VARIABLE SELECTION IN LONGITUDINAL DATA WITH APPLICATION TO EDUCATION

by

Jongmin RA

(Under the direction of Cheolwoo Park and Woncheol Jang)

Abstract

Variable selection with a large number of predictors is a very challenging and important problem in multiple linear regression. However, relatively little attention has been paid to issues of variable selection in longitudinal data with application to education. This study examines data in which reading achievement of TOEIC measured for each quarter in a year is a response variables and other predictors such as gender, socioeconomic status (SES), and majors are used as predictors. Using this longitudinal educational data, we compare multiple regression, backward elimination, group least selection absolute shrinkage and selection operator (LASSO), and linear mixed models in terms of their performance in variable selection.

In our case study, the results show that four different statistical methods contain different sets of predictors in their models. The linear mixed model (LMM) provides the smallest number of predictors (4 predictors among a total of 19 predictors). In addition, LMM is the only appropriate method for the repeated measurement and is the best method with respect to the principal of parsimony. We also provide interpretation of the selected model by LMM in the conclusion.

Index words: Education, Group LASSO, Linear Mixed Model, Longitudinal Data, Marginal R^2, Multiple Regression

VARIABLE SELECTION IN LONGITUDINAL DATA WITH APPLICATION TO EDUCATION

by

Jongmin RA

B.A., Hannam University, Korea, 2000

M.A., Southern Illinois University, Edwardsville, IL, 2003

POST BACCALAUREATE., Southern Illinois University, Edwardsville, IL, 2003

A Thesis Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Master of Science

Athens, Georgia

2011

VARIABLE SELECTION IN LONGITUDINAL DATA WITH APPLICATION TO EDUCATION

by

Jongmin RA

Approved:

Major Professors: Cheolwoo Park
Woncheol Jang

Committee: Abhyuday Mandal
Yehua Li

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2011

Dedication

Mysterious Joviality

Mirage whirling within hearts
Inserts puffs of loving touch to souls
Now, in the hours of grey the moment

Jadegreen eyes whispering lovable smiles
Usher us into mosaic roads
Numbed hearts stuck in the dark
Glide into the rainbow

Seize the moment
Of missing, loving touch
Never forget the moments
Gracious tears flowed into sun,

and to whom I loved.

Acknowledgments

I would like to sincerely thank my advisors, Professor Cheolwoo Park and Woncheol Jang for their guidance, understanding, patience, and supports for my thesis. I have been amazingly fortunate to have advisors who gave me the freedom to explore on my own research interests, and at the same time all kinds of supports. I am deeply grateful to them for the long discussions that helped me sort out technical details of my works.

Table of Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

It is of great interest to find important predictors in explaining and predicting response variables (e.g., English achievement) in educational settings. For instance, the most common example is that repeatedly measured individual's reading achievement can be explained and predicted by various predictors: socioeconomic status (Krashen and Brown, 2005; Sui-Chu and Willms, 1996), effect of age (Krashen, 1982), gender difference and other different types of anxiety on reading achievement (Aida, 1994; Goshi, 2005; Horwitz, Horwitz and Cope, 1986; Matsuda and Gobel, 2004). In parallel, it is important to choose appropriate predictors in the model because including irrelevant predictors increases standard error of parameter estimates and reduces efficiency of the estimates whereas omitting relevant variables results in biased coefficient estimates for the predictors (Menard, 1995). Despite many previous studies investigating how predictors affect achievement, the question of choosing appropriate statistical methods to select important and meaningful predictors under repeated measurements is still not clearly answered.

The characteristic of a longitudinal study is that individuals are measured repeatedly through different time points. Longitudinal data require special statistical methods because the set of observations on the same individual tends to be intercorrelated and can be explained by both fixed and random effects. In random effects, levels can be thought of as being randomly sampled from a population. Variation in response variables across levels of random factors can be estimated and assessed.

As longitudinal data are common in educational settings, the linear mixed model (LMM) has emerged as an effective approach since it can model within- and between-subject heterogeneity (Vonesh, Chinchilli and Pu, 1996). LMM also attempts to account for within-subject dependency in the multiple measurements by including one or more subject-specific variables in a regression model (Laird and Ware, 1982; Gilks, Wang, Yvonnet and Coursaget, 1993).

In terms of model selection criteria, many of previous studies in the language assessment use traditional statistics, such as R^2, Akaike's information criterion (AIC: Akaike, 1974), and the Bayesian information criterion (BIC: Schwarz, 1978), for model selection without considering data structure. Parallel with this, model selection criteria for linear mixed models have received little attention (Orelien and Edwards, 2008; Vonesh et al., 1996). However, several studies (Vonesh and Chinchilli, 1997; Vonesh, et al., 1996, Zheng, 2000) recently suggest model fit indices which are useful for mixed effect models. More specifically, studies in Vonesh and Chinchilli (1997) and Vonesh, et al.(1996) show that marginal R^2 is preferred when only fixed-effect components are involved in the predicted values, but conditional R^2 is preferred for random effects (Vonesh et al., 1996).

1.1 STATEMENT OF PROBLEMS

It is not uncommon to collect a large number of predictors to model an individual's reading achievement more accurately in educational and psychological fields. Thus, it is fundamental to select meaningful variables in multivariate statistical models (Zhang, Wahba, Lin, Voelker, Ferris, Klein, and Klein, 2004) to increase prediction accuracy and to provide better understanding of concepts. However, it is challenging to select important variables when a response variable is measured repeatedly over a certain period of time because it is known that the selection process of statistically significant variables is hindered by the correlation among the repeated measurements.

Furthermore, classical variable selection methods, such as the forward selection and the backward elimination methods are time consuming and, unstable, and sometimes unreliable for making inferences. Although there is a great deal of extent research examining issues of variable selection in linear regression, little research has been done investigating how differently and similarly different statistical methods perform within a longitudinal data.

1.2 THE PURPOSE OF THE STUDY

This study aims to investigate how similarly and differently various statistical methods perform in the presence of the repeated measurements in the data. Hence, this study compares four different statistical methods, multiple regression, backward elimination, the group LASSO, and linear mixed models, using a test of English as International Communication (TOEIC) data as individuals' reading achievement. For the linear mixed model, marginal R^2 for remaining variables is used to provide a better understanding of the impacts of selected predictors in the longitudinal data.

