905*	0.008157	1.204789

Main effects	Levels	Odds Ratio	95 % confidence interval	
			Lower bound	Upper bound
English	S	1.617691	1.553011	1.682371
	E	1.905987*	0.743307	2.068667
	C	2.382142	2.309622	2.454662
	B	3.20913	3.11505	3.30321
	A	5.114093	5.004333	5.223853
Medium	2	1.141108	1.013708	1.268508
	3	0.573498	0.420618	0.726378
Gender	1	2.212228	2.114228	2.310228
Year	2007	1.089806	1.044726	1.134886

* Non-significant odds ratios at 5% level

According to Table 4, it can be seen that approximately 72 % of the data is correctly classified. Hence the chosen model from the above procedures seems to be giving satisfactory results.

Interpretation and calculation of the parameter estimates

Having fitted the final model it is of importance to interpret the results obtained.

Consider first the binary predictor,
Where;

Qual~ Binomial (denom, π_{ij})

$$\text{Qual} = \begin{cases} 1, & \text{if a student is eligible} \\ 0, & \text{if a student is not eligible} \end{cases}$$

To interpret the particular parameter of interest, the odds ratios are calculated and presented in Table 5. The following section will discuss the impact from all significant terms under the final fitted model on the response variable.

The results of Table 5 indicate the following important conclusions for the 3 districts studied. It was found that the students who are following Commerce in girls' schools have 11 times more odds to enter a university than Bio Science students from girls' schools, while students who are following Arts stream in girls' schools too showed a 6 times higher odds of entering a state university than Bio Science students from girls' schools. However students who are following Combined Mathematics in girls' schools have a 0.69 times less odds to enter to a university compared to those who followed Bio Science in girls' schools. Students who are following Commerce and Arts in boys' schools have 3.5 and 1.5 times more odds, respectively of being university eligible than the students following Bio Science from boys schools. Students who are following Combined Mathematics in boys' schools have 0.56 times less odds to enter a university compared to those who followed Bio Science in boys' schools. The above results may be due to that both female and male students find Science subjects more competitive and challenging than Commerce and Arts subjects.

Moreover, students from girls' schools following Combined Mathematics or Arts as a subject have 0.57 and 0.6 times less odds, respectively to be eligible to enter a university than those from mixed schools. Also students from boys' schools following Combined Mathematics have 1.5 times more odds to be eligible to enter into a university than those from mixed schools. The students from boys' schools following Arts as a subject have 0.5 times less odds to be eligible to enter to a university than those are from mixed schools.

Furthermore it was found that the odds of eligibility of national schools compared to provincial schools and private schools are higher for all significant IQ score bands. This may be due to the national schools getting more government facilities and support compared to other types of schools i.e. A/L examination based model papers, educational seminars and having more scholarship students. The odds can be quantified as before by using Table 5.

Students who have higher IQ score bands [marks (55–65) and marks (65–74)] in girls' schools are 1.37 times more likely to be eligible to enter a university than who are in mixed schools. Also students who are in the IQ score bands of marks 35–54 and 65–74 in boys' schools are 2.4 and 1.4 times respectively more likely to be eligible to enter a university than who are in mixed schools.

Students who are in provincial girls' schools and provincial boys' schools, private girls' schools and private boys' schools with higher IQ score bands [marks (35–54), marks (55–64) and marks (65–74)] are more likely to be qualified than those with an IQ score band of marks 0–34 and absent for the examination in an increasing order. The odds can be quantified as before by using Table 5.

Accordingly, the students who gained A, B, C and S grades for the English examination have respectively 5, 3, 2 and 1.5 times more odds of being eligible to enter into a university than who failed (F). This may be due to that students who are more knowledgeable in English tend to use additional reading materials/publications, so their ability of acquiring further knowledge may reflect the increased chance to enter a university.

Furthermore, the students who followed their examination in English medium have 1.15 higher odds to be eligible to enter a university than the students who did their examination in Sinhala medium. However those students who did their examination in Tamil medium have 0.57 less odds to be eligible to enter a university than the students who did their examination in Sinhala medium.

Female students have a 2.2 higher odds of being eligible to enter to a university than male students.

Students who sat for their Advanced Level examination in the year 2007 have a 1.09 times higher odds of being eligible for university, compared to the

students who did their Advanced Level examination in the year 2006. This may be due to the fact that during this academic year there were some syllabus changes and some changes in the difficulty level of question papers.

