A COMPARATIVE STUDY OF SOME VARIABLE SELECTION TECHNIQUES IN LOGISTIC REGRESSION

Ijomah Maxwell . A. & Nwali Obinna
Department of Mathematics/Statistics
University of Port Harcourt, **NIGERIA**

ABSTRACT

Several research works have studied on the performance of variable selection techniques in logistic regression but were limited to models without interaction. In this research, we considered a comparative study of some variable selection techniques in logistic regression for models with and without interaction. Newton Raphson iteration method was applied to obtain coefficients of the variables in the full model (model without interaction). The performance of each technique was judged by their Akaike Information Criterion (AIC) value and the value of the Area under Reciever's Operating Characteristic (AROC) curve. Our findings show that for models without interaction, the forward stepwise, backward stepwise and best subset methods gave same result. Also, for model with interaction, Best subset method outperformed the other two methods. The AROC also revealed that the model fitted using these three methods have an excellent discrimination ability.

1. INTRODUCTION

Logistic Regression is an approach to studying relationship among variables when the dependent variable is categorical (dichotomous, polytomous or ordinal). Binomial logistic regression or binary logistic regression is an aspect of logistic regression that deals with a dependent variable with dichotomous outcome (pass or fail, success or failure, dead or alive, etc). Statistical modeling is aimed at fitting a model with a minimized number of variables which gives a better description of the data and also results in numerical stability. Some commonly used methods for selecting variables in logistic regression include forward selection, backward elimination, stepwise selection, best subset selection, purposeful selection, tabu search, and Bayesian model averaging.

Wang et al. (2004) compared the performance of Bayesian Model Averaging method and stepwise selection method. Their work result in a conclusion that the Bayesian Model Averaging is better that the stepwise selection method. Saker at el. (2009) conducted a study which was aimed at selecting variables for fitting a model for the explanatory variable. They used both stepwise selection method and best subset selection method for variable selection. Their findings revealed that both methods gave same results, but they did not consider comparing the performance of these methods when there are interaction factors present in the model. Hosmer and Lemeshow (2000) also showed that both stepwise selection method and Best subset selection method selected same variables using the UIS data. They also did not consider variable selection when interaction factors are present.

In the absence of comparison of the performance of these three variable selection methods when

interaction factors are present in the model, we are studying the performance of these selection methods to identify the one which is more reliable in the presence of interaction using the model's AkaikeInformation Criterion (AIC) value and the Area Under the Reciever's Operating Characteristic (ROC) curveas a measure for this comparison.

This work is limited to the use of these variable selection techniques in Binary Logistic Regression and did not extend beyond models with two factor interactions.

2. METHODS

A model (without interaction) containing all variables is fitted and these three selection techniques are used to select variables that are considered important. Following this selection is a selection of variables using these three selection techniques when two factor interactions are present. The preferred technique resulting in the suitable fitted model is judged using the Akaike Information Criterion (AIC) and the Area under Receiver's Operating Characteristic (AROC) curve. We performed Newton Raphson Iteration to obtain the coefficient of the variables in the full model (model with all main effect factors). Best subset selection method was done with XLMINER and SPSS 20 was used for forward stepwise, backward stepwise selection and Area under Receiver's Operating Characteristic (AROC) curve. To compare the performance of Best Subset method to other methods, the set of variables resulting in a model with minimum AIC among the selected subsets of models is compared with the AIC resulting from the set of variables selected by other methods.

2.1 Data

The data is a primary data collected (using a questionnaire) from commercial motorcycle operator who carry out their commercial motorcycle operation within some (Aluu, Omagwa, Isiokpo and Elele) areas of Ikwerre Local Government Area. Data was collected from loading points in Aluu, Omagwa, Isiokpo and Elele using a questionnaire. Out of a total of 705 motorcycle operators who are members of the commercial motorcycle operator's union in Ikwerre Local Government Area, a total of 303 motorcycle operators took part in the survey and 274 filled their questionnaire correctly, while 29 of them had some issues with their response. The 274 motorcycle operators whose questionnaires were correctly filled were used as the cases for the study. There are seven independent variables with Crash Involvement as the dependent variable. Dummy variable coding method was used for coding the variables. The variables and variable coding are shown in Table 1.

Table 1. Categorical variable coding

Variable	Measurement/category	Parameter coding	
		1	2
Crash Involvement	Not-involved (0)	0	
	Involved (1)	1	
Motorcycle Ownership	Rented (0)	0	
	Owner-operator (1)	1	
Age	Less than 30 years (1)	0	0

	30-40 years (2)	1	0
	Above 40 years (3)	0	1
Possession of Valid Driver's License	Have no valid license (0)	0	
	Have valid license (1)	1	
Knowledge of Road Signs	Have no knowledge of road signs (0)	0	
	Have knowledge of road signs (1)	1	
Alcohol Intake	No (0)	0	
	Yes (1)	1	
Marital Status	Single (0)	0	
	Married (1)	1	
Educational Status	Have no formal education (1)	0	0
	Primary education (2)	1	0
	Secondary education and above (3)	0	1

2.2 Logistic Regression Model

Let N be the number of subjects/population in the dataset,

$X = (X_1, X_2, X_3, \dots, X_N)^T$, where X_i is a collection of the outcomes of the i^{th} subject associated with the k independent random variable and a constant ($x_{i0} = 1\ \forall i$) , i.e, $X_i = (x_{i0}, x_{i1}, x_{i2}, \dots, x_{ik})$

Let$Q = (Q_1, Q_2, Q_3, \dots, Q_N)^T$, where

Q_i , $i = 1,2,3, \dots, N$ are binomial random variables (dependent variables) with values one for success and zero for failure.

Let $X_{ij} = (X_{i1}, X_{i2}, \dots, X_{ik})^T$ be a column vector with the k independent variables as its elements.

Let$n = (n_1, n_2, n_3, \dots, n_N)^T$, where n_i denotes the number of observation for the i^{th} subject.

The probability of success occurring in the i^{th} population is

$$\pi(X_i) = E(Q_i = 1|X_i) = \frac{e^{\beta_0 x_{i0} + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}}}{1 + e^{\beta_0 x_{i0} + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}}} \quad (1.1)$$

The Multiple Logistic Regression Model is defined as

$$\pi(X) = \frac{e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}}} \quad (1.2)$$

β_0 is the constant term and $\beta_j, j = 1,2, \dots, k$ is the Logistic Regression coefficient for the kth variable.

The Multiple Logistic Regression Model with two factor interaction is defined as

$$\pi(X) = \frac{e^{\beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} + \sum_{j=1}^{k-1} \beta_{1j+1} X_{i1} X_{ij+1} + \cdots + \beta_{k-1k} X_{ik-1} X_{ik}}}{1 + e^{\beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} + \sum_{j=1}^{k-1} \beta_{1j+1} X_{i1} X_{ij+1} + \cdots + \beta_{k-1k} X_{ik-1} X_{ik}}},$$

(1.3)

where $\beta_{jj+1}, j = 1,2,3, \dots, k-1,$ is the Logistic Regression coefficient of the interaction between the $(k-1)$th variable and the kth variable.

The logit of (3.3) is defined as

$$G(X) = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} + \sum_{j=1}^{k-1} \beta_{1j+1} X_{i1} X_{ij+1} + \sum_{j=2}^{k-1} \beta_{2j+1} X_{i2} X_{ij+1} + \beta_{k-1k} X_{ik-1} X_{ik}$$

(1.4)

In (1.3) and (1.4), $X_{ik-1}X_{ik} = 0, if\ the\ (k-1)th$ variable and the kth variable are design variables of same factor.

2.3 Parameter estimation

The method of estimation adopted for the estimation of the Logistic Regression model parameters is the method of maximum likelihood.

