Estimating the True Value of Merit-Based Financial Aid in Higher Education

Michael Johnston

A.J. Palumbo School of Business Administration

Duquesne University

Economics Senior Seminar

Abstract

Higher education institutions use merit-based financial aid as an important tool in their efforts to compel quality students to enroll in their college. While the importance is unquestioned, this paper examines the statistical significance and validity of using merit-based aid as a tool to persuade students to matriculate. Additionally, the paper examines whether different segments of students react differently to aid offers based upon academic qualifications or preferred academic discipline. Finally, that information is then used to quantify the effect aid has on students in the construction of a probit model to explain student behavior and estimate the probability of enrollment given certain individual characteristics in the context of a merit-based financial aid offer. By developing those estimations, the price elasticity is then approximated regressing probability of enrollment against scholarship awards and comparing the subsequent slopes against the different student segments. Merit-based aid is in fact found to be statistically significant and its degree is found to vary across different segments explained by a variety of factors.

I. Introduction

College admissions has grown into a long and very structured process in which students, who can be viewed as the consumers of high education as a good, search within a market for the supplier which best meets their needs. Those needs are categorized in a variety of ways including academic, athletic, spiritual, social, and demographic preferences which all drive the choice function of the consumer. The interesting dynamic within the market is that the supplier, or any given university, wishes to qualify its consumers and only provide its services to students it deems fit within previously outlined and developed standards. These standards generally revolve around academic quality but also include the well-rounded features of a student such as athletic prowess, dedication to community service, musical talents, and others. In terms of objective qualification however, the university is limited to using academic indicators such as SAT scores or GPA to gain a sense of the overall academic quality of a student.

While a school is trying to limit the students it accepts to maintain certain academic standards, it also needs to mix that with an ability to attract the students it has interest in. Therefore, even though the student can be identified as the consumer, and the school as the supplier, both parties take part in some degree of selling themselves to the other. As discussed, students use their talents and experiences, specifically captured within their objective academic measurements to prove their worth to a school. The school on the other hand, will boast characteristics such as academic quality, school pride, and athletic ability, but in many cases their most attractive tool to persuade individual students may be the amount of merit based financial aid it can offer.

At issue then is the ability of an institution to control the likelihood of enrollment of the students it accepts using its merit based financial aid as the driving force. While there are certain externalities whose existence is known but is immeasurable by institutions, this paper examines the ability of a college admission office to control its designation of funds to most efficiently understand the likelihood that the students they wish to matriculate do indeed enroll. Furthermore it segments students into demographic groups to isolate the way in which different groups respond to the aid.

II. Literature Review

The market for higher education exists in a very complex manner which has led to significant research within the subject, especially the causes surrounding a student's tendency to enroll in a particular university. Before one even questions the effect of financial aid that a school can offer, it is important to note the reasons why it is offered. "College financial aid offers depend on the characteristics of the student; students who are considered more desirable by the college administration may receive more generous offers. Hence, college costs and financial aid packages are likely correlated with student characteristics" (Linsenmeier, 2001). College financial aid can be used as a tool in an effort to persuade students to consider a particular university that might otherwise seem unattractive without the possibility of receiving a scholarship (Avery, 2002).

Merit based scholarships are becoming increasingly popular for a variety of reasons. In 2000, over $100 billion was spent on the federal and state level to subsidize higher education, with a large part of that money being designated as Pell Grants. While there is such a large amount of money being devoted to supporting college attendance, Dynarski contends there is little evidence that these subsidies serve their goal of increasing attendance (Dynarski, 2001). There is a strong sense that these need based grants are not serving their purpose and thus the trend is towards more effective merit based aid (Dynarski, 2000; van der Klaauw, 2001). In addition to the need based aid being flat ineffective in the opinion of many economists, Edlin contends that it is also distributed in a way that is both inequitable and inefficient. He claims that financial aid in higher education is actually a tax on savings. By rights, two families of identical size and income could pay entirely different sums for higher education thanks to need based aid that considers net wealth in the determination of allocation. If one family goes on extravagant vacations every year he says, while the other consistently saves money instead of spending on things like vacations, the one that has been saving will receive less need based aid because it has more wealth at the time of matriculation into school (Edlin, 1993). Together, the thoughts of ineffectiveness, inequality, and inefficiency all play parts in a shift towards merit-based awards.

Heller describes the market for higher education by identifying students as human capital who are looking to optimize their future value through the college choice. Within that choice they must then consider price as a constraint since that will directly impact their expected future value of the education (Heller, 2001). Essentially, the student is forced to make an optimal decision based on the utility he or she will receive from any schools that may be possible destinations. They are presented with the choices: (i) enroll at College X or enroll at another college (van der Klaauw, 2001). Heller says that specific studies have shown that aid does have a significant impact on college enrollment, especially at lower-income levels (Heller).

Such theory and studies lead universities to believe that they possess aid as a tool to use in the competition for the country's best students and that belief is validated by a few examples. Several states have begun to offer significant merit-based scholarships to students in their state to increase college attendance and completion. Although the source of this funding originates at the state level rather than the university, its success is noteworthy. Specifically, the college HOPE program in Georgia has become a prominent example of how a state-run merit-based scholarship program should be run. Dynarski finds that for every $1,000 of subsidy granted by the state for this program, college attendance raises by approximately 4 to 6% (Dynarski, 2002).

Linsenmeier also finds similar results in his study of a Northeastern University that changed its financial aid structure to target money that was designated in loan form to be given as grants. He finds that adding significant grant money increased the likelihood of enrollment approximately 3%. Like many other studies, his identified students within an income structure and found the increase in attendance across the low-income student segment (Linsenmeier).

In an attempt to look at the market in a broad context removed from state grants, Avery and Hoxby developed a survey and administered it to high school seniors in the process of selected a college. It was administered in three phases to a sample size randomly selected by high school college guidance counselors from the top ten percent of the high school's students. The survey gauged where students applied, their acceptance status, their financial aid grants, and ultimately their decision on whether or not to enroll. Characteristics that increased probability of attendance included financial aid, the selectivity of the school, and whether or not family members ever attended the university. High tuition, room and board, a mean SAT score lower than their individual score, and less selectivity were all causes to lower the probability of enrollment. Specifically the data suggests that every $1,000 in aid grant money increases the probability of enrollment by 11% while a similar decrease in tuition only decreases the probability by 2% indicating an uneven reliance towards scholarships than lower tuition (Avery).

The most difficult concept within the education market is the incomplete information problem that exists. van der Klaauw explains that a student makes decisions on which school to attend based on a variety of reasons. The most difficult part for colleges though is the lack of information concerning a prospective student's alternative options, including other schools or employment opportunities. Specifically within the context of aid, he says that it is nearly impossible to gauge the level of aid another school is willing to offer and how a student will subsequently react to that (van der Klaauw).

III. Data Specification

To examine the ability of an institution to affect the probability of enrollment after acceptance, a one year sample was used; the 2003 fall accepted applicants from Duquesne University, a medium-sized private University in Pennsylvania. The data consists of whether or not a student enrolled, their would-be resident status in the university, whether they are a male or female, the school into which they were accepted and its tuition, zip-code, SAT score, high school GPA, and merit based scholarship awarded from the Office of Admissions. Additional information is available from within the provided data. The different academic schools into which students were accepted can be roughly ranked by competitiveness within the market for higher education programs by the flat-rate tuition which is constant to all students within a school but varies University-wide based on demand of schools that offer varying programs.

IV. Difference of Means Solution

The preliminary examination of the impact of merit based scholarships is done via a difference of means test across the different controlled group of students. The sample of students is segmented into groups related to their academic qualifications including: top third SAT, bottom third SAT, top third GPA, and bottom third GPA. The sample is also segmented into eight different academic school criteria. A paired two sample for means t-test was conducted across each group. The null hypothesis stated that the mean merit based scholarship was the same for students who did enroll and those who did not for each of the different segments. The results are displayed in Table 1.