This study will be useful both to those whose interests focus on assessing effects of predictors and to those who want to build a flexible but parsimonious model for prediction in a longitudinal data setting. The next chapter presents statistical methods employed in this study. Chapter 3 describes a case study in education. Chapter 4 summarizes the results of statistical tests, and Chapter 5 gives conclusions.

Chapter 2

VARIABLE SELECTION

2.1 LINEAR REGRESSION

Multiple linear regression is a flexible method of data analysis that may be appropriate whenever a response variable is to be examined in relation to any other predictors (Cohen, Cohen, West, and Aiken, 2003). For instance, if a multiple regression method is used for predicting and explaining an individual's English achievement, variables such as gender, age, and socio-economic status (SES) might all contribute toward English achievement. If all of these variables are collected, it is possible to see which of these variables give results in the most accurate prediction of English achievement. A multiple regression method for predicting English achievement, Y, with the observed data $(X_{1i}, \cdots, X_{pi})$, $i = 1, \cdots, n$, is as follows:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + \epsilon_i.$$

This equation shows the relationship between p predictors and a response variable Y, all of which are measured simultaneously on the subject. This method is called linear because the effects of the various predictors are treated as additive. For example, Y_i is composed of a linear combination of regression parameters. The regression parameters, $\beta_1, \ldots, \beta_p$, are weights which are determined from the sample data and used to optimally predict Y_i. The intercept, β_0, is a scaling constant which absorbs the differences in the scales used to measure Y and p predictors. The ϵ_i represents the error term. For a given set of data, the coefficients for β_j are determined mathematically to minimize the sum of squared deviations between predicted $\widehat{Y}$ and the actual Y values.

Also, much efforts has been put to estimate the performance of different models and choose the best one by using fit indices such as AIC (Akaike, 1973), BIC (Schwarz, 1978), Mellow's C_p (Mallows, 1973), R^2, and adjusted R^2. AIC and BIC are based on the penalized maximum likelihood estimates. AIC is defined as -2 $\times$ log(L)+2p, where log(L) is the log likelihood function of the parameters in the model evaluated at the maximum likelihood estimator while the second term 2p is a penalty term for additional parameters in the model. Therefore, as the number of independent variables p included in the model increases, the first term decreases while the penalty term increases. Conversely, as variables are dropped from the model, the lack of fit term increases while the penalty term decreases. BIC is defined as -2 $\times$ log(L) + p $\times$ log(n). The penalty term for BIC is similar to AIC but uses a multiplier of log(n) instead of a constant 2 by incorporating the sample size n. In general, AIC tends to choose overly complex models when sample size is large and BIC tends to choose overly simple model when sample size is small and also choose the correct model when sample size approaches infinity.

Mallow's C_p is also commonly used to investigate how well a model fits data and can be defined as $C_p = \frac{SSE_p}{\widehat{\sigma^2}} + 2p - n$. In this equation, $\hat{\sigma}$ represents the estimate of σ^2 and SSE_P is defined as $\sum_{i=1}^{n}(\mathbf{Y}_i - \sum_{j=1}^{p} \mathbf{X}_{ji}\widehat{\boldsymbol{\beta}})^2$, where $\widehat{\boldsymbol{\beta}}$ is the estimator of $\boldsymbol{\beta}$. Mallow's C_p is calculated for all possible subset models. The model with the smallest value of C_p is deemed to be the best linear model. As the number of independent variables p increases, an increased penalty term 2p is offset with a decreased SSE.

Another commonly used fit indices for model selection are R^2 or adjusted R^2. Either R^2 or adjusted R^2, in general, is often used to investigate how well the model fits data. Both R^2 and adjusted R^2 represent the percentage of the variability of the response variable that is explained by the variation of predictors. R^2 is a function of the total sum of squares (SST) and SSE, and the formula is given by $(1 - \frac{SSE}{SST})$.

Adjusted R^2 takes into account the degrees of freedom used up by adding more predictors. Even though adjusted R^2 attempts to yield a more robust value to estimate R^2, there is little difference between adjusted R^2 and R^2 when a large number of predictors are included in a model. Adjusted R^2 is computed using the formula, $1-((1-R^2)\times\frac{(n-1)}{(n-p-1)})$. When the number of observations is very large compared to the number of predictor variables in a model, the value of R^2 and adjusted R^2 will be much closer because the ratio of $(n-1)/(n-p-1)$ will approach 1.

Despite the practical advantages of using a multiple regression method, it is difficult to build multiple regression models for repeatedly measured responses. Therefore, a multiple regression method is not appropriate for correlated response variables as in longitudinal data without accounting for correlation within response variables.

2.2 BACKWARD ELIMINATION

It is a common and important practice in multiple linear regression to select relevant variables among a large number of predictors. A subset selection method is one of the most widely used variable selection approaches in which one predictor at a time is added or deleted based on the F statistic iteratively (Bernstein, 1989). Subset selection methods, in general, provide an effective means to screen a large number of variables (Hosmer and Lemeshow, 2000).

Since there is a possibility of emerging a suppressor effect in the forward inclusion method (Agresti and Finlay, 1986), the backward elimination method is usually preferred method of exploratory analysis (Agresti, 2002; Hosmer and Lemeshow, 2000; Menard, 1995) and follows three steps. First, obtains a regression equation which includes all p predictors. Second, conducts a partial F-test for each of the p predictors which indicates the significance of the corresponding predictor as if it is the last variable entered into the equation. Finally, selects the lowest partial F value and compares it with a threshold partial, F_α, the value set equal to some predetermined level of significance, α. If the smallest partial F is less than F_α, then deletes that variable and repeats the process for $p-1$ predictors. This sequence continues until the smallest partial F at any given step is greater than F_α. The variables that are remained in the model are considered as significant predictors. In general, the backward elimination method is computationally attractive and can be conducted with an estimation accuracy criterion or through hypothesis testing.

The backward elimination method, however, is far from perfection. This method often leads to locally optimal solutions rather than the globally optimal solution. Also, the backward elimination method yields confidence intervals for effects and predicted values that are far too narrow (Altman and Andersen, 1989). The degree of correlation among the predictors affects the frequency with which authentic predictors find their way into the final model in terms of frequency of obtaining authentic and noise predictors (Derksen and Keselman, 1992).

