International Journal of Mathematics and Statistics Studies

Vol.3, No.4, pp.27-41, July 2015

Published by European Centre for Research Training and Development Studies (www.eajournals.org)

APPLICATION OF LOGISTIC REGRESSION MODEL TO ADMISSION DECISION OF FOUNDATION PROGRAMME AT UNIVERSITY OF LAGOS.

[1]Osibanjo F. S., [2]Olalude G. A. [2]Akintunde M. O.and [2]Ajala A. G.

[1]Department of Industrial Relations and Personnel Management, University of Lagos, Nigeria

[2]Department of Statistics, Federal Polytechnic Ede, Ede, Osun State, Nigeria.:

ABSTRACT: *This paper considers the application of logistic regression model to admission process from foundation to degree of university of Lagos. The choice of this model becomes imperative as a result of dichotomous relationship existing in the model (either recommended for the admission or not). 395 students of foundation programme was selected from faculty of business administration. Statistical package for social scientist (SPSS) was used for the analysis. The results show that type of secondary school attended, mode of school fees payment at first registration, sponsor and first semester grade point average contribute significantly to the chance of gaining admission to the degree programme of the institution. The fitness of the model was assessed using Hosmer and Lemeshow test, split-sample approach and other supplementary indices to validate the model. The fitted model indicated that fitted binary logistic regression model could be used to predict the future admission process.*

KEYWORDS*:* Logistic regression model, Direct entry admission, Foundation programme Likelihood ratio, Odds ratio, Cross-validation, Roc curve.

INTRODUCTION

Many statistical problems call for the analysis and prediction of a dichotomous outcome: whether a graduate will succeed in the college or not, whether a child should be classified as learning disabled or not, whether an applicant should be employed/promoted or not and so on. These type of questions were addressed by either ordinary least squares (OLS) regression or linear discriminant function analysis. Both techniques were subsequently found not to be ideal for handling dichotomous outcomes due to their strict statistical assumptions of linearity, normality, and continuity for ordinary least squares (OLS) regression and multivariate normality with equal variances and covariance for discriminant analysis (Peng, et.al 2001, Cabrera, 1994; Cleary and Angel, 1984; Cox and Snell, 1989; Efron, 1975; Lei and Koehly, 2000; Press and Wilson, 1978; Tabachnick and Fidell, 1996).

This paper considers situations where the response variable is a categorical, attaining only two possible outcomes. Since response variables are dichotomous, it is inappropriate to assume that they are normally distributed. Despite the popularity of logistic regression modeling and the ease with which researchers are able to apply this technique using statistical software, confusion

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

continues to exist over terms, concepts, modeling approaches, and interpretations (Peng and So, 2002).

It is our intention in this paper to perform analysis in which the outcome variable follows Bernoulli distribution, which is a special case of the binomial distribution rather than a normal distribution in order to model binary outcomes and to apply logistic regression to estimate the probability of the individual contributions of demographic-academic variables (department, gender, age, type of secondary school attended, difference between year of completion of secondary school and year of admission into the programme, total grade points in English, Mathematics and Economics at O'level, entrance examination score, mode of school fees payment at first registration, sponsor, first semester grade point average, and area of residence during the programme) recommending graduates of Foundation Programme into Direct Entry Admission into University of Lagos. So also to apply logistic regression to estimate the probability of the collective contributions of the listed demographic-academic variables recommendation of graduates of Foundation Programme into Direct Entry Admission into University of Lagos. And to predict the categorical outcome for individual cases, assess the goodness-of-fit of a given model and determine whether the model can be used to predict outcome.

Logistic regression was first proposed in the 1970s as an alternative technique to overcome limitations of ordinary least squares (OLS) regression in handling dichotomous outcomes. It became available in statistical packages in the early 1980s (Ahani, et.al, 2010). Logistic regression model has become, in many disciplines, the standard method of data analysis. Since then, logistic regression has been used in many disciplines including medical studies (Sanchez et al., 2008; Kaufman et al., 2000; Rubino et al., 2003) and biomedical research mainly to formulate models sorting the factors that might determine whether or not an outcome happens (Sharareh R. et al, 2010). In the social research (Ingles et al., 2009; King and Zeng, 2002; Saijo et al., 2008; and Garcia-Ramirez et al., 2005), in market research (Neagu and Hoerl, 2005; Kleijnen et al., 2004; Barone et al., 2007; Sallis and Sharma, 2009; and Kirkos, 2009), in educational research, (Austin, et al, 1992) such as dropping out of college as a function of students characteristics, also become an important tool at the commercial applications (Erhart, Hagquist, Auquier, Rajmil, Power & Ravens-Sieberer, 2009; O'Leary, 2009; Weber & Michalik, 2008).