Category	t-Stat	p-Value	Enroll_Mean	Non_Mean
Bottom SAT	7.1657	0.0000	5008.37	2604.92
Top SAT	2.0139	0.0448	8312.86	7701.64
Bottom GPA	6.7151	0.0000	4993.69	2735.05
Top GPA	2.6168	0.0093	8032.31	7229.88
Education	1.9638	0.0528	6113.79	4768.39
Business	6.2886	0.0000	6929.15	4095.81
Health Science	1.3333	0.1856	6366.49	5727.61
Liberal Arts	2.8483	0.0047	5907.85	4757.92
Music	1.6008	0.1161	6686.67	5331.77
Nursing	2.6184	0.0127	6088.16	4092.11
Science	1.6916	0.0930	6275.93	5479.26
Pharmacy	-0.5333	0.5950	6419.58	6686.56

Table 1. Difference of Means Tests.

Significant differences exist within several of the segments. Notably, while scholarship significance varied across the academic schools the differences were all significant across the qualification segments at the 5% level. More specifically, the lesser qualified students were more influenced by the amount of scholarship offered than were the better qualified students. The implication associated with this difference is that a more qualified student is less impacted by financial aid because he or she is more likely to have significant aid offers from several schools and is free to make their decision based more upon other characteristics. The lesser qualified students then would be more likely to be influenced by a scholarship offer given the assumption they do not have as many offers from other schools.

Additionally, three schools have significant mean differences at the 5% level: Business, Liberal Arts, and Nursing. All three of these schools are at the bottom of the tuition fees for the University thus implying they are the least competitive schools for students to be admitted into. The other school with the same tuition price at the bottom of the pricing structure is the School of Education, which has a p-value of just slightly over .05. The significant differences in scholarships then exist because the University is forced to offer scholarships to persuade students to enroll in each of these schools more so than they would need to compel a student to enroll in the Pharmacy school, for example. In fact, the Pharmacy school was the only school to actually have a higher mean scholarship for students who did not enroll than those who did, even though the difference was not significant. Students applying for the more competitive schools are intrinsically more interested in simply being accepted into the program than having to be swayed by a significant financial aid offer.

V. Model Specification

The difference of means test indicated that the level of scholarship is indeed significant, at least in certain segments. The question then remains of how to best utilize that information to make intelligent decisions involving the allocation of scholarship funds. To specify how each dollar of scholarship money affects the probability a given student will enroll after being accepted in the University, a probit model was developed to explain the behavior of students based upon their different characteristics. The probit model used is defined as EQ.1:

$$\hat{p} = \alpha + \beta_1(male) + \beta_2(resident) + \beta_3(\ln(scholarship) * dum) + \beta_4(\ln(abs(15219 - zip))) + \beta_5(SAT) + \beta_6(GPA)$$

Variable Definitions

$\hat{p}$	Probability of enrollment
Male	1 = male; 0 = female
Resident	1 = resident ; 0 = commuter
ln(scholarship)*dum	Log of the dollar amount of merit based scholarship times dummy variable set to 1 = receiving some amount of merit scholarship; 0 = receiving no merit based scholarship
ln(abs(15219-zip))	Log of the absolute value of the difference between Duquesne's zipcode and student's zipcode[1]

[1] Used as a proxy for approximate distance from the University since actual mileage distance is unavailable. Measures the log of the number of zip codes away rather than absolute distance.

SAT	Student’s SAT score
GPA	Student’s high school GPA

VI. Results

Segment		Base	Bottom GPA	Top GPA	Bottom SAT	Top SAT	Education	Business
LR Stat		75.2742	116.7585	36.4775	86.3781	37.0534	26.5460	41.3018
LR Prob		0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0000
McFadden R-Square		0.0551	0.1094	0.0338	0.0809	0.0346	0.0990	0.0668
A	coefficient	-1.3402	3.4135	0.2891	2.6398	1.4069	2.4257	2.3242
	z-statistic	-3.0569	4.9600	0.3005	3.7132	1.6731	2.6437	3.6469
	p-value	0.0022	0.0000	0.7638	0.0002	0.0943	0.0082	0.0003
β1(male)	coefficient	-0.1567	-0.2199	0.0307	0.0091	-0.0824	-0.0067	-0.1221
	z-statistic	-1.7699	-2.1728	0.3076	0.0903	-0.8566	-0.0318	-0.9209
	p-value	0.0767	0.0298	0.7584	0.9281	0.3917	0.9746	0.3571
β2(resident)	coefficient	-0.4164	0.3678	0.1870	0.3896	0.3066	0.5199	0.2376
	z-statistic	-2.8966	2.4217	1.0801	2.7906	1.5610	1.9787	1.2507
	p-value	0.0038	0.0154	0.2801	0.0053	0.1185	0.0478	0.2111
β3(ln(scholarship) *dum)	coefficient	-0.0569	0.0915	0.1157	0.0873	0.0431	0.1001	0.0989
	z-statistic	-3.4509	7.1692	0.0413	6.8175	1.3709	2.8788	5.0733
	p-value	0.0006	0.0000	0.0051	0.0000	0.1704	0.0040	0.0000
β4(ln(abs(15219-zip)))	coefficient	0.0952	-0.1047	-0.0840	-0.0715	-0.0913	-0.0910	-0.0585
	z-statistic	5.4422	-5.4390	-4.2136	-3.6539	-4.7927	-2.4178	-2.4325
	p-value	0.0000	0.0000	0.0000	0.0003	0.0000	0.0156	0.0150
β5(SAT)	coefficient	0.0005	-0.0010	-0.0006	-0.0012	-0.0007	-0.0011	-0.0015
	z-statistic	1.4500	-2.4475	-1.4541	-1.7726	-1.0265	-1.2099	-2.6723
	p-value	0.1471	0.0144	0.1459	0.0763	0.3047	0.2263	0.0075
β6(GPA)	coefficient	0.2163	-0.6884	-0.0575	-0.4624	-0.1259	-0.4172	-0.3045
	z-statistic	2.0347	-3.6372	-0.2701	-3.8108	-1.1204	-1.6595	-2.1976
	p-value	0.0419	0.0003	0.7871	0.0001	0.2625	0.0970	0.0280

Segment		HealthSci	Liberal Arts	Music	Nursing	Science	Pharmacy
LR Stat		12.1567	64.9883	19.8430	24.0338	12.7643	41.2415
LR Prob		0.0586	0.0000	0.0030	0.0005	0.0469	0.0000
McFadden R-Square		0.0370	0.0727	0.1435	0.1880	0.0316	0.1195
A	coefficient	0.9418	2.1517	-0.4954	3.5846	0.7388	5.6942
	z-statistic	1.0463	4.1807	-0.2891	2.3038	0.8283	4.7658
	p-value	0.2954	0.0000	0.7725	0.0212	0.4075	0.0000
β1(male)	coefficient	0.0562	-0.0686	0.3408	0.0949	-0.0985	0.0414
	z-statistic	0.2656	-0.6349	1.1605	0.1719	-0.6388	0.2216
	p-value	0.7905	0.5255	0.2458	0.8635	0.5230	0.8246
β2(resident)	coefficient	0.0852	0.6863	0.2221	0.1183	0.1797	0.1004
	z-statistic	0.2179	4.0128	0.3866	0.1836	0.7521	0.1610
	p-value	0.8275	0.0001	0.6991	0.8543	0.4520	0.8721
β3(ln(scholarship) *dum)	coefficient	0.0560	0.0553	-0.0352	0.1831	0.0477	-0.0376
	z-statistic	1.1232	3.3097	-0.5042	3.1272	1.5108	0.7327
	p-value	0.2614	0.0009	0.6141	0.0018	0.1308	0.4637
β4(ln(abs(15219-zip)))	coefficient	-0.1009	-0.0619	-0.1468	-0.2314	-0.0803	-0.1696
	z-statistic	-2.6691	-2.9114	-2.5333	-2.9068	-2.3996	-4.3555
	p-value	0.0076	0.0036	0.0113	0.0037	0.0164	0.0000
β5(SAT)	coefficient	-0.0008	-0.0008	0.0027	-0.0011	0.0004	-0.0024
	z-statistic	-1.0019	-1.9374	2.0734	-0.6994	0.5604	-2.8220
	p-value	0.3164	0.0527	0.0381	0.4843	0.5752	0.0048
β6(GPA)	coefficient	-0.0164	-0.4144	-0.4822	-0.5066	-0.3075	-0.4481
	z-statistic	-0.5324	-3.2237	-1.2942	-1.1147	-1.5997	-1.7664
	p-value	0.5945	0.0013	0.1956	0.2650	0.1097	0.0773

The model was applied to a range of segmented groups: the university as a whole, the top third of students according to SAT and GPA, the bottom third of students according to SAT and GPA, and each individual academic school. The model was applied to individual schools rather than utilizing dummy variables in a single use model to adequately isolate the differences between groups. Additionally, because the schools are treated individually in terms of operation within the university, examining each group alone with the model is reasonable to understand the students individual to that school. By separating the sample into classifications according to their academic ability, the model can also discover whether students at different academic quality levels react in similar ways to differences in merit-based aid.