Let the likelihood function be denoted by $L(\beta)$.

$$L(\beta) = \prod_{i=1}^{N} \binom{n_i}{Q_i} \pi(X_i)^{Q_i} \left(1 - \pi(X_i)\right)^{n_i - Q_i} \quad (1.5)$$

Although $\binom{n_i}{Q_i}$ is a constant term in the likelihood function and maximizing

$L(\beta) = \prod_{i=1}^{N} \pi(X_i)^{Q_i} \left(1 - \pi(X_i)\right)^{n_i - Q_i}$ gives same result as maximizing

$$L(\beta) = \prod_{i=1}^{N} \binom{n_i}{Q_i} \pi(X_i)^{Q_i} \left(1 - \pi(X_i)\right)^{n_i - Q_i}$$

since $\binom{n_i}{Q_i}$ does not contain $\pi(X_i)$, for

$n_i = 1\ and\ Q_i = 0\ or\ 1$, $\binom{n_i}{Q_i} = 1.$

Therefore,

$$L(\beta) = \ln\left[\prod_{i=1}^{N} \pi(X_i)^{Q_i} \left(1 - \pi(X_i)\right)^{1 - Q_i}\right] \quad (1.6)$$

$$\frac{\partial \ln L(\beta)}{\partial \beta_j} = \sum_{i=1}^{N} x_{ij}\left(Q_i - \pi(X_i)\right) \quad (1.7)$$

$$\frac{\partial^2 \ln L(\beta)}{\partial \beta_j \partial \beta_{j'}} = -\sum_{i=1}^{N} x_{ij} x_{ij'} \pi(X_i)\left(1 - \pi(X_i)\right) \quad (1.8)$$

TheNewton-Raphson equation for iteration is defined as

$$\beta^{(i+1)} = \beta^{(i)} + (X^T T X)^{-1} X^T \left(Q - \mu_Q\right) \quad (1.9)$$

$$where\ \left(\beta^{(i)}\right)^T = (\beta_0 \quad \beta_1 \quad \beta_2 \quad \cdots \quad \beta_K)$$

$$Q = \begin{pmatrix} Q_1 \\ Q_2 \\ Q_3 \\ \vdots \\ Q_N \end{pmatrix}, \mu_Q = \begin{pmatrix} \pi(X_1) \\ \pi(X_2) \\ \pi(X_3) \\ \vdots \\ \pi(X_n) \end{pmatrix},$$

$$T = \begin{pmatrix} t_1 & 0 & 0 & \cdots & 0 \\ 0 & t_2 & 0 & \cdots & 0 \\ 0 & 0 & t_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & t_N \end{pmatrix} \quad (1.10)$$

where $t_i = \pi(X_i)\left(1 - \pi(X_i)\right)$

and $X = \begin{pmatrix} x_{10} & x_{11} & x_{12} & & x_{1k} \\ x_{20} & x_{21} & x_{22} & \cdots & x_{2k} \\ x_{30} & x_{31} & x_{32} & & x_{3k} \\ & \vdots & & \ddots & \vdots \\ x_{N0} & x_{N1} & x_{N2} & \cdots & x_{Nk} \end{pmatrix}$

For the first step of the iteration ($i = 0$), we guess the elements of $\beta^{(i)}$ ($i.e.\beta^{(0)}$) and use our guess to find $\beta^{(i+1)}(i.e.\beta^{(1)})$. In the next iteration ($i = 1$), the result for β^{i+1} which is $\beta^{(1)}$ is substituted for $\beta^{(i)}$ in equation (1.9) and then we solve for$\beta^{(i+1)}$($i.e.\beta^{(2)}$).The iteration process continues until $\beta^{(i)} \cong \beta^{i+1}$. Note that at any step of the iteration, the elements of $\beta^{(i)}$ are used to solve for $\pi(X_i)$ which is further used for finding $(X^T TX)^{-1} and\ X^T(Q - \mu_Q)$ at that step.

2.3.1 Variance and covariance

The variance covariance matrix denoted by $\mathbb{V}$ is defined as

$$\mathbb{V} = (X^T TX)^{-1} \qquad (1.11)$$

The diagonal elements in $\mathbb{V}$ are the variance of the Logistic Regression coefficients while the off diagonal elements are the covariance. To compute t_i in T, we make use of our maximum likelihood estimates.

2.3.2 Standard Error and Confidence Interval

The standard error of the jth Logistic Regression coefficient denoted as $\widehat{SE}(\hat{\beta}_j)$ is defined as

$$\widehat{SE}(\hat{\beta}_j) = [Var(\hat{\beta}_j)]^{\frac{1}{2}} \qquad (1.12)$$

The $100(1-\alpha)\%$ confidence interval estimate of the kth Logistic Regression coefficient, denoted as CI is

$$\widehat{CI} = \hat{\beta}_j \pm Z_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{\beta}_j)} \qquad (1.13)$$

Where $Z_{1-\frac{\alpha}{2}}$ is the normal critical value of a two-tail test of size α.

2.4 Likelihood ratio test

The likelihood ratio test is used for comparing the likelihood ratio of one Logistic Regression model to another. Let G denote the statistic used for this comparison.

$$G = [-2(loglikelihood\ of\ the\ reduced\ model - loglikelihood\ of\ the\ full\ model)] \qquad (1.14)$$

$$\Rightarrow G = -2\ln\left[\frac{likelihood\ of\ the\ reduced\ model}{likelihood\ of\ the\ full\ model}\right] \qquad (1.15)$$

Supposing that X_{i1}, X_{i2}, X_{i3} are independent variables of a Logistic Regression model,

To test $H_0: \beta_1 = 0$, we fit

$G(X) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3}$ and

$G(X') = \beta_0 + \beta_2 X_{i2} + \beta_3 X_{i3}$and then compute the log-likelihood of each model.

Likelihood of the full model is$\prod_{i=1}^{N} \pi(X_i)^{Q_i}\left(1 - \pi(X_i)\right)^{1-Q_i}$, where

$$\pi(X_i) = \frac{e^{\beta_0 x_{i0} + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}}}{1 + e^{\beta_0 x_{i0} + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}}}$$

Also, the likelihood of the reduced model is$\prod_{i=1}^{N} \pi(X_i')^{Q_i}\left(1 - \pi(X_i')\right)^{1-Q_i}$, where

$$\pi(X_i') = \frac{e^{\beta_0 x_{i0} + \beta_2 x_{i2} + \beta_3 x_{i3}}}{1 + e^{\beta_0 x_{i0} + \beta_2 x_{i2} + \beta_3 x_{i3}}}$$

$$G = \left[-2\ln\left[\frac{\prod_{i=1}^{N} \pi(X_i)^{Q_i}\left(1-\pi(X_i)\right)^{1-Q_i}}{\prod_{i=1}^{N} \pi(X_i')^{Q_i}\left(1-\pi(X_i')\right)^{1-Q_i}}\right]\right]$$

If $P\left(\chi^2_{\ v} > G\right) < \alpha$, we reject the null hypothesis and conclude at α level of significance that the variable X_{i1}contributes significantly to the prediction of the dependent variable if it is included in the model.

2.5 Area under the Receiver Operating Characteristic (ROC) curve.

Classification is done using a cut-off point between zero and one which is chosen by the researcher. This cut-off point if not well selected may lead to an inaccurate classification. The area under the ROC curve gives a very good description of classification accuracy. To produce an ROC curve, we plot the probability of detecting a true signal (sensitivity) and false signal (1-specificity) for an entire range of cut-off points. The area under the ROC curve ranges from zero to one and it gives a measure of the model's ability to discriminate between those subjects who experience the outcome of interest versus those who do not.

According to Hosmer and Lemeshow (2000), if area under $ROC\ curve \approx 0.5$, this suggests no discrimination, if $0.7 \leq area\ under\ ROC\ curve < 0.8$, this is considered an acceptable discrimination if $0.8 \leq area\ under\ ROC\ curve \leq 0.9$, this is considered an excellent discrimination, if $area\ under\ ROC\ curve > 0.9$, this is considered an outstanding discrimination.

2.5.1 Estimate of area under ROC curve

Let N_c be the set of subjects with $Q_i = 1$ and N_C' the set of subjects with $Q_i = 0$.

Let $\mathbb{N}_c = N_c \times N_C' = \{n_1, n_1, n_3, \ldots, n_n\}$, where $n_i$in $\mathbb{N}_c$ is a pair of an element in N_c and an element in N_C', and n is the number of elements in N_c multiplied by the number of element in N_C' or the total number of pairs of subjects with $Q_i = 1$ and subjects with $Q_i = 0$.

For each pair, we compare the estimated odd of the subject with $Q_i = 1$ and the estimated odd for the subject with $Q_i = 0$.

Let us define$R = \{r_1, r_1, r_3, \ldots, r_n\}$, where r_i is a value assigned to n_i based on comparison of the estimated odd of subjects in n_i.