On the other hand university of Lagos offers a year (two-Semester Foundation Programme) across all departments. This is open to candidates who possessed at least five credits pass in O' level including Mathematics and English Language. Students who have two A-level passes, HND holders and graduates of recognized institution may also be considered into the Foundation Programme and satisfy all other conditions as stated in the general regulations governing University of Lagos. Also, credit pass in Economics in O'level is required for candidates seeking admission into Faculty of Business Administration. Students are required to satisfy the university in a selections process. Highly successful graduates are eligible for consideration for direct entry into the 200 Level of the Degree Programmes of the University through Joint Admissions and Matriculation Board (JAMB) and satisfy all other conditions as stated in the general regulations governing undergraduate degrees of the University of Lagos. Foundation Programme currently runs on a full-time basis and this is strictly non-residential programmes. The university provides an intensive one-year taught Foundation courses of study in the following Faculties: Arts, Business

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

Administration, Education, Engineering, Environmental Sciences, Law, Science, Social Sciences, Pharmacy and College of Medicine. The highly successful graduates of Foundation Programme, University of Lagos are also eligible for consideration for direct entry into the University of Bedford, the University of Central Lancashire (UK) and other prospective partnering Universities.

Specification of the model

The logistic regression equation bears many similarities to the regression equation. In its simplest form, when there is only one predictor variableX_1, the logistic regression equation from which the probability of Y is predicted is given by:

$$P(Y) = \frac{1}{1+e^{-(\beta_o+\beta_1 X_1)}} = \frac{e^{(\beta_o+\beta_1 X_1)}}{1+e^{(\beta_o+\beta_1 X_1)}} \qquad 1$$

In which $P(Y)$ is the probability of Y occurring, e is the base of natural logarithms, and the other coefficients form a linear combination much the same as in simple regression. Just like linear regression, it is possible to extend this equation so as to include several predictors. When there are several predictors the equation becomes:

$$P(Y) = \frac{1}{1+e^{-(\beta_o+\beta_1 X_1+\beta_2 X_2+\beta_i X_i)}} = \frac{e^{(\beta_o+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_i X_i)}}{1+e^{(\beta_o+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_i X_i)}} \qquad 2$$

The above equation (1) is the same as the equation (2) used when there is only one predictor except that the linear combination has been extended to include any number of predictors. Specifically, the values of the parameters are estimated using *maximum-likelihood estimation*, which selects coefficients that make the observed values most likely to have occurred.

The Binary Logistic Regression Model

In a variety of regression applications, the response variable of interest has only two possible qualitative outcomes, and therefore can be represented by a binary indicator variable taking values 0 and 1. A binary response variable, taking on the values 0 and 1, is said to involve binary responses or dichotomous responses.

The dependent variable of the logistic model is classified into two basic types (Afifi, et.al, 2004);

- Continuous Variable: can assume any value within a specified range.
- Discrete Variable: can only assume certain values and there are usually “gaps” between values (categorical response has two main categories: success(occurrence) and fail (no occurrence))

Generalised Linear Models (GLMs) are able to model non-normally distributed dependent variables, and thus overcome the problems of the assumptions of regular linear regression models (Venables and Ripley 2002). Quinn and Keough (2002) note that GLMs have three components: (i) a response (dependent) variable with population distribution belonging to the exponential family,

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

(ii) the predictor (independent) variable, and

(iii) a "link function" that links (i) and (ii). For example, the logistic model (for multiple predictors) is:

$$E(Y_i) = \pi_i = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots\beta_i x_i}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots\beta_i x_i}} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots.. \quad 3$$

where π_i is the probability that the response variable y = 1, β_0 is the equation constant, and β_i is the coefficient of predictor variable x_i. Thus, the binary response variable is modeled as an odds ratio, i.e. the probability that y will be a member of one class relative to the other class (Trexler and Travis 1993). Rather eqn. 3 directly, the link function $g(x)$ allows the response variable to be modeled as:

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_i x_i \dots\dots\dots\dots\dots \quad 4$$

For a binomial response variable the logistic (logit) link is the natural logarithm of the odds ratio:

$$g(x) = In\left[\frac{\pi_i}{1-\pi_i}\right] \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \quad 5$$

Thus, the logistic regression model is more easily solved as:

$$g(x) = In\left[\frac{\pi_i}{1-\pi_i}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_i x_i \dots\dots\dots\dots\dots\dots. \quad 6$$

The equation intercept (constant) $\beta_0$and variable coefficients β_i are estimated from calibration data using maximum likelihood techniques. Once these are known, the model can be applied to estimate future states based on an alternative data set considering the same variables.