The results for the model varied across different segments. The LR probability for every model suggests that jointly the coefficients are significantly different from zero. When the model was applied to Health Sciences and Sciences the LR probability is close to a 5% level while the remainders are clearly significant at that level. Similar to the results observed through the difference of means tests, the model seems to be most significant with the lesser achieving segments, along with the Business and Liberal Arts schools.

Specifically, each individual variable reacts differently within the various segments. Upon the original specification of the model within a broad University-wide sample, the male variable was significant at the 10% level with a negative coefficient. The direction of the coefficient is consistent with the makeup of the University demographics suggesting males are less likely to enroll, supported by a much larger percentage of females in the University population as a whole. When applied to every other segment of the applicant pool, sans the bottom GPA segment, the male variable becomes insignificant when examining different subsets of the sample.

Resident status is a similarly varying variable across the segments. Significant for the University-wide sample, the lesser-achieving segments, as well as Education and Liberal Arts students, resident status is not significant in the other groups. Within the significant groups, the coefficient is generally positive, another characteristic supported by the makeup of the University community which observes approximately 85% of incoming students residing in on-campus housing.

Of primary interest within the model to describe probability of enrollment, the scholarship variable produced expected results. The natural log function was used to transform the data because the relationship was found to be non-linear. The first dollar of scholarship money seems to be more significant than the last dollar of the scholarship. This would support the idea that students expect to pay some portion of their tuition but indeed need and expect a significant base grant. The natural log of the scholarship was multiplied by a dummy variable set to zero if the applicant received no scholarship money in order to ensure that all observations were included in the analysis since the natural log of zero is invalid. Significant within all segments other than Top SAT, Health Science, Music, Science, and Pharmacy, the scholarship variable coefficient is expectedly positive in most cases indicating that additional scholarship dollars do increase the probability of enrollment. Interesting to note is that although the variable and mean difference is insignificant, the coefficient within the Pharmacy segment is negative, which is a consistent with the mean of the award offered to a student who does not enroll exceeding the award offer to students who do enroll. Additionally, the scholarship seems to affect most significantly the lesser-achieving students whose segments have z-statistics of approximately seven, supporting the theory that this group most likely has fewer options and is more likely to react to scholarship dollars.

Most significant in the model seems to be the proxy variable for distance from the University. The school, which has a large portion of its enrollment focused around the local Pittsburgh area, is most likely to have a student enroll the closer he or she lives to the University. Like the scholarship variable, this variable is also transformed into the natural log which fits the distribution more accurately than a standard linear variable. The variable, defined as the natural log of the absolute value of the difference between 15219[2] and the student's zip code is not an absolute measure of distance, rather a proxy for distance counting the number of zip codes away from the school the student lives. While the variable is not perfect in its measure of distance, it does significantly describe the tendency of students to enroll given their geographic location at the 5% level. The direction of the variable is negative indicating the more removed the student is from the area the less likely they are to enroll.

The final two variables examine the academic qualities of accepted students and their impact on the probability that they will enroll. SAT and GPA scores are the most common quantitative variables assigned to evaluate academic talent of applicants. The significance of these variables differs over segments, but the groups where both SAT and GPA are significant are: the lesser-achieving students, Business, Liberal Arts, and Pharmacy. Those variables are all significant at the 10% level, and most at the 5% level. The sign for almost every coefficient is negative which would indicate that the more advanced a student is, the less likely they will enroll in the University given the existence of so many other options they have which are unidentified by the University.

The coefficients for the model run over the different groups were very consistent, both with expectations and across the board as a whole. The main differences appeared with the base sample of the University as a whole and all of the different segments. For the resident,

[2] 15219 is the zip code for Duquesne University.

scholarship, distance, and achievement variables, the coefficient for the University sample is inverted from that of the consistent sign shown across the other segments. The main explanation for this variance is the individuality of each specific school. Because the students that are admitted to each school are so different, as evidenced through the differences in significance of explanatory variables, by combining them into one sample together, their expected behavior becomes difficult to gauge. In fact, it is not that surprising that the variables produce opposite than expected coefficients because the correlations of variables differ so greatly for each individual through the sample.

The overall effectiveness of the model was measured using an Expectation-Prediction Table. Coefficients were transformed using the model's density function to forecast the probability of enrollment for the entire sample. A cut-off value of .5 was set within the expectation test to measure how accurate the model is at predicting behavior. The table compares how many probabilities greater than .5 actually enrolled and how many less than .5 did not enroll. The percentages estimated correctly by the model over the different segments are displayed in Table 2.

Category	**% Correct**
Bottom SAT	62.91%
Top SAT	59.95%
Bottom GPA	65.71%
Top GPA	57.83%
Education	62.56%
Business	62.86%
Health Science	62.45%
Liberal Arts	66.87%
Music	68.00%
Nursing	68.42%
Science	62.46%
Pharmacy	68.90%

Table 2. Estimation Accuracy Percentages

Once the probability estimates were forecast for the sample, the forecasted probabilities were regressed against scholarship awards to establish a line of best fit through the observations that had been graphed. By establishing such a regression line, the model was then used to estimate a price elasticity based on how additional scholarship dollars affected the likelihood of enrollment. While the estimates are more accurate for the segments in which the scholarship variable had been decidedly significant, the results are displayed in Table 3 for all segments. The constant suggests a probability of enrollment based on no scholarship award while the slope captures the change in probability for every $1,000 of merit-based award offered.

Category	**Scholarship Constant**	**Slope**
Bottom SAT	0.4498	0.0180
Top SAT	0.4152	0.0018
Bottom GPA	0.4217	0.0198
Top GPA	0.4096	0.0045
Education	0.5117	0.0074
Business	0.4060	0.0122
Health Science	0.3888	0.0011
Liberal Arts	0.4127	0.0026
Music	0.3949	0.0130
Nursing	0.5215	0.0152
Science	0.4335	0.0039
Pharmacy	0.6441	-0.0098

Table 3. Best Fit Regression Lines

In terms of achievement, the slopes are clearly steeper for the lesser-achieving students further supporting the theory that that group is more likely to make decisions based on scholarship awards. The starting probability of enrollment is also higher in those groups. The variation within the slopes for the individual schools reflects the differing impacting characteristics across the segments. Liberal Arts for instance had a high significance for the scholarship variable in the original probit model, but has a small slope in terms of reaction to each additional $1,000 of aid. This is reflective upon the fact that Liberal Arts also had other significant deciding characteristics like resident status and distance. The same reasoning applies across each segment where the expected slope may be higher than the actual observed. The regression lines are graphed against each other for the schools and the achievement characteristics in the appendix.

VII. Implications of Results

The long-range importance of the model is to more efficiently analyze and predict the behavior of possible students given their characteristics and the award offered by the institution. Essentially, the model can be used as a predictor in order to best price scholarship awards which will compel enrollment of the students the institution wishes to attract. By understanding the probabilities associated with behavior, the institution can then limit the amount that it may overpay to certain students and use that money to offer additional awards to students previously not persuaded by limited resources. An efficient strategy can be implemented using information gathered through the model. Because the institution understands that not all students will enroll after an acceptance offer has been tendered, it must over-admit and potentially over-commit merit-based awards. The model can be used to admit an accurate amount in order to reach set quotas of students in the most efficient allocation of resources. The institution should select a probability of enrollment in which to set all students equal to. For instance, if the school decides to offer acceptance at a 60% probability of enrollment by offering the corresponding scholarship based on the model estimation, the institution will then admit exactly their enrollment quota for the given year divided by 60%. Additionally, because the institution understands that only 60% of their offers will be accepted, it can consequently offer scholarship awards in the amount of the total resources available divided by 60%. The institution could similarly use this information over individuals on average and accept at an average probability rate over a number of people equal to the quota divided by the average rather than setting a single probability in a blanket way.