More specifically, the number of candidate predictors affects the number of noise predictors that gains entry to the model. Furthermore, it is well known that the backward elimination method will not necessarily produce the best model if there are redundant variables (Derksen and Keselman, 1992). It also yields R^2 values that are badly biased upward and has severe problems in the presence of collinearity. Since the backward elimination method gives biased regression coefficient estimates, they need to be shrunk because the regression coefficients for remaining variables are too large. Besides well-known inherent technical problems, it is time consuming when a large number of predictors are included in the model and cumbersome to choose appropriate variables manually when categorical variables are included in the model as a dummy variable.

2.3 GROUP LASSO

A number of shrinkage methods are developed to overcome the inherent problems shown in traditional variable selection methods (Bondell and Reich, 2008; Forster and George, 1994; George and McCulloch, 1993; Tibshirani, 1996). Among many suggested shrinkage methods, the least absolute shrinkage and selection operator (LASSO) suggested by Tibshirani (1996) is one of well-known penalized regression approaches (Bondell and Reich, 2008; Meier, van de Geer and Bühlmann, 2008; Tibshirani, 1996). The LASSO method minimizes the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant (Tibshirani, 1996). It is also well known that all the variables in LASSO type methods such as the standardized LASSO and group LASSO (Yuan and Lin, 2006) need to be standardized before performing analysis. The LASSO method is defined as follows:

$$\boldsymbol{\beta}_{LASSO}(\lambda) = \arg\min_{\boldsymbol{\beta}} \sum_{i}^{n} (\mathbf{Y} - \sum_{j=0}^{p} \mathbf{X}_{ji}\beta)^2 + \lambda \sum_{j=1}^{p} \|\beta_j\|$$

In this equation, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ and λ is a penalty or a tuning parameter. The parameter λ controls the amount of shrinkage that is applied to the estimates. The solution paths of LASSO are piecewise linear, and thus can be computed very efficiently. The variables selected by the LASSO method are included in the model with shrunken coefficients. The salient feature of the LASSO method is that it sets some coefficients to be 0 and shrinks others. Furthermore, the LASSO method has two advantages compared to the traditional estimation method. One is that it estimates parameters and select variables simultaneously (Tibshirani, 1996; Fan and Li, 2001). The other is that the solution path of the LASSO method moves in a predictable manner since it has good computational properties (Efron, Hastie, Johnstone, and Tibshirani, 2004). Thus, the LASSO method can be used for high-dimensional data as long as the number of predictors, p is smaller than or equal to n, $p \leq n$.

The LASSO method, however, has some drawbacks (Yuan and Lin, 2006). If the number of predictors, p, is larger than the number of observations, n, the LASSO method at most select n variables due to the nature of the convex optimization problem. Also, the LASSO method tends to make selection based on the strength of individual derived input variables rather than the strength of groups of input variables, often resulting in selecting more variables than necessary. Another drawback of using the LASSO method is that the solution depends on how the variables are orthonormalized. That is, if any variable X_i is reparameterized through a different set of orthonormal contrasts, there is a possibility of getting different set of variables in the solution. This is undesirable since solutions to a variable selection and estimation problem should not depend on how the variables are represented. In addition, the LASSO solutions bring another problem when categorial variables in the model are included to represent the effect of a single nominal variable. The LASSO method treats categorical variables as an individual variables rather than a group (Meier, van de Geer and Bühlmann, 2008).

A major stumbling block for the LASSO method is that if there are groups of highly correlated variables, it tends to arbitrarily select only one from each group. This makes models difficult to interpret because predictors that are strongly associated with the outcome are not included in the predictive model.

To remedy the shortcomings of the LASSO method, Yuan and Lin (2006) suggested the group LASSO in which an entire group of predictors may drop out of the model depending on λ. The group LASSO is defined as follows:

$$\boldsymbol{\beta}_{LASSO}(\lambda) = \arg\min_{\boldsymbol{\beta}}(\|\mathbf{Y} - \sum_{\ell=1}^{L} \mathbf{X}_{\ell}\boldsymbol{\beta}_{\boldsymbol{\ell}}\|^2 + \lambda \sum_{\ell=1}^{L} \sqrt{P_\ell}\|\boldsymbol{\beta}_\ell\|_1).$$

In this equation, $\mathbf{X}_\ell$ represents the predictors corresponding to the ℓth group, with corresponding coefficient subvector, and $\boldsymbol{\beta}_\ell$. P_ℓ takes into account for the different group sizes. If $\mathbf{x} = (x_1, \ldots, X_k)^T$, then, $\|\mathbf{x}\|_1^2 = \sum_{i=1}^{k} X_i^2$. The group LASSO acts like the LASSO at the group level; depending on λ, an entire group of predictors may drop out of the model. The group LASSO takes two steps. First, a solution path indexed by certain tuning parameter is built. Then, the final model is selected on the solution path by cross validation or using a criterion such as the Mallow's C_p.

This gives group LASSO tremendous computational advantages when compared with other methods. The group LASSO makes statistically insignificant variables become zero by incorporating shrinkage as the standard LASSO does. Overall, the group LASSO method enjoys great computational advantages and excellent performance, and a number of nonzero coefficients in the LASSO and the group LASSO method are an unbiased estimate of the degree of freedom (Efron et al., 2004).

Even though the group LASSO is suggested for overcoming drawbacks of the standard LASSO, the group LASSO method still has some limitations. For example, the solution path of the group LASSO is not piecewise linear which precludes the application of efficient optimization methods (Efron et al., 2004). It is also known that the method tends to select a larger number of groups than necessary, and thus includes some noisy variables in the model (Meier et al., 2008). Furthermore, the group LASSO method is not directly applicable to longitudinal data and needs further study for being suitable for the repeated measurement.

2.4 LINEAR MIXED MODEL

The linear mixed model (LMM) is very useful in longitudinal studies to describe relationship between a response variable and predictors. LMM has been called in different fields. In economics, the term “random coefficient regression models” is common. In sociology, “multilevel modeling” is common, alluding to the fact that regression intercepts and slopes at the individual level may be treated as random effects of a higher level. In statistics, the term “variance components models” is often used in addition to mixed effect models, alluding to the fact that one may decompose the variance into components attributable to within-groups versus between-groups effects. All these terms are closely related, albeit emphasizing different aspects of LMM.

