AC 2009-100: USING A MULTIVARIATE APPROACH TO INVESTIGATE THE FACTORS CONTRIBUTING TO MINORITY STUDENTS' APPLYING TO GRADUATE SCHOOL

Xiaochun Jiang, North Carolina A&T State University

Jerry Watson, North Carolina A&T State University

Marcia Williams, North Carolina A&T State University

A multivariate Approach to Investigate the Factors Contributing to Minority Students Applying for Graduate School

Abstract

The goal of this research is to study what, if any, factors that may help minority students make his or her decision to apply to graduate school in science, technology, engineering, and mathematics areas. In the past four years, a longitudinal study has been conducted at North Carolina A & T State University to assess the effectiveness of the North Carolina Louis Stokes Alliance for Minority Participation project. Key information such as demographic information (e.g., gender, ethnicity), and academic information (e.g., high school GPA, SAT score) about the participants in the project was recorded. Over the years, student performance was tracked during their stay at the university and information about whether those participants went to graduate school was recorded. Given the complex nature of the problem, a multivariate statistical approach - multiway frequency analysis was used to investigate the association among the variables. Results indicated that there is a statistical significant relationship between the student's intent to apply to graduate school and his/her research experience as well as his/her high school GPA. Findings from this research will provide lights on minority students' decision to pursue advanced degrees and may help recruit minority students to apply to graduate school.

Introduction

The shortage of minority master and doctorate students is very challenging [1]-[4]. The North Carolina Louis Stokes Alliance for Minority Participation (NC-LSAMP), funded by the National Science Foundation, is a comprehensive, multidisciplinary, undergraduate program designed to increase substantially the quantity and quality of students, especially African American, Hispanic, and Native American students, who successfully complete science, technology, engineering, and mathematics (STEM) baccalaureate degree programs, and increasing the number of students interested in, and academically qualified for and matriculating into programs of graduate study. The NCLSAMP project involves eight institutions (four minority schools and four majority schools) within the University of North Carolina system. Partner institutions include North Carolina A&T State University as lead campus, Fayetteville State University, North Carolina Central University, North Carolina State University, University of North Carolina at Chapel Hill, University of North Carolina at Charlotte, University of North Carolina at Pembroke and Winston- Salem State University.

The primary objective of the NC-LSAMP project is to substantially increase the number of minority students graduating with B.S. degrees, and subsequently pursuing M.S. and Ph.D. degrees in science, technology, engineering, and mathematics (STEM) disciplines. The Alliance is achieving this goal by systematically enhancing recruitment, retention, access, and opportunities to education, internships, and research in these fields. One of the efforts that the alliance is actively pursuing is involving students in research. There is an annual research conference sponsored by the Alliance that showcases faculty-mentored research projects completed by the students.

As one of the nation's largest Historically Black College and Universities (HBCU), North Carolina Agricultural and Technical State University (NCA&T) was established in 1891 and is a public, comprehensive, land–grant university committed to fulfilling its fundamental purposes through exemplary undergraduate and graduate instruction, scholarly and creative research, and effective public service. NCA&T is located in Greensboro, North Carolina and enrolls over 10,000 students with about 89% of them being African Americans. NCA&T is the leading producer of African-American engineers in the nation. The university offers programs at the baccalaureate, masters and doctoral levels with emphasis on engineering, science, technology, business, education, agriculture, and other academic areas. Basic and applied research is conducted by faculty in university centers of excellence, in interinstitutional relationships, and through significant involvement with several public and private agencies. The university also conducts major research through engineering, transportation, and its extension programs in agriculture.

As part of the assessment effort, a longitudinal study is being and will continue to be conducted at NCA&T [5]. Demographic information such as age, gender, and race, and academic information such as SAT score, High school GPA of all students participating in the project at NC A&T since year 2003-2004 has been entered to a database developed to help the assessment efforts.

In order to assess the effectiveness of the project, control groups consisting of students that have no involvement in the project were carefully chosen to match the experimental groups. Statistical analysis was performed to ensure the compatibility between the control groups and the experimental groups. Specifically, demographic similarities between the two groups were established using the Fisher's Exact Test while academic credential similarities were compared using the Wilcoxon rank-sum test. In order to evaluate the effectiveness of the project, student cumulative GPAs were recorded every semester. Statistical analysis indicated that the NC-LSAMP project has had a very significant positive impact on student performance. Participants of the NC-LSAMP project performed significantly better than their counterparts in the control group for each semester [5]-[6].