VIII. Suggested Future Research

Higher education as a market is a highly studied phenomenon. Using the model developed, future research can make extensions on the findings here. In terms of the specific model itself, additional work can attempt to tighten the explanation of variance by attempting to identify additional student characteristics. These characteristics could include family demographics, including number of children in the family, whether parents attended college or not, and if so whether the student is a legacy at the given institution. Further examination of the quality of student would be advisable as well, attempting to capture additional extra-curricular tendencies in addition to the simple raw SAT and GPA data. By developing a variable that could capture well-roundedness through including participation in sports teams, community service, etc. the model could gather a more specific grasp of the student. Lastly, financial data specifying the background of the student's financial position could play a significant role. Using this additional data the probit model may be able to better predict probable behavior of the individuals.

Additional research can be done in regards to an individual institution's competitors. Duquesne for instance can profile students based on what type of individual applies to which of its competitors. If the institution can reasonably asses the likelihood a student has applied to a given competitor, it can use that information to transform its estimated probability of enrollment by a degree assigned to each school.

Lastly, the model can be used across samples from other Universities to discover how the applicants at different schools react to scholarship dollars. Most interestingly will be to examine the differences across different types of universities, varying in characteristics such as size, location, public or private status, etc. Then, the specific institution in question could use the model in their own admissions practices.

IX. Conclusions

This study focused on answering several questions. First, it is clear that there are significant differences in the mean scholarship award offered between the students who decide to enroll in the University and those who choose another alternative. Secondly, it is clear that those differences are affected by the segment of student being examined. Students who have achieved less academically are more prone to make decisions based on financial aid grants than those who are in the top third in terms of SAT scores and GPA. This suggests that students who have achieved more academically have more options in terms of their future prospects and therefore scholarship money makes less of a difference in the decision making. These findings are then supported once the probit model quantified specifically the impact financial aid dollars have on the probability an accepted student will enroll in the University. In addition to there being clear differences in terms of academically defined segments, schools with a less competitive nature, demonstrated by their lower tuition, have a more significant reliance on tuition as a predictor of behavior.

Practically, the developed probit model can be used to more efficiently dictate allocation of aid resources to best persuade students to enroll in the University. Because the school is forced to over-accept under the knowledge that students faced with alternative choices will in some cases chose those alternatives, the ability to predict the likelihood of such an occurrence leaves the institution at a great advantage. With specific indications of how students behave based on their academic prowess and their potential academic interest, the model creates a clear directive as to the most efficient allocation of resources. While the market exists in a constant flux of give and take between consumers and suppliers of higher education, the suppliers clearly have a significant tool in the form of merit-based aid to effectively develop a student body consisting of the students it wishes to have matriculate. The question then shifts from the importance of the tool to most effective usage, which can be answered by clearly identifying characteristics that explain behavior and utilizing that knowledge efficiently.

References

Avery C. and C. Hoxby 2002, Do and Should Financial Aid Packages Affect Students' College Choices?, paper presented at the NBER Conference on College Decisions, How Students Actually Make Them and How They Could (Bretton Woods, NH)

Dynarski, Susan. "Hope for Whom? Financial Aid for the Middle Class and Its Impact on College Attendance." National Tax Journal. September 2000, *53*(3), pp. 629– 61.

Dynarski, Susan. "Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion", working paper, Kennedy School of Government, Harvard University, 2001.

Dynarski, Susan. "The Behavioral and Distributional Implications of Aid for College." American Economic Review. May 2002.

Edlin, Aaron. "Is College Financial Aid Equitable and Efficient?" The Journal of Economic Perspectives, Volume 7, Number 2, Spring 1993, pp. 143-158.

Heller, Donald E. "The Effects of Tuition and State Financial Aid on Public College Enrollment" The Review of Higher Education - Volume 23, Number 1, Fall 1999, pp. 65-89.

Linsenmeier, David M., Harvey Rosen and Cecilia Rouse "Financial Aid Packages and College Enrollment Decisions: An Econometric Case Study" Princeton University Industrial Relations Section Working Paper No. 459, November 2001.

van der Klaauw, H. Wilbert, "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach" . International Economic Review, Vol. 43, pp. 1249-1287, 2002.

Appendix

Base University Model

Dependent Variable: ENROLL
Method: ML - Binary Probit
Date: 12/03/05 Time: 17:21
Sample: 1 1000
Included observations: 988
Excluded observations: 12
Convergence achieved after 6 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-1.340191	0.438419	-3.056870	0.0022
MALE	-0.156686	0.088528	-1.769902	0.0767
RESIDENT	-0.416422	0.143762	-2.896604	0.0038
LOG(SCHOLARSHIP)*DUM	-0.056908	0.016491	-3.450872	0.0006
LOG(ABS(15219-ZIP))	0.095171	0.017488	5.442229	0.0000
SAT	0.000523	0.000361	1.449972	0.1471
GPA	0.216273	0.106290	2.034743	0.0419

Mean dependent var	0.527328	S.D. dependent var	0.499505
S.E. of regression	0.482297	Akaike info criterion	1.321287
Sum squared resid	228.1906	Schwarz criterion	1.355973
Log likelihood	-645.7159	Hannan-Quinn criter.	1.334478
Restr. log likelihood	-683.3530	Avg. log likelihood	-0.653559
LR statistic (6 df)	75.27421	McFadden R-squared	0.055077
Probability(LR stat)	3.38E-14		

Obs with Dep=0	467	Total obs	988
Obs with Dep=1	521		

Bottom_Third GPA

Dependent Variable: ENROLL
Method: ML - Binary Probit
Date: 12/03/05 Time: 16:29
Sample: 1 788
Included observations: 770
Excluded observations: 18
Convergence achieved after 6 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	3.413491	0.688208	4.959966	0.0000
MALE	-0.219895	0.101206	-2.172754	0.0298
RESIDENT	0.367817	0.151885	2.421683	0.0154
LOG(SCHOLARSHIP)*DUM	0.091503	0.012763	7.169166	0.0000
LOG(ABS(15219-ZIP))	-0.104675	0.019245	-5.438981	0.0000
SAT	-0.001017	0.000416	-2.447519	0.0144
GPA	-0.688374	0.189261	-3.637166	0.0003

Mean dependent var	0.497403	S.D. dependent var	0.500318
S.E. of regression	0.465190	Akaike info criterion	1.252815
Sum squared resid	165.1145	Schwarz criterion	1.295055
Log likelihood	-475.3337	Hannan-Quinn criter.	1.269071
Restr. log likelihood	-533.7129	Avg. log likelihood	-0.617316
LR statistic (6 df)	116.7585	McFadden R-squared	0.109383
Probability(LR stat)	0.000000		

Obs with Dep=0	387	Total obs	770
Obs with Dep=1	383		

Bottom_Third SAT

Dependent Variable: ENROLL
Method: ML - Binary Probit
Date: 12/03/05 Time: 17:12
Sample: 1 788
Included observations: 771
Excluded observations: 17
Convergence achieved after 6 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	2.639808	0.710932	3.713165	0.0002
MALE	0.009059	0.100340	0.090283	0.9281
RESIDENT	0.389585	0.139604	2.790640	0.0053
DUM*LOG(SCHOLARSHIP)	0.087258	0.012799	6.817458	0.0000
LOG(ABS(15219-ZIP))	-0.071478	0.019562	-3.653881	0.0003
SAT	-0.001192	0.000673	-1.772630	0.0763
GPA	-0.462441	0.121349	-3.810839	0.0001

Mean dependent var	0.518807	S.D. dependent var	0.499971
S.E. of regression	0.474781	Akaike info criterion	1.291004
Sum squared resid	172.2188	Schwarz criterion	1.333201
Log likelihood	-490.6819	Hannan-Quinn criter.	1.307242
Restr. log likelihood	-533.8710	Avg. log likelihood	-0.636423
LR statistic (6 df)	86.37809	McFadden R-squared	0.080898
Probability(LR stat)	2.22E-16		