For eachn_i: $r_i = 1$ if the estimated odd of the subject with $Q_i = 1$ is greater than the estimated odd of the subject with $Q_i = 0; r_i = 0.5$ if the estimated odd of the subject with $Q_i = 1$ is equal to the estimated odd of the subject with $Q_i = 0; r_i = 0$ if the estimated odd of the subject with $Q_i = 1$ is less than the estimated odd of the subject with $Q_i = 0$.

The estimate of the area under the ROC curve is $\frac{\sum_i^n r_i}{n}$.

2.6 Stepwise Logistic Regression

In regression analysis, a collection of variables (independent) are studied to know the association of such variables with a particular variable (dependent). This collection of independent variables contains important and unimportant ones. Stepwise Regression is employed to carry out a stepwise selection procedure aimed at screening this collection of variables, and fitting several Logistic Regression equations simultaneously. The selection procedure is based on a statistical

algorithm which carries out a check for important variables, and either add or delete them in accordance with a decision rule. The importance of an independent variable depends on a measure of the significance of the coefficient of that variable. In Logistic Regression, significance of a variable is determined using the likelihood ratio chi-square test. The most important variable is usually the variable with the largest change in log-likelihood relative to the model not containing the variable. The steps for Forward Stepwise selection and Backward Stepwise selection are as follow:

2.6.1 Forward Stepwise selection

STEP 0

Let L_0 denote the log-likelihood of the intercept only model,

$L_{ij}^{(0)}$ denote the log-likelihood of the model containing the jth independent variable,

$G_{ij}^{0} = -2\left(L_0 - L_{ij}^{(0)}\right), P_{ij}^{(0)} = Pr\left(\chi^2(v) > G_{ij}^{(0)}\right)$, where $v = 1$ if X_{ij} is continuous and $k-1$ if X_{ij} is polytomous,

$P_{ie_1}^{(0)} = min\left(P_{ij}^{(0)}\right)$,

X_{ie_1} denote the most important variable at step zero,

P_E denotes the $\alpha - level$ to judge the importance of a variable

Compute $P_{ie_1}^{(0)}$ and move to step 1 if $P_{ie_1}^{(0)} < P_E$, if otherwise, terminate the process.

STEP 1

$L_{ie_1}^{(1)}$ denote the log-likelihood of the model containing X_{ie_1},

$L_{ie_1 ij}^{(1)}$ denote the log-likelihood of the model containing $X_{ie_1} and\ X_{ij}$,$j = 1,2,\dots,k, j \neq e_1$

$G_{ij}^{(1)} = -2\left(L_{ie_1}^{(1)} - L_{ie_1 ij}^{(1)}\right)$, be the likelihood statistic of the model containing X_{ie_1} and X_{ij} versus the model containing only X_{ie_1},

$P_{ij}^{(1)} = Pr\left(\chi^2(v) > G_{ij}^{1}\right)$,

X_{ie_2} denote the variable with minimum p-value when added to the model containing X_{ie_1},

$P_{ie_2}^{(1)} = min\left(P_{ij}^{(1)}\right)$. Compute $P_{ie_2}^{(1)}$ and proceed to step 2 if $P_{ie_2}^{(1)} < P_E$, stop if otherwise.

STEP 2

Let $L_{-ie_j}^{2}$ denote the log-likelihood of the model without X_{ie_j}

$G_{-ie_j}^{(2)} = -2\left(L_{-ie_j}^{(2)} - L_{ie_1 ie_2}^{(2)}\right)$, where $L_{ie_1 ie_2}^{(2)}$ is the likelihood of the model containing $X_{ie_1} and\ X_{ie_2}$,

$P_{-ie_j}^{(2)} = Pr\left(\chi^2(v) > G_{-ie_j}^{2}\right)$.

Let X_{ir_2} denote the variable that yields the maximum p-value when removed from the model containing $X_{ie_1} and\ X_{ie_2}$,

$P_{r_2}^{(2)} = max\left(P_{-ie_1}^{(2)}, P_{-ie_2}^{(2)}\right)$.

Remove X_{ir_2} if $P_{r_2}^{(2)} > P_R$, where P_R denote the $\alpha - level$ for removal of a variable. If $P_{r_2}^{(2)} < P_R$, X_{ir_2} remains in the model and then we fit a model containing

$X_{ie_1}, X_{ie_2}, and\ X_{ij}, j = 1,2,\dots,k, j \neq e_1, e_2.$

Let $L^{(2)}_{ie_1\, ie_2\, ij}$ denote the log-likelihood of the model containing$X_{ie_1}, X_{ie_2}, and\ X_{ij}, j = 1,2,\dots,k, j \neq e_1, e_2,$

$G^{(2)}_{ij} = -2\left(L^{(2)}_{ie_1 ie_2} - L^{(2)}_{ie_1 ie_2\, ij}\right),$

$P^{(2)}_{ie_3} = min\left(P^{(2)}_{ij}\right).$

Compute $P^{(2)}_{ie_3}$ and move to step 3 if$P^{(2)}_{ie_3} < P_E$, if otherwise, terminate the process.

STEP 3

The procedure in STEP 3 is similar to the procedure in STEP 2. The process continues in this manner until the last step, STEP S.

STEP S

The final step occurs when all k variables have entered the model or when all the variables that are in the model have p-values to remove less than P_R, and the variables not in the model have p-value to enter greater than P_E.

2.6.2 Backward Stepwise selection

STEP 0

Let L_f denote the log-likelihood of the full model,

$L^{(0)}_{-ij}$denote the log-likelihood of the model not containing the jth independent variable,

$G^{0}_{-ij} = -2\left(L^{(0)}_{-ij} - L_f\right),$

$P^{(0)}_{-ij} = Pr\left(\chi^2(v) > G^{(0)}_{-ij}\right), where\ v = 1\ if\ X_{ij}$ is continuous and $k - 1$ if X_{ij} is polytomous,

$P^{(0)}_{ir_1} = max\left(P^{(0)}_{ij}\right),$

X_{ir_1}denote the most unimportant variable at step zero,

P_R denote the $\alpha - level$ to judge the unimportance of a variable

Compute $P^{(0)}_{ir_1}$ and move to step 1 if$P^{(0)}_{ir_1} > P_R$, if otherwise, terminate the process.

STEP 1

$L^{(1)}_{-ir_1}$denote the log-likelihood of the model not containing X_{ir_1},

$L^{(1)}_{-ir_1 ij}$denote the log-likelihood of the model not containing $X_{ir_1} and\ X_{ij}$,$j = 1,2,\dots,k, j \neq r_1$

$G^{(1)}_{-ij} = -2\left(L^{1}_{-ir_1 ij} - L^{1}_{-ir_1}\right)$, be the likelihood statistic of the model not containingX_{ir_1} and X_{ij}versus the model not containing onlyX_{ir_1},

$P^{(1)}_{-ij} = Pr\left(\chi^2(v) > G^{1}_{-ij}\right),$

X_{ir_2}denote the variable with maximum p-value when removed from the model not containing X_{ir_1},

$P^{(1)}_{-ir_2} = max\left(P^{(1)}_{-ij}\right)$. Compute $P^{(1)}_{-ir_2}$ and proceed to step 2 if $P^{(1)}_{-ir_2} > P_R$, stop if otherwise.

STEP 2

Let $L^{(2)}_{ir_k}$denote the log-likelihood of the model containing X_{ir_k} $(k = 1,2)$ and other variables not

removed in step one.

$G_{ir_k}^{(2)} = -2\left(L_{-ir_1ir_2}^{(2)} - L_{ir_k}^{(2)}\right)$, where $L_{-ir_1ir_2}^{(2)}$ is the likelihood of the model not containing $X_{ir_1} and\ X_{ir_2}$,

$P_{ir_k}^{(2)} = Pr\left(\chi^2(v) > G_{ir_k}^2\right)$.

Let X_{ie_2} denote the variable that yields the minimum p-value when entered into the model not containing $X_{ir_1} and\ X_{ir_2}$,

$P_{e_2}^{(2)} = min\left(P_{ir_1}^{(2)}, P_{ir_2}^{(2)}\right)$.

Enter X_{ie_2} if $P_{e_2}^{(2)} < P_E$, where $P_E$denote the $\alpha - level$ for entering of a variable. If $P_{e_2}^{(2)} > P_E$, X_{ie_2} is removed from the model and then we fit a model not containing $X_{ir_1}, X_{ir_2}, and\ X_{ij}$, $j = 1,2,\dots,k, j \neq e_1, e_2$.