In the context of repeated measure, let $\mathbf{Y}_i$ is an $n_i \times 1$ vector of observations from the ith subject. Then, the mixed model (Laird and Ware, 1982) is as follows:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i.$$

In this model, $\mathbf{X}_i^T(X_{1i}, X_{2i}, \cdots, X_{pi})^T$, where $\mathbf{X}_i$ is an $n_i \times p$ fixed-effect design matrix whereas $\mathbf{Z}_i$ are $n_i \times q$ known constant design matrices. $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_p)^T$ is an p-dimensional vector and unknown coefficients of the fixed effects and $\mathbf{b}_i$ are q-dimensional vector and unknown individual-specific random effects. Here, $\mathbf{b}_i$ is assumed to be multivariate normally distributed with mean vector $\mathbf{0}$ and variance matrix $\boldsymbol{\Psi}$. Thus, the random effects vary by group. In addition, variance-covariance matrix $\boldsymbol{\Psi} = \mathrm{diag}(\Psi, \cdots, \Psi)$ should be symmetric and positive semi-definite (Laird and Ware, 1982).

The $\boldsymbol{\varepsilon}_i$ are $n_i \times 1$ vectors of error term and assumed to follow a normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Omega}$ which are same for all subjects. It is also commonly assumed that $\mathbf{\Omega}$ is diagonal and all diagonal values are equal, σ^2. However, instead of assuming equal variance in grouped data, it is possible to extend to allow unequal variance and correlated within-group errors. The vectors $\mathbf{b}_i$ and $\boldsymbol{\varepsilon}_i$ are assumed to be independent.

Once statistically significant predictors in the model are obtained, goodness-of-fit for LMM can be considered. Among different types of R^2 such as unweighted concordance correlation coefficient (CCC: Vonesh et al, 1996), and proportional reduction in penalized quasi-likelihood (Zheng, 2000), the marginal R^2 (Vonesh and Chinchilli, 1997) is easy to compute and interpret in that it is a straightforward extension of the traditional R^2 (Orelien and Edwards, 2008). We use the marginal R^2 in our analysis for selecting relevant variables, which is defined as follows:

$$R_m^2 = 1 - \frac{\sum_{i=1}^{n}(\mathbf{Y}_i - \hat{\mathbf{Y}}_i)'(\mathbf{Y}_i - \hat{\mathbf{Y}}_i)}{\sum_{i=1}^{n}(\mathbf{Y}_i - \bar{\mathbf{Y}}\mathbf{1}_{pi})'(\mathbf{Y}_i - \bar{\mathbf{Y}}\mathbf{1}_{pi})}.$$

Given the equation shown above, $\mathbf{Y}_i$, n of observations, is a observed response variable and $\widehat{\mathbf{Y}}_i$ is a predicted response variable. $\bar{\mathbf{Y}}$ is the grand mean of the Y_{ij}, and $\mathbf{1}_{pi}$ is an $p_i \times 1$ vector of 1's. This equation implies $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$ and considers only fixed effects. In addition, marginal R^2 models the average subject ($\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$) leads to the term average model R_m^2 (Vonesh and Chinchilli, 1997) where R_m^2 is the proportionate reduction in residual variation explained by the modeled response of the average subject. Thus, when important predictors in the model are not included, the values of marginal R^2 decrease sharply. If the random effects are excluded in the computation of the predicted values that lead to the residuals, the marginal R^2 is able to select the most parsimonious model.

Chapter 3

CASE STUDY

This study takes place in a public university in the Republic of Korea, between the years 2009 and 2010, over two semesters. Participating students (n=243) enrolled in an TOEIC class for four hours a week. The TOEIC dataset records 243 students and 19 predictors. Among 19 predictors, 12 are continuous: age, father's education level, mother's education level, SES, English study time, reading time, level of reading competence, level of computer skill, length of private tutoring, three mean-centered cognitive assessment scores (STAS: State and trait anxiety scale, FLCAS: Foreign language classroom anxiety scale, FRAS: Foreign language reading anxiety scale); and 7 are categorical: major, gender, experience of private tutoring, experience of having foreign instructors, living areas, length of staying at aborad, experience of staying English speaking countries. The waves 2, 3, and 4 data are collected every three months after collecting wave 1 data. The longitudinal nature of the data would be very useful in estimating the impact of changes in education. In addition, the efforts for avoiding test translation effect as well as for increasing empirical validity evidence are made. The three original cognitive assessments (STAS, FLCAS and FRAS) written in English are first translated into Korean version of them which are later re-translated into the English version of it. Later, careful comparison between two forms of three original cognitive assessments are scrutinized. FLCAS, one of three anxiety scales, is shown in Table A.5. The predictors are as follows:

1. MAJORS: ME - Medical major, NE - Nursing major, EB - E-Business major, PO - Police administration major, CE - Child Education major, CC - Child Care program,

TS - Tourism major, OT - Occupational Therapy major, EFL - English as a Foreign Language major (Reference program).

2. GENDER: GE - Female=1, Male=0.

3. AREA: AR1 - Seoul and Kyounggi, AR2 - Others, Chung Nam (Reference area).

4. PLACE (Countries in which students staying for studying) : PL1 - English Speaking Countries, PL2 - Non English Speaking Countries, No experience (Reference group).

5. EF (experience of having foreign instructors): EF - Yes=1, No=0.

6. TU (experience of having tutoring): TU - Yes=1, No=0.

7. AGE.

8. INCOME.

9. FAEDU (Father's educational level). 1 - Middle school, 2 - High school, 3 - College or University, 4 - Graduate school.

10. MOEDU (Mother's educational level). 1 - Middle school, 2 - High school, 3 - College or University, 4 - Graduate school.

11. ES (Studying time for English).

12. EM (Studying reading materials written in English).

13. ENL (Levels of Reading Competence in English).

14. TIME (Length of staying in abroad).

15. TTIME (Length of having private tutoring).

16. COM (Level of Computer Skills). 1 - Very low, 2 - Low, 3 - Average, 4 - High, 5 - Very high.

17. STAI (State Trait Anxiety Inventory). 1 - Strongly disagree, 2 - Disagree, 3 - Neutral, 4 - Agree, 5 - Strongly agree.

18. FLCAS (Foreign Language Classroom Anxiety Scale). 1 - Strongly disagree, 2 - Disagree, 3 - Neutral, 4 - Agree, 5 - Strongly agree.

19. FRS (Foreign Language Reading Anxiety Scale). 1 - Strongly disagree, 2 - Disagree, 3 - Neutral, 4 - Agree, 5 - Strongly agree.