Obs with Dep=0	371	Total obs	771
Obs with Dep=1	400		

Business

Dependent Variable: ENROLL
Method: ML - Binary Probit
Date: 11/19/05 Time: 22:09
Sample: 1 451
Included observations: 447
Excluded observations: 4
Convergence achieved after 6 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	2.324210	0.637307	3.646924	0.0003
MALE	-0.122089	0.132578	-0.920883	0.3571
RESIDENT	0.237616	0.189993	1.250657	0.2111
DUM*LOG(SCHOLARSHIP)	0.098898	0.019494	5.073309	0.0000
LOG(ABS(15219-ZIP))	-0.058451	0.024029	-2.432493	0.0150
SAT	-0.001461	0.000547	-2.672270	0.0075
GPA	-0.304508	0.138566	-2.197559	0.0280

Mean dependent var	0.472036	S.D. dependent var	0.499777
S.E. of regression	0.480454	Akaike info criterion	1.322087
Sum squared resid	101.5679	Schwarz criterion	1.386333
Log likelihood	-288.4864	Hannan-Quinn criter.	1.347416
Restr. log likelihood	-309.1373	Avg. log likelihood	-0.645384
LR statistic (6 df)	41.30176	McFadden R-squared	0.066802
Probability(LR stat)	2.53E-07		

Obs with Dep=0	236	Total obs	447
Obs with Dep=1	211		

Education

Dependent Variable: ENROLL
Method: ML - Binary Probit
Date: 11/19/05 Time: 22:06
Sample: 1 196
Included observations: 195
Excluded observations: 1
Convergence achieved after 6 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	2.425742	0.917570	2.643660	0.0082
MALE	-0.006734	0.211526	-0.031835	0.9746
RESIDENT	0.519945	0.262769	1.978712	0.0478
DUM*LOG(SCHOLARSHIP)	0.100077	0.034764	2.878764	0.0040
LOG(ABS(15219-ZIP))	-0.091030	0.037649	-2.417843	0.0156
SAT	-0.001062	0.000878	-1.209925	0.2263
GPA	-0.417221	0.251420	-1.659462	0.0970

Mean dependent var	0.553846	S.D. dependent var	0.498372
S.E. of regression	0.473085	Akaike info criterion	1.310336
Sum squared resid	42.07621	Schwarz criterion	1.427828
Log likelihood	-120.7577	Hannan-Quinn criter.	1.357907
Restr. log likelihood	-134.0307	Avg. log likelihood	-0.619270
LR statistic (6 df)	26.54600	McFadden R-squared	0.099030
Probability(LR stat)	0.000176		
Obs with Dep=0	87	Total obs	195
Obs with Dep=1	108		

Health Sciences

Dependent Variable: ENROLL
Method: ML - Binary Probit
Date: 11/19/05 Time: 22:03
Sample: 1 246
Included observations: 245
Excluded observations: 1
Convergence achieved after 6 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.941782	0.900106	1.046301	0.2954
MALE	0.056230	0.211696	0.265618	0.7905
RESIDENT	0.085177	0.390898	0.217901	0.8275
DUM*LOG(SCHOLARSHIP)	0.056011	0.049869	1.123161	0.2614
LOG(ABS(15219-ZIP))	-0.100913	0.037807	-2.669136	0.0076
SAT	-0.000811	0.000809	-1.001868	0.3164
GPA	-0.016367	0.030744	-0.532367	0.5945

Mean dependent var	0.395918	S.D. dependent var	0.490048
S.E. of regression	0.484040	Akaike info criterion	1.350168
Sum squared resid	55.76204	Schwarz criterion	1.450203
Log likelihood	-158.3955	Hannan-Quinn criter.	1.390452
Restr. log likelihood	-164.4739	Avg. log likelihood	-0.646512
LR statistic (6 df)	12.15671	McFadden R-squared	0.036956
Probability(LR stat)	0.058563		
Obs with Dep=0	148	Total obs	245
Obs with Dep=1	97		

Liberal Arts

Dependent Variable: ENROLL
Method: ML - Binary Probit
Date: 11/19/05 Time: 22:01
Sample: 1 666
Included observations: 655
Excluded observations: 11
Convergence achieved after 6 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	2.151654	0.514660	4.180726	0.0000
MALE	-0.068602	0.108049	-0.634910	0.5255
RESIDENT	0.686328	0.171035	4.012804	0.0001
DUM*LOG(SCHOLARSHIP)	0.055308	0.016711	3.309714	0.0009
LOG(ABS(15219-ZIP))	-0.061896	0.021260	-2.911446	0.0036
SAT	-0.000848	0.000438	-1.937350	0.0527
GPA	-0.414435	0.128560	-3.223670	0.0013

Mean dependent var	0.425954	S.D. dependent var	0.494865
S.E. of regression	0.471665	Akaike info criterion	1.286438
Sum squared resid	144.1593	Schwarz criterion	1.334365
Log likelihood	-414.3083	Hannan-Quinn criter.	1.305021
Restr. log likelihood	-446.8025	Avg. log likelihood	-0.632532
LR statistic (6 df)	64.98834	McFadden R-squared	0.072726
Probability(LR stat)	4.34E-12		
Obs with Dep=0	376	Total obs	655
Obs with Dep=1	279		

Music

Dependent Variable: ENROLL
Method: ML - Binary Probit
Date: 11/19/05 Time: 21:57
Sample: 1 110
Included observations: 100
Excluded observations: 10
Convergence achieved after 6 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.495407	1.713626	-0.289099	0.7725
MALE	0.340819	0.293680	1.160512	0.2458
RESIDENT	0.222095	0.574531	0.386568	0.6991
DUM*LOG(SCHOLARSHIP)	-0.035217	0.069852	-0.504171	0.6141
LOG(ABS(15219-ZIP))	-0.146759	0.057932	-2.533303	0.0113
SAT	0.002725	0.001314	2.073438	0.0381
GPA	-0.482151	0.372553	-1.294181	0.1956

Mean dependent var	0.470000	S.D. dependent var	0.501614
S.E. of regression	0.466867	Akaike info criterion	1.324262
Sum squared resid	20.27074	Schwarz criterion	1.506624
Log likelihood	-59.21311	Hannan-Quinn criter.	1.398067
Restr. log likelihood	-69.13461	Avg. log likelihood	-0.592131
LR statistic (6 df)	19.84299	McFadden R-squared	0.143510
Probability(LR stat)	0.002953		
Obs with Dep=0	53	Total obs	100
Obs with Dep=1	47		

Nursing

Dependent Variable: ENROLL
Method: ML - Binary Probit
Date: 11/19/05 Time: 21:55
Sample: 1 97
Included observations: 95
Excluded observations: 2
Convergence achieved after 7 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	3.584550	1.555907	2.303833	0.0212
MALE	0.094911	0.552092	0.171911	0.8635
RESIDENT	0.118325	0.644352	0.183635	0.8543
DUM*LOG(SCHOLARSHIP)	0.183059	0.058537	3.127237	0.0018
LOG(ABS(15219-ZIP))	-0.231352	0.079590	-2.906815	0.0037
SAT	-0.001112	0.001590	-0.699442	0.4843
GPA	-0.506639	0.454508	-1.114698	0.2650

Mean dependent var	0.600000	S.D. dependent var	0.492497
S.E. of regression	0.446454	Akaike info criterion	1.240405
Sum squared resid	17.54026	Schwarz criterion	1.428585
Log likelihood	-51.91922	Hannan-Quinn criter.	1.316444
Restr. log likelihood	-63.93611	Avg. log likelihood	-0.546518
LR statistic (6 df)	24.03378	McFadden R-squared	0.187952
Probability(LR stat)	0.000515		

Obs with Dep=0	38	Total obs	95
Obs with Dep=1	57		

Pharmacy

Dependent Variable: ENROLL
Method: ML - Binary Probit
Date: 11/19/05 Time: 21:51
Sample: 1 256
Included observations: 254
Excluded observations: 2
Convergence achieved after 7 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	5.694163	1.194800	4.765787	0.0000
MALE	0.041441	0.186978	0.221638	0.8246
RESIDENT	0.100385	0.623539	0.160992	0.8721
DUM*LOG(SCHOLARSHIP)	0.037638	0.051368	0.732712	0.4637
LOG(ABS(15219-ZIP))	-0.169570	0.038933	-4.355465	0.0000
SAT	-0.002355	0.000834	-2.821964	0.0048
GPA	-0.448056	0.253655	-1.766403	0.0773