Let $L_{-ir_1ir_2\,ij}^{(2)}$ denote the log-likelihood of the model not containing$X_{ir_1}, X_{ir_2}, and\ X_{ij}, j = 1,2,\dots,k, j \neq r_1, r_2$,

$G_{-ij}^{(2)} = -2\left(L_{ir_1ir_2\,ij}^{(2)} - L_{ir_1ir_2}^{(2)}\right)$, $P_{-ij}^{(2)} = Pr\left(\chi^2(v) > G_{-ij}^{(2)}\right)$

$P_{-ir_3}^{(2)} = max\left(P_{-ij}^{(2)}\right)$.

Compute $P_{-ir_3}^{(2)}$ and move to step 3 if$P_{-ir_3}^{(2)} > P_R$, if otherwise, terminate the process.

STEP 3

The procedure in STEP 3 is similar to the procedure in STEP 2. The process continues in this manner until the last step, STEP S.

STEPS

The final step occurs when all k variables have been removed from the model or when all the variables that are not in the model have p-values to enter greater than P_E, and the variables in the model have p-value to be removed less than P_R.

2.7 Best subset Logistic Regression

Best subset logistic regression is another method used for variable selection. This method of variable selection searches for the best model among models with equal number of variables, based on some criterion (Mallow's C_q, AR^2, etc.).

Best subset selection method is used to select the model with the minimum Cq or the maximum AR^2 from the set of models with one variable, two variables, three variables, etc.

Hosmer et al. (1989) showed how to conduct best subset selection in logistic regression using any software capable of performing best subset linear regression analysis when weights are involved. To conduct best subset logistic regression using a linear regression program, it is required that we already know the coefficients of the logistic regression variables.

$$Z = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_N \end{pmatrix}$$

$\hat{\beta} = (X^T T X)^{-1} X^T T (X\hat{\beta} + T^{-1}(Q - \mu_Q))$

$= (X^T T X)^{-1} X^T T Z$, where

$$Z = (X\hat{\beta} + T^{-1}(Q - \mu_Q))$$

$$z_i = (1 \quad x_{i1} \quad x_{i2} \quad \cdots \quad x_{ik}) \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} + \frac{Q_i - \pi(X_i)}{\pi(X_i)(1 - \pi(X_i))}$$

Let $X = (\mathbb{X}_1 \quad \mathbb{X}_2)$, where $\mathbb{X}_1$ is an $N \times q + 1$ design matrix for the subset of q variables in the model and the constant, and $\mathbb{X}_2$ is an $N \times k - q$ matrix containing the remaining $k - q$.
Let us also partition the vector of coefficients for $\mathbb{X}_1$ and $\mathbb{X}_1$ as
$\beta^T = (\beta_1^T \quad \beta_2^T)$. The resulting information matrix (I) is
$\mathrm{I} = \begin{pmatrix} \mathrm{I}_{11} & \mathrm{I}_{12} \\ \mathrm{I}_{21} & \mathrm{I}_{22} \end{pmatrix}$, where $\mathrm{I}_{11} = (\mathbb{X}_1^{\mathrm{T}} \mathrm{T} \mathbb{X}_1)$, $\mathrm{I}_{12} = (\mathbb{X}_1^{\mathrm{T}} \mathrm{T} \mathbb{X}_2) = \mathrm{I}_{21}^{\mathrm{T}}$, and $\mathrm{I}_{22} = (\mathbb{X}_2^{\mathrm{T}} \mathrm{T} \mathbb{X}_2)$.
The estimator of the coefficient vector β_1 obtained from the linear regression of Z on $\mathbb{X}_1$ using weighted matrix V is
$\widetilde{\beta_1} = (\mathbb{X}_1^T T \mathbb{X}_1)^{-1} \mathbb{X}_1^T T Z$(1.16)
The vector of fitted values for $\mathbb{X}_1$ obtained using linear regression is
$\tilde{Z}(q) = \mathbb{X}_1 \widetilde{\beta_1}$. (1.17)
The residual sum of squares for the fitted model containing variables in $\mathbb{X}_1$ is
$SSE(q) = [Z - \tilde{Z}(q)]^T T [Z - \tilde{Z}(q)]$
$= Z^T T Z - \widetilde{\beta_1}^T \mathbb{X}_1{}^T T \mathbb{X}_1 \widetilde{\beta_1}$(3.38)
$SSE(k) = [Z - \hat{Z}(k)]^T T [Z - \hat{Z}(k)]$
$= Z^T T Z - \beta^T X^T T X \hat{\beta}$(1.18)
The Mallow's Cq statistic for a particular subset of q variables, using linear regression is
$C_q = \frac{SSE(q)}{(SSE(k)/N-k-1)} + 2(q+1) - N$(1.19)

2.8 Akaike Information Criterion (AIC)

Akaike Information Criterion (AIC) is a measure which enables the comparison of a set of statistical models to identify the model which minimizes information lost. From a set of statistical models, AIC gives an estimate of the quality of each of the model relative to each of the other model.
AIC is defined as
$AIC = 2k - 2loglikelihood$, where k is the number of estimated parameters in the model
Assuming that there are five models for comparison, if we define the AIC's of these models as AIC_1, AIC_2, AIC_3, AIC_4, and AIC_5, then the model having min(AIC_1, AIC_2, AIC_3, AIC_4, AIC_5) is the best model among these five models.

3. RESULTS

3.1 The Logistic Regression Model with only the constant term

The Logistic Regression model with only the constant term is defined as

$$\pi(X) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$$

3.1.1 Parameter Estimation

$$\hat{\pi}(X_i) = \frac{e^{\beta_0}}{1+e^{\beta_0}} = \frac{\sum_{i=1}^{n} Q_i}{\sum_{i=1}^{n} n_i} = \frac{204}{274} = 0.7445 \; for \; all \; i = 1,2,\dots,274$$

$$1 - \hat{\pi}(X_i) = 0.2555 \; for \; i = 1,2,\dots,274 \; , \; \widehat{\beta_0} = \ln\frac{0.7445}{0.2555} = 1.0695$$

$$, \; \widehat{var}(\widehat{\beta_0}) = \frac{1}{\sum_{i=1}^{n} \hat{\pi}(X_i)(1-\hat{\pi}(X_i))} = 0.01919, \quad \widehat{SE}(\widehat{\beta_0}) = \sqrt{var(\beta_0)} \approx 0.139$$

$$\widehat{CI} = 1.07 \pm 1.96 \times 0.139 = 1.07 \pm 0.27244 = (0.79756 \; , 1.34244)$$

3.1.2 Model Likelihood

$$\mathbf{L}(\boldsymbol{\beta_0}) = \prod_{i=1}^{n} \pi(X_i)^{Q_i} \left(1 - \pi(X_i)\right)^{1-Q_i}$$

$$= 2.365 \times 10^{-68}$$

$$\ln L(\beta_0) = \ln(2.365 \times 10^{-68}) = -155.7060 \; , \; -2 \ln L(\beta_0) = 311.4119$$

3.2 Logistic Regression model without interaction

The Logistic regression model without interaction, described by Equation 1.2 is defined as follow

X_{i1} − Motorcycle Ownership (MO)
X_{i2} −Possesion of valid driver's license (POVDL)
X_{i3} − Alcohol intake (AI)
X_{i4} − Knowledge of Road Signs (KORS)
X_{i5} − Marital Status (MS)
X_{i6} − (30 − 40 years) (AGE1)
X_{i7} − (above 40 years) (AGE2)
X_{i8} − Primary education (EDU1) ,
X_{i9} −Secondary education and above (EDU2)

$$\pi(X) = \frac{e^{\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+\beta_3 X_{i3}+\beta_4 X_{i4}+\beta_5 X_{i5}+\beta_6 X_{i6}+\beta_7 X_{i7}+\beta_8 X_{i8}+\beta_9 X_{i9}}}{1+e^{\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+\beta_3 X_{i3}+\beta_4 X_{i4}+\beta_5 X_{i5}+\beta_6 X_{i6}+\beta_7 X_{i7}+\beta_8 X_{i8}+\beta_9 X_{i9}}}$$

