Factors Affecting an Admitted Student's Decision to Enroll

Bakeerathan GUNARATNAM[1], Harold S. HALLER[2], Wojbor A. WOYCZYNSKI[2], John G. HALLER[3]

[1]Department of Bioinformatics and Biostatistics, University of Louisville
485 E Gray Street, Louisville, KY, 40202, USA
[2]Department of Mathematics, Applied Mathematics, and Statistics,
Center for Stochastic and Chaotic Processes in Science and Technology
Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA
[3]Vice President of Enrollment Management, University of Miami
1306 Stanford Drive, Coral Gables, FL 33124, USA

Corresponding Author

Bakeerathan Gunaratnam, PhD
Department of Bioinformatics and Biostatistics, University of Louisville
485 E Gray Street, Louisville, KY, 40202
email: b0guna01@louisville.edu

Keywords: Admitted students, Enrollment management, Logistic regression

 Abstract. This paper presents a process for and the development of a "Choice" model for predicting student enrollment in a professional school. Our process emphasizes the importance of using subject-matter experts to identify factors that potentially influence students' enrollment decisions. Our analysis focuses on how to identify the statistically significant factors that impact admitted student decisions to enroll or not to enroll. As a case study, this theory was applied to professional school enrollment at a private law school at a major research university in the Midwest. The research process involved personnel from the Law School's enrollment office and statisticians who developed a model using a logistic multiple regression method by analyzing 2008, 2009 and 2010 enrollment data for admitted students to the Law School in order to predict the probability of students enrolling in the Law School. The model was validated with 2011 enrollment data using a classification table with 71.4 % accuracy.

Introduction

Since the devastating recession that began in 2008, students are hesitating to enroll in professional schools because of reduced employment opportunities and the costs associated with matriculation, even though they are offered admission with a partial scholarship or financial aid. At the same time, schools are facing significant challenges in evaluating who would and would not enroll, due to the uncertain nature of human selection patterns, family commitments, geographical preferences, and professional interests and specialization. Relative to enrollment, Hossler, Bean, and Associates (Hossler 1990) discussed how students select colleges in Chapter 4 of *The Strategic Management of College Enrollments*.

This paper presents a process for and the development of a "Choice" model for predicting student enrollment in a professional school. Our process emphasizes the importance of using subject-matter experts to identify the factors that potentially influence a student's enrollment decisions. Our analysis focuses on how to identify the statistically significant factors

that impact admitted student decisions to enroll or not to enroll. In this way the analysis quantifies the marginal factors from consumer choice theory that influence enrollment decisions from the applicant's perspective (Hossler 2015). As a case study, we have applied this theory to professional school enrollment at a private law school at a major research university in the Midwest.

The "Choice" Model Process

The first step in applying our "Choice" model process requires a face-to-face meeting with the subject matter experts for the professional school in order to understand how the "Choice" model will be used. For example, the professional school's objective, in part, could be to determine predictors for the minimum amount of financial aid for each admitted student that is needed to enroll the highest profile prospective students in the program.

Toward meeting the professional school's objective, a retrospective study can be conducted beginning with a brainstorming session with the school's subject-matter experts. This session leads to the development of a cause and effect diagram (fishbone diagram). The diagram identifies data for analysis that is available, can be obtained with time and effort, and is unobtainable. The advantage of using such an approach is that all of the potential causes are identified in an organized and logical manner. Based on the "Choice" model outlined by Hossler, Bean, and Associates (Hossler 1990), the generic fishbone diagram should be completed during the brainstorming session. Factors identified on the fishbone diagram for which data are currently available should be circled and for which data are obtainable with some effort should be put in a square. Noting the factors for which information is available or can be obtained often indicates the likelihood of developing a meaningful "Choice" model.

It is recommended that the process for developing the "Choice" model proceed sequentially because as more of the obtainable variable data listed in the cause and effect diagram become available, additional analyses can be performed to improve the model. As in all modeling, the objectives of sequential analyses are to:

1.) Identify factors that do and do not influence an admitted student's decision or choice to enroll. Note, the statistically significant factors must make sense.
2.) Finalize a model for predicting the odds of any admitted student enrolling given a specified level of financial aid.
3.) Apply the model to students in the next academic year to verify the "Choice" model's accuracy.