Mean dependent var	0.582677	S.D. dependent var	0.494091
S.E. of regression	0.457305	Akaike info criterion	1.251576
Sum squared resid	51.65460	Schwarz criterion	1.349062
Log likelihood	-151.9502	Hannan-Quinn criter.	1.290794
Restr. log likelihood	-172.5709	Avg. log likelihood	-0.598229
LR statistic (6 df)	41.24147	McFadden R-squared	0.119491
Probability(LR stat)	2.60E-07		

Obs with Dep=0	106	Total obs	254
Obs with Dep=1	148		

Science

Dependent Variable: ENROLL
Method: ML - Binary Probit
Date: 12/03/05 Time: 17:27
Sample: 1 296
Included observations: 293
Excluded observations: 3
Convergence achieved after 6 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.738801	0.891931	0.828316	0.4075
MALE	-0.098529	0.154250	-0.638764	0.5230
RESIDENT	0.179718	0.238939	0.752147	0.4520
DUM*LOG(SCHOLARSHIP)	0.047740	0.031599	1.510801	0.1308
LOG(ABS(15219-ZIP))	-0.080252	0.033444	-2.399606	0.0164
SAT	0.000378	0.000675	0.560385	0.5752
GPA	-0.307473	0.192202	-1.599740	0.1097

Mean dependent var	0.457338	S.D. dependent var	0.499029
S.E. of regression	0.493070	Akaike info criterion	1.383223
Sum squared resid	69.53184	Schwarz criterion	1.471145
Log likelihood	-195.6421	Hannan-Quinn criter.	1.418437
Restr. log likelihood	-202.0243	Avg. log likelihood	-0.667720
LR statistic (6 df)	12.76434	McFadden R-squared	0.031591
Probability(LR stat)	0.046935		
Obs with Dep=0	159	Total obs	293
Obs with Dep=1	134		

Top_GPA

Dependent Variable: ENROLL
Method: ML - Binary Probit
Date: 12/03/05 Time: 17:28
Sample: 1 788
Included observations: 785
Excluded observations: 3
Convergence achieved after 8 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.289138	0.962334	0.300455	0.7638
MALE	0.030652	0.099656	0.307576	0.7584
RESIDENT	0.186989	0.173119	1.080119	0.2801
DUM*LOG(SCHOLARSHIP)	0.115654	0.041263	2.802822	0.0051
LOG(ABS(15219-ZIP))	-0.084025	0.019942	-4.213561	0.0000
SAT	-0.000570	0.000392	-1.454069	0.1459
GPA	-0.057494	0.212866	-0.270094	0.7871

Mean dependent var	0.443312	S.D. dependent var	0.497093
S.E. of regression	0.487332	Akaike info criterion	1.344779
Sum squared resid	184.7690	Schwarz criterion	1.386384
Log likelihood	-520.8257	Hannan-Quinn criter.	1.360776
Restr. log likelihood	-539.0644	Avg. log likelihood	-0.663472
LR statistic (6 df)	36.47751	McFadden R-squared	0.033834
Probability(LR stat)	2.23E-06		
Obs with Dep=0	437	Total obs	785
Obs with Dep=1	348		