One of the ultimate goals of the NC-LSAMP project is to attract more minority students to pursue advanced degrees upon successful completion of their undergraduate studies. Over the years, this has been proven to be a very difficult task. One question is what helps a minority student decide to go to graduate school. For instance, does gender, ethnicity, SAT score or High school GPA have any impact on their decisions? How about undergraduate research experience? This research is a first attempt to address some of the questions for the NC-LSAMP project.

The objective of the study is to study what, if any factors that may help minority students make his or her decision to apply to graduate school in science, technology, engineering, and mathematics areas.

Methodology

In this research, only 32 students who participated in the NC-LSAMP project since year 2003 were included. The reason for this is two fold. First, only those who applied to graduate school

were included in the study, those who recently joined the program (for instance, a sophomore) were excluded. Second, it would be very usefully to include students from the control group in this study. Unfortunately, it is very difficult to obtain data after students graduated. In this case, only a handful of students from the control group provided such information after they graduated. Therefore, it will not provide any meaningful comparison. Consequently, they were not included in this study. The following sections will describe the details of the participants, hypotheses and statistical analysis that were used in the study.

Participants

Thirty two students who participated in the NC-LSAMP project were included in the study. Six discrete variables were used in this research. They are gender (Male, Female), ethnicity (African American, Caucasian, Hispanic American, and Asian), SAT score, High School GPA, research experience (Yes, No), and application to graduate school (Yes, No). Tables 1-2 show the gender and ethnicity information of the participants. Since NCA&T is a historically black university, all but one participant were African Americans. Therefore, this variable was not included in the statistical analysis. Academic background information such as SAT score and High School GPA is continuous in nature and is categorized as a dichotomous variable to facilitate the statistical analysis. SAT score is converted to "L-low" if the score is below 1000, and "H-high" if it is more than or equal to 1000. High school GPA is converted to "L-low" if it is less than 3.0 and "H-high" if it is greater than or equal to 3.0. Research experience is recorded as "Yes" when a student participated in an undergraduate research or "No" if the student did not have such experience. Finally, application to graduate school is recorded as "Yes" if a student went pursue advance degrees and "No" if the student did not. Tables 3-6 provide frequency information for these four variables.

Table 1 Gender

Gender	Count	Percentage
Female	15	47%
Male	17	53%

Table 2 Ethnicity

Ethnicity	Count	Percentage
African American	31	97%
Hispanic American	1	3%

Table 3 SAT

SAT	Count	Percentage
Low	22	69%
high	10	31%

Table 4 High School GPA

High School GPA	Count	Percentage
Low	15	47%
High	17	53%

Table 5 Research Experience

Research Experience	Count	Percentage
Yes	18	56%
No	14	44%

Table 6 Graduate School

Graduate School	Count	Percentage
Yes	18	56%
No	14	44%

Statistical Analysis

This research problem is very complex in nature. It involves multiple variables, specifically, all of them are discrete. In order to investigate the relationship among those variables, a multivariate statistical technique is needed. Multiway frequency analysis is designed to discover associations among discrete variables [7]. In multiway frequency analysis, a log-linear model is used to develop the best model for predicting frequencies in each cell of the design. This model includes only the associations that are needed to reproduce the observed frequencies [7]. This process is similar to multiple regressions. The full model includes all possible effects in a multiway frequency analysis. An example full model of three variables (A, B, C) is shown in (1):

$$ln\ F_{ejk} = \theta + \lambda_{Ai} + \lambda_{Bj} + \lambda_{Cik} + \lambda_{ABij} + \lambda_{ACik} + \lambda_{BCjk} + \lambda_{ABCijk} \tag{1}$$

Where Fe is the expected frequency for each cell ijk, λ is the effect parameter, and θ is a constant.

Modeling fitting is usually done using the G^2 statistic for a particular incomplete model and evaluating its significance [7]. Since the full (saturated) model fits the data perfectly, the goal of the log-linear model is to identify the incomplete model with the fewest effects that still can describe the observed frequencies.

Hypotheses

H0: the aforementioned incomplete model is statistically the same as the saturated model (2)

Ha: the aforementioned incomplete model is statistically different from the saturated model (3)

To test this hypothesis, an incomplete log-linear model containing only high school GPA and research experience was developed. The test statistic G^2 for this model was computed and evaluated its significance. Since G^2 is a test of fit between observed frequencies and expected frequencies, an insignificant G^2 is expected in order to have a good model.