20. Y (Reading Achievement): Y1 - (Reading Achievement score for the first wave), Y2 - (Reading Achievement score for the second wave), Y3 - (Reading Achievement score for the third wave), Y4 (Reading Achievement score for the fourth wave).

Chapter 4

RESULTS

4.1 DESCRIPTIVE ANALYSIS

For descriptive analysis, frequencies and percentages of all variables are calculated. Among these variables, only categorical variables are shown in Table A.1. As noted from the table, there are 9 different majors and similar number of students are participated in this study except two majors (CE and OT) which consist of less than 10 % of total sample sizes, respectively. Also, there are the smallest number of students ($n = 14$) in Child Education major compared to other majors. Relatively a large number of students ($n = 135$) from the ChungNam ares are participated. Concerning experiences of having classes with foreigner instructors, about 44.9% of students do not have any experience. About 15.7% of them have experience studying abroad and 29.2% are male. Furthermore, almost 60 % of students never have a private tutoring.

For continuous variables, the mean and standard deviation of continuous variables are calculated and more detailed information is available in Table A.2. In terms of outcomes across 4 wave points, reading scores of TOEIC are increased as time increases. However, scores of TOEIC are slightly dropped between wave 2 and wave 3. The average age of students is 20.11 years old. The average education level of fathers (3.42) is little higher than that of mothers (3.13). Furthermore, the significantly different TOEIC scores across four waves are shown among different majors. For instance, Figure 4.1 shows that students in Medical major has high initial TOEIC scores compared to those in Child Care major. Fore more detailed information is available in Table A.2.

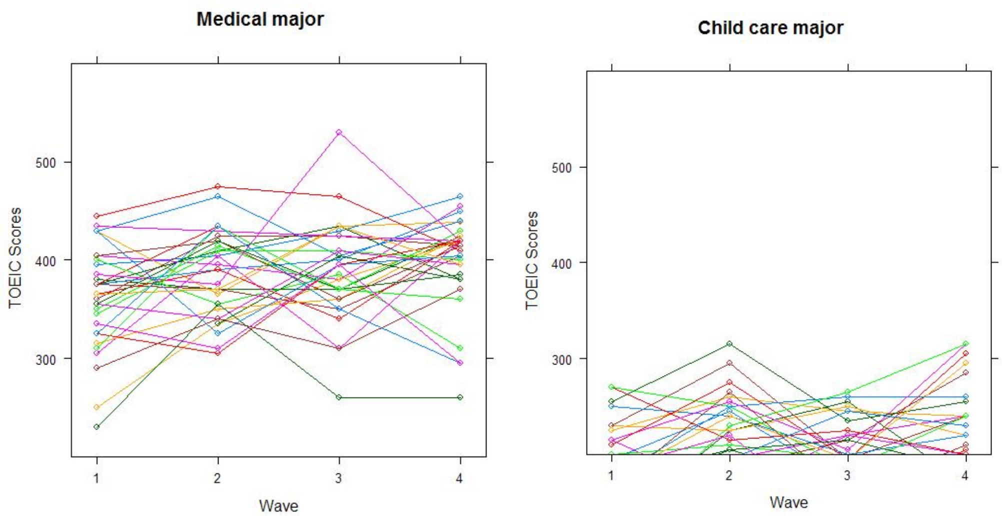

Figure 4.1: Individual TOEIC scores between Medical and Child Care majors

4.2 MULTIPLE REGRESSION

The existence of relationship between reading achievement and predictor variables across wave 1, wave 2, wave 3, and wave 4 is analyzed using four separate multiple regression runs. A EFL (English as a Foreign Language) major is chosen as a baseline. In terms of the significance of predictors, four separate analyses result in the consistent outcomes. The results show that there are four majors statistically significant: ME (Medical), NU (Nursing), EB (E-Business), and TS (Tourism). Furthermore, three majors (e.g., ME, NU, OT:Occupational Therapy) have higher positive coefficients across four wave time points compared to EFL. Rest of majors such as EB, PO (Police administration), CE (Child education), TS, and CC (Child care) have negative coefficients. Besides MAJOR variable, ENL (Levels of Reading Competence in English) is only one covariate included in the model. It implies that ENL has statistically significant and positive relationship across four separate multiple-regression analyses (Table A.3).

4.3 BACKWARD ELIMINATION

To select important and meaningful predictors, separate backward elimination procedures are conducted for each wave. Nineteen predictors are entered into the full model. As shown in Table A.4, there are five majors (ME, NU, EB, TS, and CC) statistically significant across four wave points. Besides MAJOR variable, there are seven significant predictors across four separate analyses; two variables at the first and second wave points, one variable at the third, and six variables at the fourth wave point. Among seven selected predictors, only one categorical variable, GEN (Gender) is included in the model at the fourth wave point. Among statistically significant continuous predictors such as FAEDU (Father education level), MOEDU (Mother education level), ES (Studying time for English), EM (Studying materials written in English), STAI (State Trait Anxiety Inventory) and ENL, the predictor, ENL, is the only one variable which is statistically significant across four separate backward elimination procedures. The interesting point is that four separate backward elimination procedures contain different sets of predictors in the model. It might imply that the backward elimination method is not suitable for dealing with the repeated measurement.

4.4 GROUP LASSO

When compared to the backward elimination method, the group LASSO contains more predictors in the model (Table A.4). The four separate group LASSO procedures contain different types of predictors. Besides MAJOR, total 17 predictors are included across four separate models; 14 variables are selected in the first wave, then 10 variables in the second and the third waves, and 9 variables in the fourth wave. Compared to multiple regression and backward elimination methods, the group LASSO includes more categorical variables, such as area, place, and length of staying abroad in the finalized model. This is not surprising as discussed in Section 2.3. Besides MAJOR variable, variables such as age, EM, and ENL, and STAI were included across the four separate models. However, The results show that four separate group LASSO methods contain different sets of predictors in the model. This might suggest inappropriateness of using the group LASSO to the repeated measurement.

4.5 LINEAR MIXED MODEL

The results show that all the majors and four continuous explanatory variables (e.g., MOEDU, ES, EM, and ENL) are included in LMM (Table A.3). The results reveal that TOEIC achievement is positively related with MOEDU ($p < 0.05$), ES ($p < 0.01$), and ENL ($p < 0.01$) but negatively related to EM ($p < 0.01$). Interesting finding is that EM affects the TOEIC achievement positively in univariate analysis but affects TOEIC achievement negatively when considered EM conditional on major, ENL, ES, and MOEDU variables.