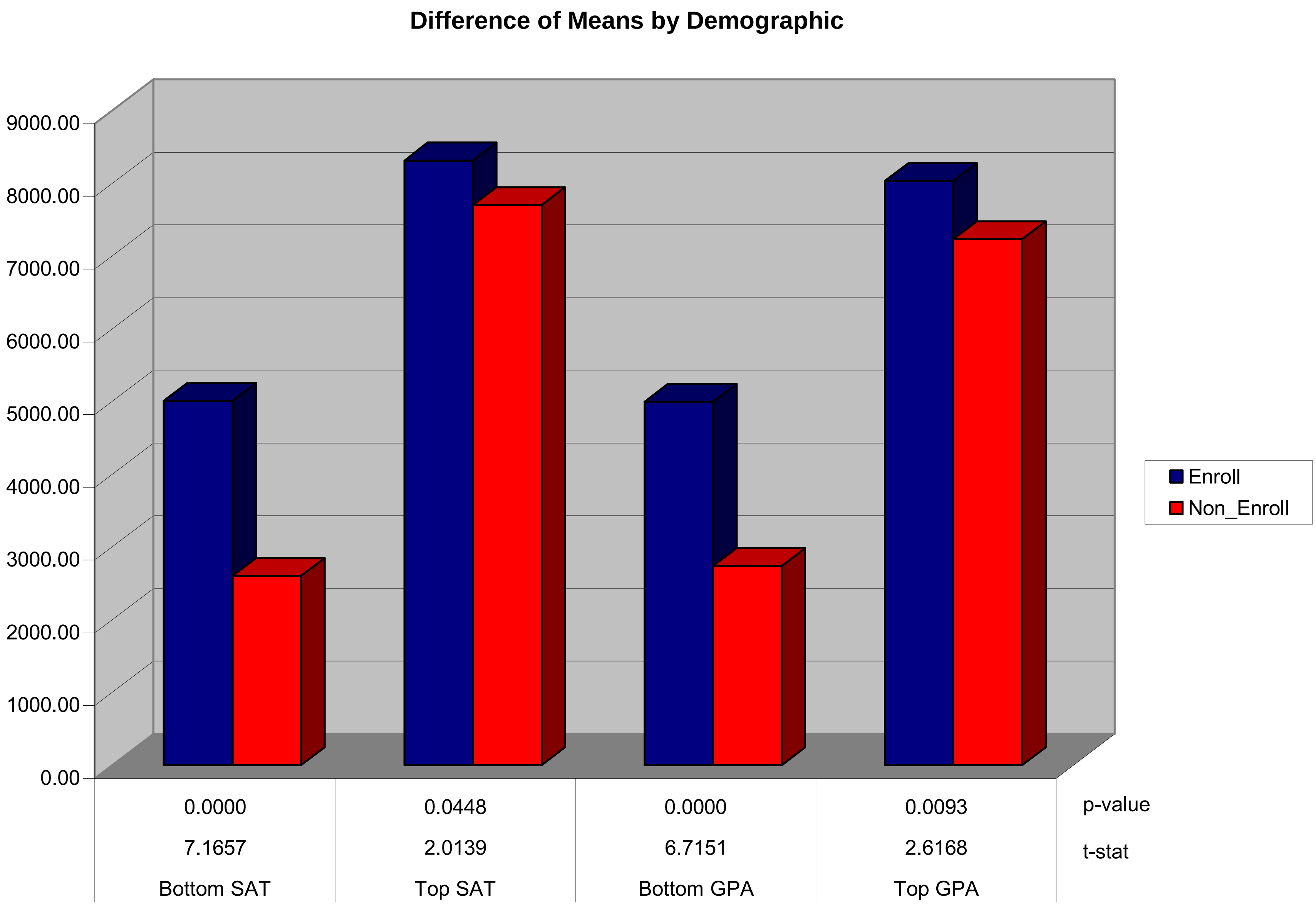
Top_SAT

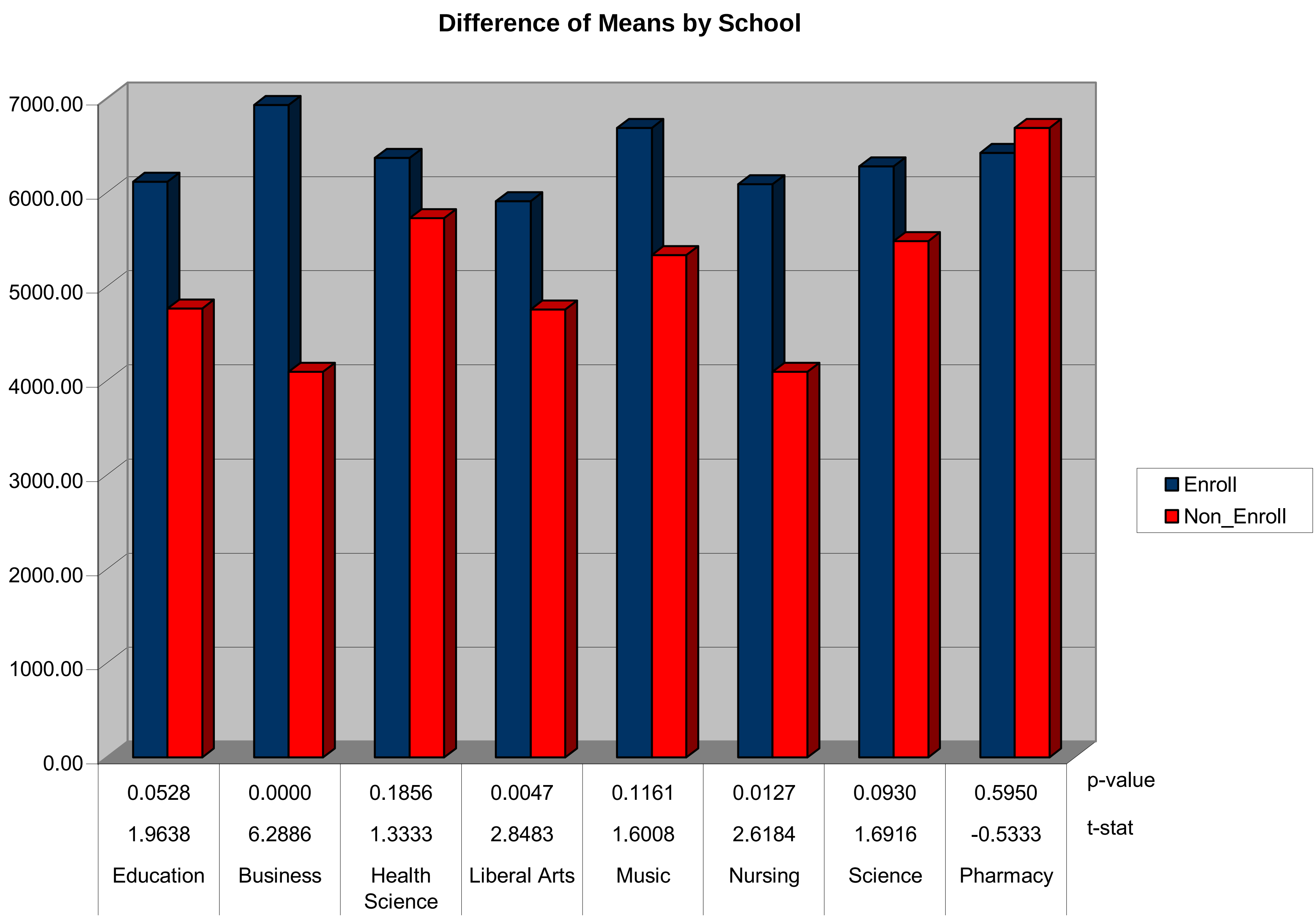
Dependent Variable: ENROLL
Method: ML - Binary Probit
Date: 12/03/05 Time: 17:30
Sample: 1 788
Included observations: 784
Excluded observations: 4
Convergence achieved after 7 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	1.406935	0.840927	1.673076	0.0943
MALE	-0.082373	0.096160	-0.856626	0.3917
RESIDENT	0.306605	0.196415	1.561006	0.1185
LOG(SCHOLARSHIP)*DUM	0.043111	0.031448	1.370865	0.1704
LOG(ABS(15219- ZIP))	-0.091278	0.019045	-4.792690	0.0000
SAT	-0.000662	0.000645	-1.026501	0.3047
GPA	-0.125926	0.112393	-1.120414	0.2625

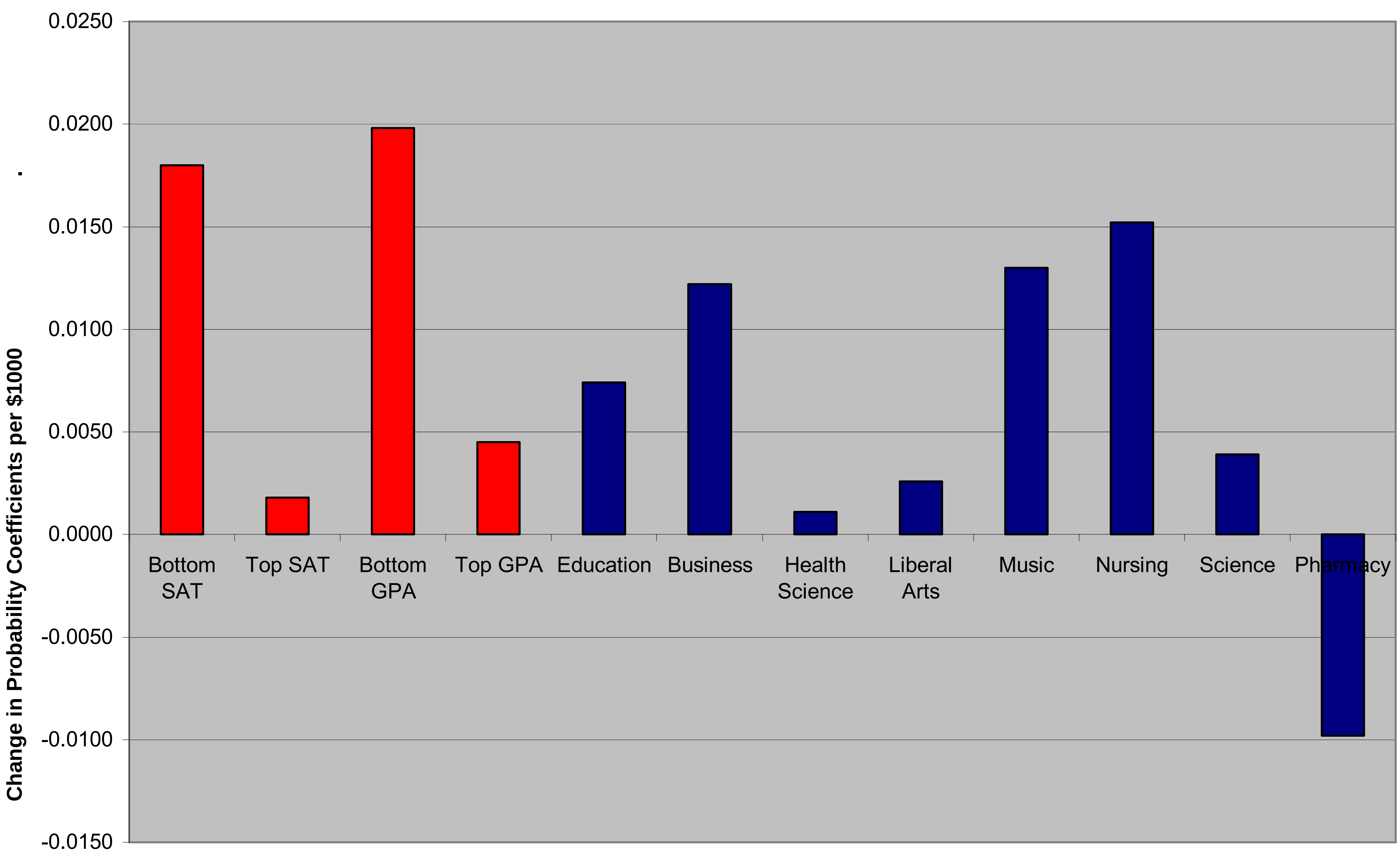
Mean dependent var	0.429847	S.D. dependent var	0.495370
S.E. of regression	0.485593	Akaike info criterion	1.337139
Sum squared resid	183.2170	Schwarz criterion	1.378785
Log likelihood	-517.1583	Hannan-Quinn criter.	1.353153
Restr. log likelihood	-535.6850	Avg. log likelihood	-0.659641
LR statistic (6 df)	37.05337	McFadden R-squared	0.034585
Probability(LR stat)	1.72E-06		