Case Study Introduction

As an application of our professional school "Choice" model, a law school at a Midwestern research university was chosen, which we will refer to as the "Law School" in this case study. A law school degree has been viewed as a ticket to success and a solid career. Following the 2008 recession, while some law school graduates found positions in their chosen legal field many others struggled to find positions utilizing their law school training. Thus the supply and demand has been out of balance as indicated by the National Center for Education Statistics report that there was no change in the number of jobs for lawyers from 2010 to 2013

despite the fact that there have been about 45,000 regional completions of law degrees from 2007 to 2011. Also, the cost of law schools has been increasing consistently per year for the past years.

In addition, students' debt loans have been increasing at similar rates. Because of these situations applicants are hesitating to enroll even though they are offered admission into law school with a partial scholarship or financial aid. At the same time, law schools are facing significant challenges in evaluating who would and would not enroll, due to the uncertain nature of human selection patterns, family commitments, geographical preferences, and legal interests and specialization.

Using the cause and effect diagram as discussed from brainstorming session, the Enrollment Office gathered and prepared a spreadsheet of data file suitable for statistical method for logistic regression. Using this approach we were able to:

1) Identify factors that do and do not influence an admitted student's decision to enroll in the Law School.
2) Determine that the factors in the "Choice" model were both consistent with Enrollment Department experience and were logical or rational.
3) Determine the odds of an admitted student enrolling based on each admitted student's academic ability, ethnicity, socioeconomic, racial, and geographical preferences with various amounts of scholarship awarded.
4) Use the "Choice" model for students applying to the Law School for the academic year starting in September 2011 to verify the model's accuracy and determine the net tuition revenue for an optimized enrollment class based on a total financial aid budget.

Case Study Data

The learning data set consisted of year 2008, 2009 and 2010 admitted students to the Law School who had and had not enrolled as well as academic and demographic information about each student identified in the cause and effect diagram. The numbers of admitted students in 2008, 2009, and 2010 were 829, 1169 and 852 respectively, and 23% of these students enrolled out of the total 2,850. The racial and ethnic diversity figures indicated that there were 70% Caucasian, 11% Asian, 6% African American, 3% Hispanic and 1% Native American. The average age of the students for the three learning years at matriculation was 24.75 years (the standard deviation SD = 3.47). Their average grade point average (GPA) score was 3.49 (on a 4.0 scale) (SD = 0.31), average law school admission test (LSAT) score was 159.9 (on a 120 to 180 scale) (SD = 4.54), and 57% were male. Also, the data set had 8% missing values of which most were demographics for international students.

Case Study Methods and Materials

The response variable is categorical (admitted student enrolls yes or no) and the following variables were available as predictors, which were based on the brainstorming sessions with the subject-matter experts: scholarship, GPA, LSAT, ethnicity, degree, region, gender, travel stipend used to visit the school, scholarship awarded, age, permanent residence country, current residence country, permanent residence distance from campus, current residence distance from campus, legacy, area of student's academic interest, area population in 2010, area population change in 2005, area population change in 2010, area personal crime risk, area

property crime risk, socioeconomic region, area unemployment rate, area household income, area household spending, area median age, area average household size, area household children, area temperature in June, area temperature in July, area rain fall and area snow fall. A final model was obtained using a backward stepwise method eliminating variables that were not statistically significant at a level 0.05. Additionally, the estimated values of coefficients of the predictor variables were calculated along with p-values for the most robust model.

The accuracy for the chosen model was measured by the area under curve (AUC) in the receiver operating characteristic (ROC) curve. Also, a statistical cross validation (CV) technique was used in order to check over-fitting in the model. In addition, the p-value for the Homer-Lemeshow goodness-of-fit test was calculated to measure lack-of-fit. Finally, the model was validated with 2011 data from admitted students using a two-way classification table.