DISCUSSION

Multilevel modelling is an increasingly popular approach to modelling hierarchically-structured data, outperforming classical regression in predictive accuracy. This paper highlights the important research findings in the educational context. Although multilevel data structures are commonly encountered in social and medical research, the modelling of such data has been confined mostly to continuous responses. Hence, modelling multilevel data in the presence of a binary response and categorical factors proved to be a new and challenging experience. A complete analysis of the data was carried out by fitting a binary logistic multilevel Bayesian model using the accepted methodology.

Summary of conclusions

Having established the final model, the following are some of the major findings obtained with respect to the districts concerned.

University eligibility for both girls' and boys' schools are highest for the Commerce stream followed by Arts, Bio Science and Physical Science in this order. When the Physical Science stream is considered, boys' schools have the highest odds of university eligibility followed by mixed schools and girls' schools, respectively. In the Arts stream, mixed schools have the highest odds of university eligibility followed by girls' and boys' schools in that order. For many IQ bands the students from national schools show higher odds of eligibility compared to both provincial and private schools. For the middling and higher IQ levels, the students in boys' schools have a higher odds of being eligible to enter university, followed by girls' schools and mixed schools in this order. Students from all school types having higher IQ levels have more odds of university eligibility compared to the lowest IQ levels. The odds of university eligibility for a student increases with increasing performance in the English paper. The odds of university eligibility are highest for English medium students followed by Sinhala and Tamil medium students in this order. Female students have a higher odds of eligibility compared to male students.

Students from the 2007 A/L batch have a higher odds of university eligibility compared to students from the 2006 A/L batch.

Recommendations

Considering the variables studied, the results in Table 5 indicate that the main effects of English and Medium are significant. This indicates that the students who perform better in the English paper and the students who study in the English medium have a higher eligibility of university entrance. Another important effect as shown in Table 5 is the school sector and IQ interaction. This combined effect indicates that students with moderate IQ in national schools have a higher eligibility of university entrance than the students with moderate IQ from provincial schools. The students with higher IQ from national schools have a higher eligibility of university entrance than the students with higher IQ from private schools. While taking these important effects into account, it should also be stressed that there could be other factors, which have not been considered in this study that could be important.

Thus the students intent on entering a university can be advised to improve their English and study in English medium. If the student's IQ is moderate or high, studying in a national schools is advised. University eligibility for both boys and girls are highest for the Commerce stream. Thus, those intent on going through the national universities have a high chance if they select the Commerce stream.

Problems encountered

Due to reasons of confidentiality the University Grants Commission provided data for only 3 districts in Sri Lanka. Thus these results cannot be generalized to the entire island. Also as data is available only for 3 districts, district could not be used as the third level.

Many problems were also encountered at each stage of this study mainly due to the inability of using traditional univariate techniques in the presence of multilevel data structure. Most traditional techniques were built upon a heavy assumption of independence, while in multilevel data structures this assumption does not hold. Hence one of the main challenges that had to be dealt with was to find a suitable univariate test to analyze multilevel level data structure. The Generalized Cochran-Mantel-Haenszel Test for correlated categorical data, proposed by Zhang and Boos (1997) was used for the univariate analysis. This involved in several re-categorizations of variables in order to deal with sparse data problem. For example, different categorizations as 20^{th}, 25^{th} percentiles were evaluated simultaneously.

Particular difficulties faced during the model fitting involved non convergence problems of estimates and the estimated variance converging to negative values. The latter phenomenon however is commonly encountered in multilevel modelling as suggested by Brown and Prescott (2012). But negative variance problem had a major impact in identifying confidence intervals and so on. Also the model selection phase involved in identifying significant variables with unequal number of levels. Hence the use of only Wald statistic resulted in indecisive situations. Thus, another measurement together with Wald test was needed. The Bayesian deviance information criterion (DIC) was found to be a satisfactory measure in this case, however, in order to obtain this statistic it required the model to be refitted under the MCMC method. This required considerable time to get a converged result.

Despite the limitations discussed, the findings of this multilevel study showed promising results, which one should take into consideration to evaluate university entrance eligibility for the selected group of Sri Lankan students.

Further research

The dataset used in this study was restricted to three districts for the reason of confidentiality. It is recommended to apply the techniques described here to a sample of students drawn from all districts in Sri Lanka in order to generalize the findings to the entire Sri Lankan context. This will then be a 3 level study with districts making up the third level.

This research can also be considered as the basis for gaining insights to the need of a sound multilevel goodness-of-fit technique. As discussed in this study, goodness-of-fit tests available in the multilevel context is mostly basic graphical techniques.