To assess the significance of the logistic regression coefficients the Wald statistic and likelihood ratio test are used (Afifi et al., 2004).

$$\frac{\hat{\beta}^2}{[s.e(\hat{\beta})]^2}$$

Where $\hat{\beta}$ represents the estimated coefficient $\beta$and $s.e(\hat{\beta})$ is its standard error. Under the null hypothesis of zero slope and based on asymptotic theory, this quantity follows a chi-square distribution with one degree of freedom. If the estimated value of the slope is small and its estimated variability is large, then we cannot conclude that the slope is significantly different from zero and vice versa (Afifi et al., 2004).

The likelihood ratio test for overall significance of the beta's coefficients for the independent variables in the model is used (Hosmer and lemeshow, 2000; Fienberg,1998). The test based on the statistic" G" under the null hypothesis that the beta's coefficients for the covariates in the model are equal to zero. G statistic takes the form:

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

$$G = -2In\left[\frac{Likelihood\ without\ the\ variable}{Likelihood\ with\ the\ variable}\right]$$

The distribution of "G" is a chi-square with q degree-of-freedom, where q is the number of covariates in the logistic regression equation. Hauck and Donner (1977) and Jennings (1986) examined the performance of the Wald test and found that the test often failed to reject the null hypothesis when the coefficient was significant. They recommended that the likelihood ratio test to be used.

The likelihood statistic L is used to assess the fitness of the model. The sampling distribution of the – 2 log L has a chi-square distribution with q degrees of freedom under the null hypothesis that all regression coefficients of the model are zero (Fienberg, 1998). A significant p-value provides evidence that at least one of the regression coefficients for an explanatory variable is non zero.

Hosmer and Lemeshow (2000) developed a goodness-of-fit test for logistic regression models with binary responses. They porposed grouping based on the value of the estimated probabilities. This test is obtained by calculating the Pearson chi-square statistic from the 2×g table of observed and expected frequencies, where g is the number of groups. The statistic is written

$$x_{HW}^2 = \sum_{i=1}^{g} \frac{(o_i - N_i\pi_i^2)}{N_i - \pi_i(1-\pi_i)}$$

Where;

N_i Is the number of observation in the ithgroup.

o_i Is the number of event outcomes in the ithgroup.

π_i Is the average estimated probability of an event outcome for the ith group.

The odds ratio

The odds ratio is a measure of association for 2×2 contingency table (Agresti, 2007). In 2×2 tables the probability of "success" is π_1 in row 1 and π_2 in row 2. Within row 1, the odds of success are defined to be:

$$odds_1 = \frac{\pi_1}{1-\pi_1}\ and\ odds_2 = \frac{\pi_2}{1-\pi_2}$$

Evaritt (1998) and Agresti (2002) define the odds ratio in two groups of subjects as "the ratio of odds". Thus;

$$\theta = \frac{odds_1}{odds_2} = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$$

For the binary regression model, the odd ratio is the exponent (e^{β_i}) is the ratio of odds for a one-unit change in (Hosmer and Lemeshow, 2000). The change in Log odds, and the corresponding change in the odds ratio, for a c units is estimated $\exp[c\beta_i]$ (Fleiss, 1981). When the two groups of odds are identical then the odds ratio is equal to one.

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

International Journal of Mathematics and Statistics Studies
Vol.3, No.4, pp.27-41, July 2015
Published by European Centre for Research Training and Development Studies (www.eajournals.org)

The corresponding lower and upper confidence limits for odds ratio for a c units change are $\exp[cL_i]$ and $\exp[cU_i]$, respectively, for (c>0), or $\exp[cU_i$ and $\exp[cL_i]$ respectively, for (c<0), where (L_i, U_i); can be either the likelihood ratio-based confidence interval or the Wald confidence interval for β_i (Agresti,2002).

Generally, logistic regression is well suited for describing and testing hypotheses about relationships between a categorical or continuos predictor variables. In the simplest case of linear regression for one continuous predictor X and one dichotomous outcome variable Y, the plot of this data results in sigmodal or S-shaped which is difficult to describe with a linear equation for two reasons. First, the extreme do not follow linear trend. Second, the error are neither normally distributed nor constant across the entire range of data (Peng, Manz, & Keck, 2001). Logistic regression solves these problems by appling the logit transformation to the independent variable.