Results

SAS ® 9.1 was used to perform the statistical analysis. Prior to modeling, frequency tables were generated for screening purposes.

<u>Frequency Tables</u>

Table 7 shows one of the two way tables – between research experience and application to graduate school. For instance, the percent of students who had research experience and went to graduate school is 34.38% while the percent of students who had no research experience and went to graduate school is only 9.38%. This indicates that whether a student decides to go to graduate school may have some association with his or her research experience. Similarly, the percent of students who had high high school GPA and went to graduate school is 46.88% and the percent of students who had low high school GPA and went to graduate school is only 9.38% as shown in table 8. It seems that high school GPA may also be associated with a student's decision to pursue advanced degrees. A further look at the rest of the tables reveal that gender might not be associated with a student's intent to graduate school, neither did SAT score. However, statistical significance of these conclusions needs to be tested using inferential statistics which will be described in the next section.

Higher order frequency tables are much harder to analyze, especially when there are many cells without frequency values. Since there are only thirty two data points in this study, there are quite a few cells with zero frequency when five-way table is established. Therefore, it is very difficult to draw any meaningful conclusions reglated to the high order association. However, it appears that high order associations (three way, four way, and five way) are very weak, if any.

Table 7 Frequency Table for Research Experience and Graduate School

Frequency (Percentage %)		Graduate School		
		Yes	No	**Total**
Research Experience	Yes	11 (34.38)	3 (9.38)	**14 (43.75)**
	No	3 (9.38)	15 (46.88)	**18 (56.25)**
	Total	**14 (43.75)**	**18 (56.25)**	**32 (100)**

Table 8 Frequency Table for High School GPA and Graduate School

Frequency (Percentage %)		Graduate School		
		Yes	No	Total
Research Experience	Yes	3 (9.38)	11 (34.38)	**14 (43.75)**
	No	15 (46.88)	3 (9.38)	**18 (56.25)**
	Total	**18 (56.25)**	**14 (43.75)**	**32 (100)**

Maximum Likelihood Analysis of Variance

SAS ® Proc CATMOD was used to develop the incomplete log-linear model. There are many possible incomplete models available, especially when a large number of variables are included in the multiway frequency analysis.

For instance, to study the association with three discrete variables, there may be 126 ($_7C_1$ + $_7C_2$ + $_7C_3$ + $_7C_4$ + $_7C_5$ + $_7C_6$) possible incomplete models. To avoid those tedious work, screening process can be used to narrow the possible choices. Based on the screening process, it was hypothesized that an incomplete model with high school GPA and research experience fits the observed frequencies as described in the hypotheses section. A log-linear model with only high school GPA * Graduate school and Research experience * Graduate school was developed. Results can be seen in the maximum likelihood analysis of variance table as shown in Table 9. The p value for the goodness of fit test is 0.1462. At the significance level of 0.05, there is not enough evidence to reject the null hypothesis. Therefore, the incomplete model with only high school GPA and research experience can be used to predict expected frequencies in each cell.

Results from the maximum likelihood analysis of variance table also revealed that both graduate school / research experience association and graduate school / high school GPA association are significant ($p<0.05$), indicating that research experience and high school GPA are related to the student's decision to pursue advanced degrees. Table 10 shows the maximum likelihood estimates of the unsaturated log-linear model. It can be seen that coincidently, the impact of the research experience and high school GPA in this model are the same since they have the same estimates.

Table 9 Maximum Likelihood Analysis of Variance

Source	DF	Chi Square	P Value
Graduate School*Research Experience	**1**	**5.76**	**0.0164**
Graduate School*High School GPA	**1**	**5.76**	**0.0164**
Likelihood Ratio	**3**	**5.38**	**0.1462**

Table 10 Analysis of Maximum Likelihood Estimates

Parameter			Estimate	Standard Error
Graduate School*Research Experience	**NO**	**NO**	**0.5618**	**0.2341**
Graduate School*High School GPA	**NO**	**NO**	**0.5618**	**0.2341**

Discussion

To attract more minority students to pursue advanced degrees has been proven to be a very difficult task. In order to address this issue, it is necessary to investigate what, if any factors that might be related to a student's decision to go to graduate school. One of the goals of the NC-LSAMP project is to substantially increase the number of minority student pursuing M.S. and Ph.D. degrees in science, technology, engineering, and mathematics (STEM) disciplines. As a first attempt, this research applied a multivariate statistical technique to investigate the relationship among several discrete variables.