3.2.1 Parameter estimation

3.2.1.1 Logistic Regression coefficient estimation using Newton-Raphson iteration method.

$\beta^{(i)} = \beta^{(i-1)} + (X^T T X)^{-1} X^T (Q - \mu_Q)$. Let $\beta^{(0)}$ be a zero vector,

$$\beta^{(1)} = \beta^{(0)} + (\mathrm{X^T TX})^{-1}\mathrm{X^T}(\mathrm{Q} - \mu_Q)$$

$$\pi(X_i) = \frac{e^{(0)x_{i0}+(0)x_{i1}+(1)x_{i2}+(0)x_{i3}+(0)x_{i4}+(0)x_{i5}+(0)x_{i6}+(0)x_{i7}+(0)x_{i8}(0)x_{i9}}}{1+e^{(0)x_{i0}+(0)x_{i1}+(1)x_{i2}+(0)x_{i3}+(0)x_{i4}+(0)x_{i5}+(0)x_{i6}+(0)x_{i7}+(0)x_{i8}(0)x_{i9}}}, t_i$$

$$= \pi(X_i)(1 - \pi(X_i))$$

$$(X^T TX)=\begin{pmatrix} \sum_{i=1}^{274} t_i & \sum_{i=1}^{274} x_{i1}t_i & \sum_{i=1}^{274} x_{i2}t_i & & \sum_{i=1}^{274} x_{i9}t_i \\ \sum_{i=1}^{274} x_{i1}t_i & \sum_{i=1}^{274} {x_{i1}}^2 t_i & \sum_{i=1}^{274} x_{i1}x_{i2}t_i & \cdots & \sum_{i=1}^{274} x_{i1}x_{i9}t_i \\ \sum_{i=1}^{274} x_{i2}t_i & \sum_{i=1}^{274} x_{i1}x_{i2}\,t_i & \sum_{i=1}^{274} {x_{i2}}^2 t_i & & \sum_{i=1}^{274} x_{i2}x_{i9}t_i \\ & \vdots & & \ddots & \vdots \\ \sum_{i=1}^{274} x_{i9}t_i & \sum_{i=1}^{274} x_{i1}x_{i9}\,t_i & \sum_{i=1}^{274} x_{i2}x_{i9}\,t_i & \cdots & \sum_{i=1}^{274} {x_{i9}}^2 t_i \end{pmatrix}$$

$$= \begin{pmatrix} 68.50 & 17.509.000 & 38.5037.50 & 38.0022.75 & 6.50049.75 & 13.75 \\ 17.50 & 17.505.000 & 9.25010.25 & 8.7504.250 & 1.50011.50 & 3.500 \\ 9.000 & 5.0009.000 & 4.5005.500 & 4.0001.000 & 1.0005.950 & 2.750 \\ 38.50 & 9.2504.500 & 38.5021.00 & 22.0013.00 & 3.75026.70 & 8.250 \\ 37.50 & 10.255.500 & 21.0037.50 & 19.0011.75 & 2.50026.75 & 8.000 \\ 38.00 & 8.7504.000 & 22.0019.00 & 38.0017.75 & 6.50026.50 & 9.250 \\ 22.75 & 4.2501.000 & 13.0011.75 & 17.7522.75 & 0.00017.25 & 4.500 \\ 6.500 & 1.5001.000 & 3.7506.500 & 6.5000.000 & 6.5003.500 & 1.750 \\ 49.75 & 11.505.750 & 26.7526.50 & 26.5017.25 & 3.50049.75 & 0.000 \\ 13.75 & 3.5002.750 & 8.0009.250 & 9.2504.500 & 1.7500.000 & 13.75 \end{pmatrix}$$

$$X^T(Q-\mu_Q) = \begin{pmatrix} 67 \\ -6 \\ -2 \\ 47 \\ 29 \\ 30 \\ 15.5 \\ 5 \\ 51.5 \\ 10.5 \end{pmatrix} \text{ and}$$

$$(X^T TX)^{-1}X^T(Q-\mu_Q) = \begin{pmatrix} 2.04 \\ -1.65 \\ -0.95 \\ 0.49 \\ -0.43 \\ -0.42 \\ -0.63 \\ -0.08 \\ -0.10 \\ -0.16 \end{pmatrix}$$

$$\Rightarrow \beta^{(1)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 2.04 \\ -1.65 \\ -0.95 \\ 0.49 \\ -0.43 \\ -0.42 \\ -0.63 \\ -0.08 \\ -0.10 \\ -0.16 \end{pmatrix} = \begin{pmatrix} 2.04 \\ -1.65 \\ -0.95 \\ 0.49 \\ -0.43 \\ -0.42 \\ -0.63 \\ -0.08 \\ -0.10 \\ -0.16 \end{pmatrix}$$

$$\beta^{(2)} = \begin{pmatrix} 3.139 \\ -2.233 \\ -1.308 \\ 0.757 \\ -0.729 \\ -0.544 \\ -1.141 \\ -0.784 \\ -0.322 \\ -0.460 \end{pmatrix} \beta^{(3)} = \begin{pmatrix} 3.513 \\ -2.455 \\ -1.436 \\ 0.859 \\ -0.849 \\ -0.593 \\ -1.352 \\ -0.985 \\ -0.359 \\ -0.513 \end{pmatrix}$$

$$\beta^{(4)} = \begin{pmatrix} 3.558 \\ -2.482 \\ -1.455 \\ 0.870 \\ -0.864 \\ -0.605 \\ -1.376 \\ -1.004 \\ -0.359 \\ -0.511 \end{pmatrix} \beta^{(5)} = \begin{pmatrix} 3.559 \\ -2.482 \\ -1.455 \\ 0.870 \\ -0.865 \\ -0.606 \\ -1.377 \\ -1.005 \\ -0.359 \\ -0.512 \end{pmatrix}$$

Since $\beta^{(4)} \cong \beta^{(5)}$ the iteration converges at the fifth iteration and $\beta^{(5)}$ is the vector of our maximum likelihood estimates of the Logistic Regression Model coefficients.

$$\Rightarrow \begin{pmatrix} \boldsymbol{\beta_0} \\ \boldsymbol{\beta_1} \\ \boldsymbol{\beta_2} \\ \boldsymbol{\beta_3} \\ \boldsymbol{\beta_4} \\ \boldsymbol{\beta_5} \\ \boldsymbol{\beta_6} \\ \boldsymbol{\beta_7} \\ \boldsymbol{\beta_8} \\ \boldsymbol{\beta_9} \end{pmatrix} = \begin{pmatrix} 3.559 \\ -2.482 \\ -1.455 \\ 0.870 \\ -0.865 \\ -0.606 \\ -1.377 \\ -1.005 \\ -0.359 \\ -0.512 \end{pmatrix}$$

3.2.1.2 Variance and Standard error of estimate

At the fifth iteration,

$$(X^T TX) = \begin{pmatrix} 35.060 & 13.436 & 6.0680 & 16.980 & 21.575 & 22.686 & 14.000 & 4.5000 & 24.717 & 7.4340 \\ 13.436 & 13.436 & 3.2580 & 7.4270 & 7.5750 & 6.6150 & 2.5310 & 1.3270 & 8.7910 & 2.5090 \\ 6.0680 & 3.2580 & 6.0680 & 3.0680 & 3.6790 & 2.8960 & 0.3530 & 0.9290 & 3.9070 & 1.7530 \\ 16.980 & 7.4270 & 3.0680 & 16.980 & 10.696 & 11.328 & 6.5860 & 2.4100 & 11.407 & 4.0620 \\ 21.575 & 7.5750 & 3.6790 & 10.696 & 21.575 & 13.084 & 8.3670 & 1.8970 & 14.988 & 4.8050 \\ 22.686 & 6.6150 & 2.8960 & 11.328 & 13.084 & 22.686 & 11.793 & 4.4990 & 15.816 & 5.7080 \\ 14.000 & 2.5310 & 0.3530 & 6.5860 & 8.3670 & 11.793 & 14.000 & 0.0000 & 10.674 & 2.9930 \\ 4.5000 & 1.3270 & 0.9290 & 2.4100 & 1.8970 & 4.4990 & 0.0000 & 4.5000 & 2.4360 & 1.2350 \\ 24.717 & 8.7910 & 3.9070 & 11.407 & 14.988 & 15.816 & 10.674 & 2.4360 & 24.717 & 0.0000 \end{pmatrix}$$