Once selecting statistically significant predictors in the model, changes of marginal R^2 across all possible combinations of predictors are calculated in Table 4.1. Model 1 only contains MAJOR and ENL predictors in the model. Model 2 includes MAJORS and ENL with other three predictors (EM, ES, MOEDU) to identify which predictors mostly affect the marginal R^2. More specifically, marginal R^2 calculated for all possible combinations within Model 2 are also considered. However, there is less variations among all possible combinations. Values of the marginal R^2 for all possible combinations of the selected predictors range from 0.518 to 0.527. Model 3 contains MAJOR, ENL, ES, EM, and MOEDU predictors selected from the LMM. Model 4 includes all 19 predictors in the model. Then, marginal R^2 is calculated for each model. Values of three different marginal R^2s for Model1, Model 2, Model 3 and Model 4 were 0.513, 0.527, 0.539, and 0.544 respectively.

Changes of marginal R^2 for Model 1, Model 2, Model 3, and Model 4 are graphically presented (Figure 4.2). As shown in Figure 4.2, there is less changes of marginal R^2 (0.005) between Model 4 including 19 predictors and Model 3 including 5 predictors. However, compared to changes of marginal R^2 from Model 4 to Model 3, changes of marginal R^2 from Model 3 to Model 2 is relatively large, 0.012. This results suggest that four continuous variables (ENL, EM, ES, MODED) should be included in the model. LMM is somewhat different from other

Table 4.1: Marginal R^2 for All Possible Combinations

MODEL	PREDICTORS	Marginal R^2
MODEL 1	MAJOR, ENL	0.513
MODEL 2	MAJOR, ENL, EM	0.518
	MAJOR, ENL, ES	0.521
	MAJOR, ENL, MODEDU	0.519
	MAJOR, ENL, EM, ES	0.527
	MAJOR, ENL, EM, MOEDU	0.527
	MAJOR, ENL, ES, MODEDU	0.527
MODEL 3	MAJOR, ENL, ES, EM, MOEDU	0.539
MODEL 4	ALL PREDICTORS	0.544

statistical techniques mentioned in this study because repeated measurements across four waves are accounted for and analyzed at once in LMM.

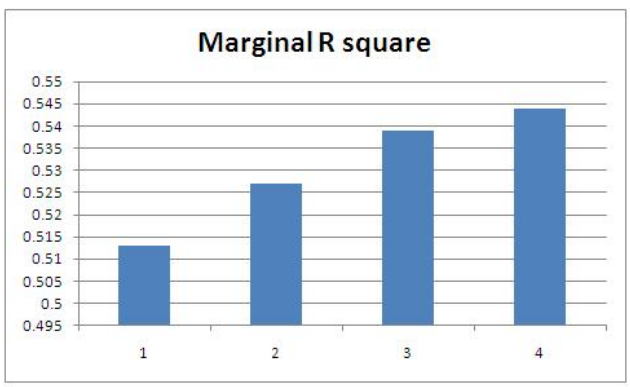

Figure 4.2: Changes of Marginal R^2

Chapter 5

CONCLUSIONS

This study examines the relation of TOEIC achievement and nineteen predictors under four different statistical methods. Different sets of predictors are selected in four different statistical methods. The results show that there is strong evidence to support the existence of relation between TOEIC achievement and some predictors included in this study. Without considering other predictors, there are much variation in TOEIC reading achievement among 9 different majors. As expected, students in medial program have high TOEIC scores compared to others in different majors. Thus, it is necessary to investigate predictors which affect growth of TOEIC scores.

The results also show that ENL (the level of English ability) is a useful variable to explain and predict TOEIC achievement. Interestingly, ENL is significant across four different statistical methods. It makes sense since the level of English ability is a crucial factor for English achievement. There is also another interesting finding in this study. The EM (Studying reading materials written in English) affects TOEIC reading achievement positively across four waves. However, when negative relationship between EM and TOEIC achievement has emerge when considered EM predictor conditional on other predictors (MAJOR, ENL, ES, MOEDU) in the model.

The results obtained from LMM reveal that there is little variation in the values of marginal R^2 across all possible combinations of predictor variables included in the final model. Among four different statistical methods, LMM seems to be most effective and useful to build a parsimonious model with important and meaningful predictors because it takes into account the repeated measurements, is flexible, and powerful to analyze balanced and unbalanced grouped data. However, these results must be regarded as very tentative and inconclusive because this is a search for plausible predictors, not a convincing test of any theory. Further development based on these results would require replication with other data and explanation of why these variables appear as predictors of continuity of achievement.

Moreover, this study has some limitations. Besides simply finding important variables, it is necessary to deal with other considerations such as optimal size of variables, interaction effects, and ratio of variables and observations (O'Hara and Sillanpaa, 2009; Zhang and Xu, 2005). Another limitation is that the best-fit model among four statistical models is not pursued since the objective of this research is to test hypotheses based on theories.

Concerning the LASSO methods, the group LASSO method enjoys great computational advantages and excellent performance, and a number of nonzero coefficients in the LASSO and the group LASSO method are an unbiased estimate of the degree of freedom (Efron et al., 2004). However, it is necessary to consider the LASSO method in the hierarchical structure for further studies since experiment and survey designs should be included in the model. Then, the LASSO method in an LMM framework is useful to explain random effects. Despite the limitations listed above, this thesis contributes to the field of education as a better way of explaining of relationship between personal predictors and English achievement.

Appendix A

Tables

Table A.1: Descriptive Statistics for Categorical Variables

Variables		N	%
Major	Medical	31	12.8 %
	Nursing	35	14.1 %
	EFL	25	10.3 %
	Child Care	25	10.3 %
	Child Education	14	5.8 %
	Tourism	32	13.2 %
	Occupational Therapy	23	9.5 %
	E−BUSINESS	31	12.8 %
	Police Administration	27	11.1 %
Gender	Female	171	70.4 %
	Male	72	29.6 %
Area	Seoul and Kyounggi	52	21.4 %
	ChungNam	135	55.6 %
	Others	56	23.0 %
Place	Non−English Speaking Countries	27	11.1 %
	English speaking Countries	11	4.5 %
	No Travel	205	84.4 %
EF	Yes	134	55.1 %
	No	109	44.9 %
Tutor	Yes	95	39.1 %
	No	148	60.9 %

