A Logistic Approach To Predicting Student Success In Online Database Courses

George Garman, Metropolitan State College of Denver, USA

ABSTRACT

This paper examines the affects of reading comprehension on the performance of online students in a beginning database management class. Reading comprehension is measured by the results of a Cloze Test administered online to the students during the first week of classes. Using data collected from 2002 through 2008, the significance of the Cloze Test score is analyzed with respect to the three different assessment methods used in the class as well as to the overall average score of the students in the class. The data are tested using a binary logistic model that analyzes the data on a success (improvement) or failure (no improvement) basis. The analysis finds that reading comprehension has a significant impact on the scores students earn on examinations and on the final class average. However, the reading comprehension score has no significant impact on assessments that are more under the control of the student such as online open-book quizzes and projects.

Keywords: Cloze Test; Logistic Regression; Online Education

INTRODUCTION

etropolitan State College of Denver is an urban institution located in downtown Denver. The College shares a campus, called Auraria, with the University of Colorado at Denver and the Community College of Denver. By charter, the campus is strictly commuter with large sections of the available space dedicated to parking lots and parking garages.

The institutions housed on the Auraria campus have grown significantly over the past thirty years with over 40,000 students currently enrolled. The physical campus is surrounded by the city of Denver and cannot effectively expand. The stresses placed on the Auraria campus are enormous. Classes must be scheduled from 6:00am until after 10:00pm. Weekend classes are necessary and common. Perpetual traffic congestion is common on the streets surrounding the campus. Parking lots and garages generally fill early in the day.

Metropolitan State College of Denver is a modified open enrollment college that has been given the mission of educating all those who seek higher education studies. To effectively fulfill this mission, the College must continue to accept new applicants. During the past five years, the College has embarked upon an initiative to deliver many of its courses online. This initiative allows students to work and learn from home and avoid the trappings of an overextended campus. The College also benefits as it can attract new students from both within the Denver metropolitan area and from other areas throughout the state. Online delivery of courses also allows the College to keep pace with its competitors who are also delivering courses online.

Predicting success in online courses can be difficult. Students must be self-motivated and willing to take on responsibility to succeed in online courses. They must be willing and able to read and comprehend significant amounts of complex material without the benefit of an explanation by an instructor. Those students who can demonstrate an ability to read and comprehend will be expected to perform better in online courses than those who have more difficulty reading. This paper measures success of students who are just beginning an introductory database course as it relates to the student's ability to read.

It is an extension of an earlier paper that employed simple linear regression to predict the success of students in an online course. The data are updated through the Spring semester of 2008 and a logistic model is used to estimate the predictors of student success. [see Garman].

ONLINE COURSE ASSESSMENT

The beginning database course covers both the theoretical foundations of database theory and the basic elements of the SQL programming language. Students are supported by a required textbook, an SQL guide that provides a self-guided approach to database programming, an extensive set of online tutorial slides, and access to an instructor. Students have access to an Oracle 9i database housed on the College's mainframe with permissions to create a variety of Oracle objects under their own schemas. They can access this database from any on-campus computer lab or from any off-campus computer that has internet access.

Student work is assessed by three different methodologies. All students are required to complete three examinations that are proctored on campus by the College's Testing and Assessment Center. The examinations have a strict time limit and are closed book and closed notes. This assessment provides an assurance student work is not cooperative. Students' are assessed by a series of online quizzes. These quizzes contain exclusively objective questions with the answers found in the textbook. Although the quizzes are open book, they are restricted by time limits. Students are instructed to become familiar with the material before starting the quiz since time limits will not allow them to successfully search for all of the answers. The final assessment is a series of projects that are submitted and graded. The projects are assigned early in the course and students have ample time to complete them.

CLOZE TEST

Reading ability is frequently measured by a Cloze Test (or Deletion Test). The Cloze Test requests that a person read a passage of text with some words replaced by blanks. The person is asked to discern what word is missing and to fill in the blank with that particular word. The test can be structured so that key words are omitted or every nth word is omitted. The validity of the Cloze Test is widely accepted as a measure of ones ability to comprehend language. As stated in [Aitkens, 59] "that the Cloze procedure is a valid and reliable measure of both specific and general reading comprehension.

The Cloze Test used in this paper is structured from a passage describing the basic reasoning behind database theory. It comes from an older database management book that has long been out of print so there would be a minuscule likelihood that any students would have access to the passage. After providing several full sentences, every fifth word has been converted to a blank. The Test contains thirty-three blanks. Online students are required to take the quiz during the first week of class. Although the test scores are not reported to the students, students receive a grade for taking the test. The score is the percentage of the thirty-three questions the student has answered correctly. Students are also asked to report an estimate of the amount of time spent on the Test.