$$\Rightarrow (X^T TX)^{-1} = \begin{pmatrix} 0.6040 & 0.1390 & 0.0440 & 0.0390 & -0.106 & -0.044 & -0.086 & -0.134 & -0.386 & -0.365 \\ 0.1390 & 0.1540 & 0.0070 & 0.0170 & 0.0210 & 0.0110 & 0.0520 & 0.0460 & 0.0470 & 0.0480 \\ 0.0040 & 0.0070 & 0.2280 & 0.0010 & 0.0060 & 0.0120 & 0.0550 & 0.0110 & -0.025 & -0.043 \\ 0.0390 & 0.0170 & 0.0010 & 0.1170 & -0.006 & -0.007 & -0.003 & -0.008 & 0.0020 & -0.007 \\ 0.1060 & 0.0210 & 0.0060 & 0.0060 & 0.1280 & 0.0080 & 0.0170 & 0.0370 & 0.0080 & 0.0000 \\ 0.0440 & 0.0110 & 0.0120 & 0.0070 & 0.0080 & 0.1810 & -0.071 & -0.106 & -0.037 & -0.056 \\ 0.0860 & 0.0520 & 0.0550 & 0.0030 & 0.0170 & -0.071 & 0.2010 & 0.1270 & -0.004 & -0.002 \\ 0.1340 & 0.0460 & 0.0110 & 0.0080 & 0.0370 & -0.106 & 0.1270 & 0.3780 & 0.0730 & 0.0640 \\ 0.3860 & 0.0470 & 0.0250 & 0.0020 & 0.0080 & -0.037 & -0.004 & 0.0730 & 0.4260 & 0.3880 \\ 0.3650 & 0.0480 & 0.0430 & 0.0070 & 0.0000 & -0.056 & -0.002 & 0.0640 & 0.3880 & 0.5300 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} VAR(\beta_0) \\ VAR(\beta_1) \\ VAR(\beta_2) \\ VAR(\beta_3) \\ VAR(\beta_4) \\ VAR(\beta_5) \\ VAR(\beta_6) \\ VAR(\beta_7) \\ VAR(\beta_8) \\ VAR(\beta_9) \end{pmatrix} = \begin{pmatrix} 0.604 \\ 0.154 \\ 0.228 \\ 0.117 \\ 0.128 \\ 0.181 \\ 0.201 \\ 0.378 \\ 0.426 \\ 0.530 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} SE(\beta_0) \\ SE(\beta_1) \\ SE(\beta_2) \\ SE(\beta_3) \\ SE(\beta_4) \\ SE(\beta_5) \\ SE(\beta_6) \\ SE(\beta_7) \\ SE(\beta_8) \\ SE(\beta_9) \end{pmatrix} = \begin{pmatrix} 0.777 \\ 0.392 \\ 0.478 \\ 0.342 \\ 0.358 \\ 0.425 \\ 0.448 \\ 0.615 \\ 0.653 \\ 0.728 \end{pmatrix}$$

3.2.2 Variable selection for model without interaction

For Forward and Backward stepwise logistic regression, $P_E = 0.15$ and $P_R = 0.2$.

Table 2. Summary of Mallow's Cq, AIC and AROC for selected models (model without interaction)

Selection Method	Variables in the model	Cq	AIC	AROC
Forward stepwise	$X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7}$	6.500	238.493	0.845
Backward stepwise	$X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7}$	6.500	238.493	0.845
Best subset	$X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7}$	6.500	238.493	0.845

3.3 The Logistic regression model with two factor interaction

The Logistic regression model with interaction,described by Equation 1.3 is defined as follow

X_{i1} − Motorcycle Ownership(MO)
X_{i2} −Possesion of Valid Driver's License (POVDL)
X_{i3} − Alcohol intake (AI)
X_{i4} − Knowledge of Road Signs (KORS)
X_{i5} − Marital Status (MS)
X_{i6} − (30 − 40 years) (AGE1)
X_{i7} − (above 40 years) (AGE2)

$$\pi(X) = \frac{e^{\beta_0+\sum_{j=1}^{7}\beta_j X_{ij}+\sum_{j=1}^{6}\beta_{1j+1}X_{i1}X_{ij+1}+\sum_{j=2}^{6}\beta_{2j+1}X_{i2}X_{ij+1}+\ldots+\beta_{67}X_{i6}X_{i7}}}{1+e^{\beta_0+\sum_{j=1}^{7}\beta_j X_{ij}+\sum_{j=1}^{6}\beta_{1j+1}X_{i1}X_{ij+1}+\sum_{j=2}^{6}\beta_{2j+1}X_{i2}X_{ij+1}+\ldots+\beta_{67}X_{i6}X_{i7}}}, X_{i6}X_{i7} = 0$$

3.3.1 Variable selection for models with two factor interaction.

For Forward and Backward stepwise logistic regression, $P_E = 0.15$ and $P_R = 0.2$.

Table 3. Summary of AIC and AROC for selected models (model with interaction)

Selection Method	Variables in the model	AIC	AROC
Forward stepwise	X_{i1}, X_{i3}, X_{i5}, $X_{i1}X_{i4}$, $X_{i2}X_{i3}$, $X_{i3}X_{i6}$, $X_{i4}X_{i7}$ $X_{i5}X_{i6}$	226.512	0.862
Backward stepwise	X_{i1}, X_{i3}, X_{i4}, X_{i6}, X_{i7}, $X_{i2}X_{i3}$,$X_{i3}X_{i5}$,$X_{i4}X_{i6}$	220.568	0.868
Best subset	X_{i1},X_{i3},X_{i4},X_{i5},X_{i6},X_{i7}, $X_{i2}X_{i3}$,$X_{i4}X_{i6}$	220.185	0.871

4. DISCUSSION

The data analyzed contains seven dependent variables with 274 observations. The independent variables are Motorcycle Ownership (X_{i1}), Possession of Valid Driver's License (X_{i2}), Alchohol Intake (X_{i3}), Knowledge of Road Signs (X_{i4}) Marital Status (X_{i5}), AGE1 [30-40 (X_{i6})], AGE2 [above 40 (X_{i7})], EDU1 [Primary education (X_{i8})], EDU2 [Secondary education and above (X_{i9})]. We used the Newton Ralphson iteration method to obtain coefficients of the variables in the full model (model without interaction factors). We also performed variable selection to fit a model not containing interaction factors and a model containing two factor interactions using Forward Stepwise, Backward Stepwise, and Best Subset methods of variable selection.

4.1 Model without interaction

The variables considered for selection are X_{i1},X_{i2},X_{i3},X_{i4},X_{i5},X_{i6},X_{i7},X_{i8}, and X_{i9}. The three selection techniques employed excludedX_{i8} and X_{i9}. Table 2 gave a summary of the Mallow's Cq, AIC and AROC for selected models. The information on Table 2 show that Best subset method, Forward stepwise method and backward stepwisemethod selected same set of variables (X_{i1},X_{i2},X_{i3},X_{i4},X_{i5},X_{i6},X_{i7})[Mallow's Cq (6.500) AIC (238.493), and AROC (0.845)]. This shows that forward stepwise method, backward stepwise method and best subset method has equal performance for selecting variables when interaction factors are not present.
The resulting AROC values for the models in Table 2 are all between 0.8 and 0.9, which indicates that the fitted models have excellent discrimination ability.

4.2 Model with interaction

For model with interaction, there are differences in the variables selected by the three variable selection methods adopted in this research work.

Best Subset method selected 25 models from which we chose the model [variable in the model (X_{i1}, X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7}, $X_{i2}X_{i3}$,$X_{i4}X_{i6}$)] with the smallest AIC value (220.185) (see Table 5).
Backward Stepwise method selected a model [variable in the model (X_{i1}, X_{i3}, X_{i4}, X_{i6}, X_{i7}, $X_{i2}X_{i3}$, $X_{i3}X_{i5}$,$X_{i4}X_{i6}$)] with larger AIC value (220.568) compared to the AIC value (220.185) of the selected model with minimum AIC value among models selected using Best subset method (see Table 3).
Forward Stepwise method selected a set of variables resulting in a model [variable in the model (X_{i1}, X_{i3}, X_{i5}, $X_{i1}X_{i4}$, $X_{i2}X_{i3}$, $X_{i3}X_{i6}$, $X_{i4}X_{i7}$, $X_{i5}X_{i6}$)]
with the largest AIC value (226.512) and the smallest AROC value (0.862) (see Table 3).
The model with the smallest AIC value among models selected using Best Subset method have AIC value smaller than the model fitted with the variables selected using Forward Stepwise selection method and Backward Stepwise selection method. Table 3 shows the AIC values and the AROC values for the models selected by each selection technique.
From Table3, the method with the best performance for fitting a model containing two factor interactions is the Best subset method which fitted a model [variables in the model (X_{i1}, X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7}, $X_{i2}X_{i3}$,$X_{i4}X_{i6}$)] with the minimum AIC value (220.185) and maximum AROC value (0.871) when compared to that of Backward Stepwise selection method and Forward Stepwise method.