Acknowledgement

The authors wish to thank The University Grants Commission of Sri Lanka for providing the much required data for the study. Thanks is also due to Ms Hansapanie Rodrigo for sharing her data.

REFERENCES

1. Agresti A. (1990). *Categorical Data Analysis*. John Wiley & Sons, New York, USA.
2. Aitkin M., Anderson D. & Hinde J. (1981). Statistical modeling of data on teaching styles (with discussion). *Journal of the Royal Statistical Society B* **144** (A): 419 − 461.
3. Aitkin M. & Longford N. (1986). Statistical modeling in school effectiveness studies (with discussion). *Journal of the Royal Statistical Society* **149** (A): 1 − 43.
4. Browne W.J. (1998). Applying MCMC methods to Multilevel models. *Ph.D. thesis,* University of Bath, UK.
5. Browne W.J. (2012). *MCMC Estimation in MLwiN, v2.25*. Centre for Multilevel Modelling, University of Bristol, UK.
6. Brown H. & Prescott R. (2012). *Applied Mixed Models in Medicine*. John Wiley Press, New York, USA.
7. De Silva D.B.U.S. & Sooriyarachchi M.R. (2012). Generalized Cochran-Mantel-Haenszel test for multilevel correlated data: an algorithm and R function. *Journal of the National Science Foundation of Sri Lanka* **40** (2) : 137 − 148. DOI: http://dx.doi.org/10.4038/jnsfsr.v40i2.4441
8 Fielding A. & Yang M. (2005). Generalized linear mixed models for ordered responses in complex multilevel structures. *Journal of the Royal Statistical Society* **168**: 159 −183. DOI: http://dx.doi.org/10.1111/j.1467-985X.2004.00342.x
9. Gelman A. & Rubin D. (1992). Inference from iterative simulation using multiple sequence. *Statistical Science* **7**: 457 − 511. DOI: http://dx.doi.org/10.1214/ss/1177011136
10. Goldstein H. (1986). Multilevel covariance component models. *Biometrika* **74** (2): 430 − 431. DOI: http://dx.doi.org/10.1093/biomet/74.2.430
11. Goldstein H. (1995). *Multilevel Statistical Models*, 1st edition. Edward Arnold Publishers, London. UK.
12. Goldstein H. (2003). *Multilevel Statistical Models*, 3rd edition, Edward Arnold Publishers, London. UK.
13. Goldstein H. & Rasbash J. (1992). Efficient computational procedures for the estimation of parameters in multilevel models based on iterative generalised least squares. *Computational Statistics and Data Analysis* **13**: 63 − 71.
14. Hedeker D. & Gibbons R.D.A. (1994). Random-effects ordinal regression model for multilevel analysis. *Biometrics* **50** (4): 933 − 944. DOI: http://dx.doi.org/10.2307/2533433
15. Polit D. (1996). *Data Analysis and Statistics for Nursing Research*. Appleton and Lange, Stamford, USA.
16. Pinheiro J.C. & Bates D.M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics* **4** (1): 12 − 35.
17. Rashbash J., Steele F., Browne W. J. & Goldstein H. (2004). *A User's Guide to MLwiN*. University of Bristol, UK.
18. Roudenbush S.W., Yang M.C. & Yosef M. (2000). Maximum likelihood for generalized linear models with nested random effects via high order, multivariate laplace approximation. *Journal of Computational and Graphical Statistics* **9**(1): 141 − 157.
19. Spiegelhalter D.J., Best N.G., Carlin B.P. & Van Der Linde A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64** (4): 583 − 639. DOI: http://dx.doi.org/10.1111/1467-9868.00353

20. Steedman J. (1983). *Examination Results in Selective and Nonselective Schools: Findings from the National Child Development Study.* National Children's Bureau, London, UK.
21. Tinkalin T. (2005). Gender differences and high attainment (with discussion). *British Educational Research Journal* **29**(3): 307 − 325.
DOI: http://dx.doi.org/10.1080/01411920301854
22. Zeger S.L. & Karim M.R. (1991). Generalized linear models with random effects; a Gibbs Sampling approach. *Journal American Statistical Society* **86**: 79 − 102.
DOI: http://dx.doi.org/10.1080/01621459.1991.10475006
23. Zhang J. & Boos D.D. (1997). Generalized Cochran-Mantel-Haenszel Test Statistics for correlated categorical data. *Communications in Statistics - Theory and Methods* **26**(8): 1813 − 1837.
DOI: http://dx.doi.org/10.1080/03610929708832016