Table A.2: Descriptive Statistics for Continuous Variables

Variables	Mean	S.E	Skewness	Kurtosis
Age	20.05	0.76	3.21	16.64
Income	2.55	1.14	0.85	0.03
FAEDU	3.42	0.77	0.45	-0.18
MOEDU	3.13	0.55	0.06	0.11
ES	1.09	0.29	2.79	5.81
EM	1.02	0.16	6.61	36.29
ENL	3.27	0.65	-0.33	-0.71
TIME	0.19	0.92	2.68	6.31
TTIME	7.19	12.61	2.12	4.68
COM	2.05	0.57	0.01	0.14
STAI	115.53	12.42	0.74	1.28
FLCAS	98.03	10.74	-0.13	1.28
FLRAS	90.89	11.97	-0.08	0.71
Y1	234.69	80.68	0.47	-0.21
Y2	274.94	75.11	0.23	-0.34
Y3	264.75	80.63	0.18	-0.63
Y4	284.03	80.57	0.16	-0.74

Table A.3: Comparison of Estimated Coefficients for Regression and Linear Mixed Methods

Variables	RE1	RE2	RE3	RE4	LMM
Intercept	46.63	188.96	101.80	280.88	178.20
	(101.70)	(99.83)	(103.08)	(101.96)	(64.72)
ME	99.27**	80.27**	91.35**	100.86**	96.66**
	(16.35)	(16.05)	(16.58)	(16.39)	(8.83)
NU	31.69**	38.65**	43.52**	46.48**	43.96**
	(13.14)	(12.89)	(13.31)	(13.17)	(7.28)
EB	-66.32**	-52.68**	-44.39**	-49.60**	-51.25**
	(14.02)	(13.76)	(14.21)	(10.05)	(7.56)
PO	-24.36	-41.17*	-51.97**	-43.27	-26.46**
	(13.59)	(13.34)	(13.77)	(13.62)	(7.64)
CE	-6.01	-27.63	-29.16	-16.92	-18.00*
	(16.36)	(16.06)	(16.59)	(16.41)	(9.16)
TS	-59.65**	-38.28**	-56.19**	-42.91**	-45.43**
	(13.57)	(13.32)	(13.76)	(13.61)	(7.48)
OT	6.71	29.67*	13.21	22.86	16.17*
	(14.50)	(14.23)	(14.70)	(14.54)	(8.20)
CC	-37.34	-37.21*	-20.53*	-56.16**	-39.66**
	(14.29)	(14.02)	(8.83)	(14.32)	(7.95)
MOEDU	12.89	6.61	2.06	2.00	10.19*
	(7.94)	(7.79)	(8.04)	(7.96)	(4.40)
ES	0.29	0.26	0.08	0.34	19.26**
	(0.19)	(0.19)	(0.19)	(0.19)	(6.38)
EM	-0.10	-0.62	-0.01	-0.63	-38.38**
	(0.37)	(0.37)	(0.38)	(0.37)	(12.42)
ENL	30.65**	35.07**	30.27**	28.70**	30.03**
	(6.14)	(6.03)	(6.23)	(6.16)	(3.304)

$*$ Coefficient Statistically significant at $\alpha = 0.05$
$**$ Coefficient Statistically significant at $\alpha = 0.01$

Table A.4: Comparison of Estimated Coefficents for Backward and Group LASSO

Variables	BK1	BK2	BK3	BK4	GLA1	GLA2	GLA3	GLA4
Intercept	82.34	198.59	151.81	257.44	-112.06	27.26	34.06	97.28
	(44.77)	(42.18)	(41.69)	(58.80)				
ME	46.21**	41.72**	46.59**	47.17**	164.35	146.58	163.77	230.71
	(7.45)	(7.06)	(7.83)	(8.04)				
NU	6.59***	19.12**	22.17**	22.83**	111.27	125.39	142.82	184.82
	(2.69)	(2.10)	(7.10)	(6.94)				
EB	-35.08**	-31.06**	-27.37**	-25.32**	89.17	100.48	92.11	138.05
	(6.86)	(6.58)	(7.31)	(7.40)				
PO	-7.13	-16.67**	-21.22**	-12.87	68.71	78.28	89.19	120.66
	(7.04)	(6.80)	(7.55)	(7.37)				
CE	-6.59	-16.44*	-13.36	-9.08	23.12	57.21	66.05	93.85
	(8.44)	(8.17)	(9.07)	(8.89)				
TS	-31.26**	-20.99**	-29.53**	-20.39**	36.25	40.22	53.82	77.66
	(6.79)	(6.56)	(7.29)	(7.15)				
OT	0.96	15.29	8.00	15.13	15.86	40.88	25.65	51.72
	(7.29)	(7.09)	(7.86)	(7.76)				
CC	-16.44*	-20.15	-29.27**	-25.60**	31.64	25.56	45.25	45.15
	(7.27)	(6.976)	(6.98)	(7.63)				
GEN				-11.03*	3.59	3.19	2.86	
				(4.46)				
AGE					4.48	2.70	4.42	1.42
AR1					12.58			-1.57
AR2					9.05			-1.31
PL1						2.53		
PL2						10.95		
INCOME							7.14	
FAEDU				13.20**	3.48			4.32
				(5.10)				
MOEDU	22.09**				18.85	5.17	3.98	
	(6.45)							
EF					11.73	0.23		
ES				25.35*	9.18			0.96
				(12.24)				
EM		-50.07*		-65.45**	-10.92	2.26	-10.66	6.26
		(21.30)		(25.33)				
ENL	30.27**	37.50**	27.80**	29.50**	13.59	20.43	11.15	4.50
	(5.62)	(5.55)	(6.02)	(6.06)				
LEN1					-5.72	0.81	0.17	
LEN2					-0.45	10.66	0.51	
TU					-3.99		-6.38	
TTIME						-0.01		
COM					-10.69		-5.12	-1.38
STAI				-0.63*	0.78	0.36	0.28	-0.15
				(0.28)				
FLRAS					0.12			0.30