5. CONCLUSION

In conclusion,

- The three methods of variable selection considered in this study have same performance in selecting variables for fitting a model without interaction.
- Best subset method outperformed Backward Stepwise method and Forward stepwise method in selecting variables for fitting a model with two factor interaction.
- All the models selected have excellent discrimination ability.

REFERENCE

1. Agresti, A. (2002). Categorical data analysis, Second Edition. John Wiley & Sons, Inc., Hoboken, New Jersey.
2. Berry, Michael J. A. (1997). Data Mining Techniques for Marketing, Sales and Customer Support. Wiley. p. 10.
3. Biondo S., Ramos E., Deiros M. et al. (2000). Prognostic factors for mortality in left colonic peritonitis: a new scoring system, J. Am. Coll. Surg. Vol. 191, № 6. – p. 635-642.
4. Boyd, C. R.,Tolson, M. A., and Copes, W. S. (1987). Evaluating trauma care: The TRISS method. Trauma Score and the Injury Severity Score. The Journal of trauma. 27 (4): 370–378. PMID 3106646. doi:10.1097/00005373-198704000-00005.
5. Bursac, Z., Gauss, C. H., Williams, D. K., and Hosmer, D. W. (2008). Purposeful selection of variables in logistic regression. Source Code for Biology and Medicine 2008, 3:17 doi:10.1186/1751-0473-3-17
6. Cox, D. R. & Snell, E. J. (1989). The analysis of binary data, Second Edition, London: Chapman and Hall.
7. Dabbour, E. (2012). Using Logistic Regression to identify risk factors causing rollover

collisions. International Journal for Traffic and Transport Engineering, 2(4): 372 – 379, DOI: p:/http/dx.doi.org/10.7708/ijtte.2012.2(4).07

8. Dobson, A. J. (2002). An Introduction to Generalized Linear Models. 2nded. Boca Raton, FL:Chapman& Hall/CRCSchafer, J. L. (2001). Lecture Notes for Statistics 544: Categorical Data Analysis I, Fall 2001. Penn State Univ. http://www.stat.psu.edu/_jls/
9. Draper, N. R. and Smith, H. (1981). Applied Regression Analysis. 2nded. New York: John Wiley.
10. Hosmer, D. W., Jovanovic, B and Lemeshow, S. (1989). Best Subsets Logistic Regression. Biometrics, Vol. 45, No. 4 (Dec., 1989), pp. 1265-1270.
11. Hosmer, D. W. &Lemeshow, S. (2000). Applied Logistic Regression, Second edition. John Wiley & Sons, Inc. Hoboken, New Jersey.
12. Kleinbaum, D. G. & Klein, M. (2010). Logistic regression a self learning text (3rd ed.). Springer New York Dordrecht Heidelberg London.
13. Kologlu M., Elker D., Altun H., Sayek I. (2001). Validation of MPI and OIA II in two different groups of patients with secondary peritonitis, Hepato-Gastroenterology. Vol. 48, № 37. – pp. 147–151.
14. Lee K. I. and Koval J. J. (1997). Determination of the best significance level in forward stepwise logistic regression, Communications in Statistics - Simulation and Computation, 26:2, 559-575.
15. Lihui, Z., Chen, Y. &Schaffner, D. (2001), Comparison of logistic regression and linear regression in modeling percentage data. Appl. Environ. Microbiol, 67(5):2129. DOI: 10.1128/AEM.67.5.2129-2135,2001.
16. Marshall J. C., Cook D. J., Christou N. V. et al. (1995). Multiple Organ Dysfunction Score: A reliable descriptor of a complex clinical outcome, Crit. Care Med. Vol. 23. – pp. 1638–1652.
17. Nagelkerke, N. J. (1991). A note on general definition of coefficient of determination. Biometrika, 78: 691-692.
18. Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized Linear Models, Journal of the Royal Statistical Society, Series A, vol. 135, pp.370-384.
19. Oluwadiya, K. S., Kolawole, I. K., Adegbehingbe, O. O.,Olasinde, A. A., Olaide A., &Uwaezuoke S. C. (2008). Motorcycle crash characteristics in Nigeria: Implication for control. Accident Analysis and Prevention 41 (2009) 294-298.
20. Pacheco, J., Casado, S., Núñez, L. (2009). A variable selection method based on Tabu search for logistic regression models. European Journal of Operational Research 199 (2009) 506–511
21. Palei, S. K. Das, S. K. (2009). Logistic regression model for prediction of roof fall risks in bord and pillar workings in coal mines: An approach. Safety Science. 47: 88–96. doi:10.1016/j.ssci.2008.01.002.Quin, G. &Keough. M. (2001). Generalized Linear Models and logistic regression. Experimental Design and Data Analysis for Biologists, Cambridge University Press.
22. Pohar, M. Blas, M. Turk, S. (2004). Comparison of Logistic Regression and Linear Discriminant Analysis: A Simulation Study. (PDF). Metodološkizvezki. 1 (1).
23. Sarkar, S.K. and Habshah, M. (2009). Optimization Techniques for Variable Selection in Binary Logistic Regression Model Applied to Desire for Children Data. Journal of Mathematics and Statistics 5 (4): 387-394, 2009. ISSN 1549-3644.
24. Shuhua, H. (2007). Akaike Information Criterion. Center for research in scientific

computation, North Carolina State University, Raleigh, NC.

25. Strano, M., Colosimo, B.M. (2006). Logistic regression analysis for experimental determination of forming limit diagrams. International Journal of Machine Tools and Manufacture. 46 (6): 673–682. doi:10.1016/j.ijmachtools.2005.07.005.
26. Wang, D., Zhang, W. and Bakhai, A. (2004). Comparison of Bayesian model averaging and stepwise methods for model selection in logistic regression. Statist. Med. 2004; 23:3451–3467, Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.1930

APPENDIX

Table 4. Summary of all selected models (model without interaction) using best subset method

Table 5. Summary of all selected models (model with interaction) using best subset method