The major findings of this research are that research experience and high school GPA are related to a student's decision to go to graduate school. The significance of the finding, especially the impact of research experience, is very important. The NC-LSAMP project is very active in involving participants in undergraduate research. Every year, the alliance hosts a conference that showcases research conducted by the participants under the supervision of faculty advisors. The NC-LSAMP program also very encourages students to present their research work in other regional or national conferences. Feedback from the students is very positive. Most participants had no research experience prior to participating in the NC-LSAMP project. Some were even intimidated by research and, hence, not very confident of attending graduate school. The NCLSAMP project provided opportunities to these minority students to gain experience and conduct research. Such experience not only increases their level of interest in research but, more importantly, also increases their level of confidence of doing well in graduate school. In addition, these students often would influence their friends to get involved in research activities. So the potential of the impact of the NC-LSAMP program is huge.

It is not surprising to see that high school GPA is related to a student's intent to go to graduate school. It is surprising, though, to see that SAT score is not associated with the intent statistically. From these results, it looks that high school GPA is a better indication when looking at a student's intent to go to graduate school. However, it needs to be pointed out that the sample size of this study is very small; only thirty two students were included. Therefore, extra caution needs to be taken when drawing conclusions about SAT scores. Similar situation applies to gender. The insignificance of gender might be the result of the small sample size. Another possible reason is that even though there is a trend that more male students are enrolled in STEM areas than the female students nationwide, it is not true at NCA&T.

Another related constraint of this study is the lack of involvement of students from the control group. It would be interesting to compare the experimental group where students participated in the NC-LSAMP project and the control group where students did not. Unfortunately, it is very

difficult to track students after their graduation. At this time, there are too few data from the control group to make any meaningful comparison.

In the future, a much larger sample size needs to be used in the study. Given sufficient number of students, more levels can be categorized for high school GPA and SAT scores (There are only two levels for each variable in the current study). In addition, other variables such as whether a student has a summer internship and whether a student participated in a graduate school workshop can also be included.

Besides using multiway frequency analysis, it is also possible to develop a logistic regression model where the probability of a student going to graduate school can be predicted based on the key factors that influence his or her decision in the future research. Such a model can provide assistance to recruiting undergraduate students to go to graduate school.

Conclusion

This research is part of the ongoing effort of the North Carolina Louis Stokes Alliance for Minority Participation to attract more minority students to pursue advanced degrees in STEM fields. A multivariate statistical method, multiway frequency analysis was used to study what, if any factors that may help minority students at North Carolina A & T State University make his or her decision to apply to graduate school in STEM areas. Results indicate that there are statistical significant relationships between the student's intent to apply to graduate school and his/her research experience, and between the student's intent to apply to graduate school and his or her high school GPA. However, this research has limitations due to its small sample size. A much larger sample size needs to be used in the future studies. Another limitation is the lack of control group information. Even though there was a control group in the study, getting the information that is relevant to this study prove to be very difficult. Nevertheless, findings from this research will provide lights on minority students' decision to pursue advanced degrees and may help recruit minority students to apply to graduate school.

Acknowledgment

Bibliography

1. Carnevale, A. P., and Fry, R. A., "Crossing the great divide: Can we achieve equity when generation Y goes to college?" 2000, Princeton, NJ: Educational Testing Service.
2. Oguntimein, G. B., Leigh-Mack, P., Davy, B., and Wheatland, J., "Research experience program for undergraduates in a Historically Black College and University." *Proceedings of the 2004 American Society for engineering Education Annual Conference & Exposition*, 2004.
3. Mervis, J., "Wanted: A better Way to boost number of minority Ph.D.s", *Science*, 281, 1998, 1268-1270.
4. Smith, S., Working recruitment miracles. Black Issues in Higher Education, 16 (170), 1999, 40-41.

5. Jiang, X., Sarin, S., and Williams, M., "Assessment of NC-LSAMP project: A longitudinal Study", *Proceedings of the 2005 American Society for engineering Education Annual Conference & Exposition,* 2005, Portland, OR.
6. Williams, M., Watson, J., Jiang, X., and Sarin, S., "Preliminary Results of the NC-LSAMP project", *Proceedings of the 2006* Results of the NC-LSAMP project", *Proceedings of the 2006American Society for engineering Education Annual Conference & Exposition*, 2006, Chicago, IL.
7. Tabachnick, B.G., Fidell, L. S., "Using multivariate statistics," fifth edition, 2007, Allyn and Bacon.