Case Study Statistical Analysis and Results

The logistic regression model developed for discriminating between admitted students who enroll and do not enroll is shown in Table 1 along with the estimated values of coefficients of the statistically significant predictor variables, standard errors of the estimators, and p-values for the most robust model. The model coefficients summarized in Table 1 can be interpreted as follows. First, applicants with exceptionally high GPA and LSAT scores are less likely to seek enrollment at the Law School, unless scholarship awards approach full tuition. Second, the further the applicant's current residence is from the Law School the less likely he or she is to enroll. Finally, the Stipend Award is a travel reimbursement for a visit to the Law School. Note that if this award is actually used, i.e. the applicant visits the Law School, the odds are much higher that he or she will enroll, which emphasizes the importance of arranging for the applicant to visit the campus.

Table1: Selected Variables in the Logistic Regression Model

Variable	Estimate of coefficients	Standard Error	p-value
Intercept	26.8288	2.7441	0.0000
GPA	-1.2358	0.1783	0.0000
LSAT	-0.1400	0.0158	0.0000
Scholarship/1000	0.0212	0.0075	0.0048
log(Current Distance + 1)	-0.2915	0.0313	0.0000
Stipend Awarded	-0.5877	0.1276	0.0000
Stipend Used	1.7691	0.1805	0.0000

Moreover, the area under ROC curve reveals the accuracy of the model. To describe the accuracy of the model, the AUC was also calculated. The model accuracy was good with an area under the receiver operating system of 0.743 (Hosmer and Lemeshow, 2000). The Hosmer-Lemeshow goodness of fit (p-value 0.6127) shows that the model fits very well at 95% confidence level and, according to the Brier scores (0.1527), the model has high accuracy (Glenn W Brier, 1950). In addition, the model was validated using CV. The value was 0.1527. The following figures show the results of the predictive model for the year 2008, 2009 and 2010 enrollment data. These figures are: 66.8% - Sensitivity, 71.3% - Specificity, 28.7% - Type I Error and 33.2% -Type II Error. Using the "Choice" model based on enrollment data from

admitted students for the years 2008, 2009 and 2010 to predict 2011 enrollment decisions: 138 students (70.4% of actual enrolled) were predicted to enroll and actually enrolled, while 413 students (71.7% of actual did not enroll) were predicted not to enroll and did not enroll. The overall accuracy of the model is 71.4%, which is a significant improvement over the prediction possible without this logistic regression model.

Case Study and Conclusions

A "Choice" model process was used and a statistical model was developed using a logistic multiple regression method for analyzing 2008, 2009 and 2010 enrollment data from admitted students to the Law School for predicting the probability of students enrolling. The predictors: GPA, LSAT, scholarship awarded, and distance from current residence to campus were significant in the model. The p-value for Hosmer and Lemeshow goodness of fit test was 0.6127 and the sensitivity and specificity of the fitted model were 71.3% and 66.8 % respectively. The area under the ROC curve was 0.743 and the Brier score was 0.1527. Also, the model was validated with 2011 enrollment data using a classification table with 71.4 % accuracy. The research process and modeling can be applied to any organizational enrollment application where a fixed, finite amount of funds are available to encourage applicants to enroll and quantitative and subjective criteria are used to define the optimal class.

Since enrollment patterns may change if institutional policies are altered, this model should be validated each year with new data and updated in order to improve its predictive power. By using both the research process outlined here and logistic multiple regression methods with new data to augment the existing data, improved predictive enrollment models with increased precision can be generated.

References

[1] Berge, Diane and Hendel, Darwin (2003), "Using Logistic Regression to Guide Enrollment Management at a Public Regional University", *Association for Institutional Research Professional File*, 86, 1-11.

[2] Brier, Glenn (1950), "Verification of Forecasts Expressed in terms of Probability", *Monthly Weather Review*, 78, 1-3.

[3] Hossler, Don, Bontrager, Bob, and Associates, *Handbook of Strategic Enrollment Management*, Josey – Bass 2015

[4] Hosmer, David and Lemeshow, Stanley (2000) Applied Logistic Regression, John-Wiley & Sons Inc., New York.

[5] Kohavi, Ron (1995), "A Study of Cross Validation and Bootstrap of Accuracy Estimation and Model Selection", *Proceeding of the Fourteenth International Joint Conference of Artificial Intelligence*, 2(12), 1137-1143.

[6] Thomas, Emily, Dawes, William, and Reznik, Gayle (2001), "Using Predictive Modeling to Target Student Recruitment: Theory and Practice", *Association for Institutional Research Professional File*, 78, 1-15.