Cross Validation Techniques

Cross-validation is a general procedure used in statistical model building. It can be used to decide on the order of a statistical model including time series models, regression models, mixture distribution models, and discrimination models (Chernick, 2008). Cross validation is performed in different ways, some of them are:

1. Take two random subsets of the data. Models are fit or various statistical procedures are applied to the first subset and then are tested on the second subset. This is known as data-splitting
2. Leave-one-out technique is performed by fiting to all but one observation and then testing on the remaining one and has also been called “cross-validation by Efron” (1983), but it does not provide an adequate test.
3. Fit the model n times, each time leaving out a different observation and testing the model on estimating or predicting the observation left out each time. This provides a fair test by always testing an observations not used in the fit. It also is efficient in the use of the data for fitting the model since n – 1 observation are always used in the fit.

Hit ratio is the percentage of objects (individuals, respondents, firms, etc.) correctly classified by the logistic regression model. It is calculated as the number of objects in the diagonal of the classification matrix (H_o) divided by the total number of objects (N). Also known as the percentage correctly classified (Hair et al. 2009).

This can be compared with maximum chance and proportional chance criterion to determine the discriminating power of the function. Maximum chance criterion is the percentage of the total sample represented by the larger of the two groups (H_C). The proportional chance criterion is obtained from the actual direct entry admission decision by the equation $\pi^2 + (1-\pi)^2$, where π is the proportional of individuals in group (recommended for direct entry) and $1-\pi$ is the proportion of individuals in group (not recommended for direct entry). According to Marcoulides (1997) the difference between H_o and H_c may be tested by the following statistic

$$z = \frac{H_o - H_c}{\sqrt{H_c(N - H_c)/N}}$$

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

International Journal of Mathematics and Statistics Studies
Vol.3, No.4, pp.27-41, July 2015
Published by European Centre for Research Training and Development Studies (www.eajournals.org)

Where the significance of z is found by comparison with a critical value from a standard normal distribution.

DATA AND METHOD OF ANALYSIS

The data used for the analysis comprised 1,822 graduates of Foundation Programme in Faculty of Business Administration, University of Lagos for four academic sessions: 2007/2008 - 2010/2011 sessions out of which 395 graduates were randomly selected. it was obtained through the students' record in the Foundation Programme Office, Faculty of Business Administration, University of Lagos. The data collected include department (dept), gender (gender), age during the registration (age), type of secondary school attended(tssa), difference between year of completion of secondary school and year of admission into the programme, total grade points in English, Mathematics and Economics at O'level, entrance examination score, mode of school fees payment at first registration, sponsor, first semester grade point average, and area of residence during the programme that concerned the graduates admitted into the programme during these aforementioned sessions and completed the programme for one academic session.

The outcomes are coded in binary response: $Y = 1$ if the graduate was recommended for direct entry and $Y = 0$ if the graduate was not recommended for direct entry as shown below:

$$Y = \begin{cases} \textit{1, Recommended for direct entry} \\ \textit{0, Not recommended for direct entry} \end{cases}$$

Data analysis was carried out with the latest version of SPSS.

RESULTS

Descriptive Analysis

The cross tabulation of categorical variables show that departments distribution are 25.8% (n=102) Accounting, 17.0% (n=67) Actuarial Science & Insurance, 23.0% (n=91) Finance, 20.8% (n=82) Business Administration and 13.4% (n=53) Industrial Relations & Personnel Management. The gender distibution are 52.4% (n=207) female and 47.6% (n=188) male. The distribution according to secondary school attended are 28.9% (n=114) public and 71.1% (n=281) private. The distribution by mode of school fees payment at first registration are 34.9% (n=138) full payment and 65.1% (n=257) part payment. The distrution of graduate according to sponsor are 80.0% (n=316) parent and 20.0 (n=79) not parent. The distribution of graduates by the area of residence during the programme are 47.1% (n=186) within Akoka/Yaba and 52.9% (n=209) outside Akoka/Yaba.

The age ranged from 16 to 28 years with a mean of 18.93 years and standard deviation of 2.50. The diffrerence between year of completion of secondary school and year of admission ranged from 1-11 years with a mean of 2.51 and standard deviation of 1.743. The total points of credit

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

passed in English, Mathematics and Economics ranged from 6-18 points (i.e. A1=1, B2=2, B3=3,C4=4,C5=5,C6=6) with a mean of 12.05 and standard deviation of 2.88. The entrance examination score before admision into Foundation Programme for these graduates ranged from 74-137 with average of 101.41 and standard deviation of 12.00 while the first semester GPA ranged from 0.00 to 5.00, with a mean of 3.12 and standard deviation of 1.12.