k-1	Variable in the model	RSS	C_q	-2LL	AIC
1	X_{i1}	287.434	12.7613	261.788	265.788
2	X_{i1}, X_{i5}	280.269	7.7136	253.913	259.913
3	$X_{i1}, X_{i5}, X_{i2}X_{i4}$	276.164	5.6748	240.251	248.251
4	$X_{i1}, X_{i3}, X_{i2}X_{i3}, X_{i4}X_{i7}$	271.112	2.7049	229.406	239.406
5	$X_{i1}, X_{i3}, X_{i2}X_{i3}, X_{i4}X_{i7}, X_{i5}X_{i6}$	268.323	1.9618	219.652	231.652
6	$X_{i1}, X_{i3}, X_{i5}, X_{i1}X_{i4}$, $X_{i2}X_{i3}, X_{i3}X_{i6}$	264.976	0.6691	213.569	227.569
7	$X_{i1}, X_{i3}, X_{i4}, X_{i6}, X_{i7}$, $X_{i2}X_{i3}, X_{i4}X_{i6}$	262.185	-0.0770	204.704	220.704
8	$X_{i1}, X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7}$, $X_{i2}X_{i3}, X_{i4}X_{i6}$	**260.019**	**-0.2074**	**202.185**	**220.185**
9	$X_{i1}, X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7}$, $X_{i2}X_{i3}, X_{i4}X_{i6}, X_{i3}X_{i6}$	258.287	0.0887	200.393	220.393
10	$X_{i1}, X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7}$, $X_{i2}X_{i3}, X_{i4}X_{i6}, X_{i3}X_{i6}, X_{i1}X_{i7}$	257.148	0.9684	199.213	221.213
11	$X_{i1}, X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7}$, $X_{i2}X_{i3}, X_{i4}X_{i6}, X_{i3}X_{i6}, X_{i1}X_{i6}, X_{i1}X_{i7}$	255.605	1.4503	197.522	221.522
12	$X_{i1}, X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7}$, $X_{i2}X_{i3}, X_{i4}X_{i6}, X_{i3}X_{i6}, X_{i1}X_{i6}, X_{i1}X_{i7}, X_{i4}X_{i7}$	255.063	2.9166	196.951	222.951
13	$X_{i1}, X_{i3}, X_{i4}, X_{i6}, X_{i7}$, $X_{i1}X_{i6}, X_{i1}X_{i7}, X_{i2}X_{i3}, X_{i2}X_{i6}, X_{i2}X_{i7}, X_{i3}X_{i5}$, $X_{i3}X_{i6}, X_{i4}X_{i6}$	254.576	4.4381	196.453	224.453
14	$X_{i1}, X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7}$, $X_{i1}X_{i6}, X_{i1}X_{i7}, X_{i2}X_{i3}, X_{i2}X_{i6}, X_{i2}X_{i7}, X_{i3}X_{i6}$, $X_{i4}X_{i6}, X_{i4}X_{i7}$	254.015	5.8857	195.241	225.241
15	$X_{i1}, X_{i3}, X_{i4}, X_{i6}, X_{i7}$, $X_{i1}X_{i6}, X_{i1}X_{i7}, X_{i2}X_{i3}, X_{i2}X_{i6}, X_{i2}X_{i7}, X_{i3}X_{i5}, X_{i3}X_{i6}$, $X_{i4}X_{i5}, X_{i4}X_{i6}, X_{i4}X_{i7}$	253.519	7.3980	194.796	226.796
16	$X_{i1}, X_{i3}, X_{i4}, X_{i6}, X_{i7}$, $X_{i1}X_{i6}, X_{i1}X_{i7}, X_{i2}X_{i3}, X_{i2}X_{i6}$, $X_{i2}X_{i7}, X_{i3}X_{i4}, X_{i3}X_{i5}, X_{i3}X_{i6}, X_{i4}X_{i5}, X_{i4}X_{i6}, X_{i4}X_{i7}$	253.264	9.1474	194.486	228.486
17	$X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i6}, X_{i7}$, $X_{i1}X_{i2}, X_{i1}X_{i6}, X_{i1}X_{i7}, X_{i2}X_{i3}, X_{i2}X_{i6}, X_{i2}X_{i7}$, $X_{i3}X_{i5}, X_{i3}X_{i6}, X_{i4}X_{i5}, X_{i4}X_{i6}, X_{i4}X_{i7}$	252.854	10.7437	194.266	230.266
18	$X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i6}, X_{i7}$, $X_{i1}X_{i2}, X_{i1}X_{i6}, X_{i1}X_{i7}, X_{i2}X_{i3}, X_{i2}X_{i6}$, $X_{i2}X_{i7}, X_{i3}X_{i4}, X_{i3}X_{i5}, X_{i3}X_{i6}, X_{i4}X_{i5}, X_{i4}X_{i6}, X_{i4}X_{i7}$	252.599	12.4923	193.936	231.936

19	$X_{i1}, X_{i2,} X_{i3}, X_{i4}, X_{i6}, X_{i7},$ $X_{i1}X_{i2}, X_{i1}X_{i6}, X_{i1}X_{i7}, X_{i2}X_{i3},\ X_{i2}X_{i6},\ X_{i2}X_{i7},$ $X_{i3}X_{i4}, X_{i3}X_{i7},$ $X_{i3}X_{i5}, X_{i3}X_{i6},\ X_{i4}X_{i5}, X_{i4}X_{i6},\ X_{i4}X_{i7}$	252.455	14.3511	193.743	233.743
20	$X_{i1}, X_{i2,} X_{i3}, X_{i4}, X_{i6}, X_{i7},$ $X_{i1}X_{i2}, X_{i1}X_{i3}, X_{i1}X_{i6}, X_{i3}X_{i4}, X_{i1}X_{i7}, X_{i3}X_{i7}, X_{i2}X_{i3},\ X_{i}$ $,\ X_{i2}X_{i7},\ X_{i3}X_{i5}, X_{i3}X_{i6},\ X_{i4}X_{i5}, X_{i4}X_{i6},\ X_{i4}X_{i7}$	252.323	16.2217	193.692	235.692
21	$X_{i1}, X_{i2,} X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7},$ $X_{i1}X_{i2}, X_{i1}X_{i3}, X_{i1}X_{i6}, X_{i3}X_{i4}, X_{i1}X_{i7}, X_{i3}X_{i7}, X_{i2}X_{i3},\ X_{i}$ $,\ X_{i2}X_{i7},\ X_{i3}X_{i5}, X_{i3}X_{i6},\ X_{i4}X_{i5}, X_{i4}X_{i6},\ X_{i4}X_{i7}$	252.198	18.0990	193.659	237.659
22	$X_{i1}, X_{i2,} X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7},$ $X_{i1}X_{i2}, X_{i1}X_{i3}, X_{i1}X_{i6}, X_{i3}X_{i4}, X_{i1}X_{i7}, X_{i3}X_{i7}, X_{i2}X_{i3},\ X_{i}$ $,\ X_{i2}X_{i7},\ X_{i3}X_{i5}, X_{i3}X_{i6},\ X_{i4}X_{i5}, X_{i4}X_{i6},\ X_{i4}X_{i7}$ $X_{i5}X_{i6}$	252.114	20.0161	193.612	239.612
23	$X_{i1}, X_{i2,} X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7},$ $X_{i1}X_{i2}, X_{i1}X_{i3}, X_{i1}X_{i6}, X_{i3}X_{i4}, X_{i1}X_{i7}, X_{i3}X_{i7}, X_{i2}X_{i3},\ X_{i}$ $,\ X_{i2}X_{i7},\ X_{i3}X_{i5}, X_{i3}X_{i6},\ X_{i4}X_{i5}, X_{i4}X_{i6},$ $X_{i4}X_{i7}, X_{i5}X_{i6}$	252.100	22.002	193.582	241.582
24	$X_{i1}, X_{i2,} X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7}, X_{i1}X_{i2}$ $,X_{i1}X_{i3}, X_{i1}X_{i4}, X_{i1}X_{i6}, X_{i3}X_{i4}, X_{i1}X_{i7}, X_{i3}X_{i7}, X_{i2}X_{i3},\ X$ $,\ X_{i2}X_{i7},\ X_{i3}X_{i5}, X_{i3}X_{i6},$ $X_{i4}X_{i6}, X_{i4}X_{i5}, X_{i4}X_{i7}, X_{i5}X_{i6}$	252.098	24.0003	193.560	243.56
25	$X_{i1}, X_{i2,} X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7}, X_{i1}X_{i2}, X_{i1}X_{i3},$ $X_{i1}X_{i4}, X_{i1}X_{i5}, X_{i1}X_{i6}, X_{i3}X_{i4}, X_{i1}X_{i7}, X_{i3}X_{i7}, X_{i2}X_{i3},\ X_{i}$ $,\ X_{i2}X_{i7},\ X_{i3}X_{i5}, X_{i3}X_{i6},$ $X_{i4}X_{i6}, X_{i4}X_{i5}, X_{i4}X_{i7}, X_{i5}X_{i6}$	252.098	26	193.549	245.549

k-1	Variable in the model	SSE	C_q	-2LL	AIC
1	X_{i1}	293.9354	25.749	261.788	265.788
2	X_{i1}, X_{i6}	288.0576	21.834	251.658	257.658
3	X_{i1}, X_{i2}, X_{i6}	279.6204	15.345	241.144	249.144
4	$X_{i1}, X_{i2}, X_{i3}, X_{i6}$	274.6962	12.391	234.867	244.867
5	$X_{i1}, X_{i2}, X_{i3}, X_{i5}, X_{i6}$	270.1564	9.823	230.078	242.078
6	$X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i5}, X_{i6}$	265.3246	6.961	224.898	238.898
7	$X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7}$	**262.8789**	**6.500**	**222.493**	**238.493**
8	$X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7}, X_{i9}$	262.6844	8.305	222.296	240.296
9	$X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7}, X_{i8}, X_{i9}$	262.3816	10	221.985	241.985