$*$ Coefficient Statistically significant at $\alpha = 0.05$

$**$ Coefficient Statistically significant at $\alpha = 0.01$

Table A.5: Foreign Language Classroom Anxiety Scale

1. I never feel quite sure myself when I am speaking in my foreign language class.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
2. I don't worry about making mistakes in language class	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
3. I tremble when I know that I'm going to be called on in language class.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
4. It frightens me when I don't understand what the teacher is saying in the foreign language.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
5. It wouldn't bother me at all to take more foreign language classes.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
6. During language class, I find myself thinking about things that have nothing to do with the course.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
7. I keep thinking that the other students are better at language than I am.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
8. I am usually at ease during tests in my language class.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
9. I start to panic when I have to speak without preparation in language class.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
10. I worry about the consequences of failing my foreign language class.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
11. I don't understand why some people get so upset over foreign language classes.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
12. In language class, I can get so nervous I forget things I know.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
13. It embarrasses me to volunteer answers in my language class.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
14. I would not be nervous speaking the foreign language with native speakers.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
15. I get upset when I don't understand what the teacher is correcting.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
16. Even if I am well prepared for language class, I feel anxious about it.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
17. I often feel like not going to my language class.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
18. I feel confident when I speak in foreign language class.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
19. I am afraid that my language teacher is ready to correct every mistake I make.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
20. I can feel my heart pounding when I'm going to be called on in language class.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
21. The more I study for a language test, the more confused I get.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
22. I don't feel pressure to prepare very well for language class.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
23. I always feel that the other students speak the foreign language better than I do.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
24. I feel very self-conscious about speaking the foreign language in front of other students.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
25. Language class moves so quickly I worry about getting left behind.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
26. I feel more tense and nervous in my language class than in my other classes.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
27. I get nervous and confused when I am speaking in my language class.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
28. When I'm on my way to language class, I feel very sure and relaxed.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
29. I get nervous when I don't understand every word the language teacher says.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
30. I feel overwhelmed by the number of rules you have to learn to speak a foreign language.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
31. I am afraid that the other students will laugh at me when I speak the foreign language.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
32. I would probably feel comfortable around native speakers of the foreign language.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
33. I get nervous when the language teacher asks questions which I haven't prepared in advance.	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree

Bibliography

[1] Agresti, A. (2002). *Categorical data analysis* (2nd ed.), Boboken, NJ: John Wiley & Sons.

[2] Agresti, A. & Finlay, B. (1986). *Statistical method for the social sciences* (2nd ed.), San Francisco, CA: Dellen.

[3] Aida, Y. (1994). Examination of Horwitz, Horwitz, and Cope's construct of foreign language anxiety; The case of students of Japanese. *The Modern Language Journal, 78*, 155–168.

[4] Akaike, H. (1973). Information theory and an extenstion of the maximum likelihood principle. In B. N. Petrov and F. Csaki (Eds), *Second international symposimum on information theory*, 267–281. Budapest: Academiai Kiado.

[5] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716–723.

[6] Altman, D. G., & Andersen, P. K. (1989). Bootstrap investigation of the stability of a Cox regression model. *Statistics in Medicine, 8*, 771–783.

[7] Bernstein, I., H. (1989). *Applied multivariate analysis*. New York: Springer-Verlag.

[8] Bondell, H. D., & Reich, B. J. (2008). Simultaneous regression shrinkage, variable selection and clustering of predictors with OSCAR. *Biometrics, 64*, 115–123.

[9] Cohen, J., Cohen, P., West S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.), Mahwah, NJ: Lawrence Erlbaum Association.

[10] Derksen, S., & Keselman, H. J. (1992). Backward, forward and stepwise automated subset selection algorithms. *British Journal of Mathematical and Statistical Psychology, 45*, 265–282.

[11] Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics, 32*, 407–489.

[12] Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association, 96*, 1348–1360.

[13] Foster, D. P., & George, E. I. (1994). The risk inflation criterion for multiple regression. *The Annals of Statistics, 22*, 1947–1975.

[14] George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association, 88*, 881–889.

[15] Gilks, W. R., Wang, C. C., Yvonnet, B., & Coursaget, P. (1993). Random effects models for longitudinal data using Gibbs sampling. *Biometrics, 49*, 441–453.

[16] Goshi, M. (2005). Foreign language classroom anxiety: How should classroom teacher deal with it? *Journal of the School of Marine Science and Technology, 3*, 61–66.

[17] Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*, 125–132.

[18] Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*, New York: John Wiley & Sons.

[19] Krashen, S. (1982). *Principles and practice in second language acquisition*, New York: Pergamon.

[20] Krashen, S., & Brown, C. L. (2005). The ameliorating effects of high socioeconomic status: A secondary analysis. *Bilingual Research Journal, 29*, 184–196.

[21] Laird, N., & Ware, J. H. (1982). Random effect models for longitudinal data. *Biometrics, 38*, 963–974.

[22] Matsuda, S., & Gobel, P. (2004). Anxiety and predictors of performance in the foreign language classroom. *System, 32*, 21–36.

[23] Meier, L., van de Geer, S., & Bühlmann, P. (2008). The group lasso for logistic regression. *Journal of Royal Statistical Society, B, 70*, 53–71.

[24] Menard, S. (1995). *Applied logistic regression analysis* (Sage University Papers Series on Quantitative Application in the Social Sciences, series no. 106) (2nd Ed.), Thousand Oaks, CA: Sage.

[25] Orelien., & Edwards, L. J. (2008). Fixed effect variable selection in linear mixed models using R^2 statistics. *Computational Statistics & Data Analysis, 52*, 1896–1907.

[26] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461–464.

[27] Sui-Chu, E. H., & Willms, J. D. (1996). Effects of parental involvement on eighth-grade achievement. *Sociology of Education, 69*, 126–141.

[28] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of Royal Statistical Society, B, 58*, 267–288.

[29] Vonesh, E. F., & Chinchilli, V. M. (1997). *Linear and Nonlinear models for the analysis of repeated measurement*, New York: Marcel Dekker.

[30] Vonesh, E. F., Chinchilli, V. M., & Pu, K. W. (1996). Goodness-of-fit in generalized nonlinear mixed-effects models. *Biometrics, 52*, 572–587.

[31] Yuan, M ., & Lin, Y. (2006). The composite absolute penalties family for grouped and hierarchical variable selection. *Journal of the Royal Statistical Society, B, 68*, 49–67.

[32] Zhang, H. H., Wahba, G., Lin, Y., Voelker, M., Ferris, M., Klein, R., & Klein, B. (2004). Variable selection and model building via likelihood basis pursuit. *Journal of the American Statistical Association, 99*, 659–672.

[33] Zheng, B. Y. (2000). Summarizing the goodness of fit of generalized linear models for longitudinal data. *Statistics in Medicine, 19*, 1265–1275.