Stepwise logistic regression analysis was used to reduce number of covariates results as presented in table1 below.

Table 1

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
tssa(1)	-1.058	.546	3.761	1	.052	.347	.119	1.011
mode_pay(1)	-1.623	.511	10.086	1	.001	.197	.073	.537
sponsor(1)	-1.931	.518	13.906	1	.000	.145	.053	.400
firstgpa	3.119	.403	59.983	1	.000	22.617	10.272	49.795
Constant	-4.164	1.037	16.140	1	.000	.016		

a. Variable(s) entered on step 1: tssa, mode_pay, sponsor, firstgpa.

It is noted that the covariates (type of secondary school attended, mode of school fees payment at first registration, sponsor and first semester GPA) of the graduates are statistically significant; while the covariates (department, gender, age, difference between year complete secondary school and year of admission into Foundation Programme, total grade points in three compulsory subject at O'level results, entrance examination score and area of residence during the programme) are statistically non-significant.

The Wald test is obtained by comparing the maximum likelihood estimate of the beta's$\hat{\beta}_i$, to its standard error. The resulting ratio, under the hypothesis that $\beta_i = 0$ are given in Table 1. But the covariate type of secondary school attended (tssa) is statistically non-significant and statistically significant using Wald test and likelihood ratio test respectively. However, Hauk and Donner (1977) and Jennings (1986) examined the performance of the Wald test and found that the test often failed to reject the null hypothesis when the coefficient was significant. They recommended that the likelihood ratio test to be used. Therefore, it is evident that the covariate (type of secondary school attended) is statistically significant.

And the logit is:

$$\boldsymbol{Y} = \boldsymbol{logit}(\boldsymbol{Direct\ Entry}) = -4.164 - 1.058X_1 - 1.623X_2 - 1.931X_3 + 3.119X_4$$

Here, the relationship between logit(Direct Entry) and X_1, X_2, X_3&X_4 is linear. Hence,

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

$$\boldsymbol{Y = logit(Direct\ Entry)} = -4.164 - 1.058tssa(1) - 1.623mode\_pay(1) - 1.931sponsor(1) + 3.119firstgpa$$

In other words

$$\boldsymbol{Y = logit(Direct\ Entry)} = -4.164 - 1.058private - 1.623part\ payment - 1.931not\ parent + 3.119firstgpa$$

The (Y) above indicates that: graduates attended private secondary school are less possibility to be recommended for direct entry; graduates with part payment of school fees at first registration are less possibility to be recommended for direct entry; a graduate that is not sponsor by parent are less likelihood than a graduate sponsored by parent of being recommended for direct entry. First semester GPA increases the chance of being recommended for direct entry, in other words, the higher the first semester GPA, the higher the chances that a graduate of Foundation Programme would be recommended for direct entry.

Moreso, the "B" values are the logistics coefficients that can be used to create a predictive equation. In this research:

$$P(Y|X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4) = \frac{e^{(\beta_o+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4))}}{1 + e^{(\beta_o+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4)}} = \frac{1}{1 + e^{-(\beta_o+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4)}}$$

Here, the relationship between the outcome and the predictors is non-linear.

$$P(Direct\ Entry) = \frac{e^{(-4.164-1.058private-1.623part\ payment-1.931not\ parent+3.119firstgpa)}}{1 + e^{(-4.164-1.058private-1.623part\ payment-1.931not\ parent+3.119firstgpa)}}$$

$$= \frac{1}{1+e^{-(-4.164-1.058private-1.623part\ payment-1.931not\ parent+3.119firstgpa)}}$$

The exponent (Exp (B)) in table 1 above is the odds ratio, thus:

- The odds ratio for private secondary school to public secondary school graduates to be recommended for direct entry are 0.347
- The odds ratio for part payment to full payment to be recommended for direct entry are 0.197
- The odds ratio for graduates not sponsor by parent to graduates sponsor by parent is 0.147

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

- The odds associated with first semester GPA is 22.617. Hence when GPA is raised by 1.00 point the odds is 22.62 times as large and therefore graduates are 23 more times to be recommended for direct entry.