METHODOLOGY

Data have been collected from the Spring Semester of 2002 through the Spring Semester of 2008. Students who did not take the Cloze Test or who did not complete the course were eliminated. The cleaned data resulted in 235 data points. A baseline of performance was created for each student which was the overall GPA earned for a previous courses taken at MSCD. The baseline GPA was always the GPA earned by the student at the end of the semester prior to entering the Database course. The final course grade and each of the assessments were transformed into an integer value ranging from 0 to 4. The average of the course and the average of each of the assessments was evaluated as 90% and above equaling 4 points, 80% to 89% equaling 3 points, 70% to 79% equaling 2 points, 60% to 69% equaling 1 point, and below 60% equaling 0 points. The final course grade and each of the three course assessments were assigned a score as defined above.

The Binary Logit Model is a model based upon the cumulative logistic probability function that defines a successful event (Binary Response) as a 1 and a failure as a 0 [Pindyck and Rubinfeld, 287]. The successful event that a student improves for the final course grade of in any assessment area means that the student will receive a

score as computed above that exceeds the baseline GPA. The successful event will evaluate to a Binary Response as one and the failure to perform above the baseline GPA will evaluate to a Binary Response of zero. The odds that a student's outcome will result in a success is p/(1-p) where p is the success probability [Theil, 632]. The logistic model is defined as:

$$p(1\text{-}p) = \text{Exp}^{\,B_0 + B_1 X_1 + B_2 X_2 + B_3 X_3}$$

Where Exp is the base of the natural log and X_i 's are the independent variables. The equation above is linear in the log as shown below:

$$\text{Log}(\,(p/(1\text{-}p)) = B_0 + B_1 X_1 + B_2 X_2 + B_3 X_3$$

■■■

Reading proficiency, as measured by the Cloze Test, may impact the performance of the students on different assessment measures. The semester the course was taken may also affect student performance. Four models are specified below.

(1) $\text{Log}\,(p_1/(1\text{-}p_1)) = B_0 + B_1\,\text{ClozeScore} + B_2\,\text{SpringBinary} + B_3\,\text{SummerBinary}$
 Where p1 is the Estimated Final Course Average Proportion

(2) $\text{Log}\,(p_2/(1\text{-}p_2)) = B_0 + B_1\,\text{ClozeScore} + B_2\,\text{SpringBinary} + B_3\,\text{SummerBinary}$
 Where p1 is the Estimated Assignment Average Success Proportion

(3) $\text{Log}\,(p_3/(1\text{-}p_3)) = B_0 + B_1\,\text{ClozeScore} + B_2\,\text{SpringBinary} + B_3\,\text{SummerBinary}$
 Where p1 is the Estimated Project Average Success Proportion

(4) $\text{Log}\,(p_4/(1\text{-}p_4)) = B_0 + B_1\,\text{ClozeScore} + B_2\,\text{SpringBinary} + B_3\,\text{SummerBinary}$
 Where p1 is the Estimated Quiz Average Success Proportion

Each of the models has the ClozeScore as an independent variable with SpringBinary being one if the course was taken during the spring semester and SummerBinary being one if the course was taken during the summer term. The fall semester would be identified if both the SpringBinary and SummerBinary were zero.

Equation (1) measures the independent variables against a student's overall average for the course. Equations (2), (3), and (4) measure each of the assessments (examinations, projects, and quizzes) against the three independent variables described above. In all cases, the null and alternative hypotheses are indicated below:

Ho: $B_1 = 0$
Ha: $B_1 \neq 0$

A statistically significant coefficient for the ClozeScore would indicate evidence that the score on the Cloze Test is a predictor of the dependent variable.

RESULTS

The tables below summarize the results of the four regression equations. The coefficients for the ClozeScore was significant for the FinalAvg and the QuizAvg and not significant for the AssginmentAvg and the ProjectAvg. The semester binaries were not statistically significant in any model. Two of the Goodness of Fit tests for FinalAvg were statistically significant as was one of the tests for QuizAvg. The Goodness of Fit tests for the AssignmentAvg and the ProjectAvg were all statistically insignificant. The last log-likelihood from the maximum log-likelihood iterations is shown along with its associated G-statistic. This statistic tests the hypotheses that the slope of all coefficients is equal to zero. This test is rejected only for the QuizAvg regression.

Table 1

Predictor	(Equation 1 Final Course Average)						(Equation 2 Assignment Average)					
	Coef	Z	P	Odds Ratio	CI Lower	CI Upper	Coef	Z	P	Odds Ratio	CI Lower	CI Upper
Constant	-1.68	-3.47	0.001				0.130	0.29	0.77			
ClozeScore	0.028	2.48	0.013	1.03	1.01	1.05	0.004	0.42	0.68	1.00	0.98	1.02
SpringBinary	-0.12	-0.34	0.736	0.89	0.46	1.74	-0.62	-1.93	0.05	0.54	0.28	1.01
SummerBinary	-0.18	0.350	0.612	0.84	0.42	1.66	-0.09	-0.29	0.77	0.91	0.48	1.72
Goodness of Fit	ChiSq	DF	P				ChiSq	DF	P			
Pearson	58.46	49	0.167				46.21	49	0.587			
Deviance	68.33	49	0.035				56.91	49	0.204			
Hosmer-Lemeshow	16.88	7	0.018				9.00	8	0.342			
	Log-L'hood = -145.995 G=6.708 P=0.082						Log-L'hood = -160.659 G=4.355 P=0.226					