Obs with Dep=0	447	Total obs	784
Obs with Dep=1	337		

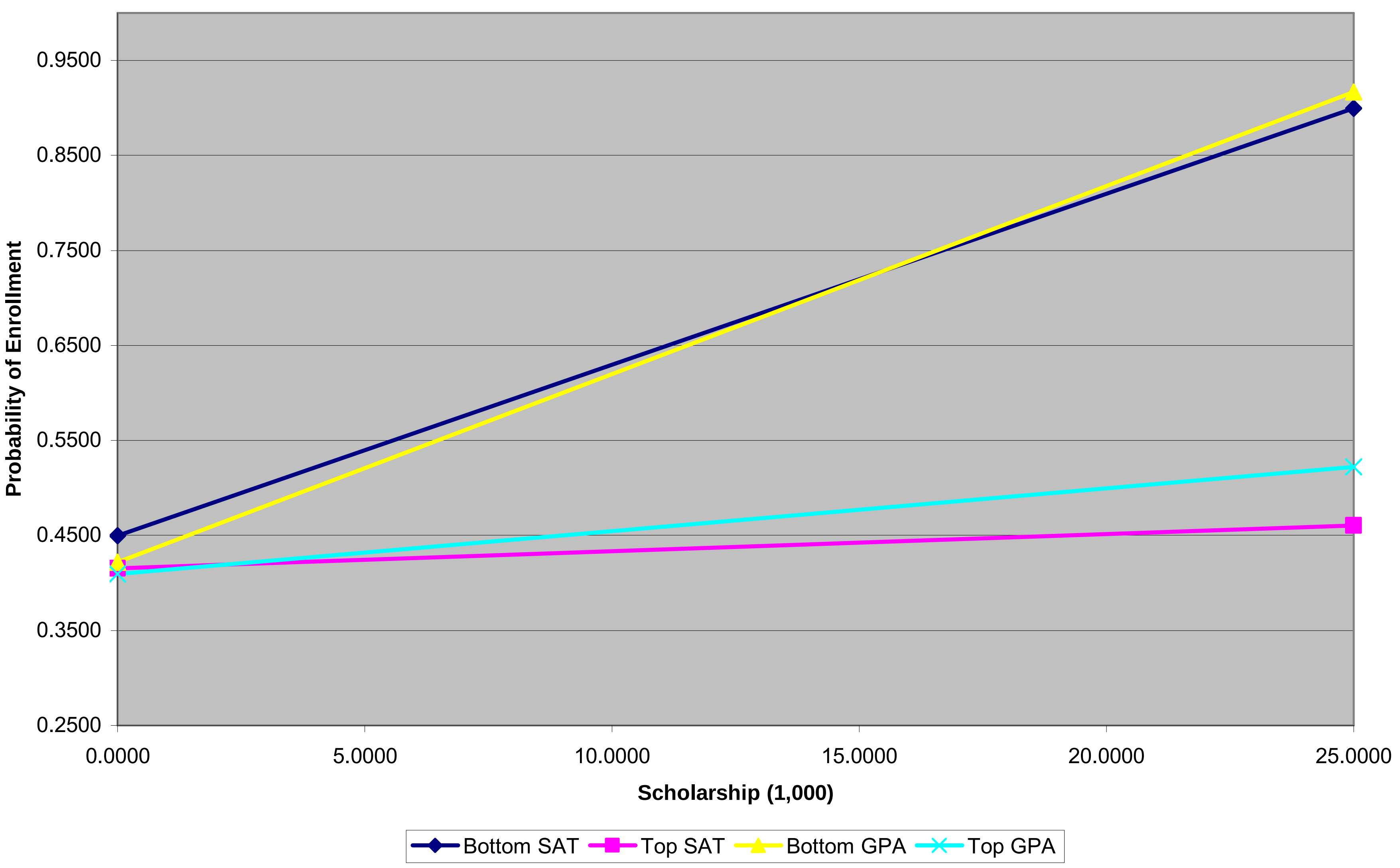

Difference of Means by Demographic
9000.00
8000.00
7000.00
6000.00
5000.00
4000.00
3000.00
2000.00
1000.00
0.00
Enroll
Non_Enroll
0.0000
0.0448
0.0000
0.0093
p-value
7.1657
2.0139
6.7151
2.6168
t-stat
Bottom SAT
Top SAT
Bottom GPA
Top GPA

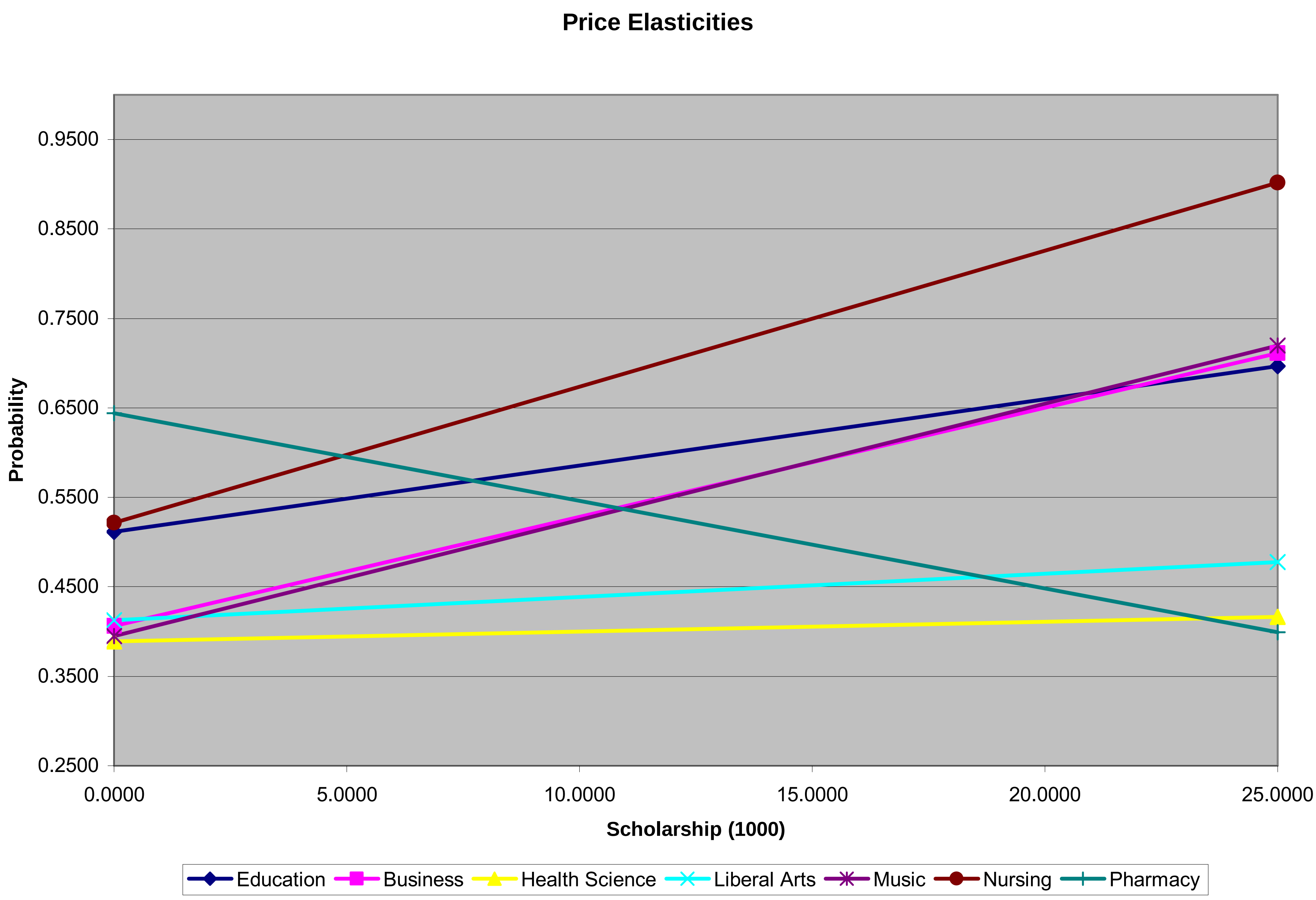

Difference of Means by School
7000.00
6000.00
5000.00
4000.00
3000.00
2000.00
1000.00
0.00
Enroll
Non_Enroll
p-value
t-stat
0.0528 0.0000 0.1856 0.0047 0.1161 0.0127 0.0930 0.5950
1.9638 6.2886 1.3333 2.8483 1.6008 2.6184 1.6916 -0.5333
Education
Business
Health Science
Liberal Arts
Music
Nursing
Science
Pharmacy

Sensitivity to Merit Based Aid



Price Elasticities




Price Elasticities
Probability
0.9500
0.8500
0.7500
0.6500
0.5500
0.4500
0.3500
0.2500
0.0000
5.0000
10.0000
15.0000
20.0000
25.0000
Scholarship (1000)
Education
Business
Health Science
Liberal Arts
Music
Nursing
Pharmacy