Table 2 shows the classification table. Using the obtained Y function observations which are classified as follows, using a prior probability of 0.50

Table 2

Classification Table[a]

Observed		Predicted: Recommended for Direct Entry? Not Recommended	Predicted: Recommended for Direct Entry? Recommended	Predicted: Percentage Correct
Recommended for Direct Entry?	Not Recommended	71	15	82.6
	Recommended	11	298	96.4
Overall Percentage				93.4

a. The cut value is .500

(a) 82.6% of all graduates of Foundation Programme not recommended for Direct Entry were correctly classified and 17.4% were incorrectly classified.
(b) 96.4% from all graduates who wererecommended for Direct Entry were correctly classified, 3.6% were incorrectly classified.
(c) The overall correct percentage was 93.4%, which reflects the model's overall explanatory strength.

Table 3

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	137.859[a]	.503	.774

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test

Step	Chi-square	Df	Sig.
1	6.070	8	.639

From Table 3, Cox & Snell R-Square indicating that 50.3% of the variation in the independent variable is explained by the logistic model. Nagelkerke R Square is indicating a moderately strong relationship of 77.4% between the predictors and the prediction.The value of the Hosmer-Lemeshow goodness-of-fit statistic for the full model was Chi-square = 6.070 and the corresponding p-value from the chi-square distribution with 8 degree of freedom is 0.639 which means that it is not statistically significant and therefore our model is quite good.

Cross Validation Results

Using cross validation goodness-of-fit statistics the results for the full model was (using prior probability of .50) summarized in table 3

Table 4: Predicted classification table base on Training sample and Validation sample taking 0.5 as cut off

Training Sample					**Validation Sample**			
Observed (Y)	Expected (Y)				Observed (Y)	Expected (Y)		
	Not recommended	Recommended	Total			Not recommended	Recommended	Total
Not recommended	48	8	56		Not recommended	22	8	30
Recommended	3	181	184		Recommended	5	120	125

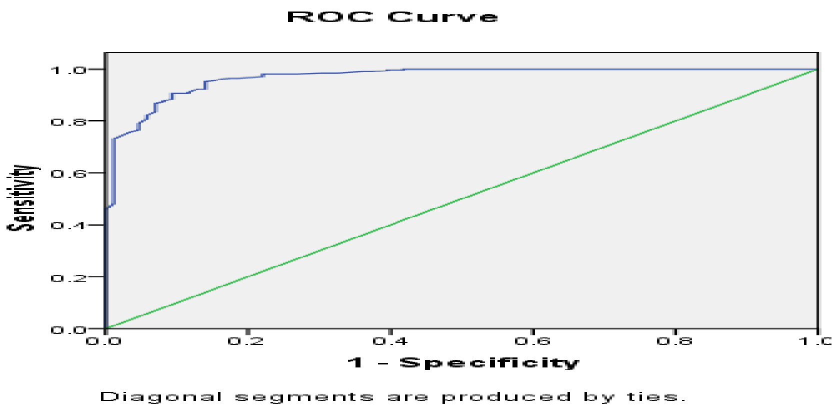
The classification matrix shows the accuracy of direct entry admission decision prediction in the cross validation as presented in table 4. In this validation sample of the 155 graduates that 120 or 96.00% were correctly classified into group recommended for direct entry while 22 or 73.33% were correctly classified into group not recommended for direct entry. The total correctly classified was 142 of 155 or 91.61%. In the training sample of the 240 graduates that 181 or 98.40% were correctly classified into group recommended for direct entry while 48 or 85.71% were correctly classified into group not recommended for direct entry. The total correctly classified was 229 of 240 or 95.42%. The percent prediction error rate for validation sample is 8.39% while the rate for the training sample was 4.58%. Thus the total prediction error rate for the validation sample is not considerably higher than training sample and we conclude that it is a reliable indicator of the predictive capability of the fitted logistic regression model.

ROC Curve

Using ROC curve for the classification accuracy, it is found that the area under the ROC curve, which ranges from zero to one, provides a measure of the model's ability to discriminate between those subjects who experience the response of interest versus those who do not.

Plotting sensitivity versus (1 – specificity) over all possible cut-points is shown in the Figure below. The area under the ROC curve for the final model is 0.968. The asymptotic significance of the model is less than 0.005 (Asymptotic Sig = .000), so all are doing better than guessing. This considered excellent discrimination.

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)



DISCUSSIONS OF FINDINGS

Significance testing for the logistic coefficients using likelihood ratio and stepwise regression method show that, at the 0.05 level of significance, type of secondary schools attended by the graduates, mode of school fees payment at first registration, sponsor and first semester GPA were significant predictors. But department, gender, age, difference between year completed secondary school and year of admission into foundation programme, total grade point in English, Mathematics and Economics at O'level results, entrance examination score and area of residence during the programme were non-significant predictors.