Table 2

Predictor	(Equation 3 Project Average)						(Equation 4 Quiz Average)					
	Coef	Z	P	Odds Ratio	CI Lower	CI Upper	Coef	Z	P	Odds Ratio	CI Lower	CI Upper
Constant	-1.11	-0.26	0.798				-2.10	-4.12	0.000			
ClozeScore	0.014	1.34	0.180	1.01	0.99	1.03	0.325	2.80	0.005	1.03	1.01	1.06
SpringBinary	-0.48	-1.47	0.141	0.62	0.33	1.17	-0.01	-0.03	0.977	0.99	0.50	1.98
SummerBinary	-0.36	-1.11	0.265	0.69	0.37	1.32	0.060	0.17	0.867	1.06	0.53	2.15
Goodness of Fit	ChiSq	DF	P				ChiSq	DF	P			
Peason	46.78	49	0.563				61.56	49	0.107			
Deviance	58.42	49	0.168				71.87	49	0.018			
Hosmer-Lemeshow	9.16	8	0.329				6.695	8	0.570			
	Log-L'hood = -160.535 G=3.751 P=0.290						Log-L'hood = -140.577 G=8.449 P=0.038					

CONCLUSIONS

The results are somewhat mixed but seem to be reasonable. The reading scores are clearly a significant predictor of a student's ability to succeed in an online course. High performance on the Cloze test indicates a level of performance higher than performance in prior courses. When completing activities that are not time bound, the Cloze score was a somewhat poor predictor of the student's score. However, open book quizzes that are time bound and require immediate reading comprehension correlate strongly with the Cloze score. When immediate knowledge and comprehension really matter, as when taking a time-bound open book quizzes, the Cloze score is a highly significant predictor. Perhaps not surprisingly, the semester in which the course was taken was not significant.

An extension of this analysis would involve comparing the affects of the Cloze Test on the scores of classroom lecture students to those who took the course online. One might expect that lecture students would be an entirely different group from online students. They probably would be less autodidactic and have a more passive approach to learning than the online group.

AUTHOR INFORMATION

Dr. George Garman is currently a professor of Computer Information Systems at Metropolitan State College of Denver. His primary teaching responsibilities are in the database field but he also teaches courses in statistics and management science. Prior to coming to Denver, Dr. Garman was a senior systems analyst at Blue Cross and Blue Shield Association in Chicago. He holds a bachelors degree from Indiana University and earned his Ph.D. degree from the University of Notre Dame in 1977. He has published numerous academic articles and has presented papers at several academic conferences.

REFERENCES

1. Aitken, Kenneth G., (1977) Using Cloze Procedure as an Overall language Proficiency Test, *TESOL Quarterly*, 11(1), March, 59-67.
2. Cheung, L and Kan, A. (2002) Evaluation of Factors Related to Student Performance in a Distance-Learning Business Communication Course. *Journal of Education for Business*, 77(5), 257-263.
3. Didia, D. and Hasnat, B. (1998) The Determinants of Performance in the University Introductory Finance Cource, *Financial Practice and Education*, 8(1), 102-107.
4. Dutton, John, Dutton, Marilyn, and Perry, Jo, (2002) How Do Online Students Differ From Lecture Student?, *Journal of Asynchronous Learning Networks*, 6(1), 1-20.
5. Garman, George. (2007) Evaluating Reading Skills as a Predictor of Success in Online Courses, International Business and Economic Research Conference Proceeding, Las Vegas, NV (distributed on CD ROM).
6. James, David and Chilviers, Clair (2001), Academic and Non-Academic Predictors of Success on the Nottinghom Undergraduate Medical Course: 1970-1995, 35(11), 1056-1064.
7. Kobayashi, Miyoko, (2002) Cloze Tests Revisited: Exploring Item Characteristics with Special Attention to Scoring Methods. *The Modern Language Journal*, 86(iv).
8. Minaei-Bidogoli, Behrouz (et. Al.) (2003) "Predicting Student Performance: An Application of Data Mining Methods with an Educational Web-Based System," *Frontiers in Education*, Vol. 1, 5-8.
9. Parker, Angie, (1999) A Study of Variables that Predict Dropout from Distance Education, International Journal of Educational Technology," *International Journal of Educational Technology*.
10. Pindyck, Robert S, and Rubinfeld, (1981) Daniel L., *Econometric Models and Economic Forecasts*. McGraw- Hill Book Company, New York.
11. Sweet, R., (1986) "Student Dropout in Distance Education: An Appreciation of Tinto's Model," *Distance Education* (2), 201-213.
12. Theil, Henri, (1971) *Principles of Econometrics*, John Wiley and Sons, Inc., New York.
13. Wojciechowshi, Amy and Palmer, Louann Dierlein, Individual Student Characteristics: Can Any Be Predictors of Success in Online Classes?

NOTES