The odds ratio for private to public secondary school ranges between 0.12 times to 1.01 times with confidence of 95%. The odds ratio for part payment to full payment ranges between 0.07 times to 0.54 times with confidence 95%. The odds ratio for graduates not sponsored by parents to graduates sponsored by parent ranges between 0.05 times to 0.40 times with confidence of 95%. The odds (likelihood) associated with first semester GPA ranges between 10.27 times to 49.80 times with confidence 95%. To assess the fitness of the model the maximum likelihood test and Hosmer Lemeshow goodness-of-fit test suggest that the fitted logistic regression model has significant predictive ability for future subjects. Prediction error rate for validation of the model is not so high. The area under the ROC curve for the final model was 0.968 and asymptotic significance of the model is less than 0.005 which indicates that the predictive ability of the fitted model is good.

Thus, different summary measures of goodness-of-fit and others supplementary indices of predictive ability of the fitted model indicate that the fitted binary logistic regression model can be used to predict the future subjects.

CONCLUSION

Arising from the study, we conclude based on the model used that graduates that attended private secondary school have less possibility to be recommended for direct entry to graduates that attended public secondary school. Graduates with part payment of school fees at first registration are less possibility to be recommended for direct entry to graduates with full payment of school

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

fees at first registration while a graduate that is not sponsor by parent are less probability to graduates sponsored by parent of being recommended for direct entry. The first semester GPA reported in interval data ranging from 0.00 – 5.00. We conclude that the probability of recommended for direct entry increases as points of First Semester GPA increases. In other words, the higher the first semester GPA, the high likely a graduate of Foundation Programme would be recommended for direct entry. However, department, gender, age, difference between year complete secondary school and year of admission into foundation programme, total grade point in English

RECOMMENDATION

Based on the findings of this paper we recommend that the same study to all Foundation programme in University of Lagos with increase sample size be carried out. Develop a logistic regression model that contains repeated measures. Apply Classification and Regression Tree (CART) and compare the result with binary logistic regression model.

REFERENCES

Afifi, A., Clark, V. A., and May, S. (2004). Computer- Aided Multivariate Analysis. Fourth Edition, Champman and Hall/CRC.

Andy F. (2009). Discovering Statistics Using SPSS. Third Edition, Oriental Press, Dubai.

Austin, J. T., Yaffee, R. A., & Hinkle, D. E. (1992).Logistic regression for research in higher education.*Higher Education: Handbook of Theory* and Research, 8, 379-410.

Barone S., Lombardo A. and Tarantino P. (2007).A weighted logistic regression for conjoint analysis and Kansei Engineering.Quality and Reliability Engineering International, Vol. 23 Issue 6, pp. 689 – 706, John Wiley & Sons, Ltd.

Bradley, A.P. (1997). *The use of the area under the roc curve in the evaluation of machine learning algorithms*.Pattern Recognition.

Cabrera, A. F. (1994). Logistic regression analysis in higher education: An applied perspective. *Higher Education: Handbook of Theory and Research, Vol. 10,* 225–256.

Chernick, M.R. (2008). Boostrap Method: A Guide for Practitioners and Researchers. Second Edition, Wiley, Inc., New York.

Christensen R. (1997): Log-Linear Models and Logistic Regression. Springer publisher.

Cleary, P. D., & Angel, R. (1984).The *analysis of relationships involving dichotomous dependent variables*.*Journal of Health and Social Behavior*, *25*, 334–348.

Cox D.R. and Snell E.J (1989), *The Analysis of Binary Data*, 2nd edition, Chapman and Hall, London.

E.Ahani, O. Abass, and Ray O. Okafor (2010). *Application of Logistic Regression Model to Graduation CGPA of University graduates. International Journal of Statistics and Systems, Vol. 5, pp. 417-426.*

Efron, B. (1983). *Estimating the error rate of a prediction rule: improvements on cross validation. Journal of the American Statistical Association, Vol. 78, pp. 316-331*

Erhart, M., Hagquist C., Auquier P., Rajmil L., Power M., Ravens-Sieberer U. and the European KIDSCREEN Group. (2009*). A comparison of Rasch item-fit and Cronbach's alpha item reduction analysis for the development of a Quality of Life scale for children and adolescents. Child: Care, Health and Development, Blackwell Publishing Ltd.*

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

Ferri C.,Flach P., Hernandez-Orallo J. (2002). Learning Decision Trees Using the Area under the ROC Curve. Nineteenth International Conference on Machine Learning (ICML 2002); Morgan Kaufmann; pp. 46-139.

Fleiss, J., L. (1981). Statistical Methods For Rates And Proportions. Second Edition.John Wiley & Sons, Inc.

Hair, J.F., Anderson, R.E., Babin, B.J., Black, W.C. (2009) Multivariate Data Analysis.Seventh Edition.Maxwell MacmillqnInternation, New York.

Hauck, W.W., and Donner, A. (1977). *Wald's test as applied to hypotheses in logit analysis. Journal of the American Statistical Association, Vol. 72, pp.851-853.*

Hosmer, D. W., Lemeshow, S. (2000). Applied Logistic Regression, Second Edition, Wiley, Inc., New York.

Hosmer, D.W. &Lemeshow, S. (2000) *Applied logistic regression, 2nd edition.* Wiley, New York.

Ingles, C., J.; Garcia-Fernandez, j., M.; Castejon,J., L. ; Valle Antonio, B., D., and Marzo, J., C. (2009). Reliability and validity evidence of scores on the Achievement Goal Tendencies Questionnaire in a sample of Spanish students of compulsory secondary education. Psychology in the Schools, Vol. 46 Issue 10, pp. 1048 – 1060, Wiley Periodicals, Inc., A Wiley Company

Jennings, D.E. (1986a). *Judging inference adequacy in logistic regression. Journal of the American Statistical Association, 81, pp. 471-476*

King G. and Zeng L. (2002).Estimating risk and rate levels, ratios and differences in case-control studies.Statistics in Medicine, Vol. 21 Issue 10, pp. 1409 – 1427, John Wiley & Sons, Ltd.

Kirkos E., Spathis C., and Manolopoulos Y. (2009). Audit-firm group appointment: an artificial intelligence approach. Article on line in advance of print, Intelligent Systems in Accounting, Finance & Management, John Wiley & Sons, Ltd.

Kleijnen M., De Ruyter K., Wetzels M. (2004). *Consumer adoption of wireless services: Discovering the rules, while playing the game. Consumer adoption of wireless services: Discovering the rules, while playing the game. Journal of Interactive Marketing, Vol. 18 Issue 2, pp. 51 – 61, Wiley Periodicals, Inc., A Wiley Company, and Direct Marketing Educational Foundation, Inc.*

Lei, P.-W., &Koehly, L. M. (2000, April).*Linear discriminant analysis versus logistic regression: A comparison of classification errors.* Paper

Neagu R. and Hoerl R. (2005). A Six Sigma Approach to Predicting Corporate Defaults. Quality and Reliability Engineering International.Vol. 21 Issue 3, pp. 293-309, John Wiley & Sons, Ltd.

O'Leary, D., E. (2009). Downloads and citations in Intelligent Systems in Accounting. Finance and Management. Intelligent Systems in Accounting, Finance & Management, Vol. 16 Issue 1-2, pp. 21 – 31, John Wiley & Sons, Ltd.

Peng, C. Y., Manz, B. D., & Keck, J. (2001).*Modeling categorical variables by logistic regression.American Journal of Health Behavior, 25(3), 278–284.*

Peng, C. Y., & So, T. S. H. (2002). Logistic regression analysis and reporting: A primer. Understanding Statistics, 1(1), 31–70.presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Rosenbaum P.R. and Rubin D.B. (1983). *"The Central Role of the Propensity Score in Observational Studies for Causal Effects", Biometrika, 70, 41-55.*

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)

Saijo, Y., Ueno T., Hashimoto, Y. (2008). *Twenty-four-hour shift work, depressive symptoms, and job dissatisfaction among Japanese firefighters. American Journal of Industrial Medicine, Vol. 51 Issue 5, pp. 380 – 391. Wiley-Liss, Inc., A Wiley Company.*

Sallis, J., E. and Deo Sharma , D.(2009). Knowledge seeking in going abroad. Thunderbird International Business Review, Vol. 51 Issue 5, pp. 441 – 456, Wiley Periodicals, Inc., A Wiley Company.

Tabachnick , B., and L. Fidell. (1996). *Using multivariate statistics Third edition.* New York, USA: Harper Collins.

Weber, S., O.; and Michalik G., W., (2008).Incorporating sustainability criteria into credit risk management. Business Strategy and the Environment , Vol. 19 Issue 1, pp. 39 – 50, John Wiley & Sons, Ltd. and ERP Environment.

Venables W.N. & Ripley B.D. (2002). Modern applied statistics with S. Fourth edition. Springer: New York.

ISSN 2053-2229 (Print), ISSN 2053-2210 (Online)